Kaplan & Sadock's Comprehensive Textbook of Psychiatry (2 Volume Set)
by Benjamin J. Sadock (Editor), Virginia A. Sadock (Editor)
Lippincott Williams & Wilkins Publishers; 7th edition (January 15, 2000)

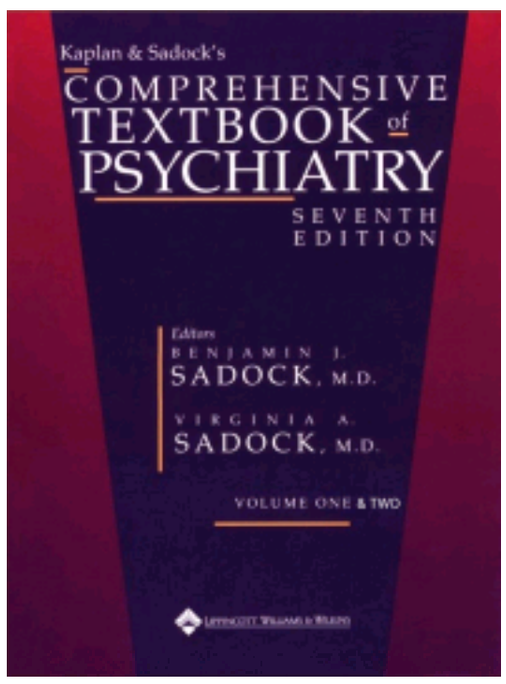

By OkDoKeY

Kaplan & Sadock's Comprehensive Textbook of Psychiatry

TABLE OF CONTENTS

VOLUME I

CHAPTER 1. NEURAL SCIENCES

CHAPTER 2. NEUROPSYCHIATRY AND BEHAVIORAL NEUROLOGY

I´m sorry, but this page is missing!

- OkDoKeY

ABOUT THE EDITORS

BENJAMIN J. SADOCK, M.D.

Benjamin James Sadock, M.D., is the Menas S. Gregory Professor of Psychiatry and Vice Chairman of the Department of Psychiatry at the New York University (NYU) School of Medicine. He was graduated from Union College, received his Doctor of Medicine degree from New York Medical College, and did his internship at Albany Hospital. He completed his residency at Bellevue Psychiatric Hospital and then entered military service where he served as assistant chief and acting chief of neuropsychiatry at Sheppard Air Force Base, Texas. He has held faculty and teaching appointments at Southwestern Medical School and Parkland Hospital in Dallas, and at New York Medical College, St. Luke's Hospital, the New York State Psychiatric Institute, and Metropolitan Hospital in New York City. He joined the faculty of the NYU School of Medicine in 1980 and served in various positions: director of medical student education in psychiatry, co-director of the Residency Training Program in Psychiatry, and director of Graduate Medical Education. Currently, Dr. Sadock is director of Student Mental Health Services, psychiatric consultant to the Admissions Committee and co-director of Continuing Education in Psychiatry at the NYU School of Medicine. He is on the staff of Bellevue Hospital and Tisch Hospital and is consultant psychiatrist at Lenox Hill Hospital. Dr. Sadock is a diplomate of the American Board of Psychiatry and Neurology and served as an assistant and associate examiner for the board for over a decade. He is a Fellow of the American Psychiatric Association, a Fellow of the American College of Physicians, a Fellow of the New York Academy of Medicine, and a member of Alpha Omega Alpha Honor Society. He is active in numerous psychiatric organizations and is president and founder of the NYU-Bellevue Psychiatric Society. Dr. Sadock was a member of the National Committee in Continuing Education in Psychiatry of the American Psychiatric Association, served on the Ad Hoc Committee on Sex Therapy Clinics of the American Medical Asso ciation, was delegate to the Conference on Recertification of the American Board of Medical Specialists, and was a representative of the American Psychiatric Association's Task Force on the National Board of Medical Examiners and the American Board of Psychiatry and Neurology. In 1985 he received the Academic Achievement Award from New York Medical College. He is author or editor of over 100 publications, a book reviewer for psychiatric journals, and lectures on a broad range of topics in general psychiatry. Dr. Sadock maintains a private practice for diagnostic consultations, psychotherapy, and pharmacotherapy. He has been married to Virginia Alcott Sadock, M.D., clinical professor of psychiatry at NYU School of Medicine, since completing his residency. Dr. Sadock enjoys opera, skiing, and traveling, and is an avid fly-fisherman.

VIRGINIA A. SADOCK, M.D.

Virginia Alcott Sadock, M.D., joined the faculty of the New York University (NYU) School of Medicine in 1980 where she is currently clinical professor of psychiatry and attending psychiatrist at Tisch Hospital and Bellevue Hospital. She is director of the Program in Human Sexuality and Sex Therapy at the NYU Medical Center, one of the largest treatment and training programs of its kind in the United States. She is the author of over 50 articles and chapters on sexual behavior and was the developmental editor of *The Sexual Experience*, published by Williams & Wilkins—one of the first major textbooks on human sexuality. She serves as referee and book reviewer for several medical journals including the *American Journal of Psychiatry* and the *Journal of the American Medical Association*. She has had a long-standing interest in the role of women in medicine and psychiatry and was a founder of the Committee on Women in Psychiatry of the New York County District Branch of the American Psychiatric Association. She is active in academic matters, has served as an assistant and associate examiner for the American Board of Psychiatry and Neurology for over 15 years, and was also a member of the test committee in psychiatry for both the American Board of Psychiatry and the Psychiatric Knowledge and Self-Assessment Program (PKSAP) of the American Psychiatric Association. She served as chairperson of the Committee on Public Relations. New York County District Branch of the American Psychiatric Association and also participated in the National Medical Television Network series *Women in Medicine* and the PBS television documentary *Women and Depression*. She has been vice-president of the Society of Sex Therapy and Research, a regional council member of the American Association of Sex Education Counselors and Therapists, and is president of the Alumni Association of Sex Therapists. She lectures extensively both in this country and abroad on sexual dysfunction, relational problems, and dep ression and anxiety disorders. She is a Fellow of the American Psychiatric Association, a Fellow of the New York Academy of Medicine and a diplomate of the American Board of Psychiatry and Neurology. Dr. Sadock was graduated from Bennington College, received her M.D. from New York Medical College, and trained in psychiatry at Metropolitan Hospital. She lives in Manhattan with her husband, Dr. Benjamin Sadock, where she maintains an active psychiatric practice that includes individual psychotherapy, couples and marital therapy, sex therapy, psychiatric consultation, and pharmacotherapy. They have two children, James and Victoria, both physicians. In her leisure time Dr. Sadock enjoys theater, film, reading fiction, and travel.

CONTRIBUTORS

Henry David Abraham, M.D.
Clinical Associate Professor of Psychiatry, Brown University School of Medicine; Chief of Alcohol and Drug Treatment Services, Butler Hospital, Providence, Rhode Island; Lecturer in Psychiatry, Harvard Medical School; Consultant in Psychiatry, Massachusetts General Hospital, Boston, Massachusetts.
11.7. Hallucinogen-Related Disorders

Russell L. Adams, Ph.D.
Professor of Psychiatry and Behavioral Science, University of Oklahoma College of Medicine; Director of University Hospital Neuropsychology Assessment Laboratory; Director of Psychology Internship and Post Doctoral Neuropsychology Training Program, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma.
7.5. Personality Assessment: Adults and Children

W. Stewart Agras, M.D., F.R.C.P.(C)
Professor of Psychiatry and Behavioral Sciences, Stanford University School of Medicine; Director of Stanford University Psychiatric Clinics, Stanford, California.
3.3. Learning Theory

Hagop S. Akiskal, M.D.
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Director of Outpatient Psychiatric Services, San Diego Veterans Affairs Healthcare System, San Diego, California.
14.1. Mood Disorders: Introduction and Overview
14.6. Mood Disorders: Clinical Features

George S. Alexopoulos, M.D.
Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University, New York, New York; Director of Cornell Institute of Geriatric Psychiatry, New York Hospital-Cornell Medical Center, Westchester Division, White Plains, New York.
51.3d. Geriatric Psychiatry: Mood Disorders

Anne Alonso, Ph.D.
Clinical Professor of Psychology, Harvard Medical School; Director of Center for Psychoanalytic Studies at the Massachusetts General Hospital, Boston, Massachusetts; Professor, The Fielding Institute, Santa Barbara, California.
30.4. Group Psychotherapy, Combined Individual and Group Psychotherapy

Kenneth Z. Altshuler, M.D.
Stanton Sharp Professor of Psychiatry and Chairman, Department of Psychiatry, University of Texas Southwestern Medical School, Dallas, Texas.
30.10. Other Methods of Psychotherapy

Lisa Amaya-Jackson, M.D.
Assistant Professor of Psychiatry and the Behavioral Sciences, Duke University School of Medicine; Director of Trauma Evaluation, Treatment and Research Program, Center for Child and Family Health, Durham, North Carolina.
46.2. Posttraumatic Stress Disorder in Children and Adolescents

Daniel G. Amen, M.D.
Director of Amen Clinic for Behavioral Medicines, Fairfield, California.
2.13. Neuroimaging in Clinical Practice

George W. Arana, M.D.
Professor of Psychiatry and Behavioral Sciences, Associate Dean of College of Medicine, Medical University of South Carolina; Director of Mental Health Service, Ralph H. Johnson Veterans Affairs Medical Center, Charleston, South Carolina.
31.9. Barbiturates and Similarly Acting Substances
31.14. Chloral Hydrate

Lesley M. Arnold, M.D.
Assistant Professor of Psychiatry, Director of Division of Psychosomatic Research, University of Cincinnati College of Medicine, Cincinnati, Ohio.
25.7. Psychocutaneous Disorders

J. Hampton Atkinson, Jr., M.D.
Adjunct Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Staff Psychiatrist, San Diego Veterans Affairs Healthcare System, San Diego, California.
2.8. Neuropsychiatric Aspects of HIV Infection and AIDS

F. M. Baker, M.D., M.P.H.
Professor of Psychiatry, University of Hawaii John A. Burns School of Medicine; Chief of Inpatient Psychiatry, Hawaii State Hospital, Kaneohe, Hawaii.
51.6e. Geriatric Psychiatry: Minority Issues

Glen B. Baker, Ph.D.
Professor of Psychiatry, Co-Director of Neurochemical Research Unit, University of Alberta Faculty of Medicine and Dentistry, Edmonton, Alberta, Canada.
31.20. Monoamine Oxidase Inhibitors

Lorian Baker, Ph.D.
Former Research Professor of Child Psychiatry, University of California at Los Angeles School of Medicine, Los Angeles, California.
35.1. Reading Disorder
35.2. Mathematics Disorder
35.3. Disorder of Written Expression and Learning Disorder Not Otherwise Specified
36. Motor Skills Disorder: Developmental Coordination Disorder

James C. Ballenger, M.D.
Professor of Psychiatry and Chairman, Department of Psychiatry and Behavioral Sciences, Director of Institute of Psychiatry, Medical University of South Carolina, Charleston, South Carolina.
31.9. Barbiturates and Similarly Acting Substances
31.10. Benzodiazepine Receptor Agonists and Antagonists
31.12. Buspirone
31.14. Chloral Hydrate

Jay M. Baraban, M.D., Ph.D.
Associate Professor of Neuroscience Psychiatry, and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland.
1.8. Intraneuronal Signaling Pathways

Ida Sue Baron, Ph.D.
Clinical Associate Professor of Pediatrics, Georgetown University School of Medicine and Health Sciences, Washington, D.C.

7.6. Neuropsychological and Intellectual Assessment of Children

William W. Beatty, Ph.D.
Professor of Psychiatry and Behavioral Science, University of Oklahoma College of Medicine, Oklahoma City, Oklahoma.
2.7. Neuropsychiatric Aspects of Multiple Sclerosis and Other Demyelinating Disorders

Linda Beauclair, M.D.
Assistant Professor of Psychiatry
McGill University Faculty of Medicine, Montreal, Quebec, Canada.
31.24. Selective Serotonin-Noradrenaline Reuptake Inhibitors

Aaron T. Beck, M.D.
Professor Emeritus of Psychiatry and Director of Psychopathology Research Center, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
30.6. Cognitive Therapy

Anne E. Becker, M.D., Ph.D.
Assistant Professor of Medical Anthropology and Psychiatry
Harvard Medical School, Boston, Massachusetts.
4.1. Anthropology and Psychiatry

Joseph H. Beitchman, M.D.
Professor of Psychiatry and Public Health Science, University of Toronto Faculty of Medicine; Head of Child and Family Studies Centre, Clarke Institute of Psychiatry, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
37.1. Expressive Language Disorder
37.2. Mixed Receptive-Expressive Language Disorder
37.3. Phonological Disorder
37.4. Stuttering
37.5. Communication Disorder Not Otherwise Specified

Marilyn B. Benoit, M.D.
Clinical Associate Professor of Psychiatry, Georgetown University School of Medicine and Health Sciences, Washington, D.C.
49.3. Foster Care

Kate A. Berg, Ph.D.
Co-Director of Office of Bioethics and Special Populations Research, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland.
28.4. Genetic Counseling

Sarah L. Berga, M.D.
Associate Professor of Obstetrics, Gynecology, Reproductive Sciences, and Psychiatry, University of Pittsburgh School of Medicine; Attending Physician, Magee-Women's Hospital, Pittsburgh, Pennsylvania.
28.2. Psychiatry and Reproductive Medicine

R. Lindsey Bergman, Ph.D.
Clinical Instructor of Psychiatry and Biobehavioral Science, University of California at Los Angeles School of Medicine; Associate Director of Child and Adolescent Obsessive-Compulsive Disorder and Anxiety Program, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
46.1. Obsessive-Compulsive Disorder in Children

William Bernet, M.D.
Associate Professor of Psychiatry, Vanderbilt University School of Medicine; Medical Director of Psychiatric Hospital at Vanderbilt, Nashville, Tennessee.
49.4. Child Maltreatment

Garth Bissette, Ph.D.
Professor of Psychiatry, University of Mississippi School of Medicine, Jackson, Mississippi.
1.6. Neuropeptides: Biology and Regulation

Ira B. Black, M.D.
Professor and Chair, Department of Neuroscience and Cell Biology, University of Medicine and Dentistry of New Jersey Robert Wood Johnson Medical School, Piscataway, New Jersey.
1.7. Neurotrophic Factors

Deborah Blacker, M.D., Sc.D.
Assistant Professor of Psychiatry, Harvard Medical School; Assistant Professor of Epidemiology, Harvard School of Public Health; Clinical Staff, Massachusetts General Hospital; Boston, Massachusetts.
7.8. Psychiatric Rating Scales

Barry Blackwell, M.D., F.R.C.Psych.
Clinical Professor of Psychiatry, University of Wisconsin Medical School, Milwaukee Campus, Milwaukee, Wisconsin.
27.1. Treatment Compliance

Ray Blanchard, Ph.D.
Associate Professor of Psychiatry, University of Toronto Faculty of Medicine; Head of Clinical Sexology Program, Clarke Institute of Psychiatry, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
19.3. Gender Identity Disorders

Dan G. Blazer, II, M.D., Ph.D.
J.P. Gibbons Professor of Psychiatry and Behavioral Sciences, Professor of Community and Family Medicine and Dean of Medical Education, Duke University School of Medicine, and Durham, North Carolina.
14.2. Mood Disorders: Epidemiology

Bennett Blum, M.D.
Clinical Instructor of Geriatric Psychiatry, University of California at Los Angeles School of Medicine; Director of Forensic Gero-Psychiatry Division, Park Dietz and Associates, Newport Beach, California.
51.6b. Geriatric Psychiatry: Forensic Issues

Michael P. Bogenschutz, M.D.
Assistant Professor of Psychiatry, University of New Mexico School of Medicine; Clinical Director of Dual Diagnosis Treatment Program, Mental Health Center, University of New Mexico Health Sciences Center, Albuquerque, New Mexico.
9.1. Classification of Mental Disorders

Kyle Brauer Boone, Ph.D.
Associate Professor-in-Residence, University of California at Los Angeles School of Medicine, Los Angeles, California; Director of Neuropsychological Services,

Department of Psychiatry, Harbor-UCLA Medical Center, Torrance, California.
51.2d. Geriatric Psychiatry: Neuropsychological Evaluation

Neil W. Boris, M.D.
Assistant Professor of Community Health Sciences, Tulane University School of Public Health and Tropical Medicine; Assistant Professor of Pediatrics, Psychiatry, and Neurology, Tulane University School of Medicine, New Orleans, Louisiana.
44.1. Reactive Attachment Disorder of Infancy and Early Childhood

Soo Borson, M.D.
Professor of Psychiatry and Behavioral Sciences, University of Washington School of Medicine; Director of Geropsychiatry Services and Director of University of Washington Medical Center Alzheimer's Disease Research Center, Seattle, Washington.
51.3a. Geriatric Psychiatry: Psychiatric Problems in the Medically Ill;
51.3b. Geriatric Psychiatry: Sleep Disorders

H. Stefan Bracha, M.D.
Research Physician, National Center for Posttraumatic Stress Disorder, Honolulu Veterans Affairs Medical and Regional Office Center, Honolulu, Hawaii.
2.13. Neuroimaging in Clinical Practice

Olga Brawman-Mintzer, M.D.
Assistant Professor of Psychiatry and Behavioral Sciences, Medical University of South Carolina, Charleston, South Carolina.
31.12. Buspirone

Kirk J. Brower, M.D.
Associate Professor of Psychiatry, Director of Alcohol Research Center, University of Michigan Medical School, Ann Arbor, Michigan.
11.13. Anabolic-Androgenic Steroid Abuse

Kelly D. Brownell, Ph.D.
Professor of Psychology, Master of Silliman College, Yale University; Professor of Epidemiology and Public Health and Director of Clinical Training, Director of Yale Center for Eating and Weight Disorders, Yale University School of Medicine, New Haven, Connecticut.
25.3. Obesity

Robert W. Buchanan, M.D.
Associate Professor of Psychiatry
University of Maryland School of Medicine, Baltimore, Maryland; Chief of Outpatient Research Program, Maryland Psychiatric Research Center, Catonsville, Maryland.
12.1. Schizophrenia: Introduction and Overview

Oscar G. Bukstein, M.D., M.P.H.
Associate Professor of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.
49.11. Adolescent Substance Abuse

Jack D. Burke, Jr., M.D., M.P.H.
Professor of Psychiatry and Head, Department of Psychiatry and Behavioral Science, Texas A & M University College of Medicine; Chairman, Department of Psychiatry, Scott and White Memorial Hospital and Clinic, Temple, Texas.
5.1. Epidemiology

Vivien K. Burt, M.D., Ph.D.
Associate Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Women's Life Center, UCLA Neuropsychiatric Institute and Hospital; Medical Director, Mental Health Clinic, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
22. Impulse-Control Disorders Not Elsewhere Classified

Juan Bustillo, M.D.
Assistant Professor of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.
12.9. Schizophrenia: Psychosocial Treatment

Eric D. Caine, M.D.
John Romano Professor of Psychiatry and Chair, Department of Psychiatry, University of Rochester School of Medicine and Dentistry, Rochester, New York.
10. Delirium, Dementia, and Amnestic and Other Cognitive Disorders

Robert Jean Campbell, III, M.D.
Clinical Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University; Medical Director, New York Gracie Square Hospital; Attending Psychiatrist, New York Hospital, New York, New York.
52.2. Managed Care

Robert Cancro, M.D., Med.D.Sc.
Lucius N. Littauer Professor of Psychiatry and Chairman, Department of Psychiatry, New York University School of Medicine; Director of Department of Psychiatry, Tisch Hospital, New York, New York; Director of Nathan Kline Institute for Psychiatric Research, Orangeburg, New York.
12.7. Schizophrenia: Clinical Features

Dennis P. Cantwell, M.D.***
Joseph Campbell Professor of Child Psychiatry, Department of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
35.1. Reading Disorder;
35.2. Mathematics Disorder;
35.3. Disorder of Written Expression and Learning Disorder Not Otherwise Specified
36. Motor Skills Disorder: Developmental Coordination Disorder

William T. Carpenter, Jr., M.D.
Professor of Psychiatry and Pharmacology, University of Maryland School of Medicine, Baltimore, Maryland; Director of Maryland Psychiatric Research Center, Catonsville, Maryland.
12.1. Schizophrenia: Introduction and Overview

Christine K. Cassel, M.D.
Professor of Psychiatry and Chairman, Department of Geriatrics and Adult Development, Mount Sinai School of Medicine, New York, New York.
51.6c. Geriatric Psychiatry: Ethical Issues

Irene Chatoor, M.D.
Professor of Psychiatry and Behavioral Sciences, George Washington University School of Medicine and Health Sciences; Interim Chair and Director of Infant Psychiatry, Children's National Medical Center, Washington, D.C.
41. Feeding and Eating Disorders of Infancy and Early Childhood

Guy Chouinard, M.D., M.Sc.
Professor of Psychiatry, McGill University Faculty of Medicine; Professor of Psychiatry, University of Montreal Faculty of Medicine; Director of Clinical Psychopharmacology Unit, Allan Memorial Institute; Head of Clinical Psychopharmacology Inpatient Unit, Fernand-Seguin Research Centre, Louis-H. Lafontaine Hospital, Montreal, Quebec, Canada.
31.24. Selective Serotonin-Noradrenaline Reuptake Inhibitors

Tiffany W. Chow, M.D.
Clinical Instructor of Neurology, University of California at Los Angeles School of Medicine; Director of Frontotemporal Dementia Clinic, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
2.1. Neuropsychiatry: Clinical Assessment and Approach to Diagnosis

Domenic A. Ciraulo, M.D.
Professor of Psychiatry and Chairman, Boston University School of Medicine; Psychiatrist-in-Chief, Boston Medical Center; Chief of Psychiatry Service, Veterans Affairs Boston Clinic, Boston, Massachusetts.
11.12. Sedative-, Hypnotic-, or Anxiolytic-Related Abuse

James L. Claghorn, M.D.
Clinical Associate Professor of Psychiatry, University of Texas Medical School at Houston, Houston, Texas.
31.19. Mirtazapine

C. Robert Cloninger, M.D.
Wallace Renard Professor of Psychiatry, Washington University School of Medicine, Saint Louis, Missouri.
24. Personality Disorders

Barbara J. Coffey, M.D.
Assistant Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts; Director of Tourette's Clinic, Director of Pediatric Psychopharmacology Clinic, McLean Hospital, Belmont, Massachusetts.
48.6. Pediatric Psychopharmacology

Lee S. Cohen, M.D.
Associate Professor of Psychiatry, Harvard Medical School; Director of Perinatal and Reproductive Psychiatry, Clinical Research Program, Clinical Psychopharmacology Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts.
13.4. Postpartum Psychiatric Syndromes

Calvin A. Colarusso, M.D.
Clinical Professor of Psychiatry and Director of Child Psychiatry Residency Training Program, University of California at San Diego School of Medicine, La Jolla, California; Training and Supervising Analyst, San Diego Psychoanalytic Institute, San Diego, California.
50. Adulthood

Ralph Colp, Jr., M.D.
Assistant Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Senior Attending Psychiatrist, St. Luke's-Roosevelt Hospital Center, New York, New York.
55.1. History of Psychiatry

Philip M. Coons, M.D.
Professor of Psychiatry, Indiana University School of Medicine; Attending Psychiatrist, Indiana University Hospitals, Indianapolis, Indiana.
18.2. Dissociative Fugue

Jeremy D. Coplan, M.D.
Associate Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Associate Director of Biological Studies Unit, New York State Psychiatric Institute, New York, New York.
15.3. Anxiety Disorders: Biochemical Aspects

G. Rees Cosgrove, M.D., F.R.C.S.(C)
Associate Professor of Surgery, Harvard Medical School, Boston, Massachusetts.
31.32. Neurosurgical Treatments

Jorge Alberto Costa e Silva, M.D.
Professor of Psychiatry and Director of International Center for Mental Health Policy and Research, New York University School of Medicine; Director of Division of Mental Health and Prevention of Substance Abuse, World Health Organization, Geneva, Switzerland.
55.2. World Aspects of Psychiatry

Paul T. Costa, Jr., Ph.D.
Clinical Professor of Psychiatry, Georgetown University School of Medicine and Health Sciences, Washington, D.C.; Associate Professor of Medical Psychology, Johns Hopkins University School of Medicine; Chief of Laboratory of Personality and Cognition, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland.
6.4. Approaches Derived From Philosophy and Psychology

Nancy S. Cotton, Ph.D.
Instructor of Psychology, Department of Psychiatry, Harvard Medical School, Boston, Massachusetts.
32.3. Normal Adolescence

Thomas J. Crowley, M.D.
Professor of Psychiatry, University of Colorado School of Medicine; Director of Division of Substance Dependence, University of Colorado Health Sciences Center, Denver, Colorado.
11.8. Inhalant-Related Disorders

Jan L. Culbertson, Ph.D.
Associate Professor of Pediatrics and Clinical Associate Professor of Psychiatry and Behavioral Sciences, University of Oklahoma College of Medicine; Director of Neuropsychology Services, Child Study Center, Children's Hospital of Oklahoma, Oklahoma City, Oklahoma.
7.5. Personality Assessment: Adults and Children

Jeffrey L. Cummings, M.D.
Augustus S. Rose Professor of Neurology and Professor of Psychiatry and Biobehavioral Science, University of California at Los Angeles School of Medicine, Los Angeles, California.
2.1. Neuropsychiatry: Clinical Assessment and Approach to Diagnosis

John F. Curry, Ph.D.
Associate Professor of Psychiatry and Behavioral Sciences, Duke University School of Medicine; Associate Professor of Psychology, Duke University, Durham, North Carolina.

3.2. Extending Piaget's Approach to Intellectual Functioning

David Gordon Daniel, M.D.
Clinical Professor of Psychiatry and Behavioral Sciences, George Washington University School of Medicine and Health Sciences, Washington, D.C.
2.6. Neuropsychiatric Aspects of Movement Disorders

Garrett C. Daum, M.D.
Assistant Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Medical Director of Outpatient Geriatric Psychiatry Programs, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
51.4j. Geriatric Psychiatry: Cognitive-Behavioral Therapy

John M. Davis, M.D.
Gilman Professor of Psychiatry, University of Illinois at Chicago College of Medicine, Chicago, Illinois.
31.2. Pharmacokinetics and Drug Interactions

Kenneth L. Davis, M.D.
Esther and Joseph Klingenstein Professor of Psychiatry and Chairman, Department of Psychiatry, Mount Sinai School of Medicine, New York, New York.
31.15. Cholinesterase Inhibitors

Lori L. Davis, M.D.
Assistant Professor of Psychiatry, University of Alabama School of Medicine, Birmingham, Alabama.
28.3. Premenstrual Dysphoric Disorder

Mark DeAntonio, M.D.
Assistant Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Inpatient Adolescent Service, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
48.8. Child Psychiatry: Residential and Inpatient Treatment

Charles DeBattista, M.D., D.M.H.
Clinical Fellow in Affective Disorders, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California.
31.33. Other Pharmacological and Biological Therapies

Rebecca F. Detweiler, Ph.D.
Assistant Professor of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada.
37.5. Communication Disorder Not Otherwise Specified

Lynn H. Deutsch, D.O.
Clinical Assistant Professor of Psychiatry, Georgetown University School of Medicine and Health Sciences; Medical Officer, District of Columbia Commission on Mental Health Services, Saint Elizabeth's Campus, Washington, D.C.
7.7. Medical Assessment and Laboratory Testing in Psychiatry

Stephen I. Deutsch, M.D., Ph.D.
Professor and Associate Chairman of Clinical Neurosciences, Department of Psychiatry, Georgetown University School of Medicine and Health Sciences; Chief of Psychiatry Service, Veterans Affairs Medical Center, Washington, D.C.
7.7. Medical Assessment and Laboratory Testing in Psychiatry

Leah J. Dickstein, M.D.
Professor of Psychiatry and Associate Chair for Academic Affairs, Director of Division of Attitudinal and Behavioral Medicine, Department of Psychiatry and Behavioral Sciences, University of Louisville School of Medicine; Staff Psychiatrist, University of Louisville Hospital, Louisville, Kentucky.
27.5. Other Additional Conditions That May Be a Focus of Clinical Attention

Joel E. Dimsdale, M.D.
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California.
25.9. Stress and Psychiatry

Nancy S. Downs, M.D.
Assistant Clinical Professor of Psychiatry, Associate Director of Residency Training, University of California at San Diego School of Medicine, La Jolla, California.
28.5. Death, Dying, and Bereavement

Martin J. Drell, M.D.
Professor of Clinical Psychiatry and Head of Infant, Child, and Adolescent Psychiatry, Louisiana State University Medical School; Clinical Director of New Orleans Adolescent Hospital, New Orleans, Louisiana.
49.5. Children's Reaction to Illness and Hospitalization

Jennifer J. Dunkin, Ph.D.
Assistant Professor of Psychiatry and Director of Geriatric Psychology and Biobehavioral Sciences Fellowship Program, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.2c. Psychological Changes with Normal Aging

Elisabeth M. Dykens, Ph.D.
Associate Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Attending Psychiatrist, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
34. Mental Retardation

James C. Edmondson, M.D., Ph.D.
Clinical Assistant Professor of Neurology, State University of New York at Brooklyn College of Medicine; Assistant Attending, Department of Neurology, Brooklyn Hospital Center; Assistant Attending, Department of Neurology, Long Island College Hospital, Brooklyn, New York.
2.11. Neuropsychiatric Aspects of Neuromuscular Disease;
2.12. Neuropsychiatric Aspects of Child Neurology
28.6. Chronic Pain and the Placebo Effect

Michael F. Egan, M.D.
Director of Clinical Research, Neuroscience Research Center at Saint Elizabeth's Hospital, and Acting Branch Chief of Clinical Research Services Branch, National Institute of Mental Health, National Institutes of Health, Washington, D.C.
2.6. Neuropsychiatric Aspects of Movement Disorders;
12.4. Schizophrenia: Neurobiology

John Richard Elpers, M.D.
Professor of Clinical Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California; Vice Chairman, Department of Psychiatry for Planning and Development, and Director of Ambulatory Psychiatric Services, Harbor-UCLA Medical Center, Torrance, California.

52.1. Public Psychiatry

Norbert B. Enzer, M.D.
Professor of Psychiatry and Interim Associate Dean for Community Programs and Graduate Medical Education, Michigan State University College of Human Medicine, East Lansing, Michigan.
49.15. Psychiatric Prevention in Children and Adolescents

Brenda R. Erickson, M.D.
Former Assistant Professor of Psychiatry, University of Nevada School of Medicine, Reno, Nevada.
13.1. Schizoaffective Disorder, Schizophreniform Disorder, and Brief Psychotic Disorder

Spencer Eth, M.D.
Professor of Psychiatry, New York Medical College; Vice Chairman and Clinical Director, Department of Psychiatry, St. Vincents Hospital, New York, New York.
51.6b. Geriatric Psychiatry: Forensic Issues

Dwight L. Evans, M.D.
Professor of Psychiatry, Chair of Department of Psychiatry, Professor of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
25.10. Behavior and Immunity

Brian Anthony Fallon, M.D., M.P.H.
Associate Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Attending Psychiatrist, New York State Psychiatric Institute, New York, New York.
2.9. Neuropsychiatric Aspects of Other Infectious Diseases

Beverly J. Fauman, M.D.
Associate Professor of Psychiatry, University of Maryland School of Medicine; Senior Psychiatrist, Walter P. Carter Center, University of Maryland Medical System, Baltimore, Maryland.
29.2. Other Psychiatric Emergencies

Jan Fawcett, M.D.
Stanley G. Harris, Sr., Professor of Psychiatry and Chairman, Department of Psychiatry, Rush Medical College; Director, Rush Institute for Mental Well-Being, Chicago, Illinois.
31.27. Sympathomimetics

Marc D. Feldman, M.D.
Associate Professor of Psychiatry, Vice Chairman for Clinical Services, Medical Director of Center for Psychiatric Medicine, and Director of Division of Adult Psychiatry, Department of Psychiatry and Behavioral Neurobiology, University of Alabama School of Medicine, Birmingham, Alabama.
17. Factitious Disorders

S. Shirley Feldman, Ph.D.
Senior Research Scientist, Division of Child Psychiatry and Child Development, Stanford University School of Medicine; Associate Director of Human Biology Program at Stanford University, Stanford, California.
49.7. Childhood or Adolescent Antisocial Behavior

Eileen B Fennell, Ph.D.
Professor of Clinical and Health Psychology, University of Florida; Professor of Neurology, University of Florida College of Medicine, Gainesville, Florida.
7.6. Neuropsychological and Intellectual Assessment of Children

Wayne S. Fenton, M.D.
Associate Clinical Professor of Psychiatry and Behavioral Sciences, George Washington University School of Medicine and Health Sciences, Washington, D.C.; Director of Research, Chestnut Lodge Hospital, Rockville, Maryland.
12.10. Schizophrenia: Individual Psychotherapy

Calvin J. Flowers, M.D.
Research Fellow, Department of Psychiatry and the Behavioral Sciences, University of Southern California School of Medicine, Los Angeles County and USC Medical Center, Los Angeles, California.
31.5. b-Adrenergic Receptor Antagonists

Charles V. Ford, M.D.
Professor of Psychiatry, Director of Neuropsychiatry Clinic, Department of Psychiatry and Behavioral Neurobiology, University of Alabama School of Medicine, Birmingham, Alabama.
17. Factitious Disorders

Victor M. Fornari, M.D.
Associate Professor of Clinical Psychiatry, New York University School of Medicine, New York, New York; Associate Chairman for Education and Training, Department of Psychiatry, North Shore University Hospital, Manhasset, New York.
49.10. Child Psychiatry: Identity Problem and Borderline Disorders

Sydney Frederick-Osborne, Ph.D.
Postdoctoral Fellow, Department of Psychiatry, University of California at San Francisco School of Medicine, San Francisco, California.
1.11. Psychoneuroendocrinology

Nelson B. Freimer, M.D.
Associate Professor of Psychiatry, University of California at San Francisco School of Medicine, San Francisco, California.
1.18. Genetic Linkage Analysis of the Psychiatric Disorders

M. Gerard Fromm, Ph.D.
Director of Therapeutic Community Program, Austen Riggs Center, Stockbridge, Massachusetts; Faculty, Massachusetts Institute for Psychoanalysis, Cambridge, Massachusetts.
30.9. Eriksonian Clinical Theory and Psychiatric Treatment

Abby J. Fyer, M.D.
Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Attending Psychiatrist, New York State Psychiatric Institute, New York, New York.
15.4. Anxiety Disorders: Genetics

Glen O. Gabbard, M.D.
Bessie Walker Callaway Distinguished Professor of Psychoanalysis and Education, Karl Menninger School of Psychiatry, The Menninger Clinic; Clinical Professor of Psychiatry, University of Kansas School of Medicine, Wichita, Kansas; Director of Topeka Institute for Psychoanalysis, Topeka, Kansas.
6.1. Psychoanalysis
14.5. Mood Disorders: Psychodynamic Aspects;

15.5. Anxiety Disorders: Psychodynamic Aspects;
30.1. Psychoanalysis and Psychoanalytic Psychotherapy;
30.12. Combined Psychotherapy and Pharmacotherapy

Steven J. Garlow, M.D., Ph.D.
Assistant Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia.
31.22. Nefazodone
31.29. Trazodone

Thomas R. Garrick, M.D.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Staff Psychiatrist, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
25.6. Endocrine and Metabolic Disorders

Nori Geary, Ph.D.
Associate Research Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University, New York, New York; Research Professor, E. W. Bourne Behavioral Research Laboratory, New York Hospital-Cornell Medical Center, Westchester Division, White Plains, New York.
1.20. Appetite

J. Christian Gillin, M.D.
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Director of Mental Health Clinical Research Center, San Diego Veterans Affairs Healthcare System, San Diego, California.
1.19. Basic Science of Sleep

Michael J. Gitlin, M.D.
Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Mood Disorders Program and Medical Director of Schizophrenia Aftercare Clinic, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
8. Clinical Manifestations of Psychiatric Disorders

Robert N. Golden, M.D.
Professor of Psychiatry and Chair, Department of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, North Carolina.
31.11. Bupropion

Marion Zucker Goldstein, M.D.
Associate Professor of Psychiatry, State University of New York at Buffalo School of Medicine and Biomedical Sciences; Director of Division of Geriatric Psychiatry, University Psychiatry Practice, Erie County Medical Center, Buffalo, New York.
51.6f. Geriatric Psychiatry: Gender Issues;
51.6g. Geriatric Psychiatry: Elder Abuse, Neglect, and Exploitation

Daniel Goleman, Ph.D.
Former Science Editor, *New York Times*, New York, New York.
3.6. Emotional Intelligence

Maureen Fulchiero Gordon, M.D.
Assistant Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
32.2. Normal Child Development

Jack M. Gorman, M.D.
Professor of Psychiatry, Columbia University College of Physicians and Surgeons, New York, New York.
15.1. Anxiety Disorders: Introduction and Overview

Irving I. Gottesman, Ph.D., F.R.C.Psych.(Hon)
Sherell J. Aston Professor, Departments of Psychology and Pediatrics (Medical Genetics), University of Virginia School of Medicine, Charlottesville, Virginia.
1.17. Population Genetic Methods in Psychiatry

Gary L. Gottlieb, M.D., M.B.A.
Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts.
51.5a. Geriatric Psychiatry: Financial Issues

Igor Grant, M.D., F.R.C.P.(C)
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Chief of Ambulatory Care for Psychiatry and Staff Psychiatrist, San Diego Veterans Affairs Healthcare System, San Diego, California.
2.8. Neuropsychiatric Aspects of HIV Infection and AIDS;
5.2. Statistics and Experimental Design

Jack A. Grebb, M.D.
Clinical Professor of Psychiatry, New York University School of Medicine, New York, New York.
1.1. Neural Sciences: Introduction and Overview;
31.1. General Principles of Psychopharmacology

Richard Green, M.D., J.D.
Visiting Professor of Psychiatry, Imperial College School of Medicine; Head of Gender Identity Clinic, Charing Cross Hospital, London, England; Professor of Psychiatry Emeritus, University of California at Los Angeles School of Medicine, Los Angeles, California; Senior Research Fellow, Institute of Criminology, and Affiliated Lecturer, Faculty of Law, University of Cambridge, Cambridge, England.
19.3. Gender Identity Disorders

Marcia Greenleaf, Ph.D.
Assistant Professor of Psychology, Department of Psychiatry, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York.
30.3. Hypnosis

Stanley I. Greenspan, M.D.
Clinical Professor of Psychiatry and Behavioral Sciences and Pediatrics, George Washington University School of Medicine and Health Sciences; Supervising Child Psychoanalyst, Washington Psychoanalytic Institute, Washington, D.C.
3.2. Extending Piaget's Approach to Intellectual Functioning

John H. Greist, M.D.
Clinical Professor of Psychiatry, University of Wisconsin Medical School; Distinguished Senior Scientist. Co-Director of Lithium Information Center, Madison Institute of Medicine, Madison, Wisconsin.
31.18. Lithium

Roland R. Griffiths, Ph.D.

Professor of Psychiatry and Behavioral Sciences, Department of Neurosciences, Johns Hopkins University School of Medicine, Baltimore, Maryland.
11.4. Caffeine-Related Disorders

Lawrence S. Gross, M.D.
Associate Professor of Clinical Psychiatry and Behavioral Sciences, University of Southern California School of Medicine; Director of Outpatient Psychiatry, Cedars-Sinai Medical Center, Los Angeles, California.
31.8. Antihistamines

George T. Grossberg, M.D.
Samuel W. Fordyce Professor of Psychiatry and Chairman, Department of Psychiatry, Saint Louis University School of Medicine, Saint Louis, Missouri.
51.4g. Geriatric Psychiatry: Electroconvulsive Therapy

Peter B. Gruenberg, M.D.
Associate Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
54.2. Ethics in Psychiatry

Frederick G. Guggenheim, M.D.
Marie Wilson Howells Professor of Psychiatry and Chair, Department of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences, Little Rock, Arkansas.
16. Somatoform Disorders

Raquel E. Gur, M.D., Ph.D.
Professor of Psychiatry and Director of Neurospsychiatry, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
12.3. Schizophrenia: Brain Structure and Function

Ruben C. Gur, Ph.D.
Professor of Psychiatry and Director of Neuropsychology, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
12.3. Schizophrenia: Brain Structure and Function

Alan S. Gurman, Ph.D.
Professor of Psychiatry, Director of Couple-Family Clinic, Chief Psychologist, University of Wisconsin Medical School, Madison, Wisconsin.
30.5. Family Therapy and Couple Therapy

Kathleen Y. Haaland, Ph.D.
Professor of Psychiatry, University of New Mexico School of Medicine; Staff Psychologist, Albuquerque Veterans Affairs Medical Center, Albuquerque, New Mexico.
7.4. Clinical Neuropsychology and Intellectual Assessment of Adults

Katherine A. Halmi, M.D.
Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University, New York, New York; Director of Eating Disorders Program, New York Hospital-Cornell Medical Center, Westchester Division, White Plains, New York.
20. Eating Disorders

M. Jackuelyn Harris, M.D.
Associate Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Co-Director of Geriatric Psychiatry Program, San Diego Veterans Affairs Healthcare System, San Diego, California.
51.3f. Geriatric Psychiatry: Schizophrenia and Delusional Disorders

Jennifer F. Havens, M.D.
Assistant Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Director of Special Needs Clinic, Pediatric Psychiatry, Columbia-Presbyterian Medical Center, New York, New York.
49.6. Child Psychiatry: Psychiatric Sequelae of HIV and AIDS

Donald P. Hay, M.D.
Associate Professor of Psychiatry, Director of Geriatric Pscychiatry Programs, University of Colorado Health Sciences Center, Denver, Colorado.
51.4g. Geriatric Psychiatry: Electroconvulsive Therapy

Stephaine L. Heard, M.D.
Clinical Assistant Professor of Psychiatry, Michigan State University College of Human Medicine, Kalamazoo Center for Medical Studies, Kalamazoo, Michigan.
49.15. Psychiatric Prevention in Children and Adolescents

A. Scott Henderson, M.D., D.Sc., F.R.A.C.P., F.R.A.N.Z.C.P., F.R.C.P., F.R.C.Psych.
Professor, Australian National University; Director of National Health and Medical Research Council, Psychiatric Epidemiology Research Centre, Canberra, Capital Territory, Australia.
51.1b. Geriatric Psychiatry: Epidemiology of Psychiatric Disorders

Victoria C. Hendrick, M.D.
Assistant Professor of Psychiatry and Biobehavioral Sciences, Director of Pregnancy and Postpartum Mood Disorders Program, University of California at Los Angeles School of Medicine, Los Angeles, California.
25.6. Endocrine and Metabolic Disorders

Hugh C. Hendrie, M.B., Ch.B.
Albert E. Sterne Professor of Psychiatry and Chairman, Department of Psychiatry, Indiana University School of Medicine, Indianapolis, Indiana.
51.6d. Geriatric Psychiatry: Sociocultural Issues

Stephen P. Herman, M.D.
Associate Clinical Professor of Psychiatry, Mount Sinai School of Medicine, New York, New York.
49.12. Forensic Child and Adolescent Psychiatry

Jerry D. Heston, M.D.
Associate Professor of Psychiatry, University of Tennessee at Memphis College of Medicine; Medical Director of Child and Adolescent Day Treatment Services, University of Tennessee Medical Center, Memphis, Tennessee.
48.7. Child Psychiatry: Partial Hospital and Ambulatory Behavioral Health Services

Euthymia D. Hibbs, Ph.D.
Chief of Child and Adolescent Psychosocial Treatment Research, Child and Adolescent Treatment and Preventive Intervention Research Branch, Division of Services and Intervention Research, National Institute of Mental Health, National Institutes of Health, Rockville, Maryland; Adjunct Associate Professor of Psychiatry; George Washington University School of Medicine and Health Sciences, Washington, D.C.
48.2. Child Psychiatry, Short-Term Psychotherapy

Robert M. A. Hirschfeld, M.D.
Titus H. Harris Distinguished Professor of Psychiatry and Chair, Department of Psychiatry and Behavioral Sciences, University of Texas Medical School at Galveston,

Galveston, Texas.
14.9. Mood Disorders: Psychotherapy

Max Hirshkowitz, Ph.D.
Associate Professor of Psychiatry, Associate Director of Sleep Disorders Center, Baylor College of Medicine; Director of Sleep Research Center, Houston Veterans Affairs Medical Center, Houston, Texas.
21. Sleep Disorders

Robert M. Hodapp, Ph.D.
Associate Professor of Education and Psychological Studies in Education, University of California at Los Angeles, Los Angeles, California.
34. Mental Retardation

Ralph E. Hoffman, M.D.
Associate Professor of Psychiatry, Yale University School of Medicine; Medical Director, Yale Psychiatric Institute, New Haven, Connecticut.
12.6. Schizophrenia: Psychodynamic to Neurodynamic Theories

Jimmie C. Holland, M.D.
Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University; Chief of Psychiatry Services and Wayne E. Chapman Chair in Psychiatric Oncology, Memorial Sloan-Kettering Cancer Center, New York, New York.
25.11. Psycho-Oncology

Eric Hollander, M.D.
Professor of Psychiatry, Director of Psychopharmacology
Mount Sinai School of Medicine, New York, New York.
18.5. Dissociative Disorders Not Otherwise Specified

Nancy L. Hornstein, M.D.
Attending Staff, Condell Medical Center, Libertyville, Illinois; Attending Staff, Victory Memorial Hospital, Waukegan, Illinois.
49.8. Dissociative Disorders in Children and Adolescents

Thomas B. Horvath, M.D., F.R.A.C.P.
Professor of Psychiatry, Baylor College of Medicine; Chief of Staff, Houston Veterans Affairs Medical Center, Houston, Texas.
51.5c. Geriatric Psychiatry: Veterans Affairs Medical Centers and Psychogeriatric Services

Ewald Horwath, M.D., M.Sc.
Associate Clinical Professor of Psychiatry, Columbia University College of Physicians and Surgeons; Director of Intensive Care Unit, Washington Heights Community Service, New York State Psychiatric Institute, New York, New York.
15.2. Anxiety Disorders: Epidemiology

Kenneth I. Howard, Ph.D.
Professor of Psychology, Northwestern University, Evanston, Illinois; Professor of Psychiatry, Northwestern University Medical School, Chicago, Illinois.
30.11. Evaluation of Psychotherapy

Charles Campbell Hughes, Ph.D.*
Professor of Anthropology, University of Utah; Professor of Family and Preventative Medicine, University of Utah School of Medicine, Salt Lake City, Utah.
13.3. Acute and Transient Psychotic Disorders and Culture-Bound Syndromes

John R. Hughes, M.D.
Professor of Psychiatry, Psychology, and Family Practice, University of Vermont School of Medicine, Burlington, Vermont.
11.9. Nicotine-Related Disorders

Thomas M. Hyde, M.D., Ph.D.
Special Expert in Neurology, Neurology Consultation Service, Clinical Brain Disorders Branch, National Institute of Mental Health Neuroscience Center at Saint Elizabeth's Hospital, National Institutes of Health, Washington, D.C.
12.4. Schizophrenia: Neurobiology

Steven E. Hyman, M.D.
Director of National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.
1.10. Basic Molecular Neurobiology

Robert B. Innis, M.D., Ph.D.
Professor of Psychiatry and Pharmacology, Director of Neurochemical Brain Imaging Program, Yale University School of Medicine, New Haven, Connecticut.
1.15. Principles of Neuroimaging: Radiotracer Techniques

Keith E. Isenberg, M.D.
Associate Professor of Psychiatry, Washington University School of Medicine; Research Associate Professor, Saint Louis College of Pharmacy; Director of Electroconvulsive Therapy (ECT) Service, Barnes-Jewish Hospital, Saint Louis, Missouri.
1.9. Basic Electrophysiology
31.31. Electroconvulsive Therapy

Rolf G. Jacob, M.D.
Associate Professor of Psychiatry and Otolaryngology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania.
30.2. Behavior Therapy

Ari B. Jaffe, M.D.
Clinical Instructor of Psychiatry, New York University School of Medicine, New York, New York.
11.10. Opioid-Related Disorders

Jerome H. Jaffe, M.D.
Clinical Professor of Psychiatry
University of Maryland School of Medicine, Baltimore, Maryland.
11.1. Substance-Related Disorders: Introduction and Overview;
11.3. Amphetamine (or Amphetamine-like)-Related Disorders;
11.6. Cocaine-Related Disorders;
11.10. Opioid-Related Disorders

Philip G. Janicak, M.D.
Professor of Psychiatry
Medical Director of the Psychiatric Clinical Research Center, University of Illinois at Chicago College of Medicine, Chicago, Illinois.
31.2. Pharmacokinetics and Drug Interactions

Lissy F. Jarvik, M.D., Ph.D.
Professor Emeritus of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Distinguished Physician (Emeritus), UCLA Neuropsychiatric Institute and Hospital, Staff Psychiatrist, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
51.1a. Geriatric Psychiatry: Introduction;
51.2a. Psychiatric Examination of the Older Patient;
51.4a. Geriatric Psychiatry: Treatment of Psychiatric Disorders: Introduction and Overview

James W. Jefferson, M.D.
Clinical Professor of Psychiatry, University of Wisconsin Medical School; Distinguished Senior Scientist, Co-Director of Lithium Information Center, Madison Institute of Medicine, Madison, Wisconsin.
31.18. Lithium

Dilip V. Jeste, M.D.
Professor of Psychiatry and Neuroscience, University of California at San Diego School of Medicine, La Jolla, California; Director of Geriatric Psychiatry Clinical Research Center, San Diego Veterans Affairs Healthcare System, San Diego, California.
51.3f. Geriatric Psychiatry: Schizophrenia and Delusional Disorders

Russell T. Joffe, M.D.
Professor of Psychiatry and Chair, Department of Psychiatry, McMaster University Faculty of Health Sciences; Psychiatrist-in-Chief, Hamilton Psychiatric Hospital, Hamilton, Ontario, Canada.
31.28. Thyroid Hormones

Carla J. Johnson, Ph.D.
Associate Professor of Speech-Language Pathology, University of Toronto, Toronto, Ontario, Canada.
37.1. Expressive Language Disorder;
37.2. Mixed Receptive-Expressive Language Disorder;
37.3. Phonological Disorder

Ricardo E. Jorge, M.D.
Associate Research Scientist, Department of Psychiatry, University of Iowa College of Medicine; Attending Psychiatrist, University of Iowa Hospitals and Clinics, Iowa City, Iowa.
2.5. Neuropsychiatric Aspects of Traumatic Brain Injury

Allan M. Josephson, M.D.
Professor of Psychiatry and Chief of Psychiatry and Health Behavior, Section of Child, Adolescent, and Family Psychiatry, Medical College of Georgia; Director of Clinical Services for Psychiatry, Medical College of Georgia Hospital and Clinics, Augusta, Georgia.
48.5. Child Psychiatry: Family Therapy

Robert M. Kaplan, Ph.D.
Professor of Family and Preventive Medicine and Chief of Division of Health Care Sciences, University of California at San Diego School of Medicine, La Jolla, California.
5.2. Statistics and Experimental Design

Craig N. Karson, M.D.
Professor of Psychiatry and Pathology, University of Arkansas for Medical Sciences; Chief of Staff, John L. McClellan Memorial Hospital, Little Rock, Arkansas.
1.16. Principles of Neuroimaging: Magnetic Resonance Techniques

Julia E. Kasl-Godley, Ph.D.
Research Associate, Department of Psychiatry, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.2c. Psychological Changes With Normal Aging

Ira R. Katz, M.D., Ph.D.
Professor of Psychiatry, University of Pennsylvania School of Medicine; Director of Section of Geriatric Psychiatry, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania.
51.6a. Geriatric Psychiatry: Psychiatric Aspects of Long-Term Care

Jeffrey William Katzman, M.D.
Assistant Clinical Professor of Psychiatry, University of New Mexico School of Medicine; Acting Chief of Albuquerque Veterans Affairs Medical Center, Albuquerque, New Mexico.
22. Impulse-Control Disorders Not Elsewhere Classified

Paul E. Keck, Jr., M.D.
Professor of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio.
31.7b. Valproate

Francis J. Keefe, Ph.D.
Adjunct Professor of Psychiatry and Biobehavioral Sciences, Duke University School of Medicine, Durham, North Carolina.
25.9. Stress and Psychiatry

Samuel J. Keith, M.D.
Professor of Psychiatry and Chairman, Department of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.
12.9. Schizophrenia: Psychosocial Treatment;
13.1. Schizoaffective Disorder, Schizophreniform Disorder, and Brief Psychotic Disorder

Jeffrey E. Kelsey, M.D., Ph.D.
Assistant Professor of Psychiatry and Behavioral Sciences, Director of Mood and Anxiety Disorders Clinical Trials Program, Emory University School of Medicine, Atlanta, Georgia.
31.25a. Selective Serotonin Reuptake Inhibitors: Introduction and Overview;
31.25b. Citalopram;
31.25c. Fluoxetine;
31.25d. Fluvoxamine;
31.25e. Paroxetine;
31.25f. Sertraline

John R. Kelsoe, M.D.
Associate Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California.
14.3. Mood Disorders: Genetics

Kenneth S. Kendler, M.D.
Professor of Psychiatry and Human Genetics, Virginia Institute for Psychiatric and Behavioral Genetics, Medical College of Virginia at Virginia Commonwealth

University, Richmond, Virginia.
12.5. Schizophrenia: Genetics

Sidney H. Kennedy, M.D.
Professor of Psychiatry, University of Toronto Faculty of Medicine; Head of Mood and Anxiety Disorders Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
31.20. Monoamine Oxidase Inhibitors

Ronald C. Kessler, Ph.D.
Professor of Sociology, Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts.
4.2. Sociology and Psychiatry

Bryan H. King, M.D.
Professor of Psychiatry and Pediatrics, Dartmouth Medical School, Hanover, New Hampshire; Director of Child and Adolescent Psychiatry, Children's Hospital at Dartmouth, Dartmouth Hitchock Medical Center, Lebanon, New Hampshire.
34. Mental Retardation

Deborah A. King, Ph.D.
Associate Professor of Psychiatry (Psychology), University of Rochester School of Medicine and Dentistry; Staff Psychologist and Director of Geriatric Psychiatry Services, Strong Memorial Hospital, Rochester, New York.
51.4k. Interventions and Consultation With Families of Older Adults

Robert A. King, M.D.
Professor of Child Psychiatry, Medical Director of Tic Disorder/Obsessive-Compulsive Disorder Specialty Clinic, Yale Child Study Center, Yale University School of Medicine; Associate Director of Child Psychiatry Consultation-Liaison Service in Pediatrics, Yale-New Haven Hospital, New Haven, Connecticut.
33. Psychiatric Examination of the Infant, Child, and Adolescent

Darrell G. Kirch, M.D.
Professor of Psychiatry and Health Behavior and Dean, Schools of Medicine and Graduate Studies, Medical College of Georgia, Augusta, Georgia.
28.4. Genetic Counseling

Thomas J. Kiresuk, Ph.D.
Professor of Health Psychology, Department of Psychiatry, University of Minnesota Medical School; Chief Clinical Psychologist, Hennepin County Medical Center; Director of Center for Addiction and Alternative Medicine Research, Minneapolis Medical Research Foundation, Minneapolis, Minnesota.
28.8. Alternative and Complimentary Health Practices

Laurel J. Kiser, Ph.D., M.B.A.
Professor of Psychiatry, University of Tennessee at Memphis College of Medicine; Executive Director of Child and Adolescent Day Treatment Services, University of Tennessee Medical Center, Memphis, Tennessee.
48.7. Child Psychiatry: Partial Hospital and Ambulatory Behavioral Health Services

Herbert D. Kleber, M.D.
Professor of Psychiatry, Director of Division on Substance Abuse, Department of Psychiatry, Columbia University College of Physicians and Surgeons; Executive Vice President and Medical Director, National Center on Addiction and Substance Abuse at Columbia University, New York, New York.
31.23. Opioid Agonists

Arthur Kleinman, M.D.
Presley Professor of Medical Anthropology and Psychiatry, Chairman, Department of Social Medicine, Harvard Medical School, Boston, Massachusetts; Professor of Social Anthropology, Harvard University, Cambridge, Massachusetts.
4.1. Anthropology and Psychiatry

Ami Klin, Ph.D.
Harris Associate Professor of Child Psychology and Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
38. Pervasive Developmental Disorders

Alex Kopelowicz, M.D.
Assistant Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California; Director of San Fernando Mental Health Center, Mission Hills, California.
52.4. Psychiatric Rehabilitation

Kathryn J. Kotrla, M.D.
Assistant Professor of Psychiatry and Behavioral Sciences, Baylor College of Medicine; Chief of Psychiatry, Ben Taub General Hospital, Houston, Texas.
1.3. Developmental Neurobiology

Peter D. Kramer, M.D.
Clinical Professor of Psychiatry, Brown University School of Medicine, Providence, Rhode Island.
55.3. The Future of Psychiatry

Ronald F. Krasner, M.D.
Assistant Professor of Psychiatry, Director of Psychiatric Education, and Vice Chairman, Department of Psychiatry and Behavioral Sciences, Northwestern University Medical School, Chicago, Illinois.
30.11. Evaluation of Psychotherapy

Suchitra Krishnan-Sarin, Ph.D.
Assistant Professor of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
31.21. Naltrexone

Robert Kroll, Ph.D.
Assistant Professor of Speech-Language Pathology, University of Toronto; Assistant Professor of Psychiatry, University of Toronto Faculty of Medicine; Director of Speech Foundation of Ontario Stuttering Centre; Toronto, Ontario, Canada.
37.4. Stuttering

Anand Kumar, M.D.
Associate Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Attending Psychiatrist, Geriatric Service, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
51.2f. Geriatric Psychiatry: Neuroimaging: Special Issues

Lawrence A. Labbate, M.D.
Associate Professor of Psychiatry and Behavioral Sciences, Medical University of South Carolina; Director of Outpatient Mental Health Service, Ralph H. Johnson Veterans Affairs Medical Center, Charleston, South Carolina; Associate Professor of Clinical Psychiatry, Uniformed Services University of the Health Sciences F. Edward Hébert School of Medicine, Bethesda, Maryland.

31.9. Barbiturates and Similarly Acting Substances
31.14. Chloral Hydrate

Asenath LaRue, Ph.D.
Professor of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.
7.4. Clinical Neurology and Intellectual Assessment of Adults

John Lauriello, M.D.
Assistant Professor of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.
12.9. Schizophrenia: Psychosocial Treatment;
13.1. Schizoaffective Disorder, Schizophreniform Disorder, and Brief Psychotic Disorder

Eleanor P. Lavretsky, M.D., Ph.D.
Research Psychopharmacologist, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
51.2a. Psychiatric Examination of the Older Patient

Helen Lavretsky, M.D.
Assistant Professor of Psychiatry, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.4e. Geriatric Psychiatry: Psychopharmacology: Antipsychotic Drugs

Lawrence W. Lazarus, M.D.
Assistant Professor of Psychiatry, Director of Geropsychiatry Fellowship Program, Rush Medical College, Chicago, Illinois.
51.4i. Geriatric Psychiatry: Individual Psychotherapy

Paul Leber, M.D.
Associate Clinical Professor of Psychiatry and Behavioral Science, George Washington University School of Medicine and Health Sciences, Washington, D.C.; Former Director, Division of Neuropharmacological Drug Product, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland.
31.3. Drug Development and Approval Process in the United States

Jay L. Lebow, Ph.D.
Senior Therapist and Research Consultant, Family Institute, Northwestern University, Evanston, Illinois.
30.5. Family Therapy and Couple Therapy

Barry D. Lebowitz, Ph.D.
Chief of Adult and Geriatric Treatment and Preventive Intervention, National Institute of Mental Health, National Institutes of Health, Rockville, Maryland; Adjunct Faculty, Department of Psychiatry, Georgetown University School of Medicine and Health Sciences, Washington, D.C.
51.5d. Community Services for the Elderly Psychiatric Patient

Marguerite S. Lederberg, M.D.
Clinical Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University; Attending Psychiatrist, Memorial Sloan-Kettering Cancer Center, New York, New York.
25.11. Psycho-Oncology

Sing Lee, M.D.
Associate Professor of Psychiatry, Chinese University of Hong Kong, Hong Kong, China; Lecturer of Social Medicine, Harvard Medical School, Boston, Massachusetts.
9.2. International Psychiatric Diagnosis

Heinz E. Lehmann, M.D.*
Professor Emeritus of Psychiatry, McGill University Faculty of Medicine, Montreal, Quebec, Canada; Deputy Commissioner for Research, Office of Mental Health, Albany, New York.
12.7. Schizophrenia: Clinical Features

Fredda L. Leiter, M.D.
Clinical Instructor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Attending Psychiatrist, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
51.4a. Geriatric Psychiatry: Treatment of Psychiatric Disorders: Introduction and Overview

Henrietta L. Leonard, M.D.
Professor of Psychiatry and Human Behavior, Brown University School of Medicine; Director of Training, Child and Adolescent Psychiatry Residency Program and Combined Pediatrics-Psychiatry-Child Psychiatry Residency Program, Rhode Island Hospital, Providence, Rhode Island.
46.4. Selective Mutism

Ira M. Lesser, M.D.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California; Director of Residency Training, Vice Chair for Academic Affairs, Department of Psychiatry, Harbor-UCLA Medical Center, Torrance, California.
51.3c. Geriatric Psychiatry: Anxiety Disorders

Molyn Leszcz, M.D., F.R.C.P.(C)
Assistant Professor of Psychiatry, Head of Psychotherapy Program, University of Toronto Faculty of Medicine, Toronto, Ontario, Canada.
51.4l. Geriatric Psychiatry: Group Therapy

Eric S. Levine, Ph.D.
Assistant Professor of Pharmacology, University of Connecticut School of Medicine, Farmington, Connecticut.
1.7. Neurotrophic Factors

Stephen B. Levine, M.D.
Clinical Professor of Psychiatry
Case Western Reserve University School of Medicine, Clinical Staff, University Hospitals of Cleveland, Cleveland, Ohio; Co-Director of The Center for Marital and Sexual Health, Beachwood, Ohio.
19.2. Paraphilias

David A. Lewis, M.D.
Professor of Psychiatry and Neuroscience, University of Pittsburgh School of Medicine; Associate Director for Basic Research, Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania.
1.2. Functional Neuroanatomy

Owen Lewis, M.D.
Associate Clinical Professor of Psychiatry, Columbia University College of Physicians and Surgeons; Director of Residency Education in Child Psychiatry, New York State Psychiatric Institute, New York, New York.

48.1. Child Psychiatry: Individual Psychodynamic Psychotherapy

Robert Paul Liberman, M.D.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Clinical Research Center for Serious Mental Illnesses, Los Angeles, California.
52.4. Psychiatric Rehabilitation

Keh-Ming Lin, M.D., M.P.H.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
13.3. Acute and Transient Psychotic Disorders and Culture-Bound Syndromes

Mark S. Lipian, M.D., Ph.D.
Assistant Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California; Associate Clinical Professor of Psychiatry and Human Behavior, University of California at Irvine College of Medicine, Irvine, California; Medical Director of Conditional Release Program of Orange County, Santa Ana, California.
27.2. Malingering

Mack Lipkin, Jr., M.D.
Professor of Clinical Medicine and Director of Division of Primary Care, New York University School of Medicine, New York, New York.
28.1. Primary Care and Psychiatry

Richard J. Loewenstein, M.D.
Medical Director, Trauma Disorders, Sheppard Pratt Health System, Baltimore, Maryland.
18.3. Dissociative Identity Disorder

Joan L. Luby, M.D.
Assistant Professor of Psychiatry (Child), Director of Preschool and Infant Clinic, Department of Psychiatry, Washington University School of Medicine, Saint Louis, Missouri.
44.2. Stereotypic Movement Disorder of Infancy and Disorders of Infancy and Early Childhood Not Otherwise Specified

R. Bruce Lydiard, Ph.D., M.D.
Professor of Psychiatry and Behavioral Sciences, Director of Psychopharmacology Unit and Clinical Psychopharmacology Research Division, Medical University of South Carolina, Charleston, South Carolina.
31.12. Buspirone

Jeffrey M. Lyness, M.D.
Assistant Professor of Psychiatry, University of Rochester School of Medicine and Dentistry, Rochester, New York.
10. Delirium, Dementia, and Amnestic and Other Cognitive Disorders

Wayne Macfadden, M.D.
Clinical Assistant Professor of Psychiatry, University of Pennsylvania School of Medicine; Chief of Inpatient Dual Diagnosis Unit, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania.
11.5. Cannabis-Related Disorders

Robert T. Malison, M.D.
Assistant Professor of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
1.15. Principles of Neuroimaging: Radiotracer Techniques

Facundo F. Manes, M.D.
Research Fellow, Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa.
2.3. Neuropsychiatric Aspects of Brain Tumors

Myrl R. S. Manley, M.D.
Associate Professor of Clinical Psychiatry, Director of Medical Student Education in Psychiatry
New York University School of Medicine, New York, New York.
7.1. Psychiatric Interview, History, and Mental Status Examination

Theo C. Manschreck, M.D., M.P.H.
Professor of Psychiatry and Human Behavior, Director of Laboratory for Clinical and Experimental Psychopathology, Director of Division of Public Psychiatry, Director of Schizophrenia and Related Psychosis Research, Brown University School of Medicine, Providence, Rhode Island.
13.2. Delusional Disorder and Shared Psychotic Disorder

John S. March, M.D., M.P.H.
Associate Professor of Psychiatry and Director of Programs in Pediatric Anxiety Disorders and Psychopharmacology, Division of Child and Adolescent Psychiatry, Department of Psychiatry, Duke University School of Medicine; Associate Professor of Psychology: Social and Health Sciences, Duke University, Durham, North Carolina.
48.3. Child Psychiatry: Cognitive-Behavioral Psychotherapy

Stephen R. Marder, M.D.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Mental Health Services and Chief of Psychiatry, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
12.8. Schizophrenia: Somatic Treatment;
31.17. Dopamine Receptor Antagonists (Typical Antipsychotics);
31.26. Serotonin-Dopamine Antagonists

Deborah B. Marin, M.D.
Assistant Professor of Psychiatry and Geriatrics, Mount Sinai School of Medicine; Director of Division of Geriatric Psychiatry, Mount Sinai Medical Center, New York, New York.
51.6c. Geriatric Psychiatry: Ethical Issues

John C. Markowitz, M.D.
Associate Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University, New York, New York.
30.7. Interpersonal Psychotherapy

Carol A. Mathews, M.D.
Research Fellow, Department of Psychiatry, University of California at San Francisco School of Medicine, San Francisco, California.
1.18. Genetic Linkage Analysis of the Psychiatric Disorders

Richard E. Mattison, M.D.
Clinical Professor of Psychiatry and Director of School Consultation, Department of Psychiatry and Behavioral Science, State University of New York at Stony Brook School of Medicine, Stony Brook, New York.

49.14. School Consultation

Jeffrey E. Max, M.B.B.Ch.
Associate Professor-In-Residence, University of California at San Diego School of Medicine, La Jolla, California; Director of Neuropsychiatric Research, Children's Hospital and Health Center, San Diego, California.
2.5. Neuropsychiatric Aspects of Traumatic Brain Injury

Jon M. McClellan, M.D.
Assistant Professor of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington; Medical Director, Child Study and Treatment Center, Lakewood, Washington.
47. Early-Onset Schizophrenia

James T. McCracken, M.D.
Professor-in-Residence of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Division of Child and Adolescent Psychiatry, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
39.1. Attention-Deficit Disorders
42. Tic Disorders

Robert R. McCrae, Ph.D.
Research Psychologist, Personality, Stress and Coping Section, Gerontology Research Center, National Institute on Aging, National Institutes of Health, Baltimore, Maryland.
6.4. Approaches Derived From Philosophy and Psychology

John Stephen McDaniel, M.D.
Associate Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia.
25.1. History, Classification, and Current Trends in Psychosomatic Medicine

Susan L. McElroy, M.D.
Professor of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio.
31.7b. Valproate

Thomas H. McGlashan, M.D.
Professor of Psychiatry, Yale University School of Medicine; Executive Director, Yale Psychiatric Institute, New Haven, Connecticut.
12.6. Schizophrenia: Psychodynamic to Neurodynamic Theories;
12.10. Schizophrenia: Individual Psychotherapy

James J. McGough, M.D.
Associate Clinical Professor of Psychiatry and Biobehavioral Sciences, Division of Child and Adolescent Psychiatry, University of California at Los Angeles School of Medicine; Associate Director of Outpatient Service, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
27.4. Borderline Intellectual Functioning and Academic Problem

Michael T. McGuire, M.D.
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Attending Psychiatrist, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
4.3. Evolutionary Biology and Psychiatry

Kevin F. McKenna, M.D., Ph.D.
Clinical Associate Professor of Psychiatry, University of Alberta Faculty of Medicine and Dentistry, Edmonton, Alberta, Canada.
31.20. Monoamine Oxidase Inhibitors

William T. McKinney, Jr., M.D.
Helen and Norman Asher Professor of Psychiatry and Behavioral Sciences, Northwestern University Medical School; Director of The Asher Center for the Study and Treatment of Depressive Disorders; Clinical Staff, Northwestern Memorial Hospital, Chicago, Illinois.
5.4. Animal Research and Its Relevance to Psychiatry

Claude Mellins, Ph.D.
Assistant Professor of Clinical Psychology in Psychiatry, Columbia University College of Physicians and Surgeons; Research Scientist, HIV Center for Clinical and Behavioral Studies, New York State Psychiatric Institute; Co-Director of Speech Needs Clinic, Pediatric Psychiatry, Columbia-Presbyterian Medical Center, New York, New York.
49.6. Child Psychiatry: Psychiatric Sequelae of HIV and AIDS

Mario F. Mendez, M.D., Ph.D.
Associate Professor of Neurology, Psychiatry, and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Neurobehavioral Unit, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
2.4. Neuropsychiatric Aspects of Epilepsy

W. Walter Menninger, M.D.
J. Cotter Hirschberg Professor and Former Dean, Karl Menninger School of Psychiatry and Mental Health Sciences, Chief Executive Officer of The Menninger Foundation and Clinic; Clinical Professor of Psychiatry, University of Kansas School of Medicine, Kansas City, Kansas; Instructor, Topeka Institute for Psychoanalysis, Topeka, Kansas.
52.3. Role of the Psychiatric Hospital in the Treatment of Mental Illness

James R. Merikangas, M.D.
Lecturer of Psychiatry, Yale University School of Medicine; Attending in Neurology, Yale-New Haven Hospital, New Haven, Connecticut.
2.10. Neuropsychiatric Aspects of Headache

Kathleen Ries Merikangas, Ph.D.
Professor of Epidemiology and Psychiatry and Director of Geriatric Epidemiology Research Unit, Yale University School of Medicine, New Haven, Connecticut.
2.10. Neuropsychiatric Aspects of Headache

Jonathan M. Meyer, M.D.
Adjunct Assistant Professor of Psychiatry, Oregon Health Sciences University School of Medicine, Portland, Oregon.
31.6. Anticholinergics and Amantadine

Juan E. Mezzich, M.D., Ph.D.
Professor of Psychiatry, Mount Sinai School of Medicine; Director of Division of Psychiatric Epidemiology and International Center for Mental Health, Mount Sinai Medical Center, New York, New York.
9.2. International Psychiatric Diagnosis;
13.3. Acute and Transient Psychotic Disorders and Culture-Bound Syndromes;

23. Adjustment Disorders

Edwin J. Mikkelson, M.D.
Associate Professor of Psychiatry, Harvard Medical School; Medical Director, Mentor Clinical Care; Consultant, Massachusetts Department of Mental Retardation, Boston, Massachusetts.
43. Elimination Disorders

Andrew H. Miller, M.D.
Associate Professor of Psychiatry and Behavioral Sciences and Emory University School of Medicine, Atlanta, Georgia.
1.12. Immune System and Central Nervous System Interactions

Marvin J. Miller, M.D.
Assistant Professor of Psychiatry, Indiana University School of Medicine; Staff Psychiatrist, Larue Carter Memorial Hospital, Indianapolis, Indiana.
7.9. Computer-Based Testing of the Psychiatric Patient

Mark J. Mills, J.D., M.D.
Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
27.2. Malingering

Klaus Minde, M.D., F.R.C.P.(C)
Professor of Psychiatry and Pediatrics and Chairman, Division of Child Psychiatry, McGill University Faculty of Medicine; Director of Department of Psychiatry, Montreal Children's Hospital, Montreal, Quebec, Canada.
49.1. Psychiatric Aspects of Day Care

Paul C. Mohl, M.D.
Professor of Psychiatry, University of Texas Southwestern Medical School; Director of Psychiatric Residency Training, University of Texas Southwestern Medical Center, Dallas, Texas.
6.3. Other Psychodynamic Schools

Steven O. Moldin, Ph.D.
Chief of Genetics Research Branch, Division of Basic and Clinical Neuroscience Research, National Institute of Mental Health, National Institutes of Health, Rockville, Maryland.
1.17. Population Genetic Methods in Psychiatry

Constance A. Moore, M.D.
Associate Professor of Psychiatry and Director of Sleep Disorders Center, Baylor College of Medicine; Director of Sleep Diagnostic Center, Houston Veterans Affairs Medical Center, Houston, Texas.
21. Sleep Disorders

Michael G. Moran, M.D.
Associate Professor of Psychiatry, University of Colorado School of Medicine; Director of Adult Psychosocial Medicine, National Jewish Medical and Research Center, Denver, Colorado.
25.5. Respiratory Disorders

James Morrison, M.D.
Clinical Professor of Psychiatry, Temple University School of Medicine, Philadelphia, Pennsylvania; Chief of Staff, Coatesville Veterans Affairs Medical Center, Coatesville, Pennsylvania.
53.2. Examining Psychiatrists and Other Trainees

Rodrigo A. Muñoz, M.D.
Clinical Professor of Psychiatry, University of California at San Diego; Attending Psychiatrist, UCSD Medical Center; Attending Psychiatrist, Scripps Memorial Hospital, La Jolla, California; Attending Psychiatrist, Mercy Hospital and Medical Center, San Diego, California.
53.2. Examining Psychiatrists and Other Trainees

William E. Narrow, M.D., M.P.H.
Senior Advisor for Epidemiology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.
12.2. Schizophrenia: Epidemiology

J. Craig Nelson, M.D.
Professor of Psychiatry, Yale University School of Medicine; Director of Inpatient Psychiatry Service and Geriatric Psychiatry Programs, Yale-New Haven Hospital, New Haven, Connecticut.
31.30. Tricyclics and Tetracyclics

Charles B. Nemeroff, M.D., Ph.D.
Reunette W. Harris Professor of Psychiatry and Chairman, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia.
1.6. Neuropeptides: Biology and Regulation;
31.22. Nefazodone;
31.25a. Selective Serotonin Reuptake Inhibitors: Introduction and Overview;
31.25b. Citalopram;
31.25c. Fluoxetine;
31.25d. Fluvoxamine;
31.25e. Paroxetine;
31.25f. Sertraline;
31.29. Trazodone

John C. Nemiah, M.D.
Professor of Psychiatry, Dartmouth Medical School, Hanover, New Hampshire; Clinical Staff, Mary Hitchcock Memorial Hospital, Lebanon, New Hampshire; Professor of Psychiatry Emeritus, Harvard Medical School, Boston, Massachusetts.
Introduction

Eric J. Nestler, M.D., Ph.D.
Jameson Professor of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
1.10. Basic Molecular Neurobiology

Jeffrey H. Newcorn, M.D.
Associate Professor of Psychiatry and Pediatrics, Mount Sinai School of Medicine; Director of Child and Adolescent Psychiatry, Mount Sinai Medical Center, New York, New York.
23. Adjustment Disorders

Dorian S. Newton, Ph.D.

Affiliate Member, San Francisco Psychoanalytic Institute, San Francisco, California; Director of Mills College Counseling and Psychological Services, Oakland, California.
6.2. Erik H. Erikson

Peter M. Newton, Ph.D.
Professor of Psychology, Wright Institute, Berkeley, California.
6.2. Erik H. Erikson

Cynthia T. M. H. Nguyen, M.D.
Postdoctoral Research Fellow, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California.
51.4d. Geriatric Psychiatry: Psychopharmacology: Antianxiety Drugs

Linda M. Nicholas, M.D.
Assistant Professor of Psychiatry, University of North Carolina School of Medicine, Chapel Hill, North Carolina.
31.11. Bupropion

Steven L. Nickman, M.D.
Clinical Assistant Professor of Psychiatry, Harvard Medical School; Assistant in Psychiatry, Massachusetts General Hospital, Boston, Massachusetts; Assistant in Psychiatry, McLean Hospital, Belmont, Massachusetts.
49.2. Adoption

Ruta Nonacs, M.D., Ph.D.
Instructor of Psychiatry, Harvard Medical School, Boston, Massachusetts.
13.4. Postpartum Psychiatric Syndromes

Grayson S. Norquist, M.D., M.S.P.H.
Director of Division of Services and Intervention Research, National Institute of Mental Health, National Institutes of Health, Rockville, Maryland.
12.2. Schizophrenia: Epidemiology

Ann E. Norwood, M.D.
Associate Professor of Psychiatry and Associate Chairman, Department of Psychiatry, Uniformed Services University of the Health Sciences F. Herbert School of Medicine, Bethesda, Maryland.
30.8. Brief Psychotherapy

H. George Nurnberg, M.D.
Professor of Psychiatry and Vice Chair for Clinical Programs, Department of Psychiatry, University of New Mexico School of Medicine; Medical Director, University of New Mexico Health Sciences Center, Albuquerque, New Mexico.
9.1. Classification of Mental Disorders

M. Kevin O'Connor, M.D.
Assistant Professor of Psychiatry, Mayo Foundation, Rochester, Minnesota.
25.8. Musculoskeletal Disorders

Stephanie S. O'Malley, Ph.D.
Associate Professor of Psychiatry, Director of Division of Substance Abuse Research, Yale University School of Medicine, New Haven, Connecticut.
31.21. Naltrexone

Mark Olfson, M.D., M.P.H.
Associate Professor of Clinical Psychiatry, College of Physicians and Surgeons of Columbia University; Attending Psychiatrist, New York State Psychiatric Institute, New York, New York.
5.3. Mental Health Services Research

Angel A. Otero-Ojeda, M.D.
Professor of Psychiatry, Havana University; Chairman, Executive Committee, Cuban Glossary of Psychiatry, Havana, Cuba.
9.2. International Psychiatric Diagnosis

Michael J. Owens, Ph.D.
Associate Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia.
1.6. Neuropeptides: Biology and Regulation;
31.22. Nefazodone

Ken A. Paller, Ph.D.
Assistant Professor of Psychology, Northwestern University, Evanston, Illinois.
3.4. Biology of Memory

Laszlo A. Papp, M.D.
Associate Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Director of Biological Studies Unit, New York State Psychiatric Institute, New York, New York; Director of Anxiety Disorders Program, Long Island Jewish-Hillside Medical Center, Queens, New York.
15.7. Anxiety Disorders: Somatic Treatment

Carmine M. Pariante, M.D.
Medical Research Council Clinical Training Fellow, Section of Clinical Neuropharmacology, Institute of Psychiatry, London, United Kingdom.
1.12. Immune System and Central Nervous System Interactions

Barbara L. Parry, M.D.
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Director of Psychiatric Emergency Room, UCSD Medical Center, San Diego, California.
28.2. Psychiatry and Reproductive Medicine

Caroly S. Pataki, M.D.
Associate Clinical Professor of Psychiatry and Biobehavioral Sciences and Associate Director of Training and Education for Child and Adolescent Psychiatry, University of California at Los Angeles School of Medicine; Attending Psychiatrist, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
32.1. Child Psychiatry: Introduction and Overview;
45. Mood Disorders and Suicide in Children and Adolescents

Robert H. Paul, Ph.D.
Fellow, Department of Psychiatry and Behavioral Sciences, University of Oklahoma College of Medicine, Oklahoma City, Oklahoma.
2.7. Neuropsychiatric Aspects of Multiple Sclerosis and Other Demyelinating Disorders

Bradley D. Pearce, Ph.D.
Assistant Professor of Psychiatry and Behavioral Sciences, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta,

Georgia.
1.12. Immune System and Central Nervous System Interactions

David Pelcovitz, Ph.D.
Clinical Associate Professor of Psychology in Psychiatry, New York University School of Medicine, New York, New York; Chief Psychologist, Division of Child and Adolescent Psychiatry, North Shore University Hospital, Manhasset, New York.
49.10. Child Psychiatry: Identity Problem and Borderline Disorders

William H. Pelham, Ph.D.
Professor of Psychology and Director of Clinical Training, State University of New York at Buffalo, Buffalo, New York.
30.2. Behavior Therapy

Bradley S. Peterson, M.D.
House Jameson Assistant Professor in Child Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
33. Psychiatric Examination of the Infant, Child, and Adolescent

John M. Petitto, M.D.
Associate Professor of Psychiatry, Neuroscience, and Pharmacology, University of Florida College of Medicine, Gainesville, Florida.
25.10. Behavior and Immunity

Cynthia R. Pfeffer, M.D.
Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell University, New York, New York.
48.10. Psychiatric Treatment of Adolescents

Kemuel L. Philbrick, M.D.
Assistant Professor of Psychiatry, Mayo Foundation, Rochester, Minnesota.
25.8. Musculoskeletal Disorders

Edmond Hsin-tung Pi, M.D.
Executive Vice Chair, Department of Psychiatry, King/Drew University of Medicine and Science, Los Angeles, California.
31.4. Medication-Induced Movement Disorders

John Piacentini, Ph.D.
Assistant Professor-in-Residence of Psychiatry and Biobehavioral Science, University of California at Los Angeles School of Medicine; Director of Child and Adolescent Obsessive-Compulsive Disorder and Anxiety Program, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.
46.1. Obsessive-Compulsive Disorder in Children

Daniel S. Pine, M.D.
Associate Professor of Clinical Psychiatry, Division of Child and Adolescent Psychiatry, College of Physicians and Surgeons of Columbia University; Attending Psychiatrist, New York State Psychiatric Institute, New York, New York.
15.6. Anxiety Disorders: Clinical Features

Bruce G. Pollock, M.D., Ph.D.
Professor of Psychiatry and Pharmacology, University of Pittsburgh School of Medicine; Director of Geriatric Psychopharmacology Program, Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania.
51.4b. Geriatric Psychiatry: Psychopharmacology: General Principles

Harrison G. Pope, Jr., M.D.
Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts; Chief of Biological Psychiatry Laboratory, McLean Hospital, Belmont, Massachusetts.
11.13. Anabolic-Androgenic Steroid Abuse;
31.7b. Valproate

Robert M. Post, M.D.
Chief of Biological Psychiatry Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.
14.8. Mood Disorders: Treatment of Bipolar Disorders;
31.13. Calcium Channel Inhibitors

Karl H. Pribram, M.D., Ph.D.
Director of Center for Brain Research and Informational Sciences, Radford University, Radford, Virginia; Professor Emeritus, Stanford University, Stanford, California.
3.5. Brain Models of Mind

Patricia N. Prinz, Ph.D.
Professor of Biobehavioral Nursing and Health Systems, University of Washington School of Nursing; Adjunct Professor of Psychiatry, University of Washington School of Medicine, Seattle, Washington.
51.3b. Geriatric Psychiatry: Sleep Disorders

David B. Pruitt, M.D.
Professor of Psychiatry and Vice Chairman, Department of Psychiatry, University of Tennessee at Memphis College of Medicine, Memphis, Tennessee.
48.7. Child Psychiatry: Partial Hospital and Ambulatory Behavioral Health Services

Andres J. Pumariega, M.D.
Professor of Psychiatry and Chair, Department of Psychiatry and Behavioral Sciences, James H. Quillen College of Medicine, East Tennessee State University, Johnson City, Tennessee.
48.9. Child Psychiatry: Community-Based Treatments

Frank W. Putnam, M.D.
Chief of Unit on Developmental Traumatology, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland; Professor of Pediatrics, Ohio State University College of Medicine and Public Health; Professor of Psychiatry, University of Cincinnati College of Medicine, Cincinnati, Ohio.
18.3. Dissociative Identity Disorder

Denisa Radoi-Andraous, M.D.
Assistant Professor of Psychiatry, University of Montreal Faculty of Medicine, Montreal, Quebec, Canada.
31.24. Selective Serotonin-Noradrenaline Reuptake Inhibitors

Scott L. Rauch, M.D.
Associate Professor of Psychiatry, Harvard Medical School; Associate Chief of Psychiatry for Neuroscience Research, Massachusetts General Hospital, Boston, Massachusetts.
31.32. Neurosurgical Treatments

Darrel A. Regier, M.D., M.P.H.
Associate Director for Epidemiology and Health Policy Research, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland; Clinical

Professor of Psychiatry, Georgetown University School of Medicine and Health Sciences, Washington, D.C.
5.1. Epidemiology

Edward L. Reilly, M.D.
Professor of Psychiatry and Director of Residency Training, University of Texas Medical School at Houston; Director of Mental Sciences Institute, University of Texas-Houston Health Science Center, Houston, Texas.
1.14. Applied Electrophysiology

Eric M. Reiman, M.D.
Professor, Associate Head of Psychiatry
University of Arizona College of Medicine, Tucson, Arizona; Scientific Director of Positron Emission Tomography Center, Good Samaritan Regional Medical Center; Director of Arizona Alzheimer's Disease Research Center, Phoenix, Arizona.
51.2e. Geriatric Psychiatry: Neuroimaging: Overview

Perry F. Renshaw, M.D., Ph.D.
Research Director, Brain Imaging Center, McLean Hospital, Boston, Massachusetts.
1.16. Principles of Neuroimaging: Magnetic Resonance Techniques

Victor I. Reus, M.D.
Professor of Psychiatry, University of California at San Francisco School of Medicine; Director of Psychiatry, Langley Porter Psychiatric Institute, San Francisco, California.
1.11. Psychoneuroendocrinology

Charles F. Reynolds, III, M.D.
Professor of Psychiatry, University of Pittsburgh School of Medicine; Director of Mental Health Clinical Research Center for the Study of Late-Life Mood Disorders, Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania.
51.4c. Geriatric Psychiatry: Psychopharmacology: Antidepressants and Mood Stabilizers

Robert G. Robinson, M.D.
The Paul W. Penningroth Professor of Psychiatry and Head, Department of Psychiatry, University of Iowa College of Medicine, Iowa City, Iowa.
2.2. Neuropsychiatric Aspects of Cerebrovascular Disorders;
2.3. Neuropsychiatric Aspects of Brain Tumors;
2.5. Neuropsychiatric Aspects of Traumatic Brain Injury

Robert M. Rohrbaugh, M.D.
Assistant Clinical Professor of Psychiatry, Yale University School of Medicine, New Haven, Connecticut; Director of Education in Psychiatry, Connecticut Veterans Affairs Healthcare System, West Haven, Connecticut.
51.3g. Geriatric Psychiatry: Personality Disorders

Richard B. Rosse, M.D.
Associate Professor of Psychiatry, Georgetown University School of Medicine and Health Sciences; Chief of Georgetown University Teaching Unit, Veterans Affairs Medical Center, Washington, D.C.
7.7. Medical Assessment and Laboratory Testing in Psychiatry

Bruce J. Rounsaville, M.D.
Professor of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
31.21. Naltrexone

Alec Roy, M.D.
Professor of Psychiatry, New Jersey Medical School, University of Medicine and Dentistry of New Jersey, Newark, New Jersey; Assistant Chief of Psychiatry for Substance Abuse, East Orange Campus of the Veterans Affairs New Jersey Health Care System, East Orange, New Jersey.
29.1. Suicide

Teresa A. Rummans, M.D.
Associate Professor of Psychiatry, Mayo Foundation, Rochester, Minnesota.
25.8. Musculoskeletal Disorders

A. John Rush, M.D.
Professor of Psychiatry, Betty Jo Hay Distinguished Chair in Mental Health, and Rosewood Corporation Chair in Biomedical Science, University of Texas Southwestern Medical School, Dallas, Texas.
14.7. Mood Disorders: Treatment of Depression;
30.6. Cognitive Therapy

Sheila Ryan, C.S.W.
Program Director of Special Needs Clinic, Pediatric Psychiatry, Columbia-Presbyterian Medical Center, New York, New York.
49.6. Child Psychiatry: Psychiatric Sequelae of HIV and AIDS

Joel Sadavoy, M.D., F.R.C.P.(C)
Associate Professor of Psychiatry and Head, Division of General Psychiatry, University of Toronto Faculty of Medicine; Psychiatrist-in-Chief, Mount Sinai Hospital; Clinical Director, Joint General Psychiatry Program, Mount Sinai Hospital and Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
51.4h. Geriatric Psychiatry: Psychosocial Treatments: General Principles
51.4i. Geriatric Psychiatry: Individual Psychotherapy

Benjamin J. Sadock, M.D.
Menas S. Gregory Professor of Psychiatry and Vice Chairman, Department of Psychiatry, New York University School of Medicine; Attending Psychiatrist, Tisch Hospital; Attending Psychiatrist, Bellevue Hospital Center; Consultant Psychiatrist, Lenox Hill Hospital, New York, New York.
7.2. Psychiatric Report and Medical Record;
7.3. Signs and Symptoms in Psychiatry

Virginia A. Sadock, M.D.
Clinical Professor of Psychiatry and Director of Program in Human Sexuality and Sex Therapy, New York University School of Medicine; Attending Psychiatrist, Tisch Hospital; Attending Psychiatrist, Bellevue Hospital Center, New York, New York.
19.1a. Normal Human Sexuality and Sexual Dysfunctions

Rafael J. Salin-Pascual, M.D., Ph.D.
Professor of Psychiatry and Physiology, Department of Physiology, Universidad Nacional Autonomona de Mexico, Mexico City, Mexico.
1.19. Basic Science of Sleep

Steven C. Samuels, M.D.
Assistant Professor of Psychiatry, Mount Sinai School of Medicine; Director of Outpatient Geriatric Psychiatry Program, Department of Psychiatry, Mount Sinai Medical Center, New York, New York.

31.15. Cholinesterase Inhibitors

Ofra Sarid-Segal, M.D.
Assistant Professor of Psychiatry, Boston University School of Medicine; Staff Psychiatrist, Department of Veterans Affairs Outpatient Clinic, Boston, Massachusetts.
11.12. Sedative-, Hypnotic-, or Anxiolytic-Related Abuse

Stephen M. Saunders, Ph.D.
Assistant Professor of Psychology, Marquette University, Milwaukee, Wisconsin.
30.11. Evaluation of Psychotherapy

S. Alan Savitz, M.D.
President and Chief Executive Officer, PacifiCare Behavioral Health, Laguna Hills, California.
51.5b. Geriatric Psychiatry: Managed Care

Alan F. Schatzberg, M.D.
Kenneth T. Norris Jr., Professor of Psychiatry and Chairman, Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California.
31.33. Other Pharmacological and Biological Therapies

Stephen C. Scheiber, M.D.
Adjunct Professor of Psychiatry, Northwestern University Medical School, Chicago, Illinois; Adjunct Professor of Psychiatry, Medical College of Wisconsin, Milwaukee, Wisconsin; Executive Vice President American Board of Psychiatry and Neurology, Deerfield, Illinois; Senior Attending Physician, Evanston Hospital, Evanston, Illinois.
53.1. Graduate Psychiatric Education

Diane H. Schetky, M.D.
Associate Clinical Professor of Psychiatry, University of Vermont College of Medicine, Burlington, Vermont; Attending Psychiatrist, Maine Medical Center, Portland, Maine.
49.13. Ethical Issues in Child and Adolescent Psychiatry

Lon S. Schneider, M.D.
Professor of Psychiatry, Neurology, and Gerontology, University of Southern California School of Medicine, Los Angeles, California.
51.4f. Geriatric Psychiatry: Psychopharmacology: Antidementia Drugs

Richard S. Schottenfeld, M.D.
Professor of Psychiatry, Director of Substance Abuse Residency Training
Yale University School of Medicine; Director of Substance Abuse Treatment Unit and Associate Clinical Director of Addiction Services, Connecticut Mental Health Center, New Haven, Connecticut.
31.23. Opioid Agonists

Marc A. Schuckit, M.D.
Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California; Director of Alcohol Research Center, San Diego Veterans Affairs Healthcare System, San Diego, California.
11.2. Alcohol-Related Disorders

Mary E. Schwab-Stone, M.D.
Associate Professor of Child Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
33. Psychiatric Examination of the Infant, Child, and Adolescent

Erich Seifritz, M.D.
Professor of Psychiatry, Psychiatric University Hospital, Basel, Switzerland.
1.19. Basic Science of Sleep

Alberto C. Serrano, M.D.
Professor of Psychiatry, University of Hawaii John A. Burns School of Medicine; Director of Consortium of Education and Training, Kapi'olani Medical Center for Women and Children, Honolulu, Hawaii.
48.4. Child Psychiatry: Group Psychotherapy

Richard P. Shank, Ph.D.
Senior Research Fellow, CNS Drug Discovery, Robert Wood Johnson Pharmaceutical Research Institute, Spring House, Pennsylvania.
1.5. Amino Acid Neurotransmitters

Edward R. Shapiro, M.D.
Medical Director and Chief Executive Officer, Austen Riggs Center, Stockbridge, Massachusetts; Director of Erik H. Erikson Institute for Education and Research of the Austen Riggs Center; Associate Clinical Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts.
30.9. Eriksonian Clinical Theory and Psychiatric Treatment

Peter A. Shapiro, M.D.
Associate Professor of Clinical Psychiatry, Columbia University College of Physicians and Surgeons; Assistant Director, Consultation-Liaison Psychiatry Service, Columbia-Presbyterian Medical Center, New York, New York.
25.4. Cardiovascular Disorders

M. Tracie Shea, Ph.D.
Associate Professor of Psychiatry and Human Behavior and Director of Clinical Assessment and Training Unit, Brown University School of Medicine; Staff Psychologist, Providence Veterans Affairs Medical Center, Providence, Rhode Island.
14.9. Mood Disorders: Psychotherapy

Javaid I. Sheikh, M.D., M.B.A.
Associate Professor of Psychiatry, Stanford University School of Medicine, Stanford, California; Chief of Psychiatry, Palo Alto Veterans Affairs Health Care System, Palo Alto, California.
51.4d. Geriatric Psychiatry: Antianxiety Drugs

Cleveland G. Shields, Ph.D.
Associate Professor of Family Medicine and Psychiatry, University of Rochester School of Medicine and Dentistry; Family Therapist, Highland Hospital; Family Therapist, Strong Memorial Hospital, Rochester, New York.
51.4k. Interventions and Consultation With Families of Older Adults

Daniel J. Siegel, M.D.
Associate Professor of Clinical Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Medical Director of Infant and Preschool Service, UCLA Neuropsychiatric Institute and Hospital, Los Angeles, California.

3.1. Perception and Cognition

Daphine Simeon, M.D.
Assistant Professor of Psychiatry and Director of Medical Student Education in Psychiatry, Mount Sinai School of Medicine, New York, New York.
18.5. Dissociative Disorders Not Otherwise Specified

Robert I. Simon, M.D.
Clinical Professor of Psychiatry and Director of Program in Psychiatry and Law, Georgetown University School of Medicine, Washington, D.C.
54.1. Legal Issues in Psychiatry

George M. Simpson, M.D.
Professor of Research Psychiatry and Director of Clinical Research, Department of Psychiatry and the Behavioral Sciences, University of Southern California School of Medicine, Los Angeles County and USC Medical Center, Los Angeles, California.
31.4. Medication-Induced Movement Disorders;
31.5. b-Adrenergic Receptor Antagonists;
31.6. Anticholinergics and Amantadine;
31.8. Antihistamines

Gary W. Small, M.D.
Professor of Psychiatry and Biobehavioral Sciences, Director of Geriatric Psychiatry and Psychology Fellowship Program, Director of Center on Aging, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.1a. Geriatric Psychiatry: Introduction;
51.3e. Geriatric Psychiatry: Alzheimer's Disease and Other Dementias

Buster D. Smith, M.D.
Clinical Associate in Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
51.6a. Geriatric Psychiatry: Psychiatric Aspects of Long-Term Care

Gerard P. Smith, M.D.
Professor of Psychiatry, The Joan and Sanford I. Weill Medical College of Cornell Univerity; Attending Psychiatrist (Behavioral Science), New York Hospital, New York, New York.
1.20. Appetite

Thomas E. Smith, M.D.
Assistant Professor of Psychiatry, New York, New York.
52.4. Psychiatric Rehabilitation

Virginia L. Smith-Swintosky, Ph.D.
Senior Scientist, CNS Drug Discovery, Robert Wood Johnson Pharmaceutical Research Institute, Spring House, Pennsylvania.
1.5. Amino Acid Neurotransmitters

Solomon H. Snyder, M.D.
Distinguished Service Professor of Neuroscience, Pharmacology, Molecular Sciences, and Psychiatry, Johns Hopkins University School of Medicine, Baltimore, Maryland.
1.21. Future Directions in Neuroscience and Psychiatry

Michael E. Spagna, Ph.D.
Associate Professor of Special Education, California State University at Northridge, Northridge, California.
35.1. Reading Disorder;
35.2. Mathematics Disorder;
35.3. Disorder of Written Expression and Learning Disorder Not Otherwise Specified;
36. Motor Skills Disorder: Developmental Coordination Disorder

David Spiegel, M.D.
Professor of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California.
30.3. Hypnosis

Herbert Spiegel, M.D.
Special Lecturer in Psychiatry, Columbia University College of Physicians and Surgeons, New York, New York.
30.3. Hypnosis

Larry R. Squire, Ph.D.
Professor of Psychiatry and Neurosciences, University of California at San Diego School of Medicine; Professor of Psychology, University of California at San Diego, La Jolla, California; Research Career Scientist, San Diego Veterans Affairs Healthcare System, San Diego, California.
3.4. Biology of Memory

Sergio E. Starkstein, M.D., Ph.D.
Director of Neuropsychiatry, FLENI, Buenos Aires, Argentina.
2.2. Neuropsychiatric Aspects of Cerebrovascular Disorders

Murray B. Stein, M.D.
Associate Professor of Psychiatry, University of California at San Diego School of Medicine, La Jolla, California.
25.9. Stress and Psychiatry

Terry S. Stein, M.D.
Professor of Psychiatry, Michigan State University College of Human Medicine, East Lansing, Michigan.
19.1b. Homosexuality and Homosexual Behavior

Marlene Steinberg, M.D.
Associate Professor of Psychiatry, University of Massachusetts Medical School, Worcester, Massachusetts; Research Affiliate, Department of Psychiatry, Yale University School of Medicine, New Haven, Connecticut.
18.1. Dissociate Amnesia;
18.4. Depersonalization Disorder

Hans Steiner, M.D.
Professor of Psychiatry and Behavioral Sciences, Stanford University School of Medicine; Director of Training, Division of Child Psychiatry and Human Development, Stanford, California.
40. Disruptive Behavior Disorders;
49.7. Childhood or Adolescent Antisocial Behavior

Alan Stoudemire, M.D.

Professor of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia.
25.1. History, Classification, and Current Trends in Psychosomatic Medicine

Eric C. Strain, M.D.
Associate Professor of Psychiatry and Behavioral Science, Johns Hopkins School of Medicine, Baltimore, Maryland.
11.4. Caffeine-Related Disorders

James J. Strain, M.D.
Professor of Psychiatry, Mount Sinai School of Medicine; Director of Division of Behavioral Medicine and Consultation-Psychiatry, Mount Sinai Medical Center, New York, New York.
23. Adjustment Disorders;
25.12. Consultation-Liaison Psychiatry

Joel E. Streim, M.D.
Associate Professor of Psychiatry, University of Pennsylvania School of Medicine; Director of Geriatric Psychiatry Fellowship Program, Ralston-Penn Center, Philadelphia, Pennsylvania.
51.6a. Geriatric Psychiatry: Psychiatric Aspects of Long-Term Care

Gregory M. Sullivan, M.D.
Postdoctoral Fellow, Department of Psychiatry, Columbia University College of Physicians and Surgeons; Clinical Assistant in Psychiatry, Columbia-Presbyterian Medical Center, New York, New York.
15.3. Anxiety Disorders: Biochemical Aspects

David L. Sultzer, M.D.
Associate Clinical Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine; Director of Gero-Neuropsychiatry Division, West Los Angeles Veterans Affairs Ambulatory Care Center, Los Angeles, California.
51.4e. Geriatric Psychiatry: Psychopharmacology: Antipsychotic Drugs

Norman Sussman, M.D.
Clinical Professor of Psychiatry, New York University School of Medicine; Director of Psychopharmacology Research and Consultation Service, Bellevue Hospital Center, New York, New York.
31.7c. Other Anticonvulsants;
31.16. Clonidine

Dragan M. Svrakic, M.D., Ph.D.
Assistant Professor of Psychiatry, Washington University School of Medicine, Saint Louis, Missouri.
24. Personality Disorders

Rex M. Swanda, Ph.D.
Adjunct Assistant Professor of Psychiatry, University of New Mexico School of Medicine; Staff Psychologist, Veterans Affairs Medical Center, Albuquerque, New Mexico.
7.4. Clinical Neuropsychology and Intellectual Assessment of Adults

Carrie Sylvester, M.D., M.P.H.
Associate Professor of Psychiatry, Behavioral Sciences, and Pediatrics and Director of Education in Child and Adolescent Psychiatry, Northwestern University School of Medicine; Attending Psychiatrist, Children's Memorial Hospital; Attending Psychiatrist, Northwestern Memorial Hospital, Chicago, Illinois.
46.3. Separation Anxiety Disorder and Other Anxiety Disorders

Kenneth Tardiff, M.D., M.P.H.
Professor of Psychiatry and Public Health, The Joan and Sanford I. Weill Medical College of Cornell University; Attending Psychiatrist and Medical Director of Payne Whitney Clinic, New York Hospital, New York, New York.
27.3. Adult Antisocial Behavior and Criminality

Laurence H. Tecott, M.D., Ph.D.
Assistant Professor of Psychiatry, University of California at San Francisco School of Medicine, San Francisco, California.
1.4. Monoamine Neurotransmitters

Michael E. Thase, M.D.
Professor of Psychiatry, University of Pittsburgh School of Medicine; Research Director of Division of General Adult Psychiatry, Western Psychiatric Institute and Clinic, Pittsburgh, Pennsylvania.
14.4. Mood Disorders: Neurobiology

Armin Paul Thies, Ph.D.
Associate Clinical Professor of Psychology, Yale University School of Medicine, New Haven, Connecticut.
33. Psychiatric Examination of the Infant, Child, and Adolescent

Mauricio Tohen, M.D., Dr.P.H.
Associate Clinical Professor of Psychiatry, Harvard Medical School, Boston, Massachusetts.
31.7a. Carbamazepine

Alan Trachtenberg, M.D., M.P.H.
Adjunct Associate Professor of Community Medicine and Health Care Sciences, George Washington University School of Medicine and Health Sciences, Washington, D.C.; Medical Director, Office of Pharmacologic and Alternative Therapies, Center for Substance Abuse Treatment, Substance Abuse and Mental Health Services Administration, United State Public Health Service, Rockville, Maryland.
28.8. Alternative and Complimentary Health Practices

Alfonso Troisi, M.D.
Professor of Psychopathology, University of Rome Tor Vergata, Rome, Italy.
4.3. Evolutionary Biology and Psychiatry

Manuel Trujillo, M.D.
Professor of Clinical Psychiatry and Vice Chairman of Psychiatry, New York University School of Medicine; Director of Psychiatry, Director of Division of Cultural Psychiatry, Bellevue Hospital Center, New York, New York.
4.4. Cultural Psychiatry

Roy E. Twyman, M.D.
Associate Professor of Neurology and Pharmacology, University of Utah School of Medicine, Salt Lake City; Director, Global CNS Research and Development, Robert Wood Johnson Pharmaceutical Research Institute, Spring House, Pennsylvania.
1.5. Amino Acid Neurotransmitters

Jurgen Unützer, M.D., M.P.H.

Assistant Professor-in-Residence, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.3a. Geriatric Psychiatry: Psychiatric Problems in the Medically Ill

Robert J. Ursano, M.D.
Professor of Psychiatry and Neuroscience and Chairman, Department of Psychiatry, Uniformed Services University of the Health Sciences F. Edward Hébert School of Medicine, Bethesda, Maryland.
30.8. Brief Psychotherapy

Jerome V. Vaccaro, M.D.
Associate Clinical Professor Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California.
51.5b. Geriatric Psychiatry: Managed Care

Bessel A. van der Kolk, M.D.
Professor of Psychiatry, Boston University School of Medicine, Boston, Massachusetts; Chief of Trauma Center, HRI Hospital, Brookline, Massachusetts.
28.7. Physical and Sexual Abuse of Adults

Daniel P. van Kammen, M.D., Ph.D.
Professor of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania; Director of Global Clinical Research and Development, Robert Wood Johnson Pharmaceutical Research Institute, Springhouse, New Jersey.
31.17. Dopamine Receptor Antagonists (Typical Antipsychotics);
31.26. Serotonin-Dopamine Antagonists

William W. Van Stone, M.D.
Emeritus Clinical Associate Professor of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, California; Chief of Treatment Services, Mental Health and Behavioral Sciences Services, Veterans Affairs Medical Center, Washington, D.C.
51.5c. Geriatric Psychiatry: Veterans Affairs Medical Centers and Psychogeriatric Services

Johan M. F. Verhulst, M.D.
Associate Professor of Psychiatry and Behavioral Sciences, University of Washington School of Medicine, Seattle, Washington.
26. Relational Problems

Jeff Victoroff, M.D.
Associate Professor of Clinical Neurology, University of Southern California School of Medicine, Los Angeles, California; Director of Neurobehavior, Co-Director of Geriatric Neurology Clinic, Director Near Medicine Clinic, Rancho Los Amigos Medical Center, Downey, California.
51.2b. Central Nervous System Changes With Normal Aging

Michael V. Vitiello, Ph.D.
Professor of Psychiatry and Behavioral Sciences, University of Washington School of Medicine; Adjunct Professor of Psychology, and Biobehavioral Nursing and Health Systems, University of Washington School of Nursing, Seattle, Washington.
51.3b. Geriatric Psychiatry: Sleep Disorders

Fred R. Volkmar, M.D.
Harris Associate Professor of Child Psychiatry, Pediatrics, and Psychology, Yale University School of Medicine, New Haven, Connecticut.
38. Pervasive Developmental Disorders

Thomas A. Wadden, Ph.D.
Professor of Psychology and Director of Weight and Eating Disorders Program, Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania.
25.3. Obesity

Thomas A. Wehr, M.D.
Chief of Clinical Psychobiology Intramural Research Program, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.
1.13. Chronobiology

Daniel R. Weinberger, M.D.
Chief of Clinical Brain Disorders Branch, National Institute of Mental Health, National Institutes of Health, Bethesda, Maryland.
1.3. Developmental Neurobiology

Myron F. Weiner, M.D.
Professor of Psychiatry and Vice Chair, Department of Psychiatry, University of Texas Southwestern Medical School; Chief of Geropsychiatry, University of Texas Southwestern Medical Center, Dallas, Texas.
6.3. Other Psychodynamic Schools

Myrna M. Weissman, Ph.D.
Professor of Epidemiology in Psychiatry, Columbia University College of Physicians and Surgeons; Chief of Division of Clinical and Genetic Epidemiology, New York State Psychiatric Institute, New York, New York.
15.2. Anxiety Disorders: Epidemiology;
30.7. Interpersonal Psychotherapy

Lawrence A. Welkowitz, Ph.D.
Assistant Professor of Psychology, Keene State College, Keene, New Hampshire.
15.8. Anxiety Disorders: Psychological Treatments

Paul H. Wender, M.D.
Distinguished Professor of Psychiatry, University of Utah School of Medicine, Salt Lake City, Utah.
39.2. Adult Manifestations of Attention-Deficit Disorders

Louis J. West, M.D.*
Professor of Psychiatry and Biobehavioral Sciences, University of California at Los Angeles School of Medicine, Los Angeles, California

28.9. Nonprofessional Therapies, Quacks, and Cults

Tonya Jo Hanson White, M.D.
Assistant Professor of Psychiatry, University of Iowa College of Medicine; Attending Psychiatrist, Eating Disorders Clinic, University of Iowa Hospital and Clinics, Iowa City, Iowa.
49.5. Children's Reaction to Illness and Hospitalization

Robert L. Williams, M.D.
Professor Emeritus of Psychiatry, Baylor College of Medicine, Houston, Texas.

21. Sleep Disorders

G. Terence Wilson, Ph.D.
Professor of Psychology, Rutgers University, Piscataway, New Jersey.
3.3. Learning Theory

Eve J. Wiseman, M.D.
Associate Professor of Psychiatry and Behavioral Sciences, University of Arkansas for Medical Sciences; Chief of Special Treatment Section, John L. McClellan Memorial Hospital, Little Rock, Arkansas.
51.3h. Geriatric Psychiatry: Drug and Alcohol Abuse

Steven S. Wolf, M.D.
Co-Director of Neurology Consultation Service, National Institute of Mental Health Neuroscience Center at Saint Elizabeth's Hospital, National Institutes of Health; Director of Neurology Service, District of Columbia Commission on Mental Health Services, Washington, D.C.
2.6. Neuropsychiatric Aspects of Movement Disorders

George E. Woody, M.D.
Clinical Professor of Psychiatry, University of Pennsylvania School of Medicine; Chief of Substance Abuse Treatment Unit, Philadelphia Veterans Affairs Medical Center, Philadelphia, Pennsylvania.
11.5. Cannabis-Related Disorders

Joseph C. Wu, M.D.
Associate Professor of Psychiatry, Clinical Director of Brain Imaging Center, University of California at Irvine College of Medicine, Irvine, California.
2.13. Neuroimaging in Clinical Practice

Lyman C. Wynne, M.D., Ph.D.
Professor of Psychiatry Emeritus, University of Rochester School of Medicine and Dentistry, Rochester, New York.
51.4k. Interventions and Consultation With Families of Older Adults

Joel Yager, M.D.
Professor of Psychiatry and Vice Chair for Education, Department of Psychiatry, University of New Mexico School of Medicine, Albuquerque, New Mexico.
8. Clinical Manifestations of Psychiatric Disorders

Alayne Yates, M.D.
Professor of Child and Adolescent Psychiatry, University of Hawaii John A. Burns School of Medicine; Director of Division of Child and Adolescent Psychiatry, Kapi'olani Medical Center for Women and Children, Honolulu, Hawaii.
49.9. Child Psychiatry: Gender Identity and Sexual Issues

William R. Yates, M.D.
Professor of Psychiatry and Family Medicine, Chairman of Department of Psychiatry, University of Oklahoma College of Medicine-Tulsa, Tulsa, Oklahoma.
25.2. Gastrointestinal Disorders

Kimberly A. Yonkers, M.D.
Associate Professor of Psychiatry, Obstetrics, and Gynecology, University of Texas Southwestern Medical School, Dallas, Texas.
28.3. Premenstrual Dysphoric Disorder

Carlos A. Zarate, Jr., M.D.
Assistant Professor of Psychiatry, University of Massachusetts Medical School; Director of Bipolar and Psychotic Disorders Program, University of Massachusetts Medical Center, Worcester, Massachusetts.
31.7a. Carbamazepine

Elsa M. Zayas, M.D.
Instructor, Director of Geriatric Consultation, Department of Psychiatry, Director of Partial Hospitalization Program, Saint Louis University School of Medicine, Saint Louis, Missouri.
51.4g. Geriatric Psychiatry: Electroconvulsive Therapy

Charles H. Zeanah, M.D.
Professor of Psychiatry and Chair of Division of Child and Adolescent Psychiatry, Tulane University School of Medicine, New Orleans, Louisiana.
44.1. Reactive Attachment Disorder of Infancy and Early Childhood

Sidney Zisook, M.D.
Professor of Psychiatry and Director of Residency Training, University of California at San Diego School of Medicine, La Jolla, California; Attending Psychiatrist, San Diego Veterans Affairs Healthcare System, San Diego, California.
28.5. Death, Dying, and Bereavement

Rebecca K. Zoltoski, Ph.D.
Assistant Professor of Biological Sciences, Illinois College of Optometry, Chicago, Illinois.
1.19. Basic Science of Sleep

Charles F. Zorumski, M.D.
Professor of Psychiatry and Neurobiology, Washington University School of Medicine; Attending Psychiatrist, Barnes-Jewish Hospital, Saint Louis, Missouri.
1.9. Basic Electrophysiology;
31.31. Electroconvulsive Therapy

Stephen R. Zukin, M.D.
Director of Division of Clinical and Services Research, National Institute on Drug Abuse, National Institutes of Health, Bethesda, Maryland; Clinical Professor of Psychiatry, Albert Einstein College of Medicine of Yeshiva University, Bronx, New York.
11.11. Phencyclidine (or Phencyclidine-like)-Related Disorders

---

*deceased

EDITORS

SENIOR CONTRIBUTING EDITOR

CONTRIBUTING EDITORS

SECTION EDITORS

Clinical Professor of Psychiatry
University of Maryland School of Medicine
Baltimore, Maryland

Samuel J. Keith, M.D.
Professor of Psychiatry and Chairman
Department of Psychiatry, University of New Mexico
School of Medicine Health Sciences Center
Albuquerque, New Mexico

Caroly S. Pataki, M.D.
Associate Clinical Professor of Psychiatry and Biobehavioral Sciences
Associate Director of Training and Education for Child and Adolescent Psychiatry, University of California at Los Angeles School of Medicine
Attending Psychiatrist
UCLA Neuropsychiatric Institute and Hospital
Los Angeles, California

Gary W. Small, M.D.
Professor of Psychiatry and Biobehavioral Sciences
Director of Geriatric Psychiatry and Psychology Fellowship Program
Director of Center on Aging
University of California at Los Angeles School of Medicine
Los Angeles, California

Alan, Stoudemire, M.D.
Professor of Psychiatry and Behavioral Sciences
Emory University School of Medicine
Atlanta, Georgia

---

*deceased

PREFACE

This seventh edition of *Kaplan and Sadock's Comprehensive Textbook of Psychiatry* is being published on the threshold of the twenty-first century. For over 30 years it has helped educate generations of psychiatrists, other physicians, and mental health professionals from all fields—psychology, social work, and nursing, among others. Its goal has always been to foster professional competence and ensure the highest quality of care based upon humanistic and scientific principles. The textbook has earned a reputation both in the United States and around the world, as a reliable, consistent, and accurate compendium of psychiatric knowledge. We believe this millennium edition meets the high standards set by its predecessors.

EDITORSHIP

The task of continuing the *Comprehensive Textbook of Psychiatry* without the collaboration of Harold I. Kaplan, M.D. (1927–1998), was made possible with the able assistance of Virginia A. Sadock, M.D., who was often referred to as the "unsung heroine" of this and other *Kaplan and Sadock* books. Her role was described in the preface of the last edition of the textbook as follows:

> We especially want to thank Virginia Alcott Sadock. M.D., Clinical Professor of Psychiatry and Director of Graduate Education in Human Sexuality at New York University School of Medicine. As in all our previous books, she has served as an assistant to the editors and actively participated in every editorial decision. Her enthusiasm, sensitivity, comprehension, and depth of psychiatric knowledge were of immeasurable importance to the editors. She has ably represented not only the viewpoint of women in medicine and psychiatry but has also made many contributions to the content of this textbook. We are deeply appreciative of her outstanding help and assistance.

She now joins as coeditor of the *Comprehensive Textbook of Psychiatry* for this edition and those to follow.

TEACHING SYSTEM

This textbook forms one part of a comprehensive system developed to facilitate the teaching of psychiatry and the behavioral sciences. At the head of the system is *Comprehensive Textbook of Psychiatry*, which is global in depth and scope. It is designed for and used by psychiatrists, behavioral scientists, and all workers in the mental health field. *Kaplan and Sadock's Synopsis of Psychiatry* is a relatively brief, highly modified, original, and current version useful for medical students, psychiatric residents, practicing psychiatrists, and mental health professionals. The *Concise Textbook of Clinical Psychiatry,* derived from the *Synopsis*, emphasizes clinical psychiatry and includes extensive case studies useful for students and clinical practitioners from all fields. Another part of the system is *Study Guide and Self-Examination Review for Kaplan and Sadock's Synopsis of Psychiatry,* which consists of over 1600 multiple-choice questions and answers including detailed case histories. It is designed for students of psychiatry and for clinical psychiatrists who require a review of the behavioral sciences and general psychiatry in preparation for a variety of examinations. The questions are modeled after and consistent with the format used by the United States Medical Licensing Examination. Other parts of the system are the pocket handbooks: *Pocket Handbook of Clinical Psychiatry, Pocket Handbook of Psychiatric Drug Treatment, Pocket Handbook of Emergency Psychiatric Medicine,* and *Pocket Handbook of Primary Care Psychiatry*. These books cover the diagnosis and treatment of psychiatric disorders, psychopharmacology, psychiatric emergencies, and primary care psychiatry, respectively, and are compactly designed and concisely written to be carried in the pocket of clinical clerks and practicing physicians, whatever their specialty, to provide a quick reference. Finally, *Comprehensive Glossary of Psychiatry and Psychology* provides simply written definitions for psychiatrists and other physicians, psychologists, students, other mental health professionals, and the general public.

Taken together, these books create a multiple approach to the teaching, study, and learning of psychiatry.

CHANGES IN THIS EDITION

New Contributors A tradition of inviting a certain number of new authors to write sections written by prior authors began with the second edition of the textbook. That was done for the same reasons as in other great textbooks of medicine—to ensure a fresh approach to each topic and to keep the *Comprehensive Textbook of Psychiatry* vital and current. Over 60 percent of the contributors to this edition are new. The editors are deeply grateful to the more than 1500 psychiatrists and behavioral scientists who contributed to previous editions, all of whom maintained the highest standards of scholarship. Many of their sections remain classics in the field and are accessible to the interested reader. We especially wish to thank John Nemiah, M.D., editor emeritus of the *American Journal of Psychiatry*, for agreeing to write the Introduction. He has contributed to every edition of this book since its beginning.

Major Changes in This Edition Almost every chapter in this edition has been completely rewritten or revised, and many new sections have been added. Some of the major additions to the text-book and other highlights are listed below.

Neural Science, Neuropsychiatry, and Behavioral Neurology A major expansion in the chapters covering neural science, neuropsychiatry, and behavioral neurology reflects the rapid advances in these fields. New sections include Developmental Neurobiology, Neurotrophic Factors, Appetite, and expanded sections on radiotracing imaging and magnetic resonance imaging. Also added to this edition are sections on Neuropsychiatric Aspects of Neuromuscular Diseases, Neuropsychiatric Aspects of Traumatic Brain Injury, Neuropsychiatric Aspects of Brain Tumors, and a special section on Psychiatric Aspects of Child Neurology. Recent advances are included in Genetic Linkage Analysis of Psychiatric Disorders, Neuropsychiatric Aspects of HIV and AIDs, and Neuropsychiatric Aspects of Other Infectious Diseases. These sections chapters were completely revised and updated.

Psychological, Sociocultural, and Experimental Sciences The sections Perception and Cognition, Learning Theory, and Brain Models of Mind were extensively revised. Recent advances are included in Biology of Memory and a new section Emotional Intelligence has been added. Evolutionary Biology and Psychiatry is included in a separate section for the first time. Health Services Research and Outcomes Research in Psychiatry was added to the chapter on experimental methods, and the section Epidemiology is thoroughly updated. The chapter Theories of Personality and Psychopathology has been revised, and the section Psychoanalysis is updated to reflect changing theories in the field.

Clinical Psychiatry A newly written version of Psychiatric Interview, History, and Mental Status was prepared for this edition. The section Psychiatric Report was expanded to include the medical record and third-party documentation requirements. Signs and Symptoms in Psychiatry was revised to include a glossary of psychiatric terms. A new section, Computer-Based Testing of the Psychiatric Patient, is included in this edition.

Anxiety Disorders is greatly expanded with newly written sections that include Introduction and Overview, Epidemiology, Biochemical Aspects, Genetics, Psychodynamic Aspects, Clinical Features, Somatic Treatment, and Psychological Treatments.

The chapter Substance-Related Disorders was expanded to include the separate sections Caffeine-Related Disorders and Nicotine-Related Disorders and a new section Anabolic-Androgenic Steroid Abuse. The treatment of depressive disorders has been enlarged to include the separate sections Psychotherapy, Treatment of Bipolar Disorders, and Treatment of Depressive Disorders. The dissociative disorders are now covered in five new sections: Dissociative Amnesia, Dissociative Fugue. Dissociative Identity Disorder (formerly called multiple personality disorder), Depersonalization Disorder, and Dissociative Disorders Not Otherwise Specified. The coverage of attention-deficit/hyperactivity disorder now includes a new section. Adult Manifestations of Attention-Deficit/Hyperactivity Disorder. Other new sections include Primary Care Psychiatry and Cults, Quacks, and Nonprofessional Therapies. A new section, Alternative and Complementary Health Practices, is included for the first time in a major psychiatric textbook.

All sections dealing with clinical disorders have been thoroughly updated and follow a similar outline, which includes an introduction and definition of the disorder; a history of the disorder including comparative nosology, epidemiology, and etiology; diagnosis and clinical features; pathology and laboratory examination; differential diagnosis; and course and prognosis. Treatment strategies for all clinical disorders are presented eclectically to include biological, pharmacological, psychosocial, and psychotherapeutic approaches. The area of psychiatric treatment has been expanded with the addition of three new sections: Interpersonal Therapy. Eriksonian Clinical Theory and Psychiatric Treatment, and Combined Psychotherapy and Psychopharmacotherapy.

Because of the increased importance of certification and qualifications of physicians generally, the editors included a new section. Examining Psychiatrists and Other Trainees. The chapter Ethics and Forensic Psychiatry is completely updated and a new section, World Aspects of Psychiatry, was written to reflect the fact that mental

illness is a worldwide problem that requires global solutions.

A new section on Chronic Pain and the Placebo Effect is included to reflect the editors' belief that psychiatrists be involved in the emerging clinical specialties of palliative care and pain control. In the Spring of 1999, the American Board of Psychiatry and Neurology (ABPN) and the American Board of Physical Medicine and Rehabilitation (ABPMR) joined the American Board of Anesthesiology (ABA) in recognition of pain management as an interdisciplinary subspecialty. The respective Boards have agreed on a single standard of certification.

Geriatric Psychiatry The chapters on geriatric psychiatry continue to expand in each edition, and we thank Lissy Jarvik, M.D., for her extraordinary help over the years as contributing editor of this section. With the assistance of Gary Small, M.D., she helped coordinate the content of this important subspecialty. Both the editors and the field of psychiatry owe her a debt of gratitude. We especially wish to thank Dr. Small for integrating the many sections in this chapter into a coherent whole. New sections written by new contributors in geriatric psychiatry for this edition include Special Issues in Neuroimaging, Psychopharmacology: General Principles, Antidepressants and Mood Stabilizers, Antianxiety Drugs, Antipsychotic Drugs, and Antidementia Drugs. Other new sections include Psychosocial Treatment: General Principles, Cognitive Behavior Therapy, Gender Issues, and the section Health Care Delivery Systems, which includes the separate subsections Medicare and Medicaid and Managed Care. All the geropsychiatry sections were revised and updated.

Child and Adolescent Psychiatry The editors owe a great debt to Dennis Cantwell, M.D. (1939–1997), who served as contributing editor to the child and adolescent psychiatry section. He was responsible for organizing and inviting the contributors to this chapter—all of whom he valued as experts in their respective fields. The editors wish to dedicate this area of the *Comprehensive Textbook of Psychiatry* to his memory and as a testimonial to this outstanding and much admired child psychiatrist. Dr. Cantwell had been assisted by Caroly Pataki, M.D., who was able to step into the breach to complete the work he had begun. We thank her deeply for her efforts.

Many new sections covering child and adolescent psychiatry were added to this edition. New sections written by new contributors include Obsessive-Compulsive Disorder in Children, Posttraumatic Stress Disorder in Children, Short-term Psychotherapy, Cognitive and Behavioral Therapy, Family Therapy, Psychiatric Sequela of HIV and AIDS, Dissociative Disorders, Gender Identity and Sexual Issues, Adolescent Substance Abuse, Forensic Child and Adolescent Psychiatry, Ethical Issues in Child and Adolescent Psychiatry, School Consultation, Community-Based Treatments, and Psychiatric Prevention.

Psychopharmacology The editors continue to use the unique format of discussing drugs used in the treatment of mental disorders on a pharmacological basis rather than under the rubric of antidepressant, antipsychotic, and the like. Thus the clinician can use a psychotherapeutic drug on the basis of its pharmacological activity, which may make it equally effective in depression, anxiety, and other disorders as well as being specific for a particular condition.

A thoroughly updated section General Principles of Psychopharmacology was written for this edition, and two new sections, Drug Development and Approval and Pharmacokinetics and Drug Interactions, were added. New sections on classes of drugs with unique pharmacological properties were written, including Cholinesterase Inhibitors and Other Anticonvulsants. The most recently developed drugs are covered in detail, and all discussions of other drugs have been thoroughly updated.

New Format This seventh edition uses color for the first time to highlight figures, tables, and case studies and to help differentiate the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) and fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) diagnostic criteria. Like all *Kaplan and Sadock* books, this edition includes color plates of major psychiatric drugs and their various dosage forms including those recently released. This edition is heavily illustrated; photographs enhance the learning experience and keep the reader from being lost in a sea of type.

Following the style of other major medical textbooks, internal literature citations were eliminated, and the number of references at the end of each section reduced. Contributors were asked to limit themselves to 30 to 40 major books, monographs, and review articles and to include current references where possible; thus some citation lists are not as long as some of the authors would have wished. In addition, 1999 references were added by the editors to alert the reader to the most current literature. Contributors were also asked to note the five most important references with an asterisk.

Case histories are cited extensively to add clarity to the clinical disorders. They are derived from the DSM-IV and ICD-10 casebooks and the clinical and research experience of the contributors. We wish to thank the American Psychiatric Association (APA) and the World Health Organization (WHO) for permission to use some of their material.

DSM-IV AND ICD-10

In 1994 DSM-IV was published by the APA. DSM-IV contains the official nomenclature used by psychiatrists and other mental health professionals; thus DSM-IV terminology is used throughout the *Comprehensive Textbook*. In the year 2000, according to treaties between the United States and WHO, DSM-IV classifications must be identical with those used by ICD-10. Accordingly, this is the first U.S. textbook to include the definitions and diagnostic criteria of mental disorders used in ICD-10. Readers can find tables comparing the DSM-IV and ICD-10 diagnostic criteria in Chapter 10.

The psychiatric disorders discussed in this textbook are consistent with the nosology in DSM-IV; however, some of our contributors maintain reservations about the changes introduced into the various editions of DSM. In several sections of the book the reader will find these objections clearly stated. DSM-IV is a diagnostic and statistical manual; it is *not* and never claimed to be a textbook. Unfortunately, it is used as a text by some groups, including insurance companies who believe it to be a comprehensive source of information about mental illness.

As future editions of DSM appear—and the editors believe they are in the offing—the *Comprehensive Textbook of Psychiatry* will continue to allow room for dissent, before and especially after every new edition of the manual appears. It will continue to provide a forum for discussion, evaluation, criticism, and disagreement, while duly acknowledging the official nomenclature.

THE CONTINUING CRISIS IN THE FUTURE OF PSYCHIATRY

The last edition of the *Comprehensive Textbook of Psychiatry,* published in 1995, included the following commentary on the crisis of the future of psychiatry. This crisis continues and is far from resolution.

The introduction of the American Health Security Bill (the Clinton plan) in 1993 served as a catalyst for dramatic change in the delivery of health care in the United States even though the bill was not enacted into law. In the vanguard of change were the insurance companies and the health maintenance organizations (HMOs), which are, in the main, managed care programs run by large profit-seeking corporations. Managed care has had serious and adverse effects on the practice of psychiatry. For example, most managed mental health care (MMHC) plans restrict the number of outpatient visits for psychotherapy to a small and unpredictable number of sessions, usually 5 to 20 a year. Although some types of psychotherapy can be conducted within that framework, other types (insight oriented) require frequent visits over an extended period. Before a patient can be referred to a psychiatrist, many HMOs require that the patient see a primary care physician (the so-called gatekeeper), sometimes for several weeks; during that time, the doctor may prescribe pharmacotherapy about which he or she may have limited knowledge. Drugs, rather than psychotherapy, become the treatment of choice even though many studies have found the superior efficacy of psychotherapy used in conjunction with drugs in the treatment of most mental disorders, particularly depressive disorders and schizophrenia. Persons who are emotionally well make fewer general medical visits than do persons with emotional disorders. Providing timely psychotherapy results in savings in the overall cost of general medical care.

Many HMOs require preauthorization by a panel of so-called behavioral health experts. This panel requires information about the intimate and private details of a person's life to authorize therapy. If the patient or doctor refuses to comply, permission for psychiatric treatment is usually denied. And even if the patient is permitted to enter therapy, the psychiatrist must send frequent written reports to the HMO about the treatment, which breaches the confidentiality and trust of the doctor-patient relationship. Patients usually must be treated by psychiatrists who are enrolled in their particular HMO. They forfeit the right to see a doctor of their own choosing. In the traditional fee-for-service system patients can seek treatment from any psychiatrist they choose and can seek a second or even a third opinion if they so desire. In an HMO the patient does not have these options. Capitation, another method of payment used by HMOs, is untested in psychiatry and may mean "de-capitation" of the field.

HMOs use from 15 to 30 percent of their revenues to pay for marketing, administration, and the distribution of profits to owners and investors—money that

would otherwise be available for clinical care, research, and medical education. Health care in America is being “corporatized,” and HMOs reap profits by often eliminating laboratory tests, referrals to specialists, and reducing length of hospital stay to questionable and dangerous proportions. For example, patients with major psychiatric disorders are being forced out of the hospital, often against their will and against the recommendation of their psychiatrists. HMOs also increase their profit margin by paying lower fees to doctors, and since the HMOs control the supply of patients, price control rules the system.

The issue of financial liability is another area of danger to doctors who work for HMOs. Psychiatrists (and other physicians) who sign contracts with HMOs must agree to accept complete liability if any adverse effects to the patient occur during the course of treatment. Consider this example: A psychiatrist wants to hospitalize a potentially suicidal patient, but the HMO refuses to pay for hospitalization or limits the number of days allowed in the hospital to fewer than the psychiatrist deems necessary. The psychiatrist can be sued for malpractice if the patient ultimately commits suicide because of premature hospital discharge mandated by the HMO. The HMO accepts no liability for any adverse outcome based on their decisions. The only alternative is for the psychiatrist to treat the patient for no fee or for the patient to pay for treatment out-of-pocket. Neither option is satisfactory.

Currently, the future of psychiatric treatment is of concern. Unfortunately, prejudice toward mental illness still exists in many quarters—political policy makers, insurance companies, the general public, and, sadly, the medical profession itself. Psychiatry and medicine are at a crossroad. It would be tragic to take the path that discards and negates the humanism that psychiatry has brought to medicine and the great advances that have been made over the past hundred years by Sigmund Freud and other great psychiatric clinicians and researchers.

A new concept of medical services as *market driven* now dominates the health care industry and will do so for the foreseeable future. Paradoxically, the role of government must increase to regulate this new industry whose preoccupation is the *cost* of health care. For example, some states recently passed legislation allowing patients to sue health maintenance organizations (HMOs). Ultimately the U.S. Congress will become the arbiter between the consumers of health care (patients), the providers (physicians and other health professionals), and the payers (insurance companies and HMOs). In this sense, society and the body politic will determine the nature and quality of health care in the United States.

Physicians, especially psychiatrists, have a special obligation to be informed about sociopolitical issues affecting the physical and psychological well-being of their patients. The spirit of the Hippocratic oath written over 2000 years ago continues to inspire the ethics of the medical profession: *To act for the good of my patients according to my ability and my judgment*. As medicine changes, physicians (and other health care professionals) are the last stronghold for humanitarian and compassionate care that stresses the inherent dignity and worth of each person.

CONTRIBUTING AND SECTION EDITORS

In the preparation of this textbook we have been helped immensely by our distinguished section and contributing editors.

The section editors worked closely with the authors and suggested modifications when necessary. We are deeply appreciative of their efforts. They include Hagop S. Akiskal, M.D., who covered the area of mood disorders; Jack M. Gorman, M.D., who covered anxiety disorders; Katherine A. Haaland, Ph.D., who covered neuropsychological assessment; Jerome Jaffe, M.D., who covered substance-related disorders; Samuel Keith, M.D., who covered schizophrenia and other psychotic disorders; Caroly Pataki, M.D., who covered child and adolescent disorders; Gary W. Small, M.D., who covered geriatric psychiatry; and Alan Stoudemire, M.D., who covered psychosomatic medicine.

The contributing editors kept us apprised of new advances in the field and helped obtain contributors with expertise in their respective areas. We thank them for their help and cooperation. They include Glen Gabbard, M.D., who contributed to the areas of psychodynamics and psychoanalysis; Lissy Jarvik, M.D., who contributed to geriatric psychiatry; Joel Yager, M.D., who contributed to clinical psychiatric disorders; and the late Dennis Cantwell, M.D., who contributed to child and adolescent psychiatry. Our special thanks are extended to Jack Grebb, M.D., a past coauthor of *Kaplan and Sadock’s Synopsis of Psychiatry* who, as contributing editor, assisted us in many editorial decisions, particularly in the area of biological psychiatry.

Together this admirable and distinguished group of men and women helped integrate an immense amount of material into a balanced and consistently styled work. The editors and the field of psychiatry owe them a debt of gratitude for their outstanding help.

ACKNOWLEDGMENTS

In addition to our contributing and section editors mentioned above, we want to extend our appreciation to Justin Hollingsworth, who served as project editor in New York and helped us immensely. He was ably assisted by Linda Kenevich. Margaret Cuzzolino, David Abramson, and Angela Barbuscia, all of whom worked with alacrity. Our editors Anne Schwartz and Radhika Rao Gupta provided invaluable assistance.

We especially want to thank our dear friend Nancy Barrett Kaplan, who gave us her complete support throughout this project. Over the years she has been an invaluable source of encouragement and understanding. Others who helped us were Norman Sussman, M.D., Victoria Sadock, M.D., James Sadock, M.D., Chris Marra, D.O., and Ben Hammer, M.D.

We take this opportunity to acknowledge those who have translated this and other *Kaplan and Sadock* books into foreign languages. Current translations include Croatian, French, German, Greek, Indonesian, Italian, Japanese, Polish, Portuguese, Russian, Spanish, and Turkish, in addition to a special Asian and international student edition.

The staff at Lippincott Williams & Wilkins was most efficient. We especially wish to thank Katey Millet, developmental editor, who was extraordinarily helpful and who has been an inspiration to us all. We also thank Susan Gay, former vice-president and publisher of Williams & Wilkins, for her encouragement and support over the years and we want to welcome Charley Mitchell.

Robert Cancro, M.D., Professor and Chairman of the Department of Psychiatry at New York University School of Medicine, participated as Senior Contributing Editor of this edition. Dr. Cancro’s commitment to psychiatric education and psychiatric research is recognized throughout the world. He has been a source of great inspiration to the editors and contributed immeasurably to this and previous editions. He is a much valued and highly esteemed colleague, and it is very special privilege to work closely with him. Dr. Cancro has developed a department that represents the very best in American psychiatry. Our collaboration and association with this outstanding American educator has contributed immeasurably to the ideas and directions shaping this textbook.

Finally, we want to express our deepest thanks to our contributors who were extraordinarily cooperative in every aspect of this textbook.

NYU School of Medicine, New York, New York

Benjamin James Sadock, M.D.

Virginia Alcott Sadock, M.D.

INTRODUCTION: TWO FACES OF PSYCHIATRY

JOHN C. NEMIAH, M.D.

Over 30 years have elapsed since the publication of the first edition of *Comprehensive Textbook of Psychiatry* (CTP-I). In comparison with its successors, it was a modest work of only some 1600 pages contained within the covers of a single volume. Equally striking is the difference in its contents from what the modern reader will find in this seventh edition of this now classic text. CTP-I, in line with the standard diagnostic manual of its day, referred to “psychiatric reactions” rather than to “disorders” and construed their causes and treatment in psychological terms. Twelve pages, for example, ere devoted to the “psychological treatment” of schizophrenia in contrast to a mere two-and-a-half to its “organic treatment.” “Serotonin” had but a single entry in the index, “dopamine” was not listed at all, and only a handful of pages were allotted to the budding clinical experience with the use of chloropromazine and the antidepressants. It needs only the most cursory glance at the pages of the latest edition to discover what dramatic changes 30 years have wrought. “Psychiatric reactions” are now full-fledged illnesses, their cause and treatment are seen as being primarily biological in nature, and psychological factors are minor actors on the clinical stage.

A shift of outlook of such magnitude occurs, of course, for many reasons. It is notable, however, that similar shifts have happened before in the development of psychiatric theory and practice, and the brief account that follows of a curious chapter of psychiatric history may, therefore, give us useful insights into the nature and implications of the changes that have taken place in our own era.

“The 11th of August,” wrote Frank Podmore in *From Mesmer to Christian Science*,

> should be observed as a day of humiliation in the civilised world, for on that date in 1784 a Commission, consisting of the most distinguished representatives of Science in the most distinguished capital in Europe, pronounced the rejection of a pregnant scientific discovery—a discovery possibly rivaling in permanent significance all the contributions to the physical sciences made by the two most famous members of the Commission—Lavoisier and Benjamin Franklin.

The oft-told events to which Podmore alludes may be quickly recounted. In the early 1770s the Viennese physician Franz-Anton Mesmer had developed a remarkable treatment for human illness that became a highly popular therapeutic procedure among the citizens of Paris after he established a clinical practice in that ity in 1778. Indeed, so thronged was Mesmer's clinic with patients from all levels of society that a commission of inquiry was established by the authorities to examine his procedures and their results.

Mesmer's explanation of the pathogenesis of human discases was based on his concept of the nature and function of what be called “animal magnetism.” There exists, he proposed, an invisble fluid spread throughout the universe with properties similar to mineral magnetism whose steady flow through the bodies of living beings sustains their healthy physiological functioning. However, as the result of a pathological blockage of that flow, a localized organ dysfunction occurs that is manifested in the symptoms of a specific bodily disease. Treatment was aimed at restoring the normal flow of the fluid by overcoming the blockage. That was accomplished by the production of a powerful surge of fluid in the patient's body by a magnetic therapist, who induced the flow of magnetic fluid by repetitious passes of his hands over the patient's thorax and abdomen. The same therapeutic effect could be produced in several patients at once by placing them in contact with the *baquet*, a large, circular wooden tub filled with “magnetized” water. The high drama of such a group treatment is evident in the commissioners' description of a typical clinical session in their final report.

> The patients are arranged in large numbers in several rows around the baquet and are exposed to the therapeutic flow of the magnetic fluid over several channels at once—by its transmission over the iron rods protruding from the baquet, by a long cord entwined around their bodies connecting the patients to one another, through the fluid's transmission to their neighbors by the mutual contact of their thumbs, and by the sounds of a piano or a pleasing voice that spreads it through the atmosphere. Patients are also magnetized directly when the magnetist passes his fingers or an iron wand over their faces, over the back or top of their heads, or over the diseased parts of their bodies. It is particularly to be noted that magnetization is produced by contact with the magnetist's hands as be applies pressure to the upper and lower abdomen for a long stretch of time, often several hours in duration.
>
> The patients present a spectacle of wide-ranging conditions and behavior. Some are calm and tranquil and have no untoward sensations. Other are coughing and spitting, or suffer from mild pains and complain of localized or generalized burning accompanied by bouts of sweating. Yet others are agitated and wracked by convulsions that are notable for their frequency, duration, and violence. Moreover, when one patient has a convulsion, many others follow suit. The Commissioners have witnessed some seizures that lasted more than three hours and were accompanied by the violent expectoration of a foul, viscous fluid, sometimes flecked with blood. The convulsions are further characterized by precipitous involuntary movements of all the limbs and the entire body, by a construction of the throat, by spasmodic jerks in the hypochondrium and the epigastric region, by a dimming and clouding of vision, by piercing shrieks, tears and hiccoughs, and by immoderate laughter. All of these events are preceded or followed by a state of languor, dreaminess, and prostration, and even by drowsiness.

After extensive observations of these remarkable phenomena the commissioners proceeded to investigate the nature of the animal magnetism alleged to be their cause. In a series of clever experiments, whose design might well be admired by modern investigators, they determined conclusively that there was no evidence whatsoever for the existence of a magnetic fluid, and that the dramatic, often wildly violent behavior attributed to it was entirely due to the effect of imagination and imitation on suggestible patients in a group setting.

The commissioners in the concluding paragraph of their report wrote.

> We have determined that the magnetic fluid cannot be detected by any of the five senses and that it has no effect whatsoever either on the Commissioners themselves or on the patients they exposed to it. They have ascertained that the practice of compressing or touching the body brings about changes that are rarely favorable and arouses excitations of the imagination that are invariably vexatious. Finally, they have demonstrated by definitive experiments that imagination without magnetism produces convulsions and that magnetism without imagination produces nothing. Therefore, the Commissioners have unanimously concluded regarding the question of the existence and the effectiveness of animal magnetism that there is no evidence whatsoever of a magnetic fluid and that such a non-existent fluid is consequently without useful effect. They have further determined that the observable violent results of public treatment are the product of touching, of an aroused imagination, and of an instinctive tendency automatically to imitate whatever strongly impresses the senses. At the same time, the Commissioners feel obliged to add the important observation that the practice of touching and the repetitious arousal of the imagination designed to bring on the magnetic crisis can be injurious. Furthermore, the public spectacle of these crises is equally dangerous because of that element of imitativeness that nature appears to have designed as a basic law. As a consequence, public treatment employing magnetism can in the long run have only disastrous effects.

Such, then, were the commissioners' conclusions and recommendations to which Podmore took such vigorous objection. It was not so much their denial of the existence of Mesmer's magnetic fluid that troubled him as the fact that having recognized the central importance of imagination and suggestion in producing the therapeutic convulsions (and the cures that frequently followed them), the commissioners then proceeded to dismiss magnetic phenomena as being the result of mere imagination and to condemn treatment as harmful. Mesmer's magnetic fluid, Podmore agreed,

> was a chimera and Mesmer . . . perhaps three parts a charlatan, [but] his claim to our remembrance lies in this—that he wrested the privilege of healing from the Churches and gave it to mankind as a universal possession. In rejecting that gift for themselves and their successors to the third and fourth generation, Bailly and his colleagues rejected more than they knew.

If Podmore's charges against the commissioners are perhaps a bit overblown, he does point to an arresting fact: Having skillfully demonstrated the fallacy of Mesmer's hypothesis of a physical magnetic fluid, the commissioners perceptively recognized the psychological basis for the patients' dramatic response to magnetic procedures and for the significant number of genuine cures that resulted from their application. But having made that discovery, far from being intrigued by such a remarkable finding and advising its further exploration, they dismissed it as not only being nugatory but as resulting in dangerous clinical practices that should be

curtailed.

The conclusions of the commissioners' report seem to have had little or no effect on the continued development of animal magnetism. Indeed, although the French Revolution temporarily slowed its progress, it became a vigorous movement throughout France in the early decades of the nineteenth century, spread beyond French borders to the rest of the Continent, crossed the Channel to England, and ultimately reached the United States.

The large number of individuals engaged in the clinical application of animal magnetism (or “mesmerism,” as it came to be called) ignored the commissioners' evidence against the existence of a magnetic fluid and held fast to Mesmer's mechanistic view that it was the active agent in the response to magnetic treatment. The discovery in 1785 by the Marquis de Puységur of the phenomenon of somnambulistic trance (which thereafter replaced seizures as the primary response to magnetic passes) did not alter adherence to the fluidic hypothesis. Nor was it in any way modified by the proposal of Alexandre Bertrand in the 1820s that the magnetic fluid was indeed nonexistent and that the effects of mesmeric treatment were entirely due to the psychological agency of suggestion. Not until the 1840s did the first serious challenge to the fluidic hypothesis arise.

In the late autumn of 1841 James Braid, an English surgeon in Manchester, attended a public demonstration of magnetism by the traveling French mesmerist, Charles Lafontaine. Impressed, despite his initial skepticism, by what he observed, Braid embarked upon a series of mesmeric experiments of his own. In the course of these he ultimately discovered that he could produce all the phenomena of mesmeric trance by merely arousing the idea of them in his subjects' minds. Hypnosis (as Braid termed it) was, he proposed, a process arising entirely within the hypnotic subject and set in motion by the hypnotist's verbal instructions. Braid's procedures made it clear that neither a magnetic fluid nor any other external physical agency was responsible for the observed phenomena, but that they resulted from those very psychological agencies of suggestion and imagination that had been so vigorously belittled and rejected by the French commission a half-century earlier.

Although Braid's studies made no immediate impact on his English colleagues, by a curious turn of fate his ideas returned across the Channel to France where they caught the attention of the French physician Eugene Azam. Azam is perhaps best remembered for his description of the phenomenon of “double consciousness” in his famous patient, Félida X, but he was also notable for the fact that, as Charcot wrote, “he was the first person in France to verify Braid's findings by his own experiments.” Indeed, Azam's publication in 1860 of a description of Braid's hypnotic experiments helped set the stage for three important developments in France during the ensuing decade.

1. Charcot's major interest in the nature and clinical application of hypnosis, which led to its extensive use at the Salpêtrière in Paris, gave the subject a legitimacy it had not had before. From being a phenomenon viewed by most of the orthodox scientific and medical world as mere quackery, if not outrightly fraudulent, with the blessing of a person of Charcot's academic and scientific reputation it became for the first time in its history a sanctioned procedure worthy of clinical application and serious study. The major thrust of Charcot's research was aimed at the clinical application of hypnosis to patients with major hysterical disorders. In this context, he came to the conclusion that hypnosis and hysteria were closely related phenomena whose manifestations were the result of poorly defined pathological processes in the central nervous system. Charcot, that is, held to a physical rather than a psychological view of the hypnotic process.
2. At the same time that Charcot's studies were under way in Paris, Hippolyte Bernheim in Nancy had joined with Ambroise Liábeault, a general practitioner in the neighboring countryside, in a study of the medical uses of therapeutic hypnotism. Influence by their reading of Bertrand and Braid. Liábeault and Bernheim advanced the hypothesis that the basic factor in the production of hypnotic phenomena was the subject's psychological state of suggestibility. Suggestibility, they held, was an entirely normal trait that was to be found to a greater or lesser degree in the vast majority of human beings; and it was, furthermore, the primary agency in affecting cures by hypnotic therapy. It is evident from the nature of their views that Liábeault and Bernheim, and the so-called Nancy School that gathered around them, were completely at odds with Charcot and the “Salpêtrière School” with regard to every aspect of therapeutic hypnosis. Nor is it surprising that a protracted controversy raged between them, ultimately to be settled in favor of Bernheim and his colleagues.
3. Although Pierre Janet came from a background of philosophical rather than medical studies, it is evident that he was thoroughly familiar with the literature of clinical hypnosis. “M. Despine,” he wrote in 1889 in his first major publication. *L'Automatisma Psychologioue*,

   has maintained that psychology has no application whatsoever to somnambulism and that it can be explained only by physiology. In fact, far from explaining it, it cannot even be defined in physiological terms. Many authors, such as Bertrand and Braid, and more recently Gurney and Bernheim, have distinguished themselves by having recognized that somnambulism is a psychological phenomenon and can be defined only by its uniquely psychological characteristics.

It was with this intellectual background and predilection for psychological concepts that in the early 1880s Janet embarked on the hypnotic investigation of hysterical patients that led to his important discovery of psychological dissociation and psychological automatism—a discovery that led him to formulate the psychological explanation of hysterical symptoms as originating from dissociated traumatic memories and to his use of hypnotic suggestion in their treatment.

Almost simultaneously with Janet, Sigmund Freud, stimulated by his observations of Charcot's work at the Salpêtrière and Bernheim's practice in Nancy, began his own investigation of hysterical patients in conjunction with Josef Breuer. Like Janet (with whose early studies he was familiar) Freud focused his attention on the role of unconscious traumatic memories in the production of hysterical symptoms. He went beyond Janet, however, in pointing to the important psychogenic role of the unconscious painful feelings associated with the traumatic memories and the necessity for their conscious expression in the course of therapy. With these original formulations Freud laid the groundwork for the concepts of psychological conflict and psychic structure that became the hallmark of psychoanalytic theory and practice.

While Janet and Freud were thus pursuing the medical aspects of dissociation and hypnosis, a less well known group of lay investigators, working in London under the aegis of the recently created Society for Psychical Research, were studying the nature and extent of hypnotic and dissociative phenomena in normal subjects. Particularly prominent among these investigators was Frederic W. H. Myers, one of the first of their contemporaries to recognize and publicize the significance of Freud's and Janet's early findings. Myers's own investigations led him to a demonstration of the presence of dissociative elements in a variety of unusual but not necessarily pathological phenomena, such as artistic creation, nonpsychotic hallucinations, automatic writing, medium trance-states, and spirit possession. His postulate of a “subliminal mind” expanded the scope and meaning of the process of dissociation far beyond the narrower pathological concept of clinical investigators. So impressed was William James by Myers's contributions that (as he wrote in a eulogy published after Myers's death in 1901) he “was disposed to think it a probability that Frederic Myers will always be remembered in psychology as the pioneer who staked out a vast tract of mental wilderness and planted the flag of genuine science upon it.”

James himself, of course, was a major player in the explorations of the reaches of the mind, and his *The Varieties of Religious Experience* remains a classic contribution to the psychology of religion. His profound interest in and commitment to the development of depth-psychology is evident throughout his writings. As he commented in the volume just mentioned,

> I cannot but think that the most important step forward that has occurred in psychology since I have been a student of that science is the discovery . . . that, in certain subjects at least, there is not only the consciousness of the ordinary field . . . but an addition thereto in the shape of a set of memories, thoughts, and feelings which are extra-marginal and outside of the primary consciousness altogether. . . . I call this the most important step forward because, unlike the other advances which psychology has made, this discovery has revealed to us an entirely unsuspected peculiarity in the constitution of human nature.

It is evident that as it evolved 100 years ago the study of psychology, both normal and abnormal, was satisfying, rewarding, and stimulating for those engaged in it. Moreover, they were optimistic about its future and looked forward confidently to further explorations and major discoveries in the still uncharted land of the unconscious. As the philosopher Henri Bergson wrote at the dawn of the twentieth century,

> To explore the most sacred depths of the unconscious, to labor in . . . the subsoil of consciousness, that will be the principal task of psychology in the century which is opening. I do not doubt that wonderful discoveries await it, as important perhaps as have been in the preceding centuries the discoveries of the physical and natural sciences.

Brave words! Bold predictions! And yet how widely off the mark they appear in the face of modern developments in psychiatry, especially on the North American continent as it approaches the new millennium. Our current major emphasis on the biological, neurological, and phenomenological aspects of psychiatric disorders is a far cry from the prominent concern with their psychological attributes among investigators 100 years ago. The briefest perusal of our contemporary psychiatric literature will readily indicate the extent of that difference. The space current major psychiatric publications allot to the elucidation of the unconscious, of psychological conflict, and of psychodynamic processes is miniscule compared with that devoted to the description of psychiatric syndromes and their neurobiological underpinnings. It is clear that there have been major changes in the conception of psychiatric illness over the course of the century now ending—changes whose

details are within the memory of many living psychiatrists and need be only briefly reviewed.

From the earlist years of the twentieth century. American psychiatry was significantly guided by two major influences—the psychobiological concepts of Adolf Meyer and the psychoanalytic theories of Freud. Both strongly underscored the vital importance of including psychosocial factors in the clinical understanding and treatment of psychiatric illnesses. The intellectual stimulation and the therapeutic optimism aroused by that humanistic approach brought psychiatrists away from their isolation in asylums for the insane back into the mainstream of medicine. By midcentury, the number of physicians choosing the practice of psychiatry was beginning to swell; general hospitals throughout the country were opening psychiatric inpatient units and outpatient clinics, and under the banners of “psychosomatic medicine” and “consultation-liaison psychiatry,” psychiatric clinicians and investigators were applying their psychological knowledge and procedures beyond the realm of psychiatric disorders to an exploration of the psychosocial aspects of the conventional medical and surgical illnesses as well.

At the height of this movement. Carl Binger, for many years the editor-in-chief of the journal *Psychosomatic Medicine*, wrote in *The Two Face of Medicine*,

> The principal contribution of psychiatry to medicine lies in its humanizing influence. It has kept man himself as the center of its concern rather than his enzymes or the chains of proteins of which he is composed. It has thus far managed to escape the automation that threatens to swallow up clinical medicine. . . . Our view is the holistic view. . . . We are less preoccupied with machine processed data than are those in other medical disciplines. The unprocessed reality of our world is . . . the human animal in his [inner] environment . . . which we approach through a study of personality. This latter has become of paramount importance not only for psychiatry but in the management of all sick people.

Binger's comments should make starkly clear the magnitude of the revolution that has occurred in the three decades since their utterance. The scientific advances in the knowledge of brain structure and function that have been achieved since that time, the development of effective psychopharmacotherapeutic agents, and the radical revision of the concept of psychiatric disorders and their diagnosis have now to a large degree supplanted attention to the clinical applications of psychodynamic understanding. However, no matter how strongly the biological view of psychiatric disease dominates the current scene, it constitutes only a cross-sectional view of a more extended historical process.

A look backward over the past 200 years reveals an interesting pattern in the evolution of ideas about mind and body and their role in the production of illness. Mesmer, of course, was not aware that his animal magnetism and the conditions he treated with it would evolve into the modern discipline of psychiatry. Indeed, he viewed his procedure as a contribution to the general practice of medicine and ascribed the therapeutic effect of his magnetic treatment to the direct mechanical action of a physical magnetic fluid on the patient's bodily organs. His materialistic hypothesis persisted as the accepted view for some 70 years until it was at last superseded by the psychological hypothesis of suggestion that emerged from the experiments and writings of James Braid. As the discipline of psychiatry gradually crystallized from those early clinical studies, it was dominated by psychological theories in one form or another for nearly a century until they were replaced by the biological hypotheses now in the ascendancy.

A historical perspective, in other words, reveals a curious alternation between psychological and biological explanations of illness. History, moreover, makes evident the fact that although one outlook may be dominant, it is not necessarily exclusive. On the contrary, its obverse is always to be found in a small but devoted minority. Indeed, throughout the two centuries that began with Mesmer, whatever the prevailing theoretical view of psychiatric illness, there has been, as Stanley Jackson has demonstrated, a constant demand for psychological understanding and treatment, while at the same time the quest for biological knowledge has continued at a steady pace.

History shows as well that there has always been a degree of antagonism between the proponents of the psychological and the biological view of psychiatric phenomena. The reasons for this seeming incompatibility are not entirely clear, but perhaps it has something to do with differences in human temperament. As Samuel Taylor Coleridge long ago suggested, each of us is born either an Aristotelian or a Platonist—each of us, that is, inherently prefers either to limit knowledge to what is ascertainable by the five senses alone or to include within its compass the psychological insights derived from introspection, intuition, and inspiration. Thus, among psychiatrists, there are those who would restrict the understanding of psychiatric disease to their biological aspects alone and those who would include their intangible, but no less real, psychological components as well.

The inherent tendency toward misunderstanding and antagonism between the proponents of these separate views hardly gives logical grounds for concluding that the concepts and practical clinical procedures associated with each are mutually exclusive, or even incompatible. On the contrary, biological and psychological knowledge each have vital contributions to make to the understanding and treatment of psychiatric disease. To borrow Binger's metaphor, they are the two faces of psychiatry, and both are essential for the fullest development of psychiatry as a discipline.

That is not to say, however, that any one individual can become completely familiar with both of these basic pathways to knowledge. A mere glance at the overwhelming wealth of information contained in this new edition of CTP demonstrates the impossibility of that feat. Instead of straining after the unattainable, each of us must choose that portion of the whole that is of most interest and master that more limited domain. At the same time, we must move outside our own parochial perceptions and predilections to a genuine understanding and appreciation of the pursuits of those who are following different pathways from our own. It is a major virtue of the volumes in hand that they greatly facilitate that task. The various contributors to their pages have delineated aspects of one or the other of the two faces of psychiatry with a clarity and precision that will enable their readers to achieve a broader vision of the nature and treatment of psychiatric disorders. Therein lies the basis for the ultimate integration of our knowledge of body and mind.

SUGGESTED CROSS-REFERENCES

The neural sciences are discussed in Chapter 1, neuropsychiatry in Chapter 2, and biological therapies in Chapter 3. Psychoanalytic theories and other psychological theories are covered in Chapter 6, and psychological therapies are covered in Chapter 30. The history and future of psychiatry is discussed in Chapter 55.

CHAPTER REFERENCES

Azam E: *Hypnotisme at Double Conscience*. Félix Alcan, Paris. 1893.

Bailly J: Rapport des commissaires chargés par le roi de l'examen du magnétisme animal. Imprimerie Royale, Paris, 1784.

Bergson H: *Dreams*. Huebsch, New York, 1914.

Bertrand A: *Du Magnétisme Animal en France*. Baillière. Paris. 1826.

*Binger C: *The Two Faces of Medicine*. Norton. New York, 1967.

Coleridge S: *Table Talk*. Oxford University Press, London, 1917.

*Jackson S: *Care of the Psyche*. Yale University Press, New Haven, 1999.

James W: Frederic Myers's service to psychology. Proc Soc Psychical Res *17:13*, 1901–1903.

*James W: *The Varieties of Religious Experience*. Longmans, Green, New York, 1903.

Janet P: *L'Automatisme Psychologique*. Félix Alcan, Paris. 1889.

Mesmer F-A: *Le Magnétisme Animal* (R Amadou, editor). Payot, Paris, 1971.

*Myers W: *Human Personality*. Longmans, Green, London, 1903.

*Podmore F: *From Mesmer to Christian Science*. University Books, New Hyde Park, NY, 1963.

DRUGS USED IN PSYCHIATRY

This guide contains color reproductions of some commonly prescribed major psychotherapeutic drugs. This guide mainly illustrates tablets and capsules. A † symbol preceding the name of the drug indicates that other doses are available. Check directly with the manufacturer. (*Although the photos are intended as accurate reproductions of the drug, this guide should be used only as a quick identification aid.*)

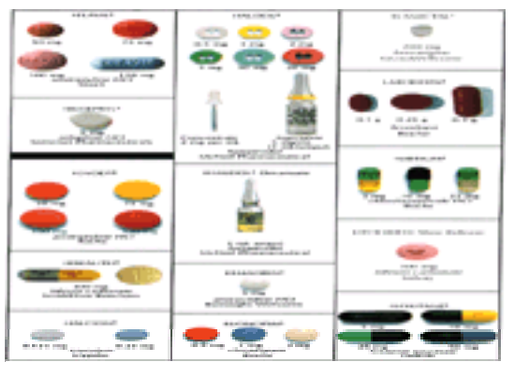

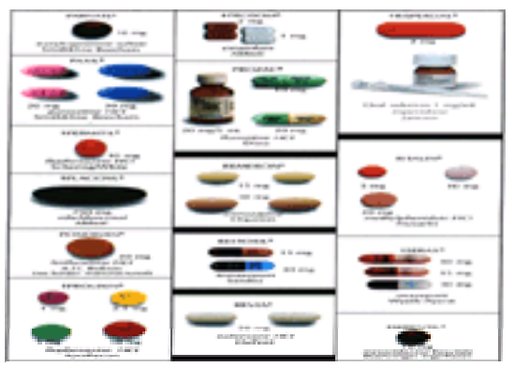

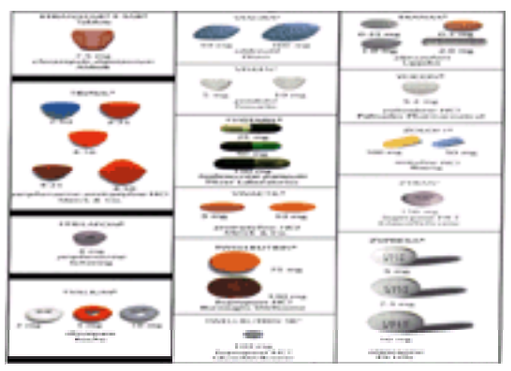

Textbook of Psychiatry

COLOR PLATES

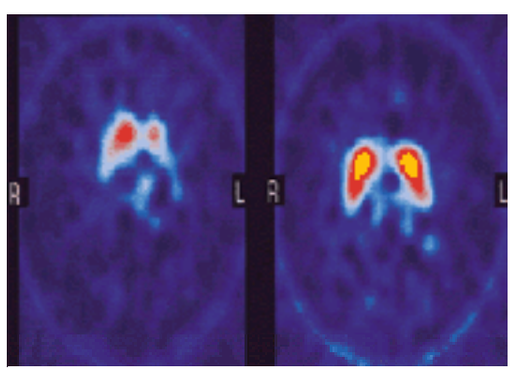

FIGURE 1.15–4 SPECT images of the distribution of [^{123}I]b-CIT (cocaine-iodo-tropane) in a healthy subject and a patient with Parkinson's disease. [^{123}I]b-CIT is a radiolabeled cocaine analogue and is a probe of dopamine transporters in the striatum. These transporters are located presynaptically on terminals of dopamine neurons projecting from the substantia nigra to the striatum. These transverse images show a high density of dopamine transporters in striatum and a marked reduction of these sites in an age- and sex-matched patient with idiopathic Parkinson's disease. The transporters are lost because the entire dopaminergic neuron, including its terminal projections in the striatum, degenerate in this disorder. (Courtesy of John Seibyl, Yale University.)

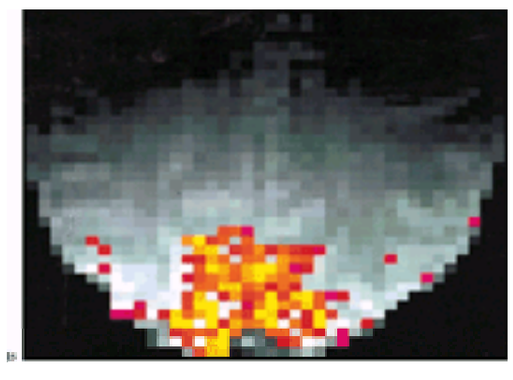

FIGURE 1.16–6 fMRI studies of photic stimulation. **A**, Change in image signal intensity in a single pixel during four alternating, 30-second epochs of flash photic stimulation. **B**, Activation maps demonstrating statistically significant increases in image signal intensity in primary visual cortex. Image data is acquired with a surface coil placed at the back of the head, thus explaining the decrease in image intensity with increasing distance from the occipital pole of the head. (Courtesy of Jonathan M. Levin, M.D., and Luis C. Maas.)

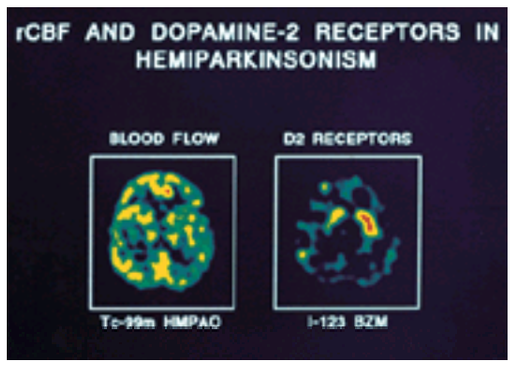

FIGURE 2.6–2 Regional cerebral blood flow, visualized with Technetium-99m HMPAO (**left panel**), and dopamine D_2-receptor binding, visualized with ^{123}I-iodobenzamide (I-123 BZM) (**right panel**), in a patient with hemiparkinsonism. Individuals with Parkinson's disease may present with markedly asymmetric symptoms. In this case, I-123 BZM uptake is markedly increased in the basal ganglia contralateral to the affected limb, whereas blood flow is unaffected. The increased uptake reflects increased D_2-receptor availability, either because of receptor upregulation or depletion of competing dopamine for the D_2 sites. (Courtesy of Michael Knable, D.O., and Daniel R. Weinberger, M.D.)

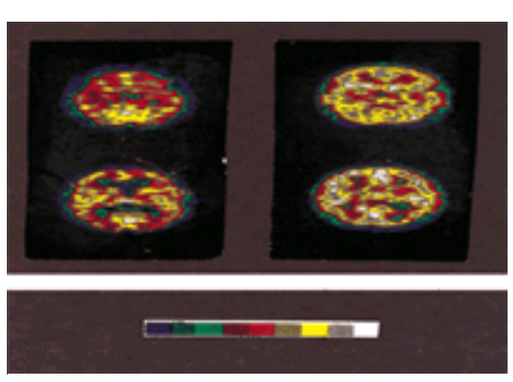

FIGURE 2.9–1 Transverse views obtained with Technetium-99m HMPAO SPECT. The two views on the left are of the brain of an adolescent with Lyme encephalopathy and demonstrate moderate heterogeneous hypoperfusion. The views on the right are of the brain of an adolescent without encephalopathy and demonstrate a normally perfused scan. The color spectrum scale, from purple to white, represents low to normal perfusion.

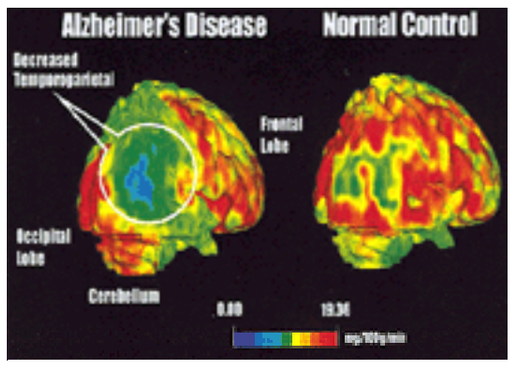

FIGURE 2.13–2 Three-dimensional PET FDG images demonstrate markedly lower glucose metabolism in temporoparietal region in patient with Alzheimer's disease than in normal control.

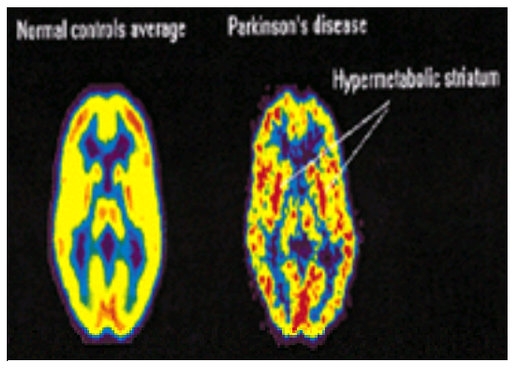

FIGURE 2.13–3 Transaxial section of PET FDG demonstrates hypermetabolism in striatum in patient with Parkinson's disease compared with normal control average image.

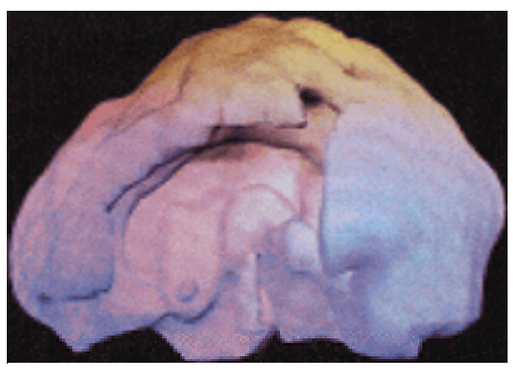

FIGURE 2.13–4 Stroke. Left side surface view shows marked decreased left frontal, temporal, and parietal lobe. (Note: SPECT cortical surface images are rendered by setting the threshold at 55%, looking at most active 45% of brain activity; SPECT active images are rendered by setting blue color threshold set at 55%, looking at average brain activity in the brain compared to red [or white] threshold set at 85%, looking at the most active 15% of brain activity.)

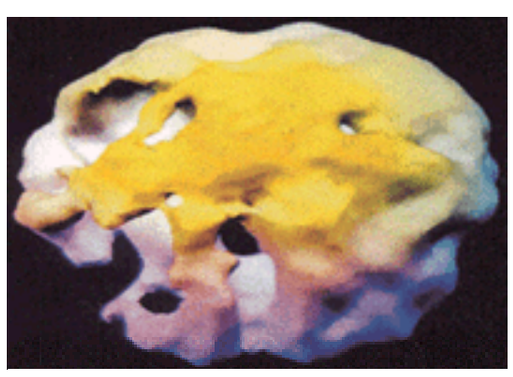

FIGURE 2.13–5 Two right hemisphere strokes with depression as presenting symptom. Top-down surface view shows marked decreased right frontal, temporal, and parietal lobe.

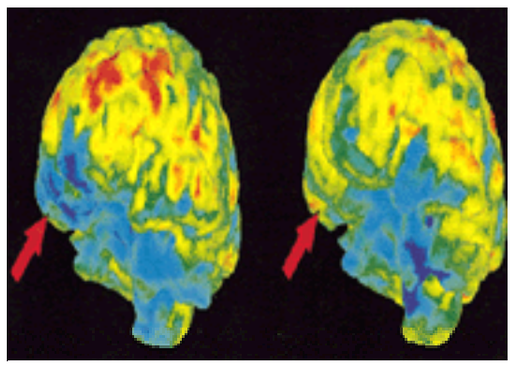

FIGURE 2.13–6 Three-dimensional PET FDG images demonstrate lower orbitofrontal metabolism in orbital frontal lobe syndrome in patient secondary to traumatic brain injury than in normal control.

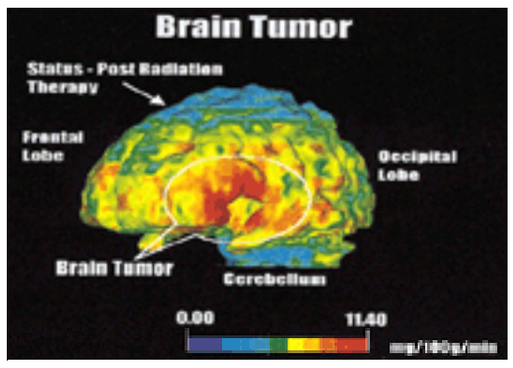

FIGURE 2.13–7 Three-dimensional PET FDG image demonstrate hypermetabolism in region of brain tumor in frontal cortex and decreased metabolism in parietal cortex secondary to postradiation injury.

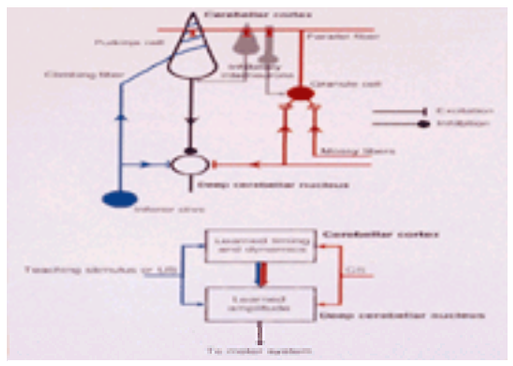

FIGURE 3.4–2 A schematic diagram of the circuitry of the mammalian cerebellum (**top**). In the classically conditioned blink response, input from the air-puff unconditioned stimulus and input from the auditory conditioned stimulus comes in through parallel pathways to the cerebellar cortex and to the deep cerebellar nucleus, and plasticity occurs in both pathways (**bottom**). (Reprinted with permission from Raymond JL, Lisberger SG, Mauk MD: The cerebellum: A neuronal learning machine? Science *272:*1126, 1996. ©1996 American Association for the Advancement of Science.)

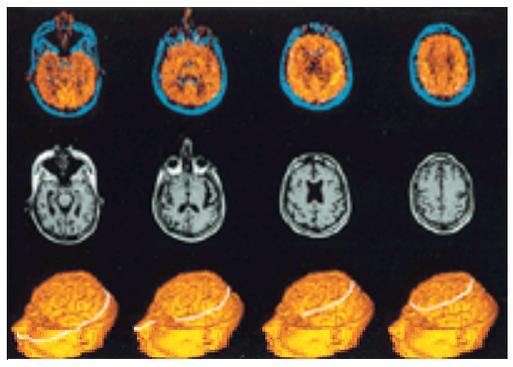

FIGURE 3.4–5 PET and MRI scans in a patient with Korsakoff's syndrome. Neural dysfunction was evident as reduced glucose utilization in multiple cortical regions in the frontal and parietal lobes, and in the cingulate. Functional neuroimaging can reveal brain dysfunction that might otherwise not be evident if limited to structural neuroimaging results. In Korsakoff's syndrome, the memory impairment probably reflects a disruption of thalamocortical circuitry. (Reprinted with permission from Paller KA, Acharya A, Richardson BC, Plaisant O, Shimamura AP, Reed BR, Jagust WJ: Functional neuroimaging of cortical dysfunction in alcoholic Korsakoff's syndrome. J Cogn Neurosci *9:*277, 1997.)

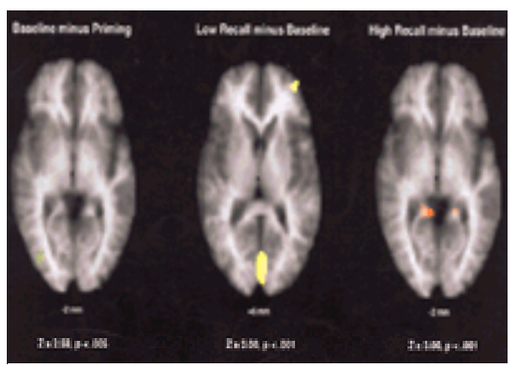

FIGURE 3.4–9 PET activations superimposed over averaged transverse MRI scans with the distance shown representing the distance from the line connecting the anterior and posterior commissure. Words were studied under strong or weak learning conditions (high recall or low recall) and then both declarative memory (cued recall) and nondeclarative memory (priming) were tested. The Baseline minus Priming subtraction showed an area of decreased blood flow (green) in right visual association cortex thought to be related to the greater ease of processing primed words. The Low Recall minus Baseline subtraction showed an area of increased blood flow (yellow) in secondary visual cortex and left prefrontal cortex thought to be related to the effort involved in deliberate, effortful retrieval. The High Recall minus Baseline subtraction showed a region of increased blood flow (red) in bilateral hippocampal regions thought to be related to successful retrieval of recently acquired information. (Reprinted with permission from Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS: Conscious recollection and the human hippocampal formation: Evidence from positron emission tomography. Proc Natl Acad Sci USA *93:*321, 1996. ©1996 National Academy of Sciences, U.S.A.)

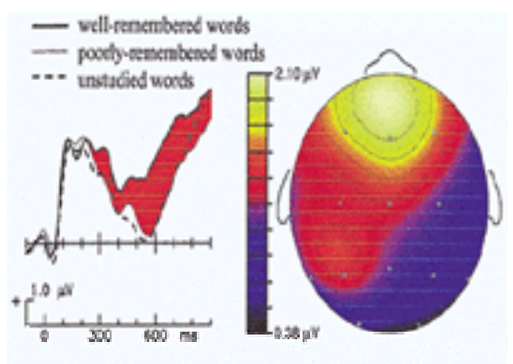

FIGURE 3.4–10 Brain potentials showing a differential response based on the extent to which subjects engaged in recollection following word presentations. Potentials shown at the left were recorded from a frontal scalp electrode. Measurements from multiple electrodes over the 400- to 800-ms latency range were used to

generate the topographical map at the right, showing that the neural correlate of recollection was broadly distributed across the scalp, with largest responses over frontal cortex. (Adapted with permission from Paller KA, Kutas M: Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. J Cogn Neurosci *4:*375, 1992.)

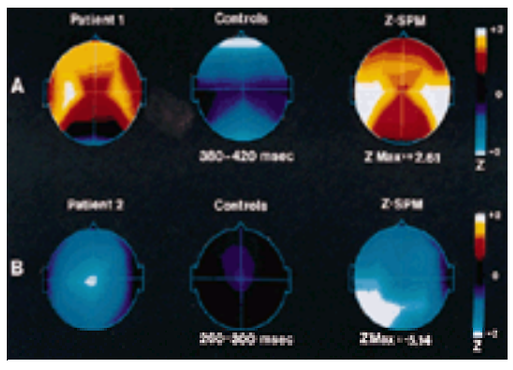

FIGURE 11.7–3 Illustrative cases of quantitative EEG abnormalities in hallucinogen persisting perception disorder (flashback). **Patient 1:** A 26-year-old computer programmer used LSD at the age of 18 on 15 occasions; 31 months later the patient experienced the abrupt onset of intense, LSD-like set of visual and affective disturbances lasting all night. At 25 he suffered the spontaneous onset of hourly flashing white lights centrally and black dots in his peripheral fields, which have continued for the past 10 years. Topographic brain maps are shown during the 380–420 msec epoch of the visually evoked potential in row **A**. The upper left map represents the subject's data. The upper middle map shows control subjects for the same poststimulus latency epoch. The right upper image is a significance probability map (SPM) showing *Z*-scores resulting from a comparison of the data from the left and middle upper maps. The patient shows an enhancement in the visually evoked signal involving both temporal regions of the cerebrum. **Patient 2:** A 23-year-old musician used LSD on 16 occasions over a 4-month period at the age of 20. Within 2 months he began to notice a progressive, continuous visual disorder characterized by flashes of color, persisting afterimages, haloes around objects, a grainy texture to the sky, and the lingering trails of objects as they passed through his visual field. The graininess in the visual field interfered with night vision. Topographic brain mapping is illustrated during the 260–300 msec epoch of auditory evoked potentials in row **B**. Note the region of reduced electrical activity in the left posterior temporal region in the lower right map. (Reprinted with permission from Abraham HD, Duffy FH: Stable quantitative EEG difference in post-LSD visual disorder by split-half analysis: Evidence for disinhibition. Psychiatry Res *67:*173, 1996.)

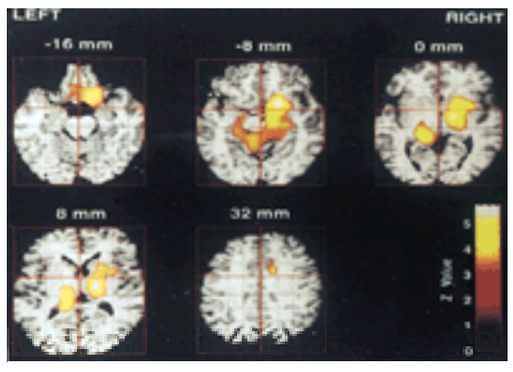

FIGURE 12.1–1 Axial sections demonstrating brain areas with significantly increased activity during auditory verbal hallucinations in the group study. Functional PET results (threshold at $Z > 3.09$, $P < 0.001$, by reference to the unit normal distribution) are displayed in color, superimposed upon a single structural T1-weighted magnetic resonance imaging (MRI) scan that has been transformed into the Talairach space for anatomical reference. Section numbers refer to the distance from the anterior commissure-posterior commissure line, with positive numbers being superior to the line. The areas of activation extend into the amygdala bilaterally, and into the right orbitofrontal cortex. Although these regions of extension are consistent with the limbic paralimbic component of activity during hallucinations, and may contribute to drive and affect in this context, definitive statements cannot be made in the absence of discrete maxima. (Reprinted with permission from Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, Seaward J, McKenna P, Chua SE, Schnorr L, et al: A functional neuroanatomy of hallucinations in schizophrenia. Nature *378:*1769, 1995.)

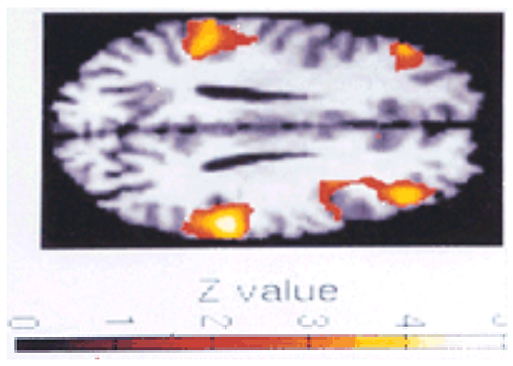

FIGURE 12.1–2 There is a significant difference in O_{15} activity in the prefrontal and parietal cortex during the performance of an auditory discrimination task in deficit and nondeficit patients, with deficit patients having decreased activity in these regions. (Courtesy of A. Lahti, Maryland Psychiatric Research Center, Baltimore, MD.)

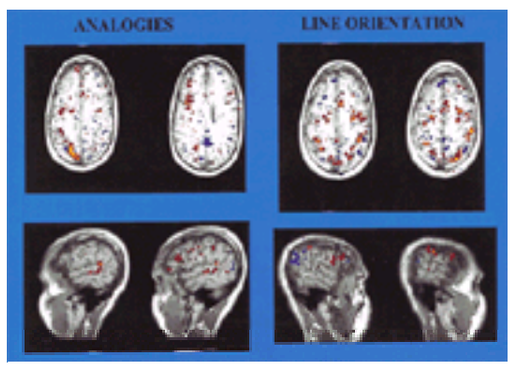

FIGURE 12.3–6 Illustration of functional imaging data obtained in healthy people: **A**, sex differences in local glucose metabolism; **B**, activation with verbal and spatial tasks as seen by functional MRI.

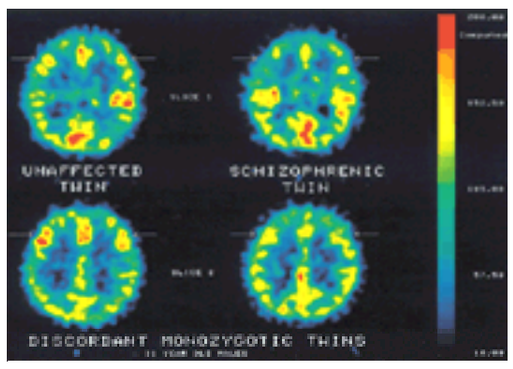

FIGURE 12.4–2 PET scans using H_2O_{15} of two monozygotic twins, one with (right) and one without (left) schizophrenia. Top and bottom scans show two levels through the dorsolateral prefrontal cortex. At the time of scanning, subjects are performing a cognitive task that typically requires prefrontal cortical function. The affected twin blood flow to the dorsolateral prefrontal cortex is markedly reduced compared to the unaffected twin. (Courtesy of R. Berman and D. Weinberger.)

FIGURE 25.8–2 Woman with lupus erythematosus malar rash. (Courtesy of M. Kevin O'Connor, M.D.)

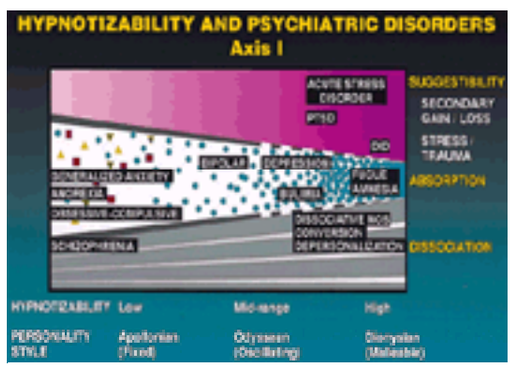

FIGURE 30.3–1 Model of the shift from normal to hypnotic attention. Hypnosis involves narrowing the focus of attention, with concomitant increases in dissociation of thoughts, perceptions, and feelings at the periphery and increased suggestibility.

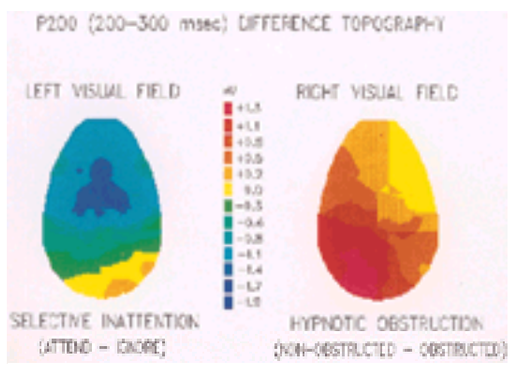

FIGURE 30.3–2 Brain electrical activity mapping of visual event-related potentials comparing the effects of selective inattention to a visual stimulus (attending to the other visual hemifield) and hypnotically hallucinating an obstruction to that stimulus. Selective inattention involves increased amplitude anteriorly, while hypnotic hallucination produces decreased amplitude in the occipital cortex.

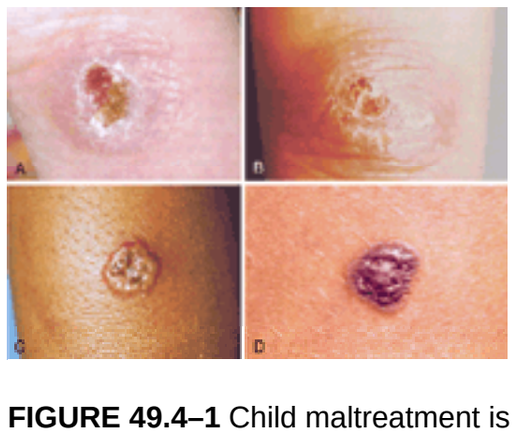

FIGURE 49.4–1 Child maltreatment is frequently manifested by bruises, burns, and other skin lesions. The diagnostic problem is that abusive injuries may be mistaken for medical conditions and vice versa. **A**, Physical abuse, a cigarette burn on the sole of the foot of a 5-month-old infant. **B**, A swimming pool granuloma overlying an interphalangeal joint. **C**, Tinea corporis. **D**, Clark's nevus, a benign, common acquired nevus. [Reprinted with permission from Reece RM: *Child Abuse: Medical Diagnosis and Management*. Lea & Febiger, Philadelphia, 1994 **(A)**; Caputo R: *Pediatric Dermatology and Dermatopathology*, vol 4. Williams & Wilkins, Baltimore, 1996 **(B)**; Caputo R: *Pediatric Dermatology and Dermatopathology*, vol 2. Lea & Febiger, Philadelphia, 1993 **(C)**; Caputo R: *Pediatric Dermatology and Dermatopathology*, Lea & Febiger, Philadelphia, 1990 **(D)**.]

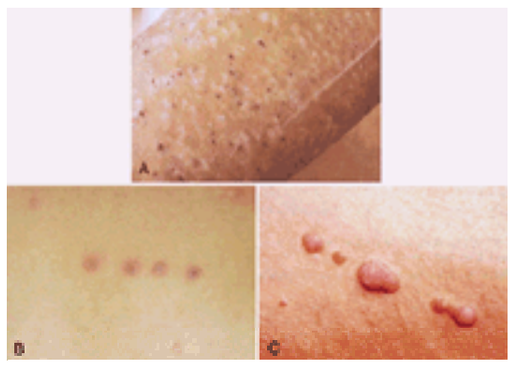

FIGURE 49.4–2 These linear skin lesions illustrate both child maltreatment and other conditions not related to abuse. **A**, Physical abuse, superficial ulcers caused by applying sandpaper to a child's skin. **B**, A series of insect bites distributed in a linear fashion, indicating that a single insect took bites in sequence. **C**, Common warts that are linear because lesions are induced when old lesions are scratched. [Reprinted with permission from Caputo R: *Pediatric Dermatology and Dermatopathology*, vol 2. Lea & Febiger, Philadelphia, 1993 **(A)**; Caputo R: *Pediatric Dermatology and Dermatopathology*, vol 4. Williams & Wilkins, Baltimore, 1996 (**B** and **C**).]

CHAPTER 1. NEURAL SCIENCES

1.1 NEURAL SCIENCES: INTRODUCTION AND OVERVIEW

JACK A. GREBB, M.D.

The first two chapters of this textbook—Neural Sciences and Neuropsychiatry and Behavioral Neurology—provide the reader a very complete primer and reference resource on the biological basis of normal behaviors and mental illnesses. Both chapters focus on the human brain, which is the biological substrate for all of our emotions, cognitive abilities, and behaviors—that is, everything we feel, think, and do. This chapter describes the basic biology of the brain and presents information on the interactive systems within the brain, the imaging of brain function, and the genetics of mental illnesses. Chapter 2, Neuropsychiatry and Behavioral Neurology, focuses on the neuropsychiatric aspects of the more classical neurological diseases, the study of which provides valuable information regarding how identified lesions in the nervous system are associated with disorders of affect, cognition, and behavior.

NEUROANATOMY

Functional neuroanatomy (discussed in Section 1.2) is the study of interacting and interdependent neurons, groups of neurons (e.g., nuclei), and brain regions. The three neural systems of most interest in psychiatry are the thalamocortical system, the basal ganglia, and the limbic system. In the 1970s, the thalamus was a focus of psychiatric research; however, the thalamus and its interactions with the cerebral cortex have primary importance in the sensory, motor, and associative functions of the brain. Moreover, the thalamus and its interactions with the cerebellar cortex are now known to be involved in cognition. The basal ganglia, once thought of only as part of the motor system, is now known to be a complex system within itself, and plays a key role in the cognitive functions of the brain. In contrast to the thalamus and basal ganglia, the limbic system has long been associated with psychiatric symptoms because of its clear involvement in the experience and expression of emotions.

The rapidly expanding understanding of the adult human brain in recent years has allowed more questions to be asked about the developing human brain (Section 1.3). Because psychiatric disorders that become symptomatic only in adulthood may be caused by genetic or environmental events that occur during conception or early in development, the study of human brain development has the potential to provide neurobehavioral scientists with a breakthrough in understanding mental illnesses.

NEUROTRANSMITTERS

There are four broad classes of neurotransmitter and neuromodulator substances in the brain: monoamines, amino acids, peptide neurotransmitters, and the much more recently discovered neurotrophins (also known as *neurotrophic factors*). Two additional neurotransmitters that do not fit into these four major classes are the gas nitric oxide and the purine-related neurotransmitters adenosine and adenosine triphosphate (ATP). Any single neuron can release multiple different types of neurotransmitters or neuromodulators, and also have receptors for multiple different receptor types and subtypes, thus making each individual neuron capable of exquisite integration and modulation of incoming and outgoing signals. There are five classic monoamine neurotransmitters, which are serotonin, the three catecholamines (epinephrine, norepinephrine, and dopamine), acetylcholine, and histamine (Section 1.4). The monoamine neurotransmitters, although present in only a small percentage of neurons localized in small nuclei of the brain, have enormous impact on total brain functioning because the diffuse projections of axons from these monoaminergic neurons can affect virtually every brain region. In contrast to the monoamine neurotransmitters, the amino acid neurotransmitters are widely distributed in the brain, and it is possible to conceptualize the brain as reflecting the balance between the excitatory amino acid glutamate, and the inhibitory amino acid g-aminobutyric acid (GABA) (Section 1.5). In contrast to the relatively small number of different monoamine and amino acid neurotransmitters, over one hundred different neuropeptide neurotransmitters have already been identified (Section 1.6). Because virtually all existing drugs for psychiatric conditions act through monoamine or amino acid neurotransmitter systems, the development of drugs that would have specific agonist or antagonist properties on neuropeptide systems offers great hopes for the development of new pharmacological treatments.

The first neurotrophic factor, nerve growth factor (NGF), was discovered by the Nobel Prize laureate, Rita Levi-Montalcini. The neurotrophic factors are polypeptides (i.e., proteins), and thus have the same basic structure as the neuropeptide neurotransmitters. Less is known about neurotrophic factors than about the more classic neurotransmitter substances, but neurotrophic factors can be conceptualized as differing from neuropeptide neurotransmitters in having longer-term regulatory functions (Section 1.7). The identification of additional neurotrophic factors and neurotrophic factor receptors provides researchers with many new hypotheses for disease etiologies and new hopes for disease interventions. Clinical trials of specific neurotrophic factors (e.g., brain-derived neurotrophic factor [BDNF], ciliary neurotrophic factor [CNTF]) in neurological diseases have already been conducted on neurodegenerative diseases such as amyotrophic lateral sclerosis.

Nitric Oxide Nitric oxide (NO) acts in the brain as both an intraneuronal second messenger and as a neurotransmitter. NO is formed from the amino acid arginine by the actions of NO synthase. NO and NOS synthase have been described in the brain. NO synthase is present in discrete regions of the brain, particularly the striatum, hypothalamus, basal forebrain, and cerebellum. The best understood pathway resulting in the generation of NO begins with activation of the *N*-methyl-D aspartate (NMDA) subtype of the glutamate receptor. Activation of the NMDA receptor results in the influx of calcium into the neuron, and the calcium-mediated activation of NO synthase, which generates NO from arginine. Intraneuronal NO then acts on the iron molecule contained in guanylyl cyclase and results in the formation of cyclic guanylyl monophosphate (cGMP), a potent second-messenger molecule. Because of its gaseous properties, NO can also diffuse to adjacent neurons, in which it can also result in the formation of cGMP. Unlike other neurotransmitters, NO is not stored in synaptic vesicles and is not necessarily released only on depolarization. Its receptors are iron and possibly other reactive metals, and not the conventional protein neurotransmitter receptors. Currently available data suggest that NO may be involved in some aspects of learning and memory. In addition, inhibitors of NO synthase may be effective in reducing ischemic damage after cerebrovascular events.

Adenosine and ATP Adenosine is a purine and ATP is synthesized from adenosine. P_1 receptors, which have a high affinity for adenosine, and P_2 receptors, which have a high affinity for ATP, have been found in the brain. The P_1 receptors are blocked by xanthines, such as caffeine and theophylline. Three subtypes of the P_1 receptor are the adenosine A_1, A_2 and A_3 receptors, all of which are G protein-linked receptors. Adenosine is concentrated in discrete regions of the brain and appears to have the general effect of inhibiting the release of most other neurotransmitters. This property has led to research activities to study adenosine analogues for use as anticonvulsants or sedatives. ATP is stored along with catecholamines in synaptic vesicles, and is released when the catecholamines are released. ATP preferentially acts on P_2 receptors, and data show that at least one function of ATP is the excitatory activation of sodium-potassium and calcium ion channels. A small body of evidence suggests that nucleotides and nucleosides may have trophic actions on glial and neuronal cells.

s-Receptors Another molecule that does not fit easily into the general scheme is the s-receptor, for which a natural ligand has not been identified, although in the past the sigma receptor was erroneously thought to be the primary receptor for phencyclidine (PCP, also known as “angel dust”). The previous association between s-receptors and PCP and the discovery that a number of antipsychotic drugs bind with high affinities to the s-receptors have caused researchers and pharmaceutical companies to develop sigma antagonists as potential antipsychotic compounds. So far, however, clinical trials with these compounds have not been successful.

SIGNAL TRANSDUCTION AND MOLECULAR GENETICS

The process of *chemical neurotransmission* refers strictly to the release of a neurotransmitter by a presynaptic neuron, the travel of that neurotransmitter across some space (e.g., the synaptic cleft), and the binding of that neurotransmitter to its specific receptor on a postsynaptic neuron (or an autoreceptor on a presynaptic neuron). The process of *signal transduction,* however, refers to the process by which an electrical signal (e.g., the action potential) in the presynaptic neuron is translated into a chemical signal (e.g., the release of a neurotransmitter), and the process by which the chemical signal (e.g., the interaction of a neurotransmitter and its receptor) is translated back into an electrical signal in the postsynaptic neuron. The study of intraneuronal signaling pathways and the regulation of neuronal ion channels

provides the basics for understanding signal transduction.

The initial step in intraneuronal signaling is most often the generation of second-messenger molecules (e.g., ATP) following the activation of a neurotransmitter receptor by its specific neurotransmitter (Section 1.8). In this schema, the first messenger is considered to be the neurotransmitter, peptide, or hormone that activated the receptor. The generation of the second messenger, however, can lead to a cascade of intraneuronal third, fourth, and more messenger molecules. One example of this cascade is the process of protein phosphorylation, which is a reversible, posttranslational modification of a protein. The deletion or the addition of one or more phosphate groups to a protein results in a change in the function of that protein. Thus, protein phosphorylation can serve as a type of molecular on-off switch for protein function. Protein phosphorylation more often modulates the function of a protein than it turns a specific function completely on or off. The identification of multiple biochemical steps, such as protein phosphorylation, offers researchers novel opportunities to identify pathophysiological processes as well as to develop therapeutic approaches.

Although psychiatrists and other mental health professionals may be more familiar with neurotransmitters and intraneuronal messengers than with ion channels, it is the balance between external and internal concentrations of ions that actually fuels the activities of the brain (Section 1.9). This balance is achieved by a wide array of ion channels, some of which are regulated by neurotransmitters and others by voltage gradients directly. Many of the drugs of interest in psychiatry act directly on ion channels. The benzodiazepines act on GABA type A ($GABA_A$) receptors that are chloride ion channels. Phencyclidine acts on the NMDA subtype of glutamate receptors that are calcium ion channels. Nicotine, the active ingredient in tobacco, acts on nicotinic acetylcholine receptors that are sodium and potassium ion channels. As with the neurotransmitter receptors, the delineation of ion channel subtypes and the modulation of ion channel function by processes such as protein phosphorylation are among the most active areas of neuropsychiatric and neuropharmacological research.

Driving the development of the brain as well as the daily maintenance and regulation of brain function is the process of genetic expression (Section 1.10). The basic process of genetics involves the transcription of deoxyribonucleic acid (DNA) into ribonucleic acid (RNA) and the translation of RNA into a protein. A complex system of regulation exists for transcription and translation, and the newly discovered molecules and pathways for this regulation are sites of investigation for discoveries in the etiology, pathophysiology, and treatment of mental disorders. Alterations in gene expression occur both during development and in adulthood and may be the bases for abnormal and normal development, and for abnormal and normal adaptation to stress. Psychiatric research in the twenty-first century is likely to investigate neurotrophic factors and molecular genetics.

INTEGRATION AND TIME

In addition to the central nervous system, the human body contains two other systems that have complex, internal communicative networks—the endocrine system and the immune system. Mostly because of the discoveries of the involved molecular signals, it is now known that these three systems are integrated with each other, which has given birth to the sciences of psychoneuroendocrinology (Section 1.11) and psychoneuroimmunology (Section 1.12). The interactions between the neuroendocrine and central nervous systems can most easily be seen in the psychiatric symptoms that can accompany some hormonal disorders (e.g., depression in Cushing's syndrome), and also in the identification of disorders of neuroendocrine regulation as potential markers for state or trait variables in psychiatric conditions.

In addition to the property of internal communication, another property shared by the neuroendocrine, immune, and central nervous systems is that they undergo regular changes with time. The study of these changes with time and disorders of time regulation are encompassed in the field of chronobiology (Section 1.13).

IMAGING BRAIN FUNCTION

Although X-rays and computed tomography (CT) can provide images of the skull, these are often of more interest to neurologists and neurosurgeons than to psychiatrists because psychiatry is more focused on the function of the brain itself. The first approach to be developed for measuring and imaging brain function was the field of applied electrophysiology, using the tools of electroencephalography (EEG) and evoked potentials (EPs) (Section 1.14). More recently, computerized approaches to these data have yielded more sophisticated analyses and presentations of information. The recent development of magnetoencephalography may expand the ability of this field to measure the activity of deeper brain structures.

Of most value in psychiatric research today for the visualization of brain function are the modalities of positron emission tomography (PET) and single photon emission computed tomography (SPECT), utilizing either radiotracer techniques (Section 1.15) or the differing magnetic resonance properties of the molecules in the brain (Section 1.16). These techniques can measure and visualize brain function during increasingly shorter time periods, allowing researchers to ask increasingly specific questions about specific brain regions and neural networks and their relationship to specific emotional, cognitive, and behavioral states and activities.

APPLIED GENETICS

The application of the techniques of population genetics to the study of mental illness provided some of the first objective data that mental illnesses were biological illnesses, thereby helping to destigmatize these human conditions. The use of population genetic methods in psychiatric research continues to yield important data and insights into mental illnesses (Section 1.17). The tools of population genetics have been more recently supplemented by the application of molecular neurobiological tools to this discipline, and the ability to study specific genetic linkages among individuals and groups of individuals (Section 1.18). The application of these techniques can lead to the identification of a specific gene or genes as causative agents for specific mental disorders. The example of the discovery of the gene for Huntington's disease is described in a subsequent chapter (Section 1.21).

SLEEP AND APPETITE

Sleep and appetite are just two examples of complex behaviors that are the observable results of the summations of neural processes. Sleep (Section 1.19) and appetite (Section 1.20) are regulated by specific brain regions and specific neurotransmitters, modulated by intraneuronal signals and ion channels, affected by the immune and neuroendocrine systems, altered with time, visualizable through brain-imaging techniques, subject to genetic regulation, and often affected by neurological or psychiatric disorders. Other complex brain activities such as perception, cognition, and memory are described elsewhere in this textbook.

SUGGESTED CROSS-REFERENCES

Neuropsychiatry and behavioral neurology are discussed in Chapter 2; the neuropsychological and psychiatric aspect of AIDS are discussed in Section 2.8; the neurochemical, viral, and immunological studies of schizophrenia are discussed in Section 12.4; the biochemical aspects of mood disorders are discussed in Section 14.3; biological therapies are discussed in Chapter 31; and Alzheimer's disease is discussed in Section 51.3. The future of psychiatry is discussed in Section 55.3.

SECTION REFERENCES

Barnard EA, Burnstock G, Webb TE: G protein-coupled receptors for ATP and other nucleosides: A new receptor family. Trends Pharmacol Sci *15:*67, 1994.

Erfurth A: Adenosine and neuropsychiatric disorders. Implications for treatment. CNS Drugs *2:*184, 1994.

*Iadecola C: Bright and dark sides of nitric oxide in ischemic brain injury. Trends Neurosci *20:*132, 1997.

Kandel ER: A new intellectual framework for psychiatry. Am J Psychiatry *155:*457, 1998.

*Moncada S, Higgs A: The L-arginine-nitric oxide pathway. N Engl J Med *329:*2002, 1993.

Neary JT, Rathbone MP, Cattabeni F, Abbracchio MP, Burnstock G: Trophic actions of extracellular nucleotides and nucleosides on glial and neuronal cells. Trends Neurosci *19:*13, 1996.

*Nelson RJ, Demas GE, Huang PL, Fishman MC, Dawson VL, Dawson TM, Snyder SH: Behavioral abnormalities in male mice lacking neuronal nitric oxide synthase. Nature *378:*383, 1995.

*Porka-Heiskanen T, Strecker RE, Thakkar M, Bjorkum AA, Greene RW, McCarley RW: Adenosine: A mediator of the sleep-inducing effects of prolonged wakefulness. Science *276:*1265, 1997.

*Shihabuddin LS, Ray J, Gage FH: Stem cell technology for basic science and clinical applications. Arch Neurol *56:*29, 1999.

*Vile RG, Sunassee K, Diaz RM: Strategies for achieving multiple layers of selectivity in gene therapy. Mol Med Today *4:*84, 1998.

*Weinberger DR: Cell biology of the hippocampal formation in schizophrenia. Biol Psychiatry *45:*395, 1999.

*Young GB, Pigott SE: Neurobiological basis of consciousness. Arch Neurol *56:*153, 1999.

*Zigmond MJ, Bloom FE, Landis SC, Roberts JL, Squire LR, editors: *Fundamental Neuroscience*. Academic Press, San Diego, 1999.

1.2 FUNCTIONAL NEUROANATOMY

DAVID A. LEWIS, M.D.

The normal affective, cognitive, and behavioral processes that are disturbed in different psychiatric disorders arise because of specific patterns of activation in networks of neurons that are distributed through the central nervous system. These patterns of activation are mediated by the connections among specific brain structures. Consequently, understanding the neurobiological bases for psychiatric disorders requires an appreciation of the major principles governing the functional organization of these structures and connections in the human brain. The thalamocortical, basal ganglia, and limbic systems are of particular relevance to neuropsychiatric disorders.

Those systems are formed by extensive and highly specific connections among certain anatomical structures, and the activation of those multiple connections gives rise to distinct behaviors, cognitive abilities, and emotional states. Thus, knowledge of the anatomical organization of the functional systems is crucial for the development and testing of hypotheses regarding the biological bases of the signs and symptoms of neuropsychiatric disorders.

PRINCIPLES OF BRAIN ORGANIZATION

Cells The human brain contains approximately 10^{11} nerve cells or *neurons*. In general, neurons are composed of four morphologically identified regions (Fig. 1.2-1): (1) the cell body or *soma*, which contains the nucleus and can be considered the metabolic center of the neuron; (2) the *dendrites,* processes that arise from the cell body, branch extensively, and serve as the major recipient zones of input from other neurons; (3) the *axon,* a single process that arises from a specialized portion of the cell body (the *axon hillock*) and conveys information to other neurons; and (4) the *axon terminals,* fine branches near the end of the axon that form contacts *(synapses)* generally with the dendrites or the cell bodies of other neurons, release neurotransmitters, and thereby provide a mechanism for interneuronal communication.

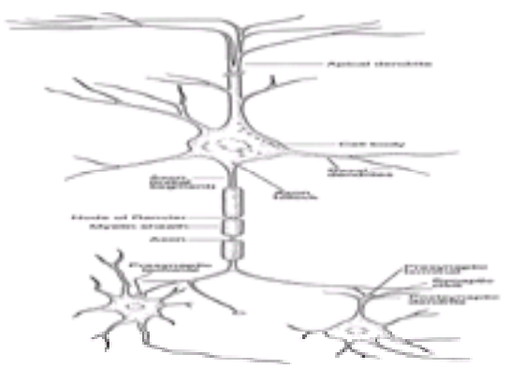

FIGURE 1.2-1 Drawing of the major features of a typical neuron. (Adapted from Kandel ER: Nerve cells and behavior. In *Principles of Neural Science,* ed 3, E Kandel, J Schwartz, T Jessell, editors. Elsevier, New York, 1991.)

The majority of neurons in the human brain are considered to be multipolar in that they give rise to a single axon and several dendritic processes. Although there are a number of classification schemes for neurons in different brain regions, almost all neurons can be considered either projection or local circuit neurons. *Projection neurons* have long axons and convey information from the periphery to the brain (sensory neurons), from one brain region to another, or from the brain to effector organs (motor neurons). In contrast, *local circuit neurons* or *interneurons* have short axons and process information within distinct regions of the brain.

Neurons can also be classified according to the neurotransmitters they contain (e.g., the dopamine neurons of the substantia nigra). Identification of neurons by their neurotransmitter content in anatomical studies provides a means for correlating the structure of a neuron with certain aspects of its function. However, neurotransmitters have defined effects on the activity of neurons, whereas complex brain functions, such as those disturbed in psychiatric disorders, are mediated by the coordinated activity of ensembles of neurons. Thus, the effects of neurotransmitters (or of pharmacological agents that mimic or antagonize the action of neurotransmitters) on behavioral, emotional, or cognitive states must be viewed within the context of the neural circuits that they influence.

In addition to neurons, the brain also contains several types of *glial* cells, which are at least ten times more numerous than the neurons. Although glial cells are not directly involved in information processing, they play several important roles in the nervous system. *Oligodendrocytes* and *Schwann cells*, found in the central and peripheral nervous systems, respectively, are relatively small cells that wrap their membranous processes around axons in a tight spiral. The resulting *myelin sheath* facilitates the conduction of action potentials along the axon. *Astrocytes,* the most numerous class of glial cells, appear to serve a number of functions, including participation in the formation of the blood-brain barrier, removal of certain neurotransmitters from the synaptic cleft, buffering of the extracellular potassium (K^+) concentration, and, given their close contact with both neurons and blood vessels, possibly a nutritive function as well. The third class of glial cells, the *microglia*, are actually derived from macrophages and function as scavengers, eliminating the debris resulting from neuronal death and injury.

Architecture Neurons and their processes form groupings in a number of different ways, and those patterns of organization can be evaluated by several approaches. The pattern of distribution of the neurons, called *cytoarchitecture,* is revealed by aniline dyes that stain ribonucleotides, Nissl substance, in the nucleus and the cytoplasm of neuronal cell bodies. The Nissl stains demonstrate the relative size and packing density of the neurons and consequently reveal, for example, the organization of the neurons into the different layers of the cerebral cortex. In certain pathological states, such as Alzheimer's disease, neuronal degeneration and loss results in striking changes in the cytoarchitecture of some brain regions (Fig. 1.2-2).

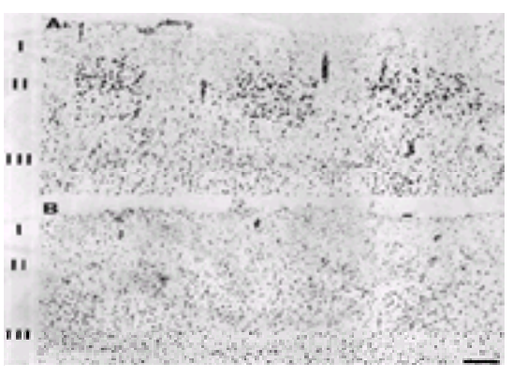

FIGURE 1.2-2 Nissl-stained sections of the superficial layers of the intermediate region of human entorhinal cortex. In the control brain **(A)**, layer II contains clusters or islands of large, intensely stained neurons. In Alzheimer's disease **(B)**, these layer II neurons are particularly vulnerable to degeneration, and their loss produces a

marked change in the cytoarchitecture of the region. Roman numerals indicate the location of the cortical layers. Calibration bar (200 μm) applies to **A** and **B**.

Other types of histological techniques, such as silver stains, selectively label the myelin coating of axons and, consequently, reveal the *myeloarchitecture* of the brain. For example, certain regions of the cerebral cortex—such as area MT, a portion of the temporal cortex involved in processing visual information—can be identified by a characteristic pattern of heavy myelination in the deep cortical layers. The progression of myelination is highly region-specific, may not be complete for years after birth, and may be a useful anatomical indicator of the functional maturation of brain regions.

Immunohistochemical and other related techniques—which identify the location of neurotransmitters, their synthetic enzymes, or other molecules within neurons—can be used to determine the *chemoarchitecture* of the brain (Fig. 1.2-3B). In some cases, these techniques reveal striking regional differences in the chemoarchitecture of the brain that are difficult to detect in cytoarchitecture.

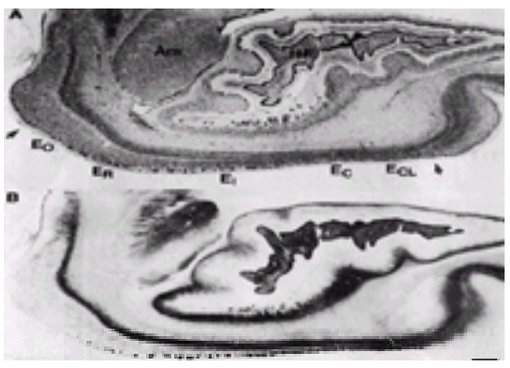

FIGURE 1.2-3 Adjacent sagittal sections through the medial temporal lobe of the human brain labeled to reveal the cytoarchitecture (**A**—Nissl stain) and chemoarchitecture (**B**—nonphosphorylated neurofilament protein immunoreactivity) of the entorhinal cortex. Arrows indicate the rostral (left) and caudal (right) borders of the entorhinal cortex, and letters indicate some of its subdivisions. Calibration bar (2 mm) applies to both panels. (Reprinted with permission from Beall MJ, Lewis DA: Heterogeneity of layer II neurons in human entorhinal cortex. J Comp Neurol *321:*241, 1992.)

Connections Every function of the human brain is a consequence of the activity of specific neural circuits. The circuits form as a result of several developmental processes. First, each neuron extends an axon, either after it has migrated to its final location or, in some cases, before. The growth of an axon along distinct pathways is guided by molecular cues from its environment and eventually leads to the formation of synapses with specific target neurons. Although the projection of axons is quite precise, some axons initially produce an excessive number of axon branches or *collaterals* and thus contact a broader set of targets than are present in the adult brain. During later development the connections of particular neurons are focused by the pruning or elimination of axonal projections to inappropriate targets.

Within the adult brain the connections among neurons or neural circuits follow several important principles of organization. First, many but not all connections between brain regions are *reciprocal;* that is, each region tends to receive input from those regions to which it sends axonal projections. In some cases the axons arising from one region may directly innervate the reciprocating projection neurons in another region; in other cases local circuit interneurons are interposed between the incoming axons and the projection neurons that furnish the reciprocal connections. For some projections the reciprocating connection is indirect, passing through one or more additional brain regions and synapses before innervating the initial brain region.

Second, many neuronal connections are either divergent or convergent in nature. A *divergent* system involves the conduction of information from one neuron or a discrete group of neurons to a much larger number of neurons that may be located in diverse portions of the brain. The locus ceruleus, a small group of norepinephrine-containing neurons in the brainstem that sends axonal projections to the entire cerebral cortex and other brain regions, is an example of a divergent system. In contrast, the output of multiple brain regions may be directed toward a single area, forming a convergent system.

Third, the connections among regions may be organized in a hierarchical or parallel fashion or both. For example, visual input is conveyed in a *serial* or *hierarchical* fashion through several populations of neurons in the retina to the lateral geniculate nucleus, to the primary visual cortex, and then progressively to the multiple visual association areas of the cerebral cortex. Within the hierarchical scheme, different types of visual information (e.g., motion, form) may be processed in a parallel fashion through different portions of the visual system.

Finally, regions of the brain are specialized for different functions. For example, lesions of the left inferior frontal gyrus (Broca's area, Fig. 1.2-4A) produce a characteristic impairment in speech production. However, speech is a complex faculty that depends not only on the integrity of Broca's area but also on the distributed processing of information across a number of brain regions through divergent and convergent, serial and parallel interconnections. Thus, the role of any particular brain region or group of neurons in the production of specific behaviors or in the pathophysiology of a given neuropsychiatric disorder cannot be viewed in isolation but must be considered within the context of the neural circuits connecting those neurons with other brain regions.

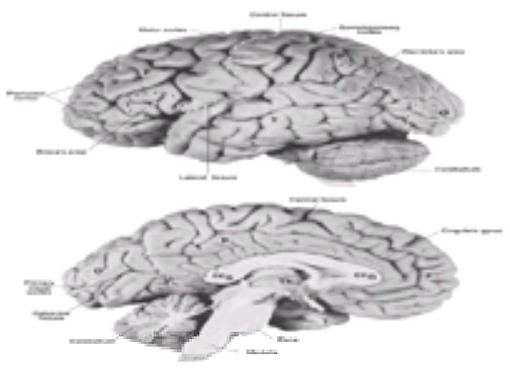

FIGURE 1.2-4 Photographs of the lateral **(top)** and medial **(bottom)** of the left hemisphere of a human brain indicating the location of major surface landmarks. F indicates frontal lobe, O indicates occipital lobe, P indicates parietal lobe, T indicates temporal lobe, Th indicates thalamus, cc_G indicates the genu of the corpus callosum, cc_S indicates the splenium of the corpus callosum.

Distinctiveness of the Human Brain The human brain is substantially larger than the brains of other primate species and certain areas of the human brain have expanded disproportionately. For example, the prefrontal cortex has been estimated to occupy only 3.5 percent of the total cortical volume in cats and 11.5 percent in monkeys but close to 30 percent of the much larger cortical volume of the human brain. Conversely, the relative representation of other regions is decreased in the human brain; for example, the primary visual cortex accounts for only 1.5 percent of the total area of the cerebral cortex in humans, but in monkeys a much greater proportion (17 percent) of the cerebral cortex is devoted to that region. Thus, the distinctiveness of the human brain is attributable both to its size and to the differential expansion of certain regions, particularly those areas of the cerebral cortex devoted to higher cognitive functions.

The expansion and the differentiation of the human brain is associated with substantial differences in the organization of certain elements of neural circuitry. For example, when compared with rodents, the dopaminergic innervation of the human cerebral cortex is much more widespread and regionally specific. The primary

motor cortex and certain posterior parietal regions receive a dense dopamine innervation in both monkeys and humans, but those areas receive little dopamine input in rats. These types of species differences indicate that there are limits to the accuracy of the generalizations made about human brain function when studies on rodents or even nonhuman primates are used as the basis for the inference. However, direct investigation of the organization of the human brain is obviously restricted and complicated by a number of factors. As indicated above, the expansion of the human brain is associated with the appearance of additional regions of the cerebral cortex. For example, the entorhinal cortex of the medial temporal lobe is sometimes considered to be a single cortical region, yet in the human brain the cytoarchitecture and the chemoarchitecture of that cortex differs substantially along its rostral-caudal extent (Fig. 1.2-3). It is tempting to identify those regions by their location relative to other structures, but sufficient interindividual variability exists in the human brain to make such a topological definition unreliable. In the case of the entorhinal cortex, the location of its different subdivisions relative to adjacent structures, such as the amygdala and the hippocampus, varies somewhat across human brains. Therefore in all studies, particularly those using the human brain, areas of interest must be defined such that investigators can accurately identify the same region in all cases.

An additional limitation to the study of the human brain concerns the changes in morphology and biochemistry that can occur during the interval between the time of death and the freezing or fixation of brain specimens. In addition to the influence of the known postmortem interval, such changes may begin to occur during the agonal state preceding death. When comparing aspects of the organization of the human brain with that of other species, the researcher must try to account for changes that may have occurred in the human brain as a result of postmortem delay or agonal state. Furthermore, in the study of disease states appropriate controls must be used because differences in neurotransmitter content or other characteristics among cases could be a result of factors other than the disease state. Studies of the human brain in vivo—using such imaging techniques as positron emission tomography, magnetic resonance imaging, and magnetic resonance spectroscopy—circumvent many of those problems but are limited by a level of resolution that is insufficient for the study of many aspects of human brain organization.

STRUCTURAL COMPONENTS

Major Brain Structures In the early stages of the development of the human brain, three primary vesicles can be identified in the neural tube: the *prosencephalon,* the *mesencephalon,* and the *rhombencephalon* (Table 1.2-1). Subsequently, the prosencephalon divides to become the *telencephalon* and the *diencephalon*. The telencephalon gives rise to the cerebral cortex, the hippocampal formation, the amygdala, and some components of the basal ganglia. The diencephalon becomes the thalamus, the hypothalamus, and several other related structures. The mesencephalon gives rise to the midbrain structures of the adult brain. The rhombencephalon divides into the *metencephalon* and the *myelencephalon*. The metencephalon gives rise to the pons and the cerebellum; the medulla is the derivative of the myelencephalon.

Primary Vesicles	Secondary Vesicles	Brain Components
Prosencephalon	Telencephalon	Cerebral cortex
		Hippocampus
		Amygdala
		Striatum
	Diencephalon	Thalamus
		Hypothalamus
		Epithalamus
Mesencephalon	Mesencephalon	Midbrain
Rhombencephalon	Metencephalon	Pons
		Cerebellum
	Myelencephalon	Medulla

Adapted from Noback J: *The Human Brain: An Introduction to Its Functional Anatomy*, ed 3. Mosby, St. Louis, 1991.

Table 1.2-1 Derivatives of the Neural Tube

The cerebral cortex of each hemisphere is divided into four major regions; the *frontal, parietal, temporal*, and *occipital* lobes (Fig. 1.2-4). The frontal lobe is located anterior to the central sulcus and consists of the primary motor, premotor, and prefrontal regions. The primary somatosensory cortex is located in the anterior parietal lobe; in addition, other cortical regions that are related to complex visual and somatosensory functions are located in the posterior parietal lobe. The superior portion of the temporal lobe contains the primary auditory cortex and other auditory regions; the inferior portion contains regions devoted to complex visual functions. In addition, some regions of the superior temporal sulcus receive a convergence of input from the visual, somatosensory, and auditory sensory areas. The occipital lobe consists of the primary visual cortex and other visual association areas.

Beneath the outer mantle of the cerebral cortex are a number of other major brain structures, such as the caudate nucleus, the putamen, and the globus pallidus (Fig. 1.2-5 and Fig. 1.2-6). Those structures are components of the basal ganglia, a system involved in the control of movement. The hippocampus and the amygdala, which are components of the limbic system, are located deep in the medial temporal lobe (Fig. 1.2-6 and Fig. 1.2-7). In addition, the derivatives of the diencephalon, such as the thalamus and the hypothalamus, are prominent internal structures; the thalamus is a relatively large structure composed of a number of nuclei that have distinct patterns of connectivity with the cerebral cortex (Fig. 1.2-6 and Fig. 1.2-7). In contrast, the hypothalamus is a much smaller structure that is involved in autonomic and endocrine functions.

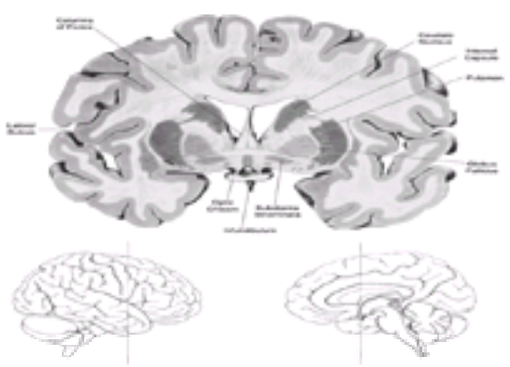

FIGURE 1.2-5 Drawing of a coronal section through the optic chasm of a human brain. The inset below indicates the level of the section. (Adapted from Nieuwenhuys R, Voogd J, van Huijzen C: *The Human Central Nervous System: A Synopsis and Atlas,* ed 3. Springer, New York, 1988.)

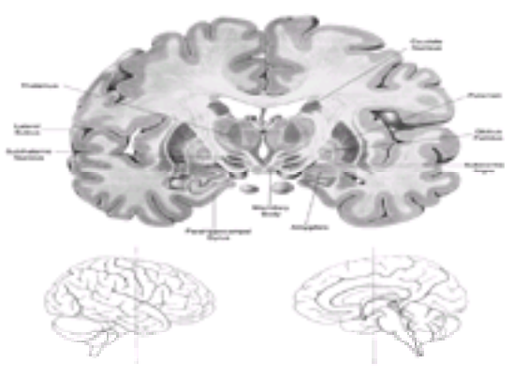

FIGURE 1.2-6 Drawing of a coronal section at the level of the mamillary bodies. The inset below indicates the level of the section. (Adapted from Nieuwenhuys R,

Voogd J, van Huijzen C: *The Human Central Nervous System: A Synopsis and Atlas,* ed 3. Springer, New York, 1988.)

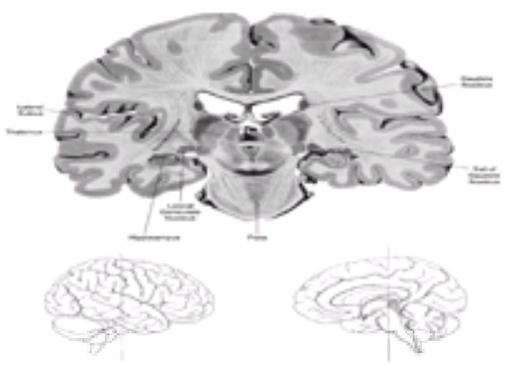

FIGURE 1.2-7 Drawing of a coronal section through the posterior thalamus. The inset below indicates the level of the section. (Adapted from Nieuwenhuys R, Voogd J, van Huijzen C: *The Human Central Nervous System: A Synopsis and Atlas*, ed 3. Springer, New York, 1988.)

Ventricular System As the neural tube fuses during development, the cavity of the neural tube becomes the ventricular system of the brain. It is composed of two C-shaped *lateral ventricles* in the cerebral hemispheres that can be further divided into five parts: the anterior horn (which is located in the frontal lobe), the body of the ventricle, the inferior or temporal horn in the temporal lobe, the posterior or occipital horn in the occipital lobe, and the atrium (Fig. 1.2-8). The *foramina of Monro* (interventricular foramina) are the two apertures that connect the two lateral ventricles with the *third ventricle,* which is found on the midline of the diencephalon. The *cerebral aqueduct* connects the third ventricle with the *fourth ventricle* in the pons and the medulla.

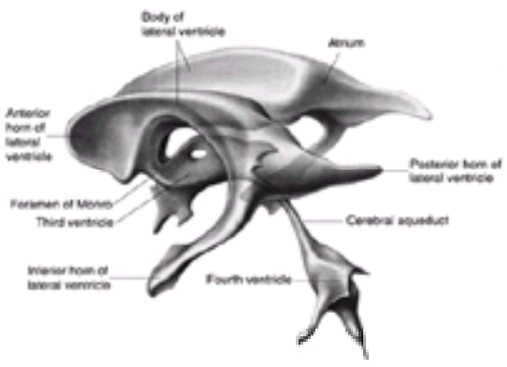

FIGURE 1.2-8 Drawing of a cast of the ventricular system of the human brain. The C shape of the lateral ventricles within the cerebral hemispheres is shown. (Adapted from Nieuwenhuys R, Voogd J, van Huijzen C: *The Human Central Nervous System: A Synopsis and Atlas,* ed 3. Springer, New York, 1988.)

The ventricular system is filled with cerebrospinal fluid (CSF), a colorless liquid containing low concentrations of protein, glucose, and potassium and relatively high concentrations of sodium and chloride. The majority (70 percent) of the CSF is produced at the choroid plexus located in the walls of the lateral ventricles and in the roof of both the third and fourth ventricles. The *choroid plexus* is a complex of ependyma, pia, and capillaries that invaginate the ventricle. In contrast to other parts of the brain, the capillaries in the choroid plexus are fenestrated, which allows substances to pass out of the capillaries and through the pia mater. The ependymal or choroid epithelial cells, however, have tight junctions between cells to prevent the leakage of substances into the CSF; that provides what is sometimes referred to as the *blood-CSF barrier*. In other parts of the brain, the endothelial cells of the capillaries exhibit tight junctions that prevent the movement of substances from the blood to the brain; that is referred to as the *blood-brain barrier*.

The CSF is produced constantly and circulates through the lateral ventricles to the third ventricle and then to the fourth ventricle. The CSF then flows through the medial and lateral apertures to the cisterna magna and pontine cistern and finally travels over the cerebral hemispheres to be absorbed by the arachnoid villi and released into the superior sagittal sinus. Disruptions in the flow of the CSF usually cause some form of hydrocephalus; for example, if an intraventricular foramen is occluded, the associated lateral ventricle becomes enlarged, but the remaining components of the ventricular system remain normal.

Several functions are attributed to the CSF: it serves to cushion the brain against trauma, to maintain and control the extracellular environment, and to spread endocrine hormones. Since the CSF bathes the brain and is in direct communication with extracellular fluid, it is possible to measure the amount of certain compounds in the CSF as a correlate of the amount of that substance in the brain. For example, concentrations of homovanillic acid (HVA), a metabolite of the neurotransmitter dopamine, are thought to reflect functional activity of that neurotransmitter. Thus, the concentration of HVA in samples of the CSF taken in a lumbar puncture may provide a picture of brain dopaminergic function. However, because the CSF bathes the entire brain, the concentrations of HVA in CSF may not be a valid indicator of the activity of dopamine neurons in any particular brain area. Consequently, caution must be exercised in interpreting the findings of investigations that rely on CSF measurements as indicators of neurotransmitter activity.

FUNCTIONAL BRAIN SYSTEMS

The relation between the organizational principles and the structural components of the human brain are illustrated in three functional systems—the thalamocortical, basal ganglia, and limbic systems.

Thalamocortical Systems

Thalamus The largest portion of the diencephalon consists of the *thalamus,* a group of nuclei that serve as the major synaptic relay station for the information reaching the cerebral cortex. On an anatomical basis the thalamic nuclei can be divided into six groups: anterior, medial, lateral, reticular, intralaminar, and midline nuclei. A thin Y-shaped sheet of myelinated fibers, the *internal medullary lamina,* delimits the anterior, medial, and lateral groups of nuclei (Fig. 1.2-9). In the human thalamus the anterior and medial groups each contain a single large nucleus, the *anterior* and *medial dorsal nuclei*. The lateral group of nuclei can be further subdivided into dorsal and ventral tiers. The dorsal tier is composed of the *lateral dorsal,* the *lateral posterior,* and the *pulvinar nuclei;* the ventral tier consists of the *ventral anterior,* the *ventral lateral,* the *ventral posterior lateral,* and the *ventral posterior medial nuclei*. The lateral group of nuclei are covered by the *external medullary lamina,* another sheet of myelinated fibers. Interposed between those fibers and the internal capsule is a thin group of neurons forming the *reticular nucleus* of the thalamus. The *intralaminar nuclei,* the largest of which is the *central median nucleus,* are located within the internal medullary lamina. The final group of thalamic nuclei, the *midline nuclei*, cover portions of the medial surface of the thalamus. The midline nuclei of each hemisphere may fuse to form the interthalamic adhesion, which is variably present.

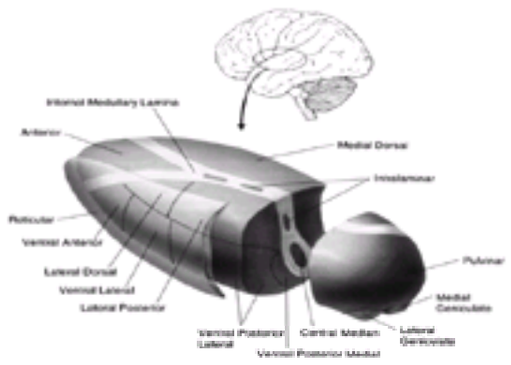

FIGURE 1.2-9 Drawing of the nuclei of the thalamus as seen on the left side of the brain. (Adapted with permission from Kelly JP: The neural basis of perception and movement. In *Principles of Neural Science,* ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Elsevier, New York, 1991.)

Thalamic nuclei can also be classified into several groups based on the pattern and information content of their connections (Table 1.2-2). For example, *relay nuclei* project to and receive input from specific regions of the cerebral cortex. Those reciprocal connections apparently allow the cerebral cortex to modulate the thalamic input it receives. *Specific relay nuclei* process input either from a single sensory modality or from a distinct part of the motor system. For example, the lateral geniculate nucleus receives visual input from the optic tract and projects to the primary visual area of the occipital cortex. As summarized in Figure 1.2-10, neurons of the thalamic relay nuclei furnish topographically organized projections to specific regions of the cerebral cortex, although some cortical regions receive input from more than one nucleus.

Type	Nuclei	Principal Afferent Inputs	Major Projection Sites
Specific relay	Anterior	Mamillary body of hypothalamus	Cingulate cortex
	Ventral anterior	Globus pallidus	Premotor cortex
	Ventral lateral	Dentate nucleus of cerebellum	Motor, premotor cortices
	Ventral posterior lateral	Medial lemniscal and spinothalamic pathways	Somatosensory cortex
	Ventral posterior medial	Sensory nuclei of trigeminal nerve	Somatosensory cortex
	Medial geniculate	Inferior colliculus	Auditory cortex
	Lateral geniculate	Optic tract	Visual cortex
Association relay	Lateral dorsal	Unknown	Cingulate cortex
	Lateral posterior	Superior colliculus	Parietal cortex
	Pulvinar	Superior colliculus	Temporal, parietal, occipital cortices
	Medial dorsal	Amygdala and hypothalamus	Prefrontal cortex
Diffuse projection	Midline	Reticular formation, hypothalamus	Basal forebrain, cortex
	Intralaminar	Reticular formation, spinothalamic tract, globus pallidus	Basal ganglia, cortex
	Reticular	Cerebral cortex, thalamus	Thalamus

* This table does not include the cortical inputs to each thalamic nucleus.
Adapted from Kelly JP: The neural basis of perception and movement. In Principles of Neural Science, ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Elsevier, New York, 1991.

Table 1.2-2 Connections of Thalamic Nuclei*

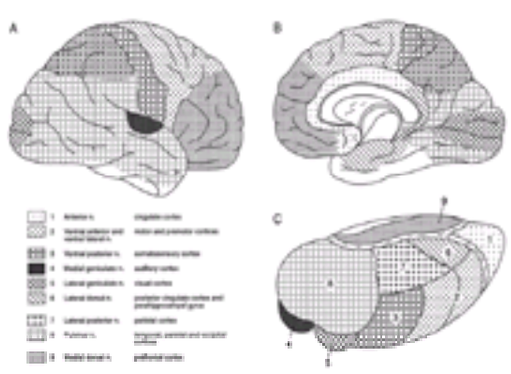

FIGURE 1.2-10 Schematic drawings of the lateral **(A)** and medial **(B)** surfaces of the right cerebral hemisphere and the right thalamus **(C)**. Shading patterns depict the cortical projection zones of some thalamic relay nuclei. (Adapted from Burt AM: *Textbook of Neuroanatomy*. Saunders, Philadelphia, 1993.)

In contrast, *association relay nuclei* receive highly processed input from more than one source and project to larger areas of the association cortex. or example, the medial dorsal thalamic nucleus receives input from the hypothalamus and the amygdala and is reciprocally interconnected with the prefrontal cortex and certain premotor and temporal cortical regions. In contrast to relay nuclei, *diffuse-projection nuclei* receive input from diverse sources and project to widespread areas of the cerebral cortex and to the thalamus. The divergent nature of the cortical connections of those nuclei indicates that they may be involved in regulating the level of cortical excitability and arousal. Finally, the reticular nucleus is somewhat unique in that it receives input from collaterals of the axons that reciprocally connect other thalamic nuclei and the cerebral cortex. Each portion of the reticular nucleus then projects to that thalamic nucleus from which it receives input. The pattern of connectivity may indicate that the reticular nucleus samples both cortical afferent and efferent activity and then uses that information to regulate thalamic function.

Cerebral Cortex The *cerebral cortex* is a laminated sheet of neurons, several millimeters thick, that covers the cerebral hemispheres. More that 90 percent of the total cortical area consists of the *neocortex,* which has a six-layered structure (at least at some point during development). The remainder of the cerebral cortex is referred to as the *allocortex* and consists of the *paleocortex* and the *archicortex*, regions that are restricted to the base of the telencephalon and the hippocampal formation, respectively.

Within the neocortex, the two major neuronal cell types are the pyramidal and stellate or nonpyramidal neurons. *Pyramidal neurons*, which account for approximately 60 percent of all neocortical neurons, usually have a characteristically shaped cell body that gives rise to a single apical dendrite that ascends vertically toward the cortical surface. In addition, the neurons have an array of short dendrites that spread laterally from the base of the cell. The dendrites of pyramidal neurons are coated with short protrusions, called *spines,* that are the sites of most of the excitatory inputs to these neurons. Most pyramidal cells are projection neurons that are thought to use excitatory amino acids as neurotransmitters. In contrast, *nonpyramidal cells* are generally small, local-circuit neurons, many of which use the inhibitory neurotransmitter g-aminobutyric acid (GABA). These neurons can be divided into distinct functional subclasses based on their biochemical and morphological features (Fig. 1.2-11). For example, the chandelier class of GABA neurons contains the calcium-binding protein parvalbumin, and exerts powerful inhibitory control over pyramidal neurons through synaptic inputs to the axon initial segment of pyramidal cells. In contrast, double bouquet cells contain the calcium-binding protein calbindin and provide inhibitory synapses to the dendritic shafts of pyramidal neurons, as well as to other local circuit neurons.

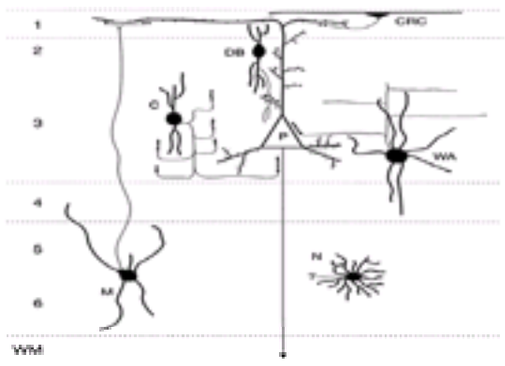

FIGURE 1.2-11 Schematic drawings of different morphological subclasses of GABA-containing local circuit neurons in the primate prefrontal cortex. As illustrated, the axons of some subclasses of GABA neurons selectively target different portions of pyramidal neurons (P). C indicates chandelier neuron, CRC indicates Cajal-Retzius cell, DB indicates double-bouquet cell, M indicates Martinotti cell, N indicates neurogliaform neuron. (Adapted with permission from Condé F, Lund JS, Jacobowitz DM, Baimbridge KG, Lewis DA: Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: Distribution and morphology. J Comp Neurol *341:*95, 1994.)

Neocortical neurons are distributed across six layers of the neocortex; those layers are distinguished by the relative size and packing density of their neurons (Fig. 1.2-12 and Fig. 1.2-13). Each cortical layer tends to receive particular types of inputs and to furnish characteristic projections. For example, afferents from thalamic relay nuclei terminate primarily in deep layer III and layer IV, whereas corticothalamic projections originate mainly from layer VI pyramidal neurons. These laminar distinctions provide important clues for dissecting possible pathophysiological mechanisms in psychiatric disorders. For example, reports of decreased somal size and diminished spine density on deep-layer pyramidal neurons in the prefrontal cortex of schizophrenic subjects suggests that these changes may be related to abnormalities in afferent projections from the medial dorsal thalamic nucleus. Consistent with this interpretation, the number of neurons in the medial dorsal nucleus has been reported to be decreased in schizophrenia.

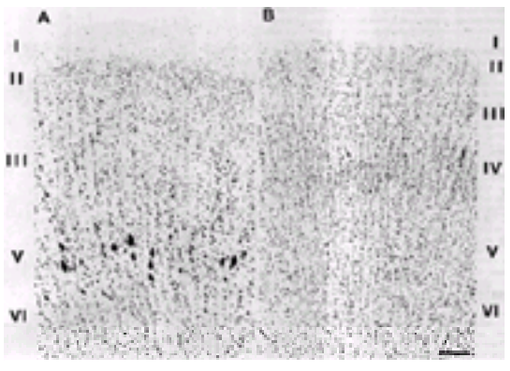

FIGURE 1.2-12 Nissl-stained sections of **(A)** Brodmann's area 4 (primary motor cortex) and **(B)** area 41 (primary auditory cortex) from a control human brain. Roman numerals indicate the cortical layers. Marked differences in neuronal size and packing density across the layers of the two regions are evident. Calibration bar (200 μm) applies to **A** and **B**.

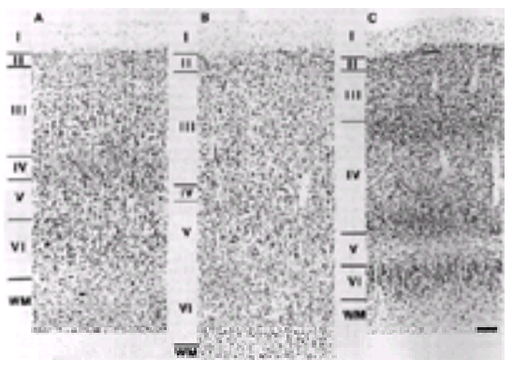

FIGURE 1.2-13 Nissl-stained sections of **(A)** Brodmann's area 46 (dorsolateral prefrontal cortex), **(B)** area 9 (dorsomedial prefrontal cortex) and **(C)** area 17 (primary visual cortex) from a control human brain. Roman numerals indicate the cortical layers. Note the marked differences in laminar organization between the prefrontal and visual areas, but the more subtle differences between the two prefrontal regions. Calibration bar (200 µm) applies to **A-C**.

In addition to the horizontal laminar structure, many aspects of cortical organization have a vertical or columnar characteristic. For example, the apical dendrites of pyramidal neurons and the axons of some local-circuit neurons have a prominent vertical orientation, indicating that those neural elements may sample the input to or regulate the function of neurons in multiple layers, respectively. Afferent inputs to the neocortex from other cortical regions also tend to be distributed across cortical layers in a columnar fashion. Finally, physiological studies in the somatosensory and visual cortices have shown that neurons in a given column respond to stimuli with particular characteristics, whereas those in adjacent columns respond to stimuli with different features.

The neocortex can be divided into two general types of regions. Regions with a readily identifiable six-layer appearance are known as the *homotypical cortex* and are found in association regions of the frontal, temporal, and parietal lobes. In contrast, some regions of the neocortex do not retain a six-layer appearance. Those regions, called the *heterotypical cortex*, include the primary motor cortex, which lacks a defined layer IV, and primary sensory regions, which exhibit an expanded layer IV. The neocortex can be further divided into discrete areas, each area having a distinctive architecture, a certain set of connections, and a role in particular brain functions. Most subdivisions of the human neocortex have been based on cytoarchitectural features; that is, subdivisions differ in the size, packing density, and arrangement of neurons across layers (Fig. 1.2-12 and Fig. 1.2-13). The most widely used system is that of Korbinian Brodmann (Fig. 1.2-14), who divided the cortex of each hemisphere into 44 numbered areas. Some of these numbered regions correspond closely to functionally distinct areas, such as area 4 (primary motor cortex in the precentral gyrus) and area 17 (primary visual cortex in the occipital lobe). In contrast, other Brodmann's areas appear to encompass several cortical zones that differ in their functional attributes. Although Brodmann's brain map has been extensively used in postmortem studies of psychiatric disorders, many of the distinctions among regions are quite subtle (Fig. 1.2-13A and Fig. 1.2-13B), and the locations of the boundaries between regions may vary among persons.

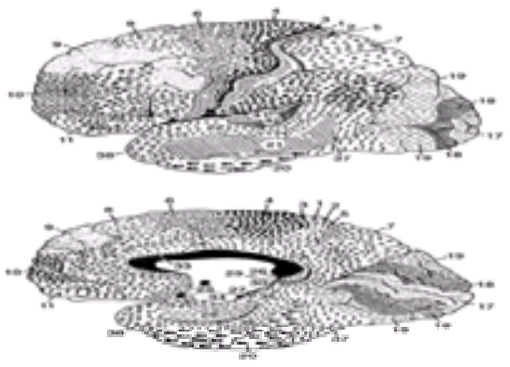

FIGURE 1.2-14 Drawing of the cytoarchitectonic subdivisions of the human brain as determined by Brodmann. **Top,** lateral view; **bottom,** medial view.

Although a given cortical area may receive other inputs, it is heavily innervated by projections from particular thalamic nuclei and from certain other cortical regions either in the same hemisphere *(association fibers)* or the opposite hemisphere *(commissural fibers)*. The patterns of connectivity make it possible to classify cortical regions into different types. *Primary sensory areas* are dominated by inputs from specific thalamic relay nuclei and are characterized by a topographic representation of visual space, the body surface, or the range of audible frequencies on the cortical surface of the primary visual, primary somatosensory, and primary auditory cortices, respectively. Those regions project in turn to nearby *unimodal association regions,* which are also devoted to processing information from a particular sensory modality. Output from those regions converges in *multimodal association areas,* such as the prefrontal cortex or the temporoparietal cortical regions. Neurons in those regions respond to complex stimuli and are thought to be mediators of higher cognitive functions. Finally, those regions influence the activity of the *motor areas* of the cerebral cortex that control behavioral responses.

Although this classification scheme of cortical regions is accurate in many respects, it fails to account for some of the known complexities of cortical information processing. For example, somatosensory input from the thalamus projects to several distinct, topographically organized maps in the cerebral cortex. In addition, information flow within the cortex is not confined to the serial processing route implied in the classification scheme but also involves parallel processing streams, such as sensory input from the thalamus to both the primary and association areas.

Although this discussion has not distinguished between the cerebral hemispheres, certain brain functions, such as language, are localized to one hemisphere (Fig. 1.2-15). The structural bases for the lateralization of function have not been determined, but some anatomical differences between the cerebral hemispheres have been observed. For example, a portion of the superior temporal cortex, called the *planum temporale,* is generally larger in the left hemisphere than in the right hemisphere. That cortical area, which is located close to the primary auditory cortex and includes the region called *Wernicke's area* (Fig. 1.2-4A), appears to be involved in receptive language functions that are localized to the left hemisphere. In addition, Brodmann's area 44, in the left inferior frontal cortex contains larger pyramidal neurons than the homotopic region of the right hemisphere, a difference that may contribute to the specialization of Broca's area for motor speech function.

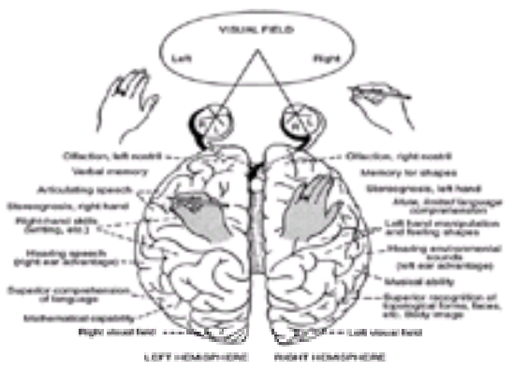

FIGURE 1.2-15 Drawing of the dorsal surface of the human brain showing the tendency for certain functions to be preferentially localized to one hemisphere. However, it is important to note that the intact brain may not be as lateralized as some studies (e.g., of patients with commissurotomies) suggest, that the degree of lateralization differs across individuals, and that in the intact brain it is rare that one hemisphere can mediate a function that the other hemisphere is completely unable to perform. (Reprinted with permission from Fuchs AF, Phillips JO: Association cortex. In *Textbook of Physiology,* HD Patton, AF Fuchs, B Hillie, AM Scher, R Steiner, editors, vol 1, ed 21. Saunders, Philadelphia, 1989.)

Functional Circuitry The connections between the thalamus, the cortex, and certain related brain structures comprise three thalamocortical systems, each with different patterns of functional circuitry. Those three systems—sensory, motor, and association systems—are described separately here but are heavily interconnected.

THALAMOCORTICAL SENSORY SYSTEMS Several general principles govern the organization of the thalamocortical sensory systems. First, sensory receptors transduce certain stimuli in the external environment to neural impulses. The impulses ascend, often through intermediate nuclei in the spinal cord and the medulla, and ultimately synapse in specific relay nuclei of the thalamus.

Second, projections from peripheral sensory receptors to the thalamus and the cortex exhibit topography; that is, a particular portion of the external world is mapped onto a particular region of the brain. For example, in the somatosensory system, axons carrying information regarding a distinct part of the body synapse in a discrete part of the ventral posterior nucleus of the thalamus. Specifically, the ventral posterior medial nucleus receives inputs regarding the head, and the ventral posterior lateral nucleus receives inputs regarding the remainder of the body. The nuclei then project topographically to the primary somatosensory cortex, where several representations of the contralateral half of the body can be found. Those representations are distorted; regions heavily innervated by sensory receptors, such as the fingers, are disproportionately represented in the primary somatosensory cortex.

Third, in some cases, sensory inputs travel to the thalamus in a segregated manner according to the submodality of the information conveyed. The inputs are then processed in a parallel fashion; particular pathways may be exclusively devoted to processing a submodality. An example of such segregation is evident in the somatosensory system, where most fibers carrying tactile and proprioceptive information travel in the medial lemniscus, but pain and temperature information is conveyed to the ventral posterior thalamic nuclei through the spinothalamic tract. Although some tactile information is carried in the spinothalamic tract, the submodalities of pain and temperature are largely segregated from tactile and proprioceptive inputs as they ascend to the thalamus.

Finally, sensory pathways exhibit convergence; that is, primary sensory areas process sensory information and then project to unimodal association areas. Subsequently, the unimodal areas project to and converge in multimodal association areas. An illustration of convergence in sensory pathways is found in the somatosensory system. The primary somatosensory cortex, located in the anterior parietal lobe (Fig. 1.2-16), has been divided into four regions on the basis of cytoarchitecture. Each of the cytoarchitectonic regions—numbered 1, 2, 3a, and 3b by Brodmann—contains a topographical representation of the body. The regions are heavily interconnected, and all project to the next level of somatosensory processing in area S-II. That type of projection, from one level of processing to a more advanced level, is termed a *feedforward projection*. The reciprocal connection, from the more advanced processing level back to the simpler level, is called a *feedback projection*. Both projections have distinct patterns of laminar termination: feedforward projections originate in the superficial layers of cortex (layer III) and terminate in layer IV; feedback projections originate in layers III, V, and VI and terminate outside layer IV. Further processing of somatosensory information occurs in higher-order somatosensory areas, such as area 7b of the posterior parietal cortex, which receive feedforward projections from S-II. Lesions of the posterior parietal cortex reflect the complexity of the information processed there; after a person has sustained a posterior parietal lesion, the ability to understand the significance of sensory stimuli is impaired, and extreme cases result in contralateral sensory neglect and inattention. However, the processing of somatosensory information within

the cortex is clearly much more complex than what has been described here (Fig. 1.2-17).

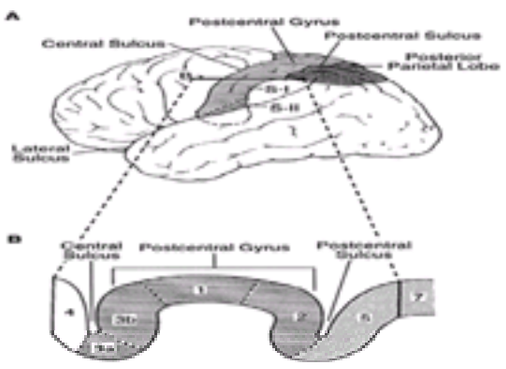

FIGURE 1.2-16 Drawing of the location of the somatosensory cortices in the human brain. **A:** Somatosensory cortices are located in the anterior and posterior parietal cortex. **B:** Primary somatosensory cortex (S-I) is divided into four cytoarchitectonic regions, as shown on the drawing of the section taken at the level depicted in **(A)**. (Reprinted with permission from Martin JH, Jessell TM: Anatomy of the somatic sensory system. In *Principles of Neural Science,* ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Elsevier, New York, 1991.)

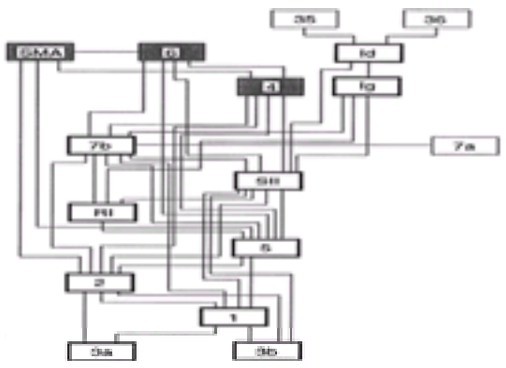

FIGURE 1.2-17 A proposed organizational scheme of the connectivity among cortical areas involved in somatosensory information processing. Hierarchical assignments were made on the basis of feedforward and feedback patterns of connections, as described in the text. (Reprinted with permission from Felleman DJ, Van Essen DC: Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex *1:*36, 1991.)

THALAMOCORTICAL MOTOR SYSTEMS The thalamocortical motor systems exhibit some unique organizational principles but also share many of the features present in the sensory systems. First, in contrast to sensory systems, which primarily ascend from sensory receptor to cortical association areas, motor systems descend from association and motor regions of the cortex to the brainstem and the spinal cord. For example, the corticospinal tract originates in layer V neurons of the premotor and primary motor cortices of the frontal lobe and terminates in the spinal cord to influence motor behavior.

Second, motor systems exhibit strong topography at both the thalamic and the cortical levels. For example, the corticospinal tract is organized so that a topographical representation of the contralateral half of the body is evident in the primary motor and premotor cortices. The representation of the body is disproportionate, with large regions of the motor cortex devoted to areas of the body involved in fine movement, such as the face and the hands.

Finally, there is a convergence of projections from several sensory association regions to the motor regions of the frontal cortex. For example, the premotor cortex receives a convergence of afferents from higher-order somatosensory and visual areas of the posterior parietal cortex, whereas afferents from the primary somatosensory cortex converge on the primary motor cortex. In addition to cortical input, the primary motor cortex receives afferents from the ventral lateral nucleus of the thalamus; that nucleus receives afferents predominantly from the cerebellum. The premotor cortex receives input from the ventral anterior thalamic nucleus, which receives much of its input from the globus pallidus.

THALAMOCORTICAL ASSOCIATION SYSTEMS The multimodal association areas of the cortex are organized according to several general principles. First, association regions receive a convergence of input from a variety of sources, including unimodal and multimodal association regions of the cortex, the association nuclei of the thalamus, and other structures. For example, the prefrontal cortex receives afferents from higher-order sensory cortices of the parietal and temporal lobes, the contralateral prefrontal cortex, the cingulate cortex of the limbic system, the medial dorsal nucleus of the thalamus, an association relay nucleus, and portions of the amygdala. The medial dorsal nucleus receives highly processed inputs from many sources, including some regions, such as the amygdala, that project directly to the prefrontal cortex. The redundant (direct and indirect) projections may serve to attach additional significance to certain inputs received by the prefrontal cortex. The significance of these inputs may also be influenced by their temporal and spatial coincidence with modulatory inputs from brainstem nuclei that utilize the monoamine neurotransmitters dopamine, norepinephrine, or serotonin. Interestingly, the density of the monoamine afferent systems differs substantially across cortical regions of the human brain (Fig. 1.2-18), suggesting that the relative influence of these systems differs with the functional characteristics of the cortical region.

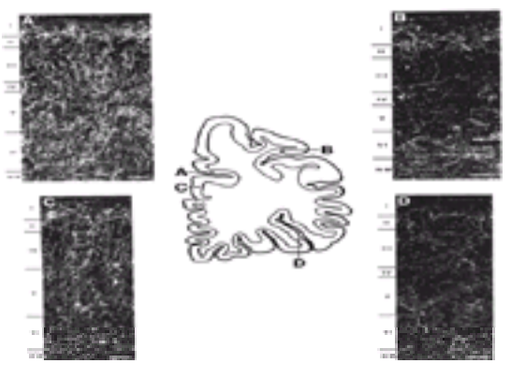

FIGURE 1.2-18 Photomicrographs illustrating the differential distribution of dopamine-containing axons in different regions of the human prefrontal **(A, B, D)** and anterior cingulate **(C)** cortices. The center panel represents a coronal section through the human prefrontal cortex at the level of the genu of the corpus callosum (cc), and the surrounding panels show the density of dopamine axons (white lines) at the indicated locations. Roman numerals indicate the cortical layers, and WM indicates white matter. Calibration bars equal 200 μm. (Reprinted with permission from Lewis DA: The catecholaminergic innervation of primate prefrontal cortex. J Neural Transm *36:*179, 1992.)

Second, the projections that terminate in multimodal association regions exhibit a topographical organization. Since the information conveyed in those projections is highly processed, it does not appear that the topographical organization of the afferents is a representation of the external world. Nonetheless, a distinct pattern is present in the afferents received by association areas. For example, different cytoarchitectonic regions of the medial dorsal nucleus project to discrete regions of the prefrontal cortex. In addition, some of the cortical afferents received by the prefrontal cortex are topographically organized; certain regions of the prefrontal cortex

predominantly receive highly processed information from one modality.

The patterns of connectivity are clearly related to some of the functional characteristics attributed to the prefrontal cortex. For example, in monkeys, lesions of the dorsolateral prefrontal cortex consistently produce an impairment in the monkey's ability to perform spatial delayed-response tasks. Those tasks require the monkey to maintain a spatial representation of the location of an object during a delay period in which the object is out of sight; it has been suggested that the prefrontal cortex plays a role in maintaining the spatial representation of the object. Such a function would require that the prefrontal cortex receive information regarding the location of objects in space; indeed, the dorsolateral prefrontal cortex is innervated by afferents from association regions of the parietal cortex that convey such information. Although the dorsolateral prefrontal cortex is necessary for the performance of delayed-response tasks in the monkey, it is not sufficient for the performance of the task. For example, lesions of the medial dorsal nucleus in the monkey result in similar impairments on the performance of spatial delayed-response tasks. Thus, the functions attributed to the prefrontal cortex are a result of the neural circuitry involving the region.

Knowledge of the integration of afferent inputs into the neural circuitry of certain prefrontal regions may also be important for understanding the nature of prefrontal cortical dysfunction in schizophrenia. Schizophrenic patients perform poorly on tasks that are known to be mediated by the prefrontal cortex. Those findings have been correlated with other measures to suggest, albeit indirectly, that the dopamine projections to the prefrontal cortex are impaired in schizophrenia. For example, studies in nonhuman primates have shown that performance of delayed-response tasks, the same type of behaviors that are impaired in schizophrenic subjects, requires an appropriate level of dopamine input to the dorsolateral prefrontal cortex.

CEREBELLO-THALAMOCORTICAL SYSTEMS The cerebellum has traditionally been considered to be involved solely with motor control, regulating posture, gait, and voluntary movements. However, recent studies indicate that the cerebellum may also play an important role in the mediation of certain cognitive abilities through inputs to portions of the thalamus that project to association regions of the cerebral cortex.

The cerebellum is located in the posterior cranial fossa, inferior to the occipital lobes (Fig. 1.2-4). The external surface of the cerebellum, the cerebellar cortex, is composed of small folds, termed *folia,* separated by sulci. Viewed from the dorsal surface, the cerebellum contains a raised central portion, called the *vermis,* and lateral portions called the *cerebellar hemispheres* (Fig. 1.2-19). Located within the cerebellum are the deep cerebellar nuclei that are arranged as follows: the fastigial nucleus is located next to the midline; located slightly more lateral are the globose and emboliform nuclei; and the largest nucleus, the dentate, occupies the most lateral position. In general, the cerebellar cortex can be considered to process the inputs to the cerebellum, and the deep nuclei to process the outputs.

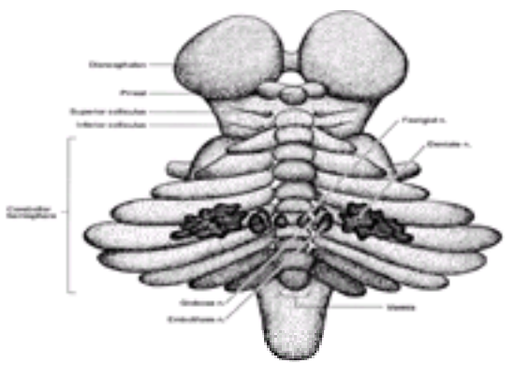

FIGURE 1.2-19 Schematic drawing of the dorsal view of the cerebellum showing the relative location and size of the cerebellar nuclei situated deep within the cerebellum. (Adapted from Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.)

Although many portions of the cerebellum are interconnected with brain regions that regulate motor actions, from the standpoint of psychiatric illness the circuitry of the cerebellum involved in cognitive functions is of greatest interest. For example, the lateral cerebellar cortex and the dentate nucleus are markedly expanded in the primate brain. It has been suggested that these changes are associated with an increase in the size of cortical areas (especially the prefrontal regions) influenced by cerebellar output and an expanded role of the cerebellum in cognitive functions. Recent studies in nonhuman primates have shown that the dorsolateral prefrontal cortex receives inputs from two ipsilateral thalamic nuclei (medial dorsal and ventral lateral), which in turn receive inputs from the contralateral cerebellar dentate nucleus. The cells of the dentate nucleus involved in these connections are distinct from those that influence the motor and premotor regions of the cerebral cortex. Interestingly, functional imaging studies in schizophrenic subjects have revealed abnormal patterns of activation in the cerebellum, thalamus, and prefrontal cortex, suggesting that dysfunction of this circuitry might be associated with the disturbances in cognitive processes exhibited by these patients.

Basal Ganglia System The basal ganglia are a collection of nuclei that have been grouped together on the basis of their interconnections. These nuclei play an important role in regulating movement and in certain disorders of movement (dyskinesias), which include jerky movements (chorea), writhing movements (athetosis), and rhythmic movements (tremors). In addition, recent studies have shown that certain components of the basal ganglia play an important role in many cognitive functions.

Major Structures The basal ganglia are generally considered to include the *caudate nucleus,* the *putamen,* the *globus pallidus* (referred to as the *paleostriatum* or *pallidum*), the *subthalamic nucleus,* and the *substantia nigra* (Fig. 1.2-20). The term *striatum* refers to the caudate nucleus and the putamen together; the term *corpus striatum* refers to the caudate nucleus, the putamen, and the globus pallidus; and the term *lentiform nucleus* refers to the putamen and the globus pallidus together.

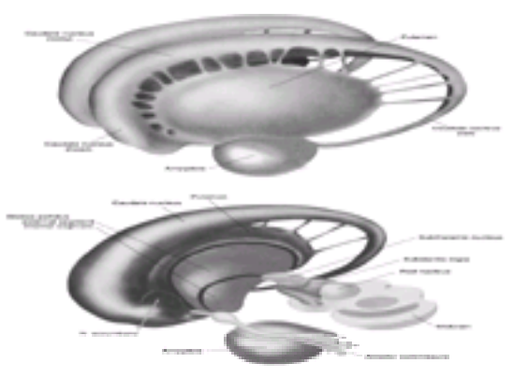

FIGURE 1.2-20 Schematic drawing of the isolated basal ganglia as seen from the dorsolateral perspective, so that the caudate nucleus is apparent bilaterally. In the bottom panel, the basal ganglia from the left hemisphere has been removed, exposing the medial surface of the right putamen and globus pallidus, as well as the subthalamic nucleus and substantia nigra. (Adapted from Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.)

Although these nuclei are generally agreed to belong to the basal ganglia, some controversy exists concerning whether other nuclei should be included in the definition of the basal ganglia. Some investigators believe that additional regions of the brain have anatomical connections that are similar to other components of the basal ganglia and should, therefore, be included in the definition. Those additional regions are usually termed the *ventral striatum* and the *ventral pallidum*. The ventral striatum includes the nucleus accumbens (Fig. 1.2-20), which is the region where the putamen and the head of the caudate nucleus fuse, and the olfactory tubercle. The ventral pallidum is a region that receives afferents from the ventral striatum and includes but is not limited to a group of neurons termed the *substantia innominata* (Fig. 1.2-5). This section focuses on the structures generally accepted as belonging to the basal ganglia but also discusses additional structures when relevant to the functional anatomy of the system.

CAUDATE NUCLEUS The caudate nucleus is a C-shaped structure that is divided into three general regions. The anterior portion of the structure is referred to as the

head, the posterior region is the *tail*, and the intervening region is the *body* (Fig. 1.2-20). The caudate nucleus is associated with the contour of the lateral ventricles: the head lies against the frontal horn of the lateral ventricle, and the tail lies against the temporal horn (Fig. 1.2-5, Fig. 1.2-6, and Fig. 1.2-7). The head of the caudate nucleus is continuous with the putamen; the tail terminates in the amygdala of the temporal lobe.

PUTAMEN The putamen lies in the brain medial to the insula and is bounded laterally by the fibers of the external capsule and medially by the globus pallidus (Fig. 1.2-5 and Fig. 1.2-6); the putamen is continuous with the head of the caudate nucleus (Fig. 1.2-20). Although bridges of neurons between the caudate nucleus and the putamen show the continuity of the nuclei, the two structures are separated by fibers of the anterior limb of the internal capsule.

GLOBUS PALLIDUS In contrast to the caudate nucleus and the putamen, which are telencephalic in origin, the globus pallidus is derived from the diencephalon. The globus pallidus constitutes the inner component of the lentiform nucleus (Fig. 1.2-20, bottom panel); with the putamen it forms a conelike structure, with its tip directed medially (Fig. 1.2-5 and Fig. 1.2-6). The posterior limb of the internal capsule bounds the globus pallidus medially and separates it from the thalamus; the putamen borders the globus pallidus laterally. In the human the medial medullary lamina divides the globus pallidus into external (lateral) and internal (medial) segments (Fig. 1.2-5 and Fig. 1.2-6).

SUBTHALAMIC NUCLEUS The subthalamic nucleus (of Luys) is also derived from the diencephalon. The large-celled nucleus lies dorsomedial to the posterior limb of the internal capsule and dorsal to the substantia nigra (Fig. 1.2-6 and Fig. 1.2-20). Discrete lesions of the subthalamic nucleus in humans lead to hemiballism, a syndrome characterized by violent, forceful choreiform movements that occur on the side contralateral to the lesion.

SUBSTANTIA NIGRA The substantia nigra is present in the midbrain between the tegmentum and the basis pedunculi and is mesencephalic in origin (Fig. 1.2-6). The substantia nigra consists of two components: a dorsal cell-rich portion referred to as the *pars compacta* and a ventral cell-sparse portion called the *pars reticulata*. Most of the neurons in the pars compacta of the substantia nigra in humans are pigmented because of the presence of neuromelanin; those cells contain the neurotransmitter dopamine. The dendrites of the pars compacta neurons frequently extend into the pars reticulata, where they receive synapses from the neurons of the pars reticulata that use the inhibitory neurotransmitter GABA.

In rodents, the dopamine-containing neurons of the substantia nigra (A9 region) have been distinguished from those located in the ventral tegmental area (A10 region) and the retrorubral field (A8 region), but recent studies in monkey and human brains suggest that dopamine neurons can be more meaningfully parcellated at a functional level into dorsal and ventral tiers (Fig. 1.2-21). The dorsal tier is formed by a medially and laterally oriented band of neurons that includes the dopamine-containing cells that are (1) located in the medial ventral mesencephalon, (2) scattered dorsal to the dense cell clusters in the substantia nigra, and (3) distributed lateral and caudal to the red nucleus. The ventral tier is composed of the dopamine neurons that are densely packed in the substantia nigra and the cell columns that penetrate into the substantia nigra pars reticulata. Dorsal-tier dopamine neurons express relatively low concentrations of the messenger ribonucleic acids (mRNAs) for the dopamine transporter and the dopamine type 2 (D_2) receptor, contain the calcium-binding protein calbindin, and send axonal projections to the regions of the striatum that are dominated by input from limbic-related structures and association regions of the cerebral cortex. In contrast, ventral-tier neurons contain high concentrations of the mRNAs for the dopamine transporter and the D_2 receptor, typically lack calbindin, and send axonal projections to the sensorimotor regions of the striatum. Each of these features may contribute to the greater vulnerability of ventral-tier neurons to the pathology of Parkinson's disease, whereas dorsal tier neurons may be more likely to be involved in the pathophysiology of schizophrenia.

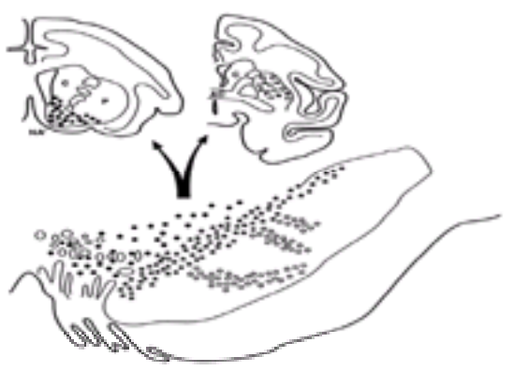

FIGURE 1.2-21 Schematic drawing of the topographic organization of dopamine-containing neurons in the mesencephalon and their projections to the ventral striatum and the sensorimotor-related dorsal striatum. All areas of the ventral striatum receive inputs from the dorsal tier neurons, with the shell region of the nucleus accumbens innervated almost exclusively by dorsal tier neurons (filled circles). In contrast, the ventral columns of cells (open circles) in the ventral tier send projections selectively to the sensorimotor-related striatum. The neurons of the densocellular zone (stars) of the ventral tier are unique in that they project to both the ventral and sensorimotor-related striatum. AC, anterior commissure; C, caudate nucleus; NA, nucleus accumbens; P, putamen. (Adapted from Lynda-Balta E, Haber SN: The organization of midbrain projections to the ventral striatum in the primate. Neuroscience *59:*625, 1994.)

Internal Organization The caudate nucleus and the putamen are frequently referred to together because of their common characteristics. For example, in the rodent these nuclei are a continuous structure, and in all mammals, they are composed of histologically identical cells. The majority of neurons in the striatum are medium-sized cells (10–20 μm in diameter) that possess spines on their dendrites; these so-called *medium spiny neurons* are known to send their axons out of the striatum. In addition to medium spiny neurons, medium-sized cells without spines (*medium aspiny neurons*) are present, as are large cells with and without spines (*large spiny neurons* and *large aspiny neurons*). With the exception of the medium and large spiny cells, most other striatal neurons are local circuit neurons.

Immunohistochemical and receptor-binding studies have shown a discontinuity in the distribution of certain neurotransmitter-related substances that form the functional circuitry of the basal ganglia. For example, in the striatum, zones that contain a low density of acetylcholinesterase enzymatic activity are surrounded by regions rich in acetylcholinesterase activity. The acetylcholinesterase-rich regions are referred to as the *matrix,* and the acetylcholinesterase-poor zones are termed either *striosomes* in the primate or *patches* in the rodent. The organization of several neuropeptide systems follows this system. For example, the distributions of enkephalin, substance P, and somatostatin immunoreactivity show the compartmentalization of the striatum. In addition, in the rodent certain subtypes of dopamine receptors are present predominantly in one compartment as compared with the other. In addition, the distribution of some afferent systems terminating in the striatum follows the striosome-matrix organization. For example, afferents from the thalamus terminate preferentially in the matrix rather than in the striosome.

Functional Circuitry Projections into, within, and out of the basal ganglia are topographically organized and maintain that topography throughout he processing circuits of the basal ganglia. The existence of such patterns of connectivity has resulted in the hypothesis that parallel, independent circuits in the basal ganglia process information from different regions of the brain and subserve separate complex functions.

INPUTS TO THE BASAL GANGLIA The striatum is the major recipient of the inputs to the basal ganglia. Three major afferent systems are known to terminate in the striatum: the corticostriatal, the nigrostriatal, and the thalamostriatal afferents (Fig. 1.2-22). The corticostriatal projection originates from all regions of the neocortex, arising primarily from the pyramidal cells of layers V and VI, which utilize the excitatory neurotransmitter glutamate. A topography governing corticostriatal projections has been found in the monkey. Afferents from the sensorimotor cortex terminate predominantly in the putamen; association regions of the cortex terminate preferentially in the caudate nucleus. The prefrontal regions, in particular, provide a heavy input to the head of the caudate nucleus. In addition, afferents from the limbic cortical areas and from the hippocampus and the amygdala terminate in the ventral striatum. The second major class of afferents utilize the neurotransmitter dopamine. In Figure 1.2-22, these projections are shown arising from the substantia nigra pars compacta, but as noted previously (Fig. 1.2-21), different portions of the striatum receive input from either the dorsal- or ventral-tier dopamine-containing neurons of the ventral mesencephalon. Electron microscopy studies have shown that many of the synapses formed by dopamine axon terminals on medium spiny neuron dendrites are immediately adjacent to the synapses provided by corticostriatal axons, suggesting that dopamine may play an important role in modulating the excitatory influence of cortical projections on striatal neurons. The third afferent system originates in the thalamus. The thalamic nuclei providing the projections are the intralaminar nuclei, particularly the central median nucleus.

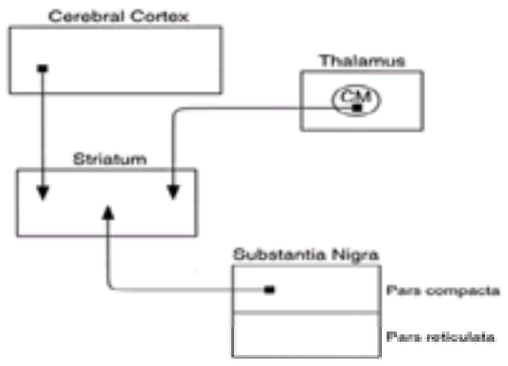

FIGURE 1.2-22 Diagram of the inputs to the basal ganglia system. Three major afferent systems have been identified: the corticostriatal, thalamostriatal, and nigrostriatal pathways.

Disruption of the input pathways of the basal ganglia has been associated with some movement disorders, such as Parkinson's disease, which is characterized by muscular rigidity, fine tremor, shuffling gait, and bradykinesia. The most consistent neuropathological feature of Parkinson's disease is a degeneration of the dopamine neurons in the substantia nigra pars compacta, accompanied by a loss of dopamine terminals in the striatum. Levodopa (Larodopa, Dopar), a precursor in the biosynthesis of dopamine, is used as a treatment for Parkinson's disease because of its ability to augment the release of dopamine from the remaining terminals. Conversely, the administration of dopamine receptor antagonists (so-called typical antipsychotics) agents in the treatment of schizophrenia is frequently associated with Parkinsonian features and other motor-system abnormalities; the fact that these agents are D_2-receptor antagonists is thought to explain their movement-related adverse effects.

INTERNAL PROCESSING The major processing pathways within the basal ganglia are summarized in Figure 1.2-23. Within the striatum, the subclass of medium spiny neurons that contain the neuropeptide substance P send inhibitory projections to the internal segment of the globus pallidus in what is termed the *direct pathway*. In contrast, the subpopulation of medium spiny neurons that contain the neuropeptide enkephalin provides inhibitory projections to the external segment of the globus pallidus, which in turn sends inhibitory afferents to the internal segment of the globus pallidus in what is termed the *indirect pathway*. The globus pallidus external also projects to the pars reticulata of the substantia nigra. The topography found in the afferent projections to the striatum appears to be maintained in that processing pathway. For example, the sensorimotor territories of the striatum project most heavily to the ventral portion of the globus pallidus, whereas association territories project to the dorsal regions of the globus pallidus.

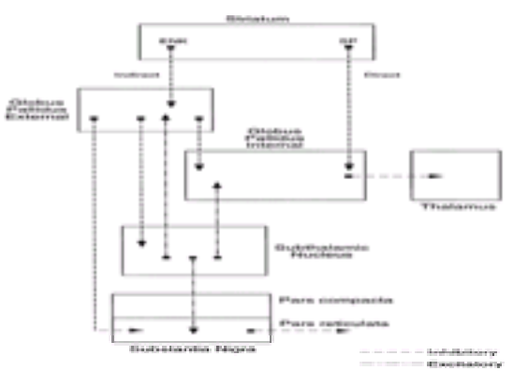

FIGURE 1.2-23 Diagram of the intrinsic circuitry of the basal ganglia. Substance P (SP)-containing striatal neurons send an inhibitory projection directly to the internal segment of the globus pallidus, whereas those containing enkephalin provide an inhibitory projection to GABA neurons in the external segment of the globus pallidus, which in turn project to the internal segment of the globus pallidus. The subthalamic nucleus receives a projection from the external segment of the globus pallidus and projects back to both segments. Finally, the subthalamic nucleus and globus pallidus external project to the substantia nigra pars reticulata.

The external segment of the globus pallidus also gives rise to an inhibitory projection that terminates in the subthalamic nucleus. In contrast, neurons in the subthalamic nucleus provide excitatory projections that terminate in both segments of the globus pallidus and in the pars reticulata. Although most connections within the basal ganglia are unidirectional, a reciprocal projection is found between the external segment of the globus pallidus and the subthalamic nucleus.

The intrinsic circuitry of the basal ganglia is disrupted by a severe loss of neurons in the striatum in Huntington's disease. This autosomal-dominant disorder is characterized by progressive chorea and dementia. Although the gene for Huntington's disease has been identified, how the excessive number of trinucleotide repeats in this gene leads to the selective degeneration of striatal cells is currently a matter of intense investigation. Interestingly, recent studies indicate that cortical neurons are also subject to degeneration in Huntington's disease.

OUTPUT OF BASAL GANGLIA The internal segment of the globus pallidus is the source of much of the output of the basal ganglia (Fig. 1.2-24). That segment of the globus pallidus provides a projection to the ventral lateral and ventral anterior nuclei of the thalamus and to the intralaminar thalamic nuclei—in particular, the central median nucleus. The pars reticulata of the substantia nigra also provides a projection to the ventral anterior and ventral lateral thalamic nuclei. Those portions of the ventral lateral and ventral anterior thalamic nuclei then project to the premotor and prefrontal cortices. As a result of the projections of the premotor and prefrontal cortices to the primary motor cortex, the basal ganglia are able to indirectly influence the output of the primary motor cortex. In addition, the cortical output of the basal ganglia exhibits marked convergence, that is, although the striatum receives afferents from all regions of the neocortex, the eventual output of the globus pallidus and the pars reticulata is largely conveyed through the thalamus to a much smaller portion of the neocortex, the premotor and prefrontal regions.

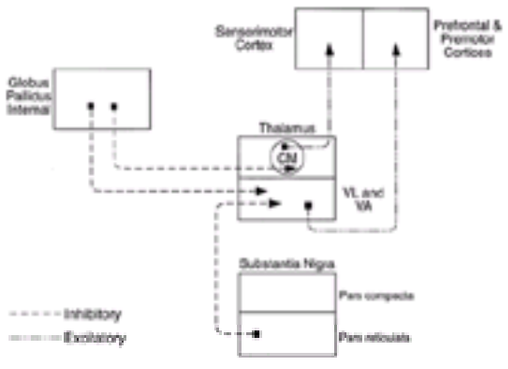

FIGURE 1.2-24 Diagram of the output of the basal ganglia system. The internal segment of the globus pallidus projects to the central median (CM), ventral lateral (VL), and ventral anterior (VA) nuclei of the thalamus. Those nuclei then project to sensorimotor, prefrontal, and premotor cortices. The substantia nigra pars reticulata also projects to the VL and VA nuclei.

The functional consequences of the neural circuitry of the basal ganglia can also be considered in the context of some of the neurotransmitters used (Fig. 1.2-23 and Fig. 1.2-24). Since the afferents from the cortex are thought to use glutamate, which is an excitatory neurotransmitter, cortical afferents presumably excite the

structures of the basal ganglia in which they terminate. Many of the processing pathways within the basal ganglia use the inhibitory neurotransmitter GABA. Finally, the output pathways of the basal ganglia—namely, the globus pallidus and the substantia nigra pars reticulata—use GABA as well. Thus, excitation from cortical afferents eventually disinhibits the target structures of the basal ganglia because of the back-to-back inhibitory pathways of the basal ganglia.

Historically, motor systems have been divided into pyramidal (corticospinal) and extrapyramidal (basal ganglia) components; that division is based on clinical findings suggesting that lesions of each system result in distinct motor syndromes. For example, lesions of the extrapyramidal system result in involuntary movements, changes in muscle tone, and slowness of movement; lesions of the pyramidal system lead to spasticity and paralysis. Because of these findings the pyramidal and extrapyramidal systems were thought to independently control voluntary and involuntary movement, respectively. However, for several reasons that division is no longer accurate. First, other structures of the brain outside the traditional pyramidal and extrapyramidal systems, such as the cerebellum, are involved in the control of movement. Second, the pyramidal and extrapyramidal systems are not independent—their neural circuits are interconnected. For example, the basal ganglia influence motor behavior through certain regions of the cerebral cortex, which then directly (through the corticospinal tract) or indirectly (through specific brainstem nuclei) produce motor activity.

Finally, although the basal ganglia are important in the control of movement, they also appear to be involved in other functions of the brain. For example, recent studies of the connections of the basal ganglia in nonhuman primates also support a role for these structures in cognitive functions. The dorsolateral prefrontal cortex has been shown to receive inputs from portions of the thalamus that are the targets of projections from specific locations within the internal segment of the globus pallidus, providing evidence for a distinct pallido-thalamo-cortical pathway. Thus, in addition to linking association regions of the cerebral cortex, such as the prefrontal and posterior parietal areas, with the control of motor activity in the primary motor cortex, some of the output of the basal ganglia appears to be directed back to regions of the prefrontal cortex. These findings suggest that “closed” loops exist between the prefrontal cortex and basal ganglia, which presumably have a cognitive rather than a motor function.

Limbic System The concept of the limbic system as a neural substrate for emotional experience and expression has a rich but controversial history. More than 100 years ago Pierre Broca applied the term “limbic” (from the Latin *limbus* for border) to the curved rim of the cortex, including the cingulate and the parahippocampal gyri, located at the junction of the diencephalon and the cerebral hemispheres (Fig. 1.2-25). In 1937, primarily on the basis of anatomical data James Papez postulated that these cortical regions were linked to the hippocampus, the mammillary body, and the anterior thalamus in a circuit that mediated emotional behavior (Fig. 1.2-26). That concept was supported by the work of Heinrich Klüver and Paul Bucy, who demonstrated that temporal lobe lesions, which disrupt components of the circuit, alter affective responses in nonhuman primates. In 1952, Paul MacLean coined the term *limbic system* to describe Broca's limbic lobe and related subcortical nuclei as the neural substrate for emotion.

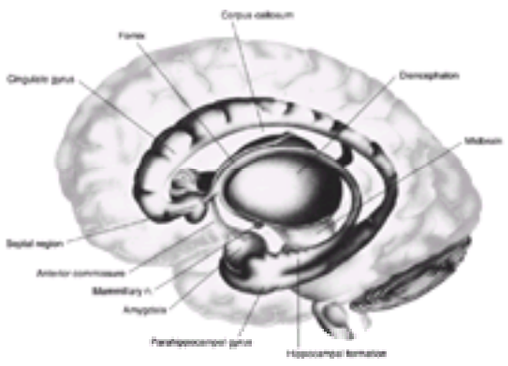

FIGURE 1.2-25 Schematic drawing of the major anatomical structures of the limbic system. Note that the cingulate and parahippocampal gyri form the “limbic lobe,” a rim of tissue located along the junction of the diencephalon and the cerebral hemispheres. (Adapted from Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.)

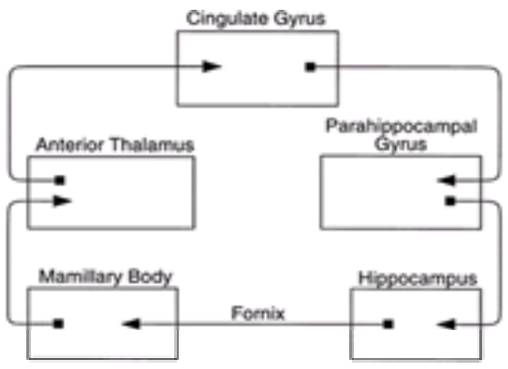

FIGURE 1.2-26 Diagram of the neural circuit for emotion as originally proposed by James Papez.

However, over the last 45 years it has become clear that some limbic structures (e.g., the hippocampus) are also involved in other complex brain processes such as memory. In addition, expanding knowledge of the connectivity of traditional limbic structures has made it increasingly difficult to define the boundaries of the limbic system. Despite those limitations, the concept of a limbic system may still be a useful way to describe the circuitry that relates certain telencephalic structures and their cognitive processes with the hypothalamus and its output pathways that control autonomic, somatic, and endocrine functions.

Major Structures There is no unanimity on which brain structures constitute the limbic system. This section includes the brain regions that are most commonly listed as components of the limbic system: the cingulate and parahippocampal gyri (limbic cortex), the hippocampal formation, the amygdala, the septal area, the hypothalamus, and related thalamic and cortical areas.

LIMBIC CORTEX The limbic cortex is composed of two general regions, the cingulate gyrus and the parahippocampal gyrus (Fig. 1.2-25). The *cingulate gyrus,* located dorsal to the corpus callosum, includes several cortical regions that are heavily interconnected with the association areas of the cerebral cortex. As the cingulate gyrus travels posteriorly, it becomes continuous (via the cingulum bundle of fibers in the white matter) with the *parahippocampal gyrus,* located in the medial temporal lobe, which contains several distinct cytoarchitectonic regions. One of the most important of those regions is the *entorhinal cortex,* which not only funnels highly processed cortical information to the hippocampal formation but is also a major output pathway from the hippocampal formation.

HIPPOCAMPAL FORMATION Three distinct zones—the dentate gyrus, the hippocampus, and the subicular complex—constitute the hippocampal formation, which is located in the floor of the temporal horn of the lateral ventricle (see Fig. 1.2-7). Those zones are composed of adjacent strips of cortical tissue that run in a rostral-caudal direction but fold over each other mediolaterally in a spiral fashion, resulting in a C-shaped appearance. The *dentate gyrus* is composed of three layers: an outer, acellular molecular layer, which faces the subarachnoid space of the hippocampal fissure; a middle layer composed of granule cells; and an inner polymorphic layer (Fig. 1.2-27). The granule cells extend their dendritic trees into the molecular layer and give rise to axons that form the mossy fiber projection to the hippocampus.

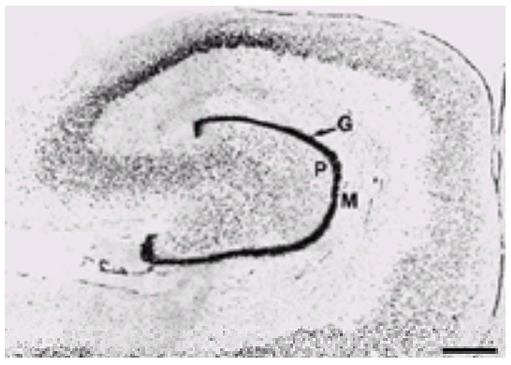

FIGURE 1.2-27 Nissl-stained coronal section through the dentate gyrus of the human hippocampal formation. Medial is to the left. M, molecular layer; G, granular layer; P, polymorphic layer. Calibration bar equals 1.0 mm.

The *hippocampus* is also a trilaminate structure composed of molecular and polymorphic layers and a middle layer that contains pyramidal neurons. On the basis of differences in cytoarchitecture and connectivity, the hippocampus can be divided into three distinct fields, which have been labeled CA3, CA2, and CA1. (*CA* is derived from the term *cornu ammonis* after the Egyptian deity Ammon, who was depicted with ram's horns, which some early investigators thought described the shape of the hippocampus.) The white matter adjacent to the polymorphic layer of the hippocampus is known as the *alveus*. The axons in that structure contribute to the fimbria, which at the caudal end of the hippocampus becomes the crus of the *fornix*. Those bilateral structures converge to form the body of the fornix, which travels anteriorly and then turns inferiorly to form the columns of the fornix, which pass through the hypothalamus into the mammillary bodies (Fig. 1.2-28). The *subicular complex* is generally considered to have three components—the presubicular, the parasubicular, and the subiculum—which together serve as transition regions between the hippocampus and the parahippocampal gyrus.

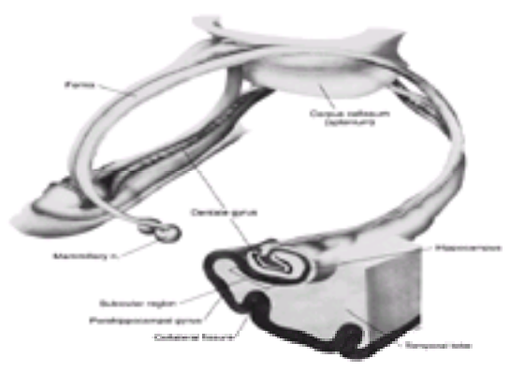

FIGURE 1.2-28 Schematic drawing of a cross-sectional view of the hippocampal formation and the path of the fornix running between that structure and the mammillary bodies. (Adapted from Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.)

The components of the hippocampal formation have a distinct pattern of intrinsic connectivity that is largely unidirectional and provides for a specific flow of information (Fig. 1.2-29). The major input to the hippocampal formation arises from neurons in layers II and III of the entorhinal cortex that project through the *perforant path* (i.e., through the subiculum and the hippocampus) to the outer two thirds of the molecular layer of the dentate gyrus, where they synapse on the dendrites of granule cells. The mossy fiber axons of the granule cells then provide a projection to the pyramidal neurons of the CA3 field of the hippocampus. Axon collaterals from CA3 pyramidal neurons project within CA3 and, through the so-called Schäffer collaterals, to the CA1 field of the hippocampus. That region in turn projects to the subicular complex, which provides output to the entorhinal cortex, completing the circuit.

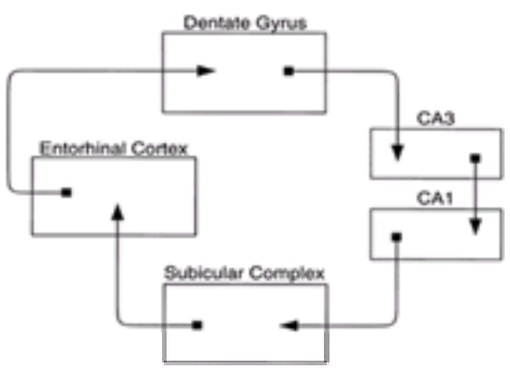

FIGURE 1.2-29 Diagram of the intrinsic neural circuitry of the hippocampal formation.

AMYGDALA Located in the medial temporal lobe just anterior to the hippocampal formation are a group of nuclei referred to as the *amygdala* (see Fig. 1.2-6). Those nuclei form several distinct clusters: the basolateral complex, the centromedial amygdaloid group, and the olfactory group, including the cortical amygdaloid nuclei. The *basolateral complex,* the largest of the three groups, differs from the remaining amygdaloid nuclei in a number of respects. Although the basolateral complex is not a laminated structure, its connectivity and some other anatomical characteristics are more similar to cortical regions than to the remaining amygdaloid nuclei. For example, the basolateral nuclei are directly and reciprocally connected with the temporal, insular, and prefrontal cortices. In addition, like some cortical regions, the basolateral complex shares bidirectional connections with the medial dorsal thalamic nucleus, and it receives projections from the midline and intralaminar thalamic nuclei. Finally, neurons of the basolateral complex with a pyramidal-like morphology appear to furnish projections to the striatum that use excitatory amino acids as neurotransmitters. Thus, on the basis of those anatomical characteristics, the basolateral complex can be said to function like a multimodal cortical region.

In contrast, the *centromedial amygdala* appears to be part of a larger structure that is continuous through the sublenticular substantia innominata with the bed nucleus of the stria terminalis. That larger structure, which has been termed the *extended amygdala*, consists of two major subdivisions. The central subdivision of the extended amygdala includes the central amygdaloid nucleus and the lateral portion of the bed nucleus of the stria terminalis. That subdivision is reciprocally connected with brainstem viscerosensory and visceromotor regions and with the lateral hypothalamus. In addition, it receives afferents from cortical limbic regions and the basolateral amygdaloid complex. In contrast, the medial subdivision of the extended amygdala, composed of the medial amygdaloid nucleus and its extension into the medial part of the bed nucleus of the stria terminalis, is distinguished by reciprocal connections with the medial or endocrine portions of the hypothalamus.

SEPTAL AREA The septal area is a gray matter structure located immediately above the anterior commissure (Fig. 1.2-30). The septal nuclei are reciprocally connected with the hippocampus, the amygdala, and the hypothalamus and project to a number of structures in the brainstem.

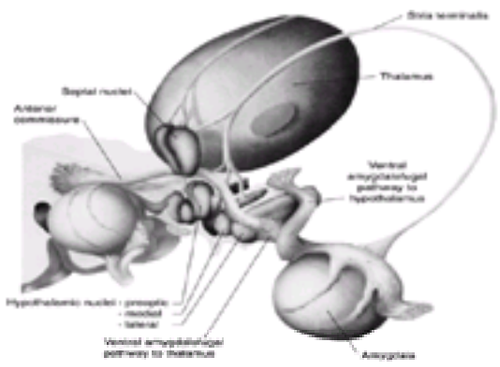

FIGURE 1.2-30 Schematic drawing of some components of the limbic system showing the major output pathways of the amygdala, the stria terminalis and the ventral amygdalofugal pathway. (Adapted from Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.)

HYPOTHALAMUS The hypothalamus, a relatively small structure within the diencephalon, is a crucial component of the neural circuitry regulating not only emotions but also autonomic, endocrine, and some somatic functions. In addition to its relations with other components of the limbic system, it is interconnected with various visceral and somatic nuclei of the brainstem and the spinal cord, and it provides an output that regulates the function of the pituitary gland. On its inferior surface the hypothalamus is bounded rostrally by the optic chiasm and caudally by the posterior edge of the mammillary bodies. The area of the hypothalamus between those two structures, called the *tuber cinereum,* gives rise to the median eminence, which is continuous with the infundibular stalk and then the posterior lobe of the pituitary gland (Fig. 1.2-31). On the basis of these features the hypothalamus is subdivided from anterior to posterior into three zones: the supraoptic region, the tuberal region, and the mammillary region. (In addition, the preoptic area, a telencephalic structure located immediately anterior to the supraoptic region is usually considered part of the hypothalamus.) These three zones are also divided on each side into medial and lateral areas by the fornix as it travels through the body of the hypothalamus to the mammillary bodies. As shown in Table 1.2-3, the six parts of the hypothalamus contain different nuclei.

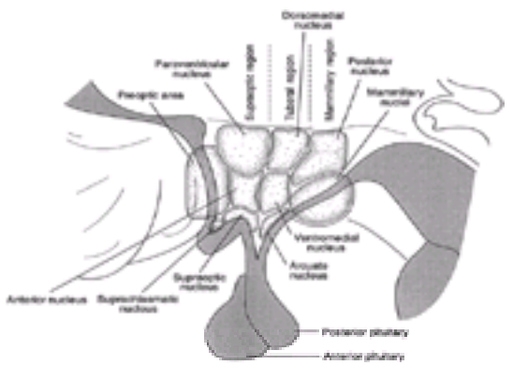

FIGURE 1.2-31 Schematic drawing of the nuclei in the medial hypothalamus. (Adapted from Burt AM: *Textbook of Neuroanatomy*. Saunders, Philadelphia, 1993.)

Region	Medial Area	Lateral Area
Supraoptic	Supraoptic nucleus Paraventricular nucleus Anterior nucleus Suprachiasmatic nucleus	Lateral nucleus Part of supraoptic nucleus
Tuberal	Dorsomedial nucleus Ventromedial nucleus Arcuate nucleus	Lateral nucleus Lateral tuberal nuclei
Mammillary	Mammillary body Posterior nucleus	Lateral nucleus

Adapted from Nolte J: *The Human Brain: An Introduction to Its Functional Anatomy*, ed 2. Mosby, St. Louis, 1990.

Table 1.2-3 Hypothalamic Nuclei

Those different nuclei subserve the diverse functions of the hypothalamus. For example, the *suprachiasmatic nucleus* receives both direct and indirect projections from the retina and appears to be important in the regulation of diurnal rhythms. The *supraoptic* and *paraventricular nuclei* contain large cells (magnocellular neurons) that send oxytocin and vasopressin-containing fibers to the posterior neural lobe of the pituitary. In addition, some neurons of the paraventricular nucleus project to the median eminence, where they release neuropeptides, such as corticotropin-releasing factor, into the portal blood system. These neuropeptides then control the synthesis and the release of anterior pituitary hormones. The paraventricular nucleus also gives rise to descending projections that regulate the sympathetic and parasympathetic autonomic areas of the medulla and the spinal cord.

Within the medial tuberal region of the hypothalamus, the *ventromedial* and *arcuate nuclei* also participate in the regulation of the anterior pituitary function. In addition, the ventromedial nucleus may play an important role in reproductive and ingestive behavior. The medial posterior section of the hypothalamus contains the *posterior nucleus* and the *mammillary bodies*. Within the mammillary bodies, the lateral and medial mammillary nuclei receive hippocampal input through the fornix (Fig. 1.2-28) and project to the anterior nuclei of the thalamus. The posterior nucleus shares reciprocal connections with the extended amygdala. That nucleus appears to be more developed in primates than in rodents, suggesting that it plays an important role in the human brain, one that has still to be clarified.

The lateral portions of the hypothalamus contain a relatively low density of neurons scattered among longitudinally running fibers of the medial forebrain bundle, which is interconnected with multiple regions of the forebrain, the brainstem, and the spinal cord.

Functional Circuitry The major structures of the limbic system are interconnected with each other and with other components of the nervous system in a variety of ways. However, several major output pathways of the limbic system are clearly defined. In one pathway (Fig. 1.2-32) highly processed sensory information from the cingulate, the orbital and temporal cortices, and the amygdala is transmitted to the entorhinal cortex of the parahippocampal gyrus and from there to the hippocampal formation. After traversing the intrinsic circuitry of the hippocampal formation, information is projected through the fornix either to the anterior thalamus, which in turn projects to the limbic cortex or to the septal area and the hypothalamus. Those latter two regions provide feedback to the hippocampal formation through the fornix. In addition, the mammillary bodies of the hypothalamus project to the anterior thalamus. Finally, the hypothalamus and the septal area project to the brainstem and the spinal cord.

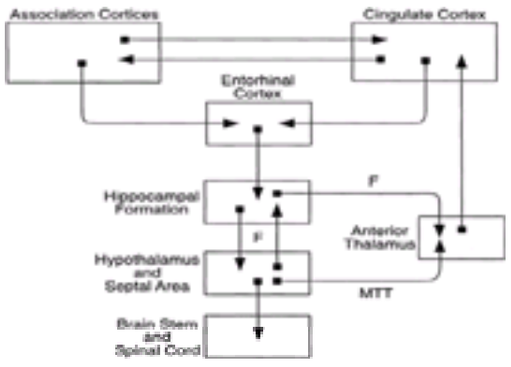

FIGURE 1.2-32 Functional neural circuitry of the limbic system. This diagram illustrates the manner in which the hippocampal formation and the anterior thalamus provide a mechanism for the integration of information between the cerebral cortex and the hypothalamus. F, fornix; MTT, mammillothalamic tract. (Adapted from Nolte J: *The Human Brain: An Introduction to Its Functional Anatomy,* ed 3. Mosby, St. Louis, 1993.)

Another major pathway within the limbic system centers on output from the amygdala (Fig. 1.2-33). Highly processed sensory information, primarily from the association regions of the prefrontal and temporal cortices, projects to the amygdala. Output from the amygdala is conducted through two main pathways (Fig. 1.2-30). A dorsal route, the *stria terminalis,* accompanies the caudate nucleus in an arch around the temporal lobe and contains axons that project primarily to the septal area and the hypothalamus. The second major output route, the *ventral amygdalofugal pathway* passes below the lenticular nucleus and contains fibers that terminate in a number of regions, including the septal area, the hypothalamus, and the medial dorsal thalamic nucleus. The medial dorsal nucleus in turn projects heavily to prefrontal and some temporal cortical regions.

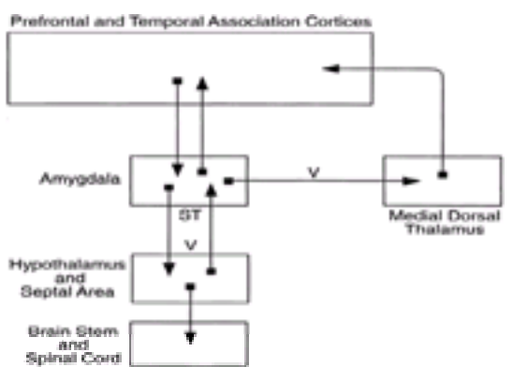

FIGURE 1.2-33 Functional neural circuitry of the limbic system. This diagram illustrates how the amygdala and the medial dorsal thalamus serve to integrate information processing between prefrontal and temporal association cortices and the hypothalamus. V, ventral amygdalofugal pathway; ST, stria terminalis. (Adapted from Nolte J: *The Human Brain: An Introduction to Its Functional Anatomy,* ed 3. Mosby, St. Louis, 1993.)

Both these pathways reveal how the limbic system is able to integrate the highly processed sensory and cognitive information content of the cerebral cortical circuitry with the hypothalamic pathways that control autonomic and endocrine systems. In addition, the limbic system interacts with components of the basal ganglia system (Fig. 1.2-34). For example, the ventral amygdalofugal pathway also projects to the nucleus accumbens (ventral striatum), the area where the head of the caudate nucleus fuses with the putamen (Fig. 1.2-20). That region sends efferents to the ventral palladium, an extension of the globus pallidus, which in turn projects to the medial dorsal thalamic nucleus. The pathway indicates that the functions of the basal ganglia extend beyond the regulation of motor activities and shows the necessity of considering the function or dysfunction of particular brain regions in the context of all aspects of their circuitry.

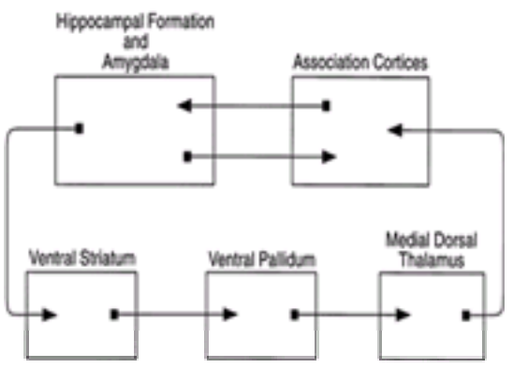

FIGURE 1.2-34 Functional neural circuitry of the limbic system. This drawing illustrates the interaction between the limbic system and certain components of the basal ganglia. (Adapted from Nolte J: *The Human Brain: An Introduction to Its Functional Anatomy,* ed 3. Mosby, St. Louis, 1993.)

SUGGESTED CROSS-REFERENCES

Section 1.4 discusses monoamine neurotransmitters, Section 1.5 discusses amino acid neurotransmitters, Section 1.8 discusses intraneural signaling pathways, Section 2.5 discusses movement disorders, Section 3.5 discusses the brain circuitry that subserves memory, Section 12.3 discusses brain structure and function in schizophrenia, Chapter 37 discusses communication disorders, Section 51.2b discusses central nervous system changes in normal aging, Section 51.2e and Section 51.2f discuss neuroimaging, and Section 51.3e discusses Alzheimer's disease and other dementing disorders.

SECTION REFERENCES

Alexander GE, Crutcher MD: Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. Trends Neurosci *13:*266, 1990.

Beall MJ, Lewis DA: Heterogeneity of layer II neurons in human entorhinal cortex. J Comp Neurol *321:*241, 1992.

Bloom FE, Björklund A, Hökfelt T, editors: *Handbook of Chemical Neuroanatomy,* vol 13. New York, Elsevier, 1997.

Burt AM: *Textbook of Neuroanatomy*. Saunders, Philadelphia, 1993.

Calabresi P, De Murtas M, Bernard G: The neostriatum beyond the motor function: Experimental and clinical evidence. Neuroscience *78:*39, 1997.

Castro-Alamancos MA, Connors BW: Thalamocortical synapses. Prog Neurobiol *51:*581, 1997.

Chesselet M-F, Delfs JM: Basal ganglia and movement disorders: An update. Trends Neurosci *19:*417, 1996.

Condé F, Lund JS, Jocobowitz DM, Baimbridge KG, Lewis DA: Local circuit neurons immunoreactive for calretinin, calbindin D-28k or parvalbumin in monkey prefrontal cortex: Distribution and

morphology. J Comp Neurol *341:*95, 1994.

Felleman DJ, Van Essen DC: Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex *1:*36, 1991.

Gloor P: *The Temporal Lobe and Limbic System*. Oxford, New York, 1997.

Guillery RW, Feig SL, Lozsádi DA: Paying attention to the thalamic reticular nucleus. Trends Neurosci *21:*28, 1998.

*Heimer L, Harlan RE, Alheid GF, Garcia MM, De Olmos J: Substantia innominata: A notion which impedes clinical-anatomical correlations in neuropsychiatric disorders. Neuroscience *76:*957, 1997.

Hendelman WJ: *Student's Atlas of Neuroanatomy*. Saunders, Philadelphia, 1994.

Kandel ER, Schwartz JH, Jessel TM, editors: *Principles of Neural Science,* ed 3. Elsevier, New York, 1991.

Lewis DA: Development of the prefrontal cortex during adolescence: Insights into vulnerable neural circuits in schizophrenia. Neuropsychopharmacology *16:*385, 1997.

Lewis DA: The catecholaminergic innervation of primate prefrontal cortex. J Neural Transm *36:*179, 1992.

*Lewis DA, Sesack SR: Dopamine systems in the primate brain. In *Handbook of Chemical Neuroanatomy,* FE Bloom, A Björklund, T Hökfelt, editors, vol 13. Elsevier, New York, 1997.

Lynda-Balta E, Haber SN: The organization of midbrain projections to the ventral striatum in the primate. Neuroscience *59:*625, 1994.

Nieuwenhuys R, Voogd J, Van Huijzen C: *The Human Central Nervous System: A Synopsis and Atlas,* ed 3. Springer, New York, 1988.

*Nolte J: *The Human Brain: An Introduction to its Functional Anatomy*. Mosby, St. Louis, 1993.

O'Donnell P, Grace AA: Dysfunctions in multiple interrelated systems as the neurobiological bases of schizophrenic symptom clusters. Schizophr Bull, in press.

Parent A, Hazrati L-N: Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res Rev *20:*91, 1995.

Parent A, Hazrati L-N: Functional anatomy of the basal ganglia. II. The place of subthalamic nucleus and external pallidum in basal ganglia circuitry. Brain Res Rev *20:*128, 1995.

Raymond JL, Lisberger SG, Mauk MD: The cerebellum: A neuronal learning machine? Science *272:*1126, 1996.

*Ricci PT, Zelkowitz BJ, Nebes RD, Meltzer CC, Mintun MA, Becker JT: Functional neuroanatomy of semantic memory: Recognition of semantic associations. Neuroimage *9:*88, 1999.

Risold PY, Thompson RH, Swanson LW: The structural organization of connections between hypothalamus and cerebral cortex. Brain Res Rev *24:*197, 1997.

*Salloway S, Cummings J: Subcortical structures and neuropsychiatric illness. Neuroscientist *2:*66, 1996.

Shink E, Bevan MD, Bolam JP, Smith Y: The subthalamic nucleus and the external pallidum: Two tightly interconnected structures that control the output of the basal ganglia in the monkey. Neuroscience *73:*335, 1996.

*Stahl SM: Substance P and the neurokinins: Novel peptide neurotransmitters in psychopharmacology. J Clin Psychiatry *60:*77, 1999.

*Young PA, Young PH: *Basic Clinical Neuroanatomy*. William & Wilkins, Baltimore, 1997.

1.3 DEVELOPMENTAL NEUROBIOLOGY

KATHRYN J. KOTRLA, M.D. AND DANIEL R. WEINBERGER, M.D.

Human behaviors, thoughts, and emotions reside in the interface between the mind and the brain. The mind is experienced; the brain is the substrate on which mind depends. Basic neuroscience research provides an impressive array of information about the brain. The genes and molecules required for normal brain formation and functioning are being discovered, as are the many neural networks that subserve different aspects of cognition, memory, and mood. The practicing psychiatrist does not need to know the details of neuronal connectivity and functioning. However, moving from a knowledge of the concepts of basic neuroscience to an understanding of the mind is within the domain of psychiatry and will impact psychiatric care substantially.

Three threads of basic neuroscience research are particularly important. Developmental neurobiology explores the mechanisms underlying prenatal and postnatal brain development. Psychiatric conditions like schizophrenia and autistic disorder are likely to have their origins in utero. Understanding development will illuminate the etiology of such disorders, and may suggest novel treatment and prevention strategies. During postnatal development there are critical periods in which the fine-tuning of cortical connections occurs. Understanding critical periods during early childhood may help to explain the lifelong difficulties of individuals who suffer trauma, abuse, or neglect in early life. Lastly, neuroscience is investigating the continued plasticity of the adult brain. Understanding how the adult brain changes with time and experience has implications for how psychiatric disorders change in their expression and treatment responsiveness over a lifetime.

PRENATAL DEVELOPMENT

The wonder of development is that a structure as complex as the human brain originates from a flat sheet of embryologic ectoderm. The final, formed brain shows remarkable order in its predictable cortical layering, its diversity of cortical areas, and the numerous networks linking specific cortical areas and subcortical structures. To have cells choosing to become a certain neuronal type, attaining the correct laminar position, finding the correct target, and expressing the correct neurotransmitters at first seems overwhelmingly difficult. However, the final, breathtakingly complex set of connections in the human brain depends on a series of much simpler decisions as neurons become progressively more restricted in the choices they make. These decisions require the subtle interplay of genetic and environmental factors; much has been learned at a molecular level about these processes. At first glance this information seems most relevant to mental retardation or autistic disorder, in which abnormal brain development results in lifelong disability. However, even schizophrenia is believed to originate in subtle aberrant brain development, and understanding it requires an understanding of its etiology.

Neurogenesis and Neural Identity The cerebral cortex possesses an orderly six-layered array of neuronal and glial cell types; layer I is the most superficial layer closest to the meninges, layer VI lies deeper, closest to white matter. Each layer has characteristic interconnections, with the superficial layers connecting with other cortical areas, and the deeper layers connecting to subcortical structures. Across the cortex there are microscopic variations in the appearance of the array, corresponding to areas of cortex with distinct functions, connections, receptors, and neurotransmitters (Fig. 1.3-1). Such complexity arises via a series of progressive restrictions in cell fate.

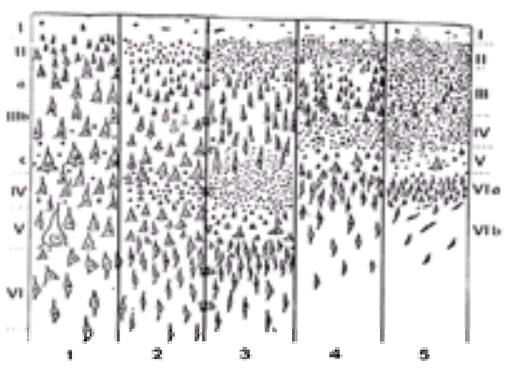

FIGURE 1.3-1 The cerebral cortex contains discrete areas with differing cytoarchitecture. This drawing shows the six layers of the cortex (I-VI). The five different panels illustrate marked differences in cortical structure in different brain areas. Note the variations in neuronal size and morphology. (Reprinted with permission from Carpenter MB, Sutin J: *Human Neuroanatomy,* ed 8. Williams & Wilkins, Baltimore, 1983.)

The initial decision to form a brain depends on the embryonic mesoderm inducing the overlying ectoderm to become nervous system. This sheet of ectoderm, the neural plate, invaginates to form the neural tube as cells are progressively determined to form forebrain, midbrain, and hindbrain even before any neurons have been generated.

On the inner surface of the neural tube there is a rapidly dividing pseudostratified epithelium that forms a ventricular zone. From the neural ectoderm arise neuroblasts, which are precursor cells located in the ventricular zone. Neuroblasts divide and produce a lineage of daughter cells that migrate into the developing cortex to form cortical neurons. In humans, cortical neurons are born from around gestational day 40 until day 125.

The first postmitotic neurons leave the neuroepithelium and accumulate beneath the pial surface to form the preplate. This is split by later-generated neurons, the cortical plate, into a superficial, marginal zone (future layer I) and a deep subplate zone. As cortical plate neurons are born, they populate layers VI through II of the adult cortex, between the marginal and subplate zones.

Normally, the six cortical layers are generated in an orderly sequence, with first-born neurons residing in the deeper cortical layers, and later-born neurons residing in more superficial layers. One early decision a neuron faces is to which cortical layer to belong. This decision is reached before the final cell division producing the daughter neuron, and is based on cues in the neuron's microenvironment, including previously generated cells. This was shown experimentally by transplanting younger cortex making deep-layer neurons into older cortex making superficial-layer neurons. If neuroblasts are transplanted before daughter neurons are born, the daughter neurons become superficial-layer neurons, consistent with the local environment. However, if the progenitor neuroblasts are transplanted closer to the daughter neurons' birth, the daughter neurons migrate to deep cortical layers, true to their original laminar fate. Laminar identity is probably marked molecularly, and this marker may be used during later neuronal migration and target finding.

The microenvironment may also provide other information about a neuron's fate, such as what kind of cortex to become. For example, limbic cortex is marked by a protein that distinguishes it from other cortical areas. Very soon after neurogenesis, the cells destined to form the prefrontal, cingulate, perirhinal, and hippocampal cortices are marked on their surfaces by a specific molecule, called *limbic system-associated membrane protein* (LAMP). The neocortex as a whole is also distinct from other forebrain regions in its expression of specific regulatory genes. Later differences in neocortical areas, like whether to become visual cortex or sensory cortex, are likely to be determined by epigenetic factors and afferent inputs.

Other localizing information, which may guide afferent ingrowth, seems to be provided around neurogenesis. Positional information in the neuroepithelium is probably imparted to preplate cells around the time they are generated; this information is used in the subplate to control the targeting of ingrowing thalamocortical axons, at

least in the neocortex. Abnormalities in the assignment of laminar identity or cortical area markers could result in disordered cortical layers or in the formation of incorrect afferent and efferent connections.

Neuronal Migration Once neurons are born in the ventricular zone, they migrate past earlier born neurons to assume their final laminar position. The formation of the six cortical layers is complete between gestational age 26 and 29 weeks. To reach their laminar location, neurons migrate along radial glial fibers that stretch from the ventricular to superficial surface, a journey that may take place over tens of millimeters. Neurons must travel through a complex, rapidly expanding zone containing afferents from the thalamus and other cortical areas (Fig. 1.3-2).

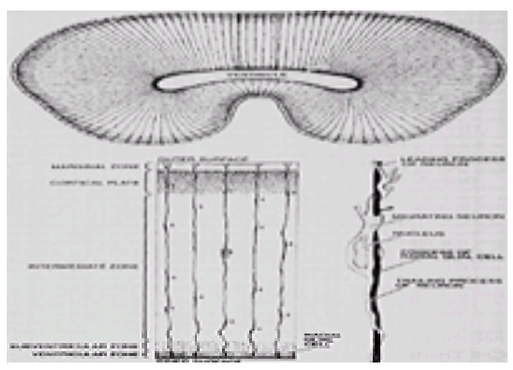

FIGURE 1.3-2 To form the cortex, neurons are born in the ventricular zone and migrate along radial glial cells to assume their position in the correct cortical layer. The upper figure clarifies the relationship between the ventricular surface, where neurons are born, and their final destination in the cortical plate. More detail is provided in the lower left drawing. This figure labels the ventricular zone, the large expanse of the intermediate zone through which the neuron migrates, and the final developing cortical layers in the cortical plate. The drawing on the lower right demonstrates the morphology of a migration neuron as it climbs along a radial glial cell through the intermediate zone. (Reprinted with permission from Maxwell Cowan W: The development of the brain. Sci Am *241:*124, 1979.)

During migration, there appear to be transient synapses and expression of neurotransmitters and neuromodulators, suggesting a chemical interaction between the migrating neurons and the embryonic environment, which influences the rate of neuronal migration. Also, contact interaction between migrating neurons and the surfaces of neighboring cells plays a crucial role in selecting migratory pathways, and in choosing to detach from the radial glial fiber and stop migrating. Stopping migration is probably an active process, and may depend on cell surface markers conferring laminar identity. The vast majority of neurons find the correct position; the small percentage of neurons that migrate to the wrong location mostly degenerate during the later phase of naturally occurring cell death.

Establishing Connections After finding the correct layer comes the intricate decision to establish connections between different brain regions. This process occurs between subcortical and cortical areas, within cortical areas, and between cortical areas. The final fate and functioning of neocortex are very dependent on the connections it establishes. How does a neuron extend its axon through the complex embryonic environment to find its correct target?

The leading edge of the axon, the growth cone, has an array of molecules on its surface. It interacts with molecules on other cells, axons, and the extracellular matrix in the embryonic environment. The molecules in the growth cone's environment are differentially adhesive or repellent. They cause the growth cone to grow where there are adhesive molecules and to avoid repulsive molecules that cause the growth cone to collapse. To choose among the number of possible pathways in the embryonic environment, the growth cone samples among the choices by means of its filopodia, and grows along specific pathways to which it is adherent or does not encounter repulsive molecules. There also appear to be diffusible gradients towards which some growth cones grow.

To illustrate the complexity of pathfinding, axons from the lateral geniculate nucleus, the visual thalamus, must navigate long distances to find the visual cortex, and then grow specifically into layer IV to synapse. Thalamic afferents presumably select their cortical targets based on molecular information on target cells or in the extracellular matrix. The subplate plays a crucial role during development, allowing afferents from thalamic neurons to wait under the developing cortex until layer IV neurons are born and migrate to their final location. In the absence of the subplate, ingrowing thalamic afferents grow past the visual cortex and fail to find their appropriate target.

Finding the appropriate cortical layer is likely to depend on laminar addresses conferred when neurons are born. For example, just as thalamic afferents to the visual cortex find layer IV, axons of layer III cortical neurons bypass layer IV to make specific connections with layer V. Even if layer V neurons are transplanted to foreign locations, they still send their axons to appropriate targets.

The molecules guiding pathway selection are being identified and characterized rapidly. One of the first discovered was the neural cell adhesion molecule (NCAM). NCAM is one of a family of molecules that mediates cell-cell or cell-substrate adhesion, and is found in many parts of the developing nervous system. Deficits in NCAM expression result in subtle cytoarchitectural abnormalities in specific brain areas. Similarly, growth-associated protein (GAP-43) is a molecule that plays a key role in guiding axon growth and modulating new connections. If GAP-43 is overexpressed, aberrant extra connections are formed in the hippocampus and other areas of the central nervous system and the peripheral nervous system. Conversely, in the absence of GAP-43, grossly abnormal connections are established.

However, there are molecules that provide more specific pathway selection information. The LAMP molecule, which marks the identity of limbic and prefrontal cortices, is needed for the formation of axon pathways between neurons in the prefrontal, cingulate, and limbic areas. It is also required for the thalamus to send correct inputs into these cortical areas, completing a network linking cortical and subcortical structures.

Cell Death Once cortical neurons are created and assume their connections, there is a period of naturally occurring cell death (*apoptosis*) in widespread areas. Apoptosis is a complex cellular process, including the expression of specific gene sequences, that ultimately leads to deoxyribonucleic acid (DNA) fragmentation and nuclear dissolution. It is distinctly different from necrosis in that cells are not damaged by external agents but die after the activation of an internal program. Also, apoptosis does not produce an inflammatory response like necrosis does. Neurons that make an adequate number of appropriate connections do not express the genes leading to apoptosis, presumably because of the trophic factors they receive from their connections, which prevent expression of the apoptotic program.

Apoptosis occurs mostly during the second half of gestation, and may result in the loss of 25 to 40 percent of the neurons in different cortical layers. The extent of cell death is controlled by influences from the neuron's synaptic targets, afferent input to the neurons, and local glial-derived and extracellular matrix molecules. If the amount of target or afferent input is diminished, more cell death results; similarly, if target or afferents are increased, the number of surviving neurons increases. Notably, even in adults neurons continue to depend on their efferent and afferent connections both for their survival and the maintenance of normal morphology and biosynthetic events.

Neuronal Differentiation and Neurotransmitter Selection The cues governing final neuronal phenotype and neurotransmitter selection also arise from multiple sources. These include previously generated neurons, glia, the extracellular matrix, various trophic factors, growth factors, and local neurotransmitters. One critical factor in neuronal differentiation appears to be cortical afferents. The differentiation of neurons in layers VI, V, and IV coincides with the ingrowth of thalamocortical fibers, that of layer III with the arrival of interhemispheric fibers, and that of layer II with the arrival of cortico-cortical fibers. Postsynaptic targets also produce differentiation factors, which alter gene expression in the presynaptic neuron and influence the choice of neurotransmitter and neuropeptides synthesized. A neuron's target can even influence which other cells connect with the neuron's dendrites. Because differentiation factors can be communicated through efferent and afferent connections, these factors regulate neuronal phenotype very precisely. They can influence the functioning of a limited number of neurons linked together in a network, without causing changes in the large numbers of nearby cells. Even in postembryonic, functional neurons, normal fluctuations in neuronal activity or hormone levels or an insult to the system can alter transmitter and neuropeptide expression.

Implications for Psychiatry What happens when the developmental plan goes awry? Aberrations in the formation of neural ectoderm or in the formation of neuroblasts are likely to result in gross abnormalities like anencephaly. A generalized failure of the migration of daughter neurons into cortical layers is seen in a lissencephalic brain, characterized by an agyric (smooth) cerebral surface. Mutations in certain cell adhesion molecules affect neural migration or axonal outgrowth, and are associated with inherited hydrocephalus. Such gross pathology is unlikely to result in an illness appreciated as psychiatric because children with these

disorders have severe mental retardation and neurological syndromes. Abnormalities of neural migration have been implicated in radiation exposure, fetal alcohol syndrome, epilepsy, reading disorder, autistic disorder, and schizophrenia. A relatively well-researched example is the radiation exposure during Hiroshima and Nagasaki, which caused a disruption of neural migration in fetuses exposed during midgestation (weeks 10 to 17). Postmortem examination showed massive heterotopia, attenuation of superficial cortical layers, and a reduction in cortico-cortical connections; epilepsy and mental retardation were frequent medical sequelae. However, given the intricate interplay of molecules and cells during development, the potential for even more subtle aberrations in cortical development and connections abounds.

Localized Abnormalities If abnormalities develop within certain neural networks, specific behavioral, cognitive, or mood symptoms recognized as a psychiatric disorder could conceivably result. Examples of localized abnormalities abound in developmental neurobiology. If monkeys are enucleated in utero, developing visual cortex diminishes in size, and surrounding cortical areas show abnormal cytoarchitecture; even localized neuronal morphologies and neurotransmitter receptor patterns show changes. Efferent connections can also be altered by abnormal afferent connections. If somatosensory input from the thalamus is routed to and synapses with visual cortex, primary visual cortical neurons retain a projection to the spinal cord, acting like somatosensory cortex.

Such subtle developmental anomalies are being discovered in many animal species. In the fruitfly, when a cell surface adhesion molecule is deleted genetically, the fly develops a characteristic set of behavioral dysfunction. The fly can walk and jump, but not fly; it has abnormal visual orientation and drinking behavior. This suggests that the molecule is crucial for the development of neural circuits controlling a subset of behavior.

There are mutant mice strains characterized by abnormal neural migration localized to the hippocampus. The hippocampus is divided into areas such as CA1, CA2, and CA3. The neurons of the hippocampus are generated during characteristic prenatal intervals in the ventricular zone lining the lateral ventricle. As in other cortical areas, the neurons migrate along primarily radial glial fibers to reach their final location. The migratory path followed by neurons destined for areas CA1 and CA2 is fairly direct. The path followed by neurons destined for area CA3, in contrast, is tortuous and lengthens during the migratory process because of hippocampal growth. One autosomal-dominant mutation, the hippocampal lamination defect mutation, affects only late-generated pyramidal cells destined for area CA3. In this mutation, this very discrete population of pyramidal cells stops in its migration and resides in an ectopic position below earlier-generated cells. This mutant demonstrates that heterotypic neuronal lamination can be limited to certain cell types in specific cortical areas.

Kallman's syndrome, an inherited human disorder characterized by hypogonadism and anosmia, is thought to result from the lack of a substrate adhesion molecule. During normal development, olfactory and gonadotropin-releasing hormone neurons migrate along a common pathway. In Kallman's syndrome this migration appears to be arrested, possibly because of a failure in neuronal interaction or synaptogenesis of olfactory neurons and their target in the olfactory bulb. Neuroanatomical findings include aplasia of the olfactory gyri and absent olfactory tracts and bulbs, which implies that an abnormality of a specific molecule causes dysfunction in discrete neuronal areas.

Neurodevelopmental Hypothesis of Schizophrenia Although future investigations may suggest that the seeds of other psychiatric illnesses are sown during development, the implications that abnormal brain development has for psychiatry are best illustrated in schizophrenia. The etiology of schizophrenia has been explored using in vivo neuroimaging, and postmortem neuropathological examinations. The neurodevelopmental hypothesis states that schizophrenia results from abnormal brain development, which manifests characteristic symptoms during early adulthood.

One dysfunctional neural network in schizophrenia links the association cortices of the frontal, parietal, and temporal lobes and the limbic cortex and subcortical structures. This network contains cortices that are the most evolutionarily advanced, and are critical to executive functioning, memory, and attention (Fig. 1.3-3) functions that are particularly impaired in persons with schizophrenia. Because these cerebral areas are well defined, it is possible to investigate their integrity, both with in vivo neuroimaging techniques and via postmortem neuropathological examinations.

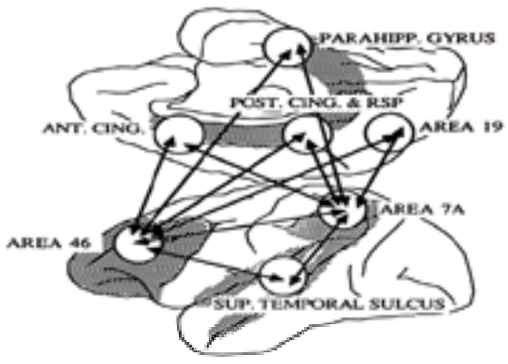

FIGURE 1.3-3 Cortical and subcortical areas are physically connected into discrete neural networks that subserve distinct aspects of cognition, affect, and behavior. A crucial goal of development is the establishment of the correct neural connections to form these networks. This drawing is a simplification of the multiple pathways connecting the frontal, parietal, and temporal cortices, areas involved in attention and memory, and implicated in schizophrenia. The hatched areas reveal inputs from one thalamic nucleus. (Reprinted with permission from Goldman-Rakic PS: Parallel distributed networks in primate association cortex. Annu Rev Neurosci *11:*143, 1988.)

NEUROIMAGING Imaging the brain with magnetic resonance imaging (MRI) allows for volumetric measurements of brain structures. In schizophrenia, reductions on the order of 10 to 15 percent are reported in overall temporal lobe size, in temporal lobe gray matter, and in specific mesial and lateral temporal lobe structures. Recent MRI studies suggest subtle volumetric reductions in widespread cortical areas, including the frontal and parietal secondary association areas, and in the thalamus.

Neuroimaging with single photon emission computed tomography (SPECT) and positron emission tomography (PET) can investigate the functioning of these brain regions in vivo. By imaging subjects during tasks, cerebral activity patterns reflect the functioning of neural networks necessary to perform the tasks. In numerous studies individuals with schizophrenia show dysfunction within frontal-parietal-temporal networks, even during different tasks that utilize these cerebral areas. An example is the Wisconsin Card Sorting Test, an abstract problem-solving test requiring attention and working memory. Monozygotic twins discordant for schizophrenia underwent PET blood flow scans while performing this test. In all but one pair the ill twin had relatively decreased prefrontal cerebral blood flow; the mesial temporal lobe limbic region in the ill twin was invariably hyperactive. In vivo studies do not speak to when or how the functional abnormality arose. However, finding functional consequences in cerebral areas defined as abnormal in neuropathological investigations is essential to correlate neurodevelopmental abnormalities with the clinical symptoms of schizophrenia.

NEUROPATHOLOGY Postmortem morphometric studies of brains from patients with schizophrenia are consistent with in vivo imaging studies. Schizophrenia is associated with enlarged ventricles and apparently focal decreases in the size of mesial temporal lobe structures, including the amygdala, hippocampus, parahippocampal gyrus, and entorhinal cortex. Similarly, reduced neuronal counts and decreased neuropil have been reported in selected cortical and periventricular regions. One crucial observation is the lack of gliosis in schizophrenia. A proliferation of glial cells is seen in most degenerative brain conditions and encephalopathies that arise after birth. This suggests that whatever causes the brain abnormalities in schizophrenia does so before the third trimester of gestation, when glial cells become responsive to injury.

Of considerable interest in understanding the developmental neurobiology of schizophrenia are studies exploring cortical cytoarchitecture, particularly in the networks showing in vivo dysfunction. In the entorhinal cortex of individuals with schizophrenia, heterotopic groups of neurons belonging to layer II are found displaced into layer III. This may indicate abnormal neuronal migration that results in heterotopic neuronal islands, and abnormal cytoarchitecture; unfortunately, not all studies replicate this finding. However, other investigators report reduced numbers of small neurons and abnormal neuronal aggregates in the anterior cingulate. Layer II of the prefrontal cortex of persons with schizophrenia shows reduced numbers of small neurons and higher densities of pyramidal neurons in layer V.

A neuron-specific stain for nicotinamide-adenine dinucleotide phosphate-diaphorase (NADPH-d) has been used to study brains from patients with schizophrenia. NADPH-d-staining cells are seen in cortex and in subcortical white matter and are likely to be the remnants of subplate neurons. In the prefrontal and lateral temporal

lobe cortex of some patients with schizophrenia these neurons are decreased in the cortex and its subjacent white matter, but are present in abnormally high numbers in deeper white matter. This is consistent with the hypothesis of aberrant neuronal migration in schizophrenia.

CONSEQUENCES OF ABNORMAL CYTOARCHITECTURE The number of studies exploring the extent, type, and location of cytoarchitectural abnormalities in schizophrenia is limited. Nevertheless, each positive finding implicates a failure of neurons assuming the correct laminar location, a process that occurs during the second trimester of intrauterine brain development. This could reflect a defect not only in migration, but in the ability to attract the correct afferents or make normal efferent connections.

If one area of cortex is abnormal, there will be ramifications in the organization and function of cortical and subcortical areas from which it receives afferents and to which it sends efferents. Because cortical areas depend on their afferent and efferent connections to survive cell death, the reports of reduced cell numbers in schizophrenia could be due to the effects of disordered cortex elsewhere. This is consistent with the subtle cortical volume reductions found in MRI studies of schizophrenia.

Not surprisingly, multiple postmortem biochemical and molecular abnormalities are also associated with schizophrenia. These include mesial temporal lobe abnormalities of choline acetyltransferase, glutamate decarboxylase, glutamate, aspartate, serotonin, neurotensin, and substance P. Abnormalities in receptor subtypes include k-opioid, b-adrenergic, serotonin (5-hydroxytryptamine [5-HT]) type 1A [HT_{1A}], $5\text{-}HT_2$, g-aminobutyric acid (GABA), as well as alterations in glutamate reuptake sites and adenylate cyclase. In the prefrontal cortex, reported abnormalities include serotonin reuptake sites and met-enkephalin, $5\text{-}HT_{1A}$ and $5\text{-}HT_2$ muscarinic cholinergic, glutamate and $GABA_A$ receptors.

Molecules crucial to normal brain development and postnatal plasticity are being investigated in schizophrenia. In brains of persons with schizophrenia abnormalities of NCAM have been reported. GAP-43 is essential for the initial establishment and reorganization of synaptic connections, and remains high throughout life in the limbic system and neocortex, areas involved in the processing and storage of information. In frontal cortex and visual association areas of persons with schizophrenia, levels of GAP-43 protein are increased, which is perhaps a reflection of the plasticity of inputs to prefrontal cortex. GAP-43 messenger ribonucleic acid (mRNA), a measure of intrinsic prefrontal efferent plasticity, is reduced. Levels of synaptophysin also are reduced; synaptophysin is a synaptic–vesicle-associated protein involved in neurotransmitter release that is used as a marker of presynaptic terminal density. Reduced synaptophysin mRNA concentrations are also found in specific regions of the mesial temporal lobe. Although the precise mechanisms controlling these molecular changes remain unknown, the results suggest abnormal synaptic connectivity.

If abnormal brain development causes schizophrenia, why does onset of symptoms occur in late adolescence or early adulthood? Frontal–lobe-linked networks, which appear to be dysfunctional in schizophrenia, undergo substantial postnatal maturation. Schizophrenic symptoms may not appear until a critical network is tested at a particular time in postnatal maturation. However, it should be noted that individuals destined to develop schizophrenia showed subtle but significant delays in early-childhood milestones like walking and speech development. This suggests that abnormalities are present virtually from birth, but that their clinical manifestations may vary over time.

EARLY POSTNATAL DEVELOPMENT

The early postnatal years are marked by a rapid maturation of cognitive, social, and behavioral abilities as infants progress from helplessness to autonomy, and children and adolescents develop more sophisticated ways of thinking. The information and abilities acquired by infants, children, and adolescents are staggering. The impact that these early years have on personality development and behavior is profound, longlasting, and at times, refractory to treatment interventions. The consequences of physical or emotional childhood trauma are seen in every psychiatric practice. For such dramatic cognitive, behavioral, and emotional changes, there must be an underlying neurobiological substrate. Neuroscience is exploring the structural and functional foundations of normal postnatal maturation and how it is impacted by the environment.

Postnatal Cortical Maturation The number of cortical synapses changes dramatically in early postnatal life. Newborn monkeys have approximately the same number of synapses as adult monkeys do. Synaptic density increases in the first few months of life until it is about 40 percent higher than adult values. Synaptic density remains high until adolescence when it decreases to adult levels, and then remains fairly constant throughout life. The pruning of cortical synapses appears to involve primarily excitatory connections. Local circuit interneurons, which are GABAergic, appear to maintain stable synapses from childhood into adulthood. This suggests that exuberant excitatory synaptic connections are selectively remodeled into adultlike connectivity by experience.

The maturation of neuronal structure was investigated by a postmortem examination of prefrontal cortex from 10 weeks' gestational age through adulthood. Neurons were stained, and their location and morphology was studied. Distinct cortical areas and layers assume adult morphology at different rates. The subplate has disappeared by birth in the visual and somatosensory cortices, but is present in the prefrontal cortex of the newborn and gradually declines during the first 6 months of life. It may continue to serve a role in the development of postnatal prefrontal cortico-cortical projections.

Many neuronal types alter their shape postnatally, corresponding to altered synaptic connections. In several cortical layers in the prefrontal cortex and other association cortices, pyramidal neurons continue dendritic outgrowth and spine formation throughout the first 2 postnatal years. Pyramidal neurons in layers III and V of the prefrontal cortex may not develop adultlike dendritic fields until adolescence. The pyramidal neurons of layer III are a major source of cortico-cortical connections, and they show intensive spine growth during early childhood, possibly as targets of thalamocortical and cortico-cortical fibers. Similarly, the interneurons that modulate pyramidal neurons show postnatal changes.

Implications for Psychiatry The continued maturation of prefrontal cortex in early life offers ample opportunities for childhood experience to permanently shape the cortex that controls thought, behavior, and mood. At a gross level, if the processes governing synaptic remodeling and maturation go awry, severe conditions can occur. Fragile X syndrome, the second most common inherited form of mental retardation, is associated with abnormal synapses. Individuals with fragile X show thin, elongated dendritic spines with smaller synaptic contacts, much like the appearance of immature synapses during normal neocortical development. However, environmental input to otherwise normal cortex can also produce dramatic changes in cortical structure and function.

Environmental Effects on Cortical Connections During early life the cortex is fine-tuning its connections dependent on patterns of neural activity caused by environmental input. An example is human infants with congenital cataracts. The cataracts must be removed before 4 to 6 months of age, or permanent visual impairment results. Also, if strabismus is not corrected by about 7 years of age, the squinting eye is permanently visually handicapped. The timing of these critical periods corresponds with the normal timing of exuberant synapse elimination. In the human visual cortex, synapse elimination begins at about 6 months and is complete by age 6 or 7 years.

The impact of left-hemisphere damage on language development reveals an analogous process. If the brain damage occurs before 8 years of age, language development may recover and seem normal; after 8 years of age, aphasia results. These clinical examples illustrate that the availability of exuberant synapses facilitates functional plasticity. Similarly, normal cortical language representation can be visualized using functional MRI. Individuals who learned two languages during infancy show similar cortical activity when exposed to either language. In contrast, individuals who learned a second language in early adulthood show distinct Broca's area representations of the native and second languages. This illustrates how cortical organization and plasticity is dependent on the time of the environmental influence. However, to more fully understand the mechanisms underlying this plasticity requires experimental animal models.

David Hubel and Torsten Wiesel described the impact of environmental manipulations on the visual system in cats and monkeys. Analogous to humans, there is a critical period during visual cortical development during which the cortex is exquisitely sensitive to changes in environmental input. Once the critical period passes, cortical abnormalities are irreversible. In contrast, adult visual cortex is much less sensitive to environmental input; a year of monocular lid closure in an adult cat leads to no detectable cortical effects. If one eyelid is sutured during the critical period, there is little visual input from that eye; if the eye is opened later, vision remains permanently impaired. This is reflected cortically by reduced representation from the sutured eye and expanded representation from the open eye (Fig. 1.3-4). In the lateral geniculate nucleus, the synaptic waystation between the retina and visual cortex, there is profound atrophy in the geniculate layers receiving input from the covered eye. In the visual cortex, the terminals from visually deprived afferents are smaller, have fewer mitochondria, and make immature-appearing synapses. GAP-43, a molecule implicated in axonal connectivity, has its highest expression in primary visual cortex during the critical period as synapses are remodeled.

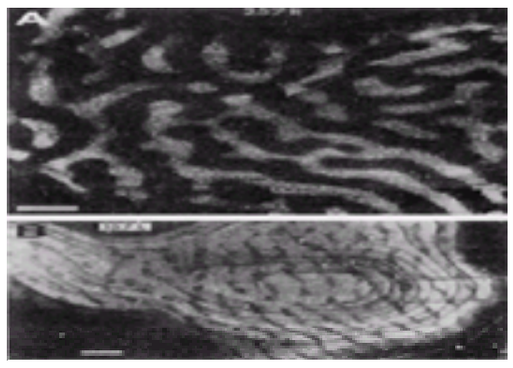

FIGURE 1.3-4 Early postnatal environmental inputs dramatically alter cortical structure. The top picture **(A)** shows an autoradiograph from layer IV of the primary visual cortex of a normal monkey. The light areas reveal inputs from one eye; the dark bands are inputs from the other eye. Equal cortical representation from both eyes is evident. Picture **B**, an autoradiograph from a monkey whose right eye was closed during infancy, illustrates what happens after monocular deprivation. Note the marked expansion of the light areas, and the shrinkage of the dark bands. The light areas are the inputs from the eye that was open during the experiment. (Reprinted with permission from LeVay S, Wiesel TN, Hubel DH: The development of ocular dominance columns in normal and visually deprived monkeys. J Comp Neurol *191:*11, 1980.)

If kittens are deprived of visual input altogether, compensatory changes are seen in the visual areas; there are now neurons that respond to auditory and somatosensory stimulation. Within altered cortex there are alterations in the concentration of serotonin, dopamine, norepinephrine, glutamate, and acetylcholinesterase. This demonstrates that the morphological synaptic, and molecular plasticity of the critical period is crucial for normal cortical connectivity and functioning.

Strabismus has also been experimentally induced in cats and monkeys by unilateral ablation of one extraocular muscle. Although these animals are born with a normal CNS, the abnormal visual input and neural activity induced by strabismus causes structural CNS changes. The morphology of retinogeniculate axons is altered, with smaller arbors and fewer boutons compared to normal. The lateral geniculate nucleus contains smaller neuronal cell bodies. The anatomical organization of the visual cortex is also altered with a reduction in the proportion of cells that respond to binocular input and a loss of the normal orientation selectivity.

Early environmental manipulations may also impact higher cognitive functioning. This has been elegantly explored in a series of experiments investigating monkey visual recognition memory. Visual recognition memory is the ability to see an object and to consciously remember having seen it before. This capacity requires the interaction of visual pathways and the limbic system. In adult monkeys, area TE is critical for this interaction. Area TE is a cortical region on the lateral temporal lobe that receives projections from multiple visual areas in prestriate cortex, and sends direct and indirect projections to the amygdala, perirhinal cortex, and hippocampus. Adult lesions in area TE significantly impair visual recognition memory.

In contrast, infant lesions in area TE leave visual recognition memory intact. This sparing of function is explained by the enhanced plasticity of the immature brain that results from the redundancy of connections. In infants, normally transient projections distribute memory functions throughout several visual association areas. During normal maturation, visual memory becomes more localized to area TE as projections to other areas retract. In animals with TE lesions during infancy, these immature projections are maintained. Moreover, visual areas not normally involved in recognition memory take over that function after early TE lesions. These findings offer evidence for substantial remodeling of cognitive neural networks in response to early environmental manipulations. Additionally, the observation that a specific set of neural connections involved in memory processing normally regresses after infancy has potentially fascinating implications for understanding phenomena such as infantile amnesia.

Critical Periods for Cognition and Emotion The importance of critical periods does not apply solely to visual and language cortices. It has stunning implications for how early childhood experiences can leave brain traces that affect brain function and behavior throughout adult life. Just as visual cortex is shaped by experience, neural systems subserving cognition and affect may be equally impacted. For example, animals reared in complex environments have a greater width of the cortical mantle with increased dendritic arborizations, increased numbers of synapses per neuron, and longer postsynaptic densities compared to animals reared in standard cages. Early postnatal environmental manipulations result in long-lasting changes in hippocampal and prefrontal cortical glucocorticoid receptors, which may permanently alter an animal's stress response.

Early brain manipulations can also have delayed effects on the regulation of neurochemical systems implicated in adult-onset psychiatric illness. For example, scientists have recently shown that neonatal rat pups who undergo a hippocampal lesion during the first week of life appear relatively normal during their equivalent of childhood. However, during early adulthood they manifest dramatically abnormal behavioral responses to environmental and pharmacological stresses. Antipsychotic medications ameliorate some of these abnormal responses. The abnormalities appear to be modiated by rewiring of the connections involved in the regulation of mesolimbic dopaminergic function. Rats with this developmental lesion have been studied as a potential animal model of a number of phenomena associated with schizophrenia.

Behavioral studies in monkeys have shown that early childhood experience can have a profound impact on adult adaptation, especially in the context of social stress. When infant monkeys are removed from their mothers, even for relatively brief periods, their tolerance for stress during adulthood may be reduced. Moreover, the genetic tendency of an animal to be stress sensitive interacts with maternal separations. If a monkey is innately hyperresponsive to environmental stress, childhood maternal separation leads to exaggerated stress responses as an adult. Recently, the molecular consequences of early stress are being identified. In rats, variations in maternal behavior predict alterations in the expression of synaptic markers in the brain of adult animals. Thus, animal studies suggest that early environmental influences have a permanent impact on underlying cerebral structure and function.

In humans, there is evidence that early environmental stimulation, even before preschool, improves learning with an effect that lasts for years. Preliminary work suggests that children who have survived global environmental neglect may have underdevelopment of cerebral structures. Similarly, children exposed to chronic trauma have behavioral impulsivity, cognitive distortions, and difficulties with cognitive organization. These children also evidence an array of physiological abnormalities, including hyperarousal, an increased startle response, sleep difficulties, and affect-regulation problems. The plasticity of postnatal cerebral systems provides a neurobiological explanation for the major impact that childhood experiences have on adult functioning. It also offers a substantial opportunity for early childhood therapeutic interventions in vulnerable populations.

NEURONAL PLASTICITY

Synaptic plasticity is a property of adult as well as developing or young cortex, and reflects how synaptic strength changes with experience. Its relevance to psychiatry is seen in the course of the illnesses psychiatrists treat. Clinical research supports the notion that psychiatric illnesses progress and become more refractory to treatment over time. This has been demonstrated most clearly in bipolar disorders and schizophrenia. The expression or severity of an illness changing over time implies an underlying change in the neurobiology of the illness. Neuroscience studies of learning and memory have helped to illuminate the plasticity of adult cortex, which can be used as a blueprint for brain changes associated with psychiatric illnesses. What evidence is there for structural brain changes with learning?

Cortical Remodeling Human functional neuroimaging studies demonstrate changes in neural activity patterns as a behavior or a response is learned. In nonhuman primates, this can be investigated by recording the firing pattern of cortical neurons. When an adult monkey learns a task, its behavior reflects alterations of neuronal firing patterns distributed in the cortical regions involved in the task. An altered neuronal firing pattern can lead to cortical remodeling; for example, if an adult monkey attends to a tactile stimulus to its finger, the cortical representations from that digit increase.

How can neuronal firing patterns lead to a remodeled cortex? Short-term changes, or memory, result from the strengthening of existing synapses. This occurs via the covalent modification of existing proteins after the activation of second-messenger systems. Long-term memory requires neuronal gene expression and protein synthesis, resulting in the growth of new synaptic connections (Fig. 1.3-5).

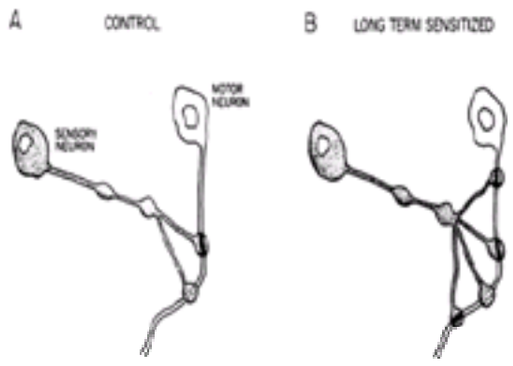

FIGURE 1.3-5 Neurons show plasticity associated with learning and experience. This drawing is of identified neurons from the sea slug *Aplysia*. The sensory neuron and motor neuron are responsible for a behavior that can be observed. The control panel **(A)** represents the synaptic connections in an untrained animal. After learning trials the number of synaptic connections increases, as demonstrated in **B**. (Reprinted with permission from Kandel ER: Genes, nerve cells, and the remembrance of things past. J Neuropsychiatry Clin Neurosci *1:*118, 1989.)

Long-Term Potentiation (LTP) Learning is being actively explored in mammals, using a model called long-term potentiation in which a prolonged excitatory stimulus delivered to presynaptic hippocampal neurons leads to a long-lasting increased response in postsynaptic neurons. Somehow, the postsynaptic neuron senses the coincidence between its own and presynaptic activity and sends a signal back to concurrently active presynaptic inputs to selectively increase their strength.

Neurotransmitters modulate the changes associated with learning and synaptic strengthening. Dopamine is a good candidate as a neuromodulator in the hippocampus, because late LTP can be blocked by dopamine type 1 receptor (D_1) antagonists. Acetylcholine and norepinephrine probably play major roles in facilitating cortical plasticity, and serotonin plays a role in invertebrate systems.

LTP involves several phases. Initially, presynaptic stimulation leads to a transient increase in postsynaptic calcium. The increase in intracellular calcium leads to a brief presynaptic increase in neurotransmitter release. However, calcium returns to resting levels within minutes; therefore long-lasting changes in synaptic strength involve additional processes in the presynaptic and postsynaptic neurons.

The signal that goes from the postsynaptic to the presynaptic neuron is still being explored, although it may involve nitric oxide. The signal activates second messengers presynaptically, which induce autonomous protein kinase activation. Protein kinases are enzymes that phosphorylate multiple substrates, altering their functioning, altering neuronal physiology, and increasing synaptic strength. When protein kinases become autonomously activated, they are independent of the second messenger. This explains how a short-lived increase in second messengers results in longer-lasting changes in synaptic strength.

Protein kinases involved in LTP include protein kinase C (PKC), calcium/calmodulin-dependent kinase II (CaMKII), the cyclic adenosine monophosphate (cAMP) dependent protein kinase (PKA), and protein tyrosine kinase (PTK). Substrates for phosphorylation implicated in long-term potentiation include GAP-43 for PKC, the glutamate receptor by CaMKII, and synaptophysin, the nicotinic acetylcholine receptor, and the glutamate receptor by PTKs. Phosphorylating GAP-43 may increase presynaptic neurotransmitter release; it also is associated with inducing a neuronal state of growth and synaptic sprouting.

Following activation of protein kinases, LTP depends on RNA transcription and protein synthesis and cAMP plays a central role in this process. cAMP concentrations increase in LTP due to calcium and calmodulin stimulation of adenylyl cyclase, the enzyme that produces cAMP. Additionally, D_1 and D_5 receptors are coupled to a G protein that increases cAMP concentration when activated.

cAMP concentration is crucial to LTP because of its effect on gene transcription. There is a cAMP response element (CRE), which when phosphorylated activates gene transcription. The mRNAs of several CRE-linked genes, including the immediate early gene c-*fos,* increase during long-term potentiation. The synaptic remodeling underlying learning depends on gene expression.

Learning and Gene Expression Immediate early genes (IEG) encode nuclear regulatory proteins that cause stable alterations in the transcription of late genes. Triggering IEG expression can lead to a cascade of gene transcription and synaptic remodeling. In addition to LTP IEGs are implicated in more behaviorally relevant learning paradigms. One-day-old chicks learn to avoid a bitter-tasting bead after a single training exposure. Training induces the IEGs c-*fos* and c-*jun* in certain brain regions. After c-*fos* and c-*jun* induction, there is enhanced synthesis of a variety of proteins, and increased fucosylation of membrane glycoproteins, including NCAM. Structurally, training results in a 60 percent increase in the density of dendritic spines, increases in synaptic number, and a 60 percent increase in the numbers of synaptic vesicles per synapse. The net result is synaptic remodeling, with active synapses showing increased numbers of boutons.

GAP-43 is another molecule important in adult cortical plasticity. In the neocortex of adult humans and nonhuman primates, GAP-43 remains present in limbic and associative areas, suggesting that these areas continue to undergo structural changes associated with learning and memory. Even when networks are formed, the precise connections within them remain responsive to environmental input. The guiding principle, both in final synaptic selections in development and in adult cortical plasticity is summed up in the adage “neurons that fire together wire together.”

Implications for Psychiatry Because psychiatric illnesses are thought to be associated with a pattern of neuronal activity, this activity may result in a type of learning that strengthens activity patterns representing the ill state. If the principle “neurons that fire together wire together” operates in psychiatric illnesses, this predicts that, if untreated, psychiatric illnesses would somehow become more entrenched neurobiologically and perhaps become more refractory to treatment. Alternatively, episodes of illness could become more frequent or more severe.

Progression of Illness In schizophrenia, the longer a person goes without seeking treatment, the more refractory the illness becomes, requiring more time on and higher doses of medication before symptoms remit. Similar observations are true of mood disorders. Most patients who have one episode of a mood disorder will have repeat episodes, and, especially in bipolar disorder, the intervals between episodes may become shorter. Although the initial episodes of mood disorders are often associated with stressors, later episodes can arise spontaneously. These observations support the hypothesis that the neurobiological underpinnings of these disorders are dynamic, changing with time and as a result of cortical remodeling.

Two experimental models have been used to explain the progression of bipolar disorder. One is kindling, in which electrical current in the amygdala induces seizures in rats. After a number of current-induced seizures, spontaneous epilepsy develops in the absence of stimulation. The second is stimulant-induced sensitization, where an animal shows greater behavioral changes to a consistent amount of stimulant after repeated exposures. In both these models, the environmental manipulations rapidly activate neurotransmitter pathways, and over the long term, also activate gene expression, including of *c-fos*. The abnormal neural activity induced by seizures results in axonal growth and synaptic reorganization in limbic structures, analogous to the changes seen in memory.

Treatment Implications The clinical implications of these models are critical. Neuroscience predicts that preventing subsequent episodes of a major psychiatric illness will positively alter the course of the disorder. This argues that pharmacological prophylaxis is essential. There is supporting evidence in this regard for schizophrenia, major depression, and bipolar disorder. However, there is ample room to speculate about parallel neurobiological processes in posttraumatic stress disorder, obsessive-compulsive disorder, panic disorder, and substance-related disorders.

Additionally, if abnormal cellular processes contribute to the etiology of major psychiatric disorders, how do pharmacological interventions normalize these processes? Are medications effective only because of their neurotransmitter receptor profiles? Or do they influence second-messenger systems, gene transcription, and protein translation, as in learning? The time course of pharmacological efficacy suggests that medications do not merely affect neurotransmitters, but also alter underlying cellular functioning. Dopamine, norepinephrine, acetylcholine, glutamate, and opiates have all been associated with changing gene expression.

Not surprisingly, antipsychotic medications induce the expression of c-*fos* and c-*jun*. In rats, the short-term administration of clinical doses of antipsychotic agents, such as haloperidol (Haldol), clozapine (Clozaril), and olanzapine (Zyprexa) induce c-*fos* in the nucleus accumbens. Antipsychotic agents that clinically cause extrapyramidal symptoms (e.g., haloperidol) also induce c-*fos* in the dorsal striatum. Clozapine, which clinically does not cause extrapyramidal adverse effects, does

not induce c-*fos* in this region. This suggests that extrapyramidal and antipsychotic effects can be dissociated in terms of their respective molecular neuroanatomies. Antidepressant treatments increase intracellular phosphorylation of cAMP-responsive proteins (e.g., CRE) that lead to transcription of specific genes. One of these genes, brain-derived neurotrophic factor, has also been implicated in neuronal plasticity and may play a role in the potential restitutive effects of treatment with antidepressant medications.

These recent observations on the intracellular effects of psychiatric medications have potentially far-reaching implications for the understanding of mental illness and approaches to their treatments. The traditional view that psychiatric medications can be characterized by their actions at synapses where they bind to cell surface proteins (e.g., receptors) is yielding to a more enlightened vision that the synapse is only the tip of the iceberg of cellular function and plasticity that is of relevance to psychiatry.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2, perception and cognition in Section 3.1, and psychopharmacology in Chapter 31.

SECTION REFERENCES

Bachus SE, Kleinman JE: The neuropathology of schizophrenia. J Clin Psychiatry *57:*72, 1996.

Barbe MF, Levitt P: Attraction of specific thalamic input by cerebral grafts depends on the molecular identity of the implant. Proc Natl Acad Sci USA *89:*3706, 1992.

Barbeau D, Liang JJ, Robitaille Y, Quiron R, Srivastava LK: Decreased expression of the embryonic form of the neural cell adhesion molecule in schizophrenic brains. Proc Natl Acad Sci USA *92:*2785, 1995.

Benowitz LI, Routtenberg A: GAP-43: An intrinsic determinant of neuronal development and plasticity. Trends Neurosci *20:*84, 1997.

Bymaster FP, Rasmussen K, Calligaro DO, Nelson DL, DeLapp NW, Wong DT, Moore NA: In vitro and in vivo biochemistry of olanzapine: A novel, atypical antipsychotic drug. J Clin Psychiatry *58*(Suppl):28, 1997.

Comery TA, Harris JB, Willems PJ, Oostra BA, Irwin SA, Weiler IJ, Greenough WT: Abnormal dendritic spines in fragile X knockout mice: Maturation and pruning deficits. Proc Natl Acad Sci USA *94:*5401, 1997.

Duman RS, Heninger GR, Nestler EJ: A molecular and cellular theory of depression. Arch Gen Psychiatry *54:*597, 1997.

Glantz LA, Lewis DA: Reduction of synaptophysin immunoreactivity in the prefrontal cortex of subjects with schizophrenia. Arch Gen Psychiatry *54:*660, 1997.

*Goodman CS, Shatz CJ: Developmental mechanisms that generate precise patterns of neuronal connectivity. Cell *72:*77(Suppl), 1993.

Jessell TM, Kandel ER: Synaptic transmission: A bidirectional and self-modifiable form of cell-cell communication. Cell *72*(Suppl):1, 1993.

Kim KHS, Relkin NR, Lee K-M, Hirsch J: Distinct cortical areas associated with native and second languages. Nature *388:*171, 1997.

Kotrla KJ: Functional neuroimaging in psychiatry. In *Textbook of Neuropsychiatry,* ed 3, RE Hales, S Cyudofsky, editors. American Psychiatric Press, Washington, DC, 1997.

Kotrla KJ, Weinberger DR: Brain imaging in schizophrenia. In *Annual Review of Medicine,* vol 46, WP Creger, CH Coggins, EW Hancock, editors. Annual Reviews, Palo Alto, CA, 1995.

Lander AD: Understanding the molecules of neural cell contacts: Emerging patterns of structure and function. Trends Neurosci *12:*189, 1989.

Lieberman J, Jody D, Geisler S, Alvir J, Loebel A, Szymanski S, Woerner M, Borenstein M: Time course and biologic correlates of treatment response in first-episode schizophrenia. Arch Gen Psychiatry *50:*369, 1993.

Liu D, Kiorio J, Tannenbaum B, Caldji C, Francis D, Freedman A, Sharma S, Pearson D, Plotsky P, Meaney M: Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. Science *277:*1659, 1977.

Maletic-Savatic M, Malinow R, Svoboda K: Rapid dendritic morphogenesis in CA1 hippocampal dendrites induced by synaptic activity. Science *283:*1923, 1999.

Marin-Padilla M: Early ontogenesis of the human cerebral cortex. In *Cerebral Cortex,* vol 7, A Peters, EG Jones, editors. Plenum, New York, 1988.

McConnell SK: The generation of neuronal diversity in the central nervous system. Annu Rev Neurosci *14:*269, 1991.

Merzenich MM, Sameshima K: Cortical plasticity and memory. Curr Opin Neurobiol *3:*187, 1993.

Milleret C: Visual callosal connections and strabismus. Behav Brain Res *64:*85, 1994.

Mrzljak L, Uylings HBM, Van Eden CG, Judas M: Neuronal development in human prefrontal cortex in prenatal and postnatal stages. In *Progress in Brain Research,* vol 85, HBM Uylings, CG Van Eden, JPC De Bruin, MA Corner, MGP Feenstra, editors. Elsevier Science, New York, 1990.

Nowakowski RS: Some basic concepts of the development of the central nervous system. In *Fetal Neural Development and Adult Schizophrenia,* SA Mednick, TD Cannon, CE Barr, M Lyon, editors. Cambridge University Press, Cambridge, England, 1991.

*O'Leary DDM, Schlaggar BL, Tuttle R: Specification of neocortical areas and thalamocortical connections. Annu Rev Neurosci *17:*419, 1994.

Oppenheim RW: Cell death during development of the nervous system. Annu Rev Neurosci *14:*453, 1991.

Patterson PH, Nawa H: Neuronal differentiation factors/cytokines and synaptic plasticity. Cell *10:*123, 1993.

Perrone-Bizzozero NI, Sower AC, Bird ED, Benowitz LI, Ivins KJ, Neve RL: Levels of the growth-associated protein GAP-43 are selectively increased in association cortices in schizophrenia. Proc Natl Acad Sci USA *93:*14182, 1996.

Perry WS, Neal R, McDowell JE, Braff DL: Schizophrenia and frontal lobe functioning: Evidence from neuropsychology, cognitive neuroscience, and psychophysiology. In *The Human Frontal Lobes: Functions and Disorders,* BL Miller, JL Cummings, editors. Guilford, New York, 1999.

*Post RM: Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. Am J Psychiatry *149:*999, 1992.

*Rakic P: Development of the cerebral cortex in human and nonhuman primates. In *Child and Adolescent Psychiatry. A Comprehensive Textbook,* ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Roberson ED, English JD, Sweatt JD: A biochemist's view of long-term potentiation. Learning Memory *3:*1, 1996.

Rose SPR: Cell-adhesion molecules, glucocorticoids and long-term memory formation. Trends Neurosci *18:*502, 1995.

Rugarli EI, Ballabio A: Kallman syndrome: From genetics to neurobiology. *JAMA 270:*2713, 1993.

Toldi J, Feher O, Wolff J-R: Neuronal plasticity induced by neonatal monocular and binocular enucleation. Prog Neurobiol *48:*191, 1996.

Webster MJ, Bachevalier J, Ungerleider LG: Development and plasticity of visual memory circuits. 1-14. In *Maturational Windows and Adult Cortical Plasticity, SFI Studies in the Sciences of Complexity,* vol 24. Addison-Wesley, 1995.

Weiler IJ, Hawrylak N, Greenough WT: Morphogenesis in memory formation: Synaptic and cellular mechanisms. Behav Brain Res *66:*1, 1995.

*Weinberger DR: Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry *44:*660, 1987.

Weinberger DR, Lipska BK: Cortical maldevelopment, anti-psychotic drugs, and schizophrenia: A search for common ground. Schizophr Res *16:*87, 1995.

Wong EV, Kenwrick S, Willems P, Lemmon V: Mutations in the cell adhesion molecule L1 cause mental retardation. Trends Neurosci *18:*168, 1995.

1.4 MONOAMINE NEUROTRANSMITTERS

LAURENCE H. TECOTT M.D., PH.D.

Historically the monoamine neurotransmitters and acetylcholine have been strongly implicated in the etiology and treatment of a wide variety of neuropsychiatric disorders. The monoamines serotonin, norepinephrine, dopamine, and histamine are widely distributed throughout the central nervous system (CNS). The activity of each monoamine neurotransmitter system modulates multiple neuronal pathways that mediate diverse behavioral and physiological processes. Conversely, each CNS function is regulated by multiple interactive neurotransmitter systems. In light of this complexity, determining the mechanisms through which monoamine systems impact the etiology and treatment of psychiatric disorders poses a major challenge. Advances in molecular neurobiology provide powerful new tools to aid in this endeavor. Molecular cloning studies have led to the identification and functional characterization of gene products that contribute to monoaminergic transmission, such as monoamine receptors, transporters, and synthetic and degradative enzymes. More recently, these genes have provided targets for powerful techniques that enable the precise introduction of mutations into the mouse genome. Thus, the functional consequences of perturbing gene function may be examined in the context of the intact, behaving organism. The molecular cloning of genes involved in monoaminergic transmission has also led to the identification of allelic variants of human genes, which raises the intriguing prospect that the inheritance of particular variants may contribute to disease susceptibility and to the efficacy of therapeutic agents.

ANATOMY OF MONOAMINE SYSTEMS

The anatomical organization of monoaminergic systems shares a number of common features. Monoaminergic systems are strikingly divergent: monoaminergic cell bodies are generally found in aggregates located in a few restricted subcortical brain regions. Individual monoaminergic neurons typically possess long and extensively branched axonal processes, innervating a large number of postsynaptic cells. This organization may permit monoaminergic systems to exert control in a coordinated manner over diverse brain regions. The actions of monoamines in particular brain areas are determined not only by the extent of monoamine innervation, but also by the receptor subtypes expressed in these regions. Monoaminergic receptors are diverse with regard to their regional and synaptic localization within the brain, and to the intracellular signaling systems to which they couple. This receptor diversity provides a means through which a single signaling molecule may produce effects that vary in different postsynaptic neurons.

Serotonin Although approximately one in a million brain neurons are serotonergic, serotonin systems influence CNS activity at all levels of the neuraxis. Serotonergic neurons are clustered in midline raphe nuclei of the midbrain, pons, and medulla. These neurons project extensively throughout the brain and descend to the spinal cord (Fig. 1.4-1). The majority of the serotonergic innervation of the forebrain arises from the dorsal and median raphe nuclei of the midbrain. Ascending projections from these nuclei course through the medial forebrain bundle before diverging to many target regions. Whereas the median raphe nucleus provides the majority of the serotonergic innervation of the limbic system, the dorsal raphe nucleus provides the primary innervation of the striatum and thalamus. In addition to the differences in target areas innervated by the median raphe and dorsal raphe, structural differences in the axonal projections from these nuclei have been observed. Whereas fibers from the dorsal raphe are fine, with small vesicle-containing swellings called *varicosities,* median raphe axons are beaded, with large spherical varicosities. These axons show differential sensitivity to the neurotoxic effects of the amphetamine analog 3,4-methylene-dioxymethamphetamine (MDMA, “ecstasy”). This agent produces a selective loss of fine axons while sparing the larger beaded projections derived from the median raphe. The significance of the morphological differences in these projection fibers remains to be determined. Both types of fibers are found in the neocortex, which receives a rich serotonergic innervation derived from both nuclei. The divergent nature of serotonergic projections is illustrated by this innervation; it has been estimated that each serotonergic neuron may influence 500,000 target neurons. Furthermore, each cortical neuron may be associated with over 200 serotonergic varicosities, which provides a means through which serotonin could affect widespread and coordinated modulation of cortical function. The caudal raphe serotonergic neurons project to the medulla, cerebellum, and spinal cord.

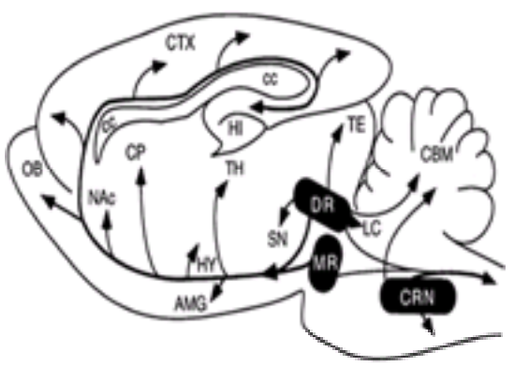

FIGURE 1.4-1 Brain serotonergic pathways (in rats). Serotonergic neurons are located in brainstem midline raphe nuclei and project throughout the neuraxis. (There is an approximate similarity between monoamine pathways in rats and in humans.) Abbreviations: AMG, amygdala; BFC, basal forebrain complex; CBM, cerebellum; cc, corpus callosum; CP, caudate putamen; CRN, caudal raphe nuclei; CTX, neocortex; DR, dorsal raphe nucleus; HI, hippocampus; HY, hypothalamus; LC, locus ceruleus; LTN, lateral tegmental noradrenergic nuclei; MPC, mesopontine complex; MR, median raphe nucleus; NAc, nucleus accumbens; OB, olfactory bulb; PFC, prefrontal cortex; PI, pituitary; SNC, substantial nigra pars compacta; TE, tectum; TH, thalamus; TM, tuberomammillary nucleus of hypothalamus; VTA, ventral tegmental area.

Dopamine Dopamine neurons are more widely distributed than those of other monoamines, residing in the midbrain substantia nigra and ventral tegmental area, and in the periaqueductal gray, hypothalamus, olfactory bulb, and retina. Of particular relevance are three dopamine containing systems: (1) the nigrostriatal, (2) mesocorticolimbic, and (3) tuberohypophyseal system (Fig. 1.4-2). The nigrostriatal dopamine system has been the most extensively studied of the dopaminergic pathways. Dopamine cell bodies located in the pars compacta division of the substantia nigra send ascending projections to the dorsal striatum, particularly the caudate and putamen. This projection modulates motor function, as highlighted by the motor disturbances of Parkinson's disease, a disorder characterized by degeneration of the nigrostriatal system. In addition, the extrapyramidal adverse effects of antipsychotic drugs are believed to result from the blockade of striatal dopamine receptors.

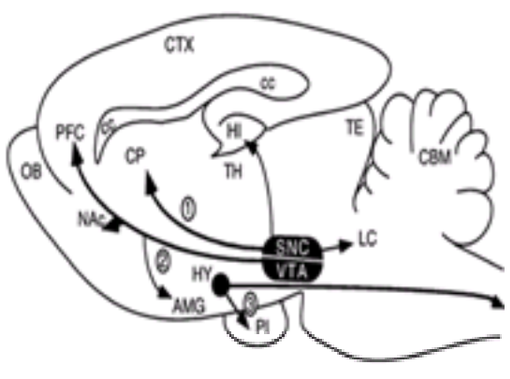

FIGURE 1.4-2 Brain dopaminergic pathways (in rats). The three principal dopaminergic pathways: (1) nigrostriatal pathway, (2) mesocorticolimbic pathway, and (3) tuberohypophyseal pathway.

The midbrain ventral tegmental area lies medial to the substantia nigra and contains dopaminergic neurons that give rise to the mesocorticolimbic dopamine system. These neurons send ascending projections that innervate limbic structures, such as the nucleus accumbens and amygdala, as well as associated cortical structures, particularly the prefrontal cortex. The mesoaccumbens projection is believed to regulate the rewarding properties of a wide variety of stimuli, including drugs of abuse. The mesocortical projection is believed to be a major target for the antipsychotic properties of dopamine receptor antagonist drugs. In this regard, the decreased predisposition of clozapine (Clozaril) to produce extrapyramidal adverse effects has been attributed to relatively selective actions on the mesocortical system. The tuberohypophyseal system consists of dopaminergic neurons in the hypothalamic arcuate and periventricular nuclei, and their projections to the pituitary gland. These projections provide inhibitory regulation of prolactin release. The administration of dopamine receptor antagonist antipsychotic drugs may lead to a disinhibition of release, resulting in galactorrhea.

Norepinephrine and Epinephrine Norepinephrine-producing neurons are found in the pons and medulla, in two major clusterings: the locus ceruleus and the lateral tegmental noradrenergic nuclei (Fig. 1.4-3). Noradrenergic projections from both these regions ramify extensively as they project throughout the neuraxis. In humans the locus ceruleus is found in the dorsal portion of the caudal pons, and contains approximately 12,000 tightly packed neurons on each side of the brain. These cells provide the major noradrenergic projections to the neocortex, hippocampus, thalamus, and midbrain tectum. The activity of locus ceruleus neurons varies with the sleep/wake cycle and is responsive to sensory stimuli, indicating a role for this structure in arousal state and vigilance. The projections from lateral tegmental noradrenergic nuclei neurons, which are loosely scattered throughout the ventral pons and medulla, partially overlap with those of the locus ceruleus. Fibers from both cell groups innervate the amygdala, septum, and spinal cord. Other regions, such as the hypothalamus and lower brainstem, receive adrenergic inputs predominantly from the lateral tegmental noradrenergic nuclei. The relatively few neurons that utilize epinephrine as a neurotransmitter are located in the caudal pons and medulla, intermingled with noradrenergic neurons. Projections from these groups ascend to innervate the hypothalamus, locus ceruleus, and visceral efferent and afferent nuclei of the midbrain.

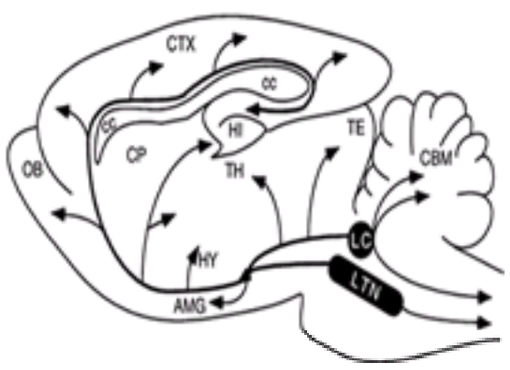

FIGURE 1.4-3 Brain noradrenergic pathways (in rats). Projections of noradrenergic neurons located in the locus ceruleus and lateral tegmental noradrenergic nuclei.

Histamine Central histaminergic neural pathways have only recently been characterized by immunocytochemistry using antibodies to the synthetic enzyme histidine decarboxylase and to histamine. Histaminergic cell bodies are located within the tuberomammillary nucleus of the posterior hypothalamus. As with other monoaminergic systems, histaminergic fibers project diffusely throughout the brain and spinal cord (Fig. 1.4-4). Ascending projections course with other monoaminergic fibers in the medial forebrain bundle, and descending projections travel through the midbrain central gray to the dorsal hindbrain and spinal cord. The fibers have varicosities that are seldom associated with classical synapses. The hypothalamus receives the densest histaminergic innervation, consistent with a role for this transmitter in the regulation of autonomic and neuroendocrine processes.

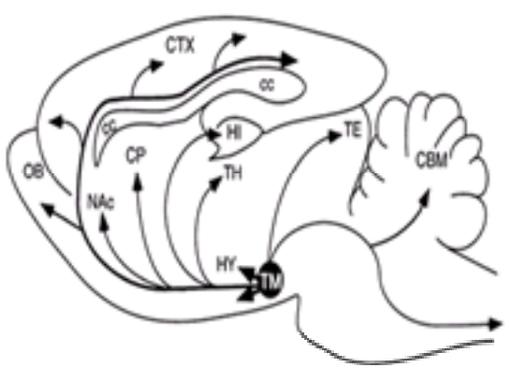

FIGURE 1.4-4 Brain histaminergic pathways (in rats). Histaminergic neurons are located in the tuberomammillary nucleus of the caudal hypothalamus and project to the hypothalamus and more distant brain regions.

Acetylcholine The axonal processes of cholinergic neurons may either project to distant brain regions (projection neurons) or contact local cells within the same structure (interneurons). Two large clusters of cholinergic projection neurons are found within the brain: the basal forebrain complex and the mesopontine complex (Fig. 1.4-5). The basal forebrain complex provides the vast majority of the cholinergic innervation to the nonstriatal telencephalon. It consists of cholinergic neurons within the medial septal nucleus, the nucleus of the diagonal band, the substantia innominata, the preoptic field, and the nucleus basalis of Meynert. In Alzheimer's disease the majority of nucleus basalis neurons are lost, leading to substantial impairments in the cortical cholinergic innervation. These impairments are believed to contribute to the symptoms of the disease and to correlate with the severity of dementia. The mesopontine complex consists of cholinergic neurons within the pedunculopontine and laterodorsal tegmental nuclei of the midbrain and pons. Regions innervated by these projections include the lateral hypothalamus, thalamus, tectum, substantia nigra, raphe nuclei, locus ceruleus, and cranial nerve nuclei. Acetylcholine is also found within interneurons of several brain regions, including the striatum. The modulation of striatal cholinergic transmission has been implicated in the antiparkinsonian actions of anticholinergic agents. Within the periphery, acetylcholine is a prominent neurotransmitter, located in preganglionic autonomic neurons, postganglionic parasympathetic neurons, and motoneurons innervating skeletal muscle.

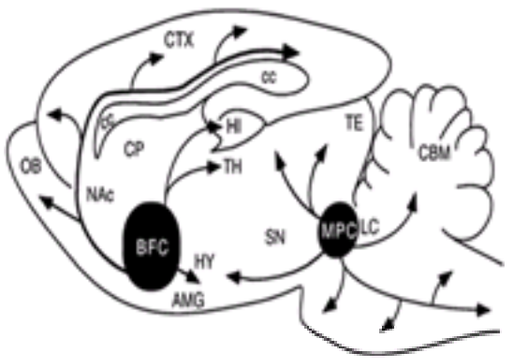

FIGURE 1.4-5 Brain cholinergic projection pathways (in rats). The majority of cholinergic projection neurons are located in the basal forebrain complex and the mesopontine complex.

MONOAMINE SYNTHESIS, STORAGE, AND DEGRADATION

In addition to similarities in neuronal organization, monoaminergic systems are similar with regard to their synthesis, storage, and degradation (Fig. 1.4-6). Monoamines are synthesized within neurons from common amino acid precursors (Fig. 1.4-6, Step 1) and taken up into synaptic vesicles via a vesicular monoamine transporter (Fig. 1.4-6, Step 2). Upon stimulation, vesicles within nerve terminals release neurotransmitter into the synaptic cleft (Fig. 1.4-6, Step 3). Once released, the monoamines interact with postsynaptic receptors to alter the excitability of postsynaptic cells (Fig. 1.4-6, Step 4). Monoamines may also interact with presynaptic autoreceptors located on the nerve terminal to suppress further release (Fig. 1.4-6, Step 5). In addition, released monoamines may be taken back up from the synaptic cleft into the nerve terminal by plasma membrane transporter proteins (Fig. 1.4-6, Step 6). Reuptake plays an important role in limiting the magnitude and duration of action of synaptically released monoamines. Once monoamines are taken up, they may be subject to enzymatic degradation (Fig. 1.4-6, Step 7) or they may be protected from degradation by uptake into vesicles. The processing of acetylcholine differs from this scheme, and is described below.

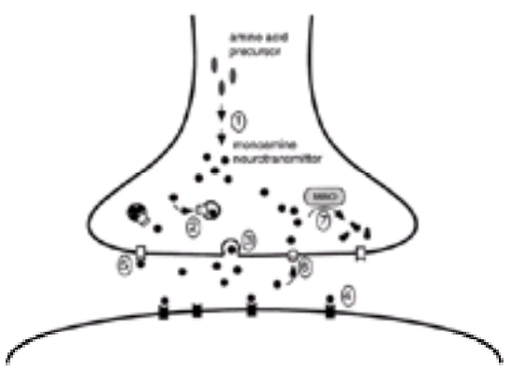

FIGURE 1.4-6 Schematic diagram of monoaminergic synapse. Steps involved in synaptic transmission are described in the text.

Serotonin The CNS contains less than 2 percent of the serotonin in the body; peripheral serotonin is located in platelets, mast cells, and enterochromaffin cells of the gastrointestinal system. Despite the abundance of peripheral serotonin, its inability to cross the blood-brain barrier necessitates the synthesis of serotonin within the brain. Serotonin is synthesized from the amino acid tryptophan, which is derived from the diet. The rate-limiting step in serotonin synthesis is the hydroxylation of tryptophan by the enzyme tryptophan hydroxylase to form 5-hydroxytryptophan (Fig. 1.4-7). Under normal circumstances this enzyme is not saturated by substrate, so tryptophan concentration can impact the rate of serotonin synthesis; therefore much attention has focused on the factors that determine tryptophan availability. Unlike serotonin, tryptophan is taken up into the brain via a saturable active carrier mechanism. Because tryptophan competes with other large neutral amino acids for transport, brain uptake of this amino acid is determined both by the amount of circulating tryptophan and by the ratio of tryptophan to other large neutral amino acids. This ratio may be elevated by carbohydrate intake, which induces insulin release and the uptake of many large neutral amino acids into peripheral tissues. Conversely, high-protein foods tend to be relatively low in tryptophan, thus lowering this ratio. The administration of specialized low-tryptophan diets has been found to produce significant declines in brain serotonin levels. Following tryptophan hydroxylation, 5-hydroxytryptophan is rapidly decarboxylated by aromatic amino acid decarboxylase to form serotonin.

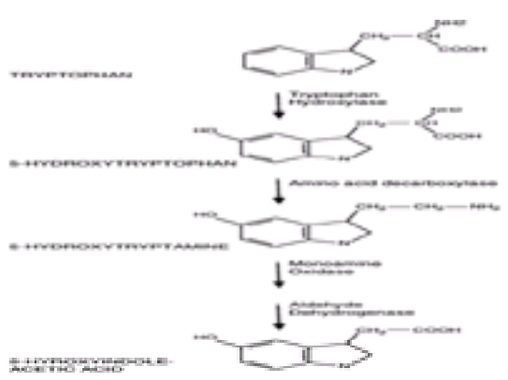

FIGURE 1.4-7 Synthesis and catabolism of serotonin.

The first step in the degradation of serotonin is mediated by monoamine oxidase (MAO) type A. A (MAO_A), which oxidizes the amino group to form an aldehyde. MAO_A is located in mitochondrial membranes and is nonspecific in its substrate specificity; in addition to serotonin, it oxidizes norepinephrine. The elevation of serotonin levels by MAO inhibitors (MAOIs) is believed to underlie the antidepressant efficacy of these drugs. Following oxidation by MAO_A, the resulting aldehyde is further oxidized to 5-hydroxyindoleacetic acid (5-HIAA).

Catecholamines The catecholamines are synthesized from the amino acid tyrosine, which is taken up into the brain via an active transport mechanism (Fig. 1.4-8). Within catecholaminergic neurons, tyrosine hydroxylase catalyzes the addition of a hydroxyl group to the meta position of tyrosine, yielding dopa. This rate-limiting step in catecholamine synthesis is subject to inhibition by high levels of catecholamines (end-product inhibition). Because tyrosine hydroxylase is normally saturated with substrate, manipulation of tyrosine levels does not readily impact the rate of catecholamine synthesis. Once formed, dopa is rapidly converted to dopamine by dopa decarboxylase, which is located in the cytoplasm. It is now recognized that this enzyme acts not only on dopa, but on all naturally occurring aromatic l-amino acids, including tryptophan. Thus, this enzyme is more accurately termed *aromatic amino acid decarboxylase*. In noradrenergic and adrenergic neurons, dopamine is oxidized in catecholaminergic vesicles by dopamine b-hydroxylase to form norepinephrine. This promiscuous enzyme will oxidize most phenylethylamines, producing metabolites that may replace norepinephrine at nerve terminals. These metabolites may act as false neurotransmitter, producing minimal postsynaptic effects. In adrenergic neurons, norepinephrine is converted to epinephrine by phenylethanolamine *N*-methyltransferase (PNMT), which is located within the cytoplasmic compartment.

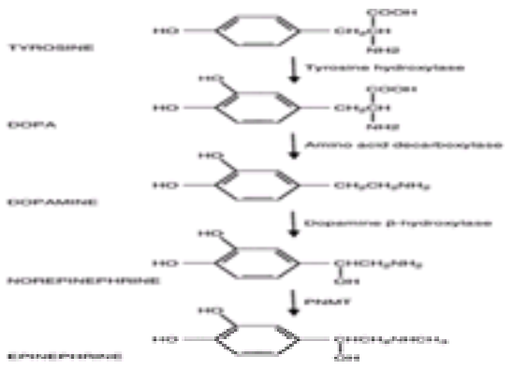

FIGURE 1.4-8 Synthesis of catecholamines.

Two enzymes that play major roles in the degradation of catecholamines are monoamine oxidase and catechol *O*-methyltransferase (COMT). Monoamine oxidase is located on the outer membrane of mitochondria and oxidatively deaminates catecholamines to their corresponding aldehydes. Two MAO isozymes with differing substrate specificities have been identified: MAO_A, which preferentially deaminates serotonin and norepinephrine, and MAO type B (MAO_B), which deaminates dopamine, histamine, and a broad spectrum of phenylethylamines. The blockade of monoamine catabolism by MAOIs produces elevations in brain monoamine levels. MAO is also found in peripheral tissues such as the gastrointestinal tract and liver, where it prevents the accumulation of toxic amines. MAO inhibitors may produce adverse effects by elevating the levels of these peripheral amines. For example, peripheral MAO degrades dietary tyramine, an amine that can displace norepinephrine from sympathetic postganglionic nerve endings, producing hypertension. Thus, patients treated with MAOIs are cautioned to avoid pickled and fermented foods that typically have high levels of this amine. COMT is widely distributed throughout the brain and peripheral tissues. It has a wide substrate specificity, catalyzing the transfer of methyl groups from *S*-adenosyl methionine to the *m*-hydroxyl group of most catechol compounds. The catecholamine metabolites produced by these and other degradative enzymes are frequently measured as indicators of the activity of catecholaminergic systems. In humans the predominant metabolites of dopamine and norepinephrine are homovanillic acid (HVA) and 3-methoxy-4-hydroxyphenylglycol (MHPG), respectively.

Histamine As is the case for serotonin, the brain contains only a small portion of the histamine found in the body. Histamine is distributed throughout most tissues of the body, predominantly in mast cells. Because it does not readily cross the blood-brain barrier, histamine is believed to be synthesized within the brain where it is formed by the decarboxylation of the amino acid histidine by a specific histidine decarboxylase. As this enzyme is not normally saturated with substrate, histamine synthesis is sensitive to histidine levels, which is consistent with the observation that the peripheral administration of histidine elevates brain histamine levels. Histamine is metabolized in the brain by histamine methyltransferase, producing methylhistamine; in turn, methylhistamine undergoes oxidative deamination by MAO_B.

Acetylcholine Acetylcholine is synthesized by the transfer of a methyl group from acetyl coenzyme A to choline in a reaction mediated by choline acetyltransferase (CAT). Choline availability is the rate-limiting factor in acetylcholine synthesis. The majority of choline within the brain is transported from the blood, and a high-affinity transport mechanism exists within cholinergic nerve terminals, where CAT is also found. The rate of choline transport is regulated such that increased cholinergic neural activity is associated with enhanced choline uptake. Following synthesis, acetylcholine is stored in synaptic vesicles through the action of a vesicular acetylcholine transporter. Following vesicular release into the synaptic cleft, acetylcholine is rapidly hydrolyzed to choline by acetylcholinesterase located in the synaptic cleft. Choline is then taken back into the presynaptic terminal via the choline transporter. One strategy in the treatment of dementia due to Alzheimer's disease is the augmentation of cholinergic transmission using cholinesterase inhibitors, such as tacrine (Cognex).

Plasma Membrane Transporters A great deal of progress has been made in the molecular characterization of the monoamine plasma membrane transporter proteins. These membrane proteins mediate the reuptake of synaptically released monoamines into the presynaptic terminal. The reuptake of monoamines across the presynaptic membrane into the nerve terminal is an energy-requiring process dependent upon the activity of a Na^+, K^+-activated adenosine triphosophatase (ATPase). Monoamine reuptake is an important mechanism for limiting the extent and duration of activation of monoaminergic receptors; also, transporters serve as molecular targets for a number of antidepressant drugs, psychostimulants, and monoaminergic neurotoxins. Whereas transporter molecules for serotonin, dopamine, and norepinephrine have been well characterized, transporters selective for histamine and epinephrine have not been demonstrated.

The molecular cloning of serotonin, dopamine, and norepinephrine transporter molecules has confirmed that all belong to a common gene family of transporter molecules that also includes those for g-aminobutyric acid (GABA), glycine, and choline. These proteins share strong sequence homologies and are believed to be integral membrane proteins that span the plasma membrane twelve times. In contrast to monoaminergic receptors, there is evidence for only a single type of transporter molecule for serotonin, dopamine, and norepinephrine. The expression of these proteins appears to be restricted to the corresponding class of monoaminergic neurons. For example, the messenger ribonucleic acid (mRNA) encoding the serotonin transporter molecule is restricted to serotonergic neurons, the one coding the dopamine transporter molecule is restricted to dopaminergic neurons, and the one coding the norepinephrine transporter molecule is restricted to noradrenergic neurons.

Monoaminergic transporters are molecular targets for psychotherapeutic drugs as well as substances of abuse. The therapeutic effects of tricyclic drugs, such as amitriptyline and imipramine, have been associated with their blockade of the serotonin transporter molecule and the norepinephrine transporter molecule although these drugs also interact directly with several monoaminergic receptor subtypes. More blockers of serotonin transporter molecules, such as the selective serotonin receptor inhibitors (SSRIs) (e.g., citalopram [Celexa], fluoxetine [Prozac], fluvoxamine [Luvox], paroxetine [Paxil], and sertraline [Zoloft]), are used in the treatment of depressive, anxiety, and a variety of other disorders. Conversely, compounds with relative selectivity for the norepinephrine transporter molecules, such as nortriptyline (Pamelor) and desipramine (Norpramin), also have antidepressant efficacy. Among drugs of abuse, cocaine binds with high affinity to all three known monoamine transporters, although the rewarding and stimulant properties of the drug have been attributed primarily to its blockade of the dopamine transporter molecule. This view has been recently supported by the absence of cocaine-induced locomotor stimulation in a strain of mutant mice engineered to lack this molecule. In fact, psychostimulants produce a paradoxical locomotor suppression in these animals that has been attributed to the serotonin molecule transporter-blocking properties of these compounds. Transporters may also provide routes that allow neurotoxins to enter and damage monoaminergic neurons; examples include the dopaminergic neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) and the serotonergic neurotoxin MDMA.

Vesicular Monoamine Transporter In addition to the reuptake of monoamines into the presynaptic nerve terminal, a second transport process serves to concentrate and store monoamines within synaptic vesicles. The transport and storage of monoamines in vesicles may serve several purposes: (1) to enable the regulated release of transmitter under appropriate physiological stimulation, (2) to protect monoamines from degradation by MAO, and (3) to protect neurons from the toxic effects of free radicals produced by the oxidation of cytoplasmic monoamines. In contrast with the plasma membrane transporters, a single type of vesicular monoamine transporter is believed to mediate the uptake of monoamines into synaptic vesicles within the brain. Consistent with this, blockade of this vesicular monoamine transporter by the antihypertensive drug reserpine (Serpasil) has been found to deplete brain levels of serotonin, norepinephrine, and dopamine. The molecular cloning of this transporter has revealed it to have 12 putative membrane-spanning domains; however, it does not display sequence homology to the plasma membrane transporters. Moreover, it utilizes a H^+ gradient rather than Na^+/Cl^- gradients. Vesicular transport requires a H^+ pumping ATPase, which establishes a concentration gradient of H^+ across the vesicle membrane. In the presence of this gradient, the vesicular monoamine transmitter takes up neurotransmitter in a manner that is coupled to the release of luminal protons; the activity of the transporter is altered by amphetamine-like agents. These compounds are taken up via plasma membrane transporters into monoaminergic terminals, where they act as weak bases to disrupt pH gradients. This produces a reversal of vesicular monoamine transmitter activity, leading to monoamine release from vesicles and reversal of plasma membrane transporter activity. The resulting release of monoamines from presynaptic terminals contributes to the stimulant properties of these compounds. The anorectic agent fenfluramine is believed to stimulate serotonin release in an analogous manner. Recently a separate vesicular transporter for acetylcholine has been molecularly cloned; its structure is homologous to that of the vesicular monoamine transmitter, and both are believed to have a common bioenergetic mechanism.

RECEPTORS

Ultimately the effects of monoamines on CNS function and behavior depend upon their interactions with receptor molecules. The binding of monoamines to these plasma membrane proteins initiates a series of intracellular events that modulate neuronal excitability. Unlike the transporters, multiple receptor subtypes exist for each monoamine transmitter (Table 1.4-1). The initial classification of many receptor subtypes was based on radioligand binding studies. Receptor binding sites were identified on the basis of the rank order of binding affinities for a number of agonist and antagonist compounds. More recently, the molecular cloning of monoamine

receptors has confirmed that many of the sites initially defined by these binding studies did indeed correspond to distinct receptor proteins encoded by unique genes. Molecular cloning has also led to the identification of previously unknown receptors, and to the introduction of powerful tools to characterize receptor structure and function.

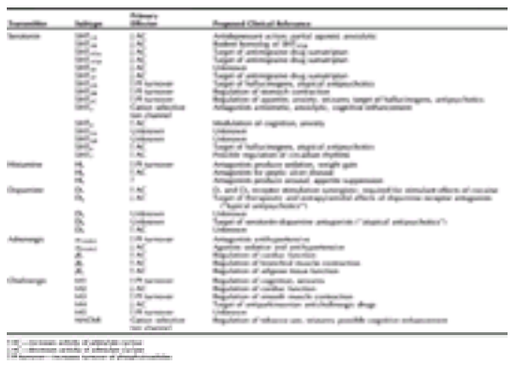

Table 1.4-1 Monoamine Receptors: Overview

Neurotransmitter receptors produce intracellular effects by one of two basic mechanisms: (1) via interactions with G proteins that couple receptors to intracellular effector systems, and (2) by providing channels through which ions flow when transmitters bind (ligand-gated ion channels). With the exception of the serotonin type 3 (5-HT_3) receptor subtype (a ligand-gated ion channel), all known monoaminergic receptors belong to the superfamily of G-protein–coupled receptors. However, within each monoaminergic receptor family, the subtypes are heterogeneous with regard to the G proteins with which they interact, and to the second-messenger effects that they produce. Monoaminergic receptors are also diverse in their regional patterns of expression within the brain, their neurotransmitter binding affinities, and in their synaptic localization. Whereas many receptor subtypes are located exclusively in postsynaptic membranes, others are located presynaptically in axon terminal membranes. Some receptors on the presynaptic terminal respond to monoamines that are released by that neuron and these presynaptic autoreceptors often act to inhibit neurotransmitter release. Several monoaminergic receptor subtypes are located presynaptically in some brain regions and postsynaptically in others.

In the wake of the recent proliferation of known receptors subtypes, much work needs to be done to determine the functional roles of individual receptors. In many instances this effort is hampered by the paucity of selective agonist and antagonist drugs. Recently, a molecular genetic approach to examining receptor function has been applied to complement pharmacological studies. Gene targeting procedures have enabled the generation of mouse strains with disruptions in genes that encode individual receptor subtypes; the resulting mutant mice have a complete and specific absence of the targeted receptor. Studies in these animals are providing clues to receptor function and to the contributions of the targeted receptors to the actions of nonspecific drugs. Molecular and pharmacological approaches will guide the generation of subtype-selective compounds and facilitate the development of therapeutic agents that will alter monoaminergic transmission in a more refined manner.

Serotonin Receptors Brain serotonin receptors were initially characterized on the basis of radioligand binding studies into two classes: serotonin type 1 (5-HT_1) receptors, to which [$_3$H]5-HT bound with high affinity, and 5-HT_2 which were labeled with high affinity by [^{3}H]spiperone. Subsequent binding studies revealed that each of these classes consisted of multiple subtypes. The application of molecular cloning techniques has produced a proliferation in the number of known subtypes. The great diversity of serotonin receptors provides a means whereby a single neurotransmitter may produce a wide variety of cellular effects in multiple neuronal systems. At present, at least 14 distinct serotonin receptor subtypes have been identified and molecularly cloned, which has led to rapid advances in determining the structure, pharmacology, brain distribution, and effector mechanisms of these receptors and to a more precise classification of serotonin receptor subfamilies on the basis of their structural homologies and primary effector mechanisms.

The 5-HT_1 receptors comprise the largest serotonin receptor subfamily, with human subtypes designated: 5-HT_{1A}, 5-HT_{1Da}, 5HT_{1Db}, 5-HT_{1E}, and 5-HT_{1F}. All five 5-HT_1 receptor subtypes display intronless gene structures and high affinities for serotonin and adenylate cyclase inhibition. The most intensively studied of these has been the 5-HT_{1A} receptor, a subtype that is found on postsynaptic membranes and serotonergic neurons, where it functions as a somatodendritic autoreceptor. The stimulation of these autoreceptors suppresses the activity of serotonergic neurons. The desensitization of 5-HT_{1A} autoreceptors by the chronic administration of SSRIs has been implicated in their antidepressant effects. An additional role for this receptor subtype in the regulation of anxiety is suggested by the anxiolytic properties of partial 5-HT_{1A} receptor agonists such as buspirone (Buspar) and by the enhancement of anxiety-related behaviors in a strain of mutant mice lacking 5-HT_{1A} receptors. The 5-HT_{1Da} and 5-HT_{1Db} receptors resemble each other in structure and brain localization. 5-HT_{1D} receptors are found on axon terminals of serotonergic and nonserotonergic neurons, where they act to reduce neurotransmitter release. The study of these receptors has been hindered by a lack of selective pharmacological tools. This is also true for the 5-HT_{1E} and 5-HT_{1F} receptor subtypes. The highest levels of 5-HT_{1E} receptor expression are found in the striatum and entorhinal cortex whereas 5-HT_{1F} receptor expression is highest in the dorsal raphe nucleus, hippocampus, cortex, and striatum. In addition, 5-HT_{1D} and 5-HT_{1F} receptors are expressed in cerebral vessels, and are stimulated by the antimigraine drug sumatriptan (Imitrex). The relative importance of these receptors in the therapeutic efficacy of this drug remains to be determined.

At least three receptor subtypes mediate the effects previously attributed to a single 5HT_2 receptor subtype. The classic 5-HT_2 receptor has thus been renamed 5-HT_{2A} to indicate that it is a member of a serotonin receptor subfamily. A second receptor initially termed 5-HT_{1C} has been renamed 5-HT_{2C} to indicate that it belongs within this subfamily. The third known 5HT_2 receptor, termed 5-HT_{2B}, contributes to the contractile effects of serotonin in the stomach fundus. All three subtypes exhibit high sequence homology, similar pharmacological binding profiles, and stimulation of phosphoinositide turnover. High levels of 5-HT_{2A} receptors are found in the neocortex and in peripheral locations such as platelets and smooth muscle. Much recent attention has focused on the contributions of 5-HT_{2A} and 5-HT_{2C} receptors to the actions of serotonin-dopamine antagonists, such as clozapine (Clozaril), risperidone (Risperdal), and olanzapine (Zyprexa). Analysis of the receptor-binding properties of these drugs has led to the hypothesis that 5-HT_{2A} receptor blockade correlates with the therapeutic effectiveness of these antipsychotic agents. The relative importance of related serotonin receptors to the efficacy of these agents remains to be determined. The 5-HT_{2C} receptor is expressed at high levels in many CNS regions including the hippocampal formation, prefrontal cortex, amygdala, striatum, hypothalamus, and choroid plexus. Stimulation of 5-HT_{2C} receptors has been proposed to produce anxiogenic and anorectic effects. Accordingly, a transgenic mouse strain lacking this receptor subtype exhibits an obesity syndrome associated with overeating; these animals also display an enhanced susceptibility to seizures, implicating this receptor in the regulation of neuronal network excitability. A variety of antidepressant and antipsychotic drugs antagonize 5-HT_{2C} receptors with high affinity. Conversely, hallucinogens such as lysergic acid diethylamide (LSD) display agonist activity at 5-HT_2 serotonin receptor subtypes, among others.

The 5-HT_3 receptor is unique among monoaminergic receptors in its membership within the ligand-gated ion channel superfamily. Rather than activating G proteins, the binding of serotonin to this receptor permits passage of Na^+ and K^+ ions through an ion channel located within the 5-HT_3 receptor complex. This produces rapid excitatory effects in postsynaptic neurons. The receptor is expressed within the hippocampus, neocortex, amygdala, hypothalamus, and brainstem motor nuclei. Outside the brain, it is found in the pituitary gland, enteric nervous system sympathetic ganglia, and in sensory ganglia. 5-HT_3 receptor antagonists such as ondansetron (Zofran) have been used as antiemetic agents and are under evaluation as potential antianxiety and cognitive-enhancing agents.

Investigations into the functions of recently identified 5-HT_4; 5-HT_5, including subtypes 5-HT_{5A} and 5-HT_{5B}; 5-HT_6 and 5-HT_7 receptors are hindered by a lack of selective agonist and antagonists. Studies of the cloned receptors reveal that all but the 5-HT_5 receptor are linked to the stimulation of adenylate cyclase. The primary effector mechanisms of the 5-HT_5 receptors remain to be determined. The 5-HT_4 receptors are expressed in the hippocampus, striatum, substantia nigra, and superior colliculus, and have been implicated in the serotonergic regulation of cognition and anxiety. The two 5-HT_5 receptor subtypes are highly homologous, and are expressed in the neocortex, hippocampus, raphe nuclei, and cerebellum. 5-HT_6 receptors may contribute to the actions of the several antidepressant, antipsychotic, and hallucinogenic drugs that bind with high affinity. These receptors are expressed in the neocortex, hippocampus, striatum, and amygdala; highest concentrations of 5-HT_7 receptor expression are found in the hypothalamus and thalamus. These receptors have been proposed to contribute to the serotonergic modulation of circadian rhythms. Although it is premature to assign functional roles to these new receptor subtypes with confidence, it is likely that these receptors will ultimately provide targets for the development of useful therapeutic compounds.

Dopamine Receptors In 1979 it was clearly recognized that the actions of dopamine are mediated by more than one receptor subtype. Two dopamine receptors, termed type 1 (D_1) and type 2 (D_2) were distinguished on the basis of differential binding affinities of a series of agonists and antagonists, distinct effector mechanisms, and distinct distribution patterns within the CNS. It was subsequently found that the therapeutic efficacy of antipsychotic drugs correlated strongly with their affinities for the D_2 receptor, implicating this subtype as an important site for the action of antipsychotic drugs. Until recently, these were the only two identified dopamine receptors; however, molecular cloning studies have revealed additional receptor heterogeneity. Three additional dopamine receptor genes have been identified, encoding the D_3, D_4, and D_5 dopamine receptors. Based on their regional brain distributions and primary effector mechanisms, the D_3 and D_4 receptors are considered to be D_2-like, and the D_5 receptor is considered to be D_1-like. The functional roles of these subtypes remain to be determined, although several intriguing possibilities are under investigation.

The D_1 receptor was initially distinguished from the D_2 subtype by its high affinity for the antagonist SCH 23390 and its relatively low affinity for butyrophenones such as haloperidol (Haldol). Whereas D_1 receptor activation stimulates cAMP formation, D_2 receptor stimulation produces the opposite effect on cAMP formation. In addition to the stimulation of adenylate cyclase, D_1 receptors may also stimulate phosphoinositide turnover. D_1 receptor mRNA is expressed in the terminal fields of the nigrostriatal and mesocorticolimbic pathways, with high levels in the dorsal striatum, nucleus accumbens, and amygdala. In contrast, little D_1 mRNA expression is found in dopamine cell body regions such as the substantia nigra pars compacta and the ventral tegmental area. This finding, and the persistence of D_1 receptor binding following lesions of dopaminergic neurons, suggests that this receptor subtype is not found on dopaminergic neurons and is therefore not an autoreceptor. Electrophysiological studies have indicated that D_1 receptor activation is required for striatal D_2 receptor activation to produce its maximal effect. The proposed synergistic effects of striatal D_1 and D_2 receptor activation have recently received support from studies in a mouse strain with a targeted elimination of D_1 receptors. The effects of both D_1 and D_2 receptor activation were attenuated in these animals; moreover, these mice were resistant to the hyperlocomotor effects of cocaine, indicating that D_1 receptors contribute significantly to the effects of cocaine on the CNS.

The D_5 receptor was molecularly cloned on the basis of its sequence homology with the D_1 receptor. The two receptors have a higher degree of homology with each other than with the D_2 and D_4 subtypes. This structural similarity is reflected in the similar affinities of a wide variety of dopaminergic drugs for these two receptors. The main distinguishing feature of their binding profiles is that the binding affinity of dopamine is higher for D_5 than for D_1 receptors. Not surprisingly, these two receptors are also similar in that they both stimulate adenylate cyclase activity. However, these receptors differ in their regional distributions within the CNS: the expression of D_5 receptors appears to be highly restricted to limbic system structures such as the hippocampus and hypothalamus. The lack of selective drugs hinders attempts to distinguish the functional roles of D_1 and D_5 receptors.

The dopamine D_2 receptor was initially distinguished from the D_1 receptor on the basis of its high affinity for butyrophenones. Moreover, D_2 receptor stimulation was observed to inhibit rather than to stimulate adenylate cyclase activity. Subsequently, the D_2 receptor subtype was found to display interactions with a variety of G proteins, leading to diverse second-messenger effects such as modulation of Ca^{++} and K^+ channel function and the alteration of phosphoinositide production. The intracellular consequences of D_2 receptor activation appear to depend upon the cell type in which the receptor is expressed. In addition to D_2 receptor mRNA expression in brain regions that receive dopaminergic innervation, D_2 transcripts are found in dopaminergic neurons of the ventral tegmental area and substantia nigra. The D_2 receptor may have either a postsynaptic function or an autoreceptor function. D_2 autoreceptors may be found on dopaminergic terminals or on the cell bodies and dendrites of dopaminergic neurons. D_2 receptors are also expressed in the anterior pituitary and mediate the dopaminergic inhibition of prolactin and melanocyte-stimulating hormone release. Molecular cloning has revealed a long and short form of the D_2 receptor that differ in length by 29 amino acids, products of alternative splicing of a single gene; differences in the functional roles of the long and short forms remain to be determined.

A great deal of attention has focused on the clinical correlates of D_2 receptor function. It has been proposed that the brains of untreated schizophrenia patients have elevations in D_2 receptor density. Furthermore, radioligand binding studies have revealed a correlation between the clinical efficacy of antipsychotic drugs and their antagonist affinities for this receptor subtype. This finding has contributed significantly to the “dopamine hypothesis” of schizophrenia. The extrapyramidal adverse effects of antipsychotic drugs, such as dystonia and parkinson-like symptoms, have been attributed to the blockade of striatal D_2 receptors. A significant contribution of D_2 receptors to the dopaminergic regulation of motor function is further highlighted by a parkinson-like movement disorder observed in a mutant mouse strain that lacks this receptor subtype.

The D_3 and D_4 receptors are considered to be D_2-like on the basis of similarities in their gene structures, sequence homologies, and pharmacology. These receptors are expressed in lower abundance than the D_2 receptor and their regional distributions are distinct. Whereas D_3 receptor expression is high in the nucleus accumbens, highest levels of D_4 receptors are expressed in the frontal cortex, midbrain, amygdala, and medulla. Whereas little D_3 receptor expression has been detected outside the nervous system, D_4 receptors are more abundant in the heart than in the brain. The primary effector mechanisms associated with the stimulation of brain D_3 and D_4 receptors remain to be determined. Particular attention has been paid to a potential role of D_4 receptors in schizophrenia. On postmortem, elevated D_4 receptor levels have been found in the brains of patients with schizophrenia. Moreover, the atypical antipsychotic drug clozapine (Clozaril) has a high affinity for the D_4 receptor; this receptor is highly polymorphic in humans and at least 25 distinct alleles have been identified. Studies are therefore under way to determine whether particular D_4 alleles are associated with psychotic disorders or with responsiveness to antipsychotic drugs.

Adrenergic Receptors Adrenergic receptor heterogeneity was first discovered in the 1940s, when a and b subtypes were identified in pharmacological studies of isolated peripheral tissues. Subsequently, radioligand binding and molecular cloning studies have identified three main adrenergic receptor subfamilies; a_1, a_2, and b. Each subfamily consists of at least three distinct receptor subtypes. Receptors within each subfamily share similar sequence homologies, pharmacological binding profiles, and effector mechanisms. The activation of a_1 receptors (subtypes designated a_{1A}, a_{1B}, and a_{1C}) stimulates phosphoinositide turnover. These receptors are believed to play a significant role in regulating smooth muscle contraction and have been implicated in the control of blood pressure, nasal congestion, and prostate function. Although all three subtypes are expressed in the brain, their contributions to the central actions of norepinephrine remain to be determined.

The a_2 receptors (designated including subtypes a_{2A}, a_{2B}, and a_{2C}) have been implicated in the regulation of cardiovascular function, autonomic nervous system activity, and arousal. The functions of a_2 receptor subtypes have been difficult to determine because of a lack of selective agonists and antagonists; a_2 receptors display both presynaptic autoreceptor and postsynaptic actions, and all appear to inhibit cAMP formation. The stimulation of a_2 autoreceptors inhibits firing of the noradrenergic neurons of the locus ceruleus, which have been implicated in arousal states. This mechanism has been proposed to underlie the sedative effects of the a_2 receptor agonist clonidine (Catapres). In addition, the stimulation of brainstem a_2 receptors has been proposed to reduce sympathetic nervous system activity and to augment parasympathetic nervous system activity, which may relate to the utility of clonidine in lowering blood pressure and in suppressing the sympathetic hyperactivity associated with opiate withdrawal. The a_2 receptor antagonist yohimbine (Yocon) is used in the treatment of erectile disorder (impotence); the neural mechanisms that underlie this effect remain to be determined.

Like the a-adrenergic receptors described, b-adrenergic receptors (designated including subtypes b_1, b_2, and b_3) are found both in the brain and in many peripheral tissues. The functional roles of the peripheral b-adrenergic receptors are better understood than are its central functions. Cardiac b_1 receptors play a major role in the regulation of heart function, and b_2 receptors regulate bronchial muscle contraction. b_3 receptors are found in adipose tissue, where they stimulate fat catabolism. Although b_1 and b_2 receptors are widely distributed in the CNS, their contributions to catecholamine function are not well understood. Propranolol (Inderal) is a widely used nonspecific antagonist of both b_1 and b_2 receptors. In addition to its utility for the treatment of hypertension and arrhythmias, its effectiveness in blunting autonomic symptoms underlies its utility in the management of social phobia; also, through mechanisms that are currently unknown, it is also effective in the treatment of akathisia.

Histamine Receptors Histaminergic systems have been proposed to modulate arousal, wakefulness, feeding behavior, and neuroendocrine responsiveness. Three histaminergic receptor subtypes have been identified. Histamine type 1 (H_1) receptors are expressed throughout the body, particularly in smooth muscle of the gastrointestinal tract and blood vessel walls. H_1 receptors are widely distributed throughout the CNS, with particularly high levels in the forebrain and cerebellum. H_1 receptor activation stimulates phosphoinositide turnover, and tends to increase excitatory neuronal responses. These receptors are the targets of antihistaminergic agents used in the treatment of allergic rhinitis and conjunctivitis. The well-known sedative effects of these compounds have been attributed to their actions in the CNS. In addition, the sedation and weight gain produced by a number of antipsychotic and antidepressant drugs have been attributed to H_1 receptor antagonism;

conversely, H_1 receptor agonists stimulate arousal and suppress food intake in animal models.

H_2 receptors are also widely distributed throughout the body, and are found in gastric mucosa, smooth muscle, cardiac muscle, and cells of the immune system. Within the CNS, H_2 receptors are abundantly expressed in the neocortex, hippocampus, amygdala, and striatum. Activation of these receptors stimulates adenylate cyclase and produces excitatory effects in neurons of the hippocampal formation and thalamus. H_2 receptor antagonists are widely used in the treatment of peptic ulcer disease. The functional significance of central H_2 receptors is unclear, although several studies indicate that stimulation of these receptors produces antinociceptive effects.

Unlike H_1 and H_2 histamine receptors, H_3 receptors are located presynaptically on axon terminals. Those located on histaminergic terminals act as autoreceptors to inhibit histamine release. H_3 receptors are also located on nonhistaminergic nerve terminals, where they inhibit the release of a variety of neurotransmitters. Particularly high levels of H_3 receptor binding are found in the frontal cortex, striatum, amygdaloid complex, and substantia nigra. Lower levels are found in peripheral tissues such as the gastrointestinal tract, pancreas, and lung. Antagonists of H_3 receptors have been proposed to have appetite suppressant, arousing, and cognitive-enhancing properties.

Cholinergic Receptors Two major classes of cholinergic receptors exist: G protein–coupled muscarinic receptors and nicotinic ligand-gated ion channels. Muscarinic receptors have been implicated in learning and memory, sleep regulation, pain perception, and the regulation of seizure susceptibility. The five known subtypes of muscarinic receptors are heterogeneous with regard to regional brain distribution and primary effector mechanisms. The muscarinic type 1 (M_1), M_3, and M_5 receptors stimulate phosphatidylinositol (PI) turnover, and the M_2 and M_4 receptors inhibit adenylate cyclase. The functional roles of the individual subtypes within the CNS are not well understood because highly subtype-selective agonists and antagonists have been unavailable. Based on brain distribution and the actions of nonspecific drugs, M_1 receptors have been implicated in learning and memory processes. Striatal M_4 receptors have been implicated as putative targets for anticholinergics used as antiparkinson agents. Recently, a line of mice lacking functional M_1 receptors has been generated and found to be resistant to the convulsant effects of muscarinic agonists. In the periphery, M_2 receptors regulate heart rate and contractility and M_3 receptors mediate smooth muscle contraction and glandular secretion.

Nicotinic acetylcholine receptors, like 5-HT_3 receptors, are members of the ligand-gated ion channel superfamily. They are composed of a pentameric complex of membrane protein subunits radially arranged around a central ion pore. The binding of acetylcholine to this receptor permits passage of Na^+ and K^+ ions through the ion channel, depolarizing the postsynaptic cell. Nicotinic acetylcholine subunits are heterogeneous and associate in varied combinations. Thus, the properties of an individual complex will depend upon its particular subunit composition. The relative frequencies of various subunit combinations in discrete brain regions and their functional significance remain to be fully clarified.

Within the human brain, nicotinic acetylcholine receptors are found at highest densities within the hippocampal formation, neocortex, substantia nigra, ventral tegmental area, dorsal raphe nucleus, periaqueductal gray, and the basal forebrain cholinergic complex. Consistent with this distribution, these receptors have been implicated in cognitive function. Cortical acetylcholine receptors are diminished in Alzheimer's disease, and nicotine administration improves attention defects in some patients. In some schizophrenia patients nicotine administration improves measures of sensory gating. This and other data have led to the proposal that the high incidence of cigarette smoking in this population represents a form of self-medication. Recently, some rare familial epilepsy syndromes have been associated with mutations of nicotinic acetylcholine receptors. Finally, the reinforcing properties of tobacco use are proposed to involve the stimulation of nicotinic acetylcholine receptors located in mesolimbic reward pathways.

SUGGESTED CROSS-REFERENCES

The intracellular consequences of receptor activation are discussed in Section 1.8. The electrophysiological effects of brain monoamines are described in Section 1.9. The basic concepts of molecular biology that are relevant to current monoamine research are presented in Section 1.18. Alzheimer's disease and seizure disorders are covered in Chapter 10. Section 31.5 presents b-adrenergic receptor antagonists, Section 31.6 presents anticholinergics, Section 31.7 presents antihistamines, Section 31.15 presents cholinesterase inhibitors, Section 31.17 presents dopamine receptor antagonists, 31.26 presents serotonin-dopamine antagonists, and Section 31.25 presents selective serotonin reuptake inhibitors.

SECTION REFERENCES

Baik JH, Picetti R, Saiardi G, Thiriet A, Dierich A, Depaulis A, Le Meur M, Borrelli E: Parkinsonian-like locomotor impairment in mice lacking dopamine D_2 receptors. Nature *377:*424, 1995.

*Barker EL, Blakely RD: Norepinephrine and serotonin transporters: Molecular targets of antidepressant drugs. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

*Cooper JR, Bloom FE, Roth RH: *The Biochemical Basis of Neuropharmacology,* ed 7. Oxford University Press, New York, 1996.

*Gainetdinov RR, Wetsel WC, Jones SR, Levin ED, Jaber M, Caron MG: Role of serotonin in the paradoxical calming effect of psychostimulants on hyperactivity. Science *283:*397, 1999.

Giros B, Jaber M, Jones SR, Wightman BM, Caron MG: Hyperlocomotion and indifference to cocaine and amphetamine in mice lacking the dopamine transporter. Nature *379:*606, 1996.

Gotti C, Fornasari D, Clementi F: Human neuronal nicotinic receptors. Prog Neurobiol *53:*199, 1997.

Hamilton SE, Loose MD, Qi M, Levey AI, Hille B, McKnight GS, Idzerda RL, Nathanson NM: Disruption of the m1 receptor gene ablates muscarinic receptor-dependent M current regulation and seizure activity in mice. Proc Natl Acad Sci USA *94:*13311, 1997.

*Hartman DS, Civelli O: Dopamine receptor diversity: Molecular and pharmacological perspectives. Prog Drug Res *48:*173, 1997.

Heisler L, Chu HM, Brennan T, Danao J, Bajwa P, Parsons L, Tecott LH: Elevated anxiety and antidepressant-like responses in serotonin 5-HT_{1A} receptor mutant mice. Proc Natl Acad Sci USA *95:*15049, 1998.

Hu XT, Moratalla R, Graybiel AM, White FJ, Tonegawa S: Elimination of cocaine-induced hyperactivity and dopamine-mediated neurophysiological effects in dopamine D_1 receptor mutant mice. Cell *79:*945, 1994.

Knable MB, Weinberger DR: Dopamine, the prefrontal cortex and schizophrenia. J Psychopharmacol *11:*123, 1997.

Leurs R, Smit MJ, Timmerman H: Molecular pharmacological aspects of histamine receptors. Pharmacol Ther *66:*413, 1995.

Lindvall O, Bjorklund A: Dopamine- and norepinephrine-containing neuron systems: Their anatomy in rat brain. In *Chemical Neuroanatomy,* PC Emson, editor. Raven, New York, 1983.

*Liu Y, Edwards RH: The role of vesicular transport proteins in synaptic transmission and neural degeneration. Ann Rev Neurosci *20:*125, 1997.

MacKinnon AC, Spedding M, Brown CM: Alpha 2-adrenoceptors: More subtypes but fewer functional differences. Trends Pharmacol Sci *15:*119, 1994.

Minneman KP, Esbenshade TA: Alpha 1-adrenergic receptor subtypes. Ann Rev Pharmacol Toxicol *34:*117, 1994.

Ramboz S, Oosting R, Amara DA, Kung HF, Blier P, Mendelsohn M, Mann JJ, Brunner D, Hen R: Serotonin receptor 1A knockout: An animal model of anxiety-related disorder. Proc Natl Sci USA *95:*14476, 1998.

*Schwartz JC, Arrang JM, Garbard M, Traiffort E: Histamine. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Tecott LH: Serotonin receptor diversity: Implications for psychopharmacology. In *American Psychiatric Press Review of Psychiatry,* vol 15, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1996.

Tecott LH, Sun LM, Akana SF, Strack AM, Lowenstein DH, Dallman MF, Julius D: Eating disorder and epilepsy in mice lacking 5HT2$_C$ serotonin receptors. Nature *374:*542, 1995.

Tork I: Anatomy of the serotonergic system. Ann NY Acad Sci *600:*9, 1990.

Weiner N, Molinoff PB: Catecholamines. In *Basic Neurochemistry,* ed 5, GH Siegel, BW Agranoff, RW Albers, PB Molinoff, editors. Raven, New York, 1994.

1.5 AMINO ACID NEUROTRANSMITTERS

RICHARD P. SHANK, PH.D., VIRGINIA L. SMITH-SWINTOSKY, PH.D. AND ROY E. TWYMAN, M.D.

Throughout the first half of the twentieth century there was a prevailing view that neurons within the vertebrate central nervous system (CNS) are electrically coupled at their synapses and that synaptic transmission therefore could only be mediated electrically. About 1950 John Eccles and his colleagues realized that this could not be not true, and they were forced to conclude that synaptic transmission must be primarily a chemically mediated process. This realization stimulated research to identify the chemical messengers (neurotransmitters or transmitters). Electrophysiological experiments revealed that several amino acids either increased or decreased the excitability of neurons when applied to the extracellular surface. The amino acids that increased neuronal excitability include glutamate, aspartate, cysteate, and homocysteate, which are all similar structurally. Those that decreased neuronal excitability include g-aminobutyric acid (GABA), glycine, b-alanine and taurine, which also exhibit structural similarity.

After the neuronal effects of these amino acids were reported, a debate arose as to whether these amino acids were really neurotransmitters or served some other undefined function in regulating or modulating the excitability of neurons. This debate is still not fully resolved, in part because there is not yet a precise understanding of all the chemical events that regulate neuronal excitability and their physiological significance. Nevertheless, rigorous criteria have been developed that can be used to determine whether an amino acid is actually a neurotransmitter. Currently, three amino acids (glutamate, GABA, and glycine) meet the criteria sufficiently to allow a consensus conclusion that they are true neurotransmitters. Furthermore, there is now little doubt that glutamate is the major excitatory neurotransmitter and that GABA and glycine are the major inhibitory neurotransmitters in the brain and spinal cord, respectively. Two more, aspartate and taurine, are likely to serve a significant, but as yet undefined, role in modulating the excitability of neurons. A role for cysteate, homocysteate, and b-alanine is less certain.

NEUROPHYSIOLOGICAL SIGNIFICANCE OF EXCITATORY AMINO ACIDS

Nontransmitter Functions of Glutamate and Aspartate Glutamate and aspartate serve several functions in neural tissues unrelated to their role as neurotransmitters. Both are major constituents in proteins and several peptides, including *N*-acetylaspartylglutamate, a putative transmitter or cotransmitter. Aspartate is a precursor of *N*-acetylaspartate, a primary organic anion in the CNS. Glutamate and aspartate also have a role in energy metabolism; they are intermediates in the malate-aspartate shuttle, a metabolic pathway that reoxidizes cytosolic nicotine adenine dinucleotide (NADH) formed by aerobic glycolysis and captures the energy within mitochondria for adenosine triphosphate (ATP) formation (Fig. 1.5-1). Because glucose is rapidly consumed via aerobic metabolism in neural tissues, this pathway is quite active.

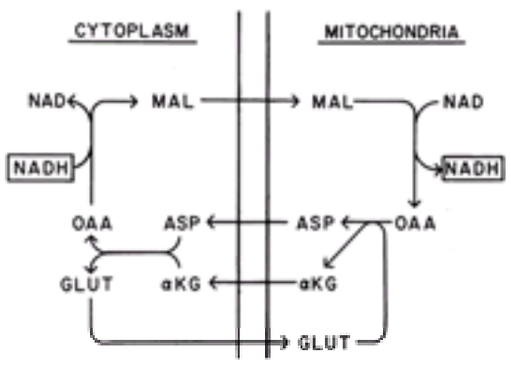

FIGURE 1.5-1 The metabolic components of the malate-aspartate shuttle. Although it is not evident from the diagram, the flux of aspartate (ASP) across the mitochondrial membrane is coupled to the flux of glutamate (GLUT) in the reverse direction. The fluxes of malate (MAL) and a-ketoglutarate (a-KG) may be similarly coupled. OAA, oxaloacetate.

Glutamate is also the immediate metabolic precursor of nearly all GABA synthesized in neural tissues, which presumably occurs only in neurons that use GABA as a neurotransmitter (i.e., GABAergic neurons). Glutamate also can serve as an intermediate in the detoxification of ammonia.

Glutamate as a Neurotransmitter Glutamate is the principal excitatory neurotransmitter in the mammalian central nervous system (CNS). Most neurons that use glutamate as a neurotransmitter (termed either *glutamatergic* or *glutaminergic*) are projection neurons—exemplified by the pyramidal neurons that arise in the cerebral cortex and project to various subcortical regions or other cortical areas—somatic primary afferent sensory neurons, and ganglion neurons that arise in the retina and project to the lateral geniculate (Table 1.5-1). Glutamatergic interneurons include cerebellar granule cells, bipolar cells in the retina, and granule cells in the hippocampus. Glutamatergic neurons are particularly prevalent in the hippocampus, where they appear to have a significant role in memory formation.

Anatomical Pathway	Type of Neuron	Behavioral Role
Corticoamygdala	Pyramidal	Emotion
Corticocortical	Pyramidal	Association, perception
Corticonucleus tractus solitarius	Pyramidal	Regulation of breathing, heart rate, taste, olfaction
Corticospinal	Betz	Movement, pain sensation
Corticostriate	Pyramidal	Motor
Corticothalamic	Pyramidal	Relay, integration
Primary sensory afferents	Aδ, C fibers	Sensory input
Intercerebellar	Granule cells	Motor
Entorhinal-hippocampal	Perforant path	Learning, memory, emotion
Intrahippocampal	CA3 pyramidal cells	Learning, memory (Schaffer collaterals)
Dentate-hippocampal	Granule cells (mossy fibers)	Learning, memory, association
Hippocampal-tuberculoseptal-mammillary	CA1 pyramidal	Learning, memory, emotion
Interstitia	Photoreceptors, bipolar	Primary visual input
Retinotectal	Ganglion cells	Vision—sensory
Visual geniculocortico	Optic radiation	Vision—integration
Olfactory bulb-olfactory cortex	Mitral cells, tufted cells	Olfaction—integration
Cochlea	Hair cells, spiral cells	Primary auditory ganglion

*Pathways and behavioral roles are more complex than indicated.

Table 1.5-1 Principal Glutamatergic Pathways, Neuron Types and Behavioral Roles*

Aspartate as a Neurotransmitter Although aspartate has been shown to be a potent stimulatory factor in the CNS, its role in the brain is unclear. No specific receptors for aspartate have been identified, but it may be a physiologically relevant agonist (ligand) at some types of glutamate receptors. Compelling evidence indicates that some interneurons within the hippocampus and spinal cord concentrate aspartate within their synaptic terminals and release this amino acid during

membrane depolarization. However, it has not been established that aspartate is concentrated in specific types of synaptic vesicles released via exocytosis.

NEUROPHYSIOLOGICAL SIGNIFICANCE OF INHIBITORY AMINO ACIDS

GABA as a Neurotransmitter GABA is prevalent throughout the CNS but at comparatively low concentrations brainstem and spinal cord, where glycine concentration is high. GABA is generally not found in peripheral neurons, but it is detected in some endocrine tissues, including b islets in the pancreas and in ovaries. GABA is not an essential amino acid and is not used as a building block for protein, but it is a constituent of some peptides, such as homocarnosine, a dipeptide of GABA and histidine. GABA is now recognized as the major inhibitory neurotransmitter within the CNS, but actual proof of this awaited identification of specific receptors for GABA in the postsynaptic membranes adjacent to presynaptic terminals and a demonstration that vesicles within the synaptic terminals of neurons that synthesize GABA (GABAergic neurons) selectively accumulate and concentrate GABA. The latter criterion was essential because the release of neurotransmitter molecules from synaptic terminals is quantal, which implies that transmitter molecules are released from synaptic vesicles into the synaptic cleft by an exocytotic process.

Within the CNS GABA is synthesized only in select types of neurons, which are likely to GABAergic neurons. These are usually small interneurons with short axons, but some projection neurons, such as the Purkinje cells in the cerebellum and the striatonigral and pallidonigral neurons in the basal ganglia are GABAergic (Table 1.5-2). Physiologically, the most common function of inhibitory GABAergic neurons appears to be to focus and refine the firing pattern (nerve-impulse generation) of the projection neurons that transfer neural information from one functional unit to another. An example of this is surround inhibition, which occurs when excitatory projection neurons activate inhibitory interneurons via collateral axons; which in turn inhibit surrounding projection neurons via synaptic contacts on the cell soma near the axon hillock. GABAergic neurons can also facilitate the output of excitatory projection neurons by a process of disinhibition, which occurs when two GABAergic neurons are linked synaptically in series. The first inhibits the ability of the second to depress the activity of an excitatory neuron, thereby increasing the excitatory output. GABA also mediates presynaptic inhibition, which occurs when a small presynaptic terminal impinges on a larger one and inhibits the release of transmitter molecules from it.

Anatomical Location	Type of Neuron	Neural Systems Influenced
Cerebral cortex	Inhibitory interneurons	Cortical efferents
Basal ganglia	Neostriatal interneurons	Striatal efferents
	Striatonigral neurons	Extrapyramidal output
	Pallidonigral neurons	Extrapyramidal output
Cerebellum	Purkinje	Cerebellar output
	Basket cells	Cerebellar efferents
	Golgi cells	Cerebellar efferents
	Stellate cells	Cerebellar efferents
Hippocampus	Basket cells	Hippocampal efferents
Retina	Amacrine cells	Retina efferents
Olfactory bulb	Inhibitory interneurons	Olfactory bulb efferents
Brainstem, spinal cord	Inhibitory interneurons	Central integration of sensory efferents; motorneuron output

Table 1.5-2 Principal GABAergic Neuron Anatomical Locations, Neuron Types, and Neural Pathway Influenced

Glycine as a Neurotransmitter Glycine is present throughout the CNS but is more prevalent in the brainstem and spinal cord, which are the primary areas where it appears to serve a transmitter function. In all tissues, glycine is a major constituent of protein and several peptides such as glutathione. Virtually all criteria required to establish glycine as a neurotransmitter have been met. These include the identification of specific receptors in the postsynaptic membrane adjacent to presynaptic terminals and a demonstration that vesicles within the synaptic terminals of presumptive glycinergic neurons selectively accumulate and concentrate glycine. Anatomically, glycinergic neurons are usually small interneurons, predominantly located in the brainstem and spinal cord, and often functionally associated with a-motoneurons.

Taurine Taurine, a sulfonated amino acid, is prevalent throughout the CNS and many other tissues. Taurine is highly concentrated (10 to 20 mmol/L) in the immature brain and declines during synaptogenesis. It is also abundant in white matter and therefore does not have a tissue distribution expected of a neurotransmitter. There is no compelling evidence that taurine activates specific receptors localized to synaptic terminals or that it accumulates into synaptic vesicles. Therefore, although the inhibitory effects of taurine on neuronal excitability suggest that this amino acid does function to depress neuronal excitability, it is not likely to be involved in synaptic signaling.

AMINO ACIDS AS COTRANSMITTERS

Some glutamatergic and GABAergic neurons also use another agent, usually a peptide, as a neurotransmitter. For example, glutamatergic primary afferent neurons also release substance P. In addition, enkephalin and cholecystokinin have been colocalized with glutamate in the perforant path fibers of the hippocampus. Some glutamatergic neurons may also use aspartate or *N*-acetylaspartylglutamate as a cotransmitter. Cholecystokinin-8 (CCK-8), neuropeptide Y, and galanin are each colocalized with GABA in some cortical, limbic, and spinal neurons, respectively. Enkephalin and thyrotropin-releasing hormone (TRH) are also localized in some spinal GABAergic neurons. Some studies suggest that the peptide neurotransmitters are released from synaptic terminals primarily during periods of high-frequency neuronal firing.

BIOCHEMICAL PROCESSES THAT MEDIATE TRANSMITTER FUNCTION

Synthesis and Regulation of the Neurotransmitter Pool of Glutamate Glutamate is rapidly synthesized from glucose in neural tissues, including synaptic terminals, but this reflects the role of glutamate in energy metabolism. The biochemical process used to replenish the neurotransmitter pool appears to involve a net synthesis of glutamate precursors in astrocytes, which are released and subsequently taken up into glutamatergic synaptic terminals and converted to glutamate. Glutamine is the most firmly established precursor serving this function, but a-ketoglutarate and malate may also be used in this capacity (Fig. 1.5-2).

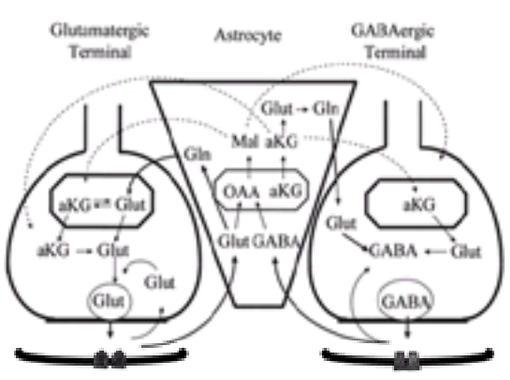

FIGURE 1.5-2 Some biochemical events associated with the neurotransmitter functions of glutamate and GABA. Many of the molecules of glutamate and GABA released from presynaptic terminals are transported into astrocytes and therein converted to glutamine and possible an intermediate of the tricarboxylic acid cycle, which are released and transported into the nerve terminals for subsequent conversion back into glutamate or GABA. MAL, malate, aKG, a-ketoglutarate.

The transmitter pool of glutamate may be defined as the molecules stored within synaptic vesicles. Glutamate is concentrated within vesicles by an ATP-dependent transporter specific for glutamate and present only in the membrane of vesicles in glutamatergic terminals. The vesicle transporter can concentrate glutamate approximately tenfold above the cytosolic level (~10 mmol/L). By comparison, the concentration of glutamate in the synaptic cleft when the synaptic terminal is inactive (resting state) is approximately 1 μmol/L. Sodium ion (Na^+)-Dependent transporters with high affinity for both glutamate and aspartate present in the

cytoplasmic membrane also contribute to the regulation of the transmitter function of glutamate and aspartate. Specific types of transporters are expressed in neurons and astrocytes. Transporters in astrocytes and in neuronal synaptic membranes contribute to neurotransmitter inactivation by maintaining low extracellular concentrations around synapses, whereas neuronal transporters in nonsynaptic regions may contribute to regulating the excitability of neurons by controlling the ratio of the intracellular to extracellular concentration within neurons. In this regard, although the transporters generally function as uptake systems (sometimes termed *reuptake*), they can also serve an export function during membrane depolarization.

Regulation of the neurotransmitter pool of aspartate is presumed to be similar to that for glutamate. A major difference is that relatively few neurons appear to contain synaptic vesicles with ATP-dependent transporters specific for aspartate. Therefore aspartate may function as a neurotransmitter in comparatively few neurons.

Synthesis and Regulation of the Neurotransmitter Pool of GABA Glutamate is the major metabolic precursor of GABA, although a small amount of GABA is derived from polyamines. Glutamate decarboxylase (GAD) catalyses the reaction, and two distinct forms of this enzyme exist—GAD I and GAD II. The enzyme requires pyridoxal phosphate (vitamin B_6) as a coenzyme, which also serves a regulatory function. GABA is synthesized from a pool of glutamate that is rapidly formed from a-ketoglutarate, a tricarboxylic acid cycle intermediate. Much of the GABA synthesized via this pathway is probably not in synaptic terminals and therefore is not within the transmitter pool. Although some of the GABA molecules formed via this pathway may be released into extracellular fluid via a transport process, most is likely converted immediately to succinate by the enzyme GABA transaminase. The primary metabolic pathway by which the transmitter pool of GABA is replenished appears to be one in which glutamine and possibly a-ketoglutarate are supplied by astrocytes to GABAergic synaptic terminals. These are transported into the synaptic terminals and metabolized to glutamate, which is then converted to GABA, and transported into the synaptic vesicles (Fig. 1.5-2).

Release of Transmitter Molecules From Synaptic Terminals Amino acid neurotransmitters are released from synaptic terminals by a process similar to that for acetylcholine and the monoamine neurotransmitters. Glutamate, GABA, glycine, and aspartate are each accumulated in their own specific type of synaptic vesicle via a Na^+-independent vesicular ATP-dependent transporter (pump). Neurotransmitters are concentrated within synaptic vesicles at concentrations about tenfold that the cytoplasmic concentration, resulting in an intravesicular concentration of nearly 100 mmol/L. Because the diameter of synaptic vesicles is about 10^{-7} meter, one vesicle can contain approximately 5000 molecules.

Transmitter release is triggered by depolarization of the presynaptic membrane by an action potential. This results in the opening of voltage-dependent calcium ion (Ca^{2+}) channels, allowing calcium to flow into the presynaptic terminal. The rise in intracellular Ca^{2+} concentration initiates a cascade of events involving at least eight proteins (mostly protein kinases) that enable the vesicles to dock at specific sites within the presynaptic membrane, whereupon the vesicles fuse with the membrane and release their contents into the synaptic cleft. Although the presynaptic terminal of neurons within the CNS may contain numerous vesicles, a single action potential typically induces transmitter release from only a few (1 to 3). This suffices to raise the concentration of neurotransmitter molecules within the synaptic cleft to nearly 1 millimeter for a few milliseconds, which will effectively saturates adjacent postsynaptic receptors.

Removal of Transmitter Molecules From the Synaptic Cleft (Transmitter Inactivation) Transmitter molecules within the cleft can readily diffuse into extracellular fluid adjacent to the synapse, where the concentration of transmitter molecules is at least 1000-fold lower. As long as the 1000-fold gradient is maintained, diffusion can restore the resting concentration within the cleft to micromolar concentrations within a few milliseconds. Uptake of transmitter molecules back into the presynaptic terminal or into the postsynaptic neuron also contributes to the inactivation process. A key element in the inactivation of the transmitter function of glutamate, GABA, and aspartate is the vigorous uptake of the amino acids by transporters in the cytoplasmic membrane of astrocytes. This is primarily how a low (£ μmol/L) extracellular concentration (£μmol/L) of these amino acids is maintained, thereby allowing diffusion to remove molecules from the synaptic cleft rapidly. The molecules of glutamate, GABA, and aspartate taken up by astrocytes are metabolized to glutamine and intermediates of the tricarboxylic acid cycle (Fig. 1.5-2).

AMINO ACID NEUROTRANSMITTER RECEPTORS

Glutamate Receptors Glutamate receptors are found throughout the brain and are expressed on both neurons and glia, although not all glutamate receptor subtypes are found on both cell types. Glutamate receptors, sometimes referred to as *excitatory amino acid receptors*, were initially classified into *N*-methyl-D-aspartate (NMDA), quisqualate, and kainate receptors on the basis of their preferential activation by these exogenous agonists. More recently, five categories of glutamate receptors (NMDA, kainate, a-amino-3-hydroxy-5-methylisoxazole-4-propionic acid [AMPA], L-2-amino-4-phosphonobutyrate (L-AP4), and trans-1-aminocyclopentane-1,3-dicarboxylic acid [ACPD] receptors) have been established on the basis of pharmacological, electrophysiological, and molecular biological criteria. The L-AP4 receptor type is defined by its agonist and acts as an inhibitory autoreceptor, while the quisqualate receptors of the previous classification have been subdivided by means of more-specific agonists into AMPA and ACPD receptors. AMPA and Kainate receptors are sometimes collectively referred to as non-NMDA receptors. NMDA, kainate and AMPA receptors are ionotropic glutamate receptors; the L-AP4 and ACPD receptors are grouped as metabotropic receptors. Ionotropic receptors are ligand-gated cation-specific channels that are activated rapidly (milliseconds), whereas metabotropic receptors coupled to G proteins and second-messenger systems function more slowly on a scale of several hundred milliseconds to seconds. Representative agonists and antagonists for each of these glutamate receptor classes are shown in Table 1.5-3.

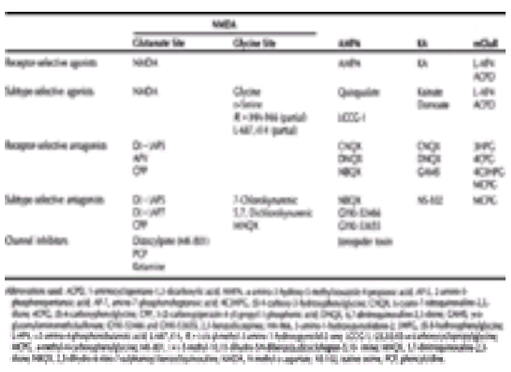

Table 1.5-3 Common Glutamate Receptor Agonists and Antagonists

AMPA Receptors Recent cloning efforts have clearly demonstrated that AMPA and kainate receptors are distinct receptor complexes, although they can be activated by the same agonists. Four genes encode the AMPA receptor (*GluR*1 through *GluR*4 [*GluR*-A through *GluR*-D]) and five genes encode the kainate receptor (*GluR*5 through *GluR*7 and *KA*1 and *KA*2). The four AMPA receptor subunits are similar in size and amino acid sequence. Each subunit exists in two different forms, “flip” and “flop,” created by alternative splicing. AMPA receptor subunits are expressed predominantly in the flip form in embryonic brains and gradually change over to the flop form, which dominates in the adult brain. The AMPA receptor channels are permeable to Na^+ and potassium (K^+) predominantly but will allow permeability to Ca^{+2} in the absence of the *GluR*2 (*GluR*-B) subunit. This however is only known to occur in a small population of hippocampal and neocortical nonpyramidal neurons in the brain. The AMPA receptor has at least three binding sites at which agonists or antagonists can interact: glutamate binding, desensitization and intra-ion-channel binding sites. Recently, a family of drugs called ampakines have been shown to selectively increase AMPA receptor-gated currents and improve cognitive performance in a series of behavioral tasks.

Kainate Receptors Although kainate is an effective agonist of AMPA receptors, it also activates its own distinct class of ionotropic receptors, the kainate-preferring receptors. The five subunits are divided into two groups; *GluR*5 through *GluR*7 represent the low-affinity kainate-binding site (K_d = 50 nmol/L), whereas *KA*1 and *KA*2 correspond to the high-affinity kainate-binding site (K_d = 5 nmol/L). Each group is of similar size and amino acid sequence identity, with the *KA*1 and *KA*2 subunits being slightly larger than the *GluR*5 through *GluR*7 subunits. Despite their wide distribution throughout the CNS, the physiological significance of kainate receptors remains largely unknown, although they have been implicated in fast glutamatergic transmission in hippocampal neurons. In addition to postsynaptic functions, kainate receptors act presynaptically on mossy fiber terminals within the hippocampus to generate epileptiform activity.

NMDA Receptors NMDA receptors mediate excitatory neurotransmission in the CNS in different ways from AMPA and kainate receptors. They are characterized by voltage-dependent block by magnesium ions (Mg^{2+}), a high permeability to Ca^{2+}, and slow gating kinetics. The NMDA receptor is a ligand-gated ion channel composed of two different subunits: *NMDAR*1 and *NMDAR*2. *NMDAR*1 can exist in seven splice variants, and there are four different genes encoding variants of *NMDAR*2 (2A, B, C, D). At present it is not clear how many R1 and R2 subunits are present in each functional NMDA receptor or if additional subunits exist, although

the receptor is thought to be pentameric. However, it is known that *NMDAR*1 serves as the core subunit of a functional NMDA receptor, with the *NMDAR*2 subunits acting as modulatory components of the receptor. Both *NMDAR*1 and *NMDAR*2 are required for receptor function.

NMDA receptors have a number of distinct recognition sites for endogenous and exogenous ligands, each with discrete binding domains. At present there are at least seven pharmacologically distinct sites through which compounds can alter the activity of this receptor (Fig. 1.5-3). Drugs that affect NMDA receptor function are divided into four groups, those acting at (1) the glutamate and NMDA recognition site, which is highly conserved on the NMDA NR2 subunits; (2) the strychnine-insensitive glycine binding site, where glycine is required as a coagonist for channel opening; (3) the intraion-channel binding site, where Mg^{2+} sits blocking ionic currents through the receptor at resting potentials; and (4) modulatory sites such as the redox modulatory site, the proton sensitive site, the zinc (Zn^{2+}) site, and the polyamine site.

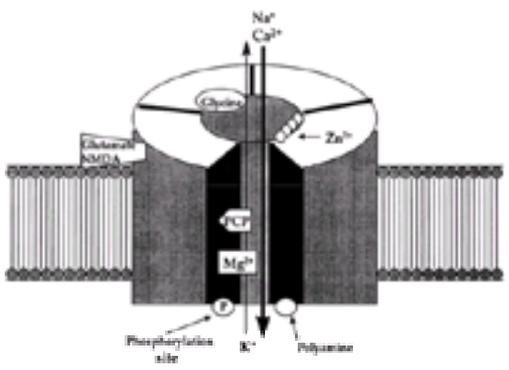

FIGURE 1.5-3 Diagrammatic illustration of the NMDA receptor–ion-channel complex and its modulatory sites. PCP, phencyclidine.

The NMDA receptor has three characteristic features: (1) at resting potentials it remains blocked by Mg^{2+}, and ionic currents flow through the receptor only if the neuronal membrane is partially depolarized; (2) significant amounts of extracellular Ca^{2+} enter the cell interior during activation of the receptor; and (3) NMDA receptor–mediated neurotransmission occurs slowly and lasts for a prolonged period. Because of these properties, it has been shown that the NMDA receptor serves a critical role in synapse development and plasticity, including the phenomena of long-term potentiation and long-term depression. Abnormal functioning of the NMDA receptor can lead to a variety of neurological disorders. Overactivation of the NMDA receptor has been indicated in ischemic insults, head trauma, and epileptic seizures, triggering a cascade of cellular events that culminate in neuronal cell death. On the other hand, hypofunction of the NMDA receptor elicits a psychomimetic state that closely resembles schizophrenia.

Metabotropic Receptors Not as much is known about the last group of glutamate receptors, the metabotropic receptors. The metabotropic receptor proteins belong to the superfamily of G-protein–coupled receptors, all of which contains seven-transmembrane domains. So far, the metabotropic receptor gene (*mGluR*) family has been shown to contain eight members, which are closely related in primary structure and can be divided into three groups on the basis of amino acid sequence homology, agonist sensitivity, and associated signal-transduction mechanisms. Group I receptors (mGlu1 and mGlu5) are coupled to the inositol (1,4,5) triphosphate-Ca^{2+} cascade, while group II (mGlu2 and mGlu3) and group III (mGlu4 and mGlu6 through mGlu8) lead to the inhibition of adenylate cyclase. Some members of the *mGluR* families exist in alternatively spliced variants. Several lines of evidence suggest that the N-terminal large extracellular domain of the receptors contains glutamate binding sites, while the C-terminal domain plays a role in determining the potency of agonists regulating the transduction mechanisms of the metabotropic receptors. Glutamate, quisqualate, and ibotenate activate both mGluRs and ionotropic receptors. Some glutamate analogues, such as ACPD and L-AP4, are specific for metabotropic receptors, while their potencies differ for each mGluR subtype (Table 1.5-3). Generally, metabotropic receptor agonists induce a slow membrane depolarization (a rise time of around 5 seconds and lasting up to 60 seconds, which is about 1000-fold slower than ionotropic receptors) accompanied by an increase in firing rate in many neurons. These effects are attributed to direct inhibitory effects on K^{+} channels. In addition to direct excitatory postsynaptic effects, metabotropic receptor activation suppresses both excitatory and inhibitor transmission at synapses by presynaptic mechanisms via an autoreceptor-type mechanism, thereby modulating presynaptic activity. Several metabotropic receptors have been implicated in synaptic plasticity that occurs in learning and memory.

GABA RECEPTORS

$GABA_A$ Receptor The $GABA_A$ receptor is a member of the superfamily of ligand-activated ion channels in the cell membrane. GABA type A ($GABA_A$) receptors are most closely related to strychnine-sensitive glycine receptors, more distantly related to acetylcholine nicotonic receptors and serotonin 5-hydroxytryptamine (5-HT) [5-HT] type receptors, and even more distantly related to glutamate ionotropic receptors (AMPA and kainate receptors and NMDA receptors). $GABA_A$ receptors are heteropentameric protein complexes, which when activated undergo a series of conformational changes that form an open channel (pore) selectively permeable to anions, specifically chlorine anion (Cl^-) and to a lesser degree (HCO^{-3}). Receptor activation normally results in an influx of Cl^- which rapidly and transiently hyperpolarizes the membrane, a process generally referred to as the generation of an inhibitory postsynaptic potential. The increase in Cl^- flux also decreases the resistance of the membrane, which acts as a shunt to impede the ability of depolarizing excitatory postsynaptic potentials to elicit action potentials (nerve impulses). For this reason, inhibitory synapses are most effective when located near the point at which action potentials are initiated, usually the axon hillock. Therefore, it is not surprising that GABAergic inhibitory synapses are often concentrated on neuronal cell bodies near the axon hillock.

$GABA_A$ receptors are heteromeric in that the receptor can comprise at least four types of subunit proteins, termed a, b, g, and d. It is pentameric in that each receptor has a total of five proteins; therefore all $GABA_A$ receptors have more than one copy of at least one type of subunit protein. There are multiple subtypes of three of the subunit proteins, at least six subtypes of a, three of b, and two of g. Although theoretically there could be many thousands of $GABA_A$ receptor subtypes, it is likely that fewer than 20 exist naturally. Most often the receptors contain two a, two b, and one g, or one a, two b, and two g subunits. The different subunits and the different subtypes of each subunit from which a particular type of receptor is formed can influence the physiological properties of the receptor (e.g., channel open time and rate of desensitization) as well as susceptibility to pharmacological agents as discussed below.

$GABA_A$ receptors are regulated by phosphorylation of some serine hydroxyl residues in the inner loop of the b subunits. Phosphorylation can be mediated by protein kinase A or protein kinase C. Depending on the type of b subunit, phosphorylation can affect the channel gating properties (e.g., channel open time and rate of receptor desensitization) either positively or negatively.

A variety of pharmacological agents can influence the activity of $GABA_A$ receptors. At least five separate drug binding sites have been identified (Fig. 1.5-4 and Fig. 1.5-5). Many clinically useful drugs are known to bind to benzodiazepine or barbiturate sites. These sites are allosteric to the GABA binding site. Drugs that bind to them influence the ability of GABA to activate the receptor by either altering the affinity between GABA and its binding sites ($GABA_A$ receptors probably possess two GABA binding sites) or by altering the channel open time and rate of receptor desensitization. It now seems unlikely that either of these sites serves a physiological purpose. An unusual characteristic of the benzodiazepine site is that drugs binding to it can exert either a positive modulatory (an agonist) or negative modulatory (an inverse agonist) effect, or no effect at all (an antagonist). The steroid site may have some yet-undetermined physiological relevance. Several of the pharmacological effects of ethanol are mediated through effects on $GABA_A$ receptors, but it may not act at a specific and unique site.

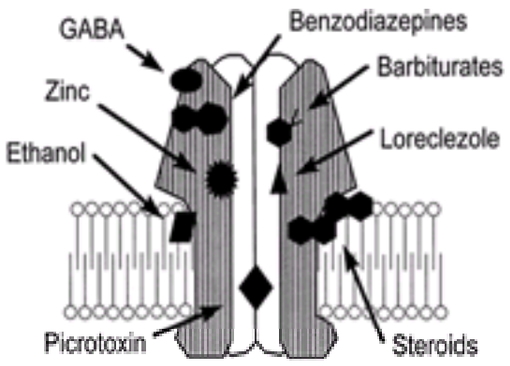

FIGURE 1.5-4 Diagrammatic illustration of the $GABA_A$, receptor–ion-channel complex and its modulatory sites.

FIGURE 1.5-5 Structure of the prototypical $GABA_A$ and benzodiazepine receptor agonists and antagonists. (Courtesy of Daniel C. Javitt, M.D., Ph.D., and Stephen R. Zukin, M.D.)

$GABA_B$ Receptors The metabotropic $GABA_B$ receptors are a member of the superfamily of G-protein-coupled receptors expressed in the cell membrane. These receptors generally exert an inhibitory effect on neuronal excitability by generating hyperpolarizing potentials that are much slower in onset and longer in duration than those mediated by $GABA_A$ receptors. Because $GABA_B$ receptors are G-protein coupled, receptor activation initiates guanosine triphosphate (GTP) hydrolysis and thereby causes dissociation of the G-protein subunits (a, b, g) from the receptor. $GABA_B$ receptors are coupled to a G_i subtype of G protein, and the dissociated a subunit activates a type of K^+ channel, thereby hyperpolarizing the membrane. $GABA_B$ receptors are often located on presynaptic terminals where they serve to inhibit transmitter release by reducing the efficacy of action potentials to activate Ca^{2+} influx.

CLINICAL CONSIDERATIONS

Glutamate and Aspartate Glutamate and aspartate function effectively as excitatory neurotransmitters because their extracellular and intracellular concentrations are tightly regulated by several biochemical processes, most of which require a constant source of energy (e.g., ATP). These include the vigorous transport systems in the membrane of astrocytes and neurons that rapidly remove glutamate from extracellular fluid (e.g., synaptic cleft) and transport systems within the synaptic vesicles from which the transmitter molecules are released into the synaptic cleft. Glutamate and aspartate can be rapidly metabolized, but only intracellularly. Therefore any condition that impairs the ability of the membrane transport systems to maintain these amino acids at very low extracellular levels can have a severe impact on the physiological activity of neurons.

Ischemia-Induced Neurodegeneration The events that cause neuronal cell death as a result of stroke, head trauma, or any condition involving a dramatic loss of oxygen or glucose supply to the brain are complex and not clearly elucidated. However, it is likely that a breakdown in the regulation of glutamate is a major factor. Early metabolic consequences of ischemia or hypoxia include accumulation of lactic acid, a concomitant decrease in pH, and a decrease in energy storage molecules (phosphocreatine and ATP). This has an immediate deleterious effect on the activity of the sodium-potassium ion pump, which accounts for more than 50 percent of ATP-supplied energy use in the CNS. This in turn results in a dissipation of transmembrane gradients for K^+ and Na^+ and a concomitant depolarization of the cell membrane. Dissipation of the Na^+ gradient inhibits the removal of glutamate from extracellular fluid, and membrane depolarization activates voltage-sensitive Ca^{2+} channels in synaptic terminals, thereby promoting excessive glutamate release. Consequently, high levels of glutamate accumulate in the synapse, causing excessive activation of NMDA and AMPA receptors. Because of the prevalence of these receptors, the intracellular accumulation of Ca^{2+} is greatly exacerbated. This pathological accumulation of Ca^{2+} promotes a cascade of events that can result in neuronal cell death. Neurological disorders in which this pathological cascade may be involved include global and focal ischemia (stroke) but also hypoglycemia, head trauma, spinal injury, status epilepticus, drug abuse, and certain food toxicities (e.g., monosodium glutamate, and mussel poisoning).

Chronic Neurodegenerative Disorders Dysregulation of glutamate and aspartate and overactivation of their receptors may contribute neuronal cell loss in chronic disorders such as acquired immune deficiency syndrome (AIDS) dementia, Parkinson's disease, motor neuron disease (including amyotrophic lateral sclerosis), Huntington's disease, and Alzheimer's disease. Tissue-specific defects in glial transporter genes resulting in impaired glutamate uptake have been identified in several cases of the sporadic form of amyotrophic lateral sclerosis. Lathyrism, which is clinically similar to amyotrophic lateral sclerosis, has been linked to the ingestion of cycad beans, which contain an excitotoxin b-*N*-methylamino-L-alanine. Abnormal activation of excitatory pathways within the basal ganglia appears to play a part in the symptom expression of parkinsonism in animal models. In primates, NMDA and non-NMDA antagonists increase the therapeutic efficacy of the dopaminergic drug, levodope (Larodopa).

Epilepsy Epilepsy is a group of neurological disorders characterized by spontaneous recurrent seizures. A seizure is an abnormal paroxysmal firing of cerebral neurons in synchronous fashion and is often associated with motor signs and sensory, autonomic, or psychic symptoms. Loss or impairment of consciousness often occurs. Epileptic syndromes are defined on the basis of clusters of signs and symptoms that generally occur together in a patient with recurrent seizures. They are classified on the basis of the seizures: localization related, generalized, undetermined, or a special syndrome (1989 International League Against Epilepsy classification). The epileptic syndromes are further divided by etiology: idiopathic, cryptogenic, or symptomatic. Epilepsy occurs in 1 to 2 percent of the population worldwide, and epileptic patients account for a major proportion of the return visits in neurological clinics. Although compounds that antagonize the action of glutamate at NMDA receptors or AMPA and kainate receptors are generally effective in blocking seizures, the only marketed antiepileptic drug in which glutamate antagonism is thought to be a prominent component of its antiseizure activity is topiramate (Topamax). Topiramate selectively antagonizes AMPA and kainate receptors.

Although many neurobiological factors may contribute to seizure formation, a prominent feature of most seizures is an abnormal and excessive firing of glutamatergic neural pathways. Therefore, abnormalities in the regulation of glutamate may be a factor in the initiation, spread and maintenance of seizure activity in some types of epilepsy. The involvement of glutamatergic receptors in seizures and epilepsy is widely accepted on the basis of evidence that injections or focal applications of glutamatergic agonists at NMDA receptor or AMPA and kainate receptors seem to produce seizures or epileptic-like activity in numerous in vitro and animal models of epilepsy. Furthermore, selective antagonists to these receptors reduce epileptic activity or are potent anticonvulsants in several models. Elevated plasma concentrations of glutamate have been observed in a epileptic-like mouse strain and in probands with generalized epilepsy. Some studies using in vitro or ex vivo techniques and studies in animal models indicate that glutamate or aspartate release increases during seizure activity. A study using chronically implanted microdialysis probes in bilateral hippocampi of epileptic humans showed marked increases in extracellular glutamate concentrations immediately prior to the onset of seizures on the side of seizure origin.

Many patients with temporal lobe or complex partial epilepsy have neuronal loss and sclerosis, particularly in mesial hippocampus. Mesial temporal sclerosis is a common finding in surgical specimens removed from patients suffering from chronic refractory complex partial seizures. Neuronal loss is prominent in CA1, CA3 and

dentate hilus. Interestingly, the pathology closely resembles the findings seen in model systems of prolonged seizure activity or in those induced by application of excitotoxins.

Kindling, which is a gradual induction of a hyperexcitable neuronal state, can occur by focal repetitive subconvulsive stimulation of the hippocampus, amygdala, or some other brain areas. Kindling results in increased susceptibility toward seizures and has been studied extensively in animals, particularly rodents. The basis for the persistent hyperexcitable state remains controversial, and the actual involvement of excitatory amino acid receptors in this hyperexcitable state remains unclear. However, there is good evidence that glutamatergic receptors, particularly NMDA types, play a role in the development and enhancement of the kindled or seizure-prone state. NMDA receptor antagonists can prevent the kindling phenomenon despite the expression of seizurelike discharges in in vitro models such as hippocampal slices. A variety of NMDA antagonists, including those that act as channel inhibitors or compete with the glutamate recognition or glycine recognition sites, appear to be highly effective in blocking the development of kindling. They do not appear to be as effective as anticonvulsants in fully expressed seizures unless they are used at dosages that produce significant toxic effects, such as neurological or behavioral impairment. In contrast, AMPA receptor antagonists are highly effective in blocking the expression of seizures, but they seem to have little effect on the induction of the kindled state. Further research is needed to define the role of excitatory amino acids in kindling and hippocampal injury. Much of the research has focused on ionotropic glutamate receptors. With the recent discovery of more-selective metabotropic glutamate receptor agonists and antagonists, along with the cloning of metabotropic receptor genes, the role of metabotropic receptors is now being explored.

Schizophrenia The cause of schizophrenia remains elusive, although several observations suggest that it may involve abnormalities in glutamatergic transmission. This hypothesis has been spurred on by the knowledge that phencyclidine (PCP, “angel dust”), a noncompetitive NMDA receptor antagonist, generates psychomimetic behaviors similar to those seen in schizophrenia. Clinical data supporting a role for glutamate in the etiology of schizophrenia are sparse but interesting. In the 1980s it was observed that glutamate concentration was lower in the cerebrospinal fluid (CSF) of schizophrenic patients than in that of control patients. These results have not been replicated successfully. Postmortem studies of schizophrenics have shown a consistent increased density of glutamate receptors (NMDA and non-NMDA) within the frontal and temporal cortex, together with some evidence for decreased production or release of glutamate, or both. Such results suggest that schizophrenia involves decreased glutamatergic transmission.

Neuropathic Pain Activation of afferent C fibers with nociceptive stimuli produces pain sensations that are enhanced during pathological conditions. Activity-dependent increases in excitability are induced in the spinal dorsal horn neurons by repetitive stimulation of C fibers. This is thought to contribute to the development and maintenance of chronic pain symptoms. NMDA antagonists, ketamine (Ketalar) and D-amino-propyl-valeric acid, have consistently reduced this activity in the rat dorsal horn nociceptive neurons, suggesting that the NMDA receptor contributes to this phenomenon. AMPA and kainate receptors may also play a role in modulating pain. In situ hybridization studies have revealed that expression of kainate receptor gene *GluR*5 is particularly prominent in dorsal root ganglion neurons. Animal studies have indicated that kainate receptor antagonists significantly reduce nociception, and early human trials with some of these agents show them to be promising for analgesia.

Substance Abuse Several lines of evidence indicate that one of the acute effects of ethanol is to inhibit glutamate receptor function, particularly NMDA and kainate receptors. Such inhibition leads to depressed synaptic transmission and may result in ethanol-induced cognitive deficits. Indeed, low concentrations of ethanol are known to inhibit long-term potentiation in the hippocampus.

GABA

Epilepsy Like the glutamatergic system, GABA and GABA receptor subtypes play a central role in the expression of seizures. GABA, as the major inhibitory neurotransmitter in the CNS, can be found in up to 30 percent of CNS synapses. In general, for the mature brain, loss or blockade of GABA inhibition can result in increased hyperexcitability and expression of seizures. Unlike with the glutamatergic system, several GABAergic drugs are widely used in the treatment of epilepsy. Clinically effective benzodiazepines and barbiturates likely act at $GABA_A$ receptors to enhance inhibition. Both have been shown to be effective in the control of partial, complex partial, and generalized tonic-clonic seizures. Benzodiazepines are also effective in the short-term treatment of generalized absence, but a functional tolerance tends to develop, thereby reducing their efficacy. Barbiturates may exacerbate generalized absence seizures. Benzodiazepines are also effective in the treatment of atypical absence and myoclonic seizures. Anticonvulsant benzodiazepines and barbiturates are highly sedating, which limits their use. Phenobarbital (Solfoton) an anticonvulsant barbiturate, has been used for the treatment of epilepsy since the early 1900s, and it is still the most widely used anticonvulsant in the world, primarily because of its low cost.

Other evidence that the GABAergic system is important in the expression of seizures is that manipulation of $GABA_A$ receptor function can cause, exacerbate, or reduce seizure activity. The mushroom poisons picrotoxin and bicucculine antagonize $GABA_A$ receptors noncompetitively and competitively, respectively, and can elicit seizures. Penicillin given at high doses (especially in renal failure patients or intrathecally) can result in partial or generalized seizures. Penicillin reduces GABA-induced chloride current flow by blocking the ion channel pore. The $GABA_A$ receptor has numerous modulatory sites that can allosterically increase or decrease the chloride ion channel current flow. For example, negative modulators that act at the benzodiazepine site (e.g., b-carbolines) or the steroid site (e.g., pregnenolone sulfate) can lower seizure thresholds.

In the mature or adult brain, enhancement of $GABA_A$ receptor function generally raises the seizure threshold. Certain naturally occurring and synthetic pregnane-derived steroids are potent positive modulators, now called neuroactive steroids, or neurosteroids. Some data support the role of the endogenous neurosteroids in cyclical changes in behavior and in catamenial or menstrually related epilepsy. A pregnane-derived synthetic neurosteroid is in clinical trials for treatment of epilepsy.

Not only does positive modulation of these allosteric sites result in anticonvulsant activity, increases in GABA availability also seem to be of clinical benefit. Reduction of GABA clearance by inhibition of GABA uptake or reduction of GABA degradation by poisoning GABA transaminase are both effective. A GABA uptake inhibitor has been recently approved for treatment of partial seizures, but it may exacerbate generalized absence seizures. An irreversible inhibitor of GABA transaminase also effective in the treatment of partial seizures is in the late stages of development. Finally, some evidence in animal models suggests that modulation of $GABA_B$ receptors may play a role in the treatment of generalized absence seizures.

Anesthesia Pentobarbital (Nembutal) has been a popular drug for induction of anesthesia. Like phenobarbital, pentobarbital allosterically enhances $GABA_A$ receptor function, but over a narrow concentration range it also can directly activate $GABA_A$ receptors. Pentobarbital also has activity in blocking glutamate receptors and voltage-gated calcium channels. The potent benzodiazepines midazolam (Versed) and lorazepam (Ativan) have replaced diazepam (Valium) as a drug of choice for induction of anesthesia.

In the 1990s pentobarbital was replaced by propofol (Diprivan), which is more easily titrated and is especially useful in the neurosurgical setting. Propofol, which has a short duration of action, allosterically enhances $GABA_A$ function and directly activates these receptors. Several neurosteroids also can directly activate $GABA_A$ receptors, and it has been proposed that this direct effect is in part responsible for the anesthetic qualities of these agents.

In the 1940s, parenteral administration of cholesterol was shown to induce sedation in animals, and this subsequently led to the development of synthetically derived steroids for sedation. Although its mechanism of action was unknown at the time, one derivative, alfaxalone (Alfathesin) was a popular agent during the early 1970s for the induction of anesthesia. It has been subsequently discovered that alfaxalone's pregnane-based structure is similar to that for the neurosteroids and that it is a potent enhancer of $GABA_A$ receptor function.

Anxiety- and Sleep-Related Disorders Benzodiazepines and barbiturates have a long history in the treatment of anxiety and insomnia. Diazepam was once the most prescribed medication in the United States. With the cloning of GABA receptor genes, it is now recognized that the hints of subtypes of $GABA_A$ receptors from previous pharmacological experiments are true. This is especially true for subclasses of benzodiazepine-sensitive receptors. Benzodiazepines with greater anxiolytic than sedative-hypnotic properties are in use clinically, and their efficacy may be related to preferential binding to particular subtypes of $GABA_A$ receptors.

Spasticity Loss of spinal and supraspinal inhibition may result in spasticity or hyperreflexic states. One particular disorder, stiff person syndrome, is associated with increased reflexivity and muscle spasms and occasionally with seizures, diabetes or both in some patients. The disorder is frequently associated with circulating antibodies to GAD, the GABA synthesis enzyme. Interestingly, antibodies to GAD are associated with type I diabetes (autoimmune-associated juvenile form). Benzodiazepines, especially diazepam, and baclofen are mainstays in the treatment of spasticity. However, these agents are often only moderately effective,

especially in supraspinal forms of spasticity.

Substance Abuse Ethanol enhances GABA receptor function in some in vitro preparations potentially via a protein-binding site. Although studies have been controversial, many accept that ethanol may selectively alter function of some subtypes of GABA receptors. $GABA_A$-active benzodiazepines and barbiturates are known for their development of tolerance and potential addictive activity during long-term administration.

Glycine Glycinergic neurotransmission is important in the circuits for local inhibitory control in the spinal cord. Genetic defects in glycine subunit genes have been identified as causes of hypersensitive reflexes and spasticity in both humans and animals. Opportunities for pharmacotherapeutic modulation of glycine receptors have been rather limited by the lack of readily identifiable allosteric regulatory sites. Strychnine, a potent antagonist, has long been used as a poison for rodents and in humans, where it has been used as an instrument of malicious intent. Clinically relevant positive modulators of the glycine receptor include ethanol and the anesthetic propofol.

SUGGESTED CROSS-REFERENCES

Further information about the neuroanatomy of specific excitatory and inhibitory projections can be found in Section 1.2 on neuroanatomy. Further information on the receptor transduction mechanisms can be found in Section 1.9 on electrophysiology. Information regarding the contributions of specific cortical regions and pathways in schizophrenia can be found in Section 12.3 on brain structure and function and Section 12.4 on neurochemical, vital, and immunological studies in schizophrenia. The role of GABA benzodiazepine receptors in mood disorders is discussed in Chapter 14, and their role in anxiety disorders is discussed in Chapter 15. The clinical use of benzodiazepines is discussed in Section 31.10. Epilepsy is covered in Section 2.4.

SECTION REFERENCES

*Barnard EA, Skolnick P, Olsen RW, Mohler H, Sieghart W, Biggio G, Braestrup C, Bateson AN, Langer SN: International Union of Pharmacology. XV. Subtypes of g-aminobutyric acid A receptors: Classification on the basis of subunit structure and receptor function. Pharmacol Rev *50:*291, 1998.

*Betz H, Kuhse J, Fischer M, Schmieden V, Laube B, Kuryatov A, Langosch D, Meyer G, Bormann J, Rundstrom N, Matzenbach B, Kirsch J, Ramming M: Structure, diversity and synaptic localization of inhibitory glycine receptors. J Physiol (Paris) *88:*243, 1998.

Bigge CF, Boxer PA, Ortwine DF: AMPA/kainate receptors. Curr Pharm Design *2:*397, 1996.

Cacabelos R, Takeda M, Winblad B: The glutamatergic system and neurodegeneration in dementia: Preventive strategies in Alzheimer's disease. Int J Geriatr Psychiatr *14:*3, 1999.

Cooper JR, Bloom FE, Roth RH: Amino acid neurotransmitters. In *The Biochemical Basis of Neuropharmacology,* ed 7. Oxford University Press, New York, 1996.

Dingledine R, Borges K, Bowie D, Traynelis SF: The glutamate receptor ion channels. Pharmacol Rev *51:*7, 1999.

Javitt DC, Zukin SR: Amino acid neurotransmitters. In *Comprehensive Textbook of Psychiatry/VI,* vol 1, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Kvamme E, editor: *Glutamine and Glutamate in Mammals*. CRC Press, Boca Raton, FL, 1988.

*McGeer PL, McGeer EG: Amino acid neurotransmitters. In *Basic Neurochemistry: Molecular, Cellular, and Medical Aspects*, GJ Siegel, BW Agranoff, RW Albers, PB Molinoff, editors. Raven, New York, 1989.

*Michaelis EK: Molecular biology of glutamate receptors in the central nervous system and their role in excitotoxicity, oxidative stress and aging. Prog Neurobiol *54:*415, 1998.

Millhorn DE, Hokfelt T: Chemical messengers and their coexistence in individual neurons. News Physiol Sci *3:*1, 1988.

*Newcomer JW, Farber NB, Jevtovic-Todorovic V, Selke G, Melson AK, Hershey T, Craft S, Olney J-W: Ketamine-induced NMDA receptor hypofunction as a model of memory impairment and psychosis. Neuropsychopharmacology *20:*106, 1999.

Ozawa S, Kamiya H, Tsuzuki K: Glutamate receptors in the mammalian central nervous system. Prog Neurobiol *54:*581, 1998.

Seeburg PH: The TINS/TIPS Lecture: The molecular biology of mammalian glutamate receptor channels. Trends Neurosci *16:*359, 1993.

Shepherd GM: *Neurobiology,* ed 3. Oxford University Press, New York, 1994.

*Tamminga CA: Schizophrenia and glutamatergic transmission. Crit Rev Neurobiol *12:*21, 1998.

1.6 NEUROPEPTIDES: BIOLOGY AND REGULATION

MICHAEL J. OWENS, PH.D., CHARLES B. NEMEROFF, M.D., FH.D. AND GARTH BISSETTE, PH.D.

The past several decades have witnessed a veritable explosion of knowledge about the central nervous system (CNS), and in no area has this been as impressive as in peptide neurobiology. Numerous peptide neurotransmitter candidates have been identified and characterized, their CNS distributions mapped, and their genes cloned. The tenet “one neuron-one transmitter” erroneously attributed to Dale has been convincingly refuted with numerous demonstrations of neurons containing multiple peptides or combinations of peptide and nonpeptide neurotransmitters. Additionally, since the early 1980s there has been an embarrassment of riches in the form of knowledge about neurotransmitter receptor diversity, diversity of receptor-effector coupling, and neurotransmitter transporters. These discoveries have not yet been fully integrated into what is known about normal or aberrant CNS function, although dysfunction at virtually any level could conceivably lead to neuropsychiatric deficits. This chapter summarizes, with several examples, the general principles governing neuropeptide biology and includes a review of several neuropeptides that have been unequivocably demonstrated to be involved in the pathophysiology of one or another neuropsychiatric disorders.

By definition, a neuropeptide is a chain of two or more amino acids linked by peptide bonds, and differs from other proteins only in the length of the amino acid chain. Over 100 unique biologically active peptide sequences have been purified from biological sources; their sizes range from two (carnosine and anserine) to over 40 amino acids (corticotropin-releasing factor [CRF] and growth hormone-releasing factor [GRF]) (Table 1.6-1). Most of the other known active peptides fall within these size limits. By convention, peptides greater than 90 amino acids in length (about 10,000 molecular weight) are considered proteins. The five examples discussed here are somatostatin (somatotropin release-inhibiting factor [SRIF]), thyrotropin-releasing hormone (TRH), neurotensin neuromedin N, and CRF. Their structures are illustrated Table 1.6-2, and are written by convention from the amino terminus (NH_2–) beginning on the left to the carboxyl terminus (–COOH) on the right.
Somatostatin, CRF, and TRH are hypothalamic hypophysiotropic hormones and, like neurotensin and neuromedin N, function in the CNS as neurotransmitters in ways that are often quite distinct from their effects on the endocrine axes. Neuropeptides have been implicated as chemical mediators in pathways subserving a variety of behavioral and physiological effects, including such diverse behaviors as thermoregulation, food and water consumption, sex, sleep, locomotion, memory, learning, responses to stress and pain, and emotion. Those actions have stimulated interest in the contribution of the peptidergic neuronal systems to the symptoms and behaviors exhibited in such major psychiatric illnesses as psychotic disorders, mood disorders, and dementia. In addition to their endocrine and neurotransmitter roles, many peptides and their receptors apparently play active roles in development and often appear transiently in various anatomical regions or in such abundance that a trophic effect is postulated, an effect that does not necessarily persist beyond early development. An example of this ontogenetic shift is found in the distribution of messenger ribonucleic acid (mRNA) for SRIF in developing rats. The mRNA exhibits more regional heterogeneity in developing rats than it does in adult animals, with the hypothalamic mRNA concentrations being much higher than the cerebrocortical expression.

Adrenocorticotropin hormone (ACTH)
Angiotensin
Atrial natriuretic peptide
Bombesin
Calcitonin
Calcitonin gene-related peptide (CGRP)
Cholecystokinin (CCK)
Corticotropin-releasing factor (CRF)
Dynorphin
β-endorphin
Leu-enkephalin
Met-enkephalin
Galanin
Gastrin
Gonadotropin-releasing hormone (GnRH)
Growth hormone
Growth hormone-releasing hormone (GHRH)
Insulin
Motilin
Neuropeptide Y
Neurotensin
Orphanin FQ/Nociceptin
Oxytocin
Pancreatic polypeptide
Prolactin
Secretin
Somatostatin (SRIF)
Substance K
Substance P
Thyrotropin-releasing hormone (TRH)
Vasoactive intestinal polypeptide (VIP)
Vasopressin

Table 1.6-1 Selected Neuropeptide Transmitters

Name	Amino Acid Sequence
Thyrotropin-releasing hormone (TRH)	pGlu-His-Pro-NH_2
Somatostatin (SRIF-28)	NH^2-Ser-Ala-Asn-Ser-Asn-Pro-Ala-Met-Ala-Pro-Arg-Glu-Arg-Lys-Ala-Gly-Cys-Lys-Asn-Phe-Phe-Trp-Lys-Thr-Phe-Thr-Ser-Cys-OH
Neurotensin	pGlu-Leu-Tyr-Glu-Asn-Lys-Pro-Arg-Arg-Pro-Tyr-Ile-Leu-OH
Neuromedin N	NH_2-Lys-Ile-Pro-Tyr-Ile-Leu-OH
Corticotropin-releasing factor (CRF)	Ser-Glu-Glu-Pro-Pro-Ile-Ser-Leu-Asp-Leu-Thr-Phe-His-Leu-Leu-Arg-Glu-Val-Leu-Glu-Met-Ala-Arg-Ala-Glu-Gln-Leu-Ala-Gln-Gln-Ala-His-Ser-Asn-Arg-Lys-Leu-Met-Glu-Ile-Ile-NH_2

Subscript numbers refer to points of known peptidase cleavage
1. Pyroglutamate aminopeptidase (E.C. 3.4.19.3)
2. Prolyl endopeptidase (E.C. 3.4.21.26)
3. Carboxypeptidase (E.C. 3.4.17.10)
4. Metalloendopeptidase (E.C. 3.4.24.15)
5. Neutral endopeptidase (E.C. 3.4.24.11)
6. Metalloendopeptidase (E.C. 3.4.24.16)

Table 1.6-2 Selected Neuropeptide Structures

Many of the known behavioral effects of neuropeptides are observed only after their direct injection into the CNS because most peptides do not penetrate the blood-brain barrier in amounts sufficient to produce effects before being inactivated by serum and tissue enzymes that degrade them. The degradation is usually the result of the cleavage of specific amino acid sequences targeted by a specific peptidase designed for that purpose. Peptide fragments often possess full or partial biological activity at peptide receptors, and the relatively lengthy period of peptide-induced effects may be attributable to this delayed cessation of activity. How much neuropeptide must be injected to elicit a physiological or behavioral effect is debatable because endogenous neuropeptides exist in picomolar to femtomolar concentrations whereas nanomolar concentrations are often required to produce such effects. Such differences should be expected, however, when it is realized that the anatomical substrate mediating the behavior or effect may be quite far from the site of injection, that the preexisting endogenous peptide signal and its regulatory and feedback systems must be overwhelmed, that a gauntlet of peptidases must be negotiated, and that synaptic concentrations may indeed reach high nanomolar concentrations. Another disconcerting fact about neuropeptides is the propensity for a property or effect to be elicited by one or more apparently unrelated peptides. Although often frustrating and potentially confounding to researchers, that property can be explained by the ability of more than one sequence of amino acids to assume similar three-dimensional shapes and charge distributions, all of which may be recognized by the receptor, by the presence of multiple receptor types contained on neurons initiating the behavior, and by possible similarities between second-messenger transduction mechanisms activated by discrete and distinct peptide receptor types. The tertiary structure for recognition is also used by the immune system for the production of specific antibodies, as well as by biological receptors. Because both descriptive (immunohistochemical) and quantitative (radioimmunoassay) methods for detecting the small concentrations of endogenous peptides rely on immunological techniques, it is always possible that a similarly shaped structure could confound the interpretation of immunoassay results. Thus, descriptions of antisera specificity and amount of cross-reactivity with other potential ligands are crucial to the proper evaluation of results. The dichotomy in methodology for detecting neuronal localization of peptide versus regional peptide concentration is also seen in the tools for peptide receptor identification by

autoradiography (localization) and regional membrane preparations for receptor quantification (“grind and bind”). The analogous procedures for detecting peptide mRNA are in situ hybridization (localization), Northern blot analysis, ribonuclease protection assay, and the polymerase chain reaction for regional mRNA quantification. Measurements of neuropeptide concentration changes do not indicate which of several mechanisms may be mediating the observed changes. Increases in concentration may represent increased synthesis and release, decreased release with continued synthesis, or decreased degradation. Attempts to verify peptide turnover may be made if the mRNA concentration, peptide concentration, receptor up-regulation or down-regulation, and degradative activity are known. Although methods to achieve each of these goals are now available, they have not generally been applied in combination to the same tissue sample.

Whereas the differences between neuropeptides and the classic monoamine and amino-acid neurotransmitters are often striking, their CNS effects are similar in that they primarily excite or inhibit discrete neurons upon direct application. Because both these effects may be observed among neurons from the same region, effects observed in one location cannot be generalized to either immediate or distal neurons. However, the onset of activity is often delayed for neuropeptides (seconds) as compared with the classic transmitters (milliseconds), whereas the duration of activity can be relatively delayed for neuropeptides (minutes) as compared with most of the classic transmitters (seconds).

DISTRIBUTION

Neuropeptides are found throughout the CNS, as well as in various peripheral organs, such as the gastrointestinal tract, pancreas, and adrenal glands. Many CNS peptides, such as neurotensin and SRIF, play dual roles in the brain and gut. The full extent of any communication between the CNS and gut systems employing the same peptide is not known with certainty, but may be considerable. Although CNS neuropeptides are found predominantly in neurons, peptide receptors have been reported in glia.

Neuropeptides were originally purified from hypothalamic extracts and thus it is not surprising that some of the highest concentrations of certain neuropeptides are found in the hypothalamus. This is true for all four of the example peptides: TRH, CRF, SRIF, and neurotensin. These many other neuropeptides are also widely distributed in extrahypothalamic brain areas and may occur in either intrinsic interneurons or in longer projection neurons. Immunohistochemical and retrograde tracing studies have focused on the locations and morphological types of neurons containing SRIF. In the hypothalamus most of the SRIF-containing neurons that project to the median eminence have been shown to emanate from cell bodies mainly in the rostral periventricular nucleus, with some in the paraventricular nucleus and none in the arcuate nucleus. Thus, the other hypothalamic regions (arcuate, suprachiasmatic, ventromedial) containing SRIF neurons probably do not project to the median eminence and may perform a regulatory or feedback function on neurons containing other hypothalamic releasing factors, such as GRF, CRF, TRH, or their afferents. In the cortex of rats SRIF is found in some of the large stellate-shaped neurons and in abundance among the fusiform-shaped, nonpyramidal neurons of layers II to V, and particularly in layer V of the sensory cortex. In monkeys, however, layer III is where SRIF is predominantly located in visual, auditory, or association cortex, and cortical neurons containing SRIF are usually oriented vertically in layers II to V and horizontally in layer VI. In human entorhinal cortex, SRIF neurons are abundant in the white matter underlying the cortex and are relatively uniformly distributed throughout the cortical layers, being absent only in the outer molecular layer. A recent comparison of human and monkey distribution of the prosomatostatin-derived peptides SRIF-28, SRIF-14, and SRIF-28 $_{1\text{-}12}$ in the prefrontal cortex found reduced staining of SRIF-28 in unperfused monkeys and human brain obtained 5 hours after death, indicating that the processing of peptides may continue after death. In the rat striatum SRIF is extensively colocalized in neuropeptide-Y–containing neurons, and although not influenced by lesions of the dopamine neurons innervating the striatum, SRIF concentrations are increased by the destruction of cortical inputs. In the human hippocampus, SRIF neurons are arranged in a manner similar to that in rats and nonhuman primates, with cell bodies in the deep layers of the dentate gyrus projecting fibers to the outer two thirds of the molecular layer. In the basal forebrain region of nonhuman primates, SRIF is contained within small neurons of the nucleus basalis of Maynert, which apparently communicate with the cholinergic neurons there. Those data indicate that the neuronal cell types and afferent and efferent connections of SRIF-containing neurons vary widely among different brain regions and that species differences occur with enough frequency to render direct extrapolations between species difficult.

Many neuropeptides have now been shown to be colocalized in neurons that also contain classic transmitters, other neuropeptides, or both. Neurotensin is found in neurons containing the dopamine synthetic enzyme tyrosine hydroxylase in the ventral tegmental area and arcuate nucleus of the hypothalamus of rats. Neurotensin is found in dense-core vesicles only in tyrosine hydroxylase-positive staining cell bodies in ventral tegmental area. Other nerve terminals that were stained for neurotensin did not contain tyrosine hydroxylase. These findings are discordant with other evidence showing decreases in both dopamine and neurotensin concentrations after reserpine and the dual release of frontal cortex dopamine and neurotensin after electrical stimulation of the median forebrain bundle. Another subset of ventral tegmental area neurons projecting to the frontal cortex has been shown to contain neurotensin, cholecystokinin (CCK), and tyrosine hydroxylase. Many other examples of colocalization have been cited, including reports of three to six peptides in a single neuron. TRH colocalization with another peptide, substance P, and a classic transmitter, serotonin, has been described in a population of neurons on the median raphe nucleus and spinal cord. Corticotropin-releasing factor has been reported to be colocalized with three other neuropeptides (vasopressin, oxytocin, and neurotensin) in some neurons of the hypothalamic paraventricular nucleus in both rats and humans. Somatostatin has been found in g-aminobutyric acid (GABA) neurons of the thalamus of cats and the cortex of rats, and with neuropeptide Y in the striatum, hippocampus, and cortex. Colocalization reports may be more species specific than is generally realized, as the neurotensin-dopamine colocalization in the ventral tegmental area and frontal cortex was clearly demonstrated to be absent in humans and nonhuman primates.

The two main methods for mapping peptides, immunohistochemistry and radioimmunoassay, are complementary in their determination of neuropeptide locations and concentrations, respectively, but they do not indicate which immunoreactive neuronal cell bodies are connected to the various immunoreactive terminal fields. Through the use of retrograde tracing methods and dual staining techniques, several pathways for certain peptides have now been delineated. They include projections of amygdala neurons containing neurotensin, SRIF, or CRF to the parabrachial nucleus of the mesencephalon, and a neurotensin-containing projection from the lateral parabrachial nucleus back to the central amygdala has also been described. Two other neurotensin projections that have been observed in rats are those from the ventral tegmental area to the nucleus accumbens and from the endopiriform nucleus and prepiriform cortex to the diagonal band area. This methodology has also led to the identification of TRH neurons in the paraventricular nucleus and bed nucleus of the stria terminalis as the origin of projections to the median eminence and septum, respectively. Lesion studies with excitotoxic amino acids or electrocoagulation have also demonstrated putative connections between discrete anatomical loci, such as the increased TRH immunoreactive concentrations in the nucleus of the solitary tract after bilateral electrolytic lesions of the TRH-containing paraventricular nucleus in rats. Although such work is beginning to elucidate neuropeptide wiring diagrams in mammalian brain, the association between discrete anatomical pathways containing a neuropeptide and the behaviors or effects observed after neuropeptide administration remains nascent. One of the best examples of this kind of association is seen in the response of CRF neuronal systems to stressful stimuli. The distribution of CRF neurons in the rat CNS is illustrated in Figure 1.6-1. Radioimmunoassay studies have documented increased CRF content in the locus ceruleus and decreased CRF concentrations in the median eminence after a regimen of acute or chronic stress in rats. Other studies have shown that CRF-containing nerve terminals impinge upon noradrenergic neurons of the locus ceruleus and that exogenous CRF applied to those neurons alters their firing rate. Some of the noradrenergic locus ceruleus neurons, in turn, project to the hypothalamic paraventricular nucleus where their input increases CRF synthesis and release. Because CRF injection into the locus ceruleus elicits fearful or anxious behavior, one could postulate that stress activates the CRF neurons terminating in the locus ceruleus noradrenergic neurons and that the increased CRF content in the locus represents an increased release of the CRF in this region onto the noradrenergic cell bodies. One can further postulate that the resulting increased noradrenergic signal, and perhaps other inputs to the paraventricular nucleus of the hypothalamus, mediates the stress-induced increased release of CRF from the median eminence, which is detected as decreased CRF concentrations. Thus, both an observed increase and decrease in regional CRF content can be hypothesized as resulting from an increased release of CRF with or without concomitant new synthesis of CRF to replace the released peptide. An alternative explanation for the apparent decrease in median eminence CRF concentrations after stress versus increased concentrations in the locus ceruleus is that they are both released, but the CRF released from the median eminence is removed by the pituitary portal system whereas that in the locus ceruleus remains in the tissue that is dissected. Similar studies utilizing mRNA measures are ongoing under the hypothesis that changes in mRNA production might more accurately reflect biosynthetic rates of neuropeptide production.

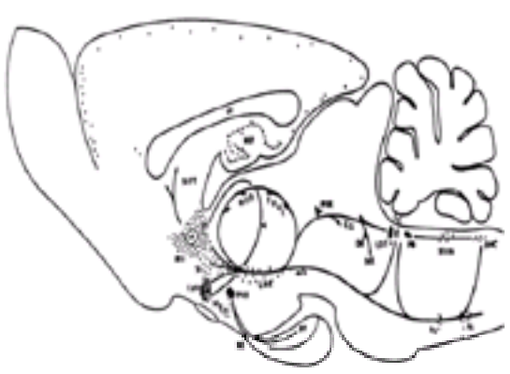

FIGURE 1.6-1 Major CRF-stained cell groups (dots) and fiber systems in the rat brain. CC, corpus callosum; HIP, hippocampus; SEPT, septal region; AC, anterior

commissure; BST, bed nucleus of the stria terminals; SI, substantia innominata; CcA, central nucleus of the amygdala; MPO, medial preoptic area; PVH, PVN of hypothalamus; ME, median eminence; PP, posterior pituitary; LHA, lateral hypothalamic area; mfb, median forebrain bundle; MID THAL, midline thalamic nuclei; ST, stria terminalis; POR, perioculomotor nucleus; CG, central gray; DR, dorsal raphe; MR, median raphe; LDT, laterodorsal tegmental nucleus; LC, locus ceruleus; PB, parabrachial nucleus; MVN, medial vestibular nucleus; DVC, dorsal vagal complex; A_5, A_1 noradrenergic cell groups. (Reprinted with permission from Swanson LW, Sawchenko PE, Rivier J, Vale WW: Organization of ovine corticotropin-releasing factor immunoreactive cells and fibers in the rat brain: An immunohistochemical study. Neuroendocrinology *36:*165, 1983.)

BIOSYNTHESIS

The biosynthesis of neuropeptides involves the transcription of mRNA sequences from deoxyribonucleic acid (DNA) templates contained on the appropriate genes. Since the 1980s the application of molecular biological techniques has allowed the genes of many of the various peptides to be cloned and the complementary DNA probes constructed that allow mapping of the regions where the mRNA's coding for the peptide prohormone is located. A good example of the exploitation of such techniques is provided by TRH. Although TRH was the first of the hypothalamic releasing factors to be chemically identified in 1969, the TRH precursor was the last of the releasing-factor prohormones to be described. The gene for TRH in humans resides on chromosome 3; in the rat it consists of three exons (coding regions) separated by two introns (noncoded sequences). The first exon contains the 5' untranslated end of the mRNA encoding the TRH preprohormone, the second exon contains the signal sequence and much of the remaining amino terminal end of the precursor peptide, and the third contains the remainder of the sequence, including five copies of the TRH precursor sequence, the carboxy terminal region, and the 3' untranslated region. Regions in the 5' flanking sequence correspond to promoter regions and have sequence homologies with a glucocorticoid receptor binding site and the thyrotropin b subunit gene that may regulate the expression of the TRH gene. Although some disagreement about the precise sizes of the TRH precursors exists, the TRH prohormone has been mapped immunohistochemically to regions previously shown to exhibit TRH-containing cell bodies, including the paraventricular nucleus of the hypothalamus and the raphe nuclei, whereas the axons and terminals that have been identified as containing TRH do not stain as intensely for the precursor.

The neurotensin-neuromedin N gene was originally cloned from canine ileal mucosa, and complementary deoxyribonucleic acid (cDNA) probes constructed against this form were used to clone the rat gene. The rat gene contains four exon sequences separated by three introns and spans approximately 10.2 kilobases. In the rat the neurotensin-neuromedin N sequence is contained in the fourth exon and the single copies of each peptide sequence are separated by a pair of dibasic residues. The human neurotensin gene has been localized to chromosome 12 (12q21). In pheochromocytoma (PC-12) neurons in culture, the neurotensin-neuromedin N gene is regulated by lithium, nerve growth factor, cyclic adenosine monophosphate (AMP) activators, and dexamethasone through their effects on a 5' cis-regulating region. The distribution of the neurotensin-neuromedin N mRNA is generally the same as described for neurotensin-containing neuronal cell bodies, except in the hippocampus and subiculum, where few neurons stain immunohistochemically for neurotensin and yet an abundance of the neurotensin-neuromedin N mRNA is found.

The role of 5' regulatory sequences on peptide genes has been well described for the SRIF gene. The human SRIF gene is located on chromosome 3. The regulatory region in the SRIF gene is upstream from the sequence coding the SRIF mRNA and contains the palindrome sequence of eight base pairs that is found in other genes regulated by cyclic AMP. The promoter region acts on the downstream sequence as an enhancer for transcription and exhibits both distance and orientation sensitivity for the sequence being enhanced. This cyclic AMP response element demonstrates recognition sites for protein kinases A and C and casein kinase II, which may in turn regulate that activity.

PEPTIDE PROCESSING

Because neuropeptides are first synthesized as larger precursor molecules, a wide variety of processes can come into play in the cleavage of the active peptide forms from the precursor (Fig. 1.6-2). For example, SRIF is first produced as a 116 amino acid prohormone called preprosomatostatin, and it contains a 24 amino acid signal sequence that is removed in the formation of the 92 amino acid prosomatostatin. It is further processed to either SRIF-28 or SRIF-14, and a major site of the processing step has been identified as the Golgi apparatus. The last processing step shows significant species differences, which should be considered when extrapolating between human and animal studies. For example, the 12 amino acid sequence of SRIF-28 that is cleaved in the formation of SRIF-14 is much more abundant in rodent than in human brain. The joint actions of a basophilic aminoprotease and an endoprotease contained in secretory granules cleaves SRIF-14 from SRIF-28. Most active peptide sequences are flanked by dibasic amino acids (Arg and Lys), which act as cleavage sites for the carboxypeptidase-B types of enzymes. However, SRIF-28 is cleaved at a single arginine from its prosomatostatin precursor. Related peptides are often contained in the same prohormone sequence, as is the case for neurotensin and neuromedin N. Those peptides are separated by a single pair of dibasic residues on their common mRNA and yet have distinctly different distribution patterns in the brain. Other tissues may also exhibit processing that is different from that of the brain, as is seen for neuromedin N in the mouse ileum. Multiple active peptide copies can also be contained in the prohormone structure as is noted with TRH, which has five complete copies in the mammalian 285 amino acid prohormone. Studies using antisera that recognize the intervening sequences between the five copies of TRH within the prohormone indicate that all five copies are liberated during processing. Regional differences of prohormone processing have been demonstrated within the brain, as is clearly seen for TRH. In the hypothalamus, the main storage forms of TRH are TRH, pre-pro TRH (160-169), pre-pro TRH (178-199), and two additional forms that are found in the olfactory bulb region. The differences in the ratio of TRH to its prohormone precursor in various extraneuronal tissues also indicates widely varying regional differences in processing of the TRH precursors. Processing also differs across the life cycle, as has been reported for TRH. Hypothalamic TRH prohormone processing in mice was observed to accelerate during development based on the ratio of TRH to its precursors, and immunohistochemical staining or in situ hybridization autoradiography indicated that a significant amount of processing occurred during post-Golgi transport and storage.

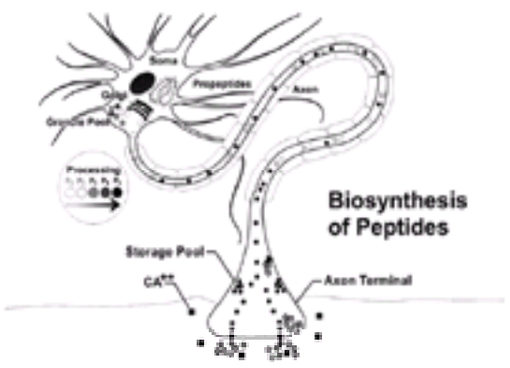

FIGURE 1.6-2 The peptide neuron. The figure shows the main steps in the chain of events from the information stored in the DNA molecule to the peripherally detected peptide fragments. The DNA sequence in the nucleus is transcribed to the mRNA molecule for further transport to the endoplasmic reticulum, where a translation takes place to form a large precursor protein (preproprotein). That protein is prepared for axonal transport by packaging into neurosecretory vesicles within the Golgi complex. During transport, the precursor protein is processed by specific cleavage enzymes into active and inactive peptide fragments. After release, the peptides are further degraded into smaller peptide fragments or constituent amino acids. (Courtesy of Thomas Davis, Ph.D.)

Although many known peptides are complete and biologically active when cleaved from the prohormone, many others are subjected to posttranslational processing. Certain peptides have a metabolically blocked carboxy terminus that is often amidated. A glycine residue in the prohormone sequence often acts as the amide donor, and in the case of TRH is attacked by a monoxygenase that is contained in secretory granules. TRH is further processed on the N-terminus where glutamine is cyclized by a glutamylcyclase. These alterations are usually effective in reducing susceptibility to degradation, and are often required for biological activity, as is the case for TRH, which is rendered inactive when the C-terminal amide is removed by proline endopeptidase to generate the free-acid structure. Other posttranslational processing events for active peptides include glycosylation and cyclization, which are often required for either biological activity or transport.

RECEPTORS AND SECOND MESSENGERS

Neuropeptide receptors have undergone the same process of discovery and characterization that receptors for other neurotransmitters have enjoyed. The process begins with the pharmacological characterization of the receptor's physicochemical binding properties by assessing the affinity of various metabolically derived and synthetic peptide fragments, and the native molecule, for the receptor binding site found in membrane preparations. Peptide receptor locations are mapped with radioactive or fluorescent tags that are inserted into peptide molecules, which often contain substituted amino acids at the most vulnerable peptidase cleavage sites. Previously, once the peptide receptor was characterized pharmacologically, it was usually purified from some relatively enriched biological tissue source or brain region by affinity column chromatography. After it had been purified, binding parameters and activity were recharacterized for the reconstituted purified receptor protein and structural information obtained by X-ray crystallography. This process was closely followed in the purification of the neurotensin-neuromedin N receptor. Because neurotensin and neuromedin N share significant sequence homology, the latter is active in displacing ligands from the neurotensin receptor, but with approximately 20 times less potency. The neurotensin receptor was first characterized by photoaffinity labeling and cross-linking of radioiodinated ligands, which resulted in two labeled subunits of about 49 Kd and 51 Kd from rat brain synaptosomes. The receptor was next solubilized and characterized for ligand affinity and binding capacity in mouse brain, which was followed by affinity column chromatographic purification and confirmation of an aggregate molecular weight of approximately 100,000. However, similar work with bovine cerebral cortex yielded a purified neurotensin receptor of approximately 72,000, indicating that significant species differences may exist. The neurotensin receptor mRNA has been cloned, and regional in situ hybridization mapping studies indicate that its distribution is generally the same as was shown for the receptor using radioactive ligands and autoradiography. The location is particularly rich in dopamine cell body regions and some dopamine terminal regions, and thus it is not surprising that dopamine neuronal activity seems to regulate neurotensin receptor expression. Neurotensin receptors have been shown to be colocalized with enzymes that degrade both neurotensin and neuromedin N in primary cultures of neurons from the forebrains of 14-day-old mouse embryos.

The much more powerful tools of molecular biology have been utilized more recently. Utilizing expression cloning techniques or a variety of low-stringency hybridization screening methods, numerous receptors and peptides have been cloned. Transfection studies allow for the production of highly enriched cell preparations expressing a variety of peptides or peptide receptors. Moreover, now that it is possible to purify mRNAs for peptide receptors, the induction of mutations allows for the identification of the regions controlling ligand binding. The distribution of receptors mapped with autoradiographic techniques has been largely verified by in situ hybridization using receptor mRNA probes. That information will make it possible to design drugs specifically to fit those binding sites on the receptor, leading to the ability to manipulate peptide systems in ways that are currently enjoyed by the more classic neurotransmitters. Finally, transgenic techniques such as targeted gene overexpression and gene knockouts or knockdowns will contribute to our understanding of the physiological roles of these peptides and their receptors.

Receptor populations for peptides exhibit changes in the numbers of binding sites on the basis of the magnitude of transmitter signal received and the input from second messenger feedback regulation. The up-regulation and down-regulation of peptide receptors has been most often demonstrated in the anterior pituitary, but has also been described in the cerebral cortex and other brain regions. Peptide receptor affinity for ligands usually remains stable in the face of this regulation of receptor number. Receptor expression fluctuates in various brain regions during development as well. Autoradiography has demonstrated high concentrations of SRIF receptors in rat somatosensory cortex at day 16 of the embryo in the intermediate zone and a transient decrease in cortical plate SRIF receptors at birth. Decreases to adult levels of SRIF receptors in somatosensory cortex are achieved by postnatal day 21. Somatostatin receptors in the cerebellum of 13-day-old rats have been shown to be pharmacologically similar to those of adults in binding parameters. Several research groups have described different classes of SRIF receptors on the basis of selective binding by various pharmacological ligands. Regional differences in the binding of SRIF-28 and SRIF-14 and their inability to desensitize each other's binding site has indicated that separate receptor populations may exist for these two forms. Further evidence for distinct populations of SRIF-28 and SRIF-14 receptors is provided by their production of different second messenger effects, and their opposite effects on potassium conductance in rat cortex. Five different SRIF receptor subtypes have been identified with molecular techniques.

Neuropeptide receptors have been associated with just about every type of second messenger signal transduction system that has been identified. Mechanisms using cyclic AMP; cyclic guanosine monophosphate (cGMP); protein kinases A and C; sodium, potassium, and calcium channels; and inositol phosphate and diacylglycerol have all been identified as neuropeptide receptor signal transduction mechanisms. Such mechanisms offer a myriad of possible modulatory effects, from the amplification to the attenuation of postsynaptic signals, and contribute greatly to the integrative power of neural networks. Both the neurotensin and the TRH receptors are internalized within the postsynaptic cell upon binding of their endogenous ligand or the appropriate agonist, where portions of the complex may eventually be transported to the receptive cell's nucleus with subsequent regulatory effects. Specific peptide receptor antagonists have been difficult to develop, with opioid antagonists being the most successful. Altered molecular forms of native peptides, such as a-helical CRF $_{9\text{-}41}$, have been used but are not ideal owing to size constraints on diffusion and lack of ability to penetrate the blood-brain barrier.

The inability to block specific neuropeptide signals pharmacologically has severely hindered research into the roles of the endogenous peptides in various behaviors and physiological effects. The disadvantages of trying to decipher a substance's role in neurotransmission by examining only the effects of excess concentrations should be obvious to even casual observers. As noted previously, this is an area in which the use of transgenic laboratory animals can be of great service, although the results are potentially confounded by differing roles for neuropeptides and receptors in early development and adulthood. More recently, several pharmaceutical companies have synthesized nonpeptidergic, lipophilic compounds that act as neuropeptide agonists or antagonists.

PEPTIDASES

Peptides are degraded to smaller fragments, and eventually to single amino acids, by specific enzymes termed *peptidases*. As yet, peptides or their fragments have not been shown to be actively taken up by presynaptic nerve terminals, as is the case for the monoamines. The enzymes may be found bound to post- or presynaptic neural membranes or in solution in the cytoplasm and extracellular fluid, and they are distributed widely in peripheral organs and serum as well as in the CNS. They often have a metal ion among their subunit components; those components form the active site for cleavage of the target peptide sequence, and that active site often forms a three-dimensional cleft where the specific peptide bond cleavage occurs. There are several general classes of peptidases, with several distinct enzymes in each class. Those classes include the serine endopeptidases containing such enzymes as trypsin and chymotrypsin; the thiol peptidases, such as pyroglutamate amino peptidase and cathepsin B and C; the acid proteases, such as pepsin and renin; the metalloendopeptidases, such as neural endopeptidase and angiotensin-converting enzymes; and the metalloexopeptidases, such as the aminopeptidases and the carboxypeptidases such as enkephalin-convertase and carboxypeptidase A and B. These degradative enzymes are often the same as those used in processing but have different subcellular locations. An example is carboxypeptidase B, which cleaves the dibasic amino acid residues flanking the active peptide sequence in the prohormone during processing, or reduces activity at the receptor if the peptide contains dibasic amino acids in the active sequence, such as neurotensin. Peptidases have pH and temperature optimums for activity and can be inhibited by various chemicals or chelators or by amino acid substitution at vulnerable points in the peptide chain. Alterations in peptidase activity or concentration can contribute to alterations in the synaptic availability of a peptide, and the regulation of peptidase levels may be as exquisitely controlled as receptor number and peptide synthesis and release. Table 1.6-2 shows the potential cleavage points for neuromedin N, neurotensin, SRIF, and TRH. Cleavage of the actively released form of the peptide usually ends or significantly reduces biological activity, but examples abound of partial or complete receptor activation by partially metabolized peptides or their fragments.

The metabolism of TRH has been investigated fairly completely, principally because of the limited number of fragments that can be generated from a tripeptide. The principal cleavage enzymes are pyroglutamyl amino peptidase, which cleaves the cyclized glutamyl residue from the C-terminus and generates a histidine-proline (His-Pro) fragment. That fragment spontaneously cyclizes into a diketopiperazine, the so-called cyclo His-Pro, after the N-terminal amide has been removed by the action of the proline endopeptidase. The active site of the pyroglutamyl amino peptidase enzyme has been shown to contain tyrosine; histidine, arginine, and possibly lysine residues, but does not contain serine, cysteine, aspartate, or glutamate. Regional differences in TRH degradation have been described, with spinal cord metabolism of TRH generating more deamidated TRH than cerebral cortex degradation. The half-life of TRH in serum is estimated at only 2 to 3 minutes, and CSF is now known to contain pyroglutamyl amino peptidase activity. Neonatal CSF has less of activity of this enzyme when compared to the CSF of adults, and differences in subcellular localization of the enzyme in the adult hypothalamus (soluble fraction) and cerebral cortex (membrane bound) has been reported as well, with the brain activity of both forms decreasing during development. Both of those TRH peptidases have been detected in the cytosol of brain homogenates, but are found only in trace amounts in the soluble fraction of synaptosomes; most of their activity is associated with synaptosomal membranes. Thyroid hormones have been shown to regulate pyroglutamyl amino peptidase in the membrane-bound fraction but not in the soluble form, although in serum the peptidase does not appear to be influenced by thyroid hormones. Thus peptidases offer yet another opportunity for the integration and regulation of neuropeptide transmitter actions and synaptic availability. Because the present peptidase inhibitors are relatively nonspecific in their ability to inhibit various peptidases, there have been few attempts to influence peptide concentrations by pharmacological blockade of their associated peptidases; however, the angiotensin-converting enzyme (ACE) inhibitors such as captopril are one exception. It is expected that second- and third-generation peptidase inhibitors, with discrete peptidase and possibly regional specificity, will be developed that eventually may allow the truly elegant manipulation of endogenous neuropeptide concentrations. In addition, the genes coding for the peptidases are being cloned, and knockout experiments will provide novel and valuable information.

NEUROENDOCRINE SECRETION

With the exception of neuromedin N, each of the example peptides are known to play major roles in pituitary-target endocrine organ regulation, including CRF-induced release of proopiomelanocortin products, such as adrenocorticotropic hormone (ACTH) and b-endorphin; TRH release of thyrotropin (thyroid-stimulating hormone [TSH]) and prolactin; and SRIF-induced inhibition of the release of growth hormone, thyrotropin, gonadotropins, and ACTH. Neurotensin, which is abundant in the hypothalamus and median eminence, may mediate the preovulatory release of luteinizing hormone and receive feedback for the induction of mRNA synthesis by estrogen, but it is not a hypothalamic hypophysiotropic hormone in the classic sense. A sexually dimorphic distribution of the neurotensin-neuromedin N mRNA in the preoptic hypothalamus also supports such a role for neurotensin in rodents.

The peptides involved in neuroendocrine regulation have cell bodies residing in the hypothalamus that receive feedback from all levels of the endocrine axes. The complexity of those interactions has been well demonstrated for the hypothalamic-pituitary-thyroid axis and the hypothalamic-pituitary-adrenal axis and have now been extended to the molecular level. The regulatory feedback of thyroid hormones onto the TRH-synthesizing neurons of the paraventricular nucleus was first demonstrated with evidence of TRH concentration changes, reported to be reduced in the median eminence after thyroidectomy, but not in the rest of the hypothalamus, and which could be prevented by thyroid hormone replacement. The treatment of normal rats with exogenous thyroid hormone decreases TRH concentration in the paraventricular nucleus and the posterior nucleus of the hypothalamus. That effect was corroborated for the TRH prohormone as well, with median eminence levels of TRH prohormone being reduced by thyroidectomy and the precursor levels increasing toward normal concentrations after thyroxine treatment. The TRH mRNA also exhibits such regulation by thyroid hormone as expected, with increased mRNA concentration in the paraventricular nucleus 14 days after thyroidectomy. Unilateral tri-iodothyronine implants prevent the increase in TRH mRNA that is seen on the contralateral untreated side in propylthioruracil-induced hypothyroidism. The effects of thyroid hormones on TRH expression in the paraventricular nucleus of developing rats are not observed until between embryo day 20 and 7 days after birth, although TRH mRNA is evident as early as embryo day 16. The ability of thyroid hormones to regulate TRH mRNA can be superseded by other stimuli that activate the hypothalamic-pituitary-thyroid axis. In that regard repeated exposure to cold (which releases TRH from the median eminence) induces increases in the levels of TRH mRNA in the paraventricular nucleus despite concomitantly elevated concentrations of thyroid hormones. Further evidence of the different levels of communication of the hypothalamic pituitary-thyroid axis are seen in the ability of TRH to regulate the production of mRNA for the pituitary TRH receptor and for TRH concentrations to regulate the mRNA coding for both the a and b subunits of the TSH molecule. The latter effect has been shown to be dependent on intracellular calcium and protein kinase C. The regulatory interplay also extends to the accessible pools of second messenger phosphoinositides, whose pool size is regulated by TRH receptor number. TRH-containing synaptic boutons have been observed in contact with TRH-containing cell bodies in the medial and periventricular subdivisions of the paraventricular nucleus, thus providing anatomical evidence for ultrashort feedback regulation of TRH concentrations there.

Regional differences in CRF receptor regulation by corticosterone have also been reported, which have been shown to partly result from differential glycosylation of the CRF receptor. The regulation of neuropeptide mRNA concentrations may be influenced by other neuropeptides, as well as by components of the particular endocrine axis normally associated with the particular peptide, as demonstrated by the ability of neuropeptide Y to increase hypothalamic CRF mRNA.

Because many endocrine systems are cyclic in their regulatory functions, it is not surprising that neuropeptides often exhibit rhythms in concentrations that are based on diurnal, lunar, and circannual periodicities. Hypothalamic and certain extrahypothalamic regional concentrations of CRF exhibit increased concentrations in the afternoon as compared to morning concentrations, and this increase can be attenuated by corticosterone only in certain brain regions such as the hypothalamus. Somatostatin, CRF, and TRH concentrations in the CSF of nonhuman primates exhibit daily fluctuations, and the monthly cycles in gonadotropins of mammals exhibiting estrus are well recognized. Circannual rhythms of neurotensin and SRIF concentrations that are 180 degrees out of phase in rodent hypothalamus have been noted. Changes in mRNA expression during development have been seen in CRF mRNA, which is present at gestational day 17, but decreases from day 19 to day 21, when concentrations again rise to attain adult levels by 4 days after birth. Other peptide mRNAs, such as SRIF, do not exhibit such fluctuations during development, but do show differential distribution during ontogeny. Daily fluctuations in rat paraventricular nucleus CRF mRNA expression are lowest during the period of highest plasma corticosterone levels during the 24-hour cycle.

NEUROPEPTIDES IN PSYCHIATRIC DISORDERS

Humans are less than ideal subjects for neuropeptide research for several reasons. The peripheral sources of many peptides, the relatively high concentration of serum peptidases, and the blood-brain barrier all conspire to render serum concentrations of CNS neuropeptides difficult to interpret at best. The use of biopsy to assess tissue concentrations directly is not routinely repeatable, is limited to superficial structures, suffers from potential morbidity, and would provide only limited information. However, CSF has been shown to reflect extracellular fluid concentrations of transmitter substances, is in direct contact with the CNS, is screened from peripheral serum sources by the blood-brain barrier, and may be sampled across time. The limitations of CSF studies include a lack of information about the regional CNS source of any concentration changes detected, the use of lumbar CSF, which is somewhat removed from higher CNS sources of peptides and subject to spinal cord peptide contributions, and the potentially confounding effects of previous drug treatments or disease episodes. Postmortem tissue studies of neuropeptide concentration changes in psychiatric disease are affected by agonal state, postmortem delay, previous drug treatment, and coexisting illnesses. Most of the data on CSF concentration changes or tissue concentration changes of neurotransmitters have been derived from comparisons between diagnostically defined psychiatric groups and control groups. However, the controls may be so-called neurologically or psychiatric controls, not healthy volunteers, and the accuracy and consistency of the diagnoses may be less than optimal. In addition, the etiology of a diagnostic class of disease may differ among subjects in the same diagnostic group. Even after matching for age, gender, or other demographic variables, heterogeneity among human research populations results in individual variations of absolute peptide values that are often quite wide. Such variations severely reduce the power of group comparisons to detect alterations in peptide concentrations. The use of pretreatment and posttreatment CSF samples, or of samples taken during the active disease state versus when the patient is in remission, has begun to address the serious limitations in study design. For such progressive diseases as schizophrenia or Alzheimer's disease, serial CSF samples may be a valuable indicator of disease progression or response to treatment. Even with these constraints, significant progress has been made in describing the effects of various psychiatric disease states on neuropeptide systems in the CNS.

Alzheimer's Disease Dementia of the Alzheimer's Type represents up to two thirds of the demented population encountered in clinical practice, and over half of the nursing home beds in the United States are currently occupied by such patients. The disease is characterized by a progressive, gradually worsening dementia that cannot be ascribed to metabolic disorders, pharmacological treatment, or infectious agents and is neuropathologically associated with the pathological presence of senile plaques and neurofibrillary tangles within the CNS. The first described specific neurochemical deficit to be associated with Alzheimer's disease was reduced amounts of choline acetyltransferase-containing nerve terminals in cortical regions as a result of degeneration of cholinergic neuronal perikarya in the nucleus basalis of Meynert in the substantia innominata region of the basal forebrain. Within a few years of that finding, SRIF was found to be markedly reduced in concentration in the cerebral cortex of Alzheimer's disease patients. Subcortical regions containing SRIF, such as the substantia innominata, hypothalamus, and bed nucleus of the stria terminalis, were spared whereas SRIF receptors in the cortex were decreased in number. In regions such as the hippocampus, findings of SRIF depletions were less consistent than in the cortex, but when depleted in the hippocampus, the SRIF neurons colocalized with neuropeptide Y were spared. Somatostatin concentration in the CSF of Alzheimer's disease patients has also been consistently found to be decreased, and this decrease has been correlated with the magnitude of cognitive impairment. Therapies that slow or partially reverse the dementia associated with Alzheimer's disease have been reported to also partially reverse the decrease in CSF SRIF. However, CSF SRIF concentrations are also decreased in delirium, major depressive disorder, schizophrenia, multiple sclerosis, and dementia associated with Parkinson's disease. Increased activity of an SRIF cleaving peptidase have been described in certain cortical regions of the brain tissue of persons with Alzheimer's disease, raising the possibility that increased SRIF degradation may contribute to the decreases in SRIF concentration observed. Treatment of Alzheimer's disease patients with SRIF infusion systemically, however, has not been successful in reversing the dementia, probably because of poor penetration across the blood-brain barrier. In experimental animals cysteamine depletion of hippocampal SRIF leads to deficits in performance on tasks requiring retention of information. It is unclear whether the neuropeptide deficits precede, succeed, or occur in tandem with the cholinergic deficits seen in Alzheimer's disease, and whether the neurochemical systems and regions first exhibiting deficits are the site of the onset of pathology.

The CRF-containing interneurons of the cortex are also consistently depleted in Alzheimer's disease (Fig. 1.6-3). As with SRIF, subcortical areas containing CRF neurons may be spared, but unlike SRIF, CRF receptors are increased in number (up-regulated) with no change in affinity. Various research groups have reported the CRF concentrations in the CSF of patients with Alzheimer's disease to be increased, decreased, or unchanged; these inconsistencies are likely due to when the CSF sample was obtained. Other peptides have been shown to be altered less consistently in Alzheimer's disease, such as substance P or neurotensin, whereas most peptides are reported to be unchanged, including TRH, vasoactive intestinal peptide, CCK, and the enkephalins. Only one peptide, galanin, is reported to be reliably increased in concentration in Alzheimer's disease. Novel agonists at peptide receptors such as CRF or SRIF may allow for the development of new treatments; currently, compounds that increase the availability of CRF by competing for the binding site on the CRF-binding protein are being planned for clinical trial in Alzheimer's disease patients.

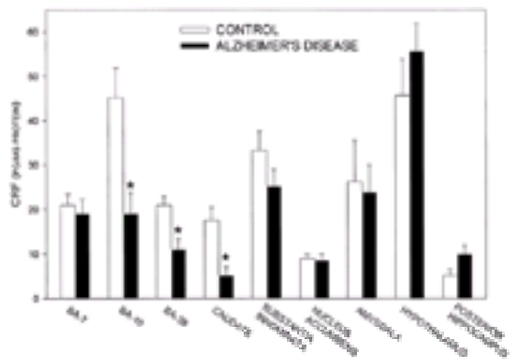

FIGURE 1.6-3 Regional brain concentrations of CRF in patients with dementia of the Alzheimer's type and in controls. *N* = 7-13 per group. Data presented as mean ± SE 7 * P < 0.01 by Student's *t*-test. (Data derived from Bissette G, Reynolds GP, Kilts CD, Widerlov E, Nemeroff CB: Corticotropin-releasing factor like immunoreactivity in senile dementia of the Alzheimer type. JAMA *254:*3067, 1985.)

Mood Disorders

Corticotropin-Releasing Factor After a search spanning nearly three decades, CRF was isolated and characterized in 1981 as a 41-amino acid peptide. CRF is the primary hypothalamic ACTH-secretagogue in most species; it also functions as an extrahypothalamic neurotransmitter in a CNS network that apparently coordinates global responses to stressors. CRF and its homologs represent an ancient family of peptides subserving numerous functions. In higher organisms, including mammals, convincing evidence supports the hypothesis that CRF plays a complex role in integrating the endocrine, autonomic, immunological, and behavioral responses of an organism to stress.

Hyperactivity of the hypothalamic-pituitary-adrenal axis in major depressive disorder remains one of the most consistent findings in biological psychiatry. The reported hypothalamic-pituitary-adrenal axis alterations in major depression include hypercortisolemia, resistance to dexamethasone suppression of cortisol secretion, blunted ACTH responses to intravenous CRF challenge, and elevated concentrations of CRF in CSF. The exact pathological mechanisms underlying hypothalamic-pituitary-adrenal axis dysregulation in major depressive disorder and other mood disorders remains to be elucidated. Defects have been postulated to exist at corticolimbic loci, hypothalamic loci, or both.

Once the phenomenon of hypothalamic-pituitary-adrenal axis hyperactivity in patients with major depression was established, many clinical research groups utilized various provocative neuroendocrine challenge tests as a “window into the brain” in attempts to elucidate pathophysiological mechanisms. In normal subjects the CRF stimulation test, using either rat or human CRF or ovine CRF, yields robust ACTH, b-endorphin, b-lipotropin, and cortisol responses following intravenous or subcutaneous administration. However, in patients with major depressive disorder, blunting of ACTH or b-endorphin secretion with a normal cortisol response has been repeatedly reported. Patients with posttraumatic stress disorder, 50 percent of whom also fulfill the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria for major depressive disorder also show blunted ACTH secretion in response to a CRF challenge. Importantly, researchers have reported normalization of the ACTH response to CRF following clinical recovery from depression, suggesting that the blunted ACTH response, like dexamethasone nonsuppression, may be a state marker for depressive disorders.

Mechanistically, two hypotheses have been advanced to account for the ACTH blunting following exogenous CRF administration. The first hypothesis suggests that pituitary CRF receptor down-regulation occurs as a result of hypothalamic CRF hypersecretion. The second hypothesis postulates altered sensitivity of the pituitary to glucocorticoid negative feedback. Substantial support has accumulated favoring the first hypothesis. It should be kept in mind that neuroendocrine studies represent a secondary measure of CNS activity; the pituitary ACTH responses principally reflect the activity of hypothalamic CRF rather than that of the corticolimbic CRF circuits, which are most likely to be involved in the pathophysiology of depression.

A potentially more direct method for evaluation of extrahypothalamic CRF tone may be obtained from measurements of CRF concentrations in CSF. A marked dissociation between CSF and plasma neuropeptide concentrations has been described, thus indicating that neuropeptides are secreted directly into CSF from brain tissue as opposed to being derived from plasma-to-CSF transfer. Evidence that CRF concentrations in CSF originate from nonhypophysiotropic CRF has been obtained from studies in which CRF concentrations in CSF were repeatedly measured over the course of the day. Two independent research groups reported that CRF concentrations in the CSF of in rhesus monkeys are not entrained with pituitary-adrenal activity. The proximity of corticolimbic, brainstem, and spinal CRF neurons to the ventricular system suggests that these areas make substantial contributions to the CRF pool in CSF.

A series of studies have demonstrated significant elevations of CRF concentrations in the CSF of drug-free patients with major depression or following suicide (Fig. 1.6-4). Additionally, severity of depression appears to correlate significantly with CRF concentrations in the CSF of patients with anorexia nervosa, multiple sclerosis, and Huntington's disease. The elevation of CRF concentrations in the CSF of patients with anorexia nervosa reverts to the normal range as these patients approach normal body weight. No alterations of the concentrations of CRF in CSF have been reported in other psychiatric disorders including mania, panic disorder, and somatization disorders as compared to controls.

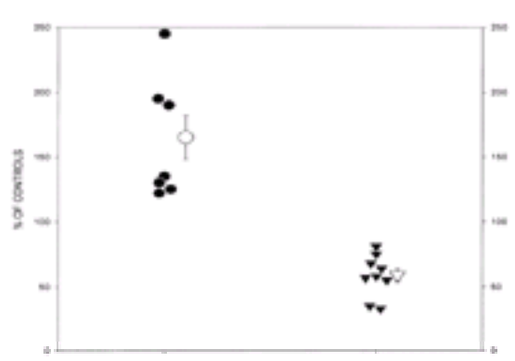

FIGURE 1.6-4 Scatterplot showing the mean cerebrospinal fluid neuropeptide values in patients with major depression. Each individual point represents an individual study. Mean % Control ± SEM for all studies shown by hollow symbols. (Adapted from Owens MJ, Plotsky PM, Nemeroff CB: Peptides and affective disorders. In *Biology of Schizophrenia and Affective Disorders,* SJ Watson, editor. American Psychiatric Press, Washington, DC, 1996.)

Of particular interest is our demonstration that the elevated concentrations of CRF in CSF in drug-free depressed patients are significantly decreased 24 hours after their final ECT treatment, indicating that CSF CRF concentrations, like hypercortisolemia, represent a state rather than a trait marker; other recent studies have confirmed the normalization of CRF concentrations in CSF following successful treatment with fluoxetine. One group demonstrated significant reduction of elevated CRF concentrations in the CSF of 15 female patients with major depression who remained depression free for at least 6 months following antidepressant treatment as compared to little significant treatment effect on the CRF concentrations in the CSF of 9 patients who relapsed in this 6-month period. This suggests that elevated or increasing CRF concentrations in CSF during antidepressant drug treatment may be the harbinger of a poor response in major depressive disorder despite early symptomatic improvement. Interestingly, treatment of normal subjects with desipramine fluoxetine-containing antidepressants is associated with a reduction in the concentration of CRF in CSF.

In preclinical studies, CRF hypersecretion is associated with CRE-receptor down-regulation. Depression is a major determinant of suicide more than 50 percent of completed suicides are accomplished by patients with major depressive disorder. If CRF hypersecretion is a characteristic of depression, evidence of related CRF-receptor down-regulation should be evident in the CNS of depressed suicide victims. Indeed, it has been reported that there is a marked (23 percent) decrease

in the density of CRF receptors in the frontal cortex of suicide victims as compared to matched control samples; these findings have been confirmed in a second study.

Somatostatin Like a number of other neuropeptides, somatostatin was serendipitously discovered during attempts to purify growth hormone-releasing factor (GRF). As the name implies, SRIF inhibits the release of growth hormone from the anterior pituitary. Since its structural identification 20 years ago, SRIF has been unequivocally shown to be the major inhibitory influence on growth hormone secretion. Also, SRIF fulfills a number of criteria for neurotransmitter status within the CNS. The acceptance of a role for SRIF as a neurotransmitter has led to its investigation in a number of psychiatric and neurological diseases. Nonhyphysiotropic SRIF-containing neurons may play a role in a number of disorders including but not limited to depressive disorders, dementia, and epilepsy.

Like many other neuropeptide transmitters, central administration of SRIF produces a variety of behavioral and physiological effects. Briefly, the peptide produces a non–opioid-mediated analgesia in animals and man. Sleep patterns, food consumption, locomotor activity, and memory processes are also altered by central SRIF administration. Of particular interest is the fact that the sleep, eating, psychomotor activity, and anterior pituitary hormone secretion are all altered in depressed patients. Investigation of a role for this peptide in mood disorders was therefore of interest. The clearest evidence for involvement of SRIF in psychiatric illness has come from studies of major depression. A consistent decrease has been reported in SRIF concentrations in the CSF of drug-free depressed patients (Fig. 1.6-4). Research has revealed that a number of neuropeptides in CSF are almost exclusively of central origin, though the actual sites of production remain obscure. Decreases in SRIF concentrations in CSF are proposed to be the result of decreased neuronal synthesis and release. Whether this is a primary or secondary effect of the illness is unknown. Moreover, although SRIF levels in CSF did not correlate with severity of depression, clinically improved patients exhibited a rise in SRIF towards normal concentrations. In nine patients with bipolar I disorder followed longitudinally, SRIF concentrations in CSF were significantly decreased during the major depressive episode state versus mood-improved manic episodes.

Although the data is still relatively limited, overview of the extant literature suggests that decreases in concentrations of SRIF in CSF are a consistent state-dependent finding in depression. Indeed, secondary to the hypercortisolemia associated with depression, this is one of the more consistent findings in biological psychiatry. However, the finding has little apparent diagnostic usefulness because similar changes are observed in a number of neurological disorders without psychiatric comorbidity. However, reductions in SRIF concentrations in CSF do appear to be associated with impairment in cognitive function. Some investigators have suggested that the decrease in SRIF concentrations in CSF may be related to the hypothalamic-pituitary-thyroid axis overactivity commonly found in patients with depression. Whether one is responsible for the other, or whether both are responses to dysregulation of other neurotransmitter systems associated with depression is unknown. Rational design of peptide or nonpeptide-based drugs selectively active at different SRIF receptor subtypes will certainly aid in understanding its role in behavior, and may ultimately lead to novel therapeutic agents.

Thyrotropin-Releasing Hormone The early availability of adequate tools such as assays and synthetic peptides to assess hypothalamic-pituitary-thyroid axis function, coupled with observations that primary hypothyroidism is associated with depressive symptomatology, ensured extensive investigation of the involvement of this axis in mood disorders. Indeed, TRH, a pyroglutamylhistidylprolinamide tripeptide, was the first of the hypothalamic-releasing hormones to be isolated and characterized.

Early studies established the hypothalamic and extrahypothalamic distribution of TRH. The extrahypothalamic presence of TRH quickly led to speculation that TRH might function as a neurotransmitter or neuromodulator; indeed, a large body of evidence supports such a role for TRH.

Interest in putative CNS actions of TRH were stimulated by studies of the hypothalamic-pituitary-thyroid axis and depression by Prange and colleagues. In the 1970s it was hypothesized that thyroid function was integral to the pathogenesis of, and recovery from, mood disorders because of the copious interactions among thyroid hormones, catecholamines, and adrenergic receptors in the CNS. Overall, these studies suggest a role for thyroid dysfunction in refractory depression and are consonant with clinical studies that suggest the existence of an increased rate of hypothyroidism among patients with refractory depression.

The use of TRH as a provocative agent for assessment of hypothalamic-pituitary-thyroid axis function evolved rapidly after its isolation and synthesis. Clinical use of a standardized TRH stimulation test revealed blunting of the TSH response in approximately 25 percent of euthyroid patients with major depression; these data have been widely confirmed.

The observed TSH blunting in depressed patients does not appear to be the result of either excessive negative feedback resulting from hyperthyroidism or to SRIF hypersecretion. In fact, depressed patients exhibit reduced CSF concentrations of SRIF. It is possible that TSH blunting is a reflection of pituitary TRH receptor down-regulation as a result of median eminence hypersecretion of endogenous TRH. Indeed, the observation that concentrations of TRH in CSF are elevated in depressed patients as compared to controls supports the hypothesis of TRH hypersecretion but does not elucidate the regional CNS origin of this tripeptide. These elevations may be relatively specific to depression because no such alteration has been reported in patients with Alzheimer's disease, anxiety disorders, or alcoholism. Some investigators have suggested that the development of autoimmune thyroiditis gives rise to hypersecretion of hypothalamic TRH as a compensatory mechanism to maintain normal plasma triiodothyronine (T_3) and thyroxine (T_4) concentrations. Clearly, further studies in which CSF TRH concentrations in CSF are measured are needed.

Schizophrenia Both clinical and postmortem investigations of schizophrenia patients, as well as animal studies, have sought to elucidate the role of neuropeptides in the pathological manifestations of schizophrenia. Although constrained by diagnostic uncertainties and drug treatment effects, the research to date on the postmortem brain tissue of schizophrenia patients has not revealed major alterations of neuropeptide systems. A number of peptides (endogenous opioids, substance P, cholecystokinin [CCK], SRIF) have been reported to be altered in the CSF of schizophrenia patients, but many of these findings either have not been independently reproduced or describe a marginally statistically significant difference in CSF peptide concentrations among groups of patients and controls with variance of over 100 percent around the mean. Further confounds are the effects of treatment with antipsychotic drugs on peptide systems; such drugs have been described for regional neurotensin, CCK, substance P, and SRIF concentrations in laboratory animals. How much time is necessary to abolish those drug-induced alterations of neuropeptide system concentration changes is not known for humans, but may significantly exceed the 2 to 3 weeks of drug holiday used in most clinical studies.

The most likely candidate neuropeptide with evidence of selective alteration in schizophrenia is neurotensin. It was first shown to have pharmacological interactions with dopamine while undergoing characterization of its potent hypothermic and sedative potentiating activity. Subsequent work indicated that neurotensin possessed many properties that were also shared by antipsychotic drugs, including the ability to inhibit avoidance responding, but not escape, in a conditioned active avoidance task; the ability to block the effects of indirect dopamine agonists or endogenous dopamine in the production of locomotor behavior; and the ability to elicit increases in dopamine release and turnover. Unlike antipsychotic drugs, neurotensin is not able to displace dopamine from its receptor; neurotensin is colocalized in certain subsets of dopamine neurons and is coreleased with dopamine in the mesolimbic and medial prefrontal cortex dopamine terminal regions that are implicated as the site of dopamine dysregulation in schizophrenia. Antipsychotic drugs that act at dopamine type 2 (D_2) and D_4 receptors increase the synthesis and concentration of neurotensin in those dopamine terminal regions but not in others. That effect of antipsychotic drugs in increasing neurotensin concentrations persists after months of treatment and is accompanied by the expected increase in neurotensin mRNA concentrations, as well as expression of the “immediate early gene” c- *fos* and the transcription factor Fos within hours of initial drug treatment. The altered regulation of neurotensin expression by antipsychotic drugs apparently extends to the peptidases that degrade the peptide; recent reports have revealed decrease neurotensin metabolism in rat brain slices 24 hours after the administration of haloperidol.

Decreased neurotensin concentrations in CSF have been reported in several populations of patients with schizophrenia when compared to controls or patients with other psychiatric disorders. Although treatment with antipsychotic drugs has been observed to increase neurotensin concentrations in the CSF, it is not known whether this increase is causal or merely accompanies the decrease in psychotic symptoms seen with successful treatment. Postmortem studies have shown an increase in neurotensin concentrations in the dopamine-rich Brodmann's area 32 of the frontal cortex, but that result may have been confounded by premortem antipsychotic treatment. Other researchers have found no postmortem alterations in neurotensin concentrations of a wide sampling of subcortical regions. A comparison of the genomic sequence of the neurotensin neuromedin N gene in schizophrenia patients compared with age- and sex-matched controls found no differences in the gene sequence in the coding region. A critical test of the hypothesis that neurotensin may act as an endogenous antipsychotic-like substance awaits the development of a neurotensin receptor agonist that can penetrate the blood-brain barrier.

SUGGESTED CROSS-REFERENCES

Section 1.10 discusses basic molecular neurobiology, Section 1.11 discusses psychoneurocadocrinology, and the psychiatric aspects of endocrine disorders are discussed in Section 25.6.

SECTION REFERENCES

Betancur C, Azzi M, Rostène W: Nonpeptide antagonists of neuropeptide receptors: Tools for research and therapy. Trends Pharmacol Sci *18:*372, 1992.

Bissette G, Reynolds GP, Kilts CD, Widerlov E, Nemeroff CB: Corticotropin-releasing factor-like immunoreactivity in senile dementia of the Alzheimer's type. J Am Med Assoc *254:*3067, 1985.

*Bissette G, Nemeroff CB: The neurobiology of neurotensin. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Bissette G: Neuropeptides and Alzheimer's disease pathology. Ann NY Acad Sci *814:*17, 1997.

*De Souza EB, Grigoriadis DE: Corticotropin-releasing factor: Physiology, pharmacology, and role in central nervous system and immune disorders. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

*Engstroem G, Westrin A, Ekman R, Traeskman-Bendz L: Relationships between CSF neuropeptides and temperament traits in suicide attempters. Pers Individ Dif *26:*13, 1999.

Heit S, Owens MJ, Nemeroff CB: Corticotropin-releasing factor, stress, and depression. Neuroscientist *3:*186, 1997.

*Hökfelt TGM, Castel M-N, Morino P, Zhang X, Dagerlind A: General overview of neuropeptides. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Le F, Cusack B, Richelson E: The neurotensin receptor: Is there more than one subtype? Trends Pharmacol Sci *17:*1, 1996.

*Mason GA, Garbutt JC, Prange AJ Jr: Thyrotropin-releasing hormone: Focus on basic neurobiology. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Nemeroff CB, editor: *Neuropeptides in Psychiatric Disorders*. American Psychiatric Press, Washington, DC, 1991.

Nemeroff CB, editor: *Neuroendocrinology*. CRC Press, Boca Raton, 1992.

Owens MJ, Nemeroff CB: The physiology and pharmacology of corticotropin-releasing factor. Pharmacol Rev *43:*425, 1992.

Owens MJ, Plotsky PM, Nemeroff CB: Peptides and affective disorders. In *Biology of Schizophrenia and Affective Disorders,* SJ Watson, editor. American Psychiatric Press, Washington, DC, 1996.

Patel YC: Molecular pharmacology of somatostatin receptor subtypes. J Endocrinol Invest *20:*348, 1997.

Perone MJ, Windeatt S, Castro MG: Intracellular trafficking of prohormones and proneuropeptides: Cell type-specific sorting and targeting. Exper Physiol *82:*609, 1997.

Reichlin S: Neuroendocrinology. In *Williams Textbook of Endocrinology,* ed 8, JD Wilson, DW Foster, editors. WB Saunders, Philadelphia, 1992.

*Rubinow DR, Davis CD, Post RM: Somatostatin in the central nervous system. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

1.7 NEUROTROPHIC FACTORS

ERIC STEPHEN LEVINE, PH.D. AND IRA B. BLACK, M.D.

Trophic factors are survival molecules that are essential for the development of the nervous system. These factors regulate a wide range of developmental events that transcend their survival-promoting effects, including modulation of neurite outgrowth, phenotypic differentiation, and synaptogenesis. In addition to their roles during development, trophic factors regulate neuronal function and response to injury throughout life. The neurotrophin gene family is a widely distributed group of factors that subserves multiple functions in the nervous system. Neurotrophic factors have been implicated in pathophysiological mechanisms underlying neuropsychiatric disease, including schizophrenia and depression. These factors may also be important in neurodegenerative disorders such as Alzheimer's disease, Huntington's disease, and Parkinson's disease; each involves the selective loss of a neurotrophin-sensitive neuronal population.

NEUROTROPHINS AND THEIR RECEPTORS

The prototypical trophic factor is nerve growth factor (NGF), discovered by Rita Levi-Montalcini and Viktor Hamburger in the 1950s. NGF was originally isolated based on its ability to dramatically promote neurite outgrowth from sympathetic and sensory ganglia. It was later discovered to be essential for the survival and development of peripheral sympathetic and certain sensory neuronal populations. In the absence of NGF or when its action is blocked, there is a virtually complete loss of these responsive neuronal populations. Importantly, specific neuronal systems in the brain also respond to NGF and related trophic factors, greatly expanding their functional roles.

NGF is one member of a gene family of closely related trophic factors known as *neurotrophins*. This family includes brain-derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3) and NT-4/5, each of which shares high-sequence homology with NGF. These factors are expressed in neurons and glial cells throughout the nervous system. BDNF, NT-3, and NT-4/5 are found in virtually all areas of the brain, with highest levels of expression in the cerebral cortex and hippocampus. The distribution of NGF is much more restricted, with expression in cortex and hippocampus, but low levels in most other areas. Neurotrophin expression reaches its peak during development, but persists throughout life. Neurotrophin expression is also increased following specific neuronal insults, suggesting roles in responses to injury.

The biological effects of neurotrophins are mediated via binding to specific transmembrane receptors that trigger changes in intracellular second messengers. Neurotrophins bind to two distinct classes of receptor. Members of the trk tyrosine kinase receptor family appear to mediate most of the biological actions of these factors. The trk family consists of three related receptors, trkA, trkB, and trkC. TrkA is the primary receptor for NGF, trkB for BDNF and NT-4/5 and trkC for NT-3, although NT-3 also interacts at lower affinity with trkA and trkB as well (Fig. 1.7-1). These receptors have intrinsic tyrosine kinase activity and become autophosphorylated upon binding of the appropriate ligand. Phosphorylated tyrosine residues, in turn, provide binding sites for several downstream second messenger signals that attach to recognition sites (e.g., *src* homology 2 [SH-2] domains) on the intracellular domain of the trk receptor. The diverse second messengers include various protein kinases and transcription factors. These multiple, parallel signaling pathways evoke distinct biological effects on neurons, such as neurite outgrowth or survival itself.

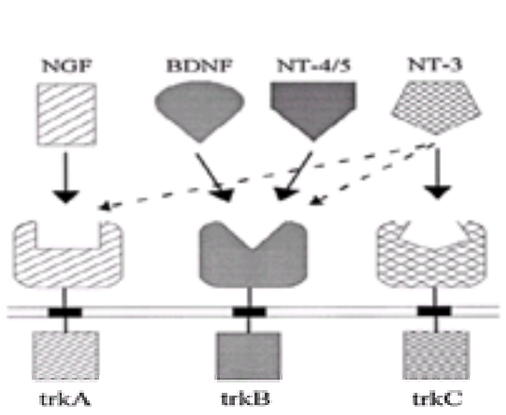

FIGURE 1.7-1 Schematic representation of the ligand-receptor relationship between neurotrophins and the trk family of tyrosine kinase receptors. Primary ligand-receptor interactions are indicated by thick arrows; secondary interactions are indicated by thin, dashed arrows. Only full-length receptor isoforms are depicted in this diagram. (Adapted from Barbacid M: J Neurobiol *25:*1386, 1994.)

The distribution of trk receptors is widespread in the nervous system; trkB and trkC, in particular, are expressed in virtually all regions of the brain, with highest levels in cortex and hippocampus. They are also found in thalamus, hypothalamus, cerebellum, basal ganglia, and brainstem, implying functional roles within neuronal populations subserving diverse functions, such as sensory perception, motor activity, endocrine regulation, cognition, motivation, and emotion. Distribution of the trkA receptor is more restricted, with high levels of expression only in the striatum and basal forebrain, suggesting that NGF acts on a limited population of central neurons.

Several different isoforms of the trk receptors are expressed in the nervous system. These splice variants are encoded by alternate messenger ribonucleic acid (mRNA) derived from the *trk* genes. In addition to the full-length, signaling receptors, *trk*B and *trk*C receptors are expressed in truncated forms. These truncated receptors contain the extracellular, ligand-binding domain, but lack the intracellular tyrosine kinase signaling domain. Functional roles of these receptors are unclear, but they may act as dominant-negative inhibitors of trk signaling or they may play a role in ligand clearance or presentation. The truncated receptors are the major variant expressed by glial cells. Expression of truncated forms is also higher during maturity than during development. These alternate receptor forms may provide added flexibility in neurotrophin signaling.

In addition to high-affinity binding to respective trk receptors, all the neurotrophins bind with relatively equal (and lower) affinity to the common neurotrophin receptor, known as *p75*. The functional role of this receptor is unclear. It is a member of the tumor necrosis factor family and may play roles in cell death. Signaling mechanisms mediated by p75 have not been readily identified but are thought to include activation of the sphingomyelin pathway. In some cellular populations, (e.g., oligodendrocytes), activation of p75 can trigger death, in contrast to the survival-promoting effects associated with neurotrophin signaling via trk receptors. Thus, differential expression of these two receptor systems may provide a mechanism for balancing opposing neurotrophin effects.

REGULATION OF GENE EXPRESSION

To appreciate the physiological significance of neurotrophin action in the brain, it is essential to understand the regulation of availability. In peripheral target organs, neurotrophins are synthesized and released constitutively by nonneuronal cells, independent of the influence of innervating neurons. In the brain, however, neuronal expression of neurotrophin genes is modulated by impulse activity itself. For example, trophin expression in vivo in the hippocampus and neocortex is increased by limbic seizures. Dramatic increases in NGF and BDNF mRNA concentrations are produced by different methods of seizure induction, including dentate gyrus lesions, kainic acid injections, and electrical stimulation. Increased neurotrophin expression may facilitate further seizure activity, suggesting potential roles in mechanisms

underlying epilepsy and kindling.

Specific mechanisms governing activity-dependent regulation of trophin gene expression have been characterized in vitro. Direct pharmacological depolarization markedly increases hippocampal neuron NGF and BDNF mRNAs, potentially leading to increased protein concentrations. BDNF expression increases almost fourfold within 3 hours and attains a maximal increase within 6 hours whereas NGF mRNA exhibits a smaller and more sluggish response. Activation of specific neurotransmitter receptors mimics these effects. In particular, excitatory glutamatergic and cholinergic inputs increase neurotrophin mRNAs and inhibitory g-aminobutyric acid (GABA)-ergic inputs decrease message levels. The persistent activation of neurotrophin genes by transient electrical activity may be critical in triggering changes in downstream gene expression. In turn, these sequelae may result in long-term changes in neuronal function.

Neurotrophins and Long-Term Potentiation Direct stimulation of intrinsic anatomical pathways also regulates neurotrophin gene expression, providing a critical physiological context. In vivo, pharmacological activation of hippocampal afferent pathways increases BDNF and NT-3 mRNAs. Electrical stimulation evokes similar effects in vitro. Of particular interest is the relationship between neurotrophin gene expression and long-term potentiation, which is a form of activity-dependent synaptic plasticity most commonly studied in the hippocampus; it has been invoked as a cellular substrate for learning and memory. Stimulation patterns that induce long-term potentiation increase both BDNF and neurotrophin-3 messages in the hippocampus. Moreover, gene knockout animals with null mutations (i.e., inactivation) of the BDNF gene exhibit impaired long-term potentiation, which supports the contention that neurotrophins play critical roles in synaptic plasticity and perhaps in memory.

Neurotrophins and Stress Neurotrophin gene expression in the brain is modulated by stressful stimuli, with potential relevance to neuropsychiatric disease. Physiological responses to stress resemble symptoms of depression, and chronic stress exacerbates clinical depression. Chronic stress also increases neuronal vulnerability to a wide variety of insults (e.g., metabolic toxins, seizure activity), mediated at least in part by increased circulating glucocorticoids. BDNF expression is decreased in response to short-term or long-term stressors (e.g., forced immobilization), which may contribute to resulting hippocampal damage. NT-3 mRNA levels increase in response to stress whereas treatment with antidepressant agents produces a decrease. It is not yet clear whether these changes in neurotrophin availability play a role in adaptive changes. These complex interactions between environmental stimuli and neurotrophins may mediate cognitive changes that occur in response to stress and in the course of psychiatric or degenerative disease, an area that is being actively studied.

Although the regulation of neurotrophin gene expression has been intensely studied, much less is known about the mechanisms underlying neurotrophin release. This is crucial, however, because it is only after release and subsequent receptor binding that neurotrophins exert biological effects. Recent studies have begun to identify the mechanisms that control neuronal neurotrophin release. In the hippocampus, both NGF and BDNF are released from neuronal axon terminals and dendritic processes in an activity-dependent manner. Atypically, release is dependent on sodium, but independent of extracellular calcium, although intracellular calcium stores are involved. Much remains to be elucidated concerning the spatial localization and kinetics of trophin release. Nevertheless, activity-dependent regulation of neurotrophin gene expression and release provides a set of mechanisms through which the external or internal environment may regulate availability.

ROLES OF NEUROTROPHINS DURING DEVELOPMENT

Neurotrophins play an essential role in coordinating diverse developmental processes. The most dramatic effects of NGF are on the survival of peripheral sensory and sympathetic neurons. During development, there is widespread, naturally occurring cell death, and the survival of many neuronal populations appears to be dependent on access to target-derived trophic factors. In many targets, neuronal innervation density is proportional to NGF synthesis. Thus, the neurotrophic hypothesis states that competition for limiting amounts of neurotrophins provides a mechanism to match the number of afferent neurons to target size. Lack of NGF results in the loss of most sympathetic and some sensory neurons. Exogenous NGF, on the other hand, rescues a larger number of neurons than normally survive during the period of developmental cell death.

The effects of neurotrophins on survival are complex and appear to be critically dependent on the specific cellular context. Neuronal cell death arises from two distinct processes: apoptosis or necrosis. Apoptotic or programmed cell death is an active process requiring new protein synthesis, and can be induced in vitro by serum deprivation or calcium channel blockers. This type of cell death is commonly associated with development as well as with neurodegenerative disease. Apoptosis can be markedly reduced or prevented by treatment with neurotrophic factors, consistent with their role as survival-promoting agents. Neurotrophins also prevent cell death after axotomy, an apoptotic type of cell death caused by trophic factor deprivation. However, necrotic cell death often occurs after an acute insult or injury, such as oxygen-glucose deprivation; this type of cell death can be enhanced by neurotrophins under some conditions. The situation is complicated by the fact that both necrotic and apoptotic cell death occur under most complex conditions, such as stroke and brain trauma. Thus, the therapeutic potential of trophic factors after neuronal injury may be critically related to the specific nature of the insult, the temporal profile of the derangement, and its physiological context.

Trophic factors exert widespread actions throughout the nervous system and are not limited to the foregoing examples. Other neurotrophin family members act on specific neuronal populations during development. For example, BDNF, NT-3, and NT-4/5 promote the survival of developing motor neurons. BDNF and NT-4/5 acting through the trkB receptor, promote the survival of nodose ganglion neurons. BDNF also increases expression of substance P in neural–crest-derived sensory neurons. Many other neurotransmitter and neuropeptide systems are also targets for neurotrophin modulation.

NEUROTROPHINS IN THE CENTRAL NERVOUS SYSTEM

Neurotrophins and their receptors are widely distributed in the central nervous system, with distinct differences in modes of action from the peripheral model. Trophic support in the central nervous system can be derived from target neurons as well as afferent neurons and local support cells, including astrocytes, oligodendrocytes, and microglia. In addition, brain neurons are not critically dependent on a single factor for survival, but instead have overlapping trophic dependence. This has been most clearly shown in animals with targeted inactivation of specific neurotrophin genes or their receptors. Although specific peripheral neuronal populations are severely depleted in these animals, there is little dramatic neuronal loss in the brain.

One of the best-characterized NGF-responsive cell populations in the brain are the cholinergic neurons of the basal forebrain. These neurons provide a dense modulatory input to the hippocampus. Taken together, the basal forebrain-hippocampal system plays a central role in attention, learning, and memory, with a particular emphasis on spatial learning and declarative memory. The prominent degeneration of basal forebrain cholinergic neurons in dementia of the Alzheimer's type is thought to be responsible for many of the ensuing cognitive deficits of the disorder. NGF enhances the survival of these neurons after axotomy or injury in animal models, and modulates neuronal function by increasing activity of choline acetyltransferase, an essential enzyme in acetylcholine biosynthesis. Thus, NGF may be clinically efficacious in rescuing this degenerating neuronal population and enhancing function of surviving neurons in disease states, with potential therapeutic implications.

NGF also has actions in the spinal cord that play a role in nociception. NGF promotes the survival of pain fibers and induces expression of voltage-gated ionic conductances in spinal nociceptive neurons, including induction of the peripheral nerve-type PN-1 sodium channel, which may play a role in the regulation of nociception. Animals with a targeted deletion of the NGF gene or chronic exposure to NGF blocking antisera exhibit decreased responsiveness to pain and temperature and a selective loss of small-diameter dorsal root ganglion neurons. Trophic factor interventions may therefore be useful in alleviating chronic pain syndromes.

BDNF and NT-4/5 acting via the trkB receptor, have a range of effects on multiple neuronal populations implicated in neuropsychiatric and neurodegenerative disease. These include the basal forebrain cholinergic neurons as well as dopaminergic neurons of the substantia nigra. Degeneration of nigral dopamine neurons results in the motor deficits observed in Parkinson's disease. BDNF enhances expression of dopamine synthetic enzymes, and protects these cells from exogenous neurotoxins. Trophic factors also increase axonal sprouting, resulting in increased connectivity with target neurons that may compensate for neuronal loss. This range of actions provides potential therapeutic opportunities for treating degenerative disease.

BDNF has also been implicated in the pathophysiology of mood disorders, including depression. Current treatment for clinical depression centers on enhancement of serotonergic neurotransmission, through the use of serotonin-specific reuptake inhibitors. Administration of BDNF also evokes antidepressant-like effects in some animal models, which may result from changes in the activity of serotonergic systems. BDNF promotes sprouting of serotonergic axons, increases serotonin synthesis, and dramatically enhances the survival of these neurons after exposure to neurotoxins. Because the typical therapeutic effects of antidepressant medications take several weeks to achieve clinical efficacy, it is possible that growth–factor-induced changes in neuronal morphology or connectivity might play a role in these processes. Thus, BDNF may ameliorate neuropsychiatric conditions that involve dysfunction of monoaminergic (e.g., serotonergic, noradrenergic) systems.

Neurotrophins and Synaptic Transmission Synapses are the main communicative junctions between neurons, and modulation of synaptic transmission is thought

to be central to learning and memory, cognition, and plasticity of circuits underlying motivation and emotion. In addition to well-established effects on neuronal survival, growth, and differentiation, recent work indicates that neurotrophins also acutely regulate synaptic efficacy, greatly expanding their functional relevance. Neurotrophins affect synaptic transmission through multiple mechanisms of action, including induction of ion channel expression, enhancement of presynaptic neurotransmitter release, and modulation of postsynaptic receptor responsiveness. These newly discovered effects have been identified both in the developing and mature nervous systems, indicating that mechanisms responsible for the formation of neural circuits during development also play roles in the dynamic regulation of activity within established circuits.

Direct presynaptic effects on neurotransmitter release have been best characterized at the neuromuscular junction, one of the more accessible and easily studied synapses in the nervous system. At developing neuromuscular synapses in culture, BDNF and NT-3 potentiate evoked synaptic currents within minutes of exposure by enhancing acetylcholine release from presynaptic nerve terminals. Increased transmitter release results from local effects at the terminal and does not require cell body signaling. In addition to this short-term effect, long-term exposure to BDNF or NT-3 for several days enhances synaptic maturation through effects on presynaptic processes. Similar effects are observed in mammalian systems in vitro, where NGF and BDNF directly potentiate the release of acetylcholine and glutamate in the hippocampus and cortex. Moreover, this enhanced transmitter release augments synaptic transmission in a subpopulation of neurons.

Neurotrophins also modulate synaptic activity via postsynaptic mechanisms. In the hippocampus, BDNF elicits sustained enhancement of action–potential-driven synaptic activity within minutes of exposure. This increase results from modulation of postsynaptic responsiveness to excitatory input via phosphorylation-dependent mechanisms (Fig. 1.7-2). TrkB receptor activation is critical for this response because neurotrophin-4, another trkB ligand, elicits similar effects. In contrast, the related neurotrophins NGF and neurotrophin-3, as well as the unrelated growth factors epidermal growth factor and basic fibroblast growth factor, do not share this effect. A postsynaptic locus of action is substantiated by the finding that the full-length trkB receptor is an intrinsic component of the postsynaptic density, a specialization of the postsynaptic membrane that anchors neurotransmitter receptors and second messenger signaling molecules. Furthermore, BDNF acutely enhances phosphorylation of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptor in the isolated postsynaptic density, leading to altered receptor function. BDNF-induced modulation of NMDA receptor activity appears to play a critical role in the modulatory effects of BDNF on synaptic transmission.

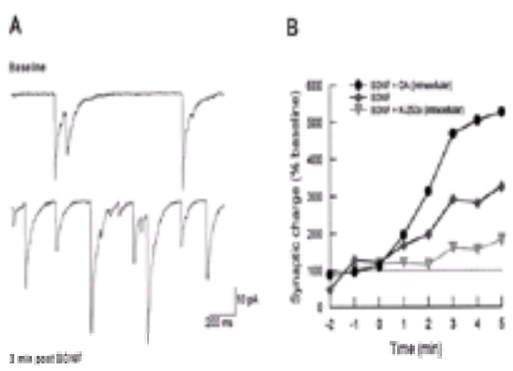

FIGURE 1.7-2 A. BDNF-induced potentiation of synaptic transmission. Whole-cell voltage clamp recordings from a hippocampal neuron before and 3 minutes after exposure to BDNF. Downward deflections represent excitatory postsynaptic currents. **B**. The postsynaptic component of BDNF modulation. Time course of a recording from an individual neuron shows the effect of bath-applied BDNF (50 ng/ml; diamonds). The effect of BDNF is decreased by postsynaptic injection of the trk tyrosine kinase inhibitor K-252a (200 nM; triangles), and enhanced by injection of the phosphatase inhibitor okadaic acid (OA; 0.5 μM; circles). Each point represents the average synaptic activity for a 1-minute period. (Reprinted with permission from Levine ES, Dreyfus LF, Black IB, Plummer MR: Proc Natl Acad Sci USA *92:*8074, 1995.)

Thus, neurotrophins potentiate synaptic transmission via both presynaptic and postsynaptic effects, increasing the efficacy of synaptic transmission and the probability of triggering postsynaptic action potentials. The effects of neurotrophins on synaptic activity may have specific consequences for activity-dependent plasticity in the hippocampus, including long-term potentiation, a potential cellular substrate of learning and memory. Animals with a targeted deletion of the BDNF gene have deficits in long-term potentiation that can be restored by reintroduction of BDNF. Endogenous BDNF enhances long-term potentiation, although specific mechanisms are controversial. Modulation of postsynaptic glutamate receptors by neurotrophins may contribute to changes in long-term potentiation. Similar mechanisms are thought to occur in widespread areas of the brain, subserving multiple adaptive processes. Taken together, these data point to an important role for neurotrophins in the regulation of synaptic plasticity and potentially for the learning and memory processes that are dependent on this type of plasticity. Increasing neurotrophin availability may represent a novel approach towards reversing cognitive deficits resulting from injury, disease, or aging.

In addition to direct effects on synaptic efficacy, neurotrophins also regulate expression of voltage-gated ion channels that regulate neuronal excitability. This has been most closely examined in a pheochromocytoma cell line (PC-12) that differentiates into a sympathetic neuron-like phenotype in response to NGF. In these cells NGF induces the expression of multiple types of sodium, potassium, and calcium channels. The increase in sodium channel expression is mediated at least in part by cyclic adenosine-monophosphate–dependent protein kinase (PKA). NGF also increases expression of voltage-gated sodium, calcium, and potassium currents in a neuroblastoma cell line. Whereas continuous exposure to NGF causes the induction of a family of sodium channels, brief exposure selectively induces expression of the peripheral nerve-type sodium channel gene *PN-1*. Although these effects have been typically characterized in clonal cell lines, similar mechanisms occur in primary neurons. For example, functional expression of N-type calcium (Ca^{2+}) channels is greatly enhanced by NGF in sympathetic neurons; in dorsal root ganglia, NGF accelerates the acquisition and diversity of sodium currents. Thus, modulation of ionic currents represents another important facet of neurotrophin modulation.

These mechanisms also occur in the brain, with functional implications for psychiatric and neurodegenerative disease resulting from basal forebrain-hippocampal dysfunction. In basal forebrain cholinergic neurons, NGF increases both L-type and N-type components of voltage-gated calcium currents. This effect is specific to NGF because BDNF does not have a similar effect. Calcium entry via these voltage-dependent channels regulates neurotransmitter release and neuronal firing patterns and also modulates neuronal gene expression. Thus, calcium influx via these channels may mediate some of the well-known trophic actions of NGF. Since these neurons provide a dense cholinergic input to the hippocampus, this effect of NGF may also play a role in modulation of hippocampal synaptic transmission. Regulation of these ionic currents could therefore have long-lasting consequences for neuronal responsiveness to synaptic inputs. Importantly, the memory loss and cognitive impairment that accompanies dementia of the Alzheimer's type is thought to result from deficits in the basal forebrain-hippocampal system. Neurotrophins therefore are potential candidates for underlying pathophysiology and as a potential treatment opportunity.

Regulation of Neuronal Connectivity In addition to dynamic changes in electrical activity, modulation of neuronal function can result from changes in neural architecture and circuit formation. Along these lines, neurotrophins may influence synaptic transmission in the brain via direct effects on axonal outgrowth, dendritic morphology, and synaptic connectivity. In addition to promoting survival (neurotrophism), these factors also have neurotrophic effects, guiding the direction of neurite growth. For example, growth cones of sensory neurons in vitro grow towards a source of NGF. It is not known whether NGF acts in vivo to guide axons to their proper targets, but once they reach their targets, NGF can regulate the shape and arborization of terminal contacts. NGF promotes neurite elaboration in developing cerebellar Purkinje neurons, and these effects are dependent on neuronal activity, requiring simultaneous exposure to NGF and depolarizing agents. NT-3 also enhances neurite outgrowth and branching in cultures of embryonic rat hippocampus. In developing cortex, endogenous neurotrophins regulate the development of characteristic dendritic branching patterns and laminar organization. Specific neurotrophins increase the length and complexity of dendrites of cortical neurons. Basal dendrites of neurons in each cortical layer respond most strongly to a single neurotrophin, whereas apical dendrites respond to several neurotrophins. These effects are also dependent on interactions with ongoing neuronal activity.

The physiological relevance of these effects may relate to activity-dependent formation of neuronal circuitry during development and its reorganization after injury. For example, during the development of the visual system, axons from the thalamus carrying visual information become segregated into eye-specific patches (ocular dominance columns) within their target in the primary visual cortex. This reorganization results from activity-dependent synaptic competition between axons representing the two eyes. Infusion of BDNF or NT-4/5 into cat primary visual cortex blocks column formation. Neurotrophin-receptor antagonists, which block the actions of endogenous neurotrophins, also inhibit the formation of these columns. These data suggest that neurotrophins, normally present in limiting amounts within visual cortex, are necessary for the selective growth and segregation of thalamic axons into ocular dominance columns.

OTHER NEUROTROPHIC FACTOR FAMILIES

Many other growth and trophic factor families are expressed in the CNS, with specific neuronal targets that overlap and are distinct from neurotrophin targets. Examples include glial-derived neurotrophic factor (GDNF), a recently discovered trophic molecule isolated from an astrocyte cell line, that is a distant member of the transforming growth factor-b (TGF-b) family. GDNF is the most potent survival factor yet identified for dopamine neurons, which degenerate in Parkinson's disease, and motor neurons, which degenerate in amyotrophic lateral sclerosis. Thus, trophic factors may represent a promising therapeutic approach to rescuing these vulnerable neuronal populations. Dopamine neurons are also implicated in the pathophysiology of schizophrenia, and most antipsychotic medications target specific dopamine receptor subtypes. Dopaminergic neurotransmission may also be acutely modulated by trophic factors, providing therapeutic alternatives for clinical intervention. Other important growth factors include ciliary neurotrophic factor, which in addition to promoting survival of cultured ciliary neurons also acts on a wide range of neurons, including motor neurons and basal forebrain cholinergic neurons. The fibroblast growth factor family promotes cellular proliferation (mitogenesis) in a variety of cell types. In addition, basic fibroblast growth factor and acidic fibroblast growth factor, members of a gene family that now numbers seven related factors, also promote the survival and differentiation of numerous cellular types. Thus, disorders involving various transmitter systems may be potential targets for intervention using different trophic factors.

Many questions regarding the functional roles of trophic factors remain unanswered. Are there discrete families of factors that preferentially act on specific neuronal systems? What is the function of the overlapping responsiveness to different neurotrophins? Are there distinct signaling pathways mediating different downstream trophic effects? The recently discovered neuromodulatory effects of trophic factors suggest that targeting these molecules may represent a novel strategy for subtly fine-tuning nervous system function. The growing knowledge of neurotrophin roles in the adult nervous system will have important consequences for understanding brain function and dysfunction as it relates to neuropsychiatric and neurodegenerative disease.

SUGGESTED CROSS-REFERENCES

Neuroanatomy is discussed in Section 1.2 and Section 1.3. Excitatory amino acid neurotransmitters are presented in Section 1.5. Section 1.8 discusses intraneuronal signaling pathways, and Section 1.9 covers basic electrophysiology.

SECTION REFERENCES

Barbacid M: The trk family of neurotrophin receptors. J Neurobiol *25:*1386, 1994.

Bothwell M: Functional interactions of neurotrophins and neurotrophin receptors. Annu Rev Neurosci *18:*223, 1995.

Cabelli RJ, Hohn A, Shatz CJ: Inhibition of ocular dominance column formation by infusion of NT-4/5 or BDNF. Science *267:*1662, 1995.

*Chao MV, Hempstead BL: p75 and trk: A two-receptor system. Trends Neurosci *18:*321, 1995.

Cohen-Cory S, Dreyfus CF, Black IB: NGF and excitatory neurotransmitters regulate survival and morphogenesis of cultured cerebellar Purkinje cells. J Neurosci *11:*462, 1991.

DiCicco-Bloom E, Friedman WJ, Black IB: NT-3 stimulates sympathetic neuroblast proliferation by promoting precursor survival. Neuron *11:*1101, 1993.

Elliott RC, Inturrisi CE, Black IB, Dreyfus CF: An improved method detects differential NGF and BDNF gene expression in response to depolarization in cultured hippocampal neurons. Brain Res Mol Brain Res *26:*81, 1994.

Gall CM, Isackson PJ: Limbic seizures increase neuronal production of messenger RNA for nerve growth factor. Science *245:*758, 1989.

Kang H, Schuman EM: Long-lasting neurotrophin-induced enhancement of synaptic transmission in the adult hippocampus. Science *267:*1658, 1995.

Kang HJ, Welcher AA, Shelton D, Schuman EM: Neurotrophins and time: Different roles for trkB signaling in hippocampal long-term potentiation. Neuron *19:*653, 1997.

*Kaplan DR, Stephens RM: Neurotrophin signal transduction by the trk receptor. J Neurobiol *25:*1404, 1994.

Korsching S: The neurotrophic factor concept: A reexamination. J Neurosci *13:*2739, 1993.

*Levi-Montalcini R: The nerve growth factor 35 years later. Science *237:*1154, 1987.

Levine ES, Crozier RA, Black IB, Plummer MR: Brain-derived neurotrophic factor modulates hippocampal synaptic transmission by increasing N-methyl-D-aspartic acid receptor activity. Proc Natl Acad Sci USA *95:*10235, 1998.

Levine ES, Dreyfus CF, Black IB, Plummer MR: Brain-derived neurotrophic factor rapidly enhances synaptic transmission in hippocampal neurons via postsynaptic tyrosine kinase receptors. Proc Natl Acad Sci USA *92:*8074, 1995.

Levine ES, Dreyfus CF, Black IB, Plummer MR: Differential effects of NGF and BDNF on voltage-gated calcium currents in embryonic basal forebrain neurons. J Neurosci *15:*3084, 1995.

Lindholm D, Castren E, Berzaghi M, Blochl A, Thoenen H: Activity-dependent and hormonal regulation of neurotrophin mRNA levels in the brain—implications for neuronal plasticity. J Neurobiol *25:*1362, 1994.

Lo DC: Neurotrophic factors and synaptic plasticity. Neuron *15:*979, 1995.

Lu B, Yokoyama M, Dreyfus CF, Black IB: Depolarizing stimuli regulate nerve growth factor gene expression in cultured hippocampal neurons. Proc Natl Acad Sci USA *88:*6289, 1991.

Meakin SO, Shooter EM: The nerve growth factor family of receptors. Trends Neurosci *15:*323, 1992.

*Mufson EJ, Kroin JS, Sendera TJ, Sobreviela T: Distribution and retrograde transport of trophic factors in the central nervous system: Functional implications for the treatment of neurodegenerative diseases. Prog Neurobiol *57:*451, 1999.

Patterson SL, Grover LM, Schwartzkroin PA, Bothwell M: Neurotrophin expression in rat hippocampal slices: A stimulus paradigm inducing LTP in CA1 evokes increases in BDNF and NT-3 mRNAs. Neuron *9:*1081, 1992.

Siuciak JA, Lewis DR, Wiegand SJ, Lindsay RM: Antidepressant-like effect of brain-derived neurotrophic factor (BDNF). Pharmacol Biochem Behav *56:*131, 1997.

*Snider WD: Functions of the neurotrophins during nervous system development: What the knockouts are teaching us. Cell *77:*627, 1994.

Snider WD, Lichtman JW: Are neurotrophins synaptotrophins? Molec Cell Neurosci *7:*433, 1996.

Snider WD, Wright DE: Neurotrophins cause a new sensation. Neuron *16:*229, 1996.

Suen PC, Wu K, Levine ES, Mount HTJ, Xu JL, Lin SY, Black IB: Brain-derived neurotrophic factor rapidly enhances phosphorylation of the postsynaptic N-methyl-D-aspartate receptor subunit 1. Proc Natl Acad Sci USA *94:*8191, 1997.

*Thoenen H: Neurotrophins and neuronal plasticity. Science *270:*593, 1995.

Thoenen H, Bandtlow C, Heumann R: The physiological function of nerve growth factor in the central nervous system: Comparison with the periphery. Rev Physiol Biochem Pharmacol *109:*145, 1987.

Wetmore C, Olson L: Expression and regulation of neurotrophins and their receptors in hippocampal systems. Hippocampus *3:*171, 1993.

Wu K, Xu JL, Suen PC, Levine ES, Huang YY, Mount HTJ, Lin SY, Black IB: Functional trkB neurotrophin receptors are intrinsic components of the adult brain postsynaptic density. Mol Brain Res *43:*286, 1996.

Zafra F, Castren E, Thoenen H, Lindholm D: Interplay between glutamate and gamma-aminobutyric acid transmitter systems in the physiological regulation of brain-derived neurotrophic factor and nerve growth factor synthesis in hippocampal neurons. Proc Natl Acad Sci USA *88:*10037, 1991.

1.8 INTRANEURONAL SIGNALING PATHWAYS

JAY M. BARABAN, M.D., FH.D.

Prior to delineating the organization of specific intraneuronal signaling pathways, it is important to consider, in general terms, their role in helping neurons interpret and respond to the barrage of afferent stimulation impinging on them continuously. From an evolutionary perspective, second messenger systems predate neurotransmitters and neurotrophins, examples of first messengers detected by cell surface receptors. Before the advent of neurotransmitters, prokaryotic organisms relied on cyclic adenosine monophosphate (cAMP) and other intracellular signaling pathways to coordinate diverse responses located in disparate parts of these unicellular organisms to changes in ambient nutrients or conditions. Neurotransmitters and neurotrophins have evolved subsequently to take advantage of these internal signaling pathways that have undergone a parallel growth process.

Intraneuronal signaling pathways do more than merely enlarge the sphere of influence of afferent stimuli beyond the local environment of the cell surface receptor. With the aid of these internal signaling pathways, the postsynaptic neuron partakes actively in amplifying or muting the initial signal conveyed by a particular receptor and is not relegated to the subservient status of a totally passive partner in responding to these external cues. Although each neurotransmitter receptor is dedicated to recognizing a specific molecular structure, the flexibility inherent in the organization of intraneuronal signaling pathways allows that recognition event to have entirely different meanings depending on the current context of the signal as well as the neuron's past experience. In other words, these internal signaling pathways empower neurons to shape their responses to incoming stimuli; collectively, these changes in ensembles of neurons are of paramount importance in enabling the nervous system to adapt to its environment and learn from experience.

The concept that neurons or networks of neurons have the ability to modify their responses to a given stimulus is also directly relevant to understanding the actions of psychotropic drugs. In the absence of second messenger systems, drugs would be expected to exert consistent effects with repeated administration. Thus, all of the complex time-dependent changes in psychotropic drug action, such as tolerance to opiates or benzodiazepines, or the delayed therapeutic response to antidepressant or antipsychotic drugs, ultimately result from the ability of neurons, through their internal signaling pathways, to mount a compensatory response to this form of stimulation. In some instances, this adaptation may run counter to the desired clinical effect, such as tolerance to the analgesic effects of opioids or substance dependence. Alternatively, this active adaptation can be highly desirable because it presumably underlies the therapeutic action of antidepressant or antipsychotic agents, which develop after a lag of days to weeks following the onset of treatment. These examples underscore the direct relevance of intraneuronal signaling pathways to the most challenging problems facing psychiatry.

Although the overwhelming majority of psychiatric drugs target extracellular receptors or uptake sites, the explosion of information on intraneuronal signaling pathways suggests that these may represent suitable drug targets. In particular, the availability of transgenic approaches to examine the phenotype caused by deleting a specific component of a signaling pathway will be invaluable in developing a new generation of psychiatric drugs aimed at signaling pathways that function beneath the neuronal surface. Furthermore, insights gleaned from the ongoing search for genes involved in inherited psychiatric syndromes may help to focus attention on specific pathways as therapeutic avenues that are capable of compensating for these pathological defects.

Psychiatrists have long been taught that a true understanding of normal and abnormal behavior requires an appreciation of the interplay of forces lurking beneath the surface. Advances in defining the signaling pathways mediating neurotransmitter and neurotrophin action indicate that this notion rings true at the cellular level. Accordingly, deciphering the logic of intraneuronal signaling pathways is of paramount importance in understanding how neurons behave and represents a new frontier in dissecting the molecular and cellular basis of behavior and of the action of psychiatric drugs.

MAJOR SIGNALING PATHWAYS

Cyclic AMP System The classic cross-perfusion experiments conducted by Otto Loewi at the turn of the century led to the identification of acetylcholine as a neurotransmitter and revolutionized the conception of synaptic transmission. In a similar manner, the discovery of cAMP by Sutherland and Rall nearly half a century later using an analogous approach established the principle that intracellular transmitters or second messengers are instrumental in conveying information from cell surface receptors to their targets within the cell. The accumulation of decades of research on this system has revealed its operating principles in great detail and has served as a blueprint for deciphering other signaling pathways as well.

For this signaling pathway, generation of cAMP is controlled by the balance between the activity of its synthetic enzyme, adenylyl cyclase, which converts adenosine triphosphate (ATP) to cAMP, and phosphodiesterase, which cleaves cAMP to an inactive breakdown product, AMP. Adenylyl cyclase is regulated by cell surface receptors, via a family of adapter proteins referred to as *G proteins* because they bind guanosine 5'-triphosphate (GTP) when the receptor is activated. In this activated configuration, the a subunit of the G protein complex dissociates from the bg subunits, enabling it to regulate cyclic AMP formation. The a subunit possesses an intrinsic GTPase activity that converts GTP to guanosine diphosphate (GDP), which allows the subunits to reassemble into an inactive configuration.

Neurotransmitter receptors may couple to adenylate cyclase via different classes of G proteins, referred to as G_s or G_i, depending on whether they stimulate or inhibit cyclic AMP formation (Fig. 1.8-1). In this way, the net effect of the transmitter on a given neuron is determined by the specific receptor subtypes expressed on its surface. For example, norepinephrine stimulates adenylate cyclase via its interaction with b-adrenergic receptors, the type that speed heart rate, and it inhibits adenylate cyclase via the muscarinic cholinergic receptor subtype.

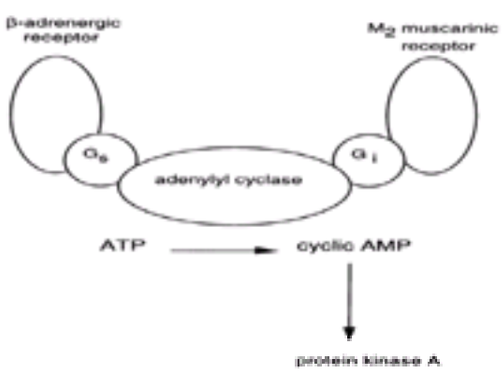

FIGURE 1.8-1 Organization of the cyclic AMP system. Generation of cyclic AMP from ATP by adenylyl cyclase can be stimulated or inhibited by receptor-G protein linkages. In the example shown, the b-adrenergic receptor coupled to G_s stimulates cyclic AMP synthesis whereas an M_2 muscarinic receptor inhibits this process via its linkage to G_i. Cyclic AMP exerts many of its cellular actions via its ability to activate protein kinase A.

Cyclic AMP exerts a wide variety of actions on neuronal function via its stimulatory effect on cyclic AMP-dependent kinase, which has a broad range of substrate proteins involved in regulating virtually every aspect of neuronal function from ion channel gating to axonal transport, much as the autonomic nervous system influences the activity of diverse organ systems to achieve a cohesive response. One target of the cAMP system that has been the focus of attention in recent years is a transcription regulatory factor that enables elevations in cyclic AMP to regulate gene expression. This factor, referred to as cAMP response element binding (CREB)

protein, binds to a short sequence of deoxyribonucleic acid (DNA) in the regulatory regions of its target genes. Phosphorylation of CREB on a specific serine residue is required for it to promote transcription of target genes. Thus, alterations in cyclic AMP levels can affect neuronal function over a broad range of time scales. Rapid effects can be induced by targeting ion channel gating or transmitter release machinery. On a more sluggish time scale, cyclic AMP can influence neurotransmitter synthesis or energy metabolism. Furthermore, longer-lasting changes in neuronal function can be achieved by this messenger as a result of its ability to control the expression of specific target genes.

Cyclic GMP Besides cyclic AMP, another cyclic nucleotide, cyclic guanosine monophosphate (GMP) has been identified as a second messenger regulated by neurotransmitter receptor stimulation. The discovery of cyclic GMP-dependent protein kinases suggested that both cyclic nucleotide systems followed similar blueprints. However, subsequent studies have revealed startling contrasts between these systems. The link between neurotransmitter receptor activation and stimulation of guanylate cyclase does not appear to rely primarily on G protein coupling. Instead, the available evidence indicates that elevations in intracellular calcium trigger increases in nitric oxide production, which in turn activate guanylyl cyclase (Fig. 1.8-2). This complex cascade introduces a novel element, (i.e., a gaseous second messenger that is capable of diffusing both within cells as well as across membranes to neighboring cells, blurring the semantic distinction between first and second messengers). Nitric oxide has the ability to coordinate responses in clusters of cells within its sphere of diffusion, ensuring that neighboring neurons can be made aware of the status of postsynaptic neurons in their vicinity, as well as the activity of afferents emanating from distant parts of the nervous system. Furthermore, nitric oxide has the ability to influence afferent terminals, a retrograde feedback function that may be important in regulating the activity of incoming stimuli. Thus, the discovery of nitric oxide as a neuronal messenger breaches the classical notion that synapses convey information in only one direction. Recent studies have also suggested that another diffusible gas, carbon monoxide, which is also capable of activating guanylate cyclase, may also function in an analogous fashion to nitric oxide.

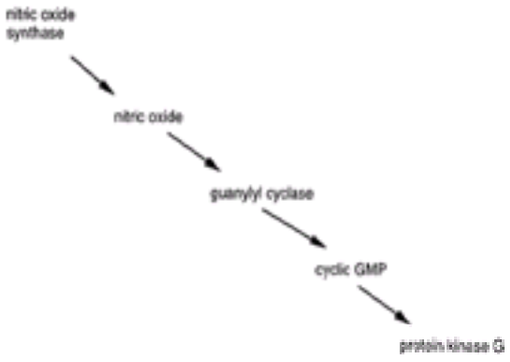

FIGURE 1.8-2 Organization of the cyclic GMP system. In contrast to cyclic AMP, cyclic GMP synthesis is regulated by stimulation of guanylyl cyclase by nitric oxide. Beyond this point in the pathway, cyclic GMP mimics cyclic AMP as it acts by stimulating its cognate kinase, protein kinase G.

Phosphoinositide (PI) Characterization of the neurotransmitter receptors coupled to the cAMP system revealed that there were many receptors that did not act via this second messenger pathway. This discrepancy generated interest in the possible existence of other second messenger systems operating in parallel with the cAMP system. This line of research came to fruition in the early 1980s with the emergence of a coherent view of the phosphoinositide second messenger system. This second messenger system parallels many aspects of the cAMP system (Fig. 1.8-3). Neurotransmitter receptor stimulation is coupled, via G proteins, to activation of a second messenger generating enzyme, phospholipase C. This enzyme cleaves inositol-containing phospholipids located in the plasma membrane into two second messengers, diacylglycerol and inositol trisphosphate (IP_3). Thus, activation of neurotransmitter receptors linked to the phosphoinositide system generates a pair of second messenger signals that can affect cellular responses via distinct pathways.

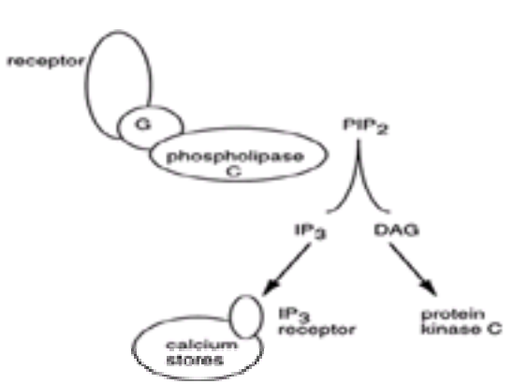

FIGURE 1.8-3 Organization of the PI system. In this system neurotransmitter receptor stimulation leads via G protein coupling to activation of phospholipase C. This enzyme cleaves the membrane phospholipid, PIP_2, into the second messengers, IP_3 and diacylglycerol (DAG). IP_3 releases calcium from intracellular stores; DAG stimulates protein kinase C.

Because the effects of cAMP are mediated to a large extent via activation of a kinase, it was generally assumed that each of these second messengers acted in a similar fashion. This turned out to be true of diacylglycerol, which activates protein kinase C, a kinase that is highly enriched in the brain. In contrast, IP_3 acts much like an intracellular transmitter that has its own receptor present on the cytoplasmic face of intracellular organelles that store calcium. Binding of IP_3 to its receptor triggers release of calcium from these intracellular stores. Calcium, in turn, is a second messenger in its own right that is capable of regulating many intracellular processes, ranging from regulation of ion channel activity to gene expression. Previous studies had led to an appreciation of the vital role calcium plays as a mediator of transmitter release following action potential invasion of nerve terminals. In that situation, calcium enters the nerve terminal via opening of voltage-dependent calcium channels. Thus, it had been assumed that neurotransmitter regulation of intracellular calcium levels was mediated by their ability to elicit depolarization and subsequent activation of voltage-dependent calcium channels. In contrast, the IP_3 system provides an alternative route by which neurotransmitters can regulate intracellular levels of calcium, an important determinant of cellular function.

It is noteworthy that lithium (Eskalith) played an important role in the studies that led to the current understanding of the phosphoinositide second messenger system. It had been noted in studies aimed at defining the effects of lithium on the central nervous system that lithium caused a modest decrease in concentrations of inositol, a sugar closely related to glucose. In pursuing the basis for this effect, investigators noted that lithium was an effective inhibitor of a phosphatase that converted inositol phosphate into inositol. Lithium could then be used to force accumulation of inositol phosphates, providing a convenient means of measuring activation of the phosphoinositide system. As a result of lithium treatment, researchers were able to detect the presence of inositol trisphosphate and realized that it was generated by phospholipase C acting on PI biphosphate (PIP_2), producing both diacylglycerol and IP_3 in one reaction.

The discovery that lithium acts as an inhibitor of a key enzyme that generates free inositol needed to replenish inositol phospholipids has prompted the hypothesis that depletion of inositol and subsequent rundown of the PI cycle may underlie lithium's therapeutic action. However, this view has been challenged recently in light of animal studies demonstrating that inositol levels in brain are unaffected by lithium concentrations within its therapeutic range.

Direct Coupling Between G Proteins and Ion Channels The elucidation of the cAMP, cyclic GMP, and PI second messenger systems focused attention on the importance of diffusible small molecules, which acted much like intracellular neurotransmitters. Instead of being released from a presynaptic site and diffusing across the synaptic cleft to act at receptor sites on the postsynaptic side, they diffused from their site of formation on the cytoplasmic side of the plasma membrane to act on their receptors within the cell whether they be kinases, as in the case of cAMP, calcium, or diacylglycerol, or receptors, as found for IP_3. However, diffusible second messengers are not universal components of the signaling pathways that mediate the actions of G–protein–coupled neurotransmitter receptors. In many important situations, the G proteins themselves link neurotransmitter receptor activation to ion channels shortcircuiting the rest of the cascade. Prominent examples of this type

of arrangement are provided by opioid receptors and muscarinic receptors involved in vagal slowing of the heart. As these receptors had been shown to be linked to G_i and cause inhibition of adenylate cyclase, it had been taken for granted their important effects on ion channel regulation, in particular opening of potassium channels causing hyperpolarization, was a result of lowering cAMP concentrations. However, this theory was shattered by experiments in which restoration of cAMP concentrations was ineffective in reversing this effect. Analysis of this paradox revealed that even though G_i was involved in mediating this response, it was due to a direct effect of G_i on the potassium channels (Fig. 1.8-4). Subsequently, this direct coupling has been found to be a common type of linkage between neurotransmitter receptors and ion channels.

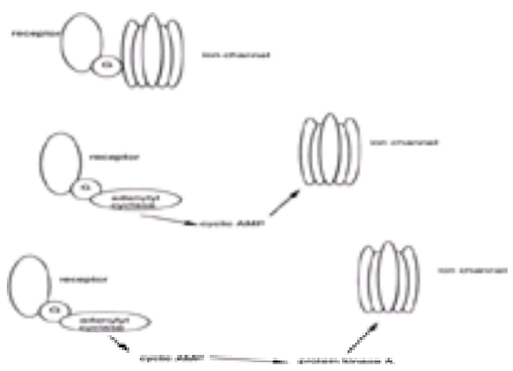

FIGURE 1.8-4 Regulation of ion channels by G-protein–coupled receptors. Not all the effects of cyclic nucleotides are mediated by protein kinases. As shown in upper panel, G proteins can directly link neurotransmitter receptors to ion channels. This configuration underlies regulation of potassium channels that slow the heart in response to vagal stimulation of muscarinic cholinergic receptors. Alternatively, cyclic nucleotides can directly affect ion channels in a kinase-independent fashion, as illustrated in the middle panel. Cyclicnucleotide–gated channels play a key role in photoreceptor responses to light. The conventional kinase-dependent pathway is shown in the bottom panel. This type of arrangement mediates the well-known ability of sympathetic stimulation to speed the heart rate via activation of b-adrenergic receptors.

In this regard, it is interesting to note that the generalization that cyclic nucleotides always act via kinases has also been debunked by studies in sensory neurons demonstrating that these signaling molecules interact directly with ion channels, without kinases acting as intermediaries. These alternate arrangements emphasize the notion that intracellular signaling cascades have evolved in ways that heighten their versatility, with each of the components having multiple signaling capacities.

Tyrosine Phosphorylation The landmark discovery that viral oncogenes were mutant versions of cellular genes that had been hijacked by tumor viruses has had a major impact on the entire field of tumor biology. In addition, characterization of the mode of action of several of these proto-oncogenes yielded another surprise—they represent a novel class of kinases that phosphorylate tyrosine residues, rather than serine or threonine residues that are targeted by all previously described mammalian kinases.

Since the late 1970s intense exploration of this novel form of phosphorylation has provided compelling evidence that it can exert powerful influences on neuronal function. Initially it was assumed that because tyrosine kinases were proto-oncogenes involved in regulating cellular proliferation, this type of phosphorylation would have little relevance to the nervous system, which has extremely low levels of cellular proliferation. Contrary to this presumption, surveys of the level of tyrosine phosphorylation present in various tissues revealed that the brain has one of the highest concentrations of phosphotyrosine-containing proteins as well as of tyrosine kinases.

Studies of tyrosine kinase signaling cascades have provided important new sights into the mode of action of nerve growth factor and other neurotrophins. Binding of growth factors to the extracellular portion of these receptors leads to activation of a tyrosine kinase domain located in the cytoplasmic tail of the receptor protein. It is thought that growth factor binding brings together two receptor molecules that then phosphorylate each other on tyrosine residues. This simple modification of the cytoplasmic tail converts it into a magnet for an array of signaling proteins that are brought together at the inner surface of the membrane. This arrangement triggers multiple divergent signaling cascades from this nidus. One of these branches can lead to activation of the PI system, as one of the isoforms of phospholipase C contains a domain that is attracted to the activated growth factor receptor. Another branch leads to activation of *ras*, a proto-oncogene that has G-protein–like properties.

Another family of tyrosine kinases has been identified that differs from the receptor tyrosine kinases in that it contains only the cytoplasmic domain. The absence of an extracellular ligand binding domain has prompted questions about how members of this family are regulated; these questions are still largely unresolved. Nevertheless, evidence is accumulating that these non-receptor tyrosine kinases, such as *fyn, src,* and *yes* play a critical role in multiple neuronal responses, including ion channel regulation.

The organization of the tyrosine kinase pathways provides an interesting contrast to the classical, second–messenger-based pathways outlined previously. First, the tyrosine kinase pathways do not utilize small, diffusible second messengers. In theoretical terms, this may have the advantage of maintaining a restricted spatial domain of signal propagation, as opposed to the classical second messenger systems, which have a larger sphere of influence. Second, the ability of multiple adapter proteins to interact with the activated cytoplasmic tail of tyrosine kinase receptors confers a remarkable degree of divergence as multiple signaling pathways can be engaged simultaneously. Thus, the importance of tyrosine phosphorylation signaling does not reside simply in the availability of another residue that is amenable to modification by phosphorylation; rather, it represents an intracellular signaling system built on an alternative set of architectural principles.

Identification of the signaling pathway downstream of Ras has opened up a new avenue to understanding the mechanism of action of growth factors and neurotransmitters. Members of the Ras family are referred to as *small G proteins* because of their lower molecular weight compared to their large G protein relatives linked to the cyclic AMP and PI systems. Like their cousins, Ras and other small G proteins bind GTP following activation of growth factor receptors and slowly hydrolyze it to GDP with the help of accessory proteins called *GTPase-activating proteins* (*GAP*). In contrast to the large G proteins that are involved in regulating the synthesis of small second messenger molecules or in regulating ion channel activity, activated *ras* has the ability to stimulate a cascade of kinases arranged in series, that lead to activation of mitogen-activated protein (MAP) kinase (Fig. 1.8-5). Unlike protein kinase A (PKA) or protein kinase C (PKC), each of these kinases is regulated by phosphorylation by an upstream activator kinase, instead of by small second messengers. However, like PKA or PKC, this kinase cascade has numerous targets within the cell ranging from influencing the organization of the cytoskeleton in the cytoplasm to control of gene expression in the nucleus. The pervasive influence of this signaling pathway helps to explain many of the remarkable effects of neurotrophins on neuronal growth and differentiation.

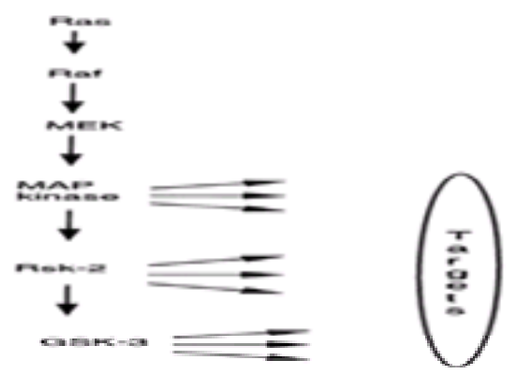

FIGURE 1.8-5 Kinase cascade activated by Ras. Ras, a member of the small G protein family, has been identified as an important mediator of growth factor responses. Activation of Ras stimulates a kinase, Raf, which regulates a series of downstream kinases. In contrast to the “classical” second messenger systems described above, these kinases are directly regulated by upstream kinases rather than by second messengers.

MAP kinase is particularly interesting from the perspective of neuronal signal transduction because it was initially identified using *MAP-2*, a neuron-specific cytoskeletal protein, as a preferred substrate. Once investigators realized that the kinase was widely distributed and also activated by mitogens and other extracellular agonists, its name was changed from *MAP-2 kinase* to *mitogen activated protein kinase* (MAP kinase) or *extracellular agonist regulated kinase* (ERK). Recent studies have demonstrated that MAP kinase in neuronal dendrites is associated with microtubules and indicate that MAP-2 is a physiological substrate of this kinase in vivo. Other substrates that have been identified include tyrosine hydroxylase and transcription regulatory factors, as well as another downstream kinase referred to as *Rsk-2*, which is capable of phosphorylating and activating CREB. The length of the kinase cascade distal to MAP kinase has also been stretched to add yet another kinase called (glycogen synthase kinase-3), which is downstream of Rsk-2. GSK-3 has received attention recently as the target of lithium, accounting for its teratogenic effects in several model systems of development. Further studies are needed to help elucidate whether this kinase may also play a role in mediating lithium's therapeutic action in mood disorders.

Cross-Talk Among Signaling Pathways The organization of intraneuronal signaling pathways allows for a high degree of interaction or cross-talk among pathways. For example, growth factor receptor activation can also engage the PI system initially identified as a target of neurotransmitter receptors. Conversely, activation of neurotransmitter receptors can also stimulate MAP kinase initially linked to growth factor receptor activation. Thus, neurotransmitters have the ability to influence the response to neurotrophins and vice versa. In this context it is noteworthy that neurotransmitters can also regulate the MAP kinase pathway. This interplay among signaling pathways increases their versatility by providing alternative routes for regulating many of the same effector proteins.

Although kinase regulation has been focused on, it is important to emphasize that there are analogous regulatory cascades involved in controlling dephosphorylation. In this way, the longevity of phosphorylation on a specific residue can vary dramatically depending on the substrate protein involved and the context it which it occurs. For example, PKA phosphorylation of dopamine regulated phosphoprotein-32 (DARPP-32), a protein phosphatase inhibitor that is highly enriched in neurons that receive dopaminergic innervation, is required for it to be functional. Thus, if a neuron receives simultaneous stimulation of both a calcium-dependent kinase and PKA, activation of DARPP-32 may greatly prolong the duration of phosphorylation of residues normally removed by the phosphatase inhibited by DARPP-32. An analogous cascade involves another phosphatase inhibitor, referred to as *I-2*, which is inhibited by GSK-3. Thus, in this case, activation of the MAP kinase/GSK-3 pathway would also affect the longevity of phosphorylation on residues regulated by the I-2–sensitive phosphatase. As a result of cross-talk between systems, coordinate activation of multiple pathways can have important synergistic effects.

Another level of cross-talk has been observed at the level of specific target proteins (Fig. 1.8-6). Rather than being substrates for specific kinases, the more common situation is that a given target is phosphorylated by multiple kinases. This overlapping of substrate specificity allows for complex patterns of regulation. For example, phosphorylation of a specific substrate by both protein kinase A and protein kinase C may have qualitatively different effects than modification by either alone. In addition, there are intriguing examples of conditional phosphorylation in which a protein only becomes a substrate for a protein kinase when it is first phosphorylated on a nearby residue by a distinct kinase. This sequential form of interaction confers a conditional switching mechanism that will only allow a substrate protein to be modified when a specific pre-condition is met. Furthermore, there are situations that represent the equivalent of “or” operations, when multiple kinases can substitute for each other in phosphorylating a specific residue. Therefore, if any one of the corresponding pathways is activated, the response will be triggered. A specific example of the “or” situation is presented by CREB. Phosphorylation on serine 133 by PKA converts this transcription factor into its active form. In addition, this same residue can be phosphorylated by calcium on calmodulin-dependent kinases, as well as Rsk-2, a kinase downstream of MAP kinase. Thus, activation of any one of these signaling pathways is sufficient to trigger CREB activation, allowing these distinct pathways to converge on CREB to induce changes in gene expression. In summary, multiple levels of cross-talk between signaling pathways transforms them from isolated pathways into a highly integrated network that possesses a high degree of sophistication and versatility in detecting and responding to incoming stimuli.

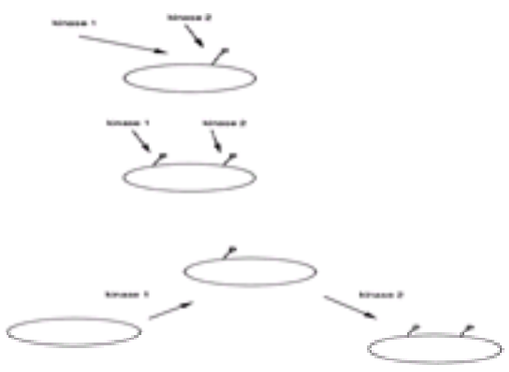

FIGURE 1.8-6 Cross-talk between kinases. As multiple kinases target individual protein substrates, multiple types of interaction can occur at this level. As shown in the top panel, there are instances where multiple kinases are capable of targeting the same residue. This arrangement is analogous to an “or” circuit, since activation of either kinase 1 or kinase 2 is sufficient to influence the target protein. In contrast, phosphorylation of a substrate by two different kinases on distinct residues may be needed to elicit a functional change in activity of the substrate (middle panel). As both kinases are necessary, this configuration is analogous to an “and” circuit. As illustrated in the bottom panel, the sequence of phosphorylation can also be crucial. In the example shown, both kinases are necessary, but they must be activated in the correct order because kinase 2 cannot act on the target unless it has already been phosphorylated by kinase 1.

SYNAPTIC PLASTICITY

In its simplest form, the postsynaptic response to neurotransmitter release can be mediated by a single protein complex. For example, nicotinic acetylcholine receptors are self-contained stimulus-response modules that both detect a stimulus, acetylcholine, and generate a response, passage of ion currents. In a similar vein, other members of this superfamily of ionotropic receptors, including g-aminobutyric acid (GABA) and glutamate receptors, have the ability to function in a manner that is independent of the intracellular signaling pathways discussed. Thus, in contrast to growth factor or G-protein–coupled receptors, which often recruit elaborate cascades to elicit a response, the simplicity of self-sufficient ionotropic receptor complexes represents an optimal design for achieving reliability, precision, and speed. However, this view of ionotropic receptors as insulated from their social environment has had to be abandoned in the face of overwhelming evidence that this class of receptors is dynamically regulated by intraneuronal signaling pathways. Although these receptors do not rely on intraneuronal signaling pathways to operate ion channels, because these channels are an intrinsic feature of the receptor complex the linkage between ligand binding and ion channel gating is nevertheless subject to regulation by the network of intraneuronal signaling pathways just described. For example, phosphorylation of the GABA or glutamate receptors modulates their response to ligand exposure.

Long-Term Depression The principle that ion channels are regulated by second messenger pathways is of central importance in considering how neuronal responses are altered by experience. Perhaps, one of the best examples for which the intracellular pathways involved have been worked out is the paradigm of long-term depression induced in Purkinje cells of the cerebellum. In this model of synaptic plasticity, the responses of Purkinje neurons to activation of ionotropic glutamate receptors is reduced for extended periods of time, by coincident activation of multiple signaling pathways. Even though the direct response to glutamate receptor activation does not rely on second messenger systems, its amplitude is indirectly modulated by a network of intraneuronal signaling pathways. This arrangement appears to be a general feature of synaptic transmission in which slower, second-messenger–based signaling pathways have a major impact on the fast modes of synaptic transmission.

In the specific example of long-term depression in cerebellar Purkinje cells, the plasticity observed is triggered by coordinate activation of two distinct classes of afferents to these neurons. Purkinje cells receive a major input from both climbing fibers arising from the inferior olive and parallel fibers that emanate from granule cells of the cerebellum. The parallel-fiber response in an individual Purkinje cell is decreased if that cell is coincidentally activated by a climbing fiber input. Analysis of this phenomenon has revealed that it is dependent on coincident activation of two distinct types of glutamate receptors: an a-amino-3-hydroxyl 5-methyl-4-isoxazole propionic acid (AMPA) type of ionotropic glutamate receptors and a G-protein–coupled (metabotropic) glutamate receptor that linked to the PI system. In addition, the climbing fiber input produces a massive depolarization of the Purkinje cell that opens voltage gated calcium channels. The coordinated presentation of these three signals: (1) depolarization with subsequent calcium entry, (2) activation of the PI/PKC system, and (3) AMPA receptor stimulation conspire to produce a prolonged decrease in responses elicited by AMPA receptor stimulation. Any combination of only two of these signals is insufficient to trigger this form of synaptic plasticity. The exact mechanism by which these signals are integrated or detected simultaneously has not yet been worked out. However, the requirement for simultaneous

activation of multiple pathways provides a compelling example of how these pathways can act synergistically to modulate neuronal responses. This example also underscores the importance of intraneuronal signaling pathways in regulating the responsiveness of ionotropic receptor channels.

Long-Term Potentiation The notion that coactivation of multiple second messenger pathways can have a qualitatively different impact than any one individually is also borne out in another well-known model of synaptic plasticity, *long-term potentiation*. In this paradigm, which has received intense attention because it provides a model of associative learning, modification of the synaptic responses to glutamate are also dependent on co-activation of multiple second messenger pathways. This requirement for coordinate activation of multiple pathways presumably represents a form of safeguard against changing synaptic weight inadvertently, which could have devastating effects on the nervous system. Accordingly, the requirements for synapse modification that underlie learning appear to have evolved in a way that entails the approval of multiple branches of the signaling network as a means of checking that the pattern of synaptic activity is of sufficient importance to warrant a long-term change in the synapses to be modified.

Role of NMDA Receptor Activation Prior to outlining the intracellular signaling pathways involved in long-term potentiation, it is first important to understand the key synaptic events that trigger plasticity in this paradigm. Studies of long-term potentiation of inputs to CA1 hippocampal neurons have highlighted the role of *N*-methyl-D-aspartate (NMDA) receptors as coincidence detectors. In this system basal levels of synaptic activity are mediated by activation of AMPA receptors. Even though the same synapses also express NMDA receptors, under quiescent conditions these do not open in response to glutamate because NMDA receptors have an additional requirement that must be met before they open. Unlike AMPA receptors, NMDA receptors will only pass current if they detect glutamate and membrane depolarization simultaneously. Thus, NMDA receptors represent an unusual hybrid between ligand-gated ion channels and voltage-gated ion channels. In contrast to AMPA glutamate receptors or nicotinic acetylcholine receptors, which open automatically whenever they detect glutamate or acetylcholine, NMDA receptors are indifferent to the presence of glutamate unless they simultaneously detect that the neuronal membrane is depolarized. Only when both these conditions are met does the NMDA receptor open its channel, allowing influx of depolarizing current.

This unusual property of NMDA receptors provides a molecular mechanism for conferring associative properties on long-term potentiation. When a glutamatergic input is of sufficient strength to meet both requirements of NMDA receptor activation, (i.e., glutamate and depolarization), then it will trigger a persistent potentiation of the response to that input. In contrast, an input that is too weak to elicit sufficient depolarization for NMDA receptor activation on its own will not undergo potentiation, unless it is presented simultaneously with other inputs that induce depolarization. Thus, the special requirements of the NMDA receptor allow weak inputs to be strengthened as long as they are presented simultaneously (i.e., associated) with strong stimulation of other inputs. This associative property of long-term potentiation has many of the same formal features as classical associative conditioning. For example, the conditioned stimulus, a bell in Pavlov's famous experiment, when presented by itself is insufficient to trigger a response, salivation. However, when the bell is sounded together with an unconditioned stimulus, such as food, then the temporal pairing of the two is capable of altering the response to the weak bell stimulus.

Role of Phosphorylation The associative property of this model of synaptic plasticity has focused attention on deciphering the intraneuronal signaling pathways that mediate the long-term change in synaptic transmission triggered by NMDA receptor stimulation. NMDA receptor activation leads to transient rises in intracellular levels of calcium making this second messenger an attractive candidate. Experiments demonstrating that intracellular calcium chelators block this form of long-term potentiation corroborated the critical role of calcium in this process. Subsequent studies investigated whether either of the calcium-sensitive kinases that are highly enriched in neuronal dendrites, calcium/calmodulin-dependent kinase II and protein kinase C, were critical for this process. Unexpectedly, inhibition of either kinase blocked long-term potentiation, indicating that activation of both is necessary to trigger this persistent change in synaptic responsiveness. Thus, in both long-term potentiation and long-term depression simultaneous activation of multiple signaling pathways is used to detect unusual forms of stimulation that warrant an adjustment of synaptic responses.

To underscore this general point, subsequent studies have also demonstrated that activation of the cyclic AMP system is also necessary for long-term potentiation to occur. The role of cyclic AMP in this paradigm may be related to its ability to suppress phosphatase activity. A key effect of PKA is to phosphorylate and thereby inactivate phosphatase inhibitors, such as DARPP-32. In this way, PKA on the one hand phosphorylates effector substrates and at the same time inactivates the phosphatase that erases its effects or the effects of other kinases.

Given the data linking phosphorylaton or dephosphorylation to synaptic plasticity, there is intense interest in defining the substrates involved. Evidence suggesting that the potentiation of synaptic responses reflects an increase in responsiveness of AMPA receptors located in the postsynaptic membrane has focused attention on these receptors themselves as candidate substrates. In this regard, it is interesting that the AMPA receptor is a substrate for PKA, PKC, and calcium/calmodulin-dependent kinase. On the other hand, there are those who subscribe to the alternate view that long-term potentiation may reflect a presynaptic alteration in transmitter release that is produced by a retrograde messenger, such as nitric oxide, generated postsynaptically.

Although the formal similarities between the associative properties of long-term potentiation and classical conditioning have provided compelling support for the hypothesis that this form of synaptic plasticity underlies associative learning, it has been difficult to gain experimental support linking this electrophysiological response to the behavioral phenomena. An important breakthrough in this area has been the utilization of transgenic animals with targeted mutations in genes encoding signaling molecules involved in long-term potentiation. The ability to examine the effect of these genetic alterations on behavior in the intact animal as well as on long-term potentiation in vitro has provided a means of bridging the gap between intraneuronal signaling pathways and behavior. Ongoing experiments in this area have validated the hypothesis that long-term potentiation and long-term depression represent the synaptic counterparts of learning. Application of this powerful genetic strategy to analysis of behavior provides an unprecedented means of dissecting the signaling pathways involved in modifying behavior in response to naturally occuring stimuli as well as to psychiatric drugs. As the technology employed by this approach is still developing rapidly, its application can be expected to become more widespread in the near future and provide a wealth of insights into the molecular substrates of a wide variety of behavioral responses.

Actions of Psychotropic Drugs In addition to providing insight into the molecular mechanism underlying synaptic plasticity, studies of intraneuronal signaling pathways are also directly relevant to deciphering the mode of action of psychotropic drugs. Modern psychopharmacology has made tremendous advances in defining the primary site of action of the major classes of psychiatric drugs. For example, benzodiazepines act via GABA receptors, cocaine blocks the reuptake of monoamines, and opioids act via an array of opiate receptor subtypes. However, these important advances have not always provided an adequate explanation of the delayed effects elicited by these agents. For example, the primary site of action of the most widely used class of antidepressant drugs is the serotonin uptake site. However, it is clear that blockade of this uptake pump is not sufficient to elicit an antidepressant effect because this blockade occurs rapidly whereas the therapeutic response is delayed. This paradox has focused attention on the neuronal adaptation to serotonin uptake blockade to look for clues to the basis for the antidepressant response. A leading hypothesis to account for the delay in antidepressant response hinges on the observation that prolonged uptake blockade leads to downregulation of inhibitory presynaptic receptors on serotonin nerve terminals. This adaptation may then allow for increased release of serotonin, with each action potential invading the serotonin nerve terminals. According to this theory, the intracellular signaling pathway regulating the responsiveness of presynaptic receptors is the critical determinant of the antidepressant response.

Another example is provided by recent studies implicating the cAMP system in mediating responses elicited by long-term opiate administration. Many of the acute effects of opiate receptor activation on ion channel function are mediated directly by G_i proteins, independent of cAMP. However, in parallel with these electrophysiological effects, opioid receptor activation also suppresses cAMP concentrations. With long-term administration of opioids, neurons adapt to this persistent suppression of cAMP concentrations by altering the expression of several components of the cAMP system, yielding a net increase in cAMP tone to compensate for the chronic negative influence of opioids. This adaptive shift in the cAMP system brings the system back into balance as long as opioid stimulation persists. However, if opioids are withdrawn, the neuron is left with an overly active cAMP system that may underlie many of the behavioral phenomena of withdrawal. Although it is unclear how long-term suppression of cAMP concentrations leads to a compensatory upregulation of this system, available evidence suggests that CREB, a transcription factor sensitive to cyclic AMP, mediates this response. Evidence supporting this theory has been provided by recent studies demonstrating that opiate withdrawal is attenuated in transgenic animals that are deficient in CREB. These landmark results provide important confirmation that intraneuronal signaling pathways play a central role in this important aspect of opiate action. In addition, these findings exemplify how application of molecular biological approaches to deciphering intraneuronal signaling pathways will provide important insights into the mode of action of currently used psychiatric drugs and pave the way for a new generation of improved treatment approaches.

SUGGESTED CROSS-REFERENCES

The role of intraneuronal signaling pathways in mediating the effects of neurotransmitters on ion channels and gene expression are also discussed in Section 1.3, Section 1.4, Section 1.7, and Section 1.14; the cellular events underlying memory are discussed in Section 3.5.

SECTION REFERENCES

*Bear MF, Malenka RC: Synaptic plasticity: LTP and LTD. Curr Opin Neurobiol *4:*389, 1994.

Bear MF, Abraham WC: Long-term depression in hippocampus. Ann Rev Neurosci *19:*437, 1996.

Berridge MJ: Inositol trisphosphate and calcium signaling. Nature *361:*315, 1993.

Blenis J: Signal transduction via the MAP kinases: Proceed at your own RSK. Proc Natl Acad Sci USA *90:*5889, 1993.

Blier P, De Montigny C: Current advances and trends in the treatment of depression. Trends Pharmacol Sci *15:*220, 1994.

*Bliss TVP, Collinridge GL: A synaptic model of memory: Long-term potentiation in the hippocampus. Nature *361:*31, 1993.

Blitzer RD, Wong T, Nouranifar R, Iyengar R, Landau E: Postsynaptic cAMP pathway gates early LTP in hippocampal CA1 region. Neuron *15:*1403, 1995.

Chen C, Tonegawa S: Molecular genetic analysis of synaptic plasticity, activity-dependent neural development, learning and memory in the mammalian brain. Annu Rev Neurosci *20:*157, 1997.

Dohlman HG, Thorner J: RGS proteins and signaling by heterotrimeric G proteins. J Biol Chem *272:*3871, 1997.

Duman RS, Heninger GR, Nestler EJ: A molecular and cellular theory of depression. Arch Gen Psychiatry *54:*597, 1997.

Exton JH: Regulation of phosphoinositide phospholipases by hormones, neurotransmitters and other agonists linked to G proteins. Ann Rev Pharmacol Toxicol *36:*481, 1996.

Frank DA, Greenberg ME: CREB: A mediator of long-term memory from mollusks to mammals. Cell *79:*5, 1994.

*Ghosh A, Greenberg ME: Calcium signaling in neurons: Molecular mechanisms and cellular consequences. Science *268:*239, 1995.

Greengard P, Valtorta F, Czernik AJ, Benfenati F: Synaptic vesicle phosphoproteins and regulation of synaptic function. Science *259:*780, 1993.

Gudermann T, Kalkbrenner F, Schultz G: Diversity and selectivity of receptor-G protein interaction. Ann Rev Pharmacol Toxicol *36:*429, 1996.

Hedgepeth CM, Conrad LJ, Zhang J, Huang H-C, Lee VMY, Klein PS: Activation of the Wnt signaling pathway: A molecular mechanism for lithium action. Dev Biol *185:*82, 1997.

Huang C, Hepler JR, Gilman AG, Mumby SM: Attenuation of Gi- and Gq-mediated signaling by expression of RGS4 or GAIP in mammalian cells. Proc Natl Acad Sci USA *94:*6159, 1997.

Huganir RL, Greengard P: Regulation of neurotransmitter receptor desensitization by protein phosphorylation. Neuron *5:*555, 1990.

*Hyman SE, Nestler EJ: *Molecular Foundations of Psychiatry*. American Psychiatric Press, Washington, DC, 1993.

Jaffrey SR, Snyder SH: Nitric oxide: A neural messenger. Ann Rev Cell Dev Biol *11:*417, 1995.

Jope RS, Williams MB: Lithium and brain signal transduction systems. Biochem Pharmacol *47:*429, 1994.

Klein PS, Melton DA: A molecular mechanism for the effect of lithium on development. Proc Natl Acad Sci USA *93:*8455, 1996.

Linden DJ: Long-term synaptic depression in the mammalian brain. Neuron *12:*457, 1994.

Maldonado R, Blendy JA, Tzavara E, Gass P, Roques BP, Hanoune J, Schutz G: Reduction of morphine abstinence in mice with a mutation in the gene encoding CREB. Science *273:*657, 1996.

Mansuy IM, Winder DG, Moallem TM, Osman M, Mayford M, Hawkins RD, Kandel ER: Inducible and reversible gene expression with the rtTA system for the study of memory. Neuron *21:*257, 1998.

Mauk MD, Garcia KS, Medina JF, Steele PM: Does cerebellar LTD mediate motor learning? Toward resolution without a smoking gun. Neuron *20:*359, 1998.

Mulkey RM, Endo S, Shenolikar S, Malenka RC: Involvement of a calcineurin/inhibitor-1 phosphatase cascade in long-term depression. Nature *369:*486, 1994.

Nestler EJ: Under siege: The brain on opiates. Neuron *16:*897, 1996.

*Nester EJ, Aghajanian GK: Molecular and cellular basis of addiction. Science *278:*58, 1997.

Nicoll RA: The coupling of neurotransmitter receptors to ion channels in the brain. Science *241:*545, 1988.

Nishizuka Y: Intracellular signaling by hydrolysis of phospholipids and activation of protein kinase C. Science *258:*607, 1992.

Schlessinger J: SH2/SH3 signaling proteins. Curr Opin Genet Dev *4:*25, 1994.

Segal RA, Greenberg ME: Intracellular signaling pathways activated by neurotrophic factors. Ann Rev Neurosci *19:*463, 1996.

Sunahara RK, Dessauer CW, Gilman AG: Complexity and diversity of mammalian adenylyl cyclases. Ann Rev Pharmacol Toxicol *36:*461, 1996.

*Takahashi M, Terwilliger R, Lane C, Mezes PS, Conti M, Duman RS: Chronic antidepressant administration increases the expression of cAMP-specific phosphodiesterase 4A and 4B isoforms. J Neurosci *19:*610, 1999.

*Vanhoutte P, Barnier JV, Guibert B, Pages C, Besson MJ, Hipskind RA, Caboche J: Glutamate induces phosphorylation of Elk-1 and CREB, along with c-fos activation, via an extracellular signal-regulated kinase-dependent pathway in brain slices. Mol Cell Biol *19:*136, 1999.

Yang X, Diehl AM, Wand GS: Ethanol exposure alters the phosphorylation of cyclic-AMP-responsive element-binding protein and cyclic-AMP-responsive element-binding activity in rat cerebellum. J Pharmacol Exper Ther *278:*338, 1996.

Zheng F, Gingrich MB, Traynelis SF, Conn PJ: Tyrosine kinase potentiates NMDA receptor currents by reducing tonic zinc inhibition. Nature Neuroscience *1:*185, 1998.

1.9 BASIC ELECTROPHYSIOLOGY

CHARLES F. ZORUMSKI, M.D., AND KEITH E. ISENBERG, M.D.

Neurons use electrical signals to send and receive information. These electrical signals determine local and network properties of the central nervous system (CNS) and result from the flow of ions across cell membranes through macromolecular pores called *ion channels*. Neurons possess two classes of ion channels, *gated* and *nongated*. Nongated ion channels open spontaneously and contribute to the cellular resting membrane potential. The opening and closing of most ion channels is regulated (gated) by changes in transmembrane voltage or neurochemicals. Certain voltage-gated sodium channels allow very rapid movement of ions and provide the basis for communication within and between neurons. These rapid signals *(action potentials)* are generated at or near the neuronal cell body and are transmitted to nerve terminals with little decrement in amplitude. The high-fidelity propagation of action potentials results from *saltatory conduction*, which is the ability of electrical signals to "jump" rapidly between axonal nodes of Ranvier.

At a nerve terminal an action potential causes a depolarization that opens voltage-gated calcium channels. The influx of calcium promotes the release of a chemical neurotransmitter into the extracellular space where the transmitter is able to influence a receiving cell. Neurotransmitters bind to specific protein receptors and alter neuronal excitability via actions on ion channels. There are two broad classes of neurotransmitter receptors: *ligand-gated ion channels* are directly opened by the binding of a transmitter whereas G-*protein–coupled receptors* influence the function of ion channels indirectly via guanine nucleotide binding proteins (G proteins) or chemical second messengers.

PRINCIPLES OF CELLULAR ELECTROPHYSIOLOGY

Resting Membrane Potential In nerve cells, potassium ions (K^+) are at higher concentration inside the membrane than outside whereas the opposite is true for sodium (Na^+), calcium (Ca^{2+}), and chloride (Cl^-) ions (Fig. 1.9-1). The bulk solutions on either side of the membrane are electrically neutral, with most of the intracellular negative charge being contributed by large organic anions (acids and proteins). The differential distribution of ions across neuronal membranes results in part from the action of membrane pumps that use energy from adenosine triphosphate (ATP) to drive ions against a concentration gradient into or out of the cell. The best characterized pump is the Na^+-K^+ adenosine triphosphatase (ATPase) that transports 3 Na^+ out of and 2 K^+ into the cell during each cycle. Because an unequal amount of charge is moved during each cycle, the pump is electrogenic and produces an electrochemical potential across the membrane that makes the inside of the membrane negative with respect to the outside. Na^+-K^+ ATPase activity is a major contributor to brain energy utilization, with as much as 40 percent of brain oxygen consumption resulting from pump activity required to reestablish ionic homeostasis following action potential firing and synaptic transmission. The cardiac glycosides digoxin (Lanoxin) and ouabain are effective inhibitors of Na^+-K^+ ATPase in the heart and improve myocardial contractility by depolarizing cardiac myocytes and increasing intracellular Ca^{2+}.

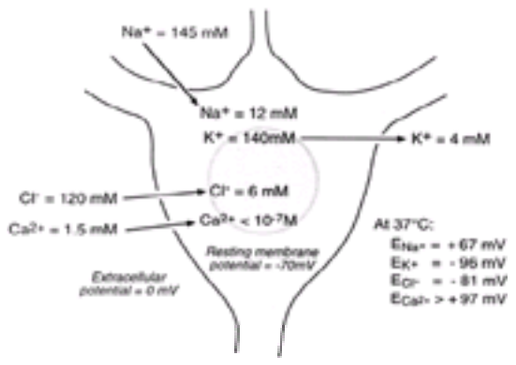

FIGURE 1.9-1 The distribution of Na^+, K^+, Ca^{2+}, and Cl^- across the membrane of a typical neuron. Using these ion concentrations, the equilibrium (Nernst) potentials for these ions at 37°C are shown at the lower right.

At rest, neuronal membranes are permeable to K^+ and Cl^- and to a lesser extent to Na^+, partly because of the flow of ions through nongated leakage channels. K^+ and Cl^- flow down their concentration gradients making the inside of the neuronal membrane negative with respect to the outside. The separation of charge establishes a voltage, called a *(potential difference)* across the membrane. The presence of a transmembrane potential difference creates an electrical gradient for the movement of ions in addition to the concentration gradient established by the differences in ion concentrations inside and outside the cell. By convention, the bulk extracellular solution is at 0 mV, making the resting membrane potential about –70 mV inside the neuron. The bulk extracellular and intracellular solutions are electrically neutral and the charge separation that produces the membrane potential occurs in the immediate vicinity of the membrane. The number of ions needed to change the membrane potential is very small relative to concentrations in the bulk solutions. For example, a potential change of 100 mV across a 1 cm^2 area of membrane requires the movement of only about 10^{-12} moles of a monovalent ion. By comparison, Na^+ and K^+ are present at about 10^{-1} M in the extracellular and intracellular fluids, respectively.

For each ion in solution there is a specific membrane potential at which the opposing forces of the electrical gradient and concentration gradient are balanced. This potential (known as the *Nernst potential* or *equilibrium potential*) can be calculated based on the ion concentrations on either side of the membrane. For K^+, the Nernst potential (designated E_K) is expressed as: $E_K = (RT/zF) \cdot \ln([K]_o/[K]_i)$, where R is the ideal gas constant (8.31 joules/degree/mole), T is the temperature in degrees Kelvin, z is the valence of the ion, F is Faraday's constant (96,500 coulombs/mole, the charge on a mole of monovalent ions), and $[K]_o$ and $[K]_i$ are the concentrations of K^+ outside and inside the cell. At 37° C, the Nernst potential for K^+ is –96 mV, while E_{Na} is +67 mV, E_{Cl} is –81 mV and E_{Ca} is greater than +97 mV. The importance of these equilibrium potentials comes from the fact that when an ion channel that is permeable to a specific ion opens, it drives the membrane potential towards the equilibrium potential for that ion. For example, when K^+-selective ion channels open, the neuronal membrane potential moves toward –96 mV. This makes the inside of the cell more negative, an effect that is termed *hyperpolarization*. Na^+ and Ca^{2+} channel opening has the opposite effect, making the inside of the cell less negative *(depolarization)*. At any time, the membrane potential is a weighted average of the equilibrium potentials of the ions to which the membrane is permeable.

Passive Membrane Properties To understand how ion concentration gradients, electrical gradients, ion channels, and the distribution of charges across the membrane are related, it is helpful to describe the cell membrane as an electrical circuit consisting of resistors (conductors), batteries, and capacitors. Because ions do not directly penetrate the lipid membrane but rather flow through ion channels, the ion channels can be thought of as variable resistors. Physiologists describe ion channels in terms of their ion selectivity (which ions flow through the channel) and their conductance (relative ease of passing ions). Conductance (g) is the inverse of resistance (R) in an electrical circuit (g = 1/R). The presence of a voltage across the membrane provides an electrical driving force for the flow of ions through ion channels resulting in a transmembrane current. The relationship among voltage (V), ionic current (I), and resistance (conductance) is given by the physiologists version of Ohm's law: $I_{ionic} = g \cdot (V_m – E_{rev})$ where V_m is the membrane potential, E_{rev} is the Nernst potential for the ions flowing through the channel, and $(V_m – E_{rev})$ represents the driving force for ion flow.

Another important passive electrical property is capacitance. A capacitor is an electrical device consisting of two conductors separated by an insulating material that is

capable of storing charges of opposite sign on the two conductors. In the case of neurons, the conductors are the extracellular and intracellular fluids while the lipid membrane is the insulator. Whenever current flows through the membrane, some current must flow to charge the membrane capacitance (C_m). The expression describing this capacitive current is: $I_{cap} = C_m \cdot (dV/dt)$. Note that capacitive current flows only when the membrane potential is changing (i.e., there is some change in voltage [dV] as a function of time [dt]. The total current flowing across a membrane at any given time is a sum of I_{cap} and I_{ionic}. One of the major tools used by physiologists to study ionic currents is a *voltage clamp* (or more recently a patch clamp). These techniques employ electrical devices to keep the membrane potential constant and eliminate the contribution of capacitive currents during physiological studies, thus making it possible to measure ionic currents directly.

One way to view the operation of an ion channel is as a battery (voltage source) in series with a conductor (resistor). The different types of ion channels can be viewed as being in parallel with each other and with the membrane capacitance. The net result is that the neuronal membrane can be represented by an equivalent electrical circuit (Fig. 1.9-2), which can be used to describe how current flows when ions enter and exit the cell in response to various stimuli.

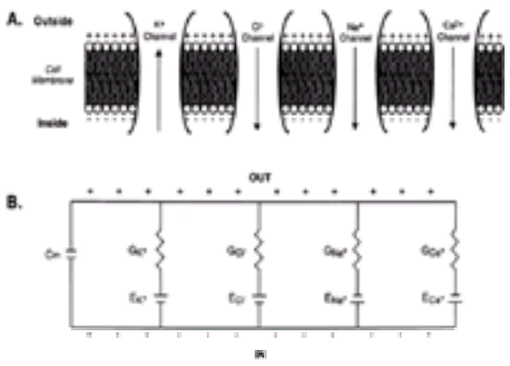

FIGURE 1.9-2 Ion channels form proteinaceous pores that traverse the lipid bilayer of the cell membrane. Because of the action of membrane pumps, the extracellular surface of the membrane has a net positive charge with respect to the intracellular surface. As a result of the transmembrane potential and the presence of ion channels, the neuronal membrane can be depicted as an equivalent electrical circuit in which each ion channel is a resistor (conductor, G_x) in series with a battery (E_x). Different ion channels are shown parallel with each other and parallel with the membrane capacitance (C_m).

Active Membrane Properties: Action Potentials Changes in membrane potential have important effects on excitability because certain ion channels are activated (gated) by voltage changes. When neurons are depolarized with respect to the resting potential, specific Na^+ channels open rapidly and drive the membrane potential towards the Na^+ equilibrium potential (+66 mV). Because of the leakage channels that are open at rest, there is initially a balance between the leakage currents and the currents flowing through Na^+ channels that are opened by depolarization. However, at a certain membrane potential the current flowing through Na^+ channels exceeds the current through the leakage channels. The membrane potential at which Na^+ currents exceed the leakage currents is called the *threshold potential*. Importantly, at potentials that are depolarized with respect to threshold the entry of more Na^+ into the neuron produces further depolarization, which in turn opens more Na^+ channels in a regenerative fashion. During this process the neuronal membrane potential depolarizes to potentials >0 mV but never reaches the Na^+ equilibrium potential for two reasons. First, during the depolarization, Na^+ channels not only activate but they also rapidly inactivate. *Inactivation* refers to a process by which voltage-gated ion channels enter a nonconducting state despite the continued presence of the activating stimulus (depolarization). Second, the depolarization produced by Na^+ entry also opens voltage-gated K^+ channels, which drive the membrane potential towards the K^+ equilibrium potential (–96 mV). The net effect of the activation and inactivation of Na^+ channels and the delayed opening of K^+ channels is that the neuronal membrane potential rapidly changes to values >0 mV and then returns rapidly to the resting membrane potential. This rapid sequence occurs over several milliseconds and is referred to as an *action potential* (or *spike*) (Fig. 1.9-3). The fact that the membrane potential transiently exceeds 0 mV is called an *overshoot*. Action potentials represent all-or-none increases in electrical excitability and are important contributors to information transfer within and between neurons, allowing the neuronal cell body to communicate rapidly with its terminals, and in the terminals providing the depolarization that promotes the Ca^{2+}-dependent release of neurotransmitters.

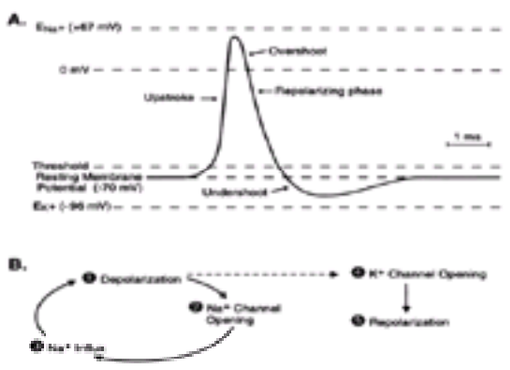

FIGURE 1.9-3 A. A neuronal action potential as recorded by an intracellular microelectrode. The portions of the action potential are described in the text. **B**. The sequence of events underlying the action potential.

In most neurons the K^+ equilibrium potential is negative with respect to the resting membrane potential. Thus, the action potential is often followed by a transient afterhyperpolarization (or undershoot) that decays back to the resting potential as the voltage-sensitive K^+ channels responsible for action potential repolarization close (Fig. 1.9-3). Following an action potential, there is a time during which stimulation either cannot elicit an action potential or during which it takes a very strong stimulus to evoke an action potential. These are called the absolute and relative refractory periods, respectively. The *absolute refractory period* results from the increased K^+ conductance that repolarizes the action potential and produces the undershoot; the *relative refractory period* reflects the time it takes for Na^+ channels to recover from inactivation.

Action Potential Conduction in Axons Action potentials are typically generated in the neuronal cell body or in the initial segment of the axon (also called the *axon hillock*) where Na^+ channels are densely collected. Because action potentials are generated at a distance from the nerve terminals where neurotransmitters are released, an important question concerns how action potentials are transmitted to the synaptic terminals. In a strictly passive nerve fiber, leakage of current across the membrane results in *decremental conduction* with the signal fading over a distance that is determined by the longitudinal (axial) resistance of the fiber, the membrane capacitance, and the transmembrane resistance. Decremental conduction is more typical of the spread of electrical signals along dendrites back to the neuronal cell body, although recent studies have shown that dendrites also have voltage-gated ion channels that play important roles in modifying electrical synaptic inputs to the dendrites.

Many axons are encased in myelin sheaths that allow them to send action potentials over longer distances. As a result of myelination, axons are electrically insulated except at Ranvier's nodes where there are collections of voltage-gated Na^+ channels involved in action potential generation (Fig. 1.9-4). The myelin sheath greatly increases the transmembrane resistance and diminishes current leakage from the axon, making it easier for current to flow down the length of the axon. Once generated, action potentials propagate rapidly and the wave of depolarization jumps from node to node in a form that transmits the signal faithfully to the nerve terminals. This process of action potential spread through axons is referred to as *saltatory conduction* (derived from the Latin word *saltare* meaning “to jump”) and is important because of the speed and fidelity with which electrical information is passed from a nerve cell body to its terminals. The importance of saltatory conduction can be readily appreciated when considering the distances over which impulses must travel from the CNS to cause movement in the toes. In several human illnesses, including multiple sclerosis and Guillain-Barré syndrome, demyelination of axons produces changes in axon conduction and specific neurological defects.

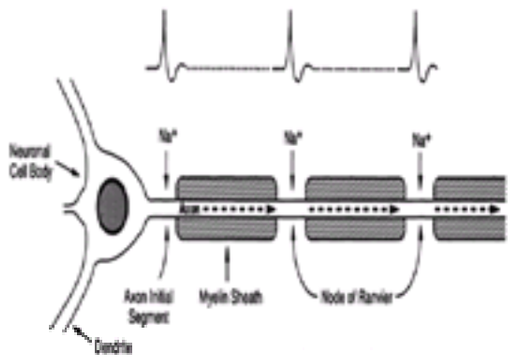

FIGURE 1.9-4 An example of saltatory conduction of an action potential in a neuron with a myelinated axon. The action potential is generated in the initial segment of the axon. As the signal moves along the axon, current tends to leak from the cell diminishing the amplitude. However, myelin insulates the axon and markedly diminishes current leakage out of the axon, thus enhancing flow down the axon to the first node of Ranvier. At the node of Ranvier, Na^+ channels of the type involved in action potentials open in response to the wave of depolarization and reproduce the all-or-none action potential. The sequence is repeated at subsequent nodes of Ranvier until the action potential reaches the nerve terminal.

ION CHANNELS

Structure and Function of Voltage-Gated Ion Channels Voltage-gated ion channels allow the flow of ions in response to changes in membrane voltage and are key elements in neuronal excitation and inhibition. Although ion channels can usually pass more than a single type of ion, voltage-gated channels are named according to the predominant ion that flows when the channel is open. Ion channels that are selective for Na^+, K^+, Ca^{2+}, or Cl^- have been described in neuronal membranes. Certain ion channels that are gated directly by chemical neurotransmitters such as glutamate and acetylcholine are selective for Na^+, K^+, and Ca^{2+} but exclude Cl^- and are called *nonselective cationic channels*.

Sodium (Na^+) Channels Na^+ channels are primarily responsible for the fast upstroke of action potentials, although in some neurons Na^+ channels also contribute to lower-level depolarizations and pacemaker firing. Pacemaker activity refers to the ability of certain neurons to depolarize spontaneously and to drive activity in a system of connected cells. Na^+ channels *activate* (open) rapidly in response to depolarization and also *inactivate* rapidly and nearly completely in response to prolonged depolarizations.

Cloning studies have provided important information about the structure of Na^+ channels. Na^+ channels cloned from rat brain contain three protein subunits—a main (or a) subunit with a molecular weight of 240 to 280 kd and two minor subunits with molecular weights of 30 to 40 kd (termed b *1* and b*2*) that appear to assemble in a 1:1:1 ratio. The a-subunit is a glycoprotein consisting of four structurally similar (*homologous*) domains that each have six proposed membrane spanning regions, referred to as S1 through S6 (Fig. 1.9-5). The properties of voltage dependence, ion permeation, activation, and inactivation are conferred by specific regions of the Na^+ channel protein. However, the exact manner in which the proteins assemble in the lipid membrane remains a matter of active study.

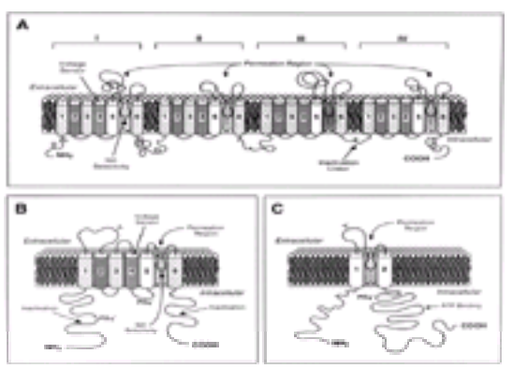

FIGURE 1.9-5 A. The diagram shows the proposed secondary structure of voltage-gated Na^+, Ca^{2+}, and K^+ channels based on analysis of the primary amino acid sequences. Na^+ and Ca^{2+} channels consist of four homologous domains (I,II,III, and IV), each of which has six membrane spanning regions (S1-S6). Both the amino (NH_2) and carboxy (COOH) terminals are located intracellularly. A stretch of amino acids between S5 and S6, called the p-loop, is believed to form two antiparallel b-sheets that line the channel pore. Positive charges in the S4 region are believed to comprise the voltage sensor. **B**. Some K^+ channels also have 6 transmembrane regions but lack the four homologous repeats seen in Na^+ and Ca^{2+} channels. **C**. Inwardly rectifying K^+ channels, including K-ATP channels, have a structure that differs from the scheme described above. These channels have two membrane spanning regions and a pore-lining p-loop. In contrast to channels gated by extracellular ATP (Fig. 1.9-7C), the bulk of the protein is located intracellularly.

Relations between primary protein structure and ion channel function in Na^+ channels have been examined using mutations of specific amino acid residues. It appears that both the amino- and carboxy-terminals of the a-subunits are located intracellularly. The fourth membrane spanning region (S4) plays a key role in sensing the transmembrane voltage changes that allow channel gating. Between the S5 and S6 membrane spanning regions there is a segment of hydrophobic amino acids that does not completely cross the lipid membrane bilayer. This reentrant loop of amino acids (called a *p loop*) is a feature shared by other voltage-gated ion channels and appears to form the lining of the ion channel pore.

Na^+ channels contain several sites at which neurotoxins and drugs act to influence excitability. Most but not all Na^+ channels contain an extracellular site at which tetrodotoxin (TTX) and saxitoxin (STX) act to block ion flow. TTX is a neurotoxin isolated from puffer fish that is used experimentally to block Na^+ channel function. At a site on Na^+ channels that is distinct from the TTX site, certain scorpion and sea anemone toxins act to modify gating properties. The a-scorpion toxin slows inactivation of Na^+ channels while b-scorpion toxins shift the voltage of activation and allow channels to open at voltages closer to the resting membrane potential. The net effect of the scorpion toxins is to enhance excitation, contributing to the increased firing in pain fibers and paralysis (tetany) that are associated with a scorpion sting. Mutations in the a-subunit of skeletal muscle Na^+ channels cause the human disorder hyperkalemic periodic paralysis. Like the anemone and a-scorpion toxins, these mutations slow channel inactivation.

Other toxins isolated from the buttercup family (aconitine), the lily family (veratridine), and frogs that are used for arrow poisons in South America (batrachotoxin) promote the direct opening of Na^+ channels and prolong the duration that the channels stay open. The net effect is similar to the scorpion toxins. Finally, certain local anesthetic drugs, including lidocaine and procaine, block Na^+ channels by binding reversibly to sites within the hydrophobic regions of the ion channel. The blockade of Na^+ channels is likely to contribute to local anesthetic effects as well as to the antiarrhythmic effects of these drugs in the heart. The rich pharmacology of voltage-gated Na^+ channels provides a means for understanding how certain neurotoxins act as well as a means by which Na^+ channel function can be manipulated therapeutically. It is important to emphasize that not all Na^+ channels in neurons are sensitive to all these agents. It is clear that TTX-insensitive Na^+ channels exist in a variety of excitable cells, although their function is not well understood at present.

Potassium (K^+) Channels K^+ channels represent the most diverse family of voltage-gated ion channels in excitable cells and are important participants in determining the firing properties of neurons. For example, the fast repolarization of neurons produced by certain K^+ channels allows an increased rate of action potential firing, which can then be used in frequency-dependent information coding. Most neurons express multiple types of K^+ channels that differ in their activation and inactivation kinetics, voltage dependence, and pharmacology. Because the equilibrium potential for K^+ is about –90 mV in most neurons, the opening of K^+ channels allows K^+ to flow out of the cell, resulting in membrane hyperpolarization and a decrease in excitability. The first K^+ channel to be identified was called a

delayed rectifier. These channels derive their name from the experiments of Hodgkin and Huxley on squid giant axons and are so named because the currents gated by these channels activate more slowly than the Na^+ channels that produce the upstroke of the action potential (i.e., the K^+ channel opening is delayed). A rectifier (or diode) is an electrical device that passes current better in one direction than another. The K^+ current is described as a rectifier because the channel is more effective in allowing K^+ ions to exit than to enter the cell.

Delayed-rectifier channels open slowly and show little inactivation during prolonged depolarizations. It appears that these channels help to determine the frequency with which neurons fire action potentials. Early experiments in squid giant axons indicated that delayed rectifier currents were the primary K^+ currents involved in action potential repolarization. However, in neurons the situation is more complex, with several more rapidly activating K^+ channels contributing significantly. These include two classes of calcium-activated K^+ channels that are opened by increases in intracellular Ca^{2+} and are called SK *small conductance Ca^{2+}-activated K^+* (*SK*) *channels* and *big conductance* (*BK*) *channels*, based on the relative ease (conductance) with which they pass K^+ when open. SK channels activate with relatively small intracellular Ca^{2+} increases to levels on the order of 100 nM whereas BK channels require Ca^{2+} levels to rise to 1 to 10 µM. In neurons, intracellular Ca^{2+} buffering mechanisms usually keep free Ca^{2+} concentrations <100 nM at rest. SK and BK channels are also distinguished by distinctive voltage-dependence and pharmacology. The bee toxin apamin blocks SK channels and the scorpion toxin charybdotoxin blocks BK channels. BK channels contribute to action potential repolarization in neurons, whereas SK channels produce a slow afterhyperpolarization responsible for the process of *accommodation* (adaptation) that diminishes repetitive action potential firing during prolonged depolarizations (Fig. 1.9-6).

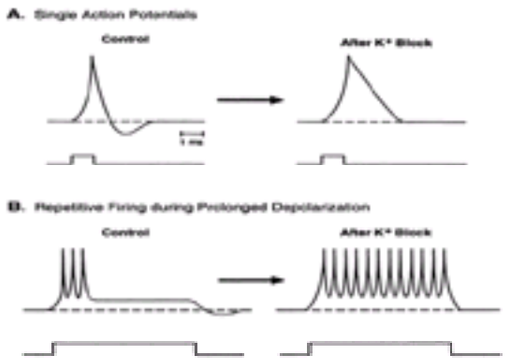

FIGURE 1.9-6 A. The traces show the effect of inhibiting K^+ channels involved in action potential repolarization. Following K^+ channel block, action potentials are broadened and show a diminished undershoot. **B**. During prolonged depolarizations, Ca^{2+}-activated K^+ conductances diminish repetitive firing by the process of accommodation (or spike adaptation). When these K^+ conductances are inhibited, prolonged depolarization results in repetitive action potential firing for the duration of the depolarizing stimulus.

A-type K^+ channels rapidly activate with depolarizations to potentials greater that –60 mV and rapidly inactivate at depolarized potentials; hence A channels are also referred to as transient (inactivating) K^+ channels. A channels are involved in setting the interspike frequency with which neurons can fire and contribute to action potential repolarization. D-type K^+ channels are inhibited by a-dendrotoxin and also appear to help action potential repolarization in neurons. These channels show slower activation and less inactivation than A-type channels.

M channels represent a class of K^+ channels that are activated in a time- and voltage-dependent fashion but are blocked by the neurotransmitter, acetylcholine, acting at muscarinic receptors. These channels contribute to action potential repolarization and help to slow repetitive firing. In the sea snail, *Aplysia californica*, S channels are K^+ channels that contribute to action potential repolarization are inhibited by the neurotransmitter serotonin. Importantly, the activity of these S channels is diminished during acute behavioral sensitization of the gill-withdrawal reflex in *Aplysia*, and studies of the role of these channels in synaptic function have provided important insights into the cellular basis of certain forms of learning and memory.

Some ion channels are opened by hyperpolarization instead of depolarization. These include anomalous (inward) rectifier and H channels that allow K^+ to enter rather than exit the cell. The name “anomalous rectifier” (also called “inward rectifier”) indicates that, in contrast to the delayer rectifier, this channel passes K^+ much better in an inward than in an outward direction. The anomalous rectifier and H channels (hyperpolarization activated channels) are believed to contribute to the neuronal resting membrane potential and to pacemaker firing in certain neurons. H channels are permeable to both K^+ and Na^+ and show little inactivation. Thus, H channels gate a persisting current at membrane potentials near rest. The Na^+ current contribution from H channels may contribute to the fact that the resting membrane potential in most neurons is depolarized with respect to the K^+ equilibrium potential.

In peripheral tissues and in some neurons a class of K^+ channels is regulated by intracellular ATP (K_{ATP}). In the pancreas these K_{ATP} channels are important because they are involved in controlling the release of insulin and are a site of action of the hypoglycemic sulfonylurea drugs, such as tolbutamide (Tolbutamide), that are used to treat patients with diabetes mellitus. The hypoglycemic drugs promote the release of insulin by blocking the ATP-sensitive K^+ channels, which in turn leads to membrane depolarization, calcium influx, and release of the hormone. Diazoxide (Hyperstat), an antihypertensive drug that has the adverse effect of increasing blood glucose concentrations and has the opposite effect on pancreatic (K_{ATP}) channels, opening the channels and diminishing the release of insulin.

K_{ATP} channels exist in the CNS and appear to be involved in regulating the release of certain neurotransmitters and perhaps in determining the response of some neurons to changes in intracellular energy levels. Although the role of K_{ATP} channels in regulating transmitter release in neurons is not well understood, the effects of the hypoglycemic drugs and diazoxide on the release of insulin from pancreatic islet cells raise instructive points concerning how K^+ channels help to control the release of transmitters and hormones. During action potentials, voltage-gated Ca^{2+} channels open and the resulting increases in intracellular Ca^{2+} promote transmitter release. When K^+ channels that normally contribute to the repolarization of action potentials are blocked, either by neurochemicals or drugs, the action potential duration is prolonged and more Ca^{2+} enters the cell, leading to enhanced transmitter release.

In the late 1980s several groups used the *Shaker* behavioral mutants in the fruit fly, *Drosophila melanogaster*, to clone specific K^+ channels. *Shaker* is a mutant fly that develops hyperexcitability (shaking) following exposure to specific anesthetics. The enhanced excitability results from abnormal function of A-type K^+ channels. Cloning studies of K^+ channels from *Shaker* indicate that some K^+ channels, like Na^+ channels, are proteins that have six putative membrane spanning regions (S1 through S6), a reentrant p loop between S5 and S6 that lines the ion channel pore, and a voltage sensor in the S4 region (Fig. 1.9-5). These K^+ channel proteins are only about one fourth the size of the Na^+ channel a-subunit and have only one of the homologous internal repeats seen in Na^+ channels. It also appears that there are multiple distinct subfamilies of K^+ channels based on genetic studies in *Drosophila*, where the most detailed analysis has been done. Presently, four subfamilies of K^+ channels have been identified (Kv1 through Kv4), each of which has multiple subtypes. Different K^+ channels within a subfamily and perhaps across subfamilies may combine to produce channels with diverse functions, but it is presently unclear how subunits assemble to produce functional K^+ channels. Best estimates suggest a structure with four pore-forming (a) subunits that are analogous to the four homologous repeats of Na^+ channels. K^+ channels also contain cytoplasmic (b) subunits that appear to be important in channel inactivation and in controlling cell surface expression of the a-subunits.

An important principle in K^+ channel physiology is that certain neurotransmitters alter the function of these channels. Acetylcholine, acting at muscarinic receptors, blocks several K^+ currents leading to enhanced neuronal excitability. In the hippocampus and other CNS regions, the neurotransmitters g-aminobutyric acid (GABA), serotonin, and adenosine open the same class of inwardly rectifying K^+ channels by activating a G protein. Similarly, acetylcholine activates inwardly rectifying K^+ channels in a variety of tissues including the heart and CNS. These G–protein–regulated, inwardly rectifying K^+ channels (GIRKs) allow divergent synaptic inputs to a single neuron to exert regulatory influences over neuronal firing through a single class of ion channels. The inwardly rectifying K^+ channels, including K_{ATP} channels, differ structurally from the Kv (*Shaker*) family described previously in that they have only two membrane spanning regions and a pore-forming loop (see Fig. 1.9-5); GIRKs also lack the voltage sensor found in the Kv family. K_{ATP} channels appear to be heteromultimers that contain an inwardly rectifying K^+ channel and a large subunit that binds ATP and sulfonylurens.

Calcium (Ca^{2+}) Channels Because Ca^{2+} is involved in numerous cellular events including enzyme activation, gene expression, and neurotransmitter release, the

regulation of intracellular Ca^{2+} levels is of major importance to neurons. Furthermore, excessive and prolonged increases in intracellular Ca^{2+} concentrations contribute to neuronal death in acute and chronic human neurodegenerative conditions. Voltage-activated Ca^{2+} channels provide a major source of the Ca^{2+} signal that is used to activate cellular processes, and neurons possess multiple classes of voltage-gated Ca^{2+} channels.

There are several ways to classify voltage-gated Ca^{2+} channels based on biophysical and pharmacological properties. Some Ca^{2+} channels are activated by relatively small depolarizations over the range of –80 to –50 mV and are called *low voltage-activated* (*LVA*) *Ca^{2+} channels*. These LVA channels inactivate rapidly and are relatively insensitive to dihydropyridine calcium channel inhibitors, such as nifedipine (Adalat) and nimodipine (Nimotop). LVA channels are also called *T-type Ca^{2+} channels* because of their transient (inactivating) currents. LVA Ca^{2+} channels contribute to burst firing and oscillatory activity in neurons because they are activated at membrane potentials near rest. Oscillatory neuronal firing may be important in driving coordinated movements and in maintaining wakefulness. LVA channels also appear to be present in neuronal dendrites and may contribute to synaptic integration and synaptic plasticity. These channels do not appear to participate in transmitter release.

A second group of Ca^{2+} channels is activated by larger depolarizations to membrane potentials above –50 mV and are called *high voltage-activated* (*HVA*) *Ca^{2+} channels*. In many neurons, even when Na^{+} channels that are typically involved in the upstroke of action potentials are blocked, HVA Ca^{2+} channels can produce regenerative spikes. These Ca^{2+} spikes are typically slower in onset and longer in duration than Na^{+} spikes, reflecting the kinetics of HVA channels. HVA Ca^{2+} channels are heterogeneous and several classes contribute to HVA Ca^{2+} currents. *L-type Ca^{2+} channels* (named for their "*long*-lasting" responses) show slow inactivation during sustained depolarizations and are sensitive to blockade by dihydropyridines. L-type Ca^{2+} channels provide sufficient Ca^{2+} influx during action potentials to activate Ca^{2+}-dependent second-messenger systems. *N-type Ca^{2+} channels* (named historically because they were neither L nor T type) are also HVA channels that appear to be involved in providing the Ca^{2+} signal necessary for the release of neurotransmitters from presynaptic terminals. N-type channels are blocked by w-conotoxin GVIA, a poison derived from the snail, *Conus geographicus. P-type Ca^{2+} channels* represent a third class of HVA channels and are so named because of their presence in Purkinje cells of the cerebellum. P-type channels are also found in pyramidal neurons of the hippocampus and cortex, and are insensitive to dihydropyridines and w-conotoxin GVIA but are blocked by a toxin from the funnel web spider, *Agelenopsis aperta*, that is designated w-*Aga-IVA*. P-type channels, like N-type channels, participate in the release of neurotransmitters at specific sites in the CNS. Recent evidence indicates that there are other classes of HVA Ca^{2+} channels that contribute to CNS function (designated *Q-type* and *R-type*), but their functions are not well understood at present. Q-type channels are potently blocked by w-conotoxin-MVIIC isolated from *Conus magus* and appear to participate in transmitter release. R-type channels are *r*esistant to the Ca^{2+} channels antagonists described above.

Most structural information about Ca^{2+} channels comes from skeletal muscle HVA Ca^{2+} channels. These channels consist of five distinct subunits that are termed a1 (165–195 kDa), a2/d (~170 kDa), b (50–60 kDa), and g (25–35 kDa). The a1-subunits show considerable sequence homology (~30 percent) to voltage-gated Na^{+} channels and form the ion channel pore. A recurring theme in the a1-subunits is the existence of four homologous internal repeats that each contain six putative membrane spanning regions and a p-loop. HVA a1 Ca^{2+} channel subunits contain the dihydropyridine binding site and show structural heterogeneity, with current evidence suggesting the existence of several different a1 subunits (designated $a1_{A-E}$). Additionally, an a, subunit for LVA Ca^{2+} channels has been cloned from rat brain (designated a1G). Point mutations in the a1-subunit of skeletal muscle T tubule Ca^{2+} channels appear to cause the human disorder hypokalemic periodic paralysis. The causative mutations occur in the S4 region of the channel involved in voltage sensing.

In some regions of the CNS, particularly retinal photoreceptors and olfactory epithelial cells, intracellular cyclic nucleotides (e.g., cyclic AMP and cyclic GMP) gate specific classes of ion channels. These cyclic-nucleotide gated (CNG) channels have structural features that are reminiscent of voltage-gated calcium channels, including the presence of six membrane spanning regions and a p-loop that lines the ion channel. CNG channels are nonselectively permeable to cations, but like voltage-gated calcium channels, the flow of monovalent cations through CNG channels is blocked by calcium.

Chloride (Cl^{-}) Channels In most neurons, Cl^{-} is present at higher concentrations extracellularly than intracellularly and the equilibrium potential fo r Cl^{-} is near the cell resting membrane potential. Thus, the opening of Cl^{-} channels tends to keep the neuronal membrane potential near rest, and in conjunction with K^{+} channels, serves as a mechanism to dampen neuronal excitability. Cl^{-} channels contribute significantly to the resting membrane potential in certain neurons and muscle cells. These channels are spontaneously open at resting membrane potentials and exhibit weak voltage and time-dependence. In certain muscle fibers the background Cl^{-} conductance is the largest resting conductance and the distribution of Cl^{-} is near equilibrium. In the human illness myotonia congenita, an abnormality of muscle Cl^{-} channels leads to increased muscular excitability and symptoms of fatigue and cramping during exercise. The rat muscle Cl^{-} channel involved in myotonia is a single polypeptide with 12 putative membrane spanning regions.

Some Cl^{-} channels are activated by increases in intracellular Ca^{2+} and others are activated by membrane depolarization. The Ca^{2+}-activated Cl^{-} channels may help to determine the interspike frequency with which neurons can fire. In addition to their roles in neuron excitability, Cl^{-} channels serve important functions in secretory cells providing the major source of Cl^{-} in tears, sweat, and digestive juices. A defect in secretory Cl^{-} channels that renders the channels insensitive to normal activating stimuli appears to be a major factor in the pathophysiology of cystic fibrosis. The cystic fibrosis transmembrane conductance regulator (CFTR) has a structure that is distinctly different from the muscle Cl^{-} channels involved in myotonia but is similar to ATPase-linked transporters with 12 transmembrane regions.

NEUROTRANSMITTERS AND ION CHANNELS

Classes of Neurotransmitters Much of the information transfer between neurons in the CNS occurs via chemical synapses. These synapses use a variety of messengers (neurotransmitters) that are released in a Ca^{2+}-dependent fashion from presynaptic terminals and act on specific protein receptors to produce biochemical and excitability changes in the receiving cell. There are two primary groups of neurotransmitters—low–molecular-weight amines and neuroactive peptides. These agents act on two classes of receptors, ligand-gated ion channels, at which the binding of the transmitter directly opens ion channels in the membrane, and G protein coupled receptors. The activated G protein then acts on ion channels or alters biochemical second-messenger systems. Physiologists classify synaptic transmission according to the speed of transmission (fast or slow) and according to the nature of the response (excitatory or inhibitory). Fast synaptic transmission occurs on a time scale of up to several hundred milliseconds and is mediated primarily by amine neurotransmitters acting at ligand-gated ion channels. Slow synaptic communication occurs on the scale of seconds to minutes or longer through the actions of either amines or peptides acting on G protein coupled receptors. Different ion channels determine whether transmitter effects are *excitatory* (depolarizing) or *inhibitory* (hyperpolarizing). Moreover, an excitatory synaptic input can exert an inhibitory influence on the firing characteristics of a region. For example, the release of an excitatory neurotransmitter onto an inhibitory interneuron can result in the inhibitory neuron diminishing the activity of a population of cells. Conversely, inhibition of inhibitory interneurons can enhance regional excitability. This provides a great deal of flexibility in controlling and fine-tuning the inputs and outputs of a region.

Currently, there are nine low–molecular-weight amines that serve as neurotransmitters. These include glutamate, the major fast excitatory transmitter in the mammalian CNS, acetylcholine, the excitatory transmitter at the vertebrate neuromuscular junction, g-aminobutyric acid (GABA), and glycine, the major fast inhibitory transmitters in the brain and spinal cord, respectively, and the biogenic amines, dopamine, norepinephrine, epinephrine, serotonin, and histamine. It also appears that the purines, adenosine and adenosine triphosphate (ATP), act as transmitters in some regions. A large number of neuroactive peptides alter neuronal excitability. However, it is uncertain whether all of these substances function as neurotransmitters. Many of these peptides, including vasopressin and cholecystokinin, were first identified as hormones in the vasculature and gut. ATP and certain neuroactive peptides coexist with amine neurotransmitters in some nerve terminals, and there is evidence for co-release of these agents at particular synapses. These observations suggest that interactions between classes of neurotransmitters may be important in determining the ultimate effects of a presynaptic neuron on its postsynaptic target.

Conductance Mechanisms Underlying Neurotransmitter Actions Neurotransmitter actions are often described in terms of effects on membrane conductances. The transmitters that act at ligand-gated ion channels increase the conductance of the cell membrane to specific ions. Excitatory transmitters, such as acetylcholine and glutamate, directly activate nonselective cationic channels, increasing the conductance to Na^{+}, K^{+}, and in some cases Ca^{2+}, while the inhibitory transmitters, GABA and glycine, increase the conductance to $C1^{-}$. A second group of transmitters increases membrane conductance, but does so indirectly through a G protein. For example, GABA, serotonin, and adenosine promote G-protein–mediated opening of inwardly rectifying K^{+} channels in a variety of neurons. A third set of transmitter actions results from indirect effects on voltage-gated or leakage ion channels. These transmitters typically decrease membrane conductance by activating chemical second messenger systems via G-protein coupled receptors. Certain voltage-gated K^{+} and Ca^{2+} channels are specific targets of this inhibition, resulting in excitation or inhibition, respectively. Most transmitters that act on G protein coupled receptors exert at least some of their effects by these decreased conductance mechanisms. The electrical principles underlying synaptic excitation or inhibition are identical to those described for voltage-gated ion channels and are based on the relative permeabilities of the ion channels and the Nernst potentials of the ions involved.

Several transmitters (e.g., GABA, glutamate, acetylcholine, serotonin) act at both ligand-gated ion channels and G-protein coupled receptors. This raises the point that receptors for almost all neurotransmitters, and consequently the effects of these transmitters, are heterogeneous, with the nature of the transmitter effect depending on the specific receptor to which the transmitter binds. Molecular cloning studies have demonstrated that receptors for most neurotransmitters are structurally complex, with multiple receptor subtypes being the rule rather than the exception. At the receptor level there is tremendous flexibility in determining the effects of a given neurotransmitter on a single neuron or on a set of neurons in a CNS region.

Structure of Neurotransmitter Receptors Considerable information now exists about the primary structure of neurotransmitter receptors. Most transmitter-gated ion channels are multimeric proteins consisting of several (usually 5) subunits that each have multiple (2 to 5) membrane spanning regions (Fig. 1.9-7). The functional receptor typically has a large amino-terminal region that extends into the aqueous extracellular environment. In this extracellular region are sites at which neurotransmitters bind and at which sugar molecules are attached to the receptor (glycosylation sites). The function of receptor glycosylation is poorly understood, but presumably plays a role in determining optimal conformations for channel gating. The intracellular regions of the receptor often contain sites at which a phosphate group can be attached. Phosphorylation represents an important mechanism by which second messenger systems modulate the function of receptors and ion channels and is likely to be involved in certain forms of short-term learning and memory.

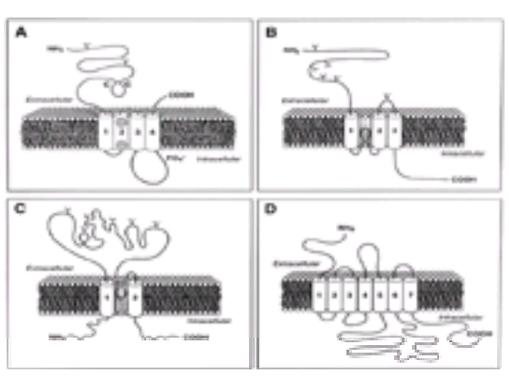

FIGURE 1.9-7 The diagrams show the proposed secondary structure of receptors for several neurotransmitters, including a GABA, receptor **(A)**, an ion-channel-linked glutamate receptor **(B)**, a channel gated by extracellular ATP **(C)**, and a G-protein coupled receptor **(D)**.

The first transmitter-gated channel to be cloned was the muscle-type nicotinic acetylcholine receptor. To date, five neuromuscular nicotinic receptor subunits have been identified (a, b, g, s, and e). Each of these subunits has four membrane spanning regions and a pair of cysteine residues located 15 amino acids apart in the extracellular region of the protein (Fig. 1.9-7). These cysteine residues form a disulfide bridged loop that may contribute to transmitter binding. The nicotinic ion channel is a nonselective cation channel that is permeable to Na^+, K^+, and Ca^{2+}. The membrane spanning regions of the subunits form the ion channel with the second transmembrane region forming the lining of the channel pore. The nicotinic receptor subunits assemble to form a pentamer with the stoichiometry of a $_2$bg and d or e depending on the age of the animal. Subsequent studies found that muscle-type nicotinic receptors are part of a superfamily that includes neuronal nicotinic, GABA type A ($GABA_A$), glycine, and serotonin (5-hydroxytintamine [5-HT]) type 3 ($5\text{-}HT_3$) receptors. Interestingly, $GABA_A$, and glycine receptors are anion selective, passing primarily Cl^- in physiological solutions, whereas nicotinic and $5\text{-}HT_3$ receptors are cation selective. Differences in charges on amino acids at the entrance to the ion channel pore determine whether the channel passes cations or anions. The presence of the paired cysteine residues in the extracellular domain and four membrane spanning regions is characteristic of this class of receptors, which is sometimes referred to as the *cysteine-loop superfamily of receptors*.

The ligand-gated ion channels gated by extracellular ATP (called *P2X receptors*) are exceptions to the scheme described above and have structures more typical of the inwardly rectifying K^+ channels (Fig. 1.9-7). ATP receptors have two membrane spanning regions and a pore forming region (p loop) that are connected by a large loop of extracellularly located amino acids. A major difference between the P2X receptors and the inwardly rectifying K^+ channel is that the bulk of the P2X receptor is extracellular whereas the majority of the K^+ channel is intracellular. P2X channels are cation selective and have a relatively large permeability to calcium. These receptors appear to play a role in fast excitatory synaptic transmission in certain regions of the CNS, including the thalamus. Native ATP receptors may consist of combinations of P2X subunits.

Ion channel-linked glutamate receptors also appear to be exceptions to the structural scheme proposed for $GABA_A$, and nicotinic receptors. The glutamate receptor subunits have three membrane spanning regions and a reentrant p loop that does not completely span the membrane (Fig. 1.9-7). The p loops in the glutamate-gated and ATP-gated channels are similar to those found in voltage-gated ion channels.

G-protein coupled receptors have a distinctly different structure from the ligand-gated ion channels. These receptors are typically single subunit proteins with seven membrane spanning regions (Fig. 1.9-7). Transmitter binding is believed to occur in a pocket formed by the intramembranous portions of the receptor. The coupling of the receptor to a G protein occurs at intracellular loops of the receptor. G-protein coupled receptors also have sites for glycosylation and phosphorylation.

PSYCHIATRIC ASPECTS OF ION CHANNELS

The successful application of electrophysiological methods to problems of interest in psychiatry makes it likely that principles of cellular physiology will be increasingly important to the understanding and treatment of psychiatric disorders. The ability to describe the effect of drugs on specific ion channels has had a major impact on psychopharmacology. Illuminating the mechanisms of action of commonly used psychotherapeutic agents by exploring the molecular and cellular effects of these drugs is likely to have further effects on the clinical use of the compounds. The interaction of anxiolytics, such as the benzodiazepines, with the $GABA_A$ receptor is an example. A molecular understanding of the effects of current psychotherapeutic agents offers the hope for the design of new agents with superior therapeutic properties. The ability to study the electrophysiological properties of individual neurons shows that the function of the cells partially reflects the selective expression of ion channels. Cells expressing the *N*-methyl-D-aspartate (NMDA) type of glutamate receptor develop enhanced responsivity when given high-frequency stimulation, a phenomenon called *long-term potentiation* (LTP). This process is a model for the synaptic plasticity that might be responsible for memory and learning. Other nerve cells exhibit intrinsic repetitive firing that drives certain neural networks. The sustained activity of some neural systems may be responsible for behavioral states such as sleep and wakefulness. These examples and others illustrate the importance of electrophysiological principles and will be presented briefly below. Ion channel defects also have been linked directly to several human disorders; examples are described to illustrate the pathological impact of channel defects.

GABA Receptors and Anxiolytic Drugs $GABA_A$ receptors are a type of ligand-gated ion channel that pass chloride ions selectively and are inhibitory in most CNS regions. Additionally, $GABA_A$ receptors are a site of action of many CNS depressants including barbiturates, neurosteroids, loreclezole, benzodiazepines, and probably ethanol. These agents appear to act at distinct sites within the $GABA_A$ receptor complex. Benzodiazepines, in particular, have complex effects on $GABA_A$ receptors. Agonists at benzodiazepine sites (e.g., diazepam [Valium], chlordiazepoxide [Limbrium], and alprazolam [Xanax]) increase the apparent affinity of $GABA_A$ receptors for GABA, enhancing responses to low GABA concentrations but not altering the peak responses to maximal GABA concentrations. The net effect of benzodiazepine agonists on $GABA_A$ ion channels is to increase the frequency of channel openings in the absence of changes in intrinsic channel kinetics (e.g., open channel durations). These effects differ from the effects of barbiturates and neurosteroids, which prolong channel open times and are capable of directly opening $GABA_A$ channels in the absence of GABA. A second class of drugs, called *benzodiazepine inverse agonists*, bind to benzodiazepine sites but inhibit Cl^- flux. Certain b-carbolines are examples of inverse agonists at benzodiazepine receptors. The inverse agonists have clinical effects that are the opposite of the effects of the benzodiazepine agonists (e.g., anxiogenic and convulsant activity in contrast to sedation). Flumazenil (Romazicon) binds to benzodiazepine sites but does not alter the flow of Cl^-. Rather, flumazenil blocks the effects of benzodiazepine agonists and inverse agonists, and is referred to as a *benzodiazepine receptor antagonist*. Flumazenil is clinically useful in reversing the effects of benzodiazepine receptor agonists in anesthetic or overdose situations.

The effects of benzodiazepine agonists are heterogeneous and depend on the region of the CNS and the subunit composition of the $GABA_A$ receptors. Certain sites have a high affinity for the hypnotic drug, zolpidem (Ambien) and are referred to as benzodiazepine *type 1 receptors* (BZ_1). Other sites have a lower affinity for this drug and are called *type 2 receptors* (BZ_2). BZ_1 receptors are the predominant $GABA_A$-benzodiazepine receptors in the CNS whereas BZ_2 sites are found primarily in

the hippocampus, neocortex, striatum, and spinal cord.

Studies using recombinant $GABA_A$ receptors have demonstrated that the distinction of the two benzodiazepine receptors results from differences in the subunit composition of $GABA_A$ receptors. So far 15 different $GABA_A$ receptor subunits have been cloned (termed a1-6, b1-4, g1-3, d, and e). Functional GABA-gated ion channels appear to be pentamers composed of various combinations of these subunits. GABA most likely binds to b-subunits, while the presence of a g-subunit is critical for benzodiazepine sensitivity. The distinction between BZ_1 and BZ_2 pharmacology depends on which a-subunit is expressed with b-and g-subunits; a1-containing $GABA_A$ receptors exhibit BZ_1 pharmacology whereas a2-, a3- and a5-containing receptors have BZ_2 responses. Receptors containing a6-subunits (and perhaps a4-subunits) are insensitive to benzodiazepine agonists. The prevalence of sleep and anxiety disorders, coupled with the important clinical effects of benzodiazepines as hypnotics, anxiolytics, anticonvulsants, and muscle relaxants has made $GABA_A$-benzodiazepine receptors a major target for drug development. Current efforts in this direction include the development of receptor-subtype selective agents (such as zolpidem [Ambien]) and partial agonists that have diminished intrinsic efficacy at $GABA_A$ receptors, fewer adverse effects, and perhaps less abuse potential than traditional benzodiazepine agonists.

NMDA Receptors and PCP The effects of neurotransmitters on ion channels is of considerable importance in understanding the mechanisms of actions of a variety of psychoactive drugs. An intriguing observation is that the NMDA class of glutamate receptors may be an important site of action for the street drug, phencyclidine (PCP, “angel dust”). PCP is abused for its hallucinogenic and dissociative (feelings of unreality) properties. PCP and its structural analogs, dizocilpine (MK-801) and ketamine (Ketalar), bind to a site within the NMDA channel and block ion flow. The block of NMDA channels by PCP-like drugs is long-lived, with the ion channel closing around the PCP molecule. Relief of PCP block requires that NMDA channels open at depolarized potentials. It is presently uncertain how the NMDA-channel–blocking effects contribute to the psychotomimetic effects of PCP, although understanding this interaction remains an area of intense investigation. The finding that PCP-like drugs produce pathomorphological changes in posterior cingulate cortical neurons suggests the involvement of specific limbic circuits. NMDA receptors also appear to be an important site of action of alcohol in the CNS; ethanol inhibits NMDA receptors through an unknown mechanism.

An important aspect of NMDA receptors is the role that these ligand-gated channels play in synaptic plasticity. Although the cellular mechanisms underlying learning and memory in the human brain are unknown, the changes are believed to be responsible for certain forms of memory reside in longer-term alterations in synaptic transmission. When glutamate synapses are used at high frequency, they undergo a persistent enhancement of responsivity, a process referred to as *long-term potentiation* (LTP). LTP induction depends critically upon activation of NMDA receptors and requires the coincident detection of changes in presynaptic function (glutamate release) and postsynaptic membrane depolarization. NMDA receptors have several unique properties that make them potential molecular switches for altering synaptic function. First, NMDA ion channels are highly permeable to calcium ions and when these channels open they provide a large calcium signal to neurons. Calcium, in turn, is an important messenger that drives a host of cellular biochemical changes that include activation of specific protein kinases, phospholipases, and other cellular synthetic enzymes. Thus, calcium influx mediated by NMDA channels serves as an important trigger for producing changes in synaptic function. Second, NMDA ion channels are effectively inhibited at membrane potentials near the neuronal resting membrane potential because of a voltage-dependent block by physiological concentrations of extracellular magnesium ions. The magnesium-dependent block of NMDA channels is relieved when the neuronal membrane potential is depolarized. In effect, NMDA receptors serve as coincidence detectors requiring for activation both the binding of glutamate and postsynaptic membrane depolarization. When these conditions are met, NMDA receptors participate in synaptic transmission and play key roles in the induction of LTP. Interestingly, NMDA receptors also participate in some forms of long-term synaptic depression as well, but in the case of “long-term synaptic depression” it appears that the degree of postsynaptic membrane depolarization is less than that which accompanies long-term potentiation possibly resulting in a smaller calcium signal and the preferential activation of protein phosphatases in postsynaptic cells. Presently, LTP and long-term depression are the leading candidates for cellular mechanisms underlying certain forms of learning in the mammalian nervous system. Given that PCP and ethanol inhibit NMDA receptors, it is tempting to speculate that the amnestic effects of these drugs (called “blackouts” in the case of alcohol) result from blockade of NMDA receptors.

Depolarization Block and Antipsychotic Drugs Although the mechanisms involved in the actions of antipsychotic drugs remain a matter of debate, it is clear that these agents alter dopamine function in the CNS, blocking dopamine receptors and increasing dopamine turnover. An interesting finding that may have relevance to the mechanism of action is that antipsychotic drugs alter the firing of dopaminergic neurons. During initial short-term administration, the antipsychotics increase firing of dopamine neurons. However, long-term administration over several weeks leads to a progressive decrease in firing frequency that is accompanied by persistent depolarization of dopamine neurons. The mechanisms responsible for this depolarization block are not well understood, but could involve depolarization-induced inactivation of voltage-gated sodium and calcium channels. The induction and maintenance of the block are known to depend upon afferent inputs to the dopamine neurons from striatum and nucleus accumbens. Recent evidence suggests that depolarization block may involve activation of glutamate-mediated excitatory feedback as a result of chronic blockade of postsynaptic dopamine receptors.

Both dopamine receptor antagonists and serotonin-dopamine antagonists depolarize neurons in the ventral tegmental area neurons that project to cortical and limbic sites whereas only the dopamine receptor antagonists depolarize neurons in the substantia nigra that project to the striatum. This suggests that the lack of effect of the serotonin-dopamine antagonists antipsychotics on substantia nigra dopamine neurons may account for the diminished frequency of extrapyramidal adverse effects associated with these agents. Furthermore, there is a correlation between the ability of antipsychotic drugs to produce depolarization block of mesolimbic dopamine cells and their therapeutic benefits.

Interpretation of the depolarization block hypothesis has been complicated by the fact that only 50 to 80 percent of dopamine cells exhibit depolarization block; the remaining cells show altered firing. Thus, despite the net decrease in firing rate, extracellular dopamine levels do not change. Thus, it may be possible that the mechanisms that compensate for depolarization block are keys to understanding the clinical effects of these drugs.

Oscillatory Neuronal Firing and Behavioral States Certain behavioral states, including wakefulness, attention, mood, and sleep appear to require sustained activity in specific neuronal circuits. Activity in these circuits sets a background tone against which phasic inputs can be evaluated for information content. How this background tone arises is not certain but is believed to involve an interplay of the intrinsic electrophysiological properties of specific neurons and sustained effects of more diffusely acting neurotransmitter systems such as the muscarinic and monoaminergic systems. Certain neurons have voltage-gated ionic conductances that allow them to fire rhythmically and spontaneously, thus having the behavior expected of a pacemaker or oscillator. For example, neurons in the inferior olivary nucleus fire action potentials spontaneously and sustain this firing for relatively long periods in the absence of outside inputs. These inferior olivary neurons fire conventional fast Na^+ spikes that provide the depolarization needed to open high-voltage–activated (HVA) Ca^{2+} channels. In turn, Ca^{2+} influx activates a Ca^{2+}-dependent K^+ conductance that rapidly and effectively hyperpolarizes the membrane. When hyperpolarized, low-voltage–activated (LVA) Ca^{2+} channels open and bring the membrane potential back to the threshold for firing Na^+ spikes, which then activate another cycle. In the case of the inferior olivary neurons, it is the properties of the LVA Ca^{2+} channels that allow oscillatory firing. LVA channels are inactivated at the neuronal resting membrane potential but can be activated only when the membrane is hyperpolarized with respect to rest. In effect, hyperpolarization becomes an activating stimulus that allows the LVA channels to open. The oscillatory firing of inferior olivary neurons then drives Purkinje neurons in the cerebellum at the inferior olivary neurons' preferred firing frequency. The Purkinje neurons thus resonate in response to the inferior olivary input. This resonating circuit is believed to contribute to the physiological resting tremor, which oscillates at about 10 cycles per second; in this circuit the inferior olivary neurons are considered pacemakers.

Pacemaker activity is found in other CNS cells, including thalamic neurons, where similar but not necessarily identical mechanisms are used to drive oscillatory firing. In the thalamocortical system, changes in neuronal activity are associated with the state of arousal and network activity in thalamic and cortical neurons, mediated by both intrinsic neuronal conductances and synaptic connections, drive specific changes in the electroencephalogram (EEG) during different stages of sleep and wakefulness. It is now clear that thalamocortical neurons exhibit two distinct activity states. During sleep, the cells exhibit synchronized rhythms that resemble delta, spindle, and other slow waves on the EEG; during awake states and REM sleep, these neurons show tonic activity. LVA calcium channels are important participants in thalamocortical network activity. The transition from sleep to wakefulness is mediated by the depolarization of thalamic reticular neurons and inactivation of LVA calcium channels. It also appears that specific abnormalities in thalamocortical neurons may be critical in the generation of the 3-Hz spike-and-wave activity seen during absence seizures.

In some regions of the CNS the immediate outputs of the pacemaker cells are mediated by fast excitatory or inhibitory transmitters. However, certain oscillatory cells are capable of firing in bursts of action potentials. Bursts are periods of frequent spike firing, followed by relatively long quiescent periods. This type of firing can be used to drive activity in a local or distributed neural network. Additionally, burst-like firing can provide the increased intracellular Ca^{2+} signal that stimulates release of peptide transmitters from certain cells. In turn, the slow synaptic actions of the peptides in combination with or independent of other G-protein coupled receptor systems can alter the frequency of oscillatory firing and bursting. A clear example of this is the repeated firing that occurs when spike-frequency adaptation is inhibited by blocking Ca^{2+}-activated K^+ conductances. In this fashion, the intrinsic electrical properties of neurons and the effect of modulatory transmitters conspire to determine a background level of activity (or tone) in specific neuronal systems.

Ion Channels and Disease It is now clear that certain clinical disorders result from heritable defects in ion channels. Although most of these illness would not be

considered psychiatric, the ion-channel disorders provide important models for understanding how defects in cellular excitability and ion transport produce pathological effects. One of the best examples of a disorder involving an ion channel is cystic fibrosis, an illness that results from a defect in a specific type of chloride channel protein. The cystic fibrosis transmembrane regulator (CFTR) is a chloride channel that is activated by ATP and phosphorylation. Numerous mutations in the CFTR protein have been described and these typically result in a failure of chloride ions to cross cell membranes. In more than 70 percent of cystic fibrosis cases, the defect is a deletion of a phenylalanine residue at position 508 that results in a failure of chloride channel insertion into cell membranes. The loss of these chloride channels results in marked abnormalities in various bodily secretions.

A second human disorder caused by defective ion channels is the long QT syndrome that is characterized by the presence of prolonged QT intervals on an electrocardiogram (ECG). Individuals with long QT intervals are predisposed to develop malignant cardiac arrhythmias (e.g., torsade des pointes) either spontaneously or during treatment with certain drugs, including tricyclic antidepressants. Several inherited mutations of cardiac ion channels have been implicated in the long QT syndrome. These include mutations in a protein (Kv1QT1) that contributes to the main delayed rectifier potassium current responsible for repolarization in the heart. A second mutation associated with long QT syndrome is in the *human ether-a-go-go-related gene (HERG)* that encodes a rapid potassium channel involved in cardiac repolarization. Interestingly, terfenadine (Seldane) and ketoconazole (Nizoral) block the *HERG* potassium channel, presumably accounting for the development of long QT intervals in patients treated with these drugs. A third ion channel defect associated with long QT syndrome involves a cardiac sodium channel (SCN5A) that participates in the fast upstroke of action potentials. The identified mutations alter sodium channel inactivation and cause prolonged cardiac action potentials.

In addition to disorders that involve ion channels, a number of therapeutic drugs, including psychotherapeutic agents, may alter voltage-gated ion channel function. Several examples have been cited earlier in this chapter, but it is important to emphasize that some anticonvulsants, including carbamazepine (Tegretol) and valproic acid (Depakene) inhibit the function of sodium channels. A dampening of excitability is likely to participate in the anticonvulsant actions, but may also be important in the mood-stabilizing effects of these drugs. Other agents [e.g., verapanil (Calan)] that dampen excitability by blocking voltage-activated calcium channels may also be useful in certain psychiatric disorders.

In the future it is likely that the diversity of voltage-gated and ligand-gated ion channels will be better understood at structural, biophysical, and genetic levels. Although the electrical events underlying neuronal excitability are relatively stereotyped, the various ion channels contributing to neuronal firing offer a great deal of flexibility in the control of cellular activity. Furthermore, the diversity of ion channels involved in electrical signaling provides a complex and powerful means by which excitability can be modulated by neurotransmitters and therapeutic drugs. Determining how subtle alterations in ion channel function contribute to behavior and clinical syndromes will remain a major goal of work in this field.

SUGGESTED CROSS-REFERENCES

Applied electrophysiology is discussed in Section 1.14, monoamine neurotransmitters are discussed in Section 1.4, amino acid neurotransmitters are discussed in Section 1.5, neuropeptides are discussed in Section 1.6, and intraneuronal signaling pathways in Section 1.8. Section 1.15 and Section 1.16 discuss neuroimaging, and basic molecular neurobiology is discussed in Section 1.10. The biology of memory is discussed in Section 3.5.

SECTION REFERENCES

Ackerman MJ, Clapham DE: Ion channels—basic science and clinical disease. N Engl J Med *336:*1575, 1997.

Aguilar-Bryan L, Clement JP, Gonzalez G, Kunjilwar K, Babenko A, Bryan J: Toward an understanding of the assembly and structure of K_{ATP} channels. Physiological Reviews *78:*227, 1998.

Armstrong CM, Hille B: Voltage-gated ion channels and electrical excitability. Neuron *20:*371, 1998.

*Barnard EA: The transmitter-gated channels: A range of receptor types and structures. Trends Pharmacol Sci *17:*305, 1996.

Barry DM, Nerbonne NM: Myocardial potassium channels: Electrophysiological and molecular diversity. Ann Rev Physiol *58:*363, 1996.

Bertolino M, Llinas RR: The central role of voltage-activated and receptor-activated calcium channels in neuronal cells. Ann Rev Pharmacol Toxicol *32:*399, 1992.

Cannon SC: Sodium channel defects in myotonia and periodic paralysis. Ann Rev Neurosc *19:*141, 1996.

Catterall WA: Structure and function of voltage-gated ion channels. Trends Neurosci *16:*500, 1993.

Collo G, North RA, Kawashima E, Merlo-Pich E, Neidhart S, Suprenant A, Buell G: Cloning of $P2X_5$ and $P2X_6$ receptors and the distribution and properties of an extended family of ATP-gated ion channels. J Neurosc *16:*2495, 1996.

*Dingledine R, Borges K, Bowie D, Traynelis SF: The glutamate receptor ion channels. Pharmacolog Rev *51:*7, 1999.

Fozzard HA, Hanck DA: Structure and function of voltage-dependent sodium channels: Comparison of brain II and cardiac isoforms. Physiol Rev *76:*887, 1996.

George AL: Molecular genetics of ion channel diseases. Kidney International *48:*1180, 1995.

*Grace AA, Bunney BS, Moore H, Todd CL: Dopamine-cell depolarization block as a model for the therapeutic actions of antipsychotic drugs. Trends Neurosci *20:*31, 1997.

*Hille B: *Ionic Channels of Excitable Membranes*. Sinauer Associates, Sunderland, MA, 1992.

Hodgkin AL, Huxley AF: A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol (London) *117:*500, 1952.

*Huguenard JR: Low-threshold calcium currents in central nervous system neurons. Ann Rev Physiol *58:*329, 1996.

Jan LY, Jan YN: Cloned potassium channels from eukaryotes and prokaryotes. Ann Rev Neurosci *20:*91, 1997.

Koester J: Membrane potential, passive properties of the neuron, and voltage-gated ion channels and the generation of the action potential. In *Principles of Neural Science,* ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Elsevier, New York, 1991.

*Lee J-H, Daud AN, Cribbs LL, Lacerda AE, Pereverzev A, Klockner U, Schneider T, Perez-Reyes E: Cloning and expression of a novel member of the low voltage-activated T-type calcium channel family. J Neurosci *19:*1912, 1999.

Lingrel JB, Kuntzweiler T: Na^+, K^+-ATPase. J Biol Chem *269:*19659, 1994.

*Lisman JE: Relating hippocampal circuitry to function: Recall of memory sequences by reciprocal dentate-CA3 interactions. Neuron *22:*233, 1999.

Marsh D: Peptide models for membrane channels. Biochem J *315:*345, 1996.

McCormick DA, Bal T: Sleep and arousal: Thalamocortical mechanisms. Ann Rev Neurosci *20:*185, 1997.

Miljanich GP, Ramachandran J: Antagonists of neuronal calcium channels: Structure, function and therapeutic implications. Ann Rev Pharmacol Toxicol *35:*707, 1995.

Miller RJ: Calcium channels prove to be a real headache. Trends Neurosci *20:*189, 1997.

Nicholls JG, Martin AR, Wallace BG: *From Neuron to Brain,* ed 3. Sinauer Associates, Sunderland, MA, 1992.

Pape H-C: Queer current and pacemaker: The hyperpolarization-activated cation current in neurons. Ann Rev Physiol *58:*299, 1996.

Perez-Reyes E, Cribbs LL, Davd A, Lacerda AE, Barclay J, Williamson MP, Fox M, Rees M, Lee JH: Molecular characterization of a neuronal low-voltage-activated T-type calcium channel. *391:*896, 1998.

Siegel GJ, Agranoff BW, Albers RW, Molinoff PB: *Basic Neurochemistry*. Raven Press, New York, 1994.

Sieghart W: Structure and pharmacology of g-aminobutyric acid $_{A}$ receptor subtypes. Pharmacol Rev *47:*181, 1995.

Walker D, De Waard M: Subunit interaction sites in voltage-dependent Ca $^{2+}$ channels: role in channel function. Trends Neurosci *21:*148, 1998.

Zagotta WN, Siegelbaum SA: Structure and function of cyclic nucleotide-gated channels. Ann Rev Neurosci *19:*235, 1996.

1.10 BASIC MOLECULAR NEUROBIOLOGY

STEVEN E. HYMAN, M.D. AND ERIC J. NESTLER, M.D., PH.D.

Deoxyribonucleic acid (DNA) contains the genetic blueprint of nearly all living organisms, from the most primitive bacteria to humans. Only among some viruses is ribonucleic acid (RNA) used instead of DNA for the storage of genetic information. During development the information contained in DNA is read out in a series of steps regulated partly by extracellular signals and partly by hierarchies of information encoded in the DNA itself. In mature organisms, even in postmitotic cells, such as neurons, DNA remains a reservoir of information to be called upon for maintenance of cellular integrity as well as to respond to environmental challenges. The regulation of gene expression by environmental signals is a fundamental mechanism of homeostasis, adaptation, and in the nervous system, of learning and memory. Activation and suppression of the expression of specific genes is involved, for example, in such diverse processes as adaptations to a changing diet, muscle hypertrophy or atrophy in response to changing physical loads, expansion of lymphocyte clones in response to an infection, or the production of long-term memories in the brain in response to environmental stimuli.

Gene-environment interactions are also crucial to understanding the pathogenesis of mental disorders. The eventual discovery of genes that confer vulnerability to (or that protect against) mental disorders will not supersede research on environmental factors; indeed, studies of the inheritance of mental disorders, including studies of monozygotic twins (i.e., twins with identical genomes), have demonstrated that genes confer vulnerability to but not the certainty of developing a mental disorder. To produce illness, and even a particular pattern of illness, genes collaborate with nongenetic factors, some perhaps the result of random events occurring during development, others representing identifiable environmental “second hits.” Increasingly, however, research on the genetics of behavior has found that the environmental factors that interact with genes to produce patterns of behavior are unshared among siblings, that is, they are not the result of readily identified shared factors such as socioeconomic status or a generalized parenting style. This makes the identification of vulnerability genes even more important for the generation of hypotheses or as prospective markers of vulnerability. The contribution of genes and nongenetic factors to illness is well known from general medicine; for example, based on prior gene-environment interactions that govern vulnerability a cigarette smoker may or may not develop lung cancer. The rule for most human illness—including mental illness—is that both genes and environment are likely to play at least some role.

NUCLEIC ACIDS: DNA AND RNA

The human genome contain approximately 80,000 genes, which comprise about 3 percent of the total 3 billion base pairs of DNA that make up the human genome. DNA is a double helix in which each of the two strands is an extended linear polymer synthesized from building blocks called *nucleotide bases*. DNA is synthesized out of four types of nucleotide bases, the purines, adenine (A) and guanine (G), and the pyrimidines, cytosine (C) and thymine (T), each connected to a deoxyribose sugar group.

In order to express the information contained in DNA, the DNA must be transcribed to produce ribonucleic acid (RNA). RNA has some independent functions within cells (e.g., it is a key structural element in ribosomes, the organelles on which proteins are synthesized), but its most prominent function is to serve as an intermediate in the synthesis of proteins, which are the most important structural and functional molecules in cells. RNA differs by one nucleotide building block from DNA; in RNA the pyrimidine uracil (U) takes the place of thymine. In addition, the attached sugar group is ribose instead of deoxyribose. Individual nucleotides are joined into strands of DNA or RNA via phosphate groups that form a phosphodiester linkage with a sugar group. The alternating deoxyribose sugar and phosphate groups that connect the bases of each DNA strand form a sugar-phosphate backbone on the outside of the double helix with the bases arrayed on the inside. In DNA, the nucleotide base A is always paired with or is complementary to T on the opposite strand (A·T), and G is paired with C (G·C). In RNA, U, which is structurally similar to T, is also complementary to A (T·A) (Fig. 1.10-1). The rules of base pairing observed in DNA result from the fact that only complementary pairs of nucleotides form a maximum number of stabilizing hydrogen bonds. Any other arrangement of bases destabilizes the structure of the DNA. If there are too many mismatched bases the two strands will not anneal to form a double helix; areas that are not annealed are targets for DNA repair enzymes.

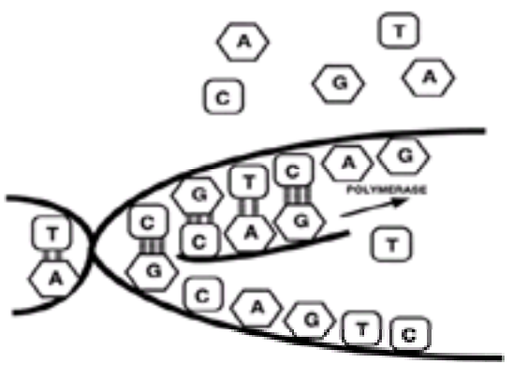

FIGURE 1.10-1 DNA replication. For replication or transcription into RNA, DNA is first unwound and after a variety of biochemical steps an appropriate polymerase adds nucleotides to a growing nucleic acid chain based on the principle of complementary base pairing with the existing strand used as a template. A new double helix is stabilized when hydrogen bonds form between complementary bases of the two strands. Two hydrogen bonds form when A is across from T; three hydrogen bonds form when G is across from C. Other appositions of bases are destabilizing and are repaired by other enzymes; for example, a noncomplementary nucleotide is replaced by a complementary one.

A critical property of such a linear polymer as DNA or RNA is that it can serve as a template for the processive synthesis of other macromolecules. The principle of complementary base pairing provides the mechanism by which information can be transferred. An enzyme, a DNA polymerase in the case of DNA replication, or RNA polymerase in the case of transcription of DNA into RNA, can move down a template strand of DNA adding sequential nucleotide bases complementary to the bases on the template strand (Fig. 1.10-1).

Although the actual enzymatic steps involved in the replication of DNA are complex, the overall principles are simple. Replication begins with separation of the two strands of the double helix in a local region. Each strand then serves as a template for a new DNA molecule by the sequential addition of nucleotides to a growing complementary strand (Fig. 1.10-1). Eventually the replication process generates two complete DNA double helices, each identical in sequence to the original. DNA replication is said to be semiconservative because each daughter DNA molecule contains one of the original parental strands plus one newly synthesized strand. Transcription of DNA into RNA is conceptually similar except that only one DNA strand, not both, serves as a template and when synthesis of the RNA is complete it is released and the DNA strands reanneal into a double helix.

INFORMATION FLOWS FROM DNA TO RNA TO PROTEIN

DNA carries information in its linear sequence of nucleotides. Although the linear polynucleotide structure of DNA is well suited for the stable storage of information and for self-replication, its chemical simplicity and its relatively rigid helical structure limit its biological functions. Thus, the information contained within DNA must be

read out to yield RNAs and proteins. Like DNA, RNA is chemically quite simple (i.e., composed of four nucleotides), but since it is a flexible single strand, free to fold into a variety of conformations, it is functionally more versatile than DNA.

Messenger RNA (mRNA) functions as an intermediate between the sequences of DNA that comprise the transcribed regions of gene, and the sequence of proteins (Fig. 1.10-2). Not all RNA serves as mRNA, however. Other RNAs serve distinct functional roles in cells. Ribosomes, the organelles on which proteins are synthesized, are constructed out of complexes of various ribosomal RNAs (rRNAs) together with proteins. Transfer RNAs (tRNAs) transport specific amino acids to the ribosomes for incorporation into proteins during the process in which mRNA is translated into protein.

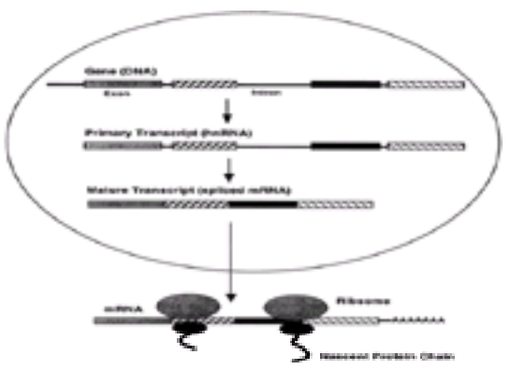

FIGURE 1.10-2 Processing of RNA. Gene transcription produces an RNA that is colinear with the transcribed region of the DNA. This complementary copy of the DNA contains sequences destined to be exported from the nucleus to be translated into protein (exons) and intervening sequences that will remain in the nucleus and be degraded (introns). The primary transcript is often described as heteronuclear RNA (hnRNA). Only after the introns are spliced out and other chemical modifications occur is the messenger RNA (mRNA) ready to exit the nucleus to be translated into protein within the cytoplasm. The most significant chemical modifications are aimed at increasing the stability of the mRNA. These include the addition of multiple A nucleotides at the end of the RNA (a so-called *poly-A tail*). Particular exons within some genes in some cells may be excluded from the mature mRNA. This process of alternative splicing permits the generation of more than one mRNA and therefore more than one protein from a single gene. Within the cytoplasm, multiple ribosomes can process down the mRNA at once.

As in the case of DNA, mRNA carries information encoded in its linear sequence of nucleotides. DNA and mRNA specify amino acid building blocks for proteins in linear stretches of three nucleotides. Proteins are comprised of unbranched chains of amino acid building blocks. An amino acid is a small molecule that contains an amino group (NH_2) and a carboxylic acid or carboxy group (COOH), plus a variable side chain. The side chains used by the 20 common amino acids differ markedly in size, shape, hydrophobicity, and charge. Amino acids are linked to each other by peptide bonds that join the amino group of one amino acid to the carboxy group of another amino acid.

Within the portion of an mRNA that is translatable into a protein, each successive group of three nucleotide bases specifies either one amino acid or termination of the protein chain and is called a *codon*. The rules specifying the correspondence between each trinucleotide combination (codon) and an amino acid is called the *genetic code*. Since RNA is a linear polymer of 4 nucleotides, there are 4^3 (i.e., 64) possible codons, but only 20 amino acids. As a result, although each codon specifies only a single amino acid, most amino acids are specified by more than one codon. The genetic code is therefore said to be degenerate. With only a few minor exceptions the code has been conserved across evolution. The codons in an mRNA molecule do not interact directly with the amino acids they specify; the translation of mRNA into protein depends on the presence of tRNAs, which serve as adapter molecules that recognize a specific codon in the mRNA as well as the corresponding amino acid (Fig. 1.10-3).

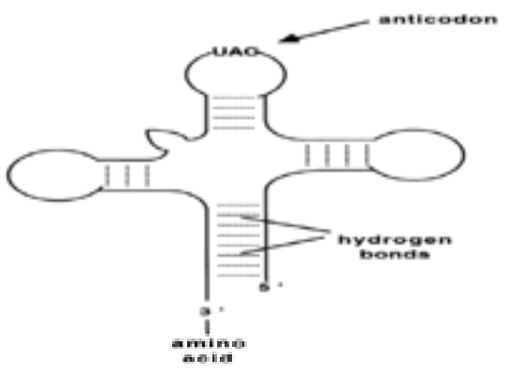

FIGURE 1.10-3 Transfer RNA. Transfer or tRNA is a single strand of RNA that folds on itself based on the formation of complementary base pairs. This hybridization provides hydrogen bonds—indicated by the dotted lines—that stabilize the secondary structure of the molecule. One of the loops that is formed contains the anticodon, the sequence of three nucleotides on the tRNA that binds to the complementary codon on a messenger RNA molecule. For the anticodon UAC shown in the figure, the corresponding codon on the mRNA would be AUG. The free end of the tRNA binds to a specific amino acid, in this case, methionine. Each tRNA with a given anticodon binds only one type of amino acid. (Adapted from Hyman SE, Nestler EJ: *Molecular Foundations of Psychiatry*. American Psychiatric Association, Washington, DC, 1993.)

The ribosome is a structure composed both of proteins and structural RNAs. These organelles provide a structure on which tRNAs can interact via their anticodons with the codons of an mRNA in sequential order. The ribosome finds a specific start site on the mRNA that sets the reading frame, and then moves along the mRNA molecule progressively, translating the nucleotide sequence one codon at a time, using tRNAs to add amino acids to the growing end of the polypeptide chain (Fig. 1.10-2 and Fig. 1.10-3).

Unlike nucleic acids, which are constructed of four bases that are chemically quite similar, proteins are constructed out of twenty quite different amino acids. By incorporating so many different types of amino acids, each with their chemically diverse side chains, proteins have extraordinary functional versatility, unlike DNA or RNA. The specific properties of proteins depend not only on the linear sequence of their amino acid building blocks (primary structure), but also on the tendency of certain combinations of amino acids to form intrinsic structural motifs (secondary structures, e.g., a helixes or b sheets), and by their folded three-dimensional characteristics (tertiary structure). In addition, proteins form stable interactions with other proteins (quaternary structure). In such cases the individual polypeptide chains are called *subunits*. The folding of proteins, and the interactions of proteins with each other and with other molecules such as nucleic acids, may also be regulated by chemical modification of the protein, most often of particular side chains. For example, one ubiquitous mechanism of regulation of protein function is by the covalent addition of a phosphate group (by specific enzymes called *protein kinases*) to the hydroxyl groups found in serine, threonine, or tyrosine side chains. Cells may contain tens of thousands of distinct proteins, each with unique structural and functional properties. Some examples include neurotransmitter receptors, neuropeptides, ion channels and pumps, cytoskeletal and contractile proteins, and many different types of enzymes.

GENES AND CHROMOSOMES

As a first approximation, genes can be defined as stretches of DNA that encode a single protein or a single functional RNA, such as an rRNA or tRNA. There are exceptions to this rule because there are mechanisms, such as alternative splicing of the primary RNA transcript into different mRNAs, that may intervene between a given gene and a finished protein. As a result, in some cases a single gene may actually encode multiple proteins.

Genes are arrayed on extremely long chains of DNA called *chromosomes*. Chromosomes of eukaryotic organisms also contain large amounts of intergenic DNA. Indeed, within the human genome the 80,000 genes take up perhaps only 3 percent of chromosomal DNA. Moreover, genes are not distributed equally along the

chromosomes, but are often found in clusters. Some chromosomal regions are gene rich and others relatively gene poor. Within intergenic regions there is a great deal of DNA with unique sequences of unknown function, but also long stretches of tandemly repeated sequences, known as *satellite DNA*. These tandem repeats, which show a great deal of variability in length from person to person, are the basis of commonly used genetic markers for research on the pattern of heritability of traits, such as in linkage analysis, as well as for the identification of individuals in forensic investigations.

The chromosomes of eukaryotic cells are so long that they would not fit in the nucleus in their extended form. Stretches of the DNA that are not being actively transcribed are tightly packed into a conformation described as a *coiled coil*, which permits the chromosomes to fit within the nucleus. To create packed conformations, lengths of DNA are coiled around structural and regulatory proteins of which the most important are the histones. Regions of DNA that are being actively transcribed into RNA may be greater than 1000-fold more extended than regions that are transcriptionally quiescent.

GENE EXPRESSION

Proteins are not synthesized directly from the DNA that encodes them, but in two sequential processes: transcription of DNA into mRNA, which occurs in the nucleus, and translation of the mRNA into protein according to the rules of the genetic code, which occurs in the cytoplasm. Transcription of protein-coding genes can be divided into three major steps. First, the enzyme primarily involved in RNA synthesis, RNA polymerase, must interact with the gene at an appropriate transcription start site and begin transcribing (*initiation*). Second, the RNA polymerase must successfully transcribe an appropriate length of RNA (*elongation*). Third, transcription of the RNA must terminate appropriately. The resulting RNA is then posttranscriptionally processed. It receives a stretch of adenine nucleotides, called a *poly-A tail*, which makes the RNA more stable in the cell (Fig. 1.10-2). The primary transcript, which is an exact complement to the DNA, is also called *heteronuclear RNA*.

In higher eukaryotes, not all transcribed sequences contribute to the formation of the mRNA, which will eventually be translated into protein. There are also intervening sequences of unknown function that interrupt the continuity of the gene. DNA sequences within a gene that code for a segment of mRNA are called *exons* because the information in those sequences will be exported from the nucleus; the intervening sequences, which remain in the nucleus are called *introns*. When a protein coding gene is first transcribed, a long RNA, the primary transcript, is produced that is colinear with the DNA and therefore contains both exons and introns (Fig. 1.10-2). Before the transcript exits the nucleus, the introns are removed and the exons are spliced together to form a mature mRNA. Precise RNA splicing depends on three types of DNA sequence. Two mark either side of the splice junction; the third is located within the intron itself and is called the *branch point*. This sequence plays a critical role in the chain of reactions that produce accurate splicing. The spliced mRNA leaves the nucleus and binds to a ribosome in the cytoplasm where it can direct the synthesis of a protein; however, the entire mature mRNA is not translated. All mRNAs contain untranslated flanking sequences at their ends.

Many genes contain multiple introns and exons that may not be spliced identically in every cell type or in a given cell type at every stage of development. This mechanism, alternative splicing, can produce functionally very different forms of a protein or even entirely different proteins from a single gene. One of the first examples of alternative splicing discovered was that of calcitonin and calcitonin gene-related peptide (CGRP), two different peptides derived from alternative splicing of the same primary transcript.

TRANSCRIPTIONAL CONTROL

It is at the step of transcription initiation that environmental signals most often exert powerful regulatory control on gene expression, both during development and in mature cells. In addition to containing information that ultimately directs the synthesis of proteins (i.e., the genetic code), genes also contain regulatory information. Every cell in our bodies contains in its nucleus all 80,000 or so of our genes, but not every gene is active in every cell. Our cells differ—indeed, their identity is defined by the fact that each type of cell in the body expresses only a subset of the entire complement of genes. In any given cell, some genes are “on,” being read out to make RNA and hence proteins, and the rest are silent. For example, in the red blood cell precursors of the bone marrow, the genes that encode globins, the proteins that form the critical building blocks of hemoglobin, are highly active making globin-encoding mRNA. In the midbrain a subset of cells make a protein called tyrosine hydroxylase, which is the rate-limiting enzyme for the synthesis of dopamine and other catecholamines. It would not be adaptive for midbrain neurons to be synthesizing large amounts of hemoglobin, just as red blood cell precursors might find it deleterious to express the enzyme activity conferred by the presence of tyrosine hydroxylase.

Regulatory sequences within DNA control the expression of genes by virtue of their ability to bind specific regulatory proteins. Certain regulatory sequences of DNA specify the beginnings and ends of DNA segments that can be transcribed into RNA. Other regulatory sequences determine in what cell types and under what circumstances the gene to which they are linked can be read out. DNA sequences that subserve control functions are often called *cis*-regulatory elements (Fig. 1.10-4). The term *cis* refers to the fact that the relevant regulatory sequences are physically linked on the DNA to the region being controlled. The proteins that bind to *cis* elements have been described as *trans*-acting factors because they may be encoded anywhere in the genome rather than on the same stretch of DNA that they regulate. Proteins that are involved in specifying whether and under what circumstances a gene will be transcribed are more generally known as *transcription factors*. Many transcription factors bind DNA directly; others interact only indirectly via protein-protein interactions with other factors that do bind DNA.

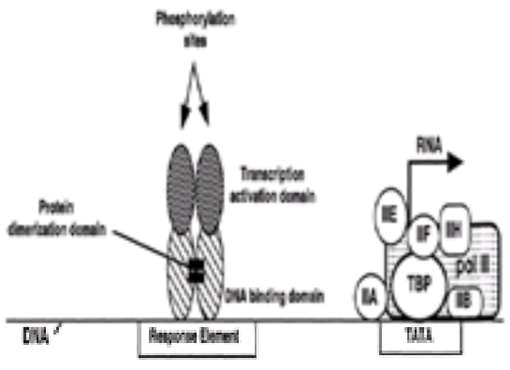

FIGURE 1.10-4 *Cis* and *trans* regulation. The figure shows two *cis*-regulatory elements (open rectangles) along the stretch of DNA (thin line). The element to the left represents a response element that serves as a binding site for a hypothetical transcription factor that binds as a homodimer. The other is the TATA element, shown binding the TATA binding protein TBP. Multiple general transcription factors and RNA polymerase II (pol II) associate with TBP. This basal transcription apparatus recruits RNA polymerase II into the complex and also forms the substrate for interactions with activator proteins, such as those binding to the activator elements shown. The transcription factor shown binding to the response element is a substrate for a protein kinase that phosphorylates its activation domain. This in turn would alter the ability of the transcription factor to interact with the basal transcription factor apparatus. (Adapted from Hyman SE: Regulation of gene expression by neural signals. Neuroscientist *2:*217, 1996.)

Those *cis*-regulatory elements that specify the site within a gene at which transcription starts, and upon which the complex of proteins that forms the basal transcription apparatus is assembled, are called the *basal* or *core promoter*. The core promoters for most protein-encoding genes are comprised of a sequence motif rich in the nucleotide bases A and T, and are thus commonly called the *TATA box*. The TATA box is generally located between 25 and 30 nucleotides upstream of the actual site at which transcription of DNA into RNA is initiated. In eukaryotes, transcription of protein-coding genes is carried out by the enzyme RNA polymerase II, which does not directly contact DNA but interacts with the complex of proteins that assembles at the TATA box (Fig. 1.10-4).

Transcription Factors The basal transcription complex, which is assembled at the TATA box, is sufficient to set the start site and direction of transcription, but it is not adequate to initiate more than low levels of transcription of DNA into RNA. To achieve significant levels of gene expression the basal transcription complex requires help from additional transcription factors that bind to *cis*-regulatory elements found elsewhere within the gene. *Cis*-elements that exert control near the core promoter itself have been called *promoter elements* and those that act at a distance—often several hundred to more than a thousand nucleotide base pairs away—have been called *enhancer elements*, but the commonly made distinction between promoter and enhancer elements is artificial. Both are composed of small “modular” sequences of DNA (generally 7 to 12 base pairs in length), each of which is a specific binding site for one or more transcription factors. Multiple *cis*-regulatory elements arrayed throughout the control regions of a gene, and the proteins they bind, act in a combinatorial fashion to give each gene its distinct patterns of expression and regulation.

Transcription factors that are tethered to DNA by binding *cis*-elements often have a modular structure comprised of physically separate domains: a domain that recognizes and binds a specific DNA sequence, an activation domain that interacts with the basal transcription complex or an adapter protein to form an active complex, and a oligomerization domain that permits the formation of homomeric complexes or heteromeric complexes with other transcription factors (Fig. 1.10-5A and Fig. 1.10-5B, and Fig. 1.10-6). Many transcription factors are active only as dimes or higher-order complexes. Within transcription factor dimers, whether homodimers or heterodimers, it is common for both partners to contribute jointly to both the DNA-binding domain and to the activation domain. Some transcription factors contact the basal transcription complex directly; others interact through the mediation of adapter proteins. In either case, transcription factors that bind *cis*-regulatory elements at a distance from the core promoter can interact with the basal transcription apparatus because the DNA forms loops that bring distant regions in contact with each other (Fig. 1.10-5B).

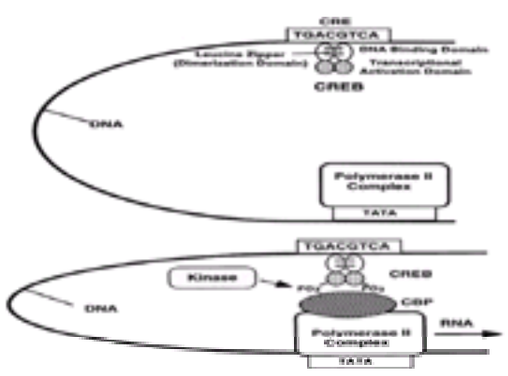

FIGURE 1.10-5 A and B. Looping of DNA permits activator (or repressor) proteins binding at a distance to interact with the basal transcription apparatus. The basal transcription apparatus is shown as a single box bound at the TATA element. The activator protein, CREB, is shown bound as a homodimer to its cognate *cis*-regulatory element, the CRE, at a distance from the TATA element (**A** and **B**). CREB dimerization occurs at a specialized domain called a leucine zipper, because several leucine residues on one CREB molecule bind (via hydrophobic interactions) with several leucine residues on the other CREB molecule. Upon phosphorylation many activators such as CREB are able to recruit adapter proteins that mediate between the activator and the basal transcription apparatus. An adapter protein that binds phosphorylated CREB is called CREB binding protein (CBP). With the recruitment of the adapter, a mature transcription complex forms that permits the synthesis of RNA by RNA polymerase II (**B**). (Adapted from Hyman SE: Regulation of gene expression by neural signals. Neuroscientist *2:*217, 1996.)

Regulation of Gene Expression by Extracellular Signals The nervous system responds to salient environmental stimuli, drugs, and changes in the internal milieu with long-term changes in neural function, which may result in alterations in behavior. A high degree of plasticity is characteristic of the mammalian brain, including its ability to form long-term memories within many different types of circuits. Plasticity is required for normal development and learning, but also contributes to the pathophysiology of psychiatric disorders and also to their treatment. Many neural mechanisms contribute to plasticity, but the formation of long-term memories and long-lived adaptations to changes in the environment depends on the regulation of gene expression.

Environmental stimuli are transduced by neurons into neurotransmitter signals. The binding of neurotransmitter to their receptors initiates cascades of intracellular signals, most notably second messengers and signal-regulated protein kinases. Along with neurotransmitters another important intracellular signaling mechanism are the steroid hormones, such as cortisol, which are released in response to stress. Both second messenger systems and steroid hormones regulate the expression of genes in the cell nucleus. Some second messengers may actually enter the cell nucleus. In most cases, however, they modify the function of another protein such as a protein kinase, which enters the nucleus and interacts with transcription factors that directly or indirectly bind to DNA. Steroid hormones are lipophilic molecules that cross the cell membrane and bind a receptor in the cytoplasm. The activated steroid-receptor complex then translocates to the nucleus where it binds to specific *cis*-regulatory elements. Many and possibly most genes contain *cis*-regulatory elements that confer responsiveness to such physiological signals. Such *cis*-elements are called *response elements*. Response elements work by binding transcription factors that are activated (or inhibited) in response to specific physiological signals, such as second messenger–dependent phosphorylation, steroid hormone binding, or drugs. Many examples of transcription factors activated by physiological and pharmacological signals are now known. A small number of these will be described because of their relevance to neural plasticity, including the action of psychotropic drugs.

Cyclic Adenosine Monophosphate (cAMP) Response Elements in the Regulation of Neural Genes One of the first types of response elements to be characterized in detail was an element that permits genes to respond to stimulation of receptors, such as b-adrenergic, dopamine (D_1), or type 1 serotonin (5-hydroxytryptamine [5-HT]) type 4 ($5\text{-}HT_4$), $5\text{-}HT_6$, and $5\text{-}HT_7$ receptors, that activate the cyclic AMP (cAMP) pathway (see Fig. 1.10-6). Such cAMP response elements were subsequently shown to confer responsiveness to calcium (Ca^{2+}) as well. Ca^{2+} is an important second messenger in the nervous system. Neurons express a variety of receptors, such as *N*-Methyl-D-aspartate glutamate receptors, that permit Ca^{2+} entry, and also express a large number of voltage-sensitive Ca^{2+} channels that are activated when a neuron is depolarized. Most recently, cAMP response elements have been shown to be responsive to several types of growth factors, which produce their effects on neurons via complex signal transduction pathways. Thus, there are many intercellular signals that could ultimately converge to produce activation of gene expression via cAMP response elements.

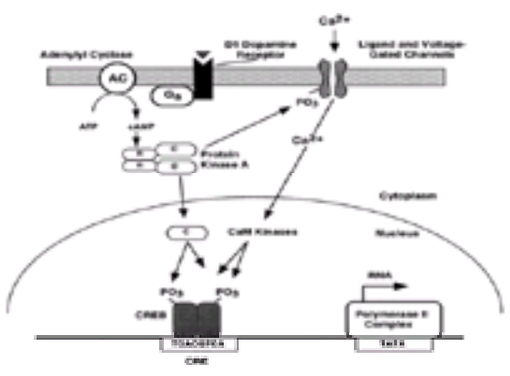

FIGURE 1.10-6 Summary of cyclic AMP-mediated signaling to the nucleus. Many neurotransmitters (shaded triangle) bind to receptors (black) that are linked via the G protein G_s to activation of adenylyl cyclase (AC). Adenylyl cyclase catalyzes the synthesis of the second messenger cAMP from adenosine triphosphate (ATP). While cAMP has several functions in cells, its major effect is to bind to the regulatory subunits, R, of the cyclic AMP-dependent protein kinase (also known as protein kinase A or PKA). Following cAMP binding to R, the catalytic subunits, C, are freed to phosphorylate substrate proteins. For example, they may phosphorylate L-type Ca^{2+} channels, increasing Ca^{2+} entry into cells. The C subunits can also directly enter the cell nucleus where they phosphorylate the cyclic AMP response element binding protein, which binds DNA at cyclic AMP response elements that contain the DNA sequence motif CGTCA. This converts CREB into a transcriptional activator, Ca^{2+} may also lead to CREB phosphorylation. The precise nuclear translocation step is unclear, but may involve the direct transport of Ca^{2+} into the nucleus where it can activate Ca^{2+}/calmodulin-dependent protein kinases (especially CaMK-IV) found in the nucleus. (Adapted from Hyman SE: Regulation of gene expression by neural signals. Neuroscientist *2:*217, 1996.)

Among the first genes in which cAMP response elements were identified were genes encoding the neuropeptides, somatostatin, proenkephalin, and vasoactive intestinal polypeptide. However, cAMP response elements have now been identified in a large number of genes expressed in the nervous system, ranging from enzymes such as tyrosine hydroxylase to transcription factors such as c-*fos*.

CRE binding protein (CREB) is the major protein that binds cAMP response element in most cell types that have been investigated. The cAMP response element is a palindromic DNA sequence that most often binds two molecules of CREB associated with each other as a homodimer (Fig. 1.10-6). However, CREB is a member of a

family of related proteins that can bind cAMP response elements. Many transcription factors are members of families, permitting subtle forms of positive and negative regulation. CREB is closely related to proteins called the *activating transcription factors* and the cAMP response element modulators (CREMs). Many of these proteins can dimerize with CREB and regulate its activity.

CREB is constitutively synthesized in most cells, and, under basal conditions, is already bound to cAMP response elements. Bound CREB becomes a transcriptional activator when it is phosphorylated on a critical amino acid (Ser_{133}) by any of several protein kinases, including cAMP-dependent or Ca^{2+}-dependent protein kinases or certain growth–factor-activated protein kinases. Such phosphorylation permits CREB to interact with adapter proteins, such as CREB binding protein *(CBP)* or a protein called p300 (Fig. 1.10-5A and Fig. 1.10-5B), which in turn contacts the basal transcription apparatus and permits transcriptional activation.

Because CREB is already located within the nucleus under basal conditions, CREB activation requires the translocation of protein kinases (or regulators of protein kinases) to the nucleus (Fig. 1.10-6). The stimulation of receptors, such as b-adrenergic or D_1 receptors, that are coupled to the stimulatory G protein G_s activates adenylyl cyclase, leading to generation of cAMP at the cell membrane. The cAMP-dependent protein kinase is activated by binding of cAMP to its regulatory subunits, which leads to release of active catalytic subunits that can translocate into the nucleus where they can phosphorylate CREB (Fig. 1.10-6). Phosphorylation of CREB by Ca^{2+}-dependent protein kinases involves a somewhat different mechanism or translocation. It appears that Ca^{2+} interacts with its binding protein, calmodulin in the cytoplasm. The calcium/calmodulin complex then translocate to the nucleus where it activates Ca^{2+}/calmodulin dependent protein kinase type IV. This kinase phosphorylates CREB.

The convergence of multiple signaling pathways (the cAMP and Ca^{2+} pathways) on a single transcription factor has important implications. Such convergence is particularly important in the nervous system because this is a candidate mechanism for the formation of associative memories. Associative memory depends on the temporally closely coordinated arrival of two different signals that must then be integrated within target cells and circuits. The attractiveness of hippocampal long-term potentiation as a mechanism of memory is based in part on the need for simultaneous activation of two different inputs for its initiation. There is strong evidence from a variety of model systems that the conversion of short-term to long-term memory requires new protein synthesis and therefore new gene expression. Indeed, there is good evidence that certain kinds of long-term plasticity are dependent on cAMP response elements and CREB. For example, experiments in *Drosophila* and mice in which CREB activity was altered gene targeting yield organisms with deficits in long-term memory.

CREB has also shown to be phosphorylated in the striatum, including the nucleus accumbens, in response to the administration of cocaine and amphetamine, and to be activated in the locus nucleus coeruleus during opiate withdrawal. CREB-deficient mice, and animals with more selective reductions in CREB levels in specific brain regions achieved by use of antisense oligonucleotide methods, exhibit attenuated behavioral withdrawal from opiates, further suggesting that CREB is involved in adaptation to long-term opioid administration. Understanding how CREB activation by extracellular stimuli leads to the altered expression of target genes that lead to the formation of memory in the hippocampus, or aspects of addiction in the nucleus accumbens and locus coeruleus, are currently areas of intense research. Overall, the mechanisms underlying the many forms of plasticity of which the nervous system is capable are likely to be quite complex, involving a multitude of different target genes and many different types of regulation. Nevertheless, CREB appears to be important, and also provides a good model to understand the role of gene expression in long-term changes in neural function.

AP-1 Family of Proteins In addition to the CREB family, another family of transcription factors that appears to play a crucial role in the regulation of neural gene expression is the *activator protein-1 (AP-1) protein* family that binds as heterodimers (and sometimes as homodimers) to the DNA sequence TGACTCA, the AP-1 sequence. The AP-1 sequence differs from the sequence of cAMP response elements by only a single base. Yet this one base difference strongly biases protein binding away from CREB (which requires an intact CGTCA motif) to the AP-1 proteins. Thus, under most circumstances this sequence will not confer cyclic AMP responsiveness on linked genes. Instead, AP-1 sequences confer responsiveness to certain growth–factor-stimulated signaling pathways as well as several other pathways. The first signaling pathway shown to activate transcription via AP-1 sequences was the protein kinase C pathway; the AP-1 DNA sequence is sometimes described as a 12-O-tetradecanoyl-phorbol-13-acetate (TPA), *response element* (TRE) because the phorbol ester TPA, which activates protein kinase C, can induce gene expression via the AP-1 site.

The AP-1 proteins generally bind DNA as heterodimers comprised of one member each of two different families of related proteins, the Fos family and the Jun family. The known members of the Fos family are c-Fos, Fra-1 (Fos-related antigen-1), Fra-2, and FosB; The known members of the Jun family are c-Jun, JunB, and JunD.

Unlike CREB, which is constitutively synthesized, members of the Fos family are generally induced from low basal levels upon neuronal stimulation, and rapidly return to basal levels shortly after stimulation. It has been shown, for example, that following a seizure c-Fos is induced from very low levels in the hippocampus, with the other Fras induced similarly with a slightly delayed time course. As another example, acute administration of cocaine or of haloperidol [Haldol] rapidly and transiently induces c-Fos and the various Fras from very low concentrations in different cells in the striatum and prefrontal cortex. A very different pattern is seen, however, after long-term stimulation. After repeated administration of cocaine, the ability of a subsequent acute stimulus to induce c-Fos and the various Fras shows desensitization. Instead, there is a gradual accumulation of truncated products of the *fos*B gene (also termed *chronic Fras*) that have a very long half-life. This change over time in the members of the Fos family that are expressed is likely to have important consequences for the regulation of target genes. Alterations in AP-1 complexes over time with long-term stimuli, such as long-term cocaine administration, might permit cells to activate specific programs of adaptation appropriate to the strength and time course of the stimulus.

Immediate Early Genes Genes that are transcriptionally activated by synaptic activity, drugs, and growth factors have often been classified into two main groups. Genes, such as the c-*fos* gene, that are activated rapidly (within minutes), transiently, and without requiring new protein synthesis are frequently referred to as *cellular immediate early genes* (IEGs). Genes that are induced or repressed more slowly (over hours), and are dependent on new protein synthesis, may be described as *late-response genes*. The term *IEG* was initially applied to describe viral genes that are activated immediately upon infection of eukaryotic cells by commandeering host cell transcription factors for their expression. Viral immediate early genes generally encode transcription factors needed to activate viral “late” gene expression. This terminology has been extended to cellular (i.e., nonviral) genes. The terminology is somewhat problematic because many cellular genes are induced independently of protein synthesis, but with a time course intermediate between classic IEGs and late-response genes. Some genes may be regulated with different time courses or requirements for protein synthesis in response to different extracellular signals. Moreover, many cellular genes regulated as IEGs encode proteins that are not transcription factors. Despite these caveats, the concept of IEG-encoded transcription factors in the nervous system has provided a useful heuristic for understanding cascades of gene regulation. The protein products of those cellular IEGs that function as transcription factors bind to *cis*-regulatory elements contained within a subset of late-response genes to activate or repress them. IEGs such as c- *fos* have therefore been termed “third messengers” in signal transduction cascades, with neurotransmitters designated as intercellular first messengers, and small intracellular molecules, such as cyclic AMP and Ca^{2+}, as second messengers. Overall, neural genes that are regulated by extracellular signals are activated or repressed with varying time courses by reversible phosphorylation of constitutively synthesized transcription factors and by newly synthesized transcription factors, some of which are regulated as IEGs.

Because of their rapid induction from low basal levels in response to neuronal activation (the critical signal being Ca^{2+} entry) and second messenger and growth factor pathways, several IEG products, including c-Fos, c-Jun, and Zif268 (Egr-1) have been used widely as cellular markers of neural activation. IEG mapping is now one of the fundamental tools of functional neuroanatomy. As long as a stimulus is excitatory or activates the appropriate second messengers, it is likely to activate several IEGs, such as c-*fos*, in cells that process the stimulus. Thus, IEG expression has been used to identify cells in the nervous system that respond to a wide variety of perturbations: for example, sensory stimuli (e.g., stress, nociceptive stimuli, conditioning), drugs (e.g., cocaine, antipsychotic drugs), and immunological challenges (e.g., bacterial endotoxin).

Steroid Hormones Another important family of physiologically regulated *cis*-regulatory elements are the glucocorticoid response elements (GREs). Glucocorticoid and other steroid hormones bind to and thereby activate specific receptors within the cytoplasm of cells. These receptors are held in the cytoplasm of cells by a variety of proteins (called *chaperone proteins*) until hormone binding occurs. Once activated by steroid hormone binding, the receptors are freed from their cytoplasmic anchoring proteins and translocate into the nucleus where they bind to GREs or other steroid hormone response elements contained within particular genes. Such binding then increases or decreases the rate at which these target genes are transcribed, depending on the precise nature and DNA sequence context of the element. Most of the known effects of glucocorticoids, gonadal steroids, thyroid hormone, retinoic acid, and vitamin D on cellular function are mediated via such actions on gene expression. Of great interest, steroid hormone receptors appear to be able to form protein-protein interactions with AP-1 proteins and with members of the CREB family, creating the possibility of very complex regulatory mechanisms.

FUTURE DIRECTIONS

Information encoded in the approximately 80,000 genes in the human genome is read out during development and in mature cells in response to information from the environment. An understanding of how variations in gene sequences lead to disease vulnerability or resistance, and an understanding of how disease, drugs, and

salient environmental stimuli lead to activation or suppression of the expression of genes are among the most important goals of current biology. Both aspects of molecular genetics will play a central role in the eventual understanding of mental illness and its treatment.

SUGGESTED CROSS-REFERENCES

Intraneuronal signaling pathways are discussed in Section 1.8, neurotransmitters in Section 1.4 and Section 1.5, and genetics in Section 1.17 and Section 1.18.

SECTION REFERENCES

Bourtchuladze B, Frenguelli B, Blendy J, Cioffi D, Schutz G, Silva A: Deficient long-term memory in mice with a targeted mutation of the cAMP responsive element binding (CREB) protein. Cell *79:*59, 1994.

Carew TJ: Molecular enhancement of memory formation. Neuron *16:*5, 1996.

Carey M: The enhanceosome and transcriptional synergy. Cell *92:*5, 1998.

Cole RL, Konradi C, Douglass J, Hyman SE: Neuronal adaptation to amphetamine and dopamine: Molecular mechanisms of prodynorphin gene regulation in rat striatum. Neuron *14:*813, 1995.

Darnell JE Jr: STATs and gene regulation. Science *277:*1630, 1997.

Deisseroth K, Bito H, Tsien RW: Signaling from synapse to nucleus: Postsynaptic CREB phosphorylation during multiple forms of hippocampal synaptic plasticity. Neuron *16:*89, 1996.

Deisseroth K, Heist EK, Tsien RW: Translocation of calmodulin to the nucleus supports CREB phosphorylation in hippocampal neurons. Nature *392:*198, 1998.

DeZazzo J, Tully T: Dissection of memory formation: From behavioral pharmacology to molecular genetics. Trends Neurosci *18:*212, 1995.

Duman RS, Heninger GR, Nestler EJ: A molecular and cellular hypothesis of depression. Arch Gen Psychiatry *54:*597, 1997.

*George R: Molecular genetics of substance abuse vulnerability: A current approach. Neuropsychopharmacology *20:*3, 1999.

Gingrich JR, Roder J: Inducible gene expression in the nervous system of transgenic mice. Annu Rev Neurosci *21:*377, 1998.

Ginty DD, Kornhauser JM, Thompson MA, Bading H, Mayo KE, Takahashi JS, Greenberg ME: Regulation of CREB phosphorylation in the suprachiasmatic nucleus by light and circadian clock. Science *260:*238, 1993.

Hagiwara M, Brindle P, Hartounian A, Armstrong R, Rivier J, Vale W, Tsien RY, Montminy MR: Coupling of hormonal stimulation and transcription via the cyclic AMP responsive factor CREB is rate-limited by nuclear entry of protein kinase A. Mol Cell Biol *13:*4852, 1994.

*Hill CS, Treisman R: Transcriptional regulation by extracellular signals: Mechanisms and specificity. Cell *80:*199, 1995.

Hiroi N, Brown J, Haile C, Ye H, Greenberg ME, Nestler EJ: FosB mutant mice: Loss of chronic cocaine induction of Fos-related proteins and heightened sensitivity to cocaine's psychomotor and rewarding effects. Proc Natl Acad Sci USA *94:*1037, 1997.

Hope B, Kosofsky B, Hyman SE, Nestler EJ: Regulation of immediate early gene expression and AP-1 binding in the rat nucleus accumbens by chronic cocaine. Proc Natl Acad Sci USA *89:*5764, 1992.

Hope BT, Nye HE, Kelz MB, Self DW, Iadarola MJ, Nakabeppu Y, Duman RS, Nestler EJ: Induction of a long-lasting AP-1 complex composed of altered Fos-like proteins in brain by chronic cocaine and other chronic treatments. Neuron *13:*1235, 1994.

*Hyman SE: Regulation of gene expression by neural signals. Neuroscientist *2:*217, 224, 1996.

*Hyman SE, Nestler EJ: *The Molecular Foundations of Psychiatry*. American Psychiatric Association, Washington, DC, 1993.

*Hyman SE, Nestler EJ: Initiation and adaptation: A paradigm for understanding psychotropic drug action. Am J Psychiatry *153:*151, 1996.

Karin M: The regulation of AP-1 activity by mitogen-activated protein kinases. J Biol Chem *270:*16483, 1995.

Kwok RS, Lundblad JR, Chrivia JC, Richards JP, Bachinger HP, Brennan RG, Roberts SGE, Green MR, Goodman RH: Nuclear protein CBP is a coactivator for the transcription factor CREB. Nature *370:*223, 1994.

Martin KC, Kandel ER: Cell adhesion molecules, CREB, and the formation of new synaptic connections. Neuron *17:*567, 1996.

McGue M, Bouchard TJ Jr: Genetic and environmental influences on human behavioral differences. Annu Rev Neuroscience *21:*1, 1998.

Montminy M: Transcriptional regulation by cyclic AMP. Annu Rev Biochem *66:*807, 1997.

Morgan JI, Curran T: Stimulus-transcription coupling in the nervous system: Involvement of the inducible proto-oncogenes fos and jun. Annu Rev Neurosci *14:*421, 1991.

Morgan JI, Curran T: Immediate-early genes: Ten years on. Trends Neurosci *18:*66, 1995.

*Nestler EJ, Aghajanian GK: Molecular and cellular basis of addiction. Science *35:*521, 1997.

Silva AJ, Kogan JH, Frankland PW, Kida S: CREB and memory. Annu Rev Neurosci *21:*127, 1998.

Son JH, Joh TH: Genetically engineered neural transmission. Mol Psychiatry *2:*26, 1997.

1.11 PSYCHONEUROENDOCRINOLOGY

VICTOR I. REUS, M.D. AND SYDNEY FREDERICK-OSBORNE, ꟻH.D.

The term “psychoneuroendocrinology” expresses an appreciation of inextricable structural and functional relationships between the hormonal system and the central nervous system (CNS) and the behaviors that modulate and are derived from both. Endocrine disorders are frequently associated with secondary psychiatric symptoms such as depressed mood and disturbances in thought, while a significant percentage of patients suffering from defined psychiatric syndromes display regular patterns of endocrine dysfunction. In the classic conception hormones were defined as products of endocrine glands that were transported by blood to exert their action at sites distant from their release. However, advances in neuroscience have shown that the brain not only serves as a site of regulatory control but also has secretory functions of its own, and that classic distinctions between the origin, structure, and function of nerve cells as opposed to endocrine cells are not valid.

Over the course of evolution, as organisms have increased in complexity, hormones that first appeared in unicellular organisms have been recruited to serve a multiplicity of functions, a quality referred to as *pleiotropy*. A single hormone may act at multiple sites, including binding to receptors on the membrane, cytoplasm, or nucleus, each with different effects, and subtle differences in molecular structure or metabolic processing can have profound physiological consequences. Hormones are thus ideally suited to regulate complex behavioral activities. This chapter emphasizes neuroendocrine systems that have been most directly linked to specific behavioral functions, particularly those involved in the organism's response to stress, and the principles underlying their regulation and interaction.

HORMONE SECRETION

Hormones are divided into two general classes: (1) proteins, polypeptides, and glycoproteins and (2) steroids and steroidlike compounds (Table 1.11-1) that are secreted by an endocrine gland into the bloodstream and are transported to their site(s) of action. When a hormone is colocalized and cosecreted with a neurotransmitter (e.g., norepinephrine), it may be referred to as a *neuromodulator*, although some hormones or neuromodulators have been shown to meet criteria for neurotransmitters themselves.

Structure	Examples	Storage	Lipid Soluble
Proteins, polypeptides, glycoproteins	ACTH, β-endorphin, TRH, LH, FSH	Vesicles	No
Steroids, steroid-like compounds	Cortisol, estrogen, thyroxine	Diffusion after synthesis	Yes
Functions			
Autocrine	Self-regulatory effects		
Paracrine	Local or adjacent cellular action		
Endocrine	Distant target site		

ACTH, adrenocorticotropin; TRH, thyrotropin–releasing hormone; LH, leutinizing hormone; FSH, follicle-stimulating hormone

Table 1.11-1 Classifications of Hormones

Hormone secretion is stimulated by the action of a neurohormone, a neuronal secretory product of neuroendocrine transducer cells of the hypothalamus. Neurohormones (Table 1.11-2) include corticotropin-releasing hormone (CRH), which stimulates adrenocorticotropin (adrenocorticotropic hormone [ACTH]), thyrotropin-releasing hormone (TRH), which stimulates release of thyroid-stimulating hormone (TSH), gonadotropin-releasing hormone (GnRH), which stimulates release of leutinizing hormone (LH) and follicle-stimulating hormone (FSH), and somatostatin (somatotropin release-inhibiting factor [SRIF] and growth–hormone-releasing hormone (GHRH) both of which stimulate growth hormone release. Chemical signals cause the release of these neurohormones from the median eminence of the hypothalamus into the portal hypophyseal bloodstream and their transport to the anterior pituitary to regulate the release of target hormones. Pituitary hormones in turn act directly on target cells (e.g., ACTH on the adrenal gland) or stimulate release of other hormones from peripheral endocrine organs. In addition, these hormones have feedback actions that regulate neurohormone secretion and effects in the brain itself, both directly and as modulators of neurotransmitter action (neuromodulation).

Neurohormone	Hormone Stimulated
Corticotropin-releasing hormone (CRH)	Adrenocorticotropic hormone (ACTH)
Thyrotropin-releasing hormone (TRH)	Thyroid-stimulating hormone (TSH)
	Luteinizing hormone (LH)
Gonadotropin-releasing hormone (GnRH)	Follicle-stimulating hormone (FSH)
Somatostatin (SRIF)	Growth hormone (GH)
Growth–hormone-releasing hormone (GHRH)	GH
Oxytocin	Prolactin
Arginine vasopressin (AVP)	ACTH

Table 1.11-2 Neurohormones

HORMONE SYNTHESIS AND STRUCTURE

Peptide hormones represent subsections of larger amino acid chains or polypeptides called *prohormones*. Production of a peptide hormone occurs by the cleavage of its prohormone chain at a given site on the chain by the appropriate enzyme. Proopiomelanocortin is a prohormone that contains the sequences for ACTH, b-endorphin, b-lipotropin, and a–melanocyte-stimulating hormone (MSH). Some hormones, called *dimers*, contain two or more peptide chains (e.g., FSH, LH, TSH). Further cleavage of these hormone peptide chains in the course of metabolism creates in some cases yet other biologically active peptides that may have different effects than those of the parent peptide. Even minor modifications of structure can drastically change the binding properties and metabolic processing of peptides.

CELLULAR MODE OF ACTION

Intracellular The primary mode of action of steroid hormones (glucocorticoids, estrogen, testosterone) and thyroid hormones (triiodothyronine or [T_3] and thyroxine or [T_4]) is by binding to intracellular receptors in the cytoplasm or nucleus. The hormone-receptor complex in turn binds to common response elements on chromosomal deoxyribonucleic acid (DNA) and alters transcription through a confirmational change that unmasks the binding site. The steroid hormone complex can also interact with transcription factors such as those produced by c-*fos*, c-*jun*, or activator protein-1 (AP-1) to either amplify or inhibit gene expression. By these mechanisms, hormones regulate the induction of such gene products as enzymes and other cell proteins that effect metabolic change. Based on their amino acid structure, steroid receptors are classified into two groups, with glucocorticoid, mineralocorticoid, progesterone, and androgen receptors in one class, and estrogen, thyroid, cholecalciferol (vitamin D_3), and retinoic acid receptors in the other.

Membrane Receptor Binding Alternatively, like peptides or other hormones steroids may exert a physiological effect within minutes, a time course precluding a genomic mechanism and deriving instead from binding to membrane-bound receptors and a resultant alteration in membrane permeability, ion flow, or neurotransmitter release.

CHARACTERISTICS OF ENDOCRINE ACTIVITY

In general, most hormonal compounds exert their effect in a tonic rather than phasic fashion, being diffused in a less precise manner than a neurotransmitter and over a longer time period. Theoretically, such a characterization would allow hormones to be more closely linked to integrated behavioral responses. Release of many hormones is pulsatile, and the pattern of these pulses (i.e., duration interpulse interval, slope of increase or decrease in rate, amplitude) is crucial to their effects. Other factors that can influence the effect of a hormone in a given individual include a history of exposure during critical developmental encoding periods, the frequency and chronicity of past exposure, the time since last exposure, and the status of other influences on the target system. A decrease in the amplitude of the stress response after repeated exposure is referred to as *habituation* and an enhancement is termed *sensitization*. Facilitation of a habituated response after exposure to a novel stimulus or more severe stressor is called *dishabituation*. In the case of the hypothalamic-pituitary-adrenal system, the release of cortisol by the adrenal gland is dependent on an integration of three separate control systems. These include an underlying circadian rhythm regulated by the suprachiasmatic nucleus; a stress-responsive circuit involving inputs to the hypothalamus from the brain stem, limbic system, and cerebral cortex; and a feedback control system exerted principally through two classes of corticosteroid receptor, mineralocorticoid (type I) and glucocorticoid (type II) located most importantly in the hippocampus, but also (for type II) in other brain regions as well. These receptors each have different functions: the mineralocorticoid is most densely concentrated in hippocampus and septum and regulates glucocorticoid levels at low concentration with a “high affinity-low capacity” system; the glucocorticoid is more widely distributed and binds glucocorticoids with lower affinity but higher capacity, such as would occur in stressful states.

DEVELOPMENTAL PSYCHONEUROENDOCRINOLOGY

Although a review of the effect of hormones on brain development is beyond the scope of this chapter, it is important to note that hormones can have organizational as well as activational effects. Exposure to gonadal hormones during critical stages of neural development directs changes in brain morphology and function (e.g., sex-specific behavior in adulthood). Similarly, thyroid hormones are essential for the normal development of the CNS, and thyroid deficiency during critical stages of postnatal life will severely impair growth and development of the brain, resulting in behavioral disturbances that may be permanent if replacement therapy is not instituted.

HYPOTHALAMIC-PITUITARY-ADRENAL AXIS

Since the earliest conceptions of the stress response by Hans Selye and others, investigation of hypothalamic-pituitary-adrenal function has occupied a central position in psychoendocrine research. CRH, ACTH, and cortisol are all elevated in response to a variety of physical and psychic stresses and serve as prime factors in the maintenance of homeostasis and the development of adaptive responses to novel or challenging stimuli. The hormonal response is dependent not only on the characteristics of the stressor itself, but also on how the individual assesses and is able to cope with it. Aside from generalized effects on arousal, distinct effects on sensory processing, stimulus habituation and sensitization, pain, sleep, and memory storage and retrieval have been documented. In primates, social status can influence adrenocortical profiles and in turn be affected by exogenously induced changes in hormone concentration.

Exposure to chronic stress produces increased concentrations of corticotropin-releasing factor (CRF) in the paraventricular nucleus of the hypothalamus as well as of colocalized arginine vasopressin (AVP), but to a reduction in CRF receptor number in the anterior pituitary because of downregulation over time. Release of CRF results in a simultaneous activation of the locus-ceruleus noradrenergic circuit, which functionally increases arousal and selective attention, and decreases vegetative functions such as appetite and sex drive. ACTH concentrations are increased in acute stress, but diminish over time in sustained stress, suggesting that corticosteroid receptor downregulation at the level of the hippocampus leads to an inhibition of negative feedback on CRF.

These systems are amplified, in turn, by serotonergic and cholinergic input and inhibited by g-aminobutyric acid (GABA) and opioids. Three types of inhibitory feedback of glucocorticoids on CRH and ACTH have been characterized. Fast, rate-sensitive feedback occurs while plasma concentrations of the glucocorticoid are rising, and regulates release rather than synthesis of CRH and ACTH. Intermediate, delayed feedback occurs from 1 to 12 hours following steroid administration, is dose-sensitive and duration-sensitive rather than rate-sensitive, and inhibits release of both CRH and ACTH as well as synthesis of CRH. Slow feedback is similar to intermediate and is distinguished by decreased synthesis of both CRH and ACTH (and other proopionielanocortin derivatives).

Similarities in the changes in neuroendocrine systems following chronic stress and in depressive disorders suggests that some psychiatric syndromes may not be distinct disease states per se, but rather exist on a continuum with normal functioning. In support of such a conception are a variety of studies showing that although immediate release of glucocorticoids serves homeostatic needs, more prolonged activation can result in structural neuropathology, glutamate toxicity, damage to CA1 and CA4 hippocampal neurons and speculatively, lasting behavioral change. Some of the neuropathological sequelae of normal human aging parallel such stress-induced adaptations in the hypothalamic-pituitary-adrenal system and may either produce or be secondary to changes in neuroendocrine function and stress responsivity.

Pathological alterations in hypothalamic-pituitary-adrenal function have been associated primarily with mood disorders, posttraumatic stress disorder, and dementia of the Alzheimer's type, although recent animal evidence also points toward a role of this system in substance use disorders as well. Disturbances of mood are found in more than 50 percent of patients with Cushing's syndrome (characterized by elevated cortisol concentrations), with psychosis or suicidal thought apparent in more than 10 percent of cases studied. Cognitive impairments similar to those seen in major depressive disorder principally in visual memory and higher cortical functions, are common and relate to the degree of hypercortisolemia present, and possible reduction in hippocampal size. In general, reductions in cortisol level result in a normalization of mood and mental status. Conversely, in Addison's disease (characterized by adrenal insufficiency), apathy, social withdrawal, impaired sleep, and decreased concentration frequently accompany prominent fatigue. Replacement of glucocorticoid but not of electrolyte results in resolution of behavioral symptomatology. In a similar fashion, hypothalamic-pituitary-adrenal abnormalities are reversed in individuals who are successfully treated with antidepressant medications, and these drugs have been found to stimulate corticosteroid receptor gene expression. A failure to normalize hypothalamic-pituitary-adrenal abnormalities is a poor prognostic sign. Alterations in hypothalamic-pituitary-adrenal function associated with depression include elevated cortisol concentrations, failure to suppress cortisol in response to dexamethasone, increased adrenal size and sensitivity to ACTH, a blunted ACTH response to CRH and, possibly, elevated CRH concentrations in brain. In addition to altered slow feedback, several groups have demonstrated decreased sensitivity to glucocorticoid fast-feedback as well. The pattern of these abnormalities has not thus far led to a definitive theory of mechanism. It may be that other elements such as AVP, which is synergistic with CRH, or MSH, which is synergistic with ACTH, are important to understanding the change in homeostasis in depression. For instance, it has recently been noted that individuals with depression have impaired secretion of pro-g MSH and b-endorphin, but not ACTH, following metyrapone inhibition of cortisol synthesis in individuals

with depression. The finding that corticosteroids have multiple regulatory effects on serotoninergic function, particularly on the serotonin (5-hydroxytryptamine [5-HT]) subtype 1A (5-HT_{1A}) receptor may also be relevant, as well as the state-dependent stimulant-like effects that glucocorticoids can exert on mesencephalic dopamine transmission.

ENDOGENOUS OPIOIDS

Since the discovery of endogenous opioid receptors and their endogenous ligands in the early 1970s, research into the possible behavioral roles of such compounds has grown at a rapid pace. The term “opioid” was initially introduced to differentiate endogenous opioid peptides from other exogenous opiate drugs but is now used to refer to all drugs with opioid activity. b-endorphin is the principal opioid peptide prototype and, like ACTH, MSH, and b-lipotropin, is derived from proopiomelanocortin. Methionine enkephalin (Met-entephalin) and leucine enkephalin (Leu-enkephalin) are two small pentapeptides that also possess direct opioid activity. Met-enkephalin is contained in both proopiomelanocortin and another precursor, proenkephalin, and Leu-enkephalin is contained in the prohormones proenkephalin and prodynorphin. At least three different receptor systems for these ligands have been identified (μ, d, and k), and it is likely that each of these has subtypes. The best-documented function of the endogenous opiates is in analgesia and alteration of pain perception, but effects on stress, appetite regulation, learning and memory, motor activity, and immune function have significantly expanded scientific consideration of the functional principles that govern opioid localization and release.

In animal models a number of stressors, including those that are purely psychological, induce opiate-mediated effects such as analgesia and hypomotility that are reversed by the opiate antagonist naloxone (Narcan). Several studies have found that concentrations of plasma b-endorphin in humans are correlated with measures of stress elicited by surgery, exercise, parachuting, or pain. Opioids also affect eating behavior. Most commonly, short-term administration of opioid agonists increases eating whereas antagonists reduce food intake by up to 30 percent, diminish intake of fats and highly palatable foods, and increase caloric expenditure. Thus far, however, their long-term use in obesity and eating disorders has not proven clinically useful. Some studies have found certain binge parameters (e.g., duration) to be reduced in bulimia, but no studies have demonstrated weight loss in obese subjects. Naltrexone (ReVia) has recently been demonstrated in double-blind trials to be an effective adjunct in the treatment of alcohol dependence, reducing drinking, craving, the high derived from drinking alcohol, and the likelihood that sampling alcohol would precipitate a relapse.

It is well known that exogenous opioids (e.g., heroin, morphine) can induce a euphoric mood state and that exercise increases release of endogenous opioids and is associated with mood enhancement; these observations, together with findings that exercise-induced mood enhancement is blocked by naloxone, suggest that endogenous opioids are also involved in the mediation of mood. Such conclusions must be moderated, however, by the recognition that additional specific and nonspecific effects on other neurochemical systems are possible contributors to exercise-related mood effects.

Early enthusiasm for the idea that a dysfunctional opiate system was etiologically related to mental illness, particularly schizophrenia, has waned in the face of contradictory findings. Increases in various endorphin compounds have been reported in plasma as well as in postmortem brain tissue of patients with schizophrenia but studies of short-term and long-term treatment with opioid antagonists show no consistent or reproducible effects on psychopathology. In rhesus monkeys, naloxone reduces maternal affect and social grooming.

NEUROPEPTIDE Y, GALANIN, AND INSULIN

Neuropeptide Y and the related peptide bind at three receptors, neuropeptide Y type 1 (Y_1) (postsynaptic), Y_2 (presynaptic and postsynaptic), and Y_3. Neuropeptide Y is synthesized in the arcuate nucleus of the hypothalamus, has a mutually inhibitory relation with insulin, and is stimulated by stress and corticosteroids. Neuropeptide Y has been studied for its potential anxiolytic, antinociceptive, antihypertensive, and memory-enhancing effects and for a possible role in schizophrenia and depression. Immunoreactive neuropeptide Y is found in the serotonin-containing raphe nucleus, and treatments for depression (antidepressants, lithium [Eskalith], electroconvulsive therapy [ECT]) increase neuropeptide Y concentrations in many brain areas in rats. Reduced concentrations of neuropeptide Y have been found in the brains of suicide victims and of depressed individuals who died of other causes. Significantly low levels of neuropeptide Y have also been reported in the temporal cortex of patients with schizophrenia. Neuropeptide Y has been found to increase feeding in a variety of animal studies, particularly with regard to carbohydrate ingestion; it also counteracts leptin effects on feeding. In a related function, neuropeptide Y suppresses the thermogenic activity of brown fat and increases the storage of white fat.

Galanin is an inhibitory peptide that is stimulated in a coordinated fashion with gonadal steroid release. Its documented actions include antinociception; increased release of growth hormone; memory impairment; and inhibition of insulin release, locus ceruleus noradrenergic firing, and acetylcholine release. Galanin can dramatically increase food consumption, partly because of an increased preference for fats over carbohydrates. Work on identifying receptor subtypes and developing specific antagonist ligands for them is in the early phase but holds promise for the treatment of obesity and eating disorders.

Insulin Increasing evidence indicates that insulin may be integrally involved in learning and memory. Insulin receptors occur in high density in the hippocampus and are thought to help neurons metabolize glucose. Patients with Alzheimer's disease have lower CSF insulin concentrations than controls, and both insulin and glucose dramatically improve verbal memory. Depression is frequent in patients with diabetes, as are indices of impaired hormonal response to stress. It is not known whether these findings represent direct effects of the disease or are secondary, nonspecific effects.

Hypothalamic-Pituitary-Gonadal Axis GnRH is a decapeptide that was sequenced and synthesized by Andrew Schally and colleagues in 1971. GnRH administration results in the rapid release of LH and FSH from the pituitary in healthy subjects and in some pathological states, such as acromegaly, an abnormal release of growth hormone or prolactin. The cell bodies of GnRH are located principally over the optic chiasm in the arcuate area with projections to the median eminence, and in the lamina terminalis. GnRH release is stimulated by norepinephrine and inhibited through negative feedback of gonadal steroids. Administration of GnRH can result in a depressive-like state, characterized by hot flashes, anxiety, insomnia, decreased libido, and fatigue in euthymic subjects, but it is not known whether this is a direct effect of the agent or is caused by the hypoestrogenic state that is produced when GnRH is given continuously. On this basis, a GnRH analogue has been found to have some efficacy in the treatment of paraphilia. In contrast to other neuroendocrine systems, there is relatively little evidence for a disturbance in the hypothalamic-pituitary-gonadal axis in depression; normal LH and FSH responses have been observed after even submaximal doses of GnRH.

The gonadal hormones (progesterone, androstenedione, testosterone, estradiol, and others) are steroids that are secreted principally by the ovary and testis, but significant amounts of androgens arise from the adrenal cortex as well. The prostate gland and adipose tissue are also involved in the synthesis and storage of dihydrotestosterone and contribute to individual variance in sexual function and behavior.

The timing and presence of gonadal hormones play a critical role in the development of sexual dimorphisms in the brain. Developmentally, these hormones direct the organization of many sexually dimorphic CNS structures and functions, such as the size of hypothalamic nuclei (INAH3), and corpus callosum, neuronal density in the temporal cortex, the organization of language ability, and responsivity in Broca's area. Women with congenital adrenal hyperplasia, a deficiency of the enzyme 21-hydroxylase, which leads to high exposure to adrenal androgens in prenatal and postnatal life, have in some studies been found to be more aggressive and assertive and less interested in traditional female roles than control female subjects. Sexual dimorphisms may also reflect acute and reversible actions of relative steroid concentrations (e.g., higher estrogen levels transiently increase CNS sensitivity to serotonin).

TESTOSTERONE

Testosterone is the primary androgenic steroid, having both androgenic (i.e., facilitating male gonadal development) and anabolic (i.e., facilitating linear body growth and somatic growth) functions. Testosterone is associated with increased violence and aggression in animals and in correlations studies in humans, but anecdotal reports of increased aggression with testosterone treatment have not been substantiated in human scientific investigations. Correlational reports may be confounded, however, by factors such as past history, and social factors, which are particularly important determinants of the effects of hormones in primates and in humans. For instance, in the cynomolgus monkey, administration of testosterone increases dominant behavior in dominant monkeys and submissive behavior in submissive monkeys; in hypogonadal men it improves mood and decreases irritability. Varying effects of anabolic-androgenic steroids on mood have been noted anecdotally. In the only prospective placebo-controlled study of anabolic-androgenic steroid administration in normal subjects, positive mood symptoms including euphoria, increased energy, and sexual arousal were reported, in addition to increases in negative mood symptoms of irritability, mood swings, violent feelings, anger, and hostility.

Testosterone is important for sexual desire in both men and women. In males, muscle mass and strength, sexual activity, desire, thoughts, and intensity of sexual feelings are dependent on normal testosterone levels, but these functions are not clearly augmented by supplementing testosterone in those with normal androgen levels. Addition of small amounts of testosterone to normal hormonal replacement in postmenopausal women has, however, proven to be as beneficial as its use in

hypogonadal men.

Dihydroepiandrosterone (DHEA) is an adrenal androgen and the most abundant circulating steroid. It has many physiological effects, but behavioral interest has centered on its steady decrement over the life span in humans, and its possible involvement in memory. Several controlled trials of DHEA administration point to an improvement in well-being and functional status in depressed as well as normal individuals. Its effects may result from its transformation into estrogen or testosterone or from its antiglucocorticoid activity.

ESTROGEN AND PROGESTERONE

The primary estrogens are estradiol (E_2), estrogen (E_1), and estriol (E_3), of which estradiol is the major secretory product of the ovaries. Two different estrogen receptors have been identified, each with different anatomical distribution and physiological effects. Estrogens can influence neural activity in the hypothalamus and limbic system directly through modulation of neuronal excitability and have complex multiphasic effects on nigrostriatal dopamine receptor sensitivity. Accordingly, there is evidence that the antipsychotic effect of psychiatric drugs may change over the menstrual cycle and that the risk of tardive dyskinesia is partly dependent on estrogen concentrations. Several studies have suggested that gonadal steroids modulate spatial cognition and verbal memory and are involved in impeding age-related neuronal degeneration. There is also increasing evidence that estrogen administration decreases the risk and severity of dementia of the Alzheimer's type in postmenopausal women. Estrogen has mood-enhancing properties and can also increase sensitivity to serotonin possibly by inhibiting monoamine oxidase. In animal studies long-term estrogen treatment results in a decrease in 5-HT_1 and an increase in 5-HT_2 receptors. In oophorectomized women, significant reductions in tritiated imipramine binding sites (which modulate presynaptic serotonin uptake) were restored with estrogen treatment.

Progesterone, the primary progestin, is produced by the corpus luteum of the ovary, and is a precursor for estrogens, androgens, and adrenocortical steroids. Although progesterone itself may be anxiogenic, metabolites of progesterone (allopregnanolone and pregnanolone) appear to have anxiolytic and hypnotic properties via (g-aminobutric acid [GABA] type A [$GABA_A$]) agonistic activity. Progesterone also modulates serotonin; it is colocalized with serotonin in cells of the median raphe in the macaque monkey; it causes increased serotonin uptake and turnover in several brain areas in rats.

The association of these hormones with serotonin is hypothetically relevant to mood change in premenstrual and postpartum mood disturbances. Premenstrual dysphoric disorder is a disorder in which a constellation of symptoms resembling major depressive disorder occurs in most menstrual cycles, appearing in the luteal phase and disappearing within a few days of the onset of menses. No definitive abnormalities in estrogen or progesterone levels have been demonstrated in women with premenstrual dysphoric disorder, but decreased serotonin uptake with premenstrual decreases in steroid levels have been correlated with severity of some symptoms. The rise and fall of these symptoms reflect progesterone levels, with a delay of 4 days, although evidence does not support the idea that progesterone per se is responsible for the symptoms. However, progesterone downregulates the estrogen receptor, and it has been suggested that, despite high circulating concentrations of estrogen, the luteal phase is a period of functional estrogen withdrawal, with concomitant effects on the serotonergic system. Recent evidence indicates that the abrupt decline of progesterone and allopregnanolone in the luteal phase results in increased production of the a4 subunit of the $GABA_A$ receptor and changes in receptor sensitivity that could account for the typical behavioral symptoms noted. Although the specific susceptibility that characterizes women with premenstrual dysphoric disorder remains to be described, treatments based on these hypothetical neurotransmitter effects are available. Selective serotonin reuptake inhibitors (SSRIs), particularly fluoxetine (Prozac), have demonstrated efficacy in the majority of women studied, and up to 50 percent of women may respond to fluoxetine administered only in the second half of each cycle. Alprazolam (Xanax), a $GABA_A$ agonist, is significantly more effective than placebo in women who do not respond to treatment with SSRIs. In women with severe symptoms that are not responsive to these treatments, the long-term use of a GnRH agonist to abolish menstrual cycling, with added estrogen-progestin, may be required.

The bulk of psychological symptoms associated with the menopause are actually reported during perimenopause rather than after complete cessation of menses. Although studies suggest no increased incidence of major depressive disorder, reported symptoms include worry, fatigue, crying spells, mood swings, diminished ability to cope, and diminished libido or intensity of orgasm. Hormone replacement therapy is the treatment of choice at this time. Although most advocate estrogen replacement, combination androgen-estrogen replacement therapies may be superior for reinstating energy, a sense of well-being, and libido. In women with an intact uterus, the addition of progestin is necessary to protect against endometrial hyperplasia; this can reportedly attenuate the beneficial effects of estrogen on mood, so the lowest possible progestin dosage necessary is recommended.

PROLACTIN

Since its identification in 1970, the anterior pituitary hormone prolactin has been examined as a potential index of dopamine activity, dopamine receptor sensitivity, and antipsychotic drug concentration in studies of CNS function in psychiatric patients, and as a correlate of stress responsivity. The secretion of prolactin is under direct inhibitory regulation by dopamine neurons located in the tuberoinfundibular section of the hypothalamus. Prolactin also inhibits its own secretion by means of a short-loop feedback circuit to the hypothalamus. In addition, prolactin-releasing factors have been identified, including estrogen, serotonin, opiates, TRH, and oxytocin.

Prolactin is primarily involved in reproductive functions. During maturation, prolactin secretion participates in gonadal development. In adults, prolactin contributes to the regulation of the behavioral aspects of reproduction and infant care, including estrogen-dependent sexual receptivity and breastfeeding. In female rats, prolactin secretion is strongly stimulated with exposure to pups. In women, basal prolactin levels are elevated in the postpartum period prior to weaning, and prolactin release is stimulated by suckling. In line with these functions, hyperprolactinemia is associated with low testosterone in men and reduced libido in both men and women. In rodents, prolactin is increased along with corticosterone in response to such stressful stimuli as immobilization, hypoglycemia, surgery, and cold exposure. A recent report suggests that prolactin increase may be specifically associated with the use of passive coping in the face of a stressor.

Although prolactin metabolism is not clearly altered in psychiatric disorders, hyperprolactinemic patients often complain of depression, decreased libido, stress intolerance, anxiety, and increased irritability. These behavioral symptoms usually resolve in parallel with decrements in serum prolactin when either surgical or pharmacological treatments are employed. Serum prolactin concentrations have also been positively correlated with severity of tardive dyskinesia, particularly in women who have been exposed to antipsychotic medication. Studies in psychiatric populations have attempted to use prolactin response to infusions of dopaminergic agonists as an index of central neurotransmitter activity. Thus far, the conclusions to be drawn from this strategy are not clear because widely discrepant and contradictory results have been reported.

HYPOTHALAMIC-PITUITARY-THYROID AXIS

Thyroid hormones are involved in the regulation of nearly every organ system, particularly those integral to the metabolism of food and the regulation of temperature, and are responsible for optimal development and function of all body tissues. Moreover, rates of secretion and metabolism of all other major hormones and catecholamines (cortisol, gonadal hormones, insulin) depend on thyroid status. The thyroid gland secretes two thyroid hormones: T_3 and T_4. T_3 is the more potent of the two, and most of the T_3 circulating in the blood is created by the peripheral metabolism of T_4. The brain relies on its own conversion of T_4 to T_3, rather than on circulating T_3. The hypothalamus secretes TRH into the capillaries of the pituitary portal venous system, and the pituitary responds with synthesis and secretion of TSH, which stimulates thyroid cells. Negative feedback regulation occurs when T_3 and T_4 act in the pituitary and hypothalamus to inhibit TSH and TRH respectively. Finally, a corticotropinrelease–inhibiting factor (CRIF) has recently been identified in the rat that inhibits both synthesis and secretion of ACTH. This peptide, prepro-TRH 178-199, is derived from the prohormone TRH and may play a role in integrating the regulation of the hypothalamic-pituitary-adrenal and hypothalamic-pituitary-thyroid axes.

There is general agreement that central noradrenergic systems are primarily stimulatory to TSH secretion and that central dopamine neurons inhibit TSH release. Thyroid hormones in turn are important regulators of central adrenoreceptor function, generally decreasing presynaptic noradrenaline release and increasing postsynaptic b-adrenergic receptor number. Hypothyroidism is conversely associated with decreased b-receptor number. These changes parallel the alteration in a- and b-receptor sensitivity associated with pharmacological and electroconvulsive antidepressant treatments and may explain the therapeutic efficacy of supplemental thyroid hormone in treatment-resistant depression. In addition to its prime endocrine function, TRH has direct effects on neuronal excitability, behavior, and neurotransmitter regulation, particularly on central cholinergic systems located in the septo-hippocampal band and on mesolimbic and nigrostriatal dopamine systems. In lower animals, TRH possesses mild stimulant properties. Initial reports of its mood-elevating effects in normal human subjects led to a number of projects investigating its short-term and long-term antidepressant effects in clinical populations. Despite some initial enthusiasm, the degree of mood alteration does not seem to be great nor is its occurrence reliable.

Given these observations, it is not surprising that alterations in behavioral function have been observed in patients with primary thyroid gland dysfunction, beginning with the earliest reports in the medical literature. It has been noted that thyroid disorders may induce virtually any psychiatric symptom or syndrome, although regular associations of specific syndromes and thyroid conditions are not consistently found. Hyperthyroidism is commonly associated with fatigue, irritability, insomnia, anxiety, restlessness, weight loss, and emotional lability; marked impairment in concentration and memory may also be evident. Such states can progress into delirium or mania or they can be episodic in ature. On occasion, a true psychosis develops, with paranoia being a particularly common presenting feature. In some cases psychomotor retardation, apathy, and withdrawal rather than agitation and anxiety are the presenting features. Symptoms of mania have also been reported following rapid normalization of thyroid status in hypothyroid individuals and may covary with thyroid level in individuals with episodic endocrine dysfunction. In general, behavioral abnormalities resolve with a normalization of thyroid function and are responsive symptomatically to traditional psychopharmacological regimens. Caution should be exerted, however, regarding use of antidepressant medications in hyperthyroid states because of possible synergistic cardiotoxicity. In several case reports, haloperidol (Haldol) has been linked to increasing thyrotoxicity, and hyperthyroidism has been linked to an enhancement of the neurotoxic effects of antipsychotic medications.

The psychiatric symptoms of chronic hypothyroidism are generally well recognized. Most classically, fatigue, decreased libido, memory impairment, and irritability are noted, but a true secondary psychotic disorder or dementia-like state can also develop. Suicidal ideation is common and the lethality of actual attempts is profound. In milder, subclinical states of hypothyroidism, the absence of gross signs accompanying endocrine dysfunction may result in its being overlooked as a possible cause of a mental disorder. Accordingly, the evaluation of basal TSH concentration or the TSH response to TRH infusion is necessary to arrive at the proper diagnosis. Figure 1.11-1 illustrates a characteristic physical sign of advanced hypothyroidism.

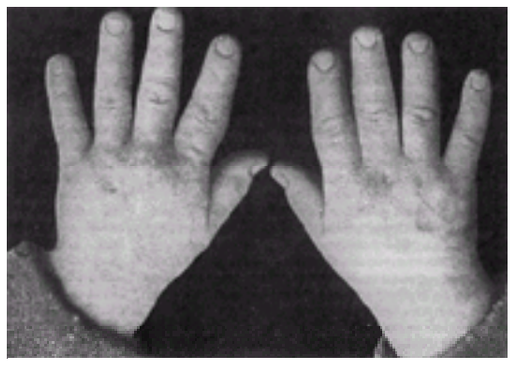

FIGURE 1.11-1 Hands of a patient suffering from hypothyroidism (myxedema), illustrating the swelling of the soft parts, the broadening of the fingers, and their consequent stumpy or pudgy appearance. (Reprinted from Waterfield RL: Anæmia. In *French's Index of Differential Diagnosis,* ed 7, AH Douthwaite, editor. Williams & Wilkins, Baltimore, 1954.)

A blunted response of TSH to TRH infusion has been found in a significant percentage of patients with a variety of disorders including eating disorders, alcoholism, and schizophrenia, and, most commonly, major depressive disorder, and probably reflects a transient hyperthyroxinemia. Large-scale studies suggest that such subjects are in fact euthyroid, and predictive sensitivity of the test is low. Antithyroid antibodies are found more frequently in women with depression than control subjects, and may contribute to relative treatment resistance. In recent years, several investigators have presented evidence that attention-deficit/hyperactivity disorder may be associated with a generalized resistance to thyroid hormone due to a genetic mutation of the b-thyroid receptor.

Most antidepressant therapies have some influence on thyroid concentrations at baseline; T_4 and T_3 concentrations have been correlated with antidepressant response, as have antidepressant-induced changes in thyroid hormones as well as changes induced by ECT. Lithium increases antithyroid antibodies, and inhibits iodine uptake into the thyroid, iodination of tyrosine, release of T_3 and T_4 from the thyroid, and peripheral breakdown of thyroid hormones. It also blocks the thyroid-stimulating effects of TSH through interference with adenylate cyclase and may, in certain circumstances, precipitate a rebound thyrotoxicosis. About 30 percent of patients receiving lithium will have an elevated TSH level during treatment, and approximately one sixth of these will go on to develop frank hypothyroidism. Attention to subtle alteration in thyroid status induced by lithium treatment is important in the clinical evaluation of symptomatic complaints such as fatigue, memory impairment, and anhedonia. Carbamazapine (Tegretol), an anticonvulsant shown to have antimanic properties akin to lithium, also decreases peripheral thyroid hormone concentrations while increasing TSH. Administration of T_3 is sometimes helpful in patients with treatment-resistant depression, while adjunctive T_4 contributes to decreasing cycling in patients with rapid-cycling bipolar I disorder.

PARATHYROID HORMONE

Parathyroid hormone (PTH) was originally isolated as an endocrine factor having effects on bone, gut, and kidney and contributing to calcium and phosphorus homeostasis. However, the frequent and often profound neuropsychiatric changes that can result from altered parathyroid gland function are consistent with other central actions of PTH that have been described in recent years.

Although PTH is known to increase calcium uptake leading to enhanced neurotransmitter release, PTH administration can impair the active uptake and release of both norepinephrine and dopamine through a variety of other actions. Consequences of exogenously administered PTH include adrenergic-like effects (not blocked by b-adrenergic antagonist), learning and memory problems, and a state of hyperalgesia.

Lithium treatment can raise the concentrations of serum calcium and may increase PTH. When such effects are associated with somatic or behavioral change, discontinuation of lithium should result in rapid symptomatic improvement. When this does not happen, a parathyroid adenoma is sometimes discovered fortuitously. Primary hyperparathyroidism most commonly occurs secondary to a single parathyroid adenoma, the removal of which almost invariably results in a lysis of behavioral symptoms, regardless of severity or chronicity. Animal studies suggest that long-term lithium administration can stimulate the development of extant parathyroid tumors but does not induce tumors in normal parathyroid tissue. Thus, reinstitution of lithium treatment after surgical removal of the tumor should be possible.

GROWTH HORMONE

Most psychiatric studies of the regulation of somatotropin or growth hormone (GH) have utilized strategies similar to those described for prolactin. Accordingly, studies of GH response to various provocative stimuli, such as to GHRH or psychotherapeutic drugs, have been seen as a means to evaluate central neurotransmitter function. Augmentation of GH secretion in response to GHRH, LH-releasing hormone (LHRH), and TRH in patients with schizophrenia or dementia of the Alzheimer's type has been interpreted as reflecting an alteration in catecholamine and possibly prostaglandin regulation, which facilitate the secretion of human GH. In general, however, there is a large variation in GH response to GHRH; a blunted response has been variably linked to length of illness, presence of negative symptoms, and platelet monoamine oxidase activity, but the validity of the conclusions drawn from this test is controversial.

The stress responsivity of somatotrophs is well established but species dependent, with increases in circulating GH noted in humans and inhibition of secretion noted in rodents. Case reports have documented reversible GH deficiencies and marked growth retardation and delay of puberty secondary to stressful experience. Administration of GH to individuals with GH deficiency has a beneficial effect on cognitive function in addition to its more obvious somatic effects, but there is evidence for poor psychosocial adaptation in adulthood for children who were treated for GH deficiency. Nearly one third of patients with major depressive disorder and dysthymic disorder may have a GH deficiency. Some prepubertal as well as adult patients with diagnoses of major depressive disorder show hyposecretion of GH-releasing hormone during an insulin tolerance test, a deficit that has been interpreted as reflecting alterations in both cholinergic and serotonergic mechanism. A number of GH abnormalities have been noted in patients with anorexia nervosa. Secondary factors such as weight loss, however, in both major depressive disorder and eating disorders, may be responsible for such alterations in endocrine release. Nonetheless, at least one study has reported that GHRH stimulates food consumption in patients with anorexia nervosa and attenuates elevated food consumption in patients with bulimia. Administration of GH to elderly men results in an increase in lean body mass and improved vigor.

SOMATOSTATIN

SRIF is a hypothalamic tetradecapeptide that is located principally in the nerve endings of the median eminence and in neurosecretory neurons located in the paraventricular nucleus. SRIF inhibits anterior pituitary secretion of ACTH, thyrotropin, GH, and prolactin and alters release of catecholamine neurotransmitters. Five receptor subtypes have been cloned and receptor-specific ligands developed. SRIF was so named because of its action in inhibiting the release of immunoreactive GH, a function that is subserved by somatostatin-2 receptors.

In rats, SRIF delays the extinction of active avoidance behavior and antagonizes amnesia induced by electric shock. Alterations in the concentration of SRIF have been associated with a number of conditions in which cognitive dysfunction is present, including Huntington's disease, Parkinson's disease, multiple sclerosis, and particularly Alzheimer's Disease. Decreases in SRIF are highly correlated with decreases in acetylcholinesterase, suggesting a close relationship between the cholinergic and somatostatinergic systems. Decreased concentrations of SRIF in CSF are inconsistently found in patients with depression, and central injection of SRIF in rats causes decreased slow-wave and rapid eye movement (REM) sleep, altered appetite and locomotor activity, impaired cognition, and decreased sensitivity to pain. Early stressful experiences have also been related to sustained elevations of CRF and somatostatin in the CSF of adult primates.

CHOLECYSTOKININ

In addition to its presence in pancreas and the gastrointestinal tract, cholecystokinin (CCK) has been identified in mammalian brain, with high concentrations found in the cerebral cortex, limbic system, and hypothalamus. In animal studies CCK is involved in the regulation of such behavioral functions as inhibition of intake of both solid and liquid food, production of satiety, and pain relief. Of the two identified receptor subtypes, CCK type A (CCK-A) is found primarily in the periphery and in some discrete brain areas, whereas CCK-B is plentiful in the brain. The primary form of CCK, CCK-8S (a sulphated octapeptide), coexists with dopamine in the ventral tegmental area and substantia nigra, and its interactions with dopamine are to be context- and location-specific. Both CCK-A and CCK-B receptors have a high affinity for CCK-8S. CCK-B receptors also have high affinity for CCK-8US (unsulphated), CCK-4, and pentagastrin, whereas CCK-A receptors do not.

Of particular interest to psychiatry is the coexistence of CCK with dopamine in mesolimbic and mesocortical but not in nigrostriatal systems. This suggests that CCK-8 activity might either be dysregulated in psychiatric syndromes thought to involve altered dopamine transmission or that CCK-8 analogues could be used therapeutically in the treatment of these syndromes. CCK-A receptor antagonists has been proposed for the treatment of schizophrenia, and initial evidence suggests that CCK agonists may be useful for decreasing the severity of parkinsonian symptoms. Evidence is stronger for a role of CCK-B receptor antagonists in the treatment of anxiety. Several small-scale human studies have demonstrated that administration of CCK-4 or pentagastrin can induce panic attacks in a significant percentage of normal volunteers. Panic attacks in subjects with preexisting panic disorder can be elicited at doses of CCK-4 that do not reliably induce panic in normal subjects, indicating the enhanced sensitivity of patients with panic disorder to CCK-4. Not only can selective CCK-B antagonists completely abolish the anxiogenic effects of CCK-4, but in animal models of anxiety used to evaluate the efficacy of benzodiazepines CCK-B antagonists also demonstrate anxiolytic properties.

MELATONIN

Melatonin is a pineal hormone that is derived from the serotonin molecule and controls photoperiodically mediated endocrine events (particularly those of the hypothalamic-pituitary-gonadal axis). It also modulates immune function, mood, and reproductive performance and is a potent antioxidant and free-radical scavenger. Melatonin has a depressive effect on CNS excitability and is an analgesic and has seizure-inhibiting effects in animal studies. These effects may occur via a GABA-potentiating action, although they may be secondary to the antioxidant or antiexcitotoxic actions of melatonin. Although some researchers have found low melatonin in depression, confounding factors exist, and suppression of melatonin is not necessary for the efficacy of light therapy in seasonal affective disorder. However, melatonin can be a useful therapeutic agent in the treatment of circadian phase disorders such as jet lag. Intake of melatonin increases the speed of falling asleep, as well as its duration and quality.

ARGININE VASOPRESSIN

AVP or (antidiuretic hormone [ADH]) is a posterior pituitary hormone that maintains plasma osmolarity through regulation of renal water excretion. AVP release is triggered by pain, emotional stress, dehydration, increased plasma osmolarity, or decreases in blood volume.

Profound alterations in fluid ingestion and excretion have been observed in psychiatric patients throughout most of the twentieth century. Polydipsia occurs in 10 to 15 percent of hospitalized psychiatric patients and is unrelated to diagnosis. In many cases the syndrome is secondary to inappropriate secretion of AVP, which occurs as a feature of the altered behavioral state itself and resolves with treatment or, conversely, is precipitated by a variety of antidepressant or antipsychotic agents.

Animal and normal human studies of AVP administration (or longer-acting synthetic analogue compounds) have indicated that the hormone may enhance both the consolidation and retrieval of memory, particularly that associated with aversive learning. Clinical studies in assessments of attention, concentration, and memory in the elderly, those with depression, or those with dementia have had more mixed results. Although positive results exist, the effect is small and has yet to be consistently reproduced; it may be that such reported effects are secondary to a general arousal effect. However, AVP has also been shown to prevent the loss of tolerance to alcohol and to delay the loss of sexual behavior following castration, both of which suggest involvement of a learning mechanism.

Altered AVP function has been reported in depression and in eating disorders. Anorexic patients show hypersecretion of centrally directed AVP, and patients with bulimia nervosa or depression may have an attenuated AVP response to hypertonic saline. An inverse relation between AVP concentration and motor activity in depression and an increased number of vasopressin and oxytocin neurons have also been reported in the hypothalamus of depressed patients.

OXYTOCIN

Oxytocin is also a posterior pituitary hormone that is involved in osmoregulation, the milk ejection reflex, food intake, and female maternal and sexual behaviors. The majority of studies of oxytocin have so far been conducted in animals. Convergent evidence using a range of methodologies indicates that oxytocin inhibits food and sodium intake. Oxytocin binding in the hypothalamus is increased by estrogen and glucocorticoids, and in estrogen-primed females, oxytocin has anxiolytic activity and promotes the initiation of maternal behaviors. Oxytocin induces a variety of reproductive (grooming, arousal, lordosis, orgasm, nesting, birthing) and maternal behaviors (breast-feeding, mother-infant bonding), although the former may be restricted to nonmonogamous species. Infusion of oxytocin in females of monogamous species facilitates pair bonding in the absence of mating, and administration of an oxytocin antagonist prevents pair bonding. AVP serves a similar pair-bonding function in males, and promotes monogamy and paternal behavior in male prairie moles. Oxytocin can also act as a neuromodulator of limbic dopamine concentrations and thus may be involved in adaptation to substances of abuse.

SUBSTANCE P

Substance P, an 11 amino acid peptide discovered originally in 1930, has neurotrophic effects and acts as an excitatory transmitter in primary afferent nerve terminals in mammalian spinal cord, helping to regulate sympathetic noradrenergic function. Depending on the pain paradigm utilized, administration of substance P can produce either hyperalgesia or analgesia. Substance P also has been found to have memory-promoting and reinforcing effects, which may be produced by the N- and C-terminal fragments respectively. Although substance P has been implicated in the pathogenesis of several neuropsychiatric disorders, evidence for any specific role is mixed.

NEUROTENSIN

Neurotensin is a tridecapeptide in the substantia nigra, limbic system, and hypothalamus that has a close neuroanatomical relation with dopaminergic pathways. It is involved in the control of anterior pituitary activity and affects release of prolactin and TSH at both hypothalamic and pituitary concentrations. Several lines of evidence suggest that neurotensin may have a possible antipsychotic role. Psychotogenic drugs (e.g., methamphetamine [Desoxyn]) inhibit the release of neurotensin via an inhibitory effect of the dopamine type 1 (D_1) receptor. All antipsychotic drugs increase neurotensin concentrations in the nucleus accumbens and dopamine receptor antagonists also increase concentrations in the caudate nucleus. Central administration of neurotensin produces a range of pharmacological effects similar to those produced by antipsychotic drugs in animals, and such effects may be specific to the mesolimbic-cortical and not the nigrostriatal-dopamine system.

Neurotensin may exert its antipsychotic action through intramembrane receptor interactions that reduce affinity of the D_2-agonist binding site.

MELANOCYTE-STIMULATING HORMONE

Melanocyte-stimulating hormone (MSH), an anterior pituitary peptide, controls the secretion of melatonin (and melanin) and, in some paradigms, exerts opposite effects on behavior. Intraperitoneal administration of a-MSH delays extinction in a passive-avoidance paradigm and increases emotional response. In a double-blind crossover trial in humans, an infusion of a-MSH resulted in a significant improvement in verbal memory but little change in mood. Because phenothiazines increase pituitary MSH secretion and pigmentation in patients in proportion to their therapeutic potency, it has been suggested that MSH peptides may possess some therapeutic properties. A dose-related biphasic effect on mood has also been reported for MSH-release–inhibiting factor. Recent data indicate that MSH interacts with leptin to counteract Neuropeptide Y, decrease food intake, and increase energy expenditure.

SUGGESTED CROSS-REFERENCES

Section 1.12 on the immune system contains information on the interaction between endocrine, immune, and neural systems. Section 2.5 provides more detail on the psychiatric effects of endocrine disorders. Chapter 14 discusses further aspects of mood disorders and section 31.28 presents the therapeutic use of thyroid hormones.

SECTION REFERENCES

*Arborelius L, Owens MJ, Plotsky PM, Nemeroff CB: The role of corticotropin-releasing factor in depression and anxiety disorders. J Endocrinol *160:*1, 1999.

Baker DG, West SA, Oath DN, Hill KK, Nicholson WE, Ekhator NN, Bruce AB, Wortman MD, Keck PE Jr, Geracioti TD Jr: Cerebrospinal fluid and plasma betaendorphin in combat veterans with post-traumatic stress disorder. Psychoneuroendocrinology *22:*517, 1997.

Barenbaum SA, Resnick SM: Early androgen effects on aggression in children and adults with congenital adrenal hyperplasia. Psychoneuroendocrinology *22:*505, 1997.

Berr C, Lafont S, Debuire B, Dartigues J-F, Baulieu E-E: Relationships of dehydroepiandrosterone sulfate in the elderly with functional, psychological, and mental status, and short-term mortality: A French community-based study. Proc Natl Acad Sci USA *93:*13410, 1996.

*Besedovsky HO, Rey AD: Immune-neuro-endocrine interactions: Facts and hypotheses. Endoc Rev *17:*64, 1996.

*Brzezinski A: Melatonin in humans. N Engl J Med *336:*186, 1997.

Burleson MH, Malarkey WB, Cacioppo JT, Poehlmann KM, Kiecolt-Glaser JK, Berntson GG, Glaser R: Postmenopausal hormone replacement: Effects on autonomic, neuroendocrine, and immune reactivity to brief psychological stressors. Psychosom Med *60:*17, 1998.

*Coplan JD, Andrews MW, Rosenblum LA, Owens MJ, Friedman S, Gorman JM, Nemeroff CB: Persistent elevations of cerebrospinal fluid concentrations of corticotropin-releasing factor in adult nonhuman primates exposed to early-life stressors: Implications for pathophysiology of mood and anxiety disorders. Proc Natl Acad Sci *93:*1619, 1996.

Deijen JB, de Boer H, van der Veen EA: Cognitive changes during growth hormone replacement in adult men. Psychoneuroendocrinology *23:*45, 1998.

De Souza EB: Corticotropin-releasing factor receptors: Physiology, pharmacology, biochemistry and role in central nervous system and immune disorders. Psychoneuroendocrinology *20:*789, 1995.

Deuschle M, Blum WF, Strasburger CJ, Schweiger U, Weber B, Korner A, Standhardt H, Gotthardt U, Schmider J, Pflaum C-D, Heuser I: Insulin-like growth factor-I (IGF-I) plasma concentrations are increased in depressed patients. Psychoneuroendocrinology *22:*493, 1997.

Gudmundsson A, Carnes M: Pulsatile adrenocorticotropic hormone: An overview. Biol Psychiatry *41:*342, 1997.

*Insel TR, O'Brien DJ, Leckman JF: Oxytocin, vasopressin, and autism: Is there a connection? Biol Psychiatry *45:*145, 1999.

Kawata M: Roles of steroid hormones and their receptors in structural organization in the nervous system. Neurosci Res *24:*1, 1995.

Lamberts SWJ, van den Beld AW, van der Lely A-J: The endocrinology of aging. Science *278:*419, 1997.

*Liu D, Diorio J, Tannenbaum B, Caldji C, Francis D, Freedman A, Sharma S, Pearson D, Plotsky PM, Meaney MJ: Maternal care, hippocampal glucocorticoid receptors, and hypothalamic-pituitary-adrenal responses to stress. Science *277:*1659, 1997.

Lopez JF, Chalmers DT, Little KY, Watson SJ: Regulation of serotonin $_{1A}$, glucocorticoid and mineralocorticoid receptor in rat and human hippocampus: Implications for the neurobiology of depression. Biol Psychiatry *43:*547, 1998.

*Lyons DM, Wang OJ, Lindley SE, Levin S, Kalin NH, Schatzberg A: Separation induced changes in squirrel monkey hypothalamic-pituitary-adrenal physiology resemble aspects of hypercortisolism in humans. Psychoneuroendocrinology *24:*131, 1999.

Marangell LB, George MS, Callahan AM, Ketter T, Pazzaglia PJ, L'Herrou TA, Leverich GS, Post RM: Effects of intrathecal thyrotropin-releasing hormone (protirelin) in refractory depressed patients. Arch Gen Psychiatry *54:*214, 1997.

McCarty R, Gold PE: Catecholamines, stress, and disease: A psychobiological perspective. Psychosom Med *58:*590, 1996.

McLay RN, Freeman SM, Harlan RE, Ide CR, Kastin AJ, Zadina JE: Aging in the hippocampus: Interrelated actions of neurotrophins and glucocorticoids. Neurosci Biobehav Rev *21:*615, 1997.

Mongeau R, Marsden CA: Effect of central and peripheral administrations of cholecystokinin-tetrapeptide on panic-like reactions induced by stimulation of the dorsal periaqueductal gray area in the rat. Biol Psychiatry *43:*335, 1997.

Nelson JC, Davis JM: DST studies in psychotic depression: A meta-analysis. Am J Psychiatry *154:*1497, 1997.

O'Brien JT: The ‘glucocorticoid cascade' hypothesis in man. Br J Psychiatry *170:*199, 1997.

Pietrowsky R, Thiemann A, Kern W, Fehm HL, Born J: A nose-brain pathway for psychotropic peptides: Evidence from a brain evoked potential study with cholecystokinin. Psychoneuroendocrinology *21:*559, 1996.

Rosch PJ: Stress and memory loss: Some speculations and solutions. Stress Med *13:*1, 1997.

Sapolsky RM: Stress, glucocorticoids, and damage to the nervous system: The current state of confusion. Stress *1:*1, 1996.

*Seeman MV: Psychopathology in women and men: Focus on female hormones. Am J Psychiatry *154:*1641, 1997.

Toren P, Dor J, Rehavi M, Weizman A: Hypothalamic-pituitary-ovarian axis and mood. Biol Psychiatry *40:*1051, 1996.

Wang C, Alexander G, Berman N, Salehian B, Davison T, McDonald V, Steiner B, Hull L, Callegari C, Swerdloff RS: Testosterone replacement therapy improves mood in hypogonadal men—a clinical research center study. J Clin Endocrinol Metab *81:*3578, 1996.

Wise PM, Krajnak KM, Kashon ML: Menopause: The aging of multiple pacemakers. Science *273:*67, 1996.

Yaffe K, Sawaya G, Lieberburg I, Grady D: Estrogen therapy in postmenopausal women. JAMA *279:*688, 1998.

*Yates WR, Perry PJ, MacIndoe J, Holman T, Ellingrod V: Psychosexual effects of three doses of testosterone cycling in normal men. Biol Psychiatry *45:*254, 1999.

Zweifel JE, O'Brien WH: A meta-analysis of the effect of hormone replacement therapy upon depressed mood. Psychoneuroendocrinology *22:*189, 1997.

1.12 IMMUNE SYSTEM AND CENTRAL NERVOUS SYSTEM INTERACTIONS

ANDREW H. MILLER, M.D., BRADLEY D. PEARCE, PH.D. AND CARMINE M. PARIANTE, M.D.

An ever-growing database demonstrates that interactions between the immune system and the central nervous system (CNS) play a critical role in the maintenance of bodily homeostasis and the development of diseases, including psychiatric disease. Alterations in CNS function brought about by a variety of stressors have been clearly shown to influence both the immune system as well as diseases that involve the immune system. Moreover, many of the relevant hormonal and neurotransmitter pathways that mediate these effects have been elucidated. Of considerable interest is recent data outlining the impact of soluble factors (cytokines) derived from immune cells and glia, which in turn have profound effects on the CNS. The relative roles of these factors in the various psychiatric diseases is an area of active investigation, as is the role of infectious and autoimmune diseases in the pathophysiology of psychiatric disorders. Taken together, these findings highlight the importance of interdisciplinary efforts involving the neurosciences and immunology for gaining new insights into the etiology of psychiatric syndromes.

OVERVIEW OF IMMUNE SYSTEM

The immune system has the capacity to protect the body from the invasion of foreign pathogens, such as viruses, bacteria, fungi, and parasites. In addition, the immune system can detect and eliminate cells that have become neoplastically transformed. These functions are accomplished through highly specific receptors on immune cells for molecules derived from invading organisms and a rich intercellular communication network that involves direct cell-to-cell interactions and signaling between cells of the immune system by soluble factors called *cytokines*. The body's absolute dependence on the efficient functioning of the immune system is illustrated by the less than 1-year survival rate of untreated infants born with severe combined immunodeficiency disease and the devastating opportunistic infections and cancers that arise during acquired immune deficiency syndrome (AIDS).

CELLS AND TISSUES

The immune system must be able to survey all tissues of the body for the presence of infectious agents or neoplastic cells and to mobilize its effector components to specific sites in the body where infectious agents may invade. Therefore, an important requirement of the immune system is that it be systemic and mobile. Cells of hematopoietic origin largely accomplish this function. Like all other blood cells, immune cells are derived from hematopoietic precursor stem cells, which in the adult originate in the bone marrow. The stem cells are pluripotent and are capable of differentiating into any one of the various mature hematopoietic cells. There are two major paths of differentiation that are regulated in part by cytokines and other factors (Fig. 1.12-1). The lymphoid path leads to the formation of the mature lymphocytes—B cells, T cells, and natural killer cells; the myeloid path of differentiation leads to the other cells of the blood (some of which participate in the immune response), including monocytes and granulocytes, which include neutrophils, eosinophils, and basophils. Monocytes and basophils may further differentiate into macrophages and mast cells, respectively, which take up residence in tissues throughout the body.

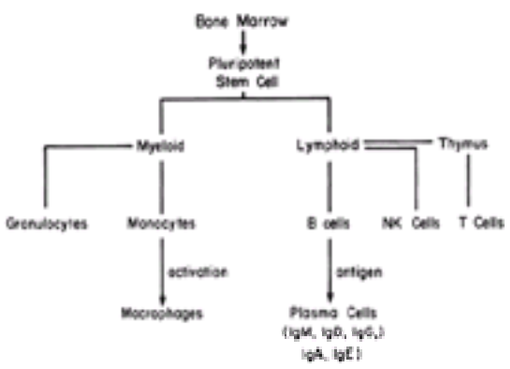

FIGURE 1.12-1 All cells are ultimately derived from a pluripotent stem cell of bone marrow, which gives rise to myeloid or lymphoid progenitors that undergo further differentiation to mature cell types indicated in the schematic diagram. NK = natural killer; Ig = immunoglobulin. (Reprinted with permission from Norin AJ: Introduction to immunobiologic concepts. In *Depressive Disorders and Immunity*, AH Miller, editor. American Psychiatric Press, Washington, DC, 1989.)

Lymphocyte maturation occurs in *primary immune tissues*. In humans the bone marrow serves as the primary site for B-cell maturation, and the thymus is the primary site for T-cell maturation. An important part of the maturation process is the screening out of cells that are reactive to the body's own constituents *(self-reactive)*. On maturation, lymphocytes exit the primary immune tissues and circulate through the bloodstream and the lymphatic system into and out of the secondary immune tissues, including the spleen and widely distributed lymph nodes. *Secondary immune tissues* provide a structure for interactions between different immune cells and circulating pathogens.

NATURAL AND ACQUIRED IMMUNITY

The immune system is often divided on a functional basis into two separate categories: natural or innate immunity and specific or acquired immunity (Table 1.12-1). The components of natural immunity act in a relatively nonspecific manner against pathogens or infected cells and may be evolutionarily more primitive than the specialized T and B lymphocytes that mediate specific immunity; operationally, however, the two modes of immunity interact and cooperate.

Feature	Natural	Specific (Acquired)
Physicochemical barriers	Skin, mucous membranes	Cutaneous and mucosal immune systems; antibody in mucosal secretions
Circulating molecules	Complement	Antibodies
Cells	Phagocytes (macrophages, neutrophils), natural killer cells	Lymphocytes
Soluble mediators active on other cells	Macrophage-derived cytokines, (e.g., interferon alpha and beta, tumor necrosis factor)	Lymphocyte-derived cytokines, (e.g., interferon gamma)

Reprinted with permission from Abbas AK, Lichtman AH, Pober JS: Cellular and Molecular Immunology, ed 2. Saunders, Philadelphia, 1994.

Table 1.12-1 Features of Natural and Specific (Acquired) Immunity

Natural Immunity The cells mediating natural immunity do not require prior activation to be functional; they therefore provide an important first line of defense against infectious agents during the early stages of an immune response. Mononuclear phagocytic cells and natural killer cells are examples of immune cells that mediate nonspecific immunity. Mononuclear phagocytic cells, such as macrophages, microglia, certain endothelial cells, and reticular cells of lymphoid organs, are collectively referred to as the *reticuloendothelial system*, and these cells destroy extracellular pathogens (e.g., bacteria and parasites) by engulfing and degrading them. Mononuclear phagocytes also release type I interferons (IFN), which have direct antiviral properties, and proinflammatory cytokines including tumor necrosis factor (TNF), interleukin (IL-1), and IL-6. TNF is the principal mediator of the response to gram-negative bacteria and is one of the earliest cytokines released in the proinflammatory cascade that includes IL-1 followed by IL-6. TNF is an endogenous pyrogen that along with IL-1 is capable of inducing fever by increasing the synthesis of prostaglandins by cytokine-stimulated hypothalamic cells. TNF also leads to cachexia, characterized by wasting of muscle and fat cells, in part secondary to appetite suppression.

The combination of liver-derived plasma proteins induced by TNF and IL-1 with those induced by IL-6 constitutes the acute-phase response (Fig. 1.12-2). The acute-phase response is designed to limit tissue damage, isolate and destroy invading organisms, and set repair functions in motion. These objectives are achieved by rapid changes in plasma protein composition characterized by increases in the so-called “acute phase reactants”: C-reactive protein (which coats bacteria to facilitate phagocytosis-opsonization), macroglobulin and other antiproteases (which neutralize tissue destructive proteases), and the clotting protein fibrinogen. Albumin and transferrin, the iron transport protein, decline during the acute-phase response and are therefore called *negative acute phase reactants*.

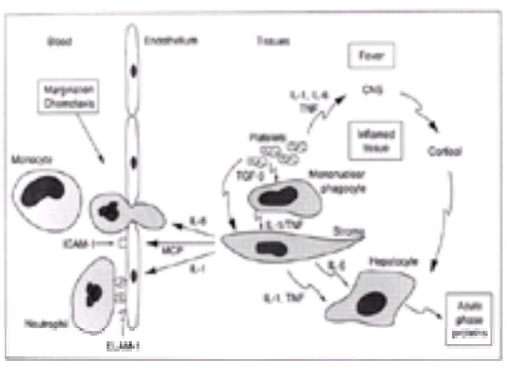

FIGURE 1.12-2 Scheme for cell and cytokine interaction in the acute phase response. Mononuclear phagocytes release early cytokines, such as IL-1 and TNF, at the site of tissue damage. These cytokines activate adjacent stroma and endothelium to elicit other chemotactic cytokines and initiate the accumulation of inflammatory cells. The hepatic response is activated by other cytokine mediators, as is the activation of the adrenal-pituitary axis to release ACTH and subsequently cortisol. (Reprinted with permission from Baumann H, Gauldie J: The acute phase response. Immunol Today *15:*74, 1994.)

Complement factor proteins, which are produced by the liver, provide another important humoral component of nonspecific immunity. These functionally linked proteins interact with one another in a highly regulated manner and subserve many of the effector functions of the immune system, including cell lysis, opsonization, activation of inflammation by attracting inflammatory cells (chemotaxis) and stimulating immune cells to release chemical mediators of inflammation, and neutralization of antigen-antibody complexes that can damage tissues.

Natural killer cells are also an important component of natural immunity. These cells can destroy virally infected cells by binding to them and releasing cytolytic factors, including perforin. Natural killer cells also have the ability to recognize and destroy neoplastically transformed host cells, especially those of hematopoietic origin, thus providing protection against some cancers.

Acquired Immunity T and B lymphocytes are the crowning achievement of immune cell specialization and evolution. These cells account for both the diversity and the specificity of the immune response and for the adaptive aspect of the immune system. Furthermore, T and B cells are responsible for directing the immune response against foreign targets, rather than self components. An effective specific immune response includes three conceptually separate phases: an induction phase, in which the presence of an infectious agent or antigen is detected; an activation phase, which includes the proliferation and the mobilization of the immune cells relevant to the eradication of the infectious agent; and an action or effector phase, in which the infectious agent is neutralized and eliminated.

INDUCTION PHASE Recognition of pathogens or neoplastically transformed cells is achieved through specialized receptors for antigens on the surface of B and T lymphocytes. Antigens are foreign substances that induce specific immunity and typically include molecules derived from pathogens, such as viral subunits, enzymes, and bacterial cell wall glycoproteins. The B-cell antigen receptor is a membrane-bound form of immunoglobulin. A related form of immunoglobulin is secreted as antibody by mature B cells (plasma B cells). Antibodies play a central role in humoral immunity and help to kill a variety of pathogens. Each immunoglobulin or antibody molecule has two identical antigen-binding sites, and those binding sites can recognize the tertiary structure of specific proteins and other molecules, such as polysaccharides and lipids, which are important components of infectious agents (Fig. 1.12-3).

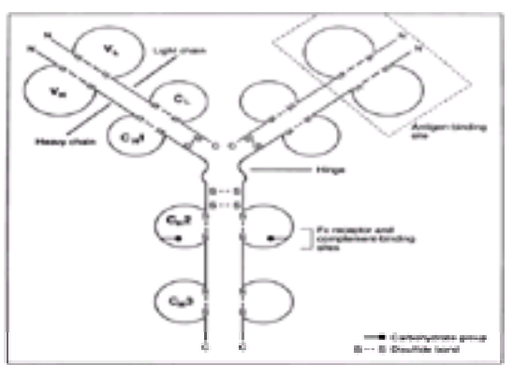

FIGURE 1.12-3 Schematic diagram of an immunoglobulin molecule. In this drawing of an IgG molecule, the antigen binding sites are formed by the juxtaposition of V_L and V_H domains. The locations of complement and Fc receptor-binding sites within the heavy chain constant regions are approximations. S--S refers to intrachain and interchain disulfide bonds; N and C refer to amino and carboxy termini of the polypeptide chains, respectively. (Reprinted with permission from Abbas AK, Lichtman AH, Pober JS: *Cellular and Molecular Immunology*, ed 2. Saunders, Philadelphia, 1994.)

The T-cell antigen receptor has a single antigen-binding site. Antigen recognition by T cells takes on an added level of complexity not inherent to B cells. The T-cell receptor recognizes only fragments of protein antigens. In addition, the antigen fragments must be present in association with a class of cell surface molecules called *major histocompatibility (MHC) molecules* (Fig. 1.12-4). Virtually all nucleated cells of the body express MHC molecules on their surface. Most cells express class I MHC molecules; some cells, usually of immune system origin, also express class II MHC molecules. The induction of a T-cell response depends on the ability and the effectiveness of MHC molecules to bind and present antigen. Therefore, the repertoire of MHC molecule genes that a person inherits can contribute significantly to antigen presentation and ultimately to susceptibility to infectious diseases and autoimmune disorders. T cells are MHC class-restricted—that is, T-cell receptors recognize antigen only in association with one or the other class of MHC molecules. All MHC class I-restricted T cells also express an invariant surface glycoprotein, referred to as *clusters of differentiation (CD) 8*. All MHC class II-restricted T cells express a different invariant surface glycoprotein, referred to as *CD4*. CD8 and CD4 molecules assist the binding of T cells to antigen-MHC complexes and assist in subsequent T-cell activation. Most $CD8^+$ T cells are cytolytic T lymphocytes (CTLs) that have the ability to lyse cells to which they bind, whereas most $CD4^+$ T cells are helper T cells that secrete cytokines on activation. Antigen-presenting cells (APCs), including macrophages, B cells, dendritic cells, Langerhans' cells of the skin and human endothelial cells, all have the ability to engulf, degrade, and process extracellular proteins and then display the processed bits of protein in conjunction with surface class II MHC molecules (Fig. 1.12-4). Therefore, these cells play an important role in presenting antigen to $CD4^+$ T cells.

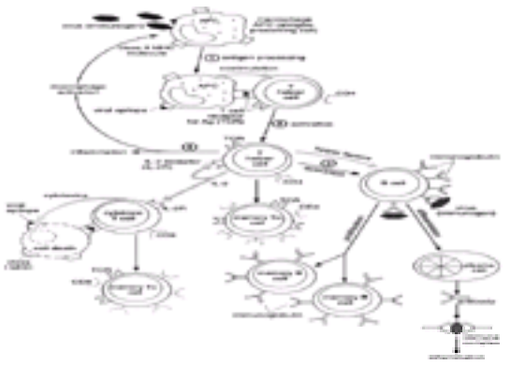

FIGURE 1.12-4 Sequence of events in a prototypical acquired immune response. Antigen presenting cells (APCs) present processed immunogen to helper T cells, which are central to the development of acquired immune responses. Through their T cell receptors (TCRs), T cells recognize particular epitopes of the immunogen in association with a Major Histocompatibility Complex (MHC) molecule. T helper cells in turn can help B cells make antibody and activate other effector cells including cytotoxic T cells, T and B memory cells, and natural killer cells, macrophages, granulocytes and antibody dependent cytotoxic (K) cells (not pictured). (Reprinted with permission from Stites DP, Terr AI, Parslow TG: *Medical Immunology*, ed 9. Appleton & Lange, Stamford, CT, 1997.)

All the receptors for an antigen on a particular B or T cell are identical and unique to that cell and its decedents *(clones)*. A family of lymphocytes with identical antigen receptor specificity is called a *clonal line*. Diversity in antigen recognition is derived from the vast number of different B-cell and T-cell clonal lines present in each person. Such diversity is achieved by semirandom mixing and matching of sequences of deoxyribonucleic acid (DNA) that code for the antigen-binding portion of the T-cell and B-cell receptors. That process of genetic diversity has been estimated to have the capacity to generate in each person more than 10^8 different receptors with functionally distinct antigenic specificity. Such high diversity makes it likely that each person possesses some clonal line of lymphocytes with an antigen receptor specificity capable of binding to a portion of any pathogen that may be encountered. Thus, the specific recognition of pathogens by the immune system entails the clonal selection of lymphocytes that are specifically responsive to the infectious agent.

ACTIVATION PHASE The binding of foreign antigens by B cells and T cells is usually not sufficient to produce cell activation; an accessory signal must also be provided. Important accessory signals are generated by a group of cytokines called *interleukins* that are secreted by T helper cells and antigen presenting cells, such as macrophages. T helper cells and APCs cooperate (Fig. 1.12-3); APCs secrete IL-1 and other cytokines that stimulate T helper cells to secrete a host of cytokines including interferon gamma (IFN-g), which then increases the phagocytic ability of APCs and increases their class II MHC expression, thus improving their antigen-presenting capacity. IFN-gamma also has direct antiviral properties. In addition, IL-1 stimulates T cells to produce IL-2 and express IL-2 receptors on their surface; IL-2 is an important cytokine that activates multiple lymphocyte functions (Table 1.12-2).

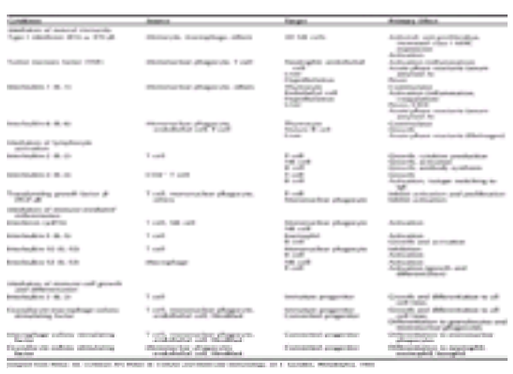

Table 1.12-2 Cytokines and Their Effects

Recent evidence indicates that two subclasses of T helper cells secrete different cytokine profiles after stimulation. These T helper subsets have been best characterized in the mouse. T helper-1 (Th-1) cells primarily secrete IL-2 and IFN-g and are involved in cell-mediated inflammatory reactions and result in activation of CTL. T helper-2 (Th-2) cells primarily secrete IL-4, IL-5, IL-6, and IL-10, and encourage antibody formation, particularly immunoglobulin E (IgE) responses. Thus, Th-2 cytokines are found in association with strong antibody responses and allergic responses. Various cytokines and their effects are listed in Table 1.12-2.

After binding antigen in the presence of stimulatory cytokines, T and B lymphocytes with the appropriate binding sites are activated, leading to cell growth, division, and proliferation. Activation also results in the clonal expansion of immune cells with the identical high-affinity specificity for the foreign antigen. Some of the progeny during clonal expansion undergo further differentiation into mature effector cells, such as antibody-secreting plasma B cells, and CTLs. By contrast, some descendents of activated B cells or T cells become memory cells that are primed for activation on further stimulation by the same antigen. Reexposure to that antigen results in a secondary immune response *(acquired immunity)*, which is usually more rapid and more robust than the first or primary immune response to that antigen. Memory cells may live for many years, thus providing long-lasting acquired immunity, as is evident in individuals who received a vaccine or had their first contact with a specific infectious agent during infancy.

EFFECTOR PHASE The ultimate aim of an immune response is the neutralization and elimination of pathogens. The principal effector mechanisms of acquired immunity are mediated by antibodies *(humoral immunity)* secreted from B cells, and by cytolytic T cells *(cellular immunity)*. Humoral immunity is especially effective in combating extracellular pathogens, such as bacteria and parasites; cellular immunity is effective in protecting against viral infection and, as with natural killer cells, may provide some protection against tumor cells.

The effector components of natural immunity are also recruited, enhanced, and directed toward specific pathogens as a result of the actions of B and T cells. For example, circulating antibodies can neutralize pathogens by binding to and coating the pathogens *(opsonization)*. Pathogens that are opsonized are made susceptible to lysis by complement factors and phagocytosis (Fig. 1.12-2). Natural killer cells and *phagocytic cells*, such as neutrophils and macrophages, have receptors for the

Fc fragment of antibodies. Furthermore, complement proteins bind to and are activated by the Fc fragments of some types of antibodies. Thus, antibodies can link effector cells and cytolytic proteins of natural immunity with pathogens, lending a level of specificity that is not inherent in the effector processes themselves. Although a large diversity of antigen-binding domains are present in different antibodies, the remainder of the molecule is highly conserved, and there are a limited number of different types of Fc fragments *(isotypes)* of the molecule. Because the Fc fragment of the antibody is the effector portion of the molecule, antibodies with different isotypes have different effector features (Table 1.12-3).

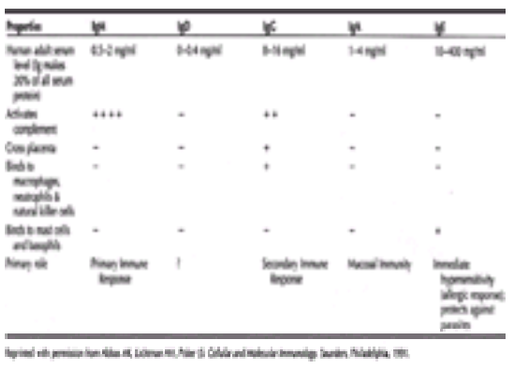

Table 1.12-3 Antibody Isotypes and Their Features

IMMUNE SYSTEM AND DISEASE

The general effectiveness of the immune system in protecting the body against pathogens has been made dramatically clear by the extensive pathology that characterizes AIDS in persons infected by the human immunodeficiency virus (HIV). HIV selectively binds to the CD4 molecule on T-helper cells via the gp120 protein on its membrane envelope and thereby gains entry and inhibits T-helper cell function. Since T-helper cells play a critical role in facilitating all aspects of specific immunity, the incapacitation of T-helper cells by HIV has catastrophic effects on the immune system. AIDS patients become susceptible to a wide spectrum of pathogens, such as protozoa *(Pneumocystis)*, bacteria *(Mycobacterium tuberculosis)*, fungi *(Candida)*, and viruses (herpes simplex). Furthermore, AIDS patients have a high incidence of malignant tumors, especially those known to result from virally induced cellular proliferation and transformation. The nervous system is also affected in many AIDS patients, as evidenced by memory loss and other nonspecific neuropsychiatric disorders. No evidence indicates that HIV directly infects neurons; however, the infection of macrophages in neural tissues may lead to the impairment of neuronal function through the release of cytokines.

At the other end of the spectrum from immunodeficiency is autoimmunity. A number of relatively common diseases—such as type I diabetes, rheumatoid arthritis, and systemic lupus erythematosus—have been shown to result from a specific autoimmune response directed against self-antigenic components. Clear genetic links to the expression of autoimmune disorders are often associated with specific types of MHC molecules. In most cases, however, a genetic background is not sufficient for the expression of disease. For example, the much greater prevalence of rheumatoid arthritis and systemic lupus erythematosus in women than in men suggests that, at least in some cases, there may be a hormonal component to the expression of these diseases.

The extent to which the immune system provides protection against cancer is still undetermined. Several effector mechanisms of the immune system are capable of destroying tumor cells in vitro (natural killer cells, CTLs, and TNF). The relatively rare occurrence of nonvirally induced tumors in immunodeficient patients suggests that the immune system plays a role primarily in protecting against tumor-inducing viruses, rather than providing widespread tumor surveillance and elimination. Nevertheless, the extensive use of a variety of cytokines and other immune modulators in the treatment of neoplastic diseases underlines the importance of these factors in cancer therapy.

METHODS USED IN STUDYING THE IMMUNE SYSTEM

In Vitro Assays Much of the understanding of the immune system has been derived from in vitro studies. In vitro assays may be especially useful in dissecting the direct and indirect mechanisms by which neurally controlled factors can influence immune cell function. Two of the most widely used assays in the study of neural-immune interactions assess the proliferative capacity of lymphocytes and determine the cytolytic capability of CTL and natural killer cells.

For proliferative assays, mononuclear cells, including lymphocytes, are removed from the experimental subject and are challenged in vitro with a mitogenic stimulus. Commonly used mitogenic stimuli (including concanavalin A, phytohemagglutinin, pokeweed mitogen, and lipopolysaccharide) are glycoproteins derived from plant lectins or bacterial cell walls that have been found to polyclonally stimulate lymphocyte proliferation. Proliferation is monitored by the incorporation of 3H-thymidine into the DNA of the dividing cells. The limitations of proliferative assays are their notorious interassay variability and a limited understanding of the relationship between the polyclonal proliferative response to a mitogen and the clonally selective proliferative response to a specific antigen or pathogen.

NK cell assays have been widely applied to studies of neural-immune interactions. Typically, immune cells isolated from a subject are incubated in vitro with chromium 51-labeled target cells. Lysis of target cells results in the release of chromium 51 into the incubation medium, which is then collected and measured.

Flow Cytometry The development of monoclonal antibodies against specific immune cell surface markers, such as the various CD determinants, has been useful in monitoring and sorting subclasses of immune cells (Table 1.12-4). Fluorescently tagged monoclonal antibodies and the cells to which they bind can be detected by a laser-controlled flow cytometer. Clinically, flow cytometry has important applications in monitoring the proportion of subsets of immune cells in patients' peripheral blood. For example, a diagnostic feature of the onset of AIDS is the precipitous decline in the proportion of circulating $CD4^+$ cells. Experimentally, flow cytometry may be useful in studying the effects of various treatments and environmental factors on the proportion or number of immune cell subpopulations present in the various immune compartments. However, changes in the number and the percentage of a given subset are independent phenomena and may involve different mechanisms: for example, the increased percentage of one subset may actually be related to a decrease in the percentage of other subsets of lymphocytes.

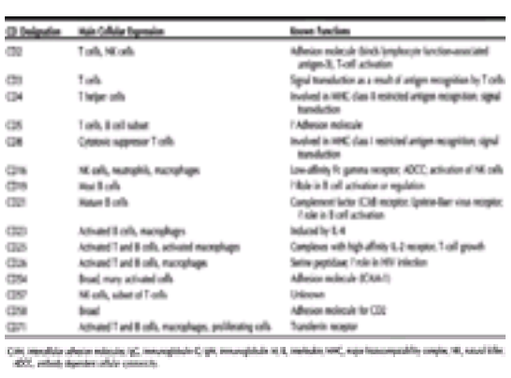

Table 1.12-4 Some Defined Clusters of Differentiation (CD) Molecules and Their Known Functions

In Vivo Assays Results from in vitro assays can be difficult to interpret because the relation between changes in these assays and the capability of the immune system to exert effective responses in vivo is still unclear. To address this issue, studies exploring the immune system in humans (e.g., evaluating the consequences of stressors) have used at least three different assays for the examination of the immune system in vivo: (1) the antibody response to antigen challenge such as a vaccination, (2) antibody titers to latent viruses, and (3) cutaneous delayed-type hypersensitivity. Examining the antibody response to a vaccine allows evaluation of all phases of the immune response including antigen presentation, B- and T-cell cooperation, and humoral immunity. Alterations (notably increases) in antibody titers

to latent viruses (herpes viruses in particular) has been considered a marker of viral reactivation secondary to impairment in cellular immunity. Finally, cutaneous delayed-type hypersensitivity is based on the intradermal injection of one or more protein antigens (e.g., from tetanus, diphtheria, streptococcus, tuberculin, or candida). The antigens induce a rapid accumulation of neutrophils followed by a more persistent infiltration of T cells and monocytes, and the reaction is interpreted by measuring the induration at 48 hours.

REGULATION OF THE IMMUNE RESPONSE

An effective immune response requires the cooperation of many components of the immune system, often resulting in the augmentation of each component's contribution to the overall immune response. However, the simultaneous indiscriminate amplification of all aspects of the immune system would not be efficient and could even be disastrous. An overactive immune system may contribute to autoimmunity; furthermore, the inflammatory component of immune responses can be damaging if not controlled, as is seen in immune complex diseases and septic shock. Therefore, regulation of the immune response is necessary to make sure that the response is energy efficient, focused on the infectious agent, counterbalanced in a fashion that does not cause self-damage, and reversible once the pathogen has been eliminated.

Probably the most important form of *intrinsic regulation* of the immune system is mediated by the various cytokines. Several examples of the facilitatory effects of cytokines have been cited; however, some cytokines, such as transforming growth factor b (TGF-b), also produce potent inhibitory effects on lymphocyte activation and proliferation (Table 1.12-2). In addition, there is clear evidence for the existence of a subclass of CD4+ and CD8+ T cells (T suppressor cells) that act primarily to suppress the function of other T cells, either by secreting inhibitory cytokines or by cytolysis. However, the isolation of a unique subset of T cells that have a predominant suppressor action has been elusive. Another important mode of intrinsic regulation results from the production of antibodies or T cells that bind to determinants (idiotypes) in the antigen-binding domain of other antibodies or T-cell antigen receptors and serve to influence (inhibit) further antigen-antibody interactions.

The relative significance of *extrinsic regulation* of the immune response remains to be fully established. However, the increasing evidence of neural-immune interactions indicates that extrinsic factors of central nervous system origin plays an important role in the modulation of the immune system.

EVIDENCE OF NERVOUS SYSTEM AND IMMUNE SYSTEM INTERACTIONS

Immune Cell Receptors As outlined in Table 1.12-5, cells from the immune system express receptors for a wide variety of molecules that are, in part, regulated by or derived from the nervous system. One of the first receptors to be characterized in lymphocytes was the b-adrenergic receptor. Subsequently, receptors for the other classic neurotransmitters have been described. As in the nervous system, receptors for the neurotransmitters in immune cells are located in the cell membrane and in most cases are coupled to G proteins and associated second-messenger pathways. Nevertheless, the biochemical mechanisms for the receptor-mediated activity of the majority of the molecules listed in Table 1.12-5 has yet to be fully elucidated.

Neurotransmitters	Neuropeptides	Steroid Hormones
β-adrenergic	Growth hormone	Glucocorticoids
α-adrenergic	Prolactin	Mineralocorticoids
Dopamine	Corticotropin	Estrogens
Serotonin	CRH	Progesterones
Acetylcholine	Substance P	Androgens
Histamine	Somatostatin	
Vasoactive intestinal peptide		
Endorphins and enkephalins		
Vasopressin	Thyroid hormone	

CRH, corticotropin releasing hormone.

Table 1.12-5 Receptors for Chemical Messengers in Cells and Tissues of the Immune System

Several important concepts from research on receptors in immune cells and tissues are central to understanding the effects of neurally derived molecules on immune function. First, the expression of receptors is heterogeneous. For example, of the two types of receptors for adrenal steroids, mineralocorticoid receptors and glucocorticoid receptors, only glucocorticoid receptors are expressed in the thymus, whereas both glucocorticoid and mineralocorticoid receptors are expressed in the spleen. Related to heterogeneity in receptor expression in immune cells and tissues is heterogeneity in receptor density. For example, of the subsets of T cells, the number of b-adrenergic receptors is highest for T suppressor cells (CD8+, CD28–), followed by T cytotoxic cells (CD8+, CD28+) and then T helper cells (CD4+).

Heterogeneity of both receptor expression and receptor density is relevant for understanding the sensitivities of the various immune cells and tissues to circulating hormones and is important for determining the net effect of these agents on immune function in vivo. Thus, even though b-adrenergic receptor agonists generally inhibit lymphocyte function in vitro, the increased susceptibility of T suppressor cells to the inhibitory effect (secondary to increased receptor density) may paradoxically lead to an enhancement of the immune response in vivo.

Another important concept is that a change in circulating concentrations of a hormone or transmitter is not necessarily reflected equally in all immune compartments. For example, stress-related increases in glucocorticoids are more effective in activating glucocorticoid receptors in the peripheral blood and the thymus than in the spleen. Thus, the microenvironment of any given tissue is critical in determining hormonal or neurotransmitter influences on immune function. Taken together with the heterogeneity in receptor density and expression, the data demonstrate that the influence of any given molecule on the immune system is a function of (1) the type of cell that exhibits the relevant receptor, (2) the density of the receptors on that cell, and (3) whether that cell is located in an immune compartment that allows access of the relevant molecule to the receptor under the conditions being studied.

Cross talk between receptor-associated second-messenger systems is another important mechanism by which neurally derived or regulated molecules influence the immune response (Fig. 1.12-5). For example, dual signaling through the b-adrenergic and T cell receptors can lead to a synergistic rise in cyclic adenosine monophosphate (cAMP), which may alter the early events in T cell activation. Changes in these early events may then influence the transcription of multiple genes, including the genes for IL-2 and the IL-2 receptor, as well as genes for other cytokines.

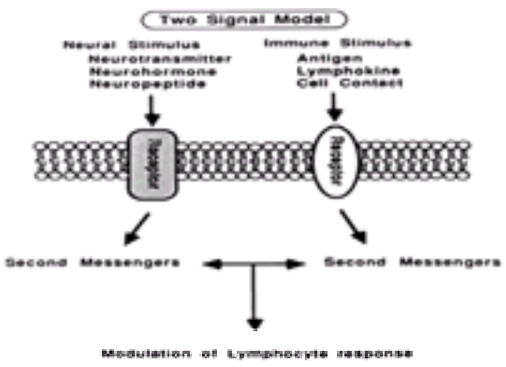

FIGURE 1.12-5 Schematic representation of two signal model of the interaction of neurally and immune-derived substances with their respective receptors on lymphocytes. The figure shows the potential crosstalk between second messengers with resulting modulation of lymphocyte function. These models assume that the lymphocytes are in G_0 of the cell cycle. (Reprinted with permission from Rozman TL, Carlson SL: Neural-immune interactions: Circuits and networks. Prog Neuroendocrin Immunol *4:*22, 1991.)

IN VITRO AND IN VIVO EFFECTS

Numerous chemical messengers derived from or regulated by the nervous system are capable of altering immune cell function and distribution. Table 1.12-6 provides a necessarily simplified, representative listing of selected molecules and their effects. The immunological effects of the agents depend on a number of factors aside from those involving the relevant expression, density, and activation of receptors on target immune cells. For example, the effect of any given molecule on the immune system depends on the phase of the immune response that is involved. As previously noted, the major phases of an immune response include the induction phase, the activation and proliferation phase, and the effector phase. Norepinephrine, for example, promotes immune function during the induction phase, both potentiates and inhibits immune function during the activation and proliferation phase, and inhibits the effector phase (Fig. 1.12-6). The potentiation of the proliferative phase occurs at low concentrations of norepinephrine, but inhibition occurs at high norepinephrine concentrations. These findings indicate that both the timing of exposure as well as the dose are important.

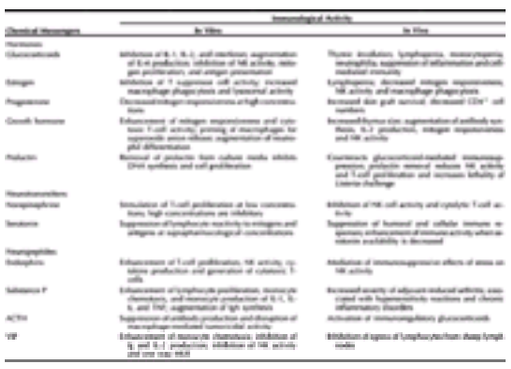

Table 1.12-6 Representative Listing of Chemical Messengers and Their Immunological Effects

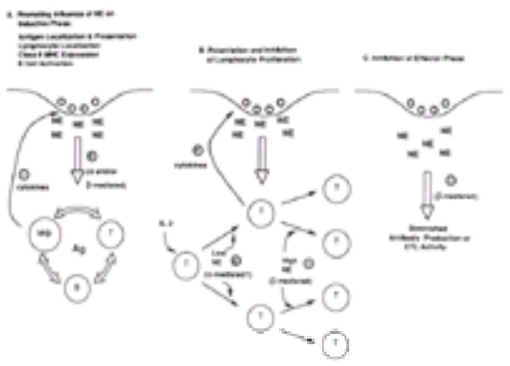

FIGURE 1.12-6 Working model for bidirectional communication between norepinephrine and cells of the immune system. This diagram depicts the potential modulatory effects of norepinephrine (NE) during different phases of an immune response. The capacity for cellular responsiveness to NE will depend on the presence of adrenoceptors, on cell type, and on the activational state. (A) In the presence of antigen, NE may have an enabling influence on lymphocyte and accessory cell function. To achieve the proper balance between a- and b-mediated effects, regulation of NE levels may be critical. (B) Low levels of NE may stimulate a-adrenoceptors and thus potentiate lymphocyte proliferation. (B,C) With increasing NE, b-adrenoceptor-mediated inhibition of lymphocyte proliferation and effector function may occur. The response to NE may be influenced by, and alter the production of cytokines (such as IL-2) and neuroendocrine peptides (not shown for clarity). Finally, products of the immune system may modulate NE availability by altering release or turnover. Ag, antigen; MO, macrophage. (Reprinted with permission from Madden KS, Livnat S: Catecholamine action and immunologic reactivity. In *Psychoneuroimmunology*, ed 2, R Ader, DL Felten, N Cohen, editors, Academic Press, New York, 1991.)

Issues of timing are also relevant in terms of development and aging. In aged rats, for example, a progressive loss of noradrenergic innervation of the spleen is accompanied by a progressive increase in the density of b-receptors on splenic lymphocytes. However, there is also an age-related dysfunction that involves impaired coupling between the b-receptor and adenylate cyclase, indicating that noradrenergic agents may have variable, unpredictable effects on immune function in old animals.

Related to the phase of the immune response and developmental stage of the animal is the type of immune response as it relates to pathophysiology. Substances that are primarily inhibitory to immune function may promote tumor development in animals with cancer but may attenuate the development of autoimmune disease. For example, the administration of glucocorticoids accelerates the growth of tumors in mice. But glucocorticoids inhibit the development of several types of autoimmune disorders, including experimental allergic encephalitis (a model of multiple sclerosis) and streptococcal cell wall-induced polyarthritis (a model of rheumatoid arthritis). Another important factor in determining the immunological effect of a particular molecule is its indirect effects, as well as its direct effects, on the immune system. In vitro studies provide important information on the direct effects of the various chemical messengers, but the influence of those agents in vivo may be completely different. For example, a number of in vitro studies have shown that opioid peptides are capable of enhancing natural killer cell activity. However, in vivo, opioid peptides play an important role in mediating the inhibitory effects of shock stress on natural killer cell activity, most likely through effects in the brain. In vivo, neurally derived molecules act against a complicated background of multiple hormones that may have synergistic or antagonist effects or both. For example, prolactin antagonizes the effects of glucocorticoids on spleen mitogen responses in mice. Furthermore, many of the hormones and transmitters influence other bodily systems, including the cardiovascular system, which may influence the traffic of immune cells to various organs, immunological and otherwise. Changes in immunocyte distribution may ultimately have effects on cellular function.

NEURAL INNERVATION OF LYMPHOID TISSUES

Historically, the identification of nerve fibers derived from the sympathetic nervous system in immune tissues was one of the first indications that communication between the CNS and the immune system was possible. Sympathetic nerve fibers have been identified in organs that are responsible for the development, education, and function of lymphocytes. Specifically, nerve fibers are found in the bone marrow, thymus, spleen, and lymph nodes. The nerves that innervate the thymus gland are derived from the vagus, phrenic, and recurrent laryngeal nerves and from the stellate and other small ganglia of the thoracic sympathetic chain. The nonmyelinated nerves that innervate the bone marrow arise from the level of the spinal cord associated with the location of the bone. The spleen obtains its sympathetic nerves from the celiac ganglion. Autonomic nervous system innervation of the lymph nodes is not as dense or as uniquely distributed as noted for the spleen and the thymus. In general, sympathetic nerve fibers enter the lymphoid tissues in association with the vascular supply. Because those nerves play an important role in vascular tone, their presence in association with the smooth muscle cells of the blood vessels is not unexpected. However, the nerve fibers travel with small blood vessels devoid of smooth muscle cells and are present in the parenchyma of the lymphoid tissue, associated not only with blood vessels but with lymphocytes and other immune cells (Fig. 1.12-7). The existence of noradrenergic nerve terminals in the parenchyma of lymphoid tissues suggests that the release of norepinephrine in these areas may interact with the neurotransmitter receptors on nearby cells and ultimately influence their function.

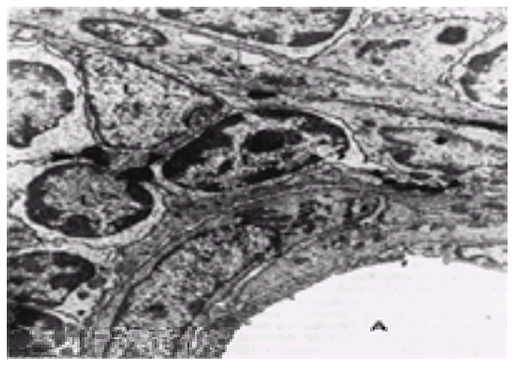

FIGURE 1.12-7 Tyrosine hydroxylase-immunoreactive nerve processes (small arrowheads) in contact with the smooth muscle (S) of the central arteriole (A), and nerve processes (large arrowheads) in direct contact with lymphocytes (L) in the periarteriolar lymphatic sheath of the rat spleen. Transmission electron micrograph, X6732. (Courtesy of Denise L. Bellinger, Department of Neurology and Anatomy, University of Rochester School of Medicine and Dentistry, Rochester, NY.)

Since catecholamines (including norepinephrine) have stimulatory effects on immune function at low concentrations and inhibitory effects at high concentrations, differential sympathetic nervous system effects on the immune response are possible, depending on local concentrations of catecholamines and the location of the immune cell relative to the point of neurotransmitter release. Chemical sympathectomy has variable effects on immune function, depending in part on the phase of the immune response studied. The reported effects of sympathectomy include suppressed antibody responses to sheep red blood cells, suppressed cytolytic T-cell activity, and enhanced natural killer-cell activity. Splenic sympathectomy also leads to an upregulation of b-adrenergic receptors on lymphocytes and a decrease in suppressor lymphocyte function. Aside from the phase of the immune response, other factors that may influence the effects of nervous innervation on immune function include the animal's age, sex, and strain.

Colocalization of neurotransmitters with a variety of neuropeptides—including neuropeptide Y, substance P, vasoactive intestinal peptide, and calcitonin gene-related peptide—has also been identified in nerves innervating immune tissues. The distribution of these peptides have not been extensively studied, and their relative significance has yet to be determined.

Finally, data suggests that local immune responses within the microenvironment of the immune tissues may be able to interact directly with nerve fibers through the effects of cytokines on neurotransmitter release. Neurotransmitters may in turn regulate immune function or the sensitivity of cells to local immunomodulatory factors. Although earlier studies have focused on efferent nerve fibers, more attention is now being paid to sensory afferent fibers and their relevance to immune system communication with the brain. For example, sensory afferent fibers appear to relay immune signals through the vagus nerve, to the nucleus of the solitary tract. From the nucleus of the solitary, there are pathways to other brain regions that may be relevant to the CNS response to immune system activation.

CNS LESIONS AND CELLULAR IMMUNITY

Some of the earliest studies demonstrating CNS involvement in the regulation of immune phenomena were those involving lesions in specific areas of the brain in laboratory animals. In a series of experiments in the early 1960s, lesions of the anterior hypothalamus were found to protect guinea pigs from death by anaphylactic shock when compared with sham and control lesions. Such protection was not apparent with median and posterior hypothalamic lesions. Lesions of the anterior hypothalamus were also accompanied by decreases in thymus and spleen cell number, splenic mitogen responsiveness, antigen responsiveness, and natural killer cell activity. In contrast, bilateral electrolytic lesions of the hippocampus and the amygdala have been associated with increases in thymic and splenic mitogen responsiveness and no change in thymus and spleen cell number. Left-sided neocortical lesions have been associated with decreases in spleen T-cell numbers, T-cell mitogen proliferation, T-cell cytotoxicity (measured in mixed lymphocyte cultures), and natural killer cell activity; these immune parameters were either unchanged or enhanced by right-sided lesions. B-cell and macrophage responses were not affected by right-sided or left-sided neocortical ablations.

Some investigators have reported an increased prevalence in humans of immune disorders, especially those involving the thyroid and the gastrointestinal tract, in association with atypical cerebral dominance (left-handedness). The relation between cerebral dominance and immune dysfunction in humans lends support to the notion that lateralized alterations in CNS function may have differential effects on the immune response. However, some patients with brain tumors and head injuries have gross disturbances of in vitro lymphocyte function, and there is no evidence that the immune changes are related to the laterality of the lesion.

BEHAVIORAL CONDITIONING

Demonstration that learning processes are capable of influencing immunological function is another example of interactions between the immune system and the nervous system. Several classical conditioning paradigms have been associated with the suppression or enhancement of the immune response in various experimental designs. The conditioning of immunological reactivity provides further evidence that the CNS can have significant immunomodulatory effects.

Some of the first experiments on immunological conditioning were derived from the serendipitous observation that animals undergoing extinction in a taste-aversion paradigm with cyclophosphamide, an immunosuppressive agent, had unexpected mortality. In that taste-aversion paradigm, the animals were simultaneously exposed to an oral saccharin solution (the conditioned stimulus) and an intraperitoneal injection of cyclophosphamide (unconditioned stimulus). Since the animals experienced considerable physical discomfort from the cyclophosphamide injection, through the process of conditioning they began to associate the ill effects of cyclophosphamide with the taste of the oral saccharin solution. If given a choice, the animals avoided the saccharin solution (taste aversion). Conditioned avoidance can be eliminated or extinguished if the saccharin is repeatedly presented in the absence of cyclophosphamide. However, it was observed that animals undergoing extinction of cyclophosphamide-induced taste aversion unexpectedly died, leading to the speculation that the oral saccharin solution had a specific conditioned association with the immunosuppressive effects of cyclophosphamide. Repeated exposure to the saccharin-associated conditioned immunosuppression during extinction might explain the unexpected death of animals. To test that hypothesis the researchers conditioned the animals with saccharin (conditioned stimulous) and intraperitoneal cyclophosphamide (unconditioned conditioned stimulous) and then immunized them with sheep red blood cells. At different times after immunization the conditioned animals were reexposed to saccharin (conditioned stimulous) and examined. The conditioned animals exhibited a significant decrease in mean antibody titers to sheep red blood cells as compared with the control animals. Thus, the evidence demonstrated that immunosuppression of humoral immunity was occurring in response to the conditioned stimulous of saccharin alone.

Because the immunological effects of conditioned immunosuppression were not large, the influence of immunological conditioning on the development of a spontaneously occurring autoimmune disease in New Zealand mice was investigated. These animals provide a standard model for the study of systemic lupus erythematosus, a fatal autoimmune disorder that is similar to that found in humans. Death in the New Zealand mice can be delayed by weekly injections of cyclophosphamide. In the initial studies the animals were first conditioned with saccharin and cyclophosphamide and then divided into three groups: (1) saccharin only (conditioned stimulous group), (2) saccharin and cyclophosphamide (conditioned stimulous plus unconditioned conditioned group), and (3) no treatment. As shown in Figure 1.12-8, animals given saccharin alone had a mortality rate as low as the animals receiving saccharin plus weekly injections of cyclophosphamide; these findings supported the notion that conditioned immunosuppression was occurring in response to saccharin alone, and the effects were of sufficient magnitude to powerfully influence disease expression.

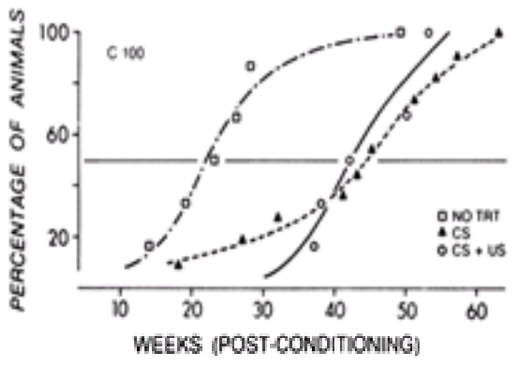

FIGURE 1.12-8 Mortality rate in NZB × NZW F_1 female mice treated with saccharin and CY weekly and then continued on a regimen of saccharin and CY (group CS + US, N = 6), continued on saccharin alone (group SD, N = 11), or deprived of both saccharin and CY (no treatment, N = 6). (Reprinted with permission from Ader R: Behaviorally conditioned modulation of immunity. In *Neural Modulation of Immunity*, R Guillemin, M Cohen, T Melnechuk, editors. Raven, New York, 1985.)

The ability to condition immunosuppression using T–cell–independent antigens and a graph-versus-host response (T cells present in transplanted bone marrow attack the host) has indicated that conditioned immunosuppression generalizes to both humoral and cell-mediated immunity. Furthermore, conditioned enhancement of natural killer cell activity in response to the conditioned stimulous, camphor, has been found after repeated pairing of the immunostimulant poly I:C with camphor odor. Finally, Studies of conditioning of immune responses have been expanded to include demonstrations that environmental stimuli, such as those inherent in passive avoidance paradigms, can be associated with conditioned immunosuppression.

STRESS AND THE IMMUNE RESPONSE

Interest in the effects of stress on the immune system grew out of a series of animal and human studies that suggested that stressful stimuli can influence the development of immune-related disorders including infections, cancer, and autoimmune diseases. Experiments conducted on laboratory animals in the late 1950s and the early 1960s, for example, indicated that a wide variety of stressors—including isolation, rotation, crowding, exposure to a predator, and electric shock—increased morbidity and mortality in response to several types of tumors and infectious diseases caused by viruses and parasites. Critical variables in the effects of stress on illness in animals involved the timing of the stressor application and the type of infectious agent or tumor. For example, mice subjected to electric grid shock 1 to 3 days before the injection of Maloney murine sarcoma virus-induced tumor cells exhibited a decreased tumor size and incidence. In contrast, mice exposed to grid shock 2 days after tumor cell injection exhibited an increase in tumor size and number.

Fewer studies have been carried out on the relation between stress and immune-related illnesses in humans, and in general they are difficult to interpret because of the many factors that can influence illness and illness behavior. Nevertheless, from prospective studies on upper respiratory infections verified either by physician diagnosis or by biological methods, evidence indicates that stressful life events can increase the susceptibility to infectious diseases in humans. For example, investigators have found that infection rates by five separate rhinoviruses administered intranasally are significantly greater in persons experiencing a high degree of psychological stress than in those under low stress. Prospective studies in humans on the development of cancer are mixed. Some studies have indicated a relation between depressive symptoms (presumably secondary to increased stress and inability to cope) and cancer development; others have been unable to replicate these findings. Once cancer has developed, however, data on women with metastatic breast cancer indicate that supportive group therapy may increase the time of survival by more than 1 year, although group therapy did not alter the mortality rates in the patients. Similar findings of increased survival have been found in patients diagnosed with malignant melanoma who were provided with a 6-week structured psychiatric group intervention.

In an attempt to understand the potential mechanisms of the effect of stress on illness involving the immune system, researchers have focused considerable attention on the effects of stress on a variety of immune parameters in laboratory animals and humans.

Animal Studies Early studies using 19 hours of tail shock have shown that intense stress inhibits mitogen-induced lymphocyte proliferation in the peripheral blood of rats. Subsequent studies have indicated that this stressor also inhibits natural killer cell activity and the production of interferon and IL-2 in the blood and spleen of rats. Since the early studies on tail shock, there has been an explosion of reports on the effects of many types of stress on virtually every aspect of immune function; in general, the stressors inhibit immune responses (Table 1.12-7). However, several studies using either chronic stressors or mild or brief stressors have noted an enhancement of both cellular and humoral aspects of the immune system. For example, in a study of mice exposed to 0.3, 0.8, or 1.2 mA of electric foot shock over a 1-hour period, there was a graded enhancement of the proliferative response of T cells to the mitogen concanavalin A as the shock level increased. Interestingly, proliferative assays using the B-cell mitogen poke weed did not show any significant effect of the shock treatments, suggesting that the type of immune response may dictate the relative influence of stress. Differential effects of stress on various measures of the immune system have been further demonstrated in a study in which physical restraint in mice significantly depressed the cellular immune response to influenza virus infection without affecting the humoral response as measured by antiviral IgG antibody levels.

Stressor	Subject	Immunological Effect
Restraint	Rat	Involution of the thymus
Restraint	Mouse	Decreased macrophage cytotoxic activity
Noise	Mouse	Decreased lymphocyte proliferation to mitogenic stimulation, reduced lymphocyte cytotoxicity; prolonged noise-enhanced immune responses
Noise	Rat	Leukopenia
Isolation	Rat	Decreased lymphocyte proliferation to antigenic stimulation
Crowding	Rat	Increased lymphocyte stimulation to antigenic stimulation
Avoidance learning task	Mouse	Decreased interferon production, delayed allograph rejection, decreased antibody response to challenge, increased death rate following antigenic challenge
Uncontrollable shock	Rat	Decreased NK cell activity, decreased interleukin-2 production, decreased IL-2 receptor expression
Heat stress	Mouse	Diminished delayed-type hypersensitivity

Adapted from Weiss JM, Sundar S: Effects of stress on cellular immune responses in animals. In *Annual Review of Psychiatry*, vol 11. American Psychiatric Press, Washington, DC, 1992.

Table 1.12-7 Stress-Induced Changes in the Immune System of Rodents

Stress-induced alterations in immune function may be related to the psychological state of the animal. One study suggested that coping is an important variable in stress effects on the immune system. Whereas phytohemagglutinin and concanavalin A stimulation of lymphocytes was suppressed in rats exposed to inescapable, uncontrollable electric shock for 80 minutes, followed by several minutes of tail shock 24 hours later, animals receiving the same total amount of shock using a yoked paradigm but able to terminate the stressor did not have decreased lymphocyte activity compared with nonstressed controls. These findings are consistent with the hypothesis that the ability to cope with a stressor may mitigate against some of its noxious effects.

Finally, different mild stressors like electric foot shock and restraint have been found to elevate plasma levels of cytokines, including IL-6 in rats. Peripheral IL-6 release during stress may be related to catecholamines, given that epinephrine administration is also capable of inducing IL-6 secretion. Of note, is that stressor administration found to elevate IL-1b protein levels in the brain of stressed animals who were rendered glucocorticoid deficient by adrenalectomy. The ability of stressors to stimulate cytokine production and the ability of cytokines to alter CNS function in turn suggests that the immune system may contribute to the behavioral and neuroendocrine response to stress, even in the absence of a formal immune challenge (i.e., a pathogen).

Human Studies The effect of stress on measures of the immune system in humans has attracted considerable attention. In studies on academic stress among medical students, a decrease in natural killer cell activity was found during the final examination period as compared with a preexamination baseline. Examination stress has also been associated with decreased number of T cells, mitogen responses, interferon production, and antibody responses to recombinant hepatitis B vaccine. In addition, increases in antibody titers to latent herpes viruses, presumably secondary to impaired cellular immunity have been observed. Investigators have

also reported decreased measures of immune function in persons exposed to chronic life stressors, such as divorce and the taking care of patients with Alzheimer's disease. For example, caregivers to Alzheimer's disease patients showed alterations in lymphocyte subpopulations, increased antibody titers to herpes simplex virus, decreased proliferative response to mitogens, more days of illness from infectious disease, impaired antibody responses to an influenza virus vaccine, and longer latency for wound healing.

Conjugal bereavement is one of the most stressful of commonly occuring life events, and has been associated with increased medical morbidity and mortality. One group investigated the effects of bereavement on immune measures in a prospective longitudinal study of spouses of women with advanced breast carcinoma. Lymphocyte stimulation was measured in men before and after the deaths of their wives. Lymphocyte stimulation responses to phytohemagglutinin, concanavalin A, and pokeweed mitogen were significantly lower during the first 2 months after bereavement compared with prebereavement immune responses. However, the number of peripheral blood lymphocytes and the percentage and the absolute number of T and B cells during the postbereavement period were not significantly different from those in the prebereavement period. Follow-up during the remainder of the postbereavement year revealed that lymphocyte stimulation responses had returned to prebereavement levels for the majority of the men. Prebereavement mitogen responses did not differ from those of age-matched and sex-matched controls.

Advancing age has been shown to increase responsivity to the physiological effects of stress and depression on the neuroendocrine and immune systems. Age has been shown to interact with chronic stress and social support in determining the cardiovascular reactivity to stress, and the largest study to date investigating immune abnormalities in major depressive disorder revealed significant age-related differences between depressed patients and age- and sex-matched controls. For example, in contrast to age-related increases in mitogen responses in the controls, the depressed patients did not show increased lymphocyte responses with advancing age. Similar age-related differences between depressed patients and controls were found for $CD4^+$ lymphocytes. Finally, in a recent study on female caregivers of handicapped people, caregivers had a significantly lower percentage of T cells, a significantly higher percentage of $CD8^+$ cells, and a significantly lower T $CD4^+/CD8^+$ ratio; however, older caregivers (>45 years, median of age) also had lower numbers of T cells and $CD4^+$ cells and higher antibody titers to cytomegalovirus. The interaction between stress, depression, and age is important, since older people also show age-related alterations in immune function *(immunosenescence)*, which may leave them more vulnerable to stress-induced immune alterations and the development of immune-related diseases including infectious diseases and cancer. Whether stress-induced immune alterations will include clinically relevant changes in the ability of the organism to respond to infections or other in vivo immune challenges affecting health outcome is an active area of investigation. Several recent animal studies have begun to address the issue of direct evidence of a causal relation among stress, immune alterations, and disease states. One study found that rats exposed to an acute swimming stress exhibited decreased natural killer cell activity against a mammary tumor in vitro and an increase in tumor lung metastasis when the same tumor cells were injected in vivo. In addition, investigators have shown that restraint stress in mice not only suppressed measures of the cellular immune response to a specific herpes simplex viral infection but also was associated with an increased herpes simplex virus local infection. These animal studies support the hypothesis that stress effects on tumors and infection are a result of stress-induced suppression of the immune system.

Immunological Mechanisms Only recently have investigators begun to examine the immunological mechanisms through which stress may affect the immune system. In general, a stressor can alter immune function in two major ways. First, the stressor can lead to changes in the distribution of immune cells in any given part of the body, second, stress can alter the function of the immune cells themselves. Because the immune response depends on the interplay of various immune cell subtypes, a redistribution of relevant cell types into or out of a particular immune compartment can directly influence the local immune response. For example, significant and selective changes in peripheral blood cell distribution have been described in rats undergoing a mild acute stress (2 hours of restraint), including decreased numbers of white blood cells and decreased numbers and percentages of monocytes and lymphocytes including B cells, natural killer cells, and, to a lesser degree, T cells. Interestingly, cells leaving the blood during stress return when the stressor is discontinued; also, based on stress-induced enhancement of immune responses in the skin it appears that cells leaving the blood during stress may migrate to the skin where they would be more likely to encounter pathogens. Studies also have shown that stressors can alter a variety of cell-specific activities, including IL-2 production and the expression of the IL-2 receptor gene. In a study examining the biochemical mechanisms of stress-induced impairment of T-cell mitogenesis, spleen cells isolated from rats exposed to two brief stressors (5 minutes of restraint or 2 minutes of foot shock) exhibited a diminished response to both T-cell mitogens and a combination of the phorbol ester tetradecanoylphorbol acetate and the calcium ionophore ionomycin. Since stimulation with the latter two agents mimics early signals generated by mitogen surface receptor binding, including increased intracellular calcium and protein kinase C activation, the data indicate that stress-related defects in T-cell proliferation occur at sites beyond or in addition to the early events in cellular activation. Production of nitric oxide (NO) by macrophages has also been recently shown to be involved in the biochemical mechanisms underlying the stress-induced decrease in the lymphocyte proliferative response; both the depletion of macrophages and the inhibition of NO synthesis have been shown to attenuate stress-induced immune changes.

PUTATIVE MECHANISMS AND MEDIATORS

A number of studies have focused on the neuroendocrine mechanisms by which stress or alterations in CNS function may influence the immune response. The two systems that have received the most attention are the endocrine system, especially the hypothalamic-pituitary-adrenal axis, and the autonomic nervous system (Fig. 1.12-9). These two systems are intimately associated with the organism's response to stress, and immune cells and tissues not only express receptors for the transmitters and hormones emanating from those systems but also receive direct innervation from autonomic nervous system fibers.

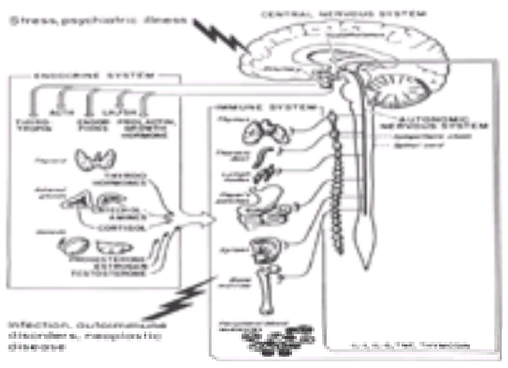

FIGURE 1.12-9 Diagram of bidirectional communication between the central nervous system and the immune system. The endocrine system and the autonomic nervous system are depicted as important mediators of CNS effects of the immune system. Inflammatory cytokines including IL-1, IL-6, and tumor necrosis factor (TNF) are shown closing the loop and interacting with the brain. CRH - Corticotropin-releasing hormone, ACTH - corticotropin, LH - luteinizing hormone, FSH - follicle stimulating hormone. (Drawn by Ellen Felten, Medical Arts Department, Mount Sinai Medical Center, NY.)

The interplay of the hypothalamic-pituitary-adrenal axis and autonomic nervous system with the immune system has been the focus of numerous recent experiments. The data indicate that both the hypothalamic-pituitary-adrenal axis and the autonomic nervous system have specific and selective effects on the immune system that are determined in part by which immune compartment and which immune response is examined. For example, after mild foot shock, investigators have found suppression of both splenic and peripheral blood lymphocyte responses to nonspecific T-cell mitogens. Removing endogenous corticosteroids by adrenalectomy prevents shock-induced suppression of the proliferative response of T cells in the peripheral blood. However, adrenalectomy has no effect on the stress-induced suppression of T-cell proliferation in the spleen. The b-adrenergic receptor antagonists propranolol (Inderal) and nadolol (Corgard) attenuates the shock-induced suppression of T-cell proliferation in the spleen but has no effect on stress-induced suppression of T-cell mitogen responses in the peripheral blood. The findings that stress-induced changes of immune parameters in the peripheral blood are adrenal dependent or pituitary dependent whereas stress effects on immune parameters in the spleen are related to catecholamine release has been demonstrated in a number of laboratories using a variety of paradigms.

A central factor in the regulation of the response to stress is *corticotropin-releasing hormone* (CRH). Aside from its ability to modify the expression of a number of animal behaviors, CRH is a pivotal molecule in mediating the hypothalamic-pituitary-adrenal axis and the sympathetic nervous system (SNS) outflow of the stress response to peripheral tissues. The immunological effects of CRH on a wide range of immune functions have been well characterized. CRH-administered icv was first shown to have a powerful immunosuppressive effect on NK-cell activity in the rat spleen. Follow-up studies have shown that CRH is also capable of inhibiting in vivo antibody formation, specifically the generation of an IgG response to immunization with keyhole limpet hemocyanin. The influence of CRH on antibody responses is also apparent in CRH-overproducing mice whose immune deficits are characterized by a profound decrease in B-cell number and severely diminished primary and memory antibody responses. Chronic icv administration of CRH and acute infusion of CRH into the locus ceruleus have been shown to suppress

lymphocyte-proliferative responses to nonspecific mitogens and antibody to the T cell receptor (anti-CD3).

Interestingly, CRH has also been found to stimulate the release of proinflammatory cytokines in laboratory animals and humans. Long-term intracerebroventricular administration of CRH to rats led to induction of IL-1b messenger ribonucleic acid (mRNA) in splenocytes, and intravenous infusion of CRH in humans led to an almost fourfold induction of IL-1a. Both treatments also led to significant increases in the immunoregulatory cytokine, IL-2. Finally, the proinflammatory effects of CRH are manifest in the direct autocrine or paracrine inflammatory actions of this peptide at peripheral sites of inflammation such as in an inflamed arthritic joint. Taken together, these results demonstrate that CRH has well-documented immunosuppressive effects on in vivo cellular and humoral responses while having a stimulatory effect on cytokine production and local inflammation.

The mechanism by which centrally administered CRH influences the immune response has been considered. Two of the major outflow pathways that are stimulated by CRH in the hypothalamus are the hypothalamic-pituitary-adrenal axis, which ultimately releases glucocorticoids, and the SNS, which releases catecholamines. Both these factors (glucocorticoids and catecholamines) are known to influence multiple aspects of the cellular and humoral immune response as well as the production and release of cytokines. Activation of the SNS by intracerebroventricular CRH has been found to be a major regulator of the effects of CRH on splenic natural killer activity, in particular by use of the sympathetic ganglionic blocker, chlorisandamine, which reverses the inhibitory effect of intracerebroventricular CRH on natural killer activity in the spleen. The hypothalamic-pituitary-adrenal axis is also involved in CRH immune effects as demonstrated by the finding that the effects of chronic icv CRH on splenocyte proliferative responses were eliminated by adrenalectomy. In addition, the B-cell decreases found in CRH-overproducing mice are very consistent with the marked reduction of rodent B cells found following chronic exposure to glucocorticoids.

Endorphins also appear to play a role in stress effects on natural killer cell activity in the spleen. Rats subjected to a foot-shock paradigm known to be associated with opioid analgesia exhibited a decreased natural killer cell activity that was prevented by injections of the long-acting opioid antagonist naltrexone (ReVia).

Much of the focus of CNS-immune system interactions has been on the inhibitory immunological effects of neurally derived or regulated molecules, but several pituitary hormones, including prolactin and growth hormone, seem to have immune-enhancing or immunoprotective properties. For example, removal of the pituitary before stress results in an augmented inhibitory effect of stress on a variety of immune parameters, suggesting that pituitary hormones may restrain stress-induced effects.

Finally, the effects of cytokines on the nervous system and the endocrine system close the loop between the brain and the immune system and indicate that neural-immune interactions are bidirectional (Fig. 1.12-9).

IMMUNE SYSTEM EFFECTS ON CNS AND ENDOCRINE FUNCTION

A tremendous amount of interest has been generated in the neurosciences by the discovery that cytokines are capable of exerting profound effects on the CNS and neuroendocrine system. Acting at the level of the CNS, the pituitary, and the adrenal glands, cytokines appear to play a role in the regulation of sleep, temperature, feeding behavior, neurotransmission, and the secretion of multiple peptides and hormones, including CRH, ACTH, prolactin, leutinizing hormone, follicle stimulating hormone, growth hormone, thyroid stimulating hormone and glucocorticoids.

By making electrophysiological recordings from individual neurons in the ventromedial hypothalamus at various time intervals after an antigenic challenge, investigators in the early 1980s found that the hypothalamic neuronal firing rate was maximally increased at the time of the peak of the immune response, when presumably the release of immunological mediators was greatest. Furthermore, blood concentrations of glucocorticoids were noted to be increased at the time of peak immune responsivity, whereas norepinephrine concentrations in the hypothalamus were found to be decreased. To determine whether cytokines were influencing hypothalamic norepinephrine concentrations and ultimately glucocorticoid secretion, the researchers injected supernatants from mitogen-stimulated lymphocytes into rats and measured hypothalamic norepinephrine concentrations; a significant drop in norepinephrine concentrations 2 hours after injection was seen. That decrease was similar to the decrease in norepinephrine seen during a normal immune response except that, as anticipated, the kinetics were much accelerated. Those studies have been repeated, using IL-1, with similar results, and subsequent studies have characterized the capacity of numerous cytokines to influence a host of monoamine and other neurotransmitter pathways. IL-1 also has been shown to exhibit potent stimulation of the secretion of hypothalamic CRH and therefore is capable of activating the neuroendocrine cascade, resulting in increased glucocorticoid release. In addition, IL-1 b and its mRNA (as well as other cytokines including TNF-a, IL-6 and IL-2) have been found in nerve cell bodies and nerve fibers within the hypothalamus, the hippocampus, and other regions in human and rodent brain, suggesting that IL-1 and other cytokines may play a role in neurotransmission.

Further evidence that the immune system may exert a powerful regulatory influence on neuroendocrine function is provided by data showing the effects of cytokines on the pituitary gland. For example, in vitro studies have shown that IL-1 is capable of enhancing the release of corticotropin (adrenocorticotrophic hormone [ACTH]), luteinizing hormone, growth hormone, and thyroid-stimulating hormone while inhibiting the release of prolactin from rat pituitary cells. Those effects occur at concentrations between 10^{-9} and 10^{-12} mol and are eliminated by incubation with antibody to IL-1.

Evidence of direct interaction between the immune system and the adrenal gland also exists. Virus-infected lymphocytes produce hormones concomitantly with human leukocyte interferon. Two of the products are structurally similar to pituitary-derived ACTH and b-endorphin on the basis of antigenicity, molecular weight, and retention time on reverse-phase high-pressure liquid chromatography. In addition, the two products demonstrate the appropriate biological activity—that is, stimulating primary cultures of adrenal tumor cells to secrete glucocorticoids, binding opiate receptors in vitro, and causing analgesia and catatonia in mice. The simultaneous release of ACTH-like and b-endorphin-like products in response to a variety of stimuli, including CRH, indicates that immunocytes (probably macrophages), like pituitary cells, are capable of transcribing the proopiomelanocortin gene, which is responsible for coding the precursor protein from which ACTH and b-endorphin are derived. Lymphocyte production of an ACTH-like hormone suggests that immunocytes may be capable of tapping directly into the hypothalamic-pituitary-adrenal axis at the level of the adrenal gland, giving rise to a so-called lymphoid-adrenal axis. Nevertheless, the physiological relevance of the lymphoid-adrenal axis has not been established. Other hormones found to be secreted by immunocytes include somatostatin, vasoactive intestinal polypeptide, thyrotropin, prolactin and growth hormone.

Finally, based on recent data from mice infected with a virus, it appears that depending on the type of immune stimulus (e.g., bacterial versus viral), different cytokine pathways may be involved in neuroendocrine activation. Early glucocorticoid responses to viral infections and virus-type stimuli, for example, seem to preferentially involve IL-6, whereas TNF and IL-1 b may play a greater role in glucocorticoid release during bacterial infections.

Aside from the potent effects of cytokines (especially IL-1) on neuroendocrine function, mounting data has indicated that cytokines play an important role in the mediation of ‘sickness behavior.' Sickness behavior is a syndrome that includes weakness, malaise, listlessness, anhedonia, hypersomnia, anorexia, social isolation, hyperalgesia, and poor concentration. This syndrome typically occurs during infections, but may occur in a wide variety of clinical settings, including any medical condition or treatment that leads to significant inflammation or the release of proinflammatory cytokines (e.g., TNF, IL-1, and IL-6).

One of the first indications that cytokines might be involved in this syndrome in humans came from the behavioral effects of cytokines administered in clinical trials for immunological conditions including cancer. The ability of cytokine antagonists such as the IL-1 receptor antagonist (which is also expressed in the brain) to block symptoms of sickness behavior in laboratory animals has confirmed the role of IL-1 (and possibly other cytokines) in this syndrome. a-Melanocyte–stimulating hormone and vasopressin are capable of opposing the ability of cytokines to induce sickness behavior, suggesting that the brain exhibits several opposing processes for limiting the effects of these factors on CNS function.

Many of the behavioral effects of IL-1 are mediated within the CNS and since cytokines do not cross the blood-brain barrier, considerable attention has been paid to the mechanism by which peripherally released cytokines communicate with the brain. Four major hypotheses have been advanced including (1) active transport of cytokines across the blood-brain barrier, (2) access of cytokines to the brain in areas where the blood-brain barrier is weak, such as the organum vasculosum of the lamina terminalis, (3) conversion of cytokine signals into secondary signals, such as prostaglandin or nitric oxide signals, by endothelial cells that line the blood vessels of the brain, and (4) transmission of cytokine signals (through cytokine receptor binding) along sensory afferents to the nucleus of the solitary tract and then via ascending catecholaminergic pathways onto relevant brain regions including the paraventricular nucleus of the hypothalamus. There are data to support each of these mechanisms, and it appears that the various effects of cytokines on the CNS may each be mediated by one or more of these pathways. Once peripheral cytokine signals reach the brain, there is reason to believe that in many instances these signals are turned back into cytokine signals (peripheral cytokines beget new central cytokines). For example, peripheral administration of IL-1–inducers or IL-1 itself, is associated with an increase in IL-1 immunoreactivity and bioactivity in several brain regions, including the hippocampus and hypothalamus. Cytokines mediate their effects through specific receptors, that are members of several classes of receptor gene families, including the immunoglobulin family (for IL-1 and IL-6), the hemopoietin family (for Ils 2-7), and the nerve growth factor (NGF) family (for TNF and nerve growth factor). There is widespread and unique distribution of cytokine receptors in the brain. For instance, especially high concentrations of IL-1 and

its receptors are found in the rat hippocampus. Of note is that soluble receptors exist for many cytokines as well as decoy (inactive) receptors (such as the type II IL-1 receptor) both of which serve to limit cytokine action. A specific endogenous IL-1 receptor antagonist has also been described.

RELEVANCE OF IMMUNE-CNS INTERACTIONS TO PSYCHIATRY

There are at least two major considerations when examining the potential psychiatric relevance of nervous system-immune system interactions. First, it is important to consider the potential impact that psychiatric disorders may have on immune function. Given that CNS processes appear to be involved in immune regulation, it is logical to consider that disorders of CNS function as manifested by psychiatric disease may contribute to the development, course, or outcome of diseases involving the immune system. Second, given the potential impact the immune system may have on the nervous system both through direct (e.g., infections or autoimmune disease) or indirect (release of cytokines) actions, it is important to consider that the pathophysiology of some psychiatric disorders may involve significant contributions from the immune system. This area of inquiry is often referred to as *neuroimmunology* and involves the investigation of the potential causative role of infectious, paraneoplastic, and autoimmune processes in psychiatric and neurological syndromes.

PSYCHIATRIC DISORDERS, IMMUNOCOMPETENCE, AND HEALTH AND ILLNESS

The psychiatric disorder that has been most investigated in terms of the competence of the immune system is major depression. More than 30 studies and several meta-analyses have examined immune parameters in depression. Although a number of investigators have reported depression-related alterations in peripheral blood immune cell numbers and decreases in peripheral blood mitogen responses and natural killer cell activity, these findings have not been reliably replicated from study to study. Nevertheless, when results are pooled across studies more consistent results are apparent. Depressionrelated immune changes may occur in association with other variables that characterize persons with depression, such as age, sex, and symptom severity. For example, in a large study of 91 depressed patients and 91 controls, investigators found no mean differences in immune measures between the groups. However, in that study both advancing age and severity of depression were associated with decreases in $CD4^+$ -cell numbers and mitogen responsiveness of peripheral blood lymphocytes in depressed patients as compared to controls. Moreover, studies have reported sex differences in the relationship between immune function and depression. One study reported decreased natural killer activity in males and no changes in females; two other studies have reportedly found increased natural killer activity in depressed women compared with controls. Other important factors that appear to significantly influence immunologic responses in depressed patients are sleep and cigarette smoking. For example, disruption of sleep has been found to have a suppressive effect on natural killer cell activity. Nutritional considerations are also important.

Because depressed patients frequently exhibit alterations in cortisol secretion and since cortisol has potent immunoregulatory effects, a number of studies have examined the relation between measures of cortisol secretion and the immune response in depressed patients; however, no clear association between these variables has emerged.

Finally, epidemiological data have provided only limited support for the idea that depressive symptoms or mood disorders are associated with an increased mortality from immune-related diseases. For example, several studies have examined mortality in large numbers of patients with psychiatric disorders; in general, the results suggest that psychiatric disorders are significantly associated not with death from natural causes but rather with unnatural mortality such as suicide and accidental death. Furthermore, a 1994 meta analysis of 7 large prospective studies on the relative risk of developing cancer as a function of baseline depression found a small but marginally statistically significant increase (1 to 2 percent) in cancer rates in individuals with depressive symptoms versus nondepressed controls. Taken together with the inconsistent findings across studies of immune dysfunction in depressed patients, these results indicate that persons with depression do not clearly exhibit evidence of impaired immunocompetence and are not clearly at high risk for immune-related conditions. Nevertheless, these reports primarily focused on persons free of physical illness at the time of study, and considerable evidence suggests that persons with physical illness complicated by depression have a worse outcome than persons without depression. This evidence of increased morbidity and mortality in depressed patients with comorbid medical illnesses (e.g., neoplastic and cardiovascular disorders), emphasizes the notion that the effects of depression on disease development and outcome are ultimately determined by the genetic predisposition of the individual. Based on this predisposition, the pathophysiology of depression serves to interact and exacerbate disease-specific processes as opposed to having a specific association with a given disease.

Finally, a great deal of attention has been directed to the notion that stress and depression may influence immunocompetence in HIV-seropositive persons, thereby serving as cofactors in the progression of HIV infection to AIDS. The scientific evidence for an association between psychosocial factors and disease progression, however, remains somewhat elusive, and contrasting findings have been reported on the effects of stress on immune parameters in this population. For example, several studies have explored the relationship between psychosocial factors and the number of T-cell subsets among HIV-seropositive adults. In at least two studies no relation was found between life stress or depression and counts of $CD4^+$ cells, $CD8^+$ cells, or helper-suppressor T-cell ratios in a cohort of HIV-seropositive homosexual men. Furthermore, stress and depression measures were not related to advanced illness or mortality over time. In contrast, other studies have found that HIV-positive subjects who experienced severe stress have relevant changes in immune parameters, including lower $CD8^+$ and lower natural killer cell counts, increased levels of serum neopterin (a marker of immune activation), and decreased proliferative responses to the mitogen phytohemagglutinin. To complement these findings, a well-controlled 1997 study reported that severe life stress in people infected with HIV increased the odds of undergoing HIV disease progression nearly fourfold. This finding supports the notion that the previously observed stress-related immune changes translate into greater vulnerability for disease development. Ongoing studies examining psychosocial variables and immunological and clinical endpoints in individuals with HIV infection may shed more light on these important issues.

NEUROIMMUNOLOGY

Disorders that are typically associated with the immune system, including infectious diseases and autoimmunity, can influence CNS function. Microbial pathogens can invade the brain and either infect neural cells directly, such as the infection of neurons by polio virus, or grow locally outside the brain as in bacterial meningitis. If the infection results in the destruction of neurons, the consequences can be particularly devastating because, unlike glia or cells in the periphery, neurons have minimal capacity for regeneration. Thus, the brain has developed a specialized immune system to maintain the tenuous balance between protecting neurons from invading pathogens on the one hand, and preventing damage mediated by the immune response on the other. As an initial defense, the blood-brain barrier prevents the entry of microbial organisms, which may explain why most infections with blood-borne neurotropic viruses do not result in encephalitis.

The brain has historically been considered “immune privileged” because it lacks conventional lymphatics, has extremely low levels of MHC expression, and is resistant to the transmigration of immune cells. Nevertheless, a pronounced inflammatory response can be elicited by invading pathogens or other forms of injury. During CNS infection extracellular antigens are thought to drain to cervical lymphatics, glia are induced to express MHC class I and II, and activated T cells enter the CNS. Local induction of cytokines helps to orchestrate the immune response, and adhesion molecules are upregulated on endothelium and perivascular glia to facilitate the entry of macrophages and other leukocytes. Besides being a major source of proinflammatory cytokines, astrocytes and microglia become activated during inflammation and can produce other diffusible mediators such as NO, prostagladins, and excitatory amino acids. Such inflammatory reactions can occur even in the absence of an infectious agent as exemplified in postischemic brain injury in which there is a rapid induction of cytokines, extravasation of leukocytes, and gliosis. Which aspects of this immune reaction are deleterious and which aspects are the immunological accouterments of repair remains to be determined.

The idea that alterations of CNS function result from a combination of the direct effects of an injurious event on various cell types and the effects of cytokines and other inflammatory mediators on neurons and supporting cells is a cornerstone of neuroimmunology. The idea that infectious agents can lead to psychiatric disorders has been well established. Obvious examples include the mental retardation that may develop after congenital infection with rubella or cytomegalovirus, the delirium that accompanies acute meningoencephalitis after CNS infection by herpes simplex virus type I, dementias due to slow viruses, such as kuru and Creutzfeldt-Jakob disease, and the neuropsychiatric manifestations that occur during neurosyphilis.

Schizophrenia Several lines of evidence suggest that virus infection during neural development may be involved in the pathogenesis of some cases of schizophrenia. The data include (1) an excess number of patient births in the late winter and early spring, suggesting possible exposure to viral infection in utero during the fall and winter peak of viral illnesses; (2) an association between exposure to viral epidemics in utero and the later development of schizophrenia; (3) an increased likelihood for schizophrenic patients to have had older siblings in the household (a potential source of viral infections) as compared to controls; and (4) geographical variation of prevalence, with schizophrenia being more common at greater distance from the equator.

Investigators have reported various alterations in immune markers in patients with schizophrenia including increased levels of interferon, decreased IL-2 production, and increased IL-2 receptors. Cerebrospinal fluid (CSF) immunoglobulins have also been increased in some studies. Although these immune findings in schizophrenia patients may indicate evidence of immune system activation secondary to infection, it should be noted that they may also indicate that an autoimmune process is involved in the disorder. Attempts to isolate infectious agents, especially viruses and viral DNA, from schizophrenia patients have been unsuccessful. For

example, the polymerase chain reaction has been used to amplify virus-specific nucleic acids for cytomegalovirus, Borna disease virus, HIV, bovine viral diarrhea virus and influenza A in brain tissues from schizophrenia patients; no evidence of virus-specific nucleic acids was found in any of the tissue samples. Nevertheless, because the initial neuronal abnormalities in schizophrenia have been proposed to arise during neurodevelopment, a perinatal viral infection could insidiously disrupt development and then be cleared by the immune system prior to clinical diagnosis. In such a scenario, host factors such as cytokines could be responsible for causing the developmental abnormality by interacting with growth factors or adhesion molecules. Viral infections can also cause alterations in neurotransmitters and neuronal circuitry. For example, neonatal infection of rats with the lymphocytic choriomeningitis virus manifests in a loss of inhibitory neuron function in the hippocampus. This decrement of inhibitory control could release excitotoxic influences impinging on other neuron populations, and result ultimately in a complex cascade of selective neurodegeneration.

As previously noted, an autoimmune cause has been suspected in some patients with schizophrenia. Nevertheless, attempts to isolate autoantibodies to CNS tissue constituents in schizophrenia patients have failed to yield consistent results. Furthermore, since schizophrenia may involve various forms of CNS tissue damage, with the resultant release of brain antigens, autoantibodies to CNS tissues in those instances may be the result of CNS pathology rather than the cause. Nevertheless, several autoimmune disorders, including autoimmune disorders of the thyroid and such collagen vascular diseases as systemic lupus erythematosus and Behcet's syndrome, are capable of indirectly or nonspecifically altering CNS function, but only a few autoimmune conditions directly involve brain antigens. Neural cells are the target for autoantibodies in the paraneoplastic syndromes. For example, autoantibodies to cytoplasmic proteins of Purkinje cells are associated with subacute cortical cerebellar degeneration, which is a rare complication of breast or ovarian cancers. Autoantibodies to g-aminobutyric acid (GABA)-ergic neurons in the serum and the CSF appear to be the mechanism behind the stiff-man syndrome, a rare disorder characterized by progressive rigidity, accompanied by recurrent painful muscle spasms. Antineuronal antibodies can also arise following infection with group A b-hemolytic streptococcal infections, as exemplified by Sydenham's chorea. Considering that children with Sydenham's chorea frequently exhibit obsessive-compulsive symptoms, emotional lability, and hyperactivity, there may be a spectrum of pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections (PANDAS). In particular, sudden onset of obsessive-compulsive disorder, tics, attention-deficit/hyperactivity disorder, and other psychiatric syndromes has been characterized in children following infection with group A b-hemolytic streptococcus. These findings in conjunction with the indication of a genetic vulnerability to this condition (high frequency of binding of a monoclonal antibody designated D8/17) represents an exciting new development in the etiology and possible treatment of these disorders.

Major Depressive Disorder There has been increasing interest in the possibility that immune activation may contribute to the pathophysiology of depression. For example, elevated serum concentrations of the proinflammatory cytokines IL-1 and IL-6 as well as increased acute-phase proteins including C-reactive protein, haptoglobin, and alpha 1-acid glycoprotein have been found in patients with major depressive disorder. In addition, cellular markers of immune activation have been described. The source of immune activation in major depressive disorder is unknown, although studies have shown that both stress and CRH are capable of inducing proinflammatory cytokines in the absence of a formal immune challenge. Administration of a variety of cytokines in clinical trials also has been associated with the development of depressive syndromes (sickness behavior).

It has been suggested that the acute-phase response may contribute to decreases in the availability of L-tryptophan, thus contributing to reduced serotonin in the brain. The findings that stressors may be able to activate the release of proinflammatory cytokines and an acute-phase response in the absence of an immune challenge further supports the notion that the immune system may be recruited to participate in the behavioral response to stress, and therefore may contribute to the biochemical and molecular biological changes that characterize depression.

As for infectious diseases contributing to the pathophysiology of major depressive disorders, the potential contribution of a virus to the etiology of depression has long been considered. There has been renewed interest in this area based on the isolation of Borna disease virus from immune cells of patients with depression. Borna disease virus was primarily of interest as a pathogen of horses and sheep until recent serological and virological data demonstrated that it can infect humans and may be associated with psychiatric diseases. The neurotrophic and immunopathogenic capabilities of Borna disease virus have been well studied in the rat model. During the initial phase of Borna disease in rats, the infection of neurons and glia is accompanied by a robust inflammatory response with participation of T cells, B cells, and macrophages. Rats that recover from the initial phase may develop chronic symptoms including behavioral abnormalities and obesity; further studies are needed to clarify the importance of Borna disease virus in human neuropsychiatric disorders such as major depression.

Alzheimer's Disease Although Alzheimer's disease is not considered primarily an inflammatory disease, emerging evidence indicates that the immune system may contribute to its pathogenesis. The discovery that amyloid plaques are associated with acute-phase proteins such as complement proteins, a1-antichymotrypsin, and c-reactive protein, suggests the possibility of an ongoing immune response. Furthermore, gliosis and increased levels of proinflammatory cytokines are found in and around plaques. Interestingly, the induction of IL-6 has been proposed to precede neuritic degeneration in nascent (diffuse) plaques, and IL-1b can increase mRNA for amyloid precursor protein. Thus, immune mediators have been theorized to have an early role in plaque formation. The idea that inflammatory processes are involved in Alzheimer's disease has been bolstered by recent studies showing that the long-term use of nonsteroidal antiinflammatory drugs is negatively correlated with the development of Alzheimer's disease. Studies are underway to determine whether immune suppression with glucocorticoids can improve disease outcome.

HIV Infection Infection with HIV is an immunological disease associated with a variety of neurological manifestations including dementia. Although some neurological symptoms may be a consequence of opportunistic infections, accumulating evidence indicates that HIV itself can produce encephalitis. Infected microglia can be readily identified in the brain while infection of neurons does not appear to occur in vivo. Nevertheless, HIV encephalitis results in synaptic abnormalities and loss of neurons in the limbic system, basal ganglia, and neocortex. Current research examining the mechanism of this neurodegeneration has focused on a network of interactions between viral products (e.g., gp120, tat), glia, macrophages, and neurons. In this regard, soluble factors such as cytokines (IL-1, IL-6, TNF, TGF-b), free radicals, and excitatory amino acids have all been proposed as intermediaries in HIV-induced neuropathology. Studies in rodents have demonstrated that viral gp120 can induce IL-1b expression, activate glia, and cause neuronal damage. Antagonists of the *N*-methyl-D-aspartate (NMDA) excitatory amino acid receptor can ameliorate this damage although many aspects of this neuropathogenic pathway are unclear. Neuroendocrine abnormalities have also been described following HIV infection, perhaps as a result of cytokine activation. Thus, while the complex interactions between viral and host factors in HIV encephalitis may be perplexing, current research is revealing a variety of potential targets for therapeutic intervention in the disease.

Multiple Sclerosis Multiple sclerosis is a demyelinating disease characterized by disseminated inflammatory lesions of white matter. Since the mid-1970s considerable progress has been made in elucidating the immunopathology of myelin destruction that occurs in multiple sclerosis, and the animal model for the disease, *experimental allergic encephalomyelitis*. Although the initial step in lesion formation has not been determined, disruption of the blood-brain barrier, and the infiltration of T cells, B cells, plasma cells, and macrophages all appear to be associated with lesion formation. Proinflammatory cytokines, notably TNF-a, IFN-g, and IL-1 are believed to participate in the immunopathology by activating glia, inducing MHC class II, or mediating cytotoxicity by stimulation of free radical formation. Conversely, treatment with IFN-b decreases lesions and has been used clinically in multiple sclerosis. Current therapeutic strategies are focusing on preventing the transmigration of leukocytes into the brain by targeting chemokines (a type of cytokine that mediates chemoattraction), or by blocking adhesion molecules on endothelium or glia. Several viruses have been suggested to trigger multiple sclerosis, most recently human herpesvirus-6. One mechanism by which a virus could trigger multiple sclerosis is by molecular mimicry in which a shared epitope between a virus and the host allows for self tolerance to myelin to be broken. The selective depletion of autoreactive T cells or induction of immune tolerance by oral administration of myelin proteins is also under investigation.

Finally, there are several disorders in which neural-immune interactions are suspected but not well documented. Chronic fatigue syndrome is an illness with controversial etiology and pathogenesis. Besides persistent fatigue, symptoms frequently include depression and sleep disturbances. Tests of immune function have found indications of both immune activation and immunosuppression. Neuroendocrine assessments indicate that patients with chronic fatigue syndrome may be hypocortisolemic because of impaired activation of the hypothalamic-pituitary-adrenal axis. Although an acute viral infection frequently precedes the onset of CFS, no infectious agent has been definitively identified as causing the disease. In contrast, Lyme disease, in which sleep disturbances and depression are also common, is clearly caused by infection with the tick-borne spirochete, *Borrelia burgdoferi*. The organism can invade the CNS and cause encephalitis and neurological symptoms. Lyme disease is remarkable because it appears to produce a spectrum of neuropsychiatric disorders including anxiety, irritability, obsessions, compulsions, hallucinations, and cognitive deficits. Immunopathology of the CNS may be involved because symptoms can persist or reappear even after a lengthy course of antibiotic treatment, and the spirochete is frequently difficult to isolate from the brain. Gulf War syndrome is a controversial condition with inflammatory and neuropsychiatric features. The condition has been attributed variously to combat stress, chemical weapons (e.g., cholinesterase inhibitors), infections, and vaccines. Given the impact of stress on neurochemistry and immune responses, these pathogenic mechanisms are not mutually exclusive.

IMMUNOCYTES AS TOOLS FOR EXPLORING THE NEUROBIOLOGY OF PSYCHIATRIC SYNDROMES

Unlike brain cells, immune cells are readily available from the peripheral blood and express receptors for hormones and transmitters that are altered in a variety of psychiatric disorders. Immune cells may therefore provide useful receptor models for exploring the molecular and biochemical mechanisms that may be involved in altered neuroendocrine or neurotransmitter function.

For example, adrenal steroid receptors are expressed in immune tissues and, after long-term exposure to glucocorticoids in rats, those receptors in immune tissues exhibit downregulation in parallel with adrenal steroid receptors in multiple brain regions, including the hypothalamus and the hippocampus. During major depression, abnormalities in hypothalamic-pituitary-adrenal axis activity, including hypercortisolemia and nonsuppression of the hypothalamic-pituitary-adrenal axis by dexamethasone, are believed to be partly related to reduced glucocorticoid receptor responsiveness to feedback inhibition by glucocorticoids at the level of the hippocampus, the hypothalamus, and the pituitary. Therefore, several studies have been conducted investigating glucocorticoid receptors in lymphocytes from depressed patients. Nevertheless, results have been inconsistent; in some studies, decreased receptor number in depressed patients has been observed; others have obtained discordant results. Fewer studies have investigated the functional sensitivity of cells to the inhibitory effects of glucocorticoids, but they have been more consistent, showing increased resistance of cells from depressed patients to the inhibitory effects of glucocorticoids on immune function. Evidence of altered glucocorticoid receptor function in lymphocytes from depressed patients indicates that the glucocorticoid resistance in this disorder may be reflected in the readily accessible cells from the immune system. Abnormalities in lymphocyte glucocorticoid receptor expression have also been found in patients with posttraumatic stress disorder. In contrast to patients with depression, patients with posttraumatic stress disorder exhibit significantly higher numbers of glucocorticoid receptors than in controls; these findings are accompanied by significantly lower concentrations of plasma cortisol in posttraumatic stress disorder patients. Immune cells may therefore provide a useful receptor model both for identifying persons with hypothalamic-pituitary-adrenal axis alterations and exploring the molecular mechanisms involved.

Another lymphocyte receptor that has been evaluated in the depressive disorders is the b-receptor. In the majority of studies, b-adrenergic receptors of peripheral immune cells from depressed patients have exhibited evidence of diminished b-adrenergic responsiveness. The decreased responsivity may be the result of desensitization (diminished function, normal number) or downregulation (diminished number) of leukocyte b-receptors. Because postmortem brain b-receptors from depressed patients (death by suicide) have been shown to be upregulated (increased receptor binding), brain and lymphocyte b-receptors appear to have an inverse relationship in the depressive disorders. However, whether the decreased responsivity of immune cell b-receptors is caused by an increase in plasma catecholamines or whether a decrease in peripheral sensitivity to catecholamines leads to an increase in brain catecholamine production is unknown. Nevertheless, as with glucocorticoid receptors, immune cells may be an important tool for identifying and evaluating b-receptor alterations in the depressive disorders.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2; and neuroimaging is covered in Section 1.15 and Section 1.16. Schizophrenia is covered in Chapter 13, mood disorders in Chapter 14, and anxiety disorders in Chapter 15. Behavior and immunity is discussed in Section 25.10. Alzheimer's disease is presented in Chapter 10 and Section 51.3e.

SECTION REFERENCES

*Abbas AK, Lichtman AH, Pober JS: *Cellular and Molecular Immunology*, ed 2. WB Saunders, Philadelphia, 1994.

Ackerman KD, Bellinger DL, Felten SY, Felton DL: Ontogeny and senescence of noradrenergic innervation of the rodent thymus and spleen. In *Psychoneuroimmunology*, ed 2, R Ader, DL Felten, N Cohen, editors. Academic Press, New York, 1991.

Ader R, Felten DL, Cohen N, editors: *Psychoneuroimmunology*, ed 2. Academic Press, New York, 1991.

Bellinger DL, Felten SY, Lorton D, Felton DL: Neural modulation of the immune system. In *Immunology of the Nervous System*, RW Keane, W Hickey, editors. Oxford University Press, New York, 1996.

Ben-Eliyahu S, Yirmiya R, Liebeskind JC, Taylor AN, Gale RP: Stress increases metastatic spread of a mammary tumor in rats: Evidence for mediation by the immune system. Brain Behav Immun *5:*193, 1991.

Berkenbosch F, Oers JV, Del Rey A, Tilders F, Besedovsky H: Corticotropin-releasing factor-producing neurons in the rat activated by interleukin-1. Science *238:*524, 1987.

Besedovsky H, Del Rey A, Sorkin E: The immune response evokes changes in brain noradrenergic neurons. Science *221:*564, 1983.

*Besedovsky HO, Del Rey A: Immune-neuro-endocrine interaction: Facts and hypotheses. Endocrine Rev *17:*64, 1996.

Bonneau RH, Sheridan JF, Feng N, Glaser R: Stress-induced modulation of the primary cellular immune response to herpes simplex virus infection is mediated by both adrenal-dependent and independent mechanisms. J Neuroimmunol *42:*167, 1993.

Breder CD, Dinarello CA, Saper CB: Interleukin-1 immunoreactive innervation of the human hypothalamus. Science *240:*321, 1988.

Cohen S, Tyrell D AJ, Smith AP: Psychological stress and susceptibility to the common cold. N Engl J Med *325:*606, 1991.

*Connor TJ, Leonard BE: Depression, stress and immunological activation: the role of cytokines in depressive disorders. Proc Natl Sci Counc Repub China B *62:*583, 1998.

Cunnick JE, Lysle DT, Kucinski BJ, Rabin GS: Evidence that shock-induced immune suppression is mediated by adrenal hormones and peripheral beta-adrenergic receptors. Pharmacol Biochem Behav *36:*645, 1990.

Dhabhar FS, McEwen BS: Enhancing versus suppressive effects of stress hormones on skin immune function. Proc Nat Acad Sci USA *96:*1059, 1999.

Dhabhar FS, Miller AH, McEwen BS, Spencer RL: Stress-induced changes in blood leukocyte distribution. Role of adrenal steroid hormones. J Immunol *157:*1638, 1996.

Ericsson A, Kovacs KJ, Sawchenko PE: A functional anatomical analysis of central pathways subserving the effects of interleukin-1 on stress-related neuroendocrine neurons. J Neurosci *14:*897, 1994.

Evans DL, Leserman J, Perkins DO, Stern RA, Murphy C, Tamul K, Liao D, van der Horst CM, Hall CD, Folds JD, Golden RN, Petitto JM: Stress-associated reductions of cytotoxic T lymphocytes and natural killer cells in asymptomatic HIV infection. Am J Psychiatr *152:*543, 1995.

Fawzy FI, Fawzy NW, Hyun CS, Elashoff R, Guthrie D, Fahey JL, Morton DL: Malignant melanoma: Effects of an early structured intervention, coping, and affective state on recurrence and survival 6 years later. Arch Gen Psychiatry *50:*581, 1993.

*Goehler LE, Gaykema RPA, Nguyen KT, Lee JE, Tilders FJH, Maier SF, Watkins LR: Interleukin-1B in immune cells of the abdominal vagus nerve: A link between the immune and nervous systems? J Neurosci *19:*2799, 1999.

Justice A: Review of the effects of stress on cancer in laboratory animals: Importance of time of stress application and type of tumor. Psychol Bull *98:*108, 1985.

Karalis K, Muglia LJ, Bae D, Hilderbrand H, Majzoub JA: CRH and the immune system. J Neuroimmun *72:*131, 1997.

Keller S, Schleifer SJ, Liotta AS, Bond RN, Farhoody N, Stein M: Stress-induced alterations of immunity in hypophysectomized rats. Proc Natl Acad Sci USA *85:*9297, 1988.

Keller SE, Weiss JM, Schleifer SJ, Miller NE, Stein M: Suppression of immunity by stress: Effects of a graded series of stressors on lymphocyte stimulation in the rat. Science *213:*1397, 1981.

Kent S, Bluthe RM, Kelley KW, Dantzer R: Sickness behavior as a new target for drug development. Trends Pharmacol Sci *13:*24, 1992.

Kiecolt-Glaser JK, Glaser R, Gravenstein S, Malarkey WB, Sheridan J: Chronic stress alters the immune response to influenza virus vaccine in older adults. Proc Natl Acad Sci USA *93:*3043, 1996.

Leserman J, Petitto JM, Perkins DO, Folds JD, Golden RN, Evans DL: Severe stress, depressive symptoms, and changes in lymphocyte subsets in human immunodeficiency virus-infected men. A 2-year follow-up study. Arch Gen Psychiatry *54:*279, 1997.

*McEwen BS, Biron CA, Brunson KW, Bulloch K, Chambers WH, Dhabhar FS, Goldfarb RH, Kitson RP, Miller AH, Spencer RL, Weiss JM: The role of adrenocorticoids as modulators of immune function in health and disease: Neural, endocrine and immune interactions. Brain Res Brain Res Rev *23:*79, 1997.

Miller AH: Neuroendocrine and immune system interactions in stress and depression. Psychiatr Clin North Am, in press.

Nguyen KT, Deak T, Owens SM, Kohno T, Fleshner M, Watkins LR, Maier SF: Exposure to acute stress induces brain interleukin-1b protein in the rat. J Neurosci *18:*2239, 1998.

Pearce BD, Steffensen SC, Paoletti AD, Henriksen SJ, Buchmeier MJ: Persistent dentate granule cell hyperexcitability after neonatal infection with lymphotic choriomeningitis virus. J Neurosci

*16:*220, 1996.

Riley V: Psychoneuroendocrine influences on immunocompetence and neoplasia. Science *212:*1100, 1981.

*Rothwell NJ: Annual review prize lecture: Cytokines—Killers in the brain? Physiol *514:*3, 1999.

Ruzek MC, Miller AH, Opal SM, Pearce BD, Biron CA: Characterization of early cytokine responses and an interleukin (IL)-6-dependent pathway of endogenous glucocorticoid induction during murine cytomegalovirus infection. J Exper Med *185:*1185, 1997.

*Sheridan JF, Dobbs C, Brown D, Zwilling B: Psychoneuroimmunology: Stress effects on pathogenesis and immunity during infection. Clin Microbiol Rev *7:*200, 1994.

Schleifer SJ, Keller SE, Camerino M, Thornton VC, Stein M: Suppression of lymphocyte stimulation following bereavement. JAMA *250:*374, 1983.

Steel DM, Whitehead AS: The major acute phase reactants: C-reactive protein, serum amyloid P component and serum amyloid A protein. Immunol Today *15:*81, 1994.

Spiegel D, Kraemer HC, Bloom JR, Gotheil E: Effects of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet *2:*888, 1989.

Stein M, Miller AH, Trestman RL: Depression, the immune system and health and illness. Arch Gen Psychiatry *48:*171, 1991.

Swedo SE, Leonard HL, Garvey M, Mittleman B, Allen AJ, Perlmutter S, Lougee L, Dow S, Zamkoff J, Dubbert BK: Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections: Clinical description of the first 50 cases. Am J Psychiatry *155:*264, 1998.

*Turnbull AV, Rivier CL: Regulation of the hypothalamic-pituitary-adrenal axis by cytokines: Actions and mechanisms of action. Physiol Rev *79:*1, 1999.

Yehuda R, Boisoneau D, Mason JW, Giller EL: Glucocorticoid receptor number and cortisol excretion in mood, anxiety, and psychotic disorders. Biol Psychiatry *34:*18, 1993.

Zhou D, Kusnecov AW, Shurin MR, DePaoli M, Rabin BS: Exposure to physical and psychological stressors elevates plasma interleukin 6: Relationship to the activation of the hypothalamic-pituitary-adrenal axis. Endocrinology *133:*2523, 1993.

1.13 CHRONOBIOLOGY

THOMAS A. WEHR, M.D.

Chronobiology is the study of biological rhythms. Such rhythms span a spectrum of periods from the fraction-of-a-second fluctuations of the electroencephalogram to multiyear cycles in populations of predators and their prey. Two types of biological rhythms, circadian (daily) rhythms and seasonal rhythms, enable the organism to anticipate and adapt to corresponding cycles in the environment.

PHYSIOLOGY

Circadian Rhythms Because the Earth rotates on its axis, it presents two different environments—a daytime world and a nighttime world—to the organisms that live on it. In the course of their evolution, most animals have specialized in the active engagement of only one of those two environments. Because adaptations that make animals fit for one environment are liable to make them unfit for the other, animals have also adopted ways of withdrawing from the unfavorable environment (e.g., by sleeping in secure refuges) to avoid danger and to prevent the inefficient use of energy.

Thus, as the world alternates between phases of light and of darkness, animals alternate between phases of active engagement and of withdrawal. That alternation of behavioral states is not simply a passive response to the alternation of day and night. Animals possess clocklike mechanisms that automatically switch them from one state to the other in anticipation of the transitions between day and night. Those mechanisms also switch back and forth between contrasting physiological states that are geared to engagement or to withdrawal. For that reason, circadian rhythms in animals' behavior and physiology often exhibit two distinct phases, a diurnal phase and a nocturnal phase, with relatively discrete transitions between them. In humans, for example, the *diurnal phase* is one of active engagement of the environment, high body temperature, and decline or cessation of the secretion of several hormones, including pineal melatonin, pituitary thyrotropin, pituitary prolactin, pituitary corticotropin, and adrenocortical steroids. Conversely, the *nocturnal phase* is one of withdrawal from the environment, rest, sleep, low body temperature, and augmentation of the secretion of the above hormones.

The best evidence that circadian rhythms are not simply passive responses to cyclic changes in the external environment but are generated by endogenous processes is the fact that the rhythms persist when organisms are placed in constant conditions. In that situation the rhythms continue to oscillate but with an intrinsic period that differs slightly from 24 hours (hence the name “circadian,” about daily). In humans, for example, the period is slightly longer than 24 hours. In a normal environment the circadian system responds to external stimuli that serve as time cues, *zeitgebers*, in such a way that the period of its oscillations is adjusted to 24 hours and a characteristic *phase relationship* to the day-night cycle is maintained. Thus, in humans external time cues advance the phase of their longer-than-24-hour circadian rhythms several minutes each day, so that sleep recurs every 24 hours and recurs at night. Otherwise, humans would go to sleep later and later each day and would make a full circuit around the clock every few weeks.

Light, acting on the retina, is the most important stimulus for phase-resetting of the circadian system. The direction and the magnitude of the phase-resetting effect of light depends on the circadian phase at which it is applied. For example, early-morning light shifts the oscillations of the circadian system to an earlier *advanced-phase position*; late-evening light shifts the oscillations to a later *delayed-phase position*. Light in the middle of the day has little or no effect. Thus, in humans, evening light delays the time of sleep onset, but morning light advances it. The dependence of light's phase-resetting effects on the circadian phase at which it is applied can be shown quantitatively in a *phase-response curve* that indicates the magnitude and the direction of phase shifts induced by light pulses at each circadian phase (Fig. 1.13-1). The phase-resetting properties of light have led to its use in the treatment of certain types of sleep and mood disorders and of jet lag.

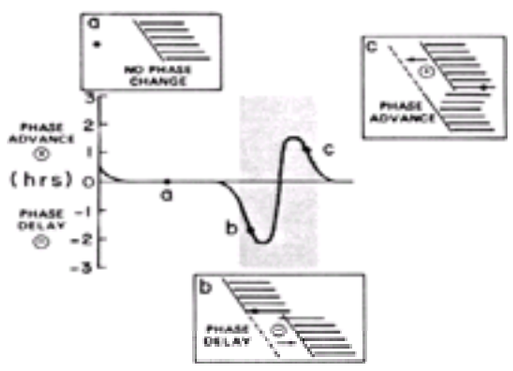

FIGURE 1.13-1 The derivation of a phase-response curve for the phase-resetting effect of light pulses in an animal whose circadian locomotor activity rhythms are free-running in constant darkness. Light pulses result in no phase change **(a)**, a phase delay **(b)**, or a phase advance **(c)**, depending on the phase of the circadian cycle at which the stimulus was presented. (Reprinted with permission from Czeisler CA, Richardson GS, Coleman RM, Zimmerman JC, Moore-Ede MC, Dement WC, Weitzman ED: Chronotherapy: Resetting the circadian clocks of patients with delayed sleep phase insomnia. Sleep *4:*1, 1981.)

Light-Oriented Rhythms and Behavior-Oriented Rhythms Circadian rhythms in animals appear to be of two types: (1) rhythms whose phase relations to the external light-dark cycle are similar in day-active and night-active species, but whose phase relations to the endogenous activity-rest cycle are opposite (Type I) and (2) rhythms whose phase relations to the external light-dark cycle are opposite in day-active and night-active species but whose phase relations to the endogenous activity-rest cycle are similar (Type II).

Type I rhythms have the same orientation to the light-dark cycle in both day-active and night-active species. They include the rhythms in the circadian pacemaker's responses to light pulses, which reaches its maximum at night (Fig. 1.13-1); the rhythm in the hypothalamic suprachiasmatic nucleus's electrical multiple-unit activity and glucose utilization (Fig. 1.13-2), which reaches its maximum in the daytime; the rhythm in retinal disk shedding, which reaches its maximum in the morning; the rhythm in melatonin secretion, which reaches its maximum at night (Fig. 1.13-3); and the rhythm in cerebrospinal fluid arginine vasopressin secretion, which reaches its maximum in the daytime. Most of these rhythms are intimately connected with mechanisms responsible for the generation and entrainment of circadian rhythms, and one of their most important functions appears to be the transmission of information about the timing of day and night to other parts of the organism.

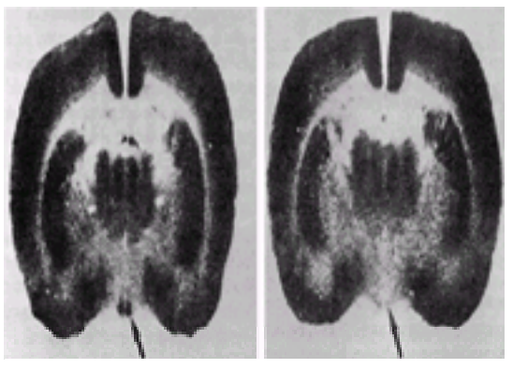

FIGURE 1.13-2 In autoradiographs of coronal sections of rat brain, 2-deoxyglucose uptake in the suprachiasmatic nuclei of the hypothalamus (indicated by arrows) is higher during the light phase **(left)** than during the dark phase **(right)** of a 24-hour light-dark cycle. (Reprinted with permission from Schwartz WJ, Smith CB, Davidsen LC: In vivo glucose utilization of the suprachiasmatic nucleus. In *Biological Rhythms and Their Central Mechanism*, M Suda, O Hayaishi, H Nakagawa, editors. Elsevier, New York, 1979.)

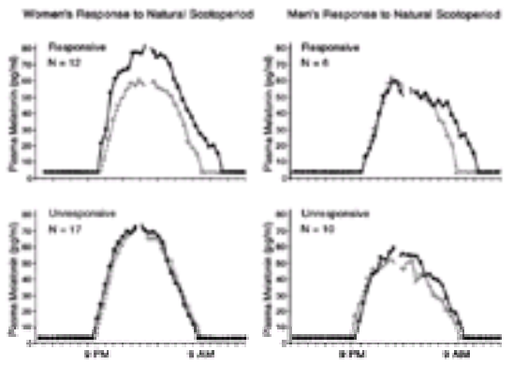

FIGURE 1.13-3 The intrinsic duration (the duration in constant dim light) of nocturnal melatonin secretion becomes longer in winter (black lines) than in summer (hatched lines) in some individuals living in a modern urban environment **(top)** but not in others **(bottom)**. In that environment, women are more likely to respond to seasonal changes in duration of the natural scotoperiod than men (maximum melatonin level also increases in winter compared with summer in women, but not in men). (Reprinted with permission from Wehr TA: Melatonin and seasonal rhythms. J Biol Rhythms *12:*517, 1997.)

Type II rhythms have the same orientation to the activity-rest cycle in both day-active and night-active species. They include the rhythm in sleep, which reaches its maximum during the rest phase; the rhythm in body temperature, which reaches its maximum during the activity phase; and the rhythms in pituitary thyrotropin secretion, pituitary corticotropin secretion, and adrenocortical steroid secretion, which reach their maximums during the rest phase. These rhythms seem to be directly connected with the maintenance of internal conditions that serve active engagement of and withdrawal from the environment.

If Type I rhythms are directly connected with clock mechanisms and Type II rhythms are directly connected with the expression of behavioral rhythms that are driven by those clock mechanism, it is likely that Type II rhythms occupy a more subordinate position that Type I rhythms in the hierarchical organization of the circadian system.

Sunrise and Sunset To coordinate the timing of the diurnal and nocturnal periods of circadian rhythms with the timing of day and night, organisms must track and anticipate the transitions between day and night. That task is complicated by the fact that the Earth's axis of rotation is tilted, causing the interval between dawn and dusk to vary during the course of the year. Animals appear to deal with this situation by using two separate mechanisms to track and to anticipate the changing times of dawn and dusk. In that way they are able to modify the durations of diurnal and nocturnal periods of their circadian rhythms to match the changing durations of day and night outside. According to one model, this is accomplished by a circadian pacemaker that consists of two coupled oscillators. One oscillator, entrained to sunset, controls the timing of the transition from the diurnal period to the nocturnal period of the circadian system; a second oscillator, entrained to sunrise, controls the timing of the transition from the nocturnal period to the diurnal period of the circadian system. According to this model, in humans the dusk-entrained oscillator controls the onset of sleep and the onset of nocturnal melatonin secretion, and the dawn-entrained oscillator controls the offset of sleep and the offset of nocturnal melatonin secretion. Thus, when nights become longer in winter, the sleep period and the melatonin secretion period become longer, too (Fig. 1.13-3).

The fact that activity or sleep in animals and humans tends to be bimodally distributed, with a peak near dusk and a peak near dawn (Fig. 1.13-4) seems consistent with a two-oscillator model. The strongest support for the model comes from the phenomenon of splitting of activity-rest cycles: when nocturnal rodents are exposed to constant light, their activity sometimes splits into two circadian components that seem to oscillate independently. A phenomenon that resembles splitting has also been described in human sleep-wake cycles.

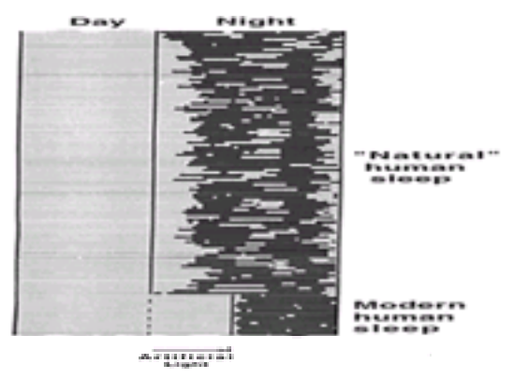

FIGURE 1.13-4 In long (14-hour) nights, human sleep is no longer consolidated, but divides into an evening bout and a morning bout that are separated by a period of quiet wakefulness. The graph shows nightly patterns of sleep (dark, horizontal bars) and wakefulness (horizontal lines), with data for each night plotted successively beneath the night that preceded it, in a young woman who rested and slept in 14-hour nightly dark periods for 15 weeks and then returned to a conventional 8-hour sleep schedule. The fact that many types of animals exhibit similar bimodal patterns of distribution in their nocturnal behavior suggests that this pattern of human nocturnal behavior is a natural one. Consistent with this idea, diaries and other historical records from the preindustrial era typically contain references to a “first sleep” and a “morning sleep.” Modern humans use artificial light to extend their waking activities into the first hours of the night (arrow). This practice causes their sleep to become consolidated and reduced in duration. (Reprinted with permission from Wehr TA: Seasonal photoperiodic responses of the human circadian system. In *Handbook of Behavioral Neurobiology: Circadian Clocks*, JS Takahashi, FW Turek, RY Moore, editors. Plenum, New York, 1998.)

Seasonal Rhythms The system of dawn-tracking and dusk-tracking circadian mechanisms that adjusts the duration of diurnal and nocturnal periods of circadian rhythms to match the changing duration of day and night has a second important function. The timing of dawn relative to the timing of dusk is an indicator of the time of year. Taking advantage of that fact, animals also use dawn-tracking and dusk-tracking mechanisms as a kind of biological calendar so that they can anticipate conditions that will prevail at particular times of the year and can initiate responses that are likely to be adaptive. For example; animals respond to critical changes in

day length *(photoperiod)* by storing energy in anticipation of seasonal scarcity of food or by conceiving offspring in anticipation of seasonal abundance of food.

Melatonin as a Chemical Transducer of Change of Season The pineal hormone *melatonin* is an important chemical mediator of the effects of seasonal changes in the photoperiod on reproduction and other functions. In some animals the duration of nocturnal melatonin secretion appears to be regulated by separate circadian processes that are synchronized with dusk and dawn and that trigger the onset and the offset of secretion. Thus, in winter, during which nights are long, the duration of nocturnal melatonin secretion is long. In summer, when nights are short, the duration of melatonin secretion is short (Fig. 1.13-3). In many species these changes in the duration of melatonin secretion serve as chemical signals that convey information about the duration of night to target sites in the body that induce season-specific changes in behavior and physiology. For example, numerous experiments have shown that the effect of the photoperiod on breeding is mediated by changes in the duration of nocturnal melatonin secretion.

Elements of the reproductive system that respond to the seasonal melatonin message appear to possess *interval timers* that measure the duration of melatonin secretion. When the duration increases by a certain amount, winter-type reproductive responses are triggered. When the duration declines by a certain amount, summer-type responses are triggered. Further research has shown that the direction of change in the duration of melatonin secretion is also important in determining the nature of the response. Intermediate durations of melatonin secretion can elicit either summer-type responses or winter-type responses, depending on whether the durations to which an animal was previously exposed were long or short.

The exact nature of an animal's reproductive responses to photoperiod-induced changes in melatonin secretion depends on its species and the length of its gestational period. For example, short photoperiods and long melatonin durations induce breeding in sheep, but long photoperiods and short melatonin durations induce breeding in hamsters. Because the gestational period of sheep is long, the gestational period of hamsters is short, the offspring of each species tend to be born at approximately the same time of year, the late spring—a time when conditions are favorable for their survival. Thus, melatonin is generically neither progonadal nor antigonadal. Rather, it conveys information about the duration of the photoperiod, and that information is used in different ways by different animals.

A 420-amino acid melatonin receptor has been identified and cloned. Structure analysis revealed that the receptor protein is a member of the guanine nucleotide binding protein-coupled receptor family. Melatonin receptors are concentrated in the suprachiasmatic area of the hypothalamus and the pars tuberalis of the pituitary. It seems likely that the effects of melatonin on reproductive function are mediated, at least in part, by interval timers associated with melatonin receptors in those areas.

The retinohypothalamic-pineal axis's responses to light are highly conserved in humans. Like other animals, humans secrete melatonin exclusively at night, and they interrupt its secretion when they are exposed to light during the period of nocturnal secretion. In many individuals, the retinohypothalamic-pineal axis also is capable of detecting changes in the length of the night and making proportional adjustments in the duration of nocturnal melatonin secretion, producing the type of melatonin message that other animals use to trigger seasonal changes in their behavior. This response has been shown both in naturalistic studies and in experiments in which melatonin profiles were compared after chronic exposure to long and short artificial “nights.”

The effect of changes in photoperiod on the melatonin message do not depend solely on the acute melatonin-suppressing effects of light; that is, differences in the duration of nocturnal melatonin secretion that are seen under different photoperiods are not simply a function of differences in the extent to which the beginning and end of the light period encroaches on the nocturnal period of melatonin secretion and directly suppresses its initial and terminal phases. Changes in the length of the night to which an individual is exposed alter the intrinsic duration of nocturnal melatonin secretion that is programmed by processes within the organism (presumably within the suprachiasmatic nucleus). That this is the case can be seen when the duration of melatonin secretion is measured in constant dim light following exposures to photoperiods of different duration. Under these conditions the aftereffects of the photoperiod to which an individual has been exposed can be detected in the duration of melatonin secretion even though the individual is no longer being exposed to that photoperiod.

Individuals who live in modern urban environments differ in the degree to which, or even whether, the intrinsic duration of melatonin secretion (the duration measured in constant dim light) responds to seasonal changes in the length of the solar night. Approximately 1 in 3 healthy women living in the District of Columbia area respond unequivocally to seasonal changes in the length of the solar night. By contrast, only 1 in 8 healthy men respond unequivocally. Women also differ from men in that the maximum level of their nocturnal melatonin secretion increases in winter compared to that in summer.

Response to morning light seems to be critical in determining an individual's response to seasonal or experimental changes in the length of the night. Changes in the intrinsic duration of melatonin secretion that are induced by changes in night length are highly correlated with changes in the intrinsic timing of the morning offset of secretion. By contrast, they are only weakly correlated with changes in the intrinsic timing of evening onset of secretion.

Animals use photoperiod-induced changes in the duration of nocturnal melatonin secretion as a chemical signal to trigger seasonal changes in reproduction and other functions. Although humans can produce this signal, it is unclear whether they are capable of responding to it in similar ways. It is difficult to know whether the human reproductive system, for example, is able to respond to the seasonal melatonin message given that only a minority of modern humans who live in urban environments unequivocally respond to seasonal changes in the length of the night by making proportional adjustments in the intrinsic duration of melatonin secretion; investigators who have looked for seasonal influences on human reproduction have studied populations that were not selected for this response. Nevertheless, there are some indications that some humans may respond to this signal. For example, rates of human conception exhibit clear and reproducible patterns of seasonal variation.

Unicellular algae also produce melatonin at night, and they use photoperiod-induced changes in the duration of their nocturnal production to trigger encystment, an adaptive response to conditions that prevail in winter. The fact that humans and algae both use changes in the duration of nocturnal melatonin secretion to code for seasonal changes in the duration of the night points to an ancient evolutionary origin for the mechanism.

Effect of Night Length on Sleep Because modern humans use artificial light to extend their period of wakefulness and activity into the evening hours, they adhere to a short-night sleep schedule throughout the year for most of their lives. In these circumstances, individuals fall asleep shortly after lying down and sleep without interruption until they arise in the morning. This type of sleep, which we tend to regard as the only normal type of sleep, is highly consolidated and efficient, occupying almost all of the nightly period of bed rest.

The sleep of most other animals is quite different. Their sleep and activity are polyphasic, exhibiting multiple bouts per day. Moreover, in many species, including insects, birds, and mammals, average daily profiles of sleep or activity exhibit two prominent peaks, one near dusk and one near dawn. Such bimodal patterns are not necessarily present in all circumstances; they tend to emerge when the relevant behavioral phase (sleep or activity) expands, as can occur after a change in photoperiod.

Until recently, humans were considered to be fundamentally different from all other animals in having sleep that is consolidated into one continuous nocturnal bout. In long nights, however, human sleep resembles that of other mammals to a much greater extent than has been appreciated. In these conditions, humans typically lie awake for one or more hours and then sleep in bouts of several hours' duration that are separated by extended periods of quiet wakefulness. The bouts of sleep are distributed in such a way that when nightly profiles of sleep are averaged, a remarkably symmetrical bimodal pattern emerges, as occurs in other animals.

Circannual Rhythm Certain types of seasonal rhythms in animals, such as rhythms in hibernation, persist in experimental conditions when the duration of the photoperiod and all other conditions are held constant. Apparently those seasonal rhythms are not simply passive responses to seasonal changes in the photoperiod; they are endogenous and capable of self-sustainment. In constant conditions, their intrinsic period is slightly different from one year (hence the name “circannual,” about yearly). In the natural environment the circannual rhythm responds to zeitgebers in such a way that its period is adjusted to 1 year, and a characteristic phase relationship to the seasons is maintained. Annual changes in light, temperature, and diet have been shown to serve as zeitgebers for various types of circannual rhythms.

Feedback Some outputs of the circadian system feed back onto its rhythm-generating mechanisms and modify their oscillatory behavior. For example, pharmacological studies suggest that the onset of melatonin secretion in the evening, which is triggered by the circadian system, may feed back on the system and phase advance it. This effect of melatonin may be mediated by high-affinity melatonin-binding sites that are found in the suprachiasmatic nucleus of the hypothalamus, which is the site of a circadian pacemaker. Experiments have shown that arousing stimuli—such as forced exercise, social interactions, and novel environments—when applied near the time of onset of the activity phase, advance circadian rhythms. This finding suggests that the daily onset of the activity phase, which is triggered by the circadian system, may also feed back on the system and phase advance it.

NEUROBIOLOGY

The circadian system consists of three elements: (1) a rhythm generator or pacemaker, (2) visual inputs to the pacemaker that mediate its entrainment to the photoperiod, and (3) efferent connections that drive and coordinate numerous physiological and behavioral output rhythms. Research over the past three decades has revealed that the retinohypothalamic-pineal axis contains all those elements and is the core of the mammalian circadian system (Fig. 1.13-5).

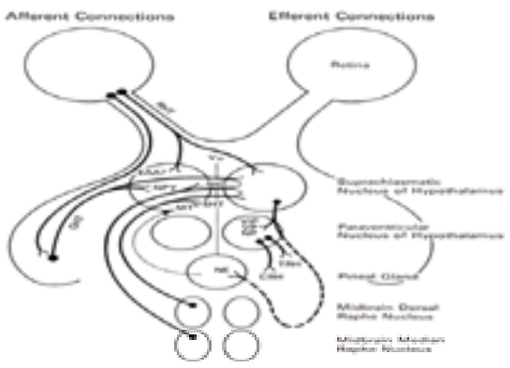

FIGURE 1.13-5 Important pathways and neurotransmitters of the retinohypothalamic-pineal/midbrain raphe axis (schematic drawing, not to scale). Anatomical structures: V_m, third ventricle; RHT, retinohypothalamic tract; GHT, geniculohypothalamic tract. Neurotransmitters: ACh, acetylcholine; EAA, excitatory amino acids; GABA, g-amino butyric acid; NPY, neuropeptide Y; 5HT, serotonin; VIP, vasoactive intestinal peptide; VP, vasopressin; NP, neurophysin; NE, norepinephrine; OXT, oxytocin. Hormones: MT, melatonin; CRH, corticotropin-releasing hormone; TRH, thyrotropin-releasing hormone.

Pacemaker The mammalian brain has a least two circadian pacemakers, but only one, the suprachiasmatic nucleus of the hypothalamus, has been identified. (A second oscillator, which can be entrained by feeding and which survives suprachiasmatic nucleus lesions, has not been localized.) The suprachiasmatic nucleus consists of bilateral compact cell groups lying above the caudal third of the optic chiasm on either side of the third ventricle (Fig. 1.13-2 and Fig. 1.13-5). The evidence that the suprachiasmatic nucleus generates and coordinates circadian rhythms in mammals includes the following:

1. The suprachiasmatic nucleus is directly innervated by the retinohypothalamic tract, a pathway that originates in the retina and is necessary and sufficient for the entrainment of circadian rhythms by light.
2. Suprachiasmatic nucleus lesions disrupt many types of circadian rhythms.
3. The suprachiasmatic nucleus exhibits circadian rhythms in multiunit electrical activity, even when it is isolated from the rest of the brain by knife cuts that create a hypothalamic island in vivo.
4. Suprachiasmatic nucleus explants continue to exhibit circadian rhythms in multiunit electrical activity, vasopressin release, and metabolic activity in vitro.
5. Fetal suprachiasmatic nucleus transplants restore circadian rhythmicity to animals whose circadian rhythms have been disrupted by suprachiasmatic nucleus lesions.
6. When such suprachiasmatic nucleus transplants are obtained from mutant donors whose circadian rhythms oscillate unusually rapidly, the recipient's circadian rhythms also oscillate unusually rapidly.

Intact suprachiasmatic nucleus's have also been shown to be necessary for the expression of photoperiodic seasonal rhythms.

The suprachiasmatic nucleus of a variety of mammalian species share a number of common properties. The suprachiasmatic nucleus can be divided into dorsomedial and ventrolateral subdivisions. The *dorsomedial* subdivision contains neurons exhibiting vasopressinlike and somatostatinlike immunoreactivity. Most vasopressin-containing cells and somatostatin-containing cells seem to be local circuit neurons that do not project beyond the confines of the suprachiasmatic nucleus. The *ventrolateral* subdivision contains a large number of vasoactive intestinal polypeptide-immunoreactive neuronal perikarya. The majority of those cells also appear to be local-circuit neurons confined to the boundaries of the suprachiasmatic nucleus, although there is a small projection to the periventricular and anterior hypothalamic area.

The normal neuroanatomical organization of the suprachiasmatic nucleus is not necessary for its cells to generate circadian rhythms. Dispersed cultures of suprachiasmatic nucleus cells are capable of generating circadian rhythms in vasopressing release, and transplants of those cultures are capable of restoring circadian rhythms in animals whose circadian rhythms have been disrupted by suprachiasmatic nucleus lesions. Those results suggest that circadian rhythm generation in the suprachiasmatic nucleus may be a property of individuals cells. The fact that all successful transplants included vasoactive intestinal polypeptide-containing cells may indicate that those particular suprachiasmatic nucleus cells play an important role in the generation of circadian rhythms.

Circadian rhythm generation by the suprachiasmatic nucleus pacemaker is reflected in changes in multiple-unit electrical activity and metabolism, which increase in the daytime in both day-active and night-active mammals (Fig. 1.13-2). Furthermore, circadian rhythms in that activity persist when knife cuts isolate tissue containing the suprachiasmatic nucleus from the rest of the brain. When splitting occurs in hamsters, the pattern of glucose metabolism becomes asymmetrical in the two suprachiasmatic nucleus.

Clock Genes Recently, two genes that govern the frequency of oscillations of the circadian pacemaker in animals were identified and cloned in humans. One of the genes, *per*, had already been identified and extensively investigated in *Drosophila*. In the fly, positive and negative feedback between the *per* gene and the product of its expression, the *per* protein, appear to be part of a mechanism that generates circadian rhythms. The other gene, *clock*, was first identified in mice. Flies and mice bearing mutations in these genes exhibit alterations in the timing of their daily activity-rest cycles that are phenotypically similar to circadian sleep disorders in humans. Daily rhythms in behavior are abnormally delayed in animals with some types of mutations and abnormally advanced in animals with other types of mutations.

Afferents to the Suprachiasmatic Nucleus Pacemaker The suprachiasmatic nucleus receives three major afferents, a direct visual pathway, an indirect visual pathway, and a pathway from the raphe nuclei of the brainstem (Fig. 1.13-5). All three pathways terminate on cells in the ventrolateral subdivision of the suprachiasmatic nucleus, and all three pathways have been shown to affect the expression of rhythms generated by the suprachiasmatic nucleus.

Direct Visual Pathway Entrainment of circadian rhythms in animals requires intact visual pathways. If the eyes are removed or the optic nerves cut, the circadian pacemaker expresses its own intrinsic rhythms and free runs, even in the presence of a light-dark cycle. The pathways responsible for photic entrainment are unusual in some respects. The threshold for photic effects on entrainment is rather high and resembles that of cone photoreceptors. Genetic mutant mice that lack rods can be entrained perfectly normally to light-dark cycles, suggesting that cones or some novel photoreceptors are responsible for phase resetting by light. The photic entrainment system is also unusual in that it can integrate light over long durations; it responds optimally to pulse durations of 5 minutes.

The retinohypothalamic tract—which passes from the retina along the optic nerve, exits from the optic chiasm, and enters the suprachiasmatic nucleus—is both necessary and sufficient for photic entrainment of circadian rhythms. If the optic tracts, which are distal to the chiasm, are cut, entrainment is preserved.

Neurophysiological studies have shown that the suprachiasmatic nucleus contains many cells whose electrical activity is modified when the retina is exposed to light. The visual fields of those cells are wide, and they code exclusively for luminance (as opposed to spatial characteristics of light that be useful for image processing). With light stimulation, electrical activity is enhanced in some cells and suppressed in others.

Which neurotransmitters mediate retinohypothalamic tract input to the suprachiasmatic nucleus is still not known. Currently, several lines of evidence point to excitatory amino acids. They have been found in the retinohypothalamic tract and its terminals, and their levels are reduced by optic nerve ransection. Furthermore, an *N*-methyl-D-aspartic acid (NMDA) glutamate receptor antagonists can block light-induced phase shifts in rodents.

Photic entrainment may depend on transcriptionally regulated signal transduction processes in the suprachiasmatic nucleus. Light triggers production of mRNA and protein products of the proto-oncogene c-*fos* in the suprachiasmatic nucleus. That stimulation exhibits the same circadian-phase dependence and the same photic threshold as circadian-rhythm phase-shifting effects of light, and the stimulation can be blocked with an NMDA antagonist. Light also stimulates the *jun*-B protooncogene family, which forms a dimer with c-*fos*, called *activator protein-1* (AP-1), which serves as a transcription factor.

Indirect Visual Pathway An indirect visual pathway reaches the suprachiasmatic nucleus through the lateral geniculate nucleus and the intergeniculate leaflet. That pathway is not necessary for entrainment; nevertheless, it appears to be capable of modifying the responses of the suprachiasmatic nucleus circadian pacemaker to photic and nonphotic entraining stimuli. In hamsters, destruction of the intergeniculate leaflet prevents lengthening of the intrinsic period of circadian rhythms that is normally caused by constant light, reduces the magnitude of phase shifts that are induced by light pulses, and advances the time of onset of the nocturnal activity period under conditions of entrainment to light-dark cycles. Furthermore, electrical stimulation of the lateral geniculate nucleus causes phase shifts of circadian rhythms.

Geniculohypothalamic tract input to the suprachiasmatic nucleus appears to be mediated by neuropeptide Y, which is present in cells of the tract and is released in terminals in the suprachiasmatic nucleus when the lateral geniculate nucleus is stimulated. Moreover, injections of neuropeptide Y into the suprachiasmatic nucleus produce phase shifts that resemble those produced by electrical stimulation of the lateral geniculate nucleus. In nocturnal rodents the phase shifts produced by electrical stimulation of the lateral geniculate nucleus and by applications of neuropeptide Y to the suprachiasmatic nucleus resemble phase shifts produced when dark pulses are applied against a background of constant illumination.

In humans neuropeptide Y cell bodies do not appear to be present in the lateral geniculate nucleus. Instead, they are found within the suprachiasmatic nucleus itself, which may indicate that in humans the indirect visual input to the suprachiasmatic nucleus is circumscribed and bypasses the lateral geniculate nucleus.

Pathway From the Raphe Nuclei From the serotonergic raphe nuclei, extensive projections terminate in the suprachiasmatic nucleus, mainly on the vasoactive intestinal polypeptide-containing cells. Further, the suprachiasmatic nucleus show a diurnal rhythm of serotonin uptake and sensitivity. In contrast to excitatory amino acids and neuropeptide Y, which are excitatory neurotransmitters, serotonin is inhibitory and suppresses spontaneous electrical activity of suprachiasmatic nucleus cells.

Although lesion studies have shown that raphe input is not necessary for photic entrainment, the raphe cells do appear to influence the rhythmic behavior of the suprachiasmatic nucleus pacemaker and its responses to light-dark cycles. For example, in rodents housed in constant darkness, lesioning of the raphe causes activity rhythms to disintegrate, which suggests that raphe input may be necessary to maintain the coherent expression of suprachiasmatic nucleus rhythms.

The raphe nuclei also appear to modulate animals' responses to the photoperiod. Animals adjust the durations of the diurnal and nocturnal periods of the circadian systems to match seasonal changes in durations of the photoperiod and scotoperiod. In rodents exposed to a fixed light-dark cycle, lesioning of the raphe nuclei causes the onset of locomotor activity to advance to an earlier time and the offset to delay to a later time such that the duration of the nocturnal activity period increases (as also occurs when the photoperiod is shortened). Lesion studies suggest that the effects of serotonergic input on the times of onset and offset of activity may be mediated by separate tracts—a tract from the dorsal raphe nucleus that modulates the time of activity offset and a tract from the median raphe nucleus, lateral to the other tract, that modulates the time of activity onset.

The idea that serotonergic input from the raphe nuclei modulates or reinforces the SCN circadian pacemaker's responses to the entraining light-dark cycle is also supported by observations that serotonin and the serotonin agonists, quipazine and buspirone, can induce phase shifts in the circadian rhythm of spontaneous electrical activity of the suprachiasmatic nucleus in vitro. Because increases in serotonergic activity have been shown to be correlated with increased arousal in rats, phase shifts induced by arousing stimuli may be mediated by serotonergic input to the suprachiasmatic nucleus from the raphe nuclei. The effects of serotonin on the circadian pacemaker appear to be mediated by a novel receptor, the serotonin (5-hydroxytryptamine [5-HT]) type 7 (5-HT $_7$) receptor.

Other afferent connections to the suprachiasmatic nucleus have been described, but their function and physiological significant are obscure. A few dopamine-containing terminals are present in the suprachiasmatic nucleus, and there are afferents from the anterior hypothalamic nucleus, ventromedial nucleus, arcuate nucleus, and paraventricular thalamic nucleus. There are also numerous commissural projections between the suprachiasmatic nucleus.

Efferent Connections The suprachiasmatic nucleus drives circadian rhythms in locomotor activity, food intake, water intake, sexual behavior, deep body temperature, sleep, adrenocorticotropic hormone (ACTH) secretion, prolactin secretion, melatonin secretion, and gonadotropin secretion. The imposition of rhythmic variation on those functions is probably mediated by efferent pathways from the suprachiasmatic nucleus to networks that regulate those functions.

Efferent projections of the suprachiasmatic nucleus are confined mainly to the hypothalamus. Besides the commissural connections, six efferent pathways to fiber terminals can be identified.

The principal efferent from the suprachiasmatic nucleus is a dense plexus of fibers that terminates in the subparaventricular zone, below the posterior part of the paraventricular nucleus of the hypothalamus. A few of those fibers continue through the paraventricular nucleus and midline thalamic nuclei and end in the paraventricular nucleus of the thalamus. Fibers from the suprachiasmatic nucleus to the paraventricular nucleus and subparaventricular zone contain vasoactive intestinal polypeptide, vasopressin, and neurophysin. The hypothalamic paraventricular nucleus is a way station in the multisynaptic pathway that connects the suprachiasmatic nucleus to the pineal gland and that mediates the effects of light and pacemaker activity on nocturnal melatonin secretion (Fig. 1.13-3). The pathway proceeds through the intermediolateral cell column of the spinal cord and the superior cervical ganglion and terminates in noradrenergic sympathetic fibers on the pineal gland. At night the suprachiasmatic nucleus triggers the release of norepinephrine, which stimulates the synthesis and the release of melatonin by the pineal gland. Lesions of the paraventricular nucleus and the superior cervical ganglion abolish the nocturnal rhythm of melatonin secretion and abolish seasonal photoperiodic reproductive rhythms. The fact that melatonin inhibits spontaneous electrical activity and glucose metabolism in the suprachiasmatic nucleus suggest that the SCN modulates its own activity through a pineal neuroendocrine feedback loop.

Since the paraventricular nucleus contains cell bodies of thyrotropin-releasing hormone neurons and corticotropin-releasing hormone neurons that regulate pituitary thyroid-stimulating hormone secretion and pituitary ACTH secretion, the connections between the suprachiasmatic nucleus and the paraventricular nucleus may also be the route through which the suprachiasmatic nucleus pacemaker imposes daily rhythms on the secretion of thyroid-stimulating hormone and ACTH by the pituitary. Paraventricular nucleus lesions do not interfere with the expression of activity and sleep circadian rhythms.

Other pathways consist of relatively fewer fibers. One pathway runs rostrally and ends in the ventral part of the medial preoptic area. Probably through this connection the SCN imposes circadian rhythms on water intake, deep body temperature, and reproductive behavior. Another pathway runs rostrally through the medial preoptic nucleus and ends in the intermediate lateral septal nucleus. Other fibers terminate in the bed nucleus of the stria terminalis, the parataenial nucleus, the paraventricular nucleus of the thalamus, the midbrain central gray nuclei and raphe nuclei, the lateral geniculate nucleus, a zone near the arcuate nucleus, and the lateral hypothalamic area. Interruption of this last pathway by lesions of the dorsomedial nucleus interrupts the circadian control of food intake without affecting rhythms in body temperature, drinking, and locomotor activity.

Efferent projects from the suprachiasmatic nucleus terminate in the intergeniculate leaflet and in the raphe nuclei, areas from which afferents to the suprachiasmatic nucleus arise. Through those feedback loops, the SCN may modify its own inputs.

Parallels exist between the closely related indoles serotonin and melatonin (serotonin is a precursor of melatonin). The suprachiasmatic nucleus drives daily rhythms in the synthesis and the release of each indole, and each feeds back on the suprachiasmatic nucleus—one acting as a neurotransmitter, the other as a hormone. Each appears to be capable of inducing phase shifts in rhythms generated by the suprachiasmatic nucleus pacemaker, and each appears to play a role, as mediator or modulator, in the pacemaker's responses to seasonal changes in the photoperiod.

BIOLOGICAL RHYTHMS IN PSYCHIATRY

Delayed Sleep Phase Syndrome and its Treatment With Light Abnormalities in the timing of the oscillations of the circadian pacemaker may play an important role in the pathogenesis of some sleep and mood disorders. The evidence for such a role is clearest in the case of *delayed sleep phase syndrome*, a type of circadian rhythm sleep disorder in which individuals have difficulty falling asleep until late at night, but once asleep, tend to sleep soundly and have difficulty arising in the

morning. If they force themselves to rise early, they function poorly during the morning hours and become increasingly sleep-deprived as the week progresses. On weekends, they may sleep late into the afternoon to recover the sleep that they have lost during the week. These individuals appear to suffer not from a disorder of sleep per se, but from a disorder in the timing of the sleep-wake cycle, which remains stubbornly delayed relative to the sleep and work schedule that they wish to follow. Such individuals report that it is difficult or impossible to advance the timing of their sleep to earlier hours. In these individuals the circadian pacemaker seems unable to adopt a normal phase position when it entrains to the day-night cycle; instead, the pacemaker adopts an abnormally delayed phase position. Using principles derived from the human phase response curve for light, clinical researchers have successfully treated delayed sleep phase syndrome by exposing patients to bright light in the morning, which advances their rhythms, and by shielding them from light in the evening, which delays their rhythms.

Some individuals, especially the elderly, suffer from an opposite condition, *advanced sleep phase syndrome*, in which they find it difficult to remain awake in the evening, fall asleep early, and wake up early in the morning. In principal, advanced sleep phase syndrome could be treated by an opposite approach that employs light treatment in the evening and avoidance of light in the morning.

Animals bearing mutations in clock genes exhibit abnormalities in the phasing of their circadian rhythms that are phenotypically similar to circadian sleep disorders in humans. Now that the genes *per* and *clock* have been identified and cloned in humans, it will be possible to examine the frequencies of polymorphisms in these genes and determine whether delayed sleep phase syndrome and advanced sleep phase syndrome are associated with specific variants that affect the frequency of oscillations of the circadian pacemaker.

Treatment of Depression by Phase-Advancing the Sleep-Wake Cycle It is well known that depression and mania are often accompanied by dramatic changes in the amount and timing of sleep. For example, many patients with bipolar I disorder awaken later and sleep more when depressed and awaken earlier and sleep less when manic. In some of these cases, the timing of the sleep-wake cycle shifts later in depression (as in delayed sleep phase syndrome) and shifts earlier in mania (as in advanced sleep phase syndrome).

Experiments have shown that these state-dependent changes in the amount and timing of sleep are not merely symptoms of the illness, but feed back on the illness and play a role in its pathogenesis. For example, it is well known that sleep deprivation can temporarily improve depression and can sometimes trigger mania. Conversely, sleep can trigger or worsen depression, at least in individuals whose depressions improved after sleep deprivation. Less well known is the fact that experimentally altering the timing of sleep, without altering its duration, can also alter mood. Shifting the timing of sleep earlier than usual (e.g., sleeping from 5 PM to 1 AM) can improve depression and can sometimes trigger mania. Conversely, scheduling sleep in the morning can trigger depression, at least in patients whose depressions have improved after sleep deprivation (sleep in the afternoon and evening is much less likely to have this effect). These findings seem to indicate that sleep is depressant and wakefulness antidepressant, especially when they occur in the morning hours. Thus, the timing of the sleep-wake cycle, relative to other circadian rhythms, appears to influence the clinical state of patients with mood disorders.

These observations led to the development of an adjunctive treatment for depression. Patients who have just begun treatment with antidepressant medications are sleep-deprived for one night, then allowed to return to sleep but on an altered sleep schedule in which they sleep from 5 PM to 1 AM (Fig. 1.13-6). Among patients whose depressions improve after a night of sleep deprivation (about 60 percent), the relapse that typically occurs when they resume their customary sleep schedule seems to be prevented when their sleep is shifted to the earlier hours. On subsequent nights, their sleep can gradually be shifted later until they are again sleeping at their customary hours. As patients shift back to their normal schedule, the medication begins to exert its antidepressant effects. This approach to treatment combines the immediate antidepressant effects of total deprivation of sleep, the gradual antidepressant effects of phase-advance of sleep, and the long-term antidepressant effects of medications, so that the patient experiences an improvement that is both rapid and sustained. Some depressed patients respond to the sleep manipulations alone, without adjunctive antidepressant medications.

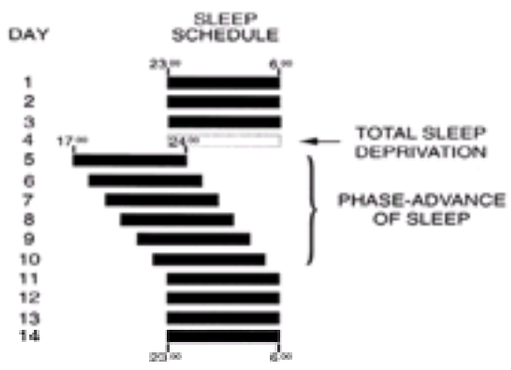

FIGURE 1.13-6 Treatment of depression by shifting the sleep-wake cycle earlier (phase-advancing it), relative to other circadian rhythms, then gradually shifting it back to a normal schedule. The phase-advance treatment is based on experimental observations that sleep is depressant when it coincides with late night and early morning circadian phases, but not when it coincides with late afternoon and early evening circadian phases. (See Wehr TA, Wirz-Justice A, Duncan WC, Gillin JC, Goodwin FK: Phase-advance of the circadian sleep-wake cycle as an antidepressant. Science *206:*710, 1979; Berger M, Vollmann J, Hohagen F, König A, Lohner H, Voderholzer U, Riemann D: Sleep deprivation combined with consecutive sleep phase advance as a fast-acting therapy in depression: An open pilot trial in medicated and unmedicated patients. Am J Psychiatr *154:*870, 1997.)

Light, Melatonin, and Seasonal Affective Disorder In animals, photoperiod-induced changes in the duration of nocturnal melatonin secretion can induce changes not only in reproductive function, but also in activity, aggression, sleep, food intake, weight, metabolism, and other behaviors and functions. Thus, at one time of year an animal may become more active and aggressive, eat more, sleep less, become interested in sex, and interact more with its physical and social environment. At another time of year, it may become less active and aggressive, eat less, sleep more, lose interest in sex, and withdraw from its physical and social environment. It may be relevant that similar seasonal changes in behavior occur as pathology in humans who are afflicted with *seasonal affective disorder* (classified as depressive disorder, seasonal type, in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV]).

Seasonal affective disorder is characterized by recurrent depressions that usually begin in November and end in March. Symptoms of winter depression commonly include lethargy, decreased activity, social withdrawal, loss of interest in sex, oversleeping, overeating, carbohydrate cravings, and weight gain. Sometimes the picture changes to an opposite one in the spring, with increased energy, increased activity, social engagement, increased libido, decreased need for sleep, decreased appetite, and weight loss. The ubiquitous role of melatonin as a chemical mediator of seasonal change in animals and the similarities between patients' symptoms and behaviors that melatonin triggers in animals led clinical investigators to hypothesize that photoperiod-induced changes in the duration of melatonin secretion are part of the mechanism that drives the seasonal cycles of this disorder.

This hypothesis has been tested in a number of naturalistic studies and experiments. For the hypothesis to be tenable, however, its assumption that the duration of nocturnal melatonin secretion is longer in winter than in summer in patients with seasonal affective disorder would have to be shown to be true. Recent research shows that this assumption is true for men with seasonal affective disorder but not for women with seasonal affective disorder. Thus, there is a gender difference in responsiveness of the retinohypothalamic-pineal axis to seasonal changes in the natural scotoperiod in patients with seasonal affective disorder. Curiously, this gender difference is opposite to the one found in healthy volunteers. The results in men are consistent with (but do not prove) the classic melatonin hypothesis of seasonal affective disorder. The results in women, however, appear to be inconsistent with the hypothesis as originally formulated.

The classic melatonin-duration hypothesis of the pathogenesis of seasonal affective disorder has been tested in a number of experiments, with inconsistent results. Supporting the hypothesis are findings that exposure to light improves winter depression, and that exposure of the eyes (not the skin) to light is necessary for this improvement to occur. Also consistent with the hypothesis is the finding that morning treatment with the beta blocker drug propranolol, which suppresses the terminal portion of nocturnal melatonin secretion and thereby shortens its duration, is associated with improvement of winter depression, and the finding that administration of melatonin to patients who have improved during light treatments induces some of the most characteristic symptoms of winter depression. Observations that suppression of melatonin secretion by light is not necessary for improvement to occur during light treatment, however, are inconsistent with the hypothesis.

In future studies of the melatonin-duration hypothesis seasonal affective disorder, it will be necessary to investigate men and women separately, and to determine why patients respond to seasonal changes in the natural photoperiod differently than their sex-matched controls. In the latter effort, the effect of morning light on

processes that regulate the timing of offset of nocturnal secretion are of particular interest, because changes in the intrinsic duration of nocturnal melatonin secretion are mostly a function of changes in the timing of offset of secretion (Fig. 1.13-3).

Monitoring Bipolar Disorders With Continuous Recording of the Activity-Rest Cycle Instruments and data displays that were developed for research on biological rhythms in animals are proving to be valuable tools for automated assessment of the course of bipolar I disorder. To study the behavior and physiology of the circadian pacemaker, researchers have relied primarily on analyses of the expression of its rhythm in longitudinal recordings of locomotor activity (the activity-rest cycle) in animals. These researchers found it convenient to display activity data in the form of a raster plot, in which horizontal 24-hour segments of the recordings are shown consecutively beneath one another. Typically, the data are double-plotted to the right to compensate for the artificial interruption that is created when the data are cut into segments and to facilitate visual inspection of the courses of the activity-rest cycle. This format yields a series of stacked horizontal lines in which Line 1 consists of data for Day 1 plus Day 2, Line 2 consists of data for Day 2 plus Day 3, and so on (Fig. 1.13-7).

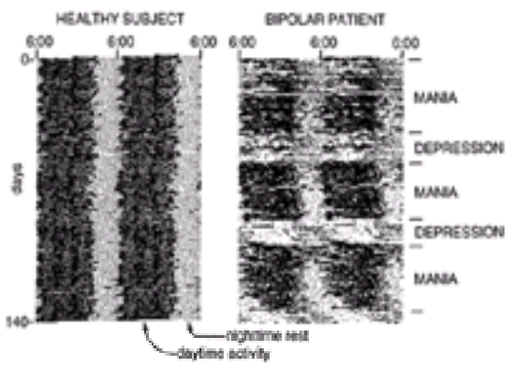

FIGURE 1.13-7 Longitudinal raster plots of wrist motor activity in a healthy individual **(left)** and in an individual with rapid cycling bipolar illness **(right)** (see description of format of plot in text). Activity levels are unchanging in the healthy individual, but change dramatically in the patient, as he switches back and forth between manic (dark areas) and depressive (light areas) phases of his illness. Changes in the timing of the activity-rest/sleep-wake cycle (phase-advanced in mania and phase-delayed in depression) and in the duration of rest and sleep (shorter in mania and longer in depression) can also be detected in the tracing. (Data from Wehr TA, Turner EH, Clark CH, Barker C, Leibenluft E: Treatment of a rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. Biol Psychiatry *43:*822, 1998.)

Portable electronic devices can be used to obtain similar motor activity recordings in patients with bipolar I disorder (Fig. 1.13-7). Such recordings provide a continuous, quantitative, and objective recording of the course of illness that reveals the longitudinal changes in the activity-rest and sleep-wake cycles that are such characteristic features of the disorder. A clinician becomes familiar with gestalt of the raster plot and can readily detect and monitor these features in the compressed and ordered display in which the data are presented. Recently, electronic wrist-activity recording devices that are small, light, and inconspicuous have become commercially available. These devices probably represent a methodological breakthrough for the clinical assessment of bipolar disorders. With their capacity to continuously and inexpensively monitor the course of illness, they could prove highly useful in diagnosis and in the evaluation of response to treatment.

SUGGESTED CROSS-REFERENCES

Sleep and the REM-non-REM sleep cycle are discussed in Section 1.19 on the basic science of sleep. The menstrual cycle and the neuroendocrine systems are discussed in Section 1.11 on psychoneuroendocrinology. Excitatory amino acids are discussed in Section 1.5 on amino acid neurotransmitters; neuropeptide Y is discussed in Section 1.6, and the serotonergic system is discussed in Section 1.4 on monoamine neurotransmitters. Sleep disorders are discussed in Chapter 21 and mood disorders are discussed in Chapter 14.

SECTION REFERENCES

Bartness TK, Goldman BD: Mammalian pineal melatonin: A clock for all seasons. Experientia *45:*939, 1989.

Cagnacci A, Volpe A: Influence of melatonin and photoperiod on animal and human reproduction. J Endocrinol Invest *19:*382, 1996.

*Dijk D-J, Czeisler CA: Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. J Neurosci *15:*3526, 1995.

Gilette MU: Cellular and biochemical mechanisms underlying circadian rhythms in vertebrates. Curr Opin Neurobiol *7:*797, 1997.

Hall JC: Tripping along the trail to the molecular mechanisms of biological clocks. Trend Neurosci *18:*230, 1995.

King DP, Zhao Y, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TDL, Vitaterna MH, Kornhauser JM, Lowrey PL, Turek FW, Takahashi JS: Positional cloning of the mouse circadian clock gene. Cell *89:*641, 1997.

*Klein DC, Moore RY, Reppert SM, editors: *Suprachiasmatic Nucleus: The Mind's Clock*. Oxford University Press, New York, 1991.

Miyamoto Y, Sancar A: Vitamin B_2-based blue-light photoreceptors in the retinohypothalamic tract as the photoactive pigments for setting the circadian clock in mammals. Proc Nat Acad Sci USA *95:*6097, 1998.

Morgan PJ, Williams LM: Central melatonin receptors: Implications for a mode of action. Experientia *45:*955, 1989.

Pittendrigh CS: The photoperiodic phenomena: Seasonal modulation of the ‘day within'. J Biol Rhythms *3:*173, 1988.

Roenneberg T, Aschoff J: Annual rhythm of human reproduction. J Biol Rhythms *5:*217, 1990.

*Rosenthal NE, Joseph-Vanderpool JR, Levendosky AA, Johnston SH, Allen R, Kelly KA, Souetre E, Schulz PM, Starz BK: Phase-shifting effects of bright morning light treatment in patients with delayed sleep phase syndrome. Sleep *13:*354, 1990.

Rusak B: The mammalian circadian system: Models and physiology. J Biol Rhythms *4:*121, 1989.

Rusak B, Bina KG: Neurotransmitters in the mammalian circadian system. Annu Rev Neurosci *13:*387, 1990.

Sassone-Corsi P: Molecular clocks: Mastering time by gene regulation. Nature *392:*871, 1998.

Sun ZS, Albrecht U, Zhuchenko O, Bailey J, Eichele G, Lee CC: *RIGUI*, a putative mammalian ortholog of the Drosophila period gene. Cell *90:*51003, 1997.

Tei H, Okamura H, Shigeyoshi Y, Fukuhara C, Ozawa R, Hirose M, Sakaki Y: Circadian oscillation of a mammalian homologue of the Drosophila *period* gene. Nature *389:*512, 1997.

*Vondrasova-Jelinkova D, Hajek I, Illnerova H: Adjustment of the human melatonin and cortisol rhythms to shortening of the natural summer photoperiod. Brain Res *816:*249, 1999.

*Wehr TA: Effects of sleep and wakefulness on depression and mania. In *Sleep and Biological Rhythms*, J Montplaisir, R Godbout, editors. Oxford University Press, London, 1990.

Wehr TA, Rosenthal NE: Seasonality and affective illness. Am J Psychiatry *146:*829, 1989.

Wehr TA: Melatonin and seasonal rhythms. J Biol Rhythms *12:*517, 1997.

Wehr TA: A ‘clock for all seasons' in the human brain. In *Progress In Brain Research*, vol 2, RM Buijs, A Kalsbeek, JH Romihn, CMA Pennartz, M Mirmiran, editors. Elsevier, Amsterdam, 1996.

*Wehr TA, Turner EH, Clark CH, Barker C, Leibenluft E: Treatment of a rapidly cycling bipolar patient by using extended bed rest and darkness to stabilize the timing and duration of sleep. Biol

Psychiatry *43:*822, 1998.

Textbook of Psychiatry

1.14 APPLIED ELECTROPHYSIOLOGY

EDWARD L. REILLY, M.D.

HISTORY AND OVERVIEW

Electrophysiological changes in biology are now taken for granted by the medical community but the first recognition that there was an electrical action potential in nerves was made by Dr. Du Bois-Reymond in 1848. The first report of electrical activity of the brain was made in August 1875 by Richard Caton, who recorded electrical activity from the brain of a rabbit and a monkey. It would be more than another 50 years before a relatively unknown psychiatrist in Jena, Germany, would record electrical activity from the brain of humans as seen from recordings from the skull surface. Hans Berger first wrote on the electroencephalogram (EEG) of man in 1929, but his work remained essentially unrecognized until the far better known Lord Adrian reported at several meetings in 1934 that Berger's work was to be taken seriously and that his “Berger rhythm” was not just artifact or noise.

Expected frequencies and standards of normal were developed in the 1930s and early 1940s. These standards as currently viewed were frequently overly rigid and led to a higher “abnormal” rate than such observation of normal range of variation yields now.

In some cases it took 20 or 30 years to make a decision as to whether infrequent patterns were indeed abnormal or merely rare. As one looks at patterns such as the 14 and 6 per second positive spike or the 6 per second spike and wave it is clear that this is not a debate that ends easily or with precision.

Alpha rhythm was defined, but its place in normalcy would remain debatable and many studies would be attempted to see whether or not the actual frequency in the alpha range correlated with some aspect of personality, neurosis, introversion, extroversion, intelligence, and other specific features. No replicated correlation was very robust.

Studies were made of normal groups and psychiatric patients or target populations on various features. Sometimes the trait of aggressiveness could be definable or viewed analytically even though observable behavior did not show this feature. In many of these studies various differences were found and correlations were made that other studies were unable to replicate.

Some of these early studies suggested that as much as 15 percent of a normal population had a nonspecific abnormal EEG. In the 1940s better screened, well-controlled populations were found to have a 5 percent nonspecific abnormality rate. Hospitalized psychiatric populations have a higher rate of abnormality (generally twice the rate of a control population) but no specific or consistent feature “identifies” a psychiatric patient, much less a particular diagnosis.

ELECTROPHYSIOLOGY

Reading an electrocardiogram (ECG) of heart action is entirely different from eading an EEG or brainwave test. This is also true for the ability to read sleep studies (polysomnograms), evoked potentials (EPs), or electromyograms (EMGs) (electrical activity generated by muscular activity).

Between the 1940s and the 1960s when there were very few neurologists, it was much more common for psychiatrists to know how to read an EEG, but they often did not perform or read EMGs. In the mid-1940s evoked potentials were recorded. Quantified EEGs are a much newer development, and research is being conducted by psychiatrists, neurologists, and Ph.D. recipients in a number of areas.

Some neurologists are skilled at reading EMGs, EEGs, evoked potentials, and polysomnograms, but not all neurologists can read EEGs. Neurologists are much more likely to have received training in electroencephalography than are psychiatrists, which gave rise to the perception that all neurologists could read EEGs and that all psychiatrists could not. From the 1950s to the 1970s the American Board of Psychiatry and Neurology (ABPN) did not wish to have further subspecialization and still refuses to qualify individuals in regard to a test rather than an area of skill.

The need to demonstrate training and skill in electroencephalography was so great that the American EEG Society in the 1950s developed an independent examination by a separate American Board of Qualification in EEG, now called the American Board of Clinical Neurophysiology, to demonstrate competency in reading EEGs. The exam requires understanding of EEG, EPs, polysomnography, and the basic science and clinical implications of these tests.

In 1972 the American Medical EEG Society, which is more clinically oriented, developed a separate Board to test competency in EEG reading skills. The American Board of EEG and Neurophysiology expects all readers to have an understanding of clinical EEGs and has added qualifications for those interested only in polysomnography, EPs, or quantitative EEG. In the author's opinion the latter Board probably meets the style of training and interest in both clinical practice and research for psychiatrists.

These two Boards, spin-offs from medical organizations that started them, are not sanctioned by the American Board of Medical Specialists (ABMS). Both Boards test candidates predominantly on central nervous system (CNS) electrophysiology. There are differences in the criteria and expectation of breadth of the ability to read CNS activity or more general neurophysiological activity between the two non-ABMS Boards.

In 1997 the American Board of Psychiatry and Neurology decided that neurophysiology did merit a special added qualification. The ABPN developed a written neurophysiology examination leading to an added qualification for candidates who have passed the American Board of Psychiatry and Neurology in either Psychiatry or Neurology. Very soon it will be open only to people who have taken a postgraduate 5th year fellowship in a Neurology program. The exam covers operating room neurophysiology, electromyography, polysomnography, evoked potentials, and EEG. The ABPN examination is a written exam only, and the candidates do not have to prove that they can actually interpret a record that they had never seen before nor do they have to defend how good or bad their readings are as is required in both the non-ABPN examinations.

In the author's opinion it was a mistake for the ABPN, which received little or no supervision from the chairs of psychiatry departments until after the decision was inevitable, to require a postgraduate 5th year of training as a fellow in a Neurology program in order to take the examination. Electrophysiology training leading to Board eligibility should be allowed in Psychiatry Accreditation Council for Graduate Medical Education (ACGME) programs.

ELECTROENCEPHALOGRAPHY

An EEG is a test producing an analog graph of very low voltage electrical activity produced by the brain as viewed from the surface of the skull. The value of EEG increases as the physician ordering and using it recognizes the strengths of the test and learns how to compensate for its weaknesses.

Many new tests are available to evaluate brain structure but EEG remains one of the very few objective tests of brain function. Mental status examination, EPs, positron emission tomography, single photon emission-computed tomography (SPECT), functional magnetic resonance imaging, and psychological tests are others,

but only the mental status examination is quicker than and as safely repeatable as the EEG.

The EEG is produced by cortical activity near the electrodes. Each wave is probably the cumulative result of the activity from 7 square centimeters of tissue. The waves are believed to be a reflection of the excitatory and inhibitory postsynaptic potentials in the dendrite portion of the neurons. The exact source and direction of spread of the measured activity remains controversial. The resultant activity recorded is dependent on the output of the surface gray matter, which is a layer of neurons varying in thickness from 1.5 to 4 mm.

To be reflected in the EEG, disturbance of function must occur or be reflected in the surface cortical neuronal activity. An early tumor or vascular disorder involving the cortex of the convexity of the cerebral hemispheres may cause clear EEG abnormality in the presence of slight clinical signs. A subcortical lesion such as a hemorrhage involving only the internal capsule may produce profound clinical change (dense hemiplegia) in the face of a normal or minimally disturbed EEG. EEG change is manifest early after stroke or trauma and usually subsides; computed tomography (CT) change may not be seen as early as EEG change.

Because the electrical signal is small and depends greatly on the application of the electrode, technique is a crucial issue in the reliability of the activity recorded. A year of training is recommended for the technologist and the interpreting electroencephalographer, but frequently this is not the case. A recent marketing campaign informs physicians that they and an employee can be trained in hours or days to use new and usually expensive equipment to run examinations that the untrained physician will then “read.”

The quality of EEG procedure varies greatly in different laboratories. The American EEG Society and the American Medical EEG Society have outlined minimum standards on EEG but well-defined higher expectations exist and should be used.

Recording

A. Position of Electrodes. The positions internationally recommended are those of the International 10/20 System as outlined by Jasper in 1958 (Fig. 1.14-1). Measurement is made from the front at the nasion to the inion (a ridge below the occipital portion of the skull), and electrode points are at 10 percent and then 20 percent intervals along the total distance, ending with another 10 percent jump in distance. A similar measurement across the coronal or transverse plane from the preauricular notch of each ear allows placement of mid-temporal and central electrodes. Measurement around the head along the 10 percent line locates the occipital, posterior temporal, anterior temporal, and frontal-polar electrodes. Filling in this system over the convexity of the hemisphere allows us to compare a chain of parasagittal electrodes and the midline electrodes to a chain over the temporal region on each side. Odd numbers are on the left and even numbers on the right. The letter refers to the area involved. *C* stands for *central* and the lateral central electrodes are over the central sulcus or Rolandic fissure. This places those two electrodes (C3 and C4) over the motor and sensory strip and on the anterior border of the parietal lobe and posterior border of the frontal lobe. Because the fissure also runs in a postero-superior direction, the midline central (vertex) electrode (CZ) is anterior to the central fissure

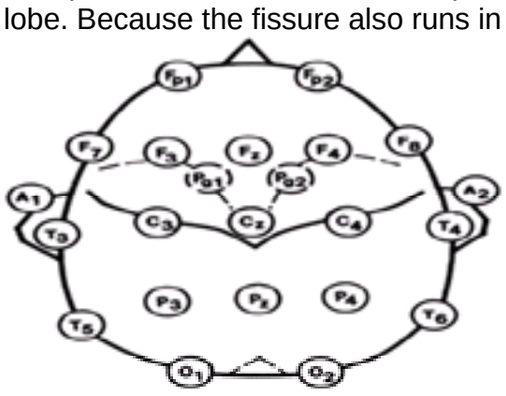

FIGURE 1.14-1 International 10/20 system.

The P3 and P4 electrodes are placed over the posterior parietal lobe. The middle and posterior temporal electrodes are not a source of confusion but the electrodes referred to as *F7* and *F8* (postero-inferior-frontal) are referred to clinically as *anterior temporal electrodes*. They are referenced as such because the activity recorded in abnormal records is generally reflected from the anterior temporal region in spite of the fact that the electrodes are placed over the posterior frontal lobe, which is rarely as active or abnormal as the temporal region.

The increasing use of computer-assisted evaluation of tests in quantified electroencephalography has led to more situations where a greater number of electrodes is believed to be useful. The Electrode Position Nomenclature Committee of the American EEG Society has developed a nomenclature for an expanded number of electrodes that they have designated as “modified combinatorial nomenclature” (Fig. 1.14-2). It will take some time for any standardization on how many additional electrodes will be used in particular tests, such as Fourier Transforms and digital EEGs. The development of the modified nomenclature provides a specific position and name for 75 electrodes that can be placed on the scalp.

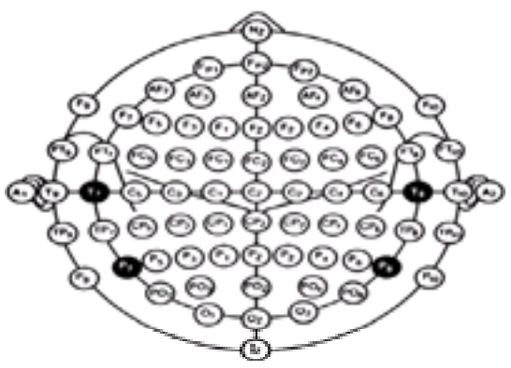

FIGURE 1.14-2 Modified Combinatorial Nomenclature. In this designation four electrode positions in black are given a new name. This system allows for 75 electrode placements clearly designated and defined. The T7 on the left and T8 is a change from the previous T3 and T4. The previous T5 and T6 in this nomenclature is P7 and P8.

For the moment, the use of more than the standard 21 electrodes in the 10/20 system will probably be more research oriented than routine, but that may change as the value of using additional positions becomes better understood.

B. Electrode Type. Cerebral activity is picked up from the skull with metal electrodes, several types of which are in general use.
   1. Disc—silver or gold. These can be attached with paste or glued in place with collodion. The first is removed with hard rubbing and the second with acetone.
   2. Needles—small platinum. These offer fast application but they fall out if the head is moved or sometimes just by their own weight. At one time these were the most commonly sold type of electrodes, but fewer are used now because of concern about the transmission of the human immunodeficiency virus (HIV).
   3. Special
      a. Nasopharyngeal—inserted in the nostril and closer to the temporal region than scalp electrodes. No actual penetration of tissue occurs.
      b. Sphenoidal—needles are inserted between the zygoma and the sigmoid notch until they come in contact with the base of the skull close to the foramen ovale. The needle insertion can reach unintended areas and produce complications. These electrodes are generally used only when surgery for seizure disorders is being considered.
      c. Cortical—surgical implantation of electrodes may involve a subdural grid of electrodes that can be left in place for several days or weeks for further study. Electrodes have been stereotactically planted at various locations in the depth of the brain. Again such surgical involvement most commonly is related to seizure disorder patients who may require surgery for intractable seizures.

C. Record. The EEG machine compares the difference in electrical voltage and amplitude between electrodes; clinical machines do not measure absolute values at a single electrode. Newer computer methods attempt to calculate such absolute value at individual electrodes with an attempt to get a reference-free estimate of activity at each electrode.

Each laboratory sets up patterns of electrodes in runs that are called *montages* (Fig. 1.14-3). These let the reader see activity from different areas of the cortex. Rarely is actual abnormal activity specific to a single electrode.

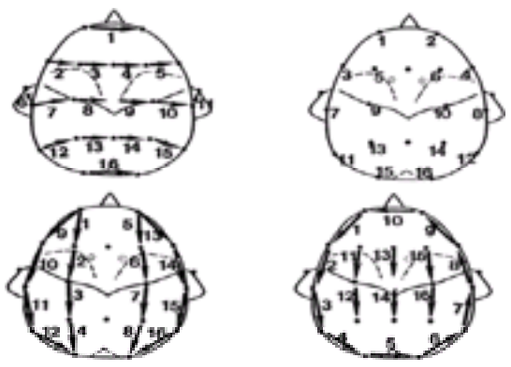

FIGURE 1.14-3 These are four common montages for 16 channel machines. The upper left is considered a “coronal” montage, which is especially useful in seeing subtle temporal asymmetries. The upper right montage is referential, generally done either to ipsilateral ears or linked ears or run to the vertex electrode. When run in left-right pairs, as in this diagram, subtle differences in the same electrode on each side of the head are evident. At lower left a longitudinal montage gives the reader an excellent view of the left and right parasagittal chain of electrodes and compares left and right temporal regional differences. The lower right montage covers the circumference of the head and simultaneously covers the parasagittal electrodes. It is particularly useful in seeing temporal and occipital asymmetries differences while not losing sight of the overall surface activity.

The page of EEG seen in most laboratories moves at 30 mm a second and has dark lines at 1-second intervals and light lines every 200 mm. The wave height indicates the µV involved. The number of μV per mm of pen deflection depends on the setting of the amplifier. The reader cannot identify the amplification unless the record is marked in some way. In most laboratories the letter “S” stands for sensitivity and the number following it indicates the μV per mm. Thus, a small number indicates that it takes fewer microvolts to move the pen one mm. The smallest numbers indicate the highest amplification. Normally, S7 is used to indicate 7 μV per mm but the machine can reach S1 or if needed, as in a “cerebral inactivity” recording, one may make the pen move 1 mm for a 0.5 μV response with S0.5 μV per mm as the amplification.

D. Recording Features. The components of value in an EEG record are:
   1. Frequency (mainly 0.5 cps and 40 cps)
   2. Symmetry
   3. Amplitude (recorded in μV—usually less than 100 μV)
   4. Distribution
   5. Reactivity
   6. Specific patterns

Most abnormal records have some variation in the first four features.

Frequency is the single most important feature of the listed variables.

Electroencephalographers use a shorthand for frequency bands.

Alpha (8 to 13 cps waves over the posterior regions with the patient awake but relaxed with eyes closed)

Beta (any rhythm over 13 cps)

Theta (waves 4 cps to under 8 cps)

Delta (less than 4 cps)

Alpha waves are the most common normal dominant activity in awake adults. Children have slower posterior dominant activity until 7 to 9 years of age, and continue to have slower activity mixed with alpha waves into the early 20s. Alpha waves are the dominant activity in 95 percent of normal adults in the awake resting state.

If the patient is asleep, concentrating, or even focusing with eyes open, alpha activity is absent or at least attenuated.

Beta activity is identified as 14 cps or more. Five percent of normal adults have beta activity as their dominant posterior frequency; it may be widespread, especially if drug induced, in which case it occurs with highest voltage anteriorly. A number of medications can add beta activity to an EEG. In particular, meprobamate (Equanil, Miltown), chlordiazepoxide (Librium), diazepam (Valium), flurazepam (Dalmane), and barbiturates do this. Drug-induced beta activity can turn a search for small spikes into a difficult task.

Theta waves refer to activity from 4 cps to less than 8 cps. Such activity is seen in limited amounts in normal awake adults, especially in the temporal and central region, but becomes more evident in drowsiness. In children under 8 years of age, theta activity is often the posterior dominant frequency.

Delta waves have a frequency less than 4 cps. Delta waves are seen normally in adults only in sleep but occur in a variety of disease states. Delta waves may be from a limited area or “focus.” Focal delta waves in an adult raises the possibility of a destructive process. Such destruction could result from any one of many causes such as infarction, hemorrhage, embolism, tumor, or abscess. All these causes may produce similar EEG changes, but the presence of a focus can be sufficient confirmation to be helpful in considering further, more expensive, and often more invasive testing.

The most common abnormality is slowing of the dominant posterior frequency. In diffuse disease, the frequency of activity slows over the entire brain. (Fig. 1.14-4); in a focal destructive process, the slowing is often localized or lateralized (Fig. 1.14-5).

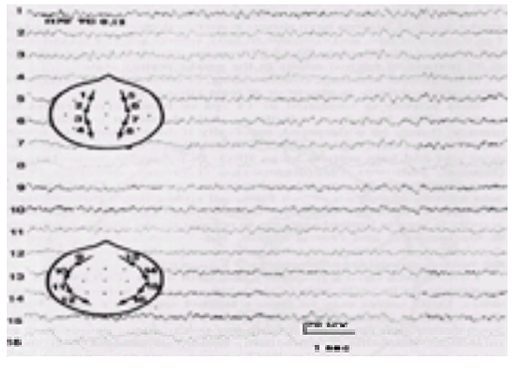

FIGURE 1.14-4 This is a record that is symmetrical but the dominant posterior activity is below normal limits. This is one of the most common abnormalities in EEG and is particularly seen in toxic or metabolic processes or CNS degenerative diseases.

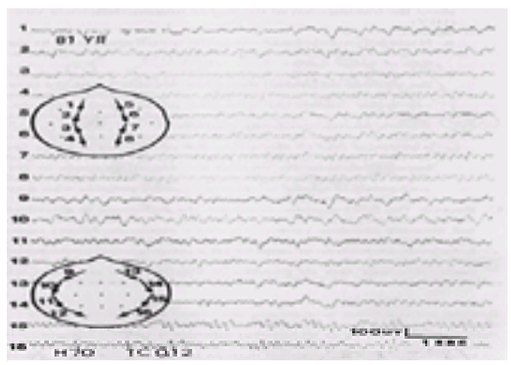

FIGURE 1.14-5 This woman had an infarction and in channels 1 and 2 higher-voltage delta waves are evident from the frontal central region than from the comparable area in channels 5 and 6 on the right. Similarly, there is delta activity from the left anterior temporal region reflected in channels 9, 10, and 11 with higher voltage and persistence than is seen on the right. This record is showing a lateralized abnormality maximal in the left anterior quandrant.

In an area of damaged cortical tissue the slow activity may be constant. If it recurs intermittently with normal activity seen between bursts, it raises the possibility of a subcortical source. As used in EEG, *subcortical* begins below the surface 4 mm of gray matter and extends caudally to the medulla. The EEG reader cannot define the subcortical distance because all activity seen is cortical but may be modified by different subcortical influences. The activity seen is compared with activity expected for the level of consciousness at the time of the EEG; there is a great difference between waking and sleeping activity.

Clinically, sleep is important as an activator of potential abnormalities, especially those seen in patients who have seizures. A great many individuals will show abnormality in sleep but not in the awake record. The abnormality most sensitive to sleep activation is epileptiform activity. *Epileptiform* means “the activity seen is similar to patterns observed between or during seizures in patients who have epilepsy” (Fig. 1.14-6). This type of activity is seen in a high percentage of patients with seizures and a very low percentage of normal individuals. The evaluation of a patient with possible seizures is not complete without sleep activation as well as awake recording.

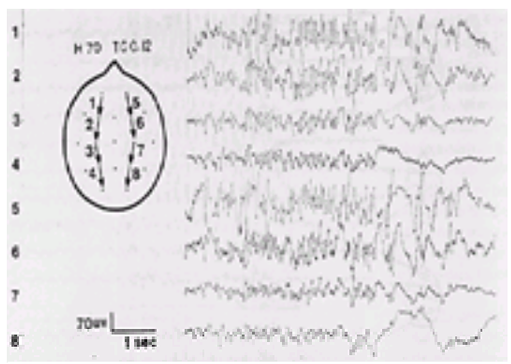

FIGURE 1.14-6 A sleeping patient develops bilaterally synchronous multiple-spike discharges developing over a 4-second period. The multiple spikes are what are seen in the tonic phase of a seizure and alternating spike and slow waves are what are seen in the clonic phase. This very short burst in a patient with such seizures was short and this is only epileptiform activity with no actual seizure and no motion occurring. Actual seizures are associated with a good deal of muscle artifact and the pattern is not as clearly readable during a clinical seizure.

Some reports imply that sleep deprivation is a better activator, but a well-controlled study suggests that recording sleep (natural or from sedation) rather than sleep deprivation is the important factor.

Epileptiform activity generally reflects cerebral irritative processes but does not have a one-to-one correlation with clinical epilepsy or seizures. The absence of epileptiform activity does not mean a patient does not have epilepsy. Persons with epilepsy show epileptiform activity in 50 to 60 percent of initial EEGs if sleep is recorded, and less if it is not. Serial records will demonstrate epileptiform activity in at least one record in 80 to 98 percent of seizure patients, depending on seizure type and if repeated testing is spread over a year or more. Special electrodes (such as nasopharyngeal electrodes for temporal lobe epilepsy) or reduction of anticonvulsant medication may be necessary to demonstrate epileptiform activity.

There are other ways to activate the EEG besides sleep. The two most common in routine testing are hyperventilation and photic stimulation. *Hyperventilation* consists of asking the patient to breathe rapidly with mouth open for a 3-minute period and then to breathe normally. Hyperventilation may stress the brain function just sufficiently to demonstrate a lateralized change or in children with an absence seizure may actually provoke a seizure with its characteristic changes in the record (Fig. 1.14-7).

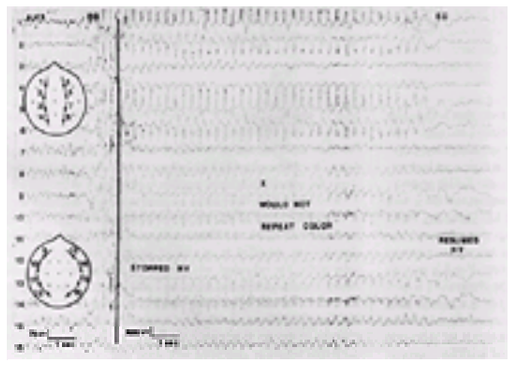

FIGURE 1.14-7 The slide begins with a patient hyperventilating with eyes closed. As the epileptiform burst begins, the patient's eyes open without command. There is increased voltage and amplification is rapidly decreased so it takes 200 μV, not 75 μV to move the pen 1 cm. The patient does not respond to the command. When the burst ends, the patient's eyes close and the patient begins hyperventilating again without command. This is a clinical absence seizure documented by EEG.

Photic stimulation similarly can provoke normal “driving” responses where the occipital dominant activity matches or is a harmonic of the frequencies. Flashes may provoke muscle responses, which are rare but not abnormal, or epileptiform discharges, which if prolonged beyond the flash tend to strongly indicate the possibility

that the patient has a potential for seizures and in most instances has a history of seizures (Fig. 1.14-8).

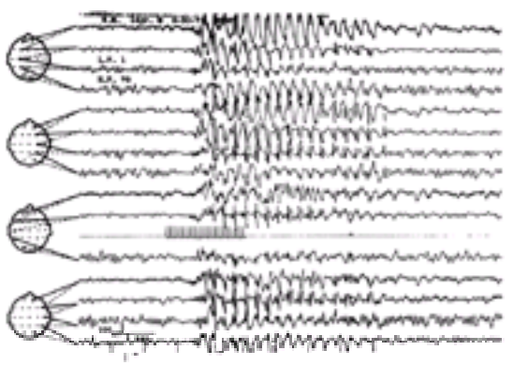

FIGURE 1.14-8 This is a patient with a paroxysmal response to flashes at a different rate than the flash rate. When the flash stops, the paroxysmal activity continues. Prolonged paroxysmal response to photic stimulation beyond the flash is highly correlated with seizure disorder but can be seen before the first spontaneous seizure in some patients.

Since the EEG is a reflection of function, it can change from day to day and at times from hour to hour. Serial tracings assist in sorting out the picture. Diffuse processes such as encephalopathies may show focal areas of slow activity superimposed on diffuse slow activity. Such lateralization raises the question of focal problems but in these patients the asymmetry is not consistent and it usually changes from record to record and day to day if it is really a diffuse process.

The basic posterior dominant frequencies tend to show a trend toward or away from normalization that is helpful in judgment about effectiveness of therapy.

Abnormality may be:

Diffuse—suggesting a process that involves the entire cortex or perhaps the thalamus or brainstem as suggested by “bilateral simultaneous slow waves anteriorly.” Only cortex activity is being recorded but bilateral symmetry suggests it is driven from subcortical sites (Fig. 1.14-9).

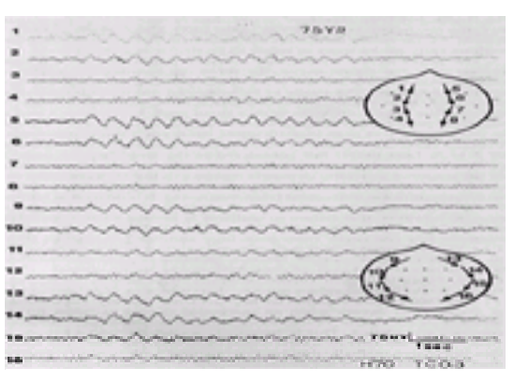

FIGURE 1.14-9 This record demonstrates bilaterally synchronous frontal central trains of 21/2- to 3-second delta activity referred to as *frontal intermittent rhythmic delta activity* (FIRDA). This synchronous activity in an awake individual raises the possibility of the involvement of deep midline structures somewhere from the medulla to the thalamus.

Focal—the majority of times the EEG can lateralize a focal process to a hemisphere. Often it can be further localized.

a. Quadrants—a lesion may involve the anterior or posterior half of one hemisphere and the lesion will be reported as maximal in a particular quadrant.
b. Temporal—parasagittal distinctions. Most often the temporal area reflects abnormality more than the parasagittal area. As a result it becomes necessary to weigh minor parasagittal changes heavily to compensate for this tendency of the temporal area to seem more involved.

Structural lesions have more stable asymmetries that persist in the same location; they may be increasing in serial records. An example of the latter would be an abscess that becomes more evident in serial records.

Another source of generally progressive focal change are tumors, but changes can be masked with treatments such as steroids. Tumors often show focal slowing. The abnormalities in the EEG cannot be distinguished in any definitive way from the changes from focal vascular problems. As a class, however, tumors are more apt to have sharp waves or spikes mixed with the slow activity than are vascular diseases (15 percent of tumor patients have seizures and only 7 percent of cases of even acute cerebral vascular accidents have them). Often the first EEG record, CT scan, and initial brain scan is abnormal in patients who have tumors. After 2 weeks the EEG is generally (but not always) worse or not improved in contrast to what is seen in patients with vascular disorders where the EEG generally shows a rapid trend toward improvement and perhaps normalization.

The EEG is immediately at its worst after a thrombosis or hemorrhage. It generally shows improvement with time. Possibly normalization will occur in a 2- to 6-week period. The abnormality is most often focal slow waves, which may be evident early at a point when the CT scan and brain scan results are still normal. In a day or so the CT scan and by 10 days or 2 weeks, a brain scan is apt to become abnormal and the EEG will already be improved. It is easy to suggest a vascular cause for any focal change in an adult population. It is better for the EEG reader to provide the characteristics and changes of the record and let a clinician knowledgeable in such subtle changes compare them to the clinical findings.

INITIAL USE OF EEG

It was recognized early in the recording of EEG that the patterns changed when patients went to sleep. The first classification of sleep was first suggested in 1937. In 1953 researchers noted periods of rapid eye movements, which were soon referred to as *REM*. It is when this activity is present that an individual experiences dream activity. Arousal during REM is usually associated with an awareness of the dream content but if the REM period is not interrupted, the dream will not be remembered later. Individuals who have been asked by their therapist to report dream activity and write down the dream as soon as they wake up often find that they wake up during dreams more frequently. They also find that they may not remember the dream content in the morning until they refer to their notes. Following the identification of this stage, much work was done on issues related to REM sleep versus non-REM (NREM) sleep. At one point there was some concern that lack of REM sleep might in fact provoke psychotic behavior. But those early suggestions were not replicated, and it is recognized that a number of medications can eliminate REM and not be associated with ill effects. Stages 3 and 4 sleep are characterized by high-voltage delta activity and these two stages are sometimes lumped together as *slow-wave sleep*.

Specialized sleep laboratories are used to identify disorders such as narcolepsy, sleep apnea, and nocturnal myoclonus (periodic leg movements of sleep). In 1986 an entity called *REM sleep behavior disorder* was identified as an unusual situation occurring often in older adults. The individual, while dreaming, does not have the usual loss of muscle activity and therefore could and often does act out the dream with activity involving jumping, pushing, hitting, or kicking. Sleep laboratories identify these syndromes and can then suggest effective treatments to minimize the loss of rest and the sometimes dangerous behaviors associated with some sleep

syndromes.

Sleep research is being done to compare medication effects in double-blind studies of long-term use of various treatments. Nortriptyline (Pamelor) used at maintenance doses in elderly patients with depression demonstrated increased REM activity generation and increase delta-wave production in the first NREM period of the night when compared to patients on placebo. It has become apparent that even antidepressant medications in the same class may have different effects on the sleep pattern. Better understanding and awareness of these differences will be useful in drug selection in the future.

In addition to laboratories designed to look at “disturbed sleep patterns,” there are laboratories designed to record ongoing EEG activity in the awake and sleep state as part of the investigation of epileptiform activity. More importantly, attempting to get better localization of the onset of actual clinical seizures. These laboratories tend to be distinct from the typical sleep laboratory, but a 24-hour EEG monitoring may be done in these specialized departments.

EVOKED POTENTIALS: BRAIN RESPONSES TO SPECIFIC STIMULI

Localization and lateralization are important aims of structural tests such as CT scanning and functional tests such as the EEG and psychological tests. Development of localizing and lateralizing techniques that are noninvasive, independent of patient cooperation, and specific as to location as well as cause has been a goal, but for electrophysiology this is only minimally realized.

The measurement of electricity from the surface of the skull provides a large array of ongoing activity of very low voltage. Buried under that already low-voltage signal there are some additional changes of extremely low voltage that can be demonstrated to occur in a very time-related fashion after a specific external or internal stimulus. The evoked potential after such a stimulus meets some localization goals. The response to a specific stimulus is called an *evoked response* (ER), an EP, or an *event-related potential* (ERP). The latter has a longer latency and is influenced more by level of arousal and medication than the shorter-latency ER or EP responses. The evoked potential is an attempt to separate specific reactions of the CNS to a stimulus from ongoing nonstimulus-related EEG activity. If one samples regularly from EEG and averages each time epoch recording, there is as much chance that the EEG wave will go up as down at any point in the epoch because EEG is a random wave form; it will average out to a flat line or 0. If the samples are not random but triggered by a stimuli, the event following and repeating each stimulus may stand out from the larger EEG signal, which averages out. Recently the increased degree of computer sophistication to average out some of these responses has made it possible to record these events easily. Newer methods have provided recordings as clinically relevant modalities. An EEG tends to be 30 to 60 μV, and the small evoked potentials are in the 0.1 to 0.5 μV range. Taken in the perspective that less than 2 μV of electricity was taken 20 years ago as the level suggesting no brain function or brain death because machine limits were reached, better amplification and artifact reduction was needed to demonstrate these responses. The measurement of this tiny activity is much easier to demonstrate with computers that average many responses, better amplifiers, and better techniques to avoid artifact.

The EEG depends on surface cortical activity, but the evoked potential can demonstrate transmission time through the brainstem with the *brainstem auditory evoked potential* (BAER) and the *somatosensory evoked potential* (SEP). Transmission through hemisphere pathways is measured by the *pattern reversal visual evoked potential* (PRVEP). Thus, EEGs that measure only surface function and the changes following stimulation, sleep, arousal, or drug effect can be complemented by an accurate measurement of the transmission time, and the consistency of the response to a specific sensory stimulus.

There are three major stimuli being used in the clinical electrophysiology setting.

1. pattern—reversal of a checkerboard—VEP (PRVEP)
2. click stimulus by earphone—BAER
3. electrical stimulation of the median nerve—SEP

These tests provide a measure of functional integrity or lack of it in situations where documentation often has been lacking in the past. Patients with multiple sclerosis display delayed transmission because of myelin loss, and the visual EP may show prolonged latency even in the absence of visual complaints; other causes affect these potentials as well.

There is a great rush to sell “easy, foolproof” equipment. Some of the machines are so good that they are easily and reliably operated, but no machine is immune to the “garbage in, garbage out” rule. If the data are bad, the most skilled reader is blind.

A basic problem is that the responses change with each variation of stimulus or equipment setting. The visual evoked potential with bright flashes to closed eyes is more complex and less reproducible than the pattern-reversal visual response, which is simpler in its shape and more replicable in repeat collections. This does not mean that the flash response is useless, but the sparse plain response to pattern reversal is easier to quantify as normal or abnormal (Fig. 1.14-10). It is valuable in clinical use, and standard deviations from normal by age exist.

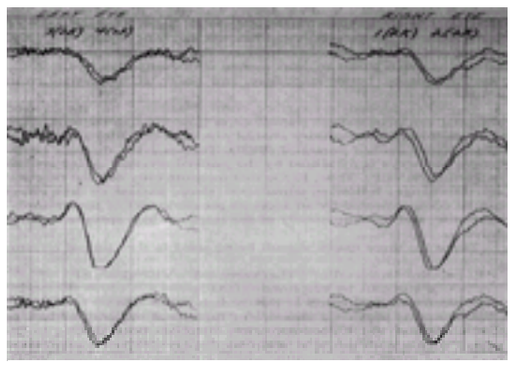

FIGURE 1.14-10 This is a visual evoked response to pattern reversal of a checkerboard. Notice that there are two lines in that the series of stimuli were given twice to each eye to see if the waves reproduced. The reproducibility is excellent and the latencies in this evoked potential are normal. The total field seen is 200 ms.

After the introduction of PRVEPs, the next clinically useful EP followed the observation that very early responses to clicks had identifiable waves that corresponded with anatomical generators in the auditory pathway. The response is suitable for measurement of brainstem function even more than it is a measure of hearing. This modality brought new terminology into the language. The visual EP is clearly dependent on the cortex underneath the electrode. Visual EPs in the midline occipital region are often different from lateral electrodes both in wave form and amplitudes, although not necessarily in latency. The cortex-dependent responses are “near field” and most responses 60 ms or more after a stimulus are near field, depending on the effect on the cortex under the electrode.

In contrast, the response averaged in the first 10 ms after auditory clicks is similar when recorded from almost any cortical location. The peaks are generated from sources so far from the cortex that they are equally far from all the electrodes and these are “far field” potentials. Far field potentials are short in latency and more reproducible than long latency responses. The BAER is much less sensitive to change from level of consciousness or drug effect. At the extreme, the BAER has been preserved in overdose situations severe enough to cause electrocerebral inactivity in the EEG. The anatomical correlations are far better than with the near field potentials in that the various early peaks seem to relate to particular brainstem areas.

The peaks in the BAER that are most replicable are the one representing the eighth nerve, the three representing the superior olivary level of the pons, and the five representing the inferior colliculus in midbrain. The peaks to be measured are those on the side stimulated. Because of cancellation effects, the wave referenced to the opposite side is different with a small peak 1 and peak 3 and a separated peak 4 and 5 (Fig. 1.14-11). The opposite response is useful to record to help in peak identification, but is not actually measured. Items of value are the latency between peak 1 and 3, and the latency of 3 to 5. Less useful but considered is the amplitude of peak 5 compared to peak 1. Reproducibility is critical. The significance of the response generally demands reproducibility between separate collections. Movement distorts the waves as does muscle activity without significant movement (clenched teeth). It follows that failure to replicate raises the possibility of significant artifact. It is not proven, however, that failure to replicate is always artifact; patients with stable early components followed by variability of later peaks suggests this can occur from pathology and not just artifact (Fig. 1.14-12). Differentiation is difficult, and extreme care must be taken to attempt to get records of relaxed, motionless patients.

Sedation may be necessary. Level of consciousness is not as critical as for EEG and long-latency EP because sedation effect does not change the BAER wave. Normal values must be determined in each laboratory.

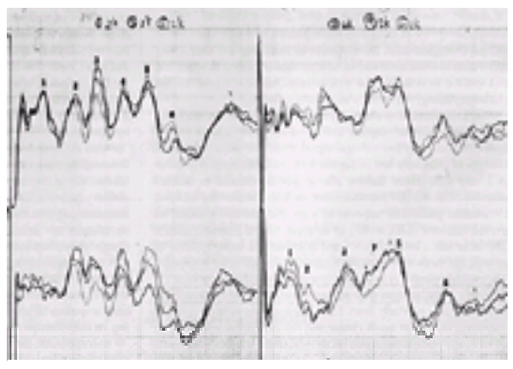

FIGURE 1.14-11 A brainstem auditory evoked potential (BAER). The top line is recorded from the side where the clicks occurred and the bottom line is from the opposite side of the head. The AER is a far-field response. The waves measured are on the side where the clicks occurred and the other side is used to help identify peak 1 and peaks 4 and 5, which look different on the two sides. Recording both sides makes peaks more easily identifiable. The recording is of the 10 ms after the clicks.

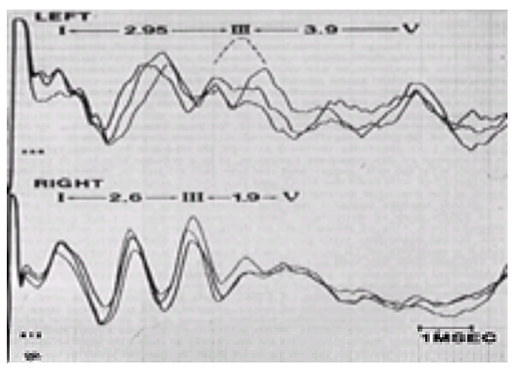

FIGURE 1.14-12 Abnormal BAER on one side. On one side there is a variable response. The early components replicate but the variability of later peaks is abnormal and the latency from peak 3 to peak 5 is prolonged compared to the right where reproducible waves occur at normal latency.

The P300 commonly used EP in psychiatric research populations is a wave that occurs when an individual is told to listen for or look for an infrequent stimuli mixed in with a much more frequent common stimulation. This is frequently a click or tone when done with sounds and may be a different symbol or visual cue. The attention to the rare oddball stimulus is associated with a widely distributed wave occurring about 300 ms after the stimulus. In alert attentive individuals, this is a bigger and earlier wave form than is seen in the populations that are less alert or not really interested. Testing for this particular wave has been done to assess for mental function as well as to test for interest and attention.

There is preliminary work that suggests that the P300 may have reduced amplitude in some patients with schizophrenia and seems to be associated with structural abnormality in the medial temporal lobe.

In a group of medication-free men with chronic alcohol dependence as compared to a control group, there was reduction in the P300 amplitude that may reflect a deficiency in an inhibitory mechanism that may underlie P300 generation.

In special instances, the following more limited and sometimes more controversial stimuli may be used:

1. tones at various frequencies (audiology)
2. electrical stimulation of peroneal and tibial nerves
3. visual stimulation with flashes, hemifields, or different colors.

There are major technical differences in machines and in the way responses are recorded. Machines designed by engineers tend to put positive up; EEG uses negative up. Both engineers and neurophysiologists put positive up in some specific instances. The BAER is usually positive up because the latencies used relate to the surface positive peaks.

The presence and absence of similar components that do or do not cancel is a major factor in what is displayed. Technique includes decisions about filters. A move from one high filter (eliminates higher-frequency activity from analysis) to another can cause a shift of several milliseconds in the apparent latency of evoked potentials.

Terminology may name a peak by its sequence or by its expected latency. As latencies change, it is difficult to identify the particular peaks and this remains a major problem at times in interpretation of long latency near field responses.

The P300 wave has been studied in many psychiatric and organically afflicted patients. Because of its long latency, it can be considered an ERP. The response is effected by incentive, significance of the stimulus, and other factors but can be recorded even in infancy (Fig. 1.14-13).

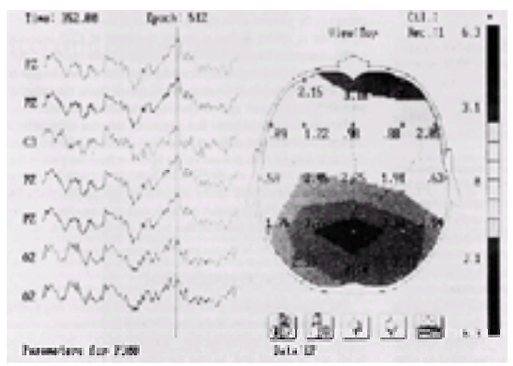

FIGURE 1.14-13 Brain map of the P300 peak. To the left is the evoked potential, to the right a map at the msec where a line is drawn on the potential. The p300 response to an oddball signal is fairly symmetrical and widespread. To read an evoked potentials look at the evoked potential to the left, it is not necessary to have a map. The map gives the viewer an idea of what is going on over the entire head and not just at the recording electrodes but it can be deceptive because it is a particular millisecond view of what is going on over 200 to 600 ms. Moving back 10 ms or forward 10 ms and the picture may change considerably; laterality can change a little bit. The skill of the person picking the point for the illustration is crucial in determining clinical accuracy or error.

Latency is prolonged in patients with dementia but amplitude changes are suggested in uninterested, depressive, and alcoholic patients. It is still of greater research than clinical value, but a useful tool.

MEDICATION

Medication changes in the EEG have been measured for many years, and some of the first computer analysis of drug effects on EEG by the early computers was done with ‘period analysis.' This methodology measured baseline crossings. The actual wave (how high it went, how fast it went, and whether it was pointed at the top or not) was not measured but a baseline of the EEG was determined for the electrical activity. Each time a wave passed above or below that baseline, a notation was made. With this period analysis computer methodology, changes were made before and after drug administration. This method used far less computer energy and memory than is now available. The methodology looked only at frequency change and yet was able to provide signatures of the effect an antidepressant drug had on the EEG as opposed to the effect of an antianxiety drug. By current standards this is a rather simplistic methodology, but it allowed drug companies to look at changes in the EEG in test animals and to identify potential antidepressants and antianxiety drugs to screen further. This EEG screening of drug effects on the brain by period analysis helped to enlarge the role of EEG in psychiatry. Period analysis suggested that changes in the electrophysiology of the cortex in response to drugs was of value.

Using a single-dose drug effect and looking at EEG changes to determine whether to use the medication or try another has not really been proven to be of significant value in clinical practice. Robust evidence that this can be done with a high enough sensitivity (the percentage of people with the disease who are so identified by the test) plus sufficient specificity (the percentage of people free of the disease to be so identified by the test) to be clinically useful does not yet exist.

However, some prediction can be achieved. The attempts to predict response to treatment, whether it be to medication or to cognitive therapy or to any other kind of therapy, certainly need to be continued as a source of interest and study. It is possible that in the future electrophysiological changes may prove to be helpful in determining whether a psychopharmacological drug that is known not to show significant clinical effect for 10 days to 2 weeks or 6 weeks is likely to be effective.

QUANTITATIVE OR COMPUTERIZED ELECTROENCEPHALOGRAPHY

The term “quantitative EEG” is generic and is probably currently the best term to apply to a wide range of analysis of EEG. Some of this analysis such as fourier transforms and spatial analysis has been conducted for several years but is now accomplished faster and for more patients by advanced computer analysis and digital recording so it is more cheaply available for the study of EEG and evoked potentials.

Some of the early studies were conducted without computers and with much more laborious quantifications on small numbers of patients. Much of that work was called *topographical mapping*. Currently, quantitative EEG includes evoked potentials analysis, spectral analysis, fourier transforms, digital EEG, discriminant function analysis, and other entities. Some of the changes in this part of the field are the result of technological changes in collecting data plus changes related to computer analysis of the recorded data.

Some mapping data is merely recording data that was available in the EEG since the 1930s, but such data is displayed as a color picture like a weather map. The latter is useful, probably accurate, but leaves the naive viewer in a position to misinterpret or to overvalue the changes if the display makes the changes look more important than they really are. Some early color maps were deceptive because the color changes were so abrupt as to appear more significant than the change between bands really represented. Some companies advertising their machines encourage psychiatrists in particular to use the responses as a way to make psychiatric diagnosis. These companies should warn physicians that they are required to have a fairly significant understanding of electroencephalography and be able to review the EEG and not just the printed map or discriminate function.

The misuse of quantitative EEG for unnecessary billing, or the misuse of discriminate function as a screen rather than a way to separate two carefully defined groups should not be interpreted as a weakness of discriminate function. It is a rebuke leveled at practitioners who do not understand the test and its limits. Assuming that a technician will read the EEG and that the psychiatrist only needs to look at the machine output is not good medicine or good psychiatry.

Some studies suggest that the combination of quantitative EEG and ERPs may contribute to better discrimination in schizophrenia populations. Use of these tests to differentiate between schizophrenia, mood disorders with psychotic features, and controls presently has a sensitivity of 78 percent and a specificity of 85 percent, which is encouraging but still not definitive in a clinical population. With additional components perhaps better sensitivity and specificity can be reached in some psychiatric populations.

While digital recording of EEG lends itself as a first step for further analysis in a variety of ways, digital EEG recording may be of value to a practitioner who wishes to use it for no purpose other than reading standard EEGs. With adequate size and detail from the monitor screen, the record can be recorded at one site and stored in nonpaper form such as on an optical disc. It can be read at another site either at the time recorded or some later date. Once the record is digitized, other manipulations such as montage change are available.

As psychiatrists become more knowledgeable about the strengths and limitations of these tests and the interpretation of them, quantitative EEG will add to the evaluation of psychiatric patients, but as yet it is not likely to determine an Axis I or Axis II diagnosis; it has been known for years that an EEG may suggest an Axis III component.

It is becoming apparent that the very sophisticated analysis of EEG changes in frequency where there is only medication or state change of the patient can be used much more in some patient populations than in others. Before a fourier transform of EEG becomes useful, it requires totally artifact-free recording of the cerebral activity. Chewing movement, eyeblink, and drowsiness all must be absent from the EEG that is to be interpreted, which limits the value of such tests in confused, agitated, heavily medicated patients.

Psychiatrists should know that a patient's behavior changes with different illnesses and different phases of emotional disturbances. Research designs must demand only a behavior from the patient that is reasonable for each emotional state. Inability to control a patient in the test situation may render an excellent test inappropriate for certain patients. The physician ordering such tests in clinical or research situations must consider that before any testing is done.

MAGNETOENCEPHALOGRAM

The magnetoencephalogram measures ionic currents originating from biochemical sources at the cellular level in the central nervous system. Instead of measuring the “electric fields,” the magnetic fields are measured and, as was the case with EEGs in the early years, there are technical problems. Sufficiently sensitive transducers were not available until the late 1960s. Present-day equipment and techniques do not draw great clinical utilization and may not do so until a number of questions are answered.

In the early 1970s a superconducting quantum interference device (SQUID) was developed that avoided the use of any electric reference and led to more rapid development. In the early 1990s there were only about 50 laboratories with magnetoencephalography capability and 24 channel magnetometers were in use. Helmets that cover the whole head involving as many as 122 channels are being developed and possibly as many as 150 channels could be used to cover the whole cortex in the not-too-distant future. The strengths of magnetoencephalography are:

1. Magnetoencephalography appears to be a better way to find the precise source of a response to stimulus or the cortical activating area provoking some sort of change.
2. Magnetoencephalography evaluations help to better demonstrate the origin of some of the electroencephalographic activity than can be determined by an EEG from the surface of the skull.
3. Some patterns may be specific to magnetoencephalography and are not seen in the EEG or at least are seen quite differently.
4. At present the clinical value of magnetoencephalography is relatively limited but its research value appears to be unlimited.
5. The greater accuracy in identifying source is going to make magnetoencephalography a major advance in identifying specific areas and in separating areas of

function such as speech that are near an area that needs to be excised because it is a seizure focus.

The so-called inverse problem is one particular area where magnetoencephalography seems to have a definite advantage over electroencephalography. It is crucial to find the intercranial sources of potentials recorded at the scalp and particularly to be able to relate them to particular neural generators and neural events. Magnetoencephalography techniques seem to be more accurate at finding the source on the cortex that is affected by a particular stimulus response or motor event. This alone makes this modality a significant and major tool to demonstrate true cortical response areas for sensory and motor events.

Magnetoencephalography is not yet a great threat to the usual clinical EEG recording because of the former's cost and limited availability, but it certainly offers the potential for greater localization when properly applied to research and clinical situations.

SUGGESTED CROSS-REFERENCES

Basic electrophysiology is discussed in Section 1.9, and Section 2.3 is on the neuropsychiatric aspects of epilepsy. Section 7.7 is on medical assessment and laboratory testing in psychiatry. Chapter 10 discusses delirium, dementia, and amnestic and other cognitive disorders and mental disorders due to a general medical condition.

SECTION REFERENCES

American Clinical Neurophysiology Society: 1998 Guidelines in EEG. J Clin Neurophysiol *15:*377–423, 1998.

American Clinical Neurophysiology Society: 1998 Guidelines in EEG. J Clin Neurophysiol *15:*464, 1998.

American Electroencephalographic Society, Guidelines Four: Standards of practice in clinical electroencephalography. J Clin Neurophysiol *11:*14, 1994.

American Electroencephalographic Society, Guidelines Fourteen: Guidelines for recording clinical EEG on digital media. J Clin Neurophysiol *11:*114, 1994.

American Electroencephalographic Society, Guidelines One: Minimum technical requirements for performing clinical electroencephalography. J Clin Neurophysiol *11:*2, 1994.

American Electroencephalographic Society, Guideline Thirteen: Guidelines for standard electrode position nomenclature. J Clin Neurophysiol *11:*111, 1994.

American Medical Electroencephalographic Association: Electroencephalography standards. *1*, 1988.

American Psychiatric Association Task Force on Quantitative Electrophysiological Assessment, Quantitative Electroencephalography: A report on the present state of computerized EEG techniques. Am J Psychiatry *148:*961, 1991.

Antonowicz JL: Missed diagnoses in consultation liaison psychiatry. Psychiatr Clin North Am *21:*705, 1998.

Callicott JH, Egan MF, Bertolino A, Mattay VS, Langheim FJ, Frank JA, Weinberger DR. Hippocampal N-acetyl aspartate in unaffected siblings of patients with schizophrenia: A possible intermediate neurobiological phenotype. Biol Psychiatry *44:*941, 1998.

Collura TF: History and evolution of computerized electroencephalography. J Clin Neurophysiol *12:*214, 1995.

Degen R, Degen HE: The diagnostic value of the sleep EEG with and without sleep deprivation in patients with atypical absences. Epilepsia *24:*557, 1983.

*Drake ME: Clinical utility of event-related potentials in neurology and psychiatry. Semin Neurol *10:*196, 1996.

*Duffy FH, Hughes JR, Miranda F, Bernad P, Cook P: Status of quantitative EEG in clinical practice. Clin Electroencephalogr *25:*6, 1994.

Garber JH, Weilburg JB, Duffy FH, Manschreck TC: Clinical use of topographic brain electrical activity mapping in psychiatry. J Clin Psychiatry *50:*205, 1989.

Gloor P, Berger H: *On the Electroencephalogram of man: Electroencephalography and Clinical Neurophysiology*, supplement 28. Elsevier, Amsterdam, 1969.

Hari R: Magnetoencephalography as a tool of clinical neurophysiology. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

*John ER, Prichep LS: Principles of neurometric analysis of EEG and evoked potentials. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Knott V, Mohr E, Haché N, Mahoney C, Mendis T: EEG and the passive P300 in dementia of the Alzheimer type. Clin EEG *30:*64, 1999.

Marsan CA, Zivin LS: Factors related to the occurrence of typical paroxysmal abnormalities in the EEG records of epileptic patients. Epilepsia *11:*36, 1970.

Morihisa JM, Duffy FH, Wyatt RJ: Brain electrical activity mapping (beam) in schizophrenic patients. Arch Gen Psychiat *40:*719, 1983.

Myslobodsky MS, Coppola R, Jacob B, Weinberger DR: Adequacy of the international 10-20 electrode system for computed neurophysiologic topography. J Clin Neurophysiol *7:*507, 1990.

Niedermeyer E: Dipole theory and electroencephalography. Clin Electroencephalogr *27:*121, 1996.

Niedermeyer E: The normal EEG of the waking adult. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, EL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Pascual-Leon A, Meador KJ: Transcranial magnetic stimulation. J Clin Neurophysiol *15:*22, 1998.

Polich J: P300 in clinical applications: Meaning, method and measurement. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Prichep LS, Lieber AL, John ER, Alper K, Gomez-Mont F, Essig-Peppard T, Flitter M: Quantitative EEG in depressive disorders. In *Brain Electrical Potentials and Psychopathology: Proceedings of the VII International Symposium on Clinical Neurophysiological Aspects of Psychiatric Conditions*, September, 1985, Philadelphia, PA, C Shagass, RC Josiassen, RA Roemer, editors. Elsevier, New York, 1986.

Reilly EL: EEG recording and operation of the apparatus. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Reilly EL: Nasopharyngeal, sphenoidal, and other electrodes. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Reilly EL, Peters JF: Relationship of some varieties of electroencephalographic photosensitivity to clinical convulsive disorders. Neurology *23:*1050, 1973.

Reisberg B, Burns A, Brodaty H, Eastwood R, Rossor M, Sartorius N, Winblad B: Diagnosis of Alzheimer's disease. Int Psychogeriatr *9:*11, 1997.

*Selemon LD, Goldman-Rakic PS. The reduced neuropil hypothesis: A circuit based model of schizophrenia. Biol Psychiatry *45:*17, 1999.

Sharbrough FW: Nonspecific abnormal EEG patterns. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Shahrokhi F, Chiappa KH, Young RR, Pattern shift visual evoked responses: Two hundred patients with optic neuritis and/or multiple sclerosis. Arch Neurol *35:*65, 1978.

*Small JG: Psychiatric disorders and EEG. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Small JG, Milstein V, Kellams J, Miller MJ, Boyko OB, Small IF: EEG topography in psychiatric diagnosis and drug treatment. Ann Clin Psychiatry *1:*7, 1989.

Spitzer RA, Cohen LG, Fabrikant J, Hallett M: A method for determining optimal interelectrode spacing for cerebral topographic mapping. Electroencephalogr Clin Neurophysiol *72:*355, 1989.

Weickert CS, Weinberger DR. A candidate molecule approach to defining developmental pathology in schizophrenia. Schizophr Bull *24:*303, 1998.

Wieser H: Stereoelectroencephalography and foramen ovale electrode recording. In *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, ed 3, E Niedermeyer, FL Da Silva, editors. Williams & Wilkins, Baltimore, 1993.

Winterer G, Dorn H, Herrmann WM, Gallhofer B, Bauer U, Hegerl U, Ihl R, Dierks T, Maurer K: The AMDP modules I-IV: Recommendations for a standardized acquisition of EEG data in psychiatry. Neuropsychobiology *36:*100, 1997.

Wyllie E, Glazer JP, Benbadis S, Kotagal P, Wolgamuth B: Psychiatric features of children and adolescents with pseudoseizures. Arch Pediatr Adolesc Med *153:*244, 1999.

1.15 PRINCIPLES OF NEUROIMAGING: RADIOTRACER TECHNIQUES

ROBERT T. MALISON, M.D., AND ROBERT B. INNIS, M.D., Ph.D.

Neuroimaging methodologies allow measurement of the structure, function, and chemistry of the living human brain. Although some of the methods such as structural magnetic resonance imaging (MRI) have been a standard component of the clinical assessment of patients with neurological disorders for more than a decade, similar applications have not yet been found for patients with psychiatric disorders; any methods are largely restricted to research studies. However, following the path established in neurology, neuroimaging of psychiatric disorders can be expected to have a similarly useful clinical application in psychiatry. In addition to the potential clinical applicability of neuroimaging, the current attraction of these methods in psychiatry is based on very promising results from all the methods described in this chapter to examine the physiology of healthy subjects and the pathophysiology of patients with psychiatric illnesses. These relatively recent studies, the vast majority of which have been performed since the late 1980s, have demonstrated that reliable measurements are feasible and that selection of the appropriate target parameters will provide key information on the pathophysiology of psychiatric illnesses. Such information may be useful for diagnosing and monitoring the illness and for developing new treatments. Neuroimaging is also attractive to the field of psychiatry because it provides direct measurements of the brain and thereby bypasses the limitations of peripheral measures (e.g., concentrations of plasma hormones or urinary catecholamine metabolites), which generate only indirect assessments of central nervous system function.

The two radiotracer methods of neuroimaging, positron emission tomography (PET) and single photon emission computed tomography (SPECT) entail the injection of radioactively labeled drugs. The imaging and measurement over time of the distribution of these radiotracers is used to assess the neurochemistry, blood flow, or metabolism of the brain. The three methods based on nuclear magnetic resonance are MRI, functional MRI, (fMRI), and magnetic resonance spectroscopy (MRS). In general terms, MRI provides exquisitely detailed images of brain structure; fMRI provides images of local neuronal activity with high spatial and temporal resolution; and MRS provides measurements of the concentrations of numerous chemicals in the brain without the radiation exposure of PET and SPECT but with much lower sensitivity.

BASIC PRINCIPLES

In PET and SPECT a biological process of interest is studied by synthetically incorporating a radionuclide into a molecule of known physiological relevance. The so-called radiopharmaceutical is then administered to a patient either by inhalation, ingestion, or most commonly by intravenous injection. As radioactivity distributes within the subject, the radiotracer's uptake into the brain is measured over time and is used to obtain information about the physiological process of interest. Because of the high-energy (g-ray) emissions of the specific isotopes employed and the sensitivity and sophistication of the instruments used to detect them, the two-dimensional distribution of radioactivity within a brain slice may be inferred from information obtained outside the head. For this reason, PET and SPECT are referred to as *emission tomographic* (from the Greek *tomos* for *cut*) *techniques*. In contrast to more conventional radiographic methods, like a chest X-ray, where an external source of radiation merely casts a shadow of the body's organs and cavities onto a planar film, PET and SPECT rely on more sophisticated principles to produce three-dimensional information. In order to understand this process, a basic understanding of the physics of photon emission is required.

Physics of Photon Emission Radioactive decay is a process whereby unstable nuclides, those possessing an excess number of protons, reduce their net nuclear positivity. For the radionuclides used in PET, a proton is converted to a neutron and a particle called a *positron* (denoted e_+ or b_+) is emitted. A positron may be thought of as the antimatter equivalent of an electron, possessing identical mass but opposite charge. When ejected from the nucleus, a positron travels until it collides with an electron. This collision results in the annihilation of both particles and the conversion of mass into energy (Fig. 1.15-1). The energy produced has a characteristic profile consisting of two g-photons (rays) of equivalent energy (511 keV) and opposite trajectory (180° apart). These dual photons distinguish PET from SPECT and have implications for camera design. The most commonly used positron-emitting nuclides in PET include 11-carbon (^{11}C), 13-nitrogen (^{13}N), 15-oxygen (^{15}O), and 18-fluorine (^{18}F).

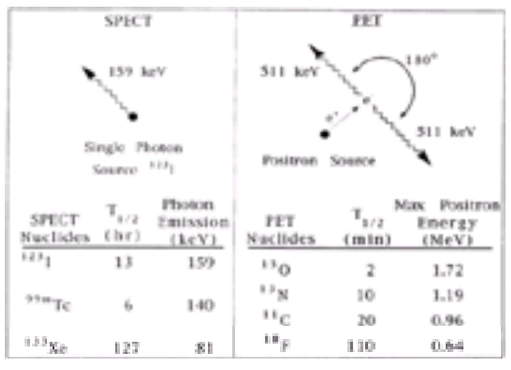

FIGURE 1.15-1 The decay of a SPECT radiopharmaceutical results in the emission of a high-energy photon directly from the radionuclide. In contrast, the decay of a PET radiopharmaceutical results in the emission of a positron (e^+) that travels a variable distance before annihilating with an electron (e^-), which then yields two 511 keV photons at an angle of 180 degrees to each other. The distance traveled by the positron decreases the resolution of PET images when using the typical nuclides listed by 0.2 to 1.3 mm, with resolution measured as FWHM. The longer-lived SPECT radionuclides emit single photons of different energies, whereas the PET radionuclides consistently yield two photons of 511 keV.

PET scanners take advantage of the unique spatial signature of back-to-back photons by using a method known as *coincidence detection* to locate the source of an annihilation event (Fig. 1.15-2). Coincidence detection is an efficient technique and contributes to PET's superior sampling rates and sensitivity. In a typical configuration, a PET scanner consists of a circular array of highly sensitive scintillation detectors that surround the head. These detectors are made of dense crystalline materials (e.g., bismuth germanium oxide, sodium iodide, or cesium fluoride) that trap the invisible, high-energy g-rays and convert them to visible light. This brief flash of light is then converted into an electrical pulse by an immediately adjacent photomultiplier tube, and the electrical pulse is then registered by the scanner's computer. When the scanner detects two electronic signals from two radiation detectors that coincide (to within 3-10 nsec), an annihilation event is presumed to have occurred at some point along an imaginary line connecting the detectors. In contrast, single events are ignored. Although any two crystal detectors may be activated by coincident photons, the most straightforward conceptual configuration for a PET camera is one in which only opposing detectors are electronically connected. Although it is true that two unrelated photons from spatially separate annihilation events can reach opposing detectors concurrently, such accidental coincidences are much less frequent than true ones. Nevertheless, random coincidences constitute one source of the background noise in PET images.

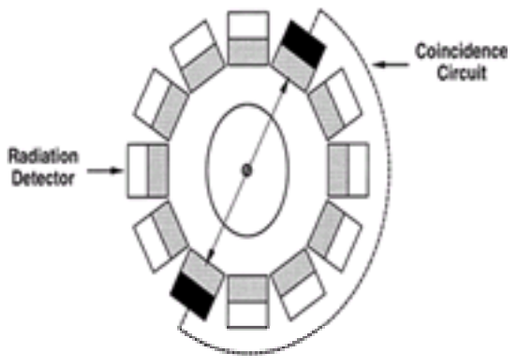

FIGURE 1.15-2 A PET scanner consists of a ring of radiation sensors that are designed to detect the simultaneously emitted, characteristically back-to-back (180 degrees apart) dual photons that are created by the annihilation of a positron and an electron. Opposing detectors are electronically coupled to form a coincidence circuit. Thus, when separate scintillation events in paired detectors coincide, an annihilation event is presumed to have occurred at some point along a line connecting the two. This information is registered by a computer and later used to reconstruct images using the principles of computed tomography. (Reprinted with permission from Malison RT, Laruelle M, Innis RB: Positron and single photon emission tomography: Principles and applications in psychopharmacology. In *Psychopharmacology: The Fourth Generation of Progress*, F Bloom, D Kupfer, editors. Raven, New York, 1995.)

Because PET detects the site at which a positron annihilates and not the site of its emission, there exists an intrinsic theoretical limit on the spatial resolution of PET. Specifically, a positron generally travels a finite distance before coming to rest in a tissue and colliding with an electron. Thus, an annihilation typically occurs some distance away from the site of radioactive decay. This distance is proportional to the positron's average kinetic energy as it is emitted from the nucleus and is characteristic of the specific isotope employed (Fig. 1.15-1). For example, the range for ^{11}C decay is roughly 2 mm. An additional limitation placed on PET is that sometime photons are emitted at an angle slightly more or less than 180 degrees. This typically occurs if the positron is not entirely at rest when it annihilates. Thus, both remote positron annihilation and photon noncollinearity are factors that theoretically limit PET's achievable spatial resolution (approximately 2 to 3 mm); modern PET instruments fall within this realm.

In the case of SPECT, the opposite occurs. Instead of a proton-rich radionuclide ejecting a positron (i.e., e+), it captures an orbiting electron (denoted e–). Once again, the net result is transformation of a proton into a neutron. Most commonly, the radioactive progeny of this process remains in a residually excited, so-called metastable state. With the dissipation of this metastable arrangement, the daughter nucleus achieves a ground state, and a single g-photon is produced. Thus, SPECT employs isotopes that decay by electron capture or g-emission or both, including both 123-iodine (^{123}I) and the long-lived metastable nuclide 99m-technetium (^{99m}Tc). No comparable theoretical limit on spatial resolution exists for SPECT because the site of g-emission and the site of radioactive decay are synonymous.

The emission of only a single photon fundamentally distinguishes SPECT from PET and necessitates an intrinsically different approach to ascertaining the origin of a decay event and therefore of camera design. Specifically, SPECT utilizes a method know as *collimation* (Fig. 1.15-3). In a manner analogous to the effects of a polarizing filter for visible light, a collimator is a physical filter that permits only g-rays of a specific spatial trajectory to reach the SPECT scanner's detector. Most commonly, a collimator is a lead structure that is interposed between the subject and the radiation detector. The collimator contains many holes of sufficiently long and narrow dimension so that only photons of a parallel trajectory are allowed through. In contrast to parallel photons, g-rays that deviate slightly are absorbed by the lead and go undetected (Fig. 1.15-3). Different collimators (e.g., parallel, fan-beam, and cone-beam) have holes of differing orientations (e.g., perpendicular to the detector, focused in two dimensions, and focused in three dimensions, respectively). Given a known geometric configuration for the specific collimator's holes, the original path of a detected photon is linearly extrapolated. Collimation is less efficient than coincidence detection because many potentially informative photons are lost. However, the sensitivity of SPECT has been largely enhanced by advances in collimator design and an increase in the number of detectors surrounding the body; SPECT is now sufficiently sensitive for routine use in nearly all the same applications as PET.

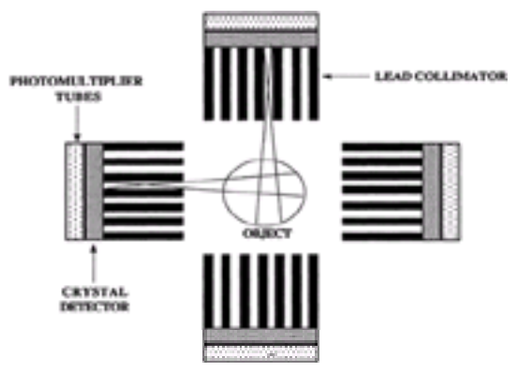

FIGURE 1.15-3 The method of image reconstruction from back projection in SPECT uses a collimator placed between the object and the crystal detector. The area of the object that is viewed by the underlying detector is decreased by having longer and narrower holes in the collimator. By moving the detector-collimator complex around the object, multiple views are obtained and provide the primary data for image reconstruction.

Image Reconstruction Although the nature of photon emission and detection are different in PET and SPECT, both techniques rely on the same principles of computed tomography when translating information about photon paths into brain images. Briefly, computed tomography is based on the premise that an appreciation of an object's two- or three-dimensional distribution in space may only be inferred by viewing it from multiple vantage points. Since information about a photon's direction, not depth, is known, views of photon trajectories from multiple angles around the entire head are required. In PET and SPECT, such a set of measurements from a given angle or viewpoint is referred to as a *projection*. A ring of essentially contiguous radiation detectors in a PET scanner provides multiple projections in this modality. In contrast, SPECT cameras usually rely on several (typically 2 to 4) detector “heads” that rotate around the subject in synchrony, collecting data over 360 degrees. A picture of the distribution of radioactivity within a given brain slice is then inferred by retracing or backprojecting the trajectories (typically thousands) of g-rays across the field of view for every imaging angle. Conceptually analogous to the simple childhood puzzle in which numbers in a square grid (e.g., 3 × 3) are inferred from their sums along each row, PET and SPECT images require fast computer coprocessors and efficient mathematical algorithms (fast Fourier transformations) to handle the considerably larger matrices (e.g., 128 × 128 or 256 × 256 elements) of radiation density values and the correspondingly more intensive calculations. In this manner, individual radiation values (i.e., counts of detected events) are determined for each cell of the matrix (also known as a *picture element* or *pixel*), corresponding shades of color assigned, and an image of the distribution of radioactivity within the brain produced (Fig. 1.15-4).

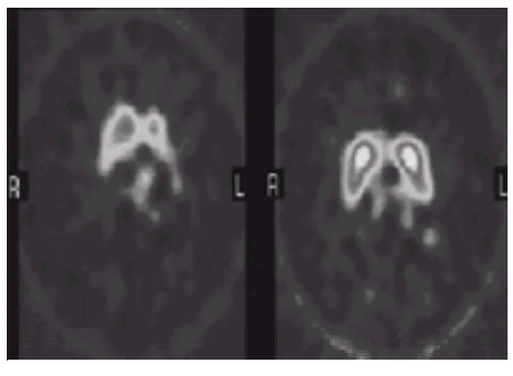

FIGURE 1.15-4 SPECT images of the distribution of [^{123}I]b-CIT (cocaine-iodo-tropane) in a healthy subject and a patient with Parkinson's disease. [^{123}I]b-CIT is a

radiolabeled cocaine analogue and is a probe of dopamine transporters in the striatum. These transporters are located presynaptically on terminals of dopamine neurons projecting from the substantia nigra to the striatum. These transverse images show a high density of dopamine transporters in striatum and a marked reduction of these sites in an age- and sex-matched patient with idiopathic Parkinson's disease. The transporters are lost because the entire dopaminergic neuron, including its terminal projections in the striatum, degenerate in this disorder. (Courtesy of John Seibyl, Yale University.) (See Color Plate 1.)

Despite its complexity and computational intensity, backprojection is an imperfect process and introduces known artifacts into the images themselves. As the backprojection algorithm retraces a photon's path, it cannot be sure of the actual point of decay. The algorithm is thus forced to assume an equal probability of radioactive decay and hence of radiation value for every point along the line of trajectory. Areas of the brain in which radioactivity is highly concentrated will stand out as many trajectories from multiple projections are superimposed and their probability values are summed. In the process, however, those areas containing no radioactivity now bear the statistical imprint of the algorithm's guess. Thus, small but finite values are ascribed to areas where none should exist. By increasing the density of spatial sampling through greater numbers of projections, the impact of these spurious values on image quantitation can be minimized but not eliminated. Therefore, a filter is still required to restore quantitative accuracy to images by “erasing” counts in those areas that should have none. Several filters have been developed (e.g., Ramp, Butterworth, and Hanning) in an effort to overcome these limitations, and these techniques remain the mainstay of the field of image reconstruction. Trade-offs exist with respect to the relative impact of filters on spatial resolution and noise amplification, and filter selection depends on the imaging context. Alternative reconstruction methods (e.g., restorative and iterative techniques) are the current focus of much research, and simple filtered backprojection is likely to be superseded by more quantitatively accurate methods in the near future.

Factors Affecting Image Quantitation Several physical factors affect the quantitative accuracy of PET and SPECT images. Among these are the statistics of radioactive decay, photon scatter, limited spatial resolution, and partial volume effects.

Statistics of Radioactive Decay Mathematically, radioactive decay is described by an exponential curve. The rate at which a specific radionuclide decays is expressed in terms of its radioactive half-life, or $T_{1/2}$ value, a parameter defined as the time required for half of the radioactive atoms to decay. Values of $T_{1/2}$ vary between species and are characteristic of a given nuclide. The characteristic $T_{1/2}$ values of several commonly used PET and SPECT isotopes are listed in Figure 1.15-1. Although a given isotope's half-life is constant, the nature of radioactive decay is intrinsically statistical. This phenomenon is most readily conceptualized by imagining an isotope of infinite half-life (i.e., unchanging levels of radioactivity). In struggling to measure the precise amount of radioactivity during a fixed period of time, variations in the individually recorded values invariably result. Only by taking the statistical average of multiple measurements is the true amount of radioactivity (and $T_{1/2}$ value) inferred. The variation in sampling derives from the intrinsic probabilistic nature (mathematically described by a Poisson distribution) of radioactive decay and the random fluctuation in individual decay events from moment to moment, and it occurs irrespective of detection method.

The manifestations of this effect are most readily appreciated by imaging an object which contains a uniform concentration of radioactivity. The swiss cheese appearance of the resulting images is the spatial equivalent of this temporal variation in PET and SPECT images. The probabilistic inaccuracies (or statistical noise) introduced is by virtue of its random nature easily surmountable through the collection counts. Longer sampling times and greater instrument sensitivity are the principle ways in which counting statistics are improved. Longer acquisition times improve statistical noise at the expense of temporal resolution. Conversely, increased sensitivity (e.g., larger collimator holes in SPECT) are traded for poorer spatial resolution (because slightly less than parallel photons are detected).

Photon Attenuation Although the high energies of photons emitted by PET and SPECT nuclides enable their penetration of brain structures, a significant number of g-rays escape detection by both types of scanners based on their interactions with surrounding tissues. These interactions fall into two general categories—Compton scattering and photoelectric absorption. In *Compton scattering*, a collision occurs between the photon and an atomic electron. The photon is deflected from its original trajectory and, in the process, loses a fraction of its original energy. Alternatively, in *photoelectron absorption*, the photon's energy is completely absorbed by the atom, and an electron may be ejected from its orbit. For this reason, g-radiation is said to be ionizing.

Because the chances of scatter or absorption decrease with increasing photon energy and increase with distance, photon attenuation is both energy- and depth-dependent. On both counts, PET has distinct advantages. Since photons in PET have higher energies (i.e., 511 keV) than those in SPECT (typically 80 to 160 keV), they are less prone to attenuation (Fig. 1.15-1). Moreover, linear attenuation is largely depth independent in PET because of coincidence detection. Nevertheless, activity at the brain's center is disproportionately underestimated (roughly fourfold to fivefold) in comparison to its surface for both PET and SPECT. Compensating for undetected photons is therefore crucial for comparing radioactive densities in different brain regions. The most commonly employed method with SPECT is uniform attenuation correction. In *uniform attenuation correction*, an ellipse is fitted to the brain's contour and the same attenuation value (typically equal to that of water) is assigned to all points within the ellipse. A commonly used method for attenuation correction in PET (and with some recent SPECT devices) is *nonuniform attenuation correction*, which relies on a preceding transmission study similar to a computed tomography (CT) scan. An external source of radiation is transmitted through the subject's head, creating a precise attenuation map for that individual. Because the sizes and shapes of patients' heads vary and because the attenuation properties of bone, tissue, fluid, and air differ, such an approach has clear theoretical advantages. Although commercially available PET and SPECT scanners now support this technology, the conditions under which nonuniform methods offer clear advantages over uniform approaches remain to be established for practical brain-imaging applications.

Photon Scatter In both PET and SPECT, instrumentation and image reconstruction are based on the underlying assumptions that detected photons retain their linear paths. However, Compton effects cause photons to deviate from their original trajectories. Although these photons lose energy to atoms in the tissue, many scattered photons retain sufficient energy to enable their escape from the brain. The detection of scattered events therefore leads to errors in image reconstruction as a result of false assumptions about the photon's original path. Much like accidental coincidences, scattered photons increase the background noise and compromise image contrast.

Because radionuclides emit photons of a known energy, scattered photons may be distinguished from true ones by the loss of energy they sustained from collisions with electrons. In an attempt to exploit this principle, PET and SPECT cameras measure the energy spectrum of their detected photons. In practice, however, accurately discriminating between true and scattered photons is often difficult because the energy resolution of current PET and SPECT scanners is limited. Also, the photopeak energies of true photons are not identical, but rather normally distributed about a mean value. Thus, scattered and photopeak photons inevitably overlap in their energy distributions. Current algorithms that subtract a *scatter fraction* from the photopeak counts are an attempt to compensate for this problem; however, these methods have obvious limitations. As for attenuation correction, advances in scatter correction offer the promise of incorporating *a priori* information about the head's structure and density in achieving more faithful image reconstruction.

Spatial Resolution In contrast to the fine visual detail seen in magnetic resonance images, pictures created using SPECT and PET appear blurred. The visual sense of imprecision is the qualitative consequence of limited spatial resolution. Equally important, however, is the quantitative impact of finite resolution on the measured radioactivity in individual brain regions. The latter partial volume effects have important consequences for image quantitation and require a clearer understanding of spatial resolution and its definition.

In PET and SPECT, spatial definition is defined in practical terms, the distance by which two objects must be separated to perceive them as discrete (Fig. 1.15-5). In a SPECT or PET camera with perfect resolution, a point source of radioactivity would be depicted as a vertical line of infinitely narrow width. In such an ideal device, two point sources could be distinguished from each other as long as they were not superimposed. In the real world, however, PET and SPECT scanners perceive the radioactivity from such a point source as a Gaussian curve and the radioactivity from the point is spread out. This *point spread function* characterizes a camera's resolving capacity. The spatial diffusion of imaged radioactivity is expressed in terms of the *full-width-at-half-maximum* (FWHM), or the width of the Gaussian curve at half of the curve's peak activity. The FWHM is the parameter most commonly used to define resolution in emission tomography because this is the distance at which the peaks of both sources become distinguishable from one another (Fig. 1.15-5).

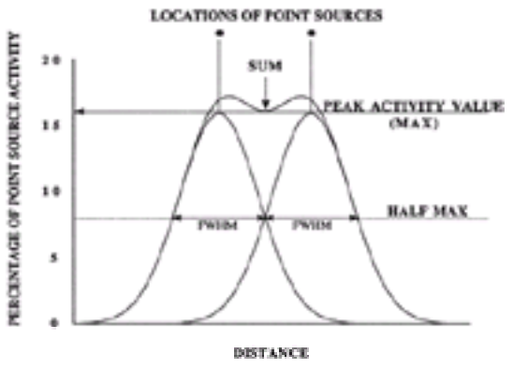

FIGURE 1.15-5 The limited resolution of PET and SPECT cameras blurs the activity of single point sources into adjacent regions with no activity. Viewed in just one dimension, a point source is visualized by the camera as a gaussian curve. Resolution is defined as the width of the curve at half maximum measured peak levels (FWHM). For two point sources of equal intensity separated by a distance equal to the FWHM of the camera, the sum of the activities begins to show a modest decrease at the midpoint. Thus, two point sources separated by a minimum distance equal to the FWHM will begin to appear as two separate points rather than just one.

Excluding issues of positron range, image resolution is primarily influenced by issues of instrumentation. For example, the precision of collimation, the number and size of detectors, and the accuracy with which scintillation events are localized within the crystalline elements all contribute to limited spatial resolution. In the case of PET, state-of-the-art devices have yielded resolutions close to the theoretical limits of accuracy (3 mm). However, the average PET and SPECT camera is currently capable of 5 to 6 mm and 7 to 9 mm FWHM, respectively.

Partial Volume Effects In the simplest terms, partial volume effects create one of two problems for image quantitation: the appearance of radioactivity where there was none, and the impression of less radioactivity than truly exists. For example, just as the brightness in a part of a room depends on the intensity and distance separating two lamps, so too will the measured radioactivity in a given brain region reflect the relative activity and proximity of nearby structures. Thus, brain regions with relatively lower concentrations of radioactivity will appear “hotter” in PET and SPECT images as imaged activity “spills over” from adjacent (more active) areas. Conversely, as the size of a radioactive region becomes smaller than two to three times the FWHM resolution, true activity is effectively diluted by nonradioactive areas within the field of resolution. In the latter case, regions containing equal concentrations of radioactivity will appear to have declining levels with decreasing size. Metaphorically speaking, together these two effects result in sharp peaks and steep canyons of brain activity being rendered as short hills and shallow valleys in PET and SPECT images.

Several approaches are currently taken to compensate for errors resulting from limited spatial resolution and partial volume effects. One method attempts to simulate errors created by partial voluming by creating a plastic model or phantom. Models can be designed to approximate the structures or activity distributions of interest in the brain. For example, finely machined, polycarbonate brain phantoms are commercially available and can recreate the geometry of gray and white matter. Once imaged, regionally specific correction factors, or *recovery coefficients*, are derived that relate units of measured activity to known activity. Such methods, however, are unable to account for inter-subject variations in brain anatomy, whether pathological or nonpathological. For this reason, researchers are now beginning to use structural information (e.g., CT or MRI scans) to quantify functional information. More specifically, partial volume errors may be mathematically compensated for by registering subjects' MRI scans with their own PET or SPECT scans and incorporating a priori functional (e.g., relative blood flow ratios in gray and white matter) and physical information (e.g., a PET or SPECT camera's three-dimensional point spread function). The latter approach is more complicated than the former but has obvious advantages for conditions such as Alzheimer's disease, in which cortical atrophy is present.

Radiopharmaceuticals The versatility and sensitivity of PET and SPECT arise largely from the ability of talented radiochemists to synthesize a radiopharmaceutical of high chemical purity, high radioactive yield, and small mass dose. Expressed differently, in order to ensure that a specific biological system of interest is adequately measured yet unperturbed by the tracer, high purity and high specific activity (expressed in units of radioactivity per chemical quantity: Ci/mmol) are paramount. However, the physical nature of radioactive decay and the short half-lives of most suitable radionuclidic species (Fig. 1.15-1) constantly challenge the radiochemist's efforts. Chemical yield generally improves with increasing reaction times; however, radioactivity and specific activity diminishes with increasing decay times. Thus, an optimal synthetic scheme is a balanced one in which chemical yield is maximized, radioactive byproducts are minimized, and the final product is capable of prompt purification. Given the high affinity of many radiopharmaceuticals (e.g., neurotransmitter receptor ligands) for their physiological targets, specific activities of greater than 2000 Ci/mmol are generally required. Most radiopharmaceuticals are still manually prepared by radiochemists racing against the clock of a nuclide's decay; however, a limited number of radiochemical syntheses are now automated and performed in robotically controlled hot cells (e.g., [^{18}F]-2-fluoro-2-deoxyglucose; [^{18}F]FDG).

In the case of positron-emitting radionuclides (e.g., ^{15}O, ^{13}N, ^{11}C, and ^{18}F), the particularly short half-lives (2, 10, 20, and 109 minutes, respectively) have special implications for the design of PET imaging facilities. Most PET centers have an on-site cyclotron that generates radionuclides for “real-time” utilization. An exception to this is ^{18}F, whose nearly 2-hour half-life permits a “local” regional facility to produce quantities for large metropolitan centers. The significant expense of a cyclotron (typically $1 to $2.5 million) and its highly skilled support staff are relative disadvantages for PET. In contrast, SPECT isotopes like ^{99m}Tc ($T_{1/2}$ = 6 hours) may be obtained from inexpensive molybdenum generators located in many hospital radiopharmacies. Alternatively, ^{123}I has a sufficiently long half-life (13 hours) to permit centralized production at distant (>3000 miles) commercial reactors. The radionuclide may then be delivered via overnight express mail and still meet the radiochemical needs of high specific activity.

The choice of a candidate molecule for radiopharmaceutical development depends primarily upon the physiological process under investigation. In the case of regional cerebral blood flow, relatively nonspecific and often nonorganic, diffusible tracers may be employed (e.g., the gaseous tracer ^{133}Xe). In contrast, the measurement of aspects of brain neurochemistry require much greater biochemical selectivity. Thus, PET and SPECT radiopharmaceuticals are most often naturally occurring substances, structural analogues, or ligands that selectively label a particular brain target. In this regard, PET has significant advantages over SPECT because ^{11}C can be directly substituted for ^{12}C in existing organic molecules without altering their intrinsic biochemical properties. Alternatively, fluorine is frequently substituted for native hydrogen atoms without significant isotopic effects (e.g. [^{18}F]FDG). In contrast, SPECT nuclides (i.e., ^{123}I and ^{99m}Tc) are uncommon elements of organic substrates. The metallic nature and multiple valence states of ^{99m}Tc necessitate bulky complexing groups for its molecular stabilization. These barriers have largely limited the initial uses of ^{99m}Tc to nonselective processes (e.g., the blood flow agent [^{99m}Tc]-hexamethyl propyleneamine oxime; [^{99m}Tc]HMPAO). However, in 1997, ^{99m}Tc-labeled probes of the dopamine transporter were developed, and SPECT imaging has demonstrated appropriate labeling in human and monkey brain. Extension of these efforts is certain to result in the development of many other ^{99m}Tc-labeled probes in the future.

Many ^{123}I-containing radiopharmaceutical agents have been developed as a result of rapid advances in iodode metallation procedures and increasing knowledge of the structure-activity relationships of pharmacologically active compounds. The lipophilic nature of ^{123}I may actually facilitate transfer across the blood-brain barrier and, in some instances, improve affinity of the parent compound at its site of action. In particular, SPECT imaging of brain receptors and uptake sites with ^{123}I-containing radiopharmaceuticals is routine at several university medical centers.

Successful in vivo radiopharmaceuticals must fulfill several stringent pharmacokinetic criteria. Because a radiopharmaceutical must easily enter the brain, tracer binding to plasma proteins must be readily reversible, and its transport across the blood-brain barrier must be favorable. Although some tracers (e.g., [^{18}F]-FDG) may have facilitated carriers, most ligands must be sufficiently lipid-soluble to permit passive diffusion across the blood-brain barrier. However, as the tracer's lipophilicity increases, its “signal-to-noise” properties may be degraded because nonspecific binding increases. Thus, lipophilicity (nonspecific binding) and affinity (specific binding) are important factors influencing an imaging agent's signal-to-noise ratio. Lastly, tracer metabolism may also limit a ligand's in vivo utility. For example, rapid degradation, lipophilic radioactive metabolites, and pharmacologically active metabolites may all confound central measurements.

Safety With regard to studies in humans, PET and SPECT methods give rise to similar safety concerns for radiation exposure and pharmacological toxicity of the injected radiopharmaceutical. The radiation exposures from typical PET and SPECT scans are thought to be reasonably safe within the context of the present knowledge of radiation biology. The Food and Drug Administration (FDA) has established limits of radiation exposure to various organs of the body and to the body as a whole, these limits are applied to research studies and are often lower than exposures in routine clinical nuclear medicine procedures. Although the FDA limits are

presently thought to provide adequate safety, the long-term biological effects of ionizing radiation are an area of active investigation and even controversy. The estimation of the dose received by the body depends on multiple factors (including the amount of activity, the type of emission, and the residence time in the body), but the shorter half-lives of PET radionuclides and the higher sensitivity of the method generally yield lower radiation burdens than did a comparable SPECT study. A useful guideline is to employ doses of radiotracer that are "as low as reasonably allowable" to provide useful results.

Fortunately, the pharmacological toxicity of radiopharmaceuticals is usually not a significant issue. The sensitivity of functional imaging is so high that minuscule mass doses of compound may be injected, although that small mass is associated with significant levels of radioactivity. For example, some radiopharmaceuticals are injected at mg/kg doses that are a millionfold lower than the minimal dose required to have any pharmacological effect. In such situations no pharmacological toxicity would be expected and only an unusual immunological adverse effect could be anticipated. Nevertheless, the potential pharmacological effects and toxicity of radiopharmaceuticals needs to be evaluated relative to previously established criteria for nonradioactive pharmaceuticals. The final formulation of any radiotracer must meet established guidelines for purity, sterility, and lack of pyrogenicity.

RESEARCH APPLICATIONS

The uses of PET and SPECT brain imaging can be roughly divided into measurements of local neuronal activity, neurochemistry, and in vivo pharmacology.

Local Neuronal Activity Local neuronal activity is associated with energy consumption and can be directly measured with glucose metabolism or indirectly with cerebral blood flow. Local cerebral blood flow is coupled with glucose metabolism and neuronal activity via mechanisms that are not completely elucidated. PET tracers for measurement of local neuronal activity include [^{18}F]FDG (fluoro-deoxyglucose for glucose metabolism) and [^{15}O]H_2O (blood flow). SPECT does not have a comparable tracer for glucose metabolism but does have ^{99m}Tc- and ^{123}I-labeled agents, as well as ^{133}Xe to provide measures of blood flow.

Neuronal metabolic demands are believed to primarily reflect terminal rather than cell body activity. In any given volume of brain, the majority of [^{18}F]FDG uptake is believed to be in terminals rather than in cell bodies, a conclusion that is based upon a limited number of studies in which the cell bodies are anatomically distant from their terminals. Metabolic rate will not distinguish activity of excitatory and inhibitory neurons. Although increased [^{18}F]FDG uptake is usually interpreted as increased functional activity of a region, it may reflect an overall decreased activity based upon increased firing of inhibitory interneurons.

The clinical uses of PET and SPECT imaging to measure local neuronal activity are largely restricted to neurological disorders and include localization of cerebral ischemia, localization of epileptic foci, and distinguishing radiation necrosis from tumor growth. These imaging results can directly impact clinical care. For example, the neurosurgical treatment of patients with medically refractory epilepsy critically depends upon accurate localization of the seizure focus, which is often distant from the surface of the brain and poorly localized by scalp electrode EEG. The seizure focus is hypometabolic with decreased blood flow during the interictal period, but is hypermetabolic with increased blood flow during the ictal period. PET and SPECT imaging has been used either as a primary means of localization or as a confirmation of other diagnostic tests to select the portion of the brain that is subsequently resected.

PET imaging using [^{15}O]H_2O has been elegantly combined with neuropsychological activation studies to localize cognitive and sensory functions, including reading, speaking, word associations, visual identification, and spatial localization. The short half-life of ^{15}O ($T_{1/2}$ of 2 minutes) allows multiple (often 8 to 10) bolus injections of the tracer in one experimental session. Thus, both baseline scans and those following neuropsychological tasks can be repeated and averaged.

Recently developed techniques in functional fMRI offer great promise to provide measures of local neuronal activity similar to that from PET and SPECT imaging. The primary signal from functional MRI is believed to derive from the concentration of deoxyhemoglobin. Functional MRI is superior to PET and SPECT in that it involves no radiation exposure and has greater temporal (<1 second) and spatial (<1 mm) resolutions. If these methods are fully developed with adequate quantitation, they may supplant PET and SPECT for measures of local neuronal activity.

Neurochemistry Two major attributes of both PET and SPECT are high sensitivity and chemical selectivity, both of which are critically important for in vivo neurochemical measurements. The sensitivity of PET and SPECT to detect radiotracers is less than 10^{-12} M, which is several orders of magnitude greater than that of nuclear magnetic resonance (NMR) methods. "Sensitivity" refers to the minimal concentration of the target compound that can be reliably measured. For example, the minimal concentration of [^{11}C]chlorpromazine that can be measured with PET in the human brain within a reasonably acceptable imaging time (e.g., 15 to 30 minutes) is less than 10^{-12} mol. In contrast, the minimal concentration of g-aminobutyric acid (GABA) that can be measured with MRS is about 10^{-4} mol.

In a manner exactly analogous to nonradioactive drugs, radiotracers that label specific target sites in the brain can be developed. These specific tracers can thereby provide measures of multiple neurochemical pathways in the brain, including synthesis and release of transmitters, receptors, reuptake sites, metabolic enzymes, and possibly even second messenger systems. Of these multiple neurochemical systems in the brain, the greatest effort has been devoted to imaging of dopaminergic transmission; it is an example of the types of measurements provided by these methods.

Synthesis 6-[^{18}F]Fluoro-L-3,4-dihydroxyphenylalanine ([^{18}F]-FDOPA) has been successfully used in animal and human studies to provide a measure of dopamine terminal innervation of the striatum. These studies have demonstrated decreased striatal uptake in parkinsonian patients as compared to healthy subjects. Furthermore, these studies have questioned the widely held notion that symptoms develop only after 85 to 90 percent depletion of endogenous dopamine levels. Imaging studies of patients with early signs of the disorder suggest that symptoms may begin with only a 50 to 60 percent decrease in striatal dopamine terminal innervation.

Release A potential method for the measurement of transmitter release involves the displacement of receptor radiotracers by the endogenous transmitter. On first consideration, this displacement may seem impossible because the endogenous transmitter tends to have a much lower affinity than the tracer for the receptor. For example, [^{11}C]raclopride has a IC_{50} value (concentration of drug required to occupy 50 percent of sites) for the dopamine type 2 (D_2) receptor of approximately 1 nM. In contrast, the IC_{50} value for dopamine itself may be as high as 1 μm. How then could dopamine effectively compete with [^{11}C]raclopride for binding to the D_2 receptor? Different affinities, which are inversely related to IC_{50}, will influence the speed at which equilibrium is achieved. However, after equilibrium is achieved, dopamine demonstrates effective competition with high-affinity tracers. In an equilibrium state, if the displacer dopamine is present at a concentration equal to its IC_{50} value, then 50 percent of receptors will be occupied by the drug and 50 percent of the radioligand (which is associated with only tracer receptor occupancy) will be displaced. Thus, the real questions of feasibility will be determined by physiological concentrations of dopamine in the synapse, the in vivo inhibition constant (K_i) of dopamine for the receptor, and whether adequate time has elapsed to establish equilibrium binding conditions. Several investigators have provided evidence from in vivo labeling studies in rodents that both the resting levels of synaptic dopamine and stimulant-induced dopamine release are associated with significant D_2 receptor occupancy, which is mirrored by comparable displacement of radiotracer from the receptor. Recent PET and SPECT D_2 receptor imaging studies in humans and monkeys have been combined with a pharmacological challenge of dextroamphetamine (Dexedrine), which causes a massive release of dopamine into the synapse. The release of dopamine can be monitored by displacement of tracer binding to the D_2 receptor. These studies have shown that the amount of displacement in healthy subjects is correlated with psychological reports of activation and feeling high. Similar studies in patients with schizophrenia have shown that the amount of dopamine released is two and a half times higher and that the amount of the release is correlated with a transient increase in positive symptoms. These dopamine receptor imaging studies combined with amphetamine challenge have provided further evidence of dopamine playing a role in psychotic symptoms.

Receptors Receptor studies have probably received the greatest effort among the various targets of neurochemical imaging. If a receptor is selectively altered in a specific disease, then imaging of this site may provide diagnostic information about the disorder. For the dopamine receptor system, the Johns Hopkins PET group has reported that drug-naive schizophrenia patients have an elevation of D_2 receptor density in the striatum of two and a half times as measured by the virtually irreversible tracer [^{11}C]N-methylspiperone and kinetic modeling. In contrast, the PET group at the Karolinska Institute, Sweden has reported normal levels of D_2 receptor densities in drug-naive schizophrenia patients using the reversible radiotracer [^{11}C]raclopride and an equilibrium approach. Reasons for these disparate results have been investigated but remain disappointingly elusive. Elevated D_2 receptors are certainly not a sine qua non of schizophrenia but may be associated in an as yet unknown way with aspects of psychosis in subsets of patients. Research imaging of the D_2 receptor has more recently focused on the interaction with synaptic dopamine, both basal levels and stimulant-induced release.

Transporters The transporter is located presynaptically on terminals of dopamine projections from substantia nigra to striatum. Thus, the transporter is a marker for dopamine terminal innervation, which is decreased in patients with idiopathic Parkinson's disease. Several radiotracers for the dopamine transporter have been developed: [^{11}C]cocaine, [^{11}C]methylphenidate, [^{11}C]CFT (also designated WIN 35,428), and [^{123}I]b-CIT (also designated RTI-55). The striatal uptake of both [^{11}C]CFT

and [^{123}I]b-CIT have recently been shown to be markedly decreased in patients with Parkinson's disease in comparison to healthy subjects of similar mean age (Fig. 1.15-4). Imaging with these tracers may be useful research tool for early diagnosis and for monitoring the progression of the disorder.

Metabolism The fate of a neurotransmitter can be studied by injection of selective inhibitors of the catabolic enzymes. For example, selegiline (Edepryl) is an irreversible inhibitor of monoamine oxidase (MAO) type B, and imaging with ^{11}C-labeled selegiline has been reported to provide a measure of regional enzyme activity in the brain. PET scanning with [^{11}C]selegiline may provide useful dosage-response measurements in patients treated with MAO inhibitors (MAOIs). Furthermore, reversible MAOIs such as [^{11}C]Ro 19-6327 may have advantages relative to the irreversible agents in terms of data analysis and ease of performing in vivo occupancy studies.

In Vivo Pharmacology Since receptors are frequently the targets of therapeutic medications, several investigators have argued that receptor imaging may be used to monitor drug treatment more accurately than is possible with measurement of plasma levels of the medications. However, the rationale for this argument is flawed from a theoretical perspective. Under steady-state conditions achieved with long-term treatment, the level of free (i.e., not protein bound) drug in plasma should be in equilibrium with the free level of drug in the extracellular space of the brain. Thus, under steady-state conditions there is little apparent value in performing expensive neuroreceptor imaging studies instead of simple measurements of the free level of drug in plasma. However, for non–steady-state conditions (e.g., beginning or discontinuing treatment), receptor imaging can provide valuable kinetic information. The brain uptake and washout of many psychoactive agents can be markedly delayed compared to a rapid peak and fast clearance of the drug from plasma. For example, the maximal brain uptake of the potent cocaine analogue cocaine-iodo-tropane occurs about 12 hours after intravenous administration as compared to plasma levels, which peak at 2 minutes. In addition, significant D_2 receptor occupancy has been reported to last for several weeks following discontinuation of antipsychotic agents, even when plasma levels are almost undetectable.

Several pharmaceutical companies and academic researchers have begun to explore the role of receptor imaging in new drug development. The two basic methods are the radiolabeling of the target compound (e.g., with ^{11}C) or the in vivo screening of the effects of the intravenously administered nonradioactive compound with previously developed radiotracers. An example of the first method would be the use of ^{11}C-labeled fluoxetine (Prozac); an example of the second method would be the use of nonradioactive fluoxetine to interact with a different radiolabeled probe (e.g., [^{11}C]citalopram) for the 5-HT transporter. The first method is probably better suited to PET radiochemistry, which can more easily provide a pharmacologically identical radiolabeled form of the target compound than SPECT, which would probably use an iodinated analogue. However, the second method may be equally well performed with PET or SPECT provided that an appropriate radiotracer has been developed for each method.

Brain imaging studies of antipsychotic medications provide an example of the advantages and limitations of the PET and SPECT methods. Several pharmaceutical companies are trying to develop so-called atypical medications like clozapine (Clozaril) and risperdol (Risperidone) that would have superior efficacy and fewer adverse effects than older antipsychotic medications. Studies with [^{11}C]clozapine have been relatively disappointing because of the high nonspecific uptake of the radioactivity. These results may prove to be typical, since only a small percentage of potential compounds prove to be useful in vivo radiotracers with low nonspecific binding. Studies of nonradioactive antipsychotic medications with established and selective receptor tracers have provided valuable information on the receptor occupancy profiles and the pharmacokinetics of brain uptake. For example, researchers have shown that, in comparison to several typical antipsychotic medications, clozapine is associated with a disproportionately high occupancy of D_1 relative to D_2 receptors. Novel therapeutic compounds could be examined for both the pharmacokinetics of entry into the brain and their receptor occupancy profiles, which will provide the combined effect of the parent compound and any active metabolites; this potential application is receiving growing attention.

SUGGESTED CROSS-REFERENCES

Brain-imaging techniques, including electroencephalography and magnetoencephalography, are discussed in Section 1.16. Neuroimaging in clinical practice is discussed in Section 2.13, and neuroimaging in geriatric assessment is discussed in Section 51.2e and Section 51.2f. The other sections of Chapter 1 discuss related neural sciences, particularly Section 1.2 on functional neuroanatomy and Section 1.14 on applied electrophysiology.

SECTION REFERENCES

Calne DB, Langston JW, Martin WRW, Stoessl AJ, Ruth TJ, Adam MJ, Pate BD, Schulzer M: PET after MPTP: Observations relating to the cause of Parkinson's disease. Nature *317:*246, 1985.

Carson RE, Channing MA, Blasberg RG, Dunn BB, Cohen RM, Rice KC, Herscovitch P: Comparison of bolus and infusion methods for receptor quantification: Application to [^{18}F]-cyclofoxy positron emission tomography. J Cereb Blood Flow Metab *13:*24, 1992.

*Childress AR, Mozley PD, McElgin W, Fitzgerald J, Reivich M, O'Brien CP: Limbic activation during cue-induced cocaine craving. Am J Psychiatry *156:*11, 1999.

*Ernst M, Zametkin AJ, Jons PH, Matochik JA, Pascualvaca D, Cohen RM: High presynaptic dopaminergic activity in children with Tourette's disorder. J Amer Acad Child Adolesc Psychiatry *38:*86, 1999.

Farde L: Quantitative analysis of D_2 dopamine receptor binding in the living human brain with PET. Science *231:*258, 1986.

*Farde L, Nordstrom A-L, Wiesel F-A, Pauli S, Halldin C, Sedvall G: PET analysis of central D_1 and D_2 dopamine receptor occupancy in patients treated with classic neuroleptics and clozapine—relationship to extrapyramidal side effects. Arch Gen Psychiatiatry *49:*538, 1992.

Farde L, Wiesel F-A, Stone-Elander S, Halldin C, Nordstrom A-L, Hall H, Sedvall G: D_2 dopamine receptors in neuroleptic-naive schizophrenic patients. Arch Gen Psychiatry *47:*213, 1990.

Fowler JS, Volkow ND, Wolf AP, Dewey SL, Schlyer DJ, MacGregor RR, Hitzemann R, Logan J, Bendriem B, Gatley SJ, Christman D: Mapping cocaine binding sites in human and baboon in vivo. Synapse *4:*371, 1989.

Frost J, Wagner H: *Quantitative Imaging: Neuroreceptors, Neurotransmitters, and Enzymes*. Raven Press, New York, 1990.

Frost JJ, Douglass KH, Mayber HS, Dannals RF, Links JM, Wilson AA, Ravert HT, Crozier WC, Wagner HN Jr: Multicompartmental analysis of [^{11}C]carfentanil binding to opiate receptors in humans measured by positron emission tomography. J Cereb Blood Flow Metab *9:*398, 1989.

Innis RB. Single photon emission computed tomography imaging of dopaminergic function: presynaptic transporter, postsynaptic receptor, and “intrasynaptic” transmitter. Adv Pharmacol *42:*215, 1998.

Innis RB, Malison RT, Al-Tikriti M, Hoffer PB, Sybirska EH, Seibyl JP, Zoghbi SS, Baldwin RM, Laruelle MA, Smith E, Charney DS, Heninger G, Elsworth JD, Roth RH: Amphetamine-stimulated dopamine release competes in vivo for [^{123}I]IBZM binding to the D_2 receptor in nonhuman primates. Synapse *10:*177, 1992.

*Ito H, Halldin C, Farde L: Localization of 5-HT_{1A} receptors in the living human brain using [carbonyl-^{11}C]WAY-100635: PET with anatomic standardization technique. J Nucl Med *40:*102, 1999.

Kung H, Kim H, Kung M, Meegalia S, Plossl K, Lee H: Imaging of dopamine transporters in humans with technetium-99m TRODAT-1. Eur J Nucl Med *23:*1527, 1996.

Laruelle M, Abi-Dargham A, van Dyck C, Gil R, D'Souza C, Erdos J, McKance-Katz E, Rosenblatt W, Fingado C, Zoghbi S, Baldwin R, Seibyl J, Krystal J, Charney D, Innis R: SPECT imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Natl Acad Sci USA *93:*9235, 1996.

*Laruelle M, D'Souza CD, Baldwin RM, Abi-Dargham A, Kanes SJ, Fingado CL, Seibyl JP, Zoghbi SS, Bowers MB, Jatlow P, Charney DS, Innis RB: Imaging D_2 receptor occupancy by endogenous dopamine in humans. Neuropsychopharmacology *17:*162, 1997.

Logan J, Fowler J, Volkow N, Wolf A, Dewey S, Schlyer D, MacGregor R, Hitzemann R, Bendriem B, Gatley S, Christman D: Graphical analysis of reversible radioligand binding from time-activity measurements applied to [N-^{11}C-methyl]-(-)cocaine PET studies in human subjects. J Cereb Blood Flow Metab *10:*740, 1990.

Malison RT, Best S, van Dyck C, McCance E, Wallace E, Laruelle M, Baldwin R, Seibyl J, Price L, Kosten T, Innis R. Elevated striatal dopamine transporters in acute cocaine abstinence as measured by [^{123}I]b-CIT SPECT. Am J Psychiatry *155:*832, 1998.

Malison RT, McCance E, Carpenter L, Baldwin R, Seibyl JP, Price L, Kosten T, Innis R: [^{123}I]b-CIT SPECT imaging of dopamine transporter availability after mazindol administration in human cocaine addicts. Psychopharmacology *137:*321, 1998.

Malison RT, McDougle CJ, van Dyck CH, Scahill L, Baldwin RM, Seibyl JP, Price LH, Leckman JF, Innis RB: [^{123}I]b-CIT SPECT imaging demonstrates increased striatal dopamine transporter binding

in Tourette's syndrome. Am J Psychiatry *152:*1359, 1995.

*Phelps M, Mazziotta J, Schelbert H: *Positron Emission Tomography and Autoradiography: Principles and Applications for the Brain and Heart*. Raven Press, New York, 1986.

*Seibyl JP, Marek KL, Quinlan D, Sheff K, Zoghbi SS, Zea-Ponce Y, Baldwin RM, Fussell B, Smith EO, Charney DS, Hoffer PB, Innis RB: Decreased single-photon emission computed tomographic [^{123}I]b-CIT striatal uptake correlates with symptom severity in idiopathic Parkinson's disease. Ann Neurol *38:*589, 1995.

Sorensen JA, Phelps ME: *Physics in Nuclear Medicine*, ed 2. WB Saunders, Philadelphia, 1987.

*Ungerleider L: Functional brain imaging studies of cortical mechanisms for memory. Science *270:*769, 1995.

Volkow ND, Wang G-J, Fischman MW, Foltin RW, Fowler JS, Abumrad NN, Vitkun S, Logan J, Gatley SJ, Pappas N, Hitzemann R, Shea CE: Relationship between subjective effects of cocaine and dopamine transporter occupancy. Nature *386:*827, 1997.

Wong DF, Gjedde A, Wagner HN: Quantification of neuroreceptors in the human brain. I. Irreversible binding of ligands. J Cereb Blood Flow Metab *6:*137, 1986.

Wong DF, Gjedde A, Wagner HN, Dannals RF, Douglass KH, Links JM, Kuhar MJ: Quantification of neuroreceptors in the living brain. II. Inhibition studies of receptor density and affinity. J Cereb Blood Flow Metab *6:*147, 1986.

Wong DF, Wagner HN, Tune LE, Dannals RF, Pearlson GD, Links JM, Tamminga CA, Borussolle EP, Ravert HT, Wilson AA, Toung T, Malat J, Williams MA, O'Tuama LA, Snyder SH, Kuhar MJ, Gjedde A: Positron emission tomography reveals elevated D_2 dopamine receptors in drug-naive schizophrenics. Science *234:*1558, 1986.

1.16 PRINCIPLES OF NEUROIMAGING: MAGNETIC RESONANCE TECHNIQUES

CRAIG N. KARSON, M.D., AND PERRY F. RENSHAW, M.D., ĦH.D.

Most atoms have a nuclear magnetic moment. A strong external magnetic field will align the magnetic moment of atoms along the axis of the magnetic field. The magnetic moment of atoms aligned with a strong external field can be deflected off axis by a radiofrequency pulse at the *nuclear resonant spin frequency* or *Lamour frequency*. In turn, a detectable radiofrequency signal will be generated by the realignment of magnetic moments with the external magnetic field once the radiofrequency pulse is turned off. The detection of this signal forms the basis of all magnetic resonance experiments. In *magnetic resonance imaging* (MRI), magnetic field gradients are used to encode the locations from which these radiofrequency signals arise. In *magnetic resonance spectroscopy* (MRS), slight differences in the resonance frequencies of particular atoms in different molecular environments can be used to measure the concentrations of specific neurochemicals.

Once the magnetic moment of an atom has been deflected off the axis of the main magnetic field, it will return to its equilibrium orientation. The rate at which the magnetic moment returns to equilibrium is described by two relaxation time constants. The realignment of atomic spins with the external magnetic field is characterized as T1 (spin-lattice) relaxation; T2 relaxation describes the rate at which the coherence of the magnetization decays after an initial radiofrequency pulse deflects the spins off axis. In MRI, images may be collected using different acquisition parameters and the resulting images may be described as being either T1-or T2-weighted.

MRI first became available for clinical studies in the 1980s and allows for detailed in vivo study of brain structure. Today, the use of MRI to rule out overt structural changes in the brain resulting from neurological illness remains the only agreed-upon clinical use for any magnetic resonance technique in the practice of psychiatry.

Magnetic resonance spectroscopy constitutes a second group of relevant magnetic resonance techniques. The first magnetic resonance experiments involved the spectroscopic detection of MRS signals from in vitro samples more than 50 years ago. At the present time, both in vitro and in vivo MRS experiments are feasible, and clinical scanners may be used to measure the concentration of many molecules in virtually any organ system based on the detection of radiofrequency signals from magnetic resonance-visible nuclei (Table 1.16-1). An example of an in vitro application would be the study of extracts from frozen brain tissue. In vivo MRS studies to evaluate human neurochemistry typically exploit the magnetic properties of the principle isotopes of hydrogen (^{1}H or proton) or phosphorus (^{31}P) atoms. Proton MRS can provide information about the concentration of several neurochemicals in tissue, including *N*-acetyl aspartate (NAA), an established neuronal marker. Phosphorus MRS provides data about the concentration of bioenergetic phosphates such as phosphocreatine (PCr) and adenosine triphosphate (ATP) in the brain. Finally some MRS-visible atoms like fluorine (^{19}F) and lithium (^{7}Li) are found in the molecular structure of psychotropic medications, that is, fluoxetine (Prozac) and lithium (Eskalith), respectively, and the concentration of these medications and their metabolites may be determined spectroscopically.

Nucleus	Spin Quantum Number	NMR Frequency at 10 kG (Mhz)	Relative Sensitivity at Constant Field
^{1}H	1/2	42.58	1.00
^{19}F	1/2	40.05	0.83
^{7}Li	3/2	16.55	0.29
^{23}Na	3/2	11.26	0.09
^{31}P	1/2	17.25	0.06
^{13}C	1/2	10.71	0.02
^{39}K	3/2	1.99	0.0005

Table 1.16-1 Relative NMR Sensitivities

Since the late 1980s ultrafast scanning techniques such as echo planar imaging that allow the collection of complete image planes in less than 100 ms have become clinically available. It has been demonstrated that by exploiting the high-temporal resolution of these methods, it is possible to construct images that are sensitive to changes in regional cerebral metabolism. This family of MR techniques is often referred to as *functional MRI* (fMRI). Current interest in fMRI methods for psychiatric research reflects the opportunity to probe brain activity in vivo without the need for ionizing radiation. Repeated studies of individual subjects, as well as studies of women and children, are now feasible. Additionally, there will be widespread availability of fMRI because every high-field MRI scanner can be equipped with a high-speed imaging coil.

Clinical MRI, MRS, and fMRI have evolved as separate disciplines. The rate of technical development continues at a rapid pace. Improvements in signal-to-noise ratios continue, not only with the fast scanning techniques, but with the introduction of human magnets that operate at higher field strengths, such as 3 or 4 tesla. And, unlike other neuroimaging techniques, the end of this evolution is not in sight, suggesting that MRI may eventually dominate psychiatric neuroimaging in a way that could not have been imagined only a short time ago.

MAGNETIC RESONANCE IMAGING

Psychiatry was introduced to MRI via the striking images of the human brain obtained in vivo by MRI scans. The appearance of these images is heavily dependent upon the scan acquisition parameters. In practice, images are often referred to as being either T1-weighted or T2-weighted. In T1-weighted images, water is dark; in T2-weighted images, water is bright (Fig. 1.16-1). While examinations of brain involve both T1- and T2-weighted images, T1-weighted images are best for anatomical studies. Neoplasia and inflammation are often associated with increased water content, most vividly rendered on T2-weighted images. Magnetic resonance images may be acquired both before and after the administration of a contrast agent, which is typically a strongly paramagnetic gadolinium chelate. These agents work by altering the local magnetic environment and, by extension, tissue relaxation times. Magnetic resonance contrast agents will diffuse from the intravascular to the extracellular space when there is some compromise of the blood-brain barrier (e.g., in the case of cerebral neoplasm or inflammation).

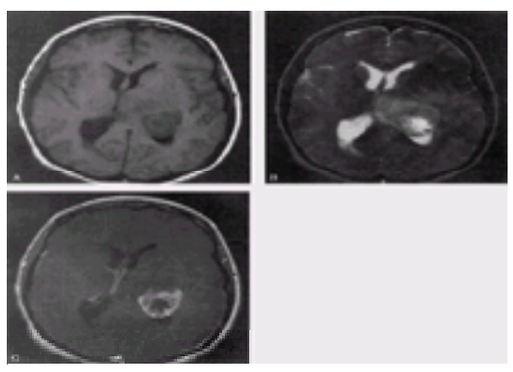

FIGURE 1.16-1 Three axial images from a 46-year-old woman who was hospitalized for the first time for depression and suicidality following the end of a longstanding relationship. A malignant neoplasm extending into the posterior aspect of the left lateral ventricle is clearly seen in all three images. Images **A** and **B** are T1-and T2-weighted, respectively. Image **C** demonstrates the effects of postcontrast enhancement.

Relatively few studies have substantiated the utility of MRI for the evaluation of organic brain pathology in patients with mental illness. Initial studies have suggested that 10 to 17 percent of subjects with a primary psychiatric disorder have evidence of a focal abnormality on brain MRI. More recently, investigators focused on unanticipated findings that might lead to a change in treatment in a cohort of 6200 consecutively referred psychiatric inpatients in a single institution. Ninety-nine subjects (1.6 percent) were noted to have findings that might lead to a change in clinical management. The most common new diagnosis was that of multiple sclerosis, which was present in approximately 0.8 percent of all subjects, a rate that is roughly 13 times higher than the prevalence of multiple sclerosis in the general population. The majority of subjects with unsuspected multiple sclerosis had been diagnosed as having refractory affective illness.

On the basis of existing data it has been suggested that structural brain imaging may be most appropriate for subjects with a new-onset psychosis, a new-onset dementia or delirium, or for subjects with other acute mental status changes in the presence of specific risk factors (i.e. advanced age, history of head trauma, or neurological signs). Data supporting the use of MRI for the detection of intracranial pathology in patients with longstanding psychotic disorders comes from large-scale autopsy studies of patients who had been hospitalized long-term. Intracranial neoplasms are approximately 60 percent more common at death in patients with psychotic disorders (3.7 percent) than in age-matched individuals who die of general medical illness (2.3 percent). A recent review of reversible dementia concluded that treatable conditions were likely to be found in approximately 1 percent of patients. Although normal-pressure hydrocephalus, neoplasm, and subdural hematoma may be detected on brain MRI in patients with dementia, the most common reversible causes of new-onset memory loss are depression, medication effects, and metabolic derangements.

In clinical practice structural brain imaging may be performed using either computed tomography (CT) or MRI. Given the fact that MRI provides more detail with regard to different brain tissues (e.g., gray versus white matter), it probably provides a higher yield when investigating possible central nervous system pathology. CT may be more widely available than MRI and is less expensive. In some circumstances, such as acute trauma, possible acute intracranial bleeding, or when MRI is relatively contraindicated, CT may be the preferred imaging modality. For each individual case it is most helpful if the referring psychiatrist specifies the clinical questions to be addressed by the scan to the local neuroradiologist. The better the information provided, the better the chances that an optimal study will be performed.

Neuromorphometry of Psychiatric Disorders MRI has largely succeeded CT as the psychiatric research tool of preference for the evaluation of structural changes in the brain in patients with psychiatric disorders, particularly schizophrenia. Software improvements have now made such neuromorphometric studies easier to perform, although many of the limitations of the older CT research still present confounds for MRI studies. Specifically, there may be evidence of diffuse change, such as atrophy, and many pathological findings on MRI, such as white-matter hyperintensities, that do not suggest specific causes.

Schizophrenia CT studies of patients with schizophrenia provided compelling evidence that a substantial fraction of these patients had reduced cerebral mass, as evidenced by enlargement of the brain ventricles and cortical sulci. Because of its very high spatial resolution and absence of exposure to radiation, MRI was introduced as a psychiatric research tool that might better resolve the specific nature of the structural abnormalities of the brain in schizophrenia. Since the late 1980s, improvements in image-processing tools have made in vivo computerized neuromorphometry a reality.

However, even with the increased resolution provided by MRI scanning, it remains unclear whether there are relatively specific structural brain changes in substantial fractions of patients with schizophrenia or if cerebral mass reduction is nonspecific. Several studies have provided convincing data regarding reduced volume in specific cortical regions, including prefrontal cortex and temporal cortical regions. Such changes have even been found in patients early in the course of illness, suggesting that the structural changes observed in the cortex are present at the onset of the illness and do not stem from progression of the illness or from iatrogenic factors. Enlargement of caudate volume has also been reported early in the course of the illness, but multiple studies suggest that caudate enlargement may be secondary to treatment with typical dopamine receptor antagonist agents. More recently it has been suggested that this enlargement recedes with clozapine (Clozaril) treatment.

Eloquent studies now combine structural imaging through MRI scans with functional imaging techniques, including positron emission computed tomography (PET) and single photon emission computed tomography (SPECT), using computerized coregistration of the image sets. For example, one study found that a sample of patients with schizophrenia had reduced metabolic activity in the anterior left portion of the thalamus as measured by ^{18}F-fluorodeoxyglucose PET and also had reduced volume in the same area as measured by MRI scan. As the thalamus has extensive afferent and efferent projections throughout cortex and receives massive input from structures in the brainstem reticular formation, these observations of altered thalamic architecture and activity are difficult to interpret in isolation. Resulting neural network hypotheses about schizophrenia have gained primacy and require that changes in volume (usually reduction) and activity be demonstrated consistently over a constellation of structures in the disorder.

Mood Disorder Although there have been fewer studies of mood disorders than of schizophrenia, some consistent findings have emerged. In major depressive disorder, smaller volumes of the frontal lobe, the cerebellum, the caudate, and the putamen have been reported. On the basis of these observations, neuroanatomic models of mood regulation involving specific frontosubcortical circuits have been proposed.

In patients with bipolar I disorder, the most commonly reported finding is the presence of an enlarged third ventricle. As the third ventricle lies adjacent to the hypothalamus, some researchers have suggested that hypothalamic dysfunction may play a role in the pathogenesis of bipolar I disorder. Less commonly, decreased volumes of the cerebellum and temporal lobes have also been noted.

In terms of pathological findings, several groups have reported an increased incidence of white matter hyperintensities in individuals with bipolar I disorder; unfortunately, neither the cause nor the clinical significance of these hyperintensities has been clearly established.

Obsessive-Compulsive Disorder Because of theoretical constructs suggesting caudate nucleus involvement and an early report of right caudate enlargement in patients with obsessive-compulsive disorder several MRI studies have focused on this structure, but compelling evidence for any change in caudate volume does not exist. A suggestion that this disorder is associated with nonspecific reductions in white matter as well as increases in cortical volume awaits replication.

Substance Abuse Investigations of structural brain changes in the various forms of substance abuse are confounded by two main issues. First, many types of substance abuse also involve behavior patterns that place patients at substantial risk for serious brain injury (i.e., fighting, motor vehicle accidents, and severe malnutrition). Separating the relatively subtle brain-toxic effects of a single substance, such as alcohol, from such direct and major brain injury is very difficult. In addition, polysubstance abuse tends to prevail in patients who are dependent on multiple illicit or prescription drugs, making it very difficult to attribute a subtle change to a single agent.

Of the substance abuse disorders, alcohol dependence has been the most thoroughly studied using computerized MRI scan neuromorphometry. Current evidence suggests that alcohol dependence leads to generalized reductions in brain mass as reflected by decreases in cortical gray and white matter and increased cerebrospinal fluid volume of the cerebral ventricles, particularly the lateral ventricles. Many of the changes may be at least partially reversed during sustained periods of abstinence. Reduction in the mass of the corpus callosum has been reported in one study, and a reduction of volume in the anterior lobe of the cerebellar vermis in the offspring of alcoholic mothers has also been noted.

Psychiatric Disorders in Childhood To date, the number of MRI studies in childhood psychiatric disorders is relatively small, and the early studies clearly suffered because they did not employ standardized morphometric analysis methods. In turn, some early findings, such as reduced volume of the cerebellar vermis in childhood autism, proved difficult to replicate. Disorders that have received special attention include autistic disorder, fragile X syndrome, Down syndrome, schizophrenia, obsessive compulsive disorder, Tourette's disorder, attention-deficit/hyperactivity disorder, and reading disorder.

Early studies of patients with attention deficit hyperactivity disorder suggested that the disorder was associated with reduced volume in the right prefrontal cortex and corpus callosum. A well-controlled study of 57 patients with attention deficit hyperactivity disorder and 55 age-matched controls using computerized morphometry finds attention deficit hyperactivity disorder associated with reduced volume of the right anterior frontal cortex and globus pallidus and loss of the normal right-left

asymmetry of the caudate. Based on these data, dysfunction of right prefrontal-striatal neural systems is hypothesized for attention deficit hyperactivity disorder.

In childhood schizophrenia the structural abnormalities in the brain mimic those seen in the adult disorder. These include lateral ventricular enlargement and a trend towards decreased cerebral volumes.

MAGNETIC RESONANCE SPECTROSCOPY

Under the right conditions, the actual concentration of an element with certain magnetic properties can be measured in specific brain regions by MRS. In human studies, several issues determine whether the capacity for MRS detection can be translated into an actual application. Identification and reliable measurement of cerebral metabolites requires the compound of interest to have a sufficiently high concentration and to give rise to a distinct MRS resonance. The concentration of the molecule also determines the spatial resolution that may be obtained.

When an element of interest not normally present in the human body (no background signal) is introduced into the human body and has a strong radiofrequency signal, MRS resonances may be detectable. The most common route of introduction for such elements is as a constituent element of pharmacotherapeutic molecules. For example, ^{19}F has a very high MRS sensitivity (see Table 1.16-1). Trifluorinated medications, including fluoxetine, trifluoperazine (Stelazine), and fluphenazine (Permitil) reach sufficiently high concentrations during routine pharmacotherapy to yield a signal of sufficient intensity for MRS detection in vivo in human brain. Polycyclic fluorinated drugs typically achieve brain concentrations on the order of 1 to 50 mmol, and ^{19}F MRS signals are usually recorded from the whole brain. ^{7}Li has a modest MRS sensitivity, but is present in brain at concentrations ranging from 0.1 to 1.0 millimolar when plasma lithium concentrations are within the therapeutic range. In a typical ^{7}Li MRS study, signals may be detected from brain regions ranging in volume from 100 to 1000 cm^3. Historically, one of the first research applications of MRS in psychiatry was to examine the pharmacokinetics of fluoxetine, antipsychotic medications, and ^{7}Li in the human brain.

Different issues arise when attempting to apply MRS to elements that are intrinsic to the human body. The contrast between signal from the molecule of interest and other molecules containing the MRS nucleus may be limited. For example, sodium (^{23}Na) has a strong magnetic moment, but is present at similar concentrations throughout the human body, and very few MRS applications have been developed for measuring ^{23}Na concentration in the human brain. ^{1}H is virtually ubiquitous in the human body, both in water as well as in a wide range of biomolecules. By suppressing the ^{1}H MRS signal from water protons, signals from endogenous metabolites with concentrations in the millimolar range may be detected in brain regions as small as 0.25 cm^3. Compared to ^{23}Na MRS, in which only a single resonance line is observed, hydrogen atoms that are present in different magnetic environments have slightly different resonance frequencies. Thus, the concentrations of several different compounds are typically detected in a ^{1}H MRS study. In addition, a family of sophisticated data acquisition techniques, known as *spectral editing*, facilitates the observation of resonance lines from some compounds, such as g-aminobutyric acid (GABA), which might not be detected otherwise. The concentration of proton MRS visible compounds vary throughout different brain regions and can be affected by disease processes or medications.

In vivo MRS may also be used to measure the concentration of high-energy phosphate metabolites using ^{31}P MRS. In studies of neuropsychiatric disorders, it is of relevance to explore which brain regions have altered concentrations of nucleoside triphosphate, ATP in brain, and phosphocreatine. Although the signal contrast principles are similar to proton MRS, the relatively weak contrast achieved has limited the use of ^{31}P MRS to very large brain regions. At 1.5 tesla, ^{31}P MRS data is usually obtained from brain regions on the order of 25 to 50 cm^3.

MRS has several unique features in contrast with other neuroimaging techniques, some of these are advantageous, others are not. One advantage is that the potential for signal detection for an element at a given concentration in the brain can be modeled by dissolving the molecule of interest in water and then placing it in a 1-liter flask, which is studied as if it is the human head; this is called a *phantom*. Another advantage is that MRS can be applied directly in animal or in vitro experiments that are related to in vivo work in human subjects. If the effect of a medication on the proton spectrum is being examined in a group of human subjects, it would be possible to examine the proton spectrum in the brain of a rat treated with the same medication or in the extracted cytosolic or membrane fractions of brain tissue taken from the same treated rat. Post mortem samples of human brain tissue taken from a deceased subject who was treated with the same medication could also be studied in vitro. The final and perhaps most important advantage of MRS when compared to other in vivo neurochemical measurement techniques is that the patient is not exposed to ionizing radiation. In subjects free of body metal or those implanted with biomedical devices such as pacemakers, MRS can be performed an unlimited number of times and at any interval required by the clinical or experimental situation.

A disadvantage of MRS is the requirement for specially built detection coils; only the in vivo proton MRS uses the standard clinical MRI apparatus for signal detection. ^{19}F MRS, ^{7}Li MRS, and ^{31}P MRS require the construction of a separate radiofrequency detection coil for each separate element. This time-consuming exercise requires considerable resources and expertise that are not normally available in clinical MRI units. Thus, much of the current focus in MRS is in the area of proton MRS.

Proton MRS The hydrogen atom is the most sensitive nucleus for MRS (Table 1.16-1). Proton MR spectra document signals from a number of compounds including NAA, cytosolic choline-containing compounds, Inositol and lactate (Fig. 1.16-2 and Fig. 1.16-3). Studies designed to identify alterations in brain levels of each of these compounds have been reported in patients with neuropsychiatric disorders.

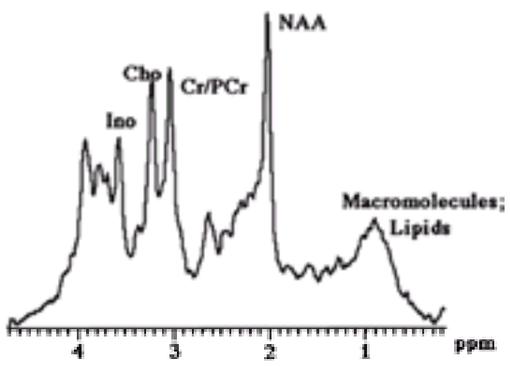

FIGURE 1.16-2 Proton magnetic resonance spectra. A is a short echo time (TE = 30 msec) STEAM spectrum from a 32-year-old healthy volunteer. Data was acquired from an 8-cm^3 volume located in the left, anterior frontal lobe. Note the prominent Inos resonance. Cho = cytosolic, choline-containing compounds; Cr/PCR = creatine and phosphocreatine; NAA = *N*-acetyl-L-aspartate.

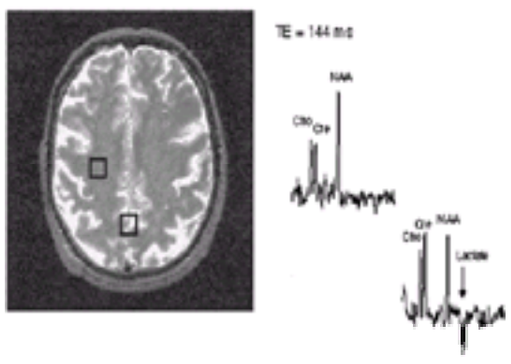

FIGURE 1.16-3 B displays two long echo time (TE = 144 ms) spectra from two locations (**A**) obtained from a 56-year-old man with mitochondrial myopathy, encephalopathy, lactacidosis, and stroke (MELAS). This disorder has been associated with increased brain lactate concentrations, best seen in the inverted pair of resonance lines at 1.4 PPM in the occipital region. (Courtesy of Lawrence L. Wald, Ph.D.)

Schizophrenia The most intensively studied ^{1}H MRS resonance line arises from NAA. One of the most consistently reported ^{1}H MRS findings in the psychiatric literature is the observation of decreased NAA resonance intensities in the hippocampus and frontal lobes of patients who have schizophrenia. Decreased NAA concentrations have also been observed in the temporal lobes of individuals with a first episode of psychosis, suggesting that this finding is present early in the course of illness. Further studies are necessary to clarify the degree to which this deficit increases with illness progression and to determine the degree to which this finding is specific to schizophrenia. An in vitro study of postmortem brain samples done on a 7-T unit suggests that concentrations of some amino acids such as glutamate, are reduced in schizophrenia. In vivo study of these lower concentration compounds will be facilitated by new human magnets that are 3 and 4 T.

Mood Disorders To date, most MRS studies relevant to mood disorders have employed either ^{19}F or ^{7}Li MRS in order to evaluate brain levels of pharmacological agents such as fluoxetine and lithium. However, another prominent brain ^{1}H MRS resonance line arises from cytosolic, choline-containing compounds. The vast majority of brain choline is present in the phospholipids phosphatidylcholine and sphingomyelin, which have restricted molecular motion and so do not contribute to the ^{1}H MRS resonance.

A role for altered choline metabolism in the pathophysiology of mood disorders was first proposed as an “adrenergic-cholinergic balance” manifested in mania by cholinergic underactivity and adrenergic overactivity. Several clinical studies have also demonstrated that the oral administration of both phosphatidylcholine and choline provides a moderate antimanic effect. Additionally, ^{7}Li exerts a potent and specific inhibitory effect on membrane Cho transport in man. Accumulating evidence has demonstrated that phosphatidylcholine hydrolysis by phospholipases A $_{2}$, C, and D is a widespread response elicited by many factors. These include growth factors, cytokines, neurotransmitters, hormones, and other extracellular signals. The molecules produced by this cleavage play an important role in intracellular message transduction.

Proton MRS provides a window on choline metabolism in the human brain that may have special relevance for the study of mood disorders illness. In this regard, a recent study of individuals with major depression has suggested that the intensity of the ^{1}H MRS choline resonance is altered in the basal ganglia and frontal lobes of depressed individuals. Future studies will be needed in order to determine the extent to which the ^{1}H MRS choline resonance is altered in other brain regions, as well as the degree to which these alterations are mood-state dependent.

Substance Abuse A limited number of ^{1}H studies relevant to substance use disorders have been conducted. Most attention has been paid to the fact that brain ethanol can be readily observed using ^{1}H MRS and that the relaxation or visibility of ethanol seems to be greater in tolerant than in nontolerant individuals. Extensions of this work might ultimately lead to diagnostic tests for ethanol tolerance and, by implication, for ongoing consumption. Elevated concentrations of creatine and myoinositol in abstinent cocaine users suggest that ^{1}H MRS may have a role in studying cocaine-related disorders, as well.

Panic Disorder A series of carefully designed ^{1}H MRS experiments in subjects with panic disorder have been undertaken that demonstrate that the intravenous infusion of lactate is associated with an excess increase in brain lactate levels in patients who experience a panic attack during the infusion. Later studies find that controlled hyperventilation increases brain lactate and does so disproportionately in subjects with panic disorder. Taken together, these studies have provided direct evidence that an important pathophysiological component of panic disorder may consist of an elevated cerebrovascular response to hypocapnia.

Dementia A number of research groups have used ^{1}H MRS to investigate possible neurochemical changes in patients with probable Alzheimer's disease. The most consistently reported observation has been that of decreased brain NAA concentrations. This decrease has been noted in many brain regions and may be most severe in the medial temporal lobe. Decreased concentrations of NAA in subjects with probable Alzheimer's disease most likely reflect a loss of neurons, which is consistent with structural neuroimaging studies that have generally noted progressive brain atrophy as being characteristic of the illness.

A more surprising finding has been the observation of increased brain inositol concentrations in patients with Alzheimer's disease. In occipital lobe gray matter, concentrations of inositol may be increased by as much as 50 percent relative to values obtained from matched comparison subjects. Although the cause of the observed increase in inositol is not clear, some clinical imaging centers now perform ^{1}H MRS as a diagnostic test for patients with memory loss. Additionally, some investigators have suggested that the ratio of NAA to inositol may serve as a useful measure of disease progression for following subjects with probable Alzheimer's disease over time.

Finally, a selective decrease in cholinergic neurons has been a consistent finding in postmortem studies of Alzheimer's disease brain tissue. Given that it is used for the synthesis of membrane lipid as well as of the neurotransmitter acetylcholine in cholinergic neurons, brain choline is almost exclusively dietary in origin. Recent studies have demonstrated that brain choline uptake may be directly observed by ^{1}H MRS following the oral administration of choline salts and that this uptake decreases markedly with increasing age, an observation that may provide a rationale for the fact that of the Alzheimer's dementia type is a disorder that almost always begins in late life. Additional studies are required to determine the relationship between decreases in brain choline uptake with advancing age and the development of dementing illness.

^{31}P MRS: DIRECT LOOK AT BRAIN BIOENERGETICS Much of the energy for biochemical reactions in the human brain is supplied when ATP is metabolized to adenosine diphosphate, a phosphodiester, in a reaction that also releases inorganic phosphate. ^{31}P MRS is able to detect and differentiate signal characteristics from the phosphorus moiety on each of these molecules and therefore to provide a window into the bioenergetic state of the brain (Fig. 1.16-4). In addition, this technique can also detect resonance lines arising from phosphomonoesters, which serve as membrane precursors, and from phosphodiesters, which consist largely of membrane catabolites. Alterations in the relative levels of phosphomonoesters and phosphodiesters provide some insight into the state of the brain membranes. However, MRS studies are severely limited by the problem of poor signal-to-noise contrast. Until recently, the most practical method for increasing ^{31}P MRS signal involved increasing the volume of measurement. As a result, the first generation of ^{31}P MRS studies did not provide the spatial resolution to make compelling anatomic correlates, and most neuropsychiatric studies focused on very large brain areas such as the entire frontal cortex. The recent development of methods for proton decoupling makes it possible to obtain brain ^{31}P MRS spectra with narrower resonance lines and increased signal-to-noise ratios. Additionally, the installation of three and four tesla whole-body magnets will also increase the utility of ^{31}P MRS for studies of cerebral metabolism.

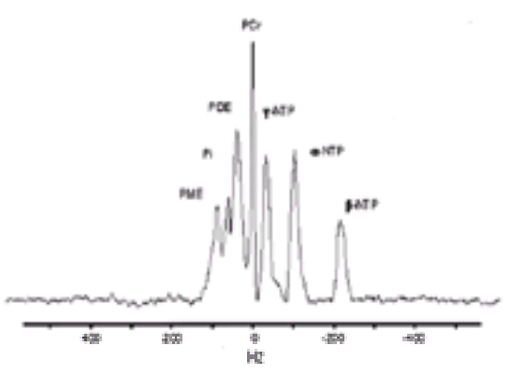

FIGURE 1.16-4 Phosphorus (^{31}P) MR spectrum. Data is acquired from an axial brain slice 50 cm in thickness. PME, phosphomonoesters; PDE, phosphodiesters; Pi, inorganic phosphate; PCr, phosphocreatine; NTP, nucleoside triphosphate. (Courtesy of Marc J. Kaufman, Ph.D.)

Schizophrenia Many studies suggest that schizophrenia seems to be associated with an increase in the concentration of phosphodiesters in the broadly defined frontal lobes. Based on the complementary finding of reduced phosphomonoester concentration in the frontal region, it has been suggested that the increase in phosphodiesters reflects a decrease in their activity and an increase in the activity of some phospholipase species. Alternatively put, schizophrenia patients may have decreased synthesis and increased breakdown of membrane phospholipids. These results, obtained in patients who had never been medicated as well as those who had been medicated recently suggest that the changes of the ^{31}P spectrum are present early in the course of the disease and may not relate to iatrogenic factors such as long-term treatment with antipsychotic medications. Increased concentration of phosphodiesters is also found in Alzheimer's disease, leading some to speculate about potential similarities in some aspects of the pathophysiology of Alzheimer's disease and schizophrenia. Although the neuropathological changes of Alzheimer's

disease are not found in schizophrenia, the high incidence of severe dementia in elderly institutionalized schizophrenia patients suggests that it is premature to dismiss the notion of pathophysiological commonalities between these two disorders.

Mood Disorder Low concentrations of phosphomonoesters and increased concentrations of phosphodiesters also occur in samples of patients with bipolar I disorder. Whether this suggests that at least in some cases schizophrenia and bipolar I disorder may share common pathophysiological markers or that treatment with ^{7}Li, which strongly impacts ^{31}P metabolism via its actions on inositol metabolism and choline transport, may alter other elements of phosphate metabolism is unclear.

Alcohol Abuse A limited number of ^{31}P MRS studies in patients with substance dependence limits the capacity to draw firm conclusions. Preliminary results may suggest that, in contrast to the other neuropsychiatric disorders noted, alcohol-dependent patients may have reductions in concentrations of phosphodiesters in the brain.

^{19}F MRS: TOWARDS BRAIN PHARMACOKINETICS OF ANTIDEPRESSANT AND ANTIPSYCHOTIC AGENTS Serum concentrations of either antidepressant or antipsychotic medications have limited clinical utility. Moreover, these classes of medications often have long half-lives and are highly lipid soluble. These factors suggest that the ability to measure the concentration of either of these classes of pharmacotherapeutic agents directly in the brain may provide useful information about clinical efficacy and the incidence of adverse effects.

Animal studies first suggested the possibility that ^{19}F MRS could be used to measure the concentration of trifluorinated antipsychotic drugs such as fluphenazine and trifluoperazine. Shortly thereafter, in vivo detection of such drugs in the human brain was accomplished by this technique. However, enthusiasm was limited by the very low signal-to-noise ratios and the unstable signal achieved in vivo: patients on low and moderate dosages of trifluoperazine often had large, easily detectable signals in contrast to patients on oral dosages of 120 mg a day who had no signal.

Brain Fluoxetine and Fluvoxamine Concentrations In contrast to the inconsistent signals achieved from the trifluorinated antipsychotic agents, a very consistent and robust signal was obtained from the ^{19}F constituent of the trifluorinated selective serotonin reuptake inhibitor, fluoxetine, the most prescribed antidepressant in the United States. In the first large study involving 21 subjects on routine doses ranging from 20 to 40 mg a day, brain concentrations initially reflected cumulative dose and reached steady state only after 6 months or more of treatment at concentrations ranging up to 11 μg per ml, about 20 times the concentration detected in plasma. These results have been widely replicated and probably reflect the long half-life of fluoxetine and its highly lipophilic nature.

Later studies have focused on the T1 and the anatomic origin of the nuclear magnet resonance (NMR) signal from ^{19}F in fluoxetine. To do this requires a specialized quadrature volume coil built in the laboratory. This contrasts to earlier work using a flat surface 16 cm in diameter placed on the front of the head, which was later modified to conform to the surface of the forehead.

In an experiment involving five subjects treated with fluoxetine, there was substantial variation in the T1 values, which apparently stems from real interindividual differences. Unfortunately, T1 is correlated with the corresponding brain fluoxetine concentration (C_F) at a Pearson r of 0.57 (mean T1 = 240 mx ± 75), suggesting a possible confound for the determination of fluoxetine concentrations as T1 appears in both the numerator and denominator of the calculation equation. This correlation of T1 with brain C_F may stem from fast exchange between unbound and bound fluoxetine. In brain, fluoxetine will undoubtedly bind to a variety of sites, which can be assumed to be relatively immobile. Binding fluoxetine, a small molecule, to macromolecular sites will resist ^{19}F's ability to align with a magnetic pulse and will lower the T1; unbound fluoxetine will have a longer T1 by comparison.

To demonstrate the utility of in vitro MRS to address questions about the in vivo state, in vitro ^{19}F MRS was performed on extracted brain tissue obtained at autopsy from a patient who was being treated with fluoxetine at the time of his death from natural causes. In these cytosolic (unbound) extracts, fluoxetine concentrations ranged up to fourfold higher than those seen in the intact tissue, supporting the notion that a relatively long T1 in some subjects may be related to the bound fraction of fluoxetine or its metabolites. These fundamental questions about previously “hard” assumptions, such as that T1 is similar from subject to subject, illustrate that MR remains in its early development.

In vivo methods similar to those described for fluoxetine, have now been applied to fluvoxamine (Luvox). Measured concentrations range from 3.0 to 12.7 μmol in a sample of patients receiving a mean dosages of 217 ± 68 mg a day. Moreover, for this fast-acting drug, the mean elimination half-life from the brain is 58 hours and the mean brain to plasma elimination half-life ratio is 2.4.

Lithium (^{7}Li) MRS in the Brain Like the ^{19}F atom, ^{7}Li is an element found in a frequently used psychotropic medication that is also not normally present in the human brain. Exogenously administered, ^{7}Li easily creates a sufficient brain concentration to be measured by MRS. Current MRS studies find that brain ^{7}Li concentrations are about half those in serum and peak at around 4 hours after the last dose of ^{7}Li, compared with at around 2 hours in serum. However, the extent to which altered ^{7}Li pharmacokinetics contributes to the variable clinical response to ^{7}Li therapy in patients with bipolar I disorder has not been determined. Additionally, studies of clinical populations that may have altered ^{7}Li pharmacokinetics, such as children, the elderly, or those with medical illness, have not been concluded.

Also promising are animal studies that suggest that the strong ^{7}Li signal makes it possible to ultimately map ^{7}Li distribution in human brain, potentially identifying the site or sites of ^{7}Li action. Another approach may be to combine localized ^{7}Li MRS with proton MRS to relate the actions of ^{7}Li to changes in brain Cho metabolism.

FUNCTIONAL MAGNETIC RESONANCE IMAGING

Functional brain imaging studies have historically been limited both by the need to use radioactive tracers as well as by poor temporal resolution. Recent developments in the area of MRI may largely surmount these limitations. First the development of high-speed, echo planar imaging devices has greatly enhanced the temporal resolution of MRI. With echo planar imaging single-image planes can be acquired in 50 to 100 msec or multiple-image planes can be acquired each second. fMRI studies, which may be performed with or without a high-speed MR scanner, selectively detect image parameters that are proportional to cerebral blood flow or blood volume. This strategy capitalizes on the fact that focal changes in neuronal activity usually are closely coupled to changes in cerebral blood flow and blood volume.

fMRI studies may be divided into two separate classes: those that make use of endogenous physiological factors to detect changes in cerebral activation: the *noncontrast* techniques, and those that require the intravenous administration of a paramagnetic agent: the *contrast* techniques. Noncontrast techniques make use of either T1-weighted pulse sequences to detect changes in blood flow or, more commonly, T2-weighted pulse sequences to detect changes in the local concentration of paramagnetic deoxyhemoglobin. The latter method has been referred to as *blood oxygen-level dependent* imaging. In a blood oxygen-level dependent imaging experiment, regional brain activation is associated with changes both in blood flow and blood volume, generally leading to a washout of paramagnetic deoxyhemoglobin and an increase in local signal intensity.

Blood oxygen-level dependent imaging studies are limited in several important aspects from the perspective of performing clinical research studies. At 1.5 tesla, the magnitude of the observed changes in signal intensity is relatively small. For instance, photic stimulation, which induces a substantial increase in occipital cortical blood flow, produces only a 2 to 4 percent increase in magnetic resonance signal intensity. Recent work also suggests that the magnitude of changes in blood oxygen-level dependent signal intensity may vary with subject's age and gender. Further, many medications directly affect vascular tone and modify changes in blood oxygen-level dependent signals presenting an important confound for studies of subjects with psychiatric illness. Finally, the uncoupling of cerebral blood flow and volume that occurs following cerebral activation and produces the blood oxygen-level dependent effect appears to resolve with prolonged stimulation. In response to these problems, many research groups are developing noncontrast fMRI methods that have a greater sensitivity to changes in cerebral blood flow.

The contrast method is a tracer kinetic technique that utilizes the bolus injection of a paramagnetic contrast agent to produce changes in tissue magnetic susceptibility and magnetic resonance image intensity. During the first pass of the contrast agent, magnetic resonance signal intensity may decrease by as much as 20 to 40 percent. This method may be used to map the distribution of cerebral volume at rest or to measure changes in response to cerebral activation. Resting maps of cerebral volume (Fig. 1.16-5 and Fig. 1.16-6) have been shown to correlate well with PET images of fluorodeoxyglucose uptake and with SPECT images of cerebral blood flow. Additionally, the recent development of a multiple bolus method for performing dynamic susceptibility contrast studies is likely to facilitate the measurement of the drug's effects on cerebral hemodynamics.

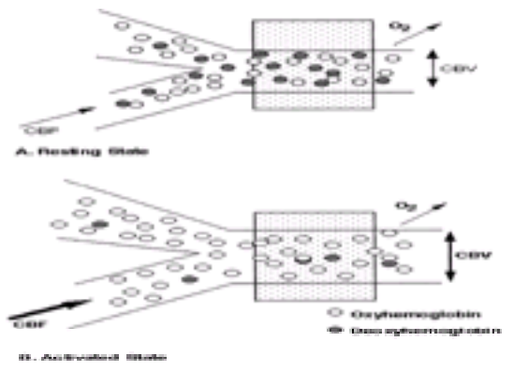

FIGURE 1.16-5 A, B. Mechanism of blood oxygen level dependent fMRI contrast. Regional brain activation is associated with disproportionate increases in blood flow and blood volume, leading to a washout of paramagnetic deoxyhemoglobin. (Courtesy of Jonathan M. Levin, M.D.)

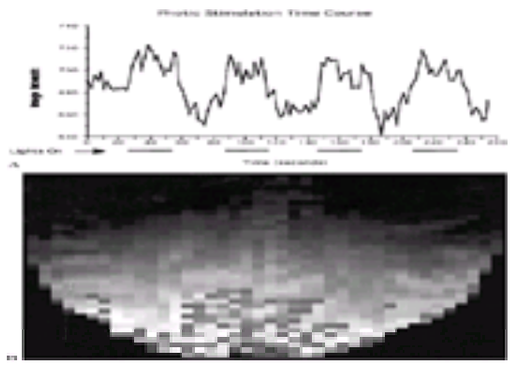

FIGURE 1.16-6 fMRI studies of photic stimulation. **A**, Change in image signal intensity in a single pixel during four alternating, 30-second epochs of flash photic stimulation. **B**, Activation maps demonstrating statistically significant increases in image signal intensity in primary visual cortex. Image data is acquired with a surface coil placed at the back of the head, thus explaining the decrease in image intensity with increasing distance from the occipital pole of the head. (Courtesy of Jonathan M. Levin, M.D., and Luis C. Maas.) (See Color Plate 1.)

Schizophrenia The promise that fMRI studies hold for the evaluation of patients with mental illness has recently been reviewed and schizophrenia is the disorder that has been studied the most intensively. Subjects with schizophrenia appear to have an increased blood oxygen-level dependent signal change in response to photic stimulation when compared to healthy volunteers. This finding, which was not anticipated, has been substantiated by recent PET studies in which an augmented cerebral blood flow increase was noted in schizophrenia subjects relative to comparison subjects during photic stimulation across a range of different frequencies. Additionally, a preliminary dynamic susceptibility contrast fMRI study has suggested that subjects with schizophrenia have an increased resting occipital blood volume. It is not clear to what extent these early findings result from altered vascular anatomy in the occipital cortex, altered cortical activation in response to photic stimulation, or the vasoactive effects of antipsychotic medication. Blood oxygen-level dependent signal changes in response to motor tasks have also been evaluated in schizophrenia patients. Finger-to-thumb opposition was associated with decreased activation of sensorimotor cortices and the supplementary motor area in subjects with schizophrenia relative to comparison subjects.

Although the use of blood oxygen-level dependent fMRI as a research tool is complicated by the large number of factors that can influence the magnitude of the signal change, altered patterns of cortical activation may serve as a more promising target for research. As an example, investigators have recently used blood oxygen-level dependent fMRI to evaluate changes in frontal and temporal lobe activation during word production in subjects with schizophrenia and in matched comparison subjects. Whereas the frontal lobe has decreased activation in subjects who have schizophrenia, the temporal lobe demonstrates increased activation. Similar findings have recently been demonstrated using PET blood flow studies and the identification of altered patterns of frontal-temporal circuitry in subjects with schizophrenia remains an active area of research.

Obsessive-Compulsive Disorder and Tourette's Disorder Patients with obsessive-compulsive disorder have been studied during symptom provocation using blood oxygen-level dependent fMRI. Brain regions that activated in the patient group in this study included the medial orbitofrontal, lateral frontal, anterior temporal, anterior cingulate, and insular cortices. These results are reasonably consistent with those reported by other groups using both PET and SPECT imaging.

fMRI has also been used to contrast the cerebral blood in Tourette's disorder patient, while they are suppressing tics versus when they allow the spontaneous expressions of their tics. Significant changes in signal intensity between the two states were demonstrated in the thalamus, and basal ganglia, and these changes correlated inversely with tic severity.

Alzheimer's Disease Dynamic susceptibility contrast MRI has also been used to evaluate changes in regional cerebral perfusion in patients with Alzheimer's disease. Based on the data from these studies, cerebral blood volume maps appear to provide data that is equivalent to that obtained by SPECT for establishing a diagnosis of probable Alzheimer's disease. However, data to generate cerebral blood volume maps is acquired over the course of 1 to 2 minutes and without the use of radiotracers. Thus, susceptibility contrast MRI may be particularly well suited, as compared to SPECT, for the evaluation of subjects with dementia. The ongoing development of therapeutic agents for Alzheimer's disease suggests that functional brain-imaging studies may play an important role in the evaluation of these drugs on cerebral metabolism. Additionally, it has recently been demonstrated that asymptomatic individuals at increased risk of developing Alzheimer's disease have decreased temporoparietal metabolism. These findings, if replicated, would bode very well for the use of dynamic susceptibility contrast fMRI as a safe diagnostic tool for the very large population of older adults at risk for developing dementing illness. Susceptibility contrast MRI is also being actively investigated as a tool for the evaluation of patients with a range of neurological disorders, including cerebral ischemia and infarction, cancer, and epilepsy.

Over time, fMRI will come to play an important role in psychiatric neuroimaging research. It offers an unprecedented opportunity for clinical neuropsychiatrists to directly evaluate many aspects of brain activity in both healthy and ill individuals. However, as with any new methodology, the promise that is so readily apparent must be balanced against the obstacles that will be identified with further research.

SUGGESTED CROSS-REFERENCES

Brain mechanisms of schizophrenia are discussed in Section 12.3, bipolar I disorder is discussed in Chapter 14, and obsessive-compulsive disorder is presented in Chapter 16. Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39, Chapter 40, Chapter 41, Chapter 42, Chapter 43, Chapter 44, Chapter 45, Chapter 46, Chapter 47, Chapter 48 and Chapter 49 discuss childhood psychiatric disorders, substance abuse is discussed in Chapter 11, and Alzheimer's disease is discussed in Section 51.3e.

SECTION REFERENCES

Aylward EH, Harris GJ, Hoen-Saric R, Barta PE, Machlin SR, Pearlson GD: Normal caudate nucleus in obsessive-compulsive disorder assessed by quantitative neuroimaging. Arch Gen Psychiatry *53:*577, 1996.

Belliveau JW, Rosen BR, Kantor HL, Rzedzian RR, Kennedy DN, McKinstry RC, Vevea JM, Cohen MS, Pykett IL, Brady TJ: Functional cerebral imaging by susceptibility contrast imaging. Magn Reson Med *14:*538, 1990.

Bertilino A, Callicott JH, Elman I, Mattay VS, Tedeschi G, Frank JA, Breier A, Weinberger DR: Regionally specific neuronal pathology in untreated patients with schizophrenia: a proton magnetic

resonance spectroscopic imaging study. Biol Psychiatry *43:*641, 1998.

Bertolino A, Nawroz S, Mattay VS, Barnett AS, Duyn JH, Moonen CTW, Frank JA, Tedeschi G, Weinberger DR: Regionally specific pattern of neurochemical pathology in schizophrenia as assessed by multislice proton magnetic resonance spectroscopic imaging. Am J Psychiatry *153:*1554, 1996.

Breier A, Buchanan RW, Elkashef A, Munson RC, Kirkpatrick B, Gellad F: Brain morphology and schizophrenia: A magnetic resonance imaging study of limbic prefrontal cortex and caudate structures. Arch Gen Psychiatry *49:*921, 1992.

*Breiter HC, Rauch SL, Kwong KK, Baker JR, Weisskoff RM, Kennedy DN, Kendrick AD, Davis TL, Jiang A, Cohen MS, Stern CE, Belliveau JW, Baer L, O'Sullivan RL, Savage CR, Jenike MA, Rosen BR: Functional magnetic resonance imaging of symptom provocation in obsessive-compulsive disorder. Arch Gen Psychiatry *53:*595, 1996.

Buchsbaum MS, Someya T, Teng CY, Abel L, Chin S, Najafi A, Haier RJ, Wu J, Bunney WE: PET and MRI of the thalamus in never-medicated patients with schizophrenia. Am J Psychiatry *153:*191, 1996.

*Cardoza JD, Herfkens RJ: *MRI Survival Guide*. Raven Press, New York, 1994.

Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Vaituzis AC, Dickstein DP, Sarfatti SE, Vauss YC, Shell JW, Lange N, Kaysen D, Krain AL, Ritchie GF, Rajapakse JG, Rapoport JL: Quantitative brain magnetic resonance imaging in attention-deficit hyperactivity disorder. Arch Gen Psychiatry *53:*607, 1996.

Chang L, Mehringer CM, Ernst T, Melchor R, Myers H, Forney D, Satz P: Neurochemical alterations in asymptomatic abstinent cocaine users: A proton magnetic resonance spectroscopy study. Biol Psychiatry *42:*1105, 1997.

Cohen BM, Renshaw PF, Stoll AL, Wurtman RJ, Yurgelun-Todd DA, Babb SM: Markedly decreased brain choline uptake in older adults: An in vivo proton magnetic resonance spectroscopy study. JAMA *274:*902, 1995.

*Dager SR, Friedman SD, Heide A, Layton ME, Richards T, Artru A, Strauss W, Hayes C, Posse S: Two-dimensional proton echo-planar spectroscopic imaging of brain metabolic changes during lactate-induced panic. Arch Gen Psychiatry *56:*70, 1999.

*Dager SR, Layton ME, Strauss W, Richards TL, Heide A, Freidman SD, Artru AA, Hayes DE, Posse S: Human brain metabolic response to caffeine and the effects of tolerance. Am J Psychiatry *156:*229, 1999.

*Dager SR, Strauss WL, Marro KI, Richards TL, Metzger GD, Artru AA: Proton magnetic resonance spectroscopy investigation of hyperventilation in subjects with panic disorder and comparison subjects. Am J Psychiatry *152:*666, 1995.

Exton JH: Messenger molecules derived from membrane lipids. Curr Opin Cell Biol *6:*226, 1994.

Harris GJ, Lewis RF, Satlin A, English CD, Scott TM, Yurgelun-Todd DA, Renshaw PF: Dynamic susceptibility contrast MRI of regional cerebral blood volume in Alzheimer's disease. Am J Psychiatry *153:*721, 1996.

Janowsky DS, El-Yousef MK, Davis JM, Sekerke HJ: A cholinergic-adrenergic hypothesis of mania and depression. Lancet *23:*632, 1972.

Karson CN, Lyon N, Bracha HS, Guggenheim FG: The profile of cognitive impairment in elderly dyskinetic subjects. J Neuropsychiatry Clin Neurosci *5:*61, 1993.

*Karson CN, Newton JEO, Livingston R, Jolly JB, Cooper TB, Sprigg J, Komoroski RA: Human brain fluoxetine concentrations. J Neuropsychiatry *5:*322, 1993.

Kaufman MJ, Levin JM, Christensen JD, Renshaw PF: Magnetic resonance studies of substance abuse. Semin Clin Neuropsychiatry *1:*61, 1996.

Komoroski RA, Newton JEO, Cardwell D, Sprigg J, Pearce J, Karson CN: In vivo ^{19}F spin relaxation and localized spectroscopy of fluoxetine in human brain. Magn Res Med *31:*204, 1994.

Komoroski RA, Newton JEO, Sprigg JR, Cardwell D, Mohanakrishnan P, Karson CN: In vivo ^{7}Li nuclear magnetic resonance study of lithium pharmacokinetics and chemical shift imaging in psychiatric patients. Psychiatry Res *50:*67, 1993.

Kwon JS, McCarley RW, Hirayasu Y, Anderson JE, Fischer IA, Kikinis R, Jolesz FA, Shenton ME: Left planum temporale volume reduction in schizophrenia. Arch Gen Psychiatry *56:*142, 1999.

Kwong KK, Belliveau JW, Chesler DA, Goldberg IE, Weisskoff RM, Poncelet BP, Kennedy DN, Hoppel BE, Cohen MS, Turner R, Cheng HM, Brady TJ, Rosen BR: Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. Proc Natl Acad Sci USA *89:*5675, 1992.

Levin JM, Ross MH, Renshaw PF: Clinical applications of functional MRI in neuropsychiatry. J Neuropsychiatry Clin Neurosci *7:*511, 1995.

Meyerhoff DJ, MacKay S, Sappey-Marinier D, Deicken R, Calabrese G, Dillon WP, Weiner MW, Fein G: Effects of chronic alcohol abuse and HIV infection on brain phosphorus metabolites. Alcohol Clin Exper Res *19:*685, 1995.

Miller BL, Moats RA, Shonk T, Ernst T, Woolley S, Ross BD: Alzheimer disease: Depiction of increased cerebral myo-inositol with proton MR spectroscopy. Radiology *187:*433, 1993.

Nopoulos P, Torres I, Flaum M, Andreasen NC, Ehrhardt JC, Yuh WTC: Brain morphology in first-episode schizophrenia. Am J Psychiatry *152:*1721, 1995.

Omori M, Pearce J, Komoroski RA, Griffin WST, Mrak RE, Husain MM, Karson CN: In vitro 1H-magnetic resonance spectroscopy of postmortem brains with schizophrenia. Biol Psychiatry *42:*359, 1997.

Patton RB, Sheppard JR: Intracranial tumors at autopsy in mental patients. Am J Psychiatry *113:*319, 1956.

Peterson BS: Neuroimaging in child and adolescent neuropsychiatric disorders. J Am Acad Child Adolesc Psychiatry *34:*1560, 1995.

Peterson BS, Skudlarski P, Anderson AW, Zhang H, Gatenby JC, Lacadie CM, Leckman JF, Gore JC: A functional magnetic resonance imaging study of tic suppression in Tourette syndrome. Arch Gen Psychiatry *55:*326, 1998.

Pettegrew JW, Keshavan M, Panchalinga MK: Alterations in brain high-energy phosphate and membrane phospholipid metabolism in first-episode, drug-naive schizophrenics. Arch Gen Psychiatry *48:*563, 1981.

Pfefferbaum A, Sullivan EV, Mathalon DH, Shear PK, Rosenbloom MJ, Lim KO: Longitudinal changes in magnetic resonance imaging brain volumes in abstinent and relapsed alcoholics. Alcohol Clin Exper Res *19:*1177, 1995.

Rauch SL, Renshaw PF: Clinical neuroimaging in psychiatry. Harvard Rev Psychiatry *2:*297, 1995.

Ramaprasad S, Komoroski RA: NMR imaging and localized spectroscopy of lithium. Lithium *5:*127, 1994.

Renshaw PF, Yurgelun-Todd DA, Cohen BM: Greater hemodynamic response to photic stimulation in schizophrenic patients: An echo-planar MRI study. Am J Psychiatry *151:*1493, 1994.

Satlin A, Bodick N, Offen WW, Renshaw PF: Brain proton magnetic resonance spectroscopy (1H-MRS) in Alzheimer's disease: changes after treatment with Xanomeline, an M1 selective cholinergic agonist. Am J Psychiatry *154:*1459, 1997.

Schroeder J, Wenz F, Schader LR, Baudendistel K, Knopp MV: Sensori-motor cortex and supplementary motor area charges in schizophrenia. Br J Psychiatry *167:*197, 1995.

Soares JC, Mann JJ: The anatomy of mood disorders—Review of structural neuroimaging studies. Biol Psychiatry *4:*86, 1997.

Strauss WL, Layton ME, Dager SR: Brain elimination half-life of fluvoxamine measured by ^{19}F magnetic resonance spectroscopy. Am J Psychiatry *155:*380, 1998.

Tsai G, Coyle JT: N-acetylaspartate in neuropsychiatric disorders. Prog Neurobiol *46:*531, 1995.

Weytingh MD, Bossuyt PMM, van Crevel H: Reversible dementia: More than 10% or less than 1%? J Neurol *242:*466, 1995.

Woodruff PWR, Wright IC, Bullmore ET, Brammer M, Howard RJ, Williams SCR, Shapleske J, Rossell S, David AS, McGuire PK, Murray RM: Auditory hallucinations and the temporal cortical response to speech in schizophrenia: A functional magnetic resonance imaging study. Am J Psychiatry *154:*1676, 1997.

*Yurgelun-Todd DA, Waternaux CM, Cohen BM, Gruber SA, English CD, Renshaw PF: Functional magnetic resonance imaging of schizophrenic patients and comparison subjects during word production. Am J Psychiatry *153:*200, 1996.

Textbook of Psychiatry

1.17 POPULATION GENETIC METHODS IN PSYCHIATRY

STEVEN O. MOLDIN, PH.D., AND IRVING I. GOTTESMAN, FH.D., F.R.C. PSYCH. (HON.)

The human genome consists of between 50,000 and 100,000 genes, of which over 20,000 have been identified. Over 5000 genetic disorders, each transmitted through a single mutant gene, have been characterized. The application of more powerful quantitative methods of analysis, new molecular technologies, and more detailed maps of the human genome have permitted localization to chromosomal regions of over 400 of these disease genes, with precise identification of more than 80.

There are major public health implications of identifying genes that influence an individual's risk for developing the more common familial mental disorders like schizophrenia, bipolar I disorder, alcoholism (alcohol abuse or dependent) and obsessive-compulsive disorder. Such findings may ultimately be of relevance to many affected individuals and their relatives, given the potential for developing a genetic test to identify individuals at risk and, most importantly, for providing the pharmaceutical industry with new drug therapy targets. Clinicians and researchers must understand the basic mathematical principles and quantitative methods of population genetics and genetic epidemiology so that they will be able to appreciate the relevance of new data derived from the genetic analysis of mental disorders.

MAIN SUBDIVISIONS OF GENETICS

The scientific study of heredity, which began with Gregor Mendel's work on peas in 1865, gradually developed into five major disciplines. *Biochemical genetics* is concerned with the biochemical reactions by which genetic determinants are replicated and produce their effects. *Developmental genetics* is concerned with the study of mutations that produce developmental abnormalities so that researchers can gain an understanding of how normal genes control growth and other developmental processes. *Molecular genetics* studies the structure and the functioning of genes at the molecular level. *Cytogenetics* deals with the chromosomes that carry those determinants. *Population genetics* may be subdivided into the partially overlapping fields of evolutionary genetics, genetic demography, quantitative genetics, and genetic epidemiology. Quantitative genetics and genetic epidemiology provide the mathematical methods to aid in the identification of genetic factors that influence risk to mental disorders.

DEFINITIONS

Population genetics deals with the mathematical properties of genetic transmission in families and populations. The primary goal of evolutionary genetics is to understand changes in gene frequency across generations. Genetic demography is primarily concerned with differential mortality and fertility (fitness) in human populations. Quantitative genetics and genetic epidemiology are the fields of population genetics that are of most relevance to the study of mental disorders. The goal of quantitative genetics is to partition variation in the observed differences between phenotypes into its genetic and environmental components. This area of study was developed largely to improve animals and plants through artificial selection and usually deals with continuous traits (e.g., milk yield or egg size), rather than discrete traits. Genetic epidemiology is explicitly directed toward understanding the causes, distribution, and control of disease in groups of relatives and the multifactorial causes of disease in populations. The mathematical principles of genetic epidemiology and quantitative genetics are central to *risk analysis*, which is the essential element in genetic counseling of familial disease, and to *linkage analysis*, which is a statistical procedure used to implicate a particular chromosomal region as containing a putative disease susceptibility gene. Figure 1.17-1 shows the focus of genetic epidemiology to be on genetic and environmental factors that interact in determining observed behavioral outcomes (disease). Differences between common and individual-specific environmental effects, and between genes of major versus minor effect, are discussed in the section on multilocus genetic models. *Psychiatric genetics* involves the specific application of genetic principles and methods to the study of mental disorders.

FIGURE 1.17-1 The network of genetic epidemiology.

BASIC ELEMENTS

A fundamental distinction in population genetics dating to Wilhelm Johannsen's work in 1909 is between *genotype* (an inferred set of genes) and *phenotype* (an observed effect of those genes); the distribution of the frequencies of the various phenotypes constitutes the essential description of a population. Gene frequencies are used to predict frequencies of genotypes and phenotypes and vice versa under a set of assumptions that include the following: (1) From the pattern of familial inheritance, the genotypes can be distinguished unequivocally, such that the frequencies of phenotypes are the same as those of the underlying genotypes. This relation between phenotypic and genotypic frequencies requires the related assumptions of negligible mutation rates and the occurrence of segregation of genes according to Mendel's laws. (2) There is no selection—that is, the expected number of fertile progeny from a mating that reaches maturity does not depend on the genotypes of the mates. (3) The population structure is such that all matings take place at random with respect to the genetic differences being considered in a population of infinite size. Consequently, the probability of mating between persons is in no way influenced by their genotype at a given locus.

A general theorem formulated in 1908 independently by Godfrey Harold Hardy and Wilhelm Weinberg is derived from these assumptions and fits the facts well in many cases. In its simplest form the *Hardy-Weinberg law* states that if respective gene frequencies of two alternative forms of a gene *(alleles) A* and *a* are *p* and *q*, the respective genotypic frequencies among progeny with genotypes *AA, Aa*, and *aa* are p^2, *2pq*, and q^2. This relation between gene frequencies and genotype frequencies is of the greatest importance because many of the deductions in quantitative and population genetics are based on it.

A basic distinction in population genetics of direct relevance to the analysis of mental disorders is that between *quantitative* and *qualitative* phenotypes. Can persons be classified to one of a small number of discrete classes, or can they be assigned a continuous score on an observed continuum? Disease phenotypes are qualitative—persons are classified as affected or unaffected. Contemporary genetic analysis usually posits an underlying but so far unobservable liability to affection that is continuous, with affected cases consisting of persons at an extreme end of the continuum. This is analogous to having height as the phenotype but only being able to measure tall versus nontall, rather than height in centimeters. Quantitative phenotypes are those in which an observable continuous scale is used to measure

the variable under consideration, e.g., enzyme levels, intelligence quotient (I.Q.) scores, and blood pressure.

A variable measured on a continuous scale has inherently greater information content than one measured as a dichotomous variable; therefore, quantitative traits that are highly correlated with liability to an illness can make important contributions to genetic analysis. Highly specific and sensitive biological measures of quantitative processes have not yet been found for mental disorders; rather, qualitative determinations (affected versus unaffected status) established through a structured diagnostic interview are the typical source of phenotypic data for genetic analysis.

GENETIC MODELS

Mathematical models are required in population genetics to represent the ways in which genes and the environment interact to form complex phenotypes transmitted within families (Table 1.17-1).

		Source of Familial Resemblance		
Genetic Model	Genes of Major Effect (N)	Genes of Minor Effect	Common Environment	Individual-Specific Environment
Single Major Locus	Yes (1)	No	No	Yes
Allelic Heterogeneity	Yes (1)	No	No	Yes
Locus Heterogeneity	Yes (>1)	No	No	Yes
Multifactorial	No	Yes	Yes	Yes
Mixed	Yes (1)	Yes	Yes	Yes
General multilocus	Yes (>1)	Yes	Yes	Yes

Table 1.17-1 Genetic Models of Disease Transmission

Single Major Locus Model The single major locus model assumes that all relevant genetic variation results from the presence of alleles at a single locus, and that environmental variation is unique to an individual. With two alleles, *A* and *a*, with respective frequencies *p* and *q*, three genotypes are possible: *AA, Aa*, and *aa*. When both alleles are the same, it is a *homozygous* genotype; when two alleles are different, it is a *heterozygous* genotype. The sum of the allele frequencies totals unity, so by definition, $p + q = 1$; thus, $1 - q = p$. If the environment is constant, such that each genotype corresponds to only one phenotype, the gene at a given locus is *completely penetrant*.

Diseases transmitted through a single major locus are referred to as *Mendelian diseases*, as the pattern of inheritance in families follows the rules of Mendelian segregation and can usually be recognized through visual inspection of pedigrees. Characteristic single locus diseases include retinitis pigmentosa, Duchenne dystrophy, Huntington's disease, phenylketonuria, and cystic fibrosis. The important discovery in 1991 of intraallelic expansion of highly unstable trinucleotide (triplet) repeat sequences helps to explain the variations in age of onset and severity without invoking an additional modifying locus. Huntington disease, fragile X syndrome, myotonic dystrophy, spinobulbar muscular atrophy, spinocerebellar ataxia type 1, and Machado-Joseph disease are all caused by the expansion of unstable repeat sequences in single major genes.

Familial patterns of simple Mendelizing inheritance can be characterized by whether the disease gene is on an autosome or on a sex chromosome and by whether both alleles are required for expression (recessive disease) or only one allele is sufficient (dominant disease). The liability distributions in the general population resulting from a diallelic major locus in Hardy-Weinberg equilibrium are shown in Figure 1.17-2.

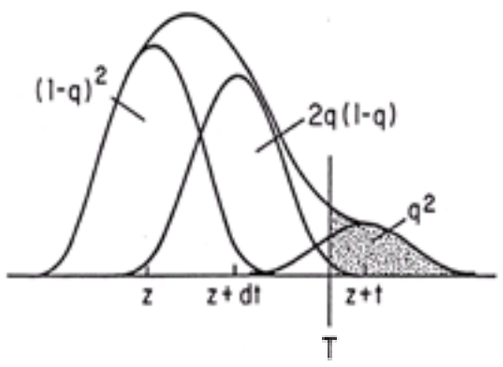

FIGURE 1.17-2 Liability distributions resulting from a single major locus in Hardy-Weinberg equilibrium. The locus has two alleles *A* and *a*, with frequency *p* and *q*. Given that $p + q = 1$, $p = (1 - q)$. The three genotypes *(AA, Aa, aa)* have respective means of z, z + dt, and z + t; d = 0 results in a recessive locus while d = 1 results in a dominant locus. The respective genotypic frequencies are $(1 - q)^2$, $2q(1 - q)$, and q^2. The shaded area gives the lifetime morbid risk of the disease (K_p). The proportion of persons with a given genotype who are above the threshold T gives the penetrance of that genotype.

Single Major Locus Modes of Inheritance The following criteria for different single locus models of disease transmission are required: (1) *Autosomal dominant*—(a) transmission continues from generation to generation without skipping; (b) except for freshly mutated cases (or nonpaternity), every affected child has an affected parent; (c) the two sexes are affected in equal numbers; and (d) in marriages of an affected heterozygote to a normal homozygote, the probability that a child born into that family will be affected (the segregation ratio) is [fr1,2]; (2) *Autosomal recessive*—(a) if the disease is rare, parents and relatives (except siblings) are usually normal; (b) all children of two affected parents are affected; (c) in marriages of two normal heterozygotes, the segregation ratio in offspring is ¼ and (d) the two sexes are affected in equal numbers; (3) *Sex-linked recessive*—(a) if the disease is rare, parents and relatives (except maternal uncles and other male relatives in the female line) are usually normal; (b) hemizygous affected men do not transmit the disease to children of either sex, but all their daughters are carriers; (c) heterozygous carrier women are normal but transmit the disease to their sons with a segregation ratio of ½, and half of the daughters are normal carriers; and (d) except for mutants, every affected male child comes from a carrier mother.

The concept of *incomplete penetrance* refers to persons with identical genotypes who have different phenotypes because of variability in nontransmissible environmental factors that contribute to the phenotype. The penetrance is the probability that a person with a given genotype will manifest the illness; probabilities are usually denoted by the letter *f*. The lifetime cumulative incidence or morbid risk of a disease is frequently denoted by the letters K_p. In the case of a disorder caused by a diallelic autosomal single major locus the respective gene frequencies of the *A* and *a* alleles are *p* and *q* and the respective penetrances associated with the *AA, Aa*, and *aa* genotypes are f_1, f_2, and $f_3 - K_p = f_1p_2 + f_2 2pq + f_3q_2$. Current generalized single major locus models allow for incomplete penetrance (i.e., one or more *f*s are not equal to 0 or 1), with transmission of a fully penetrant major locus contained as a submodel.

Elucidation of abnormal protein products and subsequent resolution of pathophysiology is theoretically more straightforward in the case of disease transmission through a single major gene than in the case of disease transmission through polygenes. However, the genetics of single major locus diseases can still be complicated, as exemplified by Huntington disease. Clinical characterization of this disease first occurred in the 1800s, a dominant disease locus was linked to genetic markers on chromosome 4 in 1983, and the precise gene was only identified in 1993.

Multilocus Models These models specify the effects of multiple loci, with or without contributions from environmental factors, on susceptibility to disease. It is useful to distinguish *familial (common) environmental effects* from *individual-specific (idiosyncratic) environmental effects*. The latter refer to environmental experiences

unique to the individual and not shared among family members; this is also called the *within-family environment*. The former refer to environmental influences that are common to, or shared by, family members; this is also called the *between-family environment*.

Genetic influences in multilocus models may arise from the effects of *genes of major effect* versus *genes of minor effect*. The former refer to genes that make a large relative contribution to the total variance in the disease attributable to genetic influences; the latter are genes that each make a small relative contribution to the total variance attributable to genetic factors. The individual unit for each—a gene—remains the same; the distinction between genes of major versus minor effect refers exclusively to the relative degree of influence that they have on the final behavioral outcome. Most common human diseases are inherited under a multilocus model; examples include hypertension, insulin-dependent diabetes mellitus, pyloric stenosis, rheumatoid arthritis, peptic ulcer, most cases of breast cancer, coronary artery disease, spina bifida, coronary artery disease, dementia of the Alzheimer's type late onset, with multiple sclerosis, and many mental disorders. Multilocus model variants may be distinguished in regard to the number of loci, if any, that exert a larger influence on the phenotype relative to the influence of other genes.

Multifactorial Model The multifactorial model assumes that all genetic variance is attributable to genes, called *polygenes*, that each exert a small relative effect. All relevant genetic and environmental contributions to variation can be combined into a normally distributed variable termed I>liability. There are one or more threshold values on the liability scale such that affected individuals are those with liability values that exceed the threshold. Familial inheritance is modeled through correlations in liability between family members, with the following assumptions: (1) relevant genes act additively and are each of small effect in relation to the total variation; (2) environmental contributions result from many events whose effects are additive; and (3) there may be multiple thresholds, such that individuals with scores between threshold values represent milder phenotypic or spectrum cases. When all transmissible effects are genetic (i.e., common environment exerts no influence), this is called the *polygenic model*. One or more threshold values are on the liability scale; affected persons are those with liability values that exceed the threshold. Familial inheritance is modeled through correlations in liability between family members. There may be multiple thresholds, such that persons with scores between threshold values represent mild phenotypic or spectrum cases. Normal traits inherited in this way include intelligence, stature, skin color, total dermal ridge count, and probably blood pressure. When the phenotype is qualitative (presence or absence of disease), a continuous liability distribution is unobservable but is assumed to underlie the discrete phenotypic events that are observed. Liability distributions in the general population for single-threshold and two-threshold multifactorial models are shown in Figure 1.17-3.

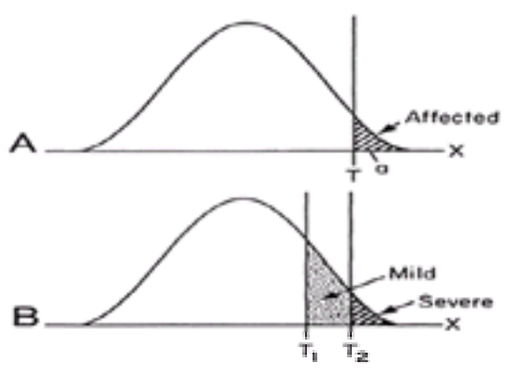

FIGURE 1.17-3 The unobserved liability *(x)* underlying a multifactorial disease with **(A)** a single threshold T and mean liability of affected persons denoted by *a*; **(B)** two thresholds T_1 and T_2, used to model severity, such that persons with a liability score above T_2 have a core (severe) phenotype, and persons whose liability is between T_1 and T_2 have a milder form of the illness (spectrum condition). The shaded areas give the lifetime cumulative incidences (K_p).

Mixed Model The mixed model is a marriage of the single major locus model and the multifactorial model. A distribution of liability is determined by the effects of a major locus, a multifactorial transmissible background (polygenes or environmental factors), and residual individual-specific environmental factors. The mixed model differs from the multifactorial model regarding the presence of a single genetic locus of major effect. Since both the single major locus model and the multifactorial model are submodels, the mixed model provides a statistical advance in permitting the rigorous testing of whether a single major locus or a multifactorial component (or both) contributes to familial resemblance.

General Multilocus Model The general multilocus model differs from the mixed model by the specification of more than one locus of relative major effect. The major assumption is that the marginal effects of these genes are etectable and separable from the background effects of other genetic loci of minor effect or environmental contributors. When multiple loci of small relative effect influence a phenotype measured on a continuous scale, they are commonly referred to as *quantitative trait loci*. Complex interactions among loci of major or minor effect *(epistasis)* may occur. Alternatively, each locus of relative major effect may be an independent and sufficient cause of disease; such multilocus models are termed *genetic heterogeneity models*—individuals can be affected if they possess a predisposing genotype at any one of the loci of relative major effect. Consequently, disease transmission in one set of families may be attributable to a locus that is different from the one predisposing to illness in another set of families.

The general multilocus model is the most comprehensive and realistic transmission model because common illnesses are likely to be influenced by major and minor genetic effects and by common environment. One way to quantify the relative effect of one locus versus others in multilocus models is to consider the proportion of disease risk that is attributable to that locus. This may be accomplished through consideration of the risks to different classes of relatives (i.e., siblings, monozygotic twins, cousins) of affected individuals conferred by that specific locus as compared to the population prevalence. Locus-specific risk ratios (frequently denoted as l) on the order of 1.5 to 2 may be expected for many mental disorders and other complex diseases. For example, analyses of recurrence risk patterns in first-, second-, and third-degree relatives and in monozygotic twins have suggested that multilocus models with gene effects of l = 2 are consistent with the familial transmission of schizophrenia and with cleft palate. Recently, variance-component methods for linkage analysis, which permit estimation of the magnitude of the effect of a detected locus within a multilocus system, have been developed.

RESEARCH DESIGNS

Population, family, twin, adoption, linkage, and association studies can each contribute evidence to evaluate the involvement of genetic factors in the cause of an illness (Table 1.17-2).

Study	Unit of Analysis	Goal
Population	Subjects in the general population	Establish lifetime incidence
Family	Pedigrees	Establish familiarity; estimate mode of transmission and risks to relative classes
Twin	Monozygotic and dizygotic twins	Distinguish genetic from environmental effects
Adoption	Adoptees; adoptive and biological relatives of adoptees	Distinguish genetic from environmental effects
Linkage	Nuclear or extended pedigrees or both	Establish chromosomal location of a disease susceptibility gene
Association	Unrelated affected individuals and controls	Identify a specific disease susceptibility gene
Transgenic	Gene expression and function in animals	Specify developmental outcomes/pathways

Table 1.17-2 Study Designs for Genetic Research on Mental Disorders

Population Studies Prevalence and incidence rates of mental and other disorders derived from community-based surveys have important scientific and health policy implications. Variations in such rates can provide clues to causes and can be used as base rates for comparative purposes in family genetic studies.

Family Studies If genetic factors are involved in illness transmission, the illness should occur in the families of affected members at a higher rate than in appropriate control populations. However, relatives who share a number of genes also tend to share common environments, so familial aggregation by itself does not necessarily implicate a genetic mechanism; culture, family environment, or infectious agents may be responsible. Family studies for mental and other disorders begin with affected persons *(probands)* selected from consecutive hospital inpatient admissions or a psychiatric case registry. Available relatives are located and assessed for psychopathology with structured or semistructured diagnostic instruments. Countries with national health insurance and psychiatric registers can provide morbidity information across generations. Recurrence risks are expected to increase as the degree of genetic overlap between relatives increases. The closest familial relationship is that of monozygotic twins, who share all their genes in common. Dizygotic twins, full siblings, parents, and children are first-degree relatives, who share half of their genes on average. Second-degree relatives of probands—grandparents, grandchildren, uncles, aunts, nieces, nephews, and half-siblings—share one fourth their genes.

A variety of factors tend to make comparisons of familial risk to mental disorders across studies difficult. Those factors include differences in sample characteristics, methods of age correction, ascertainment schemes, and diagnostic procedures. Such methodological concerns can explain how in different studies the risk for depression in first-degree relatives of depressive probands varies between 11 and 18 percent whereas the risk to relatives of normal controls varies between 0.7 and 7.3 percent. Comparison of normal controls and high-risk relatives by similar case-finding and diagnostic methods are essential when interpreting mean risk estimates. Similarly, the ideal family study uses double-blind, case-controlled methods in which diagnoses of relatives are made independently of the proband's diagnosis.

Age Correction Family studies permit determination of morbid risk estimates in different relative classes. A simple tally of the frequency of a disorder in relatives will underestimate the true morbid risks, because not all unaffected individuals have passed through the period of risk for mental disorders at the time of examination. Quantitative methods have been developed that permit estimation of morbid risks with suitable *age correction*, that is, morbid risk estimation that takes into account the fact that some of the unaffected individuals now observed as unaffected will develop illness at a later point in time. Wilhelm Weinberg's short method of age correction was the first devised early in the twentieth century; this simple procedure assigns weights to the number of unaffected individuals in different age groups. Weinberg's method was followed by one developed by Eric Strömgrer, and independently by Eliot Slater, which uses the ages at onset in the proband samples to obtain an age-at-onset distribution. All unaffected relatives are weighted by the proportion of the risk period through which they have passed, and the lifetime morbid risk is then computed as the number of affected individuals divided by the number at risk (the number affected plus the weighted sum of unaffected relatives).

Survival curve estimation methods are now applied to determine age-corrected lifetime morbid risks in relatives. *Survival analysis* is a mathematical technique that models time to an event (e.g., illness onset) while paying special attention to incomplete (censored) data in which the event is not observed for all individuals. Covariates that influence the time to the event may be modeled in the Cox proportional hazards model. The nonparametric Kaplan-Meier estimate of time to onset of illness is typically used to estimate lifetime morbid risk; only onsets in relatives (and not in the probands) are considered.

Twin Studies The twin method has been a popular research design to implicate or exclude genetic factors in the cause of a disease. Since monozygotic twins have identical genotypes, any dissimilarity between pair members must result from the action of the environment, either prenatally or postnatally. Consequently, anything less than 100 percent concordance among monozygotic pairs living through the period of risk excludes genetic factors as sufficient determinants of that disease.

If genetic differences are not important for the familial clustering of a disease, no differences should be seen in the monozygotic and dizygotic concordance rates. This is what occurs in twin studies of diseases caused by infectious agents (e.g., measles). Conversely, if genes are important in causing a disease, the monozygotic concordance rate is significantly higher than the dizygotic rate. A genetic basis is the most likely explanation for the higher monozygotic concordance rate if monozygotic twins are not more predisposed to having the disease and if monozygotic twin environments are not more alike in features that cause the disease. The twin method in psychiatry has also been useful in identifying spectrum conditions that are alternative manifestations of the disease genotype that occur in monozygotic twins discordant for the core illness. A variant of the twin design is to include the offspring of concordant versus discordant twins.

Critics of the twin method have argued that monozygotic pairs share more similar environments than do dizygotic pairs, and that is responsible for the higher monozygotic concordance rate for mental disorders. Three ways in which environmental factors may increase the rate have been advanced: (1) monozygosity per se, (2) the effects of identification by one twin with another, and (3) the sharing of a similar ecology, with enhanced exposure to triggering events. No conclusive evidence exists that those limitations have substantially or consistently biased the results of twin studies of mental disorders.

Adoption Studies Whereas monozygotic and dizygotic twin studies endeavor to hold the family environment constant to compare the resemblances between persons with the same and different genotypes, adoption studies permit the comparison of the effects of different types of rearing on groups who are assumed to be similar in their genetic predispositions. Such studies attempt to separate the effects of genes and the familial environment by capitalizing on the adoption process, in which children receive their environment from a source different from their gene source. Consequently, adoption study designs permit the disentangling of genetic and environmental factors that contribute to the familial aggregation of a disease. The ability to draw inferences from an adoption study is strongest when the adopted children are separated from their biological parents at birth.

Potential problems of the research design are that any parent-child interaction from the time of birth to the separation confounds a clear demarcation of genetic and environmental aspects and the environmental circumstances of biological parents may be associated with prenatal and perinatal events relevant to the cause of the disease.

Three major designs of adoption studies have been used to study mental disorders: (1) the parent-as-proband design compares the rate of illness in the adopted offspring of ill and well persons. Support for a genetic component is indicated if the risk of illness among adopted children of ill parents is greater than the risk of illness among adopted children of well parents; (2) the adoptee-as-proband design uses ill and well adoptees as probands. Genetic factors are implicated if (a) the risk of illness in the biological relatives of ill probands is greater than that in the adoptive relatives of ill probands and (b) the risk of illness is greater in the biological relatives of ill probands than that in the biological parents of well adoptees; and (3) the seldom used cross-fostering approach, which compares rates of illness in two groups of adoptees. One group has ill biological parents and are raised by well adoptive parents; the other group has well biological parents and are raised by ill adoptive parents.

A famous adoption study in psychiatry was started in the 1960s by David Rosenthal, Seymour Kety, Paul Wender, and their colleagues to study schizophrenia in Denmark. Major accomplishments of the project were to rule out some alleged environmental factors (e.g., being reared by a schizophrenic parent) as either necessary or sufficient for the development of schizophrenia in the offspring of parents with schizophrenia, and to confirm the validity of family and twin results in implicating genes. The data have held up remarkably well, even after probands and relatives were classified with criteria used in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III). The data also provided an opportunity to develop operational criteria for schizotypal personality disorder as a spectrum condition genetically related to the core schizophrenic phenotype, since it occurred at a higher rate in the biological relatives of adopted-away schizophrenia patients than in the adoptive relatives of schizophrenia patients and the relatives of control adoptees.

Association Studies The standard method for mapping disease loci applies classic parametric or nonparametric methods to family data in a search for linkage between the disease and a marker locus (a gene or deoxyribonucleic acid [DNA] sequence of known location). An alternative approach is to look for statistical associations in the general population between the disease and a marker. Linkage analysis of family data implicates a chromosomal region by assuming a relationship between a disease locus and marker loci in that region; association analysis implicates a specific gene by assuming a relationship between a disease and alleles at a specific genetic locus.

Population associations can generally arise for three reasons: (1) the implicated locus is itself a disease susceptibility locus—possession of the particular allele associated with the disease is neither necessary not sufficient, but it increases the likelihood of becoming ill; (2) a disease locus and the associated marker locus are tightly linked, that is, physically close to each other. Nonrandom association of alleles at tightly linked loci is commonly called *linkage disequilibrium*. Linkage disequilibrium can persist for many generations as a function of the physical distance between the disease locus and the marker locus; or (3) people with the disease and those without may be genetically different subsets of the population that coincidently differ in allele frequencies *(population stratification)*—in this case, the implicated locus is likely to be unrelated to the disease.

Classic disease-marker studies have been conducted by studying a sample of unrelated affected persons and comparing the frequency of a particular marker allele in the affected group versus its frequency in a control sample. This is a *population-based case control study* of disease-marker association. Very strong associations have been found between the human lymphocyte antigen (HLA) system on chromosome 6 and a few diseases (e.g., ankylosing spondylitis); weaker associations

have more typically been found for complex diseases like insulin-dependent diabetes mellitus and multiple sclerosis. However, causal inferences based on genetic differences between cases and controls drawn from a heterogeneous population have been difficult to replicate or interpret in the study of mental disorders because: (1) problems with selections of controls lead to difficulties in distinguishing linkage disequilibrium from population stratification; and (2) inadequate statistical correction for the testing of association at many loci leads to an increased type I error rate and chance findings (i.e., falsely concluding that a disease-marker association exists when there truly is none).

Selecting suitable controls for population-based association studies is crucial to minimize the chances that the study and control groups are drawn from genetically distinct subpopulations. Methods have been recently applied to circumvent this problem, and the focus is on the use of association tests as a means of locating disease genes through detection of linkage disequilibrium. *Family-based association studies* use disease and marker data within families. Comparisons are made by comparing genetic material from a sample of patients with genetic material obtained from their two parents. A single affected individual and the individual's two parents identified for family-based association studies are frequently referred to as a *trio*. Given that data from parents may be difficult or impossible to obtain, methods for family-based association analysis have been extended to permit use of data from unaffected siblings.

QUANTITATIVE METHODS

Data from many of these research designs can be analyzed by using sophisticated mathematical models and high-speed computers. The methods most typically used in the study of genetic factors are presented in Table 1.17-3.

Method	Data Source	Goal
Path analysis	Twin, adoption	Distinguish transmissible environment from polygenes
Segregation analysis	Pedigree	Distinguish a major locus from polygenes or transmissible environment
Linkage analysis	Pedigree	Establish chromosomal localization of a putative disease susceptibility locus
Association analysis	Unrelated affecteds, controls	Implicate a specific gene as a disease susceptibility locus, given linkage disequilibrium

Table 1.17-3 Quantitative Methods of Genetic Analysis

Path Analysis Path analysis was introduced as a technique to explain the interrelations among variables by analyzing their correlational structure and to evaluate the relative importance of varying causes influencing a certain variable. The primary goal of path analysis in genetic epidemiology is to distinguish genetic effects from common environmental effects that contribute to the familial aggregation of a disease. Twin and adoptive data are necessary to separate nature from nurture in path analysis. When genetic transmission is present, additive genetic effects cannot be distinguished from other genetic effects (i.e., single major locus models and multilocus models cannot be distinguished).

Familial correlations are estimated through *maximum likelihood techniques*, statistical procedures for estimating parameters, such that the best-fitting estimates are those that maximize the probability of the observations. Comparisons of competing models are made by fitting a general model and alternative submodels. Since log likelihoods are calculated for the general model (L_1) and the submodel (L_2), $-2(L_1-L_2)$ is approximately distributed as a chi-square with the degrees of freedom equal to the difference in the number of estimated parameters. This is the *likelihood ratio test*, the test statistic for comparing alternate models. Both qualitative (affected or unaffected status) and quantitative phenotypes may be analyzed, and examples of the results of applying path models of multifactorial transmission to analyze several traits are given in Table 1.17-4.

Trait	Genes	Common or Family Environment	Individual-Specific or Unique Environment
Intelligence	0.52	0.34	0.14
Personality (extroversion)	0.60	0.00	0.40
Religious devotion	0.29	0.24	0.47
Bipolar I disorder	0.86	0.07	0.07
Major depressive disorder	0.45	0.00	0.55
Neurotic depression	0.08	0.54	0.38
Schizophrenia	0.63	0.29	0.08
Bulimia nervosa	0.04	0.24	0.72
Alcoholism (women)	0.50	0.00	0.50
Late-onset Alzheimer's disease	0.58	0.39	0.03
Tuberculosis	0.06	0.62	0.32

Table 1.17-4 Genetic and Environmental Contributions to the Variance in Liability of Several Traits, Assuming Etiological Homogeneity

A useful application of complex path models was exemplified by the analysis of twin and family data from a variety of published sources for tuberculosis and schizophrenia. The results showed that the major contribution to phenotypic variance for tuberculosis came from shared family environment, rather than from genes; that result is expected for an illness caused by an infectious agent. Results from twin data alone would have been misleading in implicating a significant genetic effect. By contrast, the largest contribution to the variance in schizophrenia and bipolar I disorder comes from genes, with suggestion of a modest role for the common environment.

Complex Segregation Analysis Complex segregation analysis is a method for resolving a single major locus effect, but it leaves cultural inheritance confounded with additive polygenes. The unit of analysis is an entire pedigree, and the goal is to statistically assess evidence for the segregation of a major gene in the presence of other sources of familial resemblance. The application of complex segregation analysis to psychiatric family data has led to disappointing results, in the sense that no single gene effect has been consistently identified for any mental disorder.

Linkage Analysis Complex segregation analysis is an analytic tool for identifying a major locus, but phenotypic segregation patterns alone do not provide opportunities for its localization and ultimate identification. Linkage analysis is a statistical procedure by which pedigree data are examined to determine whether a disease phenotype is co-segregating with a genetic marker of known chromosomal location. Linkage analysis allows an investigator to infer that two loci (a genetic marker locus and a putative disease susceptibility locus) are located close enough together on the same chromosome that their alleles tend to be transmitted together from parent to child more frequently than would occur by random assortment. The demonstration of linkage between a putative disease susceptibility locus and one or more genetic markers thus determines on which chromosomal region the disease locus lies.

Genetic markers are entities known to follow a simple Mendelian mode of inheritance with an identified chromosomal location. At least one parent must be doubly heterozygous for that mating to be informative for linkage. Therefore, the usefulness of a genetic marker for linkage depends on the number of alleles and the gene frequencies (its degree of *polymorphism*) in the sense that an increased polymorphism leads to an increased probability of heterozygosity.

In the late 1970s the number of available polymorphic genetic markers was severely limited to blood cell antigen loci (ABO, Rh, and HLA blood groups) now known to lie on chromosome 1, 6, or 9. Some of those markers were highly polymorphic, but their limited number and restricted coverage of the genome meant that even linkage studies with excellent family data had little prospect for success. However, in the late 1970s and early 1980s geneticists proposed to treat differences in the DNA sequence as allelic variants and to use them as genetic markers. Through molecular genetic techniques *restriction fragment length polymorphisms* (RFLPs) were

obtained and were well suited as genetic markers in linkage analysis. RFLP markers are highly polymorphic and are available in great numbers that saturate the genome. A variety of other types of genetic markers (e.g., microsatellite markers, variable number tandem repeats) have since been developed. The most common type of human genetic variation is the *single-nucleotide polymorphism*. There has been growing recognition that large collections of mapped single-nucleotide polymorphism would provide powerful information for genetic studies. Single-nucleotide polymorphism can serve as genetic markers for establishing linkage.

A common measure of the degree of marker polymorphism often used is the *polymorphic information content* (PIC) value, which is defined as the probability that the marker genotype of a given offspring will allow deduction of whichever one of the two marker alleles of the affected parent that were received. If a PIC value approaches 1, virtually all matings are informative for linkage; a PIC value of ½ means that 50 percent of matings are informative. Another common measure of the degree of polymorphism is called *heterozygosity*, which is the probability that a random individual is heterozygous for any two alleles at a given locus. A genetic marker's usefulness for linkage analysis is reflected in its PIC and heterozygosity values; the higher the values, the greater its utility.

A classic approach in human genetics has been to search for single locus effects underlying disease transmission once a Mendelian mode of transmission has been identified and then to apply linkage analysis. Chromosomal localization through linkage analysis is the first essential step in the process of identifying, isolating, and cloning a disease susceptibility locus.

Three analytical strategies (Table 1.17-5) are used to search for linkage to mental disorders and locate disease susceptibility genes: (1) apply parametric maximum likelihood methods to calculate LOD scores in small or extended pedigrees; (2) apply nonparametric methods to study allele sharing among affected sibling pairs; or (3) apply the transmission/disequilibrium test (TDT) or other family-based association tests to detect linkage given linkage disequilibrium.

Method	Unit of Analysis	Mode of Inheritance	Multipoint Analysis	Computational Load	Typical Statistic	Target
Parametric	Pedigree	Required	Yes (<14)	Intensive	LOD score	Chromosomal region
Allele-sharing	Affected sib pairs or pedigrees or both	Not required	Yes	Not intensive	IBD-sharing probability; nonparametric linkage statistic	Chromosomal region
Linkage disequilibrium	Unrelated affecteds, controls	Not required	No	Not intensive	χ^2	Individual gene

IBD, identity-by-descent. Multipoint analysis refers to the simultaneous use of information provided by more than one genetic marker.

Table 1.17-5 Quantitative Methods Used to Locate Disease Susceptibility Genes

Logarithm of Odds (LOD) Score Methods The alleles at different loci received by a person from one parent are called a *haplotype*. In principle, a doubly heterozygous person—for example, *Aa, Bb*—received the *A* allele with either the *B* or *b* allele from one parent. If two loci (with alleles *A* or *a* at one locus and *B* or *b* at the other) are inherited independently of each other, a father passes the four haplotypes *AB, ab, Ab*, and *aB* to his offspring in the ratio of 1:1:1:1. If the *AB* and *ab* haplotypes look the same as the ones he received from his parents, the children receiving them are called *parental types*. The other two haplotypes (*Ab* and *aB*) in this case are unlike any haplotypes received by the father from the grandparents of the child and contain one allele from each grandparent (a recombination of grandparental alleles must have occurred in the father). The nonparental types (*Ab* and *aB*) are called *recombinants*, and the other two haplotypes are called *nonrecombinants*. A recombination between two genes denotes the event that two different grandparents contribute one allele at each of the two loci to a haplotype in a person, whereas a nonrecombination is said to have occurred when a haplotype in a person contains two alleles that originated from the same grandparent of that person. A mating is potentially informative for linkage between two specific genes when at least one of the parents is a double heterozygote.

When two genes are inherited independently of each other, recombinants and nonrecombinants are expected in equal proportions among the offspring. Some pairs of genes consistently deviate from the 1:1 ratio of recombinant to nonrecombinant offspring; in other words, alleles of different genes appear to be genetically coupled. That is called *genetic linkage*. The extent of genetic linkage is measured by the *recombination fraction*, which is the probability that a gamete produced by a parent is a recombinant. The recombination fraction is frequently denoted by the Greek letter theta (q). Genes segregating independently are unlinked with q = ½, whereas linked genes are characterized by q < ½. Some pairs of genes are tightly linked, so that q approaches 0—that is, only rarely does a recombination occur between them. The estimation of q and the test of the hypothesis of free recombination (q = ½) versus linkage (q < ½) are the goals of linkage analysis.

Recombination fractions reflect genetic distance on a chromosome, which is not exactly the same as physical distance. Genetic distance refers to a mathematical estimate of distance between markers and reflects the number of recombinant events (values of q greater than ½ are not meaningful); physical distance reflect the actual number of base pairs on the chromosome. Two loci that show 1 percent recombination are defined as being 1 centimorgan (cM) apart on a genetic map. (100 cM define a Morgan, which was named in honor of Thomas Hunt Morgan; these are the units that measure genetic distance). However, for distances above 5 cM, genetic distance is not a simple reflection of the number of recombinant events. A mathematical equation called the *mapping function* defines the relationship between the recombination fraction and genetic distance.

A nonconstant relationship exists between genetic distance, as measured in cM, and physical distance, which is measured in DNA base pairs or megabases (Mb; 1 Mb = 1,000,000 base pairs). The entire human genome is 3000 Mb, or three billion base pairs. A sex-averaged figure that relates physical and genetic distance is 1 cM = 0.9 Mb, but the actual correspondence varies widely for different chromosomal regions.

Figure 1.17-4 shows in a simplified manner the chromosomal interpretation of recombination. In *meiosis* (cell division leading to the formation of gametes) homologous chromosomes pair up. At that point each homologous chromosome consists of two strands (*chromatids*), so that a chromosome pair consists of four. In the course of meiosis, the two homologous chromosomes separate from each other at most places but maintain one (1) or two (2) zones of contact *(chiasmata)*. Chiasmata reflect the occurrence of *crossing over* between chromatids. A single crossover generates two recombinant and two nonrecombinant chromatids (1) whereas a two-strand double crossover leaves four nonrecombinant chromatids (2). The overall effect averaged over all double crossovers is to generate 50 percent recombinants.

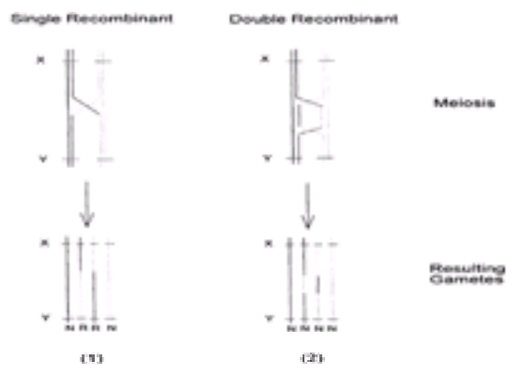

FIGURE 1.17-4 Schematic representation of a pair of homologous chromosomes, each consisting of two strands. Single (1) and two-strand double (2) crossovers involve two of the four chromatids; the solid line chromosome carries alleles X_1 and Y_1, at two loci, while the dotted one carries alleles X_2 and Y_2. Gametes (sperm, ovum) in which the chromatid is the same line type at the two loci are nonrecombinant (*N*) for these loci; those where the chromatids are different line types are recombinant (*R*).

Since recombination events can be recognized only on the basis of haplotypes passed from parents to children, linkage analysis requires phenotypic observations on

pedigree members. Estimating q is carried out by using the method of maximum likelihood. A relevant quantity is the likelihood ratio that is obtained by dividing the likelihood of a given family [L(q)] by its value under free recombination (L(½)). It is usually convenient to work not with the likelihood ratio but with its logarithm to the base 10. This is the *LOD score Z (q)*, such that $Z(q) = \log_{10}[L(q)/L(½)]$. The LOD score serves as a measure of the weight of the data in favor of the hypothesis of linkage. The critical value generally adhered to as the criterion for significant evidence for linkage to simple, monogenic diseases with unambiguous phenotypes and established modes of transmission is 3 for autosomal loci. The one-sided type I error rate associated with that value when q is the only estimated parameter (i.e., the mode of disease transmission is known) and when the linkage test is conducted at a single marker is 0.0001 in large samples.

The traditional advantages of LOD score methods for linkage analysis include the following: (1) because it is a parametric approach, LOD score methods have high power to detect a true linkage given knowledge of the true mode of disease inheritance; (2) if affected sib pairs are rare and etiological heterogeneity is likely, multigenerational pedigrees with multiple affected relatives of various classes (uncles, grandparents, cousins, and so forth) can be analyzed; (3) linkage to a particular chromosomal region can be excluded; and (4) a measure of the distance between two loci—the recombination fraction q—can be estimated.

A consideration when applying LOD score methods for linkage analysis is that the mode of inheritance is assumed to be known. When single major locus inheritance parameters (gene frequencies and penetrances) are estimated jointly with q in linkage analysis, the LOD score value does not have the same statistical meaning. One conservative correction for maximizing a LOD score over *t* different transmission models proposed by Neil Risch is to subtract $\log_{10}(t)$ from the result, (e.g., a LOD score of 3 maximized over 5 transmission models is reduced to about 2.3).

Allele-Sharing Methods Allele-sharing methods assume that a disease susceptibility locus can be identified given that a pair of affected relatives—typically a sibling (sib) pair, but other relative pairs may also be considered—will tend to inherit the same allele more often than expected under random Mendelian assortment. Each pair shares either 0, 1, or 2 alleles *identical by descent* (IBD) at a given locus, and the *allele-sharing proportion* is defined as the proportion of affected relative pairs that share a single allele IBD at that locus. Genotyping parents and other individuals in a pedigree allow determination of whether given alleles are actually inherited from a common ancestor, that is, IBD status can be deduced. In practice, the situation is more complicated because one cannot unambiguously determine the number of alleles shared IBD at all of the loci in the genome. In such cases, inferences are drawn based only on identity-by-state (IBS) information—that is, it is determined whether individuals show the same allele at a given locus, regardless of whether the allele came from a common ancestor. IBD methods have greater power to detect linkage and are less sensitive to misspecification of marker allele frequencies (which can lead to false-positive results) as compared to IBS methods.

Tracing the inheritance pattern in the affected sib pair method uses perturbations in the distribution of IBD scores at a marker locus to detect the presence of a linked disease locus. In the absence of linkage, the probability that two siblings share neither, one, or both marker haplotypes identical by descent is independent of their disease phenotypes. Consequently, if pairs of siblings are studied because they are both affected, the distribution of their IBD scores will be ¼, ½, and ¼ for IBD = 2, 1, and 0, respectively. The allele-sharing proportion z_{all} for sibs is $z_{all} = (z_1 + 2z_2)/2$, where z_1 and z_2 are the proportions of sib pairs sharing 1 and 2 alleles IBD; thus, $z_{all} = ½$ for sib pairs in the absence of linkage, i.e., $(½ + ¼\cdot2)/2$. When a disease locus is linked to a marker locus and pairs of affected sibs are studied, there is a perturbation in the IBD score distribution at that marker locus. New variance-component linkage analysis methods that utilize multipoint IBD information in pedigrees of arbitrary size have recently been developed.

The specific advantages of affected sib pair methods in the study of mental disorders are the following: (1) these semiparametric or fully nonparametric methods specification of complex, non-Mendelian modes of transmission is not required. Concomitantly, several confounding factors that make it difficult to accurately estimate the mode of transmission in segregation analysis (e.g., complex ascertainment strategies, cohort effects, sex effects, variable age of onset) do not have to be modeled; (2) the practice of testing for linkage under several transmission models—which necessitates some downward correction to the linkage statistic to prevent inflation in the type I error rate for testing across multiple disease transmission models—is now unnecessary; (3) large multigenerational families with many affected persons, which are typically difficult to locate and study, are not required; and (4) complete extraction of multipoint inheritance information and estimation of disease gene location is now possible using new nonparametric methods.

Linkage Disequilibrium Mapping Parametric and nonparametric linkage methods are very useful in the search for disease susceptibility genes. However, these methods may require very large sample sizes to detect the modest gene effects that are most likely to be operative in most mental disorders (recurrence risk ratios on the order of 1:6; Table 1.17-6). In addition, isolation of a specific disease gene knowing only its subchromosomal location *(positional cloning)* depends on high-resolution mapping of the chromosomal regions detected through linkage analysis. With current technologies, in the absence of fortuitous events like the presence of a trinucleotide repeat expansion or gross deletion, a gene most likely must be localized to a region of about 1 million base pairs before it is practical to identify it. Given the magnitude of the genetic effects found in mental disorders, localization to regions of this size may require 800 to 1600 affected sib pairs.

	Recurrence Risk Ratio*						
N	1.0	1.2	1.4	1.6	1.8	2.0	3.0
100	0	0	2	5	9	15	45
200	0	1	6	19	36	51	90
400	0	3	23	59	82	94	100
800	0	12	69	97	100	100	100

* Morbid risk conferred by a specific susceptibility locus to first-degree relatives of an affected individual, divided by the lifetime cumulative incidence of the disease. Figures displayed are percentages. (Adapted from Hauser ER, Boehnke M, Guo S-W, Risch N: Affected-sib-pair interval mapping and exclusion for complex genetic traits: Sampling considerations. Genet Epidemiol 13:117, 1996.)

Table 1.17-6 Power to Detect Linkage to a Disease Susceptibility Gene for *N* Affected Sib Pairs

The initial detection and positional cloning of genes of modest effect may be facilitated through application of family-based association tests. Linkage disequilibrium mapping using family-based association tests can also be accomplished using unrelated affected individuals and their parents. Such methods can have tremendous power in genetically isolated populations, where a significant fraction of affected individuals have inherited the disease susceptibility allele from a common ancestor. Such a strategy has been successfully applied to a simple monogenic disease, diastrophic dysplasia.

The *haplotype relative risk* (HRR) method has been proposed as a family-based test of association. The control sample is the alleles at different loci received from one parent (the parental haplotype) not present in the affected person, which represents a random sample of haplotype pairs from the same genetic population. These researchers did not focus on the use of the HRR method as a test for linkage. Richard Spielman and colleagues developed a related method—the TDT—as a test for linkage between a complex disease and a marker given an established disease-marker association (linkage disequilibrium).

The TDT employs the alleles not transmitted by parents to an affected offspring as the controls. Thus, DNA needs to be collected from unrelated affected subjects and their two biological parents. Table 1.17-7 shows the simple 2 × 2 contingency table that can be constructed given a marker with two alleles A^1 and A^2; the significance test employs the simple X_2 statistic. The TDT has been generalized to the case of an arbitrary number of marker alleles.

Transmitted Allele	Nontransmitted Allele		
	A_1	A_2	Total
a_1	a	b	a+b
A_2	c	d	c+d
Total	a+c	b+d	2n

* Combinations of transmitted and nontransmitted marker alleles A_1 and A_2 among 2n biological parents of n affected individuals. The notation is as follows: a, the number of times that an A_1A_1 parent transmits A_1 to affected offspring; b, the number of times that an A^1A_2 parent transmits A_1 to affected offspring; c, the number of times that an A_1A_2 parent transmits A_2 to affected offspring; and d, the number of times that an A^2A_2 parent transmits A_2 to affected offspring. For the hypothesis of no linkage (and no allelic association), χ^2 (one degree of freedom) = $(b - c)^2 / (b + c)$.

Table 1.17-7 Transmission/Disequilibrium Test: Detecting Linkage Given a Population Association*

Use of the TDT does not solve the problem of inadequate statistical correction given conduction of a large number of association tests at many loci. A conservative approach is to divide the desired type I error probability by the number of tests conducted. For example, maintenance of a 5 percent false-positive rate (significance level = 0.05) when 50 independent tests have been conducted would require a significance level of 0.001 for each test. Genomic screens have been recommended based on association studies using the TDT and the intriguing possibility of eventually testing for disease associations at all of the human genes that will eventually be identified by the beginning of the twenty-first century has been discussed. An efficient design for psychiatric genetic studies may be to employ a multi-stage strategy for genotyping and analysis of nuclear or small extended pedigrees that incorporates linkage and family-based association analysis.

Statistical Criteria for Declaring Linkage to Mental Disorders It is crucial in the genetic investigation of mental disorders that a sufficiently stringent standard be adopted for the declaration of linkage, in order to maintain a high likelihood that the assertion will be true and stand the test of time. The LOD score criterion for declaring a linkage is 3 in the study of classical Mendelian diseases with known modes of familial transmission.

Ever-evolving genetic methods and technologies now permit systematic screening of the entire human genome. The increased number of markers being tested inflate the type I error rate. A set of guidelines have been proposed for interpreting linkage results of complex disease. These guidelines distinguish the nominal significance level, which is the probability of encountering a linkage statistic of a given magnitude at one specific locus, from the genome-wide significance level, which is the probability that one would encounter such a deviation somewhere in a whole genome scan. A given linkage statistic like a LOD score has a corresponding nominal *P* value and a genome-wide *P* value.

It has further been proposed that genome-wide *P* values be interpreted to evaluate the magnitude of linkage evidence and classify it as “suggestive,” “significant,” or “highly significant.” Suggestive linkage reports will often reflect chance findings and will often be wrong, but are worth reporting as tentative findings. Table 1.17-8 shows equivalent LOD score values and associated nominal and genome-wide *P* values for these different categories. A more stringent LOD score criterion (3.3 for LOD score methods; 3.6 for allele-sharing methods) than the traditional value of 3 is required to claim significant linkage evidence in the analysis of mental disorders and other complex diseases. Confirmation of linkage requires a two-step process: (1) significant linkage is found in at least one study; and (2) evidence of linkage to the same region is obtained by an independent group of investigators in an independent sample.

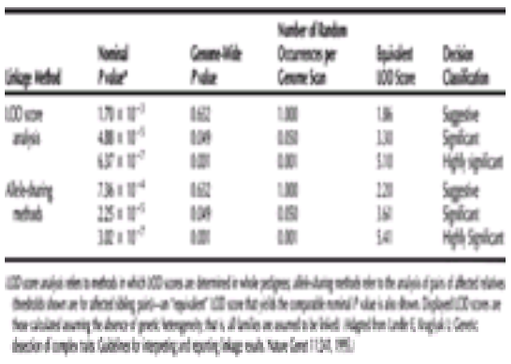

Linkage Method	Nominal P value*	Genome-Wide P value	Number of Random Occurrences per Genome Scan	Equivalent LOD Score	Decision Classification
LOD score analysis	1.70×10^{-2}	0.632	1.000	1.86	Suggestive
	4.88×10^{-5}	0.049	0.050	3.30	Significant
	6.37×10^{-7}	0.001	0.001	5.10	Highly significant
Allele-sharing methods	7.26×10^{-4}	0.632	1.000	2.20	Suggestive
	2.25×10^{-5}	0.049	0.050	3.61	Significant
	3.02×10^{-7}	0.001	0.001	5.41	Highly Significant

LOD score analysis refers to methods in which LOD scores are determined in whole pedigrees, allele-sharing methods refer to the analysis of pairs of affected relatives [illegible] affected sibling pairs—an "equivalent" LOD score that yields the comparable nominal *P* value is also shown. Displayed LOD scores are those calculated assuming the absence of genetic heterogeneity; that is, all families are assumed to be linked. Adapted from Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genet 11:241, 1995.)

Table 1.17-8 Criteria for Evaluating Reports of Linkage to Mental Disorders

MULTIVARIATE ANALYSIS INCORPORATING QUANTITATIVE TRAITS

Traditional approaches in genetic epidemiology focus on the qualitative determination of disease status as the exclusive source of data for genetic analysis. Such approaches are problematic in the case of mental disorders, in which phenotypic assessment through structured and semistructured interviews is potentially complicated by diagnostic error on misclassification. Quantitative traits that cosegregate with the disease phenotype provide greater information content than do groupings of individuals into affected or unaffected classes. The usefulness of pedigrees is increased because a greater range of information is available on unaffected persons who are not yet through the risk period. Thus, the power to initially map a susceptibility locus of small relative effect may be enhanced through consideration of the effects of such loci on quantitative traits correlated with disease. The application of such multivariate methods in the genetic analysis of mental disorders has yet to reach its full potential. Examples of such traits include psychophysiological abnormalities in alcoholism and attentional impairments in schizophrenia.

FUTURE DIRECTIONS

Table 1.17-9 shows a variety of relevant resources on the World Wide Web. As progress in the construction and completion of detailed genetic and physical maps continues, the ability to detect disease susceptibility genes for mental disorders will be enhanced. Complicating factors like multiple genes of small relative effect, common and individual-specific environmental effects, diagnostic error on misclassification, clinical heterogeneity, and etiological heterogeneity present challenges that can may be overcome in the following ways:

Electronic Address	Description
http://genetics.nature.com	Nature Genetics
http://www.faseb.org/genetics/ashg/pubs.htm	American Journal of Human Genetics
http://www.cup.org/journals/jhl/LSCAT/hge/hge.html	Annals of Human Genetics
http://www.faseb.org/genetics/ashg/ashgmenu.htm	American Society of Human Genetics
http://www.bga.org	Behavior Genetics Association
http://darwin.cwru.edu/ige/s.html	International Genetic Epidemiology Society
http://www.sph.umich.edu/group/statgen/	Statistical genetics at the University of Michigan
http://www.sfbr.org/online/departments/genetics/population.html	SFBR Population Genetics Laboratory
http://www2.ncsu.edu/ncsu/CIL/stat_genetics	North Carolina State University Statistical Genetics Program
http://lotka.stanford.edu	Stanford University Human Population Genetics Laboratory
http://www.mc.duke.edu/depts/genetics/course	Genetic analysis methods course at Duke University
http://linkage.rockefeller.edu/soft/list.html	Linkage analysis software
http://nitro.biosci.arizona.edu/ebook/book.html	*Fundamentals of Quantitative Genetics*
http://www3.ncbi.nlm.nih.gov/omim	Online Mendelian Inheritance in Man
http://www.nih.nimh.nih.gov/gg/gg.html	Genetics Research Branch NIMH

Table 1.17-9 Scientific Resources Related to Quantitative Genetics on the World Wide Web

1. Collection of a sufficiently large sample of pedigrees is required to obtain sufficient power to detect linkage and map loci to small chromosomal regions. Large-scale family-based association tests across the whole genome should be conducted in the sample;
2. Independent confirmation by at least one independent group of investigators remains the standard to establish the validity of an initial linkage report. The linkage evidence in at least one report needs to be significant, and must remain after sample augmentation (i.e., more families, or more subjects per family, are analyzed);
3. Resolution of genetic heterogeneity requires further work on phenotypic classification in order to identify clinical characteristics that can delineate genetically distinct subgroups;
4. Strict blindness on a project must be maintained as clinicians gather pedigree and diagnostic data and laboratory personnel establish genotypes. This necessitates close collaborative relationships between mental health clinicians, human biologists, genetic epidemiologists, and molecular geneticists in future

large-scale studies of mental disorders;
5. Diagnostic models and hierarchies, as well as a justification for classifying specific genotypes as “affected,” must be clearly specified prior to statistical genetic analyses;
6. Diagnostic and genotyping data should be made widely available to the scientific community so as to facilitate efforts to understand discrepancies between studies;
7. State-of-the-art techniques for molecular and quantitative analysis must continue to be rapidly applied to the genetic analysis of mental disorders. These include new nonparametric methods for multipoint linkage analysis, advances in high-throughput genotyping, the development and application of microchip DNA arrays, and the generation of abundant new genetic markers that reflect common human genetic variation.

SUGGESTED CROSS-REFERENCES

Epidemiological principles and methods are discussed in Section 5.1; mathematical concepts useful in understanding the fundamental principles of population genetics can be found in Section 5.2; and findings related to the epidemiology of schizophrenia and mood disorders are presented in Section 12.2 and Section 14.2, respectively. Section 12.5 and Section 14.3 present the findings from the study of the genetics of schizophrenia and mood disorders, respectively.

SECTION REFERENCES

Almasy L, Blangero J: Multipoint quantitative-trait linkage analysis in general pedigrees. Am J Hum Genet *62:*1198, 1998.

Boehnke M, Langefeld CD: Genetic association mapping based on discordant sib pairs: the discordant-alleles test. Am J Hum Genet *62:*950, 1998.

Bouchard TJ, Propping P: *Twins as a Tool of Behavioral Genetics*. Wiley, Chichester, England, 1993.

*Chakravarti A: Population genetics—Making sense out of sequence. Nat Genet *21:*56, 1999.

Chee M, Yang R, Hubbell E, Berno A, Huang XC, Stern D, Winkler J, Lockhart DJ, Morris MS, Fodor SPA: Accessing genetic information with high-density DNA arrays. Science *274:*610, 1996.

Cheung VG, Gregg JP, Gogolin-Ewens KJ, Bandong J, Stanley CA, Baker L, Higgins MJ, Nowak NJ, Shows TB, Ewens WJ, Nelson SF, Spielman RS: Linkage-disequilibrium mapping without genotyping. Nat Genet *18:*225, 1998.

Collins FS, Guyer MS, Chakravarti A: Variations on a theme: Cataloging human DNA sequence variation. Science *278:*1580, 1997.

Elston RC: Algorithms and inferences: The challenge of multifactorial diseases. Am J Hum Genet *60:*255, 1997.

Gottesman II, Bertelsen A: Confirming unexpressed genotypes for schizophrenia: Risks in the offspring of Fischer's Danish identical and fraternal discordant twins. Arch Gen Psychiatry *46:*867, 1989.

Gottesman II, Moldin SO: Schizophrenia genetics at the millennium: Cautious optimism. Clin Genet *52:*404, 1997.

*Greenwood CMT, Bull SB: Analysis of affected sib pairs, with covariates—With and without constraints. Am J Hum Genet *64:*871, 1999.

Hauser ER, Boehnke M, Guo S-W, Risch NJ: Affected-sib-pair interval mapping and exclusion for complex genetic traits: Sampling considerations. Genet Epidemiol *13:*117, 1996.

Kendler KS: Genetic epidemiology in psychiatry: Taking both genes and environment seriously. Arch Gen Psychiatry *52:*895, 1995.

Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, Eaves LJ: Stressful life events, genetic liability, and onset of an episode of major depression in women. Am J Psychiatry *152:*833, 1995.

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: Parental treatment and the equal environment assumption in twin studies of psychiatric illness. Psychol Med *24:*579, 1994.

*Khoury MJ, Beaty TH, Cohen BH: *Fundamentals of Genetic Epidemiology*. Oxford University Press, New York, 1993.

Kidd KK: Can we find genes for schizophrenia? Am J Med Genet *74:*104, 1997.

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES: Parametric and nonparametric linkage analysis: A unified multipoint approach. Am J Hum Genet *58:*1347, 1996.

Kruglyak L, Lander ES: High-resolution genetic mapping of complex traits. Am J Hum Genet *56:*1212, 1995.

*Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genet *11:*241, 1995.

*McGuffin P, Owen MJ, O'Donovan MC, Thapar A, Gottesman II: *Seminars in Psychiatric Genetics*. Royal College of Psychiatrists, London, 1994.

Moldin SO: Detection and replication of linkage to a complex human disease. Genet Epidemiol *14:*1023, 1997.

*Moldin SO: Focus chapter. Research methods in behavioral genetics. In *Handbook of Research Methods in Clinical Psychology*, ed 2, PC Kendal, JN Butcher, editors. Wiley, New York, 1999.

Moldin SO: The maddening hunt for madness genes. Nat Genet *17:*127, 1997.

Moldin SO, Gottesman II: Genes, experience, and chance in schizophrenia: Positioning for the 21st century. Schizophr Bull *23:*547, 1997.

Plomin R, DeFries JC, McClearn GE, Rutter M: *Behavioral Genetics*, ed 3. Freeman, New York, 1997.

Rice JP: The role of meta-analysis in linkage studies of complex traits. Am J Med Genet *74:*112, 1997.

Risch NJ, Botstein D: A manic depressive history. Nature Genet *12:*351, 1996.

Risch NJ: Linkage strategies for genetically complex traits: I. Multilocus models. Am J Hum Genet *46:*222, 1990.

Risch NJ: A note on multiple testing procedures in linkage analysis. Am J Hum Genet *48:*1058, 1991.

*Risch NJ, Merikangas K: The future of genetic studies of complex human diseases. Science *273:*1516, 1996.

Sherman SL, DeFries JC, Gottesman II, Loehlin JC, Meyer JM, Pelias MZ, Rice JP, Waldman I: Recent developments in human behavioral genetics: past accomplishments and future directions. Am J Hum Genet *60:*1265, 1997.

*Sing CF, Haviland MB, Reilly SL: Genetic architecture of common multifactorial diseases. In *Variation in the Human Genome*. Ciba Foundation Symposium 197, Wiley, Chichester, UK, 1996.

Spielman RS, Ewens WJ: The TDT and other family-based tests for linkage disequilibrium and association. Am J Hum Genet *59:*983, 1996.

Spielman RS, Ewens WJ: A sibship test for linkage in the presence of association: The sib transmission/disequilibrium test. Am J Hum Genet *62:*450, 1998.

Turkheimer E, Gottesman II: Simulating the dynamics of genes and environment in development. Dev Psychopathol *8:*667, 1996.

Vogel F, Motulsky AG: *Human Genetics: Problems and Approaches*, ed 3. Springer-Verlag, New York, 1997.

Wang DG, Fan J-B, Siao C-J, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E, Robinson E, Mittmann M, Morris MS, Shen N, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES: Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280:*1077, 1998.

1.18 GENETIC LINKAGE ANALYSIS OF THE PSYCHIATRIC DISORDERS

CAROL A. MATHEWS, M.D., AND NELSON B. FREIMER, M.D.

In recent years genetic mapping techniques have been used to identify the underlying defects responsible for a number of important diseases, such as cystic fibrosis and Huntington's disease; such techniques are becoming increasingly important in studies of the psychiatric disorders as well, partly because attempts to find causes for these disorders using biochemical approaches have been largely unsuccessful. Genetic mapping studies are based on the idea that specific inherited mutations in the genetic material (the *genome*) of an individual are ultimately responsible for disease causation; such studies aim to isolate the responsible genes by identifying their chromosomal location within the genome through studies of affected individuals and their families. Each cell of all individuals contains two copies of each chromosome (called *homologues*), one inherited from the mother and one inherited from the father. During meiosis, the parental homologues cross over, or recombine, creating unique new chromosomes that are then passed on to the progeny. Genes that are physically close to one another on a chromosome are said to be *genetically linked*; those that are physically farther apart or are on different chromosomes are said to be *genetically unlinked*. Genes that are unlinked will recombine at random (i.e., there is a 50 percent chance of recombination). Genes that are linked will recombine less frequently than expected by random segregation; how much less often is proportional to the physical distance between them. Genetic mapping studies identify regions of the genome where recombination between two genetic loci is less than would be expected by chance, that is, areas of linkage. The chromosomal location of disease genes is identified by observation of inheritance patterns of diseases in conjunction with inheritance patterns of segments of deoxyribonucleic acid (DNA) of known chromosomal location (genetic markers).

Genetic markers are segments of DNA that contain known variations, called *polymorphisms*. Historically, a number of different types of markers have been used in genetic mapping studies, including the blood group antigens and other antigenic markers, as well as restriction fragment length polymorphisms (RFLPs), which are naturally occurring genomic sequence differences resulting in variations in restriction enzyme cutting sites in the genome. However, the low variability, small number, and uneven distribution of these markers has limited their power in linkage studies. Currently, microsatellite markers, or *simple sequence length polymorphisms* (SSLPs), are in wide use. Microsatellite markers are stretches of variable numbers of repeat units two to four base pairs in length, usually cytidine (C) - adenosine (A) repeats. Microsatellite markers have multiple advantages over other available genetic markers; they are distributed widely throughout genome, are highly polymorphic (the number of repeat units varies a great deal), and can be amplified with a technique called *polymerase chain reaction*.

Before the development of microsatellite markers, linkage was discovered for only a few Mendelian traits (i.e., traits for which a specific genotype at one particular locus is both necessary and sufficient to cause the trait). However, most medical and psychiatric diseases do not follow simple Mendelian inheritance patterns, but rather are examples of complex traits. A complex trait is any phenotype that does not exhibit classic inheritance (usually recessive or dominant) that can be attributed to a single locus. In other words, there is no one-to-one relationship between the genotype at a single locus and the phenotype in these disorders. This is caused by many factors, including incomplete penetrance (expression of the phenotype in only some of the individuals carrying the disease genotype), the presence of phenocopies (forms of the disease that are not caused by genetic factors), locus heterogeneity (different genes causing the same disease in different families or populations), or polygenic inheritance (the presence of multiple genes acting together to cause disease). Because it allows for close examination of virtually all chromosomal regions of interest using multiple markers, the construction of a genetic map that densely covers the genome with a large number of highly polymorphic markers (thousands are currently available) has been a pivotal development in the study of complex traits. Mapping a complex disorder involves several components; these include epidemiological studies to determine the likely mode of inheritance, definition of the phenotype to be studied, choice of an appropriate study population, and determination of correct technical and statistical approaches.

EPIDEMIOLOGICAL APPROACHES

The first step in identifying the gene or genes for a disorder is to determine whether or not the disease in question is in fact inherited. This is done using epidemiological approaches such as twin studies, relative risk determinations, and segregation analyses. In addition to indicating whether or not a disease is likely to have a genetic etiology, these studies are useful for estimating both the likely mode of transmission and the degree of genetic contribution, as well as for selecting appropriate families for further genetic studies.

Twin studies examine the concordance rates of a particular disorder (the percentage of twin pairs where both twins have the disorder) in monozygotic and dizygotic twins. For a disorder that is strictly determined by genetic factors, the concordance rate should be 100 percent in monozygotic twin pairs (because monozygotic twins share 100 percent in their genetic material in common) and 25 or 50 percent in dizygotic twin pairs (who are no more closely related than any siblings), depending on whether the disease is recessive or dominant, respectively. For a disorder where genetic factors are important but not exclusive, the concordance rates should be greater for monozygotic than for dizygotic twins; when genetic factors do not play a role the concordance rates should not differ between the twin pairs. This is based on the assumption that the environment for monozygotic twin pairs is no more similar than that for dizygotic twin pairs. Concordance rates for bipolar I disorder range between 65 percent and 100 percent in monozygotic twins and between 10 percent and 30 percent for dizygotic twins, indicating that it is highly genetic; concordance rates for schizophrenia, in contrast, are approximately 40 percent to 50 percent in monozygotic twins and 9 percent to 10 percent for dizygotic twins, indicating that although there is a genetic contribution to the disorder, it is weaker than that for bipolar I disorder.

Relative risk is defined as the rate of occurrence of the disease for the relatives of an affected individual divided by the rate of occurrence of the disease for the general population. A relative risk of greater than one suggests a genetic etiology; the magnitude of the measure gives an estimate of the genetic contribution to the disease. Relative risks can be calculated for sib pairs, parent-offspring pairs, and various other types of family relationships. The likely mode of transmission can be assessed by comparing the degree of relative risk for each type of relationship.

Segregation analysis compares the observed number of affected individuals within a given family to the expected number of affected individuals within that family, assuming a particular mode of inheritance, and then compares the likelihood of observing these data under the given genetic model to the likelihood of observing the data under other genetic models. These studies establish whether there is a genetic component to the disorder in question, and can also be used to estimate the likely genetic model (i.e., dominant or recessive). Because segregation analysis can be used to calculate the likelihood of a particular genetic model for individual families under specified parameters, it is useful both for choosing families for a genetic study when previous studies have indicated that the disease is Mendelian in only a proportion of affected families, and for choosing appropriate statistical models when studying a particular set of families.

Defining a Phenotype for Genetic Studies Because complex diseases, by definition, are not inherited in a simple Mendelian manner, they are often difficult to map. One approach to simplifying the genetic mapping process is to define the disease phenotype carefully. This is a particularly important step in the study of psychiatric disorders because they are defined empirically rather than on the basis of anatomic or physiological indices.

There are several elements to phenotype definition. The first involves deciding on the appropriate diagnostic criteria for the study in question, and deciding how these criteria will be applied to individuals in the study. For the psychiatric disorders there are several different sets of diagnostic criteria that are commonly used; these include the *Diagnostic and Statistical Manual of Mental Diagnoses* (DSM) diagnoses, the International Classification of Disease (ICD) schemas, and various research criteria such as the Research Diagnostic Criteria (RDC). Although these schemas overlap a great deal, there are significant differences between them. Because of the difficulty in precisely defining a disease phenotype for psychiatric disorders, the choice of a diagnostic schema can significantly influence the success of the study. It is preferable to use the most rigorous criteria available so as to minimize the likelihood of including individuals who do not actually have the disorder in question or who have a different form of the disorder.

It is also important to standardize the diagnostic procedures used to identify and classify individuals. One way of doing this is to use only experienced clinicians in the diagnostic process, and to train them in the administration of the instruments and the diagnostic criteria to be employed. Additionally, a best estimate procedure, a consensus diagnosis, or both are helpful. The best estimate process involves making use of every piece of information available, including medical records, interviews, and videotapes, where appropriate, to arrive at a diagnosis. For a consensus diagnosis, two or more diagnosticians independently review the material and make a diagnosis for each individual. The diagnoses are then compared; individuals for whom an agreement in diagnosis cannot be reached are not entered into the study.

The second element in phenotype definition involves making use of all available information about the genetic epidemiology of the disorder in order to choose a sample of affected individuals to study. It is often the case that a subset of affected individuals or families carry the disorder in what appears to be a simple Mendelian pattern whereas the inheritance pattern is less clear for other families or groups. When there are likely to be multiple genes contributing to a phenotype or disorder, it makes sense to begin with a study sample where there may be major loci. Redefining the disease phenotype can often simplify the mapping process by identifying such groups or families. For example, in the search for a genetic defect for Alzheimer's disease type, the process was advanced enormously by limiting the study population to those individuals who had early age of onset (before age 65); the early-onset trait segregated in an autosomal dominant fashion. Other ways of redefining the phenotype include focusing on specific ethnic groups, using family history as a prerequisite (it is common to require that there be at least two other affected individuals in a family), or restricting the study population to those with a particularly severe form of the disease. The rationale is that the severe forms of the disorder (early onset is thought to be an indicator of severity in many disorders) may be more genetically homogenous than the less severe forms and thus may be easier to map. This approach was successful in linkage studies of the long Q-T syndrome, in which the study group was limited to those affected individuals with a very severe form of the disorder.

Narrowing the phenotype is an effective way of increasing the chances of finding a genetic defect in complex diseases; however, it can greatly reduce the power of the study by limiting the number of affected individuals available. For this reason it has been argued that broadening the phenotype is an appropriate strategy for some disorders. The suggestion is that for some complex diseases the phenotype of interest may represent the extreme end of a spectrum, and that in order to have enough power to identify a gene, other disorders within the spectrum must also be included. This argument has been particularly favored in the study of schizophrenia. Family studies have indicated that there is an increased incidence of schizoaffective disorder, schizotypal personality disorder, and paranoid personality disorder in the biological families of schizophrenia patients, and individuals with these disorders are often included as affected subjects in genetic studies of schizophrenia. The same argument has been used in genetic studies of bipolar I disorder (using bipolar II disorder and major depressive disorder as alternate phenotypes), and Tourette's disorder, where chronic motor tic disorder and obsessive-compulsive disorder are considered alternate phenotypes. Although this strategy is often used, expanding the phenotype presumably leads to an increased rate of phenocopies, making detection of linkage more difficult. In addition, the theory that a disorder such as schizophrenia represents the extreme end of a spectrum of disorders also implies that the genetic effect for such a disorder may be too weak to be detected in mapping studies; this is supported by the concordance data from twin studies, which indicate a weak genetic effect for schizophrenia. So far, the strategy of using a broadly defined phenotype has not directly led to the identification of any disease genes. Another strategy that is similar to expanding the phenotype is to examine characteristics believed to be related to the disease in question, although not part of the diagnostic criteria. This technique was used in a study of non–insulin-dependent diabetes mellitus, which was mapped to chromosome 12 in a sample of affected individuals only after the phenotype was modified to account for mean insulin levels; although the overall genome screen provided no significant evidence of linkage, significant evidence of linkage was found on chromosome 12 when the study was limited to individuals with the lowest mean insulin levels.

Although the two approaches of narrowing the disease phenotype and broadening the disease phenotype may seem to be mutually exclusive, many groups studying complex disorders are attempting to incorporate both strategies into their study designs. One way to do this is to create stratified diagnostic categories, ranging from a narrow diagnostic category to a broad diagnostic category, and to test for genetic linkage under each of these schemas. There is currently a great deal of debate about whether this strategy is useful or not. Some investigators argue that for complex diseases that may actually be part of a spectrum, this strategy decreases the rate of false-negative results, that is, of missing an existing linkage because of misspecification. Others argue that using several models and picking the one that yields the highest scores statistically greatly increases the rates of false positives, that is, of identifying an area of linkage where none exists. One problem that clearly exists with the use of multiple diagnostic categories is that the level of significance attached to a particular statistical result drops with the use of more models.

The issues involved in phenotype definition are complicated, and within the field of genetics there is a great deal of disagreement about the best way to approach studies of complex traits. Because of the number of unknowns in other aspects of these studies, however, it is clear that the problem of how to define the disease to be studied is a serious one that should be addressed carefully.

Genome Screen Versus Candidate Gene Approach Once genetic epidemiology has provided as much information as possible about the genetic basis of a trait and the phenotype to be studied has been defined, the next step is to search for regions of the genome that may be linked to the disorder, either by examining candidate regions or by performing a screen of the entire genome. For disorders whose pathophysiology is known or partially understood, one or several candidate genes for the disorder may have already been identified. A candidate gene is usually a gene whose location has already been determined, whose protein product is known, and that is hypothesized to play a role in the pathophysiology of the disorder in question, although sometimes genes with unknown functions that are located in a particular chromosomal region of interest are also considered in this approach. Candidate genes for the neuropsychiatric disorders have traditionally been derived from obvious pharmacological information or from cytogenetic data (including translocations and deletions) indicating a candidate region. Examples of such potential candidate genes for neuropsychiatric disorders include the dopamine receptors and the serotonin receptors, as well as chromosomal translocations found in one or more individuals with the disease of interest.

The other approach to searching for linkage is to conduct a whole genome screen. This approach assumes that there are no known specific areas of interest within the genome (i.e., there are no candidate regions or any potential candidate genes have been excluded), and therefore involves screening the entire genome with evenly spaced markers in an attempt to find areas of potential linkage. Genome screens are becoming increasingly common for linkage studies involving psychiatric disorders, because of limited success with the candidate gene method in those cases where there are candidate genes, as well as a relative lack of candidate areas.

Several strategies have been developed to map disease genes; these methods are all based on identifying genetic marker alleles that are shared by affected individuals who are identical by descent. Marker alleles that are identical by descent are inherited from a common ancestor along with a linked disease gene. The optimal linkage method and type of analysis for a study sample is determined by the number of meioses separating the affected individuals in the sample via the common ancestor. These methods range from sib-pair studies, where affected individuals are separated from each other via a common ancestor by two meioses, to association studies in isolated populations, where affected individuals are separated from each other by up to about 100 meioses (Fig. 1.18-1). Alleles or haplotypes that are shared identical by descent must be differentiated from those that are shared identical by state; rather than being inherited from a common ancestor, these alleles are shared by affected individuals for a variety of other reasons, such as the same mutation in an allele occurring twice in a population. Dense, closely spaced marker maps are crucial for determining whether an allele is shared identical by descent or identical by state; markers closely surrounding an allele shared identical by descent will also be shared identical by descent, whereas markers closely surrounding an allele shared identical by state will not show sharing among those affected.

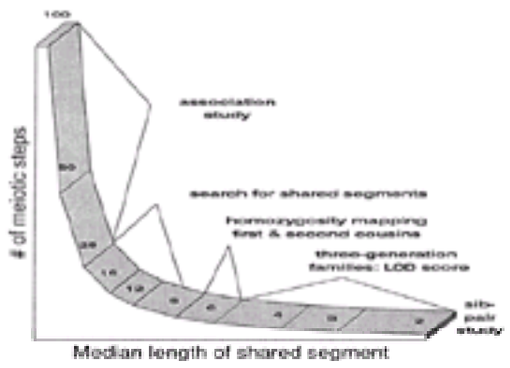

FIGURE 1.18-1 Depiction of the median lengths of the chromosomal segments inherited identical by descent around a disease gene by two individuals separated from a common ancestor by differing numbers of meiotic steps. (Reprinted with permission from Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB: Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nature Genet *8:*380, 1994.)

PEDIGREE ANALYSIS

Searching for areas of linkage to the disorder of interest, either in a whole genome screen or in a candidate area, has traditionally been done using pedigree analysis or logarithm of the odds (LOD) score analysis. This type of study is best done in families comprising three or four generations, where affected individuals are separated by two to six meiotic steps. The primary goal of pedigree analysis is to determine if two or more genetic loci (i.e., a genetic marker and the disease of interest) are cosegregating within a pedigree. For each marker locus the number of recombinants within the pedigree is compared to the number of nonrecombinants within the pedigree and a recombination fraction (q) is estimated. The recombination fraction represents the percentage of recombinations between two loci—in this case the hypothetical disease locus and the marker locus, and is equal to the genetic distance between the two loci [1 percent recombination is equivalent to 1 centimorgan (cM) in genetic distance and, on average, covers a physical distance of about 1 megabase (mB) of DNA]. A recombination fraction of 0.5 or 50 percent indicates that the two loci are not linked, but rather that they are segregating independently. An LOD score is then calculated to determine the likelihood that the two loci are linked at the distance suggested. The LOD score is calculated by dividing the likelihood of acquiring the data if the loci are linked at a given recombination fraction by the likelihood of acquiring the data if the loci are unlinked (q = 0.5), to obtain an odds ratio, and by then taking the $\log_{10}$ of this odds ratio to obtain an LOD score. The LOD score is calculated using the following equation:

$$\text{LOD score } Z(\theta) = \log_{10}[L(\theta)/L(0.5)]$$

where q is the recombination fraction from 0.0 to 0.5, $L(q)$ is the likelihood of the occurrence of the data if the loci are linked at a distance q, and $L(0.5)$ is the likelihood of the occurrence of the data if the loci are unlinked.

An LOD score can be obtained for various values of the recombination fraction, from 0.0 (completely linked) to 0.5 (unlinked). In addition, data from multiple family pedigrees can be combined to derive a cumulative LOD score (designated by Z). The value of q that gives the largest LOD score is considered to be the best estimate of the recombination fraction between the disease locus and the marker locus. This recombination fraction can then be converted into a genetic map distance. In general, the likelihood that any two given loci are linked (the prior probability of linkage) is expected to be approximately 1:50, based on the genetic length of the known genome. In order to compensate for the prior probability of linkage and bring the posterior (or overall) probability of linkage to about 1:20, the commonly accepted significance level of $p = 0.05$, a conditional probability of 1000:1 odds in favor of linkage (corresponding to an LOD score of 3 and a p value of 0.0001) is required. Therefore, under ideal conditions, for simple Mendelian disorders an LOD score of 3 is considered to be significant evidence of linkage.

A pedigree analysis consists of scanning the genome or a portion of the genome with a series of markers in one or more affected pedigrees, calculating an LOD score at each marker position, and identifying the chromosomal regions that show a significant deviation from what would be expected under the conditions of independent assortment. Because of the frequent use of multiple markers and multiple genetic models, as well as uncertainties in estimates of penetrance level and allele frequencies for many disorders, the question arises as to what represents acceptable evidence for linkage. In the instances where multiple genetic markers are tested, an LOD score of 3.4 has been suggested as providing sufficient evidence for linkage. The problem of using multiple genetic models is somewhat more difficult. One suggestion has been to use the formula 3 + log (t), where t is the number of tests performed. Under these parameters, an LOD score of 3.4 or higher would be required to provide significant evidence of linkage if more than one model is used, and an LOD score of 4 would be required if ten models are used. Many investigators argue that these criteria are too stringent because they assume that each model is independent of the others, which is not the case in mapping studies.

It is important, however, to account for the increased likelihood of obtaining a false positive in the context of multiple tests. In a standard genome screen of several hundred markers, false positives (regions that are significant at $p < 0.05$, corresponding to an LOD score of >1), can be expected approximately 24 times (once per chromosome) by chance alone (Fig. 1.18-2). This is important because some investigators have argued that an LOD score of >1 is a significant LOD score in genetic studies of a complex disease. For this reason, a classification system for reporting linkage based on dense genome scans has been proposed (Table 1.18-1). Although used by some investigators, these criteria are not universally accepted. Because the techniques and statistical methods used in the study of complex diseases are being developed at a rapid pace, there is currently no widely accepted standard for evaluating the significance of the findings of such studies.

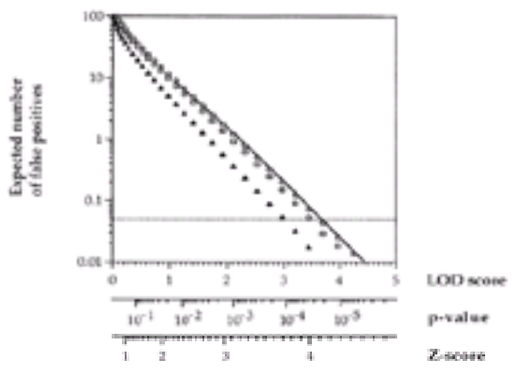

FIGURE 1.18-2 Number of false positives expected in a whole genome scan for a given threshold of LOD score. Solid line represents the expectation for a perfect genetic map. Symbols represent the results for 100 sib pairs using genetic maps with markers spaced every 0.1 cM (circles), every 1 cM (squares), and every 10 cM (triangles). The dotted line indicates the 5 percent genome-wide significance level. (Reprinted with permission from Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genet *11*:244, 1995.)

Suggestive linkage Statistical evidence expected to occur once at random in a genome screen	$p \leq 0.001$
Significant linkage Statistical evidence expected to occur about 0.05 times in a genome screen	$p \leq 0.0001$
Highly significant linkage Statistical evidence likely to occur 0.001 times in a genome screen	$p < 0.00002$
Confirmed linkage Significant linkage from one or a combination of initial studies that has subsequently been confirmed in a further sample.	

Reprinted with permission from Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genet 11:242, 1995.

Table 1.18-1 Suggested Criteria for Statistical Evidence of Linkage in Genetic Studies of Complex Traits

Although these criteria suggest that replication of the finding in a separate sample is necessary to confirm linkage, it is important to note that failure to replicate a linkage finding in another population does not necessarily mean that there is no linkage at that locus. Failure to replicate the finding may be caused by genetic heterogeneity, diagnostic differences between the study populations, or statistical fluctuation; in such cases, extension studies in the original population or a meta-analysis of all studies may help to confirm or refute the initial finding.

NONPARAMETRIC APPROACHES

In addition to traditional parametric approaches such as the pedigree method, multiple nonparametric approaches have been developed to address the problem of searching for linkage in the presence of genetic heterogeneity, incomplete penetrance, and unknown modes of transmission. These model-free tests for linkage are so

named because they do not require the specification of a model or the attendant parameters (mode of inheritance, type of penetrance, allele frequencies, and mutation and recombination rates). Nonparametric methods are less powerful in their ability to detect linkage than a correctly specified linkage model, but are often more robust in the face of unknown parameters than linkage analysis; these methods should find allele sharing in the presence of incomplete penetrance, phenocopies, genetic heterogeneity, and high-frequency disease alleles. For this reason, they are best used when the genetic etiology of the disorder is complex (i.e., when simple Mendelian inheritance patterns are not observed and linkage analysis is not a feasible option). Another advantage of nonparametric methods is that LOD score analysis of a small number of large families can be very sensitive to small changes in a few data points (e.g., a change from unaffected status to affected status), whereas nonparametric analyses, because they are usually done on a large number of small families, tend toward normal distributions and are thus less affected by such changes. The most commonly used nonparametric methods in the study of psychiatric disorders are sib-pair analysis and affected pedigree member analysis.

Sib-Pair Analysis Sib-pair analysis, first proposed in 1935, is a nonparametric method that is widely used in genetic studies of complex traits. Sib-pair analysis examines the frequency with which sib pairs concordant for a trait share a particular region of the genome compared with the frequency that is expected under random segregation. Although it is possible to do this test using both affected and unaffected siblings, usually only affected-affected sib pairs are evaluated because little information is gained by including unaffected sib pairs; excluding these individuals alleviates the problem of incomplete penetrance.

Sib-pair analysis is based on the concept that siblings share approximately 50 percent of their genomes identical by descent. Therefore, if a group of unrelated sib pairs with a certain trait shares a particular area of the genome significantly more than 50 percent of the time (the proportion of sharing expected under conditions of random segregation), that area of the genome is considered likely to be linked to the trait in question. In this method, siblings are genotyped, and population frequencies and parental genotypes are used to estimate the proportion of genes shared identical by descent at each site for each sib pair. Those pairs concordant and discordant for each genetic locus can be compared via three statistical tests, the means test, the proportions test, and the goodness-of-fit test.

Affected Pedigree Member Analysis The *affected pedigree member method* uses relatives other than sib pairs (although it can also include sib pairs) to search for evidence of an association between a disease and specific alleles shared by affected individuals. The affected pedigree member method examines all affected individuals in a pedigree to determine the number of alleles shared identical by state in excess of what would be expected merely on the basis of allele frequencies and the relationships between family members. The percentage of alleles shared identical by descent and thus evidence for linkage can then be inferred from the percentage of alleles shared identical by state. Because allele frequencies can have an impact on the percentage of alleles shared identical by state, they must be estimated cautiously; fluctuation in these frequencies can have a major influence on significance levels.

Association Studies Whereas linkage studies attempt to find cosegregation of a genetic marker and a disease gene within a family or families, association studies examine whether a particular allele occurs more frequently in affected individuals within a population. In general, association studies are based on the idea that certain alleles at markers closely surrounding a disease gene will be in linkage disequilibrium with the disease gene; that is, that these alleles will be carried more often than expected by random segregation in affected individuals because they are inherited identical by descent. Association studies of candidate genes make a slightly different assumption—that a particular variant allele is actually causative of the disease, as with the association studies of dopamine receptors in thrill-seeking behavior.

Linkage disequilibrium studies have traditionally been used to complement traditional pedigree analyses, either to further narrow a region of interest that has already been identified or to confirm a previous finding. However, these methods are now being used for whole genome screens, particularly for diseases where traditional linkage studies have been unsuccessful. These studies have one great advantage over a traditional pedigree analysis: because affected individuals are chosen from an entire population rather than from one or a few pedigrees, the number of potential subjects is limited only by the size of the population and the frequency of the disease. For disorders where genetic heterogeneity or incomplete penetrance are likely to be factors, maximizing the potential number of affected individuals that can be included in the analysis is extremely important. In addition, expanding the potential sample pool allows investigators to define the disease phenotype more rigidly, thereby increasing the likelihood that the phenotype under investigation is caused by a major gene effect.

Linkage disequilibrium studies are best done in populations that are descended from a small number of original founders. As a population gets older, recombination across the generations will gradually diminish the length of the chromosomal regions that are shared identical by descent, and individuals several generations away from the founding of the population should therefore share very few regions of the genome by chance. Individuals who share a common trait or disease in such a population are likely to have inherited the same genetic defect as well as specific alleles for markers in the region surrounding the putative gene. In general, such linkage disequilibrium approaches require that the population under investigation has undergone rapid expansion from a small original founding population and that there has been minimal immigration subsequent to the founding. For a genome screen attempting to find shared segments of the genome, in order to maximize the likelihood of finding an association in modern-day affected individuals given current marker map densities, the ideal population should be approximately six to twenty generations distant from the founders. A population that is older than this is likely to have conserved areas that are too small to be identified using currently available marker maps (which have a resolution of a few centiMorgans), increasing the chances of a false-negative result, whereas a younger population is likely to have more areas that are conserved identical by descent by chance, increasing the likelihood of a false-positive result. For studies that attempt to further delineate an area already identified by a traditional linkage study by examining associations between the disease gene and particular alleles at one or a few markers, an older population is appropriate. It is likely, however, that by the early twenty-first century a new marker map will be available at very high resolution (<1 cM); it will then be possible to consider genome-wide association studies in a number of different types of populations.

The degree of linkage disequilibrium in a particular area of the genome can be assessed in several ways. These include haplotype inspection, where segments of the genome are visually inspected for segments shared by all or most affected individuals, and two statistical tests, the transmission disequilibrium test and the haplotype relative risk statistic, which examine the alleles or series of alleles at different markers (haplotypes) that are transmitted from a parent to an affected individual under the assumption that a particular allele or haplotype is associated with the disease.

For association studies, as for linkage and other types of studies, there is some risk of obtaining a positive result purely by chance. Whereas in linkage studies each marker or genetic model tested is not completely independent of the other markers and models tested, for association studies each test performed carries an independent risk of a false-positive result. Therefore, the statistical correction that is applied to prevent such false positives should be more straightforward. It has been suggested that the threshold for significance for these tests should be set at $p = 0.05/n$, where n is the number of independent tests, or in the case of testing for multiple alleles at multiple markers $p = 0.05/n(m-1)$, where n is the number of markers and m is the number of alleles per marker. Again, as for linkage studies, association studies should be confirmed in an independent sample whenever possible.

Genetic mapping has undergone a number of important advances in the 1980s and 1990s, making it possible to find the genetic basis for many complex traits, including familial breast cancer, diabetes mellitus, and, more recently, neuropsychiatric disorders such as dementia of the Alzheimer's type. For the neuropsychiatric disorders, as for other complex traits, however, geneticists are no longer searching for one definitive genetic defect, as was the case with simple Mendelian diseases. Rather, the search has shifted to susceptibility genes, with the understanding that there are likely to be several genes for each clinical entity, acting either in concert with one another within a population, or independently in different populations, and that these gene effects are likely to be modified by environmental or developmental phenomena. Several neuropsychiatric diseases that have been extensively studied genetically, with varying degrees of success, include Alzheimer's disease bipolar I disorder, schizophrenia, and Tourette's disorder. Each of these disorders has provided a somewhat different insight into the difficulties inherent in the genetic mapping of complex disorders.

ALZHEIMER'S DISEASE

Alzheimer's disease is an excellent example of how advances in available technologies and methods have aided genetic mapping studies of complex diseases. The search for the genetic basis of Alzheimer's disease began with a traditional linkage study using a candidate area on chromosome 21; this effort evolved over time into a study of a number of other areas of the genome using several different approaches, culminating in the identification of defects in three unique genes that are responsible for Alzheimer's disease.

Dementia of the Alzheimer's type characterized by progressive impairment of memory and intellectual functioning. The clinical signs and symptoms, although characteristic, are not limited to this disorder but are also found in several other types of dementia. For this reason, the diagnosis of Alzheimer's disease can only be confirmed histopathologically at autopsy; the presence of senile plaques (made up of a core of beta-amyloid fibrils surrounded by dystrophic neurites), neurofibrillary tangles, and congophilic angiopathy in the brain parenchyma and associated blood vessels are pathognomonic.

Alzheimer's disease occurs in both familial and sporadic forms; approximately 60 percent of cases appear to be sporadic (i.e., do not have a family history). A variable

age of onset has been noted, ranging from as early as age 50 to as late as age 95. The concordance rate for Alzheimer's disease in monozygotic twin pairs is approximately 50 percent, indicating a reasonably strong genetic component.

Although phenotypically Alzheimer's disease appears to be fairly homogenous, epidemiological studies have indicated that it is genetically heterogeneous. In order to more clearly define a phenotype that would be amenable to genetic studies, investigators have divided dementia of the Alzheimer's type into early-onset and late-onset categories; early-onset Alzheimer's disease consists of individuals affected before age 65, whereas late-onset Alzheimer's disease consists of individuals affected after age 65. No consistent phenotypic differences between the early and late-onset categories have been noted; however, in a significant number of families the early-onset form of the disorder appears to be inherited in an autosomal dominant fashion with age-dependent penetrance. In late-onset Alzheimer's disease the etiology is less clear—although some cases are probably inherited as an autosomal dominant, most are almost certainly not.

Chromosome 21 It has long been recognized that many individuals with Down's syndrome (trisomy 21) develop an early-onset dementia that is clinically and histopathologically indistinguishable from Alzheimer's disease. For this reason, chromosome 21 was considered to be an excellent candidate region for initial genetic studies of this disorder. In 1987 a group of investigators, using four large families that showed evidence of autosomal dominant transmission of early-onset Alzheimer's disease, looked at chromosome 21 using 10 restriction fragment length polymorphisms (RFLPs) as genetic markers. They found evidence for linkage (multipoint analysis gave an LOD score of 4.06) near two marker loci close to the centromere in the 21q11.2-21q22.2 region. Because amyloid b-peptide (a proteolytic product of amyloid precursor protein) is present in senile plaques and was known to map to chromosome 21, it was hypothesized that a defect in this gene, leading to either overexpression of normal b-amyloid, or expression of an abnormal form of b-amyloid, was responsible for Alzheimer's disease. Although the original linkage finding was later found to be a false positive, this hypothesis was strengthened when the gene coding for amyloid precursor protein was identified later in 1987 and shown to be located at 21q11.1-21q21.2. Mutations were subsequently found both in coding regions within the amyloid precursor protein gene, resulting in amino-acid substitutions, and in areas that lie outside the b-amyloid sequence, resulting in overproduction of b-amyloid or production of longer fragments in individuals with Alzheimer's disease. Transgenic mice with mutated amyloid precursor protein complementary DNA (cDNA) were shown to produce b-amyloid deposits and senile plaques, as well as showing synapse loss, astrocytosis, and microgliosis, all part of the pathology of the disorder. Although no neurofibrillary tangles were observed in these mice, these findings nevertheless represent striking evidence that mutations in the b-amyloid gene are indeed responsible for at least some of the histopathological elements of Alzheimer's disease. To date, there have been six missense mutations identified near or within the b-amyloid sequence of the amyloid precursor protein on chromosome 21 that are present in the families with early-onset Alzheimer's disease.

Chromosome 14 It is interesting to note that none of the families that were originally used in the studies of chromosome 21 and provided the first suggestion of linkage in this area ultimately showed linkage to the amyloid precursor protein gene. For this reason, and because linkage to chromosome 21 was excluded in number of other families evidencing early-onset Alzheimer's disease, further genome screening was done to search for evidence of linkage elsewhere. As result of these studies, linkage on chromosome 14 was independently identified in such families by four teams studying a large number of families from the general population, and was found to be the strongest at 14q24.3. Using traditional linkage studies, investigators were able to narrow the region of interest on chromosome 14 to 8.9 cM; haplotype inspection further delineated the region, and a physical map of this area was constructed using yeast artificial chromosomes. Exon trapping and direct selection of cDNA from the yeast artificial chromosomes then identified several new genes. Five missense mutations were found in early-onset Alzheimer's disease families in one of these transcripts, containing a previously uncharacterized gene that was subsequently named *presenilin-1 (PS-1)*. It is an integral transmembrane protein with at least seven transmembrane domains; its function is currently unknown. The *PS*-1 gene on chromosome 14 has since been recognized as a major locus for Alzheimer's disease in the early-onset familial form of the disorders contributing to up to 70 percent of cases. This area of chromosome 14 has also been examined in late-onset Alzheimer's disease families; however, there is currently no evidence for linkage in these families.

Chromosome 1 Although the *PS*-1 gene accounts for a significant proportion of families with early-onset Alzheimer's disease, it does not account for all such cases. In particular, *PS*-1 does not account for this disorder in a group of families of Volga-German descent. These families, which have a high incidence of the early-onset form of the disease, are known to be descended from one German family that emigrated to Russia from Germany, and then to the United States. In this group Alzheimer's disease segregates in an autosomal dominant fashion, and was hypothesized to result from a single gene defect inherited from a common ancestor. Initial linkage results from a genome search in several families of Volga-German descent did not show linkage to the chromosome 14 site or to the chromosome 21 site, but did identify yet another region, this time on chromosome 1 in the area of 1q31-q42. Following this linkage study and the identification of *PS*-1, a search of a publicly available cDNA database identified a cDNA sequence showing high homology to the *PS*-1 gene on chromosome 14, and which was found to map to within the chromosome 1 region of interest for Alzheimer's disease. This gene was subsequently named *presenilin-2 (PS-2)*. Two missense mutations have been found in *PS*-2 in the Volga-German families and in several talian families with early-onset Alzheimer's disease.

Presenilin-1 and Presenilin-2 Mutations The genes for *PS*-1 and *PS*-2 are expressed in most tissues in the body and share approximately 67 percent overall sequence homology, with the highest similarity being in putative transmembrane domains. The presence of high-sequence homology suggests that *PS*-1 and *PS*-2 have similar functions. However, the intronic sequences of *PS*-2 are not similar to those of *PS*-1, indicating that these genes are not the result of a recent gene duplication, but rather are two separate genes. In addition, as a mutation in only one of the two is sufficient to cause disease, it is clear that one gene cannot compensate for the other.

PS-1 mutations account for approximately 30 to 70 percent early-onset Alzheimer's disease families. To date, 47 missense mutations and one point mutation in *PS*-1 have been found; all but one of the mutations involve amino acids that are conserved between *PS*-1 and *PS*-2. *PS*-1 mutations are for the most part fully penetrant with a narrow age-of-onset range (35 to 55 years); it appears that for Alzheimer's disease caused by this gene, age of onset is determined by both the nature of the mutation and by the position of the mutation within the gene. In addition, although most mutations result in early-onset Alzheimer's disease, some *PS*-1 mutations may also result in late-onset Alzheimer's disease. Although less is currently known about *PS*-2 mutations, they do appear to result in a later and more variable age of onset (up to the 70s and 80s) than do *PS*-1 mutations.

Chromosome 19 For early-onset Alzheimer's disease traditional linkage studies were successful in identifying linkage to three genes, in part because careful phenotype definition allowed investigators to identify large families in which the disease appeared to segregate in a Mendelian fashion. However, these approaches were unsuccessful in studies of late-onset Alzheimer's disease in part because the mode of inheritance was less clear in individuals with this disorder. For this reason, investigators used a nonparametric approach, the affected pedigree member method, to search for evidence of linkage in a large number of small families with late-onset Alzheimer's disease. In 1991, the results of a linkage study using 36 markers in such families using the affected pedigree member method provided evidence (p <0.001) for a susceptibility gene on the long arm of chromosome 19. Subsequent multipoint LOD score analysis gave an LOD score of 4.60 on chromosome 19 between two markers in the area of 19q13.2. There were several candidate genes known to map to this area, including growth factors, proteases, and apolipoprotein E. The major known function of apolipoprotein E is in lipid transport, but other functions have also been suggested. Although it has no known specific causal relationship to Alzheimer's disease, it is found on postmortem in senile plaques, neurofibrillary tangles, and vascular amyloid deposits in the brain of individuals with this disorder. In addition apolipoprotein binds to b-amyloid in the cerebrospinal fluid with high affinity. For this reason, studies were undertaken to determine if an association between Alzheimer's disease and any of the expressed alleles of apolipoprotein E was present.

Apolipoprotein E In 1993 several epidemiological studies identified the e4 allele of the apolipoprotein E gene as a genetic risk factor for late-onset Alzheimer's disease. There are three known alleles of this gene—e2, e3, and e4. In most populations, the e3 allele is the most common. However, in the familial late-onset form of the disease the incidence of e*4* is approximately 50 percent and in sporadic late-onset Alzheimer's disease it is 40 percent, compared with 16 percent in normal controls. Studies have demonstrated that there is a gene dose effect for the e4 allele in families with late-onset disease, with the risk for disease increasing and the mean age of onset decreasing with increasing number of e4 alleles. In the late-onset families, those who are homozygous for e4 are affected 91 percent of the time, with a mean age of onset at 68.4 years. Those who are heterozygous for e4 are affected 47 percent of the time, with a mean age of onset at 75.5 years, and those with no e4 are affected 20 percent of the time with a mean age of onset at 84.3 years.

It has been shown that *apoE*-e4 binds more readily to b-amyloid than does *apoE*-e3, and that patients homozygous for *apoE*-e4 have more b-amyloid deposited in senile plaques and cerebral blood vessel walls than do those homozygous for *apoE*-e3. Although it seems likely that *apoE* is directly involved in the pathogenesis of Alzheimer's disease, it should also be noted that the possibility remains that *apoE*-e4 is a genetic marker for another gene located close by that is actually the susceptibility gene Alzheimer's disease, rather than having a role in and of itself.

Alzheimer's disease is a neuropsychiatric disorder with multiple causes, presumably both genetic and environmental. Although the genetics of this complex disorder are not fully elucidated, the history of the search for disease genes for this disorder illustrates how important each phase of the approach can be, from carefully defining the phenotype to making use of available epidemiological studies, as well as the importance of using all available data, including the public databases. To date, genes have been identified on chromosomes 21, 14, and 1, coding for the b-amyloid peptide of amyloid precursor protein, *PS*-1 and *PS*-2. Together, mutations in these genes account for about 10 percent of those affected, primarily those with familial early-onset disease. An additional 15 percent of affected individuals have sporadic early-onset disease, while 30 percent has the familial late-onset form and 45 percent the sporadic late-onset variant. In addition, a promising area on the

long arm of chromosome 19 has been identified as being potentially linked to the late-onset familial form of the disease. This area is promising because it is the location of the gene coding for apolipoprotein E, which has been previously shown to be related to Alzheimer's disease. Presence of the e4 allele of this gene in these families correlates both with increased risk of getting the disease and for decreased age of onset of the symptoms in a dose-effect manner (Fig. 1.18-3).

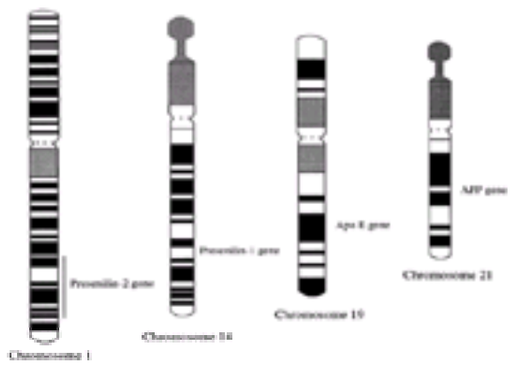

FIGURE 1.18-3 Chromosomal location of the genes implicated in Alzheimer's disease. Apo E = apolipoprotein E. APP = amyloid precursor protein.

Although the pathophysiology of the disease remains a mystery, b-amyloid appears to play a significant role in the course of the disease. The role of the presenilin proteins is unclear at this point; however, studies are currently under way to investigate the function of the presenilin proteins in normal individuals, as well as the effect at a cellular level of mutations in these genes in mice.

BIPOLAR I DISORDER

The search for the genetic basis of bipolar I disorder has been less successful than that for Alzheimer's disease; fraught with missteps and partial answers, the history of genetic mapping attempts for this disease illustrate not only the extreme complexity of the psychiatric disorders, but also the evolution of linkage studies of such diseases. Bipolar I disorder is an episodic illness characterized by recurrent periods of both mania and depression; psychotic symptoms are often a part of the clinical picture, particularly in more severely affected individuals. Although there continues to be disagreement the disorder is considered by many to be part of an affective disorders spectrum ranging from major depressive disorder, bipolar I and II disorders to schizoaffective disorder, bipolar type. Some epidemiological studies have indicated that there may be a single major gene with incomplete penetrance responsible for the disorder in some families; the mode of inheritance in these studies is inconsistent—some have indicated autosomal dominant inheritance whereas others have suggested X-linked or recessive inheritance. However, more recent studies indicate that the genetics of bipolar I disorder are more complex; these studies provide evidence for genetic heterogeneity, multifactorial inheritance, or both. Ongoing problems with linkage studies of bipolar I disorder have included difficulties with specification of genetic parameters such as phenocopy rate, model specification, and phenotype definition. Because using a narrowly defined phenotype often means using fewer individuals in a family or population and thus can result in an unacceptable loss of power, many studies done so far have used a broader definition of affected status that includes bipolar II disorder or major depressive disorder. Although intended to increase the ability to detect linkage, these methods often ultimately result in a decreased ability to detect linkage because of incorrectly specified parameters.

First-Generation Studies The first set of genetic studies following the introduction of DNA markers to report positive linkage to bipolar I disorder were published in 1987; they reported linkage on chromosomes X and 11. Like the initial studies of Alzheimer's disease, these studies used information about the clinical presentation of the disorder to identify likely areas of interest in the genome. In the first study of bipolar I disorder, investigators noted that in several families this and other mood disorders appeared to be inherited in an X-linked fashion and that they appeared to cosegregate with color blindness and G6PD deficiency, which map to the X chromosome. Linkage studies in several Israeli pedigrees with bipolar I disorder as well as color blindness or G6PD deficiency gave LOD scores between 4.75 and 9.17. The study of chromosome 11, which was based on previous studies that indicated a possible linkage to markers on this chromosome, was done in a population of Old Order Amish; two-point linkage analysis gave LOD scores of 4.08 and 2.63 with two RFLP markers on chromosome 11, and multipoint linkage analysis gave a peak LOD score of 4.9. Additionally, visual inspection of haplotypes showed that specific alleles for these two markers segregated with the illness in the three available generations.

Not surprisingly, these findings generated a great deal of interest; both studies showed high LOD scores and seemed to provide clear evidence for linkage. However, replication studies in other populations failed to produce positive results for either the X chromosomes or for chromosome 11. Following these attempts at replication, some members of each group extended and reanalyzed the data from the original studies, incorporating several important changes, in an attempt to confirm the linkage findings. The pedigrees in the Israeli study were reevaluated, and linkage studies were done using several DNA markers in the area. Using these markers, evidence for linkage disappeared in all but one family; in this family the LOD score dropped to 2.09. The most likely explanation for this is that the initial linkage studies used variations in phenotype (color blindness and G6PD deficiency) as markers rather than DNA polymorphisms; such markers are less informative and recombinations are frequently missed. In addition, the affection status of some individuals changed between the first and the second study; it was noted that some errors had been made regarding G6PD status in some individuals, and some individuals previously considered unaffected with bipolar I disorder later developed the disorder. These changes in phenotype also affected the linkage results. Although one family continued to have a suggestive LOD score, the pedigree could not be extended for further study, making confirmation impossible.

When the Amish pedigrees were reevaluated, the most significant change was a change in status from unaffected to affected in two members of the original pedigree. These two individuals did not share the alleles of interest in the original two markers, and this single change lowered the LOD scores for the few markers from 4.9 and 2.63 to 1.75 and 1.29, respectively. Adding data from 10 unaffected individuals who had not been previously included also decreased the LOD scores, although to a lesser extent.

In addition, in an attempt to extend the study, a new branch of the pedigree had been identified that contained both affected and unaffected individuals. As there was evidence that bipolar I disorder might segregate separately in this new branch, it was analyzed both alone and as a part of the larger kindred. In both instances, linkage of bipolar I disorder to the original area on chromosome 11 was excluded.

The failure to replicate or extend these linkage findings, which initially appeared to be so promising, illustrates the vulnerability of such studies to relatively small changes in affection status, as well as the problems inherent in using markers with low informativeness or only a few markers in an area of interest. The next generation of studies attempted to rectify these problems by using multiple markers in what could be considered partial genome screens. In a partial screen, markers widely spaced across the genome are tested, and any giving LOD scores above a certain low threshold are further investigated. Often in these studies, only a few markers are initially tested, and if these markers give evidence for areas of interest the areas are further investigated and the genome screen is postponed.

Second-Generation Studies The first partial genome screen that gave evidence for linkage to bipolar I disorder was completed in 1994 and gave initial results on chromosome 21. This study initially gave high LOD scores for ten markers on 21q22.3 in one family (the maximum LOD score was 3.41) under an autosomal dominant model, as well as significant results in this region using the affected pedigree member method. However, attempts at replication or confirmation of this finding were unsuccessful, as were those with chromosome 11 and X chromosome. Although both the LOD score analysis and the affected pedigree member method gave evidence for linkage that was significant or at least suggestive, they were found in only one pedigree, and attempts to strengthen the finding by adding other families to the sample significantly diminished evidence for this finding rather than enhancing it. In addition, there was no consistent haplotype for this region in the affected individuals. Other investigators have been unable to replicate this finding in other samples.

Another partial genome screen led to a possible area of linkage near the centromere of chromosome 18, although the results of this study are somewhat harder to interpret. This study examined 11 markers on chromosome 18 in 22 families and found several positive two-point LOD scores in individual families; the highest LOD score was 2.38. However, of these positive LOD scores, no clear pattern was observed; some of the markers gave positive LOD scores under a recessive model in some families, some gave positive LOD scores under a dominant model in other families, and some markers gave positive LOD scores in some families under both models. Nonparametric tests provided somewhat stronger evidence for linkage, with p values of less than 0.05 for 7 of the 11 markers, but again, no clear patterns

emerged.

Although these data may represent evidence for a susceptibility locus for bipolar I disorder in the pericentromeric region of chromosome 18, indicating either a small gene effect present in many families or a major locus responsible for the phenotype in a few particular families, the data do not clearly point to either a specific mode of inheritance or to linkage to specific markers in the area. Because multiple markers and multiple genetic models were tested, a statistical correction is necessary, reducing the statistical significance of the findings. In addition, the LOD scores were highest under a broad diagnostic category, which included bipolar I disorder, bipolar II disorder, schizoaffective disorder, and recurrent major depressive disorder; the LOD scores were lower when a narrow diagnostic category was used, calling into question the hypothesis that these LOD scores represent a susceptibility gene specific to bipolar I disorder. The fact that nonparametric methods supported the positive LOD scores suggests that the region should be further explored, but clearly these results in themselves are not sufficient evidence of linkage in this area.

Attempts to replicate this finding in other populations have for the most part been unsuccessful. One group was able to exclude linkage in the pericentromeric area of chromosome 18 in five families with bipolar I disorder using both linkage studies and nonparametric methods. Another group used one of the families from the original study (the same Old Order Amish kindred reported in the linkage studies) as well as five additional Amish pedigrees, and found no evidence for linkage except in the original kindred, where they obtained a maximum LOD score of 1.94. Nonparametric methods provided no evidence for linkage.

One group, however, did find evidence to support linkage to the chromosome 18 area. Hypothesizing that a parent-of-origin effect might play a role in the genetics of bipolar I disorder, the researchers divided their sample into those families where the disorder appeared to be inherited from the paternal side and those where the disorder appeared to be inherited from the maternal side, and analyzed these groups together and separately. Using nonparametric methods, the researchers found evidence for excess allele sharing in several markers in the pericentromeric area of chromosome 18 (18p11). This group also analyzed markers all along chromosome 18, and found excess allele sharing in the area of 18q as well. Two-point LOD score analysis, as well as multipoint LOD score analysis, provided some support for linkage in both areas, with LOD scores of 1.45 in one marker each in both regions. Interestingly, the evidence for linkage in the 18p area was obtained under a broad diagnostic category, which included bipolar I and II disorders and recurrent major depressive disorder, whereas the evidence for linkage in the 18q region was found under a narrow diagnostic category, which included only bipolar I and bipolar II disorder. The results for 18q were somewhat strengthened when the pedigrees were divided into paternal and maternal samples; two-point LOD scores in the area of 18q were between 2.6 and 3.5 in the paternal pedigrees, the peak multipoint LOD score in this area was 3.11, and there was evidence of excess allele sharing for several markers in the area. Again, however, the evidence for 18p was weaker, with evidence for linkage at only one marker. This group attempted to replicate their findings in a new sample of thirty pedigrees with bipolar disorders, using the same diagnostic criteria as the original study, they found evidence for excess allele sharing among affected sib pairs in two markers in the region of interest on 18q, but little evidence for linkage on 18p. In the new sample they found reduced evidence for the parent-of-origin effect that they noted in their original sample.

Third-Generation Studies The next step in the progress of linkage studies of BP was complete genome screening; because the results of the partial genome screens were difficult to interpret, and because of the ongoing difficulties in replication and extension of these studies, a more systematic search for susceptibility genes for bipolar disorder was initiated by several investigators. A complete genome screen provides an opportunity to compare the LOD scores (or *p* values) in potential areas of interest with those in other areas of the genome, thus giving an estimate of the level of background “noise” and the chances that the results of interest represent a false positive rather than a true area of linkage. For example, an LOD score of 1.5 would be of more interest in a genome screen where a few LOD scores are above 0.5 than it would in a genome screen where a large number of LOD scores are above 1.0.

Results of a complete genome screen in a kindred of Old Order Amish that subsumed the pedigrees from the earlier studies showed some promising sites, including markers in the areas of 6pter-p24, 13q13, and 15q11-qter, although there was no evidence for linkage on chromosomes 11 or X. One marker in each of these three areas gave *p* values of 0.001 or less using nonparametric methods; however, under LOD score analysis none of the markers had an LOD score greater than 3, and none of the flanking markers were strongly positive. In addition, there was no evidence of shared haplotypes. Although it may be very likely that these findings represent false positives, these areas should be more closely examined using additional markers and additional families in order to further explore the possibility that they contain susceptibility loci.

The genetic studies of bipolar disorders in the Amish illustrate many of the inherent difficulties of doing genetic analysis on complex disorders. The Amish appear to be a perfect population for such investigations; they are an isolated population with known genealogies, clean phenotypes with little comorbidity, and large extended families. However, the findings for bipolar disorders in this group have so far been inconsistent or contradictory, and even in this ideal population there has been no strong evidence for linkage that has withstood attempts at confirmation or replication.

Yet another susceptibility locus for bipolar disorders has been potentially located on chromosome 4. In 1996, as part of a genome search in twelve families with bipolar I, bipolar II, and recurrent major depressive disorder, investigators found evidence for linkage on chromosome 4p in one family, with a two-point LOD score of 4.1 at one marker, as well as positive LOD scores in the flanking markers under an autosomal dominant model. Three-point analysis with neighboring markers gave an LOD score of 4.8. In all cases, the LOD score was highest under the narrow phenotypic model, which included bipolar I and bipolar II disorder subjects only. Haplotype analysis showed a haplotype of seven markers that were shared by all 11 affected individuals and by 14 of 16 of those affected with recurrent major depressive disorder, and extended-relative pair analysis showed increased allele sharing at two markers that was suggestive but not significant.

When the other 11 families were added, the maximum LOD score, assuming genetic homogeneity, dropped to between 2.9 and 3.3. There was, however, evidence for locus heterogeneity in these families, with only approximately 30 to 35 percent of the families likely to be linked to this locus. Under the assumption of genetic heterogeneity, the maximum LOD score for the combined families was 4.1. Although it is important to be cautious in interpreting these results, the LOD scores in this study indicate that there may be a locus for bipolar disorders in some families at 4p16. The fact that haplotype analysis and extended-relative pair analysis support this finding to varying degrees, and that extending the study to several other families seemed to confirm the results, makes it worth further examination.

In 1996 a third genome screen, this time in two large Costa Rican pedigrees, gave evidence for linkage on the long arm of chromosome 18 somewhat below the area previously reported on 18q, as well as two other areas of potential linkage on 18p and 11p (Fig. 1.18-4). Under a narrow diagnostic model three consecutive markers on chromosome 18 in the area of 18q22-q23 had LOD scores that exceeded 1.6. When this area was further examined using haplotype evaluation, association tests, and joint linkage association tests, a haplotype (composed of several alleles that are relatively rare in the general population of Costa Rica) was found in 16 of 17 affected individuals in one family. A distinct but similarly sized haplotype in the same region was found in 7 of 9 affected individuals in the other family; 3 of the 9 individuals in this family also shared some portion of the first haplotype. In all, 23 of the 26 affected individuals in the two families shared some portion of a core haplotype in this area. Association-linkage scores (which provide joint evidence for linkage and association compared with the assumption of no linkage and no association) were significant for two markers, with LOD scores of 3.70 and 4.06. An independent sample of affected individuals in Costa Rica is currently under investigation in an attempt to confirm this finding and to further investigate the areas on 18p and 11p.

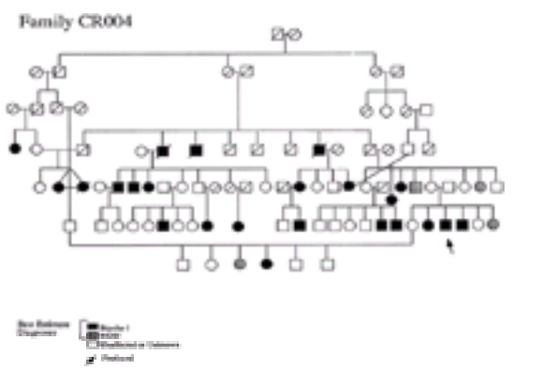

FIGURE 1.18-4 Pedigree of a large Costa Rican family containing multiple members affected with bipolar I disorder and major depressive disorder (MDD). (Reprinted with permission from Freimer NB, Reus VI, Escamilla M, Spesny M, Smith L, Service S, Gallegos A, Meza L, Batki S, Vinogradov S, Leon P, Sandkuijl L: An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. Am J Med Genet *67:*258, 1996.)

Because of the continued suggestion of one or more susceptibility loci located on chromosome 18, other groups have also examined this area for evidence of linkage.

Several of these groups have found some evidence for linkage, although weak, at or near 18q23; support for linkage at 18p has been even more uncertain.

The combined evidence of these studies, although somewhat contradictory and confusing, appears to point to two different susceptibility loci on chromosome 18: one on 18p and one on 18q. These loci, if they do indeed represent two independent true areas of linkage, are different in intriguing ways. The 18p finding is perhaps the most difficult to explain because for most studies a single marker explains the majority of the positive findings, and each study points to a different marker. In addition, the evidence for a locus in this area appears to be strongest in individual families—the evidence actually diminishes when families are combined, and it is strongest under a broad phenotype definition that includes not only bipolar I disorder but also major depressive disorder.

In contrast, the 18q finding extends over several markers, and appears to be robust to the addition of new families. Although different markers were tested in the different studies, they all lie within the area of the larger haplotype noted in the Costa Rican families. Finally, this finding appears to be more robust when the phenotype is narrowly defined—the evidence weakens or disappears when the phenotype is expanded to include major depression.

Unlike Alzheimer's disease, where at least part of the genetic story has been successfully elucidated, bipolar I disorder has been resistant to ongoing attempts at clarification. This is in part because of the fact that not only is there likely to be genetic complexity in bipolar I disorder (i.e., locus heterogeneity, incomplete penetrance and so forth), but there is a greater risk for phenotypic complexity as well. There is a great deal of symptom overlap between bipolar I disorder and other mood and psychotic disorders, and currently, as the different inclusion criteria in the studies implies, there is also a great deal of disagreement about the degree of relationship between these disorders. The difficulty in interpreting these studies points to the importance of carefully following up isolated positive results by examining flanking markers, by adding new families or affected individuals, and by attempting replication of significant or near-significant results in an independent sample. Nothwithstanding the uncertainties and contradictions, these studies indicate several promising areas of the genome that warrant further examination. Specifically, the areas of 4p, 18p, and 18q, which either show significant LOD scores or nonparametric results in a single sample or ongoing suggestive evidence in a number of samples, are promising. In addition, new approaches to the study of bipolar I disorder, such as population-based studies and large-scale sib-pair analyses, may allow for the implementation of strategies such as stricter phenotype definition without a concomitant loss of power. These strategies in turn may increase the likelihood of finding loci that either have a small gene effect or are present in only a small proportion of affected individuals.

SCHIZOPHRENIA

Schizophrenia is another psychiatric disorder that has for a variety of reasons defied most attempts at genetic mapping. Perhaps the most significant problem that has been encountered in genetic studies of schizophrenia is the difficulty in phenotypic characterization. The phenotype for schizophrenia is broad, and the cardinal symptoms, delusions and hallucinations, are also found in a number of other psychiatric disorders. More specific characteristics of schizophrenia such as so-called negative symptoms (e.g., flat or inappropriate affect, social isolation, and poverty of thought) and a progressive downhill course, can be difficult to recognize or quantify. Schizophrenia is thought by some investigators to represent the extreme end of a spectrum of psychotic disorders ranging from schizoid personality disorder, schizotypal personality disorder, and paranoid personality disorder to schizoaffective disorder. The diagnosis of schizophrenia is further complicated by the fact that the definition of the disorder has varied dramatically, both across years and between psychiatric centers, leading to significant problems in constructing and interpreting epidemiological studies. However, epidemiological studies done using relatively recent, widely accepted operationalized criteria have indicated that there is likely to be a genetic component for the susceptibility to schizophrenia. Relative risk studies have shown a 10-fold increase in the risk of developing schizophrenia among relatives of affected individuals as compared with relatives of controls, and twin studies show a concordance rate of 40 to 50 percent for monozygotic twins and about 10 percent for dizygotic twins. These studies indicate that although schizophrenia is at least in part genetic, it is not likely to be caused by a single major gene but rather by several genes acting in concert with each other and the environment. Research also indicates that the genetic contribution to schizophrenia is likely to be weaker than that for bipolar I disorder, potentially making it a much more difficult disease to study. Attempts to use measures that are thought to be related to schizophrenia, such as abnormalities in smooth-pursuit eye movements, have also been unsuccessful. Although for the most part linkage and association studies of schizophrenia have been unsuccessful, several recent studies have provided evidence that one or more susceptibility genes for this disorder may soon be within reach.

Early Studies Early studies of schizophrenia were, like those of Alzheimer's disease, based on observed cytogenetic abnormalities in individuals with schizophrenia-like symptoms. In the late 1980s a report of a chromosomal translocation on chromosome 5 in an uncle-nephew pair with schizophrenia prompted several linkage studies in the area of 5q11-q13. Using varying diagnostic models ranging from narrow (including schizophrenia, schizophreniform disorders, and unspecified functional psychosis) to very broad (including all psychiatric disorders seen in the sample) and various genetic models, investigators obtained maximum LOD scores of 2.45 to 6.49 in a group of seven large pedigrees. The highest LOD scores (4.33 to 6.49) were obtained using an extremely broad definition of affected status that included essentially all psychiatric disorders; LOD scores for the narrower phenotypic categories were lower (between 3.22 and 4.33). Although these LOD scores are significant, the fact that they are lower in the narrow than in the broad phenotypic categories suggests that this area of potential linkage may not be specific to schizophrenia, and thus may be difficult to replicate or confirm. Subsequent groups have failed to find evidence of linkage to the 5q11-q13 are or have excluded linkage to this region in their study families; association studies have also produced negative results. In addition, there has been no follow-up to this finding in the form of extension or fine mapping studies in the original population, indicating that it is likely to represent a false positive.

The next area of the genome that gave evidence for a possible susceptibility locus was on chromosome 22. An initial partial genome screen using 240 markers in a subset of families with multiple cases of schizophrenia or schizoaffective disorder, using an autosomal dominant model, identified an area of interest on 22q13. Although the LOD scores were fairly low (three contiguous markers gave LOD scores between 0.94 and 1.54), they are of interest because velo-cardio-facial syndrome (VCFS), a disorder in which affected individuals have been described as having schizophrenia-like symptoms beginning in adolescence or early adulthood, is known to be caused by microdeletions in 22q11, near the area identified in the genome screen for schizophrenia. However, attempts to replicate these linkage results using an additional 217 families from multiple sites and the same parameters as the original study have provided no further evidence for linkage.

Genome Screens Although earlier linkage studies of schizophrenia have pointed to some potentially interesting areas, as is true for bipolar I disorder, the most promising evidence of linkage for a susceptibility gene for schizophrenia has come from complete genome screens. At least two independent genome screens of families with schizophrenia have pointed to one area in particular that is suggestive for linkage—the short arm of chromosome 6. These studies have been supported by follow-up replication and extension studies. Following up on positive LOD scores on 6p25-p24 obtained in an initial genome screen, one group of investigators obtained a maximum LOD score of 3.51 in this area under the assumption of 50 percent heterogeneity (i.e., under the assumption that only 50 percent of the study families were likely to be linked to the locus). Sib-pair analysis provided additional support for linkage in the study families, although the overall results were weaker. Again, all the results were strongest for broad diagnostic categories; a narrow phenotype definition gave nonsignificant LOD scores. In addition, although suggestive for linkage in the region of 6p24-22, this study provided little information about the likely mode of inheritance; for a given diagnostic category, the LOD scores for the codominant, dominant, and recessive models were often similar.

A follow-up study that attempted to replicate the finding of linkage in the 6p area using a sib-pair analysis and a narrow diagnostic phenotype (schizophrenia and schizoaffective disorder only) found excess allele sharing in affected sib pairs at one of the markers that originally showed linkage and at several additional markers in this region, with a maximum LOD score of 2.2. These results provide support for linkage to schizophrenia on chromosome 6, although they do not provide clear confirmation of the original finding; although the regions of interest overlapped in the two studies, they were not identical. Further support for a susceptibility locus on 6p came from a second complete genome screen. Investigators in this study found preliminary evidence of linkage on 6p in five Icelandic pedigrees using a nonparametric method called weighted-rank pairwise correlation. Follow-up of this finding in an additional 65 families also pointed to the 6p25-p24 region; when the two samples were combined, a slight increase in the probability of linkage in this area was obtained but did not meet stringent criteria for significant linkage. Because of the previous work on chromosome 6 and these positive findings, however, 6p25-p24 was examined in a third sample using association methods; in this sample evidence of linkage disequilibrium on 6p was obtained. These data were once again combined with the data from the original study, and gave significant evidence of linkage ($p = 0.00004$) at one marker. However, attempts at replication in a fourth sample were unsuccessful, and a separate replication study in a sample of 57 pedigrees gave only mildly positive evidence of linkage (maximum LOD score of 1.17, sib-pair analysis p value <0.004). Another study that focused on the short arm of chromosome 6 took a slightly different approach. After traditional pedigree and sib-pair analysis produced no significant evidence for linkage on 6p in 10 large Canadian families with schizophrenia, this group used a quantitative trait approach to examine the hypothesis that their study families carried a susceptibility locus specific to the positive symptoms of schizophrenia, including disorganized thinking, hallucinations, and delusions. Using this approach they found significant evidence for linkage (nominal $p = 1.2 \times 10^{-5}$ and 5.4×10^{-6}, and empirical $p = 0.34$ and 0.0085, respectively) in two markers in the 6p11 to 6p21 region, slightly more centromeric than the area implicated in the other studies, but adjoining it.

These studies all point to a possible susceptibility locus on chromosome 6p in or near the area of 6p25-p24, although the evidence is not conclusive. In addition, at least two studies provided no evidence for a susceptibility locus for schizophrenia on 6p using several of the original markers and the same genetic models and parameters used in the first study. Interpretation of the results of these studies is complicated, and there is as yet no clear consensus. Although it is possible that these results represent a false positive, it is also possible that they represent the presence of genetic heterogeneity, that is, that a susceptibility locus for

schizophrenia does indeed exist on 6p, and possibly on 22q, but that these loci are relevant in only a proportion of affected families or individuals. Because the study of the genetics of schizophrenia necessarily involves many complicating factors, a logical next step would be to perform association studies in an appropriate population or to attempt large-scale sib-pair studies. These approaches allow for careful phenotype definition and selection of appropriate study subjects, as well as increasing the number of possible subjects, thereby increasing the chances of finding linkage or association in the presence of complicating factors such as genetic heterogeneity or incomplete penetrance.

TOURETTE'S DISORDER

Although the data is confusing at best and contradictory at worst, there currently appears to be evidence for at least one susceptibility locus for schizophrenia, perhaps more. In contrast, there have been no reports of positive linkage findings to Tourette's disorder from several ongoing genetic studies. Tourette's disorder is a neuropsychiatric disorder with onset in childhood that is characterized by chronic intermittent motor and vocal tics. Twin studies have shown it to have a monozygotic concordance rate of 50 to 70 percent, as compared to a dizygotic concordance rate of 8 percent, indicating that there is a strong genetic component. As with other neuropsychiatric disorders, phenotype definition has been a critical part of the genetic study of Tourette's disorder. Because of the strong phenotypic similarities between it and other tic disorders such as chronic motor tic disorder, many investigators have suggested that these disorders lie along a contiuum, with transient tic disorders at one extreme and Tourette's disorder at the other. Segregation analyses have supported this theory to some extent, indicating that chronic motor tic disorder and obsessive-compulsive disorder, which has been noted to cosegregate in families with Tourette's disorder, may actually be phenotypic variants of it. Although the spectrum approach is not universally accepted and brings with it the problem of potential phenocopies, because of the difficulty of obtaining families with sufficient numbers of affected individuals most studies done to date have included individuals with chronic motor tic disorder and obsessive-compulsive disorder in addition to those with Tourette's disorder.

As noted previously, however, linkage studies have been unsuccessful despite these efforts to increase the power of the study by expanding the phenotype. Early suggestions that susceptibility loci might be located on chromosome 18 or chromosome 9 based on observed cytogenetic abnormalities in a few individuals with Tourette's disorder were not supported by linkage studies in these areas. In addition, although early segregation analysis indicated that the inheritance pattern of the disorder was most likely to be a single major gene transmitted in an autosomal dominant fashion with incomplete penetrance, complete genome screening efforts have also failed to find evidence of linkage. An ongoing collaborative genome screen for Tourette's disorder, which currently encompasses more than 600 markers, has so far indicated exclusion of linkage in up to 95 percent of the genome. Another, more recent study has reported preliminary LOD scores of 1.57 and 2.21 at two individual markers (one on chromosome 8 and one on chromosome 17) in a few families; further work in these areas using additional markers and additional families is currently under way. Candidate gene studies examining the dopamine type 1 (D_1) type 2 (D_2), type 3 (D_3), type 4 (D_4), and type 5 (D_5) receptors, the dopamine transporter protein, prodynorphin, proopiomelanocortin, gastrin-releasing peptide, tyrosine hydroxylase, and dopamine hydroxylase have also been negative. Because of the lack of success with linkage and candidate gene studies despite evidence of a strong genetic contribution to Tourette's disorder, investigators are currently reexamining their assumptions about the disorder in an attempt to identify methods of investigation that are likely to be more successful than the traditional approaches have been.

Segregation Analyses Early segregation analyses indicated that the mode of inheritance of Tourette's disorder was most likely to be autosomal dominant with incomplete penetrance (with a lower penetrance for females than for males) and variable expression of phenotype that included chronic motor tic disorder and obsessive-compulsive disorder. Because linkage analysis using this model has consistently failed to uncover any areas suggestive of linkage, however, the data from these studies have been recently reanalyzed and the underlying assumptions closely examined.

In the first such study, a single large pedigree was reanalyzed under an assumption of assortative mating when it was noticed that a higher-than-expected number of married-in spouses were also affected (36 percent). Some clinical studies have suggested that families in which both parents are affected with tic disorders or Tourette's disorder have children with more severe symptoms than those in which only one parent has a tic disorder. Reanalysis of this pedigree indicated that the correlation resulting from assortative mating was significant; the results also suggested an additive model of inheritance in which heterozygotes (who carry one copy of the disease allele) have a lower risk of developing Tourette's disorder than do homozygotes (who carry two copies of the disease allele), and in which nongene carries have no risk of developing the disorder. A second group also suggested an additive rather than dominant or recessive model of inheritance as a result of their segregation analysis of several families with the disorder; this study additionally provided evidence for the presence of a strong multifactorial background. Extrapolating from the results of their study, the investigators hypothesized that under this model all individuals homozygous for Tourette's disorder allele would be affected, but only 2.2 percent of males and 0.3 percent of females heterozygous for the allele were likely to be affected.

At the same time as some investigators were reevaluating the assumptions underlying the presumed genetic models of Tourette's disorder, others were reevaluating the assumptions underlying the presumed phenotype. The assumption that chronic motor tic disorder or obsessive-compulsive disorder are variant phenotypes of Tourette's disorder for the purpose of genetic studies, although supported by early segregation analyses, ignores the possibility that they may also result from different genetic or nongenetic forms in some or all cases. For this reason, the question of appropriate phenotype specification was examined in a series of simulations using two different phenotypic approaches. The first approach was designed to minimize the risk of obtaining false positives by including only the most severe phenotypes (i.e., individuals with severe Tourette's disorder only); it has the potential drawback of a much smaller sample size, leading to unacceptable loss of power. The second approach was designed to effectively expand the pool of available subjects by doing multiple linkage tests using gradually broadened disease phenotypes; this strategy potentially increases the rate of false-positive results, particularly if corrections for multiple testing are not made. These strategies were compared using simulations in which the true linkage in the population was hypothesized to be for either Tourette's disorder only, for Tourette's disorder chronic motor tic disorder, or for Tourette's disorder, chronic motor tic disorder, and obsessive-compulsive disorder. Like the earlier studies, an autosomal dominant model with incomplete penetrance and a low population gene frequency was assumed. Using the first approach an LOD score of greater than 3 was found for only one family (from a total of 32); in contrast, the LOD score was greater than 3 for four families using the second approach, indicating that the power of detection of linkage was greater for the broader models. When the simulations were repeated using an incorrect model (i.e., assuming a broad phenotype when in fact the true phenotype was narrow or vice versa), the results indicated that if several diagnostic models were tested, the chances of finding linkage were always >70 percent, whereas if only the narrower phenotype was used, the likelihood of finding linkage was high only if chronic motortic disorder and obsessive-compulsive disorder were not in reality variant forms of Tourette's disorder; if they were variant phenotypic expressions, the first approach was much less likely to detect linkage. These results suggest that doing linkage studies using a hierarchical system of phenotype definitions ranging from narrow to broad (with appropriate corrections for multiple testing) could substantially increase the chances of finding linkage to Tourette's disorder.

Unlike Alzheimer's disease, bipolar I disorder, and schizophrenia, there are currently no significant or even suggestive linkage findings for Tourette's disorder. Segregation analyses continue to suggest a strong genetic component, and to support the hypothesis of a single major locus effect, perhaps with an additional environmental component. Given this, and despite evidence that the available family data should be sufficiently informative for linkage studies, the continued difficulty in finding linkage using traditional methods is surprising. Thus, attention has turned to other ways of elucidating the genetic etiology of this disorder. One such strategy is to assume that Tourette's disorder is a heterogeneous disorder and that those with the most severe symptoms are the most likely to have a clear genetic component. Because of the potential loss of power associated with restricting the phenotype to only severely affected cases of the disorder, thereby decreasing the pool of potential subjects, attention is shifting to population-based studies, such as linkage disequilibrium and sib-pair analyses, where larger samples sizes can be collected than are possible in family studies. At least two such studies are currently under way. If they are successful, they will have a significant impact on the way the study of complex diseases is approached.

AGGRESSION

The genetics of aggression is difficult to study because it is poorly defined as a phenotype. Rather than being a syndrome of collected symptoms with a generally agreed-upon definition, aggression is a behavior, with a phenotype that is open to interpretation. Some scientists studying aggression consider only acts of violence against others in their definition of the phenotype; others add acts of violence against oneself, such as suicide attempts, self-mutilatory behavior, and violent thoughts. One definition that is in widespread use and can be applied to both animals and to humans states that aggression is behavior that involves an intent to harm someone else. The key to this definition is in the concept of *intent*, which can be difficult to define in both humans and in animals but provides a way of differentiating aggression from accidental injury. A somewhat narrower definition that applies to humans exclusively is that of violence; *violence* is commonly accepted to mean aggressive behavior by humans towards humans. Measures of aggression vary widely, and range from self-reports to family reports to observations of behavior. Most assessment instruments measure aggression as a relatively permanent personality trait (i.e., as a propensity towards violent behavior) rather than measuring the behavior itself, although newer instruments are beginning to measure this too. These instruments commonly measure either feelings such as anger, resentment, and suspicion; behaviors; or a combination of the two.

Despite these difficulties, several investigators have attempted to measure a genetic contribution to aggressive behavior. A meta-analysis of data from 24 genetic epidemiological studies, mainly twin and adoption studies, suggested that up to 50 percent of the variance in aggressive behavior results from a genetic effect, even

after controlling for the effects of common environments. The results of the individual studies in this set were quite variable, ranging from no evidence of a genetic effect, to a concordance for aggression between identical twin pairs of up to 0.85. The meta-analysis also examined the differences between studies that used self or family report measures of aggression and those that used observational ratings and found that the use of observational ratings greatly reduced the effect of genetic factors and greatly increased the effect of environmental factors. The results of the meta-analysis also showed that both genetics and common environment were equally important in determining aggressive behavior in youths, whereas the impact of common environment decreased and the effect of genetics increased in adults, suggesting that if there is a genetic component to aggression, its impact may change over time.

To date there has only been one study reporting convincing linkage for aspects of aggression in humans. This study examined a large Dutch kindred in which male family members had borderline mental retardation and abnormal behavior, including impulsive aggression, attempted rape, arson, and exibitionism. A genetic defect was identified in p11-p21 of the X chromosome, in the region near the monoamine oxidase type A (MAO_A) and type B (MAO_B) genes, with an LOD score of 3.55. Subsequent mutation analysis revealed a mutation (changing a glutamine to a termination codon) that was present in all affected individuals and in the two obligate carries studied, but was not seen in the unaffected members of the family. Functional studies, in which fibroblast cells were grown in the presence of tryptamine and then exposed to dexamethasone (which produces a 6-fold to 14-fold increase in MAO_A production in normal cells), noted a significant decrease of MAO_A activity in affected individuals and a normal level of MAO_A activity in unaffected individuals.

As a result of this study, and because of reports indicating that aggressive behavior is correlated with lowered concentrations of 5-hydroxyindoleacetic acid (5-HIAA), a serotonin metabolite, in the cerebrospinal fluid of monkeys and humans, several mouse knockout models have been made to examine the relation between genes involved in the serotonin and MAO pathways and aggression. Mice in whom the gene that codes for MAO_A was knocked out were found to have behavioral abnormalities as pups, and increased aggressiveness as adults; these mice showed increased levels of serotonin and norepinephrine and decreased levels of 5-HIAA as pups. Interestingly, the levels of both serotonin and its metabolite returned to normal in the adults. Both pups and adult mice showed persistent changes in the somatosensory cortex on histological examination. Similarly, mice lacking serotonin (5-hydroxytryptamine [5-HT] subtype 1B (5- HT_{1B}) receptor (the homologue of the $5\text{-}HT_{1DB}$ receptor in humans) showed an increase in aggressive behavior when compared to wild-type mice. A third knockout mouse model, deficient in a-calmodulin-dependent protein kinase II (CaMK II), also showed an increase in aggressive behavior, although in this instance the abnormality was seen only in those mice heterozygous for the mutation. The homozygous mutant mice, although not aggressive, showed a large number of other behavioral abnormalities. CaMK-II–deficient mice also showed decreased serotonin release. CaMK-II–mediated phosphorylation is required for the activation of tryptophan hydroxylase, the rate-limiting enzyme in the synthesis of serotonin. The results of these studies appear to be somewhat contradictory at first; the MAO_A knockout mice, which showed an increase in central serotonin concentrations, had increased aggressiveness, as did the $5\text{-}HT_{1B}$ and CaMK-II knockout mice, who showed a decrease in serotonin availability. One explanation that has been put forward is that in order to produce a change in behavior such as an increase in aggression, the genetic defect must affect the developmental process of the animal; thus, it is not the measurable concentrations of available serotonin in adult animals that is important, but the effects of altered concentrations on early development. Although not entirely satisfactory, this hypothesis is particularly appealing in light of the observation that serotonin concentrations, although abnormal in the genetically altered mouse pups, returned to normal in adult mice, and may explain why some genetic mutations associated with increased aggression in adults cause lowered serotonin concentrations whereas others cause increased serotonin concentrations.

At least two other genetic abnormalities, not known to be associated with the serotonin pathways, have also been found to be associated with increased aggression in mice. These abnormalities are in the nitric oxide synthase *(NOS)* gene and in the adenosine receptor a_{2a}. These mice were similar to the other knockout mice discussed in that the males showed increased aggression under testing circumstances when compared to wildtype mice. In addition, the mice with defects in the *NOS* gene showed an abnormal increase in the amount of sexual behavior with nonreceptive females, and the mice with defects in the adenosine receptor showed less exploratory behavior as well as having prominent hypoalgesia. The mechanisms behind these behavioral abnormalities are currently unknown.

The results of these human and mouse studies suggest that aggression is likely to have a complex causation including multiple, perhaps additive, genetic effects. Currently, only a single small kindred with multiple other behavioral and cognitive abnormalities provides evidence for a specific genetic contribution to aggression in humans. Although it is a promising start, this finding by no means explains the etiology of aggression.

The field of genetics has developed a great deal in the last several years, and continues to grow at a rapid pace. For many diseases of interest, it is no longer possible to rely exclusively on traditional linkage studies assuming Mendelian inheritance. New approaches and methodologies that address or circumvent the problems inherent in the study of complex disorders, such as genetic heterogeneity, unknown genetic parameters, uncertain inheritance patterns, and variable phenotypes, have previously led to the identification of genetic defects for such complex disorders such as breast cancer and diabetes. These methods are also successfully being applied to the genetic study of neuropsychiatric disorders, such as Alzheimer's disease, bipolar I disorder, schizophrenia, and Tourette's disorder. Neuropsychiatric disorders have not only provided the impetus but have also served as models for the development and use of new genetic approaches such as complete genome screens using linkage disequilibrium and sib-pair analyses, as well as emphasizing the importance of epidemiological studies, phenotype definition, and careful selection of the study sample. In particular, the genetic studies of Alzheimer's disease have clearly illustrated the value of careful phenotype definition and the use of a variety of methodological approaches, as well as the importance of incorporating all available information, such as epidemiological studies and molecular and genetics databases, into the study. In the next several years the use of these tools will help to elucidate the genetic bases for a number of other neuropsychiatric diseases, including those discussed here.

SUGGESTED CROSS-REFERENCES

Section 1.10 on basic molecular neurobiology and section 1.17 on population genetics provide basic information on the mechanisms of cell division and genetic recombination as well as the required elements of genetic study. Chapter 9 describes the classification of mental disorders, including the use of diagnostic schemas and research instruments. Chapter 12 describes the clinical diagnosis and epidemiology of schizophrenia, Chapter 14 describes the clinical diagnosis and epidemiology of mood disorders, Chapter 42 describes the clinical diagnosis and epidemiology of tic disorders, and section 51.3e and Chapter 10 describe the clinical diagnosis and epidemiology of Alzheimer's disease.

SECTION REFERENCES

Baron M, Freimer NB, Risch N, Lerer B, Alexander JR, Straub RE, Asokan S, Das K, Peterson A, Amos J, Endicott J, Ott J, Gilliam TC: Diminished support for linkage between manic depressive illness and X-chromosome markers in three Israeli pedigrees. Nature Genet *3:*49, 1993.

Berrettini WH, Ferraro TN, Goldin LR, Detera-Wadleigh S, Choi H, Muniec D, Guroff JJ, Kazuba DM, Nurnberger JI, Hsieh W-T, Hoehe MR, Gershon ES: A linkage study of bipolar illness. Arch Gen Psychiatry *54:*27, 1997.

Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JI, Gershon ES: Chromosome 18 DNA markers and manic-depressive illness: Evidence for a susceptibility gene. Proc Natl Acad Sci USA *91:*5918, 1994.

Blackwood DHR, He L, Morris SW, McLean A, Whitton C, Thomson M, Walker MT, Woodburn K, Sharp CM, Wright AF, Shibasaki Y, St. Clair DM, Porteous DJ, Muir WJ: A locus for bipolar affective disorder on chromosome 4p. Nature Genet *12:*427, 1996.

Brzustowicz LM, Honer WG, Chow WC, Hogan J, Hodgkinson K, Bassett AS: Use of a quantitative trait to map a locus associated with severity of positive symptoms in familial schizophrenia to chromosome 6p. Am J Hum Genet *61:*1388, 1997.

Cruts M, Hendriks L, Van Broeckhoven C: The presenilin genes: A new gene family involved in Alzheimer's disease pathology. Hum Molec Genet *5:*1449, 1996.

*Dracopoli NC, Haines JL, Korf BR, Moir DT, Morton CC, Seidman CE, Seidman JG, Smith DR: *Current Protocols in Human Genetics*. Wiley, Bethesda, MD, 1996.

Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Rojas E, Gallegos A, Meza L, Fournier E, Baharloo S, Blankenship K, Tyler DJ, Batki S, Vinogradov S, Weissenbach J, Barondes SH, Sandkuijl LA: Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nature Genet *12:*436, 1996.

Freimer NB, Reus VI, Escamilla M, Spesny M, Smith L, Service S, Gallegos A, Meza L, Batki S, Vinogradov S, Leon P, Sandkuijl L: An approach to investigating linkage for bipolar disorder using large Costa Rican pedigrees. Am J Med Genet *67:*254, 1996.

Hen R: Mean genes. Neuron *16:*17, 1996.

Hendriks L, Van Broeckhoven C: The bA4 amyloid precursor protein gene and Alzheimer's disease. Euro J Biochem *237:*6, 1996.

Heutink P, van de Wetering BJM, Pakstis AJ, Kurlan R, Sandor P, Oostra BA, Sandkuijl LA: Linkage studies on Gilles de la Tourette's syndrome: What is the strategy of choice? Am J Hum Genet *57:*465, 1995.

Houwen RHJ, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, Sandkuijl LA, Freimer NB: Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nature Genet *8:*380, 1994.

Kelsoe JR, Ginns EI, Egeland JA, Gerhard DS, Goldstein AM, Bale SJ, Pauls DL, Long RT, Kidd KK, Conte G, Housman DE, Paul SM: Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. Nature *342:*238, 1989.

*Lander E, Kruglyak L: Genetic dissection of complex traits: Guidelines for interpreting and reporting linkage results. Nature Genet *11:*241, 1995.

Lander E, Shork NJ: Genetic dissection of complex traits. Science *265:*2037, 1994.

Leckman JF, Sholomskas D, Thompson WD, Belanger A, Weissman MM: Best estimate of lifetime psychiatric diagnosis. Arch Gen Psychiatry *39:*879, 1982.

*Lichter DG, Dmochowski J, Jackson LA, Trinidad KS: Influence of family history on clinical expression of Tourette's syndrome. Neurology *52:*308, 1999.

McMahon FJ, Hopkins PJ, Jianfeng X, McInnis MG, Shaw S, Cardon L, Simpson SG, MacKinnon DF, Stine OC, Sherrington R, Meyers DA, DePaulo JR: Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. Am J Hum Genet 61:1397, 1997.

Miles DR, Carey G: Genetic and environmental architecture of human aggression. J Pers Soc Psychol *72:*207, 1997.

*Ott J: *Analysis of Human Genetic Linkage*. Johns Hopkins University Press, Baltimore, 1992.

Pauls DL: Issues in genetic linkage studies of Tourette's syndrome: Phenotypic spectrum and genetic model parameters. Adv Neurol *58:*151, 1992.

Pericak-Vance MA, Bebout JL, Gaskell PC, Yamaoka LH, Hung W-Y, Alberts MJ, Walker AP, Bartlett RJ, Haynes CA, Welsh KA, Earl NL, Heyman A, Clark CM, Roses AD: Linkage studies in familial Alzheimer's disease: Evidence for chromosome 19 linkage. Am J Hum Genet *48:*1034, 1991.

Pulver AE, Karayiorgou M, Lasseter VK, Wolyniec P, Kasch L, Antonarakis S, Housman D, Kazazian HH, Meyers D, Nestadt G, Ott J, Liang K-Y, Lamacz M, Thomas M, Childs B: Follow-up of a report of a potential linkage for schizophrenia on chromosome 22q12-q13.1: Part 2. Am J Med Genet *54:*44, 1994.

Reus VI, Freimer NB: Understanding the genetic basis of mood disorders: Where do we stand? Am J Hum Genet *60:*1283, 1997.

*Risch N: Genetic linkage: Interpreting lod scores. Science *255:*803, 1992.

Roses AD: Apolipoprotein E alleles as risk factors in Alzheimer's disease. Ann Rev Med *47:*387, 1996.

St. George-Hyslop PH, Tanzi RE, Polinsky RJ, Haines JL, Nee L, Watkins PC, Myers RH, Geldman RG, Pollen D, Drachman D, Growdon J, Bruni A, Foncin J-F, Salmon D, Frommelt P, Amaducci L, Sorbi S, Placentini S, Stewart GD, Hobbs WJ, Conneally M, Gusella JF: The genetic defect causing familial Alzheimer's disease maps on chromosome 21. Science *235:*885, 1987.

Schellenberg GD, Bird TD, Wijsman EM, Orr HT, Anderson L, Nemens E, White JA, Bonnycastle L, Weber JL, Alonso ME, Potter H, Heston LL, Martin GM: Genetic linkage evidence for a familial Alzheimer's disease locus on chromosome 14. Science *258:*668, 1992.

Schwab SG, Albus M, Hallmayer J, Hönig S, Borrmann M, Lichtermann D, Ebstein RP, Ackenheil M, Lerer B, Risch N, Mair W, Wildenauer DB: Evaluation of a susceptibility gene for schizophrenia on chromosome 6p by multipoint affected sib-pair linkage analysis. Nature Genet *11:*325, 1995.

Sherrington R, Brynjolfsson J, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H: Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature *336:*164, 1988.

*Strachan T, Read AP: *Human Molecular Genetics.* Bios Scientific Publishers, New York, 1996.

Straub RE, Lehner T, Luo Y, Loth JE, Shao W, Sharpe L, Alexander JR, Das K, Simon R, Fieve RR, Lerer B, Endicott J, Ott J, Gilliam TC, Baron M: A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. Nature Genet *8:*291, 1994.

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS: A potential vulnerability locus for schizophrenia on chromosome 6p24-22: Evidence for genetic heterogeneity. Nature Genet *11:*287, 1995.

*Tecott LH, Barondes SH: Behavioral genetics: Genes and aggressiveness. Curr Biol *6:*238, 1996.

Tsuang MT, Faraone SV: *The Genetics of Mood Disorders*. Johns Hopkins University Press, Baltimore, 1990.

Van Broeckhoven CL: Molecular genetics of Alzheimer's disease: Identification of genes and gene mutations. Eur Neurol *35:*8, 1995.

van de Wetering BJM, Heutink P: The genetics of the Gilles de la Tourette's syndrome: A review. J Lab Clin Med *121:*638, 1993.

Volavka J: *Neurobiology of Violence*. American Psychiatric Press, Washington, DC, 1995.

Walkup JT, LaBuda MC, Singer HS, Brown J, Riddle MA, Hurko O: Family study and segregation analysis of Tourette's syndrome: Evidence for a mixed model of inheritance. Am J Hum Genet *59:*684, 1996.

*Yasuda M, Maeda K, Hashimoto M, Yamashita H, Yoshitaka H, Ikejiri Y, Bird TD, Tanaka C, Schellenberg GD: A pedigree with a novel presenilin 1 mutation at a residue that is not conserved in presenilin 2. Arch Neurol *56:*65, 1999.

1.19 BASIC SCIENCE OF SLEEP

J. CHRISTIAN GILLIN, M.D., ERICH SEIFRITZ, M.D., REBECCA ZOLTOSKI, PH.D., AND RAFAEL J. SALIN-PASCUAL, M.D., PH.D.

Sleep disturbances are particularly pertinent to psychiatrists and other mental health professionals because these complaints are frequently associated with depression, anxiety, substance abuse, obsessive-compulsive disorders, psychosis, acute and chronic stress, and many other disorders. The neuroscience of sleep regulation, in addition, is important to understand the pathophysiology of psychiatric disorders and the mechanism of action of psychoactive drugs. New insights into the regulation of sleep and wakefulness emerging from molecular biology, genetics, and neuroimaging techniques promise exciting progress in the field of sleep physiology research in particular, and in the pathophysiology of psychiatric disorders in general. The association between sleep and mood disorders is, for instance, intriguing if one considers the powerful antidepressant effect of one night of sleep deprivation, and the recurrence of depressive symptoms after one night of recovery sleep. Furthermore, the sleep electroencephalogram (EEG) has some value as a predictor for treatment outcome as well because it may be used as a biological variable to follow patients during treatment. In addition, exciting new findings on the association between memory function and sleep also emphasize the importance of sleep in neuroscience and in psychiatry.

PHENOMENOLOGY OF SLEEP AND WAKEFULNESS IN NORMAL HUMANS

Sleep Stages and Their Normal Organization Most mammals have two major phases of sleep: *rapid eye movement (REM) sleep* and *non-rapid eye movement (NREM) sleep*. REM sleep is sometimes called *dreaming* or *D sleep* because it is associated with dreaming, or *paradoxical sleep*, because the electroencephalogram EEG becomes activated during this state of sleep. NREM sleep, on the other hand, conforms to traditional concepts of sleep as a time of decreased physiological and psychological activity; it is therefore sometimes called *orthodox sleep* or *slow wave sleep* (SWS), *EEG synchronized sleep*, or *S sleep*.

NREM sleep is further divided into four sleep stages on the basis of visually scored EEG patterns (Table 1.19-1, Fig. 1.19-1).

Table 1.19-1 Technical Terms Commonly Used in Characterizing Sleep

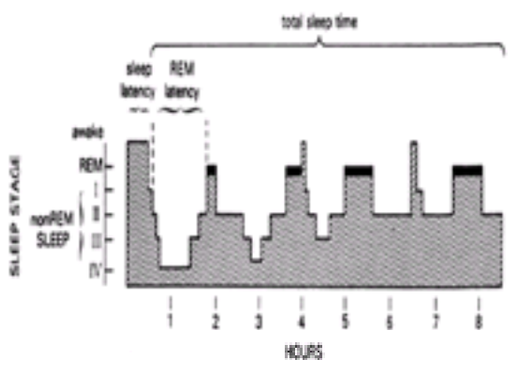

FIGURE 1.19-1 Sleep pattern in a young healthy subject with power spectral analysis of low-frequency band across the night.

Sleep normally begins with *stage 1*, a brief transitional phase, before progressing successively into stages 2 through 4. *Stage 2 sleep* is defined by the presence of sleep spindles and K-complexes in the EEG. *Stage 3* and *Stage 4 sleep*, or *delta sleep* are defined by the presence of delta waves in the EEG, a moderate (20 to 50 percent) and large (>50 percent) proportion of an epoch (usually 30 seconds long) of sleep, respectively.

REM sleep is characterized by an activated EEG, loss of tone in the major antigravity muscles, and periodic bursts of rapid eye movements. The REM latency is usually about 70 to 100 minutes in normal subjects but may be significantly shortened in some patients with depressive disorder, eating disorders, borderline personality disorder, schizophrenia, alcohol use disorder, or other psychiatric disorders. Thereafter, NREM and REM sleep oscillate with a cycle length of roughly 90 to 100 minutes. When it does occur, delta sleep is highest in the first NREM period of the night and declines with each successive NREM period. Most of the delta sleep occurs in the first half of the night and most of the REM sleep occurs in the last half of the night.

REM sleep and NREM sleep differ from each other in many psychological and physiological domains (Table 1.19-2). On a psychological level, REM sleep is associated with dreaming mentation, while NREM sleep is more likely to be associated with abstract thinking. Autonomic functioning is often highly variable during REM sleep; it is usually slow and steady in NREM sleep.

	NREM Sleep	REM Sleep
EEG	Spindles, K complexes, delta waves, synchronized	Low-voltage, mixed frequency, sawtooth waves, activated
EOG	Quiescent or slow movements	Bursts of fast movements
EMG	Partial relaxation	Atonia of antigravity muscles
Intercostal muscles	Partial relaxation	Atonia
Genioglossus	Partial relaxation	Hypotonic
Blood pressure	Decreased, steady	Variable
Heart rate	Decreased, steady	Variable
Cardiac output	Decreased	Decreased
Cerebral glucose metabolism	Decreased	Unchanged or increased
Brain temperature	Decreased	Increased
Respiratory rate	Decreased	Variable
Ventilatory response to CO_2	Intact	Partially impaired
Genitalia	Infrequent tumescence	Tumescence
Mentation	Conceptual, abstract; infrequent dreaming	Perceptual, frequent dreaming
Pathology	Night terrors, somnambulism, panic attacks	Nightmares, REM sleep behavior disorder

Table 1.19-2 Comparison of NREM and REM Sleep

Commonly Used Terms in Sleep Laboratory Studies Many different measurements may be made of sleep, sleepiness, wakefulness, physiology, or performance in a sleep laboratory, depending on the purpose of evaluation. In order to measure the stages of sleep, the subject is usually monitored by electrodes that are attached to the scalp for recording EEG, near the eyes to measure eye movements *(electrooculogram* [EOG]), and on the chin (submental) muscles to record muscle tone *(electromyogram* [EMG]). In the clinical evaluation of a patient suspected of having specific clinical sleep disorder, recordings may be made of nasal and oral air flow by means of thermistors, respiratory movements of the chest and abdomen, electrocardiogram, oxygen saturation, core body temperature (often via the rectal route or the tympanic membrane), muscle twitches of the legs (usually the anterior tibial muscles), penile erections, or esophageal pH (Table 1.19-1).

Although most sleep laboratories around the world use a common set of definitions for sleep stages as defined by Allan Rechtschaffen and Anthony Kales, no commonly accepted definitions exist for some specific measures. For example, depending on the laboratory, sleep onset is defined as the first epoch of stage 1, of stage 2, of REM sleep, or of persistent sleep (e.g., 9 minutes of sleep within a 10-minute period). Likewise, the first REM period may be defined as either the first epoch of REM sleep or the first period of persistent REM sleep (i.e., 3 minutes). REM latency may either include or exclude wakefulness between sleep onset and the first REM period. These variations in definitions do not usually make an appreciable difference, but clinicians and researchers need to be aware of these distinctions when comparing results and reports from different laboratories.

Computer-Assisted Analysis of Sleep Records Although visual scoring is still the standard for the identification of specific sleep stages, considerable progress has been made with computer scoring of sleep stages. Some commercial companies now market hardware and software designed to replace or assist the sleep technician, and further progress can be expected in the coming years. However, there is no generally accepted algorithm of automatic scoring reliable enough to replace visual scoring.

For research purposes, two methods of quantitative EEG wave form analysis have received particular attention: *spectral analysis* and *period-amplitude analysis*. Power spectral analysis uses fast Fourier transformation to compute cumulative or average power (uV^2) of specific frequency bands (Fig. 1.19-1). Period-amplitude analysis computes the peak amplitude of each EEG wave and the time interval between zero (baseline) crossings. It is particularly useful in estimating the amplitude and temporal distribution of transient phenomena in the EEG, such as delta waves, K complexes, and sleep spindles, but it may miss frequency components of the EEG that fail to produce zero crossing; these may be estimated by the calculation of derivatives. Both methods measure slow-wave activity well. Power spectral analysis has also been used to map the topographical distribution of specific frequency bands throughout the progression of sleep. It has, for example been shown that the spatial distribution of sleep spindles and delta waves changes from the beginning to the end of the night.

Measures of Daytime Sleepiness Excessive sleepiness is a common complaint in clinical practice and may be a sign of treatable conditions, such as chronic partial sleep deprivation, certain sleep disorders (sleep apnea, narcolepsy, frequently interrupted sleep, sleep drunkenness), circadian rhythm disorders (phase advance and phase delay syndromes, or a consequence of shift work), certain medical conditions, and the effect of sedating medications. It should be distinguished from fatigue, which often is experienced as a tiredness and a desire to sleep but does not usually lead to sleep. The typical patient with insomnia, for example, complains about fatigue but has difficulty initiating and maintaining sleep. In contrast, the patient with true excessive daytime sleepiness readily and unwillingly falls asleep quickly at inopportune times and places, such as when driving, working, or meeting with friends. Persistent pathological sleepiness is clinically serious and may be associated with significant morbidity and mortality, including poor work performance, accidents, poor judgment, and medico-legal problems for clinicians, patients, and employers who fail to deal with these difficulties appropriately.

Daytime sleepiness can be assessed by either subjective or objective criteria. During the routine review of symptoms, clinicians should inquire about pathological daytime sleepiness, both from the patient and others who have had an opportunity to observe the patient's levels of alertness and sleepiness. For a more detailed evaluation across the day, clinicians can give the patient the Stanford Sleepiness Scale, a seven-point subjective scale ranging from 1 (maximum alertness) to 7 (fighting to stay awake); it can be filled out by the patient, for example, each hour during the day to estimate subjective sleepiness and alertness. Alternatively, for an estimate of average sleep propensity, the Epworth Sleepiness Scale is frequently used.

Of the objective laboratory-based methods for measuring daytime sleepiness, the Multiple Sleep Latency Test (MSLT) is the best known. Mean MSLT is about 11 to 15 minutes for most healthy adults and about 2 to 4 minutes in patients with sleep apnea or narcolepsy. A second objective method for assessing alertness is the Maintenance of Wakefulness Test (MWT), in which the subject is asked to maintain wakefulness during 4 or 5 periods while lying in bed in the dark. Again, the latency to sleep is measured. Normal subjects can usually maintain wakefulness for about 18 minutes before falling asleep. Based on the findings that power in the theta and alpha range of the waking EEG increases progressively during sustained wakefulness, this measure has been proposed as an objective estimate of sleepiness. However, the clinical merit of theta and alpha power assessment needs to be evaluated further.

Circadian and Homeostatic Processes The daily sleep-wake cycle is an example of a *circadian rhythm* (from the Latin *circa*, meaning “about,” and *dies*, meaning “day”). Circadian rhythms are endogenously regulated by a biological clock located in the suprachiasmatic nuclei of the anterior hypothalamus, which in turn is synchronized with the environment by visual or other, nonphotic time clues (*Zeitgebers*, meaning “time givers”). For example, if humans and animals are kept in an environment devoid of time clues, the period length of the sleep-wake cycle and other biological rhythms, such as core body temperature or cortisol, increases from about 24 hours to about 24.2 hours; this condition is called free-running. If, however, a 24-hour light-dark cycle of sufficient amplitude is imposed on this environment, subjects revert to a 24-hour sleep-wake cycle, that is, the subjects and their rhythms are entrained to the 24-hour day.

Sleep and wakefulness are strongly influenced by two separate processes: (1) an endogenous biological clock, which drives the circadian rhythm of the propensity for sleep and the characteristics of sleep across the 24-hour day; and (2) a homeostatic process that increases sleep propensity the longer the period of wakefulness prior to sleep. There is evidence that the duration of wakefulness prior to sleep onset is directly related to the amount of delta activity during NREM sleep, and that this homeostatic process during wakefulness increases in a monotonic function and decays in a negative exponential function when sleep is initiated. It is postulated that the amount of delta activity during NREM sleep is a measure of the sleep pressure accumulated during prior wakefulness, and this dissipates during sleep. The interaction between the circadian process C and the homeostatic process S have been formalized by Alexander Borbely and colleagues in the *two-process-model* of sleep regulation. Because these two processes constantly interact, it is difficult to study the contribution of each of these factors to sleep separately. An elegant method to disentangle the influence of process C and S is the *forced desynchrony* protocol. Thus, if subjects are subjected to a 28-hour light-dark cycle, the sleep-wake period is too long to synchronize with the internal clock, which has a cycle length of about 24.2 hours. This experimental situation leads to a desynchronization of circadian and homeostatic factors, and permits the study of sleep and wakefulness at different circadian phases. Such studies have supported the hypothesis that slow-wave activity is mainly driven through homeostatic processes whereas REM sleep as well as sleep spindles are driven by the circadian system. Humans appear to have two peaks of daytime sleepiness. The first, and obvious one, is at night in the normally entrained individual; the second is in midafternoon (“siesta hour”). Not surprisingly, automobile accidents, in which the driver falls asleep at the wheel, peak during the last half of the night and in the midafternoon. If the circadian temperature curve is used to index the phase position of the circadian pacemaker, it is seen that the major period of sleepiness occurs near the nadir of temperature, that is, at about 3 to 5 AM in normal circumstances. Both the duration of sleep and the type of sleep are strongly influenced by the phase position of sleep onset. People tend to awaken on the rising limb of the temperature curve; thus, sleep tends to be longest when it starts near the peak of the temperature curve. Furthermore, REM sleep is most likely to occur near the temperature nadir; thus, REM latency is shorter and REM time higher during morning naps compared with afternoon or evening naps.

The 24-hour sleep-wake rhythm is driven by the suprachiasmatic nuclei synchronized to the environmental light-dark cycle by Zeitgebers. Lesions of the

suprachiasmatic nuclei in animals result in arrhythmic rest-activity patterns, which no longer follow a circadian rhythm but are distributed in numerous short bouts of sleep and wakefulness during the 24-hour period. Despite the dramatic alteration of the temporal organization of sleep and wakefulness, the total amount of sleep and wakefulness in 24 hours remains fairly constant in the suprachiasmatic nuclei-lesioned animal. Indeed, a suprachiasmatic nuclei-lesioned animal deprived of sleep shows the normal compensatory increase in sleep during the recovery period, thereby demonstrating that the homeostatic and circadian processes can be separated. Although the major circadian pacemaker in mammals is the suprachiasmatic nuclei, a second circadian clock may be located in the retina. Cultured retina cells from wild-type golden hamsters exhibit a circadian 24-hour rhythm of melatonin synthesis, and the rhythm of melatonin production by retina cells of *tau*-mutant hamsters, which have a 20-hour rhythm, has a period of 20 hours. This suggests that the period of the circadian pacemakers is genetically determined.

A number of sleep disorders specifically involve disturbances of the circadian sleep-wake system: jet lag, shift-work schedules, phase-advanced sleep (in which the patient is an extreme lark), phase-delayed sleep (in which the patient appears to be an extreme night owl), and non–24-hour syndrome (in which the patient appears to free-run, with a period length of about 25 hours between each sleep onset). A subjective scale for estimating morningness and eveningness is available for clinical and research purposes.

Because the endogenous period in humans is slightly longer than 24 hours, it is easier to shift sleep-wake rhythms to later rather than earlier. For example, most individuals cope with jet lag more easily when traveling west than when traveling east. It usually takes about one day to accommodate for each time zone traveled when moving in an easterly direction, somewhat faster in a westerly direction. In addition, shift workers seem to feel better and perform better when going from a day to an evening to a night shift *(forward shifting)* than from a day to a night to an evening shift *(backward shifting)*. Appropriate administration of bright light and darkness may ameliorate jet lag and shift-work problems by hastening the resynchronization of the endogenous clock modulating sleep-wakefulness, and controlling temperature and other psychobiological rhythms. For example, exposure to bright light in the evening and darkness in the morning may help with jet lag when traveling westward or shifting from an evening to a night schedule of work. In addition, the administration of exogenous melatonin may shift the phase position of the clock and offers hope for treating jet lag, shift work, and other circadian sleep-wake disorders in which rapid resynchronization of the internal clock with the sleep-wake cycle is required. At this time, however, neither the efficacy nor the safety of melatonin administration for more than very short-term use has been demonstrated.

Free-running in the natural environment has been implicated in two clinical sleep disorders. First, in certain blind persons the circadian system free-runs even though the individual tries to remain on a synchronized sleep-wake pattern. The drifting phase position of the circadian system is shown by the free-running pattern of melatonin secretion (i.e., the onset of melatonin appears approximately 45 minutes later each subjective day). As would be expected, these patients have a cyclic pattern of sleep disturbance with a period of about 3 to 4 weeks as the circadian system goes in and out of synchronization with the environment. Second, patients with the non–24 hour-day syndrome typically go to bed and arise about 30 to 45 minutes later each day (i.e., they are free-running with a sleep-wake cycle length of about 24.5 to 24.75 hours even though they live in the natural environment).

The flexibility of the circadian system appears to diminish with age in most persons. Thus, the elderly usually suffer more from jet lag and shift-work schedules than do the young. In addition, the endogenous circadian pacemaker tends to be phase advanced and of lower amplitude in the elderly compared with the young. For example, teenagers typically like to go to bed around 2 AM and may sleep solidly until noon whereas grandparents retire by 9 PM, arise around 6 AM after a night of shallow sleep, and nap in the midafternoon.

The cellular mechanisms eventually yielding clockwise oscillations in the suprachiasmatic nuclei are currently under intensive investigation. Modern molecular genetic techniques identified genetic loci that modulate the internal pacemaker. In the fruit fly *(Drosophila)*, the genes *period (per)* and *timeless (tim)*, appear to be critical for circadian rhythm generation. The interaction of their products PER and TIM appears to be essential for a 24 hours cycle. Perhaps coupled together, TIM and PER in the cytoplasma are involved in an negative autoregulatory feedback loop with their own genes. High levels turn the genes off; with low levels messenger ribonucleic acid (mRNA) is produced again. In the morning, *tim* and *per* are turned on and mRNA is produced. In the evening, on the other hand, high levels of TIM and PER shut off *tim* and *per*. This autoregulatory feedback loop is an important part of the circadian machinery, and lends new and exciting insight into the way light may reset the clock. For instance, TIM is degraded by light: if light exposure occurs during the early part of the night (i.e., during the production phase of TIM and PER), the level of these proteins is decreased and the phase of the cycle is delayed. On the other hand, if light exposure occurs later in the night (i.e., when TIM and PER levels are falling), these natural processes are accelerated and the clock is advance phased. These interactions are pertinent if one considers that light is the major external clue for resetting the clock, mediated through the *circadian visual system* involving projections from the retina to the suprachiasmatic nuclei via the retinohypothalamic tract, and indirectly to the intergeniculate leaflet and via the geniculohypothalamic tract to the suprachiasmatic nuclei. In mammals, mutant mice are important tools to study the internal clockwork; the golden hamster has a single-gene *(tau)* mutation leading to a reduced period (t) of 20 hours. Other genes that affect period are *clock* and *wheels*, which are located on chromosome 5 and 4, respectively.

PHYLOGENY AND ONTOGENY OF SLEEP

Phylogeny Nearly all animals, even single-cell organisms and plants, have a rest-activity cycle, but circadian timing, amount, and type of sleep varies dramatically between species. A few species, such as the sloth, sleep as much as 20 hours per day; others, such as the shrew, may not sleep at all. Some animals, including humans, are diurnal and sleep predominantly at night. Other animals, such as the laboratory rat, are nocturnal and sleep predominantly in the light. Still other species, such as the cat, are crepuscular and are active predominantly at dawn and dusk.

Among the higher animals, slow-wave sleep is present in birds, mammals, and some reptiles, suggesting an obligatory role for this sleep state. REM sleep occurs in all marsupials and birds as well as in some placental mammals and reptiles, but it may be absent in the monotreme echidna, a primitive egg-laying mammal. Thus, NREM sleep may have evolved earlier than REM sleep, the former in association with homeothermy, the latter with viviparity (live bearing). Recent data, however, indicating the occurrence of REM sleep also in the monotreme platypus suggests that REM sleep was present also in the earliest mammals.

Sleep patterns have evolved as surviving species adapted to their environments. For instance, predators that have secure environments (such as lions) sleep more deeply than do prey that sleep in insecure environments (such as rabbits). Marine mammals, which must breathe as they spend their lives in the water, have evolved interesting ways of sleeping. The blind Incus dolphin, which lives in turbid water, has to maintain almost continual activity; therefore, it apparently sleeps in numerous short periods, each averaging about 90 seconds. Some other dolphin species exhibit unihemispheric sleep: while one hemisphere has slow-wave sleep, the other is awake with low-voltage and fast-frequency EEG waves. If one hemisphere is experimentally deprived of NREM sleep, it shows the expected increases in total sleep and delta sleep, which accompany recovery from sleep deprivation of the whole brain. Similarly, some species of seals have also evolved hemispheric sleep so that one side of the body maneuvers the head near the water surface in order to permit the animal to breathe occasionally as the other half of the brain sleeps. Further understanding of the anatomy and chemistry of these and other species will provide insight into the mechanisms and functions of sleep.

Ontogeny Sleep-wake states change dramatically across the lifespan, not only in the amount of sleep but in its ultradian and circadian timing. The full-term newborn human infant may average about 16 hours of sleep per day, of which about 50 percent is REM sleep. The duration of the REM sleep cycle of the infant is relatively short compared with that of the adult, and the sleep-wakefulness cycle is polyphasic, with numerous short bouts of sleep and wakefulness over the 24-hour period. During the first months of life, sleep-wake cycles gradually change as sleep at night and wakefulness by day become consolidated, although napping may continue into childhood. By age 3 or 4, the percentage of REM sleep falls to adult levels of about 20 to 25 percent and remains in this range for the rest of the individual's life. Nevertheless, REM latency tends to decrease and length of the first REM period tends to increase in later life as stages 3 and 4 sleep decline.

The amount of time spent in delta sleep (stages 3 and 4) each night peaks in early adolescence and gradually falls with age until it nearly disappears at about the age of 60. Young adults typically spend about 15 to 20 percent of total sleep time in delta sleep. By the age of 60 or 70, few individuals have any delta-wave activity during sleep. Interestingly, men tend to lose delta sleep at an earlier age than women. The loss of delta sleep results more from a reduction in amplitude than from fewer slow (0.5-2 Hz) waves. Some researchers have speculated that this might indicate accelerated aging in men.

Sleep in the Elderly After the age of 65, about one third of women and one fifth of men report that they take over 30 minutes to fall asleep. Wake time after sleep onset (WASO) tends to increase with age, perhaps because of the greater incidence of sleep-related breathing disorders (i.e., mild apnea) and nocturnal myoclonus. Using the Apnea Index of £5 apneic episodes per hour as a cut-off criterion, prevalence rates range from 27 to 75 percent for older men and from 0 to 32 percent for older women. In general, the severity of apnea in these older persons is mild compared with that seen in patients with clinical sleep apnea. Periodic limb movements during sleep are also common in the elderly, with prevalence rates ranging from 25 to 60 percent in various studies of the healthy elderly. Individuals with periodic limb movements are reported to sleep about an hour less per night than controls without periodic limb movements. Perhaps as a result, napping also increases with age, although it rarely accounts for a large proportion of total sleep time in healthy individuals.

Average daily total sleep time actually increases slightly after the age of about 65. Greater numbers of elderly individuals fall into either long-sleeping (>9 hours) or short-sleeping (<5 hours) subgroups. It is noteworthy that death rates are higher in both long-sleeping and excessively short-sleeping individuals. The reasons for that are still unknown, although there has been speculation that sleep apnea might contribute to increased mortality in the long-sleeping group.

Although the incidence of insomnia and certain other sleep-wake disorders tends to increase with age, clinicians should not assume that age explains these common complaints. Rather, the clinician must search for underlying conditions that can be treated, such as medical, neurological, psychiatric, situational, pharmacological, or circadian factors.

NEUROBIOLOGICAL BASIS OF SLEEP

Neuroanatomy of Sleep-Wakefulness The neuroanatomy and neurophysiology of sleep and wakefulness remain incompletely understood. Behavioral and cortical EEG activation depend upon the reticular activating system, arising within the brainstem, and more rostral thalamic, hypothalamic, limbic, and cortical systems. Moreover, the isolated brainstem itself can generate components of REM sleep. NREM sleep, on the other hand, is controlled by widespread anatomic areas. In any case, no specific neuroanatomical sleep center has yet been identified with certainty for the entire constellation of REM or NREM sleep.

Reticular Activating System Components of the ascending reticular activating system (ARAS) are critically important for the generation and maintenance of waking states or arousal (Fig. 1.19-2). They reside in the oral pontine and mesencephalic tegmentum, including both noradrenergic cell bodies, such as the locus ceruleus, and cholinergic cell bodies, such as the pedunculopontine tegmental and lateral dorsal tegmental nuclei. The locus ceruleus neurons project rostrally through the mesencephalic tegmentum directly and diffusely to cortical areas and subcortical way stations; in addition, noradrenergic neurons project caudally into the brainstem and spinal cord. The pedunculopontine tegmental and lateral dorsal tegmental form the largest collection of cholinergic cells in the pontine tegmentum. They innervate the thalamus, hippocampus, hypothalamus, and cingulate cortex. In addition, histaminergic neurons in posterior hypothalamus project to the cortex and maintain arousal. Glutaminergic neurons in subcortical and cortical structures may also play an important role in wakefulness and arousal.

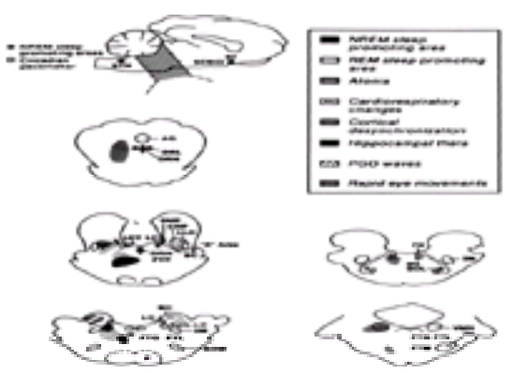

FIGURE 1.19-2 Localization of NREM and REM sleep promoting areas in the brain.

Anatomical Control of REM Sleep At least five anatomical sites have been implicated in the generation of NREM sleep: the basal forebrain area, thalamus, hypothalamus, dorsal raphe nucleus, and nucleus tractus solitarius of the medulla. For example, lesions of the preoptic basal forebrain area produce hyposomnia lasting 4 to 6 weeks in rats and cats whereas electrical stimulation and local warming of this region elicits both EEG and behavioral signs of sleep. In addition, some noncholinergic neurons in the basal forebrain discharge selectively during NREM sleep. The thalamus in general and the reticular nucleus of the thalamus in particular appear to play an important role in the generation of cortical sleep spindles (12 to 14 Hz) and delta waves (0.5-3 Hz, 75 μV or greater in humans) during NREM sleep. Recently Mircea Steriade and colleagues hypothesized that thalamocortical cells are hyperpolarized by corticothalamic cells and are depolarized by cholinergic input from basal forebrain and the lateral dorsal tegmental and pedunculopontine tegmental, as well as through noradrenergic, serotonergic, and excitatory amino-acid input. The thalamocortical cells, which drive and synchronize the ensembles of cortical cells to produce the major EEG rhythms, change their rhythms from spindle frequencies to delta frequencies as they hyperpolarize. This theory also explains the fact that immediately prior to and during REM sleep, when cholinergic activity in the lateral dorsal tegmental and pedunculopontine tegmental is high, the EEG becomes gradually desynchronized. While histaminergic cells in posterior hypothalamus maintain arousal, ventrolateral preoptic neurons in the anterior hypothalamus may be involved in the induction of slow-wave sleep. The dorsal tegmental, the origin of the serotonergic projections of the brain, may be involved in the induction of sleep, at least insofar as either selective lesions or depletion of serotonin induce a dramatic insomnia lasting several days. Finally, the nucleus tractus solitarius was implicated by experiments in which medullary anesthesia or cooling of the fourth ventricular floor caused EEG activation. Low-frequency stimulation of this region produced EEG synchronization and behavioral sleep while cells in this region increase their discharge rate during NREM sleep. These regions all appear to facilitate sleep; however, none have been found to be essential.

In terms of pathophysiological concepts of mood disorders and the proposed mechanisms of action of antidepressant treatment it is pertinent to emphasize the role of serotonin on sleep regulation. Monosynaptic projections from the dorsal raphe nucleus to the cholinergic-cholinoceptive lateral dorsal tegmental and pendunculopontine tegmental produce a hyperpolarization of these so-called REM-on cells through serotonin (5-hydroxytryptamine [5-HT]) subtype 1A (5-HT $_{1A}$) receptors and finally inhibit REM sleep. On the other hand, activation of the self-inhibitory 5-HT $_{1A}$ autoreceptors in the dorsal raphe nucleus lead to a disinhibition of REM sleep. While the 5-HT$_{1A}$ agonist ipsapirone inhibits REM sleep and enhances NREM sleep EEG power in the lower frequencies, pindolol (Visken), a b-adrenergic receptor antagonist with 5-HT$_{1A}$ antagonistic properties produces a reduction of both REM sleep and of EEG power in the lower EEG frequencies. In combination these findings suggest that pindolol preferentially blocks 5-HT $_{1A}$ presynaptic, rather than 5-HT$_{1A}$ postsynaptic receptors, which is particularly pertinent for the proposed mechanism of action of how pindolol augmentation of a selective serotonin reuptake inhibitor (SSRI) might work.

Neurological lesions associated with pathological somnolence, stupor, or coma have been noted in midbrain and diencephalon, the oral pontine and midbrain tegmentum, and the posterior hypothalamus and subthalamus. A recently described syndrome of “fatal insomnia” is associated with degeneration of the thalamus, although it is doubtful that the insomnia resulted directly from degeneration of the thalamus. Data in the prion knock-out mice suggest that the sleep abnormalities in fatal familial insomnia are related to the function of this membrane surface protein. Genetic factors also account for narcolepsy; there is evidence of an involvement of human lymphocyte antigen (HLA) DQA1*0102 and DQB1*0602 in narcolepsy as candidate susceptibility genes in the HLA class II region. Much effort has been made to identify endogenously produced “sleep substances” that induce or promote sleep. Generally, brain extracts from sleep-deprived animals have been screened for such substances. Although none of the candidates have satisfied the criteria required for such a factor, a recently described membrane lipid (cerebrodiene) that accumulates during wakefulness may be involved in the induction and maintenance of sleep. Another interesting somnogenic candidate is adenosine, which may increase in brain tissue with increasing duration of wakefulness; infusion of adenosine into basal forebrain promotes delta sleep.

Anatomic Sites for Control of REM Sleep Transection and lesion studies suggest that the midpons is necessary for REM sleep. Transection at the midpontine level produces a preparation with local signs of REM sleep (rapid eye movements, muscle atonia) caudal to the transection, whereas transection at the pontomedullary junction produces a preparation that does not show REM sleep signs in the medulla, but does show signs in the forebrain. Electrolytic and kainic acid lesions of the oral and caudal pontine tegmentum, including the lateral dorsal tegmental and pendunculopontine tegmental, eliminate or markedly reduce the amount of REM sleep. After suffering a shrapnel wound with a sliver lodging in the pontine mesencephalic region, a soldier apparently ceased to have REM sleep. Despite the documented loss of REM sleep over a period of several years, the individual has no specific problems related to the loss of REM sleep.

As proposed in the reciprocal interaction model of REM sleep, cholinergic REM-on neurons in lateral dorsal tegmental and pendunculopontine tegmental orchestrate the various events of REM sleep. They are inhibited by REM-off neurons within locus ceruleus and DRN. The REM-off cells decrease their discharge activity preceding a REM sleep episode and are at a very low level of activity or are silent during REM sleep, thus disinhibiting the spontaneously active cholinergic neurons in lateral dorsal tegmental and pendunculopontine tegmental. Serotonergic terminals, acting on 5-HT $_{1A}$ receptors on cholinergic cells, apparently hyperpolarize cholinergic burst cells, which may be responsible for pontine-geniculate-occipital spikes, large monophasic electrical waves appearing just before and during REM sleep in cats and, presumably, in many other species. Those data are consistent with the hypothesis that serotonergic neurons within DRN inhibit pontine-geniculate-occipital spikes. For example, they cease firing just before the appearance of pontine-geniculate-occipital spikes. Noradrenergic neurons within locus ceruleus may also inhibit REM sleep: not only are there REM-off cells within locus, but administration of metyrosine (Demser), an inhibitor of tyrosine

hydroxylase, the rate-limiting enzymatic step in the synthesis of catecholamines, shortens REM latency and increases REM sleep time in humans and animals. Consistently, in 1988 it has been shown that a depleting serotonin from the brain by using the rapid tryptophan depletion challenge (i.e., a short-term) diet with high loading of amino acids, but without the essential amino acid tryptophan) leads to a disinhibition of REM sleep. This finding might be particularly pertinent for the association of depression with short REM latency and increased REM sleep.

Muscle Atonia of REM Sleep The muscle atonia of REM sleep is abolished by bilateral lesions of the pontine reticular region, just lateral to the locus ceruleus (peri-locus ceruleus a-region), and its descending pathway to the bulbar reticular formation. Animals can ambulate during REM sleep and may show oneiric behavior, including locomotion, attack, and flywatching, as if the animal were acting out a dream. Thus, the muscle atonia of REM sleep is apparently mediated through nonmonoaminergic neurons of the peri-locus ceruleus a-region, which project via tegmentoreticular tract projections to the bulbar magnocellular field and excite neurons in the reticular zone. This area in turn, via projections in the ventrolateral reticulospinal tract, may induce muscle atonia by hyperpolarizing spinal motoneurons. Those mechanisms for atonia may be clinically relevant to two conditions characterized by abnormalities of muscle atonia: (1) *narcolepsy*, in which patients experience sudden brief episodes of *cataplexy*, or loss of muscle tone while awake; and (2) *REM sleep behavior disorder*, in which human subjects maintain muscle tone during REM sleep and may act out their dreams.

Biochemistry of Sleep and Wakefulness The biochemistry of sleep also remains incompletely understood, although dramatic progress has been made in recent years (Table 1.19-3).

Substance	Possible Roles in Regulation of Sleep-Wakefulness
Serotonin	L-tryptophan has hypnotic effects, increases delta sleep. Serotonergic neurons in DRN cease firing in REM sleep and may inhibit cholinergic neurons in LDT-PPT. PGO waves, and REM sleep
Norepinephrine	Noradrenergic neurons in LC cease firing in REM sleep and may inhibit REM sleep. Arousal
Acetylcholine	Cholinergic neurons in dorsal tegmentum orchestrate REM sleep, and together with basal forebrain inhibit cortical EEG synchronization through influence on thalamus
Dopamine	Mediates alerting effects of amphetamine and cocaine and sedating effects of antipsychotics. Sleepiness of narcolepsy may be related to decreased dopamine turnover
γ-aminobutyric acid (GABA)	Hypnotic and other effects of benzodiazepines may be mediated through enhancement of GABA
Adenosine	Appears to promote sleep. Alerting effects of caffeine may be mediated by blockade of adenosine receptors
Interleukins and other immune modulators	Interleukin promote slow-wave sleep in animals, and immune modulators may be increased in plasma at sleep onset in normal controls. NREM sleep measures may correlate with natural killer cell activity in humans
Prostaglandins	PGD_2 and PGE_2 increase sleep and wakefulness, respectively, in animals
Endogenous sleep factors	Putative hypnotoxins include delta-sleep-inducing substances—peptide (DSIP), uridine, arginine vasotocin, muramyl peptides, and others

Table 1.19-3 Neurotransmitters and Neuromodulators That May Regulate Sleep-Wake States

Cortical and behavioral arousal mechanisms involve dopaminergic, noradrenergic, histaminergic, glutaminergic, and cholinergic neurons. Data implicating the latter systems have been briefly mentioned already. In addition, antihistamines and b-adrenergic receptor antagonists, and, to some extent, anticholinergic agents have sedative effects. Neurotransmitter regulation of NREM sleep and REM sleep is complex.

Consistent with the concept that cholinergic neurons mediate cortical arousal and REM sleep, administration of carbachol, neostigmine, and other cholinergic agonists within the medial pontine reticular formation induce REM sleep in cats. Recently, a “hot spot” within the peribrachial region of the cat has been described where a single injection of a cholinergic agonist will, after the delay of a day, significantly increase pontine-geniculate-occipital activity and REM sleep for over 10 days.

Although there is a large database on how aminergic projections from the DRN and the locus ceruleus turn off REM sleep by hyperpolarization of cholinergic REM-on cells, how the REM-off neurons are inhibited is just beginning to be understood. It appears, that GABAergic, glycinergic, and nicotinic neurons and receptors within these monoaminergic nuclei play an important role.

It is of particular interest to mental health professionals that REM sleep can be promoted in humans by cholinergic agonists, such as physostigmine, arecoline, pilocarpine, and RS 86 (an experimental muscarinic agonist). Physostigmine-induced REM periods appear to be completely normal in their EEG, EOG, and EMG patterns, and, moreover, are associated with dreams that appear to be indistinguishable from those collected by arousing subjects from spontaneous REM periods. Furthermore, cholinergic agonists appear to substantially affect NREM sleep, even causing a decrease in EEG power in the delta frequencies.

How is it that a few molecules of a cholinergic agonist can induce or release the fascinating process we know of as dreaming? If these turn on the stage lights of the theater of the mind, who or what determines which play will be enacted? It is possible that neuroimaging studies during sleep and, particularly during REM sleep, might reveal new insights into the neuronal basis of dreaming. Based on emission tomography (PET) studies it has been suggested that the amygdala cingulate gyrus and other areas associated with emotions are activated during REM sleep.

Brain Metabolism Consistent with the popular hypothesis that sleep is a time of rest and restoration, whole body and brain metabolic rates decrease during NREM sleep compared with wakeful states and during REM sleep. However, metabolic rate is about the same or slightly increased during REM sleep as compared with wakefulness. Local cerebral metabolism has been studied in humans utilizing the ^{18}F-fluoro-deoxyglucose method with PET or by measuring cerebral blood flow. No specific sleep center has yet been identified by those techniques.

The relationships between environmental and core body temperature, metabolism, and sleep have been investigated. Some investigators hypothesize that sleep and wakefulness have homeostatic effects on body and brain temperature. For example, it is known that the onset of NREM sleep tends to depress metabolic rate in both the brain and the body, which may explain why delta sleep is increased after subjects are warmed in a hot tub or sleep in a moderately warm environment. Increased wakefulness in relatively cold environments may be explained by the body's efforts to warm itself. That the nocturnal core body temperature of depressed patients drops below that of normal controls is a related finding.

Neuroendocrinology and Neuroimmunology Some hormones and immune modulators are affected by the sleep-wake cycle in that the time course of their plasma levels may be either circadian or sleep related. For example, plasma thyroid-stimulating hormone (TSH) rises during the evening hours and is inhibited by sleep onset. Growth hormone, on the other hand, is stimulated by sleep onset, particularly in association with delta sleep, whereas prolactin levels are increased throughout most of the sleep period. In adolescents, luteinizing hormone is also increased during sleep. While cortisol secretion increases towards the end of the sleep period, this increase is primarily driven by an endogenous circadian rhythm whether or not sleep occurs. In contrast, sleep onset actually reduces secretion of adrenocorticotropic hormone (ACTH) and cortisol. Melatonin is released at night in the dark in entrained subjects, but its release also occurs independently of sleep and wakefulness. In depression, the composition of hormonal secretion during nocturnal sleep appears to be altered, including an increased level of cortisol and a blunted sleep–onset-associated peak of growth hormone and prolactin concentrations.

In keeping with traditional advice to “get a lot of sleep when you are sick,” sleep and the immune system apparently have positive effects on each other. Certain immune modulators such as interleukins and tumor necrosis factor may promote sleep. Sleep deprivation may depress immune measures in man and animals. The interrelationships between the brain and the immune system are exciting areas of research; particularly interesting is the finding that sleep architecture and continuity changes are possibly among the first biological signs to change after seroconversion in human immunodeficiency virus (HIV)-infected subjects.

DREAMING AND OTHER PSYCHOLOGICAL EXPERIENCES DURING SLEEP

Dreams are reported about 70 to 80 percent of the time when subjects are awakened from REM sleep. Nevertheless, dreaming is not confined to REM sleep only; it is reported during awakenings from NREM sleep about 10 to 20 percent of the time, and is reported especially frequently after a few minutes of stage 1 sleep at sleep onset. Some type of mental activity appears to occur much of the time throughout sleep. Indeed, it may be the inability to “turn off the mind” during sleep that leaves many ruminative insomniac patients feeling as if they never slept at all.

By most definitions, dreams involve a visual experience with a plot that evolves over time. Dreams are experienced, if not always remembered, in full color and appear to take place in real time. Nevertheless, about one third of dreams often involve rapid shifts in time and place from one scene to the next; about 30 percent of dreams entail normally impossible acts such as flying or talking with animals. Despite the often bizarre experiences of dreaming, most dreamers seem to accept these

experiences uncritically as real. Some persons, however, so-called “lucid dreamers” are consciously aware that they are dreaming and can change the “script” of their dream as it takes place.

Although averaging about 10 hours of REM sleep a week, most people spontaneously remember only about one or two dreams during that time, and the variation in dream recall between people is impressive. Many people say they rarely if ever dream; others say they dream nearly every night. Why such variation exists is not known. Dream recall does not appear to be related to the amount of REM sleep, but may be correlated with the number or duration of waking episodes during the night. The memory for dreams is evidently increased if subjects are awakened directly from REM sleep and have an opportunity to think consciously about their dreams. The memory for dreams is strangely fleeting: nearly everyone has had the experience of awakening from a dream, thinking about it, and then later failing to recall the content despite remembering having had a dream. While it is plausible to invoke psychological defenses—repression, suppression, and denial—for forgetting dreams, such explanations seem forced for dreams that are inherently interesting, bizarre, exciting, or psychologically benign.

The classic nightmare occurs during REM sleep, but frightening experiences may also arise out of NREM sleep. The night terror or incubus, for example, typically occurs during delta sleep in the first or second NREM period. This disorder is characterized by a sudden attack of fear, screaming, shortness of breath, and tachycardia; it may last for several minutes and may be unresponsive to attempts to console the victim. In contrast to a nightmare, the patient rarely recalls specific content or visual images when awakened from a night terror and often has total amnesia for the entire episode the next day.

Nearly half of all patients with panic disorder experience classic panic attacks during sleep, typically in NREM sleep during transitions between stages 2 and 3. While patients with posttraumatic stress disorder often experience terrifying nightmares and flashbacks, some of these events apparently occur during NREM sleep and are not typical REM sleep nightmares. This may explain why monoamine oxidase inhibitors, which can totally eliminate REM sleep, have had only modest success in treating these patients.

Interesting new approaches, based on functional brain imaging, the emerging neurophysiology of sleep, and cognitive neuroscience are now available to study the phenomenology and neuroscientific basis of dreaming. Functional imaging is certainly an approach that has the potential to provide more insight into brain areas activated during sleep, and in particular, during dreaming and REM sleep. So far, PET studies have yielded the most precise neuroanatomical information. Future research will certainly profit from the improved temporal and spatial resolution of functional magnetic resonance imaging (fMRI) techniques. Unfortunately, it has not been possible yet to coregister EEG and fMRI because of mutual interference.

SLEEP DEPRIVATION

Historical evidence suggests that the general population of the United States sleeps less now than it did at the beginning of this century. With the invention of the electric light bulb and central heating, the move from rural to suburban and urban life styles, and the growth of international commerce and communication, periods of work and play have expanded to fill the last temporal frontier, nighttime. Does this mean that the general population, or at least large segments of it, is chronically sleep deprived? There is no doubt that many individuals are pathologically sleepy, either acutely or chronically because of partial sleep deprivation or because of clinical disorders. Excessive sleepiness is an important public health issue because it is responsible for many car and occupational accidents, and impaired performance, for example, by sleep-deprived physicians. Although the subjective report of insomnia should not be equated with objective sleep loss, prospective epidemiological studies have shown that persistent complaints of insomnia predict the later onset of depressive disorders.

Partial and selective deprivation of sleep and its different stages has been used to investigate the functions of sleep. Unfortunately, this strategy has not forced sleep to reveal all its secrets. For example, volunteers have been kept awake for as long as 12 days. As one might expect, total or partial sleep deprivation makes one very sleepy, but it has not provoked either insanity or physical illness. The major effects were found to be psychobiological, primarily reduced alertness, attention, information processing, motivation, and performance, especially for sustained, boring tasks or for highly complex tasks. Many antidepressant drugs and benzodiazepines selectively reduce REM sleep and delta sleep, respectively, but without harmful effects that are directly attributable to the loss of REM or delta sleep.

In humans, interruption of sleep continuity may be more disturbing than loss of total sleep time. Frequent short breaks in sleep, for example, for as short as 15 seconds every 10 minutes, produces significant daytime sleepiness and dysphoria even though total sleep time is preserved. Such interference occurs in patients with bad coughs, painful arthritis, sleep apnea, or nocturnal myoclonus. Since many patients do not remember these frequent short interruptions, it is useful for the clinician to talk with a bedpartner who can comment on sleep continuity.

In keeping with the homeostatic function of sleep, a rebound in total sleep time or specific sleep stages usually follows deprivation of sleep, or of a specific stage of sleep. After total and partial sleep deprivation, recovery sleep is characterized by short sleep latency and increased sleep efficiency, delta sleep, and total sleep time. Objective and subjective measures of daytime sleepiness are also increased. Similarly, following REM sleep deprivation, REM rebound occurs, including shortening of REM latency and increase of REM sleep time for a few nights. In general, however, recovery from sleep deprivation does not require minute-by-minute compensation for lost sleep.

Overnight sleep deprivation depresses cerebral glucose metabolic rates within specific subcortical areas in human brain during wakefulness, particularly the white matter areas of the thalamus and midbrain. This observation suggests that sleep deprivation depresses general arousal systems. However, selected areas of the brain may be spared when they are engaged in specific tasks at the time of the study, for example, visual areas when the subject has to perform a visual continuous performance task. In other words, sleep deprivation apparently reduces the number of parallel processing operations the brain can carry on simultaneously. This model is consistent with the general observation that sleep-deprived subjects can usually perform relatively well on focused tasks for short periods of time if they are motivated sufficiently. Less important activities or competing duties, however, may be neglected.

In contrast to the previously described effects of relatively short-term sleep deprivation in humans, prolonged sleep deprivation in rats results in death, requiring about 3 weeks of total sleep deprivation or about 5 weeks of selective REM deprivation. In these studies experimental animals were awakened whenever they entered sleep or REM sleep respectively, by turning a disc on which they lived; control animals who lived on the other half of the same disc, however, did not die. The cause of death was apparently related to a significant increase in overall metabolism in the face of falling body temperature and increased food intake; the exact mechanisms remain unknown.

Whether the results of those rat studies are applicable to humans is unknown. For instance, a 72-year-old woman who claimed to have slept only an hour a day since her late teens has been studied. Another case was reported of a man who apparently stopped sleeping altogether for months and responded to treatment with serotonin precursors. Subjects with either brainstem lesions or those taking monoamine oxidase inhibitors (MAOIs), and who had no REM sleep over a period of years have also been studied without apparent difficulties related to the loss of REM sleep.

Thus, several questions about sleep remains unanswered: Why do we sleep? What functions does sleep serve? Various theories have stressed the themes of rest, recovery, or restoration, energy conservation and metabolic modulation, ecological and ethological processes that favor the survival of the species, instinctive behaviors, and enhancement of cognition and memory. Nevertheless, scientific support for these theories is still unavailable.

IMPLICATIONS FOR PSYCHIATRIC DISORDERS

Psychiatrists have long had a strong interest in sleep and dreaming, especially in whether sleep disturbances cause mental disorders or vice versa? In the last century, the famous English neurologist Hughlings Jackson said “Find out about dreams and you will find out about psychosis.” Later, Freud made dream interpretation one of the key elements of psychoanalytic theory and practice. Following the discovery of REM sleep in the early 1950s, psychiatrists and sleep researchers studied the sleep patterns of schizophrenic and depressed patients, testing the hypothesis that hallucinations might represent waking dreams or some derangement of REM sleep, and determining whether or not deprivation of REM sleep or sleep in general would be psychologically deleterious or beneficial.

The new neurobiology of sleep has important implications for depression, some types of alcoholism and schizophrenia, eating disorders, borderline personality disorder, and other clinical conditions associated with short REM latency, increased REM sleep, increased REM sleep density, and loss of stages 3 and 4 sleep.

The emerging concepts of sleep neurophysiology are consistent with the cholinergic-aminergic imbalance hypothesis of mood disorders, which proposes that depression is associated with an increased ratio of central cholinergic to aminergic neurotransmission. The characteristic sleep abnormalities of depression may reflect a relative predominance of cholinergic activity, originating within the lateral dorsal tegmental and pendunculopontine tegmental, in relationship to noradrenergic and serotonergic activity, originating within the locus ceruleus, and the DRN, respectively. Cholinergic projections from dorsal tegmentum or basal forebrain to

thalamus may also suppress delta sleep. Consistent with the role of cholinergic mechanisms, depressed patients are significantly more sensitive, compared with normal controls, to the REM-sleep–inducing effects of muscarinic agonists, such as arecoline or RS 86, but (not of pilocarpine, another selective muscarinic agonist). On the other hand, central depletion of either serotonin or catecholamines shortens REM latency and increases REM sleep. Antidepressant medications presumably reduce REM sleep by either their anticholinergic properties or by enhancing aminergic neurotransmission. Intense and prolonged dreams often accompany abrupt withdrawal from antidepressant drugs, a reflection of a REM rebound following drug-induced REM deprivation.

Short REM latency has been found in some patients with schizophrenia, obsessive-compulsive disorder, alcoholism, eating disorders, narcolepsy, and borderline personality disorder. Both short REM latency and decreased delta sleep appear to be state and trait characteristics of depression. Short REM latency may also be a genetic marker for depressive disorders within families, and an indicator of a poor prognosis in recovered depressed and alcoholic patients. For example, among the unaffected relatives of depressed patients with short REM latency, the presence of short REM latency and deficits of slow wave sleep may be risk factors for the later development of a mood disorder. However, although none of the sleep parameters are specific for mood disorders as has been proposed for a long time, it appears that “multivariate” clusters of sleep changes might be more specific for particular disorders.

Although the morbidity associated with sleep deprivation has been emphasized, the paradoxical finding that total sleep deprivation, partial sleep deprivation (especially in the last half of the night), and selective REM sleep deprivation have antidepressant effects in depressed patients and must not be overlooked; unfortunately, the beneficial effect of total and partial sleep deprivation only lasts until the next sleep period, after which the patient typically awakens depressed again. Hence, sleep may be depressogenic in some patients. There is some evidence suggesting that selective REM sleep deprivation repeated over about 2 weeks, or a continued sleep-phase advance protocol after total sleep deprivation, might result in a sustained antidepressant effect. However, so far the latter treatment modes have scientific rather than practical implications because these protocols are extremely costly and time-consuming.

Many of the following observations suggest that some depressed patients are overaroused, at least in some areas of the brain: the antidepressant effects of sleep deprivation, loss of delta sleep and sleep continuity, and increased core body temperature during sleep. Moreover, a preliminary PET study showed that cerebral glucose metabolism during the first NREM period was significantly elevated in depressed patients compared with normal controls and that depressed patients had some of the abnormalities during sleep that had previously been reported in other patients while awake, such as decreased relative metabolic activity in the anterior cingulate and ventral-medial prefrontal cortex. Consistent with the overarousal hypothesis, depressed patients who responded to sleep deprivation showed significantly elevated metabolic rates within the anterior cingulate gyrus before sleep deprivation as compared with nonresponders and normal controls. After clinical improvement, metabolic activity normalized in responders, but did not change in the two other groups.

If the clinical applications of sleep deprivation therapy in depression remain limited, the experimental and theoretical implications of sleep deprivation in depression are fascinating. In what other ways can depression be “turned” off and on so quickly and predictably as sleep deprivation and recovery sleep?

On a more practical clinical level, as understanding of the basic mechanisms of arousal and sedation, increases the clinician will manage better not only symptoms of insomnia and hypersomnia associated with psychiatric illness but iatrogenic sleep problems often associated with medications. For example, many patients complain of insomnia during treatment with SSRIs. Interestingly, depletion of brain serotonin, with the tryptophan free amino acid drink challenge, actually reverses the poor sleep efficiency in euthymic patients treated with SSRIs. These observations suggest that serotonin has an arousing effect under some circumstances. This arousing effect of serotonin may be mediated by the 5-HT_2 receptor, because SSRIs that antagonize this receptor tend to increase sleep efficiency and sleep quality more than SSRIs without this property.

The clinical wisdom of the centuries, linking sleep to psychiatric disorders, has been reinforced by the revelations of modern neurobiology. Sleep remains one of the “royal roads” to understanding psychiatric disorder, and treating the patients who suffer from them.

SUGGESTED CROSS-REFERENCES

Section 1.13 is devoted to the related issue of chronobiology. Chapter 21 discusses sleep disorders. Section 31.33 includes a discussion of sleep deprivation and sleep delay therapies. Section 51.3b focuses on sleep disorders in the elderly.

SECTION REFERENCES

Benca RM, Obermeyer WH, Thisted RA, Gillin JC: Sleep and psychiatric disorders: A meta-analysis. Arch Gen Psychiatry *49:*651, 1992.

Berger M, Vollmann J, Hohagen F, König A, Lohner H, Voderholzer U, Riemann D: Sleep deprivation combined with consecutive sleep phase advance as a fast-acting therapy in depression: An open pilot trial in medicated and unmedicated patients. Am J Psychiatry *154:*870, 1997.

Bhatti T, Gillin JC, Seifritz E, Moore P, Clark C, Golshan S, Stahl S, Rapaport M, Kelsoe J: Effects of a tryptophan-free amino acid drink challenge on normal human sleep electroencephalography and mood. Biol Psychiatry *43:*52, 1998.

Buchsbaum MS, Gillin JC, Wu J, Haslett E, Sicotte N, Dupont R, Bunney WE Jr: Regional cerebral glucose metabolic rate in human sleep assessed by positron emission tomography. Life Sci *45:*1349, 1989.

Cajochen C, Kräuchi K, Wirz-Justice A: The acute soporific action of daytime melatonin administration: Effects on the EEG during wakefulness and subjective alertness. J Biol Rhythms *12:*636, 1997.

*Daan S, Beersma DGM, Borbély AA: Timing of human sleep: Recovery process gated by a circadian pacemaker. Am J Physiol *246:*R161, 1984.

*Dijk DJ, Czeisler CA: Contribution of the circadian pacemaker and the sleep homeostat to sleep propensity, sleep structure, electroencephalographic slow waves, and sleep spindle activity in humans. J Neurosci *15:*3526, 1995.

Eaton WW, Badawi M, Melton B: Prodromes and precursors: Epidemiologic data for primary prevention of disorders with slow onset. Am J Psychiatry *152:*967, 1995.

Giles DE, Kupfer DJ, Rush AJ, Roffwarg HP: Controlled comparison of electrophysiological sleep in families of probands with unipolar depression. Am J Psychiatry *155:*192, 1998.

Gillin JC, Rapaport M, Erman MK, Winokur A, Albala BJ: A comparison of nefazodone and fluoxetine on mood and on objective, subjective, and clinician-rated measures of sleep in depressed patients: A double-blind, 8-week clinical trial. Clin Psychiatry *58:*185, 1997.

Gillin JC, Smith TL, Irwin M, Butters N, Demodena A, Schuckit M: Increased pressure for rapid eye movement sleep at time of hospital admission predicts relapse in nondepressed patients with primary alcoholism at 3 month follow-up. Arch Gen Psychiatry *51:*189, 1994.

Gottschalk LA, Buchsbaum MS, Gillin JC, Wu JC, Reynolds CA, Herrera DB: Positron emission tomographic studies of the relationship of cerebral glucose metabolism and the magnitude of anxiety and hostility experienced during dreaming and waking. J Neuropsychiatr Clin Neurosci *3:*131, 1991.

Hong CC, Gillin JC, Dow BM, Wu J, Buchsbaum MS: Localized and lateralized cerebral glucose metabolism associated with eye movements during REM sleep and wakefulness: A positron emission tomography (PET) study. Sleep *18:*570, 1995.

Irwin M, Smith TL, Gillin JC: Electroencephalographic sleep and natural killer activity in depressed patients and control subjects. Psychosom Med *54:*10, 1992.

*Kruyger MH, Roth T, Dement WC: *Principles and Practice of Sleep Medicine*. WB Saunders, Philadelphia, 1994.

Lavie P, Pratt H, Scharf B, Peled R, Brown J: Localized pontine lesion: Nearly total absence of REM sleep. Neurology *34:*118, 1984.

Lerner RA, Siuzdak G, Prospero-Garcia O, Henriksen SJ, Boger DL, Cravatt BF: Cerebrodiene: A brain lipid isolated from sleep-deprived cats. Proc Natl Acad Sci USA *91:*9505, 1994.

Luebke JI, Greene RW, Semba K, Kamondi A, McCarley RW, Reiner PB: Serotonin hyperpolarizes cholinergic low-threshold burst neurons in the rat laterodorsal tegmental nucleus in vitro. Proc Natl Acad Sci USA *89:*743, 1992.

Maquet P, Peters JM, Aerts J, Delfiore G, Degueldre C, Luxen A, Franck G: Functional neuroanatomy of human rapid-eye-movement sleep and dreaming. Nature *383:*163, 1996.

*Marshall L, Derad I, Strasburger CJ, Fehm HL, Born J: A determinant factor in the efficacy of GHRH administration in promoting sleep: High peak concentration versus recurrent increasing slopes. Psychoneuroendocrinology *24:*363, 1999.

Meddis R, Pearson AJD, Lanford G: An extreme case of healthy insomnia. Electroencephalogr Clin Neurophysiol *35:*213, 1973.

Mignot E, Kimura A, Lattermann A, Lin X, Yasunaga S, Mueller-Eckhardt G, Rattazzi C, Lin L, Guilleminault C, Dement WC, Underhill P: Extensive HLA class II studies in 58 non-DRB1*15 (DR2) narcoleptic patients with cataplexy. Tissue Antigens *49:*329, 1997.

Moore P, Gillin JC, Bhatti T, DeModena A, Seifritz E, Clark C, Stahl S, Rapaport M, Kelsoe J: Rapid tryptophan depletion, sleep EEG and mood in men with remitted depression on serotonin reuptake inhibitors. Arch Gen Psychiatry *55:*534, 1998.

Oleksenko AI, Mukhametov LM, Polyakova IG, Supin AY, Kovalzon VM: Unihemispheric sleep deprivation in bottlenose dolphins. J Sleep Res *1:*40, 1992.

Porkka-Heiskanen T, Strecker RE, Thakker M, Bjorkum A, Greene RW, McCarley RW: Adenosine: A mediator of one sleep-inducing effect of prolonged wakefulness. Science *276:*1265, 1997.

Portas CM, Thakkar M, Rainnie D, McCarley RW: Microdialysis perfusion of 8-hydroxy-2-(di-n-propylamino) tetralin (8-OH-DPAT) in the dorsal raphe nucleus decreases serotonin release and increases rapid eye movement sleep in the freely moving cat. J Neurosci *16:*2820, 1996.

Seifritz E, Gillin JC, Rapaport MH, Kelsoe JR, Bhatti T, Stahl SM: Sleep electroencephalographic response to muscarinic and serotonin $_{1A}$ receptor probes in patients with major depression and in normal controls. Biol Psychiatry *44:*21, 1998.

Seifritz E, Stahl SM, Gillin JC: Human sleep EEG following the 5-HT1A antagonist pindolol: Possible disinhibition of raphe neuron activity. Brain Res *759:*84, 1997.

Sherin JE, Shiromani PJ, McCarley RW, Soper CB: Activation of ventrolateral preoptic neurons during sleep. Science *271:*216, 1996.

Skaggs WE, McNaughton BL: Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. Science *271:*1870, 1996.

*Steriade M, Contreras D, Amzica F: Synchronized sleep oscillations and their paroxysmal developments. Trends Neurosci *17:*199, 1994.

Thase ME, Buysse DJ, Frank E, Cherry CR, Cornes CL, Mallinger AG, Kupfer DJ: Which depressed patients will respond to interpersonal psychotherapy? The role of abnormal sleep EEG profiles. Am J Psychiatry *154:*502, 1997.

Tobler I, Gaus SE, Deboer T, Achermann P, Fischer M, Ruelicke T, Moser M, Oesch B, McBride PA, Manson JC: Altered circadian activity rhythms and sleep in mice devoid of prion protein. Nature *380:*639, 1996.

Tosini G, Menaker M: Circadian rhythms in cultured mammalian retina. Science *272:*419, 1996.

White JL, Darko DF, Brown SJ, Miller JC, Hayduk R, Kelly T, Mitler MM: Early central nervous system response to HIV infection: Sleep distortion and cognitive-motor decrements. AIDS *9:*1043, 1995.

Wu JC, Gillin JC, Buchsbaum MS, Hazlett E, Sicotte N, Bunney WE Jr: The effect of sleep deprivation on cerebral glucose metabolic rate in normal humans assessed with positron emission tomography. Sleep *14:*155, 1991.

Wu JC, Gillin JC, Buchsbaum MS, Hershey T, Johnson JC, Bunney WE Jr: Effect of sleep deprivation on brain metabolism of depressed patients. Am J Psychiatry *149:*538, 1992.

1.20 APPETITE

NORI GEARY, PH.D., AND GERARD P. SMITH, M.D.

Recent methodological innovations in the behavioral, physiological, biochemical, and genetic sciences have led to fundamental changes in the understanding of the controls of eating. The traditional view, which held that food intake was tightly controlled by hypothalamic neural centers reacting to the state of energy balance, has been replaced by a more complex concept of control by a distributed neural network in the brain that integrates oropharyngeal food stimuli (including those giving rise to hedonic perceptions), gastrointestinal signals, hormonal and metabolic signals, and environmental and experiential contingencies.

An important part of this change in perspective has been to focus on the individual meal as the functional unit of analysis for the investigation of eating rather than on measures of food intake over extended periods (e.g., kcal/day). This change is based on the realization that understanding the controls of individual meals is necessary for understanding disordered as well as normal patterns of eating: abnormally large meal sizes are the crucial behavioral change in binge eating, and abnormally small meals are the crucial behavioral change in restricted eating and anorexia nervosa. Therefore, this chapter reviews the basic science of eating from the perspective of the normal and disordered controls of individual meals.

MEALS

Meal as the Unit of Analysis The meal is the biological and functional unit of eating. It is the biological unit because in the vast majority of animal species and in humans, ingestive behavior is organized as discrete bouts, or meals, separated by intervals of noneating. Furthermore, this basic pattern is maintained despite great variations in food availability, food quality, opportunities or pressures to engage in nonfeeding behaviors, and the presence of physiological challenges, such as infection. The meal is the functional unit of eating because the timing, size, and content of meals provides a complete description of the basic nutritional functions of food, namely what to eat and how much to eat. The same may be said of other functions of food, for example, the psychological function of maximizing the pleasures of eating.

The phases of the meal that require analysis are the initiation of eating, the maintenance of eating during the meal, and the termination of eating (Fig. 1.20-1A). The intermeal interval is bounded by the end of one meal and the beginning of the next. Ingestion of food during the meal stimulates positive feedback signals from the mouth that contribute to the maintenance of eating and negative feedback signals from the mouth, stomach, and small intestine that ultimately terminate eating (Fig. 1.20-1B).

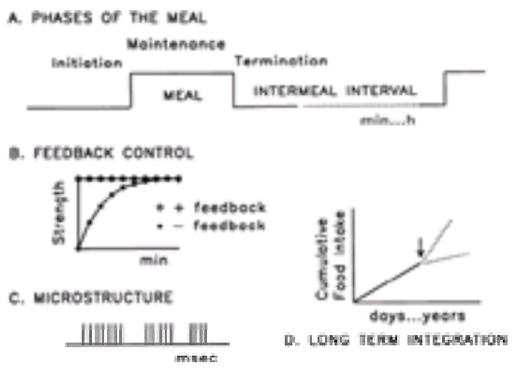

FIGURE 1.20-1 A, The meal is the functional unit of eating. Direct and indirect controls of the initiation, maintenance, and termination of eating determine the size of meals and the duration of intermeal intervals. **B**, Meal size is a function of positive and negative feedbacks elicited by preabsorptive food stimuli arising during meals. Positive feedback, shown in hypothetical units, begins as soon as eating begins and is thought not to decay significantly during the meal. Negative feedback increases in strength gradually as eating continues. When the strength of negative feedback equals that of positive feedback, eating stops and the meal ends. **C**, Microstructural analyses are based on the patterns of the individual movements of eating, for example, the temporal organization of individual licks of liquid foods on an event recorder. **D**, Amount eaten in the long term is simply the sum of the sizes of many individual meals. This schematic demonstrates the typically constant increase in cumulative intake evident on a macroscopic scale (e.g., kcal/d). This rate is increased or decreased if energetic demands are increased or decreased, for example, by changing physical activity, as shown to the right of the arrow.

Microstructure Eating employs a variety of movements. Food is licked, sucked, bitten, or masticated prior to being moved by lingual and palatal movements to the oropharynx and swallowed. These are all largely automatic rhythmic movements organized by central pattern generators in the hindbrain. Therefore, it is not surprising that the hindbrain also contains sufficient neural machinery to initiate and terminate eating movements in response to various afferent stimuli.

The temporal organization of eating movements is the microstructure of eating (Fig. 1.20-1C). The chief advantage of the analysis of microstructure is its potential to track behavior back through the motor neurons producing it. This analysis has produced very promising results in highly restricted experimental situations. It is not yet possible, however, to link individual licks and bites with the controls of meal timing, size, and content.

Meal and the Scale Controls of meals are sometimes considered short-term mechanisms and are supposed to be under fundamentally different constraints than the “long-term” mechanisms presumed to be related to more biological variables. For this reason, long-term studies of eating often do not involve any direct measure of behavior, but rely on indirect measures, such as amounts eaten per day or week. This strategy can provide a net accounting of long-term changes of eating within the larger control system of energy balance (Fig. 1.20-1D), but it has serious disadvantages. Most importantly, the lack of a behavior makes it a dead end for the brain sciences. The brain cannot be understood without an understanding of its behavioral output. In contrast, meal analysis not only provides behaviors to track back into physiology, but also connects to all long-term controls because the amount eaten over any period of time is completely determined by the number and size of meals. Therefore, all influences on eating must affect the timing or size of meals. Thus, the meal is also the appropriate unit of analysis for studying the contribution of eating to the regulation of energy balance.

Subjective Experience of Eating The subjective experiences associated with meals are scientifically accessible in humans, and analysis of them has produced many important insights into normal and disordered eating. This work has been most productive when it has been based on operationally defined categories of subjective experience. Theoretical categorizations of subjective experience, in contrast, have been less useful. For example, the traditional distinction between appetite as an urge to eat elicited by anticipation of the pleasures of eating, and hunger as an urge to eat elicited by the vegetative consequences of nutrient depletion has not been illuminating.

CONTROLS OF MEAL FREQUENCY

Because the frequency of eating is a determinant of intake, it is important to identify the adequate stimuli and mechanisms of the controls of the initiation of eating.

The adequate stimuli are numerous and diverse. They include olfactory, visual, auditory, temporal, circadian, metabolic, cognitive, and social stimuli. Most of these are conditioned on the basis of prior eating experience. Despite conventional wisdom, gastric contractions are not adequate stimuli for meal initiation. Stimuli resulting from inhibition of glucose utilization in the brain or liver and the inhibition of fatty acid oxidation in the liver may be unconditioned stimuli for meal initiation, but they probably do not operate when food is readily available.

A new physiological stimulus for the initiation of eating related to metabolism is a transient decline in plasma glucose that has been recorded prior to spontaneous meals in rats and humans (Fig. 1.20-2). The decline is too small to influence cellular glucose availability; rather, its pattern appears crucial: pharmacologically stimulated declines that are too small or too large do not stimulate feeding. The relative importance of this stimulus in controlling meal initiation under normal environmental conditions remains to be determined. The mechanisms of how these arious stimuli initiate eating are unknown. The lone exception is that hepatic vagal afferents are necessary for the initiation of eating by inhibition of metabolism in the liver.

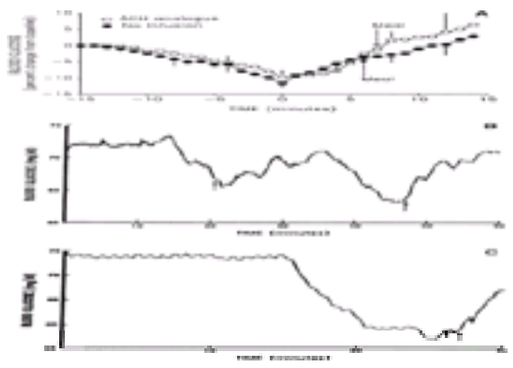

FIGURE 1.20-2 The premeal transient decline in blood glucose in rats and humans. **A**, Filled circles, percent deviations from baseline blood glucose during undisturbed spontaneous feeding in rats. Note that meals begin several minutes after the nadir. Open circles, intravenous infusions of an insulin secretogogue that produced similar declines also elicited meal initiation. (Reprinted with permission from Campfield LA, Smith FJ: Meal initiation occurs after experimental induction of transient declines in blood glucose. Am J Physiol *265:*R1423, 1993.) **B, C**. Blood glucose changes preceding requests for morning meals (arrows) in two normal weight volunteers spending the night in a metabolism laboratory. (Reprinted with permission from Campfield LA, Smith FJ, Rosenbaum M, Hirsch J: Human eating: Evidence for a physiological basis using a modified paradigm. Neurosci Biobehav Rev *20:*133, 1996.)

CONTROLS OF MEAL SIZE

Direct Controls The most experimentally accessible, and consequently the best understood, controls of eating are the direct controls of meal size. These have two defining characteristics. First, these signals originate from the action of ingested food acting on preabsorptive receptors during the meal. The critical receptors are localized in the oropharynx, stomach, and upper small intestine. The preabsorptive location and defined temporal domain of their action make these signals very accessible, and probably also accounts for their prepotency in the physiological control of eating. These receptors transduce food stimuli into changes in peripheral neural activity and changes in local or systemic levels of chemical signals. Information encoded in these ways functions as feedback controls that affect meal size by influencing either the maintenance of eating during the meal or the termination of eating at the end of the meal.

The second defining characteristic of direct controls of meal size is that their afferent signals are initially processed in the hindbrain. Local interneuronal networks in the hindbrain consist of neurons that receive afferent stimuli from the gut, interneurons that integrate these stimuli, and premotor and motor neurons that produce the rhythmic oral movements of eating. These hindbrain networks provide the basic physiological control of eating. Their function has been elegantly demonstrated in decerebrate rats, in which supracollicular transections of the brainstem disconnect the hindbrain from the diencephalon and telencephalon (Fig. 1.20-3). Decerebrate rats do not search for food, but when liquid food is delivered into the mouth, they eat and swallow discrete meals that are terminated by passive rejection of delivered food. Furthermore, meal size is normally responsive to many of the same direct controls of meal size that operate in the intact rat (Fig. 1.20-4).

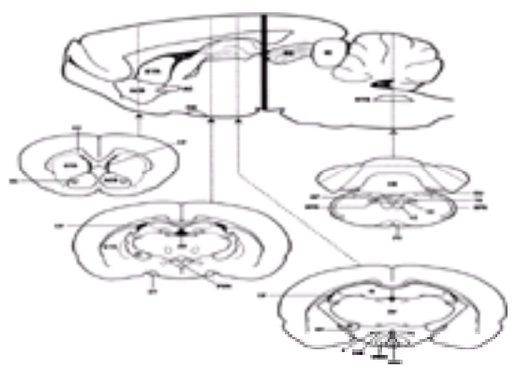

FIGURE 1.20-3 Schematic drawing of sagittal and coronal sections of the rat brain indicating several structures implicated in the central control of eating. The dotted lines indicate the planes of section of the coronal sections. The thick vertical line in the sagittal section indicates the level of the cerebral disconnection used in the studies of the chronic decerebrate rat discussed in the text. AC, anterior commissure; ACB, nucleus accumbens; AP, area postrema; ARC, arcuate nucleus; CB, cerebellum; CC, corpus callosum; CU, cuneate nucleus; DM, dorsomedial hypothalamic nucleus; F, fornix; H, hippocampus; IC, inferior colliculus; NTS, nucleus of the solitary tract; OT, optic tract; OX, optic chiasm; PVN, paraventricular hypothalamic nucleus; PY, pyramidal tract; SC, superior colliculus; SP5, spinal tract of the trigeminal nerve; STR, striatum; TS, solitary tract; VMH, ventromedial hypothalamic nucleus; LV, lateral ventricle; 3V, third ventricle; 4V, fourth ventricle; 10, dorsal motor nucleus of the vagus nerve; 12, hypoglossal nucleus. (Reprinted with permission from Paxinos G, Watson C: *The Rat Brain in Stereotaxic Coordinates*. Academic Press, New York, 1982.)

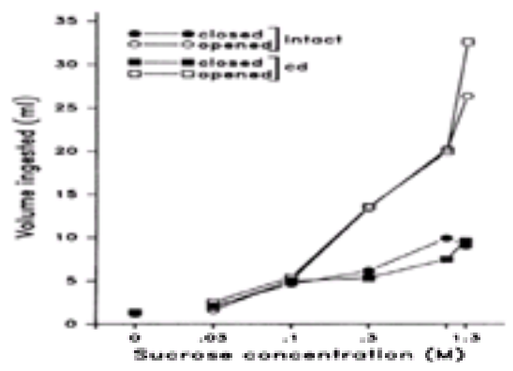

FIGURE 1.20-4 Intakes of various concentrations of sucrose by intact and chronic decerebrate (cd) rats. Sucrose was delivered by intraoral catheters until rejected by the animal. Intake is an increasing function of sucrose concentration in cd rats that feed normally (closed condition). The gain of this function is dramatically increased in sham feeding rats in which postingestive controls of meal size are minimized by opening a gastric cannula (opened condition). Note the close correspondence of both the stimulatory effect of increasing sucrose concentration and the interaction of this effect with the inhibitory effect of normal postingestive food stimuli in intact and cd rats. (Reprinted with permission from Grill HJ, Kaplan JM: Sham feeding in intact and chronic decerebrate rats. Am J Physiol *262:*R1070, 1992.)

Indirect Controls Indirect controls of meal size are formally distinguished from direct controls in that they do not arise from the preabsorptive action of food during meals. They also appear to be distinct from direct controls in several other ways. For example, they tend to act for longer periods than the durations of individual meals and intermeal intervals. Second, the neural mechanisms of indirect controls are localized predominantly in the forebrain. Forebrain networks mediate the higher functions of eating, including the appetitive behaviors involved in food search and selection as well as the subjective experience of eating. Forebrain systems receive ascending hindbrain afferents and forebrain sensory inputs, and are interconnected with each other and with the hindbrain systems in complex but poorly understood distributed eural networks. These neurological facts mean that forebrain systems control eating by modulating the direct controls of eating. Thus, although forebrain systems are necessary for many controls of eating, they are not sufficient for any of them.

Table 1.20-1 lists several indirect controls of eating. Given their variety, it is not surprising that the experimental analysis of indirect controls of eating has proceeded at an uneven pace. For example, identification of any normal metabolic controls of eating has been surprisingly difficult. Indeed, whether energy metabolism exerts any physiological control over meal size except after long periods of food deprivation is controversial. Similarly, until 1994, little was known about the mechanism by which adiposity influenced eating.

Type*	Examples
Metabolic	Changes in availability of substrates, changes in adipose tissue mass
Cyclic	Circadian and ovarian hormonal rhythms
Immune system	Neoplasms, infections
Conditioned	Learned preferences, aversions, and satieties

*Types are neither mutually exclusive nor exhaustive. Note that conditioned controls are produced by physiological contingencies, such as illness; by social or ecological contingencies, such as group sizes or the relative densities of food sources or predators; and, in humans, by cultural and esthetic contingencies.

Table 1.20-1 Indirect Controls of Meal Size

In contrast to the difficulty in identifying a specific metabolic control of eating, learned controls of eating are prominent and pervasive. With the exception of a few unconditioned gustatory preferences and aversions (such as for sweet and bitter taste), all food selection appears to be learned. Controls of meal initiation and meal size are also readily conditionable in animals and humans. Flavor provides the *conditioned stimuli* for these learned controls. The *unconditioned stimuli* for many learned controls, such as satiety conditioned by the nutrient density of a food, are postingestive consequences of eating. Other controls may involve higher-order conditioning. For example, cultural socialization may result in a preference for the flavor of capsaicin (chili), which all infants avoid.

Some forms of learned controls of eating are special forms of conditioning. Flavor aversions conditioned by upper gastrointestinal food poisoning, for example, can be learned after only one *conditioned stimuli-unconditioned stimuli* pairing despite extraordinarily long *conditioned stimuli-unconditioned stimuli* delays (hours), and are extremely resistant to extinction. Carefully designed tests are required to reveal learned controls of meals, and much remains to be done on this problem. But this work should be a high priority in the investigation of disordered eating because the number of unconditioned hindbrain mechanisms through which all learned controls of meal size operate is small enough to make it possible to achieve control of them in humans. Furthermore, behavior therapy programs should increase the accessibility of conditioned physiological controls of eating because they are highly structured and can present specific food stimuli.

SIGNALS AND PERIPHERAL MECHANISMS FOR CONTROL OF MEAL SIZE

The size of an individual meal is controlled by the central neural integration of positive and negative feedback signals that affect the maintenance or termination of eating (Fig. 1.20-1B). Thus, increases in meal size result if there is an increase in positive feedback signals, a decrease in negative feedback signals, or both, and decreases in meal size result if there is a decrease in positive feedback signals, an increase in negative feedback signals, or both (Table 1.20-2). This simple syllogism has provided a powerful strategy in the analysis of normal eating, and should be equally useful in the analysis of disordered eating.

	Signal Type*	
Change of Meal Size	Positive	Negative
Increase	↑	—
Increase	—	↓
Increase	↑	↓
Decrease	↓	—
Decrease	—	↑
Decrease	↓	↑

*Signals are the hypothetical neural representations of the direct controls of meal size that originate from preabsorptive actions of food stimuli during meals. Positive signals tend to increase meal size and negative signals tend to decrease meal size. Changes in their strength can result from changes in afferent activity or changes in the central neural processing of that afferent information, for example as the result of influence of indirect controls. ↑ indicates increased strength, ↓ indicates decreased strength, and — indicates no change.

Table 1.20-2 Changes in Positive and Negative Feedback Signals that Lead to Changes in Meal Size

Direct Controls of Meal Size

Maintenance of Eating Flavors are the only known signals facilitating eating once it has begun. These positive feedback controls of meal size arising from oropharyngeal food stimuli reach the brain through the olfactory (via retronasal stimulation), trigeminal, facial, glossopharyngeal, and vagus nerves. This afferent information (except for the olfactory stimuli) is initially processed in the hindbrain. For example, sweet tastes elicit smiles and ingestive responses in normal and mesencephalic human infants, whereas bitter tastes elicit grimaces and rejection. Similarly, test meal size is a graded function of sucrose concentration in the chronic decerebrate rat. The identity of this function in intact and decerebrate rats (Fig. 1.20-4) demonstrates that hindbrain processing of this signal is sufficient for this control of meal size without any contribution from forebrain neural networks involved in the perception of sweetness.

If inhibitory postingestive signals are eliminated by use of the sham feeding technique, the gain of the intake function increases dramatically (Fig. 1.20-4). Indeed, as originally observed by Ivan Pavlov, after a day of food deprivation, animals sham feed nearly continuously for hours.

Termination of Eating Signals contributing to the normal termination of meals are referred to as *satiety signals*. Oropharyngeal food stimuli, gastric, and intestinal food stimuli all give rise to satiety signals. The oropharyngeal signals are presumably neural and are not well understood; some may involve the cephalic phase release of gut peptides.

Control of meal size by gastric volume has been demonstrated in rats with pyloric cuffs that prevent ingested food from passing into the small intestine. The mechanoreceptors initiating this control apparently activate vagal and splanchnic visceral afferents, possibly through the release of gastrin-releasing peptide or neuromedin B.

Food stimuli arising from the intestinal phase of digestion activate chemoreceptors that initiate a number of potent satiety signals. Many of these signals reach the hindbrain via vagal afferents, which are stimulated directly by small intestinal luminal contents or indirectly by gut peptides that are released by the preabsorptive actions of food. Intestinal infusions of various food stimuli that match the normal rates of emptying from the stomach have convincingly demonstrated the physiological

role of intestinal satiety signals in animals and humans.

A simple method to study postingestive satiety mechanisms is the administration of preloads. Oral preloads, sometimes administered in ways that minimize oropharyngeal stimuli, have been used extensively in animals and humans to determine the contributions of a food's metabolic energy content, nutrient composition, colligative and osmotic properties, and weight, to its satiating potency. This work suggests that there is no general quantitative relation among these variables, although some interesting points have emerged. When mixed nutrient preloads are used, the decrease in eating is usually proportional to the preload's metabolizable energy content, although the compensation is often incomplete. When the preload's macronutrient composition is varied, protein preloads are typically more satiating than isoenergetic loads of carbohydrate or fat. Finally, liquid foods, especially soups, are typically relatively more satiating than foods in other physical forms, especially for women.

The relative contribution of satiating food stimuli that activate receptors in different loci in animals can be isolated experimentally by techniques such as the chronic gastric cannula and the inflatable pyloric cuff. Alternative strategies are required in humans. One is to vary the time between preloads and test meals. Such studies indicate that preloads are more satiating when eating begins sooner after loading (£ 30 min) than when eating begins later (£ 60 min), suggesting that preabsorptive signals are more important than postabsorptive signals in mediating the satiating effects of preloads. Another strategy to identify the site of action of food stimuli is to compare the effects of infusions made into successive functional compartments. This method has also produced strong evidence for the primacy of intestinal over postabsorptive signals in satiety. For example, a recent study compared the satiating effects of intraduodenal glucose infusions and intravenous glucose infusions that produced identical increases in systemic glucose levels. Intraduodenal, glucose decreased premeal hunger ratings and decreased meal size; intravenous glucose did not.

GUT PEPTIDES Gut peptides appear to be an especially important class of direct satiety signals elicited by food stimuli in the small intestine (Table 1.20-3). One of these, cholecystokinin (CCK), is now proven to be a physiological control of meal size, based on fulfillment of five criteria: (1) CCK is released during meals; (2) administration of CCK in amounts that reproduce prandial levels is sufficient to inhibit meal size; (3) CCK reduces meal size in the absence of abnormal or adverse behavioral, physiological, or subjective effects; (4) administration of antagonists of the type A CCK (CCK $_A$) receptor before meals increases meal size; and (5) CCK$_A$ receptor antagonists block the satiating effect of food stimuli that release endogenous CCK. All these criteria have been met in rats; further work is required to establish the last two criteria in humans.

Physiological Status	In Animals	In Humans
Cholecystokinin (CCK)	Y	+
Pancreatic glucagon	Y	+
Insulin	+	?
Gastrin-releasing peptide	?	?
Neuromedin B	?	?
Amylin	+	?
Somatostatin	?	?
Neurotensin	?	?
Apolipoprotein A-IV	?	?
Enterostatin	?	?

*Each peptide listed is directly or reflexively released during meals and each has been implicated in the mechanism of satiation in animals. Physiological status in animals and humans is rated as proven (Y), probable (+), or unclear (?), based on comparisons of prandial levels with threshold doses of exogenous peptide for inhibition of meal size and on tests of antagonism of endogenous peptide.

Table 1.20-3 Peripheral Peptides Hypothesized to Signal Satiety*

Pancreatic glucagon has also fulfilled these criteria in rats. Figure 1.20-5 shows an example of the specific satiating actions of CCK and glucagon in young men of normal weight.

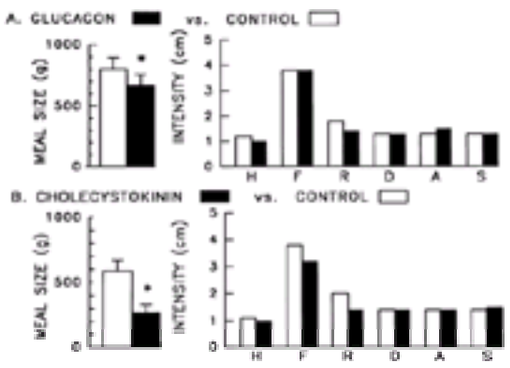

FIGURE 1.20-5 Cholecystokinin and pancreatic glucagon are released during meals in response to preabsorptive food stimuli. Intravenous infusion of either peptide during meals in normal human volunteers reduces meal size without physical or subjective side effects. **A**, Effects of intravenous infusion of pancreatic glucagon (3 ng/kg^{-1}/min^{-1}) on the size of macaroni and beef lunches and postprandial subjective ratings in 12 normal weight men. **B**, Effects of similar infusions of cholecystokinin octapeptide (CCK-8, 2 ng kg^{-1} min^{-1}) in 8 of the same men. Psychophysical ratings were made with 15-cm visual analog scales (0 cm, “not at all”; 15 cm, “most possible”) of “hungry” (H), “stomach full” (F), “restless” (R), “dizzy” (D), “stomach ache” (A), and stomach sick (S). Scales were completed just after the meal. Both peptides significantly decreased meal size (*) without changing the normal pattern of postprandial subjective reports. (Reprinted with permission from Geary N, Kissileff HR, Pi-Sunyer FX, Hinton V: Individual, but not simultaneous, glucagon and cholecystokinin infusions inhibit feeding in men. Am J Physiol *62:*R975, 1992.)

SYNERGY Many of the direct controls of meal size appear to interact in a functionally synergistic fashion, that is, their actions combine in a supra-additive manner. For example, the satiating potency of exogenous CCK is increased by small gastric preloads, even of substances that are not CCK secretagogues. Because such synergistic interactions present the possibility for maximal changes in meal size by minimal manipulations, they are an especially attractive avenue for development of therapeutic controls of meal size.

Indirect Controls of Meal Size Conditioning modulates both meal timing and meal size; other indirect controls appear to act mainly on meal size. For example, increased eating in cold environments, during lactation, after ovariectomy, or after body weight loss caused by food deprivation as well as the decreased eating after forced weight gain are mediated primarily by changes in meal size.

Fat Mass, Insulin, and OB Protein Since the classic experiments by G. C. Kennedy in the early 1950s, it has been known that the mass of fat in the body is correlated inversely with meal size and food intake. Recent work has demonstrated that this lipostatic control is mediated in rodents by actions in the brain of circulating insulin and OB protein (also called leptin), the product of the *oB* gene.

Insulin is a negative-feedback signal of lipostatic control because: (1) basal concentrations of plasma insulin are positively correlated with adipose mass; (2) plasma insulin is actively transported into the brain by a vascular receptor transport mechanism; and (3) intracerebral infusion of insulin in amounts too small to have peripheral effects by diffusion into the blood decreases meal size and food intake. It is clear that obesity involves some dysfunction of lipostatic control because it is characterized by hyperinsulinimia. Resistance to the inhibitory effect of insulin on eating has been demonstrated directly after central administration of insulin in genetically obese *(fa/fa)* rats.

OB protein is a secretory product of the adipose tissue discovered in 1994. It is a negative-feedback signal of lipostatic control because plasma concentration of OB protein is a linear function of fat mass, and peripheral or central administration of OB protein decreases meal size in a potent, dose-related manner. Receptors for OB protein are widespread in the periphery and the brain, and the physiology of these systems is under intense investigation in animals and humans. It is already clear

that increased meal size accounts for the hyperphagia of genetically obese rodents that fail to produce biologically active OB protein (*ob*/*ob* mouse) or lack functional OB receptors (*db*/*db* mouse and *fa*/*fa* and *cp*/*cp* rats) and that decreased meal size accounts for the hypophagia of normal animals treated with OB protein. A report that infusion of specific antibodies to the OB protein into the lateral ventricle of rats increased food intake suggests that OB protein may exert a tonic indirect inhibitory control of meal size. As a rule, the concentration of circulating OB protein is positively correlated with adiposity in humans, suggesting that decreased sensitivity may contribute to obesity. This hypothesis is supported by a demonstration that mice made obese by feeding a very high-fat diet were less hypophagic than control mice following OB protein administration. Isolated instances of human obesity may arise from mutations in the genes expressing OB protein or its receptor. Three such families have been identified, two with mutations resulting in deficiencies of functional OB protein and one with a deficiency in functional OB protein receptors.

Estradiol Estradiol exerts phasic and tonic inhibitory effects on eating. The phasic effect is a cyclical decrease in food intake that occurs during the periovulatory period. The tonic effect is revealed by ovariectomy, which in animals elicits a dramatic increase in the basal level of feeding and abolishes the cyclical feeding rhythm. These changes are best correlated with hanges in circulating estradiol, and estradiol is the only hormone of the hypothalamic-pituitary-ovarian axis that normalizes eating and body weight in ovariectomized animals. All these effects are accounted for by changes in meal size. One of the mechanisms by which estradiol decreases meal size is by increasing the potency of cholecystokinin. Estradiol's potential role in the increased vulnerability of women to eating disorders has not been established. It is possible that estradiol's inhibitory influence on eating may be part of the reason that anorexia nervosa often develops near menarche.

Learning Gastric and intestinal satiety signals, the major direct negative-feedback controls of meal size, also provide important unconditioned stimuli for indirect, conditioned controls of meal size. The importance of conditioned controls can be demonstrated by experimental manipulations that prevent the normal actions of the postingestive controls of meal size and lead to their gradual extinction (Fig. 1.20-6). This has been modeled in the sham feeding rat and probably occurs in patients who vomit or purge after binge-eating.

FIGURE 1.20-6 The potency of learned controls of meal size based on postingestive food stimuli is revealed by the progressive increase in test meal size during repeated sham feeding trials. During sham feeding liquid food drains from open gastric cannulas without significant accumulation in the stomach or small intestine, so that learned controls based on gastrointestinal food stimuli extinguish. In this experiment rats were offered a sweet liquid diet once daily, after 3 hours of deprivation of their maintenance diet. During week 1, rats fed normally (real feeding, RF); during weeks 2 to 7, sham feeding (SF) tests were alternated with real feeding tests, and during weeks 8 to 11, rats only sham fed. The figure shows the average real and sham meal sizes in each week. During the first sham feeding test, rats still ate well-defined meals terminated by behavioral signs of normal satiety, indicating that after this short period of food deprivation, pregastric food stimuli can elicit satiety. Meal size, however, nearly doubled during this test, because of the absence of direct, unconditioned gastric and postgastric controls of eating. Sham meal size doubled during weeks 3 to 7, when sham and real feeding tests were alternated, whereas real meal size increased only a small amount, and sham meal size almost doubled a third time during weeks 8 to 11, when there were no real feeding tests. These increases reflect the extinction of conditioned satiety. Subsequent studies indicate that this conditioned satiety is based on both gastric and intestinal food stimuli.

CENTRAL MECHANISMS OF THE CONTROL OF MEAL SIZE

Hypothalamic Centers and a Distributed Network Experimental manipulations of the hypothalamus produce large effects on eating. Lateral hypothalamic damage produces hypophagia; electrical stimulation in this area elicits eating. Lesion and stimulation of the ventromedial hypothalamus have the opposite results. Originally presented as explanations of how eating was organized neurologically into a lateral hunger center reciprocally connected with a medial satiety center, these hypothalamic effects are now seen as problems to be solved in terms of hypothalamic functions in the distributed neural network that controls eating. Figure 1.20-3 indicates several sites that are functional nodes in this network in animals. Direct knowledge of the network that controls human eating is extremely limited, but increased progress in this area may be expected as evolving techniques in brain imaging reinvigorate functional neuroanatomy.

Neuropharmacology, Neurotransmitters, and Neuromodulator Neuropharmacological methods have largely replaced classical lesion and stimulation techniques in the study of eating. In a few instances it is now possible to link particular direct controls of meal size with the specific central neurochemical systems that are necessary for their normal processing.

In addition, a number of amines and peptides reliably increase or decrease eating (Table 1.20-4). Although it is reasonable to assume that the stimulants of eating enhance the orosensory positive-feedback effect of the test diets and that inhibitors of eating enhance the postingestive negative-feedback effects of the test diets, further work is required to demonstrate this and to eliminate the alternative possibilities (Table 1.20-2). Those endogenous substances for which there is the most evidence of function under physiological conditions (Table 1.20-4) are discussed. Again, this progress results mainly from work in animals, in which it is now possible to manipulate and measure brain neurochemicals at the sites of their biological action during meals. In contrast, only very indirect methods, such as measurements of the concentrations of neurotransmitters or metabolites in the cerebral spinal fluid, are possible in humans and these cannot usually be done in the context of eating.

Increase Eating		Decrease Eating	
Chemical	Physiological Status	Chemical	Physiological Status
Dopamine	Y	Serotonin	Y
Neuropeptide Y	Y	Corticotropin-releasing factor	Y
Norepinephrine	Y	Gastrin-releasing peptide	Y
Endogenous opioids	Y	Glucagon-like peptide 1	+
Galanin	+	Oxytocin	+
Growth-hormone-releasing factor	+	Neuromedin B	?
Melanin-concentrating hormone	+	Enterostatin	?
Orexins A and B (hypocretins 1 and 2)	+	α-melanocyte-stimulating hormone	+

Each neurochemical listed occurs in the brain and has been implicated in the central physiology of eating in animals. Endogenous opioids include beta-endorphin and dynorphin. Physiological status in animals is rated as present (Y), probable (+), or unclear (?), based on tests of agonists and antagonists of the endogenous components.

Table 1.20-4 Central Neurotransmitters and Neuromodulators Implicated in the Control of Eating

Stimulants of Eating

DOPAMINE Dopamine is thought to play an important role in mediating the rewarding effects of food that maintain eating, produce pleasure, and reinforce learning about food. In animals, dopamine antagonists markedly decrease the appetitive and consummatory responses maintained by food. For example, dopamine antagonists selectively reduce the orosensory positive-feedback effect of sucrose solutions that stimulate ingestion without altering the sensory intensity of the sucrose or its postingestive satiating potency, and without impairing oromotor movements, the animals' motor capacity, or creating an aversion. Dopamine is released in several brain areas that are thought to mediate reward functions, including the nucleus accumbens, medial hypothalamic areas, and the amygdala.

Dopamine may also be involved in the central processing of the satiating effect of peripheral CCK, and dopamine inhibits intake when injected into perifornical sites in the hypothalamus. This is a good example of how the effect of a transmitter on eating is determined by the site and functional contingencies of the network it acts in.

NOREPINEPHRINE Noradrenergic neurons in the locus ceruleus of the hindbrain project to the hypothalamus and many other brain sites. Numerous studies support a role for norepinephrine in the control of feeding. Administration of norepinephrine into the paraventricular nucleus and perifornical hypothalamus increases meal size and preferentially stimulates carbohydrate intake via an a_2-adrenergic receptor mechanism. Corticosterone also upregulates these receptors. Furthermore, spontaneous feeding increases paraventricular nucleus norepinephrine release and also upregulates a_2-adrenergic receptors. The effect of norepinephrine infusion into the paraventricular nucleus shows rapid tolerance, however, so that chronic infusions do not affect body weight.

NEUROPEPTIDE Y Neuropeptide Y is the most potent known centrally acting orexigen. Short-term central administration of neuropeptide Y increases feeding in numerous animal species, although it had no effect in the single test in primates, done in baboons. Long-term central administration of neuropeptide Y leads to sustained hyperphagia and increased body weight in rats. The paraventricular nucleus and adjacent perifornical hypothalamus, which receive projections from neuropeptide Y cells in the hindbrain and in the arcuate nucleus of the hypothalamus, are considered the most likely sites for a physiological action of neuropeptide Y. Concentrations of neuropeptide Y in this area change during meals, and administration of neuropeptide Y antibodies or antisense neuropeptide Y deoxynucleotides here decreased feeding. Several lines of evidence suggest that both insulin and OB protein may inhibit feeding in part by down-regulating neuropeptide Y.

OPIOIDS Administration of opioid antagonists such as naloxone (Narcan) reduce meal size under a variety of conditions. This inhibitory effect appears to be relatively selective for preferred foods, such as sweet or high-fat foods, suggesting that opioids are particularly relevant to the maintenance of eating by preferred foods. Ligands for the μ-, k-, and d-opioid receptor subtypes have all been reported to stimulate feeding, and several brain areas, including the ventral tegmentum, various hypothalamic nuclei, the amygdala, and the nucleus accumbens, have been implicated in opioid-induced feeding. Endogenous opioids are also necessary for the orexic effect of neuropeptide Y. The most important site of this opioid mediation is the nucleus tractus solitarius of the hindbrain. This is a clear demonstration of the dependence of forebrain effects on eating on hindbrain mechanisms proximal to the final motor neurons that control eating.

Inhibitors of Eating

SEROTONIN The prototypical serotonin agonist dexfenfluramine (withdrawn from the U.S. market), which releases synaptic serotonin and blocks serotonin reuptake, inhibits eating in animals and humans. The inhibitory effect in rodents can be blocked by antagonism of two serotonin receptor subtypes, serotonin (5-hydroxytryptamine [5-HT]) subtype 2C (5-HT_{2C}) are 5-HT_{1B}; 5-HT_{1B} is the homologue of the 5-HT_{1D} subtype in other mammals. Complementary work with serotonin agonists indicates that 5-HT_{1B} receptors primarily affect meal size and 5-HT_{2C} receptors affect eating rate. Endogenous serotonin appears to have a physiological role in satiety because both intraventricular injection of 5-HT_{2C} receptor antagonists and direct injection of 5-HT_{1A} receptor antagonists in the brainstem raphe, where they stimulate autoreceptors on serotonin perikarya, increase meal ize. Additionally, mutant mice without 5-HT_{2C} receptors are overweight and unresponsive to serotonin agonists. There is also evidence that part of CCK's satiating action is mediated by central 5-HT_{2C} receptors. Serotonin may be also be involved in the stimulatory effect of sweet taste on eating. The hypothalamic paraventricular nucleus appears to be an important site of serotonin effects on eating in animals but it is not the only one.

CORTICOTROPIN-RELEASING FACTOR This peptide is synthesized by neurons in many brain areas, including the medial parvocellular subnucleus of the paraventricular nucleus, which is the source of the projection to the median eminence, the medial preoptic area, the nucleus tractus solitarius, and the amygdala. Administration of CRF into the cerebral ventricles or into the paraventricular nucleus has a potent anorectic effect, and CRF appears to be involved in many circumstances in which eating is inhibited. Interestingly, there appears to be a reciprocal relationship between many controls of CRF and neuropeptide Y activity, for example, has opposite effects OB protein on neuropeptide Y and CRF. Furthermore, the two peptides have opposite effects on sympathetic nervous activity related to metabolism as well as on eating.

DISORDERED CONTROLS OF MEAL SIZE IN BULIMIA NERVOSA

The quantitative approach to the controls of meal size that has fueled the recent progress in the basic science of eating has begun to be applied to patients with eating disorders. The most significant results have come from studies in patients with bulimia nervosa when it was discovered that patients ate much larger meals under laboratory conditions. This difference can be obtained in individual test meals of a single test diet or in residential laboratory settings in which patients have access to a variety of foods for 1 or more days. An important aspect of the results in the latter condition is that patients with bulimia nervosa did not necessarily eat more meals than controls, but dramatically increased the size of about one quarter of their meals.

There is considerable evidence that the postingestive negative-feedback effects of food are less satiating in patients with bulimia nervosa. Equivalent preloads of food decrease intake less in patients with bulimia nervosa than in controls, particularly when the patients are eating large meals. Furthermore, patients with bulimia eat a larger amount of food to produce equivalent self-reports of fullness during a meal. This suggests that ingested food has a less satiating effect because the peripheral, preabsorptive, negative-feedback mechanisms are decreased or that the central processing of this negative-feedback information reduces the potency of the negative feedback; there is growing evidence to support both these possibilities.

Two peripheral, negative-feedback signals are decreased in bulimia patients. First, volume distention of the stomach produces a decreased perceptual and mechanical response, presumably because the stomachs are larger than normal as the result of accommodation to the frequent ingestion of large meals. Second, food-stimulated release of CCK is less in patients with than bulimia nervosa than in controls (Fig. 1.20-7).

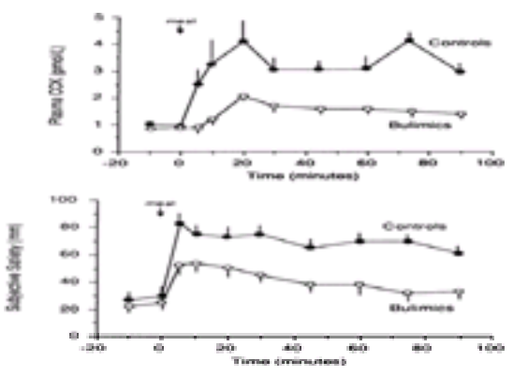

FIGURE 1.20-7 Prandial concentrations of cholecystokinin (CCK; **A**) and the subjective experience of satiety (**B**) are reduced in patients with bulimia nervosa. Fourteen patients and 10 control women matched for age and weight were offered a 400-ml liquid meal after an overnight fast (arrows) and ate it in 1 to 2 minutes. Plasma CCK was measured with a selective bioassay. Satiety was measured by a 100-mm visual analog scale (0 = empty, 100 = full). Both the peak CCK concentration and the integrated CCK response (area under the curve) were significantly reduced in patients with bulimia and this correlated with reports of significantly less satiety beginning 5 minutes after meal onset. (Reprinted with permission from Geracioti TD Jr, Liddle RA: Impaired cholecystokinin secretion in bulimia nervosa. N Engl J Med *319:*683, 1988.)

In addition to the diminished negative-feedback potency of these direct controls, there is evidence in support of abnormal indirect control and central processing. The most significant abnormality of the indirect controls is cognitive. This is demonstrated by the much larger meals eaten by bulimia patients when they are instructed to binge compared to when they are instructed not to binge. Thus, the instructions decrease the potency of the defective, negative-feedback direct controls. It is possible that instructions also enhance positive feedback directly (Table 1.20-2), but there is no evidence for this. The conditioning process resulting in this abnormal cognitive control is unknown.

The abnormal central processing of the negative-feedback information may result from a decrease in central serotonergic function that has been demonstrated under

a variety of conditions. If central serotonin function is decreased in bulimia patients, they should be more vulnerable than controls to a further decrease in serotonin function produced by serotonin depletion. This prediction has been confirmed: acute tryptophan depletion that probably decreased central serotonin activity increased meal size in patients with bulimia, but not in controls. Because abnormally large meals are characteristic of obesity as well as bulimia nervosa, it is interesting to note that facilitation of serotonin synaptic activity is the primary action of the most efficacious antiobesity drugs, including fenfluramine, dexfenfluramine (withdrawn from the U.S. market), and sibutramine (Meridia).

Finally the defects in central serotonin, gastric distention, and peripheral CCK associated with bulimia may be synergistic, (i.e., they may produce supraadditive effects). This is because in animals the satiating potency of CCK is synergistic with gastric distention and is reduced by decreased central serotonergic function.

SUGGESTED CROSS-REFERENCES

Section 1.4 contains further information of the physiology of monoamine neurotransmitters, and Section 1.6 contains similar information about neuropeptides. Basic learning theory is presented in Section 3.3. Eating and feeding disorders are reviewed in Chapter 20 and Chapter 41; Section 25.3 includes background information on obesity.

SECTION REFERENCES

Anderson GH, Kennedy SH, editors: *The Biology of Feast and Famine: Relevance to Eating Disorders*. Academic Press, New York, 1992.

Booth DA, editor: *Neurophysiology of Ingestion*. Pergamon Press, New York, 1993.

Bouchard C, Bray GA, editors: *Regulation of Body Weight: Biological and Behavioral Mechanisms*. Wiley, New York, 1996.

*Brownell KD, Fairburn CG, editors: *Eating Disorders and Obesity:A Comprehensive Handbook*. The Guilford Press, New York, 1995.

Brunner L, Nick H-P, Cumin F, Baum H-P, Whitebread S, Stricker-Krongrad A, Levens N: Leptin is a physiologically important regulator of food intake. Int J Obes Relat Metab Disord *21:*1152, 1997.

Comuzzie GA: The search for human obesity genes. Science *280:*1374, 1998.

*Devlin MJ, Walsh BJ, Guss JL, Kissileff HR, Liddle RA, Petkova E: Postprandial cholecystokinin release and gastric emptying in patients with bulimia nervosa. Am J Clin Nutr *65:*114, 1997.

Eckel LA, Langhans W, Kahler A, Campfield LA, Smith FJ, Geary N: Chronic administration of OB protein decreases food intake by selectively reducing meal size in female rats. Am J Physiol *275:*R186, 1998.

Flier JS, Maratos-Flier E: Obesity and the hypothalamus: Novel peptides for new pathways. Cell *92:*437, 1998.

Geliebter A, Melton PM, McCray RS, Gallagher DR, Gage D, Hashim SA: Gastric capacity, gastric emptying, and test-meal intake in normal and bulimic women. Am J Clin Nutr *56:*656, 1992.

Greenberg D, Smith GP: The controls of fat intake. Psychosom Med *58:*559, 1996.

Grill HJ, Kaplan JM: Caudal brainstem participates in the distributed neural control of food intake. In *Handbook of Behavioral Neurobiology, vol 10, Neurobiology of Food and Fluid Intake*, EM Stricker, editor. Plenum Press, New York, 1990.

*Harris RB: Parabiosis between db/db and ob/ob or db/ + mice. Endocrinology *140:*138, 1999.

*Kalra S, Dube MG, Pu S, Xu B, Horvath TL, Kalra PS: Interacting appetite-regulating pathways in the hypothalamic control of body weight. Endocr Rev *20:*68, 1999.

Kaye WH, Weltzin TE, McKee M, McConaha C, Hansen D, Hsu LKG: Laboratory assessment of feeding behavior in bulimia nervosa and healthy women: Methods for developing a human-feeding laboratory. Am J Clin Nutr *55:*372, 1992.

Kennedy GC: The role of depot fat in the hypothalamic control of food intake in the rat. Proc R Soc London B Biol Sci *140:*578, 1953.

Kissilef HR, Wentzlaff TH, Guss JL, Walsh BT, Devlin JJ, Thornton JC: A direct measure of satiety disturbance in patients with bulimia nervosa. Physiol Behav *60:*1077, 1996.

Langhans W: Role of the liver in the metabolic control of eating: What we know—and what we do not know. Neurosci Biobehav Rev *20:*145, 1996.

Langhans W: Bacterial products and the control of ingestive behavior: Clinical implications. Nutrition *12:*303, 1996.

Langhans W, Scharrer E: Metabolic control of eating. In *Metabolic Control of Eating, Energy Expenditure and the Bioenergetics of Obesity*, AP Simopoulos, editor. S. Karger, Basel, Switzerland, 1992.

Lavin JH, Wittert G, Sun W-M, Horowitz M, Morley JE, Read NW: Appetite regulation by carbohydrate: Role of blood glucose and gastrointestinal hormones. Am J Physiol *271:*E209, 1996.

Leibel RL, Chung WK, Chua SC: The molecular genetics of rodent single gene obesities. J Biol Chem *272:*31937, 1997.

*Mantzoros CS: The role of leptin in human obesity and disease: A review of current evidence. Ann Intern Med *130:*671, 1999.

Rolls BJ: Effects of food quality, quantity, and variety on intake. In *Not Eating Enough*, BM Marriott, editor. National Academy Press, Washington, DC, 1995.

Schwartz GJ, Moran TH: Integrative gastrointestinal actions of the brain-gut peptide cholecystokinin in satiety. In *Progress in Psychobiology and Physiological Psychology*, SJ Fluharty, AR Morrison, editors. Academic Press, New York, 1995.

Sclafani A: How food preferences are learned: Laboratory animal models. Proc Nutr Soc *54:*419, 1995.

*Smith GP, editor: *Satiation from the Gut to the Brain*. Oxford University Press, New York, 1998.

Smith GP: Control of food intake. In *Modern Nutrition in Health and Disease*, ed 9, ME Shils, JA Olsen, M Shike, AC Ross, editors. Williams & Wilkins, Baltimore, 1998.

Smith GP: Dopamine and food reward. In *Progress in Physiological Psychology and Psychobiology*, vol 16, SJ Fluharty, AR Morrison, J Sprague, editors. Academic Press, San Diego, 1995.

*Smith GP: The direct and indirect controls of meal size. Neurosci Biobehav Rev *20:*41, 1996.

Weltzin TE, Fernstrom MH, Fernstrom JD, Neuberger SK, Kaye WH: Acute tryptophan depletion and increased food intake and irritability in bulimia nervosa. Am J Psych *152:*1668, 1995.

*Weltzin TE, Fernstrom MH, Kaye WH: Serotonin and bulimia nervosa. Nutr Rev *52:*399, 1994.

Woods SC, Seeley RJ, Porte D, Schwartz MW: Signals that regulate food intake and energy homeostasis. Science *280:*1378, 1998.

1.21 FUTURE DIRECTIONS IN NEUROSCIENCE AND PSYCHIATRY

SOLOMON H. SNYDER, M.D.

The rapid advances in molecular neurosciences over the past two decades are continuing to accelerate as the tools provided by the field of molecular biology become ever more powerful. These tools will enable the development of more selective therapeutic agents with fewer adverse effects. More tantalizing and less certain is the possibility that the abnormal genes that play a role in major mental illness will be deciphered. The major genes involved in schizophrenia and mood disorders will be pinned down well before the end of 2010. However, this does not mean that the "cause" of these diseases will be fully elucidated; most investigators suspect that there will be as many different causes of the major psychotic disorders as there are of pneumonia.

ABNORMAL GENES

The technique of *reverse genetics* in which deoxyribonucleic acid (DNA) sequences and blood cells of extended families are analyzed to look for correlations with the disease process, has achieved great success recently. Specific molecular genetic abnormalities have been detected for cystic fibrosis, Duchenne's muscular dystrophy, and several other conditions. Numerous groups have been employing reverse genetic techniques to seek specific abnormalities in patients with mood disorders and schizophrenia. There have been reports of "clues," some of which have been replicated in independent investigations. Unique patterns on chromosome 18 have been detected by several researchers into mood disorders and there are more preliminary hints of abnormalities in schizophrenia. Advances in Alzheimer's disease have been substantial and may bolster efforts to pin down the molecular genetics of the principal psychiatric disorders. The presence of amyloid plaques has long been known to be a hallmark of Alzheimer's disease. However, amyloid deposits occur in numerous diseases and seem to represent a form of "scar" following neuronal damage. Hence, one school of thought has argued that amyloid deposition is merely secondary to some other primary event. With the first elucidation of specific abnormal genes in certain hereditary forms of Alzheimer's disease came the "Eureka" discovery that some of these mutations involved steps in processing the amyloid precursor protein. Although these abnormalities account for only a small percentage of the total population of Alzheimer's disease patients, amyloid is increasingly believed to be causally involved in a major portion of the disease. Because the specific lesion need not be in the amyloid protein precursor itself, the most prevalent aberration discovered in Alzheimer's disease occurs in the presenilins, membrane-associated proteins whose normal biological role is unknown and may not relate directly to amyloid. However, presenilin mutations lead to disordered amyloid processing. Analogously, some cases of depression may derive from abnormalities in serotonin metabolism whereas others could involve distinct systems whose malfunction secondarily disrupts serotonin mechanisms.

Similar reasoning might be applied to the dopamine hypothesis of schizophrenia. The major antipsychotic drugs act by blocking dopamine receptors, and drugs such as amphetamines that release dopamine can exacerbate psychotic symptoms. Based on these findings, researchers have looked for specific dopamine abnormalities in schizophrenia, and there have been reports of increased levels of certain forms of dopamine receptors in drug-free patients with schizophrenia. However, the fact that dopamine-related drugs modulate symptoms in no way directly implicates dopamine in causation. Nonetheless, like amyloid in Alzheimer's disease the impressive body of circumstantial evidence on dopamine might turn out to be of fundamental importance.

Even if the specific abnormality in a disease is found understanding of causation or any tools to treat patients may remain elusive. In 1983 restriction fragment length polymorphism studies pointed toward the abnormal gene in Huntington's disease, and after a decade-long search, this gene was identified and was found to code for a large protein designated *huntingtin*. Unfortunately, nothing about huntingtin has explained the pathophysiology of Huntington's disease. The caudate nucleus of patients with Huntington's disease shrivels selectively. However, huntingtin concentrations are uniform throughout the brain and the protein occurs in high densities all over the body. There are now some valuable clues that might contribute to our understanding of psychiatric disorders. The molecular disturbance in Huntington's disease involves increased numbers of glutamine molecules occurring in repeats of up to 80 or more in huntingtin, whereas normal subjects have only a few of these repeats. Researchers have speculated that the glutamine repeats cause huntingtin to bind to other proteins that are the real neurotoxic culprits. One protein that binds selectively to huntingtin, especially forms of huntingtin that have larger numbers of glutamine repeats, is designated *huntingtin-associated protein* (HAP). In contrast to the ubiquitous distribution of huntingtin, HAP occurs only in the brain and displays discrete regional localizations similar to those of neuronal nitric oxide synthase, which has been implicated in neurotoxicity. The glutamine repeats in huntingtin also bind to glyceraldehyde-3-phosphate dehydrogenase. This enzyme is usually thought of only as a housekeeping enzyme of the glycolytic pathway, but it also appears to be a messenger protein that interacts with numerous other macromolecules in the cell, including serving as a transcription factor in the nucleus. Thus HAP, glyceraldehyde-3-phosphate dehydrogenase, or another protein may be the ultimate mediators of damage, and nitric oxide synthase may also play a role. Such a chain of functionally and physically associated proteins may help to explain seemingly conflicting findings in various diseases.

Identifying protein-protein interactions has become increasingly important in all aspects of biology. Various proteins, such as actin and myosin, are known to bind to each other, but the extent of large multiprotein complexes has not been adequately appreciated. The yeast-two hybrid system is an extremely powerful tool that enables scientists to identify novel protein-protein interactions by evaluating the ability of the proteins synthesized from the complementary deoxyribonucleic acid (cDNA) of a known protein to interact with a large universe of proteins synthesized from a cDNA library of any chosen tissue such as the brain; HAP was discovered by the yeast-two hybrid approach.

Investigating the neuropathology of mental illness was long regarded as a dead-end enterprise but newer imaging techniques have revealed specific abnormalities; for example, increased lateral ventricular size has been observed repeatedly in chronic schizophrenia and specific abnormalities have been observed in portions of the temporal lobe as well as in the more primitive areas of the frontal lobes. The patterns of disturbance suggest an aberration of development rather than a neurodegenerative process, which fits with the contrast between schizophrenia and Huntington's disease, the latter being a classic neurodegenerative condition. Patients with Huntington's disease seem to be completely normal until degeneration sets in; by contrast, children destined to develop schizophrenia appear to be "not quite right" from earliest childhood. One might speculate that abnormal genes in schizophrenia could influence neural development of the temporal lobe

MAKING NERVES GROW

Given the likelihood that major mental illnesses are neurodegenerative or neurodevelopmental, can anything be done therapeutically for such disturbances? Although drugs can improve symptoms in neurodegenerative disease (such as cholinesterase-inhibiting drugs in Alzheimer's disease), the present generation of drugs is probably incapable of restoring disordered neuronal patterns. In the ideal situation nerves that never developed properly or those that degenerated would grow; this concept may no longer be pure fantasy.

Research on neurotrophic factors has burgeoned in recent years. Nerve growth factor (NGF) was the first to be discovered, followed by brain-derived neurotrophic factor (BDNF), neurotrophin 3 and 4 (NT3 and NT4), and, more recently, glial-derived neurotrophic factor (GDNF). Receptors for these factors have been characterized at a molecular level as deriving from the family of genes designated *trk*, which are related to growth-stimulating oncogenes.

These neurotrophic factors cannot readily be employed as therapeutic agents, because they do not penetrate blood-brain barrier and are extensively metabolized in the body. However, recombinant techniques make possible the synthesis of large amounts of these proteins, which are now in advanced clinical trials (e.g., the insulin-like growth factor [IGF-1, myotrophin] for amyotrophic lateral sclerosis). An initial clinical trial of NGF administered subcutaneously has revealed its therapeutic efficacy in diabetic neuropathy. In the brain NGF selectively stimulates the growth of cholinergic neurons. Loss of the cholinergic pathway to the cerebral cortex whose cell bodies are in the basal nuclei of the forebrain accounts for much of the memory deficit in patients with Alzheimer's disease. Lesions of this pathway produce memory losses in animals. Injection of NGF into the cerebral ventricles stimulates the regrowth of the cholinergic neurons selectively and alleviates the memory deficit. Clinical trials of NGF administered by an in-dwelling cannula into the brains of Alzheimer's patients are anticipated.

Neurotrophic actions of small-molecule drugs that pass the blood-brain barrier and are orally available have been observed recently in the course of research on the immunophilins. Immunophilins are the protein receptors for classic immunosuppressant drugs such as cyclosporin-A and FK506. Immunosuppression derives from the

binding of the drug-immunophilin complex to the calcium-activated phosphatase calcineurin, inhibiting its ability to remove phosphate groups from certain proteins. One of these proteins, *nuclear factor of activated T cells* (NFAT), is a transcription factor that can only enter the nucleus when it is dephosphorylated. By blocking calcineurin, phosphatase activity, treatment with the immunosuppressant drugs increases phosphorylated levels of NFAT; the latter cannot enter the nucleus to stimulate formation of the cytokine interleukin-2, which normally initiates the immune response of T cells.

Immunophilin levels are up to fifty times higher in the brain than in tissues of the immune system and are especially enriched in the growth cones of regenerating neurons. These findings led to experiments revealing the very potent neurotrophic actions of immunosuppressant drugs in stimulating the growth of various neuronal systems such as dorsal root ganglia. The drugs stimulate the growth of all types of damaged neurons, including the sciatic and facial nerves and dopamine and serotonin neurons in the brain. The neurotrophic actions also occur with drug derivatives whose complexes with the immunophilin do not bind to calcineurin so that the drugs are not immunosuppressant. Although these drugs show promise in animal models of damaged neurons, clinical trials have not yet been conducted to assess their therapeutic relevance.

SUGGESTED CROSS-REFERENCES

Section 1.4 and Section 1.5 discuss neurotransmitters, Section 1.6 discusses neuropeptides, Section 1.14 discusses molecular neurobiology, and Section 1.18 discusses genetic linkage analysis. Chapter 12 covers Schizophrenia. Chapter 10 discusses cognitive disorders and medical conditions, including Huntington's disease and Alzheimer's disease.

SECTION REFERENCES

*Alzheimer's Disease Collaborative Group: The structure of the presenilin 1 (S182) gene and identification of six novel mutations in early-onset AD families. Nature Genet *11:*219, 1995.

Andreason NC: Pieces of the schizophrenia puzzle fall into place. Neuron *16:*697, 1996.

*Andreasen NC: Understanding the causes of schizophrenia. New Engl J Med *340:*645, 1999.

*Barondes SH: *Molecules and Mental Illness*. Scientific American Press, New York, 1993.

Fields S, Song O: A novel genetic system to detect protein-protein interactions. Nature *340:*245, 1989.

*Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72:*971, 1993.

Karayiorgou M, Gogos JA: A turning point in schizophrenia genetics. Neuron *19:*967, 1997.

Li XJ, Li SH, Sharp AH, Nucifora FC, Schilling G, Lanahan A, Worley P, Snyder SH, Ross CA: A huntington-associated protein enriched in brain with implications for pathology. Nature *378:*398, 1995.

*Ross CA, Pearlson GD: Schizophrenia, the heteromodal association neocortex and development: Potential for a neurogenetic approach. Trends Neurol Sci *19:*171, 1996.

Scheuner D, Eckman C, Jensen M, Song X, Citron M, Suzuki N, Bird TD, Hardy J, Hutton M, Kukull W, Larson E, Levy-Lahad E, Viitamen M, Peskind E, Poorkaj P, Schellenberg G, Tunzi R, Wasco W, Lannfelt L, Selkoe D, Younkin S: Secreted amyloid b-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease. Nature Med *2:*864, 1996.

*Sterner JR, Connolly MA, Valentine HL, Hamilton GS, Dawson TM, Hester L, Snyder SH: Neurotrophic actions of nonimmunosuppressive analogues of immunosuppressant drugs FK506, rapamycin and cyclosporin A. Nature Medicine *3:*421, 1997.

CHAPTER 2. NEUROPSYCHIATRY AND BEHAVIORAL NEUROLOGY

2.1 NEUROPSYCHIATRY: CLINICAL ASSESSMENT AND APPROACH TO DIAGNOSIS

TIFFANY W. CHOW, M.D., AND JEFFREY L. CUMMINGS, M.D.

Neuropsychiatry is the clinical discipline devoted to understanding and treating brain-based behavioral disturbances. It encompasses human behavior, emotion, cognition, and their physical bases. The traditions of biological psychiatry gave rise to neurology when clinicians observed abnormal human behaviors and searched for causes based on neuroanatomy. The neuropsychiatrist must master psychopharmacology and neuroscience in order to understand the physical basis of behavior and to benefit patients with behavioral disturbances produced by central nervous system disorders. This scientific pursuit combines research, education, and clinical patient care. Neurologists must heed the behavioral aspects of neurological disorders whereas psychiatrists cannot expect to fully comprehend behavior without appreciating its physical underpinnings; all effective therapeutic interventions for abnormal behavior must eventually be translated into neurological events. Contemporary neuropsychiatry investigates not only the neurological basis of psychiatric disorders such as schizophrenia, bipolar illness, depression, anxiety, and obsessions and compulsions but also explores the behavioral alterations exhibited by patients with overt brain disorders such as stroke, epilepsy, and multiple sclerosis. This introductory section reviews behavioral aspects of neurological disorders and presents a comprehensive approach to neuropsychiatric assessment.

MENTAL STATUS EXAMINATION

The key to accurate neuropsychiatric diagnosis is a comprehensive mental status examination. Mental status testing, augmented by the clinical history, will lead to a specific diagnosis or to the generation of diagnostic hypotheses. The neurological examination, selected laboratory studies, the success of treatment interventions, and repeated evaluation over time will eventually complete the diagnostic process. Effective management demands accurate diagnosis. The neuropsychiatrist must make an accurate behavioral assessment in order to identify and manage the underlying brain disorder.

The elements of a comprehensive mental status examination include observation of the dress and spontaneous demeanor of the patient, evaluation of mood, assessment of thought form (including formal thought disorder, obsessional thinking, perseveration, and others) and thought content (delusions, hallucinations, preoccupations, and others). This is followed by a more structured probing of cognitive domains that include attention, language, memory, visuospatial and constructional abilities, calculation, abstraction and judgment, and frontal-systems skills. Interview schedules and rating scales may be used to standardize and quantify the observations. Referral to a neuropsychologist or speech pathologist may be indicated for further testing of neuropsychological and communicative abilities. Table 2.1-1 provides an outline of a thorough mental status examination.

Table 2.1-1 Neuropsychiatric Comprehensive Mental Status Evaluation

In neuropsychiatric assessment, behavioral phenomena are explored and cognitive functions probed as a means of investigating brain function. Although complex behavior is a product of integrated brain activity, individual brain regions make specialized contributions to each human capacity and may produce recognizable syndromes when injured. Structured neuropsychiatric assessment of the major brain regions synthesizes the information into a comprehensive behavioral and etiological formulation. Cortical or subcortical, right hemisphere or left hemisphere, and anterior or posterior hemispheric location and emotional and cognitive dimensions are systematically evaluated.

Observational Assessment Even a casual interaction with the patient provides an opportunity for neuropsychiatric data collection. Patients with dementia, delirium, or frontal lobe syndromes may appear disheveled and unkempt whereas patients with schizophrenia-like illnesses or secondary mania may be eccentric or flamboyant. Patients with unilateral neglect following lateralized brain injury may ignore objects or people in one hemi-universe and fail to dress or groom one half of the body. Signs of incontinence may be evident in patients with frontal lobe syndromes or dementia. The patient's spontaneous demeanor may evidence anxiety (e.g., hyperthyroidism), impulsivity (e.g., orbitofrontal syndromes), or apathy (e.g., medial frontal lobe syndromes). Social interactions are revealing: patients with delusional disorders may be suspicious and establish poor rapport, patients with orbitofrontal dysfunction may be unduly familiar with the examiner, and patients with delirium or frontal lobe disorders may be disinhibited and distractible. Observing the patients' interactions with their spouses is critical: patients with dementias such as dementia of the Alzheimer's type (also called Alzheimer's disease) become increasingly dependent on their spouse for information and automatically turn to them for answers during the examination. Observation of the patient's affect may provide additional information. A variety of context-incongruent affects may be manifested by patients with brain dysfunction including jocularity, sadness, irritability, anger, affective lability, and perplexity. Each of these affective abnormalities may have diagnostic or localizing significance. Verbal output including speech (the mechanical aspects of communication) and language (the propositional content of communication) also provides diagnostic information. Abnormalities of speech include mutism, dysarthria, increased or decreased voice volume, stuttering, abnormal rate of speech, and disturbed speech prosody (loss of melody or inflection). Language disturbances in neuropsychiatric syndromes include aphasia (fluent or nonfluent), echolalia (echoing what the examiner says), and palilalia (repeating oneself). Patients with Tourette's disorder will have involuntary vocalizations ranging from subtle sighs, sniffs, grunts, and barks to shouting and coprolalia (swearing). A wide variety of diagnostically important motor behaviors may be evident during the examination, including movement disorders (tremor, chorea, parkinsonism, tics, myoclonus, tremor, gait abnormalities), compulsions, stereotypies, hyperactivity, or absence spells. Alterations of the patient's posture may also be diagnostically meaningful. Patients with Parkinson's disease have a flexed, stooped posture whereas patients with progressive supranuclear palsy have an extended, upright posture. Finally, the patient's psychomotor speed should be noted. Response latencies may be prolonged or the patient may interrupt the examiner impulsively, anticipating the question. Speech and movement may be slowed or executed with abnormal rapidity. Slowing of central processing is typical of subcortical neurological disorders, medial frontal syndromes, and depression. Patients with secondary mania or disinhibited orbitofrontal syndromes may have accelerated speech and increased motor activity.

Neuropsychiatric Interview The neuropsychiatric interview includes a more formal evaluation of the patient's mood and affect as well as thought form and thought content. The interview is conducted in the same way as a conventional psychiatric assessment, but a wider variety of phenomena may be observed in neuropsychiatry patients because of the expression of abnormalities resulting from the concomitant delirium, dementia, aphasia, or amnesia.

Disturbances of Mood *Mood* refers to an internally experienced pervasive emotion; *affect* is the outward emotional display. Mood disturbances commonly occur in the course of neurological disease. Depression is the most frequent mood disturbance, attending stroke, Parkinson's disease and other movement disorders, epilepsy, and multiple sclerosis. Patients often evidence depressive symptoms of sadness and feelings of hopelessness, helplessness, worthlessness, and guilt. Loss

of interest in work, hobbies, or recreational activities may be marked; patients may acknowledge loss of energy and fatigue. Associated disorders include insomnia (difficulty falling asleep, frequent nocturnal awakenings, early-morning awakening), alterations in appetite, loss of libido, and motor behavior changes (agitation, slowed movement and speech, reduced voice volume). Suicidal ideation may range from passive wishes for death to directed suicidal activities. Among neuropsychiatric illnesses, Huntington's disease and complex partial epilepsy have substantially increased suicide rates. Typical changes of affect accompanying depressed mood include tearfulness and sobbing, but the overall range of mood and affect may be reduced.

Euphoria, hypomania, or mania may also occur in neuropsychiatric disorders. Euphoria is most common with frontal lobe dysfunction (trauma, frontal lobe tumors, frontal lobe degenerative diseases, frontal infections) and with secondary mania. The patient with mania or hypomania may exhibit an elevated mood, undue optimism, inflated self-esteem, grandiosity, exaggerated confidence, and motoric hyperactivity. Distractibility, flight of ideas, racing thoughts, and pressured speech may be evident. Insomnia, increased appetite, and exaggerated libido may also occur.

Anxiety occurs in a variety of neuropsychiatric conditions including metabolic encephalopathies (e.g., hyperthyroidism, anoxia), toxic disorders (e.g., lidocaine toxicity), and degenerative disorders (e.g., Alzheimer's disease, frontotemporal dementia, Parkinson's disease). The patient exhibits excessive worry and apprehension, undue pessimism, and feelings of impending death or disaster. Somatic symptoms including palpitations, tachycardia, shortness of breath, lightheadedness, fleeting pains, dry mouth, nausea, diarrhea, and perspiration are often prominent. The patient may appear tense with furrowed brow and worried expression; tremor, tachypnea, pupillary dilatation, and restless fidgeting may be observed.

Mood lability is a disorder of mood regulation observed in neuropsychiatric syndromes. Patients may shift rapidly from sadness to jocularity. Typically, patients with marked lability are irritable and alternate among anger, apathy, and euphoria. The emotional outbursts are usually shortlived. Lability is common in patients with frontal lobe dysfunction or mental retardation. In the mood lability syndromes, mood and affect are congruent (i.e., the patient both feels anger and appears angry). On the other hand, the affective instability of pseudobulbar palsy is marked by incongruent mood and affect, as described below.

Mood and affect may be dissociated in neuropsychiatric disorders. Pseudobulbar palsy is a disorder of affect in which the patient's emotional expression is a gross exaggeration of or completely at variance with the patient's mood. The patient may weep in minimally sad situations or in the absence of an appropriate stimulus. Alternatively, the patient may laugh when sad or when no humorous event has occurred. The abnormal pseudobulbar affect is accompanied by the loss of other bulbar functions normally under supranuclear control and manifested by dysarthria, increased gag response, increased jaw jerk and facial muscle stretch reflexes, and facial weakness with hypomimia. Congruence between mood and affect should be explored in patients with neuropsychiatric illness, and the associated neurological signs of pseudobulbar palsy must be sought during the physical examination. In most cases patients with pseudobulbar palsy have bilateral lesions of the descending corticobulbar tracts. Common causes are cerebrovascular disease, multiple sclerosis, amyotrophic lateral sclerosis, traumatic brain injury, and brainstem or skull base tumors.

Disorders of Thought Form Formal thought disorders are less common than delusions as psychotic manifestations of neuropsychiatric diseases, but classic thought disorders have been observed in the schizophrenia-like illnesses accompanying epilepsy, Huntington's disease, and idiopathic basal ganglia calcification. Tangentiality, circumstantiality, loose associations, illogicality, derailment, and thought blocking may be observed. Flight of ideas occurs in secondary mania; circumstantiality is evident in some patients with temporal lobe epilepsy or alcohol-induced persisting dementia.

The thought form disorders common to neuropsychiatric conditions are perseveration, intrusions, and incoherence. *Perseveration* refers to the inappropriate continuation of an act or thought after its proper context has passed. *Intrusions* are late recurrences of words or thoughts from an earlier context. Perseverations and intrusions are seen in aphasias, frontal lobe syndromes, and dementing illnesses. *Incoherence* occurs when there is no logical association between words or ideas. Incoherence occurs in extreme cases of psychosis but is more common in delirium and cortical dementia such as dementia of the Alzheimer's type.

Disorders of Thought Content Many disorders of thought content may be observed in neuropsychiatric diseases. Delusions are the principal manifestation of psychosis in neurological disorders. A delusion is a false belief based on incorrect inference about external reality, and the delusion is firmly held in spite of evidence to the contrary. The most common types of delusions are persecutory beliefs involving the conviction of personal endangerment, theft of personal property, spousal infidelity, or entry by unwelcome strangers into one's home. Content-specific delusions such as *Capgras's syndrome* may also be observed in neurological illnesses, however, and the type of delusion is not helpful in distinguishing schizophrenia or other idiopathic psychotic illnesses from the delusions of neurological disease. Table 2.1-2 lists and defines the content-specific delusions that have been reported in neuromedical disorders with delusions.

Delusion	Content
Capgras's syndrome	A significant other (usually a family member) has been replaced by an identical-appearing imposter
Erotomanic (de Clerambault's syndrome)	The individual is secretly loved by another, usually someone of higher social or economic status
Dorian Gray	Others are aging while the individual appears to remain the same age
Persecutory (Fregoli's phenomenon)	A persecutor is able to assume the appearance of others
Doppelganger	The individual has a twin or second self
Delusional incubus or succubus	Demon or phantom sits on top of or has sex with the individual during sleep
Intermetamorphosis	Familiar persons takes on the appearance of tormentors
Lycanthropy (werewolfism)	The individual periodically turns into an animal (werewolf)
Jealous (Othello syndrome)	The individual's spouse is unfaithful
Parasitosis	The individual is infested with insects, worms, lice, vermin
Delusional phantom boarder	Unwelcome guests are living in the individual's home

Table 2.1-2 Content Specific Delusions Observed in Neuropsychiatric Disorders

Schneiderian first-rank symptoms are delusional experiences that include hearing one's thoughts spoken aloud, hearing voices arguing about oneself, hearing voices commenting on one's actions, having bodily sensations imposed from outside, attributing one's feelings to external sources, experiencing one's drive and actions as controlled from the outside, having one's thoughts inserted or withdrawn from the mind, broadcasting one's thoughts, and attributing special delusional significance to one's perceptions. Although most common in schizophrenia, first-rank symptoms have been reported in a number of neurological illnesses with schizophrenia/like manifestations including epilepsy, Huntington's disease, and idiopathic basal ganglia calcification.

Hallucinations are common manifestations of neuropsychiatric illness. They are sensory perceptions occurring without appropriate external stimulation of the relevant sensory organ. Hallucinations and delusions may occur together as elements of a psychotic process, or hallucinations may be recognized by the nondelusional patient as unreal. Hallucinations by themselves are neither necessary nor sufficient evidence of psychosis. Hallucinations may occur in any sensory modality, including visual, auditory, tactile (formication), gustatory, and olfactory. Somatic hallucinations are sensory experiences involving the patient's body and internal organs. Hallucinations may be formed, well-defined and recognizable sensations (e.g., visual hallucinations of people or animals; auditory hallucinations of voices) or unformed, elementary sensations (visual hallucinations of flashing light or colors; auditory hallucinations of unrecognizable noises).

Obsessions are another type of anomalous thought content. The patient experiences recurrent, intrusive thoughts, images, or impulses that are ego-dystonic and involuntary. The obsessional thoughts typically involve violent, sexual, or visceral-eliminative themes. Patients with obsessions frequently also have compulsions—repetitive, stereotyped behaviors that are recognized as senseless but cannot be resisted. Washing and checking compulsions are the most common, although nearly any action can be ritualized and become the object of compulsive repetition.

Depersonalization and *derealization* also are examples of altered thought content that refer to dissociative experiences. In derealization the patient may feel as if in a dream. Depersonalization is a state of altered experience in which the feeling of one's own reality is temporarily lost. There may be a loss of ability to experience emotion and a disturbed perception of time. Neuropsychiatric disorders to be considered in the differential diagnosis of depersonalization include epilepsy, migraine, encephalitis, and systemic metabolic disorders.

Mental Status Examination Systematic probing of the patient's mental state is a crucial part of neuropsychiatric assessment. Diagnostic hypotheses generated while taking the history and conducting the neuropsychiatric interview are further refined by mental status testing. The examination should be conducted methodically and should comprehensively assess the major domains of neuropsychological function (Table 2.1-1). Where possible, the examination should be quantified to facilitate

comparisons among patients and documentation of changes in the patient's function over time. The patient's age, handedness, educational level, primary language, and sociocultural background may all influence the mental status examination and should be determined prior to initiating the evaluation.

Alertness and Attention The patient's level of arousal and attention are evaluated by observation and testing. Levels of hypoarousal descend from drowsiness to obtundation, stupor, and then coma. The *drowsy* patient is fatigued and falls asleep when unstimulated. *Obtundation* is a state of moderately reduced alertness with slow responses and diminished interest in the environment. *Stupor* is the next most severe level of impaired consciousness. Stuporous patients must be vigorously stimulated to be aroused and engaged in conversation. Stupor returns when stimulation ceases. *Coma* is state of unarousable unresponsiveness. Comatose patients have no psychological response to external stimulation or internal needs. Impaired arousal occurs in toxic-metabolic encephalopathies, increased intracranial pressure, large strokes, encephalitis, and traumatic brain injury. Alternately, *hyperarousal* with anxiety, autonomic hyperactivity (tachycardia, tachypnea, hyperthermia), and exaggerated startle responses may occur in metabolic disorders, particularly during withdrawal from alcohol, opiates, or sedative-hypnotic agents.

Stupor must be distinguished from akinetic mutism, locked-in syndrome, and catatonic unresponsiveness. *Akinetic mutism* is a state of silent, alert-appearing immobility. The patient's eyes are open and may follow environmental events; there are regular sleep-wake cycles. The patient may be completely inert or may have occasional brief movements to adjust posture spontaneously or in response to vigorous stimulation. The disorder has occurred with large frontal lobe injuries, bilateral cingulate gyrus damage, and midbrain pathology. *Locked-in syndrome* occurs with bilateral pontine lesions that render the patient mute and paralyzed. Intellectual function, however, is not impaired and the patients can communicate by eye movements or eye blinks (e.g., blinking once for “yes” and twice for “no”). In *catatonic stupor* the patient is mute and motionless. A psychological pillow and waxy flexibility may be elicited. Brief impulsive acts may punctuate the catatonic period. Catatonia occurs in a wide variety of neurological, toxic-metabolic, and psychiatric disorders including frontal lobe syndromes, hypoglycemia, schizophrenia, and depressive disorders.

Attention is impaired in disorders of arousal but may also be abnormal in patients who are fully alert. Delirium and acute confusional states are characterized by abnormal attention in alert patients. Two types of tests are particularly useful in detecting compromised attention: *forward digit span* and *continuous performance tests.* In testing digit span, the patient must repeat increasingly long series of numbers spoken aloud by the examiner. The numbers are given at a rate of one per second in a monotone voice. A normal forward digit span is seven digits long; fewer than five is abnormal and indicates impaired attention. Sustained attention is evaluated by a continuous performance test such as the “A” test, in which the clinician says a series of random letters at a rate of one per second. The patient is asked to signal each time the letter “A” occurs. The test continues for 1 minute and assesses the patient's ability to sustain concentration over time. Distractible patients make errors of omission, failing to detect one or more of the *A*s. Individuals with normal attention perform the test perfectly. Acute confusional states and impaired attention are usually indicative of a toxic or metabolic disorder but also occur with increased intracranial pressure, frontal lobe disorders, or focal lesions of the posterior right hemisphere.

Mental control tests are tasks that require intact attention but also depend on other intellectual resources. Serial subtractions, spelling words backwards, reverse digit span, and reciting the days of the week or months of the year in reverse order are examples of tests used to investigate mental control; both the accuracy and the speed of performance are observed. Attention, language, frontal lobe functions, and calculation skills contribute to mental control abilities. When instrumental functions (language, calculations) are intact, impaired mental control is usually indicative of a frontal systems disorder.

Impaired attention conditions all aspects of behavior and intellectual function. Patients who are drowsy or in acute confusional states will perform badly on comprehension, repetition, drawing, and calculation tests because of the inability to sustain attention. In general, conclusions regarding a potential underlying dementia or memory disturbance cannot be drawn while the patient is in a confusional episode. Management efforts are directed to controlling the associated disorder and restoring attention; reexamination will then reveal the baseline mental condition.

Memory Aspects of memory to be assessed in the course of mental status examination include learning, recall, recognition, and memory for remote information. Learning, recall, and recognition can be assessed in both verbal and nonverbal domains.

Inquiring about personal orientation is one means of assessing recent memory. Temporal and spatial orientation must be continuously updated and patients with memory abnormalities may fail to establish the current day, date, or place. Orientation to self is rarely lost and has little value as a memory test.

Word-list memory tests are a common means of assessing learning, recall, and recognition. The patient is given three words to remember and then is asked to recall them 3 minutes later. The examiner notes how much difficulty the patient has in learning the three words initially and how many times the words must be presented before the patient can repeat all three. A few minutes later, the patient is asked to remember the three words. If he or she can do so, the performance is normal. If the words cannot be remembered, the patient is given clues (e.g., the category of items to which the word belongs or a multiple-choice list of words containing the target and several foils) to distinguish between storage and recall deficits. Patients with encoding or registration deficits (e.g., amnesia) will not find prompting and clues helpful; patients with intact storage but impaired retrieval (e.g., patients with frontal-subcortical circuit disturbances) are aided by these maneuvers. Memory performance patterns are often more rigorously elicited with a ten-word list rather than a three-word list.

Nonverbal memory can also be tested to distinguish between encoding and access impairments. Patients are asked to copy several figures (usually as part of the constructional task described below); a few minutes later the patients must reproduce the drawings from memory; if they cannot spontaneously remember the drawings, the patients may choose from a group of figures containing the targets and several foils. Patients who cannot recognize the original drawings have a nonverbal encoding disturbance as part of an amnestic disorder; patients who cannot recall the drawings but are able to recognize them within a group of stimuli and foils have a nonverbal memory access abnormality resulting from a frontal-subcortical circuit disorder.

Remote memory testing evaluates the patient's ability to remember events and individuals of the remote past. This aspect of memory function may be gauged while taking a history of the patient's illness, by inquiring about the patient's significant life events (e.g., dates of marriage, birth dates of children), and by asking about political leaders and important historical events. Although remote information is not stored in specific focal brain areas and recall failures do not provide localizing information, the temporal profile of remote memory may be diagnostically revealing. Amnestic disorders are characterized by a period of anterograde amnesia following the onset of the disorder (this may continue indefinitely if the patient fails to recover or time-limited if memory function returns), a period of retrograde amnesia extending backwards in time for a few seconds to a few years from the time of onset of the amnesia, and intact remote recall for material beyond the period of the retrograde amnesia. Patients with dissociative amnesia usually complain of profound retrograde but not anterograde amnesia. Patients with Alzheimer's disease demonstrate a temporal gradient, progressively losing the episodic memories from more recent to more remote.

Language Language assessment includes observing the patient's spontaneous speech and testing language comprehension, repetition, naming, reading, and writing. Abnormalities of language must be distinguished from disorders of speech *(dysarthria)*, abnormalities of inflection and melody *(aprosodia)*, and absence of linguistic output *(mutism).* Language-related abilities that may be evaluated include speech prosody, automatic speech, singing, and right-left orientation. Language disturbances are usually observed in the course of taking the history and conducting the neuropsychiatric interview, and these patients will require the most detailed testing. When the verbal exchange has been unremarkable, testing of a few difficult items of the major linguistic categories is adequate to establish normal language function.

Spontaneous speech reveals abnormalities of thought and language and the two types of abnormalities must be carefully distinguished. Disorders of thought form and content have been discussed above. Aphasic disturbances of spontaneous verbalization are characterized as fluent or nonfluent. *Fluent aphasias* feature normal or excessive amounts of speech, normal phrase length with complex sentence structures, preserved speech melody, no dysarthria, a paucity of information content with excessive dependence on words of indefinite reference such as “thing” and “it,” and paraphasic errors. Paraphasias may involve the substitution of one phoneme for another (e.g., “nook” for “book”) in phonemic (literal) paraphasias; the replacement of one word with another in verbal (semantic) paraphasias; or the construction of new words in neologistic paraphasias. *Nonfluent aphasias* exhibit essentially the opposite features of fluent aphasias. The patients have reduced verbal output, shortened phrase lengths or one-word replies, agrammatism with a tendency to omit the syntactic functor words (e.g., “and,” “but,” “the”), increased articulatory effort and impaired speech initiation, reduced speech melody, and dysarthria. There are few or no paraphasias and information content is relatively normal. Inflection of the remaining limited verbal output may be preserved. The individual aphasia syndromes are described below.

Language comprehension is tested by asking the patient to follow verbal commands, respond to “yes” and “no” questions, or decipher complex linguistic constructions. The easiest commands are one-step orders involving whole body responses such as “stand up” and “turn around.” Next in order of difficulty are single-step commands involving midline body parts such as “open your mouth,” “close your eyes,” and “stick out your tongue.” Pointing to room objects and body parts assesses the next level of comprehension. Patients are asked to point to one, two, or three objects visible in the examination room (e.g., “point to the ceiling, then to the floor, then to the desk”) or to one, two, or three body parts (e.g., “point to your nose, then to your shoulder, then to your knee”). Pointing to objects in an array of items laid in front of the patient presents a igher level of difficulty for most patients than sequential pointing to room objects or body parts. An array such as a comb,

pen, coin, and keys is provided, and the patient is asked to point to several of the items in turn (e.g., “point to the keys, then to the comb, then to the pen”). Each time the test is repeated, the item order is changed. Pointing tasks may be made more difficult by asking the patient to point to an item that is described rather than named (e.g., “point to the entrance to the room” or “point to the source of illumination”). Pointing is impossible for some patients because of paralysis or apraxia, and “yes” and “no” questions may be used to minimize the motor demands of the task. These may range from easy (e.g., “are the lights on in this room?”) to modestly difficult (e.g., “do you put our shoes on before your socks?”). Finally, complex questions such as “Is my wife's brother a man or a woman?” or “if a lion is killed by a tiger, which animal is dead?” allow the examiner to determine if the patient can follow complex linguistic formulations. Impaired comprehension usually implies dysfunction of posterior structures of the dominant or left hemisphere. Comprehension is abnormal in many aphasic syndromes, dementia, and delirium.

Repetition is assessed by asking the patient to repeat increasingly long phrases or sentences beginning with a single word and progressing to sentences eight to ten words in length (e.g., “the quick brown fox jumped over the lazy dog”). Abnormalities include omissions and paraphasic substitutions. The repetition span will typically be one or two words longer than the digit span, and patients with shortened digit spans will be able to repeat only brief phrases. Conversely, the digit span is a repetition test, and patients with repetition disturbances (e.g., several of the aphasias) will have abnormal digit spans.

Assessment of naming involves asking the patient to name objects and parts of objects, body parts, pictures or drawings of objects, and colors. Naming is compromised systematically with impairment of rarely used low-frequency words impaired prior to commonly used high-frequency words. Thus, a thorough language examination must include naming of low- and high-frequency words. As a rule of thumb, objects are of higher frequency than the parts of the objects (e.g., “wristwatch” is a more common word than “watchband” or “crystal,” and “glasses” is a higher-frequency word than “lens” or “frame”). Error types include failure to respond, production of paraphasic errors, and circumlocutory responses where the patient describes the object or its use but does not name it. Anomia occurs in aphasias, dementias, and deliria. Adequate vision must be ensured before errors are ascribed to naming deficits.

Evaluation of reading entails assessment of the patient's ability to read aloud and to comprehend what is read. These two aspects of reading may be impaired independently. The most elementary assessment of reading involves asking the patient to read aloud and point to an object the name of which has been written (e.g., “nose” or “door”). Comprehension of more complex written material is evaluated by asking the patient to read and follow written commands (e.g., “raise your right hand”) or to read and fill in the blanks in incomplete written sentences (e.g., “a man who repairs cars and trucks is known as a _____”). Again, adequate vision must be demonstrated before failures are ascribed to an alexia. Abnormalities of spontaneous speech are usually recapitulated in reading aloud, and deficits in reading comprehension typically co-occur with disturbances of comprehension of spoken language. In some cases—alexia with agraphia, alexia without agraphia, and simultanagnosia—reading abnormalities occur in patients without aphasic disturbances.

Writing may be disturbed on a mechanical or aphasic basis. Mechanical agraphias occur in patients with limb paresis, limb apraxia, or movement disorders such as tremor and chorea. Micrographia is a characteristic aspect of parkinsonism in which the script becomes progressively smaller as the patient writes a sentence or extended series of numbers or letters. Aphasic agraphias accompany aphasic syndromes and the same types of errors noted in verbal output are reproduced in written form. For example, patients with Broca's aphasia have impoverished agrammatic output, whereas those with Wernicke's aphasia have fluent paraphasic writing. Agraphia occurs without aphasia in Gerstmann's syndrome (agraphia without alexia, acalculia, right-left disorientation, finger agnosia), alexia with agraphia, and callosal agraphia (left-handed agraphia in patients with injury of the corpus callosum). Agraphia also occurs in dementia and is a sensitive index of the attentional deficit in delirium.

Word-list Generation Word-list generation tasks involve asking the patient to think of as many members of a specific category (category fluency) or as many words beginning with a specific letter (semantic fluency) as possible in 1 minute. Typically, the patient is asked to produce as many animals names as possible in 1 minute or is asked to think of as many words beginning with the letter “F” as possible in 1 minute (excluding numbers, proper names like France or Frank, and repetitions of words with the same core such as *fix, fixed, fixing*, and *fixation*). Normal individuals produce between 12 and 24 animals names in 1 minute (mean: 18, less than 12 is abnormal) and between 10 and 20 words beginning with the letter “F” (mean: 15;less than 10 is abnormal). Word-list generation is sensitive to anomia and word-finding disturbances, lexical search abnormalities associated with frontal-subcortical systems dysfunction, and psychomotor retardation.

Visuospatial Skills There are a number of visuospatial abilities including spatial attention, perception, construction, visuospatial problem solving, and visuospatial memory. Constructional tasks are the most widely used screening tests of visuospatial ability. Clock drawing and copying figures of increasing complexity are useful assessment techniques. In the clock drawing test, the patient is asked to draw a clock and fill in the clock numbers; when that task is complete, the patient is asked to draw in the clock hands to indicate that the time is 10 minutes after 11 o'clock. Patients with poor planning skills will draw a clock face that is too small to conveniently contain the required numbers or may space the numbers inappropriately. Patients with unilateral neglect will ignore half of the clock face. Patients with frontal lobe disorders may have stimulus boundedness and will place one hand on the 10 and one on the 11 when attempting to set the clock for 10 minutes after 11:00.

Tests of copying begin by having the patient reproduce an elementary shape such as a circle; then the patient is asked to copy two or three more challenging figures (e.g., intersecting circle and triangle, overlapping pentagons, cube). Abnormalities include distortions of size, failure to reproduce the shapes accurately, absence of perspective, perseveration on individual elements, drawing over the stimulus figure, or unilateral neglect. Hemi-neglect is a valid indicator of contralateral brain injury, but other types of constructional abnormalities do not have precise localizing value. Injury of the posterior aspect of the right hemisphere produces the most severe and most enduring visuospatial deficits, but dysfunction of frontal, occipital, or parietal lobes of either hemisphere (cortical or subcortical) can adversely affect drawing abilities. Preliminary evidence suggests that the two hemispheres have complementary roles in constructional abilities—the right hemisphere mediates the external configuration whereas the left hemisphere mediates the internal details (Fig. 2.1-1). Drawing disturbances are common with focal brain damage, degenerative disorders, and toxic and metabolic encephalopathies. The term “drawing apraxia” is avoided here; *apraxia* is applied to disorders characterized by an inability to do on command tasks that can be performed spontaneously. Using this definition, drawing disturbances are not apraxias.

FIGURE 2.1-1 Hemineglect. A patient with a right middle cerebral artery infarct was asked to draw a flower. The placement of all features of the flower on the right side demonstrates left hemineglect.

Calculation Calculation skills tested on mental status examination include addition, multiplication, subtraction, and division. Higher-order mathematical skills are not assessed except in specific circumstances where premorbid mathematical abilities are well documented and the examiner is mathematically sophisticated. Patients are asked to perform one- and two-digit tasks (e.g., 16 + 32, 15 × 3) mentally or to address more demanding problems with the aid of pencil and paper. Calculation abilities are determined by the individual's education and occupational history. Acalculias occur in aphasic syndromes where aphasic substitutions of one digit for another lead to inaccurate answers, visuospatial disorders leading to incorrect alignment of columns of numbers, and primary anarithmetias where basic mathematical processes are disrupted. Anarithmetias occur with damage to the posterior aspect of the left hemisphere.

Abstraction Abstracting skills refer to the patient's ability to derive a general principle from a specific example. Similarities, differences, idioms, and proverbs are used to assess abstracting capacity. Similarities require the patient to identify the class or category of which two items are members (e.g., rose and tulip, bicycle and train, watch and ruler). Differences require the patient to identify the salient distinguishing feature between two similar items (e.g., child and midget, canal and river, lie and mistake). Idioms are metaphorical statements or aphorisms that require the patient to generalize to a larger meaning (e.g., “seeing eye to eye,” “level headed,” and “eyes peeled”). Proverbs are usually double metaphors that require the patient to ignore the immediate meaning and derive a lesson or maxim (e.g., “don't cry over spilled milk,” “people who live in glass houses shouldn't throw stones,” “the tongue is the enemy of the neck”). Success on abstraction tasks depends on the individual's level of education. In addition, these tests are culturally biased and proverbs from the individual's own culture must be used to make a valid assessment of

abstraction abilities. Abstraction is sensitive to many types of brain dysfunction and task failure is a nonspecific indicator of cerebral dysfunction. Abstraction deficits are particularly common with frontal systems disorders.

Judgment and Insight Assessing judgment is an important aspect of mental status testing. It should be done so as to yield ecologically valid results that meaningfully relate to the individual's everyday decisions. Questions regarding the patients' insight into their own conditions, their plans for the future, and their understanding of their own limitations best demonstrate their insight and judgment. Questions such as “What would you do if you were walking beside a lake and saw a 2-year-old child playing alone at the end of a pier?” also help to explore the patient's interpersonal and social judgment. Judgment is impaired in many neurological conditions. Orbitofrontal damage produces particularly marked alterations in social judgment.

Frontal-Subcortical Systems Tasks Frontal-subcortical systems tests assess the integrity of a complex circuitry involving the frontal cortex, caudate and thalamic nuclei, and the connecting white matter tracts. Frontal-subcortical functions are among the most difficult to examine either with bedside techniques or formal neuropsychological measures. Some areas of the frontal lobe such as the orbitofrontal cortex produce few identifiable cognitive changes when injured but have major behavioral ramifications with obvious disinhibition and impulsiveness. The principal abnormalities that are observed in patients with frontal-subcortical systems dysfunction include perseveration, motor programming deficits, apathy, poor word list generation, impaired set maintenance, abnormal recall with preserved recognition memory, stimulus boundedness, concrete interpretation of proverbs, imitation of actions of the examiner, and impulsive responses to environmental events.

Motor programming tasks sensitive to injury to the convexities of the frontal lobe and related subcortical structures include alternating programs (Fig. 2.1-2) (e.g., alternating shapes or Ms and Ns in series are copied from a sample provided by the examiner and then extended across a page of paper), reciprocal programs (e.g., the patient holds the right hand in a fist close to his or her own face and extends the left, making a ring with the thumb and first finger, then the positions are repeatedly reversed), and the go-no go test (e.g., when the examiner taps under the table once, the patient taps twice; when the examiner taps twice, the patient makes no response). Another task, copying multiple loop figures, often elicits perseveration (Fig. 2.1-3). Execution of serial hand sequences frequently is disrupted in patients with frontal-subcortical systems abnormalities. In this task, the examiner demonstrates a sequence of three hand positions (e.g., fist, slap, side of hand) and the patient is asked to learn and perform the sequence. If the patient fails, the patient is then asked to say aloud “fist, slap, side.” If the patient fails to guide the hand sequence correctly despite learning the verbal progression, verbal-motor dissociation and hence frontal-subcortical dysfunction is implicated.

FIGURE 2.1-2 Alternating programs. The examiner's model is above, the patient's attempt to duplicate the pattern is below. Failure to alternate between the two shapes is evident.

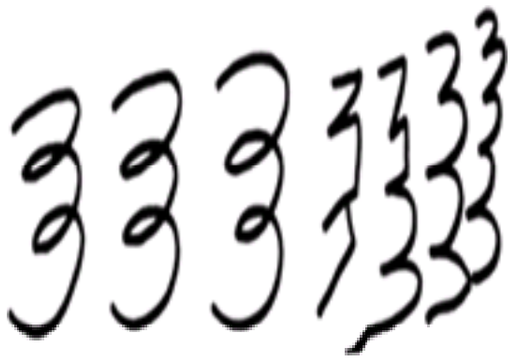

FIGURE 2.1-3 Multiple loops. The three figures on the left were drawn by the examiner. When the patient attempts to copy the figures, perseveration with production of additional loops is demonstrated.

Difficulty maintaining a mental set and limitations in appropriately shifting between tasks also is typical of patients with frontal-subcortical systems dysfunction. Tests to elicit these deficits include trail-making tests (e.g., asking the patient to sequentially alternate between numbers and letters—1-A-2-B-3-C, and so on) and card sorting tests (e.g., the Wisconsin Card Sorting Test) where the patient sorts cards according to a series of criteria designated by the examiner such as color, shape, or number. The patient must discover and apply the sorting criteria through trial-and-error observations. These tests are typically administered and interpreted by neuropsychologists.

Environmental dependency and poor response inhibition can be demonstrated in patients by creating situations with competing demands. One such neuropsychological test (e.g., the Stroop Test) involves writing the name of a color in a different color (e.g., writing the word “blue” in the color red) and asking the patient to ignore the color name and read the color of the print. Patients with frontal-subcortical systems deficits often cannot inhibit the tendency to read the word rather than state the color.

In addition to these more specific frontal systems tests, poor abstracting abilities, stimulus boundedness on clock drawing tests, poor word-list generation, and adopting poor strategies when drawing complex figures (Fig. 2.1-4) aid the examiner in recognizing a frontal-subcortical systems disturbance. These deficits are common in patients with head trauma, frontal lobe degenerations, frontal lobe neoplasms, chronic multiple sclerosis, Huntington's disease and other basal ganglia disorders, multiple subcortical infarctions, and some brain infections such as syphilis.

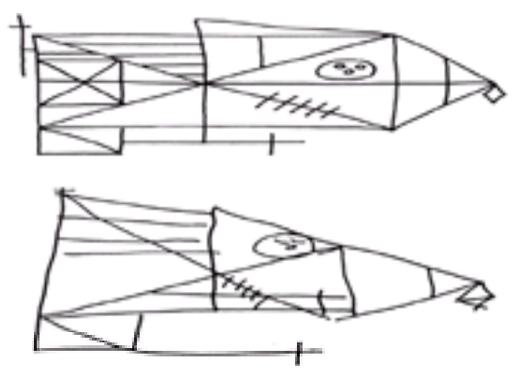

FIGURE 2.1-4 Rey-Osterreith Figure. The figure above was drawn by the examiner; the figure below is an attempt to copy the modal by a patient with frontal lobe

dysfunction.

Miscellaneous Tests A variety of miscellaneous tests may aid diagnosis or localization when added to the mental status examination in specific circumstances. Testing of praxis, prosody, right-left orientation, finger identification, singing, automatic speech, and visual recognition are among these specialized tasks. *Ideomotor apraxias* are abnormalities of learned motor behavior that occur in the absence of motor, sensory, or comprehension deficits and involve motor tasks that the patient can perform spontaneously but cannot execute on command. Apraxis of limbs and buccofacial structures may be detected. Limb commands include performing an action that ordinarily requires no object (e.g., thumb a ride, wave goodbye) or may involve pantomiming the use of an object (e.g., throw a ball, hammer a nail). Typical oral-lingual commands include pretending to blow out the flame of a burning match, coughing, licking the lips, and pretending to sniff a flower. Failures may involve substituting the limb for the imaginary object (e.g., using the hand for the hammer instead of pretending to hold the hammer) or, in more severe cases, simply failing to perform the act adequately. Ideomotor apraxias of the type described here occur with lesions of the left hemisphere or corpus callosum.

Aprosodias are disorders of speech melody and inflection. They assume particular importance in neuropsychiatry because prosody is the means by which emotion is communicated and any disruption of this mechanism impairs the patient's ability to relay his emotional state. Executive aspects of prosody are tested by having the patient inflect a neutral sentence (e.g., “I'm going to the store”) to provide a variety of different meanings such as anger, happiness, and surprise. Prosodic comprehension is assessed by having the patient guess the emotion when the same or a similar neutral test phrase is differentially inflected by the examiner. Executive prosody is disrupted by lesions of the right premotor cortex or the basal ganglia, receptive prosody is impaired by lesions of the posterior superior right temporal lobe (equivalent of Wernicke's area in the right hemisphere).

Disturbances of *right-left orientation* and *finger identification* occur in conjunction with agraphia and acalculia as part of *Gerstmann's syndrome.* Right-left orientation is assessed by asking the patients to point to their own body parts using various combinations of right and left (e.g., “Point to your right shoulder”; “With your right hand, touch your left knee”). If they perform these less demanding tasks accurately, they are asked to point to specific parts of the examiner who is seated facing the patient. The patient is thus required to perform a mental reversal to correctly respond to the commands (e.g., “Which is my right hand?” “Where is my left shoulder?”). The examiner may make the task more difficult by crossing his own hands, thus requiring another reversal by the patient. Finger identification is also tested with a hierarchy of increasingly complex tests. At the most elementary level, the patients are asked to name each finger (e.g., thumb, forefinger, middle finger, ring finger, pinkie or little finger). If the patients are able to perform, they are asked to point to the equivalent finger of one hand while the examiner touches a finger of the other hand held out of sight above the patient's head. Another taxing test of finger identification involves having the patients close their eyes and then state how many fingers are between the fingers of one hand touched by the examiner (e.g., if the examiner touches the thumb and pinkie, there are three fingers between). Gerstmann's syndrome occurs with lesions of the left angular gyrus. All four components must be present for the syndrome to have localizing significance.

Singing and *automatic speech* may be useful adjuncts in language assessment, particularly in the patient with very limited verbal output. In some cases, patients who are mute and incapable of propositional speech nevertheless are able to sing. This occurs in extrapyramidal syndromes such as advanced Parkinson's disease and in patients with nonfluent aphasias. Automatic speech refers to recitation of overlearned sequences such as the alphabet, days of the week, months of the year, nursery rhymes, or prayers. Automatic speech, like singing, may be preserved when little other verbal output can be demonstrated. Preserved automatic speech despite severely impaired spontaneous verbal output is characteristic of *isolation aphasia* (also known as *mixed transcortical aphasia*) and may also be demonstrable in patients with mental retardation syndromes and advanced dementias.

Visual recognition skills may be compromised by specific lesions of the nervous system. Visual perception is intact and naming skills are normal but the patient cannot recognize the object, place, or person. When there are visual complaints, recognition of objects, places, and faces should be tested. *Visual object agnosia* occurs with bilateral lesions of the inferior longitudinal fasciculi or adjacent cortex of the medial occipito-temporal regions. The patients can see the objects, have intact language skills, and can accurately identify the objects by palpation but cannot identify them visually. *Prosopagnosia* refers to impaired recognition of familiar faces. The patient can tell that faces are not identical but cannot recognize familiar individuals such as friends, family members, and celebrities. Auditory recognition skills are intact and the patient will recognize familiar individuals by their voices. Prosopagnosia can occur with lesions limited to the right posterior hemisphere, but many patients have bilateral posterior hemispheric injuries. *Environmental agnosia*, the inability to recognize familiar places such as the patient's own home, may accompany prosopagnosia. *Phonagnosia* refers to the inability to recognize familiar voices, and *auditory agnosia* involves the loss of recognition of all sounds. Stroke is the most common cause of all types of agnosia.

NEUROLOGICAL AND PHYSICAL EXAMINATION

The neurological examination consists of assessment of cranial nerves, motor systems, sensory abilities, muscle stretch reflexes, and primitive reflexes. The findings of the neurological examination are augmented by selected aspects of a general physical examination. The details of the neurological and physical examinations are well described elsewhere and will not be recapitulated here, but the observations most useful in neuropsychiatric assessment are reviewed.

Cranial Nerve Examination Function of the *cranial nerve I (olfaction)* is tested by asking the patient to identify a variety of distinctive odors (e.g., perfume, coffee, chocolate, soap, spices). Requiring patients to close their eyes during the testing will help to ensure that olfaction *per se* is being assessed. The olfactory nerves, bulbs, and tracts are located on the inferior service of the frontal lobe where they are vulnerable to damage by contusions in the course of close head trauma and compression by subfrontal tumors. Olfaction may also be compromised by smoking, sinusitis with rhinitis, and nasal trauma.

Examination of *cranial nerve II (optic)* includes visual inspection of the nerve head, assessment of visual acuity, and mapping of the visual fields. The optic nerve is the only nerve directly accessible to visual inspection by the examiner. By means of ophthalmoscopy, the clinician can detect optic atrophy (a pale disc), papilledema reflecting increased intracranial pressure, or papillitis associated with optic nerve ischemia or inflammation. Macular degeneration, pigmentary retinopathy, retinal infarction, narrowing of retinal vessels, and retinal hemorrhages or exudates may be seen. Visual acuity is tested with a Snellen chart held 18 inches from the face: each eye is tested separately with glasses or pin hole to minimize visual disturbances produced by abnormalities of the shape of the optic globe (astigmatism). Acuity should be correctable to 20/20 if optic nerve function is normal. Impaired acuity after correction implies the presence of prechiasmatic optic nerve dysfunction such as ischemic optic neuritis or inflammation associated with multiple sclerosis or other inflammatory disorder. Papilledema creates an enlarged blind spot but does not impair visual acuity in the remaining visual fields. Visual field defects are identified by covering one of the patient's eyes and comparing the patient's visual field to that of the clinician's by determining when the patient detects a stimulus (moving finger or small stimulus item) entering the visual field from the periphery. The shape of the visual field is determined by repeating this maneuver in all quadrants. Homonymous visual field defects do not disturb visual acuity. Bilateral double simultaneous visual stimulation allows detection of visual neglect: if the patient has full visual fields with monocular testing but extinguishes stimuli in one hemi-field during double simultaneous stimulation, then a neglect syndrome is present. Damage to the optic nerve will produce a monocular visual field defect of the ipsilateral eye. Postchiasmatic injury between the optic chiasm and the occipital lobe usually produces a homonymous visual field defect. Occipital injuries produce highly congruent field defects of similar size and shape in both fields, whereas more anterior lesions (e.g., damage to the geniculocalcarine radiation in the temporal lobe) produce incongruent field defects. Midline tumors in the region of the optic chiasm such as craniopharyngiomas or chromophobe adenomas may compress the chiasm medially, producing a bitemporal visual field defect. Impaired vision of any cause may be associated with release hallucinations (formed or unformed visual hallucinations occurring in the blind field). Visual neglect occurs most often with lesions of the parietal cortex but has also been reported with frontal lobe and subcortical lesions.

Ocular motility, pupillary responses, and eyelid position are mediated by *cranial nerves III, IV, and VI* (*oculomotor, trochlear, and abducens*, respectively). The rectus externus moves the eye laterally and is mediated by the abducens nerve; the superior oblique moves the eye downward when the globe is in the adducted position and is mediated by the trochlear nerve; all the remaining ocular movements, the pupillary responses, and position of the eyelid are determined by oculomotor nerve function. Pursuit movements are tested by having the patient follow the examiner's finger through seven positions (adduction, adducted and up, abducted and up, abducted, abducted and down, adducted and down, convergence). The patient's saccades (volitional movements) are then tested by asking the patient to look to the right, left, up, and down without following a visual target. Conjugate gaze is mediated by gaze centers in the pons (lateral movements) and the midbrain (vertical movements). Brainstem lesions, extrapyramidal movement disorders (e.g., Parkinson's disease and Huntington's disease), and bilateral upper motor neuron lesions may produce supranuclear gaze palsies manifested by incomplete ocular movements, hypometric (abnormally small) saccades, and irregular pursuit movements. Pupillary changes occur in response to light stimulation and in conjunction with convergence (constriction of the pupil with near vision). Cranial nerve III is responsible for pupillary constriction through its parasympathetic branches whereas pupillary dilatation is a function of the sympathetic nervous system. Lid position is determined by the balanced input from the third nerve responsible for volitional eye opening, the seventh cranial nerve mediating lid closure, and sympathetic input that also participates in keeping the lid open. Pupillary abnormalities occur with brainstem lesions, syphilis (*Argyll-Robertson pupil*, in which the reaction to light is lost while the convergence-accommodation mechanism is intact), diabetes, and compression by aneurysms or other mass lesions. Third-nerve palsy results in ptosis, a nonreactive

pupil, and an inability to adduct, elevate, or depress the eye. Lid retraction may occur with lesions of the collicular plate or in parkinsonian syndromes. Horner's syndrome includes ptosis, pupillary constriction, and diminished sweating on the ipsilateral side of the face. It results from sympathetic paralysis caused by lateral medullary infarcts, cervical cord lesions, pulmonary apical or mediastinal lesions, or neck trauma.

The *cranial nerve V (trigeminal)* innervates the muscles of mastication and is responsible for facial and corneal sensation. The fifth cranial nerve is involved in tic doloreux and other facial pain syndromes.

Cranial nerve VII (facial) supplies the facial musculature including the platysma muscles of the anterior neck, orbicularis oris and buccinator muscles, orbicularis oculi, and the frontalis. In addition, its branches innervate the lacrimal and salivary glands, muscles of the stapedius, and taste fibers of the anterior two thirds of the tongue. Fibers supplying the upper facial muscles receive supranuclear input from ipsilateral and contralateral motor cortex, whereas fibers innervating the lower facial muscles receive only contralateral input. Thus, peripheral nerve lesions result in ipsilateral paralysis of upper and lower face whereas upper motor neuron lesions above the level of the seventh nerve nucleus produce only contralateral lower facial paresis. All human faces are asymmetrical and facial asymmetry by itself should not be regarded as evidence of paresis; asymmetries, however, can result from weakness and their detection should lead to careful testing for evidence of muscle paresis. The facial nerve receives input from both descending pyramidal neurons mediating volitional movements and nonpyramidal limbic system connections mediating emotional responses. This dual innervation results in the preservation of emotional responses in spite of paralysis of volitional facial movements in patients with pyramidal tract lesions (most evident in the syndrome of pseudobulbar palsy) and a lack of emotional facial responses with intact volitional movements in patients with lesions of the limbic system. The latter has been observed, for example, in patients with seizure foci arising from lesions of the temporal lobes. The facial musculature is involved in generalized movement disorders (e.g., chorea, tics, parkinsonism), which are discussed below.

The *cranial nerve VIII* is responsible for hearing (*auditory branch*) and balance (*vestibular branch*). Auditory compromise is associated with deafness and can produce auditory hallucinations. Vestibular end organ or nerve lesions produce nystagmus and vertigo and, when acute, may be associated with the experience of sudden shifts in the environment (e.g., vertical inversion). Vestibular nystagmus is horizontal or combined horizontal-rotatory and is typically accompanied by vertigo and nausea. Lesions disrupting vestibular connections in the central nervous system can produce nystagmus in any direction and are usually not associated with vertiginous or nauseous sensations. Vestibular nystagmus should be distinguished from gaze-paretic nystagmus resulting from inadequate ocular deviation and occurring with paresis of cranial nerves III, IV, or VI. The quick or jerk phase of gaze-paretic nystagmus occurs in the direction of the weakened muscle. Nystagmus also occurs with a variety of medications, particularly anticonvulsant and sedative-hypnotic drugs. Table 2.1-3 presents the differential diagnosis of nystagmus and related disorders of ocular movement.

Table 2.1-3 Types of Nystagmus and Related Ocular Movements

Cranial nerves IX and X (*glossopharyngeal and vagus*, respectively) control pharyngeal and laryngeal function, taste, and the gag reflex. The glossopharyngeal nerve conveys sensory information (including taste sensation from the posterior third of the tongue) and the vagus nerve is primarily responsible for the motor aspects of these structures. Dysfunction of the vagus nerve results in a hoarse voice or aphonia and dysphagia. Supranuclear lesions in the syndrome of pseudobulbar palsy produce an exaggerated gag reflex.

Cranial nerve XI (*spinal accessory*) innervates the upper half of the trapezius muscle and the sternocleidomastoid muscle. *Cranial nerve XII* (*hypoglossal*) innervates the tongue; tongue weakness is common in pseudobulbar palsy.

Motor System Examination Motor system examination includes assessment of muscle bulk, strength, and tone as well as tests of coordination and identification of movement disorders. Normal motor function depends on the functional integrity of the pyramidal motor system, extrapyramidal nuclei, cerebellum, spinal cord, peripheral nerve, myoneural junction, and muscle. Assessment of the motor system is particularly important in neuropsychiatry; psychiatric disturbances frequently produce associated motor system manifestations, movement disorders are often accompanied by neuropsychiatric abnormalities, and many of the agents used to treat psychiatric diseases adversely affect motor system function.

Muscle *bulk* is examined by visual inspection and palpation and occasionally by measurement. The dominant limbs, particularly the upper extremity, have slightly greater bulk than the nondominant extremity. Muscle wasting may occur with volitional disuse, muscle disease, nerve or spinal cord disease, and in generalized weight loss secondary to malnutrition, systemic illness, or advanced brain diseases.

Muscle *strength* is tested by asking the patient to exert maximal power of specified muscle groups. Strength is graded as 0—no evidence of muscle contraction, 1—muscle contraction without movement of the limb, 2—limb movement after gravity eliminated, 3—limb movement against gravity, 4—limb movement against partial resistance, 5—normal strength. Upper and lower limbs and both proximal and distal musculature should be assessed. Distal weakness is characteristic of peripheral neuropathies; proximal weakness is a manifestation of primary muscle disease. Weakness below a specific level of innervation occurs with spinal cord disease (*paraparesis* or *quadriparesis*), and fatigability is most characteristic of muscle and myoneural junction disorders (e.g., myasthenia gravis). Lateralized weakness (*hemiparesis*) is a manifestation of lesions of the contralateral brainstem or above. Pyramidal tract pathology produces a “predilection” pattern of weakness affecting the extensor muscles of the upper limbs and the flexor muscles of the lower limbs; this is known as a *decorticate posture* and imitates the antigravity posture of the biped. Pyramidal lesions of the pons result in flexor weakness of all limbs and consequent extensor posturing; extension of all limbs is known as the decerebrate posture and recapitulates the antigravity posture of the quadruped.

Muscle *tone* is evaluated by passively manipulating the limbs or neck while the patient relaxes. Additional information can be garnered by asking the patient to perform motor activities with the contralateral limb while tone in the ipsilateral limb is determined (e.g., the patient is asked to draw a large square in the air with one arm while the examiner is testing the tone of the other). Activation of tone by this maneuver (*Froment's sign*) is characteristic of parkinsonianism. *Hypotonia* is relatively unusual in neuropsychiatric disorders. Muscle tone is decreased in muscle and peripheral nerve disease, cerebellar syndromes, early in the course of many choreiform disorders, and in the short term following an upper motor neuron lesion. Episodic hypotonia occurs in narcoleptic patients and accounts for their cataplectic attacks. Increased muscle tone (*rigidity*) is encountered in a variety of neuropsychiatric conditions. *Spasticity* is a form of rigidity in which the muscle is flaccid at rest and reacts with increasing tone to slow passive stretch. In many cases, the tone abruptly decreases after reaching a certain magnitude resulting in a sudden loss of resistance (the clasp-knife phenomenon). Spasticity results from upper motor neuron lesions that result in disinhibition and hyperexcitability of alpha and gamma motor neurons of the spinal cord. Muscle stretch reflexes are mediated through the same mechanism as spasticity and are exaggerated in spastic disorders. Parkinsonian extrapyramidal syndromes produce *plastic rigidity* (also known as “*lead pipe” rigidity*) characterized by an increased resistance to passive movement that is independent of the velocity of limb movement. The rigidity is produced by dysfunction of the nigrostriatal projection or the globus pallidus. Plastic rigidity may involve primarily the appendicular musculature (in Parkinson's disease) or may be more pronounced in the axial structures such as the neck and trunk (in progressive supranuclear palsy). *Cogwheel rigidity* observed in some extrapyramidal disorders results from the occurrence of tremor that is palpated as intermittent resistance when manipulating the limbs; it may occur with or without plastic rigidity. “ *Gegenhalten*” (also known as *paratonia*) refers to the resistance to movement encountered in patients with advanced brain diseases. It is characterized by active resistance to passive limb movements in all directions and at any velocity. It may result from release of primitive protective or position holding mechanisms similar to the released motor programs responsible for grasp and suck reflexes. *Waxy flexibility* is classically associated with catatonia. The patient's limb has moderate resistance to passive movement and maintains the end position of the movement (like candle wax). Catatonia occurs in a variety of psychiatric, frontal lobe, extrapyramidal, and toxic-metabolic disorders.

Coordination may be disrupted by many types of motor and sensory abnormalities but is primarily dependent on intact cerebellar function. Disturbances of coordination must be interpreted with caution to avoid ascribing the effects of weakness, spasticity, or sensory loss to cerebellar disease. Tests of coordination used in the standard neurological examination include rapid alternating movements (alternating supination and pronation movements of the hand), fine finger movements (repeated apposition of the thumb and first finger), finger-to-nose movements (the patient alternates between touching his own nose and the examiner's finger), heel-knee-shin maneuvers (the patient touches the knee of one leg with the heel of the other and then gently slides the heel down the shin), and rebound check tests (the clinician suddenly releases as the patient is pushing his extended arms upward against the examiner; the clinician observes whether the patient can quickly arrest the upward movement after release of the limbs). Gait, eye movement, and tremor-related observations aid in identifying a cerebellar disorder. Dysdiadochokinesia (abnormal rapid alternating movements), ataxia, loss of rebound check, and intention tremor are characteristic cerebellar disturbances.

Gait and posture are examples of motor acts dependent on successful motor and sensory integration. Systematic gait observations should assess step initiation, stride length, step height, base width, step symmetry, path deviation, trunk stability, speed, turning, and adventitious movements. *Festination of gait* refers to increasingly fast forward movement with an inability to stop. Asking the patient to perform tandem toe-to-heel walking makes cerebellar, sensory, and vestibular disturbances of gait and balance more evident. Table 2.1-4 summarizes the principal types of gait abnormalities.

Observation	Abnormality	Cause
Step initiation	Start hesitation	Parkinsonism
Stride length	Shortened	Parkinsonism
Step height	Decreased	Parkinsonism
	Increased	Sensory abnormality
Base width	Increased	Cerebellar ataxia, chorea, sensory abnormalities
Step symmetry	Asymmetric	Pain with unilateral avoidance
Path course	Deviation	Cerebellar or vestibular ataxia
Trunk stability	Unsteady	Cerebellar or vestibular disorders
Speed	Slowed	Parkinsonism
Turning	En bloc	Parkinsonism
Adventitious movements	Present	Chorea, athetosis, dyskinesia, tics, tremor, myoclonus, dystonia

Table 2.1-4 Abnormalities of Gait

Posture may be extended or flexed. In addition, sensory abnormalities, cerebellar disorders, and extrapyramidal disturbances may contribute to postural instability. Decorticate (flexed upper limbs, extended lower limbs) and decerebrate postures (all limbs extended) have already been described. Parkinson's disease and many parkinsonian syndromes produce a tendency to flexion of all limbs; progressive supranuclear palsy causes a characteristic extended posture. Postural stability is tested by asking the patient to stand with legs comfortably apart and eyes open; the examiner then stands behind the patient and pulls backward on the patient's shoulders with a short rapid jerk. Patients with extrapyramidal disorders fall backwards or must take a compensatory backward step. Sensory input to postural mechanisms can be tested by asking the patient to stand with feet together and eyes closed. Patients with sensory abnormalities from peripheral neuropathy disorders or the spinal cord affecting its posterior columns can stand with eyes open but not with eyes closed (*Romberg's sign*); patients with cerebellar ataxia cannot stand with feet together whether the eyes are open or closed.

Soft signs are minor motor and sensory abnormalities that are normal in the course of early development but abnormal when present beyond childhood. They have been reported with increased frequency in a variety of psychiatric disorders including schizophrenia and sociopathy. Soft signs lack definitive localizing significance but are indicative of subtle brain dysfunction. Table 2.1-5 lists the main soft signs elicited in the course of the neurological examination.

Test	Soft Sign
Articulation	Mild dysarthria
Finger tapping	Slow, clumsy, irregular
Foot tapping	Slow, clumsy, irregular
Rapid alternating movements	Slow, clumsy, irregular
Eye closure	Cannot maintain on command
Tongue extrusion	Cannot maintain on command
Arms extended	Minor choreiform movements
Finger-to-nose	Jerky, clumsy
Heel-to-shin	Jerky, clumsy
Heel walking	Unsteady, posturing of upper limbs
Toe walking	Unsteady, posturing of upper limbs
Standing on one foot	Cannot balance
Hopping on one foot	Unsteady
Tandem walking	Unsteady
Tandem walking backwards	Unsteady
Face-hand test	Distal stimulus extinguished when distal and proximal stimulus delivered simultaneously
Two-point discrimination	Perceives two separate points only when more widely separated than normal
Graphesthesia	Errors in recognizing numbers written on fingertips by the examiner
Romberg sign	Patient feels unstable when standing with feet together, eyes closed

Table 2.1-5 Neurological Soft Signs

A wide variety of *movement disorders* may be observed in the course of the neuropsychiatric examination. Four general classes of movement disorders can be distinguished—parkinsonism, dystonia, hyperkinesias, and tremor. The first three are irregular disorders, and tremor is a regular hyperkinesia. Parkinsonian syndromes are characterized by rigidity and bradykinesia. Dystonic disorders produce sustained postural deviations. Hyperkinesias include chorea, dyskinesias, athetosis, tics, and myoclonus. Tremors are regularly alternating movements. Several different types of movement disorders (e.g., parkinsonism, dystonia, and tremor) can occur simultaneously.

Parkinsonism may be produced by degenerative disorders (Parkinson's disease, progressive supranuclear palsy), multiple small strokes (lacunar state), metabolic disorders (hypothyroidism, hypoparathyroidism), head trauma, CNS infections (Creutzfeldt-Jakob disease, HIV encephalopathy), CNS tumors, hydrocephalus, and drug treatment. Medications most likely to induce parkinsonism include dopamine receptor antagonists, such as phenothiazines and butyrophenones. Table 2.1-6 summarizes the differential diagnosis of parkinsonism. The parkinsonian syndrome features bradykinesia and plastic rigidity with or without a rest tremor. Manifestations of bradykinesia include a masked face, reduced spontaneous blinking, diminished spontaneous swallowing with sialorrhea, start hesitation when initiating movement, shuffling gait (combination of reduced stride length and decreased step height), decreased arm swing when walking, en bloc turns, slowed movement, and reduced spontaneous gesturing.

Degenerative disorders
Parkinson's disease
Progressive supranuclear palsy
Cortico-basal ganglionic degeneration
Idiopathic basal ganglia calcification
Dementia with Lewy bodies
Rapid Huntington's disease (Westphal variant)
Parkinsonism-dementia amyotrophic lateral sclerosis complex of Guam
Wilson's disease
Hallervorden-Spatz disease
Azorean disease (Machado-Joseph disease)
Striatonigral degeneration
Olivopontocerebellar atrophy
Shy-Drager syndrome
Pallidal atrophies
Vascular disorders
Binswanger's disease
Lacunar state
Metabolic conditions
Hypoparathyroidism
Hypothyroidism
Drugs and toxins
Antipsychotics
Lithium carbonate
Reserpine
Methyl-phenyl-tetrahydropyridine (MPTP)
Manganese
Reserpine
Trauma
Dementia pugilistica
Postconcussive encephalopathy
Infections
Encephalitis lethargica
HIV encephalopathy
Creutzfeldt-Jakob disease
Syphilis
Miscellaneous
Carbon monoxide intoxication
Carroll-Hallervorden (Wolf's disease)
Dyke-Lee syndrome
Hydrocephalus
Meningioma
Neoplastic metastasis/carcinomatosis

Table 2.1-6 Causes of Parkinsonism

Hyperkinetic movement disorders include athetosis (observed primarily in cerebral palsy), chorea (Huntington's disease, Sydenham's chorea, neuroacanthocytosis), tardive dyskinesia, myoclonic disorders (Creutzfeldt-Jakob disease, myoclonic epilepsy, hereditary myoclonus, drug-induced myoclonus), ballismus (usually occurring as hemiballismus following infarction of the subthalamic nucleus), and tic disorders (Gilles de la Tourette syndrome, simple tic syndromes). *Hemiballismus* is a dramatic disorder in which the patient manifests large-amplitude flinging movements of the arm and leg contralateral to a lesion of the subthalamic nucleus; with time, the movement may evolve into hemichorea. *Chorea* is an irregular hyperkinesia intermediate in speed between the slower athetosis and the faster myoclonus. The choreic movements resemble normal movements and may be incorporated into semipurposeful gestures to partially obscure their occurrence. Electromyographic recordings, however, reveal that antagonistic muscles are active at inappropriate times. In Sydenham's postinfectious chorea the movements involve the distal musculature more than the proximal; in Huntington's disease the proximal musculature is more affected. *Dyskinetic movements* observed in tardive dyskinesia are a special instance of chorea. *Myoclonus* refers to random asynchronous brief muscle jerks large enough to move a limb. *Tics* are similar to myoclonus in amplitude and duration but are stereotyped, repetitively involving the same muscles, and they are subject to partial temporary volitional suppression. Table 2.1-7 summarizes the principal hyperkinetic movement disorders. *Athetosis* is a slow, writhing hyperkinesia.

Table 2.1-7 Differential Diagnosis of the Principal Hyperkinetic Disorders

Dystonia is a movement disorder in which there is tonic contraction of antagonistic muscle groups leading to sustained postural abnormalities. Initially, the muscle contractions may be intermittent, giving rise to an irregular tremor-like disorder. Dystonia may be (1) focal: involving a single body part (e.g., torticollis), (2) segmental: involving adjacent body parts (e.g., torticollis plus involvement of one shoulder), or (3) generalized. Table 2.1-8 provides the differential diagnosis of dystonia. Idiopathic dystonias appear to have few neuropsychiatric associations; secondary dystonias occur in conjunction with many of the other hyperkinesias (including Huntington's disease and Parkinson's disease); and dystonia is a frequent adverse effect of treatment with neuroleptic agents (e.g., dopamine receptor antagonists). Drug-induced dystonias may be acute (neuroleptic-induced acute dystonia) occuring within a few hours of initiating treatment or increasing the drug dosage, or they may be manifestations of tardive dyskinesia.

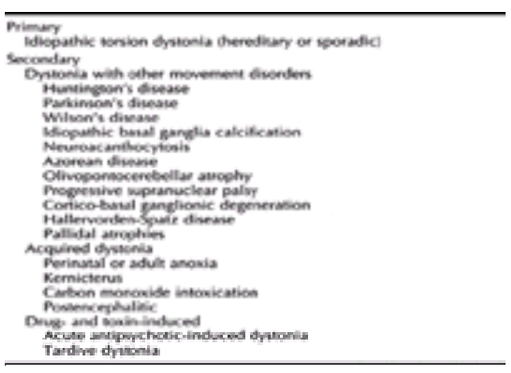

Table 2.1-8 Differential Diagnosis of Dystonia

Tremor is an involuntary, regular oscillation of a body part around a fixed point. Table 2.1-9 presents a classification of tremors including rest tremors, postural tremors, and kinetic tremors. *Rest tremors* include the typical large-amplitude, low-frequency (approximately 4–6 Hz) tremor of Parkinson's disease and other parkinsonian states. This tremor disappears with sleep and is better characterized as a tremor of alert repose, occurring when the patient is alert but inactive. Rest tremor is suppressed by volitional movement. The principal *postural tremors* include exaggerated physiological tremors and essential tremors. These are small-amplitude, high-frequency (10–2 Hz) tremors that are absent at rest and occur during postural maintenance activities (holding the arms outstretched in front of the body, writing and so on). Exaggerated physiological tremors occur with fatigue, excitement, hypoglycemia, and thyrotoxicosis as well as in metabolic encephalopathies and drug-induced toxic states (caused by lithium, tricyclic agents, sympathomimetics, and so on). Essential tremors are usually inherited but may occur sporadically. Kinetic tremors are those that worsen with movement. Cerebellar intention tremors are the classic example of kinetic tremors and are observed frequently in patients with multiple sclerosis.

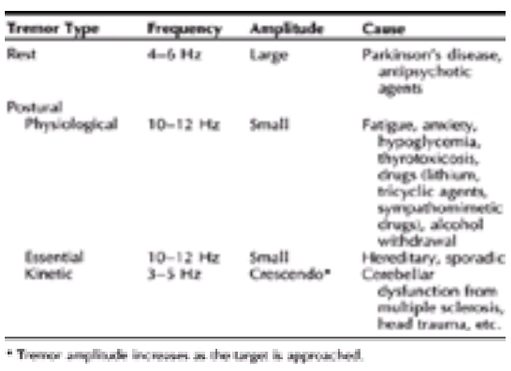

Tremor Type	Frequency	Amplitude	Cause
Rest	4–6 Hz	Large	Parkinson's disease, antipsychotic agents
Postural			
Physiological	10–12 Hz	Small	Fatigue, anxiety, hypoglycemia, thyrotoxicosis, drugs (lithium, tricyclic agents, sympathomimetic drugs), alcohol withdrawal
Essential	10–12 Hz	Small	Hereditary, sporadic
Kinetic	3–5 Hz	Crescendo*	Cerebellar dysfunction from multiple sclerosis, head trauma, etc.

* Tremor amplitude increases as the target is approached.

Table 2.1-9 Classification of Tremor

Sensory Examination Examination of sensory function includes assessment of primary sensory modalities mediated at the thalamic level as well as secondary or cortical sensation. Sensory information is projected from peripheral receptors to the post-Rolandic sensory cortex via peripheral nerves, spinal cord, brainstem, and thalamus; damage at any of these levels may produce a sensory deficit. Primary modalities include light touch, pain (tested by examining the patient's ability to distinguish a sharp pin from a dull object), temperature, and vibration. Vibration should be tested with a 128-Hz tuning fork. Cortical sensory modalities include joint position sense, two-point discrimination (ability to discriminate two closely spaced areas of stimulation from a single stimulation), graphesthesia (ability to identify numbers “written” on the tips of the fingers), and stereognosis (ability to recognize by touch objects placed in the hand). Tactile double simultaneous stimulation is a sensitive and valuable means of detecting unilateral neglect. The patient is told that they will be touched in one or two areas including the face, dorsum of the hands, or front of the lower legs. The patient's ability to perceive bilateral simultaneous stimuli is then systematically assessed. Patients with neglect will fail to perceive

stimuli on one side of the body when they are simultaneously stimulated on the other side. Primary sensory abilities (e.g., light touch) must be intact to perform the test. Patients with unilateral neglect may manifest the related behavioral disorder of *anosognosia* (denial of illness).

Muscle Stretch Reflexes Muscle stretch reflexes are monosynaptic spinal cord reflexes that are modulated by descending connections. Muscle stretch reflexes are decreased in muscle, peripheral nerve, and nerve root disorders and increased in the presence of upper motor neuron lesions. Every muscle has an associated muscle stretch reflex and the examiner chooses a small number that can routinely be obtained in the screening evaluation and selects more specific reflexes to explore hypotheses generated in the course of the examination. A reasonable screening examination would include the following muscle stretch reflexes: biceps, brachioradialis, triceps, finger flexors, knee jerks, and ankle jerks. Reflexes are graded as 0—absent or unobtainable, 1—decreased, 2—normal, 3—brisk with spread of the contraction response to adjacent muscles, or 4—clonus. *Clonus* refers to the series of rhythmic reflex contractions observed when a muscle is abruptly subjected to sustained stretch. Clonus is most easily elicited at the ankles but may be observed in the knees, wrists, and jaw. Lateralized increased reflexes in conjunction with weakness, spasticity, and the Babinski sign comprise the upper motor neuron syndrome indicative of a contralateral lesion of the descending pyramidal system. Reflexes are diminished in the acute phase of an upper motor neuron lesion but gradually increase during the first few weeks following the injury.

Pathological and Primitive Reflexes *Babinski's sign* is the most important pathological reflex sought in the neurological examination. It is typically elicited by stroking the lateral aspect of the plantar surface of the foot from back to front with a semi-sharp object (Figure 2.1-5). A normal response consists of plantar flexion of the great toe; an abnormal response (Babinski's sign) is comprised of dorsiflexion of the great toe with or without fanning of the other toes. It is produced by upper motor neuron lesions and is a fragment of a protective flexion reflex. In its completely developed form (observed in infants and in some patients with severe spinal cord injuries), there is a flexor synergy with flexion of hip, knee, and toe removing the foot from the noxious stimulus. Thus, Babinski's sign like other primitive reflexes represents the release of more primitive motor programs from suprasegmental control. There is no equivalent reflex in the upper extremity, but *Hoffman's sign* (thumb flexion elicited by briskly flexing the distal phalanx of the middle finger while holding the digits in a dorsiflexed position) may be used to assess hyperreflexia in the upper extremities and is frequently present in patients with upper motor neuron lesions and Babinksi signs in the lower limbs.

FIGURE 2.1-5 Babinski reflex. (Reprinted with permission from Major RH, Delp MH: *Physical Diagnosis*, ed 5. Saunders, Philadelphia, 1956.)

The *glabella tap reflex (Myerson's sign)* is often mistakenly considered a primitive reflex similar to grasp and suck responses, but it is best considered a sign of extrapyramidal dysfunction. Continued blinking with tapping of the glabellar region is associated with movement disorders and is seen in Parkinson's disease and medication-induced parkinsonism. To be abnormal the blinking must continue to occur after four or more glabella taps.

The *grasp reflex* refers to involuntary gripping of objects in or near the patient's hand. Four types of grasp response have been described: (1) the simple grasp response occurs when the patient's thenar eminence is stimulated; (2) the hooking response occurs when the patient's flexed fingers are gently extended; (3) if the hooking response is marked, there is loss of voluntary relaxation and the patient grips steadily as the examiner withdraws his hand (traction grasp); and (4) in patients with advanced disease, the sight of the examiner's hand or light touch of the hand causes the patient to move his hand to grasp the examiner's hand (groping response or magnetic grasp). Grasp reflexes occur in patients with advanced brain disease and with more restricted lesions limited to the medial frontal lobes. Grasp responses can be elicited in the feet as well as in the hands.

The *sucking reflex* is comprised of sucking movements of the lips, tongue, and jaw elicited by gentle stimulation of the lips. Like the grasp reflex, these are primitive motor programs that occur normally in infants and reappear as age-inappropriate signs in patients with frontal lobe and diffuse brain dysfunction.

The *palmomental reflex* features ipsilateral contraction of the mentalis muscle in response to stroking of the thenar eminence of the hand. This reflex can be seen in a range of ages in individuals and may be regarded as pathological when it is unilateral or when it does not fatigue with repeated stimulation.

Physical Examination Selected aspects of the general physical examination are particularly important in neuropsychiatric assessment. Dysmorphological features are important in the recognition of congenital and chromosomal syndromes. The manifestations of fetal alcohol syndrome, for example, may be evident in adults and include microcephaly, short palpebral fissures, maxillary hypoplasia, short nose, smooth philtrum with thin and smooth upper lip, small distal phalanges, and small fifth fingernails.

Examination of the head and neck is essential. Patients may have enlarged heads from compensated hydrocephalus (adult males should have a head circumference less than 58.4 cm and adults females less than 57.5 cm) or small heads from inadequate brain development. Palpation of the head may reveal evidence of a previous craniotomy not revealed or recalled by the patient. A short neck is characteristic of platybasia associated with congenital deformities of the posterior fossa *(Arnold-Chiari malformation)* and of basilar impression (occurring, for example, with Paget's disease of the skull) that may give rise to hydrocephalus.

The neuro-ophthalmological examination may contribute essential information in the neuropsychiatric examination. Examination of the optic nerve, visual fields, extraocular movements, lid position, and pupillary responses have already been discussed. In addition, inspection of the eyes may reveal a Kayser-Fleischer ring at the corneal limbus in Wilson's disease. Retinal inspection may show macular degeneration (associated with blindness and visual hallucinations), and retinal degenerations (occurring with phenothiazines, particularly thioridazine, and some hereditary CNS disorders). Cataracts may be present in diabetes, hypoparathyroidism, and Wilson's disease and after treatment with chlorpromazine (Thorazine) and corticosteroids.

Cardiovascular assessment is of obvious interest in neuropsychiatry. Cardiac and carotid auscultation as well as blood pressure measurement are necessary in most examinations. Abdominal palpation may occasionally reveal organomegaly in neuropsychiatric illnesses including hepatomegaly in alcoholism. Skin changes in neuropsychiatric disorders include pigment deposition after long-term treatment with chlorpromazine, small pigmented papules (angiokeratoma corporis diffusum) in Fabry's disease, and a variety of anomalies in the phakomatoses such as tuberous sclerosis and von Recklinghausen's disease.

NEUROBEHAVIORAL SYNDROMES

Behavioral disturbances associated with neurological disorders will be divided arbitrarily into neurobehavioral syndromes and neuropsychiatric syndromes. Neurobehavioral syndromes include deficit disorders such as aphasia, amnesia, frontal lobe disorders, and anosognosia. Neuropsychiatric syndromes include productive disorders such as delusions, hallucinations, depression, mania, anxiety, personality changes, obsessive-compulsive disorders, and paraphilias.

Aphasia *Aphasia* refers to the loss of language abilities produced by brain dysfunction. The most common cause of isolated aphasia is stroke, but it may also occur with brain tumors, infections, trauma, and in the course of degenerative diseases such as Alzheimer's disease.

There are eight major aphasia syndromes, each identifiable by a distinctive combination of changes of fluency, comprehension, and repetition (Table 2.1-10). *Wernicke's aphasia* has a fluent output with poor comprehension and impaired repetition. Naming is abnormal, reading comprehension is impaired, reading aloud is compromised, and writing contains errors similar to the fluent paraphasic output. *Broca's aphasia* is characterized by a terse nonfluent output, intact comprehension, and poor repetition. Reading comprehension is preserved and reading aloud has the same nonfluent characteristics as spoken language; naming and writing are impaired. *Global aphasia* is the most severe aphasic syndrome: spontaneous output is nonfluent and often severely limited, comprehension is abnormal, and

repetition is disturbed. Naming, reading comprehension, reading aloud, and writing are all abnormal. *Anomic aphasia* is a mild aphasic syndrome with fluent output, preserved comprehension, and intact repetition. The principal disturbances are a word-finding deficit and abnormalities of naming. Reading aloud and writing contain the same abnormalities noted in spontaneous speech. *Conduction aphasia* resembles Wernicke's aphasia in having fluent output and impaired repetition, but comprehension is preserved. Conduction aphasia patients read with comprehension but cannot read aloud. Writing contains paraphasic errors. The typical paraphasias of conduction aphasia are phonemic substitutions, whereas patients with Wernicke's aphasia have verbal and neologistic paraphasias. *Transcortical sensory aphasia* is similar to Wernicke's aphasia (fluent output, poor comprehension) with intact repetition. Reading aloud, reading comprehension, and writing are compromised. *Transcortical motor aphasia* has features similar to Broca's aphasia (nonfluent output, intact comprehension) with normal repetition. Reading comprehension is intact; reading aloud and writing are abnormal. *Mixed transcortical aphasia* resembles global aphasia (nonfluent, poor comprehension) but repetition is intact.

Aphasia Type	Fluency	Comprehension	Repetition	Naming	Reading Comprehension	Reading Aloud	Writing
Wernicke's	F	Impr	Impr	Impr	Impr	Impr	Impr
Anomic	F	Normal	Normal	Impr	Normal	Impr	Impr
Conduction	F	Normal	Impr	Impr	Normal	Impr	Impr
Transcortical sensory	F	Impr	Normal	Impr	Impr	Normal	Impr
Broca's	NF	Normal	Impr	Impr	Normal	Impr	Impr
Transcortical motor	NF	Normal	Normal	Impr	Normal	Impr	Impr
Mixed transcortical	NF	Impr	Normal	Impr	Impr	Impr	Impr
Global	NF	Impr	Impr	Impr	Impr	Impr	Impr

F, fluent; NF, nonfluent; Impr, impaired.

Table 2.1-10 Characteristics of the Major Aphasia Syndromes

Each of the aphasic syndromes is associated with a different region of erebral cortical dysfunction. Wernicke's aphasia is produced by a lesion of the posterior superior left temporal gyrus (Wernicke's area). Broca's aphasia reflects pathology of the middle third of the left inferior frontal gyrus (Broca's area). Global aphasia is associated with a large left-hemisphere lesion involving the anterior and posterior lateral convexity. Global aphasia is occasionally produced by a combination of two lesions, one involving Wernicke's area and another affecting Broca's area. Anomic aphasia occurs with lesions of the posterior inferior left temporal gyrus, left angular gyrus, or left temporal pole. Conduction aphasia is seen with disruption of the arcuate fasciculus traversing from Wernicke's area to Broca's area. Transcortical sensory aphasia is caused by lesions of the posterior inferior parietal region. Transcortical motor aphasia may occur with lesions of the left medial frontal lobe or the convexity area superior to Broca's area. Mixed transcortical aphasia is produced by lesions combining the syndromes of transcortical motor and transcortical sensory aphasia or by large lesions of the anterior and posterior medial left hemisphere.

Localizing principles regarding lesion location in aphasic patients can be derived from these observations. Syndromes with impaired repetition all have lesions bordering directly on the Sylvian fissure of the left hemisphere (Wernicke's, Broca's, and conduction aphasias), whereas those with intact repetition (transcortical aphasias) have lesions that spare the immediate peri-Sylvian structures. Disorders with impaired comprehension have lesions affecting posterior hemispheric structures (behind the Rolandic fissure), whereas aphasic syndromes with intact comprehension typically have lesions limited to anterior hemispheric structures. These localization rules apply only to right-handed adults. Children often exhibit nonfluent aphasias regardless of the lesion localization in the left hemisphere, and left-handed individuals may have mixed cerebral dominance resulting in atypical aphasia syndromes following localized left hemisphere injury.

Etiological diagnosis is aided by aphasia syndrome identification. Wernicke's aphasia and conduction aphasia are most often caused by emboli, usually arising from the heart. Transcortical aphasias arise from lesions of the vascular border zone regions and are typically seen with large-vessel disease involving the extracranial carotid arteries. Transcortical sensory aphasia is commonly observed in the middle phases of Alzheimer's disease. Global aphasia usually occurs with thromboembolic occlusion of the left middle cerebral artery.

Amnestic and Memory Disorders *Amnesia* refers to an inability to learn new information despite normal attention, language, and other intellectual abilities. Amnestic disorder is typically characterized by a period of *retrograde amnesia* extending back from the time of onset of the amnesia and encompassing a few seconds to a few years of the patient's life. Retrograde amnesia does not include all of the patient's past life and personal identity is not forgotten. Persons with amnesia who claim to have forgotten all personal history are usually found to be malingering or are suffering from an idiopathic psychiatric disorder. Amnestic disorders also have an *anterograde amnesia* that extends forward from the time of onset of the memory disorder. Anterograde amnesia may be temporary, lasting until memory functions recover, or it may be permanent. In patients exhibiting recovery of memory function, there may a concomitant shrinking of retrograde amnesia as the anterograde amnesia resolves. The memory dysfunction in amnesia is an inability to store new information for later recall. Memory for newly learned information is not aided by clues or prompting strategies. *Confabulation* refers to untrue responses to questions that are formulated by amnestic patients without a deliberate attempt to mislead. The confabulated responses are often drawn from the patient's actual experiences but are produced out of context. For example, when asked about their occupation, the patient may produce a detailed but untrue description, the details of which were taken from a job they had 10 or 20 years previously. Confabulation is common in the acute phases of amnesia but is rare in chronically amnestic patients.

Amnestic disorder is associated with a restricted set of lesions involving limbic structures of the medial aspect of the cerebral hemispheres. The damage that produces amnesia affects the hippocampus, fornix, mammillary bodies, or medial thalamic nuclei. Severe amnesia usually reflects bilateral damage, whereas unilateral injury results in relatively modest and hemisphere-specific deficits with left-sided lesions producing verbal amnesia and right-sided lesions causing deficits in learning nonverbal and visuospatial information.

The differential diagnosis of amnesia includes pathological processes that predominantly affect the medial hemispheric limbic areas. Table 2.1-11 lists the most common causes of amnesia and the site of the associated lesion. Hippocampal injury occurs with closed head trauma, occlusion of the basilar or posterior cerebral artery, anoxia, hypoglycemia, tumors, and herpes simplex encephalitis. The hippocampus is involved early in the course of Alzheimer's disease, and an amnestic disorder is occasionally the presenting manifestation of this degenerative disorder. The fornix can be involved by surgery, brain tumors, and rupture of anterior cerebral artery aneurysms. Mammillary bodies are affected in the Wernicke-Korsakoff syndrome and by brain tumors. The medial thalamic nuclei are involved in Wernicke-Korsakoff syndrome, brain tumors, and some strokes (occlusion of thalamoperforant branches of the posterior communicating and posterior cerebral arteries).

Syndrome	Anatomical Site of Pathology
Wernicke-Korsakoff syndrome	Mammillary bodies, medial dorsal nucleus of thalamus
Head trauma	Hippocampus
Stroke (posterior cerebral artery)	Hippocampus
Subarachnoid hemorrhage (rupture of aneurysm of anterior cerebral artery)	Fornix
Anoxia	Hippocampus
Hypoglycemia	Hippocampus
Herpes simplex encephalitis	Hippocampus
Alzheimer's disease	Hippocampus
Neoplasms	Temporal lobe, fornix, hypothalamus, thalamus
Surgery (lobectomy)	Temporal lobe
Transient global amnesia	Hippocampal ischemia (in some cases)
Electroconvulsive therapy	Uncertain

Table 2.1-11 Differential Diagnosis of Amnesia and the Associated Site of the Lesion

A few clinical signs help to distinguish among the amnestic disorders. The Wernicke-Korsakoff syndrome classically evolves through an acute phase of Wernicke's

encephalopathy characterized by ophthalmoplegia, ataxia, and delirium. If the responsible thiamine deficiency is not reversed, the patient will progress into the Wernicke-Korsakoff syndrome that includes persistent amnesia, nystagmus, and ataxia. A peripheral neuropathy is usually present and a history of alcohol abuse is common, although any cause of thiamine deficiency (i.e., starvation, prolonged vomiting, gastric carcinoma) can produce the Wernicke-Korsakoff syndrome. Lesions of the hippocampus (e.g., posterior cerebral artery occlusion, brain tumors, trauma) commonly impinge on the adjacent fibers of the optic radiation and cause a superior quadrantanopsia or homonymous hemianopsia. Frontal–lobe-type behavioral disturbances commonly accompany posttraumatic amnesias because of orbitofrontal injury occurring in conjunction with the temporal lobe damage.

Amnesia must be distinguished from several other types of memory abnormalities. *Age-cognitive decline* (also known as *benign senescent forgetfulness*) occurs in the course of normal aging. It is characterized by impaired spontaneous recall, retained recognition skills, an absence of deficits in other neuropsychological domains, and an acute awareness of the loss of timely recall. Age-associated cognitive decline is not the harbinger of a dementing illness. The characteristic memory disturbance of *Alzheimer's disease* features both difficulty in learning new information and impaired remote recall. Throughout most of the course of the illness there is an inability to remember remote as well as newly learned material. The memory abnormality of the frontal lobe syndromes and *subcortical dementias* such as Huntington's disease and Parkinson's disease is a retrieval deficit with poor recall and intact recognition. Patients learn and retain information but have difficulty spontaneously accessing it. Unlike patients with amnesia, they are aided by clues, prompting strategies, and encoding enhancement techniques such as embedding the to-be-remembered material in a story.

Frontal Lobe Disorders Frontal lobe disorders are among the most important syndromes in neuropsychiatry. Frontal lobe dysfunction is present in many idiopathic psychiatric diseases (depression, schizophrenia, obsessive-compulsive disorder), and frontal lobe lesions produce profound behavioral alterations. Lesions of the deep subcortical as well as the superficial cortical structures of the frontal lobe produce behavioral disorders and “frontal systems” dysfunction more accurately reflects the anatomy of the conditions with frontal-lobe–type behaviors.

Three behavioral complexes are currently recognized with frontal systems dysfunction: an orbitofrontal disinhibition syndrome, a medial frontal apathetic-akinetic syndrome, and a frontal convexity neuropsychological deficit and motor programming disorder (Table 2.1-12). Combinations of symptoms are the rule rather than the exception in most patients with frontal lobe disorders.

Behavioral Syndrome	Location of Major Abnormality	Features
Disinhibition	Orbitofrontal	Impulsiveness, loss of tact, may be facetious
Apathetic	Medial frontal	Limited motivation and initiative
Motor programming deficits	Frontal convexity	Perseveration, impersistence, impaired serial-order behavior
Depression	Frontal polar and subcortical (caudate nucleus)	Major depressive symptoms

Table 2.1-12 Principal Frontal Lobe Syndromes

The *orbitofrontal syndrome* is the most well known and flamboyant of the frontal lobe behavioral disorders. The patients are disinhibited, making tactless remarks and acting on impulse. They are coarsened, lack empathy, and show little concern for the feelings of others. They fail to plan ahead and exhibit little concern about their illness or future. Mood is typically irritable and labile with a fatuous euphoria. Inappropriate jocularity with an insensitive humor (*witzelsucht*) may be observed. There is a lack of social restraint and an undue familiarity with strangers. Despite the marked behavioral alterations, these patients may have few or no neuropsychological deficits and exhibit intact language, memory, and visuospatial skills. The orbitofrontal syndrome occurs with posttraumatic encephalopathy, frontal brain tumors (particularly inferior frontal meningiomas), anterior cerebral artery stroke or aneurysm rupture, multiple sclerosis, and degenerative frontal lobe diseases such as Pick's disease. Deficits in olfaction commonly accompany the syndrome because of the proximity of the olfactory nerves, bulbs, and tracts to the orbital surface. This syndrome has been observed with inferior caudate lesions as well as disturbances of the orbitofrontal cortex.

The *medial frontal lobe syndrome* is also a striking behavioral disorder. The severity of the syndrome extends from *akinetic mutism* at the most severe end of the spectrum to a mild lack of motivation on the other. Akinetic mutism occurs with bilateral medial frontal lesions and is characterized by mutism and very limited spontaneous movement. The patients, however, appear alert, have their eyes open, exhibit ocular following movements of environmental events, will eat when fed, and may move with persistent stimulation or pain. The syndrome has been reported with bilateral anterior cerebral artery occlusions, trauma, hydrocephalus, bilateral thalamic infarction, and midline tumors of the thalamus, third ventricle, hypothalamus, and pituitary. Partial amelioration of the syndrome by dopaminergic agents (levodopa [Larodopa], bromocriptine [Parlodel]) suggests that interruption of ascending dopaminergic fibers may play a role in the pathophysiology of the disorder. An attenuated form of this syndrome occurs with unilateral medial frontal lobe lesions. Loss of initiative, poor motivation, limited gesturing, and apathy characterize the spontaneous behavior of patients with medial frontal lesions. A variety of neurobehavioral and neuropsychological abnormalities may accompany the neuropsychiatric disorders. Patients with left-sided lesions manifest transcortical motor aphasia with little speech initiation, nonfluent output, intact comprehension, and preserved repetition. With lesions of either hemisphere, a callosal apraxia with inability to perform learned motor acts on command with the left hand may also be present. If the lesion extends posteriorly to the medial central gyrus, paresis of the contralateral leg and foot will be evident.

Frontal convexity lesions produce abnormalities of sequential behavior including perseveration (abnormal continuation of behavior), impersistence (abnormal early termination of behavior), deficits in set shifting, and disturbances of programming sequential motor acts such as alternating programs, reciprocal programs, multiple loops, and serial hand sequences (Fig. 2.1-1 and Fig. 2.1-2). Patients also have difficulty developing strategies that will facilitate copying complex figures (Fig. 2.1-3), exhibit poor judgment, and are concrete on tests of abstraction. There may be no elementary neurological abnormalities. Tumors, stroke, trauma, and frontal degenerations account for most cases of the frontal convexity syndromes. The condition has been observed with dorsal caudate lesions as well as frontal convexity lesions.

Neglect and Anosognosia Unilateral *neglect* refers to a hemi-inattention disorder in which the patient fails to attend to or act on stimuli in one hemi-universe. Visual neglect is most striking and easily demonstrable, but in its fully developed form hemi-sensory neglect includes all sensory modalities (e.g., hearing, touch). Hemi-motor neglect may also occur, with the patient failing to act in one hemi-space. The patient with neglect fails to draw or copy one half of objects (Fig. 2.1-1), may not dress or groom one half of the body, will not see or hear individuals and events occurring on one side of the body, and will extinguish (fail to perceive) stimuli on the side contralateral to the lesion when both sides of the body or visual field are stimulated. Hemi-sensory neglect is more common and more enduring with lesions of the right parietal lobe or right thalamus than with left-sided lesions, but it occurs with injury to either hemisphere. Motor neglect occurs with frontal or subcortical damage.

Anosognosia refers to the denial of hemiparesis that occurs in some patients with unilateral neglect. In some usages, the term has been extended to encompass all forms of denial of illness. A variety of anosognosic phenomena has also been described, enlarging the syndrome to include all types of abnormal regard for the affected limbs. *Hemiasomatognosia* refers to neglect of one side of the body when there is no hemiparesis, and *somatophrenia* is the term for denial of ownership of one's paretic limbs. When limb weakness is acknowledged but minimized, the term *anosodiaphoria* is used; hatred of a weak limb is called *misoplegia.* Preoccupation and personification (e.g., naming the limb, etc) are two other variants of anosognosia.

NEUROPSYCHIATRIC SYNDROMES

Delusions Delusions, false beliefs held despite evidence to the contrary, occur in a variety of neurological disorders. They are most common in Alzheimer's disease, vascular dementia, Huntington's disease, temporal lobe epilepsy, and multiple sclerosis. They also occur in posttraumatic encephalopathy, Wilson's disease and other degenerative extrapyramidal disorders, and brain tumors, but they are not common in these conditions. Delusions are rare in Parkinson's disease, except as a manifestation of dopamine or anticholinergic toxicity. Delusions are frequent in delirium. When delusions occur with focal lesions, the injury is usually in the left or right temporal-parietal region. With left-sided lesions, delusions occur in conjunction with Wernicke's aphasia; with right-sided lesions, prominent visual hallucinations

commonly accompany the delusional disorder. Together, these clinical observations indicate that delusions are most common in diseases affecting the temporal lobe cortex or the basal ganglia, particularly the caudate nucleus. Delusions are more frequent in disorders with bilateral than unilateral damage.

Several principles have evolved that aid in characterizing and understanding delusions in neurological diseases. First, delusions are not reactions to declining intellectual function; the more intact the delusional patient's cognition, the more complex the delusions tend to be. Patients with dementing illnesses have simple delusions, whereas patients with single strokes or other limited lesions exhibit complex, well structured, and firmly held delusional beliefs. The intellect is in the service of the delusional process. Second, no specific neuropsychological correlates of delusions (e.g., deficits in memory, language, visuospatial functions, frontal lobe abilities) have been consistently identified. Cognitive abnormalities indicative of frontal lobe or temporal lobe dysfunction are common in delusional patients. Third, there is no delusional content that distinguishes neurological illnesses from idiopathic psychotic processes such as schizophrenia. Most delusions in neurological illness tend to be persecutory beliefs, but any type of delusional content can be observed (Table 2.1-2). Fourth, visual hallucinations are more common in conjunction with delusions in neurological illness than with idiopathic psychoses. Auditory hallucinations occur with delusions in both neurological and idiopathic psychiatric disorders. Fifth, delusions are most common in diseases affecting both hemispheres such as the degenerative disorders and vascular dementia. Sixth, when delusions follow unilateral lesions, the laterality of the damage may influence the delusion content. Schneiderian first-rank symptoms are more common with left-sided lesions, and delusions of substitution and misidentification syndromes such as the Capgras's symptom are more common with right-hemisphere lesions. Seventh, delusions are not invariably linked to lesions in specific anatomical structures in the same way that a hemiparesis predictably follows a pyramidal tract lesion, and other nonanatomical factors (e.g., biochemical changes, developmental factors, environmental conditions) must participate in determining the occurrence of delusions. Eighth, it is common for the onset of delusions to be delayed for considerable periods after the occurrence of a brain insult. In temporal lobe epilepsy, for example, several years may elapse between the occurrence of the brain injury with the onset of the seizures and the first appearance of delusions. Ninth, the approach to treatment of delusions in neurological illness involves addressing both the delusions and the underlying illness. Delusions may respond to neuroleptic agents and rarely improve with anticonvulsants or other nonneuroleptic medications.

Hallucinations Hallucinations are sensory perceptions occurring without the appropriate stimulation of the corresponding sensory organ. Visual hallucinations are common in neurological illness. The differential diagnosis of visual hallucinations is presented in Table 2.1-13. Ocular diseases that produce total or partial blindness—enucleation, cataracts, macular degeneration, retinal diseases, or vitreous traction—are common causes of hallucinations. Vitreous traction produces unformed flashes of light; the other processes may produce formed or unformed hallucinations. Inflammation of the optic nerves from ischemic optic neuritis or optic nerve demyelination can cause phosphenes or unformed flashes of light.

Condition/Lesion Location	Cause
Ocular abnormality	Enucleation Cataracts Macular degeneration Vitreous traction
Optic nerve abnormality	Optic neuritis (ischemic, demyelinating, glaucoma)
Brainstem abnormality	Peduncular hallucinosis (infarction, tumor)
Cerebral hemispheric lesions	Geniculocalcarine lesions (infarction, tumor) Epilepsy
Neurological illnesses	Migraine Narcolepsy
Medical illnesses	Delirium
Toxic disorders	Delirium Withdrawal Hallucinogens
Psychiatric disorders	Schizophrenia Mania Depression
"Normal" conditions	Hypnagogic hallucinations Imaginary companions of childhood Sensory deprivation Sleep deprivation

Table 2.1-13 Differential Diagnosis of Visual Hallucinations

Lesions of the midbrain produce the syndrome of *peduncular hallucinosis* characterized by visual hallucinations typically of a formed type occurring in the evening and associated with a benign affect. A sleep disturbance is usually present. The condition rarely lasts more than a few days. Peduncular hallucinosis has been produced by midbrain strokes and brainstem tumors.

Lesions of the geniculocalcarine radiations produce a visual field defect. Hallucinations within the field abnormality are common in the early period soon after the injury occurs. Known as *release hallucinations*, these are usually formed, are not necessarily stereotyped from episode to episode, last from minutes to hours at a time, and may be modified by moving or closing the eyes.

Focal seizures may produce visual hallucinations. Occipital foci tend to produce unformed hallucinations while parietal and temporal foci produce formed hallucinations. *Ictal hallucinations* are brief and stereotyped. They may be associated with other seizure phenomena (e.g., interruption of consciousness, head or eye deviation) and are generally not lateralized in the visual field. *Micropsia* (things look small and far away), *macropsia* (things look large and close), and *metamorphopsia* (distortions) may also occur in the course of seizures.

Hallucinations in *migraine* vary from simple light flashes, to scintillating scotomata, to fully formed complex visions. The classical hallucination of migraine is a fortification spectra that looks like the jagged top of the wall of a fort or castle. Hallucinations precede the headache and may be the dominant aspect of the migrainous attack. Micropsia and macropsia may occur in migraine where they are termed the “Alice in Wonderland syndrome.”

Narcolepsy is associated with the tetrad of sleep attacks, visual hallucinations, cataplexy (loss of muscle tone with subsequent falling), and sleep paralysis (momentary inability to move or speak on awakening from sleep). The hallucinations occur on falling asleep (hypnagogic) or on awakening (hypnopompic) and are usually terrifying in nature. Documentation of sleep-onset rapid eye movement (REM) sleep with multiple sleep latency polysomnography confirms the diagnosis.

Visual hallucinations are common in delirium associated with toxic or metabolic encephalopathies, in alcohol or sedative-hypnotic withdrawal, and with ingestion of hallucinogens such as lysergic acid diethylamide (LSD), psylocibin, and mescaline, and dissociative agents such as phencyclidine (PCP). Visual hallucinations occur in up to 30 percent of patients with Parkinson's disease treated with dopaminergic agents (e.g., levodopa bromocriptine pergolide [Permax]). Dementia is a common predisposing factor in Parkinson's disease patients with hallucinations.

Visual hallucinations occur in idiopathic psychoses including schizophrenia, depression, and mania. They are rarely the dominant type of hallucination in these circumstances. Normal individuals may experience visual hallucinations on falling asleep or in the course of sensory or sleep deprivation. The imaginary companions of childhood often have a visual aspect.

Hallucinations in other sensory modalities are less common than visual hallucinations in neurological illnesses. *Auditory hallucinations* occur in conjunction with persecutory delusions in the delusional disorders described earlier. They may also occur with deafness, brainstem lesions, and epilepsy. *Tactile hallucinations* are reported in delirium, withdrawal (particularly opiate withdrawal), and in association with delusions of infestation. *Gustatory* and *olfactory hallucinations* occur in psychoses and epilepsy.

Depression Depressive symptoms are common in neurological diseases. In some cases, the mood changes represent grief for lost function, altered role and status, and increased dependence. In many, however, the depression is more severe than anticipated for the functional disability and greater than experienced by patients with nonneurological disabling diseases. In these conditions the depression is a behavioral manifestation of the brain dysfunction. Neurological diseases in which depression has been identified as a prominent abnormality include Alzheimer's disease, Parkinson's disease, Huntington's disease, Wilson's disease, idiopathic basal ganglia calcification, stroke and vascular dementia, multiple sclerosis, and epilepsy. Endocrine disorders, systemic illnesses, and a wide variety of medications are also capable of inducing a depressive disorder.

The diagnosis of mood disorders in patients with neurological illness presents a difficult challenge. Many neurological diseases without mood changes produce symptoms that are characteristic of depressive episodes—diminished pleasure and interest, weight loss, insomnia, agitation, retardation, fatigue, impaired concentration—and may lead to the misdiagnosis of depression. Other neurological illnesses, such as the parkinsonian disorders, produce psychomotor slowing, stooped posture, fatigue, and impaired concentration resembling depression; depression in these disorders can be overlooked or overdiagnosed. Experiential manifestations of depression including feelings of sadness, worthlessness, hopelessness, and recurrent thoughts of death or suicide are the most dependable

indicators of a depressive syndrome and should be carefully sought. Dementia patients, however, may be unable to describe these subjective symptoms and depression must be inferred from the associated symptoms.

A few observations pertain to most neurological disorders manifesting depression as part of their symptom complex. First, the entire spectrum of depression severity may be observed, with some patients manifesting few mood symptoms and others meeting all criteria for a major depressive episode. In some disorders such as Alzheimer's disease, major depressive disorder is rare while depressive symptoms are common. Second, frontal-subcortical systems structures are the sites most often implicated in the depression syndromes. Depression may occur with lesions elsewhere in the nervous system, but they are less common and less severe. Third, there is preliminary evidence that depressive symptoms may differ subtly in different neurological disorders. Suicide is common in Huntington's disease and epilepsy, but rare in Parkinson's disease; psychosis is common in depressed patients with epilepsy and rare in Parkinson's disease; the depression of Parkinson's disease is characterized by more anxiety and less guilt than primary depressive disorders; psychomotor retardation is more extreme in post-stroke depression than in other depression syndromes. Fourth, depression in neurological diseases usually responds to treatment with psychopharmacological agents or electroconvulsive therapy.

Mania Mania is much less common than depression in the course of neurological illness. Maniform symptoms have been observed in Huntington's disease, Wilson's disease, idiopathic basal ganglia calcification, stroke, trauma, multiple sclerosis, general paresis, viral encephalitis and postencephalitic syndromes, frontal degenerative disorders, and following thalamotomy. A variety of medications have also been reported to produce manic behavior including steroids, triazolobenzodiazepines, dopaminergic agents, thyroid preparations, sympathomimetics, and stimulants. Antidepressant agents may precipitate manic episodes in depressed patients.

Structural lesions producing mania usually involve the basotemporal region, parathalamic structures, or the inferior medial frontal lobe. When lateralized, the lesions have had a marked right-sided predominance. A family history of psychiatric illness is more common in patients with secondary mania than in normal controls or in patients with secondary depression, and genetic vulnerability may facilitate the appearance of mania in the setting of brain dysfunction. The natural history of secondary mania is variable, patients may experience a single manic episode, recurrent mania, or alternating periods of depression and mania. The optimal treatment of secondary mania has not been determined. Tranquilization may be necessary in the acute phases of the manic episode; lithium (Eskalith), carbamazepine (Tegretol), clonazepam (Klonopin), and valproate (Depakene) may be used in the treatment of secondary mania; these agents may also be used prophylactically in patients who require treatment with agents that have previously precipitated manic behavior.

Personality Alterations *Personality* refers to stable patterns of behavior that include the way one relates to, perceives, and thinks about the environment and oneself. This is the least explored area of neuropsychiatry. Personality changes have been difficult to quantify for study, and personality alterations may be difficult to distinguish from the delusional and mood disorders that also occur in neurological illness. Table 2.1-14 summarizes the personality alterations observed in specific neurological diseases and conditions.

Neurological Disease	Personality Alteration
Alzheimer's disease	Disengagement, indifference
Frontal lobe degenerations	Disinhibition or apathy
Huntington's disease	Irritability, impulsiveness
Human immunodeficiency virus (HIV) encephalopathy	Apathy
Focal lesions	
Orbitofrontal	Irritability, impulsiveness, "pseudopsychopathic"
Medial frontal	Apathy, loss of motivation and initiative
Right hemisphere (childhood)	Schizoid
Right hemisphere (adult)	Alexithymia
Left hemisphere (Wernicke's area)	Suspiciousness, irritability
Bilateral temporal lobe	Placidity (component of the Kluver-Bucy syndrome)
Ventromedial hypothalamus	Rage
Temporal lobe epilepsy	Suspiciousness, Geschwind's syndrome (hyperreligiosity, circumstantiality, hypergraphia, viscosity, hyposexuality)

Table 2.1-14 Personality Alterations Described in Neurological Illnesses

Alzheimer's disease has a profound impact on personality and the behavioral changes may predate the neuropsychological deficits. Early in the illness, patients become disengaged and indifferent, showing little concern about their own disease or insight into the feelings of family members. Late in the course, impulsiveness and aggression are often exhibited. Patients with other dementia syndromes exhibit contrasting personality alterations. Frontotemporal dementias such as Pick's disease usually produce disinhibition, impulsiveness, and facetiousness characteristic of orbitofrontal dysfunction or apathy suggestive of medial frontal pathology. Huntington's disease combines marked irritability with impulsiveness sometimes leading to aggression, violence, and suicide. Many subcortical dementias (e.g., Parkinson's disease, progressive supranuclear palsy, HIV encephalopathy) cause apathy and indifference.

Focal lesions also may be associated with personality alterations. The marked personality alterations following orbitofrontal and medial frontal lesions have already been described. Patients with Wernicke's aphasia are often suspicious, demanding, aggressive, and irritable. Preliminary studies of adults with right hemisphere lesions suggest that they may exhibit alexithymia with reserved expression of emotions and a tendency to forego symbolic thought. Right hemisphere damage sustained in childhood may result in a schizoid type of behavioral pattern, perhaps because the inability to perceive or execute emotional cues limits the child's ability to engage in interpersonal relationships.

Limbic system lesions have profound effects on personality. Ventromedial hypothalamic lesions produce a syndrome of dementia, hyperphagia, and rage. Bilateral medial temporal lobe lesions produce the Kluver-Bucy syndrome of placidity, hypersexuality and altered sexual behavior, visual agnosia, hypermetamorphosis (compulsive exploration of environmental stimuli), and hyperorality.

A controversial area of research in neuropsychiatry concerns personality changes in epilepsy. Personality inventories reveal that the most common type of personality change in epilepsy patients is increased suspiciousness and paranoia. An uncommon syndrome (known as the *temporal lobe epilepsy personality* or *Geschwind's syndrome*) occurring in patients with partial complex seizures consists of hypergraphia, circumstantiality, interpersonal viscosity, hyperreligiosity, and hyposexuality. This personality style is not unique to epilepsy but occurs with increased frequency in patients with epilepsy and can be the presenting manifestation of the limbic seizure disorder.

Anxiety Anxiety is a state of apprehension, tension, or uneasiness that occurs in anticipation of internal or external danger. The anxiety syndrome includes motor tension, autonomic hyperactivity, apprehensive expectation, and heightened vigilance. Anxiety occurs in a variety of neurological and medical disorders and can be precipitated by drugs. Neurological diseases causing anxiety include brain tumors (particularly in the regions of the temporal lobe or third ventricle), trauma, stroke, migraine, encephalitis, multiple sclerosis, epilepsy, Alzheimer's disease, Parkinson's disease, Huntington's disease, and Wilson's disease. The principal medical conditions associated with anxiety include hypoxia, hypoglycemia, hyperthyroidism, Cushing's disease, and mitral valve prolapse. Medications capable of causing anxiety are cocaine, sympathomimetics (e.g., amphetamines), caffeine, organophosphate toxicity, lidocaine, and procaine penicillin. Alcohol and sedative-hypnotic withdrawal are frequently accompanied by anxiety. When anxiety is associated with focal brain lesions, the pathology usually involves the right temporal lobe.

Treatment of anxiety in neurological disorders requires attention to the underlying condition as well as the anxiety symptoms. When pharmacological agents are indicated, conventional anxiolytics and b-adrenergic receptor antagonists are useful.

Obsessive-Compulsive Disorder *Obsessions* are recurrent, intrusive, senseless ideas, thoughts, and images that are ego-dystonic and involuntary. *Compulsions* are repetitive activities carried out in response to an obsession or executed in a stereotyped and ritualized fashion. Lesions in basal ganglia are associated with obsessive-compulsive behavior (Table 2.1-15). Conditions affecting the basal ganglia and producing obsessive-compulsive disorder include Parkinson's disease, postencephalitic Parkinsonism, Huntington's disease, progressive supranuclear palsy, Tourette's disorder, neuroacanthocytosis, Sydenham's chorea, carbon monoxide poisoning, neonatal hypoxia, bilateral caudate infarctions, cardiopulmonary arrest, and manganese poisoning. Obsessional thinking has also been observed in conjunction with neuroleptic-induced and postencephalitic oculogyric crises. Treatment of obsessions and compulsions in neurological disorders entails treatment of the underlying disease and symptomatic relief of the obsessive-compulsive disorder.

Acute dystonia with oculogyric crises
Anoxia with bilateral globus pallidus lesions
Anoxic injury to caudate and putamen
Bilateral caudate nucleus infarctions
Carbon monoxide intoxication with bilateral globus pallidus injury
Carbon monoxide poisoning with bilateral caudate nucleus lesions
Huntington's disease
Manganese intoxication
Neuroacanthocytosis
Parkinsonism with compulsions during "on" period in patients experiencing on-off swings
Postencephalitic parkinsonism
Progressive supranuclear palsy
Sydenham's chorea
Tourette's disorder

Table 2.1-15 Neurological Disorders Associated With Obsessive-Compulsive Disorder

Paraphilias and Altered Sexual Drive Neurological disorders can produce paraphilic behavior and alterations in sexual drive. Markedly diminished sexual interest occurs with temporal lobe epilepsy, hypothalamic lesions, and right hemisphere brain injuries. Heightened sexual drive occurs in secondary mania, postictally (i.e., after a seizure), following markedly improved seizure control in patients with epilepsy (e.g., after temporal lobectomy or with improved anticonvulsant control of seizures), with introduction of levodopa or other dopaminergic agents in Parkinson's disease, with diencephalic or frontal lobe lesions, after septal injury, and in the Kluver-Bucy syndrome. Hypersexuality has also been induced by amphetamines, cocaine, hyperthyroidism, hypercortisolism, and androgen administration.

Paraphilias are characterized by intense sexual urges and sexually arousing fantasies usually involving nonhuman objects or animals, the suffering or humiliation of oneself or one's partner, or children or other nonconsenting persons. Examples of paraphilias include exhibitionism, fetishism, frotteurism, pedophilia, sexual masochism, sexual sadism, transvestitism, voyeurism, bestiality (zoophilia), and telephone scatologia. Paraphilic behavior has been described in association with temporal lobe epilepsy, postencephalitic states, Tourette's disorder, frontal lobe disorders, Huntington's disease, brainstem tumors, bilateral temporal injury, septal injury, and multiple sclerosis. Paraphilic behavior has occasionally occurred with dopaminergic therapy of Parkinson's disease.

Ganser's Syndrome Ganser's syndrome is a syndrome of approximate answers that may be a manifestation of malingering, confusional state, or disinhibition syndromes. Patients with Ganser's syndrome will respond to items on mental status examinations in such an approximately correct fashion as to raise suspicion of duplicity (e.g., “five quarters in a dollar, 13 months in one year.”

Table 2.1-16 lists acquired neuropsychiatric syndromes that have lateralized origins, arising from left, right, or bilateral hemisphere damage.

Left Hemisphere	Right Hemisphere	Bilateral
Aphasia	Delusions with Schneiderian first-rank symptoms	Gerstmann's syndrome (acalculia, finger agnosia, right/left disorientation, agraphia without alexia)
Ideomotor apraxia	Primary progressive aphasia	Verbal amnesia
Depression (LR)	Alexithymia Aprosodia Capgras syndrome Mania Nonverbal amnesia Social impropriety Visuospatial abnormalities (RL)	Abulia Kluver-Bucy syndrome (placidity, altered sexual behavior, visual agnosia, hypermetamorphosis, hyperorality) Severe amnesia

Table 2.1-16 Etiological Lateralization of Neuropsychiatric Syndromes

PRINCIPLES OF NEUROPSYCHIATRY

This summary of neuropsychiatric syndromes leads to several generalizations that are applicable to clinical practice. First, essentially all types of behavior observed in idiopathic psychiatric syndromes can occur in association with a neurological disorder and neurological diseases are an important part of the differential diagnosis of behavioral disturbances. Second, the phenotypic identity of psychiatric and neurological disorder does not necessarily imply pathogenetic identity, but shared pathophysiological mechanisms in similar disorders is likely and provides a framework for treatment and research. Third, diagnosis informs therapy and a thorough diagnostic evaluation should be a part of every psychiatric assessment. Fourth, deficit syndromes frequently accompany productive symptoms and a probing mental status examination aids in the identification and characterization of neuropsychiatric syndromes. Fifth, laboratory tests and neuroimaging play an important role in confirming clinical diagnoses. Sixth, late-onset disorders are more likely to be related to brain diseases, but neuropsychiatric disorders are not limited to the elderly. Seventh, patients with neuropsychiatric illnesses usually lack the premorbid psychiatric or personality alterations that occur in many idiopathic psychiatric disorders. Eighth, the presence of neuromedical illness in a patient with a behavior disturbance should lead the clinician to explore potential relationships between the two. Ninth, patients with neuropsychiatric illness often lack the family history of psychiatric disability common in many with idiopathic psychiatric disorders; however, hereditary neuropsychiatric illnesses such as Huntington's disease are exceptions to this rule. Tenth, pharmacotherapy relies on intervention in brain processes and is applicable in neuropsychiatric as well as idiopathic psychiatric disorders. Neuropsychiatric syndromes are more common in patients who are older or who have sustained brain injuries and drug dosages will usually have to be adjusted downward to avoid pharmacotoxicity.

SUGGESTED CROSS-REFERENCES

Section 1.2 on Functional Neuroanatomy contains a review of brain function localization. Chapter 3 discusses aspects of cognition and memory as viewed from a psychological perspective. Section 2.8 on Neuropsychiatric Aspects of HIV Infection and AIDS contains more information on the impact of HIV infection on cognition. Clinical assessment and diagnosis are also discussed in Chapter 7. A complete list of signs and symptoms in psychiatry are in Section 7.3. Chapter 51 on Geriatric Psychiatry contains sections addressing the dementias and neurological changes seen in aging.

SECTION REFERENCES

Adams RD, Victor M: *Principles of Neurology*, ed 5. McGraw-Hill, New York, 1993.

Alexander MP: Mild traumatic brain injury: Pathophysiology, natural history, and clinical management. Neurology *45:*1253, 1995.

Cummings JL, Mega MS, Gray K, Rosenberg-Thompson S, Carusi DA, Gornbein J: The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. Neurology *44:*2308, 1994.

Cummings JL: Neuropsychiatric manifestations of right hemisphere lesions. Brain and Language *57:*22, 1997.

*Cummings JL, *Trimble M: Concise Guide to Behavioral Neurology and Neuropsychiatry*. American Psychiatric Association, Washington, DC, 1994.

Cutting J: *The Right Cerebral Hemisphere and Psychiatric Disorders.* Oxford Medical Publications, New York, 1990.

*Devinsky O: *Behavioral Neurology: 100 Maxims*. Mosby Yearbook, St Louis, 1992.

Edwards-Lee T, Miller BL, Benson DF, Cummings JL, Russell GL, Boone K, Mena I: The temporal variant of frontotemporal dementia. Brain *120:*1027, 1997.

Folstein MF, Folstein SE, McHugh PR: “Mini-Mental State.” A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res *12:*189, 1975.

*Gifford DR, Cummings JL: Evaluating dementia screening tests. Methodologic standards to rate their performance. Neurology *52:*224, 1999.

Kiernan RJ, Mueller J, Langston JW, Van Dyke C: The Neurobehavioral Cognitive Status Examination: A brief but differentiated approach to cognitive assessment. Ann Intern Med *107:*481, 1987.

Laplane D, Levasseur M, Pillon B, Dubois B, Baulac M, Mazoyer B, Tran Dinh S, Sette G, Danze F, Baron JC: Obsessive-compulsive and other behavioural changes with bilateral basal ganglia lesions. Brain *112:*699, 1998.

Levy ML, Miller BL, Cummings JL, Fairbanks LA, Craig A: Alzheimer disease and frontotemporal dementias. Behavioral distinctions. Arch Neurol *53:*687, 1996.

Lishman WA: *Organic Psychiatry*, ed 3. Blackwell Scientific Publications, Boston, 1997.

Litvan I, Mega M, Cummings JL, Fairbanks L: Neuropsychiatry aspects of progressive supranuclear palsy. Neurology *47:*1184, 1996.

*Mega MS, Cummings JL: Frontal-subcortical circuits and neuropsychiatric disorders. J Neuropsychiatry Clin Neurosci *6:*358, 1994.

Mega MS, Cummings JL, Fiorello T, Gornbein J: The spectrum of behavioral changes in Alzheimer's disease. Neurology *46:*130, 1996.

Mendez MF, VanGorp W, Cummings JL: Neuropsychiatry, neuropsychology, and behavioral neurology: A critical comparison. Neuropsychiatr Neuropsychol Behav Neurol *8:*297, 1995.

Mendez MF, Grau R, Doss RC, Taylor JL: Schizophrenia in epilepsy: Seizure and psychosis variables. Neurology *43:*1073, 1993.

*Miller BL, Cummings JL: *The Human Frontal Lobes: Functions and Disorders*. Guilford, New York, 1999.

Miller BL, Cummings JL, McIntyre H, Ebers G, Grode M: Hypersexuality or altered sexual preference following brain injury. J Neurol Neurosurg Psychiatry *49:*867, 1986.

Pearlman AL, Collins RC, editors: *Neurobiology of Disease*. Oxford University Press, New York, 1990.

*Plum F, Posner JB: *The Diagnosis of Stupor and Coma*, ed 3. FA Davis Company, Philadelphia, 1980.

Robinson RG, Starkstein SE: Current research in affective disorders following stroke. J Neuropsychiatry Clin Neurosci *2:*1, 1990.

Shukla S, Cook BL, Mukherjee S, Godwin C, Miller MG: Mania following head trauma. Am J Psychiatry *144:*93, 1987.

*Squire LR, Zola-Morgan S: The medial temporal lobe memory system. Science *253:*1380, 1991.

Starkstein SE, Mayberg HS, Bertheir ML, Fedoroff P, Price TR, Dannals RF, Wagner HN, Leiguarda R, Robinson RG: Mania after brain injury: Neuroradiological and metabolic findings. Ann Neurol *27:*652, 1990.

Strub RL, Black FW: *The Mental Status Examination in Neurology*, ed 2. FA Davis, Philadelphia, 1985.

Sultzer DL, Cummings JL: Drug-induced mania–causative agents, clinical characteristics and management. Med Toxicol Adverse Drug Exp *4:*127, 1989.

Tandberg E, Larsen JP, Aarsland D, Cummings JL: The occurrence of depression in Parkinson's disease. A community-based study. Arch Neurol *53:*175, 1996.

Tom T, Cummings JL: Depression in Parkinson's disease: Pharmacological characteristics and treatment. Drugs Aging *12:*55, 1998.

Wright MT, Cummings JL, Mendez MF, Foti DJ: Bipolar syndromes following brain trauma. Neurocase *3:*111, 1997.

2.2 NEUROPSYCHIATRIC ASPECTS OF CEREBROVASCULAR DISORDERS

ROBERT G. ROBINSON, M.D., AND SERGIO E. STARKSTEIN, M.D., FH.D.

Cerebrovascular disease represents one of the major health problems in the United States, with an estimated annual incidence for thromboembolic stroke between 300,000 and 400,000. Since the late 1970s, however, there has been a steady decline in the incidence of stroke, presumably related to the improved control of hypertension. Nevertheless, among adults over age 50, stroke remains the third leading cause (after heart disease and cancer) of mortality and morbidity in the United States.

Although virtually any psychiatric disorder may occur in patients with cerebrovascular disease, the only disorder listed in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) that is specific to cerebrovascular disease is vascular dementia (Table 2.2-1). Other disorders that may be related to cerebral infarction or hemorrhage are shown in Table 2.2-1. These diagnoses require that “there is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of the cerebrovascular disorder.”

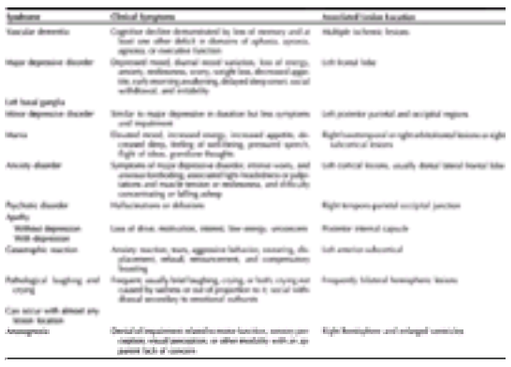

Table 2.2-1 Clinical Syndromes Associated With Cerebrovascular Disease

HISTORY

The first reports of emotional reactions following brain damage (usually caused by cerebrovascular disease) were made by neurologists and psychiatrists in case descriptions. For example, Adolf Meyer proposed that insanities following brain injury may be related to specific locations and causes of brain injury although he believed that most of these disorders resulted from a combination of psychological, social, and biological factors. Joseph Babinski noted that patients with right hemisphere disease frequently displayed the symptoms of anosognosia, euphoria, and indifference. Eugen Bleuler recognized that “melancholic moods following stroke may last for months and sometimes longer.” Emil Kraepelin stated that cerebrovascular disorder may either accompany manic depressive disease (bipolar I disorder) or may itself engender states of depression.

Kurt Goldstein was the first to describe an emotional disorder that he termed *catastrophic reaction* thought to be uniquely associated with brain disease. Catastrophic reaction is an emotional outburst involving various degrees of anger, frustration, depression, tearfulness, refusal, shouting, swearing, and sometimes aggression. Goldstein ascribed this reaction to the organism's inability to cope when faced with a serious defect in its physical or cognitive functions.

A second emotional abnormality, also thought to be characteristic of brain injury, was the *indifference reaction* described by Derek Denny-Brown. Associated with right hemisphere lesions, this reaction consists of symptoms of indifference toward failure, lack of interest in family and friends, enjoyment of foolish jokes, and minimization of physical difficulties. In the late nineteenth century, Leonore Welt first described euphoria and loquaciousness associated with orbital frontal injury. Hermann Oppenheim used the term *Witzelsucht* to refer to the inappropriate sense of humor in these patients and Karl Kleist distinguished the orbital frontal cortex as the center of emotional life and the dorsal lateral frontal cortex as the source of psychomotor and intellectual activity.

A third emotional disorder that has been historically associated with brain injury such as cerebral infarction is pathological laughter or crying in which patients' emotional displays are characteristically unrelated to their inner emotional state. For example, crying might occur spontaneously or after some seemingly minor provocation.

COMPARATIVE NOSOLOGY

The most pragmatic way to classify cerebrovascular disease is to examine the means by which parenchymal changes occur in the brain. The first of these, ischemia, may occur either with or without infarction of parenchyma, and includes transient ischemic attacks, atherosclerotic thrombosis, cerebral embolism, and hemorrhage (Table 2.2-2). Hemorrhage may cause either direct parenchymal damage by extravasation of blood into the surrounding brain tissue, as in intracerebral hemorrhage, or indirect damage by hemorrhage into the ventricles, subarachnoid space, extradural area, or subdural area (Fig. 2.2-1). These changes result in a common mode of expression: a sudden, convulsive, focal neurological deficit or stroke.

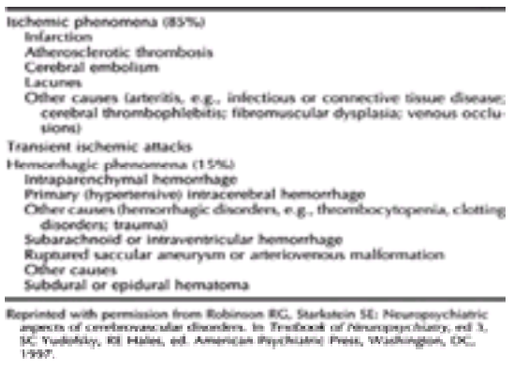

Ischemic phenomena (85%)
- Infarction
- Atherosclerotic thrombosis
- Cerebral embolism
- Lacunes
- Other causes (arteritis, e.g., infectious or connective tissue disease; cerebral thrombophlebitis; fibromuscular dysplasia; venous occlusions)

Transient ischemic attacks

Hemorrhagic phenomena (15%)
- Intraparenchymal hemorrhage
- Primary (hypertensive) intracerebral hemorrhage
- Other causes (hemorrhagic disorders, e.g., thrombocytopenia, clotting disorders; trauma)
- Subarachnoid or intraventricular hemorrhage
- Ruptured saccular aneurysm or arteriovenous malformation
- Other causes
- Subdural or epidural hematoma

Reprinted with permission from Robinson RG, Starkstein SE: Neuropsychiatric aspects of cerebrovascular disorders. In Textbook of Neuropsychiatry, ed 3. SC Yudofsky, RE Hales, ed. American Psychiatric Press, Washington, DC, 1997.

Table 2.2-2 Classification of Cerebrovascular Disease

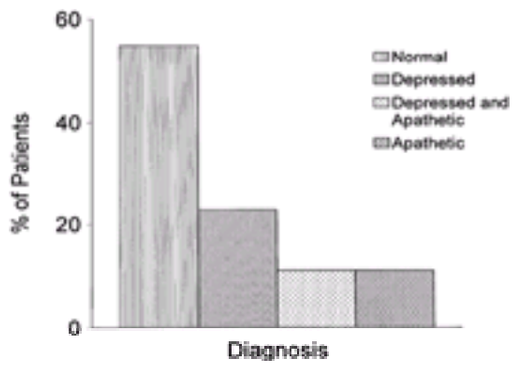

FIGURE 2.2-1 The percentage of patients with apathy, depression, or both following acute stroke among 80 consecutive patients. About half of the patients with apathy had an accompanying depressive disorder. (Data from Starkstein SE, Fedoroff JP, Price TR, Leiguarda R, Robinson RG: Apathy following cerebrovascular lesions. Stroke *24:* 1625, 1993.)

There are four major categories of cerebrovascular disease (Table 2.2-1): atherosclerotic thrombosis, cerebral embolism, lacunae, and intracranial hemorrhage. In various studies of the incidence of cerebrovascular disease, the ratio of infarcts to hemorrhages has been shown to be about 5 to 1. Atherosclerotic thrombosis and cerebral embolism each account for approximately one third of all strokes. There are many other less common causes of cerebrovascular disease.

EPIDEMIOLOGY

Vascular Dementia In a review of population-based studies the European Community for Concerted Action on Epidemiology and Prevention of Dementia found a consistent increase in the lifetime prevalence of vascular dementia with advancing age. Prevalence rates ranged from 1.5 per 100 for women ages 75 to 79 years in the United States to 16.3 per 100 for men older than age 80 in Italy. In most age groups men had a higher prevalence of vascular dementia than did women. Vascular dementia is the most common type of dementia in Japan, representing up to 50 percent of all clinical cases and from 54 to 65 percent of all autopsy-confirmed cases of dementia (Table 2.2-3).

Disorder	Prevalence rate	Studies
Vascular dementia	1.5/100; women 75–79 yrs; USA	Rocca et al. (1991)
	16.3/100 men >80 yrs; Italy	Rocca et al. (1991)
Major depressive disorder	10/100 dementia cases in USA	Kaszniel et al. (1988)
	15/100 strokes in community	Burvill et al. (1995)
	28/100 strokes in acute-care hospital	Robinson & Starkstein (1987)
Mania	<1/100 strokes in acute-care hospital	Robinson & Starkstein (1987)
Anxiety	6/100 strokes in acute-care hospital	Castillo et al. (1993)
	27/100 anxiety + acute depression	Castillo et al. (1993)
	28/100 anxiety ± acute depression	Astrom et al. (1993)
	3.5/100 anxiety neurosis in community	House et al. (1992)
Psychosis	<1/100 strokes in acute-care hospital	Robinson & Starkstein (1987)
Apathy	11/100 apathy only–acute-care hospital	Starkstein et al. (1992a)
	11/100 apathy ± depression in acute-care hospital	Starkstein et al. (1992a)
Catastrophic reaction	19/100 strokes in acute-care hospital	Starkstein et al. (1992b)
Pathological emotions	15/100 acute stroke in community	House et al. (1991)
	21/100 6-month poststroke	House et al. (1991)
	12/100 1-year poststroke	House et al. (1991)
	18/100 rehabilitation hospital	Robinson & Starkstein (1987)
Anosognosia	24/100 strokes in acute-care hospital	Starkstein et al. (1992)

Table 2.2-3 Prevalence Studies of Psychiatric Disorders After Stroke

In two autopsy series, stroke accounted for approximately 20 to 25 percent of all dementia cases, and 10 to 15 percent of cases were thought to be the result of a combination of vascular disease and dementia of the Alzheimer's type. In a clinical series using in vivo imaging, the proportion of dementia that was directly attributable to stroke was 10 to 15 percent.

Depression Depressive disorders are probably the most common emotional disorder associated with cerebrovascular disease. The prevalence depends upon whether community-based samples or hospitalized patients are studied or whether patients with acute stroke or those with chronic stroke are evaluated. Approximately 15 to 25 percent of community-based samples of acute stroke patients and 30 to 40 percent of patients hospitalized with acute stroke have a clinically diagnosable major or minor depressive disorder (Table 2.2-3).

Mania Mania occurs much less frequently than depression following stroke (only 3 cases were identified among a consecutive series of more than 300 acute stroke patients including 143 patients with longitudinal assessment). Although numerous case reports and empirical studies document that stroke is associated with mania, there are no epidemiological studies that document the incidence or prevalence of this condition.

Anxiety In a consecutive series of 309 patients with first-episode acute stroke lesions, only 6 percent met the criteria for generalized anxiety disorder in the absence of any other mood disorder. On the other hand, 27 percent of the 309 met the criteria of the revised third edition of DSM (DSM-III-R) for generalized anxiety disorder (excluding the 6-month duration criteria). The majority of patients with generalized anxiety disorder also had major or minor depression (i.e., 21 percent of the 27 percent with generalized anxiety disorder). A Swedish study also found that 28 percent of 80 patients with acute stroke had this disorder; 55 percent also had major depression (Table 2.2-3). A community-based study reported that 3.5 percent of 89 patients with stroke met of the ninth revision of *International Statistical Classification of Diseases* (ICD-9) criteria for anxiety neurosis at 1 month poststroke. Thus, anxiety disorder following stroke is usually comorbid with a depressive disorder but a small number of patients will have anxiety alone.

Psychosis Although it is rare, case reports and empirical studies have documented that psychosis may occur after stroke. No epidemiological study has documented the incidence or prevalence of psychosis following stroke.

Apathy Apathy is the absence or lack of feeling, emotion, interest, concern, or motivation and has been reported frequently among patients with brain injury. Using an apathy scale, of 80 consecutive patients with single stroke lesions, 9 (11 percent) showed apathy as their only psychiatric disorder whereas another 11 percent had both apathy and depression (Fig. 2.2-1; Table 2.2-3).

Catastrophic Reaction *Catastrophic reaction* is a term coined by Kurt Goldstein to describe anxiety, tears, aggressive behavior, swearing, displacement, refusal, renouncement, and compensatory boasting that he attributed to an “inability of the organism to cope when faced with physical or cognitive deficits.” Using a Catastrophic Reaction Scale (CRS) developed to assess the existence and severity of the catastrophic reaction, 12 of 62 consecutive patients (19 percent) with acute stroke lesions were found to have catastrophic reactions (Table 2.2-3).

Pathological Emotions *Pathological emotions* are characterized by episodes of laughing, or crying or both, that are not appropriate to the context. These episodes may appear spontaneously or may be elicited by nonemotional events and do not correspond to underlying emotional feelings (Table 2.2-2). This condition was found in 13 of 89 patients (15 percent) seen at 1 month poststroke, in 21 percent at 6 months, and in 12 percent at 1 year (Table 2.2-3). Other studies have reported frequencies of 18 percent in patients at a rehabilitation hospital and of 14 percent in patients in a community-based study.

Anosognosia The term *anosognosia* was first used by Joseph Babinski in 1914 to indicate the lack of awareness of hemiplegia. It has also been used to refer to unawareness of other poststroke deficits such as cortical blindness, hemianopia, and amnesia. Among 80 acute stroke patients, 24 percent had moderate or severe anosognosia (Table 2.2-3).

ETIOLOGY

Poststroke Depression Although the cause of poststroke depression is unknown, a number of studies have suggested that location of brain injury may play an important role. The first study to report a significant role for lesion location in poststroke depression examined 29 patients with left-hemisphere brain injury secondary to stroke (*N* = 18) or traumatic brain injury (*N* = 11) and found a significant inverse correlation between the severity of depression and the distance of the anterior border of the lesion from the frontal pole (Fig. 2.2-2). This surprising finding led to a number of subsequent examinations of this phenomenon in other populations; all these studies found a significant correlation between severity of depression and proximity of the lesion to the frontal pole.

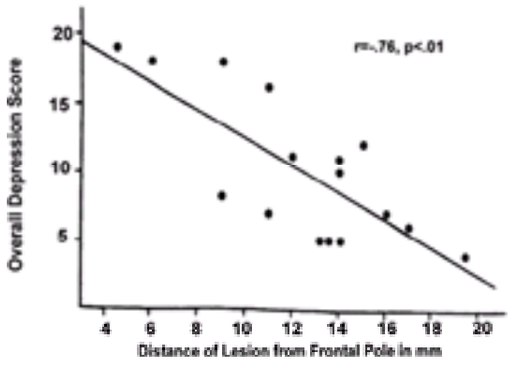

FIGURE 2.2-2 Scatter plot of the relationship between severity of depression in patients with stroke or traumatic brain injury and distance of the anterior border of the lesion from the frontal pole as measured on CT scan. For both stroke and trauma lesions, depression increased with proximity of the lesion to the frontal pole. (Reprinted with permission from Robinson RG, Szetela B: Mood change following left hemispheric brain injury. Ann Neurol *9:*447, 1981.)

Other studies of patients with acute stroke have also found that hemispheric lesion location plays an important role in poststroke depression. For example, a study of 45 patients with single lesions restricted to either cortical or subcortical structures in the left or right hemisphere found that significantly more (44 percent) patients with left cortical lesions or left subcortical lesions (39 percent) were depressed compared to patients with right cortical lesions (11 percent) or patients with right subcortical lesions (14 percent) (Fig. 2.2-3). Although some studies have not replicated this finding, most of the studies with negative findings examined patients who were beyond the acute stroke period. Other studies of patients within the first few weeks poststroke have found lateralized effects of infarction on major depressive disorder.

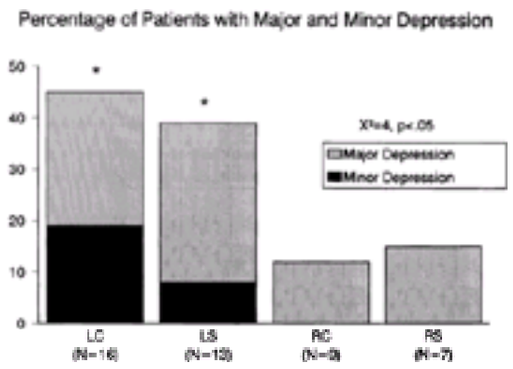

FIGURE 2.2-3 The frequency of depression in patients with acute first-ever stroke lesions restricted to cortical or subcortical brain regions. Both cortical and subcortical lesions of the left hemisphere were associated with greater frequency of depression than patients with right hemisphere lesions. (Data from Starkstein SE, Robinson RG, Price TR: Comparison of cortical and subcortical lesions in the production of poststroke mood disorders. Brain *110:*1045, 1987.)

Although significantly more patients with left anterior lesions developed poststroke depression compared with other lesion locations, not every patient with a left anterior lesion developed a depressive disorder. That raised the question of why some but not all patients with lesions in these locations develop depression. Thirteen patients with major poststroke depression were compared to 13 stroke patients without depression who were matched for lesion size and location. Patients with major depressive disorder, however, had significantly more subcortical atrophy as measured both by the ratio of third ventricle to brain (i.e., the area of the third ventricle divided by the area of the brain at the same level), and the ratio of lateral ventricle to brain (i.e., the area of the lateral ventricle contralateral to the brain lesion divided by the brain area at the same level) (Fig. 2.2-4). It is likely that the subcortical atrophy preceded the stroke because it was visible within a few days after the stroke and was found on the side of the brain opposite the lesion.

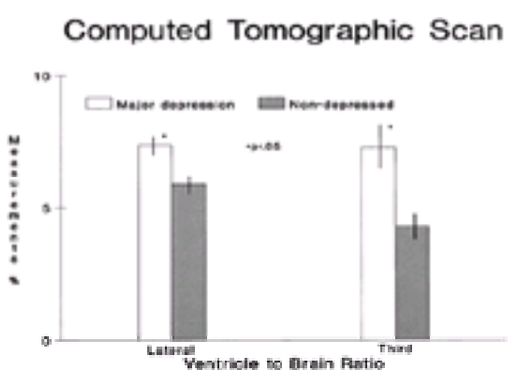

FIGURE 2.2-4 Ventricular to brain ratios in patients with and without major depression who were matched for lesion size and location. Patients with depression had larger ventricles, suggesting a mild degree of subcortical atrophy perhaps occurring prior to stroke, making them more vulnerable to developing depression. (Data from Starkstein SE, Robinson RG, Price TR: Comparison of patients with and without poststroke major depression matched for size and location of lesion. Arch Gen Psychiatry *45:*247, 1988.)

Several studies have reported that patients with depression were more likely than patients without depression to have either a previous personal history or a family history of psychiatric disorders. In addition, a low but significant correlation has been found between depression and functional physical impairment (i.e., activities of daily living); this association, however, was not found by all investigators. Thus, the severity of physical impairment does not appear to be a major determinant of depression. Several studies have found that depression may adversely affect physical recovery from stroke. One study examined a consecutive series of 63 stroke patients with subsequent major or minor depressive disorder compared to nondepressed stroke patients during a 2-year follow-up. Although both groups had similar impairments in activities of daily living during the time they were in the hospital, the depressed patients had significantly less improvement by 2 years follow-up than the nondepressed patients (Fig. 2.2-5). This finding held true after controlling for important variables such as the type and extent of in-hospital and rehabilitation treatment, the size and location of the lesion, the patient's demographic characteristics, the nature of the stroke, the occurrence of another stroke during the follow-up

period, and medical history.

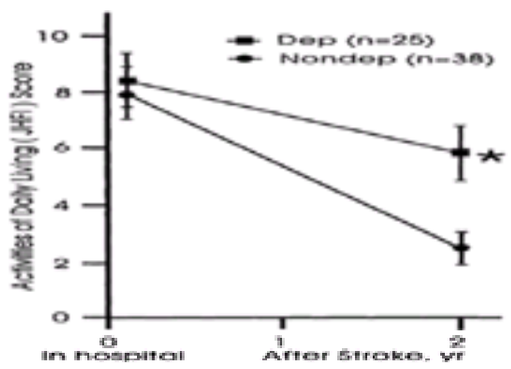

FIGURE 2.2-5 Change in activities of daily living scores (as measured by JHFI) for in-hospital depressed patients (either major or minor depression) and in-hospital nondepressed patients. At the time of the in-hospital evaluation, cognitive impairment associated with major depression and left hemisphere stroke lasted for about 1 year. (Reprinted with permission from Parikh RM, Robinson RG, Lipsey JR, Starkstein SE, Fedoroff JP, Price TR: The impact of post-stroke depression on recovery in activities of daily living over two-year follow-up. Arch Neurol *47:*785, 1990.)

Depression has also been found to be more common in nonfluent aphasic patients as compared to fluent aphasic patients. Although it may be that the higher frequency of depression among nonfluent aphasic patients is related to their greater awareness of their impairment, aphasia and depression do not appear to be causally related. The association between nonfluent aphasia and poststroke depression may be explained by the fact that the lesion location that produces nonfluent language may also produce depression.

It has been suggested that poststroke depression may be the consequence of severe depletions of norepinephrine or serotonin produced by frontal or basal ganglia lesions. In support of this hypothesis, a positron emission tomography (PET) study found that patients with left hemisphere stroke showed a significant inverse correlation between the amount of *N*-methylspiperone binding (presumably serotonin [5-hydroxytryptamine (5-HT)] type 2 [5-HT $_2$] receptor binding) in the left temporal cortex and severity of depression (i.e., higher depression rating scale scores were associated with lower serotonin receptor binding). Patients with right hemisphere lesions on the other hand had an increase in 5-HT $_2$ receptor binding in the temporal and parietal cortex. Thus, a greater depletion of biogenic amines in patients with right hemisphere lesions as compared with those with left hemisphere lesions could have led to a compensatory upregulation of receptors that might protect against depression. Also, patients with left hemisphere lesions may have had moderate depletions of biogenic amines but without a compensatory upregulation of serotonin receptors, therefore having a dysfunction of biogenic amine systems in the left hemisphere. This dysfunction ultimately may have led to the clinical manifestations of depression.

Mania A study of 17 patients with secondary mania (mood disorder due to stroke, with manic features) found that most had right hemisphere lesions. The frequency of right hemisphere lesions was significantly higher among 28 patients with major depressive disorder, who tended to have left frontal or basal ganglia lesions (Fig. 2.2-6). These findings were replicated in other patients with secondary mania. Lesions associated with mania were either cortical (basotemporal cortex or orbitofrontal cortex) or subcortical (frontal white matter, basal ganglia, or thalamus). A PET study with fluorine-18 deoxyglucose (FDG) showed a focal hypometabolic deficiency in the right basotemporal cortex in 3 patients with right subcortical lesions.

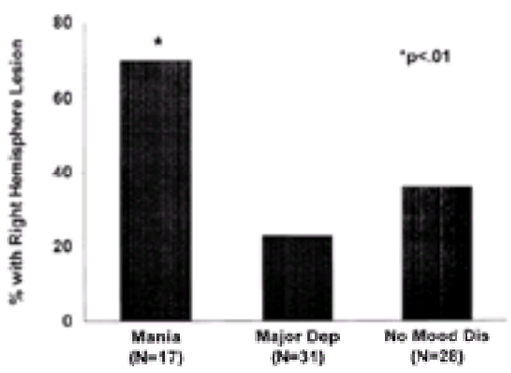

FIGURE 2.2-6 The percent of patients with mania, major depressive disorder, or no mood disorder following brain injury divided by lesion location as visualized on CT scan. Mania was strongly associated with right hemisphere lesions and major depressive disorder with left hemisphere injury. (Data from Robinson RG, Boston JD, Starkstein SE, Price TR: Comparison of mania with depression following brain injury: Causal factors. Am J Psychiatry *145:*172, 1988.)

Thus, right hemisphere lesions that lead to mania appear to be in specific right hemisphere structures that have connections to the limbic system. The right basotemporal cortex may be particularly important because direct lesions as well as distant hypometabolic effects (diaschisis) of this cortical region were frequently associated with secondary mania.

Not every patient with a lesion in limbic areas of the right hemisphere develops secondary mania, which suggests that risk factors must exist for this disorder. One study found that patients with secondary mania had a significantly higher frequency of positive family history of mood disorders than did depressed patients or patients with no mood disturbance. Another study compared patients with secondary mania to patients with no mood disturbance who were matched for size, location, and etiology of brain lesion. Patients with secondary mania had significantly greater degree of subcortical atrophy than no mood disturbance patients, as measured by bifrontal and third ventricular to brain ratio. Of the patients who developed secondary mania, those who had a positive family history of psychiatric disorders had significantly less atrophy than those without such a family history, suggesting that genetic predisposition to mood disorders and brain atrophy may be independent risk factors for poststroke mania.

Although the mechanism of secondary mania remains unknown, lesion studies as well as metabolic studies have suggested that the right basotemporal cortex may play an important role. A combination of biogenic amine system dysfunction and release of tonic inhibitory input into the basotemporal cortex and lateral limbic system may lead to mania.

Poststroke Psychosis Information about the mechanism of poststroke psychosis (psychotic disorder due to stroke) is derived from anecdotal or small case series. One study of 5 patients with psychosis following stroke found that all patients had right hemisphere lesions, primarily involving frontoparietal regions. When compared with 5 patients matched for age, education, and lesion size and location but no psychosis, patients with secondary psychosis had significantly greater subcortical atrophy, as manifested by larger areas of both the frontal horn of the lateral ventricle and the body of the lateral ventricle (measured on the side contralateral to the brain lesion). Several investigators have also reported a high frequency of seizures among patients with secondary psychosis. These seizures usually started after the brain lesion but before the onset of psychosis. Three of 5 control subjects matched for psychosis and lesion were found to have seizure disorder, as compared to 0 of 5 poststroke nonpsychiatric control subjects.

It has been hypothesized that three factors may be important in the mechanism of organic hallucinations, namely a right hemisphere lesion involving the temporoparietal cortex, the presence of seizures, and subcortical brain atrophy.

Apathy A study of 80 patients with single stroke lesions found that patients with apathy showed a significantly higher frequency of lesions involving the posterior limb of the internal capsule as compared to patients with no apathy. Lesions in the internal globus pallidus and the posterior limb of the internal capsule have been

reported to produce behavioral changes, such as motor neglect, psychic akinesia, and akinetic mutism. The ansa lenticularis is one of the main internal pallidal outputs, and it ends in the pedunculopontine nucleus after going through the posterior limb of the internal capsule. In rodents, this pathway has a prominent role in goal-oriented behavior and dysfunction of this system may explain the presence of apathy in patients with lesions of the posterior limb of the internal capsule.

Catastrophic Reaction In a study of 62 patients with acute stroke, those demonstrating catastrophic reactions had a significantly higher frequency of lesions involving the basal ganglia compared to acute stroke controls. When 10 depressed patients with a catastrophic reaction were compared to 10 depressed patients without a catastrophic reaction, the group with catastrophic reaction had significantly more anterior lesions, mostly located primarily in subcortical regions (i.e., 8 of 9 depressed patients with catastrophic reaction had subcortical lesions; 3 of 9 depressed patients without catastrophic reaction had subcortical lesions).

Based on these findings, the catastrophic reaction may result from neurophysiological dysfunction rather than realization of intellectual impairment. Catastrophic reaction occurred predominantly in patients with major depressive disorder associated with anterior subcortical lesions. Subcortical damage has also been hypothesized to underlie the “release” of emotional display by removing inhibitory input to limbic areas of the cortex.

Pathological Emotions Pathological emotions have classically been explained as secondary to the bilateral interruption of descending neocortical upper motor neuron innervation of bulbar motor nuclei. Some patients with pathological emotions have bilateral lesions and pseudobulbar palsy but others do not. One study found that patients with frontal or temporal lesions in either hemisphere had a significantly increased frequency of pathological emotions. Examination of lesion size and location in 12 patients with pathological crying found that the patients with the most frequent crying episodes had relatively large bilateral pontine lesions; the intermediate group had large bilateral lesions. The least affected patients had relatively large unilateral subcortical lesions. It was hypothesized that pathological emotions may arise from partial destruction of raphe serotonergic neurons or their projections.

DIAGNOSIS AND CLINICAL FEATURES

Vascular Dementia Dementia is a syndrome that includes deterioration of intellectual ability as well as alterations in the patient's emotional and personality functions. Multi-infarct dementia is characterized by an abrupt onset, by stepwise deterioration of intellectual function, and gradual accumulation of neuropsychological deficits in which some cognitive functions are more impaired than others; it results from ischemic injury in multiple brain regions. To make the diagnosis, these deficits must not be limited to a period of depression or delirium and must be of sufficient degree to impair work, usual social activities, or interpersonal relations. Finally, there must be evidence of organic factors related to these disturbances.

For the diagnosis of vascular dementia using DSM-IV criteria, cognitive decline should be demonstrated by loss of memory and at least one other deficit, including aphasia, apraxia, agnosia, or deterioration of executive function. Multifocal deficits are expected, and single defects in cognition, such as amnestic states, aphasia, and apraxias, do not fulfill the criteria. Single lesions may produce vascular dementia if they lead to loss of memory as well as some other cognitive function, deficits that are severe enough to produce impairment in daily living.

Diagnosing vascular dementia can be difficult. A particularly difficult differential diagnosis occurs in patients with cerebral infarction but without prominent (focal) neurological symptoms. This is frequently the case with thalamic or parietal lobe infarctions that present with confusion and drowsiness as the only clinical manifestations. Diagnostic problems may also occur when nonvascular diseases, intracranial neoplasms, or subdural hematoma present with focal neurological deficits of acute onset.

Poststroke Depression A comparison of depressive symptoms among patients with acute stroke with and without depression was undertaken to determine whether depression or acute physical illness was the cause of depressive symptoms. A consecutive series of 85 stroke patients who acknowledged the presence of a depressed mood (no other symptom was required) were compared with 120 stroke patients without a depressed mood. The study found that, except for early morning awakening, all the affective and autonomic symptoms of depression were significantly more frequent among patients with a depressed mood than among patients without a depressed mood ($p < .01$; (Fig. 2.2-7). Moreover, the presence of nonspecific symptoms of depression (i.e., the frequency of depressive symptoms in the nondepressed group) may have led to false-positive diagnoses in only 3 percent of patients, and only 5 percent of patients had all the symptoms necessary for a diagnosis of major depressive disorder except for feelings of sadness (i.e., possible false-negative cases). Therefore, the use of DSM-IV criteria in an acutely medically ill population does not appear to produce significant numbers of false-positive or false-negative cases.

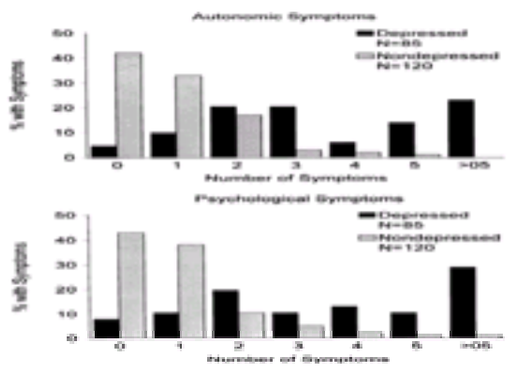

FIGURE 2.2-7 Autonomic and psychological symptoms of depression among patients with depressed mood after stroke. (Reprinted with permission from Fedoroff JP, Starkstein SE, Parikh RM, Price TR, Robinson RG: Are depressive symptoms non-specific in patients with acute stroke? Am J Psychiatry *148:*1172, 1991.)

Mania The symptoms of mania examined were in a series of 12 consecutive patients who met third edition of DSM (DSM-III) criteria for an organic affective syndrome, manic type. These patients, who developed mania after a stroke, traumatic brain injury, or tumors, were compared to patients with primary mania (i.e., no known neuropathology). Both groups of patients showed similar frequencies of elation, pressured speech, flight of ideas, grandiose thoughts, insomnia, hallucinations, and paranoid delusions. Thus, the symptoms of mania that occurred after brain damage (secondary mania) appeared to be the same as those found in mania without brain damage (primary mania).

Anxiety The diagnosis of generalized anxiety disorder based on DSM-IV criteria is termed “anxiety disorder due to stroke with generalized anxiety.” It requires the presence of anxiety and worry the majority of the time for 6 or more months with three or more of the following symptoms: restlessness or being keyed up, early fatigue, difficulty concentrating, irritability, muscle tension, or sleep disturbance. Generalized anxiety disorder was associated with a prior history of alcohol abuse significantly more frequently among patients with acute stroke lesions than among depressed or control patients. A subsequent in-hospital study found that patients with generalized anxiety disorder and major depressive disorder were significantly more impaired in their activities of daily living and social functioning at 1 to 2 years follow-up than patients with major depressive disorder alone. Hospitalized patients with generalized anxiety disorder, however, were not more impaired in their activities of daily living or social functioning than patients without this disorder. These findings suggest that impairment does not cause the disorder but the disorder impacts on physical and social recovery from stroke.

Apathy In a study of 9 patients with poststroke apathy compared to 71 stroke patients without apathy, the only demographic correlate of apathy was age. Apathetic patients (with or without depression) were significantly older than nonapathetic patients. Also, apathetic patients showed significantly more severe deficits in activities of daily living (ADL) and there was a significant interaction between depression and apathy on scale scores, with the greatest impairment found in patients who were both apathetic and depressed.

Catastrophic Reaction Catastrophic reactions occurred in 12 of 62 patients (19 percent) admitted to the hospital with stroke. Patients with catastrophic reactions were found to have a significantly higher frequency of familial and personal history of psychiatric disorders (mostly depressive disorder) than patients without catastrophic reactions. Catastrophic reactions, however, were not significantly more frequent among aphasic compared with nonaphasic patients. This finding did not support the contention that catastrophic reactions represent an understandable psychological response of “frustrated” aphasic patients. Furthermore, 9 of the 12 patients with catastrophic reaction also had after their strokes major depressive disorder, 2 had minor depressive disorder and only 1 was not depressed. Thus,

catastrophic reaction was significantly associated with major depressive disorder.

Pathological Emotions A study of the clinical correlates and treatment of emotional lability (including pathological laughter and crying) was conducted in 28 patients with either acute or chronic stroke. A Pathological Laughter and Crying Scale (PLACS) was developed to assess the existence and severity of emotional lability. There are no generally accepted criteria for the diagnosis of pathological emotions. Patients with this condition, however, acknowledged an inability to control their crying or laughter, an increased frequency of emotional display, and a recognition that the emotional display was inconsistent with or excessive in terms of their underlying emotional feelings.

PATHOLOGY AND LABORATORY EXAMINATION

Vascular Dementia The clinical identification of vascular dementia requires a medical history, neurological examination, psychiatric interview, and psychometric testing using a dementia rating scale. Structural imaging studies using computed tomography (CT) or magnetic resonance imaging (MRI) scans should document the existence of one or more cerebral infarctions (Fig. 2.2-2). Laboratory data that can be helpful are blood chemistries (including B_{12}, folate, thyroid function), cerebrospinal fluid analysis, an electroencephalogram (EEG), and an EEG with evoked responses, CT, MRI, and, in certain cases, cerebral angiography. Data from these laboratory studies will usually identify cases of potentially treatable forms of dementias caused by tumor, vascular malformation, cerebral hematoma, normal-pressure hydrocephalus, infections, and metabolic, toxic, and drug-induced encephalopathy, as well as dementia due to vitamin or endocrine deficiencies.

Poststroke Depression The dexamethasone (Decadron) suppression test (DST) has been investigated as a possible biological marker for major depressive disorder with melancholic features. Several studies have demonstrated a statistical association between major poststroke depression (mood disorder due to stroke with major depressive-like feature) and failure to suppress serum cortisol in response to administration of dexamethasone. The specificity of the test, however, has generally been insufficient to allow it to be diagnostically useful. In a study of 65 patients whose acute strokes had occurred within the preceding year, 67 percent of the patients with major poststroke depression failed to suppress serum cortisol, compared to 25 percent of patients with minor depressive disorder, and 32 percent of nondepressed patients. The sensitivity of the DST for major depression was 67 percent but the specificity was only 70 percent. False-positive tests, found in 30 percent of patients, seemed to be related to large lesion volumes.

A study of growth hormone response to desipramine found that growth hormone responses were significantly blunted in patients with poststroke depression, suggesting that diminished a_2 adrenergic receptor function may be an important marker for poststroke depression. The sensitivity of the test was 100 percent and the specificity was 75 percent. Future studies may further examine the validity of the endocrine response as a marker of poststroke depression.

Other Disorders The utility of laboratory examinations in the diagnosis or prognosis of mania, anxiety disorder, psychosis, apathy, catastrophic reactions, or anosognosia have not been established except as discussed in the section on etiology.

COURSE AND PROGNOSIS

Vascular Dementia The course of vascular dementia is characterized by current stroke with associated deterioration of cognitive function. The probability of recurrent stroke is about 7 percent per year. The course and prognosis, however, can be influenced by prevention. A longitudinal study of 173 patients examined the frequency of risk factors for stroke and cerebral atherosclerosis among patients with vascular dementia. Although hypertension was the single most potent risk factor for cerebral atherosclerosis and stroke, hypotension, present in 66 percent of cases, was by far the most common risk factor for vascular dementia in this sample. Heart disease of the atherosclerotic type, with or without cardiac arrhythmia, was also represented in the majority of cases of vascular dementia. Cardiac disease may provide a source of cerebral emboli leading to vascular dementia. Smoking of one or more packs of cigarettes per day was a risk factor in 21 percent of the patients. Hyperlipidemia of the type 4 form (hypertriglyceridemia) was present in 29 percent of cases. Diabetes mellitus of sufficient clinical severity to require medical treatment was found in 20 percent of the cases, and symptomatic peripheral vascular disease with ischemic symptoms referable to the lower extremities was present in 6 percent of the cases. The association with limited education suggested that education-induced factors (perhaps neurobiological or social factors) place patients at greater risk for developing dementia from a vascular cause. Thus, the course and prognosis of vascular dementia is dependent upon factors that place patients at risk for vascular dementia.

Depression In a 2-year longitudinal study, a consecutive series of 103 acute stroke patients were examined for depression at 3, 6, 12, and 24 months follow-up. At the time of the initial in-hospital evaluation, 26 percent of the patients had the symptom cluster of major depressive disorder whereas 20 percent had the symptom cluster of minor depressive disorder (Fig. 2.2-8). Although both major and minor depressive disorders were still present in 86 percent of patients at 6 months follow-up, only 1 of 5 patients with in-hospital major depression continued to have the condition at 1 year follow-up (2 cases had minor depression). Patients with minor depression, however, had a less favorable prognosis, with only 40 percent having no depression at 1 year and 30 percent having no depression at 2 years follow-up. In addition, about 30 percent of patients who were not depressed in the hospital became depressed after discharge. Thus the natural course of major depression appeared to be between 6 months and 1 year whereas the duration of minor depression was more variable and in many cases the patients appeared to be chronically depressed.

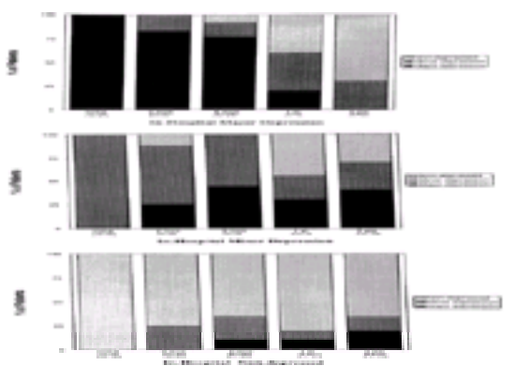

FIGURE 2.2-8 Diagnostic outcome in patients with acute stroke. The upper panel shows the diagnoses at 3, 6, 12, and 24 months follow-up for patients with an in-hospital diagnosis of major depressive disorder. The middle and lower panels show similar data for patients with in-hospital minor depressive disorder or no depression. Patients with in-hospital minor depressive disorder had a greater frequency of depressive disorders at 2 years than patients with major depressive disorder or nondepressed patients. (Data taken from Robinson RG, Bolduc P, Price TR: A two-year longitudinal study of poststroke depression: Diagnosis and outcome at one- and two-year follow-up. Stroke *18:*837, 1987.)

Other studies have also found that major depression had a duration of less than 1 year in the majority of cases, but a minority may develop chronic depression. A 10-year follow-up of patients with major or minor depression in the hospital found that depressed patients had a fourfold higher likelihood of dying than patients without depression. This finding held up even when other factors related to prognosis such as concurrent medical illness were controlled.

The nature of poststroke mood disorders is exemplified by the following case history.

CENTER>

A 52-year-old married man suffered a heart attack while playing basketball. Two nights later he had a thromboembolism that produced a large right middle cerebral artery infarct with left hemiparesis and sensory deficit. Within a couple of months he developed depression that lasted for several months. After a course of antidepressant medication, he became manic with grandiose ideas, pressured speech, decreased sleep, and increased energy. Those symptoms were effectively treated with lithium (Eskalith), but depression ensued, with prominent suicidal thoughts and delusional beliefs. This was treated effectively with electroconvulsive therapy and the patient remained euthymic on lithium and nortriptyline (Pamelor).

Mania The course of mania following stroke has not been systematically examined. Anecdotal cases have been reported that recurrent episodes of mania or

depression may occur in these patients. The response to medication and the course of disorder for individual patients with single or recurrent episodes of mania are shown in Figure 2.2-9.

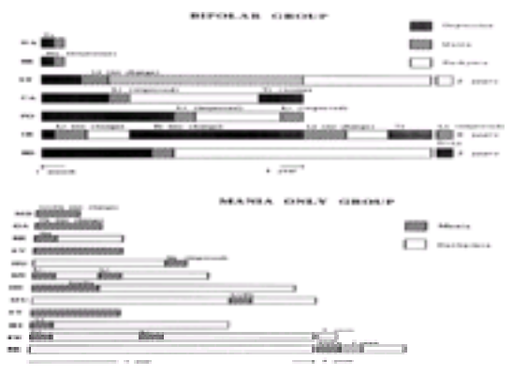

FIGURE 2.2-9 A, Longitudinal evolution of mood disorder for individual patients with bipolar I disorder. **B**, Longitudinal evolution of mood disorder for individual patients with unipolar mania (mania without depression). Tr, tricyclic antidepressant; Ne, nortryptyline; Li, lithium; Ca, carbamazepine. (Reprinted with permission from Robinson RG, Starkstein SE: Neuropsychiatric aspects of cerebrovascular disorders. In *Textbook of Neuropsychiatry*, ed 3, SC Yudofsky, RE Hales, editors. American Psychiatric Press, Washington, DC, 1997.)

Anxiety A 2-year follow-up of 142 patients with acute stroke found that 23 percent of patients developed anxiety disorder due to stroke with generalized anxiety after the initial in-hospital evaluation (i.e., between 3 and 24 months poststroke). Early-onset but not late-onset was associated with prior history of psychiatric disorder, including alcohol abuse. Early-onset anxiety disorder had a mean duration of 1.5 months whereas the delayed-onset form of the disorder had a mean duration of 3 months. In addition, the existence of anxiety disorder also influenced the duration of depression. Patients with generalized anxiety and major depressive disorder had a mean duration of depression that was significantly longer than the duration of depression without anxiety disorder.

Other Disorders The course and prognosis of patients with psychosis, apathy, catastrophic reaction, pathological emotion, and anosognosia have not been systematically studied.

TREATMENT

Vascular Dementia Some of the risk factors for stroke can be effectively treated, which raises the hope that the natural progression or even pathogenesis of vascular dementia might be effectively treated. Stroke of cardioembolic origin is responsible for about 15 percent of all ischemic strokes; this percentage is even higher among younger patients. Following cardioembolic stroke, anticoagulation is an effective treatment to reduce the risk of recurrence.

Since the late 1980s, treatment with antiplatelet aggregate drugs has reduced the number of repeated ischemic vascular episodes in patients with transient ischemic attacks. The efficacy of these drugs suggests that aspirin treatment might also modify the course of vascular dementia. Aspirin combined with other antiplatelet drugs has been shown to be effective in secondary prevention of stroke. The United Kingdom-Transient Ischemic Attack (UK-TIA) Aspirin Trial, with 2435 patients using two different dosages of aspirin, found a 21.7 percent and 25.1 percent reduction (depending on dosage) in the risk of nonfatal strokes, myocardial infarction, or death compared with placebo treatment.

Other therapeutic measures that may be helpful in vascular dementia include smoking cessation, prevention or careful management of diabetes mellitus, potassium supplementation (because of its vascular protective effect), estrogen replacement (in postmenopausal women), antihypertensive medications (angiotensin-converting enzyme inhibitors and calcium channel inhibitors), lipid-lowering agents, anticoagulants (for patients in whom emboli are likely), and aspirin. For patients who are in a “predementia” stage (i.e., those who have a history of transient ischemic attacks, stroke, previous cognitive impairment, or silent cerebral infarctions, but without global cognitive impairment), prevention may include carotid endarterectomy (when carotid stenosis is from 70 to and 99 percent), anticoagulants, aspirin, ticlopidine (Ticlid), agents that interfere with amyloid depositions in vessels (in the case of stroke secondary to lupus), and calcium channel inhibitors (pretreatment to attenuate the effect of infarcts). Finally, for patients in the dementia stage (i.e., patients with atherosclerosis of the extracranial arteries, cardiac embolism, and intracranial small vessel disease who have already shown evidence of cognitive decline in several areas of intellectual functioning), treatment measures may include antidepressant drugs, antihypertensive agents, cholinergic agonists, nerve growth factor, aspirin, or ticlopidine.

Depression Several randomized double-blind treatment studies have been published on the efficacy of antidepressant treatment of poststroke depression. The first study examined 14 patients treated with nortriptyline and 20 patients given a placebo. The 11 patients treated with nortriptyline who completed the 6-week study showed significantly greater improvement in their scores on the Hamilton Rating Scale for Depression (HAM-D) than did the 15 placebo-treated patients ($p < .01$; Fig. 2.2-10). Successfully treated patients had serum nortriptyline levels between 50 and 150 ng/mL. Three patients experienced adverse effects (including delirium, confusion, drowsiness, and agitation) that were severe enough to require the discontinuation of nortriptyline. A recent double-blind controlled trial using the selective serotonin reuptake inhibitor (SSRI) citalopram (Celexa) found that HAM-D scores were significantly more improved over 6 weeks in patients receiving active ($N = 27$) compared with placebo ($N = 32$) treatment. At both 3 and 6 weeks, the active group had significantly lower HAM-D than the placebo group. This study was the first to establish the efficacy of SSRIs in the treatment of poststroke depression.

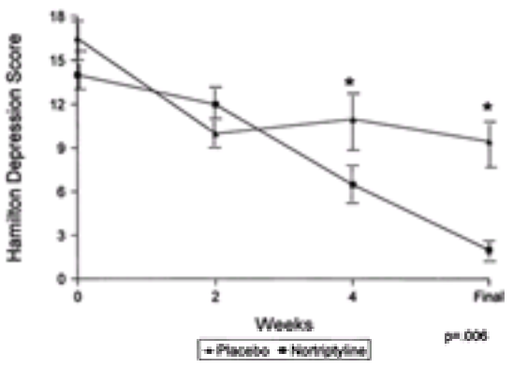

FIGURE 2.2-10 Hamilton depression scores during 6 weeks of double-blind treatment of poststroke depression with nortriptyline or placebo. Repeated measures analysis showed significant treatment effects with nortriptyline being superior to placebo at 4 and 6 weeks of treatment. (Reprinted with permission from Lipsey JR, Robinson RG, Pearlson GD, Rao K, Price TR: Nortriptyline treatment of poststroke depression: A double-blind treatment trial. Lancet *i*:297, 1984.)

Electroconvulsive therapy has also been reported to be effective for treating poststroke depression; it causes few adverse effects and no neurological deterioration. Finally, psychological treatment, including group and family therapy, has also been reported to be useful. However, controlled studies have not been conducted for these treatment modalities.

Anxiety Disorders There have been no systematic studies of the treatment of anxiety disorders following stroke. For guidance, the clinician must look to studies performed on patients with anxiety disorders not associated with brain injury. Benzodiazepines are the most commonly used medications in generalized anxiety disorders. Short-acting benzodiazepines are a prudent choice because longer-acting agents may place the patient at a greater risk for adverse effects, such as sedation, ataxias, disinhibition, and confusion. As with tricyclic drugs, very conservative dosage and careful monitoring must be employed.

In light of the comorbidity of anxiety and depression, antidepressants (tricyclics or SSRIs) may also be of benefit for the poststroke patient with generalized anxiety disorder. Tricyclic drugs may be efficacious in the treatment of generalized anxiety disorder even in the absence of depression; however, the specific needs of the poststroke patient have not been investigated. In one multicenter trial that excluded patients with comorbid depression, both chlordiazepoxide (Librium) and imipramine (Tofranil) were found to be superior to placebo in treatment of this disorder over an 8-week period. Finally, buspirone (BuSpar) may be useful in reducing anxiety without many of the adverse effects such as sedation and without the risk of development of tolerance.

Psychosis There are no controlled treatment trials among patients with delusions or hallucinations following stroke. Anecdotal reports have suggested two basic approaches to treatment, one utilizing anticonvulsant medications and the other antipsychotic agents. The use of anticonvulsants has its rationale in the frequent coexistence of seizures with psychotic disorders following stroke.

Apathy There have been no controlled trials of treatment of apathy following stroke. Patients have been treated with nortriptyline, apomorphine, and amphetamines with some success. Since this is a relatively common consequence of stroke, treatment trials are urgently needed to address a problem that can be devastating to the resumption of physical and social activities following stroke.

Pathological Emotions The treatment of pathological laughter and crying in patients with stroke has been assessed in two double-blind placebo-controlled trials. Using a standardized Pathological Laughter and Crying Scale (PLACS), a double-blind drug trial of nortriptyline versus placebo was conducted. The dosage of nortriptyline was titrated from 20 mg on week 1, to 50 mg on weeks 2 and 3, to 70 mg on week 4, to 100 mg on weeks 5 and 6. Twenty-eight patients completed the 6-week protocol (4 patients dropped out). Patients on nortriptyline showed significantly greater improvement in PLACS scores compared with placebo-treated controls (Fig. 2.2-11). These group differences were statistically significant at weeks 4 and 6. Although a significant improvement in depression scores was also observed, improvements in PLACS scores were significant for both depressed and nondepressed patients who displayed pathological emotion. This indicates that treatment response was not simply related to treatment of depression.

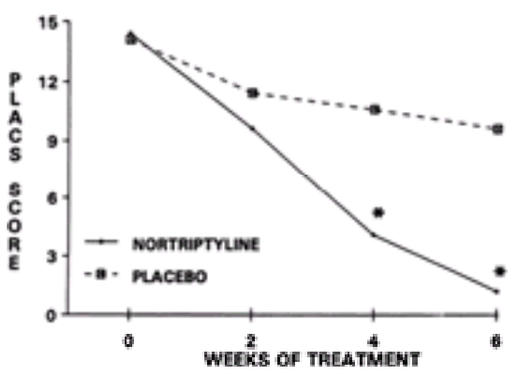

FIGURE 2.2-11 Pathological laughter and crying scale (PLACS) scores during 6 weeks of double-blind treatment of poststroke pathological emotions with nortriptyline or placebo. Repeated measures analysis showed significant treatment effects with nortriptyline being superior to placebo at 4 and 6 weeks of treatment. (Reprinted with permission from Robinson RG, Parikh RM, Lipsey JR, Storkstein SE, Price TR: Pathological laughing and crying following stroke: Validation of measurement scale and double-blind treatment study. Am J Psychiatry *150:* 286, 1993.)

Citalopram has also been evaluated in the treatment of pathological emotion following stroke. In a double-blind placebo-controlled crossover study, 16 patients were evaluated and treatment was given for 3 weeks after a week of washout. All of the citalopram-treated patients reported a greater than 50 percent reduction in the number of crying episodes. Eight patients responded within 24 hours of taking citalopram (20 mg), 3 patients responded within 3 days, and only 4 patients took more than a week to respond. None of the patients had major depressive disorder, but HAM-D scores dropped significantly during treatment with citalopram.

Thus, citalopram as well as nortriptyline appears to be an effective method of treatment for pathological crying following stroke. In addition, poststroke depression and pathological laughing and crying appear to be independent phenomena, although they may coexist. Both depression and pathological laughing and crying, however, are amenable to treatment.

A 64-year-old right-handed, married woman with no history of stroke suffered a thrombotic right hemisphere stroke with a hemiparesis but no sensory deficit. Beginning within a few days after the stroke, the patient had uncontrollable crying episodes that occurred 5 to 10 times per day and lasted for about 1 to 2 minutes. She and her husband were retired and had an active social life. In addition to the crying episodes, the patient had major depressive disorders with a HAM-D score of 19. She stated that she felt sad but the crying was greatly in excess of her sadness at the time and she had no sense of being able to control the crying. Her PLACS score was 24, which was severe. She showed no improvement over 6 weeks of treatment with placebo but improved greatly after a course of nortriptyline. The pathological emotions were more troublesome to her than the depression. She stopped seeing any friends or even leaving the house for fear of being embarrassed socially by these crying episodes.

Other Disorders Effective treatments have not been established for catastrophic reactions or anosognosia.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2, neuroimaging in Section 1.15 and Section 1.16, schizophrenia in Chapter 12, mood disorders in Chapter 14, anxiety disorders in Chapter 15, and Alzheimer's disease and other dementing disorders in Section 51.3e.

SECTION REFERENCES

Alexopoulos GS, Meyers BS, Young RC, Campbell S, Silbersweig D, Charlson M: 'Vascular depression' hypothesis. Arch Gen Psychiatry *54:*915, 1997.

Andersen G, Vestergaard K, Lauritzen L: Effective treatment of poststroke depression with the selective serotonin reuptake inhibitor citalopram. Stroke *25:*1099, 1994.

Astrom M, Adolfsson R, Asplund K: Major depression in stroke patients: A 3-year longitudinal study. Stroke *24:*976, 1993.

Astrom M: Generalized anxiety disorder in stroke patients: A 3-year longitudinal study. Stroke *27:*270, 1996.

Babinski J: Contribution a l'etude des troubles mentaux dans l'hemiplegie organique cerebrale (anosognosie). Rev Neurol (Paris) *27:*845, 1914.

Bleuler EP: *Textbook of Psychiatry.* Macmillan, New York, 1951.

Burvill PW, Johnson GA, Jamrozik KD, Anderson CS, Stewart-Wynne EG, Chakera TMH: Prevalence of depression after stroke: The Perth Community Stroke Study. Br J Psychiatry *166:*320, 1995.

Castillo CS, Starkstein SE, Fedoroff JP, Price TR, Robinson RG: Generalized anxiety disorder following stroke. J Nerv Ment Dis *181:*100, 1993.

Denny-Brown D, Meyer JS, Horenstein S: The significance of perceptual rivalry resulting from parietal lesions. Brain *75:*434, 1952.

Gainotti G, Azzoni A, Gasparini F, Marra C, Razzano C: Relation of lesion location to verbal and nonverbal mood measures in stroke patients. Stroke *28:*2145, 1997.

*Ghika-Schmid F, van Melle G, Guex P, Bogousslavsky J: Subjective experience and behavior in acute stroke: The Lausanne Emotion in Acute Stroke Study. Neurology *52:*22, 1999.

Goldstein K: *The Organism: A Holistic Approach to Biology Derived from Pathological Data in Man.* American Books, New York, 1939.

Hermann M, Walesch C-W: Depressive changes in stroke patients. Disability and Rehabilitation *15:*55, 1993.

*Herrmann N, Black SE, Lawrence J, Szekely C, Szalai JP: The Sunnybrook Shoke Study, a prospective study of depressive symptoms and functional outcome. Stroke *29:*618, 1998.

House A: Depression associated with stroke. J Neuropsychiatry *8:*453, 1996.

House A, Dennis M, Mogridge L, Warlow C, Hawton K, Jones L: Mood disorders in the year after stroke. Br J Psychiatry *158:*83, 1991.

Katzman R, Lasker B, Bernstein N: Advances in the diagnosis of dementia: Accuracy of diagnosis and consequences of misdiagnosis of disorders causing dementia. In *Aging and the Brain*, RD Terry, editor. Raven, New York, 1988.

Kotila M, Numminen H, Waltimo O, Kaste M. Depression after stroke: Results of the FINNSTROKE Study. Stroke *29:*368, 1998.

Kraepelin E: *Manic Depressive Insanity and Paranoia.* E & S Livingstone, Edinburgh, 1921.

Lipsey JR, Robinson RG, Pearlson GD, Rao K, Price TR: Nortriptyline treatment of post-stroke depression: A double-blind treatment trial. Lancet *1:*297, 1984.

Meyer A: The anatomical facts and clinical varieties of traumatic insanity. Am J Insanity *60:*373, 1904.

*Paradiso S, Robinson RG: Gender differences in poststroke depression. J Neuropsychiat Clin Neurosci *10:*41, 1998.

*Plate KH: Mechanism of angiogenesis in the brain. J Neuropathol Exp Neurol *58:*313, 1999.

Price TR: Pathological laughing and crying following stroke: Validation of measurement scale and double-blind treatment study. Am J Psychiatry *150:*286, 1993.

Robinson RG: *The Clinical Neuropsychiatry of Stroke.* Cambridge University Press, Cambridge, United Kingdom, 1998.

*Robinson RG, Paradiso S: Depression, psychosis, and agitation in stroke. In *Primer on Cerebrovascular Diseases*, KMA Welch, LR Caplan, DJ Reis, BK Siego, B Weir, editors. Academic Press, New York, 1997.

Robinson RG, Schultz SK, Paradiso S: Treatment of poststroke psychiatric disorders. In *Geriatric Psychopharmacology*, JC Nelson, editor. Marcel Dekker, New York, 1998.

Robinson RG, Starkstein SE: Neuropsychiatric aspects of cerebrovascular disorders. In *Textbook of Neuropsychiatry*, ed 3, SC Yudofsky, RE Hales, editor. American Psychiatric Press, Washington, DC, 1997.

Schultz SK, Castillo CS, Kosier JT, Robinson RG: Generalized anxiety and depression. Am J Geriatr Psychiatry *5:*229, 1997.

Shimoda K, Robinson RG: Effect of anxiety disorder on impairment and recovery from stroke. J Neuropsychiat Clin Neurosci *10:*34, 1998.

*Starkstein SE, Fedoroff JP, Price TR, Leiguarda R, Robinson RG: Anosognosia in patients with cerebrovascular lesions: A study of causative factors. Stroke *23:*1446, 1992.

Starkstein SE, Fedoroff JP, Price TR, Leiguarda R, Robinson RG: Apathy following cerebrovascular lesions. Stroke *24:*1625, 1993.

Starkstein SE, Fedoroff JP, Price TR, Leiguarda R, Robinson RG: Catastrophic reaction after cerebrovascular lesions: Frequency, correlates, and validation of a scale. J Neurol Neurosurg Psychiatry *5:*189, 1993.

UK-TIA Study Group: United Kingdom transient ischaemic attack (UK-TIA) aspirin trial: Interim results. Br Med J *296:*316, 1988.

*Vigliani MC, Duyckaerts C, Hauw JJ, Poisson M, Magdelenat H, Delattre JY: Dementia following treatment of brain tumors with radiotherapy administered alone or in combination with nitrosourea-based chemotherapy: A clinical and pathological study. Neurooncology *41:*137, 1999.

Welt L: Ueber charakterveranderungen des menschen infolge von lasionen des stirnhirms. Dtsch Arch Klin Med *42:*339, 1888.

2.3 NEUROPSYCHIATRIC ASPECTS OF BRAIN TUMORS

FACUNDO F. MANES, M.D., AND ROBERT G. ROBINSON, M.D.

Brain tumors can be divided into primary tumors (those arising from the brain and its coverings) and metastatic tumors (those arising elsewhere in the body). Table 2.3-1 lists the most clinically important tumors. In children, astrocytomas and medulloblastomas are the most common tumors; in adults, the most common are metastatic tumors, astroglial neoplasms (including glioblastoma multiforme), meningiomas, and pituitary adenomas.

TABLE cellSpacing=0 cellPadding=0 align=left border=0 hspace="10" vspace="5">

Histologically benign
Meningioma
Pituitary adenomas
Acoustic neuromas
Craniopharyngiomas
Pilocytic astrocytomas
Epidermoid tumors
Colloid cysts
Choroid plexus papillomas
Hemangioblastomas
Histologically malignant
Gliomas
Astroglial neoplasms
Oligodendrogliomas
Gangliogliomas
Ependymomas
Lymphomas
Medulloblastomas
Primitive neuroectodermal tumors
Germ cell tumors
Pineal cell tumors
Chordomas
Choroid plexus carcinomas
Metastatic brain tumors
Single or multiple metastases
Meningeal carcinomatosis

Table 2.3-1 Most Frequent Brain Tumors

Patients with brain neoplasms characteristically have focal neurological disturbances. However, primary and metastatic intracranial tumors produce direct and indirect effects on brain functions that could result in behavioral alterations. Before neuroimaging and modern neurosurgical techniques became available it was well recognized that some patients with tumors would first present with neurobehavioral or psychiatric manifestations. The current knowledge on neuropsychiatric aspects of brain tumors is scanty and based on clinical case reports and large, uncontrolled case studies from the older literature.

Intracranial tumors have a high incidence of mental symptoms, consisting of personality changes, emotional disturbances, and intellectual defects. One study found that the prevalence of mental symptoms in patients with tumors of the temporal lobes was 94 percent, with neoplasms of the frontal lobes it was 90 percent, and with infratentorial tumors it was 47 percent. Another study found neuropsychiatric symptoms in 51 percent of patients with brain tumors, and yet another showed that in 18 percent of patients, behavioral changes were the first clinical manifestation of a brain tumor. The significant advances that have occurred in the diagnosis and surgical management of brain tumors, however, have decreased the frequency of mental symptoms presenting as the first manifestation of a brain tumor.

The frequency of intracranial tumors at autopsies conducted in psychiatry hospitals is reported to be about 3.5 percent. Based on two studies, the frequency of intracranial tumors in general hospital autopsies is between 3.7 and 5 percent, similar to that in chronic psychiatric hospitals.

PRINCIPAL INTRACRANIAL TUMORS

Gliomas Gliomas include astroglial neoplasms (which are thought to have astrocytic precursors), oligodendrogliomas arising from oligodendrocytes, and ependymomas apparently derived from ependymal cells. Mixed types also occur. The glioblastoma multiforme is a primitive and highly malignant form of glioma whose cell of origin is uncertain. Gangliomas are tumors with both neoplastic glia and neoplastic neurons. The pathological distinction between subtypes of gliomas is usually made on descriptive grounds. These distinctions are helped by the presence of glial fibrillary acidic protein in astroglial tumors, myelin basic protein in oligodendrogliomas, and ciliary bodies in ependymomas.

Glioblastoma Multiforme and Anaplastic Astrocytoma (Malignant Astrocytoma) Glioblastoma multiforme and anaplastic astrocytoma account for about 20 percent of all intracranial tumors, about 55 percent of all tumors of the glioma group, and more than 90 percent of gliomas of the cerebral hemispheres in adults. Although predominantly cerebral, these tumors may be observed in the brainstem, cerebellum, or spinal cord. The peak incidence is in middle adult life, but no age group is exempt. Glioblastoma multiforme (Fig. 2.3-1) and anaplastic astrocytoma differ in age of onset and response to treatment. The mean age of patients with glioblastoma is 56 years and of those with anaplastic astrocytoma, 46 years. Postoperative survival is higher in patients with anaplastic astrocytoma.

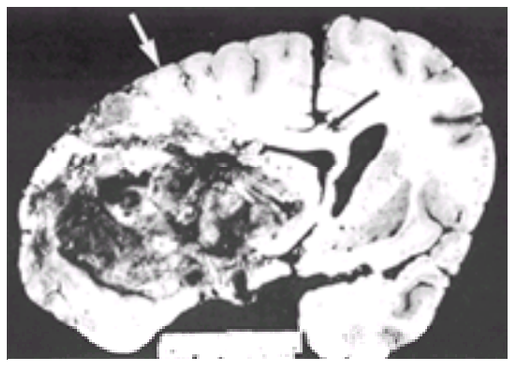

FIGURE 2.3-1 Glioblastoma multiforme. This large necrotic, hemorrhagic neoplasm has infiltrated the temporal and frontal lobes, basal ganglia, and internal capsule. Note the marked enlargement of the right hemisphere and the right cingulate gyrus herniation (*black arrow*). Just superior to the neoplasm is an area of infarction (*white arrow*) caused by compression of the middle cerebral artery by the neoplasm. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

Clinically, diffuse cerebral symptoms and seizures usually give way in a few weeks or months to a more definite frontal, temporal, parietooccipital, or callosal syndrome. The treatment of these tumors is complex and includes surgery, radiation, chemotherapy, and immunotherapy.

Astrocytoma (Low-Grade Astroglial Neoplasms) Grades 1 and 2 astrocytomas (Fig. 2.3-2), which constitute between 25 and 30 percent of cerebral gliomas, may

occur anywhere in the brain or spinal cord. In children the tumors are usually in the cerebellum. By far the most common presenting symptom is an epileptic seizure. Excision of part of the cerebral astrocytoma may allow survival in a functional state for many years. The role of radiation in patients with low-grade astroglial neoplasms is controversial, and chemotherapy has no proven benefit in the treatment of patients with low-grade astrocytomas.

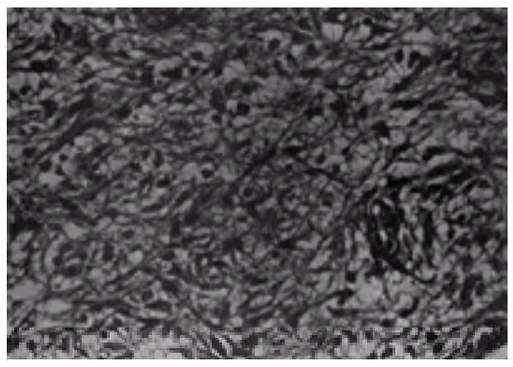

FIGURE 2.3-2 Astrocytoma. These fibrillary astrocytes form an irregular meshwork. Hematoxylin-phloxin-saffron, ×350. (From Robertson DM, Dinsdale HB: *The Nervous System.* Williams & Wilkins, Baltimore, 1972.)

Oligodendrogliomas Oligodendrogliomas (Fig. 2.3-3) represent the third most common glioma, following glioblastoma and astrocytoma. They occur most frequently within the cerebral hemispheres, often in a frontal location. Surgical resection is the most common treatment.

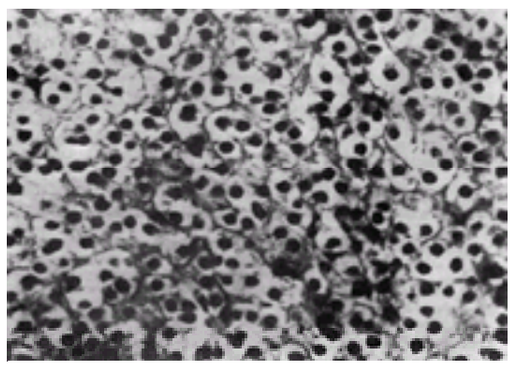

FIGURE 2.3-3 Oligodendroglioma. The prominent nuclei are surrounded by unstained cytoplasm. Hematoxylin and eosin stain (Reprinted with permission from Hirano A: *A Guide to Neuropathology.* Igaku-Shoin, New York, 1981.)

Meningiomas Meningiomas are usually benign growths that rarely invade the substance of the brain, thus presenting the potential for curative surgery. Meningiomas represent approximately 15 percent of all brain tumors. There is some evidence of a hormonal component in their growth: they may enlarge during pregnancy, they occur twice as often in women as in men, and they are more common in patients with breast carcinoma. Of all intracranial meningiomas, 85 to 90 percent are located supratentorially. Children show an increased incidence of intraventricular and posterior fossa meningiomas, as well as meningiomas without dural attachment. Radiation may provide important adjunctive therapy for surgically unresectable meningiomas (Fig. 2.3-4).

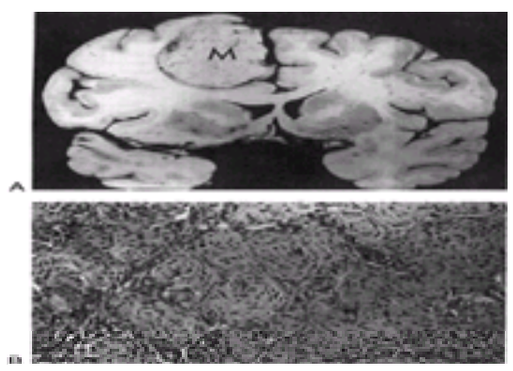

FIGURE 2.3-4 A, Parasagittal meningioma. This well-circumscribed, globoid, discrete meningioma (*M*) is attached to the dura and falx. It compresses rather than invades the underlying superior frontal gyrus. **B**, Meningioma. The tumor cells appear uniform and are arranged in whorls. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

Metastatic Brain Tumors Brain metastasis is a common complication of systemic cancer and an important cause of morbidity and mortality in cancer patients. It is the most common type of intracranial tumor, and its incidence may be rising. The frequency of metastasis to the brain varies dramatically from one type of cancer to another. Clinical presentations of metastatic disease include progressive neurological deficit, seizure, and hemorrhage. Intracerebral hemorrhage is especially likely to be the presenting event for melanomas, renal cell carcinomas, and choriocarcinomas. Surgery is the best treatment for single metastatic lesions.

Meningeal Carcinomatosis Cancer metastatic to the cerebral meninges represents approximately 5 percent of metastatic tumors; breast, lung (Fig. 2.3-5), and melanoma are the most common primary tumors. Multiple deficits of cranial nerves and spinal nerve roots are common presenting syndromes and are sometimes difficult to distinguish from paraneoplastic states.

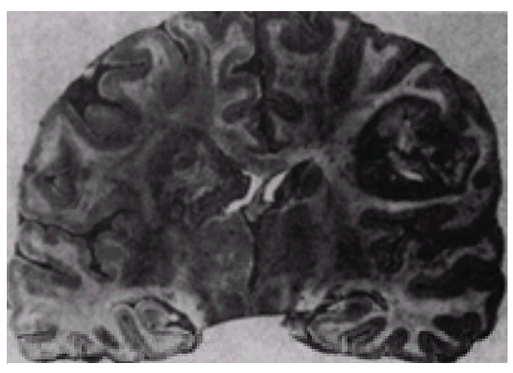

FIGURE 2.3-5 Metastatic carcinoma (from the lung). The well-circumscribed masses contain necrotic cysts. (Reprinted with permission from Russell DS, Rubenstein

LL: *Pathology of Tumors of the Nervous System*, ed 3. Williams & Wilkins, Baltimore, 1971.)

EPIDEMIOLOGY

New brain tumors develop in approximately 35,000 adult Americans each year. Recent evidence indicates that the incidence of primary tumors among elderly persons is increasing. More adults die each year of primary brain tumors than of Hodgkin's disease or multiple sclerosis.

Primary brain tumors account for about 2 percent of cancer deaths but are responsible for 7 percent of the years of life lost from cancer before the age of 70. They are responsible for 20 percent of malignant tumors diagnosed before the age of 15. About 30 percent of deaths are due to cancer in Western society, and 1 in 5 of those who die of cancer exhibit intracranial metastatic deposits at autopsy.

Among the causes of death from intracranial disease, tumor is exceeded in frequency only by stroke. Primary tumors of the brain represent 22 percent of all childhood neoplasms, second in frequency only to leukemia. In the United States, the annual incidence of all brain tumors is 46 per 100,000, and that of primary tumors is 15 per 100,000.

Given the current early diagnosis and treatment of brain tumors, formal studies of psychiatric symptoms of brain tumors are difficult to carry out, and the prevalence is difficult to estimate.

ETIOLOGY

The production of symptoms by tumor growth is governed by certain principles of physics and physiology. The cranial cavity has a restricted volume, and the three elements contained therein—the brain (about 1400 mL), cerebrospinal fluid (CSF) (140 mL), and blood (150 mL)—are relatively incompressible. According to the Monro-Kellie doctrine, the total bulk of the three elements is constant, and any increase in the volume of one of them must be at the expense of one or both of the others. Tumors growing in one part of the brain compress and destroy brain tissue and displace CSF and blood; once the limit of this accommodation is reached, the intracranial pressure rises. Elevation of the intracranial pressure and perioptic pressure impairs axonal transport in the optic nerve and venous drainage from the optic nerve head and retina, exhibited as papilledema. Once pressure is raised in a particular compartment of the brain, it begins to displace tissue; eventually there tissue at a distance from the tumor is displaced, resulting in false localizing signs.

DIAGNOSIS AND CLINICAL FEATURES

The physician should be aware of the subtlety of presentation of brain tumors in many patients and the need for efficient and expeditious diagnosis. A single seizure, loss of interest in usual activities, or loss of hearing for high frequencies may all signal a tumor, and the physician should have a high index of suspicion in these settings.

Patients with brain tumors typically present with headaches, seizures, nonspecific cognitive or personality changes, or focal neurological signs. As a result of the widespread availability of sensitive imaging techniques such as magnetic resonance imaging (MRI), tumors are being detected at an earlier stage, and patients with subtle clinical signs and symptoms are being diagnosed.

Headaches Headaches are the presenting symptom in approximately 35 percent of patients with brain tumors and develop during the course of the disease in 70 percent. Most of these headaches are intermittent and nonspecific. They are generally dull, nonthrobbing, and often indistinguishable from tension headaches. The headaches are usually on the same side as the tumor. Supratentorial tumors usually produce headaches with a frontal location, because most supratentorial pain-sensitive structures are supplied by the trigeminal nerve. The posterior fossa is innervated by cranial nerves IX and X and upper cervical nerves, which usually produce pain in the occipital region and neck. Certain features of a headache increase the possibility of a tumor. These include headaches that wake the patient at night or are worse on waking and improve over the course of the day; headaches exacerbated by postural change, coughing, or exercise; headaches of recent onset that differ from the patient's usual headaches; and the presence of nausea or vomiting, papilledema, and focal neurological signs.

Seizures Seizures are the presenting symptom in approximately one third of patients with brain tumors and are present at some stage of the illness in 50 to 70 percent of patients. Approximately half of the patients have focal seizures and the other half have generalized seizures. As a rule the seizures respond to standard anticonvulsant medications and may also improve following surgery.

Papilledema Papilledema is an important sign of increased intracranial pressure transmitted through the optic nerve sheath. Advances in neuroimaging have resulted in many patients being diagnosed at an earlier stage, and the frequency of papilledema in patients with brain tumor today is probably much lower than it was in the past.

Vomiting Vomiting appears in a relatively small number of patients with a tumor syndrome and usually accompanies the headache. It is more frequent in tumors involving the posterior fossa.

Altered Mental Status Mental status changes are the initial symptom in 15 to 20 percent of patients with brain tumors. These changes may range from subtle problems with concentration, memory, affect, personality, initiative, and abstract reasoning to severe cognitive problems and confusion. With increasing intracranial pressure, there is also depression of the levels of consciousness, resulting in drowsiness and eventually to stupor and coma if treatment is not given.

Regional Symptoms and Signs Sooner or later, patients with psychomotor asthenia, headaches, and seizures display focal cerebral signs (localized weakness, localized sensory loss, and paresthesias or dysesthesias, ataxia, and incoordination). Some patients may present with such signs. Nearly always, however, the focal signs at first are slight and subtle.

Cranial Nerves Tumors of the base of the skull frequently affect cranial nerves. The pattern of involvement and the presence of accompanying neurological signs depends upon the location of the tumor.

NEUROLOGICAL AND NEUROPSYCHIATRIC SYMPTOMS AND BRAIN TUMOR LOCATION

Brain neoplasms pose some difficulties for examining clinical-anatomical correlations. For example, they may not be confined to discrete areas of the brain, so a precise correlation of lesion location and behavioral change is often not possible. Particulary with infiltrating gliomas, associated edema, mass effect, hydrocephalus, and the progressive growth of the neoplasm can complicate clinical-anatomical relationships. Unlike studies of mood disorders after cerebrovascular lesions, there are relatively few studies of psychiatric disorders associated with brain tumors (Table 2.3-2 and Table 2.3-3).

Tumor Type	Lesion Location	Treatment	References
Metastases	Brainstem	Haloperidol, lithium, radiation	Greenberg D, Brown C. J Nerv Ment Dis 171:454, 1983
Glioma	Diencephalon		Stern K, Dancey T. Am J Psychiatry 98:716, 1942
Meningioma	Frontoparietal		Ogden W. Arch Neurol Psychiatry 64:457, 1950
Teratoma	Intraventricular		Alpers. Arch Neurol Psychiatry 38:291, 1937
Craniopharyngioma	Hypothalamus		Mahmud N, et al. Arch Neurol 17:112, 1967
Metastases	Parieto-temporo-occipital		Jamieson R, Wells C. J Clin Psychiatry 40:280, 1979
Metastases	Brainstem and thalamus		Greenberg D, Brown C. J Nerv Ment Dis 171:454, 1983
Meningioma	Intraventricular		Binder R. J Clin Psychiatry 44:94, 1983
Meningioma (3 cases)	Frontal	Surgical resection, lithium	Starkstein S, et al. J Nerv Ment Dis 176:87, 1988
Meningioma	Temporoparietal	Surgical resection	Starkstein S, et al. J Nerv Ment Dis 176:87, 1988
Adenoma	Pituitary	Surgical resection	Starkstein S, et al. J Nerv Ment Dis 176:87, 1988
Astrocytoma	Temporal	Radiotherapy	Starkstein S, et al. J Nerv Ment Dis 176:87, 1988
Glioblastoma multiforme	Temporal	Surgical resection	Filley C, et al. Neurol J Med 162:119, 1995

Table 2.3-2 Case Reports of Secondary Mania After Brain Tumors

Psychiatric Disorder	Tumor Type	Lesion Location	Treatment	Reference
Schizophrenia 6 patients	Astrocytoma (2 patients; oligodendroglioma (1 patient), ganglioglioma (1 patient), epidermoid cyst (1 patient)	Hippocampal gyrus (3 patients), uncus-amygdala (1 patient), cingulate gyrus and frontal (2 patients)	Electroshock, psychotherapy	Malamud N, et al. Arch Neurol 17:113, 1967
Anxiety	Chilostoma multilocular	Hippocampal gyrus		Malamud N, et al. Arch Neurol 17:113, 1967
Catatonia	Astrocytoma	Frontal		Gelenberg A, et al. Lancet 1:1339, 1976
Personality changes	Thalamic infiltration	Thalamus	Radiotherapy, dexamethasone	Gainam O, et al. J Neurosurg 4:263, 1990
Anorexia nervosa syndrome (21 cases were found in a review article)	Prolactinoma, astrocytomas, craniopharyngiomas, gliomas	Hypothalamus		Chiplonkar S. Brain Dev 16:175, 1994 (review article)
Panic Disorder	Astrocytoma	Insula, claustrum, and parahippocampal gyrus	Antidepressants	Dubovsky D, et al. Neuropsychiatry Neuropsychol Behav Neurol 2:262, 1989
Panic Disorder	Pituitary adenoma	Temporal, parahippocampal gyrus	Surgical resection	Filley C, et al. West J Med 163:59, 1995
Abulia	Meningioma	Bilateral medial frontal lobe	Surgical resection	Filley C, et al. West J Med 163:59, 1995
Auditory hallucinations	Oligoastrocytoma (1 patient; Oligodendroglioma (1 patient)	Temporal	Surgical resection	Filley C, et al. West J Med 163:59, 1995

Table 2.3-3 Case Reports of Psychiatric Changes After Brain Tumors

Frontal Lobe Initially, frontal lobe tumors are often clinically silent. As the tumor enlarges, however, there may be personality changes such as disinhibition, irritability, impaired judgment, and lack of initiative (abulia). In addition, hemiparesis, seizures, aphasia, urinary frequency and urgency, and gait difficulties may occur. Gaze preference and primitive reflexes (e.g., forced grasping and snout) may also be present. Meningiomas of the olfactory groove may produce anosmia.

Manic behaviors may also follow frontal damage (Table 2.3-2). Predisposing factors to secondary mania are either a family history of mood disorders or atrophy in frontal and diencephalic regions, probably existing before the brain tumor. Lesion location seems to be an important factor in secondary mania, because most patients reported in the literature had lesions involving ventral brain areas of the right hemisphere (mainly right orbitofrontal and basotemporal cortices). Dysfunction of these heteromodal ventral brain regions may release archicortex brain regions, resulting in disinhibited behaviors. Conversely, patients with injury to the dorsolateral prefrontal cortex often present with apathy, indifference, and psychomotor retardation. Patients with anterior cingulate lesions may be akinetic, with mutism and inability to respond to commands. Euphoria is more common in patients with right frontal damage; those with left frontal lesions tend to exhibit akinesia, abulia, and depressed affect (Table 2.3-4).

Tumor Type	Lesion Location	Treatment	Authors
Epidermoid cyst	Hippocampus and third ventricle		Malamud N, et al. Arch Neurol *17*: 113, 1967
Astrocytoma (2 patients)	Temporal	Electroshock therapy	Malamud N, et al. Arch Neurol *17*: 113, 1967
Glioblastoma multiforme	Temporal		Malamud N, et al. Arch Neurol *17*: 113, 1967
Glioblastoma (3 patients), glioma (2 patients)	Anterior half of corpus callosum (5 patients)		Nasrallah H, et al. Biol Psychiatry *16*: 663, 1981
Metastases	Frontal	Surgical resection	Filley, et al. West J Med 163:59, 1995

Table 2.3-4 Case Reports of Depression After Brain Tumors

Temporal Lobe Temporal lobe tumors frequently cause seizures. These include simple partial seizures characterized by olfactory and gustatory hallucinations, déjà vu, and feeling of fear and pleasure and complex partial seizures characterized by impairment of consciousness, repetitive psychomotor movements, and automatic behavior. Temporal lobe tumors may also cause memory disturbances, visual field defects (superior quadrantanopsia) and, when the dominant temporal lobe is involved, aphasia. Temporolimbic tumors tend to cause psychosis and schizophrenia-like illnesses (Table 2.3-3) perhaps because of disruption of limbic structures, including components of the Papez circuit: the hippocampus, fornix, mammillary bodies, mammillothalamic tract, anterior thalamic nucleus, cingulate gyrus, and parahippocampal gyrus. Mania is also associated with temporal lobe lesions, frequently involving the right side. Irritability, panic disorder, and changes in personality have also been described in patients with brain tumors in the temporal lobe.

Parietal Lobe Parietal lobe tumors are not frequently associated with psychiatric symptoms. Tumors of the parietal lobe can produce contralateral sensory loss, particularly involving joint position sense, two-point discrimination, stereognosis, and graphesthesia as well as other modalities. Lesions in the dominant parietal lobe are associated with aphasia, while lesions in the nondominant parietal lobe may result in neglect of the contralateral side and the loss of ability to acknowledge deficits (anosonognosia). Parietal lobe lesions may also lead to hemiparesis, homonymous visual deficits (or neglect), agnosia, apraxias, sensory seizures, and disturbance of visual spatial ability.

Occipital Lobe Occipital lobe astrocytomas may cause homonymous hemianopsia, visual hallucinations, and (less commonly) visual seizures characterized by lights, colors, and formed geometric patterns. Tumors at the parietooccipital junction may produce visual agnosias such as aprosopagnosia (inability to recognize faces).

Diencephalic Tumors and Third Ventricular Tumors Tumors around the third ventricle may produce symptoms resulting from hypothalamic dysfunction, autonomic dysfunction, and impaired memory. Tumors of the thalamus can produce hemiparesis, hemisensory deficits, visual disturbances, and frank alterations in mood and personality. Lesions in this region interrupt cortical-stratial-pallidal-thalamic-cortical loops, which affect many frontal lobe functions. Hypothalamic tumors can reproduce several symptoms observed in anorexia nervosa.

Corpus Callosum Tumors of the corpus callosum have been reported to produce a variety of psychiatric manifestations. The location of the lesion on the corpus callosum may be important for the production of certain psychotic symptoms. Depressive and cognitive symptoms may be produced by lesions anywhere along the anterior half of the corpus callosum.

Posterior Fossa Tumors Headaches and ataxia are the two most common symptoms. Midline cerebellar lesions may produce truncal ataxia, while lesions in the cerebellar hemispheres may cause appendicular ataxia. No correlation has been established between tumors in the posterior fossa and development of psychiatric symptoms.

Brainstem Tumors Brainstem tumors produce cranial neuropathies, weakness, numbness, ataxia, and occasionally vertigo, nausea, and vomiting. Mania has been reported after brainstem tumor.

Pituitary Tumors Pituitary tumors have been reported to produce a wide variety of neuropsychiatric symptoms from depression and apathy to paranoia. Tumors in this region can produce endocrine alterations, which can cause neuropsychiatric symptoms.

INFLUENCE OF LESION LATERALITY

A frequent observation is that patients with left hemisphere lesions are prone to depressive or catastrophic reactions, whereas patients with right hemisphere lesions

often show undue cheerfulness and indifferent emotional reactions. An emotional valence hypothesis posits that the right hemisphere processes negative emotions and the left hemisphere, positive emotions. Vascular lesions in the left anterior cortex were found to be associated with a significantly higher frequency of major depression than lesions in other locations. Another study found that left-side lesions were commonly associated with akinesia and depression, whereas right side lesions were more often associated with euphoria and underestimation by patients of the seriousness of the illnesses. More studies of tumor location and its effects on mood changes are needed.

IMAGING AND LABORATORY EXAMINATION

The introduction of computed tomography (CT) scanning and, more recently, MRI has revolutionized the diagnosis and management of brain tumors.

Skull X-Rays Plain skull films are rarely necessary today with the widespread availability of CT and MRI. Occasionally they may be useful in demonstrating calcification, bony erosion, or hyperostosis.

Computed Tomography CT scans can detect over 90 percent of brain tumors. Small tumors (<0.5 cm), tumors adjacent to bone (e.g., pituitary adenomas, clival tumors, and acoustic neurinomas), brainstem tumors, and low-grade astrocytomas may be missed and are better detected by the more sensitive MRI. CT scans tend to be better tolerated than MRI because of their shorter scanning time, and they are more sensitive in detecting calcification and bony involvement. Contrast enhancement is indispensable in the evaluation of brain tumors and may help disclose their presence (Fig. 2.3-6).

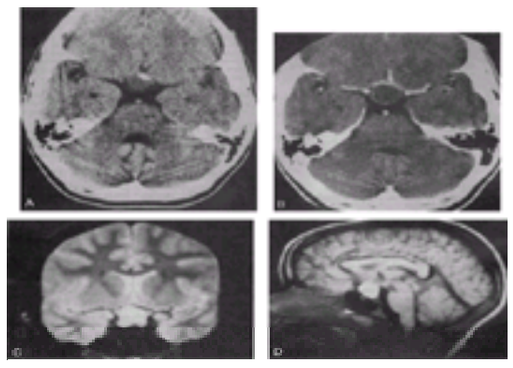

FIGURE 2.3-6 Craniopharyngioma. **A**, Axial CT scanning demonstrates a soft tissue mass in the suprasellar cistern precontrast with calcification. **B**, Rim enhancement noted postcontrast typical of a craniopharyngioma. **C**, A coronal T2-weighted MRI scan demonstrates the cystic nature of the lesion, which elevates the anterior cerebral artery A_1 segments. Note signal void in the vascular structures owing to flow. **D**, A sagittal T1-weighted MRI scan demonstrates the craniopharyngioma; typical hyperintensity on the T1 image is attributed to the cyst contents. (Reprinted with permission from Zimmerman EA, Cohen BH: Congenital tumors. In *Merritt's Textbook of Neurology*, ed 9, LP Rowland, editor. Williams & Wilkins, Baltimore, 1995.)

Magnetic Resonance Imaging MRI provides much greater anatomical detail in multiple planes and is especially useful for visualizing skull base, brainstem, and posterior fossa tumors. MRI is superior to CT scans in detecting hemorrhage and solid and cystic components within tumors and in demonstrating the relationship of the tumors to intracranial vessels. As with CT scans, the use of contrast agent greatly increases the sensitivity of the test. Newer techniques such as magnetic resonance spectroscopy, which allows direct investigation of tumor metabolism, and echoplanar MRI which scans images in less than 100 ms and provides information on tumor perfusion and diffusion, are currently being evaluated for possible use in separating recurrent tumors from radiation change.

Angiography Magnetic resonance angiography provides a noninvasive means of displaying blood vessels in the brain. It is increasingly replacing conventional angiography, although angiography has better resolution and is still necessary in certain situations.

Electroencephalography (EEG) EEG in patients with brain tumors may reveal nonspecific electrical abnormalities. However, the EEG is often normal in patients with brain tumors and thus has limited value as a screening test.

Cerebrospinal Fluid Analysis Examination of the CSF can be useful in the diagnosis of certain brain tumors and in evaluating the extent of leptomeningeal spread. Lumbar puncture holds definite risks for individuals with increased intracranial pressure and should be avoided in these patients.

DIFFERENTIAL DIAGNOSIS

The physician should be aware of the subtlety of presentation of brain tumors. A single seizure, headaches, nausea and vomiting, sensory changes, a loss of interest in usual activities, and other focal neurological signs and symptoms may all be signs of a tumor. In addition, certain situations should make the clinician suspect the presence of a brain tumor. Organic causes should be considered when a patient presents with psychiatric symptoms without previous psychiatry history, without family history of psychiatric disorder, or with atypical features. The evaluation of an intracranial mass lesion involves careful neurological and medical examinations. Neuroimaging such as CT or MRI with contrast is indicated. The differential diagnosis for brain tumors includes subdural hematoma, aneurysm, stroke, infection, demyelinative diseases, degenerative neurological diseases, hydrocephalus, and pseudotumor cerebri.

COURSE PROGNOSIS AND TREATMENT

The course and ultimate prognosis for a patient with an intracranial tumor depend upon the tumor's histological type, rate of growth, invasiveness, and response to treatment. Given the early diagnosis of brain tumors and their surgical treatment, formal treatment studies of mood disorders after brain tumors are difficult to carry out. Most patients with neuropsychiatric symptoms after brain tumors were treated with drugs frequently used among patients with functional psychiatry symptoms, such as neuroleptics, anticonvulsants, antidepressants, and lithium. These drugs should be used with care in patients with brain tumors, because of their adverse effects, such as excessive sedation and lowered seizure threshold. Whether the agents used in the treatment of mood disorders will also be useful for treating mood disorders and other psychiatric disorders associated with brain tumors remains to be empirically established. At the present time, there are no controlled studies of the treatment of psychiatric consequences of brain tumors.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2, neuroimaging in Section 1.15 and Section 1.16, schizophrenia in Chapter 12, and mood disorders in Chapter 14.

SECTION REFERENCES

*Adams RD, Victor M: *Principles of Neurology*, ed 6. McGraw-Hill, New York, 1997.

Azizi SA, Miyamoto C: Principles of treatment of malignant gliomas in adults: An overview. J Neurovirol *4:*204, 1998.

Black PM: Brain tumors (first of two parts). N Engl J Med *324:*1471, 1991.

Black PM: Brain tumors (second of two parts). N Engl J Med *324:*1555, 1991.

Buatti JM, Meeks SL, Marcus RB Jr, Mendenhall NP: Radiotherapy for pediatric brain tumors. Semin Pediatr Neurol *4:*304, 1997.

Galasko D, Kwo-On-Yuen PF, Thal L: Intracranial mass lesions associated with late-onset psychosis and depression. Psychosis Depression Elderly *11:*151, 1988.

Greenberg DB, Brown GL: Mania resulting from brain stem tumor. J Nerv Ment Dis *173:*434, 1985.

Hall WA: The safety and efficacy of stereotactic biopsy for intracranial lesions. Cancer *82:*1749, 1998.

*Irle E, Peper M, Wowra B, Kunze S: Mood changes after surgery for tumors of the cerebral cortex. Arch Neurol *51:*164, 1994.

*Kaye A, Laws E: *Brain Tumors.* Churchill Livingstone, New York, 1995.

*Keschner M, Bender MB, Strauss I: Mental symptoms associated with brain tumors: A study of 530 verified cases. JAMA *110:*714, 1938.

Lowry JK, Snyder JJ, Lowry PW: Brain tumors in the elderly: Recent trends in a Minnesota cohort study. Arch Neurol *55:*922, 1998.

Mickle JP: Neurosurgery for pediatric brain tumors. Semin Pediatr Neurol *4:*273, 1997.

Muir DF: Translational research models in neuro-oncology. Semin Pediatr Neurol *4:*292, 1997.

*Nakawatase TY: Frontal lobe tumors. *The Human Frontal Lobes: Functions and Disorders*, BL Miller, JL Cummings, editors. Guilford, New York, 1999.

Nasrallah H, McChesney CM: Psychopathology of corpus callosum tumors. Biol Psychiatry *16:*663, 1981.

Obbens EA, Shapiro WR: Brain tumors. Cancer Chemother Biol Response Modif *17:*619, 1997.

*Price TR, Goetz KL, Lowell MR: Neuropsychiatric aspects of brain tumors. In *The American Psychiatric Press Textbook of Neuropsychiatry*, ed 3, SC Yudofsky, RE Hales, editors. American Psychiatric Press, Washington, DC, 1997.

Robinson RG, Kubos KL, Starr LB, Rao K, Price TR: Mood changes in stroke patients: Relationship to lesion location. Compr Psychiatry *24:*555, 1983.

Sawaya R, Hammoud M, Schoppa D, Hess KR, Wu SZ, Shi WM, Wildrick DM: Neurosurgical outcomes in a modern series of 400 craniotomies for treatment of parenchymal tumors. Neurosurgery *42:*1044, 1998.

Shapiro WR: Management of primary brain tumors. Ann NY Acad Sci *835:*132, 1997.

Shi WM, Wildrick DM, Sawaya R: Volumetric measurement of brain tumors from MR imaging. J Neurooncol *37:*87, 1998.

Starkstein S, Boston J, Robinson RG: Mechanisms of mania after brain injury. 12 case reports and review of the literature. J Nerv Ment Dis *176:*87, 1988.

Tomita T: Neurosurgical perspectives in pediatric neurooncology. Childs Nerv Syst *14:*94, 1998.

2.4 NEUROPSYCHIATRIC ASPECTS OF EPILEPSY

MARIO F. MENDEZ, M.D., FH.D.

Epilepsy is associated with a range of biologically based psychopathology. Studies from communities, epilepsy clinics, and psychiatric hospitals demonstrate an increased prevalence of psychiatric problems among patients with epilepsy as compared to patients without epilepsy. Among the former group, much of the psychopathology results from electrophysiological, structural, or chemical changes in the temporal limbic system and possibly in the frontal lobes.

Although most patients with epilepsy are normal, one quarter or more have schizophreniform psychoses, depression, personality changes, or hyposexuality. These behavioral changes are chronic and present between seizure episodes. Other behaviors are more episodic and directly related to the seizure discharges. Patients may have auras with psychic content, nonconvulsive status epilepticus, prodromal symptoms, postictal confusion, and a periictal psychosis that usually occurs in the postictal period and is distinct from the schizophreniform psychosis of epilepsy.

There are important implications for the management of epilepsy patients who manifest psychopathology. These include maximizing the mood-stabilizing and other psychotropic effects of anticonvulsant drugs such as carbamazepine (Tegretol), valproate (Depakene), and clonazepam (Klonopin). Secondly, psychotropic medications can lower the seizure threshold, an important consideration when choosing a psychotropic drug for a brittle epileptic patient. Finally, the clinician must consider the potential interaction of anticonvulsant and psychotropic medications and monitor their respective drug levels.

DEFINITION

Epileptic *seizures* are sudden, involuntary behavioral events associated with either excessive or hypersynchronous electrical discharges in the brain. The seizure itself is known as the *ictus.* The *interictal period* refers to the period between the postictal abnormalities and the next ictus, and *periictal* refers to the period just before or after the ictus and is applied when there is insufficient information to know where the ictus ends or begins. Epileptic seizures can be primary, secondary to a neurological condition, or reactive to a situational factor such as sleep deprivation or drug withdrawal. *Epilepsy* is the current tendency to seize, and *status epilepticus* is prolonged or repetitive seizures without intervening recovery.

In epilepsy, abnormal electrical discharges are caused by hyperexcitable neurons with sustained postsynaptic depolarization. Proposed mechanisms for this sustained depolarization include changes in ionic conductance, decreased g-aminobutyric acid (GABA) inhibition of cortical excitability, and increased glutamate-mediated cortical excitation. In animals, alumina-induced membrane changes alter the ratio of intracellular-extracellular ionic concentrations and result in abnormal neuronal firing. Anticonvulsants such as phenytoin (Dilantin), carbamazepine, and valproate reduce this repetitive firing through effects on sodium channels. Ethosuximide (Zarontin) works through blockage of calcium currents. Penicillin-induced cortical injury causes seizures through decreased GABA inhibition. Barbiturates and benzodiazepines may reduce seizures by enhancing GABA-receptor current; valproate may reduce seizures through blockage of GABA catabolism. Kainic acid, a glutamate agonist, induces seizures through increased synaptic action at its *N*-methyl-D-aspartate (NMDA) receptors. Much work is under way on potential anticonvulsants that may work through inhibition of this excitatory receptor mechanism.

The electroencephalogram (EEG) is a surface recording of brain wave activity used in the evaluation of seizures. Basic waves include normal waking alpha (8 to 13 Hz), which is most prominent over the occipital region, high-frequency beta (above 13 Hz), and theta (4 to 7.5 Hz) and delta slowing (3.5 Hz or less). Seizures are manifest as multiple spikes of spike and wave discharges on the EEG (Fig. 2.4-1). A spike is a sharp transient with a duration of 20 to 70 msec. Interictally single spikes and other markers of abnormal electrical activity may be seen, often emanating from a temporal lobe.

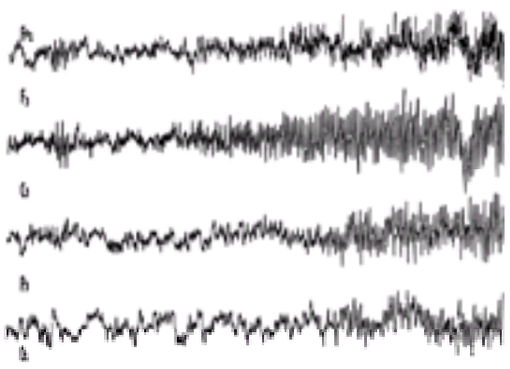

FIGURE 2.4-1 Electroencephalogram demonstrating the focal onset of seizure discharges from the left frontotemporal region consistent with the onset of complex partial seizures.

HISTORY

In his book on epilepsy, *On The Sacred Disease*, Hippocrates (460 to 377 bc) attacked the prevailing belief that those afflicted with epilepsy were possessed by gods or goddesses. He proposed that epilepsy was a brain disease caused by the blockage by phlegm of air-carrying vessels to the brain. Despite this initial view, throughout most of human history epilepsy was believed to indicate demonic possession or the accumulation of bad humors, and attempts at exorcism involved trephination, cautery of the back of the skull, diuretics, emetics, bloodletting, purging, sweating, and even intercourse to release sperm. In the eighteenth century the first scientific treatise on epilepsy since ancient times attributed seizures to masturbation. By happenstance, bromides, which were introduced to diminish libido and masturbation, proved to be the first successful anticonvulsant medication. With the development of effective anticonvulsants and the introduction of EEG, the pendulum has swung back to Hippocrates' belief that epilepsy is rooted in organic brain disease.

The purported association of epilepsy with behavioral disorders also dates to antiquity. The brain was the seat of both the falling sickness and madness and both were believed to be related to phlegm. With demonic possession as a form of punishment, unusual or abnormal behaviors became associated with seizure patients even during their seizure-free periods. At the turn of the nineteenth century the psychiatric writings of Emil Kraepelin emphasized that epilepsy patients underwent personality changes and had a predisposition to psychosis. With the greater understanding of the physical basis of epilepsy, many clinicians sought to protect patients from the demonic stigma of their disease; in their view, psychiatric problems resulted from the psychosocial difficulties associated with having seizures rather than any unique relationship of epilepsy with psychiatric illness. The current age was initiated by the definition of temporal lobe epilepsy and the concept of a physiological disturbance in the limbic or emotional brain.

NOSOLOGY

The International Classification of Epileptic Seizures divides seizures into *generalized* and *partial* (Table 2.4-1). Generalized seizures are those with an initial widespread bihemispheric involvement, and partial seizures are those that emanate from a focus limited to part of one hemisphere. In adults, most generalized seizures are tonic-clonic seizures (*grand mal seizures* or *convulsions*) and characterized by an abrupt loss of consciousness with tonic rigidity followed by a synchronous, clonic release. Partial seizures are either complex partial seizures (*psychomotor* or *temporal lobe epilepsy*) or simple partial seizures, depending on whether or not there is complex symptomatology such as an alteration of consciousness or psychic symptoms (Table 2.4-2). Simple partial seizures produce isolated motor, sensory, autonomic, psychic, or mixed symptoms in a clear sensorium. Simple partial seizures that evolve to complex partial seizures are considered auras. Complex partial seizures are usually characterized by motionless staring combined with simple automatisms (automatic motor activity) and last about 1 minute. Complex partial seizures that evolve to generalized tonic-clonic seizures are "secondarily generalized." Finally, there is a second form of generalized seizures, *absence (petit mal) seizures*, which occur less commonly in adults and are characterized by brief lapses of consciousness. Absence seizures differ from complex partial seizures in being short (10 seconds) and very repetitive; in lacking auras, postictal confusion, or complex automatisms; and in having characteristic 2 to 4 cps spike-and-wave discharges on EEG.

I. Partial (focal, local) seizures
  A. Simple partial seizures
     Motor, somatosensory, autonomic, or psychic symptoms
  B. Complex partial seizures
     1. Begin with symptoms of simple partial seizure but progress to impairment of consciousness
     2. Begin with impairment of consciousness
  C. Partial seizures with secondary generalization
     1. Begin with simple partial seizure
     2. Begin with complex partial seizure (including those with symptoms of simple partial seizures at onset)
II. Generalized seizures (convulsive or nonconvulsive)
  A. Absence (typical and atypical)
  B. Myoclonus
  C. Clonic
  D. Tonic
  E. Tonic-clonic
  F. Atonic/akinetic
III. Unclassified

Table 2.4-1 International Classification of Epileptic Seizures

Type	Symptoms	Probable Source
Dysphasic[1]	Nonfluent Impaired comprehension	Left perisylvian language areas
Dysmnestic	Déjà vu, déjà vécu, déjà pensé, déjà entendu, jamais vu, prescience, illusion of memory	Mesobasal temporal,[2] especially on right
Cognitive	Dreamy state, altered time sense, derealization, depersonalization	Mesobasal temporal and temporal neocortex
	Forced thinking, forced actions, and abstract or obscure thoughts	Frontal association cortex
Affective	Fear, anxiety, apprehension, depression, pleasure, unpleasure	Mesobasal temporal and temporal neocortex
Illusions[3]	Macropsia, micropsia, telopsia, movement, metamorphopsia, increased color intensity, increased stereopsis intensity	Lateral superior temporal neocortex, especially on right for visual illusions
Hallucinations[3]	Structured, hallucinatory remembrances, autoscopy	Mesobasal temporal and temporal neocortex

[1] Does not include speech arrest or simple vocalizations.
[2] Includes hippocampus, amygdala, and the parahippocampal gyrus.
[3] Includes within integrative (eg, emotion, illusion, stereopsis) or experiential (memories of past experience or involvement).

Table 2.4-2 Psychic Auras

EPIDEMIOLOGY

Seizure disorders are common and usually have an early onset. Epilepsy affects 20 to 40 million people worldwide, has a lifetime prevalence of at least 0.63 percent, and an annual incidence of about 0.05 percent. The overall incidence is high in the first year, drops to a minimum in the 30s and 40s, then increases again in later life. More than 75 percent of patients have their first seizure before age 18, and 12 to 20 percent have a familial incidence of seizures. Among adults, the most common seizures are complex partial seizures and generalized tonic-clonic seizures.

Psychopathology Epidemiological studies from communities, psychiatric hospitals, and epilepsy clinics have indicated a higher prevalence of psychiatric problems among epilepsy patients as compared to normal controls. About one quarter of all those with epilepsy and one half of those specifically with complex partial seizures had psychosis, depression, personality disorders, or hyposexuality. These problems are about equally divided between those that occur ictally or periictally and those that occur interictally or are variably related to the ictus. The percentage of epilepsy patients in psychiatric hospitals was also higher than the general prevalence of epilepsy and ranged from 4.7 percent of all inpatients in a British psychiatric hospital to 9.7 percent in a U.S. Veterans Affairs psychiatric facility. Among patients attending epilepsy clinics, about 30 percent had a prior psychiatric hospitalization and 18 percent were on at least one psychotropic drug. Furthermore, epidemiological studies indicate an increased interictal psychopathology among head-injured patients with epilepsy, compared to head-injured patients without epilepsy. Despite criticisms of selection bias, these studies constitute a broad spectrum of sources that indicate greater overall psychopathology in epilepsy patients.

Do epilepsy patients have greater psychopathology than other similarly impaired patients? If this were so, it would suggest that the psychopathology is of biological origin rather than a less specific reaction to chronic disease. Although disputed by some investigators, several studies report more psychopathology among epilepsy patients than among patients with long-term diseases that do not directly affect the brain. Furthermore, the pattern of behavioral changes in patients with seizures appear specific to epilepsy. For example, on the Minnesota Multiphasic Personality Inventory-2 (MMPI-2), despite a lack of difference in overall psychopathology, patients with epilepsy have higher schizophrenia scale and paranoia scale scores than patients with other neurological disabilities.

Many studies found a special relationship to psychopathology in patients whose seizures emanated from mediobasal temporal lesions. Psychiatric disturbances, primarily psychosis and personality disorders, are two to three times more common in patients with complex partial seizures, most of whom have a temporal focus, compared to those with generalized tonic-clonic seizures; other studies have failed to find a difference. Nevertheless, 60 to 76 percent of adults with epilepsy, regardless of seizure type, have a temporal lobe focus, and many generalized tonic-clonic seizures are secondarily generalized from a temporal lobe focus without a preceding complex partial seizure. Therefore, the presence of mediobasal temporal limbic pathology cannot be excluded in comparison groups with generalized tonic-clonic seizures.

Psychosis Psychosis is the specific psychiatric disorder most clearly associated with epilepsy. The lifelong prevalence of all psychotic disorders among epilepsy patients ranges from 7 to 12 percent. In a follow-up of 100 children with complex partial seizures for up to 30 years, of the 87 who survived to adulthood and did not have mental retardation, 9 (10 percent) experienced a psychotic illness. Moreover, in temporal lobectomy studies, where there is surgical removal of an epileptic focus, psychosis occurred in 7 to 8 percent of patients, even long after the seizures were arrested. That represents about a twofold or greater risk of psychosis for patients with epilepsy than for the general population; patients whose epilepsy has a mediobasal temporal focus are especially at risk.

Studies on the laterality of the seizure focus suggest an association of a left-sided focus with psychosis. Although conclusions derived from surface EEG recording are open to criticism, depth recording of presurgical patients show that twice as many patients with left temporal lesions have psychosis. Positron emission tomography (PET) scans and single-photon emission computerized tomography (SPECT) scans may show predominant left temporal hypometabolism in epilepsy patients with psychosis.

Depression The prevalence of depression in various studies varies and may range from 7.5 to 34 percent of patients with epilepsy. Those with complex partial seizures and poor seizure control are more likely to suffer from mood disorder. Psychological studies also suggest a greater incidence of ideational orientation, self-criticism, and depression among epilepsy patients with a left hemisphere focus. In comparison, other studies have not established the proposed association of

right hemisphere foci with mania and other disorders. Patients with complex partial seizures of temporal lobe origin have a higher incidence of depression than do patients with other types of seizure disorders.

Other Behaviors The prevalence of other specific behavioral disorders among patients with epilepsy is less well established. There is convincing evidence, however, that personality disorders, suicidal behavior, and hyposexuality are more prevalent among epilepsy patients than among those without seizure disorders.

ETIOLOGY

Most new-onset epilepsy is idiopathic, but other frequent causes include trauma in the third and fourth decades, neoplasms in the fifth and sixth decade, and cerebrovascular disease in the elderly. Although some complex partial seizures originate from frontal or temporal neocortex and other areas, at least two thirds of complex partial seizures and generalized tonic-clonic seizures originate from the mediobasal temporal limbic structures (hippocampus, amygdala, and parahippocampal gyrus).

Psychopathology The relationship of seizures, psychiatric syndromes, and the mediobasal temporal lobes implies that many behavioral changes are more than psychological reactions to the psychosocial stressors of epilepsy. Stimulation and ablation studies in man and animals link temporal limbic structures to emotional behavior.

For example, temporal limbic stimulation of humans evokes psychic auras and automatisms; amygdalar stimulation and ablation of animals results in either aggression or placidity. Moreover, psychotic behavior in cats occurs when their limbic structures undergo kindling (i.e., the repeated application of epileptic agents in order to induce lasting behavioral changes).

There are several potential organic causes of psychiatric disturbances in epilepsy (Table 2.4-3). First, the pathology itself could be the source of both the seizures and the behavioral changes. Left hemisphere and temporal lobe lesions may be associated with a schizophreniform psychosis, and psychosis in epilepsy may be particularly frequent if there is specific underlying pathology or ventricular enlargement. Second, ictal or subictal epileptiform activity may promote behavioral changes by facilitating distributed neuronal connections, increasing limbic-sensory associations, or changing the overall balance between excitation and inhibition. Third, the absence of function, such as the interictal hypometabolism observed on PET scans (Fig. 2.4-2), may lead to depression or other interictal behavioral changes. Among epilepsy patients with a schizophreniform psychosis, SPECT scans have shown reductions in cerebral blood flow in the left medial temporal region. Fourth, seizures may result in neuroendocrine or neurotransmitter changes such as increased dopaminergic or inhibitory transmitters, decreased prolactin, increased testosterone, or increased endogenous opioids, all of which can affect behavior. Furthermore, neurobiological factors may be potentiated by psychodynamic factors such as feelings of helplessness, learned helplessness, dependency, low self-esteem, and the disruption of reality testing. In sum, the psychiatric manifestations of epilepsy are heterogeneous disorders with a multiplicity of causes.

Common neuropathology, genetics, or developmental disturbance
Ictal or subictal discharges potentiate abnormal behavior
Kindling or facilitation of a distributed neuronal matrix
Changes in spike frequency or inhibitory-excitatory balance
Altered receptor sensitivity (e.g., dopamine receptors)
Secondary epileptogenesis
Absence of function at the seizure focus
Inhibition and hypometabolism surrounding the focus
Release or abnormal activity of remaining neurons
Dysfunction or downregulation of associated areas
Neurochemical
Dopamine and other neurotransmitters
Endorphins
Gonadotrophins and other endocrine hormones
Psychodynamic and psychosocial effects of living with epilepsy
Dependence, learned helplessness, low self-esteem, weak defense mechanisms
Disruption of reality testing
Neurobiological and psychodynamic factors potentiate each other
Sleep disturbance
Anticonvulsant drug-related

Table 2.4-3 Proposed Relations of Psychiatric Disturbances to Epilepsy

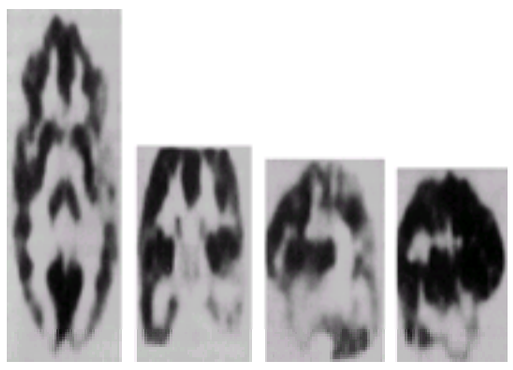

FIGURE 2.4-2 ^{18}Fluorodeoxyglucose positron emission topography (PET) scans demonstrating interictal hypometabolism in the left temporolimbic region. (Reprinted with permission from Engel J Jr: *Seizures and Epilepsy.* FA Davis, Philadelphia, 1989.)

DIAGNOSIS

In epilepsy, psychiatric behaviors can be conceptualized in relation to the ictus or seizure discharges. These behaviors occur as part of the ictus, periictally, or during the interictal period (Table 2.4-4). Moreover, a range of other behaviors appear to have some relation to the ictus but do not clearly fall into one of the former three categories.

I. Ictal
A. Ictal psychic symptoms
B. Nonconvulsive status: simple partial seizures, complex partial seizures, and periodic lateralizing epileptiform discharges
II. Periictal (includes predromal, postictal, and mixed ictal)
A. Prodromal symptoms: irritability, depression, headache
B. Postictal delirium
C. Periictal psychotic symptoms
1. Concomitant with increased seizure frequency
2. Concomitant with decreased seizure frequency
3. The postictal psychoses
III. Interictal
A. The schizophreniform psychosis
B. Personality disorders
C. The Gastaut-Geschwind syndrome
IV. Variably related to ictus
A. Mood disorders (depression and mania)
B. Dissociative states
C. Aggression
D. Hyposexuality
E. Suicide
F. Other behaviors

Table 2.4-4 Behavioral Disorders in Epilepsy

Ictal Seizure discharges can produce semi-purposeful automatisms and psychic auras such as mood changes, derealization and depersonalization, and forced thinking. Ictal fear, which ranges from a vague apprehension to abject fright, has occurred without any other seizure manifestation, and ictal depression has extended

days or longer after the seizure has passed. Some patients have pleasurable auras. Fyodor Dostoyevsky had “ecstatic auras” during which he felt in perfect harmony with the entire universe and “would give 10 years of this life, perhaps all of it, for a few seconds of such bliss.” The experience of epileptic derealization or depersonalization could impair reality testing. Another psychic aura is “forced thinking,” characterized by recurrent intrusive thoughts, ideas, or crowding of thoughts. Forced thinking must be distinguished from obsessional thoughts and compulsive urges. Epilepsy patients with forced thinking experience their thoughts as stereotypical, out-of-context, brief, and irrational, but not necessarily as ego dystonic.

A 36-year-old right-handed man presented with frontal headaches and 5 years of complex partial seizures. His seizures began with 15 seconds of a sense of “impending doom,” speech arrest, and orobuccal movements followed by 30 seconds of altered consciousness. At seizure onset, he felt forced to think the phrase “tell me yes.” The phrase repeated itself several times in his mind without his being able to control it. Concomitantly, his mouth would open, and he would attempt to say the phrase but could utter only unintelligible sounds. The patient interpreted this phrase as a call for help. On examination, he had a mild memory deficit, normal language testing, a right facial droop, and brisk right-sided reflexes with a right Babinski response. Neuroimaging revealed a left frontal mass lesion. EEGs showed amplitude attenuation and polymorphic delta in the left frontal area, and intraoperative electrocorticography disclosed polyspike and spike-wave discharges and impaired language from just below the lesion. The patient underwent subtotal resection of a 4.3 × 3 × 3 cm oligodendroglioma. Postoperatively his forced thinking and seizures resolved, but he had a nonfluent aphasia.

Cognitive disorders follow status epilepticus with simple partial seizures, complex partial seizures, or absence seizures. Recurrent or prolonged simple partial seizures do not result in alteration of consciousness or invariable abnormalities on EEG, and, if manifested by psychic auras, simple partial seizures may be difficult to distinguish from primary psychiatric disturbances. Status epilepticus from complex partial seizures and absence seizures results in prolonged alterations of responsiveness. With the addition of various ictal auras, complex partial status epilepticus can appear psychotic. Occasionally, EEGs and a therapeutic trial of anticonvulsant medications may be the only way to distinguish behavioral disturbances caused by nonconvulsive status epilepticus. Finally, recurrent EEG complexes known as *periodic lateralizing epileptiform discharges* may also be associated with prolonged confusional behavior and focal cognitive changes.

A 68-year-old man had a left temporal-parietal hemorrhagic stroke. An initial fluent aphasia and right hemiparesis completely resolved, but he developed poststroke epilepsy. His seizures began with speech arrest and were followed by secondary generalization to tonic-clonic seizures. The postictal periods lasted for days because of continued left hemisphere periodic lateralizing epileptiform discharges. During these prolonged postictal periods, he was confused, placid, and had a return of his aphasia. One year later, after achieving seizure control, the patient developed mania for the first time in his life. His mania was in a clear sensorium without a change in his neurological examination or epileptiform activity on EEG. He did not sleep, had flight of ideas, and had grandiose ideation including beliefs that he was a three-star general, had killed Hitler, and was now a millionaire. He exposed himself to everyone including his daughter and inserted pencils up his penis because he believed that he needed catheterization. His psychosis lasted for 3 months until he had two generalized tonic-clonic seizures. Postictally, for 10 days he remained placid, confused, and aphasic, with a right beating nystagmus and periodic lateralizing epileptiform discharges maximal in the left temporal region (Fig. 2.4-3). With a new anticonvulsant medication he returned to normal with total resolution of his mania.

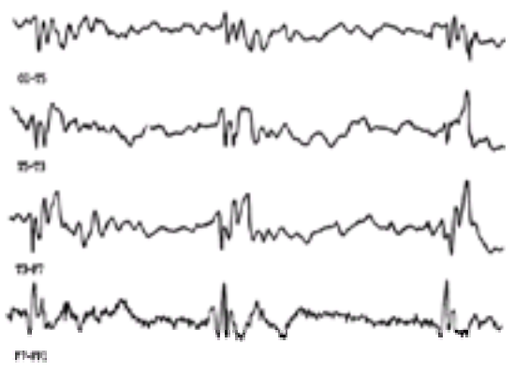

FIGURE 2.4-3 Periodic lateralizing epileptiform discharges (PLEDs).

Periictal Features Psychiatric disturbances can occur before seizures *(prodromal)*, after seizures *(postictal)*, or during intermittent seizure activity. Some patients experience prodromal symptoms that begin at least 30 minutes before seizure onset, last 10 minutes to 3 days, and are continuous, with irritability, depression, headache, confusion, and other symptoms. The postictal period is characterized by a delirium lasting minutes to hours or, occasionally, days. Prolonged postictal confusion may particularly follow right temporal complex partial seizures. Some twilight states result from a protracted period of intermixed ictal and postictal changes.

Periictal psychotic symptoms often worsen with increasing seizure activity. Rarely, psychotic symptoms alternate with seizure activity. In this “alternating psychosis,” when patients are having seizures they are free of psychotic symptoms, but when they are seizure free and their EEG has “forced or paradoxical normalization,” they manifest psychotic symptoms. This alternating pattern is much less common than the increased emergence of psychotic behavior with increasing seizure activity.

An important periictal psychiatric disorder consists of brief psychotic episodes that follow clusters of generalized tonic-clonic seizures, (i.e., postictal psychosis). These psychotic episodes occur in patients who have complex partial seizures with psychic auras, frequent secondary generalization to tonic-clonic seizures, and bilateral interictal discharges. The postictal psychosis of epilepsy emerges after a lucid interval of 2 to 72 hours (mean of 1 day) during which the immediate postictal confusion resolves and the patient appears to return to normal. The postictal psychotic episodes last 16 to 432 hours (mean of 3½ days) and often include grandiose or religious delusions, elevated moods or sudden mood swings, agitation, paranoia, and impulsive behaviors but no perceptual delusions or voices heard. The postictal psychoses remit spontaneously or with the use of low-dose psychotropic medication.

A 33-year-old man with a 15-year history of generalized tonic-clonic seizures and a 4-year history of periictal psychotic episodes had several hospitalizations for recurrent postictal psychosis. The initial flurry of generalized tonic-clonic seizures was followed by a 24- to 48-hour latency period, and, subsequently, 2 to 7 days of delusions, hallucinations, and disordered thought processes. He believed that people could transmit messages and read his thoughts, and voices commanded him to love his neighbor. The patient claimed to be able to predict the future and to communicate with a dead grandfather who voiced dissatisfaction with things on earth. During these episodes, the patient had loose associations, euphoria, agitation, and occasional spike and waves on EEGs. Between psychotic episodes, he was psychiatrically and neurologically normal, and his EEGs showed left temporal interictal spikes. After the postictal psychosis, the patient returned to baseline without residual changes in behavior.

Interictal Features

Schizophreniform Psychosis Most epilepsy patients with a schizophreniform psychosis have a chronic interictal illness without known direct relationship to seizure events or ictal discharges. Many of these patients, however, develop worsening psychotic symptoms concomitant with an increase in seizure frequency or with anticonvulsant withdrawal and a few others have worsening psychotic symptoms on control of the seizures (“alternating psychosis”). The terms *alternating psychosis* and *forced normalization* refer to this demonstrable antagonism between the psychosis and the seizures or EEG discharges. Epilepsy patients with this long-term interictal psychosis often have an 11- to 15-year history of seizures, usually poorly controlled complex partial seizures with secondary generalized tonic-clonic seizures. However, seizure control with anticonvulsants or removal of the seizure focus does not prevent the development of this psychosis, which occasionally emerges for the first time after successful seizure treatment. This disorder sometimes resembles a schizoaffective psychosis with intermixed affective symptoms. In addition, there are prominent paranoid delusions, relative preserved mood, normal premorbid personality, and no family history of schizophrenia. Other reported differences with idiopathic schizophrenia are outlined in Table 2.4-5.

Epilepsy Characteristics:
Complex partial seizures with secondary generalized tonic-clonic seizures
More auras and automatisms than in nonpsychotic epilepsy patients
Epilepsy present for 11 to 15 years before psychosis
Long interval of poorly controlled seizures
Recently diminished seizure frequency, especially generalized tonic-clonic
Left temporal focus
Mediobasal temporal lesions, especially tumors
Psychosis Characteristics:
Atypical paranoid psychosis—parania with sudden onset
Psychosis alternating with seizures
Preserved affective warmth
Failure of personality deterioration
Less social withdrawal than schizophrenia
Less systematized delusions than schizophrenia
More hallucinations and affective symptoms than schizophrenia
More religiosity than schizophrenia
More positive as opposed to negative symptoms
Few Schneiderian first-rank symptoms
Negative family history for psychosis
No family history of schizophrenia

Table 2.4-5 Proposed Predisposing Factors for the Interictal Schizophreniform Psychosis of Epilepsy

A 23-year-old man developed paranoid delusions after his daily complex partial seizures were controlled for the first time. His seizures dated from age 8 and consisted of a rising epigastric sensation and facial flushing followed by a motionless stare and automatisms, often culminating in secondary generalized tonic-clonic seizures. Prior to initiating anticonvulsant therapy, there was no history of paranoid or psychotic behavior. Afterwards, he believed that people were sending energy to him through small concealed batteries. He felt able to work this energy off with his fluorescent watch dial and a one-armed plastic crucifix in his boot. The patient also believed that people were watching him, trying to manipulate him, and threatening him through telephone lines and telephone poles. His examination was remarkable for the degree of emotion when relating his bizarre ideas. He had a lesion in the left anterior temporal area, probably consistent with an old calcified cyst, and left temporal spikes on EEG. His paranoid delusions subsequently abated with antipsychotic therapy.

Personality Disorders Among patients with epilepsy there is a high prevalence of personality disorders including borderline, atypical or mixed, histrionic and dependent disorders. Patients with personality disorders show dependent and avoidant personality traits. The most common personality disorder in epilepsy is a borderline personality disorder. Not surprisingly, patients with epilepsy frequently lack a stable character structure and can be immature and impulsive. This personality constellation partially explains the increased incidence of irritability, suicide attempts, and intermittent explosive disorder. Those with epilepsy are stigmatized, feared, and subject to difficulties in obtaining a job, driving an automobile, and maintaining a marriage. These psychosocial difficulties along with any associated mental retardation contribute to the dependency, low self-esteem, and overall borderline personality traits present in many such patients. In addition, the experience of epileptic auras may contribute to the development of personality disorders.

Gastaut-Geschwind Syndrome Although there is no general personality profile of an epilepsy patient, a group of traits termed the *Gastaut-Geschwind syndrome* occur in a subset of patients with complex partial seizures. Some epilepsy patients with a temporal limbic focus develop a sense of the heightened significance of things. These patients are serious, humorless, overinclusive, and have an intense interest in philosophical, moral, or religious issues. Occasionally, patients with epilepsy experience multiple religious conversions or experiences. In interpersonal encounters, they demonstrate “viscosity,” the tendency to talk repetitively and circumstantially about a restricted range of topics. They can spend a long time getting to the point, give detailed background information with multiple quotations, or write copiously about their thoughts and feelings *(hypergraphia).* Viscosity may particularly occur in patients with left-sided or bilateral temporal foci.

A 39-year old man developed seizures following a contusion of the left temporal region. His seizures began with stereotypical voices and generalized to tonic-clonic seizures. He was extremely circumstantial and tangential, stressed every detail, and had difficulty getting to the point. Ironic and minor philosophical insights were fascinating to him. He wrote 30-page rambling letters to his physician, and his writings were full of metaphors and quotes. An example of his writing was as follows: “I became overwhelmed by the sentiment of a letter composed in my head before reaching paper. The sentiment of this letter continued to expand in all dimensions until it seemed no longer connected to any specific ideal, but more to an all-pervasive color, yellow, and a smell, like burning leaves. I felt deliriously happy, but I felt in danger as well. Afterwards I got an acute attack of aphasia and could do nothing but shrug. My prior prophet voices, which went away with the Dilantin, were saying something profound that I needed to get down on paper. It seems as though I am a prophet and I will never have another problem for the rest of my life.”

Conclusive proof that epilepsy patients with a temporal lobe foci are disproportionately prone to the Gastaut-Geschwind syndrome has remained elusive. Most of the early studies used the MMPI, a test that proved insensitive to most of the specific traits attributed to epilepsy. Studies with the Bear-Fedio Inventory, an MMPI-like instrument developed to assess these “epileptic traits,” found that epilepsy patients with temporal lobe foci were sober and humorless, dependent, circumstantial, and had strong philosophical interests. In addition, those with a left-sided focus had a more reflective ideational style and maximized their problems whereas those with a right-sided focus had emotional tendencies and minimized their problems. Further investigations with the Bear-Fedio Inventory described these seizure patients as having viscosity in interactions, prominent religious interests, a pronounced sense of personal destiny, and deepened affect. However, other applications of this inventory found the same characteristics in patients without epilepsy who had psychiatric disorders or comparable physical disabilities. Although these personality characteristics do occur in some epilepsy patients, they may not be specific for patients with seizure disorders.

Variably Related to Ictus

Mood Disorders Depressive disorder is the most prevalent neuropsychiatric disorder in epilepsy, occurring in 7.5 to 34 percent of epilepsy patients. Depression is also the main diagnosis among patients with epilepsy in mental hospitals. Depression is twice as common in epilepsy patients as in comparably disabled populations, suggesting that much of the depression in epilepsy patients is more than just a psychological reaction to a disability.

Most patients with epilepsy have a chronic interictal depression or dysthymia. These patients frequently have accompanying paranoia and hallucinations, emphasizing the continuum with psychotic disorders. These depressed patients tend to have frequent complex partial seizures, particularly with a left-sided temporal focus. The experience of certain psychic auras, especially those with cognitive content, may predispose to interictal depression. Several investigators also report increased seizure control or a decrease in seizures prior to the onset of interictal depressive symptoms. Patients with this alternating depression experience relief with a seizure or electroconvulsive therapy.

There are other associations of depression with epilepsy. The rare occurrence of ictal depression may not only outlast the actual ictus but may lead to suicide. Depression also occurs periictally. Episodic mood disturbances, often with agitation, suicidal behavior, and psychotic symptoms, may occur with increasing seizure activity. Finally, postictal depression is common, and, a prolonged depressive state occasionally follows complex partial seizures even when ictal experiences do not include depression.

Mood disorder due to epilepsy with manic features or with mixed features is much rarer than mood disorder due to epilepsy with depressive features or with major depressive-like features. Rarely, manic symptoms may emerge with an increase in seizure frequency or after seizure control. Although a right temporal focus was suggested as the source of mania in epilepsy, this laterality has not been established.

Dissociative States A specific association of epilepsy with dissociative identity disorder, depersonalization disorders, possession states, fugue states, and psychogenic amnesia is intriguing but unresolved. Studies of patients with dissociative identity disorder reveal frequent EEG changes but few actual seizures. It is conceivable that temporary personality disintegration occurs in some patients as part of postictal confusion or periictal psychosis, particularly in those with a right temporal focus. In one intracarotid amobarbital (Amytal) study of multiple personality in two seizure patients, the different personalities were precipitated without seizure activity. Persistent alterations in the experience of self and feelings of being taken over by others may occur in patients with auras of derealization and depersonalization. In epilepsy, prolonged periods of compulsive wandering with amnesia have resulted from an admixture of ictal and postictal changes and has been termed *poriomania.* Finally, some patients may have periods of amnesia or “lost time,” possibly caused by complex partial seizures without surface EEG abnormalities.

Aggression Lay people have accredited to epilepsy aggressive and violent acts and have even used the epilepsy defense in criminal proceedings. This belief peaked in the nineteenth century when the criminologist Cesare Lombroso promoted the association of epilepsy with aggressive, sociopathic tendencies. Investigators have bolstered this association with studies showing aggressive verbalizations with stimulation of the amygdala and interictal defensive rage in cats with epileptic hippocampal lesions. Among patients in a maximum-security mental hospital, the violent patients had focal temporal slowing or sharp waves on EEG and dilated

temporal horns or small temporal lobes on computerized tomography (CT) scans. These results suggest that high violence rating scores are associated with abnormal temporal electrical discharges on EEG and temporal lobe abnormalities on CT. Moreover, patients with left temporal lobe seizure foci have higher scores on hostile feelings than other patients with epilepsy.

A 37-year-old left-handed man with epilepsy presented with aggressive episodes. The seizures consisted of an olfactory aura followed by “spacing out” or alteration of consciousness for approximately 1 minute. In addition to these complex partial seizures the patient had occasional secondary generalized tonic clonic seizures with urinary incontinence and tongue-biting. During the postictal period, as he began to recover consciousness he experienced an overwhelming sense of threat or of having been harmed. These feelings became focused on any individual who was in his immediate environment. That person was believed to have beaten or otherwise hurt him and was going to harm him further. The patient felt compelled to attack these individuals, often inflicting significant physical injury. Although his postictal confusion would clear in about 1 hour, his sense of being harmed or threatened slowly diminished over about 24 hours after a seizure. After the resolution of these feelings, he felt great remorse over the harm that he had done. Nevertheless, on several occasions he was charged with aggravated assault. Sleep-deprived EEGs confirmed the presence of left anterior temporal epileptiform activity. The patient's aggressive postictal episodes abated with control of his complex partial seizures with carbamazepine.

Although aggression can occur in relation to an ictus, as in this patient, most aggression among epilepsy patients is not related to epileptiform activity. Aggression in epilepsy is usually associated with psychosis or intermittent explosive disorder and correlates with subnormal intelligence, lower socioeconomic status, childhood behavior problems, prior head injuries, and possible orbital frontal damage. Moreover, although the prevalence of epilepsy among prison inmates has been 2 to 4 times that among the general population, studies from the United Kingdom and the United States have not found more violent crimes among prisoners with epilepsy than prisoners without epilepsy.

Can violence itself be a seizure? Following the 1976 case of a New York city policeman who had never had seizures and successfully claimed the epilepsy defense, criteria for ictal violence were proposed that included video-EEG telemetry (Table 2.4-6). Since then, epilepsy has rarely if ever been proved to directly result in premeditated violence. Such acts require a series of coordinated steps that rarely occur as manifestations of seizures. Simple violent automatisms, such as spitting or flailing the arms, can occur at the onset of complex partial seizures, and secondary violent automatisms can occur as a response to an unpleasant or emotional aura or periictal sensation (Table 2.4-7). More commonly, nondirected violent movements, aimless destructive behavior, or angry verbal outbursts occur during postictal delirium when patients misinterpret attempts to protect or restrain them.

1. The diagnosis of epilepsy is established by at least one specialist in epilepsy.
2. The presence of epileptic automatisms are documented by history and by closed-circuit TV-EEG telemetry.
3. The presence of violence during epileptic automatisms is verified in a videotape-recorded seizure in which ictal epileptiform patterns are also recorded on the EEG.
4. The aggressive act is characteristic of the patient's habitual seizures, as elicited by history.
5. A clinical judgment is made by the epilepsy specialist attesting to the possibility that the aggressive act was part of a seizure.

Table 2.4-6 Criteria for Assessing Ictal Violence in Epilepsy

Period	Cause
Interictal	Impulse-control disorder
	Mental retardation or cognitive impairments
	Personality disorders
	Schizophrenia-like psychosis of epilepsy
	Medication-related
Prodrome	Mounting tension, irritability
Ictal	A direct manifestation of the seizure
	Violent automatism
	Reaction to a negative aura
	Subtle, seizure equivalents
Postictal	Resistive
	Poriomania and somnambulism
	Postictal psychosis
	Postictal persistent perception of threat

Table 2.4-7 Mechanisms of Aggression Among Epilepsy Patients

Sexuality Patients with epilepsy tend to be hyposexual. Both men and women experience disturbances of sexual arousal and a lower sexual drive. Some patients have a disinterest in all the usual libidinous aspects of life, including loss of erotic fantasies or dreams, and may suffer from impotence or frigidity. Other patients with epilepsy claim sexual desire but experience physiological signs of decreased sexual arousal. Men have an increased risk of erectile dysfunction, suggesting a neurophysiological component, and studies of sex hormones suggest the possibility of a subclinical hypogonadotropic hypogonadism. Substantial improvement to the point of public hypersexuality can occur after seizures are brought under control. Moreover, prior to temporal lobectomy, most epilepsy patients are hyposexual, but nearly a third of them have an increase in libido after the operation, providing that their seizures are controlled.

Other sexuality changes are rare. Individual cases of transvestism, fetishism, and gender dysphoria are not frequent enough to exclude a coincident association. True ictal sexual manifestations are also unusual; however, libidinous feelings, erotic sensations, sexual remembrances, and even orgasm do occur, although rarely and primarily in women, probably from seizure discharges in the amygdala. In addition, ictal masturbation has occurred with absence status. A woman with nymphomania proved to have incidental sexuality from sensory simple partial seizures caused by a tumor in the sensory cortex, representing her genital region.

Suicide The risk of completed suicide in epilepsy patients is four to five times greater than among the nonepileptic population, and those with complex partial seizures of temporal lobe origin have a particularly high risk, up to 25 times greater. A comparison of suicide attempts among patients with epilepsy and comparably handicapped controls without epilepsy has reported that 30 percent of those with epilepsy had attempted suicide as compared to 7 percent of the controls. This increased risk of suicide continues even long after temporal lobectomy and successful control of seizures. Most suicidal behavior among epilepsy patients is not directly caused by reactions to the psychosocial stressors of having a seizure disorder. Rather, these patients are likely to attempt suicide in conjunction with borderline personality behaviors and are likely to complete suicide because of psychosis. Contributors to successful suicides include paranoid hallucinations, agitated compunction to kill themselves, and occasional ictal command hallucinations to commit suicide.

A 26-year-old woman had her initial seizure during her first pregnancy at age 18. Her seizures included echoing sounds “like walking in a cave,” a motionless stare with stereotypical automatisms, postictal confusion, and occasional secondary generalized tonic-clonic seizures. Because of a variable response to anticonvulsant medications she underwent EEG and closed-circuit television video-EEG (CCTV-EEG) telemetry that documented complex partial seizures from a right temporal focus as well as nonepileptic seizures. The patient, who had six children by six different men, had prominent feelings of inadequacy and isolation, and was considering cutting her wrists “just to see if anyone cared.” Her multiple suicide attempts and threats resulted in five psychiatric hospitalizations. During one period of time, she complained of decreased menses, weight gain, stretch marks, increased appetite and sleep, and exhaustion. She insisted that she was pregnant despite six negative pregnancy tests and multiple evaluations.

Other Behavioral Changes Other psychiatric disorders may be associated with epilepsy or epileptiform EEG activity. Anxiety disorders occur among epilepsy patients and must be distinguished from simple partial seizures manifesting as anxiety or panic attacks. Among the impulse-control disorders, intermittent explosive disorder is characterized by a prodromal mounting tension and irritability, postictal remorse, and increased temporal spikes on EEG. Among the somatoform

disorders, some epilepsy patients have a conversion disorder, often manifest as nonepileptic seizure events. Finally, patients with epilepsy are subject to other behavioral difficulties stemming from their epilepsy such as adjustment disorders, subtle cognitive effects of seizures, and the potential behavioral effects of anticonvulsant medications.

PATHOLOGY AND LABORATORY EXAMINATION

Neurodiagnostic Tests In addition to the routine laboratory data and toxicology screens used to exclude reactive seizures, several neurodiagnostic tests are useful in the assessment of epilepsy. EEG is the most widely used confirmatory test for seizures; however, single EEGs are frequently normal and must be repeated, particularly with provocative maneuvers such as sleep. Occasionally, CCTV-EEG telemetry for an extended period of time is necessary to capture seizure activity. Neuroimaging procedures such as CT scans and magnetic resonance imaging can more precisely visualize a seizure focus or even a mesial temporal sclerosis (Fig. 2.4-4). Other tests that occasionally aid in localizing the seizure focus include quantitative EEG, SPECT scans, and PET scans. PET scans may show interictal hypometabolism around the temporal seizure focus and are also useful in the presurgical assessment of patients with medically intractable seizures. Neuropsychological examinations, particularly during a Wada test, further help in localizing and lateralizing memory and language prior to surgery.

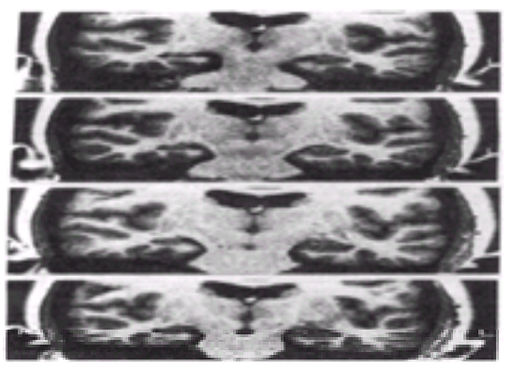

FIGURE 2.4-4 A series of magnetic resonance imaging scans demonstrating mesial temporal sclerotic changes in the left hippocampal region. (Reprinted with permission from Engel J Jr: In vivo imaging of the temporal lobe limbic system. In *The Temporal Lobes and the Limbic System*, Trimble MR, Bolwig TG, editors. Wrightson Biomedical Publishing, Petersfield, England, 1992.)

Neuropathology The common pathological findings in epilepsy are mediobasal temporal lobe lesions. About two thirds of adults with epilepsy have a temporal lobe focus and two thirds of these have mesial temporal sclerosis with pyramidal cell loss in the hippocampus. Theories about the cause of mesial temporal sclerosis include perinatal insults, dysgenesis, and kindling from reactive seizures. Another 20 to 25 percent of those with temporal lobe lesions have tumors such as hamartomas and gangliogliomas. The rest have scars from trauma and other causes or lack a distinct histological lesion.

DIFFERENTIAL DIAGNOSIS

Clinicians must distinguish epileptic seizures from two other transient behavioral events, *syncope* and *nonepileptic seizures (pseudoseizures).* Syncope is a loss of consciousness usually with premonitory lightheadedness, autonomic reactivity, a brief atonic ictus, and little or no postictal confusion. Syncope lacks the many characteristic features of seizures as well as a clear epileptiform EEG. Nonepileptic seizures, on the other hand, are involuntary, psychogenically induced spells that mimic many epileptic behaviors.

Differentiating epileptic seizures from nonepileptic seizures can be extremely difficult, and even epileptologists are incorrect 20 to 30 percent of the time. Patients with nonepileptic seizures are most commonly women between the ages of 26 and 32 years with psychological stressors and poor coping skills. Fifteen percent or more of these patients have a true seizure disorder as well, and nonepileptic seizures may result from the elaborating or “highlighting” of their epileptic seizures. Nonepileptic seizures are most commonly characterized by unresponsiveness with motor activity that does not fit a typical complex partial seizure or a generalized tonic-clonic seizure (Table 2.4-8). In children, pseudoseizures are usually characterized by unresponsiveness, with violent and uncoordinated movements of the whole body. However, every epileptic behavior can occasionally occur, including tongue biting and incontinence, and nonepileptic events are especially difficult to differentiate from the atypical motor behavior of frontal lobe epilepsy. The most helpful differentiation feature may be an ictal duration of 2 or more minutes. In addition, nonepileptic seizures usually occur in the presence of a witness; can often be induced with injections, hypnosis, or suggestions; and are poorly responsive to treatment with anticonvulsant medications. Ultimately, the differentiation may require CCTV-EEG telemetry along with the assessment of the absence of a seizure-induced rise in serum prolactin levels.

Table 2.4-8 Nonepileptic Seizures Versus Epileptic Seizures

Nonepileptic seizures result from a variety of psychiatric conditions. The most common psychiatric disturbance among these patients is conversion disorder. Patients with nonepileptic seizures who have conversion disorder have a high incidence of prior trauma or sexual or physical abuse. The remaining patients with nonepileptic seizures have depression, anxiety disorders, posttraumatic stress disorder, or borderline or other personality disorders. Additional diagnoses associated with nonepileptic seizures are psychosis, impulse control problems, and mental retardation.

One additional differentiation is that of most nonepileptic seizures from those specifically due to the malingering or feigning of epilepsy for secondary gain (Table 2.4-9). Epileptic seizures lend themselves to malingering because of their behavioral and episodic nature and the lack of consistent physical or diagnostic findings.

Malingered Seizures	Nonmalingered Nonepileptic Seizures
Preceding	Ictus
More common in males	Marked female predominance
Less likely to obtain prior abuse history	Prior history of physical or sexual abuse
Less likely to obtain prior psychiatric history	Prior psychiatric history
Evident secondary gain	No clear secondary gain*
No clear emotional precipitants	Frequent emotional precipitants
Seizures are not suggestible	Seizures may be easily suggested
During	Ictus
Seizures under volitional control	Seizures not under volitional control
Conscious awareness of seizures	Subconscious awareness of seizures only
Cannot maintain deficits over time	Able to maintain over time
Errors in seizure behavior are likely to be major distortions	Errors in seizure behavior are likely to be omissions, perseverations, near-misses
Following	Ictus
Angry, anxious on confrontation with the lack of evidence for epileptic seizures	Indifferent, detached
Uncooperative, including circumstantial and evasive answers; may leave against medical advice	Cooperative with the workup, but answers may be devoid of content

* In factitious disorder, gain is the internal behavior itself with the consequent medical attention.

Table 2.4-9 Malingered Seizures Versus Nonmalingered Nonepileptic Seizures

A 33-year-old veteran of the Persian Gulf War presented with a complaint of epileptic spells, beginning 3 years after returning home. The patient claimed that the stress of the war induced his seizure disorder and he requested disability compensation. He described his seizures as the abrupt loss of consciousness associated with jerking movements of his extremities. His episodes occurred irregularly with a frequency of 2 to 4 a week. On admission to the hospital, medical staff observed several seizurelike spells where the patient assumed a flexed posture of both his upper and lower extremities and then shook them uncontrollably and in an asynchronous fashion. During this ictal period, the patient had normal pupillary and corneal reflexes. His seizures lasted nearly 5 minutes and then immediately resolved without postictal confusion. Postictally, he recalled comments and other environmental events that occurred during his seizurelike episodes. EEGs obtained immediately after an event did not reveal postictal slowing, and prolactin levels obtained 15 minutes after a seizure episode were not significantly elevated over baseline levels. His seizures did not respond to anticonvulsant medications, but they abated after he changed his strategy and began to explore compensation for other reasons.

COURSE AND PROGNOSIS

Most patients with epilepsy have a good prognosis. The majority of seizures can be controlled with anticonvulsant medications sufficiently that the patient can live a productive life. Some seizures, such as absence seizures, tend to disappear by adulthood. For patients whose seizures are medically intractable, epilepsy surgery offers a good alternative (i.e., temporal lobectomy or corpus callosotomy) as long as the focus can be localized or lateralized. In addition, most epilepsy patients will not have psychiatric disorders, and others have psychiatric difficulties only if they endure many years of poorly controlled seizures. For those with behavioral problems, anticonvulsant drugs or epilepsy surgery may relieve some symptoms, such as hyposexuality and aggression, but may not affect the emergence of other symptoms, such as psychosis and suicidal behavior.

TREATMENT

Anticonvulsant Medications In the treatment of psychiatrically disturbed patients with epilepsy, a first consideration is the use of psychoactive anticonvulsant medications. Carbamazepine, valproate, gabapentin (Neurontin), and lamotrigine (Lamictal) have significant antimanic and modest antidepressant properties, probably through mood-stabilization effects. They have some efficacy in the long-term prophylaxis of manic or depressive episodes. Carbamazepine and valproate may also ameliorate some dyscontrolled, aggressive behavior in brain-injured patients. Clonazepam, in addition to its anxiolytic properties, can serve as a supplement to other antimanic therapies. Gabapentin also decreases anxiety and improves general well-being in some epilepsy patients. Both carbamazepine and ethosuximide may have value for borderline personality disorder.

Encephalopathic changes occur at toxic levels of all anticonvulsant drugs. Even at therapeutic levels, barbiturates may need discontinuation because of drug-induced depression, suicidal ideation, sedation, psychomotor slowing, and paradoxical hyperactivity in the very young and the very old. Gabapentin may induce aggressive behavior or hypomania, and vigabatrin (Sabril) may precipitate depression. In addition, clinicians need to be aware of the potential emergence of psychopathology on withdrawal of anticonvulsant medications. Anxiety and depression are the most common emergent symptoms, but psychosis and other behaviors may also occur.

Psychotropic Medications A second consideration is the seizure-threshold lowering effect of psychotropic medications (Table 2.4-10). This is usually not a problem but can occasionally reach clinical significance in poorly controlled epilepsy. Psychotropic drugs are most convulsive when the drug is introduced rapidly and in high doses. Clozapine (Clozaril), for example, has induced seizures in 1 to 4.4 percent of patients, particularly when the dose was rapidly increased. When initiating therapy with psychotropic agents it is best to start low and go slow while monitoring anticonvulsant levels and EEGs.

Potential	Antipsychotic	Antidepressants	Other Psychotropics
High	Chlorpromazine Clozapine	Buproprion Imipramine Maprotiline Amitriptyline Amoxapine Nortriptyline	
Moderate	Most piperazines Thiothixene	Protriptyline Clomipramine	Lithium
Low	Fluphenazine Haloperidol Loxapine Molindone Pimozide Thioridazine Risperidone Olanzapine	Doxepin Desipramine Trazadone Trimipramine Selective serotonin reuptake inhibitors	Ethclorvynol Glutethimide Hydroxyzine Meprobamate Methaqualone

Table 2.4-10 Seizure Threshold Lowering Effect of Psychotropic Medications

Drug Interactions A third treatment consideration is the potential for interaction of anticonvulsant and psychotropic medications (Table 2.4-11). Most commonly, an anticonvulsant drug increases the metabolism of a psychotropic drug with a consequent decrease in its therapeutic efficiency. Conversely, withdrawal of anticonvulsant drugs can precipitate rebound elevations in concentrations of the psychotropic medication. Moreover, the initiation of a psychotropic drug may result in competitive inhibition of anticonvulsant metabolism, with elevations of anticonvulsant levels to toxicity.

Anticonvulsant	Indication*	Effects of Psychotropic Drug on Anticonvulsant Drug**	Effects of Anticonvulsant Drug on Psychotropic Drug**
Carbamazepine	SPS, CPS, GTCS	Potentially decreased	Decreased
Phenytoin	SPS, CPS, GTCS	Potentially decreased or increased, rarely toxic levels	Decreased
Phenobarbital and primidone	SPS, CPS, GTCS	Potentially decreased	Significantly decreased
Valproic acid	CPS, GTCS, absence	Potentially increased, rarely toxic levels	Potentially decreased
Ethosuximide	Absence	None known	None known
Clonazepam	Myoclonic	Potentially decreased	Potentially decreased
Gabapentin	Add-on: CPS, SPS, ± 2nd GTCS	No significant interactions known	
Lamotrigine	Add-on: CPS, SPS, ± 2nd GTCS	No significant interactions known	
Vigabatrin	Add-on: CPS, SPS, ± 2nd GTCS	No significant interactions known	
Tiagabine	Add-on: CPS, SPS, ± 2nd GTCS	No significant interactions known	

* SPS, simple partial seizure; CPS, complex partial seizure; GTCS, generalized tonic-clonic seizure.
** Antipsychotic and antidepressant drugs. Lithium and the minor tranquilizers have few drug interactions with anticonvulsant drugs.

Table 2.4-11 Anticonvulsant-Psychotropic Drug Effects on Blood Levels

Compared to older drugs, the new anticonvulsant medications have fewer potential interactions with psychotropic medications. Gabapentin, lamotrigine, vigabatrin, and tiagabine are relatively free of enzyme-inducing or enzyme-inhibiting properties. Felbamate (Felbatol), however, has been withdrawn in the United States because some patients developed fatal aplastic anemia and liver disease.

Surgery Epilepsy surgery is a fourth treatment consideration and is limited to patients with medically intractable seizures. The main operation involves resection of epileptogenic tissue by removal of 4 to 6 cm of the anterior temporal lobe. Over 80 percent of temporal lobectomy patients experience some reduction in their seizure frequency and over 50 percent are entirely seizure-free. Removal of the amygdala and most of the hippocampus may have postoperative behavioral effects. Some patients have an anomia or a verbal memory deficit after resection of the dominant hemisphere, and patients occasionally develop a transient postoperative depression. Others experience a reduction in preoperative depression and relief of hyposexuality, but patients with epilepsy may continue to develop psychosis, personality changes, and suicidal behavior even long after the temporal lobectomy. Moreover, patients with preoperative psychotic symptoms are at higher risk for a poor surgical outcome and postoperative psychosis.

Less common epilepsy surgeries include resection of extratemporal lesions, removal of the epileptogenic hemisphere, and ligation of the corpus callosum. Corpus callosotomy, which aims to prevent the interhemispheric spread of seizures, results in a unique, transient disconnection syndrome of mutism, apathy, agnosia, apraxia of the nondominant limbs, difficulty naming, and difficulty writing with the nondominant hand.

Seizure Management In treating the neuropsychiatric disorders of epilepsy, a final consideration is altering the seizure management itself. In addition to the occasional behavior alleviated by strict seizure control, allowing seizures under carefully controlled conditions, much like ECT, relieves some cases of periictal psychosis, depression, or other behaviors.

SUGGESTED CROSS-REFERENCES

Most of the specific psychiatric syndromes associated with epilepsy are discussed in more detail in the appropriate sections devoted to them. Personality disorders are discussed in Chapter 24, mood disorders are discussed in Chapter 14, and sexual disorders are discussed in Chapter 19. The rest of the neuropsychiatric sections of Chapter 2 are also pertinent to epilepsy.

SECTION REFERENCES

Ahern GL, Herring AM, Tackenberg J, Seeger JF, Oommen KJ, Labiner DM, Weinand ME: The association of multiple personality and temporolimbic epilepsy. Intracarotid amobarbital test observations. Arch Neurol *50:*1020, 1993.

Alper K, Devinsky O, Perrine K, Vazquez B, Luciano D: Psychiatric classification of nonconversion nonepileptic seizures. Arch Neurol *52:*199, 1995.

*Bear D, Fedio P: Quantitative analysis of interictal behavior in temporal lobe epilepsy. Arch Neurol *34:*454, 1977.

Blumer D, Montouris G, Hermann B: Psychiatric morbidity in seizure patients on a neurodiagnostic monitoring unit. J Neuropsychiatry Clin Neurosci *7:*445, 1995.

Cockerell OC, Moriarty J, Trimble M, Sander JW, Shorvon SD: Acute psychological disorders in patients with epilepsy: A nation-wide study. Epilepsy Res *25:*119, 1996.

Devinsky O, Abramson H, Alper K, FitzGerald LS, Perrine K, Calderon J, Luciano D: Postictal psychosis: A case control series of 20 patients and 150 controls. Epilepsy Res *20:*247, 1995.

Dongier S: Statistical study of clinical and electroencephalographic manifestations of 536 psychotic episodes occurring in 516 epileptics between clinical seizures. Epilepsia *1:*117, 1959.

Gibbs A: Ictal and non-ictal psychiatric disorders in temporal lobe epilepsy. J Nerv Ment Dis *113:*522, 1951.

Gudmundsson G: Epilepsy in Iceland: A clinical and epidemiological investigation. Acta Neurol Scand *25*(Suppl):1, 1966.

Kanemoto K, Kawasaki J, Kawai I: Postictal psychosis: A comparison with acute interictal and chronic psychoses. Epilepsia *37:*551, 1996.

Kanner AM, Stagno S, Kotagal P, Morris HH: Postictal psychiatric events during prolonged video-electroencephalographic monitoring studies. Arch Neurol *53:*258, 1996.

Ketter TA, Malow BA, Flamini R, White SR, Post RM, Theodore WH: Anticonvulsant withdrawal-emergent psychopathology. Neurology *44:*55, 1994.

Krahn LE, Rummans TA, Peterson GC: Psychiatric implications of surgical treatment of epilepsy. Mayo Clin Proc *71:*1201, 1996.

Lancman ME, Asconape JJ, Graves S, Gibson PA: Psychogenic seizures in children: Long-term analysis of 43 cases. J Child Neurol *9:*404, 1994.

Lesser RP: Psychogenic seizures. Neurology *46:*1499, 1996.

*Lindsay J, Ounsted C, Richards P: Long-term outcome in children with temporal lobe seizures: III. Psychiatric aspects in childhood and adult life. Dev Med Child Neurol *21:*630, 1979.

Manchanda R, Schaefer B, McLachlan RS, Blume WT, Wiebe S, Girvin JP, Parrent A, Derry PA: Psychiatric disorders in candidates for surgery for epilepsy. J Neurol Neurosurg Psychiatry *61:*82, 1996.

*Mandelbaum DE, Burack GD: Pre-existing or epilepsy related problems have attributed to AEDs. Antiepileptic drugs. Epilepsia *40:*389, 1999.

Mathews WS, Barabas G: Suicide and epilepsy: A review of the literature. Psychosomatics *22:*515, 1981.

*Mendez MF, Cummings JL, Benson DF: Depression in epilepsy, significance and phenomenology. Arch Neurol *43:*766, 1986.

Mendez MF, Doss RC, Taylor JL: Interictal violence in epilepsy. Relationship to behavior and seizure variables. J Nerv Ment Dis *181:*566, 1993.

Mendez MF, Engebrit B, Doss R, Grau R: The relationship of epileptic auras and psychological attributes. J Neuropsychiatry Clin Neurosci *8:*287, 1996.

Morrell MJ, Guldner GT: Self-reported sexual function and sexual arousability in women with epilepsy. Epilepsia *37:*1204, 1996.

Murialdo G, Galimberti CA, Fonzi S, Manni R, Costelli P, Parodi C, Solinas GP, Amoretti G, Tartara A: Sex hormones and pituitary function in male epileptic patients with altered or normal sexuality. Epilepsia *36:*360, 1995.

Perez MM, Trimble MR: Epileptic psychosis—diagnostic comparison with process schizophrenia. Br J Psychiatry *37:*245, 1980.

Perini GL, Tosin C, Carraro C, Bernasconi GFG, Canevini MP, Canger R, Pellegrini A, Testa G: Interictal personality disorders in temporal lobe epilepsy and juvenile myoclonic epilepsy. J Neurol Neurosurg Psychiatry *61:*601, 1996.

Pollack MH, Scott EL: Gabapentin and lamotrigine: Novel treatments for mood and anxiety disorders. CNS Spectrums *2:*56, 1997.

Pond DA, Bidwell BH: A survey of epilepsy in 14 general practices: II. Social and psychological aspects. Epilepsia *1:*285, 1959–1960.

Rao SM, Devinsky O, Grafman J, Stein M, Usman M, Uhde TW, Theodore WH: Viscosity and social cohesion in temporal lobe epilepsy. J Neurol Neurosurg Psychiatry *55:*149, 1992.

Robertson MM, Channon S, Baker J: Depressive symptomatology in a general hospital sample of outpatients with temporal lobe epilepsy: A controlled study. Epilepsia *35:*771, 1994.

Slater E, Beard A: The schizophrenia-like psychosis of epilepsy: Psychiatric aspects. Br J Psychiatry *109:*95, 1963.

*Smith DB, Treiman DM, Trimble MR: *Neurobehavioral Problems in Epilepsy.* Raven Press, New York, 1991.

Smith PF, Darlington CL: The development of psychosis in epilepsy: A re-examination of the kindling hypothesis. Behav Brain Res *75:*59, 1996.

Swanson SJ, Rao SM, Grafman J, Salazar AM, Kraft J: The relationship between seizure subtype and interictal personality. Results from the Vietnam Head Injury Study. Brain *118:*91, 1995.

*Trimble MR: *The Psychosis of Epilepsy.* Raven Press, New York, 1991.

Williams D: The structure of emotions reflected in epileptic experiences. Brain *79:*29, 1956.

Wong MT, Lumsden J, Fenton GW, Fenwick PB: Electroencephalography, computed tomography and violence ratings of male patients in a maximum-security mental hospital. Acta Psychiatr Scand *90:*97, 1994.

2.5 NEUROPSYCHIATRIC ASPECTS OF TRAUMATIC BRAIN INJURY

RICARDO E. JORGE, M.D., JEFFREY E. MAX, M.B.B.CH. AND ROBERT G. ROBINSON, M.D.

The neuropsychiatric consequences of *traumatic brain injury* (sometimes referred to as TBI) may be divided into disorders that are also seen in patients without brain injury and those that are unique to patients with brain damage. The disorders also seen in patients without brain injury cover the whole spectrum of psychiatric disorders including cognitive, substance abuse, mood, anxiety, psychotic, somatic, and personality disorders. Many of these disorders are included in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as disorders due to a medical condition—traumatic brain injury, in this case (Table 2.5-1). Most of these disorders have not been extensively studied in the population with traumatic brain injury, and much research is still needed in this area. The disorders that are unique to brain injury also cover a wide range of disorders including pathological laughing or crying, apathy, indifference, denial, anosognosia, aprosody, and neglect. Most of these disorders have not been extensively examined in patients with traumatic brain injury.

Delirium due to traumatic brain injury
Amnestic disorder due to traumatic brain injury
Transient
Chronic
Dementia due to traumatic brain injury
Personality change due to traumatic brain injury
Labile type
Disinhibited type
Aggressive type
Apathetic type
Paranoid type
Combined type
Unspecified type
Mood disorder due to traumatic brain injury
With depressive features
With major depressive-like episode
With manic features
With mixed features
Anxiety due to traumatic brain injury
With generalized anxiety
With panic attacks
With obsessive-compulsive symptoms
Psychotic disorder due to traumatic brain injury
With delusions
With hallucinations

Table 2.5-1 DSM-IV Classification of Some Behavioral Syndromes Occurring After Traumatic Brain Injury

HISTORY

The earliest physical evidence of traumatic brain injury due to assault occurred 1 million years ago. A damaged skull from an early hominid found in South Africa showed two posterior fractures that matched with the condylar surfaces of an antelope humerus discovered nearby. The earliest written evidence of brain injuries was found on the Edwin Smith Papyrus, dated 5000 years ago, which contained the first 27 head-injury records. The Hippocratic Corpus included a treatise on head injury with thoughtful comments on skull fractures, delirium, seizures, and coma.

Associations between traumatic brain injury and a variety of neuropsychiatric disorders have been reported in the medical literature for many years. In 1904 Adolf Meyer, for example, identified a number of disorders that he referred to as the *traumatic insanities.* Although he believed that these disorders represented a combination of psychological, social, historical, and biological factors, he suggested that there may be some unique associations between these disorders and specific lesion locations. Studies of war-related head injuries identified the high incidence of psychiatric complications following brain injury. Several of these studies emphasized the importance of frontal lesions in the pathogenesis of behavioral disturbances. The most famous case of frontal lobe injury, however, was that of Phineas Gage, who suffered a penetrating frontal brain injury after an explosion that shot an iron bar through his skull. After the injury, he was described as childish, capricious, inconsiderate, profane, and having poor judgment.

Analysis of large series of cases like the Oxford Collection of Head Injury Records suggest that biological variables such as the extent of brain damage, lesion location, and the presence of posttraumatic epilepsy are important etiological factors in determining the type and duration of psychiatric syndromes.

COMPARATIVE NOSOLOGY

The neurological and neurosurgical literature abounds with clinical descriptions of early and delayed behavioral abnormalities that follow traumatic brain injury. Acute syndromes include confusional states, apathy, agitation, restlessness, irritability, and posttraumatic amnesia. Delayed, often irreversible, consequences of such injury include cognitive disorders (e.g., amnesia, dementia, executive dysfunction), and personality change due to a general medical condition.

The spectrum of psychiatric disorders attributable to traumatic brain injury span almost the entire spectrum of psychiatric disorders. According to DSM-IV, these disorders are categorized as being caused by traumatic brain injury if there is evidence from the history, physical examination, or ancillary studies that the disturbance is the direct physiological effect of brain trauma (Table 2.5-1).

EPIDEMIOLOGY

In the United States the annual incidence of traumatic brain injury can be conservatively estimated as 200 per 100,000 population. Figure 2.5-1 shows the annual incidence of new cases of traumatic brain injury leading either to death or admission to hospital. Most of these injuries occurred among adolescents and young adults, with a second peak occurring among elderly subjects. Approximately 20 percent of hospital admission due to traumatic brain injury were of patients under the age of 15. There was also a significant sex difference—males were two to three times more likely to suffer brain injury than females. African-Americans also had higher rates of such injuries than other groups, a finding that may be explained by increased firearm exposure and higher homicide rates among African-Americans.

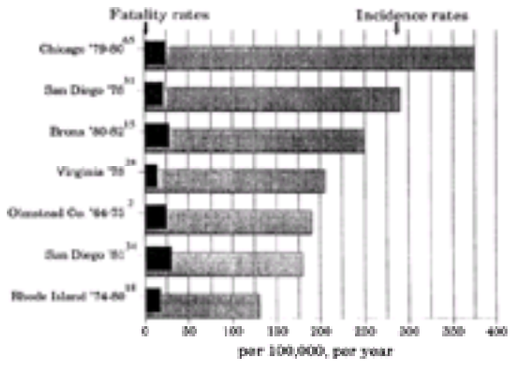

FIGURE 2.5-1 Brain injury incidence and fatality rates from selected U.S. studies. (Reprinted with permission from Kraus JF, McArthur DL: Epidemiology of brain injury. In *Neurology and Trauma*, Evans RW, editor. Saunders, Philadelphia, 1996.)

Low socioeconomic status constitutes another independent risk factor for traumatic brain injury. The single greatest risk factor, however, is substance abuse. Close to one third of brain injury patients had an identifiable alcohol use disorder before trauma, and more than 50 percent were intoxicated at the time of injury. Risk factors for such injury in children include hyperactivity and aggression. In addition, there may be a high rate of preinjury psychiatric disorder in these children. Transport-related cases (i.e., motor vehicle accidents and pedestrians hit by vehicles) are the most important cause of injury, particularly in younger patients. Falls associated with the elderly are the second most prevalent cause of injury. Assaults (especially penetrating injuries involving firearm use) as well as sports- and recreation-related injuries are the next most common causes.

Case-fatality rates vary from 3 to 8 percent among different incidence studies. Since 1979, however, the death rate associated with traumatic brain injury has decreased about 20 percent, probably from multiple causes including legislation, improved safety systems, and enhanced neurosurgical intensive care. Unfortunately, this reduction has been mirrored by an increase in the number of brain-injured survivors who sustain some type of chronic disability.

CLINICAL FEATURES

Acute Behavioral Consequences Head injury encompasses a wide range of severity from patients who die at the moment of trauma to those who do not require medical evaluation or assistance. Most patients admitted to hospital with a head injury diagnosis have a mild injury. A minority of these mildly affected patients will develop acute complications (e.g., brain swelling, intracranial infection) or prolonged postconcussional symptoms. Neuroimaging studies such as computed tomography (CT) and magnetic resonance imaging (MRI) have demonstrated the presence of structural brain lesions in some patients with mild head injury who have not experienced clinical complications.

The most common consequence of head injury is impairment of consciousness, ranging from transient confusion to protracted coma. The Glasgow Coma Scale is commonly used to grade the severity of traumatic brain injury. The scale gives a quantitative estimate of level of consciousness and neurological status based on patterns of eye opening, as well as best verbal and motor responses (Table 2.5-2). Glasgow coma scale scores between 13 and 15 define mild brain injury, scores between 9 and 12 define moderate brain injury, and scores between 3 and 8 define severe injury.

Response	Score*
Eye Opening	
None	1
To pain	2
To sound	3
Spontaneous	4
Motor Response	
None	1
Extension	2
Abnormal flexion	3
Withdrawal	4
Localizes	5
Obey commands	6
Verbal Response	
None	1
Unintelligible sounds	2
Inappropriate words	3
Confused	4
Oriented	5

*Glasgow Coma Scale Score = Eye Opening + Motor Response + Verbal Response (3 to 15).
Adapted from Teasdale G, Jennett B: Assessment of coma and impaired consciousness: A practical scale. Lancet 2:81, 1974.

Table 2.5-2 Glasgow Coma Scale

The early phase of recovery from traumatic brain injury is characterized by disorientation, confusion, and impaired memory function. Apathetic withdrawal, agitation, or severe delirium may also be observed in these patients. *Posttraumatic amnesia* occurs during the period when the patient (who is usually emerging from coma) is disoriented, confused, and has disrupted memory functioning. Deficits are observed in declarative memory (i.e., memory of recent events and times), affecting both anterograde and retrograde processes; procedural memory, in contrast, appears to be relatively spared. Duration of posttraumatic amnesia has been widely used as a measure of severity of the brain injury. It may be assessed using the Galveston Orientation and Amnesia Test (GOAT) or the Children's Orientation and Amnesia Test (COAT), which evaluate orientation to person, place, and time, as well as awareness of the accident and its consequence. Duration of posttraumatic amnesia has proved to be a good predictor of the degree of disability, vocational outcome, and severity of personality change following the injury.

Clinical features of the early phase of recovery from traumatic brain injury are not exclusively characterized by memory impairment. Patients frequently have a decreased level of consciousness and meet DSM-IV criteria for delirium. In addition, patients may present with perceptual disturbances (i.e., illusions or hallucinations), delusional thoughts, psychomotor agitation or retardation, affective lability, and neurovegetative symptoms (e.g., tachycardia, hypertension, diaphoresis, and sleep-wake cycle disruption). Symptoms usually have a sudden onset and a fluctuating course; delirium is most frequently observed in severe cases.

> A 19-year-old man was admitted to a trauma center after a motorcycle accident. He presented with a right epidural hematoma and bilateral contusions in the anterior temporal lobes. He had also a scalp laceration and right maxillary and zygomatic fractures. The postresuscitation Glasgow Coma Scale score was 8. The hematoma was surgically evacuated and the patient was transferred to the intensive care unit. Two days later, the patient became restless and agitated. He removed his intravenous lines and monitoring devices and tried to get out of bed. He was disoriented, incoherent, and aggressive. His behavior suggested that he was experiencing frightening visual hallucinations. A coarse, rapid tremor was noted in both hands and he was diaphoretic and mildly hypertensive. He responded well to a short course of high-potency antipsychotic agents.

Multiple conditions can contribute to the development of delirium in patients with traumatic brain injury. These include structural brain damage, cerebral edema, secondary brain hypoxia, seizures, electrolyte imbalance, infections, and substance (e.g., barbiturates, opioids, steroids, alcohol) withdrawal. Old age, coexistent severe medical disease, polypharmacy, and lesions of the basal ganglia and right hemisphere have also been shown to be significant risk factors.

Chronic Behavioral Consequences

Cognitive Disorders Cognitive disturbances are one of the most important long-term sequelae of severe traumatic brain injury. A 1990 study reported on the cognitive outcome of 127 severely brain-injured patients who were capable of completing serial neuropsychological assessments during a 1-year follow-up period (i.e., excluding those patients with a persistent vegetative state or with very severe intellectual impairment). At 1 year follow-up the brain-injured patients showed slower information processing and greater impairment in memory function compared with a neurologically intact control group. In contrast, linguistic and visuospatial abilities were found to be within the normal range. Patients with mild or moderate head injuries may also show cognitive impairment following brain trauma. These patients complain of lack of concentration and memory deficits during the first weeks following injury; however, spontaneous recovery is the rule for the majority of these patients.

Memory functions are also distinctively impaired in traumatically brain-injured children and adults. Memory deficits are the most frequent chronic cognitive disturbance reported by patients and relatives in traumatic brain injury. Memory dysfunction is characterized by both anterograde and retrograde deficits, faulty sequencing of events, and inefficient encoding and storage strategies. A DSM-IV diagnosis of amnestic disorder due to traumatic brain injury, chronic, may be made for those patients without dementia in whom the memory disturbance causes significant impairment in social or vocational functioning and represents a significant decline with respect to previous levels of performance. However, patients manifesting an isolated memory deficit are rare.

Attentional deficits are among the most frequent neuropsychological symptoms observed in brain-injured patients following resolution of posttraumatic amnesia. Attention consists of multiple processes subserved by interrelated neural networks. Patients may present with restricted verbal or visuospatial attentional span, altered

vigilance patterns (i.e., sustained attention deficits), or slowed information processing. The most consistent findings, however, are associated with performance in the most demanding tasks (e.g., in divided attention paradigms such as the Paced Auditory Serial Addition Task).

Linguistic competence is also frequently affected by traumatic brain injury. Approximately one third of severely brain-injured patients admitted to a rehabilitation facility showed fluent (51 percent), nonfluent (35 percent), or global (14 percent) aphasic syndromes. Aphasia tends to resolve in the majority of cases during the first year following trauma. Anomia, however, constitutes the most prevalent chronic linguistic deficit. Brain-injured patients may also have high-order language alterations and present with a defective narrative discourse, a lack of semantic coherence, aprosody, and impaired pragmatics of communication, all of which result in impoverished and disorganized language and in a reduced communication proficiency.

A prominent defect in control or executive functions has been consistently described in patients who survive severe head injury. Executive functions include goal formation, planning, selection of adequate response patterns, and monitoring of ongoing behavior. When confronted with a demanding environment, the adaptive functioning of brain-injured patients is often impaired. The patient is dismally affected by this lack of initiative, rigid thinking, and faulty problem-solving ability. Several neuropsychological tasks were specifically designed to quantify these deficits. These include the Wisconsin Card Sorting Test, the Goldberg Executive Control Battery, the Tower of London test, and the trail making test. The executive dysfunction observed in patients with brain injury is strongly associated with dysfunction of fronto-subcortical pathways.

In contrast to what happens with memory and control functions, visuospatial and praxic abilities are usually preserved in the long-term following traumatic brain injury, probably because of the relative sparing of posterior association cortices.

Finally, unawareness *(anosognosia)* or denial of deficits is a cognitive disorder frequently observed in brain-injured patients, particularly in those who have suffered extensive frontal lobe damage. This constitutes a severe behavioral sequel that impedes realistic goal-setting and interferes with the rehabilitation process.

DEMENTIA *Dementia* is a syndrome defined in DSM-IV by impairment of memory and at least one other cognitive domain in the absence of an alteration of consciousness. The cognitive defect must have a significant impact on the social and occupational functioning of the involved subject. Dementia due to head trauma is characterized by prominent memory and executive dysfunction with relatively preserved visuospatial, praxic, and primary linguistic functions. In addition, these patients may be severely apathetic and withdrawn and demonstrate markedly slow information-processing. Physical examination may reveal the presence of extrapyramidal signs.

> A 40-year-old white male was severely injured in a motor vehicle accident and experienced protracted coma of 1-month duration. An MRI performed at 2 weeks from injury revealed the presence of widespread diffuse axonal injury. At 6-month follow-up, the patient presented with a mild left hemiparesis, right hemidystonic symptoms, and a left peripheral facial palsy. Neuropsychological testing disclosed substantial memory deficits, frontal lobe dysfunction, and significantly impaired problem-solving ability. Visuospatial and linguistic skills ranked within the lower average range. His hygiene and self-care were poor and he hoarded garbage in his pockets and under his bed. He had frequent bursts of severe aggressive behavior, but overall he remained abulic and withdrawn. Lithium (Eskalith) was effective in controlling his aggressive behavior.

A chronic subdural hematoma in the elderly may present as a progressive dementia. In addition, there is some evidence that previous traumatic episodes constitute a risk factor for the development of Alzheimer's disease.

PERSONALITY CHANGES Traumatic brain injury patients may experience significant personality changes. These patients have been described as irritable, childish, inconsiderate, capricious, anxious, or aggressive. They lack foresight and misjudge the consequences of their actions. Disinhibition is a frequent and striking clinical feature that may lead to antisocial behavior; conversely, patients may become apathetic, abulic, and withdrawn.

Some investigators group these changes into two distinct syndromes: a *pseudodepressed personality syndrome* (personality change due to traumatic brain injury, apathetic type), which is characterized by apathy and blunted affect, and a *pseudopsychopathic personality syndrome* (personality change due to traumatic brain injury, disinhibited type), characterized by disinhibition, egocentricity, and sexual inappropriateness.

DSM-IV defines personality change due to traumatic brain injury as a persistent personality disturbance that represents a change from the individual's previous personality profile (or a deviation of normal development in children) and is attributable to the pathophysiological changes triggered by brain trauma (Table 2.5-1). The disturbance must not occur exclusively during the course of delirium and cannot be diagnosed if dementia is present. In addition, the disturbance must not be better accounted for by another mental disorder (e.g., mood disorder or substance abuse).

> A 42-year-old construction worker fell from the second floor of a new building. He was in coma for 3 days and remained amnestic and disoriented for about 3 weeks. A CT scan showed bilateral orbitofrontal and anterior temporal hemorrhagic contusions.
>
> Six months after the injury he had undergone a significant personality change. He spent most of his day watching television and refused to reinitiate his usual activities. He ate and gained excessive weight. His wife complained of his frequent and often inappropriate sexual demands and stressed his lack of intimacy. He was also easily upset, shouting and making threats when he felt provoked. He was less sensitive to other people's feelings. A trial of carbamazepine (Tegretol) with therapeutic blood concentrations and the maintenance of a numerical record of outbursts resulted in significantly reduced irritability and outbursts.

DSM-IV further categorizes this condition into the following types: labile, disinhibited, aggressive, apathetic, paranoid, combined, and unspecified type. Disinhibition, antisocial conduct, and hypersexuality have been linked to the occurrence of orbitofrontal lesions; apathy has been linked to medial frontal lesions; and aggression and poor impulse control have been associated with anterior temporal lesions.

A common co-occurrence of disorders involves personality change due to traumatic brain injury and attention deficit/hyperactivity disorder in children. Occasionaly, this co-occurrence may be associated with secondary mania. This constellation of symptoms may be related to a neuroanatomical model of disinhibition implicating ventral frontal and temporal lobe structures.

> A 7-year-old male presented to a pediatric brain injury clinic almost 2 years after a severe traumatic brain injury in which he was struck by a car. The injury resulted in encephalomalacia of a portion of the left frontal lobe and ventriculomegaly. The child's development until the injury was apparently normal. After the injury he required a 1-month inpatient rehabilitation. He had trouble sustaining attention and following instructions. Less problematic, although present, were symptoms of distractibility, interrupting others, impulsivity, and disorganization. He was fidgety, had trouble playing quietly, talked excessively, and shifted from one uncompleted activity to another in rapid succession. He was treated with a hydantoin (Dilantin) prophylactically for 7 months. Within 1 month of discontinuing this hydantoin he became violent toward family members and school personnel in response to minimal provocation, mostly when he was tired. This would last for up to 20 minutes and he usually required restraint. These episodes declined in frequency in response to initiation of carbamazepine, but the child remained potentially dangerous. Cognitive testing revealed intelligent quotient (I.Q.) in the borderline range, with deficits in attention, memory, perceptual-motor integration, and processing speed. Complex partial seizures were also noted. He has been followed for 2 years while on 250 to 300 mg of carbamazepine a day. Outbursts continued to be problematic but less frequent for 1 year, after which they no longer occurred. Attention-deficit/hyperactivity disorder symptomatology was dominated by inattention and disorganization with no evidence of impulsivity or hyperactivity any longer.

Depressive Disorders Mood disorders appear to be a frequent psychiatric complication of traumatic brain injury and they may play a relevant role in shaping long-term outcome. The reported frequency of depressive disorders following such injury has varied from 6 to 77 percent. This variability may be due to the lack of uniformity in the psychiatric diagnosis. Review of the literature on this issue reveals that most of the studies had relied on rating scales (e.g., Hamilton Rating Scale for Depression) or a relative's reports rather than on structured interviews and established diagnostic criteria. Studies in the early 1990s, however, have used structured interviews and criteria from the revised third edition of DSM (DSM-III-R) to ascertain the prevalence of mood disorders in this population.

A 1993 study found that 17 out of 66 patients (26 percent) developed major depression while 2 patients (3 percent) developed minor depressive disorders immediately following traumatic brain injury (i.e., approximately 1 month after brain injury). There were also 11 patients who developed a delayed-onset major depressive episode point during a 1-year follow-up period. Thus, 47 percent of patients with follow-up data developed major depressive disorder during the first year after the traumatic

episode. Minor depressive disorder was diagnosed in another 8 patients during the course of the year (Fig. 2.5-2).

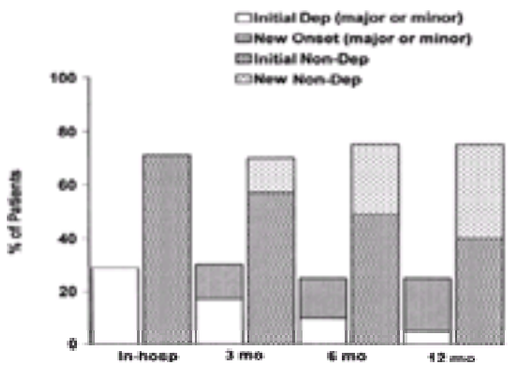

FIGURE 2.5-2 The percent of traumatic brain injury cases (N = 66) who were depressed (major or minor) in the immediate short-term period following injury and at each follow-up time period. The patients depressed initially improved over time but new cases developed so that the overall rate of depression remained relatively stable. (Data from Jorge RE, Robinson RG, Amdt SV, Starkstein SE, Forrester AW, Geisler F: Depression following traumatic brain injury: A 1 year longitudinal study. J Affect Disord *27:*233, 1993.)

When compared with their nondepressed counterparts, patients with major depressive disorder did not show significant differences in demographic variables or in the type or severity of brain injury. There was, however, significantly greater frequency of previous personal history of psychiatric disorders in the group with major depression. In addition, these patients had a significantly poorer premorbid social functioning than patients without depression. Cross-sectional analysis at 3, 6, and 12 months follow-up evaluations showed that poor social functioning was the strongest clinical correlate of major depressive disorder (Fig. 2.5-3). Although patients with depression showed poor premorbid social functioning levels, the fact that the occurrence of major depression was significantly associated with a poor psychosocial outcome suggests that it may negatively influence posttraumatic psychosocial adjustment.

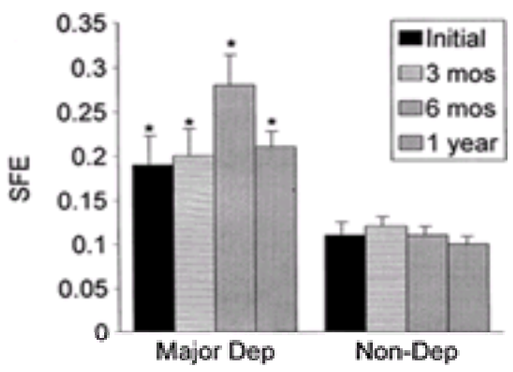

FIGURE 2.5-3 Social function exam (SFE) scores over the first year following brain injury in patients with major depressive disorder or no mood disturbance. Social functioning during the initial evaluation reflected impairment prior to the injury, while the follow-up evaluations reflected social functioning in the month prior to evaluation. Patients with depression consistently had more impairment in social functioning than nondepressed patients. (Data from Jorge RE, Robinson RG, Arndt SV, Starkstein SE, Forrester AW, Geisler F: Depression following traumatic brain injury: A 1 year longitudinal study. J Affect Disord *27:*233, 1993.)

A 42-year-old engineer had a motor vehicle accident when returning from a convention. He suffered multiple injuries, including a diaphragmatic rupture and a left fronto-temporo-parietal subdural hematoma. When admitted to the hospital, the patient was hypotensive and hypoxic. His diaphragm was repaired and the subdural hematoma was evacuated with the urgent intervention of two surgical teams. The patient remained in coma during the following 72 hours. Posttraumatic amnesia lasted for almost 3 weeks. At this point, the neurological exam disclosed a right hemiparesis and a left lateral rectus palsy. The patient was mildly hypophonic and dysarthric. Forty days after the accident he was transferred to a specialized rehabilitation unit. A neuropsychiatric evaluation was completed once his posttraumatic amnesia had cleared. Neuropsychological tests were within normal limits. The patient conveyed a profoundly depressed mood and feelings of hopelessness. He stated that he would never be able to recover, that his career was ruined, and that it would have been better if he had died in the accident. He had no appetite and refused to participate in physical rehabilitation. He also had significant sleep problems. Treatment of depression was initiated with paroxetine (Paxil) at a dosage of 20 mg a day. After 3 weeks, the patient's mood was significantly improved and he became involved in the rehabilitation program. At 6-month follow-up, he was no longer depressed and had returned to work.

There is empirical evidence supporting an association between depression due to traumatic brain injury and specific lesion locations. Studies of war injuries found that several years after penetrating brain injury, depressive symptoms were more common among patients with right than left hemisphere lesions. Depressive symptoms were also more frequent among patients with frontal and parietal lesions than among patients with other lesion locations. Recent studies have reported that lesion location may also play an important role in the development of mood disorders in the immediate posttrauma period. In one series, major depression following traumatic brain injury was associated with the presence of left dorsolateral frontal or left basal ganglia lesions (Fig. 2.5-4).

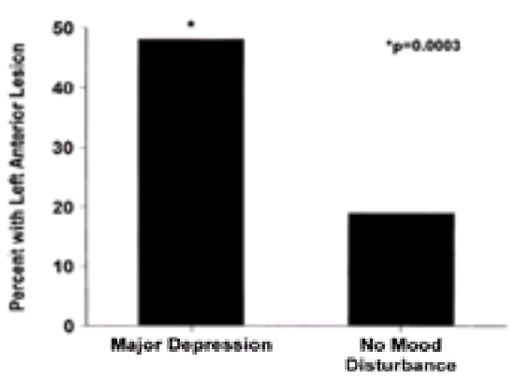

FIGURE 2.5-4 Percent of patients with major depression or no mood disturbance at 1 month following head injury who had evidence on CT scan of injury involving the left dorsolateral frontal cortex or the left basal ganglia. Because patients with traumatic brain injury frequently have multiple areas of injury, a logistic regression analysis was used to examine the independent effects of each area of injury. The strongest effect of lesions on major depression was found in this left anterior brain region (Wald c^2 = 12.97, df = 1, *p = 0.0003). (Data from Fedoroff JP, Starkstein SE, Forrester AW, Geisler F, Jorge RE, Robinson RG, et al: Depression in patients with acute traumatic brain injury. Am J Psychiatry *149:*918, 1992.)

The association between left anterior lesions and major depressive disorder held up only during the period immediately after the injury because delayed-onset depression was not related to the presence of specific lesion locations. Late-onset depressions were associated, however, with a personal history of psychiatric disorder and with poor social functioning levels. Thus, psychosocial factors appear to be more important in the long-term effects of raumatic brain injury.

Secondary Mania Secondary mania (mood disorder due to a general medical condition, with manic features) and hypomania have been reported in a number of

medical conditions such as thyroid disease, uremia, vitamin B_{12} deficiency, or following open heart surgery. Mania has also been associated with traumatic brain injury.

A recent study reported on 20 patients who developed mania after closed head trauma. There was a significant association between mania and the presence of posttraumatic seizures, predominantly of the partial complex type *(temporal lobe epilepsy).* There was no association, however, with family history of bipolar I disorder among first-degree relatives.

A prospective study found that 6 out of 66 patients with traumatic brain injury (9 percent) developed secondary mania at some point during the first year following the injury. Although manic episodes were short-lasting (approximately 2 months), the presence of an expansive mood had a mean duration of 5.7 months. Secondary mania was not related to the type or severity of brain injury, the presence of posttraumatic epilepsy, the degree of physical or intellectual impairment, level of social functioning, or the presence of family or personal history of psychiatric disorder. Secondary mania was, however, associated with the presence of basotemporal lesions (Fig. 2.5-5). This is consistent with the findings in patients with mania secondary to stroke. A small series of secondary mania has been described in consecutively hospitalized children and adolescents. The syndrome was associated with severe brain injury, a family history of mood disorder, and possibly with anterolateral temporal lobe lesions. The development of abnormal electrical activation patterns in limbic networks, functional changes in aminergic inhibitory systems, and the presence of aberrant regeneration patterns may all play a role in the genesis of these syndromes.

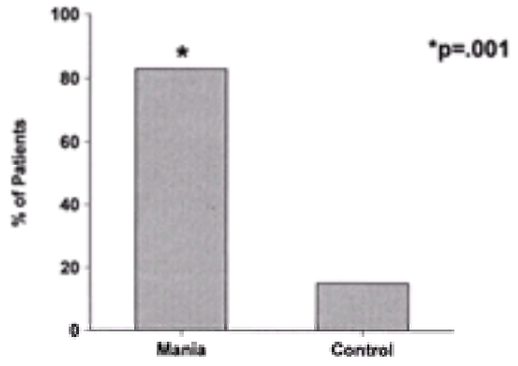

FIGURE 2.5-5 The frequency of basotemporal lesions among patients with traumatic brain injury who developed mania sometime during the first year following stroke compared with patients who did not develop mania. (Data from Jorge RE, Robinson RG, Starkstein SE, Arndt SV, Forrester AW, Geisler FH: Secondary mania following traumatic brain injury. Am J Psychiatry *150:*916, 1993.)

The diagnosis of mania following brain injury is based on DSM-IV criteria for mood disorder due to traumatic brain injury with manic features or with or mixed features (Table 2.5-1). As in the case of the depressive disorders, the mania must be thought to be a direct physiological consequence of the brain injury. The diagnosis only requires that the predominant mood be elevated, euphoric, or irritable. This obviously does not require the same specificity of syndrome as a manic episode or a major depressive episode. The aforementioned prospective study required that patients meet the more rigorous criteria for a manic episode.

A 31-year-old married man with two children was struck in the head when a welding tank exploded, sending large and small pieces of metal in all directions. He suffered an intracerebral bleed in the right basal ganglia and white matter surrounding the anterior horn of the right lateral ventricle. Within a few days he became manic, staying up all night, selling things he owned for ridiculously low prices, and buying a new car he could not afford. He believed that he would become a famous songwriter and was hypersexual. He was admitted to hospital where he was treated successfully with lithium and low-dose haloperidol (Haldol). He remained euthymic for about 6 months and he discontinued his medications. Although he continued to experience significant behavioral changes, manic symptoms did not recur during a 2-year follow-up period.

The differential diagnosis of mania due to traumatic brain injury should include substance-induced mood disorder, psychotic syndrome disorder due to epilepsy, and personality change due to traumatic brain injury. Substance-induced mood disorder occurs following intoxication or withdrawal from drugs and is another important differential diagnosis among patients with traumatic brain injury who frequently have a history of substance abuse.

Anxiety Disorders There is a paucity of studies on the prevalence and clinical correlates of anxiety disorders in the traumatically brain-injured population. The available data, however, suggest that anxiety disorders are a frequent psychiatric complication of traumatic brain injury (Fig. 2.5-6). DSM-IV classifies anxiety disorders due to a medical condition according to the predominance of either generalized anxiety, panic attacks, or obsessive-compulsive symptoms (Table 2.5-1).

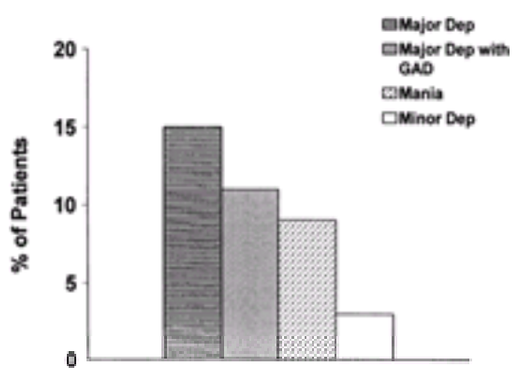

FIGURE 2.5-6 The frequencies of several neuropsychiatric disorders during the hospitalization immediately following traumatic brain injury. Most of the patients had moderate injury (Glasgow Coma Score of 8 to 12) but some had severe injury. Major depressive disorder was the most common disorder particularly when it is combined with group having generalized anxiety disorder plus major depression. (Data from Robinson RG, Jorge RE: Mood disorders. In *Neuropsychiatry of Traumatic Brain Injury.* Silver JM, Yudofsky SC, editors. American Psychiatric Press, Washington, DC, 1994.)

Anxiety disorders following traumatic brain injury may manifest themselves in a variety of ways. Pathological worrying, free-floating anxiety, and autonomic symptoms are commonly associated with anxious depressions. Rigidity of thinking, obsessions, and ritualistic compulsions may also be observed following brain trauma. The onset of obsessive-compulsive symptoms, however, is usually preceded by generalized anxiety. There are also case reports of patients experiencing panic attacks after traumatic brain injury. Posttraumatic stress disorders are characterized by recurrent intrusive recollections, distressing dreams, and flashbacks of the traumatic event. Patients with posttraumatic stress disorder show features of increased arousal such as difficulty falling or staying asleep, exaggerated startle response, and irritability. They lose interest in social activities and avoid thoughts, feelings, or circumstances associated with the event.

A 33-year-old certified nurse was assaulted while walking to her job in a small hospital. She received several blows to her head and was admitted unconscious to the emergency room. A CT scan showed a depressed skull fracture and right temporal and parietal contusions. She regained consciousness during the first 12 hours of admission and was confused and disoriented for the following 3 days. After this, she had an uneventful course and a good recovery. One month after trauma she received a complete neuropsychiatric evaluation. She complained of intrusive recollections of the moments previous to the assault, which were associated with great psychological distress, sweating, and palpitations. She was also disturbed by nightmares in which she suddenly found herself in an intensive care unit connected to a ventilator, incapable of talking or asking for help. She avoided the hospital's surroundings and was unable to walk alone. She also complained of decreased concentration and memory impairments. The results of her neuropsychological tests, however, were within the normal range.

A prospective study of 66 patients with traumatic brain injury found that 7 patients (11 percent) met DSM-III-R criteria for both generalized anxiety disorder and major depression whereas 10 patients met criteria for major depression alone, but none had generalized anxiety disorder alone. The duration of the generalized anxiety disorder was 1.5 months on average, but the mean duration of depression associated with generalized anxiety disorder was 7.5 months whereas depression alone had a mean duration of 1.5 months (Fig. 2.5-7). The presence of anxiety disorder was not associated with any background characteristic; severity of head injury; or severity of social, physical, or cognitive impairment. Anxiety disorder plus depression, however, was significantly associated with right hemisphere lesions while depression alone was associated with left anterior lesions. In addition, a study of veterans of the Vietnam war with penetrating brain injuries found that anxiety symptoms were associated with the presence of right orbitofrontal cortex lesions. These findings suggest that anxiety disorder due to traumatic brain injury is often accompanied by depression and that the mechanism of anxiety may be different than that of depression alone. Anxiety may be mediated by pathophysiological changes provoked by right hemisphere lesions.

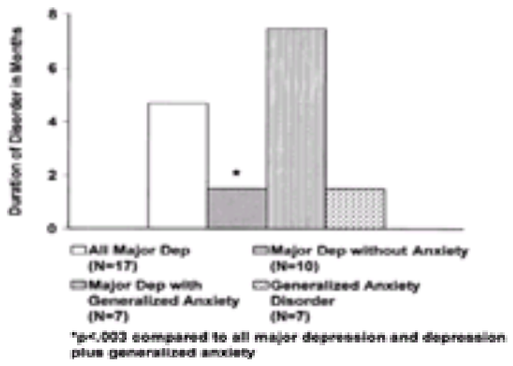

FIGURE 2.5-7 Duration of depression among patients with major depressive disorder or generalized anxiety disorder following traumatic brain injury. Patients with major depressive disorder without anxiety disorder had depressions significantly shorter in duration than patients with major depression and generalized anxiety disorder. Patients with anxiety disorder alone had relatively brief anxiety disorders. (Data from Jorge RE, Robinson RG, Starkstein SE, Arndt SV: Depression and anxiety following traumatic brain injury. J Neuropsychiatry Clin Neurosci *5:*43, 1993.)

Psychotic Disorders Psychotic disorders due to traumatic brain injury are often associated with seizure disorder. In addition, the epileptic foci are often located in limbic or paralimbic cortical regions. Psychotic episodes may be temporally linked to seizures or may have a more prolonged interictal course. In the latter case, the clinical picture is characterized by the presence of partial seizures.

Electroencephalographic and functional neuroradiological studies (e.g., single photon emission computed tomography [SPECT] and positron emission tomography [PET]) will usually define ictal and interictal disturbances. Clinical manifestations include personality changes, mood instability, paranoid ideation, poor control over aggression, disinhibited behavior, and hypersexuality. However, these patients lack the pervasive alteration of mood that characterizes major depressive episodes or manic episodes.

PATHOLOGY

The first level of categorization of head injuries divides them into *closed* or *penetrating* injuries depending upon the integrity of the meningeal covering. Missile wounds are the most frequent cause of penetrating brain injuries. They tend to produce discrete lesions that may be complicated by infection or hemorrhage. On the other hand, motor vehicle accidents are the most frequent cause of closed trauma, which represents the majority of traumatic brain injuries. Diffuse brain damage is a prominent feature of closed head injury (Table 2.5-3).

1. Focal Lesions
Contusions and lacerations
Intracerebral hemorrhage
Intracranial-extracerebral hemorrhage (epidural and subdural hematomas, subarachnoid hemorrhage)
Focal ischemic lesions
2. Diffuse Lesions
Diffuse axonal injury
Diffuse ischemic damage

Table 2.5-3 Neuropathological Classification of Traumatic Brain Injury Lesions

Primary brain damage is produced by contact and inertial forces that occur at the time of injury. Contact forces may result in laceration to the scalp, skull fractures, intracranial hemorrhages, contusions, and intracerebral hemorrhages. Inertial loading consists of acceleration on deceleration and rotational forces, which result in diffuse axonal injury and eventually in acute subdural hematoma from the tearing of subdural bridging veins (Figure 2.5-8).

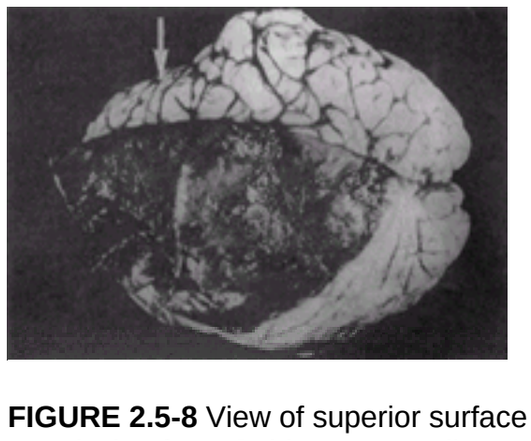

FIGURE 2.5-8 View of superior surface of brain showing subdural hematoma (dura folded back) and compression of right frontal region (*arrow*). (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

Secondary brain damage is produced by pathological processes that are initiated at the moment of injury but span a variable period following the traumatic episode. These include brain damage secondary to ischemia (e.g., that resulting from associated hypotension or hypoxia), brain swelling, raised intracranial pressure (with consequent reduction of cerebral perfusion pressure), and infection.

Focal lesions consist of contusions and lacerations that usually occur at the surface of the brain (Table 2.5-3). They are more prominent at the crest of cerebral gyri

and have a predilection for the frontal and temporal poles. Lacerations are usually accompanied by extracerebral hemorrhages (burst lobe). Intracranial-extracerebral hemorrhages occur in the epidural space (i.e., between the skull and the dura), the subdural (i.e., between the dura and the arachnoid), or the subarachnoid space. Intracerebral hemorrhages are often multiple, involving the frontal and temporal lobes and basal ganglia and may have a delayed onset (i.e., hours or days after trauma). Finally, ischemic damage may produce focal lesions from traumatic vascular lesions (e.g., arterial dissection, vascular distortion and compression, and venous thrombosis) or from arterial vasospasm.

Diffuse lesions include diffuse axonal injury that occurs preferentially within the corpus callosum, thalamus, and dorsolateral quadrants of the upper brainstem. Pathological processes include fragmentation of the axolemma, axonal transport disruption, axonal bulb formation, astrogliosis, and microglial activation. Although usually associated with immediate and persistent coma, lesser degrees of axonal disruption may also be seen in patients with a lucid interval or even in patients who suffered mild brain injuries. Diffuse injury also includes ischemic damage, which is highly prevalent among patients with severe head injuries. A seminal neuropathological study reported that severe ischemic damage was present in 27 percent of 151 cases, moderately severe ischemic damage was present in 43 percent, and mild damage was found in another 30 percent of cases.

The influence of these different patterns of injury on the psychiatric disorders following brain trauma has not been extensively studied. Previous studies have stressed the importance of diffuse brain injury as a predictor of long-term disability of such patients. Another study, however, did not find significant differences in neuropsychological outcome between two groups of patients with focal or diffuse patterns of injury. Furthermore, both focal and diffuse lesions usually coexist in traumatically brain-injured patients. A recent study suggests that adaptive dysfunction in children is significantly correlated with the depth of the lesion, a finding previously observed in other age groups.

LABORATORY TESTS

Neurochemical Changes Since the late 1980s there has been intensive research on the complex neurochemical changes occurring after brain trauma. This was undertaken in an attempt to identify specific interventions that could modify the pathological processes resulting in diffuse neuronal death. For example, increased activity of basal forebrain cholinergic systems has been identified in patients with traumatic brain injury. Pathological excitation of basal forebrain nuclei following acetylcholine release, however, may result in structural damage of these nuclei and persistent behavioral deficits. Furthermore, such brain injury activates the sympathoadrenomedullary axis as well as the ascending cathecholaminergic pathways. Circulating levels of catecholamines have been shown to significantly correlate with severity of injury as measured by Glasgow Coma Scale scores (Fig. 2.5-9).

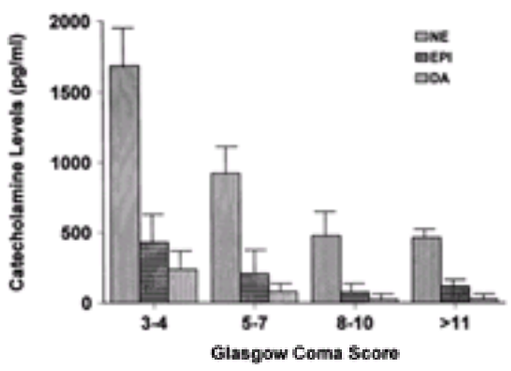

FIGURE 2.5-9 Mean ± SEM catecholamine concentrations within 48 hours of head injury in patients grouped by initial Glasgow Coma Scale score. (Reprinted with permission from Hamill RW, Woolf PD, McDonald JV, Lee L, Kelly M: Cathecholamines predict outcome in traumatic brain injury. Ann Neurol *21:*438, 1987.)

The neurotoxic effects of the excitatory amino acids glutamate and aspartate have been studied at length. Overactivity of these neurotransmitter systems plays a role in neuronal damage associated with diverse types of injury. These neurotoxic effects are mediated through changes in the membrane conductance of sodium and calcium ions. Traumatic injury is associated with the formation of highly reactive oxygen free radicals, which in turn effect the peroxidation of cell membranes, cytosolic proteins, and nucleic acids. There is also evidence of activation of complement proteins as well as different cytokines such as interleukin-1b and tumor necrosis factor.

Patients with traumatic brain injury may also present with structural and functional changes of the hypothalamic-pituitary axis. Hormonal responses to trauma include increases in adrenocorticotropin, cortisol, growth hormone, and prolactin concentrations. In contrast, gonadotropins, sex steroid hormones, and thyroid hormone concentrations decrease following trauma. The presence of elevated prolactin levels is strongly associated with hypothalamic damage and it has been hypothesized that the changes in adrenocorticotropic hormone (ACTH) and cortisol concentrations that occur in TBI patients may mediate changes in mood. The magnitude, regional distribution, and relation of the temporal course of these neurochemical changes to mood disorders remains largely unexplored and constitutes an interesting field for further investigation.

Neuroimaging Recent advances in neuroimaging techniques have certainly increased the understanding of the relationship between structural abnormalities and behavioral disturbances following traumatic brain injury. CT and MRI are routinely used for the evaluation of these patients. CT is still the most efficient means of detecting surgically treatable hematomas and is the study of choice for evaluating patients with rapid changes in their neurological status. Figure 2.5-10 shows the sites of most common injury as seen on neuroimaging. MRI, however, is more sensitive in detecting the more prevalent posttraumatic nonhemorrhagic lesions (e.g., cortical contusions and deep white matter lesions) and in identifying small subdural collections. A delayed MRI scan (i.e., 2 weeks after injury) would be indicated in those patients whose initial instability favored an emergency CT scan but were later treated medically or were left with persistent neurological deficits after a neurosurgical procedure. There is also some evidence that functionally significant lesions may be better evaluated by late MRI scanning. Posttraumatic atrophic changes can be quantified using MRI volumetric studies; for example, changes in left hippocampal volume may be indicative of verbal memory functioning.

FIGURE 2.5-10 Area of maximum shear strain and expected distribution of traumatic intra-axial lesions for the three orthogonal planes of rotational acceleration as predicted by Holbourn. The expected location of traumatic lesions are shown for rotation in the sagittal **(A)**, coronal **(B)**, and axial **(C)** planes. A greater frequency of lesions is to be anticipated in the more darkly shaded regions, which represent areas of maximum shear strain. (Reprinted with permission from Gentry LR, Godersky JC: Head trauma. In *MRI of the Brain and Spine*, SW Atlas, editor. Raven, New York, 1991.)

Functional neuroimaging studies, for example, PET and SPECT, provide additional information with regard to the metabolic rates of cortical and subcortical structures as well as of the status of specific neurotransmitter systems. Both PET and SPECT have greater sensitivity than CT and MRI in detecting local abnormalities in brain-injured patients with long-term deficits. The correlations between these radiological abnormalities and specific neurobehavioral deficits, however, have generally

been weak and require further study.

Electrophysiological Studies Conventional electroencephalogram (EEG) recordings are currently used in trauma intensive care units for monitoring procedures and to diagnose brain death. The EEG is also invaluable in the diagnosis of status epilepticus and is the primary diagnostic procedure to localize a posttraumatic epileptic focus. The chance of finding an abnormality on routine EEG can be increased by using nasopharyngeal, anterior temporal, or sphenoidal leads. Video-EEG monitoring and 24-hour ambulatory recordings may be useful in the differential diagnosis of patients presenting with unclear paroxysmal behavioral disturbances. Polysomnography permits the diagnosis of atypical sleep disturbances that may occur in brain-injured patients. These include atypical night terrors, sleep apnea, nocturnal myoclonus, and restless leg syndrome. Visual, auditory, and somatosensory evoked potentials are useful to localize the level of injury to the central nervous system; auditory evoked potentials are particularly effective in detecting brainstem pathology.

Finally, quantitative EEG is currently used as an adjunctive diagnostic technique in the evaluation of slow-wave abnormalities associated with brain injuries and in the diagnosis of posttraumatic temporal lobe epilepsy.

COURSE AND PROGNOSIS

Course and prognosis of patients with traumatic brain injury involves the longitudinal analysis of neurological, neuropsychological, psychiatric, and psychosocial variables. Some of these variables are readily operationally defined, while others are more elusive and difficult to quantify. The Glasgow Outcome Scale has been widely used as a measure of the long-term outcome of these patients. It consists of five levels of outcome: (1) death, (2) persistent vegetative state, (3) severe disability (conscious but dependent in activities of daily living, (4) moderate disability (disabled but living independently), and (5) good recovery (mild neuropsychiatric sequela but able to resume an otherwise normal life). Although crude, the scale is appreciated for both its validity and high reproducibility.

The long-term outcome of patients with brain injury is primarily related to the severity of the injury, the type and location of the intracranial lesion, and the efficacy of the immediate medical and surgical treatment. Outcome is also influenced by concurrent factors that include age, socioeconomic status, educational level, previous psychiatric disorders (e.g., history of alcohol or drug abuse, personality disorders, etc.), and premorbid social functioning levels. Finally, the quality and extent of rehabilitation services and the availability of social and vocational support also play a significant role in outcome.

The National Institutes of Health (NIH) Traumatic Coma Data Bank was initiated by the National Institute of Neurological Disorders and Stroke in order to characterize the natural history of traumatic head injury and to evaluate the determinants of recovery. Of 746 severely head-injured patients studied in this cohort, 243 patients (32.5 percent) died, 325 patients (42 percent) were either severely disabled or in a vegetative state, 138 (18 percent) had moderate disability, and the remaining 50 (7 percent) had a good recovery as measured by the Glasgow Outcome Scale at the time of hospital discharge. Another study analyzed the outcome of 170 patients with moderate head injuries at 3 months following traumatic brain injury. According to the Glasgow Outcome Scale, 38 percent of patients with moderate head injury made a good recovery, 49 percent were left with moderate disability, and 10 percent of patients had severe disability; the mortality rate was 3 percent. The same authors found that the majority of patients with mild head injury (78 percent) made a good recovery, 22 percent of patients were left with moderate disability, and there were no patients experiencing severe disability.

Cognitive disturbances are among the most important factors in long-term disability following severe traumatic brain injury. A 1992 study reported on a group of 50 patients with mild and moderate head injury and correlated different neuropsychological measures obtained in hospital and at 1 and 3 months follow-up with serial MRI findings. The group with traumatic brain injury had significantly greater neuropsychological impairment than a control group. These cognitive deficits showed an impressive recovery during follow-up and were not consistently related to specific lesion locations. MRI abnormalities were present at baseline in 40 (80 percent) of these patients and gradually resolved by 3 months. Other studies have reported that 18 percent of patients with severe head injury had returned to gainful employment and 62 percent of former students had returned to school by 6 months follow-up. For those not back to work or school at 6 months, 31 percent of the former workers and 66 percent of the former students had returned by 12 months. The three most significant predictors for returning to work or school were age, intact verbal abilities, and speed of information processing. A series of moderate head injury cases revealed that 69 percent of patients who had been gainfully employed before the injury were unemployed at 3 months' follow-up.

The NIH traumatic coma data bank found that outcome following severe brain injury in children and adolescents was poorest in the 0 to 5 age range, followed by the 11 to 15 age range, and the 6 to 10 age range. These findings suggest that younger children are at greater risk following the relatively diffuse type of injury associated with traumatic brain injury, which runs counter to the findings of relative sparing of function (plasticity) following more focal childhood brain lesions. Factors predictive of the development of a new psychiatric disorder in children during the first 2 years after the injury include increasing severity of injury, preinjury family dysfunction, preinjury psychiatric disorder in the patient and in family members, and lower socioeconomic status.

Furthermore, behavioral changes associated with brain injury may disrupt interpersonal relationships and pose a great burden to family members. During the first year after the injury, more than two thirds of relatives experienced moderate to severe degrees of burden as a consequence of the behavioral changes in the family member. Significant predictors of family functioning outcome at various points during the first 2 years after childhood brain injury include preinjury family functioning and the development of a new psychiatric disorder in the child. Thus, it is not surprising that psychosocial adjustment and community reentry have become the targets of rehabilitation efforts. The patient's motivation and a history of alcohol or drug abuse as well as awareness of cognitive and physical impairments have been shown to exert a significant effect on rehabilitation outcome.

TREATMENT

Biological Treatment Patients with brain injury are more sensitive than normals to the side effects of medications, especially psychotropic agents. Dosages of these drugs must be prudently increased so as to minimize the adverse effects (i.e., start low, go slow). However, the patient must receive an adequate therapeutic trial with regard to dosage and duration of treatment. Brain-injured patients must be reassessed frequently in order to determine changes in treatment schedules. Special care must also be taken in monitoring drug interactions. In addition, if there is evidence of a partial response to a specific medication, augmentation therapy is warranted, considering mechanisms of action and adverse effects of the second drug.

Agitation and violent behavior may also occur during the early stage of recovery (e.g., when the patient is emerging from coma) and may require the use of sedative medications. Low doses of high-potency antipsychotic medications (e.g., haloperidol) may be useful in this context, particularly when the agitated behavior stems from psychotic ideation or hallucinatory experiences. Besides their neurological and cardiovascular adverse effects, there is experimental evidence that these drugs may be detrimental for recovery; thus, they must be promptly tapered as soon as the patient stabilizes.

Benzodiazepines [e.g., lorazepam (Ativan)] are also helpful in the management of acute agitation and aggression. Oversedation, poor coordination, and cognitive disturbances are their most common adverse effects; paradoxical rage is a rare complication. Dosages should be progressively reduced as symptomatic relief is obtained. There are also reports on the use of carbamazepine, amantadine (Symmetrel), and clonidine (Catapres) for the treatment of acute agitated states.

There is also some evidence of the clinical efficacy of dopamine agonists (e.g., bromocriptine [Parlodel]) or in patients with poor motivation and frontal lobe deficits. Chronic aggressive disorders may be treated with anticonvulsants (e.g., carbamazepine or valproic acid [Depakene], b-adrenergic receptor antagonists (e.g., propranolol [Inderal] or pindolol [Visken]), antidepressants (e.g., amitriptyline [Elavil] or sertraline [Zoloft]), or lithium.

There is only one small double-blind, placebo controlled study of the efficacy of pharmacological treatments of depression in traumatically brain-injured patients. In this study the antidepressant effectiveness of desipramine (Norpramin) was significantly greater than placebo. Selection among competing antidepressants is usually guided by their adverse effect. Mild anticholinergic activity, minimal lowering of seizure threshold, and low sedative effects are the most important factors to be considered in the choice of an antidepressant drug in this population. Tricyclic drugs have important anticholinergic effects that may interfere with cognitive and memory functions; in addition, they may lower the seizure threshold. Selective serotonin reuptake inhibitors (SSRIs) are antidepressants that appear to have milder adverse effects. The most common adverse effects include headache, gastrointestinal complaints, and insomnia. Diminished libido and sexual dysfunction may be an additional concern. Fluoxetine (Prozac) (starting at 10 mg/day), sertraline (starting at 25mg/day), or paroxetine (starting at 5 to 10 mg/day) are among the most useful drugs in this group.

Trazodone (Desyrel) is an antidepressant that also inhibits serotonin reuptake. Treatment is started at low doses (50 to 100 mg) at bedtime following a snack; the dosage may be gradually increased every 3 to 4 days up to 400 mg. The most troublesome adverse effects are sedation and orthostatic hypotension.

There have been no systematic studies of the treatment of secondary mania. There are, however, several reports of potentially useful treatment modalities. A double-blind, placebo-controlled study in a single patient with secondary mania following traumatic brain injury concluded that clonidine (600 µg/day) was effective in reverting manic symptoms, carbamazepine (1200 mg/day) did not elicit mood changes and levodopa (Dopar) and benserazide (375 mg/day) resulted in an increase of manic symptoms. Lithium, carbamazepine, and valproate therapies have also been reported to be efficacious in individual cases. Lithium has been reported to impair cognitive performance in traumatically brain-injured patients; in addition, it may lower seizure threshold. Some authors limit its use to patients in whom bipolar disorder preceded the onset of the injury.

Carbamazepine is a mood-stabilizing drug that should be gradually increased to obtain therapeutic blood levels (8 to 12 µg/ml). Complete blood counts should be obtained every 2 weeks for the first 2 months of therapy and every 3 months thereafter. Liver function tests should be obtained every 3 months. Frequent adverse effects include sedation, dry mouth, gastrointestinal upset, drowsiness, impaired concentration, ataxia, nystagmus, and rash. Severe complications include pancytopenia, aplastic anemia, and cholestatic jaundice.

Valproic acid is another mood-stabilizing medication that can be progressively increased from 500 mg a day up to the dosage necessary to obtain plasma concentrations between 50 and 100 µg per ml. The maximum recommended dosage is 60 mg per kg daily divided into two to four doses. Valproic acid has potentially serious adverse effects. These include hepatotoxicity, which ranges from a discrete elevation of transaminases and serum ammonia levels to irreversible liver failure; hemorrhagic pancreatitis has also been reported. Less serious adverse effects include drowsiness, tremor, gastritis, and increased weight. Liver function tests and serum amylase levels should be monitored.

For the treatment of anxiety disorders, buspirone (BuSpar), a drug that has an agonist effect on serotonin (5-hydroxytryptamine [5-HT]) subtype 1A (5-HT $_{1A}$) receptors and an antagonist effect on dopamine type 2 (D_2) receptors, has proved to a safe and efficacious anxiolytic agent. The initial dosage is 15 mg a day given in three divided doses and it may gradually be increased (5 mg every 4 days) to up 50 mg a day. The most common adverse effects are dizziness and headaches.

Pathological laughing and crying may respond to treatment with antidepressant medications. There is, however, a great variability in treatment response among brain-injured patients with some showing a rapid response at relatively low doses and others requiring more time and a full dose schedule. Furthermore, there are case reports of successful treatment of apathetic states with psychostimulants. These include dextroamphetamine (Dexedrine) (8 to 60 mg/day), methylphenidate (Ritalin) (10 to 60 mg/day), and pemoline (Cylert) (56.25 to 75 mg/day). They are given twice a day, with the last dose given at least 6 hours before sleep in order to prevent initial insomnia. Treatment is begun at lower dosages, which are then gradually increased. Patients taking stimulants need close medical supervision to prevent abuse or toxic effects. The most common adverse effects are anxiety, dysphoria, headaches, irritability, anorexia, insomnia, cardiovascular symptoms, dyskinesias, or even psychotic symptoms.

Electroconvulsant therapy (ECT) is not contraindicated in brain-injured patients and may be considered if other methods of treatment prove to be unsuccessful. ECT should be administered with the lowest possible effective energy, using pulsatile currents, with an interval of 2 to 5 days between treatments and a reduced number of treatments for a complete course (i.e., 4 to 6 treatments). Nondominant unilateral ECT is preferred. Even transcranial magnetic stimulation has been reported to be successful in four patients with major depression developing after brain injury.

Behavioral Treatment Behavioral deficits in self-care habits (e.g., feeding or personal hygiene), interpersonal skills (e.g., disinhibited behavior), problem solving, or response to environmental stress may be amenable to behavioral intervention. Behavioral rehabilitation programs shape behavior based on the principles of operant conditioning (e.g., contingency contracts and token economies). Their goal is to increase the patient's repertoire of social and independent living skills, generalizing their use from the rehabilitation environment to the more demanding conditions of community life.

Controlled treatment trials of psychotherapy in any of the neuropsychiatric disorders resulting from brain injury have not been conducted. There are, however, reports of the utility of cognitive behavioral therapy, group therapy, and family therapy in the treatment of depression or other neuropsychiatric disorders. The utility of specific psychological interventions for each neuropsychiatric disorder needs to be developed.

SUGGESTED CROSS-REFERENCES

Basic neurological issues are discussed in Section 1.1. Neuroimaging is covered in Section 1.15 and Section 1.16. Neuropsychological tests used to evaluate neurological and psychiatric patients are described in Section 7.4 and Section 7.5.

Delirium, dementia, and amnestic disorders are covered in Chapter 10. The neuropsychiatric complications of stroke and epilepsy are discussed in Section 2.2 and Section 2.4, respectively.

SECTION REFERENCES

Baker Price LA, Persinger MA: Weak but complex pulsed magnetic fields may reduce depression following traumatic brain injury. Percept Mot Skills *83:*491, 1996.

Bigler ED, Blatter DD, Anderson CV, Johnson SC, Gale SD Hopkins RO, Burnett B: Hippocampal volume in normal aging and traumatic brain injury. AJNR Am J Neuroradiol *18:*11, 1997.

Chestnut RM: Secondary brain insults after head injury: Clinical perspectives. New Horizons *3:*366, 1995.

*Deb S, Lyons I, Koutzoukis C, Ali I, McCarthy G: Rate of psychiatric illness 1 year after traumatic brain injury. Am J Psychiatry *156:*374, 1999.

Grafman J, Vance SC, Swingartner H: The effects of lateralized frontal lesions on mood regulation. Brain *109:*1127, 1986.

*Graham DI, McIntosh TK: Neuropathology of brain injury. In *Neurology and Trauma*, RW Evans, editor. WB Saunders, Philadelphia, 1996.

Gualtieri CT: Pharmacotherapy and the neurobehavioral sequelae of traumatic brain injury. Brain Injury *2:*101, 1988.

Hamill RW, Woolf PD, McDonald JV, Lee L, Kelly M: Catecholamines predict outcome in traumatic brain injury. Ann Neurol *21:*438, 1987.

Hibbard MR, Uysal S, Kepler K, Bogdany J, Silver J. Axis I psychopathology in individuals with traumatic brain injury. J Head Trauma Rehabil *13:*24, 1998.

Jennett B, Bond M: Assessment of outcome after severe brain damage: A practical scale. Lancet *1:*480, 1975.

Jorge RE, Robinson RG, Arndt SV, Starkstein SE, Forrester AW, Geisler F: Depression following traumatic brain injury: A 1-year longitudinal study. J Affect Disord *27:*233, 1993.

Jorge RE, Robinson RG, Starkstein SE, Arndt SV: Depression and anxiety following traumatic brain injury. J Neuropsychiatry Clin Neurosci *5:*43, 1993.

Jorge RE, Robinson RG, Starkstein SE, Arndt SV, Forrester AW, Geisler FH: Secondary mania following traumatic brain injury. Am J Psychiatry *150:*916, 1993.

Katz DI: Neuropathology and neurobehavioral recovery from closed head injury. Head Trauma Rehab *7:*1, 1992.

Kraus JF, McArthur DL: Epidemiology of brain injury. In *Neurology and Trauma*, RW Evans, editor. WB Saunders, Philadelphia, 1996.

Levin HS, Gary HE Jr, Eisenberg HM, Ruff RM, Barth JT, Kreutzer J, High WM Jr, Portman S, Foulkes MA, Jane JA: Neurobehavioral outcome 1 year after severe head injury. Experience of the Traumatic Coma Data Bank. J Neurosurg *73:*699, 1990.

Levin HS, Williams DH, Eisenberg HM, High WMJ, Guinto FCJ: Serial MRI and neurobehavioral findings after mild to moderate closed head injury. J Neurol Neurosurg Psychiatry *55:*255, 1992.

*Levin HS, Aldrich EF, Saydjari C, Eisenberg HM, Foulkes MA, Bellefleur M, Luerssen TG, Jane JA, Marmarou A, Marshall LF: Severe head injury in children: Experience of the Traumatic Coma Data Bank. Neurosurgery *31:*435, 1992.

*Lishman WA: Brain damage in relation to psychiatric disability after head injury. Br J Psychiatry *114:*373, 1968.

Max JE, Smith WL, Sato Y: Traumatic brain injury in children and adolescents: Psychiatric disorders in the first three months. J Am Acad Child Adolesc Psychiatry *36:*96, 1997.

Max JE, Robin DA, Lindgren SD, Ardnt S: Traumatic brain injury in children and adolescents: Psychiatric disorders at two years. J Am Acad Child Adolesc Psychiatry *36:*1278, 1997.

Max JE, Castillo CS, Lindgren SD, Ardnt S: The Neuropsychiatric Rating Schedule: reliability and validity. J Am Acad Child Adolesc Psychiatry *37:*297, 1998.

Prigatano GP: Psychiatric aspects of head injury: Problem areas and suggested guidelines for research. In *Neurobehavioral Recovery from Head Injury.* HS Levin, J Grafman, HM Eisenberg, editors. Oxford University Press, Oxford, England, 1987.

Rimel RW, Giordani B, Barth JT, Jane JA: Moderate head injury: Completing the clinical spectrum of brain trauma. Neurosurgery *11:*344, 1982.

Rosenthal M, Christensen BK, Ross TP: Depression following traumatic brain injury. Arch Phys Med Rehabil *79:*90, 1998.

Ruff RM, Marshall LF, Crouch J, Klauber MR, Levin HS, Barth J, Kreutzer J, Blunt BA, Foulkes MA, Eisenberg HM: Predictors of outcome following severe head trauma: Follow-up data of the Traumatic Coma Data Bank. Brain Injury *7:*101, 1993.

Shukla S, Cook BL, Mukherjee S, Godwin C, Miller MG: Mania following head trauma. Am J Psychiatry *144:*93, 1987.

*Silver JM, Yudofsky SC: Psychopharmacology. In *Neuropsychiatry of Traumatic Brain Injury*, JM Silver, SC Yudofsky, editors. American Psychiatric Press, Washington, DC, 1994.

Starkstein SE, Robinson RG: Mechanism of disinhibition after brain lesions. J Nerv Ment Dis *185:*108, 1997.

Teasdale G, Jennett B: Assessment of coma and impaired consciousness: A practical scale. Lancet *2:*81, 1974.

Van Hoesen GW: Ventromedial temporal lobe anatomy, with comments on Alzheimer's disease and temporal injury. Neuropsychiatry Clin Neurosci *9*(3):331, 1997.

Wilson JTL, Hadley DM, Wiedmann KD, Teasdale GM: Neuropsychological consequences of two patterns of brain damage shown by MRI in survivors of severe head injury. J Neurol Neurosurg Psychiatry *59:*328, 1995.

Wroblewski BA, Joseph AB, Cornblatt RR: Antidepressant pharmacotherapy and the treatment of depression in patients with severe traumatic brain injury: A controlled, prospective study. J Clin Psychiatry *57:*582, 1996.

2.6 NEUROPSYCHIATRIC ASPECTS OF MOVEMENT DISORDERS

DAVID GORDON DANIEL, M.D., MICHAEL F. EGAN, M.D. AND STEVEN S. WOLF, M.D.

In common to all extrapyramidal movement disorders is dysfunction of circuits involving cortico-basal ganglia-thalamic loops or the cerebellum. The former circuits funnel information from anterior cortical regions into the striatum. This information is funneled to the thalamus and then the circuit is completed back to the cortex, particularly anterior, motor, and multimodal cortical regions. The basal ganglia and thalamic influences modulate cortical motor activity and integrate input from the cerebellum. In motor function the basal ganglia have “switching” and “braking” functions that orchestrate smooth movement by coordinating facilitation and suppression of motor activity. The frontal corticostriatothalamic loops can be divided by type into separate parallel circuits with motor, oculomotor, limbic, and prefrontal targets. This explains, in part, the association of basal ganglia disease with cognitive, emotive, motor, and oculomotor features.

The primary components of the basal ganglia are the caudate nucleus, the putamen, and the globus pallidus. The caudate and putamen nuclei are sometimes referred to as the *striatum.* The striatum receives afferent fibers from the cortex. The output of the striatal regions can be divided into two segregated circuits, the direct, dopamine type 1 (D_1) receptor-mediated striatonigral loop, and the indirect, dopamine type 2 (D_2) receptor-mediated striatopallidal loop (Fig. 2.6-1). The former projects directly to the internal segment of the globus pallidus (GPi) and substantia nigra pars reticulata (SNr). The indirect loop first goes to the external segment of the globus pallidus (Gpe), then to the subthalamic nucleus (STN), and then to the globus pallidus and substantia nigra pars reticulata GPi/SNr. Normal motor function is produced by the coordinated action of both in regulating thalamic excitatory influences on the cortex. Relative overactivation of the direct pathway leads to hyperkinetic movements such as are seen in Huntington's chorea, Levodopa (Larodopa)-induced dyskinesias, tardive dyskinesia, and tic disorders. Relative overactivation of the indirect pathway leads to hypokinetic symptoms, such as are seen in parkinsonian syndromes.

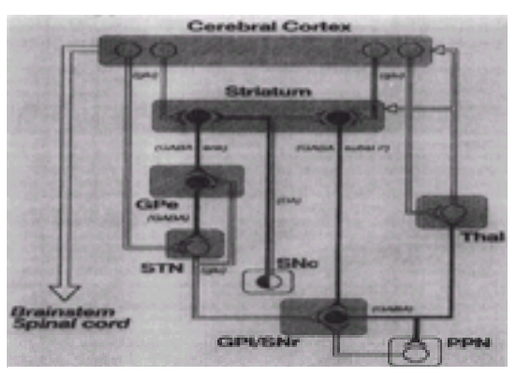

FIGURE 2.6-1 Schematic diagram of the circuitry and neurotransmitters of the basal gangliathalamocortical circuitry, indicating the parallel ‘direct' and ‘indirect' pathways from the striatum to the basal ganglia output nuclei. Inhibitory neurons are shown as filled symbols, excitatory neurons as open symbols. Abbreviations: DA, dopamine; enk, enkephalin; GABA, g-aminobutyric acid; Gpe, external segment of globus pallidus; Gpi, internal segment of globus pallidus; glu, glutamate; PPN, pedunculopontine nucleus; SNc, substantia nigra pars compacta; SNr, substantia nigra pars reticulata; subst P, substance P; STN, subthalamic nucleus; Thal, thalamus. (Reprinted with permission from Alexande GE, Crucher MD: Functional architecture of basal ganglia circuits: Neural substrates of parallel process. Trends Neurosci *13:*266, 1990.)

Cortico-basal ganglia-thalamic loop function requires balanced activity of several neurotransmitters and neuromodulators (Fig. 2.6-1). Cortical areas send excitatory glutaminergic projections to input areas of the striatum. Inhibitory nigrostriatal dopamine projections facilitate conduction through the direct pathway, and inhibit the indirect pathway. Excitatory cholinergic projections balance the inhibitory effects of dopamine. g-Aminobutyric acid (GABA), acting as an inhibitory transmitter, is released at terminals of the striatonigral pathway. Substance P and enkephalin act as neuromodulators. In addition to modulating disordered movement, most of these substances are considered important in the pathophysiology of disorders of reality testing, cognition, and emotion.

PARKINSON'S DISEASE

Introduction Parkinsonism describes characteristic clinical motor manifestations of disrupted nigrostriatal-modulated basal ganglia activity. The complex interconnections of the basal ganglia and cortical, limbic, and thalamic circuitry intimately link parkinsonism with disorders of cognition, mood, and reality testing. Treatment of parkinsonism with dopaminergic or anticholinergic agents may further disrupt normal homeostatic balances related to psychiatric, cognitive, and motor function. Parkinsonism is a syndrome of multiple etiologies in which at least two of the following four cardinal clinical signs are present: resting tremor, bradykinesia, rigidity, and postural instability (Table 2.6-1). The terms “primary parkinsonism,” “idiopathic parkinsonism,” and “paralysis agitans” are used interchangeably with *Parkinson's disease.*

Table 2.6-1 Known Causes of Parkinsonism

Definition The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) recognizes several psychiatric phenomena that occur secondary to Parkinson's disease. A clinically significant mood disorder believed to be secondary to Parkinson's disease is referred to as *mood disorder due to Parkinson's disease.* Types include, with depressive features; with major depressive-like features; with manic features; and with mixed features. Dementia due to Parkinson's disease is characterized by “the presence of a dementia that is judged to be the direct pathophysiological consequence of Parkinson's disease.” Prominent delusions

or hallucinations that are judged to result from the direct physiological effects of Parkinson's disease or its treatment are referred to as *psychotic disorder due to Parkinson's disease.* Types include with delusions and with hallucinations. Prominent delusions or hallucinations judged to be the direct physiological effect of a drug treatment for Parkinson's disease are termed *substance-induced psychotic disorder.* Delirium caused by Parkinson's disease or its treatment is termed by DSM-IV *delirium due to Parkinson's disease* and *substance intoxication delirium*, respectively.

History Parkinsonian-like syndromes were described in ancient Egyptian and Indian texts dated to the twelfth century bc and circa 1000 to 2500 bc respectively. The latter described treatment with a levodopa-containing herb, *Mucuna puriens.* Shortly after John Parkinson's description in 1817, Jean Martin Charcot documented the therapeutic effects of anticholinergic belladonna alkaloids. The role of the substantia nigra was first proposed in 1895, and in 1912, Friedrich Lewy first described what came to be known as the *Lewy body.* Around 1957 it was noted that dopaminergic precursors reversed reserpine-induced parkinsonism; shortly thereafter, reductions in dopamine concentrations were demonstrated in postmortem brain tissue of patients with parkinsonism. In 1960 efforts began to treat Parkinson's disease with levodopa (Laradopa) and within the decade clinically significant results had been achieved.

Epidemiology Most estimates of the prevalence of Parkinson's disease range between 1 and 2 per 1000. Primarily a disease of later life, Parkinson's disease occurs in less than 1 percent of cases before age 30. The prevalence increases markedly with age to around 1 per 100 after age 65; males and females are affected approximately equally. The disease is more common in industrialized countries. Depressed mood may be the most common psychiatric complication of Parkinson's disease, with prevalence estimated between 20 and 63 percent, depending on the criteria for depression and the subpopulation studied. Cognitive impairment, ranging from mild to severe in intensity, is present in the majority of patients. DSM-IV notes that prevalence estimates for dementia in Parkinson's disease range from 20 to 60 percent. Increased risk of dementia is associated with older age, advanced disease, atypical neurological features, and depressed mood. Estimates of the prevalence of psychotic symptoms vary widely with stage of illness, setting, degree of cognitive impairment, and past history of psychosis. The prevalence of delusions and visual hallucinations during dopamine replacement therapy has been estimated at 15 and 20 percent respectively, but may be much higher in the later stages of Parkinson's disease.

Etiology In Parkinson's disease, hypokinetic motor manifestations result from overactivation of the indirect feedback loop secondary to loss of the modulating effects of the pars compacta of the substantia nigra dopaminergic projections to the neostriatum (caudate-putamen nuclei of the basal ganglia) (Fig. 2.6-1). The disease becomes clinically detectable when approximately 70 percent reduction in neostriatal dopamine levels occurs. The best explanation for its late onset may be a combination of normal depletion of dopamine with aging, which occurs at a rate of approximately 13 percent by decade, combined with vulnerability from an earlier disease-related insult that lowers dopaminergic reserve. Experimental models of parkinsonism suggest that earlier onset may be delayed by compensatory mechanisms such as increased activity and dopamine release by surviving neurons, reduction in dopamine reuptake sites in the synaptic cleft, and receptor supersensitivity. One of the best animal and human models of parkinsonism is exposure to a byproduct of synthetic narcotic synthesis, *N*-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP). MPTP is a neurotoxin that selectively damages the substantia nigra and causes motor abnormalities similar to those seen in Parkinson's disease that are partially reversible with levodopa. Free radicals and other toxic metabolites are prime suspects in the pathogenesis of Parkinson's disease. Dopaminergic neurons are considered vulnerable to oxidative stress via a variety of pathways. However, degeneration may occur in multiple other neurotransmitter and anatomical systems, contributing to a wide range of dysfunction. For example, cholinergic imbalances and degeneration of neurons in the hypothalamus, sympathetic ganglia, dorsal motor nucleus of the vagus, and myenteric plexus may contribute to autonomic dysfunction.

Psychiatric Manifestations Depression has been noted prior to observation of extrapyramidal signs in 34 to 43 percent of patients with Parkinson's disease. Theoretically, deficits in mesocortical or mesolimbic dopamine systems could affect mood before the clinical threshold for motor signs from nigrostriatal dopamine degeneration is attained. Impairment of serotoninergic or noradrenergic function could also contribute. Psychosocial reaction to disability is likely to play an etiological role in depression in Parkinson's disease. The cognitive deficits follow a primarily “subcortical” pattern believed to be associated with degenerative processes in the basal ganglia, thalamus, and brainstem. Dopaminergic deficits and Lewy bodies are seen in cell bodies in the ventral tegmental areas that project to areas related to motion and cognition, including the nucleus accumbens, olfactory tubercle, central amygdaloid nucleus, cingulate, prefrontal lobes, entorhinal cortex, and hippocampus. Damage to the mesocortical dopamine projections may be particularly injurious to executive functions linked to the frontal lobe, and appears sensitive to dopamine replacement. Cholinergic deficits could impair memory. Psychosis in Parkinson's disease may have several concurrent causes including dopamine imbalances innate to Parkinson's disease, dopamine replacement therapy, anticholinergic therapy, dementia, and delirium. In Parkinson's disease, delusions and hallucinations in the context of a clear sensorium were reported prior to the dopamine replacement era. In theory, damage to mesocortical-cortical-subcortical feedback loops could interfere with cortical modulation of limbic activity and lower the threshold for psychosis. The most common cause of psychotic symptoms in Parkinson's disease is dopamine replacement therapy, presumably by stimulation of postsynaptic D_2 receptors in the mesolimbic tracts. Because cerebral compensatory mechanisms are tenuous, patients with Parkinson's disease are vulnerable to decompensation into delirium by a wide variety of medical, surgical, and toxic insults. Any medication that alters central nervous system (CNS) function may cause delirium, even at what appear to be subtherapeutic dosages. Sleep disturbance in Parkinson's disease may have multiple concurrent causes including age-associated sleep patterns, depression, delirium-related disturbance in sleep-wake cycle, the rigidity and bradykinesia of Parkinson's disease, recurrence of tremor with awakening, and dopaminergic drugs.

Diagnosis and Clinical Features

Motor Manifestations The clinical diagnosis of Parkinson's disease is confirmed at autopsy in around 80 percent of cases. The diagnosis of Parkinson's disease requires an unexplained progressive condition with at least two of four essential features: bradykinesia, tremor, rigidity, and postural instability. Bradykinesia is a state of globally decreased motor activity manifest by slowness, decreased spontaneity, and subjective feelings of apathy. Parkinsonian tremor is rhythmic, relatively slow (3 to 6 cycles per second), present at rest, worsened by stress, and affects the limbs, head, lips, and tongue. Parkinsonian muscle rigidity may be continuous (“lead-pipe”) or ratchetlike (“cog-wheel”) when the joint is passively moved. Subjectively, muscle stiffness or muscle aching and incoordination may be experienced. Bradykinesia and rigidity may result in drooling, micrographia, hypophonia, and a festinating gait, consisting of small shuffling steps, bent posture, and decreased arm swing. Diminished postural stability may lead to falls; the motor manifestations generally progress with aging.

Clinical Manifestations Depressed mood often precedes the onset of motor symptoms and is accompanied by sleep and appetite disturbance, anhedonia, psychomotor retardation, and anxiety. Guilt and remorse are often less prominent than in other forms of depression. Depression in Parkinson's disease is associated with increased cognitive deficits. Although the pattern of cognitive deficits is conceptualized as primarily subcortical in Parkinson's disease, there is much overlap with executive, visuospatial, and other cortical functions. The cardinal features, predominantly subcortical patterns of cognitive impairment, include slowed information processing, apathy, depression, memory deficits, and difficulty manipulating acquired knowledge. Compared to patients with Alzheimer's disease, in patients with Parkinson's disease word fluency, attentional skills, information-processing speed, and motor aspects of speech are more prominently impaired and loss of recognition memory, loss of intellectual aspects of language, aphasia, agnosis, and apraxia are less prominent. Psychotic syndromes may occur in clear or impaired sensoriums, with or without dementia. Common symptoms include hallucinations, vivid dreams, nightmares, and paranoid ideation. In contrast to schizophrenia, visual hallucinations are more common than auditory ones; paranoia is generally not systematized; the sleep-wake cycle is generally disturbed; and formal thought disorder, loose associations, blocking, and bizarre affect are uncommon. In the so-called benign hallucinosis of Parkinson's disease, the patient's insight into the disease origin of the hallucinations is preserved. It is not unusual, however, for this state to progress to more florid psychotic states in which the patient lacks insight. In psychosis associated with dementia of Parkinson's disease, disorientation and confusion may be prominent. Clinical features of delirium in Parkinson's disease may include relatively rapid onset (hours to weeks), confusion, altered sensorium, fluctuating levels of consciousness and confusion, disturbed sleep-wake cycle, and global cognitive deficits. Patients may be agitated, or in states of “quiet delirium,” may exhibit psychomotor retardation or depressed levels of consciousness. Delirium may be superimposed on preexisting dementia or psychotic states. Early stages of delirium may be characterized by a nocturnal pattern called *sundowning* in which agitation and confusion become worse at night. Recognition of delirium is important because it may be reversible, or if undetected progress to death. Physical and neurological examinations, blood chemistries, and reevaluation of all medications with potential for direct or indirect central nervous system effects are essential. Radiography (X-ray) of the chest, computed tomography (CT) scan or magnetic resonance imaging (MRI) of the brain, or lumbar puncture may be appropriate diagnostic interventions to determine the underlying cause.

Pathology and Laboratory Examination Postmortem studies suggest that approximately 50 percent of nigral neurons must be lost before motor manifestations of Parkinson's disease develop. Damage is primarily to the ventrolateral portion of the nigra, which loses its pigmentation and appears pale. Reductions in dopamine and its metabolite correlate with degeneration of the dopamine-releasing neurons, which have their bodies in the pars compacta of the substantia nigra and whose neurons terminate in the neostriatum. In Parkinson's disease characteristic eosinophilic cytoplasmic inclusions called Lewy bodies are found in the substantia nigra, locus ceruleus, dorsal vagal nucleus, and nucleus basalis of Meynert, and in lesser concentrations elsewhere in the nervous system. The potential primary versus secondary roles of Lewy bodies in the pathophysiology of Parkinson's disease is unknown. In intellectually impaired patients with Parkinson's disease reduced choline acetyl transferase in the cerebral cortex and substantia innominata and cell loss in the nucleus basalis of Meynert have been observed, but not uniformly, and unlike in patients with Alzheimer's disease are not consistently accompanied by neurofibrillatory tangles and amyloid plaques. In a minority of cases the neuropathology of Parkinson's disease may be similar to that seen in Alzheimer's disease; this is particularly likely in the most severely demented Parkinson's disease patients, who also

may share metabolic patterns (i.e., temporo-parieto-occipital deficits) seen in Alzheimer's disease.

Differential Diagnosis Potentially preventable or reversible causes of parkinsonism and their treatment include Wilson's disease (copper chelation), hydrocephalus (drainage), subdural hematoma (drainage), tumor (surgical recision), hypoparathyroidism (hormone replacement), non-wilsonian hepatocerebral degeneration (treatment of hepatic failure), neuroleptics (reduction of dosage or change of agents), calcium channel blocker intoxication (reduction of dosage or change of agents), multiple lacunar infarctions (treatment of hypertension), toxin intoxication such as manganese, MPTP, or carbon monoxide; and pugilistic parkinsonism. In Parkinson's disease MRI may reveal thinning of the low signal band produced by the substantia nigra but the usefulness of this finding for early diagnosis is limited because it may not be evident early in the disease. However, the diagnosis of several secondary causes of parkinsonism may be assisted by MRI, including vascular parkinsonism (high T2 signal in striatum or brainstem), multiple system atrophy (low signal in putamen), progressive supranuclear palsy (midbrain atrophy-enlarged posterior third ventricle), olivopontocerebellar atrophy (atrophy of cerebellum and pons), Fahr's disease (parathyroid dysfunction, basal ganglia calcifications), and hydrocephalus (ventricular enlargement without gyral atrophy).

In vivo brain imaging techniques that examine blood flow, metabolism, dopamine binding, or ^{18}F-fluorodopa (FDG) uptake are being increasingly utilized to study Parkinson's disease (Fig. 2.6-2). These techniques may have predictive value for presymptomatic states or to confirm diagnoses in the future. Positron emission tomography (PET) studies of Parkinson's disease using a labeled dopamine precursor, ^{18}F-flurodopa, indicate an overall reduction in dopaminergic innervation in the striatum and are consistent with the pattern of dopamine deficits seen in postmortem studies. Striatal uptake of FDG correlates with motor performance in Parkinson's disease, particularly bradykinesia. FDG PET scanning has been shown to demonstrate preclinical dopamine deficts in at-risk patients exposed to MPTP and in a normal control who later developed parkinsonian symptoms. A theoretical application would be to identify potential candidates for protective therapy, if such could be established. PET and singal photon emission computed tomography (SPECT) studies of glucose metabolism and cerebral blood flow in patients with dementia due to Parkinson's disease indicate that some patients may share patterns of temporal-parietal deficits with patients with Alzheimer's disease. However, such patients do not necessarily have Alzheimer's disease-like pathology upon postmortem examination. At this time PET scanning cannot distinguish with certainty among causes of dementia due to Parkinson's disease and is not considered to be useful as a routine clinical tool.

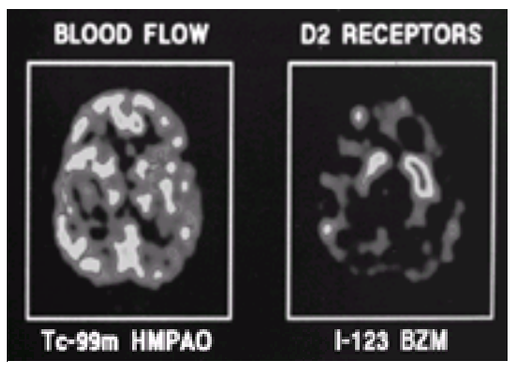

FIGURE 2.6-2 Regional cerebral blood flow, visualized with Technetium-99m HMPAO (**left panel**), and dopamine D_2-receptor binding, visualized with ^{123}I-iodobenzamide (I-123 BZM) (**right panel**), in a patient with hemiparkinsonism. Individuals with Parkinson's disease may present with markedly asymmetric symptoms. In this case, I-123 BZM uptake is markedly increased in the basal ganglia contralateral to the affected limb, whereas blood flow is unaffected. The increased uptake reflects increased D_2-receptor availability, either because of receptor upregulation or depletion of competing dopamine for the D_2 sites. (Courtesy of Michael Knable, D.O., and Daniel R. Weinberger, M.D.) (See color Plate 2.)

Differential diagnosis of depression from manifestations of Parkinson's disease and dopamine replacement therapy is difficult because of shared psychomotor retardation, psychomotor agitation, sleep disturbances, sexual disturbances, anorexia, and fatigue. Anhedonia may be difficult to distinguish from facial masking and bradykinesia.

Course and Prognosis Although motor manifestations of Parkinson's disease usually progress with age, heterogeneity and irregularity in rate is the rule. Tremor-dominant patients generally have a later onset, less cognitive disturbance, and a more favorable prognosis. The rate and outcome of cognitive deterioration is exceedingly variable, ranging from subtle to severe.

Treatment Aims of treatment may be protective, symptomatic, or restorative. Protective therapy aims to prevent or slow the destruction of dopaminergic neurons by hypothetical mechanisms of oxidative stress. Identification of risk factors and presymptomatic markers for Parkinson's disease are essential components of this strategy. Clinically unsubstantiated interventions include facilitation of processes that lead to deactivation of free radicals and augmentation of mitochondrial and other cellular defenses and reparative processes via antioxidants, free radical scavengers, and inhibitors of monoamine oxidase (MAO) and glutamate antagonists. Vitamin E therapy has produced modestly encouraging but inconclusive benefits. Large double-blind clinical trials of MAO type B (MAO_B) inhibitors in Parkinson's disease found an increase in the time from diagnosis to dopamine replacement therapy. Although the interpretation that $MAOI_B$ inhibitors may prevent destruction of dopaminergic neurons or otherwise alter the course of Parkinson's disease independent of symptomatic improvement remains controversial, they are frequently used in practice, particularly early in the disease. Ongoing investigational approaches seek relatively high MAO_B selectivity without amphetamine-like metabolites.

Symptomatic treatment consists primarily of strategies that enhance dopaminergic transmission by increased dopamine synthesis, dopamine agonists, enhanced dopamine release inhibition of dopamine uptake, and inhibition of dopamine catabolism. Amantadine (Symmetrel) is sometimes used as an initial monotherapy. Levodopa, a precursor of dopamine, is the mainstay of treatment. Carbidopa, a peripheral dopa decarboxylase inhibitor, increases the CNS bioavailability of levodopa and reduces peripheral adverse effects. When in treatment levodopa should be initiated remains controversial because motor fluctuations, dyskinesias, and treatment resistance are frequent late complications of therapy and may be related to cumulative levodopa exposure. Dopamine agonists such as bromocriptine (ParLODel) or pergolide (Permax) are commonly used as adjunctive treatment, often with a view to prevent or reduce motor complications from levodopa therapy. Anticholinergic therapy seeks to obviate the effects of dopaminergic imbalances in producing overactivity of cholinergic neurons and is sometimes used as an initial monotherapy for mild symptoms, or as an adjunct for sialorrhea or tremor. Peripheral and central anticholinergic side effects are a source of toxicity in the elderly.

Recent advances in symptomatic treatment of Parkinson's disease have occurred along two primary lines. Newer agents such as pramipexole (Mirapex), ropinirole, and cabergoline selectively target dopamine D_2 and D_3 receptors in an attempt to ameliorate parkinsonian motor symptoms with fewer adverse effects. Of these, the nonergoline agents pramipexole and ropinirole have been the most extensively studied as means of delaying the introduction of levodopa and onset of dyskinesias early in treatment and as adjunctive therapies later in the illness. Although, cabergoline has the potential advantage of relatively long action and once daily dosing, no application has been filed in the United States for use in Parkinson's disease. A second line of available new therapies are useful in delaying the wearing-off effects of levodopa. Tolcapone and entacapone increase the bioavailability of levodopa by inhibiting a major route of catabolism via plasma catechol- *O*-methyltransferase (COMT).

The frequency of surgical procedures for Parkinson's disease patients declined dramatically after the introduction of dopamine replacement therapy. Recently interest in surgery has increased with improvements in stereotactic instrumentation, MRI localization, and electrophysiological monitoring procedures. Indications for possible surgery include a combination of unilateral tremor with relatively little bradykinesia; cognitive, speech, or gait disturbance; treatment resistance or intolerance to medicinal treatment; and absence of medical risk factors. Tremor may be reduced by lesions of cortical areas 4 or 6, the globus pallidus, the ventrolateral thalamus, and the ansa and fascicularis lenticularis. In theory, grafting of dopamine-releasing living tissue could have reparative effects in Parkinson's disease. Despite initial optimism, autopsies have shown that grafted adrenal medullary autographs do not survive. The favorable clinical response seen in a third of patients may result from nigral sprouting related to the procedure itself or nerve growth factors released from the dying graft. Fetal midbrain tissue has the theoretical advantage of differentiating and forming synaptic contacts with striatal tissue. Genetically engineered donor tissue and brain-derived neurotrophic factors are investigational approaches.

Treatment of depression due to Parkinson's disease should be accompanied by an assessment of psychological reactions to the illness, psychosocial stressors, potential medical factors such as hypothyroidism and vitamin B_{12} or folic acid deficiency, and consideration of iatrogenic causes. The effect of dopamine replacement

therapy on mood in Parkinson's disease has yet to be fully delineated. If a dopaminergic agent is suspected to exacerbate depression, the lowest effective dosage may be used or substitution of another agent considered. Placebo-controlled evidence suggests that tricyclic antidepressants including desipramine (Norpramin, Pertofrane), impramine (Imipramine), and nortriptyline (Pamelor, Aventyl) may have useful antidepressant properties in some patients with Parkinson's disease. The anticholinergic effects of such agents may have a salutory effect on motor function. Selective serotonin reuptake inhibitors (SSRIs), presumably by feedback on dopaminergic systems through serotoninergic circuits originating in the dorsal raphe circuits, have the potential to produce extrapyramidal side effects. MAO_B inhibitors have not been established as effective treatments for depression in Parkinson's disease. Electroconvulsive therapy has been reported to produce rapid but short-lived improvement in mood, bradykinesia, and rigidity in some patients with Parkinson's disease; proposed mechanisms include increased postsynaptic dopaminergic activity and decreased presynaptic autoreceptor sensitivity. Early in the illness dopamine replacement therapy may remediate some cognitive loss in Parkinson's disease, particularly in memory, executive function, and general intellectual ability. Later in the illness pharmacological intervention may emphasize minimization of drugs such as anticholinergic and sedative-hypnotic agents because they may exacerbate cognitive deficits.

Pharmacotherapy of agitation requires a risk-benefit assessment of the advantages and adverse effects of the medication employed. Sedative-hypnotics, including benzodiazepines, may worsen confusion or disinhibit patients and tolerance may develop to soporific effects; these agents may be associated with daytime sedation, falls, and hip fractures. Short-acting agents may lack propensity to accumulate high blood levels and multiple metabolites, but may cause sharper rebound anxiety. Propranalol (Inderal), buspirone (BuSpar), trazodone (Desyrel), and anticonvulsant agents such as valproic acid (Depakene) may reduce agitation in some patients. Tricyclic antidepressants with sedating and anticholinergic properties may provide sedation and have motor benefits. However, anticholinergic antidepressant and antiparkinsonian agents have the potential to exacerbate memory and other deficits.

In the treatment of psychosis related to dopaminergic replacement therapy patients may be exquisitely sensitive to small doses of high-potency antipsychotics. In contrast, low-potency antipsychotics such as thioridazine (Mellaril) may cause fewer motor effects but more anticholinergic-induced cognitive impairment. Newer serotonin-dopamine antagonists with relative specificity for mesolimbic (A10) compared to nigrostriatal (A9) dopaminergic tracts may broaden the therapeutic window of antipsychotic treatment of Parkinson's disease. Multiple investigations have found that clozapine (Clozaril), at lower dosages and blood levels than used in the treatment of schizophrenia, is efficacious and well tolerated in the treatment of dopamine–replacement-induced psychosis. The use of clozapine in Parkinson's disease has been limited by the necessity for weekly blood monitoring, cost, and its nonmotor adverse effects. There is currently optimism over potential utility of serotonin-dopamine antagonists such as risperidone (Risperdal), olanzapine (Zyprexa), quetiapine (Seroquel) and investigational agents such as sertindole (Serlect) in the treatment of dopaminergic psychosis. Preliminary evidence suggests substantial variation in their motor therapeutic windows in this context. New investigational treatments for psychotic disorders such as highly selective 5-HT_{2A} receptor antagonists, raise optimism for even larger therapeutic windows. In delirium the primary intervention is correction of the underlying systemic, neurological, or iatrogenic cause. Judicious use of antipsychotic medication may be required for control of agitation.

Nonpharmaceutical interventions may helpful. Behavioral disorders such as wandering, agitation, disturbed sleep-wake cycle, and aggressive and hypersexual behavior may respond to familiar routines; clues to time, place, and situation; consistent caregivers, and social interactions. Caregiver education and support groups are essential to maintaining morale, compliance, and consistency of approach.

PARKINSONIAN SYNDROMES

Secondary parkinsonism should be suspected clinically when clinical features atypical to Parkinson's disease are seen such as paucity of tremor, prominent or early gait or autonomic disturbance, pyramidal or cerebellar signs, or a cortically dominant pattern of dementia (Table 2.6-1). With CT, MRI, or postmortem examination a wider pattern of degenerative changes is seen, which explains the additional clinical features. Unlike in Parkinson's disease, SPECT and PET scanning may indicate decreased density of postsynaptic dopamine receptors, explaining in part poor response to levodopa. Parkinsonism-plus syndrome describes multisystem degenerative disorders characterized by both parkinsonian and other neurological features.

Diffuse Lewy Body Disease Lewy body diseases are a spectrum of neurodegenerative disorders associated with motor, cognitive, or psychiatric dysfunction. After Alzheimer's disease, diffuse Lewy body disease is the second most common form of dementia, but its prevalence is difficult to estimate because of the lack of standardized diagnostic criteria. Most cases develop later in life, typically between ages 50 and 80, with males affected more often than females. Diffuse Lewy body disease typically presents with neuropsychiatric features such as hallucinations (often visual), delusions (often paranoid), and memory loss. An insidious, progressive dementia with typically cortical features such as apraxia, aphasia, and agnosia usually ensues. A fluctuating course is considered characteristic. In contrast to Parkinson's disease, motor symptoms, especially rigidity and shuffling gait, usually in the absence of prominent tremor, appear relatively late. Lewy bodies are characteristic eosinophilic intracytoplasmic inclusion bodies found most abundantly in the brainstem in Parkinson's disease. In diffuse Lewy body disease, Lewy bodies are found throughout the brainstem, striatum, thalamus, and cortex. Characteristic patterns of ubiquitin-immunoreactive, *tau* protein-negative degenerative changes in the CA 2-3 region of the hippocampus distinguish diffuse Lewy body disease from Alzheimer's disease. However, in approximately half of cases of diffuse Lewy body disease, typical neuropathological changes of Alzheimer's disease are seen as well; choline acetyltransferase activity is frequently low. The course is progressive with an average survival of less than 7 years following diagnosis. As with Parkinson's disease, extrapyramidal features frequently benefit from dopamine replacement and are exacerbated by neuroleptic agents; some patients appear to benefit from cholinomimetic drugs.

Supranuclear Palsy Progressive supranuclear palsy, first described by Steele, Richardson, and Olszewsky in 1964, is a neurodegenerative syndrome characterized by parkinsonian motor features, impaired ocular motility, prominent postural instability, subcortical dementia pattern with prominent frontal lobe pattern, pseudobulbar signs, and poor response to dopamine replacement.

The prevalence of diagnosed supranuclear palsy is only approximately 1 percent that of Parkinson's disease. A community survey suggests a prevalence of approximately 1.5 per 100,000 for males and 1.2 per 100,000 for females. In approximately 95 percent of cases initial diagnosis occurs after age 50 and mean age of onset is around age 62. Although supranuclear palsy patients are more likely than controls to have lived in rural areas, no other geographic, toxic, infectious, or genetic factors have been established. Neurochemically, marked reductions are seen in striatal dopamine as in Parkinson's disease, but in contrast to Parkinson's disease striatal D_2 receptors are reduced as well. The latter may in part contribute to the characteristically poor responsiveness to dopamine replacement. In contrast to Parkinson's disease mesolimbic dopamine systems are relatively spared and there is widespread cholinergic neuronal damage subcortically, including to the nucleus basalis of Meynert and striatal interneurons. The earliest presenting symptom is often postural instability. Characteristic vertical gaze disturbances, particularly downward, appear relatively late. Rigidity of facial muscles may lead to a characteristic “worried” or “astonished” expression. Pseudobulbar palsy in supranuclear palsy may be characterized by dysarthria, dysphagia, and emotional lability. In contrast to Parkinson's disease, predominance of axial rigidity, minimal resting tremor, symmetric presentation, and unresponsiveness to dopamine replacement are common.

The most consistently affected areas on postmortem examination are the substantia nigra, subthalamic nucleus, and globus pallidus. Marked atrophy of the midbrain and pontine tegmentum are seen with relative sparing of the cerebrum. In contrast to Parkinson's disease, the entire nigra is damaged in patients with supranuclear palsy. Microscopically, neuronal loss with granulovacuolar degeneration and gliosis are seen. Unlike in Alzheimer's disease patients, the neurofibrillatory tangles seen in patients with supranuclear palsy are more likely to be straight than helical and are not found in association with beta amyloid. After the earliest stages, MRI or high-resolution CT scanning may indicate atrophy of the midbrain and anterior temporal lobes, and enlargement of the third ventricle without cerebellar atrophy. PET measures of cerebral glucose metabolism and SPECT measures of cerebral blood flow indicate a bilateral superior frontal dominant pattern of hypometabolism, in contrast to the medial frontal, temporal–parietal-dominant, and diffuse patterns typical of Pick's disease, Alzheimer's disease, and Parkinson's disease, respectively.

Psychiatric Features The psychiatric and cognitive manifestations of supranuclear palsy are highly variable. Estimates of the prevalence of significant cognitive impairment vary from approximately 60 to 80 percent depending on length of illness and are predominantly subcortical in character with slowed thought, prominent frontal lobe features including perseveration; impaired verbal fluency, executive functioning, and mental flexibility; and emotional lability. Visually related cognitive tasks are especially impaired. Aphasia, apraxia, and agnosia are usually not prominent. Affective lability and depression are common. The latter may be difficult to distinguish from apathy and psychomotor retardation. Mania and schizophrenia-like psychoses are rare. When delusions occur they are usually neither bizarre nor systematized. Insomia, multiple awakenings, and abnormal sleep architecture are common. The progression of supranuclear palsy is little altered by currently available therapeutic interventions. Response of motor and ocular manifestations to levodopa and dopamine agonists is usually transient and very mild. Because postsynaptic D_2 receptors are lost, dopamine replacement–induced dyskinesias are rare. Mesolimbic dopamine tracts are relatively spared by supranuclear palsy and dopaminergic psychosis has been reported.

Secondary and tertiary tricyclic drugs and SSRIs have been reported to improve emotional lability or motor symptoms in some patients with supranuclear palsy. The experimental a_2-antagonist idaxozan was reported to produce improvement in motor function and balance in some patients. Nonspecific cholinomimetic agents have

produced cognitive improvement in some patients.

WILSON'S DISEASE (PROGRESSIVE HEPATOLENTICULAR DEGENERATION)

Progressive hepatolenticular degeneration is an autosomal, recessive, systemic disorder of metabolism in which copper accumulation is associated with cirrhosis of the liver, degenerative changes in the basal ganglia, and damage to other organs. Following its initial description by Wilson in 1912, progressive hepatolenticular degeneration became known as *Wilson's disease.* Evidence of excessive copper deposition in the brain was noted around 1940, and successful chelation therapy was described by Cummings in 1951. Copper is primarily removed by biliary excretion; Wilson's disease has been linked to a defect in a recessive gene on chromosome 13 involving copper transport by hepatic lysosomes. Copper accumulates systemically and interferes in local metabolism, causing cell necrosis and death. The prevalence of Wilson's disease is estimated at 30 per million. The gene frequency has been estimated between 1 in 100 and 1 in 180. Hepatic presentations are usually seen between ages 8 and 16. It occurs among all races, but is most prevalent among Eastern European Jews, southern Italians, and residents of the smaller Japanese islands. The onset of neurological symptoms is usually in the second decade, and rarely after the third decade, although onset in the fifth decade has been reported. Slightly more cases are reported in men than women. Women are more likely to present with liver disease and men with neurological disease.

The most common presentations, in descending order, are neurological, hepatic, psychiatric, orthopedic, ophthalmological, hemolytic anemia, and cardiac. Neurological presentations are protean, but typically begin with basal ganglia signs of tremor or rigidity. The classic “wingbeating” tremor is localized to the arms, usually absent at rest, and is elicited by extension. Parkinsonian or intention tremors may also be seen. Parkinsonian rigidity, dystonic postures, and occasionally choreo-athetotic movements are seen. Dystonia and rigidity of laryngeal, pharyngeal, and facial musculature may result in dysarthria, dysphagia, hoarseness, drooling, and the classic vacuous or fixed open-mouthed smile. The tendency for motor disorders to predominate in the bulbar musculature contrasts with parkinsonism. Loss of coordination, convulsions, coma, and delirium may occur. Liver involvement may present with asymptomatic elevation of transaminases, hepatosplenomegaly, jaundice, or hepatitis, which eventually progress to cirrhosis. Treatment for psychiatric manifestations prior to recognition of Wilson's disease is common. Symptoms are usually in the affective and behavioral realm, including depression, mania, irritability, disinhibition, and nonspecific personality changes; schizophrenia-like psychoses are rare, Although dementia may occur, cognitive changes are not consistent, occur relatively late, and are relatively mild compared to motor changes. Kayser-Fleischer rings, which are not specific to Wilson's disease, are usually present by the time neurological symptoms occur and consist of copper-colored granules of copper on Descemet's membrane at the limbus of the cornea. Although early in the course a slit lamp may be required for detection, later they may be seen by ophthalmoscope or even by the naked eye. Diagnostically, serum free copper and urinary copper concentrations are increased, and serum ceruloplasmin, which binds copper, is decreased. Liver biopsy reveals elevated copper content. CT scanning and MRI often but not invariably indicates enlarged ventricles, and both cortical and brainstem atrophy is present. MRI also typically reveals abnormal signals of the lenticular, caudate, dentate, and thalamic nuclei. Without treatment, Wilson's disease is an inevitably progressive, fatal disease. Dietary restriction of copper, potassium sulfide, or zinc to reduce absorption, and chelating agents such as penicillamine may result in clinically significant improvement of neurological and hepatic signs, disappearance of Kayser-Fleischer rings, reversal of hypodense areas on CT, and improvement in cognition and behavior. The earlier treatment is initiated, the better the prognosis for preventing or reducing damage. Attempts at symptomatic relief of motor symptoms by dopaminergic or anticholinergic agents have produced inconsistent results.

MULTIPLE-SYSTEM ATROPHY

Multiple-system atrophy is a neurodegenerative syndrome in which parkinsonian, autonomic, cerebellar, and pyramidal signs are prominent. Compared to Parkinson's disease, autonomic dysfunction is more pronounced, the distribution of neuropathology is wider than in Parkinson's disease, and Lewy bodies are not prominent. Three variants of multiple-system atrophy are recognized: olivopontocerebellar atrophy, striatonigral degeneration, and Shy-Drager syndrome. In a given patient, the clinical and neuropathological features of these variants often overlap. Preliminary evidence suggests that a characteristic neuronal cytoplasmic inclusion may be found in multiple-system atrophy.

Olivopontocerebellar Atrophy First described in 1900 by Dejerine and Thomas, olivopontocerebellar atrophy is clinically distinguished by ataxia and other prominent cerebellar features. The prevalence of olivopontocerebellar atrophy is estimated at 2 per 100,000, although it is higher in some areas. Men are affected almost twice as frequently as women. The average age of onset is approximately 28 years and 50 years for the familial and sporadic varieties, respectively. The cause of olivopontocerebellar atrophy is unknown; deficiencies in glutamate dehydrogenase have been detected but are not specific to olivopontocerebellar atrophy.

The most prominent clinical manifestations are relatively late onset of progressive ataxia of the trunk and limbs; dysarthria; scanning speech and other cerebellar signs; parkinsonian symptoms including akinesia, rigidity, and resting tremor; loss of equilibrium; loss of sphincter control; mood syndromes; a variety of personality changes; sleep disturbances; and deterioration in cognitive function. The frequency and severity of cognitive deficits is uncertain; when present, they usually include memory deficits and are predominantly subcortical in nature. Aphasia, apraxia, and agnosia are usually absent.

On postmortem examination, marked atrophy of the cerebellum, ventral pons, and olive are seen. Degenerative changes in the substantia nigrum and striatum are seen when parkinsonian symptoms are prominent. Degeneration of the posterior columns, spinocerebellar tracts, vagal nuclei, and sacral autonomic nuclei may also be seen. In contrast to Parkinson's disease, Lewy bodies are not prominent. Reduced activity of choline acetyltransferase and acetylcholinesterase, and degenerative changes in the nucleus basalis of Meynert have also been reported.

Polysomnography may reveal abnormal sleep architecture and not infrequently, sleep apnea. Abnormalities may be missed on CT scanning early after onset, but eventually include increased fourth-ventricular width, enlarged cerebellopontine cistern size, and cerebellar atrophy. T2 weighted MRI images may indicate demyelination of transverse pontine fibers. Reduction in PET measures of local cerebral metabolic rates for glucose in the pons and cerebellum may precede discernible structural changes on CT or MRI.

Striatonigral Degeneration First described by Adams and colleagues in 1961, striatonigral degeneration is a rare form of multiple-system atrophy in which rigidity is very prominent. In contrast to Parkinson's disease autonomic dysfunction including orthostatic hypotension and incontinence tremor occur commonly in striatonigral degeneration; tremor is mild or absent, levodopa response is poor, and progression is rapid. Attention and some frontal lobe cognitive functions are impaired, but memory and verbal intelligence measures are usually preserved. Neuropathologically, in contrast to patients with Parkinson's disease, the substantia nigra of patients with striatonigral degeneration is relatively well preserved and marked degenerative changes with gliosis are seen in the putamen, and to a lesser extent, the caudate. The putamen is hypodense on T2-weighted MRI, and PET scanning indicates reduced glucose metabolism in the striatum and frontal cortex.

Shy-Drager Syndrome First described in 1960, this syndrome usually presents with autonomic dysfunction, which may precede motor signs by up to three years. It occurs more frequently in men and is usually diagnosed after age 50. Preganglionic lateral horn degeneration may be seen at autopsy in thoracic spinal segments.

The signs and symptoms may include orthostatic hypotension and erectile, bladder, and rectal sphincter dysfunction. Orthostatic hypotension may cause light-headedness or fainting and is not accompanied by the expected compensatory tachycardia. Supine blood pressure is usually unaffected. Extrapyramidal signs, cerebellar signs, and corticospinal signs appear later. Sleep disturbances occur frequently and may include sleep apnea. Dementia may occur late but is not a regular or predictable manifestation. Laryngeal stridor and breathing irregularities may occur. Death usually occurs within 15 years of diagnosis and may result from pneumonia due to inspiration.

NEUROLEPTIC-INDUCED PARKINSONISM

The form of secondary parkinsonism most commonly seen by psychiatrists is neuroleptic-induced parkinsonism, defined in DSM-IV as “parkinsonian tremor, muscular rigidity, or akinesia developing within a few weeks of starting or raising the dose of a neuroleptic medication (or reducing a medication used to treat extrapyramidal symptoms).” Approximately 50 percent of outpatients undergoing long-term treatment with conventional neuroleptics will develop parkinsonian symptoms at some time. Neuroleptic-induced parkinsonism is presumed to be caused by occupancy of D_2 receptors in the striatum. It occurs most commonly with high-potency antipsychotics with relatively weak anticholinergic affinity, and less frequently with low-potency conventional antipsychotics and atypical antipsychotics, both of which have relatively high affinity for serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptors compared to D_2 receptors. Varying degrees of typical parkinsonian tremor, muscular rigidity, akinesia, and postural instability may occur. Psychiatric complications may include depression and apparent intensification of deficit symptoms. Neuroleptic-induced parkinsonism must be distinguished from other psychiatric states such as catatonia, posturing, and the psychomotor retardation of depression. It may coexist with or be confused with neuroleptic malignant syndrome, Parkinson's disease, and other movement disorders. Although neuroleptic-induced parkinsonism typically begins 2 to 4 weeks after initiation of neuroleptic treatment, symptoms can begin rapidly after initiation of treatment or a dosage increase, or develop very slowly. Treatment consists of dosage reduction, addition of an antiparkinsonian agent such as an anticholinergic or amantadine, or transition to an atypical agent or one with lower potency. If after discontinuation of an antipsychotic neuroleptic-induced parkinsonism persists in excess of 3 months

for an oral agent or 12 months for a decanoate formula, a noniatrogenic etiology is likely.

HUNTINGTON'S DISEASE

Definition Huntington's disease, first reported in the medical literature by George Huntington in 1872, is a familial form of a progressive movement disorder characterized by chorea and dementia.

Epidemiology Family studies demonstrate Huntington's disease to be an autosomal dominant disorder. Although it affects only about 2 to 9 per 100,000 persons, within families this rate is 50 percent. This devastating disorder typically has an age of onset of around 35 and is relentlessly progressive, ending in death within 13 to 15 years after diagnosis.

Etiology Huntington's disease is one of the bright spots in the molecular biology revolution. Linkage to chromosome 4 was reported in 1983, and the gene, named *IT-15*, was definitively identified in 1993. The disease-bearing allele is not a simple, single mutation, but is characterized by an unstable region of deoxyribonucleic acid (DNA) that has many repeated segments of a simple 3 base pair sequence, CAG. People without Huntington's disease typically have 11 to 34 repeats of the CAG trinucleotide sequence; patients with Huntington's disease have 37 to 86 CAG trinucleotide repeats. Interestingly, the number of repeats can vary from generation to generation. For example, if a father with Huntington's disease has 50 repeats, his son could inherit the Huntington's disease gene from his father but show 60 or 70 repeats. This phenomenon of repeat expansion may explain *genetic anticipation*, a phenomenon whereby off-spring develop a disorder earlier than their parent. It appears that the number of trinucleotide repeats affects age of onset, probably because of increased neurotoxicity from increased CAG repeat number.

The IT-15 gene product, termed the *Huntington gene*, is a previously unidentified protein whose function remains unknown but is widely expressed throughout the brain and in other body regions. Animal “knockout” models suggest the protein is essential in early development. The CAG repeats appear in the 5' region of the first exon and are translated into an extended polyglutamine region in the protein. This additional sequence is thought to produce a gain of function, in contrast to the loss of protein function that probably occurs with autosomal recessive disorders. The toxic effects of the mutant Huntington gene have been hypothesized to occur through altered interactions with other proteins, and may induce neurotoxicity via effects on glutamate neurotransmission, apoptosis, free radical damage, or some other mechanism.

Diagnosis and Clinical Features Traditionally the diagnosis of Huntington's disease was made in the context of a positive family history and the onset of characteristic symptoms, and was usually confirmed by noting caudate atrophy on MRI. Now, definitive diagnosis can be made before the onset of symptoms and even in utero, via genetic testing for the number of trinucleotide repeats in IT-15. Clinical features are dominated by motor, cognitive, and psychiatric disturbances. Motor symptoms are classically choreiform, with gradual onset and marked progression. Other motor symptoms, including parkinsonism, myoclonus, dystonia, dysarthria, and tics are also observed. In early-onset Huntington's disease, parkinsonism, ataxia, seizures, and dementia are characteristic. The motor abnormalities frequently result in falls with head trauma and aspiration caused by dysphagia. On neurological examination primitive reflexes are found.

Psychiatric Features A variety of cognitive changes have been noted in patients with Huntington's disease, and are commonly present early in the course. According to DSM-IV the “essential feature of dementia due to Huntington's disease is the presence of a dementia that is judged to be the direct pathophysiological consequence of Huntington's disease.” Huntington's disease has been described as a prototypical subcortical dementia, although the validity of this concept of cortical and subcortical dementias has been debated. Typically, rather than symptoms of cortical dementias (agnosia, apraxia, and poor comprehension), Huntington's disease is characterized by abnormalities in problemsolving, insight, judgment, abstraction, attention, and motivation. Memory problems include difficulties involving retrieval. Language impairment is usually characterized by dysarthria and perseveration.

Psychiatric presentation depends to some extent on age at onset, which is typically the third or fourth decade. Symptom onset can also be seen in childhood and early adolescence or as late as after age 50. Earlier onset is usually accompanied by depression and characterological manifestations, including apathy, irritability, and emotional lability. Mood disorders affect about 30 percent of patients, and typically develop well before the onset of the movement disorder. The most common mood disorder is major depression, although mania is also seen in up to 10 percent of cases. Suicide, a complication of depression, affects about 6 percent of patients with Huntington's disease, which is four to six times higher than in the general population. Psychosis has been reported in 6 to 25 percent of patients with Huntington's disease. Hallucinations and paranoid delusions are most frequent and can develop prior to the onset of motor symptoms.

Pathology and Laboratory Examination Early in the course of Huntington's disease reduced metabolism is seen in the caudate nucleus and the putamen. Later, as symptoms develop, the classic reduction and disappearance of the head of the caudate is seen. On neuropathological examination atrophy is found in the caudate nucleus and the putamen, with the frontal and occipital cortex being variably affected. Tissue loss is observed less frequently in the globus pallidus, substantia nigra, and thalamus. On microscopic examination, loss of medium-sized spiny GABAergic neurons is most prominent, while some sparing of interneurons, such as those containing somatostatin, is observed. Some postmortem and PET studies suggest that D_1 and D_2 receptors are both reduced to similar degrees, suggesting that both striatonigral and striatopallidal pathways are involved in the neurodegeneration. The enkephalin-containing neurons of the striatopallidal pathway may be the first to die, perhaps accounting for the initial presentation with chorea. Loss of this pathway should lead to relative overactivation of the D_1-mediated direct pathway. In early-onset cases, both pathways may be equally affected, leading to predominant parkinsonian symptoms.

Differential Diagnosis The differential diagnosis of Huntington's disease has been simplified by the identification of the mutant gene. The primary challenge for the clinician is thus to know when to order the test. A family history of Huntington's disease in the face of new-onset movement disorder, psychological changes, or cognitive problems should lead one to the diagnosis. For adopted persons, the diagnosis of Huntington's disease must be entertained for new-onset movement disorders. A list of other disorders that also produce chorea and dyskinesia is provided in Table 2.6-2.

Table 2.6-2 Differential Diagnosis of Chorea/Dyskinesia

Course and Prognosis The course of Huntington's disease is relentlessly progressive with increasing cognitive deficits and neurological dysfunction. Typically, patients survive 10 to 15 years after the first clinical manifestations. Death results from head trauma, cachexia and attendant pneumonia, or suicide. There is no cure and the illness is uniformly fatal. Age of onset and progression appears to be strongly correlated with the number of trinucleotide repeats. Patients with 47 or more repeats have an earlier age of onset and more rapid deterioration of both neurological and cognitive function.

Treatment There is no currently accepted treatment that will ameliorate or limit the progressive neuronal degeneration that ultimately leads to the demise of patients with Huntington's disease. Trials of antioxidants, such as idebenone, have not produced significant results, although larger studies are needed. Fetal striatal tissue transplants are currently in development and may offer some hope in the future. Symptomatic relief for chorea can be obtained with neuroleptics, such as haloperidol (Haldol) and risperidone (Risperdal). Cholinergic agonists have not been particularly helpful for treatment of chorea, although trials of newer drugs such as tacrine (Cognex) that cross the blood-brain barrier more easily have not been published.

Treatment of secondary psychiatric symptoms is similar, for the most part, to that of primary disorders. Depression can be treated reasonably well with antidepressants; manic symptoms are more easily controlled with anticonvulsants than with lithium. Psychotic symptoms may not respond as well to neuroleptics as

do primary psychoses. Agitation and other behavioral abnormalities that frequently accompany dementing illnesses can be managed by standard psychiatric interventions.

B.W., aged 41, was admitted to a forensic unit for his first psychiatric hospitalization for court-ordered competency evaluation after assaulting his ex-wife. He was a vague historian. He lost his full-time position as a carpenter 3 years before admission, but was evasive regarding the circumstances. He acknowleged previous treatment with neuroleptics for paranoid ideation, but duration of treatment was unknown. The patient reported a history on the paternal side of "some kind of mental disorder," which resulted in the institutionalization of several relatives, but he had no additional information. Outpatient records revealed that increasingly explosive behavior led to a diagnosis of antisocial personality disorder when the patient was in his mid-30s. The admitting psychiatrist observed choreiform movements of the limbs and trunk, and referred the patient for MRI (Fig. 2.6-3) to distinguish between a diagnosis of tardive dyskinesia versus Huntington's disease. This case exemplifies the diagnostic dilemma that arises when a patient presents with psychiatric symptoms, is placed on neuroleptics, and develops a movement disorder. The psychiatrist was alerted to the possibility of a neurodegenerative disorder by the constellation of severe familial psychiatric illness, the patient's apparently unremarkable premorbid history until he was in his 30s, and the choreiform movements, deoxyribonucleic acid (DNA) testing was not yet available; however, the MRI scan (Fig. 2.5-2) was virtually pathognomonic for Huntington's disease.

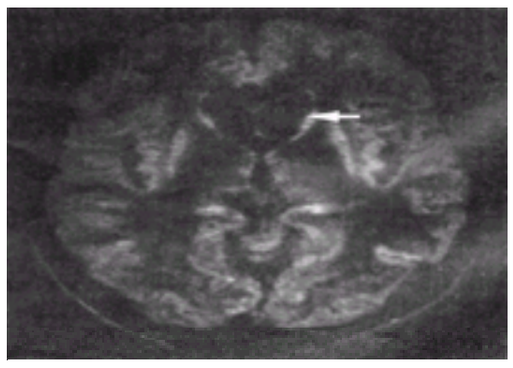

FIGURE 2.6-3 MRI of patient in case report of Huntington's disease (see text). White arrow delineates atrophic head of caudate, with characteristic "butterfly" appearance of the anterior horns of the lateral ventricles. (Courtesy of Steven S. Wolf, M.D.)

NEUROLEPTIC-INDUCED TARDIVE DYSKINESIA

Tardive dyskinesia is a movement disorder that develops during the course of long-term treatment with neuroleptic agents. It is characterized primarily by choreiform and athetotic movements; dystonia, tics, and parkinsonism are also seen. The onset of tardive dyskinesia is typically delayed, with symptoms first appearing only after years of treatment. Because tardive dyskinesia can be disfiguring, persistent, and even permanent, it has been a major limiting factor in the use of antipsychotic medications. Recently, the introduction of clozapine (Clozaril) in the United States has offered hope that tardive dyskinesia may be avoided. Clozapine is a unique, "atypical" neuroleptic that does not appear to cause tardive dyskinesia. Unfortunately, its use has been severely restricted because of its other side effects, the most serious of which is agranulocytosis. Clozapine-induced agranulocytosis has caused a number of fatalities; its use in the United States is now contingent on weekly white blood cell counts. Since the introduction of clozapine and the recognition of its unusual features, intensive efforts have been directed at developing clozapine-like medications that do not cause agranulocytosis. It is hoped that the emerging generation of putative atypical antipsychotics, such as risperidone, olanzapine, and sertindole will, like clozapine, have a reduced liability to produce tardive dyskinesia.

Definition The research criteria for neuroleptic-induced tardive dyskinesia require choreiform, athetoid, or rhythmic involuntary movements to develop in association with use of neuroleptic agents for at least 3 months, persist for at least 4 weeks, and not be attributable to another disorder.

Epidemiology Although estimates vary widely, the best studies suggest that tardive dyskinesia affects about 20 to 30 percent of patients treated with long-term neuroleptic medications. About 5 percent of patients develop tardive dyskinesia for each year they remain on these drugs. A number of variables have been associated with increased liability for developing tardive dyskinesia, including increased age, a diagnosis of mood disorder, brain damage, dosage and duration of treatment, and number of drug-free periods. Sex is often mentioned as a factor, but results from many studies are often confounded by other variables, such as age, and dosage of the neuroleptic drug. Nevertheless, older women seem to be more susceptible to developing tardive dyskinesia. The incidence of tardive dyskinesia during long-term treatment with the newer, putatively atypical agents, such as risperidone, olanzapine, and sertindole, is not known.

Etiology By definition, tardive dyskinesia is caused by chronic exposure to neuroleptic medications (i.e., drugs that potently inhibit D_2 receptors). Neuroleptic drugs do not produce tardive dyskinesia in all exposed persons, however, which means that substantial differences in liability exist. The additional factors that predispose persons to tardive dyskinesia suggest that a number of neurobiological processes may be involved. For example, the association of tardive dyskinesia with cortical brain damage implies that altered corticostriatal input could play a role. The possible reduced liability to tardive dyskinesia associated with atypical agents—the serotonin-dopamine agonists—implies that serotonin may play a role. Finally, the appearance of tardive dyskinesia-like movements in drug-naive patients with schizophrenia raises questions of whether the tardive dyskinesia-inducing processes are fundamental to the neurobiology of this illness.

Diagnosis and Clinical Features The diagnosis of tardive dyskinesia is made by excluding other organic causes of hyperkinetic movement disorders. In general, patients who present with mild to moderate choreo-athetotic symptoms during long-term treatment with neuroleptic drugs will invariably be diagnosed as having tardive dyskinesia. The yield of extensive medical evaluations on such patients is very low, and it is unclear whether this is cost effective. However, a neurological examination and family history are easily obtained and represent a reasonable initial workup. Additional laboratory testing can be considered, particularly for patients with moderate to severe symptoms, functional impairment, marked dystonic features, significant asymmetry, medical complications, or rapid progression. Such presentations sometimes indicate an underlying neurological disorder. A reasonable laboratory screening battery for such cases would include blood cell count (to look for acanthocytes), thyroid indices, liver function tests, electrolytes, renal function tests, antinuclear antibodies and lupus erythematosus preparation, rapid plasma reagin (RPR), serum copper and ceruloplasmin, and pregnancy test (where appropriate). Additional testing for suspicious cases would also include slit lamp examination of the eyes, MRI of the brain, and electroencephalogram (EEG).

Neurological Features Tardive dyskinesia is characterized by chorea and athetosis. The distribution of these movements tends to fall into two areas: (1) the oral, buccal, and facial regions, and (2) the neck, trunk, and extremities. Some authors have proposed that orofacial and limb-truncal distributions may have different underlying neurobiological correlates. Frequently, patients have symptoms both in orofacial regions and in the extremities. Dystonia is a second type of movement abnormality frequently seen in tardive dyskinesia, when it predominates, it is referred to as *tardive dystonia.* Dystonic symptoms usually affect the neck and hands, but can involve other regions as well. Other types of movement abnormalities can be part of the tardive dyskinesia syndrome; these include akathisia, tics, and even parkinsonism. Patients with tardive dyskinesia also have more neurological soft signs as compared to patients without tardive dyskinesia. The motor symptoms of tardive dyskinesia should be routinely looked for using the Abnormal Involuntary Movement Scale (AIMS) examination, or one similar to it, in all patients treated with long-term neuroleptic agents.

Psychiatric Features Patients with tardive dyskinesia tend to demonstrate a number of neuropsychiatric features as compared to patients without tardive dyskinesia. These differences include more negative symptoms (i.e., alogia, avolition, and affective flattening), more prominent ventricular enlargement, and greater deficits in neuropsychological testing. Such patients may be more likely to develop tardive dyskinesia. It has been argued, for example, that patients with schizophrenia have deficits in cortical function from a variety of different regions (e.g., frontal and temporal regions). Patients with more severe cortical deficits may be more likely to develop tardive dyskinesia because of changes in subcortical function that are secondary to cortical abnormalities. Another possibility is that the constellation of negative symptoms, cognitive deficits, and ventriculomegaly is related to tardive dyskinesia itself.

Pathology and Laboratory Examination No characteristic feature is noted in laboratory evaluations of tardive dyskinesia. A variety of postmortem studies have suggested alterations in GABAergic neurons in several basal ganglia structures, but these findings are experimental and have no diagnostic value. Neuroimaging studies have suggested a slight increase in ventricular volume in patients with tardive dyskinesia, but not all studies have found this.

Differential Diagnosis Neuroleptic-induced tardive dyskinesia must be distinguished from other hyperkinetic movement disorders such as are described in Table 2.6-2. Other causes should be sought when symptoms preceded treatment with neuroleptics or if other neurological signs are evident. Moreover, neuroleptic-induced tardive dyskinesia may coexist with other neuropathology.

Course and Prognosis Once tardive dyskinesia appears, it is thought that symptoms are not relentlessly progressive but seem to plateau in severity; however, clear documentation of this is limited. Experts typically recommend discontinuing neuroleptics if possible, but this is usually very difficult. Patients probably have a greater chance of complete remission if they stay off neuroleptics, but increases in relapse and even suicide complicate this therapeutic option. The symptoms of those who remain on neuroleptics will wax and wane over the course of years. Marked progression in severity appears to be relatively uncommon. When severe tardive dyskinesia occurs, it is thought to develop fairly quickly over the course of months. Severe tardive dyskinesia is seen in a fairly low percentage of the population, estimated to be 5 to 7 percent in some studies. Extreme cases, which are even more uncommon, can lead to a variety of complications. These include disfigurement, marked weight loss, respiratory distress (from diaphragmatic involvement), and difficulties with walking, eating, and even sitting. Death has been reported from aspiration and general cachexia. Prognosis may be improved by switching patients to atypical neuroleptics, although studies in this area are scarce. Clozapine, in particular, is clearly associated with a dramatically reduced incidence of tardive dyskinesia. Switching patients with moderate to severe tardive dyskinesia to clozapine may therefore improve their chances of remission, although this has not been clearly documented.

Treatment Treatment options abound, largely because of their fairly low efficacy. The development of a new class of putative atypical, clozapine-like agents is sure to have an impact on how tardive dyskinesia is dealt with. Many practitioners are switching patients to risperidone, olanzapine, or clozapine with the hope of improving chances of remission. It is too early to say whether the currently available atypical neuroleptics are effective in suppressing tardive dyskinesia symptoms or improving long-term outcome. A number of other medications have been used to suppress moderate to severe symptoms. These include benzodiazepines (particularly clonazepam [Klonopin]), clonidine (Catapres), propranolol (Inderal), vitamin E, reserpine (Diupres), and calcium channel inhibitors. The selection of suppressive agents depends on carefully weighing the benefits and risks for individual patients.

TOURETTE'S DISORDER

Introduction Tourette's disorder occupies a unique position in the field of neuropsychiatry. Although often classified as a neurological movement disorder, Tourette's disorder is characterized by abrupt movements and vocalizations, or tics, usually performed in response to compelling urges. The phenomenology of tics sets them apart from the majority of movement disorders, in which the abnormalities of movement are strictly involuntary. Moreover, obsessive-compulsive symptoms are intimately linked with Tourette's disorder, and other forms of psychopathology are frequently comorbid. For these reasons individuals with Tourette's disorder may seek treatment from neurologists or psychiatrists; thus, Tourette's disorder clearly bridges these two disciplines and may be considered a model neuropsychiatric disorder.

Definition Criteria for Tourette's disorder are provided in DSM-IV (see Table 42-1) and include the core features of multiple motor tics and one or more vocal tics, not necessarily concurrent, usually occurring in bouts throughout the day, and persisting for at least 1 year. Further, symptoms should appear by age 18, must cause marked distress or impairment in some important aspect of daily life, and must not be attributable to a substance or a general medical condition. Other proposed criteria are less stringent and include individuals with symptoms regardless of degree of impairment of social-occupational function.

Epidemiology and Etiology Once considered rare, Tourette's disorder is now recognized as a relatively common disorder. Although an accurate prevalence rate has not been established, an earlier figure of 0.5 per 1000 is likely to be a significant underestimation. Tourette's disorder has been identified in all major cultural and racial groups, and is three to four times more common in males than in females. Georges Gilles de la Tourette provided the first comprehensive description of the condition that bears his name in 1885, yet its etiology remains elusive over a hundred years later. High concordance rates for Tourette's disorder within monozygotic twin pairs (approximately 75 percent) clearly suggests a strong genetic burden, although the lack of complete concordance and existence of sporadic cases also implies a contributory role for nongenetic factors. Multiple family members may be afflicted with either Tourette's disorder, chronic motor or vocal tic disorder, obsessive-compulsive behaviors, or combinations thereof. Thus, Tourette's disorder may simply be the most pronounced manifestation of a group of related tic disorders, ranging from transient tic disorder at one end of the spectrum to Tourette's disorder at the other end, which share a common genetic substrate. Most evidence points to a major autosomal-dominant gene for these tic spectrum disorders, with penetrance ranging from approximately 70 percent in females to 99 percent for males. Despite overwhelming evidence of a genetic substrate, attempts to identify a major gene effect by linkage analysis have excluded a significant portion of the genome and a number of candidate genes. Multicenter studies are also adopting methodologies other than linkage analysis to address the genetic contribution to Tourette's disorder.

Clinical Features Tics usually first appear in childhood (mean age of onset is 7 years), with motor tics most often predating vocal tics. Motor tics may range from simple eye blinks, grimaces, neck jerking, or shoulder shrugs to complex tics consisting of touching, stamping the feet, squatting, smelling objects, echopraxia (imitation), or copropraxia (obscene gesturing). Similarly, vocal tics range from simple utterances such as throat clearing, sniffing, humming, grunting, or barking, to complex vocal tics, including blurting out words or phrases, echolalia (repeating the words of others), palilalia (repeating one's own words), or coprolalia (obscenities or socially inappropriate language). The latter has lent a certain notoriety to Tourette's disorder, although it appears in 20 percent or less of patients, a figure that may prove to be an overestimation as milder cases of Tourette's disorder are diagnosed. Coprolalia is rare in Japan compared with Western countries, which suggests a significant cultural influence upon the manifestation of these complex tics. Tics of all types characteristically disappear during concentration on a task or enjoyable activity and exacerbate with stress, boredom, or excitement. Another common feature is that tics can often be suppressed for variable periods of time, usually associated with mounting inner tension until the tics are released.

Psychiatric Features Beginning with the early historical accounts, reports of associated psychopathology in Tourette's disorder have been both numerous and contentious, some suggesting no specific links and others implicating a host of allied psychopathology. The most frequently encountered symptoms include obsessive-compulsive behaviors, hyperactivity, depression, and anxiety. Learning disabilities and self-injurious, oppositional, aggressive, and antisocial behaviors have all been reported. Vulnerability to depression and anxiety may be a consequence of the chronic, socially stigmatizing nature of the disorder for many patients. Overall neuropsychological test performance is unremarkable in patients with Tourette's disorder, although subtle deficits in visual-motor skills have been noted. There is no evidence of progressive deterioration in cognitive function.

OBSESSIVE-COMPULSIVE BEHAVIORS Complex tics may be difficult to distinguish from obsessive-compulsive behaviors; it has been argued that the distinction is merely one of degree. Typical symptoms include smelling or touching rituals, need for symmetry, checking behaviors, and obsessive thoughts of aggressive or sexual acts. Certainly, there is a close link between obsessive-compulsive symptoms and Tourette's disorder; the prevalence of obsessive-compulsive disorder is much higher in patients with Tourette's disorder than in the general population—estimates vary, but an estimate approaching 50 percent is typical. However, studies of patients with comorbid obsessive-compulsive disorder and Tourette's disorder suggest some differences from obsessive-compulsive disorder without associated tic disorder. Obsession with symmetry, “just right” phenomena (the need to perform a behavior repetitively until it feels just right), and certain rituals such as tapping, touching, and counting, are more common in patients with obsessive-compulsive disorder and Tourette's disorder, whereas pure obsessive-compulsive disorder patients have more contamination fears and cleaning rituals. Compulsive behaviors in Tourette's disorder are more frequently driven by a need for completeness or perfection, whereas compulsions in pure obsessive-compulsive disorder are more commonly driven by a sense of fear, dread, or worry. These observations have raised suspicions that obsessive-compulsive disorder might be a heterogeneous condition, one subtype being especially prevalent in Tourette's disorder patients. Twin, family, and complex segregation analysis studies also support the proposition that some obsessive-compulsive behaviors are in fact an alternative clinical manifestion of the Tourette's disorder gene.

ATTENTION-DEFICIT/HYPERACTIVITY DISORDER Frequently comorbid with Tourette's syndrome, attention-deficit/hyperactivity disorder has a prevelance of 20 to 90 percent within clinic populations. However, the relationship between the two conditions is unclear and a common genetic substrate appears unlikely. Like obsessive-compulsive disorder, attention-deficit/hyperactivity disorder may be genetically heterogeneous and one form might represent part of the Tourette's syndrome diathesis.

Pathology and Laboratory Examination Like other movement disorders, Tourette's syndrome has long been suspected to involve pathology within the basal ganglia; however, neuropathological and biochemical studies have been small in number and inconclusive. Progress has recently been made with the creation of a consortium of academic centers to systematically collect and examine postmortem specimens. Evidence of subtle structural pathology was found in several recent MRI volumetric studies, revealing decreased size or alteration of the normal asymmetries of the basal ganglia. Among identical twins with Tourette's syndrome lower birth weight is associated with greater tic severity in most cases, indicating that prenatal events may influence the eventual severity of the condition. The dopaminergic system has also been implicated by the observed response of Tourette's syndrome symptoms to D_2 receptor antagonists, such as haloperidol (Haldol),

and the frequent exacerbation of symptoms by dopamine-stimulating agents, such as methylphenidate (Ritalin). Thus, it has been proposed that increased dopamine-receptor sensitivity, triggered perhaps by some prenatal event, may modulate clinical expression of the Tourette's syndrome diathesis. In vivo neuroimaging studies have failed to find an alteration in D_2 receptors in Tourette's syndrome patients compared to normal control subjects. However, a recent SPECT study of identical twins discordant for Tourette's disorder severity revealed a small but significant increase in binding of ^{123}I-iodobenzamide (a D_2-receptor antagonist) in the more affected twin of each pair. Differences in binding between twins almost completely predicted differences in tic severity, thus supporting the dopamine-receptor supersensitivity hypothesis. Elevations in the striatal dopamine transporter have also been detected in postmortem samples and by SPECT. All these studies were based on small samples and require larger scale replication. A post-infectious hypothesis for the pathophysiology of some cases of Tourette's disorder and related disorders has been proposed, based on descriptions of obsessive-compulsive symptoms in Sydenham's chorea, a sequela of untreated group A b-hemolytic streptococcal infection, and observation of new-onset tics after recent streptococcal infection. It is hypothesized that such infection may trigger an autoimmune response with antibodies directed against neurons in the basal ganglia. Studies are in progress to characterize antineuronal antibodies in Tourette's disorder. Preliminary reports have also described increased expression of a B lymphocyte surface antigen, highly associated with rheumatic fever, in children with Tourette's disorder, suggesting a potential susceptibility marker, at least in some individuals. Currently, there are no diagnostic neuroimaging, blood, or cerebrospinal fluid assays useful in establishing a diagnosis of Tourette's disorder.

Differential Diagnosis Tourette's disorder may be differentiated from other hyperkinetic movement disorders by the history of both motor and vocal tics and in the absence of other CNS symptoms, pathology, or exposure to neuroleptics or stimulants. Inquiry will frequently reveal a premonitory “irresistible urge.”

Course and Prognosis Over time, specific tics often disappear, only to be replaced by new tics in different locations; however, in about two thirds of individuals the trend is towards a gradual reduction in symptom severity as the individual enters adulthood, although only a minority of patients will experience a complete remission.

Treatment Treatment begins with education of the patient, family, and teachers about the condition. This will often suffice when Tourette's syndrome symptoms are mild. For more severe cases, and those with associated symptoms of obsessive-compulsive disorder or attention-deficit/hyperactivity disorder, pharmacological intervention may be required. Low doses of neuroleptics, particularly haloperidol and pimozide (Orap), have been used successfully for tic suppression, with adequate control achieved in up to 70 percent of patients. Most recently, risperidone has been tried with promising results. However, all the usual risks attendant to antipsychotic use for other neuropsychiatric disorders exist here as well. Clonidine (Catapres) has also proved modestly beneficial although less effective than the antipsychotic agents. Other agents, such as clonazepam (Klonopin), may help a smaller number of patients. Concomitant obsessive-compulsive disorder is treated with typical antiobsessional agents, such as clomipramine (Anafranil), fluoxetine (Prozac) or fluvoxamine (Luvox). Use of sympathomimetics such as methylphenidate (Ritalin) for concomitant attention-deficit/hyperactivity disorder may exacerbate tics or precipitate the first symptoms of a tic disorder; however, some patients may tolerate stimulant medications without worsening of tics, or only temporary exacerbation.

M.S. developed eye blinking and grimacing at age 4, followed soon thereafter by excessive throat clearing. He was initially treated for allergies. His symptoms evolved into abrupt exhalation tics, neck jerks, and touching tics, with the compulsion to perform the same maneuver with the opposite hand “to even things up.” He could briefly suppress the tics, but at the expense of mounting inner tension to release them. He also reported checking rituals (repetitively checking the knobs of the washing machine, checking to be sure appliances were unplugged, checking door locks), but was not severely bothered by them. His symptoms frequently vanished in physicians' offices, leading to consideration of a psychogenic basis. Other children mimicked and ostracized him and by adolescence he experienced periods of depressed mood and panic attacks. At the age of 32 the possibility of Tourette's disorder was raised by a friend who had seen a daytime television talk show on the subject. A professional referral was made and Tourette's disorder was diagnosed. It was noted that M.S.'s father had had facial grimacing and peculiar “mannerisms,” as did a paternal uncle. The patient elected not to take medication but benefited from education about Tourette's disorder and participation in a local Tourette's disorder support group. His function improved and he was able to attain a degree in engineering while working full time. This case is typical of many patients who remain undiagnosed or are misdiagnosed for years, further compounding a stressful situation. Clinicians may be misled by the suppressibility to suspect that the condition is under pure voluntary control, and exacerbation under stress may lead to suspicion that the symptoms have an emotional or neurotic basis. Some clinicians erroneously believe that motor and vocal tics must be present together, or that coprolalia is required for the diagnosis. Tourette's disorder has recently received widespread attention in the media, which may eventually lead a patient to the proper diagnosis, 30 years after symptom onset as in this case.

SUGGESTED CROSS-REFERENCES

Neuroimaging is discussed in Section 1.16, Section 2.13, and Section 51.2f. Cognitive disorders are discussed in Chapter 10, mood disorders in Chapter 14, obsessive-compulsive disorder in Chapter 15, attention-deficit/hyperactivity disorder in Section 51.3e. Psychotherapies are discussed in Chapter 30 and biological therapies in Chapter 31.

SECTION REFERENCES

Aarsland D, Tandberg E, Larsen JP, Cummings JL: Frequency of dementia in Parkinson's disease. Arch Neurol *53:*538, 1996.

Adams RD, Victor M: *Principles of Neurology*, ed 2. McGraw-Hill, New York, 1997.

Adler CH, Sethi KD, Hauser RA, Davis TL, Hammerstad JP, Bertoni J, Taylor RL, Sanchez-Ramos J, O'Brien CF, the Ropinirole Study Group: Ropinirole for the treatment of early Parkinson's disease. Neurology *49:*393, 1997.

Alexander GE, Crucher MD: Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. Trends Neurosci *13:*266, 1990.

Altzer HY, Kennedy J, Dai J, Parsa M, Riley D: Plasma clozapine levels and the treatment of L–dopa-induced psychosis in Parkinson's disease: A high potency effect of clozapine. In *Yearbook of Psychopharmacology*, vol 12, 1995.

Chiu HF: Psychiatric aspects of progressive supernuclear palsy. Gen Hosp Psychiatry *17:*135, 1995.

Cohen S, Freedman M: Cognitive and behavioral changes in the Parkinson's-plus syndromes. Adv Neurol *65:*139, 1995.

Cornford ME, Chang L, Miller BL: The neuropathology of parkinsonism: An overview. Brain Cogn *28:*32, 1995.

*Cummings JL, McPherson S: Neuropsychological aspects of Parkinson's disease and parkinsonism. In *Neuropsychological Assessment of Neuropsychiatric Disorders*, ed *2*, I Grant, KM Adams, editors. Oxford University Press, New York, 1996.

de la Tourette G: étude sur une affection nerveuse caractérisée par de l'incoordination motrice accompagnée d'écholalie et de coprolalie. Arch Neurologie *9:*158, 1885.

*DeLong Mahlon R: Primate models of movement disorders of basal ganglia origin. Trends Neurosci *13:*281, 1990.

Dogali M, Sterio D, Fazzini E, Kolodny E, Eidelberg D, Barrack A: Effects of posteroventral pallidotomy on Parkinson's disease. Adv Neurol *69:*585, 1996.

*Donaldson I, Anderson T: Now you see it; now you don't—The episodic movement disorders. NZ Med J *112:*23, 1999.

Doraiswami M, Martin W, Metz A, Deveaugh-Geiss J: Psychosis in Parkinson's disease: Diagnosis and treatment. Prog Neuropsychopharmacol Biol Psychiatry *19:*835, 1995.

Egan MF, Apud J, Wyatt RJ: Treatment of tardive dyskinesia. Schizophr Bull *23:*583, 1997.

Golbe LI, Davis PH: Progressive supernuclear palsy. In *Parkinson's Disease and Movement Disorders*, J Jankovic, E Tolosa, editors. Williams & Wilkins, Baltimore, 1993.

*Golbe LI, Sage JI: Medical treatment of Parkinson's disease. In *Treatment of Movement Disorders*, R Kurlan, editor. Lippincott, Philadelphia, 1995.

Gottwald MD, Bainbridge JL, Dowling GA, Aminoff MJ, Alldredge BK: New Pharmacotherapy for Parkinson's disease. Ann Pharmacother *31:*1205–1217, 1997.

Grossman RG, Hamilton WJ: Surgery for movement disorders. In *Parkinson's Disease and Movement Disorders*, J Jankovic, E Tolosa, editors. Williams & Wilkins, Baltimore, 1993.

Gusella JF, MacDonald ME, Ambrose CM, Duyao MP: Molecular genetics in Huntington's disease. Arch Neurol *50:*1157, 1993.

Harvey NS: Psychiatric disorders in Parkinsonism: Functional illnesses and personality. Psychosomatics *27:*91, 1996.

Howser RA, Freeman TV, Olanow CW: Surgical therapies for Parkinson's disease. In *Treatment of Movement Disorders*, R Kurlan, editor. JB Lippincott, Philadelphia, 1995.

Hyde T, Hotson JR, Kleinman JE: Differential diagnosis of choreiform tardive dyskinesia. J Neuropsychiatry *3:*255, 1991.

Huntington G: *On Chorea.* Lea & Blanchard, Philadelphia, 1872.

*Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72:*971, 1993.

Jankovic J: Treatment of Parkinsonian syndromes. In *Treatment of Movement Disorders*, R Kurlan, editor. JB Lippincott, Philadelphia, 1995, p 95.

Kalra S, Vergeron C, Lang AE: Lewy body disease and dementia: A review. Arch Intern Med *156:*487, 1996.

Klockgether T, Loschman PA, Wullner U: New medical and surgical treatments for Parkinson's disease. J Curr Opin Neurol *7:*346, 1994.

LeWitt PA: New options for the treatment of Parkinson's disease. Baillieres Clin Neurol *6:*109, 1997.

*Lombroso PJ, Scahill LD, Chappell PB, Pauls DL, Cohen DJ, Leckman JF: Tourette's syndrome: A multigenerational neuropsychiatric disorder. In *Behavioral Neurology of Movement Disorders*, WJ Weiner, AE Lang, editors. Raven Press, New York, 1995.

Marsden CD: Clinical experience with cabergoline in patients with advanced Parkinson's disease treated with levodopa. Drugs *55*(Suppl):17–22, 1998.

Martindale D, Hackam A, Wieczorek A, Ellerby L, Wellington C, McCutcheon K, Singaraja R, Kazemi-Esfarjani P, Devon R, Kim SU, Bredesen DE, Tufaro F, Hayden MR: Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. Nat Genet *18:*150, 1998.

Murphy TK, Goodman WK, Fudge MW, Williams RC Jr, Ayoub EM, Dalal M, Lewis MH, Zabriskie JB: B lymphocyte antigen D8/17: A peripheral marker for childhood-onset obsessive-compulsive disorder and Tourette's syndrome? Am J Psychiatry *154:*402, 1997.

Nomoto M, Kita S, Iwata SI, Kaseda S, Fukuda T: Effects of acute or prolonged administration of cabergoline on parkinsonism induced by MPTP in common marmosets. Pharmacol Biochem Behav *59:*717, 1998.

Nutt JG, Hammerstead JP, Gancher ST: *Parkinson's Disease: One Hundred Maxims.* Mosby Yearbook, St Louis, 1992.

Radunovic A, Porto WG, Zeman S, Leigh PN: Increased mitochondrial superoxide dismutase activity in Parkinson's disease but not amyotrophic lateral sclerosis motor cortex. Neuroscience Letters *239:*105, 1997.

Rathbun JK: Neuropsychological aspects of Wilson's disease. Int J Neurosci *85:*221, 1996.

Rinne UK, Bracco F, Chouza C, Dupont E, Gershanik O, Marti Masso JF, Montastruc JL, Marsden CD: Early treatment of Parkinson's disease with cabergoline delays the onset of motor complications. Results of a double-blind levo-dopa controlled trial. The PKDS009 Study Group. Drugs *55*(Suppl):23, 1998.

Robertson MM, Yakeley J: Gilles de la Tourette syndrome and obsessive-compulsive disorder. In *Neuropsychiatry*, BS Fogel, RB Schiffer, SM Rao, editors. Williams & Wilkins, Baltimore, 1996, p 827.

Shults CW, Beal MF, Fontaine D, Nakano K, Haas RH: Absorption, tolerability, and effects on mitochondrial activity of oral coenzyme Q-10 in parkinsonian patients. Neurology *50:*793, 1998.

Snyder SH, D'Amato RJ: MPTP: A neurotoxin relevant to the pathology of Parkinson's disease. Neurology *35:*250, 1986.

Starosta-Rubenstein S: Treatment of Wilson's disease. In *Treatment of Movement Disorders*, R Kurlan, editor. JB Lippincott Co, Philadelphia, 1995.

Swedo SE, Leonard HL, Mittleman BB, Allen AJ, Rapoport JL, Dow SP, Kanter ME, Chapman F, Zabriskie J: Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. Am J Psychiatry *154:*110, 1997.

Taylor AE, Saint-Cyr JA: Neuropsychology of Parkinson's disease. Brain Cogn *28:*281, 1995.

Wolf SS, Jones DW, Knable MB, Gorey JG, Lee KS, Hyde TM, Coppola R, Weinberger DR: Tourette syndrome: Prediction of phenotypic variation in monozygotic twins by caudate nucleus D_2 receptor binding. Science *273:*1225, 1996.

2.7 NEUROPSYCHIATRIC ASPECTS OF MULTIPLE SCLEROSIS AND OTHER DEMYELINATING DISORDERS

WILLIAM W. BEATTY, PH.D., AND ROBERT H. PAUL, PH.D.

Myelin is a specialized extension of the central nervous system (CNS) oligodendrocyte plasma membrane. Composed of about 70 percent lipid and 30 percent protein, the known function of myelin is to make saltatory conduction possible. If damage to the myelin surrounding a nerve is limited, propagation of action potentials will be slowed; more extensive damage may result in conduction block. Because myelinated axons carry important components of the sensory inputs to the brain (e.g., the optic nerves and tracts), the motor outputs from the brain (e.g., the corticospinal tracts), as well as communications within the brain (e.g., in the corpus callosum), the symptoms of demyelination vary greatly from one patient to another.

A large number of CNS disorders can affect the integrity of the white matter (myelin). These include traumatic brain injury, especially closed head injury, vascular disorders, especially Binswanger's disease, infectious diseases such as acquired immune deficiency syndrome (AIDS); exposure to various neurotoxins such as alcohol or volatile inhalants; various inherited disorders of lipid metabolism (e.g., the leukodystrophies); and multiple sclerosis.

Because the disorders that affect the white matter vary greatly in their etiology and epidemiology, there are marked differences in their neurological, cognitive, and psychiatric consequences. However, white matter disorders share two common features: (1) they are rarely focal; instead there are usually multiple sites of injury to the CNS and, (2) there are enormous differences from one patient to another in the mixture of neurological, cognitive, and psychiatric effects of white matter diseases.

MULTIPLE SCLEROSIS

History and Epidemiology Multiple sclerosis, the most common demyelinating disease, affects 250,000 to 350,000 persons in the United States. Jean-Martin Charcot, the French physician who made pioneering contributions to both neurology and psychiatry, provided a remarkably accurate description of the major neurological, cognitive, and psychiatric features of multiple sclerosis. Charcot realized that multiple sclerosis is often characterized by exacerbations in which old neurological symptoms worsen and new symptoms appear, followed by periods of remission in which symptoms abate, sometimes to the point of complete recovery. Because neuropathological studies revealed the presence of multiple lesions distributed throughout the white matter of the CNS, the name *multiple* or *disseminated sclerosis* was applied. To distinguish multiple sclerosis from other conditions with similar symptomatology, the clinical diagnosis requires evidence (radiological or by examination or history) of at least two distinct episodes of neurological impairment that cannot be explained by a single CNS lesion.

Multiple sclerosis affects about twice as many women as men and is most prevalent among whites of northern European descent. Prevalence also increases with increasing distance from the equator. The latitude of birth and childhood residence rather than of adult residence seems more important for determining risk of developing the disease.

About 85 percent of cases of multiple sclerosis are diagnosed between age 20 and 50; however, with more widespread availability of magnetic resonance imaging (MRI), a small but increasing number of cases among children and teenagers are being reported.

MRI has greatly increased the precision of diagnosing multiple sclerosis. MRI can reveal evidence of lesions that are disseminated in space, even when clinical evidence to identify multiple distinct episodes separately in time is lacking. However, the white matter changes (Fig. 2.7-1) observed on MRI scans are not specific to multiple sclerosis. Similar changes occur in other neurological conditions and to some extent in the general population, especially the elderly. Consequently, MRI, even if evidence of lesions is found in the optic nerves, long tracts of the spinal cord, or periventricular white matter (the most frequent sites of multiple sclerosis lesions) is only suggestive. Confirmation of the diagnosis by clinical and other laboratory measures is essential. The most useful laboratory tests are visual or somatosensory evoked potentials and analysis of g-globulin and myelin basic protein in the cerebrospinal fluid. In multiple sclerosis patients sensory evoked potentials are slowed or absent and g-globulin and myelin basic protein levels are elevated, especially during an exacerbation. As with MRI, none of these laboratory measures is specific for multiple sclerosis and not all multiple sclerosis patients show any or all of these changes.

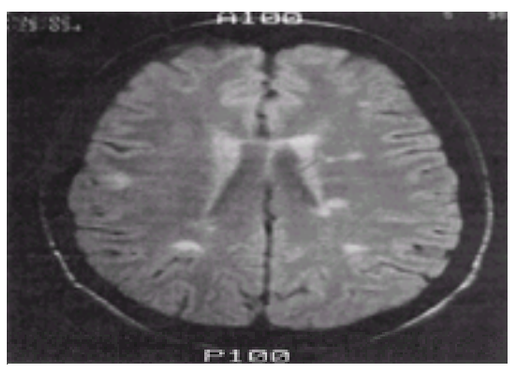

FIGURE 2.7-1 T2-weighted axial MRI scan from a 30-year-old female patient with multiple sclerosis. Many areas of increased signal intensity, which appear white, are scattered throughout the white matter, which appears gray. This is the typical appearance of multiple sclerosis lesions on MRI. (Courtesy of Don Wilson, M.D.)

The course of multiple sclerosis is just as variable and unpredictable as the presenting symptoms. In addition to the *relapsing-remitting* type in which relapses (exacerbations) that can affect sensory or motor systems alternate with periods of remission in which recovery to nearly premorbid functioning occurs, there are three other forms or subtypes of the disease. In the *benign* form, relapses are infrequent, affect mostly sensory functions, and show almost complete recovery between attacks. In the *chronic progressive/secondary progressive type* there is either a slowly deteriorating course from disease onset (chronic progressive) or a relapsing-remitting course with incomplete recovery after exacerbations (secondary progressive). In the *primary-progressive* type there is a steadily and often rapidly worsening course from the onset of the disease. The modest reduction in average life expectancy of multiple sclerosis patients as a whole is mainly accounted for by patients with this disease type. There is some evidence that the distribution of lesions is different in the primary-progressive type than in other forms of the disease, but there is no good evidence that diagnostic type (other than primary-progressive), even if assigned according to operational criteria, is of much value in predicting the future disease course or level of functional impairment of individual patients. One reason why this is so is that it is not uncommon for patients whose disease has initially followed a true relapsing-remitting course (i.e., full recovery after exacerbations) to enter a period of slow decline. Also, patients with chronic progressive multiple sclerosis may enter periods of stability in which further functional losses do not occur; such periods may last for months or years.

The consequence is that most multiple sclerosis patients experience a great deal of uncertainty about both the short-term and long-term courses of their disease.

Etiology and Mechanism of Demyelination Multiple sclerosis is now widely accepted to be an autoimmune disorder in which T lymphocytes (and a few B lymphocytes) invade the CNS and attack the myelin. The details of initiation and termination of the attack are still unknown, but there is considerable evidence that the proinflammatory cytokines, interleukin-1 (IL-1), tumor necrosis factor, and interferon gamma play an important role. Other regulatory cytokines such as interferon beta

(IFN-b) may counteract the effects of the proinflammatory interferons.

Research in this area is progressing rapidly, aided by the availability of a useful animal model. Experimental allergic encephalomyelitis is produced by autoimmunizing animals against myelin basic protein. Animals with encephalomyelitis exhibit the motor symptoms and demyelination typical of human multiple sclerosis.

The final pathway of myelin breakdown in the CNS involves splitting the tightly compacted membrane lamellae and enzymatic attack of the myelin proteins and lipids. The proteolytic and lipolytic enzymes may invade the CNS from the blood through gaps in the blood-brain barrier, be secreted by inflammatory cells, or originate from lysosomes from macrophages, microglia, or astrocytes at the target site in the CNS. The consequence is the formation of a lesion (plaque) as shown in Figure 2.7-2 and Figure 2.7-3.

FIGURE 2.7-2 Two cuts in the axial plane from the formalin-fixed brain of a 38-year-old man with multiple sclerosis who died of sepsis secondary to a urinary tract infection. Several areas of demyelination are visible as gray areas within the white matter. (Courtesy of Roger Brumback, M.D.)

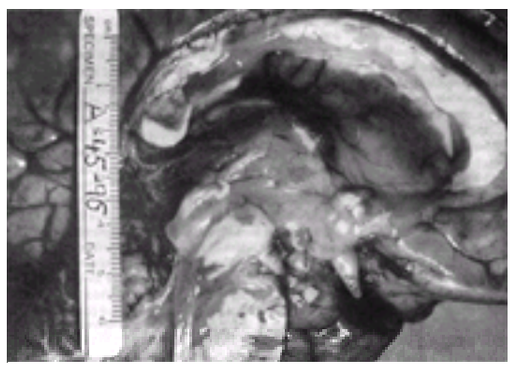

FIGURE 2.7-3 Midsaggital section through the brain of the same patient as in Figure 2.7-2. Extensive demyelination of the corpus callosum is evident that is complete in the posterior part of the body of that structure. Note that demyelination of the ventral part of the corpus callosum is more extensive, reflecting the predilection of multiple sclerosis lesions to involve periventricular regions. (Courtesy of Roger Brumback, M.D.)

Plaques may be immunologically active, characterized by cuffs of lymphocytes and macrophages, or immunologically silent in terms of inflammation. MRI technology has developed to the point where it is possible to distinguish active lesions from older inactive lesions, which are comprised of astrocytes. This permits in vivo monitoring of disease activity.

Susceptibility to multiple sclerosis seems to be influenced by a genetic factor or factors because concordance rates for the disease are about 30 percent for identical twins compared to 2 percent for fraternal twins or siblings of the same sex. Several potential markers have been proposed, but no specific gene has been identified. Based on the twin studies it is clear that any genetic influence is likely to affect susceptibility to multiple sclerosis, while also interacting with other factors.

Since the time of Charcot the possibility that exposure to stressful events might precipitate the initial episodes of multiple sclerosis or trigger exacerbations has attracted considerable interest. The idea is plausible because multiple sclerosis patients are especially susceptible to the debilitating effects of heat. Further, in multiple sclerosis, like other neurological diseases, symptoms may worsen during times of minor infection. Addressing the first question (i.e., that stress provokes the initial multiple sclerosis attack) requires a retrospective design that is subject to recall bias. Moreover, there are difficulties in establishing the time of disease onset, in defining and measuring stressful experiences, and in selecting an appropriate comparison group. Several older as well as two recent studies published since 1980 have reported that patients with multiple sclerosis recalled experiencing more stressful events than matched controls in the period (up to 2 years) before disease onset.

Determining if a relationship between stress and exacerbations exists is potentially more feasible because a prospective design can be used. In general, no study has found that patients who exhibited exacerbations were exposed to more stress than those who did not experience exacerbations and that periods of increased exposure to stress preceded exacerbations. Despite the absence of empirical evidence to support a link between stress and the onset of multiple sclerosis or the triggering of exacerbations, it is fair to say that many patients and health professionals believe such an association exists.

Psychiatric Features Psychiatric disturbance in multiple sclerosis patients has been consistently reported since Charcot's initial description of the disease. In his seminal work Charcot noted that depression, pathological laughter, and “stupid indifference” were psychiatric symptoms characteristic of multiple sclerosis. Current research continues to identify the presence of these and similar psychiatric disorders in the multiple sclerosis population.

Most reported prevalence rates of depression in multiple sclerosis range from 45 to 62 percent. This variability probably results from the use of assessment techniques that differ qualitatively. Studies that assess depression in multiple sclerosis using symptom inventories (e.g., the Beck Depression Inventory) typically report high prevalence rates. However, these studies probably overestimate the prevalence of depression because many symptom inventories inadvertently assess physical and cognitive symptoms of multiple sclerosis (e.g., fatigue, memory disturbances). In support of this contention, studies have shown that when multiple sclerosis patients are assessed using structured clinical interviews or symptom inventories that exclude nonmood items the prevalence of depression drops substantially. This does not imply, however, that depression is uncommon in multiple sclerosis patients. Even the most conservative prevalence estimates report rates of depression that are higher than those typically found in the general population.

The cause of depression in multiple sclerosis remains unclear. Typically, depression in multiple sclerosis is believed to represent either a primary manifestation of the disease caused by cerebral involvement, or a psychological reaction to living with a debilitating and chronic illness; research does not support a genetic basis for depression in multiple sclerosis.

Several lines of evidence indicate that depression in multiple sclerosis patients results from cerebral demyelination. First, multiple sclerosis patients report high rates of depression when compared to other populations of patients with chronic illnesses. Second, cerebral involvement is more closely related to depression in multiple sclerosis than is spinal cord involvement. Third, depression in multiple sclerosis is unrelated to neurological status (i.e., disease duration, course, and severity of disability), and to degree of cognitive impairment.

It should be noted, however, that most studies cited as evidence for a biological basis of depression in multiple sclerosis have assessed depression with symptom inventories, which might estimate the prevalence of depression inaccurately. Further, measures of patients' perceptions regarding their disabilities are notably lacking from most studies. This is an important limitation because, depending on the patient, multiple sclerosis-induced cognitive impairment might be more distressing than

physical disability.

Reactive explanations for depression in multiple sclerosis are based on the theory of learned helplessness, which states that depression may develop when people are confronted with aversive situations that are unpredictable or uncontrollable. Because multiple sclerosis patients have limited control over the course of their illness and the disease affects them both physically and cognitively, learned helplessness theory is a plausible explanation for the depression that occurs in multiple sclerosis. The validity of this explanation is supported by the finding that patients' degree of social support significantly predicts their mental health status.

Early studies investigating psychiatric disturbance in multiple sclerosis reported high prevalence rates of pathological laughter and crying, but more recent studies have not confirmed this finding. Pathological laughter and crying in multiple sclerosis is phenomenologically similar to pseudobulbar palsy, a disorder that results from damage, usually bilateral, to fiber tracts connecting the cortex to subcortical forebrain structures. In pseudobulbar palsy, the behavioral expression of a mood state is incongruent with the patient's cognitions; for example, a person might cry uncontrollably, yet deny feeling sad.

Contemporary reviews of multiple sclerosis psychopathology have considered “stupid indifference” to be similar to *la belle indifférence*, a feature that is usually associated with conversion disorder. Patients with *la belle indifférence* fail to appreciate the significance of their impairments. Psychodynamic theories contend that *la belle indifférence* represents patients' denial or repression of their illness. From a neurological perspective, however, *la belle indifférence* is similar to *anosodiaphoria*, a condition in which patients with cerebral injuries are unconcerned about their neurological deficits. Because multiple sclerosis usually affects cerebral tissue, anosodiaphoria is probably a more accurate account of “stupid indifference” in multiple sclerosis than is *la belle indifférence.* Anosodiaphoric multiple sclerosis patients (often referred to as *euphoric* patients) are typically described as “content and comfortable” despite having serious physical and mental disabilities. The prevalence of anosodiaphoria in multiple sclerosis is uncertain, but research suggests that most patients are aware of their disabilities. MRI studies in patients who do not express anosodiaphoria have shown that it is related to the extent of lesion involvement in the frontal lobes.

Epidemiological studies have found that multiple sclerosis patients are twice as likely as healthy controls to have bipolar I disorder. Although the cause of this comorbidity has not been definitively determined, some evidence indicates that both disorders might have a shared genetic basis. For example, in one study family histories of bipolar I disorder were investigated in three groups of multiple sclerosis patients: (1) patients with bipolar I disorder, (2) patients with depression or mania, and (3) patients with no psychiatric disturbance. Results of the study indicated that family histories of both multiple sclerosis and bipolar I disorder were more common in patients with bipolar I disorder than either of the other two groups.

Fatigue Fatigue is one of the most common and debilitating symptoms of multiple sclerosis. Surveys have shown that more than 80 percent of multiple sclerosis patients report experiencing fatigue on a daily basis. Although most studies have focused on physical fatigue, evidence suggests that cognitive fatigue might also be problematic for some individuals.

Symptoms of fatigue in multiple sclerosis patients tend to worsen as the day progresses, making late afternoon a particularly difficult time of day. Heat, physical exertion, and stress also exacerbate the severity of fatigue in multiple sclerosis. Because of the nature of fatigue in multiple sclerosis many patients must complete activities early in the day and avoid some events altogether. For these reasons, it is not surprising that many patients consider fatigue to be the most disabling symptom of their disease.

The cause of fatigue in multiple sclerosis is unknown. Although fatigue is a common symptom of depression, studies have shown that fatigue in multiple sclerosis is qualitatively distinct from the fatigue that is characteristic of mood disorders. Results from several studies indicate that fatigue is more common in multiple sclerosis patients with greater physical disability and a progressive disease course; however, some patients present with fatigue as the initial symptom of the disease. MRI studies have recently shown that fatigue in multiple sclerosis is associated with greater lesion volume in the brainstem and midbrain. In addition, fatigue in multiple sclerosis has been shown to be inversely related to pineal gland calcification, which implies that fatigue in multiple sclerosis might result from altered melatonin secretion.

Cognitive Impairment Charcot accurately described three of the major aspects of cognitive impairment that affect some multiple sclerosis patients—poor memory, poor abstracting, and slowed information processing. Yet as recently as 1982, even research neurologists believed that these impairments were uncommon and confined to multiple sclerosis patients with severe physical disabilities. More recent research indicates that 40 to 65 percent of multiple sclerosis patients will exhibit some kind of cognitive deficit on neuropsychological testing. Further, simple screening tests like the Mini Mental State Exam are insensitive to the cognitive impairments most common in multiple sclerosis. Likewise, measures of neurological impairment (e.g., the widely used Expanded Disability Status Scale) or disease duration are not useful predictors of the cognitive status of individual patients. Mild neurocognitive disorder may be evident after the first exacerbation (i.e., before the clinical criteria for diagnosis are fulfilled).

Intelligence More than 35 years ago it was reported that men who developed multiple sclerosis while in military service showed a drop of 13.5 I.Q. points over a period of about 4 years. Subsequent longitudinal studies that lacked the advantage of premorbid data observed much smaller declines or even modest increases in I.Q. over intervals of 1 to 3 years. Because most multiple sclerosis patients referred for neuropsychological evaluation perform within normal limits on standard I.Q. tests, many concluded that I.Q. was unaffected in multiple sclerosis. More recent work indicates that, on average, multiple sclerosis patients scored somewhat lower than demographically matched controls on verbal I.Q. and significantly lower on performance I.Q. The difference between verbal and performance I.Q. probably arises because the performance tests are timed whereas the verbal tests generally are not. Hence, the performance tests require both speed and accuracy whereas the verbal tests emphasize accuracy. In general, verbal I.Q. is only slightly affected in multiple sclerosis and patients with very low scores are likely to exhibit global impairments on other cognitive tests.

Language Language is also well preserved in multiple sclerosis. Aphasia, agraphia, and alexia are extremely rare, although problems of speech production (hypophonia, dysarthria) are common. Most patients perform normally on tests of confrontation naming; patients with significant naming impairments are likely to exhibit deficits on several other cognitive tests. By contrast, impairments on verbal fluency tests, which require the subject to generate words that begin with a particular letter or exemplars of a certain category, are typically observed in patients with multiple sclerosis. Although verbal fluency tests are often considered language tests, it is probably better to think of them as measures of speed of access to established knowledge about words.

Attention and Information Processing The status of attention in multiple sclerosis patients is unclear. Many studies have reported significant attentional impairments in multiple sclerosis; however, other studies have not found similar results. Whether or not multiple sclerosis patients exhibit attentional deficits is partly determined by the type of attention tested. For example, multiple sclerosis patients typically perform within the normal range on attention tests that require automatic processing (e.g., recognition of letters). On more demanding attention tests, such as those that require working memory (i.e., controlled processing tasks), multiple sclerosis patients have been found to perform significantly worse than controls, but deficits are not consistently observed on all tasks. For example, only half of the published studies that have examined controlled processing in multiple sclerosis with the Wechsler Adult Intelligent Scale-Revised (WAIS-R) Digit Span subtest have found multiple sclerosis patients to perform significantly worse than healthy controls.

Deficits are more consistently observed on tests that require multiple sclerosis patients to focus on two or more stimuli simultaneously (i.e., divided-attention tasks). Many studies have shown that multiple sclerosis patients perform significantly worse than controls on the Paced Auditory Serial Addition Task, which requires subjects to add digits to numbers previously given. Similar deficits are observed on tests of comparable difficulty (e.g., the Stroop test, which requires subjects to state aloud as quickly as possible the color of the ink in which the name of a color is printed).

Information-processing speed is also impaired in multiple sclerosis patients. One of the most commonly employed instruments to examine information-processing speed in multiple sclerosis is the Symbol Digit Modalities Test, a verbal analogue of the Digit Symbol subtest of the WAIS-R. On this task, subjects orally substitute numbers for symbols according to a key that is kept in the subject's view. Performance on this task depends on how quickly subjects can make the correct substitutions in a specified time period. Studies that have investigated information-processing efficiency with this task have consistently reported that multiple sclerosis patients perform significantly worse than healthy controls.

Memory Impairments in memory are the most extensively studied of the cognitive deficits that accompany multiple sclerosis. Memory difficulties also constitute the class of cognitive dysfunction that multiple sclerosis patients most often complain about. Problems in concentration are a close second, and the two deficits may be closely linked. Modern neuropsychological theory makes distinctions between several different types of memory and the empirical evidence indicates that certain forms of memory are much more likely to be affected by multiple sclerosis than others.

Explicit memory tests involve conscious recollection of specific events or more general knowledge acquired in the recent or distant past. Examples include reciting a

list of words, recalling the name of a famous person when shown a photograph, or naming the high school that one attended. On implicit memory tests, subjects can display evidence of retention without conscious awareness. For example, patients with amnesia from diencephalic or temporal lobe injury learn new motor skills normally, but are unable to remember having performed the tasks previously.

Memory tests may include tasks designed to study learning and memory for information acquired after the onset of disease *(anterograde memory)* or for information acquired before disease onset (*remote* or *retrograde memory*). There are an enormous variety of tests devised to study anterograde memory. Examples of explicit memory tests include word and picture lists, short stories, and production of drawings of forms or shapes that were previously copied. Anterograde implicit memory tests include learning new motor skills and various priming tasks. On priming tests, exposure to stimuli alters the speed or accuracy of subsequent responses, but subjects do not appear to consciously recollect the original exposure. Remote memory is usually tested by asking subjects to identify pictures of people who were famous in the past, to recall past public events, or to answer structured questions about episodes or events that occurred during various periods in their own lives. There are no standard tests for remote implicit memory.

Anterograde memory tests are also classified in terms of the length of the retention interval. For present purposes, *immediate memory* will refer to tests on which the retention interval is 0 to 60 seconds in length; *recent memory* will refer to retention intervals that are longer than 60 seconds. In the research literature *short-term memory* and *working memory* are approximately synonymous with *immediate memory*, and *long-term memory* means roughly the same as *recent memory.*

IMMEDIATE MEMORY The most widely used measure of immediate memory is the Digit Span test from the WAIS-R, which is also a measure of attention and concentration. The test is comprised of two subtests: *Digits Forward* (subjects are read a string of digits that they attempt to recite in order) is a relatively pure measure of attention and immediate memory; *Digits Backwards* (subjects attempt to recite the list of digits in reverse order) also requires manipulation of information in memory.

The performance of multiple sclerosis patients on the Digit Span test has been inconsistent from study to study, with mild deficits being the most common finding. The most likely explanation for the conflicting results is the heterogeneity of the samples studied. However, illness variables cannot account for the inconsistencies in results because deficits have been reported for patients with very mild physical disabilities and intact performance has been exhibited by more severely disabled patients.

On other tests of immediate memory, deficits by multiple sclerosis patients have been observed consistently. These tests have the common property of requiring subjects to divide their attention between the items to be remembered and some other task.

RECENT MEMORY When groups of multiple sclerosis patients are asked to recall information presented minutes, hours, or days earlier, impairments are generally evident. Comparable deficits have been observed using a variety of verbal stimuli including short prose passages, associated word pairs, and lists of words as well as nonverbal stimuli such as figures and the location of stimuli in spatial arrays. On multi-trial learning tests deficits are usually evident on the first trial. The deficits that are readily apparent when memory is tested by recall methods are less clear when recognition methods are employed.

REMOTE MEMORY The most common clinical test of remote memory requires subjects to recite the names of recent heads of state beginning with the current office holder and working backward in time. On this test (the *presidents test* in the United States) multiple sclerosis patients perform as well as demographically matched controls. However, on tests requiring identification of famous persons from photographs or recall of past public events multiple sclerosis showed consistent impairments that were equally severe for all past decades. Less severe impairments for recognition memory of the same information were also seen.

Such tests of remote memory require display of knowledge that is assumed to be widely shared among members of a culture. However, on these tests it is impossible to tell if a failed item was ever known by an individual patient. Tests of autobiographical memory circumvent this problem because they tap information that was certainly known by all subjects. Recent research indicates that multiple sclerosis patients exhibit mild impairments in autobiographical memory; however, their deficits are so modest that they are almost certainly clinically insignificant. The existence of small deficits that extend back into childhood memories indicates that the impairments seen on other tests are probably real deficits in remote memory.

IMPLICIT MEMORY The little information that is available indicates that implicit memory is intact in multiple sclerosis patients. This is true for motor-skill learning and several priming paradigms.

AWARENESS OF MEMORY DEFICITS Clinical experience suggests that some patients with multiple sclerosis are aware of their impairments in memory and adopt effective strategies for minimizing the impact of these difficulties in their everyday lives. Other patients with equally serious impairments are either unaware of their memory deficits or deny their existence.

Metamemory is a term that refers to knowledge of one's own memory. Experimental studies of metamemory require subjects to predict their own ability to recall or recognize newly acquired information or items of general knowledge. Subjects' predictions are then compared to their performances on memory tests. Thus, it is possible for a patient to exhibit impaired memory, but accurate (or inaccurate) metamemory. It is also possible for a patient with normal memory to show impaired (or intact) metamemory.

Studies of metamemory in multiple sclerosis indicate that the severity of cognitive impairment and the type of memory task influences the accuracy of patients' metamemory performances. For anterograde memory tests, patients with deficits in both memory and problem solving showed the largest deficits in metamemory; these patients overestimated their actual memory performance by more than 50 percent. Less serious metamemory deficits were shown by multiple sclerosis patients with either memory or problem-solving impairments; patients without either deficit were as accurate as controls in forecasting their anterograde memory performance. Although the subgroups of multiple sclerosis patients varied in the accuracy of performance on the tests of general knowledge, regardless of the extent of their cognitive impairments all patient groups could identify which items they would recognize as accurately as normal controls.

MODELS OF MEMORY DYSFUNCTION The anterograde memory deficits associated with multiple sclerosis have usually been attributed to patients' difficulties in retrieving information that has been encoded and stored. Support for this idea comes from the usual finding that memory impairments are more conspicuous on recall than on recognition tests, the absence of a temporal gradient of retrograde amnesia on remote memory tests, consistent deficits on verbal fluency tests, and evidence that fundamental mechanisms such as use of imagery and the capacity to encode verbal stimuli semantically are intact. Studies using the *selective reminding* technique provide direct evidence that the major problem in verbal memory for multiple sclerosis patients is an inability to retrieve information from long-term storage consistently.

An alternative explanation maintains that the verbal memory impairment in multiple sclerosis patients is simply a failure to learn the information adequately. When multiple sclerosis patients were given additional trials to bring their performance up to the level of controls, there was no deficit in delayed recall.

Both these theories are based on studies that compared multiple sclerosis patients to demographically matched controls, a tactic that ignores the heterogeneity of multiple sclerosis patients. When the statistical technique of cluster analysis is applied, quite a different picture emerges. In a recent study using the Selective Reminding technique about 25 percent of the multiple sclerosis patients had no memory deficits at all; another 25 percent had severe memory deficits that resembled those of patients with amnesia. The remaining patients on average showed patterns of performance consistent with the theory that retrieval failure was the major source of their memory problems. However, closer examination of individual patient profiles revealed that only 2 of 51 patients with mild memory deficits met stringent operational criteria for retrieval failure. The acquisition deficit hypothesis fared only a little better: 4 patients out of 51 met stringent operational criteria for acquisition failure.

As a practical matter, the embarrassment of these theories is of little consequence. Patients with an amnestic-like pattern are also likely to suffer impairments in a number of other cognitive domains. Attempts at memory rehabilitation by compensatory techniques are less likely to succeed with these individuals. Patients with milder memory deficits are also less likely to have global cognitive impairments and are more likely to learn to circumvent their memory problems.

Problem Solving Patients with multiple sclerosis exhibit a consistent pattern of deficits on both verbal and nonverbal tests of abstraction and problem solving. Measures that reveal deficits include the Category Test and the Wisconsin Card Sorting Test, two standard measures widely used in clinical neuropsychology, as well as in several experimental tasks. Analysis of patterns of errors, especially on the Wisconsin Card Sorting Test, has emphasized the importance of perseverative responding (i.e., inability to abandon incorrect hypotheses) in the problem-solving impairment. However, multiple sclerosis patients also exhibit deficits on problem-solving tasks on which perseverations do not occur. Thus, difficulties in forming concepts, an impairment described by Charcot, seem to be the most general

source of problem-solving difficulties.

Visual Processing Tasks Patients with multiple sclerosis are often mildly impaired in judging the orientation of lines, discriminating faces, and comprehending emotions conveyed by facial expressions. These deficits are commonly observed on the Performance scale of the WAIS-R, which emphasizes visual processing. There is no convincing evidence that any of these impairments reflect anything more than sensory or motor disturbances.

Impact on Work and Social Functions Surveys of multiple sclerosis patients in North America and Europe indicate that as many as 80 percent retire from gainful employment prematurely. Physical impairment and susceptibility to fatigue are obvious factors that contribute to premature retirement, but cognitive impairment, especially deficits in anterograde memory and information-processing speed, is an important and independent predictor of early retirement. Because the age of onset of the disease is early and the impact on life expectancy is slight, premature retirement caused by multiple sclerosis imposes a significant financial burden on patients and society.

Like any chronic disease, multiple sclerosis also places many burdens on spouses and families. These include loss of income, loss of partnership, and increased responsibilities for providing care to the patient and parental supervision of children. Although patients' physical limitations are usually obvious to family members, their susceptibility to fatigue and the cognitive and emotional difficulties that often accompany the disease may not be properly comprehended by family members. Each of these disease features has been thoroughly documented in the research literature, yet physicians and other professionals who care for multiple sclerosis patients often do not take the time to ensure that family members understand the array of changes that can accompany multiple sclerosis. Lack of understanding by family members may lead them to misattribute forgetfulness or chronic tiredness to deliberate attempts by multiple sclerosis patients to avoid responsibility or to solicit undeserved attention.

Screening for Cognitive Impairment Cognitive impairments can occur very early in the course of multiple sclerosis and contribute significantly to a multiple sclerosis patient retiring prematurely from the work force. In an ideal world all patients with multiple sclerosis would have regular neuropsychological evaluations to detect cognitive impairments early when compensatory strategies might be most likely to be effective. However, a complete neuropsychological evaluation is expensive and 25 to 40 percent of multiple sclerosis patients do not have cognitive impairments that are severe enough to impair their ability to function in everyday life.

Since 1992, several brief screening examinations that have high sensitivity and specificity for detecting cognitive impairments in multiple sclerosis patients have been developed. One of these batteries can be easily administered and scored by personnel with little or no formal training in neuropsychological testing. The proper use of a cognitive screening examination is similar to any medical screening test: to determine which patients should receive more extensive and more expensive diagnostic procedures. More widespread use of screening tests should lead to more efficient use of neuropsychological resources, reserving these relatively expensive tools for patients who are most likely to benefit from a complete assessment of cognitive function.

Mary D. (not her real name) was a 36-year-old woman who worked as a psychiatrist at an urban medical school when she volunteered as a subject in one of the authors' studies. Her first symptom (blurred vision) had occurred 12 years earlier while she was a medical student. Subsequently, she developed numbness and weakness and was diagnosed with multiple sclerosis 4 years later. Her symptoms remained fairly mild and relapses were infrequent throughout medical school, two residencies, and the first few years as an academic psychiatrist. However, Mary D. noticed that her increasing memory problems were compromising the quality of her work. Specifically, after seeing several patients consecutively, she could not remember important details of the interviews and she feared that she was confusing information about one patient with that of another. She volunteered as a research subject in a project on cognition in MS, in part to determine whether her suspicions about her own memory loss were valid.

During neuropsychological testing, which included a 2-hour battery of measures known to be sensitive to the impairments often seen in multiple sclerosis, Mary D. exhibited only mild walking difficulties, normal vision, and essentially no depression. In addition to memory problems, she complained of fatigue for which she was taking fluoxetine (Prozac) and amantadine (Symmetrel).

Testing revealed clear deficits in both verbal and nonverbal memory, information-processing speed, and abstraction problem solving; other cognitive functions were within normal limits.

As a result of the testing, Mary D. made a minor change in her professional life. She insisted that a 10-minute break be imposed between patients seen in her practice. This allowed her to dictate initial notes about each interview, circumventing her memory difficulties. At this writing (3 years after testing), Mary D. continues to work full time as a psychiatrist, although she left the medical school to take a more lucrative position in the private sector.

Mary D.'s case is noteworthy because she has been able to perform an intellectually demanding job capably despite deficits in three cognitive domains. It might be argued that the neuropsychological tests that she failed are irrelevant to job performance, but this is unlikely because the same memory and speed of information-processing tests she performed poorly on were significant predictors of employment status in the larger group of multiple sclerosis patients in the study in which she participated. Her mild sensory and motor symptoms no doubt made it easier for her to continue to work, but she experienced the considerable fatigue that is common in multiple sclerosis.

Three other factors seem important. First, she liked her job and was highly motivated to continue working; second, she suspected her cognitive deficits and accepted the test results as confirmation and an indication of the need to modify her work schedule in order to accommodate her deficits; third, she had the power to make the necessary adjustments in the way she conducted her work in order to circumvent her cognitive limitations.

Mary D.'s case illustrates what can happen and what health professionals could help to happen more frequently. What is needed is an appraisal of patients' capabilities as well as their deficits, an analysis of requirements of their jobs, and the possibility that each patient's job could be reconfigured or done in a different way. The nature of work in developed countries has changed enormously since Charcot first described multiple sclerosis. At that time most jobs required physical strength or fine-motor skill to execute highly overlearned responses. Memory and the ability to process information quickly were not very important for job performance, which may explain why neurology almost “forgot” that multiple sclerosis causes cognitive impairments for more than 100 years. Today the relative importance of cognitive abilities for employment success is clearly greater than it was 100 or even 25 years ago. The wide availability of computers and the increasing interest of many businesses in developing programs in which employees do most of their work at home offer enormous opportunities for allowing patients with multiple sclerosis (and other chronic conditions) to maintain employment.

Multiple Sclerosis as Subcortical Dementia Several authorities have suggested that multiple sclerosis gives rise to a pattern of cognitive and emotional deficits typical of the subcortical dementias associated with disorders like Huntington's disease. Consistent with this idea, many multiple sclerosis patients display cognitive deficits in processing information quickly, solving novel problems, visuospatial skills, and anterograde memory recall, and have personality and mood disturbances. The agnosias, apraxias, and aphasias often observed in cortical dementias such as dementia of the Alzheimer's type rarely occur in multiple sclerosis; as noted earlier, multiple sclerosis patients sometimes perform poorly on tests of confrontation naming. Such patients are likely to exhibit deficits in other cognitive domains, but they rarely show unusual difficulties in finding words to communicate in ordinary conversation.

Although these considerations can be taken to support the idea that multiple sclerosis gives rise to a form of subcortical dementia, in reality the situation is more complicated. First, while group studies depict a pattern of impairments and strengths that is prototypical of subcortical dementia, only a small proportion of patients, considered as individuals, meet suggested criteria for subcortical dementia (e.g., deficits in three or more different cognitive domains); many patients show no cognitive impairments at all. Second, as the case of Mary D. demonstrates, some multiple sclerosis patients who meet operational criteria for subcortical dementia can make minor adjustments to compensate for their deficiencies and lead relatively normal and productive lives.

The concept that multiple sclerosis can give rise to a form of subcortical dementia can be seen as a template to plan the initial evaluation of a patient; in this way the concept has some utility for clinicians. Any other application of the concept is likely to lead to misleading and potentially harmful inferences.

Progression of Cognitive and Emotional Deficits In numerous cross-sectional studies the correlations between disease duration and cognitive performance, when adjusted for chronological age, are uniformly low in size and often statistically nonsignificant. Longitudinal data are sparse, but the available information indicates that there are no meaningful declines in average performance of multiple sclerosis patients on neuropsychological tests over study periods of 1 to 3 years. The exception to this generalization is for tests that have a large sensory or motor component on which performance tends to decline with time.

The implications of these findings for cognitive rehabilitation of multiple sclerosis patients are clearly positive, because gains from therapy would not be cancelled by

progressive cognitive deterioration as occurs in dementia of the Alzheimer's type. Undoubtedly, some multiple sclerosis patients do experience significant cognitive decline, but at present there is no way to identify these patients.

Even less is known about the possible progression of emotional disturbances such as depression in multiple sclerosis. Again, cross-sectional studies do not reveal simple relationships between disease duration (or disability status) and measures of depression or aspects of psychological distress.

Neuroimaging Correlates Structural MRI, which revolutionized the clinical diagnosis of multiple sclerosis, continues to be the main research tool to examine correlates of cognitive impairment and to monitor the CNS effects of new pharmacological treatments.

Various measures have been employed to evaluate overall brain disturbance. The earlier studies used neuroradiological ratings of atrophy, lesion size, and extent of ventricular enlargement. More recent work measures total lesion area or volume or the area of certain structures that are often affected in multiple sclerosis, such as the corpus callosum. Regardless of the anatomical measure used, patients with greater brain disturbance perform more poorly on a variety of cognitive measures.

More specific relationships have also been reported. Studies of patients who received surgical section of the corpus callosum to control medically intractable seizures perform poorly on a variety of tasks that require transfer of information between the two cerebral hemispheres. In multiple sclerosis patients, the degree of impairment on tasks that require interhemispheric transfer is positively correlated with the extent of atrophy of the corpus callosum, but not with other anatomical measures such as total lesion area.

The Wisconsin Card Sorting Test is a problem-solving task that is known to be sensitive to focal lesions of the frontal lobes and to diffuse cerebral injury; as a group, multiple sclerosis patients display impairment on this test. However, there is great variability within the multiple sclerosis group and many patients perform normally. Recently, it has been shown that multiple sclerosis patients who performed poorly on the Wisconsin Card Sorting Test had significantly larger lesions in the white matter of the frontal lobes than patients who performed normally. There were no differences in total lesion load between the frontal and posterior groups.

These results encourage the view that it will be possible to establish more specific brain-behavior relationships and possibly to predict patterns of cognitive impairment. Functional neuroimaging would be an obvious way to approach this problem, but to date few studies with multiple sclerosis patients have been reported. None has employed an activation paradigm of the sort that has proved so powerful in studying neuroimaging correlates of cognitive processes in normal subjects.

Fewer studies have examined neuroimaging correlates of psychiatric disturbance in multiple sclerosis, but one study described a correlation between depression and the extent of lesions in the temporal lobes.

Treatment Because multiple sclerosis patients may exhibit a bewildering variety of physical symptoms, they often receive numerous medications to alleviate these conditions. If patients also require drugs to relieve psychiatric disorders, the possibility of drug interactions should be carefully considered.

Physical Symptoms Spasticity is a common problem in multiple sclerosis. Baclofen is most commonly prescribed, but various benzodiazepines, especially clonazepam (Klonopin) are also administered. Various other drugs such as cyclobenzaprine (Flexeril) are used less often.

Urinary incontinence is also common in multiple sclerosis. Peripherally active anticholinergics such as oxybutynin (Ditropan) are typically prescribed. For chronic pain, opiates are usually given with varying success.

Fatigue Amantadine and pemoline (Cylert) are given for the relief of fatigue. Amantadine was shown to be effective for some patients in a recent clinical trial, but pemoline was no more effective than placebo. Some physicians regularly prescribe corticosteroids to alleviate fatigue and elevate mood. Corticosteroids have an established role in reducing inflammation and shortening the duration of relapses in multiple sclerosis, but long-term administration of these drugs must be questioned on several grounds. Corticosteroids have unpredictable and occasionally dramatic effects on mood and behavior and increase susceptibility to hip fracture in multiple sclerosis patients.

Psychiatric Symptoms Despite the high prevalence of psychiatric disturbance in multiple sclerosis patients, few controlled studies have focused primarily on treatment-related issues. Although some authors have cautioned that multiple sclerosis patients might not respond to traditional pharmacotherapy or psychotherapy because of the demyelination process of the disease, no empirical studies have supported these claims. The few available studies have found that MS patients respond to traditional psychiatric therapy as well as other patient populations do.

Depression in patients who have multiple sclerosis has been found to respond to cognitive therapy, cognitive-behavior therapy, and group psychotherapy. Studies have shown that psychotherapy benefits patients and their families by helping them cope with the unpredictable and variable nature of the disease. For example, educating patients about their disease and providing them with strategies to facilitate continued employment and social involvement tends to reduces stress and helps patients adjust to living with a chronic illness.

Pharmacological treatments have also demonstrated efficacy in managing depression in multiple sclerosis. Specifically, tricyclic medications such as amitriptyline (Elavil), nortriptyline (Pamelor), and desipramine (Norpramin) have produced good clinical results, although the anticholinergic adverse effects of these drugs may limit their utility. Some authors have suggested that imipramine might be most appropriate for depressed multiple sclerosis patients who also experience bladder-control difficulties. The selective serotonin reuptake inhibitors (specifically sertraline [Zoloft] and fluoxetine) have also been shown to effectively treat depression in multiple sclerosis.

Pathological laughter and crying can usually be managed successfully with amitriptyline. Patients who fail to respond to amitriptyline might improve following a trial of levodopa (Larodopa). Multiple sclerosis patients with bipolar I disorder have been found to respond well to lithium treatment.

Drugs That May Affect the Course of the Disease Two synthetic analogues of natural human interferon beta have recently become available for the treatment of relapsing-remitting multiple sclerosis. Interferon beta-1b (Betaseron) reduced the frequency of relapses, and blocked the increase in lesion load as measured by serial MRI; over a 5-year trial, patients receiving this preparation tended to show smaller increases in disability, but the trend was not statistically significant.

Interferon beta-1a (Avonex) reduced the proportion of multiple sclerosis patients whose disability status progressed to a predetermined level, suggesting that interferon beta-1a slowed the course of the disease.

A slightly different form of interferon beta-1a (Rebif) is approved in Europe and Canada and is under review in the United States. This form reduced relapse rate, slowed progression of disability, and reduced total lesion load on MRI.

The most common adverse effects of all three drugs are local irritation at the injection site and flu-like symptoms, which usually disappear in a few months. Some patients experience increases in depression or chronic fatigue while taking interferon beta-1b. These adverse effects have not been reported for the interferon-1a forms, but experience with them is more limited than with interferon beta 1a.

Interferon beta has a number of immunomodulating actions that could contribute to its therapeutic action. Among the more interesting possibilities are the inhibition of interferon gamma from activated lymphocytes and the inhibition of production of tumor necrosis factor by macrophages. Both interferon gamma and tumor necrosis factor are hypothesized to play roles in demyelination. Once considered a potential therapy, interferon gamma is now known to worsen disease symptoms.

The fourth new drug, glatiramer (Copaxone), was formerly called copolymer 1. The drug is the acetate salt of synthetic polypeptides composed of four amino acids: L-alanine, L-glutamic acid, L-lysine, and L-tyrosine. Copaxone moderately reduced relapse rate and disability progression.

The effects, if any, of these four drugs on cognitive functions in multiple sclerosis are not known. Similarly, there is no evidence at present that any of the many other drugs given to multiple sclerosis patients worsens (or improves) cognitive functions at the doses utilized. However, strongly anticholinergic antidepressant medications such as amitriptyline and many of the benzodiazapines can impair anterograde memory at therapeutic doses. There are numerous other effective medications without these adverse effects so the problem of iatrogenic memory impairment can usually be avoided.

OTHER DEMYELINATING DISORDERS

In addition to common conditions such as traumatic brain injury that directly damage the white matter there are a number of rare genetic diseases that affect the integrity of myelin. Most of these disorders exhibit an autosomal recessive or X-linked pattern of inheritance and involve genes that control the production of enzymes that influence the synthesis or assembly of the lipids and proteins that comprise the myelin sheath. These reactions, including the production of the catalytic enzymes, occur in multiple steps. Depending on the stage of the biochemical chain that is affected by the gene defect, the enzyme may be completely absent, present in reduced concentration, or present in normal concentration with reduced activity. The nature of the change in enzyme function determines the age of onset of the disease and, to some extent, the pattern of neurological, cognitive, and psychiatric symptoms.

The *gangliosidoses* result from alterations in the enzyme hexosaminase A. Complete absence leads to *Tay-Sachs disease*, which begins in early infancy as a severe neurodegenerative disorder that progresses to death in several years. Less severe enzyme deficits are associated with onset in adolescence as in *Sandhoff's disease*, which often presents as bipolar I disorder with later development of neurological symptoms.

Neuronal ceroid lipofuscinosis is associated with motor dysfunction including facial dyskinesias as well as progressive dementia with personality changes. The name applied to this disorder depends upon the age of onset. In early infancy, it is called *Haltia-Santavuori disease*, but with adult onset it is called *Kufs disease.*

Another group of metabolic disorders includes *metachromatic leukodystrophy, adrenoleukodystrophy, globoid cell leukodystrophy* , and *Hurler syndrome.* These disorders may first appear in childhood, adolescence, or early adulthood. Psychiatric and cognitive symptoms are often prominent early and the progression of these symptoms is highly variable.

Diagnosis of these conditions is improving rapidly. Already there are clinical tests that are specific for some of the implicated enzymes; rapid advances in the molecular biology of the human nervous system will undoubtedly yield additional tests of high sensitivity and specificity.

Until recently, treatment for these genetic storage diseases has been mostly supportive and palliative. A recent review of more than 200 patients who received bone marrow transplantation for metachromatic leukodystrophy, globoid cell leukodystrophy, adrenoleukodystrophy, or Hurler syndrome indicated significant improvement in survival, performance on neuropsychological tests, MRI and spectroscopic measures of brain structure and metabolism, and quality of life. In contrast, the nutritional treatment for adrenoleukodystrophy portrayed in the recent motion picture, *Lorenzo's Oil*, has not proved generally effective.

SUGGESTED CROSS-REFERENCES

Neuroimaging is described in Section 2.13, neuroanatomy in Section 1.2 and Section 1.3, schizophrenia in Chapter 12, mood disorders in Chapter 14, acquired immune deficiency syndrome in Section 2.8, exposure to alcohol in Section 11.2, and exposure to volatile inhalants in Section 11.8.

SECTION REFERENCES

Abdul-Ahad AK, Galazka AR, Revel M, Biffoni M, Borden EC: Incidence of antibodies to interferon-beta in patients treated with recombinant human interferon-beta-1a from mammalian cells. Cytokines Cell Mol Ther *3:*27, 1997.

Arnett PA, Rao SM, Bernardin L, Grafman J, Yetkin FZ, Lobeck L: Relationship between frontal lobe lesions and Wisconsin Card Sort Test performance in patients with multiple sclerosis. Neurology *44:*420, 1994.

Beatty WW, Goodkin DE, Hertsgaard D, Monson N: Clinical and demographic predictors of cognitive performance in multiple sclerosis: Do diagnostic type, disease duration and disability matter? Arch Neurol *47:*305, 1990.

Beatty WW, Monson N: Metamemory in multiple sclerosis. J Clin Exp Neuropsychol *13:*309, 1991.

*Beatty WW, Paul RH, Wilbanks SL, Hames KA, Blanco CR, Goodkin DE: Identifying multiple sclerosis patients with mild or global cognitive impairments using the Screening Examination for Cognitive Impairment (SEFCI). Neurology *45:*718, 1995.

Beatty WW, Blanco CR, Wilbanks SL, Paul RH, Hames KA: Demographic, clinical and cognitive characteristics of multiple sclerosis patients who continue to work. J Neurol Rehab *9:*167, 1995.

Beatty WW, Wilbanks SL, Blanco CR, Hames KA, Tivis RD, Paul RH: Memory disturbance in multiple sclerosis: Reconsideration of patterns of performance on the Selective Reminding Test. J Clin Exp Neuropsychol *18:*56, 1996.

*Brosnan CF, Raine CS: Mechanisms of immune injury in multiple sclerosis. Brain Pathol *6:*243, 1996.

Charcot J-M: *Lectures on the Diseases of the Nervous System.* New Syndenham Society, London, 1877.

Cuzner ML, Norton WT: Biochemistry of demyelination. Brain Pathol *6:*231, 1996.

D'Esposito M, Onishi K, Thompson H, Robinson K, Armstrong C, Grossman M: Working memory in multiple sclerosis: Evidence from a dual-task paradigm. Neuropsychology *10:*51, 1996.

*Diaz-Olavarrieta C, Cummings JL, Velazquez J, Garcia de al Cadena C: Neuropsychiatric manifestations of multiple sclerosis. J Neuropsychiatry Clin Neurosci *11:*51, 1999.

Filippi M, Horsfield MA, Tofts PS, Barkhof F, Thompson AJ, Miller DH: Quantitative assessment of MRI lesion load in monitoring the evolution of multiple sclerosis. Brain *118:*1601, 1995.

Grant I, Brown GW, Harris T, McDonald WI, Patterson T, Trimble MR: Severely threatening events and marked life difficulties preceding onset or exacerbation of multiple sclerosis. J Neurol Neurosurg Psychiatry *52:*8, 1989.

Hachinski V: Interferons in the treatment of multiple sclerosis. Arch Neurol *55:*1583, 1998.

Haegert DG, Maruso MG: Genetic susceptibility to multiple sclerosis. Ann Neurol *36:*5204, 1994.

Herndon RM, Jacobs L: Interferons should be used to treat most patients with MS. Arch Neurol *55:*1581, 1998.

*IFBN Multiple Sclerosis Study Group and the University of British Columbia MS/MRI Analysis Group: Interferon beta-1b in the treatment of multiple sclerosis: Final outcome of the randomized controlled trial. Neurology *45:*1277, 1995.

Jacobs L, Cookfair D, Rudick R, Herndon R, Richert J, Salazar A, Fischer J, Granger C, Simon J, Goodkin D, and the Collaborative Multiple Sclerosis Research Group: Results of a phase III multicenter trial of intramuscular recombinant beta interferon as treatment for multiple sclerosis. Ann Neurol *36:*259, 1994.

Johnson SK, Lange G, Deluca J, Korn LR, Natelson B: The effects of fatigue on neuropsychological performance in patients with chronic fatigue syndrome, multiple sclerosis and depression. Appl Neuropsychol *4:*145, 1997.

Kujala P, Portin R, Revonsuo A, Ruutiainen J: Automatic and controlled information processing in multiple sclerosis. Brain *117:*1115, 1994.

Krivit W, Lockman LA, Watkins PA, Hirsch J, Shapiro EG: The future for treatment by bone marrow transplantation for adrenoleukodystrophy, metachromatic leukodystrophy, globoid cell leukodystrophy and Hurler syndrome. J Inherit Metab Dis *18:*398, 1995.

Krupp LB, Alvarez LA, LaRocca NG, Scheinberg LC: Fatigue in multiple sclerosis. Arch Neurol *45:*435, 1988.

Kujala P, Portin R, Ruutiainen J: The progress of cognitive decline in multiple sclerosis: A controlled 3-year follow-up. Brain *120:*289, 1997.

Lublin FD, Whitaker JN, Eidelman BH, Miller AE, Arnason BGW, Burks JS: Management of patients receiving interferon beta-1b for multiple sclerosis: Report of a consensus conference. Neurology *46:*12, 1996.

Miller DJ, Asakura K, Rodriguez M: Central nervous system remyelination: Clinical application of basic neuroscience principles. Brain Pathol *6:*331, 1996.

Minden SL: Psychotherapy for people with multiple sclerosis. J Neuropsychiatry *4:*198, 1992.

*Minden SL, Schiffer RB: Affective disorders in multiple sclerosis: Review and recommendations for clinical research. Arch Neurol *47:*98, 1990.

Nyenhuis DL, Rao SM, Zajecka JM, Luchetta T, Bernardin L, Garron DC: Mood disturbance versus other symptoms of depression in multiple sclerosis. J Int Neuropsychol Soc *1:*291, 1995.

Paul RH, Beatty WW, Schneider R, Blanco CR, Hames KA: Cognitive and physical fatigue in multiple sclerosis: Relationships between self-report and objective performance. Appl Neuropsychol, *5* in press.

Paul RH, Blanco CR, Hames KA, Beatty WW: Autobiographical memory in multiple sclerosis. J Int Soc *3:*246, 1997.

Pelosi L, Geesken JM, Holly M, Hayward M, Blumhardt LD: Working memory impairment in early multiple sclerosis: Evidence from an event-related potential study of patients with clinically isolated myelopathy. Brain *120:*20, 1997.

*Rao SM, editor: *Neurobehavioral Aspects of Multiple Sclerosis.* Oxford University Press, New York, 1990.

Rao SM, Leo GJ, Bernardin L, Unverzagt F: Cognitive dysfunction in multiple sclerosis: I. Frequency, patterns, and prediction. Neurology *41:*685, 1991.

Rao SM, Leo GJ, Ellington L, Nauertz T, Bernardin L, Unverzagt F: Cognitive dysfunction in multiple sclerosis: II. Impact on employment and social functioning. Neurology *41:*692, 1991.

Rice G, Ebers G: Interferons in the treatment of multiple sclerosis: Do they prevent the progression of the disease? Arch Neurol *55:*1578, 1998.

Rudick R, Goodkin DE, Ransohoft R: Pharmacotherapy of multiple sclerosis: Current status. Cleve Clin J Med *59:*267, 1992.

Sandroni P, Walker R, Starr A: “Fatigue” in patients with multiple sclerosis: Motor pathway conduction and event-related potentials. Arch Neurol *49:*517, 1992.

Schiffer RB, Herndon RM, Rudick RA: Treatment of pathologic laughing and weeping with amitriptyline. N Engl J Med *312:*1480, 1985.

Söderström M, Ya-Ping J, Hillert J, Link H: Optic neuritis: Prognosis for multiple sclerosis from MRF, CSF and HLA findings. Neurology *50:*708, 1998.

2.8 NEUROPSYCHIATRIC ASPECTS OF HIV INFECTION AND AIDS

IGOR GRANT, M.D., F.R.C.P.(C), AND J. HAMPTON ATKINSON, JR., M.D.

Acquired immune deficiency syndrome (AIDS) is a lethal neuromedical disorder associated with infection by viruses of the *Retroviridae* class known as *human immunodeficiency viruses* (HIV). Although the central feature of HIV infection involves gradual collapse of the body's ability to mount an appropriate cell-mediated immune response, with attendant medical complications, neuropsychiatric phenomena can also be prominent. These phenomena can be considered under three major headings: neurobiological, psychobiological, and psychosocial. This section provides an overview of the biomedicine of HIV infection, followed by a comprehensive discussion of the neuropsychiatric complications.

HIV INFECTION

Acquired immune deficiency results from infection with human immunodeficiency viruses, of which two are currently known— HIV-1 and HIV-2. Globally, the vast majority of AIDS cases result from HIV-1 infection, whereas AIDS secondary to HIV-2 appears to be confined primarily to a region in West Africa. There may be other subtypes of HIV, which are now classified as HIV-O.

The critical events in HIV-1 disease are associated with infection of a subset of lymphocytes known as the *T4 (helper)* or *cluster designation 4 positive (CD4+) lymphocytes.* CD4+ cells play a central role in cell-mediated immunity and coordinate other critical immune events. Steady depletion of CD4+ cell numbers throughout the course of HIV-1 disease leads to a catastrophic collapse of cell-mediated immunity, as well as to immune dysregulation. In turn, these events lead to death from overwhelming infections, neoplasms, or wasting syndrome.

HIV-1 HIV-1 belongs to the family of retroviruses. Other retroviruses include human pathogens, such as HIV-2, and human T-cell leukemia viruses (HTLV). Pathogenic retroviruses have also been identified in several animal species, including simian immunodeficiency virus (SIV) in macaques, chimpanzee virus, sooty mangabey virus, African green monkey virus, and mandrill virus. Among other animals, there are retroviral infections of sheep (visna virus), goats (caprine arthritis encephalitis virus), horses (equine infectious anemia virus), cats (feline immunodeficiency virus [FIV]), and cows (bovine visna).

Retroviruses have several features in common. First, they tend to be transmitted sexually and by vertical transmission (mother-fetus). Second, they tend to produce disease in the immune and central nervous systems. Third, the time from infection to the development of clinical disease can be protracted. And fourth, the immune response that is mounted by the host appears to be ineffective in preventing disease progression or eliminating the virus.

It is generally believed that HIV-1 arose from some common viral ancestor that itself may have been a retrovirus infecting animals. However, only HIV-2 has strong genetic similarity to an animal retrovirus, the sooty mangabey virus. It is interesting that both HIV-2 and the sooty mangabey are found in West Africa, where those monkeys are commonly eaten by people. The closest animal virus relative of HIV-1 at present may be the chimpanzee virus isolated from animals in Gabon.

There are many genetic strains of HIV-1. At present, these are classified into two major groups, *M* and *O.* Most disease is attributed to M subtypes, of which there are ten (A to J). The B subtype predominates in North America, western Europe, and Australia and New Zealand; the A and D types are prevalent in Africa; the C and E types are common in Southeast Asia. This marked genetic variability of HIV-1 poses challenges to developing universally effective vaccines.

Epidemiology At the end of 1997, more than 600,000 cases of AIDS were reported (including almost 400,000 deaths). Initially a disease primarily of men, women comprised 16 percent of AIDS cases by 1997. The proportion of African-Americans with AIDS is rising in the United States. For example, although approximately 12 percent of the U.S. population is African-American, approximately 36 percent of AIDS cases reported by the end of 1997 were among African-Americans. Minority women have been especially impacted. In the United States, African-American women accounted for 56 percent of AIDS cases reported in women by the end of 1997.

Worldwide, the World Health Organization in 1997 estimated that 29.4 million adults and 1.1 million children under the age of 15 were living with HIV-1 infection, and that 16,000 people became infected each day. Over 30 million people have developed HIV infection since the epidemic began; approximately 90 percent of these infections occurred in developing countries, with 65 percent in sub-Saharan Africa and 20 percent in Southeast Asia. It has been estimated that 11.7 million people have now died from HIV-associated illnesses worldwide. Based on current trends, estimates of the number of persons likely to be infected by the year 2000 have ranged from 40 million to over 100 million. The scope of the problem is therefore highly significant, especially in developing countries.

Mode of Transmission The modes of transmission include heterosexual and homosexual intercourse, vertical transmission from infected mother to fetus or newborn, and instrumental transmission, which involves introduction of HIV-contaminated fluids or materials into the body by means of needles, blood products, or various medical accidents. Receptive fellatio involving ejaculation of HIV-infected semen is another potential mode, but the actual risk is not known. Kissing is not considered a risk unless there is extensive oral disease with open sores. (Because this chapter concerns infection with HIV-1, not HIV-2, the term *HIV* is used to refer to HIV-1–related phenomena.) Worldwide the sexual mode of transmission is the most important, with homosexual sex accounting for the majority of cases in North America, Europe, and the Antipodes, heterosexual transmission being the most common in Africa and Asia, and both patterns occurring in South America.

In North America there appears to be a steady shift in risk patterns. Although the incidence of AIDS associated with homosexual transmission has stabilized or even declined somewhat, the incidence of cases related to heterosexual transmission and injection drug use continues to rise. In the period from 1988 to 1995, the proportion of AIDS cases attributable to injection drug use rose from approximately 15 to 23 percent, and that attributable to heterosexual contact grew from approximately 5 to 18 percent. About 50 percent of women with AIDS acquire HIV either directly or indirectly through injection drug use (i.e., either through being injection drug users themselves, or from having sex with a man who is an injection drug user).

These various data suggest that clinicians in the future will be treating increasing numbers of HIV-infected persons who have acquired their disease through injection drug use, heterosexual transmission, or some combination of these two, and that the rates among women, Hispanics, and blacks will continue to rise.

Process of HIV Infection The cellular and viral dynamics of HIV infection are not fully understood. In order to become infective, once HIV gains entry to the body either through disruption in the mucous membranes or directly by injection, it must enter cells that can support its replication. Entry into cells generally requires that the host cell express the CD4 receptor as well as a co-receptor, in the form of a chemokine receptor. *Chemokines* are proteins secreted by many types of cells in response to inflammatory triggers. They induce leukocyte migration and may be involved in other aspects of host defense against infection, and in the pathophysiology of inflammatory diseases. Over 40 chemokines and more than a dozen chemokine receptors have been identified. Two chemokine receptors—CCR5 and CXCR4—are known to be important for HIV entry into cells. CCR5 is expressed by monocyte-macrophage cells, CXC4 by T cells, and both are found on dendritic cells of lymph nodes. Fusion of HIV with the host cell membrane requires the presence of these chemokine receptors. In their absence, viral attachment can occur,

but fusion and viral entry are blocked. Interestingly, a small proportion of people have a genetic variation that alters the molecular structure of CCR5 so that it is not expressed on the cell surface; such individuals appear to be resistant to infection with HIV.

Once HIV fuses with a host cell membrane, it becomes uncoated and viral ribonucleic acid (RNA) and proteins are released into the cell. HIV reverse transcriptase then creates a deoxyribonucleic acid (DNA) copy of viral RNA. Viral DNA enters the nucleus of the host cell, where it is spliced into the host DNA and is termed a *provirus.* The proviral DNA then directs the formation of messenger RNA (mRNA) in a process called *transcription.* The mRNA enters the cell cytoplasm where it harnesses the protein-making machinery to generate new viral proteins and enzymes in a process termed *translation.* These proteins then accumulate inside the cell membrane and ultimately bud off as a new viral particle. However, the virus is not yet infectious, because the large viral polyproteins have not yet been cleaved into functional units. This cleavage involves the viral enzyme protease. Once the polyproteins are properly clipped, the new virion becomes a fully infectious HIV particle. Understanding the mechanisms of viral attachment, entry, and replication has been crucial to the development of new antiretroviral drugs.

Once infection is established, an extremely active process ensues wherein infected cells constantly release new virions into tissue compartments and the circulation, followed by the body's response to clear these viruses. It is now thought that several billion new virus particles may be generated daily with millions of CD4 cells destroyed every day, being replaced by comparable numbers of new cells generated by the host. The half-life of virus in the circulation is thought to be 6 hours, while that of productively infected CD4 lymphocytes is probably 1.6 days. These CD4 cells are the main reservoirs of virus, but other cells and tissue compartments participate as well. For example, some HIV is harbored by longer-lived cells of the macrophage line; it is also known that viruses are sequestered in the dendritic follicular cells of lymph nodes and microglia in the brain.

The early period of infection may be characterized by high plasma viral load and increasing efforts by the host's immune system to clear the virus. In the typical case, an equilibrium is established wherein viral load drops, but remains detectable. The period between initial exposure and this equilibrium state (approximately 6 months) is termed the *phase of primary infection.*

Clinical Picture HIV can be detected in blood, cerebrospinal fluid (CSF), lymph nodes, and other tissues shortly after infection. Antibodies to various HIV constituents typically appear after 3 to 12 weeks via a process called *seroconversion.*

Seroconversion occurring 3 months or more after exposure has been reported, but appears to be quite rare. Antibodies to HIV can be detected with the enzyme-linked immunoabsorbent assay (ELISA), an inexpensive procedure that is universally available, highly sensitive, and reasonably specific. However, because false-positive test results occasionally occur, confirmation must always be made through Western blot analysis before a person is notified that he or she is seropositive.

During the period of seroconversion, perhaps one half to two thirds of infected persons experience a seroconversion illness, which may resemble the flu. The symptoms include low-grade fever, myalgia, headache, fatigue, gastrointestinal upset, and a skin rash. Rarely, neurological complications can occur at that point, which may include aseptic meningitis, encephalopathy, or acute demyelinating neuropathy (Guillain-Barré syndrome).

Following the period of primary infection, there typically follows an asymptomatic period that may continue for many years, until medical complications develop that permit the diagnosis of AIDS. During the protracted asymptomatic stage (which in North America has a median duration of 10 years), there is typically a steady depletion of the CD4+ lymphocyte number from a preinfection level of approximately 1000 per mm_3 or more to a level of 200 per mm_3 or less. The likelihood of AIDS-defining complications increases markedly when CD4+ cells drop to below 200 per mm_3. Why some people develop AIDS within a few years whereas some others remain nonprogressors for 15 or more years is not known. Some reports have noted that nonprogressors have higher numbers of cytotoxic T lymphocytes and increased neutralizing antibodies, suggesting a better immune response; other data indicate that nonprogressors may be infected with less pathogenic strains of HIV.

Prior to the development of opportunistic infections or other complications, many HIV-infected persons develop *persistent generalized lymphadenopathy.* The prognostic significance is unclear, and it should not be assumed that persons with this condition are necessarily more advanced in HIV illness than other asymptomatic persons.

In North America and Europe the most common complication leading to the diagnosis of frank AIDS (other than drop in CD4 cell count below 200) is pneumonia caused by the protozoan *Pneumocystis carinii.* Patients present with chronic nonproductive cough and dyspnea, and they may develop hypoxemia. Diagnosis usually requires fiberoptic bronchoscopy with alveolar lavage, from which specimens can be suitably stained for the *Pneumocystis* organism. Treatment is with trimethoprim-sulfamethoxazole (Bactrim, Septra) or pentamidine (Pentam, Nebupent). In the United States it is now commonplace to initiate prophylactic treatment for *Pneumocystis* in those with CD4 cell count below 200.

Other common infections that herald the onset of AIDS include tuberculosis, candidiasis, and repeated herpes zoster. Additional complications include Kaposi's sarcoma (linked to human herpesvirus-8 [HHV-8] infection) and oral hairy leukoplakia.

Later AIDS-related events, signifying more profound immunodeficiency, include cytomegalovirus (CMV) infection (including CMV retinitis and gastroenteritis), various atypical infections (e.g., *Mycobacterium avium intracellulare*), as well as lymphomas and a severe wasting syndrome (which may be caused by various infections or by HIV itself).

DIAGNOSIS

CDC Classification System The Centers for Disease Control has developed a classification system for HIV infection comprised of two dimensions: history of clinical conditions and degree of immunosuppression. HIV-related clinical conditions are subdivided into three categories. *Category A* represents patients who have remained medically asymptomatic or have had only a transient illness at seroconversion or persisting lymphadenopathy. *Category B* identifies those patients with more serious HIV-related conditions (minor opportunistic infections). Lastly, *Category C* is indicative of patients who have had more serious AIDS-defining illnesses such as *Pneumocystis carinii* pneumonia. These classifications are currently all-absorbing, such that patients who develop the relevant symptoms but then recover do not revert to a less advanced classification.

The second dimension is degree of immunocompromise. This is based on the individual's CD4+ T-lymphocyte cell count per microliter of blood. Patients with CD4 counts greater than 500 are the healthiest, and fall into *Category 1.* Those with CD4 cell counts from 200 to 499 are in *Category 2*, and patients with CD4 counts below 200 are in *Category 3.* AIDS is diagnosed in those with CD4 <200, a Category C complication, or both (Table 2.8-1).

CD4+ Cell Count Categories	Clinical Categories: A Asymptomatic or Lymphadenopathy	B Symptomatic, but not A or C Conditions	C AIDS-Indicator Conditions
1. >500/mm^3	A-1	B-1	**C-1**
2. 200–499/mm^3	A-2	B-2	**C-2**
3. <200/mm^3	**A-3**	**B-3**	**C-3**

Bold entry means CDC 1993 AIDS indicator conditions (A-3, B-3, C-1, C-2, C-3).

Table 2.8-1 CDC 1993 Classification System for HIV Infection

Neurobiological Aspects of HIV Infection The neurobiological complications of HIV infection can be either *primary* or *secondary.* The primary neurobiological complications are those that can be attributed directly to HIV infection of the central nervous system or to immunopathological events (e.g., autoimmune phenomena) precipitated by HIV infection. Secondary neurobiological complications include infections and neoplasms facilitated by the immunodeficiency state, cerebrovascular

complications, and toxic states produced either by HIV-associated medical illnesses (e.g., hypoxemia due to *Pneumocystis* pneumonia) or by toxic effects of various therapeutic agents (e.g., myopathy due to zidovudine [Retrovir]). Some of the more common complications are listed in Table 2.8-2.

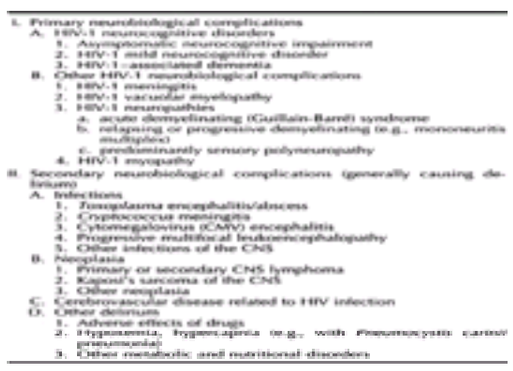

I. Primary neurobiological complications
  A. HIV-1 neurocognitive disorders
    1. Asymptomatic neurocognitive impairment
    2. HIV-1 mild neurocognitive disorder
    3. HIV-1–associated dementia
  B. Other HIV-1 neurobiological complications
    1. HIV-1 meningitis
    2. HIV-1 vacuolar myelopathy
    3. HIV-1 neuropathies
      a. acute demyelinating (Guillain-Barré) syndrome
      b. relapsing or progressive demyelinating (e.g., mononeuritis multiplex)
      c. predominantly sensory polyneuropathy
    4. HIV-1 myopathy
II. Secondary neurobiological complications (generally causing delirium)
  A. Infections
    1. Toxoplasma encephalitis/abscess
    2. Cryptococcus meningitis
    3. Cytomegalovirus (CMV) encephalitis
    4. Progressive multifocal leukoencephalopathy
    5. Other infections of the CNS
  B. Neoplasia
    1. Primary or secondary CNS lymphoma
    2. Kaposi's sarcoma of the CNS
    3. Other neoplasia
  C. Cerebrovascular disease related to HIV infection
  D. Other delirium
    1. Adverse effects of drugs
    2. Hypoxemia, hypercapnia (e.g., with *Pneumocystis carinii* pneumonia)
    3. Other metabolic and nutritional disorders

Table 2.8-2 Neurobiological Complications of HIV-1 Infection

HIV-Associated Neurocognitive Disorders From a psychiatric perspective the most important of the primary neurobiological complications are the HIV-associated neurocognitive disorders. The cardinal feature of these disorders is impairment in cognitive functioning; associated features include motor slowing or incoordination and, sometimes, mood disturbances.

Classification and Diagnostic Criteria There has been considerable confusion about the occurrence and features of the neurocognitive disorders, primarily because precise diagnostic criteria were not established until recently. For example, one of the earliest terms was *HIV encephalopathy*, a designation derived from early pathological studies that noted the presence of encephalitic features in the brains of many patients dying with AIDS. Another term was *AIDS dementia complex*, which suggested a constellation of cognitive, motor, and affective-behavioral complications. Many clinicians and investigators tended to use this designation somewhat indiscriminately—not bearing in mind, for instance, that the term *dementia* connotes severe disability. As a consequence of this inexact terminology clinicians failed to recognize milder cognitive disorders or were reluctant to diagnose patients for fear of stigmatizing them with the label of AIDS dementia complex. Richard Price and Bruce Brew provided a framework for staging AIDS dementia complex into five levels of severity, including a subclinical stage; however, it is preferable that the neurocognitive complications be classed as two syndromes differing in level of severity: *HIV-associated dementia* and the *HIV-associated mild neurocognitive disorder.*

The diagnostic criteria for the two syndromes, as used in this section, are presented in Table 2.8-3. The principal differentiating point between dementia and mild neurocognitive disorder is the extent to which the cognitive disorder interferes with a person's day-to-day functioning. Thus, a diagnosis of dementia is reserved for severe cognitive disorders that interfere substantially with work, home life, and social activities. On the other hand, mild neurocognitive disorder can be diagnosed if cognitive deficits that do not interfere in a major way with life functioning are reliably identified. Some people manifest mild neuropsychological deficits that are entirely subclinical (i.e., there is no documented interference in day-to-day functioning). Because such deficits do not reach the threshold for being termed a disorder, the term *asymptomatic neurocognitive impairment* may be used (Table 2.8-2).

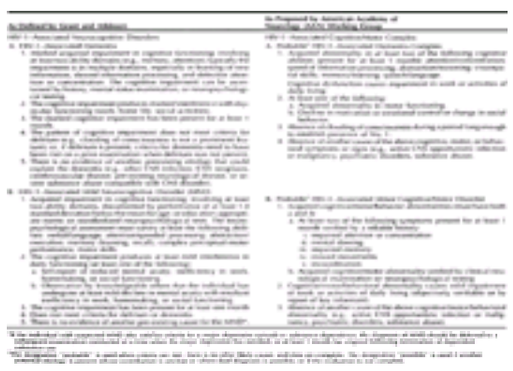

Table 2.8-3 HIV-1–Associated Cognitive Disorders

These definitions of HIV-associated dementia and HIV-associated mild neurocognitive disorder are conceptually similar to the nomenclature proposed by the American Academy of Neurology (AAN) AIDS Task Force. The definition of HIV-associated dementia differs from the AAN's HIV–1–associated dementia complex in not requiring a disturbance in either motor functioning or motivation, emotional control, or social behavior in addition to the cognitive deficit. Also, the emphasis is on marked cognitive disturbance and marked interference in daily functioning as key criteria for dementia. Mild neurocognitive disorder differs from the comparable AAN HIV–1–associated minor cognitive/motor disorder in specifying neuropsychological criteria for establishing the cognitive disorder. The purpose is to avoid inappropriately labeling persons as impaired without hard data. Like the criteria for dementia, the critical symptoms and signs are restricted to the cognitive sphere. Thus, the definitions used in this section reflect what are believed to be the central features of the disorders—cognitive impairment.

MILD NEUROCOGNITIVE DISORDER

Signs and Symptoms A person experiencing mild neurocognitive disorder associated with HIV infection will typically have some difficulty concentrating, may experience unusual fatigability when engaged in demanding mental tasks, may feel subjectively slowed down, and may notice difficulty in remembering. Such persons may say that they are not as sharp or as quick as they once were.

Such a set of presenting complaints, especially in younger individuals who may be struggling to accept their seropositive status, may lead the clinician to conclude that anxiety, depression, or hypochondriasis are responsible. Although affective features are occasionally the best explanation for such complaints, that is not generally the case. Rather, comprehensive neuropsychological testing may reveal that the individual does indeed have difficulties with speeded information processing, divided attention, and sustained effortful processing as well as deficiencies in learning and recalling new information.

Some individuals with mild neurocognitive disorder also have difficulties with tasks involving problem solving and abstract reasoning, and there may also be slowing of simple motor performance (e.g., speed of finger tapping). Verbal skills are less affected, although there may be some decrement in fluency (e.g., quickly reciting as many animals as possible or as many words beginning with a particular letter as possible).

These neuropsychological findings, which emphasize attentional problems, slowing of information processing, and deficiencies in learning, are reminiscent of neuropsychological patterns seen in patients with so-called subcortical dementias (e.g., Huntington's disease and Parkinson's disease). Fine-grained analysis of memory breakdown in HIV-infected persons also confirms a subcortical pattern. For example, persons with HIV-associated cognitive disorders have difficulty recalling words from a list but do not make intrusion errors (i.e., confabulatory responses) the way patients with cortical dementias (e.g., dementia of the Alzheimer's type) tend to do. Neuropsychological features that suggest selective involvement of subcortical structures are consistent with neuropathological findings. It is important to stress that these mild neurocognitive deficits occur independently of depression, anxiety, and other non-HIV sources of cognitive deficit.

Associated Features There are no laboratory findings specific to mild neurocognitive disorder. For example, the CD4+ lymphocyte number tends at most to be very weakly related to the severity of cognitive disturbance. However, the rate of CD4 decline over time may be related to rate of cognitive deterioration. Other immunological indicators (e.g., serum b_2-microglobulin) also relate only weakly to neuropsychological performance. Cerebrospinal fluid studies may reveal increased

mononuclear cell number, protein elevation, mild elevation in b 2-microglobulin, and some increase in quinolinic acid (QA), a putative excitotoxin. HIV RNA can often be detected in CSF using sensitive techniques (polymerase chain reaction [PCR]). There is some evidence that viral load in CSF, but not in plasma, is correlated with severity of neuropsychological impairment, especially in AIDS patients.

Neurological examination tends to be unremarkable; occasionally there may be frontal release signs (e.g., glabella response), but such findings also occur in those without mild neurocognitive disorder.

Structural brain imaging (e.g., computed tomography [CT] scanning, magnetic resonance imaging [MRI]) tends to be noncontributory. Dynamic brain imaging (e.g., single photon emission computed tomography [SPECT]) can reveal reduction in the uptake of tracer substances, but once again the relationship between reduced uptake and neuropsychometric findings is not strong.

The results of electrophysiological studies are often completely normal, although there may be evidence of slowing of central conduction time as indicated by slowed latencies of median nerve or tibial nerve somatosensory evoked potentials. Event-related potentials are typically normal. Disturbance in smooth-pursuit eye movement has occasionally been noted.

HIV-ASSOCIATED DEMENTIA

Signs and Symptoms As might be expected, the cognitive abnormalities in persons with dementia are more profound and more generalized than in patients with mild neurocognitive disorder. Patients report severe forgetfulness, difficulty concentrating, problems with naming and word finding, marked mental slowness, and disorientation. Bedside mental status examination may initially show only modest abnormality (e.g., Mini-Mental State Examination [MMSE] scores in the mid-20s). Nevertheless, comprehensive neuropsychological testing usually reveals marked impairment in speeded information processing, attention, abstracting ability, complex perceptual motor skills, learning of new information, and recall.

As the dementia progresses, the patient may become more apathetic, severely disoriented, and frankly confused and may have difficulty with independent living. Some patients become markedly irritable or labile in mood. Periods of delirium, often related to intercurrent medical events or treatment, may be superimposed on the progressive dementing picture. A small proportion of patients with dementia develop psychotic symptoms that typically have a paranoid flavor and are accompanied by auditory and visual hallucinations.

Associated Features Frank neurological findings are typical in HIV-associated dementia, including presence of frontal release signs, exaggerated deep tendon reflexes, disturbed smooth pursuit eye movement, incoordination, and motor weakness; the latter two findings tend to be worse in the lower than in the upper limbs. In advanced dementia, the patient may become mute and unresponsive and may develop urinary and fecal incontinence. The dementing disorder may be complicated by other neurological syndromes, such as sensory neuropathy and myelopathy.

Neuroradiological examination usually reveals cerebral atrophy (dilation of sulci and ventricles; Fig. 2.8-1 and Fig. 2.8-2 compare normal brain with brain that has atrophy). There may be lucencies in the white matter on CT scanning, and MRI may reveal abnormal areas of high signal, especially on T2-weighted images. Those parenchymal abnormalities may range from multifocal punctate abnormalities to larger confluent areas of abnormal signal. Figure 2.8-2,Figure 2.8-3, and Fig. 2.8-4 provide examples of such patterns.

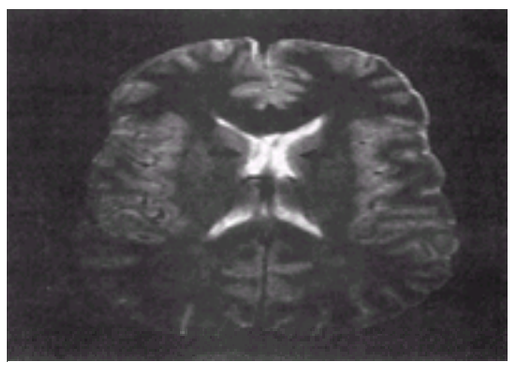

FIGURE 2.8-1 Normal MRI from a 35-year-old man with HIV infection. This T2-weighted axial section shows lateral ventricles of normal size (solid arrows) and no enlargement of sulci. (Courtesy of John R. Hesselink, M.D.)

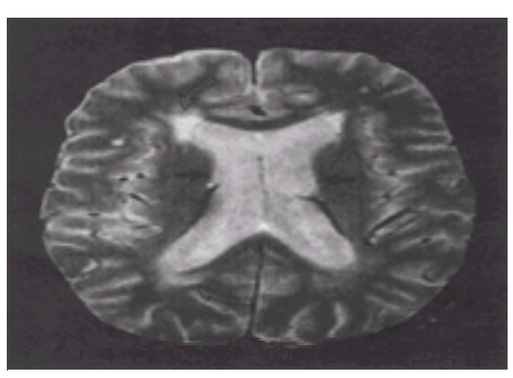

FIGURE 2.8-2 This axial T2-weighted image reveals moderate enlargement of ventricles (solid arrows) consistent with moderate central parenchymal volume loss in a man in his late 30s with AIDS. There are also several small (open arrows) and large patches of increased signal, the latter situated periventricularly around the frontal and occipital horns of the lateral ventricles (curved open arrows). The hyperintensities are consistent with HIV encephalitis, but could also represent another infectious etiology. (Courtesy of John R. Hesselink, M.D.)

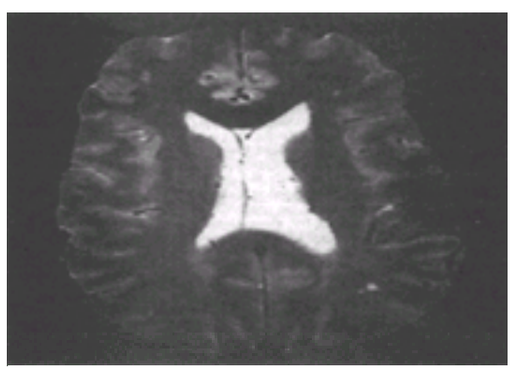

FIGURE 2.8-3 This T2-weighted axial image from a 40-year-old man with HIV infection demonstrates scattered punctate high-signal abnormalities (solid arrows). (Courtesy of John R. Hesselink, M.D.)

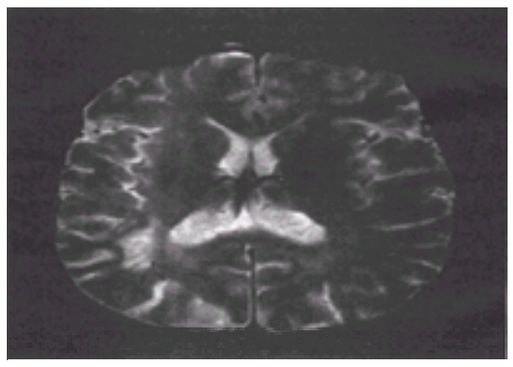

FIGURE 2.8-4 This axial T2-weighted image from a 48-year-old man with AIDS reveals a confluent area of high signal, consistent with HIV encephalitis in the right posterotemporal region (curved arrow). (Courtesy of John R. Hesselink, M.D.)

Dynamic brain imaging (e.g., positron emission tomography [PET]) may reveal increased uptake of 18-fluorodeoxyglucose (FDG) at onset of dementia, especially in the subcortical gray structures; as the dementia progresses, the metabolic abnormality tends to subside and may be replaced with hypometabolism. Studies with magnetic resonance spectroscopy have detected reduction in *N*-acetylaspartate, a putative neuronal marker, suggesting damage to or loss of neurons.

Electrophysiological abnormalities are common at this stage, and they may include slowing (and rarely, spike discharges) on the electroencephalogram (EEG), slowing of latencies on evoked potential studies, and abnormalities in event-related potentials. Ophthalmological examination will often reveal cotton wool spots in the retina, indicating small areas of infarction. Eye movements tend to be grossly ataxic.

Laboratory findings associated with dementia include very low CD4 numbers (not an invariable association), high serum and CSF b $_2$-microglobulin, increased CSF neopterin, and increased CSF quinolinic acid. Increasing CSF viral load is related to severity of dementia.

Neuropathology In the first few years after AIDS was first defined as a clinical entity, it was speculated that neuropsychiatric complications might result from CMV encephalitis. In the mid-1980s, however, it became clear that HIV itself frequently invaded the central nervous system (CNS), with specific neuropathological results.

On microscopic examination, the characteristic neuropathological changes in the brains of those dying from HIV-associated dementia include myelin pallor, microglial nodules, multinucleated giant cells, perivascular infiltrates, and foci of demyelination. Typical inflammatory changes are minimal or absent. Figure 2.8-5 and Figure 2.8-6 are photomicrographs illustrating some of those changes. Although the neuropathology can be widespread, it is typically concentrated in subcortical gray and deep white matter.

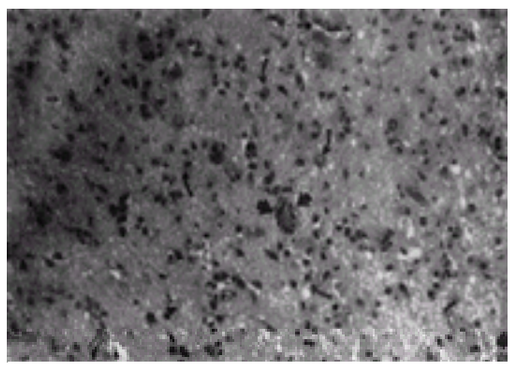

FIGURE 2.8-5 Neuropathology of HIV. This section illustrates two common findings: multinucleated giant cells (one such indicated by two straight arrows) and infiltration by mononuclear macrophages (curved arrows). (From an immunoperoxidase-stained paraffin-embedded section of brain tissue. Original magnification ×300.) (Courtesy of Clayton A. Wiley, M.D., Ph.D.)

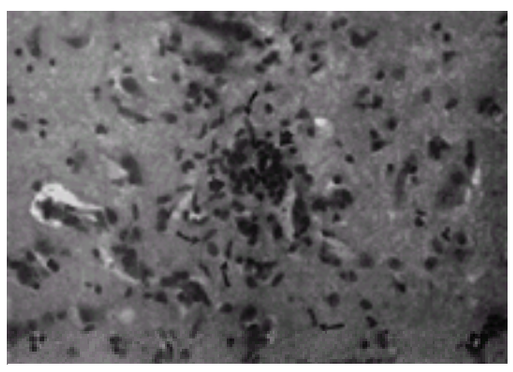

FIGURE 2.8-6 Neuropathology of HIV. This section illustrates a microglial nodule, frequently noted (along with multinucleated giant cells) in brains of persons dying with HIV-1-associated dementia. The nodule (indicated by two straight arrows) is filled with and surrounded by microglia, which are mononuclear cells with elongated nuclei (curved arrows). (From a hematoxylin- and eosin-stained paraffin-embedded section of brain tissue. Original magnification ×600.) (Courtesy of Clayton A. Wiley, M.D., Ph.D.)

Immunocytochemical or molecular biological studies on the brains of patients dying with presumed HIV-associated dementia usually demonstrate infection of the CNS by HIV-1. As detection techniques have improved, so has the likelihood of detecting HIV in the brain. For example, introduction of the PCR-technique to amplify viral products has facilitated more frequent and more widespread identification of HIV in the brain than was possible through immunocytochemical or *in situ* hybridization approaches.

Considerable effort has been devoted to cellular localization of HIV in the brain. It appears that cells of the macrophage-microglia line are principally infected. There have been no confirmed reports of productive HIV infection of neurons or astroglia; several unconfirmed reports suggest that oligodendroglia may occasionally be infected.

Although HIV is not found in neurons, neuronal loss can occur in the neocortex and deep gray structures. Furthermore, there is evidence that neurons can be damaged in HIV encephalitis. Figure 2.8-7 shows neurons that have lost their dendritic spines and others with abnormal vacuolation. Those observations raise questions about the mechanisms of CNS damage in HIV infection.

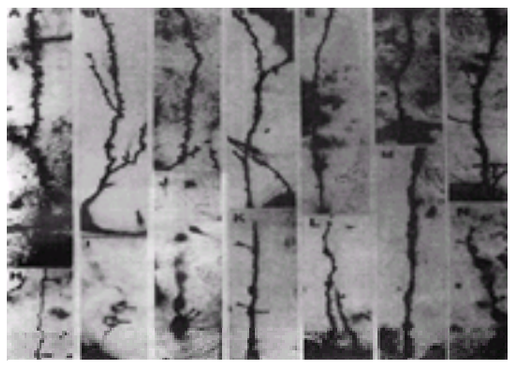

FIGURE 2.8-7 Comparative morphology of different Golgi impregnated dendritic segments in control **(A - C)** and HIV encephalopathy (HIVE) **(D - N)** frontal cortex. Medial **(A)**, proximal **(B)**, and distal **(C)** apical segments in control tissues showed abundant spines. Proximal (**D** and **E**), middle **(F)**, and distal **(G)** apical segments in HIVE cases showed decreased number of spines. **H**, typical varicosities (*arrowheads*) observed in neuritic processes of HIVE cases. **I**, vacuolated (*arrows*) apical dendrite in HIVE. **J**, vacuolated dendritic segment with occasional spines (*arrowheads*). Middle **(K)** and distal **(L)** dendritic segments were swollen throughout their length; occasional spines (arrowheads) and abnormal spines (arrow) were present. **M**, extensively distorted apical dendrite. **N**, dendrite with segmented dilatations (*stars*) and a few abnormal spines (*arrowheads*) were observed. (Magnification ×495.) (Reprinted with permission from Masliah E, Ge N, Morey M, DeTeresa RM, Terry RD, Wiley CA: Cortical dendritic pathology in HIV encephalitis. Lab Invest *66:*288, 1992.)

Pathogenesis Much recent research on neuro-AIDS has focused on possible mechanisms underlying neuronal injury in HIV infection. The macrophage microglia cells play a central role in current concepts of neuropathogenesis. It has been proposed that there are two stages of HIV infection of brain macrophages. The first stage involves binding of HIV to the cell, leading to virus internalization. During this stage, the macrophage may release toxins or stimulate other cells to do so. In the second stage HIV is integrated into the macrophage genome, leading to active viral reproduction accompanied by heightened macrophage activation. During this stage the macrophage responds in an excessive manner to secondary activators, which can include the brain's own cellular regulatory system, other infectious organisms, cytokines produced by other immune cells, and viral envelope glycoprotein gp120. Activated in this manner, macrophages are thought to produce a variety of neurotoxins including glutamate-like molecules, cytokines, free radicals, and arachidonic acid. Whereas several theories of neuropathogenesis vary in certain details, the general consensus is that neuronal injury in AIDS involves overactivation of channels permeable to calcium, including voltage-dependent channels and those operated by *N*-methyl-D-aspartate (NMDA) receptors. Indeed, the final common pathway of neuronal injury in AIDS has been found to resemble the mechanism of damage in stroke, trauma, epilepsy, neuropathic pain, and several neurodegenerative diseases. The theory is that neurotoxins activate NMDA receptors, thereby leading to influx of calcium, which in turn causes the neuron to release glutamate. The glutamate excites other neurons, causing further calcium influx and leading ultimately to cellular injury and death.

Course and Prognosis The prevalence of mild neurocognitive disorder increases with disease progression. For example, approximately 50 percent of patients diagnosed with AIDS have either asymptomatic neuropsychological impairment or mild neurocognitive disorder. Among patients who have no medical symptoms or who have early medical symptoms of insufficient severity to merit the diagnosis of AIDS, the rate of mild neurocognitive disorder is highly variable. Some investigators have found no increases in this disorder among asymptomatic seropositive persons compared to suitably matched controls; other studies have reported that subtle cognitive changes do occur in the early and middle phases of disease. Twenty to 30 percent of such asymptomatic or early symptomatic persons will have either mild neurocognitive impairment or mild neurocognitive disorder.

It is not known whether mild neurocognitive disorder predisposes patients to the development of frank dementia. In one series, a 2-year follow-up indicated that neuropsychological assessment at entry was not predictive of later dementia, but that neuropsychological worsening in a 6-month period was.

Given what is known thus far it is probably prudent for the clinician to consider that mild neurocognitive disorder can complicate the early natural history of HIV infection in some individuals. Therefore, those presenting with cognitive complaints or experiencing behavioral abnormalities should undergo comprehensive neuropsychological testing, neurological examination, and, as indicated, brain imaging studies.

Dementia occurs almost exclusively in more advanced AIDS. Although there have been some reports of dementia as the first presenting feature of AIDS, it is an uncommon initial finding. The exact prevalence of dementia among AIDS patients is not clearly known. Estimates in the earlier literature ranged from 15 to 66 percent of more advanced cases; the current data suggest a prevalence of 4 to 7 percent of those with AIDS. There is a general impression that the incidence of frank dementia has declined substantially in the last several years with the general availability of antiretroviral therapy.

The prognosis of asymptomatic neurocognitive impairment and mild neurocognitive disorder is not known. At present, they appear to be long-term stable conditions or only very slowly progressive. The advent of frank dementia is a poor prognostic sign. In one series about three quarters of the patients died within 6 months of developing frank dementia; presence of mild neurocognitive disorder is also associated with earlier time to death.

Other Primary Neurobiological Disorders Although the cognitive disorders are arguably the most important of the primary HIV-associated neurobiological conditions, there are several others of which the clinician needs to be aware.

Meningitis At about the time of seroconversion, a small proportion of individuals experience symptoms of meningeal irritation, usually in the context of a seroconversion illness (fever, achiness, weakness, other flulike symptoms). Symptoms may include photophobia, headache, and neck stiffness; very rarely, a more floridly encephalitic picture evolves with confusion and clouding of consciousness.

Periseroconversion meningoencephalitis is typically self-limited, although a few patients develop a chronic aseptic meningitis syndrome, with chronic headaches lasting many months. Such headaches may respond to low doses (e.g., 25 to 50 mg) of amitriptyline (Elavil) or nortriptyline (Aventyl, Pamelor).

Vacuolar Myelopathy Vacuolar myelopathy is probably the second most prevalent and serious primary HIV-related CNS complication. It is a disorder that occurs typically in later stages of the disease, with a prevalence of 20 to 30 percent in AIDS patients. Subclinical spinal cord disease probably exists in a much higher proportion of infected persons, in perhaps 50 percent of patients with AIDS and in a variable proportion of asymptomatic carriers.

The presenting features are spastic paraparesis, a predominantly posterior column sensory loss (i.e., greater loss of vibration and proprioception versus other sensory loss), which may lead to bowel and bladder incontinence. Sometimes the myelopathy coexists with dementia, although the two conditions are thought to differ in etiopathogenesis. For example, whereas HIV is typically detected in the CNS of HIV encephalitis cases, there is less evidence for productive HIV infection in relation to the demyelinating and gliotic lesions found in myelopathy. Furthermore, the myelopathy does not seem to respond favorably to zidovudine.

Neuropathies Acute demyelinating neuropathies can occasionally develop early in the course of illness, such as at the time of seroconversion. A generalized form of acute demyelinating neuropathy is the HIV-associated Guillain-Barré syndrome. Presenting features include a rapidly evolving symmetrical paralysis with variable sensory signs, often beginning in the lower limbs. Although the syndrome is an uncommon complication of HIV infection, it is evident that HIV serotesting is an important aspect of the differential diagnosis of Guillain-Barré syndrome, especially in younger males.

Although the acute demyelinating neuropathies that occur in the early stages of HIV disease tend to be self-limiting, other more persistent kinds of neuropathy can occur during the middle and later phases. These include a predominantly sensory polyneuropathy characterized by onset of numbness, tingling, burning, or pain in a primarily distal and relatively symmetrical distribution. Symptoms are especially common in the feet and include painful sensations and contact sensitivity without motor involvement.

The prevalence of these disorders is not well understood, although it has been estimated that 10 percent of asymptomatic carriers and about a third of patients with AIDS experience sensory neuropathy. Zidovudine treatment is only sometimes effective. Low doses of noradrenergic antidepressant medications, such as desipramine (Norpramin), nortriptyline, or amitriptyline less than 50 mg, and carbamazepine (Tegretol) are useful in suppressing symptoms.

Patients with more advanced immunodeficiency can also experience chronic progressive or relapsing inflammatory demyelinating polyneuropathies, of which mononeuritis multiplex may be a variant. The various neuropathies may also be complicated by a primary myopathy, which has been associated with HIV infection.

Secondary Neurobiological Complications The most important secondary neurobiological complications of HIV infection are brain infection by *Toxoplasma gondii, Cryptococcus neoformans*, primary CNS lymphoma, and delirium secondary to metabolic derangements or pharmacological toxicity.

Toxoplasmosis Toxoplasmosis is the most prevalent cause of space-occupying brain lesions in AIDS. The disease represents reactivation of a highly prevalent pre-existing infection by the protozoan *T. gondii* in the setting of advanced immunosuppression. It is seen almost exclusively in patients with fewer than 200 CD4+ T lymphocytes per microliter. Cerebral toxoplasmosis typically presents with seizure or focal neurological disturbances, such as hemiparesis, in combination with subacute or acute changes in mental status. MRI brain scanning reveals single or multiple lesions of high-signal intensity on T2-weighted images (Fig. 2.8-8). The lesions enhance with administration of gadolinium or another contrast agent. In an AIDS patient with a positive serological test for toxoplasma and a contrast-enhancing cerebral mass lesion, empiric treatment with pyrimethamine (Daraprin) and sulfadiazine (Microsulfon) is indicated, and will typically produce both clinical and radiographic improvement in 7 to 14 days. The prior probability of toxoplasmosis is markedly diminished, however, in patients whose toxoplasma serology is negative, or in those who have been taking a sulfa antibiotic reliably. Under these circumstances, a brain biopsy may be obtained to diagnose other potential causes, principally primary CNS lymphoma. Since the widespread use of trimethoprim-sulfamethoxazole for prophylaxis against pneumocystic pneumonia in patients with CD4+ lymphocytes <200 in the United States, the incidence of cerebral toxoplasmosis has dropped considerably.

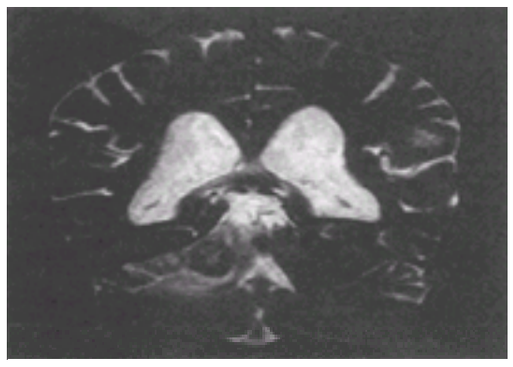

FIGURE 2.8-8 This coronal T2-weighted image from a 49-year-old man demonstrates a right cerebellar hemispheric mass, which was determined to be due to Toxoplasma infection. This scan also reveals brain atrophy with moderate enlargement of the lateral ventricles and increased fluid in the sulci (fluid is white on this T2-weighted image). (Courtesy of John R. Hesselink, M.D.)

Cryptococcus Meningitis *C. neoformans* is a yeast capable of infecting the meninges and brain of immunocompromised persons; perhaps 10 percent of persons with AIDS develop this serious complication. The presenting features are fever with meningeal signs (headache, neck stiffness, nausea, and delirium). On examination, there may be signs of cranial nerve involvement and papilledema. CT and routine MRI scan are sometimes unremarkable, although evidence of meningeal involvement may be seen on MRI enhanced with gadolinium, which is helpful in visualizing areas of blood-brain barrier permeability.

Although there is generally a favorable response to specific treatment, which combines amphotericin B (Fungizone) and flucytosine (Ancobon), relapse is common and early death is a frequent outcome. As with toxoplasmosis, initial treatment needs to be followed by lifelong suppressive therapy with agents such as fluconazole (Diflucan) or amphotericin B; treatment with the latter entails placement of an indwelling catheter since it is given intravenously.

Cytomegalovirus (CMV) Encephalitis Herpes viruses can produce encephalitis in immunocompromised persons. CMV commonly coinfects the brain in late-stage AIDS. If encephalitis develops, it can present with fever, rapidly progressive delirium leading to obtundation, and seizures. The condition may be difficult to distinguish from HIV-associated dementia, although the MRI brain scan, if positive, may show typical areas of periventricular and subependymal abnormal signal. Genciclovir and foscarnet (Foscavir) are virustatic, but long-term efficacy has yet to be established. By enhancing the return of immune function, protease inhibitors appear to be lowering the prevalence of CMV disease in the CNS.

Progressive Multifocal Leukoencephalopathy Many healthy persons are asymptomatic carriers of papovaviruses (human papilloma or wart viruses), but in the setting of severe immunodeficiency these viruses can produce a catastrophic encephalopathy. The patient may present with multifocal neurological signs, including hemiparesis, language disturbance, gait difficulty, and a progressive delirium.

CT scanning may show zones of subcortical lucency, and MRI scan may reveal patchy areas of high signal in the subcortical white matter (Fig. 2.8-9). Historically, the disease progresses rapidly to death but reports of stabilization and marked improvements in clinical neurological status in patients with progressive multifocal leukoencephalopathy have appeared since the introduction of highly active combination antiretroviral therapies.

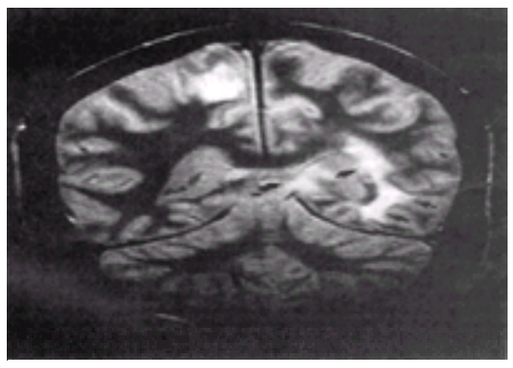

FIGURE 2.8-9 This coronal MRI taken from a 48-year-old HIV positive man with AIDS demonstrates a hyperintensity in the subcortical white matter in the left parietal lobe due to progressive multifocal leukoencephalopathy (PML) (arrows). From other images it was observed that the lesions were multifocal, extending from frontal to occipital lobes, and involving the basal ganglia. (Courtesy of John R. Hesselink, M.D.)

Lymphoma AIDS is probably the most common condition associated with primary lymphoma of the CNS. These B-cell lesions tend to arise multifocally and produce a slowly deteriorating neurocognitive disorder that may at first be confused with HIV-associated dementia, but later progresses to multifocal signs and signs of increased intracranial pressure. On MRI, lymphoma is difficult to differentiate from toxoplasmosis; increased uptake of thallium-201 on SPECT is more common in lymphoma than toxoplasmosis. Brain biopsy is usually required to make a firm diagnosis (Fig. 2.8-10). Many clinicians empirically treat for toxoplasmosis and reserve invasive diagnostic procedures until such treatment clearly fails. There is no effective treatment for the lymphoma although radiation therapy can reduce intracranial pressure.

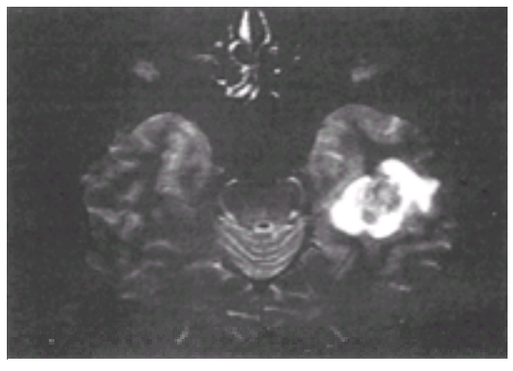

FIGURE 2.8-10 Axial T2-weighted image from a 41-year-old man with AIDS demonstrating a mass in the left temporal lobe (arrows) later determined to be a lymphoma. (Courtesy of John R. Hesselink, M.D.)

Delirium Delirium exists when a patient experiences rapidly evolving cognitive decline, especially in the areas of attention, learning and remembering new information, and orientation with reduced or fluctuating alertness (clouding of consciousness). Although intrinsic disorders (i.e., infections, neoplasms, and so on) are common causes, delirium can also reflect metabolic derangements or pharmacological toxicity.

Epidemiology The prevalence and incidence of delirium in HIV illness are unknown, although delirium (along with adjustment disorder) is generally one f the most frequent diagnoses made by psychiatric consultation services when evaluating hospitalized HIV patients. It is likely that delirium is as underdiagnosed and undertreated in HIV units and in general surgical and medical wards.

As treatments for advanced HIV disease have improved, an increasing number of patients who otherwise would have been hospitalized are now being managed in ambulatory settings. Therefore, physicians working in outpatient clinics and in private consultation offices need to maintain a high index of suspicion to the possible emergence of delirium in symptomatic HIV patients, regardless of the setting.

Etiology The combination of systemic illness, neurocognitive impairment, and multiple medications with CNS effects place the person with symptomatic HIV disease at high risk for delirium. Therapeutic agents likely to be associated with delirium are described in Table 2.8-4. Particularly noteworthy are those agents with anticholinergic properties, such as amitriptyline and chlorpromazine (Thorazine), along with antiemetic agents (scopolamine, prochlorperazine [Compazine], antihistamines, and benzodiazepines. Opportunistic or other infections, metabolic factors (hypoxemia, hypercarbia, electrolyte imbalances, hypoglycemia), hepatic and renal dysfunction, surgical intervention, and psychoactive substance use and withdrawal states can all contribute to likelihood of delirium.

Table 2.8-4 Neuropsychiatric Side Effects of Therapeutic Drugs for HIV

Pneumocystis pneumonia can be complicated by hypoxemia, which can lead to clouding of consciousness. Similarly, delirium has occasionally been reported as a complication of zidovudine and ganciclovir. Other causes of delirium in late-stage AIDS include severe nutritional deficiencies (e.g., vitamin B_{12}) and electrolyte imbalance (e.g., hyponatremia).

Diagnosis and Clinical Features Both hyperactive (vigilant) and hypoactive (withdrawn) deliria occur. Detecting the essential features of delirium, including impairment of sustaining and shifting attention, disorganized thinking, altered level of alertness, increased or reduced psychomotor activity, and memory impairment requires careful interviewing and discipline on the part of the examiner, who may be conducting an assessment in a busy medical unit, or who may be distracted by the patient's flamboyant, delusional, or hallucinatory symptomatology. In outpatient settings deliria may be even more difficult to detect, based on incipient or subtle presentations or on an interviewer's lower index of suspicion. Detecting delirium is crucial because it is a medical emergency requiring prompt assessment for potentially lethal conditions (e.g., opportunistic brain infection, tumor, or cerebrovascular events).

Differential Diagnosis The differential diagnosis includes substance intoxication delirium, substance withdrawal delirium, and delirium due to a general medical condition; deliria can also be imposed on dementia. The standard neuromedical workup for deliria should be pursued, along with toxicology screening and brain imaging (MRI or CT), if possible.

Course and Prognosis The course and prognosis of deliria in the HIV-positive patient appear to mimic the course and prognosis of other medical illnesses. There is a prodromal phase, which may be brief or of several days' duration, an acute phase, when the diagnosis is made, and then either rather prompt resolution or a more persisting subacute phase, which may last for several days or weeks. Excess morbidity from delirium beyond its neuromedical etiology may result from suicide, injury while fleeing delusional dangers, or assault based on paranoid perception of caretakers. If delirium is associated with an underlying HIV dementia, the prognosis may be especially poor.

PSYCHOBIOLOGICAL ASPECTS OF HIV INFECTION

The fundamental premise of modern psychiatry is that emotional life and behaviors have biological underpinnings that develop and are shaped within a broader social context. Perhaps that interplay is nowhere more evident than in the psychobiology of HIV infection. The HIV pandemic in Western industrialized nations now includes both the initial risk groups (homosexual men and injection drug users), growing number of women, and adolescents; the pandemic crosses diverse ethnic and socioeconomic strata. Further, dramatic medical advances have begun to alter the natural history of the illness from one of accelerating deterioration to a more chronic course resembling that of chronic cardiac and neurological diseases. The changing face of the HIV pandemic suggests that the role of clinical psychiatry must expand to encompass new populations and to adapt to changing neuromedical aspects of illness. The psychiatrist involved in the care of HIV-infected patients will be asked to address patients' premorbid psychiatric disorder and as well as new-onset emotional distress, which may complicate a patient's response to infection, disease, and treatment.

Etiology Figure 2.8-11 illustrates several models for understanding the origin of psychiatric conditions observed in the course of infection and illness from HIV. The transition model proposes that adverse psychological phenomena (e.g., adjustment disorders, mood disorders, anxiety disorders) may be anticipated at key points, such as discovery of seroconversion, initiation of antiretroviral treatment, onset of physical symptoms, advance in HIV illness stage, and HIV-related bereavement, and that they reflect a breakdown in coping capacities in the face of HIV.

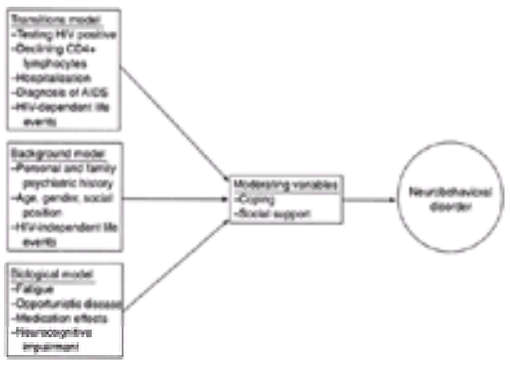

FIGURE 2.8-11 Models for understanding neurobehavioral conditions in HIV infection and illness.

The biological model suggests that HIV involvement of the CNS may be associated directly with depressive or manic syndromes or psychotic phenomena. Patients may respond to their awareness of HIV-associated neurocognitive decline with psychological distress, or the neurocognitive decline itself may increase the likelihood of such symptoms.

A third conceptualization, which may have been underestimated, is the so-called background model. It emphasizes that a psychiatric disorder that preceded the HIV epidemic may emerge during the course of an HIV illness for reasons not directly associated with HIV transitions; this model takes note of evidence from several medical illness populations that stressful life events not directly attributable to the medical condition may determine the level of emotional distress more strongly than do illness-related events. In sum, there is more to HIV psychiatry than HIV.

Social Factors No other epidemic of the twentieth century has been associated with as much social stigmatization and political debate as has HIV. Fear of HIV and AIDS is widespread in the general community (including among many medical and mental health caregivers) and is based on concerns about contagion as well as on bias. Although there is evidence of increased tolerance, the practical consequences of infection still include loss of employment, denial of medical benefits or life insurance, and derailment of career aspirations. Clinicians treating HIV-infected persons must be careful not to underestimate these social phenomena and their impact on individual patients, and they should be willing to examine their own fears and biases about HIV and their own mortality.

Apart from the effects of society's response to HIV, several studies suggest that the patient's immediate social support is a key moderator of the effects of illness and of life events on mood and function. Regardless of stage, HIV-seropositive persons with satisfying and stable social support, especially those who cope actively with those life events, will be significantly less distressed than those with unsatisfying or inconsistent support.

Psychological Factors In Western industrialized nations the two major groups that have been at highest risk for infection with HIV—homosexual men and intravenous drug users—may be at elevated risk for psychological difficulties that precede the date of infection. In failing to take a careful history, a psychiatrist can make the error of ascribing to HIV infection and its consequences psychological disorders that may have little direct relation to the infection as such.

A series of studies comparing the lifetime prevalence of diagnosable psychiatric disorders between seronegative homosexual and heterosexual men matched for age and social position suggest there may be more syndromal disorders in homosexual men than in heterosexual men. The lifetime prevalence of major depressive disorder in seronegative homosexual men ranged from 30 to 35 percent compared to the expected rates of about 3 percent from the Epidemiologic Catchment Area (ECA) surveys of community controls similar in age. Similarly, the lifetime prevalence of alcohol use disorders ranged from 25 to 40 percent compared to expected rates of 15 to 20 percent in epidemiological surveys. As for nonalcohol substance-related disorders (excluding intravenous drug abuse), lifetime rates in the cohorts of homosexual men ranged from 20 to almost 40 percent, whereas data from the ECA studies suggest expected rates of around 10 percent for men similar in age.

Inferences about biological predisposition and recurrent disorder must be tempered by the realization that in the majority of cases mood disorders first occurred in the context of disclosure or conflict over sexual orientation, and that the use of alcohol or other drugs may be prevalent among homosexual men as a way of coping with stigmatization. Nevertheless, little is known at present about the interaction of HIV and psychiatric history, and careful clinical follow-up for recurrence is an important research question and a prudent clinical approach.

For individuals with injection drug use disorders, studies preceding the HIV epidemic indicated high lifetime rates of personality disorders, anxiety disorders, and mood disorders. Elevated lifetime rates of antisocial personality disorder (ranging from 25 to over 50 percent), phobias (10 percent), and major depressive disorder (20 to 25 percent), are reported, with even higher rates being noted in persons who have multiple drug use disorders or who have both an alcohol use disorder and another substance-related disorder. Apart from formal psychiatric disorder, personality inventories such as the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) have revealed persisting symptoms of depression, anxiety, low self-esteem, and ego boundary disturbances among individuals with substance-related disorders.

In general, across all risk groups, women and persons of minority ethnicity are among the fastest-growing HIV risk groups. There is concern that these groups may also be at increased risk for emotional distress, given the potential socioeconomic disadvantages and the stigmatization of being female, homosexual, and of racial or ethnic minority. More research is needed to address the needs of those populations. Vulnerability to emotional distress (anxiety and depressive symptoms) appears to be associated with a personal or family psychiatric history, perception of low social support, use of passive and avoidant coping approaches, and the experience of stressful life events apart from HIV seropositivity itself, along with the demographic characteristics of youth, low education, and unemployment. HIV-seropositive persons with hemophilia without identifiable vulnerability factors show comparable levels of emotional adaptation. Those with any one risk factor were at greater risk of distress whereas those with multiple risk factors comprised the most distressed group.

Understanding background assets and vulnerabilities affords the clinician more diverse opportunities to treat psychiatric adjustment, to enhance coping capacity, to retain or provide social support, and to make a more meaningful contribution than would be possible if one's view were restricted to treating HIV-related issues alone.

TRANSITION POINTS IN HIV SYNDROMES

Seroconversion and Serotesting Phenomena The point of testing seropositive is associated with many of the same responses described for other life-threatening illness: shock, anger, anxiety, guilt, and denial. Fear of disclosure of serostatus, of loss of control, of abandonment by family and friends, of the inability to work, of medical expenses, and of pain may be evident. Anger may be directed at friends, family, lovers, physicians, and other caregivers. Loss of self-esteem and self-blame are common, as is regret or guilt over lifestyle (e.g., drug use). Conflicts about trusting and relying on physicians and others may emerge. The reality of death may be confronted. Psychological issues surrounding seroconversion continue throughout the course of infection and adaptation to illness.

The CDC guidelines recommend routine HIV testing in hospitals and clinics when seroprevalence exceeds 1 percent. The validity of this recommendation has been debated, but voluntary testing is generally advocated on the grounds that behavioral interventions are the only immediately available prevention strategies against HIV transmission, and that effective medical therapies are available for HIV-associated disease. Further, some psychiatric disorders (e.g., bipolar I, disorder, alcohol use disorders, severe personality disorders) are believed to confer increased risk for HIV because of their association with poor impulse control, impaired assessment of risk, the bartering of sex for illicit drugs, and other factors. General guidelines for voluntary HIV testing and counseling are described in Table 2.8-5.

1. Discuss meaning of a positive result and clarify distortions (e.g., the test detects exposure to the AIDS virus; it is not a test for AIDS).
2. Discuss the meaning of a negative result (e.g., seroconversion requires time, recent high-risk behavior may necessitate follow-up testing).
3. Be available to discuss the patient's fears and concerns (unrealistic fears may require appropriate psychological intervention).
4. Discuss why the test is necessary. (Not all patients will admit to high-risk behaviors.)
5. Explore the patient's potential reactions to a positive result (e.g., "I'll kill myself if I'm positive"). Take appropriate necessary steps to intervene in a potentially catastrophic reaction.
6. Explore past reactions to severe stresses.
7. Discuss the confidentiality issues relevant to the testing situation (e.g., is it an anonymous or nonanonymous setting?). Inform the patient of other possible testing options where the counseling and testing can be done completely anonymously (e.g., where the result is not made a permanent part of a hospital chart). Discuss who has access to the test results.
8. Discuss with the patient how being seropositive can potentially affect social status (e.g., health and life insurance coverage, employment, housing).
9. Explore high-risk behaviors and recommend risk-reducing interventions.
10. Document discussions in chart.
11. Allow the patient time to ask questions.

Adapted from Rogers RR, Ciesla JA, Denicoff KI, Akhtar AR: Laboratory and Diagnostic Testing in Psychiatry. American Psychiatric Press, Washington, DC, 1989.

Table 2.8-5 Pretest HIV Counseling

Contrary to what one might predict, studies using standardized questionnaires and rating scales suggest that symptoms of depression and anxiety are usually only moderately elevated from baseline in groups of individuals examined before and immediately after HIV antibody test results are announced, and the symptoms generally return to slightly below immediate pretesting levels within 3 months. Initial concerns about widespread suicidal ideation or suicide attempts following testing positive for HIV have been unfounded. The most prevalent psychiatric disorder in this phase is an adjustment disorder with depressed or anxious features, the prevalence of which has been as high as 20 percent in some samples. Estimates of the prevalence of psychiatric disorder in response to seroconversion are speculative for several reasons—it is unknown whether distress or psychiatric mood disorder contributed to the acquisition of HIV or led to a request for serotesting or whether the setting or the positive test result brought about the psychiatric disorder.

Physically Asymptomatic, Early Symptomatic, and Later Phases Psychologically, the physically asymptomatic person must balance the likelihood of a foreshortened life expectancy with the fact of living in the present, evaluating plans for the future, and confronting the need for medical follow-up. Individuals with plasma HIV RNA at £10,000 copies per mL and a CD4+ lymphocyte count of <500 are candidates for early aggressive combination antiretroviral therapy. Ambivalence or skepticism about treatment, lack of concern about HIV, and doubts about the need for medications in the absence of symptoms may be evident.

Most studies consisting of samples of homosexual men indicate that levels of anxious and depressed moods are similar across physically asymptomatic seropositive men and seronegative at-risk controls. Some investigations indicate that on an annual basis over 20 percent of seropositive individuals experience episodic anxiety symptoms lasting at least 1 month, compared with negligible rates among community low-risk controls. In early-stage illness complaints of fatigue and insomnia are likely to reflect underlying major depressive disorder. As persons become medically more symptomatic, distress may increase. There is some evidence that patients with more advanced symptoms may be more distressed, perhaps related to the progression of their AIDS or fears of transition to frank AIDS. Physical symptoms of HIV disease (e.g., loss of appetite, autonomic arousal) may increase as depressed mood deepens.

In general, patients who are younger, socioeconomically disadvantaged, or with reduced levels of social support (e.g., the “closeted” homosexual person or undisclosed drug-using individual) are most likely to experience greater psychological distress. Beyond that, the perception of physical symptoms—regardless of HIV status or CDC stage—(e.g., lymphadenopathy, skin rash, cough, fever, sore throat) appears to be strongly associated with psychological distress. Even when signs of disease cannot be documented, the patient's conviction about bodily complaints helps to determine psychological distress.

If syndromic diagnoses (as contrasted with self-report of psychological symptoms) are considered, rates of current psychiatric disorder are similar across physically asymptomatic and mildly symptomatic persons and do not differ markedly from rates in seronegative at-risk controls (such as homosexual men). The current rate of an adjustment disorder is approximately 5 percent, as is the rate of major anxiety disorder (generalized anxiety disorder or panic disorder). The point prevalence of major depressive disorder is about 8 percent, which is similar to rates found in other chronic medical illness, outpatients with end-stage renal disease, or patients with insulin-dependent diabetes mellitus. Current diagnoses of an alcohol-related disorder (in non–drug-abusing risk samples) are usually in line with the expected community rates.

As the biology of illness asserts itself forcefully, the picture may change, and in frank AIDS or in hospitalized patients higher rates of psychiatric disorder (especially depression) have been detected. Most of those data were collected early in the epidemic. It is unclear if such rates would still pertain given the prevailing therapeutic advances, the ability to maintain better overall health among those living with AIDS, and the possibilities that antiretroviral treatment may forestall neurocognitive decline. In any event, these rates presently are similar to the rates found in patients hospitalized with non-HIV medical disease.

Psychosis and manic states arising de novo are uncommon in HIV illness, but when such syndromes occur, they are usually in the context of neurocognitive disorder and frank AIDS. Psychosis and manic conditions may be associated with early mortality. Other biological psychiatric disorders encountered in later-phase illness include deliria and dementia. To these causes must be added iatrogenic phenomena from treatment, as described in Table 2.8-5.

OTHER PSYCHIATRIC DISORDERS

Adjustment Disorders

Epidemiology, Etiology, Diagnosis, and Course There are no studies yet that predict rates of adjustment disorder and attempt to link them to transition points in HIV illness or to intercurrent non-HIV events. The rates of adjustment disorder with depressed, anxious, or mixed features vary widely according to the population studied. For example, in cohorts of military personnel who have undergone mandatory testing and who are examined shortly after being informed of serostatus, rates of an adjustment disorder approach 15 to 20 percent. On the other hand, studies of men who have known their serostatus for at least 1 year after voluntary testing reveal negligible rates of adjustment disorder even though these men may be dealing with HIV staging or treatment concerns. The clinical course of these HIV-related adjustment disorders is usually benign and responsive to education and support.

ANXIETY DISORDERS

Epidemiology The rates of the major anxiety disorders, panic disorder, obsessive-compulsive disorder, and generalized anxiety disorder appear to have lifetime prevalences and current prevalences within the range expected from epidemiological studies (e.g., 1 to 2 percent).

Etiology Preoccupation with HIV can be the manifest content of panic and of obsessive-compulsive disorders or the probable latent source of generalized anxiety disorder. There is no systematic evidence that anxiety disorder due to HIV disease occurs in the context of HIV infection of the brain.

Diagnosis, Clinical Features, and Differential Diagnosis The diagnosis and clinical features of the major anxiety disorders are usually straightforward, and in many cases a history of anxiety may precede the likely time of infection. The differential diagnosis must include HIV-related and non–HIV-related endocrinological disorders and concurrent cardiopulmonary disorders, anemia associated with HIV, intoxication and withdrawal syndromes, and major depressive disorder.

Course and Prognosis Almost no data are available, but the course and outlook do not in general seem to deviate from that expected in non-HIV samples.

MOOD DISORDERS

Major Depressive Disorder

Epidemiology The lifetime prevalence of major depressive disorder among men with homosexual or injection drug use risk factors is estimated to range between 25 and 35 percent, with some studies suggesting that in the majority of cases the first episode of major depressive disorder preceded the likely date of seroconversion. The point prevalence of major depressive disorder has generally been in the range of 4 to 8 percent for seropositive homosexual or bisexual men, with rates being similar across medically asymptomatic seropositive and moderately symptomatic persons; there may well be an increase in frank AIDS. The point prevalence in

women injection drug users approaches 15 to 30 percent across all stages of disease. Some early estimates based on chart review suggested that 40 percent of hospitalized individuals with frank AIDS (more advanced disease) may meet criteria for major depressive disorder.

Etiology It is not yet determined whether HIV itself increases the likelihood of major depressive disorder beyond the expected increases found in other chronic diseases. HIV infection can be associated with anergic-apathetic-fatigue states, presumably mediated by release of somnogenic lymphokines. Decline of immunocompetence as reflected by lower CD4+ lymphocyte count is not itself associated with mood disorder. The relative etiological weight of illness factors (e.g., initiating antiretroviral therapy, noting the onset of constitutional symptoms, diagnosis of frank AIDS) and other factors, such as adverse life events independent of HIV infection or pre-existing mood disorder, are being explored in prospective studies. Previous mood disorder, substance use disorder, and neurocognitive impairment are thought to predict subsequent depressive episodes.

Diagnosis and Clinical Features The diagnosis of major depressive disorder is rather straightforward provided that the patient has few medical symptoms and generally well-preserved cognitive capacities. As the illness progresses, somatic symptoms of disease (e.g., fatigue, weight loss, insomnia) or of neurocognitive impairment (e.g., slow thinking, diminished powers of concentration, or forgetfulness) may complicate diagnosis. In the presence of somatic symptoms, the diagnosis of major depressive disorder should not be made unless there are prominent symptoms in such areas as pervasive sadness, diminished interest or pleasure, feeling of worthlessness or guilt, indecisiveness, and morbid preoccupation with death. In particular, diminished activity based on fatigue must be discriminated from withdrawal due to lack of interest or pleasure. The interviewer must be careful to distinguish a true sense of worthlessness from one of disappointment at not being able to perform at full capacity; persisting guilty rumination must likewise be distinguished from episodic guilt associated with drug use or from having a sexually transmitted disease, and from regret over lost opportunity. In terms of evaluating thoughts of death, which might be expected to occur in individuals with HIV illness, the level of morbid preoccupation and a change of focus from usual thoughts or attitudes about death must be assessed. Because many HIV-positive persons have considered suicide at some point in the illness, the interviewer must note whether suicidal ideation is held out as a plan potentially to use at some future date as a way of maintaining control, or if it represents a self-destructiveness indicative of depression, based on a morbid sense of failure, worthlessness, or past transgressions.

Differential Diagnosis The differential diagnosis includes mood disorder (depressed) due to a non-HIV medical condition, early delirium, HIV-associated neurocognitive disorders, and apathetic-anergic syndromes. Since establishing the diagnosis can be difficult in advanced illness, it is important to combine meticulous neuromedical evaluation, psychiatric interviewing, and neuropsychological assessment. Electroencephalography may help to identify a delirium and is especially useful if baseline recordings are available for comparison. Evaluation may reveal more than one disorder concurrently, and treatment will generally be directed sequentially at the most prominent symptoms.

Cognitive complaints of mental slowing and lack of concentration can be part of a depressive picture or evidence of neurocognitive disorder. In physically asymptomatic stages, cognitive complaints are especially likely to reflect depressed mood rather than cognitive impairment. Conversely, persons with mild neurocognitive disorder documented by neuropsychological testing do not have markedly high rates of mood disorder. Therefore, the full evaluation of cognitive complaints requires both detailed mood assessment and neuropsychological testing.

Course and Prognosis The annual risk of an episode is from 10 to 20 percent; the majority of cases may be classified as mild to moderate in severity. The course of individual episodes is variable and is similar to that found in major depressive disorder in psychiatric populations: most persons have infrequent, widely spaced episodes whereas a few have frequent episodes. The usual prognosis is for remission either spontaneously or with treatment, although a more chronic course will ensue in a significant minority of cases. In some studies the persistence of depressive symptoms is associated with a heightened risk of mortality from AIDS.

Mania Mood disorders with manic features, with or without hallucinations, delusions, or a disorder of thought process can complicate any stage of HIV infection, but most commonly occur in late-stage disease complicated by neurocognitive impairment.

Epidemiology Precise estimates of prevalence or incidence are not available but may approach 1 percent. Some evidence suggests the rate in AIDS is up to ten times the general population rates.

Etiology In non–HIV-infected patients, mood disorders with manic features are noted to occur at therapeutic dosages of many medications and in a wide variety of neurological conditions, such as cerebrovascular disorder, meningitis, and tumor, and these sources can also be causative in HIV illness. Steroids, zidovudine, and ganciclovir are the most frequently reported iatrogenic causes. There are two typical onsets of mania in HIV. Manic states with onset early in HIV are associated with personal or family history of bipolar I disorder. Manic syndromes in persons without previous personal or family history of mood disorder usually have their onset late in the course of HIV illness and have a higher prevalence of comorbid dementia; in these cases the etiology is presumed to be related to the pathophysiology of HIV infection of the CNS.

Diagnosis and Clinical Features Cases may first have a period of days or weeks of change in personality (e.g., become pompous, belittling, and sexually inappropriate) and then progress into a fuller manic picture. The mood disturbance is generally elevated, with associated grandiosity. Diminished need for sleep, loquaciousness, teeming thoughts, and unrealistic personal or financial plans may appear. Irritability may be prominent in AIDS. In early-stage disease the neuromedical evaluation is generally unremarkable apart from findings expected for HIV disease, but personal and family histories of mood disorder are evident. In cases first appearing in the context of AIDS, personal and family histories are usually unremarkable but there is clinical dementia or neurocognitive impairment.

Differential Diagnosis Differential diagnosis includes brief psychotic disorders, the substance-related disorders, bipolar I disorder, neuromedical causes of mania, an evolving delirium, or manic symptoms that herald HIV-associated psychosis or dementia. Brief psychotic disorder is distinguished by hallucination, delusions, and bizarre or disordered thinking. Substance-related disorders can be distinguished by history and toxicology screen; bipolar I disorder may be suspected if there is prior personal or family history of mood disorder. Discontinuation of suspect medication and medical assessment may reveal iatrogenic or neurological sources.

Psychotic Disorders Psychotic symptoms are usually later-stage complications of HIV infection. They require immediate medical and neurological evaluation and often require management with antipsychotic medications. Whereas psychotic symptoms can obviously occur in deliria or can reflect neurological or primary psychiatric disorder or iatrogenic origins, there is also considerable interest in new-onset psychosis, wherein these etiologies do not seem to be present (e.g., psychotic disorder due to HIV disease).

Epidemiology Prevalence estimates vary widely depending on methodology: large-scale surveys find a prevalence of new-onset psychosis at less than 0.5 percent whereas chart review methods find frequencies ranging from 3 to 15 percent in persons for whom obvious causes (e.g., delirium) have been excluded. Thus, psychotic symptoms may be uncommon but not rare in HIV-infected populations.

Etiology Beyond known causes of psychosis (delirium; substance-related disorders; neurological vascular disease, opportunistic infection, or neoplasm; and iatrogenic causes) several theories have been offered to explain new-onset psychosis in which no known cause can be identified: these include direct effects of HIV on the brain; other CNS viruses, such as cytomegalovirus or human herpes virus; coinfection with two or more viruses; or brief psychotic disorder.

Diagnosis and Clinical Features The clinical presentation in new-onset psychosis is extremely variable. The most prevalent symptom seems to be delusions (occurring in almost 90 percent of cases in some series) with persecutory, grandiose, or somatic components. Persecutory themes can be quite elaborate, with patients proclaiming messianic themes or believing that they have been accorded superhuman powers by God. Somatic delusions may include being shot through with electricity or lasers. Delusions of thought insertion, thought broadcasting, and of passivity and control are also described. Most patients experience auditory hallucinations, with perhaps one half of those also experiencing visual hallucinations. A majority of series also report disorders of thought process, including looseness of associations or frankly disorganized thinking. Disturbances of mood commonly coexist, with anxiety being the most prevalent symptom, followed by depressed mood, euphoria, or irritability, and mixed depressed and euphoric states. Lability, flatness, and inappropriate laughter or anger are also described; bizarre behavior is commonplace. There are reports of persons eating dirt to conquer their fear of germs, and one patient painted himself and his entire apartment, including furniture and appliances, with green paint in an effort to “celebrate life.”

Bedside examination reveals impairment in memory or other cognitive functions in perhaps one third of patients; more comprehensive and formal neuropsychological assessment would be likely to detect neurocognitive difficulties in a larger proportion of cases, but many psychotic individuals are unable to complete such examinations.

Neurological findings are infrequent and nonspecific, usually consisting of ataxia, mild increases in motor tone, hyperreflexia, and tremor, but bizarre grimacing and posturing can also be present. Cerebrospinal fluid is generally unremarkable except for the mild pleocytosis common to HIV infection. Diffuse cortical slowing has

been reported in about one half of patients on whom an EEG was performed. CT and MR scans reveal nonspecific cerebral atrophy in about one half of cases; rarely, focal abnormalities are evident, suggesting tumor, opportunistic infection, or vascular etiology.

Differential Diagnosis The differential diagnosis includes brief psychotic disorder, major depressive disorder (with psychotic features), schizophrenia, psychoactive substance-related disorders, and iatrogenic causes (Table 2.8-4), cerebrovascular disorder, CNS neoplasm, and opportunistic infection (e.g., cryptococcosis, toxoplasmosis, human herpesvirus). The history and basic laboratory evaluation (e.g., toxicology screening) will usually make psychiatric etiologies clear. A thorough neuromedical assessment must be performed in order to rule out a remediable neurological cause.

Course and Prognosis The course and prognosis are highly variable and depend partly on whether specific complicating conditions coexist. For example, concurrent dementia due to HIV disease indicates a poor prognosis. Here the course may be one of rapid deterioration, with death within 6 months or less. However, when neurocognitive abnormalities are not detected or are mild, individuals have been followed for up to 2 years in a stable, treated course. Nevertheless, there is some evidence that death occurs earlier in those experiencing psychosis than in nonpsychotic patients who have similarly advanced HIV disease. In the short term, prognosis with regard to symptom control is quite favorable, with most patients responding to low-dosage regimens of antipsychotic medications.

SUBSTANCE-RELATED DISORDERS

Epidemiology If injection drug users are excluded (for whom, by definition, lifetime prevalence of substance use disorders is 100 percent), the lifetime rates in the major risk group—homosexual and bisexual men—range from 20 to over 40 percent. While some studies show higher current (previous 1 to 6 months) rates of substance-related disorders among seropositive homosexual and isexual persons compared with those expected from community samples, others show roughly comparable or only slightly higher rates.

Substance abuse can influence the likelihood of HIV transmission in at least two ways. First, sharing of infected needles (“works”) by intravenous drug users can result in direct entry of HIV into the bloodstream. Second, even among noninjection drug users, intoxication can increase the likelihood of having sex with multiple partners, reduce the use of condoms, and, because of altered pain perception, increase the likelihood of traumatic anal intercourse. Thus, even though those at risk for HIV exposure may not abuse substances significantly more, the implications of such use may be drastically different.

Etiology The etiology of substance abuse in those at risk for HIV infection is no better understood than it is for substance abuse generally. It has been speculated that at least among homosexual and bisexual persons the developmental stresses of establishing an alternate sexual identity and of negotiating that identity within the context of family expectations and stigmatization by society may contribute to substance use as a way of mastering distress.

Course and Prognosis Several prospective studies are currently under way that will provide much-needed systematic data about the natural history of alcohol and nonalcohol substance-related disorders in various risk groups, as well as among HIV-infected persons. Studies of HIV-positive homosexual and bisexual men suggest that most incident alcohol and other substance use disorders are relapses of conditions whose original onset preceded the likely date of seroconversion. Furthermore, seropositive and seronegative homosexual and bisexual men are equally likely to experience an episode of the disorder within the follow-up period. Incident rates are generally less than 5 percent, annually, despite markedly elevated lifetime histories of substance-related disorder. It may be that within homosexual and bisexual communities the emphasis on health promotion and education about the adverse health consequences of alcohol and other drugs of abuse has had a positive effect. These preliminary data also suggest that the stress of HIV infection or of the threat of infection does not necessarily lead to an increased risk of developing an alcohol or nonalcohol substance-related disorder.

The course of substance use disorder and that of HIV may intersect, with important prognostic implications. For example, stimulants may accelerate HIV replication, especially in the CNS. These drugs (e.g., amphetamines and cocaine) may produce brain injury by vascular injury and ischemia, and by increasing dopamine and glutamatergic transmission, leading to cytotoxic injury and death of selected neurons. It is reasonable then to expect that persons with stimulant-use behavior and HIV infection would show very high rates of cognitive impairment; however, surprisingly few studies have been conducted to address this question. Rates of cognitive impairment in individuals who contracted HIV infection through injection drug use are reported to be similar to those of subjects who are infected through sexual contact. The rate and frequency of cognitive symptoms do not appear to be increased with chronic substance abuse, nor does evidence suggest that injection drug users have shorter survival or shorter time for progression to AIDS. Nevertheless, important questions remain. One of these concern possible effects of both chronic neurocognitive impairment and ongoing drug abuse on adherence to prescribed antiretroviral drug treatments. Also, because neurocognitive impairment increases risk for early death in HIV-infected people who are not substance users, additional impairment resulting from ongoing substance abuse may further influence the course of the disease.

Standard psychological therapies developed in working with patients with heart disease, cancer, or other life-threatening illnesses are generally applicable to the patient with HIV illness. Potential roles for the psychiatrist in the seroconversion and serotesting phase are to discuss the basic medical and psychological meanings of seropositivity, to help the person decide who should be informed of risk factors and serostatus, as well as the manner in which such disclosure should take place, to manage anxious or depressive symptoms, and to assist in developing a follow-up plan for HIV education and treatment, including self-help groups and safe sex practices. Many authorities believe that it is best to bring up issues of power of attorney for health care and custody of children earlier in the course of the disease rather than later. In later phases the clinician can assist in addressing issues brought up by monitoring changes in HIV RNA viral load, CD4+ lymphocyte count, watching for symptoms, and adherence to treatment, and by maintaining morale. It is appropriate to discuss treatment alternatives and issues of adapting to health change, and to address mortality and aggressiveness of end-stage care.

Prevention Efforts at primary prevention to stem the spread of the disease have been successful in some arenas but not in others. In industrialized countries, for example, screening procedures for blood banks have largely eliminated transfusion and blood-product–related spread. Public education efforts may also have altered sexual practices in some at-risk subgroups (e.g., certain homosexual and bisexual men).

Unfortunately, such efforts have been slow to evolve and have not yet been applied to the vast majority of at-risk persons. That there are difficulties in modifying the kinds of behaviors leading to HIV transmission should not come as a surprise to psychiatrists. One is dealing, generally speaking, with a biologically driven behavior (sex) that is highly motivated and strongly pleasurable, and is trying to balance that against risks that are often perceived as not highly imminent and whose consequences appear only many years later. With respect to drug abuse, behavior governed by anticipation of immediate pleasure or a sense of calm among individuals with impulsive and risk-taking habits, who may lack self-esteem and a sense of future, is difficult to modify through any conventional educational approach.

Compounding these individually determined resistances is the fractured attitude of society about issues such as sexuality, birth control, and drug abuse. Condoms may be safe and effective, but U.S. society has had difficulty implementing even this relatively simple preventive measure because of its reluctance to discuss sexuality and the specifics of modes of transmission explicitly in schools and in the media. Ambivalence about sexuality and decisions to put philosophical objections to birth control ahead of preventing the spread of a fatal disease compromise prevention efforts.

Nevertheless, evidence continues to accumulate that, if properly implemented and supported, certain techniques do lower the likelihood of HIV transmission. The elements of prevention programs that are currently thought to be efficacious were reviewed and described in the February 11 to 13, 1997 National Institute of Health (NIH) consensus statement. It presents evidence for efficacy of programs directed at men who have sex with men, adult women at risk for sexual transmission, serodiscordant couples, adolescents, and injecting drug users. For example, a program combining information on HIV transmission, with a focus on developing a responsible attitude both to self and others, including necessary skills training in negotiation of safer sex, and coupled with access to condoms, has resulted in a substantial reduction in new cases of HIV in military recruits in Thailand. It is important to note that, beyond being grounded in state-of-the-art notions of behavioral science, this program received strong unambivalent support from authorities. In contrast, in the United States, despite numerous reports and opinions from learned panels, including the NIH consensus panel, to the effect that needle exchange programs can reduce HIV transmission in injection drug users, there has been little political will to implement such programs; indeed, some jurisdictions have actually moved to criminalize such prevention efforts.

Postexposure Prophylaxis Prompt initiation of antiretroviral therapy following accidental exposure to HIV by needle stick or similar accidents can reduce the likelihood of an infection actually developing by about 80 percent. The average risk of infection from percutaneous injury is about 0.3 percent. While the seriousness of the exposure (e.g., amount of inoculum) needs to be weighed against the risks of antiretroviral therapy, most authorities recommend combination treatment with zidovudine 200 mg three times a day and lamivudine (Epivir) 150 mg two times a day for 4 weeks for all but the most trivial exposures. If the source of exposure was an AIDS patient or one with a high viral load, a protease inhibitor (e.g., indinavir [Crixivan] 800 mg three times a day can be added). Such prophylactic treatment is accompanied by monitoring for plasma viral load as well as for serostatus.

The matter of postexposure prophylaxis for cases other than accidental exposure or rape remains controversial. The risk of infection from vaginal intercourse is estimated at 0.05 to 0.15 percent, from receptive anal intercourse, 0.08 to 0.32 percent, and approximately 0.7 percent from using contaminated injection drug equipment. Treatment, as for accidental occupational exposure, can be initiated within 72 hours of exposure with zidovudine and lamivudine. However, because of costs involved and risks of developing resistant HIV strains, postexposure treatment may not be indicated for those who clearly intend to re-expose themselves by repeated high-risk behavior.

Treatment A growing list of agents that act at different points in viral replication has raised for the first time the hope that HIV might be permanently suppressed or actually eradicated from the body. At the time of this writing, the active agents were in two general classes: the reverse transcriptase inhibitors and the protease inhibitors. The reverse transcriptase inhibitors are further subdivided into the nucleoside reverse transcriptase inhibitor group and the nonnucleoside reverse transcriptase inhibitors. Table 2.8-6 lists the currently available agents in each of these three categories.

Generic Name	Trade Name	Usual Abbreviation
Nucleoside Reverse Transcriptase Inhibitors		
Zidovudine	Retrovir	AZT or ZDV
Didanosine	Videx	ddI
Zalcitabine	Hivid	ddC
Stavudine	Zerit	d4T
Lamivudine	Epivir	3TC
Abacavir	Ziagen	
Non-nucleoside Reverse Transcriptase Inhibitors		
Nevirapine	Viramune	
Delavirdine	Rescriptor	
Efavirenz	Sustiva	
Protease Inhibitors		
Saquinavir	Invirase	
Ritonavir	Norvir	
Indinavir	Crixivan	
Nelfinavir	Viracept	

Table 2.8-6 Antiretroviral Agents

Although the indications for initiating antiretroviral therapy have not been fully resolved at the time of this writing, the International AIDS Society-USA Panel (1997) recommended that therapy be initiated for individuals with plasma HIV RNA levels greater than 5000 to 10,000 copies per mL, regardless of CD4 cell count. The panel further recommended that all persons with detectable HIV in plasma (the nominal level of detection is in the range of 200 to 400 copies per mL) be provided such treatment on request. The Panel cautioned, however, that at present such treatment must be lifelong, is complex, requires strenuous effort at adherence, and results in many adverse effects ranging from unpleasant to serious. Complete and informed involvement on the part of the patient is essential because viral resistance to the various compounds can develop quickly even after brief periods of nonadherence. For these reasons, some physicians and patients may decide not to initiate antiretroviral therapy immediately and may prefer to follow the patient closely for any changes in plasma viral load.

Currently it is recommended that treatment be initiated with triple therapy, that is, a combination of two reverse transcriptase inhibitors plus one protease inhibitor. Three agents are recommended because an individual with a plasma viral load of 20,000 copies per mL is likely to harbor all possible resistant genotypes of HIV that can result from one- or two-point mutations. Thus, even the use of two agents leaves the distinct possibility that a few resistant strains will escape, leading to treatment failure. The likelihood that an infected person will harbor HIV genotypes with three independent mutants is very low; thus, triple therapy should be initiated where possible.

In choosing specific drug combinations, clinicians take into account both the mode of action of specific agents, and their drug-drug interactions. For example, zidovudine and lamivudine are both reverse transcriptase inhibitors; however, zidovudine appears to work best in actively replicating cells whereas lamivudine is most active in resting infected cells. Furthermore, zidovudine penetrates into the central nervous system; however, the penetration of lamivudine is low.

The goal of therapy is to maintain plasma viral load at an undetectable level. By early 1998 patients could expect to have to take treatment indefinitely. Indications for changing treatment include upswing in plasma viral load, toxic effect, intolerance, and nonadherence.

The antiretroviral agents have many adverse effects, too numerous to describe. Of importance to psychiatrists is that protease inhibitors are metabolized by the hepatic cytochrome P-450 oxidase system, and can therefore increase levels of certain psychotropic drugs that are similarly metabolized. These include bupropion (Wellbutrin), meperidine (Demerol), various benzodiazepines, and selective serotonin reuptake inhibitors (SSRIs). Therefore, prescribing psychotropic drugs to persons taking protease inhibitors must be done with caution.

In addition to the new nucleoside reverse transcriptase inhibitor, nonnucleoside reverse transcriptase inhibitors, and protease inhibitor agents that may soon become available, other classes of drugs are under investigation. These include agents that interfere with HIV cell binding and fusion, the action of HIV integrase, drugs that interfere with certain HIV genes such as *gag*, among others.

Beyond treatment directed specifically against HIV, many interventions are available to prevent and treat various complications of immunodeficiency caused by opportunistic viral, bacterial, fungal, and protozoan infections. There has been a substantial increase, both in survival and in quality of life, resulting from early diagnosis and treatment of these opportunistic conditions.

The use of combination antiretroviral regimens, combined with more specific treatments of complications, has prolonged the survival of asymptomatic as well as symptomatic HIV-infected persons. However, despite progress in maintaining patients longer and in better states of health, the ultimate outcome is still uncertain; that is, it is unclear at present whether any HIV-infected person can expect to escape developing AIDS and ultimately dying. HIV-infected persons are keenly aware of this prognosis and their concern sometimes takes the form of psychiatric disturbances. These phenomena, along with the primary and secondary neurocognitive complications related to HIV infection, are discussed next.

The introduction of potent antiretroviral drug combinations, and the development of sensitive quantitative assays of HIV RNA concentrations, has begun to revolutionize the treatment of HIV cognitive disorders as it has done for systemic disease. There is hope that measuring the viral load in CSF will be a window into brain infection and a guide for monitoring response to therapy.

Elevated concentrations of HIV RNA in CSF are associated with neurocognitive impairment in AIDS. High-dosage zidovudine monotherapy can ameliorate HIV-associated neurocognitive impairment, perhaps in part because of its good penetration of the blood-brain barrier. Since the introduction of protease inhibitors and combination antiretroviral therapy, zidovudine monotherapy is no longer an option for treatment; nevertheless, it remains an important component of combination regimens. Protease inhibitor drugs are not able to cross the blood-brain barrier in therapeutic amounts and therefore might not be expected to be effective in treating neurocognitive disorders. Nevertheless, combination of these drugs with other first-generation agents like zidovudine appears to prevent or even reverse the progression of these disorders. Clinical improvement both in terms of enhanced performance on standardized neuropsychological testing, and improvement in the pattern and severity of white matter signal abnormalities detected on MRI scans, can be seen within 2 to 3 months of beginning therapy. Perhaps one out of every three affected patients will show clinically significant improvement in neurocognitive function. It is not clear how protease inhibitors could arrest or even reverse HIV-associated neurocognitive impairment. It may be that a markedly reduced viral load no longer triggers the production of neurotoxic cytokines, excitatory neurotransmitters, and inflammatory agents by the immune system.

Novel treatments may also be useful. Neuronal excitotoxicity, mediated through the activation of glutamatergic receptors by the HIV envelope protein gp120, is a potentially important mechanism by which brain dysfunction might occur in HIV infection. Memantine is an open-channel antagonist of NMDA-type glutamate receptors that is generally well tolerated. It is currently being studied as a treatment for dementia in a large Phase III placebo-controlled clinical trial. On the theory that an agent that could dislodge gp120 from neural receptor sites might be promising, an octapeptide called *d-ala-peptide-t-amide (peptide t)* has been used in Phase II clinical trials. Compared to placebo, peptide t was associated with neuropsychological improvement in cognitively impaired individuals (with CD4 counts <200), and a reduced likelihood of progression of impairment on 6-month follow-up. Calcium channel inhibitors, which theoretically appeared useful, have not proven successful.

The remaining forms of treatment are principally supportive. The most important step is to exclude other potentially treatable conditions, such as secondary infections or neoplasia, metabolic abnormalities with low-grade delirium, or other psychiatric disorders (e.g., major depressive disorder). Once the diagnosis is clear, then the usual supportive measures for neurocognitively impaired persons should be employed. These include identifying areas of cognitive strength and deficit, reducing emphasis on areas that are now impaired (e.g., divided attention, speeded processing), emphasizing efforts to maintain good orientation and reality testing, and avoiding medications that may further compromise cognitive function, in particular, benzodiazepine drugs. If they need to be used, such medications should be at lower-than-usual doses. Additionally, antidepressant and antipsychotic agents, if indicated, may also have to be prescribed in much lower dosages (e.g., at 25 percent of the usual recommended dosage).

PSYCHOTHERAPY

Approaches Major psychodynamic themes for HIV-infected patients involve self-blame, self-esteem, and issues regarding death. The psychiatrist can help patients deal with feelings of guilt regarding behaviors that contributed to infection or AIDS. Some HIV and AIDS patients feel that they are being punished. Difficult health care decisions, such as whether to initiate or continue taking antiretroviral medication and terminal care and life-support systems, should be explored, and here denial of illness may be evident. Major practical themes involve employment, medical benefits, life insurance, career plans, dating and sex, and relationships with families and friends. The entire range of psychotherapeutic approaches may be appropriate for patients with HIV-related disorders. Both individual therapy and group therapy can be effective. Individual therapy may be either short-term or long-term and may be supportive, cognitive, behavioral, or psychodynamic. Group therapy techniques can range from psychodynamic to completely supportive in nature.

At the time of testing positive the HIV-infected person is almost immediately faced with questions about disclosing his or her serostatus. Most counselors suggest that recently infected individuals withhold disclosure to others until they feel well-enough informed about HIV and its treatment, and ready to answer questions that may arise. This readiness includes feeling informed enough to answer questions that others may ask in relation to their own safety, to differences between HIV and AIDS, to stage of disease and lifespan. A patient's readiness to accept the emotional consequences of disclosure—such as another's fury and rejection—must also be anticipated. In brief the HIV-infected person must assess not only his or her readiness, but the readiness of others to hear the announcement of seropositivity. Among the fears that must be confronted is the individual's concern that once serostatus has been revealed, he or she has lost control of who next knows of their seroconversion. In deciding whether or not to tell others the patient must address their sense of betrayal if they are not told. The same issues apply to the person's work environment. As a practical matter the individual may need to decide whether to tell a trusted colleague in case of a job-related accident that might put others at risk of infection. Similarly, parents must decide when or whether to tell their children. Some parents want to tell very young children as soon as possible, while other parents prefer to withhold this information until their child's teenage years, for fear of “taking away their childhood.” The question of custody of children after the parent's death will need to be considered. The same question of timing will arise about when to tell children that they are seropositive. The parent must balance fears that telling the child's school will lead to discrimination while guarding their child's and others' safety should there be an accident.

The psychiatrist may have a special role regarding HIV treatment. The advent of protease inhibitors, and the promise of additional increasingly effective therapies, has brought hopes of a “cure” to patients and physicians alike. Even patients who have failed one or more rounds of combination therapies may find that family, friends, and physicians continue to be optimistic. The psychiatrist may be the only “safe” person to whom the patient can express discouragement, weariness, fear of treatment failure, and fury or guilt for not being able to tolerate successful therapy or for not responding to regimens that have benefited others. The psychiatrist also may be the only one confronting unrealistic expectations of cure, or assumption that safer sex practices are no longer relevant. Paradoxically, the therapeutic task also may be to examine the patient's reaction to a reprieve from certain death—the so-called second life agenda.

Direct counseling regarding substance use and its potential adverse effects on the HIV-infected patient's health is indicated. Specific treatments for particular substance-related disorders should be initiated if necessary for the total well-being of the patient.

Support Systems The gay community has provided a significant support system for HIV-infected people, particularly homosexual and bisexual persons. Public education campaigns within this community have resulted in significant (more than 50 percent) reductions in the highest-risk sexual practices. Individuals are likely to practice safe sex if they know guidelines for safe sex, have access to a support group that reinforces such practices, are in a steady relationship, and have a close relationship with a person with AIDS. Unfortunately, injection drug users have been difficult to reach, but locating HIV treatment clinics in methadone (Dolophine) treatment settings and other innovative outreach efforts are in progress. Involving injection drug users in prevention efforts is particularly important because they are a major reservoir from which the virus spreads to women, heterosexual men, and children. The paradoxical effect of highly effective antiretroviral therapies may be to increase rates of high-risk sex and injection drug users, based on the belief that a cure is available or that undetectable viral load renders the person noninfectious.

Assessment The assessment of HIV-infected patients should include a complete sexual and substance-use history, a psychiatric history, and an evaluation of the support systems available to the patient. The clinician must understand the patient's history with regard to sexual orientation and substance use, and the patient must feel that the therapist is not judging past or present behaviors. A sense of trust and empathy can often be fostered by the therapist's asking specific, well-informed, and straightforward questions about the homosexual or substance-using culture. The therapist must also determine the patient's knowledge about HIV and AIDS.

Therapist-Related Issues Countertransference issues and the burnout of therapists who treat many HIV-infected patients are two key issues to evaluate on a regular basis. Therapists must acknowledge to themselves their predetermined attitudes toward sexual orientation and substance use so that those attitudes do not interfere with the treatment of the patient. Issues regarding the therapist's own sexual identity, past behaviors, and eventual death may also give rise to countertransference issues. For some psychotherapists who have practices with many HIV-infected patients, professional burnout can begin to inpair their effectiveness. Some studies have found that seeing many HIV-infected patients in a short period of time seems to be more stressful to therapists than seeing a smaller number of HIV-infected patients over a long period of time.

Involvement of Significant Others The patient's family, lover, and close friends are often important allies in treatment. The patient's spouse or lover may have guilt feelings about possibly having infected the patient or may experience anger at the patient for possibly infecting him or her. The involvement of members of the patient's support group can help the therapist assess the patient's cognitive function and can also aid in planning financial and living arrangements for the patient. The patient's significant others may themselves benefit from the attention of the therapist in helping them cope with the illness and the impending loss of a friend or family member.

Psychopharmacological Treatment The psychopharmacological treatment of patients with HIV infection is safe and effective, providing one keeps in mind the guidelines on dosing and drug interactions developed for the use of psychotropic drugs in the medically ill and the elderly. Psychotropic drugs have no known adverse effects on immunological functions.

INTERACTIONS OF PSYCHOTROPIC DRUGS AND ANTIRETROVIRAL DRUGS

Protease inhibitors are metabolized by the cytochrome p-450 oxidase system, primarily by the 3a 3/4 isoform, and secondarily by the 2D6 system. Obviously these pathways are shared by many psychotropic drugs. The clinical relevance of this potential for drug interactions has not been well studied. Drug manufacturers of the protease inhibitors have as a first step listed many psychotropic drugs as relatively or absolutely contraindicated in combination with protease inhibitors, which may be an overstatement. Because these presumptive drug interactions have not been systematically studied, it would be unfortunate to deny patients entire classes of potentially safe and beneficial psychotropic drug interventions. For now the most sensible policy is to be aware of the likely major drug-drug interactions, to monitor patients for treatment-emergent adverse effects and, wherever possible, for plasma psychotropic concentrations. The most likely interactions are as follows. All protease inhibitors will increase psychotropic drug concentrations if the major route of metabolism of the psychotropic agent is the 3 cytochrome P450 (CYP) CYP 3A system. If the CYP 2D6 system is the primary route of metabolism (e.g., tricyclic medications, SSRIs), ritonavir (Norvir) will specifically inhibit their metabolism. Thus the protease inhibitors may inhibit the metabolism of many antidepressants and antipsychotic agents as well as benzodiazepines. For example, plasma concentrations of alprazolam (Xanax), midazolam (Versed), triazolam (Halcion), and zolpidem (Ambien), may be increased and dosage reduction and careful monitoring may be required to prevent oversedation or other toxic effects. Protease inhibitors have been reported to increase concentrations of bupropion, nefazodone (Serzone), and fluoxetine (Prozac) to toxic levels, and to increase desipramine plasma concentrations by 100 to 150 percent. Drug interactions with antipsychotic agents are less well studied, but here ritonavir particularly may increase concentrations. Concentrations of methadone and meperidine are also reported to be elevated. Additionally, concentrations of some drugs of abuse such as methylenedioxymethamphetamine (MDMA) may be increased. In turn, protease inhibitors may induce the metabolism of valproate (Depakene) and of lorazepam (Ativan) and lead to lower plasma concentrations.

Some psychotropic medications may induce metabolism of protease inhibitors. Carbamazepine and phenobarbital may reduce serum concentrations of protease

inhibitors. The clinical relevance of this potential interaction is not clear, but use of an alternate mood stabilizer may be indicated. Interactions with lithium (Eskalith) and gabapentin (Neurontin) have not been reported.

Finally, psychotropic drugs may reduce the metabolism of some protease inhibitors, with an increase of protease inhibitor adverse effects; this has been reported with nefazodone and fluoxetine.

ANXIETY DISORDERS

Symptomatic anxiety that does not meet criteria for a formal diagnosis in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) is perhaps the most prevalent anxiety state. It may arise episodically or become a chronic feature of preoccupation with physical symptoms or illness. The decision to treat episodes of acute anxiety rests on whether it appears to interfere with daily functioning or quality of life. Restlessness, anxiousness, or inability to concentrate that impair one's usual social relationships or functioning for a week or more may indicate a need for psychological or pharmacological therapy. Nonpharmacological treatments may be tried first, including supportive psychotherapy or the use of stress and anxiety management approaches, such as relaxation training, education, and assisting in accessing various social services.

Psychopharmacological treatment is usually brief, lasting no more than 2 to 6 weeks, based on the view that most episodes of anxiety will remit within that time. For physically asymptomatic seropositive patients, the usual choice is an intermediate or long-acting benzodiazepine (e.g., diazepam [Valium] at 20 to 40 mg a day, or lorazepam at 0.5 mg every 8 hours). For individuals with more physically symptomatic disease, many clinicians recommend the use of agents with short to intermediate half-lives and no active metabolites, such as oxazepam (Serax) at 10 mg every 6 hours, or lorazepam at 0.5 mg every 8 hours.

In general, evidence of a therapeutic effect should be expected within 1 week of starting treatment. Failure to improve means that the diagnosis should be reconsidered and that major depressive disorder or substance-related disorder should be included in the differential diagnosis. Chronic anxiety states may be difficult to distinguish from hypochondriasis centered on HIV concerns; the latter tends to be treatment resistant and extremely difficult to manage even with combined psychotherapeutic and pharmacological approaches because obsessional or paranoid character structure may underlie symptomatology. In any event, a small proportion of patients may require maintenance therapy with anxiolytic agents.

For patients with recent histories of alcohol or nonalcohol substance-related disorders, the use of benzodiazepines may not be desirable. Buspirone (BuSpar) or hydroxyzine (Vistaril) are safe and effective, with low initial starting dosages and slow increases for individuals who are physically sicker. The delayed onset of therapeutic effect makes buspirone less suitable for cases of acute severe anxiety.

Occasionally obsessive-compulsive disorders are evident, with HIV infection or AIDS being the focus of rumination and ritualistic behavior. The differential diagnosis should include major depressive disorder. There are reports of such patients responding well to standard antidepressant therapy, such as imipramine (Tofranil) or desipramine (Norpramin). Serotonin reuptake inhibitors (e.g., fluoxetine, fluvoxamine [Luvox], sertraline [Zoloft], paroxetine [Paxil], or clomipramine [Anafranil]) are also effective; however, there is little experience with these agents in seropositive persons; however, fluoxetine was well tolerated in standard dosages in one group.

Similarly, panic disorder in physically asymptomatic or more symptomatic phases responds to standard therapy, which includes education, cognitive therapy, and pharmacological intervention (e.g., with desipramine). Starting doses of benzodiazepines or cyclic antidepressant agents in physically symptomatic persons are as low as one third of those used for physically healthy individuals, and doses should be increased cautiously.

MAJOR DEPRESSIVE DISORDER

Treatment of a major depressive episode in the physically asymptomatic patient with HIV disease is rather straightforward. Clinical experience and a limited number of investigational trials indicate that the usual therapeutic doses of standard agents are generally safe and effective in physically asymptomatic or moderately symptomatic persons, although there are almost no data on maintenance treatment and longterm follow-up. First-generation tricyclic agents with low anticholinergic and other adverse effects (e.g., desipramine), can be used. Most clinicians prefer selective serotonin reuptake inhibitors such as fluoxetine, sertraline, or paroxetine because of their more favorable adverse-effect profiles.

In more advanced HIV disease, starting dosages are usually about one fourth to one half of that in the customary psychiatric patient (e.g., fluoxetine at 5 to 10 mg daily or nortriptyline at 10 mg daily). The most prevalent adverse effects are the same as in non-HIV samples although it is noted that up to 25 percent of outpatients with frank AIDS will experience the severe adverse effects of tricyclics. HIV-infected persons are at increased risk of seizure disorder related to HIV or to opportunistic infection or neoplasm, and caution should be used with bupropion because seizures and confusional states can occur at daily dosage greater than 200 mg or if the individual doses are taken less than 6 hours apart.

Physically asymptomatic patients may show the best rates of response to medication, and there are reports of somewhat reduced efficacy in patients with frank AIDS, suggesting that overall health factors may influence response to therapy. That reduced efficacy in later-stage illness is consistent with reports of a response rate of around 40 percent in hospitalized, non-HIV medically ill patients with major depressive disorder. The possibility of more limited efficacy in later-stage disease has also been related to loss of social support and to the psychological burden of advancing physical illness.

Physically ill patients with major depressive disorder who are unable to tolerate or do not respond to cyclic antidepressant regimens may respond to psychostimulants, such as methylphenidate (Ritalin), dextroamphetamine (Dexedrine), or pemoline (Cylert). Another advantage of these agents is their rapid onset of action, measured in days as compared to weeks for conventional antidepressant agents. The usual starting dose for dextroamphetamine (Dexedrine) is 5 to 10 mg daily in a divided dose in the morning and early afternoon, with increases daily or every other day up to 20 to 30 mg a day. Some authorities prefer methylphenidate because dextroamphetamine may produce severe tremor or a persisting movement disorder in later-stage patients. Previous histories of stimulant or intravenous drug abuse, seizure disorder, opportunistic CNS disease, cardiac arrhythmias, or hypertension are considered relative contraindications to treatment with psychostimulants.

Intramuscular injections of testosterone cypionate (Depo–Testosterone), up to 400 mg biweekly for 8 weeks, have been shown in open label trials to be effective for major depressive disorder as well as for fatigue-anergia syndromes. Responses have been marked even in individuals with late-stage disease. Adverse effects have included irritability, hair loss, fatigue, tension, and breast tenderness, but in most cases these are reported to diminish without change in dosage.

Electroconvulsive therapy (ECT) is effective in patients whose depression requires urgent treatment, in those who have not responded to pharmacological therapies, or in those having major depressive disorder with psychotic features. With judicious use good outcomes may be expected even in persons with neurocognitive impairment, much in the same way that elderly patients with dementia and major depressive disorder respond. Neurological examination and brain imaging are necessary to rule out opportunistic brain disease and increased intracranial pressure before commencing ECT.

DELIRIUM, DEMENTIA, AND OTHER COGNITIVE DISORDERS

Delirium Symptoms of delirium in HIV illness can be effectively managed with low dosages of either low-potency antipsychotic agents, such as chlorpromazine at 10 to 25 mg once to three times daily, or with high-potency agents, such as haloperidol (Haldol) at 0.25 mg to 5 mg once to three times daily, or with atypical serotonin-dopamine agonists, including risperidone (Risperdal) at 0.5 mg to 2 mg daily, or olanzapine (Zyprexa) at 10 mg daily. There may well be an increased incidence of extrapyramidal symptoms associated with high-potency typical agents in advanced HIV illness, and patients with underlying HIV-associated dementia appear to be at highest risk for medication-induced movement disorders. For patients who do not respond to low-dosage oral therapy, excellent results have been reported with intravenous haloperidol given in individual boluses ranging from up to 2 to 10 mg every hour, or in extreme instances, up to 150 mg haloperidol daily. Some clinicians have also had good results with a combination of intravenous haloperidol and lorazepam, with an average daily intravenous dose of less than 50 mg of haloperidol and 10 mg of lorazepam. In general, no serious adverse effects have been noted with those more aggressive intravenous regimens, although nearly one half of the patients treated may have extrapyramidal symptoms.

Benzodiazepines alone (e.g., lorazepam) do not appear to be effective in delirious states, and they may accentuate confusion.

HIV-Associated Dementia and Mild Neurocognitive Disorder Treatment of neurocognitive impairment in HIV should ideally be both psychological and pharmacological. The patient should be informed of the strengths and the liabilities detected on neuropsychological testing and be advised of simple coping

measures, such as keeping to a routine, doing one thing at a time, and reducing external stimuli. The patient's safety is a priority, and friends, family, and service organizations should be enlisted to assist with caretaking.

Antiretroviral treatments are the drugs of choice for HIV-associated cognitive impairment. Although it is not established that antiretrovirals must penetrate well into the CNS in order to be effective, it is wise to try to include those known to cross the blood-brain barrier in such regimens. Of the currently available antiretrovirals the nucleoside reverse transcriptase inhibitors zidovudine and stavudine (Zerit) as well as the nonnucleoside reverse transcriptase inhibitor nevirapine (Viramune) show reasonable concentrations in CSF. In general, the reverse transcriptase drugs do not cross the blood-brain barrier.

Although some clinicians advocate the use of psychostimulants to improve attention, concentration, and psychomotor activity, most observations are based on data from uncontrolled studies. The one controlled study of methylphenidate did not show it to be superior to placebo. Clinicians who report positive experience with stimulants recommend methylphenidate at 2.5 to 5 mg twice a day, with breakfast and lunch, with slow increases up to 20 mg a day. Some have reported that good results can be maintained for many months, although doses may need to be increased. The relative psychiatric and medical contraindications for psychostimulant treatment apply, as discussed in reference to treatment of major depressive disorder.

Studies are currently in progress to determine if aggressive combination antiretroviral therapy can lower CSF viral load as well as improve cognitive functioning.

Fatigue-Anergia States Fatigue-anergia-apathy syndromes can also complicate advanced disease in its middle and later stages. These syndromes are not necessarily accompanied by significant depressive symptomatology or marked neurocognitive impairment. While release of various lymphokines (e.g., tumor necrosis factor, various interleukins) may be primarily responsible, some have suggested that there is a psychodynamic component—a conservation-withdrawal response in the face of loss of control. Beyond ensuring optimal nutrition and medical management, simple measures, such as encouragement and support, allowing a patient to participate in treatment decisions, and providing activities outside the hospital room, can be helpful, with a response usually occurring within 1 week. Psychostimulants at dosages used for patients with neurocognitive disorder (e.g., methylphenidate at 10 to 20 mg daily) have achieved gratifying results in other groups of patients, but there are no controlled studies available. Here too the usual contraindications for psychostimulant treatment should be considered. Testosterone cypionate 400 mg intramuscularly every 2 weeks has been shown to be effective in crossover studies, with increases in energy, motivation, and activity; adverse effects of irritability, tension, and hair loss are mild and transient.

OTHER CONDITIONS

Psychosis Symptoms of HIV psychosis respond well to antipsychotic medications. Although no comparative studies have yet been published comparing atypical and typical antipsychotic agents, these newer medications may become the drugs of first choice, based on adverse effects profiles (e.g., risperidone, olanzapine). Dosing for atypical antipsychotic drugs can be at one half (or less) the dosages used for acute psychoses in psychiatric patients.

For first-generation antipsychotic medications the usual therapeutic dose is one tenth to one third of that used for most acute psychoses in psychiatric settings, with the mean daily dose being about 150 mg in chlorpromazine equivalents (e.g., 1 to 10 mg daily of haloperidol or fluphenazine [Prolixin]). The latency of response is similar to that observed in the functional psychoses, with the majority of improvement said to occur within the first 6 weeks of therapy. It is unclear why lower-dose regimens are adequate, but that may be related to pharmacokinetic changes associated with chronic disease, or to HIV-associated damage to subcortical limbic or basal ganglia structures.

The latter phenomenon has also been thought to explain the sensitivity of patients with HIV-associated psychosis to extrapyramidal adverse effects. Movement disorders attributable to subcortical damage can appear in HIV-infected persons not taking antipsychotic medications, and they include myoclonus, paroxysmal dystonia, parkinsonian features, essential tremors, hemiballism, and hemichorea. In one study, thioridazine and haloperidol were equally effective, but thioridazine produced fewer intolerable extrapyramidal adverse effects. In general, these effects can be managed with low doses of benztropine (Cogentin), at 1 to 3 mg daily, although some authorities suggest substituting amantadine (Symmetrel) at 50 to 100 mg twice daily to avoid the anticholinergic properties of benztropine.

Neuroleptic Malignant Syndrome It is thought that AIDS patients generally and those with HIV-associated dementia in particular may be at increased risk for the most serious adverse effect of antipsychotic drugs, the neuroleptic malignant syndrome. This condition may present itself with subtle signs, and the diagnosis should be considered in any patient on antipsychotic drugs who develops rigidity, akinesia, mutism, lethargy, fever, autonomic instability, and increased creatine phosphokinase (CPK) activity. Neuroleptic malignant syndrome may be particularly associated with high-potency agents. Treatment is directed toward discontinuing the antipsychotic drug and instituting supportive care.

Manic States For immediate control of manic excitement up to 10 mg of clonazepam (Klonopin) daily is effective in many instances. If psychotic features are present, low doses of antipsychotic agents, such as risperidone at 0.5 to 2 mg daily, olanzapine up to 10 mg daily, chlorpromazine at 25 to 150 mg daily, or haloperidol at 0.5 to 5 mg daily, may be employed. For longer-term management lithium is effective but may not be as well tolerated as is carbamazepine and valproate (depakene). For example, in some studies lithium and antipsychotic medications are poorly tolerated by individuals with HIV-associated neurocognitive disorders, especially if brain MRI abnormalities are present (e.g., atrophy) whereas valproate (dosage range 750 to 1750 mg daily; plasma concentration >50 µg per mL) is fully successful. Good control is usually possible within 7 days, and treatment gains have been maintained for up to 4-year follow-up. Lithium has been used to treat patients who develop manic syndromes as an adverse effect of zidovudine, with good control of symptoms, which allows a patient to continue antiretroviral therapy. It may be that valproic acid and carbamazepine would be effective in these iatrogenic manias.

In HIV-infected patients treated with lithium for the control of bipolar I disorder care must be taken to monitor lithium concentrations closely, especially if the patient has significant gastrointestinal disturbances (e.g., vomiting and diarrhea) that may affect lithium absorption and excretion. Carbamazepine may reduce serum concentrations of protease inhibitors, and these agents themselves may lower valproate concentrations.

HIV and the Chronically Mentally Ill In metropolitan areas at the center of the AIDS epidemic, the prevalence of HIV seropositivity among patients admitted to general psychiatric hospitals may exceed 10 percent, and it is believed that HIV serotesting is underused by many physicians treating the chronically mentally ill. Individuals with schizophrenia, bipolar I disorder, or other chronic mental disorders are believed to be at increased risk for HIV acquisition and transmission, based on impulsiveness, difficulty in perceiving risk, and comorbidities, such as concurrent intravenous drug use and unsafe sexual practices. HIV education programs specifically tailored to the needs of patients with chronic mental disorders are currently being developed.

Legal Considerations: HIV Testing, Informed Consent, and Disclosure Written informed consent must be obtained from the person before testing for antibodies to HIV. Given an appropriate rationale, such as the need to provide the best possible care or to safeguard the health of others, permission is rarely denied. Although statutes vary in different jurisdictions, permission to test may be obtained from the patient's attorney-in-fact or closest available relative if the patient is incapable of giving informed consent, so long as the testing is aimed at offering the most appropriate care.

The need to protect potential sexual contacts of patients who are known to be seropositive or at high risk of HIV seropositivity may arise. Physicians and staff have the obligation to offer education to seropositive patients about acquisition and transmission of HIV and techniques to reduce transmission. In psychiatric inpatient settings the staff is obligated to protect the health of others by monitoring and, as necessary, restricting the privileges of patients known to be HIV-seropositive who are engaging or attempting to engage in high-risk (sexual) activity while hospitalized. It is generally agreed that the physician has the ethical right (but not the legal requirement) to disclose HIV seropositivity against the patient's wishes in order to safeguard the health of other selected persons. Spouses, sexual partners, and other high-risk contacts (e.g., those who share needles with the patient) may be informed, but in most jurisdictions the physician is not permitted to directly identify the seropositive person in question. In all cases the physician is obligated to tell the patient that his or her sexual contacts or other high-risk partners are being notified (Table 2.8-7 and Table 2.8-8).

1. Interpretation of test result:
   Clarify distortion (e.g., "'A negative test still means you could contract the virus at a future time; it does not mean you are immune from AIDS'").
   Ask questions about the patient's level of understanding and emotional reaction to the test result.
2. Recommendations for prevention of transmission (careful discussion of high-risk behaviors and guidelines for prevention of transmission).
3. Recommendations on the follow-up of sexual partners and needle contacts.
4. If test result is positive, recommendations against donating blood, sperm, or organs and against sharing razors and anything else that may have blood on it.
5. Referral for appropriate medical and psychological support. Activate supports available to patient (e.g., family, friends, community services).

Adapted from Rosse RB, Glese AA, Deutsch SI, Morihisa JM: *Laboratory and Diagnostic Testing in Psychiatry*. American Psychiatric Press, Washington, DC, 1989.

Table 2.8-7 Posttest HIV Counseling

Table 2.8-8 Physician Responsibilities Regarding Partner Notification

Worried Well A small group of persons, described as the "worried well," remain convinced that they have the symptoms of HIV infection despite repeated negative serological testings. Those patients present with anxiety-based physical symptoms (fatigue, diaphoresis, myalgia, sore throat, and perception of enlarged lymph nodes) and ruminate on images of HIV disease, on past risky sexual practices, and on their potential threat to current lovers or family. They may exhibit compulsive checking of their skin for lesions of Kaposi's sarcoma (such as counting moles or freckles), compulsive washing of items they believe are contaminated, such as clothes, and repeated questioning and physical scrutiny of partners for signs of HIV disease. Such individuals may travel from physician to physician seeking someone who supports their perceptions. The majority seem to qualify for a major psychiatric disorder, including major depressive disorder, delusional disorder, obsessive-compulsive disorder, or hypochondriasis. Treatment usually consists of a combination of appropriate psychopharmacological intervention for the primary psychiatric diagnosis, along with education about anxiety-based symptoms, exploration of distress related to life circumstances that might amplify physical symptoms, and self-monitoring of rituals.

Suicide Studies based on coroners' reports suggest that patients with advanced HIV disease have a thirtyfold risk of committing suicide compared to seronegative persons matched for age and social position. Some survey reports indicate that seronegative persons who are in a high-risk group for HIV infection, as well as seropositive persons at all stages of HIV infection, have an elevated lifetime prevalence of suicidal ideation and suicide attempt compared with community controls. It is important to note that both sources of data suggest that psychiatric disorder is strongly implicated in suicide, attempted suicide, and suicidal ideation. Psychological autopsies from coroners' cases have identified psychiatric histories in almost 50 percent of cases. Suicide attempt and suicidal ideation are correlated with histories of major depressive disorder or substance-related disorders, and in over half of cases these suicidal behaviors commenced before the likely date of seroconversion. Conflicts about sexual preference may be associated with suicide attempts by adolescents. This, together with the increase in HIV infection in adolescents, may place HIV-infected youths at particularly high risk. Suicide rates in women are not noted to be elevated, but the epidemic is now just starting to affect large numbers of women, and their greater vulnerability to major depressive disorder may mean that women are at risk. Advances in therapy may heighten hope and reduce the risk of suicide. However, those whose hopes are first raised but who then do not respond to or cannot tolerate these agents may require psychotherapeutic intervention. Thus, while debate over the distinction between suicide and the right to choose death or to refuse unwanted treatment are important issues, it is imperative that persons expressing suicidal behaviors or ideas be examined for a major psychiatric disorder and be offered appropriate treatment.

Bereavement Persons in social communities representing the high-risk groups for HIV infection are subject to repeated bereavement. Indeed, some surveys find that almost one of three homosexual men has suffered the loss of two or more members of his social network during the preceding 12 months, as compared to none of a heterosexual lower-risk comparison group. Furthermore, it has been noted that approximately 20 percent of these bereaved persons can be characterized as carrying unresolved grief (i.e., a persisting experience of grief for the deceased, difficulty with adjustment to the loss, and an inability to discuss the loss without distress). Persons with unresolved grief have been found to report more depressive and anxious symptoms than those with resolved grief (or those without loss), and they also demonstrate significantly higher lifetime histories of alcohol and other psychoactive substance-related disorders. In one study seropositive men were more likely than seronegative high-risk men to report unresolved grief, although there were no differences in grief resolution between physically symptomatic and physically asymptomatic seropositive persons. Given the association of grief with depression, and the possible psychological consequences of unresolved grief, clinicians should routinely inquire into their patients' history of bereavement and loss, especially because some may be experiencing multiple bereavements. The ultimate impact of bereavements on adaptation to illness, compliance with medical treatment, and adherence to safer sexual practices and health outcome is unknown.

Sleep Disorders Decreased sleep quality, difficulty falling asleep, fragmented nighttime sleep, and early-morning awakenings seem to increase as immune function and CD4+ lymphocyte counts diminish, and they may affect a substantial proportion of persons with AIDS. The reports of increased sleep difficulty in men with later-stage disease seem to be independent of zidovudine therapy or of psychological distress. Some have proposed that alterations in sleep architecture, such as increases in slow-wave sleep, may be an early marker of CNS involvement in HIV-infected persons. The therapeutic approach has generally been limited to symptomatic treatment with nonbenzodiazepine agents used as hypnotics, such as trazodone (Desyrel).

Pain Syndromes Pain probably remains the most underrecognized and undertreated symptom in HIV disease, as it is in other life-threatening illnesses, such as cancer. Undertreatment of pain is especially prevalent in injection drug users. The etiology and pathogenesis of pain in HIV-related disorders is just beginning to be understood. Psychiatric disorders may complicate persisting pain, and, because no one specialty addresses pain syndromes, patient care is often fragmented between anesthesiologists, neurologists, internists, and psychiatrists.

Neuropathic pain related to HIV usually presents as a persisting, painful sensorimotor neuropathy with dysesthesia, stocking-glove sensory loss, diminished distal reflexes, and distal weakness. Similarly postherpetic neuralgia (herpes zoster radiculitis) may involve pain of the face or trunk. Treatment of neuropathic pain syndromes is usually with low-dosage tricyclic antidepressant agents, such as desipramine or nortriptyline at 10 to 25 mg a day. The typical steady-state dosage is 50 mg a day, although some patients require higher amounts (75 to 100 mg daily). A response will often ensue within 1 to 2 weeks, but 4 to 6 weeks of treatment may be necessary before response occurs or another tricyclic agent is chosen. In general, tricyclic antidepressants probably are more effective than SSRIs for chronic neuropathic pain. Opioid analgesics are also effective. Anticonvulsants, such as phenytoin (Dilantin) or carbamazepine, at usual therapeutic concentrations required for seizure management, may also be effective. Postherpetic neuralgia may also be treated with topical capsaicin (Dolorac) and may also respond to clonazepam at 1 to 5 mg daily.

Chronic headache may appear as a residual symptom from acute aseptic meningitis in seroconversion illness or as an effect of zidovudine (which may persist after drug discontinuation). It is known that imipramine, desipramine, amitriptyline and nortriptyline can be effective in treating migraine and mixed migraine-tension headache syndromes in non-HIV populations, and there is speculation that persisting headache following aseptic meningitis may also respond to low-dosage regimens of these agents (e.g., nortriptyline or desipramine at 10 to 25 mg daily with increases up to 75 mg).

Among the rheumatological disorders are arthralgias, myalgias, and arthritides involving large joints of the leg. HIV-related arthralgias may respond to nonsteroidal

anti-inflammatory agents, although acetaminophen (Tylenol) should be avoided because it may diminish the metabolism of zidovudine. HIV may also be associated with a polymyositis, which involves pain, weakness, and elevated CPK, along with changes on electromyography indicating a myopathic process. Long-term administration of zidovudine may also produce a myositis that persists when the medication is discontinued. Psychopharmacological interventions are not of demonstrated efficacy in these states.

Multidisciplinary pain treatment approaches, which employ coordinated efforts of experts in various disciplines, may be as useful in chronic HIV-related pain as they are in chronic, non-HIV pain syndromes. Such approaches employ education about the nature of persisting pain, activity scheduling, self-monitoring and relaxation training, and cognitive therapies to reduce disability related to pain.

Finally, studies of acute postoperative pain and chronic cancer pain generally indicate that for those conditions in which opiate analgesia is indicated, those medications are often underprescribed or irrationally prescribed, in subtherapeutic doses at too extended an interval. The clinician should always be alert to that possibility in advanced HIV disease.

SPECIAL ISSUES

Children and Adolescents In approximately 30 percent of cases, infants born to untreated seropositive mothers will themselves seroconvert within the first 18 months of life, presumably from infection in utero by vertical transmission from the mother; there are some indications that breast feeding is also a risk factor. Antiretroviral treatment during pregnancy dramatically reduces the risk of transmission.

Few data are available describing the neuropsychological or clinical psychiatric findings in children with asymptomatic or mildly symptomatic infection. The predominant feature of AIDS in children consists of a developmental delay evident in infancy and progressive cognitive impairment in older children. The majority of physically symptomatic HIV-positive children evidence a developmental delay, and almost all of those are reported to have CNS dysfunction. Developmental delays are usually most evident in fine and gross motor skills and speech, and school-aged children with frank AIDS may have intellectual abilities in the mildly to severely retarded range.

The course of illness is variable, ranging from static and persisting deficits to rapid decline, with many children showing a steadily deteriorating course. The group of children most at risk for HIV infection is presently the same group most at risk for developmental delay from environmental factors, such as poor nutrition, prolonged hospitalization, and understimulation. Added to those factors must be the impact of the mother's HIV illness on caretaking.

Standard treatment for developmental delay is promptly to institute a program of rehabilitation and enrichment. All seropositive infants and children should be considered for treatment. The verbal and play psychotherapies developed for children with developmental delays and mental retardation from non-HIV sources are also thought to be applicable for children infected with HIV.

Adolescents are one of the fastest-growing risk groups, and runaways, homosexual and bisexual youths, drug-using adolescents, and sexually abused adolescents are thought to be at particularly high risk. Most of the major psychiatric syndromes reported in adults with HIV have been described in adolescents, including major depressive disorder, delirium, and anxiety and psychotic syndromes. There is intense current interest in the efficacy of educational interventions for promoting behavior change and safer sexual practices among adolescents, using peer counseling, mass media, family, and school-based interventions. The longer-term efficacy of those programs is being evaluated.

HIV infection touches every member of the family. One partner's fear of discussing his or her seropositive status with a spouse out of fear of rage and abandonment may lead to attempts to isolate the knowledge and disguise the illness. Parents may be reluctant to disclose their seropositivity or their child's serostatus to members of the extended family because of concerns over rejection and condemnation. Individual members of the family or the entire family may deny the illness of the mother or child. A parent or other family member may express guilt as anger toward the seropositive child, or the child may become the displaced focus for other family conflicts. Conjoint and family therapy interventions are often indicated, but therapy must frequently be directed toward supporting the children's primary caretaker.

Women and HIV Women, especially those of minority ethnicity, are a fast-growing risk group for HIV. Most women infected with HIV are injection drug users or partners of drug users who have become infected by heterosexual intercourse. Studies from partners of injection drug users indicate high concurrent rates of substance-related disorders, major depressive disorder (up to 20 to 25 percent), and severe personality disorders.

The cross-sectional psychiatric findings in HIV-seropositive women may depend greatly on risk group or mode of acquisition. Prospective studies are urgently needed to describe the physical, neuropsychological, and psychological course of HIV infection in women.

In earlier phases of the epidemic physicians frequently suggested that pregnancy be avoided or terminated because of its immunosuppressant effects, and the likelihood of maternal fetal transmission. However, advent of potent new therapy opens the possibility of a safer pregnancy. As of late 1998, one of the most prevalent dilemmas involves pregnancy in the severely immunocompromised woman with CD4 lymphocyte count <200. Because pregnancy may further reduce immune capacity there is enhanced risk of serious opportunistic infections as well as danger to the developing fetus from the aggressive treatment of such infections. Women who are not severely immunocompromised (e.g., CD4 lymphocyte count <200) may face another dilemma: if the woman's partner is also seropositive she risks being infected by novel strains of HIV. Because recommendations in this arena are evolving so rapidly, specialized referral and family planning counseling are usually needed.

SUGGESTED CROSS-REFERENCES

Some of the specific syndromes associated with HIV infection are discussed in Chapter 10 on delirium, dementia, and other cognitive disorders; in Chapter 11 on substance-related disorders; in Chapter 12 on schizophrenia; in Chapter 13 on other psychotic disorders; in Chapter 14 on mood disorders; in Chapter 15 on anxiety disorders; and in Chapter 19 on human sexuality. Treatment of specific disorders is reviewed in Chapter 30 on psychotherapies and in Chapter 31 on biological therapies. Detailed information on neuropsychological assessment is provided in Chapter 7 on diagnosis and psychiatry, specifically in Section 7.4 on neuropsychological and intellectual assessment of adults, and Section 7.5 on neuropsychological and intellectual assessment of children. Additional topics in neuropsychiatry are treated in Chapter 2 on neuropsychiatry and behavior neurology. Discussion of neuroimaging is provided in Section 1.15 and Section 1.16 on neuroimaging in clinical practice. Detailed discussion of life adversity and immunity are treated in Section 25.9 on stress and psychiatry and Section 25.10 on behavior and immunity.

SECTION REFERENCES

*Atkinson JJ, Grant I, Kennedy CJ, Richman DD, Spector SA, McCutchan JA: Prevalence of psychiatric disorders among men infected with human immunodeficiency virus. Arch Gen Psychiatry *45:*859, 1988.

Breitbart W, Rosenfeld B, Passik S, Kaim M, Funesti-Esch J, Stein K: A comparison of pain report and adequacy of analgesic therapy in ambulatory AIDS patients with and without a history of substance abuse. Pain *72:*235, 1997.

*Carpenter CCJ, Fischi MA, Hammer SM, Hirsch MS, Jacobsen DM, Katzenstein DA, Montaner JSG, Richman DD, Saag MS, Schooley RT, Thompson MA, Vella S, Yeni PG, Volberding PA: Antiretroviral therapy for HIV infection in 1997. JAMA *277:*1962, 1997.

Centers For Disease Control: 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. MMWR *41:*1, 1992.

Deeks SG, Smith M, Holodniy M, Kahn JO: HIV-1 protease inhibitors: A review for clinicians. JAMA *277:*145, 1997.

*Dickey WC, Dew MA, Becker JT, Kingsley L: Combined effects of HIV-infection status and psychosocial vulnerability on mental health in homosexual men. Soc Psychiatry Psychiatr Epidemiol *34:*4, 1999.

*Ellis RJ, Deutsch R, Heaton RK, Marcotte TD, McCutchan JA, Nelson JA, Abramson I, Thal LJ, Atkinson JH, Wallace MR, Grant I, and the HIV Neurobehavioral Research Center Group: Neurocognitive impairment is an independent risk factor for death in HIV infection. Arch Neurol *54:*416, 1997.

Ellis RJ, Hsia K, Spector SA, Nelson JA, Heaton RK, Wallace MR, Abramson I, Atkinson JH, Grant I, McCutchan JA, and the HIV Neurobehavioral Research Center Group: Cerebrospinal fluid human

immunodeficiency virus Type I RNA levels are elevated in neurocognitively impaired individuals with acquired immunodeficiency syndrome. Ann Neurol *42:*679, 1997.

Gendelman HD, Lipton S, Epstein L, Swindells S, editors: *Neurology of AIDS.* Chapman & Hall Publishers, New York, 1998.

Grant I, Atkinson JH, Hesselink JR, Kennedy CJ, Richman DD, Spector SA, McCutchan JA: Evidence for early central nervous system involvement in the acquired immunodeficiency syndrome (AIDS) and other human immunodeficiency virus (HIV) infections. Studies with neuropsychologic testing and magnetic resonance imaging. Ann Intern Med *107:*828, 1987.

*Grant I, Martin AM, editors: *Neuropsychology of HIV Infection.* Oxford University Press, New York, 1994.

Grant I, Olshen RA, Atkinson JH, Heaton RK, Nelson J, McCutchan JA, Weinrich JA: Depressed mood does not explain neuropsychological deficits in HIV infected persons. Neuropsychology *7:*53, 1993.

Heaton RK, Grant I, Butters N, White DA, Kirson D, Atkinson JH, McCutchan JA, Taylor MJ, Kelly MD, Ellis RJ, Wolfson T, Velin R, Marcotte TD, Hesselink JR, Jernigan TL, Chandler J, Wallace M, Abramson I, and the HNRC Group: The HNRC 500—Neuropsychology of HIV infection at different disease stages. J Int Neuropsychol Soc *1:*231, 1995.

Heyward WL, MacQueen KM, Jaffe HW: Obstacles and progress toward development of a preventive HIV vaccine. J Int Assoc Physicians in AIDS Care *August:*28, 1997.

Holmes V, Fernandez F, Levy JK: Psychostimulant response in AIDS-related complex (ARC) patients. J Clin Psychiatry *50:*5, 1989.

Katz MH, Gerberding JL: The care of persons with recent sexual exposure to HIV. Ann Intern Med *128:*306, 1998.

Kelly JA, McAuliffe TL, Sikkema KJ, Murphy DA, Somlai AM, Mulry G, Miller JG, Stevenson LY, Fernandez MI: Reduction in risk behavior among adults with severe mental illness who learned to advocate for HIV prevention. Psych Services *48:*1283, 1997.

Kelly JA, Murphy DA, Sikkema KJ, McAuliffe TL, Roffman RA, Solomon LJ, Winette RA, Kalichman SC, and the Community HIV Prevention Research Collaborative: Randomized, controlled, community-level HIV-prevention intervention for sexual risk behaviour among homosexual men in US cities. Lancet *350:*1500, 1997.

Lipton SA, Gendelman HE: Dementia associated with the acquired immuno-deficiency syndrome. N Engl J Med *332:*934, 1995.

Lobovits AH, McCarthy D, Simon R: The prevalence and management of pain in patients with AIDS: A review of 134 cases. Clin J Pain *5:*245, 1989.

Lyketsos CG, Hoover DR, Guccione M, Dew MA, Wesch JE, Bing EG, Treisman GJ: Changes in depressive symptoms as AIDS develops. Am J Psychiatry *153:*1430, 1996.

Lyketsos CG, Schwartz J, Fishman M, Treisman G: AIDS mania. J Neuropsychiatry Clin Neurosci *9:*277, 1997.

Maj M, Janssen R, Starace F, Zaudig M, Satz P, Sughondhabirom B, Luabeya MA, Riedel R, Ndetei D, Calil HM, Bing EG, St Louis M, Sartorius N: WHO Neuropsychiatric AIDS study, cross-sectional phase I. Study design and psychiatric findings. Arch Gen Psychiatry *51:*39, 1994.

Marzuk PM, Tierney H, Tardiff K, Gross EM, Morgan EB, Hsu MA, Mann JJ: Increased risk of suicide in persons with AIDS. JAMA *259:*1333, 1988.

Masliah E, Heaton RK, Marcotte TD, Ellis RJ, Wiley CA, Mallory M, Achim CL, McCutchan JA, Nelson JA, Atkinson JH, Grant I, and the HNRC Group: Dendritic injury is a pathological substrate for human immunodeficiency virus-related cognitive disorders. Ann Neurol *42:*963, 1997.

Mayne TJ, Vittinghoff E, Chesney MA, Barrett DC, Coates TJ: Depressive affect and survival among gay and bisexual men infected with HIV. Arch Intern Med *156:*2233, 1996.

McArthur JC, Grant I: HIV neurocognitive disorders. In *Neurology of AIDS*, HE Gendleman, S Lipton, L Epstein, S Swindells, editors. Chapman and Hall, New York 1997.

McArthur JC, Hoover DR, Bacellar H, Miller EN, Cohen BA, Becker JT, Graham NM, McArthur JH, Selnes OA, Jacobson LP, Visscher DR, Concha M, Saah A: Dementia in AIDS patients: Incidence and risk factors. Multicenter AIDS Cohort Study. Neurology *43:*2245, 1993.

Miller EN, Selnes OA, McArthur JC, Satz P, Becker JT, Cohen BA, Sheridan K, Machado AM, Van Gorp WG, Visscher B: Neuropsychological performance in HIV-1-infected homosexual men: The Multicenter AIDS Cohort Study (MACS). Neurology *40:*197, 1990.

Navia BA, Jordan BD, Price RW: The AIDS dementia complex: I. Clinical features. Ann Neurol *19:*517, 1986.

Perkins DO, Leserman J, Stern RA, Baum SF, Liao D, Golden RN, Evans DL: Somatic symptoms and HIV infection: Relationship to depressive symptoms and indicators of HIV disease. Am J Psychiatry *152:*1776, 1995.

Perry S, Jacobsberg L, Fishman B: Suicidal ideation and HIV testing. JAMA *263:*5, 679, 1990.

Rabkin JG, Ferrando S: A “second life” agenda: Psychiatric research issues raised by protease inhibitor treatments for people with the human immunodeficiency virus or the acquired immunodeficiency syndrome. Arch Gen Psychiatry *54:*1049, 1997.

Rabkin JG, Johnson J, Lin SH, Lipsitz JD, Remien RH, Williams JBW, Gorman JM: Psychopathology in male and female HIV-positive and negative injecting drug users: Longitudinal course over 3 years. AIDS *11:*1, 1997.

Sewell DD, Jeste DV, Atkinson JH, Heaton RK, Hesselink JR, Wiley C, Thal L, Chandler JL, Grant I, the HNRC Group: HIV-associated psychosis: A study of 20 cases. Am J Psychiatry *151:*237, 1994.

*Spiegel L, Mayers A: Psychosocial aspects of AIDS in children and adolescents. Pediatr Clin North Am *38:*153, 1991.

Stout JC, Ellis RJ, Jernigan TL, Archibald SL, Abramson I, Wolfson T, McCutchan JA, Wallace MR, Atkinson JH, Grant I, and the HIV Neurobehavioral Research Center Group: Progressive cerebral volume loss in human immunodeficiency virus infection: A longitudinal volumetric MRI study. Arch Neurol *55:*161, 1998.

Van Dyck CH, McMahon TJ, Rosen MI, O'Malley SS, O'Connor PG, Lin CH, Pearsall HR, Woods SW, Kosten TR: Sustained-release methylphenidate for cognitive impairment in HIV-1-infected drug abusers: A pilot study. J Neurol *9:*29, 1997.

White DA, Heaton RK, Monsch AU, and the HNRC Group: Neuropsychological studies of asymptomatic human immunodeficiency virus type-1 infected individuals. J Intern Neuropsychol Soc *1:*304, 1995.

Working Group of the American Academy Of Neurology Aids Task Force: Nomenclature and research case definitions for neurologic manifestations of human immunodeficiency virus-type 1 (HIV-1) infection. Neurology *41:*778, 1991.

2.9 NEUROPSYCHIATRIC ASPECTS OF OTHER INFECTIOUS DISEASES

BRIAN ANTHONY FALLON, M.D., M.P.H.

Ever since the link between severe neuropsychiatric disorders and infectious disease was established in the early 1900s by the identification of the cause of syphilis, questions have been asked about the role of other infectious organisms in the etiology of neuropsychiatric disorders. At times the link between an infectious agent and a neuropsychiatric disorder is obvious, as in the case of neurosyphilis, the viral influenza outbreak of the 1920s, and the current human immunodeficiency virus (HIV) and Lyme disease epidemics. At other times the link is less clear but strongly suspected, as has been true for chronic fatigue syndrome or in the search for bacterial or viral etiologies of obsessive-compulsive and psychotic disorders. Psychopathology may emerge as a result of direct invasion of the central nervous system (CNS) by neurotropic agents or by an indirect host-determined cellular, humoral, or cytokine immune response to infectious organisms that inadvertently damages host tissue. In its effort to protect, the immune response may thereby provoke neuropsychiatric disorders.

This section focuses on selected infectious diseases other than HIV disease that invade the CNS and that have been directly associated with neuropsychiatric syndromes. Particular attention will be paid to the neuropsychiatric aspects of Lyme disease because it has spread rapidly since the 1970s in various parts of the world and has been associated with a plethora of neuropsychological and neurobehavioral problems in both children and adults. In addition, the concluding portion of this section will briefly address a few areas of recent investigation on the overlap of infectious disease and neuropsychiatry.

SPIROCHETAL DISEASES

Under the umbrella of the order of spirochetes are three agents that are known to invade the CNS. These include borrelia, treponema, and leptospira. Borrelia, which require an arthropod vector and a mammalian or bird reservoir, are commonly known to cause relapsing fever and Lyme disease. Treponema, which are spread person to person and do not use an arthropod vector, are the spirochetes responsible for syphilis. Leptospira, which are spread by contaminated water, are the agents of Weil's disease, which can have CNS manifestations.

Lyme Disease Lyme disease (Lyme borreliosis), transmitted by the bite of an infected Ixodes tick, can cause a vast array of neuropsychiatric disorders, ranging from mild mood changes to psychosis and severe memory loss. Lyme disease has been reported throughout the United States and in many countries throughout the world. The causative agent of Lyme disease, *Borrelia burgdorferi*, is initially inoculated into the skin by an infected tick, typically inducing a local rash, known as *erythema migrans*, which is reported by approximately two thirds of infected patients. Rapidly disseminated by the bloodstream through the body, *B. burgdorferi* has been found in the CNS as soon as 3 weeks after initial skin infection. Known to be neurotropic, *B. burgdorferi* may reside in the cerebrospinal fluid (CSF) or adhere to glial cells or other brain tissue. Like its spirochetal counterpart, *Treponema pallidum, B. burgdorferi* may remain latent, causing illness months to years later. Partly because of this latency in disease expression, patients may be unable to recall the initial tick bite or rash. Antigenic variability, which refers to the ability to express different surface antigens and to thus evade the immune response, is a feature of borrelial organisms that *B. burgdorferi* shares.

Diagnosis The epidemiological surveillance criteria for the diagnosis of Lyme disease in the United States require a history of exposure to an area endemic for Lyme disease and either a physician-diagnosed erythema migrans rash or serological evidence of exposure to *B. burgdorferi* and at least one of the following three clinical features: (1) arthritis; (2) neurologic symptoms (cranial or peripheral neuropathy, meningitis, encephalomyelitis, or encephalitis with evidence of intrathecal antibody production); or (3) cardiac conduction defects. Although useful for epidemiological monitoring, these criteria are unduly restrictive and should not be used for clinical purposes, because these criteria exclude patients who might have Lyme disease, such as seropositive patients who have diffuse arthralgias but not frank arthritis or patients who have encephalopathy without objective CSF abnormalities. Further complicating the diagnosis is the unreliability of the serologic tests. False-positive results might result because of cross-reactivity with other spirochetal organisms. False-negative results may occur because the patient is tested too soon after infection and before an appropriate antibody response is mounted or because the patient's immune response has been abrogated as can occur when a patient is given antibiotic shortly after initial infection. It is not uncommon for a patient with Lyme disease to have negative or equivocal test results in one laboratory, but positive ones in another or for a patient to have negative test results initially but positive ones several months later after antibiotic treatment has been initiated. For these reasons a rational approach to the diagnosis of Lyme disease must be based upon the primary clinical presentation, followed by the supportive evidence of laboratory test results. Laboratory tests that can be helpful include indirect tests such as the enzyme-linked immunosorbent assay (ELISA) and Western blot analysis and direct tests such as the polymerase chain reaction (PCR) assay for borrelial deoxyribonucleic acid (DNA) or antigen detection assay. When Lyme disease is suspected, the clinician should order both an ELISA and a Western blot, as some patients may have a negative ELISA result but a positive Western blot result. Bands of particular significance on the Western blot include the ones identified by the Centers for Disease Control as being most frequent and specific, as well as the 31kD (OspA) and 34 kD (OspB) bands. Although highly specific for *B. burgdorferi* DNA, the PCR assay has low sensitivity. Although laboratory testing is a valuable component of the diagnostic assessment, negative test results cannot be used to exclude Lyme disease in a patient with typical clinical features and a history of exposure to a Lyme disease endemic area.

Clinical Features The erythema migrans rash is the hallmark feature of early Lyme disease; antibiotic treatment at this stage often results in cure. Although typically the rash has a bull's eye, rounded appearance, it may also have a triangular, elongated, or other shape. Because patients may not recall seeing the rash, the flu-like symptoms that often occur shortly after the rash may be ignored, only to be followed several months to years later by the emergence of a multisystem disease affecting the joints, the heart, the eyes, and the peripheral or central nervous system; 15 to 40 percent of patients may have neurologic signs as their presenting feature. Headaches may be followed by meningitis, cranial neuritis, motor or sensory radiculitis, or an encephalitis characterized by mood lability and disturbances of memory or sleep. Although suggestive of Lyme disease, Bell's palsy may occur in only 5 to 10 percent of a sample of patients with neurologic Lyme disease. Symptoms of radiculoneuropathy or peripheral nerve involvement include sharp stabbing or deep boring pains that may radiate from the spine into an extremity of the trunk; areas of numbness, burning, or tingling; weakness; and fasciculations. In later stages of Lyme disease a minority of patients may develop a chronic meningoencephalomyelitis characterized by somnolence, confusion, poor concentration, impaired memory, myoclonus, apraxia, ataxia, paraparesis, dysarthria, dysphasia, seizures, or bladder abnormalities. Some of these patients may be misdiagnosed as having multiple sclerosis because of a relapsing and remitting course and the concurrence of spinal motor signs, ataxia, bladder dysfunction, and, less often, optic neuritis.

The profile of neuropsychiatric Lyme disease typically includes disturbances of cognition and mood. On formal neuropsychological testing, more than 50 percent of patients with chronic neurologic Lyme disease will show impairment in short-term memory, processing speed, or attention. This cognitive impairment, although worsened by marked pain or mood disorders, exists independently of the physical symptoms or the severity of concurrent depression. Typical cognitive symptoms include word-finding problems, word-substitutions, new-onset dyslexia, transient episodes of geographic disorientation, marked inattention and distractibility, difficulty with organization, and the sensation that one's brain is in a fog. Less commonly, the severity of the cognitive disturbance causes a global impairment, suggestive of a new-onset dementia.

Although the full spectrum of psychiatric disorders has been associated with *B. burgdorferi* infection, by far the most frequent are disturbances of mood, characterized by irritability, mood swings, and sleep loss. The majority of controlled studies in which patients with Lyme disease are compared to healthy controls or to patients with other illnesses reveal that depression occurs more frequently in the group with Lyme disease. Children with neurological Lyme disease typically present with complaints of headaches as the most common symptom, followed by behavioral, cognitive, or mood disturbance as the next most prevalent symptom. Behavioral problems include falling asleep in class, agitation, and poor school performance; common cognitive problems include attentional and short-term memory and visospatial deficits; common mood problems include irritability and new-onset anxiety. Other less common neuropsychiatric aspects associated with Lyme disease in adults and children include panic attacks, transient paranoia, illusions or hallucinations (visual, olfactory, auditory), anorexia, depersonalization, violent outbursts, obsessive-compulsive disorder, agitated mania, hyperacute sensitivity to light or sound, and what appears to be personality change. Because of the multisystem involvement in Lyme disease and the frequent concurrence of anxiety and depression, patients may be mistakenly diagnosed as having a primary psychiatric or a

somatoform disorder before Lyme disease is even considered. If Lyme disease is considered but serological tests are equivocal despite the presence of a clinical profile typical of Lyme disease, the somatoform label may once again be mistakenly applied.

A 22-year-old previously healthy college graduate in his first few months of law school developed joint and muscle pains. A medical workup was negative, but the symptoms persisted and worsened, accompanied by fatigue so severe that he was unable to go to class or to study. Referral to a psychiatrist resulted in treatment with a selective serotonin reuptake inhibitor (SSRI) for possible depression without much benefit. Further medical workup revealed a positive result for Lyme disease ELISA with an equivocal Western blot analysis result. The diagnosis of probable Lyme disease was made and the patient was given a 6-week course of oral antibiotics, with a marked improvement in symptoms. Over the following 4 months his prior symptoms returned accompanied by headaches, word-finding problems, paresthesias, shooting and stabbing pains, and hypersensitivity to light and sound. Several consulting doctors gave conflicting opinions, with some firmly stating that this could not be persistent Lyme disease as it had been adequately treated and others stating that persistent infection was indeed possible and that additional treatment with antibiotics was warranted. The Lyme tests remained equivocal and a brain magnetic resonance imaging (MRI) revealed no abnormalities. Three additional months of oral antibiotics resulted in some improvement of arthritic symptoms, but the fatigue and cognitive problems remained. Several months later, the patient developed paranoid delusions followed by a manic episode for which he was hospitalized. Without further testing such as a spinal tap or electroencephalogram (EEG), his doctors dismissed Lyme disease as a possible cause for his new-onset mania. The patient was discharged on antipsychotic agents and lithium, with only a partial improvement in his mood lability. An outpatient internist then checked the patient's spinal fluid, which showed white blood count of 7 × 109 per liter 7 (white blood count [WBC]) and evidence of *B. burgdorferi* antibodies in both his serum and cerebral spinal fluid (CSF). The diagnosis of neuroborreliosis was made and the patient was placed on a 3-month course of intravenous antibiotics. Although the mood lability, mania, and cognitive problems dramatically resolved with long-term antibiotic treatment, 3 years later the patient had not yet returned to school because of problems with persistent fatigue. At that point, because a PCR assay of his plasma was positive for Lyme disease and a brain single photon emission computed tomography (SPECT) scan revealed global heterogeneous hypoperfusion, the patient was once again treated with antibiotics.

This case highlights the diagnostic difficulties of Lyme disease: the confusion triggered by equivocal results on serological tests, the need to consider *B. burgdorferi* as the cause for new-onset mania, the inadequate response to standard psychiatric medications, the initially robust but subsequently partial response to antibiotic regimens, and the persistence of the DNA of the organism despite aggressive antibiotic therapy.

Tests for CNS Lyme Disease Examination of the CSF is crucial to rule out other possible causes of CNS disease and to identify the presence of Lyme meningitis or encephalitis. In early neurologic Lyme disease, a spinal tap may reveal lymphocytic pleocytosis, mildly increased protein, and, in some cases, an elevated immunoglobulin G (IgG) index or the presence of oligoclonal bands. In later-stage neurological Lyme disease, however, the CSF may appear normal. MRI studies may reveal punctate white matter lesions on T2-weighted images, suggestive of a demyelinating disorder such as multiple sclerosis. EEG studies are generally normal, although diffuse slowing or epileptiform discharges may be seen. SPECT and positron emission tomography (PET) studies may be particularly helpful in late-stage Lyme disease. Recent reports indicate that many patients with Lyme encephalopathy have a pattern of either global or heterogeneous hypoperfusion, which in some cases improves after antibiotic treatment (Fig. 2.9-1). Given the difficulties facing the clinician attempting to determine whether the fatigue, mood lability, and cognitive tracking problems are caused by primary depression or by an underlying systemic disease, functional imaging studies are a valuable tool to assist in the differential diagnosis.

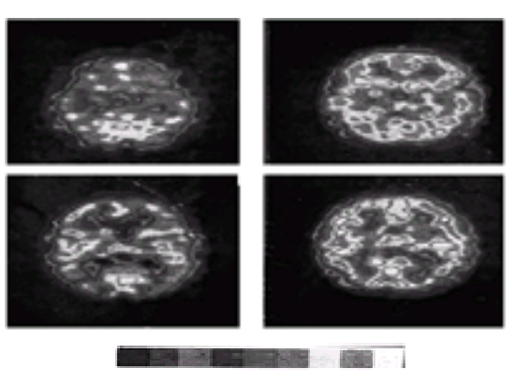

FIGURE 2.9-1 Transverse views obtained with Technetium-99m HMPAO SPECT. The two views on the left are of the brain of an adolescent with Lyme encephalopathy and demonstrate moderate heterogeneous hypoperfusion. The views on the right are of the brain of an adolescent without encephalopathy and demonstrate a normally perfused scan. The color spectrum scale, from purple to white, represents low to normal perfusion. (See color Plate 2.)

Differential Diagnosis In considering the diagnosis of Lyme disease, it is important to ask the patient about exposure to a Lyme disease-endemic area, history of a tick bite or unusual rash, and the presence of multisystemic involvement. Called the “new great imitator” (after the original great imitator, syphilis), the broad spectrum of atypical neurological manifestations of Lyme disease include strokes, Guillain-Barré syndrome, cerebellar syndromes, seizures, pseudotumor-like syndrome in children, spastic paraparesis, multiple–sclerosis–like illnesses, and progressive dementias. Similarly, other diseases that may look like neuropsychiatric Lyme disease need to be excluded, such as major depression with somatic preoccupation, panic disorder, systemic lupus erythematosus or other connective tissue diseases, chronic fatigue syndrome, endocrinological disorders, vitamin deficiencies, other infectious illnesses, vascular dementias, and other neurodegenerative disorders.

Treatment For early Lyme disease without CNS involvement, 3 to 4 weeks of oral doxycycline (Vibramycin) (100 mg twice a day), amoxicillin (Amoxil) (500 mg three times a day), or cefuroxime (Ceftin) (500 mg twice a day) is recommended. For Lyme disease with CNS involvement, an initial course of 4 to 6 weeks of intravenous ceftriaxone (Rocephin) (2 grams/day) or cefotaxime (Claforan) (2 grams/8 hours) is recommended. Symptoms may worsen during the first week of antibiotic treatment, much like the Jarisch-Herxheimer reaction during the treatment of syphilis. For patients who relapse, longer and repeated courses of antibiotics are often helpful. Failure to treat Lyme disease early in its course or for a sufficiently long duration may lead to a chronic illness characterized by persistent waxing and waning neuropsychiatric disturbances, arthralgias, myalgias, sensory-hyperacuities, and severe fatigue. In some patients these symptoms reflect the effects of persistent infection while in others the symptoms may reflect a residual postinfectious syndrome. Because the laboratory tests for chronic Lyme disease are not sufficiently reliable to document the presence or absence of persistent infection, decisions regarding treatment should be based primarily upon the physician's clinical judgment. Given the emerging literature that indicates that *B. burgdorferi* is capable of remarkable persistence in the human host despite standard courses of antibiotic treatment and clinical reports documenting improvement in chronic Lyme disease among some patients treated with long courses of antibiotics, many community physicians are now willing to treat such patients more aggressively. Other physicians, wary of the risks associated with long-term antibiotic treatment, choose not to treat these patients. Until well-controlled studies are conducted of patients with chronic symptoms, the debate on the proper antibiotic treatment for chronic Lyme disease will continue. A vaccine for Lyme disease was introduced in 1999; however, it is only effective in about 50 to 75 percent of cases.

Neurosyphilis The cause of syphilis, *Treponema pallidum*, was identified in 1905. Because of the cognitive loss and neuropsychiatric disturbances associated with tertiary neurosyphilis, these patients accounted for 5 to 15 percent of psychiatric hospital admissions and were diagnosed as having general paresis, general paralysis of the insane, or dementia paralytica. With penicillin treatment of primary and secondary syphilis, neurosyphilis is now an uncommon cause of hospital admissions.

Primary syphilis is manifest by a syphilitic ulcer, the chancre, at the site of inoculation. Secondary syphilis, a result of hematogenous dissemination of the spirochete, is characterized by flu-like symptoms followed by a skin rash, generalized lymphadenopathy, and mucosal lesions. Left untreated, both primary and secondary syphilis resolve on their own, after which the patient enters a latent period during which infection is present but clinical symptoms are not manifest. After months to years, about one third of patients with untreated latent syphilis develop tertiary syphilis that affects the brain or heart. As in neuroborreliosis, invasion of the CNS by *Treponema pallidum* occurs early in the disease and may be asymptomatic for months to years prior to clinical expression.

Clinical neurosyphilis can be divided into four types: syphilitic meningitis, meningovascular syphilis, parenchymatous neurosyphilis, and gummatous neurosyphilis. Syphilitic meningitis, the result of direct meningeal inflammation, rarely has focal findings. Meningovascular syphilis results from the ischemic changes caused by proliferative endarteritis, resulting in permanent CNS damage. In parenchymatous neurosyphilis (general paresis or tabes dorsalis), which generally starts 10 to 20 years after infection, there is direct neural destruction resulting in diminished neuron concentration, demyelination, and gliosis (Fig. 2.9-2). In gummatous

neurosyphilis, the mass effect causes neurological symptoms.

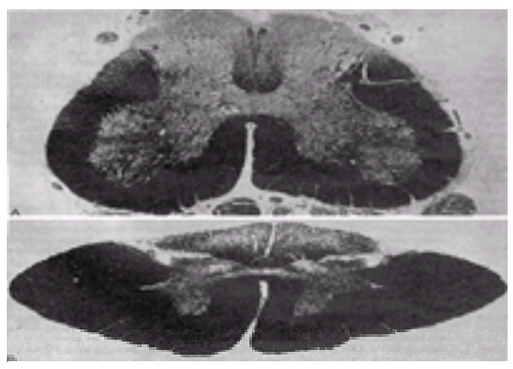

FIGURE 2.9-2 Tabes dorsalis. Degeneration of the posterior column in the sacral and thoracic cord (myelin sheath stain). Reprinted with permission from Merritt HH, Adams RD, Solomon HC: *Neurosyphilis*. Oxford University Press, New York, 1946.)

General paresis often starts with subtle cognitive and emotional changes, such as problems with concentration and irritability; if untreated, it can lead to memory loss, confabulation, anomia, apraxia, or pseudobulbar palsy. The disease may mimic any other psychiatric disorders as well. Half of the patients with neurosyphilis will manifest dementia of whom one quarter will have prominent psychiatric manifestations such as depression, paranoia, psychosis, or mania. A worsening of symptoms during the first 24 hours after the initiation of antibiotic treatment has been termed the *Jarisch-Herxheimer reaction*; in rare cases, psychosis may emerge shortly after antibiotics are started. With disease progression there is loss of muscle tone and fine motor control and seizures, spasticity, and eventually paralysis and death occur. Focal neurological findings are rare, consistent with the generalized pathophysiology. Tabes dorsalis on the other hand develops somewhat later than general paresis, 15 to 20 years after infection, and causes a more characteristic clinical picture of lancinating pains, attacks of abdominal pain, and paresthesias. Because of progressive loss of proprioception and sensation, patients compensate by walking with a broad-based, shuffling gait. Unlike patients with general paresis, not all patients with tabes will have CSF abnormalities.

Tests *T. pallidum* is difficult to demonstrate in the CSF and difficult to culture. Although PCR techniques are being developed to detect the genetic material of the spirochete, this method is currently only available in research laboratories. Clinicians must rely upon serological tests in the context of a careful history and physical examination. Serological tests for syphilis include the nontreponemal Veneral Disease Research Laboratory (VDRI) and rapid plasma reagin (RPR) tests and, for confirmatory purposes, the fluorescent treponemal antibody-absorption (FTA-ABS) test. CSF studies are useful to confirm the diagnosis of neurosyphilis if clinical findings are suggestive, to diagnose asymptomatic involvement so that treatment can be started, and to follow treatment efficacy. These CSF studies are limited by the low specificity of the typical abnormalities of elevated protein, g-globulin, and leukocyte count and the low sensitivity (but high specificity) of the VDRL test. The CSF FTA-ABS test on the other hand is thought to have excellent sensitivity but less specificity than the CSF VDRL test. Neuroradiological studies of neurosyphilis report the presence of cortical atrophy, most commonly affecting the frontal and temporal lobes.

Treatment The goal in clinical neurosyphilis is to reverse the manifestations or arrest the disease progression, although in some patients antibiotic therapy may not be able to achieve these goals. Standard courses consist of intravenous aqueous penicillin G, 12 to 24 million units daily in divided doses at 4-hour intervals for 2 weeks, or intramuscular weekly injections of 2.4 to 4.8 million units of penicillin G benzathine for 3 weeks or intramuscular injections of 2.4 million units of penicillin G procaine four times daily for 2 weeks. The likelihood of marked improvement for patients with general paresis is less than that for patients with syphilitic meningitis or meningovascular syphilis, reflecting the pathological process, which in the former is irreversible neuron damage and in the latter is CNS inflammation. During the first year after treatment, the serum and CSF should be regularly monitored for the reemergence of reactivity so that treatment can be reinitiated if necessary. Certain conditions, such as comorbid HIV infection, may place patients at greater risk for persistence of treponemal infection after antibiotic treatment. However, most neurosyphilis patients, when treated, will show improvement in the cognitive, psychiatric, and functional domains.

NON-HIV VIRAL INFECTIONS OF THE CENTRAL NERVOUS SYSTEM

Numerous viruses are invasive and neurotropic, with the extent of consequent neuronal dysfunction varying widely depending upon the virulence of the virus and the immunological response of the host. This section will focus upon agents known to cause striking neuropsychiatric diseases: herpes simplex, rabies, measles, and subacute sclerosing panencephalitis; Table 2.9-1 presents other infectious causes of neuropsychiatric disorders.

Bacterial Infections
Acute
Haemophilus, meningococcus, pneumococcus
Subacute
Brucellosis, "Group A Beta Hemolytic Streptococcus" leptospirosis, Lyme disease, syphilis, tuberculosis, Whipple's disease
Fungal Infections
Coccidioidomycosis, cryptococcosis, histoplasmosis, *Candida*
Parasitic Infections
Cysticercosis, malaria, toxoplasmosis
Prions
Creutzfeldt-Jakob disease, fatal familial insomnia, Kuru
Viral Infections
Arbovirus, coxsackievirus, cytomegalovirus, enterovirus, Epstein-Barr virus, Flavivirus, herpes simplex virus, human immunodeficiency virus, influenza virus, lymphocytic choriomeningitis virus, measles virus, mumps virus, papovavirus, poliovirus, rabies virus, rubella, togavirus

Table 2.9-1 Selected Infectious Causes of Neuropsychiatric Disorders

Herpes Viruses Included under the spectrum of herpesviruses are human herpesvirus 1 (HHV-1), that is, herpes simplex virus 1; HHV-2, that is, herpes simplex virus 2, varicella-zoster virus, Epstein-Barr virus (EBV), cytomegalovirus (CMV), HHV-6, HHV-7, and Kaposi's sarcoma herpesvirus.

Herpes Simplex Herpes simplex encephalitis is a dramatic disorder, characterized by the abrupt onset of fever, personality change, and headaches, followed by cognitive changes and focal neurological signs, such as aphasia, visual field deficits, hemiparesis, or partial seizures. Although focality is an important feature of herpes simplex encephalitis, other viruses, such as the LaCrosse virus or the nonpolio enteroviruses, may also cause focal signs. Neurobehavioral aspects of herpes simplex encephalitis such as hallucinations, memory loss, or behavioral disturbances may be the primary clinical feature, a consequence of the predilection of the virus for the temporal lobes. Although the course of illness is typically rapidly progressive, resulting in refractory seizures, coma, and death within 2 weeks, occasionally the progression may be slower with varied neuropsychiatric features.

HSV-1 is usually transmitted orally entering the CNS through sensory nerves, particularly the trigeminal ganglia. HSV-2 is transmitted genitally and may seed the sacral ganglia or disseminate hematogenously. Herpes simplex viruses typically produce a lytic infection with neuronal necrosis and tissue destruction, and intranuclear inclusion bodies in the neurons and glia. Patients who survive herpes simplex encephalitis may exhibit postencephalitic symptoms, such as amnesia, aphasia, and less commonly, the Klüver-Bucy syndrome or dementia.

Routine serological studies are of little value in suspected herpes simplex encephalitis. The CSF usually demonstrates leukocytosis (approximately 100 cells/mm ^{3}), a moderate protein elevation, and a normal or depressed glucose content. PCR analysis of the CSF to detect HSV DNA is at present the diagnostic procedure of choice because the PCR assay has high sensitivity and specificity. Recent studies indicate that approximately 80 percent of patients with biopsy-proven herpes simplex encephalitis will have focal EEG abnormalities consisting of slowing or repetitive epileptiform discharges in the frontotemporal area. MRI studies in early stages of herpes simplex encephalitis may reveal T2 prolongation in the insular cortex and cingulate gyrus. SPECT or PET imaging may show reduced blood flow in the

orbitofrontal and temporal regions. Brain biopsy can be helpful in cases that are difficult to diagnose, although the complication rate is approximately 3 percent.

If untreated, 40 to 70 percent of patients with herpes simplex encephalitis will die. Antiviral therapies include acyclovir (zovirax) and vidarbine (Vira-A); however, even with acyclovir treatment fewer than 40 percent of patients survive with minimal or no sequelae (Fig. 2.9-3).

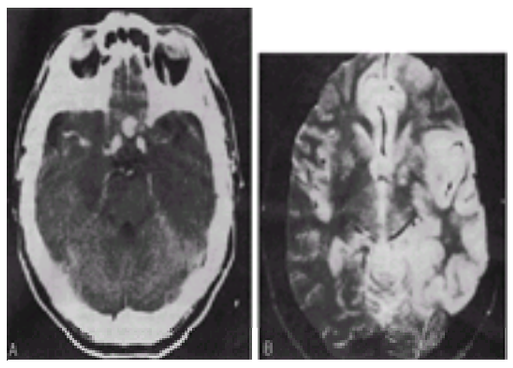

FIGURE 2.9-3 Herpes encephalitis. **A**, Contrast-enhanced axial CT scan shows diffuse decreased density of the left temporal lobe with minimal hypodensity of the medial right temporal lobe. An abnormal CT scan is usually not seen until day 6 to 7 after the onset of manifestations. Eventually, a majority of scans show gyral enhancement in the sylvian fissure area. These findings should raise the suspicion of an underlying infectious lesion such as herpes, early infarction from emboli or vasculitis, or metastatic tumors. **B**, T2-weighted axial MRI scan shows diffuse increased signal intensity along the left temporal lobe cortex as well as the posteromedial left temporal lobe. Both medial frontal lobes are involved as well. An abnormal MRI scan is usually seen by day 1 or 2 after the onset of manifestations. The patient was treated for herpes encephalitis and responded to acyclovir therapy. (Reprinted with permission from Jubelt B, Miller JR: Viral infections. In *Meritt's Textbook of Neurology*, ed 9, LP Rowland, editor. Williams & Wilkins, Baltimore, 1995.)

Epstein-Barr Virus Most adults have evidence of past exposure to EBV, with approximately 50 percent seropositivity among children over age 5. Infection in childhood is generally mild, whereas in adolescence and young adulthood it may result in infectious mononucleosis or, rarely, a fulminant life-threatening disease. EBV enters the body by infecting oral mucosal epithelial cells. The clinical symptoms of infectious mononucleosis of sore throat, headache, malaise, and fatigue are largely a result of the vigorous cellular immune response to EBV infection rather than direct cytotoxic effects. Significant neurological complications of EBV infection are rare, occurring in less than 0.5 percent of cases of infectious mononucleosis.

EBV encephalitis occurs usually within 1 to 3 weeks after the onset of clinical infectious mononucleosis. Patients with EBV encephalitis may present with cerebellar ataxia, personality changes, psychosis, transient global amnesia, perceptual distortions of size and space, focal neurological findings, seizures, or coma. EEG usually reveals generalized slowing with occasional sharp-wave activity. The diagnosis of an EBV neuropsychiatric syndrome requires an appropriate clinical history in the setting of serological evidence of acute or, rarely, chronic active infection. In cases of EBV encephalitis commonly there is a lymphocytic pleocytosis (atypical lymphocytes are particularly suggestive) with elevated protein. In most cases EBV encephalitis is self-limited, with recovery occurring within weeks to months; rarely, acute EBV infection may result in a relapsing or chronic encephalitis. Treatment is generally supportive.

Other Herpes Viruses With herpes zoster, neuropsychiatric complications occur most frequently in immunocompromised patients, resulting in encephalitis, myelitis, or leukoencephalitis. With CMV infection, encephalitis may also occur because CMV is tropic for the CNS; however, only in rare exceptions has CMV encephalitis occurred in non–HIV-infected immunocompromised individuals.

Rabies Although most cases of human rabies occur after animal bites, other sources of rabies infection include aerosols (risk for spelunkers) and person-to-person transmission following corneal transplants. The virus replicates locally at the site of inoculation and subsequently spreads to the CNS by retrograde axonal transport, infecting the lower areas of the brain most prominently, particularly the limbic system, hippocampus, brainstem, and cerebellum. Limbic system involvement may result in aberrant sexual behavior and behavioral dyscontrol, whereas brainstem involvement typically results in alterations of body temperature and respiratory control. The site and amount of inoculation is associated with morbidity. For example, multiple dog bites to the face may result in a 60 percent mortality rate without prophylactic intervention whereas multiple bites to the hand are associated with lower mortality rates of about 15 percent. The incubation period prior to symptomatic expression ranges from a few days to several years. Once symptoms emerge, the course is rapidly fatal. Most patients get the furious form characterized by agitation, hallucinations, odd behaviors, extreme excitability, and in some cases, hydrophobia. Diagnosis is based on the history of an animal bite in a patient with unexplained encephalitis that has been confirmed by the demonstration of rabies antigen on a skin biopsy of the patient or from a putatively infected animal. There is no treatment for rabies virus infection. Disease prevention is critical, aided by preexposure vaccination in high-risk individuals and postexposure prophylaxis with rabies immunoglobulin and rabies vaccine (Fig. 2.9-4).

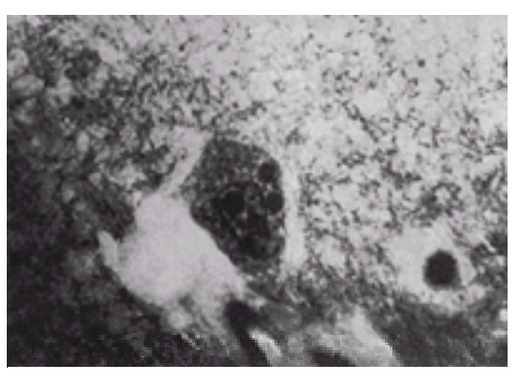

FIGURE 2.9-4 Rabies. Inclusion bodies (Negri bodies) in cytoplasm of ganglion cell of cerebral cortex. (Reprinted with permission from Jubelt B, Miller JR: Viral infections. In *Merritt's Textbook of Neurology*, ed 9, LP Rowland, editor. Williams & Wilkins, Baltimore, 1995.)

Subacute Sclerosing Panencephalitis Subacute sclerosing panencephalitis is a very rare slow infection with measles virus that causes progressive inflammation and sclerosis of the brain. Primarily affecting children and young adults, the rate of subacute sclerosing panencephalitis decreased markedly after 1960 as a result of widespread measles vaccination, with a current rate in the United States of only 1 per 100 million people per year. The onset generally occurs 7 to 12 years after measles and is subtle, characterized by gradual changes in behavior and school performance. Neuropsychological testing may demonstrate reduced overall intelligence and problems with reading, writing, and visuospatial processing. Neuropsychiatric symptoms may include hallucinations, apraxia, agnosia, and Balint's syndrome (optic ataxia, simultanagnosia, and sticky fixation). Repetitive myoclonic jerks are common, at times accompanied by movement disorders and cerebellar ataxia. In advanced stages, dementia, mutism, cortical blindness, optic atrophy, stupor, coma, and death occur. The usual course of the illness is 1 to 3 years, with rare patients surviving up to 10 years.

Serological testing may reveal unusually high titers of antibodies to measles virus. CSF studies typically show high measles antibody titers and a greatly elevated gamma globulin fraction with oligoclonal bands in a CSF with slightly elevated protein concentrations. EEG studies are essential, particularly in the myoclonic stage, when they reveal high-amplitude bilateral and stereotyped complexes that repeat every 3 to 5 seconds. MRI studies may reveal enlarged ventricles and diffuse brain atrophy, with multifocal low-density white matter lesions and lucent areas in the basal ganglia. PET and SPECT studies may reveal early subcortical hypermetabolism followed by global cortical and subcortical hypometabolism.

No treatments are known to reverse the disease, although slightly prolonged survival has been reported with isoprinosine (Inosiplex) and with intraventricular or

intrathecal injections of interferon-a.

Progressive Multifocal Leukoencephalopathy This disease affects immunocompromised subjects and is a progressive infection of oligodendroglial cells with the JC papovavirus. Typically the onset is abrupt with focal neurological or neuropsychological signs and the course is almost invariably fatal within 2 to 4 months. Definitive diagnosis requires a brain biopsy. Neuroimaging studies reveal multifocal areas of high signal intensity in the white matter. Functional imaging with PET or SPECT may reveal a heterogeneous pattern of reduced metabolic activity and perfusion.

SUBACUTE SPONGIFORM ENCEPHALOPATHIES

Included in this group are Creutzfeldt-Jakob disease; *kuru*, a dementing disease of three New Guinea tribes that is most likely spread by ritual cannibalism; *Gerstmann-Straüssler syndrome*, a familial disorder characterized by dementia and ataxia; and *fatal familial insomnia*, a disorder causing disturbances of sleep and of motor, autonomic, and endocrine function. These disorders are all slow infections caused by a transmissible agent not yet clearly described that may be a prion, a virino or an atypical virus. Prions are proteinaceous agents devoid of nucleic acid that are crucial in the pathogenesis of the spongiform encephalopathies. A virino is a small molecule (probably a nucleic acid) associated with a host protein. Characteristic of the neuropathology of these disorders is the neuronal vacuolation that leads to spongy degeneration of the cerebral cortical gray matter.

Creutzfeldt-Jakob Disease Invariably fatal, this transmissible, rapidly progressive disorder occurs mainly in middle age or older and is manifest early on by fatigue, flu-like symptoms, and mild cognitive impairment or focal findings such as aphasia or apraxia. Subsequent psychiatric manifestations include mood lability, anxiety, euphoria, depression, delusions, hallucinations, or marked personality changes. Progression of disease occurs over months leading to dementia, akinetic mutism, coma, and death. Other common neurological findings are generalized startle myoclonus, cortical blindness, and extrapyramidal and cerebellar signs.

The rates of Creutzfeldt-Jakob disease range from 0.25 to 2 cases per million persons a year worldwide. The infectious agent self-replicates and can be transmitted to humans by inoculation with infected tissues and sometimes by ingestion in food. Iatrogenic transmission has been reported via transplantation of contaminated cornea or to children via contaminated supplies of human growth hormone. Household contacts are not at greater risk than the general population, unless there is direct inoculation. Because of an epidemic of a newly recognized prion disease, bovine spongiform encephalopathy (mad cow disease), among cattle in the United Kingdom and because of the unexpected recent emergence of cases of an atypical form of Creutzfeldt-Jakob disease among teenagers in the United Kingdom, fears exist that transmission to humans may have occurred as a result of eating infected meat (Fig. 2.9-5).

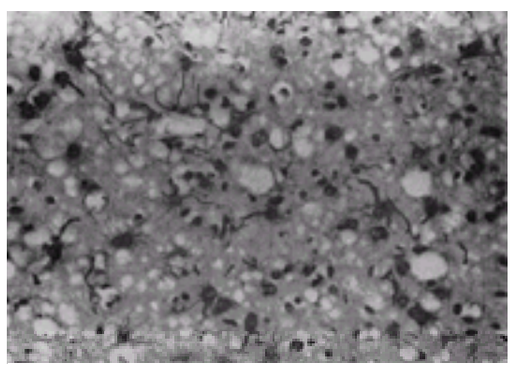

FIGURE 2.9-5 Creutzfeldt-Jakob disease. Section from cortex showing status spongiosis of the neuropil, loss of neurons, and prominent astrocytosis. (PTAH stain × 120). (Reprinted with permission from Jubelt B, Miller JR: Viral infections. In *Merritt's Textbook of Neurology*, ed 9, LP Rowland, editor. Williams & Wilkins, Baltimore, 1995.)

Diagnosis requires pathological examination of the cortex, which reveals the classic triad of spongiform vacuolation, loss of neurons, and glial cell proliferation. Genetic susceptibility is a factor in disease risk, indicated by a common polymorphism of the human prion protein. An immunoassay for Creutzfeldt-Jakob disease in the CSF is currently under development, showing promise in supporting the diagnosis of Creutzfeldt-Jakob disease in patients with dementia. EEG abnormalities, although not specific for Creutzfeldt-Jakob disease, are present in nearly all patients: a slow and irregular background rhythm with periodic complex discharges. Computed tomography (CT) and MRI studies may reveal cortical atrophy later in the course of disease; SPECT and PET reveal heterogeneously decreased uptake throughout the cortex. There is no known treatment for Creutzfeldt-Jakob disease.

OTHER INFECTIOUS CAUSES OF NEUROPSYCHIATRIC DISORDERS

A variety of bacterial, mycoplasmal, fungal, and parasitic infections can cause neuropsychiatric disturbances as a result of a chronic meningitis or sequelae from an acute infection (Table 2.9-1).

EMERGING AREAS OF INVESTIGATION

Chronic Fatigue Syndrome Chronic fatigue syndrome, more commonly referred to as *myalgic encephalomyelitis* in the United Kingdom and Canada, is a multisystem syndrome characterized by 6 months or more of severe, debilitating fatigue, often accompanied by myalgia, headaches, pharyngitis, low-grade fever, cognitive complaints, gastrointestinal symptoms, and tender lymph nodes. The search for an infectious cause of chronic fatigue syndrome has been active because of the high percentage of patients who report abrupt onset after a severe flu-like illness. In the mid-1980s chronic fatigue syndrome was linked to infection with Epstein-Barr virus. After EBV was shown in controlled studies to have no specific role in the etiology of chronic fatigue syndrome, reports have linked chronic fatigue syndrome to a variety of other agents, including enteroviruses, retroviruses, and new lymphotropic herpesviruses but these reports have not been consistently replicated in well-designed studies. Certain organisms, however, such as *B. burgdorferi* (which causes Lyme disease), can result in a chronic fatigue syndrome-like picture; however, most cases of CFS are not linked to Lyme disease. Evidence of immune dysregulation has been frequently reported among patients with chronic fatigue syndrome, but the data are not consistent across studies nor are they reflective of illness severity. Various studies have found high rates (15 to 54 percent) of depressive disorders among patients with chronic fatigue syndrome. In addition, recent research has shown that patients who are most likely to be plagued by persistent fatigue after an acute viral illness are patients with preexisting or comorbid psychiatric problems. However, other research has shown that the cognitive impairment in chronic fatigue syndrome exists even in the absence of preexisting or comorbid psychiatric disorders, thus leading to the conclusion that psychiatric disorders alone cannot account for chronic fatigue syndrome. At present, chronic fatigue syndrome is best conceptualized as a heterogeneous syndrome of uncertain etiology, most likely involving an interplay of psychiatric, infectious, neuroendocrine, and immunological factors.

Group A b-Hemolytic Streptococci Poststreptococcal autoimmunity has been postulated to be a cause of certain types of childhood-onset obsessive-compulsive disorders and Tourette's disorder based on the observation that children who develop Sydenham's chorea are often observed to have tics or obsessive-compulsive symptoms prior to the onset of the chorea. These *pediatric autoimmune neuropsychiatric disorders* are characterized by abrupt and dramatic symptom exacerbations that are temporally related to group A b-hemolytic streptococcal infections. Recent research has identified a genetic marker in pediatric autoimmune neuropsychiatric disorders that has previously been shown to be both highly specific and sensitive in identifying individuals with rheumatic fever. In one study 85 percent of children who developed streptococcal-related obsessive-compulsive disorder or tics and 89 percent of children with Sydenham's chorea carried the D8/17 monoclonal antibody marker on DR+ cells in the peripheral circulation, whereas only 17 percent of healthy controls carried this marker. Investigations are currently underway to determine whether treatments that modulate the immune response (e.g., intravenous g-globulin or plasmapheresis) are effective in eliminating obsessive-compulsive disorder and tic disorders among children with pediatric autoimmune neuropsychiatric disorders.

Borna Disease Virus Borna disease virus (BDV) is a small neurotropic ribonucleic acid (RNA) virus that infects various domestic animal species, causing disturbances in behavior and cognition and, rarely, death. Researchers have found that BDV targets cells of the limbic system in animals and compromises their neuronal function without causing direct damage. BDV has recently been linked to a wide array of neuropsychiatric disorders in humans. Evidence suggestive of BDV infection in humans has been accumulating from several research centers. One research group recently identified serum antibodies to BDV in 9.6 percent of 416 people with schizophrenia, major depressive disorders, bipolar I disorder, and other neuropsychiatric disease whereas these serum antibodies were found in only 1.5

percent of 203 healthy controls. Reverse transcriptase-PCR identified BDV RNA sequences in 13 of a subset of 26 psychiatric patients but in none of 23 healthy controls. Other reports have identified BDV antibodies in 6.8 percent of patients with psychiatric illnesses versus 3 percent of surgical controls. These studies represent provocative preliminary findings that suggest a possible role for BDV in a subset of human neuropsychiatric diseases; well-controlled microbiological and epidemiological studies are needed to determine the significance of these reports.

SUGGESTED CROSS-REFERENCES

Acquired immune deficiency syndrome is discussed in Section 2.8; interactions of the immune system and the CNS are discussed in Section 1.12; neuropsychological testing is discussed in Section 7.4; and neuroimaging is discussed in Section 2.13. Obsessive-compulsive disorder and schizophrenia are discussed in Chapter 15 and Chapter 12, respectively.

SECTION REFERENCES

Ackerman R, Rehse-Kupper B, Gollmer E, Schmidt R: Chronic neurologic manifestations of erythema migrans borreliosis. Ann NY Acad Sci *64:*506, 1988.

Ancar B, Yalaz K, Oktem F, Köse G: Long-term follow-up of patients with subacute sclerosing panencephalitis treated with intraventricular alpha-interferon. Neurology *48:*526, 1997.

Bates DW, Buchwald D, Lee J, Kith P, Doolittle T, Rutherford C, Churchill H, Schur P, Wener M, Wybenga D, Winkelman J, Komaroff AL: Clinical laboratory test findings in patients with chronic fatigue syndrome. Arch Intern Med *155:*97, 1995.

Bode L, Zimmermann W, Ferszt R, Steinbach F, Ludwig H: Borna disease virus genome transcribed and expressed in psychiatric patients. Nat Med *1:*232, 1995.

Bloom BJ, Wyckoff PM, Meissner HC, Steere AC. Neurocognitive abnormalities in children after classic manifestations of Lyme disease. Pediatr Infect Dis J *17:*189, 1998.

Brown P, Cathala F, Castaigne P, Gajdusek DC: Creutzfeldt-Jakob disease: Clinical analysis of a consecutive series of 230 neuropathologically verified cases. Ann Neurol *20:*597, 1986.

*Burrascano JJ: Lyme disease. In *Conn's Current Therapy*, Rakel RE, editor. Saunders, Philadelphia, 1997.

Collinge J: New diagnostic tests for prion diseases. N Engl J Med *335:*963, 1996.

Coyle PK: Neurologic Lyme disease. Semin Neurol *12:*200, 1992.

Coyle PK, Schutzer SE, Deng Z, Krupp LB, Belman AL, Benach JL, Luft BJ: Detection of *Borrelia burgdorferi*-specific antigen in antibody-negative cerebrospinal fluid in neurologic Lyme disease. Neurology *45:*2010, 1995.

DeLuca J, Johnson SK, Ellis SP, Natelson BH: Cognitive functioning is impaired in patients with chronic fatigue syndrome devoid of psychiatric disease. J Neurol Neurosurg Psychiatry *62:*151, 1997.

*Demitrack MA, Abbey SE, editors: *Chronic Fatigue Syndrome.* Guilford Publications, New York, 1996.

Fallon BA, Nields JA, Burrascano JJ, Liegner K, DelBene D, Liebowitz MR: The neuropsychiatric manifestions of Lyme borreliosis. Psychiatric Q *63:*95, 1992.

*Fallon BA, Nields JA: Lyme disease: A neuropsychiatric illness. Am J Psychiatry *151:*1571, 1994.

Fallon BA, Das S, Plutchok JJ, Tager F, Liegner K, Van Heertum R: Functional brain imaging and neuropsychological testing in Lyme disease. Clin Infect Dis *24*(Suppl): S57-63, 1997.

Haywood AM: Transmissible spongiform encephalopathies. N Engl J Med *337:*1821, 1997.

Heegaard ED, Hornsleth A: Parvovirus: The expanding spectrum of disease. Acta Paediatr *84:*109, 1995.

Hendler N, Leahy W: Psychiatric and neurologic sequelae of infectious mononucleosis. Am J Psychiatry *135:*842, 1978.

Hooshmand H, Escobar MR, Kopf SW: Neurosyphilis: A study of 241 patients. JAMA *219:*726, 1972.

*Huang C, Chatterjee NK, Grady LJ: Diagnosis of viral infections of the central nervous system. N Engl Med *340:*483, 1999.

*Huisman TA, Wohlab G, Nadal D, Boltshauser E, Martin E: Unusual presentations of neuroborreliosis (Lyme disease) in childhood. J Comput Assist Tomogr *23:*39, 1999.

Krupp LB, Masur D, Schwartz J, Coyle PK, Langenbach LJ, Fernouist SK, Jandorf L, Halperin JJ: Cognitive functioning in late Lyme borreliosis. Arch Neurol *48:*1125, 1991.

Lawrence C, Lipton RB, Lowy FD, Coyle PK: Seronegative chronic relapsing neuroborreliosis. Eur Neurol *35:*113, 1995.

Lecour H, Miranda M, Magro C, Rocha A, Goncalves V: Human leptospirosis—a review of 50 cases. Infection 17:10, 1989.

Logigian EL, Kaplan R, Steere AC: Chronic neurologic manifestations of Lyme disease. N Engl J Med *323:*1438, 1990.

Jones KJ, Garada BM, Holman BL, Steere AC: Reversible cerebral hypoperfusion in Lyme encephalopathy. Neurology 1997; *49:*1661, 1997.

Oksi J, Kalimo H, Marttila RJ, Marjamaki M, Sonninen P, Nikoskelainen J, Viljanen MK: Inflammatory brain changes in Lyme borreliosis. Brain *119:*2143, 1996.

Pachner AR: Borrelia burgdorferi in the nervous system: The new “Great Imitator.” In *Lyme Disease and Related Disorders.* Ann N Y Acad Sci *539:*56, 1988.

Sauder C, Muller A, Cubitt B, Mayer J, Steinmetz J, Trabert W, Ziegler, Wanke K, Mueller-Lantzsch N, de la Torre JC, Grasser FA: Detection of borna disease virus (BDV) antibodies and BDV RNA in psychiatric patients: Evidence for high sequence conservation of human blood-derived BDV RNA. J Virol *70:*7713, 1996.

*Scheld WM, Whitley RJ, Durack DT, editors: *Infections of the Central Nervous System.* Lippincott-Raven, New York, 1997.

Swedo SE, Leonard HL, Mittleman BB, Allen AJ, Rapoport JL, Dow SP, Kanter ME, Chapman F, Zabriskie J: Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. Am J Psychiatry *154:*110, 1997.

Swedo SE: Sydenham's chorea: A model for childhood autoimmune neuropsychiatric disorders. JAMA *272:*1788, 1994.

Thomas EW: Syphilis: Its course and management. MacMillan, New York, 1949.

Waltrip RW, Buchanan RW, Summerflet A, Breier A, Carpenter WT, Bryant NL, Rubin SA, Carbone KM: Borna disease virus and schizophrenia. Psychiatry Res *56:*33, 1995.

Wessely S, Chalder T, Hirsch S, Pawlikowska T, Wallace P, Wright DJM: Postinfectious fatigue: Prospective cohort study in primary care. Lancet *345:*1333, 1995.

*Yolken RH, Torrey EF: Viruses, schizophrenia, and bipolar disorder. Clin Microbiol Rev *8:*131, 1995.

2.10 NEUROPSYCHIATRIC ASPECTS OF HEADACHE

KATHLEEN RIES MERIKANGAS, PH.D., AND JAMES R. MERIKANGAS, M.D.

Headache is one of the most common human afflictions. It is the condition that most often leads people to seek medical advice. Ten percent of all people report that headache leads to impairment in their daily life. It has dramatic impact on occupational and social disability and the use of health services. Despite substantial research efforts, the etiology of headache syndromes remains poorly understood. Because headache is commonly associated with psychiatric syndromes, psychiatrists are often consulted for the evaluation and treatment of people suffering from headache. The purpose of this chapter is to: (1) provide an overview of the current diagnostic nomenclature for the most common headache syndromes, (2) describe the clinical and laboratory examination necessary for making a diagnosis of headache, and (3) summarize current treatment strategies for the major headache syndromes.

DEFINITIONS

The International Headache Society (IHS) introduced a new headache classification system in 1988 (Table 2.10-1) in order to provide specific operational criteria for the major types of headache and to facilitate international standardization of the diagnostic nomenclature for headache syndromes. This classification system consists of a classification scheme and diagnostic criteria as well as guidelines for differential diagnosis. The criteria are intended to be applied to classify headache subtypes based on information obtained from a history, a physical and neurological examination, and appropriate laboratory investigations. This system has no category for “psychosomatic” or “psychogenic” headache because there has never been a valid demonstration that mental illness or life stress causes headaches. The most frequent primary headache syndromes are *migraine, tension-type headache*, and *cluster headache.* The other headache subtypes described in the IHS system are secondary to a variety of acute and chronic conditions. For a thorough clinical evaluation of headache it is essential to perform a differential diagnosis of the multiple causes of headache.

1. Migraine
2. Tension-type headache
3. Cluster headache and chronic paroxysmal hemicrania
4. Miscellaneous headaches unassociated with structural lesions
5. Headache associated with head trauma
6. Headache associated with vascular disorder
7. Headache associated with nonvascular intracranial disorder
8. Headache associated with substances or their withdrawal
9. Headache associated with noncephalic infection
10. Headache associated with metabolic disorder
11. Headache or facial pain associated with disorder of cranium, neck, eyes, ears, nose, sinuses, teeth, mouth, or other facial or cranial structures
12. Headache associated with cranial neuralgias, nerve trunk pain, and deafferentation pain
13. Headache not classifiable

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias, and facial pain. Cephalalgia 8(Suppl):1, 1988.

Table 2.10-1 Headache Classification System of the International Headache Society

Migraine Migraine is a complex debilitating condition characterized by either the presence or absence of aura symptoms. Migraine presentation is multifaceted and symptoms emanate from multiple systems, including vascular, neurological, gastrointestinal, endocrine, and visual manifestations. These symptoms may be accompanied by a variety of changes in behavior and cognition, including mood alterations and confusion. Historically, the usual definitions of migraine included the presence of cyclic headaches associated with a variety of gastrointestinal and neurological symptoms. There is general agreement that a thorough evaluation is required for all patients who present with headache complaints.

The IHS criteria for migraine with and without aura are presented in Table 2.10-2 and Table 2.10-3. The core features of most definitions of migraine include recurrent headache, accompanied by gastrointestinal symptoms such as nausea or vomiting, and hyperesthesia manifested by photophobia or phonophobia. The headache generally has a pulsatile or throbbing quality exacerbated by routine physical activity involving movement of the head, and is often unilateral. The IHS criteria operationalize these features of headache to establish common thresholds and distinctions between migraine and other types of headache. Migraine was formerly divided into two major subtypes, *common* and *classic*, with the latter being distinguished by the presence of neurological symptoms that precede the onset of the headache. The new classification of migraine by the IHS no longer includes the common-classic distinction; instead migraine is subtyped according to the presence or absence of aura symptoms (reversible neurological dysfunction).

A. At least 5 attacks
B. Duration between 4 and 72 hours
C. At least 2 of the following:
   1. Unilateral
   2. Pulsating pain
   3. Moderate to severe intensity
   4. Worsening with exertion
D. One of the following:
   1. Nausea or vomiting
   2. Photophobia and phonophobia

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 8(Suppl):1, 1988.

Table 2.10-2 International Headache Society Criteria: Migraine Without Aura

A. At least 2 attacks that fulfill criteria in B and C
B. At least 3 of the following 4 characteristics:
   1. One or more completely reversible aura symptoms that indicate local cerebral cortical or brainstem dysfunction (or both)
   2. At least one aura symptom develops gradually over >4 min or two or more symptoms occur in succession
   3. No aura symptom; lasts >60 min
   4. Headache follows aura in <1 hour
C. No evidence of related organic disease

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 8(Suppl):1, 1988.

Table 2.10-3 International Headache Society Criteria: Migraine With Aura

Despite recent progress in the standardization of the classification of migraine by the IHS, the diagnostic criteria have not been subjected to intensive investigation with respect to reliability or validity. However, several studies currently under way are examining these issues. Several features unique to the headache syndromes constitute impediments to developing a valid set of diagnostic criteria. These impediments include: the co-occurrence of multiple headache syndromes within individual persons; the tendency for headache characteristics to change across the life span; the effects of professional and self-treatment of headache in obscuring the manifestations of the underlying headache syndromes; and the lack of generalizability of treated samples from which the diagnostic criteria were derived. Specific areas of the classification system require additional clarification; these are the specification of procedures for ensuring standardized methodology for the ascertainment of the diagnostic criteria, methods for assessing and coding multiple headache syndromes within individuals, and the development of standardized methods for discriminating between primary headache syndromes and those whose etiology is known (i.e., secondary headaches).

Tension-Type Headache The definition of tension-type headache according to IHS criteria is presented in Table 2.10-4. Tension headache is characterized by episodes of bilateral pain lasting several days at a time. It is distinguished from migraine headache by its generally longer duration, the lack of pulsating quality of the pain, the lack of worsening with physical activity, and the absence of gastrointestinal concomitants. However, migraine and tension-type headache may often coexist, either simultaneously or alternating over time. Tension-type headache is no longer believed to result from muscle tension; indeed, neck pain may result from head movement to reduce headache pain.

1. Episodic tension-type headache
   A. At least 10 previous headache episodes fulfilling criteria B-D listed below; number of days with such headache: <180/year
   B. Headache lasting from 30 min to 7 days
   C. At least two of the following pain characteristics:
      1. Pressing/tightening (nonpulsating) quality
      2. Mild or moderate intensity
      3. Bilateral location
      4. No aggravation by walking stairs or similar routine physical activity
   D. Both of the following:
      1. No nausea or vomiting (anorexia may occur)
      2. Absence of photophobia and phonophobia, or presence of one but not the other
2. Chronic tension-type headache
   A. Average headache frequency ≥15 days/month for ≥6 months (180 days/year) fulfilling criteria B-D listed below
   B. At least two of the following pain characteristics:
      1. Pressing/tightening quality
      2. Mild or moderate severity
      3. Bilateral location
      4. No aggravation by walking stairs or similar routine physical activity
   C. Both of the following:
      1. No vomiting
      2. No more than one of the following: nausea, photophobia, or phonophobia

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 8(Suppl): 1, 1988.

Table 2.10-4 International Headache Society Diagnostic Criteria for Tension-Type Headache

Cluster Headache Cluster headache is a distinct syndrome characterized by frequent attacks (often several per day) over a 1- to 2-month period, separated by headache-free intervals for as long as 1 or 2 years. Although it is commonly grouped with migraine, current evidence including epidemiological data, treatment response, and clinical features suggests that cluster headache may comprise a distinct syndrome. Table 2.10-5 shows the IHS diagnostic criteria for cluster headache.

A. At least five attacks fulfilling criteria B-D listed below
B. Severe unilateral orbital, supraorbital, or temporal pain lasting 15–180 min untreated
C. Association with at least one of the following signs, which must be present on the pain side:
   1. Conjunctival injection
   2. Lacrimation
   3. Nasal congestion
   4. Rhinorrhea
   5. Forehead and facial sweating
   6. Miosis
   7. Ptosis
   8. Eyelid edema
D. Frequency of attacks from one every other day to eight per day

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 8(Suppl): 1, 1988.

Table 2.10-5 International Headache Society Criteria: Cluster Headache and Chronic Paroxysmal Hemicrania

Cluster refers to a “clustering in time,” with the headache bouts occurring every day to several times a day over a period of days to weeks, followed by a lengthy headache-free interval. Cluster headache is generally retro-orbital in location and is accompanied by autonomic changes such as lacrimation, rhinorrhea, erythema of the eye, and agitation. Patients with cluster headache do not retire to dark rooms and lie down to avoid the stimulation; they may do quite the opposite, appearing almost manic in their agitation. The pain can be so intense that the sufferer may appear to be psychotic because of the screaming and thrashing that may be associated with the pain.

Chronic paroxysmal hemicrania is a type of cluster headache that is specifically responsive to treatment with indomethacin (Indocin) and characterized by many daily focal attacks of pain, each lasting for about 15 or 20 minutes.

Posttraumatic Headache Posttraumatic headache is variable in symptom presentation, severity, and duration. A set of diagnostic criteria was recently established by the IHS, in 1988, as presented in Table 2.10-6. The key symptoms include a headache following head trauma and accompanied by a loss of consciousness, posttraumatic amnesia, and abnormal laboratory tests.

A. Head trauma with at least 2 of the following:
   1. Loss of consciousness
   2. Posttraumatic amnesia >10 minutes
   3. Abnormal results in at least 2 of the following tests:
      a. neurological
      b. spinal fluid
      c. skull x-ray
      d. psychophysiological
      e. neuroimaging
      f. vestibular function
      g. ERP
B. Headache occurs <14 days after trauma or regaining consciousness
C. Headache continues (chronic) or disappears (acute) within 8 weeks after trauma or regaining consciousness

Reprinted with permission from Olesen J: Classification and diagnostic criteria for headache disorders, cranial neuralgias and facial pain. Cephalalgia 8(Suppl): 1, 1988.

Table 2.10-6 International Headache Society Criteria: Posttraumatic Headache

Although headache following a traumatic head injury has often been attributed to emotional factors, empirical evidence suggests that emotional factors are more likely

to be sequelae rather than causes of posttraumatic headache. Nevertheless, the etiology of posttraumatic headache is unknown. The major hypotheses of the pathogenesis of posttraumatic headache include cerebral edema, cortical spreading depression, innate vulnerability to cerebral vasospasm, and transient elevation of intracranial pressure. There is no direct relationship between the prevalence or chronicity of posttraumatic headache and several indicators of severity of head injury including duration of unconsciousness, posttraumatic amnesia, electroencephalographic abnormalities, presence of skull fracture, or the presence of blood in the cerebrospinal fluid. There appears to be an inverse relationship between the severity of the head injury and the development of postinjury headache; posttraumatic headache is more common after injuries that do not result in skull fracture.

The onset of typical migraine attacks following acute head trauma occurs so frequently that it has been hypothesized that head trauma serves as a trigger for migraine in persons with underlying susceptibility to migraine. There is a substantial body of literature suggesting that individuals with either a personal or family history of migraine are particularly susceptible to headache following head trauma. Moreover, relatives of persons who have posttraumatic migraine have a higher rate of neurological symptoms, suggesting a propensity to the neurological manifestations of migraine.

Epidemiology and Course Epidemiological studies have shown that approximately 60 percent of the general population report a history of severe headaches. Milder headaches are reported by about 80 percent of the general population. Migraine without aura and tension-type headache are the most common headache syndromes in the general population. The lifetime prevalence of migraine derived from systematic population surveys is about 12 percent. Epidemiological studies that have applied the IHS criteria reveal lifetime prevalence rates of approximately 6 percent among men (range of 4 percent to 19 percent) and 17 percent among women (range of 8 to 29 percent for women). The severity of migraine ranges from mild to nearly total disability. Over 80 percent of those with migraine report some degree of disability. Recent epidemiological studies have underscored the enormous personal and social burden of migraine in terms of both direct and indirect costs.

Migraine is more common among women and persons between the ages of 20 and 45 years, with the incidence decreasing after the fourth decade of life. Migraine may often begin in childhood when boys and girls are equally likely to suffer from migraine headache. Migraine in childhood is more likely to be associated with gastrointestinal complaints, particularly episodic bouts of stomach pain, vomiting, or diarrhea, and the duration is shorter than that commonly observed in adults. Children with migraine are often misdiagnosed as suffering from “psychosomatic headaches” or school refusal. In women, migraine is strongly associated with reproductive system function, with increased incidence during puberty and the first trimester of pregnancy; migraine is also associated with exogenous hormone use. After menopause, the frequency of migraine attacks generally decreases dramatically, unless estrogen replacement therapy is administered.

A family history of migraine is one of the most potent and consistent risk factors for migraine. Although migraine has been postulated to have an autosomal dominant mode of transmission, recent reviews of the genetic studies of migraine indicate that the mode of transmission has not been well established. However, approximately 30 percent of migraine patients have no family history of this condition.

Migraine is strongly associated with a variety of medical disorders, particularly asthma, allergies, and cardiovascular disease. Comorbid psychiatric conditions include mood disorders (particularly the bipolar subtype), phobias, and panic disorders. The course of migraine is variable. In general, the frequency and duration of migraine decrease at midlife in both men and women, and the symptomatic manifestations may change substantially over time.

Compared to studies of migraine, far fewer studies have investigated the epidemiology and risk factors associated with tension-type headache. The prevalence of tension-type headache has been estimated to range from approximately 30 to 80 percent depending on the definitions used. It is difficult to estimate the true prevalence of tension-type headache given that migraine and tension-type headaches may co-occur in the same individual. Tension-type headaches are also more common in women and young adults, but there is a less steep decrement in prevalence with age.

Although posttraumatic headache is quite rare in the general population (i.e., about 1 percent lifetime prevalence), it is not uncommon among those with a history of a concussion or head injury. The estimates of the prevalence of severe and long-term headache following severe head injury based on retrospective data range from 28 to 62 percent. Children and young adults appear to be particularly susceptible to developing headache after head trauma. The results of prospective studies of the incidence of headache following severe head injury, usually termed *postconcussion headache*, reveal that approximately 50 percent of each series of admissions continue to suffer from headache at the time of discharge from the index admission, with a gradual dissipation to 20 percent in 1 year. Persistence of headache has been related to female sex; age over 45; the presence of dizziness; lack of skull fracture; intracranial hematoma; disorders of smell, hearing, or vision; depression; and impaired concentration.

Cluster headache has a very low population prevalence (less than 1 percent of the general population) and occurs nearly exclusively in males. The age at onset of cluster headache is somewhat later than that of migraine and tension-type headache; the first attack of cluster usually begins in the late 20s or 30s and may recur intermittently throughout life; risk factors include smoking and heavy alcohol use.

ETIOLOGY

The causes of the major types of headaches are still unknown. Although progress in neuroimaging allows for the study of changes in blood flow and electrical activity during headaches, the widespread changes in neural and vascular function during the progression of headache have complicated the understanding of the primary event that triggers migraine. Migraine is believed to result from a combination of external trigger factors such as hunger, sensory input, and hormonal fluctuations that activate serotonin and norepinephrine-containing neurons in the brainstem. This activation alters the physiology of neurons, glia, and blood vessels, which provokes pain and generates neuroinflammatory mediators in the brainstem, thalamus, cerebral cortex, and supporting tissues. The activation of sensory vascular terminals within the blood vessel walls releases substance P and other neuropeptides that trigger a sterile inflammatory process, causing prolonged pain. Although disturbances in serotonin regulation are associated with migraine, numerous other brain chemicals, particularly the neuropeptides and nitric oxide, are also involved in a complex way. Far less is known about the etiology of tension-type headache. It is clear however, that the term *tension-type headache* is a misnomer because there is no evidence that muscle tension is the underlying cause of this headache subtype. Current theory of the cause of cluster headache holds that hypothalamic and central pain control regions trigger a cascade of events in the brainstem comprising afferent pain and efferent parasympathetic pathways.

DIFFERENTIAL DIAGNOSIS AND CLINICAL EVALUATION

A very skillful workup is essential because headache is a nonspecific complaint with an enormous number of causes that range from the trivial to the acutely life-threatening. A thorough examination should include a description of the type and location of pain, timing, precipitants, prodromal events, and associated symptoms. The following factors must be determined in order to define whether or not the headache is migrainous: (1) onset; (2) frequency; (3) location; (4) duration; (5) quality; (6) severity; (7) precipitants; (8) precursors; (9) triggers; (10) phenomena that worsen or relieve the pain; (11) warning signs; (12) prodromal events; (13) specific symptoms including visual changes, gastrointestinal symptoms, or neurological symptoms; (14) sensitivity to light, noise, sound, or touch; (15) mood changes; and (16) cognitive changes. In addition, it is important to obtain a detailed family history, description of course, and history of previous evaluation and treatment. Differential diagnosis of headache is based on a neurological examination to rule out pathognomonic signs that might indicate other brain disorders.

In addition to a history and physical examination, laboratory studies are crucial when metabolic, structural, vascular, or other sources of headache are suspected. Even if the results are negative and do not uncover a metabolic, endocrine, or autoimmune etiology, this information may serve as a baseline for subsequent drug therapy. Application of the IHS criteria require that all the potential causes of headache shown in Table 2.10-1 be considered. The diagnosis of headache requires the exclusion of other conditions, including structural lesion, vascular malformation, viral or bacterial meningitis encephalitis, intracranial abscess or hemorrhage, cerebral contusion, metabolic disorders (urea cycle disorders, aminoacidopathies, mitochondrial disorders), pseudotumor cerebri, vasculitis, brain tumors, sinusitis, and ocular disorders, any of which may be concurrent rather than causal.

Based on the low frequency of detection of lesions such as arteriovenous malformation or brain tumors, the American Academy of Neurology issued a practice parameter discouraging the routine use of neuroimaging procedures in patients with headaches who have normal neurological examination. However, headache experts, who often serve as tertiary referral sources, may ignore this recommendation because of the lack of diagnostic certainty in headache, lack of curative properties of current treatment, and unacceptable medical and legal risks of any missed diagnosis. Table 2.10-7 lists the headache symptoms that indicate further diagnostic workup.

• First headache
• Worst headache
• Gradual worsening over days or weeks
• Vomiting prior to headache onset
• Abnormal neurological examination
• Ongoing systemic illness
• Onset after age 50
• Accompanied by fever
• Occurs during sleep

Table 2.10-7 Headache Symptoms Indicating Further Diagnostic Workup

An image of the brain is mandatory for the evaluation of patients with severe or persistent headache, the first or worst headache, or when a subdural hematoma is suspected. Magnetic resonance imaging (MRI) is indicated when hydrocephalus, brain tumor, sinusitis, vasculitis, or posterior fossa lesions are suspected, or when exposure to electromagnetic radiation is contraindicated. X-rays of the jaw and cervical spine are useful to rule out malocclusions and the degenerative changes of arthritis.

TREATMENT OF HEADACHE SYNDROMES

Migraine The mainstay of migraine treatment is pharmacological intervention. Treatment of migraine can be prophylactic, with medication taken daily; abortive, with medication taken at the onset of an attack; or palliative, with medication taken after the pain has begun. Prophylactic treatments for migraine are of varying effectiveness. Clinical trials of migraine treatment are complicated by the high placebo response rate among subjects with migraine, the heterogeneity of diagnostic subtypes of headache, the intermittent nature of the condition, and the frequent use of additional analgesics to treat headache pain. The treatment of migraine chosen for an individual depends not only on the diagnosis of migraine headache but also on specific patient factors. Excellent reviews of both the short-term and prophylactic treatment of migraine are available. There are also a variety of nonpharmacological approaches to reduce headache. These include eliminating the triggers of attacks; maximizing the regularity of daily schedule, particularly with respect to sleeping and eating habits; biofeedback; and relaxation treatment.

Symptom Relief The nonsteroidal anti-inflammatory drugs (NSAIDs), including ibuprofen (Motrin), naproxen (Naprosyn), and indomethacin and the analgesics aspirin and acetaminophen (Tylenol) are commonly used as first-line treatment of mild to moderate migraine. In 1998 the Food and Drug Administration (FDA) approved an acetaminophen-aspirin-caffeine formulation (Exedrin) for the treatment of migraine. Other classes of drugs that are commonly prescribed for more severe attacks include ergot derivatives, serotonin agonists, and narcotics. Combination agents generally comprised of barbiturates, analgesics, and caffeine are also highly effective in the treatment of migraine episodes. Clinicians should be particularly alert to the potential of drugs such as ergotamine (Ergomar) and opioids being abused. In general, opioids should be restricted to severe attacks that are not responsive to other agents. Oxycodone-acetaminophen (Percocet) and oxycodone-aspirin (Percodan) are two of the most popular drugs among opioid addicts because oxycodone is short acting, effective orally, and a euphoriant with a high street value. With short-term use, however, opioids do not produce dependence. Ergotamine tartrate (Ergomar) and dihydroergotamine (DHE) are two of the most commonly prescribed ergot derivatives for moderate to severe attacks of migraine. In order to counterbalance the common adverse effect of nausea, metoclopramide (Reglan) or prochlorperazine (Compazine) are recommended.

Sumatriptan (Imitrex), a selective serotonin (5-hydroxytrytamine [5-HT]) subtype 1D (5-HT_{1D}) agonist, was introduced for self-administered parenteral treatment of acute, severe migraine. It is also available for oral administration and more recently as a nasal spray. Although relief from headache is almost instantaneous, the major criticism of this drug is the high frequency of rebound headache, which may be a function of the short half-life of the drug. The efficacy of sumatriptan has spurred the development of other serotonin agonists with attempts to reduce the adverse effects associated with sumatriptan. A new serotonin agonist, zolmitriptan (Zomig), administered in oral form, has recently been shown to be effective in the reduction of migraine. In general, the use of narcotics for the treatment of migraine should be restricted to a severe attack that is difficult to abort with other agents. Morphine usually is a better choice than the synthetic narcotics opioids such as meperidine (Demerol) and hydromorphone (Dilaudid).

Migraine Prophylaxis There are six major classes of drugs that have been investigated in the prophylaxis of migraine headaches. These include the b-adrenergic receptor antagonists, antidepressants, calcium channel blockers, serotonin antagonists, NSAIDs, and anticonvulsants. Table 2.10-8 presents the six classes of drugs used in migraine prophylaxis and a range of recommended daily doses. Use of the NSAIDs, particularly aspirin, on a daily basis may be highly effective in migraine prophylaxis. Similarly, antihypertensive drugs of the calcium-channel class or b-adrenergic receptor antagonists have also been widely employed to prevent migraine. Recent studies suggest that the angiotensin-converting-enzyme (ACE) inhibitors and enalapril (Vasotec) may also be effective in the prevention of migraine.

Drug	Daily Dosage (mg)
β-BLOCKERS	
Atenolol	50–150
Metoprolol	100–200
Propranolol	40–240
CALCIUM-CHANNEL INHIBITORS	
Flunarizine	5–10
Verapamil	240–320
SEROTONIN ANTAGONISTS	
Methysergide	4–8
Pizotyline	1–6
ANTIDEPRESSANTS	
Amitriptyline	10–150
Doxepin	10–150
Phenelzine	15–60
ANTIEPILEPTICS	
Sodium valproate	500–1500
NSAIDs	
Naproxen	550

Table 2.10-8 Prophylactic Treatment of Migraine

The b-adrenergic receptor antagonists are currently the most popular treatment choice in migraine prophylaxis. However, the effect of this class of drugs is moderate at best. No study has reported complete elimination of migraine; however, the average duration and severity are reduced by 50 percent in most subjects. Clinicians should be particularly cautious in prescribing this class of drugs to individuals with a history of depression, since b-adrenergic receptor antagonists are associated with the development of anhedonia, irritability, and lassitude, which may occur after many months on any of these agents. In contrast, patients with high levels of autonomic anxiety may actually benefit from this class of drugs.

Given the overlap of symptoms of the actual migraine episode, including short-term changes in energy, appetite, mood, and level of anxiety as well as those that occur between attacks, and those on the anxiety and depression spectrum, it is not surprising that similar pharmacological agents have been successfully used to treat migraine, anxiety, and depression. However, the antidepressant drugs, particularly the tricyclic drugs, have also been shown to be superior to the conventional first-line agents of migraine treatment irrespective of comorbid depression or anxiety. However, patients often report excessive sedation from tricyclic agents, as well as dry mouth, constipation, and weight gain. Combinations of these classes of drugs have also been used for patients who fail to respond to first-line treatments. Tricyclic drugs plus b-adrenergic receptor antagonists and tricyclic drugs plus monoamine oxidase inhibitors (MAOIs) have been used concomitantly in the preventive treatment of severe migraine. Use of tricyclic drugs other than amitriptyline (Elavil) should be encouraged because this agent is the most sedating of all of the drugs in this class. The secondary amines [e.g., nortriptyline (Aventil) and desipramine (Norpramin)] appear to be equally efficacious in the treatment of depression, but have fewer adverse effects than do the parent tertiary amines [e.g., amitriptyline, imipramine (Tofranil)]. However, the relative efficacy of tricyclic agents in migraine prevention has not been examined.

The MAOIs have also been reported to be efficacious in the prevention of migraine headache, particularly in patients who have been unresponsive to first-line

prophylactic treatment. Phenelzine (Nardil) has been considered to be one of the most efficacious antimigraine agents. However, clinicians have generally been reluctant to prescribe MAOIs because of the possibility of a hypertensive reaction to dietary tyramine and the other adverse effects of these agents (i.e., orthostatic hypotension, weight gain, and excessive stimulation). The use of oral calcium-channel blockers such as verapamil (Isoptin) to treat the hypertensive crisis associated with MAOIs may reduce clinicians' reservations about prescribing these agents. Addition of a b-adrenergic receptor antagonist such as atenolol (Tenormin) may reduce the cardiovascular adverse effects of phenelzine. In general, the doses of antidepressant agents that have been studied in migraine prophylaxis are far lower than those used in the treatment of depression.

Valproate (Depakene), which is currently a first-line treatment for bipolar I disorder, has also been used to treat both migraine and mood disorders. Recent evidence reveals that valproate may also be effective for migraine prophylaxis in patients who do not necessarily suffer from bipolar I disorder. Thus, valproate may be the treatment of choice in patients with migraine and bipolar I disorder or recurrent depression.

The strong association of migraine with both depression and anxiety should be considered in the treatment of individuals with migraine. Systematic evaluation of the lifetime history of depression and anxiety is necessary for determining the optimal treatment strategy. If there is a subtype of migraine associated with anxiety and depression, it is critical to treat the entire syndrome rather than to limit the treatment goal to headache cessation. In general, comorbid depression and anxiety are more important in the selection of migraine prophylaxis than in the treatment of an acute attack of migraine. The use of prophylactic medications with adverse effects of lassitude, fatigue, or depression should be avoided, if possible; if not, careful clinical evaluation of these cited manifestations of depression including anergia, hypersomnia, and irritability should be monitored. When nonpharmacological approaches have failed and the frequency and severity of migraine attacks lead to impairment in functioning, preventive treatment is indicated.

Tension-Type Headache Treatment of tension-type headache may be either pharmacological or nonpharmacological, depending on the frequency and severity of headaches. Nonpharmacological approaches such as biofeedback, massage, relaxation, cervical traction, chiropractic manipulation, hot packs, and cold packs have all been reported to be effective, although there are no convincing guidelines to suggest one modality over another. Similar to the short-term treatment of migraine, analgesics and the NSAIDs are the first-line treatment for tension-type headache. Aspirin is the most commonly used agent, followed by ibuprofen and naproxen. It is important to note that symptoms of tension-type headache may arise iatrogenically from drugs such as ergotamine used to treat migraine, often as a result of overuse of the drug. Up to 40 percent of chronic headache cases are associated with overuse of drugs intended to alleviate headache. A mixture of butalbital (a short-acting barbiturate), aspirin, and caffeine (Fiorinal) is frequently used by nonneurologists for treatment of headache. However, tolerance and drug dependence may occur. The treatment of chronic tension-type headache is similar to that of migraine; the most commonly used drug treatment for chronic tension-type headache is the tricyclic drug amitriptyline (100 to 150 mg per day).

Cluster Headache Prophylactic medicine is almost always indicated for treating cluster headache because of the extreme severity of pain induced by an acute attack, which often occurs at night. Inhaled oxygen, narcotics, and self-injected dihydroergotamine and sumatriptan are the most commonly used agents for the treatment of acute attacks. Medications that have been shown to be effective in preventing attacks of cluster headache are lithium (Eskalith), the corticosteroids, methysergide (Sansert), the calcium channel inhibitors and valproic acid (Depakote); combinations of these agents are often necessary to achieve success.

Posttraumatic Headache Posttraumatic headache is often treated on an emergency basis but long-term posttraumatic headache is commonly encountered in psychiatry. Steroids are often given immediately subsequent to the acute injury, and diagnostic imaging will reveal the presence of skull fracture or subdural hematoma. NSAIDS are the most commonly prescribed agents for headache that persists beyond the acute injury; thereafter, the prophylactic treatment approaches to migraine can be implemented.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2, neurotransmitters in Section 1.4 and Section 1.5, and the neuropsychiatric aspects of head trauma in Section 2.5. Section 30.2 discusses behavior therapy, Section 30.6 discusses cognitive therapy, and tricyclic drugs are discussed in Section 31.30.

SECTION REFERENCES

*Block AR, Kremer EF, Fernandez E, editors: *Handbook of Pain Syndromes: Biopsychosocial Perspectives*. Lawrence Erlbaum, Mahwah, NJ, 1999.

Capobianco DJ, Cheshire WP, Campbell JK: An overview of the diagnosis and pharmacologic treatment of migraine. Mayo Clin Proc *71:*1055, 1996.

Diener HC: A review of current treatments for migraine. Eur Neurol *34:*18, 1995.

Fullerton T: Recent advances in the understanding and treatment of migraine. Pharmacy Pract *6:*53, 1993.

Gladstein J, Holden EW, Winner P, Linder S, and the Pediatric Committee of the American Association for the Study of Headache: Chronic daily headache in children and andolescents: Current status and recommendations for the future. Headache *37:*626, 1997.

Goadsby PJ, Olesen J: Increasing the options for effective migraine management. Neurology *48*(Suppl):S1, 1997.

Haas DC, Lourie H: Trauma-triggered migraine: An explanation for common neurological attacks after mild head injury. Review of the literature. J Neurosurg *68:*181, 1988.

*International Headache Society: Classification and diagnostic criteria for headache disorders, cranial neuralgias, and facial pain. Cephalalgia *8*(Suppl):1, 1988.

International Headache Society: ICD-10 guide for headaches. Cephalalgia *17*(Suppl):1, 1997.

Kearney JM, Holm JE, Kearney ML: Chronic tension-type headache: An investigation of the appraisal process. Headache *34:*351, 1994.

Kubacka RT: Practical approaches to the management of migraine. Am Pharmacol *34:*34, 1994.

*Lance JW: *Mechanisms and Management of Headache*, ed 5. Butterworth Scientific, London, 1993.

Lance J: A concept of migraine and the search for the ideal headache drug. Headache *30:*17, 1990.

Lipton RB, Stewart WF, von Korff M: Burden of migraine: Societal costs and therapeutic opportunities. Neurology *48:*S4, 1997.

Mathew N: Chronic daily headache: Clinical features and natural history. In *Headache and Depression: Serotonin Pathways as a Common Clue*, G Nappi, editor. Raven Press, New York, 1991.

Merikangas JR: Headache syndromes. In *Medical Psychiatric Practice*, vol 1, A Stoudemire, BS Fogel, editors. American Psychiatric Press, Washington, DC, 1991.

Merikangas KR: *Seminars in Headache Management: Comorbidity and Migraine.* Marcel Dekker, New York, 1996.

*Merikangas KR, Merikangas JR, Angst J: Headache syndromes and psychiatric disorders: Association and familial transmission. J Psychiat Res *27:*197, 1993.

Merikangas KR, Stevens DJ, Merikangas JR: Treatment of migraine and tension-type headache with concomitant depression. Dir Psychiatry *17:*187, 1997.

Nappi G, Facchinetti F, Rossi F, editors: Headache and menstrually related disorders: In search of a consensus. Cephalalgia *17:* 1997.

Olesen J, Tfelt-Hansen P, Welch KMA, editors: *The Headaches.* Raven Press, New York, 1993.

Raskin N: Acute and prophylactic treatment of migraine: Practical approaches and pharmacologic rationale. Neurology *43*(Suppl):39, 1993.

Reichlin S: Neuroendocrine significance of vasoactive intestinal polypeptides. In *A Decade of Neuropeptides:* Past, present, and future, G Koob, K Sandman, F Strand, editors. New York Academy of Sciences, New York, 1990.

Report of the Quality Standards Subcommittee of the American Academy of Neurology: Practice parameter: The utility of neuroimaging in the evaluation of headache in patients with normal neurological examinations (summary statement). Neurology *44:*1353, 1994.

Stewart WF, Lipton RB: Migraine headache: Epidemiology and health care utilization. Cephalalgia *13:*41, 1993.

*Stewart WF, Shechter A, Rasmussen BK: Migraine prevalence: A review of population-based studies. Neurology *44*(Suppl):17, 1994.

Stewart-Johnson E, Tfelt-Hansen P: Non-steroidal anti-inflammatory drugs. In *The Headaches*, Olesen J, Tfelt-Hansen P, Welch KMA, editors. Raven, New York, 1993.

Tfelt-Hansen P: Efficacy of b-blockers in migraine: A critical review. Cephalalgia *6:*15, 1986.

*Welch KMA: Drug therapy of migraine. N Engl J Med *329:*1476, 1993.

2.11 NEUROPSYCHIATRIC ASPECTS OF NEUROMUSCULAR DISEASE

JAMES C. EDMONDSON, M.D., PH.D.

The sensory fibers of the peripheral nervous system transmit somatic and visceral sensory information from all parts of the body to the brain. The extraction of relevant sensory features from the totality of sensory experience begins in the peripheral sensory organs and is refined at all levels of the sensory pathway, most importantly in the brain. The peripheral somatic sensory fibers can be viewed as accurate point-to-point transmitters of data from the skin to the spinal cord, and the visceral sensory fibers as reporters of less precise localizing ability. Under ideal circumstances the peripheral fibers color the perception of touch as little as possible and therefore afford the central nervous system (CNS) as faithful an internal image of the external world as possible. However, peripheral fibers are susceptible to physical compression and metabolic insult, which can compromise the quality of their data transmission. For example, diabetes can slowly eliminate peripheral sensory fibers and reduce sensory acuity, and can cause aberrant sensations to be transmitted to the brain. How the CNS handles inaccurate sensory information is complex and is influenced by a person's psychological state. Depressed mood and anxiety tend to intensify unpleasant somatic sensations. Conversely, persons with somatoform disorders may perceive sensations that do not correlate with physical reality. This discussion defines the function of the peripheral nervous system and of its principal disease states, and describes their relationship to conscious experience.

ANATOMY

The peripheral nervous system contains motor and sensory fibers. Motor fibers conduct action potentials from the CNS to the muscles, where they trigger the contraction of muscle fibers. Sensory fibers carry information from the peripheral receptors to the CNS. In cross-section a peripheral nerve consists of fibers of various diameter; some are wrapped in a myelin sheath and others are unmyelinated. The fibers are grouped into fascicles and are enmeshed in connective tissue layers called *endoneurium, perineurium, and epineurium.* Nerves also contain blood vessels and small amounts of adipose tissue (Fig. 2.11-1). The functional unit is the nerve axon and its myelin sheath (Fig. 2.11-2). Without myelin, an action potential can be propagated along the nerve fiber at a rate of only 1 meter per second, whereas large myelinated fibers can conduct impulses at a rate of up to 70 meters per second. Peripheral nerve disease can affect the myelin, in which case conduction velocity is slowed, or it can affect the axon, in which case the amplitude of the action potential is reduced, or both.

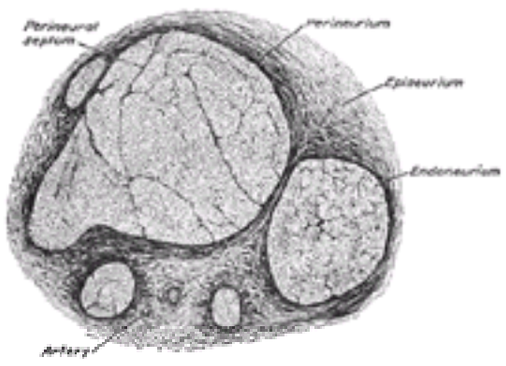

FIGURE 2.11-1 Cross-section of the sciatic nerve. Fascicles of myelinated and unmyelinated nerve fibers are enmeshed in endoneurium and are separated from other fascicles by perineurium. The outer layer of connective tissue is the epineurium, in which are embedded blood vessels and adipose tissue. (Reprinted with permission from Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.)

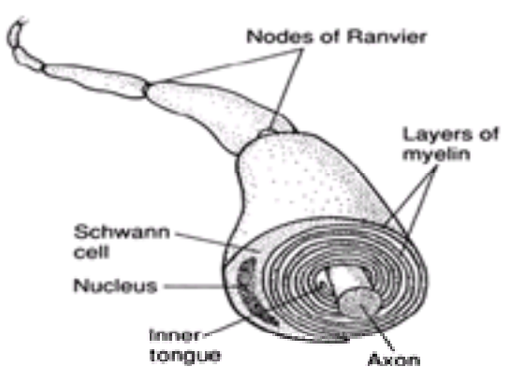

FIGURE 2.11-2 Representation of the axon and its myelin sheath. The inner tongue of Schwann cells wraps layers of hydrophobic myelin around the axon. Individual Schwann cells are separated about every 1 mm by a bare section of axon called the node of Ranvier. Action potentials jump from one node to the next, which increases the rate of propagation of the action potential. (Reprinted with permission from Kandel ER: Nerve cells and behavior. In *Principles of Neural Science*, ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Elsevier, New York, 1991. © 1996 by Appleton & Lange).

Each peripheral nerve corresponds to one or a few segments of the spinal cord. There are eight cervical spinal segments, twelve thoracic spinal segments, five lumbar spinal segments, and five sacral spinal segments. The skin monitored by nerves that enter a particular spinal segment is called a *dermatome*, and the muscles innervated by nerves exiting a particular spinal segment constitute a *myotome.* The posterior pattern of sensory dermatomes is shown in Figure 2.11-3. Once nerve roots in the cervical and lumbar regions leave the spinal canal, they converge and diverge in a plexus and emerge as the named peripheral nerves. The organization of the brachial plexus is shown in Figure 2.11-4, and a comparison of the innervation of the skin according to dermatome or named nerve is shown in Figure 2.11-5. It is usually possible to localize a lesion in the peripheral nervous system quite accurately by physical examination, noting the particular combination of sensory and motor defects produced by the lesion.

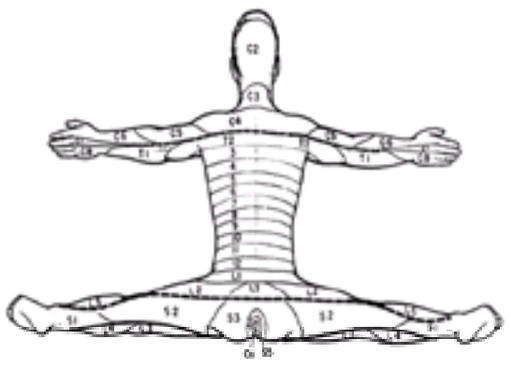

FIGURE 2.11-3 Posterior distribution of dermatomes. The arms expand the cervical dermatomes and the legs expand the lumbosacral dermatomes. (Reprinted with permission from Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.)

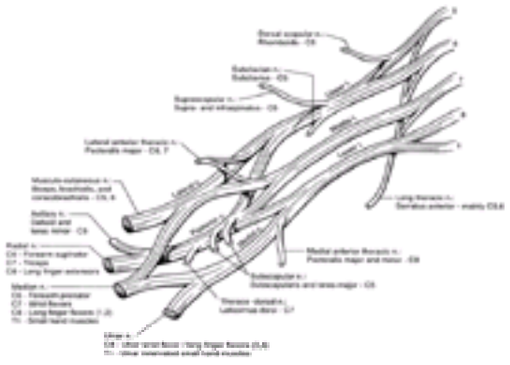

FIGURE 2.11-4 The brachial plexus. The cervical spinal roots are at the upper right. They first fuse to form three trunks *(t)*. Each trunk splits into an anterior and a posterior division, and these then fuse to form three cords *(c)*. The main named nerves of the arm *(n)* emerge from various segments of the brachial plexus. Each muscle is therefore innervated by a named nerve carrying fibers from a particular set of spinal roots. (Reprinted with permission from Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.)

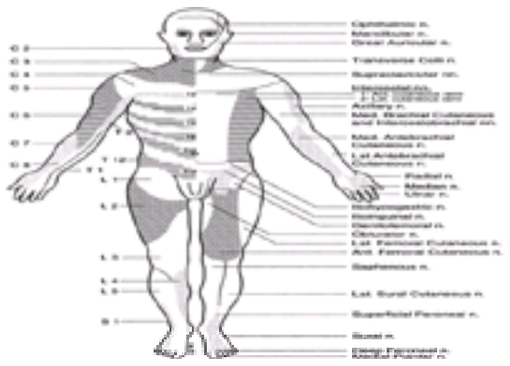

FIGURE 2.11-5 Anterior distribution of the dermatomes **(left)** and cutaneous areas supplied by named peripheral nerves **(right)**. Dermatomal patches overlap considerably with their neighboring dermatomes, whereas the distributions of peripheral nerves are sharply demarcated on the skin. (Reprinted with permission from Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.)

Motor fibers terminate on muscle fibers forming the neuromuscular junction. At the nerve terminus the action potential triggers the influx of calcium ions, which cause the release of acetylcholine from vesicular stores. Acetylcholine enters the synaptic cleft and binds to receptors on the muscle cell that trigger release of intracellular calcium and initiate contraction of the muscle fiber (Fig. 2.11-6). Acetylcholine is degraded and inactivated in the synaptic cleft by acetylcholinesterase, and the resulting acetic acid and choline are taken up by the nerve terminus and recombined into acetylcholine for storage in vesicles. Neuromuscular transmission is usually highly efficient, but it may be compromised in myasthenia gravis and other disorders of neuromuscular transmission.

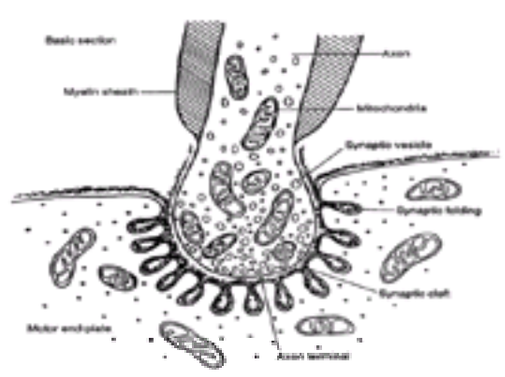

FIGURE 2.11-6 Cross-section of the neuromuscular junction. The nerve fiber terminus enters from the top and releases acetylcholine from synaptic vesicles. Acetylcholine diffuses through the matrix of the synaptic cleft to activate muscle contraction. (Reprinted with permission from Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.)

Each motor nerve branches within the muscle and innervates a handful of muscle fibers. If a nerve ending dies, the resulting denervated muscle fibers initially exhibit spontaneous activity, then eventually are reinnervated by collateral sprouts of neighboring nerves; the resulting motor units cover an unusually large area (Fig. 2.11-7).

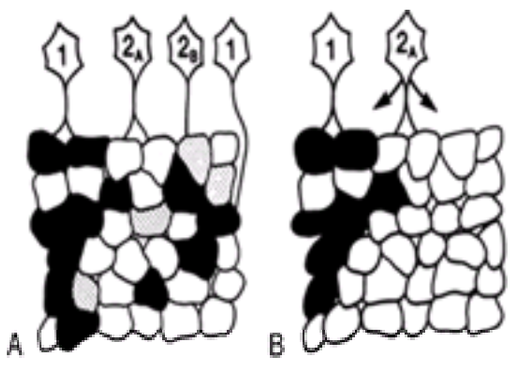

FIGURE 2.11-7 A, Normal muscle fibers have a mosaic pattern of innervation: in this figure, four individual nerve endings each branch to innervate a small subset of muscle fibers. A contiguous set of fibers constitutes a motor unit. Fibers are either type 1 (*black*), type 2_A (*white*), or type 2_B (*stippled*). The specific type of a given muscle fiber is determined by the nerve fiber that innervates it. The muscle fiber type may change if the fiber is re-innervated by a nerve fiber of a new specificity. **B**, After loss of nerve fibers, remaining nerve fibers sprout to create unusually large motor units. (Reprinted with permission from Poirier J, Gray F, Escourolle R: *Manual of Basic Neuropathology*, ed 3. Saunders, Philadelphia, 1990.)

PHYSIOLOGY

Nerve impulses are called *action potentials.* In the resting state the intracellular compartment of the neuron is negatively charged at a potential of –70 to –80 mV, but during an action potential the inside of the neuron becomes positively charged. For an action potential to be generated by a neuron, ion channels open and sodium ions begin to enter the cell, gradually making the inside of the neuron less negatively charged relative to the outside. The point at which the interior of the neuron is sufficiently less negatively charged to initiate an action potential is called the *spike threshold* and is characteristically approximately –60 mV. The action potential itself is a brief (0.1 to 2 msec) wave of reversal of membrane potential that moves along an axon (Fig. 2.11-8). During an action potential the interior of the neuron is positively charged in comparison with the outside of the neuron. The initial ion channel involved in the action potential is the Na+ channel, which, when opened, allows positively charged sodium ions to enter the neuron. The Ca_{2+} channels open next, allowing the positively charged calcium ions to enter the neuron and further contribute to the spike of the action potential. Entry of the calcium ion activates ion channels that carry an outgoing flow of potassium ions that are involved in arresting the action potential. The activation of these K+ channels results in the afterhyperpolarization of the neuron after an action potential. During the afterhyperpolarization the inside of the neuron is even more negatively charged than it was at baseline. The afterhyperpolarization contributes to the refractory period of a neuron after an action potential; during this period, another action potential cannot be generated.

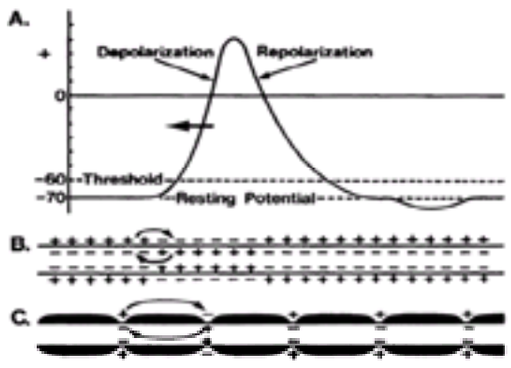

FIGURE 2.11-8 Phases of the action potential during its propagation. **A**, Changes in the membrane potential as the action potential moves from right to left. **B**, In unmyelinated fibers, an action potential triggers changes in the immediately adjacent membrane, called local circuit conduction. **C**, In myelinated fibers, the action potential jumps from one node of Ranvier to the next, called saltatory conduction. (Reprinted with permission from Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.)

ELECTRODIAGNOSIS

Nerve Conduction Studies In clinical neurophysiological testing, a nerve action potential can be initiated by application of a brief electrical stimulus to the skin directly over the nerve, and the action potential may be recorded with surface electrodes. Thus, by stimulating the nerve at one point and recording the action potential at another point a defined distance away, a nerve conduction velocity can be calculated. Extensive studies have established reference values for each of the measurable nerves against which patient values can be compared. In the case of motor nerves, recording is usually made over the muscle, which has a much larger action potential amplitude than the nerve. In this case the time interval between the nerve stimulus and the muscle action potential includes both the time of conduction of the nerve action potential and the time of neuromuscular transmission. This interval is called the *latency.* If the point of stimulation is near the muscle, a distal latency is recorded, and if the point of stimulation is closer to the spinal cord, a proximal latency is recorded. The distal latency can be subtracted from the proximal latency to eliminate the factor of the time of neuromuscular transmission and isolate the nerve conduction velocity (NCV) (Fig. 2.11-9). Slowing of nerve conduction can be either focal, multifocal, or diffuse.

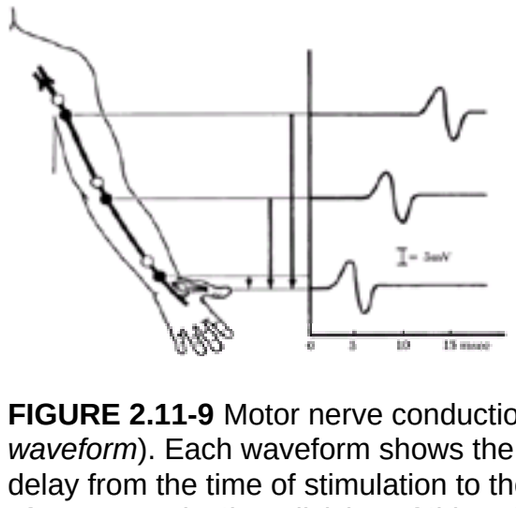

FIGURE 2.11-9 Motor nerve conduction study in the median nerve. The nerve is stimulated at the axilla (*top waveform*), elbow (*middle waveform*), and wrist (*bottom waveform*). Each waveform shows the muscle action potential recorded from the abductor pollicis brevis muscle as a function of time following nerve stimulation. The delay from the time of stimulation to the onset of the muscle action potential is the latency. Subtraction of the proximal latency from the distal latency isolates the time of nerve conduction; division of this value by the distance between the two points of stimulation gives the nerve conduction velocity. (Reprinted with permission from Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.)

Electromyography The activity of individual motor units—the muscle fibers innervated by a single nerve ending—can be recorded with an electrode placed directly into the muscle. The morphology, amplitude, and duration of the muscle action potential are noted and compared to normal values. After nerve injury the first abnormal activity, noted at 2 to 3 weeks, is spontaneous firing of denervated muscle fibers. Over the following weeks to months, nerve endings sprout and reinnervate these muscle fibers, and spontaneous activity is replaced by motor unit potentials of very large amplitude and duration. In contrast, injury to muscle fibers themselves

produces brief, low-amplitude motor potentials (Fig. 2.11-10). Examination of 20 or more individual muscle fibers in a specific muscle can reveal whether there is an injury to the innervating nerve fiber at some point along its pathway, even if the nerve conduction velocity is normal. Electromyography therefore complements the nerve conduction velocity studies, and the combination of the two techniques usually permits a more exact localization and quantification of neuromuscular disease than is possible with physical examination alone.

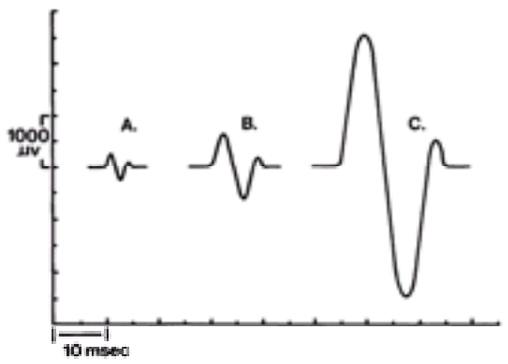

FIGURE 2.11-10 Motor unit potentials. **A**, Small amplitude and duration motor unit potential in myopathy. **B**, Normal motor unit potential. **C**,. Large amplitude and duration motor unit in neuropathy as a result of denervation, collateral sprouting, and reinnervation. (Reprinted with permission from Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.)

COMMON NERVE COMPRESSION SYNDROMES

Carpal Tunnel Syndrome The median nerve travels from the distal arm into the hand through the carpal tunnel, which is a fascia-lined passage in the ventral wrist. Compression of the nerve at the carpal tunnel produces a characteristic set of symptoms and signs in the hands and arms. Carpal tunnel syndrome is more common among people who perform repetitive hand movements, such as typing, sorting mail, playing a keyboard instrument, or lifting heavy weights, but it may also occur in the absence of such repetitive stresses. The classic symptoms include pain, numbness, and tingling in the hands, which are worse in the morning. Patients with carpal tunnel syndrome may complain that items suddenly fall out of their hands. The complaints may extend up to the ipsilateral shoulder, neck, or side of the head, or down the leg. Some patients may be only vaguely aware of tingling somewhere in their body, yet may be very unsettled by the sensations. Physical examination may show decreased light touch in the median nerve distribution in the hand, weakness of the median-innervated abductor pollicis brevis muscle, positive Tinel's sign (tingling in the fingers caused by percussion over the carpal tunnel), or positive Phalen's sign (reproduction of the symptoms of carpal tunnel syndrome by sustained palmar flexion of the wrist). Nerve conduction studies first reveal slowing and loss of amplitude of the sensory fibers, then, in more advanced cases, slowing and occasionally loss of amplitude of the motor fibers. Electromyography of the APB muscle may yield normal results or may show signs of acute or chronic denervation. Conservative treatment consists of wearing splints that prevent palmar flexion of the wrist, especially at night. Antidepressant drugs may reduce or eliminate subjective tingling. If the symptoms do not begin to improve after 3 months of wearing splints, surgical release of the trapped median nerve is beneficial in about 80 percent of cases, although the symptoms may recur months to years later.

Compression of the Ulnar Nerve at the Elbow (Cubital Tunnel Syndrome) The ulnar nerve is prone to compression in the ulnar groove between the olecranon process of the ulna and the medial epicondyle of the humerus. The clinical symptoms and signs of ulnar neuropathy at the elbow are generally less well defined than in carpal tunnel syndrome, ranging from asymptomatic to sensory loss and rarely weakness in the ulnar-innervated pattern distal to the elbow. Patients may complain of tingling and numbness throughout the arm. Nerve conduction studies show a significant slowing of the ulnar nerve conduction velocity in the segment passing through the ulnar groove, either (1) as an isolated finding or (2) in association with carpal tunnel syndrome or a more diffuse peripheral neuropathy. There is no practical way to immobilize the elbow, but patients should be advised not to lean on their elbows. The tingling and numbness may be reduced with antidepressant medications. Surgical transposition of the ulnar nerve is successful in severe cases, but the rate of surgical failures exceeds 50 percent in some series.

Peroneal Palsy The peroneal nerve is vulnerable to compression as it passes over the head of the fibula, just distal and lateral to the knee. It may be injured because of pressure of the opposite leg when the legs are crossed above the knee, during improper positioning during surgery, or during immobilization of the leg in a cast. Sensory loss extends down the anterolateral surface of the leg and foot, and there may be complete or partial weakness of foot elevation and eversion. Nerve conduction studies show slowing or loss of conduction across the fibular head, and electromyography may show evidence of denervation of the peroneal-innervated muscles. Patients with peroneal palsy should wear an ankle-foot orthosis to prevent the toes from dragging as the leg swings forward. Recovery may take as long as 2 years, and is often incomplete. Antidepressant medications often reduce neuropathic pain significantly, whereas nonsteroidal anti-inflammatory agents are generally of little benefit. Physical therapy can assist the patient to adapt to the difficulties of ambulation.

Meralgia Paresthetica The lateral cutaneous nerve of the thigh passes under the inguinal ligament just before it reaches the skin, where it may be compressed by prolonged wearing of a heavy belt, such as a carpenter's belt. Compression may also arise spontaneously, particularly in obese individuals. Meralgia paresthetica—literally ‘tingling pain in the thigh'—describes the resulting sensations in a sharply delineated patch of skin on the lateral surface of the thigh extending down to the knee. Management begins with removal of the heavy belt or loss of weight. Antidepressant medications offer relief in most cases. Gabapentin (Neurontin) has also been used.

Of interest to psychiatrists is that Sigmund Freud suffered from meralgia paresthetica at age 39. He published a paper describing his experience with the disorder in which he described his symptoms. He attributed its cause to his habit of taking cold showers daily.

Radiculopathy Commonly described as a “pinched nerve in the neck or back,” true spinal radiculopathies are in fact a relatively rare cause of neck or back pain. The sensory or motor roots (or both) may be compressed by a herniated intervetebral disk or as they leave the vertebral column by bony prominences formed by degenerative changes. Sensory loss is in a dermatomal distribution, with vague borders caused by overlapping of neighboring dermatomes. Weakness occurs in a myotomal pattern, and there may be loss of characteristic deep tendon reflexes. Nerve conduction studies are usually normal, and electromyography may show signs of denervation and reinnervation in the first 3 to 4 months of compression. Neuroimaging, either magnetic resonance imaging or myelography with computed tomography, is essential. Conservative management is always indicated first, consisting of rest, muscle relaxers such as benzodiazepines, effective analgesia, heating pads, and nonstrenuous exercises. If conservative management fails to relieve the pain, and neurological deficits persist, referral to a neurosurgeon is indicated. In carefully selected cases where a nerve root is clearly compressed, surgical removal of the offending structure can provide definitive relief. It should be noted, however, that arthritic pain and muscle pain are far more common causes of low back pain than are true radiculopathies. According to the physiatrist John Sarno, 95 percent of low back pain is psychosomatic. The first approach to such uncomplicated low back pain is conservative management. Where surgery is not appropriate, several techniques have been developed for injection of steroids or local anesthetics in the epidural space, but the pain relief afforded by these methods is usually temporary, lasting sometimes just for a few hours and only for weeks to months in the most favorable of situations. Patients who develop chronic low back pain in the absence of demonstrable radiculopathy may benefit from weight loss, physical therapy, psychiatric support, and antidepressant medications.

ACQUIRED NEUROPATHIES

Peripheral nerves are vulnerable to several metabolic and infectious conditions. Prevention is the best approach to acquired neuropathies, as the ability of the nerves to regenerate is best in cases of trauma and worst in cases of metabolic injury.

Diabetes Mellitus Diabetes promotes thrombus generation in the tiniest capillaries, and it slowly eliminates peripheral nerve endings over several years. About 15 percent of patients with diabetes have symptoms and signs of peripheral neuropathy, but 50 percent have either neuropathic symptoms or slowing of nerve conduction velocity. The longest nerves are the most vulnerable, and symptoms usually begin in the feet. Several clinical forms are seen: (1) a distal, symmetrical, predominantly sensory painful neuropathy; (2) autonomic neuropathy; (3) a painful, asymmetrical multiple neuropathy, also called *mononeuropathy multiplex*; (4) acute mononeuropathies; and (5) diabetic ophthalmoplegia. The burning pain of the sensory neuropathy may respond to antidepressant or anticonvulsant medications. Tight control of plasma blood sugar can slow the progression of the neuropathy and reduce the symptoms.

Other Acquired Metabolic Neuropathies Several medical conditions producing diffuse damage to peripheral nerves include vitamin deficiency states, heavy metal poisoning, drug intoxications (including chronic use of lithium), uremia, collagen-vascular disorders, and paraneoplastic syndromes. The principal neuropathic syndromes are listed in Table 2.11-1.

Table 2.11-1 The Principal Neuropathic Syndromes

Postherpetic Neuralgia Herpes zoster can infect and remain dormant in dorsal root ganglion cells for as long as decades, and then may be reactivated under conditions of emotional stress or immune compromise. Reactivation of herpes zoster causes several days of burning, itching, and tingling, followed by the emergence of very painful blisters in a dermatomal distribution. These resolve in 1 to 4 weeks in most patients, but up to one third of patients may suffer pain for months and sometimes for years. The antiviral agents acyclovir (Zovirax), famciclovir (Famvir), or valacyclovir hydrochloride (Valtrex) may shorten the acute course. Antidepressant medications are most effective for chronic postherpetic neuralgia, and carbamazepine (Tegretol) may be useful for lancinating pains.

Guillain-Barré Syndrome *Acute inflammatory demyelinating polyneuropathy*, or *Guillain-Barré syndrome*, is a postinfectious, autoimmune condition in which antibodies elicited by an infectious agent cross-react with the myelin sheaths of the peripheral nerves. Inflammatory loss of myelin results in slowing and blockade of nerve conduction. Initially there is tingling in the toes, followed by loss of deep tendon reflexes and motor weakness affecting the longest nerves first. The weakness thus begins in the toes and appears to ascend over the course of a few days. Once a patient cannot walk, he or she should be admitted to an intensive care unit, closely monitored for respiratory distress, and mechanically ventilated if necessary. Recovery usually begins by the fourth week and may not be complete for months. Intensive inpatient rehabilitation may be necessary and the patient may need continuous psychiatric support during the recovery. Tingling, pain, and numbness may be treated with antidepressant medications.

Autonomic Neuropathies Hereditary and acquired damage to the autonomic nervous system may produce fluctuations in blood pressure, heart rate, temperature control, bowel and bladder function, or sexual function, all of which may resemble the effects of emotional instability. Patients may develop orthostatic hypotension upon rising from a seated or lying position and complain of persistent dizziness. Blood pressure may rise excessively in response to emotional stress, such as visits to the doctor (“white coat syndrome”). Patients with these symptoms may be considered to have hysteria, anorexia nervosa, or hypochondriasis. Careful medical evaluation of somatic and visceral complaints, including autonomic electrophysiological evaluation, should be done before a strictly psychiatric condition is diagnosed. Electrical evidence of autonomic neuropathy includes decreased beat-to-beat variability of the heart rate and absence of a sympathetic skin response.

CORRELATION OF COMPLAINTS AND ELECTRODIAGNOSTIC FINDINGS

Patients are referred for electrodiagnostic evaluation because their complaints of pain, numbness, tingling, and weakness are suspected to be caused by injury to the peripheral nerves. The most common referral is to determine whether a patient's signs and symptoms of CTS, which is a clinical diagnosis, correlate with electrodiagnostic evidence of nerve injury. The majority of patients with convincing signs of CTS on examination will have at least some electrodiagnostic abnormalities, usually prolongation of the distal sensory latency and reduction of the sensory amplitude. Some patients who are asymptomatic or barely symptomatic will nevertheless have very advanced sensory and motor nerve injury on electrodiagnostic studies. Conversely, some patients with prominent complaints and classical signs of CTS will have electrodiagnostic values completely within the normal range. Questioning such patients often reveals high levels of anxiety and depression, which appear to amplify the symptoms of even extremely minor nerve compression. Such patients often respond very satisfactorily to antidepressant medications.

MANAGEMENT

Diagnostic Workup History should emphasize the location and quality of the pain, whether it is constant or intermittent, how long it has been present, and whether it is improving, remaining steady, or worsening. Past medical history should cover diabetes, thyroid disease, autoimmune disease, cancer, and infections, and a thorough history of medication use. The physical examination should test muscle bulk and strength, all sensory modalities (light touch, pain, temperature, vibration, and position sense), and deep tendon reflexes, with particular attention to patterns of abnormalities that define a dermatome or the distribution of a particular named nerve. Electrodiagnostic studies can corroborate the clinical suspicion of neuropathy and quantitate the abnormalities. If a metabolic disorder is suspected, appropriate blood and urine tests should be ordered. If a radiculopathy that may be amenable to surgery is suspected, then high–field-strength MRI scanning of the spine is indicated.

Medical Treatments Any underlying medical condition such as diabetes, vitamin B_{12} deficiency, or thyroid disease must be vigorously treated. In some cases the neuropathic symptoms and signs will resolve upon adequate management of the underlying medical problem.

Surgical Options Surgical release of entrapped nerves is most effective for carpal tunnel syndrome and certain radiculopathies, and is less uniformly effective for ulnar neuropathies. Nerve ablation has been of benefit in trigeminal neuralgia, but should be avoided for other chronic pain syndromes. Although nerve ablation may temporarily cut off sensation to a painful area, secondary sensory routes often emerge and painful sensations therefore return.

Adjunctive Medications Regardless of the cause of neuropathic symptoms, the distressing pain, numbness, and tingling can be reduced or eliminated by antidepressant medications. Neurologists have traditionally relied on low doses of tricyclic drugs, especially amitriptyline (Elavil) and nortriptyline (Pamelor). Some studies have suggested that norepinephrine reuptake inhibition is more effective for pain control than is serotonin reuptake inhibition, but other data show no therapeutic advantage of tricyclic drugs over the newer antidepressant medications. The selective serotonin specific reuptake inhibitors (e.g., fluoxetine [Prozac], sertraline [Zoloft], paroxetine [Paxil], and fluvoxamine [Luvox], bupropion (Wellbutrin), nefazodone (Serzone), trazodone (Desyrel), venlafaxine (Effexor), and mirtazapine (Remeron) are probably all equally effective for neuropathic symptoms. Low dosages usually suffice, but higher dosages may be needed. Lancinating pain, such as that seen in some types of postherpetic neuralgia, may respond to the anticonvulsants carbamazepine (Tegretol) or phenytoin (Dilantin). A few uncontrolled studies have reported that gabapentin (Neurontin) reduces neuropathic pain, but controlled trials are needed to establish a role for gabapentin in the management of neuropathic pain. Dextromethorphan has been shown in one study to provide some relief of diabetic painful neuropathy. Psychotherapy for underlying depression and anxiety may be beneficial. Vitamin therapies have been widely discussed, particularly vitamins B_6 and B_{12}, but true vitamin deficiency states are quite rare in developed countries; massive quantities of vitamin B_6 can cause a sensory neuropathy. Alternative therapies, such as herbal medicines, meditation, or acupuncture, work to the extent that they elicit the placebo effect, which may be a powerful therapeutic modality in suggestible individuals.

SUGGESTED CROSS-REFERENCES

Chronic pain is discussed in Section 28.6. Psychosomatic disorders are covered in Chapter 25. Drugs used in psychiatry (including antidepressants and benzodiazepines) are discussed and organized pharmacologically in Chapter 31.

SECTION REFERENCES

*Adams RD, Victor M, Ropper AH: *Principles of Neurology*, ed 6. McGraw-Hill, New York, 1997.

Baron R, Wasner G, Borgstedt R, Hastedt E, Schulte H, Binder A, Kopper F: Effect of sympathetic activity on capsaicin-evoked pain, hyperalgesia, and vasodilation. Neurology *52:*923, 1999.

*Block AR, Kremer EF, Fernandez E, editors: *Handbook of Pain Syndromes: Biopsychosocial Perspectives*. Lawrence Erlbaum, Mahwah, NJ, 1999.

Bromm B, Desmedt JE: Pain and the brain: From nociception to cognition. In *Advances in Pain Research and Therapy*, vol 22. Raven Press, New York, 1995.

Fields H: *Pharmacological Approaches to the Treatment of Chronic Pain.* IASP Publications, Seattle, 1994.

*Kandel ER, Schwartz JH, Jessell TM: *Principles of Neural Science*, ed 3. Elsevier, New York, 1991.

Kimura J: *Electrodiagnosis in Diseases of Nerve and Muscle*, ed 2. FA Davis, Philadelphia, 1989.

*Kruger L: *Handbook of Perception and Cognition*, ed 2. Academic Press, San Diego, 1996.

Max MB: Treatment of post-herpetic neuralgia: Antidepressants. Workshop of the VZV Research Foundation, Inc. et al: Varicella-zoster virus infection: New insights into pathogenesis and postherpetic neuralgia. Ann Neurol *35*(Suppl):S50, 1994.

Max MB, Gilron I: Sympathetically maintained pain: Has the emperor no clothes? Neurology *52:*905, 1999.

Oh SJ: *Clinical Electromyography: Nerve Conduction Studies*, ed 2. Williams & Wilkins, Baltimore, 1993.

*Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.

Poirier J, Gray F, Escourolle R: *Manual of Basic Neuropathology*, ed 3. WB Saunders, Philadelphia, 1990.

Sindrup SH, Gram LF, Brosen K, Eshoj O: The selective serotonin reuptake inhibitor paroxetine is effective in the treatment of diabetic neuropathy symptoms. Pain *42:*135, 1990.

2.12 NEUROPSYCHIATRIC ASPECTS OF CHILD NEUROLOGY

JAMES C. EDMONDSON, M.D., PH.D.

Neurological disorders during infancy, childhood, and adolescence may adversely affect the acquisition of cognitive and emotional milestones. Developmental delays may profoundly influence the psychological state of both the affected individuals and their caregivers. Many of the neurological conditions discussed in earlier sections of this chapter appear in infantile and childhood forms. In addition, a large number of rare genetic disorders appearing only in infants and children have specific effects on behavior. A neuroscientific model is evolving, in which the brain's synaptic connections are dramatically molded by environmental influences during developmental windows of time in the first years of life. A growing body of investigations has confirmed that an individual's lifelong emotional repertoire reflects an internal representation of the reactions of caregivers in the first years of life, in the form of specific synaptic networks that remain relatively stable after the developmental windows have closed. The diagnosis and treatment of biochemical and physiological disorders of the developing nervous-system—the province of child neurology—therefore has begun to yield insights into the biological basis of behavior.

The clinical neurological approach to infants and children includes localization and identification of the lesion, including determining whether a disorder is genetic or acquired. The diagnosis is synthesized from history, physical examination, neuroimaging, and laboratory tests. Many of the rules of neurological localization that have been finely determined for adults apply with equal accuracy to children, but the developing brain is generally capable of greater plasticity. For example, children may recover more rapidly and more completely from a stroke involving a specific region of the cortex, because the developing brain can more readily redistribute mental functions to unaffected regions of the brain.

PRINCIPLES OF BRAIN DEVELOPMENT

The adult brain consists of approximately 10_{11} neurons, each of which receives on average 10_3 to 10_4 synapses from 10_3 other neurons. The study of the development of the nervous system is an actively evolving body of knowledge. In the brain, the life cycle of a neuron consists of cell birth in circumscribed proliferative zones, migration to the adult position, extension of an axon, elaboration of dendrites, synaptogenesis, and finally the onset of chemical neurotransmission. At the peak of neuronal proliferation in the middle of the second trimester of gestation, 250,000 neurons are born each minute. Postmitotic neurons migrate to their adult locations in the cortex, guided by glial fibers (Fig. 2.12-1). Glial-guided neuronal migration in the cerebral cortex occupies much of the first 6 months of gestation. In the cerebellum, glial-guided neuronal migration continues through the second postnatal year. Some neurons in the prefrontal cortex migrate over a distance 5000 times the diameter of the neuronal cell body. Radial migration forms cortical columns that function as a unit in processing tasks (Fig. 2.12-2). Neuronal migration requires a complex set of cell-cell interactions, and it is susceptible to errors, in which neurons fail to reach the cortex and instead reside in ectopic positions. A group of such incorrectly placed neurons is called a heterotopia (Fig. 2.12-3). Neuronal heterotopias cause epilepsy and are highly associated with mental retardation. In a classical neuropathological study of four patients with dyslexia, for example, heterotopias were common. Heterotopias in the frontal lobe have been postulated to play a causal role in some cases of schizophrenia. Cortical neuronal migration abnormalities have been found in some people with autism, smaller cortical structures in people with Down syndrome, larger basal ganglia in people with schizophrenia, hypoplastic basal ganglia in people with Tourette's disorder, and hypoplastic cerebellar structures in numerous developmental disorders. Normal cerebral asymmetries appear to be disrupted in a number of disorders, including schizophrenia, Tourette's disorder, attention-deficit disorders, and reading disorder. Table 2.12-1 relates developmental events to severe malformations.

Days of Gestation	Event	Effect of Toxic Stimulus
0-18	Three germ layers elaborate and early neural plate forms	No effect or death
18	Neural plate and groove develop	Anterior midline defects (18–25 days)
22–23	Optic vesicles appear	"Induction" hydrocephalus (18–40 days)
24–26	Anterior neuropore closed	Anencephaly (after 25 days to 5)
26–28	Posterior neuropore closed; ventral horns form	Cranium bifidum, spina bifida cystica, spina bifida occulta (after 26 days to ?)
28–32	Anterior and posterior motor roots form	
32	Cerebellar primordium; vascular circulation	Microcephaly (30–130 days); cellular proliferation anomalies (30–175 days); migration anomalies (50 days to completion of development of each brain subdivision)
33–35	Prosencephalon cleaves to form paired telencephalon, four cerebral vesicles, choroid plexi, dorsal root ganglion development	Holoprosencephaly
41	Region of olfactory bulb appears in forebrain	Arhinencephaly
56	Differentiation of cerebral cortex, meninges, ventricular system, CSF circulation	Dandy-Walker syndrome
70–100	Corpus callosum	Agenesis of corpus callosum
70–150	Primary fissures of cerebral cortex; spinal cord ends at L3 level	Lissencephaly; pachygyria
140–175	Neuronal proliferation in cerebral cortex ends	Defects of cellular architectonics, myelin defects (175 days to 4 years postnatally)
7–9 months	Secondary and tertiary sulci	Destructive pathological changes first noted
175 days to 6 years postnatally	Neuronal migrations; glial cell proliferation; myelin formation; asymmetric and unidendritic synaptic connections; spinal cord ends L1–L2 level	

(Reprinted with permission from Menkes J (Textbook of Child Neurology, ed 5. Williams & Wilkins, Baltimore, 1995)

Table 2.12-1 Timetable of Human Central Nervous System Ontogenesis

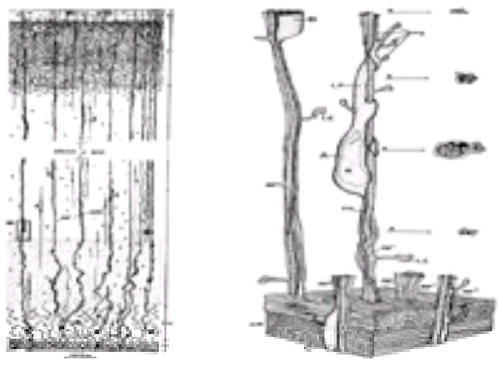

FIGURE 2.12-1 Left, Diagram of the cerebral cortex at midgestation. Radially oriented glial fibers guide the migration of neurons from the proliferative zones to the cortical plate. The rectangle marked with an asterisk shows a migrating neuron that is shown enlarged at right. *C*, cortical plate; *I*, intermediate zone; *M*, molecular layer; *MN*, migrating neuron; *RF*, radial fiber; *SV*, subventricular zone; *V*, ventricular zone. **Right**, Enlarged view of neurons migrating along glial fibers. A leading process (*LP*) precedes the nucleus as the neuron inches its way up the fiber, laying down a trailing process (*TP*). *A*, *B*, *C*, migrating neurons; *LE*, lamellate expansion; *OR*, optic radiations; *PS*, pseudopodia; *RF*, radial fiber. (Reprinted with permission from Rakic P: Mode of all migration to the superficial layers of fetal monkey neocortex. J Comp Neurol *145:* 61, 1972.)

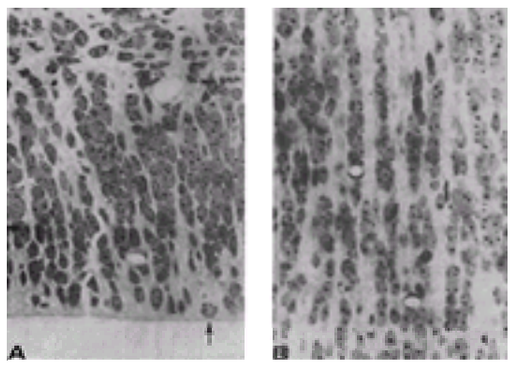

FIGURE 2.12-2 Radial organization of cortical columns. **A**. Ventricular proliferative zone, where neurons are born and begin their migration. **B**. Cortical plate. Each column responds to a specific stimulus as a functional unit. (Reprinted with permission from Rakic P: Specification of cerebral cortical areas. Science *241:*170, 1988.)

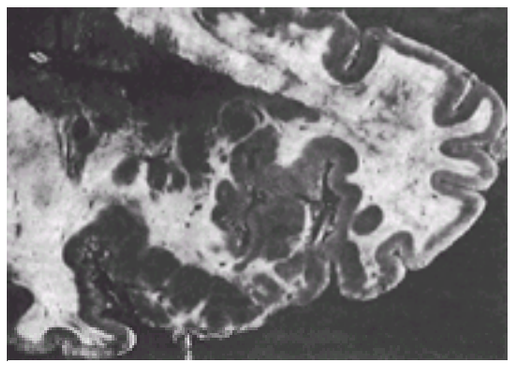

FIGURE 2.12-3 Heterotopic neurons. In addition to the neurons of cortex, visible along the gyri and sulci, large nests of neurons are located aberrantly along the ventricular wall. Heterotopic neurons may generate seizure activity. (Reprinted with permission from Layton DD: Heterotopic cerebral gray matter as an epileptogenic focus. J Neuropathol Exp Neurol *21:*244, 1962.)

Many neurons lay an axon down as they migrate; others do not initiate axon outgrowth until they have reached their cortical targets. Elaboration of a characteristic branched dendritic tree occurs once the neuron has completed migration (Fig. 2.12-4). Synaptogenesis occurs at a furious rate from the second trimester through the first 10 years or so of life. The peak of synaptogenesis occurs in the first 2 postnatal years, when as many as 30 million synapses form each second. Ensheathment of axons by myelin begins prenatally, proceeds most rapidly in early childhood, and reaches its full extent late in the third decade of life (Fig. 2.12-5). Glucose utilization, protein synthesis, and cellular metabolic rates in the brain reach their highest levels in the first decade of life, as the brain forms its adult intercellular connections (Fig. 2.12-6).

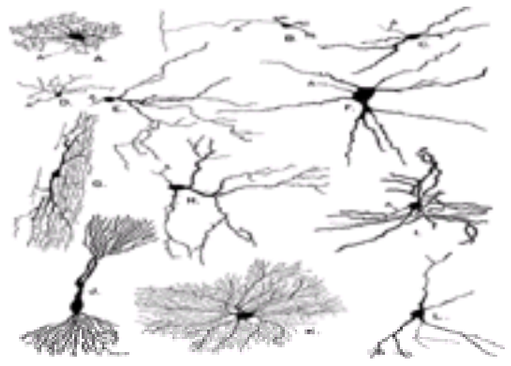

FIGURE 2.12-4 Characteristic neurons whose axons (*A*) and dendrites remain within the central nervous system. **A.** Neuron of the inferior olivary nucleus. **B.** Granule cell of the cerebellar cortex. **C.** Small cell of the reticular formation. **D.** Small gelatinosa cell of the spinal trigeminal nucleus. **E.** Ovoid cell of the nucleus tractus solitarius. **F.** Large cell of the reticular formation. **G.** Spindle-shaped cell of the substantia gelatinosa of the spinal cord. **H.** Large cell of the spinal trigeminal nucleus. **I.** Neuron of the putamen. **J.** Double pyramidal cell in Ammon's horn of the hippocampal formation. **K.** Cell from the thalamus. **L.** Cell from the globus pallidus. (Reprinted with permission from Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.)

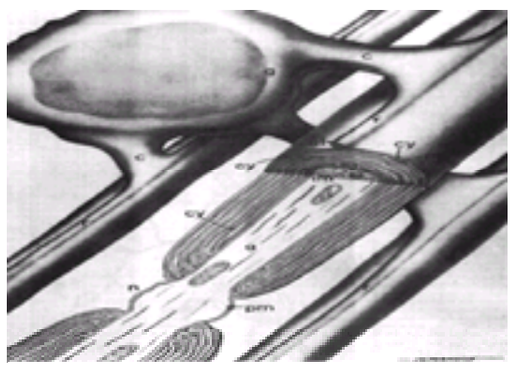

FIGURE 2.12-5 Relationship of the oligodendrocyte (*g*) and the central myelin sheath to the axon (*a*). (*c*, cytoplasmic process; *cy*, glial cell cytoplasm trapped among the layers of myelin; *im*, inner mesaxon; *n*, node of Ranvier; *ol*, outer lamina; *pm*, plasma membrane; *r*, ridge.) (Reprinted with permission from Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.)

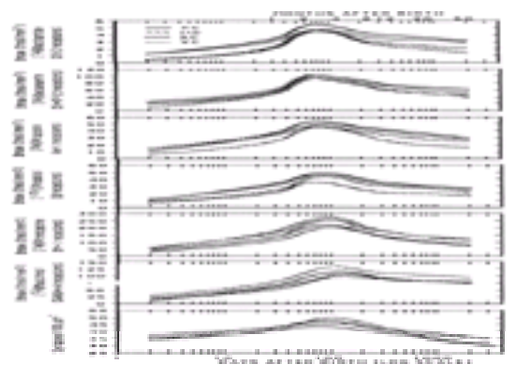

FIGURE 2.12-6 Developmental changes in the density of synapses and receptors in the prefrontal (*PC*), primary motor (*MC*), somatosensory (*SC*), and primary visual (*VC*) cortical regions. Age is presented in postnatal days on a logarithmic scale. Density of synapses is greatest at 2 to 4 months of age, then it declines as functionally irrelevant synapses are pruned according to experience. (Reprinted with permission from Rakic P: Development of the cerebral cortex in human and nonhuman primates. In *Child and Adolescent Psychiatry: A Comprehensive Textbook*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.)

The effect of experience on the formation of brain circuitry in the first years of life has generated a tremendous interest among neuroscientists. There are clear examples of the impact of early sensory experience on the wiring of cortical sensory processing areas. Similarly, early movement patterns are known to reinforce neural connections in the supplemental motor area. A common mechanism has emerged, in which neurons rapidly form a roughly fivefold excess of synaptic connections, then, through a Darwinian process of elimination, only the synapses that serve a relevant function persist. Synaptic pruning appears to preserve input in which the presynaptic cell fires in synchrony with the postsynaptic cell. This serves to reinforce neural circuits that are activated repeatedly. Neuroscientists assume that, as demonstrated experimentally in sensory and motor areas, experience plays a critical role in the modulation of the synaptic connectivity of cognitive and emotional association areas.

An emerging concept of great significance to both child and adult psychiatry is the existence of early windows of time, during which the brain establishes the basic circuitry for language, emotion, logic, mathematics, movements, and music. These windows open within the first few months of life and may close in some cases by 1 year of age. For example, a perceptual map of phonemes, the building blocks of language, is formed within the higher auditory processing areas of Wernicke's area during the first 12 months of life. The significance of this perceptual map is illustrated by the fact that certain sounds used by non–English-speaking peoples are not found in English. Studies have shown that American babies less than 6 months of age can be taught to discriminate these non-English sounds, but babies older than 6 months are no longer capable of hearing them. Moreover, the map for English differs from that for Japanese, for example, in the location of neurons that respond to the sounds “ra” and “la.” In English, these neurons are located far apart within the auditory cortex, whereas in the Japanese brain, which has difficulty distinguishing these sounds, they are so closely intertwined as to be virtually overlapping. Thus, to a Japanese individual, /la/ and /ra/ elicit nearly the same pattern of neural activity, a fact that may underlie the difficulties of the Japanese in distinguishing these sounds. In the English language, studies have shown that babies of mothers who spoke loquaciously to them acquired a larger vocabulary than babies of taciturn mothers. These findings suggest that very early experiences may establish the density and fidelity of neural circuits for specific perceptual functions, the consequences of which may affect individuals for the rest of their lives.

In the realm of emotion, early childhood experiences have been suspected to be at the root of psychopathology since the earliest theories of Sigmund Freud. Franz Alexander's psychoanalytic method aimed at tracing the threads of the earliest childhood memories to allow the patient to relive them in a less pathological environment, a process he termed a “corrective emotional experience.” While neuroscientists have no data yet to bring this method down to the level of neurons and circuits, emerging results demonstrate a profound effect of the reactions of early caregivers on an adult individual's emotional repertoire. For example, *attunement* is defined as the process by which caregivers play back a child's inner feelings. If a baby's emotional expressions are reciprocated in a consistent, sensitive manner, then certain emotional circuits are reinforced. These circuits likely include the limbic system, and in particular, the amygdala, which serves as a gate to the hippocampal memory circuits for emotional stimuli. In one study, for example, a baby whose mother repeatedly failed to mirror her level of excitement emerged from childhood an extremely passive girl, who was unable to experience a thrill or joy. The ability to learn and to respond to psychotherapy in adulthood must involve reorganization of synaptic connections. Such reorganization, as documented by functional neuroimaging, may progress over weeks to months.

The relative contributions of nature and nurture are perhaps nowhere more indistinct than in the maturation of emotional responses. In part, this is because the localization of emotion within the adult brain is only poorly understood. However, it is reasonable to assume that the reactions of one's caregivers in the first years of life are eventually internalized as distinct neural circuits, which may be only incompletely subject to modification through subsequent experience. For example, axonal connections between the prefrontal cortex and the limbic system, which very likely play a role in the modulation of basic drives, are established between ages 10 and 18 months. Recent work suggests that a pattern of terrifying experiences in infancy may flood the amygdala and drive memory circuits to be specifically alert to threatening stimuli, at the expense of circuits for language and other academic skills. Thus, infants raised in a chaotic and frightening home may be neurologically disadvantaged for the acquisition of complex cognitive skills in school.

Results documenting the organizing effect of early experiences on the internal representation of the external world have been found in the related realms of mathematics, logic, and music. A series of recent studies have shown that groups of children who were given 8 months of intensive classical music lessons in the preschool years subsequently showed significantly better spatial and mathematical reasoning than a control group. In grade school, nonmusical tasks such as navigating mazes, drawing geometric figures, and copying patterns of two-color blocks were performed significantly more skillfully by the musical children. Early exposure to music may thus be an ideal preparation for later acquisition of complex mathematical and engineering skills.

These observations suggest a neurological basis for the developmental theories of Jean Piaget, Erik Erikson, Margaret Mahler, John Bowlby, Sigmund Freud and others. Erikson's epigenetic theory states that normal adult behavior results from the successful, sequential completion of each of several infantile and childhood stages. According to the epigenetic model, failure to complete an early stage is reflected in subsequent physical, cognitive, social, or emotional maladjustment. By analogy, the experimental data discussed above suggest that early experience, particularly during the critical window of opportunity for establishing neural connections, primes the basic circuitry for language, emotions, and other advanced behaviors. Clearly, miswiring of the brain in infancy may later lead to severe handicaps when the individual attempts to relate to the world at an adult level. Brain development is achieved through the molecular interactions of thousands of gene products. Neurological disorders of infancy and childhood resulting from genetic mutations or from influences external to the brain therefore may hinder the proper emergence of mature emotional reactions.

CLASSIFICATION OF NEUROLOGICAL DISORDERS

An abnormality of mental function due to a neurological condition is called an encephalopathy. Encephalopathies are broadly defined as *static* or *progressive*. Patients with static encephalopathies gain developmental milestones at a slower rate than normal age-matched controls but do not lose abilities they have acquired. *Progressive encephalopathies* are characterized by a loss of milestones and are prognostically much more ominous. Either type of encephalopathy may be due to a genetic defect or to an acquired condition, such as infection, trauma, asphyxia, or tumor. Genetic variants that encode defective versions of crucial organellar proteins may prevent the proper maintenance of cellular homeostasis and may lead to accumulation of toxic molecules, on a time-scale of days, months, or years. Examples of these disorders include lysosomal storage diseases, mitochondrial disorders, peroxisomal disorders, and metal transport disorders. Compromise of the vascular supply to the brain caused by stroke or hypotension can interfere with cellular metabolism and produce permanent neuronal damage. Infectious agents can injure neurons directly or indirectly, as a result of an excessive immune response. Space-occupying lesions can disrupt neuronal activity by interrupting intercellular pathways or by compromising the blood supply.

Disorders of the gray matter are called *poliodystrophies* and are characterized by dementia and seizures. Disorders of the white matter are called *leukodystrophies* and are characterized by spasticity, ataxia, and optic atrophy. Many progressive encephalopathies contain features of both gray and white matter disease. Storage diseases consist of accumulations of unmetabolized waste products, which may be found in any of a large number of cell types. The combination of specific cell types in which storage products accumulate determines the neurological deficits.

The age of the patient at the time of an acquired insult and the extent of the insult are the main determinants of the prognosis of an acquired encephalopathy. Younger patients are more susceptible to infection and metabolic stress, but they are also more resilient than older patients in that they generally can recover more previous function after a comparable insult. Premature birth is a major cause of static encephalopathy, and there is a strong correlation between low birth weight and delays in cognitive and emotional development.

NEUROLOGICAL ASSESSMENT

Neurological assessment consists of a thorough history beginning with conception and proceeding through pregnancy, birth, early neonatal course, subsequent hospitalizations, chronic medical conditions, acquisition of milestones in each category of normal development, medications, allergies, family history including consanguinity, social history, and review of systems (Table 2.12-2). Neurological examination includes mental status examination, cranial nerves, motor system, associated motor system (coordination), sensory system, and reflexes. The neurological examination must be tailored to the age and abilities of the infant or child. Normal development is assessed in the categories gross motor, fine motor, speech and language, and social/adaptive. In younger patients, assessments often rely

more on observation and play interactions than on direct questioning. Subtle adaptive deficiencies may be missed in very young children even with detailed testing and may only become apparent as the age-appropriate context becomes more complex.

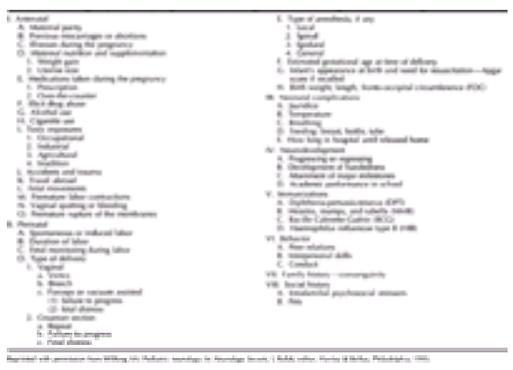

Table 2.12-2 Pediatric Neurological History

Standardized tests of neuropsychological functions may help discriminate between adaptive disorders and primary psychiatric conditions. Parents are the best sources of information about a child's behavior, and confirmatory assessment by teachers or other caregivers may yield important clues to a child's behavioral repertoire.

STATIC ENCEPHALOPATHIES

Static encephalopathies result from either acute insults to the developing brain that damage brain cells or a genetic deficiency that renders information processing abnormally inefficient but does not cause storage of metabolic products or ongoing cellular damage. Static encephalopathies are much more common than progressive encephalopathies, yet they are much less likely to have a specific identifiable cause. In the absence of neurocutaneous stigmata, a thorough diagnostic evaluation for static encephalopathies is usually fruitless.

Prematurity The ability to support the survival of premature infants who weigh as little as 450 grams at birth represents one of the technological triumphs of medicine. Longitudinal studies following the neurodevelopmental progress of premature infants for 10 and more years, however, have found very high rates of developmental delay and mental retardation among infants born prematurely. These delays have been attributed to a combination of the same high-risk factors that may have led to the premature birth (low socioeconomic status, lack of prenatal care) and, more importantly, to the intensive medical interventions necessary to sustain the extrauterine life of underdeveloped babies (high ventilatory pressures, tube feeding, and broad-spectrum antibiotics).

Classification Low birth weight is defined as less than 2500 grams, very low birth weight is defined as less than 1500 grams, and extremely low birth weight is defined as less than 1000 grams. The main factor determining the morbidity and mortality of premature infants is their pulmonary status: immature lungs can only oxygenate the blood sufficiently when ventilated at high pressures. These pressures increase the intrathoracic pressure and in turn increase the intracranial venous pressure, which may cause hemorrhage of blood from the capillaries of the germinal matrix of the brain. In other patients, the lungs cannot oxygenate the blood at any pressures, and the blood oxygen concentrations fall so low that irreversible neuronal damage occurs. A comparable mechanism of hypoxic-ischemic injury in a full-term newborn is birth asphyxia (Fig. 2.12-7). Each of these conditions is strongly associated with subsequent development of a static encephalopathy.

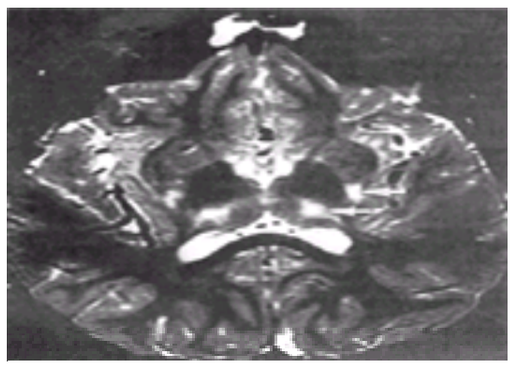

FIGURE 2.12-7 Axial T2-weighted MRI image at the level of the thalamus, illustrating bilaterally symmetrical hyperintensities in the thalamus (*arrows*). These scars are the result of severe hypoxic-ischemic injury due to birth asphyxia. At age 6 years, this child had predominantly choreoathetotic cerebral palsy and low-normal intelligence. (Reprinted with permission from Menkes J: *Textbook of Child Neurology*, ed 5. Williams & Wilkins, Baltimore, 1995.)

Intracranial Hemorrhage Four grades of intracranial hemorrhage in premature infants are recognized. *Grade I* consists of hemorrhage due to high venous pressures limited to the germinal matrix; in *grade II*, blood also bursts into the adjacent lateral ventricle but does not cause ventricular dilatation, and *grade III* occurs when the ventricles dilate because the intraventricular blood clogs the drainage of cerebrospinal fluid. *Grade IV*, in a contrasting mechanism, results from hypoxic-ischemic injury to cerebral blood vessels in states of very low blood oxygen tension and low blood pressure. When blood at normal pressures subsequently reperfuses damaged vessels, it leaks into the brain in several areas. In extreme cases, large regions of cortex are destroyed, leaving cerebrospinal fluid–filled cysts.

Neurodevelopmental Outcome Very-low-birth-weight babies are 8 times more likely to have mental retardation and 24 times more likely to have movement disorders known as cerebral palsy than are full-term babies. The types of cerebral palsy are spastic quadriparetic, spastic diplegic, hemiplegic, pseudobulbar, and choreoathetotic. Several longitudinal studies have documented an average lowering of intelligence quotient (I.Q.) scores of 6 to 14 points in cohorts of very low birth weight infants. Among extremely low birth weight infants, as many as 50 to 80 percent required some level of special education by age 10 years, compared with 15 percent of full-term infants by the same age. In infants who suffered intracranial hemorrhage, grades I and II generally were not associated with additional cognitive delays, but grade III and especially grade IV hemorrhages posed a considerable increased risk of scholastic underachievement during childhood.

Expremature infants register higher scores on scales of childhood anxiety, depression, social isolation, conduct disorder, and aggressivity. The best predictors of these outcomes is degree of prematurity, whereas neurological evaluations or cranial ultrasonograms in the neonatal period are not predictive of emotional outcome. In general, premature infants are rated less temperamental as young toddlers, but as late toddlers they are rated more withdrawn and more prone to tantrums. In childhood, 5 to 35 percent of expremature infants fit the diagnostic criteria for attention-deficit/hyperactivity disorder, compared with 3 to 15 percent of ex–full term infants. Particular reasons for this increased risk of attention-deficit/hyperactivity disorder among expremature infants may include the fact that low socioeconomic status is associated with both prematurity and attention-deficit/hyperactivity disorder. Another possible reason is that the germinal matrix in which hemorrhages occur in premature infants later gives rise to the caudate nucleus, which has been increasingly implicated in the pathogenesis of attention-deficit/hyperactivity disorder in neurobiological studies.

Strokes Prenatal strokes most commonly occur in the distribution of the middle cerebral artery and therefore damage the frontal and parietal lobes. The most common presenting features are seizures and hemiparesis. In longitudinal follow-up studies, mean I.Q. was in the normal range, but the laterality of the lesion did correlate with lower scores on specific subtests: left middle cerebral artery infarctions reduced verbal I.Q. scores, while right middle cerebral artery infarctions lowered nonverbal I.Q. scores. Studies of hemispherically injured infants have suggested, with variable consistency, that the adult model of hemispheric localization of emotional responses, in which activation of the left frontal lobe raises mood and activation of the right frontal lobe depresses mood, is evident even below the age of 1 year. Other data appear to contradict this model: lesions that inactivate the right frontal lobe in children unexpectedly elicit more-negative behavior. The frontal cortex

has various specialized regions that play largely unknown roles in the expression of emotions, and the clinically documented lesions have not yet been systematically characterized to allow specific correlation between lesion site and behavioral manifestation, even in adults. At this time, no generalizations can be made from longitudinal studies regarding the emotional consequences of prenatal and childhood ischemic injuries. Children with cognitive and motor disabilities caused by strokes use special education resources much more than children who have not had strokes, and they are more susceptible to a range of psychiatric symptoms, especially depression.

Chromosomal Anomalies The genetic basis of behavior is only beginning to be understood. Epidemiological studies have suggested a significant heritable component to all major psychiatric disorders. Most static encephalopathies are due to spontaneous genetic variations. Most children with static encephalopathies do not have characteristic dysmorphic features, and currently no genetic screening tests are available to determine the cause of their delays. There is reason to believe that such diagnostic tools may emerge in the coming decades, but the inherent variability of clinical expression of current well-characterized genetic variants must temper enthusiasm for establishment of a genetic explanation for behavior. At the same time, isolation of genes even partially responsible for specific cognitive functions and emotional repertoires is a high priority. Three chromosomal syndromes have characteristic neuropsychiatric manifestations other than simple mental retardation: fragile X syndrome, Williams syndrome, and Down syndrome.

Fragile X Syndrome *Fragile X syndrome* is characterized by mental retardation (I.Q. 35 to 70), hyperlexia without comprehension, deficits in executive function and visuospatial attention, autism, aggressivity, impulsivity, and depression. Dysmorphic features include macroorchidism, large head, elongated palpebral fissures, and large ears. There are no characteristic abnormalities in brain structure. It is the most common identified genetic cause of mental retardation (1 in 4000 people) and affects boys almost exclusively. The genetic defect is the expansion of a trinucleotide deoxyribonucleic acid (DNA) repeat in the gene *FMR*-1 on the X chromosome. This expansion causes fragility of the chromosome when cells are grown in folate-deficient medium. The mutation silences expression of the gene, resulting in absence of the FMR proteins (FMRP), a series of ribosome-associated ribonucleic acid (RNA)-binding proteins generated by alternative splicing of *FMR*-1 transcripts. Recently, a highly accurate polymerase chain reaction (PCR) test has become available to identify affected individuals and carrier females. It is reasonable to screen mildly to moderately retarded male children with the PCR test, if the family wishes to exhaust the diagnostic possibilities. However, fragile X syndrome accounts for less than 4 percent of people with mental retardation and is not a cause of academic difficulties in the absence of mental retardation. There is no specific treatment for fragile X syndrome, other than special education and psychiatric care.

Williams Syndrome Children with Williams syndrome exhibit a relative preservation of language abilities and may have an unusually large vocabulary. In marked contrast, their visuospatial abilities are severely impaired. For example, they may be able to describe the parts of a house in great detail, but cannot draw a recognizable image of a house. They can recognize individual details but cannot recognize overall patterns in an image. The deficient functions are those usually ascribed to the right hemisphere, though the correlation is not complete. Neuropathological studies show small parietal lobes, narrowing of the corpus callosum, and abnormal proportions of the cortical layers. Genetic studies have implicated a mutation in the LIM-kinase I gene as the probable cause of the perceptual disturbance. Further studies are needed to establish the role of the LIM-kinase I gene in Williams syndrome. The function of this gene is unknown. The relationship between the genotype and the phenotype is not understood, but further study promises to yield interesting insights into the biological basis of perception.

Down Syndrome In contrast to children with Williams syndrome, children with Down syndrome have relatively preserved visuospatial skills and more-impaired verbal skills. Down syndrome children are generally docile and cheerful, but they may have outbursts of anger and hyperactivity. Children with the full Down syndrome phenotype harbor a third chromosome 21, whereas a portion of the Down syndrome phenotype is seen if only part of an extra chromosome 21 is present. There are several genes in the extra segment, and efforts to correlate specific gene triplications with specific cognitive skills are ongoing.

Head Trauma The neuropsychiatric consequences of head trauma are directly related to the nature and severity of the impact. Mild injuries may or may not cause loss of consciousness, and if no bleeding occurs, such injuries are called concussions. A postconcussional syndrome consisting variously of headache, somnolence, vomiting, syncope, dizziness, irritability, and amnesia for the event may last for days, weeks, or in some cases months. A longer duration of symptoms is associated with pretrauma depression and anxiety. The vast majority of children suffering mild head trauma have complete recovery.

Severe injuries are associated with intracranial hemorrhage, intracranial swelling, or both, and children who regain consciousness after such injuries have a high incidence of long-term behavioral changes (Table 2.12-3). Cognitive and learning disabilities are common, and aggressive outbursts of behavior are characteristic, especially of frontal lobe damage. A multidisciplinary approach to the physical and mental handicaps caused by severe head trauma, including pharmacological and behavioral therapies, is available at specialized rehabilitation units.

Primary lesions
Intra-axial
Diffuse axonal injury
Cortical contusion
Subcortical matter injury
Primary brainstem injury
Extra-axial hematomas
Extradural
Subdural
Diffuse hemorrhage
Subarachnoid
Intraventricular
Primary vascular injuries
Secondary lesions
Pressure necrosis (secondary to brain displacement and herniations)
Temporal arterial infarction
Diffuse hypoxic injury
Diffuse brain swelling
Boundary and terminal zone infarction
Others
Fatty embolism
Secondary hemorrhage
Infection

Reprinted with permission from Menkes J: *Textbook of Child Neurology*, ed 5. Williams & Wilkins, Baltimore, 1995.

Table 2.12-3 Classification of Traumatic Intracranial Lesions

Acute Infections Meningitis consists of an infection in the cerebrospinal fluid, and encephalitis consists of an infection in the brain itself. The principal organisms are viruses and bacteria, and less commonly, fungi and tuberculosis are involved. When aggressively treated with antibiotics and supportive care, most children recover fully. The most common neurological sequela of bacterial meningitis is hearing loss (Table 2.12-4). A minority of patients with acute meningitis or encephalitis, especially those with hypoxic-ischemic injury, exhibit long-term cognitive deficits and may become depressed or anxious.

Complication	Number of Cases
Mental retardation	3
Seizures	3
Hemiparesis or quadriparesis	1
Bilateral deafness	4
Vestibular disturbance	1
Hydrocephalus	1
Total	13 (18%)

Reprinted with permission from Menkes J: *Textbook of Child Neurology*, ed 5. Williams & Wilkins, Baltimore, 1995.

Table 2.12-4 Major Complications Seen in 71 Children After Recovery From Meningitis

PROGRESSIVE ENCEPHALOPATHIES

The neuropsychiatric consequences of progressive encephalopathies initially affect the child, but as the child's mental functions progressively deteriorate, the caregivers increasingly demonstrate psychiatric symptoms. The stresses of watching a beloved child gradually lose skills, become vegetative, and eventually die are

extreme and try even the strongest of families. A thorough neurological evaluation is always indicated when children begin to lose developmental milestones or cease to gain milestones at their usual rate. The history and physical examination often suffice to narrow the differential diagnosis significantly, and a focused neuroimaging and laboratory evaluation frequently confirms the diagnosis. Potentially treatable diseases must be considered and prompt treatment offered to counteract the biochemical abnormality. Establishing a diagnosis even in untreatable conditions helps the family accept their situation and permits genetic testing in future pregnancies, when applicable. For a small but increasing number of diseases, presymptomatic or even prenatal treatment can avert the appearance of symptoms in patients with the disease genes.

Poliodystrophies *Poliodystrophy* refers to a disorder of the neuronal cell bodies, generally those in the cerebral cortex. The main presenting symptoms are seizures and loss of cognitive functions. Major psychiatric symptoms include loss of verbal communication, oppositional behavior, irritability, and inattentiveness. The poliodystrophies include the lipidoses, such as Tay-Sachs disease, Sandhoff's disease, Niemann-Pick disease, Gaucher's disease, Fabry's disease, and neuronal ceroid lipofuscinosis; the mucopolysaccharidoses, such as Hunter's syndrome, Hurler's syndrome, and Sanfilippo's syndrome; Alper's syndrome; the mitochondrial encephalomyopathies, such as mitochondrial encephalopathy with lactic acidoses and strokelike episodes (MELAS), Kearns-Sayre syndrome, myoclonic epilepsy with ataxia and ragged red fibers (MERRF), and Leigh's disease; the epileptic encephalopathies, such as infantile spasms, Lennox-Gastaut syndrome, and Rett's disorder, and the hereditary movement disorders with dementia, such as Huntington's disease, ataxia-telangiectasia, Wilson's disease, and Lesch-Nyhan syndrome.

Leukodystrophies *Leukodystrophy* refers to a disorder of white matter, which consists of myelinated fiber tracts. The main presenting symptoms are spasticity, ataxia, and optic atrophy. Major psychiatric symptoms include disinhibition, impulsivity, poor judgment, and emotional lability. The leukodystrophies include metachromatic leukodystrophy, globoid cell leukodystrophy, adrenoleukodystrophy, Pelizaeus-Merzbacher disease, Canavan's disease, and Alexander's disease.

Treatment Recent experimental bone marrow replacement protocols have shown prevention of progression and even reversal of disease in metachromatic leukodystrophy, globoid cell leukodystrophy, adrenoleukodystrophy, and Hurler's syndrome. Gaucher's disease has been successfully treated with enzyme replacement therapy. Several dietary restrictions or supplements have been tried with varying success for leukodystrophies.

Slow Infections

Acquired Immune Deficiency Syndrome Pediatric acquired immune deficiency syndrome (AIDS) is most frequently transmitted transplacentally. This transmission can be greatly reduced if the human immunodeficiency virus (HIV)-positive mother takes zidovudine (Retrovir) in the second half of pregnancy. Despite concerns that HIV-positive women would not seek prenatal care because they did not wish to have their HIV status entered in their medical records, it is clearly indefensible for prenatal clinics not to seek such women for testing and, if HIV-positive, to offer treatment with zidovudine. This approach has been bolstered by the recent institution of mandatory HIV testing for all newborns, positive results of which immediately identify HIV-positive mothers, albeit too late for zidovudine treatment. Perinatal transmission via breast milk or blood transfusions is much less common and tends to cause milder neuropsychiatric symptoms. Survival in congenitally or perinatally acquired HIV infection is rare beyond the age of 10 years.

Neurological symptoms, consisting of progressive motor dysfunction and loss of developmental milestones, and signs, particularly deceleration of head growth, are the presenting features of AIDS in up to 18 percent of children and adolescents. Unlike HIV-positive adults, who frequently have acute psychosis, HIV-positive children more often exhibit inattentiveness, impulsivity, depression, anxiety, and adjustment disorders as the main psychiatric symptoms. HIV-positive adolescents may also develop acute psychotic symptoms. Any acute change in mental status in an HIV-positive patient of any age requires a thorough neurological evaluation, possibly including neuroimaging, serologic testing, cerebrospinal fluid examination, and brain biopsy. Nutritional status, especially thiamine levels, electrolyte homeostasis, thyroid function, and other metabolic or endocrine indicators, should also be checked.

Any chronic infection in children may elict a reactive depression, especially if family members are ill. The organic effects of HIV tend to exacerbate depression. The best therapeutic approach is to maintain a high level of activity for as long as possible, including schooling and peer activities.

Other Slow Infections Other viral infections that may have minimal acute effects on the nervous system but may produce a slowly progressive encephalopathy include progressive multifocal leukoencephalopathy (caused by the JC virus), subacute sclerosing panencephalitis (caused by the measles or rubella viruses), and chronic enteroviral infections. Spongiform encephalopathies, such as Creutzfeldt-Jakob disease, thought to be caused by a prion, may affect children, causing anxiety, impaired judgment, and rapid cognitive decline.

Inborn Errors of Metabolism Genetic defects in cellular metabolism may affect the functioning of specific regions of the brain. Wilson's disease, a treatable disorder of copper transport, if left untreated may cause personality changes, impulsivity, social withdrawal, and flapping movements of the arms. Copper accumulates in the globus pallidus and putamen, cell loss is also seen in the substantia nigra and dentate nucleus, and astrocytes proliferate in the cerebral cortex, basal ganglia, brainstem nuclei, and cerebellum. Chelation therapy may reverse the symptoms, but the initial mobilization of copper by chelators may transiently increase neurological toxicity, sometimes for as long as several years. Infantile nephropathic cystinosis may cause visuospatial deficiencies due to loss of volume in the parietal and occipital regions prior to the initiation of therapy. Disorders of fatty acid oxidation may selectively affect verbal skills. Acute intermittent porphyria may cause panic attacks, depression, and agitated psychosis. Galactosemia, phenylketonuria, and ornithine transcarbamylase deficiency may cause mental retardation.

Acquired Metabolic Disorders Systemic metabolic disorders may affect behavior. Congenital hypothyroidism may cause severe retardation that may be only partially mitigated by postnatal thyroid replacement therapy. Acquired hypothyroidism in infants and children may cause disorders of spatial orientation, as well as general learning delays. Hyperthyroidism in children may cause hyperactivity, restlessness, and inattentiveness. Chronic hepatic and renal failure may cause deficits in visual learning and in memory.

Brain Tumors Brain tumors have been associated with the entire spectrum of behavioral disorders. The particular behavioral manifestations of a brain tumor are determined by its location, histological type, and size. Posterior fossa tumors may cause cerebellar and brainstem signs and hydrocephalus. Supratentorial infiltrative glial tumors may disable nerve fiber tracts and produce dementia.

In addition to the effects of the tumor itself, a larger risk factor for intellectual decline is the amount of cranial irradiation, which damages all cell types of the nervous system. For example, cystic cerebellar astrocytoma is a posterior fossa tumor typically treated with surgical resection but no irradiation, whereas surgical resection of another posterior fossa tumor, medulloblastoma, is usually followed by irradiation. In one long-term follow-up study, 62 percent of patients with cerebellar astrocytoma had I.Q.s above 90, whereas only 11 percent of patients with medulloblastoma and irradiation had I.Q.s above 90. Over time, the discrepancy widened. Addition of chemotherapy compounded the radiation-induced cognitive deficits.

The most common psychiatric disturbances in patients with brain tumors include reactive depression, oppositional behavior, anxiety disorders, and thought disorders. In some cases, treatment of the tumor improved these conditions, but in others, new symptoms emerged during treatment. Family members may also be severely affected, especially siblings of the patient, some of whom may view the patient as no longer part of the family. Treatment should include cognitive interventions for the patient and psychiatric support for both the patient and the family.

SEIZURE DISORDERS

Much of the study of the neuropsychiatric aspect of epilepsy applies to children, although fewer detailed studies have been done in children than in adults. Several types of epilepsy are only seen in infancy and childhood, including infantile spasms, benign rolandic epilepsy, idiopathic generalized epilepsy of childhood, idiopathic absence epilepsy, and febrile seizures (Fig. 2.12-8). Focal seizures are relatively rarer in childhood, but they may produce a wide range of behavioral manifestations (Table 2.12-5).

Seizure Manifestation	Number of Patients With Manifestations (Total 23): Age 1–5	Age 1–16	Total
Aura	6	10	16
Altered consciousness	12	13	25
Change in position of body or limbs	10	11	21
Integrated but confused activity	8	11	19
Staring or dazed expression	10	8	18
Epigastric sensation, nausea, vomiting	9	5	14
Oral movements, drooling	8	5	13
Muttering, mumbling, hissing	5	5	10
Walking, wandering	4	6	10
Pallor or flushing	5	4	9
Rubbing or fumbling	4	5	9
Speech (usually irrelevant or incoherent)	3	5	8
Affective disturbance (fear, anger)	5	3	8
Stiffening of body or limbs	5	3	8
Falling	4	3	7
Aggressive activity	4	3	7
Dreamy state	2	3	5
Forced thinking or ideational blocking	1	4	5
Searching or orienting movements	1	3	4
Abdominal pain	3	1	4
Incontinence (urinary)	2	1	3
Perceptual disturbance (visual, auditory)	0	3	3

Reprinted with permission from Aicardi J: Textbook of Child Neurology, ed 5. Williams & Wilkins, Baltimore, 1995.

Table 2.12-5 Clinical Manifestations of Complex Partial Seizures in Childhood

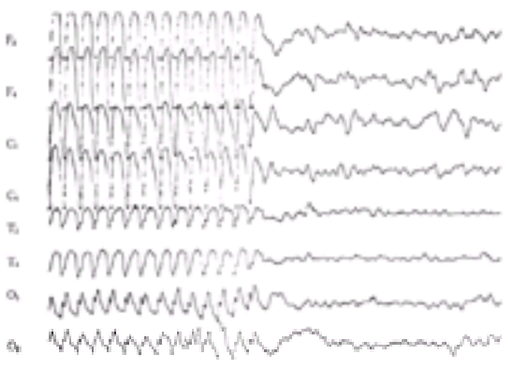

FIGURE 2.12-8 Three per second spike-wave discharges in an 11-year-old girl with frequent absence seizures induced by hyperventilation. *F3*, left frontal; *F4*, right frontal; *C3*, left central; *C4*, right central; *T3*, left temporal; *T4*, right temporal; *01*, left occipital; *02*, right occipital. (Reprinted with permission from Menkes J: *Textbook of Child Neurology*, ed 5. Williams & Wilkins, Baltimore, 1995.)

Some 70 to 80 percent of patients with seizures can expect good control with anticonvulsant medications. At a minimum, most seizure patients must take daily medication. Patients whose seizure control is inadequate live with the specter of a seizure striking them at any time. Thus, an event lasting in some cases 0.001 percent of the person's lifetime exerts a pervasive influence on the person's activities. This may lead to parental overprotection, teasing from peers, and disruption of social adjustment. If seizures have been controlled for at least 6 months and drug concentrations in serum are in the therapeutic range, then activities may usually include all but climbing, horseback riding, diving into deep water, unsupervised swimming, or use of dangerous equipment.

No single behavioral disorder is attributable to all patients with epilepsy. Children with childhood absence epilepsy tend to be passive and dependent. Children with temporal lobe focal epilepsy, in contrast, have variously been observed to be irritable, excitable, compulsive, stubborn, hyperactive, aggressive, depressed, or moody. Like adults, children with temporal lobe epilepsy are at increased risk for development of psychosis. Among children who have attempted suicide, there is a 15-fold increased appearance of children with epilepsy, especially those taking phenobarbital.

One third of children with epilepsy have psychiatric disorders, compared with 6.6 percent of healthy children and 11.6 percent of children with physical illnesses not involving the brain. Anticonvulsant medications, especially phenobarbital, may cause hyperactivity and irritability, whereas phenytoin (Dilantin) and carbamazepine (Tegretol) may cause sedation.

Pseudoseizures Pseudoseizures, also called nonepileptic seizures, are psychiatric fugue states caused by an unacceptable emotional conflict. They often occur in patients who have epilepsy or a family member with epilepsy. Nonepileptic seizures usually do not result in physical harm and are typically precipitated by a situation from which the patient would like to escape. Patients with nonepileptic seizures often have a history of physical or sexual abuse. Electroencephalograms (EEGs) show no abnormalities unless there is a concurrent epileptic seizure disorder. The events can sometimes be induced by suggestion. Once the nonepileptic nature of the events is determined, psychiatric care is important in the management of the disorder.

HEADACHE

Migraine Migraine headaches are relatively more common than tension-type headaches in children, which is the opposite of the prevalence in adults. Children may suffer from migraine with aura (classic migraine) or migraine without aura (common migraine). Association with vomiting, true vertigo, visual aura, menstrual cycle, or certain foods, and relief with sleep, permit distinction of migraine from tension-type headache. True migraine headaches rarely occur more than once per week. In some extreme cases, amnesia, confusion, and psychosis may result from migraine headaches. Frequently, a family history of migraine exists. Each patient with headaches, particularly of recent onset, should have a neuroimaging study. Migraines must be managed with patient education regarding avoidance of risk factors and prompt treatment, including sleeping.

Treatments for acute migraine include sumatriptan (Imitrex), naratriptan (Amerge), zolmitriptan (Zomig), rizatriptan (Maxalt), caffeine-ergotamine (Cafergot, Wigraine), butalbital-caffeineacetaminophen (Fioricet, Esgic), dihydroergotamine (DHE), phenothiazines, and opioids. Migraine prophylaxis has been achieved with serotonin agonists (eg., cyproheptadine [Periactin]), antidepressants, antihistamines, anticonvulsants (e.g., valproate [Depakene]), calcium channel blockers, b-adrenergic receptor antagonists, and hormones. Several flawed double-blind controlled trials have shown benefit of each of these agents, though the response in the placebo arm of many of the studies was unexpectedly low. It is difficult to separate the placebo effect from true migraine prophylaxis.

Chronic Tension-Type Headaches Tension-type headaches are frequent band-like, frontal, or shifting headaches sometimes associated with dizziness, nausea, and scotoma during the headache. Periods of daily headache may be interspersed with headache-free periods lasting days to weeks. The severity of an individual headache waxes and wanes over the course of the day. Migraine and tension-type headaches may coexist in a patient, who usually can describe distinct patterns of pain. The specter of a brain tumor looms in the minds of even young children, and a thorough history and neurological examination followed by an imaging study, which is usually normal, may allay anxieties and allow the headache to remit fully. An ophthalmological examination to rule out refractive errors, particularly astigmatism, is essential. Once structural causes of headache have been ruled out, psychological issues should be explored. Family conflicts, particularly parental conflicts or parental divorce, are at the root of tension-type headaches in a large number of children. Other factors to be addressed are difficulties meeting expected school performance, violence in the school, illness in other family members, and psychiatric conditions such as depression, anxiety, and attention-deficit/hyperactivity disorder. Both the affected child and the parents should be interviewed.

Treatment begins with reassurance that the headache does not reflect a serious medical condition. Patients should be instructed to keep a log of each headache, what triggered it, what was done to treat it, and (in girls) whether it is associated with menstrual periods. This serves to focus attention on the headaches as a valid medical condition rather than an inevitable nuisance for which there is no cure. Occasionally, an identifiable trigger can be identified and avoided. Patients should take an adequate dose of over-the counter analgesics at the earliest opportunity once a headache begins. Such medicines are much more effective at nipping a mild headache in the bud than at halting a full-blown headache. If the headaches persist despite the above interventions, prophylaxis often may be achieved with antidepressant drugs. All available antidepressants have equivalent efficacies, and the choice of drug should be based on adverse-effect profiles. Relaxation exercises or biofeedback may be effective.

Other Causes of Headache Fatigue or excessive exertion may cause an acute tension-type headache that responds to analgesics and rest. Sinusitis may cause facial pain in the context of fever and nasal congestion. Headache may be the sole manifestation of an acute viral syndrome. Dental pain, injury to the temporomandibular joint, or malocclusion that forces a child to chew on only one side of the mouth may strain one masseter muscle and produce unilateral headache.

Neck or head injury may cause muscle strain that can persist for weeks to months, especially when ongoing legal action is based on the presence of disability. Occipital headache may be caused by carpal tunnel syndrome, which can appear in adolescence. Headaches may be associated with epileptiform discharges even in the absence of clinical seizures. Twenty-four-hour ambulatory EEG may be necessary to demonstrate that spike and wave activity correlates with the duration of the headache. Headache may be the presenting symptom of collagen-vascular disease, hypersensitivity reaction, hypertension, or nervous system infection. Drugs and toxins with vasodilating properties may cause headaches. Headaches are especially associated with alcohol, caffeine withdrawal, nitrites, monosodium glutamate, and marijuana.

MOVEMENT DISORDERS

Involuntary movements are usually a sign of dysfunction of the basal ganglia. They may cause a disproportionate amount of anxiety and distress in children, who may be subject to teasing by their peers. Involuntary movements may be distinguished from seizures by their persistence with preservation of consciousness, their disappearance during sleep, and the absence of EEG abnormalities during the involuntary movements.

Classification *Chorea* is a rapid, random jerk of a limb, that may be incorporated into a movement that appears voluntary. *Athetosis* is a slow, writhing movement of the limbs, often associated with perinatal brain injury, especially hypoxic-ischemic injury or kernicterus. *Ballismus* is a high-amplitude, violent shooting of the limb from the shoulder or pelvis. It is an extreme version of chorea and is seen in Sydenham's chorea and lupus erythematosus. *Dystonia* is a persistent, simultaneous contraction of agonist and antagonist muscles. Tardive dyskinesia, dystonia, and tremor may be induced by medications in children as well as adults. *Myoclonus*, a sudden jerk of a body part, is not stereotyped, cannot be suppressed, and is nonrhythmic. *Tremor* is a continuous to-and-fro movement. *Tics* are instantaneous, stereotyped, low-amplitude movements. All movement disorders may involve one body part (focal), two or more continguous body parts (segmental), one side of the body (hemispheric), or the entire body (generalized).

Diagnosis and Treatment Movement disorders may be primary (idiopathic) or secondary (symptomatic) (Table 2.12-6). A thorough evaluation should be done, including history, especially of perinatal course, drug use, and systemic diseases; physical examination to characterize the movements, which may be difficult even for experts; neuroimaging, including magnetic resonance imaging (MRI) with contrast and possibly angiography; and laboratory examination, including toxicology screens, electrolytes, antinuclear antibodies, antistreptolysin-O titer, thyroid studies, pregnancy tests, and screening tests for specific genetic disorders. Among the genetic disorders that may cause movement disorders in children are Tourette's disorder, Hallervorden-Spatz disease, glutaric aciduria, hepatolenticular degeneration (Wilson's disease), Huntington's disease, abetalipoproteinemia, ataxia-telangiectasia, benign familial chorea, familial paroxysmal choreoathetosis, Lesch-Nyhan syndrome, ceroid lipofuscinosis, dopa-responsive dystonia, and myotonic dystrophy. Each disease has a specific treatment.

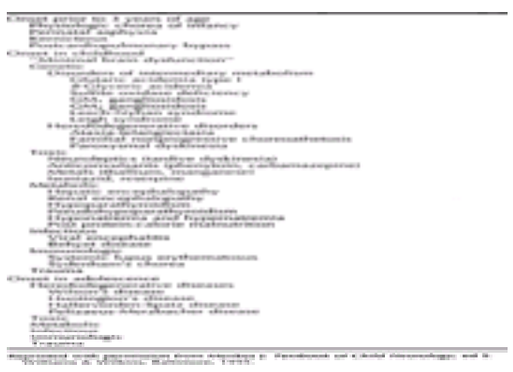

Table 2.12-6 Major Causes of Chorea

SUGGESTED CROSS-REFERENCES

Developmental neuroanatomy is described in Section 1.3. Adult manifestations of each of the conditions described in this section are discussed in the previous sections of Chapter 2. Narcolepsy is discussed in Chapter 21. Genetic counseling is discussed in Section 28.4. Normal child and adolescent development are described in Chapter 32. Mental retardation is discussed in Chapter 34. Tic disorders are discussed in Chapter 42. Psychiatric sequelae of HIV and AIDS are discussed in Section 2.8 and Section 49.6.

SECTION REFERENCES

Allen KD, Shriver MD: Enhanced performance feedback to strengthen biofeedback treatment outcome with childhood migraine. Headache *37:*169, 1997.

*Ammerman RT, Hersen M, Last CG, editors: *Handbook of Prescriptive Treatments for Children and Adolescents*, ed 2. Allyn & Bacon, Boston, 1999.

Anderson V, Bond L, Catroppa C, Grimwood K: Childhood bacterial meningitis: Impact of age at illness and acute medical complications on long term outcome. J Int Neuropsychol Soc *3:*147, 1997.

*Berg BO, editor: *Principles of Child Neurology.* McGraw-Hill, New York, 1996.

Caplan R, Arbelle S, Guthrie D, Komo S: Formal thought disorder and psychopathology in pediatric primary generalized and complex partial epilepsy. J Am Acad Child Adolesc Psychiatry *36:*1286, 1997.

Chugani DC, Muzik O, Behen M, Rothermel R, Janisse JJ, Lee J, Chugani HT: Developmental changes in brain serotonin synthesis capacity in autistic and non-autistic children. Ann Neurol *45:*287, 1999.

DeLong GR: Autism: New data suggest a new hypothesis. Neurology *52:*911, 1999.

Einfeld SL, Tonge BJ, Florio T: Behavioral and emotional disturbance in individuals with Williams syndrome. Am J Mental Retard *102:*45, 1997.

*Fenichel GM: *Clinical Pediatric Neurology*, ed 3. Saunders, Philadelphia, 1996.

*Frank Y, editor: *Pediatric Behavioral Neurology.* CRC Press, Boca Raton, FL, 1996.

Harris NGS, Bellugi U, Bates E, Jones W: Contrasting profiles of language development in children with Williams and Down syndromes. Dev Neuropsychol *13:*345, 1997.

Hokkanen L, Launes J: Cognitive recovery instead of decline after acute encephalitis: A prospective follow up study. J Neurol Neurosurg Psychiatry *63:*222, 1997.

Kumar R, Lang AE: Secondary tic disorders. Neurol Clin *15:*309, 1997.

Lewis DW, Middlebrook MT, Mehallick L, Rauch TM: Pediatric headaches: What do the children want? Headache *36:*224, 1996.

Mattia FR, deRegnier RA: Chronic physiologic instability is associated with neurodevelopmental morbidity at one and two years in extremely premature infants. Pediatrics *102:*35, 1998.

Mayes LC, Bornstein MH: Attention regulation in infants born at risk: Prematurity and prenatal cocaine exposure. In *Attention, Development, and Psychopathology*, JA Burack, JT Enns, editors. Guilford, New York, 1997.

Mazzocco MMM, Kates WR, Baumgardner TL, Freund LS: Autistic behaviors among girls with fragile X syndrome. J Autism Dev Disord *27:*415, 1997.

*Menkes JH: *Textbook of Child Neurology*, ed 5. Williams & Wilkins, Baltimore, 1995.

Minshew NJ, Luna B, Sweeney JA: Oculomotor evidence for neocortical systems but not cerebellar dysfunction in autism. Neurology *52:*917, 1999.

Muter V, Taylor S, Vargha-Khadem F: A longitudinal study of early intellectual development in hemiplegic children. Neuropsychologia *35:*289, 1997.

Parent A: *Carpenter's Human Neuroanatomy*, ed 9. Williams & Wilkins, Baltimore, 1996.

Radcliffe J, Bennett D, Kazak AE, Foley B: Adjustment in childhood brain tumor survival: Child, mother, and teacher reports. J Pediatr Psychol *21:*529, 1996.

Roberts MA, Verduyn WH, Manshadi FF, Hines ME: Episodic symptoms in dysfunctioning children and adolescents following mild and severe traumatic brain injury. Brain Inj *10:*739, 1996.

Saucerman SA: Dissociation of W/REM/NREM states may cause psychotic symptoms. Schizophr Res *25:*261, 1997.

Scott BS, Atkinson L, Minton HL, Bowman T: Psychological distress of parents of infants with Down syndrome. Am J Ment Retard *102:*161, 1997.

Slater P, Doyle CA, Deakin JF: Abnormal persistence of cerebellar serotonin-1A receptors in schizophrenia suggests failure to regress in neonates. J Neural Transm *105:*305, 1998.

Subramaniam B, Naidu S, Reiss AL: Neuroanatomy in Rett syndrome: Cerebral cortex and posterior fossa. Neurology *48:*399, 1997.

Sullivan PM, Montoya LA: Factor analysis of the WISC-III with deaf and hard-of-hearing children. Psychol Assess *9:*317, 1997.

*Swaiman KF: *Pediatric Neurology: Principles and Practice*, ed 2. Mosby, St. Louis, 1994.

Tepper VJ, Farley JJ, Rothman MI, Houck DL, Davis KF, Collins-Jones TL, Wachtel RC: Neurodevelopmental/neuroradiologic recovery of a child infected with HIV after treatment with combination antiretroviral therapy using the HIV-specific protease inhibitor ritonavir. Pediatrics *101:*7, 1998.

Waddington JL, Lane A, Scully PJ, Larkin C, O'Callaghan E: Neurodevelopmental and neuroprogressive processes in schizophrenia. Antithetical or complementary, over a lifetime trajectory of disease? Psychiatr Clin North Am *21:*123, 1998.

Wilfong AA: Pediatric neurology. In *Neurology Secrets*, LA Rolak, editor. Hanley & Belfus, Philadelphia, 1993.

Williams J, Grant M, Jackson M, Shema SJ: Behavioral descriptors that differentiate between seizure and nonseizure events in a pediatric population. Clin Pediatr *35:*243, 1996.

Wilson BA: Semantic memory impairments following non-progressive brain injury: A study of four cases. Brain Inj *11:*259, 1997.

Winders PC: *Gross Motor Skills in Children with Down Syndrome: A Guide for Parents and Professionals.* Woodbine House, Bethesda, 1997.

Wu YQ, Sutton VR, Nickerson E, Lupski JR, Potocki L, Korenberg JR, Greenberg F, Tassabehji M, Shaffer LG: Delineation of the common critical region in Williams syndrome and clinical correlation of growth, heart defects, ethnicity, and parental origin. Am J Med Genet *78:*82, 1998.

Zeanah CH, Boris NW, Larrieu JA: Infant development and developmental risk: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *36:*165, 1997.

2.13 NEUROIMAGING IN CLINICAL PRACTICE

JOSEPH C. WU, M.D., DANIEL G. AMEN, M.D., AND H. STEFAN BRACHA, M.D.

Neuroimaging has evolved tremendously, providing psychiatrists with unprecedented information about brain structure and function. Computer tomographic (CT) scanners, the first widely used neuroimaging devices, allowed assessment of structural brain lesions such as tumors or strokes. Magnetic resonance imaging (MRI) scans, developed next, allowed greater gray-white matter distinction than CT scans provided and allowed visualization of smaller brain lesions as well as white matter abnormalities. In addition to structural neuroimaging with CT and MRI, a revolution in functional neuroimaging has enabled clinical scientists to obtain unprecedented insights into the diseased human brain. The foremost techniques for functional neuroimaging include positron emission tomography (PET) and single photon emission computed tomography (SPECT). Primary observations of structural and functional brain imaging in neuropsychiatric disorders such as dementia, movement disorders, demyelinating disorders, and epilepsy have not only contributed to a greater understanding of the pathophysiology of neurological and psychiatric illnesses, but can also be of help to the practicing clinicians in difficult diagnostic situations.

STRUCTURAL NEUROIMAGING TECHNIQUES

CT Imaging CT imaging was developed in 1967 by Sir Godfrey Hounsfield using theories developed by Allan Cormack. CT imaging is based upon the use of X-ray beams that are absorbed differentially by various tissues. The attenuation of the X-ray beam caused by tissue absorption is measured by radiation detectors. Views of the tissue are obtained from different angles by rotation of the X-ray tube, the detector, or both. In the images air and cerebrospinal fluid (CSF) (which have low attenuations) are typically black; bone (which has high attenuation) is typically white; and brain tissue is a shade of gray. Dyes are sometimes used to allow better visualization of blood vessels for examination of lesions such as aneurysms. CT scans have difficulty imaging brain near bony regions, such as the basal frontal lobes. CT scans have relatively poor gray-white matter differentiation but are also relatively inexpensive.

MRI Imaging MRI is based upon the fact that certain atomic nuclei that are magnetically active (when they have an odd number of protons or neutrons) can absorb energy from radiofrequency waves that are tuned into the natural frequency of those nuclei and can reemit the absorbed energy when the radiofrequency waves are turned off. This is analogous to a vibrating tuning fork stimulating a stationary tuning fork if it shares the same resonant frequency. MRI was discovered by Edward Mills Purcell of Harvard and Felix Bloch of Stanford in 1946. In 1973, Paul Lauterbur of the State University of New York in Stonybrook suggested spatial encoding of MRI measurements, which allowed the development of visual images from MRI data.

To prepare the magnetically active nuclei to receive the energy, the nuclei must first be aligned with an externally imposed magnetic field, like a compass with the Earth's magnetic field. The magnetically active nuclei oscillate with a characteristic frequency referred to as the Larmor frequency. The frequency increases with stronger magnetic fields. There is a net magnetization vector, which is called the longitudinal vector when it is aligned with the external magnetic field. When the radiofrequency waves strike the net magnetization vector, the vector tilts away from the longitudinal vector toward the transverse vector as energy is absorbed from the radiofrequency waves. When the radiofrequency waves are stopped, the net magnetization vector relaxes back toward the longitudinal vector and releases the absorbed energy as radio waves. The time required for 63 percent recovery is called the T1 relaxation time or the spin-lattice relaxation time; it takes approximately 3 seconds (3000 ms) for water. Fat has a very short T1 time (approximately 150 ms); white matter has a moderate T1 time (approximately 750 milliseconds); and gray matter has a little longer T1 time (approximately 900 ms). T1-weighted images show CSF as darker than brain tissue and can provide good anatomical images (Fig. 2.13-1C). T1-weighted images are produced by spin echo pulse sequences with short repetition time (TR, e.g., 530 ms) and short echo time (TE, e.g., 15 ms).

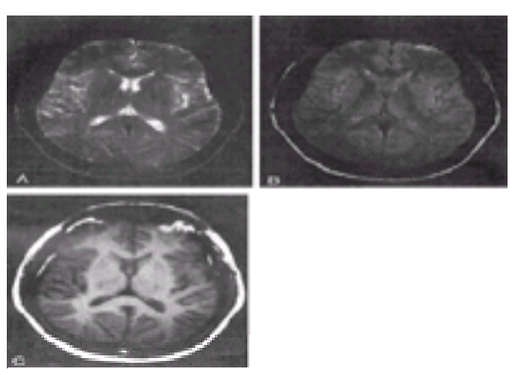

FIGURE 2.13-1 Normal MRI of a healthy elderly man: T2 weighted **(A)**, intermediate **(B)**, and T1 weighted **(C)**.

When the radiofrequency waves strike the magnetically active nuclei, they tilt toward the transverse vector, with synchronization of the phase of precession (i.e., with the magnetic nuclei spinning in alignment, or phase, around the axis like a group of gyroscopes spinning in the same direction). When the radiofrequency waves are stopped, the transverse vector disappears as the protons dephase. This is called spin-spin relaxation, and the time at which 63 percent of the transverse vector disappears is referred to as T2. Typically T2 is one-fifth to one-tenth of T1. For example, CSF has a T2 of approximately 300 ms whereas brain typically has a T2 of 70 ms. CSF is light and brain tissue is dark on a T2-weighted image (Fig. 2.13-1A). Brain lesions with high fluid content are also brighter than normal brain tissue on a T2-weighted image. T2-weighted images are visualized with spin echo pulse sequences with long TR (e.g., 3000 ms) and long TE (e.g., 100 ms).

MRI scans are more expensive than CT scans, but they can provide more information regarding gray and white matter distinctions. This can be helpful with white matter diseases such as multiple sclerosis. Bone does not affect MRI scans the way it can adversely affect CT imaging of brain regions such as the anterior area.

FUNCTIONAL NEUROIMAGING TECHNIQUES

PET Imaging PET scans rely on a variety of positron-labeled chemicals, including a glucose tracer called 18-fluorodeoxyglucose (FDG) and a blood flow tracer (e.g., oxygen-15 H_2O) to provide images of brain function or brain activity as opposed to the structural images provided by MRI. Other radioligands used in PET include 18-fluorodopa (FDOPA), a tracer that has been used to study presynaptic dopaminergic activity, and receptor ligands that are positron labeled (e.g., carbon-11 or fluorine-18) and can be used to study, for example, dopamine (e.g., raclopride, *N*-methylspiperone) and serotonin (*N*-methylspiperone) receptors. A positron emitter is created by a cyclotron via high-speed bombardment of a target material. *Positron* is a contraction for “positive electron.” Since electrons primarily carry a negative charge, a positive electron has a charge opposite that of normal electrons and is regarded as a form of antimatter. Antimatter cannot survive for long in a normal matter universe because as soon as it meets its corresponding oppositely charged matter, the two particles undergo what is referred to as annihilation decay, whereby

the mass in the matter and the antimatter particles are converted into pure energy. The positron and electron are converted into two gamma rays that emerge at a 180° angle from each other. The PET scanner is composed of crystal detectors that are tuned to these gamma rays and emit a photon when bombarded. The photons are detected by photomultiplier tubes and amplified so that a measurable signal can be generated from the positron-electron annihilation decay.

The use of positron-labeled chemicals to localize brain tumors was first described over 40 years ago in 1953 by Gordon Lee Brownell and R. D. Sweet. The first positron scanner was described in 1962 by William Higgenbotham and his colleagues. The first tomographic PET scanners were described in 1974 by Zang Hee Cho. PET scanners have improved in resolution capabilities since they were first invented approximately 30 years ago. The first scanners had a resolution of about 15 to 20 mm, whereas current state-of-the-art scanners have resolutions closer to 4 mm, which is approaching the theoretical limit. Scans can be viewed from transverse, sagittal, and coronal views, as well as with three-dimensional reconstructions. Scan images can be displayed using a color scale or a gray scale. Most clinical reports rely upon visual examination and comparison of a scan with previously studied scans to create a report that describes qualitative differences if any. In addition to these qualitative methods, quantitative methods that rely upon creating statistical probability maps, which create a z-transform of each pixel in a normalized image that is compared with a database, have been used by some centers to complement the qualitative approach. Quantitative methods can also be based upon measurements within defined brain regions of interest. Brain regions of interest can be defined by automated algorithms or can be individually drawn for specific anatomical regions.

Single Photon Emission Computed Tomography Brain SPECT imaging is a nuclear medicine technique that uses minute doses of radioactive isotopes (such as iodine-123 or technetium Tc-99m) bound to neurospecific pharmaceuticals (e.g., ^{123}I-amphetamine or ^{99}Tc-hexamethylpropyleneamineoxime [HMPAO]) to study cerebral perfusion and thus, indirectly, brain metabolic activity. SPECT can provide clinically useful information in several areas relevant to neurology and psychiatry. Currently enough published articles and extensive clinical experience exist to substantiate its use in evaluating head trauma, dementia, cerebral vascular disease, and temporal lobe epilepsy. Refractory suicidal behavior and substance abuse are other disorders about which SPECT may provide useful information. Unfortunately, the interface between psychiatry and nuclear medicine is uncommon, and SPECT is perhaps underused in clinical practice.

With continued improvement in technology, brain SPECT image resolution is nearing the same realm as PET at less cost. SPECT scans are relatively easy to perform, requiring only an intravenous line, and with multiheaded cameras the acquisition time for the scan is approximately 15 minutes. High-resolution SPECT cameras are also widely available. Unfortunately, at the time of this writing few trained physicians have extensive experience in reading brain SPECT images for neuropsychiatric indications. Some clinicians have reservations about the use of SPECT in children and teenagers, feeling it may be unsafe. The radiation exposure from SPECT, however, is about the same as that with a brain CT scan or an abdominal X-ray. As a specialty, nuclear medicine has been doing studies on children for over 30 years without identified untoward effects. Having an unresponsive or poorly understood psychiatric condition has many more risks than the risk of low-dose radiation exposure from a SPECT scan.

Brain SPECT studies can be displayed in a variety of ways. Traditionally, the brain can be examined in three different planes: horizontally, coronally, and sagittally. A normal study shows relatively symmetrical uptake of the isotope throughout the brain. Overactive areas of the brain are indicated by an increase in color (depending on the color scale used), and those with decreased activity are shown by diminished color. With advancements in computer technology the brain can be represented in three-dimensional images. Three-dimensional surface images are one way to represent brain activity, looking at the cerebral metabolism or blood flow of the brain's cortical surface. These images are helpful in evaluating strokes, brain trauma, effects from drug abuse, and so forth. A normal three-dimensional surface scan shows good, full, symmetrical activity across the brain's cortical surface. Three-dimensional active brain images compare average brain activity with that of the most active parts of the brain. These images are helpful for picking up areas of overactivity, as seen in active seizures, obsessive-compulsive disorder, and bipolar I disorder. Physicians are usually alerted that something is wrong in one of three ways: (1) too much activity in a certain area; (2) too little activity in a certain area; or (3) asymmetrical areas of activity that ought to be symmetrical or symmetrical areas of activity that are typically asymmetrical.

Magnetic Resonance Spectroscopy Magnetic resonance spectroscopy (MRS) measures the concentration of different chemicals in the brain by examining the chemical shift that occurs in different chemical environments associated with different structural moieties. The chemicals to be measured must have a concentration of at least 1 mmol per liter. MRS requires more magnetic field homogeneity than does conventional MRI. There are several different scanning protocols that are used such as stimulated echo-mode acquisition method (STEAM) and image selected in vivo spectroscopy (ISIS). The two most common forms of MRS are proton spectra and ^{31}P spectroscopy. Proton spectroscopy allows measurement of chemicals such as *N*-acetylaspartate (which is used to assess neuronal function), glutamine, glutamate, creatinine, choline, and inositol. *N*-acetylaspartate is used to assess neuronal function. The resolution for proton spectroscopy approaches 7 mm. ^{31}P spectroscopy allows measurement of adenosine triphosphate (ATP), phosphocreatinine, inorganic phosphate, phosphomonoesters, and phosphodiesters. Larger voxels are required for ^{31}P spectroscopy, and thus spatial resolution is lower. MRS provides a noninvasive way of measuring brain chemicals but has poor spatial and temporal resolution and is largely a research tool at this time.

Functional MRI Functional MRI (fMRI) allows assessment of T_2 dephasing due to changes in blood flow. Deoxygenated hemoglobin can be used as a naturally occurring contrast agent compared with oxygenated hemoglobin. Oxygenated areas have a larger signal than deoxygenated regions. Functionally activated brain regions have increased blood flow with greater oxygenation. This technique is called BOLD (blood oxygen level dependent) imaging. High-field magnets allow faster acquisition of the fMRI signals.

DEMENTIAS

Dementia or mental decline associated with aging occurs in approximately 11 percent of the population of the United States over 65 years of age and affects 2 million Americans over the age of 50. The most common form, dementia of the Alzheimer's type, accounts for approximately 70 percent of dementias. The second most common form of dementia is vascular dementia due to many small strokes. The third most common form of dementia is associated with degenerative diseases such as Huntington's disease and Parkinson's disease. The differential diagnosis of dementias now requires the use of imaging techniques to rule out treatable causes of dementias. Structural imaging techniques are primarily used by clinicians. Functional imaging techniques are more sensitive for detecting features that would help confirm the diagnosis of conditions such as Alzheimer's disease, but they are less commonly used because of issues of equipment availability, staff training, and cost containment.

Alzheimer's Disease Alzheimer's disease, the most common degenerative brain disease, results in progressive memory loss and accounts for over two thirds of dementia. Definitive diagnosis of Alzheimer's disease requires a neuropathological examination to determine whether the characteristic neurofibrillary tangles and senile plaques are present. These findings occur primarily in the medial temporal cortex and parietotemporal cortex in the early phase; in later phases the frontal lobe is also involved. Primary sensory cortex (e.g., visual cortex, somatosensory system) shows relative sparing. The onset of the disease is usually insidious and includes forgetfulness, confusion, and poor judgment. Recent memory consolidation is affected whereas long-term memory recall is initially spared. As the disease progresses, patients often experience language difficulties with failure to recall words. In later stages, patients may have difficulty forming complete sentences and with visuospatial orientation. The differential diagnosis includes treatable forms of dementia such as normal pressure hydrocephalus, nutritional deficiencies (e.g., Wernicke-Korsakoff syndrome), drug intoxication (e.g., barbiturates), brain tumors, and pseudodementia of depression.

CT and MRI CT scans are useful for screening for potentially treatable conditions that can present as dementia such as subdural hematoma, normal pressure hydrocephalus, and brain tumors in frontal or temporal regions. Alzheimer's disease CT findings overlap with those of normal control subjects and show cortical atrophy (especially in the sylvian fissure) and enlarged ventricles (especially in the anterior temporal horns). MRI T1-weighted images show generalized atrophy (as does the CT scan) but are more sensitive at showing reduced volumetric size of medial temporal regions such as amygdala and hippocampus. MRI T1-weighted images also show enlarged ventricles. MRI T2-weighted images show no difference between subcortical hyperintensities in Alzheimer's disease and normal controls.

PET and SPECT PET scans of Alzheimer's dementia patients typically show bilateral metabolic decreases in the temporoparietal association areas as well as decreases in the frontal lobe. The occipital and primary sensorimotor cortex, striatum, and cerebellum show sparing. Asymmetries are not uncommon in Alzheimer's disease. Left hemisphere decreased activity is associated with language impairment, whereas right hemisphere decreased activity is associated with greater visuospatial orientation impairment. PET scans allow high accuracy in the differentiation of normal subjects from patients with dementia of the Alzheimer type, especially in moderate and severe cases. Medications for the treatment of Alzheimer's disease (e.g., tacrine [Cognex]) have been shown to reverse the decrease in metabolism. The SPECT pattern for Alzheimer's disease is typically bilateral hypoperfusion in the parietal and temporal regions of the brain with frontal lobe hypoperfusion occurring later in the illness.

Patient A is a 70-year-old man who had been more forgetful to the point that his family was worried about him. The patient's family was interested in getting a diagnostic workup to evaluate the possible causes for his memory disorder. His PET scan showed that he had marked parietotemporal decreases (Fig. 2.13-2), which corroborated other neurological evaluations suggesting that the patient had Alzheimer's disease. The patient was treated with tacrine and benefited from some stabilization of his symptoms.

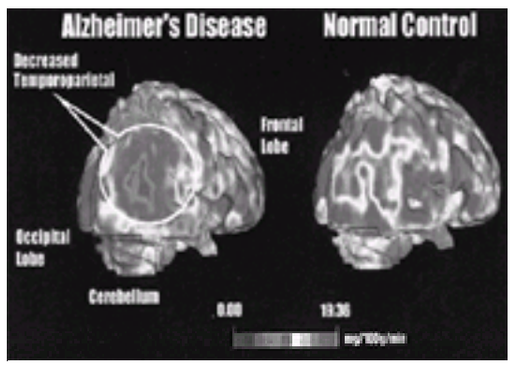

FIGURE 2.13-2 Three-dimensional PET FDG images demonstrate markedly lower glucose metabolism in temporoparietal region in patient with Alzheimer's disease than in normal control. (See Color Plate 3.)

Vascular Dementia Vascular dementia can result from multiple small infarcts that generally involve white matter lesions in deep cerebral locations. On T2-weighted images, these lesions show up as hyperintensities. These changes can be seen in normal elderly individuals, although not to the same extent as in individuals with vascular dementia. Over half of normal control subjects over age 70 have subcortical hyperintensities that typically occur in periventricular regions. The pathophysiology of these hyperintensities is not well understood but has led to theories such as leakage from vessels secondary to hypertensive damage.

Vascular dementias have a deteriorating course characterized by sudden onset with focal neurological signs and showing small stepwise decrements due to multiple small infarcts as opposed to the more insidious course of Alzheimer's disease. Lesions in specific regions of brain result in functional changes associated with that region (e.g., infarction in dominant angular gyrus can result in visuospatial defects, acalculia, agraphia). The size of the vascular lesion may be a key factor in whether or not dementia results, with a critical threshold involvement of approximately 100 cm^3 total of brain tissue.

The PET scan pattern of vascular dementia is characterized by scattered multifocal cortical and subcortical regions of hypometabolism, with metabolic defects exceeding the boundaries of the infarcted tissue. This scattered pattern differs from that of Alzheimer's disease, which is primarily localized in association cortex. The SPECT pattern in vascular dementia is characterized by multiple areas of decreased perfusion. Neuropsychological deficits correlate better with brain metabolism abnormalities found with functional neuroimaging than with white matter lesions or lacunar infarcts seen on MRI. Treatments such as propentofylline or hydergine have been found to increase brain metabolism in some studies.

Pick's Disease Patients with Pick's disease show cognitive deterioration similar to that of patients with Alzheimer's disease. It is a relatively rare cause of dementia that typically occurs before age 65, and women are more often affected than men. Patients often show greater impairment in social judgment with greater behavioral disinhibition than do patients with typical Alzheimer's disease. Apathy is a common early symptom; another symptom is difficulty performing sequences of motor tasks. In postmortem studies, argentophilic (Pick's) bodies are found in only 20 percent of patients. The atrophy encompasses both gray and white matter and was termed *lobar atrophy*, compared with Alzheimer's disease, which shows relatively greater involvement of gray matter. CT and MRI scans show frontal and temporal atrophy including white matter. Patients with Pick's disease differ from those with Alzheimer's disease by showing decreases on PET or SPECT scans primarily in frontal and temporal lobe in early stages.

Lewy Body Dementia Lewy body dementia is characterized by dementia and parkinsonian-like symptoms with fluctuations in cognitive and motoric symptoms, hypersensitivity to antipsychotic drugs, orthostatic hypotension, and psychotic symptoms. Patients with Lewy body dementia show a functional brain imaging pattern that is similar in many respects to that seen in Alzheimer's disease, with a pattern of hypometabolism in frontotemporoparietal regions. Some studies have found that Lewy body disease differs from Alzheimer's disease in showing metabolic decreases in occipital regions, which are characteristically spared in Alzheimer's disease. As with Parkinson's patients, Lewy body dementia patients show increased striatal activity on the side contralateral to parkinsonian symptoms.

Jakob-Creutzfeldt Disease Jakob-Creutzfeldt disease is caused by a prion, a proteinaceous infectious agent, and typically presents at approximately 60 years of age and then declines in incidence with age so that it is relatively rare after 80. The symptoms include rapid memory loss, myoclonic movements, cerebellar ataxia, delirium (e.g., hallucinations, confusion) and visuospatial disorders. The neuropathological findings include spongiform changes resulting in widespread vacuolization. MRI scans may be normal initially but then may show hyperintensities on T2-weighted images in striatum and cortex. The PET scans of patients with Jakob-Creutzfeldt disease show marked decreases in global regional brain metabolism.

Acquired Immune Deficiency Syndrome (AIDS) Dementia Complex Chronic human immunodeficiency virus (HIV) encephalitis can present as a form of dementia characterized by cognitive decline and motor symptoms such as ataxia, limb incoordination, leg weakness, and bowel incontinence. Patients may also have frontal release signs, seizures, or parkinsonian symptoms. Psychiatric symptoms such as depression or psychosis can sometimes be the first signs of this AIDS dementia complex. Neuropathological examinations find nonfocal abnormalities in cerebral white matter, and nearly 90 percent of AIDS patients show some neuropathological abnormalities. Twenty-five percent of patients have peripheral neuropathy in association with dementia. In the final stages, AIDS dementia complex can result in akinetic mutism with urinary incontinence.

CT or MRI scans can show ventricular enlargement with cortical atrophy, and T2-weighted MRI scans can show hyperintensities. With functional brain imaging, patients with AIDS dementia complex have been found to have hypermetabolism of the basal ganglia and limbic regions. In cortical regions, adult patients with AIDS dementia complex have been reported to have increased metabolism in frontal, temporal, and parietal cortex, whereas children with HIV have been reported to have diffuse hypometabolism in right temporo-occipital regions. Treatment with zidovudine (Retrovir) may improve functional brain imaging abnormalities.

Focal brain masses occur commonly in AIDS patients and include tumors (e.g., lymphomas), infections (e.g., cryptococcomas, toxoplasmosis), and progressive multifocal leukoencephalopathy. MRI scans are generally more sensitive at detecting these lesions than CT scans. MRS shows some promise in differentiating lesion types, with infections showing increased lactate concentrations and tumors showing increased choline concentrations. SPECT images in dementia due to HIV disease often show patchy areas of decreased activity across the cortex.

Normal Pressure Hydrocephalus The major clinical difference between normal pressure hydrocephalus and Alzheimer's disease is that normal pressure hydrocephalus has the classical triad of gait instability, bowel incontinence, and dementia. There are no signs of headache or papilledema, and normal CSF pressures are noted with lumbar punctures. In advanced stages, normal pressure hydrocephalus is associated with frontal release signs. Normal pressure hydrocephalus often occurs after head trauma, meningitis, or subarachnoid hemorrhage but may not be manifested for weeks. CT or MRI scans show lateral ventricular dilatation but normal or compressed cortical sulci. T2-weighted images show periventricular hyperintensities, but this occurs in vascular dementia as well. Nuclear medicine studies of CSF flow can show reversal of the normal ventricular to subarachnoid space flow. The PET scan of normal pressure hydrocephalus looks very similar to that of Alzheimer's disease; however, treatment of normal pressure hydrocephalus with ventricular shunts reverses the metabolic defect.

MOVEMENT DISORDERS

Parkinson's Disease Parkinson's disease occurs in middle to late life and is characterized by progressive bradykinesia-hypokinesia and other symptoms such as resting tremor, rigidity, flexed posture, loss of postural reflex, and freezing phenomenon (motor block). On neuropathological examination, patients with Parkinson's disease show a decreased number of neuromelanin-containing neurons in the substantia nigra, thus resulting in decreased dopamine release in the striatum. These substantia nigra neurons sometimes contain Lewy bodies. MRI scans can show a decrease in size of the substantia nigra, and T2-weighted images show hypointensity of the putamen in some patients.

Functional neuroimaging of patients with Parkinson's disease shows hypermetabolism in the striatum, primarily in the putamen. Treatment with pallidotomy reduces and normalizes striatal metabolism. More-advanced Parkinson's disease patients have decreases in other neocortical areas such as the frontal cortex. Assessment of brain metabolism can help provide diagnostic corroboration, especially prior to invasive treatments such as pallidotomies. Depressed parkinsonian patients have

caudate and orbitofrontal hypometabolism and sometimes medial frontal and cingulate hypometabolism as well.

Patient G is an elderly patient with increasing rigidity and tremor who initially had a good response to antiparkinsonian medication but now has increasingly less response to medications, with severe symptomatic fluctuations. He is a candidate for pallidotomy. To confirm the diagnosis of treatment-resistant Parkinson's disease (as opposed to a mimicking illness such as multiple system atrophy), a PET scan is obtained. The patient has hypermetabolism in the striatum and decreased frontal lobe activity (Fig. 2.13-3) which confirms that the patient has true Parkinson's disease. The patient undergoes a pallidotomy, which restores the capacity to respond favorably to antiparkinsonian medication.

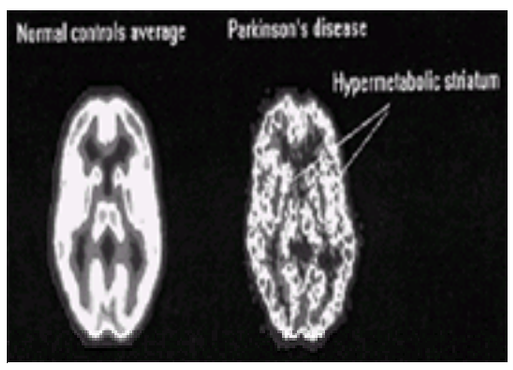

FIGURE 2.13-3 Transaxial section of PET FDG demonstrates hypermetabolism in striatum in patient with Parkinson's disease compared with normal control average image. (See Color Plate 3.)

Huntington's Disease Huntington's disease is the leading cause of adult-onset chorea, which is an irregular, rapid, involuntary movement that moves from one part of the body to another. It is a hereditary disorder with a frequency of 4 to 5 per million individuals. Typically, symptoms start in the fourth or fifth decade of life, although it can occur as early as childhood. Huntington's disease is transmitted by an autosomal dominant gene with complete penetrance. The gene for Huntington's disease has been located on the short arm of chromosome 4.

The initial symptoms may be behavioral, such as becoming very irritable, critical, or depressed. Patients show gradual deterioration of cognitive function with impaired attentional function. The first signs of motor abnormality involve the upper extremities and face (e.g., difficulty performing sequential actions). On neuropathological examination, there is a pronounced atrophy of the caudate, with loss of g-aminobutyrate (GABA)-ergic striatal efferent neurons to the lateral globus pallidus.

CT or MRI scans can show a bulge in the lateral ventricle reflecting loss of the caudate head resulting in a butterfly appearance of the lateral ventricles. T2-weighted MRI scans can show striatal hyperintensity in some forms of Huntington's disease. The most characteristic feature of regional brain metabolism with PET scans in Huntington's disease is hypometabolism in the caudate and the putamen. PET scans are more sensitive than SPECT scans for detecting striatal hypometabolism in Huntington's disease. PET scans are also more sensitive than CT scans or clinical examinations in detecting Huntington's disease. Decreased striatal metabolism is seen in presymptomatic individuals before development of caudate atrophy. Caudate metabolism is significantly correlated with bradykinesia and overall function in Huntington's disease. Gene-positive Huntington's disease patients show an annual decrease of approximately 3 percent of caudate metabolism.

Other Movement Disorders

Progressive Supranuclear Palsy Patients with progressive supranuclear palsy have symptoms such as akinetic parkinsonian symptoms, gait ataxia, pseudobulbar atrophy, and supranuclear palsy affecting vertical gaze. Patients have an intact reflex downgaze during oculocephalic maneuver despite being able to look down voluntarily. Patients typically exhibit tiredness and a tendency to fall backward. There is no familial connection. Cognitively, progressive supranuclear palsy patients show impaired verbal fluency and difficulty shifting from one task to another. CT or MRI scans show atrophy of the pons and midbrain. Progressive supranuclear palsy patients show decreases in frontal areas and striatal metabolism. FDOPA PET scans show decreased presynaptic dopamine uptake in caudate and putamen.

Hemiparkinsonism-Hemiatrophy Patients with hemiparkinsonism-hemiatrophy have unilateral motor symptoms and show decreased metabolism in the basal ganglia contralateral to the parkinsonian symptoms. Unlike Parkinson's disease, there is generally a more benign clinical course.

Multiple System Atrophy Multiple system atrophy comprises several similar disorders such as striatonigral degeneration, olivopontocerebellar syndrome, and Shy-Drager syndrome. Patients with multiple system atrophy have oligodendroglial microtubular tangles in white matter, which can result in neuronal loss and gliosis in different parts of the brain, including the neostriatum, substantia nigra, cerebellum, inferior olives, and anterior horn cells, thereby resulting in different clinical manifestations. Primary involvement of the neostriatum, referred to as *striatonigral degeneration*, results in parkinsonian symptoms without tremor and poor response to levodopa (Larodopa). Involvement of the preganglionic sympathetic neurons results in a condition known as *Shy-Drager syndrome*, with parkinsonian symptoms plus autonomic dysfunction (e.g., orthostatic hypotension, impotence, bladder and bowel dysfunction). If the degeneration occurs in the olives, pons, and cerebellum, then common associated symptoms include ataxia in addition to the parkinsonian symptoms. The most striking finding on functional imaging is the decreased metabolic activity in the striatum.

Corticobasal Degeneration Parkinsonian symptoms associated with apraxia are characteristic of corticobasal degeneration. PET scans show hypometabolism in the cortex, especially the parietal lobule, hippocampus, thalamus, and basal ganglia contralateral to the affected limb.

Hypoparathyroidism Hypoparathyroidism can result in parkinsonian symptoms or athetosis. This syndrome may occur after thyroidectomy or can be associated with primary adrenal failure due to autoimmune abnormalities. Calcification of the basal ganglia and cerebellum can occur and is detectable by CT or MRI scans. Some calcification of these structures occurs with normal aging also.

Hallervorden-Spatz Disease Hallervorden-Spatz disease is caused by excessive iron deposition in the globus pallidus, substantia nigra, and red nucleus. The symptoms include corticospinal signs (e.g., spasticity) and extrapyramidal signs (e.g., rigidity, dystonia). CT scans reveal hypodensities in the lenticular regions, and T2-weighted MRI scans show hypodensities in the pallidum except for a small area of the medial segment, thus known as the “eye of the tiger” sign. Parkinson's disease and striatonigral degeneration can also show increased iron concentration in the basal ganglia.

Hepatolenticular Degeneration (Wilson's Disease) Wilson's disease is due to abnormal metabolism of copper leading to deposition of copper in the liver, resulting in hepatitis and cirrhosis. The neurological manifestations of Wilson's disease include tremor and slowness with bulbar muscle dysfunction resulting in impairments of swallowing and speech. Emotional lability and cognitive impairments follow as the disease progresses. Characteristic corneal rings are golden brown (Kayser-Fleischer). Laboratory examinations find low serum ceruloplasmin concentration. CT scans show enlarged ventricles, brainstem atrophy, and lenticular nuclei hypodensity. MRI scans are more sensitive at detecting these abnormalities. PET scan studies of Wilson's disease show a diffuse decrease in metabolism, with marked hypometabolism in the lenticular regions and relative sparing of the thalamus.

CEREBROVASCULAR DISEASE

Strokes are the third most common cause of death (after cardiovascular disease and cancer) and are characterized by a sudden onset of an acute focal neurological event that lasts longer than 24 hours. Approximately 500,000 new or recurrent strokes occur each year, with approximately 175,000 deaths per year. Hypertension, smoking, obesity, and dyslipidemias are major risk factors for strokes. Several subtypes of stroke exist including thrombotic, hemorrhagic, embolic, and venous occlusion. Approximately 75 percent of strokes involve the internal carotid or middle cerebral artery. Embolic strokes have an immediate onset of peak symptoms. Thrombotic strokes also begin acutely but may evolve step-like over a period of time ranging from seconds to days. Symptoms such as headache or vomiting are more likely to be associated with hemorrhagic strokes than embolic strokes. Hemorrhagic strokes also have an acute beginning but then progress steadily over minutes to hours. Typically after the stroke, symptoms start to resolve gradually. The complete middle cerebral artery strokes involve the basal ganglia with a wedge-shaped region of infarction from the lateral ventricle to the cortex. Strokes are commonly preceded by transient ischemic attacks, whose symptoms often

include hemiplegia, hemianopia, diplopia, confusion, or spatial disorientation.

CT scans are often the initial imaging procedure used and can exclude conditions that mimic stroke including subdural hematomas or tumors. In the hyperacute stage (<6 hours), there may be only subtle abnormalities including mild parenchymal hypointensity. During the acute stage (12 to 24 hours), nonenhanced CT scans can show loss of gray-white matter interface and hypodensity in regions such as the basal ganglia. In the late acute stage (24 to 72 hours), nonenhanced CT scans can show wedge-shaped hypodensities with possible hemorrhagic transformation. Enhanced CT scans may highlight sites of infarction. In early subacute stage (4 to 7 days), mass effect from enlargement of the affected area can persist. In late subacute stages (1 to 8 weeks), the mass effect can resolve, although gyral enhancement may persist. In chronic stages, there may be atrophy and encephalomalacia with ventricular or sulcal enlargement.

MRI scans have greater sensitivity than CT scans, especially in the basal ganglia region, brainstem, and subcortical white matter. MRI scans can also visualize small arteriovenous malformations, angiomas, and lacunar infarcts sooner and with greater sensitivity than CT scans. In the immediate stage, T1-weighted MRI scans can show intravascular contrast enhancement. In the hyperacute stage (1 to 6 hours), T1-weighted MRI scans show gyral swelling and loss of gray-white matter interfaces. In the acute state (6 to 24 hours), T1-weighted images can show mass effects and meningeal enhancements next to the infarct; T2-weighted scans will show increased signal intensity in stroke regions. During the late acute stage (24 to 72 hours), T1-weighted images show increased mass effect, and T2-weighted images show hyperintensities. In early subacute stages (4 to 7 days), T1-weighted images show parenchymal enhancement with contrast, and T2-weighted images show several possible changes including subcortical hyperintensities, hemorrhage, or acute Wallerian degeneration. In late subacute stage (1 to 8 weeks), T-weighted image can show persistent parenchymal enhancement. In chronic stages, T1-weighted images can show atrophy in affected regions and Wallerian degeneration and T2-weighted images can show encephalomalacia.

PET PET scans are useful in assessing strokes and show metabolic disturbances that extend beyond the site of the primary lesions. PET studies of poststroke aphasia reveal that the ability to recruit right posteroinferior temporal activity is associated with recovery. Poststroke neglect is associated with hypometabolism in the hemisphere contralateral to the stroke. Homonymous hemianopsia after a middle cerebral artery stroke is associated with significant and extensive decrease in activity throughout the affected hemisphere.

SPECT SPECT is often helpful in the evaluation and management of cerebral vascular disease. In acute stroke, early SPECT depicts the area of ischemia with greater accuracy than either CT or MRI. When the perfusion defect is large, the likelihood of hemorrhagic complications or herniation increases. Reperfusion of an arterial territory after thrombolysis can be documented more conveniently with SPECT than with angiography. SPECT before and after the injection of acetazolamide has been used to assess the vascular reserve in patients with severe stenosis of the proximal vessels of the cerebrovascular tree. Figure 2.13-4 and Figure 2.13-5 show a typical stroke.

FIGURE 2.13-4 Stroke. Left side surface view shows marked decreased left frontal, temporal, and parietal lobe. (Note: SPECT cortical surface images are rendered by setting the threshold at 55%, looking at most active 45% of brain activity; SPECT active images are rendered by setting blue color threshold set at 55%, looking at average brain activity in the brain compared to red [or white] threshold set at 85%, looking at the most active 15% of brain activity.) (See Color Plate 3.)

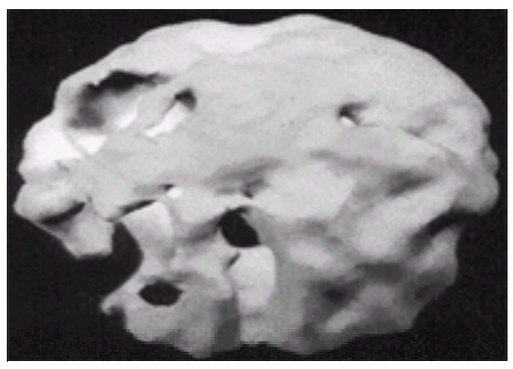

FIGURE 2.13-5 Two right hemisphere strokes with depression as presenting symptom. Top-down surface view shows marked decreased right frontal, temporal, and parietal lobe. (See Color Plate 3.)

Patient N, a 59-year-old woman suffering from severe depression, had been nonresponsive to treatment. A SPECT study was ordered to evaluate her unresponsive condition. Since she had not experienced any neurological symptoms, it was surprising to discover two large right hemisphere strokes. Her nonresponsive depression became clearer, since 60 percent of persons with frontal lobe strokes experience severe depression. As a result of the SPECT study, immediate neurological consultation was obtained and N was placed on blood-thinning medication to prevent further strokes.

MELAS Mitochondrial myopathy, encephalopathy, lactic acidosis and strokelike episodes (MELAS) is one of several mitochondrially inherited disorders. Other examples of this class of disorder includes MERRF (myoclonic epilepsy and ragged red fibers), Kearns-Sayre syndrome, and chronic progressive external ophthalomoplegia. MELAS is characterized clinically by slow growth, stroke-like episodes, seizures, and mental retardation or dementia with onset before age 40. Although the primary deficits are cerebral, muscular deficits can also occur. It is caused by mutations in the mitochondrial deoxyribonucleic acid (DNA). Imaging studies show multifocal infarcts throughout cortex and white matter. CT scans can show confluent hypolucencies consistent with infarcts. T2-weighted MRI images show hyperintensities in both cortical and subcortical regions.

DEMYELINATING DISEASES

Multiple Sclerosis Multiple sclerosis is a chronic neurological illness associated with white matter demyelination and glial scarring (sclerosis) in multiple cerebral locations, with varying symptoms such as weakness, numbness, dysesthesia, optic neuritis, diplopia, ataxia, and psychiatric symptoms that fluctuate over time. Multiple sclerosis typically strikes young adults between the ages of 20 and 40, with a predilection for Northern Europeans and a slightly greater preference for females. Familial vulnerability to multiple sclerosis is inherited. CSF examination indicates that approximately one third of multiple sclerosis patients show moderate pleocytosis and increased protein content.

CT scans can detect lesions in multiple sclerosis, which may appear as contrast-enhanced ring lesions. MRI scans, however, are more sensitive. T2-weighted images show periventricular hyperintensities and white matter lesions that are well demarcated. Atypical presentations of multiple sclerosis on MRI can mimic tumors. Contrast enhancement can be helpful in increasing sensitivity in the MR scan. Gadolinium can aid in estimating the age of a plaque with acute attacks of 8 weeks or less showing enhancement. Lesions are usually at least 3 mm in diameter on MR scan and are located in various places such as spinal cord, brainstem, and cerebellum. There are clinical correlations between the size of the lesions and cognitive impairments. MRS scans show normal *N*-acetyl asparatate (NAA) in acute

lesions but decreased NAA in chronic lesions.

Patients with multiple sclerosis show lower global metabolism on functional brain imaging studies. Multiple sclerosis patients with fatigue show reduced metabolism in the frontal, premotor supplementary motor area, and adjacent white matter and increased metabolism in the anterior cingulate and cerebellar vermis. Active white matter lesions have been reported to be hypermetabolic. Decreases in left frontal and temporal cortical metabolism have also been reported in multiple sclerosis.

Inherited White Matter Disorders (Leukodystrophies) Leukodystrophies are progressive genetic disorders that affect the metabolism of myelin. The most common is metachromatic leukodystrophy, which occurs in 1 of 100,000 births and results from deficiencies of arylsulfatase-A in the lysosome. Initial symptoms can include dementia, gait unsteadiness, progressive spasticity, and rigidity. Metachromatic leukodystrophy can also present initially as a psychiatric disorder with symptoms similar to those of schizophrenia. CT scans typically show generalized atrophy with symmetrical hypodensities in periventricular regions. T2-weighted images show symmetrical hyperintensities in periventricular white matter with cerebellar involvement. White matter in the basal ganglia region is typically not affected in the early stages of the disease.

Adrenoleukodystrophy is related to an X-linked recessive gene near chromosome region Xq28 and seen only in males. Typically, adrenoleukodystrophy is first seen between the ages of 4 to 10 years. The first symptoms can include such brain-related symptoms as cognitive decline and personality changes. Later symptoms can include pseudobulbar paralysis, cortical blindness, hemiplegia, and ataxia. Signs of adrenal insufficiency can also occur and can precede neurological signs. Laboratory findings are low serum and chloride concentrations and high potassium concentrations. CT scans typically show symmetrical hypodensities in the corpus callusom splenium and possibly puncate calcifications. MRI scans show posterior white matter involvement in 80 percent of patients and anterior white matter involvement in 15 percent of patients. MRS shows decreased NAA concentration and an increase in choline-containing compounds.

Krabbe's disease (globoid cell leukodystrophy) involves a deficiency of galactocerebroside-b-galactosidase, with several forms of clinical presentations. The adult form is characterized by dementia, pyramidal signs, and optic atrophy. CT scans show hypodensities in white matters with high attenuation of deep subcortical structures such as thalamus and basal ganglia. T2-weighted MRI scans show nonspecific, symmetrical, confluent hyperintensities that first affect the parietal region and optic pathways.

Leukoencephalopathies There are a variety of inherited leukoencephalopathies that involve subcortical white matter at an early stage including Pelizaeus-Merzbacher disease, Canavan-Van Bogert-Bertrand disease, and Alexander's disease. Spongy degeneration of infancy (Canavan-Van Bogert-Bertand disease) is an autosomal-recessive genetic disorder that is characterized by an enlarged head (macrocephaly) and regression of motor skills, optic atrophy, and corticospinal signs with hypotonia and then spasticity. T2-weighted images show hyperintensities throughout the brain and hypointensities in basal ganglia and thalamus. MRS can show significant increase in NAA.

WERNICKE-KORSAKOFF SYNDROME

Wernicke's encephalopathy, typically associated with chronic alcohol dependence, is characterized by rapid onset of ocular abnormalities (e.g., nystagmus, paralysis of movement), mental confusion (e.g., drowsiness, disorientation, amnesia), and gait ataxia. It is caused by a deficiency of thiamine, which can be reversed in the early stages by administering parenteral thiamine. Neuropathological examination reveals punctate hemorrhage around the third and fourth ventricles involving the thalamus, midbrain, and pons. If left untreated, the syndrome can become fatal or Korsakoff's syndrome can follow. T2-weighted MRI images show hyperintensities that may be enhanced in medial thalamic and periaqueductal regions and atrophy in mammillary bodies. Korsakoff's syndrome patients show decreased activity in frontal regions and in diencephalic (i.e., thalamic) regions.

HEAD TRAUMA

Trauma, the leading cause of death in individuals up to 44 years of age, accounts for almost 10 percent of all deaths. Approximately half of the deaths associated with trauma include traumatic brain injury. Those who do not die are often left with significant disabilities. The injuries most commonly involve motor vehicles. The peak age for brain trauma is young adults between 15 to 24 years of age. Head injuries can be subdivided into closed head injuries, depressed skull fractures, and compound skull fractures. Closed head injuries with no gross structural damage to the brain are called concussions.

Emergency CT scans are indicated for head injuries with loss of consciousness or amnesia, focal neurological abnormality, a Glasgow Coma Scale score of 14 or less, depressed skull fractures, and penetrating injuries and those in persons above 60 or less than 2 years of age. Common types of neurosurgically treatable lesions include acute epidural or subdural hematoma. Epidural hematomas result from laceration of a dural artery (e.g., middle meningeal artery), which collects between the outer layer of the dura and the inner table of the skull. Epidural hematomas on CT or MRI scans are indicated by a biconvex extraaxial mass. Subdural hematomas usually occur when sudden deceleration causes tearing of cortical veins that cross the subdural space. Subacute subdural hematomas may have the same density as brain tissue and can be missed on CT scan. Chronic subdural hematomas may have a variety of shapes including biconvex or crescent-shaped and may be hypodense. Acute subdural hematomas typically show up as hypointensities on T2-weighted images. Chronic subdural hematomas typically show up as hyperintensities on T2-weighted images, and these lesions are enhanced with contrast. Chronic subdural hematomas can present as psychiatric symptoms including alterations in mood, cognition, and attention.

CT scans are often normal in traumatic brain injury, especially in the early stages. In later stages, CT scans can reflect patchy hypodense or hyperdense regions. MRI scans can be more sensitive than CT scans for assessment of intraaxial brain trauma. T2-weighted images may show multifocal hyperintensities, which may be due to diffuse axonal injury that results in tearing or shearing of axons and penetrating blood vessels.

Head injuries can result in pathological sequelae following the acute injury. Cerebral herniation due to swelling and edema secondary to traumatic brain injury can be life threatening. Other potential complications include ischemia due to compression of vascular cerebral supplies such as the carotid artery or cerebral artery. Head trauma can also produce communicating hydrocephalus after subarachnoid or parenchymal bleeding. Severe head injuries can result in brain death due to severe edema that impedes blood flow to the brain. MR or CT scan can show descending transtentorial herniation. Functional imaging studies such as SPECT or PET scans show profound loss of activity in the brain.

PET or SPECT scans are useful for the assessment of traumatic brain injury. Over 45 articles on functional brain imaging studies of head trauma have been published from 1988 to 1998. Traumatic brain injury can be visualized with PET imaging and is characterized by abnormalities such as decreased metabolism in the frontal or temporal lobes relative to other brain regions. Crossed cerebellar diaschisis is seen in many patients with focal cortical injuries. These abnormalities are correlated with neuropsychological deficits. Damage to the hippocampal regions has been associated with impaired memory. In patients with concurrent evidence of damage on other imaging modalities such as MRI, the corresponding abnormalities on PET show disturbances that extend beyond structurally identifiable borders. Functional brain imaging studies are more sensitive for the detection of abnormalities than MRI or CT scan. Compensatory changes such as increased metabolism in the occipital lobe are often seen in closed head injury.

The impact of head trauma is often overlooked in psychiatry. Even minor injuries to vulnerable parts of the brain can cause problems for years to come. In addition to PET, SPECT is one of the best tools in evaluating functional deficits from head trauma that are often not seen by other studies, leading to more understanding and effective treatments for patients. Typically, SPECT findings in head trauma include focal areas of decreased activity, often in a contracoup pattern (such as decreased activity in the left anterior prefrontal cortex and right occipital lobe or the anterior and posterior aspects of a temporal lobe) and, in some cases, marked hyperactivity over the site of the injury. The authors have seen increased off-center cingulate gyrus activity after a head injury.

Documentation of head injuries is essential for several reasons. It allows school-age children and teenagers to receive more specialized services. Knowledge of the injuries is often essential for legal and insurance reasons. Patient and family understanding of the effects of brain trauma enhances treatment compliance and develops a deeper understanding from family and support systems.

Patient H was in a motor vehicle accident and struck his head on the dashboard. He had a momentary loss of consciousness. His MRI and CT scans were normal, but he complains of severely impaired ability to concentrate and remember and is no longer able to work as an accountant. He has abnormalities on neuropsychological testing including impaired performance on memory tests. A PET scan finds significantly decreased metabolic activity in orbitofrontal cortex (Fig. 2.13-6). His physician recommends rehabilitation for traumatic brain injury to help the patient develop new coping skills to compensate for the brain injury and help both him and his family to readjust their expectations of what he may be capable of achieving.

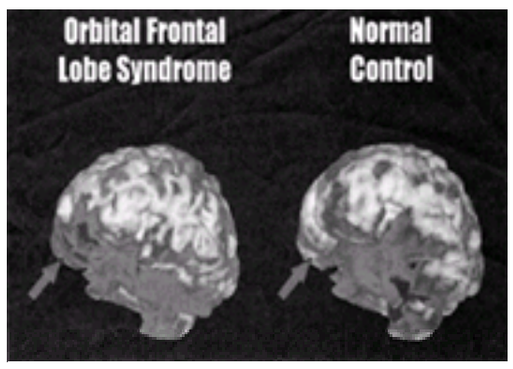

FIGURE 2.13-6 Three-dimensional PET FDG images demonstrate lower orbitofrontal metabolism in orbital frontal lobe syndrome in patient secondary to traumatic brain injury than in normal control. (See Color Plate 4.)

SOLVENT SYNDROME

Toxic exposure to organic solvents (such as toluene) exceeding OSHA limits can result in neurological symptoms (e.g., dizziness, nausea, weakness, incoordination), neuropsychological damage (e.g., cognitive deficits in memory and learning), and psychiatric symptoms (e.g., depression, anxiety) and can present as a significant problem in occupational medicine and neuropsychiatry. CT and MRI scans are typically normal. SPECT and PET brain imaging can display abnormal patterns of functional brain activity in regions such as the hippocampus, frontal and temporal regions which are consistent with predictions from neuropsychological abnormalities.

CARBON MONOXIDE POISONING

Carbon monoxide poisoning is a frequent event which can result in a wide variety of neurological and psychiatric sequelae such as mental deterioration, gait disturbance, and frontal lobe signs. Up to 40 percent of those who survive have permanent neuropsychiatric complications. Neuropathological findings include necrosis in the globus pallidum and possibly in other regions of cerebral cortex. CT scan findings can include cortical atrophy and low density lesions in the globus pallidus although these findings are often not present. SPECT and PET scans are more sensitive than CT or MRI scans at detecting damage and have reported decreased activity in basal ganglia and patchy diffuse hypoactivity in the cerebral cortex (especially in the frontal cortex) of carbon monoxide poisoning patients.

CHRONIC COCAINE USE

Chronic cocaine use has been reported to be associated with widespread vasoconstriction, neurological symptoms such as seizures and strokes; neuropsychological changes (such as abstract thinking, impairment of spatial memory and cognitive flexibility) and neurobehavioral symptoms such as paranoia and depression. CT scan studies have found evidence for atrophy which correlates with drug use. SPECT scan studies have found multifocal defects in frontal and temporal cortex and basal ganglia. PET scan studies have reported decreased frontal activity during withdrawal and a delayed decrease in FDOPA striatal uptake. Magnetic resonance spectroscopy of polydrug abusers have found 15 percent increase in phosphomonoesters and 10 percent decrease in nucleotide triphosphates. A functional MRI study of cocaine infusion reported signal increases in nucleus accumbens, striatum, thalamus, insula, hippocampus, cingulate, and prefrontal and temporal cortices and decreases in medial frontal cortex, amgydala, and temporal pole.

TUMORS

The most common initial symptoms of primary malignant brain tumors are headaches and seizures. Typically, other symptoms such as mental status change or hemiparesis occur before the tumor is diagnosed. An MRI scan is recommended for assessment of seizure onset after age 25. Symptoms consistent with increased intracranial pressure such as periodic bifrontal and bioccipital headaches, projectile vomiting, bowel incontinence, and papilledema can also be an initial presentation for tumors. Brain tumors can also present initially psychiatrically, with symptoms such as depression, nonfluent aphasia, or (if the tumor is located in the frontal lobes) impaired judgment. Visual hallucinations can occur if brain tumors are located in the occipital lobes. Temporal lobe tumor can result in symptoms such as memory impairment and linguistic deficits. Brain tumors account for 2 percent of solid tumors in adults and 15 percent of solid tumors in children. Approximately 70,000 cases of metastatic intracranial tumors and 15,000 cases of primary brain tumors are diagnosed in the United States each year. Approximately 25 percent of all patients with cancer have brain involvement during the course of their illness. Brain tumors are the second leading cause of death from neurological disease, exceeded only by strokes. Tumors located within brain parenchyma are intraaxial, in contrast to tumors located outside, which are called extraaxial. Tumors can have pathological effects by direct infiltration of brain tissue or indirectly through compression and pressure. The most common primary intracranial neoplasms are gliomas, which account for almost half of all primary intracranial tumors. The most common types of gliomas are astrocytomas, which comprise approximately 70 percent of all gliomas. Astrocytomas can vary markedly in terms of prognosis; low-grade astrocytomas have long survival (e.g., approximately 70 percent at 5 years), while high-grade astrocytomas can have low survival (e.g., approximately 10 percent at 2 years). The second most common primary intracranial neoplasms are meningiomas, which are extraaxial tumors. Modalities of treatment include surgery, radiation therapy, and chemotherapy.

CT scans show regions of hypodensity in tumor sites, with mass effects causing changes such as compression of ventricles, effacement of cerebral sulci, or midline shifts. Contrast enhancement can be very helpful in increasing the detection of brain tumors. Nonenhancing tumors are less likely to be aggressive. Calcification can occur in brain tumors and can be detected by CT scans. CT scans are less sensitive for assessing brain tumors in regions next to bone such as brainstem (e.g., cerebellopontine neurofibromas). Primary brain tumors are usually unifocal, whereas metastatic brain tumors are often multifocal. MRI scans are more sensitive than CT scans in detecting smaller tumors and assessing pathological changes such as edema. T2-weighted images show hyperintensities in tumor sites. Gadolinium can be used to enhance contrast in some tumor sites for MRI scans, although all tumors will not necessarily show enhancement. T_1-weighted scans often show decreased signal in tumor sites. Proton MRS shows a decrease in NAA, which is a neuronal marker in tumors. ^{31}P MRS can show increased phosphomonoester concentration in tumors.

Tumors are not the only reason for focal brain mass. Other causes include bacterial abscesses and cysticercal cysts. Cysticereal cysts are common in Mexico and many areas of the third world. CT scans show hypolucency at the site of the lesion, typically surrounded by enhancement. MRI can visualize the cysticercal cyst with greater sensitivity than CT scan.

FDG PET scans currently serve as the gold standard for the differentiation of recurrent brain tumors from tumor necrosis, because changes can occur in irradiated tissues months after the exposure which can be impossible to distinguish with structural imaging approaches such as MRI. However, recurrent tumors are hypermetabolic whereas radiation necrosis is hypometabolic (Fig. 2.13-7). FDG PET scans can be useful for grading tumors and predicting survival and can also be used to assess response to new approaches such as gamma-knife surgery.

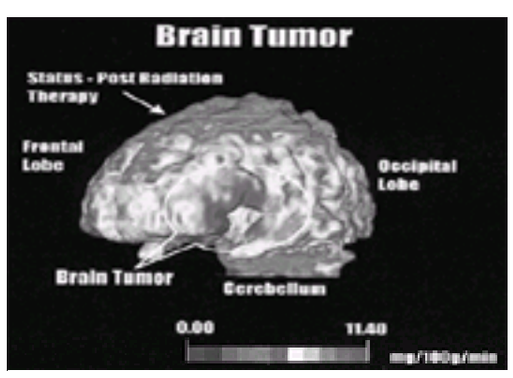

FIGURE 2.13-7 Three-dimensional PET FDG image demonstrate hypermetabolism in region of brain tumor in frontal cortex and decreased metabolism in parietal cortex secondary to postradiation injury. (See Color Plate 4.)

EPILEPSY

Epilepsy is a common neurological condition characterized by self-limited, excessive discharge of cortical neurons resulting in convulsive movements or change in consciousness and mental function. Approximately 2 million individuals in the United States have epilepsy, which is subdivided into two major categories: partial and generalized. Partial seizures usually originate in a circumscribed region of the brain; generalized seizures are associated with more diffuse abnormalities. Progression of seizures can spread and is referred to as a Jacksonian seizure. Another common type of partial seizure is psychomotor or temporal lobe epilepsy. The two major types of generalized seizures are convulsive (grand mal) and nonconvulsive (petit mal).

Clinical evaluation of seizures includes consideration of an extensive differential diagnostic list such as syncope, transient ischemic attack, migraine, and cataplexy. Different causes are associated with different ages of onset. Seizures in late life are associated with vascular disease, brain tumor, abscess, or trauma. Seizures in adolescence and early adulthood are associated with idiopathic epilepsy, trauma, or withdrawal from drugs or alcohol. Seizures in infancy are associated with congenital defects, birth trauma, metabolic disorders (e.g., hypoglycemia, phenylketonuria) or anoxia. Common neuropathological abnormalities include hippocampal sclerosis, although other causes include brain tumors, brain trauma, arteriovenous malformation, or cortical dysgenesis. The clinical workup includes neurological examination, electroencephalogram (EEG), brain imaging (structural such as enhanced CT or enhanced MRI, functional such as SPECT or PET), laboratory analysis of CSF or blood to evaluate for systemic disease (e.g., abnormal electrolyte, calcium, magnesium concentrations) and drug screenings.

MRI scans are more sensitive than CT scans at detecting lesions such as glial tumors, cortical dysplasia, hamartomas, and arteriovenous malformations. Coronal sections perpendicular to the hippocampal long axis may increase the sensitivity for detection of hippocampal lesions such as atrophy or sclerosis. Structural imaging should be done in patients over 18 years of age and in pediatric patients with abnormal development.

PET The pattern for temporal lobe epilepsy is defined by interictal temporal lobe hypometabolism. FDG PET is generally regarded as more sensitive than cerebral blood flow PET at detecting an epileptogenic focus. There is some controversy as to whether proton spectroscopy is more or less sensitive than PET for the detection of a temporal lobe epileptogenic focus.

SPECT Temporal lobe epilepsy is one of the most frequently encountered chronic epileptic disorders, and it is associated with many psychiatric symptoms such as depressive mood, anergia, irritability, euphoric mood, atypical pain, insomnia, fear, and anxiety. The imaging abnormalities in the temporal lobe typically are found in the medial aspects, which are difficult to find with routine EEG studies. SPECT findings in temporal lobe epilepsy and epilepsy in general reveal decreased activity in the interictal period and increased activity during the ictal phase. Proper diagnosis of temporal lobe epilepsy, especially when it presents with psychiatric symptomatology, is essential for effective treatment.

> A 7-year-old boy became very upset when his mother read *Alice in Wonderland.* He said that he felt like Alice. “I have weird things happen to me,” he told her. “I see things.” During the day he saw objects change shapes, often getting smaller. He also saw green, shadowy ghosts at night. He had a lot of anxiety symptoms. His mother brought him for evaluation. He had a normal EEG. A SPECT study was ordered for further evaluation. During the scan he complained of visual changes. The SPECT study showed increased activity in the medial aspects of his right temporal lobe. Within 2 weeks of divalproex (Depakote) treatment the boy's strange experiences disappeared, and over the next 6 months his anxiety significantly lessened.

FUTURE DIRECTIONS

Although PET and SPECT have made tremendous inroads into a better scientific and clinical understanding of brain disorders, there is still room for further growth and studies. These data are being correlated with other clinical variables to develop better prognostic and diagnostic perspectives. New technologies such as MRI are currently used primarily in a research context, and provide the basis for a rapidly evolving body of knowledge. This accumulating knowledge will eventually allow the use of structural and (especially) functional neuroimaging techniques in the diagnosis and assessment of psychiatric and neurological disorders.

SUGGESTED CROSS-REFERENCES

The principles of neuroimaging are discussed in Section 1.15 and Section 1.16, the neuropsychiatric aspects of multiple sclerosis and other demyelinating disorders are discussed in Section 2.7, and medical assessment and laboratory testing are discussed in Section 7.7. The psychiatric aspects of the acquired immune deficiency syndrome are discussed in Section 2.8. The use of neuroimaging in geriatric practice is discussed in Section 51.2e and Section 51.2f.

SECTION REFERENCES

Achten E, Santens P, Boon P, De Coo D, Van De Kerckhove T, De Reuck J, Caemaert J, Kunnen M: Single-voxel proton MR spectroscopy and positron emission tomography for lateralization of refractory temporal lobe epilepsy. AJNR Am J Neuroradiol *19:*1, 1998.

*Adams RD, Victor M, Ropper AH: *Principles of Neurology*, ed 6. McGraw-Hill, New York, 1997.

Barker FG II, Chang SM, Valk PE, Pounds TR, Prados MD: 18-Fluorodeoxyglucose uptake and survival of patients with suspected recurrent malignant glioma. Cancer *79:*115, 1997.

Breiter HC, Gollub RL, Weisskoff RM, Kennedy DN, Makris N, Berke JD, Goodman JM, Kantor HL, Gastfriend DR, Riorden JP, Mathew RT, Rosen BR, Hyman SE: Acute effects of cocaine on human brain activity and emotion. Neuron *19:*591, 1997.

*Datz FL: *Handbook of Nuclear Medicine*, ed 2. Mosby, St. Louis, 1993.

Eidelberg D, Takikawa S, Moeller JR, Dhawan V, Redington K, Chaly T, Robeson W, Dahl JR, Margouleff D, Fazzini E, Przedborski S, Fahn S: Striatal hypometabolism distinguishes striatonigral degeneration from Parkinson's disease. Ann Neurol *33:*518, 1993.

Engel J Jr, Henry TR, Risinger MW, Mazziotta JC, Sutherling WW, Levesque MF, Phelps ME: Presurgical evaluation for partial epilepsy: Relative contributions of chronic depth-electrode recordings versus FDG-PET and scalp-sphenoidal ictal EEG. Neurology *40:*1670, 1990.

Fink GR, Pawlik G, Stefan H, Pietrzyk U, Wienhard K, Heiss WD: Temporal lobe epilepsy: Evidence for interictal uncoupling of blood flow and glucose metabolism in temporomesial structures. J Neurol Sci *137:*28, 1996.

Grafton ST, Mazziotta JC, Pahl JJ, St. George-Hyslop P, Haines JL, Gusella J, Hoffman JM, Baxter LR, Phelps ME: A comparison of neurological, metabolic, structural, and genetic evaluations in persons at risk for Huntington's disease. Ann Neurol *28:*614, 1990.

Gross H, Kling A, Henry G, Herndon C, Lavretsky H: Local cerebral glucose metabolism in patients with long-term behavioral and cognitive deficits following mild traumatic brain injury. J Neuropsychiatry Clin Neurosci *8:*324, 1996.

Heiss WD, Kessler J, Szelies B, Grond M, Fink G, Herholz K: Positron emission tomography in the differential diagnosis of organic dementias. J Neural Transm *33*(Suppl):13, 1991.

Kao CH, Hung DZ, ChangLai SP, Lao KK, Chieng PU: HMPAO Brain SPECT in acute carbon monoxide poisoning. J Nuclear Med *39:*769, 1998.

*Krishnan KRR, Doraiswamy PM, editors: *Brain Imaging in Clinical Psychiatry*, ed 1. Marcel Dekker, New York, 1997.

Kuwabara H, Cumming P, Yasuhara Y, Leger GC, Guttman M, Diksic M, Evans AC, Gjedde A: Regional striatal DOPA transport and decarboxylase activity in Parkinson's disease. J Nucl Med *36:*1226, 1995.

*Lewis S, Higgins N, editors: *Brain Imaging in Psychiatry*, ed 1. Blackwell Scientific, Cambridge, MA, 1996.

Mielke R, Pietrzyk U, Jacobs A, Fink GR, Ichimiya A, Kessler J, Herholz K, Heiss WD: HMPAO SPET and FDG PET in Alzheimer's disease and vascular dementia: Comparison of perfusion and metabolic pattern [see comments]. Eur J Nucl Med *21:*1052, 1994.

*Osborn AG, Tong KA: *Handbook of Neuroradiology: Brain and Skull*, ed 2. Mosby, St. Louis, 1996.

*Perani DC, Stefano F: Neuroimaging methods in neuropsychology. In *Handbook of Clinical and Experimental Neuropsychology*, G Denes, L Pizzamiglio, editors. Psychology Press, Hove, England, 1999.

*Rosenberg RN, editor: *Atlas of Clinical Neurology*, ed 1. Butterworth-Heinemann, Boston, MA, 1998.

Rottenberg DA, Sidtis JJ, Strother SC, Schaper KA, Anderson JR, Nelson MJ, Price RW: Abnormal cerebral glucose metabolism in HIV-1 seropositive subjects with and without dementia. J Nucl Med *37:*1133, 1996.

*Rowland LP, editor: *Merritt's Textbook of Neurology*, ed 9. Williams & Wilkins, Baltimore, 1995.

*Tabrizi SJ, Cleeter MW, Xuereb J, Taanman JW, Cooper JM, Schapira AH: Biochemical abnormalities and excitotoxicity in Huntington's disease brain. Ann Neurol *45:*25, 1999.

Varney NR, Morrow LA, Pinkston JB, Wu JC: PET scan findings in a patient with a remote history of exposure to organic solvents. App Neuropsychol *5:*100, 1998.

Wu JC, Bell K, Najafi A, Widmark C, Keator D, Tang C, Klein E, Bunney BG, Fallon J, Bunney WE: Decreasing striatal 6-foopa uptake with increasing duration of cocaine withdrawal. Neuropsychopharmacology *17:*402, 1997.

CHAPTER 3. CONTRIBUTIONS OF THE PSYCHOLOGICAL SCIENCES

3.1 PERCEPTION AND COGNITION

DANIEL J. SIEGEL, M.D.

The fields of psychiatry and psychology have at their etymological core the term “psyche.” Efforts to understand psychological development, mental processes, psychiatric disturbance, and psychotherapeutic and psychopharmacological interventions will benefit greatly from an open approach to understanding the psyche. The functions of the mind parallel those of the psyche, which can be defined as: (1) the human soul; (2) the intellect; (3) psychiatry—the mind considered as a subjectively perceived, functional entity, based ultimately upon physical processes but with complex processes of its own: it governs the total organism and its interaction with the environment.

COGNITIVE SCIENCE

The fields relevant to this overview are a part of the interdisciplinary studies of cognitive science, which includes anthropology, cognitive psychology, developmental psychology, psycholinguistics, artificial intelligence and computational science, and neuroscience. Each of these disciplines provides an important and unique perspective on how to understand the human psyche. Biological, psychodynamic, and social psychiatry can find a common home and language within cognitive science. The common divisions of nature versus nurture and biology versus psychology disappear when the origins of mental processes are examined.

The last ten years of the twentieth century, called the “Decade of the Brain,” led to discoveries in the neurosciences that revealed a wide range of findings relevant to psychiatry. One of the major discoveries was that the brain's structure and function are a result of the transaction of several factors, including genetic, physiological, and experiential variables. In particular, brain development requires specific forms of experience in order to foster the growth of neural circuits involved in a wide array of mental processes, including attention, memory, emotion, and self-reflection. Thus, experiences shape the unfolding of genetically programmed development of the central nervous system. Genes function both as a template of information and as a mediator of transcription of the proteins that determine neural structure. Experience directly shapes the selection and timing of how the activity of genes influences the structure of the brain. In particular, the nature of the relationship between child and caregiver shapes the emerging growth of areas of the brain that are essential to cognitive and emotional processes.

The human brain at birth is immature compared to the differentiated state of other bodily organs, such as the heart or kidneys. This immaturity requires that the infant and young child's brain utilize the mature state of the adult caregiver's brain in order to organize its own functioning. These findings from developmental neuroscience have raised the important integration of interpersonal relationships and brain development. The cooperative communication of infant-caregiver attachments is thought to provide the building blocks for emotional development as well as for abstract reasoning and cognitive abilities. The patterns of interaction between child and caregiver have a direct impact on the way the brain develops and the mind of the child functions. Thus, cognitive processes need to be considered as the way in which the mind emerges from within the genetic, physiological, and experiential factors that shape the development and maintenance of mental function.

MIND AND BRAIN

A generally accepted view of the mind is that it emanates from a portion of the activity of the brain. What is this activity of the brain and how does it give rise to such mental processes as perception and cognition? How do the human experiences of sensation, thought, emotion, attention, self-reflection, and memory emerge from neural processes?

The brain is composed of about one hundred billion neurons. An average neuron is connected to about ten thousand other neurons at synaptic junctions. Release of neurotransmitters at these junctions promotes or inhibits the firing of the postsynaptic neuron's membrane with an action potential that then sends an electrical signal down its long axon to influence other neurons downstream. With hundreds of trillions of connections among a weblike neural network, there are countless combinations of possible activation profiles. The term *neural net profile* is used to describe a certain pattern of activation of the complex layers of neural circuits.

The neural net profile is the fundamental way in which mental processes are created. These activations can lead to further neural processes in a cascade of dynamic interactions that produce a range of internal events and external behaviors. The essential components of the mind come directly from how these neural events create the flow of energy and information.

ENERGY AND INFORMATION

The mind is a processor of patterns in the flow of energy and information within the brain (Table 3.1-1). Activations of individual neurons, groups of neurons, circuits or networks of neurons all involve the flow of energy through the complex system of the brain. This energy reflects the flow of ions across membranes, the consumption of oxygen and nutrients by neural cells, and the active transport of molecules into and out of nervous tissue. However, the mind is much more than some outcome of energy flow—the function and purpose of this flow of energy is to process information.

The mind is a processor of energy and information.
Energy is contained within the activations of neural circuits.
Information is contained within the patterns of activation termed a *neural net profile* or *mental representation*.
These representations serve as symbols that cause further effects in the mind, leading to the processing of information.

Table 3.1-1 Basic Ideas of the Mind

Information is contained within the brain by a process of representation. A symbol carries information about another object in the world. For example, when an individual sees the Eiffel Tower, the brain responds in particular regions of the visual system with the activation of a neural net profile. When the Tower is recalled at a later time, the visual cortex activates a similar neural net pattern and the Tower is visualized. The activation of a particular pattern of neural firing thus contains within it information about something else (the Tower). Neural net activation profiles thus represent information in the brain. Examples of representational forms include

perceptual, sensory, linguistic, and more abstract representations of concepts and categories.

Computational science reveals that the nervous system can function as a parallel and serial process of information. This processing occurs when a symbol, such as a mental representation, causes an effect. Such effects can be seen as subsequent activations of neural net profiles. Information processing becomes even more complex when the effects themselves carry information. Within cognitive psychology, these information-processing events can be seen as the contrasting, comparing, generalizing, chunking, clustering, differentiating, and extracting processes that lead to a more highly interwoven set of mental representations.

INFORMATION PROCESSING

Several elements of the brain's function as an information processor can be described (Fig. 3.1-1). At the most basic level, (A), energy leads to neural responses. This energy can be in the external form of light on the retina or soundwaves vibrating the tympanic membrane. It may also take an internal form in which the flow of energy within neural activations themselves produce subsequent neural responses.

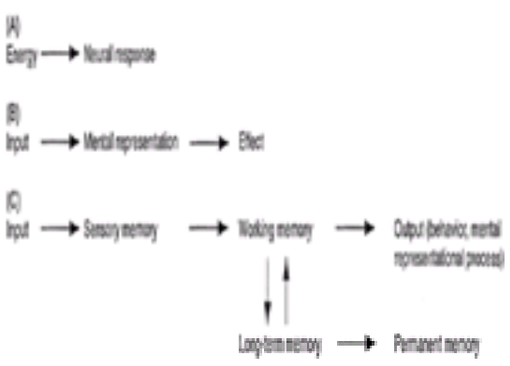

FIGURE 3.1-1 Information-processing models.

A second level of understanding information processing (B) is in the idea that an input (internal or external) leads to a representational response (a neural net profile of activation), which in turn produces a downstream effect or output. This output can be internal, such as the generation of other representations, or external, in the form of observable behavior.

A third level of viewing information processing in the mind (C) is the conceptualization of forms of sensation, perception, attention, and memory. According to this view, external energy is sensed by the peripheral nervous system and registered as sensation within the brain. The selective processing of aspects of these sensations, called “filtering,” leads to the production of perception. These perceptions are themselves subject to further filtering in which only a select few are placed within working memory. This is sometimes called the “chalkboard of the mind” and is thought to be able to handle seven plus or minus two items. It is within working memory that representations can be consciously manipulated, contrasted, clustered, and reassembled. Thus, consciousness may be intimately related to this aspect of mental functioning.

Sensation refers to the initial stages of the basic information-processing model (Fig. 3.1-1) Sensory memory lasts for about one quarter of a second. Items in sensory memory are then filtered into working or short-term memory, where they last for about one half-minute. Working memory is able to handle about seven items unless further processing creates linkages to other items within longer-term memory. Rehearsal allows these representations to remain for longer periods of time. Cognitive processes that can group bits of information into large chunks *(chunking)* can increase the capacity of working memory by making each unit more information-rich. Representations are then processed and placed within long-term memory where they can be retrieved for future use. A process called *cortical consolidation* is thought to involve some form of rehearsal and perhaps rapid eye movement sleep in which items in long-term memory are further processed and placed within permanent memory storage.

ATTENTION

Attention is the process that controls the flow of information processing. There are many aspects to attention that may derive from their neuroanatomic localization. Three components of attention have classically been used to describe certain deficits in psychiatric disorders; future studies may reveal more neurophysiologically based descriptions. These components are selectivity, capacity, and sustained concentration.

Early conceptualizations of attention were based on Donald Broadbent's idea of a filter that selects a limited amount of incoming stimuli to be further processed. Limited capacity of attention was thus attributable to the inability to process the overwhelming amount of incoming stimuli. An attention bottleneck was described as occurring either early in the sensory process and, thus, automatically or late in the perceptual processing stage and involving such processes as identification and classification. More recent views have considered the concept of general cognitive resource capacity, which limits the stimuli processed. These perspectives have been applied to understanding various psychiatric conditions, such as schizophrenia and attention-deficit/hyperactivity disorder.

Selective Attention One aspect of attention is that it focuses a metaphorical spotlight on external stimuli or internal mental representations. In Broadbent's conceptualization, selectivity has three dimensions: (1) filtering, focusing on specific attributes (e.g., large squares versus small squares); (2) categorizing, based on stimulus class (e.g., attending to letters in whatever script they are written); and (3) pigeonholing, reducing perceptual information needed to place a stimulus into a specified category (e.g., using only long hair to classify individuals as female). Each of these aspects of attention acts on incoming stimuli to make a determination of fit for the sought-after characteristic. Schizophrenia patients show greater difficulty with pigeonholing than with filtering when they are symptomatic. Another conceptualization of selective attention distinguishes between two interactive ways of processing sensory input. *Preattentive processing* (a parallel function) assesses global, holistic patterns and appears to be an early component of the perceptual process. *Focal attention* (a serial process) follows preattentive processing and involves a detailed analysis of stimuli characteristics. Focal attention can be directed at one stimulus form only and is thus limited in its capacity. In contrast, parallel (preattentive) attention processes do not appear to have limited capacity and can detect gestalt aspects of environmental stimuli from numerous sources. The ability to hear one's name called out by a nonattended voice in a crowded, noisy room is an example of an ongoing parallel process with the ability to detect gestalt features and extremely familiar (and thus automatically processed) stimuli.

Attention Capacity The concept of processing capacity involves the idea that a given task makes a demand on a limited pool of resources. A task with a high processing load draws more resources from the finite pool than does a task with a low processing load, thus inhibiting the accessibility of resources for other simultaneous functions drawing from the same pool. Focal attention requires cognitive effort and thus has a high processing load demand. Cognitive models describing several resource pools suggest an executive process that distributes resources to various cognitive functions. Serial processes that demand processing capacity inhibit the simultaneous action of other serial high-load processes. In contrast, parallel processes have low or no processing capacity demands and can function simultaneously with numerous other functions without inhibiting them.

Optimal performance is attained with moderate levels of arousal that allow for the establishment of task goals and feedback from the performance of the task, leading to appropriate resource allocation. Low levels of arousal impair those processes and lead to inadequate resource allocation. High levels of arousal may be detrimental to the performance because of poor discrimination of stimuli and diminished efficiency of allocation, resulting in poor attention functioning. Pupillary dilation is directly proportional to the level of arousal and has been used as a measure of the degree of processing demand for specific cognitive tasks.

Sustained Attention The ability to sustain attention is called *vigilance* and can be tested with task demands for alertness and concentration over a period of a few minutes to an hour. The tests usually involve detection requirements for target stimuli that occur infrequently at random intervals. An example of such a test is the continuous performance test, which has been used to study various psychiatric disorders. Important aspects of the tests are derived from signal detection theory and include the factors of sensitivity and response criterion. Sensitivity is the distinguishing of target from nontarget stimuli. The response criterion is the amount of perceptual evidence required to support the decision regarding a target versus a nontarget item.

FORMS OF REPRESENTATIONS

The essential feature of information processing in the brain is that the patterns of activation of neural circuits, the neural net profile, contains information within it. These mental representations themselves produce a neural activation effect that both contains information and produces further neural events. The localization of the neuronal firing and the particular pattern of neural activations determine the nature of what the neural net profile represents. For example, activity in the optic nerve in response to light leads to a cascade of neural responses within the visual cortex, which then gives the perceiver the sensation of a visual image. Future activation of those layers in the visual cortex in that general pattern will be the recollection of the visual image. Pattern and localization determine the kind of representation and the information it specifically contains.

SENSATION AND PERCEPTION

Forms of representations include sensory and perceptual ones that derive from input from the external world via the peripheral sensory nervous system. Sensory representations are the least processed of mental representations and are thought to be as close as the brain can get to representing the world as it is. This is a form of processing termed *bottom-up* and is in contrast to more elaborately processed representations that are directly influenced by more abstract aspects of prior experience, called *top-down* processing. As the initial sensory activations are processed (classified, compared, linked to representations from prior experience) they become influenced by “higher”-order processes and become classified as perceptual representations.

The initial stage of encoding a visual sensory representation is called an *iconic image* and is held within sensory memory for a brief period. Features of the initial stimulus, such as its linearity, size, direction, and color, are the information held within this sensory memory representation. Attentional processes at the level of sensory memory act on the initial image with higher cognitive functions such as classifications and chunking. In their essence, these top-down processes compare, contrast, and transform the initial representation to create new perceptual images within working memory. Studies of patients with schizophrenia reveal specific deficits at this early stage of perceptual processing.

Perception is created by the top-down transformations of sensory images. Perception does not necessarily involve the experience of consciousness. Blindsight patients provide an example of perception without awareness. Long-term memory can encode nonconscious perceptions. This has important clinical implications in that patients may be able to indirectly recall and be influenced by events and stimuli that they cannot consciously recall having perceived. However, if conscious, focal attention is involved in perception then the representations are processed differently. The involvement of focal attention appears to be necessary for the activation of the hippocampus in memory processing, which allows for the encoding of explicit, consciously accessible autobiographical memory. Posttraumatic and dissociative states may involve the blockage in focal processing of perceptual representations, thereby leading to psychogenic amnesia.

Imagery involves the activation of brain circuits responsible for perceptual processing. Representations (neural net profile activation patterns) can thus be created by external or internal means. Mental imagery can involve the generation, inspection, retention, and transformation of perceptual images. This processing involves similar effort and timing as when the object is perceived from an external source. Thus, complex visual images require more effort and time to rotate in internal and external reality. The ability of the mind to generate mental images is used in various forms of psychotherapy and may also be an important mechanism in the pathological production of hallucinations and illusions seen in several disorders.

MEMORY SYSTEMS

The neural networks of the brain are capable of responding to experience by the activation of particular patterns of distributed activation. Donald Hebb described a basic principle of memory that has been repeatedly supported by research: “Neurons which fire together, wire together.” Neurons that are activated in a particular pattern at one time will tend to fire together in a similar pattern in the future—this is the essence of memory.

The brain has various forms of circuits responsible for different systems of memory (Table 3.1-2). The form of memory most commonly thought of as “memory” is termed *explicit* or *declarative memory*. This form involves the conscious sensation of something being recalled at the time of retrieval and allows for the awareness of the autobiographical or factual knowledge to be shared, often verbally, with others and the self. This explicit memory system requires the involvement of focal attention and the activation of the hippocampus within the medial temporal lobe memory circuits for encoding and retrieval. Items focally attended are placed in working memory, processed further, and then placed in long-term memory. After a period of weeks to months, items are thought to undergo a process called *cortical consolidation* that places them in permanent memory where their retrieval no longer requires the hippocampus. Cortical consolidation helps to explain the phenomenon of retrograde amnesia following a physical trauma to the head. The finding that cortical consolidation may require rapid eye movement (REM) sleep also raises important questions regarding the relationships among posttraumatic stress disorder, disturbances in REM sleep, and unresolved traumatic memory.

Implicit
A behavioral, emotional, and perceptual form of memory devoid of the subjective internal experience of recalling, of self, or of past. Can include schema or mental models that are summations of representations from numerous experiences.
Also known as *early, procedural, nondeclarative memory*. Cannot be expressed in words.
Present from birth. Does not involve the hippocampus or require focal, conscious attention. Probably involves various circuits, including those of the basal ganglia, limbic system (amygdala, anterior cingulate, orbitofrontal cortex), and perceptual cortices.
Explicit
A form of memory requiring conscious awareness and involving the subjective sense of recollection and, if autobiographical, of self and past.
Also known as *late, episodic/semantic*, or *declarative memory*. Can be expressed in words or drawings.
The autobiographical component of explicit memory does not fully develop until past the first 2 years of life as the hippocampus and orbitofrontal cortex upon which it depends are maturing.

Table 3.1-2 Memory Systems

Before explicit autobiographical memory processing becomes available after the first years of life (during which time the hippocampus and orbitofrontal cortex are maturing), a form of memory called *implicit* or *nondeclarative memory* is already in place and remains active throughout the lifespan. Implicit memory involves a wide range of systems including behavioral, emotional, and perceptual memory. When these circuits are activated in retrieval, they do not have the sensation of something being recalled. For example, when riding a bicycle a person may not recall having learned to ride and may not even feel that anything is being recalled. Similarly, a person with a fear of dogs may be unable to explicitly recall (consciously) any event that may explain such an emotional response. The existence of intact implicit recollection in the absence of explicit memory is found in various conditions including surgical anesthesia, hypnotic amnesia, the adverse effects of some benzodiazepines, neurological conditions such as Korsakoff's syndrome and bilateral hippocampal strokes, divided attention, and childhood amnesia. Such a dissociation in usually associated mental processes may also occur in response to trauma and in individual with dissociative disorders. Thus, patients with post-traumatic stress disorder may have an inability to recall a traumatic event and yet may avoid contextual stimuli similar to the initial trauma, evidence startle response and anxiety, and have intrusive perceptual images each reflecting impaired explicit memory and intact elements of behavioral, emotional, and perceptual implicit memory. Supported by initial imaging studies, this explanation needs further validation through research into memory processes in the disorder.

CONSCIOUSNESS

The vast majority of mental processes are outside of conscious awareness. These processes can impact thinking, feeling, and behavior despite the lack of conscious awareness. Consciousness can be thought to include two elements: awareness and sentience, the quality of the experience. Each form of consciousness has intrigued philosophers and scientists for many years and various theories have been proposed to explain these phenomena. Little is known about the basic mechanisms that underlie the sentient experience of consciousness. Phenomenal awareness has been the focus of active research and has yielded some basic ideas about the role of consciousness in cognition. One essential issue is that the effective processing of mental representations does not require conscious awareness. However, the intentional, strategic alteration in patterns of processing may necessitate the involvement of consciousness in order to achieve a new outcome. Thus, consciousness is not required for most processes, but its involvement allows for a qualitatively different result in representational transformations. One example of this is in memory processing in which explicit memory requires focal, conscious attention or awareness in order to encode events into explicit form. Such representations

are later available for conscious retrieval when they can be examined and transformed for intentional purposes, such as the recollection of facts or autobiographical knowledge.

Conscious awareness thus is a specialized aspect of some cognitive processes. Two leading hypotheses focus on the way in which representational processes are linked or bound together during the flow of informational transformations within the mind. One hypothesis suggests that 60-cycle-per-second sweeping process that extends from the thalamus to the neocortex. This sweep may serve to bind representational processes together in the internal experience of consciousness. Processes that are active at the time of the sweep thus become linked within consciousness. Another view implicates the role of the lateral prefrontal cortex and its role in working memory. Working memory serves as the chalkboard of the mind and representational processes that become linked to the activity in this region are then a part of the attentional spotlight of conscious awareness.

Based on a biological assessment of brain function, Gerald Edelman's theory describes two forms of consciousness that derive from the resonant interactions between groups of neurons. In his model primary consciousness stems from the interaction between perceptual categorizations and conceptual categorizations. This form of consciousness, called the *remembered present*, is also found in higher animals and is unable to transcend momentary awareness. It is embedded in the present but is influenced by categorizations from the past. In human beings the capacity for lexical or language processing enables a secondary or higher-order consciousness to exist and stems from the resonance between those processes and conceptual categories. Higher-order consciousness frees inner experience from the prison of the present and allows for views of the past and plans for the future. Included in these forms of consciousness is a scene of the present situation in which the self is placed in a temporospatial context.

Cortically blind patients state that they cannot see visual stimuli, but they respond behaviorally as if they were fully sighted. They describe being unaware of visual perception, but they make eye and hand movements that reflect the processing of information about stimulus location, shape, orientation, and direction of motion. In information-processing terms behavioral tests reveal that blindsighted patients do sense and perceive visual stimuli but do not have conscious awareness of the perceptual process, an example of a dissociation of the normally associated processes of perception and consciousness or awareness of phenomena.

Misidentification syndromes are other examples of subjective, conscious experience disturbances. In prosopagnosia patients are unable to consciously access memories regarding persons familiar to them. In the Capgras syndrome patients are able to recognize a familiar person's face but feel that it is not really that person. Being certain, as in recognition, is one aspect of consciousness as a cognitive process. The pathological uncertainty of patients with obsessive-compulsive disorder theoretically can be viewed as a disturbance in that aspect of conscious functioning.

Consciousness provides a sense of continuity. Various studies find perceptual discontinuity (e.g., the presence of a blind spot in the visual field, blurred peripheral vision, saccadic eye movements that displace the visual field with sudden shifts) but the subjective experience of continuity in consciousness.

Many psychiatric patients experience a profound sense of discontinuity and confusion that may be related to a dysfunction in the sense-making, continuity-creating process of consciousness. Some psychiatric symptoms, including derealization and depersonalization, may be understood in terms of alterations in conscious functioning (as seen in some patients with schizophrenia, mood disorders, anxiety disorders, dissociative disorders, posttraumatic stress disorder, and some personality disorders), distorted body image (as in eating disorders or mood disorders), intrusive memories and flashback phenomena (as in posttraumatic stress disorder), and hallucinations (as in psychotic states).

Many psychiatric disturbances may thus involve alterations in the experience of conscious awareness and sentience. Various forms of amnesia may involve the blockage of explicit recollection and prevent certain representations from entering consciousness. Body image distortions; obsessive thinking; dissociative symptoms, such as depersonalization and derealization; and perceptual distortions, such as hallucinations and illusions may involve disturbances in conscious sentience. Primary disturbances in thought, such as loose associations and delusions, may also involve disruptions in the process of conscious awareness and sentience.

Mental Models and Schemata Studies of perception and memory support the view that the mind has organizational structures that influence the interpretation of sensory data, shape the encoding of information into long-term memory, bias the retrieval of items stored in memory, and help determine the behavioral response. Those organizing cognitive functions are called *mental models* or *schemata*.

Mental models are unconscious, highly organized structural processes that are derived from past experiences, that aid in interpreting present stimuli, and that influence the direction of future behavior. Mental models exist for various situations. When a situation is appropriate for a given mental model, that model is activated or instantiated. The process, which regulates which model is activated at a given moment, helps to carry out the appropriate information processing and subsequent behavior. The regulation process depends on the accurate reading of the situation and the selection of the appropriate mental model. Aaron Beck's theory of depression is based on the idea that mental models or schemata can produce depressive thinking and depressed moods. John Bowlby used the concept of internal working models to describe the development of early forms of schemata for attachment relationships. Difficulties in intimate relationships and related behavioral dysregulation can be seen as derivatives of models of inadequate early attachment and the presence of multiple, conflictual models. Mardi Horowitz's view of certain personality disorders and maladaptive interpersonal behavior also includes the role of mental models or person schemata. Among those schemata are models of the self, another, and the self in relationship to another in specifically defined maladaptive role relationship models. Some psychiatric signs and symptoms can be seen as derivatives of conflictual schemata and situations. Classic descriptions of interpersonal patterns in some patients with personality disorders, such as idealization and devaluation, can be seen as maladaptive schema functions.

Thought, Language, and Cognition There is no universally accepted definition of thought. Suggested basic elements include propositions (functions containing meaning), images, and lexical and semantic symbols. Cognitive processes can be carried out in parallel, simultaneously, and without consciousness. Cognitive processes, such as thoughts, are often directly known only through translation into consciousness and language. As in the study of mental models, clinical observation and experimental paradigms can infer the nature of thought processes only through indirect measures. These concepts are important in defining the term *thought disorder*.

Thinking involves the mental representation of some aspect of the world or of the self and the manipulation of those representations. Thinking depends on explicit memory and implicit memory for prior experiences. In addition, thought processes can be influenced by a person's emotional state and mental models. The basic components of thinking include categorization, judgment, decision-making, and general problem solving. The assignment of representations of events or objects to categories is important to subsequent thought processing because thoughts can act on the general class to which an item belongs rather than on individual representations; this is another example of top-down processing influences.

The ability to judge the probability of uncertain events, a primary role of judgment, is a fundamental aspect of the rationality of thought. The lack of conformity of thinking to the rules of logic is one aspect of disordered thought processing. Thinking is important in choice, the decision to choose among various options. Each of these processes contributes to the goal of problem solving; data are assessed, classified, transformed, and compared on the basis of logical rules to produce a choice that solves a problem. Failures in these steps can result in limitations and distortions in normal thought processes.

Psycholinguistics is a complex domain that focuses on the cognitive process of language formation and semantic analysis. Cognitive science views language as a dominant influence on subjective experience. It is the medium that dominates human social communication and is one of the major features distinguishing *Homo sapiens* from other species. In the human infant language shapes the ways in which the world is perceived, the manner in which desires are communicated and satiated, and the way in which society responds. Disorders in receptive and expressive language functions influence both the subjective experience of the world and social functioning; children with such disorders often have multiple levels of maladaptation. The ability to communicate through nonverbal language depends on social cognitive processes.

Modes of Processing The mind is capable of distinct modes of processing mental representations. A serial mode utilizes sequential processing in a linear fashion, which is said to be slow and energy-consuming because only a few items can be processed serially at a time. Focal, conscious attention is believed to occur in serial fashion. A parallel mode is more common throughout the brain and involves the simultaneous manipulation of large numbers of representations in a nonlinear fashion. Pattern recognition is an example of such a rapid, low-energy consuming process that can deal with a wide array of stimuli at the same time.

Jerome Bruner has described the distinction between the earlier mode of thought, called narrative cognition, versus the later mode, which is the scientific, logical, paradigmatic mode. Narrative thinking is a context-dependent form of processing that incorporates the internal experiences of the teller and the perceived expectations of the listener in the production of a story. Stories also involve the subjective experiences of the characters involved in the unfolding sequence of events. Logicoscientific paradigmatic processing is said to occur in a context-independent manner that focuses on abstract concepts and their logical, cause-and-effect

relationships. Children develop narrative thinking by the age of 2 years and the co-construction of stories between parent and child is a primary mode of communication in all cultures throughout the world.

Another distinction in contrasting modes of processing has been identified in the type of mental processes primarily attributed to the right or left cerebral hemisphere. Studies drawing on the findings of patients whose corpus callosum has been surgically severed, who have unilateral neurological lesions, or in subjects undergoing brain-imaging protocols have found a remarkable consistency in trends of left- versus right-hemisphere functioning. Some general principles from this array of studies suggest that many processes involve both hemispheres; however, there are distinct patterns primarily originating from each side of the brain. Studies of gender differences in laterality suggest that women are less lateralized in that their processes appear to be distributed across both hemispheres simultaneously whereas men appear to be more lateralized with less “cross-talk” across their hemispheres. For example, the semantic meaning of words is distributed more bilaterally in women than in men. The following generalizations relate to right-handed individuals and to most left-handed people as well. In the right hemisphere are fast-acting, parallel, holistic processes including visuospatial perception. The right side specializes in representations such as images and sensations and the nonverbal meaning of words, sometimes referred to as analogic representations. The right hemisphere is thought to work as a pattern recognition center, capable of assessing the gestalt context of a scene and providing a synthetic interpretation. On the left side are primarily more slowly acting, linear, time-dependent, serial processes. Left hemispheric processes manipulate the verbal meaning of words, known as digital representations, in a logical analytical mode of processing. A generalization from a number of studies is that the right hemisphere tends to note the patterns in the world and creates contextual meaning; the left hemisphere can only make a rationalization of the details of what it perceives in order to create a sense of meaning from a logical view that lacks context and thus may actually seem like a discontinuous and irrational set of data.

The processing of emotion also appears to have a lateralized distribution. The expression of emotion appears to be mediated primarily by the right hemisphere. Facial recognition of the affective expression of others also appears to be a specialty of the right hemisphere. Of note is that the right hemisphere appears to have a more integrated representation of the body's status, information that may be essential for individuals to know how they feel.

Metacognition Metacognition concerns processes that act on the cognitive processes themselves—thinking about thinking. The thinking can be conscious or unconscious and is revealed through either metacognitive knowledge or the regulation of cognition. Knowledge about cognition appears to develop by the age of about 6 when a child enters elementary school; this knowledge takes various forms, including what is known as the appearance-reality distinction—that things may not actually be as they appear to be. Two components of this awareness are representational diversity (the same object may appear to be different to different people) and representational change (thoughts today are different from those of yesterday and may be different again tomorrow). This form of knowledge about the person-specific meaning of cognitive representation requires some sense of the person's awareness of the separateness of minds, a theoretical domain in developmental cognitive psychology called the *theory of mind*.

The regulation of cognition, also called *metacognitive monitoring*, includes such processes as planning activities, monitoring activities, and checking outcomes. Metacognitive monitoring may involve the assessment of thinking sequences for fallacious logic, factual errors, and contradictions in the content of speech.

Social Cognition Bridging the fields of social psychology and cognitive psychology, the study of social cognition focuses on the mental processes involved in social interactions. The domains include the study of empathy, interpersonal communication (verbal and nonverbal), person perception, relationship scripts, and group processes. Other related areas include studies of attribution bias, memory for social interactions, stereotyping, mental control of social cognitive processes, and cognitive origins of a sense of self. Social cognition can be seen as a domain of social psychology that uses information-processing theory to assess the components of attention, perception, encoding, memory, retrieval, and schemata. A dominant theme in social cognition research has been that top-down, theory-driven processing influences interpretations of social situations and actions in such situations. Developmental psychologists have focused on the origins of social cognitive functioning and its deviations. For example, children with autistic disorder have significant deficits in empathic capacity and in the ability to interpret social cues. Social cognitive deficits may be present in different domains in other psychiatric disorders.

Discourse and Narrative Discourse is communication from one person to another; it is thought to involve a sense of intention or plan. Normal discourse follows a set of rules that ensure the coherence and effectiveness of communication: what is intended to be stated by the sender is understood by the listener or receiver. Some researchers support the idea that discourse is a cognitive function that follows the basic principles of information processing, including a schema for effective communication, and of social cognition, such as taking into account the listener's perspective. Incoherent discourse can be noted by analyzing unlicensed violations of the primary maxims of discourse. Another technique is that of discourse analysis, which examines the ways in which discourse deviates from an assumed discourse plan. The exact method to quantify abnormalities in discourse remains controversial, but clinical impressions of incoherence remain important for assessing deficits in social communication. The deficits may result from learned behavior, inherent cognitive abnormalities in thought or language, or deviations in social cognitive functioning. Deviations from normal discourse can be a general finding in need of further assessment. Abnormal discourse is clinically evident in psychosis, specifically in schizophrenia.

Narrative is a broad domain ranging from the literary study of fiction to the developmental psychology investigations of the origin of autobiographical accounts. From a cognitive point of view narrative is important in understanding the relations of language, memory, consciousness, mental models, self-schemata, and social cognition. Narrative can be generally defined as the way in which a person creates a verbal account of a sequence of events in the world and the internal subjective experience of the characters of the story.

Autobiographical narrative begins early in life as the capacity for language develops. Studies of early monologues find that young children interpret and assign meaning to events in their world from an early age. Narrative helps to record and make sense of the past, interpret the present, and anticipate the future. The brain has been called an “anticipation machine” and mental models, prospective memory, and narrative are the major ways in which top-down processing attempts to prepare for the possible future. The enactment of narrative themes directly affects the way in which individuals live out the story of their lives.

As with other areas in development, debate concerns whether this aspect of cognition develops from a biological imperative (i.e., the brain needs to make sense of the world, and narrative is the language output derived from that innate process) or from an internalization of social experiences (i.e., families tell stories, and children practice taking the narrator role). Anthropologists who study psycholinguistic development across cultures have described a phenomenon called *coconstruction*, in which family members collaboratively create a story of daily events in their lives. How those family behaviors influence the child's emerging capacity to organize experiences and encode them into long-term memory to be retrieved later in the production of autobiographical narrative is a fundamental question for many disciplines in cognitive science. Specific deficits in early family experiences and in innate cognitive capacities may theoretically impact the child's narrative capacity. These differences can be seen both in how different individuals tell the stories of their lives and the way they make decisions in life.

Cognitive Development Developmental theories and research can be divided into several views. Stage theories (Jean Piaget, neo-Piagetian, the sociocultural school of Alexander Luria and Lev Vygotsky) describe discontinuous periods of development, with times of stability and consolidation alternating with instability and transition. Information-processing models have not been explored in as much detail with regard to child development, but the models postulate a nonstage theory, in which the emergence of cognitive capacity is a continuous process that does not require a set of invariant sequences. Both stage and nonstage views embrace the idea that a hierarchial integration and an ongoing differentiation are fundamental aspects of cognitive development.

Another distinguishing feature is the degree to which the theories view the contributing role of innate, biological factors and the role of culturally determined social learning experiences. Do cognitive capacities emerge from a genetically determined plan, as in the Piagetian view or do they develop in response to experience, as in the sociocultural view? Developmental psychologists have found features of both views, supporting the idea of a transaction between innate factors and environmental experiences. More recent conceptualizations have drawn on the functioning of complex systems in order to conceptualize development as the continual emergence of ever-more complex capacities.

Psychiatric disturbances in cognition may reflect arrested patterns of normal cognition (as in mental retardation), deviant developmental pathways (e.g., social cognitive functioning in persons with autistic disorder), and specific cognitive impairments (e.g., schizophrenia) that may have been present early on or only became evident as life requirements, such as school, became demanding (some cases of attention-deficit/hyperactivity disorder). Investigations into the developmental features of those disorders is a major focus of the field of developmental psychopathology.

Self-Organizational Processes An understanding of both the development and subjective experience of cognitive processes has been greatly informed by the insights from the fields of evolutionary neurobiology and the nonlinear dynamics of complex systems, otherwise known as *chaos theory*. In selectionist theory, the billions of neurons become clustered into groups that have similar functions and when activated become reinforced. Neuronal groups that are not activated die off; those that are activated survive. The brain, as in species in nature, generates a diversity of activity that can then be selected by interaction with the environment. In addition, the brain has value systems that selectively reinforce the activity of neuronal groups that enhance survival. In this way the brain's neuronal groups compete

and differentiate within the brain and create an ongoing living system perpetually evolving or developing through time.

In chaos theory, complex systems from the inanimate to animate organisms have been shown to have self-organizational processes. Three principles are relevant to psychiatry and include their nonlinear nature, self-organizational processes, and movement toward complexity. *Nonlinear* refers to the finding that small changes in input or initial conditions lead to large and in the long run unpredictable changes in output. These systems function on the rules of probability, which predict that certain combinations of activity within the system are more likely than others and tend to move the system towards self-organization. This probability also predicts that the system will move itself toward increasingly complex states of functioning.

The state of activation of the various parts of the system can cluster into repeated patterns called *states*. In the brain a state of mind or mental state describes the way in which the various neuronal groups may become activated at a given time. Repeated patterns of neuronal group activation, a neural net profile, can become reinforced if they occur frequently or if the value system of the brain engrains their profile. These engrained patterns of activation are called *attractor states*; those states that are least likely to occur are called *repellor states*. The mental states are determined by the constraints on the system. Modification of constraints allows the nature of attractor and repellor states to be altered. Constraints are both external and internal. Thus, features of the external environment such as the way other people behave and relate to an individual can directly affect which mental state is more likely to be activated within the person. Internal constraints include the synaptic strengths of association as determined by both constitutional features and genetics, and those learned from experience as encoded within memory processes.

Psychiatric disturbances may be conceptualized as disturbances in self-organizational processes. Inherited and experiential internal determinants and ongoing external, environmental and social influences on the constraints of the system can thus directly affect both the development and effective use of self-regulatory mechanisms. Viewing psychiatric disturbances in this way allows for a synthesis of the unnecessarily but often historically divided views of psychodyamic, biological, and social psychiatry.

States of Mind With perhaps one hundred billion neurons, each with an average of ten thousand synaptic connections with other neurons, the brain may be capable of a huge, incomprehensible number of possible activation patterns estimated to be over ten to the millionth power. The organization of these possible neural network profiles may be the fundamental goal of cognition and perception. One way of describing the self-organizational process is in the concept of states of mind. Engrained patterns of neuronal group firing become reinforced into repeated patterns that cluster the cognitive processes of attention, perceptual bias, memory, mental models, behavioral response patterns, and emotional tone and regulation. For example, in a depressed state of mind, one may pay conscious attention to negative aspects of experience, interpret incoming stimuli in a pessimistic manner, have greater access to depressing past experiences, have the activation or instantiation of a mental model of the self and others as bad or guilty, have the behavioral pattern of withdrawl, and have a depressed mood with difficulty regulating intense affect.

States of mind are the primary mechanism by which the brain organizes its activity. Healthy mental functioning may depend on a flow of states of mind through time that are adaptive to the ever-changing environment and allow the individual to draw freely from the learning from past experience. Chaos theory suggests that nonlinear complex systems must move continually towards maximizing the complexity of the system. Achieving such a goal requires a balance between the elements of continuity, familiarity, and predictability with those of flexibility, novelty, and uncertainty. Dysfunctional mental states may be conceptualized as a disruption towards either end of this balance: with excessive rigidity, as in the case of character pathology, or excessive fluidity, as in the case of disorders of thought or of mood.

Emotion Emotion has long been considered fundamental to cognitive processing. Research paradigms have required careful control of variables and therefore have often attempted to not focus on the role of emotions in various measures of processing, such as attention, perception, and memory. However, the role of emotion in these and other aspects of cognition is now well established. What is an emotion? The answer to this question is as complex as the mind itself. One view considers emotion as a primary value system of the brain, allowing activations to be selectively reinforced. For example, emotionally charged experiences may be more readily recalled than uneventful ones. According to this view the most fundamental aspect of emotion is the arousal/appraisal system in which the brain responds to a given stimulus with the signal of “This is important—take note and pay attention now!” Emotion thus gives value to a representation by arousing attentional mechanisms and focusing a spotlight of attention on the stimulus. The second stage would then appraise the meaning of such emotional arousal by assessing its hedonic tone: “Is this good or is this bad? Should this be approached or avoided?” Emotion thus directs the flow of energy—the activations within specific circuits of the brain—as the arousal/appraisal system focuses cognitive processes on elements of the internal and external environment. A third level of emotional processing is the elaboration of this appraisal into a more specific form, called a *categorical emotion*, such as joy, interest, surprise, fear, anger, sadness, or shame. These categorical emotions have universal facial expressions of affect found in all cultures, which may have distinct psychophysiological manifestations.

An additional view examines the way in which the changes in the body's state is represented in the brain in the form of what Antonio Damasio has called a *somatic marker*. According to this view, the energizing or deenergizing bodily responses let the brain know how the individual feels about an experience. Such a somatic marker can then be used in future emotional assessments, or gut reactions, to an experience.

A part of the brain called the orbitofrontal cortex has been implicated as the sight of somatic marker processing. Allan Schore has also noted the importance of early experiences with caregivers in the maturation of this region and also its central role in coordinating self-regulatory functions early in development with basic emotional reactions and social functioning. The orbitofrontal cortex has been implicated not only to monitor but also to regulate bodily states and to possibly be involved in psychiatric disturbances ranging from autism to mood disorders. Disorders in self organization and social functioning may be better understood by examining the central role of emotion, and perhaps the orbitofrontal cortex and related regions, in the development and maintenance of dysfunctional mental states. Studies also suggest that this region is responsible for the capacity for self knowledge and the subjective experience enabling the mind to reflect on the self in the past, present, and the potential future. Both inborn and experiential factors may play important roles in allowing this region to develop the capacity to integrate a wide range of important functions of the mind including the appraisal of meaning, emotional regulation, social cognition, and autobiographical consciousness.

COGNITION AND PSYCHIATRIC DISORDERS

Since the time of Emil Kraepelin and Eugen Bleuler psychiatrists have known that certain disorders include profound deficits in mental functioning, such as attention and reasoning in the case of schizophrenia. Since the 1950s researchers have attempted to determine the exact nature of such deficits. With the advances in computer technology and the technical ability to analyze stimulus presentation and response times of the order of tens of milliseconds, cognitive psychologists have been able to devise research paradigms capable of testing when such deficits are present. Those advances led to revisions in the proposed structural models of processing.

Processing research has focused on the three domains of sensation, perception, and cognition. Sensory-processing studies focus on poststimulus events up to a maximum of 1 second, using simple stimuli. Perceptual studies examine processing after a period of up to about 5 seconds after a slightly more complex stimulus. Cognitive processing experiments can examine early processing (e.g., phenomena occurring within the first 30 seconds) of complex, long-term cognitive processes that can occur over minutes, hours, or days. Studies of long-term memory can examine processes that involve an indefinite poststimuli period. Recent attempts have been made to correlate complex cognitive findings with clinical presentations. General problems with the approach include the diversity of patients falling under the same syndrome classification. Thus schizophrenia can more appropriately be considered as the schizophrenias or schizophrenia spectrum disorder. The same may be true for other disorders, such as attention-deficit/hyperactivity disorder and major depressive disorder. An array of cognitive dysfunctions identified for certain subsets of syndromes must be interpreted in the light of the heterogeneity of some patient populations.

A related problem is the distinction between general deficits and specific deficits. Care must be taken in interpreting experimental data that show a difference in results between normal controls and a patient group. Do psychiatrically ill patients perform less well on a given paradigm because they are ill or because of a deficit specific to the disorder? The creative design of experimental tasks can help distinguish between general deficits and specific deficits. A comparison of target patient populations with matched normal persons and other psychiatric patients can help to determine disorder-specific cognitive dysfunction.

Another general issue is that of state markers versus trait markers. For example, a patient with schizophrenia may have a persistent cognitive deficit when actively psychotic (state) and also when asymptomatic (trait). These abnormal results have been found in certain cognitive tests of attention that correlate with improvement on medications; the abnormal results are also found in nonschizophrenic first-degree relatives. Is the marker of genetic vulnerability a coincidental finding or part of the core deficit in schizophrenia? An exploration of the implications of these cognitive abnormalities for the daily life of the patient is an important application of the research findings to clinical psychiatry.

Schizophrenia In the late 1890s Kraepelin described a primary attention deficit in his elaborate clinical description of patients with schizophrenia. Numerous investigators since that time have attempted to define the nature of the cognitive deficits in schizophrenia. A general approach is that an early perceptual processing

deficit leads to problems in perceptual organization. In general, information-processing models note that two things are processed: energy (in the form of external stimuli impinging on the senses) and information (a stimulus that carries a signal value based on significance derived from the prior processing of similar energy configurations). Schizophrenia patients appear to have deficits in the processing of energy as well as information.

Some cognitive tasks have been identified as trait-linked markers of schizophrenia: reaction time crossover, backward masking, dichotic listening, serial recall tasks, vigilance (sustained attention) tasks requiring high processing loads, and span-of-apprehension tests with large visual arrays. Deficits in those areas have been explored through thousands of cognitive studies examining various aspects of processing.

Crossover and Modality Shift Effects Those paradigms examine the general finding that schizophrenia patients have a slower than usual response on tasks that require rapid reaction times. A stimulus is presented with varied combinations of warning signals and preparatory intervals. Schizophrenia patients show an advantage only with short preparatory intervals and long response times with regularly spaced stimuli, a pattern distinct from that of normal controls (crossover effect). In a related paradigm, when the modality of the stimulus is varied (e.g., light is interspersed with tone), the latency (delay) of the response in schizophrenia patients, when compared with controls, is longer if the preceding stimulus was of a different modality. That is termed the *modality shift effect*, revealing a greater degree of cross-modal retardation in schizophrenia patients than in controls.

A number of theories have been proposed to explain these effects. They may be quantitative rather than qualitative distinctions from normal control groups. However, both the crossover and modality shift effects support the idea that schizophrenia patients are overly influenced by stimuli that occurred immediately before the effect. The information-processing stages that explain the persistence of prior stimulus effects is under investigation.

Visual Backward Masking, Sensorimotor Gating, and Habituation The aspect of cognitive functioning in which rapidly presented stimuli are abnormally attended to by schizophrenia patients as compared with normal controls has been examined. In visual backward masking a stimulus is followed by an interstimuli interval, and then a subsequent stimulus is presented. Figure 3.1-2 shows a typical masking experiment. The presentation of the secondary stimulus leads the schizophrenia patient to not report, to mask, the initial stimulus. Lengthening of the interstimulus interval beyond 500 milliseconds can lead to normalization, with no masking present. Thus, the rapidity of presentation of the secondary stimulus is the factor determining whether it will influence the perception or at least the reporting of the initial stimulus. Some studies find that the impairment improves with treatment by medication and can be induced in normal patients given catecholaminergic agents. Other studies find that the impairment may be a marker of increased vulnerability to schizophrenia.

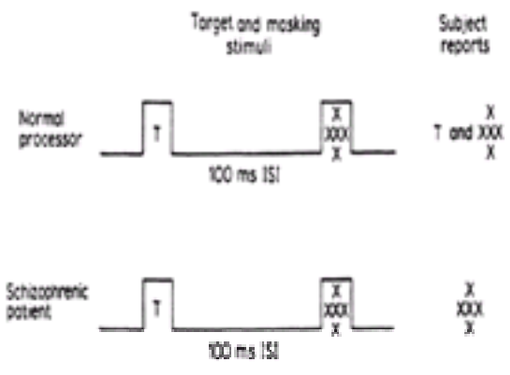

FIGURE 3.1-2 Diagram showing the difference in the verbal reports of normal versus schizophrenic subjects presented with a single backward masking trial with a 100-ms interstimulus interval (ISI). The T represents the target stimulus and the Xs represent the masking stimulus. (Reprinted with permission from Braff DL, Saccuzzo DT, Geyer MA: Information processing dysfunctions in schizophrenia: Studies of visual backward masking, sensory motor gating and habituation. In *Handbook of Schizophrenia*, vol 5, Neuropsychology and Information Processing, SR Steinhauer, JH Gruzelier, J Zubin, editors. Elsevier Science, New York, 1991.)

Sensorimotor gating and habituation are the processes by which stimuli become less attended to with rapidly repeated presentation. Habituation is believed to involve preattentive processing, and the visual masking requires higher cognitive functions. Schizophrenia patients show a markedly diminished capacity to habituate. One common study examines the persistent acoustic startle reflex as the person continues to blink with repetitive tones. Parallel findings are those of the studies that showed that lysergic acid diethylamide (LSD) administration and the intracerebral injection of dopaminergic agents in rats lead to similar findings, supporting the idea that excessive dopamine activity, thought to be central in schizophrenia, can induce those deficits.

In general, the diminished habituation and visual backward masking lend further cognitive support to the idea that schizophrenia patients have a markedly diminished capacity to regulate the flow of rapidly presented information, leading to an inundation by stimuli that in a normal person are gated out. The disruptive effects of the externally derived stimuli, as demonstrated in those experimental paradigms, are thought to occur also for internally generated stimuli.

Psychomotor Slowing A general deficit, not specific to schizophrenia, is a generalized slowing of response rate. Cognitive science is concerned with all stages in information processing, from attention, sensation, perception, and cognition to action. Multiple neural integrative systems are involved in the processes, from stimulus to response. Numerous disorders may result in a common finding of diminished reaction time. The general nature of the deficit should not lead clinicians and researchers to ignore the importance of response rate deficits in interpreting clinical and experimental data. Attempts to assess the interdependence of input, central processing, and action are part of psychomotor studies in schizophrenia patients.

Selective Attention Numerous studies have attempted to determine the nature of deficits in selective attention in schizophrenia patients. Early studies were based on the idea of a bottleneck that limits the attention-processing capacity at some stage in the information flow. These studies were based on the idea that some cognitive structure devoted to filtering is impaired. Analysis of the studies revealed consistent deficits in shadowing (the ability to repeat what the person is selectively focusing on in a dichotic condition) and in verbal recall. In general, selective attention paradigms present the person with a target stimulus and distracters to be nonattended. Findings reveal that schizophrenia patients have an impairment in their ability to avoid distracting stimuli (to filter) and to pigeonhole (to use category features to reduce stimulus qualities needed to respond). The studies found that distractibility is a core cognitive deficit, supported by its high incidence in genetically vulnerable persons, its improvement with medications, and its worsening in acutely psychotic states.

These findings were explained by using the framework of an impaired filtering structure and pigeonholing process, but recent conceptualizations have examined a generalized impairment of the information-processing capacity in schizophrenia. The capacity model examines the way in which a pool or pools of attention capacity can be allocated across mental activities. Two components are quantity of resources available (capacity) and executive allocation policy. Other areas of deficit may involve an impaired response selection process, leading to abnormal results on tasks.

Several possibilities have been proposed to explain attention deficits in schizophrenia patients on the basis of the capacity model: (1) deautomatization of normally automatic preattentive processes, (2) disproportionate allocation of attention to schema-relevant but task-irrelevant information, (3) inability to sustain controlled processes needed to maintain attention allocation without shifting, (4) inability to shift allocation biases to correct wandering attention, and (5) disorganized response selection because of heightened arousal under distracting conditions. Studies of selective attention may begin to examine these possibilities in a capacity model rather than in the previously explored structural framework.

Sustained Attention Sustained attention, or vigilance, is required to process stimuli of long duration. The most common research paradigm in cognitive studies of sustained attention and psychopathology is the Continuous Performance Test. The test is a rapidly paced set of vigilance tasks with varied spacing and timing of target and nontarget stimuli. Continuous Performance Tests have the following features: the presentation of a random sequence of visual stimuli on a small screen (numbers or letters), a fixed stimulus pace with a rapid rate (1 to 2 seconds), a brief stimulus exposure time (40 to 200 milliseconds), designated stimuli or sequences of stimuli as targets, and the use of a button with which the person can respond. The processing load for a Continuous Performance Test can be varied by blurring the stimulus presented or by changing the rapidity of presentation.

Vigilance tests, such as the Continuous Performance Test, require analysis of response features on the basis of the signal detection theory. The two elements distinguished are sensitivity and the response criterion. Diminished sensitivity is a sign of decreased vigilance and results in a high miss rate (errors of omission). The

response criterion can be diminished, leading to a high false-positive rate (error of commission). The analysis is important in the interpretation of results.

Schizophrenia patients appear to have a specific diminishment in sensitivity but not a lowering of the response criterion for both verbal and spatial stimuli. The impaired responses were significantly associated with specific clinical features in the patients and first-degree relatives. Test abnormalities, although present in other disorders, appear to be most robust in schizophrenia.

Positron emission tomography (PET) in schizophrenia patients performing a Continuous Performance Test found differential activation, with lower metabolic activity than in normal persons, in the prefrontal cortex bilaterally but normal or elevated activation in the occipital region, leading to a hypofrontal pattern. That finding is consistent with other findings supporting the idea of impaired frontal function in schizophrenia.

The vigilance studies using the Continuous Performance Test reveal that schizophrenia patients have a deficit in their ability to distinguish target stimuli from nontarget stimuli when the stimuli are presented as brief signals at a rapid pace. This finding is consistent with the other features of attention abnormalities revealed in other paradigms.

Language and Discourse Assessments of language dysfunction in schizophrenia have focused on the basic question of whether the abnormalities in speech are reflections of a core disorder of thought or an abnormality in speech production. The search for a schizophrenic language has yielded negative conclusions. Discourse analyses of conversations with schizophrenia patients suggest that they may have significant difficulties in the maintenance of a specific topic (derailment) and in the lack of a discourse plan directing speech (disorganization).

Other investigators have argued that schizophrenia patients have impaired capacities to perceive the needs of others and may be schema driven in the direction of their speech, rather than intending to communicate effectively. One study showed that delusion-consistent material presented in the nonattended channel in a dichotic listening task leads to diminished attention to the target channel. A clinical implication of this experimental finding is the possibility that schema-driven processing during speech production may be diverting attention to internal stimuli and away from the potentially confusing social demands of conversation.

Span of Apprehension Span of apprehension is an experimental paradigm that illustrates a cognitive psychological approach to the core deficit of schizophrenia patients. In the span-of-apprehension test an array of letters is displayed for a brief period (from 50 to 100 ms in most studies). One of the letters is a “t” or an “f” and the person must detect which letter is present. The number of nontarget letters is increased, and significant differences in detection are found for displays of 10 or more letters. Figure 3.1-3 provides an example of a visual display for the span-of-apprehension test.

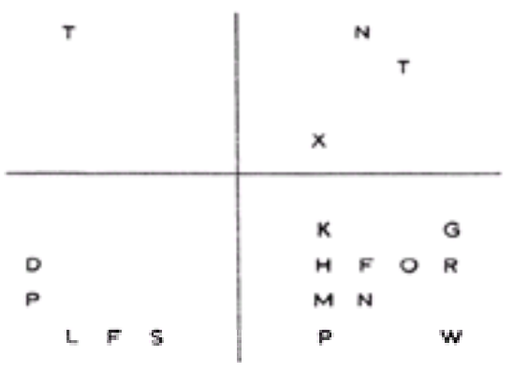

FIGURE 3.1-3 Sample arrays used in the wide-visual-angle version of the partial report span-of-apprehension task. (Figure courtesy of Robert F. Asarnow, Ph.D.)

The serial scanning process is an element of focal attention. Parallel processing in the search involves increased aspects of assessment of figure-ground and textual segregation and is thought to be an automatic process.

Studies have shown that the sequential scanning of the attention spotlight is directly affected by increasing the complexity of certain display characteristics, leading to increased errors in detection. This is logical because the iconic image has a limited display time and the scanning of the image may be incomplete by the time it decays from such an ultrabrief form of memory.

Findings in Schizophrenia Patients Schizophrenia patients show significantly increased errors in the span-of-apprehension test under conditions of increased complexity of display. Their scores are also worse in psychotic conditions and are improved with symptomatic improvement while taking medications. Increased errors are also found in nonschizophrenic mothers of children with schizophrenia. Thus the span-of-apprehension test is a measure of both state and trait in some cases of schizophrenia.

Only about half of the patients with the diagnosis of schizophrenia have abnormal results on span-of-apprehension tests. Those who do have abnormal results also have the clinical symptom of anergia. Other psychiatric disorders studied did not reveal these findings. Thus, the abnormal results on the test appear to be specific to some forms of schizophrenia. Both short-term and long-term outcome studies have found that those patients with abnormal test results whose scores improve after receiving antipsychotic medication have a good clinical response to pharmacotherapy.

The span-of-apprehension test taps into some aspect of cognitive function specific to some patients with schizophrenia, their nonschizophrenic relatives, and persons at risk for developing the symptoms of schizophrenia. To scan the iconic image the person must (1) engage attention to the iconic register, (2) move the focus of attention, and (3) disengage the focus of attention. Impairments in performing any one of those tasks can explain the test-result abnormalities. Another cause of the deficiency may be that, with each fixation of attention, less information is processed. Thus, although the individual steps of iconic scanning may be intact, less visual information is processed and more errors occur. A third possibility is that the initiation of the attention process is delayed; this possibility is consistent with the increased reaction times revealed on numerous other tasks. The delay in the face of a rapid decay rate of sensory memory places the patients at a disadvantage when rapid responses are required; this possibility is also consistent with the other forms of attention deficit described earlier.

In contrast to the structural models described, another view focuses on the capacity model of attention. With a general limit on the cognitive resources available at a given moment, tasks that require an increased processing load (as in the case of time requirement or more letters on the display) diminish the resources available for other processes (such as scanning the image for the target letter and reporting the determination). Schizophrenia patients may have a number of structural and capacity deficiencies and these hypotheses are not mutually exclusive. Further studies are needed to elucidate the nature of the cognitive state-trait marker for schizophrenia patients and persons vulnerable to the disorder.

Attention-Deficit/Hyperactivity Disorder The change in 1980 of the terms “hyperkinetic” and “minimal brain dysfunction” to “attention-deficit disorder” in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and the change in 1987 to “attention-deficit hyperactivity disorder” in the revised third edition (DSM-III-R) emphasized the role that attention factors are thought to play in children, adolescents, and adults with the disorder. In the fourth edition of DSM (DSM-IV) attention-deficit/hyperactivity disorder integrates two categories from DSM-III-R: attention-deficit hyperactivity disorder and undifferentiated attention-deficit disorder. Diagnostic criteria embrace a number of forms of the syndrome, and research into the cognitive deficits has outlined a wide array of tasks in which attention capacity is abnormal. The pervasive findings of cognitive impairment in the setting of numerous intact cognitive functions has left the research field with no clearly accepted view of a core deficit in the disorder. Clinically, child and adolescent patients present with problems in school, with peers, and at home that reflect both academic and behavioral dysfunctions. Many studies suggest that the cognitive and behavioral dysfunctions in the disorder may be independent processes with different neurophysiological bases. For example, the finding that children with attention-deficit/hyperactivity disorder are impulsive may be true behaviorally but cannot be stated as a generalization about their cognitive functioning.

Researchers have attempted to find clearcut diagnostic criteria to help clarify the disorder, but individuals classified as meeting diagnostic criteria at this point appear be quite heterogeneous. There is no definitive test for the disorder nor is a positive response to psychopharmacological intervention pathognomonic. Biochemical studies have found abnormal urinary catecholamine metabolites that normalize as the patient's behavior improves when taking psychostimulants. Those findings and

PET scan data suggest abnormal brain functions in patients with attention-deficit/hyperactivity disorder.

Data from numerous studies show that the patients with attention-deficit/hyperactivity disorder have dysfunctions on a variety of tasks ranging from those involving monitoring, perception, memory, and motor control. Intact performance has been found on a number of memory tasks requiring verbal processes (e.g., digit span, word tests, story recall) and nonverbal processes (e.g., recall, visual arrays, block series).

A number of theories have been elaborated to explain those differences between patients with the disorder and normal persons. Each theory has strengths, weaknesses, and varied support from the data, but each highlights the complexity of the cognitive dimensions of the disorder. In general, patients with the disorder evidence behavior that has been compared to that of patients with frontal lobe damage: deficits in the control of motor responses, in the execution of fine-motor movements, and in the inhibition of ongoing response patterns. Memory tasks and basic aspects of information processing are intact, but the patients have impaired performance across modalities (auditory, visual, motor, perceptual-motor), suggesting some global deficit. The patients also appear to be unusually susceptible to boredom when the required tasks are long and repetitive.

Another view examines two proposed systems—an underactive behavioral inhibition system and an impaired behavioral reward system—to explain the behavioral problems that children with the disorder often have. A related view is that the rule-governed behavioral system is not intact; patients with the disorder appear to do especially poorly in a system with delayed or nonexistent rewards as in tasks that require sustained attention, accuracy, or task-directed activity governed by another person's direction or rules. Under those conditions the patient's poor regulation and inability to meet functional demands are revealed. A related issue is a diminished motivational drive and possibly a diminished arousal regulation system.

Yet another approach supported by research data is that the patients have metacognitive deficits. According to this perspective, metacognitive processes that help plan, monitor, and regulate performance are impaired. The patient's ability to assess the task and determine strategies also has deficits. That is an example of top-down aspects of attention, with impairment of the higher cognitive processes that regulate information flow. Additional bottom-up deficiencies involve basic aspects of attention focus because of abnormalities in arousal, selectivity, and capacity.

The finding that numerous factors can influence the appearance of deficits led Virginia Douglas to the hypothesis that attention-deficit/hyperactivity disorder is a self-regulatory disorder with pervasive effects. The impairments affect each of four domains—attention, inhibition, reinforcement, and arousal—resulting in deficits in several aspects of self-regulation: (1) the organization of information processing, including planning, metacognition, executive functions, adapting appropriate cognitive sets for a given task, regulating arousal levels and alertness, and self-monitoring and self-correction; (2) the mobilization of attention, including the deployment and the maintenance of adequate attention; and (3) the inhibition of inappropriate responses such as withholding responses to extraneous stimuli and reinforcers. These deficits in self-regulation imply that increased processing demands would lead to a diffusion of attention processes and to a subsequent impairment of the indepth, coherent acquisition of knowledge and understanding. One line of research has been based on the additive factor method, in which experimenters attempt to isolate the stage of the deficit. The model for that approach entails four stages: encoding (the identification of a stimulus), serial comparison (of the stimulus with elements related to the category in long-term memory), decision (pertaining to the category into which the stimulus is stored), and translation and response organization. Studies using evoked potentials suggest that deficits are found after the search and decision stages (the first three stages). The response preparation and execution processes appear to be impaired. Features that increase the processing load—such as speed demands, complexity of stimuli, distractors leading to divided attention, and increased duration of task—reveal different areas of deficit. This fact may explain the diversity of methods and research data supporting various theories. The variables affecting outcome include information-processing demands, the availability of alternate stimuli to which to attend, and the presence of an external regulator.

The diversity of research findings and theoretical explanations is paralleled by the clinical finding that children who are severely impaired in the classroom may have no attention problems in the confined, one-on-one setting of the psychiatrist's or psychoeducational examiner's office. Such children may also be able to attend for indefinite periods to video games and yet be unable to follow complex conceptual information. The important principle is that cognitive dysfunction in psychopathological conditions may be task specific as a function of the nature of the cognitive impairment. The patient's clinical history and evaluation must consider potentially hidden domains of abnormal cognition.

Autistic Disorder Early descriptions of autism delineated the social functioning deficits that impair normal functioning. Autistic children were seen as having difficulties in normal emotional contact. Cognitive dysfunctions in autistic disorder were described later and were found to involve a number of areas, including abstraction, sequencing, language, and comprehension. Researchers are now focusing on the nature of the core deficit in the disorder. Issues of specificity and universality of areas of dysfunction in the disorder are important. How the cognitive domains relate to the social-affective deficits is of particular interest.

Seventy-five percent of patients with autistic disorder are also mentally retarded. Mental retardation involves global cognitive and language impairments that may make it difficult to distinguish autistic disorder features. General cognitive impairments may be especially difficult to assess if language functioning is severely limited. Studies of high-functioning patients with autistic disorder have permitted various deficits to be determined.

Studies of the cognitive deficits in autistic disorder have distinguished an array of dysfunctional areas, including numerous language problems, excessive or impaired responsiveness to stimuli of various modalities, different encoding of auditory stimuli, and impairment in the ability to extract important features from incoming information. This range of deficits is thought to require the transformation of symbolic representations. In contrast, some patients with autistic disorder have relatively intact visuospatial and gestalt functions, musical abilities, and rote memory. Performances on standardized tests reveal relatively good results in object assembly and block design but poor results in comprehension.

Language deficits vary and include syntactic and phonological domains. These findings initially suggested a left-hemisphere deficit, but other findings, including deficits in prosodic and pragmatic language functions, suggested right-hemisphere involvement as well. The findings can be interpreted as deviations from normal language functioning as well as delays in language functioning. The wide array of dysfunctions in autistic disorder may be due to a number of subtypes, with characteristic deficits and possibly different brain loci of dysfunction.

Other studies have focused on the social cognition of patients with autistic disorder. They have examined the nature of the patient's emotional behavior and understanding to assess the earliest clinical descriptions of an abnormal emotional connection between autistic children and their parents. Recent work in neurobiology suggests a role of the orbitofrontal cortex and the cerebellum in mediating some of these deficits. Two findings support the initial impressions: autistic children are much less likely than usual to imitate adult vocalizations and gestures, and they show much less sophisticated representational play with objects than do normal children. These findings led to the suggestion that a core deficit in autistic disorder is the representation of representations (metarepresentations), leading to deficits in symbolic play and the inability to understand the mental states of others. The inability to transfer cognitive representations into language symbols may also be a related metarepresentation deficit.

A series of studies explored the relation of these possible cognitive impairments to socioemotional behavior. In contrast to clinical lore, children with autistic disorder were found to look at their parents; they had eye contact with their parents when social interactions were parentally elicited, and they revealed normal behavioral patterns of attachment. The studies found a marked lack of social referencing (looking to parents for emotional cues in ambiguous situations) and protodeclarative gestures (pointing to objects and showing objects to familiar adults). Three domains have been proposed to explain these findings: (1) autistic children may not have the capacity to have a representation of another person as having ideas, perspectives, or emotions that can be shared. This proposal is consistent with a theory-of-mind hypothesis, in which the core deficit is believed to be the inability to have a sense of another's mind; (2) autistic patients may have an impaired ability to perceive or comprehend the emotional (usually facial) signals of others; and (3) the core deficit may involve either a lack of interest in others or an aversion to responding to others. Studies have found that although children with autistic disorder do express emotions, they have less positive affects in response to their (relatively infrequent) periods of joint attention. Furthermore, they have an impairment in their responsivity to the display of strong emotion by another, whether of distress or of pleasure. Tests of high-functioning autism patients reveal poor performance on emotion-recognition tasks, little comprehension of and empathy with depictions of social situations, and difficulty in talking about socially derived emotions such as pride and embarrassment. Autistic patients with relatively high intelligence use adaptive cognitive strategies to interpret social stimuli to compensate for impaired emotion-processing abilities. The development of social cognition and socioemotional understanding requires complex interactions among the cognitive, perceptual, and emotional processes. A series of interactive elements essential to the development of social understanding has been proposed (Fig. 3.1-4). This model describes the basic precursors of emotional responsiveness; the ability to attend to, encode, and interpret verbal and nonverbal social stimuli; the awareness of one's own and others' emotional responses; and the ability to contrast oneself with others. Out of that matrix develops the ability to understand others' views, desires, and beliefs. Accordingly, a deficit in any of those basic elements may explain the characteristic deficits observed in the social understanding of persons with autistic disorder.

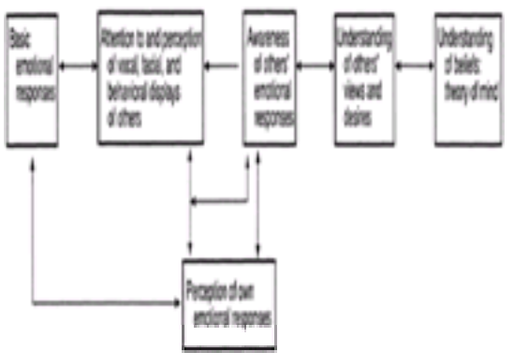

FIGURE 3.1-4 Sigman's model for the development of socioemotional understanding. (Figure courtesy of M. Sigman.)

Mood Disorders In contrast to schizophrenia, attention-deficit/hyperactivity disorder, and autistic disorder, the mood disorders do not appear to have core cognitive deficits that are diagnosis specific. Instead, cognitive abnormalities appear to be related to the degree of psychopathology and to the severity of the mood disturbance. Most studies have examined patients with depression; only a few studies have assessed cognitive functioning in patients with bipolar I disorder in a manic state. In depressed patients the severity of the depression has ranged from mild depression in students to severe illness in hospitalized patients with major depressive disorders. The studies have primarily examined attention and memory for neutral and emotionally toned stimuli. For the majority of these studies, the concept of self-organizational processes and state regulation has not been the primary focus of attention. These recent conceptualizations of the brain's functioning as a nonlinear complex system capable of self-organization may aid in the future investigation of the primary dysregulatory aspect of disorders of mood.

Depressive Disorders Depressed patients often complain of difficulties with concentrating, learning, and remembering. Studies have documented that such patients perform poorly on tasks that require sustained attention or effortful and elaborate rehearsal. Thus controlled limited-capacity processes appear to be impaired in major depressive disorder. This limitation on access to capacity-demanding resources appears to be directly related to the severity of the depression and normalizes with remission from a depressive episode.

In contrast to patients with dementia, who have impaired semantic recall, depressed patients have intact access to long-term knowledge stores. Studies have found that although explicit memory for recently learned tasks is impaired in depressive disorder as well as dementia, implicit memory (as shown by priming tasks) is intact in depression but impaired in dementia. This finding supports numerous studies showing that effort-requiring tasks are impaired in states of depression.

A dual-process model distinguishes between elaboration and activation. *Activation* makes items accessible and readily available to subsequent cognitive processes and is thought to be an automatic process fundamental to implicit memory. *Elaboration* is defined as the way in which the associative linkages are formed and strengthened between the current stimuli and other representations in memory. Elaboration occurs over an extended time and makes an item retrievable by the creation of several retrieval paths. The finding of intact implicit memory suggests that the automatic process of activation is unaffected in depression, whereas the effortful process of elaboration is impaired, thus influencing explicit memory. Other studies of memory function reveal that depressed patients are likely to remember emotionally negative words and negative events in their lives during a depressive episode. Although studies of anxious patients have found consistent attention biases favoring attention to threat-related words, studies of attention effects in depression have yielded inconsistent results. Thus processing and recall in depressed patients appear to be more related to emotional effects on cognition than to attention biasing.

Depressed patients have also been found to have an increased response criterion; they require increased supportive data from the presented stimuli to respond in a test situation. Whether this need to be certain before responding is a psychological response to being depressed or a specific feature of the depression itself has not yet been determined. The reluctance to respond needs to be considered in interpreting research data and clinical interview findings.

The attention and memory findings have suggested several theoretical frameworks that are clinically useful. A schema theory of depression outlines a positive feedback loop in which negative self-schemata prime persons to have negative thoughts, to recall negative events in their lives, and to interpret present events with a negative bias. Whether as a cause or as a maintaining influence, these depressogenic schemata are thought to create a series of cognitive functions that produce and maintain a depressed mood.

A network theory of memory and emotion has been supported by research on depressed and nondepressed persons in which mood leads to a spreading activation of items in memory that are congruent with the mood. Thus emotion directly influences retrieval by a process of state-dependent learning and memory. While depressed, patients are likely to encode items in a form that makes them readily accessible when retrieved in a depressed state; depressed mood thus becomes an internal context cue that accesses depression-related memories.

Both the network and schemata theories of depression are consonant with research findings but do not explain all the cognitive and clinical findings in depression. They provide a framework for understanding how emotions and moods influence cognitive processing, such as memory retrieval and mental models. The conceptualization may be applicable to both transient mood states and severe mood episodes. Thus emotions in both health and illness can shape the mental state instantiated at a given time and thus act as a context cue, leading to the activation of previously formed schemata for oneself and others. The activated schemata in turn can produce retrieval biasing and behavioral responses, fundamental to a mental state, that can further elicit a negative emotional response. A reinforcing loop is established to support the continuation of the depressed mood and depressive cognition.

Some patients may be especially prone to marked cognitive alterations because of an emotional state that may be a fundamental part of a clinical presentation. Thus, rapid shifts in state of mind may be a learned or constitutional feature of the individual. From the nonlinear dynamics view, these rapid shifts in mental state can be conceptualized as sudden and intense changes in the constraints on the system, which determine the flow of states of mind across time. Activating a set of constraints that establish and maintain a depressed state can then lead to the persistence of the very factors that created the shift in probabilities of that state being active. The depressed state becomes a deeply engrained attractor state that becomes difficult to alter. Pharmacological interventions may directly alter the synaptic constraints maintaining such a state. Psychotherapeutic interventions, such as cognitive behavioral and interpersonal therapy, can be proposed to produce changes in the internal processing of constraints and in external social constraints, respectively.

Bipolar I Disorder Patients with manic episodes present with a spectrum of cognitive dysfunctions that appear to be related to the severity of their mood episode and that normalize with remission. The disturbances have been described clinically as rapidly paced thinking and speech, quick associations to self-generated or other-generated stimuli, grandiosity, and increased distractibility.

Formal studies of patients with bipolar I disorder are technically difficult to carry out because of the patient's lack of cooperation and restlessness and because such patients are few in number. The studies have suggested a high rate of combinatory thinking, the inclusion of loosely associated but related intrusions, and increased distractibility. The clinical impression of humor (even in the face of an underlying dysphoria) in some patients with bipolar I disorder is corroborated by playful, extravagant, or flippant elaborations and intrusions in speech. These findings seem to indicate primarily state-dependent symptoms that improve on recovery. Thus forced-choice span-of-apprehension and backward masking tasks reveal impairments similar to those seen in actively schizophrenic patients but unlike those seen in schizophrenic patients who have persistent deficits in the remitted state. Bipolar I disorder patients who are not actively ill reveal normal information processing. Thus, bipolar disorders can also be viewed as a disorder of self-organization in which the system fluctuates between the extremes of highly activated manic and highly deactivated depressed states of mind.

Posttraumatic Stress Disorder Studies of posttraumatic stress disorder have extended the findings of information-processing abnormalities in anxiety disorders: attention biases toward fear-related and threat-related stimuli. Many of the studies were hampered by poorly defined clinical populations and control groups. However, some general trends, noted especially in the well-controlled paradigms, are promising and provide important insights into the psychopathological cognitive mechanisms in posttraumatic stress disorder.

Cognitive studies of patients with anxiety disorders have focused primarily on either attention or memory. Various research paradigms have been applied to assess attention bias and memory retrieval for neutral and emotionally activating stimuli. Studies find that anxious patients have an increased tendency to attend to

fear-related and threat-related words. One research approach includes a dichotic listening task in which an anxious patient is more easily distracted than are controls by fear-related stimuli in the nonattended channel. This finding suggests that the patients have automatic, parallel attention processes that are primed to detect certain types of stimuli. Another approach uses the Stroop paradigm, in which words are presented in different-colored inks, and the person's task is to look at the word and state the color of the ink. Anxious patients have a significant delay in their response times for fear-related words, a finding that suggests that increased attention capacity or cognitive processing is necessary when those words are perceived and the color of the ink is determined.

One theory that explains these findings is the idea of a fear network that encodes fear-related information in a memory structure that is readily accessible and able to influence cognitive, motor, and psychophysiological responses. The theoretical fear networks, which may be similar to mental models, are thought to contain three related forms of information: fear-eliciting stimulus cues, specific response patterns, and the meaning of both the cues and the responses for that particular person. According to this theory patients with anxiety disorders are believed to have fear networks that are especially coherent and stable and that require few environmental cues to become activated.

Patients with posttraumatic stress disorder have been found to have attention biases toward threat-related stimuli specific to the experienced traumatic event. Many of the patients studied were combat veterans, and further work is needed to establish the generalizability of those findings to other forms of posttraumatic stress disorder. The disorder is clinically characterized by both intrusive processes (memories, images, emotions, thoughts) and avoidance elements (psychic and emotional numbing, amnesia, behavioral avoidance of environmental cues resembling the initial trauma). The patients are thought to have a unique configuration of fear networks with stimulus cues (environmental stimuli), response components (cognitive, motoric, and psychophysiological), and meaning elements (e.g., the moral implications of the trauma, survivor guilt, and the meaning of intentional trauma versus accidental trauma).

Some theories argue that states of excessive arousal during trauma impair attention capacity and memory encoding during the event. Emotional processing during and after the traumatic experience may also be hampered by the states of psychophysiological arousal and altered memory storage. The part of the brain necessary for explicit memory processing, the hippocampus, has been shown to be abnormal in individuals suffering from chronic posttraumatic stress disorder. The subsequent clinical syndrome may have some aspects that are adaptive: cognitive attention biases that are primed to detect fear-related stimuli may permit the early detection of threatening situations that, if not avoided, would produce incapacitating psychophysiological arousal. Automatic, nonconscious behavioral avoidance response patterns, embedded in the proposed fear networks or mental models, allow the patients to minimize excessive arousal by avoiding trauma-related situations. Patients who used dissociative mechanisms during and after a traumatic experience appear to be at far greater risk for developing posttraumatic stress disorder. Dissociation is a complex process that essentially involves the dis-association of usually associated processes, such as attention, perception, memory, consciousness, and sense of self.

The psychopathological aspects of posttraumatic stress disorder can be cognitively exemplified as follows: a combat veteran with chronic posttraumatic stress disorder may have no direct recall (impaired explicit memory) of a helicopter crash in which his best friend, who was seated next to him, was killed. Years later continued avoidance of airports, amnesia for combat, general apathy, and social withdrawal (avoidance elements)—combined with startle response, panic attacks, intrusive images, and nightmares (intrusive components)—all suggest intact implicit memory for the combat trauma. The veteran is emotionally, behaviorally, and cognitively impaired.

The evaluation and treatment of patients can be greatly enhanced by an understanding of development, dissociation, and the cognitive processes (including memory) and its careful application to the assessment of presenting symptoms and signs. Several areas that have interested clinicians for decades have become of great societal concern. Two topics that inspire intense controversy are the delayed recall of repressed memories of traumatic events and the suggestibility of patients influenced by clinicians, society, or friends to believe that they are the victims of childhood trauma.

Delayed Recall Many scientists and clinicians believe in the cognitive capacity of patients to be unaware for years or decades of severely traumatic experiences that took place in their childhoods. Other researchers disagree and emphasize the paucity of studies of corroborated cases of childhood trauma that have been followed prospectively into adulthood with documentation of impaired access of consciousness to events presumably stored in memory. A cognitive science view of delayed recall can examine the role of memory processes, development, and the effect of trauma on the processing of information to describe a theoretically coherent but yet-to-be-proved set of mechanisms.

A repressed memory can be thought of as originating from the active, intentional suppression of memory from consciousness. The mechanisms underlying the process then may become automatic and the contents of memory inhibited from retrieval into consciousness. The blockage may exist to avoid flooding the person's awareness with information that is associated with excessive anxiety or fear that would impair normal functioning. This is an example of knowledge isolation in which information may be layered in the nervous system and certain aspects kept from conscious awareness.

In contrast, a traumatic event may be so overwhelming that normal processing may be impaired. If focal attention is divided, the nonfocally attended (traumatic) material will only be processed implicitly. Thus, in order to adapt to a traumatic event, some persons may have the capacity to focus their attention on a nonthreatening aspect of the environment or on their imagination during the trauma; this may be an underlying mechanism in a process called *dissociation*. Traumatic memory that has been only implicitly encoded will affect behavior and emotions and possibly will contain intrusive images and bodily sensations that are devoid of a sense of past, of self, or of something being recalled. This may partly explain the findings of posttraumatic stress disorder with amnesia (blocked conscious access to a memory or its origin) in the setting of avoidance behaviors, hyperarousal, intrusive images, and flashbacks.

Thus, two distinct mechanisms that may explain delayed recall of childhood trauma are the concept of repressed memories and the concept of dissociated memories. A person may use both mechanisms for different aspects of a traumatic event. Recollection of the traumatic memory may take different forms. Repressed memories may have been processed to some degree in narrative form whereas dissociated memories probably lack that more integrative processing. The latter form thus may be experienced as nonpast and nonself, making the intrusive retrieval of dissociated memories a confusing and frightening experience.

Studies of the development of memory in children suggest that the shared construction of narratives about experienced events is often crucial in making the events accessible to long-term retrieval. The establishment of a personal memory system appears to be a function of such memory talk in which parents discuss with children the contents of their memory. In children who have been forced to keep traumatic events a secret, as may occur with childhood abuse, the normal developmental process of narration may be blocked. This may be an additional cognitive mechanism underlying the inaccessibility of some forms of childhood trauma to consciousness in adult patients.

Suggestibility Knowledge isolation, such as in dissociation and repression, provides a theoretical scientific explanation for the underlying mechanisms that may lead to delayed recall of childhood trauma, but clinicians must also be aware of other cognitive processes that influence memory. Numerous studies have demonstrated that the human mind is easily influenced. Human suggestibility can be used to the benefit or detriment of others. Suggestibility is adaptive for a social being that relies on the experiences of others to inform its knowledge of the world and thus to increase its chance for survival. Thus listening to the stories of others, reading a textbook, and being coached in athletics all require the receiver to accept data from the sender. The learner (listener) needs to trust the reliability of the teacher (teller) in order to accept the incoming information. Critical analysis of the data received is an important component of learning. The metacognitive function of assessing the accuracy or usefulness of newly acquired information may be suspended under certain conditions, including hypnosis, drug-altered states, and conditions of severe threat.

Studies of human suggestibility indicate that postevent questioning can bias the metamemory processes that help to determine the source and accuracy of a retrieved memory. The verbal and nonverbal cues given by the interviewer may influence a person to believe that aspects of an event or an entire event that may have never happened actually took place. A person can be convinced of the accuracy of an event despite its lack of correspondence with actual experience. Factors that may influence the biasing of interviewees include a belief in the trustworthiness and authority of the interviewer, not being aware that “I don't know” is a permissible response, repetition of a question that already has been answered, and the interviewer's beliefs as communicated through emotional tone and nonverbal gestures.

It is crucial for clinicians to be aware of human suggestibility in order to avoid iatrogenic distortions. Similarly, it is important for persons who experienced severe trauma early in life to receive informed and empathetic evaluations and treatment. There is a delicate balance between supportive neutrality and active advocacy in assessment and intervention. Awareness of these fundamental cognitive processes may help guide the clinician toward achieving that goal.

Approaches to the treatment of patients with posttraumatic stress disorder need careful evaluation but generally include the view that the impaired emotional processing of the traumatic event requires the active recollection, in explicit terms, of the details of the experience. The process of effectively treating unresolved trauma usually involves the active cognitive processing of specific memories, including emotional responses, derived belief systems, and the psychophysiological

arousal at the time of the traumatic event. The provision of new cognitive information in the course of psychotherapy can in theory alter the configuration of the fear networks and allow previously inaccessible information to be explicitly processed and made available to consciousness for incorporation into an ongoing autobiographical narrative. Specific techniques, such as those which facilitate such cognitive processing of previously dissociated or in other ways isolated cognitions such as beliefs, images, and sensations, may be useful in alleviating the symptoms and dysfunction following acute or chronic trauma. Such therapy and changes in mental processing may diminish the avoidant and intrusive components of the clinical syndrome of posttraumatic stress disorder and improve social, emotional, and cognitive functioning.

FUTURE DIRECTIONS

Cognitive science offers a breadth of conceptualizations for understanding the way in which the mind functions in health and disease. The broad interdisciplinary field provides numerous research paradigms helpful in further elucidating the nature of psychopathology through techniques from the neurosciences to computer models of brain functioning and biological applications of chaos theory. The cognitive understanding of emotions and consciousness may also expand psychiatry's framework for knowing about human subjective experience. Clinical tools, from medications to in-depth psychotherapy, may also find wider application as the processes of self-organization and psychological change are better understood.

Psychiatry in turn has much to offer the field of cognitive science. The long history of descriptive psychopathology and the attempt to synthesize views of the mind and the brain can provide nonclinical cognitive scientists with unique data and relevant questions. Psychiatry is invited to join in the search for understanding the cognitive processes of the human mind.

SUGGESTED CROSS-REFERENCES

Piaget and emotional development are discussed in Section 3.2, memory is discussed in Section 3.4, cognitive disorders in Chapter 10, schizophrenia in Chapter 12, mood disorders in Chapter 14, and anxiety disorders in Chapter 15. Dissociative disorders are discussed in Chapter 18 and personality disorders are discussed in Chapter 24. Behavior therapy is discussed in Section 30.2, hypnosis in Section 30.3, and cognitive therapy in Section 30.6. Mental retardation is discussed in Chapter 34, learning disorders in Chapter 35, pervasive developmental disorder in Chapter 38, and attention-deficit/hyperactivity disorder in Chapter 39.

SECTION REFERENCES

Ali N, Cimino CR: Hemispheric lateralization of perception and memory for emotional verbal stimuli in normal individuals. Neuropsychology *11:*114, 1997.

Andreasen NC: Linking mind and brain in the study of mental illnesses: A project for a scientific psychopathology. Science *275:*1586, 1997.

Broadbent DE: *Decision and Stress*. Academic Press, New York, 1971.

Damasio AR: *Descartes' Error: Emotion, Reason and the Human Brain*. Putnam, New York, 1994.

Douglas VI: Cognitive deficits in children with attention deficit disorder with hyperactivity. In *Attention Deficit Disorder: Criteria, Cognition, Intervention*, LM Bloomingdale, JA Sergeant, editors. Pergamon, Oxford, 1988.

*Edelman G: *Bright Air, Brilliant Fire*. Basic Books, New York, 1992.

*Flavell J, Miller PH, Miller SA: *Cognitive Development*, ed 3. Prentice Hall, Englewood Cliffs, NJ, 1993.

Fogel A, Lyra MCDP, Valsiner J: *Dynamics and Indeterminism in Developmental and Social Processes*. Erlbaum, Hillsdale, NJ, 1997.

Johnson MH, Magaro PA: Effects of mood and severity on memory processes in depression and mania. Psychol Bull *101:*28, 1987.

Johnson-Laird PN: *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Harvard University Press, Cambridge, MA, 1983.

Kandel ER: A new intellectual framework for psychiatry. Am J Psychiatry *155:*457, 1998.

Kihlstrom JF: The cognitive unconscious. Science *237:*1445, 1987.

Kosslyn SM: *Image and Brain: The Resolution of the Imagery Debate*. MIT Press, Cambridge, MA, 1994.

Lewis MD: Self-organising cognitive appraisals. Cognition and Emotion *10:*1, 1996.

Lister RG, Weingartner HJ, editors: *Perspectives on Cognitive Neuroscience*. Oxford University Press, New York, 1991.

Litz BT, Keane KM: Information processing in anxiety disorders: Application to the understanding of posttraumatic stress disorder. Clin Psychol Rev *9:*243, 1989.

MacLeod C: Mood disorders and cognition. In *Cognitive Psychology: An International Review*, MW Eysenck, editor. Wiley, Chichester, England, 1990.

Main M: Metacognitive knowledge, metacognitive monitoring, and singular (coherent) vs. multiple (incoherent) models of attachment: Findings and directions for future research. In *Attachment Across the Life Cycle*, P Marris, J Stevenson-Hinde, C Parkes, editors. Routledge & Kegan Paul, New York, 1991.

*Metcalfe J, Shimamura AP: *Metacognition: Knowing About Knowing*. MIT Press, Cambridge, MA, 1994.

Mesulam MM: Review article: From sensation to cognition. Brain *121:*1013, 1998.

Milner B, Squire LR, Kandel ER: Cognitive neuroscience and the study of memory. Neuron *20:*445, 1998.

Morris RGM, editor: *Parallel Distributed Processing: Implications for Psychology and Neurobiology*. Clarendon, Oxford, 1989.

O'Mara S, Walsh V, editors: The cognitive neuropsychology of attention. Cogn Neuropsychol *11:*96, 1994.

Osherson DN, Smith EE, editors: *Thinking: An Invitation to Cognitive Science*, vol 3. MIT Press, Cambridge, MA, 1990.

*Posner MI, editor: Foundations of Cognitive Science. MIT Press, Cambridge, MA, 1989.

Schore AN: *Affect Regulation and the Origin of the Self*. Erlbaum, Hillsdale, NJ 1994.

Siegel DJ: Cognition, memory and dissociation. Child Adolesc Psychiatr Clin North Am *5:*509, 1996.

*Siegel DJ: *The Developing Mind: Toward a Neurobiology of Interpersonal Experience*. Guilford, New York, 1999.

Sigman M: What are the core deficits in autism? In *Atypical Cognitive Deficits in Developmental Disorders: Implications for Brain Function*, SH Broman, J Grafman, editors. Erlbaum, Hillsdale, NJ, 1994.

Springer SP, Deutsch G: *Left Brain, Right Brain*, ed 4. WH Freeman, New York, 1993.

Steinhauer SR, Gruzelier JH, Zubin J, editors: *Handbook of Schizophrenia*, vol 5, *Neuropsychology and Information Processing*. Elsevier Science, New York, 1991.

Watts FN, editor: *Neuropsychological Perspectives on Emotion*, vol 7, *Cognition and Emotion*. Erlbaum, Hillsdale, NJ, 1993.

Wheeler MA, Stuss DT, Tulving E: Toward a theory of episodic memory: The frontal lobes and autonoetic consciousness. Psychol Bull *121:*331, 1997.

3.2 EXTENDING PIAGET'S APPROACH TO INTELLECTUAL FUNCTIONING

STANLEY I. GREENSPAN, M.D., AND JOHN F. CURRY, ĥH.D.

Jean Piaget (1896–1980) was born in Neuchatel, Switzerland. Building on the work of the German philosopher Immanuel Kant, Piaget explored how intelligence develops from a child's actions on his world. While Piaget focused predominantly on the development of impersonal cognition, he recognized affect also had an important role. Lev Semenovich Vygotsky and Reuven Feuerstein further explored the social context of learning. More recently, the authors have developed a model to understand how affect and cognition unfold as integral parts of the same developmental process.

GENETIC EPISTEMOLOGY

Widely renowned as a child (or developmental) psychologist, Piaget referred to himself primarily as a genetic epistemologist. That self-designation reveals at once that Piaget's central project was not the articulation of a child psychology, as this term is generally understood, but rather an account of the progressive development of human knowledge.

On the classic question of the origins of knowledge Piaget was neither a nativist nor an empiricist; however, his position should not be considered an amorphous form of interactionism. Piaget stated in detail the interactionist position to which he ascribed. It is, in his words, a “constructivist structuralism,” according to which the origin of mental structures is to be sought in the actions of the subject (the child) on objects as the subject strives to adapt to its environment. Structures are constructed within the subject as a consequence of interactions between subject and object. What Piaget judged to be innate is an intelligent functioning that makes possible the production of progressively more adequate structures of knowledge on the basis of abstraction from actions performed during the stages of development.

For example, the concept of space is a fundamental mental structure developed in the earliest period of children's lives. In earliest infancy the child is aware of not one homogeneous space but several heterogeneous spaces, each centered on a certain part of the child's body (for example, visual space and tactile space). As children act on objects that may traverse these various spaces (e.g., a rattle occupying visual, tactile, and auditory spaces), they come to coordinate these individual spaces. Eventually, actions representing displacements in space are organized mentally into the general concept of space. That concept is a structure that can be described in logicomathematical terms.

EQUILIBRATION

For Piaget the general criterion for intelligent functioning is equilibration, briefly defined as “a compensation for an external disturbance.” Hans Furth described equilibration as “the factor that internally structures the developing intelligence. It provides the self-regulation by which intelligence develops in adapting to external and internal changes.” At every level of development the equilibration mechanism operates to further adaptation, but as development proceeds toward the highest level of cognitive functioning, equilibration becomes progressively more adequate in enabling the organism to adapt to a wider range of internal and external disturbances. Piaget's notion of intelligence as adaptation is therefore essentially bound to an equilibration model of intelligent functioning.

ASSIMILATION AND ACCOMMODATION

To explicate the equilibration model further requires introducing Piaget's concepts of the assimilation and accommodation processes. The biological foundation of Piaget's developmental theory is nowhere more clearly evident; these processes are considered functional invariants of all intelligent behavior. At every level of intellectual development, from infancy to adulthood, the processes operate in adaptation.

The assimilation-accommodation account of development stresses the interaction between organism and environment. A certain readiness within the organism is postulated to be necessary for change or development to occur. In Piaget's view, associationism (empiricism) in psychology has committed the fallacy of crediting only half of those conditions necessary for learning with all explanatory power. A full account of human development must include not only the influence of stimuli on respondents (S ® R) but also the influence of the responding organism on incoming stimuli (S ¬ R). Such an account is provided by Piaget's assimilation-accommodation viewpoint:

From a biological point of view, assimilation is the integration of external elements into evolving or completed structures of an organism. In its usual connotation, the assimilation of food consists of a chemical transformation that incorporates it into the substance of the organism.

Furth referred to assimilation as “an inward-directed tendency of a structure to draw environmental events towards itself.” Assimilation is the conservative side of intellectual development, ensuring continuity and coherence by incorporating new aliments into existing mental structures. However, assimilation alone cannot account for growth or change within those structures.

Accommodation occurs during the developmental periods when new data cannot be wholly assimilated into the child's existing mental structures, and yet the data are not so foreign to those structures that they can be ignored. Furth referred to accommodation as “an organism-outward tendency of the inner structure to adapt itself to a particular environmental event.” Read Tuddenham pointed out the variations in accommodation relative to levels of intellectual development as follows:

At the lowest psychological level, accommodation refers to the gradual adaptation of the reflexes to new stimulus conditions—what others have called conditioning or stimulus generalization. At higher levels it refers to the coordination of thought patterns to one another and to external reality.

This line of thought reveals in what sense intelligence is defined by Piaget in terms of equilibration. His equilibration is not a static, balanced system, but a dynamic, or mobile, equilibration between assimilation and accommodation s the child responds to the environment.

STRUCTURALISM

Intelligence has been discussed in terms of an equilibration process involving assimilation of ailments to structures and accommodation of structures to new, somewhat different ailments. Before the stages of intellectual development through which that process passes can be analyzed, one must more fully understand Piaget's notion of intellectual structure and to take a more fundamental look at the origins and developmental forms of the cognitive structures.

The term Piaget used for a cognitive structure is scheme (schema): “A scheme is the structure or organization of actions as they are transferred or generalized by repetition in similar or analogous circumstances.” Schemata exist in the infant as perceptual-motor behavior patterns (e.g., the grasping reflex). They also exist in mature intelligence, although as Furth pointed out, *schema* is more commonly used to refer to an early mental structure, whereas general schemata resulting from use of higher intelligence are referred to as *operations*.

The abstraction process that leads to the formation of cognitive structures is called *reflective* or *formal abstraction*. It is an abstraction from *actions*, according to which the similarities inherent in various behavioral acts are dissociated from their particularized contexts. (See, for example, the earlier example of the rattle in space.) According to Furth and colleagues, “More precisely, reflective abstractions are an enriching feedback into the structures of the organism from the most general coordinations of actions.”

THEORY OF STAGES

An integral part of Piaget's theory on genetic epistemology is a psychology of cognition that seeks to describe how knowledge develops and changes. The genetic framework for that and the process of intellectual adaptation during the ajor early periods of life are provided in Piaget's theory of the stages of cognitive development.

The stages of cognitive development that Piaget and his associates delineated empirically are not defined merely by the dominance of some aspect that remains present but less-dominant throughout development. Rather, each stage constitutes a structured whole that can be defined by a set of criteria.

Piaget was not entirely consistent concerning his stages of cognitive development, but the only possible source of confusion in his later writings is whether the so-called preoperational period is to be considered apart from the period of concrete operations in which it culminates. John Flavell's 1963 study and Piaget's 1983 summary define three major periods and one subperiod of intellectual development (Table 3.2-1). These periods contain subdivisions called stages. The major developmental periods are as follows:

Age (Years)	Period	Cognitive Developmental Characteristics
0–1½ to 2)	Sensorimotor	Divided into six stages, characterized by (1) inborn motor and sensory reflexes (2) primary circular reaction (3) secondary circular reaction (4) use of familiar means to obtain ends (5) tertiary circular reaction and discovery through active experimentation (6) insight and object permanence
2–7	Preoperations subperiod*	Deferred imitation, symbolic play, graphic imagery (drawing), mental imagery, language
7–11	Concrete operations	Conservation of quantity, weight, volume, length, and time based on reversibility by inversion or reciprocity; operations; class inclusion and seriation
11 through end of adolescence	Formal operations	Combinatorial system, whereby variables are isolated and all possible combinations are examined; hypothetical-deductive thinking

* This subperiod is considered by some authors to be a separate developmental period.

Table 3.2-1 Stages of Intellectual Development Postulated by Piaget

1. The sensorimotor period, which extends from birth until approximately 1½ years of age. The period is divided into six stages, which are described in general below with reference to the development of the concept of the permanent object.
2. A period of preparation for, and acquisition of, concrete operations. This period extends from the appearance (at about age 2) of the symbolic (semiotic) function to the beginning (at about age 7) of higher mental operations applied to concrete objects.
3. The period of formal operations, which begins at approximately 11 years of age. During this period, full adult intelligence develops as the operations are extended to apply to propositional, or hypothetical, thinking.

Sensorimotor Period The sensorimotor period of intelligence is so named because the child's construction of mental schemata is in no way aided by representations, symbols, or thoughts. Rather, schemata depend totally on perceptions and bodily movements.

Stage 1 of sensorimotor development is marked by a relatively few organized reflexes that stand out from the spontaneous general activity of the neonate. Among those early reflexes are the sucking reflex and the palmar reflex. These primitive reflexes take on the nature of the first schema through three types of assimilation: (1) reproductive (repeating the actions); (2) generalizing (repeating the actions on new objects); and (3) recognitory (performing different varieties of the actions on different objects).

Stage 2 contains the first habit and the primary circular reaction. The first habits develop out of the original schemata as they are applied to objects in the environment or parts of the infant's body without any differentiation between means and end. In a primitive state of consciousness, the infant is aware only of action sequences and is not even aware of self. Primary circular reactions occur when by chance the infant experiences a new consequence of a motor act and tries to repeat the act.

In stage 3 of the sensorimotor period an initial distinction between means and end becomes apparent, but in a primitive sense. The infant repeats a particular action pattern that achieved one end for the purpose of achieving other (unrelated) ends. For example, a baby who succeeds in shaking a rattle by pulling a string may repeatedly pull the string in an attempt to effect other sounds or results.

In stages 4 and 5 infants use a variety of available means to obtain particular goals. The distinction between stages 4 and 5 lies in the relative creativity or newness of the means. Stage 4 is marked by use of already familiar means. Stage 5 is marked by a search for new means based on further differentiations of already known schemata and by the tertiary circular reaction. The latter differs from a secondary circular action in that the child no longer produces schemata that were effective in one situation to produce magically efficacious results in every situation. Instead, the child explores the environment and varies means to test for effectiveness. Discovery is a hallmark of stage 5. For example, a child may use a stick to move an object not within reach.

Stage 6 is transitional, leading into the preoperational subperiod. In stage 6 the child becomes capable of inventing new means, not by direct actions on objects but by mental combination. Whereas discovery marked stage 5, insight is a characteristic of stage 6. For example, a child who has seen the father bang on a drawer to loosen it may bang on a toy box to make it easier to open.

During the sensorimotor period, a number of significant concepts are developed, including the child's concepts of space, time, and causality. These categorical concepts develop in a process parallel to the sequence of the six stages outlined above. Most important, during the sensorimotor phase, the child develops the schema of object permanence, the first major victory of conservation and the foundation of all future knowledge.

Schema of Object Permanence The knowledge that objects in the external world have an existence independent of the child's actions on them or interactions with them is a major accomplishment of the sensorimotor period. Flavell has outlined Piaget's observations and interpretations of infants' reactions to the disappearance of interesting objects, which are the foundation for the theory of development of object permanence. In stages 1 and 2, for example, a child simply continues to look at the place where the object was last seen. In stage 3, if an object such as a spoon drops to the floor, the infant will look for it (e.g., by leaning over and looking at the floor). In stage 4 if an object is repeatedly hidden at point A (in sight of the child) and then hidden at point B (also in sight of the child), the child searches for it at the original, rather than the current, hiding place (i.e., at A, not B). Stages 5 and 6 mark the child's increasing understanding of object permanence; the infant is able to follow multiple displacements of the object through points in space, even if the object is hidden within another object.

Preoperational Subperiod and Semiotic Function The advent of the preoperational subperiod is marked by the appearance of what Piaget called the *semiotic function*. This new ability was defined by Piaget and Bärbel Inhelder as follows: “It consists in the ability to represent something (a signified something object, event, conceptual scheme, etc.) by means of a signifier which is differentiated and which serves only a representative purpose: language, mental image, symbolic gesture and so on.”

During the sensorimotor period, a thing could be represented in a limited sense by a part of itself (e.g., the mother's voice might represent the presence of the mother in the room). However, such signifiers are indexes undifferentiated from their significants (the voice is part of the mother). Symbols and signs are signifiers that are differentiated from their significants. They become available to the child only with the appearance of the semiotic function, which makes representational thought possible. As Furth pointed out, representation has first of all an active meaning in Piaget's theory. The child becomes capable of summoning up a symbol or sign to

stand for a given significant. For Piaget, representation is not at the essence of thought; it has an auxiliary function.

Characteristic Behavior Patterns The semiotic function is heralded by five characteristic behavior patterns in evidence during the second year of life: (1) deferred imitation or imitation that starts after the disappearance of the model; (2) symbolic play or the game of pretending; (3) drawing or graphic imagery; (4) mental image, which appears as an internalized imitation and not as a function of perception; and (5) verbal evocation of events not occurring at the time. Each behavior pattern shows the origins of representational thought as the preoperational subperiod of cognitive development begins. For Piaget the semiotic function, which so enlarges the children's worlds—liberating them from the bonds of immediate space and time and enabling them to begin to manipulate symbols and to think rather than just to act on immediately present objects—is rooted in imitation.

IMITATION One can follow the development of imitation through the same six sensorimotor stages delineated for the concept of object permanence. Piaget did this in his volume *Play, Dreams and Imitation in Childhood*. A radically new form of imitation occurs during the second year of life: deferred imitation. For example, a child may put on Father's hat and walk as Father does, even hours after Father has gone off to work.

For Piaget, intelligence is an equilibration process in which assimilation and accommodation are in balance. However, in imitation, accommodation outweighs assimilation. According to Piaget, imitation is behavior in which “the subject's schemes of action are modified by the external world without his utilizing this external world.” In imitation the child's cognitive structures undergo temporary change without simultaneously incorporating new aliment.

SYMBOLIC PLAY A second new behavior pattern that now appears is symbolic play. In imitation the imbalance between assimilation and accommodation is weighted in favor of accommodation; however, the opposite is true in symbolic play, which is a lessening of the demand of the adaptive process.

Development of play can also be followed through the six stages of sensorimotor intelligence, but symbols are used in play only after the sensorimotor period, in the type of play characterized by games of pretending. For example, a little girl may pretend that she is asleep, that a box is her pet cat, or that she herself is a church. In each instance symbols are generated “to express everything in the child's life experience that cannot be formulated and assimilated by means of language alone.”

According to Piaget's theory, these symbols are created by the same process of imitation that gives rise to deferred imitation at this time. In fact, Piaget views imitation as the process underlying the development of the entire semiotic function. In symbolic play, then, symbols are generated by a process in which accommodation outweighs assimilation. But instead of being used accurately (i.e., to represent what they are derived), they used in a process in which a liberating assimilation outweighs accommodation.

DRAWING A third behavior pattern associated with the rise of the semiotic function is drawing, or graphic imagery. Piaget sees elements of both play and imitation in this activity. In developmental terms, he considers drawing “halfway between symbolic play and the mental image,” appearing at about age 2 or 2½ years. Drawing is playful activity, an end in itself characterized by reproductive assimilation; in other words, the child enjoys producing drawings for their own sake. However, graphic play also has accommodative elements, especially as the child grows older and attempts to draw not just formless scribble but some thing.

MENTAL IMAGE Closely related to drawing is the mental image. Piaget tied the genesis of mental imagery to accommodation and imitation. He explicitly denies that mental images can be the product of perception; they are a construction, something the child creates. The mental image is not directly given by perceptual input; it is constructed by the process of accommodation.

VERBAL EVOCATION OF EVENTS The fifth behavior pattern associated with the rise of the semiotic function concerns language, the verbal evocation of events that are not present. Piaget gave the example of a little girl saying “Anpa, bye-bye” (Grandpa went away) while pointing to the path he had taken when he left. The parallel with deferred imitation is clear, but the new representational ability is supported by the social system of language.

Concrete Operations A crucial difference between preoperational and concrete-operational thought is the presence within operative thinking of concepts of conservation. When concrete operations have been organized into a system, the child can conserve, that is, “discover what values do remain invariant . . . in the course of any given kind of change or transformation.” The progressive and continual structure building that occurs in the concrete operational period is evident in the increase, with development and age, in the scope of such concepts, such as conservation of quantity, substance, and number.

Conservation of Quantity The clearest sign that a child remains in the preoperational subperiod is the absence of the concept of conservation. For example, if liquid is poured from a short, wide glass into a tall, narrow one, the child in the preoperational stage thinks the amount of liquid has changed. At the level of concrete operations, however, children are no longer overwhelmed by the perceptual discrepancy between the two configurations. They begin to reason about the transformation, and their correct judgments regarding the conservation of quantity of liquid are accompanied by explanations grounded in logical properties. It is assumed that children are not aware of the logic they use.

When problems of conservation begin to be solved, the child passes from the preoperational subperiod into the period of concrete operations, for which the former was a long time of transition and preparation.

Conservation of Substance At the age (on average) of about 7 or 8 the child can solve the conservation-of-quantity problem and can perform similar judgments of conservation when, for example, a lump of clay is transformed in shape. Between the ages of 9 and 10, the child discovers that the weight of a given object is also conserved even when its shape is transformed. However, not until approximately age 11 or 12 do children have a logical comprehension that the volume displaced by a given object is conserved even after transformation of the object's shape. Conservation entails the logical certainty that one characteristic of an object remains invariant while the object itself undergoes some type of perceived transformation.

Concept of Cardinal Numbers The concept of cardinal numbers also develops from an initially nonconserving to a conserving stage. Children in the preoperational subperiod can be presented with two horizontal rows of colored dots in one-to-one correspondence (i.e., imaginary vertical lines could be constructed between each red dot and its corresponding blue dot). When the experimenter destroys this optical correspondence by spreading out one of the rows of dots, the child in the preoperational period thinks the larger row contains more dots. Only after conservation of cardinal number has been established as a logical necessity does the child maintain the numerical equivalence of the spread-out row. Clearly, preoperational concepts of number provide inadequate bases for arithmetic skills. It is possible that a lag in the development of number conservation could underlie certain types of arithmetic-related learning disabilities. Again, that points to the importance of children's active experience as a foundation for subsequent concept formation.

Operations Notions of conservation are the mark of well-established concrete operational thinking; thus one must understand the meaning of *operation* in Piaget's thought. Operations, themselves, constitute essential thinking. For Piaget an operation is an action that is (1) interiorized, (2) reversible, and (3) part of an organized system of such actions.

The operations that form this system are interiorized actions. In the sensorimotor period external behavior patterns gave rise through a process of abstraction to the construction of sensorimotor schemata. In similar fashion internal thinking patterns later give rise to operations. According to Furth, the generalizable aspects of actions, “those which can be found in any coordination of action,” enter into the construction of operations. Saying that the crucial aspect of actions in this regard is their generalizability explains the importance of interiorization in the construction of operations. *Interiorization* refers to “the increasing dissociation of general form from particular content.” In other words, the notions of generalizability and interiorization merely point out the process of abstraction that is occurring. For example, a child adds two apples and three apples to obtain five apples. In another instance a child adds seven blocks and one block to obtain eight blocks. In a third instance a child combines the category of fathers with that of mothers to obtain the category of parents. The operation abstracted from these three mental actions is addition or combining, without reference to the particular content of numbers, objects, or categories.

Not only must an operation be interiorized action, it must also be reversible. The action of combining (addition) is not an operation until its relationship to the action of separating (subtraction) is comprehended. To understand reversibility is to understand the third criterion of an operation, its inclusion in a system.

The reversibility essential to operatory thought may be either inversion or reciprocity. In reversibility by inversion, an action + A is reversed by – A. For example, in the above conservation-of-quantity example, pouring liquid into container 2 (+ A) may be mentally reversed, that is, mentally pouring it back into container 1 (– A). In reversibility by reciprocity a relation A < B is reversed by a relation B < A. Referring again to the conservation-of-quantity example, let A stand for container 1, and B stand for container 2. The rising height of liquid in container 2 (A < B) is offset by its narrower width (B < A).

Corresponding to these two types of reversibility are the two major categories of concrete operations: those pertaining to classes and those pertaining to relations. In the system of operations performed on classes, reversibility is by inversion; in those performed on relations, it is by reciprocity. For example, subtraction and addition relate to inversion; comparing sticks of different sizes relates to reciprocity.

Class Inclusion The concrete operation demonstrating understanding of classes is the class inclusion task. In this task a child is shown, for example, an array of pets (superordinate class) consisting of dogs and cats (subordinate classes). After counting the number of dogs, cats, and pets, the child is asked whether there are more dogs or more pets. Children in the preoperational subperiod cannot keep in mind the superordinate class while perceiving only the subordinate classes. Thus, they fail the task frequently over a series of such arrays.

Relations The concrete operation that demonstrates an understanding of relations is seriation. Children are asked, for example, to arrange a set of rods in order of increasing size. Children in the preoperational subperiod may subgroup the rods but will have difficulty completing an entire array along the required dimension. They may understand the concept of smaller versus larger but have difficulty comprehending the gradual nature of change.

Formal Operations In the third and final stage in Piaget's conception of the intellectual development in the child the logical structures of concrete operations are superseded by structures referred to as formal operations.

The relationship between the real and the possible that characterizes adolescent thinking represents a reversal of that relationship in the thinking of the concrete operational child. Inhelder and Piaget note that the real has priority for the younger child and that possibility is conceived of merely as a prolongation or extension of real operations, “as, for example, when, after having ordered several objects in a series, the subject knows that he could do the same with others.” For the adolescent, however, the possible has priority and the real is seen as a particular instance of it. “Henceforth, they conceive of the given facts as that sector of a set of possible transformations that has actually come about.” This immediately presupposes that the adolescent can take a given empirical event (such as, “the long, thin rod bends”) and categorize it within a system of possible combinations of events (e.g., long rods or short rods, thin rods or thick rods, bending or not bending). Three characteristics follow from this fundamental reorientation in thought: (1) adolescent thought is hypothetical-deductive; (2) it deals in propositions rather than in concrete events; and (3) it can isolate variables and examine all possible combinations of variables.

Hypothetical-Deductive Thought As a hypothetical-deductive form of thought, formal operational intelligence proceeds from the possible to the real. In this sense, it mirrors scientific reasoning. The implications of a propositional statement are drawn and then tested against reality. Rather than building up a proposition by induction from disparate concrete examples to a loose generalization, formal intelligence operates systematically from general statement to particular instance by means of testable hypotheses. In Flavell's words, “To try to discover the real among the possible implies that one first entertains the possible as a set of hypotheses to be successively confirmed or infirmed. Hypotheses which the facts infirm can then be discarded; those which the data confirm then go to join the reality sector.”

Propositional Thought Saying that formal operations deal in propositions rather than in concrete events, implies increased freedom from immediate content, with a correspondingly greater intellectual mobility. At one level this freedom implies the ability to manipulate abstractions that have been tied to concrete examples or events. The adolescent, for example, can perform a transitive inference (A < B, B < C; therefore, A < C) without any empirical demonstration of referents for the terms A and B. At another level this freedom implies that having performed a concrete operation, the adolescent can abstract the results of that operation and perform further operations on them. For example, an adolescent can perform the concrete operation of combining two liquids to observe the color of the resultant mix and then take the result of this operation and systematically relate it to results of all other combinations of available liquids.

Isolating Variables and Examining Combinations The example that follows helps explain the third characteristic of adolescent thought mentioned by Flavell, the isolation of variables and the examination of all possible combinations. Instead of dealing with disparate concrete experiments, hypothetical-deductive adolescents can organize their investigations into a coherent pattern a priori, then perform all relevant combinations of variables to test their hypotheses, thus isolating causal factors. Piaget's theory of formal operational cognition focused on scientific thinking. For example, the weight, speed, shape, and size of an object may all be seen to contribute to the size of a hole the object will make when hitting the ground.

Children in the preoperational substage merely describe what they see, and causal thinking is expressed in an undifferentiated form (e.g., “It has to”). The child with concrete operational thinking can categorize and order the relevant variables independently but has difficulty integrating the system of all relevant variables. The adolescent, however, can generate all possible combinations of relevant variables and then systematically test the importance of each variable.

A complete combinatorial system appears only during the period of formal operations. Instead of focusing on empirical givens, as a child in the concrete operational stage does, the adolescent using formal operational thinking constructs a hypothetical system comprising the empirical givens. Whereas the younger child could classify events according to various categories, such as length, width, and weight, the adolescent uses that classification as a basis for abstracting all possible combinations of variables. Having done this, the adolescent can then test hypotheses derived from the combinatorial system. The result of this new ability is the capacity to test the causal significance of each individual factor in succession by holding all other factors constant.

Piaget interprets the rise of formal operational thought in the context of his equilibrium model of cognitive development. Thus, he considers neurological maturation and experience of the object and interpersonal world as necessary but not sufficient to explain this qualitative improvement in thinking. In essence, the equilibration explanation is as follows: During the stage of concrete operations, a number of qualitatively heterogeneous factors are constructed by the child, resulting in the achievement of conservation of the factor in question even in the face of perceptual transformations. Such factors include quantity, weight, volume, time, and length. Eventually, the child discovers that in many concrete instances the operation of these factors is interrelated. Thus, although the factors have been constructed mentally in relative isolation from one another, their presence in real objects is mixed. Through experience with both impersonal and interpersonal objects, the child's concrete operational understanding of these factors is shown to be insufficient, and a more comprehensive, more intelligent understanding is stimulated.

EGOCENTRISM

Each major period of cognitive development is characterized by a qualitative shift toward more comprehensive and more adaptive cognitive structures. In this sense, the adolescent is more intelligent than the infant. However, each transition to a higher level of cognitive organization is initially accompanied by a lack of full differentiation between self and object. Each period has an early organizational phase that is followed by the phase of accomplishment of cognitive developmental tasks. During the early organizational phase, the child's failure to differentiate fully the self from objects is manifest in behavior reflecting stage-specific forms of egocentrism. David Elkind has summarized the process. Each developmental period has characteristic forms of egocentrism.

In the sensorimotor period, egocentrism refers literally to a lack of differentiation between self and object, as perceived in the lack of object permanence. The existence of objects independent of action patterns of the self is not acknowledged. In the preoperational subperiod the capacity to engage in symbolic thinking is accompanied by initial failure to differentiate fully between symbols and their referents. That may be manifest, for example, in failure to differentiate such mental images as dreams from real objects. In the concrete operational period the capacity to engage in logical operations is accompanied by an unrealistic certainty in which probability is not appreciated and mental construction of the self (self-definition) is not differentiated from facts. Finally, at adolescence the capacity to engage in hypothetical thinking and to understand others' points of view is accompanied by characteristic patterns of thought in which others are unrealistically presumed to be focusing on the self. As Elkind has pointed out, adolescent egocentrism is a “belief that others are preoccupied with (the adolescent's) appearance and behavior,” when, in fact, the adolescent is preoccupied with these topics.

EXTENSIONS OF PIAGET'S THEORY

In his classic statement of theory and in the bulk of his scientific research, Piaget focused on the development of logicomathematical reasoning, the operative aspects of thinking. Others, however, have extended a similar structural model to the investigation of other domains. The best known example of such an extension of Piagetian theory is probably Lawrence Kohlberg's theory of the stages of moral reasoning. Kohlberg proposed six stages in the development of moral reasoning. These begin with the moral stance of the preschool child, based on reward or punishment, and proceed through middle childhood, when moral judgment tends to be based on desire to obtain the approval of others or to uphold authority. Only with adolescence does the morality of internalized principles, such as the rights of individuals or the notion of a natural law, begin to take effect. In Kohlberg's model each major stage of moral development requires that the child attain a commensurate level of cognitive development. For example, conventional morality, which develops during middle childhood, requires concrete operational thinking, and principled morality requires formal operational thinking.

A broader area of development in which Piaget's observations have been extended has been referred to as *social cognition*. A central construct throughout Piaget's developmental model is egocentricity versus decentration. In the social domain decentration has been referred to as role taking, or the ability to take the perspective of others. That skill may be observed at a perceptual, a cognitive, or an affective level. Perceptual role-taking involves the ability to recognize how a perceptual array appears from the perspective of another when that perspective differs from one's own. Cognitive and affective role-taking involve analogous processes in which the thoughts and feelings of others are taken into account when they differ from one's own. Thus, role taking is a condition necessary for the development of empathy. Kohlberg considered it also a necessary condition for the development of higher levels of moral reasoning.

In general, social cognitive processes refer to role taking and communication skills that enable children, adolescents, and adults to understand one another's thoughts and feelings. Social cognitive processes are central to the concerns of psychiatry, because deficiencies in these processes impair communication and correlate with symptomatology and because reconstruction of social cognitive processes is a direct concern of psychosocial interventions. For example, the deficiencies in social cognition that characterize aggressive, conduct-disordered boys have been delineated by Kenneth Dodge, John Lochman, and others and then identified as targets for intervention with these youngsters.

Flavell recently reviewed three decades of work in the area of perspective taking or role taking. After a period focused on communication skills, perspective-taking researchers concentrated on cognitive processes in which children develop knowledge of themselves and others as learners and rememberers. For example, during development children become aware of strategies for encoding bits of information that need to be recalled. Studies of such metacognition are another example of the extension of a Piagetian model to domains of practical importance to psychiatry and applied psychology. Deficiencies in metacognitive skills have been identified in youngsters with attention-deficit/hyperactivity disorder or learning disorders and can serve as targets for psychoeducational intervention.

Recently, post-Piagetian research on perspective taking has focused on what has been described as children's “theory of mind.” This pertains to children's awareness that others have cognitive processes similar to their own, or that others have an internal mental state and the ability to represent the mental states of others in their own mind. An example of a theory-of-mind task is the false-belief task. In that task, a child is shown a popular, brand name candy box and asked what it contains. The child is then shown and told that the box does not contain candy but instead contains a pencil. The pencil is then put back into the box, and the child is asked to predict what another child would say is in the box and to say what is actually in the box. Thus, the task involves predicting an erroneous belief in another that differs from one's own correct belief. Young children master this task at about age 4. However, children with autistic disorder show specific deficiencies on this task, whereas children in whom language is merely delayed have no such difficulty. Thus, “theory of mind” may be an area of cognitive deficiency specific to autistic disorders.

NEO-PIAGETIAN RESPONSE TO CRITICISMS OF PIAGET'S THEORY

Major criticisms of Piaget's theory have emanated from the Anglo-American empirical tradition and from sociocultural theorists. Empiricists have pointed to findings, including the effectiveness of training interventions in accelerating Piaget's proposed pace of cognitive development and to weak correlations among scores on tasks that purportedly measure the same underlying cognitive structure. Sociocultural theorists consider Piaget's emphasis on logicomathematical reasoning and the isolation and manipulation of variables as a bias inherent in Western forms of thought. Cross-cultural research has shown that different tasks are mastered at different ages in different cultures. In addition to these external critics Piaget has been criticized within his own school of thought for failing to address individual differences in cognitive abilities, for failing to investigate cognitive development during adulthood, and for giving an insufficient explanation of the processes by which stage transitions occur.

Neo-Piagetian theorists, such as Juan Pascual-Leone and Robbie Case, attempt to address these criticisms while retaining the core assumptions of Piaget's theory. Instead of positing broad cognitive structures that subsume a variety of specific tasks, these theorists focus more on narrow-band structures and elaborate on the mechanisms or processes by which such structures change. In so doing they have moved neo-Piagetian theory closer to other branches of cognitive science and have delineated processes that are more readily studied in relation to biological or neurological variables.

Pascual-Leone, for example, has emphasized the role of attention in the process of structural changes. To move from a lower to a higher level cognitive structure a child must first inhibit the application of an old cognitive structure to an environmental stimulus or aliment and then coordinate a new set of schemata through mental effort. That effort involves deployment and maintenance of attention. On the other hand, Case emphasizes the cognitive processes of problem solving and exploration in the transition from lower to higher levels of cognitive structure. Both theorists take the view that change occurs on one or a few structures at a time and that which specific structure undergoes change is a function of individual differences or environmental opportunities. Neo-Piagetian theory is thus less universal in its model of cognitive development and more accommodating of cultural and individual differences. Nevertheless it retains the basic Piagetian concepts of an active, developing child who is constructing cognitive structures on the basis of interaction with the external and social environment. However, even neo-Piagetian theory has not solved the key challenge of how to fully integrate emotional and intellectual development.

NEW CONCEPTS OF INTELLIGENCE: EMOTIONAL BASIS OF INTELLIGENCE

Human emotions have traditionally been viewed as somewhat separate from cognition and as a minor concern to overall development. New clinical observations and theoretical formulations by the first author and emerging findings from a number of recent studies suggest that emotions are central to cognition and may actually regulate and orchestrate cognitive capacities. They may also be critical to development of cognitive capacities. It is suggested that babies' emotional exchanges with their caregivers, rather than their ability to complete cognitive tasks, should become the primary measure of developmental and intellectual competence.

Twelve-month-old Cara sits in her mother's lap at a table, eyes locked onto the psychologist, who tries to get her to follow the bean he is putting under the cup and search for it. Cara knocks over the cup. Is this little girl, as her mother fears, cognitively delayed? Does a 1-year-old child who never babbles like other children her age and violently flings food and toys away from her show signs of a significant intellectual deficit? After a battery of similarly frustrating tests, the evaluator concludes that cognitive delay is the likely diagnosis in Cara's case.

For 50 years developmental testers have expected babies to sit still in their mothers' laps, pay attention, and perform prescribed tasks while adults assess their basic intelligence. Traditional wisdom has long insisted that carefully scoring how well a tiny child fits pegs into boards, sorts cards by shape, or hunts beads under cups can reveal an accurate measure of intelligence and developmental competence. However, recent results from research and clinical practice by the first author and others suggest that this entire approach to assessing children's capacities rests on false premises and has inadvertently led to mistaken diagnoses that can stigmatize children throughout their school years.

When an evaluator schooled in this new thinking assessed Cara, he focused on her spontaneous interactions with her caregiver. He looked at each of Cara's intentional behaviors as a sign of her emotional interests. For instance, he observed her delight in yanking her mother's nose. At the assessor's suggestion the mother permitted the tugging on her nose to continue and playfully responded, “Toot, toot.” Cara smiled and pulled again. The baby was rewarded with another “Toot, toot” and a big smile from her mother. Cara soon began to copy her mother's gestures and eagerly thrust her nose towards Mom. When the mother squeezed Cara's nose, the 1-year-old girl chirped, “Mo, mo,” her very first words.

Cara showed that she could initiate social interactions and comprehend their consequences. That demonstrated degree of understanding put Cara at least at the 12-month level of cognitive development. Further observation revealed an extremely energetic, active, highly physical toddler who liked to have her way and control her surroundings. With the consultant's help, Cara's mother later altered her parenting style. She learned to follow Cara's behavioral lead, then enthusiastically engage her daughter in creative interactions while simultaneously setting firm limits. Cara's energy quickly became more focused; her babbling, richer. Before long she was saying real words and actively cooperating with her parents.

If a series of such simple, pleasurable interactions with her mother could reveal and foster Cara's language development and organizational ability, then any conception of intellect that marked her as cognitively delayed because of an inability to search for a bean has serious flaws. Those flaws are based on a long-standing mistaken belief that the intellect is superior to and supervises the passions. Clearly, as Cara's linguistic debut demonstrates, analysis of a child's early relationships and sensory and emotional experiences is a vital key to accurate assessment of intelligence and developmental competence.

Until now, however, no one has offered an explanation of how emotions give birth to intelligence. In fact, a baby's earliest feelings play a pivotal role in all later intellectual development. Unlikely as the connection between feeling states and intelligence may seem, the emotions orchestrate a vast array of cognitive operations

throughout an individual's life span. Indeed, they make possible all creative thought.

Results from four distinct lines of inquiry have recently shed new light on the importance of emotions for intelligence. In work with Arnold Sameroff, this first author has found that children with four or more family emotional risk factors had 24 times the chance of scoring an intelligence quotient (I.Q.) below 80 than children without those risks. Stephen Porges and the first author have shown that measurements in 8-month-old infants of a part of the brain that regulates emotions correlate with these same children's I.Q. scores at age 4.

The first author's work with a group of children with autistic disorder, who suffer some of the most severe thinking and language problems imaginable, has also confirmed the inextricable linkage of emotional and cognitive development. Therapeutic programs for these severely challenged children have traditionally concentrated on trying to stimulate their cognition and teach them language. However, a program based on emotional cueing (like the one that revealed Cara's true abilities) proved to be more effective for a number of these children in fostering empathy, warmth, and creative thinking.

> One young patient, Ashley, neither spoke nor made any response or eye contact with those around her. The 2-year-old child spent hours staring into space, rubbing persistently at the same patch of rug. Her abnormal repetition was viewed by the clinician observing her as more than just a distressing symptom of her autism. That symptom revealed an underlying interest and motivation that could be harnessed and redirected toward interacting with others. To initiate her cognitive progress the clinician had to first motivate her to communicate with the simplest of emotional gestures—a smile, or smirk, or purposeful hand movements. He suggested that her mother place her hand next to Ashley, on the favorite stretch of rug. When Ashley pushed it away, the mother gently put her hand back. Each time the child pushed, the mother's hand would return. A cat-and-mouse game ensued, and after three sessions of these rudimentary interactions, Ashley was looking, smiling, and anticipating.

From that tiny beginning, through a comprehensive therapeutic program, grew a bridge to emotional relationships and eventual verbal exchanges. For example, as therapy progressed, the therapist helped Ashley use her imagination by repeatedly initiating pretend play. He recognized that each time Ashley repeatedly flung herself on her mother, the child was deriving sensory-based simple pleasure from her behavior. He instructed the mother to whinny like a horse each time Ashley lunged at her. Soon Ashley imitated mother's sounds and then started initiating her own sounds and words. In that way the therapist helped the mother stretch a pleasant sensation for Ashley into a richer, more complex interaction. Over time mother and child pretended to be neighing horses, mooing cows, and barking dogs. Their social and emotional interchange grew increasingly complex, passing through the same series of developmental stages identified in children without difficulties. At age 7 Ashley now enjoys warm friendships, argues as well as her lawyer father, and scores in the low-superior I.Q. range.

A fourth line of inquiry, microscopic clinical observations of children's thinking, further clarifies the relationships between emotion and reason by revealing two necessary elements of thinking. The first process—creating a new idea—stems from the ability to use one's own emotional experience to assign meaning and significance to daily events or concepts. The second process—reflection and logical analysis—examines the newly created idea according to whatever principles of logic the person possesses and places it in a wider frame of reference.

> To understand those processes in action, the authors put a simple question to two young boys seen in therapy not long ago. When asked by the clinician, “What do you think about people who act bossy to you?” Chris replied, “Well, teachers are bosses, babysitters are bosses, policemen are bosses.” That articulate 7-year-old child lacked the emotional pathways that permit creative and intuitive thought. He could provide a formal classification of different types of bosses but could not relate these categories to his own life. However, 7-year-old Josh had no such difficulties. In response to the same question about bosses, he announced, “Most of the time I don't like being bossed, especially when my parents try to tell me when I can watch TV and when I have to go to sleep. I'm big enough to decide for myself. Sometimes when I'm being bad, I guess I need bossing, though. Maybe bosses are okay some of the time, and some of the time they're not.” Josh finds his answer in his own, apparently generally irritating, brushes with bosses. Rather than simply listing categories or incidents, he can abstract a principle from the emotional core of those incidents.

How exactly did Josh's ability to think and abstract develop? A baby's experience begins with sensations like touch and sound. Each sensation, however, also gives rise to an emotion. A toy may feel interesting or boring; a voice, soothing or jarring. Even young infants react to sensations emotionally. They prefer the sound or smell of their mother, for example, to any others and by 4 months of age can react to certain persons with fear. Furthermore, contrary to long-held assumptions, basic sensations like touch and sound can be perceived differently by different people, giving rise to emotional differences.

Emotional meaning also adheres to early concepts like “big and little,” “more and less,” “near and far,” and “now and later.” “A lot” is a bit more than makes a child happy. “Near” is snuggled next to a child in bed. “Later” is a frustrating stretch of waiting. For a child without an intuitive sense of “few and many,” numbers have no meaning. Furthermore, a young child's experience of any sensation always occurs within the context of a relationship that gives it broader meaning. Playing with mother's hair, for example, may evoke smiles and hugs or an angry scolding.

Each sensory experience has such a dual aspect and is labeled by both its physical properties and its emotional qualities. This double coding helps the child both place the memory or experience in a catalogue of experience and retrieve or reconstruct it when needed. As the child grows, emotional reactions come to operate as a sixth sense that allows the child to recognize and understand situations.

Emotion orchestrates complex judgments as well. One of modern psychology's main enigmas is how children learn to discriminate among situations (“When can I yell and kick”) and generalize from one to another (“Should I behave at school like I do at home?”). Consider how a child makes a seemingly simple judgment about when to say “hello.” She doesn't learn a set of cognitive rules like greeting only those who live on her street or only those who wave at her. Rather, from countless specific encounters she abstracts an emotional pattern; there is a feeling of warmth and friendliness in situations that rate “hello.” Her interactions create an emotional signaling system that tells her when to say hello and that it is okay to punt the football but not to kick Sarah or Charlie in the shins. That emotional signaling system, which acts like an orchestra leader for the vast array of cognitive instruments, is a quintessentially human process. No computer, for all its apparent so-called brainpower, can ever get beyond limited elements of logical analysis and think like a person. Advocates of artificial intelligence may claim that current computational capacity limits creative, humanlike thought, but the real limit is a machine's inherent inability to engage the world emotionally. No collection of microchips can ever have a child's lived emotional experience of bosses and “hello,” of noses and rugs. None, therefore, can ever create the emotionally based meaning from which creative thought grows and on which it depends.

Looking at Piaget's theory from this perspective reveals the limitations of a cognitive theory that did not adequately deal with the central role of emotions. Piaget's experiments focused on how children comprehend the relationship between physical objects, developing the ability to classify them by such parameters as shape or size. But most children can classify their emotions and emotionally relevant relationships far earlier than they can physical objects. For example, they know members of their families from those who are not members, classifying the family as a unit. Some of Piaget's observations are limited because he depended so heavily on children's perceptual and motor performance to signal cognitive advances, even though motor skills often lag behind other skills.

More important, however, is Piaget's relative lack of focus on the role of emotions. He emphasized learning through doing but did not realize that the “doing” generates formative emotional reactions as well as perceptual, motor, and cognitive ones. Consider how a child learns what an apple is. You can have her handle an apple and determine it is something red and round, bigger than a peanut and smaller than a watermelon. Alternatively, the child can observe the aspects mentioned above, experience more satisfaction eating an apple when she is hungry than when she is full, and know the pleasure of giving one to her favorite teacher.

She can also imagine how the teacher feels when she gives her the apple. She may catalogue how she herself feels when she throws one at her younger brother and gets him on the shoulder, as well as her disgust when another one rots or she discovers half a worm inside. If you ask creative adults to write an essay on apples, they will probably bring an enormous amount of personal affective experience to their reflections.

Piaget did emphasize how children's thinking comes to incorporate multiple perspectives as they grow older. A classic Piagetian experiment shows how school-age children learn to solve a problem involving weights on a seesaw. Assessing the heaviness of the weights and noticing where they are placed, they are able to figure out how a seesaw works. But what was not appreciated by Piaget was that the children's postulated perspectives incorporated the additional almost infinite number of perceptions afforded by affective experiences. To neglect this element is therefore to fail to appreciate the rich array of experiences that contribute to forming abstract concepts.

TOWARD A GENERAL DEVELOPMENTAL MODEL

The diagnosis and treatment of emotional and developmental disorders in infants and young children requires that clinician take into account all facets of the child's experience. Thus one needs a model with which to look at how constitutional-maturational (regulatory), family, and interactive factors work together as the child progresses through each developmental phase, and each phase must be viewed from both affective and cognitive perspectives.

A developmental, structuralist model formulated by the first author integrates cognitive and affective development and applies the types of structure Piaget described to a range of experience. Most important, the model lso considers individual differences in terms of biology and interaction. New findings suggest that early interaction can alter the structure and wiring of the central nervous system. In this model, the biological differences express themselves in the unique way an infant processes sensations and organizes motor patterns. Interactions harness and change these basic processes. The model can be visualized with the infant's constitutional-maturational patterns on one side and the infant's environment, including caregivers, family, community, and culture, on the other side. Both sets of factors operate through the infant-caregiver relationship, which can be pictured in the middle. Those factors and the infant-caregiver relationship in turn contribute to the organization of experience at each of six developmental levels (consistent with both cognitive and affective milestones), which may be pictured just beneath the infant-caregiver relationship.

This particular clinical and research model enables the user to look at the back-and-forth influence of highly specific, verifiable constitutional-maturational factors on interactive and family patterns and vice versa, in relationship to specific developmental processes (and to relate these processes to later developmental and psychopathological disorders).

Developmental Levels The model contains six developmental levels, which include the infant-child's ability to accomplish the following:

1. Attend to multisensory affective experience and at the same time organize a calm, regulated state and experience pleasure.
2. Engage with and evidence affective preference and pleasure for a caregiver.
3. Initiate and respond to two-way presymbolic gestural communication.
4. Organize chains of two-way communication (opening and closing many circles of communication in a row), maintain communication across space, integrate affective polarities, and synthesize an emerging prerepresentational organization of self and other.
5. Represent (symbolize) affective experience (e.g., pretend play, functional use of language), which calls for higher-level auditory and verbal sequencing ability.
6. Create representational (symbolic) categories and gradually build conceptual bridges between these categories. This ability creates the foundation for such basic personality functions as reality testing, impulse control, self-other representational differentiation, affect labeling and discrimination, stable mood, and a sense of time and space that allows logical planning. This ability rests not only on complex auditory and verbal processing abilities, but also on visual-spatial abstracting.

At each level, one looks at the range of emotional themes organized (e.g., can the child play out [symbolize] only dependency themes and not aggressive ones? is aggression behaved out and dealt with presymbolically?). One also looks at the stability of each level. Does a minor stress lead a child to lose the ability to represent, interact, engage, or attend?

In their use in day-to-day clinical work the six developmental levels can be collapsed into four essential processes that characterize development in infants and young children. These processes concern how an infant and the parents or caregivers negotiate the various phases of their early interactions, and they serve as a basis for diagnosis and treatment.

A 12-month-old infant's mother worried that, “He cries any time I try to leave him, even for a second. If I'm not standing right next to him when he is sitting on the floor, he cries and I have to pick him up. He's a tyrant. He's waking up four times at night and is a fussy eater. He eats for short bursts (breast-feeding) and then stops eating. I'm feeding him all the time.”

The mother was feeling cornered, controlled, manipulated, and bossed around. Her baby was like a “fearful dictator” (therapist's term). She said, “That's the perfect way to describe him.” The father was impatient with the mother; he felt that she indulged the baby too much. He was getting “fed up” because she had no time for him.

The baby was interactive and sensitive to every emotional nuance. As he came into the room, he immediately caught the clinician's eye. They exchanged smiles and motor gestures. He interacted with his parents with smiles, coos, and motor movements. Father intruded somewhat. He would roughhouse until the baby would cry, put the baby down, and then roughhouse again. Mother, in contrast, was ever so gentle, but long silences passed between her vocalizations. During her long silences, the baby would rev up, get more irritable, and start whining. He whined with his mother and cried fearfully with his father. Even before he could finish his motor gestures or vocalizations, his mother moved in and picked him up, or gave him a rattle, or spoke for him. In this way she undermined his initiative. Even while whining, however, he was interactive and contingent.

On physical examination, this baby was sensitive to loud noises and light touch on the arms, legs, abdomen, and back. He had a mild degree of low motor tone and was posturally insecure. He was not yet ready to crawl.

His constitutional and maturational patterns did not compromise his mastering the first developmental challenge of shared attention and engagement. He was an attentive, engaged baby. But at the second developmental stage, intentional communication and assertiveness, he was a passive reactor. He was not learning to initiate two-way communication, to be assertive and take charge of his interactions. His low motor tone was compromising his ability to control his motor movements. His sensory hyperreactivity was compromising his ability to regulate sensation. He was frequently overloaded by just the basic sensations of touch and sound, and he was not receiving support from his mother through the nurturing and rhythmic caretaking that would foster self-initiative.

This family required therapeutic work on a number of tasks simultaneously. The infant's special constitutional-maturational patterns were discussed. Hands-on practice helped the parents help their baby be attentive and calm. Those tasks included helping the mother be more patient, wait for the baby to finish what he started, and support his initiative (e.g., putting something in front of him while he was on his tummy, to motivate him to crawl and reach); getting the mother to put more affect into her voice and to increase the rhythm and speed of her vocalizations; and getting the father to be more gentle. The parents' own feelings about the interactions were explored: the father's tough-guy background, the mother's fear of her own assertiveness, her fear of her baby being injured, and their own associated family patterns.

Gradually, the baby began to sleep through the night, and he became more assertive and less clinging and fearful. He also became happier. He was slow to reach his motor milestones, so an occupational therapist worked with him and gave the parents advice on motor development and normalizing his sensory overreactivity. In 4 months, this infant was functioning in an age-appropriate manner with a tendency toward a cautious, but happy and assertive, approach to life's developmental challenges.

As developmental clinicians and researchers build on Piaget's findings and formulations, the developmental model will serve as a basis for understanding social and emotional development and provide a framework for clinical and educational intervention.

Implications for Psychotherapy Piaget was not an applied psychologist and did not develop the implications of his cognitive model for psychotherapeutic intervention. Nevertheless his work formed one of the foundations of the “cognitive revolution” in psychology. One aspect of this revolution was an increasing emphasis on the cognitive components of the therapeutic endeavor. In contrast to classical psychodynamic therapy, which focused primarily on drives and affects, and in contrast to behavior therapy, which focused on overt actions, cognitive approaches to therapy focused on thoughts, including automatic assumptions, beliefs, plans, and intentions.

Cognitive theory, including Piaget's, has influenced psychotherapeutic approaches in multiple ways. Some therapists have taken developmental notions from Piaget's work and developed intervention techniques. Others have developed cognitive models of treatment independent of Piaget but with heavy reliance upon the role of cognition. Others have included Piaget's concepts in a broader set of constructs to undergird new developmental approaches to psychotherapy.

Some psychotherapists applied Piagetian notions directly to child interventions. Susan Harter, for example, discussed techniques for helping young children become

aware of divergent or contradictory emotions and integrate these complex emotions within a more abstract or higher class of emotions. One of Harter's techniques is to ask the young child to draw different and conflicting feelings in one person. This technique represents an application of the concrete operation of class inclusion to the realm of the emotions. Harter's work applied Piagetian findings to the common therapeutic problem of helping children recognize, tolerate, and integrate mixed or ambivalent affects within stable object relations. As such, it drew upon both cognitive theory and psychodynamic theory.

Other psychotherapists developed treatment models which, although not directly dependent upon Piagetian psychology, emphasized core ideas quite similar to those Piaget discovered in his naturalistic observations of cognitive development. Aaron Beck, for example, developed an entire school of cognitive therapy that focuses on the role of cognitions in causing or maintaining psychopathology. Cognitive therapy has been shown to be an effective treatment for problems as diverse as depression, anxiety disorders, and substance abuse.

A core idea in cognitive therapy is that the patient can be assisted to identify the negative automatic thoughts and underlying dysfunctional attitudes or beliefs that contribute to emotional distress or addictive behavior. The cognitive component of the therapy begins with identification of automatic thoughts, so designated because they are rapid, overlearned responses that instantaneously mediate between an event and an affective reaction. The key therapeutic process after identification of the maladaptive thoughts is to help the patient view these thoughts more objectively, not take them in an unquestioning manner as veridical.

In helping patients take an objective or distanced perspective on their own thoughts, cognitive therapists are enhancing the patient's cognitive role-taking ability. In the realm of emotional health and personality development, they are contributing to what Piaget might have referred to as increased emotional intelligence on the part of the patient. What the cognitive therapist accomplishes through such techniques as Socratic questioning and asking if there are other ways to look at the same event is similar to what the talented teacher does in guiding children to more adequate, more intelligent understanding of operational tasks. The notion of equilibration is relevant in both instances. By helping the individual see that previous cognitive structures are in some ways inadequate, the therapist or teacher disturbs the old cognitive structure, and the patient or student experiences a disruption that leads to the search for more-adequate structures. The compensation for external disturbance is what Piaget termed *equilibration*. New structures can only be constructed through a process of accommodation, enabling the subject to assimilate a wider array of data, a new perspective, or more-complex information.

Since it requires “thinking about thinking,” cognitive therapy seems to require formal operational thinking, although to our knowledge this has not been empirically tested. At the very least, it requires the ability to recognize and articulate affects, to recognize and label events that give rise to affects, and to translate into a thought the mediating process that occurs rapidly between the event and the affect. Cognitive-behavioral models of psychotherapy include both cognitive techniques and more-behavioral, interactive techniques such as increasing pleasant activities and improving communication and problem-solving skills. Gregory Clarke and his colleagues have shown that cognitive-behavioral treatment is effective in reducing adolescent depression. It is not clear which components of the treatment are responsible for the overall positive effects, however.

It may be that adolescence is the earliest developmental stage in which a person could benefit from cognitive therapy. Even among adolescents and some adults, such therapy is difficult to conduct. In working with depressed, substance-abusing adolescents, John Curry and his colleagues have found that the cognitive components of the treatment are too difficult for adolescents who lack ready access to their own emotional states or the ability to label affects. Similar findings were reported by the first author regarding the cognitive aspects of psychodynamic therapy with some adults.

Such findings have contributed to the third approach to using Piagetian insights in psychotherapy, namely, integrating Piaget's findings into a broader model. The first author, for example, has articulated a developmentally based psychotherapy that takes account of earlier, presymbolic levels of functioning that precede the ability to recognize, label, and articulate affects and their mediating cognitions.

Developmentally Based Psychotherapy Developmentally based psychotherapy, developed by the first author, integrates cognitive, affective, drive, and relationship-based approaches with new understanding of the stages of human development. Different therapies look at different aspects of the proverbial elephant, whether from a psychodynamic, object relations, self-psychology, behavioral, or cognitive-behavioral point of view. A comprehensive, cohesive developmental framework integrates elements from these approaches with a broader understanding of the developmental processes essential for emotional or mental health. It formulates series of principles that an understanding of human development says are prerequisite for emotional growth.

Developmentally based psychotherapy constructs its therapeutic strategies from these principles of human development and growth. The clinician first determines the level of the patient's ego or personality development and the presence or absence of deficits or constrictions. For example, can the person regulate activity and sensations, relate to others, read nonverbal affective symbols, represent experience, build bridges between representations, integrate emotional polarities, abstract feelings, and reflect on internal wishes and feelings?

After determining the developmental level, the clinician looks for constitutional and maturational contributions and difficulties with sensory processing, modulation, or motor planning. The clinician looks for interactive and family contributions. Each of these is explored in the present, the past, and the anticipated future. The patient's fantasies, sense of self and others, and conflicts are understood in the context of all these influences. These are how patients make sense of their ego structure, physical makeup, family patterns, and interactions with others. Developmentally oriented therapists do not permit themselves the luxury of overfocusing on one set of variables such as inner fantasies, family dynamics, biological proclivities, or prior experience. Similarly, the formulated therapeutic strategy cannot deal only with one or two factors. It must deal with all critical factors that influence the developmental process. As collaborators in the construction of experience, therapists use their understanding of the patient's development to help the patient construct interactions that will provide growth and overcome difficulties.

Often it is assumed that critical aspects of development occur through the maturation of the nervous system along with routine, expectable experiences. It is also assumed that from these routine, expectable maturational sequences and experiences, certain psychological structures having to do with the ability to regulate, engage, interact, represent (symbolize) experience, and reflect and compare experiences, are present in most people. With these capacities in place, it is believed that the therapeutic process can focus on conflicts and anxieties and selected maladaptive behaviors or thoughts. The authors have observed, however, that only a small percentage of individuals have these core capacities. For most, such capacities must be learned as part of the therapeutic process.

The developmental perspective shows how one learns these capacities during development. It suggests strategies that can be used in the psychotherapeutic process so that adults and children who have not achieved these capacities can learn them. From a developmental point of view, the integral parts of the therapeutic process include learning how to regulate experience; to engage more fully and deeply in relationships; to read and respond to boundary-defining behaviors and affects; to perceive, comprehend, and respond to complex self- and object-defining affects, behaviors, and interactive patterns; to represent experience; to differentiate represented experience; and to form higher-level differentiations, including the capacity to engage in the ever-changing opportunities, tasks, and challenges during the course of life (e.g., adulthood and aging) and, throughout, to observe and reflect on one's own and others' experiences. Mastering these core developmental processes makes dealing with conflicts, anxieties, maladaptive behaviors, and thoughts possible.

These processes are the foundation of the ego, and more broadly, the personality. Their presence constitutes emotional health and their absence, emotional disorder. The developmental approach describes how to harness these core processes and so assist the patients in mobilizing their own growth.

SUGGESTED CROSS-REFERENCES

Perception and cognition are discussed in Section 3.1, learning theory in Section 3.3, aggression in Section 27.3, biology of memory in Section 3.4, and brain models of mind in Section 3.5. Chapter 39 addresses attention-deficit disorders. Chapter 35, Chapter 36 and Chapter 37 focuses on learning disorders, motor skills disorder, and communication disorders. Feeding and eating disorders in children are the subject of Chapter 41, and mental retardation is covered in Chapter 34.

SECTION REFERENCES

Greenspan SI: *The Clinical Interview of the Child*. McGraw Hill, New York, 1981.

Greenspan SI: *The Development of the Ego: Implications for Personality Theory, Psychopathology, and the Psychotherapeutic Process*. International Universities Press, Madison, CT, 1989.

*Greenspan SI: *Infancy and Early Childhood: The Practice of Clinical Assessment and Intervention with Emotional and Developmental Challenges*. International Universities Press, Madison, CT, 1992.

Greenspan SI: *Developmentally Based Psychotherapy*. International Universities Press, Madison, CT, 1997.

Greenspan SI: *The Growth of the Mind and the Endangered Origins of Intelligence*. Addison Wesley Longman, Reading, MA, 1997.

Hamlett KW, Pellegrini DS, Conners CK: An investigation of executive processes in the problem-solving of attention deficit-hyperactive children. J Pediatr Psychol *12:*227, 1987.

Harter S: A cognitive-developmental approach to children's expression of conflicting feelings and a technique to facilitate such expression in play therapy. J Consult Clin Psychol *45:*417, 1977.

Inhelder B, Piaget J: *The Growth of Logical Thinking from Childhood to Adolescence*. Basic Books, New York, 1958.

Kant I: *The Critique of Pure Reason*. MacMillan, London, 1963.

Lochman JE, Dodge K: Social cognitive processes of severely violent, moderately aggressive, and nonaggressive boys. J Consult Clin Psychol. *62:*366, 1994.

*Nicolopoulou A: Play, cognitive development, and the social world: Piaget, Vygotsky, and beyond. In *Lev Vygotsky: Critical Assessments: Thought and Language*, vol 2, P Lloyd, C Ferynhough, editors. Routledge, New York, 1999.

Piaget J: *Play, Dreams and Imitation in Childhood*. Norton, New York, 1951.

Piaget J: The stages of the intellectual development of the child. Bull Menninger Clin *26:*120, 1962.

*Piaget J: *The Early Growth of Logic in the Child*. Norton, New York, 1969.

Piaget J: *Structuralism*. Basic Books, New York, 1970.

Piaget J: Piaget's theory. In *Manual of Child Psychology*, P Mussen, editor. Wiley, New York, 1983.

*Piaget J, Inhelder B: *The Psychology of the Child*. Basic Books, New York, 1969.

Piaget J, Inhelder B: *The Origin of the Idea of Chance in Children*. Norton, New York, 1975.

Pinard A, Laurendeau M: Stage in Piaget's cognitive-developmental theory: Exegesis of a concept. In *Studies in Cognitive Development: Essays in Honor of Piaget*, D Elkind, JH Flavell, editors. Oxford University Press, New York, 1969.

Sameroff A, Seifer R, Barocas R, Zax M, Greenspan SI: IQ scores of 4-year-old children: Social-environmental risk factors. Pediatrics *29:*343, 1986.

Sternberg RJ, Berg C, editors: *Intellectual Development*. Cambridge University Press, Cambridge, 1992.

*Tudge J, Rogoff B: Peer influences on cognitive development: Piagetian and Vygotskian perspectives. In *Lev Vygotsky: Critical Assessments: The Zone of Proximal Development*, vol 3, P Lloyd, C Fernyhough, editors. Routledge, New York, 1999.

3.3 LEARNING THEORY

W. STEWART AGRAS M.D., F.R.C.P.(C) AND G. TERENCE WILSON PH.D.

From birth, learning plays a central role in the development of human behavior, including voluntary and involuntary motor behaviors, thinking, and emotion. The effects of the environment on the development or maintenance of disordered behaviors are, with the exception of effects such as injury, malnutrition, or infection, translated through learning; hence, learning plays an important role in the development and maintenance of psychopathology. Conversely, the basis of psychological change by means of psychotherapy depends on learning new and more adaptive behaviors. Both the theoretical basis for psychotherapy and its practical use depends on a grasp of the principles of learning.

In Russia at around the end of the nineteenth century, Ivan Pavlov (Fig. 3.3-1), a Nobel Laureate in physiology, established the foundations of classical conditioning. At roughly the same time in the United States, pioneering research on animal learning by Edwin Thorndike showed the influence of consequences (rewarding and punishing events) on behavior. Beginning in the late 1930s this process of instrumental learning was elaborated upon by B. F. Skinner in his research on operant conditioning. Research on conditioning and learning principles, conducted largely in the animal laboratory, became a dominant part of experimental psychology in the United States after World War II. In the tradition of Pavlov and Skinner, workers in this area were committed to the scientific analysis of behavior using the laboratory rat and the pigeon as their subjects.

FIGURE 3.3-1 Ivan Pavlov. (Courtesy of National Library of Medicine.)

During the first half of the twentieth century, isolated attempts were made to apply conditioning principles to a variety of problem behaviors. John B. Watson and Rosalie Rayner experimentally induced a phobic reaction in the famous case of Albert by using a simple classical conditioning procedure, and Mary Cover Jones antedated later investigations by describing the use of several different procedures for overcoming children's fears. Researchers Mowrer and Mowrer reported the successful treatment of enuresis by direct conditioning procedures.

These early applications had no impact on the practice of psychotherapy, in part because conditioning principles, which had been demonstrated with animals, were rejected as too simplistic and irrelevant to the treatment of complex human problems. Nevertheless, some attempts were made to integrate conditioning principles with psychodynamic theories of abnormal behavior, but these eclectic formulations had little effect and only obscured crucial differences between the respective behavioral and psychodynamic approaches. John Dollard and Neal Miller, for example, translated psychodynamic theory but with little consequence for any clinical innovation, since they were merely reinterpreting psychotherapy rather than advocating different concepts and procedures.

The emergence of behavior therapy as an alternative system of assessment and treatment to psychoanalysis in the late 1950s marked the first systematic extension of conditioning principles to clinical practice. Behavior therapy was originally defined as the application of “modern learning theory” to the treatment of psychiatric disorders. The phrase “modern learning theory” referred to the principles and procedures of classical and operant conditioning. In the 1950s and early 1960s, learning theory (i.e., classical and operant conditioning) was the obvious choice as a body of knowledge within experimental psychology on which to build an applied clinical science. In the 1970s experimental research on cognition (e.g., information processing) supplanted classical and operant conditioning as the major focus of theory development and research in experimental psychology. Throughout the 1970s behavior therapists increasingly incorporated cognitive processes and procedures within their clinical practice, and social-cognitive learning theory emerged as the most influential conceptual framework guiding the clinical application of learning principles and procedures. Social-cognitive theory embraces classical as well as operant conditioning procedures. However, in this approach the influence of environmental events on behavior is largely determined by cognitive processes that govern how environmental influences are perceived and how the individual interprets them. According to this view psychological functioning involves a reciprocal interaction among three interlocking sets of influences: behavior, cognitive processes, and environmental factors. In the opinion of Bandura personal and environmental factors do not function as independent determinants; rather, they determine each other. Nor can persons be considered causes independent of their behavior. It is largely through their actions that individuals produce the environmental conditions that affect their behavior in a reciprocal fashion. The experiences generated by behavior also partly determine what individuals think, expect, and can do, which in turn affects their subsequent behavior.

In social-cognitive learning theory, the person is the agent of change. The theory emphasizes the human capacity for self-directed behavior change. Strongly influenced by the social-cognitive learning model, the clinical practice of behavior therapy has increasingly included cognitive methods, especially those described by Aaron Beck. A primary focus of cognitive and behavioral techniques is to change the cognitive processes that are viewed as essential to therapeutic success. This theory assumes that it is not so much experience itself, but rather the person's interpretation of that experience, that produces psychological disturbance, a view also reflected in the work of Albert Ellis. Both cognitive and behavioral methods are used to modify misperceptions and misinterpretations of important life events; this approach is called *cognitive behavior therapy*.

It is now widely accepted that the principles of classical and operant conditioning have been extremely successful in stimulating a wealth of experimental and clinical research and innovative therapeutic techniques. A fundamental contribution of learning theory was the conceptual and methodological emphasis that the study and application of conditioning principles brought to clinical research and practice. The detailed specification of therapeutic techniques, the focus on behavior per se in assessment, treatment, and evaluation of therapy outcome, and the advances in measurement and methodology were all directly associated with the methodological behaviorism that characterized the conditioning approach. Moreover, because of the experimental, hypothesis-testing emphasis deriving from conditioning and social-cognitive theories, cognitive behavioral therapies have become the mainstream of psychotherapy research. These developments have helped to narrow the gap between the laboratory and the clinic.

Every psychotherapy session knowingly or unknowingly contains some of the procedures derived from learning theory. The most effective psychotherapies are targeted to alter a specific behavior or cluster of associated behaviors, are based on a rational model of factors maintaining the disorder, use a well-designed sequence of behavior change strategies sensitive to the individual problems presented by a particular patient, and are time-limited. Most of these psychotherapies

now appear in detailed manuals and have been rigorously tested in controlled outcome studies.

BIOLOGY OF LEARNING

Learning leads to neurochemical changes in the central nervous system. Research with simple organisms has revealed that the learning of avoidance behavior alters the chemical structure of cells in the nervous system; when the avoidance is unlearned, the chemical changes are reversed. Thus, the foundation for understanding the neurochemistry of learning has been laid, and a reciprocal interaction is evident between biological processes in the central nervous system and behavior changes resulting from environmental influences. Research conducted by Eric Kandel and his colleagues at Columbia University with *Aplysia californica* has been particularly well covered in the psychiatric literature. The aplysia, a sea mollusk, is a useful animal to study because of the simplicity of its nervous system, which contains about 20,000 neurons, many of which are large and readily identifiable. The specific behavior studied is a defensive reflex involving withdrawal of the snail's siphon when the animal is tactually stimulated. When the mollusk is touched repeatedly, it learns not to withdraw its siphon and gill, a process known as *habituation*. If the mollusk then receives a strong stimulus, such as an electric shock, it becomes sensitized, such that even a previously subthreshold tactile stimulation causes the animal to withdraw its gill and siphon. Furthermore, the snail can be classically conditioned so that it withdraws its siphon and gill to a conditioned stimulus. Habituation, sensitization, and classical conditioning of the reflex in the snail can be considered forms of learning and memory. In learning, a short-term stimulus has to be translated into long-lasting changes that involve a series of biochemical changes that are functionally interlinked and operate in overlapping time ranges.

Many but not all of the neuronal anatomical and chemical bases for the learning processes have been worked out in this animal model. Sensory neurons receiving tactile information form excitatory synapses with the gill and siphon motor neurons that cause the withdrawal activity. Habituation, sensitization, and classical conditioning all involve neurochemical changes in the sensory neuron, resulting in alterations in the amount of excitatory neurotransmitter released. The neurochemical basis of habituation is that, after repeated stimulation of the sensory neuron (e.g., repeated tactile stimulation), less calcium than usual enters the presynaptic nerve terminal, resulting in less neurotransmitter being released and, thus, less activity by the motor neurons. Sensitization requires the presence of additional neurons, called *facilitator interneurons*, that synapse onto the sensory neurons. The sensitizing stimulus, such as an electric shock, causes the facilitator interneuron to release serotonin that binds to serotonin receptors on the sensory neuron. Activation of the serotonin receptors activates adenylate cyclase, producing cyclic adenosine monophosphate (cAMP), thereby activating a cAMP-dependent protein kinase, which is believed to phosphorylate an S-type potassium channel. Phosphorylation of the potassium channel results in increased calcium influx during the action potential and increased neurotransmitter release. Although classical conditioning also results in an increased amount of neurotransmitter released by the sensory neuron, the neurochemical basis is less well understood at this time but may involve additional protein kinases.

Because psychotherapy is best viewed as a learning process, and learning produces changes in neuronal architecture, the behavior changes associated with therapy should produce anatomical changes in the central nervous system. As Eric Kandel notes, such changes should be detectable by imaging methods. In an interesting study of patients with obsessive-compulsive disorder, positron emission tomography was used to investigate changes in cerebral metabolic rates for glucose before and after treatment. The effects of two treatments were compared: fluoxetine hydrochloride, and behavior therapy consisting of exposure and response prevention. Although the study involved nine patients per group, glucose metabolic rates in the right head of the caudate nucleus changed when the obsessive-compulsive disorder was successfully treated with either fluoxetine or behavior therapy; these changes did not occur in patients who did not respond to treatment, which suggests that the changes in brain metabolism consequent upon learning-based changes in behavior may be similar to those induced by pharmacotherapy. These results underline the reciprocal interaction between learning and the central nervous system.

In the following sections classical conditioning, operant conditioning, and social learning theory are reviewed. Some common terms used in learning theory are listed and defined in Table 3.3-1.

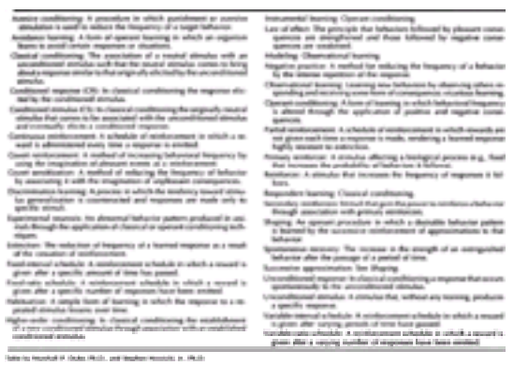

Table 3.3-1 Common Terms Used in Learning Theory

CLASSICAL CONDITIONING

History The idea that learning takes place when two events occur closely together in time has a long history, stemming from association theory developed by the British school of philosophical empiricism. It was the Russian physiologist Ivan Petrovich Pavlov and his coworkers who over many years documented the parameters of this form of learning in carefully conceived experiments. Traditional accounts of classical conditioning state that learning occurs when an initially neutral stimulus, the *conditioned stimulus*, is paired with a stimulus that naturally elicits a response, the *unconditioned stimulus*. The response elicited by the unconditioned stimulus is the *unconditioned response*. After repeated and contiguous pairing of the two stimuli, the conditioned stimulus elicits the unconditioned response, which is then called the *conditioned response*.

Pavlov's work was enthusiastically espoused by American psychologists, such as John B. Watson, who demonstrated that classical conditioning can give rise to phobia-like behavior. The subject of the experiment was Albert B., who was 11 months old. Watson demonstrated that a few pairings of a loud noise (unconditioned stimulus) with the sight of a white rat (conditioned stimulus) led Albert to avoid not only the rat, which had not caused fear before, but also similar objects, such as cotton wool and sealskin, an example of stimulus generalization in which stimuli similar to the original conditioned stimulus may elicit the conditioned response, although usually with a weakened response. According to that view, which is still widely held by psychologists and psychiatrists, classical conditioning was a rather simple, limited, and automatic form of learning. Current thinking, however, differs substantially from that traditional account.

Current Views Temporal contiguity between two stimuli is neither necessary nor sufficient for classical conditioning to take place. Two examples illustrate the point. In the first example a rat is exposed to five pairings of a tone and an electric shock in one situation. In another situation the tone is presented 10 additional times in the absence of shock. The contiguity of tone and shock is the same in both situations, but classical conditioning occurs only in the first situation because only in that situation does the tone predict or provide information about the unconditioned stimulus. In the other situation the unconditioned stimulus is equally likely whether or not the tone is sounded. Contiguity is not necessary for conditioning to occur. If the presentation of the tone and the shock is arranged so that shocks never occur in the presence of the tone, the tone comes to predict the nonoccurrence of the shock, a phenomenon called *conditioned inhibition*.

In the second example two groups of animals are exposed to a compound stimulus (tone plus light) that signals a shock. One group has a history of learning in which the light predicts a shock; the other group does not. Both groups have the same contiguous exposure to the tone-light compound stimulus, but for the group with pretraining the tone is redundant. When both groups are tested for their conditioning to the tone, the group with pretraining with the light stimulus shows significantly worse conditioning than the other group because despite equivalent contiguity, the tone conveys different information to the two groups.

Classical conditioning is not necessarily a slow process dependent on repeated pairings of stimuli. Learning is often rapid and efficient. As Robert Rescorla noted:

Pavlovian conditioning is not a process by which the organism willy-nilly forms associations between any two stimuli that happen to co-occur. Rather the organism is better seen as an information seeker using logical and perceptual relations among events, along with its own preconceptions, to form a sophisticated representation of the world. An analogy between animals showing Pavlovian conditioning and scientists identifying the cause of a phenomenon is useful. If one thinks of Pavlovian conditioning as developing between a conditioned stimulus and a unconditioned stimulus under just those circumstances that would lead a scientist to conclude that

the conditioned stimulus causes the unconditioned stimulus, one has a surprisingly successful heuristic for remembering the facts of what it takes to produce Pavlovian associative learning.

Classical conditioning has been viewed as either (1) learning that, in humans at least, requires conscious processing and awareness of the relations between events, or (2) an automatic process that occurs similarly in the mollusk and the human being, requiring little or no conscious processing or awareness. A good deal of evidence shows that awareness of a relation between events is often necessary for conditioning to occur and greatly facilitates learning. It seems reasonable to conclude that classical conditioning in humans occurs through hierarchically organized neural systems. The operation of some of these systems may predispose to automatic processing while others do not. On the other hand, evidence indicates that classical conditioning can occur in the absence of intention to learn, without awareness, and even if it is resisted. For example, both animals and humans acquire conditioned aversions to specific smells and tastes (the conditioned stimulus) if they are associated with drug-induced or illness-induced nausea (the unconditioned stimulus-unconditioned response). People develop highly specific aversions to food that they may have eaten at the onset of seasickness or of a gastrointestinal illness. The food becomes a conditioned stimulus for them, despite their knowledge that it did not cause their nausea. In those instances people seem to be biologically prepared to have some conditioned responses and not others. For example, nausea is readily conditioned to smell and taste but not to sight and sound. Some phobic reactions in humans may provide other examples of prepared learning of that sort (e.g., fears of animals and heights). The hypothesis is that people are biologically predisposed to learn certain fears as a result of their evolutionary past, when it was adaptive to do so. That explains the fact that people have phobias only to selected situations and objects, which presumably were once associated with threats to survival.

Classical conditioning may affect the responsiveness of the immune system. Animals were conditioned by pairing saccharinflavored water with injections of cyclophosphamide, an immunosuppressive drug. After a single trial, animals showed both immunosuppression and aversion to the taste of saccharin, phenomena not shown by the control group that received saccharin-flavored water paired with a placebo injection. It is well known that humans undergoing immunosuppressive therapy develop nausea that comes to be elicited by hospital cues; indeed, sometimes even the thought of the hospital elicits nausea. In a recent study women undergoing treatment with cytotoxic drugs for the treatment of ovarian cancer demonstrated nausea on return to the hospital combined with decreased immune function. This appears to replicate the work with animals and suggests that similar conditioning of immune function is found in humans. Clearly, the person is unaware of such conditioning.

Clinical Applications Classical conditioning has influenced the way in which clinical disorders have been conceptualized and has generated methods for their treatment. It has, for example, played a prominent role in the analysis and treatment of anxiety disorders. Exposure to a traumatic event is a necessary but insufficient condition for the development of posttraumatic stress disorder. Classical conditioning explains how stimuli associated with a severe trauma (the unconditioned stimulus) come to elicit stress responses that were part of the original trauma. As with phobic disorders, not everyone who experiences a traumatic event displays the stress response; predisposing genetic factors and adverse early learning experiences are determinants of posttraumatic stress disorder.

Another area of application has been in the treatment of substance abuse. Cues associated with repeated alcohol and drug use come to elicit conditioned responses. Some of the conditioned responses are opposite in nature to the unconditioned effects of the substance. They are known as *classically conditioned compensatory reactions*, which are believed to reflect homeostatic mechanisms. This conditioning process contributes to the development of behavioral tolerance to alcohol and drugs. In alcohol- or drug-dependent persons the conditioned responses are believed to be subjectively experienced as craving or anticipatory withdrawal reactions. An alternative view is that the conditioned stimulus triggers craving because it comes to signal the positively rewarding consequences of alcohol or drug use. In either case it follows that cues associated with substance abuse need to be addressed in treatment. Cue exposure treatment for alcoholic persons and drug addicts is based on the principle of *extinction*—the procedure of presenting the conditioned stimulus in the absence of the unconditioned stimulus. Doing so results in the elimination of the conditioned response when the conditioned stimulus no longer predicts its occurrence. For example, patients with alcohol-dependence are presented with alcohol-related cues (e.g., the sight and the smell of alcohol), which reliably elicit craving, without being allowed to drink alcohol (the unconditioned stimulus). Since negative emotional states can function as conditioned stimulus, cue-exposure treatment also involves the induction of relevant mood states.

Classical conditioning is no longer necessarily linked to the philosophy of behaviorism, with which it was once associated. Its concepts and methods lend themselves to current perspectives on human behavior, and they continue to play an important role in contemporary behavior therapy, which itself has become more eclectric and theoretically broader than it was during its early origins in behaviorism.

TWO-FACTOR CONDITIONING THEORY

O. H. Mowrer proposed a two-factor theory of avoidance behavior that has had a major influence on models of psychopathology and on treatment. The theory made two basic assumptions: (1) anxiety is an acquired drive established by classical conditioning in which a neutral stimulus comes to signal an aversive or traumatic event; and (2) the classically conditioned anxiety motivates escape and avoidance behavior, which is reinforced by the reduction of the underlying anxiety drive. This theory has been fruitfully applied to phobic avoidance and obsessive-compulsive behavior. The anxiety about dirt and contamination that compulsive hand washers experience is assumed to be a classically conditioned response that drives them to wash their hands repeatedly after touching anything they believe is contaminated. Hand washing removes the dirt and reduces the anxiety; the anxiety reduction then reinforces and maintains the hand-washing behavior.

Clinical Applications More important than the models of psychopathology it generated were the treatment methods that two-factor theory spawned. The techniques of exposure and response prevention followed directly from the theory. In *exposure treatment* for phobic disorders, patients are gradually and systematically exposed to the situations they fear and avoid until their phobic reactions are extinguished. In the technique of *systematic desensitization* pioneered by Joseph Wolpe, one of the founding fathers of behavior therapy, the exposure is to imagined representations of the feared stimuli while the patient is in a state of relaxation. That technique was thought to be an example of *reciprocal inhibition*, in which one response (anxiety) is inhibited by an opposite response (relaxation). However, it is more effective to conduct exposure to real-life situations (e.g., guiding an agoraphobic patient through shopping malls, supermarkets, crowded buses, and other feared situations) than to imagined scenes.

Problems Despite its historical significance in the field, two-factor conditioning theory has been criticized on several counts. One problem is the difficulty it encounters in explaining why phobic reactions are resistant to extinction. Classically conditioned fear reactions in laboratory animals extinguish rapidly when the US is omitted. But phobic fears persist, even though phobic persons frequently come into contact with the situations they fear (exposure to the conditioned stimulus) without any aversive event (the unconditioned stimulus)—by definition, the conditions for the extinction of the fear and avoidance behavior. A second problem with the theory is evidence showing that avoidance behavior is not causally mediated by fear or anxiety, whether the anxiety is defined as autonomic nervous system arousal or subjective self-report. Avoidance behavior can be eliminated without inhibiting the accompanying anxiety. Anxiety reduction may precede, accompany, or even follow changes in avoidance behavior—an impossibility if anxiety is said to cause the avoidance behavior. A more accurate conclusion is that both anxiety and avoidance behavior are correlated coeffects of a still undetermined central mediating state. Third, whereas the effectiveness of exposure treatments has been well documented, two-factor theory does not accurately account for its therapeutic mechanisms. The amount of exposure should be closely correlated with the degree of change in phobic avoidance behavior, yet it is a poor predictor of such change. An alternative explanation of exposure treatment is discussed next.

OPERANT CONDITIONING

The notion that learning occurs as a consequence of action was espoused in the pioneering work of Edward Lee Thorndike, whose learning theory dominated the field of psychology in the United States for the first half of the twentieth century. A typical experiment devised by Thorndike consisted of placing a hungry cat in a cage with some form of latching device that, when correctly manipulated, allowed the cat access to a second cage for a bite of food. Thorndike noted that the cat became efficient at opening the lock, a sequence of events termed *trial-and-error learning*, and he hypothesized that the appropriate behaviors were strengthened by the cat's experiences of success and failure. Following up on Thorndike's work, Skinner and his colleagues made the effects of environment on behavior a central aspect of learning.

The principles and procedures of operant conditioning are the product of Skinner's philosophy of *radical behaviorism*, according to which overt behavior is the only acceptable target of scientific investigation. Skinner argued that subjective experience (private events) should be included in the experimental analysis of behavior, but their role has always been restricted. Thoughts and feelings are epiphenomena in operant conditioning; they cannot exert a causal influence on behavior. Strictly speaking, aside from biological determinants (which have always been minimized), it is assumed that human behavior is exclusively a function of environmental events that are ultimately beyond personal control. Another hallmark of operant conditioning has been its emphasis on the study of the individual organism. The repeated measurement of the behavior of a person under controlled conditions is the methodological contribution of operant conditioning. Skinner rejected statistical comparisons between groups of subjects, claiming that group averages obscure what is important—namely, the behavior of individual subjects. The application of the

principles and the procedures of operant conditioning to human problems is known as *applied behavior analysis*.

Positive Reinforcement The best known learning principle contributed by operant conditioning is the principle of *positive reinforcement*—the process by which certain consequences of behavior raise the probability that the behavior will occur again. On the whole, positive reinforcers are viewed as pleasant (e.g., food, attention, praise, money). Reinforcers that affect biological processes, such as food, may be defined as *primary reinforcers*. However, events viewed as aversive by some people may be reinforcing for others. For example, the behavior of some children is reinforced by scolding, which, after all, is a form of attention. Many drugs including opioids, barbiturates, and such stimulants as amphetamine and cocaine, appear to be positive reinforcers. Animals and humans self-administer the substances, reliably discriminating between the active drug and a placebo. Complex patterns of behavior can be shaped in animals by using drugs as reinforcers.

Traditional textbook descriptions state that reinforcement of a response must be immediate. However, as with classical conditioning, temporal contiguity is not necessary for learning to occur in operant conditioning. Recent research shows that the behavior of even a simple organism, such as a laboratory rat, can be controlled by the aggregate consequences of a series of reinforcements of multiple responses over time. Revision of the requirement that reinforcement be immediate if it is to control behavior greatly extends the explanatory power of operant conditioning, but the mechanism by which the organism is able to integrate reinforcing consequences over time is never specified. Critics of operant conditioning point out that the delayed effect of reinforcement contingencies must be mediated by cognitive processes.

Negative Reinforcement Reinforcement increases the probability of behavior. *Negative reinforcement* is the process by which behavior leading to the removal of an aversive event strengthens that behavior. Negative reinforcers tend to be aversive events; for example, avoidance behavior, such as phobic reactions and compulsive rituals, is negatively reinforced because it forestalls actual or perceived aversive outcomes. Research shows that when patients with anorexia nervosa are placed in a restricted hospital environment to facilitate the use of positive reinforcement for eating and gaining weight, they work (eat and gain weight) in order to get out of the aversive environment, adding negative reinforcement effects to the positive reinforcement effects.

Punishment Punishment is the presentation of an aversive stimulus contingent on the occurrence of a particular response. The removal of a positive consequence contingent on behavior, known as *time out* from reinforcement, can also be viewed as punishment. The procedure is commonly used as a means of disciplining children with behavior problems; for example, sending a child to his or her room consequent upon misbehavior. It is necessary to distinguish punishment from negative reinforcement. Punishment decreases the probability that the behavior will occur, whereas negative reinforcement increases the probability. One must also distinguish between the usual use of the term *punishment* and the technical use of the term as meant here. In the punishment paradigm the punishing event is always delivered contingent on performance and demonstrably reduces the frequency of the behavior being punished, which is considerably different from the use of the term to denote imprisonment, for example, because the prison sentence follows long after the crime and may not affect future criminal behavior. Figure 3.3-2 summarizes the major principles of operant conditioning and the effects they have on behavior.

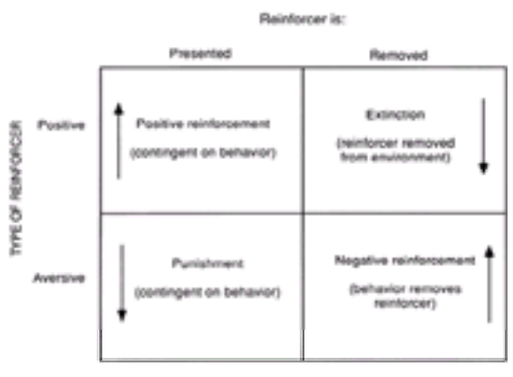

FIGURE 3.3-2 The principal procedures of operant conditioning depending on whether a positive or negative reinforcer is applied or removed after the behavior is performed. The arrows show the direction of the behavior change.

Reciprocal Influences Since much human behavior occurs within an interpersonal context, reciprocal influences occur; an example is afforded by the study of predelinquent behavior. Family studies suggest that predelinquent behavior patterns are set in motion by the excessive and inconsistent use of punishment on the part of parents. A mother may severely scold her small son, who in response may whine or have a temper tantrum. If the mother then responds by talking to the child to calm him down, the child stops whining. Thus, the child's whining punishes the mother's scolding and makes her less likely to scold in the future. The mother's attention to the child's whining reinforces that unpleasant behavior on the part of the child. Such a behavior pattern, when well established in the child, is viewed as unpleasant and aggressive by others and increases the likelihood that the child will be rejected by parents, peers, and teachers, thus initiating a complex series of events, such as poor school performance and joining a deviant peer group, which predisposes to delinquent behavior.

Clinical Applications Operant conditioning procedures have been applied to a wide range of problems in all age groups in psychiatry, education, rehabilitation, and medicine. In general, the procedures are most commonly used today to change the behavior of young children, persons with mental retardation, and institutionalized populations, such as chronically mentally ill patients. Behavior therapists in clinical practice, particularly with adult outpatient disorders, rarely describe themselves as applied behavior analysts, and they draw on broad theoretical perspectives. Reinforcement procedures may be used alone as applications of operant conditioning, or such procedures may be combined with techniques, such as extinction and punishment. However, reinforcement occurs in any form of psychotherapy because the therapist differentially attends to the verbal behavior of the patient. Moreover, most behavioral therapists today combine the use of reinforcement procedures with many other procedures derived from learning theory.

Positive Reinforcement Much is known about various *schedules of reinforcement*, defined as the patterns or frequencies with which a reinforcer is delivered as a consequence of behavior. The most frequently used schedules are listed in Table 3.3-2. One of the most frequently used schedules in clinical practice is *partial reinforcement*, in which reinforcement only occasionally results from a particular behavior. Such a pattern of reinforcement maintains the behavior at full strength. Moreover, partially reinforced behavior may be particularly resistant to extinction. Since many deviant behaviors provoke attention from others, they are maintained by the social environment. Observational and experimental work, for example, has shown that hospital staff members tend to reinforce their patients' abnormal behaviors by attending to them. When the staff members learn to stop giving such attention and to attend more frequently to adaptive behaviors, patient behavior improves. Similar findings have been made in school. Teacher attention reinforces disruptive behavior in the classroom; when such attention is withdrawn, the disruptive behavior decreases.

Fixed-ratio (FR) schedule
Reinforcement occurs after a fixed number of responses, (e.g., every 10 responses—10:1 ratio; 10 bar presses release a food pellet; workers are paid for every 10 items they make). Rapid rate of response to obtain greatest number of rewards. Animal knows that the next reinforcement depends on a certain number of responses being made.
Variable-ratio (VR) schedule
Reinforcement occurs at random intervals, (e.g., after the third, sixth, then second response, and so on). This generates fairly constant rate of response because of probability of reinforcement at any given time remains relatively stable.
Fixed-interval (FI) schedule
Reinforcement occurs at regular intervals (e.g., every 10 minutes or every third hour). Animal keeps track of time. Rate of responding drops to near 0 after reinforcement and then increases at about expected time of reward. This is known as scalloping and is seen in humans checking their mailboxes.
Variable-interval (VI) schedule
Reinforcement occurs at random intervals similar to variable ratio resulting in consistent responding. Response rate does not change between reinforcement. Animal responds at steady rate in order to get reward when it is available; common in trout fishermen, use of slot machines.

Table 3.3-2 Examples of Reinforcement Schedules in Operant Conditioning

The most used reinforcement procedure is a *shaping paradigm*, in which a behavior is changed in form by reinforcing components of the final behavior sequentially. For example, in teaching a mute patient with schizophoenia to talk, the first behavior to be reinforced may be simply looking at the therapist, followed by any mouthing movement, followed by any vocalization (perhaps in imitation of the therapist), and, finally, simple words and sentences. A *continuous reinforcement schedule* may first be used; in it reinforcement is delivered for every appropriate response. This schedule may be followed by partial reinforcement; each component behavior is first developed with continuous reinforcement and then strengthened with partial reinforcement. If speech is reinforced only in the presence of one therapist, the patient may remain mute with others, which is an example of *discriminative learning*. Similar behavior is seen in everyday life when a motorist stops at a red light and proceeds when the light changes to green—behaviors that are highly reinforced in this society. To overcome discriminative learning the therapist first establishes the beginnings of speech and then has several therapists reinforce speech to ensure generalization of the new behavior. When speech is fully developed, artificial reinforcement can be phased out because speaking should be more reinforcing than being mute. This is also an example of *chaining* of behaviors and reinforcement, because all the initial behavioral sequences are necessary for the final behavior of talking, and a complex sequence of behaviors is gradually built up and reinforced. Many problem behaviors seen in humans have been developed in animals by using various schedules of reinforcement. Thus, head banging, a behavior seen frequently in retarded and autistic children, has been developed in monkeys with the use of reinforcement such that the monkeys actually injure themselves to obtain reinforcement. Although such experiments do not prove that similar behaviors seen in humans are learned, they do call attention to the powerful effect of reinforcement in developing deviant behavior and to the fact that many behaviors are developed and maintained in this way.

Often given in the form of attention and praise contingent on certain behaviors, reinforcement is a basic ingredient of most therapies. Skilled therapists of most persuasions use contingent verbal reinforcement, as has been shown even in nondirective psychotherapy, so that certain therapeutic themes are strengthened. Other methods used in reinforcement paradigms include tokens exchangeable for goods or activities that cannot be bought or engaged in otherwise. What is reinforcing for one person may not be reinforcing for another, therefore, when reinforcement is used, the clinician must observe and measure the behavior being reinforced to make sure that it is being strengthened. The data from a clinical example of the use of token reinforcement to increase social communication is shown in Figure 3.3-3.

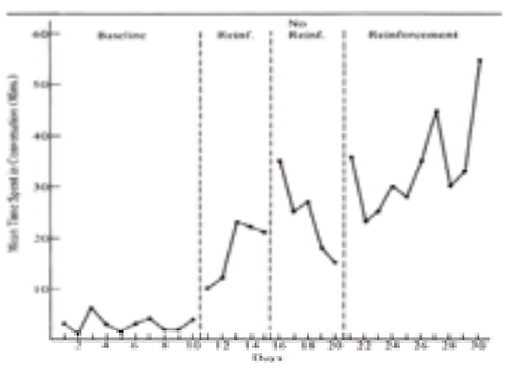

FIGURE 3.3-3 The time spent by a withdrawn patient conversing with nurses during sequential experimental phases: baseline (measurement only), reinforcement delivered in the form of tokens, after reinforcement was withdrawn, and when reinforcement was reinstated. The effect of positive reinforcement is demonstrated by an increase in conversational time during the two phases in which reinforcement was delivered, and by a decrease in conversation (extinction) when reinforcement was omitted.

The patient was a 21-year-old man who was extremely withdrawn. He spent most of his time in his hospital room, rarely approaching others or initiating conversation. Skilled psychiatric nursing care had not altered his behavior. As a first step in increasing his conversational ability, reinforcement for approaching nurses was instituted. The patient was first carefully observed, and it was noted that he enjoyed listening to the radio and watching television. He was told that for every 2 minutes during which he talked with the nurses in three daily sessions he would earn a token that could be exchanged for 3 minutes of listening to the radio or watching television, and that talking was the only way in which he would earn the right to engage in those activities. The nurses, in turn, were instructed not to approach him during the sessions but to engage in conversation only if he initiated and maintained it, thus reinforcing a chain of behaviors: approaching nurses, initiating conversation, and maintaining conversational behavior. The nurses also timed the number of minutes of conversation, using a stopwatch.

As can be seen in Figure 3.3-3, the patient engaged in little conversation with the nurses during the baseline measurement period. When the token system was introduced, he began to speak with the nurses for increasing lengths of time. In the third phase of the treatment study, the patient was given a free supply of tokens equivalent to what he had earned in the previous phase; the tokens were no longer contingent on his behavior. Under those conditions the amount of conversation gradually declined, an example of extinction. When the original reinforcement conditions were reintroduced during the final experimental phase, the patient's conversational ability improved. Similar procedures were used to generalized his newfound conversational ability to other staff members, eventually allowing the patient to engage successfully in a rehabilitation program.

EXTINCTION Disordered behavior is often developed and maintained by reinforcement in the form of attention from others. In such cases the clinician must identify the reinforcer and remove it from the patient's environment, a procedure known as *extinction*. A simple example of the use of extinction is the case of a young child who cries interminably before going to sleep. When the mother puts the child to bed, the crying begins after a few minutes, and the mother returns to find out what the problem is. She then reads to the child or engages in some other activity. When she leaves, the pattern is repeated. Clearly, maternal attention is reinforcing the crying. The treatment is to persuade the mother not to return to the child's bedroom once the child is settled for the night. However, she needs to be warned that the amount of crying may increase for two or three nights before it diminishes. That phenomenon is known as an *extinction burst:* the person exhibits more behavior in order to increase the probability of gaining positive reinforcement. When the reinforcement is not forthcoming, the behavior diminishes in frequency.

Punishment Punishment is less useful as a therapeutic procedure than either reinforcement or extinction because it may produce unwanted effects, such as aggressive behavior, and the possibility of inflicting physical damage is always present. For the most part, punishment is used only in situations in which the behavior to be changed is injurious to the patient.

A clinical example of such a condition is rumination disorder in which infants regurgitate their food mouthful by mouthful, which leads to malnutrition and dehydration and is frequently a threat to life. One treatment is to use the principle of punishment, making an unpleasant event contingent on each episode of regurgitation. In the case illustrated in Figure 3.3-4, a drop of lemon juice applied to the infant's tongue was used as the unpleasant event. During the baseline period before treatment the infant ruminated for between 40 and 70 percent of the time that he was awake. Once the lemon juice was presented contingent on spitting up food, the frequency of rumination steadily declined. Punishment was then briefly removed, and rumination returned to the baseline levels, demonstrating the efficacy of punishment. The reintroduction of punishment eventually led to the virtual elimination of the behavior and a return to normal weight with no relapse at 1-year follow-up.

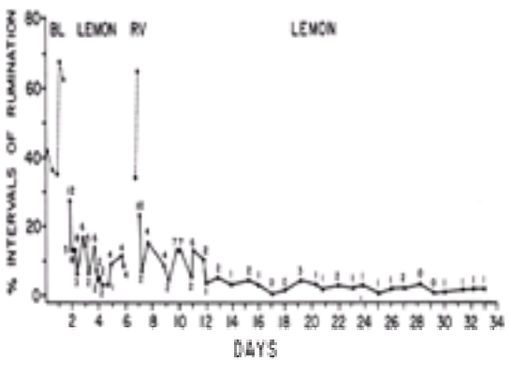

FIGURE 3.3-4 The effects of punishment in an experiment in which lemon juice was delivered contingent on ruminative vomiting in an infant. The frequency of rumination was rapidly reduced from the baseline (*BL*) and was increased only when punishment was withdrawn during the reversal phase (*RV*). The number of

applications of punishment (*lemon juice*) is shown by the number above each data point.

The use of punishment in a clinical situation should be carefully supervised and should follow certain rules. The behavior to be addressed should have been resistant to appropriate behavior-change procedures involving the use of positive reinforcement. Behaviors incompatible with the problem behavior can often be reinforced and the problem thus eliminated. In addition, the behavior to be changed should be severely incapacitating and should threaten physical integrity (e.g., the self-injurious behavior of some children with autistic disorder). Punishment procedures that themselves cause tissue damage should not be used. The behavior to be changed should be observed, measured, and recorded (as shown in Fig. 3.3-4) such that the effects of punishment and the amount of punishment used can be seen. If effectively used, punishment rapidly brings a behavior under control and behaviors can then be built up with the use of positive reinforcement.

SOCIAL-COGNITIVE LEARNING THEORY

Modes of Learning The psychologist Albert Bandura has integrated traditional classical and operant conditioning principles into a theoretically rich account of behavior and behavior change. A primary tenet of his approach is that the influence of environmental events on the acquisition and the regulation of behavior is primarily a function of cognitive processes. These processes are based on prior experience and determine what environmental influences are attended to, how they are perceived, whether they will be remembered, and how they may affect future action.

Reinforcement is regarded not as an automatic strengthener of behavior but as a source of guidance for behavior by anticipated outcomes. By observing the consequences of behavior, the person learns what action is appropriate in what situation. By symbolic representation of anticipated future outcomes of behavior, the person helps to generate the motivation to initiate and maintain behavior. Classical conditioning and operant conditioning are viewed as sources of learning about predictive relations among events. A third mode of learning is *modeling* (vicarious learning). In *modeling*, learning occurs through observation alone. The person need not exhibit any behavior or be directly reinforced for behavior. Modeling expands the scope and the complexity of learning influences on behavior. For example, it helps to explain how phobic reactions may be acquired in the absence of any direct traumatic experience. Young monkeys acquire a severe and lasting fear of snakes after observing their wild-reared parents act fearfully in the presence of a snake. Modeling has many therapeutic applications; it has prepared children for pending surgery by having them observe a film in which a child successfully copes with the novel and frightening events associated with preparation for and recovery from surgery.

The traditional emphasis on conditioning principles gives short shrift to verbal instruction as a mode of learning. The potential therapeutic effects of verbal instructions are illustrated by a study of hypertensive patients (Fig. 3.3-5). All participants in the experiment were told that relaxation training would help to lower their blood pressure. Half of the participants were also told that their blood pressure would show reductions after three sessions of relaxation training given in one morning; the other half were told that they could expect reductions only after prolonged relaxation practice. As shown in Figure 3.3-5 the groups have no difference in diastolic blood pressure readings. In systolic blood pressure, however, large and significant differences are shown; the group receiving immediate-lowering instructions showed lower blood pressure readings, but the other group did not. The mechanism underlying this effect is unknown. However, an expectancy of blood-pressure lowering may be needed to induce the biochemical changes necessary to lower blood pressure. This is an example of the complex interactions that occur between environmental events, in this case the instructions given to the patient, the patient's cognitive appraisal of the instructions, and neurochemical processes. Therapists tend to neglect the effect of therapeutic instructions, but experimental work suggests that the therapist should do everything to enhance the development of realistic outcome and efficacy expectations.

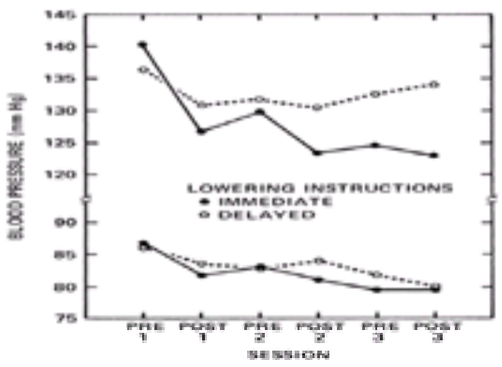

FIGURE 3.3-5 The effects of two expectancies on systolic (*top lines*) and diastolic (*bottom lines*) pressures of hypertensive persons. Systolic blood pressure was markedly reduced in the group told that relaxation training would lower blood pressure after one or two training sessions. In comparison the group told that blood pressure lowering would occur only after prolonged practice showed only slight reductions in systolic pressure.

Self-Efficacy Theory A component of social-cognitive learning theory with ramifications for the treatment of clinical problems is self-efficacy theory. *Efficacy expectations* are the degree of confidence that a person feels about coping effectively with a particular situation. *Outcome expectations* are defined as beliefs that particular actions will produce a certain outcome. *Self-efficacy* is the end product of different cognitive processes. To alter self-efficacy, people must actively appraise a specific experience and attribute successful coping to themselves as opposed to attributing it to some transient factor outside their control. Self-efficacy theory and cognitive-social learning theory in general draw heavily on the principles of attribution theory, a cognitive approach concerned with how people perceive the causes of behavior.

Self-efficacy theory predicts that the more people feel capable of predicting and controlling threatening events, the less vulnerable they are to anxiety and stress disorders in response to traumatic experiences; research with primates supports this prediction. Infant monkeys were reared in an environment in which they either exercised control over food, water, and special treats (the masters) or received reinforcers that were administered automatically and independent of their behavior (the controls). Months later, the masters were significantly less fearful in response to a threatening situation than were the controls.

The theory also predicts that psychological treatments such as exposure are effective because they enhance self-efficacy for coping with threatening events. Behavioral performance is the most powerful means of increasing self-efficacy, but self-efficacy is also influenced by sources of information derived from vicarious learning and verbal persuasion. Numerous studies have established that exposure treatment is one of the most effective methods for eliminating phobias. Exposure also produces greater increases in self-efficacy than do other methods. The greater the increase in self-efficacy, the greater the change in phobic behavior, regardless of the type of treatment. Outcome expectations (i.e., self-reports of anticipated anxiety about entering a phobic situation) tend to be correlated with efficacy expectations. If the correlation with efficacy expectations is eliminated statistically, outcome expectations fail to predict outcome, suggesting that change in self-efficacy is the important variable underlying behavior change.

Critics have charged that self-efficacy is only correlated with treatment-induced changes in phobic behavior and does not cause the behavior change. Bandura, however, points to the systematic covariation of experimentally induced levels of efficacy expectations and phobic behavior as evidence of their causal nature. Self-efficacy is related to measures of fear arousal. Increases in self-efficacy in phobic patients are associated with reductions in autonomic nervous system reactivity and neuroendocrine responses to phobic stimuli. The theory also suggests an explanation of the apparent resistance to the extinction of phobic behavior. Whether phobic reactions are eliminated by exposure to feared situations depends on the nature of the information that people derive from such experiences and not simply on the quantity of exposure. If people conclude that they can cope effectively, they no longer avoid the feared situations. But no change occurs if people conclude that they are unable to cope and hence experience unnerving anxiety. By strengthening patients' expectations that they cannot cope, exposure may even enhance their phobic sensitivity. Treatment should be aimed at increasing the patients' sense of predictability and controllability through enhanced self-efficacy. Treatment of patients with phobias along those lines produces results superior to treatment that passively exposes patients to feared stimuli without fostering coping skills, even if the length of the exposure is greater in the latter therapy.

Beyond furthering the analysis and the treatment of phobia disorders (panic disorder with agoraphobia, agoraphobia without a history of panic disorder, specific

phobia, and social phobia) self-efficacy has proved fruitful in the experimental analysis of a wide range of clinical problems, including pain management, the effects of stress on behavior and the immune system, and the prevention of relapse in substance abusers.

Self-Regulation In marked contrast to operant conditioning and radical behaviorism, social-cognitive learning theory emphasizes a capacity for self-regulation of behavior. Cognition is more than a passive conduit of external influences; it serves a generative function, allowing people to initiate thought, affect, and action to influence their circumstances, which, in turn, affect their cognition. People are neither driven inexorably by internal forces, nor are they passive reactors to environmental pressure; rather, they are both the agents and the objects of external influences.

Self-control strategies have become important components of most therapeutic interventions in behavior therapy. Among the most important elements of self-regulation of behavior are goal setting and feedback, self-monitoring, self-evaluation, and self-reinforcement.

Goal Setting and Feedback People set both long-term and short-term goals for themselves. The setting of sequential short-term goals leads to better performance than setting a distant goal. This is presumably related to the reinforcing qualities of goal attainment because reinforcing successive small steps is better than reinforcing the final behavior, at least until the behavior is well established. Goal definition is also important because the attainment of well-specified goals is more easily recognized than is the attainment of poorly defined goals. In general, the higher the goal set, the better the performance. At the highest levels of goal setting, performance begins to decline, underlining the fact that unrealistic goals undermine performance. Goal attainment enhances self-efficacy and affects future performance.

From a therapeutic viewpoint the therapist should help the patient define and set realistic, well-specified goals that signal small steps along the way to the overall goal so that demoralization can be kept at bay by success. Patients must set their own goals because self-determined goals lead to better performance than goals imposed by others. Teaching patients problem-solving strategies that they can use in many situations is a useful aspect of most therapies.

Goal attainment is not the only indicator of improvement in performance. Behavior change itself, if observed by the patient, can provide information regarding progress toward a particular goal. The process known as *informational feedback* enhances performance in a wide variety of tasks, such as learning to shoot accurately at a target, driving an automobile, and self-regulating autonomic processes. Removing informational feedback leads, at least temporarily, to setbacks. Information regarding therapeutic progress can be fed back to the patient in several ways. Patients can observe their own progress when the desired behavior change is relatively linear, such as approaching a phobic object or situation. Many behaviors, however, are complex. In such cases self-monitoring the behavior can enhance feedback. In addition, patients can plot the results of such feedback in graph form to examine their progress over long periods. Enhancement of information regarding progress is the central focus of biofeedback. In the typical biofeedback paradigm, processes that are not easily observed (e.g., blood pressure, small muscle contractions, and skin temperature) are made available for inspection by amplification. With sensitive and continuous feedback, the patient has the opportunity to learn to regulate invisible behaviors.

Self-Monitoring The basis of all self-control strategies is *self-monitoring*, which is the identification and recording of target problems and the conditions under which they occur. Typically, patients are asked to complete daily written records. For example, patients with panic disorder track all panic attacks and record the thoughts, feelings, and actions that preceded and accompanied the attacks. The goal is to identify the proximal determinants of the problem. Self-monitoring not only provides patients with an awareness of how their behavior affects or is affected by their social environment, it also prompts behavior change as patients identify specific influences on their behavior.

Self-Evaluation and Self-Reinforcement People who adopt certain standards and monitor their performances evaluate their success or failure in achieving those standards. That self-evaluation is the basis for self-reinforcement. Performances that match or exceed the standards serve as cues for rewards, and people deny themselves rewards for substandard performance. The essence of self-reinforcement is that people make freely available rewards contingent on behavior that meets preset standards. Common clinical problems, such as depression, involve the adoption of unrealistic or perfectionistic standards or excessively harsh and judgmental self-evaluation, regardless of objective performance.

Theories of Relapse One of the major problems facing therapists of all persuasions is that of patient relapse. Some instances of relapse can be easily understood; for example, the original environmental influences reinforcing symptomatic behavior may not have altered, and the patient's behavior is brought under their control once treatment has ended. This is an example of insufficient treatment; for example, perhaps the patient's family should have been brought into therapy. Sometimes, however, relapse denotes an impossible situation for the therapist, who is unable to alter a noxious psychological environment.

Less is known about the process of relapse than about the acquisition of behavior. For the most part, relapse has been studied in the addictive disorders, such as alcoholism and opiate addiction, and in related disorders, such as cigarette smoking and obesity. The basic theory concerning relapse involves situations that pose a high risk for engaging in the problem behavior, situations in which the behavior has occurred at a high frequency in the past. According to social-cognitive learning theory, if such a situation is coped with successfully, the person experiences an increase in self-efficacy, which leads to a low probability of relapse. The assumption is that the person has been taught or has developed usable coping skills. However, a former substance abuser who is deficient in coping with the high-risk situation develops a positive expectancy regarding the beneficial effects of the substance and starts using the substance again. This practice results in an *abstinence violation* effect, defined as the breaking of a self-imposed rule, leading to a diminished sense of self-efficacy.

FROM LEARNING THEORY TO PSYCHOTHERAPY

As a broad-based framework that emphasizes the multidimensional nature of psychological functioning, social-cognitive learning theory provides a flexible guide for the design of treatments for particular conditions. Treatment programs typically combine several principles and procedures. In formulating a therapy for a particular disorder, it is first necessary to develop a model, however simple, of the factors maintaining the symptoms of the particular disorder. Similarly, for the purposes of prevention, a model containing remediable risk factors for the development of the disorder is needed. The factors contained in a particular model and their relation to one another derive from a variety of sources, such as clinical observations, epidemiological studies, naturalistic observations, and laboratory experiments. The models developed at any one time may be updated when new findings concerning the disorder are made, or when controlled treatment studies suggest that additional factors need to be taken into account.

For example, the models of factors maintaining bulimia nervosa and depression differ in their details but both include behavioral antecedents and cognitive mediators of those antecedent behaviors. The treatments themselves share common elements derived from learning principles. The therapy sessions are structured to maximize learning, beginning with a review of self-monitoring, including the homework assignments. Based on this information and depending on the stage of therapy, the therapist sets an agenda for the session, which may include information on new procedures to be introduced in that session, followed by behavior change procedures deriving from the information gained from the self-monitoring records. This is followed by a summary of the session and the setting of homework. Within the therapy itself an array of procedures are used including: setting realistic outcome expectancies, strengthening efficacy expectancies, information and educational strategies, data collection, feedback regarding progress, reinforcement of adaptive behaviors, and cognitive change procedures.

Bulimia Nervosa Clinical, naturalistic, and laboratory studies all suggest that dietary restriction leads to excessive hunger and disinhibition of eating whereby individuals with bulimia nervosa develop a sense of loss of control over eating that leads to binge eating. The dietary restriction appears to be driven by overconcern about shape and weight, hence dieting (when it fails) is followed by various methods of purging to compensate for the actual or perceived excessive caloric intake. In addition, low self-esteem deriving from a multiplicity of developmental causes appears to enhance concerns about weight and shape. These considerations led to the development of the relatively simple model of factors that maintain bulimic behavior, illustrated in Figure 3.3-6.

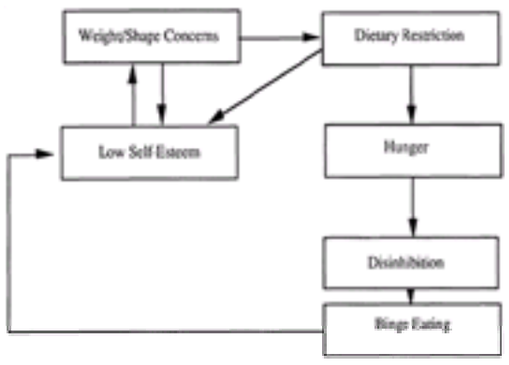

FIGURE 3.3-6 A hypothetical model of psychosocial factors maintaining bulimic behaviors.

Learning theory suggests that the antecedent events closest to the primary problem behavior should be addressed first. Hence, the first aim of cognitive-behavioral therapy for bulimia nervosa is to reduce the dietary restriction that leads to hunger and binge eating. To do this several learning principles are used. First, an educational approach is taken, presenting information concerning the factors maintaining binge eating and purging, together with a rationale for the treatment procedures stemming from the model. This is aimed at developing an expectation on the patient's behalf that reducing dieting will help to reduce the sense of loss of control and the consequent binge eating. Once the patient has accepted the rationale, the first phase of therapy begins. In this phase several additional principles are used. The patient is asked to monitor her eating behavior (i.e., meal content, binges, the circumstances surrounding binge eating and purging) on self-monitoring forms. This crucial aid to self-regulation provides information both to the therapist and the patient about the details of the problem and, in later sessions, about progress in changing the problem behaviors. The restrictive dieting is directly addressed by helping the patient to eat three meals and two snacks each day, followed by increasing the amount eaten on each occasion. Here, a process of shaping is used to allow gradual behavior change. Appropriate changes are reinforced by selective therapist attention and praise, and problems are discussed until a new solution is found. Such a solution is then tried out by the patient between sessions as homework.

Once a more stable pattern of eating has been attained the factors that drive dietary restriction such as overconcern about weight and shape are addressed. One of the learning procedures used is data gathering on the problem, including the hoped-for but unrealistic outcomes the patient believes she will attain by maintaining a thin body shape. Such outcome expectancies can be challenged regarding their reality, and alternative ways of achieving more reasonable outcome expectancies can be considered.

In the final phase of therapy, relapse prevention strategies are implemented. First, behavior change strategies that have proved successful in treatment are reviewed. Lapses occurring during this phase of treatment, for example, a small binge or even an urge to binge, are examined in detail. The events leading up to the lapse are reviewed and alternative coping strategies are considered for their potential effectiveness in preventing relapses. More generally, situations specific to the patient that are likely to precipitate relapse are identified and potential coping strategies are also identified. A case example in which cognitive-behavioral therapy was used to treat a patient with bulimia nervosa illustrates the use of these principles.

The patient was a 29-year-old married woman who, because of concerns about her weight had begun to diet in early adolescence. Although she was within the normal limits of weight for her height, her mother's concerns about her own weight and remarks about the patient's “plumpness” apparently drove her to ever-stricter dieting. This was followed later in adolescence by the occurrence of binge eating with a consequent weight gain of several pounds. To combat this weight gain the patient began to purge by inducing vomiting with her fingers following binges; later she was able to induce vomiting at will. When first seen at the clinic she was binge eating and purging almost fifteen times each week. Physical examination and electrolytes were normal except for a raised serum amylase, a finding not uncommon in patients with bulimia nervosa.

Therapy began by elaborating on her history, exploring in more detail the links between feelings of low self-esteem engendered by childhood experiences, and the importance of weight and shape to her morale, which led her to dieting and binge eating. The patient was able to perceive these links and to accept the model outlined in Figure 3.3-6. She accepted the possibility that reducing dieting would lessen binge eating and purging; however, she feared gaining weight if she gave up dieting. She was not particularly reassured by the findings from controlled trials that the average patient does not gain weight, but was willing to take an experimental approach to the problem. She was advised to weigh herself no more than once each week to reduce overreaction to natural weight fluctuations and was taught how to monitor and record her food intake, binge eating and purging, and the circumstances under which such behaviors occurred.

At the next session the food records were examined in detail with the therapist pointing out her pattern of eating little or no breakfast, a small salad for lunch, and then binge eating at least once during the evening. The link between dieting, hunger, and binge eating was again pointed out and the therapist told her to eat three meals and two snacks by the clock, rather than relying on her own hunger sensations, which were disrupted by her chaotic eating patterns. During the next three or four sessions the therapist and patient collaborated to gradually achieve this goal by using the principle of shaping and reinforcing positive changes by therapist praise and attention. By the fifth session the patient was eating three meals and two snacks on many days. The patient recognized from her food record that she was more likely to binge on days when the desired meal pattern was not maintained. This reinforced her adherence to eating three meals and two snacks daily, and also allowed her to experiment with consuming somewhat larger meals with continuing diminution of binge eating and purging. By the eighth therapy session she completed her first week free of binge eating and purging, without any noticeable weight gain. Such progress during the first half of therapy is undoubtedly a good prognostic indicator.

The next phase of treatment involved broadening food choices. The therapist recommended a behavioral task, namely, that the patient visit a supermarket and make a list of all the foods that she would *not* buy because she feared eating them. The patient complied and was asked to add some of the lesser-feared foods to her diet. With these additions continuing over the next few sessions, the bulk of the work then shifted to an examination of her distorted thinking about weight and shape. The patient was particularly prone to blame her appearance for any rejection she perceived. Such incidents were explored in detail and usually led her to accept an alternative conclusion regarding the perceived rejection. She then began to practice these exercises on her own when such thoughts arose. She was also convinced that her body was less “perfect” than the bodies of other women. Hence, she was given the task of actually looking at other women's bodies and considering their “imperfections.” She gradually became persuaded that she was not as overweight as she had thought. Although the patient occasionally lapsed into binge eating and purging, for most of the time during this phase of therapy she had been eating normally.

In the final few sessions the patient was asked to review the procedures that had helped her most in order to develop a written maintenance plan. Among the essential elements she listed eating regularly and not avoiding foods, giving up dieting, and being somewhat more accepting about her weight and shape. Circumstances that might lead to further lapses were also explored, focusing on the few lapses that had occurred during the past weeks; this allowed the patient to plan to cope better with such situations. For example, she reported being tempted to binge when she was made angry by an interpersonal interchange, particularly if she returned home alone following such an incident. She developed several alternate coping strategies such as talking the incident over with a friend or planning to eat out or go to a movie to ensure that she would not be alone under such circumstances. The systematic application of these cognitive and behavioral strategies illustrates the clinical use of relapse prevention strategies based on social-cognitive learning theory. The patient was monitored at 3-month intervals to check up on her progress and deal with any residual problems. One year following the end of treatment she had continued to abstain from binge eating and purging and treatment ended.

DEPRESSION

There are various models of the factors that maintain a depressed mood. One such model is shown in Figure 3.3-7. Distorted cognitive processes and a lack of reinforcing behavior, particularly of interpersonal behavior, are believed to maintain a depressed mood. A triad of cognitive dysfunctions is postulated; the most proximal to a depressed mood are automatic negative thoughts. Errors in information processing, which are state dependent, occurring only in the presence of a depressed mood, are the second element in the cognitive triad. Events are perceived and processed mistakenly to suggest a potential negative outcome, thus amplifying negative moods. At a deeper level, usually not easily accessible to the patient, are more general distorted attitudes and beliefs that give rise to the

automatic negative thoughts and to errors in information processing. Also, the interpersonal world of the patient is often inherently nonreinforcing, which leads to a decline in interpersonal activities that further cuts the patient off from potential reinforcers, with the potential to spiral into a deep withdrawal and a seriously depressed mood.

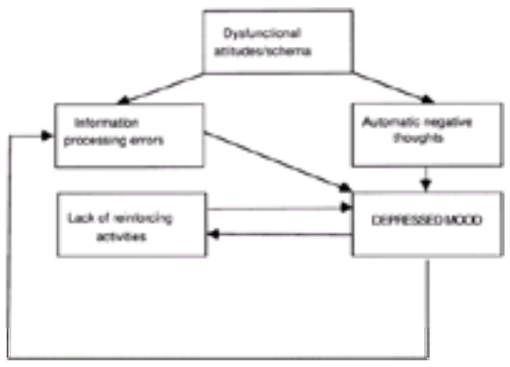

FIGURE 3.3-7 A hypothetical model of psychosocial factors maintaining depression.

Cognitive-behavioral therapy for depression begins by clarifying the patient's expectations regarding the outcome of treatment by engendering positive but realistic expectations. The relationship with the patient is described as collaborative—the patient gathers data and with the help of the therapist formulates behavior changes that are likely to ameliorate depression. Therapy usually begins with self-monitoring of mood states and the events surrounding them, and with encouragment and reinforcement of proactive behaviors that the patient will find reinforcing. The aim is to gradually shape a wider repertoire of behaviors leading to reinforcement in the patient's life by identifying behaviors associated with depressed or normal mood states, providing feedback regarding progress derived from further self-monitoring, and reinforcement of progress by the therapist. Feedback is also sought from the patient regarding the relevance and effectiveness of the procedures used in each therapy session. As the patient begins to make progress, automatic negative thoughts are examined in detail; Socratic questioning is used to teach the patient to evaluate the accuracy of the outcome expectancies involved in such thoughts. Self-monitoring is now focused on maladaptive thinking. As therapy continues, underlying attitudes and cognitive schemas are gradually uncovered and clarified in similar ways. In the final phase of treatment, relapse-prevention procedures are used, often attempting to identify broad behavior patterns associated with depressive thinking and helping patients to better cope with such problems.

SUGGESTED CROSS-REFERENCES

Applications of the principles and procedures discussed in this section, together with an assessment of the effectiveness of such therapeutic applications to various conditions, can be found in Section 30.2 on behavior therapy and Section 30.6 on cognitive therapy. Issues basic to learning are discussed in Chapter 1 on neural sciences and in Section 3.1 on perception and cognition. Derivations from the principles and procedures exemplified in this section are discussed in Chapter 6 on theories of personality and psychopathology that are derived from psychology and philosophy.

SECTION REFERENCES

Agras WS: Helping people improve their lives with behavior therapy. Behav Ther *28:*375, 1997.

Ader R, Cohen N: CNS-immune system interactions: Conditioning phenomena. Behav & Brain Sci *8:*379, 1994.

*Bandura A: *Principles of Behavior Modification*. Holt, Rinehart & Winston, New York, 1969.

Bandura A: *Self-efficacy: The Exercise of Control*. Freeman, New York, 1997.

Beck AT: *Cognitive Therapy of Depression: A Treatment Manual*. Guilford, New York, 1979.

Bovbjerg DH, Redd WH, Maier LA, Holland JC: Anticipatory immune suppression and nausea in women receiving cyclic chemotherapy for ovarian cancer. J Consult Clin Psychol *58:*153, 1990.

Brownell KD, Marlatt GA, Lichenstein E, Wilson GT: Understanding and preventing relapse. Am Psychol *41:*764, 1986.

*Clark DM, Fairburn CG: *Science and the Practice of Cognitive Behaviour Therapy*. Oxford University Press, New York, 1997.

Clark A, Hawkins RD, Kandel ER: Cell biological perspectives on learning. In *Diseases of the Nervous System*, AK Ashury, M McKhann, WH MacDonald, editors. Saunders, Philadelphia, 1986.

Clark RE, Squire LR: Classical conditioning and brain systems: The role of awareness. Science *280:*77, 1998.

Dollard J, Miller NE: *Personality and Psychotherapy*. McGraw-Hill, New York, 1950.

Hilgard ER, Bower G: *Theories of Learning*, ed 4. Prentice-Hall, Englewood Cliffs, NJ, 1975.

Hollon SD, Beck AT: Cognitive and cognitive-behavioral therapies. In *Handbook of Psychotherapy and Behavior Change: An Empirical Analysis*, ed 4, SL Garfield, AE Bergin, editors. Wiley, New York, 1994.

Kandel ER, Spencer WA: Cellular neurophysiological approaches in the study of learning. Physiol Rev *48:*65, 1968.

*Kandel ER: Genes, nerve cells, and the remembrance of things past. J Neuropsychiatry Clin Neurosci *1:*103, 1989.

Kendall PC: Healthy thinking. Behav Ther *23:*1, 1992.

Mineka S, Davidson M, Cook M, Keir R: Observational conditioning of snake fear in rhesus monkeys. J Abnorm Psychol *93:*335, 1984.

Mineka S, Gunnar M, Champoux M: Control and early socioemotional development: Infant rhesus monkeys reared in controllable versus uncontrollable environments. Child Dev *57:*1241, 1986.

Mischel W: Toward a cognitive social learning reconceptualization of personality. Psychol Rev *80:*252, 1973.

Mowrer OH: *Learning and Behavior*. Wiley, New York, 1960.

Pavlov IP: *Conditioned Reflexes*. Clarendon Press, London, 1927.

Poling A: *A Primer of Human Behavioral Pharmacology*. Plenum, New York, 1986.

Rachlin H: *Introduction to Modern Behaviorism*. Freeman, New York, 1970.

*Rauhut AS, McPhee JE, Ayres JJB: Blocked and overshadowed stimuli are weakened in their ability to serve as blockers and second-order reinforcers in Pavlovian fear conditioning. J Exp Phychol Anim Behav Process *25:*45, 1999.

*Rescorla RA: Pavlovian conditioning. It's not what you think. Am Psychol *43:*151, 1988.

Segal ZV, Williams JM, Teadale JD, Gemar M: A cognitive science perspective on kindling and episodic sensitization in recurrent affective disorder. Psychol Med *26:*371, 1996.

Seligman MEP: Phobias and preparedness. Behav Ther *2:*107, 1971.

Sherman JE, Jirenby DE, Baker TB: Classical conditioning with alcohol: Acquired preferences and aversions, tolerance, and urges/craving. In *Theories on Alcoholism*, CD Chaudron, DA Wilkinson,

editors. Addiction Research Foundation, Toronto, 1988.

*Skinner BF: *Science and Human Behavior*. Macmillan, New York, 1951.

Schwartz JC, Stoessel PW, Baxter L, Martin KM, Phelps ME: Systematic changes in cerebral glucose metabolic rate after successful behavior modification treatment of obsessive-compulsive disorders. Arch Gen Psychiatry *53:*109, 1996.

Thorndike EL: *Human Learning*. Century Company, New York, 1931.

Watson JB, Rayner R: Conditioned emotional reactions. J Exp Psychol *3:*1, 1920.

Wilson GT, Fairburn CG, Agras WS: Cognitive-behavioral therapy for bulimia nervosa. In *Handbook of Treatment for Eating Disorders*, DM Garner, P Garfinkel, editors. Guilford Press, New York, 1997.

Wolpe J: *Psychotherapy by Reciprocal Inhibition*. Stanford University Press, Stanford, CA, 1958.

3.4 BIOLOGY OF MEMORY

LARRY R. SQUIRE, FH.D., AND KEN A. PALLER, FH.D.

The topic of memory is fundamental to the discipline of psychiatry. Memory provides the essential substrate for the cognitive activities that define human experience, it allows one to connect the present moment to what came before, and it is the basis of cultural evolution.

An individual's personality reflects habits and dispositions that have developed from experience. Adaptive and maladaptive coping strategies, anxieties, and phobias are largely products of learning. Neurotic or psychotic symptoms can be the consequences of specific experiences or repeated patterns of experience. Psychotherapy is a process by which new behaviors are acquired through the accumulation of new experiences. Thus, memory is at the heart of psychiatry's concern with the effects of early experience, the development of the individual, and the possibility of change.

Disorders of memory and complaints about memory lapses are pervasive in both neurology and psychiatry. Memory problems are also of special concern as side effects of psychopharmacological treatments and electroconvulsive therapy. Accordingly, the effective clinician needs to understand memory, its psychological and neurological foundations, the varieties of memory dysfunction, and how memory can be evaluated. The biological perspective on memory developed here rests on a growing body of neuroscientific evidence that relates mental events to the functioning of the brain.

MEMORY AS SYNAPTIC CHANGE

Memory is a special case of the more general phenomenon of *neural plasticity*. Neurons can show history-dependent behavior by responding differently as a function of recent input, and this plasticity of nerve cells and synapses is the basis of memory. In the last decade of the nineteenth century, researchers proposed that the persistence of memory could be accounted for by nerve cell growth. Others have restated this idea, developing the hypothesis that the synapse is the critical site of change. In principle, there are many possible ways for such structural change to be realized, including alterations in the number of synaptic contacts or in the strength of existing contacts.

Plasticity Neurobiological evidence from animal studies supports two basic conclusions about the biology of memory. First, specific synaptic events, including an increase in neurotransmitter release, are responsible for short-lasting plasticity, which may last for seconds or minutes. Second, long-lasting memory depends on new protein synthesis, physical growth of neural processes, and an increase in the number of synaptic connections.

A major source of information has been the extended study of the marine mollusc *Aplysia californica*. A sufficient number of individual neurons and connections between neurons have been identified to allow the wiring of some simple behaviors to be diagrammed. *Aplysia* is capable of both associative learning (including classical conditioning and operant conditioning) and nonassociative learning (habituation and sensitization). Figure 3.4-1 shows the circuitry responsible for the gill-withdrawal reflex, a defensive reaction whereby tactile stimulation causes the gill and siphon to retract. When tactile stimulation is preceded by stimulation to the head, gill withdrawal is facilitated. The cellular mechanisms underlying this sensitization are based on an enhanced release of neurotransmitter by the facilitatory neuron (labeled “Int” in Fig. 3.4-1) and accompanied by covalent modifications of preexisting proteins. Under some training conditions, sensitization can persist for weeks, and these longer-lasting changes can also be produced by repeated applications of serotonin, distributed over a period of 1½ hours. Although both short- and long-lasting plasticity are based on enhanced transmitter release, the long-lasting change uniquely requires the expression of genes and the synthesis of proteins. In addition, the long-term change, but not the short-term change, is accompanied by the growth of neural processes of neurons within the reflex circuit.

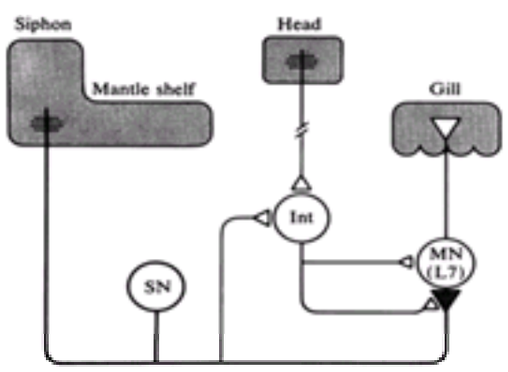

FIGURE 3.4-1 A schematic diagram of the neuronal circuit underlying behavioral habituation and sensitization of the gill-withdrawal reflect in *Aplysia*. The relative simplicity of the nervous system of *Aplysia* makes it a valuable organism for studying cellular and synaptic mechanisms of memory. The synapse between the sensory neuron (SN) and the motor neuron (MN) is an important site of habituation. Sensitization results from activation of the interneuron (Int) pathway. (Reprinted with permission from Kandel ER: *Cellular Basis of Behavior*. Freeman, San Francisco, 1976.)

In vertebrates, behavioral manipulations can also result in measurable changes in the brain's architecture. For example, rats reared in enriched environments show an increase in the number of synapses ending on individual neurons in neocortex. These changes are accompanied by small increases in cortical thickness, in the diameter of neuronal cell bodies, and in the number and length of dendritic branches. New synapses may be formed directly or synapses may be selectively preserved from a population that is continuously being replaced. Behavioral experience thus exerts powerful effects on the wiring of the brain.

Many of these same structural changes have been found in adult rats exposed to an enriched environment, and some have been found in adult rats given extensive maze training. In this case opaque contact-lens occluders were used to restrict vision to one eye, and the corpus callosum was transected to prevent information received by one cerebral hemisphere from reaching the other hemisphere. In these monocularly trained animals, increases in the size of dendritic fields of pyramidal neurons of occipital cortex were found only in the trained hemisphere. This finding rules out a number of nonspecific influences including motor activity, indirect effects of hormones, and overall level of arousal. Therefore it seems likely that long-term memory in vertebrates is generally based on specific changes within the neurons that lie along specific pathways.

Long-Term Potentiation The phenomenon of *long-term potentiation* (LTP) is a form of neural plasticity likely to be important for memory in vertebrates. LTP is observed when a postsynaptic neuron is persistently depolarized following a brief burst of high-frequency stimulation. LTP has a number of properties that make it a promising candidate as a physiological substrate of memory. First, it is established quickly and then lasts for a long time. Second, it is associative in that it depends on the co-occurrence of presynaptic activity and postsynaptic depolarization. Third, it occurs only at the potentiated synapses, not at all the synapses terminating on the postsynaptic cell. Finally, LTP occurs prominently in the hippocampus, a structure with important memory functions. The induction of LTP is known to be mediated postsynaptically and to involve activation of the *N*-methyl-D-aspartate (NMDA) receptor, which permits influx of calcium into the postsynaptic cell. The mechanism whereby LTP is maintained is not clearly established, but evidence has been presented in favor of a presynaptic locus of change (increased transmitter release).

Rapidly developing structural changes in the dendritic spines of the postsynaptic neuron have also been described in association with LTP.

A new method for studying molecular mechanisms of memory relies on introducing specific mutations into the genome. For example, by altering a single cloned gene, a mutant strain of mice can be produced with specific receptors or cell-signaling molecules inactivated or altered. This knock-out technique can provide greater specificity than pharmacological blocking methods. Recently, it has been possible to study mice with a selective deletion of one type of NMDA receptor in the CA1 field of the hippocampus. Although many aspects of CA1 physiology remain intact, the CA1 cells do not exhibit LTP. In addition, an impairment is observed on a learning task. If reversible gene knock-outs can be achieved, it will be possible to induce specific molecular changes in a developmentally normal adult.

Associative Learning Additional insights into memory have been gleaned from the study of the neural circuitry underlying classical conditioning of the eyeblink-nictitating membrane response in rabbits. Repeated pairings of a tone (conditioned stimulus) and an airpuff to the eye (unconditioned stimulus) lead to a conditioned eyeblink in response to the tone. Reversible lesions of the deep nuclei of the cerebellum eliminate the conditioned response without affecting the unconditioned response, which indicates that the cerebellum contains part of the essential circuitry for learned association, the conditioned stimulis–unconditioned stimulus link. Reversible lesions of the deep nuclei also prevent learning from occurring, and the rabbits begin learning from the naive state when the lesion is reversed. This finding does not mean that all the changes occurring in the animal during conditioning involve the cerebellum; it means only that essential neural changes responsible for the conditioned stimulus–unconditioned stimulus link depend on this circuitry. The relevant plasticity appears to be distributed between the cerebellar cortex and the deep nuclei (Fig. 3.4-2). An analogous pattern of plasticity is thought to underlie motor learning in the vestibulo-ocular reflex, and perhaps associative learning of motor responses in general. Based on the idea that learned motor responses depend on coordinated control of changes in both timing and strength of response, it has been suggested that synaptic change in the cerebellar cortex is crucial for learned timing, whereas synaptic change in the deep nuclei is crucial for learned changes in the strength of the response.

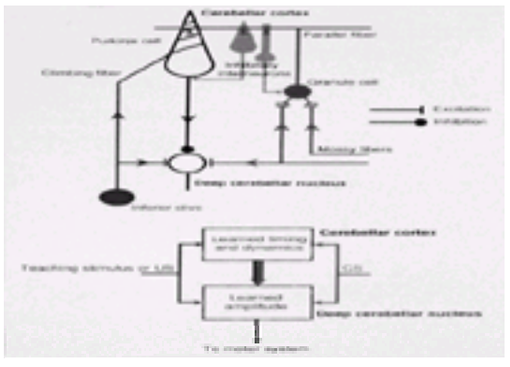

FIGURE 3.4-2 A schematic diagram of the circuitry of the mammalian cerebellum (**top**). In the classically conditioned blink response, input from the air-puff unconditioned stimulus and input from the auditory conditioned stimulus comes in through parallel pathways to the cerebellar cortex and to the deep cerebellar nucleus, and plasticity occurs in both pathways (**bottom**). (Reprinted with permission from Raymond JL, Lisberger SG, Mauk MD: The cerebellum: A neuronal learning machine? Science *272:*1126, 1996. © 1996 American Association for the Advancement of Science.) (See Color Plate 4.)

Understanding the biology of memory requires more than just an understanding of the synaptic events that store memory. It is also essential to understand how and where synaptic events are organized in the brain. Many levels of analysis can be identified between synaptic change and behavioral memory, and many important questions about memory address levels of biological analysis that are intermediate to synapses and behavior.

CORTICAL ORGANIZATION OF MEMORY

The question of where memories are stored in the brain has long been a major research issue. In the 1920s Karl Lashley carried out a series of experiments that were directed at this problem. Lashley recorded the number of trials that rats needed to relearn a preoperatively learned maze problem after removal of different amounts of cerebral cortex. The deficit was proportional to the amount of cortex removed and, further, it seemed to be qualitatively similar, regardless of what region of cortex was removed. Lashley concluded that memory for the maze habit was not localized in any one part of the brain, but instead was distributed equally over the entire cortex. Subsequent work has led to a revision of this idea. Maze learning in rats depends on many forms of information, including visual, tactual, spatial, and olfactory information. These various forms of information are processed and stored in different areas. Thus, the correlation between retention score and lesion size that Lashley observed reflects the progressive encroachment on specialized cortical areas serving the many components of cognition important to maze learning.

The specialized cortical areas responsible for processing and storing visual information have been studied most extensively in nonhuman primates. Nearly half of the primate neocortex is specialized for visual functions. Cortical pathways for visual information processing (Fig. 3.4-3) begin in primary visual cortex (V1) and proceed from there along parallel pathways or streams. One stream projects ventrally to the inferotemporal cortex (area TE in the monkey) and processes information about the quality of visual percepts. Another stream projects dorsally to the parietal cortex and processes information about spatial location. Electrophysiological studies in the monkey show that neurons in area TE register specific and complex features of visual stimuli, like shape, and may even respond selectively to patterns and objects. These specific visual processing areas, along with connections to corresponding regions in dorsolateral prefrontal cortex, are involved in the immediate experience of perceptual processing, and in what has been called *immediate memory* or *working memory*. These areas also serve as the ultimate repositories of the memories that result from their activity. Accordingly, lesions in these areas lead to impairments in visual perception as well as in visual learning and memory, although elementary visual functions such as acuity remain intact. Inferotemporal cortex can thus be thought of both as a higher-order visual processing system and a storehouse of the visual memories that result from that processing. These stored visual memories can be used and manipulated according to current processing demands, and they can also be quite long-lasting.

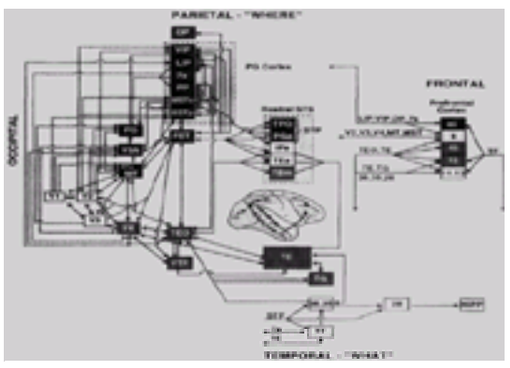

FIGURE 3.4-3 Summary of cortical visual areas and some of their connections. There are two major pathways from striate cortex (V1). The processing stream for object vision follows a ventral route into the temporal lobe via V4 (dark gray boxes) and the processing stream for spatial vision follows a dorsal route into the parietal lobe via MT (light gray boxes). Solid lines indicate connections arising from both central and peripheral visual field representations; dotted lines indicate connection restricted to peripheral field representations. Shaded region on the lateral view of the brain represents the extent of the cortex included in the diagram. Abbreviations: DP, dorsal prelunate area; FST, fundus of superior temporal area; HIPP, hippocampus; LIP, lateral intraparietal area; MSTc, medial superior temporal area, central visual field representation; MSTp, medial superior temporal area, peripheral visual field representation; MT, middle temporal area, MTp, middle temporal area, peripheral visual field representation; PO, parieto-occipital area; PP, posterior parietal sulcal zone; STP, superior temporal polysensory area; VIP, ventral intraparietal area; STS, rostral superior temporal sulcus; and VTF, visual responsive portion of area TF. (Reprinted with permission from Ungerleider LG: Functional brain imaging studies of cortical mechanisms for memory. Science *270:*769, 1995. © 1995 American Association for the Advancement of Science.)

Many parts of the nervous system participate in storing representations of an event in memory. During an event, visual information is stored in inferotemporal cortex so that the same visual material can later be recognized as familiar. Concurrently, other components of the event—including spatial, temporal, tactile, olfactory, emotional, and other sorts of information—are processed and stored separately. Memory storage in the cerebral cortex thus depends on a fractionation of experience as follows. First, any particular event or learning task is composed of a number of components. Second, each component engages a particular processing site or set of sites. Third, each processing site stores information as an outcome of the processing that is done.

Thus, memory is both distributed and localized in the nervous system. It is distributed in the sense that, as Lashley concluded, there is no unitary cortical center dedicated solely to the storage of memories. Yet, memory is localized in the sense that different aspects or dimensions of events are stored at specific cortical sites—the same regions specialized to analyze and process those particular aspects or dimensions of information.

INSIGHTS FROM AMNESIA

The idea that the functional specialization of cortical regions governs both information processing and information storage is important, but it does not provide a complete account of the organization of memory in the brain. If it did, then particular cortical injuries would disrupt only particular domains of learning and memory (i.e., visual memory or spatial memory) and no global disruption of memory would occur. Brain injury would always produce a difficulty in learning a restricted type of new information along with a loss of previously learned information of that same type. Yet common neurological syndromes of memory impairment conflict with these expectations.

The hallmark of neurological memory impairment is a profound anterograde amnesia, or loss of new learning ability, that extends across all sensory modalities. Typically, this occurs together with retrograde amnesia, a memory loss for information acquired prior to the onset of amnesia. The retrograde deficit often has a temporal gradient, such that recall for recent events is impaired, but recall for remote events is intact. Other cognitive functions are preserved, including attention, immediate memory, personality, and social skills.

The selectivity of the memory deficit in amnesia implies that the brain has isolated intellectual and perceptual functions from the ability to lay down a record of information processing. The cognitive dysfunction experienced by amnesic patients affects memory storage but does not affect a wide range of other intellectual capabilities. The fact that memory storage is affected for all sensory modalities without a parallel disruption of perception implies that the memory function is superimposed on normal cortical processing. The fact that anterograde amnesia often occurs together with intact remote memory implies that viable retrieval mechanisms are intact, and also that the brain structures damaged in amnesia are not the ultimate repositories of memory. Detailed studies of amnesic patients and models of amnesia in nonhuman animals have illuminated these issues considerably.

Specialized Memory Function Amnesia results from damage to either of two brain regions: the medial temporal lobe or the midline diencephalon. Early studies of a severely amnesic patient known as HM markedly stimulated investigation of the role of the medial temporal lobe.

H.M. became amnesic in 1953, when he sustained a bilateral resection of the medial temporal lobe to relieve severe epilepsy. The removal included approximately half of the hippocampus, most of the amygdala, and the neighboring entorhinal and perirhinal cortices. Following the surgery, H.M.'s seizure condition was much improved. Moreover, he retained normal language, normal intellectual functions, and normal immediate memory (e.g., as tested with a digit span test). However, he exhibited profound forgetfulness, and this deficit has persisted for more than 40 years.

Extensive investigations of other amnesic patients have also been used to explore the memory functions of the medial temporal lobe. For example, patient R.B. became amnesic following an episode of global ischemia. He suffered from a moderately severe anterograde amnesia with minimal retrograde amnesia. After his death 5 years later, extensive histological study of his brain revealed a circumscribed bilateral lesion of hippocampal field CA1, whereas the minor additional pathology that was found could not reasonably explain the memory impairment. Similar pathological findings in the hippocampus have also been observed in other amnesic patients (Fig. 3.4-4). Magnetic resonance imaging (MRI) with high-resolution protocols can reveal pathology in the hippocampal region of amnesic patients in vivo. Two conclusions about the anatomical correlates of amnesia follow. First, damage limited to the hippocampus itself can result in clinically significant memory impairment; second, medial temporal regions in addition to hippocampal field CA1 also make a critical contribution to memory.

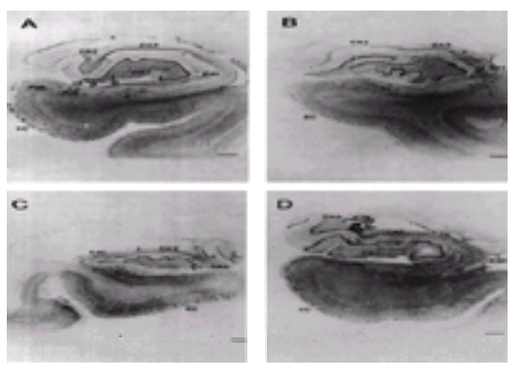

FIGURE 3.4-4 Coronal sections through the hippocampal region stained with thionin in a normal subject (**A**) and three amnesic patients with bilateral damage to the hippocampal formation (**B-D**). The hippocampus proper can be divided into three distinct fields, designated *CA1, CA2*, and *CA3*. The CA1 field extends to the subiculum (*S*). Other structures include the dentate gyrus (*DG*), presubiculum (*PrS*), parasubiculum (*PaS*), and entorhinal cortex (*EC*). In patient GD (**B**), damage included CA1; in patient LM (**C**), damage included CA1, CA2, CA3, DG, and EC; in patient WH (**D**), damage included CA1, CA2, CA3, DG, S, and EC. For additional details, see Rempel-Clower N, Zola SM, Squire LR, Amaral DG: Three cases of enduring memory impairment following bilateral damage limited to the hippocampal formation. J Neurosci *16:*5233, 1996. (Reprinted with permission from Squire LR, Zola SM: Memory, memory impairment, and the medial temporal lobe. In *Cold Spring Harbor Symposia on Quantitative Biology*, vol 61. Cold Spring Harbor Laboratory Press, Plainview, New York, 1996.)

The findings from human amnesia inspired the development of models of amnesia in experimental animals. Early animal studies yielded contradictory findings that could not be easily related to memory impairment. In part, the difficulty was that human amnesia itself was poorly understood. Memory is now known to be a collection of different abilities and not a unitary mental faculty. Human amnesia does not affect all kinds of memory. Until researchers understood this, selecting memory tasks for making comparisons across species was problematic. Indeed, obtaining memory performance measures that reflect parallel memory functions in humans and experimental animals requires a high degree of control over the cognitive strategies used.

Nonetheless, a model of human amnesia in the nonhuman primate became available in the early 1980s, and subsequent investigations identified the crucial structures and connections. In the medial temporal lobe, these include the hippocampus proper (CA fields, dentate gyrus, and subiculum) and adjacent regions of entorhinal, perirhinal, and parahippocampal cortex. Monkeys with surgical damage to specific structures were trained to perform tasks analogous to tasks sensitive to memory impairment in humans. Large medial temporal lesions intended to approximate the damage that occurred in patient H.M. caused monkeys to exhibit many features of human amnesia. For example, the impairment occurred in more than one sensory modality, short-term memory was intact, the deficit was enduring, skill learning was preserved, and retrograde amnesia was temporally graded.

Within the medial temporal lobe separate contributions can be identified for memory and emotion. The participation of the medial temporal lobe region in emotional expression was first studied systematically in 1937 by Heinrich Kluver and Paul Bucy, who found that monkeys with bilateral temporal lobectomy became tame, approached animals and objects without reluctance, examined objects by mouth instead of by hand, and exhibited abnormal sexual behavior. Subsequent studies have indicated that emotional behavior is related not to the hippocampus but to the adjacent set of nuclei known collectively as the amygdala. In addition, other work has shown that the amygdala is part of a set of structures essential for fear conditioning.

Amnesia can also result from circumscribed damage to structures of the medial diencephalon, including the mammillary nuclei, the dorsomedial nucleus of the thalamus, the anterior nucleus, the internal medullary lamina, and the mammillothalamic tract. Korsakoff's syndrome is the best studied example of diencephalic

amnesia. Patients with alcoholic Korsakoff's syndrome typically have frontal lobe pathology in addition to diencephalic damage. Frontal lobe pathology produces a pattern of cognitive impairment that is dissociable from amnesia itself. In the case of the patient with Korsakoff's syndrome, frontal lobe pathology is superimposed on severe memory impairment (Table 3.4-1).

Test	Amnesia	Korsakoff's Syndrome	Frontal Lobe Damage
Delayed recall	+	+	–
Dementia rating scale: memory index	+	+	–
Dementia rating scale initiation/ perseveration index	–	+	+
Wisconsin Card-Sorting Test	–	+	+
Temporal order memory	+	+ +	+ +
Metamemory	–	+	+
Release from proactive interference	–	+	–

Table 3.4-1 Associated and Dissociated Deficits in Amnesia

One limitation of conventional methods for assessing neuropathology is that remote functional damage may be overlooked. For example, standard MRI scans may show structural damage limited to a particular region, but this damage may lead indirectly to disrupted functioning in other regions. Accordingly, functional neuroimaging may be useful for characterizing more fully the neural dysfunction responsible for amnesia. In Korsakoff's syndrome, results from positron emission tomography (PET) have revealed functional damage in widespread cortical regions (Fig. 3.4-5). Accordingly, diencephalic amnesia may often reflect a disruption of thalamocortical connections that are critical for memory storage.

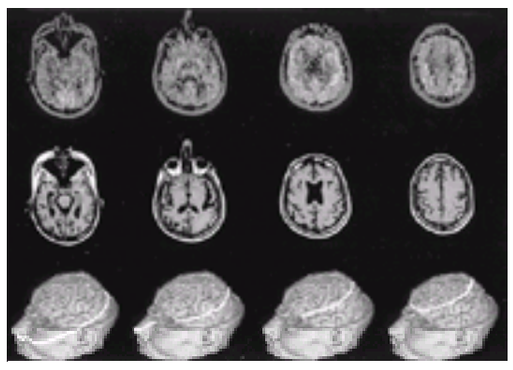

FIGURE 3.4-5 PET and MRI scans in a patient with Korsakoff's syndrome. Neural dysfunction was evident as reduced glucose utilization in multiple cortical regions in the frontal and parietal lobes, and in the cingulate. Functional neuroimaging can reveal brain dysfunction that might otherwise not be evident if limited to structural neuroimaging results. In Korsakoff's syndrome, the memory impairment probably reflects a disruption of thalamocortical circuitry. (Reprinted with permission from Paller KA, Acharya A, Richardson BC, Plaisant O, Shimamura AP, Reed BR, Jagust WJ: Functional neuroimaging of cortical dysfunction in alcoholic Korsakoff's syndrome. J Cogn Neurosci *9:*277, 1997.) (See Color Plate 5.)

Although amnesia can result from damage to either the medial temporal lobe or to the diencephalon, the distinctive functions of these two regions have been difficult to elucidate. It may be reasonable to expect the medial temporal lobe and diencephalic brain regions to make different contributions to normal memory, but there is currently no compelling evidence for a corresponding qualitative difference in memory impairment. This could be because the two regions function together as one system that facilitates the formation of links between neocortical storage sites, or because the two regions function separately but each makes an essential contribution to linking neocortical storage sites. In any event, memory clearly relies on an elaborate complex of neural circuits extending across multiple brain areas; Figure 3.4-6 shows the chief components of this circuitry. Ongoing research continues to improve our understanding of the neuroanatomy of amnesia and the normal functions of this neural circuitry.

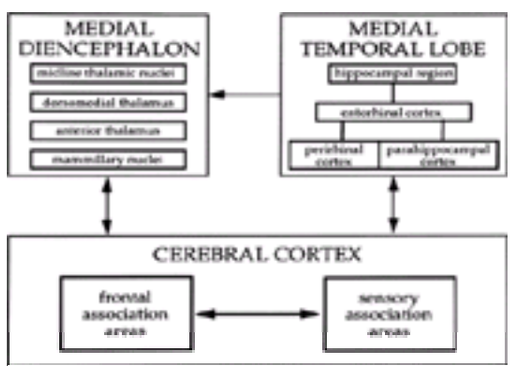

FIGURE 3.4-6 A schematic view of some of the chief brain regions critical for declarative memory. The entorhinal cortex is the major source of projections to the hippocampus, and nearly two thirds of the cortical input to the entorhinal cortex originates in the perirhinal and parahippocampal cortex. Entorhinal cortex also receives direct connections from cingulate, insula, orbitofrontal, and superior temporal cortex.

Retrograde Amnesia Memory loss in amnesia typically affects recent memories more than remote memories (Fig. 3.4-7). Temporally graded amnesia has been demonstrated retrospectively in studies of amnesic patients and prospectively in studies of monkeys, rats, mice, and rabbits. These findings have important implications for understanding the nature of the memory storage process. Memories are dynamic, not static. Apparently, memory storage can become more robust over time. As time passes after learning, some memories are forgotten while others become stronger because of a process of consolidation that depends on cortical, limbic, and diencephalic structures. The limbic-diencephalic contribution diminishes over time such that the neocortical component of the memory eventually becomes self-sufficient. In other words, the limbic-diencephalic structures are needed at the time of learning and during this gradual process. After sufficient time has elapsed, long-term memories can be retrieved whether or not limbic-diencephalic structures are intact. Thus, the permanent repositories of memory are the distributed neocortical regions, not diencephalic or hippocampal regions.

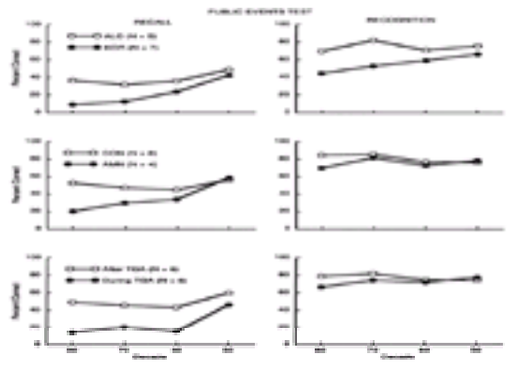

FIGURE 3.4-7 Remote memory performance of amnesic patients with Korsakoff's syndrome (KOR), alcoholic control subjects (ALC), amnesic patients with confirmed or suspected damage to the hippocampal formation (AMN), healthy control subjects (CON), and patients with transient global amnesia (TGA). The left column shows recall scores for past public events that had occurred in one of the four decades from 1950 to 1985. The right column shows performance on a multiple-choice test (four alternatives) involving the same public events. (Reprinted with permission from Kritchevsky M, Squire LR: Transient global ischemia: Evidence for extensive, temporally graded retrograde amnesia. Neurology *39:*213, 1989 and Squire LR, Haist F, Shimamura AP: The neurology of memory: Quantitative assessment of retrograde amnesia in two groups of amnesic patients. J Neurosci *9:*828, 1989.)

Atypical patterns of amnesic impairment have also been reported. Patients have been described with substantial retrograde impairments together with little or no impairment in new learning ability, a pattern termed *focal retrograde amnesia*. Evidence that focal retrograde amnesia can result from damage to portions of the anterior temporal lobes or possibly other neocortical areas has important implications for specifying the neuroanatomical substrates of memory. A likely explanation is that these neocortical areas comprise memory storage sites responsible for critical aspects of certain memories for facts and events. In some cases, retrograde deficits have been reported to span exceedingly long time periods, perhaps most of a patient's life. It thus appears possible to disrupt memory for events that occurred many years prior to the brain insult while retaining the ability to form memories for new events. The minimal problems with new learning in focal retrograde amnesia presumably relate to continued functioning in hippocampal and diencephalic regions as well as the availability of storage sites in intact parts of the cerebral cortex.

MEMORY SYSTEMS

Many kinds of memory remain intact in amnesia. The kind that is impaired is termed *declarative memory*. Declarative memory makes possible the conscious recollection of facts and events. A deficit in declarative memory presents itself as a global disorder of impaired memory for routes, lists, faces, melodies, objects, and other verbal and nonverbal material, regardless of the sensory modality in which the material is presented. This amnesic impairment typically occurs in conjunction with the preservation of a heterogeneous set of other memory capabilities, collectively termed *nondeclarative memory*. Nondeclarative memory includes skill learning, habit learning, conditioning, and the phenomenon of priming. Ample evidence has shown that, for these types of memory, amnesic patients can perform normally. A variety of skills, including perceptual, perceptuomotor, and cognitive skills, can be acquired by amnesic patients and normal subjects at equivalent rates. For example, amnesic patients can learn to read mirror-reversed text normally, they exhibit the normal facilitation in reading speed with successive readings of normal prose, and they improve as rapidly as normal subjects at speeded reading of repeating nonwords. In addition, amnesic patients can, after seeing strings of letters generated by a finite-state rule system, classify novel strings of letters as rule-based or not rule-based. Classification performance is normal despite the fact that amnesic patients are impaired at remembering the previously studied items or the events of training, which rely on declarative memory. Amnesic patients can also learn about categories by abstracting prototype information, even though they forget the examples from which the category was built.

Priming refers to a facilitation of the ability to detect or identify stimuli based on recent experience with those stimuli. Priming is another type of memory that occurs at full strength in amnesia. In one type of priming test, amnesic patients exhibited the normal tendency to complete three-letter stems with previously encountered words when they were instructed to produce the first word to come to mind (e.g., MOT__ completed to MOTEL). In another test, amnesic patients studied a list of words and then exhibited preserved priming with respect to their accuracy at identifying briefly presented words and with respect to their estimates of how long each word was presented (Fig. 3.4-8). Yet patients' ability to recognize the same words when presented later was severely impaired. Amnesic patients also exhibited priming for object names, novel objects, and novel faces. For example, in one experiment patients named pictures of previously presented objects reliably faster than they named pictures of new objects, even after a delay of a week. This facilitation occurred at normal levels, despite the fact that amnesic patients were markedly impaired at recognizing which pictures had been presented previously.

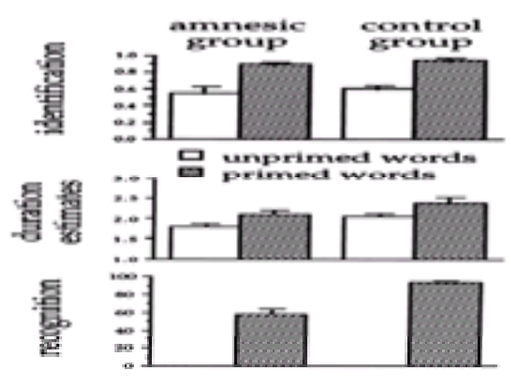

FIGURE 3.4-8 Preserved priming and impaired recognition in amnesic patients. The top graph shows the proportion of briefly presented words identified correctly by a group of amnesic patients and an age-matched control group. The priming effect took the form of more accurate identification for words presented in an earlier study phase (primed words) compared to other words (unprimed words). In the middle panel, both groups showed priming of duration estimations made on an arbitrary four-point scale, in that primed words were rated as having been presented for a longer duration compared to unprimed words. In the bottom panel, recognition performance on a forced-choice, three-alternative recognition test was impaired in the amnesic group (lower percent correct). Values are means plus standard errors of the mean. (Adapted from Paller KA, Mayes AR, McDermott M, Pickering AD, Meudell PR: Indirect measures of memory in a duration-judgment task are normal in amnesic patients. Neuropsychologia *29:*1007, 1991.)

Results from PET studies in normal volunteers suggest that priming reflects changes in early stages of the cortical pathways that participate in perceptual processing (Fig. 3.4-9). With stem-completion priming, both PET and divided visual field studies have implicated visual processing systems in extrastriate cortex, especially in the right hemisphere. In contrast, conscious recollection of remembered words normally engages a larger set of brain areas. Neural correlates of this processing have been observed as blood flow changes in hippocampal, frontal, and other brain areas. This processing simultaneously produces brain electrical activity (event-related potentials) that can be measured at the scalp (Fig. 3.4-10). A role for the frontal lobes in declarative memory has been substantiated both in patients with frontal lesions and in PET studies conducted on normal subjects. Patients with frontal lesions are not globally amnesic but tend to have certain memory problems.

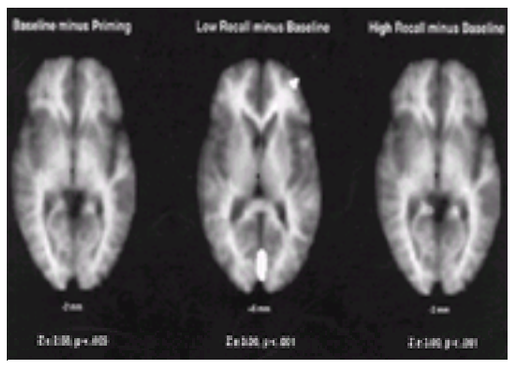

FIGURE 3.4-9 PET activations superimposed over averaged transverse MRI scans with the distance shown representing the distance from the line connecting the anterior and posterior commissure. Words were studied under strong or weak learning conditions (high recall or low recall) and then both declarative memory (cued recall) and nondeclarative memory (priming) were tested. The Baseline minus Priming subtraction showed an area of decreased blood flow (green) in right visual association cortex thought to be related to the greater ease of processing primed words. The Low Recall minus Baseline subtraction showed an area of increased blood flow (yellow) in secondary visual cortex and left prefrontal cortex thought to be related to the effort involved in deliberate, effortful retrieval. The High Recall minus Baseline subtraction showed a region of increased blood flow (red) in bilateral hippocampal regions thought to be related to successful retrieval of recently acquired information. (Reprinted with permission from Schacter DL, Alpert NM, Savage CR, Rauch SL, Albert MS: Conscious recollection and the human hippocampal formation: Evidence from positron emission tomography. Proc Natl Acad Sci USA *93:*321, 1996. ©1996 National Academy of Sciences, U.S.A.) (See Color Plate 5.)

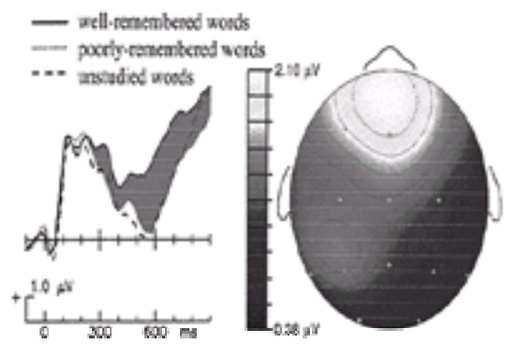

FIGURE 3.4-10 Brain potentials showing a differential response based on the extent to which subjects engaged in recollection following word presentations. Potentials shown at the left were recorded from a frontal scalp electrode. Measurements from multiple electrodes over the 400- to 800-ms latency range were used to generate the topographical map at the right, showing that the neural correlate of recollection was broadly distributed across the scalp, with largest responses over frontal cortex. (Adapted with permission from Paller KA, Kutas M: Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. J Cogn Neurosci *4:*375, 1992.) (See Color Plate 6.)

Patient B.G. suffered an infarction restricted to the right frontal lobe, and showed an abnormal tendency towards falsely recognizing as familiar items that had not been presented for study, possibly because of a tendency to rely on a general feeling of familiarity for items rather than specific memories for prior items. By virtue of its connections to posterior neocortical regions, the frontal lobe appears to play a role in the organization of retrieval. In other words, frontal-posterior networks are instrumental both in the on-line processing of information and in the retrieval of declarative memories.

One organizational scheme for categorizing multiple types of memory appears in Figure 3.4-11. Declarative memory depends on medial temporal and midline diencephalic structures along with extensive portions of the neocortex (see Fig. 3.4-6). This system provides for the rapid learning of facts (*semantic memory*) and events (*episodic memory*). Nondeclarative memory depends on several different brain systems. Habits probably depend on both the neocortex and the neostriatum, the cerebellum is important for conditioning of skeletal musculature, the amygdala for emotional learning, and the neocortex for priming. Declarative and nondeclarative memory differ in important ways; declarative memory is phylogenetically more recent than most types of nondeclarative memory. In addition, declarative memories are available to conscious recollection. The flexibility of declarative memory permits the retrieved information to be available to multiple response systems. Nondeclarative memory is inaccessible to awareness and tends to be inflexible, and it is expressed only by engaging specific processing systems. Nondeclarative memories are stored as changes within these processing systems, changes that are encapsulated such that other processing systems have limited access to the stored information.

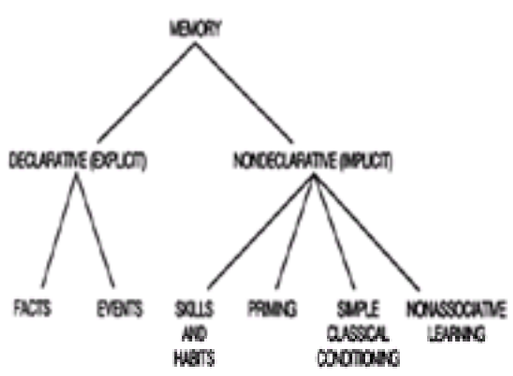

FIGURE 3.4-11 A tentative memory taxonomy. *Declarative memory* refers to conscious recollection of facts and events. *Nondeclarative memory* refers to a collection of abilities wherein performance changes as the result of experience but without affording conscious memory access to the original experience. *Nonassociative learning* includes habituation and sensitization. (Adapted from Squire LR, Zola-Morgan S: The medial temporal lobe memory system. Science *253:*1380, 1991.)

IMPLICATIONS

Current understanding of the biology of memory has significant implications for several fundamental issues in psychiatry. One example is the phenomenon of infantile amnesia, the apparent absence of conscious memory for experiences from approximately the first 3 years of life. Traditional views of infantile amnesia have emphasized repression (psychoanalytic theory) and retrieval failure (developmental psychology). A common assumption has been that adults retain memories of early events but cannot bring them into consciousness. However, another possibility is that the capacity for classical conditioning and skill learning (i.e., nondeclarative memory) emerges early in infancy, whereas declarative memory does not become available until about the third year; limbic and diencephalic structures essential for declarative memory may not be fully developed until that time. According to this view, infantile amnesia results not from the adult's failure to retrieve early memories but from the child's failure to store them in the first place. This is an intriguing suggestion, but recent evidence has not supported it. First, recall-like memory abilities have been demonstrated in young infants. In addition, it now appears that one of the most commonly employed tests of infant recognition memory, the visual-paired comparison task, depends on declarative memory.

What probably limits the development of declarative memory is not the slow maturation of the limbic-diencephalic structures essential for declarative memory, but rather the gradual development and maturation of the neocortex. As the neocortex develops, memories supported and stored there become more complex. Strategies emerge for organizing incoming information, language develops, and declarative memories become more richly encoded and interconnected, and more persistent.

The existence of multiple forms of memory and the gradual maturation of neocortex suggest an alternative to traditional views of infantile amnesia. It is not necessary to suppose that fully formed childhood memories persist but cannot be retrieved. An alternative view is that the capacity to store a viable declarative memory develops only gradually.

The existence of multiple memory systems also has implications for issues in psychoanalytic theory, including the construct of the unconscious. In considering the effects of past experience, it matters what view one takes of the nature of memory. By the traditional view, memory is a unitary faculty, and representations in memory vary mainly in strength and accessibility. According to this view, material that is unconscious is below some threshold of accessibility and could potentially be made available to consciousness. An alternative view begins with the distinction between memory that by its nature can be brought to mind (i.e., declarative memory) and other kinds of memory that are by their nature nonconscious in the sense that the knowledge is expressed through performance without affording any conscious memory content. In this view, early experience might affect later behavior, but the experience can persist to affect behavior without necessarily including a record of the event itself. Behavior can change as experience accumulates in altered dispositions, preferences, conditioned responses, habits, and skills, but these changes do not afford any actual awareness that behavior is being influenced by past experience. Nor is there any necessity that any particular past experience has been recorded. In this sense, the unconscious does not become conscious. Behavioral change can occur when new habits are acquired that supersede old ones or when sufficient awareness of a habit is gained such that it can to some extent be isolated or when the stimuli that elicit it can be processed differently. However, one does not become aware of its content in the same sense that one knows the content of a declarative memory, and one does not become aware of the early experiences that gave rise to the habit.

A better understanding of the biology of memory has also shed light on the imperfect nature of memory retrieval. It is quite possible to remember events that never happened. One way to conceptualize this problem is to consider the phenomenon of source amnesia, which occurs when an individual remembers some information without being able to remember how, where, or when the information was learned. If an individual imagines a certain event, at some later time it is possible to forget the moment when the scene was imagined while still recalling details as if they really happened. In this way, memory distortions can be caused by misattributions of source. Inferences and other thoughts invoked when recalling an experience can also be incorporated into memory such that a subsequent retrieval mistakenly includes the intervening information. Thus, the reconstructive nature of recollection poses problems for interpreting apparently “recovered” memories of traumatic events and emphasizes the importance of corroboration. As documented by experiments in adults and in children, illusory memories can be created, and children appear to be particularly susceptible to these effects. Several lines of evidence suggest that processing in the frontal lobes is essential for accurate source memory. For example, failures of source memory in the elderly appear to be associated with frontal lobe dysfunction. Patients with frontal lobe damage are highly susceptible to source errors, and they also tend to confabulate, sometimes producing bizarre responses that are inaccurate as well as implausible. Convergent results from PET studies in normal subjects imply that frontal regions, especially in the right hemisphere, play a key role in mediating accurate retrieval.

Finally, the biological understanding of declarative memory developed here is important in relation to the topic of consciousness. Declarative memory entails both a lasting influence of prior experience and an awareness of that remembering. The varieties of nondeclarative memory, such as priming, commonly occur without this awareness of remembering. To the extent that this contrast between memory with awareness versus memory without awareness can be understood at the neural level, a neuroscientific foundation can be developed to explain why certain types of neural processing lead to the subjective experience of remembering. Particular combinations of neocortical processing networks working in concert with the medial temporal lobe appear to come into play during the experience of remembering, whereas altered activity in a localized neocortical area can lead to the phenomenon of priming without explicit remembering. In this way, it may be possible to move from probing the neural substrates of declarative memory to addressing the question of why neural processing engenders consciousness at all, a problem that is central to the study of the neural basis of the mind.

ASSESSMENT OF MEMORY FUNCTIONS

A variety of quantitative methods are available to assess memory functions in neurological and psychiatric patients. Quantitative methods are useful for evaluating and following patients longitudinally as well as for carrying out a one-time examination to determine the status of memory function. If a memory problem is detected, it should be determined whether memory is selectively affected or whether memory problems are occurring against a background of intellectual deficits, as occurs in dementia. Although some widely available tests, such as the Wechsler Memory Scale-Revised, are useful measures of memory, most single tests assess memory rather narrowly, and even general-purpose neuropsychological batteries provide for only limited testing of memory functions. A complete assessment of memory usually involves a number of specialized tests that sample intellectual functions, new learning capacity, remote memory, and memory self-report.

The assessment of general intellectual functions is fundamental to any neuropsychological examination. In the case of memory testing, information about intellectual functions provides both information about a patient's general test-taking ability and a way to assess the selectivity of memory impairment. Useful tests include the Wechsler Adult Intelligence Scale-Revised, a test of object naming such as the Boston Naming Test, a rating scale to assess the possibility of global dementia, a test of word fluency, and specialized tests of frontal lobe function.

New Learning Capacity Memory tests are sensitive to impaired new learning ability when they adhere to either of two important principles. First, tests are sensitive to memory impairment when more information is presented than can be held in immediate memory. For example, one might ask patients to memorize a list of 10 faces, words, sentences, or digits because 10 items is more than can be held in mind. The paired-associated learning task is an especially sensitive test of this kind. In the paired-associate task, the examiner asks the patient to learn a list of unrelated pairs of words (e.g., queen-garden, office-river) and then to respond to the first word in each pair by recalling the second word.

Second, tests are sensitive to memory impairment when a delay, filled with distraction, is interposed between the learning phase and the test phase. In that case, examiners typically ask patients to learn a small amount of information and then distract them for several minutes by conversation, to prevent rehearsal. Recollection is then assessed for the previously presented material. Memory can be tested by unaided recall of previously studied material *(free recall)*, or by presenting a cue for the material to be remembered *(cued recall)*, or by testing recognition memory. In multiple-choice tests of recognition memory, the patient tries to select previously studied items from a group of studied and unstudied items. In yes/no recognition tests, patients see both studied and unstudied items one at a time and are asked to say “yes” if the item was presented previously and “no” if it was not. These various methods for assessing recently learned material can be ranked in terms of their sensitivity for detecting memory impairment (from most sensitive to least sensitive: free recall, cued recall, yes/no recognition, multiple-choice recognition). In practice, a cued-recall test can vary widely in its sensitivity, depending on the effectiveness of the cue.

The specialization of the cerebral hemispheres means that left and right unilateral damage to limbic-diencephalic structures is associated with different kinds of memory deficits. Accordingly, different kinds of memory tests must be used when unilateral damage is a possibility. Damage to limbic-diencephalic structures in the left cerebral hemisphere results in difficulty remembering verbal material such as word lists and stories. Damage to limbic-diencephalic structures in the right cerebral hemisphere impairs memory for faces, spatial arrangements, and other material that is typically encoded without verbal labels. For example, left medial temporal lobe damage impairs memory for both spoken and written text; right medial temporal lobe damage impairs the learning of spatial arrays, whether the layouts are examined visually or by touch. A useful way to test for anterograde amnesia for nonverbal material, for example, is to ask a patient to copy a complex geometric figure and then, without forewarning, to reproduce it after a delay of several minutes.

Remote Memory Disorders of memory are frequently accompanied by *retrograde amnesia*, that is, memory loss for events that occurred before the amnesia began. Evaluations of retrograde memory loss should attempt to determine both the severity of the loss and the time period that it covers. Most quantitative tests of remote memory are composed of material in the public domain that can be corroborated. For example, tests have involved questions about former one-season television programs, news events, or photographs of famous persons. An advantage of these methods is that it is possible to sample large numbers of events and target particular time periods. At the least, time periods before which the information could not have been learned can be identified. (For example, knowledge that Sarah Jane Moore tried to assassinate President Ford could not have been acquired prior to the 1970s). However, a disadvantage is that these tests are not particularly useful for detecting memory loss for information learned during the weeks or months immediately prior to the onset of amnesia. Most remote memory tests sample time periods rather coarsely and cannot detect a retrograde memory impairment that covers only a few months.

In contrast, autobiographical memory tests can potentially provide fine-grained information about a patient's retrograde memory. In the word-probe task, first used by Francis Galton in 1879, patients are asked to recollect specific episodes from their past in response to single word cues (e.g., *bird* and *ticket*) and to date the episodes. When normal subjects take the test, the number of episodes recalled is systematically related to the time period from which the episode is taken, and most of the memories come from recent time periods (the past 1 or 2 months). Patients with amnesia often exhibit temporally graded retrograde amnesia, drawing few episodic memories from the recent past, but producing as many remote autobiographical memories as normal subjects (Fig. 3.4-12).

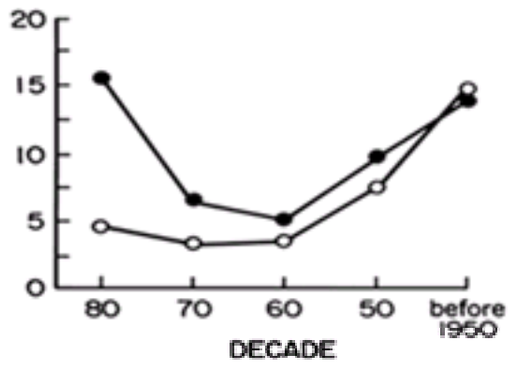

FIGURE 3.4-12 Loss of recent autobiographical memory as measured in five amnesic patients (filled circles) and five normal subjects (open circles). Well-formed memories from specific time periods were recalled in response to 75 single-word cues (e.g., *tree, flag, window*). (Reprinted with permission from MacKinnon D, Squire LR: Autobiographical memory in amnesia. Psychobiology *17:*247, 1989.)

Memory Self-Reports Patients can often supply descriptions of their memory problems that are extremely useful for understanding the nature of their impairment. Tests of the ability to judge one's own memory abilities are called *tests of metamemory*. Self-rating scales are available that yield both quantitative and qualitative information about memory impairment. As a result, it is possible to distinguish memory complaints associated with depression from memory complaints associated with amnesia. Patients with depression tend to rate their memory as poor in a rather undifferentiated way, endorsing equally all the items on a self-rating form. By contrast, patients with amnesia tend to endorse some items more than others; that is, there is a pattern to their memory complaints. Thus, patients with amnesia do not report difficulty in remembering very remote events or in following what is being said to them, but they do report having difficulty remembering an event a few minutes after it happens. Indeed, the self-reports match rather closely the description of memory dysfunction that emerges from objective tests. Specifically, new learning capacity is affected whereas immediate memory and very remote memory are intact. Some patients with amnesia, however, tend to markedly underestimate their memory impairment. In patients with Korsakoff's syndrome, for example, poor metamemory may stem from frontal lobe dysfunction. In any case, querying patients in some detail about their own sense of impairment and administering self-rating scales are valuable and informative adjuncts to more formal memory testing.

Dissociative Amnesia Differentiating dissociative amnesia from amnestic disorder, which results from neurological injury or disease, is less difficult than might be supposed. The two kinds of amnesia have markedly different characteristics. Dissociative amnesias typically do not affect new learning capacity. Patients enter the hospital able to maintain a continuing record of daily events. By contrast, new learning problems are at the core of amnestic disorder, which is much more common than dissociative amnesia. The main positive symptom in dissociative amnesia is extensive and severe retrograde amnesia. Patients may be unable to recall their own name or to recollect pertinent autobiographical information from childhood. However, patients with amnestic disorder never forget their name, and their remote memory for the events of childhood and adolescence is typically normal unless there is brain damage in the lateral temporal or frontal lobes.

Some patients with dissociative amnesia have circumscribed retrograde memory loss that covers a particular time period or that covers only autobiographical memories. One patient was reported to be able to answer questions about past public events but not questions about past personal events. Another patient scored close to zero on a test of famous photographs, far worse than any patient with neurological amnesia would score, and was also unable to identify proper names, such as *Los Angeles* and *Pontiac*. The challenge for the clinician lies not in distinguishing dissociative amnesia from amnestic disorder, but in distinguishing dissociative amnesia from malingering. Indeed, the diagnosis of dissociative amnesia can be difficult to substantiate and may be met with skepticism by hospital staff. Often, the clinical picture remains unclear until the amnesia clears. In some cases dissociative amnesia has been observed to clear after a period of days, but in other cases it has persisted as a potentially permanent feature of the personality.

SUGGESTED CROSS-REFERENCES

Functional neuroanatomy is discussed in Section 1.2; delirium, dementia, and amnestic and other cognitive disorders is discussed in Chapter 10; dissociative amnesia is discussed in Section 18.1; broad issue of neuropsychological and intellectual assessment of cognitive functions is covered in Section 7.4, Section 7.5 and Section 7.6; false memory syndrome is discussed in Section 3.4.

SECTION REFERENCES

Baddeley AD, Wilson FA, Watts N, editors: *Handbook of Memory Disorders*. Wiley, Chichester, UK, 1995.

*Bailey CH, Bartsch D, Kandel ER: Toward a molecular definition of long-term memory storage. Proc Natl Acad Sci USA *93:*13445, 1996.

Buckner RL, Tulving E: Neuroimaging studies of memory: Theory and recent PET results. In *Handbook of Neuropsychology*, vol 10, F Boller, J Grafman, editors, Elsevier, Amsterdam, 1995.

Cabeza R, Nyberg L: Imaging cognition: An empirical review of PET studies with normal subjects. J Cognitive Neurosci *9:*1, 1997.

Desimone R: Neural mechanisms for visual memory and their role in attention. Proc Natl Acad Sci USA *93:*13494, 1996.

Fuster JM: *Memory in the Cerebral Cortex: An Empirical Approach to Neural Networks in the Human and Nonhuman Primate*. MIT Press, Cambridge, MA, 1995.

Heiss W-D, Pawlik G, Holthoff V, Kessler J, Szelies B: PET correlates of normal and impaired memory functions. Cerebrovascular and Brain Metabolism Reviews *4:*1, 1992.

Kapur N: Focal retrograde amnesia in neurological disease: A critical review. Cortex *29:*217, 1993.

Kopelman MD: The Korsakoff syndrome. Br J Psychiatry *166:*154, 1995.

Lashley KS: *Brain Mechanisms and Intelligence: A Quantitative Study of Injuries to the Brain*. Chicago University Press, Chicago, 1929.

LeDoux JE: *The Emotional Brain*. Simon & Schuster, New York, 1996.

Lezak MD: *Neuropsychological Assessment*, ed 2. Oxford University Press, New York, 1983.

Mayford M, Abel T, Kandel RT: Transgenic approaches to cognition. Curr Opin Neurobiol *5:*141, 1995.

Mishkin M: A memory system in the monkey. Philos Trans R Soc Lond B Biol Sci *298:*83, 1982.

Paller KA: Consolidating dispersed neocortical memories. The missing link in amnesia. Memory *5:*73, 1997.

*Ranganath C, Paller KA: Frontal brain potentials during recognition are modulated by requirements to retrieve perceptual detail. Neuron *22:*605, 1999.

Raymond JL, Lisberger SG, Mauk MD: The cerebellum: A neuronal learning machine? Science *272:*1126, 1996.

Rugg MD: Event-related potential studies of human memory. In *The Cognitive Neurosciences*, MS Gazzaniga, editor. MIT Press, Cambridge, MA, 1995.

Schacter DL, Tulving E, editors: *Memory Systems 1994*. MIT Press, Cambridge, MA, 1994.

*Schacter DL, editor: Memory distortion: How minds, brains, and societies reconstruct the past. Harvard University Press, Cambridge, MA, 1995.

Shimamura AP: Memory and frontal lobe function. In *The Cognitive Neurosciences*, MS Gazzaniga, editor. MIT Press, Cambridge, MA, 1995.

*Squire LR: *Memory and Brain*. Oxford University Press, New York, 1987.

Squire LR: Memory and the hippocampus: A synthesis of findings with rats, monkeys, and humans. Psychol Rev *99:*195, 1992.

Squire LR, Butters N, editors: *Neuropsychology of Memory*, ed 2. Guilford, New York, 1992.

*Squire LR, Kandel ER: *Memory: From Mind to Molecule*. Freeman, New York, 1999.

Squire LR, Knowlton B, Musen G: The structure and organization of memory. Annu Rev Psychol *44:*453, 1993.

*Squire LR, Zola SM: Structure and function of declarative and nondeclarative memory systems. Proc Natl Acad Sci USA *93:*13515, 1996.

Squire LR, Zouzounis JA: Self-ratings of memory dysfunction: Different findings in depression and amnesia. J Clin Exp Neuropsychol *10:*727, 1988.

Thompson RF, Kim JJ: Memory systems in the brain and localization of a memory. Proc Natl Acad Sci USA *93:*13438, 1996.

*Ungerleider LG: Functional brain imaging studies of cortical mechanisms for memory. Science *270:*769, 1995.

Weiler IJ, Hawrylak N, Greenough WT: Morphogenesis in memory formation: Synaptic and cellular mechanisms. Behav Brain Res *66:*1, 1995.

Zola-Morgan S, Squire LR: The primate hippocampal formation: Evidence for a time-limited role in memory storage. Science *250:*288, 1990.

3.5 BRAIN MODELS OF MIND

KARL H. PRIBRAM, M.D., FH.D.

For two centuries brain models of mind have fascinated scientist and the lay public alike. This intense interest began with Francis J. Gall's pioneering correlations between brain pathology and characteristic personality histories of patients. As with every major advance in understanding the mind-brain relationship, Gall's demonstrations became a popular fad in the form of reading bumps on the skull, *phrenology*. Today a similar fad is evident in the application of the findings regarding hemispheric specialization: educators and politicians alike recommend using the right brain more lest the human race fall forever into damnation.

Brain models of mind have shown a remarkable coherence over the nineteenth and twentieth centuries despite the often acrimonious emphasis on this or that phenomenon to the exclusion of a comprehensive analysis. Further, when carefully considered, each of the often opposing views captures important aspects of the issues and that reconciliation devolves on making distinctive definitions and reading the proposals in their original form with these definitions in mind.

One definition of mind was provided by Gilbert Ryle: Mind comes from minding, paying attention. In old English the word is *gemynd*, akin to *remind*, which was derived from terms that meant *to warn* and *to intend*. The Sanskrit word *mynas* means to think.

> As a whole, the human brain is critical to minding; one case history highlights the obvious. A 14-year-old-girl who had fallen out of a rapidly moving automobile had sustained a head injury with multiple scalp lacerations. Transporting her to a hospital several hundred miles away was thought to be too risky to an already traumatized head. Her head was swathed in bandages through which some blood had oozed, making them appear bright red; in contrast, the girl looked green. When addressed: “Hello Cathy, you look like a Christmas package all dolled up in your bandages,” the girl smiled and said, “Hello, doctor.” It was evident that this girl's brain was intact. She minded, and even had a sense of humor. A thorough examination revealed a broken rib with a puncture to one lung, thus her green color. Bandages and a brief time in an oxygen tent quickly allowed healing to commence.

The diagnosis rested on the truism that scrambled brains result in scrambled minds. However, because of its pervasive validity, this truism can blind us to the more subtle aspects of the mind-brain relationship. For instance, the close association of mind to brain might lead us to suspect uncritically that mind and brain are the same, which would be as absurd as stating that the islands of Langerhans of the pancreas are the same as insulin regulation of glucose metabolism. Minding is a function of the entire organism interacting with its environment (just as glucose metabolism is a function of the organism metabolizing environmentally derived nutrients). What is common to brain and mind is their organization, much as what is common to a computer's hardware and the various levels of programming software is the *in*formation (the form within) being processed.

Thus, although the special relation between brain and conscious experience is widely acknowledged, the subtleties inherent in the nature of the relation remain debatable. In this respect apparently no progress has been made in the past two millennia.

The time is ripe for an advance in understanding. Each of the philosophical stances toward the mind-brain relationship has merit as long as it is restricted to the database that defines the stance. The set of problems that characterize the special relation between brain and the variety of mental processes is closely related and the mind-brain analysis must be anchored in an ontological neutral monism. What is ontologically neutral to the material brain and mental (psychological) processes is *order*—order as measured scientifically in terms of energy, entropy, and information.

With respect to the special relation between brain and the variety of mental processes, this ontological neutrality is expressed by showing that conscious (and unconscious) processes are coordinate with identifiable brain processes occurring in identifiable brain systems, that is, at some level the descriptions of brain processes and descriptions of mental processes become homomorphic.

An example from computer science illustrates what is meant by *homomorphic:* the computer is used as a word processor when English words and sentences are typed into it. The word processing system, by virtue of an operating system converts the keyboard input to binary, which is the language of the computer. Nothing in the description of English and of binary machine language appears to be similar, yet by virtue of the various transformations produced in the encoding and decoding operations of the various stages leading from typescript to binary, the information in the typescript is preserved in the binary language of the operation of the computing machine.

In a similar fashion, little in conscious experience resembles the operations of the neural apparatus with which it has such a special relation. However, when the various transformations, the transfer functions, the codes that intervene between experience and neural operations are sufficiently detailed, a level of description is reached in which the *transformations* of experience are homomorphic with the language used by the brain. This language is the language of the operations of a microprocess taking place in synaptodendritic fields, a mathematical language similar to that which describes processes in microphysics that is, subatomic physics.

At this microprocessing level an identity describes the relation between brain and mental processes. At more remote processing levels, encompassing larger event structures (assemblers, operating systems, or their counterparts in brain systems), pluralism, and eventually, at the level of natural language, dualism characterizes the relationship. The special relation between brain and mental processes is thus not identical, except in implementation at the microprocessing level. At the neuronal and even at the neural system level several types of relationship with psychological processes can be discerned.

First, there are neurochemical states operating in the synaptodendritic processing web that determine states of consciousness. The very active field of psychoneuropharmacology is replete with evidence of relations between catechol and indole amines acting in specified brain locations to produce states of consciousness such as wakefulness and sleep, depression, and elation, and perhaps even dissociated states such as those seen in schizophrenia. The relations between relative concentrations of blood glucose and osmolarity and hunger and thirst; between sex hormones and sexually characteristic behaviors; and between peptides such as the endorphins and enkephalins and the experiences of pain and stress are all well documented.

Second, there are detailed descriptions of the relations between the sensory systems of the brain and the sensory aspects of perception: the contents of consciousness.

States of consciousness often determine contents and as often, are determined by them. When hungry one tends to notice restaurant signs; walking past the fresh aromas emanating from a bakery whets the appetite. This connection between states and the contents of consciousness is mediated by the process ordinarily called *attention* (the control of sensory input), by *intention* (the control of motor output), and *thought* (the control of remembering). The understanding of these processes of minding is critical to understanding the special relation between brain states and the contents of conscious experience.

VARIETIES OF BRAIN ORGANIZATION

Localization and Distribution of Function Some models of brain organization are crucial for determining the organization of minding. First is the issue of localization of function. Francis Gall brought this issue to the foreground by correlating different local brain pathologies to the histories of the cadavers he autopsied. Although often wrong in detail, he was correct in the methods he carefully detailed. He was naïve in delineating the faculties of mind for which he sought localization, but systematic classification of mental functions continues to be elusive despite a half-century of operational behaviorism. Today, it is popular to discuss the modularity of mind and component systems of the brain and relate them both in the clinic and in the laboratory by crafting experimental designs and behavioral and verbal testing procedures. The use of these techniques traces its heritage directly to Gall's enterprise.

The excesses of phrenology raised the question of which brain system brought the various faculties together into a conscious self. The unity of being, the soul of mankind, was challenged when mentation was subdivided into a mere collection of faculties. Furthermore, experimental evidence accrued to demonstrate a relation between impairments in complex behaviors and verbally reported experiences and the amount of brain tissue destroyed irrespective of location. In the recent past, Karl Lashley has been an exponent of this mass action view.

However, in a letter to Fred Mettler, Lashley once stated his exasperation at being misinterpreted: “Of course I know the front of the brain does something different from the back. The visual sensory input terminates in the occipital lobes. Electrical stimulations of the pre-Rolandic areas elicit movements and the front parts are more enigmatic in their functions. But this is not the issue.” Elsewhere he states the issue clearly: “. . . certain coordinated activities, known to be dependent upon definite cortical areas, can be carried out by any part (within undefined limits) of the whole area.”

What Lashley emphasized was that certain selected mental functions appeared to be related to brain processes that are distributed. For instance, he pointed out that sensory and motor equivalences could not be accounted for even by a duplication of brain pathways: “Once an associated reaction has been established (e.g., a positive reaction to a visual pattern), the same reaction will be elicited by the excitation of sensory cells which were never stimulated in that way during training. Similarly, motor acts (e.g., opening a latch box) once acquired, may be executed immediately with motor organs which were not associated with the act during training.”

The following is example of motor equivalence: a dog was conditioned to raise his right hind leg to the sound of a tone. After this conditioned response was well established, his right motor cortex (which controls the left side of the body) was exposed. Then during the performance of the conditioned reaction a patty of strychninized filter paper (which chemically excited the cortical tissue) was placed on the area that controls the left forepaw. Immediately the dog switched the responding leg: he now raised his left forepaw to the conditioned signal. A temporary dominant focus of excitation had been established in the cortex by the chemical stimulation. E. Roy John summarizes the experiments that demonstrate such shifts in cerebral dominant foci in Figure 3.5-1.

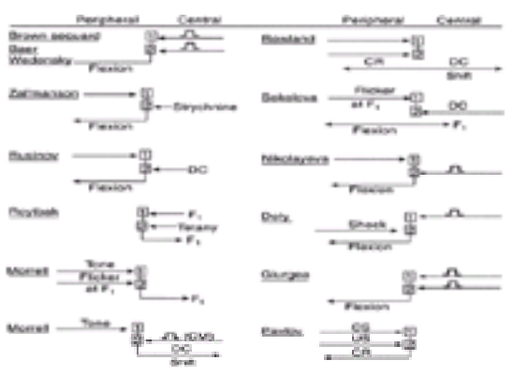

FIGURE 3.5-1 Methods of conditioning that have been used by various investigators to establish and produce shifts in cerebral dominant foci. The example in the text refers to Zal'manson's experiment. (Reprinted with permission from Pribram K: *Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology*. Random House, New York, 1971.)

The distributed aspect of brain function becomes most evident in memory storage. Even after large deletions of brain tissue such as those resulting from strokes or tumor resections, specific memories, *engrams*, are seldom lost. When amnesias do occur they are apt to be spotty and difficult to classify, which suggests that memory is stored in a distributed fashion. The storage process dismembers the input, which is then re-membered on occasions that necessitate recognition and recall. In contrast to storage, the retrieval processes are localized, at least within systems such as those that are sensory specific. When such systems are damaged, sensory-specific and even category-specific agnosias may result. Thus with regard to memory, both distributed and localized processes can be identified depending on which property of the process is being considered. This principle of analyzing a mental process to identify specific aspects will be useful in other contexts as well.

Systems in the Control of Attention and Intention Brain systems serve as controls on the processes intrinsic to minding in detail. William James noted that the delineation of minding, or consciousness, devolves on processes usually referred to as *attention* and *intention or volition*. Controls on attention determine the span of sensory processing, those on intention determine the span over which action becomes effective, and controls on thought determine the span of memories being considered.

Two decades of investigation into the neural processes involved in the control of attention discerned three such mechanisms: one deals with short phasic response to an input *(arousal)*, a second relates to prolonged tonic readiness of the organism to respond selectively *(activation)*, and a third *(effort)* acts to coordinate the phasic *(arousal)* and tonic *(activation)* processes. Separate neural and neurochemical systems are involved in the phasic (arousal) and tonic (activation) processes: the phasic process centers on the amygdala and the tonic process centers on the basal ganglia of the forebrain. The coordinating system (effort) critically involves the hippocampus, a phylogenetically ancient part of the neural apparatus.

Evidence from the analysis of changes in the electrical activity of the brain evoked by brief sensory stimulation has shown that the arousal and activation systems operate on a more basic process centered on the dorsal thalamus, the waystation of sensory input to the cerebral cortex. Brain electrical activity evoked by sensory stimulation can be analyzed into components. Early components reflect processing via systems that directly (via the thalamus) connect sensory surfaces with cortical surfaces. Later components reflect processes initiated in the thalamocortical and related basal ganglia systems that operate downward onto the brainstem (tectal region), which in turn influence a thalamic “gate” that modulates activity in the direct sensory pathways. It is the activity reflected in these later components of the brain electrical activity that constitutes activation. The thalamic gate is also regulated by input from the system centered on the amygdala—the arousal system. When stimulated, this system produces an effect on the “gate” opposite to that of the activation system.

Evidence also indicates that the coordination of phasic (arousal) and tonic (activation) attentional processes often demands effort. When attention must be paid, the hippocampal system becomes involved and influences the arousal system rostrally through frontal connections with the amygdala system and influences the activation system caudally via connections in the brainstem. At this juncture the relation of attention to intention, that is, to volition or will, comes into focus. Again, William James had already pointed out that a good deal of what is called voluntary effort is the maintaining of attention or the repeated returning of attention to a problem until it yields solution.

William James had apposed will to emotion and motivation (which he called *instinct*). Beginning with Walter Cannon's experimentally based critique of James, followed by Karl Lashley's critique of Cannon, to the anatomically based suggestions of James Papez and their more current versions by Paul MacLean, brain scientists have been deeply concerned with the processes that organize emotional and motivational experience and expression. Two major discoveries have placed the earlier more speculative accounts into better perspective. One such discovery has been of the role of the reticular formation of the brainstem and its chemical systems of brain amines that regulate states of alertness and mood. Donald Lindsley proposed an activation mechanism of emotion and motivation on the basis of the initial discovery and has more recently detailed the pathways by which such activation can exert control over brain processes. The other discovery is of the system of brain tracts that when electrically excited results in *reinforcement* (i.e., an increase in the probability of recurrence of the behavior that has produced the electrical brain stimulation) or *deterrence* (i.e., a decrease in the probability that such behavior will recur) by James Olds and Peter Milner.

To organize these discoveries and other data that relate brain mechanisms to emotion, it is necessary to distinguish clearly between data that refer to experience (feelings) and those that refer to expression, and further to distinguish emotion from motivation. Thus, feelings were found to encompass both emotional and motivational experience, emotional as affective and motivation as centered on a readiness processes. Not surprisingly, the affective process of emotion was found to be based on the process of arousal, the ability to make phasic responses to input that “stop” the motivational processes of activation that maintain selective readiness. Thus, feelings were found to be based on neurochemical states (predispositions or moods) that become organized by appetitive (motivation, “go”) and affective (emotional, “stop”) processes. Feelings of effort often are experienced as anxiety.

A wealth of new data had spawned these insights and made it fruitful to reexamine the Jamesian position with regard to his visceral theory of emotions. James is

almost universally misinterpreted as holding a peripheral theory of emotion and mind. Through his writings he emphasizes the effect that peripheral stimuli (including those of visceral origin) exert on brain processes. Nowhere, however, does he identify emotions with bodily processes; emotions are always the resultant effect on brain states. What James failed to take into account is the role of expectations (the representational role of the organization of familiarity and novelty) in the organization of emotions. It is these “neuronal models” of prior experience that were found to entail the functions of the hippocampus and of the basal ganglia, including the amygdala.

Nonetheless, James is explicit when he discusses the nature of the input to the brain from the viscera. He points out two possibilities: emotions are processed by a separate brain system or they are processed by the same systems as are perceptions. Both possibilities have been realized: parts of the frontolimbic forebrain (especially the amygdala and related systems) process visceroautonomic bodily inputs, and the results of processing become distributed via brainstem systems that diffusely influence the perceptual systems. Additionally, James clearly defines the difference between emotions and motivations (which he calls instincts): emotional processes take place primarily within the organism whereas motivations reach beyond into the organism's environment. James perhaps overemphasized the visceral determination of emotional experience, but he did occasionally include attitudinal factors as depending on sensory feedback from the somatic musculature.

The distinction between the brain mechanisms of motivation and will are less clearly enunciated by James. He grapples with the problem and establishes the questions that must be answered. These questions remained unanswered until the late 1960s when several theorists began to point out the difference between feedback, homeostatic processes on the one hand and programs, which are feedforward, homeorhetic processes, on the other. Feedback mechanisms depend on error processing and are therefore sensitive to perturbations; programs, unless completely stopped, run themselves off to completion irrespective of obstacles placed in their way.

Clinical neurology had classically distinguished the mechanisms involved in voluntary behavior from those involved in involuntary behavior. The distinction rests on the observation that lesions of the cerebellar hemispheres impair intentional (voluntary) behavior whereas basal ganglia lesions result in disturbances of involuntary movements. Damage to the cerebellar circuits are involved in a feedforward rather than a feedback mechanism. Recent microelectrode analyses suggest that the cerebellar hemispheres perform calculations in fast-time, (i.e., they extrapolate where a particular movement would end were it to be continued, and send the results of such a calculation to the cerebral motor cortex where they can be compared with the target to which the movement is directed). Experimental analysis of the functions of the motor cortex had shown that such targets are composed of “Images of Achievement” constructed in part on the basis of past experience.

Just as the cerebellar circuit has been shown to serve intentional behavior, the basal ganglia have been shown to be important to involuntary processes. These structures are also involved in the control of activation, the readiness of organisms to respond. Lesions in the basal ganglia grossly amplify tremors at rest and markedly restrict expressions of motivational feelings. Neurological theory has long held that these disturbances result from interference by the lesion of the normal feedback relationships between basal ganglia and cerebral cortex. Surgical removal of motor cortex has been performed on patients with basal ganglia lesions in order to redress the imbalance produced by the initial lesions; such resections have proved remarkably successful in alleviating the often distressing continuing disturbances of involuntary movement that characterize basal ganglia diseases.

Massively Parallel Distributed Processes Two closely related issues concerning the organization of brain function are often confounded: (1) localization versus distribution of function within each system and (2) whether processing proceeds among different localizable systems in a hierarchical fashion or whether processing proceeds in parallel and thus heterarchically.

The fact that a temporary dominant focus in the cerebral cortex can take control of the expression of a learned behavior indicates that hierarchical control operates in the central nervous system. Equally persuasive is the evidence for control over spinal cord activity by the brainstem and forebrain. Neuronal activity in the spinal cord displays an extremely high rate of spontaneous impulse generation. These generators are modulated by inhibitory local circuit neurons in such a way that the resultant activity can be modeled in terms of coupled ensembles of limit cycle oscillatory processes. In turn, these ensembles of oscillators become organized by brainstem systems that consist of cholinergic and adrenergic neurons. The cholinergic set regulates the frequency of a wide range of tonic rhythmic activities such as those involved in locomotion, respiration, cardiovascular responses, and sleep. This cholinergic system is coupled to an adrenergic set of neurons that segment the rhythmic activities into episodes. Both systems are subject to further hierarchical control by the dopaminergic system of the basal ganglia. Clinically, loss of this hierarchical control is expressed as an exaggeration of the normally present, almost subliminal tremors that under extreme conditions lead to spastic paralysis, hyperreflexia, and uncontrollable fits of oscillatory muscular spasm.

However, the evidence from the experiments that demonstrated temporary dominant foci can be viewed from the perspective: that the flexibility demonstrated by the shift from one controlling locus to another shows the organization of the cortical system to be heterarchial. Any locus within the system can become dominant if sufficiently excited. The following story, attributed to Warren McCulloch, illustrates the nature of heterarchial organization:

> After the battle of Jutland in which the British Navy took a beating, both the British and American navies reorganized to change from hierarchical to heterarchical control. Thus, battleships no longer had to await orders from a central command source to engage in defensive maneuvers. During World War II the Fifth Fleet was stationed in an only slightly dispersed mode of operation somewhere in the Pacific Ocean when it was attacked from two directions by separate air squadrons. Sightings of the attackers were made from different locations in the fleet by observers on the ships closest to one or the other of the attacking planes. The sailor who made the sighting became a dominant focus and his ship and those in his proximity took off to defend against the attackers. However, because the attack came from two different directions, two dominant foci were created, each commanding parts of the fleet to steam away in different directions. This left the ship at the center of the fleet that housed its admiral haplessly unprotected and, since no sightings were made by his ship, at a momentary loss as to what to do. Fortunately, both attacking squadrons were defeated and turned back without any damage accruing to the Fifth Fleet.

There is thus a possible penalty to be paid for the flexibility achieved by temporary dominance over processing as any person who has ever been of two minds knows well.

Ordinarily hierarchical control is conceived of as a serial process. This is because when control is direct, there is a causal connection between the controller and the controlled. Causality implies that the origination of the control signal precedes its effect on the system being controlled. Seriality remains when there are feedback loops. However, when feedforward operations are inserted into the process, seriality is no longer as clearcut. For example, lower the temperature or blood sugar on a thermostat or homeostat and the sensor responds, closes a circuit, and the effector responds; this is a serial process. Now place a control dial or other bias on the process and there are two or more ways for the sensor to become adjusted. The temperature falls, but because the heating bill was too high last month the dial is reset and warmer clothes are worn. There are parallel inputs to the sensor. Herman von Helmholtz is credited with pointing out that voluntary processes such as those by which we move our eyes are constituted of such parallel feedforward corollary discharges to the effectors. Control can be hierarchical yet dependent on a parallel process.

Processing in the cerebral cortex is massively parallel. Simulations of these parallel cortical processes have since the late 1980s become implemented on personal computers to such an extent that the endeavors have been dubbed a cottage industry. These simulations of neural networks are capable of pattern recognition, of language learning, and of decision-making that is remarkably true to life. Single-layered simulations have given way to three-layered computations that involve an input layer, an output layer, and a hidden layer. All the elements of the network are interconnected to one another. In several such simulations the input is fed forward through the net and the output is compared with one that is desired; the difference between the actual and the desired is fed back to the net. The process is repeated until the desired output is achieved. Variations on this theme abound, each variation being better adapted than its alternates for a particular purpose.

One of the most fascinating attributes of these neural networks is that the information contained in the input becomes fragmented and distributed in the elements of the layers. The simulations are therefore said to be massively *parallel distributed processes* (PDP), which makes them akin to optical information processing systems such as holography and tomography from which they were in fact derived.

Cerebral Dominance and the Unity of Consciousness Surrounding the major fissures of the primate cerebral cortex lie the terminations of the sensory and motor projection systems. These systems have been termed *extrinsic* because of their close ties by way of a few synapses with peripheral structures. The sensory surface and muscle arrangements are mapped more or less isomorphically onto the perifissural cortical surface by way of discrete, practically parallel lines of connecting fiber tracts. When a local injury occurs within these systems a sensory scotoma, or a scotoma of action, ensues. A *scotoma* is a spatially circumscribed hole in the field of interaction of organism and environment: a blind spot, a hearing defect limited to a frequency range, a location of the skin where tactile stimuli fail to be responded to. These are the systems where what Henry Head called *epicritic processing* takes place. These extrinsic sensory-motor projection systems are organized such that movement allows the organism to project the results of processing away from the sensory and muscular surfaces where the interactions take place, out into the world

external to the organism. Thus, processing within these extrinsic systems constructs an objective reality for the organism.

Between the perifissural extrinsic regions of cortex lie other regions of cortex variously named *association cortex, uncommitted cortex*, or *intrinsic cortex*. These names reflect the fact that there are few, if any, direct connections between peripheral structures and these regions of cortex that make up most of the convexity of the cerebrum. Thus, on the brain's convexity a three-tiered arrangement for these systems can be discerned. Each ajor sensory apparatus has a fairly direct input to areas in the cortex. Immediately surrounding these areas are others, which when electrically stimulated originate movements of the musculature associated with each of the sense organs (e.g., eye muscles for vision, ear muscles for hearing, and body muscles for somatic sensations). These areas are extrinsically connected to organs in the periphery of the body and therefore provide perspectives relating the body to the world beyond.

Surrounding these extrinsic areas are sensory-specific areas that are primarily connected intrinsically to other brain structures. These areas provide perspectives that are intrinsic to the entities perceived, perspectives such as those provided by color and object constancy. Finally, other areas operate on inputs from a variety of senses and relate their perspectives to each other. All these areas and the brain systems that they represent are involved in organizing phenomenal perceptions or the sensory-driven aspects of perception. Another set of systems, more noumenal in their function, are located frontally and on the limbic medial border of the brain's hemispheres.

Three important discoveries have fueled the current interest in hemispheric lateralization. One of these was actually a rediscovery during the latter half of the nineteenth century of the fact that the speech of most righthanded individuals is usually controlled by the left hemisphere. Hippocrates already knew this and may well have learned it from the Egyptians. Running from back to front, comprehension, grammar, and fluency (*semantics, syntactics*, and *pragmatics*) are affected by lesions centering on the sylvian fissure. However, dominance is not as complete in females as it is in males, nor is it as pervasive in cultures that do not use phonemic writing. It is now known that the non–speech-dominant hemisphere has its own characteristic modes of processing. With the left hemispheres of right-handed persons being taken over by an aural-oral dimension, the right hemisphere is left free to process visual-spatial relations.

A third and most pervasive and persistent focus of interest has been that of the unity of consciousness. When the corpus callosum was severed in patients who had suffered severe unilateral epileptic seizures in order to prevent involvement of the healthy hemisphere, testing revealed that what was sensed by the right hemisphere could only be expressed nonverbally by that hemisphere. The left verbal hemisphere appeared to be ignorant of what had transpired. It seemed as if consciousness had been split when the hemispheres were sundered. The assumption that there is ordinarily a unity to consciousness was bolstered precisely because this unity had been ruptured.

Taken together with the facts of hemispheric specialization and the “dominance” of the left hemisphere for language, these observations were broadened to the conception that human civilization suffered from left brain dominance and that training for greater brain balance would restore balance to civilization. However, innumerable studies have demonstrated that all but the most rudimentary processing involves both hemispheres. Even in language, the appreciation and expression of emotional communication involves the right hemisphere, and extreme specialization is limited to right-handed males raised in a phonemic literary environment.

Although the popular overgeneralization about hemispheric lateralization is to be deplored there was renewed interest in the question of whether consciousness could be divided. Sir John Eccles argued that consciousness is tied to language, an argument also made by Freud, and that therefore the right, speechless, hemisphere was to all intents and purposes essentially unconscious. However, the right hemisphere clearly communicated with left-handed, nonverbal instrumental responses that it had processed the input presented to it: the nonverbal hemisphere obviously had a mind of its own. Conscious minding is of two sorts: *instrumental* and *intentional*; thus, Eccles' proposal is tenable if what is meant is intentional consciousness. Brain facts as they relate to behavior, mind, and consciousness often spring surprises on the unwary.

VARIETIES OF CONSCIOUS EXPERIENCE

Cerebral Cortex and Reflective Consciousness The distinction between the systems that control intentional behavior and those that control involuntary behavior extends to the control of sensory input and the processing of memory. With regard to sensory input, the distinction between the contents of awareness and the person who is aware was delineated by Franz Brentano and called *intentional inexistence*. This dualism of a minding self and objective matter (e.g., brain) was already present in the writings of René Descartes. Although Cartesian dualism is perhaps the first overt nontrivial expression of the issue, the duality between subject and object and some causal connection between them is inherent in language once it emerges from simple naming to predication. John von Neumann and Julian Jaynes have suggested that a change in consciousness (i.e., in distinguishing an aware self from what the self is aware of) occurs somewhere during the eighth century BC between the time of the Iliad and the Odyssey, which links it to the invention and promulgation of phonemically based writing. Prehistory was transmitted orally and aurally; written history is visual and verbal. In an oral and aural culture a greater share of reality is carried in memory and is thus personal; once writing becomes a ready means of recording events, they become a part of extrapersonal objective reality. The shift described is especially manifest in a clearer externalization of the sources of conscience—the gods no longer speak personally to guide individual man.

This process of ever-clearer distinctions between personal and extrapersonal objective realities culminates in Brentano's intentional inexistence, which was shortened by Edmund Husserl to “intentionality.” It is this reading of the subject-object distinction that philosophers ordinarily mean when they speak of the difference between conscious and unconscious processes.

A few years ago during a seminar, the author noted that the left arm of a graduate student was moving somewhat awkwardly while arranging papers on a table in front of us. The author asked the student, Ms. C., if she was alright, while pointing to her left arm. She replied, “Oh, that's just Alice; she doesn't live here anymore.” At the end of the semester, Ms. C. presented a detailed account of her experiences with Alice.

Ms. C. experienced devastations to her locational *integrity*. Other patients, after injuries to their occipital lobes, demonstrate “blindsight,” the ability to visually identify objects in the “blind” field despite the fact that they fail to be consciously aware of these objects. Patients such as those who are blindsighted and Ms. C. who might be considered to have a tactile and kinesthetic blindsight both have damage to the cortex of the posterior convexity of their brains. Thus, they suffer disruption of their egocentric (essentially tactile and kinesthetic) and allocentric (essentially visual and auditory) organization. This is a disruption of “objective” awareness because it relates patients to their impairment as if it were a relationship among objects. The relationship is “intentional” in Brentano's sense of an ability to differentiate the perceiver from the perceived. Note also that the narrative abilities of such patients do not suffer.

In contrast is the case of a boy who was *unable* to recount his experiences. Thus, the case histories present two distinct modes of coping that are disrupted by injury to distinctly different brain systems: one articulates the organism in egocentric space and locates it allocentrically in its environment; the other evaluates and monitors experience.

According to Ms. C.:

I was doing laundry about midmorning when I had a migraine. I felt a sharp pain in my left temple and my left arm felt funny. I finished my laundry towards mid-afternoon and called my neurologist. He told me to go to the emergency room. I packed a few things and drove about 85 miles to the hospital where he is on staff (the nearest was 15 minutes away). In the E.R. the same thing happened again. And again, the next morning after I was hospitalized, only it was worse. The diagnosis of a stroke came as a complete surprise to me because I felt fine, and I didn't notice anything different about myself. I remember having no emotional response to the news. I felt annoyed and more concerned about getting home, because I was in the process of moving.

Not until several days later while I was in rehabilitation did I notice strange things happening to me. I was not frightened, angry, or annoyed. I didn't feel anything—nothing at all. Fourteen days after I was admitted to the hospital, I became extremely dizzy, and I felt I was falling out of my wheelchair. The floor was tilting to my left and the wheelchair was sliding off the floor. Any stimulus on my left side or repetitive movement with my left arm caused a disturbance in my relationship with my environment. For instance, the room would tilt down to the left, and I felt my wheelchair sliding downhill off the floor, and I was falling out of my chair. I would become disoriented, could hardly speak, and my whole being seemed to enter a new dimension. When my left side was placed next to a wall or away from any stimuli, this disturbance would gradually disappear. During this period, my left hand would contract, and the arm would draw up next to my body. It didn't feel or look like it belonged to me. Harrison moved the left arm repeatedly with the same movement, and a similar behavior occurred, except I started crying. He asked me what was I feeling, and I said anger. In another test he started giving me a hard time until the same episode began to occur, and I began to cry. He asked me what I was feeling, and I said anger. Actually I didn't feel the anger inside but in my head when I began to cry. Not until I went back to school did I become aware of having no internal physical feelings.

I call that arm *Alice* (Alice doesn't live here anymore)—the arm I don't like. It doesn't look like my arm and doesn't feel like my arm. I think it's ugly, and I wish it would go away. Whenever things go wrong, I'll slap it and say, “Bad Alice” or “It's Alice's fault.” I never know what it's doing or where it is in space unless I am looking at it. I can use it, but I never do consciously because I'm unaware of having a left arm. I don't neglect my left side, just Alice. Whatever it does, it does on its own, and most of the time, I don't know it's doing it. I'll be doing homework and then I'll take a sip of coffee. The cup will be empty. I was drinking coffee with that hand and didn't know it. Yet I take classical guitar lessons. I don't feel the strings or frets. I don't know where my fingers are nor what they are doing, but still I play.

How do I live with an illness I'm not aware of having? How do I function when I'm not aware that I have deficits? How do I stay safe when I'm not aware of being in danger?

Ms. C. is an obviously intelligent, widowed woman in her mid-50s enrolled in adult education classes and majoring in clinical psychology. She gets around splendidly despite “Alice” and despite a history of a temporary left hemiparesis. The diagnosis was damage of the right temporal-parietal cortex confirmed by an abnormal EEG recorded from that location. The damage was not sufficiently extensive to show up in a PET scan.

Contrast Mrs. C.'s story with the following observations made on an 8-year-old boy:

T.J. had an agenesis of the corpus callosum with a midline cyst at birth. During the first 6 months of his life, two surgical procedures were carried out to drain the cyst. Recently performed magnetic resonance imaging showed considerable enlargement of the frontal horns of the lateral ventricle, somewhat more pronounced on the right. The orbital part of the frontal lobes appeared shrunken as did the medial surface of the temporal pole.

T.J. appears to have no ability to quantify the passage of time and no experiential appreciation of the meaning of time units. For example, a few minutes after tutoring begins, he cannot say, even remotely, how long it has been since the session started. He is as apt to answer this question in years as in minutes. He does always use one of seven terms of time quantification (seconds, minutes, hours, days, weeks, months, or years) when asked to estimate the duration of an episode but uses them randomly. He can put these terms in order, but does not have any sense of their meaning or their numerical relationships to one another.

When T.J. returned from a trip to the Bahamas he did recall that he had been on the trip; however, the details he could recount about the trip numbered fewer than five. His estimates of how long it had been since his trip were typical in that they were inaccurate and wildly inconsistent on repeated trials. Also, on the first five tutoring sessions since his return he stated that he had not been tutored since his trip. He seems unable to place in sequence those few past events that he can recall. Nonetheless, he can answer questions correctly based on his application of general knowledge about development (e.g., he knows he was a baby before he could talk because “everyone starts as a baby”). However, one day he asked his tutor if he knew him when he was a kid, indicating his incomprehension of the duration of each of these developmental periods and his unawareness of what events constituted such a period for him.

T.J. is aware that he has a past, that events have happened to him but he cannot recollect those events. He also spontaneously speaks of events in his future such as driving an automobile and dating and growing a beard. He has playacted on separate occasions his own old age and death. He is capable of excitement about the immediate future. On the very day that he was going to the Bahamas he was very excited as he exclaimed repeatedly: “I'm going to the Bahamas.” But when his tutor asked him when, he said blankly: “I don't know.” He also displayed keen anticipation when he saw a helicopter preparing to take off from the hospital. The helicopter engines revved approximately 13 minutes before it took off and T.J. became increasingly more vocal and motorically active, laughing as he repeated “When's it going to take off?” He also anticipates future punishment when he is “bad.” He is aware, on some level of the immediate future in his constant question “What's next?” which he asks his mother at the end of each activity.

There are a variety of other occasions on which he demonstrated this capacity regarding tempo (as opposed to evaluating the duration of an experience). There have been several breaks in his usual thrice-weekly tutoring schedule. Each of four times this schedule has been interrupted, he has run to meet his tutor when he approached rather than waiting inside as he usually does. Also, on these occasions he has typically asked if his tutor missed him. However, he states he does not know how long it has been since his last session, and there is no evidence that he knew it had been longer than usual.

T.J. can compare who walks faster or who draws faster. He has a basic sense of sequencing as when he says “I'll take a turn and then you take a turn.” He also uses terms like “soon” and “quick” correctly in conversation. For example, when he wanted to do a drawing at the beginning of a session, and his tutor said that they needed begin to work, T.J. countered with “This will be quick.” Unsurprisingly, he finished his drawing at his normal pace. He somehow seems to use such terms correctly without any experiential appreciation of them.

These two case histories illuminate two very important dimensions of self. One dimension, portrayed by Mrs. C., *locates* an objective “me” in the world and with respect to the configural integrity of the body. The other dimension, highlighted by T.J. *monitors* an individual's experience. Without such monitoring, the events comprising the experience fail to be relevant and are not evaluated with respect to an autobiographical self, a narrative “I.” Kempen and van Loon have provided an excellent history of differentiating an objective “me” from a hermauentic “I.”

Location is akin to but more primitive than a spatial dimension; monitoring is akin to but more basic than a temporal dimension. However, the locational dimension includes clock time, what the Greeks called *chronos* and Albert Einstein among others related to space. A moving organism is always located in space-time. Monitoring entails not only the experiencing of duration but also the decisive moment; what the Greeks referred to as *kairos*. The dictionary defines *monitor* as a device to record or control a process; to check for significant content.

Unconscious Processing Freud had training both in medical practice and in philosophy. When he emphasized the importance of unconscious processes, was he implying the medical definition or the philosophical one? Most interpretations of Freud suggest that unconscious processes operate without one's awareness in the sense that they operate automatically much as do respiratory and gastrointestinal processes in someone who is stuporous or comatose. Freud himself seems to have promulgated this view by suggesting a horizontal split between conscious, preconscious, and unconscious processes with repression operating to push memory-motive structures into deeper layers where they no longer access awareness. Still in his “Project for a Scientific Psychology” memory-motive structures are neural programs that are located in the core portions of the brain that access awareness by their connections to cortex, which determine whether a memory-motivated wish comes to consciousness. When the neural program becomes a secondary process it comes under voluntary control, which involves reality testing and thus consciousness. To use language as an example, it is possible to know two languages but at any one time connect only one to cortex and thus the other remains

unconscious and voluntarily unexpressed.

The thrust of most recent psychoanalytical thinking as well as that of experimentalists such as Jack Hilgard is in the direction of interpreting the distinction between the conscious and unconscious in the philosophical sense. For instance, Matte Blanco proposes that consciousness be defined by the ability to make clear distinctions, to identify alternatives. Making clear distinctions would include being able to tell personal from extrapersonal reality. In contrast unconscious processes would, according to Matte Blanco, be composed of infinite sets “where paradox reigns and opposites merge into sameness.” When infinities are being computed, the ordinary rules of logic do not hold. Thus, dividing a line of infinite length results in two lines of infinite length, that is one equals two. Being deeply involved allows love and ecstasy but also suffering and anger to occur. In keeping with this, Carl Jung defined unconscious processes as those that involve feelings.

Bringing the wellsprings of behavior and experience to consciousness means making distinctions, providing alternatives, making choices, becoming informed in Claude Shannon's sense of reduction of uncertainty (Shannon noted that every binary decision reduced uncertainty by half; thus, each such decision provided one bit of information). Clarity regarding the details of how such distinctions are achieved did not come until the late 1960s when several theorists began to point out the difference between feedback, homeostatic processes on the one hand, and programs, which are feedforward, homeorhetic processes, on the other. Feedback mechanisms depend on error processing and are therefore sensitive to perturbations. Programs, unless completely stopped, run themselves off to completion irrespective of obstacles placed in their way. The difference between feedback and feed-forward processing turns out to be the same as the difference Freud drew between primary and secondary processes.

However, unconscious processes as defined by psychoanalysis are not completely “submerged” and unavailable to experience. Rather, unconscious processes produce feelings that are difficult to localize in time or in space and difficult to identify correctly. Unconscious processes construct the emotional dispositions and motivational context within which extrapersonal and personal realities are constructed. As research has shown, feelings are to a large extent undifferentiated, and are recognized and labeled according to the circumstances in which they become manifested.

It is in this sense that behavior is under the control of the unconscious processes. During angry outbursts, individuals are certainly aware of having lost their temper and of the effects of their anger on others. Despite this awareness, the anger may be uncontrolled. Only when the events leading to the anger become clearly separated into alternative or harmoniously related distinctions is the unconscious control converted into conscious control. A person with an obsession or compulsion is not unaware, in the instrumental sense, of his or her experience or behavior. The patient is very aware of it and feels awful; however, without help, he or she cannot differentiate between controls on the behavior generated by feelings.

Consequential Processing As is well known, frontal lesions were produced surgically in order to relieve intractable suffering, compulsions, obsessions, and endogenous depressions. When effective in relieving pain and depression, these psychosurgical procedures established the functional relation between frontal intrinsic cortex and the limbic forebrain in nonhuman primates. Further, frontal lesions can lead either to perseverative, compulsive behavior or to distractibility in monkeys and humans. A failure to be guided by the outcomes, that is, the consequences of behavior, can account for this effect; the opposite—the alleviation of obsessive-compulsive behavior—can also occur. Extreme forms of distractibility and obsession are caused by a lack of sensitivity of the activation (readiness) process to feedback from consequences. The results of experiments with monkeys as well as clinical observations attest to the fact that subjects with frontal lesions, whether surgical, traumatic, or neoplastic, fail to be guided by consequences.

Consequences are the outcomes of behavior. In the tradition of the experimental analysis of behavior, consequences are reinforcers that influence the recurrence of the behavior. Consequences are thus a series of events (Latin *ex-venire*, meaning *outcome*), outcomes that guide action and thereby attain predictive value (as determined by confidence estimates).

Confidence implies familiarity. Experiments on humans have shown that repeated arousal to an orienting stimulus results in habituation, that is, the orienting reaction gives way to familiarization. Familiarization is disrupted by limbic (amygdala) and frontal lesions. Ordinarily, familiarization allows continued activation of readiness; disruption of familiarization (orienting) leads to repeated distraction and thus a failure to allow consequences to form. When the process of familiarization is disrupted, the outcomes of behavior become inconsequential. When the process of familiarization is intact, it is segmented by orienting reactions into episodes within which confidence values can become established.

In such an episodic process the development of confidence is a function of coherences and correlations among the events being processed. When coherence and correlation span multiple episodes, the organism becomes committed to a course of action (a prior intention, a strategy), which then guides further action and is resistant to perturbation by particular orienting reactions (arousals). The organism is now competent to carry out the action; particular outcomes only *guide* competent performance, they no longer *produce* orienting reactions.

This cascade, which characterizes episodic processing, leads ultimately to considerable autonomy of the committed competence. Envisioned events are woven into coherent subjectivity, a story, a narrative, the myth by which the “I” lives. This narrative composes and is composed of an intention, a strategy that works for the individual in practice, a practical guide to action in achieving (temporary) stability in the face of a staggering range of options.

Consciousness is manifest (by verbal report) when familiarization is perturbed and an episode is updated and incorporated into a larger contextual scheme (the narrative) that includes both the familiar and novel episodes. Consciousness becomes attenuated when actions and their guides cohere—actions become skilled, graceful, and automatic.

Transcendental Consciousness The contents of consciousness are not exhaustively described by feelings of familiarity and novelty that are the basis for episodic and narrative consciousness, nor by those of extracorporeal allocentric and corporeal egocentric consciousness. The esoteric tradition in Western culture and the mystical traditions of the Far East are replete with instances of uncommon states that produce uncommon contents. These states are achieved by a variety of techniques such as meditation, yoga, or Zen. The contents of processing in such states appear to differ from ordinary feelings or perceptions. Experiences are described as *oceanic*, a merging of corporeal and extracorporeal reality or as *out-of-body*—that is, corporeal and extracorporeal realities continue to be clearly distinguished but are experienced by still another reality: a *meta-me* alternatively, the “I” becomes transparent, a throughput experiencing everything everywhere; there is no longer any segmentation into episodes nor do events become enmeshed in a narrative structure.

All these experiences have in common a transcendental relationship between ordinary experience and some more encompassing organizing principle. This relationship is ordinarily termed “spiritual.” The spiritual contents of consciousness can be accounted for by the effect of excitation of the frontolimbic forebrain (involved in narrative construction) on the dendritic microprocess that characterizes cortical receptive fields in the sensory extrinsic systems (involved in the construction of objective reality).

In addition to the gross topological correspondence between cortical receptive fields and the organization of sensory surfaces that gives rise to the overall characteristics of processing in the extrinsic systems, a microprocess that depends on the internal organization of each receptive field comes into play. This internal organization of receptive fields embodies, among other characteristics, a spectral domain: receptive fields of neurons in the extrinsic cortex are tuned to limited bandwidths of frequencies of radiant energy (vision), sound, and tactile vibration.

The most dramatic of these data are those that pertain to vision. The cortical neurons of the visual system are arranged, as are the other sensory systems, so as to reflect more or less isomorphically the arrangement of the receptor surfaces to which they are connected. Thus, the homunculi that Wilder Penfield and others have mapped onto the cortical surface of the extrinsic projection systems. However, within this gross arrangement lie the receptive fields of each of the neurons—a receptive field being determined by the dendritic arborization of that neuron that makes contact with the more peripheral parts of the system. Thus, the receptive field of a neuron is that part of the environment that is processed by the parts of the system to which the neuron is connected. Each receptive field is sensitive to approximately an octave of spatial frequency. This frequency-selective microprocess operates in a holographiclike manner.

Processing can thus be conceived to operate somewhat like the production of music by means of a piano; the sensory surface is analogous to a keyboard. Keyboard and strings are spatially related to provide the organization of the process. When individual strings are activated they resonate over a limited bandwidth of frequency. It is the combination of the spatial arrangement and the frequency-specific resonance of the strings that makes the production of music possible.

The gross and microorganization of the cortical neurons in the extrinsic systems resembles the organization of a multiplex hologram. A multiplex hologram is characterized by a Gabor elementary function, which Dennis Gabor called a *quantum of information*. A Gaussian envelope constrains the otherwise unlimited sinusoid described by the Fourier transform to make up the Gabor function. Experiments have shown that electrical excitation of frontal and limbic structures relaxes the

Gaussian constraints that are manifested as inhibitory surround or flanks in the receptive field architecture. When this occurs during ordinary excitation of the frontolimbic systems of the forebrain, processing leads to narrative construction. When frontolimbic excitation becomes overwhelming, experience is determined by an unconstrained holographic process.

Holograms of the type involved in brain processing are composed by converting (e.g., via Fourier transformation) successive sensory images (e.g., frames of a movie film) into their spectral representations and patching these microrepresentations into orderly spatial arrangements that represent the original temporal order of successive images. When such conversions are linear (e.g., when they employ the Fourier transform) they can readily be reconverted (e.g., by the inverse Fourier transform) into moving (i.e., successive) sensory images. The spectral domain is peculiar in that information (in the Gabor sense) becomes both distributed over the extent of each receptive field or quantum and enfolded within it. Thus sensory-image reconstruction can occur from any part of the total aggregate of receptive fields. This is what gives the aggregate its holographic, holistic aspect. All input becomes distributed and enfolded, including the dimensions of space and time and, therefore, of causality. It is this apparently timeless-spaceless-causeless aspect of processing instigated by overwhelming frontolimbic excitation that is responsible for the extrasensory dimensions of experience that characterize the esoteric traditions. Because of their enfolded property these processes tend to swamp distinctions, such as between corporeal and extracorporeal reality. In the esoteric traditions, consciousness is not limited to this type of reality.

An intriguing and related development (because it deals with the specification of a more encompassing, “cosmic” order) has occurred in quantum physics. Over the past 50 years it has become clear that there is a limit to the accuracy with which certain measurements can be made; this limit is expressed as an indeterminacy. In his description of a quantum of information, Gabor showed that a similar indeterminacy describes communication; leads to a unit of minimum uncertainty, the maximum amount of information that can be packed for processing. Thus, there is a convergence of the understanding of the microstructure of communication—and therefore of observation—and the microstructure of matter. The need to specify the observations that lead to inferring the properties of matter has led noted physicists to write representations of the observer into descriptions of the observable. Some physicists have noted the similarity of this specification to the esoteric descriptions of consciousness.

The scientific and esoteric traditions have been clearly at odds since the time of Galileo. Each new scientific discovery and the theory developed from it has thus far resulted in the widening of the rift between objective science and the subjective spiritual aspects of man's nature. The rift reached a maximum toward the end of the nineteenth century: mankind was asked to choose between God and Charles Darwin; Freud showed that heaven and hell resided within people and not in their relationship to the natural universe. However, the discoveries of twentieth-century science do not fit this mold. For once, the recent findings of science and the spiritual experiences of mankind are consonant. This augurs well for the upcoming new millennium because a science that comes to terms with the spiritual nature of mankind may well outstrip the technological science of the immediate past in its contribution to human welfare.

SUGGESTED CROSS-REFERENCES

Neuroanatomy is discussed in Section 1.2 and Section 1.3, electrophysiology in Section 1.9, perception and cognition in Section 3.1, psychoanalysis in Section 6.1, and psychosurgery in Section 31.32.

SECTION REFERENCES

Ashby WR: *Design for a Brain: The Origin of Adaptive Behavior*, ed 2. Wiley, New York, 1960.

*Avi G, Amir B: Brain organization and psychodynamics. J Psychother Pract Res *8:*24, 1999.

Bekesy Von G: *Sensory Inhibition*. Princeton University Press, New Jersey, 1967.

Bogen JE, Bogen GM: The other side of the brain III: The corpus callosum and creativity. Bull Los Ang Neurol Soc *34:*191, 1969.

Bracewell RN: The Fourier transform. Sci Am *260:*86, 1989.

Brentano F: *Psychologie vom empirischen Standpunki, vol III, Vom sinnlichen und noetischen Bewusfstsein*, ed 2, Franziska Mayer-Hillebrand, editor. Felix Meiner, Hamburg, 1968.

Bucy PC: *The Precentral Motor Cortex*. University of Illinois Press, Chicago, 1944.

Cannon WB: The James-Lange theory of emotions: A critical examination and an alternative theory. Am J Psychol *32:*106, 1927.

Efron R: *The Decline and Fall of Hemispheric Specialization*. Erlbaum, New York, 1989.

Gabor D: Theory of communication. J Inst Elect Eng *93:*429, 1946.

Gall FJ, Spurtzheim G: Research on the nervous system in general and on that of the brain in particular. In *Brain and Behavior*, KH Pribram, editor. Penguin, Middlesex, England, 1969.

Gazzaniga MS: *The Social Brain: Discovering the Network of the Mind*. Basic Books, New York, 1985.

Hermans HJM, Kempen HJG, van Loon RJP: The dialogical self: Beyond individualism and rationalism. Am Psychol *47:*23, 1992.

Hilgard ER: Divided consciousness: Multiple controls in human thought and action. Wiley, New York, 1977.

Hinton GE, Anderson JA: Parallel models of associative memory. Erlbaum, Hillsdale, NJ, 1981.

Jaynes J: The origin of consciousness in the breakdown of the bicameral mind. Houghton-Mifflin, Boston, 1990.

John ER: *Mechanisms of Memory*. Academic, New York, 1967.

Kelso JAS, Saltzman EL: Motor control: Which themes do we orchestrate? Behav Brain Sci *5:*554, 1982.

Lashley D: The thalamus and emotion. In *The Neuropsychology of Lashley*, FA Beach, DO Hebb, CT Morgan, HW Nissen, editors. McGraw-Hill, New York, 1960.

Matte Blanco I: *The Unconscious as Infinite Sets*. Gerald Duckworth Ltd, London, 1975.

McFarland DJ: *Feedback Mechanisms in Animal Behavior*. Academic Press, London, 1971.

Miller GA, Galanter A, Pribram KH: *Plans and the Structure of Behavior*. Henry Holt, New York, 1960.

Neumann E: *The Origins and History of Consciousness*. Princeton University Press, Princeton, NJ, 1954.

Olds J, Milner P: Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. J Comp Physiol Psychol *47:*419, 1954.

Peirce CS: *Pragmatism and Pragmaticism: Collected Papers*, vol 5. Harvard University Press, Cambridge, MA, 1934.

Pribram KH: *Brain and Perception: Holonomy and Structure in Figural Processing*. Erlbaum, New York, 1991.

Pribram KH: The intrinsic systems of the forebrain. In *Handbook of Physiology, Neurophysiology*, Field, HW Mogoan, VE Hall, editors. American Psychological Society, Washington, DC, 1960.

Pribram KH: The deep and surface structure of memory and conscious learning: Toward a 21st century model. In *Mind and Brain Sciences in the 21st Century*, RL Solso, editor. MIT Press, Cambridge, MA, 1997.

Pribram KH: How is it that sensing so much we can do so little? In *Central Processing of Sensory Input*, FO Schmitt, FLG Worden, editors. MIT Press, Cambridge, MA, 1974.

Pribram KH: *Languages of the Brain: Experimental Paradoxes and Principles in Neuropsychology*. Prentice-Hall, Englewood Cliffs, NJ, 1971.

Pribram KH: Limbic system. In *Electrical Stimulation of the Brain*, DE Sheer, editor. University of Texas Press, Austin, TX, 1961.

Pribram KH: Localization and distribution of function in the brain. In *Neuropsychology After Lashley*, J Orbach, editor. Erlbaum Associates, Hillsdale, NJ, 1982.

*Pribram KH: The Composition of Conscious Experience. J Conscious Studies *6:*20, 1999.

*Pribram KH: The primary of conscious experience. In *Science and the Primacy of Consciousness*, RL Amoroso, R Antunes, C Coelho, M Farias, A Leite, P Soares, editors. Noetic, Orinda, CA, 1999.

Pribram KH: What is mind that the brain may order it. In *Proceedings of the Norbert Weiner Centenary Congress, 1994*, V Mandreka, PR Masani, editors. American Mathematical Society, Providence, RI, 1997.

Pribram KH, Gill M: *Freud's "Project" Reassessed*. Basic Books, New York, 1976.

Pribram KH, McGuinness D: Arousal activation and effort in the control of attention. Psychol Rev *82:*116, 1975.

Pribram KH, Reitz S, McNeil M, Spevack AA: The effect of amygdalectomy on orienting and classical conditioning in monkeys. Pavlov J *14:*203, 1979.

Rumelhart DE, McClelland JL, and the PDP Research Group: *Parallel Distributed Processing*, vols 1 and 2. MIT Press, Cambridge, MA, 1986.

Schachter S, Singer TE: Cognitive, social and physiological determinants of emotional state. Psychol Rev *69:*379, 1962.

Searle JR: *The Mystery of Consciousness*. New York Review of Books, New York, 1997.

Thatcher RW, John ER: *Functional Neuroscience*, vol 1. Erlbaum, Hillsdale, NJ, 1977.

Weiskrantz L: *Blindsight: A Case Study and Implications*. Clarendon Press, Oxford, 1986.

3.6 EMOTIONAL INTELLIGENCE

DANIEL GOLEMAN, PH.D.

To be *emotionally intelligent* is to have the personal skills that characterize a rich and balanced personality. Emotional intelligence includes, as Aristotle put it, the rare ability “to be angry with the right person, to the right degree, at the right time, for the right purpose, and in the right way.” Emotional intelligence is distinct from intelligence quotient (I.Q.), which is the ability to perform cognitive tasks adeptly; each of these kinds of intelligence is based in differing but interlinked neural circuitry, with emotional intelligence largely mediated by limbic and prefrontal areas and I.Q. by neocortical zones alone. Emotional intelligence and I.Q. are not opposing competencies, but discrete and synergistic ones.

The theory of emotional intelligence offers a new psychological framework for primary prevention in psychiatry that integrates recent discoveries in cognitive science, neurological science, and child development. The competencies of emotional intelligence are crucial for the self-management of emotion and for the skillful handling of relationships. These abilities are learned throughout life, with primary learning occurring during childhood. Such learning shapes the underlying neurological circuitry, which continues to mature into adolescence. Emotional intelligence can be enhanced through the systematic offering of beneficial learning experiences as children grow, and deficits can be repaired through remedial learning and coaching.

Those who fail to master the competencies of emotional intelligence face a spectrum of heightened psychiatric risks, such as mood and anxiety disorders, eating disorders, and substance abuse. Because these skills of emotional intelligence are teachable, offering children and adolescents opportunities to strengthen these competencies can act as an inoculation against a spectrum of social and psychiatric risks.

COMPONENTS OF EMOTIONAL INTELLIGENCE

One commonly used version of Peter Salovey and John Mayer's 1990 definition of emotional intelligence includes abilities in five main areas:

1. *Self-awareness:* Recognizing one's feelings as they occur is the linchpin of emotional intelligence. The ability to monitor feelings from moment to moment is key to psychological insight and self-understanding. Being aware of one's emotions makes one more confident when making important personal decisions such as whom to marry or what career path to follow.
2. *Managing emotions:* Having appropriate emotional reactions is a capacity that builds on self-awareness. The ability to modulate negative affects such as anxiety, anger, and depression is a crucial emotional skill. Emotional resilience helps one to prevail over life's inevitable setbacks and upsets; those who lack emotional self-regulation are continually besieged by feelings of distress.
3. *Motivating oneself:* Being able to focus on a goal is essential for a range of accomplishments. Emotional self-control—such as delaying gratification or controlling impulsivity—is crucial in working towards such life goals. Individuals who can harness their emotions, and maintain hope and optimism despite frustrations, are generally more productive and effective in their undertakings.
4. *Recognizing emotions in others:* Empathy, another skill based in emotional self-awareness, is fundamental to interpersonal effectiveness. Those who are well attuned to subtle social cues that indicate what others feel are more successful in personal and professional relations.
5. *Handling relationships:* The art of relationships requires skill in managing others' emotions. Social competence underlies popularity, leadership, and interpersonal effectiveness.

Individuals have a profile of differing abilities in each of these areas; for instance, someone masterful at managing anger may be inept at soothing someone else's upsets. Neurological givens determine initial capacities within each domain of emotional intelligence. Each individual has underlying neurological setpoints that determine temperament—for example, the ability to control emotional impulse, shyness, or irritability. Although the underlying basis for emotional competence is neural, the brain circuitry involved is malleable. To a great extent, each of the five domains represents sets of habit and response that are learned, and so can be improved with appropriate effort.

Using a related measure, Jack Block found that the hallmarks of emotional intelligence are self-assurance, optimism, and social poise. Emotionally intelligent individuals have superior self-control and ability to motivate themselves. Life is meaningful for them; they are principled and responsible. They manage and express emotions appropriately, being assertive but sympathetic and caring in relationships. Their emotional life is rich but balanced; they are comfortable with themselves, others, and the social universe they live in. They manage stress without undue worry or rumination. They tend to be gregarious, spontaneous, playful, and open to sensual experience.

WHAT IS EMOTION?

The root of the word emotion is *motere*, the Latin verb “to move,” plus the prefix “e-” to connote “move away,” suggesting that a tendency to act is implicit in every emotion. *Emotion* refers to a feeling and its attendant thoughts, psychological and biological states, and range of impulses to act. The Oxford English Dictionary defines emotion as “any agitation or disturbance of mind, feeling, passion; any vehement or excited mental state.”

There is a long-standing debate about which emotions should be considered primary—the blue, red, and yellow of feeling from which myriad blends come—or if there are primary emotions at all. The argument for a set of core emotions is based to some extent on studies that suggest there are universally recognized facial expressions for four emotions: fear, anger, sadness, and enjoyment. The universality of facial expressions of emotion was probably first noted by Darwin, who saw it as evidence that these signals have been stamped by evolution into the central nervous system. According to some theorists the following families of emotion are universal:

Anger: fury, outrage, resentment, wrath, exasperation, indignation, vexation, acrimony, animosity, annoyance, irritability, hostility, and, perhaps at the extreme, pathological hatred and violence

Sadness: grief, sorrow, cheerlessness, gloom, melancholy, self-pity, loneliness, dejection, despair, and, when pathological, severe depression

Fear: anxiety, apprehension, nervousness, concern, consternation, misgiving, wariness, qualm, edginess, dread, fright, terror; as a psychopathology, phobia and panic

Enjoyment: happiness, joy, relief, contentment, bliss, delight, amusement, pride, sensual pleasure, thrill, rapture, gratification, satisfaction, euphoria, whimsy, ecstasy, and at the far edge, mania

Love: acceptance, friendliness, trust, kindness, affinity, devotion, adoration, infatuation, agape

Surprise: shock, astonishment, amazement, wonder

Disgust: contempt, disdain, scorn, abhorrence, aversion, distaste, revulsion

Shame: guilt, embarrassment, chagrin, remorse, humiliation, regret, mortification, and contrition

Each of these categories has a basic emotional nucleus at its core, with its variants rippling out in myriad mutations. In the outer ripples are *moods*, which, technically speaking, are more muted and last far longer than an emotion (it is relatively rare to maintain rage all day, for example, but less rare to be in an irritable mood, during which shorter bouts of anger are easily triggered). Beyond moods are *dispositions*, the temperamental proclivity to evoke a given emotion or mood such as melancholy, anxiety, or cheer. Further beyond such inclinations are the *disorders* of emotion such as clinical depressive disorders or generalized anxiety disorder, in which an individual feels chronically trapped in a toxic state.

Emotional intelligence entails the appropriate awareness, management, and expression of the range of these emotions. In this sense, many psychiatric disorders in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV)—such as the anxiety and mood disorders—bespeak a deficit in affective self-regulation, which is a key capacity of emotional intelligence. To the extent that emotional intelligence skills like affective self-regulation can be cultivated, particularly in young people, the risk of developing such psychiatric disorders should be diminished.

EMOTIONAL INTELLIGENCE AND I.Q.

In a sense the human brain contains two minds and two different kinds of intelligence: *rational* and *emotional*. These two fundamentally different modes of consciousness interact to constitute our mental life. The emotional and rational minds are semi-independent faculties, each reflecting the operation of distinct, but interconnected, circuitry in the brain. The complementarity of limbic system and neocortex, particularly of the amygdala and prefrontal lobes, means each is a full partner in mental life.

The emotional and the rational minds operate in tandem for the most part—emotion contributes to and informs the operations of the rational mind, and the rational mind refines and sometimes vetoes the input of the emotions. When these partners interact well, both emotional intelligence and intellectual ability are enhanced.

There is at best a slight correlation between I.Q. and certain facets of emotional intelligence, small enough to make it clear that these are largely independent entities. When people with high I.Q. flounder in life, and those with modest I.Q. do surprisingly well, the difference often may be attributable to emotional intelligence. Those with an extremely high I.Q. but low emotional intelligence—or low I.Q. and extremely high emotional intelligence—are relatively rare; abilities in both domains fit a bell curve.

Unlike the familiar tests for I.Q. there is as yet no corresponding test that measures emotional intelligence, although there is ample research on each of its components. Some aspects of emotional intelligence are best tested by studying an individual's ability at the task. Empathy can be evaluated, for example, by testing an individual's accuracy at interpreting another's feelings from their facial expressions.

In keeping with findings about other elements of emotional intelligence, for instance, research has demonstrated that empathy is independent of academic intelligence. Studies with school children have found only an incidental relationship between scores on measures of empathic acuity and SAT or I.Q. scores or school achievement tests. Children who showed an aptitude for reading feelings nonverbally were among the most popular in their schools, and the most emotionally stable. They did better academically, despite the fact that their I.Q.s were, on average, not higher than those of children who were less able to read nonverbal messages; this suggests that mastering this empathic ability smooths the way for classroom effectiveness (or simply makes teachers like these children more).

PRIOR THEORIES OF EMOTIONAL INTELLIGENCE

Peter Salovey and John Mayer are the first to offer a comprehensive model of emotional intelligence, but others have offered theories that overlap. In the 1920s E.L. Thorndike proposed that *social intelligence*—the ability to empathize and “act wisely in human relations”—was itself an aspect of I.Q. Other psychologists of the time took a more cynical view of social intelligence, seeing it in terms of skills for manipulating other people. Neither outlook on social intelligence convinced theorists, and the theory largely fell into disuse over the next decades. By the mid-1980s, Robert Sternberg's research on correlates of success led him back to Thorndike's conclusion: that social intelligence is distinct from intellectual capacity and is an essential component of what makes people function well in real life.

In 1983 Howard Gardner introduced the concept of multiple intelligences, proposing that rational intellect is just one of seven kinds of intelligence. Gardner's model rejects the traditional concept of I.Q. as a single, immutable factor; it recognizes that a range of skills and abilities matter over and above I.Q. His list includes the two standard academic measures of intelligence: verbal and mathematical-logical thinking, but it also includes the spatial capacity that an outstanding artist or architect might have; the kinesthetic ability necessary for dance or sports; and musical ability. Gardner also includes two facets of what he calls the “personal intelligences”: interpersonal skills—the ability to understand other people and what motivates them, and how to work cooperatively with them; and the capacity for “intrapsychic” awareness—a capacity to form an accurate, veridical model of oneself and to be able to use that model to function effectively in life. The theory of emotional intelligence expands on these two kinds of intelligence.

EMOTIONAL MIND

In recent years a cognitive science model of the emotional mind has been offered independently by Paul Ekman and Seymour Epstein. Their qualitative description sheds light on the underlying neurology; they propose the following qualities that distinguish the “emotional” from the “intellectual” mind.

Unlike the methodical and reflective intellect, the emotional mind reacts immediately, without pausing to consider consequences. Its workings are rapid, outpacing conscious thought and the deliberate, analytical reflection that is the mark of rationality. The emotional mind instantly apprehends potential danger, and can tune in to an emotional reality (e.g., he's angry with me; she's lying; this is making him sad), making snap judgments about whom to be wary of, whom to trust, who's in distress. However, these quick impressions and intuitive judgments can easily be erroneous or distorted, since the emotional mind sacrifices accuracy for speed.

Emotion-driven changes in facial expression begin within thousandths of a second after the event that triggers the reaction. The physiological changes that exemplify a given emotion—such as increased heart rate—also take only fractions of a second to begin. The reaction is particularly swift with intense emotion, like fear of a sudden threat.

This rapid-fire emotional reaction takes over in situations that have the urgency of primal survival, since emotions are a guide to action. The full blaze of intense emotion is technically very brief, lasting mere seconds rather than minutes, hours, or days. For an affect to endure, the original stimulus must be sustained, in effect continually invoking the emotion, as when the death of a loved causes prolonged bereavement. Emotions that persist for longer than a few seconds normally exist in muted forms as moods. Moods set an affective tone, but they do not determine perception or action as powerfully as brief, intense flares of emotion.

The immediate reaction to intense stimulus is emotional rather than logical because the rational mind processes stimuli more slowly than does the emotional mind. Accompanying thoughts are “automatic,” as described by Aaron Beck. However, there is a secondary, slower component to an emotional reaction. This second pathway to emotions is triggered by rumination; it is more deliberate, usually the result of conscious thought. This kind of emotional reaction entails a more extended appraisal in which cognition plays the key role in determining which emotions will be roused. Complex emotions, like embarrassment or apprehension over an upcoming exam, follow this slower route, taking seconds or minutes to unfold—these are emotions that follow from reflective thoughts.

The emotional mind is associative, and looks for symbolic significance. Art, religion, poetry, psychosis, dream, and myth are in the mode of the emotional mind, which Freud described as “primary process” thought. The primary process is the key that unlocks the meanings of art and dreams. In primary process thought, loose associations determine the flow of a narrative; one object symbolizes another; one feeling displaces another and stands for it; wholes are condensed into parts. There may be no coherence of time and no logical cause-and-effect sequence. The psychoanalytic method is in part the art of deciphering and unraveling these

substitutions in meaning.

The emotional mind has childlike qualities that intensify with the strength of the emotion. Categorical thinking typifies this, as does personalized thinking, with events perceived with a bias centering on the self.

Beliefs held by the rational mind are tentative—new evidence can refute one belief and replace it with a new one because the rational mind reasons by objective evidence. The emotional mind, however, is self-confirming, suppressing or ignoring memories or facts that would undermine its beliefs and seizing on those that support it. Emotions are self-justifying, having a set of perceptions and proofs all their own.

The emotional mind can react to the present as though it were the past. An event can bring back associated emotionally charged memories from the past. The emotional mind responds by triggering the feelings that went with the original event, whether or not or to what degree the comparison is warranted. If the evoked emotions are strong, the revived reaction is obvious. If the feelings are subtler, they can influence reactions without breaking into conscious thought. The emotional mind puts the rational mind to use, providing explanations (rationalizations) for emotional reactions without acknowledging the influence of emotional memory.

The emotional mind is largely state-specific in its operations; every strong emotion has its own distinct repertoire of thought, reactions, and memories. These state-specific repertoires become most predominant in moments of intense emotion. Selective memory, attention, or perception may signal that such an emotion-specific repertoire is active.

NEURAL ARCHITECTURE OF EMOTION

The dynamics of the emotional mind as understood through the lens of cognitive science reflect the underlying neural architecture. Understanding the interplay of brain structures that govern moments of acute, intense emotion reveals much about how emotional habits—positive and negative—are learned. The neurological data also suggest a primary prevention window of opportunity for shaping children's emotional habits.

The large size of the human neocortex makes possible the subtlety and complexity of human emotional life; the ratio of neocortex to limbic system is larger in primates than in other species and vastly larger in humans than in primates. Limbic areas that modulate emotion are connected via myriad circuits to all parts of the neocortex. Thus, the emotional centers have immense power to influence the functioning of the rest of the brain, including in its centers for thought. The neocortex, where prefrontal areas modulate emotional impulse, governs much of affective life, although in crucial, extremely emotional moments the neocortex defers to emotional impulse.

When emotions overwhelm concentration, what is being swamped is the mental capacity cognitive scientists call *working memory*, the ability to hold in mind the information relevant to the task at hand. What occupies working memory can be as mundane as the digits that compose a telephone number or as complicated as the intricate plot lines a novelist is trying to weave together. Working memory is the executive function par excellence in mental life, making possible all other intellectual efforts, from speaking a sentence to tackling a knotty logical proposition. Goldman-Rakic finds that the prefrontal cortex is the main site for working memory. When the limbic circuitry that converges on the prefrontal cortex is in the thrall of emotional distress, one cost is in the effectiveness of working memory: thinking, planning, and memory is impeded.

The amygdala is the prime candidate for the site of emotional memory. Antonio Damasio, studying patients with lesions between the amygdala and prefrontal cortex, has demonstrated the role of the amygdala and related limbic circuits in rational decisionmaking. These patients' decisionmaking was typically deeply flawed, despite their showing no deterioration in I.Q. nor in other cognitive abilities like attention and memory. They typically made disastrous choices in their business and personal lives, obsessing over decisions as simple as when to make an appointment.

Damasio argues that in the intact brain, limbic-driven surges from the viscera—“somatic markers”—function as inner alarms, calling attention to potential dangers that may result from a given course of action. These visceral sensations can cause an individual to act decisively and with confidence, immediately dropping or pursuing a course of action, thus paring down the array of choices to a more manageable decision matrix. More often than not these markers occur unaccompanied by conscious memories of the experiences that contribute to form this negative affect.

Although strong emotion can create havoc in reasoning, Damasio argues that the lack of awareness of feeling can also be ruinous, especially in weighing the decisions on which one's destiny largely depends: what career to pursue, whom to date or marry, where to live, and others. Such decisions cannot be made well through sheer rationality; they require the emotional wisdom developed through past experiences. Formal logic alone fails as the basis for deciding, for example, whom to marry or trust or even what job to take; these are realms where reason without feeling is blind.

NEURAL HIJACK

The precise neurological circuitry at play in each emotion has yet to be mapped; so far, the most extensive knowledge is for fear. Joseph LeDoux found the amygdala is central to fear and related acute emotional reactions. The amygdala stages a sort of “neural hijacking” during surges of fear and anxiety. For example, when triggered by fear, it sends urgent messages to every major part of the brain, controlling the secretion of fight-or-flight hormones, mobilizing centers for movement, and activating the cardiovascular system and muscles. Circuits from the amygdala catalyze the secretion of norepinephrine to heighten the reactivity of key brain areas, including those that make the senses more alert, in effect setting the brain on edge. The amygdala also signals the brainstem to fix the face in a fearful expression, cease unrelated muscle activity, increase heart rate and blood pressure, and slow breathing. Other circuits from the amygdala rivet attention on the source of the fear, and prepare the muscles to react.

Simultaneously, cortical memory systems are shuffled to retrieve stored knowledge relevant to the emergency at hand, taking precedence over other strands of thought. The amygdala is able to store memories and response repertoires that are marshaled without conscious thought because a direct circuit from thalamus to amygdala completely bypasses the neocortex. This bypass seems to allow the amygdala to be a repository for emotional impressions and memories that individuals are never fully aware of. Within milliseconds of the perception of something—and before there is conscious awareness of the stimulus—the “cognitive unconscious” produces an awareness of the identity of what we see, and also a liking or disliking for it—another possible indicator of the amygdala's role as a repository for emotional memory.

ROLE OF THE AMYGDALA IN FEAR

Fear is an apt case in point for understanding the neural architecture of emotion. In evolution, fear has a special prominence: more than any other emotion it is crucial for survival. In modern life, misplaced fears cause anxiety, and at a pathological extreme are at play in panic attacks, phobias, or obsessive-compulsive disorder. The neurology of fear highlights the central role of the amygdala as an alarm system.

Given a loud crash in the next room as raw physical data, the sensory circuit from, for example, the ear to the brain stem and then to the thalamus begins the first component of a fear reaction. From the thalamus, LeDoux has discovered, the path branches: a small bundle of projections leads to the amygdala and the nearby hippocampus and another, larger pathway leads to the auditory cortex in the temporal lobe, where sounds are sorted out and comprehended.

The hippocampus, a key storage site for nonaffective memory, quickly compares that “crash” to similar sounds to see if it is familiar—is this “crash” immediately recognizable? Meanwhile the auditory cortex does a more sophisticated analysis of the noise to determine its source—is it the dog? A fallen picture? An intruder? The auditory cortex makes a hypothesis about the source of the noise, and sends that message to the amygdala and hippocampus, which immediately compare it to similar memories.

If the conclusion is reassuring (the cat must have knocked a book off the desk), the general alert does not escalate to the next level. But if the noise remains unexplained, another coil of circuitry between the amygdala, the hippocampus, and the prefrontal cortex further accentuates uncertainty and fixates attention, redoubling the desire to identify the source of the sound. If no satisfying answer comes from this further keen analysis, the amygdala triggers an alarm, activating the hypothalamus, the brainstem, and the autonomic nervous system.

The amygdala's effectiveness as a central alarm system for the brain becomes evident in the ensuing sensations of apprehension and subliminal anxiety. Each of

several bundles of neurons in the amygdala has a distinct set of projections with receptors primed for different neurotransmitters. Discrete areas of the amygdala receive differing information. The amygdala's lateral nucleus receives projections from the thalamus and auditory and visual cortices. The corticomedial area of the amygdala registers smells via the olfactory bulb. The amygdala's central area receives messages from the viscera; this range of incoming signals allows the amygdala to function as a sentinel, scrutinizing every sensory experience.

The amygdala projects to most major areas of the brain. From its central and medial areas a branch goes to the parts of the hypothalamus that secrete corticotropin-releasing hormone (CRH), which mobilizes the fight-or-flight reaction via a cascade of other hormones. The amygdala's basal area sends out branches to the corpus striatum, linking into the brain's system for movement. Via the nearby central nucleus, the amygdala sends signals to the medulla in the autonomic nervous system, activating a wide range of far-flung responses in the cardiovascular system, muscles, and viscera.

The amygdala's basolateral area projects to the cingulate cortex and to the central gray cells that regulate the large muscles. These circuits also tighten the vocal cord muscles, creating the high-pitched voice of fright. Another path from the amygdala leads to the locus ceruleus in the brainstem, which in turn manufactures and disperses norepinephrine throughout the cortex, the brainstem, and the limbic system itself. Norepinephrine increases the overall reactivity of the brain areas that receive it, making the sensory circuits more sensitive.

Most of these changes go on outside conscious awareness. As unconscious anxiety begins to pierce awareness, the amygdala automatically commands a wide range of responses. It signals cells in the brainstem to put a fearful expression on the face, freezes unrelated movements the body's muscles had initiated, speeds heart rate, raises blood pressure, and slows breathing. These are just a few of the myriad changes orchestrated by the amygdala and connected areas as they commandeer the brain through a crisis.

The amygdala and the hippocampus together direct the cells that send key neurotransmitters to trigger releases of dopamine that rivets attention on the source of fear and puts the muscles at readiness to react. At the same time the amygdala signals sensory areas for vision and attention, making sure that the eyes seek out whatever is most relevant to the emergency at hand. Simultaneously cortical memory systems are reshuffled so that knowledge and memories most relevant to the particular emotional urgency will be readily recalled, taking precedence over other less relevant strands of thought.

The result of these many changes is the sensation of full-fledged fear: there is a characteristic tightness in the viscera, the heart speeds, the neck and shoulder muscles tighten, the limbs tremble; the body freezes in place, and the mind races, entertaining a host of possible dangers and appropriate responses. This entire sequence—from surprise to uncertainty to apprehension to fear—can occur within a second or so.

Damage to the amygdala also results in the loss of the ability to produce the acute physical sensation of fear. When a rare brain disease destroyed the amygdala (but no other brain structures) in the patient neurologists call “S.M.,” fear disappeared from her mental repertoire. She became unable to identify looks of fear on other people's faces, nor to make such an expression herself. As her neurologist put it, “If someone put a gun to S.M.'s head, she would know intellectually to be afraid but she would not feel afraid as you or I would.”

The competencies entailed in emotional intelligence reflect the workings of both cortical and subcortical circuits, particularly those that connect the amygdala and related structures with the prefrontal areas. Autopsy studies on children and teenagers have shown that the prefrontal zones, which modulate emotional expression and skill, are among the last parts of the brain to attain anatomical maturity; they do so around mid-adolescence, age 14 or 15.

Neuroplasticity means the brain shapes itself in part through repeated experiences, which strengthen the corresponding neural connections and weaken less used ones. The finding of relatively late maturation of the prefrontal circuitry for modulation of emotion suggests a neurological window of opportunity for helping children learn the best habits of emotional intelligence. If a child learns habits of emotional self-regulation, empathy, social skill and the like, the argument goes, then that will lower the risk of their needing remedial training (such as psychotherapy) later in life.

TRAUMA AND EMOTIONAL RELEARNING

One model for such emotional learning and relearning can be seen in the acquisition of and recovery from posttraumatic stress disorder. In posttraumatic stress disorder a terrifying moment of trauma becomes emblazoned in the emotional circuitry, continuing to trigger an emergency mobilization at the least cue of the original trauma. The acquisition of posttraumatic stress disorder and steps to recovery from it offer a paradigm of emotional learning and relearning with implications that generalize far beyond psychopathology.

Symptoms of posttraumatic stress disorder such as hypervigilance, nightmares, and overreacting to benign stimuli that in some way echo the traumatic event are, in effect, signs of an overaroused amygdala impelling vivid memories of trauma to continue to intrude on awareness. Traumatic memories encoded in the amygdala act as mental hair triggers, ready to sound a physiological alarm at the least suggestion that the original trauma is occurring again—witness the Vietnam veteran who panics at the sound of a car backfiring. This hair-trigger phenomenon may be a hallmark of emotional trauma of all kinds, including repeated physical abuse in childhood. Although posttraumatic stress disorder is typically the result of the impact of a single overwhelming episode, similar results can come from cruelties inflicted over a period of years, as can be the case with children who are sexually, physically, or emotionally abused.

Posttraumatic stress disorder represents a perilous lowering of the neural setpoint for alarm, such that an individual reacts to life's ordinary moments as though they were emergencies. The more brutal, shocking, and horrendous the events, the more indelible the memory. The neural basis for these memories appears to be a transformation in brain chemistry set in motion by a single instance of overwhelming terror. The main symptoms of posttraumatic stress disorder—the most intense kind of learned fearfulness—can be accounted for by changes in the limbic circuitry focusing on the amygdala. Some of the key changes are in the locus ceruleus, a structure that regulates the brain's secretion of catecholamines. These neurochemicals mobilize the body for an emergency; the same catecholamine surge stamps memories with special strength. In posttraumatic stress disorder this system becomes hyperreactive, secreting extra-large doses of these brain chemicals in response to situations that hold little or no threat but somehow trigger memories of the original trauma. The locus ceruleus and the amygdala are closely linked, along with other limbic structures such as the hippocampus and hypothalamus; the circuitry for the catecholamines extends into the cortex. Changes in these circuits are thought to underlie posttraumatic stress disorder symptoms like anxiety, fear, hypervigilance, being easily upset and aroused, readiness for fight or flight, and the indelible encoding of intense emotional memories. One study found that Vietnam veterans with posttraumatic stress disorder had 40 percent fewer catecholamine-inhibiting receptors than did men without the symptoms—suggesting that their brains had undergone a lasting change and that their catecholamine secretion was poorly controlled.

Other changes occur in the circuit linking the limbic brain with the pituitary gland, which regulates release of CRF, the main stress hormone the body secretes to mobilize the emergency fight-or-flight response. The changes lead this hormone to be oversecreted—particularly in the amygdala, hippocampus, and locus ceruleus—alerting the body for an emergency that really does not exist. Hypersecretion of CRF causes an overreactive startle response; individuals with too much CRF do not habituate to stressful stimuli.

A third set of changes occurs in the brain's opioid system, which secretes endorphins to blunt the feeling of pain; it also becomes hyperactive. This neural circuit also involves the amygdala, this time in concert with a region in the cerebral cortex. The opioids are brain chemicals that are powerful numbing agents, like opium and other related narcotics. When experiencing high levels of opioids, people have a heightened tolerance for pain—an effect that has been noted by surgeons who found that severely wounded soldiers needed lower doses of narcotics to handle their pain than did civilians with far less serious injuries.

Something similar seems to occur in posttraumatic stress disorder. Endorphin changes add a new dimension to the neural mix triggered by reexposure to trauma: a numbing of emotion. This appears to explain a set of negative psychological symptoms long noted in posttraumatic stress disorder: anhedonia and a general emotional numbness, a sense of being cut off from life or from concern about others' feelings. Those close to such people may experience this indifference as a lack of empathy. Another possible effect may be dissociation, including the inability to remember crucial minutes, hours, or even days of the traumatic event. The neural changes of posttraumatic stress disorder also seem to make a person more susceptible to further trauma. A number of studies have found that when animals were exposed even to mild stress when young, they were far more vulnerable than unstressed animals to trauma-induced brain changes later in life (suggesting a special urgency to treat children with posttraumatic stress disorder). This seems a reason why, exposed to the same catastrophe, one person goes on to develop posttraumatic stress disorder and another does not: the amygdala is primed to find danger, and when life presents it once again with real danger, its alarm rises to a higher pitch. All these neural changes offer short-term advantages for dealing with the emergencies that prompt them. Under duress, it is adaptive to be highly vigilant, aroused, ready for anything, impervious to pain, with the body primed for sustained physical demands, and—at least temporarily—less reactive to what might otherwise be intensely disturbing events. These short-term advantages, however, become lasting problems when the brain changes so that they become

predispositions, like a car stuck in perpetual high gear. When the amygdala and its connected brain regions take on a new setpoint during a moment of intense trauma, this change in excitability—this heightened readiness to trigger a neural hijacking—means that all of life is on the verge of becoming an emergency, and even an innocent moment is susceptible to an explosion of fear run amok.

In posttraumatic stress disorders spontaneous relearning fails to occur. This may be because of the brain changes of posttraumatic stress disorders, which are so strong that the amygdala hijacking occurs every time something even vaguely reminiscent of the original trauma comes along, thereby strengthening the fear pathway. This means that there is never a time when what is feared is paired with a feeling of calm; the amygdala never relearns a more mild reaction. Extinguishing the fear appears to involve an active adjustment process that is itself impaired in people with posttraumatic stress disorders, leading to the abnormal persistence of maladaptive responses.

But given the right experiences, even posttraumatic stress disorders can lift; strong emotional memories, and the patterns of thought and reaction that they trigger, can change with time. The original fear encoded in the amygdala does not die out completely; rather, the prefrontal cortex seems to actively suppress the amygdala's command to the rest of the brain to respond with fear.

The process whereby something that is not threatening becomes dreaded because it is associated with something frightening is called *fear conditioning*. Animal research indicates that one reason the ensuing fears can last for years is that the fear itself interferes with the ability to learn an alternative, more calm response. The key region of the brain that learns, retains, and acts on this fearful response is the circuit between the thalamus, amygdala, and prefrontal lobe—the pathway of neural hijacking. Judith Lewis Herman's therapeutic work describes distinct stages to recovery from trauma: attaining a sense of safety, remembering the details of the trauma and mourning the loss it has brought, and ultimately reestablishing a normal life. The establishment of a sense of safety as a first step may represent the necessary condition for unlearning fear conditioning. There may be a biological logic to the ordering of these steps as a sequence reflecting how the emotional brain relearns that life need not be regarded as an emergency about to happen.

PSYCHOTHERAPY AS EMOTIONAL RELEARNING

The same circuitry that can be seen so boldly imprinting traumatic memories is presumably at work in life's quieter moments as well. More ordinary childhood travails, such as being chronically ignored and deprived of attention or tenderness by one's parents, abandonment, loss, or social rejection, may never reach the fever pitch of trauma, but still leave their imprint on the emotional brain, creating distortions in intimate relationships later in life. Healing such muted emotional scars is a task of psychotherapy. In learning to deal skillfully with these loaded reactions, emotional intelligence comes into play.

The dynamic between the amygdala and the more fully informed reactions of the prefrontal cortex may offer a neuroanatomical model for how psychotherapy reshapes deep, maladaptive emotional patterns. Given the brain architecture that underlies emotional relearning in fear conditioning, what seems to remain even after successful psychotherapy is a vestigial reaction, a remnant of the original sensitivity or fear at the root of a troubling emotional pattern. The prefrontal cortex can refine or put the brakes on the amygdala's impulse, but cannot keep it from reacting in the first place. Thus while one cannot determine when one will have emotional outbursts, one can gain a degree of control over how long they last. A quicker recovery time from such outbursts may well be one mark of emotional maturity.

Over the course of therapy, what seems to change in the main are the responses that people make once an emotional reaction is triggered, but the tendency for the reaction to be triggered in the first place does not disappear entirely. Lester Luborsky's study of psychotherapy explored the main relationship conflicts that brought dozens of patients into psychotherapy—issues such as a deep craving to be accepted or find intimacy, or a fear of failure or of being overly dependent. Researchers carefully analyzed the typical (always self-defeating) responses the patients made when these wishes and fears were activated in their relationships—such as being too demanding, which created a backlash of anger or coldness in the other person, or withdrawing in self-defense from an anticipated slight, leaving the other person miffed by the seeming rebuff. During such ill-fated encounters, the patients understandably were flooded by upsetting feelings of hopelessness, sadness, resentment, anger, tension, fear, guilt, and self-blame. Whatever the specific pattern of the patient, it seemed to show up in important relationships, whether with a spouse or lover, a child or parent, or peers and supervisors at work.

Over the course of long-term therapy, however, these patients made two kinds of changes: their emotional reaction to the triggering events became less distressing, even calm or bemused, and their overt responses became more effective in getting what they truly wanted from the relationship. What did not change, however, was their underlying wish or fear, and the initial twinge of feeling. By the time the patients had but a few sessions left in therapy, the encounters they talked about showed they had only half as many negative emotional reactions compared to when they first started therapy, and were twice as likely to get the positive response they deeply desired from the other person; what did not change at all was the particular sensitivity at the root of these needs.

In brain terms the limbic circuitry may send alarm signals in response to cues of a feared event, but the prefrontal cortex and related zones would have learned a new, more healthy response. Emotional lessons—even the most deeply implanted habits of the heart learned in childhood—can be reshaped, which may explain why psychotherapy can take so long to change some of these patterns. Luborsky finds that even after successful therapy, core emotional propensities remain, although with an overlay of new insights and relearned responses.

Much successful psychotherapy is emotional relearning, as the posttraumatic stress disorders paradigm suggests. In theory, the behavior changes that ensue from psychotherapy should entail a relearning of emotional repertoires, and might mean a reshaping of the underlying neural circuitry. Suggestive evidence that this is indeed the case comes from a study of patients treated for obsessive-compulsive disorder. Half the study subjects received fluoxetine (Prozac). The other half received no medication, but rather underwent behavior therapy that systematically exposed them to the object of their obsession or compulsion without being allowed to perform it. Patients with hand-washing compulsions were put at a sink, for example, but not allowed to wash. They were taught to question the fears and dreads that motivated their compulsions, such as that failure to wash would mean they would get a disease and die. They were also encouraged to find an activity to do that they enjoyed—like listening to music—whenever they had the compulsion to wash. Over months of such sessions, therapy patients' compulsions faded, just as they did in subjects who received medication.

Positron emission tomography (PET) studies show that people with obsessive-compulsive disorder have greater than normal activity in the caudate nucleus, among other sites. The behavior therapy group had as significant a decrease in the activity of the caudate nucleus as did the patients successfully treated with fluoxetine. Systematic emotional and behavioral relearning had changed brain function and relieved symptoms as effectively as medication had.

ACQUISITION OF EMOTIONAL INTELLIGENCE

Although emotional behaviors are rooted in the operation of neurotransmitter systems, temperament is not destiny; emotional intelligence competencies are largely learned, with experience through childhood shaping and resetting of those neurotransmitter systems. Jerome Kagan proposes that genetically determined neurotransmitter setpoints probably cause individual differences in baseline temperament; for any given emotion, there are differences in how easily it is triggered, how long it lasts, and how intensely it is felt.

For example, Richard Davidson finds that the tendency toward timidity or boldness emerges within the first year of life, a fact that strongly suggests that it is genetically determined. Since the frontal lobes are still maturing in the first few months of life, their activity cannot be measured reliably until about the age of 10 months. But studies of 10-month-old infants have found that activity levels in the frontal lobes predicted whether the infants would cry when their mothers left the room. Of dozens of infants tested, every infant who cried had more brain activity in the right prefrontal area, a center for negative effect, while those who did not cry had more activity on the left; the correlation was virtually 100 percent. Timid children seem to begin life with a neural circuitry that renders them more reactive to even mild stress—from birth, their hearts beat faster than other infants' in response to strange or novel situations. Kagan posits that this signifies “behavioral inhibition,” a lifelong tendency to perceive any new person or situation as a potential threat. At 21 months, when reticent toddlers refrained from playing, heart-rate monitors showed that their hearts were racing with anxiety. This easily aroused anxiety seems to promote lifelong introversion. Middle-aged women who report having been particularly shy as children tend to experience more fears, worries, and guilt, and to suffer more from stress-related conditions such as migraine headaches, irritable bowel syndrome, and other stomach problems throughout life, compared with their more extroverted peers.

Although basic disposition appears to be laid down from birth, or very nearly from birth, those who have a tendency toward inhibition are not necessarily fated to be shy and withdrawn all their lives. The emotional lessons of childhood can have a profound impact on temperament, either amplifying or muting an innate predisposition. The great plasticity of the brain in childhood means that experiences during those years can have a lasting impact on the sculpting of neural pathways for the rest of life. Kagan finds that throughout childhood some timid children grow bolder as experience reshapes key neural circuitry. Subjects who were timid at age 4 were more likely to overcome natural inhibition by age 10 if they had a higher level of social competence: being cooperative; getting along with other children; being

empathic, prone to giving and sharing; being considerate; and being able to develop close friendships. More socially skilled children in this group were far more likely to have a succession of positive experiences with other children. The regular repetition of such social success over many years seems to reduce timidity by making the children more sure of themselves.

Subjects who remained timid at age 10 tended to be less emotionally competent: crying and falling apart under stress more easily; being emotionally inappropriate; displaying fearful, sulky, or whiny behavior; overreacting to minor frustrations; having trouble delaying gratification; being overly sensitive to criticism; or being mistrustful. These emotional lapses are likely to mean their relationships with other children will be troubled, should they be able to overcome their initial reluctance to engage.

Parental style is a powerful modifier of temperament. Kagan reports that “inhibited” children who had become less timid by kindergarten seem to have had parents who put consistent gentle pressure on them to be more outgoing, offering them repeated chances to take social risks despite fears, and so having mastery experiences. Kagan posits this behavioral change toward boldness marks an underlying shift in neurotransmitter setpoints in the amygdala. All learning—emotional, factual, or logical—implies a physical change in the brain, a strengthening of synaptic connection.

One key emotional circuit that continues to shape itself through childhood centers on the vagus nerve, which both regulates the heart and other organs and signals the amygdala via other circuits to secrete catecholamines, which initiate the fight-or-flight response. Research suggests that emotionally adept parenting leads to improved vagal-nerve function in children.

According to John Gottman, parents modify their children's vagal tone with emotional “coaching”: talking to children about their feelings and how to understand them, adopting an uncritical and nonjudgmental stance, helping children to solve emotional predicaments, and offering constructive alternatives to hitting and sulking behaviors. A 4-year longitudinal study found children of parents who did this well were better able to suppress vagal activity, had lower levels of resting catecholomines, and fewer episodes of tantrums and impulsivity.

DEVELOPMENT OF EMOTIONAL INTELLIGENCE

Emotional learning begins in life's earliest moments, and continues throughout childhood and into adulthood. Childhood and adolescence are the most critical windows of opportunity for the establishment of essential emotional habits. The primary skills of emotional intelligence each have critical periods extending over several years in childhood. Each period represents an opportunity for instilling effective emotional habits. The strengthening, sculpting, and pruning of neural circuits throughout childhood may contribute to the enduring and pervasive effects of early emotional hardships and trauma in adulthood.

Hundreds of studies suggest that how parents treat a child—with harsh discipline or emphathy, with indifference or warmth, and so on—has deep and lasting consequences for the child's emotional life. Only recently have there been hard data showing that emotionally intelligent parenting is itself of enormous benefit to a child. The ways parents handle emotions between them—in addition to their direct dealings with a child—impart powerful lessons to their children, who are astute learners attuned to the subtlest emotional exchanges in the family. Gottman's microanalysis of interactions in couples on how the partners handled their children suggested that couples who were more emotionally competent in the marriage relationship were also the most effective in helping children to cope with emotional ups and downs.

All the small interactions between parent and child have an emotional subtext; the repetition of these messages over the years forms the core of a child's emotional outlook and capabilities. These interchanges mold the child's emotional expectations about relationships, which will influence emotional functioning in all realms of life for better or worse.

The risks are greatest for children whose parents are grossly inept—immature, abusing drugs, depressed, chronically angry, or simply aimless and living chaotic lives. Such parents are far less likely to give adequate care, let alone to address their childrens' emotional needs. Simple neglect, studies find, can be more harmful than outright abuse. Studies of maltreated children find that neglected youngsters fare worst of all: they tend to be the most anxious, inattentive, and apathetic, alternately aggressive and withdrawn. In one study the rate among neglected children for having to repeat first grade was 65 percent.

Emotional competence may be decisive in determining the extent to which any given child succumbs to such hardships or responds to them with a core of resilience and thrives despite the odds. Long-term studies of children brought up in poverty, in abusive families, or by a parent with severe mental illness show that those who survive the most severe hardships tend to share key emotional skills. These include social adeptness that draws people to them, self-confidence, persistence, optimism, resilience in the face of upsets, and an easy-going nature.

PRIMARY PREVENTION

The plight of today's children can be seen in day-to-day problems that have not yet blossomed into outright crises. Perhaps the most telling data of all—a direct barometer of dropping levels of emotional competence—are from a national sample of American children, ages 7 to 16, comparing their emotional condition in the mid-1970s and at the end of the 1980s. Based on parents' and teachers' assessments, there was a steady worsening. No one problem stood out; all indicators simply crept steadily in the wrong direction. Children, on average, were doing more poorly in these specific ways:

- Withdrawal or social problems: preferring to be alone; being secretive; sulking a lot; lacking energy; feeling unhappy; being overly dependent
- Anxious and depressed: being lonely; having many fears and worries; needing to be perfect; feeling unloved; feeling nervous, sad, or depressed
- Attention or thinking problems: unable to pay attention or sit still; daydreaming; acting without thinking; being too nervous to concentrate; doing schoolwork poorly; unable to get mind off thoughts
- Delinquent or aggressive: hanging around kids who get in trouble; lying and cheating; arguing a lot; being mean to other people; demanding attention; destroying other people's things; disobeying at home and at school; being stubborn and moody; talking too much; teasing a lot; having a hot temper

No child, rich or poor, is exempt from risk; these problems are universal and occur in all ethnic, racial, and income groups. Thus, while children in poverty have the worst record on indices of emotional skills, their rate of deterioration over the decades was no worse than for middle-class children or for wealthy children: all show the same steady slide. There has also been a corresponding threefold rise in the number of children who have received psychological help, as well as a near doubling of the number of children whose emotional problems warrant help that they have not received—from about 9 percent in 1976 to 18 percent in 1989.

A more careful look at the mechanics of specific problems suggests how given deficits in emotional or social competencies lay the foundation for social or psychiatric problems and how well-aimed correctives or preventives could keep more children on track.

Aggression At the root of aggressiveness lie deficiencies in a triad of emotional intelligence competencies: empathy, self-regulation of anger, and impulsivity. Genetic givens aside, the social shaping of aggression starts early. The family life of aggressive children typically includes parents who alternately neglect their child and administer harsh and capricious punishments, a pattern that can make a child paranoid or combative. Not all angry children are bullies; some are withdrawn social outcasts who overreact to being teased or to what they perceive as slights or unfairness. The one perceptual flaw that such children share is that they perceive slights where none were intended, imagining their peers to be more hostile toward them than they actually are. This leads them to misperceive neutral acts as threatening ones—an innocent bump is seen as intentional—and to attack in return. As a result, other children shun them, isolating them further. Such angry, isolated children are highly sensitive to injustices and being treated unfairly. They typically see themselves as victims and can recite a list of instances when, for example, they were blamed for something they did not do. Another trait of such children is that once they are angry they can think of only one way to react: by lashing out.

These perceptual biases can be seen at work in an experiment in which a bully is paired with a more peaceable child to watch videos. In one video, a boy drops his books when another knocks into him, and children standing nearby laugh; the boy who dropped the books gets angry and tries to hit a child who laughed at him. When the boys who watched the video talk about it afterward, the bully always sees the boy who struck out as being justified. Even more telling, when asked to rate how aggressive the boys were, the bullies see the boy who knocked into the other as being more combative, and the anger of the boy who struck out as being justified.

This jump to judgment testifies to a deep perceptual bias in people who are unusually aggressive: they act on the basis of the assumption of hostility or threat, paying too little attention to what is actually going on. Once they assume threat, they leapfrog to action. For instance, if an aggressive boy is playing checkers with another who moves a piece out of turn, he'll interpret the move as cheating without pausing to find out if it had been an innocent mistake. His presumption is of evil intent

rather than innocence; his reaction is automatic hostility. The knee-jerk perception of a hostile act is accompanied by an equally automatic aggression; instead of pointing out to the other boy that he made a mistake, the aggressive boy begins accusing, yelling, and hitting. The more such children do this, the more automatic aggression becomes for them, and the more the repertoire of alternatives—politeness, joking—shrinks.

Such children are emotionally vulnerable in the sense that they have a low threshold for upset, getting peeved more often by more things; once upset, their thinking is muddled, so that they see benign acts as hostile and fall back on their overlearned habit of striking out. These perceptual biases toward hostility are already in place by the early grades. Most children, especially boys, are rambunctious in kindergarten and first grade, but the more aggressive children fail to learn even a modicum of self-control by second grade. Where other children have started to learn how to settle playground disagreements by negotiation and compromise, the bullies rely more and more on force and bluster. They pay a social price: within 2 or 3 hours of a first playground contact with a bully, other children say they dislike him.

Studies that have followed children from the preschool years into teenagehood find that up to half of first graders who are disruptive, unable to get along with other kids, disobedient with their parents, and resistant with teachers will become delinquents in their teen years. Not all such aggressive children are on the trajectory that leads to violence and criminality in later life. But of all children, these are the ones who are most at risk for eventually committing violent crimes.

The drift toward crime shows up surprisingly early in these children's lives. When kindergarteners were rated for hostility and troublemaking behavior, those whose behavior was most extreme at age 5 already had far greater evidence of delinquency just 5 to 8 years later in their early teens. They were about three times as likely as other children to admit they had beaten up someone who had not done anything to them, to have shoplifted, to have used a weapon in a fight, to have broken into or stolen parts from a car, and to have been drunk—and all this before they reached 14 years of age.

The prototypical pathway to violence and criminality starts with children who are aggressive and hard to handle in first and second grade. Typically, from the earliest school years their poor impulse control also contributes to their being poor students; they are seen as and see themselves as “dumb”—a judgment confirmed by their being shunted to special-education classes. Although such children may have a higher rate of “hyperactivity” or learning disorders, all of them do not. Children who on entering school already have learned a coercive style—that is, bullying—are also written off by their teachers who have to spend too much time keeping these children in line. Defying classroom rules comes naturally to these children, which means that they waste time that would otherwise be used in learning; that they are destined to academic failure is usually obvious by about third grade. While boys on a trajectory toward delinquency tend to have lower I.Q. scores than their peers, their impulsivity is more directly at cause: impulsivity in 10-year-old boys is almost three times as powerful a predictor of their later delinquency as is their I.Q.

By fourth or fifth grade these kids—now seen as bullies or just difficult—are rejected by their peers and are unable to make friends easily, if at all, and have become academic failures. Feeling themselves friendless, they gravitate to other social outcasts. Between grade four and grade nine they commit themselves to their outcast group and a life of defying the law: they show a fivefold increase in their truancy, drinking, and substance use, with the biggest boost between seventh and eight grade. By the middle-school years, they are joined by another type of “late starters,” who are attracted to their defiant style; these late starters are often youngsters who are completely unsupervised at home and started roaming the streets on their own in grade school. In the high-school years this outcast group typically drops out of school in a drift toward delinquency, engaging in petty crimes such as shoplifting, theft, and drug-dealing.

A telling difference emerges in this trajectory between boys and girls. A study of fourth-grade girls who were “bad”—getting in trouble with teachers and breaking rules, but not unpopular with their peers—found that 40 percent had a child by the time they finished the high-school years, which was three times the average pregnancy rate for girls in their schools. In other words, antisocial teenage girls don't get violent—they get pregnant.

There is, of course, no single pathway to violence and criminality, and many other factors can put a child at risk: being born in a high-crime neighborhood where they are exposed to more temptations to crime and violence, coming from a family under high levels of stress, or living in poverty. However, none of these factors makes a life of violent crime inevitable. All things being equal, the psychological forces at work in aggressive children greatly intensify the likelihood of their ending up as violent criminals.

The bent of mind that aggressive children take with them through life is one that almost ensures they will end up in trouble. A study of juvenile offenders convicted of violent crimes and of aggressive high-school students found a common mind set: when they have difficulties with someone, they immediately see the other person in an antagonistic way, jumping to conclusions about the other person's hostility toward them without seeking any further information or trying to think of a peaceful way to settle their differences. At the same time, the negative consequence of a violent solution—a fight, typically—never crosses their mind. Their aggressive bent is justified in their mind by beliefs like, “It's okay to hit someone if you just go crazy from anger”; “If you back down from a fight everyone will think you're a coward”; and “People who get beaten up badly don't really suffer that much.”

Timely help can change these attitudes and stop a child's trajectory toward delinquency; several experimental programs have had some success in helping aggressive children learn to control their antisocial bent before it leads to more serious trouble. One program for anger-ridden grade-school troublemakers conducted training sessions for 40 minutes twice a week for 6 to 12 weeks. The boys were taught, for example, to see that some of the social cues they interpreted as hostile were in fact neutral or friendly. They learned to take the perspective of other children, to get a sense of how they were being seen and of what other children might be thinking and feeling in the encounters that had made them so angry. They also got direct training in anger control through enacting scenes, such as being teased, that might lead them to lose their temper. One of the key skills for anger control was monitoring their feelings—becoming aware of their body's sensations, such as flushing or muscle tensing, as they were getting angry, and to take those feelings as a cue to stop and consider what to do next rather than strike out impulsively.

Many aggressive boys are unhappy that they lose their temper so easily, and are receptive to learning to control it. In the heat of the moment, cool-headed responses such as walking away or counting to ten so the impulse to hit will pass before they react are not automatic; the boys practice such alternatives in role-playing scenes such as riding a bus where other kids are taunting them. They can try out friendly responses that preserve their dignity while giving them an alternative to hitting, crying, or running away in shame.

Three years after the boys had been through the training, a comparison matched these boys with others who had been just as aggressive, but did not have the benefit of the anger-control sessions. It found that, in adolescence, the boys who graduated from the program were much less disruptive in class, had more positive feelings about themselves, and were less likely to drink or take drugs. Also, the longer they had been in the program, the less aggressive they were as teenagers.

Depression International data show what seems to be a modern epidemic of depression; each successive generation worldwide since the beginning of the twentieth century has lived with a higher risk than their parents of suffering a major depression (called major-depressive disorder in DSM-IV)—not just sadness, but a paralyzing listlessness, dejection, and self-pity, and an overwhelming hopelessness—over the course of life. Also, these episodes are beginning at earlier and earlier ages. Childhood depression, once virtually unknown (or unrecognized), is emerging as a fixture of the modern scene.

Although the likelihood of becoming depressed rises with age, the greatest increases are among young people. For those born after 1955 the likelihood they will suffer a major depression at some point in life is, in many countries, three times or more greater than for their grandparents. Among Americans born before 1905, the rate of those having a major depression over a lifetime was just 1 percent; for those born since 1955, by age 24 about 6 percent had become depressed. For those born between 1945 and 1954, the chances of having had a major depression before age 34 are ten times greater than for those born between 1905 and 1914. For each generation the onset of a person's first major depressive episode has tended to occur at an ever-earlier age.

Worldwide, the age when people first experience depression has dropped. The losses of stable sources of self-identification make children more susceptible to depression. A wide-scale reduction of available support from sources such as religious communities and extended families means a broad loss of resources that can buffer against setbacks and failures. Early emotional stressors may affect neuronal development, which could cause depression even decades later. Whatever the cause, depression in the young is a pressing problem. In the United States, estimates of how many children and teens are depressed in any given year vary widely, as opposed to vulnerability over their lifetime. Some epidemiological studies using strict criteria—the DSM-IV diagnostic criteria for depressive disorders—have found that for boys and girls between 10 and 13 the rate of major depressive disorder over the course of a year is as high as 8 or 9 percent, although other studies place it at about half that rate and even as low as 2 percent. Some data suggest that at puberty the rate nearly doubles for girls; up to 16 percent of girls between 14 and 16 suffer an episode of depression, whereas the rate is unchanged for boys.

Among children whose depression was severe enough that they were referred for treatment, three quarters had a subsequent episode of severe depression, according to one study of children diagnosed with depressive disorder when they were as young as 8 years old and assessed every few years until some were as old as 24. Children with major depressive disorder had episodes lasting about 11 months on average, although in one in six of them it persisted for as long as 18 months. Dysthymia, which began as early as age 5 in some children, was less incapacitating but lasted far longer—an average of about 4 years. Children who have a

dysthymia are more likely to have it intensify into major depression—a so-called *double depression*. Those who develop double depression are much more prone to suffer recurring episodes as they grow older. As children who had an episode of depression grew into adolescence and early adulthood, they suffered from a depressive disorder or bipolar I disorder, on average, 1 year in 3.

The cost to children goes beyond the suffering caused by depression itself. The sullenness or sadness such children feel causes them to avoid initiating social contacts, or to look away when another child is trying to engage them—a social signal the other child takes as a rebuff; the end result is that depressed children end up rejected or neglected on the playground. This lacuna in their interpersonal experience means they miss out on what they would normally learn in the rough-and-tumble of play. They become social and emotional laggards, with much catching up to do after the depression lifts. Indeed, when compared to those without depression, depressed children have been found to be more socially inept, to have fewer friends, to be less preferred than others as playmates, to be less liked, and to have more troubled relationships with other children.

Another cost to these children is doing poorly in school; depression interferes with their memory and concentration, making it harder to pay attention in class and retain what is taught. A child who feels no joy in anything will find it hard to marshal the energy to master challenging lessons, let alone experience flow in learning. Understandably, the longer children are depressed, the more their grades drop and the poorer they do on achievement tests, so that they are more likely to be held back in school. There is a direct correlation between the length of time a child had been depressed and his or her grade point average, with a steady plummet over the course of the episode, all of which compounds the depression. Just as with adults, pessimistic ways of interpreting life's defeats seem to feed the sense of helplessness and hopelessness at the heart of children's depression. That people who are already depressed think in these ways has long been known. However, what has emerged only recently is that children who are most prone to melancholy tend toward this pessimistic outlook before they become depressed, which suggests a window of opportunity for inoculating them against depression before it strikes.

One line of evidence comes from studies of children's beliefs about their own ability to control what happens in their lives, such as being able to change things for the better. This is assessed by children's ratings of themselves in such terms as “When I have problems at home I'm better than most kids at helping to solve the problems” and “When I work hard I get good grades.” Children who say none of these positive descriptions fits them have little sense that they can do anything to change things; this sense of helplessness is highest in those children who are most depressed. A telling study looked at fifth- and sixth-graders in the few days after they received report cards. There is a marked consequence in how children assess their role when they get a worse grade than they expected: those who see a bad grade as being caused by some personal flaw (“I'm stupid”) feel more depressed than those who explain it away in terms of something they could change (“If I work harder on my math homework I'll get a better grade”).

Martin Seligman identified a group of third-, fourth-, and fifth-graders whom classmates had rejected, and tracked the ones who continued to be social outcasts in their new classes the following year. How the children explained the rejection to themselves seemed crucial to whether they became depressed. Those who saw their rejection as due to some flaw in themselves grew more depressed. But the optimists, who felt that they could do something to change things for the better, were not especially depressed despite the continuing rejection. In a study of children making the notoriously stressful transition to seventh grade, those with the pessimistic attitude responded to high levels of hassles at school and to any additional stress at home by becoming depressed. The most direct evidence that a pessimistic outlook makes children highly susceptible to depression comes from a 5-year study of children beginning when they were in third grade. Among the younger children, the strongest predictor that they would become depressed was a pessimistic outlook coupled with a major blow such as parents divorcing or a death in the family, which left the child upset, unsettled, and, presumably, with parents less able to offer a nurturing buffer. As the children grew through the elementary school years, there was a telling shift in their thinking about the good and bad events of their lives, with the children increasingly ascribing them to their own traits: “I'm getting good grades because I'm smart”; “I don't have many friends because I'm no fun.” This shift seems to set in gradually over the third to fifth grades. As this happens those children who develop a pessimistic outlook—attributing the setbacks in their lives to some dire flaw in themselves—begin to fall prey to depressed moods in reaction to setbacks. Also, the experience of depression itself seems to reinforce these pessimistic ways of thinking, so that even after the depression lifts, the child is left with what amounts to an emotional scar, a set of convictions fed by the depression and solidified in the mind: that he cannot do well in school, is unlikable, and can do nothing to escape his own brooding moods. These fixed ideas can make the child all the more vulnerable to another depression down the road.

There are signs that teaching children more productive ways of looking at their difficulties lowers their risk of depression. In a study of one Oregon high school, about one in four students had symptoms of subclinical depression. In a special after-school class 75 of the mildly depressed students learned to challenge the thinking patterns associated with depression, to become more adept at making friends, to get along better with their parents, and to engage more in social activities that they enjoyed. By the end of the 8-week program, 55 percent of the students had recovered from their mild depression whereas only about a quarter of equally depressed students who were not in the program had begun to pull out of their depression. A year later a quarter of those in the comparison group had gone on to fall into a major depressive disorder, as opposed to only 14 percent of students in the depression-prevention program. Although the program lasted just eight sessions, it seemed to have cut the risk of depression in half.

Similarly promising findings came from a special weekly class given to 10- to 13-year-old youngsters at odds with their parents and showing some signs of depression. In after-school sessions they learned some basic emotional skills, including handling disagreements, thinking before acting, and, perhaps most important, challenging the pessimistic beliefs associated with depression—for example, resolving to study harder after doing poorly on an exam instead of thinking, “I'm just not smart enough.” Children learn in these classes that moods like anxiety, sadness, and anger do not just descend without warning; they learn that you can change the way you feel by what you think and that disputing the depressing thoughts vanquishes the gathering mood of gloom. These classes lowered depression rates by one half and continued to do so as long as 2 years later. A year after the classes ended, just 8 percent of those who participated scored at a moderate-to-severe level on a test of depression, versus 29 percent of children in a comparison group. After 2 years, about 20 percent of those in the course were showing some signs of at least mild depression, compared to 44 percent of those in the comparison group.

Learning these emotional skills at the cusp of adolescence—a crucial window for risk of depression—may be especially helpful. One experimental program focused on helping adolescent girls to learn how to handle relationships better: how to develop a friendship, how to feel more confident with other teens, how to assert limits on sexual closeness, how to be intimate, and how to express their feelings. In essence, it was a remedial tutorial in some of the most basic emotional skills.

Problems in relationships are a common trigger for depression, particularly in young people. The difficulty is as often in children's relationships with their parents as it is with their peers. Depressed children and teenagers are frequently unable or unwilling to talk about their sadness. They seem unable to label their feelings accurately, showing instead a sullen irritability, impatience, crankiness, and anger—especially toward their parents. This, in turn, makes it harder for their parents to offer the emotional support and guidance the depressed child actually needs, setting in motion a downward spiral that typically ends in constant arguments and alienation.

This new look at the causes of depression in the young pinpoints deficits in two areas of emotional competence: relationship skills and a depression-promoting way of interpreting setbacks. Although the tendency to depression is partially the result of genetic destiny, some of it seems caused by reversible, pessimistic habits of thought that predispose children to react to life's small defeats—a bad grade, arguments with parents, social rejection—by becoming depressed. Evidence suggests that the predisposition to depression, whatever its basis, is becoming ever more widespread among the young and there is a pressing need to introduce children at risk to psychological inoculation.

Eating Disorders The clinical literature on eating disorders offers a multitude of hypotheses about the causes, ranging from ever-younger girls feeling compelled to compete with unattainably high standards of female beauty to intrusive mothers who enmesh their daughters in a controlling web of guilt and blame. Most of these hypotheses suffer from one great drawback: they are extrapolations from observations of patients already diagnosed, rather than from prospective studies conducted on a large cohort before they became symptomatic. Such studies would assess for example, if having controlling parents predisposes a girl to eating disorders. Beyond that, studies can identify the cluster of conditions that leads to the problem, and distinguish them from conditions that might seem to be a cause, but that actually are found as often in people without the problem as in those who come for treatment.

When just such a study was done with more than 900 girls in the seventh through tenth grades, emotional deficits—particularly a failure to distinguish distressing feelings from one another and to control them—were found to be key among the factors leading to eating disorders. By tenth grade 61 girls in an affluent suburban high school already had serious symptoms of anorexia nervosa or bulimia nervosa. The greater the problem, the more the girls reacted to setbacks, difficulties, and minor annoyances with intense negative feelings that they could not soothe, and the less aware they were of exactly what they were feeling. When these two emotional tendencies were coupled with being highly dissatisfied with their body, the outcome was anorexia nervosa or bulimia nervosa. Overly controlling parents were found not to play a prime role in causing eating disorders. Also judged irrelevant were such popular explanations as fear of sexuality, early onset of puberty, and low self-esteem.

Instead, the causal chain this prospective study revealed began with the effects on young girls of growing up in a society that equates unnatural thinness with female

beauty. Well before adolescence, girls are already self conscious about their weight. In one study of 271 young teenagers, half the girls thought they were too fat, even though the vast majority of them were normal in weight. But the Minneapolis study showed that an obsession with being overweight is not in and of itself sufficient to explain why some girls go on to develop eating disorders.

Some obese people are unable to tell the difference between being scared, angry, and hungry and lump all those feelings together as signifying hunger, which leads them to overeat whenever they feel upset. Something similar seems to be happening in these girls. One psychologist who studied young girls and eating disorders, observed that poor awareness of one's feelings and body signals was the strongest single predictor that they would go on to develop an eating disorder within the next 2 years. Most children learn to distinguish among their sensations and can tell if they are bored, angry, depressed, or hungry, which is a basic part of emotional learning. But some have trouble distinguishing among their most basic feelings and may be unable to determine whether they are angry, or anxious, or depressed; they just experience a diffuse emotional storm that they do not know how to deal with effectively. They alleviate these symptoms with food. Other individuals, in an attempt to handle emotional confusion, refuse to eat at all. Anorexia nervosa can be a way to feel one has some control over these overwhelming feelings. The combination of poor inner awareness and weak social skills means that when upset by friends or parents these girls fail to act effectively to improve the relationship or to soothe their own distress. Instead their upset triggers bulimia nervosa, anorexia nervosa, or binge eating. Effective treatments for such girls need to include some remedial instruction in the emotional skills they lack. Researchers find that adolescents with eating disorders need to learn to identify their feelings and ways to soothe themselves or handle relationships better, without developing maladaptive eating habits.

Loneliness Socially rejected children typically are poor at reading emotional and social signals and may have limited repertoires for response. An inability to seize key cues is typical of children who are unpopular. Dropping out of school is a particular risk for children who are social rejects: between two and eight times greater than for children who have friends. One study found, for example, that about 25 percent of children who were unpopular in elementary school had dropped out before completing high school, compared to a general rate of 8 percent.

Two kinds of emotional proclivities lead children to become social outcasts. One is the propensity to angry outbursts and to perceive hostility even where none is intended. The second is timidity, anxiety, and shyness. Over and above these temperamental factors, it is children who are “off”—whose awkwardness repeatedly makes people uncomfortable—who tend to be shunned. When grade schoolers with few friends were asked to match an emotion such as disgust or anger with faces that displayed a range of emotions, they made far more mismatches than did children who were popular. When kindergarteners were asked to explain ways they might make friends with someone or keep from having a fight, it was the unpopular children—the ones others shied away from playing with—who came up with self-defeating answers (e.g., “Punch him” for what to do when both children wanted the same toy), or vague appeals to an adult for help. When asked to role-play being sad, angry, or mischievous, the more unpopular teenagers gave the least convincing performances. It is perhaps no surprise that such children believe that they are helpless to do any better at making friends; their social incompetence becomes a self-fulfilling prophecy. Instead of learning new ways to make friends, they simply keep doing the same things that have not worked for them in the past or make even more inept attempts.

In the lottery of liking, these children fall short on key emotional criteria: they are not seen as fun to be with and they do not know how to make another child feel good. Observations of unpopular children at play show that they are much more likely than others to cheat, sulk, quit when losing, or show off and brag about winning. Most children want to win at a game, but win or lose, most children are able to contain their emotions such that they do not undermine their relations with their friends.

Children who are socially tone-deaf—who continually have trouble reading and responding to emotions—end up as social isolates, but this does not apply to children who go through a temporary period of feeling left out. Those who are continually excluded and rejected feel their painful outcast status clinging to them as they continue their school years. The consequences of ending up at the social margins are potentially great as a child grows into adulthood. It is in the cauldron of close friendships and during the tumult of play that children refine the social and emotional skills that they will bring to relationships later in life. Children who are excluded from this realm of learning are inevitably at a disadvantage.

Understandably, those who are rejected report marked anxiety and are depressed and lonely. How popular a child was in third grade has been shown to be a better predictor of mental-health problems at age 18 than anything else—teachers' and nurses' ratings, school performance and I.Q., and even scores on psychological tests. Later in life people who have few friends and are chronically lonely are at greater risk for medical diseases and an early death.

Most people learn how to negotiate intimate relations—to work out differences and share their deepest feelings—in their first close friendships with same-sex friends. Children who are socially rejected are only half as likely as their peers to have a best friend during the crucial years of elementary school, and so miss out on one of the essential chances for emotional growth. One friend can make the difference even when all others turn away and even if that one friendship is not all that solid.

Friendship coaching sessions for unpopular children have shown some success. Third- and fourth-graders who were the least liked in their classes got six sessions on how to “make playing games more fun” through being “friendly, fun, and nice.” To avoid stigma, the children were told that they were acting as “consultants” to the coach, who was trying to learn what kinds of things make it more enjoyable to play games. The children were coached to act in ways typical of more popular children. For example, they were encouraged to think of alternative suggestions and compromises (rather than fighting) if they disagreed about the rules; to remember to talk with the other child while they play; to listen and look at the other child to see how he's doing; to say something nice when the other person does well; and to smile and offer help or suggestions and encouragement. The children also tried out these basic social skills while playing games such as Pick-up Sticks with a classmate, and were coached afterward on how well they did. This minicourse in getting along had a remarkable effect: a year later the children who were coached—all of whom were selected because they were the least liked in their class—were now solidly average in terms of classroom popularity: neither social stars nor outcasts.

Another program trains social outcasts to hone their ability to read and respond appropriately to other children's feelings. The children, for example, are videotaped while practicing expression of feelings such as happiness and sadness, and are coached to improve their emotional expressiveness. They then try out their newly honed skills with a child they want to make friends with. Such programs have reported a 50 to 60 percent success rate in raising the popularity of rejected children. As presently designed, these programs seem to work best for third- and fourth-graders rather than older children and were more helpful for socially inept children than for highly aggressive ones. But that is all a matter for fine-tuning; the hopeful sign is that many or most rejected children can be brought into the circle of friendship with some basic emotional coaching.

Substance Abuse In the United States the use of most substances among young people generally tapered off in the 1980s, but there is a steady trend toward more alcohol use at ever-younger ages. A 1993 survey found that 35 percent of college women said they drank to get drunk, as compared to just 10 percent in 1977; overall, one in three students drinks to get drunk. That poses other risks: 90 percent of all rapes reported on college campuses happened when either the assailant or the victim—or both—had been drinking. Alcohol-related accidents are the leading cause of death among young people between the ages of 15 and 24. Experimentation with alcohol and other substances might seem a rite of passage for adolescents, but this first taste can have long-lasting results for some. Most alcoholics and substance abusers can trace the beginning of their addiction to their teen years, though few of those who so experiment end up as alcoholics or substance abusers. By the time students leave high school, over 90 percent have tried alcohol, yet only about 14 percent eventually become alcoholics; of the millions of Americans who experimented with cocaine, fewer than 5 percent became addicted. What makes the difference?

One current theory is that those who stay with the habit, becoming increasingly dependent on alcohol or other substances, use these substances as a way to soothe anxiety, anger, or depression. Through their early experimentation they hit upon a chemical fix, a way to calm the feelings of anxiety or melancholy that have tormented them. Thus of several hundred seventh- and eighth-grade students tracked for 2 years, those who reported higher levels of emotional distress subsequently went on to have the highest rates of substance abuse. This may explain why so many young people are able to experiment with alcohol and other substances without becoming dependent, while others become dependent almost from the start: those most vulnerable to dependence seem to find in the alcohol or other substance an instant way to soothe emotions that have distressed them for years. For people who are biologically predisposed, the first drink or dose of a drug is immensely reinforcing; it stabilizes them physiologically, at least in the short term.

Certain emotional patterns seem to make people more likely to find emotional relief in one substance rather than another. For example, there may be two emotional pathways to alcoholism. Someone who was high-strung and anxious in childhood may typically discover as a teenager that alcohol can calm anxiety. Very often they are children—usually sons—of alcoholics who themselves have turned to alcohol to soothe their anxiety. One biological marker for this pattern is undersecretion of g-aminobutyric acid (GABA), a neurotransmitter that regulates anxiety: too little GABA is experienced as a high level of tension. One study found that sons of alcoholic fathers had low levels of GABA and were highly anxious, but when they drank alcohol, their GABA levels rose as their anxiety fell. These men drink to ease their tension, finding in alcohol a relaxation that they cannot seem to get otherwise. Such people may be vulnerable to abusing sedatives as well as alcohol for the same anxiety-reduction effect.

A neuropsychological study of sons of alcoholics who at age 12 showed signs of anxiety such as a heightened heart rate in response to stress as well as impulsivity found that the boys also had poor frontal lobe functioning. Thus the brain areas that might have helped ease their anxiety or control their impulsiveness helped them

less than the brain areas of other boys. The prefrontal lobes also handle working memory—which stores the consequences of various routes of action considered when making a decision—their deficit could cause these boys to slide into alcoholism by helping them ignore the long-term drawbacks of drinking, even as they found an immediate sedation from anxiety through alcohol. This craving for calm seems to be an emotional marker of a genetic susceptibility to alcohol abuse and dependence. A study of 1300 relatives of alcoholics found that the children of alcoholics who were most at risk for becoming alcoholics themselves were those who reported having chronically high levels of anxiety. Indeed, the researchers concluded that alcoholism develops in such people as "self-medication of anxiety symptoms."

A second emotional pathway to alcoholism comes from a high level of agitation, impulsivity, and boredom. This pattern shows up in infancy as being restless, cranky, and hard to handle; in grade school as being fidgety, hyperactive, and getting into trouble, a propensity that can push such children to seek out friends on the fringe and can sometimes lead to a criminal career or the diagnosis of antisocial personality disorder. Such people, who are usually men, cite agitation as their main emotional complaint; their main weakness is unrestrained impulsivity. Their usual reaction to boredom, which they often feel, is to search impulsively for risk and excitement. As adults, people with this pattern (which may be tied to deficiencies in two other neurotransmitters: serotonin and monoamine oxidase inhibitors [MAO]) find that alcohol can soothe their agitation. Not being able to stand monotony makes them ready to try anything; coupled with their general impulsivity, it makes them prone to abusing an almost random list of substances besides alcohol.

Depression can drive some to drink; the metabolic effects of alcohol often simply worsen the depression after a short reprieve. People who turn to alcohol as an emotional palliative do so much more often to calm anxiety than to alleviate depression; an entirely different class of drugs soothes the feelings of people who are depressed, at least temporarily. Feeling chronically depressed puts people at greater risk for dependence on stimulants such as cocaine, which provide a direct antidote to depression. One study found that more than half the patients being treated for cocaine addiction would have been diagnosed with major depressive disorder before they started their habit, and the deeper the preceding depression, the stronger the habit.

Chronic anger may lead to still another kind of susceptibility. In a study of 400 patients being treated for dependence to heroin and other opioids, the most striking emotional pattern was a lifelong difficulty handling anger and a quickness to rage. Some of the patients admitted that opiates made them finally feel normal and relaxed. Although the predisposition to substance abuse may in many cases be brain-based, the feelings that drive people to self-medicate themselves with alcohol or other substances can be handled without recourse to medication, as Alcoholics Anonymous and other recovery programs have demonstrated for decades. Acquiring the ability to handle those feelings—soothing anxiety, lifting depression, calming rage—removes the impetus to use drugs or alcohol in the first place. These basic emotional skills are taught remedially in treatment programs for drug and alcohol abuse; it would be far better if they were learned early in life, well before the habit became established.

PREVENTION

Over the last decade or so "wars" have been proclaimed in turn on teen pregnancy, dropping out, drugs, and violence. These campaigns come too late, after the targeted problem has reached epidemic proportions and taken firm root in the lives of the young. They are crisis intervention, the equivalent of solving a problem by sending an ambulance to the rescue rather than administering an inoculation that would prevent the disease in the first place. Instead of more such wars, the logic of prevention needs to be heeded; children need the skills to face life that will increase their chances of avoiding any of these pitfalls.

The emphasis on the role of emotional and social deficits does not deny the role of other risk factors, such as growing up in a fragmented, abusive, or chaotic family or in an poor, crime- and drug-ridden neighborhood. Poverty itself delivers emotional blows to children: poorer children at age 5 are already more fearful, anxious, and sad than their better-off peers, and have more behavior problems such as frequent tantrums and destructive spells, a trend that continues through the teen years. The press of poverty corrodes family life too: there tend to be fewer expressions of parental warmth, more depression in mothers (who are often single and jobless), and a greater reliance on harsh punishments such as yelling, hitting, and physical threats.

Emotional competence plays a role over and above family and economic forces—it may be decisive in determining the extent to which any given child or teenager is undone by these hardships or finds a core of resilience to survive them. Long-term studies of hundreds of children brought up in poverty, in abusive families, or by a parent who was severely mentally ill show that those who are resilient even in the face of the most grinding hardships tend to share key emotional skills. These include a winning sociability that draws people to them, self-confidence, an optimistic persistence in the face of failure and frustration, the ability to recover quickly from upsets, and an amiable personality. Many of these skills are innate—the luck of genes—but qualities of temperament *can* change for the better. One line of intervention is political and economic: alleviating the poverty and other social conditions that breed these problems; besides these tactics, which seem to move ever lower on the social agenda, there is much that can be offered to help children cope better with debilitating hardships.

One in two Americans experiences a mental disorder over the course of life. A study of a representative sample of 8098 Americans found that 48 percent suffered from at least one psychiatric problem during their lifetime. Most severely affected were the 14 percent who developed three or more psychiatric problems at once. This group was the most troubled, accounting for 60 percent of all psychiatric disorders occurring at any one time, and 90 percent of the most severe and disabling ones. Optimally, these problems should have been prevented in the first place. It is true that not every mental disorder can be prevented, but some, and perhaps many, can be prevented. Early intervention may help to keep one emotional problem from snowballing into a host of others—for instance, a social phobia in the sixth grade can be prevented from becoming a drinking problem in junior high school. By the late twenties, a child may have learned to alleviate her fears with substance abuse—alcohol, drugs, or both—and is liable to become depressed because her life is so messed up. What could have been done early in her life to have prevented this downward spiral?

The same holds true for dropping out, violence, or any of the perils faced by young people today. Educational programs to prevent one or another specific problem such as drug use and violence have proliferated wildly in the last decade, creating a mini-industry within the education marketplace. But many of these programs, including the most slickly marketed and most widely used, have proved to be ineffective. To the chagrin of educators, a few even seemed to increase the likelihood of the problems they were meant to head off, particularly substance abuse and teenage sex.

Active Ingredients A consortium of researchers studied outcome measures of prevention programs and distilled the active ingredients that seemed crucial to the success of those programs that worked. The list of key skills the consortium concluded should be covered, no matter what specific problem it is designed to prevent, reads like the ingredients of emotional intelligence: the skills include self-awareness; identifying, expressing, and managing feelings; impulse control and delaying gratification; and handling stress and anxiety. A key ability in impulse control is knowing the difference between feelings and actions, and learning to make better emotional decisions by first controlling the impulse to act, then identifying alternative actions and their consequences before acting. Many key competencies are interpersonal: reading social and emotional cues, listening, being able to resist negative influences, taking others' perspectives, and understanding what behavior is acceptable in a situation.

These are among the core emotional and social skills for life, and include remedies for most, if not all, of the difficulties discussed in this chapter. The specific problems these skills might prevent are nearly arbitrary; emotional and social skills could also prevent problems such as teen pregnancy and teen suicide.

The causes of all such problems are complex, interweaving differing ratios of biological destiny, family dynamics, the politics of poverty, and the culture of the streets. No single intervention can claim to do the whole job. However, to the degree that emotional deficits add to a child's risk, attention must be paid to emotional remedies, not to the exclusion of other answers, but in concert with them.

School-Based Primary Prevention Pilot interventions designed to target the specific deficits in emotional and social skills that underlie problems such as aggression or depression have been found effective, but have rarely been implemented in schools, where they could reach the maximum number of children. Emotional intelligence skill-building should be used as a primary prevention strategy; the lessons learned from such programs should be generalized and used as a preventive measure for the entire school population and taught by ordinary teachers. This more sophisticated and more effective approach to prevention should include information on acquired immune deficiency syndrome (AIDS), substance abuse, and the like, at the points in youngsters' lives when they are beginning to face these problems. Bringing emotional literacy into schools makes emotions and social life themselves topics, rather than treating these most compelling facets of a child's day as irrelevant intrusions or, when they lead to eruptions, relegating them to occasional disciplinary trips to the guidance counselor or the principal's office.

A list of the contents of an emotional literacy course almost exactly matches the core skills recommended as primary prevention for the range of pitfalls threatening children. The topics taught include self-awareness (recognizing feelings and building a vocabulary for them, and seeing the links between thoughts, feelings, and reactions), distinguishing between thoughts or feelings when making a decision; seeing the consequences of alternative choices; and applying these insights to decisions about such issues as drugs, smoking, and sex. Self-awareness also takes the form of recognizing one's strengths and weaknesses, and seeing oneself in a

positive but realistic light.

Another emphasis is on managing one's emotions: realizing what is behind a feeling (e.g., the hurt that triggers anger), and learning ways to handle anxieties, anger, and sadness. Taking responsibility for decisions and actions, and following through on commitments, is also emphasized. A key social ability is empathy, understanding others' feelings and taking their perspective, and respecting differences in how people feel about things. Relationships are a major focus, including learning to be a good listener and question-asker; distinguishing between what someone else says or does and one's own reactions and judgments; being assertive rather than angry or passive; and learning the arts of cooperation, conflict resolution, and compromise negotiation.

Typical emotional literacy curricula include, for example, learning in the earliest school years to control impulses; children who lack this ability have special trouble paying attention and so fall behind in their learning and grades. Another is recognizing their feelings; one curriculum has fifty lessons on different emotions, teaching the most basic, such as happiness and anger, to the youngest children, and later touching on more complicated feelings such as jealousy, pride, and guilt. The emotional awareness lessons include how to monitor what they and those around them are feeling, and—most important for those prone to aggression—how to recognize when someone is actually hostile, as opposed to projecting one's own hostility onto others.

Parents can benefit from being coached to be emotional mentors to their infants and toddlers, as some home-visit programs do. A strong argument can be made for emphasizing social and emotional skills more systematically in preschool programs such as Head Start; children's readiness to learn depends to a large extent on acquiring some of these basic emotional skills. The preschool years are crucial ones for laying foundation skills, and there is some evidence that, when run well, Head Start can have beneficial long-term emotional and social effects on the lives of its graduates even into their early adult years: fewer drug problems and arrests, better marriages, and greater earning power.

Such interventions work best when they track the emotional timetable of development. The newborn's repertoire of feeling is primitive compared to the emotional range of a 5-year-old, which, in turn, is limited compared to the fullness of a teenager's feelings. The timetable for emotional growth is intertwined with allied lines of development, particularly for cognition and brain and biological maturation. Emotional skills such as empathy and emotional self-regulation start to develop virtually from infancy. The kindergarten year marks a peak ripening of the “social emotions”—insecurity and humility, jealousy and envy, pride and confidence—all of which require the ability to compare oneself with others. On entering the wider social world of school, the 5-year-old also enters the world of social comparison.

David Hamburg, who has evaluated some pioneering emotional-education programs, sees the years of transition into grade school and then again into junior high or middle school as marking two crucial points in a child's adjustment. From ages 6 to 11 Hamburg says:

> School is a crucible and a defining experience that will heavily influence children's adolescence and beyond. A child's sense of self-worth depends substantially on his or her ability to achieve in school. A child who fails in school sets in motion the self-defeating attitudes that can dim prospects for an entire lifespan.

Among the essentials for profiting from school, Hamburg notes, are an ability “to postpone gratification, to be socially responsible in appropriate ways, to maintain control over their emotions, and to have an optimistic outlook”—in other words, emotional intelligence.

The transition to middle school or junior high marks an end to childhood, and is itself a formidable emotional challenge. As they enter this new school arrangement virtually all students experience a dip in self-confidence and a jump in self-consciousness; their very notions of themselves are rocky and in tumult. One of the greatest specific blows is in social self-esteem or confidence that they can make and keep friends. It is at this juncture, Hamburg points out, that it helps immensely to buttress boys' and girls' abilities to build close relationships, navigate crises in friendships, and nurture their self-confidence.

Hamburg notes that as students enter middle school, just on the cusp of adolescence, there is something different about those who have taken emotional literacy classes: they find the new pressures of peer politics, the upping of academic demands, and the temptations to smoke and use substances less troubling than do their peers. They have mastered emotional abilities that, at least for the short term, inoculate them against the turmoil and pressures they are about to face.

As family life no longer offers growing numbers of children a sure footing in life, schools are the one place communities can turn to for correctives to children's deficiencies in emotional and social competence. Since virtually every child goes to school (at least at the outset), school is the place to reach children with basic lessons for living that they may never get otherwise. Emotional literacy implies an expanded mandate for schools, taking up the slack for failing families in socializing children. This daunting task requires two major changes: that teachers go beyond their traditional mission and that people in the community become more involved with schools.

Beyond teacher training, emotional literacy expands our vision of the task of schools themselves, making them more explicitly society's agent for seeing that children learn these essential lessons for life—a return to a classic role for education. Apart from any specifics of curriculum, this enhanced role requires using opportunities in and out of class to help students turn moments of personal crisis into lessons in emotional competence. It also works best when the lessons at school are coordinated with what goes on in children's homes. Many emotional literacy programs include special classes to teach parents about what their children are learning, not just to complement what is imparted at school, but to help parents who want to deal more effectively with their children's emotional growth. The data suggest that although such courses do not change anyone overnight, as children advance through the curriculum from grade to grade, discernible improvements are seen in the emotional competence of the girls and boys who take them (Table 3.6-1).

Table 3.6-1 Outcomes: Social and Emotional Learning

Emotions and Physical Health There is another potential payoff for teaching children the basic skills of emotional intelligence: improved physical health through life. Research over the last two decades has yielded evidence of the emotions' effect on physical health. Medical science could benefit greatly from a broader knowledge and application of recent findings on the myriad ways emotions affect health and medical care.

A new area of medical research, *psychoneuroimmunology* investigates links between the immune system and the central nervous system. Recent research in psychoneuroimmunology suggests that emotions have powerful effects on the autonomic nervous system. The chemical messengers most crucial to the brain and the immune system have been found to be those that are also most dense in neural areas that regulate emotion. There are many physical meeting points where the autonomic nervous system communicates directly with lymphocytes and macrophages in the immune system.

Psychoneuroimmunology is based on a finding in 1974 by Robert Ader, who demonstrated that the immune system, like the brain, is capable of learning. Ader gave white rats a medication that artificially suppressed the quantity of T cells in their blood. The medication was given with saccharin-laced water. When the rats were given saccharin-flavored water alone, without the suppressive medication, their T-cell counts still dropped, to the point that some of the rats died; their immune systems had learned to suppress T cells in response to the sweetened water.

Until Ader's findings the prevailing wisdom had been that only the brain and central nervous system could respond to experience by changing (learning). Ader's finding led to the discovery of the myriad ways in which the central nervous system and the immune system are interconnected: biological pathways that link the mind, emotions, and body.

Research on heart patients has revealed that anger has a significant impact on heart function. When patients were asked to recount incidents that made them mad, the pumping efficiency of their hearts dropped by 5 percent. Some showed a drop in pumping efficiency of 7 percent or greater, thereby falling into a range indicating myocardial ischemia. The heart's pumping efficiency was not affected by other negative feelings such as anxiety, nor during physical exertion. This finding supports a larger network of evidence that suggests that anger can damage the heart. The theory that a hurried, high-pressure type A personality runs a greater risk of developing heart disease has become outdated, but hostility has been found to put the heart at risk.

Another study suggests that a proclivity towards anger is a better predictor of dying young than are other risk factors such as smoking, high blood pressure, and high cholesterol. Physicians who had scored highest on a test of hostility while still in medical school were seven times as likely to die by the age of 50 than those with low hostility scores. Other research has linked an inclination toward hostility in heart patients with the extent and severity of coronary artery disease.

Whereas proclivities toward chronic hostility and anger appear to put men at greatest risk for heart disease, for women the more risk-laden emotion may be anxiety. Research at Stanford University School of Medicine found that of women who had suffered a first heart attack, those who went on to suffer a second heart attack were marked by high levels of fearfulness and anxiety. In many cases the fearfulness took the form of crippling phobias: after their first heart attack the patients stopped driving, quit their jobs, or avoided going out.

An attitude of pessimism has been found to have strong negative effects on patients' recovery, and optimism has been found to offer corresponding benefits. In one study, men who had had a heart attack were evaluated on their degree of optimism or pessimism. Eight years later, of the 25 most pessimistic men, 21 had died; of the 25 most optimistic, just 6 had died. Their mental outlook proved a better predictor of survival than any medical risk factor, including the amount of damage to the heart in the first attack, artery blockage, cholesterol level, or blood pressure. Another study found that artery-bypass surgery patients who were optimistic had a faster recovery and fewer medical complications during and after surgery than did pessimistic patients.

Emotional support offers patients observable and sometimes life-saving benefits. Of bone marrow transplant patients who reported having strong emotional support from spouses, family, or friends, 54 percent were still alive after 2 years, versus just 20 percent of those who reported having little emotional support. Similarly, elderly heart-attack patients who have emotional support are more than twice as likely to survive longer than a year after an attack than are people with no such support.

Emotionally Intelligent Medicine When anger and anxiety are chronic, people are more susceptible to a range of disease. Although depression may not make people more vulnerable to ill health, it does seem to impede medical recovery and heighten the risk of death, especially in frailer patients who are seriously ill.

Medicine's inattention to the impact of emotions on illness neglects a growing body of evidence that indicates that emotions can play a significant role in vulnerability to disease and recovery from disease. Modern medical care too often lacks emotional intelligence. There are many ways medicine can incorporate the new knowledge of the impact of emotions on health into its view of patient care, which could be a new focus for consultation-liaison psychiatry.

SUGGESTED CROSS-REFERENCES

Psychoneuroimmunology is discussed in Section 25.10 on behavior and immunity. Mood disorders and anxiety disorders are covered in Chapter 14 and Chapter 15, respectively. Antisocial behavior, aggression, and violence are covered in Section 27.3. Alcohol and other substance abuse is covered in Chapter 11. School consultation is presented in Section 49.14 and psychiatric prevention in children, in Section 49.15.

SECTION REFERENCES

*Gardner H: *Multiple Intelligences: The Theory in Practice*. Basic Books, New York, 1993.

*Goleman D: *Emotional Intelligence*. Bantam Books, New York, 1995.

Gottman J: *What Predicts Divorce: The Relationship between Marital Processes and Marital Outcomes*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1993.

*Hawkins JD, Catalano RF, Kosterman R, Abbott R, Hill G: Preventing adolescent health-risk behaviors by strengthening protection during childhood. Arch Pediatr Adolesc Med *153:*213, 1999.

Kagan J: *Galen's Prophecy*. Basic Books, New York, 1994.

LeDoux J: *The Emotional Brain*. Simon & Schuster, New York, 1996.

*Martinez-Pons M: Parental inducement of emotional intelligence. Imagination Cogn Pers *18:*3, 1999.

Salovey P, Mayer JD: Emotional intelligence, imagination, cognition, and personality *9:*185, 1990.

*Salovey P, Sluyter DJ, editors: *Emotional Development and Emotional Intelligence*. Basic Books, New York, 1997.

*Sternberg RJ: *Beyond I.Q*. Cambridge University Press, New York, 1985.

Weissberg RP, Gullotta TP, Hampton RL, Ryan BA, Adams GR, editors: *Establishing Preventive Services*. Sage Publication, New York, 1997.

CHAPTER 4. CONTRIBUTIONS OF THE SOCIOCULTURAL SCIENCES

4.1 ANTHROPOLOGY AND PSYCHIATRY

ANNE E. BECKER, M.D., PH.D., AND ARTHUR KLEINMAN, M.D.

The fields of social and cultural anthropology comprise the study of various aspects of society, such as kinship, social exchange, and religious experience in the context of politics, economics, and large symbolic systems (e.g., language, aesthetic systems) in order to describe and interpret how the social world is ordered and how it responds to the disordering effects of historical change. Social and cultural anthropology include studies of local worlds—through collection and analysis of ethnographic data—and their cross-cultural comparison. Medical anthropology, a robust and expanding subfield of social anthropology, includes the comparative study of healing systems and the social, economic, political, and cultural contexts of health and health care. This field also encompasses the study of how social processes produce health effects and how cultural processes construct illness experience. This chapter reviews the major contributions of medical anthropological perspectives to psychiatry; these include a set of concepts that enhances communication and service delivery in cross-cultural settings as well as a conceptual framework alternative to biomedical theories of health and disease.

Unfortunately, much of the early interest in cross-cultural studies focused on exotic healing rituals and culture-bound diseases, thereby obscuring anthropology's theoretical and practical relevance to everyday problems and practices. The field's contributions to general medicine—for example, infectious diseases, chronic medical disorders, disability, and public health—have arguably become increasingly visible and relevant since the late 1970s. Whereas the recently emerging paradigm of neurobiological psychiatry increasingly locates mental illness in the brain and favors a universalizing theory of psychiatric illness, anthropology engages the diversity of human experience. Indeed, it is through a cross-cultural lens that universals and diversity in human experience emerge, lending insight into nonbiological factors that affect the pathogenesis, phenomenology, course, response to treatment, outcome, and lived experience of illness.

Anthropology and psychiatry have been historically linked since the late nineteenth century when W.H.R. Rivers, a psychiatrist, was joined by several other physicians on an expedition to the Torres Straits. That research trip to the South Pacific was the source of much empirical research in British social anthropology. Rivers' ethnographic and clinical expertise converged not only in his cross-cultural studies on religious healing and on pain sensation but also in the form of psychotherapy that he later developed to treat soldiers with shell shock during World War I. Cross-fertilization between the fields continued into the next century within the culture and personality school, where studies based on a psychoanalytic conceptual framework and popularized by the American anthropologists, Margaret Mead and Ruth Benedict, brought leading anthropologists and psychiatrists into a productive collaboration. Similarly, French sociologist Emile Durkheim's landmark study on suicide drew attention to the social underpinnings of seemingly individual behavior, and George Herbert Mead, the American sociologist who created the symbolic interactionist theory, created a social science approach to the study of the ways in which personhood and self processes emerge from the social context of language use.

Subsequent interface between the fields explored the phenomenological diversity of mental illness transculturally, with particular interest in culture-bound syndromes and culturally sanctioned altered states of consciousness, such as trance and possession states. These syndromes and states fascinated psychiatrists who were interested in so-called primitive culture (an idea now thoroughly discredited), yet were of only marginal theoretical or practical interest to the clinical practice of psychiatry in Western settings. As the dominant biological paradigm for mental illness emerged, however, so did a keen interest in the cross-cultural epidemiology of psychiatric disorders; indeed, recognition that certain mental disorders appear across cultures strengthened the evidence of their putative biological substrate.

Research in the cross-cultural epidemiology of mental illness has posed theoretical and methodological challenges, calling for the serious engagement of anthropological theory and methodology by the field of psychiatry in addressing the complex interplay of local biologies; the cultural patterning of experience and behavior; and the social, political, and economic contexts of disease. This application of an anthropological perspective to mental illness has illuminated a dialectic between human diversity and universals in light of the cross-cultural data.

The relevance of anthropological contributions to psychiatry is most apparent in three critical areas. First, from a theoretical standpoint, the documentation of particular mental disorders in different sociocultural contexts provides evidence for their syndromal significance and integrity in the absence of biological markers. Second, from a clinical standpoint, the elicitation of explanatory models from and the engagement of the growing number of patients from culturally and ethnically diverse backgrounds play an essential role in formulating illness and responding to it in culturally appropriate ways. Finally, from a public health perspective, understanding the social production and social course of mental illness suggests means of improving outcome or even preventing illness.

CULTURAL PSYCHIATRY

Numerous studies have sought to demonstrate the cross-cultural epidemiology of major mental illness, both to demonstrate the universality of particular syndromes and to investigate how prevalence and phenomenology may differ relative to cultural context. The interpretation of findings, however, has been limited by major methodological pitfalls in conducting transcultural research and a persistent tendency to interpret data as evidence of universality rather than of heterogeneity of disorders. Indeed, such studies have commonly stemmed from the presumption that discrete psychiatric syndromes are identifiable and comparable across varying cultural contexts. This presumption risks creating a *category fallacy*—that is, applying a category that is valid in one cultural context (in this case in Western psychiatric nosology) to a culture in which the category has no local relevance or diagnostic validity. For example, a category fallacy would occur if a Malaysian healer were to study *latah* in Boston. This healer might find fright, echopraxia, or other symptoms associated with *latah* in Malaysia, but the recognition of these symptoms as a particular syndrome with local meaning would be absent in the Boston context. The same caution would need to be exercised by a Western psychiatric researcher in investigating the prevalence of maternity blues, for instance, in Fiji where symptoms of tearfulness and anxiety may be identified in the immediate postpartum period but are not locally marked as an illness. This universalizing approach to exploring the meaning of illness and behavior transculturally contrasts with a *cultural relativist* perspective that posits that cultural values and meanings cannot be measured against a universal system but rather are relative to and fundamentally embedded in their cultural context. For example, particular behaviors (such as trance or possession states) may be judged as deviant in one society but as acceptable in another. Similarly, diverse grieving practices suggest that the duration and form of appropriate bereavement can only be understood within a particular cultural context.

To minimize the risk of creating a category fallacy, it is helpful to draw from the anthropologic concepts of *etic* and *emic* levels of analysis in investigating psychiatric disorders transculturally. *Etic*, in this case, refers to the application of Western categories to another cultural context, whereas *emic* refers to the use of locally meaningful categories. Successful research identifies indigenous syndromes and describes their phenomenology and associated local explanatory models from an ethnographic perspective. Comparison can then be made with Western psychiatric categories of disorder to establish or refute their validity locally. However, translation into the comparative cross-cultural discourse of international psychiatry should be the last, not the first, step, which unfortunately goes against the tide of much psychiatric research.

It is insufficient to use standardized assessment instruments (even if they are translated, backtranslated, and further analyzed and revised) without ensuring that they measure locally valid constructs. Anthropologists have argued that emotions, such as a depressed or anxious mood, are themselves culturally constituted and that the ways in which feelings are experienced, processed, and communicated are constrained and evolved by the cultural milieu. For example, depressed mood among the Flathead Indians of Montana is not only an indicator of distress but is experienced as a positive symbol of belongingness to the community. Among the Ifaluk in the South Pacific sadness and anger are defined as a single, complex emotional state. Soul loss is a moral-emotional state in many non-Western societies. The local meanings attributed to emotional experience and the associated somatic experience may then be recognized and labeled as illness or not, depending on the context. For this reason, assessment instruments are ideally constructed or modified to include culturally local affective terms or idioms incorporating an *emic* approach.

Several investigators have also demonstrated that a cultural response style may skew data toward either an underreporting or an overreporting of symptoms. For example, among Puerto Ricans, one study found that there was an increased tendency to respond in the affirmative to symptoms on a symptom checklist, both because of an increased willingness to report symptoms as compared with other ethnic groups and because the symptoms were judged to be less socially

stigmatizing. The latter may have been true because the symptoms overlapped with culturally appropriate means of expressing distress (ie., *nervios*). Hence, cultural norms in responding to questionnaires as well as within interview settings must be factored into both the development of appropriate instruments and interview questions and techniques and the interpretation of the data in order to enhance the validity of cross-cultural comparisons.

These problems are somewhat attenuated when local ethnopsychological or ethnomedical theories are used to establish illness or emotional categories relevant to the cultural group being studied. For example, the revised second edition of *Chinese Classification of Mental Disorders* (CCMD-2-R) represents an attempt to include Chinese mental illness nosology within a Western psychiatric nosology. The CCMD-2-R includes such illness categories as *shenjing shuairuo* (neurasthenia) and *qiqong*, induced mental disorder—and train or travel psychosis is a candidate diagnosis—while excluding certain Western disorders that hold no local relevance, such as somatoform disorders, pathological gambling, and some personality disorders.

The potential for ethnocentrism in evaluating the presence of psychopathology cross-culturally is particularly heightened in the application of Western constructs of personality. The local cultural context (including means of socialization into language use, child-rearing, adolescent initiation, opportunities for individuation and achievement) not only shapes personality development through culturally specific notions of selfhood, but also determines relative definitions of character pathology. Concepts of self in Western societies are regarded by cultural psychologists and psychological anthropologists as rather intensively *egocentric* in contrast with concepts of selfhood in the majority of other societies, which have been characterized as relatively much more *sociocentric*. In the latter, individual identity is tied to a myriad of social relationships in contrast to personal attributes. Sociocentric societies place a premium on social harmony rather than individual needs or desires. In the same vein, in contrast with Western societies in which status is ideologically *achieved*, many societies *ascribe* status, changing and constraining the landscape of behavioral motivation. Behavior that is undesirable or aberrant in Western contexts is considered appropriate in such contexts. Conversely, behavior that is supported in Western cultures, such as self-assertion or self-cultivation, is potentially problematic or disruptive in a sociocentric society in which status is ascribed.

However, as recent social theory argues, care must be taken in using these categories of cross-cultural difference for several important reasons. First, they tend to represent ideological positions, not necessarily everyday practice. For example, even though Americans, French, Italians, and Greeks may emphasize egocentricity, in actual practice these groups also show considerable sociocentricity. Second, societies are diverse; experience may be (and often is) different owing to gender, age-cohort, ethnic, religious, and class differences. Third, globalization is proceeding at an astonishing rate. Along with changes in dress, diet, entertainment, and lifestyle, the global media and political economy are also changing collective and subjective orientations and practices. A good example of this is the spread of Western fashion ideals and disorders such as anorexia nervosa to areas like China where neither the ideal nor the pathology were previously found.

The bias in cross-cultural epidemiological research in investigating and reporting universals rather than the heterogeneity of illness presentations is exemplified in the International Pilot Study of Schizophrenia (IPSS) and the Determinants of Outcome Study sponsored by the World Health Organization. The IPSS attempted to document core schizophrenic symptoms in diverse cultural contexts. Although it succeeded in establishing clusters of similar symptoms across a variety of Western and non-Western contexts, its rigorous inclusion and exclusion criteria effectively selected for homogeneity and against heterogeneity across cases. Despite this selection bias, the IPSS documented that the course of schizophrenia differed cross-culturally. However, this finding was deemphasized relative to the demonstration of core schizophrenic symptoms across cultures. Similarly, although data generated from the Determinants of Outcome Study showed impressive variation in the cross-cultural prevalence of subtypes of schizophrenia, the authors of the study chose to highlight cross-cultural similarities.

This selective attention to universals in cross-cultural psychiatric research is arguably propelled by the field's enthusiasm for establishing the neuroanatomic and neurophysiological bases of mental illness. The common characterization of mental illnesses as disorders of mind arising in the brain reflects the dominant biological paradigm in the field of psychiatry that mental disorders are grounded in the brain whereas apparent cross-cultural differences are considered mere epiphenomena. This enthusiasm remains untempered by data that continue to support the conclusion that environmental effects can override biological predisposition to illness. Notwithstanding the important advances in neuroimaging, pharmacotherapy, and genetic research that validate a biological understanding of mental illness (albeit in the absence of any biological markers for schizophrenia, depression, or anxiety disorders), the professional culture of psychiatry is itself subject to social pressures (e.g., pharmaceutical industry funding or a managed-care environment that provide incentives for pharmacological management of mental illness) that support a biologically reductionistic approach to mental illness.

Cultural differences in psychiatric illness have been commonly understood as superimposed on biological substrates. Within this model of biological primacy, culture has been conceptualized as contributing only to the particular local manifestations of an underlying disorder rather than as being inherent to the very substrate of the disorder. Anthropological research challenges this perspective by providing data on the dialectical interplay of psychobiological and sociocultural processes in the prevalence, phenomenology, course, and outcome of mental illness.

Culture of Biomedicine The professional culture of biomedicine has also been a focus of inquiry by anthropologists. As within any culture, locally held values and meanings are powerful forces in socializing individuals within the culture—in this case, clinicians. In the course of professional training—medical school, internship, and residency—clinicians learn to recognize and formulate physiological and psychological signs and symptoms into patterns that correspond to culturally salient nosological categories. In Medicine, Rationality, and Experience, Byron Good describes how students of biomedicine are taught to write and speak in ways that construct sick individuals as patients—ways, incidentally, that generally de-emphasize an individual's experience of illness. This reduction of an individual's illness experience to a diagnostic category is also supported within the professional culture of psychiatry for a variety of reasons, including the current dominance of the biological paradigm and the necessity of documenting a diagnosis from the fourth edition of Diagnostic and Statistical Manuel of Mental Disorders (DSM-IV) for reimbursement by third-party payers. The very organization of DSM-IV reflects the professional marginalization of ethnic and cultural dimensions of illness presentation and experience insofar as an outline for the cultural formulation of psychiatric illness is relegated to its ninth appendix.

Culture-Bound Syndromes *Culture-bound syndromes* are illness phenomena that appear to be restricted to a particular cultural context. Numerous culture-bound syndromes that encompass a wide range of psychological and somatic symptoms have been described around the globe (Table 4.1-1). Some examples include acute episodes of psychosis and violence such as *amok* (Malaysia) and *boufée déireante* (West Africa or Haiti), syndromes with prominent anxiety such as *ataques de nervios* (Latin America and the Caribbean), *koro* (Malaysia), or *shin-byung* (Korea), and dissociative events such as *pibloktoq* (Eskimo communities) and *qiqong* induced mental disorder (China).

Table 4.1-1 A Selection of Culture-Bound Syndromes* by Region or Ethnic Group

There has been a tendency within the field of Western psychiatry to interpret these syndromes as exotic manifestations or *formes frustes* of underlying universal disorders rather than as true culturally particular syndromes. Although somewhat controversial, it is illuminating to characterize certain disorders such as chronic fatigue syndrome and dissociative identity disorder that appear to be culturally specific to North American settings as culture-bound syndromes as well. In doing so, it becomes clearer that certain patterns of emotional and somatic experience and behavior are selected and understood as locally meaningful *idioms of distress* within their particular cultural contexts. Particular symptoms are recognized as having salience in signaling distress and channeling experience, thereby organizing a culturally mediated and specific response to distress. Individuals are socialized to attend to and interpret experience in locally meaningful ways, resulting in the culturally distinctive patterning of symptoms and attribution of relevance. This way of looking at illness is now widely accepted by medical anthropologists, whereas for many anthropologists the idea of creating a separate category of culture-bound disorders has come to represent the problem of essentialism in psychiatry, (i.e., a cultural difference is reified as primordial and definitive of some deeply natural source of group difference, when in fact what is seen as a culture is itself a blend of

subgroups and individual differences).

Cultural norms interact with pathophysiological processes not only to pattern illness experience but to sanction or discourage expression of distress in particular ways. However, they do so in ways that are less definitive than earlier believed, and in ways that are modified by intracultural and individual diversity. It is crucial to avoid unidimensional cultural or ethnic stereotypes. In societies that value autonomy, independence, and the pursuit of self-interests (such as in European-American culture), the direct expression of distress is permissible, even encouraged, in the articulation of negative or dysphoric emotional states and often in a psychological idiom; by contrast, in societies in which interdependence and support of familial or community harmony is valued (such as in many Asian societies), distress may be preferentially articulated in an indirect or somatic idiom. However, many European-Americans do not express distress directly and do somatize, and some non-Westerners have traditionally expressed emotion directly and do not somatize. Indeed, using the body as an idiom of distress is a more or less universal social psychological process.

Somatization Research has demonstrated that the somatic expression of distress occurs across all cultural groups studied, although its particular form varies with the local cultural context. The widespread notion that it is a less sophisticated form of distress reflects the western cultural predilection in the second half of the twentieth century to communicate distress in a psychological idiom; it is also a faulty and racist ideological premise in Western thought. In their review of culture and somatization, Laurence J. Kirmayer and Allan Young comment that:

Equating somatization with primitive thinking or communication is based on several doubtful premises: that greater psychological insight necessarily results in fewer somatic symptoms; that psychological idioms are inherently more “advanced” than somatic idioms; that somatic idioms are “less differentiated” and inarticulate; and, finally, that psychological (and by analogy, cultural) development moves along a one-dimensional continuum from primitive to advanced. In fact, the use of a bodily idiom may say more about the social and cultural context of communication than it does about the cognitive limitations of psychological defensiveness of the individual.

What this view fails to mention is the racist potential for deploying this theory in clinical settings; this has actually happened in the way in which ethnic minorities have been labeled by an earlier generation of psychiatrists in America.

A somatic idiom of distress may also be preferentially invoked in contexts in which mental illness is seen as a stigma or in which mental health care is not readily available. Similarly, in contexts in which emotional distress is not culturally marked, the somatic expression of distress can be understood as a means of engaging the social milieu and mobilizing support within the local cultural context. Finally, embodied distress may also have a moral context, signaling complaint, resistance, or profound demoralization in a political or social milieu in which direct articulation is an impossibility.

SOCIOSOMATIC VERSUS PSYCHOSOMATIC ILLNESS Distress may be preferentially articulated in a somatic lexicon in contexts in which a psychological idiom does not have legitimacy or in which embodied illness provides a metaphor for social disharmony. Such an idiom may rapidly mobilize social supports without jeopardizing harmonious social relations or political safety. For example, weight loss (“going thin”) indexes social loss or marginalization in Fiji. Likewise, *na tadoka ni vasucu*, a postpartum somatic syndrome, is attributed to inappropriate physical or emotional exertion by a new mother in Fiji, and is an indirect marker of social neglect. Although the syndrome, which consists of nonspecific aches and pains, commonly coexists with depressed mood, depressed mood is rarely articulated whereas this syndrome rapidly allows the new mother to elicit social supports in the postpartum period. Similarly, individuals suffering from neurasthenia in China have commonly attributed their illness to the political and social upheaval caused by the Chinese Cultural Revolution. Core symptoms of the illness, such as exhaustion and dizziness, resonated with the exhausting and oppressive political changes of that period and became a bodily metaphor for social disorder.

In these examples, pain and suffering are transduced from their social, economic, and political sources and are experienced as physical symptoms. It is arguably most appropriate to understand somatization in such instances as *sociosomatic*—that is, emanating from social contexts—rather than as psychosomatic because it is easier to identify social processes that relate to mind-body processes and because so many psychological formulations of psychosomatic illness ignore the real world of debts, obligations, threats, and opportunities, replacing it with theories like alexithymia that appear to be dubious because they forget that individuals live within a complex social context.

Cultural Construction of Illness Categories It is noteworthy that while the syndromic clustering of particular symptoms and the local attributions of the cause of the various culture-bound syndromes are unfamiliar to Western psychiatric nosology, the symptoms—including anxiety, depressive, dissociative, and psychotic—generally are not. The local interpretation of these symptoms and their formulation as syndromes is culturally constructed according to their salience in their particular context. These syndromes derive their significance from their immediate cultural context and do not necessarily correspond to the illness categories of Western nosology.

It is helpful to depart from the notion that biomedical illness categories described in DSM-IV are somehow more “real” than folk categories that are culture specific. In Medicine, Rationality, and Experience, Good argues that, “medicine formulates the human body and disease in a culturally distinctive fashion.” In other words, medicine does not simply describe biological realities but rather constitutes a particular perspective in attributing meaningfulness to symptoms and symptom clusters. Inasmuch as Western psychiatric nosology organizes disorders categorically despite considerable phenomenological overlap among syndromes, a continuum of severity, high levels of comorbidity among disorders, and the clustering of disorders with particular patterns of social problems, the boundaries between disorders reflect a schema that is culturally derived despite being based in objective symptoms. Both Western psychiatric and culture-bound syndromes are somewhat culturally constructed in that they are characterized by contextual validity and reflect a culturally specific means of organizing symptoms into syndromic significance—this is anthropological commonsense.

Cross-Cultural Variation in Psychiatric Illness Anthropologists and psychiatrists alike have been interested in how illnesses that appear to represent similar processes differ with respect to their cultural contexts. These differences have often been dismissed as relatively superficial influences on, for example, the content of delusions or hallucinations. However, substantive cross-cultural differences have been demonstrated in phenomenological variation of mental illnesses as well as in their prevalence and course. Transcultural comparison can therefore illuminate ways in which the sociocultural environment shapes disease processes, and it is as fundamental a way of knowing as an experiment in the laboratory.

Cross-Cultural Prevalence of Major Mental Disorders The major mental disorders—anxiety, mood, and psychotic disorders—have been identified around the globe but with a wide variation in prevalence. However, methodological constraints have continued to complicate the understanding of their distribution. For instance, cross-national epidemiological research had demonstrated wide variation in rates of major depressive disorder transculturally. Investigators have reported that depression was virtually unreported in China prior to 1981, whereas it has been increasingly reported in the past two decades. Similarly, depression has been documented to be on the increase in a sample of Taiwanese over 15 years. These data suggest that rates of depression may have increased with social risk factors (which almost certainly has happened), that depression has become an increasingly legitimate idiom of distress (which also appears to be the case), or that psychiatric practice may have changed (which again seems true). Thus, a change in prevalence rate is a measure not of one simple thing, but of several processes—cultural, social, and professional.

In one study on the cross-national epidemiology of major depressive disorders and bipolar I disorder, Myrna Weissman and her colleagues reported that lifetime rates of major depressive disorder ranged from 1.5 cases per 100 in Taiwan to 19.0 cases per 100 in Beirut, Lebanon. The authors drew the conclusion that “cultural differences of different risk factor profiles across countries may affect the expression of the disorder.” Indeed, the results of this study imply that either the prevalence of depression varies with culture-specific risk factors or that depression presents differently (even unrecognizably) across cultural contexts. In spite of the simplification of their analytic dichotomy and of particular interest, they also found that in most of the samples studied individuals who were divorced or separated had significantly higher rates of depression than those who were married. These data suggest that marriage may be a virtually universal marker of social support that confers significant protection against depression among the nations studied.

Also, because dysphoric mood was required to meet the criteria of major depressive disorder in this study it is quite likely that locally salient depressive syndromes may have been overlooked (e.g., those in which anhedonia was prominent), so that depression or such local depressive syndromes were underreported. Notably, this study found that the lifetime rates of bipolar I disorder were more consistent across the samples studied, ranging from 0.3 cases per 100 in Taiwan to 1.5 cases per 100 in New Zealand. This intriguing finding may relate to the local cultural response to a manic episode as socially disruptive and thus marked as deviant in contrast to a depressive episode, which may not be expressed in a way that marks it as mental illness. Thus, symptoms may or may not be recognized and labeled as pathological across differing cultural contexts. Dysphoric affect, for example, may be locally understood as emanating from demoralization or grief. This may engender a response from the individual or family to address its existential meanings, but not necessarily to seek medical care. All these possibilities may contribute to the cross-cultural differences in prevalence and presentation of the major mental disorders. The metatheoretical point is that anthropological analysis encourages the

examination and use of multiple frameworks of analysis even if they offer competitive interpretations.

SCHIZOPHRENIA Several large-scale studies have investigated the cross-cultural prevalence of schizophrenia, but interpretation of cross-cultural differences is similarly problematic. Even after the exclusion of cases that represented heterogeneous presentations of schizophrenia, researchers have reported substantial cross-cultural differences in rates of the subtypes of schizophrenia. For example, in the developing world rapid-onset of schizophrenia was more common than insidious onset, whereas in industrialized societies onset with the paranoid type of schizophrenia was more common. Catatonic schizophrenia was also more common and the hebephrenic type less common in developing countries as compared to industrialized societies. Data also reveal major differences in the clinical features and course of schizophrenia across different cultural contexts; specifically, schizophrenia is more likely to present with an insidious mode of onset and follow a worse course in Western countries.

Social Course of Illness The priority placed on identified neurophysiological and neuroanatomical correlates of psychiatric illness obscures the substantial impact of sociocultural context on course and outcome of illness. There is strong evidence that illness follows a social course that interacts with the underlying biology of the disorder to produce a range of severity and outcome. This is perhaps best clarified in data from the International Pilot Study of Schizophrenia and the Determinants of Outcome Study that indicate a more favorable prognosis for schizophrenia in nonindustrialized societies than in industrialized, Westernized societies, which is evidence that the course of schizophrenia unfolds in culturally particular ways.

Psychiatrists and anthropologists have identified several aspects of the sociocultural environment that are likely to improve the course of schizophrenia in developing societies, including the lower likelihood of social isolation and a greater degree of social support. Most developing societies have a higher degree of sociocentrism than Western cultures, placing primacy on social and familial bonds that protect individuals from social isolation. In Western contexts, a familial milieu structured on the nuclear family allows fewer resources than in extended-family situations and may engender greater intrafamilial tension, thereby acting as a potential stressor to an individual with schizophrenia. Reintegrating an individual into occupational or productive roles may be relatively easier in developing countries than in industrialized countries in which unemployment may be high and the job market and milieu may be highly competitive. Moreover, symptomatic individuals may be able to cope with work-related tasks in agricultural or village settings more easily. Furthermore, certain psychiatric symptoms may be less stigmatized in traditional societies in which individuals may be seen as victims of witchcraft or medical illness than in Western societies in which individuals with schizophrenia may be highly marginalized. Finally, the very notion of a desirable outcome is likely to be highly context dependent. Outcome variables in Western, industrialized societies may include frequency of hospitalizations, ability to care for self, or some engagement in productive activities. In developing societies, however, the social environment may provide more flexibility in supporting an individual's ability to engage in meaningful social relationships and productive labor and hence may render the illness less disabling.

Another illustration of the social course that illness takes in relation to local political economy is revealed in the different course acquired immune deficiency syndrome (AIDS) takes in America and Africa. For example, the reason the course of AIDS is several times longer in America than in Africa is not because of the biology of the virus, but because of the absence of new treatments as well as the effect of deep poverty on housing, diet, immunological status, and the presence of high rates of infectious diseases in Africa.

RELEVANCE OF SOCIOCULTURAL FACTORS TO CLINICAL PSYCHIATRIC PRACTICE

Although most practicing psychiatrists are unlikely to encounter psychiatric illness in international contexts, the rapidly expanding percentage of the United States population representing ethnically diverse populations virtually guarantees that ethnic and cultural factors will impinge upon the therapeutic encounter. Whereas only 13 percent of the United States population was comprised of ethnic minority individuals at the beginning of this century, this figure increased to 25 percent by 1990. The minority population is projected to be about one third of the total population by 2010 and to approach one half by 2050. This scenario is already in place in California, the nation's largest state, which records non-Hispanic white Americans as fewer than half the state's population.

A careful sociocultural history provides essential clinical data with which to contextualize how a patient presents with an illness. Specifically, this approach identifies crucial information regarding social supports, relevant psychosocial precipitants and stressors, explanatory models of the illness, and characteristic patterns of help-seeking behavior for a particular illness. It also assists the clinician in distinguishing abnormal from normal presentations relative to the patient's cultural background and attending to culturally salient idioms of distress that may be articulated in unconventional ways and may otherwise confuse diagnosis and delay appropriate interventions.

Ethnicity and Access to Health Care Disturbing trends in diminished access to health care among ethnic minority individuals seriously compromise their ability to receive treatment for mental illness. Some of these barriers to health care are economic; for instance, one study found that a significantly higher percentage of Latino-Americans and African-Americans were without health insurance than white Americans. Because of frequently inadequate reimbursement for mental health care by third-party payers, economically disadvantaged groups are more likely to be unable to pay for care. Other significant impediments to accessing care may include lack of available or conveniently located services; one study found that significantly lower percentages of African-Americans, Latino-Americans, and Asian-Americans had seen a physician in 1987 as compared with all Americans. Also, there is the opportunity cost for poor people to take time off from work and go without wages or to risk losing a job.

Ethnicity and Patterns of Help-Seeking Patterns of help-seeking for health care services—in particular for the treatment of mental illness—differ widely among different ethnic groups. Help-seeking behaviors include willingness to seek and initiate care in professional biomedical settings, folk or so-called alternative settings, or within the popular sector (e.g., family or friends). For example, in Canada a sample of European-Canadian patients were readily self-referred or referred by family members to professional mental health care whereas Chinese psychiatric patients delayed initiation of professional care for protracted periods of time. In this study, mental illness among Chinese patients was kept within the confines of the family, often until the symptomatology was quite advanced. Native-Americans in this study also did not voluntarily seek professional biomedical services, but came into contact passively with various health and social agencies.

Help-seeking patterns for mental illness will be predicated upon identification of symptoms as constituting an illness. The distinction between *disease* and *illness* that is often made in the medical anthropology literature is helpful in illuminating this process further. *Disease* refers to the professional's understanding of the pathophysiological processes constituting signs and symptoms whereas *illness* refers to the patient's and family's recognition, labeling, and experience of these processes as abnormal or as a departure from wellness. Even if symptoms are experienced as *illness*, help-seeking decisions rest upon local explanatory models of what, if anything, constitutes appropriate therapy. For example, an individual may prefer to treat the symptoms with home remedies or folk therapies rather than seek treatment at a biomedical facility. Even when care is sought within the professional biomedical arena, the somatic rather than affective symptoms may be selectively attended to or identified as appropriate for treatment. Additionally, local beliefs that mental illness is a stigma may discourage individuals or families from seeking professional psychiatric care.

Alternatively, persons of color may be dissuaded from seeking professional health care or from effectively engaging in the therapeutic encounter because of past experiences of having been insulted, disrespected, or otherwise discriminated against within professional settings.

Ethnicity and Psychiatric Diagnosis Ethnicity clearly makes a substantial impact upon diagnosis in the clinical encounter. Specifically, indications are that psychiatric misdiagnosis is more likely among minority individuals than nonminority individuals in the United States. A number of studies indicate that African-American patients tend to be overdiagnosed with schizophrenia and underdiagnosed with mood disorders. Some data suggest that in the United States Latinos may also be at elevated risk for a misdiagnosis of schizophrenia. The potential for misdiagnosis among other ethnic minority populations is underresearched.

CLINICIAN BIAS Clinician bias or ethnic stereotyping is likely to account for some of the psychiatric misdiagnosis associated with ethnicity. This is perhaps best evidenced by a study in which 290 clinicians responded to two clinical vignettes describing patients with symptoms, consistent with the undifferentiated type of schizophrenia and dependent personality disorder. In the United States, clinician raters were more likely to (incorrectly) diagnose black males as having paranoid schizophrenia (43 percent compared with 6 percent for white males) and paranoid personality disorder (50 percent of black males as compared with 26 percent of the total sample). In studies in the United Kingdom, persons of Caribbean origin are diagnosed with schizophrenia at least twice as often as whites, whereas the rate of schizophrenia in Trinidad and Jamaica is about the same as that of whites in the United Kingdom. These findings are evidence of racism in psychiatry. Even in the absence of clinician bias, it is important to recognize that racial discrimination in psychiatric service delivery may occur on an institutional level—for example, access to services may be differentially limited because of the distribution of clinicians or facilities or because the organization of health services delivery may favor short-term emergency care rather than preventive care among some socially disadvantaged groups.

CULTURAL FACTORS AFFECTING ACCURACY OF PSYCHIATRIC DIAGNOSIS Psychiatric diagnosis among ethnic minority patients is complicated by a variety of

cultural factors having to do with presentation of illness. Often language or communication styles differ between clinician and patient, increasing the possibility of miscommunication. Similarly, idioms of distress invoked by the patient may be unfamiliar to the clinician. Finally, powerful cultural forces provide a template for the experience of illness and individuals are socialized to attend to or disregard symptoms according to culturally salient concepts of illness. All these factors may also affect assessment of response to treatment and outcome.

For these reasons, ethnic minority patients may present “atypically” in the clinical setting or may selectively emphasize or minimize symptoms in the course of a psychiatric evaluation. In any clinical setting, illness complaints form part of a complex agenda in the negotiation of care from a practitioner. If a patient's explanatory model of the appropriate treatment of the illness diverges from the clinician's—as is more likely in a cross-ethnic or cross-cultural clinical encounter—a clinician may formulate a diagnosis and treatment plan based upon misguided assumptions about the case. A clinician may overlook important symptoms or may overpathologize behavior that is actually not considered deviant within its particular cultural context. This in turn may contribute to situations in which patients may be undertreated or overtreated for their illness, with obviously undesirable consequences. Finally, the evidence of misdiagnosis across ethnic and cultural groups again calls into question the validity of Western psychiatric diagnostic criteria among these groups; even if the illness experience is adequately communicated to the clinician, it may not neatly correspond to DSM nosological categories.

Incorporating a Cultural Formulation of Illness Into Psychiatric Assessment and Treatment Culturally sensitive and competent psychiatric assessment and treatment requires consideration and exploration of the social and cultural world of the patient. Diagnostic accuracy is at stake as is the formulation of the most appropriate treatment plan and the successful engagement of the patient in the therapeutic relationship.

Explanatory Models The concept of *explanatory models* was developed to represent the set of experiences, ideas, and expectations of an illness. Illness experience is nearly always organized around a culturally informed model of what is at stake for the patient's or the family's life and how the illness will evolve within the framework of what matters locally on personal and collective levels. Explanatory models of illness allow individuals to make sense of illness and to integrate it into their lives meaningfully. Such models are often culturally particular—for example, as in a culture-specific illness that is attributed to a spirit visitation or an episode of *nervios* attributed to stress caused by difficult life circumstances—but they also share central themes. These themes include ideas about the nature of the problem (e.g., medical, moral, or religious) and how it affects the body and sense of self (e.g., how serious it is, which parts of the body are involved). Other key questions revolve around expectations and fears regarding course and outcome—that is, what can be expected; will there be associated pain, suffering, or disability; will there be a recovery?

Explanatory models also encompass ideas about appropriate treatment. For example, treatment for an illness may begin within the family setting and only when the patient fails to improve will additional care—folk or biomedical—be sought. Individuals seeking care within a biomedical setting often have specific requests or aversions relating to treatment (e.g., the patient who arrives asking for a selective serotonin reuptake inhibitor [SSRI] but who does not want to talk about problems, the patient who does not want any “chemical” treatment despite a severe depression, the patient who prefers a “natural” remedy such as St. John's wort for depression rather than a drug or the patient who has been brought in reluctantly and who prefers to handle the problem on his or her own). Patients' ideas about appropriate therapies often stem from explanatory models or ethnotheories about illness etiology.

In addition, physical ailments, emotional problems, or unusual behaviors are often not medicalized; that is, they are frequently perceived as existential or moral problems that require intervention in religious or interpersonal realms rather than in clinical settings. In some societies illness is not perceived as a personal crisis but rather as a social indicator of disharmony. For this reason, in some traditional settings illnesses may be preferentially addressed in ritual settings if the problem is understood to emanate from a social conflict or from moral transgressions. This occurs not only in developing societies but also in industrially and technologically advanced societies. The major thematic content of explanatory models is summarized in Table 4.1-2.

What is this problem?
How does it affect the body-self?
What can be expected to happen next?
What will be the long-term outcome?
What is most to be feared about this condition?
What treatment is most appropriate?
What is to be most feared about the treatment?

Table 4.1-2 Illness Concerns and Meanings Addressed by an Explanatory Model

Explanatory Models for Illness Explanatory models of illness causation and appropriate therapy are frequently quite eclectic, incorporating both folk and biomedical therapeutic strategies. These models also often evolve as individuals consult others, build a consensus, or reject ineffective models.

A Fijian mother located in rural Fiji recounted her baby's first seizure and demonstrated her shifting model of etiology of this illness. She explained that her baby had had fever and vomiting for several days when convulsions started in the mid-afternoon. She sought advice from her grandmother, who suggested a traditional herbal remedy, which she prepared for the infant. Despite this therapy, the convulsions continued. At this juncture, her family recalled a recent altercation with a fellow villager who was known for involving evil spirits for revenge and decided that he may be responsible for the child's illness. “Once a while ago, my father and one old man in the village argued with each other,” she explained. Their dispute was on a Monday, and the child fell ill with seizures on that same Friday: “right after the fight, then came the illness.” Her family reached a consensus that they urgently needed to apologize formally for the dispute by making the ceremonial offering of kava root. She recalled that the men had been explicit in relating the argument to the seizure illness: “If [the child's] illness... is due to our argument over the marriage you must drink [the kava] so that [his] sickness will end.” She further recalled that as soon as the kava was accepted and drunk, the child's body relaxed. The child was subsequently taken a considerable distance to the nearest hospital, where he was given an injection. The convulsions finally remitted at about 9 in the evening. Although accepting that her child's epileptic seizures will recur chronically, this mother reports reflecting on her family's relational standing with other villagers with each new seizure episode. In this example, the illness was initially addressed with a popular remedy. When this proved ineffective, however, a new model of causation was formulated by consensus: that the child's seizures were a manifestation of a social conflict that required ceremonial measures. Although this model was retained, the family also sought professional treatment at a Western biomedical facility, where treatment was definitive in eradicating the symptoms. The enduring explanatory model of the illness continued to formulate it simultaneously both as a chronic seizure disorder and as a consequence of social conflict. (Adapted from Beeker AE: *Body, Self, and Society: The View for Fiji*. University of Pennsylvania Press, Philadelphia, 1995).

Patients' explanatory models tend not to be formalized, rigid, or fixed. Rather, they shift to adapt to competing or alternative models or to incorporate new information, experience, or contexts. These models are often negotiated among patients, families, and their communities as decisions about help-seeking and following treatment recommendations are made. Tacit negotiation often accompanies the clinical encounter as well. By contrast, professional explanatory models of illness are more standardized and formal, drawing from a formal body of medical knowledge and theory. However, clinicians also draw their understanding and expectations about an illness from clinical anecdotal experience and personal and cultural backgrounds.

Particularly when a clinician and a patient come from different sociocultural backgrounds, their explanatory models may diverge significantly. Such disparities in explanatory models can seriously complicate the clinical encounter, especially when they are not made explicit. The patient may leave the office with a sense of frustration at not being understood, thus compromising the therapeutic alliance; worse yet, treatment efficacy may be compromised if a patient refuses to adhere to a plan that does not fit with his or her explanatory model or if a patient entirely fails to engage in treatment. A clinician's countertransference in such a situation may include frustration with a patient who may appear to be noncompliant or condescension for unfamiliar folk ideas that do not mesh with professional medical or psychiatric knowledge.

It is common for clinicians to dismiss patients' explanatory models as folk “beliefs” that are inferior to professional clinical knowledge. Such beliefs are frequently

ascribed to mistaken or irrational ideas that can be “corrected” with appropriate education that is based on scientific knowledge. Clinicians should by all means educate patients about professional models of their illness but are strongly cautioned against presuming that patients will easily accept professional explanatory models. It is essential to appreciate that patients' explanatory models reflect a culturally particular form of rationality, one that is quite relevant to their experience of illness. Thus, when viewed anthropologically the role of the physician is to affirm the patient's experience of illness and to listen with respect to the patient and family models. The clinician may not and need not agree with those meanings, but his or her task is to negotiate with them and not to reject or discredit them.

Elicitation of a patient's explanatory models can make discrepancies and common elements in a patient's and clinician's models explicit and allow for an effective negotiation of the meaning of clinical illness and a treatment plan. Exploring a patient's explanatory model is, in effect, a strategic means of anticipating pitfalls in communication and adherence to treatment. In many cases, it is appropriate for the clinician to disclose the clinical explanatory model to allow patients to ask questions that will clarify the treatment plan or, alternatively, to facilitate negotiation of the treatment process.

Illness Meanings In addition to its critical use in psychiatric assessment, exploration of a patient's explanatory model allows a clinician to enter meaningfully into the world of the patient, develop a more empathic connection to the patient's illness experience, and identify and negotiate the most appropriate treatment plan. However, an explanatory model is only a starting point for delving further into a patient's sociocultural background. Beyond the abridged approach of eliciting an explanatory model, more extensive interviewing will be required to gain an appreciation of the social, economic, and political contexts of stressors and the cultural patterning of distress for a particular patient.

Exploring and responding to the many facets of a patient's illness experience is integral to psychotherapeutic work. Illness experience often revolves around what is at stake for a patient in personal, social (e.g., occupational, economic, and political), and cultural contexts. Besides the conventional meanings of particular symptoms and patients' explanatory models for their significance, it is helpful for clinicians to appreciate that illness may take on a variety of meanings particular to personal and cultural contexts.

Chronic mental illness with its exacerbations and remissions may especially take on personal significance for patients as they incorporate it into the trajectory of their lives and reconcile the chronicity of their illness with perceived and actual limitations and suffering. Illnesses are also invested with cultural meanings that substantially impact upon illness experience. For example, some psychiatric disorders are highly stigmatizing (e.g., psychotic disorders or pedophilia in American society) in some cultural contexts, whereas others designate the patient as a victim (e.g., posttraumatic stress disorder). The medicalization or relocation of some illnesses from one social arena to a clinical one (e.g., alcohol use disorders) may diminish their social stigma and allow mobilization of social resources in more effective ways for the patient. On the other hand, the medicalization of posttraumatic response to the violence of political oppression localizes pathology in the individual rather than in the social context that produced the symptoms.

Cultural Formulation in Psychiatric Assessment Notwithstanding the contributions of neurophysiology and psychological development to psychopathology, mental illness always unfolds in a particular social and cultural context that may both precipitate and shape it. Treatment that is decontextualized from these sociocultural influences runs the risk of being insensitive at best and ineffective or harmful at worst. In an effort to promote systematized incorporation of the sociocultural context into psychiatric assessment, DSM-IV includes an outline of the cultural formulation of a patient's illness that is meant to supplement the standard multiaxial system of evaluation. This approach is outlined in Table 4.1-3 and is a starting point for rendering social and cultural factors visible in the clinical encounter.

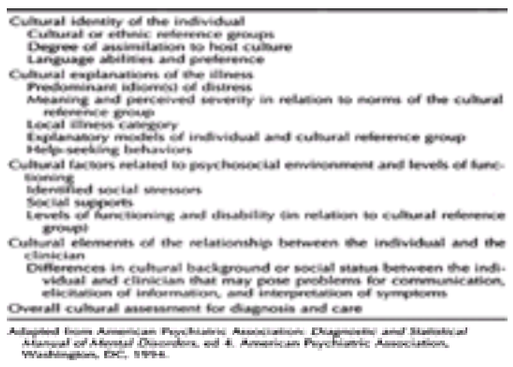

Cultural identity of the individual
- Cultural or ethnic reference groups
- Degree of assimilation to host culture
- Language abilities and preference

Cultural explanations of the illness
- Predominant idiom(s) of distress
- Meaning and perceived severity in relation to norms of the cultural reference group
- Local illness category
- Explanatory models of individual and cultural reference group
- Help-seeking behaviors

Cultural factors related to psychosocial environment and levels of functioning
- Identified social stressors
- Social supports
- Levels of functioning and disability (in relation to cultural reference group)

Cultural elements of the relationship between the individual and the clinician
- Differences in cultural background or social status between the individual and clinician that may pose problems for communication, elicitation of information, and interpretation of symptoms

Overall cultural assessment for diagnosis and care

Adapted from American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, ed 4, American Psychiatric Association, Washington, DC, 1994.

Table 4.1-3 Outline for Cultural Formulation

Although listed as the fourth component in the DSM-IV cultural formulation, the clinical assessment begins with attention to any cultural and social differences between the clinician and the patient that may introduce barriers to communication. An interpreter's services should be used in cases in which the identified patient cannot be interviewed in English. Structuring the interview setting such that the patient is comfortable may include allowing family members to remain in the room and sometimes even to speak for the identified patient when culturally appropriate. Clinicians should be aware that minority individuals are frequently demeaned or discriminated against, including in clinical encounters. They should make every effort to convey respect for patients' ethnic, cultural, and social backgrounds, particularly when it is obvious that there is a socioeconomic gap or racial difference between patient and clinician. Clinicians are often unaware of ways in which the professional culture has made them unwitting participants in a power differential between themselves and their patients; an awareness of such conventions (e.g., addressing patients by first names instead of surnames, controlling time and space in the clinical setting, limiting telephone access) can allow clinicians to identify and modify behaviors that potentially intimidate patients.

The cultural identity of the individual can be explored by asking about cultural or ethnic reference groups and about the degree of *assimilation* to the local culture. Because there are often vast intergenerational differences in assimilation that can contribute to contested ideas of illness and the need for treatment, it is helpful to ask about the degree of assimilation of family members and any tensions resulting from intrafamilial differences. In addition to assimilating to another culture, the process of *acculturation* includes the discarding of certain cultural practices, values, and traditions in the setting of another (usually more dominant) culture as elements of that culture are incorporated into attitudes and behavior. In some cases highly acculturated individuals may become more vulnerable to conventional stressors because traditional coping mechanisms are no longer available to them.

A cultural context of presenting symptoms is especially important when a patient presents with symptoms that are atypical from the perspective of the host culture. A familiarity with culture-bound syndromes and culturally specific idioms of distress is helpful in evaluating their clinical and cultural salience. Certain culture-specific presentations (e.g., *ataques de nervios* or *zar*) resemble DSM-IV diagnoses, but are not necessarily deemed pathological within their cultural reference group. Explanatory models for the illness as well as culturally salient illness meanings assist the clinician in engaging the patient and negotiating a mutually acceptable treatment plan. Also, an appreciation of the conventional coping styles and help-seeking behaviors of the reference group illuminates the significance of the patient seeking help in a biomedical setting and allows a discussion of alternative therapies a patient may be pursuing concomitant with biomedical therapy. Finally, clinicians should avoid viewing culture, ethnicity, or race as essential characteristics of the individual. Although familiarity with culturally specific idioms of distress and explanatory models of illness is indispensable in contextualizing illness, the clinician should be especially attuned to heterogeneity within the various ethnic and cultural groups and be prepared to engage individual as well as cultural diversity in the clinical encounter. Indeed, individual and cultural diversity need to be viewed as only two kinds of diversity amongst many others: gender, age-cohort, religious, class, and other sources of differentiation contribute to diversity as well.

Assessment of patients' psychosocial environment helps to identify social stressors and supports as well as to formulate what resources may be available. Levels of functioning and disability should be determined with respect to the expectations of the cultural reference group as well as to personal and clinical expectations. Therapeutic goals may differ considerably between clinician and patient and should be explored as treatment unfolds. Clinicians should be cautioned to not rigidly impose their own sociocultural values on a patient's therapeutic goals, but rather should recognize that they may not be appropriate to the patient's cultural background.

Finally, the cultural formulation of the patient integrates elements of the patient's ethnic and cultural identity, social environmental factors contributing to or buffering against distress, and the cultural construct of the illness experience into an assessment of the patient that at once allows application of Western psychiatric nosology and use of biomedical or psychiatric treatment resources in a relevant and appropriate way. Cultural elements of the relationship between the clinician and patient are

also factored into the assessment and the ongoing therapeutic relationship with the patient.

Cultural Considerations in Psychiatric Treatment After assessing and formulating a patient's illness complaint relative to personal, social, and cultural contexts, a culturally informed exploration of the patient's expectations regarding treatment will identify appropriate treatment modalities. If indicated, psychotherapeutic work will ideally integrate personal and cultural agendas for the patient. Whether applying psychodynamic techniques that focus on intrapsychic experience can be successful across cultural groups or whether use of a second or third language in exploring affective experience might affect outcome are underresearched questions. Western psychotherapeutic models assume a rather ethnocentric orientation toward an individuated self that reflects Western notions of selfhood as well as an agenda often derived from white, middle-class, male values. For patients from cultural backgrounds that value social or family harmony over the needs of the individual or that support parental authority over a child's desire to develop an independent agenda, the goals of care must be negotiated between clinician and patient to ensure that they are culturally relevant to the patient.

PSYCHOPHARMACOLOGY Medical management of mental illness takes place in the context of ethnicity-specific biologies. There is evidence that the pharmacokinetics of several psychotropic medications vary among differing ethnic groups. In addition, the volume of distribution, differential brain receptor responsivity, and exposure to environmental factors (such as herbal remedies) that alter drug metabolism may also contribute to interethnic differences in response to psychotropic medications. Specific differences that have been reported in the literature include comparably higher haloperidol plasma concentrations in Asians than in whites with the same dosage of drug and optimal response to lower plasma levels in Asians as compared with whites. Similarly, Chinese patients appear to respond to significantly lower mean lithium concentrations and to be more sensitive to benzodiazepines than whites. Practitioners who are routinely exposed to patients of different ethnic backgrounds are responsible for learning these aspects of the clinical picture.

SOCIAL CONTEXT OF MENTAL ILLNESS

Integral to the cross-cultural and cross-ethnic epidemiology of mental illness are the social, economic, and political factors that create a context for either social supports or stressors that affect risk. Studies have consistently demonstrated an inverse relationship between socioeconomic status and health and mental illness. Although there is some evidence that mental illness sometimes causes downward drift in socioeconomic status (e.g., with schizophrenia), converging lines of evidence implicate social factors in the production of mental illness mediated through social stressors. For example, high rates of suicide and substance abuse have been demonstrated among members of societies such as Native-Americans and Pacific Islanders in which rapid modernization has disrupted social structures and social supports. Powerlessness appears to also pose risk for mental illness along gender, socioeconomic, and ethnic lines. Studies have shown that, globally, women have higher rates of mental illness than men; furthermore, this research shows that women with fewer social supports and greater social stressors are at the greatest risk for depression. Moreover, victims of political violence and oppression are also at increased risk for mental illness, as evidenced by high rates of depression, anxiety, and posttraumatic stress disorder in refugee populations. Unfortunately, economic change in the current era of global capitalism is associated with increasing disparities between the highest and lowest 20 percent of populations and also increased mental health problems.

Suicide The rate of suicide varies dramatically across cultures. Both its varying prevalence and association with varying risk factors suggest that it may be a better marker of social processes than of psychiatric illness. Suicide rates are markedly higher among some age groups in the South Pacific and among women in China than elsewhere in the world. For example, within Micronesia, Trukese 15 to 24-year-old males had an annual rate of suicide of 200 per 100,000 people between the years 1974 and 1983—a rate nearly 10 times that of United States males of the same age. Notably, suicide rates in the Trukese setting underwent an eightfold increase between 1960 and 1980, a time of substantial social change locally. Moreover, precipitants to Trukese adolescent suicide attempts appeared to be related to relatively trivial parent-child conflict rather than to depression or other mental illness; such conflict, however, would have been traditionally resolved with club-houses and men's organizations no longer available in the wake of postwar economic change and restructuring of the family. In other words, social and economic changes that had redefined parenting roles and eliminated particular social supports, in combination with limited legitimate means of expressing anger toward authority figures, created a context for epidemic suicide, apparently unrelated to *psychiatric* illness.

Similarly, the rate of suicide in China, particularly among rural women, has also increased. Forty percent of the world's suicides occur in China, a nation with 21 percent of the world's population. This high rate of suicide cannot be attributed to depression because China has far lower prevalence of depression than many other areas. For example, the point prevalence of depression in China is 3 to 10 times lower than in the United States, yet China's suicide rate is twice that of the United States. Again, suicidal behavior cannot necessarily be linked with mental illness. Rather, these examples suggest that high suicide rates may accompany rapid social and economic change resulting in deep demoralization or eradication of traditional social supports in the context of limited options for resolving distress. In China, for example, rural women aged 16 to 26 years and over the age of 55 have the highest suicide rate. Although a number of social factors contribute to this, it seems as if relative powerlessness in the setting of rising economic expectation and diminishing opportunities for women are instrumental here.

Global Burden of Psychiatric Illness: Psychiatry's International Health Agenda Despite biotechnological advances that have substantially improved the overall health status in developing societies worldwide, studies have shown that there has been a paradoxical increase in the rates of mental illness in low-income societies. This increased prevalence is probably due to a variety of social factors, including rapid social change and economic restructuring that disrupt the integrity of traditional structures of authority and livelihood, thereby increasing the risk of displacement, violence, and poverty. In addition, social structures such as the extended family, which traditionally provide care-taking and other supports for the mentally ill and elderly, are undergoing rapid change in many contexts and thereby decreasing social supports that might either protect against mental illness or buffer its impact on the individual or community. These explanations are supported by the clustering of social and mental health problems. From an anthropological perspective, cases of major depressive disorder or posttraumatic stress disorder rarely occur in isolation; rather they cluster in settings of family breakdown, joblessness, community and domestic violence, and the like. Thus, mental health problems and social problems are largely inseparable.

Care and Treatment for Chronically Mentally Ill Persons in Nepal Nepal is a small, landlocked country of some 23 million people. Given that it is one of the poorest countries in the world—with a gross national product (GNP) per capita of less than U.S. $200—it is not surprising that by most measures, the health and welfare of the Nepalese are far less than optimal. Only 27.5 percent of the population is literate, with a great disparity between literacy rates for men (40.9 percent) and those of women (14 percent). With a growth rate of 2.6 percent per annum, Nepal's population is expected to double in the next 20 years, adding to the problems of an explosion that began at least two decades ago. Life expectancies are low and are at least in part accounted for by high rates of infant mortality (99 per 1000 live births), under-5 mortality (128 per 1000 live births), malnutrition (51 percent of children under 5 are underweight), and maternal mortality (830 per 100,000 live births). Although there has not been extensive psychiatric epidemiological research conducted in Nepal, the research that has been conducted indicates a prevalence rate in the general population of about 2 percent for chronic mental disorders (i.e., roughly the same as has been found elsewhere). Care and treatment services for chronically mentally ill persons in Nepal are virtually nonexistent. Of about 2500 doctors, less than 20 are trained psychiatrists and most of them are in the capital city of Kathmandu: Fewer than 10 nurses have received formal training in psychiatry; clinical psychologists are fewer still. Until recently Nepal had no training programs in psychiatry, psychology, psychiatric nursing, or social work. Individuals who wanted to be trained in these fields had to study abroad (most often in India or Pakistan). Even with the establishment of programs in the Institute of Medicine (IOM) at Tribhuvan University, Nepal's supply of mental health professionals is going to remain inadequate for the foreseeable future. The availability of psychiatric facilities is no better. The National Mental Hospital in Patan has 40 beds and the Psychiatry Department of IOM has 12 beds in the University Hospital. Beyond the Kathmandu Valley the situation is worse. The Western Regional Hospital in Pokhara has 5 psychiatric beds, and the Birendra Army Hospital has 18 psychiatric beds. The director of the National Mental Hospital is planning to open a series of day centers for chronically mentally ill persons in Kathmandu, but in the meantime there is only one rehabilitation facility in the entire country: Asha Deep in Kathmandu, originally established in 1993 by the Sisters of Maryknoll to care for mentally ill persons who had been in jail. Because of the lack of psychiatric facilities and services, many mentally disturbed persons, especially those who are violent, are taken to the police and jailed. There are between 60 and 80 mentally disturbed persons at any given time in the Dhulikhel Jail in Kathmandu where they receive a modicum of treatment. There is no formal information about private psychiatric facilities, but if the anecdotal information is accurate they are no better and some are much worse than the public facilities. Most Westerners would probably be disturbed at the conditions for psychiatric patients in Nepal's hospitals. Running water is still not available in the Mental Hospital. The wards tend to be small and bare; patients mainly stay in bed with nothing more to do than talk to relatives, friends, or one another. Psychopharmacology and electroconvulsive therapy are the primary treatments. However, given the economic constraints under which the health care system functions and the lack of trained staff, conditions in the public hospitals must be considered good. Patients are not abused, the facilities are clean if somewhat dreary, and while better and more comprehensive care can be found in many other countries, what is offered in Nepal is based on modern psychiatric practices. What of care and treatment services for chronically mentally ill individuals in the rural areas? Those who have access to transportation are sent to the hospitals in Kathmandu or Pokhara even though it might take 3 days of rough traveling on a crowded bus. Transportation other than by foot is scarce. Since 1989 His Majesty's Government of Nepal, in collaboration with United Mission to Nepal, has been training rural health care workers in the diagnosis and management of mental disorders. However, this project covers only a small part of the country, and even where it is available, health care workers can only offer a very limited range of psychopharmacological agents as treatment for chronically mentally ill persons. Recognizing the need for mental health services in primary care settings, the government has made a long-term commitment to the mental health project, and for the past decade the project has provided at least some level of treatment for chronically mentally ill persons in rural areas, extended critical knowledge to health care workers in the remote health posts, and helped to spread a concern and better understanding of mental health problems. In addition, if the rehabilitation facilities at Asha Deep are

any indication, there is a much brighter future for those in Nepal who suffer from chronic mental disease. Finally, the lack of formal psychiatric care does not mean that persons with chronic mental illness are abandoned; the great majority of them are cared for by their families. (Courtesy of Alex Cohen, Ph.D.)

In *World Mental Health*, Robert Desjarlais and his colleagues examine the relationships among poverty, urbanization, violence, diasters, displacement, and mental illness in developing societies. They point out that poor countries in Asia, Africa, Latin America, and the Pacific have a disproportionate burden of social problems (e.g., war and natural disasters) that pose risk for mental illness and that they lack adequate mental health services. Furthermore, they illustrate the interpenetration of social and mental health problems (e.g., domestic violence, substance abuse, community disintegration) that perpetuate one another in a vicious spiral in the context of key social forces, such as ethnic conflict, economic inequities, and oppression of women.

Research has shown that psychiatric illness is associated with disability across cultural contexts. Data presented in the *Global Burden of Disease* convincingly demonstrate the impact of global neuropsychiatric illness. For example, five of the ten leading causes of Disability-Adjusted Life Years (DALYs) for 15- to 44-year-olds in the world represent mental illness or self-inflicted injuries: major depressive disorder, alcohol use, self-inflicted injuries, bipolar disorders, and schizophrenia. Furthermore, the leading cause of disability-adjusted life years (DALYs) is major depressive disorder, responsible for over 10 percent of the total (Table 4.1-4). Neuropsychiatric conditions account for 26.4 percent of the total DALYs among 15- to 44-year-olds, and comprise the single largest category of illnesses. (Fig. 4.1-1). This is probably a conservative estimate of the contributions of mental illness to the burden of disease because a proportion of unintentional injuries and cases of cirrhosis of the liver are undoubtedly secondary to substance abuse. Moreover, intentional self-inflicted injuries account for 3.51 of the total DALYs among this group. These percentages reflect the high prevalence of mental illnesses as well as the inadequacy of health resources to address them, often resulting in unnecessarily prolonged disability. The high rates of mental illness and associated disability in the developing world challenge the psychiatric professional community to respond by developing effective treatments that can be applied across cultural contexts as well as by supporting social and public health policies that address the roots of the problem at the population level.

Ranking	Causes	Total DALYs (millions)	% of Total DALYs
	All causes	*419.1*	*100*
1	Unipolar major depression	43.0	10.3
2	Tuberculosis	19.7	4.7
3	Road traffic accidents	19.6	4.7
4	Alcohol use	14.8	3.5
5	Self-inflicted injuries	14.6	3.5
6	Bipolar disorder	13.2	3.1
7	War	13.1	3.1
8	Violence	13.0	3.1
9	Schizophrenia	12.5	3.0
10	Iron-deficiency anemia	12.5	3.0

Adapted from Murray CJL, Lopez AD, editors: *Global Burden of Disease*. Harvard School of Public Health on Behalf of the World Health Organization, Cambridge, 1996.

Table 4.1-4 Ten Leading Causes of DALYs at Ages 15–44 Years, the World, 1990

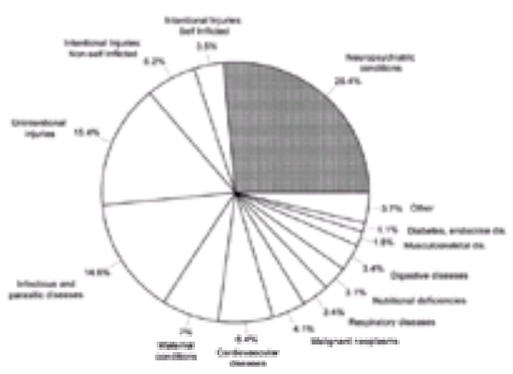

FIGURE 4.1-1 DALYs among 15–44 year olds by cause for the world, 1990. (Adapted from data in Murray CJL, Lopez AD, editors: *Global Burden of Disease*. Harvard School of Public Health on Behalf of the World Health Organization, Cambridge, 1996.)

SUGGESTED CROSS-REFERENCES

Section 4.2 covers sociology and psychiatry. An expanded review of socioeconomic aspects of health care is contained in Section 51.5a and Section 52.20. Also relevant to sociocultural issues in psychiatry are discussions of sociology and psychiatry (in Section 4.2) and public psychiatry (in Section 52.1).

SECTION REFERENCES

Andreason NC: Linking mind and brain in the study of mental illnesses: A project for scientific psychopathology. Science *275:*1586, 1997.

Becker AE: Body, Self, and Society: The View from Fiji (New Cultural Studies). University of Pennsylvania Press, Philadelphia, 1995.

Brown G, Harris T: *The Social Origins of Depression*. Free Press, New York, 1978.

Canino G, Lewis-Fernandez R, Bravo M: Methodological challenges in cross-cultural mental health research. Transcult Psychiatry *34:*163, 1997.

Chakraborty A: Culture, colonialism, and psychiatry. Lancet *337:*1204, 1991.

Cornelius LJ: Ethnic minorities and access to medical care: Where do they stand? Assoc Acad Minor Phys *4:*16, 1993.

*Desjarlais R, Eisenberg L, Good B, Kleinman A: *World Mental Health: Problems and Priorities in Low-Income Countries*. Oxford: Oxford University Press, 1995.

Dohrenwend BP, Dohrenwend BS: Social and cultural influences on psychopathology. Annual Review of Psychology *25:*417, 1974.

Eisenberg L: Psychiatry and health in low-income populations. Comprehensive Psychiatry *38:*69, 1997.

Garro L: On the rationality of decision-making studies. Part 1: Decision models of treatment choice. Med Anthropol Q *12:*319, 1998.

Garro L: On the rationality of decision-making studies. Part 2: Divergent realities. Med Anthropol Q *12:*341, 1998.

Good BJ, Good M-J D: Toward a meaning-centered analysis of popular illness categories: “fright illness” and “heart distress” in Iran. In: *Cultural Conceptions of Mental Health and Therapy*, AJ Marsella, GM White, editors. D. Reidel Publishers, Netherlands, 1982.

Good BJ: Culture, diagnosis and comorbidity. Cult Med Psychiatry *16:*427, 1993.

*Good BJ: *Medicine, Rationality, and Experience: An Anthropologic Perspective*. Cambridge University Press, Cambridge, 1994.

Guarnaccia PJ, Good BJ, Kleinman A: A critical review of epidemiological studies of Puerto Rican mental health. Am J Psychiatry *147:*11, 1980.

Hezel FX, Rubenstein DH, White GH, *Culture, Youth and Suicide in the Pacific: Papers from an East-West Conference*. University of Hawaii, Honolulu, 1984.

Hughes CC: The glossary of 'culture-bound syndromes' in DSM-IV: a critique. Transcultural Psychiatry *35:*413, 1998.

Kendler KS, Health AC, Martin NG, Eaves LJ: Symptoms of anxiety and symptoms of depression. Arch Gen Psychiatry *44:*451, 1987.

Kirmayer LJ, Young A, Hayton B: The cultural context of anxiety disorders. Cult Med Psychiatry *18:*503, 1995.

Kirmayer LJ, Young A: Culture and somatization: Clinical, epidemiological and ethnographic perspectives. Psychosom Med *60:*420, 1998.

Kleinman A: *Writing at the Margin*. University of California Press, Berkeley, 1995.

*Kleinman A: *Rethinking Psychiatry: From Cultural Category to Personal Experience*. The Free Press, New York, 1988.

Kleinman A: *The Illness Narratives: Suffering, Healing and the Human Condition*. New Basic Books, New York, 1988.

Kleinman A: *Patients and Healers in the Context of Culture*. University of California Press, Berkeley, 1980.

*Kleinman A: *Social Origins of Distress and Disease*. Yale University Press, New Haven, CT, 1986.

Kleinman A, Cohen A: Psychiatry's global challenge. Sci Am *276:*86, 1997.

Kleinman A, Good B, editors: *Culture and Depression*. University of California Press, Berkeley, 1985.

*Kleinman A, Becker AE, Dimsdale J: "Sociosomatics:" The contributions of anthropology to psychosomatic medicine. Psychosom Med *60:*389, 1998.

Lee S: Cultures in psychiatric nosology: The CCMD-2-R and international classification of mental disorders. Cult Med Psychiatry *20:*421, 1996.

Lewis-Fernandez R, Kleinman A: Cultural psychiatry. Psychiatr Clin North Am *18:*433, 1995.

Lewis-Fernandez R, Kleinman A: Culture, personality, and psychopathology. J Abnorm Psychol *103:*67, 1994.

Lin KM, Anderson D, Poland RE: Ethnicity and psychopharmacology. Psychiatr Clin North Am *18:*636, 1995.

Lin KM, Kleinman A: Recent development of psychiatric epidemiology in China. Cult Med Psychiatry *5:*135, 1981.

Lin KM, Kleinman A: Psychopathology and clinical course of schizophrenia: A cross-cultural perspective. Schizophr Bull *14:*555, 1988.

Lin T-Y, Tardiff K, Donetz G, Goresky W: Ethnicity and patterns of help-seeking. Cult Med Psychiatry *2:*3, 1978.

Lin T-Y, Rin H, Yeh EK, Hsu CC, Chu HM: Mental disorders in Taiwan, fifteen years later: A preliminary report. In *Mental Health Research in Asia and the Pacific*, W Caudell, T-Y Lin, editors. East West Center Press, Honolulu, 1969.

Link BG, Phelan J: Social conditions as fundamental causes of disease. J Health Soc Behav *36*(Suppl):80, 1995.

Loring M, Powell B: Gender, race, and DSM-III: A study of the objectivity of psychiatric diagnostic behavior. J Health Soc Behav *29:*1, 1988.

Manson SM: Culture and major depression: Current challenges in the diagnosis of mood disorders. Cult Med Psychiatry *18:*487, 1995.

Murray CJL, Lopez AD, editors: *Global Burden of Disease*. Cambridge, Harvard School of Public Health on Behalf of the World Health Organization, 1996.

Nichter M: Idioms of distress. Cult Med Psychiatry *5:*5, 1981.

O'Hare WP: America's minorities—the demographics of diversity. Pop Bull 47, 1992.

O'Nell TD: "Feeling worthless": An ethnographic investigation of depression and problem drinking at the Flathead Reservation. Cult Med Psychiatry *16:*447, 1993.

Ormel J, VonKorff M, Ustun TB, Pini S, Korten A, Oldehikei T: Common mental disorders and disability across cultures: Results from the WHO collaborative study on psychological problems in general health care. JAMA *272:*1741, 1994.

*Pierce CM, Earls FJ, Kleinman A: Race and culture in psychiatry. *The New Harvard Guide to Psychiatry*, Nicholi AM, ed 3. Harvard University Press, 1999.

Rubenstein DH: Epidemic suicide among Micronesia adolescents. Soc Sci Med *17:*657, 1983.

Sartorius N, Jablensky A, Korten A, Ernberg G, Anker M, Cooper JE, Day R: Early manifestations and first-contact incidence of schizophrenia in different cultures. A preliminary report on the initial evaluation phase of the WHO Collaborative Study on determinants of outcome of severe mental disorders. Psychol Med *16:*909, 1986.

Tousignant M: Suicide in small-scale societies. Transcult Psychiatry *35:*291, 1998.

Ware NC, Kleinman A: Culture and somatic experience: The social course of illness in neurasthenia and chronic fatigue syndrome. Psychosom Med *54:*546, 1992.

Weiss M: Explanatory Model Interview Catalogue (EMIC): Framework for comparative study of illness. Transcult Psychiatry *34:*235, 1997.

Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H-G, Joyce PR, Karam EG, Lee C-K, Lellouch J, Lepine J-P, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H-U, Yeh E-K: Cross-national epidemiology of major depression and bipolar disorder. JAMA *276:*293, 1996.

World Health Organization: The international pilot study of schizophrenia. Geneva, World Health Organization, 1973.

World Health Organization: Schizophrenia: An international follow-up study. Wiley, New York, 1979.

4.2 SOCIOLOGY AND PSYCHIATRY

RONALD C. KESSLER, PH.D.

Sociology is the study of patterns and structural determinants of social organization. Sociologists presume that there are organizing principles of social life, that these principles can be uncovered by systematic observation, and that the social forces underlying these principles influence both organizational and individual actions. Contemporary sociological studies are interested in psychiatric phenomena in a number of ways. They have focused on the structural determinants of psychiatric disorders, the social consequences of psychiatric disorders, the social factors in psychiatric help-seeking, attitudes toward the mentally ill, and mental health service organization.

SOCIAL AND CULTURAL DETERMINANTS OF MENTAL ILLNESS

Sociological research on social and cultural determinants of psychopathology is varied: one line of research investigates the effects of stressful life experiences on the onset and course of psychiatric disorders; a related line of research studies the extent to which reactivity to environmental stress is mediated or modified by social and cultural forces.

Stress and Mental Health: Effects of Life Events Although the hypothesis that stress can cause mental disorder is an old one, it has been difficult to document causal effects of this sort. Most such work has focused on the putative effects of life events. Although significant associations between presumably stressful events and mental illness have been consistently documented in these studies, the interpretation is ambiguous because this association could reflect an influence of the illness on the event. There is no certain way of discounting this possibility in the nonexperimental studies that are the mainstays of stress research.

Nonetheless, the strength and consistency of the associations documented in this literature are striking. For example, studies of job loss because of plant closings have documented rates of clinically significant anxiety and depression among unemployed workers that are two to three times higher than those found among the stably employed. Furthermore, in a few cases these studies have been able to collect pre-event data that document associations between exposure to stress and the onset of psychiatric disorders, arguing against a selection argument and in favor of the interpretation that stress is a cause of the ill health outcomes.

Elaboration of the stress-illness relationship provides information that is consistent with a causal interpretation. This can be seen in studies that attempt to sort the effects of life events into the dimensions that make them stressful. For example, job loss seems to promote anxiety and depression by increasing financial strain and heightening reactivity to unrelated stresses. As a result, the most serious psychiatric outcomes associated with job loss are found among people who lack financial reserves and who also experience some other major crisis (e.g., a life-threatening illness in a child) during the period of unemployment. Research is ongoing to delineate contextual features of this sort that account for variation in the effects of stressful events.

A recent related line of research involves the determinants of posttraumatic stress disorder following such highly stressful events as rape or combat. Although a substantial proportion of those exposed to such events develop posttraumatic stress disorder or some related anxiety or mood disorder, they typically represent only a minority of those exposed to the events. This has prompted an interest in the protective factors that allow the majority of trauma victims to avoid these disorders, an issue discussed later in this chapter.

Another recent related line of research examines the long-term effects of earlier life adversities in the context of a developmental perspective on psychopathology. Clinical studies clearly suggest that early adversities such as parental death and family violence have lifelong effects on mental health. However, a fairly new development is the systematic investigation of those effects in a representative community sample of adults who are asked retrospectively about childhood experiences. In the 1980s studies largely focused on only one type of childhood adversity, such as death of a parent, childhood family violence, or early sexual abuse, and one clinical outcome (usually major depression). These studies consistently found that the early adversities had significant effects on adult disorders. In the 1990s, however, studies have been concerned with the long-term effects of multiple childhood adversities and have shown that it is much more difficult than previously realized to pinpoint any one particular early adversity as a central risk factor for adult disorders. Instead, it appears that many early adversities cluster in the lives of particular youngsters and that these clusters, rather than the individual adversities of which the clusters are composed, are the most important determinants of psychopathology. It is likely that future work in this area will examine more closely the differential effects of various isolated early adversities and commonly occurring adversity clusters.

Research on Chronic Stress It is easier to study the effects of stressful life events than ongoing chronic stress situations because an event can be dated whereas the onset of chronic stress is more difficult to pinpoint. As a result of this difficulty, research on the effects of chronic stress is less well developed than research on the effects of life events. This does not imply that life events are more important than chronic stresses; indeed, chronic stresses are often more predictive than stressor events of psychological disorders in community surveys. Methodological research is needed before major advances can be made in this area of research.

The most advanced work on chronic stress is concerned with job stress because it is easier to measure than other types of chronic stress. Research on job stress has shown that such indicators as time pressure, closeness of supervision, and job insecurity are all associated with depression, anxiety, and substance abuse. Based on these results, more focused studies of such high-risk workers as assembly line workers and air-traffic controllers have been undertaken. These studies have described particular constellations of job conditions associated with emotional disability. For example, several large studies have linked the combination of high job demands (e.g., a job in which workers must rush to meet important deadlines) with low decision latitude (e.g., low control over the pace or organization of work) to emotional disability and cardiovascular disease. Several major corporations have now agreed to carry out job redesign experiments aimed at modifying some of these health-damaging job conditions. Those efforts, motivated partly by a desire to increase worker productivity, provide an unparalleled opportunity to study the effects of chronic stress. Such experiments should yield important knowledge about the determinants of chronic job stress and about effective strategies for changing work environments to reduce the most pernicious kinds of stress.

Research on Vulnerability Factors Although stress is strongly related to ill health, only a minority of the people who are exposed to stress develop stress-related disorders. Much research has been done on the determinants of this variation in stress reactivity and a number of determinants have been identified. Initial research on vulnerability factors was carried out as part of naturalistic studies of stress effects that focused on the ability of the putative vulnerability factors to exacerbate the impact of stress on health. For example, research showed that the impact of life events in provoking episodes of major depression is reduced among individuals who have an intimate, confiding relationship with a friend or relative. In one study, nearly 40 percent of the stressed women without a confidant became depressed compared with only 4 percent of those women with access to a confidant. This result has been replicated in several community surveys and case-control studies.

Although these studies provide suggestive evidence, methodological problems obscure results. Most seriously, in a naturalistic study the possibility that some predisposition to depression accounts for the presumed exacerbating effects of vulnerability factors cannot be ruled out. Individuals who are predisposed to becoming depressed under conditions of stress may also, for reasons related to this predisposition or its personality correlates, be less likely than others to form close, confiding personal relationships. As a result, more recent research on vulnerability factors has focused on experimental studies.

Experimental Intervention Studies As researchers have come to recognize the methodological shortcomings of naturalistic studies of vulnerability factors, experimental interventions have become popular. Most experimental interventions have examined the effects of attempting to remove vulnerability factors on such outcomes as preoperative anxiety, recovery from surgery, and compliance with medical regimens. Related interventions have also been instituted to facilitate coping with such life crises as widowhood, rape, and job loss. The vulnerability factors manipulated in these experiments have included various types of cognition, coping

strategies, and objective coping resources. The evidence from these studies suggests that a number of presumed vulnerability factors play an important part in protecting against the onset of health problems and serious illness progression. A clearer understanding of these influences will require research advances in conceptualization and measurement as well as the development of more powerful interventions aimed at modifying vulnerability factors.

GROUP DIFFERENCES IN PSYCHIATRIC DISORDER

A large part of sociological research on psychopathology has traditionally been concerned with structural correlates of psychiatric illness such as social class, sex, and age. As shown in Table 4.2-1, the associations between these variables and the prevalence of psychiatric disorders are substantial. The most obvious hypothesis to test in examining such associations is that differential exposure to stress explains group differences in mental illness. It is now clear that this hypothesis can be rejected. Although it is true that people in comparatively disadvantaged positions in society (e.g., women, lower-class persons, nonwhites) are exposed to more stress than their counterparts, differential exposure cannot totally explain their higher rates of anxiety, depression, and nonspecific distress in general population samples. As a result, vulnerability factors have taken center stage in research on group differences; such research has shown consistently that there are group differences in vulnerability to stress and that this plays an important part in explaining differences in rates of psychiatric disorder. Current research on group differences is centrally concerned with the processes that promote vulnerability to stress.

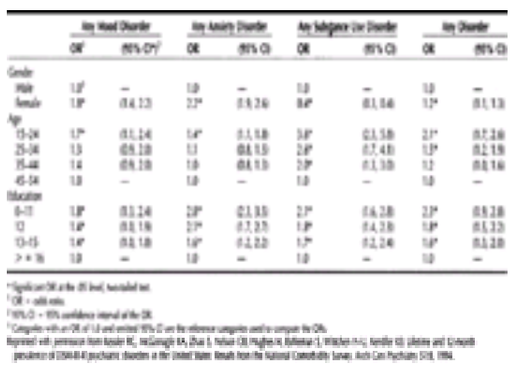

Table 4.2-1 Demographic Correlates of 12-Month DSM-III-R Psychiatric Disorders in a Nationally Representative Epidemiologic Survey of the U.S., 1990

A good example of this new work can be seen in research on the relation between social class and mental illness, one of the oldest and most firmly established associations in psychiatric epidemiology. Socially disadvantaged people have higher rates of psychiatric disorder than their more advantaged counterparts, as measured by treatment statistics, nonspecific distress in community surveys, and clinically significant psychiatric disorders in epidemiological studies. Early work on social class and psychopathology documented that lower-class people have a significantly higher probability of hospitalization and remain hospitalized longer than their middle-class counterparts; subsequent work showed that socioeconomic status is also related to psychopathology in community samples.

Until the early 1970s the dominant line of thinking in the literature on class and mental illness was that lower-class people were exposed to more stressful life experiences than those of more advantaged social status, and that this differential exposure accounted for the negative relationship between class and mental illness. This view was challenged for the first time in the Midtown Manhattan Study, in which an attempt was made to demonstrate empirically that the excess of lower-class mental health problems could be accounted for by greater exposure to stressful life experiences. Although this attempt was unsuccessful, a more complex association was documented: stressful life experiences have a greater capacity to provoke mental health problems in the lower class than in the middle class. Subsequent work has shown that this class-linked vulnerability to stress accounts for the major part of the association between social class and depression and between social class and nonspecific distress.

Differential vulnerability might arise in several ways, one of the most plausible of which is that some type of selection or “drift” of incompetent copers to the lower class might lead to the relationship between class and vulnerability. Another explanation is that one's experience as a member of a particular class leads to the development of individual differences in coping capacity as well as to differences in access to interpersonal coping resources. The available evidence supports both hypotheses. Most of the evidence for the drift hypothesis comes from studies of major mental illnesses, primarily schizophrenia. Those studies show that the early onset of a disorder can reduce one's chances of socioeconomic advancement, which seems true primarily for people who become ill before establishing a career; less severe disorders apparently do not interfere with socioeconomic achievement.

Evidence of the link between vulnerability factors and class is widespread. Lower-class people are disadvantaged in their access to supportive social relationships. Evidence also indicates that personality characteristics associated with vulnerability to stress, such as low self-esteem, fatalism, and intellectual inflexibility, are more common among lower-class people. To date, the major efforts in this area have been confined to the study of social support. The most influential work in this area has been done by the English sociologist George Brown. Brown documented that lower-class people have fewer confidants than those in the middle class, which increases their vulnerability to undesirable life events. This finding has been replicated in several investigations, but more work needs to be done to assess in parallel fashion the importance of coping strategies and personality characteristics. Also, as most investigations of class and stress have focused on life events, a more serious consideration of ongoing stressful situations may help to develop a more complete understanding of the relationship between class and psychopathology.

A related series of studies has been concerned with gender differences in anxiety and mood disorders. Community surveys show that adult women are twice as likely as men to report extreme levels of psychiatric distress and mood disorders. Although other types of psychopathology are as common among men as among women, and still others are more prevalent among men, most research has emphasized mood disorders and nonspecific distress in community samples.

Two lines of research can be identified in research on sex differences in nonspecific distress and mood disorders. The first is based on indirect assessments of role-related stress. Since the late 1980s the dominant perspective has held that women are disadvantaged relative to men because their roles expose them to more chronic stress. Because chronic stress is difficult to measure objectively, empirical analysis has used indirect assessments based on measures of objectively defined role characteristics or constellations of multiple roles to document the relation.

The second line of research on sex differences has examined stressful events. Studies have shown that there is a significant interaction between gender and undesirable events in predicting distress, with women appearing more vulnerable than men to the effects of stressful events. Several different hypotheses have been advanced to account for female vulnerability to stress, including the arguments that females are disadvantaged in access to social support, in the use of effective coping strategies, and in personality characteristics.

Although aggregate analyses of life event inventories show that women are on average more vulnerable than men, there are some events for which this is not true. Research on widows, for example, shows that women adjust to spousal death better than men; women also adjust as well as or better than men to divorce. Furthermore, women cope with financial difficulties better than do men.

A challenge for future research will be to reconcile the discrepancy between these studies of particular life events and aggregate life event surveys. The only such attempt to date, a meta-analysis of several large-scale community surveys in which the effects of different types of events were assessed separately, found no evidence that women are more distressed than men by such major life crises as job loss, divorce, or widowhood. Their greater vulnerability was primarily associated with events that happen to people close to them—death of a loved one other than a spouse being the most commonly reported event in this regard.

The greater impact of network events on women can be interpreted in several ways. One component of the difference is probably linked to the fact that women provide more support to others than men do, creating stresses and demands that can lead to psychological impairment. Also, women might be more empathic than men or might extend their concern to a wider range of people. These and other possibilities need to be investigated in the future because the role played by network events appears to account for a very substantial part of the overall gender-distress relation.

An exciting addition to the investigation of role stress and life event studies of gender difference has recently occurred as part of the larger interest in prior history of disorder. Survey research on the different predictors of onset and recurrence of depression shows clearly that while adult women are twice as likely as men to report a

recent episode of depression, there is no significant gender difference in risk of recurrence. This seeming anomaly can be explained by the fact that women are twice as likely as men to have a lifetime history of depression. Among men and women with such a history, there is no gender difference in recurrence risk; this means that an understanding of gender differences requires an understanding of the determinants of first onset. Analysis of gender-specific age-of-onset curves shows that the 2-to-1 female-to-male ratio of lifetime depression occurs by the mid-20s and that rates of first onset after that age are fairly similar for men and women. Studies of gender differences in depression need to focus not on the mid-life period, where most of the current research on gender roles is concentrated, but on late adolescence and early adulthood.

SOCIAL CONSEQUENCES OF PSYCHIATRIC DISORDERS

Sociologists have traditionally been much more concerned with the social determinants of ill health rather than with its social consequences. However, interest in the consequences of psychiatric disorders has increased over the past decade in response to the changing position of mental health treatment in managed care. Specifically, managed care plans have imposed more severe restrictions on the treatment of psychiatric disorders than physical disorders and have called for the use of evidence-based decision making in the allocation of funds for treatment. These demands necessitate an inquiry into the adverse social consequences of psychiatric disorders as well as a determination of whether these consequences can be ameliorated by treatment.

Part of this investigation has used the methods of social demography, a branch of sociology, to study the effects of early-onset psychiatric disorders on subsequent role transitions. This work has shown that early-onset disorders are powerful predictors of a wide range of adverse social consequences: school failure, teen childbearing, early marriage, marital instability, job instability, and financial adversity.

A second part of this investigation has studied role performance. This work has demonstrated that psychiatric disorders are associated with functional impairments within both family and work roles. It has been shown that people with a history of psychiatric disorders prior to marriage are substantially more likely than others to experience marital distress and marital violence, as well as to have distant or conflictual relationships with their children. As presented in Table 4.2-2, persons with a history of psychiatric disorders miss more days of work and have lower productivity while at work than workers in the same jobs who have never had a psychiatric disorder.

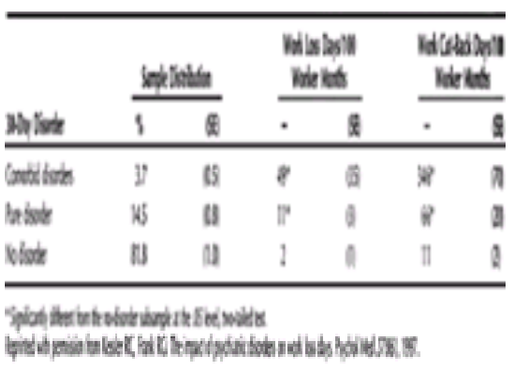

	Sample Distribution		Work Loss Days/100 Worker Months		Work Cut-Back Days/100 Worker Months	
30-Day Disorder	%	(SE)	#	(SE)	#	(SE)
Comorbid disorders	3.7	(0.5)	49*	(15)	144*	(19)
Pure disorder	14.5	(0.8)	17*	(1)	44*	(3)
No disorder	81.8	(1.0)	2	(1)	11	(2)

*Significantly different from the no-disorder subsample at the .05 level, two-tailed test.
Reprinted with permission from Kessler RC, Frank RG, The impact of psychiatric disorders on work loss days, Psychol Med 27:861, 1997.

Table 4.2-2 Multivariate Relationship Between 30-Day DSM-III-R: Disorders and Work Impairment in a Nationally Representative Epidemiologic Survey of the U.S., 1990

The studies of workplace impairments have recently caught the attention of the social policy community. Employers are being asked to reconceptualize treatment of the mental disorders of their workers as investment opportunities rather than workplace costs. Several interdisciplinary interventions are currently under way to investigate the types of workplace stresses and job demands that potentiate the effects of psychiatric disorders on performance decrements and to document the extent to which aggressive outreach and treatment of such disorders can restore productivity. The results of these studies are likely to play an important part in the evolving structure of mental health care benefits offered through the workplace.

PSYCHIATRIC HELP-SEEKING

Needs assessment surveys show that most people with serious emotional problems do not seek professional help. Furthermore, as shown in Table 4.2-3, a person seeking professional help is nearly as likely to turn to a primary care physician as to a mental health specialist. This choice is partly the result of the dearth of mental health specialists in some areas of the country, but other variables are also involved. Most of our direct knowledge about these other determinants comes from general population surveys that ask respondents about emotional problems and whether they sought professional help for these problems. These surveys have documented several consistent attitudinal, demographic, and system-dependent determinants of help-seeking behavior.

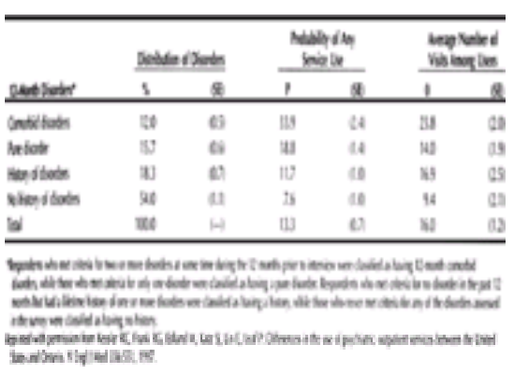

	Distribution of Disorders		Probability of Any Service Use		Average Number of Visits Among Users	
12-Month Disorders*	%	(SE)	P	(SE)	#	(SE)
Comorbid disorders	12.0	(0.7)	53.9	(2.4)	23.8	(2.0)
Pure disorder	15.7	(0.6)	18.8	(1.4)	14.0	(1.9)
History of disorders	18.3	(0.7)	11.7	(1.0)	16.9	(2.5)
No history of disorders	54.0	(1.3)	7.6	(1.0)	9.4	(2.1)
Total	100.0	(—)	13.3	(0.7)	16.0	(1.2)

*Respondents who met criteria for two or more disorders at some time during the 12 months prior to interview were classified as having 12-month comorbid disorders, while those who met criteria for only one disorder were classified as having a pure disorder. Respondents who met criteria for no disorder in the past 12 months but had a lifetime history of one or more disorders were classified as having a history, while those who never met criteria for any of the disorders assessed in the survey were classified as having no history.
Reprinted with permission from Kessler RC, Frank RG, Edlund M, Katz SJ, Lin E, Leaf P. Differences in the use of psychiatric outpatient services between the United States and Ontario. N Engl J Med 336:551, 1997.

Table 4.2-3 Probability of Any Service Use in the Previous 12 Months and Average Number of Visits by Number and Recency of DSM-III-R Disorders in a Nationally Representative Epidemiologic Survey of the U.S., 1990

Sociologists have been particularly interested in structural determinants, the strongest and most consistent of which is social class. A positive correlation between social class and help-seeking persists even though community mental health centers and other inexpensive treatment facilities have reduced the financial barriers to care. In the recent surveys carried out in the 1990s, education emerged as a stronger predictor of help-seeking than income, which suggests that some cultural facilitating factors are more important than financial resources in accounting for the influence of social class.

Women are much more likely than men to seek mental health care, even given the higher prevalence of disorder among women. Sociological research over the past few years has made considerable progress in explaining this gender difference by showing that women are more likely to recognize their problems than men, and that this recognition of a problem is critical in the decision-making process. Once they recognize that they have a problem, men and women do not differ in their likelihood of obtaining professional help. The subject of why women are more likely to recognize their problems is currently under study.

The most recent surveys of help-seeking suggest that the determinants of utilization differ markedly from one community to another, which suggests that local alternatives and barriers play a major part in determining who will seek treatment. It is likely that this result will lead to greater emphasis on comparative case studies of particular communities in the future.

It is interesting to note that most careful case studies of help-seeking document that the critical point in the process is deciding that help is needed. Most people do not know when a personal problem is big enough to warrant professional care. As social scientists come to appreciate the great importance of problem definition,

more sophisticated theories are being developed about how people make sense of symptoms. Recent research has suggested that people operate on the basis of schemas for explaining particular kinds of illnesses. These schemas contain lay accounts of the cause, course, symptoms, and prophylaxis for various illnesses. Research has already shown that a number of different schemas exist for particular medical conditions, and that characteristics of the schema held by a particular person provide important insights into the person's help-seeking and compliance habits.

Although systematic research into schemas for psychiatric disorder is in its infancy, it is likely to become a major area of investigation; this is especially true for research on psychiatric help-seeking. The results of a recent comparative general population survey of the United States and Ontario, Canada, documented that the central role of perceived need explains substantial country-to-country differences in seeking help for emotional problems. This survey found that people in the United States are much more likely than those in Ontario to seek professional help for emotional problems. However, when diagnostic data were used to investigate the relation between objective measures of need and help-seeking, the only significant between-country difference in probability of obtaining treatment was found in the subsample of respondents with low need for services, among whom more than twice as many sought help in the United States as sought help in Ontario. More focused analysis found that this difference is because of a higher level of perceived need for help in the United States than in Ontario among persons with low objective need. This difference was especially pronounced among middle-class people with insurance, raising the possibility that the demand-side controls used to limit access to mental health services in the United States are less effective in controlling unnecessary use of services than the supply-side controls used in Ontario; clearly, more detailed investigation of this possibility is needed.

Persons with comorbid conditions are more likely than those with only a single disorder to seek professional help for psychiatric difficulties. This result is consistent with the more general finding that likelihood of seeking help is positively associated with severity. Persons who have comorbid disorders are more severely distressed than those with only one disorder. Persons with comorbid conditions have a very high probability of having alcohol or drug problems (e.g., substance-related disorders) superimposed on anxiety or depression. Their help-seeking patterns direct them disproportionately to community substance abuse treatment programs where other underlying psychiatric problems are often overlooked and untreated.

Recent commentators have noted that the problems associated with comorbid diagnoses might conceivably be addressed by intervention efforts aimed at attracting psychiatrically disordered persons into treatment before they develop secondary disorders. Early interventions might be effective in preventing the onset of secondary comorbid conditions. The difficulty in implementing that approach is that the motivation to seek help often occurs only in the context of the severe role impairments associated with comorbid conditions. A future challenge will be to devise strategies to attract patients at an earlier stage in their illness, before they develop secondary comorbid disorders. Given recent evidence about the early onset of primary disorders among persons who later develop comorbid conditions, this early intervention work will almost certainly require screening in school-aged populations.

COMMUNITY RESPONSES TO THE MENTALLY ILL

Attitudes Attitudes about the mentally ill have been charted in public opinion surveys since the 1950s and dislike and fear have remained high among the attitudes surveyed. Negative attitudes are particularly pronounced among people who are poorly educated and among the elderly; men consistently report more negative attitudes than do women.

The core concerns about persons who are mentally ill revolve around their presumed unpredictability and dangerousness. These concerns have some basis in reality, as patients released from state psychiatric hospitals have comparatively high arrest rates. However, most crimes committed by released patients are property crimes that do not involve violence. Unfortunately, the mass media typically emphasize cases where violent crimes are committed by people with a history of emotional problems, thus exacerbating the problem of public misperception.

Intensely negative attitudes about the mentally ill may be a part of a larger cluster of beliefs, attitudes, and values characterized by an absence of sympathy for people who need help; a deep-seated distrust of people and institutions who are perceived as different; and a rigid outlook on what is right and wrong. Such people's views are not easily swayed by rational arguments.

Fortunately, most people have much less intensely negative feelings that can be modified on the basis of experience and as they become more knowledgeable and learn to make finer distinctions about kinds of mental illness and treatment. Visits to a psychotherapist, for example, are much less stigmatized than hospitalization for a mental illness and hospitalization in a private facility seems to be less stigmatizing than hospitalization in a public facility. Drug therapies are perceived as evidence of greater disorder and so provoke more fear and distrust of the patient than do talk therapies. For a similar reason, treatment by a psychiatrist involves more negative attitudes than consultation with a psychologist, social worker, or member of the clergy.

Survey data suggest that attitudes of community members can be influenced by contact with the mentally ill. In general, survey respondents who report knowing someone who has a history of mental illness are less negative than people who report no personal contact. It is difficult, however, to sort out cause and effect here, because negative attitudes might be associated with failure to report the mental illness of a close relative.

Family studies and studies of the reintegration of former patients into their old work roles show that contact with former coworkers and associates promotes positive attitudes about the mentally ill. Seeing a former patient perform adequately in a normal role is particularly important in this regard. Self-disclosure by the former patient about what it was like to be mentally ill and to be hospitalized also helps to promote acceptance by reducing the aura of mystery that otherwise surrounds mental illness.

Much less is known about how to change negative attitudes in the general population. Studies of the mass media show that the stereotyped depictions of former patients that commonly appear on television and in movies reinforce negative public perceptions about the mentally ill. Whether sympathetic treatment of mentally ill people in the mass media might change these negative attitudes or whether informational campaigns making use of the mass media could increase public knowledge about mental illness is less well understood.

This last issue is attracting considerable interest because several large mass media campaigns have been launched to increase public awareness, recognition, and treatment of mental illness. One such campaign was developed by the National Institute of Mental Health to increase knowledge about anxiety and depression and to encourage increased voluntary help-seeking for these disorders. Unfortunately, this campaign did not include an evaluation, so it is not known which kinds of message strategies and information channels lead most effectively to attitude and behavior changes among persons with these disorders. Judging by the experiences of health educators in conducting campaigns aimed at other public health problems, information of this sort is vitally important to successful campaign design and implementation.

A more recent related campaign has instituted an annual National Depression Awareness Day in which mass media around the country are mobilized to encourage people who might be depressed to seek treatment. Local screening sites and a toll-free number are used to facilitate screening and to encourage people who screen positively to seek treatment. Evaluations have shown that this growing campaign has been successful in bringing tens of thousands of people into treatment. Comparable programs have been established to create annual national anxiety and eating disorders screening days. Ongoing interdisciplinary research that takes into consideration the importance of social and personal barriers is refining the messages and referral strategies of these campaigns in order to increase their reach and effectiveness.

COMMUNITY REACTIONS TO SHELTER CARE HOMES

Negative attitudes about the mentally ill are important for a number of reasons, including the fact that they inhibit help-seeking for personal problems. The subject of recent sociological attention is the organizational effect of negative attitudes on attempts to establish group homes for the mentally ill.

Sociologists have done a great deal of research on collective action. Community opposition to group homes has become one of the mobilization activities studied by those working in this tradition. Research clearly indicates that middle-class neighborhoods are much more resistant to having group homes in their midst than are working-class neighborhoods; this greater resistance can be traced to effective mobilization efforts. In particular, efforts to meet and organize local opposition are accomplished more quickly in middle-class neighborhoods where a person or a committee is more likely to be selected to act on the neighborhood's behalf and where multipronged political actions are more likely to occur.

Attitudes also play an important role in the success of group homes in fostering readjustment among deinstitutionalized patients. Ethnographic research shows clearly that patients are aware of the climate of acceptance or rejection in their neighborhoods and that this influences their social functioning. The ease with which the

residents of these homes adjust to life in the community depends to a large extent on community acceptance. The conflict that can attend the creation of the home does not make a good foundation on which to build such acceptance. In general, public opinion surveys show that contact with former patients who are strangers exacerbates whatever fears and uncertainties community residents already have, particularly in neighborhoods that had previously opposed the establishment of the group home.

These issues have been neglected in most sociological studies of community opposition to group homes, which generally concentrate on structural determinants of neighborhood mobilization and on strategies available to agencies for defusing this opposition. Research is urgently needed on what happens after the home is opened and the residents must live in the neighborhood. There is evidence that contact with a former mental patient who was known prior to hospitalization can foster positive attitude changes, especially when the former patient can be seen performing adequately in normal roles. One challenge for the future will be to create structured situations that will facilitate contact between residents of sheltered care homes and their neighbors such that positive attitude changes can occur.

ORGANIZATION OF MENTAL HEALTH SERVICES

Interorganizational Coordination Research on complex organizations is one of the liveliest areas in sociology today as a result of the enormous organizational changes in American society and the innovative work of social theorists in developing new frameworks within which to understand these changes. The mental health care delivery system has been a favorite example used by these theorists to test new ideas about interorganizational linkage, because it provides unique opportunities to study a decentralized system consisting of many overlapping organizations that have complex coordinating functions.

One focus has been on the continuing diminution of state mental hospitals and the impact of downsizing on general hospitals and community-based programs. Although there is a general perception that most of the reductions in state mental hospital systems throughout the country occurred in the 1950s and 1960s, up to a 50 percent decrease in the number of inpatients occurred in many state mental health systems during the 1980s. The result has been an increased burden on general hospitals and a revolving-door policy whereby patients are treated during periods of crisis and largely ignored between admissions.

Case studies of community responses to these changes have documented enormous coordination problems and inconsistencies in organizational rationalities. Historical analyses show that these problems result from years of decision-making that lacked any overall plan or purpose. The challenge for researchers is to synthesize these case studies in order to discover the mechanisms that facilitate rationality in the relations among community organizations. Such work is currently the subject of intense interest among organizational sociologists.

A related series of studies has attempted to trace the influence of state and national policy initiatives on community-based organizations and systems. Studies have been done on how strategic decision-making in local organizations is affected by considerations concerning the future actions of state and national funding agencies. The studies show that instability of state and national initiatives to develop community-based programs leads to local processes of adaptation that were not intended by the policymakers who developed the programs. Current research is moving in the direction of comparative studies aimed at isolating characteristics of particular community systems that determine the directions of local responses.

There is also a great deal of interest in designing and evaluating organizational innovations that might improve the quality of care for the chronically mentally ill, particularly for patients unable to afford private care. Capitation programs, managed care programs, and programs that mainstream the mentally ill into existing health maintenance organizations (HMOs) created exclusively for persons with chronic mental illness are among the organizational innovations that are currently being discussed. Criteria for treatment success must be broadened for this population to include fundamental quality-of-life issues such as adequacy of housing, nutrition, employment, social integration, and other topics that are of central concern to sociologists.

Organizational Factors in Service Delivery Another kind of organizational research extends the work on job stress by studying the influence of organizational structure on the health, well-being, and productivity of its members. Some of this research has studied the structural components of mental health care organizations that affect staff satisfaction with their work; a few studies have also examined the impact of organizational structure on patient outcomes. All this work has been naturalistic rather than experimental, and comparative rather than based on case studies of individual treatment settings.

Findings include the fact that staff satisfaction and productivity are positively associated with decision latitude. Patient functioning in long-term mental hospitals is also positively associated with the decision latitude of lower-level staff. Other correlates of good patient functioning include high staff job satisfaction and high staff participation in treatment decisions. Patient functioning in short-term–care inpatient settings is positively associated with an active management style; functioning of patients in community-based shelter care homes is likely to be better when the homes are small, have flexible rules, and require patients to take some responsibility for activities of daily living.

As these results suggest, there is as yet no overarching theoretical framework that integrates the specific findings into a coherent model of organizational influence on staff and patient functioning. Integrative work of this type will be facilitated by job redesign experiments in industrial settings. Similar experiments in treatment meetings are much less common, although innovative experiments are now under way to change the structures of community-based shelter care homes in an effort to reduce the problems of staff burnout and turnover. The success of organizational redesign efforts is likely to determine whether similar experiments are carried out in a wider range of treatment settings.

Evaluation of Community Mental Health Services The development and maintenance of an effective community-based system requires a cyclical process of service planning, implementation, evaluation, and feedback. The first step in this process is usually a needs assessment, which identifies the mental health problems in the community and establishes priorities for the creation of services to address these problems. Such an assessment is vitally important to organizational success because it monitors demand for services and identifies needs not recognized by community residents.

The most direct way to conduct such an assessment is by a large-scale community survey. However, such surveys are very expensive and most local service organizations are unable to afford them. A number of innovative approaches have been devised to obtain more indirect information about need at a lower cost. These techniques include systematic interviews with key informants, the establishment of citizen advisory councils, the use of national statistics on need profiles in conjunction with small-area social indicators on community demographics, and extrapolation from data on demand for services to estimates about need for services.

Once programs are developed, research can also be important in evaluating effectiveness and targeting areas that need to be changed. Program effectiveness relies on at least two levels of research. The first focuses on success in attracting participants to the program; the second focuses on success in helping people with their problems. Behavioral scientists have been more active in the first research area than in the second area.

Research on success in attracting program participants emphasizes acceptability, accessibility, and awareness. Acceptability refers to how willing community residents are to use the new service. Accessibility involves the ease with which the program can be reached; time, distance, transportation, and financial barriers are all important considerations. Awareness relates to community knowledge that the service exists and the particular needs that it is appropriate for. An understanding of local culture is required to develop programs that are sensitive to these issues. Sociological research using ethnographic research tools or other qualitative strategies can increase the sensitivity of program staff to local norms and customs.

Research that evaluates the effectiveness of programs is much less common for several reasons: the substantial costs of implementing a carefully controlled study of treatment effectiveness, the high level of methodological sophistication required to carry out such an investigation, and the potential threat to clinicians and program administrators of openly studying the therapeutic value of their services. Although sociologists and other behavioral scientists have the expertise to do such work, this sadly remains an underdeveloped area of investigation.

Social Context of Professional Activity The medical profession is undergoing enormous changes, engendered by such things as *diagnostic-related groups* and other new payment arrangements, the shifting of care from inpatient to ambulatory settings, diversification of the medical care industry, increasingly overt competition among providers, and the growing importance of third-party payers. These changes are part of broader societal forces that include the aging of our population and cohort shifts, which have led to massive expansion in the plant facilities of the medical care industry and a marked increase in the number of physicians in the marketplace.

Sociologists have been keenly interested in the implications of these trends for the future of medicine. One perspective holds that physician domination of the health care system is too firmly established to be shaken by the changes in social context that are taking place. The legal subordination of nurses, pharmacists, and other medical care professionals to the physician is cited as critical in this regard, as are the exclusive licensing powers granted to physicians as gatekeepers of the

medical care system.

An opposing view is that the medical profession is in a period of declining power as a result of the resurgence of consumerism in medicine. The greater number of medical patients who suffer from chronic rather than acute conditions leads to the creation of interest groups made up of lay people who acquire considerable technical knowledge about their own afflictions and tend to challenge their providers. The technical diversification of medical procedures and the increasingly important contributions to health care by technician-specialists who are not physicians are also believed to play a part. With changes in the organization of professional care, new systems of ownership and management have promoted competition among physicians, which inevitably brings with it increased consumer control. Finally, the more dominant position of large insurers consolidates the bargaining position of consumers in a novel way. All these factors are particularly relevant to psychiatrists because auxiliary mental health specialists, such as clinical psychologists and psychiatric social workers, have no counterpart in the other medical specialties.

Another perspective on the changing nature of medical practice involves the proletarianization of medical work. More and more physicians, especially psychiatrists, are working as salaried employees in large, bureaucratically managed organizations. As those organizations are instituting managerial styles orchestrated by the graduates of business schools rather than of medical schools, changes in procedures for professional control will invariably occur. Formal review procedures are being applied to a wider range of professional behaviors. Within particular institutions, mechanisms are being developed to monitor and control the technical decisions of clinicians. All these trends will result in increasing external control of the domain of professional practice.

The future shape of psychiatric practice is difficult to forecast; social scientists who specialize in such prognostication have conflicting views. A more thorough understanding of the issues discussed in this section may help to modify the role of psychiatrists and future medical care systems.

SUGGESTED CROSS-REFERENCES

Other perspectives on sociocultural influences on psychiatry may be found in Section 4.1 on anthropology and psychiatry, and Section 4.3 on evolutionary biology and its applications to psychiatry. Mental health services research is discussed in Section 5.3.

SECTION REFERENCES

Katz SJ, Kessler RC, Lin E, Wells KB: Medication management of depression in the United States and Ontario. J Gen Int Med *13:*77, 1998.

Kendler KS, Davis CG, Kessler RC: The familial aggregation of common psychiatric and substance use disorders in the National Comorbidity Study: A family history study. Br J Psychiatry *170:*541, 1997.

Kessler RC: The effects of stressful life events on depression. In *Annual Review of Psychology*, Annual Reviews, Palo Alto, CA, 1997.

Kessler RC, Berglund PA, Foster CL, Saunders WB, Stang PE, Walters EE: The social consequences of psychiatric disorders II. Teenage parenthood. Am J Psychiatry *154:*1405, 1997a.

Kessler RC, Davis CG, Kendler KS: Childhood adversity and adult psychiatric disorder in the U.S. National Comorbidity Survey. Psychol Med *27:*1101, 1997.

Kessler RC, Frank RG: The impact of psychiatric disorders on work loss days. Psychol Med *27:*861, 1997.

Kessler RC, Frank RG, Edlund M, Katz SJ, Lin E, Leaf P: Differences in the use of psychiatric outpatient services between the United States and Ontario. N Engl J Med *336:*551, 1997.

Kessler RC, Olfson M, Berglund PA: Patterns and predictors of treatment contact after first onset of psychiatric disorders. Am J Psychiatry *155:*62, 1998.

Kessler RC, Zhao S, Katz SJ, Kouzis AC, Frank RG, Edlund M, Leaf P: Past-year use of outpatient services for psychiatric problems in the National Comorbidity Survey. Am J Psychiatry *156:*115, 1999.

Leong FTL, Zachar P: Gender and opinions about mental illness as predictors of attitudes toward seeking professional psychological help. Br J Guid Couns *27:*123, 1999.

Mechanic D: Organization of care and quality of life of persons with serious and persistent mental illness. In *Quality of Life in Mental Disorders*, H Katsching, H Freeman, N Sartorius, editors. John Wiley and Sons, Chichester & New York, 1997.

Mechanic D: *Improving Inpatient Treatment in an Era of Managed Care. New Directions in Mental Health Services*. Jossey-Bass, San Francisco, 1997.

Olfson M, Marchus SC, Pincus HA, Zito JM, Thompson JW, Zarin DA. Antidepressant Prescribing Practices of Outpatient Psychiatrists. Arch Gen Psychiatry *55:*310, 1998.

Olfson M, Sing M, Schlesinger HJ: Mental health/medical care costs offsets: Opportunities for managed care. Health Aff *18:*79, 1999.

Zarin DA, Pincus HA, Peterson BD, West JC, Suarez AP, Marcus SC, McIntyre JS: Characterizing Psychiatry With Findings From the 1996 National Survey of Psychiatric Practice. Am J Psychiatry *155:*397, 1998.

4.3 EVOLUTIONARY BIOLOGY AND PSYCHIATRY

MICHAEL T. MCGUIRE, M.D., AND ALFONSO TROISI, M.D.

OVERVIEW

Evolutionary studies of human behavior focus on the transmission and modification of genetically influenced behavioral traits and the systems responsible for these traits; development; function; and the environment. Because of the influence of genetic information on traits, humans, like other species, enter the world predisposed to process and interpret information in species-typical ways, to engage in certain behaviors and not others, and to work toward some goals more than others. It is unnecessary to teach infants how to bond; about emotions such as happiness, distress, disgust, anger, and fear; or about language, just as it is unnecessary to teach most adolescents about interest in the opposite sex or male and female mating strategies. This array of traits, rules, motivations, and behaviors is in part a product of the evolutionary past of all human beings, a past that determines much of what humans do, think, and feel.

When evolutionary biologists maintain that behavioral traits and internal systems cannot be adequately explained without taking into account their evolutionary history, nonbiologists often disagree. Some theorists object to what they believe are the deterministic implications of this view and disagree with its disregard for learning and culture; other theorists are concerned with the possible negative social consequences of such a view.

However, in one sense or another all sciences or purported sciences are deterministic. The sociologist who postulates that criminal behavior is caused by peer-group influence, the psychologist who predicts that a child will resent neglectful parents, or the psychiatrist who predicts that changes in central nervous system (CNS) dopamine function will hasten the onset of schizophrenia are using deterministic models in the same way as evolutionary biologists who postulate that genetic information predisposes humans to learn some things rapidly but not others. What is at issue is the *how* and the *why* of determinism and the extent of what is determined.

Concerns about learning and culture are as much a part of evolutionary explanations of human behavior as are genes. Predisposed traits are refined through learning and experience and the degree to which learning and experience influence the expression and form of traits differs across cultures because different cultures differentially value some traits more than others. Biological explanations of human behavior are incomplete unless learning and cultural influences are an integral part of the explanatory calculus—there is simply no ambiguity on this point. Social reformers often worry that if a social behavior is viewed as having genetic origins, efforts at reform will be compromised; however, the opposite is usually the case. The more that is learned about disorder etiology and pathogenesis, the more interventions are possible in the process that leads to illness and functional impairment. Evolutionary studies of human behavior have already demonstrated that a more accurate knowledge of behavior and the needs shaped by natural selection can significantly reduce mental suffering and improve psychological well-being; (e.g., John Bowlby's ethological studies of mother-child attachment have been highly influential in reforming the procedures of pediatric hospitalization).

Much of the understanding of evolution comes from the work of population biologists and their analysis of systems influencing phenotypic variation as well as from evolutionary psychologists who focus on evolved systems that guide behavior. There are important differences in these approaches. Population biologists address issues dealing with the frequency of traits in populations. Their findings permit inferences about interactions between environmental contingencies and both the adaptability and transmission of traits (e.g., walking erect, mother-infant bonding, visual acuity, increased complexity of the brain). To the degree that genetically influenced traits are associated with above-average survival and reproduction, these traits will appear at a greater frequency in subsequent generations. Studies of evolved psychological systems focus primarily on algorithms or information-processing systems that may be the postulated basis of capacities for assessing the costs and benefits of social interactions, devising novel behavioral strategies, anticipating others' needs, and altering behavior in response to self-monitored information. Both these approaches are critical to understanding and treating disorders, and both build on findings from the several branches of evolutionary biology; ethology, behavioral ecology, evolutionary anthropology, evolutionary psychology, and evolutionary psychiatry. A feature common to each of these disciplines is the critical importance placed on the theory of evolution by natural selection for understanding both past and present behavior.

Evolution by natural selection requires three conditions. First, a trait must vary within a population *(phenotypic variation)*. Second, there must be a relationship between the trait and adaptation, such as mating ability, fertility, or survivorship *(fitness variation)*. Third, there must be a consistent relationship, independent of common environmental effects, between parent and offspring for the trait *(inheritance)*. Phenotypic variation, fitness variation, and inheritance are the cornerstones of contemporary evolutionary theory, and interactions between these features are the basis of the cross-generational survival of some genes more than others.

Evolutionary theory does not assume that natural selection has any plans or goals or that the products of evolution are ideal. Environments change; a trait that is advantageous in one environment may be less advantageous in another. Also, some species eventually become extinct. Because numerous factors (e.g., genetic loading for disorders, developmental insults) influence phenotypic outcome, individuals develop variant phenotypes, and some phenotypes fare better than others. Evolutionary theory can accommodate these outcomes because it is as much about normal phenotypes as it is about variant phenotypes, and both types can be understood within the same theoretical framework.

EVOLUTIONARY BIOLOGY OR SOCIOBIOLOGY?

Following the publication in 1975 of E.O. Wilson's *Sociobiology: The New Synthesis*, and the ensuing ideological debate, many social scientists equated the term *sociobiology* with any attempt to explain human behavior using Darwin's theory. This confusion is unfortunate. What Wilson's book did was to bring together in a single volume findings from ethology, behavioral ecology, and evolutionary anthropology and combine them with many of Charles Darwin's ideas as well as refinements and changes to these ideas introduced by later investigators. The debate developed largely because Wilson speculated about the possible evolutionary bases of many human behaviors (e.g., reproduction, rituals, hierarchies) that prior to his book had been explained without consideration of possible evolutionary contributions. Times have changed somewhat since 1975. Although ideas that have their origins in evolutionary theory are now largely accepted within the social sciences, some memories of the debate and its associated confusion remain, which is why the term *evolutionary biology* is preferred over the term *sociobiology*.

KEY POINTS IN EVOLUTIONARY EXPLANATIONS OF DISORDERS

The evolutionary approach to mental disorders is based on concepts and explanations that overlap with and differ from those favored by psychiatry's prevailing explanatory models—biomedical, behavioral, sociocultural, and psychoanalytic.

Ultimate Causation Ultimate causation is the evolutionary concept that explains why in the past some traits were selected in preference to others. Current traits, such as reflexes, emotions, and the capacity to develop contingent strategies, are assumed to have been selected because they were adaptive during prior periods.

Ultimate causes differ from proximate causes such as genetic accidents, dysregulated physiological states, inadequate learning, or intrapsychic conflicts, which are the mainstay of prevailing model explanations of disorders. Ultimate and proximate causal explanations are not contradictory but complementary; both are essential for understanding disorders. For example, proximate explanations of major depressive disorder focus on putative dysfunctional biochemical states (e.g., the norepinephrine hypothesis of depression), and symptoms and signs, such as crying, feelings of despair and hopelessness, withdrawal from social interactions, and somatic indices (e.g., sleep difficulties, lack of appetite). These signs and symptoms are assumed to be the consequences of dysfunctional states. Ultimate cause explanations view many of the same symptoms and signs as evolved traits, such as signals to others that one is distressed, needs help, and is unable to engage in normal social behaviors, as well as a strategy for conserving energy *(social withdrawal)*, which reduces the frequency of costly behavior and allows restorative processes to take effect. The fact that others help those who are depressed, that persons who are depressed are not expected to carry out the normal array of social

behaviors, and that the majority of persons with depression recover on their own not only lends credence to ultimate cause explanations, but also has important therapeutic implications.

Reproduction A central theme of evolutionary theory is that traits and reproduction interact: compared to genetically influenced traits that negatively correlate with reproductive success (e.g., severe physical disabilities, hydrocephalus), traits that positively correlate with reproductive success will be present at greater frequencies in subsequent generations. The importance accorded to reproduction is sometimes taken to mean that evolutionary biologists believe that reproduction is the primary goal of all behavior; this is a misinterpretation. Human behavior (like that of closely related species), is more correctly understood as a set of acts to achieve short-term goals, such as acquiring resources, making friends, finding a mate, and reducing the effects of unpleasant emotions. These are the goals that concern humans and that are responsible for most human behavior. A consequence, often an inadvertent consequence, of successfully achieving short-term goals at a reduced average cost is increased reproduction.

Within this context, proximate factors may or may not influence reproduction and its outcomes; two such factors are of particular interest for psychiatry. First, available data indicate that the majority of persons with disorders have average or close-to-average reproductive success; that is, for most disorders, reproductive rates do not differ significantly from age- and sex-matched individuals without disorders. Disorders thus do not seem to significantly compromise reproductive capacities or reproduction, which means that disorder-influencing genetic information will be transmitted across generations and that human information-processing systems are more variant than human reproductive systems. Second, the findings that reproductive rates are not reduced among persons with disorders would seem to contradict the theory that achieving short-term goals correlates with reproductive success. Several proximate factors may be responsible for this seeming contradiction. For example, studies consistently show that assortative mating (preferential mating among persons with similar traits) occurs among persons with disorders; persons with disorders are known to cluster in enclaves in which mating options may increase; and, compared to females without disorders, those with disorders may be more susceptible to male sexual coercion.

Adaptation According to evolutionary theory any anatomical structure, physiological process, psychological process, or behavior that makes an organism more fit to survive and reproduce in comparison to other members of the same species is viewed as an adaptive trait. However, one's degree of adaptation is relative; no absolute measure exists. Although psychiatrists often use the term *adaptation*, its meaning is usually synonymous with *adjustment*. For example, according to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV) “Adaptive functioning refers to how effectively individuals cope with common life demands...” Coping may or may not be associated with improved survival or reproduction relative to other members of one's species.

Traits and Within-Trait Variation *Traits* are enduring, measurable attributes (e.g., height, baseline measures of serotonin concentration in central nervous system [CNS]) that are influenced by genetic information. *Within-trait variation* refers to the degree to which a given trait varies within a population (e.g., some people are tall, others are short; some are unusually sensitive to criticism, others can emphasize better than others). Most traits can be scaled along a continuum from suboptimal to optimal, when the measure of optimality is taken as the contribution of a trait to achieving short-term goals. Because of the genetic mixing that occurs at conception, the genetic contributions to traits as well as the strength of trait predispositions differ between parent and offspring; such differences are to be expected. Evolution has thrived on trait variation largely because different traits (e.g., sickle-cell trait) and degrees of within-trait variation (e.g., fat distribution) have different survival and reproductive value across environments.

In evolutionary analysis humans are viewed as mosaics of traits, with individual traits ranging from those that are functionally independent (e.g., visual acuity and height) to those that are highly interdependent (e.g., information-processing speed and verbal intelligence). Studies by behavioral geneticists point to three factors that influence traits and are of special interest to psychiatry: genetic predispositions, preferred environments (features of the environment to which persons preferentially respond because of their genetic makeup), and developmental environments, with the first two factors accounting for a significantly greater percentage of phenotypic variance than the third.

Many enduring features of both nondisordered and disordered behavior can be characterized in terms of an individual's location at the extremes of trait continuums: examples include impulsivity, paranoia, shyness, emotional lability, and somatization. Some disorders are best understood as clusters of suboptimal traits: mental retardation and schizoid, schizotypal, and borderline personality disorders are examples. Traits may become more obvious (e.g., impulsivity during a manic episode) or less obvious (e.g., impulsivity, depressive disorders) during disorder exacerbation.

The Social Environment According to evolutionary theory the social environment contributes to the refinement of traits and is the arena in which most short-term goals are achieved. Other individuals are essential for mating, kin investment, and nonkin reciprocal relationships (Fig. 4.3-1). Disorders occur more frequently among persons raised in adverse environments, and persons who have experienced such environments often fail to refine traits essential for social navigation in ways that optimize the achievement of short-term goals. Prevailing model explanations of disorders frequently point to adverse developmental or current environments as disorder-contributing factors. Evolutionary theory does likewise, but it also places strong emphasis on social hierarchies, the turnover rate of individuals in a social environment, and the physiological and psychological consequences of different types of social information. Social hierarchies influence social options and others' expectations of one's behavior. One's position in a hierarchy not only correlates with the type of information one receives and the benefits that accrue from social participation (e.g., high-status individuals receive a surplus of benefits), but also with the frequency of disorders (e.g., low status is associated with a decline in social influence and an increase in the frequency of depression and suicide). Rapid turnover of individuals in one's environment increases the costs and reduces the predictability of achieving goals in which others are essential participants. Social information type (e.g., negative versus positive) has a direct impact on physiological and psychological states, and repeated negative information can trigger the onset of signs and symptoms while repeated positive information can reduce their intensity.

FIGURE 4.3-1 West African Green Monkeys (“vervets”) engaged in a grooming and huddling interaction.

Sex Differences Because of their separate evolutionary histories (e.g., confronting different adaptive problems), males and females differ physically, physiologically, and psychologically in the importance they attach to goals (e.g., caring for offspring) and the strategies they utilize to achieve goals. Sex differences permit predictions about differences in the life prevalence of disorders and their manifestations. Among females disorders should be more frequent when females are unable to achieve goals that are central to female identity, such as attracting faithful mates, reproducing, and developing social support networks. Disorders associated with losses in male-male competition, failure to gain sexual access to females, failure to acquire resources, and declining social status should be more frequent among males. Epidemiological data are consistent with these predictions. Histrionic personality disorder and erotomania have strong male attraction components; anorexia nervosa and depression due to infertility have strong reproductive components. These disorders occur significantly more frequently among females than among males; for example, greater than 90 percent of the cases of anorexia nervosa occur among females, and 37 percent of females experience psychological disruptions in association with marital infertility whereas only 1 percent of males show these disruptions. For males, impulse-related disorders, depressive disorder or suicide following competitive losses, and alcohol have strong male-male competition and resource acquisition components; many types of sexual deviance and spousal abuse have female access components. These disorders are significantly more common among males. A clear implication of this fact is that differences in female-male neuroanatomy, neurochemistry, hormonal activity, goal priorities, and strategy preferences should be consistently integrated into disorder explanations and treatment plans.

Function and Functional Capacities Organisms are designed to do something, and what they do reveals their goals and provides information whereby their efficiency at achieving those goals can be assessed. For example, the function of foraging is to obtain food; the functions of social interactions include recognizing

possible mates and good reciprocators and obtaining information about resources. Some functions, such as mother-infant bonding, are closely tied to specific behaviors whereas others, such as acquiring resources, are associated with a larger number of behaviors.

Functions and the behavioral skills necessary to carry them out (functional capacities) need to be distinguished. Two persons may aspire to gain admission to medical school, yet they may differ in their capacities to achieve the goal. Functional capacities are traits that are influenced by genetic information and are subject to refinement; their goal-achieving efficiency varies across environments. Accurately assessing functions, functional capacities, and short-term goal outcomes requires the moment-to-moment assessment of behavior in the social environments in which it occurs—the success of a social signal is in part defined by the response of the person receiving the signal.

Functions and functional capacities are relevant to discussions of disorders in numerous ways: with a few exceptions, the majority of disorders are associated with both compromised functional capacities and goal achievement and improving function and functional capacities are key objectives of treatment.

Strategies Strategies are the ways in which individuals plan and carry out behavior. For all but the most rudimentary of acts, and especially when others are involved, a person's initial strategies usually require adjustment. When individuals without disorders find that a strategy is inefficient, they may devise alternative strategies. The same is true for persons suffering from minor symptoms, such as depressive disorder and anxiety: they may attempt to reduce their symptoms by avoiding persons or situations that intensify their unpleasant feelings. Both prevailing and evolutionary models of behavior agree on these points. Where evolutionary models differ from prevailing models is in the view that the presence of a disorder does not mean that there is a discontinuation of either the development or use of strategies to reduce symptoms or achieve short-term goals. For all but a few disorders, motivations to achieve short-term goals appear to vary minimally among persons with and without disorders. What often distinguishes persons with disorders is how they go about trying to achieve goals; for example, they frequently use strategies that are counterproductive and they usually have trouble changing strategies that are unproductive.

The evolutionary view that caused motivations remain intact despite the presence of disorders invites alternative interpretations of a wide array of disorder-related behavior. For example, psychic defense or denial may be viewed as selectively attending to and processing external information, while repression is selectively attending to and processing internal information. Rationalization yields self-serving explanations for one's behavior; intellectualization serves to separate emotions from cognition. These defenses take on a new meaning when they are viewed as strategies to facilitate short-term goal achievement. Signs such as social withdrawal, excessive anger, and excessive responses to specific words (e.g., observed among some persons with schizophrenia) may reflect a strategy: withdrawal removes one from painful or frightening situations and excessive anger or excessive emotional responses to normally neutral words usually changes others' behavior.

Emotions In the evolutionary context one function of emotions is as a source of information about short-term goal achievement. Depression informs one that costs have exceeded benefits in past attempts to achieve goals. Anxiety informs one that future goals are unlikely to be achieved or that achievement will result in a negative cost-benefit balance. Frustration informs one that negative cost-benefit balances exist with respect to ongoing goals. Pleasure indicates that goals have been achieved or that they are likely to be achieved with a positive cost-benefit balance.

A second function of emotions is to initiate strategy and behavior change. Evolutionary biologists view emotions as behavior-initiating systems. Pain, anxiety, and anger generally initiate behavioral responses more rapidly than cognitive assessments (e.g., acute fear of a situation triggers rapid flight-or-fight responses whereas cognitive assessments may delay responses). That excessive emotions often exceed the intensity of stimuli is thought to reflect the fact that mankind has evolved to believe that it is better to be safe than sorry.

A third function of emotion is social regulation. Emotions are signals: smiles signal pleasure and usually initiate others' social participation; anger signals that one is displeased and that interpersonal aggression may be forthcoming; depression in response to interpersonal competition signals that one has lost the competition.

More enduring and debilitating emotions require other explanations. In the absence of precipitating causes, debilitating emotions, which are observed in major depressive disorder, manic episode, and panic disorder in part reflect a suboptimal capacity to discontinue (“turn off”) emotion-initiating processes, and in part reflect a compromised ability to act in ways in which benefits exceed costs. In different combinations, genetic information, within-trait variation, physiological dysregulation, and a failure to refine traits may contribute to these types of emotions.

Information Processing Each of the preceding topics involves information processing. A feature common to most disorders is atypical or distorted information processing, such as misinterpreting reality, others' responses and motives, one's own importance, and the meaning of one's symptoms. Another common feature is the compromised ability to correct information misinterpretations. A third feature is the reduced ability to alter strategies on the basis of self-monitored information. Generally, two types of evolutionary interpretations are applied to misinterpretations. One is to ascribe information distortions to dysfunctional algorithms. Viewed this way, distortions are state conditions and they may normalize when disorders remit. The other interpretation emphasizes deficit conditions or the absence of processing capacities. Both views are clinically relevant. Tensions between algorithm capacities and the social environment are observed in highly stressful environments (e.g., dysfunctional marriages, intense parent-offspring conflict), and such tensions can result in dysfunctional algorithms that contribute to depressive disorders, adjustment disorders, amnestic disorders, and anorexia nervosa. Because each of these disorders usually improves, dysfunctionality rather than a deficit condition is implied. In the absence of precipitating causes, deficits are implicated for disorders in which emotions are debilitating and there is functional deterioration relative to predisorder states (e.g., schizophrenia).

Self-Interest and Altruism That individuals have evolved to act in their own interest and not in the interest of the species is a basic premise of evolutionary theory. Sibling rivalry and parent-offspring conflict are examples of self-interest and they illustrate some of the consequences of self-interested behavior in relationships in which genes are only partially shared. For reasons of self-interest siblings compete for parental investment. Parent-offspring conflict occurs because parents tend to limit their investment in offspring in order to achieve other goals, such as attaining resources or maintaining social support networks.

At first glance, self-interested behavior would seem to preclude the possibility of altruism. Yet altruistic behavior does occur (e.g., donating kidneys, helping friends and strangers). How is it possible that individuals have evolved to be both self-interested and altruistic? Evolution-based answers to this seeming paradox are found in the ultimate cause of explanations of kin selection and reciprocal altruism.

Kin selection explains preferential investments in kin (a cost) without an associated immediate benefit. For example, a mother's care for her offspring is both self-interested and altruistic behavior: the mother's investment (the altruistic component) increases the chances that her offspring will mature and reproduce. To the degree that her offspring reproduces, the mother's genes will be replicated in the next generation (the mother's self-interested component as well as the delayed benefit). In more complex situations, one can invest in collateral kin (e.g., sisters and brothers) and increase one's genes in subsequent generations provided that the collateral kin reproduce. *Reciprocal altruism* is defined as an exchange of favors between nonkin separated by a time interval (the time interval distinguishes reciprocal altruism from *mutualism* or *joint cooperation*): one does a favor for another (a cost and the altruistic component) and the favor is reciprocated (the benefit and the self-interested component). On average, the greater the percentage of genes shared by two individuals, the greater the probability of investment; the less offspring appear to be good reproductive bets, the less investment they will receive and nonkin nonreciprocators will be socially ostracized. Data are consistent with these predictions.

Some of the strongest evidence supporting the algorithm concept can be found in the analysis of reciprocal interactions. Most such interactions involve exchanges of different currencies (e.g., a person feeds a neighbor's dog for a week and in return receives a suggestion about how to invest in the stock market; both parties feel that the favor has been repaid). Because most reciprocal exchanges are of this type, and because precedents for specific types of exchanges are often lacking, it can be argued that algorithms are evolved capacities capable of assessing the costs and benefits of different currencies.

Kin selection and reciprocal altruism are both key to describing and understanding disorders in evolutionary context. Nondisordered behavior generally equates with high degrees of kin investment and successful nonkin exchanges. Among persons with disorders, both kin investment and reciprocal relationships are often compromised and counterproductive. For example, persons with disorders frequently seek excessive kin investment. In turn, kin reject such demands and may at times reject the persons making them. Similarly, persons with disorders often fail to reciprocate or they may reciprocate insufficiently; in turn, relationships are likely to be discontinued.

There are multiple reasons for compromised kin investment and reciprocation, including misperceptions, dysfunctional and suboptimal algorithms, as well as the behavior of kin and nonkin. Compromises reflect a breakdown in fundamental social behavior capacities, contribute to disorder onset, and extend disorder longevity by compromising the achievement of short-term goals.

Genetic Contributions to Disorders Multiple genetic factors contribute to disorders. Genes have numerous functions (e.g., structural, regulatory) and these functions may be defective; genomic imprinting (parental determination of gene expression) can occur; genetic deletions contribute to specific disorders (e.g., Prader-Willi and Angelman syndromes); genetic loading occurs for specific disorders (e.g., Huntington's disease, schizophrenia, autistic disorder); and stochastic accidents are also known—deoxyribonucleic acid (DNA) replication is not a perfect process.

The preceding list is only a sampling of possible genetic contributions to traits and disorders. Evolutionary biology introduces other possibilities. *Pleiotropy* is the control of one or more phenotypic characteristics by one gene or a set of genes. Genetic information contributing to minimally adaptive traits may be carried along from generation to generation because such traits are controlled by the same gene or genes responsible for an adaptive trait or traits. Perhaps the clearest example of pleiotropy is that of greater fecundity during reproductive years (the adaptive trait) coupled with postreproductive vulnerability to disorders such as depressive disorder (the minimally adaptive trait). Pleiotropy could qualify as an ultimate cause explanation. *Exaption* is another possibility: a trait was adaptive in the past, but it is not adaptive in the present. This possibility is the basis of what may be called the *genome-lag hypothesis:* the environment changes faster than the species genome, one result of which is that some traits selected in the past are not well adapted to the current environment. *Genetic drift* (change in gene frequencies by chance alone) is also a possibility. The contribution of genetic drift to disorders is not yet clear.

Mutation may be the basis for some disorders. Over the course of evolution, mutations are a source of genetic change and genetic variation, and some traits that currently are viewed as adaptive are possible products of mutations. Sickle cell trait is a frequently cited example of a mutation that became adaptive in malaria-prone areas. However, subsequent migration of persons with this trait to non–malaria-prone areas introduced the sickle cell gene into neighboring populations along with its undesirable consequences. Despite this, and with a few possible exceptions, mutations are an unlikely explanation of most disorders or features of disorders. As a rule, mutations are either neutral or they have deleterious effects and genes that have deleterious effects usually do not remain in populations for more than a few generations. For severely debilitating disorders that are relatively uniformly distributed throughout the world, such as the residual type of schizophrenia for which available evidence does not point to prevalence changes, the possible part played by mutations is more difficult to infer. For example, if schizophrenia is due to a mutation it would be necessary to reason that the disorder-causing mutation is very old (e.g., occurred prior to ancestral migration out of Africa or Asia); the disorder has not negatively affected reproductive success in ways that can be easily demonstrated; and selection against the mutation (if it has occurred or is occurring) is minimally influenced by such factors as the physical environment, cultural variables, or the social consequences of the disorder.

Homozygosity due to inbreeding also needs to be considered. There are a number of reports in which degree and frequency of inbreeding are associated with an increased prevalence of disorders. However, it is doubtful that any of the widely distributed, severely debilitating disorders can be explained by this theory. The near-uniform cross-cultural prevalence of many disorders, when compared to cross-cultural differences in inbreeding practices (e.g., the acceptability versus the unacceptability of marriages between first cousins), suggests as much. Minimal selection pressure is another possibility. Disorders such as reading disorder that go unnoticed in nonliterate societies or disorders that appear after the critical reproductive years, such as dementia of the Alzheimer's type, might be explained in this way. In principle, minimal selection effects are potentially distinguishable from pleiotropy effects because of the absence of associated adaptive traits.

Evolutionary Theory as a General Theory of Behavior When causal formulations are focused upon, there are a number of ways in which evolutionary theory might relate to psychiatry's prevailing models. One is as a competing theory. Viewed this way, evolutionary theory could be compared with one or more prevailing models for its explanatory power, much as the explanatory power of biomedical and psychoanalytic explanations of depression associated with loss can be compared. It is also possible to view evolutionary theory and each of the prevailing models as partial explanations of disorders. Most clinicians tend to explain disorders in this way; for example, primary explanations of bipolar I disorder are usually physiological whereas explanations for drug noncompliance are usually psychological.

A third way is to argue that evolutionary theory is a more general and inclusive explanatory system compared to any of the prevailing theories, and that it can integrate both prevailing model (proximate) and evolutionary explanations (e.g., ultimate causation, sex differences, adaptive strategies, and function). Nothing in evolutionary theory precludes explaining disorders as being partially caused by genetic loading, physiological dysfunction, dysfunctional or inadequate learning, adverse environments, or intrapsychic conflicts. Evolutionary theory can increase the scope and complexity (e.g., the addition of ultimate cause and sex difference explanations) of causal explanations and provide a structure for assessing, interpreting, and testing hypotheses. Physiological mechanisms, psychological processes, and behavioral responses are conceptualized as different components of integrated systems that have evolved because they solved adaptive problems (short-term goals). According to this view, prevailing models are more focused and explain aspects of disordered and nondisordered behavior whereas evolutionary theory integrates these findings and offers novel explanations for disorders within a theoretical framework that is relevant to the species.

IMPLICATIONS OF EVOLUTIONARY THEORY FOR CLINICAL PSYCHIATRY

The preceding discussion has numerous implications for clinical and research psychiatry in the areas of diagnosis, epidemiology, etiology, pathogenesis, and therapy.

Diagnosis The lack of balance between psychological assessment and behavioral analysis in clinical psychiatry is a major obstacle to the development of valid diagnostic criteria. The clinical phenomenology of disorders includes subjective mental experiences and objective behavioral changes, yet the diagnostic process is based primarily on the evaluation of psychological symptoms as revealed by patients. Minimal attention is paid to the detailed description and measurement of behavior and, except for periods of hospitalization, patients are rarely observed in the environments in which they live. Despite the increasing awareness of the inadequacy of disorder-based research (e.g., one disorder, one cause), psychiatry has failed to embrace a valid alternative. A major reason for this failure is that psychiatry lacks a theoretical and methodological framework for studying behavior and function. Evolutionary biology provides such a framework; for example, ethology (the evolutionary study of animal behavior) combines the direct observation of behavior with quantitative recording and a theoretical emphasis on human behaviors that are closely related to social functioning. Since the 1970s, ethological studies addressing numerous areas of ambiguity, such as prediction of drug response and definition of diagnostic boundaries, have appeared in the psychiatric literature.

For example, in 1989 the nonverbal behavior of 22 patients depressive disorders was studied to determine whether their response to antidepressant drugs (50 to 100 mg a day of amitriptyline [Elavil] for 5 consecutive weeks) could be predicted on the basis of ethological profiles at baseline. Patients' behavior during interviews was videorecorded and scored according to an ethological scoring system that included 37 different behaviors, primarily facial expressions and hand movements. For data analysis, these behaviors were combined to form eight functional categories reflecting various aspects of the patients' emotional and social functioning. At the end of the study, patients were divided into two outcome groups on the basis of their final scores on a depression rating scale. At baseline, responders and nonresponders did not differ with respect to sex, age, education; revised third edition of DSM (DSM-III-R) diagnosis, or severity of depression. However, ethological profiles of the two outcome groups did differ, with nonresponders showing significantly more assertive behavior than responders.

In an evolutionary context, suffering, statistically variant behaviors, or organic changes are neither sufficient nor necessary for defining disorders. Persons who have hypomania and antisocial personality disorder do not always suffer; neurofibrils are found at autopsy among persons who had no signs of dementia during their lifetime. Also, statistically atypical physiological measures often are not associated with disorders and statistically normal measures often are; putative atypical genetic profiles do not uniformly correlate with disorder phenotypes (e.g., nonafflicted monozygotic twin of a twin with schizophrenia). Such findings invite consideration of alternative criteria to identify disorders and explain their existence.

If evolutionary theory should be used for developing a taxonomic system for disorders, the system would primarily emphasize function and secondarily emphasize signs and symptoms. This point applies as much to dementia of the Alzheimer's type as it does to residual schizophrenia, reading disorder, amnesia, phobias, adjustment disorders, paraphilian, and mood disorders. This view clearly differs from the thinking between DSM-IV and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) thinking and its implications are unambiguous. Many persons who suffer from chronic anxiety or intermittent episodes of depression; who somaticize, who have recognizable but mild phobias; or who have mild forms of narcissistic, dependent, or obsessive-compulsive personality disorders function and achieve short-term goals at an average or above-average rate. If compromised function is the primary criterion for disorder identification, such persons would not be viewed as having a disorder, although their symptoms would not be ignored or go untreated. Conversely, many persons who are not strikingly symptomatic experience great difficulty in achieving short-term goals and function well below average; such persons would receive a diagnosis in a functionally oriented taxonomic system.

As regards common final pathways or constraints on phenotypic expression, one can learn only so fast, jump only so high, think only so rapidly, and distort reality in only so many ways. Constraints limit the phenotypic outcomes of genetic mistakes, of genetic loading for disorders, of physiological dysregulation, and of ultimately caused behavior. Social withdrawal, anxiety, depression, fear, concentration difficulties, misperceptions, erratic behavior, and delusions are all present in more than two dozen currently classified disorders. Given that these features are present across a broad array of disorders that are not necessarily obviously related, attempts to

diagnose disorders at the sign-symptom level inevitably lead to porous diagnostic categories, favor comorbidity, and may confuse intervention strategies.

Epidemiology Recent epidemiological studies suggest that the life prevalence for disorders is approximately 50 percent, and it exceeds 50 percent if the upper estimates of the disorder prevalence reported in DSM-IV are totaled. Such percentages have several possible explanations, including that disorder criteria are too broadly defined or that epidemiologists have been overzealous in their attempts to identify disorders. The preferred explanation has two parts. First, current diagnostic practices do not reflect an appreciation of the degree of within-trait variation characteristic of *Homo sapiens* and this variation is often equated with disorders. From an evolutionary perspective, trait variation equates with disorders only if traits are the basis of functional impairment. Second, some disorders and many features of disorders are adaptive and reflect adaptive strategies.

Etiology Evolutionary theories of disorder causation emphasize multiple causes; rarely are disorders explained on the basis of single causes. For example, suboptimal traits resulting from the chance effects of genetic mixing at conception, deficit conditions, genetic loading for disorders, and adverse environments can all be contributing factors to disorders. Chance genetic mixing can partially account for those disorders in which pedigree findings are negative and developmental disruptions have not occurred. The absence of an enzyme, such as monoamine oxidase type A (MAO_A), is an example of a deficit condition; the absence of this enzyme can partially account for the associated mental retardation and increased aggressivity. Genetic loading for disorders can partially account for disorders in which pedigree findings are positive but genetic penetrance is partial, such as schizophrenia, bipolar I disorder, and several personality disorders. Adverse developmental and adult environments can partially explain some features of personality disorders, adjustment disorders, the effects of partner abuse, and stress-related disorders. A lack of fit between ultimately caused behavior and current environments can partially explain childhood and adult phobias. Many of the proximate mechanisms postulated by prevailing models, such as low serotonin activity in the CNS, can partially explain impulsive and aggressive behavior.

In addition, evolutionary theory provides two other etiological explanations: some disorders represent adaptive strategies and many features of disorders are manifestations of adaptive strategies. Time-limited cases of anorexia nervosa (90 percent of which resolve), may be partially understood as an ultimately caused strategy to delay mate selection, reproduction, and maturation in situations where males are perceived as scarce. Antisocial and histrionic personality disorders, a percentage of persons with attention-deficit/hyperactivity disorder, and malingering can be partially understood as evolved, ultimately caused, high-risk strategies that often have high payoffs: for example mate and resource acquisition, reproduction, and social influence may be average or above average. Attempts to adapt are seen among persons with narcissistic and dependent personality disorders, conversion disorder, and other somatoform disorders: information is manipulated in self-interested ways (e.g., excessively personalized); and there is an excessive use of specific strategies (e.g., signaling their distress to alter others' behavior). The social withdrawal and signaling features of major depressive disorder have already been discussed; similar interpretations apply to anxiety. An erotomanic delusion in a postmenopausal single woman may represent an attempt to compensate for the painful recognition of reproductive failure. An extension of this line of reasoning suggests that a detailed ethological assessment of disorder features would reveal important differences in functional behaviors and symptom and sign configurations in association with actual or perceived adaptive problems; data are consistent with this expectation.

Pathogenesis Anyone familiar with psychiatry knows that there are different causal hypotheses for disorders. Psychoanalysts emphasize intrapsychic distortions and conflicts, behaviorists emphasize inadequate or dysfunctional learning, and biomedical psychiatrists emphasize atypical genes and physiology. Often, these hypotheses distinguish between organic and other putative causes of disorders. From an evolutionary perspective, the organic-nonorganic distinction is not useful; until proven otherwise, disorders imply different percentages of genetic, physiology, psychological, trait, and environmental contributions. For some disorders, there is strong evidence favoring the importance of a specific proximate cause. Perhaps the clearest examples are disorders that are caused by genetic mistakes or transmission (e.g., certain forms of mental retardation, Klinefelter's syndrome, Huntington's disease). However, these disorders show a wide array of clinical profiles that require explanation. For other disorders, evidence suggests that genetic information combined with some type of developmental perturbation is associated with an increased probability of disorders: monozygotic twin data addressing the occurrence of schizophrenia, alcohol use disorders, bipolar disorders, and antisocial personality disorder are consistent with this view. For disorders involving risk-taking among adolescent males, trait variation is likely; however, even in these cases there are significant differences in the degree of risk taking.

Causal interpretations become more complex when specific causal variables are considered. Physiological assessments provide a convenient example. Low concentrations of 5-hydroxyindoleacetic acid in cerebrospinal fluid (considered as a measure of serotonin release in CNS and of MAO_A activity) is associated with depression, arson-related behavior, and aggression. Further, because some disorders remit spontaneously, some physiological measures are likely to be indicators of disorder-ameliorating processes. This is also true of molecular biological approaches to disorders: DNA ® ribonucleic acid (RNA) ® protein ® enzyme ® neurotransmitter ® behavior, whereby dysfunctionality can occur at any point in the chain. As regards learning explanations, trouble can occur at any point during development. The influence of environmental information on physiological states and disorder triggering has been established and recent evidence suggests that this influence is stronger than was previously known. If algorithms are seen as evolved information-processing systems, pathogenic explanations are contingent on identifying which algorithms are dysfunctional or suboptimal. Because algorithms are thought to have anatomical, physiological, and psychological properties, it is likely that dysfunctional or suboptimal algorithms will be associated with disorder-related features within each of these parameters. For example, an individual with borderline personality disorder goes from rational reasoning to reasoning that is strongly influenced by emotions, which suggests that reasoning algorithms are unstable and that this instability is reflected in the difficulty such persons experience in their efforts to manage their emotions.

Given the preceding points, it is reasonable to assume that many factors contribute to the vast majority of currently classified disorders, that the timing of and interactions between disorder-contributing factors are crucial in disorder development and type, and that factors such as sex differences, ultimate causation, and within-trait variation can explain the features of many disorders. Postulating multiple causes provides little information on specific causes. However, this view ignores several important facts. Thus far, prevailing models have provided compelling explanations for only a very small percentage of disorders, and for the most part these disorders are caused by specific, identifiable genetic abnormalities or substance abuse. For the majority of disorders, there are simply suggestive correlations between putative causes and outcomes. Further, accepted conventions for testing hypotheses (e.g., the effect of a selective serotonin reuptake inhibitor) favor holding potential contributing or confounding variables constant while exploring the effects of a single independent variable on some outcome measure. Although this approach is methodologically defensible, it often inadvertently leads to reductionistic interpretations of the causes of various disorders.

Therapy In many instances, evolutionary-based interventions do not differ from prevailing therapeutic practices. Not only are there are a limited number of intervention options, but an overlap in interventions would be expected if features of prevailing models could be integrated into an evolutionary framework. Many of the existing therapeutic modalities (drugs, learning, discussion, environmental alterations) will be useful in treating different disorders at different points in the course of disorders. Nevertheless, treating disorders in an evolutionary context has several specific implications and priorities.

Perhaps the most obvious implication is that ultimate causes cannot be treated. However, they can be used, such as placing persons in environments in which there is an increased probability of achieving high-priority goals. Second, traits do not disappear, but in certain instances they can be used to therapeutic advantage, such as in refining suboptimal traits so that they are more useful for social navigation. The priorities of therapy can be outlined as follows: (1) If possible, treat causes, not secondary effects. (2) Improve an individual's capacities to achieve short-term goals. This often necessitates a tradeoff between symptoms and functional capacities. To the degree that immediate symptom relief necessitates the use of drugs that compromise function, function should be favored in preference to symptom reduction because improved function inversely correlates with symptom reduction. Treatment should be aimed at factors that lead to strategy failures. (3) Avoid treating adaptive strategies unless they cause great discomfort to an individual or members of the individual's social environment. Thus, histrionic personality disorder that causes distress to a patient and others would be treated whereas social withdrawal in depressive disorder and schizophrenia, or emotional responses in adjustment disorders, might not be high-priority treatment targets. (4) Optimize environmental conditions; place persons in environments in which they are most likely to achieve high-priority goals. (5) Use multiple therapeutic modalities and change these modalities over the course of treatment.

Evolutionary theory offers a framework for the improved understanding of mental illness as well as a theoretical system that can integrate findings from the neurosciences. The serious introduction of an evolutionary approach into psychiatry would allow for the integration of features of prevailing models, introduce novel causal hypotheses, and provide a structure for interpreting new data and reinterpreting old data. A clear implication of this view is that new and different types of information need to be collected and different types of analyses need to be applied in order to improve the understanding of disorders. Examples include the assessment of function, functional capacities, and strategies; timing details with respect to physiological and behavioral change; detailed assessments of environmental contingencies; and far greater consideration of within-trait variation.

SUGGESTED CROSS-REFERENCES

The neural sciences are covered in Chapter 1, genetics in Section 1.17 and Section 1.18, aggression in Section 27.3, and animal research in Section 5.4. Mood disorders are described in Chapter 14, anxiety disorders in Chapter 15, and personality disorders in Chapter 24.

SECTION REFERENCES

Alexander RD: Epigenetic rules and Darwinian algorithms: The adaptive study of learning and development. Ethol Sociobiol *11:*241, 1990.

Baron-Cohen S, editor: *The Maladapted Mind: Classic Readings in Evolutionary Psychopathology*. Psychology Press, Hove, UK, 1997.

Brothers L: *Friday's Footprint*. Oxford University Press, New York, 1997.

Brunner HG, Nelen M, Breakefield XO, Ropers HH, Van Oost BA: Abnormal behavior associated with a point mutation in the structural gene for monoamine oxidase A. Science *262:*578, 1993.

Buss DM: *The Evolution of Desire*. Basic Books, New York, 1994.

Daly M, Wilson M: *Homicide*. Aldine de Gruyter, New York, 1988.

*Darwin C: *The Expression of the Emotions in Man and Animals*. University of Chicago Press, Chicago, 1965.

Flinn MV: Culture and the evolution of social learning. Evol Hum Behav *18:*23, 1997.

Gardner RJ: Sociobiology and its applications to psychiatry. In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1993.

Grady MF, McGuire MT: A theory of the origin of natural law. Journal of Contemporary Legal Issues *8:*87–129, 1997.

Kalin NH, Shelton SE: Ontogeny and stability of separation and threat-induced defensive behavior in rhesus monkeys during the first year of life. Am Primatol *44:*125, 1998.

Kessler RC, McGonagle KA, Zaho S, Nelson CB, Hughes M, Eshelman S, Wittchen HU, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the national comorbidity survey. Arch Gen Psychiatry *51:*8, 1994.

Marks IM: *Fears, Phobias, and Rituals*. Oxford University Press, Oxford, 1987.

McEwan KL, Costello CG, Taylor PJ: Adjustment to infertility. J Abnorm Psychol *96:*108, 1987.

McGuire M, Troisi A, Raleigh M: Depression in evolutionary context. In *The Maladapted Mind: Classic Readings in Evolutionary Psychopathology*, S Baron-Cohen, editor. Psychology Press, Hore, England, 1997.

McGuire MT, Essock-Vitale SM: Psychiatric disorders in the context of evolutionary biology: The impariment of adaptive behaviors during the exacerbation and remission of psychiatric illnesses. J Nerv Ment Dis *170:*9, 1982.

McGuire MT, Marks I, Nesse RM, Troisi A: Evolutionary biology: A basic science for psychiatry? Acta Psychiatr Scand *86:*879, 1992.

*McGuire MT, Troisi A: *Darwinian Psychiatry*. Oxford University Press, New York, 1998.

McGuire MT, Troisi A: Prevalence differences in depression among males and females: Are there evolutionary explanations? Br J Med Psychol *71:*479, 1998.

Nesse RM: Evolutionary explanations of emotions. Hum Nature *1:*261, 1990.

Nesse RM, Berridge KC: Psychoactive drug use in evolutionary perspective. Science *278:*63, 1997.

Nesse RM, Williams GC: *Why People Get Sick*. Random House, New York, 1994.

Pawlowski B, Lowen CB, Dunbar RIM: Neocortex size, social skills and mating success in primates. Behavior *135:*357, 1998.

Plomin R, Owen MJ, McGuffin P: The genetic basis of complex human behaviors. Science *264:*1733, 1994.

Price J, Sloman L, Gardner R Jr, Gilbert P, Rohde P: The social competition hypothesis of depression. Br J Psychiatry *164:*309, 1994.

Rose MR: *Evolutionary Biology of Aging*. Oxford University Press, New York, 1991.

Stevens A, Price J: *Evolutionary Psychiatry*. Routledge, London, 1996.

*Trivers R: *Social Evolution*. Benjamin Cummings, Menlo Park, CA, 1985.

Troisi A, McGuire MT: Evolution and mental health. In *Encyclopedia of Mental Health*, M Friedman, editor. Academic Press, New York, 1998.

*Troisi A, Moles A: Gender differences in depression: an ethological study of nonverbal behavior during interviews. J Psychiatr Res *33:*243, 1999.

Troisi A, Pasini A, Bersani G, Grispini A, Ciani N: Ethological predictors of amitriptyline response in depressed outpatients. J Affect Disord *17:*129, 1989.

*Williams GC: *Adaption and Natural Selection*. Princeton University Press, Princeton, NJ, 1966.

*Wilson EO: *Sociobiology: The New Synthesis*. Harvard University Press, Cambridge, MA, 1975.

4.4 CULTURAL PSYCHIATRY

MANUEL TRUJILLO, M.D.

Practicing psychiatrists in the United States are keenly aware of the need for a body of knowledge that allows them to provide state-of-the-art psychiatric services across language and cultural barriers. Clinicians, particularly those who practice in urban settings, will increasingly be called to evaluate and treat patients in the many cultural and linguistic groups that constitute today's multicultural society. In treating a patient who speaks a language other than English and holds beliefs at variance with mainstream culture, the psychiatric knowledge and technical expertise of today's practitioner falls short of the goal of providing optimum psychiatric and psychological care. Different attitudes, knowledge, and skills are needed for the practicing clinician of the next century to provide technically correct and culturally competent services. Cultural psychiatry, thanks to relatively recent explosions of research knowledge, can provide the tools needed to extend accurate diagnoses and effective treatments to patients across linguistic and cultural barriers. Cultural psychiatry aims to define the impact of culture on the diagnosis and treatment of psychiatric disorders and to provide culturally syntonic guidance for the design and implementation of culturally competent systems of psychiatric care.

Cultural psychiatry draws on many basic and applied disciplines to build its essential constructs. Anthropology (both cultural and medical anthropology) supplies cultural psychiatry with essential insights into the behavior of people in their natural habitats, native views on health and illness, descriptions of indigenous healing systems, and the role of the healer and rituals of healing in different ethnic and cultural groups. Sociology elucidates the relation of basic psychological processes and psychiatric disorders to such human universals as age, gender, and social and occupational status. Epidemiology generates data about the differential incidence and prevalence of psychological distress and disorders in different cultural groups as well as comparative studies of the pathogenesis, onset, pathoplasty, course, and outcome of diagnostic entities across diverse cultural groups. Developmental psychology and psychopathology illuminate the impact of culture on normative personality development and its disorders. Cultural psychopathology investigates the various culture-bound or culture-specific syndromes, and finally, ethnic psychopharmacology studies the impact of race and ethnicity on use, metabolism, and differential effects of standard psychotropic medication and expands our knowledge of the biological and psychological effects of ancient systems of diets and curative herbs.

Cultural psychiatry's growth as a scientific discipline parallels the growth of modern clinical psychiatry. Emil Kraepelin, in an expedition to Java to study mental illness in different communities, noted lower rates of bipolar disorder among patients in Java and a lower incidence of “feelings of guilt” in Javanese depressive patients in the early days of the twentieth century. Eugen Bleuler was also keenly aware of culturally generated differences in the expression of psychopathology by English and Irish patients. The steady growth of immigrants and migrants from Hispanic and Asian backgrounds following World War II gave additional impetus to the growth of cultural psychiatry in the United States, since it exposed generations of clinicians to patterns of psychopathology and psychological distress at variance with the standard descriptions provided by emerging classification systems such as the series of diagnostic and statistical manuals published by the American Psychiatric Association.

This may explain why the clinical psychiatric literature of the 1960s and 1970s is filled with anecdotal descriptions of cross-cultural differences in the onset, pathoplasty, course, and outcome of common psychiatric disorders. Simultaneously, as the network of psychiatric services expanded exponentially following passage of the Federal Government Community Mental Health Act in 1963, differential utilization patterns of disadvantaged minorities also became apparent, along with evidence of misdiagnosis of major psychiatric disorders and poorer than expected response to mainstream treatments by socioeconomically deprived, culturally diverse populations who gained access to rapidly growing community-based systems of mental health care.

All these developments, fueled also by considerable advocacy from culturally informed and socially active clinicians and academicians, culminated in the landmark acknowledgment by the American Psychiatric Association in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) of the importance of culture and ethnicity on diagnosis and treatment of psychiatric disorders.

> Special efforts have been made in the preparation of DSM-IV to incorporate an awareness that the manual is used in culturally diverse populations in the United States and internationally. Clinicians are called on to evaluate individuals from numerous different ethnic groups and cultural backgrounds (including many who are recent immigrants). Diagnostic assessment can be especially challenging when a clinician from one ethnic or cultural group uses the DSM-IV classification to evaluate an individual from a different ethnic or cultural group. A clinician who is unfamiliar with the nuances of an individual's cultural frame of reference may incorrectly judge as psychopathology those normal variations in behavior, belief, or experience that are particular to the individual's culture. For example, certain religious practices or beliefs (e.g., hearing or seeing a deceased relative during bereavement) may be misdiagnosed as manifestations of a Psychotic Disorder. Applying Personality Disorder criteria across cultural settings may be especially difficult because of the wide cultural variation in concepts of self, styles of communication, and coping mechanisms.

A final reason for the burgeoning growth of cultural psychiatry is demographic. The U.S. Census Bureau reported in 1996 that 26.7 percent of the U.S. population comprises minorities belonging to four underrepresented racial groups: African-Americans constitute 12.1 percent; Hispanics, 10.5 percent; Asian and Pacific Islanders, 3.4 percent; and Native Americans, 0.7 percent, distributed in more than 500 tribes. Demographers project that by the year 2050, almost 50 percent of the U.S. population will be minorities: Hispanics are projected to grow to 24.5 percent, African-Americans will represent 13.6 percent, and the final 10 percent will comprise Asians, Native-Americans, and others.

These demographic trends ensure that psychiatric clinicians will increasingly face the challenge of evaluating and treating patients from many different cultural and linguistic backgrounds. Key conceptual and technical breakthroughs in cultural psychiatry will enhance their knowledge and skills and move them closer to the goal of meeting this challenge with reasonable clinical competence.

DEFINITIONS AND KEY CONCEPTS

Culture The term “culture,” a key concept for psychiatry, is plagued by lack of clarity and poor agreement by users. The social science literature in English lists over 160 definitions. Probably it was first used in its contemporary anthropological sense by the British anthropologist Sir Edward Tylor, who wrote in 1871 that culture is “the complex whole which includes knowledge, belief, art, morals, law, custom and any other capabilities and habits acquired by man as a member of society,” emphasizing a holistic conception of culture and its bonding power for society. For its application to psychiatric practice, perhaps the best definition of culture is provided by the National Institute of Mental Health's Culture and Diagnosis Group: “Culture refers to meanings, values and behavioral norms that are learned and transmitted in the dominant society and within its social groups. Culture powerfully influences cognition, feelings, and self-concept as well as the diagnostic process and treatment decisions.”

Culture is thus best conceptualized as a totality, composed of a complex system of symbols possessing subjective dimensions such as values, feelings, and ideals and objective dimensions including beliefs, traditions, and behavioral prescriptions, articulated into laws and rituals. This unique capacity of culture to bind the objective world of perceived reality to the subjective world of the personal and intimate lends it its powerful role as expressor, mediator, and moderator of psychological processes and, ultimately, emotional disorders.

A causal chain allows culturally shaped symbols to influence cognition, feelings, self-concept, and the underlying brain functions that control mood and behavior. Since humans do not exist without a culture, cultural perceptions inevitably affect how each person is viewed and what rights and obligations each individual enjoys in the different stages of the life cycle in a given society. Derived from man's unique ability to use symbols and self-awareness to transcend the instinctual biological determinism ascribed to animals, culture becomes according to C. C. Hughes a “learned configuration of images and other symbolic elements (such as language) widely shared among members of a given society or social group which, for individuals, functions as an orientational framework for behavior, and for the group, serves as the communicational matrix which tends to coordinate and sanction behavior.” As such, culture represents a normative framework that defines normality and

deviance and promotes preferred values and behavior while proscribing others and, by such means, facilitates healthy adaptations or hinders development by provoking maladjustment.

Scope of Culture Though the manifestations of culture are broad enough to be considered almost infinite, the noted American anthropologist George P. Murdock listed 68 features considered to be universally present in the hundreds of societies studied by contemporary anthropologists. In alphabetical order, they are "age-grading, athletic sports, bodily adornment, calendar, cleanliness training, community organization, cooking, cooperative labor, cosmology, courtship, dancing, decorative art, divination, division of labor, dream interpretation, education, eschatology, ethics, ethnobotany, etiquette, faith healing, family feasting, fire-making, folklore, food taboos, funeral rites, games, gestures, gift-giving, government, greetings, hair styles, hospitality, housing, hygiene, incest taboos, inheritance rules, joking, kin groups, kinship nomenclature, language, law, luck, magic, marriage, mealtimes, medicine, obstetrics, penal sanction, personal names, population policy, postnatal care, pregnancy usages, property rights, propitiation of supernatural beings, puberty customs, religious ritual, residence rules, sexual restrictions, soul concepts, status differentiation, superstition, surgery, tool-making, trade, visiting, weather control and weaving."

Obviously some of these dimensions are more central than others to the relation of culture to psychology and psychopathology. Among these are rules related to sexuality and reproduction (incest taboo and rules of marriage), community and social organizations (kinship, kin groups, power relations, and division of labor), and cosmological visions (magic, superstition, and creational myths).

Culture, Race, and Ethnicity If *culture* lacks precise definition, the use (often interchangeably) of additional terms such as *race* and *ethnicity* adds considerable complexity to an already confusing semantic field. Traditionally, *race* denotes human groupings that are biologically (and in theory genetically) determined. In fact other contemporary biologists consider race to be poorly correlated with any measurable biological or cultural phenomenon. Although some characteristics of a given "racial" group (e.g., skin color) may appear phenotypically compelling, the use of the *race* to aggregate individuals displaying that characteristic may convey a false sense of distinctiveness and may imply the existence of a biological basis for such classification systems. Such is not the case for any of the phenotypic characteristics used to establish race. On the other hand *ethnicity*, a term increasingly preferred by cross-cultural researchers, connotes groups of individuals sharing a sense of common identity, a common ancestry, and shared beliefs and history. A given patient's ethnicity can be assessed by focusing the clinical history taking on key ethnically shaped developmental experiences, such as special rituals and rites of passage, and adherence to ethnically prescribed family roles, religious observances, food preferences, and the like. Cultural identity denotes the internalized self-definition resulting from the person's selective, developmentally mediated incorporation of values, beliefs, history, and customs from those available in that person's native environment. Typically, it contains many dimensions of self-experience including age, gender, race, sexual orientation, ethnicity, language, class, and religious and spiritual beliefs.

Culture and Psychopathology Culture is an all-pervasive medium for humans. It is driven by the human brain's unique ability to create images and symbols and structure them into complex wholes that in turn can drive brain function to produce defined behaviors and modulate instinctually driven ones. The ability to mediate biological functions via symbolic (and image) representation and manipulation is dramatically expanded in humans by the function of awareness or consciousness leading to the notion of the self.

Humans structure symbols into progressively more complex sequences ranging from basic, simple ideas to complex holistic beliefs, values, and ideals. These in turn are codified by language, images, and other means and shared with others to constitute the building blocks of common culture. Culture thus becomes a hierarchical array of complex symbols that affect the individual's emotions and behaviors and, when communicated to others, affect social and group function.

All cultures develop both processes that facilitate adjustment and conflict resolution and pressures that foster conflict, deviance, and maladjustment. These pressures can act broadly on large social groups and selectively on specific cultural subgroups. All cultures define a spectrum of "normal behaviors" as well as thresholds of tolerance for diverse "abnormalities," imposing different social consequences on different patterns of deviance.

Each culture provides its own unique stresses as well as beliefs and rituals to reduce psychological tension. Ashley Montagu has indicated, for example, that cultures that provide adaptive channels for the expression of aggression and the satisfaction of dependency needs can significantly reduce personal and interpersonal conflict. In the modern era cultures and subcultures change at increasing rates in response to the adaptive demands represented by a global world in increasing economic, political, and social competition. This progressive globalization imposes additional adaptive stresses on individuals and cultural groups.

The relation between culture and psychopathology has changed significantly over the last 100 years of development of culture psychiatry. At the end of the nineteenth century and the beginning of the twentieth, cultural psychiatry was largely concerned with comparative psychiatry studies and the description (from a Western, universalistic viewpoint) of exotic features of clinical syndromes found in Africa, the Philippines, Java, the Indian subcontinent, and other non-Western locations.

During the mid-twentieth century, culture-personality studies became abundant as anthropologists such as Ruth Benedict, Margaret Mead, Bronislaw Malinowski, and others integrated psychoanalytic constructs into their work, an effort finally leading to mixed intellectual results. During the 1970s, 1980s, and in the present, efforts have focused on several fronts; the development of sound, cross-culturally valid instruments for epidemiological research, such as the Center for Epidemiological Studies Depression Scale, and the Composite International Diagnostic Interview, is among the most fruitful. Significant advances have been made in developing measures for culturally specific idioms of distress. This and other efforts may signal a healthy shift from universalistic to culturally specific perspectives.

Insofar as culture exerts broad influence on social and cognitive, affective, communicative, behavioral, and psychophysiological processes, cultural psychiatry must describe pathways through which culture can affect psychopathology, such as the following:

1. Generating psychopathology by presenting adaptive stresses that tax an individual's psychological life and behavior. Chronic social conditions such as poverty, unemployment, and pervasive racism consistently generate high rates of distress and psychopathology in those exposed to them.
2. Reducing psychopathology by enacting protective factors against external or internal stressors. Javier I. Escobar, among others, has called attention to the low prevalence of most psychiatric disorders among low-acculturation Mexican-Americans and recent Mexican immigrants, which may be due to a protective effect of their traditional culture's emphasis on, and pervasive use of, supportive extended family networks.
3. Promoting or modulating social change, thus increasing or reducing the psychological stresses associated with change.
4. Affecting the onset, pathoplasty, course, and outcome of conventional psychiatric syndromes such as schizophrenia or bipolar disorders.
5. Affecting the tolerance for certain behavioral, subjective, clinical symptoms or unusual mental states and thus promoting either pathogenic isolation or healing integration into the wider social group of the individual displaying such signs and symptoms.
6. Patterning culture-specific idioms of distress, sanctioned ways for the individual or group to express subjective distress.

Acculturation Lloyd H. Rogler defined *acculturation* as "the complex process where by the behaviors and attitudes of a significant group change toward the dominant group as a result of exposure to a cultural system that is significantly different." The impact of the acculturation process on the development of stress and psychological disorder has been studied with mixed results. While some studies support the hypothesis that high levels of acculturation decrease psychological stress, others indicate that some form of biculturalism is more conducive to mental health. Recently, studies of immigrant Mexican-Americans have established a relation between increased acculturation and increased psychological distress. Until more-definitive research establishes an empirical basis, culturally competent clinicians should be attuned to the individual experience of their patients' acculturation process. The stress of the process may be revealed by different means: an intense rejection of the subject's original and host cultures, strong resistance to integration into the host culture, or even an excessive, mechanical assimilation into the new setting, with or without rejection of the culture of origin.

GENERAL PRINCIPLES OF CROSS-CULTURAL PRACTICE

DSM-IV's outline for cultural formulation provides a comprehensive framework for the complete assessment of a patient by clinicians from a different cultural or ethnic group and ends with an overall cultural formulation that like the dynamic and psychodynamic formulation in general clinical psychiatry, summarizes the contributions of culturally derived forces and stressors to the specific clinical situation the patient faces. A culturally competent diagnostic interviewer should be aware of the following points.

Interview Across the Cultural Barrier The golden rule for cross-cultural evaluation remains the same as for any diagnostic process used in clinical psychiatry, namely, establish empathy during the clinical interview. To do so one must be aware of cultural diversity preferred communication styles. First-generation Hispanic-American and Asian-American patients may favor less-expressive styles of communication, particularly when talking to authority figures and about content that carries some stigma. Abundant tact is needed when discussing sexual themes and themes of aggression and suicide, in particular with certain Asian and Hispanic groups, if the clinician wishes to get beyond their initial discomfort and denial. Many traditional cultures expect deference and respect to elders and heads of

households and their members may respond with annoyance and emotional withholding to the more-egalitarian, informal ambiance that predominates the mainstream American medical culture.

Assessment Diagnostic assessment of patients across the language barrier is full of opportunities for misunderstanding, over- and underestimation of psychopathology, and consequent misdiagnosis. Luis R. Marcos and others have found that schizophrenic patients of Hispanic origin exhibited higher rates of psychopathology in interviews conducted in English than in interviews held in their native Spanish.

Beyond the language barrier, specific group stressors suffered by certain minority groups may add additional confusion. Cross-cultural clinicians working with refugees, for example, warn the culturally naive evaluator not to mislabel as Axis I psychopathology the sometimes vivid and highly emotional narratives of Southeast Asian and other refugees describing their witnessing or suffering terrible atrocities in their countries of origin or in the process of immigration. Fortunately clinicians practicing today across language and cultural barriers have access to the findings of numerous studies that provide guidance for accurate evaluation and diagnosis and will allow them to correct many distortions and provide culturally syntonic care.

Language Since about 50 percent of patients from different ethnic groups who present to the U.S. health care system for evaluation and treatment are monolingual (with the rest displaying various degrees of linguistic competence), the culturally competent clinician needs to appreciate the impact of language in the evaluation of psychopathology and the possibility, for a given patient, of using language as a vehicle for therapy. Marcos has established the importance of assessing the patient's language proficiency and language independence and has provided evidence of the potentially adverse effect, in the intensive psychotherapy of proficient bilingual persons, of having two relatively separate language systems with unequal access to profound emotions.

The clinician interviewing a patient using a secondary language should also realize that the secondary language is less efficient in evoking and expressing complex emotions, which may result in a more limited and less accurate psychiatric history. The loss of communicative subtleties may lead the clinician to underestimate the patient's adaptive capacity for high-level cognitive functions such as humor and irony. Use of the secondary language may also reduce the patient's ability to establish rapport and emotional communication with the evaluators.

Cultural Consultants When the language barrier proves insurmountable, the clinican may need to resort to translators or interpreters, optimally interpreters specially trained for psychiatric work. Interpretation overcomes the limitations of simple translation by conveying not only the literal meaning of the sentences uttered by the patient (denotative meaning) but also the rich background of affect and meaning accompanying that expression (connotative meaning).

Many authors have described the pitfalls of evaluating monolingual patients by inadequate translation. When relatives or other undertrained translators are used, opportunities for miscommunication are expanded by the patient-translator transference and the translator-patient countertransference, and the expected confidentiality of the clinical interview is broken. These factors may make it desirable in certain settings to have trained translators become part of the clinical multidisciplinary team. Diagnostic accuracy and completeness and culturally syntonic treatment planning often requires the use of cultural consultants familiar with the patient's cultural background, norms, idioms of distress, explanatory models of illness, and family dynamics.

Mental Status Examination The mental status examination, the key component of a psychiatric diagnostic interview, is subject to many distortions when conducted across a language and cultural barrier. The central process of the mental status examination requires observation and interpretation of the patient's appearance, behavior, language, and mental activity, both spontaneous and as elicited by the clinician's questions. In interpreting the patient's appearance, behavior, language, and thought content, the clinician must guard against what has been called “the category fallacy,” the effort to fit all conditions, whatever their cultural context, into the Western diagnostic framework without regard to validating that diagnostic framework in the local culture. Patient responses to specific mental status items are affected by the patient's culture of origin, educational level, literacy, language proficiency, and level of acculturation. Investigators consider the following items of the mental status examination to be more sensitive to cross-cultural misinterpretation.

Appearance and Behavior The mental status sections of psychiatric case reports are often replete with such expressions as “normal,” “attractive,” and “appropriate” and others that are subject to significant cultural variation and must be carefully evaluated by the clinician, with the patient's own development and culturally determined normative framework as a referent.

Relationship to the Evaluator The assessment of a patient's attitude and relationship to the clinician who is performing a psychiatric evaluation is affected by many psychosocial variables, including whether the interview is voluntary and the relative emergency of the clinical situation. Key factors of such an assessment (e.g., maintenance or avoidance of eye contact, deference, reserve, physical proximity, and physical contact) are subject to cultural prescriptions that the clinician should strive to decode.

Motor Behavior The assessment of motor activity is considered a fundamental part of the mental status assessment. In a prior publication, the author and collaborators noted that patients who communicate in a nondominant language may use extra nonverbal activity to facilitate verbalization across the language barrier. This extra activity needs to be carefully evaluated, lest it be attributed to tension, hyperactivity, or other forms of motor psychopathology.

Speech and Thought Patients communicating in a language over which they have a poor command often exhibit a high frequency of speech disturbances such as omissions, incompleteness, long pauses. These must be carefully distinguished from the impact of anxiety, depression, or emotional withdrawal.

Affect In evaluating the range, responsiveness, and quality of a patient's affect, clinicians working across the language barrier must recognize that both the spontaneous and the elicited expressions of affect are deeply patterned by cultural norms and expectations. Culturally sanctioned impassiveness should not be misinterpreted as poverty of affect any more than the ebullience attributed to persons of Mediterranean origin should be evaluated as excessively intense affect. Clinicians need to be equally sensitive to the significance of linguistic factors that may cloud the interpretation of affect.

When a mental status evaluation is performed across language and cultural barriers, certain linguistic problems may be misinterpreted as surplus psychopathology, and clinicians should guard against this error by repeating critical questions, introducing redundancies to facilitate communication, and identifying paralinguistic cues that may cloud their evaluation of mood and emotional expression, and when in doubt they should use trained translators, cultural consultants, or structured, validated interviews.

Cultural Formulation The cultural formulation is as key an instrument in the diagnosis and treatment of culturally diverse individuals as the psychodynamic formulation is for the psychodynamic psychiatrist. DSM-IV provides an outline for its completion and recommends that it includes the following five categories:

1. Cultural identity of the individual
2. Cultural explanations of the individual's illness
3. Cultural factors related to psychosocial environment and level of functioning
4. Cultural elements of the relationship between the individual and the clinician
5. Overall cultural assessment for diagnosis and care

Cultural Identity DSM-IV recommends that in assessing an individual's cultural identity, the clinician should “note the individual's ethnic or cultural reference group. For immigrants and ethnic minorities, they should assess the degree of involvement with both culture of origin and host culture.” Frances G. Lu and coworkers summarized the essential components of cultural identity (Table 4.4-1).

Ethnicity	Gender
Race	Age
Country of origin	Sexual orientation
Language	Religious and spiritual beliefs
Acculturation	Socioeconomic class and education

From Lu FG, Russell FL, Mezzich JE: Issues in the assessment and diagnosis of culturally diverse individuals. In *Annual Review of Psychiatry*, vol 14, J Oldham, M Riba, editors. American Psychiatric Press, Washington, DC, 1995.

Table 4.4-1 Aspects of Cultural Identity Development

To these factors one must add migration history, which is commonly left out of the clinical evaluation of cross-cultural patients. Culturally uninformed clinicians often treat their immigrant patients as if their lives began when they arrived in the United States, and their clinical narratives often lack key data from the patients' preimmigration experience. Careful attention must be paid to the traumas and losses encountered by refugees in their country of origin, often including exposure (as witness or victims) to physical or emotional torture or both. The process of acculturation is once again key to understanding the psychological distress and psychopathology of immigrants. Rogler has identified three major sources of stress in the migration experience: (1) insertion into the host society, frequently at lower occupational and social levels; (2) disruption of primary interpersonal networks; and (3) the stress-inducing acculturation process. The clinician can assess the degree of acculturation and the nature of the acculturation process through many indirect means. Age at immigration, number of years in the United States, occupational status, language proficiency, and participation in the host culture social networks give the clinician some idea of the rate and ease of acculturation for a given patient.

Families can also be classified by degree of acculturation. From this perspective immigrant families may be described along a continuum of acculturation as traditional, transitional bicultural, and Americanized. Each of these family structures presents different assets and vulnerabilities in relation to the immigration experience.

THERAPIES

George L. Engel provided psychiatry with a model for understanding the etiology of psychiatric disorders, the biopsychosocial model. This model emphasizes that understanding and treating psychiatric disorders requires attending to the three levels of human function, biological, psychological, and social, as well as to their systematic interrelations. In a rapidly growing global culture the social component of the biopsychosocial model must include, develop, and elaborate the cultural dimension, including the patient's religious affiliation. Culture is important to psychiatric therapy as it influences, mediates, shapes, or colors beliefs about the cause of psychological distress and preferred modes of intervention. This broad influence of culture applies to somatic interventions such as psychopharmacology as much as to psychological therapies.

The issue of what therapies to apply to patients from different ethnic or cultural groups involves three key questions. (1) can psychoanalytically based and other expressive psychotherapies be applied to patients from all cultures with comparable results? (2) Can culturally specific therapies be devised either by significant modification of mainstream psychotherapy or by crafting therapeutic interventions from the rich lode of culturally syntonic rituals and symbols? (3) Are there psychological therapies that can be specifically focused on the problems and stresses of acculturation?

Critics have argued from several perspectives that psychodynamic psychotherapy is inappropriate and ineffective with nonwhite clients. Empirically, high dropout rates and less than optimal outcomes may have helped buttress that argument. Some medical anthropologists maintain that psychodynamic psychotherapy is itself a culturally syntonic product of middle-class Western minds and as such is inappropriate for patients belonging to different cultural traditions and may even damage the patients' culture-bound view of self and their self-esteem.

Many clinicians, especially those grounded in the experience of providing psychotherapeutic services to socioeconomically deprived ethnic minorities, maintain a more balanced view, noting that psychotherapeutic success is possible and appears to be related to levels of education, acculturation, motivation, and capacity to form a therapeutic alliance. Harvey Bluestone made important recommendations for adjustments in providing psychodynamically based, insight-oriented therapy to poor Puerto Ricans in New York City. In his view, the framework for therapy must be strengthened by explicitly educating patients to keep appointments on time, to change expectations for rapid cures, and to enhance their own active participation in treatment, which challenges the culturally patterned, passive dependence on an authoritative therapist when this pattern exists.

Culturally Specific Therapies The clinical and research literature offers many examples of incorporating specific elements from the patient's native culture into the therapeutic intervention. J. J. Kreisman's account of treating two Mexican-American females with schizophrenia by incorporating folk rituals and native herbs is rather representative. Kreisman advocates integrating the patient's own cultural conceptions of illness into the conventional treatment. Jose Szapocznick has proposed a model of family therapy for Miami's Cuban families, guided by the strong familial affiliation of Cuban families, their preference for hierarchical family structures, and other empirically determined values. His model is especially attractive since it readily lends itself to empirical testing.

Original, culturally sensitive therapeutic modalities have been developed and empirically tested for the treatment of the adjustment problems of vulnerable minority populations. Giusepe Constantino and coworkers demonstrated the superiority of *cuento*, or folktale therapy, over traditional therapy group and nontherapeutic group sessions, in reducing trait anxiety in a sample of over 200 high-risk Puerto Rican children selected for maladaptive school behavior. Repeated exposure to classical Puerto Rican folktales seemed to have provided these vulnerable children with stable, culturally syntonic role models for adaptive internalization of coping skills and positive self-images.

Psychopharmacology The relation between culture and psychopharmacology is as fascinating as it is complex. A web of incompletely understood relations links purely biological factors such as genes to factors deemed more socially generated, such as customary diets, nutrients, and herbs. To complete the picture, the clinician and researcher must consider additional factors such as the patient's culturally based expectations of optimum psychiatric treatment (pharmacotherapy or psychotherapy), expected rate of recovery (fast or slow), target symptoms, and threshold and tolerance for adverse effects. If a Hispanic patient has the culturally shared expectation that psychological disorders are somatically based and best treated by an authoritative physician with medication, alternative treatment recommendations will be met with resistance and reduced compliance, unless extensive psychoeducational efforts are deployed. Alternatively, middle-class, educated, urban professionals may reject psychopharmacological prescriptions for their anxiety or depressive syndromes as simplistic, because they had expected psychotherapy or psychoanalysis. Thus the therapeutic relationship across the language/cultural barrier should be initiated by carefully eliciting the patient's explanatory framework of illness, anticipated path to recovery, and expectations for treatment. Clinicians must spend time and effort in an educational dialogue with patients and their significant others and explain the reasons for an alternative to the patient's preferred course of treatment. Slow onset of action and the frequency of adverse effects interfere with the therapeutic cooperation of some Hispanic and Asian patients who expect rapid relief, fear toxicity, or are concerned with the addictive potential of medications to be taken long term. To obtain adequate compliance across the cultural barrier the clinician must make tactful efforts to know the patient's (and the immediate family's) latent, culturally shaped beliefs about the illness and its normative treatment, to provide therapeutic options compatible with the patient's culturally prescribed explanatory models, lest hidden miscommunication hinder the necessary compliance.

Pharmacogenetics The field of pharmacogenetics grew out of observations of significant ethnic differences in response to drugs, in differential development, and in side effects profiles, leading to the discovery of defects or deficiencies in the genetically controlled activity of enzyme systems responsible for the metabolism of psychotropic medications and toxins such as alcohol.

Acetylation Status Observations of ethnic differences in the side-effects profile of the antituberculosis drug isoniazid (Nidrazid, Rifamate) led to the classification of people as slow or rapid acetylators, which, among other biological effects, determines their metabolism of psychotropic medications such as clonazepam (Klonopin) and phenelzine (Nardil).

Alcohol Metabolism P. H. Wolf, while studying racial differences in alcohol sensitivity, observed that about 80 percent of Asians and 50 percent of native Americans

exhibited the flushing response to alcohol (compared with 10 percent of whites) and concluded that these differences had a genetic basis. They have been proved to be related to genetic polymorphism of isoenzymes of alcohol dehydrogenase (ADH) and aldehyde dehydrogenase (ALDH), enzymes critical for complete metabolism of alcohol and other neurotransmitters and which play a role in development of alcoholism or its avoidance. For example, Asians who are either homozygous and heterozygous for the atypical Asian-type *ALDH*2 gene are alcohol sensitive and have a low risk for alcoholism and alcohol liver disease.

Native Americans have a high frequency of both alcohol flushing and alcohol-related problems. Akira Yoshida's research team reported in 1993 that they had practically no detectable Asian type *ADH*2 and *ALDH*2 genes, a major alcohol-rejecting genetic factor.

Cytochrome P450 Isoenzymes The cytochrome P450 enzyme system is key in the metabolism of psychotropic and nonpsychotropic drugs as well as a great variety of environmental toxins that find their way into the diets of animals and humans. The genetic defects that render these enzymes less effective and make humans poor metabolizers are unequally distributed among ethnic populations. This is particularly the case for two cytochrome P450 (CYP) isoenzymes: CYP 2D6 (debrisoquin hydroxylase), and CYP 2Cmp (mephenytoin hydroxylase). The percentage of CYP 2D6 poor metabolizers is lower for Asians (0.5 to 2.4 percent), and higher for whites (2.9 to 10 percent). Similar interethnic variance exists in the frequency of poor metabolizers of CYP 2Cmp, low among whites (3 percent), intermediate for African-Americans (18 percent), and higher (up to 20 percent) in Asian and Japanese populations.

These interethnic differences in the P450 isoenzymes are of great importance in psychiatry and psychopharmacology because of their role in the metabolism of antipsychotics, antidepressants, sedatives such as barbiturates and benzodiazepines, and b-adrenergic receptor antagonists (beta-blockers) such as propranolol (Inderal).

Environmental Factors In addition to being genetically regulated, enzymes that participate in the metabolism of psychotropic medications respond to environmental variables such as diet, alcohol, smoking status, and caffeine intake. All these factors are capable of accelerating or slowing the metabolism of drugs through enzyme induction or inhibition.

Herbal Medicines In parallel with available Western medicine–oriented psychiatric services, immigrants often retain their loyalty to ethnically based folk medicine systems. Accounts by Vivian Garrison and Allan Hardwood document extensive use of folk healers by Puerto Ricans in New York City. Other investigators have reported that Mexican-Americans are willing to accept prescribed medications from psychiatrists and herbs from a community healer, just as mainstream, young, urban professionals use natural serotonin-enhancing herbs such as St. John's wort in addition to, or instead of, the more conventional psychotropics prescribed by their psychiatrists. Culturally competent psychopharmacologists need to inquire about their patients' use of the traditional herbal medicines of Asian, African-American, Hispanics, and other immigrants living in the United States. Many of these herbs posseses high levels of psychoactive activity, such as anticholinergics (*Swertia japonica* used by Japanese patients or *Datura candida* used by Cubans), stimulants (the caffeine-loaded *Ibexguazusa* of Latin America), sedatives (*Schumanniophyton problematicans* of the Nigerians). Others, such as ginseng and glycyrrhiza, may stimulate cytochrome P450 or inhibit it.

Specific Psychoactive Medications

Antipsychotics Clinical lore has it that Asian patients require lower dosages of dopamine receptor antagonist (typical antipsychotic) medications than comparable white patients to achieve a similar clinical outcome and that when treated on a fixed-dosage schedule, Asians seem to develop significantly greater extrapyramidal adverse effects. These reports were corroborated by S. Potkin in a 1983 study that found a 52 percent higher plasma concentration of haloperidol (Haldol) in Chinese schizophrenia patients living in China than in non-Asian schizophrenia patients residing in the United States when both groups received treatment on a fixed-dosage schedule. Other investigators demonstrated that Chinese schizophrenia patients residing in Taiwan and Taipei achieve haloperidol plasma concentrations comparable to those of white, African-American, and Hispanic patients hospitalized in San Antonio while using significantly lower daily dosages of haloperidol. From their data, the authors suggest that dosages above 30 mg a day might not be necessary in Chinese patients to obtain the 20 to 25 mg/mL currently considered the upper therapeutic plateau for haloperidol plasma concentration. Dosages above 50 to 60 mg a day might not be needed in non-Chinese patients.

Survey reports indicate that African-Americans and blacks elsewhere receive significantly higher doses of antipsychotics and receive antipsychotic drug treatments significantly more often than whites, regardless of diagnosis. This fact may be explained by diagnostic and prescription biases that ascribe more psychopathology to blacks, or it may have a pharmacokinetic explanation. One study found African-American patients to have the highest concentration of plasma-reduced haloperidol for a given plasma concentration of haloperidol. Reduced haloperidol lacks antipsychotic activity and has lower binding affinity than haloperidol for the dopamine type 2 (D_2) receptor.

Tricyclic Drugs Once the cornerstone of treatment of depressive disorders, use of tricyclic drugs has been expanded considerably beyond these boundaries to include anxiety, panic disorder, obsessive-compulsive disorder, other behavior disorders, and various pain syndromes. Their cross-cultural utility has been extensively studied. As is the case with neuroleptics, studies with tricyclic drugs have shown that average dosages prescribed for Asians are significantly lower (up to 50 percent lower) than dosages prescribed in the United States for non-Asians, results that cannot be explained by ethnic differences in protein binding alone.

Ethnic psychopharmacologists have proposed that Chinese patients may respond to lower tricyclic drug dosages. Although the reason for this responsiveness has not been clearly established, preliminary evidence suggests differential responsiveness of relevant receptors, differences in plasma concentrations, or both.

Although less extensively studied, similar conclusions (equal therapeutic response to half the standard white dose) have been reached for depressed Hispanic women and African-American patients, who are reported to respond faster and more completely, albeit with higher rates of adverse effects. Sensitivity to development of anticholinergic adverse effects is important to both compliance and morbidity in accidental or intentional overdoses. Researchers have reported that delirium as a complication of tricyclic drugs appears more frequently among African-Americans than among other ethnic groups.

Benzodiazepines Studies of prescription patterns as well as those comparing the pharmacokinetics and pharmacodynamics of benzodiazepines across ethnic groups have established the enhanced sensitivity of Asians to the effects of benzodiazepines, accounting for the typically prescribed doses one half to two thirds those of similar nonminority populations. These ethnic differences in benzodiazepine metabolism are most often linked to polymorphisms in the *(S)*-mephenytoin phenotype, yielding a higher percentage of poor metabolizers in the Chinese ethnic group. Additional mechanisms, both genetic and environmental, have been invoked by other investigators.

Few studies are available to guide clinical practice with benzodiazepines in African-Americans, Hispanics, and Native Americans.

Lithium Lithium salts (Eskalith, Cibalith) are extensively used in clinical psychiatry since their clinical efficacy for the treatment of mania was first established by John Cade in 1949. Besides its core indication for the treatment of bipolar disorder, lithium is currently used for the treatment of other psychiatric disorders (e.g., intermittent explosive disorder), to treat symptoms (e.g., impulsivity of any cause), and to augment conventional strategies in the treatment of schizoaffective disorders and refractory depression.

Many studies have documented lithium's differential pharmacokinetics and pharmacodynamics in different ethnic populations. For African-Americans, in 1980 S. Okpaku and colleagues demonstrated a slow lithium efflux in red blood cells (RBCs) following laboratory loading of RBCs, as well as a higher RBC to plasma ratio, a finding later confirmed by many other investigators. Since RBC lithium concentrations are believed to be correlated with intraneuronal lithium concentration and with the incidence of adverse effects, African-Americans may need lower dosages of lithium than those currently recommended.

As with trycyclic drugs, antipsychotics, and benzodiazepines, Asians seem to achieve clinical responses comparable to those of non-Asian patients using significantly lower dosages and serum concentrations of lithium; 0.5 to 0.7 mEq/L versus the 0.8 to 1.2 mEq/L generally required by white populations.

In addition to these kinetic and dynamic factors, clinicians working across the cultural barrier need to carefully avoid the problem of misdiagnosis. Abundant evidence indicates that minority patients with bipolar disorders are at greater risk for receiving a diagnosis of schizophrenia at some point in their clinical course, and receive antipsychotic medication rather than mood stabilizers for their treatment. To avoid potential misdiagnosis, clinicians need to be especially alert to minority patients diagnosed with schizophrenia who have an atypical course, poorer than expected antipsychotic response, prominent affective symptoms, or a cyclically fluctuating course. The use of structured interview instruments may be necessary to refine the diagnosis.

CULTURE-BOUND SYNDROMES

The culture-bound syndromes have not received much attention by mainstream psychiatry in the United States. When mentioned at all, these syndromes are often discussed as exotic disturbances of thought, mood, or behavior displaying dramatic presentation, occurring in the context of specific local cultures, and at least partially understood through the lens of the psychosocial forces—often conflictual—relevant to that particular culture. For example, *taijin kyofusho*, a syndrome of some prevalence in Japan, whose sufferers have an intense concern that their body may be offensive to others because of appearance, body odor, or any other characteristic, can be best understood against the background of intense Japanese emphasis on proper social behavior and the shame attending transgressions of decorum.

The interpretation of culture-bound syndrome must be tempered by the following observations:

1. Psychiatry's neurobiological knowledge base has not developed firm predictive biological indicators, even for the major psychiatric disorders such as schizophrenic and bipolar disorder.
2. Definitions of syndromes are thus by necessity phenomenologically based and have imprecise boundaries.
3. In defining clinical syndromes, psychiatry depends on subjects who present for medical attention, a process influenced by many psychosocial and socioeconomic variables. Of all people suffering a psychological distress or disorder, only the minority who seek and find access to the health care system are represented in the official definition of the disorders.

Even in the most advanced societies, a considerable lag exists between the emergence of new forms of psychological distress or disorder and the health care system's ability to recognize, properly label, diagnose, and effectively treat it. During this time, there is frequently intense popular interest in the evolving disorder. It took time and considerable popular pressure before psychiatry was ready to respond to disorders such as the bulimia nervosa plaguing young females or the posttraumatic stress disorder of returning Vietnam veterans.

SUGGESTED CROSS-REFERENCES

Anthropology and psychiatry are discussed in Section 4.1, and sociology and psychiatry in Section 4.2. Examination of psychiatric patients is covered in Chapter 7. Psychotherapies are presented in Chapter 30 and psychopharmacological therapies in Chapter 31.

SECTION REFERENCES

Alarcón R, editor: Cultural psychiatry. Psychiatr Clin North Am *18:*3, 1995.

*Baker FM, Bell CC: Issues in the psychiatric treatment of African Americans. Psychiatr Serv *50:*362, 1999.

Bleuler E: *Dementia Praecox or the Group of Schizophrenics*. International University Press, New York, 1950.

Bluestone H, Vela RM: Transcultural aspects in the psychotherapy of the Puerto Rican poor in New York City. J Am Acad Psychoanal *10:*269, 1982.

Constantino G, Malgady RG, Rogler LH: Cuento therapy: A culturally sensitive modality for Puerto Rican children. J Consult Clin Psychol *54:*639, 1986.

Engel GL: The need for a new medical model: A challenge for biomedicine. Science *196:*127, 1977.

Escobar JI: Immigration and mental health: Why are immigrants better off. Arch Gen Psychiatry *55:*781, 1998.

*Gaw A, editor: *Culture, Ethnicity, and Mental Health*. American Psychiatric Press, Washington, DC, 1993.

Good BJ: *Medicine, Rationality and Experience: An Anthropologic Perspective*. Cambridge University Press, Cambridge, 1994.

Hughes CC, editor: *Custom-made: Introductory Readings for Cultural Anthropology*. Rand McNally, Chicago, 1976.

Johnson-Powell G, Yamamoto J, editors: *Transcultural Child Development: Psychological Assessment and Treatment*. Wiley, New York, 1997.

Kalow W: Pharmacogenetics past and future. Life Sci *47:*1385, 1990.

Kleinman A: *Social Origins of Distress and Disease*. Yale University Press, New Haven, 1986.

*Kleinman A: *Rethinking Psychiatry. From Cultural Category to Personal Experience*. Free Press, New York, 1988.

Kraepelin E: *Psychiatric, I*, ed 8. Barth, Leizpig, 1909.

Kreisman JJ: The curanderó's apprentice: A therapeutic integration of folk and medicinal healing. Am J Psychiatry *132:*81, 1975.

Lu FG, Russell FL, Mezzich JE: Issues in the assessment and diagnosis of culturally diverse individuals. In *Annual Review of Psychiatry*, vol 14, J Oldham, M Riba, editors. American Psychiatric Press, Washington, DC, 1995.

McGoldrick M, Giordano J, Pearce JK, editors: *Ethnicity and Family Therapy*, ed 2. Guilford, New York, 1996.

Marcos LR: Linguistic dimensions in the bilingual patient. Am J Psychoanal *36:*347, 1976.

Marcos LR, Alpert M, Urcuyo L: The effect of interview language on the evaluation of psychopathology in Spanish-American schizophrenic patients. Am J Psychiatry *130:*549, 1973.

Marcos LR, Cancro R: Pharmacotherapy of Hispanic depressed patients: Clinical observations. Am J Psychother *36:*505, 1982.

*Marcos LR, Trujillo M: Culture, language and communicative behavior: The psychiatric examination of Spanish-Americans. In *Latino Language and Communicative Behavior*, RP Duran, editor. Ablex Publishing, Norwood, NJ, 1981.

Mezzich JE, Kleinman A, Fábrega H, Parron DL, editors: *Culture and Psychiatric Diagnosis: A DSM-IV Perspective*. American Psychiatric Press, Washington, DC, 1996.

Montagu A: Culture and mental illness. Am J Psychiatry *118:*15, 1961.

Murdock GP: The common denominator of cultures. In *The Science of Man in the World Crisis*, R Linton, editor. Columbia University Press, New York, 1945.

Okpaku SO, editor: *Clinical Methods in Transcultural Psychiatry*. American Psychiatric Press, Washington, DC, 1998.

Okpaku S, Frazier A, Mendel J: A pilot study of racial differences in erythrocyte lithium transport. Am J Psychiatry *137:*120, 1980.

Potkin SG, Sheu Y, Pardes H: Haloperidol concentrations elevated in Chinese patients. Psychiatry Res *12:*167, 1984.

Rogler LH: The meaning of culturally sensitive research in mental health. Am J Psychiatry *146:*296, 1989.

*Rogler LH, Gurak DT, Cooney RS: The migration experience and mental health: Formulation relevant to Hispanics and other immigrants. In *Health and Behavior: Research Agenda for Hispanics*, M Gaviria, JD Arana, editors. Simon Bolivar Research Agenda for Hispanics I, University of Illinois, Chicago, 1987.

Ruiz P: Clinical care update: The minority patient. Community Ment Health J *21:*208, 1985.

*Ruiz P, editor: Cross-cultural psychiatry. In *Annual Review of Psychiatry*, vol 14, J Oldham, M Riba, editors. American Psychiatric Press, Washington, DC, 1995.

Szapocznick J, Scopetta MA, King OE: Theory and practice in matching treatment to the special characteristics and problems of Cuban immigrants. J Community Psychol *6:*112, 1978.

Tylor EB: *Primitive Culture*. Murray, London, 1871.

Vega WA, Kolody B, Aguilar-Gaxiola S, Alderete E, Catalano R, Caraveo-Anduaga J: Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. Arch Gen Psychiatry *55:*771, 1998.

Weber WW: *The Acetylator Genes and Drug Responses*. Oxford University Press, New York, 1987.

Yoshida A: Genetic polymorphisms of alcohol-metabolizing enzymes related to alcohol sensitivity and alcoholic diseases. In *Psychopharmacology and Psychobiology of Ethnicity*, KM Lin, RE Roland, G Nakasaki, editors. Progress in Psychiatry no. 35. American Psychiatric Press, Washington, DC, 1983.

CHAPTER 5. QUANTITATIVE AND EXPERIMENTAL METHODS IN PSYCHIATRY

5.1 EPIDEMIOLOGY

DARREL A. REGIER, M.D., M.P.H., AND JACK D. BURKE, JR., M.D., M.P.H.

Psychiatric epidemiology has contributed to a broad range of clinical, research, and health policy fields. Recent findings provide data on the course and co-occurrence of psychiatric disorders, identify their possible risk factors, measure the functional impairment they cause, establish a basis for policy decisions on mental health, and set a starting point for analyzing access to care and use of mental health services.

Results from epidemiological studies are routinely included in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) to describe the frequency and correlates of mental disorders. In developing the diagnostic criteria themselves, secondary analyses of many epidemiological studies assessed the frequency with which discrete symptoms appeared together, to define syndromes in large community and clinical populations. Epidemiological studies that demonstrate the significance of depression as a risk factor for death in people with cardiovascular disease and for premature death in people with breast cancer have engendered new interest in pathophysiological mechanisms that might account for the relation between these disorders.

One of the most visible indicators of the policy use of epidemiological data is found in the joint publication by the World Bank and the World Health Organization (WHO), The Global Burden of Disease. Based on data developed over the past two decades, mental disorders have been found to be among the leading causes of disability worldwide, with major depression currently the fourth leading cause of disability.

But one principal goal of epidemiology, finding ways to lessen the occurrence of new illnesses, is still in an early stage of investigation. Although some types of dementia and mental retardation caused by viral exposure (e.g., measles), toxins (e.g., lead), or metabolic or nutritional deficiencies (e.g., cretinism, pellagra) can be prevented, wide-scale prevention of other mental disorders will depend on identifying effective interventions and on closer ties to findings in basic and clinical research on mental disorders. At present the most effective methods for reducing prevalence rates of acute disorders involve early detection and treatment to reduce duration of episodes and improved maintenance treatment of chronic disorders to prevent relapse or recurrence.

Definition Mental disorder epidemiology is the quantitative study of the distribution and causes of mental disorder in human populations. This definition has several important components, including a population focus and a reliance on statistical methods to assess significant differences among population groups in their risk for developing mental disorders.

Epidemiological research focuses on population groups rather than individuals. Although individuals within population groups must be accurately diagnosed, the unique scientific questions addressed by epidemiology begin with descriptions of how mental disorders are distributed across different population subgroups. By providing a scientific method for assessing the relative health status of communities and different population groups, epidemiology serves as the basic science of public health. An observation that disease rates in one population group significantly exceed expected or average rates is the basis for identifying an epidemic within the population. Such assessments require accurate diagnosis of all cases of illness within the population and the use of statistical methods to assess differences in the frequency of disorders in population groups studied.

Identifying different rates of illness in population groups is only the first step in epidemiological investigations. Both clinicians and epidemiologists rely on observations that pathological states are differentially associated with physical, biological, social, and temporal characteristics of human beings and their environments. Where causal mechanisms are not yet clearly defined, epidemiologists attempt to identify characteristic factors that define populations with excessively high rates of illness, including mental disorders. By narrowing the range of characteristics associated with a disorder, the ultimate aim is to identify specific correlates or risk factors whose alteration will interrupt the causal network that produces a disorder.

Special Methodological Issues Advances in any field of epidemiology depend on developing methods to measure, detect, and characterize risk factors associated with specific illnesses. Such methodological issues affecting research on mental disorders include the following.

Nosology Grouping morbidity states for quantitative analysis requires explicit classification of disorders reliably applied across large populations. If mixed conditions are included in one diagnostic category, the statistical association of risk factors associated with one of the component disorders will be diluted and, thus, fail to be recognized by the epidemiological method. Psychiatric nosology has been limited by the heavy reliance on manifestational criteria (e.g., signs, symptoms, clinical course, and treatment response) rather than causal criteria (e.g., toxins, trauma, genetic vulnerability, and metabolic defects). Dependence on manifestational criteria in the absence of convincing causal factors increases the likelihood that heterogeneous groups will be combined in one diagnosis. This problem is shared by the rest of clinical medicine, which attempts to move beyond a diagnosis based on symptoms such as fever to a specific etiological agent or anatomical defect. The relatively new descriptive approach to psychiatric nosology is an important intermediate stage that will facilitate more-rigorous investigations of causal factors in clinical and epidemiological studies.

Case Identification Methods One difficulty resulting from this limitation in nosology involves developing a technique to obtain standardized psychiatric assessments of a large sample of individuals to detect clinically meaningful mental disorders. Without clearly defined nosological categories, it is difficult both to define a case and to develop case identification (diagnostic) techniques that are appropriate for large-scale population studies. Such techniques require explicit criteria if clear communication to other researchers and if reliability in administration is to be feasible. In the absence of reliably administered assessment methods, epidemiologists have often used self-report questionnaires that assess the dimensions of psychological and psychophysiological symptoms. Cutoff scores on various symptom scales have been calibrated with symptom profiles of patients under treatment. Thus, subjects in the general population could be assessed in terms of their probability of having disorders similar to those seen in clinical settings. The lack of face validity for many such questionnaires, the lack of diagnostic specificity, and the tautological assumption that mental disorders under treatment should define the full spectrum of all types of mental disorder led to clinician apathy about earlier epidemiological survey results. However, development of standardized clinical interviews for clinicians and highly structured diagnostic assessment instruments for trained survey interviewers has enabled many advances in psychiatric epidemiology.

Risk Factors At a more advanced level of investigation, another difficulty involves determining the psychosocial and biological factors likely to identify high-risk groups for particular psychiatric disorders; once identified, these factors must also be measured objectively as part of an epidemiological study. The relative paucity of clearly defined and modifiable risk factors available for study has led to overdependence on associations with descriptive and relatively unmodifiable factors such as age, gender, and ethnic status. Although presumably modifiable factors such as stressful life events have recently emerged and there is increasing evidence of molecular genetic linkage with some disorders, the mechanism of multifactorial interaction between individual susceptibility and stress or other environmental variables has not yet been fully explicated.

Chronic Disease Epidemiology Paradigm Other difficulties in psychiatric epidemiology are shared with the epidemiology of chronic medical diseases. These characteristics include a dependence on manifestational criteria and incomplete information on the subsequent temporal aspects of clinical course, uncertainty about when causative factors may have affected the individual, an undefined lag time between the onset of a pathogenic influence and the occurrence of the disorder, and the absence of any generally accepted concept of when the disorder may be considered to be eradicated in a patient with a history of the illness. Although these characteristics do not apply to all mental disorders, they are shared by many within the fields of cancer, cardiovascular disease, and rheumatological disease

epidemiology.

USES OF EPIDEMIOLOGY

The various applications of epidemiology depend on three levels of investigation that can be grouped according to their basic intent as follows:

1. *Descriptive*—studies that produce basic estimates of the rates of disorder in a general population and its subgroups
2. *Analytic*—studies that explore the basis of variations in illness rates among different groups, to identify risk factors that may contribute to development of a disorder
3. *Experimental*—studies that test the presumed association between a risk factor and a disorder and seek to reduce the occurrence of illness by controlling the risk factor

Each level of investigation yields information that can be used to improve clinical practice and plan public health policies. Addressing past difficulties with nosology, case-identification techniques, risk-factor specification, and chronic-disease paradigms has led to better action of the potential applications of epidemiology that were identified by J.N. Morris. Relating these seven basic uses to the three stages of epidemiological inquiry demonstrates the important applications of epidemiological knowledge, even in the earliest studies at the descriptive level of investigation.

DESCRIPTIVE LEVEL

Community Diagnosis Epidemiological research starts by estimating rates of illness in a defined population. This community diagnosis provides a baseline for understanding the burden of illness in the population, the mix of disorders present, and the extent to which untreated cases exist in the population. These basic rates are needed before more elaborate studies of risk factors can be undertaken, and they are important for health planners who want to know what kind of treatment services may be needed in the community, especially if untreated patients are to be brought into the health care system in the future.

Over the past 20 years, two national prevalence studies of mental disorders in the United States greatly expanded our understanding of the burden that mental disorders place on the population. The Epidemiologic Catchment Area (ECA) study in the early 1980s was a multisite, 1-year, prospective study (three interviews at 6-month intervals) of mental disorder prevalence, incidence, and mental health service use among over 20,000 adults, age 18 years and older in communities and in nursing homes, prisons, and long-stay psychiatric hospitals. It provided the first estimates of mental disorder and service use rates according to the type of explicit diagnostic criteria that were pioneered by the third edition of DSM (DSM-III) and continued in subsequent editions and in the 10th edition of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). This study was followed by a cross-sectional (single interview) National Comorbidity Survey (NCS) of over 8000 adolescents and young adults (age 15 to 54 years) in the early 1990s, which was designed to update information on prevalence rates of disorders in the revised third edition of DSM (DSM-III-R) disorders and to clarify the sequence and duration of co-occurring mental and addictive disorders in a younger population group, in which comorbid disorders were more prevalent.

Although the study designs, diagnostic criteria, diagnostic instruments, and population age groups differed somewhat in the ECA and NCS studies, the 1-year community prevalence rates for any mental and addictive disorder were remarkably similar. In the ECA this 1-year prevalence rate for any mental or addictive disorder was 28.1 percent of the population; the NCS rate was 29.5 percent.

Completing the Clinical Picture More-precise answers to such questions about community diagnosis and basic rates of illness can be obtained only with a standardized assessment procedure that can ascertain the diagnostic status of a large number of individuals. If this case ascertainment can be done well, then epidemiological studies will benefit clinical understanding and public health. The full clinical picture of the disorder can be studied by characterizing subclinical or mild cases, by studying relatives to determine the familial occurrence of the condition, and by following the further progression of the illness to determine its course and prognosis. The value of this effort rests on the fundamental rule of epidemiology, which is to sample individuals from a geographically defined population without the bias that may occur with a sample drawn only from patients seen in particular treatment settings. For example, with the new structured interviews that permit classification of anxiety disorders such as posttraumatic stress disorder by both DSM-IV, which requires a numbing criterion and ICD-10, which has no similar requirement, it will be important to determine the associated features for each definition, along with the prevalence, the prognosis, and the course of the illness in longitudinal studies.

Identification of Syndromes Similarly, some conditions may occur in a population sample that were not detected in studies with samples drawn only from clinical populations. The opportunity to identify new syndromes in the population is provided by both representative sampling from the population and the application of thorough, standardized assessment procedures for all individuals in a sample. One example is agoraphobia without prior panic disorder, which appears to occur more commonly in the general population than in clinical samples.

ANALYTIC LEVEL

Assessing Individual Risks Once the basic rates of illness are established, one can identify groups in the population with unusually high rates of illness. This comparative analysis provides a variety of hypotheses for testing to see if some characteristics of the more commonly affected group can be linked to a causal chain for the illness. For example in the 1930s psychiatric epidemiology found apparent higher rates of schizophrenia among low-income, inner-city residents. The first problem in assessing such a finding is to determine whether it reflects a potentially higher risk of developing schizophrenia among those who live in such conditions, or whether those who have schizophrenia move into such areas through downward social mobility. Recent studies by Bruce Dohrenwend and colleagues indicate that downward mobility, or social selection, is the most likely explanation for the higher rates of schizophrenia in those with lower socioeconomic status. In contrast to schizophrenia studies, studies of depression among women and of substance use and antisocial personality disorders among men have shown stronger support for social causation hypotheses.

Even if the place of residence seems to precede rather than follow the disorder, investigators must still pursue the study of risk factors that may underlie the apparent association between this group characteristic and the disorder. Some particular features of poverty, poor housing, or population density may be risk factors; genetic, nutritional, or other familial risks may be higher in persons who live in such areas; or toxic or infectious agents may be the true contributors and occur more commonly in such areas. Moving from demonstrating higher rates in a particular group to more-targeted studies of putative risk factors for individuals is an essential but difficult step in the progression of epidemiological investigations.

Historical Study Risk factor identification may be easier if some historical variation can be shown. This type of evidence may help strengthen the arguments for studying one or another putative risk factor. For example, a 25-year study of one community population in Sweden suggests that depression may be increasing and dementia may be decreasing in an industrialized Western society. Other epidemiological studies have shown that depression and drug abuse or dependence may be occurring at higher rates and with earlier onset in more recent generations.

EXPERIMENTAL LEVEL

Identifying Causes As risk factors are demonstrated, epidemiologists can help reduce the contributing causes of the disorder by intervening in the causal chain that links a risk factor to occurrence of the disorder. Studies that modify a risk factor and assess the impact of this intervention in reducing onset of illness are the long-term goal of epidemiologists. This type of investigation promises to elucidate opportunities for primary prevention of mental disorders by intervening to reduce the chances that high-risk individuals will develop an illness. Quantitative, population-based epidemiological investigation can be used to measure the impact of preventive interventions being tested. Such assessment is important both to determine the benefits of the intervention for high-risk individuals and to assess any unintended negative effects of the program.

Working of Health Services The epidemiological method is usually taken as a guide to primary prevention of disorders, but it can also be used for secondary prevention, early diagnosis, and prompt effective treatment of individuals who have already developed signs of the illness. Although the contribution of epidemiology to understanding the working of health services is often thought to refer to the descriptive level of study—with estimates of treatments provided and unmet need for service—a wider perspective demonstrates that the three-stage epidemiological method can be applied to both patient populations in clinical settings and general populations in the community.

With a parallel to primary prevention based on assessment and reduction of risk factors, a scheme for secondary prevention can also be devised. Once the prevalence of clinical disorders is determined (e.g., as with depression among primary care patients), one can assess the extent of early diagnosis and prompt, effective treatment. Although depressive disorders are common in primary care patients, recent studies in this country and abroad suggest that these disorders are not often diagnosed or treated. Interventions to improve clinical practice through secondary prevention can be tested, just as primary prevention intervention efforts would be tested in the traditional epidemiological model. Research of this type, with application of the three stages of epidemiological research to clinical populations, is known as clinical services research.

EPIDEMIOLOGICAL METHODS

Understanding the conceptual framework for applying epidemiologic methods helps one understand the findings and appreciate the potential benefits of epidemiologic research. Epidemiology has several important characteristics, including a population-centered research strategy that requires an adequate sampling plan. Especially in psychiatry, epidemiological research also required advances in the nosology of mental disorders before acceptable case-identification instruments could be developed. With these prerequisites met, quantitative research in psychiatric epidemiology has become possible.

Sampling The epidemiological method requires two desirable conditions in the sample of people under study: a suitable reference population or universe can be defined, and individuals in the study can be related to this defined universe in a specified way. If these conditions are satisfied, then results of the study can be generalized to the universe.

The *universe* is the hypothetical population to which the investigator wishes to generalize results; the study population is the operational equivalent of the universe, which is actually studied by the investigator. A universe may be all residents of a delimited area, while a study population may consist of those listed as customers of a local electric utility. The study population should be as equivalent to the specific universe as possible.

Once the study population is defined and judged equivalent to the universe, its members can be assessed to determine their risk or illness status. All members of the study population can be assessed in a total count, if money, time, and other resources permit. Usually, few representative members of the study population form the study sample.

Sampling Design The fundamental principle of sampling is to select subjects so that the sample results represent the entire study population and ensure equivalence to the intended universe. To ensure that the sample does represent the entire study population, the members of the sample must be selected without bias and with a known probability of selection. The basic case is the simple random sample: all members of the study population are listed, and a sample is selected randomly.

Sample Selection Consider, for example, a study population of 3000 people who can be listed individually. The investigator plans to examine about 300 subjects, a 10 percent sample of the study population. Using a random selection procedure (e.g., a table of random numbers), the investigator can select on average 1 in every 10 individuals to be in the study. In simple random samples, every member of the population has an equal and known probability of being selected, in this case a 1 in 10 chance. With large numbers chosen, random selection tends to produce a sample that in the aggregate reflects the composition of the study population. In this example, if the distribution of females in the 3000 population members is 60 percent, it will be roughly the same in the sample; the larger the number of people chosen for the sample, the more likely it is to reflect the characteristics of the study population.

Sample Size The drive to reproduce overall population characteristics in the sample by random selection is one reason why large sample sizes are desirable. Estimates of rates or other statistics are more precise when a large number of individuals are used in the estimation.

Point Estimate Study sample A, drawn by the investigator as described above, is clearly not the only one that could be drawn from the study population. With a different starting point in the table of random numbers, a different series of 300 individuals would be selected, called B. Its composition will also roughly reflect the characteristics of the overall population of 3000. The proportion of females in sample B will also be about 60 percent, but it probably will not be exactly the same as in sample A. If the investigator drew repeated samples and continued to calculate the proportion of females in each sample, these various estimates of the proportion of females in the overall population would form a range of values tending to center around the true population of value of 60 percent. The estimate produced from a sample is called a point estimate of the proportion of females in the study population.

Standard Error Although this repeated sampling of 300 individuals from the total population of 3000 is easily described, it is almost never feasible in a world of limited time, money, and researchers. If the first sample was drawn at random, one can calculate a statistic—the standard error of the proportion—that describes how much variation exists in a large series of repeated samples composed of 300 subjects that could theoretically be selected from a given population. With this standard error, which is easily calculated using data from one sample, one can calculate the range of values that allows investigators to say that they are 95 percent confident that the true proportion of females lies between these estimates. This range is known as the 95 percent confidence interval, and authors customarily report this range along with the point estimates of characteristics they have assessed in their samples. Larger numbers of individuals permit more-stable point estimates, which means that the confidence interval is smaller than it is with only a few subjects in the study.

Confidence Interval Uses Confidence intervals are quite helpful in comparing rates of illness. Suppose that in this sample of 300 subjects the investigator assessed the presence of major depression and wanted to compare rates for all high-school graduates with rates for those with less education. The point estimates may look different (e.g., 3.0 percent for graduates and 9.5 percent for nongraduates). A valid comparison requires considering how well these point estimates reflect the true population-based rates of illness.

Calculating 95 percent confidence intervals helps the investigator determine whether the apparent difference in rates reflects rough point estimates or is a true difference. With such a small sample the confidence intervals may be large, say 0.3 to 5.7 percent for graduates and 4.8 to 14.2 percent for nongraduates; there is an overlap, so the differences may be due to variability of point estimates associated with sampling from the population. Because the 95 percent confidence intervals are large enough to overlap, it may not be possible to say that the difference in point estimates of 3.0 percent and 9.5 percent is real. When confidence intervals overlap, a definitive comparison of point estimates can be made by a statistical test such as the chi-square (c_2) test. Larger numbers of subjects in the sample lead to smaller confidence intervals around point estimates, so there is a premium on having enough subjects. Statistical power analysis allows investigators to plan their studies so that they can be sure of a reasonable chance of having enough subjects to support confidence intervals that are smaller than the anticipated differences in the most important rates that will be compared.

Stratified Sampling In many cases, unfortunately, simple random sampling is not useful. In the example being considered, suppose that the investigator wanted to look at subgroups (e.g., college graduates) that occurred in the population much less commonly than high-school graduates and nongraduates. Rather than collapsing the college graduates into a category with all who finished high school, the investigator may want to be sure that the study sample includes enough subjects who are college graduates. With the equal probability of selection entailed in simple random sampling, college graduates would only be as common in the sample as they were in study population.

A more complicated approach to sampling lists the 3000 individuals in three separate groups, or strata, rather than in a single listing, and then samples each stratum separately. The investigator could decide to have 100 subjects from each educational stratum. If the study population contained 1500 nongraduates, the probability of selecting any one of them for the study would be 100/1500, or about 0.07. If the college graduates numbered only 500, the probability of selection would be 100/500, or 0.20.

This type of sampling is called stratified random sampling, and it requires slightly more complex statistical techniques than simple random sampling does to calculate point estimates and confidence intervals for proportions. In stratified random sampling, each subject is weighted according to the selection probability; estimates can be obtained for specific strata as well as the total study population.

Cluster Sampling One problem with these sampling strategies is that individual members of the study population must be listed to be sampled. Listing all potential subjects may be possible for small populations whose members can be identified (e.g., patients seen in a clinical setting) but is usually not possible or economically feasible for large populations, especially of community residents. An alternative is to use lists of groups, or clusters, of individuals. The only roster of units that can be sampled in a community may be a list of census tracts or a map of city blocks. Once a sample is chosen, all households within the selected block can be listed, and a sample of households can be selected. After the households are chosen, individuals who live there can be entered into the study. Cluster sampling is often the only feasible method, because lists of individuals in a population cannot always be obtained, and it is more efficient than simple or stratified random sampling of individuals

in enrolling a large sample. It produces less precise estimates, however (i.e., larger confidence intervals).

In studies of the total population, cluster sampling can become quite complicated. Clusters can be selected within several different types of strata—households and institutions, urban and rural—and some individuals can be sampled at higher rates (e.g., elderly, adults or children). In these cases, calculation of selection probabilities and the associated efforts to produce confidence intervals and more sophisticated statistical comparisons become highly complex.

Nosology Development of acceptable case-identification instruments for epidemiology requires diagnostic criteria that can be unambiguously interpreted. Once clear, explicit statements of the requirements for each disorder are available, examination protocols can be created. Proposed sets of explicit criteria for psychiatric diagnoses were first published in the 1970s, with the St. Louis research criteria and Research Diagnostic Criteria (RDC). In 1980 DSM-III adopted explicit criteria to use as the basis for clinical practice and for epidemiological and other research; this proved useful for researchers and clinicians and was extended in DSM-IV and on an international basis in ICD-10. Although major advances were made in the international standardization of explicit diagnostic criteria, substantial discrepancies in prevalence rates between similar population surveys have exposed the sensitivity of these criteria and assessment instruments to relatively small changes in content and assessment methods. Alternate approaches to adding additional information on disability and role impairment to assess the clinical significance of syndromes in community populations are now being explored.

Case Assessment Information about a subject can be collected in several ways. Medical records, often used for patients seen in clinical settings, provide essential documentation about prior course of the illness, other disorders or preexisting risks, and the type and course of treatment provided. For some studies, the diagnoses shown in these records are used to identify patients with rare disorders.

Records in central data banks can be important both for very rare disorders and for studying patterns of treatment in a defined area. Case registers are now maintained on a national basis for specific types of cancer. They provide a roster of patients with the disorder and also information that can be used to test hypotheses about risk factors or other features of the illness in the future. For more-common disorders, geographically based case registers provide a record of all those in the population who received treatment for the disorder on a longitudinal basis. For three decades, a register of all patients seen in mental health treatment facilities in Monroe County, New York, has been used to describe population treatment patterns in an urban community. With the recent rise of managed behavioral health care companies in the late 1990s, responsible for both public mental health services and large private insurance contracts, their routine clinical records may be a new source of case-register type research information.

The most important source of information about a subject in the study is often the direct interview or examination. Until the late 1970s the major obstacle in psychiatric epidemiology was the lack of an acceptable case-assessment instrument to identify subjects with mental disorders who could be used in studies of the total area population. The National Institute of Mental Health (NIMH)–Diagnostic Interview Schedule (DIS) and more recent versions of the World Health Organization Composite International Diagnostic Interview (CIDI) are now accepted for use in epidemiological studies involving over 30 countries throughout the world.

Criteria for Assessment Instrument An assessment instrument must satisfy several criteria before it can be used successfully in studies of human subjects.

Safety Assessment should not cause subjects harm; if risks do exist, the benefits of the study must outweigh the potential harm. Anyone participating in such a study must do so voluntarily after being informed of the nature and extent of potential risks and must be able to withdraw at any time. These requirements are enforced by institutional review boards for the protection of human subjects at an investigator's institution. At present, there is no reason to believe that participating in epidemiological studies that include a diagnostic psychiatric interview harms either adults or children.

Feasibility With research funds becoming more limited, it is increasingly important to develop assessment procedures that yield adequate information about large numbers of subjects without being unduly time consuming, burdensome for subjects, or dependent on a highly skilled examiner whose services are expensive. For population surveys that assess large samples, it has become prohibitively expensive to use highly trained mental health clinicians to perform psychiatric interviews. Alternatives have been needed before such major studies could begin.

Reliability and Validity One of the most important psychometric properties of assessment instruments is reliability. Although it has several aspects, reliability usually refers to the assessment instrument's capacity to give consistent results when used by different examiners or at different times. An instrument that has acceptable levels of interrater and test-retest reliability can be assessed to determine if it is measuring what it intends to measure. This property is called validity. Reliability and validity are so important that a more extensive discussion of their role in evaluating case-identification instruments is provided.

Reliability The property of yielding equivalent results when used by different examiners with the same subject is known as interrater reliability. The property of producing equivalent results when used for the same subject on different occasions is called test-retest reliability. Although these two properties are necessary, they are not sufficient for an instrument. For example, at least hypothetically a diagnostic interview could show high interrater and test-retest reliability but give consistently wrong results. Accuracy would be attenuated without consistency, however, so acceptable levels of reliability must usually be demonstrated before any assessment instrument can be used in an epidemiological study. Several measures of reliability are available; choice of the proper measure largely depends on the type of information the assessment instrument uses. In the simplest case, consider an interrater reliability pretest of a diagnostic interview that used two examiners to assess presence or absence of a disorder in a series of 100 individuals. A two-way table can be used to summarize the results from the two interviewers:

Proportion of Subjects Classified by Two Raters

		Examiner A: Present	Examiner A: Absent	Total
Examiner B	Present	$a = 0.07$	$b = 0.08$	$B_1 = 0.15$
	Absent	$c = 0.11$	$d = 0.74$	$B_2 = 0.85$
		$A_1 = 0.18$	$A_2 = 0.82$	1.00

Po = proportion of agreement observed = $a + d = 0.81$
Pc = proportion of agreement by chance = $A_1 \times B_1 + A_2 \times B_2 = 0.724$
$\kappa = (Po - Pc)/(1 - Pc) = (0.81 - 0.724)/(1 - 0.724) = 0.31$

A simple measure of agreement for these dichotomous data is proportion of observed agreement, or the percentage of patients on whom the two examiners agreed that the disorder was present or absent. In this instance, the results indicate 81 percent agreement; however, this measure is not recommended.

In comparing two judgments, some agreement can be expected to occur by chance alone, no matter how consistent the interview is. To adjust for the potential contribution of these chance agreements, a better measure of agreement is kappa. Calculation of kappa as shown in Table 5.1-1 yields a value of 0.31. Another advantage of kappa is that its confidence interval can be calculated to determine how stable the point estimate of kappa's true value is.

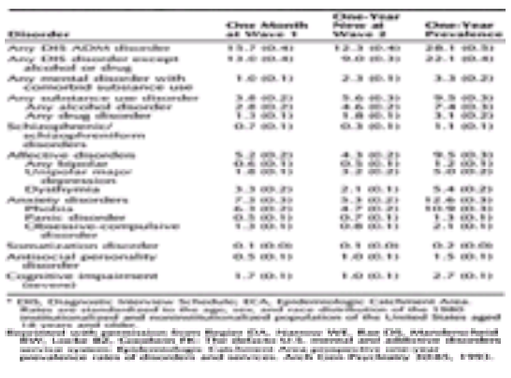

Table 5.1-1 Prevalence per 100 Persons 18 Years and Older: Five ECA Site Combined Community and Institutionalized Population* (Standard Error in Parentheses)

The kappa statistic can be used to measure agreement between the raters at different levels of information. Judgments about any disorder considered in the interview, about just one of the specific disorders (as in the example), or about an individual item for a specific disorder can be described using kappa. As the levels become broader, the reliability is likely to improve, because there are more ways to qualify as having any disorder than there are as having an individual symptom. The value of kappa depends on how common the particular condition is in the study sample. For disorders with a frequency below 5 percent, kappa's value will be so low in most cases that some investigators suggest not calculating it for these conditions.

For data expressed as ratings on an ordinal scale, other statistical measures are appropriate, including an intraclass correlation coefficient using analysis-of-variance techniques. Kappa can also can be calculated for such complex situations as shifting numbers of multiple examiners and weighted items.

Acceptable reliability has become an expected feature of psychiatric assessment instruments, especially diagnostic interviews, because substantial research has demonstrated that routine clinical diagnoses are not reliable. The classic demonstration of this inconsistency was made in the U.S.-U.K. Diagnostic Project, which showed that psychiatrists in New York and London used the same diagnostic terms in widely varying ways, both for their own patients and for patients interviewed on videotape and shown to both groups. One result of this study was to produce standardized interview schedules designed to reduce the most important sources of inconsistency in diagnostic assessments.

1. *Information variance*—examiners collect different types of information about patients
2. *Observation variance*—examiners interpret the subject's answers and nonverbal behavior in different ways
3. *Criterion variance*—examiners evaluate information about the subject according to different diagnostic rules

Standardized interviews reduce, but do not eliminate, these sources of inconsistency. Two other sources of variance remain: (1) subject variance, when the subject has different conditions at different times, and (2) occasion variance, when the subject is in a different stage of the same condition or at least reports different information about it.

Validity Reliability is assessed by comparing agreement between examiners or between examinations. Validity testing requires demonstrating that the test results are accurate, ideally by comparison with a well-known standard of truth. Naturally, a new field that uses descriptive criteria for diagnosis and has just begun developing standardized instruments because clinical examinations are unreliable has no gold standard to use as an absolute measure of truth in diagnostic assessment. In practice, choosing acceptable criterion instrument to use as a standard of comparison in validity testing of a new diagnostic instrument is one of the most difficult aspects of the study.

A simple validity study involves administering the test instrument and the criterion instrument to each subject independently, with the order of administration changed randomly so that one instrument does not influence the results of the other. Different statistical measures are calculated to measure validity than are used for reliability. Reliability tests do not assume that one examiner is more likely than the other to obtain the true answer, so a simple comparison of their agreement is made. In validity testing, however, the criterion instrument is assumed to produce the truth, and the measures calculated are designed to indicate how well results from the new instrument being tested match the results from the criterion instrument. A two-way table is constructed.

		Truth (Criterion Instrument Results)		
		Disorder Present	Disorder Absent	Total
New Instrument	Disorder present	$a = 35$	$b = 8$	$a + b = 43$
	Disorder absent	$c = 5$	$d = 52$	$c + d = 57$
		$a + c = 40$	$b + d = 60$	100

$$\text{Sensitivity} = \frac{a}{a + c} = \frac{35}{40} = 0.875$$

$$\text{False negative} = \frac{c}{a + c} = \frac{5}{40} = 0.125$$

$$\text{Specificity} = \frac{d}{b + d} = \frac{52}{60} = 0.87$$

$$\text{False positive} = \frac{b}{b + d} = \frac{8}{60} = 0.13$$

Sensitivity is a measure of the new instrument's ability to detect those with true cases of disorder identified by the criterion instrument. The *false-negative rate* is the proportion of those with true cases who are missed by the new instrument. *Specificity* is a measure of the new instrument's ability to identify those with true noncases who are identified by the criterion instrument. The *false-positive rate* is most commonly measured as the proportion of those without the disorder who are mistakenly detected as having the disorder by the new instrument.

These measures allow the investigator to judge the new instrument's ability to identify the cases it is designed to identify. Higher sensitivity and specificity values are always desirable. For a given instrument, there are trade-offs between these two values. The only way to improve both sensitivity and specificity without a trade-off is to improve the instrument itself.

Another useful measure related to validity is the proportion of those with apparent cases, as detected by the new instrument, who truly have the disorder as determined by the criterion measure. This proportion is the positive predictive value (PPV). Of the 43 individuals who appeared to have cases in the example, only 35 had true cases, so the PPV = 35/43 = 0.814. This measure is affected by the base rate, so it is especially difficult for diagnostic tests to achieve a high PPV with rare disorders.

Criterion Validity These measures are all based on a type of validity known as criterion validity because results of the test instrument are compared with results of a similar, but presumably more accurate, instrument. If the new instrument is compared with a criterion administered later, the method is called predictive criterion validity.

Face Validity Other types of validity have been described. The simplest is face validity, which refers to a judgment that the new instrument makes sense to an investigator. Standards for this type of validity are usually not clear; however, it can increase clinician and subject acceptability of a new, and otherwise unusual, instrument.

Content Validity *Content validity* refers to a systematic examination of the new instrument by an expert in the area to ensure that it covers the types of information needed for later interpretation and scoring; such instruments as standardized diagnostic interviews that are intended to collect information must be used with a set of explicit diagnostic criteria.

Procedural Validity The term “procedural validity,” recently introduced, refers to whether a particular type of examiner (e.g., nonclinicians) can use the new instrument and produce the same results as a skilled examiner (e.g., a research psychiatrist). Because this concept involves both a comparison between examiners as in reliability testing and a comparison with a criterion as in criterion validity testing, investigators have reported both kappa and sensitivity-specificity measures for

these studies.

Construct Validity The most important concept of validity is construct validity, which is the demonstration that the thing being measured exists the way the instrument assumes it does. Establishing construct validity for a mental disorder is difficult.

Ideally, safety, feasibility, reliability, and validity are examined before an instrument is used to assess subjects in an epidemiological study. In practice, the difficulty in picking a criterion instrument with established validity means that only safety, feasibility, and reliability can be well studied for psychiatric assessment instruments. In an effort to establish at least minimum standards, new instruments can be assessed for content and procedural validity and compared with similar instruments. Although these other instruments cannot be used as legitimate criterion instruments, they can provide useful comparisons if they have met the same standards, including reliability and content validity, as the new instrument being studied.

Descriptive Studies Once an appropriate sampling plan and acceptable assessment instruments are available, epidemiological investigations can begin to explore the distribution of disorders in the population. For rare disorders, a total area population survey may be unrewarding, but for most psychiatric illnesses, the starting point is an area survey to determine basic measures of frequency of the illness.

In epidemiological studies with an emphasis on careful sampling to produce results that can be extended to the universe of interest the results based on number of persons with a disorder are expressed not in raw numbers, but in terms of population rates. In general these rates are proportions that require both a numerator, the number of cases, and a denominator, the total number in the population (with or without the disorder).

$$\text{Rate} = \frac{\text{Persons with the disorder (in the population)}}{\text{Total population}}$$

This rate is actually a proportion, a fraction that includes cases in both the numerator and the denominator. Some writers reserve the term “rate” for an instantaneous change and express it using methods of calculus. To simplify that concept, the term “average rate” can be used for measures that involve rates over specified time periods.

Point Prevalence Most surveys in psychiatric epidemiology have examined the prevalence of disease in the population. The term “prevalence” encompasses several types of rates but always refers to persons who have the disorder at a specified time, regardless of when the disorder started. The basic measure is a point prevalence rate, that is, the proportion of individuals in the population who have the disorder at a specified point in time; this point can be a calendar date, such as April 1, 1990—as in the national census—or it can be defined in relation to the study assessment, such as day of interview, regardless of the calendar date.

$$\text{Point prevalence rate} = \frac{\text{Cases at } t_0}{\text{Population at } t_0}$$

Although a day is often used as the point in time, the use of diagnostic criteria that require multiple symptoms to cluster, with a minimum duration of symptoms, has led some to extend the point measure backward in time to the previous week or previous month. Moving the concept of a point to include a period longer than 1 month prior to interview is not advisable, because it can lead to confusion and raises methodological problems discussed in relation to lifetime prevalence.

To establish a point prevalence rate for depression, an investigator might sample the residents of a defined area, administer an assessment instrument that detects a current depressive illness at that time, and calculate the rate as the proportion of the sample with current depression. Because conducting such a study would likely take several weeks or months, if the sample size is large, the rate would probably be based on the time of interview rather than asking subjects to report their symptoms retrospectively for the starting date of the field survey. (Unless a simple random sample was chosen, the investigator would also need to weight the sample results to reflect the overall population figures according to an individual's probability of selection.)

Incidence In epidemiological research, *incidence* refers to a rate that includes only those whose illness started within a clearly defined time period. The most common time period is 1 year; thus, the annual incidence is usually reported. A study of incident cases is more difficult than a study of prevalent cases. To be done most carefully, a study of incidence requires at least two examinations of each subject in the sample, one at the start and one at the end of the designated time period.

The initial assessment is needed to determine which members of the sample already have the disease; they are not eligible because they already have the illness, so they omitted from the numerator (those with new disorders) at the end of the study. They are also excluded from the denominator, because they are not part of the population considered at risk of developing a new case of the disorder. At the start of the study, the numerator is set to zero; at the end of the study, individuals who have developed the illness will be counted as having new cases. If, at the second examination, the assessment instrument cannot provide retrospective coverage of the entire time period being used (e.g., 1 year), the population at risk must be surveyed more often than just at the end of the period.

For some disorders that have recurrent episodes, it is useful to use an incidence figure that includes both first-episode and later episodes of illness (e.g., a second major depressive episode). A broader concept of total incidence includes those with a new episode of illness, regardless of whether there have been previous episodes. This approach is illustrated in Table 5.1-1, in which new and recurrent episodes are combined for a 1-year total incidence rate, which when added to the point prevalence rate produces a 1-year period prevalence rate.

An alternative way to study incidence is to conduct a single examination of subjects, as if it were the end of the time period chosen for study, and ask them to recall their disease status at the beginning of the time period. Besides problems of recall, this simpler design systematically omits some categories of new cases, including those who died during the time period and cannot be sampled and those who left the circumscribed geographic area for other reasons.

This description of incidence rates is the traditional one, but more-precise concepts of incidence have been developed that take account of change over time and the possible loss to follow-up of individuals in the at-risk population. *Incidence density* (ID) refers to an average incidence rate over the time period of interest.

$$\text{ID} = \frac{\text{Number of incident cases in 1 year}}{\text{Population-time (person-years of observation)}}$$

Because some members of the at-risk population sample will develop the disease and some will drop out of the study (e.g., by moving without a forwarding address, or by dying from an unrelated illness), the denominator is calculated as a person-time figure (e.g., person-years of observation) until the individual leaves the at-risk population. This calculation depends on knowing when during the year an individual develops an illness or otherwise leaves the at-risk population. If for 1000 people being followed the only changes occur at the end of 6 months of follow-up, when 20 are lost to follow-up and another 100 develop disease; then the person-time figure for observation of at-risk, disease-free individuals is (880 × 1 year) + (20 × 1/2 year) + (100 × 1/2 year) = 940 person-years and

$$\text{ID} = \frac{100 \text{ persons}}{940 \text{ person-years}} = 0.1064 \text{ per year}$$

A contemporary example is provided in the Baltimore ECA follow-up survey conducted between 1993 and 1996 by William Eaton and coworkers, with a median time of 12.6 years after the initial interview in 1981. An incidence analysis of DSM-IV major depression identified 71 new cases among 1920 respondents with a total 23,698 person-years of exposure. An incidence density rate of 0.003 per year, or 3.0 per 1000 per year was determined for major depression. This incidence rate is considerably lower than that found during the 1-year follow-up interview in 1982, when 24 new cases were identified with 2550 person-years of exposure—resulting in an incidence density of 9.4 per 1000 per year, three times higher than the rate obtained from an interview requiring a 12.6-year median recall period. When the authors of the Baltimore analysis compared their 1-year incidence rates for major depression with those found in four of the five ECA sites, they found that the average total ECA 1-year incidence rate was 15.9 per 1000 person-years of follow-up.

The problem of recall in psychiatric evaluations is illustrated by the finding that only 5 of the 22 persons with new cases (23 percent) identified in 1982 who could be located for the 1993–1996 interviews confirmed that they had ever had a major depressive episode. Adding the 17 “forgotten” cases to the 71 identified cases would have increased the incidence rate to 3.7 per 1000 per year, and adding a similar number of presumed forgotten cases to the numerator for each year of follow-up

would match the 9 per 1000 found in the 1-year Baltimore follow-up.

One advantage of this more precise approach is that in stable populations the point prevalence rate bears a simple relation to the incidence density: Prevalence = ID × the average duration of illness before termination (e.g., death from the illness or recovery). Here, a stable population is stable in size and age distribution and has constant incidence and prevalence rates. Assuming the total ECA estimates for major depression incidence density of 15.9 per 1000 population per year and the 1 month (point) prevalence of 18 per 1000 yields an average duration of major depression of 18 ÷ 15.9 = 1.1 years. Results of the direct examination of subjects in the Baltimore study demonstrated that the median time to an episode-free year was 1 year for prevalent cases and 2 years for incident cases. Calculation of incidence density rates of mental and other disorders together with mortality, duration, and associated disability was integral to the development of disability adjusted life years (DALYs) for the World Bank and WHO Global Burden of Disease study.

Incidence rates are clearly more difficult to calculate, and they require more extensive data collection than point prevalence rates. Point prevalence rates reflect the accumulated burden of chronic cases of long duration, however, so incidence figures based on new cases in a specified time period are more useful in analytic studies of risk factors. Point prevalence surveys are most useful for descriptive purposes, as in portraying the frequency of illness in a community population and relating disease frequency to potential need for services.

Other Prevalence Rates Although point prevalence and annual incidence rates are the fundamental measures used for descriptive and analytical studies in epidemiology, other types of prevalence figures also have some usefulness.

Period Prevalence *Period prevalence rate* refers to the number of people who have a disorder at any time during a specified period. Its numerator includes any existing cases at the start of the period, plus any new cases that develop during the time period. For a 1-year period, the annual period prevalence rate is approximately equal to the point prevalence rate (existing cases/population at the start) plus the annual incidence rate (new cases in a year/population at risk). In the example of major depression noted above, the point prevalence was 1.8 percent (18 per 1000) and the new plus recurrent incidence was 3.2 percent, for a total 1-year prevalence of 5.0 percent. The period prevalence measure is less valuable than the separate expression of its two components—point prevalence and annual incidence rates. It may be useful, however, for services research studies in which annual treated prevalence rates are contrasted with annual true prevalence rates. Analyses of the ECA study have used each of these measures to describe the de facto U.S. mental and addictive disorder service system.

Lifetime Prevalence The lifetime prevalence rate is a measure of individuals considered at a point in time who have ever had the illness under study. From a clinical perspective, it may best be viewed as equivalent to a past medical history in which the patient is asked to recall any previous health-related conditions, faulty as the person's memory may be. The examiner may only be sure that the patient is most likely to recall the most salient conditions—some of which may be useful in interpreting the present illness. It may be particularly useful to describe conditions that remit but can often recur (e.g., major depression). However, from an epidemiological research perspective, at least three potential problems exist with lifetime prevalence:

1. It is almost always based on subject recall, which can be inaccurate. In addition to the recall problem noted above in the Baltimore ECA incidence study, there was a substantial difference in the reported 7.2 percent lifetime rates for major depression from the total ECA first-wave interview and the 12.5 percent rate obtained when data from both the first- and second-wave interviews are included. The use of a modified interview in the NCS with some memory-enhancing techniques produced a lifetime prevalence rate of 17.3 percent for major depression. Although the Baltimore analysts suggested that the more-severe disorders are the ones most likely to be remembered in long-term recall (lifetime prevalence studies), additional questions are raised about the clinical significance of some disorders identified even in the short-recall periods using existing diagnostic criteria and case-identification techniques in community populations.
2. It covers subjects over the full age range represented in the study population, including many who have not yet completed the age period of highest risk for onset of the disorder. Lifetime rates can be reported separately for specific age groups. This figure may be misleading, however, for some purposes, if no provision is made for the fact that people with the disease at earlier ages may have experienced different death rates than those without the disease.
3. Used in a summary way, it does not allow for the fact that incidence rates may have changed over the years. Therefore, different age groups may have true differences in rates that are obscured both by using an overall rate and by the possible differences in mortality between those with and without the disorder.

Several alternatives to the lifetime prevalence rate, as determined from a point prevalence survey, have been described. One method uses a life table approach to estimate age-specific lifetime rates based on age-specific incidence and mortality figures. This detailed age-specific information is rarely available for psychiatric disorders. Another method calculates lifetime risk by including persons in a birth cohort (with and without the disorder) who died or otherwise dropped out before the cutoff age. This latter method also can be used to ascribe probable risk to those without illness who are below the cutoff age. This approach is the most useful expression of risk, rather than actual development of illness, and is especially important for genetic studies, which must include children and deceased relatives.

Treated Prevalence Treated prevalence usually refers to either a point prevalence or an annual period prevalence rate that is determined by counting all residents in a defined geographical population who receive treatment for a given disorder.

Administrative Prevalence A more restricted form of treated prevalence has been used for studies that use registered patients at a clinical facility as their population denominator. Treated prevalence that uses registered patients as the denominator population has been called administrative prevalence. It presents difficulties, because the denominator depends on registration status, rather than being based on those who perceive the study facility as their usual source of care. With the advent of managed behavioral health care programs, treated prevalence has been used to describe the proportion of the covered population that uses at least one mental health service in 1 year. The less descriptive standard industry term has been *penetration rate*. Potential limitations of both administrative prevalence and penetration rate are the probable inclusion of some patients with no mental disorder who are evaluated or treated and the loss of some who receive treatment outside the insurance plan.

Analytical Studies Measures of disease frequency allow comparisons between groups within the study sample. If one group has higher rates of a disorder than other groups, then a search for factors that led to the higher rate can be undertaken. Many studies of disease frequency have shown that women have higher rates of major depressive disorder than men. From this consistent finding, based on group comparisons, investigators can formulate hypotheses about what factors place women at higher risk. Whether these putative risk factors are biological or psychosocial, they can next be studied in targeted investigations to see if women with the presumed risk factor have higher rates of major depressive disorder than women without them. To make comparisons among groups (and later among groups with and without suspected risk factors), investigators have several options for study design.

Cross-Sectional Prevalence Survey To examine the apparent higher frequency of major depression among women, one approach is to conduct a sample survey of the total population in an area, calculate the point prevalence rates for men and women, and compare them. For example, it may be found that women have twice the point prevalence rate of men. If the investigators examined suspected risk factors, (e.g., use of oral contraceptives), they could compare the rates for women who used this medication and those who did not.

Such a design has been used for this purpose, but it has several drawbacks. As shown in the discussion on incidence rates, point prevalence is affected both by the rate of developing an illness, as reflected in the incidence rates, and by the duration of the illness after it develops. If men and women with major depression had different average durations, they would have different point prevalence rates, even with identical incidence rates; the point prevalence rate for women could be twice the rate for men simply because women with major depressive disorder had more-chronic forms and tended to accumulate in the population. A second problem with this cross-sectional design is that it may be difficult to determine whether women with the suspected risk factor, use of oral contraceptives, were using the medication before they became ill, which would be required to show causality.

Incidence Study and Relative Risk To isolate the frequency of onset from the duration once developed and to assess possible risk factors accurately, studies of incidence rates are much more useful. The fundamental measure used to compare two groups is the relative risk (RR), a form of risk ratio:

$$\text{RR} = \frac{\text{Incidence rate in group with risk factor}}{\text{Incidence rate in group without risk factor}}$$

In terms of the usual two-way table, this risk ratio can be expressed as follows:

Incident Series of Cases	New cases	Noncases
With risk factor	a	b
Without risk factor	c	d

$$RR = \frac{a}{a+b} \div \frac{c}{c+d}$$

If the relative risk is greater than 1.0, the group with the suspected risk factor does have a higher incidence rate of the disorder. In this study design, risk factor status (e.g., use of oral contraceptives) can be determined at the start of the time period for all those without the disorder; these subjects could be reexamined during or at the end of, the period (or both) to assess development of the illness and continued use of the medication.

Prospective Cohort Study The prospective cohort study design assesses incidence rates and risk ratios. A *cohort* is a group formed by sampling from a single, well-specified population, such as residents of a circumscribed area. It allows extension of study findings to the universe of interest, and as a prospective study, it allows determination of risk factor status in persons without the disorder who are followed to identify newly developed cases.

Case-Control Study For very rare disorders (or sometimes for exploratory studies of possible risk factors), a study may use identified cases, rather than potential cases. This design is called a case-control study. It recruits patients already diagnosed, usually by monitoring records at clinical facilities such as hospitals; this method of sampling assumes that everyone with the disorder is likely to present for medical attention and to be accurately diagnosed. It also assumes that either all relevant facilities are monitored or that no differences exist in persons who stay at home or use different facilities.

One problem is that the universe from which these cases are drawn cannot be specified, at least in customary terms (e.g., geographical area of residence). Another problem is that a comparison group of persons without the disorder is not easy to specify. To compensate, often a comparison group is drawn from other patients in the same facility, and a second one is drawn from community residents in the area presumably served by the hospital. Either one of these comparison groups could be drawn from a different universe than the hypothesized one reflected in the sample of cases.

In these case-control studies, ascertaining possible risk factors is usually possible only by retrospective report of the subjects, unless prior documentation exists (e.g., in medical records).

Odds Ratio Because incidence rates cannot be calculated in such a case-control design, the typical calculation of risk ratio must be altered. Instead of an RR, the odds ratio is calculated for case control studies:

	Cases	Noncases
Possible risk factor present	a	b
Possible risk factor absent	c	d

$$\text{Odds ratio} = \frac{a}{b} \div \frac{c}{d} = \frac{ad}{bc}$$

Logistic Regression A common analytical technique in epidemiological studies is logistic regression, which provides an estimate of the probability that an event is true or false. For example, it can be used to estimate whether a disease or a treatment response is present or absent. As in multiple regression, logistic regression examines the impact of different characteristics on the outcome of interest, in this case, the probability of the event.

Analysis of data from a study using logistic regression produces a regression equation, of the form

$$Y = a + b_1X_1 + b_2X_2 + \ldots + b_nX_n$$

One goal in analyzing data with this technique is to produce the best model for predicting whether or not subjects in a study have the outcome of interest. A good model provides a statistically acceptable estimate of the probability Y. Investigators consider different characteristics (e.g., age, gender, socioeconomic status, or laboratory values) to see if they meet scientific and statistical criteria for inclusion in the model. For example, data from a survey of current major depression among patients admitted to a general medical service might be used to test a model including age, sex, and marital status in estimating the probability that major depression is present or absent. Once that model was tested, investigators might examine whether adding family history of mood disorder would improve the model's fit to the data collected from those patients.

This same modeling approach is used in other statistical techniques, but the assumptions, calculations, and interpretations of the model differ. In multiple regression, the b_n coefficient represents change in the dependent variable (Y) for each change in the corresponding independent variable (X_n). For example, a study of adult males might examine the change in weight (Y) for each change in height (X).

In logistic regression, which estimates the probability (Y) that the condition is present or not, the coefficients represent the odds of the condition's occurrence. In logistic regression, the coefficient (b_n) is the natural logarithm of the odds ratio for that independent variable (X_n).

$$b_n = \ln(\text{odds ratio } X_n)$$

Thus the odds ratio for an independent variable can be easily derived from the variable's coefficient in logistic regression models.

$$\text{Odds ratio } X_n = e^{b_n}$$

One advantage of logistic regression in estimating an odds ratio for a characteristic is that the odds ratio is adjusted for other terms present in the regression model.

Bias Analytical studies to assess risk factors can be flawed by three types of bias:

1. Selection bias. If a study sample is not drawn properly, it may not accurately reflect the study population; also, the study population may not be equivalent to the universe of interest. In case-control studies the selection of cases may be distorted if the clinical diagnoses used to determine eligibility are not accurate.
2. Observation bias. If the assessment instruments are invalid, information about subjects will be wrong. In case-control studies, examiners who know the case or noncase status of subjects can be influenced in assessing the risk factors being studied.
3. Confounding bias. In a particular study sample, some other causal factors of the disorder may be related to the risk factor being studied. This effect can often be examined, and some adjustments can be made during data analysis.

These potential flaws can affect the validity of a study's findings. A further question is whether the findings considered applicable to this particular universe also apply to other possible universes; for example, does a risk factor studied in one local community have the same association to disease in other types of communities? The answer to this second question is a matter of judgment until equally valid replications are performed.

Risk Factors and Causality Epidemiological data demonstrate an association between risk factors and a disorder. Because an association does not indicate causality in a particular direction, the desire to study human illness has led epidemiologists to consider standards of evidence that may support an interpretation of a causal connection between a risk factor and a disorder.

1. Temporality. Unless the risk factor precedes the disorder, it cannot cause the disorder. Clear demonstration of the prior occurrence of a risk factor strengthens

the possibility of causality.
2. Replication. Repeated, consistent demonstration of a risk factor's relationship to a disorder in multiple studies strengthens the possibility that a risk factor is causative.
3. Magnitude of association. A large risk ratio tends to support an interpretation of causality, as does demonstration of a doseresponse effect.
4. Plausibility. If a possible causal mechanism can be postulated within the framework of existing knowledge, it may lend credibility to the interpretation of a causal relationship; however, in new fields where knowledge is growing, this criterion may not be so relevant.
5. Specificity. If a proposed risk factor can be associated with a single disorder and no others, that association may strengthen belief in a possible causal relationship. Some factors, such as adverse life events, may lead to a variety of disorders, however, so that the failure of specificity does not discredit a potential risk factor.
6. Experimental intervention. Experiments to control a risk factor may support a causal role for the factor if the occurrence of disease is reduced. However, in such an experiment, unintended or unknown factors may have been the true responsible agents and may have been affected by the intervention.

Experimental Studies An important goal of psychiatric epidemiology is to reduce the burden of illness. With a population-based approach, epidemiological investigations are especially well suited to assessing the impact of broad efforts to reduce the occurrence of illness or its duration and associated disability. Such experimental studies are the next logical development for the WHO/World Bank Global Burden of Disease project. Interventions are now being pilot tested to reduce the burden of disease associated with major depression, which is currently identified as the fourth leading cause of disability and projected to be the second leading cause by the year 2020. By increasing screening, diagnosis, and treatment efforts in primary health care settings and using patient coordinators to ensure compliance with evidence-based protocols, changes in symptom and disability levels can be measured in large primary care population groups. The use of such interventions in community-based, rather than in specialty mental health clinical, populations will permit more-accurate assessment of their impact on changes in the burden of illness associated with major depression.

Population Attributable Risk For primary prevention—the effort to reduce onset of a disorder in persons at risk for developing it—a measure of the potential impact of controlling an established risk factor is derived from the attributable risk (AR), the risk difference measure based on incidence rates (IR) for persons with and without the risk factor.

$$\text{AR} = (\text{IR with risk factor}) - (\text{IR without risk factor})$$

The proportion of cases in the group with the risk factor that may be caused by that factor is the population AR:

$$\text{Population AR} = \frac{\text{IR (total population)} - \text{IR without}}{\text{IR (total population)}}$$

When the more precise figures of incidence density are used, this measure has been called the etiological fraction. In some instances, it may be possible to demonstrate that a “protective factor” has caused a lower rate of illness among those persons exposed to it; then the formula is reversed to show the “preventive fraction.” Studies of low dental caries rates in areas with fluoridated water provide the best example. Calculation of the etiological fraction from observational studies indicates the maximum benefit likely to occur with an intervention designed to reduce the pathogenic effect of the risk factor. Demonstration that an intervention has been effective rests on a controlled design, with calculation of a preventive fraction based on persons with a risk factor who received the intervention and those with the risk factor who did not.

Preventive Applications Over the past decade, the Institute of Medicine (IOM) issued a guide for prevention of mental disorders that separates primary prevention approaches into three types: universal, selective, and indicated. Universal interventions are applied to an entire population (e.g., water fluoridation), selective interventions are applied to populations with some risk factors, and indicated interventions are provided to populations at high risk for developing disorders. In the IOM paradigm, indicated interventions may be applied to individuals who are symptomatic but subthreshold for full diagnostic criteria as well as to asymptomatic persons with biological markers indicating a predisposition for a mental disorder. Decisions to implement any of these approaches depend on the strength of the risk factor information for each of these groups and on the power of the proposed intervention to reduce that risk.

In addition to studies that test interventions to prevent a disorder, this method can also be used for secondary and tertiary prevention. Secondary prevention is aimed at reducing the prevalence of illness by reducing its duration in those who have just developed it. Factors that tend to prolong an episode of illness, including inadequate or inappropriate treatment protocols, can be targeted in secondary prevention efforts.

Reduction of disability produced by a disorder, or tertiary prevention, can also be assessed, even if full recovery does not occur. Such studies require instruments that assess severity of illness, not simply a dichotomous present-absent rating for a disorder. Studies with this design are commonly used in controlled clinical trials to assess the effectiveness of new therapies to reduce duration until recovery or reduce severity in persons without full recovery. For more severe disorders such as schizophrenia, programs of assertive community treatment (PACTs) have been tested in many sites and shown to improve the functional capacity of patients living in community settings.

Primary prevention efforts are especially difficult to study because it is not clear how to choose large groups to study, how to demonstrate that they have received the intervention, and how to ensure that no other, unplanned intervention actually produced any improvement that occurred. Even if these conditions are met, it may be difficult to demonstrate that the intervention was effective for the reasons postulated, and that its costs (including any unintended negative consequences) did not overwhelm any benefits that did occur. However, the potential for implementing such studies will improve as more clearly defined psychosocial and biological risk factors are identified for specific disorders.

EPIDEMIOLOGICAL METHOD IN PSYCHIATRY

Special problems have delayed progress in psychiatric epidemiology. Although the general field of epidemiology has advanced to the point of being most concerned about such issues as controlling confounding bias in risk factor studies and developing ways to measure the impact of preventive interventions, methodological concerns for studies of mental disorders are more basic. They have centered on difficulties in applying complex sampling techniques, in developing acceptable case-assessment instruments, and in identifying potential risk factors that merit study in large-scale investigations. The most important of these problems has been development and application of case-assessment instruments for large-scale studies.

Case Assessment Prior to 1980 the major obstacle to valid identification of cases in large samples of nonclinical populations was the lack of a widely accepted explicit set of criteria for diagnostic classification. Researchers who had access only to earlier systems, such as the first edition of DSM (DSM-I), were forced to rely on detailed descriptions of symptom patterns formulated by the investigators to achieve a meaningful classification of cases. During the 1960s and 1970s, this effort to use detailed specifications for disorders led to more-explicit statements of empirically based criteria for specific mental disorders. Until this goal was met in DSM-III and DSM-IV, investigators had only a few alternatives in selecting instruments to identify cases.

The most common approach used a self-report questionnaire, usually with a concentration on psychophysiological symptoms, which could yield a score based on positive answers by the subject. These scores were intended to reflect the probability that a subject had a diagnosable mental disorder, with higher scores indicating a greater likelihood. Usually, a cutoff score was calculated to separate the sample into two groups—cases and noncases. (In some studies, these scales and other information were also reviewed by clinicians, who attempted to give an overall judgment about the subject's psychiatric status.)

The second major effort to develop standardized assessment instruments was to produce diagnostic interview protocols with acceptable interrater reliability. The first of these was the Present State Examination (PSE), an interview form that concentrates mainly on psychotic conditions, which has been used in major international studies, such as the U.S.-U.K. Diagnostic Project, the International Pilot Study of Schizophrenia, and other studies of schizophrenia supported by the WHO and the NIMH. The PSE is intended for use by skilled clinicians, usually psychiatrists, who have been trained in its use; after training, the interviewers are commonly expected to participate in pretesting any investigation for adequate levels of interrater reliability.

The PSE was not designed to yield diagnostic labels according to an established classification system, because the lack of explicit criteria in the early 1970s made it difficult to determine unambiguous standards for a given category. The advent of formal, detailed criteria for a range of psychiatric diagnoses made it possible to construct interviews that attempted a diagnostic assessment and assign subjects to specific categories of disorder. More recently, the PSE was updated to reflect criteria in DSM-IV and ICD-10 as part of the WHO and National Institutes of Health (NIH)–sponsored Schedules for Clinical Assessment in Neuropsychiatry (SCAN).

New versions of PSE (10th edition) and a categorical diagnostic algorithm (CATEGO) scoring program form the core of the modular SCAN examination protocol. The SCAN was recently used in the first national survey of psychiatric morbidity in Great Britain to confirm functional psychosis diagnoses of schizophrenia, manic episode, bipolar affective disorder, and others.

When the St. Louis research criteria and the RDC were formulated in the 1970s, they were accompanied by interview schedules that investigators would use to assess subjects in a study: the Renard Diagnostic Interview (RDI) for the St. Louis criteria and the Schedule for Affective Disorders and Schizophrenia (SADS) for the RDC. Because skilled clinicians had to spend up to 3 months learning these schedules and because epidemiological studies needed a large number of such high-level interviewers, they were not used much in community surveys. An important exception was use of the Schedule for Affective Disorders and Schizophrenia, Lifetime version (SADS-L) in a follow-up study of 511 community residents in New Haven. That study demonstrated that the SADS-L could be used successfully in epidemiological studies of community subjects.

Stimulated by that demonstration of feasibility and usefulness, and aware that DSM-III would adopt operationalized criteria, as in the St. Louis and RDC systems, NIMH epidemiologists sponsored development of a fully structured interview that could be used by nonclinicians to assess a large number of subjects according to DSM-III criteria. This new instrument, the NIMH DIS, was based on the RDI and SADS interviews and was written by authors of those two earlier interviews. Following extensive use of the DIS in the United States and many other nations, the WHO and three NIH Institutes agreed to sponsor international field trials of an updated and slightly modified DIS. This instrument, the CIDI, has the same structural characteristics of the DIS, with new computer scoring programs that use ICD-10 and DSM-IV diagnostic criteria. Multiple international training sites are now available to prepare research investigators for using both the CIDI and the previously described SCAN in epidemiological and clinical research settings.

Structure of the DIS and CIDI To allow nonclinicians to administer the DIS and CIDI reliably, the interview contains the exact wording to be used for each question; examiners only need to read the question aloud for each item being assessed. Once a question is answered positively, several types of probes are used to determine whether the item can plausibly be counted toward a psychiatric diagnosis. A severity criterion, based on having told a physician or other professional about the symptom or on having experienced life interference as a result of the symptom, is used to separate common and insignificant problems from clinically meaningful phenomena. A symptom that has always occurred as a result of physical illness or injury or in relation to medication, alcohol, or drugs of abuse is coded as being explained by those conditions, not by a psychiatric illness. These two features are intended to ensure consistency, accuracy, and selectivity in the assessment of positive reports of symptoms. Interviewer observations are used only for behavior-related psychotic conditions, and extensive use of marginal notes and examples is encouraged to allow editors and reviewing clinicians to judge any uncertain items.

Another effort to achieve reliability and accuracy is the use of a computer program to score the information and make diagnostic assignments; this program is written in a straightforward style, so that clinicians can easily interpret the scoring algorithms. The reliability of the administration, data entry, and scoring of the interview was greatly facilitated by development of computer-assisted interviews such as the CIDI-Auto developed by the Sydney, Australia, WHO collaborating center. Another computer-assisted interview program, using Windows technology (I-Shell), for use with the CIDI, SCAN, and Disability Assessment Schedule (DAS) was recently developed by the WHO Division of Mental Health. The CIDI and SCAN now cover both ICD-10 and DSM-IV.

Unlike the PSE, which concentrates on the 1-month period prior to interview, the DIS and CIDI assess the occurrence of symptoms throughout the patient's lifetime. This lifetime approach allows diagnosis of some disorders in which symptoms accumulate to a threshold (e.g., somatization disorder, antisocial personality), of disorders in which history is important in the diagnosis (e.g., schizophrenia, bipolar disorder), and prior episodes of recurring disorders (e.g., major depression). This approach also permits estimation of age at onset for subjects in a study as well as identification of subjects with a history of multiple disorders. Once a diagnosis is established on the lifetime basis, its most recent occurrence is dated. A modification of the DIS and CIDI permits dating the most recent experience with any positive symptoms.

Several aspects of the DIS and CIDI approach have been noted as potential weaknesses. Because it minimizes examiner observation and judgment, it is largely limited to subject self-report; self-reports may not be adequate for assessment of some disorders, such as schizophrenia and mania. For example, in the NCS, initial positive results on screening of respondents for schizophrenia with the CIDI were followed by a clinical interview using the Structured Clinical Interview for DSM (SCID). Eliciting these self-reports of symptoms on a lifetime basis means that poor recall may also affect the validity of symptom reporting, particularly for subjects with previous psychotic conditions. Although many aspects of the history of an illness can be obtained, such as age at onset and symptoms during the worst episode (for recurrent illnesses), some aspects of the course cannot be determined. Also, inevitable variations in training and editing procedures may place an effective ceiling on the degree of standardization that multisite studies can achieve.

Characteristics of the DIS and CIDI In the large epidemiological study that it was designed to support, the DIS was shown to be feasible and safe. It has been administered to about 20,000 subjects in five parts of the United States, and the refusal and dropout rate for subjects who agreed to start the DIS has been below 1 percent. The CIDI was used in the NCS to interview subjects as a random sample across the United States. Both the DIS and the CIDI have now been used internationally with many thousands of community and clinical-setting respondents.

Because the RDI and SADS interviews have been demonstrated to yield reliable current and lifetime reports by a range of subjects and because the DIS is similar to them but with more structure and a computer program to score information, it was decided to perform interrater reliability studies at a higher level, called procedural validity. In this first test, agreement was measured between two examiners, a nonclinician and a psychiatrist. Acceptable levels of agreement were reached, and further field testing was undertaken in the context of the epidemiological studies. The CIDI has been shown to have high interrater reliability across a multinational field trial conducted in varied languages.

Criterion validity for a diagnostic interview, especially one based on a new classification system, is hard to test, because no other validated interviews exist. The desire to apply a routine clinical judgment by psychiatrists using the clinical methods suffers from the possible effects of information, observer, and criterion variance that led to development of standardized interview schedules 20 years ago. One alternative is to use similar instruments that apply the same standards of evidence and use scoring procedures based on the same system of diagnostic criteria. The purpose of these studies is not to validate any single instrument, but to understand better any discrepancies with the comparison instrument. Studies comparing the CIDI and SCAN, as well as other interviews such as the SCID, are being undertaken for this purpose.

Construct Validity At least theoretically, the most important form of validity for an assessment instrument, *construct validity* shows that instrument measures what it intends to measure, that in fact the entity exists and can be quantified by the instrument. Various proposals in the psychological literature suggest ways to accomplish this task for a range of purposes. In psychiatric epidemiology, the question involves more than whether an instrument detects a given disorder as specified by a classification system's rules. In practice, these rules cannot be applied independently of a specified, reliable assessment instrument; therefore, the construct validity of the diagnostic entities in a system of classification remains to be shown for many categories for DSM-IV.

As psychiatric nosology has advanced, ways to establish validity for a diagnostic entity have been considered. The first requirement suggested was demonstration that a clinical picture with characteristic features could be described for the disorder and that these features could be used to distinguish the disorder from others that may resemble it. The second type of evidence can be derived from external criteria and includes biological variables (e.g., those obtained from biochemical and neuroanatomical studies); psychosocial and personality styles, including methods for coping with adverse life events; familial and genetic variables (e.g., patterns of transmission in relatives); clinical course and outcome variables (e.g., stability of the diagnosis over time); response to treatment; epidemiological variables (e.g., systematic occurrence in certain groups or places); and multivariate statistical analysis to demonstrate separation from, and relationship with, other disorders in terms of these variables or to examine the latent structure of the underlying category.

Unfortunately, psychiatric diagnostic assessment procedures can rarely evaluate patients with the full range of variables identified above. The differentiation of psychopathology from normal human states is usually not based on structural pathology (except for disorders such as Alzheimer's disease), statistical deviance from a physiological norm, an etiological agent, or an unambiguous pathognomonic symptom presentation. Instead, syndromal patterns of symptoms that are observable or reported by a subject are currently the basis for most diagnoses in psychiatry. The symptom clusters must not only be present, but they must be severe enough to be clinically significant. Although most medical pathology assessments can be separated from the impact of the disorder on social functioning constructs, mental disorder diagnosis requires a clinical significance judgment that implicitly includes evaluating impaired functioning.

Multiple assessment procedures are required to address all of the diagnostic dimensions needed to verify construct validity of current mental disorders. However, the future availability of genetic markers, functional imaging of brain pathophysiology, and better measures of functional impairment should greatly improve our current

efforts to separate psychopathology from normal homeostatic variations in human response to internal and external stresses.

STUDIES OF HISTORICAL IMPORTANCE

Results from epidemiological studies can be most usefully reviewed within the framework of the population studied. This framework emphasizes the primacy of the population group as the basic unit of analysis in epidemiology and illustrates that a variety of descriptive, analytical, and experimental studies may be appropriate with any population group, provided that the selection characteristics of the group are clearly defined and considered in the design. The population groups reviewed here include (1) treated populations in specialty mental health settings, (2) patients registered in general medical practice settings, (3) noninstitutionalized community populations, and (4) defined populations containing specific mental disorders.

Definition of the study population group frequency affects decisions on the type of mental disorder criteria used, interviewer selection, interview method, and the descriptive or analytical variables used. Although the public health framework of epidemiological surveys begins with a community population and proceeds to primary care medical practices and then to specialty mental health–treated populations, the reverse order has more frequently been the rule in the development of psychiatric epidemiology. This pattern is found in the rest of medicine, in which the most frequent progression is from clinical observation to the laboratory for more intensive study—laboratory, in this case, being large population groups. The choice of treated populations reflects the relative ease of defining this population group, the availability of highly trained personnel to conduct assessments, and the existence of extensive diagnostic and other related information in the medical records.

The impetus for proceeding from specialty mental health setting surveys to primary care surveys has been to make the population base more representative of the total community. In so doing, one can still retain the availability of highly trained physicians for case identification and longitudinal medical-record data to extend the pool of information about subjects beyond what is obtainable in cross-sectional studies. Selection bias may affect both the dependent variables of syndromes and diagnoses and the associated descriptive and independent risk factor variables that may lead to better understanding of etiology. Narrowing the population base from which the disorders are drawn to a clinical setting limits generalization to different populations. Likewise, the association of causal or descriptive variables with disorders may be confounded by the tendency of some variables to be strongly associated with use of services, as individuals may visit primary care practitioners or psychiatrists for reasons unrelated to whether they have a mental disorder. A few selected studies of historical importance are presented below. More-comprehensive reviews of psychiatric epidemiology studies are found in summary papers and monographs.

SPECIALTY MENTAL HEALTH SECTOR–TREATED POPULATIONS

Social Class and Mental Illness In the early 1950s August De Belmont Hollingshead and Fredrick Carl Redlich conducted one of the most prominent specialty mental health–treated prevalence surveys. By exhaustively surveying every mental health facility and private office practice psychiatrist who treated any patient in the New Haven area, the authors could define an overall mental disorder, 6-month treated prevalence rate of 8 per 1000 population in the community. The rates vary, however, by socioeconomic status. Treated prevalence rates ranged from 5 to 7 per 1000 in social classes I through IV, to a treated prevalence rate of 17 per 1000 in the lowest social class, V. Clinicians assessed these patients with a protocol report form that required unspecified clinical judgment methods and unspecified criteria for specific diagnoses. Social class was a principal descriptive variable used to stratify frequency of diagnoses and the frequency of treatment setting use. Although careful attention was given to controlling for potential confounding variables of age, sex, or mental disorder type, a strong association was found between higher social class, less severe disorders, and private office treatment settings. This association contrasted with the concentration of severe disorders in both the lower social classes and in the long-stay public mental hospitals. New cases coming into treatment settings (treated incidence) accounted for only one-eighth of the total 6-month prevalence, with a much narrower range of about 1.0 per 1000 for classes I through IV to 1.4 per 1000 for social class V. Incidence rates were less discrepant across the social class groups than prevalence rates, which indicates a longer duration of disorders in the lower social classes (prevalence/incidence = duration). This finding could be partially explained by the diagnostic distribution across social classes, which showed a gradient increase in the proportion of psychotic to neurotic disorders from class I through V.

In summary, this study may be seen as a combined descriptive and analytical epidemiological study of treated prevalence and incidence of mental disorders, in which the community population was examined only for social class and other demographic variables. As such it provided treatment rates and showed strong associations between social class and the frequency, severity, and treatment setting of patients with mental disorders. The absence of standardized assessment procedures in the total population made it impossible to determine if diagnoses were reliably assessed across treatment settings; if social class was associated with mental disorders or only with a treatment frequency; or if social class was causally related to presence of mental disorders or merely concentrated more direct risk factors within this population group.

Despite its methodological limitations, this study had a major impact on the collective social conscience of the day, illustrating that simple frequency and descriptive variable associations can have important administrative and health policy impacts while generating, but not testing, hypotheses about etiology. Implications of social class–determined treatment and the possibility that community characteristics might have an etiological role in mental disorders supported major public policy decisions in the United States to expand specialty mental health services and community mental health center legislation.

National Reporting Program Treated prevalence data are also available from the Mental Disorder National Reporting Program (NRP) transferred in 1992 from NIMH to the Center for Mental Health Services (CMHS) in the Substance Abuse and Mental Health Services Administration (SAMHSA). This program complements the U.S. Public Health Service National Health Survey, which consists of periodic health examinations and facility surveys conducted by the National Center for Health Statistics (NCHS) and other Public Health Service agencies. Components of the NRP include the state and county hospital annual census, the inventory of specialty mental health facilities, and the client/patient sample surveys drawn from the universe of facilities. Routine data on office-based psychiatrists are drawn from the NCHS National Ambulatory Medical Care Survey.

The annual census of state and county mental hospitals describes age, sex, and diagnostic composition of patients in these hospitals at the end of each calendar year and admissions during the year. A comprehensive inventory of all mental health facilities records information on their staffing and organizational characteristics of these facilities as well as the caseload and capacity, including number of episodes of care. *Episodes* are defined as residents at year's end, plus all admissions in the year. As some patients may be admitted more than once in a year, this indicator is slightly inflated over annual treated prevalence rates. Finally the sample patient surveys provide more-detailed information on patient characteristics, including diagnoses, length of stay, and type of treatment received.

Besides being the major source of national service use data, the NRP provides an invaluable source of longitudinal trend data on patients under treatment in the specialty mental health sector. For example, it documents the decrease in number of residents in state hospitals during the period of deinstitutionalization, from a 1955 high of 559,000 to a 1995 level of 69,000 (Fig. 5.1-1).

FIGURE 5.1-1 Number of residents in state hospitals (1950–1995). Patients in mental institutions, 1957–1967, NIMH. Additions and resident patients at end of year state and county mental hospitals by age and diagnosis, by state, United States, 1968–1995.

The shift in care from one type of facility to another can also be analyzed. In 1955, 1.7 million episodes of care were delivered, of which 49 percent were inpatient episodes in state and county mental hospitals, while by 1994, 9.3 million episodes were delivered and only 3 percent were inpatient episodes in state and county

mental hospitals; 12 percent were in general hospital psychiatric units, 6 percent in private psychiatric hospitals, 2 percent in veterans affairs v. medical centers, 1 percent in other inpatient settings, and 76 percent in all outpatient psychiatric services (Fig. 5.1-2).

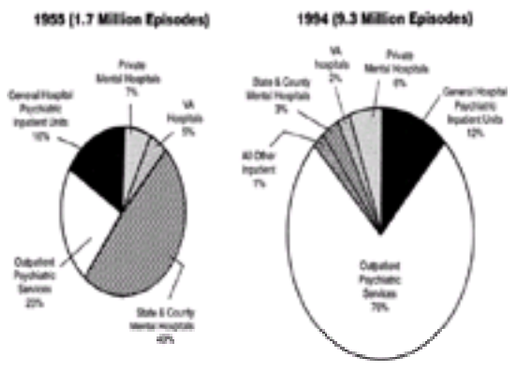

FIGURE 5.1-2 Forty year change in location of mental health treatment episodes (1955–1994). (Reprinted with permission from Redick RW, Witkin MJ, Atay JE, Manderscheid RW: *The Evolution and Expansion of Mental Health Care in the United States between 1955 and 1990*. Mental Health Statistical Note 210, Rockville, MD, 1994.)

Such information illustrates the value of treated prevalence data in describing the workings of the health care system and demonstrates how treated prevalence rates reflect administrative and policy decisions that affect patterns of treatment.

Primary Care Settings Because primary medical care settings reflect a much broader population base than specialty settings and still retain the potential for engaging patients in treatment, many epidemiological investigations have examined mental disorders in general medical populations. Various studies in the United States have indicated that at least 70 percent of the noninstitutionalized population uses general medical services over the course of a year, so general physicians are in a position to screen for mental disorders from a much larger proportion of the population than the 1 to 6 percent who may be seen in specialty mental health services in a year.

Recent studies of mental disorder in primary care patients have used contemporary diagnostic interviews to demonstrate that 15 to 30 percent of patients visiting primary care offices have a current mental disorder. However, some of these same studies suggest that few of these disorders are recognized, diagnosed, or treated by the primary care clinician. Such descriptive epidemiological studies in primary care practice settings have generated remarkable service system responses in the past decade. To address deficiencies in the recognition and treatment of depression, the Agency for Health Care and Policy Research (AHCPR) developed depression treatment guidelines for primary care physicians, and the American Psychiatric Association developed similar guidelines to assist psychiatrists with more complex cases. A promising collaborative primary care/psychiatrist treatment model that includes a patient outcome assessment/treatment compliance coordinator was developed at the University of Washington, Seattle, and the Rand Corporation, Santa Monica, California. The focus on both disability reduction and symptom reduction treatment outcomes may serve as a useful model for addressing the DALYs associated with major depression documented in the World Bank study. A major public health campaign was also developed by NIMH in 1985 to increase the awareness, recognition, and treatment of the most common of these conditions, depression. This effort, known as the Depression/Awareness, Recognition, and Treatment (DART), Project is providing educational materials to the general public, primary care clinicians, and mental health specialists. It has been replicated in the United States by the National Mental Health Association and in other countries such as Great Britain. Similar programs on panic disorder and eating disorders were also initiated in recent years by the NIMH.

Community Population Studies To overcome the problems of selection bias and unreliability of routine clinical diagnosis in treatment settings, epidemiologists rely most heavily on direct surveys of community residents. Prior to World War I, epidemiological studies used institutional records or key informants to generate prevalence estimates of mental disorders. More-recent studies have used direct interviews of community residents, with the information recorded on structured protocols by nonclinicians. Three studies conducted in the 1950s attempted to generate prevalence estimates of mental disorders in accordance with one or more of the major diagnostic practices in use at the time. Two of them (in Stirling County, Canada, and in midtown Manhattan) also examined specific hypotheses about the causal relationship of socioenvironmental factors to the occurrence of mental disorders.

Stirling County The study of Stirling County, directed by Alexander Leighton and continued by Jane Murphy sampled 1010 adults from the 20,000 residents of a rural county in Canada. Independent variables hypothesized to be causally related to mental disorders were cultural and community characteristics. The dependent variables, mental disorders defined in accordance with the newly developed DSM-I, were rated in terms of the presence or absence of 32 detailed symptom patterns, impairment levels, and the need for psychiatric attention. Trained fieldworkers surveyed this population sample with a structured psychiatric interview that was clinically evaluated later by two psychiatrists. Additional information was gathered from all county general practitioners, who supplied more-detailed medical, psychological, and social data about each sample member.

Estimates of the lifetime prevalence of these dependent variables included rates of 57 per 100 population for all DSM-I conditions, 24 per 100 for these conditions associated with significant impairment, and 20 per 100 for these conditions in need of psychiatric attention; this last rate is regarded as the most clinically meaningful estimate. Higher rates of disorders were found in communities characterized by social disintegration, although the exact causal factors within these gross environmental descriptive categories could not be specified.

Midtown Manhattan The midtown Manhattan study of Thomas Rennie, Leo Srole, and colleagues selected a sample of 1660 adult residents from an area of midtown Manhattan that had a population of 110,000 adults. The independent variables included various measures of stress, immigration status, social class, occupation, and marital status; the dependent variables measured impairment in adult life function on a 6-point gradient scale from “none” to “incapacitated.”

Psychologists and social workers used a structured psychiatric interview to obtain information from respondents, which was subsequently rated by two psychiatrists. The most celebrated findings were that 81.5 percent of the population had at least mild impairment from psychological symptoms, whereas 23.4 percent had significant impairment. Although correlations were found between sociodemographic and social stress variables and levels of impairment, one of the most significant results of this study was a debate about the dependent variable measure of mental health status. If less than 20 percent of the population could be identified as mentally healthy, the clinical usefulness of mental disorder definitions was open to question. The importance of using more-rigorous criteria for diagnoses was immediately apparent to other, more medically oriented epidemiologists.

Baltimore A third major study was undertaken in Baltimore, Maryland, where the entire noninstitutionalized population was sampled, including children. This was a two-stage morbidity survey. The first stage, household interviews by census interviewers, was followed by a second-stage clinical examination of 809 subjects, who were stratified by level of disability. General medical internists and pediatricians conducted a medical history and examination, followed by a psychiatrist's rating of all protocols that showed psychological symptoms or impairment related to mental disorders. Because this was a morbidity survey to determine the prevalence of individuals with chronic medical or mental disorders for purposes of improving services, no attempt was made to test etiological hypotheses. Diagnostic categories provided in the ICD were used for all physical and mental disorders, although no effort was made to determine the reliability or validity of assessments. A point prevalence rate of 10.9 percent was found for all ICD mental disorders, with 1.4 percent of the population exhibiting moderate or severe impairment.

Case-Control and Case-Register Studies Population samples are especially important for avoiding selection bias, but they are difficult to use for rare disorders or to generate hypotheses about possible risk factors. Case registers have been maintained for rare disorders that are serious and likely to come to medical attention, such as rare forms of cancer. Over the past three decades, case registers of psychiatric cases have been maintained at great effort in Monroe County (Rochester), New York, and abroad in Copenhagen, Denmark, and Mannheim, Germany. These have been especially useful in providing information about schizophrenia, one of the most difficult disorders to study in population surveys because it is relatively rare and difficult to diagnose accurately. However, privacy concerns and funding difficulties have jeopardized the viability of most total-population psychiatric case registers. Large public systems or private managed care firms with clinical responsibility for treatment may offer epidemiologists a similar research capacity in the future.

Danish Adoption Study One of the most prominent case-control studies in psychiatric epidemiology, that of Seymour Kety and colleagues, began in 1964. They

studied a defined population of 5483 adoptees registered from 1924 through 1947 in the city and county of Copenhagen, Denmark. At that time, the adoptees ranged from 17 to 40 years old. Of the total population, 507 were known to have been admitted to a psychiatric hospital, and 33 of this group were considered to have a diagnosis within a spectrum of schizophrenic disorders. An equal number of matched adoptees who had no history of mental hospitalization were identified for comparison.

A systematic search of available records was conducted to assess the rate of schizophrenia-like conditions in the biological relatives of these patients and comparison adoptees. Of the 150 biological relatives of the schizophrenic adoptees, 8.7 percent had a diagnosis of schizophrenia, compared with 1.9 percent of the 156 biological relatives of the comparison adoptees. The highly significant difference in these rates indicated a genetic contribution to the presence of schizophrenia in those with index cases, since environmental influences had been removed through adoption. This classic study was recently extended by examination of adoptees throughout the rest of Denmark.

An intensive worldwide search is under way for molecular genetic markers associated with schizophrenia in adoptive studies, twin studies, sibling pair studies, and informative pedigree studies. However, no genetic mechanisms that increase passive susceptibility or cause active expression of particular genes have been identified.

Experimental Epidemiology Prospective trials have been used to test the effect of treatments in identified patients, with random assignment of patients to experimental or control groups, to control for possible selection ias. Longitudinal designs allow comparison of expected outcome rates between the control and experimental subjects. Similar clinical trial designs are also being used to test preventive interventions in high-risk populations.

CONTEMPORARY EPIDEMIOLOGICAL STUDIES

Studies Using the Schedule for Affective Disorders and Schizophrenia As the SADS interview was being developed for clinical studies in the mid-1970s and shown appropriate for both patients and normal control individuals in those studies, investigators began testing the feasibility of its use in large, population-based studies of community residents and primary care patients. In a follow-up of 1095 adults first studied in New Haven in 1967, Jerome Myers and Myrna Weissman conducted interviews with 511 who were still available for study in 1975–76 and found that 15.1 percent of the sample had a definite RDC mental disorder at the time of interview. These results attracted special interest in the research community by demonstrating that community-based studies could be conducted using assessment instruments that produced clinically meaningful diagnoses.

Following this successful study, NIMH investigators designed a study for primary care settings, again using the SADS-I. This study demonstrated that nearly 28 percent of patients had a current RDC disorder and that fewer than 1 in 10 of these disorders had been recognized by primary care clinicians. Perhaps equally important, only 14 percent of this primary care population had at least minimal impairment (scored less than 70) on the Global Assessment Scale, and only 9 percent of the 28 percent with a mental disorder (about one-third) had some level of impairment.

Studies Using the Present State Examination Following successful use of the PSE in 12 centers of the WHO's International Pilot Study of Schizophrenia, in both developing and developed countries, European investigators began using the PSE as the core assessment instrument studies of the general population. Although the PSE itself does not generate diagnoses, it provides information that can be related to disease categories in the International Classification of Diseases (ICD) for the month preceding the examination. Some of the larger studies using the PSE as the core assessment instrument have been conducted in the Camberwell section of London; Edinburgh, Scotland; Canberra, Australia; and Athens, Greece.

More recently, the Office of Population Censuses and Surveys carried out the first nationally representative survey of psychiatric morbidity in Great Britain. The survey used a relatively brief (30 minute average) symptomatic screening instrument called the revised version of the Clinical Interview Schedule (CIS-R) to screen 10,000 community residents aged 16 to 64 years. To assess prevalence rates of functional psychosis, a psychosis-screening questionnaire was administered to cover symptoms of mania, thought disorder, paranoia, delusions, and auditory hallucinations in the past year. All respondents who answered this questionnaire positively, stated they or their doctor thought they had a psychotic illness, or were taking any antipsychotic medication were interviewed with the SCAN a few weeks later. To determine 1-year prevalence rates of addictive disorders, subjects were also interviewed with alcohol questions from the 1984 U.S. National Alcohol Survey and the drug use and dependence questions from the ECA study.

One-week (current) point prevalence rates of ICD-10 mood and anxiety (“neurotic”) disorders were determined from scoring algorithms of symptoms assessed by the CIS-R. A hierarchy, used to allocate individuals receiving more than one diagnosis, makes it difficult to identify prevalence rates of individual disorders lower on the following hierarchy which are independent of their co-occurrence with other disorders. The results of this major British National survey included the following diagnostic categories and prevalence rates:

ICD-10 Disorder	Rate per 100 in Past Week
Mixed anxiety and depressive disorder	7.7
Generalized anxiety disorder	3.1
Depressive episode	2.1
Phobias	1.1
Obsessive-Compulsive disorder	1.2
Panic disorder	0.8
Any neurotic disorder	16.0

In addition to the 16.0 percent with 1-week diagnosis identified by the CIS-R, the following 1-year rates were identified by means of the SCAN and separate substance use assessments:

Disorder	Rate per 100 in Past Year
Functional psychoses	0.4
Alcohol dependence	4.7
Drug dependence	2.2

Although much more detailed information is available from a series of surveys of private households, institutions, and homeless people, no annual prevalence rates for all disorders exist for comparison with the annual use of services. In addition, the policy decision to avoid use of the international CIDI instrument in favor of an older and less standardized British CIS-R instrument makes it difficult to make comparisons with U.S. surveys or other international surveys based on the DIS and the WHO CIDI.

NIMH Epidemiologic Catchment Area Program The ECA grew out of a combination of scientific developments and historical events within the mental health field. No scientific development was more critical than the operational criteria for diagnoses that produced the DSM-III and culminated in DSM-IV and ICD-10. A logical extension has used these criteria to construct standardized clinical instruments that reliably elicit information about requisite symptom patterns from research subjects.

In addition to developing operationalized diagnostic criteria and standardized instruments, the 1978 President's Commission on Mental Health demonstrated major gaps in our knowledge of specific mental disorder population rates and how such disorders are treated. Synthesis of information from the NIMH NRP, psychiatric case registers, general practice epidemiological surveys, the Baltimore morbidity survey, and various other data sources provided conservative estimates of overall annual prevalence rates and indicated where individuals with such disorders received care. This analysis of secondary epidemiological and services research data resulted in a mental disorder point prevalence rate estimate of 10 per 100 U.S. population, an annual incidence rate of 5 per 100 population, and an annual period prevalence rate of 15 per 100 population. Treated prevalence data indicated that any mental health specialist in the course of 1 year saw only 3.1 per 100 population and that most (54 percent of the 15 per 100 population) were seen exclusively in the general medical outpatient sector.

Fundamental deficits in knowledge of specific mental disorder frequency rates, the proportion of those treated and untreated, and the locus of treatment became readily apparent. The recommendation to proceed with a new wave of epidemiological and services research studies received favorable response from the NIMH leadership when the need was buttressed with evidence that the new diagnostic criteria and prototype standardized instruments could serve epidemiological

purposes.

Methodology The NIMH ECA program was a multisite epidemiological and health services research study that assesses mental disorder prevalence, incidence, and service use rates in about 20,000 community and institutional residents, using the DIS as a diagnostic assessment instrument. Stratified random probability samples were drawn from geographically defined community populations of at least 200,000 residents to obtain completed interviews from approximately 3000 adult (age 18 and over) community residents and 500 institutional residents. Five participating universities included Yale, Johns Hopkins, Washington University (St. Louis), Duke, and the University of California at Los Angeles. Several of these sites oversampled special populations, including elderly adults and minority groups.

After sample selection, trained lay interviewers used the DIS and service utilization questions to interview each subject. With a longitudinal design, each site conducted wave I and wave II face-to-face interviews (1 year apart) and one intervening telephone or face-to-face interview to assess service use and changes in symptom or diagnostic status. A total of 18,571 persons was interviewed in the first-wave community samples, and the sample was extended to 20,000 by those in prisons, nursing homes, and mental hospitals. U.S. estimates were made by first weighting each person in proportion to their probability of selection in the individual catchment area site and subsequently weighting each person in proportion to their representation in 36 subgroup defined by age, sex, and race-ethnicity characteristics in the 1980 U.S. census of adults age 18 and over.

To illustrate some of the point prevalence, incidence, and period prevalence rate measures discussed above, results are presented here from the two waves of data collection. More-detailed comparisons with data from the NCS are presented to show how different age groups and seemingly minor changes in diagnostic criteria and interview construction can affect community population prevalence rates.

Results The first column of Table 5.1-1 shows 1-month point prevalence from the initial interview, and rates of recurrent and new disorders over the year between initial and follow-up interview are shown in column two. The sum of these two rates (column three) is the 1-year period prevalence obtained from this prospective, longitudinal design. For any alcohol, drug, or mental disorder, 15.7 percent of the adult population had a mental or addictive disorder during the month before the wave-I interview. In the following year, 6.6 percent of this population developed one or more new disorders (none of whom had a lifetime history of a disorder), and 5.7 percent of the population (all of whom had a lifetime history but no current disorder at wave I) developed a recurrent or different disorder at wave II. The sum of incident and recurrent disorders yielded a total incidence rate of 12.3 percent, which combined with the 1-month point prevalence rate to produce an annual period prevalence of 28.1 percent for the adult population. Although the data were obtained from five different cluster areas, they were weighted by age, sex, and ethnicity to reflect the 1980 U.S. Census figures for both noninstitutionalized and institutional populations.

Information on both current prevalence and new and recurrent disorders over the course of 1 year tells more about the course of illness than information obtained from single-wave, cross-sectional designs. For example, the higher rate of new and recurrent disorders (5.6 percent) than current substance use disorders (3.8 percent) indicates a more chronic, relapsing clinical course than that for most other disorders. When the category of any disorder except alcohol or drug is considered, the data show a higher relative proportion of new disorders (6.3 percent) to recurrent disorders (2.7 percent) than that obtained when addictive disorders are included. Other disorders with this pattern include unipolar major depression and antisocial personality disorder; the latter is linked to substance use disorders. The episodic nature of disorders with a relatively low point-prevalence–to–incidence ratio results in significant rates of both incidence/relapse and recovery during the course of a year; thus maintaining a stable point-prevalence rate.

SUBSTANCE USE DISORDERS The 1-year prevalence of any addictive disorder was 9.5 percent, 7.4 percent had alcohol abuse or dependence and 3.1 percent had another drug abuse or dependence disorder. Almost one third of those with an addictive disorder, or 3.3 percent of the population, had a concurrent mental disorder during the same year. This dual diagnosis group has become a source of particular concern for treatment settings that often cannot provide comprehensive care for both conditions. To provide a better scientific base for understanding the relationship between the sequence and timing of the respective onsets of these disorders, the NIMH funded the NCS. Because the ECA had demonstrated that most substance use disorders occurred in the younger age groups, the NCS extended the age downward to 15 years and truncated the upper age range at 54 years. In addition, the NCS used a national probability sample (as opposed to the multisite sample used in the ECA) to obtain closer linkage to institutional and community service settings.

SCHIZOPHRENIA AND SCHIZOPHRENIFORM DISORDEF Either schizophrenia or schizophreniform disorders were identified in 1.1 percent of the population with the DIS approach, which probes each positive symptom answer about interference with life roles, relationship to drugs or alcohol, and help-seeking behavior. This rate is about half that found by the University of Michigan modified CIDI (UM-CIDI) approach of probing severity at the syndrome level rather than the symptom level. However, the NCS provided a clinical evaluation follow-up SCID interview and found a 1-year rate of 0.5 percent for nonaffective psychosis. Using a somewhat different two-stage approach, the Great Britain National Survey, with the SCAN as a follow-up interview, found a 1-year rate of 0.4 percent for schizophrenia. Different diagnostic criteria in the various DSM versions and the ICD-10 provides a source of continuing criterion variance across these studies. Standardizing epidemiological approaches for screening and confirming diagnoses of schizophrenia would be particularly helpful in future international studies. The growing interest in identifying probands for genetic studies has led NIMH-supported investigators to develop an even more comprehensive Diagnostic interview for Genetic Studies for confirmation of the diagnosis before genetic material is obtained and stored for geneotyping analyses.

AFFECTIVE (MOOD) DISORDERS The 1-year prevalence for any affective disorder was 9.5 percent; 1.2 percent met criteria for a bipolar disorder, 5.0 percent criteria for unipolar major depression, and 5.4 percent those for dysthymia. Unipolar depression has the highest annual incidence-or-recurrence rate (3.2 percent) of the affective disorders, compared with a 1.8 percent point-prevalence rate. In contrast, bipolar disorder has about equal rates of point-prevalence (0.6 percent) and incidence or recurrence (0.5 percent). Dysthymia requires a 2-year duration of symptoms and was assessed on a lifetime basis for the ECA study. However, a substantial number of subjects (2.1 percent) developed a new or recurrent episode and were not included among the 3.3 percent of wave-I respondents who were positive. Among the 5.4 percent of the population with dysthymia, 40 percent had a double depression in that they met criteria for another affective disorder during the year. Comparisons of the ECA and NCS rates for mood disorders show similar but slightly higher rates of any mood disorder in the NCS, although the differences in diagnostic interview construction tend to produce a higher proportion of unipolar major depression than dysthymia.

ANXIETY DISORDERS The 1-year prevalence of any anxiety disorder was 12.6 percent, composed primarily of phobic disorders (10.9 percent). More-detailed analyses are also available for the simple phobias, agoraphobia, and social phobia. Panic disorder was found in 1.3 percent of the population, with a slightly higher incidence-or-recurrence rate than point-prevalence rate. Obsessive-compulsive disorder was found in 2.1 percent of the population—a disorder not assessed in the NCS. Since several anxiety disorders, including posttraumatic stress disorder and generalized anxiety disorder, were not covered in all ECA sites, data from these disorders are not part of the public use core data set. However, these disorders were assessed in the NCS with a 1-year prevalence of 3.6 percent for posttraumatic stress disorder, and generalized anxiety disorder was found in 3.4 percent of adults aged 18 to 54 years in an NIMH analysis of the epidemiology of anxiety disorders from both ECA and NCS surveys.

SOMATIZATION DISORDER Somatization disorder is closely related to Briquet's syndrome in the St. Louis research criteria. It was the only somatoform disorder covered in the ECA and was not included in the NCS survey. Found in 0.2 percent of the population in 1 year, it is found predominantly in females who not only visit general medical physicians for physical symptoms, but 70 percent of whom also receive mental health services from either general medical or specialty mental health providers in 1 year.

ANTISOCIAL PERSONALITY DISORDER Antisocial personality disorder, the only Axis II disorder covered in the ECA, was found in 1.5 percent of the population, predominantly males. Criteria for this disorder include the accumulation of some conduct disorder symptoms before age 15 years with other symptom requirements after age 18. Although it is defined as a chronic disorder, there was a relatively high incidence rate, which was most likely linked to a recurrent substance abuse episode that is one criteria of this disorder.

SEVERE COGNITIVE IMPAIRMENT Although cognitive impairment is not a diagnostic category, the prevalence rate of scores between 0 and 17 on the Mini-Mental State Examination was reported to identify populations with a high probability of having organic mental disorders. A 1-year incidence rate of severe cognitive impairment of 1.0 percent provides information on the rate at which the adult population develops cognitive dysfunction in 1 year. Combining this rate with a point prevalence rate of 1.7 percent at the first interview, yields a 1-year prevalence of 2.7 percent. However, separate analyses have shown that only 0.9 percent of the population identified at wave I continued to have severe impairment for 1 year, consistent with chronic dementias such as dementia of the Alzheimer's type or vascular dementia. Slightly less than half of those with severe cognitive impairment at wave I (0.8 percent of the population) showed substantial improvement from more transient conditions.

Services research information, collected concurrently in the ECA with diagnostic information, permits estimates of both 1-year treated prevalence of these disorders and the administrative prevalence (penetration rates), which indicate the proportion of the population using any mental health services regardless of diagnosis. Figure 5.1-3 shows the relationship between the 28.1 percent of the population with any mental or addictive disorder during the year and the 14.7 percent of the population

receiving any mental health services during that year. Of the total number with any 1-year mental or addictive disorder, 28.5 percent (8 per 100 population) received some mental health services during the year. An additional 6.7 per 100 population also received some mental health services during this period, although they were below full diagnostic threshold. Although 45 percent of those using some mental health service during the year did not have a current disorder, 80 percent of the outpatients and 95 percent of the inpatients had a prior lifetime diagnosis subthreshold symptoms of severe disorders associated with significant morbidity or both in other analyses. Hence most of the total 14.7 percent using some mental health services can be considered to have some need for services.

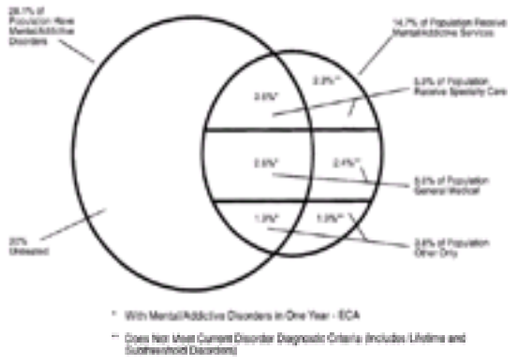

FIGURE 5.1-3 Annual prevalence of mental and addictive disorders and services. (Reprinted with permission from Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK: The defacto U.S. mental and addictive disorders service system: Epidemiologic Catchment Area prospective one-year prevalence rates of disorders and services. Arch Gen Psychiatry *50:*85, 1993.)

Specific detailed information is available from the ECA studies on service use by diagnostic category in various segments of the “defacto U.S. mental and addictive disorders service system.” About 5.9 percent of the U.S. adult population used specialty mental health services, and 6.4 percent sought mental health services from general medical physicians, for an unduplicated total of 10.9 percent using such services in the health system. Other human service professionals from social service agencies and the clergy were contacted by 3.0 percent of the population for “emotions, nerves, alcohol, or drug abuse problems.” A voluntary support network, including self-help groups, family, and friends, was also identified as the source of specific mental health support by 4.1 percent of survey respondents. The hierarchy of services shown in the figure shows that 5 percent of the population received their only mental health services from general medical physicians, and 3.8 percent received their only mental health services from human service professionals or their voluntary support network.

COMORBIDITY

Results from the ECA program demonstrated concomitant disorders were more common than expected from their prevalence rates. Of particular importance was the extensive overlap between mental disorders and alcohol or other substance use disorders. This finding led the three NIH institutes dealing with these conditions to sponsor a national survey of mental disorders co-occurring with alcohol and other substance use disorders, the NCS.

A sample of 8098 respondents aged 18 to 55 years was interviewed from 1990 to 1992 with the UM-CIDI (modified by investigators at the University of Michigan) which reinforced the instructions to subjects to try to recall their past experiences with symptoms relevant to the disorders. Higher rates of disorders resulted from the single interview in the NCS than from the first-wave interview of the ECA. Lifetime rates of mental disorders covered on the UM-CIDI for all respondents were 48.0 percent (versus 32.7 percent in the ECA wave I), and 1-year rates were 29.5 percent (versus 21.8 percent in the ECA wave I). However, when two waves of ECA data were examined and adjustments were made for common age distributions, lifetime rates (46.9 percent) and annual rates (29.8 percent) were more compatible with those found in the NCS.

Despite substantial similarities in the methods and diagnostic criteria of the ECA and NCS studies, discrepancies in prevalence rates between these similar population surveys were observed. These discrepancies and a growing clinical and health policy interest in separating clinically important disorders needing treatment from less severe syndromes prompted a more careful examination of the causes for rate differences between these studies.

A detailed collaborative analysis was undertaken to determine the potential contributions of seemingly minor differences between the two studies in demographic (including age and urban/rural residence) characteristics, use of DSM-III and DSM-III-R criteria, and DIS and UM-CIDI interview construction. By controlling for common demographic variables, standardizing the weighting of respective population groups by age, sex, and ethnicity groups to the 1990 U.S. census, and conducting a detailed comparison of questions in both surveys that addressed DSM-III criteria, rates for any disorder and for selected disorders could be compared. Table 5.1-2 displays the results of this recently published analysis comparing 1-year and lifetime rates for those 18 to 54 years for the ECA I wave, ECA II wave, and NCS I wave.

Disorder	ECA Wave I DSM-III (N = 11,432)	ECA Wave II DSM-III (N = 9,085)	NCS DSM-III-R (N = 7,999)
Any 12-month disorder	24.1 (0.5)	29.8 (0.6)	28.5 (1.0)
Any lifetime disorder	36.1 (0.6)	46.9 (0.7)	48.0 (1.1)
Any substance use or dependence			
12 month	9.5 (0.4)	10.5 (0.4)	11.5 (0.5)
Lifetime	19.9 (0.5)	24.3 (0.6)	28.1 (1.0)
Alcohol dependence			
12 month	4.1 (0.3)	4.4 (0.2)	7.4 (0.5)
Lifetime	8.6 (0.4)	11.3 (0.4)	14.9 (0.7)
Drug dependence			
12 month	2.0 (0.2)	2.4 (0.2)	2.9 (0.3)
Lifetime	4.8 (0.3)	6.4 (0.3)	7.9 (0.5)
Any affective disorder			
12 month	7.0 (0.3)	10.1 (0.4)	11.1 (0.7)
Lifetime	9.3 (0.4)	14.9 (0.4)	19.7 (0.7)
Major depressive episode			
12 month	4.7 (0.3)	6.4 (0.3)	10.1 (0.7)
Lifetime	7.2 (0.3)	12.5 (0.4)	17.3 (0.7)
Dysthymia—lifetime	3.6 (0.2)	5.5 (0.3)	6.7 (0.4)
Any anxiety disorder			
12 month	9.9 (0.4)	11.8 (0.4)	15.3 (0.7)
Lifetime	14.2 (0.4)	19.2 (0.5)	22.8 (0.8)
Panic disorder			
12 month	1.1 (0.1)	1.5 (0.1)	2.2 (0.3)
Lifetime	1.9 (0.2)	2.8 (0.2)	3.6 (0.3)
Social phobia			
12 month	1.6 (0.1)	2.1 (0.2)	7.4 (0.4)
Lifetime	2.5 (0.2)	3.7 (0.3)	13.3 (0.7)

[Reprinted with permission from Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS: Limitations of diagnostic criteria and assessment instruments for mental disorders. Implications for research and policy. Arch Gen Psychiatry 55:109, 1998].

Table 5.1-2 Comparison of One-Year and Lifetime Prevalence Rates Limited to Common Age Range of 18–54 Years (ECA and NCS)

This analysis shows that controlling for common age groups (18 to 54 years) brought results from the two surveys much closer together, despite differences in the diagnostic criteria and assessment instruments. One-year rates for any disorder were 24.1 percent from the retrospective ECA wave-I interview, 29.8 percent from the prospective wave-II ECA analysis, and 28.5 percent from the retrospective NCS single-wave interview. The lifetime rate for any disorder from the two ECA waves combined was 46.9 percent compared with the NCS rate of 48.0 percent. In the context of current health policy issues relating to medical necessity determinations in managed care, these relatively high rates have emphasized the need for greater attention to the clinical significance of these disorders and their implied need for treatment.

The NCS found generally higher rates for both annual and lifetime prevalence, particularly for substance dependence, major depression, and social phobia. However, differences between any affective disorder rate and any substance use disorder rate are not statistically significant. This emphasizes that the NCS instrument identified more symptoms among respondents and tended to move subjects from dysthymia to major depression and from substance abuse into dependence criteria. In the case of social phobia, a clear difference existed in the criteria themselves; DSM-III-R allowed avoidance or distress on exposure, while DSM-III required only avoidance. Because of the somewhat looser diagnostic criteria and additional stem questions for depression and anxiety disorders, the NCS also tended to show more comorbidity between all mental and addictive disorders. This reflects the structural difference in the UM-CIDI which placed stem (criteria A) questions for affective, addictive, and anxiety disorders at the beginning of the entire interview rather than at the beginning of each diagnostic section. This structural change was intended to prevent interviewees from recognizing that a positive response to stem questions resulted in a substantial number of additional questions, which the subject could find fatiguing later in the interview.

As a result of these and other analyses, an international group of investigators is working collaboratively with the WHO to revise the CIDI, standardizing epidemiological assessment techniques to produce more comparable rates in similar population groups. The rate of schizophrenia/schizophreniform disorder from the UM-CIDI was somewhat higher than that found from the ECA. However, a decision was made to reinterview all positive subjects in the NCS with the SCID. The reported 1 year rates of schizophrenia from the reinterview were 0.5 percent for the NCS (versus 2.8 percent who screened positive for possible psychosis on the UM-CIDI, and the 1.5 percent rate reported for the ECA).

An important issue raised by these studies is whether current diagnostic criteria are sufficiently specific to identify the full range of psychopathology—particularly in community settings where perceived disability and need for care are lower than in clinical settings. Approaches to addressing this issue include adding disability assessments to disorder assessments and improving the operational approach to determining the clinical significance of symptom clusters and syndromes. Pressure to address these issues continues from the health policy and insurance fields, which are particularly interested in the epidemiology of clinically significant disorders in need of treatment. However, patients or community residents with less severe disorders or syndromes may continue to be populations for primary or secondary prevention initiatives.

HEALTH POLICY ISSUES

Data from the ECA have been used as a basis for estimating the total direct and indirect cost of mental disorders to the United States. For 1990 the direct (treatment) cost of mental disorders was estimated to be $69.3 billion, with $78.5 billion in indirect (productivity loss) cost for a total cost of $147.8 billion. Although direct treatment cost estimation requires detailed information on service delivery costs, the cost of not treating disorders requires untreated prevalence estimates of the type available only from community population epidemiological studies. Such estimates in the future should require more-precise estimates of treatment need that are not necessarily isomorphic with DSM-IV or ICD-10 disorder prevalence rates.

Direct treatment cost findings can be related to treated prevalence rates and were initially used to estimate the costs of providing parity in coverage of severe mental disorders and physical disorders. Three health policy reports from the National Advisory Mental Health Council to the Senate Appropriations Committee on insurance parity and managed behavioral health care issues have used this epidemiological and services research data.

One of the most significant health policy documents since World War II was the Global Burden of Disease study of the World Bank and WHO, which developed DALYs to assess morbidity and mortality for all illnesses. Estimates of the incidence, prevalence, and duration of mental disorders in the United States and many other countries relied on data derived from studies using the DIS and CIDI. Future confirmation studies must add better measures of disability, such as the WHO DAS, to match prevalence rates of disorders with associated severity, duration, and levels of disability. However, current epidemiological data show the high levels of disability associated with mental and addictive disorders relative to other physical disorders. The use of this common metric for resource allocation decisions affecting services and research funding remains to be seen. Proposals for alerting ministers of health to the importance of such findings appear in World Mental Health, whose endorsement by the United Nations led to WHO sponsorship of a Nations for Mental Health program to destigmatize and increase international funding for mental health services.

Similarly, data on childhood and early adolescent onset of many disorders increased efforts to improve the delivery of mental health services to children and adolescents in the early 1990s. However, definitive epidemiologic data on rates of disorders, service use, and costs remains an unfulfilled research and public policy priority.

FUTURE DIRECTIONS

Recent advances in the fundamental scientific and technical underpinnings of mental disorder epidemiological research include the development of operational criteria that facilitate reliable application in large population groups. Although questions remain about the validity of DSM-IV diagnoses based largely on manifestational criteria instead of causal criteria, the explicit nature of these criteria permits empirical hypothesis testing. Removing presumed etiological criteria from mental disorder diagnoses has facilitated research on etiological hypotheses. Epidemiological studies using DSM-IV criteria will produce external validating criteria to assess the homogeneity of biological, demographic, social, and clinical course characteristics of groups identified with specific disorders. Stability of mental disorder rates in populations with similar characteristics will support the validity of the diagnoses, whereas variations in rates may lead to additional subtyping or even etiological clues.

An excellent cross-cultural epidemiology study was recently conducted by William Vega and coworkers with Mexican immigrants to the United States, native-born adults of Mexican origin, and current Mexican citizens. This Mexican American Prevalence and Services Survey used the UM-CIDI and found a lifetime DSM-III-R mental disorder prevalence rate of only 24.9 percent among immigrants, compared with a 48.1 percent lifetime prevalence among U.S.-born respondents. Although marked differences existed (particularly in substance use disorders), a search for the environmental and cultural causes for such dramatic differences in prevalence rates should provide the basis for careful follow-up investigations.

Improvements in nosological and case-identification tools for epidemiological research will necessitate extending the application of epidemiological methods to new population groups, including those in other cultures. Further refinements in the basic tools for childhood disorders and culture-specific disorders should expand research opportunities in these understudied groups. After 15 years of concentrated effort, investigators collaborating with NIMH have recently developed a Diagnostic Interview Schedule for Children (DISC) and field-tested it in four sites.

When the intrinsic research base of epidemiology is more secure, the interface with such other areas as service systems, clinical services, and clinical and neurobiological research can be extended. Many potential etiological or clinical service–improvement risk factors that emerge from other research areas will require validation in large population studies that use epidemiological research designs.

The full potential of psychiatric epidemiology for public health purposes remains to be realized. Improved understanding of the epidemiology of mental disorders should facilitate allocation of resources to mental health services, broaden the safety and effectiveness of specific treatments for mental disorders, and result someday in effective and reduction in the onset of new cases of these illnesses.

SUGGESTED CROSS-REFERENCES

Other quantitative and experimental methods in psychiatry are discussed in the other sections in Chapter 5. The classification of mental disorders is discussed in Section 9.1. Schizophrenia is the subject of Chapter 12; mood disorders are covered in Chapter 14; anxiety disorders are the focus of Chapter 15; substance-related disorders are discussed in Chapter 11; somatization disorder is discussed in Chapter 16; and personality disorders are the subject of Chapter 54.

SECTION REFERENCES

Andrews G, Peters L: The psychometric properties of the Composite International Diagnostic Interview. Soc Psychiatry Psychiatr Epidemiol *33:*80, 1998.

Bebbington P, Hurry J, Tennant C, Sturt E, Wing JK: Epidemiology of mental disorders in Camberwell. Psychol Med *11:*561, 1981.

Blazer DG, Kessler RC, McGonagle KA, Swartz MS: The prevalence and distribution of major depression in a national community sample: The National Comorbidity Survey. Am J Psychiatry *151:*979, 1994.

Burke JD, Burke KC, Rae DS: Increased rates of drug abuse and dependence after onset of mood or anxiety disorders in adolescence. Hosp Community Psychiatry *45:*451, 1994.

Burke KC, Burke JD, Rae DS, Regier DA: Comparing age at onset of major depression and other psychiatric disorders by birth cohorts in five U.S. community populations. Arch Gen Psychiatry *48:*789, 1991.

Cross-National Collaborative Group: The changing rate of major depression: Cross-national comparisons. JAMA *268:*3098, 1992.

Dean C, Surtees PG, Sashidharian SP: Comparison of research diagnostic systems in an Edinburgh community sample. Br J Psychiatry *142:*247, 1983.

Dohrenwend BP, Levav I, Shrout PE, Schwartz S, Naveh G, Link BG, Skodol AE, Stueve A: Socioeconomic status and psychiatric disorders: The causation-selection issue. Science *255:*946, 1992.

*Eaton WW, Anthony JC, Gallo J, Guojun C, Tien A, Romanoski A, Lyketsos C, Chen L: Natural history of diagnostic interview schedule/DSM-IV major depression. Arch Gen Psychiatry *54:*993, 1997.

Goldstein RB, Weissman MM, Adams PB, Horwath E, Lish JD, Charney D, Woods SW, Sopbin C, Wickramaratne PJ: Psychiatric disorders in relatives of probands with panic disorder and/or major depression. Arch Gen Psychiatry *51:*383, 1994.

Gonzales JJ, Magruder KM, Keith SJ: Mental disorders in primary care services: An update. Public Health Rep *109:*251, 1994.

Henderson S, Duncan-Jones P, Byrne DG, Scott R, Adcock S: Psychiatric disorder in Canberra: A standardized study of prevalence. Acta Psychiatr Scand *60:*355, 1979.

Jenkins R, Bebbington P, Brugha TS, Farrell M, Lewis G, Meltzer H: British psychiatric morbidity survey. Br J Psychiatry *173:*4, 1998.

Kahn HA, Sempos CT: *Statistical Methods in Epidemiology*. Oxford University Press, New York, 1989.

Kendler KS, Gallagher TJ, Abelson JM, Kessler RC: Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a U.S. community sample: The National Comorbidity Survey. Arch Gen Psychiatry *53:*1022, 1996.

Kessler RC, Grank RC, Edlund M, Katz SJ, Lin E, Leaf P: Differences in the use of psychiatric outpatient services between the United States and Ontario. N Engl J Med *336:*551, 1997.

*Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Witchen HU, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. Arch Gen Psychiatry *51:*8, 1994.

*Kessler RC, Zhao S, Katz SJ, Kouzis AC, Frank RG, Edlund M, Leaf P: Past-year use of outpatient services for psychiatric problems in the National Comorbidity Survey. Am J Psychiatry *156:*115, 1999.

Kety SS, Wender PH, Jacobsen B, Ingraham LJ, Jansson L, Faber B, Kinney D: Mental illness in the biological and adoptive relatives of schizophrenic adoptees: Replication of the Copenhagen study in the rest of Denmark. Arch Gen Psychiatry *51:*442, 1994.

Kleinbaum DG, Kupper LL, Morgenstern H: *Epidemiologic Research: Principles and Quantitative Methods*. Lifetime Learning Publications, Belmont, CA, 1982.

*Lyketsos CG, Chen LS, Anthony JC: Cognitive decline in adulthood: An 11.5 year follow-up of the Baltimore Epidemiologic Catchment Area Study. Am J Psychiatry *156:*58, 1999.

Meltzer H, Gill B, Petticrew M, Hinds K: *The Prevalence of Psychiatric Morbidity among Adults Living in Private Household: OPCS Surveys of Psychiatric Morbidity in Great Britain, Report 1* . HMSO, London, 1995.

Michels R, Marzuk PM: Progress in psychiatry. N Engl J Med *329:*552, 1993.

Morris JN: *Uses of Epidemiology*, ed 2. Williams & Wilkins, Baltimore, 1964.

Mrazek PJ, Haggerty RJ, editors: *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. National Academy Press, Washington, DC, 1994.

*Murray CJL, Lopez AD, editors: *The Global Burden of Disease*. Harvard University Press, Boston, 1996.

National Advisory Mental Health Council: *Parity in Financing Mental Health Services: Managed Care Effects on Cost, Access, and Quality*—An interim report to Congress by the National Advisory Mental Health Council. National Institutes of Health, Publication 98-4322, Bethesda, MD, 1998.

Pratt LA, Ford DE, Crum RM, Armenian HK, Gallo JJ, Eaton WW: Depression, psychotropic medication, and risk of myocardial infarction: Prospective data from the Baltimore ECA follow-up. Circulation *94:*3123, 1996.

Price RH, VanRyn M, Vinokur A: Impact of a preventive job search intervention on the likelihood of depression among the unemployed. J Health Soc Behav *33:*158, 1992.

Regier DA: ECA contributions to national policy and further research. Int J Methods Psychiatr Res *4:*73, 1994.

Regier DA: Distinctions between psychopathology and normality in the community (Letter to the editor—in reply). Arch Gen Psychiatry *55:*1147, 1998.

Regier DA, Boyd JH, Rae DS, Burke JD, Locke BZ, Myers JK, Kramer M, Robins LN, George LK, Karno M: One-month prevalence of mental disorders in the U.S.: Based on five epidemiologic catchment area (ECA) sites. Arch Gen Psychiatry *45:*977, 1988.

*Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS: Limitations of diagnostic criteria and assessment instruments for mental disorders: Implications for research and policy. Arch Gen Psychiatry *55:*109, 1998.

*Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK: The de facto U.S. mental and addictive disorders service system: Epidemiologic Catchment Area prospective one-year prevalence rates of disorders and services. Arch Gen Psychiatry *50:*85, 1993.

Regier DA, Rae MA, Narrow WE, Kaelber CT, Schatzberg AF: Prevalence of anxiety disorders and their comorbidity with mood and addictive disorders. Br J Psychiatry *173*(Suppl 34):24, 1998.

Robins LN, Regier DA, editors: *Psychiatric Disorders in America*. The Free Press, New York, 1990.

Rupp A, Gause EM, Regier DA: Research policy implications of cost-of-illness studies of mental disorders. Br J Psychiatry *173*(Suppl):19, 1998.

Samuels JF, Nestadt G, Romanoski AJ, Folstein MF, McHugh PR: DSM-III personality disorders in the community. Am J Psychiatry *151:*1055, 1994.

Shaffer D, Fisher P, Dulcan MK, Davies M, Piacentini J, Schwab-Stone ME, Lahey BB, Bourdon K, Jensen PS, Bird HR, Canino G, Regier DA: The NIMH Diagnostic Interview Schedule for Children Version 2.3 (DISC-2.3): Description, acceptability, prevalence rates, and performance in the MECA Study. Methods for the Epidemiology of Child and Adolescent Mental Disorders Study. J Am Acad Child Adolesc Psychiatry *35:*865, 1996.

Vega WA, Kolody B, Aguilar-Gaxiola S, Alderete E, Catalano R, Caraveo-Anduaga J: Lifetime prevalence of DSM-III-R psychiatric disorders among urban and rural Mexican Americans in California. Arch Gen Psychiatry *55:*771, 1998.

Weissman MM, Myers JK, Harding PS: Psychiatric disorders in a U.S. urban community: 1975–1976. Am J Psychiatry *135:*459, 1978.

*Wells KB: The design of *Partners in Care:* Evaluating the cost-effectiveness of improving care for depression in primary care. Soc Psychiatry Psychiatr Epidemiol *34:*20, 1999.

Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier DA: Cross-cultural feasibility, reliability, and sources of variance of the Composite International Diagnostic Interview (CIDI). Br J Psychiatry *159:*645, 1991.

5.2 STATISTICS AND EXPERIMENTAL DESIGN

ROBERT M. KAPLAN, PH.D., AND IGOR GRANT, M.D., F.R.C.P.(C)

Descriptive statistics are used to organize, summarize, and describe observations; they might include summaries of symptom checklist scores for a selected group of patients with neuroses, a summary of neuroendocrine data for a group of schizophrenia patients, or a descriptive summary of the correlation between watching television violence and behaving aggressively. In addition to descriptive statistics, *inferential statistics* are required for drawing general conclusions about populations on the basis of samples. There are many uses of inferential statistics. In some cases statistical inference is used to make statements about the eating habits of American citizens on the basis of study of a small fraction of the American people. In other instances, inferential statistics are used to decide whether to attribute differences between two groups to chance or to nonchance factors.

This section is addressed to the "informed consumer" rather than the expert biomathematician, and covers a broad range of topics.

DESCRIPTIVE STATISTICS

Descriptive statistics are used to summarize observations and to place these observations within context. The most common descriptive statistics include measures of central tendency and measures of variability.

There are three commonly used measures of central tendency: the mean, the median, and the mode. The *mean* is the arithmetic average, the *median* is the point representing the 50th percentile in a distribution, and the *mode* is the most common score. Sometimes each of these measures is the same; on other occasions, the mean, the median, and the mode can be different.

The mean, median, and mode are the same when the distribution of scores is normal. The *normal distribution*, shown in Figure 5.2-1, is a theoretical distribution of scores that is symmetrical. Under most circumstances the mean, median, and mode will not be exactly the same. The mode is most likely to misrepresent the underlying distribution and is rarely used in statistical analysis. The mean and the median are the most commonly reported measures of central tendency. The major consideration in choosing between them is how much weight should be given to extreme scores. The mean takes into account each score in the distribution; the median finds only the halfway point.

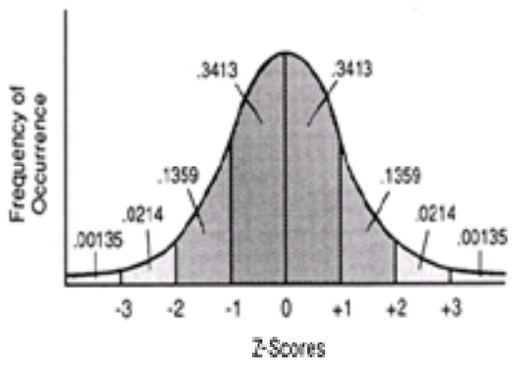

FIGURE 5.2-1 The normal distribution and areas under the normal curve for *Z*-scores from –3 to +3.

Table 5.2-1 shows how the mean and median might have very different values. The table lists the weights of two groups of college women. There are five women in each group, but four of them are in both groups (Sally, Sue, Leslie, and Dana). The fifth member in Group 1 is Allison and in Group 2 is Bertha. These two individuals have very different weights. As the example shows, the median is exactly the same in the two groups. However, the estimation of the mean is affected by the addition of the extreme case. Because the mean best represents all subjects and because of desirable mathematical properties, the mean is typically favored in statistical analysis.

Group 1		Group 2	
Woman	Weight	Woman	Weight
Sally	95	Sally	95
Sue	100	Sue	100
Leslie	105	Leslie	105
Dana	110	Dana	110
Allison	115	Bertha	275
Mean ($\bar{X}$) = 105		Mean ($\bar{X}$) = 137	
Median (Md) = 105		Median (Md) = 105	

Table 5.2-1 Mean and Median for Weights (in Pounds) of Two Groups of College Women

Despite the advantages of the mean, there are also some advantages to the median. In particular, the median disregards outlier cases, whereas the mean moves further in the direction of the outliers. Thus, the median is often used when the investigator does not want scores in the extreme of the distribution to have a strong impact. The median is also valuable for summarizing data for a measure that might be insensitive toward the higher ranges of the scale. For instance, a very easy test may have a ceiling effect but does not show the true ability of some test-takers. A ceiling effect occurs when the test is too easy to measure the true ability of the best students. Thus, if some scores stack up at the extreme, the median may be more accurate than the mean. If the high scores had not been bounded by the highest obtainable score, the mean may actually have been higher.

The mean, median, and mode are exactly the same in a normal distribution. However, not all distributions of scores have a normal or bell-shaped appearance. The highest point in a distribution of scores is called the *modal peak*. A distribution with the modal peak off to one side or the other is described as *skewed*. The word skew

literally means “slanted.”

The direction of skew is determined by the location of the tail or flat area of the distribution. Positive skew occurs when the tail goes off to the right of the distribution. Negative skew occurs when the tail or low point is on the left side of the distribution. Figure 5.2-2 illustrates the normal distribution, a distribution that is positively skewed and one that is negatively skewed.

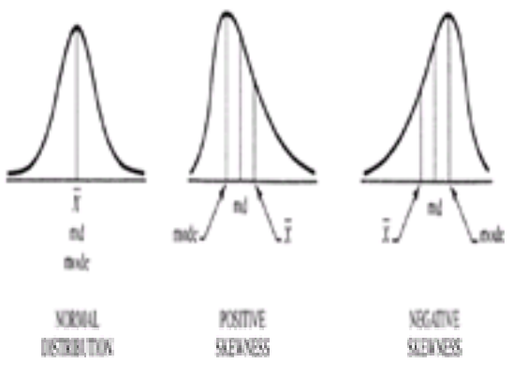

FIGURE 5.2-2 Examples of normal distribution, positive skewness, and negative skewness.

The mode is the most frequent score in the distribution. In a skewed distribution, the mode remains at the peak whereas the mean and the median shift away from the mode in the direction of the skewness. The mean moves furthest in the direction of the skewness, and the median typically falls between the mean and the mode. The relative positions of the mean, median, and mode in normal and skewed distributions are shown in the figure.

Variability Measures of central tendency, such as the mean and median, are used to summarize information. They are important because they provide information about the average score in the distribution. Knowing the average score, however, does not provide all the information required to describe a group of scores. In addition, measures of variability are required. The simplest method of describing variability is the *range*, which is simply the difference between the highest score and lowest score. Another statistic, known as the *interquartile range*, describes the interval of scores bounded by the 25th and 75th percentile ranks; the interquartile range is bounded by the range of scores that represent the middle 50 percent of the distribution.

In contrast to ranges, which are used infrequently in statistical analysis, the variance and standard deviation are used commonly. Since the mean is the average score in a distribution, the sum of the deviations around the mean will always equal zero. Yet, in order to understand the characteristic of a distribution of scores, some estimation of deviation around the mean is important. The sum of these deviations will always equal zero. However, the squared deviations around the mean can yield a meaningful index. The *variance* is the sum of the squared deviations around the mean divided by the number of cases. It is described by the formula:

$$\sigma^2 = \sum \frac{(X_i - \bar{X})^2}{N}$$

where s^2 is the variance, s means sum, X_i is the score of the i_{th} case, ($\bar{X}$) is the mean for the population, and N is the number of observations.

The variance is a very useful statistic and is commonly employed in data analysis. However, its calculation requires finding the squared deviations around the mean rather than the simple or absolute deviations around the mean. Thus, when the variance is calculated, the resulting calculation will be in units that are the natural squared units. Taking the square root of the variance puts the observations back into their original metric. The square root of the variance is known as the *standard deviation*. The standard deviation is an approximation of the average deviation around the mean. Although the standard deviation is not technically equal to the average deviation, it gives an approximation of how much the average score deviates from the mean.

Samples and Populations The methods discussed so far describe populations. *Population* is defined as the entire collection of a set of objects, people, or events, in a particular context. Population refers to the collection of all items upon which statements will be based. This might include all patients with schizophrenia in a particular hospital, or all depressed individuals in a certain community.

In statistics, means and standard deviations for populations are typically estimated from observations taken from samples. A *sample* is a subset of observations selected from the population. It might be unusual for an investigator to describe only patients with schizophrenia in a particular hospital and it is unlikely that an investigator will measure every depressed person in a community. More frequently, statements are made about populations on the basis of samples. Using a random and unbiased sample of individuals from a community, the mean of the population can be estimated. The formulas for estimating the standard deviations for samples differ slightly from those used for populations. The denominator in the radical N-1 is used for sample statistics instead of N, which is used for population statistics. Thus, the definitional formula for the standard deviation of a sample is:

$$S = \sqrt{\frac{\Sigma(X_i - \bar{X})^2}{N - 1}}$$

The computational formula for a sample standard deviation is:

$$S = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{N}}{N - 1}}$$

The computational formula does not require the calculation of each deviation from the mean. Instead, it uses the sum of the squared scores (å X_2) and the square of the summed scores (åX_2). These can be easily obtained using spreadsheet programs for microcomputers. Notice also that the Arabic “S” is used instead of the Greek “å.” It is customary to use Arabic characters for populations and Greek for samples.

Standardized Scores One of the problems with means and standard deviations is that their meanings are not independent of context. For example, a mean of 57.6 means little unless the score is known. The *Z*-score is a transformation into standardized units that provides a context for the interpretation of scores. The *Z*-score is the difference between the score and the mean, divided by the standard deviation.

$$Z = \frac{X - \bar{X}}{S}$$

In other words, a *Z*-score is the deviation of a score, X, from the mean, $\bar{X}$, expressed in standard deviation units. If a score is equal to the mean, its *Z*-score is zero. Consider the example of a score of 4 taken from a sample with a mean of 5.75 and a standard deviation of 2.11. The *Z*-score would be:

$$Z = \frac{4 - 5.75}{2.11}$$
$$= \frac{-1.75}{2.11}$$
$$= -.83$$

Thus, the observed score (4) is .83 standard deviations below the average score or that the score is below the mean, but its difference from the mean is slightly less than one standard deviation. In other words, it deviates from the mean less than the average deviation.

Now we can consider the standard normal distribution because of its central importance in statistics. The normal distribution is derived from binomial probability. A binomial event has one of two outcomes—like the results of a coin flip. Using an infinite number of binomial events, a probability distribution can be generated. Figure 5.2-1 shows the theoretical frequency distribution of "heads" in an infinite series of coin flips.

Most often, the units on the x axis of the normal distribution are in *Z*-units. Any variable transformed into *Z*-units will have a mean of 0 and a standard deviation of 1. Figure 5.2-1 shows the areas under the normal curve associated with these *Z*-units. For example, 34.13 percent or .3413 of the cases fall between the mean and 1 standard deviation above the mean. Since 50 percent of the cases fall below the mean, if a score is one standard deviation above the mean it will be in the 84th percentile rank (50 + 34.13 = 84.13). A score that is 1 standard deviation below the mean would be in the 16th percentile (50 – 34.13 = 15.87).

Translation of *Z*-scores into percentile ranks is accomplished using a table for the standard normal distribution. Certain *Z*-scores are of particular interest in statistics and psychological testing. The *Z*-score 1.96 represents the 97.5th percentile in a distribution whereas –1.96 represents the 2.5th percentile. A *Z*-score of less than –1.96 or greater than +1.96 falls outside of a 95 percent interval bounding the mean of the *Z*-distribution. Some statistical definitions of abnormality view these defined deviations as cutoff points. Thus, a person who is more than 1.96 *Z*-scores from the mean on some attribute might be regarded as abnormal. In addition to the interval bounded by 95 percent of the cases, the interval including 99 percent of all cases is also commonly used in statistics.

There are a variety of other systems by which raw scores can be transformed to give them more intuitive meaning. One common system sets the mean at 50, to correspond to the 50th percentile, and the standard deviation is set at 10. These scores are usually called *T*-scores the same as standard scores (*Z*-scores), except that the mean in the *T* system is 50 rather than 0 and the standard deviation is 10 rather than 1. Indeed, a *Z*-score can be transformed to a *T*-score by applying a linear transformation.

$$T = 10z + 50$$

The *T*-score system is commonly used in psychological tests. For example, the Minnesota Multiphasic Personality Inventory (MMPI) uses *T*-scores. Thus, a score of 60 on any subscale is one standard deviation above the mean. A score of 60 places an individual in approximately the 85th percentile relative to the standardization sample. The score of 70 or above is more than two standard deviations above the mean and thus might be considered unusual in relation to the normative population. Other simple transformations of the *Z*-score are also in common use, for example, the SAT has a mean of 500 and a standard deviation of 100.

Confidence Intervals In most statistical inference problems the sample mean is used to estimate the population mean. Each sample mean is considered to be an unbiased estimate of the population mean. Although the sample mean is unlikely to be exactly the same as the population mean, repeated random samples will form a sampling distribution of sample means. The mean of the sampling distribution is an unbiased estimate of the population mean. However, taking repeated random samples from the population is also difficult and expensive. Instead, it is necessary to estimate the population mean based on a single sample; this is done by creating an interval around the sample mean.

The first step in creating this interval is finding the *standard error of the mean*. The standard error of the mean is the standard deviation divided by the square root of the sample size. Statistical inference is used to estimate the probability that the population mean will fall within some defined interval. Because sample means are distributed normally around the population mean, the sample mean is most probably near the population value. However, it is possible that the sample mean is an overestimate or an underestimate of the population mean. Using information about the standard error of the mean, it is possible to put a single observation of a mean into context.

The ranges that are likely to capture the population mean are called *confidence intervals*. Confidence intervals are bounded by *confidence limits*. The confidence interval is defined as a range of values with a specified probability of including the population mean. A confidence interval is typically associated with a certain probability level. For example, the 95 percent confidence interval has a 95 percent chance of including the population mean. A 99 percent confidence interval is expected to capture the true mean in 99 of each 100 cases. The confidence limits are defined as the values for points that bound the confidence interval.

Creating a confidence interval requires a mean, a standard error of the mean, and the *Z*-value associated with the interval. It is accomplished using the formula:

$$CI = \overline{X} \pm Z_{\alpha} S_X$$

Consider the 95 percent confidence interval. This is obtained by taking the mean ± the standard error of the mean multiplied by the *Z*-score for the 95 percent interval, or 1.96.

If smaller sample sizes are involved, it may be advisable to use another sampling distribution instead of the standard normal distribution. For example, it is common to use the t-distribution for creation of confidence intervals for smaller samples.

***t*-tests** In experimental sciences, comparisons between groups are very common. Usually, one group is the treatment, or experimental group, while the other group is the untreated, or control group. If patients are randomly assigned to these two groups, it is assumed that they differ only by chance prior to treatment. Differences between groups after the treatment are usually used to estimate treatment effect. The task of the statistician is to determine whether any observed differences between the groups following treatment should be attributed to chance or to the treatment. The *t*-test is commonly used for this purpose. There are actually several different types of *t*-tests and they are summarized in Table 5.2-2. The table also shows the formulas used to calculate the *t*-values.

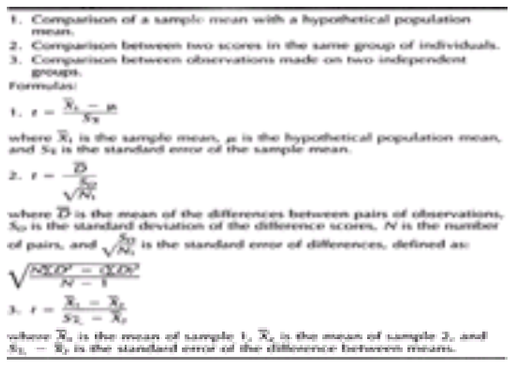

1. Comparison of a sample mean with a hypothetical population mean.
2. Comparison between two scores in the same group of individuals.
3. Comparison between observations made on two independent groups.

Formulas:

1. $t = \frac{\overline{X}_1 - \mu}{S_{\overline{X}}}$

where $\overline{X}_1$ is the sample mean, μ is the hypothetical population mean, and $S_{\overline{X}}$ is the standard error of the sample mean.

2. $t = \frac{\overline{D}}{\frac{S_D}{\sqrt{N_1}}}$

where $\overline{D}$ is the mean of the differences between pairs of observations, S_D is the standard deviation of the difference scores, N is the number of pairs, and $\frac{S_D}{\sqrt{N_1}}$ is the standard error of differences, defined as:

$$\sqrt{\frac{N\Sigma D^2 - (\Sigma D)^2}{N - 1}}$$

3. $t = \frac{\overline{X}_1 - \overline{X}_2}{S_{\overline{X}_1 - \overline{X}_2}}$

where $\overline{X}_1$ is the mean of sample 1, $\overline{X}_2$ is the mean of sample 2, and $S_{\overline{X}_1 - \overline{X}_2}$ is the standard error of the difference between means.

Table 5.2-2 Types of *t*-Tests

Most statistical tests have a common logic in that all of them are based on a ratio of observed to expected differences. The top half of the ratio is the observation. For the *t*-test the numerator is usually the observed difference between means. The denominator of the ratio is the standard error for that same observation. This is an estimate of the extent to which the means would be expected to differ by chance alone. The standard error of the mean is the standard deviation of the sampling distribution of sample means. The standard error of the mean is an approximation of the average deviation that would be expected in estimating the population mean

from a sample. Consider the *t*-test for differences between independent means. In this test the numerator is the observed difference between means of independent samples; the bottom half of the equation is the standard error of the differences between means. The observed differences between means are divided by an approximation of the average deviation that would be expected to result from sampling error. If the ratio is large, then the numerator—the observed deviation—is greater than would be expected by chance. If the ratio is small, then the numerator or difference between groups is equal to or less than what might be expected by chance. The rationale is quite similar for other types of *t*-tests. For example, the *t*-test on paired-observations compares the deviations of pairs (numerator) against the standard error for these deviations (denominator).

Analysis of Variance (ANOVA) Psychiatric and psychological studies often require comparison of the means of more than two groups, for example, a comparison of the depression scores of persons in three groups: mood disorder, chizophrenia, and nonpatient controls. The *t*-tests described above are suitable only for two sample problems. When there are three or more samples, and the data from each sample are thought to be distributed normally, *analysis of variance (ANOVA)* may be a technique of choice.

Mathematically, it can be shown that if two or more samples are drawn from the same population, then the variances (S_2) of the several samples will all be estimates of the population variance (s^2). Furthermore, the ratio of the variances derived from two samples of the same population will form a distribution known as the F-*distribution*. The shape of this distribution depends on the size of each sample, or more accurately on $N - 1$ for each sample, which represents the degrees of freedom (df) associated with each sample; df = $N - 1$ because once the mean is known and all but the very last observation is specified, then the very last observation must be invariant.

When two or more samples are being compared, two sources of variance must be considered. For simplicity, in the three-sample problem investigators may wish to compare depression scores of patients with mood disorders (X_m) to those of schizophrenia patients (X_s) and to controls (X_c). Our task is to evaluate the hypothesis (null hypothesis) that the patients with mood disorders, schizophrenia patients, and the controls have been sampled from the same population with respect to depression scores. This population will have a mean depression score X. The mean scores for the three groups will probably not be exactly the same (even if they come from the same population), but they will vary around X. This is termed *between samples variance* and this variance across means is calculated in the ANOVA procedure.

The between groups variance provides an estimate of the difference between group means. How can it be determined whether this variability is meaningful or random? In order to index between groups variance, an independent estimate of variability is used. Within each of the groups, the variability of individual scores around their own group mean provides an estimate of how depression scores are expected to vary by chance. The variance of individual scores around their group, aggregated across group, means becomes an estimate of error. This is called *within samples variance*. The *F*-ratio is a fraction in which the numerator is the between samples variance estimate and the denominator is the within samples variance estimate. The larger the *F*-ratio, the more likely it is that the null hypothesis can be rejected, that is, that the mood-disordered, schizophrenic, and control subjects came from the same population with respect to depression scores. Critical values of *F*, for various degrees of freedom at desired levels of significance, can be obtained from standard *F*-tables.

So far, the discussion of analysis of variance has considered only a one-way classification, that is, only one grouping factor (in the example above, diagnostic classification) has been considered. In much psychiatric research, subjects will actually be classified in several different ways, giving rise to a two-factor model, a three-factor model, and so on. For example, in the analysis of depression scores the subjects might be divided subjects not only by diagnostic classification (mood disorders, schizophrenia, control normals), but also by gender, giving a three-by-two factorial design (three diagnoses times two genders). *F*-ratios can then be computed for each of the two *main effects* (i.e., diagnosis and gender). A main effect is a statistical difference that is independent of the influence of other variables. It is thus possible to determine whether there are significant differences in depression between diagnostic groups and also between men and women. The factorial design allows us to make each of these assessments independently. Typically, studies are designed so that these independent variables are orthogonal or uncorrelated.

When two or more factors or classifications are being considered, an interaction may occur. Interactions are the joint effects of two or more variables. For example, female patients with mood disorders may have higher depression scores than female patients with schizophrenia or controls, but the differences between diagnostic groups may be smaller for males. In other words, some of the variation in depression scores is attributable to unique combinations of gender and diagnosis that can not be explained by either diagnosis and gender. Such a systematic variation in depression scores as a function both of sex and diagnosis is termed an *interaction*. An interaction is literally a difference of differences. In this example, the difference in depression score attributable to diagnostic groupings is larger for female than for male patients; once again, an F-ratio can be computed for the interaction effect.

Table 5.2-3 illustrates the results of a two-way ANOVA. In this study, Beck Depression scores were obtained for 10 patients with mood disorder, 10 with schizophrenia, and 10 controls. Half of the subjects in each group were women. A glance at the raw scores suggests there were differences between patients and controls and between men and women. Are these differences greater than would be expected by chance (defined as occurring less than 5 percent of the time due to chance if the null hypothesis is true)? The ANOVA computation indicates that the *F*-ratio for the diagnosis effect ($F_{2/24} = 100.79$, $p < .001$) is greater than the tabled value of *F* (5.61) at the .01 significance level for 2 and 24 degrees of freedom (df). Thus, there is a "diagnosis main effect." Similarly, for gender, the *F*-ratio exceeds the tabled value for 1 and 24 df at the .01 level (tabled value = 7.82). Finally, the *F* associated with the interaction is also larger than the tabled .05 level (3.40) for 2 and 24 df, demonstrating systematic influence of the combination of gender and diagnosis on Beck scores (i.e., women with mood disorders had significantly higher Beck scores than would be predicted from considering diagnosis and sex alone).

Data	Controls ($N = 10$)	Schizophrenic Patients ($N = 10$)	Mood Disorder Patients ($N = 10$)
	2	8	11
Men	4	9	9
($N = 15$)	3	8	16
	1	7	12
	2	6	10
	4	12	20
Women	4	11	16
($N = 15$)	2	9	18
	3	11	16
	6	13	22

ANOVA Summary Table*

Source	MS	df	F	p
Diagnosis main effect	354.433	2	100.787	<.001
Sex main effect	116.033	1	32.995	<.001
Interaction	18.433	2	5.242	<.02
Error	3.52	24		

* MS, mean square; df, degrees of freedom; F, the F-ratio; p, probability level.

Table 5.2-3 Summary of Two-by-Three ANOVA for Sex by Diagnostic Group

Chi Square The methods covered thus far are *parametric* because they are used to estimate population parameters such as the mean and the standard deviation. Using these methods requires making assumptions about population characteristics. For example, one of these assumptions is that the variable under study is normally distributed within the population from which the sample is drawn. Although for most applications in basic statistics, violations of this assumption have relatively little impact on a statistical test, there are circumstances in which substantial bias will be introduced. To address this problem, there are other statistical procedures that do *not* make assumptions about population distributions. These are called *nonparametic*, or distribution-free techniques. Many research workers prefer nonparametric methods because they rest on fewer assumptions. These techniques are also appropriate to the analysis of data that do not have continuous numerical properties, for example, data that exist as categories, ordinal data (e.g., clinical ratings of severity on a 6-point scale), or ranked data.

The most commonly used nonparametric test for categorical data is the chi-square. The chi-square statistic is used to evaluate the relative frequency or proportion of events in a population that fall into well-defined categories. For each category, there is an expected frequency that is obtained from knowledge of the population or from some other theoretical perspective. There is also an observed frequency for each category. The observed frequency is obtained from observations made by the investigator. The chi-square statistic expresses the discrepancy between the observed and the expected frequency. The formula for chi-square is:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

where, *O* is the observed frequency, and *E* is the expected frequency.

An example of a situation in which chi-square is used might be to evaluate the association between gender and diagnosis of major depression. The investigator would consider whether each male and each female in a sample did or did not carry the diagnosis. Under the null hypothesis, the expected frequency for major depression would be assumed to be equal for men and women. The chi-square test would be used to evaluate the two-by-two table.

There are also nonparametric alternatives to problems that suggest the possibility of using a *t*-test or ANOVA, but where the properties of the data render application of parametric statistics inappropriate. For a two-sample experiment the Mann Whitney *U* test may be preferable to the *t*-test, whereas, for problems involving 3 or more samples, the Kruskal-Wallis procedure can replace the ANOVA. Both these nonparametric methods involve transformation of the data into ranks.

ODDS RATIO

One common nonparametric statistic is used for the case-control design. In case-control studies, *cases* are participants classified as having the disease; *controls* are similar individuals but who do not carry the diagnosis. The relative risk may not be an appropriate representation of disease risk because both cases and controls are not representative samples from the general population. The risk of exposure among cases may be quite different from that of the general population for both cases and controls.

In order to represent risks, investigators might consider the odds of exposure to a risk factor among cases and divide it by the risk of exposure among controls. Thus, the odds ratio might be represented as:

$$\text{OR} = \frac{\text{Odds of exposure among cases}}{\text{Odds of exposure among controls}}$$

This can be calculated as:

$$\text{OR} = \frac{a/b}{c/d} = \frac{ad}{bc}$$

The odds ratio is interpreted as the chances of being exposed to a risk among those with the diagnosis divided by exposure to the risk among those without the diagnosis. An odds ratio of >1.0 suggests a positive association between the risk and the diagnosis whereas an odds ratio of <1.0 suggests a negative association with the risk factor. The confidence interval can be interpreted in the same way as the relative risk. If the interval includes 1.0, the odds ratio is assumed *not* to differ from chance.

TYPES OF ERRORS

When the null hypothesis is rejected, the observed differences between groups are deemed improbable by chance alone. For example, if drug A is compared to a placebo for its effects on depression and the null hypothesis is rejected, the investigator concludes that the observed differences most likely are not explainable simply by sampling error. The key word in these statements is *probable*. When offering this conclusion, the investigator has the odds on his or her side. However, what are the chances of the statement being incorrect?

In statistical inference there is no way to say with certainty that rejection or retention of the null hypothesis was correct. There are two types of potential errors. A *type I error* occurs when the null hypothesis is rejected when indeed it should have been retained; a *type II error* occurs if the null hypothesis is retained when indeed it should have been rejected.

Type I Error Type I errors occur when the null hypothesis is rejected but should have been retained, such as when a researcher decides that two means are different. He or she might conclude that the treatment works or that groups are not sampled from the same population whereas in reality the observed differences are attributable only to sampling error. In a conservative scientific setting, type I errors should be made rarely. There is a great disadvantage to advocating treatments that really do not work. The probability of a type I error is denoted with the Greek letter alpha (a). Because of the desire to avoid type I errors, statistical models have been created so that the investigator has control over the probability of a type I error. At the .05 significance or alpha level, a type I error is expected to occur in 5 percent of all cases. At the .01 level, it may occur in 1 percent of all cases. Thus, at the .05 a level, one type I error is expected to be made in each of 20 independent tests. At the .01 a level, one type I error is expected to be made in each 100 independent tests.

Type II Error The motivation to avoid a type I error might increase the probability of making a second type of error. In this case the null hypothesis is retained when it actually was wrong. For example, an investigator may reach the conclusion that a treatment does not work when actually it is efficacious. The probability of a type II error is symbolized by the Greek capital letter beta (*B*). Table 5.2-4 shows four possible outcomes broken down in a 2 × 2 table. The left column shows the decisions the researcher made about the null hypothesis. The hypothesis can be rejected or retained. Across the top of the table is the actual situation. The researcher does not know this, but hypothetically this is the decision that should have been made. In teaching, this is often referred to as “God's model,” implying that there is a correct decision that could have been made but it is beyond human capacity to know what it was.

		“God's Model” What Actually Is True	
		H_0 is True It should have been retained	H_0 is False It should have been rejected
Researcher's Decision	Reject H_0	Type I error $p = \alpha$	Correct decision $p = 1 - \beta$ (also called the power of the test)
	Retain H_0	Correct decision $p = 1 - \alpha$	Type II error $p = \beta$

Table 5.2-4 Four Possible Outcomes of Decisions Concerning the Null Hypothesis

Now consider the entries in Table 5.2-4. The upper-left box shows that the null hypothesis is rejected even though it is true. This is a type I error. It will occur with a probability of alpha (a), or the significance level for the test. For example, at the .05 alpha level, the probability of this type of error is 5 in 100. The other type of error is in the bottom-right box in the table. Here the decision is not to reject the null hypothesis when in actuality the null hypothesis was false. This is a type II error with the probability of beta (*B*).

Not all decisions are incorrect. The table also shows two boxes where the researcher made the correct decision. First there is the lower-left box in which the decision was to retain the null hypothesis when indeed the null hypothesis was true. This is a correct decision, and it occurs with a probability of 1 – a. Finally, there is the situation portrayed in the upper-right box. Here, the investigator decided to reject the null hypothesis when indeed the null hypothesis was false. This is a correct decision with a probability of 1 – *B*. The upper right cell is of particular interest, and defines the power of the test.

Statistical Power There are several maneuvers that will increase control over the probability of different types of errors and correct decisions. One type of correct decision is the probability of rejecting the null hypothesis and being correct in that decision. *Power* is defined as the probability of rejecting the null hypothesis when it should have been rejected. Ultimately, the statistical evaluation will be more meaningful if it has high power. It is particularly important to have high statistical power

when the null hypothesis is retained. Retaining the null hypothesis with high power gives the investigator more confidence in stating that differences between groups were non-significant. One factor that affects the power is the sample size. As the sample size increases, power increases. The larger the sample, the greater the probability that a correct decision will be made in rejecting or retaining the null hypothesis.

Another factor that influences power is the significance level. As significance increases, the power increases. For instance, if the .05 level is selected rather than the .01 level, there will be a greater chance of rejecting the null hypothesis. However, there will also be a higher probability of a type I error. By reducing the chances of a type I error, the chances of correctly identifying the real difference (power) are also reduced. Thus, the safest manipulation to affect power without affecting the probability of a type I error is to increase the sample size. The third factor affecting power is effect size. The larger the true differences between two groups, the greater the power. Experiments attempting to detect a very strong effect, such as the impact of a very potent treatment, might have substantial power even with small sample sizes. The detection of subtle effects may require very large samples in order to achieve reasonable statistical power.

It is worth noting that not all statistical tests have equal power. The probability of correctly rejecting the null hypothesis is higher with some statistical methods than with others. For example, nonparametric statistics are typically less powerful than parametric statistics, for example.

CORRELATION

Clinicians or researchers are often interested not so much in how samples differ, but rather in how variables that represent characteristics of some particular sample or population are related to each other. For example, one might wish to know the relation between oral dose of haloperidol (Haldol) and plasma level, between rapid eye movement (REM) latency and Beck Depression Inventory scores, or between volume of the third ventricle and degree of amnesia. Figure 5.2-3 provides results from a clinical study in which plasma haloperidol was determined 1 hour after each of ten single oral doses of the drug. The distribution of points in the scatter plot suggest that these data could be modeled in terms of a straight line, termed a *regression line*. How well can plasma concentration be predicted from oral dose? How much of the total variation in plasma haloperidol, is attributable to the oral dose and how much can be explained by other factors, (e.g., age, diet, body mass, and so forth)?

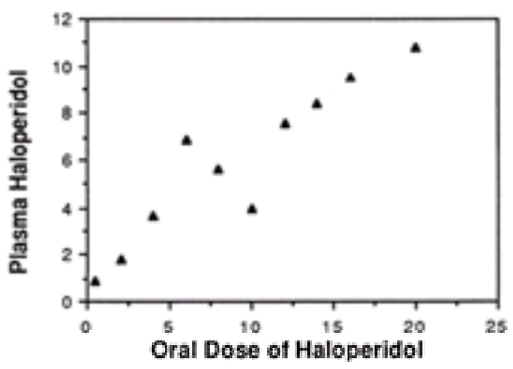

FIGURE 5.2-3 Hypothetical relationship between oral dose of haloperidol and plasma concentration of haloperidol. The linear relationship accounts for 86 percent of the variance.

The regression line that expresses the relationship between plasma haloperidol (*Y*) and oral dose (*X*) can be expressed by the equation: $Y' = a + bx$, where Y' is the predicted value of plasma haloperidol taken from the regression line, *a* is the intercept of the line on the *Y*-axis, and *b* is the slope of the regression line.

Once the regression line has been fitted to the data, what is the strength of the linear relation between plasma haloperidol and oral dose? The strength of the association is expressed as the ratio of the variance in *y* that is attributable to its relation to *x* divided by the total variance in the model. This ratio is called by r^2. In Figure 5.2-3 the r^2 was computed as .86. This means that 86 percent of the total variation in haloperidol plasma concentration can be attributed to its relation to the oral dose, and that 14 percent of the variation is the result of unexplained factors, such as measurement error or characteristics of subjects.

Customarily, the strength of the linear relationship between two variables is actually reported as the correlation coefficient (product moment correlation, Pearson's *r*, or simply *r*). The correlation coefficient is the regression coefficient when both *X* and *Y* variables are expressed in standardized or *Z*-units. The regression coefficient allows a translation between *X* and *Y* in natural units. It is the amount of expected change in *Y* for each unit change in *X*. For example, the equation $Y = 3.25 + .5X$ suggests that each unit change in *X* is expected to correspond with a .5 change in *Y*. The value 3.25 is the intercept. This is the value of *Y* when *X* is 0. The correlation coefficient is also the square root of r^2. In the above example, $r = .93$. To determine whether a product moment correlation falls outside the bounds of chance, one can compute a *t* statistic using a standard formula that takes into account the number of degrees of freedom. In the case of a correlation between two variables, the degree of freedom equals $n - 2$ where *n* is the number of pairs. In the above example *t* was computed as 7.13, which exceeds $t = 2.90$, the critical value of a = .05 for 8 degrees of freedom. Thus, it can be concluded that the *r* is significantly larger than would be expected by chance. The formulas for the correlation coefficient and the evaluation of statistical significance using the *t*-test are:

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}$$

$$t = r\sqrt{\frac{N-2}{1-r^2}}$$

where *N* is the number of cases and *X* and *Y* are scores on measured variables.

Cautions in Interpreting the Correlation Coefficient Several issues need to be borne in mind in interpreting the correlation coefficient. It is important to inspect the data that are being modeled. For example, the linear correlation between plasma concentration of an antidepressant and score on a symptom checklist is low in Figure 5.2-4; however, this is not because the variables are unrelated. The data are best modeled as a U-shaped function, rather than as a straight line.

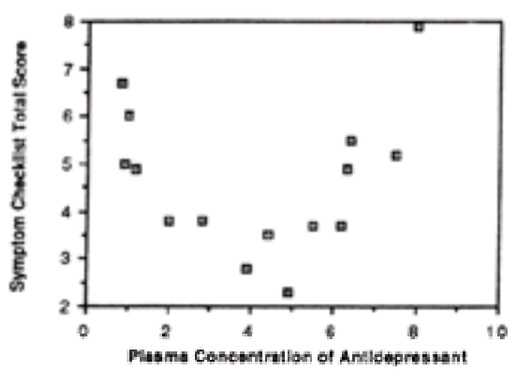

FIGURE 5.2-4 The systematic U-shaped relationship between plasma concentration of antidepressant medication and scores on a symptom checklist is not detected by Pearson product moment correlational methods, which are designed to describe linear relationships.

Correlation coefficients can be biased by a single extreme value. This is illustrated in Figure 5.2-5 and Figure 5.2-6. Figure 5.2-5 shows the correlation between scores on an inventory of life events and a self-rated depression checklist. As the figure shows, there is only a very weak correlation between the scores ($r = -.11$). Figure 5.2-6 displays the same relationship with the addition of a single outlier. This one case had many life events and a high depression score. With the addition of this one extreme point, the correlation is inflated to .97. The example illustrates the sensitivity of correlational methods to extreme scores. Investigators should inspect their data to avoid spurious high correlations caused by outliers.

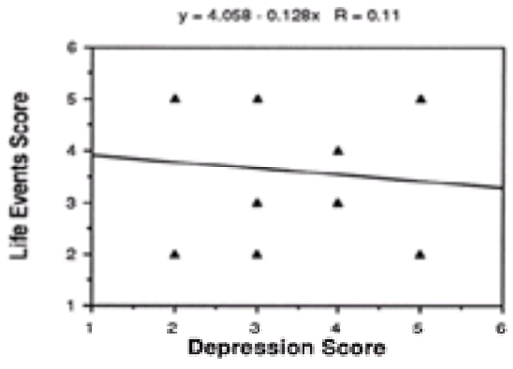

FIGURE 5.2-5 Scatterplot of depression and life events scores.

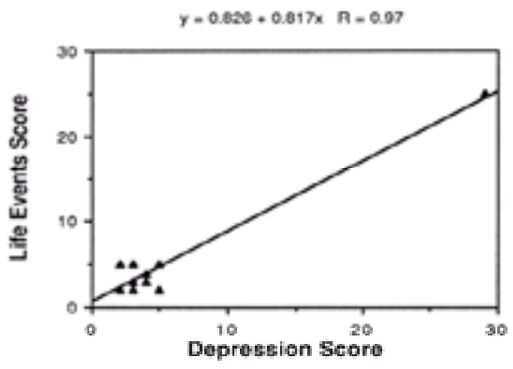

FIGURE 5.2-6 The same data as in Figure 5.2-5, with the addition of one outlier. A single outlier can have a significant effect upon estimates of linear correlation.

Figure 5.2-7 illustrates another problem of interpreting the meaning of the correlation. Here data collected from two samples were pooled and a product moment correlation was computed. This was apparently significant, but an inspection of the scatter plot shows that this apparent linear relationship is explained by the fact that the two samples differed in their mean scores for memory errors and third ventricle width (i.e., the samples came from different populations), but that within each sample there was no relationship between the two variables.

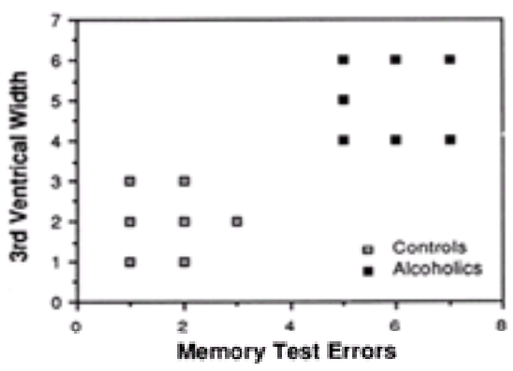

FIGURE 5.2-7 Example of ecological correlation, a correlation that must be interpreted with caution. The correlation between memory and third ventricle width is low within both groups: patients with alcoholism and controls. However, after merging both groups, a significant correlation appears. Just because the correlation from the pooled samples is significant does not demonstrate a true association between ventricle width and memory.

The range of variability in data also will determine the size of the correlation coefficient. For example, if the range of variability is very restricted (i.e., there is not much "play" in the scores), then the observed value of Pearson's r will be constrained near 0. The level of r at which a significant value is attained also depends on the size of the sample. In extremely large samples (e.g., hundreds or thousands of subjects) even tiny values of r (e.g., $r = .1$) may be statistically significant but could have little practical meaning.

Correlation When Variables Are Not Continuous The preceding discussion assumes that the two variables being correlated, x and y, were both continuous. However, correlations can also be computed even if one or both variables are not continuous. It is important to draw the distinction between dichotomous variables that are "true," such as male versus female gender, and those that have artificially been divided to form dichotomies. An artificial dichotomy might be the division of a continuous test score into normal and abnormal categories. The types of correlation coefficients used to find the relationship between dichotomous and continuous variables are summarized in Table 5.2-5. If variable y is continuous (e.g., score on a manual dexterity test), but variable x is a true dichotomy (e.g., sex of children being tested), and if one assigns the value of 0 to one sex and the value of 1 to the other, then computation of the product moment correlation will yield a special statistic called the *point biserial correlation*. The point biserial r can be interpreted in the same way as the standard product moment correlation. The relation between a continuous variable and an artificial dichotomy is evaluated using the *biserial coefficient*.

Variable Y	Variable X		
	Continuous	Artificial dichotomous	True dichotomous
Continuous	Pearson r	Biserial r	Point biserial r
Artificial dichotomous	Biserial r	Tetrachoric r	Phi
True dichotomous	Point biserial r	Phi	Phi

* The entries in the table suggest which type of correlation coefficient is appropriate given the characteristics of the two variables. For example, if variable Y is continuous and variable X is true dichotomous, the point biserial correlation will be used.

Table 5.2-5 Appropriate Correlation Coefficients for Relations Between Dichotomous and Continuous Variables*

The *phi coefficient* (f) is used for correlation's between two dichotomous variables. The phi coefficient is related to the more commonly known chi square described earlier; it can be shown that $c^2 = Nf^2$, where N is the number of cases.

The significance of phi can be determined by computing Nf^2 and referring to a chi-square table with one degree of freedom. When both variables are artificial dichotomies, another coefficient known as the *tetrachoric r* is used.

Rank Order Correlation Sometimes it is difficult to assign precise values to behavioral observations. For example, a clinician might wish to relate the amount of belligerence exhibited by drug-abusing patients to strength of their phencyclidine (PCP) abuse habit. Belligerence may be difficult to quantify, although it might be possible to rank the patients from most to least belligerent. Similarly, the “strength of drug abuse habit” may be difficult to quantify, but it may be possible to order the subjects from heaviest to lightest user. A rank order correlation coefficient, known as *Spearman's rho*, can then be computed. By computing a *Z*-value, using formulas that can be found in standard texts, it is then possible to evaluate the significance of rho for the particular sample size. It should be noted that Spearman's rho may be preferred over the product moment correlation when the variables being related are not normally distributed. For example, the distribution of alcohol consumption among those who regularly drink has a strong positive skew. Rank correlation methods may be preferred in studies of alcohol use because of this nonnormal distribution.

MULTIVARIATE ANALYSIS

Multivariate analysis considers the relation between combinations of three or more variables. For example, the prediction of number of psychiatric readmissions of schizophrenia patients from the linear combination of age, premorbid adjustment, presenting symptoms, and treatment history would be a problem for multivariate analysis. Multivariate analysis is a technical field that requires an understanding of linear and matrix algebra. A schematic representation of the techniques may help place them into context, while showing the basic similarity of approaches that might at first glance appear quite different.

Fundamentally, multivariate techniques involve manipulation of matrix data, that is, data organized in columns and rows. The data in columns are termed *variables* and those in rows are called *observations*. More specifically, multivariate analyses operate on matrix columns.

Figure 5.2-8A,Figure 5.2-8B,Figure 5.2-8C,Figure 5.2-8D,Figure 5.2-8E illustrate how this family of techniques is related. Figure 5.3-8A represents multiple regression. Here, variable Y is predicted from a linear combination (L_x) of variables X_1, X_2, and X_3. In 8b, there are two predictors (X_1, X_2) and also two outcome (dependent) variables (Y_1, Y_2). This illustrates the basis of *canonical correlation*, that is, finding the relationship of the linear combinations of two or more predictors (L_x) and two or more outcomes (L_y) simultaneously. In Figure 5.2-8C there is one outcome (Y) but rather than being continuous, it has discrete levels. This illustrates *linear discriminant analysis*. In Figure 5.2-8D the predictors X_1 and X_2 are dichotomous, but the outcomes Y_1, Y_2, and Y_3 are continuous. This is a representation of *multivariate analysis of variance* with two factors and multiple outcome variables. In Figure 5.2-8E only the relationships among the X variables are being considered, and new linear combinations of these are expressed as the factors Lx_1 and Lx_2. This illustrates *factor analysis*.

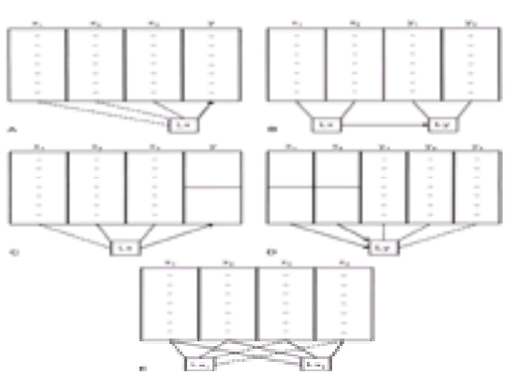

FIGURE 5.2-8 Schematic representations of multivariate models. (Adapted from Van de Geer JP: *Introduction to Multivariate Statistics for the Social Sciences*. Freeman, San Francisco, 1971.)

There is a great deal more to multivariate analysis than this set of schematics has considered. The central point is that what distinguishes these techniques is the *model* being specified more than the statistical theory or computational detail that underlies them. The various methods differ in the number and kind of predictor variables they utilize. They are the same in that they all transform groups of variables into linear combinations. A linear combination of variables is a weighted composite of the original variables. The weighting system combines the variables in order to achieve some goal. The different multivariate techniques differ according to the goal they are trying to achieve.

A linear combination of variables is expressed generally as

$$Y' = a + b_1X_1 + b_2X_2 + b_3X_3 + \ldots + b_kX_k$$

where Y' is the predicted value of Y, a is a constant, X_1 to X_k are variables and there are k such variables, and the b_1 to b_k are regression coefficients. The whole right side of the equation creates a new composite variable by transforming a set of predictor variables.

Example Using Multiple Regression Variables that are important in this combination will be associated with larger regression coefficients. As an example, suppose researchers want to predict the number of psychiatric hospitalizations from three variables: income, rating by psychiatrists, and age. This type of multivariate analysis is called *multiple regression*, and the goal of the analysis is to find the linear combination of the three variables that provides the best prediction of number of hospitalization episodes. The correlation between the criterion (hospital admissions) is found with some composite of the predictors (income plus psychiatrist rating plus age). The combination of the three predictors, however, is not just the sum of the three scores. Instead, a computer algorithm is used to find a specific way of adding the predictors together that will make the correlation between the composite and the criterion as high as possible. A weighted composite might be:

$$\text{Admissions} = .3\ (Z \text{ scores for income})$$
$$+\ .6\ (Z \text{ scores of psychiatrist ratings}) + .03\ (Z \text{ scores for age})$$

This example suggests that psychiatrist ratings are given more weight in the prediction of hospital admissions than the other variables. The rating is multiplied by .6, whereas the other variables are multiplied by much smaller coefficients. Age is multiplied by only .03, which is very close to no contribution. Because any number multiplied by 0 will be 0, age will almost drop out of the equation.

The reason for using *Z*-scores for the three predictors is that the coefficients in the linear composite will be greatly affected by the range of values taken on by the variables. Income is measured on a scale of dollars, while the range in age might be 15 to 70. To compare the coefficients to one another all the variables need to be transformed into similar units. This is accomplished by using *Z*-scores. The standardized coefficients attached to these variable *Z*-scores are termed b or b weights. When the variables are not expressed in *Z*-units, the coefficients or weights for the variables are expressed in their natural units. For example, to find an equation to estimate someone's predicted level of success on the basis of certain personal characteristics, using coefficients that applied to the untransformed values would be

useful. When this is done the weights in the model are called *raw regression coefficients* (sometimes called *b*s).

Interpreting regression coefficients can sometimes be difficult. In addition to being a reflection of the relation between a particular variable and the criterion, the coefficients are affected by the relation among the predictor variables. When the predictor variables are highly correlated with one another, it is difficult to evaluate their individual coefficients. Two predictor variables that are highly correlated with the criterion will not both receive large regression coefficients if they are highly correlated with one another.

For example, suppose that income and psychiatrist's ratings are both highly correlated with readmission. However, these two predictors also are highly correlated with one another. In effect, the two measures seem to be of the same thing (which would not be surprising because the psychiatrist might include income in his or her overall appraisal). So the psychiatrist's rating may get a lower regression coefficient because some of its predictive power is already taken into consideration through its association with income. This is known as the problem of *multicolinearity*. Regression coefficients are most easily interpreted when the predictor variables do not overlap.

The strength of the association between the predictors and the outcome is expressed, as in simple correlation, as a correlation coefficient, usually termed *multiple R*. Note that the upper case *R* is conventionally used in multiple regression rather than the lower case *r* typically used in bivariate correlation. Squaring *R* provides an estimate of the amount of variance in *Y* explained by the predictors *X*s. There are methods for determining the significance of *R*. It is also possible to compute an "adjusted *R*," which takes numbers of subjects and variables into account. This is desirable because the replicability of the regression model becomes less likely if there are too many predictor variables in relation to number of subjects from which such observations derive. Generally speaking, as the ratio of subjects to variables begins dropping below 10:1, confidence in the replicability of the regression should dwindle.

Discriminant Analysis Multiple regression is appropriate when the criterion outcome variable is continuous. However, there are many cases when the criterion is a set of categories. For example, a researcher might want to know the linear combination of demographic, historic, and symptom variables that differentiate positive symptoms and negative symptoms schizophrenic patients. The task is to find the linear combination of variables that provides a maximum discrimination between categories. One appropriate way to model this problem is through application of linear discriminant analysis. The technique attempts to find a linear combination of predictors that achieves best separation between positive and negative symptom cases. A model chi-square can be computed to assess the significance of the solution. If an apparently successful classification is achieved, it is important to determine the generalizability of the model through one of several techniques and cross-validation. *Cross-validation* is a procedure that requires two separate samples; the discriminant function equation is developed for the first sample and then tested for accuracy in the second sample.

Logistic Regression Despite the popularity of multiple regression and discriminant analysis in psychological and psychiatric research, these models have known problems. Both multiple regression and discriminant analysis can be used to predict a binary outcome on the basis of several independent variables. However, when the outcome or the dependent variable has only two values, the statistical assumptions of multiple regression are often violated. One of the most important assumptions is that the distribution of errors is normal. Another assumption, often difficult to meet, is that the independent variables will follow multivariate normality. Logistic regression requires fewer assumptions and provides an excellent alternative for estimating the effect of several independent variables upon a dichotomous outcome. The dichotomous outcome might be a diagnostic category (schizophrenic–not schizophrenic) or a treatment response (responder–nonresponder).

Logistic regression techniques allow the estimation of the probability that an event occurs. The technique differs from multiple regression because it uses the maximum-likelihood method. Linear regression uses the principle of least squares in which regression coefficients are the result of the smallest squared distances between observed and predicted values. The maximum-likelihood method uses a rule that the coefficients are based on the "most likely" result. Logistic regression also allows the characterization of nonlinear relations.

The interpretation of logistic regression coefficients is different from that for multiple regression coefficients. In regression analysis the coefficient gives the amount of change in the dependent variable for each unit change in the independent variable. Coefficients in logistic regression describe the odds of an event occurring. The odds are the probability of an event divided by the probability of no event. Coefficients in logistic regression offer the log-odds or the logarithm-of-the-odds ratio.

Log-Linear Models A variety of multivariate techniques are available for categorical data. Logistic regression is a method for providing linear models of the log-odds of an event. Using these methods the probability of an event is represented as the linear combination of predictor variables. Earlier, the chi-square statistic was described as a method for evaluating the information in contingency tables. In recent years a variety of advanced techniques for analyzing contingency tables have developed. The major limitation of the chi-square test is that it does not provide a systematic evaluation of the relationship between variables. Essentially, chi-square is a test of nonindependence but it does not provide estimates of the relationship between variables. Further, the chi-square test becomes very complicated if more than two variables are in the analysis.

Log-linear analysis allows the construction of systematic models to evaluate the relationship between categorical variables. Complex models can be constructed and specific significance tests can be ordered; tests for hierarchical relationships can also be evaluated. For instance, the relation between employment status and alcohol abuse can be evaluated separately for those with and without a diagnoses of mood disorder.

Factor Analysis and Principal Components Analysis Discriminant analysis, multiple regression, and logistic regression analysis are techniques that find linear combinations of variables that maximize the prediction of some criterion. Factor analysis and principal components analysis are used to study the interrelationships among a set of variables without reference to a criterion. Factor analysis might best be thought of as a data reduction technique. With responses to a large number of items or a large number of tests, it is often desirable to reduce all this information down into more manageable chunks. The task in correlation is to find the best-fitting line through the points created by a two-dimensional scatter diagram. As more variables are added in multivariate analysis, the number of dimensions increases. For example, a three-dimensional plot is shown in Figure 5.2-9. Scatter diagrams for more than three dimensions can only be imagined. Consider that points are plotted in the space created in these many dimensions.

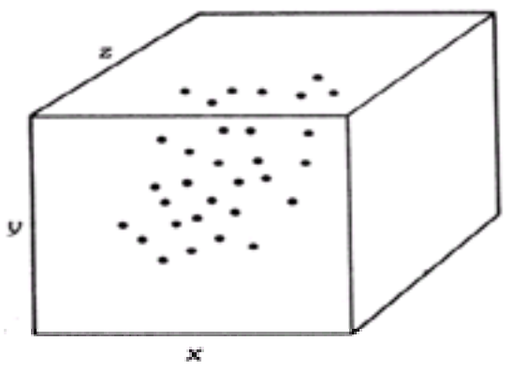

FIGURE 5.2-9 A three-dimensional scatter plot might be represented by this box. In addition to plotting points on the *X* and *Y* axes, they must be located with respect to a third *Z* axis. Although it is hard to show more than two dimensions on a flat page, a three-dimensional figure can be thought of as a box.

In principal components analysis, a matrix of correlations between every variable with every other variable is created. Then the linear combinations of the variables that describe as much of the interrelationships between the variables as possible are obtained. These linear combinations of the variables are called *principal components* and the goal of creating them is to describe as much of the association between the variables as possible. As many principal components as there are variables can be obtained; however, each principal component is extracted according to mathematical rules that make it independent or uncorrelated with all of the other principal components. The first component will be the most successful in describing the variation among the variables, and each succeeding component will be somewhat less successful. Typically, only a few components that account for larger proportions of the variation are extracted for further study.

Once the linear combinations or principal components have been found, the correlations between the original items and the factors are obtained. These correlations are called *factor loadings*. The expression "item 7 loaded highly on factor 1" means there was a high correlation between item 7 and the first principal component. By

examining which variables load highly on each factor, the factors come to be interpreted and named.

Principal components analysis and factor analysis are similar methods that differ in the matrix that is subjected to data reduction. In principal components analysis the initial correlation matrix is made up of raw correlations between variables. Thus, all values on the diagonal of the matrix have the value 1.0. In factor analysis, only the variance that each variable shares with other variables in the data set is used in the analysis. This common variance, known as *commonality*, is estimated using a variety of different methods. Different approaches to factor analysis use different commonality estimates. For example, the *principal factors method* estimates commonalties as the squared multiple correlations of each variable with all other variables. Thus, the diagonal of the initial correlation matrix includes values between 0 and 1.0 that are analogous to R^2 values for independent multiple regressions of each variable with each other variable.

Factor analysis is a complex and technical method, and there are many options the user must consider. For example, investigators frequently use methods that help them get a clearer picture of the meaning of the components by transforming the variables in a way that pushes the factor loadings toward the high or the low extreme. These transformation methods involve rotating the axes in the space created by the factors; they have therefore been labeled *methods of rotation*. These rotation methods can improve the scientific utility of the factor solution. There are several options for methods of rotation, and there are other options concerning the characteristics of the matrix that originally is entered into the analysis.

Survival Analysis Survival analysis evaluates the timing of events. In biomedical research survival analysis is typically used to evaluate life expectancy. However, a whole series of time-dependent questions can be evaluated using survival analysis techniques. For example, survival analysis can be used to evaluate age of onset of psychological illness, time to relapse for those in treatment, or the timing of developmental milestones such as first word, age of initiation of smoking, or age at marriage. Virtually any time-dependent variable can be analyzed such as duration of marriage or duration of employment.

Survival analysis requires the construction of a follow-up life table. Building the table requires a defined starting point. This could be birth, date of initiation of therapy, or date of diagnosis. Next, a follow-up interval, such as a year, is defined. The central focus of survival analysis is the time until a defined event occurs. For each time interval, survival analysis calculates the probability that the event has occurred. The following example illustrates this approach: Alcoholics had previously been reported to experience higher mortality, that is, greater proportions died at a given age compared to nonalcoholics. The alcoholics who achieved long-term sobriety had improved mortality experience compared to relapsed alcoholics. The investigators also wished to compare the mortality of the two groups of alcoholics to that of nonalcoholics. Figure 5.2-10 illustrates the results. The starting point for the analysis was date of entry into the study. The follow-up period ranged from 1 to 11 years. The numbers of deaths in each group is plotted as a function of time. Cumulatively, 19 of 101 relapsed alcoholics died, compared to 4 of 98 continuously abstinent alcoholics and 1 of 92 nonalcoholic controls. The 99 percent confidence intervals were 9.64 to 33.38 for the relapsers, 0.67 to 12.59 for the abstainers, and 0.01 to 7.43 for the nonalcoholic subjects. The confidence interval for relapsers does not overlap with that of the other two groups, so there was 99 percent confidence that these differences were not due to sampling error.

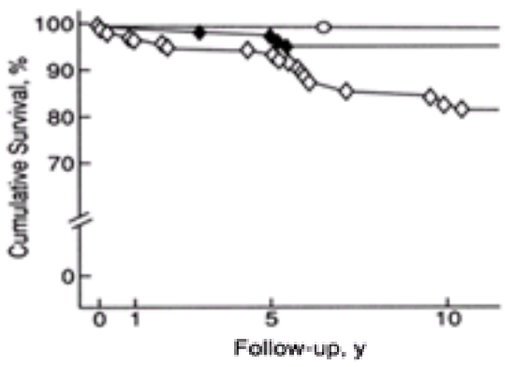

FIGURE 5.2-10 Life table survival experience for all subjects. Open diamonds (à) represent expired relapsed alcoholics (*N* = 19/101); solid diamonds (¨) represent expired abstinent alcoholics (*N* = 4/98); circles (○) represent expired nonalcoholic controls (*N* = 1/92).

Survival analysis can make comparisons between different defined groups. For example, smokers with coaddictions might be compared with those who use tobacco only. For each of these groups a *hazard rate* is created, defined as the estimated probability that an individual who has not experienced an event at the beginning of an interval will experience that event during the interval.

Survival analysis is a complex topic with many attractive features. For example, survival analysis allows the use of time-varying covariates, which means that the influence of a covariate may be different at different points in time. The effects of depression upon substance use, for example, may be stronger at some ages than at others and survival analysis allows the modeling of some of these relations. The example of alcoholic subjects considered factors such as relapse, lifetime alcohol use, neuropsychological status, and personality measures, but relapse was determined to be the only significant predictor. Unfortunately these techniques have received relatively little attention in the psychiatric literature.

EPIDEMIOLOGICAL MEASURES

A variety of epidemiological measures is commonly used in psychiatric research. Epidemiological studies typically focus on outcomes expressed as morbidity and mortality. Morbidity rates are divided into two major types: *incidence* and *prevalence. Incidence* refers to the rate at which new cases are occurring and is defined as the number of new cases that occur within a specific population within a defined time interval. Typically, the incidence rate is expressed per 1000 in the population. Conceptually, it is:

$$\text{Incidence Rate per 1,000} = \frac{\text{New Cases per Unit of Time}}{\text{Persons Exposed or at Risk per Unit Time} \times 1000}$$

Prevalence rates describe the number of diagnosed cases at a particular point in time.

$$\text{Prevalence Rate per 1,000} = \frac{\text{Cases at Specific Time Point}}{\text{Persons in Population at Specific Time Point} \times 1{,}000}$$

Prevalence rate is equal to the incidence rate times the duration of the disease. For example, if the average duration of dementia is 5 years and its incidence is 3 per 1000 per year, the prevalence would be 15 per 1000. Epidemiologists often make the distinction between *point prevalence* and *period prevalence. Point prevalence* refers to the number of cases at a defined time period, such as in the year 1998. *Period prevalence* is relevant to a time interval; it begins with an estimate of prevalence at the beginning of some time period and includes all new cases accumulated until the end of that defined period.

RESEARCH DESIGNS

Virtually all research designs common to clinical and experimental research are used in psychiatry and behavioral sciences.

Observational Studies Observational studies do not attempt to manipulate variables in a systematic fashion; instead, inferences are made on the basis of an ongoing series of observations. Some of the most common observational studies include the cohort study, the panel study, and the case-control study.

Cohort Study In a cohort study groups of people who share some common characteristics are followed over the course of time. These studies, which are often prospective, resample the same population of individuals on repeated occasions. However, the exact participants in the study may not be the same on repeated observations.

Panel Study A panel study is similar to a cohort study; however, it has the stricter requirement that exactly the same individuals who were in the original sample are followed at each repeated assessment.

Cohort and panel studies are considered to be *longitudinal* designs, which make inferences about changes over the course of time.

Cross-sectional studies differ from longitudinal studies in that they examine different groups of individuals at the same point in time. To make inferences about drug use in college, for example, the cross-sectional method would require sampling of each current class such that freshmen could be compared to sophomores, juniors, and seniors. These individuals would not be members of the same class or birth cohort.

Case-Control Study This methodology compares a group of people with a diagnosed disease (cases) with one or more groups that have not been given the same diagnosis. Case-control studies are typically retrospective because they make inferences about events that have caused currently diagnosed cases. Longitudinal studies are often prospective and have the advantage of documenting the antecedents of new cases.

Observational studies often used correlational and multivariate statistical techniques. Variables that are uncontrolled through the experimental design are often adjusted for using statistical methods. In contrast to observational studies in which important variables are not controlled, experimental studies typically involve the systematic manipulation of variables.

Experimental Studies Mental illnesses have a natural history and in most instances their course fluctuates considerably without treatment. One of the difficulties in determining the effects of an intervention is that the intervention may occur at time of crisis. For example, patients with mood disorder might seek psychiatric care on days when they are most depressed. If the exacerbation of the illness is self-limiting (as in the case of mood disorders), the patient's condition will improve spontaneously. A control group can help sort out the effect of treatment from other factors. In this way, a group receiving the intervention under study is compared with the control group to determine whether there are differences attributable to the intervention.

It is widely accepted among medical and biobehavioral scientists that a control or comparison group is required to establish causal inference. In some cases, investigators are willing to accept quasi-experimental data in which an ad hoc control is used, or where there is a stable baseline of observations prior to an intervention. However, several authors have argued that an experiment characterized by a single observation, an intervention, and a second observation is virtually impossible to interpret from a causal perspective.

For experiments using control groups, random assignment to treatment and control conditions is very desirable because randomized clinical trials remove several sources of bias. The value of randomized clinical trials has been emphasized in a series of review articles. In 1982 Harvey Sacks, Thomas Chalmers, and Henry Smith reviewed six therapies for which approximately equal numbers of randomized clinical trials and nonrandomized trials had been reported in the literature. They found that 79 percent of the studies in which patients were not randomly assigned to groups reported that the therapy was better than the control regimen. In contrast, the same therapies were found to be effective in only 20 percent of the studies in which patients had been randomly assigned to the treatment or control condition. In a related review, Chalmers and his colleagues analyzed 145 papers, which were divided into three categories: those in which randomization was blinded, those in which randomization was unblinded, and those in which assignment to treatment or control was by a nonrandom process. Review of these studies suggested that there was a systematic relationship between the rigor of the experimental design and the probability of finding a treatment benefit. There was a significant treatment benefit in 58 percent of the studies in which the subjects were not randomly assigned. The same benefit was observed in 24 percent of the unblinded randomized studies and in approximately 9 percent of the blinded randomized studies.

There are many sources of bias in studies that do not use control groups and the end result is frequently an overestimation of the effects of the therapy under study. These biases are reduced in experimental studies, but the rigor of the experimental design is systematically related to the chances of finding a treatment benefit. Valid scientific inferences must be built upon a solid experimental foundation.

ISSUES IN RESEARCH STUDIES

Reliability Attenuates Relations Reliability is the extent to which the test or measure is free of measurement error. Measurement error is the discrepancy between an observed score and the true value for a particular attribute. Under most circumstances the error is assumed to be random and independent of the true score. Reliability is estimated in a variety of different ways. If the measure is supposed to be stable over the course of time, such as a personality trait, reliability can be assessed by examining the correlation between scores on the same test for the same individuals when the test is administered at two different points in time. Other forms of reliability consider the internal consistency of a test. For example, a measure designed to tap depression may be composed of many items, each of which is an independent assessment of the general depressive attribute. A test is more reliable if responses to these independent items are correlated with one another. Measures of internal consistency, including the Kuder-Richardson 20 and coefficient alpha, are typically correlationlike indexes with ranges from 0 to 1.0. The nearer the reliability coefficient is to 1.0, the higher is the reliability of the test.

The definition of an acceptable level of reliability depends on the purpose of the test. It has been suggested that reliability estimates in the range of .70 to .80 are high enough for most purposes in basic research. In many research studies the investigator only needs an approximate estimate of whether two variables are correlated. If the result looks promising, it then may be worth the extra time and effort to make the research instruments more reliable. Increasing the reliability beyond .90 may increase expense and respondent burden.

In clinical settings, high reliability is extremely important. When tests are used to make important decisions about individual patients, it is essential that classification error be minimized. Thus, a test with a reliability of .9 might not be good enough.

A number of procedures are available to increase the reliability of a test. For example, increasing the number of items tends to increase the reliability of a measure. A prophecy formula is available that permits the estimation of the specific number of items that need to be added in order to achieve a defined level of reliability. A second strategy is to factor analyze the items, thereby obtaining homogeneous subsets of items. Selection of items from these homogeneous subsets increases the reliability of the test.

The effect of low reliability upon correlations has been well documented in the psychometric literature. Observed correlations between two variables are attenuated when either or both variables are measured with error. The expected observed correlation between two variables measured with error is defined as the true correlation times the square root of the product of their reliabilities or

$$r = tr\sqrt{r_{11}r_{22}}$$

where r is the expected observed correlation, tr is the expected true correlation, r_{11} is the reliability of the first measure, and r_{22} is the reliability of the second measure.

Consider the example of the association between a measure of life stress and a measure of social support. The Schedule of Recent Experiences (SRE) has an observed reliability of .55. The Arizona Social Support Interview Schedule has an observed reliability of .52. Suppose that the true correlation between these measures was a substantial .50. Because each measure contains measurement error, the observed correlation between the two measures would be .27. In a study with 50 participants ($r = .27$, $p > .05$), the investigator would fail to find a significant correlation between these two variables, even though there is indeed a substantial association.

Multiple Comparisons Multiple outcome measures are commonly used in psychiatric research. Investigators believe this is beneficial because psychiatric outcomes are complex and a multitude of measures are required to capture them. However, use of multiple measures creates other statistical biases. When multiple comparison are made, the probability of finding at least one difference by chance increases. At the .05 significance level, one significant difference is expected for each 20 independent tests (or 5 percent of all comparisons). Thus, a certain number of significant differences between groups are expected by chance alone. Multiple

comparisons problems occur under two circumstances. First, they may be common when the investigator is comparing multiple groups on the same outcome measure. The number of possible comparisons is equal to

$$j\,(j\,-\,1)/2$$

where *j* is the number of groups.

For example, with six groups the number of comparisons would be 15. This difficulty is avoided by using methods such as ANOVA, with appropriate follow-up tests such as the Neuman-Keuls test. These tests compare pairs of means using adjustments to protect against making a Type I error owing to multiple comparisons.

However, the problem of multiple outcome measures is more difficult. The probability of finding at least one significant difference by chance is defined as:

$$\text{Probability of 1 or more type I errors} = 1 - (1 - \alpha)^C$$

where *C* is the number of tests.

As the number of tests increases, the investigator can expect to find more spurious results. For example, the probability of finding at least one spurious statistical difference in five contrasts is .23. For 10 contrasts, the probability is .40 whereas for 20 tests the probability is .64. In other words, the chances of drawing the wrong conclusion about the null hypothesis can become quite high when multiple comparisons are performed.

There are several remedies for this situation, but the difficulties associated with multiple comparisons tests should always be kept in mind. In order to avoid false conclusions, some investigators adjust the significance level to be more conservative. This is the basis for the Bonferroni inequality, a common procedure to correct for multiple comparisons. Under this procedure, if a was originally set at .05, but 10 comparisons were performed, the adjusted probability of a Type I error would be a/*N* = .005, where *N* was the number of comparisons. However, such adjustments may be problematic because some of the adjustments are so conservative that the null hypothesis is rarely rejected (i.e., a type II error may be introduced). Another approach is to use multivariate techniques that take multiple comparisons into consideration or to reduce the data set to a smaller number of manageable dimensions. For example, if 20 neuropsychological tests were administered to 200 subjects, factor analysis might be used to reduce the neuropsychological domain to three or four factors.

Sample Size Issues Generally speaking, there are fewer biases in studies with large sample sizes in comparison with studies with small sample sizes. However, a large sample size does not necessarily ensure that the conclusions will be meaningful. Further, many studies with small sample sizes have appropriate internal validity. For studies attempting to estimate prevalence or incidence rates, representativeness is more important than sample size. In a famous case, the *Literary Digest* attempted to forecast the outcome of the 1936 presidential election between Franklin D. Roosevelt and Alfred Landon. The magazine drew its sample from its readers, from automobile registrations, and from telephone directories. In 1936, all these sources overrepresented the wealthy, most of them Republicans. The poll showed that Landon (the Republican) would win by a landslide. The results of the elections were, however, just the opposite. Roosevelt won by one of the greatest margins in American history. Thus, survey results are of little value if the sample is not representative of the population. In fact, there was no problem with the sample size in the *Literary Digest* poll; indeed, the sample size was very large. In contrast, election day polls using as few as 2000 respondents to represent all the voters in the United States have been repeatedly shown to be very accurate. Relatively small samples can be of great value if they are drawn in a random and representative fashion.

Most statistical tests take sample size into consideration. Thus, the probability of rejecting the null hypothesis by chance when the sample size is 10 is .05 if the .05 alpha level is used. The probability of rejecting the null hypothesis by chance in a sample size of 10,000 is also .05 at the .05 level. In other words, there is an inherent correction for sample size. It is commonly asserted that studies that reject the null hypothesis but have a small sample size are of little value. However, obtaining a significant difference with a small sample size often requires that the experimental effect is significantly stronger than obtaining the effect at the same alpha level with a larger sample. Thus, demonstrating treatment efficacy with a sample size of 10 per group might require that the treatment effect account for 30 percent or more of the variance in the outcome variable. Obtaining the same significance level with a large sample size (say 300 per group) might only require 1 or 2 percent of the variance to be accounted for.

Power Analysis and Sample Size Planning Many problems in the interpretation of experimental data can be avoided with appropriate planning. Statistical power defines the probability that a statistical test will produce a significant result. Biomedical research now requires sample size planning. In most research studies the investigator evaluates the null hypothesis—that is, the hypothesis of insufficient evidence for the hypothesis of interest. Typically the researcher does not want to accept the null hypothesis but prefers to reject the null hypothesis in favor of the alternative that the investigator hopes to support.

When statistical tests are nonsignificant, the investigator is typically unable to conclude that the null hypothesis is correct. Significant treatment effects can be missed because the experiment does not have sufficient power. For example, if too few subjects are in the study or if there is substantial variability, a potentially beneficial treatment might be overlooked.

In order to avoid making these type II errors, sample size planning is required. This typically requires a power calculation. Formulas for power calculations are available for most types of statistical tests. A common formula for an experiment that compares two groups is:

$$N = \frac{2s^2(Z_\alpha + Z_\beta)^2}{\Delta^2}$$

where *s* is the standard deviation of the outcome, Z_a is the *Z*-score associated with the probability of a type I error, Z_b is the *Z*-score associated with the power of the test, and D is the expected differences between the experimental and control groups.

If an investigator wants to study changes in neuropsychological functioning following the administration of a b-blocker medication. Patients will be randomly assigned to use the drug or to take a placebo. The neuropsychological test expresses outcomes in *T*-scores (mean = 50, standard deviation = 10), and the drug is expected to reduce performance in comparison to placebo by about 5 points. The investigator needs to estimate how many patients would be required to have a 90 percent chance of detecting a difference with the probability of a type I error set at .05.

The *Z*-score for the .05 significance level is 1.96 (Z_a) and the *Z*-score for a power of .90 (Z_b) is 1.28. To estimate this, the following calculations are made:

$$\begin{aligned} N &= \frac{2(10)^2(1.96 + 1.28)^2}{5^2} \\ &= \frac{20(3.24)^2}{5^2} \\ &= \frac{2100}{25} \\ &= 84 \end{aligned}$$

Thus, the experiment would require 84 subjects or 42 in each of the two groups.

Selective Biases A variety of biases are common to psychiatric research, some of which are identified as common flaws.

Selective Attrition Many studies involve follow-up of patients. However, patients may be available for follow-up on a selective basis. For example, it has been demonstrated that in treatment studies those who are available for follow-up are not a representative sample of the original population. It is common for those available for follow-up to be among those who succeeded in the treatment program. This can be a particularly severe problem in studies where the loss to follow-up is

different for treatment and control subjects.

Detection Bias A common problem on etiology studies is that those with the diagnosis may be examined in a different way than those who have not already been so diagnosed. If at all possible, it is valuable to blind the observers.

Ex Post Facto Design Problems One of the most common designs in biomedical research is the case-control study. In the research methodology literature, this is known as the *ex post facto design*. In such analyses the investigator already knows that the groups being compared differ in some respect. Typically, individuals who have a diagnosed disorder are compared to a matched group of individuals who have not been placed in the same diagnostic category. The matching occurs for a limited number of variables, and the investigator attempts to determine whether the two groups differ on various prior or current exposures to causal factors. The case-control study is also known as the *case-referent study* or the *case-comparison study* design. An example of this design might be a comparison between patients with a diagnosis of paranoid schizophrenia (cases) with people without schizophrenia (controls). Prior history for the two groups might be compared.

In contrast to true experiments, the case-control or ex post facto method has many deficiencies. In effect, the investigator clearly knows that the two groups differ on at least one variable. The observation that the groups differ on other variables implies relatively little about causation; in effect, these designs represent parallel correlational studies.

Often in case-control studies an investigator matches patients on a particular variable or uses the variable as a covariate or covariable. If the matching variable or covariate fails to change the difference between the cases and controls, it is assumed that the variable does not explain the underlying basis of the condition. However, there are many rival explanations for the failure of covariables to explain an observed relationship. One important explanation is that the covariates were measured with error or that the match was imperfect. To the extent that these problems occur, the effect of adjustment will be greatly attenuated. Thus, the fact that a covariable does not have an effect does not necessarily mean that the covariable is unimportant in the observed relationship.

Computer Applications The availability of computer software has greatly facilitated the execution of most statistical techniques. The many statistical packages run on different types of platforms or computer configurations. There are three main classes of platforms: mainframe, workstation, and microcomputer. Mainframe computers are now used less frequently for data analysis because many complex analyses can be achieved on less expensive work stations and microcomputers. The two most common types of microcomputers are IBM-compatible and Apple-Macintosh. Extensive computer software is now available for each of these platforms. For general data analysis the Statistical Package for the Social Sciences (SPSS), the BMDP series, and the Statistical Analysis System (SAS) are recommended. These are general-purpose statistical packages that perform essentially all the analyses common to biomedical research. In addition, a variety of other packages have emerged. SYSTAT runs on both IBM-compatible and Macintosh systems and performs most of the analyses commonly used in biomedical research. The popular SAS program has been redeveloped for Macintosh systems and is sold under the name *JMP*. Other commonly used programs include Stata, which is excellent for the IBM-compatible computers. The developers of Stata release a regular newsletter providing updates, which makes the package very attractive. StatView is a general-purpose program for the Macintosh computer. Newer versions of StatView include an additional program called Super ANOVA, which is an excellent set of ANOVA routines. StatView is user-friendly and also has superb graphics. For users interested in epidemiological analyses, Epilog is a relatively low-cost program that runs on the IBM-compatible platforms. It is particularly valuable for rate calculations, analysis of disease-clustering patterns, and survival analysis. GB-STAT, is a low-cost, multipurpose package that is very comprehensive.

Historically, a major concern of data analyzers was that each analysis program required its own data format. A software advance helps to alleviate this problem. A program called DBMS/Copy allows the user to move files among a number of spreadsheets, databases, and commonly used statistical packages. DBMS/Copy uses a form of artificial intelligence to create a directory of information about the data. Using this directory it can reformat the data so they can be used in most of the commonly available programs. DBMS/Copy is now distributed SPSS, Inc. of Chicago, Illinois. A summary of some available computer programs is offered in Table 5.2-6.

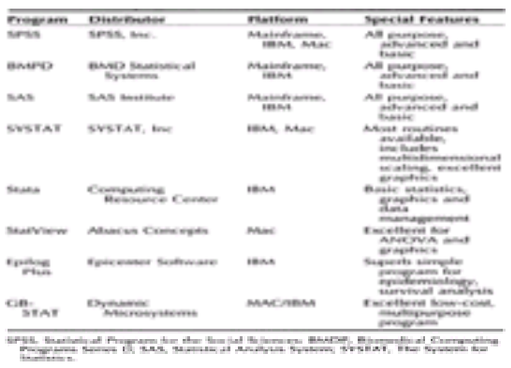

Program	Distributor	Platform	Special Features
SPSS	SPSS, Inc.	Mainframe, IBM, Mac	All purpose, advanced and basic
BMPD	BMD Statistical Systems	Mainframe, IBM	All purpose, advanced and basic
SAS	SAS Institute	Mainframe, IBM	All purpose, advanced and basic
SYSTAT	SYSTAT, Inc	IBM, Mac	Most routines available, includes multidimensional scaling, excellent graphics
Stata	Computing Resource Center	IBM	Basic statistics, graphics and data management
StatView	Abacus Concepts	Mac	Excellent for ANOVA and graphics
Epilog Plus	Epicenter Software	IBM	Superb simple program for epidemiology, survival analysis
GB-STAT	Dynamic Microsystems	MAC/IBM	Excellent low-cost, multipurpose program

SPSS, Statistical Program for the Social Sciences; BMDP, Biomedical Computing Programs, Series D; SAS, Statistical Analysis System; SYSTAT, The System for Statistics.

Table 5.2-6 Summary of Common Statistical Software

Computer software programs that provide easy access to highly sophisticated statistical methodologies represent both opportunities and dangers. On the positive side, no serious researcher need be concerned about being unable to utilize precisely the statistical technique that best suits his or her purpose, and to do so with the kind of speed and economy that was inconceivable just two decades ago. The danger is that some investigators may be tempted to employ after-the-fact statistical manipulations to salvage a study that was flawed to start with, or to extract significant findings through use of progressively more sophisticated multivariate techniques. Such ex post facto ransacking of data bases does not advance knowledge. Progress in psychiatry will depend increasingly on careful crafting of hypotheses, experimental design, and appropriate statistical modeling. A brief definition of commonly used statistical terms is listed in Table 5.2-7.

Table 5.2-7 Glossary of Statistical Terms

SUGGESTED CROSS-REFERENCES

Section 5.1 discusses epidemiology, and Section 5.3 discusses mental health services research.

SECTION REFERENCES

Bullock KD, Reed RJ, Grant I: Reduced mortality risk in alcoholics who achieve long-term abstinence. JAMA *267*(5):668, 1992.

Chalmers TC, Celano P, Sacks H, Smith H: Bias in treatment assignment in controlled clinical trials. N Engl J Med *309:*1358, 1983.

*Cohen J: *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum, Hillsdale, NJ, 1988.

Cook TD, Campbell DG: *Quasi-experimentation: Design and Analysis Issues for Field Studies*. Rand-McNally, Chicago, 1979.

Daniel WW: *Biostatistics: A Foundation for Analysis in the Health Sciences*, ed 6. Wiley, New York, 1995.

Daniel WW: *Applied Nonparametric Statistics*, ed 2. PWS-Kent, Boston, 1990.

Dawson-Saunders B, Trapp RG: *Basic and Clinical Biostatistics*, ed 2. Appleton & Lange, Norwalk, CT, 1994.

*Edwards LK, editor: *Applied Analysis of Variance in Behavioral Science.* Marcel Dekker, New York, 1993.

*Efron B, Tibshirani R: Statistical data analysis in the computer age. Science *253:*390, 1991.

*George D, Mallery P: *SPSS(R) for Windows(R) Step by Step: A Simple Guide and Reference*. Allyn & Bacon, Boston, 1999.

Hays WL: *Statistics*, ed 5. Fort Worth, TX, Harcourt Brace, 1994.

Ingelfinger JA, Mosteller F, Thibodeau LA, Ware JH: *Biostatistics in Clinical Medicine*, ed 3. McGraw-Hill, New York, 1994.

Jaccard J, Becker MA: *Statistics for the Behavioral Sciences*, ed 3. Brooks/Cole Publishing Co, Pacific Grove, CA, 1997.

Jessor R, Jessor LJ: *Problem Behavior in Psychosocial Development*. Academic Press, New York, 1977.

Kaplan RM, Sacuzzo DP: *Psychological Testing: Principles, Applications, and Issues*, ed 4. Brooks/Cole, Monterey, CA, 1997.

Keppel G. *Design and Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1991.

Keselman HJ, Keselman JC: Analysis of repeated measurements. In *Applied Analysis of Variance in Behavioral Science*. In *Statistics: Textbooks and Monographs*, vol 137, LK Edwards, editor. Marcel Dekker, New York, 1993.

*McCall R: *Fundamental Statistics for Psychology*, ed 6. Harcourt Brace, & Jovanovich, New York, 1994.

Pett MA: *Nonparametric Statistics for Health Care Research: Statistics for Small Samples and Unusual Distributions*. Sage Publications, Thousand Oaks, CA, 1997.

Sacks H, Chalmers DC, Smith H: Randomized versus historical controls for clinical trials. Am J Med *72:*233, 1982.

*Tabachnick BG, Fidell LS: *Using Multivariate Statistics*, ed 3. Harper-Collins, New York, 1994.

Van de Geer JP: *Introduction to Multivariate Analysis for the Social Sciences*. San WH Freeman, San Francisco, 1971.

*Vogt WP: *Dictionary Statistics and Methodology: A Nontechnical Guide for the Social Sciences*, ed 2. Sage, Thousand Oaks, CA, 1999.

*Ware ME, Brewer CL, editors: *Handbook for Teaching Statistics and Research Methods*, ed 2. Erlbaum, Mahwah, NJ, 1999.

5.3 MENTAL HEALTH SERVICES RESEARCH

MARK OLFSON, M.D., M.P.H.

Mental health services research is a broad multidisciplinary field that draws on a variety of methods and approaches to describe and help improve the operation of the mental health services system. Mental health services researchers are interested in defining not only who receives, provides, and pays for care, but also in the combinations of services that are most effective and the costs of these services.

Much remains to be learned about the delivery and effectiveness of mental health services. There are methodological difficulties in defining and measuring the wide variety of financial arrangements, service interventions, and patient outcomes that characterize routine clinical practice. Few randomized controlled effectiveness studies have been conducted on the various emerging forms of service delivery. In several areas, however, experimental and quasi-experimental research provides a rational basis for improving service delivery through public mental health promotion, patient and clinician educational programs, program development, and health care system and policy reform.

Services research has increased the awareness of unmet service needs by revealing that a large number of individuals with mental disorders receive no treatment. Services research has also increased awareness of the factors that govern the flow of individuals into treatment, the pivotal role of primary care providers in mental health care delivery, and the complex treatment needs of severely ill individuals.

In mental health services outcomes research, a tension often exists between “rigor” and “relevance.” Outcome studies that provide the strongest and most scientifically credible findings are often the most difficult to apply to general psychiatric practice where patients present with complex diagnostic problems and treatment needs. However, outcomes research conducted in more naturalistic settings generally yields more readily applicable but uncertain findings. Clinicians face the critical challenge of integrating findings from the full spectrum of psychiatric outcomes research into the ongoing delivery of clinical care.

No well-accepted system exists for classifying the wide array of issues that preoccupy mental health services researchers. This chapter is organized around four general areas: (1) service delivery research, (2) service utilization research, (3) financing of services, and (4) quality of care and outcomes research. Mental health services research activities often span two or more of these areas.

SERVICE DELIVERY RESEARCH

Service delivery research involves the collection, analysis, and interpretation of information from providers concerning the types of services they provide and the patients they treat. Such research uses several approaches, ranging from large-scale surveys of mental health care organizations to detailed studies of individual clinical encounters.

National Perspective A national perspective on mental health service delivery helps to define general trends in health care delivery. Nationally representative service delivery data provide insight into large shifts in practice patterns such as the ongoing decline in public mental hospital-based care, the increasing privatization of mental health services, and the rise of partial care services and other alternatives to inpatient treatment.

In the United States the most extensive national provider-based surveys of mental health services are conducted by the Substance Abuse and Mental Health Services Administration (SAMHSA) National Reporting Program (NRP). This program periodically surveys all the admissions and discharges from state and county mental hospitals. Samples are also drawn from general hospitals, private psychiatric hospitals, and other organizations that provide mental health services. These surveys provide estimates of the volume, diagnostic composition, demographic characteristics, and payment profile of patients treated in different types of mental health organizations.

NRP surveys provide a bird's-eye view of temporal changes in the pattern of service delivery. The findings also allow policy makers and program planners to examine the pattern of mental health services received by special patient groups (e.g., children, adolescents, and persons with severe and persistent mental illness) at a specific point in time.

The NRP does not capture mental health services provided in office-based practice or general medical settings. Some nationally representative mental health services data concerning these settings and providers can be gleaned by surveys conducted by the National Center for Health Statistics.

Resource utilization surveys conducted by the National Center for Health Statistics, together with population-based psychiatric epidemiological studies, have helped to define the extent to which general medical physicians provide mental health services. It is estimated that up to half of all the recipients of professional mental health care in the United States receive it in the general medical sector. This fact has focused attention on the quality of mental health services provided by primary care physicians and other nonpsychiatrist physicians, particularly on how they prescribe psychotropic medications.

Regional Perspective A regional perspective on service delivery emphasizes the composition, range, frequency, and coordination of mental health services provided to patients in a defined region such as a state, county, or catchment area. There is particular interest in measuring the extent to which specific changes in health care policy affect service delivery in a well-defined region.

If a change in policy is planned in advance, such as the scheduled closing of a state mental hospital, data collection can begin before the policy change occurs and can continue after the change in policy. In other cases retrospective designs that rely on data that was originally collected for some other purpose are used. For example, in 1987 New York state enacted triplicate benzodiazepine prescription regulations. Analysis of routinely collected prescription data clearly reveal that this policy was followed by a steep decline in the number of benzodiazepine prescriptions filled in the state.

This study is an example of an interrupted time series design, in which observations are made of a study population before and after the occurrence of a discrete change in the health care environment. Secular change, events that occur separate from the policy change under study, may confound attribution of temporal trends to the policy change.

Regression toward the mean is a second potential limitation of time series studies of policy change. Regression toward the mean is the tendency for observations of populations selected on the basis of exceeding some predetermined threshold level to return toward the mean on subsequent observations. When policies apply to specific patient groups, such as patients with high recent utilization of services, care must be taken to control or adjust for the natural tendency of these patients to return toward mean values of this variable (i.e., lower service utilization).

Policy makers and program planners are particularly interested in whether changes in policy have the intended effects on service delivery. Sections of the Omnibus Budget Reconciliation Act of 1987, for example, were intended to reduce the unnecessary use of antipsychotic drugs in nursing homes. Longitudinal analysis of prescription patterns demonstrates that a substantial decrease in antipsychotic drug use coincided with the implementation of these new federal regulations and provide evidence that the regulations had their intended effect. Detailed clinical studies are needed to evaluate whether this decrease in prescriptions for antipsychotic medications resulted in improved quality of care.

New policies may have the intended effects on service delivery, but these effects may not be sufficient to result in meaningful change in clinical outcomes. An

illustration of the importance of studying patient outcomes together with service delivery variables is provided by the Robert Wood Johnson Foundation Program on Mental Illness. This project examined the hypothesis that strengthening the role of local mental health authorities would make it possible to promote the development of comprehensive mental health care and social welfare services. Greater coordination in service delivery would in turn improve the health outcomes of patients with severe and persistent psychiatric illness. Nine cities were selected for the project. Although the project succeeded in creating structural change through the creation of centralized mental health authorities, these changes were not sufficient to promote improved patient outcomes in such crucial areas as symptomatic distress or quality of life. These findings suggest that structural change in service delivery without a complementary focus on clinical and social service processes may not be sufficient to achieve meaningful improvement in patient outcomes.

The unintended effects of policy change have also been studied. An example of unintended policy effects followed the enactment of state legislation in New Hampshire that limited Medicaid pharmacy reimbursement to three prescriptions per month. This policy sought to lower Medicaid expenditures by reducing discretionary medication costs. Analysis of clinical records revealed that the reimbursement cap led to a prompt decline in the use and costs of psychotropic drugs. However, the three-prescription cap also led to a sharp increase in the use of initial short-term mental health services. The cost of the increase in such mental health services more than offset the savings associated with the decreased use of psychotropic medications. After the reimbursement cap was repealed, use of medications and most mental health services returned to baseline levels.

Program Perspective A programmatic perspective of service delivery considers patterns of service provision from the vantage point of individual mental health programs. Researchers want to learn about the nature, frequency, and range of specific treatments provided by different programs. In some studies it is possible to link observed differences in program service delivery with patient outcomes. Research on assertive public (community) treatment programs offers an example of program level service delivery research.

Assertive public (community) treatment delivers psychiatric, social, and rehabilitative services to severely ill patients through multidisciplinary treatment teams that meet with patients in the community. There is considerable variation in the extent to which individual assertive public treatment programs implement this treatment model; this variation provides an opportunity to examine correlations between program design and patient outcome.

In one study of 18 assertive public treatment programs, total number of patient contacts, daily team meetings, and 24-hour availability were significantly correlated with reductions in number of days of patient hospitalization. Other service elements, including low client-to-staff ratio, small team size, and number of community contacts were not correlated with reduced hospital days. In a second study of three programs, reductions in rehospitalizations were achieved at two of the three programs. The program that did not reduce hospitalizations was distinguished from the other two programs by not having shared case loads, a team nurse, or daily team meetings and by not offering office visits. These observations that link specific service structures to patient outcomes help to inform program planning and development.

Patient Encounter Perspective The most microscopic service delivery research perspective examines specific aspects of individual clinical encounters. An important goal of such research is to define clinician behaviors associated with either positive or negative health care outcomes. There is a long tradition in psychotherapy research of seeking to associate specific therapist behaviors with clinical outcomes.

A related area that has received considerable attention in this regard is the assessment and diagnosis of mental disorders by primary care clinicians. Scales exist to score videotaped patient visits for clinical techniques and behaviors that may either facilitate or impede the recognition of mental disorders by primary care physicians. One study found that only a small proportion of primary care visits with depressed patients include the necessary questions to establish a diagnosis of depressive disorder. Another study found that accurate recognition of the mental disorder was increased by establishing eye contact with the patient at the outset of the interview, clarifying the presenting complaint, responding to verbal and nonverbal cues of psychological distress, and having the capacity to elicit cues from the patient that indicate distress. The results of patient encounter studies can help inform educational efforts to improve the delivery of mental health services.

SERVICE UTILIZATION RESEARCH

Health Care Seeking Service utilization research considers service provision from the perspective of the patient rather than the provider. Service utilization research emphasizes elucidating the range of factors that influence an individual's entry and use of mental health services. One important area in service utilization research concerns health–care-seeking behavior.

It is well established that most people in need of mental health care receive either inadequate professional services or none. This observation has focused attention on the processes that govern the flow of individuals into treatment and the course of patients between different levels of care. Researchers seek to identify and describe factors that either facilitate or impede the use of mental health services by persons with well-defined disorders. An understanding of these processes informs public health efforts aimed at extending treatment access to groups with high rates of unmet need for clinical care.

The main steps in the process of seeking mental health care are recognition of a mental health problem, contact with a general medical provider, and referral to a mental health specialist. In the United States these steps do not necessarily proceed in a orderly or sequential manner because many individuals bypass general medical care and present directly for specialized mental health care, sometimes on an emergency basis.

Individual determinants of the process of seeking health care are commonly conceptualized as involving interactions between health needs, predisposing factors, and enabling resources. Health needs include objective and subjective aspects of mental and physical health status. The assessment of health needs typically includes measures of the type, severity, frequency, and duration of symptoms or distress. Subjective appraisal of mental health is another important dimension of the health assessment.

Predisposing factors relate to willingness to seek care. Norms and beliefs, role obligations, and demographic characteristics are important predisposing factors. Enabling or restrictive influences, on the other hand, reflect the availability of resources that facilitate or impede service utilization. Health insurance, disposable income, and local availability of services are key enabling influences. The operation of predisposing and enabling factors is generally considered to be contingent upon the presence of some underlying clinical need for care.

Beyond individual determinants of health care seeking, health care utilization patterns are determined by provider variables such as the skill, training, attitudes, orientation, knowledge, and behavior of the health care professional; administrative controls over patient access to service are increasingly becoming another important determinant of health care utilization.

Epidemiological data indicate that greater objective illness severity is associated with higher rates of treatment. Among those who receive mental health care, individuals with more severe conditions are disproportionately served by psychiatrists and other mental health specialists as opposed to primary care physicians or other general medical professionals.

Specific aspects of psychopathology influence treatment-seeking behavior. Symptom pattern and severity play important roles in determining the timing of treatment seeking and the pathway to mental health care. In one community study of depression, depressed mood and weight loss, but not sadness or agitation, predicted mental health care from a primary care physician. In another study, chest pain, but not palpitations, during a panic attack was associated with seeking mental health treatment from a primary care physician.

Sociodemographic characteristics have been extensively studied as predisposing or restricting factors in mental health care seeking behavior. Unmet need for mental health service tends to be greatest for children, adolescents, and elderly persons as well as socioeconomically disadvantaged groups including the poor, the uninsured, blacks, Hispanics, and adults with lower levels of formal education. Although women tend to be more likely to use mental health services than men, this effect may be limited to care provided by primary care physicians and other nonmental health specialists.

The basis for the observed lower rates of mental health treatment among racial and ethnic minorities remains an important and poorly understood aspect of service utilization. Although some researchers have suggested that inter-ethnic differences in mental health utilization are largely attributable to socioeconomic differences, research demonstrates that even within privately and publicly insured populations, blacks and Hispanics have substantially lower rates of outpatient mental health service utilization than whites.

Cultural and attitudinal factors may help to explain the lower rates of professional mental health service utilization by blacks and Hispanics in relation to whites. Some

members of ethnic minorities may be reluctant to use specialized services provided by English-speaking whites, whereas others may prefer to rely on traditional, nonmedical care providers. Some racial minorities, for example, have been found to have a stronger belief than whites in the clergy's effectiveness to provide mental health care. Much remains to be learned about the role of race, ethnicity, and culture in the mental health care seeking process.

Recent research suggests that delay in health-care seeking is common and that a majority of people with the common major mental disorders eventially make treatment contact. People with early-onset psychiatric disorders may be at an especially high risk of long delays to first treatment contact.

Long delays in the initiation of treatment are especially common in first-episode schizophrenia and related disorders. Accumulating evidence suggests that failure to treat schizophrenia early in its course adversely affects long-term outcome. Several studies suggest that the duration of untreated psychotic disorders is roughly twice as long for men than women.

Treatment Noncompliance Treatment noncompliance poses a major challenge to the delivery of mental health services. High rates of treatment noncompliance have been observed in a variety of service settings. Following hospitalization, up to one third of psychiatric patients fail to make their first scheduled outpatient visit. High rates of noncompliance have also been found following referral from emergency rooms and primary care practices to outpatient psychiatric services. The literature suggests that across a variety of treatment contexts that active substance use, lack of insight into the need for treatment, and a history of treatment noncompliance—all predispose toward the premature termination of treatment.

A failure to take antipsychotic drugs as prescribed is one of the most important public health aspects of noncompliance with mental health services. In one study almost half of outpatients with schizophrenia stopped taking antipsychotic medications in the first year following hospital discharge. By 2 years following discharge, approximately three out of four outpatients had stopped taking antipsychotic medications. The clinical and public health implications of such widespread noncompliance are underscored by the high rates of relapse following experimental withdrawal of antipsychotic medications and by the close association of noncompliance with antipsychotic medication regimens and rehospitalization as determined by observational studies.

Compliance with mental health treatment is a joint function of the patient's personal characteristics, psychiatric symptoms, prescribed treatment, social context, and access to relevant services. Patient characteristics including demographic, financial, cultural, social, geographic, and attitudinal factors have been shown to influence treatment compliance. In addition, psychiatric symptoms such as pessimism, hopelessness, grandiosity, paranoia, and cognitive impairment may interfere with appropriate adherence to prescribed treatments. In one study of schizophrenia, for example, greater conceptual disorganization, emotional withdrawal, hostility, and grandiosity, but not anxiety or depression, were associated with rejection of antipsychotic medications.

Aspects of the treatment intervention, such as its effectiveness, adverse effects, complexity, and the patient-provider relationship have all been found to influence treatment compliance. Furthermore, characteristics of the service system influence treatment compliance. An increased risk of noncompliance or drop-out has been related to long clinic scheduling delays, restrictive admission and discharge policies, and a lack of access to outreach or emergency services.

Inappropriate Service Utilization Recent growth of prepaid and capitated care has increased efforts to control the costs of care and particularly to reduce care that is deemed unnecessary or inappropriate. This area is marked by considerable controversy as managed-care organizations, third-party payers, service providers, and symptomatic individuals often have very different perceptions of what constitutes inappropriate and appropriate care.

Primary care patients who use very high levels of general medical services have been studied from the standpoint of inappropriate service utilization and accumulating evidence suggests that these patients also have high rates of psychiatric disorder. In one study of distressed high utilizers of primary care, nearly one quarter of the patients met the criteria for current major depression.

It has been hypothesized that appropriate provision of psychiatric care to distressed high utilizers of general medical care will save treatment costs by reducing unnecessary and poorly focused medical expenditures. Thus far, this effect has not been convincingly demonstrated in a well-controlled study. However, a cost-offset effect has been shown following the provision of psychiatric consultation services to inpatients with general medical disorders and medical outpatients with somatization disorder. In addition, a psychiatric consultation–primary care physician collaboration model has been shown to improve the treatment of primary patients with depression.

FINANCING OF MENTAL HEALTH SERVICES

Over the past several years private-sector employers and state Medicaid agencies have increasingly enrolled in companies that provide managed mental health services. The wide appeal of these companies is that they promise to increase the availability of mental health services, contain or reduce costs, and improve the quality of care by integrating the financing and delivery of treatment.

There is general agreement that managed care has been effective in reducing or at least slowing the growth of the costs of providing mental health services. However, there is much less agreement over the effects of managed care on the quality of mental health services. Critics assert that managed care has jeopardized the quality of mental health services by underfunding services, excluding patients with more severe disorders, threatening confidentiality, and relying too extensively on less highly trained health professionals.

The debate surrounding the growth of managed mental health care has been conducted in the near absence of systematic empirical data on its actual clinical effects. Most of the empiric research on managed care has focused on cost-containment issues from the perspective of the payer, insurer, or managed-care organization. Considerably less work has been done on the effects of managed behavioral health care on the quality and effectiveness of clinical services.

In one study Medicaid patients with long-term mental illness were randomly assigned to prepaid care and followed for 1 year. All patients rejoined fee-for-service arrangements following the demonstration project. As expected, patients assigned to prepaid care had lower rates of service utilization in several areas but there were no significant differences between the groups in clinical outcomes. In multivariate analyses, however, patients with schizophrenia in the prepaid group had significantly poorer global functioning once they returned to fee-for-service care. Some indication of the longer-term effects of Medicaid managed care can be gleaned from retrospective analyses of claims data. This research suggests that Medicaid managed care for severely ill psychiatric patients is associated with a slight decreases in the rate and length of inpatient admissions, but an increase in the rate of rapid rehospitalization following hospital discharge.

Insurers and managed-care plans rely on several techniques to control access to care as well as the types, amounts, and costs of care. These include authorizing only selected clinicians to provide care, reviewing decisions concerning service utilization, closely monitoring high-cost cases, and the use of financial incentives. These cost-containment mechanisms are thought to have their greatest impact on patients who are the most severely ill and are therefore the most vulnerable to service reductions. An important challenge is to examine the effects of various evolving managed-care mechanisms on the quality of care and health outcomes of individuals with the most severe mental illnesses.

QUALITY OF CARE AND OUTCOMES RESEARCH

Quality of Care Research *Quality of care* is the degree to which treatment conforms with predefined standards of appropriate or acceptable care. Research on quality of care requires standards of care or treatment guidelines that can be operationally defined and reliably measured. Quality standards are typically developed by experimental or quasi-experimental observation, systematic review of the research literature, or expert consensus.

A distinguishing characteristic of quality-of-care outcomes research is that it does not involve experimental manipulation of service delivery. Instead, quality-of-care research studies the degree, extent, and consequences of naturally occurring variation in the organization, financing, and delivery of mental health services as it relates to adherence or departure from established quality standards. An important consequence of the observational nature of quality-of-care research is that the results readily generalize to broader practice. Three levels of mental health quality-of-care research are briefly described and illustrated.

Practice Variation At the most basic level, quality-of-care research involves documenting the extent to which treatment conforms with or deviates from accepted standards of care. In this type of research investigators observe variations in clinical practice and rely on existing treatment outcome studies to evaluate the clinical significance of the observed practice variations. The results highlight areas of concern in the clinical delivery of care. How physicians prescribe psychotropic

medications is particularly amenable to this approach to quality-of-care research.

The effects of aging on the metabolism of benzodiazepines are well known. Slower hepatic biotransformation and increased volume of distribution prolongs the half-life of long-acting benzodiazepines so drugs with shorter half-lives are recommended for the pharmacological treatment of insomnia in older patients. In one study the prescribing physicians showed little preference or none for shorter-acting agents in elderly patients which suggests that habit rather than sound pharmacokinetic principles may govern the selection of benzodiazepine medications. The findings of this study emphasize the need for targeted physician educational interventions.

Automated pharmacy records have been used to examine and evaluate patterns of psychotropic prescription in relation to established standards. A recent study of lithium (Eskalith) prescribing illustrates this analytic approach. Persuasive evidence exists that lithium has prophylactic mood-stabilizing properties in patients with bipolar I disorder. In one health maintenance organization it was estimated that the median period of continuous lithium use was less than 3 months, far shorter than the recommended duration of mood stabilization prophylaxis recommended by treatment guidelines and experts in the field. The results of this study suggest that detailed clinical studies are needed to evaluate the clinical consequences of this pattern of prescribing.

Practice Variation and Service Structure A second level of quality-of-care research attempts to link variations in processes of care with specific service structures. Service structures include physical facilities, providers, organization of services, and financing mechanisms. In this type of research, investigators seek to uncover clinically significant variations in practice and attempt to explain this practice variation based on some aspect of service structure.

An illustration of linking practice variation to service structure is provided by an analysis of discharge planning activities from a national survey of mental health facilities. It is well established that psychiatric inpatients who receive adequate discharge planning and transitional services are more likely to use outpatient mental health care services and are less likely to require rehospitalization in the months following hospital discharge.

This survey revealed marked unevenness in the availability of transitional-care services. Specifically, lower rates of telephone follow-up, staff assistance with referrals, visits to outpatient providers, and other crucial services were found at general hospitals as compared with specialized psychiatric hospitals. These differences persisted after controlling for case mix, hospital ownership, and other factors. Structural or administrative differences between the psychiatric hospitals and general hospitals might account for the differences in transitional-service availability.

Provider training is an important aspect of service structure and is increasingly becoming a subject of quality-of-care research. In a British study that compared lithium therapeutic drug-monitoring practices of general practitioners and psychiatrists, general practitioners less frequently obtained serum lithium levels from their patients and these patients were more likely to experience lithium levels above the therapeutic range. One third of the general practitioners made no change in the dosing regimen during the 6-week period following high serum lithium levels. These findings point to an area of clinical concern and suggest the need for targeted provider educational interventions.

Practice Variation, Service Structure, and Outcomes The most ambitious quality-of-care studies seek to link structures and processes of care with patient outcomes. Such research seeks to determine the clinical significance of identified practice variation by demonstrating an association between specific aspects of service delivery with patient outcomes.

Such quality-of-care studies resemble traditional epidemiological cohort studies. A well-defined group of patients is identified, receives a differential exposure to key treatment processes followed over time, and is examined for differences in outcome. This type of research tends to be expensive to conduct because very large sample sizes are required to capture sufficient variation in service delivery.

Confounding is an important limitation of research that seeks to link practice variation, service structures, and patient outcomes. It occurs when a variable other than the service and outcome variables under study is independently related to the service and outcome variables. Confounding can create an apparent association between service delivery and outcomes or mask an association that actually exists. For example, a naturalistic study of case management that fails to adjust adequately for initial patient illness severity might spuriously conclude that case management leads to poorer outcomes when in fact this observation may simply be a consequence of more severely ill patients being selectively assigned case managers.

The well-known Medical Outcomes Study provides an illustration of a health services cohort study. The Medical Outcomes Study was designed to compare the naturally occurring processes and outcomes of care for patients with depressive disorders and other chronic illnesses under fee-for-service and various prepaid service arrangements. Over 600 depressed patients of psychiatrists, psychologists, other therapists, and general medical clinicians were followed for 2 years at three urban sites. Among psychiatrists' patients with depression, those who initially received prepaid care acquired significantly more functional limitations than those who received fee-for-service care. Prepaid-plan patients of psychiatrists had fewer mental health visits than their fee-for-service counterparts and experienced a sharper reduction in antidepressant medication use over time.

These findings link treatment differences through financing arrangements to patient outcomes. Such data are valuable in guiding efforts to improve the quality of treatment of depression, particularly in the types of prepaid settings included in this study.

Experimental Outcomes Research Randomized clinical trials remain the most internally valid experimental design from which causal inferences may be drawn. Prospective randomized designs are used to study a range of mental health interventions and treatments. These investigations exist along a continuum from pure efficacy studies to applied effectiveness research. The continuum spans several related dimensions: stringency of patient selection, complexity of the interventions, and breadth of the outcome measures.

Efficacy studies typically select a narrow band of patients who possess characteristics that make them particularly responsive to the treatment and who lack characteristics that limit their responsiveness. The treatments studied in efficacy research tend to be relatively simple, time-limited, and capable of being administered in a highly reliable fashion (e.g., psychotropic medications). Outcome assessment in efficacy research tends to be narrowly focused on specific target symptoms.

Effectiveness research typically employs far less restrictive patient selection criteria and has an overriding interest in studying patients who are representative of broader patient populations. The interventions tend to be more complex and are studied for longer periods of time. Given the complex nature of the interventions studied in effectiveness research, it may be difficult to ascertain which aspects of the intervention are responsible for the observed outcomes. In effectiveness research, outcome assessments are typically broad and may extend beyond clinical symptoms to include social functioning, social relations, treatment and societal costs, work and school performance, family burden, quality of life, satisfaction with care, and other aspects of well-being.

Mental health services that seek to reduce use of acute inpatient psychiatric care have been extensively studied in randomized controlled effectiveness studies. Research has been conducted on the effectiveness of home care, partial hospital care, holding beds, and residential alternatives to inpatient treatment. This literature indicates that for many patients who were previously treated in inpatient settings, less restrictive settings are often feasible alternatives.

A variety of innovative models of continuous mental health service delivery have also been studied in randomized controlled effectiveness studies. The most well studied such model is assertive community treatment. This intervention involves using an interdisciplinary treatment team to provide a broad, individualized, aggressive, and continuous mix of outpatient rehabilitative and psychiatric services. Assertive community treatment staff assume ultimate responsibility for ensuring that patients receive food, shelter, clothing, and medical services and they help patients locate appropriate work and develop recreational interests.

Outcomes research indicates that assertive community treatment consistently increases patient satisfaction with care, usually reduces use of inpatient services, and sometimes reduces clinical symptoms and improves role functioning in relation to traditional outpatient care. The selective use of assertive community treatment has also been associated with savings in total costs. The impressive research findings from assertive community treatment are widely credited for contributing to the dissemination of this treatment model throughout the United States.

SUGGESTED CROSS-REFERENCES

Medicare is discussed in Section 51.5a and managed care is discussed in 51.5b and 52.2. Primary care psychiatry is covered in 28.1, public psychiatry is covered in Section 52.1; emergency psychiatry is covered in Chapter 29; and consultation-liaison psychiatry is covered in Section 25.12. Section 5.1 discusses epidemiology in psychiatry and Section 5.2 discusses statistics. Chapter 11 presents substance-related disorders, and Chapter 12 presents schizophrenia. Discussion of various classes of antipsychotics are found in Section 12.8, Section 31.17, Section 31.26, and Section 51.4e, and benzodiazepines are covered in Section 15.7, Section

31.10, and Section 51.4d; lithium is discussed in Section 14.7 and Section 31.18 .

SECTION REFERENCES

Badger LW, DeGruy FV, Hartman J, Plant MA, Leeper J, Anderson R, Tietze R: Patient presentation, interview content, and the detection of depression by primary care physicians. Psychosom Med *56:*128, 1994.

*Burns BJ, Santos AB: Assertive community treatment: An update of randomized trials. Psychiatric Serv *46:*669, 1995.

Dorwart RA, Hoover CW: A national study of transitional hospital services in mental health. Am J Pub Health *84:*1229, 1994.

Fenton WS, Blyler CR, Heinssen RK: Determinants of medication compliance in schizophrenia: Empirical and clinical findings. Schizophr Bull *23:*637, 1997.

Frank RG, McGuire TG: Savings from Medicaid carve-out mental health and substance abuse services in Massachusetts. Psychiatr Services *48:*1147, 1997.

Goldberg D, Jenkins R, Millar T, Faragher EB: The ability of trainee general practitioners to identify psychological distress among their patients. Psychol Med *23:*185, 1994.

Goldman HH, Morrissey JP, Ridgely MS: Evaluating the Robert Wood Johnson Foundation Program on Chronic Mental Illness. Milbank Q *72:*37, 1994.

*Howard KA, Cornille TA, Lyons JS, Vessey JT, Lueger RJ, Saunders SM: Patterns of mental health service utilization. Arch Gen Psychiatry *53:*696, 1996.

Johnson RE, McFarland BH: Lithium use and discontinuation in a health maintenance organization. Am J Psychiatry *153:*993, 1996.

*Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, Robinson P, Russo J: Collaborative management to achieve treatment guidelines. Impact on depression in primary care. JAMA *273:*1026, 1995.

Katon W, Von Korff M, Lin E, Bush T, Russo J, Lipscomb P, Wagner E: A randomized trial of psychiatric consultation with distressed high utilizers. Gen Hosp Psychiatry *14:*86, 1992.

Kehoe RF, Mander AJ: Lithium treatment: prescribing and monitoring habits in hospital and general practice. Br Med J *304:*1178, 1992.

Kessler RC, Olfson M, Berglund PA: Patterns and predictors of treatment contact after first onset of psychiatric disorders. Am J Psychiatry *155:*62, 1998.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshlmean S, Wittchen HU, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. Arch Gen Psychiatry *51:*8, 1994.

Lehman AF, Dixon LB, Kernan E, DeForge BR, Postrado LT: A randomized trial of assertive community treatment for homeless persons with severe mental illness. Arch Gen Psychiatry *54:*1038, 1997.

Lurie N, Moscovice IS, Finch M, Christianson JB, Popkin MK: Does capitation affect the health of the chronically mentally ill? Results from a randomized trial. JAMA *267:*3300, 1992.

McGrew JH, Bond GR, Dietzen L, Salyers M: Measuring the fidelity of implementation of a mental health program model. J Consult Clin Psychol *62:*670, 1994.

*Mechanic D, Schlesinger M, McAlpine DD: Management of mental health and substance abuse services: State of the art and early results. Milbank Q *73:*19, 1995.

Nichol MB, Stimmel GL, Lange SC: Factors predicting the use of multiple psychotropic medications. J Clin Psychiatry *56:*60, 1995.

Olfson M, Pincus HA: Measuring outpatient mental health care in the United States. Health Affairs *13:*172, 1994.

Pincus HA, Tanelielian TL, Marcus SC, Olfson M, Zarin DA, Thompson J, Zito JM: Prescribing trends in psychotropic medications: primary care, psychiatry, and other specialties. JAMA *279:*526, 1998.

Primm AB: Assertive community treatment. In *Integrated Mental Health Services: Modern Community Psychiatry*, WR Breakey, editor. Oxford Press, New York, 1996.

Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK: The de facto US mental and addictive disorders service system: Epidemiologic catchment area prospective 1-year prevalence rates of disorders and services. Arch Gen Psychiatry *50:*85, 1993.

Rogers WH, Wells KB, Meredith LS, Sturm R, Burnam A: Outcomes of adult outpatients with depression under prepaid or fee-for-service financing. Arch Gen Psychiatry *50:*517, 1993.

Segal SP, Cohen D, Marder SR: Neuroleptic medication and prescription practices with sheltered-care residents: A 12-year perspective. Am J Pub Health *82:*846, 1992.

Shorr RI, Fought RL, Ray WA: Changes in antipsychotic drug use in nursing homes during implementation of the OBRA-87 regulations. JAMA *271:*358, 1994.

Shorr RI, Bauwens SF: Effects of patient age and physician training on choice and dose of benzodiazepine hypnotic drugs. Arch Intern Med *150:*293, 1990.

Soumerai SB, McLaughlin TJ, Ross-Degnan DR, Casteris CS, Bollini P: Effects of limiting Medicaid drug-reimbursement benefits on the use of psychotropic agents and acute mental health services by patients with schizophrenia. N Engl J Med *331:*65, 1994.

*Sturm R, Wells KB: How can care for depression become more cost-effective? JAMA *273:*51, 1995.

Weintraub M, Singh S, Byrne L, Maharaj K, Guttmacher L: Consequences of the 1989 New York State triplicate benzodiazepine prescription regulations. JAMA *266:*2431, 1991.

Wells KB: Cost containment and mental health outcomes: Experiences from U.S. studies. Br J Psychiatry *166* (Suppl):43, 1995.

*Wells KB: Treatment research at the crossroads: the scientific interface of clinical trials and effectiveness research. Am J Psychiatry *156:*5, 1999.

5.4 ANIMAL RESEARCH AND ITS RELEVANCE TO PSYCHIATRY

WILLIAM T. MCKINNEY, JR., M.D.

Since the previous edition of this textbook the field of psychiatry-related animal research has undergone rapid development despite the continued lack of an effective, organized effort at the federal level. Increasingly attempts have been directed to conceptual clarification of the role of animal models, including their advantages and limitations in the study of human development and psychopathology, and a broader range of animal preparations with important interfaces to the basic and clinical sciences have been developed.

A theme apparent in the field of animal research is that there is no single comprehensive animal model for any clinical syndrome; all proposed animal models have their advantages and limitations. However, there is a range of models for studying specific components or aspects of psychiatric illness, which is a more useful way of viewing animal models (i.e., as experimental preparations developed in one species to understand phenomena occurring in another species).

Cross-species reasoning requires caution and scientific accuracy. Overextended comparisons have no place in comparative psychiatry and have sometimes given the field a bad name. However, certain principles guide attempts to learn something about the behavior or neurobiology of one species by studying another species, and these principles must always be respected.

Most psychiatric illness can be best understood by using a multivariate approach. Animal studies, where variables can be controlled rather precisely, have the potential to permit the study of the main effects of single variables and especially of how they interact. Such approaches are highly relevant to a biopsychosocial view of human psychopathology.

HISTORY

Ivan Pavlov is often said to have been the originator of research relevant to animal modeling of human psychopathology in general. His use of clinical terms as well as the experimental techniques he used, may seem foreign to most psychiatric clinicians. However, the fact that his work represented one of the first moves away from the correlational method of behavioral analysis to the experimental study of psychopathology is of central importance. As H. D. Kimmel has said:

> The significance of this change in direction may best be comprehended in relation to its two most important implications. First, the completely correlational method of behavioral analysis, which was the empirical foundation of all earlier systematic efforts to understand psychological abnormality, including everything from Hippocrates' humors and Franz Gall's prominences to the ingenious psychoanalytic theorizing of Sigmund Freud, could now be supplemented, if not altogether supplanted, by a direct experimental approach which was much less fraught with the dual dangers of loose conjecture and empirical untestability. Second, and historically of possible greater significance, the continuity of animal morphology, physiology, and behavior, already beginning to assume a position on center stage in man's philosophical thinking, received a new extensive thrust from the early Pavlovian findings since for the first time even such uniquely human phenomena as emotional breakdowns were seen to occur in subhuman animals.

Pavlov was followed by a number of other workers, and it is difficult to know what conclusions to draw about the early history of the field of experimental psychopathology research; some have not seen it as a particularly noteworthy beginning. However, the early pioneers may have been more successful than they seem in developing certain principles that are being rediscovered today. These include the following:

1. The demonstration that psychopathology could be experimentally studied in animals as well as in the strictly correlational studies done previously in humans.
2. The demonstration of the importance of careful behavioral observations as well as serendipity. Although most of the early workers did not use the more sophisticated and quantifiable behavioral scoring techniques now available, they were keen observers and literate in their descriptions.
3. The repeated proposal of an interactive model of psychopathology. The role of the temperament of the animals, along with a variety of social and neurobiological variables, was repeatedly stressed in the early literature. The concept of individual variability was part of the early work, and investigation of the sources of such variability continues to be an important area of research.
4. The discovery that a persistent internal response could be present even after the inducing stimulus had been withdrawn remains a major contribution to the understanding of a number of forms of psychopathology.
5. The recognition of the importance of unpredictability and uncontrollability, systematic investigations into which continue today.
6. Experimental paradigms for adaptive behavioral process that provide the foundation on which maladaptive behavior patterns are built in the presence of altered environmental demands. Adaptive mechanisms of animals and humans are fragile and share a tenuous relationship with the environment. Either internal changes in the organism (e.g., with drugs or other altered neurochemistry) or changes in the external environment (e.g., example, separation, the imposition of uncontrollability) can lead to serious behavioral changes, which can in turn lead to neurobiological changes and the establishing of a vicious circle. The study of those interactions has become the cornerstone of animal modeling research in depression and other forms of psychopathology.

One of the problems in early experimental psychopathology was that clinical terms were applied far too loosely and prematurely to a set of behavioral changes induced by methods that seemed to bear only a faint resemblance to inducing conditions for human syndromes. The result was that clinicians viewed the whole field of animal research with a certain amount of skepticism and cynicism.

The work of Harry Harlow, in particular, helped to stimulate clinicians' interest in the field of primate behavioral research. With his passing in 1981 the field of primatology lost one of its most prominent scientists. Harlow's research with primates began in about 1930 with observations at the local zoo. He soon discovered that monkeys and apes were much smarter than rats, and that tests designed to study rodent learning did not begin to tap the primates' intellectual capacities. It became apparent that more challenging or complex learning tasks and a better physical environment in which to test his primates were needed. Harlow addressed these two problems by developing the Wisconsin General Test Apparatus and a primate laboratory.

The Wisconsin General Test Apparatus brought to the study of primate learning capabilities a means by which a large number of discrete learning tests could be rapidly presented in highly standardized fashion to subject after subject. Studies of primate learning proliferated in the following years, and a battery of discrimination learning and memory tasks was developed that provided a standardized intelligence test for monkeys. Harlow then proceeded to study cortical localization of learning capabilities by lesioning different primate brain areas and noting subsequent differential patterns or deficits in their performance on the test battery.

In the late 1940s Harlow achieved a major conceptual and methodological breakthrough with his discovery of learning sets. He showed that rhesus monkeys presented with long series of six-trial, two-choice discrimination problems soon learned to achieve near-perfect performance on the second and subsequent trials of each problem. He was able to demonstrate unequivocally that the monkeys had acquired a strategy for problem solving. What the monkeys had learned was an abstract concept (“learning to learn,” in his words), rather than the product of simple associative learning.

Harlow's interest in the processes underlying primate learning extended to two lines of research. The first line involved motivation. In an effort to understand the factors that influenced learning performance, he discovered major inconsistencies with classic notions of drive reduction. It was not readily apparent why monkeys should solve puzzles more effectively when motivated by mere curiosity than when driven by hunger or thirst, but they did. Convinced that drive reduction and other motivational theories then in vogue in psychology and psychiatry could not possibly account for most of a monkey's behaviors, he began to look for alternative formulations.

At about the same time, Harlow began studies of the ontogenetic development of learning capabilities in rhesus monkeys. The task required devising a new battery of age-sensitive learning tests and acquiring suitable subjects. The latter requirement led to the establishment of a captive breeding colony and a nursery suitable for the hand rearing of large numbers of baby monkeys.

Work from Mortimer Mishkin's laboratory has demonstrated that an experience can enter into memory in monkeys in two ways: (1) as cognitive information stored in a cortico-limbo-thalamocortical system involving higher-order sensory areas of the cortex, amygdala, hippocampus and interrhinal cortex, medial thalamic nuclei, ventromedial prefrontal cortex, and basal forebrain; and (2) as a habit perhaps stored in a cortical striatal system involving sensory cortical areas, caudate, and putamen. The two systems seem to be developmentally dissociable. The system for noncognitive association (*habit system*) seems to develop early in infancy. By contrast, the system for cognitive association (*memory system*) appears to develop late in infancy, presumably because the neural circuit on which it depends has a slow ontogenetic maturation. Neonatal removal of the limbic system (i.e., combined amygdalo-hippocampal removal) produces a severe memory deficit from birth onward that is accompanied by social and emotional abnormalities that are strikingly similar to those seen in children with autistic disorder (i.e., an absence of social interactions, blank facial expression, motor stereotypies, and memory deficits). Thus, neonatal damage that leads to a severe memory disorder can have extremely serious consequences for personality and social development in part because of the cognitive memory impairment that is present from infancy onward, but also because of the direct effect of limbic lesions on mechanisms of emotionality.

Next came the cloth- and wire-covered surrogate mothers and the entrance of primatology into clinical psychiatry. Infant rhesus monkeys reared with the choice between a wire surrogate that fed them and a cloth surrogate that did not overwhelmingly preferred the cloth “mother.” Contrary to prevailing wisdom, *contact comfort*, as Harlow called it, was much more instrumental than feeding in bonding most infants to their surrogates. Harlow's discoveries with the surrogates sounded the end for drive reduction theory and revolutionized thinking about the socialization process in children; they also opened up the field of primate social development to serious scientific inquiry.

Eventually, Harlow shifted the major focus of his research to the study of social behavior and its development, both normal and abnormal. He developed the concept of affectional systems, the idea that social ontogeny involves the establishment of qualitatively different types of social relationships with a variety of others in the social network—parents, siblings, and peers as one grows up. At the same time, he studied the consequences of blocking the formation of different affectional systems through social isolation rearing or disrupting the attachment bonds once formed by experimental separations. Those studies clearly established the overwhelming importance of early social experiences for the development of species-normative adult social activities, including reproduction and maternal behavior.

Harlow spent his last years at Wisconsin expanding on the twin themes of normal and abnormal social behavior. Along with Margaret Harlow he was instrumental in establishing the nuclear family living unit, in which adult male-female pairs and various offspring could live together in a laboratory but in a situation rich in stimulation compared with the more typical laboratory environment. Harlow made a career of using rhesus monkey subjects to study human capabilities and problems not easily researched in humans themselves, and many of the fundamental concepts he developed, the source of considerable controversy at the time, are now fully incorporated into developmental theories.

RATIONALE FOR USING ANIMALS IN PSYCHIATRIC RESEARCH

The following reasons for including animal modeling research as part of a comprehensive psychopathology research program are illustrative rather than comprehensive. Issues concerning the development, practical use, and potential benefits of animal models are also summarized in Table 5.4-1.

Criteria for Development	Use in Research	Benefits
Similarity of inducing conditions	Treatment screens	New treatments
Similarity of behavioral syndromes	Study of developmental determinants	Better definition of syndromes in humans
Similarity of neurobiological mechanisms	Study of underlying mechanisms	Identification of at-risk persons
Similarity in response to clinically effective treatments	Study of interactions between social and biological components	Development of improved diagnostic tests
Other criteria appropriate for the specific model		

Table 5.4-1 Development, Practical Use, and Potential Benefits of Various Types of Animal Models

1. Many critical questions about the origins of human psychopathology cannot be studied directly in humans. By using animal preparations it is possible to control inducing conditions fairly precisely and to study the behavioral and neurobiological effects on a short-term as well as a long-term basis. For example, in the field of depression, prospective studies examining the effects of developmental events on behavior and neurobiology can be done much more easily in animals than in humans. The timing and exact nature of certain alterations in development can be specified, and the short- and long-term consequences studied. That aspect of modeling research is relevant to the question of developmental vulnerability based on early experiences and the mediating mechanisms of vulnerability.

   A particular line of research where animal preparations have a special contribution to make is in prospective studies of the effects of developmental events on behavior and neurochemistry. The interactions between those variables can be studied in a controlled and prospective manner. Animal preparations developing since the late 1980s make such investigations feasible and will facilitate the movement beyond correlation and retrospective analysis to cause-and-effect studies.
2. The underlying mechanisms associated with specific behaviors and patterns of behaviors can be studied more directly in certain animal species. Animal models potentially make possible the dissection of mechanisms in a more direct way than is possible in human clinical research, and they complement ongoing efforts in human protocols. More direct and potentially more invasive studies of neurobiological mechanisms can be performed, although such procedures need to be suited both to the species and the overall purpose of the experimental paradigm. Not all procedures are justified on ethical or economic grounds in all species. The questions have to be clear and specific, especially in proposing such studies in higher-order primates.

   The time is ripe for a vigorous effort in this area. The area of experimental psychopathology in animals has become complex enough to require the involvement of multiple laboratories, much as collaborative human studies of psychopathology often involve many centers. For example, different strategies and approaches need to be employed with several species, and techniques now generally available in only one laboratory need to be applied to many of those preparations. Attention needs to be paid to how to do what kinds of mechanism studies in a given species. Molecular or submolecular studies may be indicated in some preparations, but that may not be the only reasonable way to approach mechanism studies, for example, in a socially behaving species. The issues are complex but probably can be solved with enough discussion and with the development of some collaborative protocols across laboratories that take advantage of complementing expertise.
3. Single variables can be evaluated in terms of their main effects and in terms of the nature of their interaction with each other. For example, the nature of the interactions among genetic, developmental, social, and biological variables can be studied in various combinations in different species. In human clinical research, multiple variables interact simultaneously, and it has been impossible to sort them out in any quantifiable way.
4. The ability to isolate specific behavior patterns in animals and to study their origins, pathophysiology, and responsiveness to treatment techniques is important. Typically, in clinical work one is dealing with a broad range of behaviors that occur together, and it is impossible to study one or two in isolation and to understand them more completely. The many examples include anhedonia, stereotypic rituals, social withdrawal, and altered learning and cognitive abilities. If these and other particular aspects of psychopathological syndromes can be understood better, it might be possible to expand the understanding of situations where they typically occur together.
5. Animal models have played an important role in the preclinical evaluation of drugs. The empirical or predictive validity of animal models will be discussed further with regard to the general kinds of animal models.

   A related aspect of the use of animal models is their contribution to a better understanding of the mechanism of the action of drugs in altering specific behavior

patterns. This assistance goes beyond a mere prediction of whether drugs work or do not work and relates to the behavioral effects of agents with relatively specific mechanisms of action.

6. Animal models can also be used to understand the mechanisms of established treatment techniques. They can make possible the investigation of the mechanisms in terms of pathogenesis and of treatment responsiveness; that is, why do some drugs work whereas others do not? What are the mechanisms of action of electroconvulsive therapy in depression? Why are certain behavioral interventions effective whereas others are not?
7. Animal models also permit the understanding of a specific behavior or set of behaviors in terms of the developmental and social context as well as pathophysiologically. Rather than focusing on global syndromes, the origin, context, and responsiveness of behaviors to certain interventions can be investigated.
8. Animal modeling research, especially with primates, has led to the development of improved behavioral, ethologically based rating methods that can now be used in clinical research settings to evaluate social interactions, as between mother and infant or among peers.

Paul Wilner has described three different concepts of validity of animal models including predictive validity, face validity, and construct validity. He makes the point that, in the case of depression, animal models have been most frequently encountered in the pharmaceutical industry, where they have been used in screening tests in the development of new antidepressants. He indicates that animal models are also being used as simulations for investigating the psychobiology of depression, and that some models have been specifically developed for that purpose.

Predictive validity primarily concerns the correspondence between drug actions in the model and in the clinic. Manipulations that have certain effects in the clinic should have similar effects in the animal model for that model to be valid from this standpoint. Using this criteria, there are likely to be false positives as well as false negatives. Not all agents that work in an animal model will also work in humans and not all drugs that work in human depression will necessarily work in animal models. In terms of evaluating animal models according to this criteria it is the pharmacological profiling that is critical rather than the response to just one drug. No animal model has perfect concordance in this regard. *Face validity* refers to phenomenological similarities between the model and the disorder being studied. Realistically, not all symptoms can necessarily be modeled nor are all symptoms equally important in defining the clinical syndrome. The major point in considering the validity of animal models from this standpoint is the composite patterns of behaviors shown rather than the presence or absence of any one behavior or symptom. *Construct validity* refers to the theoretical rationale for the model, which in turn relates to the theoretical understanding of the clinical condition and its causation. Unfortunately, too many proposed animal models of depression or other psychiatric disorders use single proposed causes rather than operating on the basis of multiple risk factors. However, it should be possible in animal models to evaluate the relative contributions of various risk factors thought to be important in the human syndrome in question.

In a recent review, Mark Geyer makes the point that before criteria can be considered it is important to be explicit about the intended purpose of the model because this will determine in part the criteria that should be used to evaluate its validity. He contends that for a model to be of value in neurobiological research it must only satisfy two criteria: reliability and predictive validity. He does not think that construct validity is essential. He concludes his review by stating that there are no perfect animal models for any condition; each has advantages and limitations that need to be recognized in working with the model.

There are a number of approaches in developing animal models for any psychiatric disorder, and there is no such thing as a best model for any single syndrome. Animal models should be understood as basically experimental preparations that are developed in one species for the purpose of studying phenomena occurring in other species. The concern should be and increasingly is to develop a variety of experimental paradigms in animals to study selected aspects of human psychopathology rather than to attempt to develop a model of a given syndrome. Certain paradigms are suitable for studying certain phenomena whereas others are more suitable for studying other aspects. What is the best model for depression, schizophrenia, other mental disorders depends on what the question is.

CATEGORIES OF ANIMAL MODELS

Animal models can be divided into four categories: (1) those developed to simulate a specific sign or symptom of the human disorder (behavioral similarity models); (2) those developed to evaluate etiological theories (theory-driven models); (3) those developed with the primary purpose of studying underlying mechanisms (mechanistic models); (4) those developed to permit preclinical evaluation of treatment methods (empirical validity models). There is obvious overlap among the categories.

Behavioral Similarity Models Behavioral similarity models are designed to simulate specific symptoms of the human disorder in animals. The primary intent is to produce a particular symptom or set of symptoms, not to evaluate any specific etiological theory or to study underlying mechanisms or treatment responsiveness. The validity is judged by how closely the model approximates the human disorder from a phenomenological standpoint.

Theory-Driven Models In a theory-driven model a theory drives the development of a specific experimental paradigm. It is not necessary to assume the validity of the theory in order to proceed with the research. The attempt is made to operationalize the theory and to develop experimental paradigms to evaluate the effects of such inducing conditions. The approach is sometimes criticized on the basis that because the theory has not been substantiated in humans its use in animals is questionable. The response is that the very reason for developing the paradigm is to test or evaluate the theory. Most etiological theories of psychopathology have developed from studies of sick humans and therefore are retrospective. Using inducing conditions as an independent variable, the effects of specific conditions designed to represent certain causative theories can be evaluated prospectively.

Mechanistic Models The issue of the use of animals to study mechanisms is complex. For some, mechanism studies represent the only reason for developing animal models, which are evaluated by how well they lend themselves to those studies. With the increasing availability of high-technology methods for studying underlying neurobiological mechanisms, many researchers have become preoccupied with the molecular and submolecular basis of the altered behavior seen in many proposed animal models, and they consider useful only a model that permits those types of studies.

One must be careful, however, with the term “mechanisms.” Mechanistic studies certainly include the application of high-technology neuroscience techniques to the study of different mechanisms in the various models, but they also include social, behavioral, and developmental mechanisms. The techniques of mechanism studies cannot necessarily be transposed from rodents to monkeys or vice versa. Mechanisms must be studied as appropriate in a particular species and each approach should be recognized as having certain advantages and disadvantages. Because behavior and psychopathology occur in a social and developmental context the continued study and monitoring of social behavior is essential. A serious challenge that remains for researchers is the development of noninvasive techniques for mechanism studies in socially behaving animals that will be satisfactory to basic neuroscientists as mechanism studies. Compromises between directness of neurobiological studies and sophisticated ongoing assessment of social behaviors may be necessary.

Empirical Validity Models The best known and oldest use of animal models involves the use of animal preparation to develop and test clinically effective drugs. Basically, an ideal animal model would be one in which there are no false positives and no false negatives; that is, when a drug works in animals it is predictive of its clinical effects in humans and vice versa. Unfortunately, there is never 100 percent correspondence between the effects of a drug in an animal model and in a clinical condition, although there are a number of models with established high empirical validity. The establishment of an animal model as valid on empirical grounds (or on any other grounds) does not necessarily establish its validity on other parameters.

Genetic Models In contrast to most animal models that rely on stress-induced changes suggestive of depression, genetic models involve studying strains that exhibit spontaneous behaviors that mimic aspects of psychiatric disorders. Through selective breeding, some investigators have developed animal strains that are especially sensitive on certain tests. One example of this is the Wistar Kyoto (WKY) strain of rats, which are motorically depressed in the open field test, reveal an escape deficit when exposed to inescapable shock, and rapidly acquire a passive avoidance response and freeze in the conditioned defensive burying test. Desipramine (Norpramin) reduces immobility in the forced swim test. These animals are also sensitive to restraint-induced stress ulcer and show higher levels of adrenocorticotropic hormone in response to restraint stress, suggesting that they are hyperresponsive to stress stimulation. Other researchers have also studied various rat strains, especially the Flinders Sensitive Line that are reported to show depressive behaviors on a range of testing procedures and to respond positively to antidepressant medications.

ILLUSTRATIVE ANIMAL MODELS

Mood Disorders

Pharmacological Models In one approach drugs are administered to animals in order to reproduce some of the phenomenology of human depressive syndromes. A related approach is to use certain drugs to produce a set of changes in animals, changes that do not necessarily resemble human depression but have high empirical

validity in terms of predicting clinical drug responses. A third class of animal models is heterogeneous and, although not based on drug-induced changes, has been found useful in predicting and characterizing antidepressant activity. That type of model may have empirical validity in terms of drug screening, but in terms of induction techniques or behaviors, it bears little relationship to human depression (e.g., the bulbectomized rat syndrome and kindled amygdaloid convulsions).

Among pharmacological models, the syndrome induced by reserpine (Serpasil) and related compounds has been the most widely used. When given to animals of various species, reserpine produces a characteristic set of behaviors, including ptosis, hypothermia, inactivity, social withdrawal, and sedation. The interest in this animal model came from the clinical observation that drugs containing reserpine were reported to induce depression in humans who took them for the treatment of hypertension. At least in its initial form, this clinical observation has been called into question on the basis of recent evidence. It appears that humans who become depressed while taking reserpine-containing drugs have a history of depression and are presumably vulnerable in that area.

The reserpine syndrome as an animal model of depression has been reviewed extensively. In addition, a number of related compounds have been used to induce symptoms qualitatively similar to those produced by reserpine, although several factors are different, such as the speed of onset, duration, and central and peripheral effects. Initially most clinically active antidepressant drugs were believed to antagonize some or all of the symptoms induced by reserpine. In general this is true, but there are a number of inconsistencies. The reserpine syndrome is not a unitary entity, but involves many different effects. For example, ptosis antagonism, although probably a peripheral effect, seems to characterize the greatest number of clinically effective antidepressants, including several of the newer ones, some of which are regarded as false negatives in other reserpine procedures. In evaluating the empirical validity of the model, one must be specific about which behavior is involved. Some behaviors could have high empirical validity in terms of predicting clinical drug response, although the mediating mechanism may not even be central in origin. Reserpine has so many different neurochemical effects that one cannot reason directly from such studies to possible mechanisms associated with the behaviors produced, let alone to human depression.

Several other pharmacologically induced models have been developed, one of which is the amphetamine withdrawal model. Animals subjected to repeated amphetamine treatments, which are then stopped, evince a number of effects, including decreases in motor activity and self-stimulation behavior. Those effects have been reported to be reversed to a certain extent by amitriptyline, imipramine, mianserin, and pargyline when given on an initial short-term basis, but especially when given on a long-term basis.

Another proposed pharmacological model is clonidine (Catapres)-induced behavioral depression. Clonidine, an a-adrenergic receptor agonist, is thought to act at presynaptic receptor sites to reduce the release of norepinephrine, resulting in hypothermia, analgesia, and marked sedation. Clonidine-induced hypoactivity has been proposed as a test for antidepressant drugs. Although it is not clear which drugs will work in that model, it may play a role in screening for antidepressants that might be active in other models.

Another class of procedures widely used in pharmacology involves the potentiation by antidepressants of the behavioral and other effects of amines or their precursors. Those procedures do not attempt to mimic the clinical condition, but are based on theories about the role of the amines in depression. Thus, they are designed to show how amines interact with each other and are influenced by antidepressant drugs. By such routes, important information can be obtained that may ultimately have a bearing on mechanism questions, which can be tested in more highly developed behavioral models. Examples of such procedures include the potentiation by antidepressants of the various central effects of amphetamines and yohimbine (Yocon).

Pharmacologically induced models continue to be important for screening clinically effective drugs. Thus, in evaluating them, attention should be paid primarily to their empirical or predictive validity. Each has false negatives and false positives, but by using several such tests in a battery it may be possible to achieve an even higher degree of empirical validity.

In the case of all empirically based tests, it should be kept in mind that the mechanism by which the drug presumably acts in the animal preparation is not necessarily the same as its mechanism of action in human depression. Too many variables can intervene, and additional types of animal preparations may be necessary to assist with mechanism questions.

Separation Models Disruption of attachment bonds, whether in humans or nonhumans, has been established as a very stressful event. Humans and many animal species are in their most stable condition when they have developed secure social attachment systems. Disruption of such systems almost invariably leads to the development of grief reactions and can precipitate clinical depression in some vulnerable individuals. Many developmental, social, and neurobiological variables are known to influence the reaction to separation. However, determining the influence of those variables in humans and of how the variables interact with one another has been extremely difficult. Investigators have therefore turned to animal models for a more systematic study of the effects of separation.

BEHAVIORAL RESPONSES TO MATERNAL SEPARATION The earliest work on separation in animals began in the 1960s at a number of laboratories with the short-term separation of pigtail macaque infants from their mothers at the ages of 5 and 7 months, followed by reuniting them with their own or another mother. The behaviors seen following separation have been divided into two categories, labeled *agitation* or *protest*, and *depression* or *despair*.

The protest and despair response seen in many primate species after maternal separation has been compared with the responses in human children diagnosed with analytic depression or observed in institutions (usually hospitals or nurseries) where they were unavoidably separated from their mothers and families. The stages of response of human infants to maternal separation have been described as *protest, despair*, and *denial* (later changed to *detachment*). These stages have played a key role in the development of the theory of primary separation anxiety. As animal researchers extended the original work on the response to maternal separation, it became apparent that the response of the infants was influenced by a number of parameters, including the species, age, and social conditions.

The reaction to maternal separation in primates represents a true biobehavioral syndrome. Significant behavioral effects occur as described above, as well as major neurobiological changes.

PHYSIOLOGICAL RESPONSES TO MATERNAL SEPARATION Pigtail macaque infants undergoing maternal separation have been studied using totally implanted multichannel biotelemetry systems to monitor heart rate, body temperature, and sleep physiology before, during, and after the separation. In these studies the biological mother was usually removed from a group living situation and the infant was left in the group. Attachment bonds have been found to be as central to the development of monkeys as they are to humans, with disruption of those bonds leading to serious changes. In one study, for example, the infant's heart rate and body temperature increased significantly immediately after maternal separation. The changes were most pronounced early in separation and diminished as the separation continued. Beginning with the first night, both the heart rate and body temperature showed marked decreases from baseline levels and the behavioral patterns became more like depressive states. During reunion, the heart rate and body temperature returned to normal, although some infants exhibited a lower heart rate well into the reunion. An increased incidence of cardiac arrhythmias as a result of maternal separation has also been reported. Significant sleep changes have included increased sleep latency, more frequent arousals, less total sleep, and a disruption of rapid eye movement (REM) sleep.

In other research neurochemical effects were examined in rhesus monkey infants that were in the protest stage following maternal separation. Positive findings included elevated serotonin levels in the hypothalamus and significantly higher levels in the adrenal gland of all the major enzymes involved in catecholamine syntheses. Resting levels of norepinephrine and dopamine were unchanged in any of the brain regions examined. The study measured resting levels of those substances at one point in time and thus gives no information about possible dynamic changes occurring over time. However, it provides additional confirmation of the powerful effects of disrupting the maternal attachment bond; the syndrome is neither transient nor mild.

When separated from their mother or from a surrogate, squirrel monkeys show a marked increase in the pituitary-adrenal response. Initially it was reported that there was an identical physiological response whether the infant was separated from a mother or from a surrogate. The monkey mother also showed an elevated corticoid response to separation. The latter finding is interesting in that the mother's responses to infant removal have been very minimally described in most studies because investigators have been preoccupied with assessment of the infant. The mother has been typically described as acutely upset but as getting over it very quickly. The infant's corticoid response was felt to be caused by the separation itself rather than by the new cage in which it was housed during the separation phase, and this finding has been supported by data from a number of studies. The presence of a familiar animal during the separation phase did not alter the corticoid response, suggesting that the disruption of the specific attachment bond between mother and infant was the main cause of the increased corticoid levels.

Physiological and behavioral changes following separation may not occur simultaneously. For example, separation from the surrogate results in a behavioral response but no corticoid response. Later work shows that infants of highly dominant mothers manifest the greatest adrenocortical response to separation and that they may not always exhibit concomitant behavioral changes. This important research illustrates the complexity of the neurobiological and behavioral changes that

may accompany separation. It is important to obtain adequate baseline behavioral profiles of both the group structure and individual behavioral assessments.

Desipramine has been reported to be effective in preventing the response to maternal separation in primates, and imipramine has a similar therapeutic effect on the responses to peer separation.

OTHER RESPONSES TO MATERNAL SEPARATION When infant lagur monkeys were separated from their mothers at 6 to 8 months of age, all infants showed changes in social behavior, which varied from minimal to severe and included two deaths. All infants sought substitute caretaking during the separation and adopted a major substitute caretaker. Most infants remained with the substitute even when the mother returned. Some researchers have used distress vocalizations in various animals as an index of separation and have studied the effects of many pharmacological agents on those vocalizations. In general, the distress vocalizations are reported to be relieved by morphine and made worse by the narcotic antagonist naloxone (Narcan) when those drugs are given as single injections. A variety of opiate-like peptides have been tested, and all have been reported to be effective in decreasing distress vocalizations in separated animals when injected into the vicinity of the fourth ventricle in quite low doses. Additional maternal separation studies have been done using canine puppies, guinea pigs, and chicks. From those studies researchers have developed the theory that brain endorphins may play a critical role in the mediation of social bonds, and that when those bonds are disrupted by separation, a syndrome similar to that following narcotic withdrawal is produced.

PEER SEPARATION Rhesus monkeys and most other primate species develop strong, complex social bonds, and paradigms have been developed that involve experimental disruption of those bonds in peers of various ages, including adults. In general, the behavioral reaction to peer separation is quite similar to that following maternal separation in terms of the classic protest-despair response. Furthermore, when peer groups are formed and separations are repeated, the response is seen with each separation. Not surprisingly, a number of variables can influence the nature of the response, including age; rearing conditions; housing conditions before, during, and after each separation; and treatment with pharmacological agents. Significant individual variability can be related to a number of developmental and neurobiological variables. For example, the concentration of norepinephrine in cerebrospinal fluid (CSF) appears to be a trait-related marker predicting a more severe response to separation. Animals with lower concentrations of norepinephrine in CSF respond to separation with more huddling and self-directed behaviors than animals with higher concentrations. By contrast, CSF homovanillic acid (HVA) and 5-hydroxyindoleacetic acid (5-HIAA) levels are state-related markers that reflect the behavioral response to separation no matter how the response is obtained.

Pharmacological agents can affect the response to peer separation. Imipramine reverses the reaction to peer separation and prevents the reaction to future separations as long as the monkeys are continually administered the drug. They return to more typical separation behavior when the drug is withdrawn. Amphetamine modifies the behavioral response to separation in a very similar manner to imipramine, but the overall effects of the two drugs on group social behavior can be distinguished. a-Methylparatyrosine, which blocks tyrosine hydroxylase and thereby lowers norepinephrine and dopamine levels, can exacerbate the response to peer separation at doses so low that they have no effect when the monkeys are living as a stable social group. It is only when one combines the stress of separation with low-dose a-methylparatyrosine that one sees other effects of the drug in that paradigm. Parachlorphenylalanine, which blocks serotonin synthesis, has no effect. Low doses of alcohol alleviate the peer separation response whereas high doses make it worse.

During separation, mother-reared monkeys exhibited low and stable levels of disturbance behavior, which did not vary much from separation to reunion conditions. However, peer-reared monkeys exhibited significantly higher disturbance behavior during separations than mother-reared monkeys and the behavior was greater during separations than during reunions. Drug treatment with desipramine blocked the increase in disturbance expected by peer-reared animals during separations. Moreover, the full drug effect was not significant until 5 weeks after administration or 3 weeks after separation challenge. Fluoxetine (Prozac) had a similar effect on the response to repeated peer separation; desipramine and fluoxetine did not differ from each other.

Additional Separation Studies Based on studies in rats it has been reported that recurrent and unpredictable, brief, early maternal separations induce changes in responses to reward in adulthood and that these changes were generally opposite to those induced by isolation rearing. A reduction in the immediate exploratory response to novelty has been reported in both sexes, as well as a blunted locomotor responsivity to compound cues predictive of food reward in females. Early maternal separation did not lead to any overall alteration in consummatory behavior and sucrose preference, but did lead to a definite blunting of the responses to successive positive and negative contrast. The researchers concluded that adverse early experience could induce enduring alterations in the responding of rats to rewarding stimuli in maturity. Given the importance of anhedonia as a core clinical feature of serious depression, early maternal separation could represent a very useful paradigm to explore the neurobiological substrate of reward and of depressive disorders.

Other studies have examined the neuroendocrine correlates of separation stress in the Siberian dwarf hamster (*Phodopus sungorus*). These animals form monogamous male-female pair bonds. Disruption of the pair bond has been shown to result in increases in body mass and behavioral alterations similar to profiles seen in human atypical depression. They found that separation stress was accompanied by increased hypothalamic-pituitary-adrenal axis function, decreased peripheral sympathetic nervous system activity, and decreased reproductive profiles.

Data from a report of free-ranging rhesus monkeys to a more natural form of social separation than that used in the laboratory, namely the decreased amount of time a mother spends with its infant when they have their first postpartum estrus period and resume mating, suggest strong parallels between the behavioral response of rhesus infants to their mother's resumption of mating in the field and to forcible separation from mothers in captivity. In both situations the infants experienced increased separation from their mother and responded with distress and, in the case of a certain percentage, with depression.

RESEARCH APPLICATIONS With regard to proposed depression models, the rationale for separation studies in animals is that the bulk of evidence strongly suggests social separations as risk factors that cut across types of depressions. Animal studies represent one way of studying those risk factors. Although many factors are involved in depression, separations appear to be important events for vulnerable individuals and so are worthy of additional investigation. An even more important context in which to view separation studies is as prototypes for the study of stressful events in general. With the recent advances in the knowledge of developmental and neurobiological influences on behavior, it becomes increasingly important to have experimental paradigms in which the interactions between neurobiological factors and social risk factors can be examined. Such animal preparation provides the opportunity to control social and developmental variables and to conduct prospective studies of behavioral and neurobiological parameters. The long-term effects of early alterations can be examined in a much shorter period of time. Repeated measurements of neurobiological variables, as in the CSF, can be obtained in a way that cannot be done in humans and in relation to specific units of behavior. The effects of drugs on parameters similar to those being studied in humans can be evaluated and what those changes mean with regard to specific social behaviors can be determined. The role that specific neurotransmitter systems play in influencing specific units of social behavior, including responses to separation, can also be clarified. Humans are fundamentally social creatures, and an understanding of the social origins of psychopathology and how they are related to neurotransmitter systems becomes possible in this kind of preparation.

Learned Helplessness The animal model of learned helplessness has been extensively studied for more than 15 years. It relates closely to some important aspects of clinical depression, particularly cognitive aspects, such as a negative conception of the self, negative interpretations of one's experiences, and a negative view of the future. Those cognitive aspects are reflected in feelings of helplessness and hopelessness. Whether the phenomena are primary or secondary is a moot point for the present discussion. The phenomena occur as core aspects of depression frequently enough to be worthy of further study. Etiological theories as well as therapeutic approaches have developed from that cognitive view of depression.

In the original experimental study with animals, dogs were placed in one of three situations. In the first situation they were put in harnesses and subjected to electric shock that they could terminate by touching a panel; not surprisingly, they learned to escape the shock rather quickly. In the second situation the dogs were prepared as in the preceding situation, but when the shock was given, they were unable to terminate it. Finally, to control for the effect of the shock itself, dogs were put in harnesses but were not subjected to shock at all. In Phase 2 of the study, dogs were given electric shock while unharnessed in a shuttle box. Normally, dogs have no difficulty learning to avoid the shock by going to the other side of the box, which proved to be true for the dogs that had been exposed to escapable shock while in the harness. However, the dogs that had been exposed to inescapable shock failed to learn that they could escape from the electric shock in Phase 2 by jumping over a barrier that separated the two sides of the shuttle box. They were described as being initially agitated in reaction to the shock but, rather than to run around frantically until they discovered that they could escape the shock by crossing the barrier, they would sit or lie down, whining quietly—that is, they acted as if they were helpless and incapable of escaping. The interpretation was that the inescapable shock they had experienced earlier made them unable to cope with the present situation. One explanation was that they had learned during their initial experience that outcomes were not contingent on their behavior. No matter what they did, it did no good; they learned to be helpless.

To reverse this state, attempts were made to retain the dogs by trying to coax them across the barrier. This strategy was very difficult, and ultimately it was found that the only effective way was to drag the dogs across the barrier forcibly and thus to terminate the shock. It took many efforts before most of the dogs could learn this response and do it by themselves when placed in the shuttle box.

Recent studies suggest that cortisol may be specifically elevated in helpless animals. One study found that animals showing deficits in escape performance have norepinephrine depletions, specifically in the locus ceruleus, and the researchers have used an antagonist to manipulate that system. They suggest that the a-receptor system is involved, which has been confirmed in mice by other studies on the effects of serotonin and acetylcholine. Other investigators have suggested specific alterations in the hippocampal b-receptors that occur with the development of helpless behaviors. They suggest that those changes are reversed by antidepressant drug treatments, and that serotonin appears to regulate b-adrenergic receptor concentrations in the hippocampus. Along with lesion studies implicating the fornix and septal regions in the behavioral alterations, that study points to the septal-hippocampal pathway as an important site for serotonin regulation of norepinephrine function. The latter group has been successful in selective breeding for learned helplessness through several generations. Thus, use of the learned helplessness model that leads to some detailed hypotheses about the neurological regulation of mood is improving the understanding of human mood disorders.

RESEARCH APPLICATIONS The literature on learned helplessness is enormous and, as in the case of separation models, controversial. However, one set of recent findings merits closer scrutiny and illustrates clearly how experimental paradigms in animals can be used to investigate the interrelations between behavioral events and neurobiology, an interface crucial to understanding human psychopathology. Again, as in the separation models, one does not have to agree with the validity of the model itself to appreciate the value of that kind of work in a spectrum of research approaches aimed at understanding specific aspects of human depression.

Severe, inescapable trauma, for example, has been found to produce a deficiency in central noradrenergic activity, namely, depletion of locus ceruleus norepinephrine levels. However, studies have shown that subjects who were able to control the noxious experiences did not develop the noradrenergic deficiency and were able to respond efficiently. When the drop in noradrenergic activity was prevented by treatment with drugs, the learned helplessness phenomenon did not occur. Unfortunately, those kinds of data have sometimes been cited as evidence that the concept of learned helplessness is not valid.

A major finding, by contrast, has shown that a state that is clearly induced behaviorally is associated with major changes in certain neurobiological systems, and that if those changes can be prevented, the behavioral state that results from certain well-described behavioral manipulations can be prevented or even reversed. It would be interesting to learn if behavioral reversal of the syndrome leads to reversal of the biological changes, or if reversal of the biological changes alone (once the syndrome is set in motion) will reverse the behavioral aspects. Investigators know that in human depression not all of the significant behavioral and cognitive changes are necessarily reversed with drugs, which presumably alleviate whatever underlying biological alterations may be present. One sometimes has to deal directly with the altered cognitions and behaviors, as well as the underlying neurochemistry. The evidence of a complex interplay between a cognitive-behavioral state and neurobiology clearly warrants further investigation.

Other ongoing work in learned helplessness mainly relates to that interface between the behavioral state and neurochemical substrates. Several such substrates may prove to be important, and it will be critical to ascertain their interaction with behavioral variables. For example, the involvement of the opioid system in the learned helplessness models of depression has been investigated. Animals preexposed to inescapable shock and treated with either opioid agonists or enkephalin catabolism inhibitors evidenced a reversed escape deficit that was induced by shock pretreatment. In contrast, nalzone potentiated the effect of inescapable shocks. Imipramine reduced the number of escape failures and this effect was antagonized by naloxone (Narcan). The central histaminergic system has also been investigated in the learned helplessness paradigm. In another series of studies, earlier learned helplessness training was shown to sensitize the hippocampus to increased norepinephrine release during a subsequent less severe stressor.

Chronic Stress Models In chronic stress paradigms rats are subjected to a chronic stress regimen that is designed to be unpredictable with regard to the stimulus properties of the stress, as well as the time of stress delivery. Stressors are administered every 1 to 2 days over a period of 21 days at various points in the circadian cycle. Stressors include switching of cage mates, removal from double housing to single housing for 24 hours, 30 minutes of scrambled unpredictable foot shock, 46 hours of food deprivation, 46 hours of water deprivation, a coldwater swim, shaker stress, and tail pinch. After 21 days, when the rats are tested in an open field test situation, they do not exhibit normal open field activity or the usual response to an acute stress. The decreased exploratory behaviors can be reversed by a variety of drugs, including monoamine oxidase inhibitors (MAOIs) and tricyclic antidepressants, as well as by electroconvulsive therapy (ECT). Amphetamines and scopolamine are ineffective. Thus, the model appears to have good pharmacological specificity and is one in which studies of the neurobiological substrates could readily be done. Such an approach emphasizes the combined influence of chronicity and unpredictability in producing the behavioral alterations.

Changes in Dominance in Hierarchy Another proposed model reflects the importance of dominance in the relationships of many nonhuman primates. It has been postulated that changes in the stability of the dominance arrangement cause behavioral alterations. It is hypothesized that the behavior that occurs in association with gaining higher-dominance ranking may be elation; depression occurs with falls in one's position in the hierarchy. In this theory, depression is postulated to be adaptive because it prevents the descending animal from fighting back. It has been reasoned that, if this is so, depression could be induced by altering the dominance hierarchy in some nonhuman primates. The theory rests on limited data since the evidence for particular behavior patterns occurring in association with specific changes in the hierarchy is fragmentary. Dominance hierarchies are not easy to manipulate. However, work concerning dominance and serotonin metabolism will be interesting to follow. The work is not concerned primarily with animal models of depression, but it does involve manipulation of the dominance hierarchy, careful behavioral observations, and study of the serotonin systems.

Intracranial Self-Stimulation Models Another proposed animal model of depression is the reward-reduction model using self-stimulating animals. The involvement of catecholamines in the mediation of intracranial self-stimulation has been well established, although there is controversy about their relative importance. In general, agents that enhance the effects of catecholamines tend to increase intracranial self-stimulation responding, whereas those that impair catecholamine actions tend to decrease intracranial self-stimulation response rates. However, the actions of tricyclic antidepressant drugs in the model appear anomalous because they do not enhance intracranial self-stimulation responding despite their well-documented antidepressant action and their effects on catecholamine systems. Those drugs tend to decrease the rate of responding and to raise the reward threshold. Attempts have been made to find an animal model in which tricyclic antidepressant drugs potentiate intracranial self-stimulation responding. One such model has been suggested, in which reinforcement requires increasing effort. In those progressive fixed-ratio schedules, responding typically drops gradually to zero. Antidepressants have been reported actually to enhance responding. However, efforts to replicate the work have been unsuccessful. Rats that had electrodes implanted long term in the medial forebrain bundle were trained in progressively increasing fixed ratio schedules. Two tricyclic drugs (imipramine and protriptyline) were given, but neither resulted in response enhancement. Thus, additional work is indicated to evaluate more fully the reward-reduction model involving ICSS as an animal model of depression. It is worth pursuing, however, in view of the important finding that anhedonia is a key feature in many cases of severe depression. It would be valuable to have an animal model in which the mechanism and pharmacology of that sign could be studied.

Conditioned Motionlessness The proposed model involves pairing a buzzer (conditioned stimulus) with a tetrabenazine injection (unconditioned stimulus) for at least 11 trials. Following conditioning, some rats exhibited motionlessness after the presentation of the buzzer alone; imipramine attenuated the conditioned motionlessness. Subsequent neurochemical studies of those rat preparations supported the conclusion that the observed motionlessness was associated with an excess of functional serotonin at the synaptic cleft. Motionlessness after tetrabenazine administration is not blocked by imipramine. When the conditioned response is reversed by imipramine, however, the biochemical data resemble those for control subjects.

Behavioral Despair The model involves the use of a test of behavioral persistence in the context of an inescapable task. The test is based on the observation that when rats or mice are forced to swim in a restricted space from which they cannot escape, they eventually cease their attempts to escape and become immobile. It has been suggested that the characteristic behavioral immobility reflects a state of despair in the rats or mice. The immobility is reduced by most clinically active antidepressant medications, as well as by nonpharmacological treatments such as ECT, REM sleep deprivation, or exposure to an enriched environment. The effects can be seen after initial short-term administration, but more marked effects are seen after repeated treatments with lower doses. The drug effects do not appear to be caused by increased motor activity because the doses used generally decrease motor activity. Antidepressant medications seem to prolong the escape-directed behavior observed at the beginning of a test session, whereas psychostimulants or anticholinergic drugs cause a generalized behavioral stimulation. Those potential false positives can be distinguished from the effects of true antidepressant drugs. However, as with all models, both false positives and false negatives occur. False-positive results have been reported with antihistamines, subconvulsant doses of convulsants, and some neuropeptides. False negatives have been found with clomipramine (Anafranil) in rats and albuterol (Proventil, ventolin) in rats and mice.

The model was developed mainly for drug screening and thus must be evaluated in terms of its empirical validity, which seems at least as good as that of most drug-screening models. At a theoretical level its relationship with learned helplessness or uncontrollability models needs to be clarified and mechanism studies remain to be done. Recent years have seen a major expansion in the number of behavioral paradigms that have been proposed as animal models of depression.

Other Considerations of Animal Models of Mood Disorders Because it is ethically unacceptable to use human subjects to conduct manipulative experimental hypothesis-testing research on major depressive disorder, investigators interested in the depressive disorders have turned to observation of models with humans and to interactive models with animals. The models based on limbic dysfunction could be used to explore specific neurosubstrates of depression. In that context, the rat

with limbic dysfunction induced by olfactory bulbectomy is believed to best fulfill the requirements for a model of depression. Discovery of a genetic strain of animals that show comparable behavioral and neurovegetative deficits and similar selected drug responses to those seen in patients with depression would help to advance the field and increase the understanding of the neurobiology of the limbic system.

Comparatively little work has been done to develop animal models of bipolar disorders. Most work has involved the antagonism by lithium (Eskalith) of drug-induced states in animals and man. Table 5.4-2 summarizes those studies. Some of the newer thinking is represented by an evolutionary model of manic-depressive disorder, the essence of which is that manic and depressive behaviors are evidence of triggered fundamental alpha (a) and omega (w) states of social rank and are not by themselves pathological.

State	Lithium Effect	
	Humans	Animals
Amphetamine (low dose)		
Hyperactivity, activation	Blocked	Reduced
Amphetamine (high dose)		
Stereotypical behavior	Unchanged or prolonged	—
Morphine		
Excitement, euphoria	Reduced or unchanged	Unchanged
DMA plus tetrabenazine		
Hyperactivity	Reduced or unchanged	—
L-Dopa		
Hyperactivity, hypomania	—	Possibly reduced
Ethanol high	—	Unchanged
MAO-inhibitor		
Hyperactivity	Enhanced	—

Table 5.4-2 Antagonism by Lithium of Drug-Induced States of Animals and Humans

SCHIZOPHRENIA Is it possible to produce an animal model for schizophrenia? Several authors have suggested standards that should be satisfied if a particular preparation is to qualify as an adequate animal model. Those criteria have been mostly drug related; for instance, clinically effective antipsychotic drugs should reverse the abnormal behaviors whereas clinically ineffective drugs should not.

There should be no argument about the need for animal models of psychiatric diseases in general and for schizophrenia in particular. Although significant progress is being made in clinical research on schizophrenia, the development and use of suitable animal preparations could facilitate many kinds of studies that, for ethical and practical reasons, are impossible to do in humans.

There are no animal analogues for many of the core signs and symptoms of schizophrenia. Clinicians and animal researchers need to pay careful attention to ways of making schizophrenia's signs and symptoms operational, possibly through a more ethological analysis of human schizophrenia than has so far been done. If different types of human schizophrenia could be analyzed ethologically, the same or a similar system could be applied to the development of animal models and analogues for specific behaviors might then become more feasible. There also needs to be a major conceptual shift away from the idea that an animal model of schizophrenia can be developed. It is more realistic to develop animal preparations for studying specific aspects of schizophrenia and to better understand some fundamental issues. Figure 5.4-1 summarizes some of the existing animal models of schizophrenia along with some future approaches.

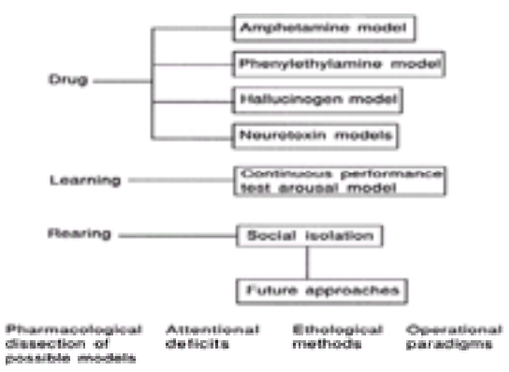

FIGURE 5.4-1 Existing animal models of schizophrenia and future approaches.

Drug-Related Animal Models

AMPHETAMINE MODEL The amphetamine model has attracted a considerable amount of attention in connection with schizophrenia and hyperactivity because amphetamine psychosis in humans can closely mimic paranoid schizophrenia. Do the data indicate, as some investigators have suggested, that animals develop schizophrenia when they are given amphetamine? Clearly, the paranoid delusions have no direct measurable analogue in the animal model, although inferences have been made from certain behaviors. Some investigators have reported that animals are hypervigilant when given amphetamine, as shown by their increased alertness and visual attention to other animals in their environment. Subordinate rats actively withdraw from social interactions, retreat to strategically defensible positions in their environment, and remain hypervigilant; some researchers have theorized that the behavior may be a manifestation of paranoia.

When given to rats, cats, and monkeys, amphetamine produces stereotypical behaviors; stereotypical behavior is not included in any list of major symptoms of schizophrenia in humans; thus, a model is being proposed based mainly on a behavior that is nonspecific and nondiscriminating for schizophrenia. However, from the standpoint of empirical validity, drugs that have antipsychotic properties in humans block the amphetamine-induced stereotypical behavior in both animals and humans. This finding has been related to the possible dopaminergic mechanisms in schizophrenia on the assumption that amphetamine-induced stereotypical behavior in animals is mediated by increased dopamine turnover.

Some investigators have tried to separate the stereotypical behavior from the increased locomotor activity produced by amphetamine by relating the behavior to the release of striatal dopamine, which is significantly increased with repeated amphetamine administration. By contrast, the increased locomotor activity is thought to be mediated by norepinephrine. Tolerance develops in locomotor activity with repeated doses of amphetamine, but not to the effects on stereotypical behaviors. Such an approach attempts to distinguish the different types of amphetamine-induced behavioral alterations in animals and to elucidate the neurochemical mechanisms that may be involved.

Other researchers have examined the effects of amphetamine on a variety of behaviors in rats, especially locomotor activity and stereotypy. A progressive augmentation of both behaviors with repeated drug administration has been reported. The duration of stereotypy was not necessarily increased, but the onset of stereotypy was. Those data were interpreted to mean that long-term amphetamine administration tended to produce increased preservation of progressively more focused and restricted behaviors; a similar phenomenon has been reported in rhesus monkeys.

PHENYLETHYLAMINE MODEL Another proposed drug-induced model of schizophrenia involves b-phenylethylamine, a neuroamine that is an endogenous component of mammalian brain and is most highly concentrated in the limbic system of the human brain; the drug phenylethylamine produces stereotypies that closely resemble those produced by amphetamines.

HALLUCINOGEN MODEL Many animal experiments using hallucinogenic agents have been very productive in promoting the understanding of the behavioral pharmacology of those compounds, but they have not provided convincing animal models of schizophrenia. As the phenomenology of hallucinogenic-induced states

in humans was more carefully compared with the symptomatology of schizophrenia, the proposed behavioral isomorphism became less persuasive.

NEUROTOXIN MODELS One hypothesis is that schizophrenia results from impairment of the structural integrity of the noradrenergic reward mechanism, and that the impairment is chronic and at least partially irreversible. 6-Hydroxydopamine (6-0HDA) is said to be the aberrant metabolite that causes schizophrenia on the basis of the following evidence obtained from animal studies:

1. When injected into certain brain sites in animals, 6-0HDA decreases stimulation and other rewarded behaviors an effect that is long-lasting.
2. Prior treatment with chlorpromazine (Thorazine) blocks the behavioral deficits as well as the depletion of norepinephrine induced by 6-hydroxyclopamine.

Nondrug Animal Models

AROUSAL It has been postulated that patients with schizophrenia operate at excessive arousal levels and that impaired attention resulting from hyperarousal constitutes a major deficit. By training rats on an operant task thought to be analogous to a test of attention (the Continuous Performance Test) used to study schizophrenia patients, attempts have been made to produce an animal model for the attentional deficit. Low levels of electrical stimulation to the reticular formation in rats cause the animals to make errors similar to those made by patients with schizophrenia. In general, antipsychotic drugs are most effective in reversing the deficit.

PRIMATE SOCIAL ISOLATION Isolation seems to share some components with schizophrenia, as illustrated in Figure 5.4-2, but one should be extremely cautions about such linkages until further studies have been done.

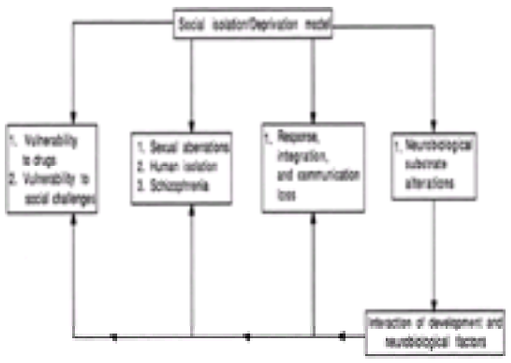

FIGURE 5.4-2 Applications of the social isolation-deprivation model of human psychopathology.

There are two potentially productive approaches toward creating animal models of schizophrenia. The ethological approach would focus on a specific behavioral analysis of schizophrenia and could in turn provide the foundation for comparisons with other species. The second approach would involve creative operational paradigms for animals based on the present level of analysis of human schizophrenia and move away from attempts to develop global models of the disorder.

Sensorimotor Gating Model of Schizophrenia Another proposed animal model of schizophrenia relates to the concept of sensorimotor gating. The background comes from reports of oversensitivity to sensory stimulation in schizophrenia patients that theoretically correlates with stimulus overload and leads to cognitive fragmentation. There have been experimental paradigms using cortical event-related potentials and the prepulse inhibition of startle responses that show that schizophrenia patients have impaired CNS inhibition (*sensorimotor gating*). Animal model studies have shown that sensorimotor gating failure similar to that seen in schizophrenia patients can be reproduced. The time course of the observed schizophrenia and animal model deficits is compatible with a temporal map of monoaminergic neuron functions. The researchers make the point that studies of sensorimotor gating would allow investigators to comment on the spatial and temporal mapping of neurons, trait and state deficits, and vulnerability factors in a schizophrenic spectrum of disorders. Such an approach that studies sensorimotor gating research in schizophrenia patients and related animal model experiments should permit one to understand the functional significance of neurotransmitter abnormalities and the relationship of those abnormalities to cognitive disturbances seen in schizophrenia. Some recent studies have shown that periods of social isolation of rats during development, but not during adulthood, produce deficits in prepulse inhibition that are partially reversible by a dopamine antagonist. Thus, this is a model in which neurodevelopmental factors can be studied in terms of their possible interactive role in schizophrenia.

ANXIETY DISORDERS It is well known to clinicians that anxiety can be either a symptom or a specific syndrome; the literature on animal models is often very confusing in that regard. In some work, anxiety seems to be used synonymously with neurosis; in other paradigms, what is being studied seems more akin to fear or to certain kinds of learning behavior. It is important to keep in mind the core features of the human syndrome.

Proposed models of anxiety must be evaluated according to how they resemble human anxiety behaviorally and how they conform to treatment responsiveness criteria. Neither in itself is completely satisfactory, but investigators do not know enough about the cause, pathogenesis, or mechanisms of human anxiety to use them as validating criteria for animal models. Some of the proposed animal models of anxiety may help to clarify those issues.

Most approaches to animal modeling of anxiety use variations of operant conditioning paradigms, and validity is evaluated on an empirical basis—that is, how well do clinically effective antianxiety agents work in the paradigms and how specific is the response? A number of the approaches have high empirical validity. Although they may not bear any relationship to the etiology or pathogenesis of human anxiety, they may still have merit in the context of that aspect of animal modeling. The assessment of behavior is of vital importance in the development of any animal model, and anxiety is no exception. What defining characteristics of the symptom or syndrome of human anxiety should researchers attempt to measure in animals? The inducing conditions for many forms of human anxiety are unknown, and so it is impossible to use those conditions as criteria. Also, not enough is known about its mechanisms to use anxiety as a criterion.

Operant Conditioning Paradigms The basic strategy is to use operant techniques to elicit a behavior with a high frequency of occurrence. After the response is well established, the behavior is suppressed by punishing it when it occurs. The analogy to fear is the conditioned association between the behavior and the punishment. Potential antianxiety drugs are evaluated according to their ability to restore responding to what it was at presuppression levels.

The Geller Conflict Test, one of the best known of such tests, is widely used to screen for potential antianxiety drugs. In the original Geller paradigm, rats were trained on a multiple variable interval 2-minute continuous reinforcement schedule for milk reinforcement; that is, there was a two-component operant behavior schedule. In the variable interval portion, signaled by one stimulus, bar pressing was reinforced at variable intervals, with the mean interval being 2 minutes. In the continuous reinforcement portion, signaled by a different stimulus, every bar pressing was reinforced. When foot shock was given concurrently with the positive reinforcement, response rates were suppressed.

Drug-induced increases in the rate of punished responding are interpreted as an index of antianxiety activity whereas decreases in unpunished responding are interpreted as indicating depressant activity. The type of behavior that originally occurs frequently but is subsequently suppressed by certain manipulations is highly sensitive to the benzodiazepines and meprobamate (Miltown) but not to chlorpromazine (Thorazine). The test, along with many modifications of it, identifies clinically active anxiolytic agents, predicts their clinical potency, and is generally insensitive to stimulant, antipsychotic, antidepressant, or analgesic drugs. It seems to work in different species and to be relatively independent of the schedules of positive reinforcement or punishment. Thus, the operant conflict approach has high empirical validity in terms of predicting clinical drug responsiveness.

The model, involving conflict behavior, has been used to evaluate several biochemical hypotheses concerning the mechanism of action for the antianxiety, “emotional analgesic” properties of benzodiazepines. With the increasing interest in the possible neurobiological substrates of anxiety, the availability of a paradigm in which the mechanism of drug action can be explored simultaneously with measures of operant behavior is necessary for future research. Over the past few years with the identification of benzodiazepine receptor sites, evidence has begun to accumulate that those receptors may mediate the therapeutic effects of such drugs. The relation of those receptor sites for benzodiazepines to the understanding of the neurobiological mechanisms of anxiety states is complex and remains an active area of investigation. The involvement of various neurotransmitter systems with those receptor sites is important to understand, and the continuing development of experimental systems in which the complex interrelationships can be studied would be helpful. A number of theories, especially with regard to the g-aminobutyric acid

(GABA) and the serotonin systems, purport to explain the effects of the anxiolytic drugs. However, anxiety has many different components, and different neurotransmitters may be involved with each. For example, muscle relaxant or anticonvulsant properties may be neurochemically mediated in one way, but anxiolytic effects as revealed in conflict paradigms may be mediated by other neurotransmitters. It is important to have careful behavioral descriptions available along with the specific neurochemical technology.

Alteration of Locus Ceruleus Function The locus ceruleus is a brain structure with a very high density of norepinephrine-containing neurons and numerous projections to other brain regions. Various techniques to alter its function provide one way to learn more about the function of one noradrenergic system in the brain. The system has been studied in the cat, macaque monkey, and squirrel monkey with such techniques as electrical stimulation, ablation, and pharmacological probes. Significant species differences are found in the catecholamine-containing cells in the brain stem, in addition to significant variations in behaviors. Cross-species reasoning from such studies is difficult, but a recent set of studies in nonhuman primates may be particularly relevant to animal models of anxiety.

Increasing locus ceruleus function, whether by electrical stimulation or with drugs, led to an increase in threat-associated behaviors, whereas decreasing it reduced those behaviors. Those behavioral effects are said to be consistent with the critical role of the locus ceruleus in mediating anxiety and fear.

It has been argued that the behavioral measures associated with the locus ceruleus function in primates are the same ones that change with environmental stimuli associated with fear in humans, and that they are decreased by diazepam (Valium). However, this approach does not permit a distinction between fear as a response to an externally threatening situation and anxiety, which typically is less related to a specific environmental precipitant. The problem arises in much of the literature on animal modeling of anxiety, where the terms “anxiety,” “fear,” and “learning” are often used interchangeably. The locus ceruleus is essential but not sufficient for the behavioral and physiological expression of anxiety and can be likened to an alarm system.

Studies of *Aplysia* It has been proposed that not only anxiety as a general state, but several specific subcategories of anxiety, can be modeled in the sea snail, *Aplysia Californica*, and that the molecular basis of anxiety can be studied in that type of animal preparation because of its relatively simple nervous system. The approach has attracted interest because it offers the possibility of more direct approaches to studying the cellular mechanisms of behavior. Behavior in such a preparation has a very different meaning from that in some of the previous models. It is not social behavior, and it is closest to a variant of conditioning paradigms that has been used for some time in the animal modeling field. In *Aplysia*, classically conditioned fear has been said to model anticipatory anxiety, and what is called long-term sensitization to model chronic anxiety.

Many workers have speculated about the relation of aversive conditioning paradigms in animals to human anxiety. The situation in which the conditioned stimulus serves as a cue for the occurrence of the unconditioned stimulus and various behavioral changes take place, presumably in anticipation of that stimulus, has been likened by investigators to anticipatory or signaled anxiety. In the sea snail, for example, exposure to the extract of shrimp elicits the withdrawal and reflex responses. The term *chronic anxiety of long-term sensitization* has been widely used to describe the state when there is repeated exposure to the unconditioned stimulus alone without any cueing or prior exposure to the conditioned stimulus. Researchers have speculated about the role of the unpredictability or uncontrollability as a factor in mediating that response.

Social Manipulations The initial stage of reaction to separation, historically labeled the *agitation* or *protest stage*, has been more recently conceptualized in the context of anxiety. The first phase is characterized by the infant's being very active behaviorally. Such infants have marked activation of the pituitary adrenal system and an increase in the enzymes involved in catecholamine synthesis. Those findings and others support the view of separation as a very powerful event from both behavioral and neurobiological standpoints, and they are consistent with a large body of literature regarding the behavioral and biological effects of a variety of stressors. To what extent the initial stage of reaction responds to anxiolytic pharmacological agents and how specific the response may be are not known.

Phobias and Other Anxiety Disorders Fear is an emotion produced by present or impending danger; the cause of fear is apparent. Anxiety, on the other hand, is an emotion of which the cause is vague or less understandable. Fear can lead to one's either freezing up or becoming mute. Much literature reports the same thing (e.g., rats freezing up in an open field) as being caused by stress.

The relation between conditioning models and theories, and anxiety and phobia models in animals has a number of shortcomings when extended to human phobias. Many animal experiments that assume conditioned fear, as well as avoidance conditioned by trauma, are models of human phobic (or anxiety) reactions. Although it is true that such induction techniques do produce fear of relatively specific stimuli and enable one to study the variables that are important for learning about fear in humans, rarely can the initiation of a human phobia (or anxiety) be ascribed to a definable event (i.e., to a definable unconditioned stimulus).

It is contended that research on animals has little to impart about human anxiety, which exists internally and symbolically, often without observable motor or autonomic concomitants. Anxiety is difficult to define. Some investigators feel that Pavlovian or Skinnerian conditioning paradigms are useless for modeling human phobias (anxieties) because human phobias and anxiety do not fit into conditioning language or paradigms. It has been argued that the language of conditioning makes assumptions about cause and treatment that are not borne out by clinical events.

One approach to the study of phobias involved the development of two lines of Pointer dogs. One line was bred for fearfulness and lack of friendliness toward people and the other for the opposite characteristics. The basic hypothesis was that heredity would largely determine many behavioral characteristics of the dog, including susceptibility to breakdown under acute and chronic stress. Through the process of selection and inbreeding, it was possible to establish the two lines of dogs and to study their behavior on a number of parameters. Throughout 10 generations, about 80 percent of each litter were similar in temperament to the parents. The phobic line of dogs was extremely timid, avoided humans, and showed decreased exploratory activity. They showed an excessive startle response, and had a slower heart rate and an increased incidence of atrioventricular heart block. Even the dogs with the most severe disturbance could learn operant conditioning bar pressing, but it was necessary to facilitate that process with benzodiazepines—the most efficacious drugs. Both amphetamines and cocaine disrupted the behavioral responses of genetically nervous dogs to a far greater extent than the stable dog.

There has been some recent work relevant to developing animal models of generalized anxiety disorder and panic disorder that should help to clarify the possible roles of neurotransmitter systems as well as to alter developmental experiences and how those two might be related in the causes of these widespread and disabling illnesses. The studies involve use of lactate to induce acute endogenous distress in nonhuman primates. Since lactate has been reported to induce panic disorder in humans, if it also does so in animals it may be possible to do studies that are more directed toward the mechanisms of those disorders. Additional animal models for anxiety have been reported in the context of the increasingly complex serotonergic receptor systems. There has also been animal work relevant to other ethologically based animal models of the various anxiety disorders.

PSYCHIATRY AND ETHOLOGY

Several authors have suggested that work in the biological science of ethology might be useful in psychiatry. John Bowlby has been taking an ethological approach to psychiatry for many years and as early as 1957 suggested that such an approach might be useful in psychiatric research. Other authors who have written about the interface between ethology and psychiatry generally approached the topic by defining ethology, presenting its vocabulary, describing its methodology, and suggesting conceptual and specific applications of ethological findings to psychiatry.

The roots of ethology lie in the natural science of biology, in particular, in zoology. The principal philosophical tenet is a naturalistic one, that is, that studies of behavior should take place in natural settings. Its origin in biology and its emphasis on the study of behavior in context have led to ethology being largely an observational, nonexperimental science. Several authors have contrasted that characteristic with the field of comparative psychology, which has different historical origins and emphases. Research in ethology has at times seemed indistinguishable from research in comparative psychology, as both have used similar techniques, such as studying naturally occurring phenomena, introducing experimental factors into a natural setting, and actually working in a laboratory. Ethology researchers usually consider those occasional excursions into the laboratory as attempts to refine mechanisms of behavior, while always being cognizant of the fact that the behavior evolves in a natural setting. A glossary of some key ethological terms is provided in Table 5.4-3.

Action-specific energy	Energy associated with the innate releasing mechanism, which builds up in a given behavioral pattern when it is not ''released'' by the appropriate releaser, and which builds up in the releasing mechanism in the absence of releasers. The behavior pattern is then said to be ''released'' by spontaneous activity.
Aggression	Intraspecific conflict mediated by physical attack or social signaling.
Appetitive behavior	Phase of behavior governing the active seeking of releasers and thought to be driven by action-specific energy. Also known as ''appetitive'' behavior pattern.
Consummatory response	Phase of behavior whereby the energy driving the appetitive phase is released. Includes the performance of the fixed action pattern (FAP).
Critical period	The time during which imprinting must occur, usually shortly after birth or early in life. Also, ''sensitive period.''
Displacement activity	A set of behavior patterns occurring alongside an unrelated set of behavior patterns. Typically, but not necessarily, these two behavioral systems are motivationally in conflict, and the displacement activity may serve to drive down another behavior system.
Ethology	The biological study of behavior. From the Greek ethos, meaning custom, usage, manner, habit. The study of behavior is considered to have a broader, more rigid concept of meaning.
Fixed action pattern (FAP)	A genetically determined behavior pattern that is released by short particular or by releasers and that consists of species-specific, stereotyped movements.
Imprinting	A specialized form of learning occurring early in life and often subserving both short- and long-term attachment to the parent. Also known as ''filial imprinting.'' The critical period, and the sometimes careful distinction between imprinting and habituation, is controversial. Imprinting is particularly resistant to extinction.
Innate	Usually describing behavior patterns, in theory not influenced by experience.
Innate releasing mechanism (IRM)	Sensory mechanisms selectively responsive to specific external stimuli and responsible for triggering the stereotyped motor responses.
Instinct	A development process resulting in species-typical behavior.
Motivation	The energy or urge that is derived from an organism's state, but simultaneously activated, driven or stimulated by stimuli in object.
Habituation	Process of a habituated pattern being interrupted through repetition itself a primary learning process. Habituation, with sensitization, and modification, is known to also sensitization.

Table 5.4-3 Selected Glossary of Ethological Terms

It has been suggested that psychiatric research and practice might benefit from the careful observational techniques ethologists employ in describing specific behavioral patterns and the context in which the behaviors occur. Those techniques would be applicable to both nonverbal and verbal behavior, as well as to the communicative aspects of each. A more general methodological issue relates to the use of the scientific method in research. In particular, it has been stressed that the ambiguity in the interpretation of findings in psychiatric research could be reduced by a greater use of operational definitions of behavior. One example is the definition of attachment behavior as any behavior that results in an increased proximity between two or more members of a species, rather than its being defined in more global terms that make inferences about the internal states of the individuals involved.

Another aspect of ethology that has been stressed in previous approaches to the interface between psychiatry and ethology is the phylogenetic origins of behavior. Ethologists are involved in the comparative study of behavior and derive hypotheses based on phylogenetic assumptions. In the same way that hypotheses are derived in the comparative study of anatomy, behavior can also be studied using the comparative method. The underlying assumption is that the behavioral patterns being studied have evolved as a result of mutation and natural selection in the same way as anatomical systems.

The more recent use of cybernetics, both control theory and information theory, in the understanding of behavioral systems in ethology has been paralleled to some extent in psychiatry; therefore, the cybernetic approach may be a useful theoretical bridge between the two fields.

The field of attachment systems has been frequently cited with regard to the application of ethological methodology and theory to psychiatry. The research on separation and the relationship between separation and depression is a corollary of the work on attachment systems. Another area of ethological research commonly suggested as potentially useful in psychiatry is that of aggression, particularly in the study of hierarchical and territorial behavior.

A natural relation between psychiatry and ethology results from their parallel positions within their respective broader disciplines. The medical sciences are based principally on the biological sciences. Psychiatry is the area within medicine most concerned with the study of behavior, and ethology has an analogous position in biology. However, ethology is still not commonly taught or used in training programs in psychiatry or child psychiatry in the United States. Nikko Tinbergen has delineated three conditions that have contributed to the relative lack of communication between psychiatrists and ethologists. The first condition is communication difficulties that arise because of differences in scientific language; the second is the obvious differences in the education of students in the two disciplines; the third is the likelihood that the people who study ethology as graduate students are different from those who enter medical school and subsequently train in psychiatry. But even if the people themselves are not particularly different, the stereotypes and the expectations of them by others can inhibit communication between psychiatrists and ethologists.

In 1973 the Nobel Prize in Physiology and Medicine was awarded to three ethologists: Karl von Frisch, Konrad Lorenz, and Nikko Tinbergen. That occasion highlighted the importance of ethology for medicine and its special relevance for psychiatry in a manner not dissimilar to the relation between molecular biology and medicine.

Karl Von Frisch Von Frisch, who was born in 1886, conducted studies on changes of color in fish and demonstrated that fish were capable of learning to distinguish among several colors and that their sense of color was fairly congruent with that of human beings. He later went on to study the color vision of bees.

Von Frisch's subsequent research was almost entirely concerned with the behavior of bees, and he is most widely known for his analyses of how they communicate with each other, that is, their language or what is known as their *dances*. His description of the complex behavior of bees prompted an investigation of information systems of other animal species.

Konrad Lorenz Born in 1903, Lorenz is known for his studies of animals, mainly birds, which he allowed to remain free but trained so that they would not be disturbed by his presence. Lorenz was a systematic observer who brilliantly described behavioral traits exhibited by animals. Many of his conclusions have been verified in experimental studies by himself and by other researchers. His interest focused from the beginning on instinctive actions, certain movements performed in a prescribed manner and provoked by certain key stimuli; these forms of behavior are now called *fixed motor patterns*. Lorenz showed that, in several animal species, a fixed motor pattern proceeds automatically once it has been provoked. It seems to be genetically programmed and, once started, is not affected by the environment. Fixed motor patterns presumably develop as a result of evolution, through the pressure of selection. Lorenz studied those patterns in several species, including jackdaws, ducks, and geese.

According to Lorenz, fixed motor patterns are provoked by stimuli specific to each pattern, which are called *key stimuli*. Those stimuli can be assumed to correspond to a particular organization of the central nervous system, originally called *das angeborene auslosende Schema*, and later, at Tinbergen's suggestion, the *innate releasing mechanism*. This mechanism is assumed to react to key stimuli by prompting or releasing corresponding fixed motor patterns. Lorenz has emphasized that not only instinctive actions but all kinds of learning have their basis in the genetically programmed equipment of the individual.

Lorenz is perhaps best known by psychiatrists for his studies of *imprinting*. During a certain short period of development, a young animal is highly sensitive to a certain type of stimulus that then, but not at other times, provokes a specific behavior pattern. Lorenz described how newly hatched goslings are programmed to follow a moving object, whereupon they rapidly become imprinted to follow it and possibly similar objects. Typically, the mother is the first moving object the young one sees, but should it see something else first, the gosling will follow it. For instance, a gosling imprinted by Lorenz followed him and refused to follow a goose (Fig. 5.4-3). Imprinting is an extremely important concept for psychiatrists to understand in their effort to link early developmental experiences with later behaviors.

FIGURE 5.4-3 In a famous experiment, Lorenz demonstrated that goslings would respond to him as if he were the natural mother. (Reprinted with permission from Hess EH: Imprinting: An effect of an early experience. Science *130:*134, 1959.)

Lorenz also studied the forms of behavior that function as sign stimuli (i.e., as social releasers) in communications between individuals of the same species. Many of the signals have the character of fixed motor patterns in that they appear automatically and provoke equally automatic reaction in other members of the species.

Lorenz is also well known for his interest in problems of aggression. He has written about the practical function of aggression, such as the defense of their territory by fish and birds. Aggression among members of the same species is common, but Lorenz has pointed out that, in normal conditions, it seldom leads to killing or even to serious injury. Although the animals attack one another, a certain balance appears between tendencies to fight and flight, with the tendency to fight being strongest in the center of the territory and the tendency to flight strongest at a distance from the center.

In many of his works Lorenz has tried to draw conclusions from his ethological studies of animals that can also be applied to human problems. Many of his suggestions are by now well known and provocative. The postulation of a primary need for aggression in humans, cultivated by the pressure of selection, is a primary example. That need might have served a practical purpose at an earlier time when human beings lived in small groups that had to defend themselves from other groups. Competition with neighboring groups became the most important factor of selection. However, Lorenz has pointed out how that need has survived the advent of weapons that can be used not merely to kill individuals but to wipe out all human beings.

Nikko Tinbergen Tinbergen, who was born in 1907, conducted a series of experiments to analyze various aspects of animal behavior. He also was successful in quantifying behavior and in obtaining measures of the power or strength of different stimuli in eliciting specific behavior. Tinbergen's first studies were of the digger wasp, *Philantus.* He determined that those insects dig individual nests into which they deposit captured bees to nourish their larvae. He was also able to show that they find their way back to their nests with the aid of various landmarks. Above all, they rely on their visual sense and learn the landmarks by means of an endogenously programmed aerial circuit.

Tinbergen's well-known studies of the various key stimuli that can provoke fixed motor patterns and of how they work together have been very valuable. Different stimuli can provoke the same motor pattern with different degrees of intensity. Particularly elegant is his analysis of the properties of the beak of the herring gull, which encourages its young to solicit food. For example, where the beak is narrow and yellow underneath, there is a contrasting red patch against which the young birds peck. By using dummies of different shapes and colors, and with different degrees of contrast, Tinbergen was able to measure the different degrees of force with which the stimuli prompted the young to peck.

Tinbergen's discovery of so-called displacement activities represents another key contribution. Those activities have been studied mainly in birds. For example, in a conflict situation, when the need for fight and the need for flight are of roughly equal strength, birds sometimes do neither. Rather, they display behavior that appears to be irrelevant to the situation; for example, a herring gull defending its territory may start to pick grass. Displacement activities of that kind vary according to the situation and the species concerned. It is well known that human beings can engage in displacement activities when under stress.

In one of his later works, Tinbergen and his wife studied early-childhood autism. They began by observing the behavior of normal children and children with autism when they meet strangers, employing the techniques used in observing animal behavior. In particular, they observed the conflict that arises in animals between fear and the need for contact and noted that it can lead to behavior that is similar to that of children with autism. They hypothesized that in certain specially predisposed children, fear can greatly predominate and can be provoked by stimuli that normally have a positive social value for most children. This innovative approach to studying infantile autism opened up new avenues of inquiry. Although the conclusions about preventive measures and treatment must be considered tentative, the methodology illustrates another way in which ethology and clinical psychiatry can relate to each other.

Future Directions Aspects of the interaction between psychiatry and ethology can be broadened in a number of ways, the first of which is methodology. Recent years have seen a beginning in the study of child behavior by ethological techniques similar to those used during the past quarter of a century in the study of primate behavior and involving operational definitions of specific behavior and the careful observation and recording of their occurrence.

The second aspect is the relative differences in interest in normal and abnormal behavior. The focus in medicine and psychiatry tends to be on the abnormal, whereas ethologists generally study the parameters of normal behavior for a given species. In a general sense, both ethologists and psychiatrists are interested in differences in adaptation—the ethologist in species differences as related to varying ecological niches and the psychiatrist in intraspecies differences in response to varying life situations. There has been a growing awareness in recent years of deficiencies in the objective understanding of normal human development, and more interaction with ethology could be useful in improving that situation.

The third issue that an interaction with ethology will force psychiatry to confront is that of the uniqueness of the human species. Psychiatry often assumes that humans are so different from other species that the study of those species is of little use in understanding human behavior. Although a phylogenetic point of view is acknowledged in human anatomy and physiology, it is certainly not emphasized in psychiatry.

The fourth aspect to be considered is the size of the behavioral units of interest to each area. Historically, ethologists have been interested in thoroughly studying the specifics of behavior to the point of a microscopic dissection, and, at the other extreme, through a more global approach. Psychiatry has been more interested in the behavior that takes place between these two extremes; that is, psychiatrists are not concerned about specific measurements of facial expression or difficulties in adaptation human beings are confronting today, but with the general behavior of individuals, families, and groups. Exceptions to the generalization include behavior modification at the more specific end of the spectrum and general systems theory at the more global end.

Inasmuch as ethology is a biological science, one might expect that the area of biological psychiatry is where the integration with ethology is most likely to occur. However, with the exception of the use of certain animal models for the understanding of the mechanism of action of a few psychopharmacological agents, biological psychiatrists have been among the least interested in the rich findings of that particular area of biology. Rather, biological psychiatry has tended to be quite reductionistic with its use of animal models. There is a growing body of literature on the effect of psychopharmacological agents in various species, but without much consideration of such factors as the phylogenetic status of the animal, its prior experience, and the effect of these agents on the animal in a natural setting.

A school of thought within psychiatry that has interesting parallels with an ethological perspective is that of psychoanalysis. The libidinous and aggressive drives about which Sigmund Freud and others have written could be viewed as analogous to some of the behavioral states studied in ethology. Much ethological literature deals with courtship and mating behavior, as well as aggression, whether expressed in hierarchical behavior or in territoriality. Much of Tinbergen's work deals with the problem of achieving reproductive success when tendencies for courtship and tendencies of aggression occur simultaneously between prospective mates. Bowlby has articulated the natural relationship between psychoanalytic and ethological thought, and his work demonstrates the usefulness of doing so. It has increased knowledge of the specific components of attachment behavior and has broadened the understanding of their existence.

Ethology seeks to understand behavior within an evolutionary or comparative framework and to use descriptive assessment techniques. There has been some interest in applying ethological descriptive techniques to assess social behaviors in psychiatric patients. For example, do patients with different types of psychiatric disorders show differences in their nonverbal social behaviors? Do those changes reflect the underlying illness or do they represent some key feature of the disorder? The use of some of the methods developed by ethologists would permit more definitive and potentially more specific characterization of disorders along those lines. Conventional psychiatric rating scales used in clinical research studies focus on individual behaviors, whereas the supplementation with ethologically based scales would permit the assessment of a broader range of key aspects of a person's functioning. For example, studies in depression have documented a reduction of the input and output of socially meaningful information, which in turn further isolates the person from others. Are such phenomena secondary to the illness (i.e., state related), or could they be evidence of traits that may characterize the baseline behaviors of such persons who may be prone to develop such disorders? These questions are parallel to the ones being asked in biological psychiatry, where state-trait distinctions are very common and where there is great interest in trait markers. The integration of data in man and animals using related rating scales and units of measurement is also important.

In a 1994 editorial in *Biological Psychiatry*, Brian Kirkpatrick highlighted certain areas as illustrating the important interface between psychiatric disease and the neurobiology of social behavior. These include:

1. Depression. Loss of significant social bonds increases the risk of major depressive disorders and there are human and animal models of the enduring neurobiological alterations associated with early social events.
2. The impairment of social relationships found in autism cannot be fully accounted for by the cognitive impairments found in these patients.
3. The neurobiology of social behavior may provide a useful conceptual framework for developing animal models relevant to the deficit syndrome of schizophrenia.
4. It may be possible to delineate a neuroanatomy of social behavior, including such fundamental behaviors as mating, paternal and maternal behavior, and the

processing of broad range of visual social stimuli.

Kirkpatrick's concluding paragraphs are worthy of a direct quote:

> Animal models of relevance to psychiatric disease are likely to account for specific aspects of human illness, and not for all the features of a disorder. For instance, models of social behavior relevant to schizophrenia might relate to the sociality found in patients with the defect syndrome, but not to the psychotic features they share with nondeficit schizophrenic patients. Similarly, social models might relate to the asociality of autism, but not to the cognitive impairments or stereotypies exhibited by autistic patients.
>
> Animal models must be used with caution. Human social behavior has unique aspects such as elaborate, flexible language, and the human brain is more complex than that of even our closest primate relatives. A superficial resemblance between human and animals behavior may also disguise great differences in the significance of the behavior. Nonetheless, current clinical research technologies, such as positron emission tomography and magnetic resonance imaging, are inappropriate for study of the structures that have been implicated in the control of social behavior in animals, due to the small size of those structures. With animal models, it is possible to use other methods to study such small structures, which are powerful determinants of behavior.
>
> As knowledge of the neural basis of social behavior progresses, this field will provide a basis for the development of rational hypothesis that could guide postmortem, pharmacological, and other clinical studies related to depression, autism, aspects of schizophrenia, and perhaps other problems.

SUGGESTED CROSS-REFERENCES

Section 1.5 is a discussion of amino acid neurotransmitters. Section 1.9 covers basic electrophysiology. Learning theory is the subject of Section 3.3. Normal child development is covered in Section 32.2. Chapter 12 focuses on schizophrenia. Chapter 14 covers mood disorders, and sleep disorders are reviewed in Chapter 21. Biological therapies are the subject of Chapter 31.

SECTION REFERENCES

Bachevalier J, Mishkin M: An early and late developing system for learning and retention in infant monkeys. Behav Brain Res *20:*249, 1986.

Braff DL: Information processing and attention dysfunctions in schizophrenia. Schizophr Bull *19:*233, 1993.

Castro WL, Matt KS: Neuroendocrine correlates of separation stress in siberian dwarf hamster (*Phodopus sungorus*). Physiol Behav *61:*477, 1997.

Coplan JD, Trost RC, Ownes MJ, Cooper TB, Gorman M, Nemeroff GB, Rosenblum L: Cerebrospinal fluid concentrations of somatostatin and biogenic amines in grown primates reared by mothers exposed to manipulated foraging conditions. Arch Gen Psychiatry *55:*473, 1998.

Clifford JJ, Waddington JL: Heterogeneity of behavioral profile between three new putative D3 dopamine receptor antagonists using an ethologically based approach. Psychopharmacology *136:*284, 1998.

Dixon AK, Fisch HU, Huber C, Walser A: Ethological studies in animals and man, their use in psychiatry. Pharmacopsychiatry *22:*44, 1989.

*Gardner R: Mechanisms in manic-depressive disorder: An evolutionary model. Arch Gen Psychiatry *39:*1436, 1982.

Gessa GL, Pani L, Serra G, Fratra W: Animal models of mania. Adv Biochem Psychopharmacol *49:*43, 1995.

Geyer MA, Athina M: Animal models of psychiatric disorders. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Ginsberg SD, Hof PR, McKinney WT, Morrison JH: Quantitative analysis of tuberoinfundibular tyrosine hydroxylase- and corticotropin-releasing factor-immunoreactive neurons in monkeys raised with differential rearing conditions. Exp Neurol *120:*95, 1993.

Harris JC: Experimental animal modeling of depression and anxiety. Psychiatr Clin North Am *12:*815, 1989.

*Henn FA, Johnson A, Edwards E, Anderson P: Melancholia in rodents: Neurobiology and pharmacology. Pharmacol Psychopharmacol Bull *21:*443, 1985.

Kirkpatrick B: Psychiatric disease and the neurobiology of social behavior. Biol Psychiatry *36:*501, 1994.

Kornetsky C, Markowitz R: Animal models and schizophrenia. In *Model Systems in Biological Psychiatry*, D Ingle, H Shein, editors. MIT Press, Cambridge, MA, 1975.

Kraemer GW: Causes and changes in brain norepinephrine systems and later effects in response to social stressors in rhesus monkeys: The cascade hypothesis. In *Antidepressants and Receptor Function, CIBA Foundation Symposium*, Nos. 126 and 127. Wiley, New York, 1986.

Kraemer GW, Ebert MH, Schmidt DE, McKinney WT: A longitudinal study of the effect of different social conditions on cerebrospinal fluid norepinephrine and biogenic amine metabolites in rhesus monkeys. Neuropsychopharmacology *2:*175, 1989.

*Kraemer GW, Ebert MH, Schmidt DE, McKinney WT: Strangers in a strange land: A psychobiological study of infant monkeys before and after separation from real or inanimate mothers. Child Dev *62:*548, 1991.

Li TK, Lumeng L, Doolittle DP: Selective breeding for alcohol preference and associated responses. Behav Gen *23:*163, 1993.

Lister RG: Ethologically based animal models of anxiety disorders. Pharmacol Ther *46:*321, 1990.

Marks I: Phobias and obsessions: Clinical phenomena in search of a laboratory model. In *Psychopathology: Experimental Models*, J Maser, M Seligman, editors. Freeman, San Francisco, 1977.

Matthews K, Wilkinson LS, Robbins TW: Repeated maternal separation of preweaning rats attenuates behavioral responses to primary and conditioned incentives in adulthood. Physiol Behav *59:*99, 1996.

Matthyse S, Haber S: Animal models of schizophrenia. In *Model Systems in Biological Psychiatry*, DI Ingle, H Shein, editors. MIT Press, Cambridge, MA, 1975.

McGuire MT, Fairbanks LA: Ethology: Psychiatry's bridge to behavior. In *Ethological Psychiatry: Psychopathology in the Context of Evolutionary Biology*, MT McGuire, LA Fairbanks, editors. Grune & Stratton, New York. 1977.

McGuire MT, Marks I, Nesse RM, Troisi A: Evolutionary biology: A basic science for psychiatry. Acta Psychiatrica Scand *86:*S9, 1992.

McKinney WT: Basis of development of animal models in psychiatry: An overview. In *Anima Models of Depression*, GF Koob, CL Ehlers, DJ Kupfer, editors. Kirkhauser, Basel, 1989.

*McKinney WT, editor: *Models of Mental Disorders*. Plenum, New York, 1988.

Mishkin M, Appenzeller T: The anatomy of memory. Sci Am *256:*80, 1987.

Mishkin M, Malamut B, Bachevalier J: Memories and habits: Two neural systems. In *Neurobiology of Learning and Memory*, G Lynch, JL McGaugh, NM Weinberger, editors. Guilford, New York, 1984.

*O'Neil BJ, Kline JA, Burkhart K, Younger J: Research fundamentals: V. The use of laboratory animal models in research. Acad Emerg Med *6:*75, 1999.

Overstreet DH: The flinders sensitive line rats: A genetic animal model of depression. Neurosci Biobehav Rev *17:*51, 1993.

Pare WP: Open field, learned helplessness, conditioned defensive burying, and forced-swim tests in WKY rats. Physiol Behav *55:*433, 1994.

Pare WP, Redei E: Depressive behavior and stress ulcer in Wistar Kyoto rats. J Physiol *87:*229, 1993.

Porsolt RD: Pharmacological models of depression. In *The Origins of Depression: Current Concepts and Approaches*, J Angst, editor. Springer-Verlag, Berlin, 1983.

Reeke GN Jr, Sporns O: Behaviorally based modeling and computational approaches to neuroscience. Annu Rev Neurosci *16:*597, 1993.

Reneric JP, Lucki I: Antidepressant behavioral effects by dual inhibition of monoamine reuptake in the rat forced swimming test. Psychopharmacology *136:*190, 1998.

Richardson JS: Animal models of depression reflect changing views on the essence and etiology of depressive disorders in humans. Prog Neuropsychopharmacol Biol Psychiatry *15:*199, 1990.

Segal DS, Geyer MA: Animal models of psychopathology. In *Psychobiological Foundations of Clinical Psychiatry*, LL Judd, PM Groves, editors. Lippincott, Philadelphia, 1985.

Siegel SJ, Ginsberg SD, Hof PR, Foote SL, Young WG, Kraemer GW, McKinney WT, Monison JH: Effects of social deprivation in prepubescent rhesus monkeys: Immunohistochemical analysis of the neurofilament protein triplet in the hippocampal formation. Brain Res *619:*299, 1993.

Snyder S: Amphetamine psychosis: A model schizophrenia mediated by catecholamines. Am J Psychiatry *130:*161, 1933.

Sunderland G, Friedman S, Rosenblum LA: Imipramine and alprazolam treatment of lactate-induced acute endogenous distress in nonhuman primates. Am J Psychiatry *146:*1044, 1989.

Swerdlow NR, Braff DL, Taaid N, Geyer MA: Assessing the validity of an animal model of deficient sensorimotor gating in schizophrenic patients. Arch Gen Psychiatry *51:*139, 1994.

Wilner P, Benton D, Brown E, Cheeta S, Davies G, Morgan J, Morgan M: Depression increases craving for sweet rewards in animal and human models of depression and craving. Psychopharmacology *136:*272, 1998.

*Wilner P: The validity of animal models of depression. Psychopharmacology *83:*1, 1984.

Wilner P: Animal models of depression: Validity and applications. Adv Biochem Psychopharmacol *49:*19, 1995.

Weed MR, Gold LH: The effects of dopaminergic agents on reaction time in rhesus monkeys. Psychopharmacology *138:*33, 1998.

Yehuda R, Antelman SM: Criteria for rationally evaluating animal models of posttraumatic stress disorder. Biol Psychiatry *33:*479, 1993.

CHAPTER 6. THEORIES OF PERSONALITY AND PSYCHOPATHOLOGY

6.1 PSYCHOANALYSIS

GLEN O. GABBARD, M.D.

Psychoanalysis has been one of the most influential intellectual movements in twentieth-century culture. It can be regarded as a theory of the personality, a method of investigation, a scientific discipline, and a form of treatment. Only a small subgroup of psychiatric patients receive formal psychoanalysis as a treatment, but the principles derived from psychoanalytic theory are broadly applicable to most patients seen in a general psychiatric practice. A systematic understanding of the unconscious mental life of the patient may illuminate reasons for noncompliance with a treatment plan, difficulties in establishing a therapeutic alliance with a clinician, and a patient's lack of interest in being helped. Hence an overarching psychoanalytically based framework is useful in the practice of psychiatry regardless of which specific treatment is being conducted.

It has been over 100 years since Sigmund Freud developed psychoanalysis as an investigative and therapeutic tool, and no body of knowledge approaches psychoanalytic theory in terms of its capacity to explain the mysteries of the human mind. Freud is ranked as the most cited twentieth-century thinker in all academic fields combined. Psychoanalysis has evolved to the point where Freud might be hard pressed to recognize it if he were alive today. Several fundamental controversies have sparked an intellectual fervor that is enriching psychoanalytic dialogues around the world. Psychoanalytic theory is now largely a pluralistic endeavor in which several useful theories compete with one another as explanations for specific human behaviors. Among the basic controversies in contemporary psychoanalysis are whether or not patients' difficulties are based in conflict or deficit, whether psychoanalysis is fundamentally a one-person (intrapsychic) or a two-person (interpersonal) psychology, whether the database in psychoanalysis is limited to a patient's subjective experience versus encompassing the analyst's experience as well, and whether psychopathology originates from traumatic reality events or largely from the role of fantasy in distorting real events.

The current controversies in theory and in the possible modifications of technique that follow from those theories can be traced to Freud's monumental work in the latter part of the nineteenth century and the first 40 years of the twentieth century. Freud himself modified his theory and practice repeatedly over the course of his career; nevertheless, he always preserved certain fundamental tenets that are still highly relevant today.

Freud articulated that any technique calling itself psychoanalytic had to involve the principles of *transference* and *resistance*. Transference is the patient's displacement onto the analyst of early wishes and feelings toward other people. Although patients may begin analysis with the understanding that their analysts are helpful professionals, they soon begin to experience the analyst as a significant figure from the past, such as a parent. Contemporary analysts also recognize that transference is partly affected by the analyst's real characteristics, so that the transference perception is an admixture of the real relationship with the analyst and relationships with figures from the past whom patients unconsciously bring into the present.

Freud originally identified resistance when he asked his patients to use the technique of *free association*, which involved saying whatever came to mind without censoring the thoughts. Freud noted that many patients could not carry out these instructions and in some cases would “go blank.” Ultimately he came to realize that such patients were unconsciously opposing the treatment every step of the way. Many patients are ambivalent about getting better and about cooperating with the clinician who is there to help them. Modern psychiatrists see resistance in the form of patients who forget to fill their prescriptions, refuse to take medications, miss their appointments, and otherwise engage in noncompliant behaviors.

These principles of technique, transference, and resistance derive from basic theoretical notions inherent in psychoanalysis.

THE UNCONSCIOUS

Foremost among these basic notions would be Freud's idea that much of mental life is unconscious. People like to think they are in control of their lives, but factors outside of conscious awareness often influence behavior. A simple example is the New Year's resolution that people have every conscious intention to keep but consistently undermine as they try to carry out their pledge. The unconscious is also revealed through slips of the tongue, termed *parapraxes*, which unwittingly exhibit to everyone what is going in the speaker's unconscious. A politician, who wants to say that his source is unimpeachable, instead blurts out, “I have this information from an impeachable source.” He blushes as everyone in the audience laughs. Nonverbal behaviors may also provide a glimpse of the individual's unconscious.

Dreams are one of the primary sources of information about unconscious conflict. Freud noted that long-forgotten memories reemerged in the process of the treatment, an observation that led him to conclude that the human mind has a form a censorship that deems certain memories, thoughts, and feelings unacceptable. The material is *repressed* or buried in the unconscious, so the person is no longer consciously aware of the phenomena that have undergone repression during conscious waking life. However, when the individual is asleep, the manifest or overt content of dreams often deals with material that would be entirely unacceptable in waking life. Freud demonstrated that a repressed childhood wish is often a motivating force that generates the content of the dream.

PSYCHIC DETERMINISM

Another central tenet of psychoanalytic theory is that behavior has meaning. Indeed, much of psychoanalytic clinical work can be described as the co-creation of meanings by patient and analyst. The meaning is partly determined by unconscious conflicts. Even when symptoms or behavior are influenced by biological factors and cultural considerations, the domain of psychoanalysis always involves meanings, and psychoanalysts are interested in how a patient uniquely interprets thoughts, feelings, and external events based on that individual's past experiences. A central feature of psychic determinism is that childhood experiences are repeated throughout life and are very much alive in terms of the adult's motivations, conflicts, and wishes. As Wordsworth noted, “The child is father of the man.”

In contemporary psychiatry, psychoanalytic theory and technique coexist with impressive advances in the neurosciences, knowledge of cultural influences on behavior, evolutionary theories, and social theories. To a large extent psychoanalytic theory has been enriched by engaging in a broad dialogue with these other disciplines, and the theory continues to evolve in an era of the greatest intellectual ferment the field has known since its inception 100 years ago.

HISTORY

The history of psychoanalysis is inextricably linked to the life of Sigmund Freud. He was born on May 6, 1856, in Freiburg, a small town in Moravia, which has since

become a part of Czechoslovakia. When Freud was 4 years old, his father, a Jewish wool merchant, moved the family to Vienna. He grew up there, was educated there, and practiced there almost his entire life. In 1938 Freud fled the Nazis by moving to England, where he died the next year (Fig. 6.1-1, Fig. 6.1-2, Fig. 6.1-3, Fig. 6.1-4, Fig. 6.1-5, Fig. 6.1-6, Fig. 6.1-7 and Fig. 6.1-8).

FIGURE 6.1-1 Sigmund Freud as a boy in 1870. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-2 Sigmund Freud and his bride, Martha Bernays, in 1886. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-3 Sigmund Freud in 1903, age 47. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-4 Sigmund Freud in 1911. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-5 Sigmund Freud with his grandson in 1922. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-6 Freud in 1936. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

FIGURE 6.1-7 Sigmund Freud at work in his office in 1936. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

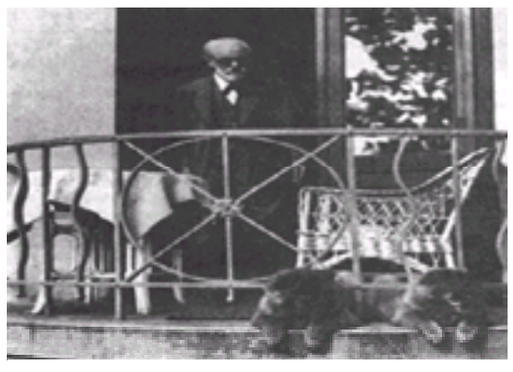

FIGURE 6.1-8 Sigmund Freud with two pet chows in 1938. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

Freud's Career Freud was first and foremost an empirical scientist. He was convinced that the key to unlocking the secret of mental processes was to be found in the study of brain physiology. The dominant approach to human physiology in Europe during Freud's formative years was the Hermann Helmholtz school of physiology. According to the deterministic principles of that school, physiological and mental processes resulted from causal laws and sequences that could ultimately be reduced to the tenets of physics, such as the principles of inertia and conservation of energy. Freud was also influenced by the theories of Charles Darwin, which were much discussed in scientific circles in his day.

Coexisting with Freud's tendency to embrace scientific empiricism was his lifelong fascination with literature, particularly the writings of Johann Wolfgang von Goethe and William Shakespeare. Freud had a deep appreciation and respect for the word as a result of being raised in the Jewish tradition. Through his immersion in literature and religion, Freud became familiar with the complexities of the human psyche and the roles of symbolism and meaning in understanding human nature. Freud's thinking always displayed a romantic element, and it was no accident that he was ultimately awarded the Goethe Prize for literature.

Medical Training Freud's experience in medical school laid the foundation for his subsequent scientific orientation. It was during those years that he was exposed to the ideas of Darwin and Helmholtz. The intellectual currency of the day was scientific empiricism; an emphasis on measurement and observation was replacing the mysticism and romanticism that had pervaded scientific thought in central Europe in the first part of the nineteenth century.

During his medical school years Freud was profoundly influenced by two mentors, Ernst Brücke and Theodore Meynert. Those two figures—along with Sigmund Exner, who worked in Brücke's laboratory—dominated the physiological research in Europe during that era. Freud also worked in Brücke's laboratory and viewed Brücke in particular as a role model whose integrity and scientific discipline Freud strove to emulate.

Brücke, Meynert, and Exner all endorsed the idea that the nervous system operates through the transmission of quantitatively variable excitations from afferent to efferent nerve endings. Because the chemical basis of neural conduction was poorly understood, Brücke assumed that conduction is electrical in nature. Central to Brücke's thinking was the notion that the mind and the body are organized along principles of psychophysical parallelism. The reflex arc was his model of neural functioning. In other words, no spontaneous central activity exists in the nervous system; it merely functions as a passive instrument that remains quiescent until stimulated by exogenous energies. The end result of the stimulus is the reduction of incoming irritation to a minimum. That physiology of force and energies, drawing heavily from, the doctrine of conservation of energy, was to have a far-reaching effect on Freud's own thinking.

Medical Career After graduation from medical school, Freud continued to work in Brücke's laboratory for a year. There Freud developed his overriding scientific ambition to apply the principles of Helmholtz and Brücke to the nervous system. He was convinced that both psychopathological manifestations and normal mental functioning could ultimately be explained in terms of the forces and energies that regulate the nervous system.

Freud's research was interrupted when he realized that laboratory work could not generate sufficient income to support a family. He had fallen in love with Martha Bernays, and he was forced to enter medical practice to provide an adequate standard of living for his new bride. In 1882 he began work as a general physician in the Vienna General Hospital. After a stint in the surgical service, he served in Theodore Meynert's psychiatric clinic. Although Freud had some reservations about Meynert's clinical competence, his study of Meynert's amentia (acute hallucinatory psychosis) had a lasting effect on him and contributed to the concept of wish fulfillment, which later became a crucial part of Freud's theory of the unconscious.

Neurological Career Freud's exposure to neuroanatomy and neuropathology under the charismatic leadership of Meynert led him to choose neurology as a specialty. In 1885 he received a grant that allowed him to study for 19 weeks at the Salpêtrière in Paris. There his career was powerfully shaped by the great French neurologist, Jean-Martin Charcot.

Under Charcot's influence, Freud became fascinated with the problem of hysteria. Whereas most neurologists did not take hysterical phenomena seriously, Charcot viewed them as worthy of careful study. He viewed them as related to a congenital degeneration of the brain, but he placed considerable emphasis on the role of hypnosis in the treatment of hysterical disorders. When Freud observed the precipitation of hysterical manifestations through the use of hypnotic suggestion, he

began entertaining the possibility that hysteria has a psychological origin.

When Freud returned to Vienna in 1886, he was determined to pursue his dual interests in hypnosis and neurology. Hungering for more knowledge about the clinical applications of hypnosis, he journeyed to Nancy, France, to study with Ambroise-Auguste Liébault, who was attempting to remove neurotic symptoms through hypnotic suggestions. Liébault's associate, Hippolyte Bernheim, was also studying the characteristics of suggestibility and noted that it was present in patients with neurotic disorders other than hysteria. Freud's observations of the therapeutic uses of hypnosis used by Bernheim and Liébault convinced him that powerful unconscious forces were involved in human motivation and behavior. He also learned that physicians themselves may play significant roles as instruments of psychotherapeutic change.

Freud's first major neurological work appeared in 1891. Entitled *Aphasia*, it offered a functional explanation that accounted for variants of aphasic disorders in terms of disruptions of the radiating associative pathways. That same year with Oscar Rie, a pediatrician, he was coauthor of a major work on unilateral paralysis in children. Two years later he wrote a massive monograph on children's paralysis that received major acclaim.

Project for a Scientific Psychology The period between 1895 and 1897 represented a turning point in Freud's career. Intent on anchoring psychology to neurophysiology, Freud struggled to apply the principles of the Helmholtz school to the workings of the mind. The result was his pivotal work, *Project for a Scientific Psychology*. Freud was disgusted with what he had written and wanted it destroyed. It was finally published posthumously because of the recognition that it had a far-reaching influence on Freud's subsequent ideas.

Freud's explanation of the mechanisms of the nervous system revolved around the principles of neuronic inertia, which asserted that neurons tend to divest themselves of Q, the quantity of excitation. Pain can be related to an excess of nervous excitation, and pleasure can be viewed as the end result of the discharge of that excitation. At the root of the theory was a model of the nervous system as a passive recipient of stimulation from external forces and a concept of motivation in terms of tension or drive reduction.

Influenced by the principle of conservation of energy, Helmholtz had insisted that the sum of forces must remain constant in any isolated system. Freud elaborated on that principle of constancy in his *Project* by developing the idea that mental processes constantly strive for equilibrium or homeostasis. In 1893 Freud formulated the principle as follows: “If a person experiences a psychical impression, something in his nervous system, which we will for the moment call the sum of excitation, is increased. Now in every individual there exists a tendency to diminish this sum of excitation once more, in order to preserve his health.” This principle of constancy served as the economic foundation for Freud's theory of instincts.

Although many of Freud's ideas in his Project have been confirmed by more sophisticated neuropsychological studies, he was unable to forge the grand synthesis between the psychological and the neurological to which he had aspired. Nonetheless, in a number of his other works, “Preliminary Communication” (1893) and *Studies on Hysteria* (1895), the notions of the discharge of affect and cerebral excitation clearly derive from ideas generated in *Project for a Scientific Psychology*. Moreover, the model of the mind set forth in Freud's magnum opus, *The Interpretation of Dreams* (1900), also had clear roots in his *Project*. Finally, his understanding of the pleasure-unpleasure principle was profoundly influenced by the basic theorems of his Project. As late as 1920, in *Beyond the Pleasure Principle*, the derivatives are clearly in evidence.

EVOLUTION OF PSYCHOANALYSIS

During the decade from 1887 to 1897, Freud's interests in clinical neurology gradually waned. His collaboration with Josef Breuer, a distinguished older colleague with whom he had worked at Brücke's institute of physiology, led him into the nether regions of the unconscious mind, a subject that attracted his interest away from neurology and toward the evolution of psychoanalysis. As the two clinicians struggled together to understand the mysteries of hysteria, they gave birth to psychoanalysis as a therapeutic technique, as a scientific discipline, and as a method of investigation.

Collaboration With Breuer: The Case of Anna O. and Studies on Hysteria The body of theory, knowledge, and technique that is now referred to as psychoanalysis had its origins in one typical case treated by Breuer. The patient, Bertha Pappenheim, consulted Breuer in December 1880 and continued in treatment with him until June 1882. Referred to by Breuer as “Anna O.,” she was an intelligent and attractive woman of 21 years who presented a plethora of hysterical symptoms in association with her father's fatal illness. The symptoms included serious disturbances of sight and speech, inability to ingest food, paralysis of three extremities with contractures and anesthesias, and a nervous cough. She also manifested two distinct states of consciousness: one a relatively normal young woman, the other a troublesome and naughty child. Breuer observed that the shift between the two discrete personalities seemed to be induced by some form of autohypnosis, and he was able to bring about the transition by placing Anna O. in a hypnotic state.

Breuer knew that Anna had been very attached to her father and had nursed him alongside her mother while he was on his death bed. During her altered states of consciousness, Anna could recall vivid fantasies and powerful feelings she had experienced as her father lay dying. Breuer was astonished to note that his patient's recollection of the affect-laden circumstances during which her symptoms first appeared led those same symptoms to disappear. Anna O. dubbed this process the “talking cure.” She was so taken by it that she continued to discuss one symptom after another. For example, she remembered sitting at her father's side while her mother was absent and having a fantasy or daydream about a snake. In her vision the snake was about to bite her father. She tried to ward off the snake, but her arm had gone to sleep as a result of having been draped over the back of her chair. The paralysis remained until she was able to recall the scene under hypnosis and regain use of her arm.

Breuer became enchanted with his extraordinary patient. He spent so much time with her that his wife grew jealous and resentful. Frightened by the sexual connotations of his wife's complaints, he abruptly terminated the treatment of Anna O. Several hours after that termination, he was called to Anna's bedside in the midst of a crisis. He found her in an agitated state in the throes of hysterical childbirth. Although he had been unaware of any sexual feelings toward him, the phantom pregnancy (pseudocyesis) reflected Anna O.'s intense erotic longings for Breuer. He calmed his patient down by inducing a hypnotic trance, and in a state of extreme agitation he arranged for an immediate departure to Venice with his wife for a second honeymoon.

Although Breuer found the whole experience with Anna O. to be highly disconcerting, Freud was intrigued by the power of unconscious memories and suppressed affects to produce hysterical symptoms. Breuer was reluctant to publish his account, but Freud insisted that he write it up from memory some 13 or 14 years after it had occurred. The collaboration of Breuer and Freud resulted in the publication of “Preliminary Communication” in 1893. In that document they articulated the causal linkage between psychic traumata and hysterical symptoms.

In 1895, with the publication of *Studies on Hysteria*, Breuer and Freud presented a much more sophisticated clinical and theoretical treatise on the pathogenesis and treatment of hysterical symptoms. The hysterical patient suffers from “reminiscences,” according to Freud. In other words, a repressed incompatible idea is the source of the symptomatic manifestations. The patient had experienced a childhood trauma that stirred up overwhelming feelings of an intensely unpleasant nature. The traumatic experience represented an incompatible idea to the patient and was, therefore, intentionally dissociated or repressed from consciousness. The nervous excitation associated with the incompatible idea was transformed or converted into somatic channels that produced hysterical symptoms. As a result, all that was left in conscious awareness was a mnemonic symbol that was only remotely connected with the traumatic event, often through disguised links. Freud thought that, if the memory of the traumatic experience could be brought back into the patient's conscious awareness, along with the strangulated affect associated with it, the symptoms would disappear as the affect was discharged.

Evolution of Freud's Technique

Use of Hypnosis Although Freud had used hypnosis since he opened his practice in 1887, he had initially used the technique simply as a means to remove symptoms through suggestion. When he encountered disappointing results, he shifted to the cathartic method as a result of Breuer's account of Anna O. In 1889 Frau Emmy von N. consulted Freud for treatment of a variety of hysterical complaints, including anesthesia and pain in her leg, an ovarian neuralgia, delirium, hallucinations, phobias, and abulia. For the first time Freud used Breuer's cathartic method, attempting to remove the woman's symptoms through a process of recovering and verbalizing suppressed feelings with which they were associated. The method came to be known as *abreaction*.

Freud soon became dissatisfied with the abreactive approach when he observed that the therapeutic benefits lasted only as long as the patient maintained contact with the physician. The personal relationship with the physician appeared to have greater therapeutic importance than the specific hypnotic technique. His understanding of the doctor-patient relationship was expanded by an incident in which one of his patients awoke from a hypnotic trance and threw her arms around Freud's neck. Experiences such as those, coupled with Breuer's report of Anna O., led Freud to realize that the patient's attachment to the physician had an erotic

component. However, instead of fleeing in panic from such developments, as Breuer had done, Freud investigated them as he did any other phenomena encountered in treatment. Those early encounters led Freud to discover *transference*, a concept that was to become a corner-stone of psychoanalytic theory and technique. Transference refers to the displacement onto the analyst of thoughts, feelings, and behavior originally associated with significant figures from the past.

Freud's discovery of transference contributed to his abandonment of hypnosis. In his view, hypnosis concealed aspects of the transference, so that they could not be investigated as part of the process. He also felt that hypnosis encouraged the patient to please the hypnotist, instead of learning about the origins and the meanings of symptoms. Freud also observed that many patients were simply refractory to hypnosis.

Concentration Method and Development of Free Association One of the patients who was particularly resistant to the hypnotic abreactive technique was Elizabeth von R. Freud evolved his method of concentration as a way of dealing with the refractory nature of that patient. He remembered that Bernheim had asserted that all forgotten memories could be recalled consciously if the physician asked appropriate leading questions and urged the patient to remember.

Freud asked his patient to lie down on a couch and to close her eyes (Fig. 6.1-9). He asked her to concentrate on a particular symptom and to recall any memories that might assist in understanding the origin of the symptom. Freud would then press his hand on the patient's forehead and reassure her that she would indeed recall relevant memories when he questioned her about them.

FIGURE 6.1-9 Sigmund Freud's couch in Freud Museum. (Courtesy of Menninger Foundation Archives, Topeka, KS.)

Much to Freud's credit, he allowed himself to learn from his patients. Elizabeth von R., for example, responded to the concentration method by telling Freud, “I could have told you that the first time, but I didn't think it was what you wanted.” Freud then modified his technique by informing patients to simply ignore all censorship. When Elizabeth von R. reproached him for interrupting her flow of thoughts with his questions, Freud modified his technique again by reducing the frequency of his questions so as not to interfere with the natural flow of the patient's associations.

By the late 1890s Freud had come to realize that the concentration technique was more an impediment than a facilitator. He abandoned the procedures of directing the patient's attention, placing pressure on the forehead, directing the patient to close her eyes, and asking probing questions. Instead, he had the patient lie on the couch and say whatever came to her mind, the method of *free association*, which remains a central part of psychoanalytic technique today.

Repression and Resistance In addition to Freud's discovery of transference, two other seminal psychoanalytic concepts—resistance and repression—grew out of his clinical investigations of hysterical patients in the 1890s. In the evolution of his technique from hypnosis to the concentration method to free association, Freud noted that certain patients *resisted* the therapeutic technique. Some could not be hypnotized. Others were unable to recall memories of causative significance. Still others encountered a mental block when they attempted to free-associate. Freud discerned that a stubborn refusal to cooperate was not at the root of the resistance. Many of his patients who were genuinely distressed by their symptoms were also the most resistant to the therapeutic techniques. Freud concluded that active forces in the patient's unconscious mind were excluding unpleasant thoughts or memories from conscious awareness. He referred to that active force as *repression*.

Freud's studies of hysterical symptoms convinced him that repression was a pivotal force in the process of symptom formation. He described the mechanism in the following manner: A traumatic experience or a series of experiences, usually of a sexual nature, that had occurred in childhood had been forgotten or repressed because of its painful or unacceptable nature, but the excitement associated with the incident was not extinguished, and traces persist in the unconscious in the form of repressed memories. The memories may remain without pathogenic effect until a contemporary event, such as a disturbing love affair, revives them. The so far successful repression fails at that point, leading the patient to experience what Freud referred to as *the return of the repressed*. The original sexual excitement is revived and comes to the surface in the form of a neurotic symptom.

On the basis of his understanding of repression, Freud postulated that symptoms arise from a compromise between a repressed impulse and the countervailing forces of repression. In cases of hysteria, Freud believed that impulses that were unacceptable and therefore repressed were diverted into somatic pathways, resulting in such symptoms as paralysis, blindness, and disturbances of sensation. He speculated that similar mechanisms are involved in the development of paranoid ideation and obsessive-compulsive symptoms. The main thrust of treatment then was to assist patients in the retrieval of repressed memories of sexual experiences so that the excitations attached to them could emerge into conscious awareness and be discharged through verbalization.

Theory of Infantile Sexuality Freud's treatment of hysterics during the early years of the 1890s convinced him that childhood sexual seduction plays a major role in causing the neuroses. Many of his patients reported such seductions by nursemaids, fathers, and caretakers, and Freud believed that repressed memories of actual sexual traumas created neurotic symptoms.

In the latter half of the 1890s Freud began to reconsider his seduction hypothesis. The first glimpse that he had doubts about the viability of his etiological theory was contained in a September 1897 letter to his confidant and colleague Wilhelm Fliess. Freud was concerned that patients who were treated according to this theory often discontinued treatment while those who continued had poor outcomes. Freud also recognized that his theory of the traumatic etiology of neurosis strained credulity in that it required all fathers of neurotics, including Freud's own father, to be sexually perverse abusers of their own children. Also, some of the reports from his patients sounded preposterous. Emma Eckstein, for example, reported a “memory” of encountering the devil and being tortured by him as he inserted needles into her fingers. Stories of torture and possession linked to witchcraft were common among patients with hysteria.

Another factor that influenced Freud was his allegiance to Fliess. A theory based purely on environmental trauma was at odds with Fliess's idea that there were inborn cycles that unfolded in the natural course of development. Fliess carried his view of biological determination to the extreme, suggesting that the precise date of a person's death was predetermined. Freud disagreed, however, because a purely constitutional etiology would make a psychoanalytic intervention highly questionable.

Freud resolved the dilemma by rebuilding his theory of neurosogenesis around masturbation as being the cause of hysteria. Influenced by Havelock Ellis and the late–nineteenth-century sexological discourse, Freud wrote in 1899: “Not only dreams are wish fulfillments, so are hysterical attacks. This... probably applies to every product of neurosis.... The symptom is the wish fulfillment of the repressing thought, for example, in the form of a punishment; self-punishment is the final substitute for self-gratification, which comes from masturbation.” The emphasis on masturbation was a natural bridge to a shift toward intrapsychic fantasy rather than external trauma.

Although Freud's subsequent writings indicate that he never entirely abandoned his belief that actual sexual seductions by parents took place, by the late 1890s he had clearly moved to a position of stressing childhood sexual fantasies as the core of the neuroses. By adopting a psychodynamic theory of infantile sexuality in which the child's psychosexual life occupies center stage, psychoanalysis had taken a new direction, coming into its own as a depth psychology. The child was no longer portrayed as an innocent victim of wicked adults but was seen as a sexual being with powerful fantasies and wishes. By the waning years of the nineteenth century, Freud had completed his self-analysis, proposed a dynamic theory of infantile sexuality, developed the technique of free association, postulated the origins and the pathogenesis of the neuroses, and designed a theory of the mind that included the unconscious as its centerpiece. Psychoanalysis as it is known today was clearly recognizable at that point, and as the twentieth century approached, Freud was immersed in perhaps his most monumental work, *The Interpretation of Dreams*.

THE INTERPRETATION OF DREAMS

Significance of Dreams Freud became aware of the significance of dreams when he noted that patients frequently reported their dreams in the process of free association. Through their further associations to the dream content, he learned that the dreams were definitely meaningful, even though that meaning was often hidden or disguised. Most of all, Freud was struck by the intimate connection between dream content and unconscious memories or fantasies that were long repressed. That observation led Freud to declare that the interpretation of dreams was the royal road to the understanding of the unconscious.

Freud's self-analysis also contributed to his appreciation of the significance of dreams. One of his principal methods in conducting his own analysis was to rely on dreams and his associative exploration of those dreams. Indeed, Freud used many of his own dreams as illustrative examples in *The Interpretation of Dreams*, which appeared in 1900. Considered by many to be Freud's greatest work, the book still informs clinical psychoanalytic work with dreams today. In his magnum opus Freud put forth the notion that a dream is the disguised fulfillment of an unconscious childhood wish that is otherwise not readily accessible to conscious awareness in waking life.

In attempting to characterize the psychology of dreaming, Freud laid the foundations for ego psychology. He suggested that unconscious childhood wishes can be transformed into disguised conscious manifestations only if a censor exists in the mind. The censor, acting in the service of the ego, functions to preserve sleep. By disguising disturbing thoughts and feelings, the censor makes sure that the dreamer's sleep will not be disturbed. Moreover, early forms of defense mechanisms in the ego were delineated by Freud's investigation of the different methods of disguise used by the ego—for example, displacement, condensation, and symbolic representation. Freud also drew beginning parallels between the dream mechanisms and the pathological thoughts of psychotic patients in the waking state.

Analysis of Dream Content The dream, as it is recalled by the dreamer, is the result of the unconscious mental activity that occurs during sleep. Freud believed that dreaming is simply the conscious experience of thoughts during sleep. Contemporary research has revealed that the cognitive activity of sleep actually varies considerably. The dreaming activity described by Freud is probably more or less restricted to the stage 1 rapid eye movement (REM) periods of the sleep-dream cycle; those periods occur approximately every 90 minutes during the night. The thoughts that occur in non-REM sleep periods tend to be more logically organized, briefer, and more realistic.

Freud distinguished between two layers of dream content. The *manifest dream* refers to the content as it is recalled by the dreamer, including the mood and sensory experiences accompanying the dream and any commentaries on the dream that occur within the dream. The unconscious thoughts and wishes that threaten to awaken the dreamer constitute the *latent* content. Freud referred to the unconscious mental operations by which the latent dream content is transformed into the manifest dream as the *dream work*. Freud's technique of dream interpretation was based on free association, in which the patient would say whatever came to mind in response to aspects of the dream. Through this approach the manifest content of the dream would gradually give way to the underlying latent content in which the unconscious meaning of the dream is contained.

Freud initially believed that internal or external stimuli initiate dreaming. Modern dream research has demonstrated that endogenous patterns of activation in the central nervous system associated with particular phases of sleep are responsible for the onset of dreaming activity. A more contemporary understanding of what Freud viewed as initiating stimuli are that such phenomena are simply incorporated into dream content and influence the material in the dream to that extent.

Nocturnal Sensory Stimuli A variety of sensory impressions—including hunger, pain, thirst, and urinary urgency—may influence dream content. For example, a man whose thirst is beginning to interfere with his sleep may dream that he has risen from bed, gone to the bathroom, drunk a glass of water, and returned to bed as a way of safeguarding the continuity of sleep. Freud's view of dreaming as the guardian of sleep is still considered a valid function of dreaming, but recent research suggests that the function of dreams is considerably more complex in that the preservation of sleep is only one of many functions.

Day Residues Freud discerned that an important influence on the dream thoughts is the residues of feelings, ideas, and thoughts associated with experiences of the preceding day. The final form of the manifest content of the dream is shaped by those day residues, as well as by sensory stimuli. Seemingly innocuous or unimportant elements from the day residues may be used as the vehicle for the disguised expression of unconscious drives and wishes of an erotic or aggressive nature. Hence, events of the preceding day may effectively conceal the infantile impulse at the core of the dream.

Repressed Infantile Drives Following the dictum that a dream represents the disguised fulfillment of an unconscious childhood wish, Freud understood the driving forces behind dream activity as impulses or wishes from the earliest years of life. Those repressed wishes, usually sexual or aggressive and stemming from oedipal and preoedipal levels of development, are melded with day residues and nocturnal stimuli to produce the end result. Despite some challenges from discoveries of recent sleep research, Freud's observations continue to have clinical utility.

Dreams by young children are particularly unique because they show little distinction between infantile and current conflicts. In general, the dreams of young children are less disguised than the dreams of older persons and show much less distinction between manifest and latent dream content because of the relative immaturity of the defensive operations of the child's ego.

Dream Work As just noted, the dream work comprises several processes that transform latent dream content into manifest dream content. The theory of the nature of dream work, which Freud first put forth in *The Interpretation of Dreams*, became the fundamental description of the operation of unconscious processes. The mechanisms that Freud elaborated from his study of dreams proved to have broad applications to neurotic symptom formation and to a general psychology of the mind.

Dream Formation The basic problem of dream formation is to determine how the latent dream content represents itself through the manifest dream. In Freud's view the state of sleep leads to a relaxation of repression. With the repressive forces weakened, unconscious impulses and wishes are allowed to press for gratification and discharge. Since an outlet in motor expression is blocked by the sleep state, those repressed wishes and impulses have to find means of representation through mechanisms of thought and fantasy. Both elements of day residues and nocturnal stimuli are appropriated to express those latent wishes or impulses in the manifest dream.

The unconscious wishes and impulses that emanate from childhood are repressed because they are painful or unacceptable. Although repression is relaxed during sleep, Freud postulated that a dream censor is still actively involved in resisting the discharge of those impulses and disguising their true nature. The dream censor serves to attach impulses and wishes to neutral or innocent images from the residues of the dreamer's current psychological experience. Trivial or insignificant images from the dreamer's day residues are presumably linked on the basis of some resemblance that allows connections to be established. To enable those neutral images to surface in the dream, the dream work uses a set of disguised mechanisms that include condensation, displacement, symbolic representation, and secondary revision.

Condensation In condensation several unconscious impulses, wishes, or feelings can be combined and attached to one manifest dream image. For example, a composite character may appear in the dream that has the name of one person in the dreamer's life, a beard like another person, and a musical instrument that reflects a third person. The feelings associated with those three persons may be disguised in the resulting amalgam and may become apparent only through analysis of the various dream elements. The converse of condensation, diffusion, can also occur in a dream when a single latent wish or impulse is distributed through multiple representations in the manifest content.

Displacement In displacement the energy or intensity associated with one object is diverted to a substitute object that is associatively related but more acceptable to the dreamer's ego. Murderous wishes toward one's mother, for example, may be redirected toward a neutral or insignificant figure in one's life. In that manner the dream censor has displaced affective energy in such a way that the dreamer's sleep can continue undisturbed.

A special instance of displacement, projection, involves the attribution of the dreamer's own unacceptable impulses or wishes to another character in the dream. For example, a dreamer who finds homosexual impulses unacceptable may attribute them to his analyst in the dream. In some cases different aspects of the dreamer are represented in several characters in the dream.

Symbolic Representation Freud noted that the dreamer would often represent ideas or objects that were highly charged by using innocent images that were in some way connected with the idea or object being represented. In that manner an abstract concept or a complex set of feelings toward a person could be symbolized by a simple, concrete, or sensory image. Freud noted that symbols have unconscious meanings that can be discerned through the patient's associations to the symbol.

However, he also believed that certain symbols have universal meanings—for example, a flower as a symbol for female genitalia, a snake as a symbol for the penis.

Secondary Revision The mechanisms of condensation, displacement, and symbolic representation are characteristic of a type of thinking that Freud referred to as *primary process*. That primitive mode of cognitive activity is characterized by illogical, bizarre, and absurd images that seem incoherent. Freud believed that a more mature and reasonable aspect of the ego is at work during the dream to organize some of those primitive aspects of the dream into a more coherent form. He called that process *secondary revision*; in it intellectual processes of a more mature nature make the dream somewhat more rational. The process is related to more mature activity characteristic of waking life, which Freud termed *secondary process*.

Typical Dreams For the most part Freud believed that the underlying meaning of a dream reveals itself only through a proper analysis of the dreamer's associations. However, just as he made exceptions to certain universal symbols, he also thought that certain dreams are universal in their meaning, and he referred to them as *typical dreams*.

Anxiety Dreams Freud's dream theory preceded his development of a comprehensive theory of the ego. Hence, his understanding of dreams stressed the importance of discharging drives or wishes through the hallucinatory contents of the dream. He viewed such mechanisms as condensation, displacement, symbolic representation, projection, and secondary revision primarily as facilitating the discharge of latent impulses, rather than protecting the dreamer from anxiety and pain. Freud understood anxiety dreams as reflecting a failure in the protective function of the dream work mechanisms. In other words, the repressed impulses succeed in working their way into the manifest content in a more or less recognizable manner. The ego reacts to the emergence of the threatening content with intense anxiety; the dreamer often experiences severe distress or may even partially awaken.

Punishment Dreams Dreams in which the dreamer experiences punishment represented a special challenge for Freud because they appear to represent an exception to his wish-fulfillment theory of dreams. He came to understand such dreams as reflective of a compromise between the repressed wish and the repressing agency or conscience. In the punishment dream the ego anticipates condemnation by the dreamer's conscience if the latent unacceptable impulses are allowed direct expression in the manifest dream content. Hence, the wish for punishment by the patient's conscience is satisfied by giving expression to punishment fantasies. In that formulation Freud anticipated the concept of the superego and the interplay between the ego and the superego, which he did not elaborate into the structural model for some 20 years.

TOPOGRAPHIC MODEL OF THE MIND

The publication of *The Interpretation of Dreams* in 1900 heralded the arrival of Freud's topographic model of the mind. That model divides the mind into three regions: the conscious system, the preconscious system, and the unconscious system, each of which has its own unique characteristics. That theory of the psyche embraces four central assumptions. The first assumption is the notion of psychological determinism derived from Freud's Helmholtzian background. The second assumption is the existence of unconscious psychological processes. The third assumption postulates that neurotic symptoms are related to unconscious psychological conflicts and forces. Finally, the topographic theory assumes that some psychological energies originate in instinctual drives.

The Conscious The conscious system in the topographic model is characterized as the part of the mind in which perceptions coming from the outside world or from within the body or mind are brought into awareness. Consciousness is viewed as a subjective phenomenon whose content can be communicated only by means of language or behavior. Freud assumed that consciousness employs a form of neutralized psychic energy that he referred to as *attention cathexis*. In other words, one is aware of a particular idea or feeling as a result of the investment of a discrete amount of psychic energy in that particular idea or feeling. Until 1923 Freud viewed the conscious mind as being in control of motor activity as well. Freud devoted little of his own energies to the conscious system, so it is relatively undeveloped in his theory of the mind, as compared with the unconscious.

The Preconscious The preconscious system comprises those mental events, processes, and contents that are capable of being brought into conscious awareness by the act of focusing attention. Although most people are not consciously aware of the appearance of their first-grade teachers, they can ordinarily bring that image to mind by the deliberate focusing of attention on the memory. Conceptually, the preconscious interfaces with both the unconscious region and the conscious region of the mind. To reach conscious awareness, the contents of the unconscious must become linked with words and thus become preconscious. The preconscious also maintains the repressive barrier and censors unacceptable wishes and desires. Unlike the unconscious, which is characterized by primary process thinking, the preconscious is characterized by secondary process thinking and the delay of drive discharge.

The Unconscious The unconscious system is a dynamic one; the mental contents and processes of the unconscious are kept out of conscious awareness through the force of censorship or repression. The repressive force, sometimes referred to as *countercathexis*, manifests itself as the resistance to remembering. The essence of the unconscious can be captured in five key features:

1. The unconscious is closely related to instinctual drives. In Freud's theory of development, instincts were then thought to consist of sexual and self-preservative drives, and the unconscious was thought to contain primarily the mental representations and derivatives of the sexual instinct.
2. The content of the unconscious is limited to wishes seeking fulfillment. Those wishes provide the motivation for dream and neurotic symptom formation. That view is now considered reductionistic.
3. The unconscious system is characterized by primary process thinking, which has as its principal aim the facilitation of wish fulfillment and instinctual discharge. Primary process thinking is governed by the pleasure principle and, therefore, disregards logical connections, has no conception of time, represents wishes as fulfillments, permits contradictions to exist simultaneously, and denies the existence of negatives. The primary process is also characterized by extreme mobility of drive cathexis, meaning that the investment of psychic energy can shift from object to object without opposition. The mental mechanisms of displacement, condensation, and symbolic representation, discussed as part of the dream work, are part of the primary process as well. That form of thinking is also associated with creativity and with severe mental illness.
4. Memories in the unconscious have been divorced from their connection with verbal symbols. Hence, when words are reapplied to forgotten memory traits, as in psychoanalytic treatment, the verbal recathexis allows the memories to reach consciousness again.
5. The contents of the unconscious can become conscious only by passing through the preconscious, where censors are overpowered, allowing the elements to enter into consciousness.

DYNAMICS OF MENTAL FUNCTIONING

Freud conceptualized the psychic apparatus in the context of the topographic model as a kind of reflex arc. The arc consists of a series of segments arranged consecutively, with the perceptual or sensory system at one end and the motor system at the other. The memory and association systems were conceived of as occupying the segments between the two extremes. Perceptions are thus modified and stored in the form of memories, and the mental energy associated with unconscious ideas seeks discharge through thought or motor activity. Hence, the normal pathway of discharge in the waking state is from the perceptual to the motor end. Freud suggested that the flow of psychic energy can be reversed in sleep, moving from the motor to the perceptual end. That reversal of the normal flow of energy, which he called *topographic regression*, is the basis for the appearance of childhood images in dreams and the presence of hallucinations in psychoses.

When Freud subsequently abandoned the reflex arc model of the mind, he nevertheless retained the concept of regression as a key component to his theory. In neurosis, for example, the patient is thought to revert to earlier modes of instinctual discharge or to earlier levels of fixation. That form of regression is termed *libidinal* or *instinctual* regression.

FRAMEWORK OF PSYCHOANALYTIC THEORY

After the development of the topographic model, Freud turned his attention to the complexities of instinct theory. The *source* of the instinct refers to the part of the body from which it arises, to the somatic process that gives rise to stimuli that are represented in mental life as emotions, or to drive representations. The *impetus* or pressure refers to the amount of force or energy created by the instinctual stimulus. The *aim* refers to the action-directed gratification of the instinct. And the *object* is the target or recipient of the action-directed gratification and may be a person or a thing.

INSTINCT OR DRIVE THEORY

Freud was determined to anchor his psychological theory in biology. That choice led to terminological and conceptual difficulties when Freud used terms derived from

biology to denote psychological constructs. *Instinct*, for example, refers to a pattern of species-specific behavior that is genetically derived and, therefore, more or less independent of learning. Used primarily by students of animal behavior, the word “instinct” applies to such phenomena as migration, nesting, and maternal behavior.

Ethological studies have increasingly demonstrated that instinctual patterns are modified through experiential learning. The effects of adaptation on instinctually derived patterns have made Freud's instinctual theory problematic. Confusion also surrounds the translation of the German word *Trieb* into English; it does not have an unequivocal correlate. Freud was referring to powerful, imperative strivings, such as self-preservation and sexuality. Hence, *Trieb* has been translated as “drive,” “instinct,” and “instinctual drive.”

Confounding matters further is the fact that Freud himself shifted his view of instinct theory over time. His clearest definition appeared in his 1915 paper “Instincts and Their Vicissitudes,” in which he referred to an instinct as “a concept on the frontier between the mental and the somatic, as the psychical representative of the stimuli originating from within the organism and reaching the mind, as a measure of the demands made upon the mind for work in consequence with its connections with the body.” The ambiguity inherent in this concept on the borderland between the biological and the psychological has led to considerable debate about whether the mental representation aspect of the term and the physiological component should be integrated or separated.

Although “drive” may have been closer than “instinct” to Freud's meaning, he preferred not to translate *Trieb* in that manner because behavior theorists of his time had already appropriated the term “drive.” In contemporary usage, “drive” and “instinct” are often used interchangeably.

Classification of Instincts Freud's instinct theory grew out of clinical observations. The sexual drive, for example, appeared to be central to the pathogenesis of hysteria. Hence, Freud was primarily preoccupied with the sexual drive during the 1890s and into the early twentieth century. Although he originally postulated a self-preservative instinct in the 1890s, he did not elaborate it until some 20 years later. As psychoanalytic theory continued to evolve, his theorizing about the role of instincts became more and more abstract and divorced from clinical data.

Libido Freud defined *libido* as “that force by which the sexual instinct is represented in the mind.” The association of libido with sexuality is somewhat misleading in that Freud's intent was to encompass the general notion of pleasure, as well as sexuality, including both the physiological underpinnings and the mental representations. The linkage of genital sexuality with libido was viewed as the end result of a course of development in which libidinal expression takes a variety of forms, as is elaborated in the discussion of libidinal zones later in this section.

Ego Instincts From 1905 on, Freud maintained a dual-instinct theory subsuming sexual instincts and ego instincts connected with self-preservation. Until 1914, with the publication of “On Narcissism,” Freud had paid little attention to ego instincts. In that communication Freud invested ego instinct with libido for the first time. He postulated an ego libido and an object libido. Freud thus viewed narcissistic investment as an essentially libidinal instinct and called the remaining nonsexual components the ego instincts. In that rather confusing model Freud suggested that the clash between libido and ego instincts may produce symptoms of neurosis. At that point ego instincts connoted both the drives of self-preservation, such as hunger and thirst, and early aspects of the superego or conscience.

Aggression When psychoanalysts today discuss the dual-instinct theory, they are generally referring to libido and aggression. However, Freud originally conceptualized aggression as a component of the sexual instincts in the form of sadism. As he became aware that sadism has nonsexual aspects to it, he made finer gradations, enabling him to categorize aggression and hate as part of the ego instincts and to categorize the libidinal aspects of sadism as components of the sexual instincts. With that differentiation, the impulse to attack in order to defend oneself can be seen as aggression associated with the ego instincts, and sadism can be regarded as a fusion of sexual and aggressive instincts.

Freud ran into much greater difficulty when he attempted to explain self-destructive behavior with his newly developed taxonomy of instincts. How, for example, could he view the self-mutilation of masochistic patients as part of the ego instincts of self-preservation? Finally, to account for the clinical data he was observing, in 1923 he was compelled to conceive of aggression as a separate instinct in its own right. The source of that instinct, according to Freud, is largely in skeletal muscles, and the aim of the aggressive instincts is destruction. He also noted that aggression in response to the frustration of libido is secondary, rather than primary.

Life and Death Instincts Before the designation of aggression as a separate instinct, Freud, in 1920, subsumed the ego instincts under a broader category of life instincts. Life instincts were juxtaposed with death instincts, and the two were referred to as *Eros* and *Thanatos* in *Beyond the Pleasure Principle*. The life and death instincts were regarded as forces underlying the sexual and aggressive instincts. Although Freud could not provide clinical data that directly verified the death instinct, he thought it could be inferred by observing the *repetition compulsion*, the tendency of persons to repeat past traumatic behavior. Freud felt that the dominant force in biological organisms had to be the death instinct. He viewed it as a tendency of all organisms and their component selves to return to an inanimate state. Here again, Freud was linking psychology to principles of hard science—specifically, the notions of entropy and constancy from physics.

In contrast to the death instinct, Eros, the life instinct, refers to the tendency of particles to reunite or bind to one another. That counterpoint to entropy leads to organization and greater unities of systems. Sexual reproduction is the most obvious example. The death and life instincts are directly analogous to catabolism and anabolism. In modern psychoanalytic thinking the death instinct has been largely discounted except by followers of Melanie Klein. The current view is that the dual instincts of sexuality and aggression are sufficient to explain most clinical phenomena.

Pleasure and Reality As noted in the discussion of Freud's *Project*, the constancy principle, the tendency of an organism to maintain a particular state or equilibrium, was one of the cornerstones of the evolution of psychoanalytic theory. Freud's notion of the death instinct was clearly linked to the constancy principle and was also associated with what he termed the *Nirvana principle*, which postulates that an organism strives to discharge internal tension and to seek a state of rest. In 1911 Freud further elaborated the nature of the human organism's need to maintain a state of equilibrium when he described two basic tenets of mental functioning: the pleasure principle and the reality principle.

Freud essentially recast the primary process and secondary process dichotomy into the pleasure and reality principles, thus taking an important step toward solidifying the notion of the ego. Both principles, in Freud's view, are aspects of ego functioning. The *pleasure principle* is defined as an inborn tendency of the organism to avoid pain and to seek pleasure through the discharge of tension. The *reality principle*, on the other hand, is considered a learned function, closely related to the maturation of the ego, that modifies the pleasure principle and requires the delay or postponement of immediate gratification.

INFANTILE SEXUALITY

In 1905 Freud published *Three Essays on the Theory of Sexuality*. He regarded that work as second only to *The Interpretation of Dreams* in importance. Contrary to popular belief, Freud's ideas did not create much of an uproar among the scientific community or even among the public. The Vienna of his day was exposed to a great deal of popular literature that openly discussed an array of sexual problems. However, Freud's notion that children are influenced by sexual drives has made it difficult for some people to accept psychoanalysis throughout its 100-year history.

Freud's book was significant because it set forth three major tenets of psychoanalytic theory. First of all, Freud broadened the definition of sexuality to include forms of pleasure that transcend genital sexuality. Second, Freud established a developmental theory of childhood sexuality that delineated the vicissitudes of erotic activity from birth through puberty. Third, he forged a conceptual linkage between neuroses and perversions.

PSYCHOSEXUAL DEVELOPMENT

Freud's ideas about psychosexual developmental stages and libidinal zones have worked their way into popular culture. It is commonplace to hear someone who is greedy or voracious referred to as “oral.” Persons who are withholding and stubborn may be regarded as “anal.” The popular usage reflects the fact that Freud's identification of psychosexual stages also laid the groundwork for the subsequent psychoanalytic theory of character.

The earliest manifestations of infantile sexuality, in Freud's view, are basically nonsexual. They are associated with such bodily functions as feeding and bowel and bladder control. As the libidinal energy shifts from the oral zone to the anal zone to the phallic zone, each stage of development is thought to build on and to subsume the accomplishments of the preceding stage. The oral stage occupies the first 12 to 18 months of life. The anal stage picks up where the oral stage leaves off and extends to approximately 36 months of age. The phallic stage is usually associated with the period of time between ages 3 and 5.

Throughout each of the psychosexual stages, specific erotogenic zones, when stimulated, give rise to erotic pleasure. During the oral stage the infant derives

pleasure from nursing, in addition to satisfying the physiological need to assuage hunger. Karl Abraham, a disciple of Freud, divided the oral period into a sucking phase and a biting phase, thus incorporating aspects of aggression and sadism into the notion of sexual pleasure. Similarly, Abraham divided the anal stage into an anal-sadistic phase and an anal-erotic phase.

The phallic stage is initially focused on urination as the source of erotic activity. That urethral phase was briefly noted by Freud but was elaborated on by a number of later writers. During the urethral phase the primary eroticism is pleasure in urination and a secondary urethral retention pleasure. One may see derivatives of that phase in paraphilic activities involving sexual excitement associated with urinating on others or being urinated on by others.

Freud suggested that phallic erotic activity in boys is a preliminary stage leading to adult genital activity. Whereas the penis remains the principle sexual organ throughout male psychosexual development, Freud postulated that the female has two principal erotogenic zones, the vagina and the clitoris. He thought that the clitoris is the chief erotogenic focus during the infantile genital period but that erotic primacy shifts to the vagina after puberty. Studies of human sexuality have subsequently questioned the validity of that distinction.

Vicissitudes *Three Essays On the Theory of Sexuality* also discusses the vicissitudes of libido, by which Freud meant the different paths that the instinct can take within the sexual life of the person. Although in normal adult functioning sexual activity is dominated by the genital zone, pregenital libidinal zones continue to play a role, specifically during foreplay. Kissing, for example, constitutes a pleasurable stimulation of the oral zone, and stimulation of the anus may also accompany foreplay or be part of lovemaking. In healthy adults who have attained mature genital potency, the sexual act culminates in the pleasure of orgasm. Freud conceptualized infantile sexuality as *polymorphously perverse*—that is, composed of partial or component instincts that allow sexual excitation to occur from multiple sources. For example, oral and anal stimulation involves such component instincts as pleasure in watching *(scoptophilia)*, pleasure in showing *(exhibitionism)*, and pleasure in cruelty *(sadism)*. Those component instincts emerge as pairs of opposites. Scoptophilia and exhibitionism are instinctual pairs, as are sadism and masochism. Freud thought that all those components are combined into a unity at puberty under the primacy of the genital zone. Unacceptable components are either repressed or confined to foreplay. If the libido becomes irrevocably attached to one of the pregenital erotogenic zones or if a part instinct predominates, the result may be a perversion in which the normal act of intercourse is replaced by some other activity, such as fetishism or voyeurism. Freud termed the persistent attachment of the libido to a particular phase of genital development a *fixation*.

Neurosis and Perversion In studying perversions, Freud noted that perverse fantasies are ubiquitous in the human unconscious. He believed neurosis to be the negative of perversion. In other words, whereas persons with perversions express their fantasies and impulses in action, neurotic persons repress similar fantasies and impulses, which manifest themselves only as neurotic symptoms. Although that conceptualization continues to be valid in some cases, it perhaps oversimplifies the pathogenesis of perversions. Freud's formulation, for example, cannot explain why a part instinct in one person is repressed and transformed into a neurotic symptom whereas in another it is expressed overtly as a perversion. More theoretical sophistication was required to explain those individual differences.

OBJECT RELATIONS IN CLASSICAL THEORY

Although Freud's thinking is usually associated with drive theory and ego psychology, he also evolved his own version of object relations theory. He always held that early experiences with parenting figures have profound influences on subsequent development. He believed that adult relationships are later editions of childhood relationships. However, in Freud's opinion, the development of object relations is intimately connected with the sexual drive. Infants, in his view, have limited awareness of the mothering figure. Their main focus is on internal states of tension and relaxation. Hence, their longing for the object is, in fact, a longing for drive discharge or tension release. The object is viewed as a vehicle to satisfy drive pressures. As a result of that perspective on the infant's world, Freud's developmental line of object relations is intimately interwoven with psychosexual stages of development.

Oral Stage In the earliest months of life, the infant experiences hunger and frustration in the absence of the breast and the need-satisfying discharge of tension when the breast is present. Tension and hunger force the recognition and acceptance of persons in the outside world. In Freud's view, then, the first psychological awareness of an object arises from the intense physiological need for a familiar experience that provides gratification and relieves tension.

The mother's responsiveness to the child is critical in laying the foundations for the most rudimentary and essential basis for the subsequent development of object relations and the capacity for entering the world of human beings. She becomes the first love object for the infant in that she is recognized as the source of gratification of hunger and the provider of the erotogenic pleasure that the infant obtains from sucking. If a warm, trusting, and affectionate relationship has been established between mother and child during the first few months of life, the stage is set for the development of trusting and affectionate relationships with others throughout life.

Conversely, when the early mother-infant bond and the feeding experience are disturbed, the groundwork is established for subsequent problems in the area of object relations. If the mother is unavailable during the early sucking phase of the oral stage, frustration may be so intense that the infant grows up with intense longings to be nurtured and taken care of by others. If the fixation occurs during the biting phase of the oral stage, the child may be plagued with oral aggressive tendencies throughout life. Those impulses may be manifested in a voracious appetite or in a tendency to make biting comments about others that is destructive to relationships.

Anal Stage After infants pass through the oral stage during the first 18 months of life, they enter the anal stage, which lasts from about 18 to 36 months of age. One striking difference is that in the anal stage the child is much less passive than in the oral stage. Moreover, the demands of toilet training during the anal period lead to a struggle of wills between mother and child. The erotization of the anal stage involves both the pleasurable sense of excretion and, later, the erotic stimulation of the anal mucosa through the retention of feces.

Freud noted the connection between anal and sadistic drives. Initially, the object of anal-sadistic activity is the feces, and pinching off is regarded as a sadistic act. As the anal stage progresses, the sadism becomes more interpersonal in nature. In the developing struggles over toilet training, the child learns to exercise power over the parents by either giving up or retaining the feces. The sense of power over the environment that comes with sphincter control represents another sadistic element.

Before toilet training, elimination and pleasurable retention are essentially autoerotic because they do not require the presence or the assistance of an outside object. The act of defecation during that early period is imbued with a sense of omnipotence as a result. The feces become libidinized because they represent pleasure. Later, the child develops an ambivalent view of feces as body contents that are both external and internal. In other words, the child regards the feces as both “me” and “not me.” On the one hand, the feces are loved and retained or reinternalized; on the other hand, they are hated and pinched off.

The ambivalence associated with the anal stage may also be transferred to objects in the external environment. The stimulation associated with having the anal area cleansed may lead to strong erotic feelings toward the mother. Later, battles over toilet training lead to aggressive and hateful feelings toward parental figures. Freud suggested that obsessive-compulsive persons had regressed to the anal stage of development. The ambivalence associated with the feces in conjunction with the parental control led those persons to become compulsively neat, rigid, domineering, and pedantic. Freud also described them as intensely ambivalent, tormented by simultaneous feelings of love and hate and by simultaneous wishes both to control and retain the object and to expel and destroy it.

Urethral Stage Although Freud did not discuss the urethral psychosexual stage in any depth, some clinicians think it has particular relevance to issues of performance and control. The urethral stage is generally viewed as transitional between the anal and the phallic psychosexual stages. Pleasure in urination is referred to as *urethral erotism*, and urethral functioning may be associated with sadistic urges carried over from the anal stage. Competitiveness and ambition are often viewed as compensation for the shame associated with loss of urethral control.

Phallic Stage Around the age of 3 the child enters the phallic stage, in which the penis or the clitoris is the primary erotogenic zone. The phallic stage of psychosexual development heralds the arrival of the oedipal level of development, in which relationships become more complicated than they were in the past. The emphasis is on triangular or three-person relationships, instead of dyadic or two-person relationships. The phallic stage is also characterized by greater tolerance of ambivalence and the ability to withstand frustration in the absence of significant objects because of the ability to maintain an internal representation of the absent object.

Another major contrast between the pregenital psychosexual stages of development and the phallic stage is the nature of the child's libidinal activity. In the oral and anal stages, such activity is, for the most part, autoerotic in that the child's sexual impulses are directed toward the child's own body. Pleasure is still derived from one's own body in the phallic phase, but that period of development is also characterized by the fundamental task of finding a love object that will establish later patterns of object choice in adult life.

Oedipus Complex The period of life between ages 3 and 5 is known as the oedipal stage of psychosexual development because the culmination of infantile sexuality—the *Oedipus complex*—occurs at that time. Freud discovered this crucial cornerstone of psychological development during his own self-analysis and in his clinical work with patients in which fantasies of incest with the parent of the opposite sex emerged with regularity in association with feelings of jealousy and murderous rage toward the parent of the same sex. Drawing an analogy between such fantasies and the Greek myth of Oedipus, who unknowingly killed his father and married his mother, Freud termed the intrapsychic constellation the Oedipus complex.

The oedipal phase of development is of central importance in the pathogenesis of neuroses and many anxiety disorders. Oedipal issues are also prominent in the psychodynamics of character neuroses and high-level personality disorders, such as hysterical personality and obsessive-compulsive personality. The Oedipus complex presents a developmental challenge for the child, and the resolution of the child's dilemma differs according to the child's gender.

RESOLUTION FOR BOYS The first love object of the male child is his mother. Unlike the little girl, the little boy does not need to shift his affection to another parent at the beginning the oedipal phase. The male child essentially falls in love with his mother. He wishes to be at the center of her world, to sleep with her, to caress her, and to have all her attention. It becomes apparent that such plans are interfered with by the relationship of his mother and father. As a result, he begins to view his father as a rival and develops murderous wishes toward him. Those thoughts create a predicament for the little boy because he also loves his father. Murderous wishes produce guilt. They also produce fear of retaliation, accompanied by anxiety about that impending retaliation.

Freud repeatedly noted that the chief source of the boy's anxiety in the Oedipus complex is that the father will retaliate by removing the child's penis. The male child's investment in keeping his genitals intact supersedes his sexual wishes for his mother, and he renounces those wishes as a result. Freud termed the renunciation of oedipal strivings secondary to castration anxiety the castration complex. Hence, the Oedipus complex in boys is resolved by the castration complex. The male child identifies with his father and decides to find a woman like his mother so that he can become like his father. That resolution is often referred to as *identification with the aggressor*. One aspect of the male resolution of the Oedipus complex is that the retaliatory father is internalized to form the superego, which Freud regarded as heir to the Oedipus complex.

Freud viewed bisexuality as a fundamental aspect of the unconscious. Loving feelings for the father coexist with the view of him as a rival who must be murdered. At times, the mother is regarded as interfering with the father-son relationship, and the male child has murderous wishes about his mother. Freud referred to that constellation as the *negative Oedipus complex*, which he considered a universal aspect of psychosexual development.

RESOLUTION FOR GIRLS Freud was frank throughout his writings about his difficulty in understanding psychological development in girls. He often dealt with that difficulty by assuming that female development occurs along lines that are essentially analogous to male development. In attempting to explain the resolution of the Oedipus complex in little girls (called the *Electra complex*), Freud noted that, although the Oedipus complex in male children is resolved by the castration complex, in female children it is promulgated by an awareness of their “castrated” state. Freud believed that little girls feel like little boys during the preoedipal psychosexual stages until the age of 3, when girls discover the existence of the penis. He postulated that the discovery leads to feelings of inferiority and narcissistic injury and to penis envy. The little girl then blames her mother for bringing her into the world less well equipped than her male counterparts and turns to her father as her love object. The wish to have a baby with her father then replaces her wish for a penis.

The discovery of the little girl's “genital inferiority,” according to Freud, leads to one of three outcomes: (1) a defiant hypermasculinity, (2) a neurotic cessation of all sexuality, or (3) normal femininity, entailing the renunciation of clitoral sexuality. Whereas the fear of castration leads to the resolution of the little boy's Oedipus complex, fear of losing the mother's love leads the little girl to a satisfactory resolution of the Electra complex. When the female child realizes that her mother disapproves of her wishes toward her father, she decides to renounce her oedipal strivings to maintain her bond with her mother.

Freud's view of female development has undergone significant revision by modern psychoanalytic authors. It is now known that the sense of being female is present long before the onset of the phallic-oedipal stage of development. Child observational research suggests that both sexes become aware of anatomical genital differences at approximately 16 to 18 months of age. Also, studies of persons with chromosomal anomalies and ambiguous genitalia have persuasively demonstrated that the primary source of femaleness is not the nature of the genitalia but the parents' conviction about the nature of the child's gender. Moreover, penis envy is no longer regarded as a bedrock phenomenon that defies further understanding and analysis. Contemporary psychoanalytic writers regard penis envy as only one aspect of the development of feminine identity, not the origin of it. They point out that following a strictly Freudian model of development may lead a clinician to help a female patient regard herself as a genitally inferior form of male. The mysterious nature of the little girl's genitalia may provoke anxiety independent of their differences from the genitals of little boys, and those anxieties can be fruitfully explored in psychoanalysis.

SIGNIFICANCE OF THE OEDIPUS COMPLEX Freud viewed the Oedipus complex, along with its successful or unsuccessful resolution, as the nucleus for the pathogenesis of adult neurosis. He also thought that personality, fixations, object choices, and identifications are significantly influenced by the oedipal phase of development. Incestuous oedipal feelings reemerge in both sexes at puberty, and one of the tasks of adolescence is to withdraw libidinal urges from parental figures and direct them toward more suitable love objects.

Latency Stage This stage of development begins with the resolution of the Oedipus complex around the age of 5 or 6 and ends with the onset of puberty, somewhere between 11 and 13 years of age. The term “latency” derives from the fact that the sexual drive is relatively quiescent during those years. The inactivity of the sexual drive can be explained by the institution of the superego as the heir to the Oedipus complex. The control of instinctual impulses by the superego allows for further integration and consolidation of sex-role identity in sex roles. The latency stage of development is also associated with the development of a sense of industry and a capacity for mastery that enhances autonomous functioning.

Genital Stage This stage of psychosexual development corresponds with the adolescent phase that begins with the onset of puberty and ends with the achievement of young adulthood. Most clinicians who work with adolescents think it is useful to subdivide the period into preadolescent, early adolescent, middle adolescent, and late adolescent subphases. The most striking influence on this period of development is the physiological maturation of hormonal systems that results in an intensification of drives, particularly sexual drives. This rather extended psychosexual stage of development, without a doubt one of the most challenging in the life cycle, requires the development of psychological mastery over drive pressures. In addition, a key developmental task associated with the genital stage is the establishment of mature object relations and genital sexuality with an appropriate partner. A major aspect of that achievement is the psychological separation from one's own parents and the establishment of an independent life-style.

Freud's stages of psychosexual development are discussed in Table 6.1-1.

Table 6.1-1 Stages of Psychosexual Development

CONCEPT OF NARCISSISM

The first systematic discussion of narcissism in Freud's work was published in 1914. Freud was dissatisfied with the paper and wrote to Karl Abraham, “The

narcissism was a difficult labor and bears all the marks of a corresponding deformation." Nevertheless, Freud's thinking about narcissism led him to important modifications in his understanding of libido and instinct theory in general.

Theoretical Basis According to the Greek myth, Narcissus was a young man who fell in love with his own reflection in the water of a pool and drowned in his attempt to embrace the beloved image. Freud used the term *narcissism* to describe situations in which a person's libido is invested in the ego itself, rather than in other people. That conceptualization of narcissism presented Freud with vexing problems for his instinct theory. It essentially violated his distinction between libidinal instincts and ego or self-preservative instincts. Freud's understanding of narcissism led him to use the term to describe a wide array of psychiatric disorders, very much in contrast to the contemporary usage of the term to describe a specific kind of personality disorder. For example, in 1908 Freud noted that in cases of schizophrenia (which was then known as *dementia precox*), the person's libido is withdrawn from objects and turned inward. He believed that the withdrawal of libidinal attachment to objects accounted for the loss of reality testing in psychotic patients. Grandiosity and omnipotence in such patients reflected excessive libidinal investment in the ego.

Freud did not limit his use of narcissism to psychoses. In states of physical illness and hypochondriasis, he observed that libidinal investment is frequently withdrawn from external objects and from outside activities and interests. Similarly, he suggested that, in normal sleep, libido is withdrawn and reinvested in the sleeper's own body. Homosexuality, which Freud regarded as a perversion, was understood as an instance of a narcissistic form of object choice, one in which a person falls in love with an idealized version of himself or herself projected onto another person. He also found narcissistic manifestations in the beliefs and myths of primitive people, especially those involving the ability to influence external events through the magical omnipotence of their own thought processes. In the course of normal development, children also exhibit a belief in their own omnipotence.

Narcissism and Object Relations According to Freud, the infant is born into a state of primary narcissism in which the libido is stored in the ego. He viewed the neonate as completely narcissistic, with the entire libidinal investment in physiological needs and their satisfaction. He referred to that self-investment as *ego libido*.

The infantile state of self-absorption changes only gradually, according to Freud, with the dawning awareness that a separate person—the mothering figure—is responsible for the gratification of the infant's needs. That realization leads to the gradual withdrawal of the libido from the self and the redirection toward the external object. Hence, the development of object relations in the infant parallels the shift from primary narcissism to object attachment. The libidinal investment in the object is referred to as *object libido*.

If the developing child suffers rebuffs or trauma from the caretaking figure, object libido may be withdrawn and reinvested in the ego. Freud called this regressive development *secondary narcissism*.

One of the confusing aspects of Freud's conceptualization of narcissism is that his use of the term varied according to context. At times, he discussed narcissism as a perversion in which persons use their own bodies or body parts as objects of sexual arousal. At other times, he used the term to describe a developmental phase, as in the state of primary narcissism, when there is no differentiation of self and other and when autoeroticism predominates. That use is particularly problematic because narcissism operates at all stages of psychosexual development, from the deprivation of oral needs during the earliest months of infancy to the losing out to a rival during the oedipal phase. In still other instances, Freud used the word "narcissism" to refer to a particular type of object choice. He distinguished love objects who are chosen "according to the narcissistic type," in which case the object resembles the subject's idealized or fantasied self-image, from objects chosen according to the "anaclitic type," in which case the love object resembles a caretaking figure from early in life. And Freud occasionally used the word "narcissism" interchangeably and synonymously with "self-esteem."

Freud's developmental view of narcissism and object relations has caused considerable controversy in the field. Part of the difficulty stems from the fact that a certain amount of narcissistic libido is necessary for healthy self-regard. Freud's maturational model implies that, in the course of healthy development, one should outgrow investment in oneself and devote one's libidinal energies primarily to others. Others have noted that Freud's model implies that investment of the libido in a love object results in depletion of libidinal investment in oneself. That conceptualization does not explain how it is that persons are likely to feel the greatest sense of self-esteem when they are in love. And Freud's model does not lend itself well to the differentiation between normal narcissism and pathological narcissism. Subsequent psychoanalytic theorists, discussed later in this section, have dealt with some of those difficulties.

EGO PSYCHOLOGY

Although Freud had used the construct of the ego throughout the evolution of psychoanalytic theory, ego psychology as it is known today really began with the publication of *The Ego and the Id* in 1923. That landmark publication also represented a transition in Freud's thinking from the topographic model of the mind to the tripartite structural model of ego, id, and superego. He had repeatedly observed that not all unconscious processes can be relegated to the instinctual life of the person. Elements of the conscience, as well as functions of the ego, are clearly unconscious as well. From 1923 to 1937 Freud developed both his theory of ego psychology and his later theory of anxiety. After his death, such psychoanalytic contributors as Heinz Hartmann, Ernst Kris, David Rapaport, and Erik Erikson continued the general thrust within psychoanalysis away from the instinctual life toward a greater emphasis on ego functioning and development and toward social and cultural perspectives.

Early Concepts of the Ego Freud's concept of the ego was ill-defined and imprecise during the formative years of psychoanalysis. He used the term to refer to the dominant place of the person's conscious thoughts and values, as opposed to the domain of repressed impulses and wishes. The ego's primary function is defensive, which at that time denoted repression. Impulses and wishes, primarily of a sexual nature, are barred from conscious awareness because of the counterforce provided by the ego. In the early years of the 1890s, Freud viewed the memories of sexual trauma as the main source of the defensive response. Those memories arouse unpleasant affect, which leads to repression, which in turn causes a damming up of energy that produces anxiety. That formulation, however, produces a contradiction in the nature of the ego's function. On the one hand, the ego reduces tension and avoids unpleasant affects through repression; on the other hand, the process of repression *produces* anxiety, which is an equally unpleasant affect. When Freud shifted in 1897 from his emphasis on actual seduction to the role of childhood fantasy, he also suspended his thinking about the function of the ego.

During the ensuing years, Freud devoted much of his attention to the instinctual drives and their vicissitudes. His studies led him back to consideration of the ego. In his 1915 work "Instincts and Their Vicissitudes," Freud noted that the sexual instinct is subjected to repression, sublimation, turning on oneself, and reversal into its opposite (a sadistic impulse transformed into a masochistic one). Those vicissitudes came to be understood as defense mechanisms of the ego.

Topographic theory failed to provide Freud with all the answers he needed to explain the functioning of the mind. For example, since neither the preconscious nor the ego instincts can be viewed as solely responsible for regression or censorship, how can repression be explained? Freud tried to explain repression by postulating that certain ideas remain in the unconscious as a result of libidinal withdrawal. He suggested that the withdrawal is a repetitive process of countercathexis that operates on an unconscious basis. He eventually assigned that function to the ego.

Freud's Ego Psychology Freud's first comprehensive theory of the ego appeared in *The Ego and the Id*. There, although the ego is primarily organized around the perceptual conscious system, it also includes the structures responsible for *unconscious* defenses and resistance in the clinical setting. The ego is regarded as the agency that organizes mental processes and functions. It is one of three major functional subdivisions of the mental apparatus—one that responds in a mostly passive way to pressures from the other two, the id and the superego. The ego has its origins in the id, in Freud's view, and is later depicted by him as a helpless rider on the id's horse, going in whichever direction the id dictates. The ego also responds to reality influences.

In 1926 Freud introduced the notion of signal anxiety in *Inhibitions, Symptoms and Anxiety*. There the ego is no longer viewed as a passive servant of the id. Signal anxiety is seen as an autonomous function of the ego that initiates defensive maneuvers in response to the threatened emergence of unacceptable wishes and impulses from the unconscious.

The concept of the ego was further expanded to include adaptation as a result of Freud's continued elaboration of the reality principle. Adaptation enables the ego to control instinctual drives when real danger prompts those drives into action. Intrinsic to the maturation of the ego is the substitution of the reality principle for the pleasure principle. In addition to reducing tension from the id, the reality principle gauges the limitations, requirements, and possibilities of the environment with an eye to attaining greater pleasure in the future. By the last decade of his life, Freud viewed the ego as the executive organ of the psyche, a powerful regulatory force for the integration and control of impulses, and as an agent of adaptation in relation to reality. In 1937 Freud made it clear that the ego evolves independently of the id.

STRUCTURE OF THE PSYCHIC APPARATUS

Contemporary psychoanalytic ego psychology is based on Freud's structural theory, which divides the psychic apparatus into the id, the ego, and the superego. The id encompasses the mental representations of the instinctual drives and some but not all the contents of the unconscious system in Freud's topographic theory. The id operates according to primary process and to the dictates of the pleasure principle. The ego occupies the position between the drives and the demands of the outer world. It evaluates, perceives, coordinates, and integrates perceptions in the service of taming the intensity of drives. The ego attempts to achieve optimal gratification of instinctual striving while maintaining good relations with the superego and with external reality. The superego assists the ego in the regulation of tensions arising from the id. It encompasses internalized moral values, prohibitions, and standards of parental figures.

Id Freud referred to the id as “the dark inaccessible part of our personality.... We approach the id with analogies: we call it a chaos, a cauldron full of seething excitation.” Freud borrowed the term from Georg Groddeck, a psychoanalytically informed internist, who coined the term “id” to mean all that was ego-alien. It was viewed as a reservoir of the unorganized instinctual drives. Operating under the domination of the primary process, the id lacks the capacity to delay or modify the instinctual drives with which the infant is born. The id should not, however, be viewed as synonymous with the unconscious because both the ego and the superego have unconscious components.

Ego The ego spans all three topographic dimensions of conscious, preconscious, and unconscious. Logical and abstract thinking and verbal expression are associated with conscious and preconscious functions of the ego. Defense mechanisms reside in the unconscious domain of the ego. The most comprehensive definition of the ego can be found in Freud's *An Outline of Psycho-Analysis*, which he wrote in 1938:

> Here are the principal characteristics of the ego. In consequence of the pre-established connection between sense and perception and muscular action, the ego has voluntary movement at its command. It has the task of self-preservation. As regards external events, it performs that task by becoming aware of stimuli, by storing up experiences about them (in the memory), by avoiding excessively strong stimuli (through flight), by dealing with moderate stimuli (through adaptation) and finally by learning to bring about expedient changes in the external world to its own advantage (through activity). As regards internal events, in relation to the id, it performs that task by gaining control over the demands of the instincts, by deciding whether they are to be allowed satisfaction, by postponing that satisfaction to times and circumstances favorable in the external world or by suppressing their excitations entirely. It is guided in its activity by consideration of the tension produced by stimuli, whether these tensions are present in it or introduced into it.

In essence, then, the ego controls motility, perception, contact with reality, and, through the mechanisms of defense available to it, the delay and modulation of drive expression.

Origins of the Ego In Freud's view, the neonate has no ego to speak of, and the coherent system of functions now associated with the concept is regarded as an outgrowth of the id. Specifically, the ego arises out of the modification of the id created by the demands of the external world. The ego forms in tandem with the increasing influence of external reality on the infant, gradually replacing the pleasure principle with the reality principle. The role of conflict, first between the id and the outside world and later between the id and the ego, Freud viewed as central to ego development.

Later contributors to ego psychology—such as Hartmann, Kris, and Loewenstein—made significant modifications to the theory of ego development. Hartmann, in particular, stressed the adaptational aspects of the ego. In his view, one can delineate a line of ego development that is independent of instinctual drives. He postulated the existence of primary autonomous ego functions that do not have their origins in the id. Those functions include perception, motility, memory, and intelligence; all are present from birth and therefore determined, at least in part, by genetic factors. In that expanded view of the ego, the origins of both the ego and the id were thought to be in a common undifferentiated matrix present at birth. That shift toward the primacy of the ego was instrumental in the movement within psychoanalytic theory away from drive theory and toward a greater focus on the complexities of the ego.

Functions of the Ego Although Freud periodically used the term “ego” to mean self, the concept is best viewed as a structure within the personality. It is an agency or an organization of functions that have in common the task of mediating between the instincts and the outside world. Ego psychologists have identified a set of basic ego functions that characterize the ego. The following descriptions reflect the activities that are generally regarded as basic and fundamental to the operation of the ego:

1. Control and regulation of instinctual drives. The development of the capacity to delay or postpone drive discharge, like the capacity to test reality, is closely related to the progression in early childhood from the pleasure principle to the reality principle. That capacity is also an essential aspect of the ego's role as mediator between the id and the outside world. Part of the infant's socialization to the external world is the acquisition of language and secondary process or logical thinking, both of which assist in the control of instinctual drives. The capacity to think in a logical and abstract manner allows for the representation of drives in fantasy, which may circumvent the need to discharge them in action.

The ego's capacity to regulate thinking and to control drive discharge is intimately connected with its defensive functioning. One example of the linkage between control of drives and defensive functioning can be seen in the ego's use of signal affects. Affect states such as guilt, anxiety, shame, and depression serve as signals of the potential breakthrough of threatening impulses from the unconscious. Those signals then act to mobilize defenses in the ego to prevent the breakthrough. That function of the ego is also instrumental in building a capacity to tolerate pain, anxiety, and frustration within manageable limits.

2. Judgment. A closely related ego function is judgment, which involves the ability to anticipate the consequences of one's actions. As with the control and regulation of instinctual drives, judgment develops in parallel with the growth of secondary process thinking. The ability to think logically allows for an assessment of how one's contemplated behavior may affect others. The consequences to oneself can also be ascertained through the use of secondary process thinking. The ego function of judgment may assist regulatory aspects of the ego in the avoidance of impulse discharge. Both those ego functions are commonly impaired in impulsive personality disorders.
3. Relation to reality. The mediation between the internal world and external reality is a crucial function of the ego. The relationship with the outside world can be divided into three aspects: the sense of reality, reality testing, and adaptation to reality. The sense of reality develops in concert with the infant's dawning awareness of bodily sensations. The ability to distinguish what is outside the body from what is inside is an essential aspect of the sense of reality, and disturbances of body boundaries, such as depersonalization, reflect impairment in that ego function.

Reality testing is an ego function of paramount importance in that it differentiates psychotic persons from nonpsychotic persons. Reality testing refers to the capacity to distinguish internal fantasy from external reality. That function of the ego gradually develops in parallel with the increasing dominion of the reality principle over the pleasure principle.

The third aspect, adaptation to reality, involves the ability to use one's resources to develop effective responses to changing circumstances on the basis of previous experiences with reality. One may perceive reality accurately but not use one's full resources to make an informed judgment about the necessary response. In that sense, adaptation is closely linked to the concept of mastery with respect to both control of drives and accomplishment of external tasks. Adaptation to reality is also intimately connected with defensive functions of the ego. One commonly calls on a variety of defensive maneuvers to master situations that may produce anxiety or other affects. To deal with overwhelming trauma, for example, one may use temporary denial to get through the crisis.

4. Object relationships. The significance of object relationships in normal psychological development and in psychiatric disorders was not fully appreciated until relatively late in the evolution of classical psychoanalysis. The capacity to form mutually satisfying relationships is in part related to patterns of internalization stemming from early interactions with parents and other significant figures. That ability is also a fundamental function of the ego in that satisfying relatedness depends on the ability to integrate positive and negative aspects of others and oneself and to maintain an internal sense of others, even in their absence. Similarly, mastery of drive derivatives is crucial to the achievement of satisfying relationships.
5. Synthetic function of the ego. First described by Herman Nunberg in 1931, the synthetic function refers to the ego's capacity to integrate diverse elements into an overall unity. Different aspects of oneself and others, for example, are synthesized into a consistent representation that endures over time. The function also involves organizing, coordinating, and generalizing or simplifying large amounts of data.
6. Primary autonomous ego functions. A direct outgrowth of the work of Hartmann, the primary autonomous functions refer to rudimentary apparatuses that are present at birth and that develop independently of intrapsychic conflict between drives and defenses, provided that what Hartmann referred to as an *average expectable environment* is available to the infant. As noted previously, those functions include perception, learning, intelligence, intuition, language, thinking, comprehension, and motility. In the course of development, some of those conflict-free aspects of the ego may eventually become involved in conflict if they

encounter opposing forces.

7. Secondary autonomous ego functions. Hartmann originally used the concept of the conflict-free sphere of ego functioning to identify areas of primary autonomy. However, that area may be enlarged by functions that originally arise in the service of defense against drives but subsequently become independent of them. Those functions are referred to as *secondary autonomous ego functions*. For example, a child may develop caretaking functions as a reaction formation against murderous wishes during the first few years of life. Later, the defensive functions of the style may be neutralized or deinstinctualized when the child grows up to be a social worker and cares for the homeless. That neutralization, leading to desexualization of libidinal drives or deaggressivization of aggressive drives, provides the ego with independent energies that were formerly used to deal with drive pressures.
8. Defensive functions of the ego. Freud acknowledged the existence of several defense mechanisms, but his writings focused predominantly on repression, which he regarded as the queen of the defenses. In many of his contributions, defense and repression are used almost synonymously. Repression provides a barrier against the direct expression of impulses and wishes from the unconscious. With the ascent of the structural model, Freud shifted the function of defense to the ego. However, defense mechanisms could not be systematically studied until after 1926, when Freud formulated his theory of signal anxiety and the mobilization of defenses in response to danger signals.

The first comprehensive study of defense mechanisms was written not by Sigmund Freud but by his daughter Anna in her landmark 1936 work, *The Ego and the Mechanisms of Defense*. Anna Freud (Fig. 6.1-10) expanded her father's work by providing detailed descriptions of a number of individual defense mechanisms, including regression, reaction formation, undoing, introjection, identification, projection, turning against the self, reversal, and sublimation. She pointed out that everyone uses a characteristic repertoire of defense mechanisms that are intimately related to that person's character. She also insisted that the ego should be the focus of psychoanalytic treatment, in addition to the uncovering of repressed drive derivatives. Her famous observation that “there is depth in the surface” reflected her appreciation of the complexity of the defensive aspects of the ego.

FIGURE 6.1-10 Anna Freud (1895-1982). (Courtesy of Dr. Saul Harrison; photograph given to Harrison by Freud in 1966.)

Evolution of Defense Mechanisms In the course of development, defenses emerge as a reflection of the ego's attempts to mediate between the pressures of the id, on the one hand, and the demands of external reality, on the other hand. At each phase of psychosexual development, then, the specific drive components evoke specific ego defenses. For example, defense mechanisms like introjection and projection are seen in concert with the development of oral incorporative impulses; reaction formations such as shame and disgust develop in response to anal impulses and pleasures. Pregenital defense mechanisms may persist into adult life, lending an infantile cast to the adult personality. In the course of normal development, mechanisms like repression and intellectualization supersede immature defenses.

The defense mechanisms most commonly used by a person to deal with unpleasant situations or distressing internal affective states constitute a significant component of that person's character. Obsessive-compulsive character traits, such as excessive orderliness and miserliness, may initially develop in response to instinctual drives, but they are subsequently divorced from them as they take a central role in the person's overall functioning, even in conflict-free situations. The defenses may be adaptive and healthy, as well as pathological; in normal functioning they are crucial to the preservation of psychological well-being.

In psychopathological states, one may see various alterations of normal defensive functioning. In hysterical and obsessive-compulsive persons, for example, one sees hypertrophied development of particular defenses, as though the dangers from infantile impulses were just as threatening in adulthood as they were in childhood. Exaggerated repression is characteristically associated with hysterical patients; such defenses as reaction formation, undoing, isolation of affect, and intellectualization are associated with obsessive-compulsive patients. Persons who are fixated at the oral psychosexual stage may rely on such defense mechanisms as denial and projection. The failure of defenses may lead to a breakthrough of direct drive expression with obvious regression in the ego's control of thought and affect, a condition most clearly manifested in the schizophrenias.

Classification of Defenses The defenses of the ego can be classified in a variety of ways. Psychoanalysts themselves disagree as to the best taxonomy, and most acknowledge that any one classification is inadequate in its capacity to take into account all the relevant factors. Defense mechanisms can be categorized according to the psychosexual stage with which they are associated, but that system does not address the fact that some defense mechanisms (such as regression) are used in every developmental stage.

Other systems assign defenses according to the nature of the psychopathology with which they are associated. According to that taxonomy, hysterical defenses include conversion, repression, somatization, and dissociation. Freud himself classified defenses according to severity of psychopathology. Denial, distortion, and projection, he believed, are defense mechanisms associated with psychosis. The neurotic defense mechanisms include repression, isolation, undoing, displacement, and reaction formation. He considered mature defenses to be sublimation, altruism, humor, and suppression. A classification that draws both on the level of psychopathology and on developmental maturity is provided in Table 6.1-2, along with definitions of commonly recognized defense mechanisms.

Table 6.1-2 Classification of Defense Mechanisms

Although analysts disagree on the total number of defense mechanisms, most agree with Freud's assessment that defense mechanisms must possess the following properties: (1) they manage instinct, drive, and mood; (2) they are unconscious; (3) they are discrete; (4) they are dynamic and reversible; and (5) they can be adaptive or pathological.

Superego The third component of the tripartite structural model is the superego. The superego establishes and maintains the person's moral conscience on the basis of a complex system of ideals and values internalized from one's parents. As noted previously, Freud viewed the superego as the heir to the Oedipus complex; it is, therefore, influenced by the resolution of oedipal conflict. It conducts an ongoing scrutiny of the behavior, thoughts, and feelings of the person; makes comparisons

with expected standards; and offers approval or disapproval. Criticisms and reproaches lead to a variety of painful feelings; praises and rewards raise the person's self-esteem. Those activities occur unconsciously to a large extent, reflecting the clinical observation that self-criticism operates as much outside conscious awareness as aggressive and sexual drive derivatives do.

The idea of the superego first appeared in 1896 with Freud's *Further Remarks on the Neuro-Psychoses of Defence*, in which he described "self-reproaches which have reemerged in transmuted form and... relate to some sexual act that was performed with pleasure in childhood." In early discussions of dreams, Freud commented on the activities of a self-criticizing agency that censors the entry of unacceptable ideas into consciousness because the ideas are morally offensive. In his 1914 paper "On Narcissism," Freud spoke of a special self-critical agency and of a hypothetical state of narcissistic perfection originating in children's idealized views of themselves. As that idealized version of self is tarnished through the slings and arrows of development, the child attempts to recover the lost narcissism by creating an ego-ideal. Having postulated that construct, Freud needed to create a special agency whose function it was to be sure that the person's behavior and thoughts measure up to the expectations of the ego-ideal. The construct of the superego grew out of the need for monitoring and preserving the ego-ideal.

Ego-Ideal The distinction between the superego and the ego-ideal has always been ambiguous. The most common usage today is to refer to the ego-ideal as an agency that *prescribes* what one should do and to refer to the superego as an agency of moral conscience that *proscribes*—that is, dictates what one should *not* do. Freud, however, did not always make that distinction. In 1921 he referred to the self-critical agency responsible for guilt and self-reproaches in melancholia and depression as the ego-ideal. In *The Ego and the Id* 2 years later, he assigned both the conscience and the ego-ideal to the superego. Today most analysts conceptualize the ego-ideal as comprising a set of functions contained within the structure of the superego.

The ego-ideal is best conceptualized as an amalgam of internalized representations. The ideal object, stemming from internal images of an admired and omnipotent parent, is one such representation. The ideal self, based on fantasies of a self-image that would result in maximal parental approval, is another representation contained within the concept. Representations of actions that should be done to attain an ideal relationship with significant figures in one's life are also involved.

The ego-ideal becomes an internal standard of what one should be. In this regard it is intimately connected with shame, which is a response to an internal perception that one has not lived up to one's ego-ideal. Although guilt is a closely related affect, it is generally regarded as a response to the transgressions of superego prohibitions, particularly sexual or aggressive wishes toward others.

Origins of the Superego As described earlier, the superego is a relatively mature structure resulting from the resolution of the Oedipus complex. In the case of the little boy, his love for his mother engenders murderous wishes toward his father. His observation of the genital differences between males and females leads him to fear that his oedipal wishes will be punished with castration. Under the influence of the perceived threat from his father, the little boy renounces his incestuous longings for his mother and identifies with his father. That resolution involves the installation of the parental moral values as the superego structure.

In the case of the little girl, she renounces oedipal strivings for her father because she realizes that a victory over her mother would result in the loss of her mother's love. Freud initially postulated that little girls grow up with less moral conviction and character as a result of the absence of the castration threat. That view is no longer held by contemporary psychoanalysts. Recent empirical studies suggest that moral developmental differences between boys and girls do exist but along different lines. Boys tend to place high moral value on achievement and fair play; girls are more likely to develop a moral code based on affiliation and relatedness to others.

Evolution of the Superego Primitive superego precursors are formed early in life from the internalization of frightening and aggressive perceptions of parental figures. In some cases those perceptions are based on real behaviors of the figures; in other cases the perceptions are influenced by distortions deriving from fantasies by the child and from projections of the child's rage and sadism onto external objects. As a child matures and resolves the Oedipus complex, identifications with both parents become integrated to form an intricate and well-rounded internal object representation within the ego that interacts with the other contents of the ego as a superego. The parental identification is further reinforced by the child's struggles to repress unacceptable sexual and aggressive wishes and by the child's identification with the parent's superegos.

The ego originates to a large extent through interactions with the external world. The superego, however, is much more closely related to the id in its origins. It emerges from a powerful struggle with the drive pressures emanating from the id and is, therefore, more internal in its origins than is the ego. The identifications constituting the superego continue to undergo modification as role models and authority figures in one's life become incorporated into a set of internal standards, values, aspirations, and ideals.

As the oedipal phase of psychosexual development is resolved, the child moves into the stage known as *latency*. That stage, ranging from age 5 or 6 until the onset of puberty, is a period of relative quiescence, when the pressures from instinctual drives are more or less under control. Punishments and restrictions previously imposed by the parents are now firmly ensconced in the superego, which judges and guides behavior even in the absence of external authorities.

With the onset of puberty and the entry into adolescence, sexual and aggressive drives resurge. The superego is undermined as abandoned incestuous strivings are rekindled. The well-known phenomenon of adolescent rebellion can be understood as both an instinctual mutiny against the prohibitions of the superego and an attempt to take flight from the forbidden impulses toward parents arising from the hormonal changes of puberty. A major developmental task of adolescence is to modify the superego from a rigid and overly punitive structure to one that allows and sanctions an adult sexual object choice.

THEORY OF ANXIETY

The development of Freud's understanding of anxiety is of fundamental importance to the theory and the practice of psychoanalysis. His clinical observations in the 1890s led him to regard anxiety as the result of "dammed-up libido." In other words, a physiological increase in sexual tension leads to a corresponding increase in libido, the mental representation of that physiological event. The normal outlet for such tension is, in Freud's view, sexual intercourse. However, abnormal sexual practices, such as abstinence and coitus interruptus, prevent tension release and result in *actual neuroses*. Those conditions of heightened anxiety related to libidinal blockage include neurasthenia, hypochondriasis, and anxiety neuroses, all of which Freud regarded as having a biological basis.

Freud differentiated the actual neuroses from the *psychoneuroses*—hysteria, phobias, and obsessional neuroses. He understood those conditions and the anxiety associated with them to be primarily related to psychological factors, rather than physiological factors. Intrapsychic conflict is responsible for anxiety in psychoneuroses, and Freud observed that the resulting anxiety is less intense and dramatic than what he observed in actual neuroses.

Freud's early theory of anxiety cannot account for objective anxiety related to the threat of real danger that is unrelated to accumulated sexual tensions. That gap in Freud's theory led him to expand his conceptual framework of the pathogenesis of anxiety to include self-preservation and the adaptive function of anxiety.

Anxiety as a Signal With the publication of *Inhibitions, Symptoms and Anxiety* in 1926, Freud created a new theory of anxiety that reflects the ascendence of the structural model over its topographic predecessor. Accordingly, the role of instinctual drives and physiological pressures is of less importance in that theory. Instead, the new theory accounts for both real external anxiety and neurotic internal anxiety as a response to a danger situation.

Freud identified two types of anxiety-provoking situations. One situation involves overwhelming instinctual stimulation, the prototype of which is the experience of birth. In such situations, the excessive amount of drive pressure penetrates the protective barriers of the ego, producing a state of helplessness and trauma. The second and more common situation involves anxiety that develops in *anticipation* of dangers, rather than as the result of those dangers. That warning to the organism, known as *signal anxiety*, operates at an unconscious level and serves to mobilize the ego's resources to avert the danger. Either external or internal sources of danger may produce such a signal that leads the ego to marshal specific defense mechanisms to guard against or to reduce the degree of instinctual excitation.

Freud's new theory of anxiety explains neurotic symptoms as a partial failure of the ego to cope with distressing stimuli. The drive derivatives associated with danger have not been adequately contained by the defense mechanisms used by the ego. In phobias, for example, Freud explained the fear of an external threat (such as dogs or snakes) as an externalization of an internal danger. In psychotic conditions the breakdown of ego-defensive functioning is so thoroughgoing that associated distortions of ego functioning take place in an effort to accomodate the idiosyncratic distortions of the external world perceived by the patient.

Although Freud focused much more of his attention on signal anxiety in the latter part of his career than in his early career, he always maintained that physiologically based actual neuroses exist alongside the psychoneuroses. Freud's differentiation of two forms of anxiety—one biologically mediated and the other psychologically generated—proved to be remarkably prophetic. That distinction parallels what we now view as panic disorder and generalized anxiety disorder. Panic disorder may be

mediated by stimulation of the locus ceruleus, the largest noradrenergic nucleus of the brain. Dysregulation of the g-aminobutyric acid system seems to trigger that form of anxiety; psychological factors, unrelated to the locus ceruleus, appear to trigger a milder form of anticipatory or generalized anxiety. It has been postulated that an underlying neural structure may lead to a diathesis for the more severe form of anxiety and that the absence of such a neural structure leads to a milder form of expression, signal or anticipatory anxiety.

Developmental Hierarchy of Anxiety Freud realized that each stage of psychological development contains a characteristic danger situation involving phase-specific issues. Hence, a developmental hierarchy of anxiety can be constructed based on the specific fears associated with each stage. The earliest danger situation is a fear of disintegration or annihilation, often associated with concerns involving fusion with an external object. As one matures to the recognition of the mothering figure as a separate person, separation anxiety or fear of the loss of the object becomes more prominent.

During the oedipal psychosexual stage, the girl child is most concerned about losing the love of the most important figure in her life, her mother. The male child is primarily anxious about bodily injury or castration. After the resolution of the oedipal conflict, a more mature form of anxiety occurs, often termed *superego anxiety*. That latency-age concern involves the fear that internalized parental representations, contained in the superego, will cease to love the child or will angrily punish the child. The symptoms of anxiety may have determinants from several of those developmental stages. A major task for the psychoanalytic clinician treating anxiety is to understand the particular danger situations and their associated developmental phases that have contributed to the anxiety of a particular patient in a particular situation.

Implications of Theory Freud's theory of signal anxiety heralded a decisive shift in his formulation of the ego. Before the introduction of the structural theory in 1923, Freud regarded the ego as a passive, weak, and fragile agency that responds to the bidding of the instinctual drives from the id and the punitive demands of the superego. In his revision of his anxiety theory, however, the ego is an executive agency that exerts control over the forces of the id. Anxiety appearing before repression is operative, and the ego is endowed with sufficient autonomy to exercise certain functions on its own.

The conceptualization of anxiety as a signal also has important clinical implications. Anxiety should not be viewed simply as a pathological symptom to be removed. A healthy ego must develop the capacity to use anxiety signals to mobilize ego resources in efforts at mastery and control. Indeed, psychoanalysts view a person's capacity to tolerate anxiety as a significant sign of ego integration and psychological health.

PSYCHOANALYTIC CONCEPT OF CHARACTER

Although the term “character” has moral overtones in popular usage, in psychoanalysis, *character* refers to the enduring patterns of functioning typical of a person. Others know a person's character by that person's habitual way of behaving, feeling, speaking, and thinking. A variety of theoretical underpinnings have been proposed to explain character formation. As previously noted, Freud linked certain character traits with certain psychosexual stages. For example, persons who are developmentally fixated at the anal stage are likely to manifest orderliness, obstinacy, and miserliness. In a classic paper on anal eroticism and character, Freud concluded that permanent character traits represent “unchanged prolongation of the original instincts, or sublimation of those instincts, or reaction formation against them.”

Freud was faced with a challenge in differentiating neurotic symptoms from character traits. He suggested that neurotic symptoms represent a “return of the repressed,” reflecting his formulation that the failure to repress drive derivatives leads to neurotic symptom formation. In contrast, he postulated that character traits are the end result of the *successful* use of repression and other defenses, particularly reaction formation and sublimation. Freud's development of the structural model allowed him to expand his understanding of character formation further by including identification as a defense. He observed that a person may be able to relinquish an object attachment only by a process of identification, in which the lost object is firmly established inside the ego. Through a series of identifications, children develop their own characters. The formation of the superego is a related identificatory process that also contributes to the end result of a unique character structure.

Karl Abraham delineated character traits linked to oral, anal, and genital roticism and, in so doing, advanced Freud's rudimentary ideas to a much more sophisticated and detailed description of character. Wilhelm Reich also made significant contributions by identifying the close relationship between resistance to psychoanalytic treatment and the nature of the patient's character traits. Reich's observations about character resistance remain useful today in the work of contemporary psychoanalytic clinicians.

Current Concepts of Character Today the concept of character and that of personality are used interchangeably, particularly in discussions concerning Axis II personality disorders and traits. Psychoanalytic interest has expanded beyond specific character traits to a broadened understanding of character as a series of compromise formations between instinctual wishes and opposing defenses, on the one hand, and constellations of internal representations of self and others, on the other hand. Increasing evidence from studies of twins and other research indicate that genetic and constitutional factors play key roles in the development of character. The unique stamp of personality of any one person appears to be the final common pathway of a complex mixture of innate biological predispositions and environmental-psychological factors. For example, recent research suggests that shyness involves an inherited diathesis that requires particular environmental circumstances to attain phenotypic manifestations.

Discussions of character lead directly to related concepts of self, ego, and identity—all of which have developed varied meanings in contemporary psychoanalytic usage. The following discussion of object relations theory attempts to explicate those terms in more detail.

OBJECT RELATIONS THEORY

The ego-psychological approach to psychoanalysis, often referred to as the classical view, is one of three major theoretical frameworks that are widely used by modern psychoanalytic clinicians. The other two theoretical schools, object relations theory and self psychology, are both emblematic of a trend within psychoanalysis away from an emphasis on drives and defenses and toward an increasing interest in relationships. Self psychology, which is discussed later in this section, derives from the interpersonal tradition established by Harry Stack Sullivan but most eloquently formulated and elaborated by Heinz Kohut and his followers. Object relations theory originated in the work of Melanie Klein and members of the British school, including D.W. Winnicott and W.R.D. Fairbairn.

Basic Concepts The ego-psychological perspective conceptualizes drives as primary and object relations as secondary. In that model the infant is motivated by the wish to discharge tension under the pressure of instinctual drives. By contrast, object relations theory stresses the fact that all drives emerge in the context of a mother-infant relationship; therefore, object seeking must be considered a motivation equal to or of even greater importance than drive discharge.

Freud made many references to the importance of internalized objects. For example, in his 1923 landmark work on the structural model, *The Ego and the Id*, he noted that “the character of the ego is a precipitant of abandoned object-cathexes and... contains the history of these object choices.” In other words, he firmly believed that a person's ego results from identification with such external objects as parental figures. Similarly, the superego and the ego-ideal are conceptualized as internalizations of aspects of one's parents.

Freud laid the groundwork for a theory of object relations within the framework of the structural model, but his followers began to notice certain limitations of the model. For example, it became increasingly apparent that the id is not simply an amorphous cauldron of explosive drives. Aggressive and sexual wishes are connected with fantasies and meanings organized around relationships. Similarly, the view of the ego as subservient to an internal critical agency called the superego is seen as having limited clinical usefulness. The superego has both loving and critical aspects to it, and the complexity of its interaction with the ego can most usefully be conceptualized as an internal object relationship. Conceptualizing an object relationship strictly in terms of its cathexis with drive energy does not do justice to the full range of needs and moods associated with relationships. Such wishes and needs as affirmation, safety, reassurance, and self-esteem are not easily subsumed under the dual-instinct theory.

In strict terms, object relations theory is not an interpersonal model of psychoanalysis. It is a theory of unconscious internal object relations involving the transformation of interpersonal relationships into internalized structures. As infants grow and develop, they internalize an experience of a self in relationship to an other. The feeding experience is perhaps the best prototype of the process. While a hungry infant is screaming for its mother, a template of unpleasant experience is laid down in its brain, involving an experience of the self as greedy, demanding, and angry and an experience of an object (the mother) as frustrating, unavailable, and inattentive. That self-object constellation is also accompanied by powerful moods.

When the mother finally feeds the baby, the self-object-moods unit is transformed into a positive experience of the self, a positive view of the object as nurturant and attentive, and a positive affective experience of pleasure and satiation. At around 16 months of age, the infant's cognitive and perceptual apparatuses have

sufficiently developed so that those two experiences are internalized as two opposing sets of self-object-mood units.

That examination of the feeding experience illustrates a basic tenet of object relations theory—namely, that object relations always involve an interaction between a self and an object associated with an affect. Internalized representations of self and object do not occur in isolation. A second point of crucial importance is that the structures laid down by the child are significantly determined by the child's fantasy life. For example, the mother who is not responding to her child's demands to be fed may actually be a perfectly competent mother who is simply occupied with other siblings. However, the infant's fantasies may involve a mother who is evil, rejecting, and abandoning, so that the internal structure laid down by the child may include such distorted characteristics. Freud was clearly aware of that phenomenon when he noted in 1940, “It is a remarkable thing that the super-ego often displays a severity for which no model has been provided by the real parents.” Hence, a one-to-one correlation may not exist between the real external object and the internalized object.

Object relations theory also views motivation differently from Freudian ego-psychological theory. For example, the infant transforms its mother into an internal soothing presence as a way of dealing with its fear of losing its mother. In that manner both the external mother and the internalized version of the mother represent safety and an enhanced sense of well-being for the child, independent of drive discharge. Internalizing bad or negative aspects of the mother is more complex and involves the fantasy of gaining control over the bad object by capturing it within oneself, developing mastery through repetitive traumatic experiences with the object, yearning to transform the bad object into a good object, and needing to hang on to the bad object because it is preferable to no object at all.

The development of the self-object units in the child's internal world reflects the operation of two distinct internalization mechanisms. *Introjection* involves the taking in of an object that continues to exist as an other, an internal presence that may be experienced as a soothing companion or a critical parent, as in the case of the superego. Introjects, then, continue to be experienced as *objects*, rather than as parts of the child's self. *Identification*, in contrast, involves a modification of the self on the basis of the internalization of a significant external figure who is used as a model. That assimilation process in identification leads to the child's experience of those parental qualities as part of the self, rather than as a foreign body, as in the case of introjection. Identification with an introject may also take place over the course of development.

Whereas ego psychology regards conflict as struggles between wishes and defenses or between intrapsychic agencies, such as the id and the superego, object relations theory views conflict quite differently. Unconscious conflict is perceived as a struggle between different self-object–affect units striving for center stage in the psyche. The notion of character is also conceptualized in a different manner. Instead of the ego-psychological view that the end result of instinctual drives and specific defensive operations deals with those drives, character is viewed as heavily influenced by the predominant constellations of self-representations and object representations deriving from introjections and identifications.

Historical Evolution Object relations theory as a separate school of psychoanalytic thought originated in the United Kingdom. The British Psychoanalytic Society is divided into three segments as a result of disagreements that took place before and during World War II. Those three discrete groups, still in existence today, include the A group, followers of Melanie Klein; the B group, followers of Anna Freud; and the independent group, who are often referred to as the middle groupers. The third group—consisting of such notable clinicians as Winnicott, Fairbairn, Harry Guntrip, Michael Balint, Margaret Little, and John D. Sutherland—are responsible for the theory of object relations as it is known today. They formed the British school of object relations out of a wish to avoid taking sides with either the Kleinians or the Anna Freudians. However, they owe a substantial debt to Melanie Klein, who is generally acknowledged to be the pivotal figure in the shift from classical theory to object relations theory.

Melanie Klein Heavily influenced by Freud, Klein evolved a theory of internal object relations that was intimately linked to drives. She arrived in England in 1926, having emigrated from Budapest and Berlin. Her unique perspective largely grew out of her psychoanalytic work with children, in which she became impressed with the role of unconscious intrapsychic fantasy. She believed that a fear of annihilation associated with Freud's death instinct was a key factor in the first few months of the infant's life. She postulated that the ego undergoes a splitting process to deal with that terror; derivatives of the death instinct—such as aggression, hatred, sadism, and all other forms of “badness”—are disavowed and projected onto the mother.

Klein viewed projection and introjection as the primary defensive operations at that stage of life. After projecting the death-instinct derivatives into the mother, the infant begins to suffer from persecutory anxiety, in which the infant lives in dread of an invasive attack from the mother. In that terrifying fantasy, the “bad mother” created by the infant's projections gets inside the infant (i.e., the “bad mother” is reintrojected) and destroys the remaining “goodness” (i.e., libidinal derivatives) that were originally protected by the splitting and projection of the “badness.”

That persecutory anxiety characterized what Klein called the *paranoid-schizoid position*, the infant's mode of organizing experience in which all aspects of the infant and the mother are split into good and bad elements. Klein thought that external and internal objects cannot be clearly distinguished because the actual figures in the external world are heavily colored by the infant's projections and ensuing perceptions are distorted. She conceptualized oscillating cycles of introjection and projection to keep bad aspects of the self separated from good aspects of the self and to keep the bad aspects of objects separated from the good aspects. In that manner the child can prevent good, loving aspects of its experience from being destroyed by bad, hateful aspects of experience. In the second half of the first year of life, according to Klein's developmental timetable, the child's part-object world (i.e., a world containing an “all-bad” mother and an “all-good” mother) begins to change. The child becomes aware that the hateful, rejecting mother and the loving, nurturing mother are simply aspects of the same person. As those disparate views are integrated, the infant becomes concerned that it may have harmed or destroyed its mother through its hostile and sadistic fantasies directed toward her. At that developmental point the child has arrived at the *depressive position*, in which the mother is viewed ambivalently as having both positive and negative aspects and as being the target of a mixture of loving and hateful feelings.

In contrast to the paranoid-schizoid position, in which the primary concern of the infant is external attack, in the depressive position the infant's primary concern is that it may harm love objects, particularly the mother. The possibility of losing the good object through sadistic and aggressive impulses produces guilt feelings in the child. Anxiety about loss of the love object through one's own destructiveness is known as *depressive anxiety*. To deal with guilt feelings, the child engages in a process of *reparation*, in which efforts are made to ensure that love prevails over hate and that damage done through one's own destructiveness is repaired through loving behavior. Because the maternal object is recognized as a whole object in the depressive position, Klein postulated that the Oedipus complex begins in the second half of the first year, and she recast the whole oedipal constellation as an effort to resolve depressive anxieties and guilt through reparation.

Kleinian theory has been criticized for its significant shortcomings. One of its primary difficulties is the assumption that an infant in its first year of life possesses a highly sophisticated capacity for abstract and conceptual thinking. The infant's perceptual-cognitive development is actually too rudimentary when that young to be capable of elaborating the sophisticated fantasies that Klein postulates. Hence, her developmental timetable, which compresses preoedipal and oedipal phases into the first year, is no longer tenable.

Klein's emphasis on envy and greed have considerable clinical utility, but her view that such states, as well as aspects of sadism and aggression, derive from the death instinct has fallen out of favor. Klein has also been criticized for relying too heavily on fantasy and thereby discounting the role of actual environmental trauma on children and the influence of real parental objects on the child's development. Post-Kleinian theorists have suggested that the paranoid-schizoid and depressive positions are not developmental phases that are outgrown but, rather, two modes of generating experience that continue in a dialectic with one another throughout adult life. Viewed from that perspective, Klein's formulations have a great deal of clinical value, particularly in the treatment of highly disturbed patients.

Independent Group The British analysts who constitute the middle group had several key features in common. First, like Klein, they were involved in the psychoanalytic treatment of patients who were more severely disturbed than the classical neurotic patient. As a consequence of that particular clinical experience, they tended to focus on preoedipal development and dyadic object relations, rather than the triangular oedipal constellation. Also, they counterbalanced Klein's focus on intrapsychic fantasy with detailed attention to the actual environment provided by the mothering figure in the earliest months of life. The group was also intensely interested in a theory of *deficit*—that is, a theory that encompassed the effects of insufficient maternal nurturance on the developing child and that conceptualized certain forms of psychopathology as involving the absence of or an insufficiently developed intrapsychic structure, rather than interagency conflict. Each of the leading figures of the British school of object relations theory made unique contributions.

W.R.D. FAIRBAIRN Although heavily influenced by Klein, Fairbairn (Fig. 6.1-11) reversed her focus on the role of fantasy in the infant's creation of its objects. He saw the causes of his patient's difficulties in their mothers' failure to provide appropriate nurturance, rather than in the frustration of drives. He thought that infants have a basic need for a kind of love and acceptance from the mother that assures the infants that they are persons in their own right. In that regard he conceptualized the need for the recognition of personhood as primary and the physiological needs related to the drives as secondary. He did not deny the existence of libido and aggression but thought that they are fundamentally object-seeking, rather than pleasure-seeking.

FIGURE 6.1-11 W.R.D. Fairbairn. (Courtesy of John D. Sutherland, Edinburgh, Scotland.)

Although Klein stressed that fantasy is the earliest and most basic activity of the infantile mind, Fairbairn regarded fantasy as substitutive rather than primary. The welfare of the developing child lies fundamentally in the experience of the mother's acceptance of it, and internal objects are elaborated only in compensation for unsatisfactory relationships with real external objects.

Fairbairn is particularly known for his insight into the primacy of the schizoid condition. He suggested that, when the mothering figure fails to fulfill the baby's basic needs, the infant splits the object and the ego into various components, leading to a schizoid state. Hence, in contrast to classical theory, the Fairbairn theory views the ego as whole at birth, only to become split as it encounters unfavorable experiences with early mothering figures. In the Fairbairn model, libidinal object-seeking is primary, and aggression occurs as a natural defensive reaction to the frustration of the libidinal drive.

D.W. WINNICOTT Of all the middle groupers, Winnicott is probably the best known. Like Fairbairn, he was a Kleinian revisionist who gave greater credence to the role of external reality in development. He was particularly interested in the creation of a *holding environment* by the mother that made it ultimately possible for the child to develop awareness as a separate person. In Winnicott's view the mother plays a key role in bringing the world to the child and in offering empathic anticipations of the infant's needs. The mother provides a fundamentally crucial relationship that enables the nascent self of the infant to emerge. Trained as a pediatrician, Winnicott gave primary attention to the mother-infant dyad. Little of importance was written about the father in Winnicott's work.

Winnicott coined the term *good-enough mothering* to designate a mothering figure who is able to provide an optimal amount of comfort and constancy for the infant, so that the infant can proceed along normal developmental lines. Such a mother tunes in to the infant and meets the child's omnipotent needs without challenging them and offers her support according to the infant's own timetable, rather than imposing herself on the child as a consequence of her own needs. She also lets the infant separate at the appropriate time and allows for an optimal disillusionment of the child's omnipotence.

Winnicott was also responsible for the concept of the *transitional object*, the infant's not-me possession, usually a reassuring blanket or toy. The object, often suffused with the mother's scent, enables the child to separate emotionally from the mother without undue anxiety. The transitional object preserves the illusion of the comforting maternal object, even in her absence. Winnicott described *transitional phenomena* as objects, sounds, sights, and smells that exist in an intermediate realm that has elements both of external reality and of the child's own subjectivity. Art, religious experience, and creativity may well derive from that sphere of experience.

Winnicott's distinction between the *true self* and the *false self* is widely used in clinical discussions. The true self refers to the inherited potential that constitutes the kernel of the child. The true self is the child's authentic being that will emerge if a good-enough mothering experience with an appropriate holding environment is offered. The false self is a facade constructed by the child in reaction to a self-involved mother who insists on certain responses from her child. The false self may develop to protect the true self by complying with the maternal demands.

Winnicott's conceptual framework of the mother-infant relationship led to a form of psychoanalytic treatment that radically departed from the classical approach. He thought that the analyst must provide the appropriate facilitating environment that was missing during the patient's early development. Only then can the true self, which was frozen in a state of developmental arrest, continue its growth and overtake the false self constructed by the patient. Like Fairbairn, Winnicott thought that psychopathology results from the mother's failures, rather than from interagency conflict or intrapsychic fantasy. His own technique took on a unique character in that he advocated a prolonged holding period before making interpretive interventions.

MICHAEL BALINT Balint thought that the search for the primary love object underlies virtually all psychological phenomena. The infant wishes to be loved totally and unconditionally, and, if the mother is not forthcoming with appropriate nurturance, the child devotes its life to a search for the love that was missed in childhood. Balint described the feeling in many patients that something is missing as *the basic fault*. Like Fairbairn and Winnicott, Balint understood that deficit in internal structure to be the result of maternal failures. He viewed all psychological motivations as stemming from that failure to receive adequate maternal love.

Unlike Fairbairn, Balint did not entirely abandon drive theory. He suggested that libido, for example, is both pleasure-seeking and object-seeking. He also worked with primitively organized patients, and, like Winnicott, he felt that certain aspects of psychoanalytic treatment occur at a more profound level than verbal explanatory interpretations. Although some material involving genital psychosexual stages of development can be interpreted from an intrapsychic conflict perspective, Balint believed, certain preverbal phenomena are reexperienced in analysis, and the relationship itself is decisive in dealing with that realm of experience.

Mixed Models When object relations theory began to cross the Atlantic and to influence North American thinkers, attempts were made to integrate American ego-psychological concepts with the object relations perspectives of Klein and members of the British school. Two prominent contributors to that mixed-model approach are Edith Jacobson and Kernberg. The melding of ego psychology and object relations theory typical of their work has led Jacobson and Kernberg, as well as those influenced by them, to be categorized as American object relations theorists.

Edith Jacobson In contrast to members of the British school, Jacobson understood the infant's disappointment with the maternal object to be not necessarily related to actual failure by the mother. In Jacobson's view, disappointment is always related to a specific, drive-determined demand, rather than to a global striving for contact or engagement. She viewed the infant's experience of pleasure or unpleasure as the core of the early mother-infant relationship. Satisfactory experiences lead to the formation of good or gratifying images, whereas unsatisfactory experiences create bad or frustrating images. Normal and pathological development are based on the evolution of these self-images and object images. As far as Jacobson was concerned, the concept of fixation refers to modes of object relatedness, rather than modes of gratification.

Jacobson believed that the structural model and an emphasis on object relations are not fundamentally incompatible. She thought that the ego and self-images and object images exert a reciprocal influence on one another's development. The ego, as it matures, integrates early pleasure-unpleasure experiences into partial primitive images of the self and the object. Subsequent events continue to be divided into either gratifying or frustrating experiences.

By the second year of life, with further maturation of the ego, the child is able to distinguish specific features of the love object and to entertain the notion of being like the admired object, rather than of becoming the object. If conditions are favorable, selective identification, the tendency to be like the object, gradually replaces the tendency to regress into merger fantasies. Eventually, competition with peers and the same-sex parent contributes to stable ego identification and the establishment of an ego-ideal. Jacobson viewed the ego-ideal as a fusion between ideal self-images and ideal object images that partially compensate for the lost fantasies of merger.

For Jacobson the superego develops in three phases. The early superego consists of archaic, sadistic images formed on the basis of introjective and projective processes. The second phase involves the fusion of ideal self-images and ideal object images into the entity of the ego-ideal. The third and final stage involves the internalization of realistic parental demands, values, and prohibitions.

Otto Kernberg Perhaps the most influential of the American object relations theorists, Kernberg was substantially influenced by both Klein and Jacobson. Much of his

theory was derived from his clinical work with patients suffering from borderline personality disorder (discussed later). In brief, Kernberg places great emphasis on the splitting of the ego and the elaboration of good and bad self-configurations and object configurations. Although he continues to use the structural model, he views the id as composed of self-images, object images, and their associated affects. Drives appear only to manifest themselves in the context of internalization of interpersonal experience. Good and bad self-relationships and object relationships become associated, respectively, with libido and aggression. Not only do object relations constitute the building blocks of structure, but they are also the building blocks of drives. Goodness and badness in relational experiences precede drive cathexis. In other words, the dual instincts of libido and aggression arise from object-directed affective states of love and hate.

The influence of Klein on Kernberg is apparent in his view of infantile development as moving from splitting toward integration. Idealized or good images of self and object are gradually integrated with devalued or bad self-images and object images, leading to ambivalent, whole-object configurations of self and other. That integration leads to feelings of guilt, concern, and mourning, much as Melanie Klein described the process. Kernberg's view of the superego follows Jacobson's model in delineating three layers based on the three developmental phases of Jacobson.

The principal features of the major object relations theorists are summarized in Table 6.1-3.

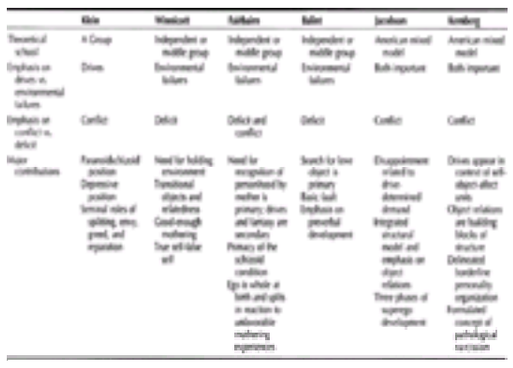

Table 6.1-3 Object Relations Theorists

Defense Mechanisms in Object Relations Theory Although the clinical work of ego psychologists has focused primarily on neurotic defense mechanisms, the object relations theorists have traditionally worked with more disturbed patients. That clinical perspective has delineated a set of primitive defense mechanisms characteristic of severe personality disorders and psychoses that are in common usage by object relations theorists.

Splitting Although Freud described splitting of the ego in his discussions of fetishism, Klein was the first analyst to recognize the universal importance of splitting in early development. It can be defined as an unconscious process that actively separates contradictory self-representations, contradictory object representations, or contradictory feelings from one another. Splitting is crucial for emotional survival because it allows the infant to separate the negative views of the mother as a rejecting, unavailable figure from the positive views of the nurturing, feeding mother, so that the feeding experience itself is not contaminated by the terrifying anxieties about the “bad mother.” Splitting helps the developing child separate love from hate, pleasure from unpleasure, good from bad, and positively colored experience from its negative counterparts. Splitting may also be viewed as a basic mode of ordering experience into separate compartments on the basis of the fear that all that is good will be destroyed by all that is bad unless the good and the bad are kept separate. Splitting is secondarily elaborated into a psychological defense.

Kernberg has been instrumental in making the defense of splitting clinically relevant to the diagnostic understanding and treatment of severe personality disorders. He has identified four common clinical manifestations of splitting: (1) the coexistence of contradictory and alternating self-representations that cause the patient to look quite different from day to day, (2) the division of persons in the environment into an idealized or all-good group and a devalued or all-bad group, (3) selective problems with impulse control, and (4) the expression of behaviors and attitudes that alternate and are contradictory but are regarded by the patient with lack of concern and bland denial. Although Kernberg has cited splitting as the key defensive operation in borderline patients, it may also be found in psychotic patients and those with neurotic or higher-level personality disorders.

Projective Identification Splitting works hand in hand with the second defense mechanism, projective identification, which is an unconscious three-step process by which object representations or self-representations are disavowed and attributed to someone else. The three steps involve the following:

1. The patient unconsciously projects an object representation or self-representation into the clinician.
2. The clinician unconsciously identifies with what is projected and begins to behave or feel like the patient's projected object representation or self-representation in response to interpersonal pressure exerted by the patient's behavior. That step in the phenomenon has been referred to as *projective counteridentification*.
3. The clinician psychologically processes the projected contents, which are, therefore, modified to some extent before finally returning to the patient through reintrojection. When those modified projected contents are returned to the patient, there is a corresponding modification of the patient's internal object relatedness.

This model is somewhat artificially linear for purposes of explication, but in the crucible of clinical work, the projective and identificatory aspects may occur virtually simultaneously. Moreover, the patient's mental contents are not mystically transferred to the clinician's internal world. A more accurate description of what actually happens is that the patient's interpersonal behavior evokes repressed aspects of the clinician that ordinarily are not regarded by the clinician as part of his or her continuous experience of self. Hence clinicians in the throes of projective identification may experience themselves as having been taken over by “alien” forces, that is, their experience is one of “I am not acting like myself.” In reality, they are simply under the sway of aspects of themselves that are ego-dystonic. In other words, there is a repressed and unconscious aspect of the clinician that is a “good fit” with what is being projected by the patient; the patient does not project into an empty container. The pre-existing aspects of the clinician may help to determine why some clinicians are more easily “provoked” by certain patients.

Projective identification as defined in this three-step model can be viewed as both an intrapsychic defense mechanism and an interpersonal process. Splitting and projective identification work in tandem to separate good and bad elements from one another. For example, a woman who has introjected a physically abusive parent may use projective identification to keep the internal bad object representation outside herself, so that she can control it. Her behavior may coerce treaters to assume the role of the abusive parent and to struggle with feelings of anger and hatred toward her. The treater's feeling of being controlled or bullied into assuming a specific role vis-à-vis the patient is a hallmark of projective identification. Although Melanie Klein originated the concept of projective identification, Wilfred Bion elaborated it to encompass a developmental situation in which the infant projects unacceptable aspects of itself into the mother, who serves as a container for the infant's projections before returning them in a modified form.

Projective identification is a controversial term, and some authors have limited its use to describe an intrapsychic defense that may or may not result in a complementary response from the clinician. In that model the identification process is regarded as occurring within the patient, rather than within the clinician who is the target of the projection. By maintaining the empathic bond or identification with the projected contents, the patient can maintain a fantasy of complete control over the target of the projection and over the projected material itself.

The distinction between projection and projective identification is also somewhat controversial. However, if one assumes that the three-step model involves both an intrapsychic defense and an interpersonal process, projective identification requires a transformation of the target of the projection. Projection does not require such a transformation. Paranoid patients, for example, may assume that people who are walking up and down the street are members of the Central Intelligence Agency, but that projection of malevolence does not alter the feelings or the behaviors of those people.

Primitive Idealization and Devaluation Primitive idealization and devaluation also occur in concert with splitting. External objects are viewed as either all good or all bad as a way of preventing the destruction of the loving aspects of others and of the self by their hateful counterparts. Hence, by attributing omnipotence and perfection to certain figures in the environment, a patient may bask in the reflected glory of the object and defend against feelings of contempt and envy toward the object. Similarly, by devaluating another person as thoroughly worthless, the patient may defend against painful feelings of envy and inferiority.

Concept of the Self Whereas ego psychologists concentrated their efforts on the ego, object relations theorists have sought to clarify the role of the self in psychoanalytic theory. The confusion between the self and the ego goes back to Freud's ambiguous usage of *Ich*, which James Strachey translated as “ego” in *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Freud used the term to connote both the person and the intrapsychic agency within the person. In other words, a term that can be literally translated from the German as “I” was used interchangeably to refer both to subjective self-experience and to a collection of organizing functions within the psyche. However, the self-experience meaning gradually became lost in the increasingly preferred usage of “ego” as an impersonal executive organ of the psyche.

Hartmann attempted to distinguish the two meanings of *Ich* by defining the interactional context. The self interacts with objects, and the ego interacts with the id and the superego. That clarification was useful because it stresses the fact that the self evolves in relational configurations to objects. Hence, one's experience of oneself is discontinuous in that the nature of the self varies with the nature of the object to which it is relating. In that regard the person's self embodies the history of many internalized relationships. The concept of the self should then refer to a multiplicity of points of view from which one experiences, observes, and feels. Everyone behaves differently from others and has a different sense of internal experience, depending on the figure with whom the person is relating.

Despite the presence of that array of selves responding to diverse environmental contexts, a person has some sense of continuity over time. The sum of the component selves is regarded as one's *identity*, a term defined by Erik Erikson as a conscious sense of inner solidarity that endures over time.

If one considers a sentence such as, “I think about myself,” one becomes aware that the self is used to describe both an object of reflection and an experiential consciousness that is doing the thinking. That distinction has been the source of another major controversy surrounding the self—namely, whether the word “self” should be used to refer to an intrapsychic representation, conscious or unconscious, or to an initiator of action in its own right. Ego-psychological thinkers have traditionally viewed the self as representational, but members of the British school have preferred a concept that includes subjective experience or personal agency. Kernberg has suggested that there is room for both the self-as-agency view and the self-as-representation perspective. The self-as-agency is embedded in the ego and may be conceptualized as the end result of the integration of a myriad of self-representation.

Contemporary psychoanalytic theory has been enhanced by conceptualizing the self as a content of the ego. The ego is useful for understanding and dealing with drive-defense conflicts, character, and compromise formations, but the self offers a subjective side to the ego that is relevant to such issues as narcissism, identity, self-other differentiation, and internal object relations. Moreover, the notion of self is important in clinical work with primitively organized patients who use such defenses as splitting and projective identification. Kohut's view of the self as a primary, supraordinate entity implying motivational status is discussed under the heading “Self Psychology.”

Infant Observation Studies Object relations theory has been bolstered by the findings of infant observation studies conducted by Margaret Mahler (Fig. 6.1-12) and her colleagues. Those investigations focused on both normal and abnormal mother-infant pairs during the first 3 years of life. The investigations were specifically designed to study the manner in which a child achieves an intrapsychic sense of separateness from its mother. As a result of the research, three developmental phases in mother-infant object relations have been described.

FIGURE 6.1-12 Margaret S. Mahler, M.D. (Courtesy of Margaret S. Mahler, M.D.)

An *autistic* phase occurs during the first 2 months of life, during which the infant spends most of its day in a half-sleeping, half-waking state, more concerned about physiological matters than relatedness. From 2 to 6 months of age, the infant enters the phase of *symbiosis*, a term Mahler used to describe an undifferentiated state of fusion between infant and mother. The onset of that phase is heralded by the infant's smile response and the baby's dawning awareness of a dimly perceived source of need satisfaction in its environment, although that maternal source is viewed as within the orbit of an omnipotent dual unity.

Most of Mahler's research focused on the third phase of the development of object relations, which she termed *separation-individuation*. That period extends from around 6 months to 36 months of age and involves the infant's psychological birth as a separate person apart from the mother. Observational data indicated that the third phase can be divided into four subphases. The first subphase, *differentiation*, occurs between 6 and 10 months of age and is linked to the child's awareness that the mother is a separate person. That awareness often results in the child's attachment to a transitional object, as described by Winnicott, to substitute for the mother when she is not available.

The second subphase, which Mahler termed *practicing*, occurs between the ages of 10 and 16 months and is characterized by elated investment in the child's newfound autonomy that results from acquiring locomotor skills. Toddlers explore their environment as though their mothers were no longer important, only to return for periodic refueling through physical contact.

The third subphase of separation-individuation, known as *rapprochement*, occurs between 16 and 24 months of age. In contrast to the toddler's relative obliviousness of the mother's presence during the previous subphase, in rapprochement the toddler shows a marked increase in its awareness of vulnerability to separations from its mother. Children at that age demonstrate a seemingly constant concern about the actual location of their mothers and display a wish for their mothers to be involved in sharing new experiences. A great need for maternal love can also be observed during the third subphase. Rapprochement is crucial in the development of object relations in which the child uses splitting of self-representations and object representations. The toddler's anger at its mother is viewed as dangerous because of the risk that the anger will destroy the positive internal representation of the mother, and the absence of the child's mother during the third subphase leads to a predominance of bad self-representations and object representations, which may be quite frightening to the child.

The fourth subphase of separation-individuation is marked by the beginnings of object constancy and the consolidation of the child as a separate person. The period corresponds to the third year of the child's life and involves the integration of the good and bad aspects of both self-representations and object representations. Moreover, the major accomplishment of the fourth subphase is the development of *object constancy*, by which the mother becomes introjected as an integrative whole object that is felt as a soothing internal presence. Because an internal image of the mother is sustained during periods of separation, the child no longer feels threatened by the comings and goings of the mother.

The empirical data gathered by Mahler and her colleagues demonstrate the crucial importance of object-seeking in the course of development. Clinicians have shown that the concepts of separation-individuation, object constancy, and, particularly, the rapprochement subphase have considerable applicability in the understanding of primitively organized patients, such as those with borderline personality disorders. The implications of that research on the understanding of pathogenesis is discussed under the heading “Psychoanalytic Psychopathology.”

SELF PSYCHOLOGY

The third major theoretical school in modern psychoanalysis is self psychology. (The other two schools are ego psychology and object relations theory, discussed above.) Derived from the work of Heinz Kohut (Fig. 6.1-13), the self psychology model of the mind regards the person as needing particular kinds of responses from others in the environment to develop and maintain a sense of self-esteem and well-being. Although object relations theory and self psychology have some similarities,

object relations theory stresses the importance of internalized relationships between self-representations and object representations, and self psychology focuses on the role of actual external relationships in creating self-cohesion and self-esteem.

FIGURE 6.1-13 Heinz Kohut. (Courtesy of AP/Wide World Photos.)

Kohut's self psychology departs from classical ego psychology in a number of ways. Defects or deficits, rather than conflicts, take center stage in self psychology. Faulty structures are viewed as responsible for faulty functioning, and the emphasis is on infantile needs, rather than repressed wishes and drives. Hence, the analyst's therapeutic goals involve understanding those needs and partially meeting them in the treatment, rather than frustrating infantile wishes that must ultimately be renounced. Building the psychic structure and repairing self-defects are seen as more important than the resolution of conflict.

Kohut's Double-Axis Theory Self psychology evolved out of psychoanalytic work with narcissistically disturbed patients. In treating those patients, Kohut noted that ego psychology based on the structural model lacks sufficient explanatory power with those patients. Instead of having discrete neurotic symptoms, those narcissistic patients had vague complaints related to disappointing patterns in relationships and a hypersensitivity to slights from others.

Kohut thought that a new theory was required to provide a conceptual framework for the analytic treatment of such patients. His theory evolved directly from clinical observations, specifically the tendency of his patients to form one of two kinds of transference: the mirror transference and the idealizing transference. The mirror-transference patients seem in desperate need of the analyst's approval and will do whatever is necessary to gain that affirmation or validation from the analyst. Those patients often appear to be performing or showing off to obtain the analyst's admiration. Kohut viewed mirror transference as a revival of an infantile situation in which the child shows off to capture the gleam in the mother's eye that makes the child feel confirmed and validated. Kohut referred to that developmental arrest as the *grandiose-exhibitionistic self*, a term he used nonpejoratively to describe a normal developmental phase in which the child's self-worth is dependent on empathic mirroring responses from the mothering figure. Without that empathy, children cannot maintain their sense of self-cohesion or wholeness, and their sense of self fragments. That sense of fragmentation of the self can be experienced along a continuum from mild anxiety or distress to full-blown panic associated with feelings of disintegration. Those children who are not provided with adequate maternal empathy are regarded as developmentally frozen and doomed to go through life seeking mirroring responses from others.

In the idealizing transference the patient regards the analyst as a perfect and omniscient parent who meets all the patient's needs. The patient may show little interest in insight or understanding in such cases because the patient's primary wish is to bask in the reflected glory of the idealized analyst. Kohut viewed that transference manifestation as similar to the developmental arrest in which the child has not been provided with a parental model worthy of idealization. As a result, the analyst is needed to perform the function that was missing in childhood.

Kohut's observations led him to postulate a double-axis theory that takes into account both narcissistic needs and object love. Kohut was impressed that narcissistic needs are never outgrown but, rather, persist throughout life in parallel development with the person's needs for object love. On the basis of that perspective, he postulated a separate axis for narcissistic development that exists alongside the classical line of development leading to object love, as depicted in Figure 6.1-14. In that model the self begins as a set of fragments that gradually achieve cohesiveness if the child's phase-appropriate developmental needs are greeted with maternal empathy.

FIGURE 6.1-14 Kohut's double axis theory (1971). (Reprinted with permission from Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice*. American Psychiatric Press, Washington, DC, 1990.)

At some point in development, a point that Kohut left unspecified, children are threatened by losing the blissful perfection of the mother-infant bond and resort to one of two strategies to recapture the perfection. In one scenario the perfection is assigned to a grandiose self within the infant, in which case the adult manifestation in the psychoanalytic setting is the mirror transference. In the other scenario the infant assigns perfection to an idealized parental imago, in which case the idealizing transference appears in the analytic setting. In the course of normal development, the grandiose self evolves into healthy ambitions, and the idealized parent imago is eventually internalized as a structure akin to Freud's ego-ideal or superego—that is, a set of values and ideals that lead to moral conduct. If, however, the child's developmental needs for mirroring or idealization or both are met with maternal failures of empathy, a developmental arrest occurs at the point of the grandiose self or the idealized parental image. Those two poles of self-development Kohut termed the *bipolar self*.

Inherent in Kohut's double-axis theory is an acceptance of narcissistic needs. He thought that Freud's model of the progression from primary narcissism to object love as an expectation in the normal maturational process contains a moralizing, pejorative tone toward narcissism. In other words, Freud's view suggests that narcissism should be outgrown and that the mature adult should be primarily devoted to the needs and concerns of others. Kohut regarded that attitude as hypocritical; after all, everyone has narcissistic needs that are crucial to one's sense of personal happiness and fulfillment. Indeed, one of Kohut's major contributions was his insistence that, in addition to the dual drives of sexuality and aggression, needs for self-esteem also occupy a place of central importance in the psyche. Self psychology, then, provides a conceptual framework for the analyst that advocates empathy for the patient's narcissistic needs, instead of viewing them as immature and contemptible. Another implication of the double-axis model is that the development of a cohesive sense of self may in and of itself be a legitimate goal for psychoanalytic treatment, regardless of the patient's capacity for object love.

Selfobject Concept Throughout the evolution of self psychology, Kohut struggled with the extent to which his new theory would depart radically from the classical model in which he had been trained. He initially tried to incorporate drive theory by suggesting that narcissistically disturbed patients invest objects with a special form of narcissistic libido. However, as his theory evolved, he eventually found that mixed models were unsatisfying to him, so he made a complete break from classical theory. He also thought that his theoretical formulations were no longer limited to narcissistic personality disorders. He greatly expanded the scope of self psychology

to include all forms of psychopathology.

Within that broadened view, the term “selfobject” became a central tenet of his theory. *Selfobjects* are viewed as fundamental needs of all persons that are required for normal development. Selfobjects may best be conceptualized as functions (such as mirroring, validating, soothing, idealizing, and affirming), rather than actual persons. As the term “selfobject” implies, other people are not viewed as separate objects with their own centers of autonomy and their own distinct needs. Rather, they are viewed as present only to gratify the needs of the nascent self. Kohut believed that selfobjects are necessary for emotional survival in the same way that oxygen in the atmosphere is needed for physical survival. Throughout life all persons are dependent on others to maintain their sense of self-esteem and well-being.

At the time of his death in 1981, Kohut had conceptualized a third area of selfobject needs, which he termed the *twinship* or *alter ego*. Originating in a childhood wish to merge with the mothering figure, this dimension of selfobject needs makes its appearance in the transference in the form of a wish or need to be exactly like the analyst. Infantile antecedents of that transference development can be found in such commonplace occurrences as a little girl's pretending to feed her doll as her mother is busy nursing a younger sibling. Hence, Kohut's bipolar self became a *tripolar self*.

As a result of this expansion of the theory, Kohut observed that all forms of psychopathology can be traced to disturbances of the self-selfobject relationship in childhood. The Oedipus complex is regarded as secondary in importance in the light of the self-psychological view that all psychiatric disorders are based on distortions of the self, weakness of the self, or defects in the structure of the self. Oedipal conflicts involving castration anxiety, inhibition of aggression, and other neurotic symptoms are regarded as breakdown products of preoedipal failures in the provision of selfobject needs. An adequate self-selfobject matrix during the early years of life should allow the developing child to weather the oedipal constellation without residuals of neurotic symptoms. Moreover, most forms of pathological behavior—including perversions, binge eating, sexual promiscuity, and drug abuse—are regarded as efforts to restore or maintain a cohesive sense of self under the threat of fragmentation. In that conceptual model, disintegration anxiety, involving the concern that the failure of adequate selfobject responses from persons in the environment will lead to fragmentation of the self, becomes the fundamental fear leading to emergency attempts to restore harmony to the self through pathological behaviors.

Despite the central importance of the self in Kohut's theory, Kohut himself never offered a simple definition of the self. He seemed to think that an overarching structure of such fundamental importance defied definition. If one follows the development of his theory from the late 1960s to the time of his death in 1981, however, his view of the self appears to have undergone a transformation from that of a self-representation to a supraordinate center of initiative and experience that was regarded as the main motivating agency in the primary constellation of the psyche. In that regard the self in Kohut's self psychology is both an experiential self and a center of initiative that establishes and repairs self-esteem. Hence, even self psychology cannot escape the dialectic of self as simultaneously subject and object. That theoretical perspective can also be characterized by a marked deemphasis on the ego and its functions and a corresponding increase in the importance of conscious subjective experience. Also, in the light of the minimization of the role of drives and defenses, aggression is viewed as a secondary reaction to selfobject failures, rather than as an instinctual drive of a primary or innate nature.

Although Kohut was always vague about the developmental timetable appropriate to his theory, certain implications can be drawn. First, Kohut believed that a psychological connection to a mothering figure is present from birth, an idea in stark contrast to Mahler's autistic phase. The infant, in Kohut's view, is born into a self-object bond that is much closer to Mahler's concept of the symbiotic phase. Second, Kohut's emphasis on the lifelong need for appropriate selfobjects suggests that psychological separation is a myth. He regarded Mahler's object relations view, which revolves around the notion that the child moves in the direction of autonomy from supportive caretakers, as containing a moralistic stance. Kohut's notion of lifelong dependence on selfobjects places him in opposition to Mahler's developmental framework. His view was that dependency itself does not diminish; instead, the quality of the selfobjects needed by the self moves from archaic to more mature and appropriate.

Infant Observation Data Infant research, particularly that of Daniel Stern, has provided considerable support for the developmental premises of self psychology. Stern's observations convinced him that the infant does not emerge from the womb into a state of autistic self-absorption. From the first days of life, the infant appears to be aware of the mothering figure. Moreover, affirmation and validation from the caretaker is crucial for the development of the infant's sense of self. In that regard Stern stressed that what develops is a sense of self-with-other, and the role of fantasy, as in Klein's theory, is viewed as of only minimal significance. The infant primarily experiences reality, in Stern's view, and relates to the real presence of the mother, whose responses allow the infant to grow. Stern believed that the infant is an adept observer of reality and that it is the older toddler who uses fantasy and distortion to alter what is perceived. He argued that secondary process thought precedes primary process thought and is the bedrock of normal development.

Stern delineated five discrete senses of self. He regarded them as different domains of self-experience, rather than as phases that are subsequently eclipsed by the succeeding phase. Each domain, once formed, remains for the entire life span and operates in concert with the other coexisting senses of self. During the first 2 months of life, an emergent self appears; it is predominantly a bodily self based on physiological needs. A core sense of self emerges from 2 to 6 months of age; it is associated with greater interpersonal relatedness. Stern viewed the third sense of self, appearing between 7 and 9 months, as a major advance. He called it the sense of subjective self because it involves the matching of intrapsychic states between mother and infant. The fourth sense of self, the verbal or categorical sense of self, emerges between 15 and 18 months of age and coincides with the ability to think symbolically and to communicate verbally. The arrival of a fifth sense of self, a narrative sense of self, appears between 3 and 5 years of age. Stern believed that the fifth sense of self, the historical view of the self, is the one encountered when patients present their life stories in the analytic setting.

Stern identified four essential features of the sense of self: agency, coherence, affectivity, and continuity (or historicity). He also noted that such fundamental issues as attachment, trust, and security transcend phase specificity and are issues for the entire life span. He also questioned the notion of symbiotic merger postulated by Mahler as occurring in the first 6 months of life. In Stern's view, such wishes for and fears of union are the product of a much more mature mind, one that is capable of thinking abstractly about such concepts as fusion and merger. A central tenet of Stern's work is that the infant develops as a result of sensitive affective attunement by the mothering figure. Kohut and other self psychologists argue that the same kind of empathic resonance is necessary in the analytic situation to repair the patient's self-defects.

Post-Kohut Contributions After Kohut's death in 1981 self psychology has continued to expand its vision. The three selfobject transferences of Kohut have come to be regarded as insufficient to account for all the transference phenomena observed in the clinical setting. Ernest Wolf, a colleague and collaborator of Kohut, has described two additional selfobject transferences. The *adversarial* selfobject transference functions so as to provide an experience with a benevolent, but opposing force while also allowing confirmation of partial autonomy for the self through active opposition. The *efficacy* selfobject transference provides an experience of having an effective impact on another individual, in that it evokes necessary selfobject functions from that individual.

Robert Stolorow and his colleagues have taken Kohut's emphasis on empathy and suggested that psychoanalytic work is confined to aspects of the patient's experience that are accessible to empathy and introspection. They emphasize the nature of the intersubjective field created by analyst and patient and view extratransference information as of minor importance to the analytic process. In this regard, they are inclined to regard psychopathology as involving a two-person interplay. Analysts who fail to meet the specific selfobject needs of the patient may significantly influence the psychopathology they perceive in the patient. Other analysts influenced by self psychology believe that information outside the empathic-introspective mode of perception must be integrated into the analyst's knowledge base. Joseph Lichtenberg regards knowledge of “model scenes” prototypical of childhood and infant experiences as extremely relevant to understanding and reconstructing the patient's early experience. He also argues that five distinct motivational symptoms must be taken into account to fully understand forces at work in the patient. These five systems include exploratory assertiveness, sensual-sexuality, aversiveness, attachment-affiliation, and most psychological aspects of physiological pressures. Lichtenberg regards mainstream self psychology as reductive in its relegation of sexuality and nonsexual pleasure to a relatively peripheral position.

In other post-Kohut developments, Howard Bacal and Kenneth Newman have sought to integrate self psychology with object relations theory. They note that although Kohut failed to acknowledge being influenced by the British School of object relations, he has much in common with their thinking. They also stress that in its applications to the clinical setting, self psychology is basically a variant of object relations theory. Whereas Kohut argued that the self alone was the center of his theory, Bacal and Newman point out that the self in connection with its object is the true basic unit of self psychology. A sophisticated understanding of the selfobject was not articulated by Kohut, so these authors have sought to fill that void by regarding the relationship between patient and analyst as a true object relationship.

POSTMODERN TRENDS

In recent years the hegemony of ego psychology in American psychoanalysis has been challenged by a number of postmodern trends in psychoanalytic thought. The unifying factors in postmodernism include the view that reality inevitably grows out of social and linguistic construction and a basic skepticism about any

unquestionable or fundamental truths. Freud is often regarded as representing modernity, characterized by a positivist orientation and a dispassionate search for an absolute or objective truth in the patient that will ultimately be revealed if the patient adheres to the method of free association. By contrast, postmodern schools of thought, such as social constructivism, perspectivism, intersubjectivity, relativism, and mutuality argue that no absolute objectivity exists and that perceptions in the analysand are inevitably influenced by the analyst's own subjectivity.

Of all the forms of postmodernism, the most extreme variation would be that of relativism. According to this school of thought, truth about the patient is impossible, and both analyst and analysand are mired in uncertainty. A more moderate view that is endorsed by a number of contemporary psychoanalytic thinkers is that of realism and perspectivism. The realist-perspectivist school differs from relativism by acknowledging the existence of a clinically useful and significant reality in the patient that requires articulation and understanding in the course of the treatment. Nevertheless, this theoretical orientation stresses that each participant in the analytic dyad has a subjectivity that ultimately results in a jointly constructed reality. In this two-person frame of reference, both analyst and analysand are regarded as having plausible perceptions of the emerging reality.

Social Constructivism The social constructivist perspective of Merton Gill and his collaborator Irwin Hoffman has been highly influential in the evolution of recent psychoanalytic thinking in America. Central to social constructivism is the notion that the analyst's *real* behavior continuously affects the patient's transference. The analyst is no longer regarded as a dispassionate observer of the intrapsychic functioning of the patient but rather a mutual and full-fledged partner in the exploration of what transpires between the two people involved in the analysis. Moreover, a central tenet of the social constructivist view is that analysts can never know the full extent of the way that their irreducible subjectivity influences patients and the analytic process.

Countertransference is viewed as continuous and unconscious so that the analyst is incapable of keeping up with unconscious emotional reactions to the patient. Another basic tenet of this perspective is that the patient may see something that the analyst is resisting in the same way that the analyst observes resistance in the patient. Hence the analyst needs to take seriously the patient's interpretations of the analyst's conscious and unconscious influences on the patient. The social constructivist view emphasizes both spontaneity and personal engagement with the patient that fully acknowledges the contributions of the analyst to the transference.

Postmodern trends in psychoanalytic thought often lead to a deconstruction of the analyst's authority, but social constructivists like Gill and Hoffman have argued that the authority of the analyst is to some extent inescapable by virtue of the asymmetry of the analytic situation. The analyst always has the benefit of knowing more of what is going on in the patient's mind than the patient knows about what is going on in the analyst's mind. In addition, by virtue of their training analysis, analysts should be more aware of biases and more capable of reflecting on those biases.

American Relational Theory Of all the postmodern schools, American relational theory has probably had the most profound impact on contemporary psychoanalytic discourse. In 1983 Jay Greenberg and Stephen Mitchell identified two distinct approaches to psychoanalytic theory: the drive-structure model and the relational model. They argue that these two perspectives are essentially incompatible because drive theory stems from a philosophical tradition that regards man as an essentially isolated individual whereas relational theory derives from a view of man as a social being. Conflict occurs in a relational context in this theory rather than between intrapsychic agencies. The similarities to the British Middle Group have been acknowledged by the fact that relational theory is sometimes regarded as the “American Middle School.”

Relational theorists acknowledge the intrapsychic but view it as the internalization of interpersonal experience that is also mediated by biologically based constraints. Physiological pressures, such as sexuality, are not ignored, but are recognized as occurring in a relational context.

Central to the relational view is the notion of mutuality (i.e., the inevitability of a two-way influence between patient and analyst). Meaning is not construed simply by the analyst's clinical acumen but through a dialogue between patient and analyst involving negotiation and joint construction. Relational theorist Lewis Aron has stressed that mutuality does *not* imply symmetry or equality. Despite the mutuality and reciprocity of the relationship, analysis is a relatively asymmetrical relationship where one person seeks help from another and pays that professional a service that is rendered under ethical guidelines.

Feminism has played an important role in the development of relational theory. Classical psychoanalysis has been viewed by feminist critics as embodying masculine ideals of objective science and detachment from the object of study. Feminist views that attachment and relatedness are as important as independence and autonomy have contributed to a shift towards a more relational perspective within analysis. Jessica Benjamin has pointed out that in most developmental theories within psychoanalysis, the mother of the infant is viewed only as an object rather than a subject. She is seen as provider, caregiver, source of nurturance, and empathic selfobject but rarely viewed as someone with a subjectivity that involves a purpose apart from taking care of the child. Benjamin has stressed that intersubjectivity is a developmental achievement involving the recognition of the mother (or other person) as a separate subject rather than merely as an object. Much like Winnicott, Benjamin emphasizes that in the course of psychoanalytic work, the analyst becomes less of an object and more of a subject to the patient, paralleling the nature of subjectivity and objectivity in the course of child development.

Lacanian Theory A discussion of postmodernism would not be complete without a survey of the theory of the French psychoanalyst Jacques Lacan (1901–1981). Although abstruse, elusive, deliberately provocative, and unorthodox, Lacan has made a lasting impression on French psychoanalysis as well as on literary and film criticism in academic departments throughout the world. Lacan's reading of Freud relies heavily on linguistics. The notion that the human being is constituted by language is one of three basic principles endorsed by Lacan. The unconscious is structured like a language that consists only of signifiers—biology and drives have no place in his theory. Second, the ego does not exist as an autonomous structure. The third principle is that an individual is inevitably embedded in political and societal structure that cannot be transcended.

Lacan preferred to speak of “orders” rather than stages of development, and he identified three orders relevant to understanding the human condition. From 6 to 18 months there is an imaginary order involving a mirror stage, during which infants delight in identifying with their reflection. This preverbal and presymbolic order gives way to a symbolic order involving the acquisition of language and a split between the inner and the outer world. The inner world is connected with a misrecognition in the mirror stage where the “I” or the self is identified with the father's laws and the cultural standards associated with those laws. The final order is that of the “real,” which is stark factual reality and exists somewhere outside of language; this is viewed as unreachable and indefinable by Lacan.

Lacan thrived on being unorthodox. He denied the significance of diagnoses, rules, or established schools of thought. He saw the analytic process as an effort to recognize the alienation from one's true self. Analysis was also designed to bring out underlying structures and contexts in the unconscious. Historical reconstruction is downplayed whereas desire for the “Other” is examined in all its many forms.

Lacanian analysis has had little influence on the clinical practice of psychoanalysis in North America, although literary critics within American universities have made many applications of Lacan in their analyses of texts. Its antiempirical basis, its denial of the importance of preverbal human experience, and the impenetrability of much of the prose written by Lacan and Lacanian disciples has made it less popular among American clinicians than those in Europe and South America.

EMPIRICAL BASIS OF PSYCHOANALYSIS

Freud and psychoanalysis are regularly pronounced dead in articles appearing in popular periodicals. In many cases these funeral notices are accompanied by challenges to the basic scientific premises of Freud's theory. Many of the attacks on Freud have centered on his early views—formulated in the 1890s—about the importance of uncovering repressed memories in the course of analysis to accomplish the therapeutic goals of the treatment. Few of the attacks have demonstrated any significant degree of understanding about the contemporary practice of psychoanalysis and how it has evolved over the last 100 years.

Clinical psychoanalysis has progressed far beyond the theory of therapeutic action advanced by Freud in the 1890s. The de-repression of long-forgotten memories shifted first to an emphasis on recovering memories of fantasies. The next major shift was towards an emphasis on exploring resistances to free association, which eventually gave way to a direct examination of transference constellations. The field has now moved into a systematic examination of both transference and countertransference because they are viewed as critical aspects of the two-person field created when an analyst and an analysand meet together in the consulting room. Critics of psychoanalysis have failed to keep up with these changes in the field so that much of the criticism is leveled at specific historical models of treatment that are no longer applicable in contemporary practice.

Many psychoanalytic critics claim that there is no empirical validation of basic psychoanalytic principles or of the psychoanalytic treatment method. Many of those critics are completely unaware of a substantial body of psychological research that has confirmed the fundamental validity of the basic psychoanalytic principles. The most central of all psychoanalytic propositions, that much of mental life is unconscious, is the most challenged by critics of psychoanalysis but ironically is best

validated by extensive literature from experimental psychology.

Subliminal presentation of stimuli that are psychodynamically meaningful to subjects can affect a wide range of behavior even though the subjects do not have any conscious awareness of the stimuli. Individuals with no capacity to recognize faces, a condition called *prosopagnosia*, can respond emotionally to someone standing in front of them about whom they consciously know nothing. Similarly, they may have differential electrophysiological responses to familiar versus unfamiliar faces. Subliminal stimuli have been studied in connection with brain evoked-related potentials, and this research has demonstrated differences in amplitude of these potentials in response to emotional versus neutral words when presented subliminally.

Recent research demonstrates that unconscious affective signals may well factor into decisionmaking before conscious thought enters in. In work by Damasio and his colleagues, the ventromedial prefrontal cortex appears to be part of a system that stores information about past punishments and rewards; signals from this area of the brain trigger the unconscious emotional responses that are registered as a “gut feeling” or intuition.

These conclusions were based on a study involving two decks of cards—one, a “good” deck, produced greater monetary rewards than the other, a “bad” deck. Normal subjects showed physiological responses (measured by skin conductance) before choosing cards from the “bad” deck. This autonomic response occurred *before* they could verbally describe their understanding of why it was better to pick from the “good” deck. A comparison group of patients with lesions in the ventromedial prefrontal cortex did not have the same physiological responses, nor did they have the “hunch” or “gut feeling” manifested by the normal subjects.

Freud's views on the role of unconscious memories have also been confirmed by experimental psychologists. His conclusions, which are often misrepresented by critics, include the following: (1) unconscious memories can indeed be recovered; (2) some recollections are true whereas others are false; (3) even memories that are essentially accurate may be distorted in various ways.

Unconscious transference processes have also been documented experimentally. In one study subjects were asked to provide a description of a significant other in their lives. A component of that description was subsequently incorporated into a depiction of a fictional character. When the subjects were later asked to recall everything they could about the fictional character, they embellished their description with other elements drawn from the characteristics of their significant other.

Although the accumulated research from the experimental psychology literature confirms that Freud was correct in his central tenet regarding the influence of unconscious thoughts, feelings, and motives on a behavior, the evidence for the efficacy of psychoanalytic treatment is more problematic. Uncontrolled studies of persons in analysis suggest that 60 to 90 percent are substantially improved after treatment. However, no randomized controlled study of analysis has yet been conducted so this evidence cannot be considered definitive. Without a control group, investigators cannot be certain if the patients would have improved to the same extent with the passage of time.

On the other hand, there are formidable methodological obstacles entailed in designing a randomized, controlled study of psychoanalysis. First, patients are not randomly assigned to analysis as a treatment. Patients who do well in analysis typically seek it out because of high motivation to understand themselves. Second, the appropriate control group for a process that takes 5 years or more is a challenge for the research design. How many patients are willing to be in a no-treatment control group for such an extended period of time? Third, if 10 percent of patients in a 16-week randomized, controlled trial of brief therapy drop out of the project, there is no substantial loss to the project because 90 percent have continued. However, if in a 5-year study of psychoanalysis, 10 percent of the patients drop out every 4 months, the sample size would be depleted to the point where the statistical power would be severely compromised in any comparisons performed by the researchers. Intercurrent life events, such as deaths of significant others, divorce, and serious illness, are much more likely to become crucial factors influencing outcome in a 5-year study than in a brief therapy project confined to 16 weeks of treatment.

An overriding problem is that a randomized, controlled study would fail to capture the essence of psychoanalytic treatment. Random assignment bypasses the fact that patient and analyst find each other through a poorly understood process of “patient-analyst matching.” For some reason, both members of the dyad feel that they can work with each other. Also, good analysis is characterized by flexibility, and a treatment manual has the potential to restrict the analyst so that the treatment being studied is different from the treatment occurring in a naturalistic setting.

Nevertheless, there are numerous studies, some using randomized, controlled design, demonstrating that psychotherapies based on psychoanalytic principles are efficacious.

PSYCHOANALYTIC PSYCHOPATHOLOGY

In an age in which neuroscience research has made giant strides in the understanding of many Axis I disorders, psychoanalytic theory still has much to contribute to the overall understanding of causes and pathogenesis. Although biological factors play major roles in such illnesses as schizophrenia and mood disorders, environmental and psychological stressors are influential in triggering the underlying biological diathesis. Moreover, the clinical management of Axis I disorders is greatly enhanced by a psychoanalytically based understanding of the personality factors that contribute to the illness. In the following survey of the psychoanalytic contributions to the understanding of psychiatric disorders, the emphasis is on psychological and environmental factors. However, whenever appropriate, the interaction between those factors and biological determinants of the illness is pointed out.

Schizophrenia Freud thought that patients with psychotic disorders are not amenable to psychoanalytic treatment, so he did not accumulate the kind of clinical experience with psychoses that he did with neuroses. Nevertheless, he attempted to place psychotic disorders in a psychodynamic conceptual framework. Freud thought that the sine qua non of schizophrenia is the withdrawal of object cathexis, by which he meant that schizophrenia patients detach themselves from any libidinal or emotional investment in external objects or their intrapsychic representations. He regarded schizophrenia as a regression from object love to an autoerotic stage of development in which one does not have to deal with the frustrations and conflicts encountered in interpersonal relationships.

Freud's decathexis theory underwent revision after the structural model had been developed. In his revised view of schizophrenia, he contrasted neurosis and psychosis on the basis of the nature of the conflict involved. He regarded neurosis as a conflict between the ego and the id, but in psychosis the conflict occurs between the ego and the external world; reality is remodeled to conform to the patient's internal distortions. Even after that substantial revision, Freud continued to write about withdrawal of object cathexis and used that concept to assert that patients suffering from schizophrenia cannot form transference attachments to treaters.

Subsequent psychoanalytic writers have demonstrated that schizophrenia patients do develop transference, although it is qualitatively different from the variety seen in the psychoanalytic treatment of neurotic patients. Harry Stack Sullivan and Frieda Fromm-Reichmann, who spent their professional careers involved in the psychoanalytic treatment of psychotic patients, regarded schizophrenia as arising from early interpersonal difficulties in the child-parent relationship that required understanding and repair in the context of a therapeutic relationship. In their view, faulty mothering led to anxiety and distrust of others, causing persons with schizophrenia to withdraw from interpersonal relatedness. This parent-blaming view has been discredited by recent research.

One controversy in the psychoanalytic literature involves whether or not the conflict theory used to explain neuroses is sufficient to explain schizophrenia as well. Jacob Arlow and Charles Brenner believed that psychosis can be explained by the same compromise-formation model that is applied to neuroses and that the two conditions may differ quantitatively, but qualitatively they are fundamentally alike. Arlow and Brenner conceptualized the quantitative differences as schizophrenia's involving greater disturbances of ego and superego functioning, more severe regression, and more intense conflicts revolving around aggression. They regarded schizophrenic symptoms, such as delusions and hallucinations, as compromise formations developed in response to conflict, much in the same way that neurotic symptoms are developed. Similarly, Arlow and Brenner viewed the withdrawal of object cathexis that Freud emphasized as simply a defensive retreat from conflict.

Biological research on schizophrenia has identified primary difficulties with information processing and attention in schizophrenia; both difficulties are improved by antipsychotic medication. Indeed, most schizophrenia patients experience themselves as being bombarded by a variety of external and internal stimuli. Approaching the illness from a deficit model, James Grotstein suggested that the schizophrenic inability to screen out incoming stimuli is the fundamental problem in the illness of schizophrenia. The defective stimulus barrier makes it extremely difficult for patients with schizophrenia to deal with the chaotic feelings within, so unacceptable impulses and affects are projectively disavowed and placed in the mothering figure.

Mood Disorders Biological factors play major roles in the causes and pathogenesis of mood disorders. A vast literature on genetic transmission and neurotransmitter changes attests to that fact. Nevertheless, research suggests that intrapsychic and environmental stressors are involved in the onset of major depressive episodes in the majority of cases. It is useful in the light of that perspective to separate cause from pathogenesis. Environmental factors, such as the loss of a loved one, and intrapsychic factors, such as falling short of internal standards, may be crucial causative determinants of a particular depressive episode. The ensuing response to

those environmental and intrapsychic triggers, however, occurs at the level of brain chemistry. Hence, the stressor interacts with a genetic vulnerability.

Depression Freud originally understood depression as internally directed anger. In his view the self-reproaches and the loss of self-esteem commonly experienced by depressed patients are directed not at the self but, rather, at an introject. He noted that in some cases the only way the ego can give up an object is to introject it, so the anger directed at the ambivalently held object takes on the clinical manifestations of depression. After his development of the structural model, he expanded his understanding of depression to include a harsh superego that punishes the person for harboring destructive wishes toward parental figures and other loved ones.

Melanie Klein suggested that depression is linked to a reactivation of the depressive position; depressed patients are convinced that they have destroyed their internal good objects because of their own aggression and greed. As a result, they feel persecuted by internal bad objects while longing for the lost love objects.

Contemporary psychoanalytic contributors have downplayed the role of aggression in the development of depression. They are likely to view depression as a disturbance of self-esteem in the context of interpersonal relationships. A consistent observation is that depressed patients feel that they have not lived up to their internal standards of conduct. The depressed patient's awareness of the disparity between their actual performance and those high internalized expectations leads them to feel helpless and powerless. Often, their internal expectations involve eliciting a certain kind of response from important persons in the environment. Depressed persons often live their entire lives for others, rather than for themselves. Depression may begin when they feel hopeless about their life plans because they realize that their efforts have been wasted in living for someone else.

From an object relations perspective, many depressed patients unconsciously experience themselves to be at the mercy of a tormenting internal object that is unrelenting in its persecution of them. In cases of psychosis, that primitive forerunner of the superego may actually be hallucinated as a voice that is unrelentingly critical. From the self-psychological point of view, depression is related to a sense of despair about ever getting one's selfobject needs met by people in the environment.

Mania Empirical research has demonstrated that genetic and biological factors play a key role in bipolar I disorder. Nevertheless, psychodynamic issues are highly relevant to the overall treatment. Psychological and environmental stressors are often involved in precipitating a manic episode, and psychotherapeutic interventions may help to identify and reduce those stressors to increase the chances that the patient will remain euthymic. Also, the psychodynamic conflicts of the patient may use the manic episodes in the service of expressing specific aspects of those conflicts. For example, manic episodes often serve a defensive function so that the patient does not get in touch with the painful affects associated with the undercurrent of depression. The variant of dysphoric mania, when depressive feelings break through the manic episodes, reflects this defensive function of mania. Melanie Klein noted that manic defenses have several functions: to deny that one is dependent on love objects, to disavow the presence of internal objects, and to restore and rescue lost love objects. Manic episodes frequently occur in response to significant losses as a way of denying the grief associated with the losses.

There are also a number of psychodynamic themes that are consistently problematic in the treatment of patients with bipolar I disorder. Denial and lack of insight into the illness is a regular accompaniment of the disorder, and generally persists even when the patient is euthymic. As a result, compliance with medication is a major problem in the management of these patients. The denial is also accompanied by a form of intrapsychic discontinuity manifested in splitting of the self-representation. In other words, euthymic patients will often disavow any connection with the aspect of themselves associated with the manic episode. They may deny that their behavior was part of an illness or completely minimize the extent to which the behavior was problematic. Moreover, since they see no connection between the euthymic self and the manic self, they may not see the need for ongoing treatment. Finally, grief and loss are often part of the aftermath of the destructive behavior connected with the manic episode. The patient may struggle with feelings of guilt and the need to do the work of mourning to assimilate the losses and work through the feeling that he or she has been destructive to loved ones.

Anxiety Disorders and Neuroses Freud's famous psychoneurotic cases are summarized in Table 6.1-4. In brief, his early view included a heavy emphasis on the traumatic consequences of childhood seduction and on the physiological buildup of libido. As Freud shifted to a model based on psychic conflict, he viewed neurotic symptoms as reflecting the breakthrough of infantile sexual impulses in response to failures of repression. With the advent of the structural model, Freud understood neurotic symptoms to be compromise formations between unconscious wishes and defenses against those wishes produced as a result of conflict between intrapsychic agencies. That view was further refined in 1926, when anxiety became understood as a signal of the presence of danger in the unconscious that mobilized ego defense mechanisms to bolster repression. The relabeling of neuroses as anxiety disorders in the third, revised third, and fourth editions of the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM-III, DSM-III-R, and DSM-IV) has led to an unfortunate tendency for clinicians to think about anxiety as an illness to be eliminated with pharmacological intervention, rather than as an overdetermined symptom of unconscious conflict that requires understanding.

	Conversion Reaction (Dora)	Phobic Reaction (Hans)	Obsessive-compulsive Reaction (Rat Man)	Mixed Psychoneurotic Reaction (Wolf Man)
Family history	Striking family history of psychiatric and physical illness	Both parents treated for neurotic conflict but not severe	No family history of mental illness	Striking family history of psychiatric and physical illness
Symptoms	Enuresis and masturbation at 6–8 yrs. Onset of neurosis at 8. Migraine, nervous cough, and hoarseness at 12. Aphonia at 16. "Appendicitis" at 16. Convulsions at 16. Facial neuralgia at 19. Change of personality at 8 from "wild creature" to quiet child	Compulsive questions at 3–2½ yrs. on regard to sex differences. Jealous reaction to sibling birth at 2½. Overt castration threat. Overt masturbation at 3½. Overeating and constipation at 4–5. Phobic reaction at 4–5. Attack of flu at 5 worsens phobia. Tonsillectomy at 5 worsens phobia	Naughty period at 3–4 yrs. Marked timidity after a beating by Father at 4. Recognizing people by their smells as a child ("Bonellius"). Precocious ego development. Onset of obsessive ideas at 6–7	Irritable and quiet up to 2½ yrs. "Naughty" period at 2½–4 yrs. Phobias at 4–5 with nightmares. Obsessional reaction at 6–7 (pious ceremonials). Disappearance of neuroses at 8
Causes	Seduction by older man. Father's illness. Father's affair	Seduction case by mother. Sibling birth at 2½	Seduction by governess at 4. Death of sibling at 4. Beating by Father at 4	Seduction by older sister at 2½. Mother's illness. Conflict between maid and governess

Adapted from Arlow (1).

Table 6.1-4 Classical Psychoneurotic Reactions

Panic Disorder Recent biological research has borne out Freud's early observations that there are two forms of anxiety, one involving a purely psychological cause and the other more clearly related to an underlying neurophysiological substrate. Psychological anxiety is now regarded as a milder form of anticipatory or signal anxiety, and neurophysiological anxiety takes the form of panic disorder. Many patients who experience episodes of panic feel that the attacks come out of the blue and are therefore psychologically contentless. Nevertheless, psychodynamic investigations of those attacks find that many of them have psychological meaning. The experience of them as coming out of the blue may only reflect the fact that the psychological triggers are unconscious. A number of studies have demonstrated a close link between separation anxiety and panic disorder. In addition, as many as 43 percent of all panic-disordered patients show improvement after taking a placebo; that finding suggests that the interpersonal involvement with a physician who is concerned about the symptoms may ameliorate certain psychological factors involved in the onset of panic attacks.

One of the clearest examples of psychologically precipitated panic attacks can be found in patients who have not attained object constancy, as described in Mahler's developmental scheme. Patients with borderline personality disorder often cannot internalize a soothing image of the therapist at times of distress. That incapacity may precipitate panic attacks that have all the physiological symptoms of the disorder, accompanied by the thought that the therapist has died or abandoned them. At such times a brief telephone conversation with the therapist reassures them that they have not been abandoned and immediately quiets the panic attack. Such clinical instances illustrate how psychodynamic factors may trigger neurotransmitter disturbances and how psychotherapeutic interventions may calm those disturbances.

Phobias The classical psychodynamic understanding of phobias is based on the mobilization of ego defense mechanisms in response to signal anxiety. An unacceptable aggressive or sexual thought begins to enter conscious awareness, and the phobic person worries about the possibility of retaliatory punishment for such thoughts. The result is the activation of signal anxiety, which enlists the ego's help in deploying three defense mechanisms—displacement, projection, and avoidance—that work in concert. The unconscious impulse is displaced and projected outward onto a neutral stimulus, such as a dog or an elevator, which is then avoided.

Obsessive-Compulsive Disorder Freud regarded obsessive-compulsive neurosis as reflecting a defensive retreat from the presence of oedipal conflicts in response to the threat of castration anxiety. In Freud's model obsessive-compulsive patients regress to the preceding psychosexual stage, the anal stage, because they feel relatively safe from the threats associated with the oedipal constellation. That regression leads to an unraveling of the smooth fusion between the sexual and the

aggressive drives, so that obsessive-compulsive neurotic patients suffer from intense ambivalence. The patients often feel paralyzed with doubt and indecision because of the simultaneous presence of hateful and loving feelings. To defend against aggressive and sexual impulses that are experienced as on the verge of being out of control, obsessive-compulsive patients use reaction formation, intellectualization, isolation, and undoing.

Recent research suggests that the psychodynamic explanation is only partially explanatory. Because patients with obsessive-compulsive disorder have smaller caudate nucleus volumes than do healthy controls and have almost no placebo responses in controlled medication trials, biological factors appear to play major roles in causing the disorder. However, even biologically based disorders have psychological meanings, and the elucidation of those meanings may be instrumental in the clinical management of those patients.

Posttraumatic Stress Disorder Freud originally placed great importance on trauma in the development of neurosis. Contemporary psychodynamic understanding of anxiety states reflects a rediscovery of the pivotal role of actual environmental trauma in many patients suffering from severe anxiety. Posttraumatic stress disorder became a legitimate diagnostic entity in the official nomenclature as a result of the substantial evidence accumulated regarding the severe and disabling effects of trauma on Vietnam War veterans and on victims of incest, sexual abuse, and rape. One contribution of psychoanalytic thinking has been to understand the meaning of the environmental stressor to a particular patient. Different persons respond differently to the same stressor. Seemingly minor stressors can produce major symptoms in one person, but overwhelming trauma may have minimal effects on another person.

Modern studies on the effects of psychic trauma in childhood have shown that an arrest of affective development occurs in response to such trauma. However, trauma in adulthood tends to produce a regression in affective development. The end result is that trauma victims are unable to use affects as signals. A kind of psychic numbing occurs because any intense emotion is internally regarded as a threat that the original trauma is returning. Consequently, affects are somatized, and many patients who have experienced trauma have psychosomatic illnesses accompanied by an inability to identify emotional states, a condition known as *alexithymia*. Associated with those developments, patients with posttraumatic stress disorder are often incapable of soothing themselves, so that they are in a permanent state of hyperarousal.

Many patients who experience childhood sexual abuse learn to dissociate under the duress of the trauma. They thus convince themselves that the trauma is happening to someone else who has a different name and a different identity. In that manner dissociative identity disorder, which many view as a form of chronic posttraumatic stress disorder, develops.

Generalized Anxiety Disorder Patients with generalized anxiety disorder may experience mild symptoms of autonomic discharge but nothing comparable to the experience of patients with panic disorder. Anxiety in cases of generalized anxiety disorder is a signal of an unconscious danger or threat to the person. The psychodynamic meaning of the signal can be discerned only by a careful evaluation of the source of the patient's fear. Anxiety stems from various developmental levels, each of which has its own characteristic fear. If anxiety stems from oedipal conflicts, a male patient may fear castration or physical harm in response to forbidden wishes or thoughts. Preoedipal concerns may lead to separation anxiety in which patients fear that an important loved one is about to abandon them. At higher developmental levels patients may suffer from superego anxiety, in which case the concern is that they are not living up to the internalized ideals and standards of their parents.

Paraphilias The classical view of the origin of the paraphilias or perversions is that they serve the function of denying castration. For example, a male exhibitionist may display his genitals in public as a way of reassuring himself that he has not been castrated. Freud noted that perversions are complex and multilayered, observing that passive and active counterparts of perverse activity ordinarily coexist in the same person. For example, sadists have a flip side that is masochistic, and exhibitionists are unconsciously voyeuristic.

From a self-psychological perspective, paraphilias are regarded as desperate responses to the absence of appropriately emphatic selfobjects. Perverse sexual activity may restore a sense of cohesiveness to the self in the paraphilic patient. Similarly, sexual acting out during a psychoanalytic process may relate to empathic failures in the analyst.

Contemporary views of the paraphilias have stressed the role of object relatedness in their cause and pathogenesis. Robert Stoller has noted that the essence of perverse sexual activity is the conversion of a passively experienced trauma into an adult act of triumph and mastery. For example, if parents dress a little boy as a girl, he may seek to master his humiliating childhood traumas by becoming a transvestite as an adult. Joyce McDougall has viewed deviant sexual behavior as resulting from unconscious scripts stemming from a complicated set of identifications and counteridentifications with the erotic desires and conflicts of one's parents. She has also noted a fear of self-disintegration involved with much perverse activity, so that certain deviant sexual practices may be experienced as medicating the internal fear of dissolving into nothingness.

Anorexia Nervosa Patients with anorexia nervosa typically lack a sense of autonomy and selfhood. Many patients with the disorder experience their bodies as somehow under the control of their parents. Self-starvation may be an effort to gain validation as a unique and special person. Only through acts of extraordinary self-discipline can the anorexic patient develop a sense of autonomy and selfhood.

Psychoanalytic clinicians who treat patients with anorexia nervosa generally agree that those young patients have been unable to separate psychologically from their mothers. The body may be perceived as if it were inhabited by an introject of an intrusive and unempathic mother. Starvation may have the unconscious meaning of arresting the growth of that intrusive internal object and thereby destroying it. Often, a projective identification process is involved in the interactions between the patient and the patient's family. Many anorexic patients feel that oral desires are greedy and unacceptable, so that those desires are projectively disavowed. Parents respond to the refusal to eat by becoming frantic about whether the patient is actually eating. The patient can then view the parents as the ones who have unacceptable desires and projectively disavow them. Others are voracious and ruled by desire but not the patient.

Bulimia Nervosa Patients suffering from bulimia nervosa lack the superego control and ego strength of their counterparts with anorexia nervosa. The bulimic patients' difficulty in controlling impulses is often manifested by chemical dependency and self-destructive sexual relationships, in addition to the binge eating and purging that are hallmarks of the disorder. Many bulimic patients have histories of difficulty in separating from caretakers manifested by the absence of traditional objects during their early childhood years. Some clinicians have observed that bulimic patients use their own bodies as transitional objects. The struggle for separation from the maternal figure is played out in the ambivalence toward food; eating may represent a wish to fuse with the caretaker, and regurgitating unconsciously expresses a wish for separation.

Psychodynamic treatment of patients with bulimia nervosa has revealed a tendency to concretize introjective and projective defense mechanisms. In a manner analogous to splitting, food is divided into two categories: those items that are nutritious and those that are unhealthy. Food that is designated as good may be ingested and retained within because it unconsciously symbolizes good introjects. But junk food is unconsciously associated with bad introjects and is, therefore, expelled through vomiting, with the unconscious fantasy that all destructiveness, hate, and badness are being evacuated. Patients may temporarily feel good after vomiting because of the fantasized evacuation, but the associated feeling of being all good is short-lived because it is based on an unstable combination of splitting and projection.

Personality Disorders Axis II in DSM-IV consists of personality disorders, an area of psychopathology in which psychoanalytic thinking makes perhaps its most significant contribution to contemporary psychiatry. Personality disorders are long-standing modes of thinking, feeling, and relating to others that are deeply ingrained within the patient. The major features of personality disorders are often *ego-syntonic*, meaning that they are not subjectively experienced by the patient as foreign or distressing in the way neurotic symptoms are. However, personality disorders do create problems for those who must relate to the patient. In many cases those with Axis II disorders do not even seek treatment until the interpersonal and social consequences of their character traits create so much concern that they feel the need to change.

Persons who suffer from personality disorders have generally resolved intrapsychic conflict through the formation of stable defensive patterns that result in severe inhibitions in work and play or that allow for partial gratification of instinctual wishes under certain conditions. The term *character neurosis* is used to describe higher-level personality disorders, such as those of a predominantly obsessive-compulsive or hysterical nature. In those conditions, discrete neurotic symptoms are absent, but certain defensive patterns and cognitive styles pervade the entire personality. Some lower-level personality disorders, such as borderline and antisocial conditions, may involve deficit, as well as certain conflict, and may manifest themselves in certain characteristic ego weaknesses, particularly in the areas of frustration tolerance, drive regulation, object relations, and affective control.

Paranoid Personality Disorder Persons suffering from paranoid personality disorder have rigid, pervasive cognitive styles involving suspiciousness, hypervigilance,

inability to trust or confide in others, pathological jealousy, and a proneness to be easily slighted. Their intrapsychic world is organized along the lines of the paranoid-schizoid position described by Melanie Klein. Splitting, projection, and projective identification are the typical defense mechanisms by which those patients separate out all badness, hatred, and aggression within themselves and deposit it projectively into others. As a result, those patients maintain a stable equilibrium but live with persecutory anxiety connected to the constant threat of attack or mistreatment by others. Moreover, because of their defensive style, the patients must be hypervigilant at all times and attempt to control others, so that the perceived threat is kept at bay. Low self-esteem and feelings of inferiority and weakness are at the core of paranoia, and the grandiosity often seen in the patients is a compensatory defense to deal with the pain caused by those unacceptable feelings.

Schizoid and Schizotypal Personality Disorders Persons with schizoid and schizotypal personality disorders are often considered misfits who seem to avoid other people. Outwardly, they may appear to prefer isolation to relatedness and show little conflict over their solitary existence. However, psychodynamic treatment of them has demonstrated that their apparent lack of concern for their mode of adjustment may mask considerable confusion and deepening longings for relatedness. From a psychoanalytic perspective, the term *schizoid* refers to a splitting or fragmentation of the self into a myriad of self-representations that produce a sense of identity diffusion. The patients' absence of conviction about who they are and what they want may cause schizoid persons to feel paralyzed and to withdraw from attempting interpersonal relatedness. Many have also experienced thoroughgoing frustration as children regarding their ability to get their emotional needs met by others, so they attempt to be self-sufficient to avoid the pain of repeated rejections and of disappointments in others. They may feel that their neediness is so voracious that it would destroy others or drain them of all they have to give. Alternatively, they may fear that they will be consumed or smothered by those from whom they seek emotional sustenance. Hence, many feel caught in a schizoid compromise in which they adjust to a solitary existence as a way of dealing with the twin fears of driving others away by their neediness and of being consumed by the demands of others.

Borderline Personality Disorder Although the borderline concept has been around for many years, the work of Otto Kernberg from his ego psychological-object relations perspective in the 1960s provided the first comprehensive psychodynamic understanding of the condition. He used the term *borderline personality organization* to describe a cluster of characteristics typical of a group of patients who appear to have a mixture of both neurotic and psychotic symptoms. Kernberg's structural analysis of that level of ego organization include the following four elements:

1. Nonspecific manifestations of ego weakness, including poor anxiety tolerance, poor impulse control, and poorly developed subliminatory channels
2. A shift toward primary process thinking, particularly under the pressure of strong affects or in unstructured situations, such as projective psychological testing
3. A pattern of specific defensive operations, including splitting, projective identification, primitive idealization, denial, omnipotence, and devaluation
4. Pathological internalized object relations based on splitting, so that representations of self and others are alternatingly all good or all bad.

Although Kernberg's psychodynamic conceptualization applies to most patients with borderline personality disorder, it was developed as a level of *ego organization* that also encompasses other personality disorders. Kernberg subsumed narcissistic, schizoid, antisocial, paranoid, and lower-level histrionic (infantile) personality disorders under the umbrella of borderline personality organization.

Kernberg also linked the cause and the pathogenesis of borderline personality disorder to a specific developmental arrest during the rapprochement subphase delineated by Mahler and her colleagues. Because those patients can distinguish self from object, Kernberg thought that they had fully traversed the symbiotic phase but had not been able to successfully negotiate separation-individuation. In Kernberg's view, something goes wrong during the rapprochement subphase of development, so that borderline patients never get over the fear that maternal figures will abandon them. He postulated that either a disturbance in the mother's emotional availability or an inborn excess of aggression in the child (or perhaps a combination of both) results in the developmental arrest. As a consequence, the children grow up with chronic anxiety about being abandoned by significant people in their lives. Similarly, they are never able to integrate positive and negative aspects of self-representations and object representations and to internalize well-rounded and complex internal images of the mother to provide object constancy.

Gerald Adler, by contrast, has approached borderline personality disorder from a deficit or insufficiency model that is influenced by self-psychological theories. He understood the borderline patient to be one who grew up in an environment where selfobject needs were not satisfied. Hence, the patients continue to search for selfobjects as adults because they have not developed a holding-soothing introject. Whereas Kernberg viewed the borderline patient as suffering from a predominance of hostile introjects, Adler regarded those same patients as suffering from an absence of introjects, so that they are prone to fragmentations of the self when others do not provide selfobject functions for them. Adler referred to that fear of self-disintegration as *annihilation panic*.

In recent years a considerable body of research has documented an extensive history of childhood trauma in many patients with borderline personality disorder. Approximately 50 percent have experienced incest or other forms of childhood sexual abuse that is etiologically relevant to the disorder; others have experienced physical and verbal abuse as well. Like other patients with a history of early trauma, abused borderline personality disorder patients may reenact characteristic patterns of object relatedness involving abuser-victim paradigms. In addition, the horrific histories of childhood trauma may stir up countertransference needs to rescue in the clinician treating the patient. Finally, the fact of abusive figures in childhood allows the clinician to empathize with the borderline personality disorder patient's need to perceive treaters as “bad objects” who are assumed to be untrustworthy.

Narcissistic Personality Disorder The two major psychodynamic theories of narcissistic personality disorder are those of Kohut and Kernberg. Kohut's theory of the cause and pathogenesis was discussed under the heading of “Self Psychology.” Kohut's view is that narcissistic patients are developmentally arrested because of parental failures of empathy in response to the child's phase-appropriate needs for idealization, mirroring, and twinship experiences. Those children then grow up treating others as though they exist only to gratify their narcissistic needs. Persons who are expected to fill selfobject functions for narcissistic persons are usually alienated by the self-centered nature of the relationship and reject the narcissistically disordered person. Hence, a patient with narcissistic personality disorder goes through life suffering repeated fragmentations of the self in response to the failure of others to fulfill those selfobject needs.

Kernberg offered a different psychodynamic formulation of narcissistic personality disorder. He saw a person with the disorder as essentially similar to a person with borderline personality disorder in that they both operate from the same level of personality organization. The corresponding defensive operations and pathological internal object relations are also present. Kernberg differentiated a narcissistic patient from a borderline patient on the basis of the presence in the narcissistic patient of an integrated but pathologically grandiose self. He conceptualized the grandiose self as being a defensive structure in narcissistic personality disorder that denies dependency on others. Kernberg saw the grandiose self as a fusion of the real self, the ideal object, and the ideal self. In other words, a narcissistic patient maintains an illusion of self-sufficiency through the grandiose self-image.

Kernberg further differentiated narcissistic personality disorder from borderline personality disorder by the presence of higher-level ego functioning in patients with narcissistic personality disorder. Narcissistic patients look much the same from day to day because their self-representation is more continuous than the alternating self-representations of the patient with borderline personality disorder.

Kohut discussed the libidinal and selfobject needs of narcissistic patients, but Kernberg stressed the underlying envy and aggression in patients with narcissistic personality disorder. Because of excessive aggression from either constitutional or environmental sources, narcissistic patients are seen as suffering from chronic intense envy that leads them to destroy and spoil the positive aspects of others whom they envy. Kohut thought that envy does not play a central role in causing narcissistic personality disorder and that aggression is secondary to narcissistic injury, rather than primary, as Kernberg argued. Although Kohut postulated that idealization, commonly seen in narcissistic patients, is a normal developmental need, Kernberg conceptualized idealization as defensive against underlying contempt, rage, and envy.

Antisocial Personality Disorder Of all the Axis II conditions, antisocial personality disorder is the one most clearly linked to biological underpinnings. Twin studies suggest that genetic factors are influential, neuropsychological testing demonstrates the presence of organically based cognitive impairment, and measures of autonomic indicators suggest that antisocial patients are autonomically hyporeactive. The normal attachment process of the infant to the mother may be compromised by those innate constitutional factors. In addition, patients with antisocial personality disorder frequently report a history of abuse or neglect by parental figures. Some authors believe that antisocial personality disorder is psychodynamically related to narcissistic personality disorder. Therefore, a pathological grandiose self is often a key feature of persons with antisocial tendencies. However, such patients differ from narcissistic patients in the nature of the parental introject. Typically, antisocial persons have internalized a cruel, malevolent parental figure, instead of the ideal object that is internalized by those with narcissistic personality disorder. Moreover, the antisocial person never really becomes sensitive to other people as separate persons with needs and feelings of their own. Those major impairments in the internalization process result in a person with no moral conscience. The only trace of superego development may be the presence of sadistic superego precursors. The only value system, if one can call it that, in patients with antisocial personality disorder is the exercise of power and destructiveness over others. Every relationship is approached as an opportunity for exploitation and cruelty. The absence of superego development in those patients makes them extraordinarily difficult or impossible to treat.

Histrionic Personality Disorder Histrionic personality disorder is best conceptualized as a continuum between a lower-level orally fixated histrionic patient on one

end and a higher-level hysterical patient with a phallic fixation on the other end. A female patient at the lower end of the continuum typically turns to her father for maternal nurturance that she feels she did not receive from her mother. She learns that dramatic affective displays are required to capture the attention of her father and other men. Genital sexual relations may be characterized by primitive oral neediness. She then discovers that sexual behavior is rarely satisfying because she is searching for a maternal breast and is receiving a penis instead, an unconscious fantasy known as the *breast-penis equation*.

At the higher end of the continuum, a hysterical female patient has typically been more successful in receiving maternal nurturance than was the female patient at the lower end. However, the hysterical woman is unable to relinquish her libidinal attachment to her father and grows up with an incompletely resolved oedipal situation. She idealizes her father as a man who has no equal. She finds herself in triangular relationships throughout her life, in rivalrous situations with other women for the affections of a man who may be forbidden or unattainable. The hysterical woman frequently chooses romantic partners who are already married or otherwise inappropriate as a way of assuring herself that she will not have to give up her attachment to her father. Because repression is a primary defense used by a patient with hysterical personality disorder, the attachment is often entirely unconscious, and only through psychoanalysis or psychotherapy does she become aware of her tendency to repeat the oedipal drama throughout her adult life.

Although hysterical personality disorder is commonly associated with women, male patients may also meet the diagnostic criteria. The dynamics of such cases is remarkably similar to those found in female patients. The histrionic male patient turns to his father for maternal nurturance. In the absence of an motionally available father, the boy child may either show a hypermasculine flight from concerns about feminine identification with the mother or show a passive effeminate identity in direct identification with his mother. At the higher end of the continuum, the male patient with hysterical personality disorder remains attached to his mother in an incompletely resolved oedipal situation, much like his female counterpart. He may find that the women he meets are disappointing because they never measure up to his mother, or he may choose a celibate or isolated life-style to maintain his unconscious attachment to his mother.

Obsessive-Compulsive Personality Disorder The early papers on the obsessive-compulsive character emphasize regression to an anal fixation in response to castration anxiety encountered in the oedipal phase. The anal traits associated with the character pattern include parsimony, excessive orderliness, and stubbornness. Much like patients with symptomatic obsessive-compulsive neurosis, anally fixated characters were viewed as victims of a harsh superego that is prone to be punitive in response to sexual and aggressive wishes.

As psychoanalytic theory has shifted more and more in the direction of the vicissitudes of object relations, the literature on obsessive-compulsive personality disorder has focused less on anal character traits and more on difficulties in self-esteem, management of dependency and anger, cognitive style, and problems with intimacy. Self-doubt is a common trait of obsessive-compulsive patients, and psychoanalytic treatment frequently reveals the entrenched conviction that they were not loved or valued by their parents. Intense anger directed at the parents and dependent longings are frequently defended against through reaction formation and isolation of affect.

Patients with obsessive-compulsive personality disorder also reveal perfectionistic strivings that are designed to transcend such unacceptable feelings as rage and dependency. They often feel that, if they could just do enough, they will finally win the parental approval that they missed in childhood. Intimacy presents a risk to obsessive-compulsive patients because they fear that their dependent yearnings and angry resentment will be triggered and then get out of control. Most patients with obsessive-compulsive personality disorder prefer work to the unpredictable intimacy of personal relationships. Spouses and significant others often complain that their obsessive-compulsive loved ones are too controlling and lack spontaneity in marriage and other close relationships.

Avoidant Personality Disorder Although it has been controversial because of its overlap with other personality disorders, avoidant personality disorder can be linked to the psychoanalytic tradition of identifying character neuroses that evolved from studying symptomatic neuroses. Avoidant personality disorder can be viewed as a characterological version of phobic neurosis. Persons with avoidant personality disorder are extremely shy and fear social situations in which they may experience humiliation or embarrassment. Although social phobias involve specific situations, the threat in avoidant personality disorder is far more generalized.

The defensive posture of shyness or avoidance typically protects against anticipated rejection, failure, humiliation, and embarrassment. Shame is often a key affective experience for patients with avoidant personality disorder. Psychodynamic exploration reveals that close relationships and social situations are avoided because of the fear that personal inadequacies will be obvious to everyone. Avoidant patients feel that they can never measure up to their own internal standard of performance. Hence, they experience a form of stage fright in interpersonal situations. Some research suggests a genetic-constitutional basis for shyness, but environmental experiences of ridicule or rejection seem to be necessary to produce the full characterological manifestations of avoidant personality disorder.

Dependent Personality Disorder Persons with dependent personality disorder often reveal family backgrounds characterized by overinvolved parents. Messages are given to the children that autonomy within the family is potentially dangerous. Loyalty to and dependency on the family are rewarded. Classical psychoanalytic explanations involving oral-stage fixation are no longer viewed as pertinent to the pathogenesis of dependent personality disorder because the stage specificity of that hypothesis is a dubious proposition. Persons who are extremely dependent have generally experienced parental messages about the dangers of separation throughout the entirety of childhood, a pervasive pattern that is not linked to one particular stage of development. Psychodynamic exploration of patients with dependent personality disorder has also shown that dependency may mask considerable hostility and aggression.

PSYCHOANALYTIC THINKING IN CONTEMPORARY PSYCHIATRY

A detailed discussion of formal psychoanalytic treatment is included in Section 30.1, Psychoanalysis and Psychoanalytic Psychotherapy. However, there are many useful applications of psychoanalytic thinking in contemporary psychiatric treatment that will be considered here. Psychodynamic and biological approaches often work synergistically in treatment settings. Despite the either-or polarization of the biological and psychodynamic orientations in some quarters, psychodynamic psychiatry is not inherently antibiological, and an integrated approach has much to offer.

For psychiatry to avoid the twin pitfalls of becoming either mindless or brainless, clinicians must acknowledge the complex relations between neurophysiological and psychosocial factors in the cause and the pathogenesis of psychiatric disorders. To say that a chemical imbalance, for example, is the cause of a disorder is reductionistic. Neurophysiological or biochemical processes in the brain are mediating mechanisms rather than causal agents. The subjective meaning of the information perceived from the environment sets the biological processes in motion.

The interface between psychological trauma and brain functioning has been amply documented in primate research. Squirrel monkeys who are removed from their mothers undergo dramatic physiological changes, some of which become permanent if the separations are repeated. Among the physiological changes produced by early separation are persistently elevated plasma cortisol, lasting alterations in the sensitivity of noradrenergic receptors, permanent changes in the sensitivity and the number of brain opiate receptors, changes in adrenal gland catecholamine-synthesizing enzymes, and changes in hypothalamic serotonin secretion. Moreover, the psychoanalytic idea that environmental insults may cause varying degrees of psychological damage, depending on the specific developmental phase during which they occur, received some support from the primate research. In one series of experiments, isolating infant monkeys from their mothers resulted in a more serious effect if the separation occurred at 90 days than if it occurred at either 60 or 120 days; perhaps the difference is related to a link between certain forms of bonding behavior and myelinization of the nervous system.

An enormous body of research since the late 1980s has amply documented the phenomenon of neural plasticity. The brain is not a static structure. It undergoes structural and functional changes almost daily. Studies on rats and marine snails have demonstrated that the number of synapses per neuron increases when learning takes place in interaction with the environment. By analogy, psychotherapy inevitably must work by altering the brain. Studies using positron emission tomography scans have demonstrated that patients with obsessive-compulsive disorder undergo similar brain changes, regardless of whether the treatment is cognitive-behavior therapy or medication.

Research has also proven convincingly that genes are far from static. Stressors from the environment appear to regulate gene expression in ways that have profound influence on the presence or absence of psychopathology. In disorders like major depressive disorder, genetic factors appear to alter an individual's susceptibility to react to environmental and psychological stressors with a depressive episode. Similarly, positive environmental influences may help an individual to overcome a problematic genetic diathesis. For example, primate studies have found that approximately 20 percent of rhesus monkeys have an innate, genetically based temperament characterized by marked social anxiety. However, when these monkeys are reared by especially nurturant mothers in the monkey colony, they overcome this genetic disadvantage and rise to the top of the social hierarchy. The fact that biological and psychosocial factors are inextricably tied together in the etiology of mental illness suggests that a combination of psychotherapeutic and psychopharmacological interventions may be the optimal treatment in many cases.

Finally, to say that psychiatric disorders have strong biological underpinnings is not devoid of meaning. Illnesses are elaborated psychologically by patients who

suffer from them, and clinicians must incorporate questions of meaning into an overall diagnostic understanding and treatment plan.

A 29-year-old single man with a 10-year history of obsessive-compulsive disorder was admitted to a psychiatric hospital. He had been housebound for the preceding 8 years because his obsessional thoughts and compulsive rituals had become incapacitating. The patient's mother had retired from her career so that she could be at home with the patient and accommodate his demands for cleanliness.

The patient lived in constant fear of contamination. Specifically, he worried that he might have semen on his hands that would impregnate women. He washed his hands compulsively to make sure that his fears would not come to fruition. He demanded that his mother be with him virtually every hour of the day. She did not enter the shower with him, nor did she sleep with him, but she helped him dress because he feared his clothes would become contaminated with germs if he touched them. The patient insisted on a complex 58-step ritual when his mother prepared his meals. If she made a small deviation from the ritual, he insisted that she throw out all the food she had prepared and start from scratch. She had wasted thousands of dollars each year in her efforts to conform to the rituals demanded by her son. The patient ordered his father to stay away from the house, lest the father bring in germs from the outside world. The father complied by staying in a separate part of the house, where he rarely had contact with his son.

The patient's early childhood had been basically unremarkable from a developmental perspective, but he did recall a vivid and frightening event when he was 5 years old. He remembered seeing his mother crying out for him to rescue her after his father had grabbed her by the breasts. He remembered trying to stop his father but could not succeed in getting him away from his mother.

By the time the patient had turned 19, he had been severely compromised by obsessive-compulsive symptoms and had been to many psychiatrists. Each time, he refused to return after one appointment. One psychiatrist had managed to obtain the patient's agreement to take clomipramine (Anafranil). However, the patient refused to take the medication after the first dose and subsequently suffered ill-defined adverse effects.

The patient was finally admitted to a hospital because he was completely incapacitated by his symptoms. On admission, his hospital psychiatrist asked him why he was seeking treatment. The patient responded, “I'm determined to be dependent—I mean, independent.” The doctor noted that he initially had said “dependent” and asked, “Is there perhaps a part of you that would like to be dependent?” The patient replied, “You mean on my mother?” The psychiatrist responded that he was sure the patient would know better than he. The patient then thought about his situation momentarily and observed, “Well, she does take pretty good care of me.” He went on to say that his mother also made him nervous when she dressed him. The patient speculated that “there was something sexual about that.”

After a week of hospitalization, the patient became less anxious and less controlled by his symptoms. He was able to touch doorknobs without fearing that he would contract germs, and he could even touch magazines that others had read. He also spent less time washing his hands than he had before admission to the hospital. Those improvements occurred without the assistance of clomipramine or other medication. The patient said he was less anxious because he was away from his mother and did not need to worry about the sexual aspects of her dressing him.

Psychodynamic-Biological Interface The interface between the psychodynamic and the biological is illustrated by the above clinical vignette. Obsessive-compulsive disorder is a condition that is biologically based to a large extent. Patients with the disorder have a smaller caudate nucleus volume than do healthy control persons. Monozygotic twins have a higher rate of concordance for the disorder than do their dizygotic counterparts. Also, patients with obsessive-compulsive disorder manifest virtually no response to a placebo.

Nevertheless, in the case of the above patient, the obsessive-compulsive symptoms were rich in psychological meaning. The patient's symptoms reflected the psychoanalytic concept known as *compromise formation:* they contained both the direct expression of an underlying wish and a defense against that wish. The patient's obsessional thoughts and compulsive rituals defended against his sexual wishes for his mother by causing him to spend every waking moment in hand-washing rituals and other activities. However, these symbolic rituals also produced a situation in which his mother showered attention on him, even to the point of dressing him, while his father stayed in another part of the house. For all practical purposes the patient had created a symbolic oedipal victory; he had successfully stolen his mother away from his father. His resistance to taking clomipramine can be understood as a refusal to give up his triumphant position with his mother. His oedipal victory also created guilt and anxiety, which led to increased reliance on the rituals and obsessions. The hospitalization removed him from the anxiety-producing situation at home; therefore, his symptoms were not necessary to deal with his anxiety.

An analogy from a common elementary school demonstration is useful in explicating the relation between the biological and the psychodynamic. If iron filings are placed on a sheet of paper with a magnet underneath, the filings line up in formation on the surface and follow the movement of the magnet as it glides beneath the paper. Similarly, psychodynamic themes often appropriate the powerful magnetlike biological forces in the brain for their own purposes. The obsessive-compulsive patient's unconscious wishes and his defenses against them appropriated the neurophysiologically driven obsessive-compulsive symptoms and used them as a vehicle for the expression of his wishes and defenses.

Noncompliance Noncompliance with pharmacotherapeutic regimens is an everyday problem in contemporary psychiatry. The psychodynamic concepts of resistance and transference may be extraordinarily useful in understanding noncompliance. In the case of the obsessive-compulsive patient above, the psychoanalytically informed understanding of his resistance to taking medication was essential to his successful treatment, which ultimately involved medication and therapeutic work with the family. The patient's slip of the tongue made him aware for the first time that any improvement in his symptoms would threaten his privileged position with his mother.

The psychodynamic understanding of the patient's obsessive-compulsive symptoms illustrates an important distinction within dynamic psychiatry, namely, the difference between causation and meaning. The obsessive-compulsive symptoms may be caused by biological forces, but they are imbued with meaning by the unconscious of the patient. Similarly, only one fourth of persons who witness a traumatic event develop symptoms of posttraumatic stress disorder. Although there may be genetic vulnerability in some cases, one can also speculate that particular forms of trauma have specific unconscious meanings to certain persons that may contribute to their becoming symptomatic.

SUGGESTED CROSS-REFERENCES

The psychoanalytic perspective is relevant to virtually every chapter of this book. Of particular interest are the discussion of Erik Erikson in Section 6.2, other psychodynamic schools in Section 6.3, approaches derived from psychology and philosophy in Section 6.4, psychodynamic to neurodynamic theories and individual psychotherapy of schizophrenia in Section 12.6 and Section 12.10, psychodynamic etiology and psychological treatments of mood disorders in Section 14.5 and Section 14.9 anxiety disorders in Chapter 15, personality disorders in Chapter 24, psychoanalysis and psychoanalytic psychotherapy in Section 30.1, evaluation of psychotherapy in Section 30.11, and psychotherapy with the elderly in Section 51.4i.

SECTION REFERENCES

Abend SA: Countertransference and psychoanalytic technique. Psychoanal Q *58:*374, 1989.

Aron L: *A Meeting of Minds: Mutuality and Psychoanalysis*. Analytic Press, Hillsdale, NJ, 1996.

Bacal H, Newman K: *Theories of Object Relations: Bridges to Self Psychology*. Columbia University Press, New York, 1990.

Balint M: *The Basic Fault: The Therapeutic Aspects of Regression*. Brunner/Mazel, New York, 1979.

Bechara A, Damasio H, Tranel D, Damasio AR: Deciding advantageously before knowing the advantageous strategy. Science *275:*1293, 1997.

Benjamin J: *Shadow of the Other: Intersubjectivity and Gender in Psychoanalysis*. Routledge, New York and London, 1998.

Benjamin J: An outline of intersubjectivity: The development of recognition. Psychoanal Psychol *7:*33, 1990.

Brenner C: *The Mind in Conflict*. International Universities Press, New York, 1982.

Chessick RD: *What Constitutes the Patient in Psychotherapy: Alternative Approaches to Understanding Humans*. Jason Aronson, Northvale, NJ, 1984.

Cooper AM: Changes in psychoanalytic ideas: Transference interpretation. J Am Psychoanal Assoc *35:*77, 1987.

Doidge N, Simon B, Gillies LA, Ruskin R: Characteristics of psychoanalytic patients under a nationalized plan: DSM-III-R diagnoses, previous treatment, and childhood trauma. Am J Psychiatry *151:*586, 1994.

Eagle M: The concepts of need and wish in self psychology. Psychoanal Psychol *7:*71, 1990.

Erdelyi MH: *The Recovery of Unconscious Memories: Hypermnesia and Reminiscence*. University of Chicago Press, Chicago, 1996.

Fairbairn WRD: *Psychoanalytic Studies of the Personality*. Routledge & Kegan Paul, London, 1952.

*Freud A: *The Ego and the Mechanisms of Defense*, rev ed. International Universities Press, New York, 1966.

*Freud S: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Hogarth Press, London, 1953–1966.

Friedman L: How and why do patients become more objective? Sterba compared with Strachey. Psychoanal Q *61:*1, 1992.

Gabbard GO: A reconsideration of objectivity in the analyst. Int J Psychonal *78:*15, 1997.

Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

Gabbard GO: Psychodynamic psychiatry in the “decade of the brain.” Am J Psychiatry *149:*991, 1992.

Gill M: *Analysis of Transference: Theory and Technique*, vol 1. International Universities Press, New York, 1982.

Gill MM: *Psychoanalysis in Transition: A Personal View*. Analytic Press, Hillsdale, NJ, 1994.

Goldberg A: Farewell to the objective analyst. Int J Psychoanal *75:*21, 1994.

Greenberg JR, Mitchell SA: *Object Relations in Psychoanalytic Theory*. Harvard University Press, Cambridge, MA, 1983.

*Greenson RR: *The Technique and Practice of Psychoanalysis*, vol 1. International Universities Press, New York, 1967.

Hartmann H: *Essays on Ego Psychology: Selected Problems in Psychoanalytic Theory*. International Universities Press, New York, 1964.

Hoffman IZ: Dialectical thinking and therapeutic action in the psychoanalytic process. Psychoanal Q *63:*187, 1994.

Hoffman IZ: The intimate and ironic authority of the psychoanalyst's presence. Psychoanal Q *55:*102, 1996.

Holzman PS, Aronson G: Psychoanalysis and its neighboring sciences: Paradigms and opportunities. J Am Psychoanal Assoc *40:*63, 1992.

Jacobson E: *The Self and the Object World*. International Universities Press, New York, 1964.

Jacobson JG: Signal affects and our psychoanalytic confusion of tongues. J Am Psychoanal Assoc *42:*15, 1994.

*Kandel ER: Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited. Am J Psychiatry *4:*505, 1999.

Kernberg OF: *Borderline Conditions and Pathological Narcissism*. Aronson, New York, 1975.

Kernberg OF: The current status of psychoanalysis. J Am Psychoanal Assoc *41:*45, 1993.

Klein M: *Envy and Gratitude and Other Works, 1946–1963*. Free Press, New York, 1975.

*Kohon G: *The British School of Psychoanalysis: The Independent Tradition*. Yale University Press, New Haven, 1986.

Kohut H: *The Analysis of the Self: A Systematic Approach to the Psychoanalytic Treatment of Narcissistic Personality Disorder*. International Universities Press, New York, 1971.

*Kohut H: *How Does Analysis Cure?* University of Chicago Press, Chicago, 1984.

Kohut H: *The Restoration of the Self*. International Universities Press, New York, 1977.

Lichtenberg J: *Psychoanalysis and Motivation*. Analytic Press, Hillsdale, NJ, 1989.

Lichtenberg J, Lachmann F, Fosshage J: *The Clinical Exchange: Techniques Derived from Self and Motivational Systems*. Analytic Press, Hillsdale, NJ, 1996.

Makari GJ: Dora's hysteria and the maturation of Sigmund Freud's transference theory: A new historical interpretation. J Am Psychoanal Assoc *45:*1061, 1997.

Mitchell SA: Contemporary perspectives on self: Toward an integration. Psychoanal Dial *1:*121, 1991.

Mitchell SA: *Hope and Dread in Psychoanalysis*. Basic Books, New York, 1993.

Ogden TH: The concept of internal object relations. Int J Psychoanal *64:*227, 1983.

Ogden TH: *Subjects of Analysis*. Aronson, Northvale, NJ, 1994.

Opatow B: The real unconscious. J Am Psychoanal Assoc *45:*865, 1997.

Parsons M: The refinding of theory in clinical practice. Int J Psychoanal *73:*103, 1992.

Renik O: Analytic interaction: Conceptualizing technique in light of the analyst's irreducible subjectivity. Psychoanal Q *62:*553, 1993.

Sandler J: On internal object relations. J Am Psychoanal Assoc *38:*859, 1990.

Scharff JS: *Projective and Introjective Identification and the Use of the Therapist's Self*. Aronson, Northvale, NJ, 1992.

Shane M, Shane E: Self psychology after Kohut: One theory or many? J Am Psychoanal Assoc *1:*777, 1993.

Simon B: In search of psychoanalytic technique: Perspectives from on the couch and from behind the couch. J Am Psychoanal Assoc *41:*1051, 1993.

Stern DN: *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. Basic Books, New York, 1985.

Stolorow R, Brandchaft B, Atwood G: *Psychoanalytic Treatment: An Intersubjective Approach*. Analytic Press, Hillsdale, NJ, 1987.

Tyson P: Theories of female psychology. J Am Psychoanal Assoc *42:*447, 1994.

Van der Kolk BA: *Psychological Trauma*. American Psychiatric Press, Washington, DC, 1987.

Wakefield JC: Freud and the intentionality of affect. Psychoanal Psychol *9:*1, 1992.

Westen D: Towards a clinically and empirically sound theory of motivation. Intern J Psychoanal *78:*521, 1997.

Winnicott DW: *The Maturational Process and the Facilitating Environment: Studies in the Theory of Emotional Development*. Hogart Press, London, 1965.

Winnicott DW: *Playing and Reality*. Basic Books, New York, 1971.

Wolf E: *Treating the Self: Elements of Clinical Self Psychology*. Guilford, New York, 1988.

6.2 ERIK H. ERIKSON

DORIAN S. NEWTON, PH.D. AND PETER M. NEWTON, PH.D.

Erik Erikson was a psychoanalyst who created an original and highly influential theory of psychological development and crisis occurring in periods that extended across the entire life cycle. His theory grew out of his work first as a teacher, then as a child psychoanalyst, next as an anthropological field worker, and finally as a biographer. Rather than starting within the nervous system of the individual as Freud had done, Erikson focused upon the boundary between the child and the environment, then graphed the evolution of the maturing ego's relations with an expanding social world. Erikson identified dilemmas or polarities in the ego's relations with the family and larger social institutions at nodal points in early, middle, and late adulthood, as well as during the years of most rapid development in childhood.

Of Erikson's many works, three are already firmly established as classics. *Childhood and Society* set forth his theory of the life cycle. *Young Man Luther*, reconstructed and analyzed the developmental crisis of identity in a very young man and its creative resolution in his 30s. *Gandhi's Truth* showed the coming together of authentic identity and calling in a man at midlife. In his psychological biographies of these world historical leaders, Erikson demonstrated the interrelations between individual psychodynamics and development on the one hand and social structure and history on the other, gracefully avoiding a reductionism in either a psychodynamic or sociological direction. Erikson showed in complex, intricate detail how great leaders solve the problem of their own identity by creating solutions for crises in their culture.

Erikson's developmental ideas are illustrated here as applied to his own life and work, primarily during youth and middle age. Some terms used here, such as *life structure*, the developmental phase of “becoming one's own man,” and “the mid-life transition” come from Erikson's student, Daniel Levinson, who, building upon his teacher's work became a leading psychosocial theorist of the life cycle.

LIFE AND WORK

Erik Homburger Erikson was born on June 15, 1902, in Karlsruhe Germany, an old capital of a Lutheran principality. His father was Protestant and his mother Jewish. His parents, both Danish, had separated before his birth; his mother was visiting friends in Germany when the baby was born. She stayed in Karlsruhe and a few years later married her child's pediatrician, a well-to-do Jewish doctor, Theodor Homburger, in whose house young Erik grew up. During an alienated adolescence the tall, blond boy found himself regarded as a gentile in his father's Jewish milieu and as a Jew at school. He remembered his mother as sad, bookish, and artistic; his adoptive father as professionally respected; and both as loving. The boy's adoption by his stepfather was the formative occasion of what became a lifelong pattern of getting himself adopted by kind men. His last name remained Homburger until he was 37, when he changed it to Erikson.

Erikson attended the *Humanistiche Gymnasium* in Karlsruhe where he studied Greek, Latin, philosophy, literature, and science. By the age of 18 he had the educational attainments of all but a few American college graduates and a stronger base than many. His primary interest was art; impatient with formal study and possessed by a restlessness he never lost, Erikson chose not to go to a university, preferring to travel about the countryside reading, drawing, and making wood carvings. Back home after a year, he tried formal art study, first in Karlsruhe and then in Munich with some success but in neither case with decisive commitment.

This sort of wandering about was not uncommon among German youth of the period, so Erikson was permitted, as his biographer Robert Coles sagely wrote, “to go through his own years of discontent and confusion without being especially singled out and thereby forced to defend behavior often best granted the limits of its own momentum.” Erikson was having what he would later term a *psychosocial moratorium*; in so doing, he was also mitigating the asperities of an *identity crisis*. At about age 21 Erikson went to live in Florence where he continued his art studies informally. There he enjoyed the friendship of his old *Gymnasium* chum Peter Blos, a writer who later became a famous American child psychoanalyst.

Becoming a Child Psychoanalyst Back in Karlsruhe Erikson felt ready to make art teaching the central component of his first adult life structure. When Erikson was 24 or 25 years of age, Peter Blos invited him to become a faculty member in a progressive grammar school in Vienna, where Blos taught language and science. Erikson's difficult transition from adolescence to early adulthood was over. The year was 1927; Sigmund Freud was 71, and his youngest child, Anna, an educator and psychoanalyst, had started a psychoanalytically enlightened school for children with an American friend named Dorothy Burlingham. Erikson joined Blos and later recalled Blos's determination to turn him into a disciplined worker: “To make a teacher of me... the highly disciplined Peter first had to teach me to keep regular work hours, a task which was initiated every morning, no matter what time of year, by a cold shower, then the preferred shock treatment for identity confusion.”

Before long, Erikson found himself not only a teacher of children but also an *analysand*, what is now called a *candidate*, at the Vienna Psychoanalytic Institute, in treatment with Anna Freud. Any sort of psychoanalytic approach to treating or educating children was then a radical idea, even an approach as cautious as Freud's. Both educationally and clinically, the Freuds' group was sufficiently deviant such that a person with Erikson's diverse identity was able to fit in. There was in 1927 a *configurational affinity* between Erikson's personal history and the history of psychoanalysis as a profession. Still there was the worry: What was a fledgling artist without a university education to do among those high-powered theorists and intellectuals at the Vienna Psychoanalytic Institute?

Erikson recalled this epiphanic exchange with his analyst, Anna Freud: “When I declared once more that I could not see a place for my artistic inclinations in such high intellectual endeavors, she said quietly: ‘You might help to make them see.’" Dreams had been Freud's royal road to the unconscious; observing children's play (and later, anthropological field studies in the role of participant observer) would be Erikson's path to understanding the ego and its development. Looking back, Erikson thought that it was Anna Freud's “simple mandate” that enabled him to succeed in combining the artistic and the theoretical in *Childhood and Society*.

Erikson remained in Vienna for 6 years, until 1933. In those years between the ages of 25 and 31, he continued as a novice adult to develop the skills and meanings of occupation in his life structure. He turned his artist's eye from the observation of nature to the analysis of children; he learned about psychoanalysis by studying children's play and his own free associations as a patient, and he learned fundamental clinical skills in the supervised treatment of others. He formed mentor relationships with Freud and Heinz Hartmann, and a more distant, admiring-inspirational relationship with Anna Freud's sick, aging, but still-productive father.

During his own late adulthood Sigmund Freud had moved beyond clinical problems more narrowly conceived to problems of the ego, society, and history in works such as *Beyond the Pleasure Principle* (written at age 64), *Group Psychology and the Analysis of the Ego* (at 65), and *The Ego and the Id* (at 66); during Erikson's years in Vienna, Freud was working on *The Future of An Illusion* (at 71) and *Civilization and Its Discontents* (at 73). Erikson seemed to identify with both Freuds, internalizing them as mentors, and they provided him with a strong, inner basis for a lifetime of psychoanalytic treatment and psychosocial research.

Equally important, during those years Erikson succeeded in committing himself to what was be the other central component of his life structure. In Vienna, he met his wife, Joan Serson, an American woman with a Masters' degree in sociology, who had a special interest in modern dance and psychoanalysis. They married, and Serson joined the faculty of Burlingham's school as an English teacher. Serson's unusual combination of interests and skills—education, psychoanalysis, and sociology—coupled with her ability as a writer, gave Erikson a skilled co-worker for a lifetime of intellectual work. Several articles were co-authored, Joan Erikson helped with most of the others, and the endowed chair at Harvard that was named after Erikson carried both their first names.

With the birth of their sons, Kai and Jon, Erikson had by 31 become a husband, a father, and a child psychoanalyst. He had come a long way from the aimless youth of his early 20s. His first adult life structure was fully formed, and he was productively engaged in the stage of ego development that he later characterized by the polarity *intimacy versus isolation*. He continued to work on that ego stage, but changes were now required in the life structure that Erikson had begun to build at 25.

Emigrating to America By 1933 the Fascist menace in Europe was growing. When Erikson graduated that year from the Vienna Psychoanalytic Institute, the Eriksons prepared to leave. They considered repatriating to Denmark but encountered sufficient annoying difficulties to allow them to justify a real emigration. Home for Erikson was always a fortuitously adoptive one, and so he allowed chance, in the form of a serendipitous meeting with Freud's disciple Hanns Sachs, to settle his

future. Sachs enthusiastically invited him to come to Boston and in 1933 the Eriksons moved to Boston. At 31, Erikson became the city's first child psychoanalyst and a member of the faculty of the Harvard Medical School. With a new home and a new profession as a professor, Erikson was, in his age 30 transition, revising the life structure that he had created in his mid-20s. At Harvard Erikson joined the group of personality researchers working under the leadership of Henry Murray, who was Director of the Harvard Psychological Clinic.

Murray's book *Explorations in Personality: A Clinical and Experimental Study of Fifty Men of College Age* was published in 1938. It is one of the few integrative masterpieces in American psychology—experimental and biographical, psychological and sociological, Freudian and Jungian, developmental and diagnostic—and it provided a luminous example of studying persons both in their origins and in their purposes. Murray's example proved seminal when Erikson later turned his hand to the biographical study of the world historical figures Sigmund Freud, Martin Luther, and Mohandas Gandhi.

Listed on the title page of Murray's book, among other research associates, was “Erik Homburger.” The next year, at age 37, during a developmental phase that Levinson called “becoming one's own man,” Homburger renamed himself Erikson, retaining Homburger as his middle name. With characteristically bold creativity, Erikson solved the problem of his paternity by adopting himself.

Psychoanalytic Anthropology In 1936 Erikson hit the road again, this time to the Institute of Human Relations of the department of psychiatry at Yale University. The interdisciplinary work of the institute further shaped Erikson's interest in cross-cultural research, and in 1938 he joined a colleague on a research expedition to the Sioux Indians in South Dakota. On the Pine Ridge reservation, he observed children and child-rearing practices and interviewed adults, resulting in his paper, “Observations on Sioux Education,” published in 1939. In the same year the peregrinating Eriksons moved to California, where he joined the faculty at the University of California at Berkeley. He remained there for a decade, the longest period in one place since his pre-adulthood years in Karlsruhe. This was the culminating period of his early adulthood, a time to realize the promise and reap the rewards of youth. At Berkeley he became his own man, visible first in the outward signs of changing his name and later in refusing to sign the university's loyalty oath of anti-Communist purity. Most important, during those 10 years, from 37 to 47, Erikson wrote his masterwork, *Childhood and Society* and began his career as a biographer.

At Berkeley Erikson continued his cross-cultural research, this time with the Yurok Indians of northern California. He also lent his research skills to the war effort in articles on submarine habitation and the interrogation of prisoners of war. He continued to do the research begun at Burlingham's experimental school on children's play. But at age 40, as Erikson was entering his mid-life transition, he was becoming more biographical and more interested in the adult years of the life cycle. Like others who manage to continue developing, he found a way to pursue his development *through* the strictures of exigent reality. World War II was under way and Erikson was deeply concerned about it. After his initial narrowly framed attempts at scholarly patriotism, he began writing his first psychobiographical essays on Adolph Hitler and the psychosocial dynamics of his appeal to young Germans—“Hitler's Imagery and German Youth” was published in 1942.

Midlife Transition The primary developmental task at midlife is to terminate the existing life structure, revise one's dream, and initiate a new life structure suitable for middle age. *Childhood and Society* was the creative product of Erikson's midlife transition. He had intended it to be a contribution to the psychiatric education of clinicians from various disciplines, but the book outgrew its author's intentions and found its way into every corner of the academy and beyond.

Erikson began work on the book when he was 42 and largely completed work on it by 46. Published in 1950, it is at once a product of early research, a prospectus of what was to come, and an initial integration of both. In it he presented clinical cases in which individual psychodynamics, society, and history are interwoven with a skill not seen before or since; analyses of children's play and development in various cultures; a theoretical sketch of the entire human life cycle; pieces on the problem of identity; and biographical essays on Adolf Hitler and the Russian writer Maxim Gorky.

As in Karl Abraham's and Freud's psychosexual stages, most of Erikson's stages in ego development occur in childhood and adolescence. Unlike the Abraham-Freud conception, however, Erikson's stages are *psychosocial*, describing crucial steps in the maturing ego's relations with the social world rather than a biological unfolding of neurophysiological capacities for excitation. Also, whereas Freud's developmental theory falls back on itself after adolescence, Erikson's continued with characterizations of developmental tasks during youth, middle age, and old age.

Erikson's emphasis, like Freud's, was both cross-cultural and universal. Erikson's eight stages are ineluctable parts of the human life cycle, yet each person goes through them in distinct ways that are determined by culture, concrete circumstance, and personality. With *Childhood and Society*, Erikson became and remained an *ego psychologist*, shifting the traditional psychoanalytic focus on the drives to adaptation and growth. In so doing, he was furthering the work of his analyst-mentor, Anna Freud, who in 1936 had written ego psychology's basic theoretical treatise, *The Ego and the Mechanisms of Defense*.

Erikson's conceptualization of psychosocial development took as its model the *epigenetic principle* of organismic growth in utero. In Erikson's view psychosocial growth occurs in phases, with individual aspects of development proceeding according to a predetermined time table. Each phase has its own time of quiescence and critical ascendancy, and each is dependent on the proper development of the other phases in the proper sequence. Work on any particular phase, Erikson theorized, is never complete and old developmental conflicts can be activated by critical life events.

Erikson's theory of the life cycle is the centerpiece of his contribution. Although the richness and complexity of Erikson's conception of life cycle development does not lend itself well to schematic presentation, the following chart is consistent with his theory.

Eight Stages of the Life Cycle Erikson's conception of the eight stages of ego development across the life cycle is the centerpiece of his life's work and he elaborated the conception throughout his subsequent writings (Table 6.2-1). The eight stages represent points along a continuum of development in which physical, cognitive, instinctual, and sexual changes combine to trigger an internal crisis whose resolution results in either psychosocial regression or growth and the development of specific *virtues*. In *Insight and Responsibility* Erikson defined virtue as “inherent strength,” as in the active quality of a medicine or a liquor. He wrote in *Identity:Youth and Crisis* that “crisis” refers not to a “threat of catastrophe, but to a turning point, a crucial period of increased vulnerability and heightened potential, and therefore, the ontogenetic source of generational strength and maladjustment.”

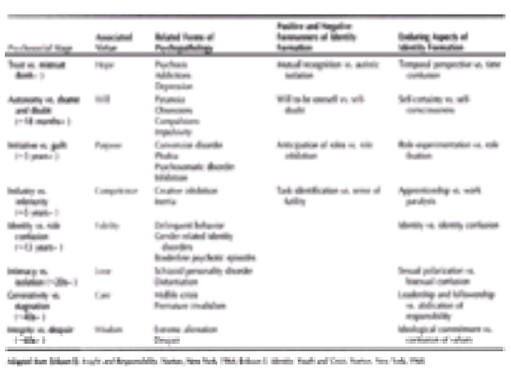

Table 6.2-1 Erikson's Psychosocial Stages

Trust Versus Mistrust (Birth to About 18 Months) In *Identity: Youth and Crisis*, Erikson noted that the infant “lives through and loves with” its mouth. Indeed, the mouth forms the basis of its first *mode* or pattern of behavior, that of incorporation. The infant is taking the world in through the mouth, eyes, ears, and sense of touch. The baby is learning a cultural modality that Erikson termed *to get*, that is, to receive what is offered and elicit what is desired. As the infant's teeth develop and it discovers the pleasure of biting, it enters the second oral stage, the active-incorporative mode. The infant is no longer passively receptive to stimuli; it reaches out for sensation and grasps at its surroundings. The social modality shifts to that of *taking and holding on* to things.

The infant's development of basic trust in the world stems from its earliest experiences with its mother or primary caretaker. In *Childhood and Society* Erikson asserts that trust depends not on “absolute quantities of food or demonstrations of love, but rather on the quality of maternal relationship.” A baby whose mother is able to anticipate and respond to its needs in a consistent and timely manner despite its oral aggression will learn to tolerate the inevitable moments of frustration and deprivation. The defense mechanisms of introjection and projection will provide the infant with the means to internalize pleasure and externalize pain such that

“consistency, continuity, and sameness of experience provide a rudimentary sense of ego identity.” Trust will predominate over mistrust, and hope will crystallize. For Erikson, the element of society corresponding to this stage of ego identity is religion, as both are founded upon “trust born of care.”

In keeping with his emphasis on the epigenetic character of psychosocial change, Erikson conceived of many forms of psychopathology as examples of what he termed *aggravated development crisis*, development that, having gone awry at one point, affects subsequent psychosocial change. A person who, as a result of severe disturbances in the earliest dyadic relationships, fails to develop a basic sense of trust or the virtue of hope may be predisposed as an adult to the profound withdrawal and regression characteristic of schizophrenia. Erikson hypothesized that the depressed patient's experience of being empty and of being no good is an outgrowth of a developmental derailment that causes oral pessimism to predominate. Addictions may also be traced to the mode of oral incorporation.

Autonomy Versus Shame and Doubt (About 18 Months to About 3 Years) In the development of speech and sphincter and muscular control, the toddler practices the social modalities of *holding on and letting go*, and experiences the first stirrings of the virtue that Erikson termed *will*. Much depends on the amount and the type of control exercised by adults over the child. Control that is exerted too rigidly or too early defeats the toddler's attempts to develop its own internal controls, and regression or false progression results. Parental control that fails to protect the toddler from the consequences of his or her own lack of self-control or judgment can be equally disastrous to the child's development of a healthy sense of autonomy. In *Identity: Youth and Crisis*, Erikson asserted: “This stage, therefore, becomes decisive for the ratio between loving good will and hateful self-insistence, between cooperation and willfulness, and between self-expression and compulsive self-restraint or meek compliance.”

Where that ratio is favorable, the child will develop an appropriate sense of autonomy and the capacity to “have and to hold”; where it is unfavorable, doubt and shame will undermine free will. According to Erikson, the principle of law and order has at its roots this early preoccupation with the protection and regulation of will. In *Childhood and Society*, he concluded, “[T]he sense of autonomy fostered in the child and modified as life progresses, serves (and is served by) the preservation in economic and political life of a sense of justice.”

An individual who becomes fixated at the transition between the development of hope and autonomous will, with its residue of mistrust and doubt, may develop paranoic fears of persecution. When psychosocial development is derailed in the second stage, other forms of pathology may emerge. The perfectionism, inflexibility, and stinginess of the person with an obsessive-compulsive personality disorder may stem from conflicting tendencies to hold on and to let go. The ruminative and ritualistic behavior of the person who suffers from an obsessive-compulsive disorder may be an outcome of the triumph of doubt over autonomy and the subsequent development of a primitively harsh conscience.

Initiative Versus Guilt (About 3 Years to About 5 Years) The child's increasing mastery of locomotor and language skills expands its participation in the outside world and stimulates omnipotent fantasies of wider exploration and conquest. Here the youngster's mode of participation is active and intrusive; its social modality is that of *being on the make*. The intrusiveness is manifested in the child's fervent curiosity and genital preoccupations, competitiveness, and physical aggression. The Oedipus complex is in ascendance as the child competes with the same-sex parent for the fantasied possession of the other parent. In *Identity: Youth and Crisis*, Erikson wrote that “[j]ealousy and rivalry... now come to a climax in a final contest for a favored position with one of the parents: the inevitable and necessary failure leads to guilt and anxiety.”

Guilt over the drive for conquest and anxiety over the anticipated punishment are both assuaged in the child through repression of the forbidden wishes and the development of a superego to regulate its initiative. This conscience, the faculty of self-observation, self-regulation, and self-punishment, is an internalized version of parental and societal authority. Initially the conscience is harsh and uncompromising; however, it constitutes the foundation for the subsequent development of morality. Having renounced oedipal ambitions, the child begins to look outside the family for arenas in which it can compete with less conflict and guilt. This is the stage that highlights the child's expanding initiative and forms the basis for the subsequent development of realistic ambition and the virtue of *purpose*. As Erikson noted in *Childhood and Society*, “[T]he ‘oedipal' stage... sets the direction toward the possible and the tangible which permits the dreams of early childhood to be attached to the goals of an active adult life.” Toward this end, social institutions provide the youngster with an economic ethos in the form of adult heroes who begin to take the place of their storybook counterparts.

When there has been an inadequate resolution of the conflict between initiative and guilt the individual may ultimately develop a conversion disorder, inhibition, or phobia. Those who overcompensate for the conflict by driving themselves too hard may experience so much stress as to produce psychosomatic symptoms.

Industry Versus Inferiority (About 5 Years to About 13 Years) With the onset of latency, the child discovers the pleasures of production. He or she develops industry by learning new skills and takes pride in the things made. Erikson wrote in *Childhood and Society* that the child's “ego boundaries include his tools and skills: the work principle... teaches him the pleasure of work completion by steady attention and persevering diligence.” Across cultures, this is a time when the child receives systematic instruction and learns the fundamentals of technology as they pertain to the use of basic utensils and tools. As children work they identify with their teachers and imagine themselves in various occupational roles.

If the child is unprepared for this stage of psychosocial development, either through insufficient resolution of previous stages or by current interference, the child may develop a sense of inferiority and inadequacy. In the form of teachers and other role models, society becomes crucially important in the child's ability to overcome that sense of inferiority and to achieve the virtue known as *competence*. In *Identity: Youth and Crisis*, Erikson noted: “[T]his is socially a most decisive stage. Since industry involves doing things beside and with others, a first sense of division of labor and of differential opportunity, that is, a sense of the technological ethos of a culture, develops at this time.”

The pathological outcome of a poorly navigated stage of industry versus inferiority is less well defined than in previous stages, but it may concern the emergence of a conformist immersion into the world of production in which creativity is stifled and identity is subsumed under the worker's role.

Identity Versus Role Confusion (About 13 Years to About 21 Years) With the onset of puberty and its myriad social and physiological changes, the adolescent becomes preoccupied with the question of identity. Erikson noted in *Childhood and Society* that youth are now “primarily concerned with what they appear to be in the eyes of others as compared to what they feel they are, and with the question of how to connect the roles and skills cultivated earlier with the occupational prototypes of the day.” Childhood roles and fantasies are no longer appropriate, yet the adolescent is far from equipped to become an adult. In *Childhood and Society* Erikson writes that the integration that occurs in the formation of ego identity encompasses far more than the summation of childhood identifications. “It is the accrued experience of the ego's ability to integrate these identifications with the vicissitudes of the libido, with the aptitudes developed out of endowment, and with the opportunities offered in social roles.”

The formation of cliques and the intolerance of individual differences are ways in which the young person attempts to ward off a sense of identity loss. Falling in love, a process by which the adolescent may clarify a sense of identity by projecting a diffused self-image onto the partner and seeing it gradually assume a more distinctive shape, and an overidentification with idealized figures are means by which the adolescent seeks self-definition. With the attainment of a more sharply focused identity, the youth develops the virtue of *fidelity*—a faithfulness not only to the nascent self-definition but to an ideology that provides a version of self-in-world. As Erikson, Joan Erikson, and Helen Kivnick wrote in *Vital Involvement in Old Age:* “[F]idelity is the ability to sustain loyalties freely pledged in spite of the inevitable contradictions of value systems. It is the cornerstone of identity and receives inspiration from confirming ideologies and affirming companionships.” Role confusion ensues when the youth is unable to formulate a sense of identity and belonging. Erikson held that delinquency, gender-related identity disorders, and borderline psychotic episodes can result from such confusion.

Intimacy Versus Isolation (About 21 Years to About 40 Years) Freud's famous response to the question of what a normal person should be able to do well, “Lieben und arbeiten” (to love and to work), is one that Erikson often cited in his discussion of this psychosocial stage, and it emphasizes the importance he placed on the virtue of *love* within a balanced identity. Erikson asserted in *Identity:Youth and Crisis* that Freud's use of the term *love* referred to “the generosity of intimacy as well as genital love; when he said love and work, he meant a general work productiveness which would not preoccupy the individual to the extent that he might lose his right or capacity to be a sexual and a loving being.”

Intimacy in the young adult is closely tied to fidelity; it is the ability to make and honor commitments to concrete affiliations and partnerships even when that requires sacrifice and compromise. The person who cannot tolerate the fear of ego loss arising out of experiences of self-abandonment (e.g., sexual orgasm, moments of intensity in friendships, aggression, inspiration, and intuition) is apt to become deeply isolated and self-absorbed. *Distantiation*, an awkward term coined by Erikson to mean “the readiness to repudiate, isolate, and, if necessary, destroy those forces and people whose essence seems dangerous to one's own,” is the pathological outcome of conflicts surrounding intimacy and, in the absence of an ethical sense where intimate, competitive, and combative relationships are differentiated, forms

the basis for various forms of prejudice, persecution, and psychopathology.

Erikson's separation of the psychosocial task of achieving identity from that of achieving intimacy, and his assertion that substantial progress on the former task must precede development on the latter have engendered much criticism and debate. Critics have argued that Erikson's emphasis on separation and occupationally based identity formation fails to take into account the importance for women of continued attachment and the formation of an identity based on relationships.

Generativity Versus Stagnation (About 40 Years to About 60 Years) Erikson asserted in *Identity: Youth and Crisis* that "[g]enerativity... is primarily the concern for establishing and guiding the next generation." The term *generativity* applies not so much to rearing and teaching one's offspring as to a protective concern for all the generations and for social institutions. It encompasses productivity and creativity as well. Having previously achieved the capacity to form intimate relationships, the person now broadens the investment of ego and libidinal energy to include groups, organizations, and society. *Care* is the virtue that coalesces at this stage. In *Childhood and Society* Erikson emphasized the importance to the mature person of feeling needed. "[M]aturity needs guidance as well as encouragement from what has been produced and must be taken care of." Through generative behavior the individual is able to pass on knowledge and skills while obtaining a measure of satisfaction in having achieved a role with senior authority and responsibility in the tribe.

When people are unable to develop true generativity, they may settle for pseudo-engagement in occupation. Often such people restrict their focus to the technical aspects of their roles at which they may be now have become highly skilled, eschewing larger responsibility for the organization or profession. This failure of generativity can lead to profound personal stagnation, masked by a variety of escapisms, such as alcohol and drug abuse, and sexual and other infidelities. Mid-life crisis or premature invalidism (physical and psychological) may occur. In this case, pathology appears not only in middle-aged persons but also in the organizations that depend upon them for leadership. Thus, the failure to develop at mid-life can lead to sick, withered, or destructive organizations that spread the effects of failed generativity throughout society; examples of such failures have become so common as to constitute a defining feature of modernity.

Integrity Versus Despair (About 60 Years to Death) In *Identity: Youth and Crisis*, Erikson defined integrity as "the acceptance of one's one and only life cycle and of the people who have become significant to it as something that had to be and that, by necessity, permitted of no substitutions." From the vantage point of this stage of psychosocial development, the individual relinquishes the wish that important people in his life had been different and is able to love in a more meaningful way—one that reflects an acceptance of responsibility for one's own life. The individual in possession of the virtue of *wisdom* and a sense of integrity has room to tolerate the proximity of death and to achieve what Erikson termed in *Identity: Youth and Crisis* a "detached yet active concern with life."

Erikson underlined the social context for this final stage of growth. In *Childhood and Society*, he wrote, "The style of integrity developed by his culture or civilization thus becomes the 'patrimony' of his soul... In such final consolidation, death loses its sting."

When the attempt to attain integrity has failed, the individual may become deeply disgusted with the external world, and contemptuous of persons as well as institutions. Erikson wrote in *Childhood and Society* that such disgust masks a fear of death and a sense of despair that "time is now short, too short for the attempt to start another life and to try out alternate roads to integrity." Looking back on the Eight Ages of Man, he noted the relation between adult integrity and infantile trust, "[H]ealthy children will not fear life if their elders have integrity enough not to fear death."

Implications for Psychotherapists Erikson's belief that psychosocial development continues throughout the life cycle led him to caution psychotherapists against a theoretical reductionism in which symptomatic behavior is viewed as being the result of a traumatic past and in which insufficient attention is paid to contemporary conflicts in the ego's relations with the world. A psychotherapy informed by Eriksonian principles emphasizes not only interpretations based on reconstruction, but also an analysis of the patient's struggles with current developmental tasks. When patients' resolutions of previous psychosocial stages have been so faulty as to seriously compromise their adult development, they have the opportunity to rework early development through the relationship with the therapist. For example, the psychotherapeutic treatment of the individual who suffers a schizophrenic break in young adulthood initially revolves around the painstaking establishment of a sense of basic trust between the patient and the therapist. However, the treatment eventually has to consider contemporary obstacles to building a sense of identity in work and intimacy in relations with others.

Psychotherapy is an intervention, sometimes an entire chapter, in the ongoing biography of the patient. The treatment uses new developmental energies to redirect the patient whose psychosocial development has been blocked. By dint of temperament, education, and conscientious self-analysis the therapist may be prepared to identify and promote a variety of directions that subsequent development may take, avoiding a countertransference rigidity that could further inhibit the patient's growth. As Erikson noted in *Identity: Youth and Crisis*, the object of psychotherapy is not to head off future conflict but to assist the patient in emerging from each crisis "with an increased sense of inner unity, with an increase of good judgment, and an increase in the capacity 'to do well' according to his own standards and to the standards of those who are significant to him."

Becoming a Biographer of Youth In 1949, as *Childhood and Society* was going to press, Erikson's stand on the University of California loyalty oath had rendered his position at the university untenable. News of his availability spread east, and other institutions vied to capitalize on Berkeley's mistake. Erikson was well known at the Menninger Clinic in Topeka, Kansas, where he had lectured, so when Robert Knight left Menninger to become the director of the Austen Riggs Center in Stockbridge, Massachusetts, he took Erikson with him. The Austen Riggs Center was devoted to psychoanalytic research and to the treatment of severely disturbed adolescents and young adults—Erikson had been happily adopted once again.

Erikson stayed at Austen Riggs from 1950 to 1960, from age 48 to 58, when he returned to the faculty of Harvard University. During those years, Erikson completed the transition to biographer, forming more fully the approach foreshadowed in his incomplete essays on Hitler, Gorky, and George Bernard Shaw. The transition was not easy; Coles describes that time as constituting a second "identity crisis." Having completed the mid-life transition and having left the era of early adulthood, Erikson faced the difficult developmental task in his late 40s and early 50s of forming and revising a life structure for middle age. He had been a clinician and a theorist of youth; he was now old enough to enact the program promised in *Childhood and Society* and to become a biographer and theorist of the whole life cycle. However, before he wrote *Young Man Luther*, he once again returned to Freud and the origins of his own professional identity. The way to the future was through the past.

In his early 50s Erikson wrote three biographical essays on Freud: "The Dream Specimen of Psychoanalysis," "Freud's 'The Origins of Psychoanalysis,'" and "The First Psychoanalyst." These essays all concern themselves with the crisis Freud suffered during his own mid-life transition as he struggled to leave neuropathology and to define a new professional identity as a psychoanalyst. Strengthened by that reexamination of Freud's successful transition (and perhaps having reassured himself of Freud's blessing), Erikson began writing one of the genuine masterworks of psychoanalysis, *Young Man Luther*.

As always, Erikson's clinical work and observations enriched his theorizing about the life cycle. He acknowledged his debt to Austen Riggs and the Western Psychiatric Institute at the University of Pittsburgh in his preface to *Young Man Luther*. His work there had allowed him to "study the afflictions of young patients as variations on one theme, namely, a *life crisis*, aggravated in patients, yet in some form normal for all youth. I could identify those acute *life tasks* that would bring young people to a state of tension in which some would become patients."

However, Erikson would neither make Martin Luther a patient nor reify the psychopathology of the young people who were his patients. Instead, he asserted that comparisons between Martin Luther and his patients were "not restricted to psychiatric diagnosis... but... oriented toward those moments when young patients, like young beings anywhere, prove resourceful and insightful beyond all professional and personal expectation. We will concentrate on the powers of recovery inherent in the young ego."

Working with the young patients at Riggs helped Erikson hone his understanding of identity formation. As he later put it in "The Problem of Ego Identity," [I]dentity... is gradually established by successive ego syntheses and re-syntheses throughout childhood. It is a configuration gradually integrating constitutional givens, idiosyncractic libidinal needs, favored capacities, significant identifications, effective defenses, successful subliminations, and consistent roles."

These "ego syntheses and re-syntheses" were about to receive great attention as Erikson attempted his biography of Luther. But to do that work, Erikson the biographer had to move further away from his own professional origins as a child psychoanalyst. He had criticized psychoanalysis for its *originology*, which he defined as the belief that something has been explained by finding an analogy to its earliest manifestations. Yet his own conception of the life cycle, with most of its stages occurring within the pre-adult era, left him still heavily rooted in childhood. In his study of Luther, Erikson was trying to understand a *life*, not just a personality; from that perspective, childhood and adolescence had to be introductory rather than the story itself.

Seeing that a child-centered view was not adequate for the tasks of biography, without either formally changing the imbalance in his theory or neglecting childhood

determinants, Erikson deftly devoted the greatest attention to an explication of development in the adult years. So successful was he in interrelating Luther's adult problems with those of his early development that *Young Man Luther* provided the seminal inspiration for the next generation of life-cycle investigators. Some of those investigators formally redressed in their own theories the originological bias vestigial in Erikson's theories.

In Luther's early or mid-20s, according to some of his contemporaries, the young monk had fallen to the floor of the choir of his monastery in a fit and shouted, “Ich bin's nit!” or “Non Sum!” “It isn't me,” or “I am not.” Erikson's task as a psychological biographer was to explain how Luther got from the identity crisis of his 20s to nailing his 95 theses to the door of the church in Wittenberg at the age of 32. Yet *Young Man Luther* is also, as it is subtitled, “a study in psychoanalysis and history.” Since it is not a psychoanalysis of history, some nonreductionist connecting concepts between the individual and the collectivity were needed. One of Erikson's connecting concepts was *ideology*, which he defined as “an unconscious tendency underlying religious and scientific as well as political thought: the tendency at a given time to make facts amenable to ideas, and ideas to facts, in order to create a world image convincing enough to support the collective and the individual sense of identity.”

Luther had personal and psychiatric problems, but in his defiance of Catholic orthodoxy he created a new religious ideology that not only supported his own identity but the identities of emerging generations of Europeans as well. “Ich bin's nit” expressed the young monk's identity crisis as he found himself “fatally over-committed” to what he was not, a young man authentically embarking on a future as an orthodox Catholic priest. “In some periods of history,” Erikson asserted, “and in some phases of his life cycle, man needs... a new ideological orientation as surely and as sorely as he must have air and food.” As such a man, Luther commanded Erikson's “sympathy and empathy” as he “faced the problems of human *existence* in the most forward terms of his era.”

The reception of the book cemented Erikson's international reputation, initially won by *Childhood and Society*, as the leader in original, humane thought about psychological development. In no other biography have the dynamics of individual conflict and development been so seamlessly interwoven with those of society and history. With *Young Man Luther*, Erikson in his late 50s ascended from the first rank of developmental psychoanalysts to that of a cultural seer.

Becoming a Biographer of Middle Age As Erikson moved through the transition from middle to old age he abandoned clinical work with patients to turn his attention fully to the study of the life cycle and the survival of the species. His emphasis, emergent in *Young Man Luther* and implicit in all his earlier work, was on the ego virtues that permit a person to live in a constructively critical relationship with the social institutions of his time. Earlier, Erikson had described the stages in the development of the ego across the life cycle. In 1964 he wrote more about the ego virtues in *Insight and Responsibility*. He was getting ready amidst the American moral and political crises of the mid-1960s to embark on his last major work, a study of the great moral leader Mohandas Gandhi. In the end, he dedicated it, *in memoriam*, to Martin Luther King, Jr.

Young Man Luther had been a story of personal choice and historical change wrought by a young man establishing and revising a first adult identity and life structure. In his mid-60s Erikson had gained sufficient perspective on the life cycle to analyze a case of decisive crisis and change in a man who was going through a mid-life transition and solving it by creating a vital life structure for middle age. Gandhi did that, as he himself put it, by realizing “his vocation in life,” leading *satyagraha* truth force campaigns in his nation, in his family, and in his own soul. Gandhi was entering the stage of the life cycle in which Erikson's ego polarity of *generativity versus stagnation* becomes active; from his developmental work on this polarity, Gandhi was learning the ego virtue of *care*.

Erikson saw that Gandhi, like Luther, was both a leader and a “religious actualist.” As a leader, he succeeded in articulating inner concerns in a way that struck a collective cord. Self-rule and home rule were inextricably combined in Gandhi's program. He would have endorsed in his own terms, Erikson believed, Luther's assertions, “Christ comes today; God's way is what makes us move; we must always be reborn, renewed, regenerated; to do enough means nothing else than always to begin again.”

Gandhi had begun and successfully completed his political novitiate as an Oxford-trained barrister in South Africa during his 20s and 30s. By his late 30s, he had developed passive resistance as a political strategy to defeat the Black Act, a law that required all Indians in the Transvaal to register with the government and to carry identification papers on their persons. At 40, Gandhi staked out his claim for leadership in his Indian Home Rule Manifesto, and during the next few years of his midlife transition he developed the device of the Satyagraha campaign, culminating in the Great March of striking mine workers in South Africa when he was 44.

When Gandhi returned to India in 1915 at the age of 45, he did so, Erikson wrote, “like a man who knew the nature and the extent of India's calamity and that of his own fundamental mission.” As a mature, middle-aged man, he was, in Erikson's view, a person who has determined what “he does and does not *care for*” as well as “what he *will* and *can* take *care of*. He takes as his baseline what he irreducibly *is* and reaches out for what only he can, and therefore, *must* do.”

What Gandhi had to do was lead a labor strike at the mill in Ahmedabad and the next year, at age 50, a national strike for independence from Great Britain. Erikson contended that Gandhi's emergence at that time as the “father of his country” underlined “the fact that the middle span of life is under the dominance of the universal human need and strength which I have come to subsume under the term *generativity*.” Erikson refers here not merely to the generativity that a parent at 20 or 30 may feel for a child. Erikson's concept of generativity in often mistakenly taken to apply to the child-rearing work of young adults. Instead, he is describing the protective concern for the generations and their retarding-facilitating social institutions, whose leadership is the essential responsibility of the middle-aged individual.

In the early 1970s Erikson returned to the San Francisco Bay Area to live and begin again. The Eriksons relocated to Cambridge, Massachusetts, some 10 years later. In 1981 the Eriksons and psychologist Helen Kivnick undertook an intensive study of 29 octogenarian parents of children whose lives had been meticulously scrutinized in the Guidance Study begun in 1928 at the Institute of Human Development at the University of California at Berkeley. The results of their research were published in *Vital Involvement in Old Age*.

In this work, Erikson, Erikson, and Kivnick revisit the eight stages of psychosocial development and their attendant virtues or strengths. They emphasize the opportunities that exist at the end of life to integrate “maturing forms of hope, will, purpose, competence, fidelity, love, and care into a comprehensive sense of wisdom.” A successful integration results from a final reworking of the earlier polarities of basic trust versus mistrust; autonomy versus shame and doubt; initiative versus guilt; industry versus inferiority; identity versus confusion; intimacy versus isolation; generativity versus self-absorption; and integrity versus despair. The authors use the term *vital involvement* to identify an involvement with the environment characterized by “actuality” and “mutuality.” The elder who is able to maintain a vital involvement with his world—no matter how small its scope—is better able to tolerate the depredations of aging: the loss of sensory acuity, physical mobility, and stamina; changes in intimate relationships as children marry and move away, and friends and partners sicken and die; the loss of social status, financial security, and sources of pride in professional competence; and the realization that relatively little time remains.

Developed in infancy, hope is the basis upon which the senescent individual attempts to integrate faith in the universe and the relative predictability of its laws, with a realistic mistrust about that which is unpredictable and unreliable. The elder may experience a need to affirm a basic faith—in religion or nature—and to understand more deeply his own place in a generational progression.

Late in life the individual must struggle to redefine a sphere of autonomy even as he becomes increasingly dependent upon others. Living independently, residing in familiar surroundings, and following established routines can be crucial symbols of autonomy and the means by which signs of impairment (e.g., a walker) may be accepted without undue shame. An individual's lifelong experience of autonomy influences his or her capacity to adjust to restrictions upon freedom. The ability to allow others to help supports the generative impulses of younger adults and reinforces for both the sense of a generational cycle.

The polarities of initiative versus guilt, first evident in the youngster's exploration and play and subsequently in adult work and recreation, must be balanced anew in old age. With the passage of time the person faces a diminution of opportunities for the exercise of initiative, and must regret those instances where he or she failed to act decisively or with sufficient concern for others. Successful adaptation to the limitations of old age must draw upon the strength of purposefulness developed in childhood and tempered by maturity.

With the natural decline in physical strength and sensory acuity, the senescent individual increasingly depends upon a lifelong sense of effectiveness. External rewards are usually not as numerous as they were in young adulthood or middle age. The feeling of competence and mastery must be sustained internally. An old person may cope with growing physical limitations by changing the criteria for a sense of accomplishment, pursuing new hobbies or modifying old ones, developing new avenues of involvement with former professions, or living vicariously through the accomplishments of children and grandchildren.

Old age provides the individual with the chance to review a lifetime of beliefs, to come to terms with choices made and opportunities lost, and to make and act upon any final commitments, which now most clearly reflect what Erikson termed “the ‘I' in the totality of life.” In striking a final balance between identity and identity confusion, the elder concerns himself with both an external image and a personal image. The external image is the way the individual may expect to be remembered;

the personal image reconciles the sense of who he or she has been with an evolving sense of who he or she may yet become. Identification with the successes of younger generations and satisfaction in one's own part in the transmission of values helps the senescent to stave off feelings of insignificance or obsolescence in the face of technological and sociological changes.

The sense of love that can emerge at the end of life is one that is built upon a foundation of love, expressed and unexpressed, throughout the life cycle. A new balance may be forged between intimacy and isolation in the face of decline. Intimate relationships may be recast in more positive terms; reminiscences provide the elder with the opportunity to draw upon and reintegrate earlier experiences of tenderness and sexuality.

In order to reconcile feelings of integrity and despair, the individual must make peace with previous choices, acknowledging what was gained and lost and accepting that it is too late to change. Relationships and values may be reviewed with more room to consider alternate points of view. In balancing the dispositions toward generativity and stagnation, the elderly person may be able to build upon years of experience as a parent, worker, and creatively productive person from middle age, which themselves were shaped by childhood experiences of caring or its absence. Grandparenthood offers another chance at generativity and a way to undertake it more robustly and less ambivalently. Erikson coined the term “grand-generativity” to distinguish the generativity of old age in which the person, freed from the direct responsibility for family, institution, and community, is better able to incorporate care for the present with concern for the future. Such concern may be expressed through advice or financial assistance to family members or on a larger scale, through commitment to religious beliefs, political involvement, or community action. Grand-generativity contributes to a feeling of immortality: in caring for the younger generation and allowing oneself to be taken care of, the senescent secures a personal place in a generational history.

The “joint reflections on old age” contained in the book are an effort to elucidate the psychosocial process of vital involvement and to extend Erikson's observations to the outer limits of the life cycle. The book also brought Erikson back full circle to his professional beginnings—having made almost 40 years earlier, play observations on the Guidance Study children.

SUGGESTED CROSS-REFERENCES

Sigmund Freud's theories are discussed most fully in Section 6.1. Other theories of personality and psychopathology are discussed in Chapter 6.3 and Chapter 6.4. Schizophrenia is discussed in Chapter 12, personality disorders in Chapter 24, and psychosomatic disorders in Chapter 25. Normal child development and adolescent development are discussed in Section 32.2 and Section 32.3, respectively, adulthood is discussed in Chapter 50, normal human sexuality in Section 19.1, and normal aging in Section 51.2c. Psychoanalysis and psychoanalytic psychotherapy are discussed in Section 30.1. Another perspective on Erikson's work is given in Section 30.9.

SECTION REFERENCES

*Coles R: *Erik H. Erikson: The Growth of His Work*. Little, Brown, Boston, 1970.

Erikson E: *Childhood and Society*. Norton, New York, 1950.

Erikson E: The dream specimen of psychoanalysis. J Am Psychoanal Assoc *2:*5, 1954.

Erikson E: The first psychoanalyst. Yale Rev *46:*40, 1956.

Erikson E: Freud's “The Origins of Psychoanalysis.” Int J Psychoanal *36:*1, 1955.

*Erikson E: *Gandhi's Truth*. Norton, New York, 1969.

Erikson E: Hitler's imagery and German youth. Psychiatry *5:*475, 1942.

Erikson E: *Identity and the Life Cycle*. Norton, New York, 1980.

Erikson E: *Identity: Youth and Crisis*. Norton, New York, 1968.

Erikson E: *Insight and Responsibility*. Norton, New York, 1964.

Erikson E: *Life History and the Historical Moment*. Norton, New York, 1975.

Erikson E: Observations on Sioux education. J Psychol *7:*101, 1939.

Erikson E: The problem of ego identity. Psychol Issues *1:*379, 1959.

*Erikson E: *Young Man Luther*. Norton, New York, 1962.

Erikson E, Erikson J, Kivnick H: *Vital Involvement in Old Age*. Norton, New York, 1986.

Evans R: *Dialogue with Erik Erikson*. Harper & Row, New York, 1967.

Freud A: *The Ego and the Mechanisms of Defense*. International Universities Press, New York, 1966.

Freud S: Beyond the pleasure principle. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 18. Hogarth Press, London, 1955.

Freud S: Civilization and its discontents. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 21. Hogarth Press, London, 1961.

Freud S: The ego and the id. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 19. Hogarth Press, London, 1961.

Freud S: The future of an illusion. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 20. Hogarth Press, London, 1961.

Freud S: Group psychology and the analysis of the ego. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 18. Hogarth Press, London, 1955.

*Friedman L: *Identity's Architect: A Biography of Erik Erikson*. Scribner, New York, 1999.

Gilligan C: Woman's place in man's life cycle. Harv Educ Rev *49:*431, 1979.

Howenstine R, Silberstein L, Newton D, Newton P: Revitalizing the life structure: An adult developmental approach to psychodynamic psychotherapy. Psychiatry *55:*194, 1992.

Kivnick H: Through the life cycle: Psychosocial thoughts on old age. In *The Course of Life: Completing the Journey*, vol 7, G Pollock, S Greenspan, editors. International Universities Press, Madison, CT, 1998.

*Levinson D, Darrow C, Klein E, Levinson M, McKee B: *The Seasons of a Man's Life*. Knopf, New York, 1978.

Levinson D, Levinson J: *The Seasons of a Woman's Life*. Knopf, New York, 1995.

McAdams D, de St. Aubin E, editors: *Generativity and Adult Development: How and Why We Care for the Next Generation*. American Psychological Association, Washington, DC, 1998.

Murray H: *Explorations in Personality: A Clinical and Experimental Study of 50 Men of College Age*. Oxford University Press, New York, 1938.

Newton P: *Freud: From Youthful Dream to Mid-Life Crisis*. Guilford Press, New York, 1995.

Newton P: Samuel Johnson's breakdown and recovery in middle-age: A lifespan developmental approach to mental illness and its cure. Int J Psychoanal *11:*93, 1984.

Newton P: Science and humanism in the biographical study of lives. Contemp Psychiatry *7:*77, 1988.

Newton P, Levinson D: Crises in adult development. In *Outpatient Psychiatry*, A Lazare, editor. Williams & Wilkins, Baltimore, 1979.

Schein S, editor: *Erik Erikson: A Way of Looking at Things*. Norton, New York, 1987.

Wallerstein R, Goldberger L, editors: *Ideas and Identities: The Life and Work of Erik Erikson*. International Universities Press, Madison, CT, 1998.

6.3 OTHER PSYCHODYNAMIC SCHOOLS

PAUL C. MOHL, M.D. AND MYRON F. WEINER, M.D.

The 11 men and women presented in this chapter—Adolf Meyer, Alfred Adler, Carl G. Jung, Sandor Rado, Otto Rank, Karen Horney, Franz Alexander, Harry Stack Sullivan, Wilhelm Reich, Erich Fromm, and Eric Berne—made their contributions in the early and middle years of the twentieth century. At that time with knowledge of the neurosciences still very primitive, the ethos of psychiatry was to use clinical observation to search for an overarching theory that would encompass not only psychopathology, but all of human behavior. None of these individuals doubted that mind has a physical basis but they all were aware of Sigmund Freud's woeful failure to develop a theory of how experience, thoughts, and feelings are transduced to and from material states. Because of this, they were forced to stand back from the physical organism and view humans strictly as psychological entities. Many doubted that biology would ever be capable of explaining human experience and respected philosophical positions of the time buttressed their opinion. Some neurologists, psychiatrists, psychologists, cognitive scientists, and computer scientists still adhere to this position.

For reasons that are complicated and not always clear, despite the availability of improving phenomenologically based nosological systems, the bias of all psychological theories of personality and psychopathology was to blur the distinction between diagnostic entities and emphasize the common elements of different psychopathologies. Carl Jung stated that with each new patient he had to abandon all theory and listen afresh. Reductionistic neurobiology was considered almost inhumane in its efforts to explain the complexities of human behavior and experience. The understanding of what it is to be human and humane has evolved, as has the understanding of what is and is not scientific. The scientific method involves observation, hypothesis-formation, hypothesis-testing, and coming to conclusions and all these individuals were scientists in that sense. They approached their clinical experience with open-minded rigor and discipline. The proof of their theories was that they fit the observations. These individuals were skeptical observers who were willing to adjust their theories if the observations failed to confirm them; their major flaw is that they overexplained mental phenomena in psychological terms.

This section addresses the contributions of 11 individuals to the present view of normal personality development, psychopathology, and the treatment of mental illnesses and minor emotional disorders. Their greatest contribution is their effort to make mental and emotional disturbances understandable as psychological phenomena. They were able to do so in large measure because of their involvement with psychoanalysis, whose basic credo is that there are laws and principles that govern mental phenomena just as there are laws that describe the physical world. Each theorist anticipated certain aspects of modern psychoanalytic theory (e.g., object relations, self psychology). The primacy of the instincts and of the Oedipal conflict in Freud's thought left many of these thinkers with the sense that something was missing. Their solutions and explanations often leave the modern psychiatrist with the very same sense. However, their struggles toward theories based on the relevance of the interpersonal world, self-esteem, and the first year of life contributed enormously to the broadening focus of psychodynamic theorizing.

All these theorists except Meyer had had experience with or trained in psychoanalysis. Alfred Adler, Fromm, Horney, Jung, Rank, Reich, Rado and Alexander had strong personal involvements with Freud. Jung was Freud's first apparent heir. Except for Rank and Fromm, all were physicians. They all had unique, strong personalities that led to schools of therapy developing around them, partly of their own devising and partly based on the needs of their followers. The best-known schools are Jung's analytic psychology, Adler's individual psychology, and Berne's transactional analysis. These theories developed and the treatment techniques based on them are widely divergent, yet each theory often began with a critique of some aspect of Freud's theories and then built a psychological edifice to correct his errors. Often, the shared foundations were not acknowledged.

The divergence of these theories and techniques and their success as explanations and treatments suggest that each is useful in some contexts and that none is applicable in every context. In attempting to employ any or all of these theories or techniques it is necessary to ask, as psychotherapy research now asks, what theory or technique is useful with what patient under what circumstance and in the hands of what therapist. In order to illustrate the different understandings and therapeutic approaches of each theorist, the following case will be used as a basis for formulation according to each theorist.

Mr. A. was a 26-year-old white male who had a history of bipolar I disorder. He was brought in for treatment after failing to complete the last required course for his advanced degree, and being arrested for disturbing the peace. He had consistently lied to his family about where he stood with his course work and about having skipped an examination that would have qualified him to use his professional degree. He had also not told them that he had been using marijuana almost daily for a number of years and occasionally used hallucinogens. His arrest for disorderly conduct had been for swimming naked in an apartment complex in the middle of the night, while under the influence of hallucinogens.

Mr. A.'s use of marijuana had begun early in college, but became daily during graduate school. He was diagnosed as having bipolar I disorder early in his senior year at college after a clear episode of mania. His mood disorder was well controlled on lithium (Eskalith). During graduate school he was episodically compliant with medications, preferring to try to maintain a state of hypomania. He saw a psychiatrist every 3 to 6 months for medication checks. During his 4 years in graduate school he had two clear episodes of depression and was begun on sertraline (Zoloft), 100 mg a day with questionable benefit. Mr. A. believed that he could be a great writer. He spent most of his time reading and trying to write. He dreamed of going to New York city and becoming part of a group of avant-garde writers that would parallel the Algonquin Club of the 1930s or the beat poets of the late 1940s. This aspiration and his marijuana abuse predated his development of bipolar I disorder. He attended class episodically, nonetheless performing adequately. His last class had no final examination, but required a paper. He planned to write this paper in the form of a play, involving a dialogue between two thinkers from different times and cultures. His professor was very excited about this idea, but Mr. A. kept postponing the task until he was forced to extend his schooling by a year. His other major interest during this time involved growing and photographing flowers.

Mr. A. was born and raised in a large city. His father had been very successful in commercial real estate, and his mother, after raising the children, used the substantial real estate holdings she inherited from her father to set up a business to manage them. Most of the money was placed in a trust for the patient and his siblings. His mother had total financial control of the trusts and doled out the proceeds to the children as they needed them. There was no family history of any psychiatric disorders.

The patient described his mother as very loving and caring, but to the point of being intrusive and controlling. For example, the mother arranged the initial treatment, but then was angry that the psychiatrist had not called her regularly to report on her adult son's progress. She was also critical of various aspects of the treatment as reported to her by her son. The patient's two older siblings had attended prestigious colleges and graduate schools, but had returned home to work in the mother's real estate management company. The 30-year-old sister was living in the parents' home. The 35-year-old brother had lived at home for a time, but then moved out to a location a few blocks away. There was a younger brother, still in college, who also smoked marijuana excessively. He had tried to minimize the patient's problems to the family, and tried to protect the patient, who desperately had not wanted to return home. Of note is that none of the children were married, although the two older ones had each had a couple of serious relationships.

The children seemed to regard the mother with affectionate amusement and bemusement. The father was seen as a very caring but not demonstrative man who put much energy into keeping the mother from becoming too upset and encouraged the children to do the same. The children often wanted to provoke the mother for her judgemental, detail-oriented intrusiveness. The father would discourage them, but occasionally found their provocations amusing.

The family viewed itself as very close, with strong values oriented towards community service and family loyalty. The family belonged to a religious community, but expressed their involvement primarily in social service and social action volunteer work, accompanied by very generous financial contributions.

The patient had been a very successful debater in high school and recalled his development as very positive, but provided few details. He tended to place himself in the role of the outsider, an observer of humanity, which he saw as consonant with the role of writer. He was proud to have bipolar I disorder, and tried to regulate his medications so that he would be hypomanic much of the time, seeing this as enhancing his creativity. He viewed his use of marijuana in the same vein. One of the most distressing aspects to him of his depressive episodes was that marijuana no longer created a feeling of well-being, but made him feel worse. His current depressive episode involved no neurovegetative symptoms. Rather, he presented as flat, numb, apathetic, ashamed, anhedonic, and anergic. He was particularly ashamed of being back in his hometown and of living with his parents.

The patient ostensibly understood and accepted his illness well and had read much about it. However, the family had responded to the information, “With proper treatment bipolars can live normal lives,” as meaning that the information should be kept secret so that he should be treated normally. Mr. A., on the other hand, was very open with friends at graduate school about his illness and his pride in it and the creativity he associated with it.

The patient had two longstanding recurrent dreams. One involved him flying. The narrative line varied but the flying theme recurred. Often he had other magical powers in his dreams such as the ability to heal, to not be killed by bullets, to save the world or some group of people from mortal danger, and so on. The other recurrent dream was of a hotel lobby. The dreams regularly began with him entering a hotel lobby to meet a group of people, accompanied by a feeling of dread.

ADOLF MEYER

Meyer (1866–1950) emigrated to the United States after having trained as a neuropathologist in Switzerland (Fig. 6.3-1). Not interested in metapsychology, he espoused a commonsense psychobiological methodology for the study of mental disorder, emphasizing the interrelationship of symptoms and individual psychological and biological functioning. His approach to the study of personality was biographical. He attempted to bring psychiatric patients and their treatment out of isolated state hospitals and into communities, and was also a strong advocate of social action for mental health. He began his career as a clinically oriented state hospital pathologist, becomming the second head of the Pathological Institute (later, the New York Psychiatric Institute, whose affiliation with the Columbia College of Physicians and Surgeons he created). He later became the President of the American Psychiatric Association and had a 32-year tenure as Chairman of Psychiatry at Johns Hopkins. His major social contribution was helping to found the National Committee on Mental Hygiene.

FIGURE 6.3-1 Adolf Meyer. (Courtesy of National Library of Medicine, Bethesda, MD.)

Psychobiology Despite his background as a neuropathologist, which was common among European psychiatrists of his generation, Meyer strongly opposed the Kraepelinian view of mental illness as having a predetermined course based on phenomenologically identified syndromes. Instead, he believed that individuals' habitual reaction patterns made them more susceptible to specific types of breakdown. Meyer used biographical study, strongly encouraging psychiatrists to take a thorough, detailed, life history of the patient to understand each individual's reaction patterns. He saw development as lifelong, thus viewing the data from adolescence onward as equally important as that from childhood in understanding the patient's narrative.

He observed reaction patterns, attempted to predict the conditions under which they might occur, and tested and validated methods for their modification. He acknowledged the contributions of Freud and Jung but thought they were too narrow. A thoroughgoing pragmatist, he preferred common sense to metapsychological constructs as the means to understand and deal with psychopathology.

Meyer believed that through a basic tendency toward integration, multiple biological, social and psychological forces contribute to personality development. The vulnerable person uses poorly planned, ill-suited means of adaptation. Meyer used the biographical approach as a practical guide to elicit information about personality development, to organize the information, and to check and reevaluate the information obtained under different circumstances. His clinical examination assessed each patient's life history; physical, neurological, genetic and social status; and the relationship between those factors and personality factors. A diagnosis and an individual treatment plan were based on this assessment.

Treatment The aim of psychobiological therapy was to help individuals make the best possible adaptation to changing environmental circumstances. It began with the development of a collaborative relationship. Out of the collaborative relationship came *distributive analysis*, an examination of the factors in patients' lives that contributed to their adjustment or lack thereof, and concluded with *distributive synthesis*; helping patients to understand themselves and to develop better coping skills. The first step in distributive analysis is the patient's own exposition of the presenting problem. Assets and liabilities are then determined by eliciting the life history in terms of the memories that are immediately available and those that are later fleshed out by reconstructing past experiences.

Treatment is initiated by focusing on patients' assets. It involves psychological, chemical, physical, and environmental measures as needed. In more severe cases, attention is paid first to patients' sleep habits, nutrition, and daily routines because these must be normalized before psychological work can be done. Patients are helped to describe their difficulties in detail. In addition to eliciting complaints or worries, patients are asked what eases or worsens their complaints and what significance they attach to their symptoms and concerns. In doing this, the therapist attempts to use the patient's own language and concepts to communicate suggestions and advice.

Meyer did not pay attention to unconscious mechanisms, but focused on patients' functioning in reality. Both present-day and long-term adaptive patterns are considered. Therapeutic sessions proceeded from immediate, obvious problems in the present to longer-term issues and historical data. With guidance, patients investigate their own personality problems, ascertain the origin of their conflicts and work to develop more useful behavior patterns. Meyer called this *habit training*, a term he may or may not have borrowed from behavioral tradition. When unhealthy adaptive patterns are modified, proper adjustment and personal satisfaction result.

Meyerians would focus first on the adequate treatment of Mr. A.'s mental illness. Because he is still dependent on his parents, they would be included in the treatment plan and might be seen separately by a social worker. Both Mr. A. and his parents would be told that failure to control the symptoms of his mental illness could cost him his life and any satisfaction that he might derive from it. Mr. A.'s mood swings would be stabilized on an appropriate medication and he would be observed closely for restoration of normal sleeping and eating habits. He would be asked to discontinue his habit of marijuana use because of its tendency to aggravate the symptoms of his mental illness and its clouding of the assessment of treatment efficacy. Random urine screening might be recommended. Mr. A. would be helped to develop a daily schedule including time for work, social interaction, and recreation. He would not be encouraged to return to school because of his tendency to use school as an escape from responsibility. He would be encouraged to develop his photography as a hobby after a period of stable work performance. Later, he would begin the distributive analysis phase of his treatment, and would be asked to reflect on the impact of his bipolar I disorder and his avoidance of responsibility on his life. In the distributive synthesis phase of treatment, he would be helped to realize that his habit of using his hypomania to heighten his enjoyment and his avoidance of responsibility would keep him from achieving independence and the pleasure of attaining any realistic goal. He would be asked to develop a plan for achieving both appropriate independence and for remaining within the family structure. Mr. A.'s parents would be seen periodically for progress reports and to urge them to push Mr. A. toward appropriate self-care, toward making use of his own resources. They might be asked to place Mr. A. on an allowance and to hold him to it strictly to reinforce his accountability for his actions. Over time, his parents would be asked to slowly withdraw financial support and to give Mr. A. more independence. They might be asked to place his inheritance in a trust designed such that he could not impulsively misspend it.

ALFRED ADLER

Adler (1870–1937) was born in Vienna and spent most of his life there (Fig. 6.3-2). A general physician, he became one of the original four members of Freud's circle in 1902. Adler never accepted the primacy of the libido theory, the sexual origin of neurosis, or the importance of infantile wishes. In 1911 he resigned as President of the Vienna Psychoanalytic Society and continued the development of his own socially and interpersonally focused theory of development. He posited a striving for self-esteem through overcoming a sense of inferiority, which he saw as an inevitable presence in the human condition as a result of our extended childhood. He equated psychological health with constructive social consciousness, developing a system that he called *individual psychology* that is still vigorous in many countries. His major social contribution was the establishment of child guidance centers in Vienna that served as a model for the rest of the world.

FIGURE 6.3-2 Alfred Adler (print includes signature). (Courtesy of Alexandra Adler.)

Personality Theory If Meyer's system were captured in a phrase, that phrase would be common sense. Adler's system might be described similarly as *Menschenkenntnis*, which is the concrete, practical knowledge of mankind. Adler saw individuals as unique, unified biological entities, all of whose psychological processes fit together into an individual lifestyle *(lebensstil)*. He also postulated a principle of dynamism, that every individual is future-directed and moves towards a goal. Once the goal is established, the psychic apparatus shapes itself toward attainment of that goal. Life goals are chosen, and are thus subject to change. Changes require modification of memories, dreams, and perceptions to fit the accomplishment of the new goal. Adler also emphasized the interface between individuals and their social environment: the primacy of action in the real world over fantasy. Community-mindedness, acceptance of the need to conform to the legitimate demands of society, is an important precept, but Adler also recognized a dialectic that occurs between individuals and their interpersonal environment, each constantly reacting to and shaping the other. Thus, Adler anticipated some of the modifications of Freud's theories introduced by Heinz Hartmann, as well as some of Sullivan's thinking.

Normal Personality and Adaptation The cornerstone of Adler's personality theory is the concept of moving from a sense of inferiority to a sense of mastery. Early in life everyone has a sense of inferiority resulting from realistic comparison with adults' size and abilities. Moving from this sense of inferiority to a sense of adequacy is the important motivational motif in life. Thus, the ideal person strives for superiority and does so through high social interest and activity; the emotionally handicapped person continues to feel inferior and reinforces that position through lack of striving and social interest.

There are many obstacles to the development of self-esteem and social interest. Prominent among them are poorly developed or “inferior” organs or systems (such as poor eyesight or poor eye-hand coordination), childhood diseases, pampering, and neglect. Physical handicaps and childhood diseases may promote self-centeredness and loss of social interest; birth order is another factor. According to Adler, first-born children, having lost their position of only child, tend not to share and become conservative. Second children change and become social activists. Youngest children feel secure because they have never been displaced. These theoretical thoughts and clinical observations by Adler anticipated recent psychological research on the importance of birth order in human behavior.

Theory of Psychopathology Emotional disorders result from mistaken lifestyles that are subject to change by will and by self-understanding. Individuals subject to emotional disorders have false ideas about themselves and the world and inappropriate goals that lead them away from constructive social interest. Individuals with a pampered lifestyle, for example, expect and demand from others, avoid responsibility, and blame others for their failures, but they feel incompetent and insecure because their well-being depends on pressing others into service. If life poses no challenge, a mistaken life style may have no consequences. When a mistaken lifestyle is ineffective, symptoms develop that protect self-esteem while helping the individual to avoid dealing realistically with the problem being confronted. The difference between *neurosis* and *psychosis*, is that neurotic individuals maintain social interest, but are blocked from life goals by symptoms, while psychotic

individuals lose social interest and retreat into their own world.

Psychotherapy Because his theory emphasized the mismatch of lifestyles with the demands of the real world, Adler focused on blocks to living productively in the real world, not on exploring unconscious conflict. His aim was to point out mistaken self-views and mistaken views of the world and then, by mobilizing will, make the needed changes, including a change in life goals.

Therapeutic Process Starting with three sessions per week and tapering off to once a week, a positive relationship with patients is established, used to lead patients to awareness of their lifestyle, how it is discordant with the demands of social reality, and ways in which it may be reoriented. Instead of striving for goals of no social value that falsely raise self-esteem, they are pushed to work toward ameliorating their own situation. Having become aware of the obstacles they have placed in their own path and of the consequences of these self-defeating behaviors, they are now helped to develop constructive interests in themselves and others. As they become less self-engaged, they find themselves better accepted by others, which reinforces their constructive efforts. Persons who have dedicated themselves to symbolically defeating others learn to cooperate and advance toward useful goals. Any endeavor in which patients can develop real competence is encouraged, whether social, work, artistic, or musical.

Patients are encouraged to remove the concrete obstacles to developing a useful lifestyle, including reading instruction for slow readers or contact lenses for persons self-conscious about their appearance. Early recollections, birth order, dreams, daydreams, and present-day interactions are all used to help patients see the inappropriateness or falseness of their ideas and life goals. Actual life events or memories of events are less important than individuals' reactions to those events or memories. Because memories are likely to be retrospective falsifications justifying an erroneous lifestyle, there is little need to verify them. There is also no need to look for latent content in dreams; they are merely an expression of present-day concerns. Nor is it necessary that therapists' interpretations be correct because they need only to help patients build a useful conception of themselves and the world. This perspective anticipated what has come to be called the *hermeneutic approach to psychotherapy*.

Several of Adler's techniques, including *reframing* and *paradoxical communication*; now enjoy wide popularity. Reframing is viewing the same data from a different point of view. Indecision, for example, was reframed from a product of mixed feelings to a wish to maintain the status quo. Failure to act keeps everything the same, which is the self-fulfilling prophecy of the discouraged person. Following this reframing statement, patients would be pushed to act constructively. Paradoxical communication is instructing patients to do the opposite of what the therapist wishes them to do. In dealing with an indecisive person, for example, Adler might caution against doing anything rash. Adler also paid attention to the impact of his patients on their environment and recognized that individuals do much to create their own interpersonal world. In response to complaints about being treated unfairly by others, Adler would ask patients how they dealt with the persons about whom they complained. Above all, Adler treated his patients as rational and as able to learn more productive ways of living. In his emphasis on practical, constructive solutions, misconstrued goals, and misperceived views of the world and the self, Adler also anticipated important elements of cognitive therapy.

As seen from the Adlerian point of view, Mr. A. has developed a mistaken lifestyle and an inappropriate life goal. He maintains himself in fantasy as a writer while failing at the accomplishments that would enable him to become a writer. Thus, he has been attempting to make the normal step of moving from inferiority to mastery in fantasy instead of through realistic achievement. Blocks to his development include pampering and his concomitant denial and abuse of his bipolar I disorder, the latter representing his organ inferiority. He has lost the social interest he had earlier in life and become extremely self-centered, unconcerned with others or with the consequences of his actions. He has used drugs and hypomania to avoid the pain of defeat. An Adlerian therapist would encourage Mr. A. to develop a realistic self-view. He has a mental illness that requires lifelong treatment. Grandiose dreams and intoxication with drugs cannot substitute for accomplishments in the real world and will always lead to defeat. He would be asked to set and to strive to accomplish small, realistic goals for himself, such as holding a steady job while enjoying photography as a hobby. His mental illness would be reframed as a challenge to his creativity and his use of marijuana as an obstacle to mobilizing his creativity. He might initially be encouraged to accept his dependency on his family and later, as he stabilizes, to reduce his dependency on his family as a means to heighten his self-esteem and to make a transition into adulthood. He would be encouraged to join the Depressive and Manic Depressive Association or other available support and educational groups as a means to better understand and accept his illness, to develop a social conscience, to stimulate his altruism. His dream about flying would be interpreted as his wish to achieve mastery, his dream about the hotel lobby would be recast his awareness that he has been trying to substitute fantasy for reality.

CARL JUNG

Jung (1875–1961) was a lifelong resident of Switzerland (Fig. 6.3-3). He trained in psychiatry under Eugen Bleuler at the Burgholzli Mental Hospital in Zurich and was strongly involved with Freud and the psychoanalytic movement from 1906 to 1914, when he resigned as president of the International Psychoanalytic Association. Following a “creative illness” that lasted from 1914 to 1918, Jung became an advocate of active introspection as the means to intrapsychic change. This episode and Jung's subsequent interpretation of it as well as that of his disciples has formed the focus of much of the recent controversy about him and his ideas. Although Jung rejected Freud's notion of libido as sexual energy and the Oedipus complex as a universal developmental stage, he believed not only in the unconscious mind, but in a shared racial and species unconscious. Jung, an intuitive introvert, was not interested in the practical aspects of living in the world; his focus was instead on individuation through becoming aware of the unconscious.

FIGURE 6.3-3 Carl Gustav Jung (print includes signature). (National Library of Medicine, Bethesda, MD.)

Personality Theory Jung developed an elaborate metapsychology that was every bit as detailed and formalized as Freud's. In addition, his theories most clearly identify most of the important fault lines in classical psychoanalysis. His construct of the psychic apparatus (Fig. 6.3-4) differed from the Freudian structure of ego, superego, id, and ego ideal. Below an outer rim of consciousness is the personal unconscious, which contains the complexes. Contained within the personal unconscious and connected to the complexes are the archetypes, the elements of the self, which in turn connect to the surface of the personality as the ego.

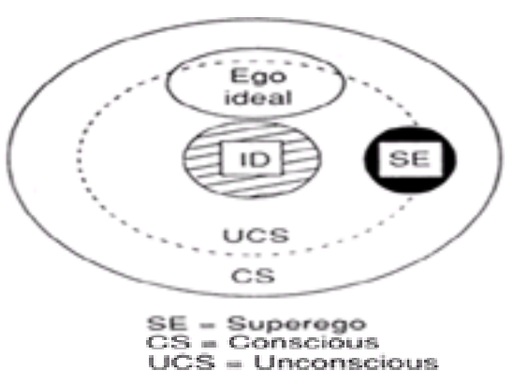

FIGURE 6.3-4 The Freudian topology of the psychic apparatus.

Complexes Complexes are groups of unconscious ideas associated with particular emotionally toned events or experiences. Jung inferred them from his early word-association studies, when he noted that certain words provoked intense reactions or produced less reaction than would be expected. Complexes are built around genetically determined intrinsic psychic structures known as *archetypes*. Complexes are also reinforced by environmental events and by selective attention or inattention and are thus self-perpetuating. They are endowed with psychic energy from their affective tone: positive, negative, mild, or strong. The more intense the complex, the greater the emotion, imagery, and tendency to action. Complexes are often stimulated by interactions with others. A father complex can be stimulated by a person who symbolizes a father (such as an older friend) or by a stimulus such as music or art that evokes father memories. The complex, formerly dormant in the unconscious, comes to the fore and tends to dominate consciousness and to displace other complexes, which then sink into the unconscious. As the father-related external stimuli diminish, the father complex, including what was thought, felt, and expressed during its ascendancy also ebbs. This is a very different model of the dynamic unconscious from Freud's. The boundary between the conscious and unconscious is far more permeable, and there is no emphasis on forces that maintain ideas and feelings in one place or another. Movement between conscious and unconscious experience is triggered by experience. Similarly, Jung's conceptualization of interpersonal boundaries was far more permeable than Freud's; Jung also placed much more emphasis on nonverbal stimulation (music, art, spiritual life) in this process than did Freud, who was the archetypal Enlightenment rationalist.

Some complexes are more conscious, better developed and more ego-syntonic; others are less conscious, poorly developed, and ego-alien. The latter are projected onto the environment, especially by the immature psyche of children, and from this, projective and introjective processes evolve. One person may introject and identify with a complex being projected by another person. Thus, therapists may become psychologically infected by their patients. It is also possible to project a complex that is not integrated within one-self onto another person and then develop a relationship with that projected complex. One can envision an interpersonal environment charged with projected complexes potentially available for introjection, thus offering endless potential for psychic mutation in an interpersonal field. Thus, for Jung, the psychological boundaries between individuals was far less clear than they were for Freud or Adler. Jung also had a parallel appreciation for the process of projective identification that Melanie Klein was developing at about the same time.

In another insight that paralleled Klein's observations, an important aspect of complexes is their bipolarity. Each complex has a positive and negative pole, such as good father and bad father or rewarding father and punishing father. One pole of a complex can be projected onto another person, who in turns acts on it in a relationship. In this way, the theory of complexes is a theory of interpersonal as well as intrapsychic relationships.

In Jungian theory, the ego is also a complex. It serves the same function as the Freudian ego, controlling conscious life and bridging the intrapsychic and external world. The other complexes that also make up the psyche may align with or be in opposition to the ego. For example, emotionally charged primitive complexes have a great tendency to become autonomous and may behave like partial personalities that oppose or control the ego. These personalities appear as images in dreams, as hallucinations, and as separate personalities in dissociative identity disorder. They also appear in seances when mediums bring forth so-called personalities from the dead. For Jung, this phenomenon also explained animism and states of possession. Jung had a great interest in mystical experience, which Freud saw as the persistence of infantile, magical thinking, primary process, or wish-fulfillment—this is another of the clear divisions in their thinking. For Freud and Freudians the world is a harsh, demanding place that forces one to give up primitive wishes for magical experiences. For Jung and Jungians magic is alive and well in the inner world, and is to be embraced.

Archetypes Complexes are connected to structures embedded more deeply in the psychic apparatus, the archetypes. Complexes, the more superficial aspect of the complex-archetype continuum, are related to events, feelings and memories from individual lives. They are the means by which archetypes express themselves in the personal psyche. Archetypes are the inherited capacity to initiate and carry out behaviors typical of all human beings, regardless of race or culture, such as nurturing and accepting nurturance; being aggressive or dealing with aggression by others. These predispositions are analogous to the organization of the cerebral cortex into the anlage for perception of visual or auditory stimuli that become the capacity to see and hear, but that require stimulation for their development. Just as vision cannot develop without visual input during physiologically critical stages, so archetypes require interactional stimulation for their elaboration into complexes. Thus, the human infant's psyche is not amorphous energy awaiting organization by the environment; it is instead a complex and organized set of potentials whose fulfillment and expression depend on the appropriate environmental stimuli. There are as many archetypes as there are prototypic human situations.

The mother complex-archetype illustrates the interrelationship of complex and archetype. All humans are born with a poorly formed but relatively clear model of an all-nurturing caretaker, which Jung referred to as the *earth mother archetype*. The mother complex emerges from this based on experiences with mothers or mother-surrogates: their attitudes, personalities, and relationship to the particular individual. The mother archetype is found in dreams or fantasies, often as a huge woman or an animal with many breasts. The motif of a many-breasted animal, found in many cultures, is that of unlimited nurturance.

Unconscious The Jungian unconscious has two layers, the more superficial being the *personal unconscious*, and the deeper layer the *collective unconscious*. The complexes exist in the personal unconscious, the archetypes in the collective unconscious or *objective psyche*. The personal unconscious is the equivalent of the Freudian unconscious, a repository of what has been repressed. The collective unconscious is the residue of what has been learned in humankind's evolution and ancestral past, much as human deoxyribonucleic acid (DNA) is an aggregate of the past. In this portion of the psyche reside instincts, potential for creativity, and the spiritual heritage of mankind. The potent synthesis of Darwinian natural selection theories with Mendelian genetics, which has had such a profound impact on all of biology, occurred in the 1920s. It is not clear whether Jung was aware of this intellectual thrust, but his understanding of the collective unconscious and of the archetypes is remarkably consonant in form if not in detail with the modern understanding of hereditary and developmental neurobiology as applied by cognitive science.

The psyche, like all other living systems, attempts to stay in balance. Jung's term for homeostasis in the relationship of conscious to unconscious life is the *law of compensation*. For any conscious attitude or experience that is overly intense, there is an unconscious compensation. A person experiencing neglect might fantasize or dream about a huge, many-breasted mother. When interpreting dreams, Jung would ask himself what conscious attitude the dream compensated for.

Symbols Although Jung accepted that certain symbols are universal, he suggested that in dealing with patients it is wisest to view symbols as expressions of content not yet consciously recognized or conceptually formulated. A tall, cylindrical object might symbolize a penis, but it could also stand for creativity or healing. Symbols are often attempts to unite and strike a balance between images from the collective unconscious with the personal unconscious. A tall, cylindrical object that symbolizes a penis in the personal unconscious might symbolize the phallic principle of creativity or fertility in the collective unconscious.

Personality Structure At the center of the conscious personality is the complex called the *ego*. Several universal complexes attend the ego. The *persona* (named after the mask worn by ancient Greek actors), or public personality, mediates between the ego and the real world. The *shadow*, a reverse image of the persona, contains traits that are unacceptable to the persona, whether they are positive or negative. A brave persona, for example, has its fearful shadow. The archetype of the shadow is the enemy or feared intruder. The *anima* is a residue of all the experiences of woman in a man's psychic heritage; the *animus*, the residue of all the experiences of man in a woman's psychic heritage. The anima or animus connects the ego to the inner world of the psyche and is projected onto others in day-to-day or intimate relationships. When connected with the shadow, a man, for example, might see attributes of woman as undesirable, and might experience guilt encountering such qualities in himself.

SELF The *self* is the archetype of the ego; it is the innate potential for wholeness, an unconscious ordering principle directing overall psychic life that gives rise to the ego, which compromises with and is partly shaped by external reality. In Jungian metapsychology, the unconscious gives rise to integration, order and individuation. The self appears from the unconscious in dreams, fantasies, and altered states of consciousness to give direction. In the first half of life, the ego attempts to identify with the self and to appropriate the power of the self in the service of the ego's growth and differentiation. During this time the ego may become inflated with an unrealistic sense of power: the arrogance of youth. If cut off from the self, there may be a sense of alienation and depression.

INDIVIDUATION In the second half of life the ego begins to attend more to the self than to the conscious realm of life. Jung called this developmental process *individuation*, the drive for individuals to become unique and to fulfill the spiritual propensities common to all humanity. Often, this process requires withdrawing from earlier identities and conventional definitions of success and seeking new paths. This change often has the paradoxical effect of leading to broader and more mature relationships in addition to greater creativity.

PSYCHOLOGICAL TYPES Jung's theory of psychological types has three axes (Fig. 6.3-5). The *extroversion-introversion* polarity refers to the two basic types of object relatedness. Extroverts are oriented to others and to the world of consciousness. Their energy flows outward first, then inward. Introverts are oriented to their inner world, their energy flowing first inward and then to outer reality. Introverts might therefore be seen as selfish and unadaptable because they attend first to their inner world and then determine how the outer world can fit them. Extreme extroverts, on the other hand, can seem insensitive to themselves and to the inner lives of

others.

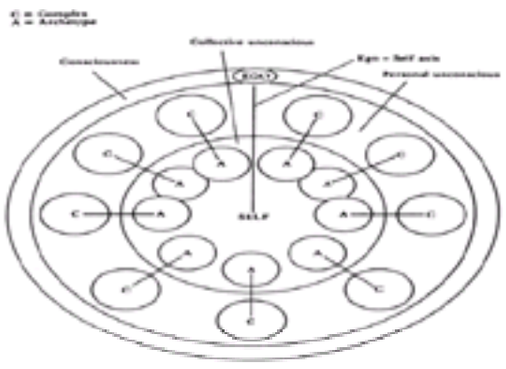

FIGURE 6.3-5 The Jungian psychic apparatus. (Reprinted with permission from Stevens A: *On Jung*. Routledge, London, 1990.)

The *sensation-intuition* polarity concerns perception. The perceptive type that Jung called sensation-oriented is stimulus-bound and attuned to the specifics of here-and-now reality. The intuitive type blurs the details but apprehends the overall picture. The sensation type comes to understand a situation by assembling the details; the intuitive type grasps the overall situation before attempting to assimilate its parts. The sensation type sees the trees first, the intuitive type sees the forest first.

The polarity of *thinking* and *feeling* deals with information processing and judgment. In the thinking mode, data is evaluated according to logical principle. Feeling, at the opposite pole, is making judgments through nonlogical processes having to do with values and understanding relationships. In social relationships, the thinking type would deal with people according to their social rank or according to the tradition of etiquette; a feeling type would deal with others in terms of their present social relationship or perceived emotional state.

Each individual has a preferred mode on each of these three polarities. By placing an individual on each of the three axes indicated in Figure 6.3-6 each individual can be identified as a type. An extroverted-sensation-thinking type is oriented to the real world, tends to perceive details, and organizes them into a logical structure. An introverted-intuition-feeling type is self-oriented, grasps situations as a whole, and is sensitive to their emotional implications.

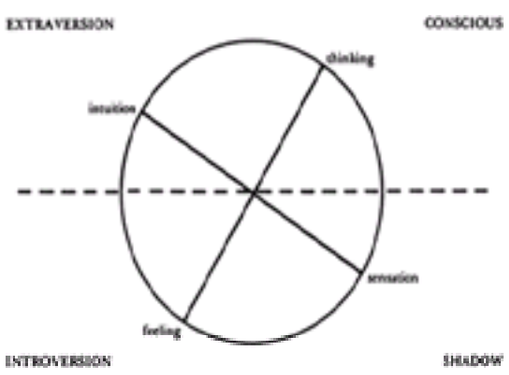

FIGURE 6.3-6 The three personality axes: extraversion-introversion, thinking-feeling, and intuition-sensation. (Reprinted with permission from Stevens A: *On Jung*. Routledge, London, 1990.)

Everyone's psyche contains all of the types. However, each person has a superior set of functions: types that are evolved from early life and that are shaped strongly by constitutional factors. In the second half of life, adults who continue the process of individuation attempt to integrate or broaden and deepen their understanding of their inferior functions. Thinking types become more aware of feelings, sensation types allow themselves to rely more on intuition, and extroverts become more interested in their own inner lives. The Myer-Briggs Type Indicator, a simple paper-and-pencil test consisting of about 40 questions, reliably places individuals on these dimensions. It has become a very popular instrument among pop psychology, management consulting, counseling, and self-improvement organizations for quickly identifying the qualities of individual participants. The typology may be the most widely known and accepted aspect of Jung's theories, although it is often used independently of its original context.

Psychopathology When another complex becomes incompatible with the ego complex, anxiety is experienced. To contain the anxiety, the incompatible complex splits off from the ego and moves unconsciously against the ego or other complexes identified with the ego. This splitting enables individuals to live out the two incompatible complexes, one more ego-identified and the other more ego-alien. The latter is often experienced as being inflicted by the outer world (“I am being mistreated” rather than “I have inner conflict”). This splitting and dissociation of consciousness is particularly evident in hysteria and is an especially good explanation of the phenomenon of multiple personalities (dissociative identity disorder). Within the psyche, ego or part personalities or complexes operate along with shadow personalities that are antithetical to them. In addition, the anima and animus and archetypal images of the self that attempt to integrate and to control the chaos. The various personalities that appear in dissociative identity disorder are manifestations of the complexes and the self.

Extroverts tend to develop hysterical or antisocial traits whereas introverts become dysthymic, anxious, and obsessional. The symptoms are often related to the attempted emergence of inferior functions. Viewed in this way, pathological conditions may contain within them the struggle toward wholeness or health; attempts by inferior functions to become integrated instead of dissociated from consciousness. The integration of these inferior functions often requires emotionally painful rearrangements of conscious thoughts, attitudes, and lifestyle.

The attempted expression of inferior functions need not result in psychopathology. Persons with highly developed thinking functions may find themselves wishing to experience life more fully and may involve themselves in highly emotional extramarital relationships. On exploration, one may find the unresolved loss of a mothering person with the resultant inability to achieve intimacy with a woman. The process of reaching for intimacy is enacted outside the marital relationship because the wife has been assigned the role of the cold, abandoning mother. This is an example of how Jungian theory can arrive at a fairly typical Freudian formulation by a rather different route.

Application Jung suggested that therapy begin with four visits per week and then be spaced out to one or two per week. Present-day Jungian practitioners work with their analysands once a week, face to face. Jung placed great emphasis on the human relationship between analyst and analysand and noted that both parties change in the course of analysis. He defined *transference* as the patient's attempt to develop psychological rapport with the doctor and held that without rapport and object relationship, the technical operations of the analyst were of no value. With its emphasis on reintegration, symbolism, and dreams, Jungian analysis seems well suited to helping educated persons deal with the developmental problems of midlife. Having achieved a professional identity, material success, and established a firm family role, they often begin to ask “Who is the real me?” “What is most meaningful to me?” and “What is my relationship to humankind and human history?” Often given strongly to self-criticism in the service of self-improvement, such individuals may be relieved to find themselves described as trying to become fully themselves instead of as being neurotic. However, the complicated, often blurred intrapsychic relationships described, the interest in mystical experience with its potential shading into magical thinking, and the blurring of boundaries between therapist and patient have made analytic psychology particularly attractive to those who are drawn to or inclined toward cults or exploitative relationships with patients. The break between Freud and Jung, which had for decades been ascribed to Freud's rigidity and authoritarianism, now appears to have had a contribution from Jung: Freud disapproved of a complicated, dependent relationship Jung is believed to have had with a former patient.

Jungians would see Mr. A.'s bipolar I disorder and use of marijuana as having unleashed conflicting archetypes, all attacking the ego complex he had grown up with. This could be seen as an opportunity for Mr. A. to claim a fuller definition of the self. Mr. A.'s aspirations to be a great writer would be taken as potentially healthy and even the content of his psychotic and substance-induced states would be analyzed so as to incorporate their meanings into his ego complex. Mr. A.'s previous psychological type would be described as extroverted, sensation, thinking—an observer of life, but not a full participant. The type emerging in his psychotic and intoxicated states would be described as introverted, feeling, intuitive—the self-absorbed, self-experiencing lover of beauty and ideas. The dreams about flying suggest an expression of the hero archetype; the anxiety dream suggests some emergence of his shadow in association with his aspirations to be a writer. His love of flowers and creativity during his periods of psychosis and intoxication suggests that the shadow is connected to his anima. The task of therapy would be to assist Mr. A. to accept these alien parts of himself and integrate them into his ego complex. Relatively little emphasis would be placed on education or specific treatment of his bipolar I disorder, except insofar as it prevented a good therapeutic alliance. His fantasies, dreams, and psychotic experiences would be investigated, discussed, and analyzed in terms of the psychological functions used and the archetypes expressed. His interpersonal issues with his family would be seen primarily as projections by him of other, unacceptable archetypes that needed identification, analysis, acceptance, and integration. For example, his relationship with his mother would be seen as a projection and fear of the earth mother archetype. The overall goal would be to harness the unleashed unconscious material to allow Mr. A. to pursue his ambitious, creative goals.

SANDOR RADO

Hungarian-born, trained in psychiatry, Rado (1890–1972) emigrated to the United States in 1931 after helping to organize the Hungarian Psychoanalytic Society and being an active member of the Berlin Psychoanalytic Institute (Fig. 6.3-7). In 1945 at Columbia University he was the director of the first psychoanalytic institute within a university medical school. Rado believed that learning, cultural, and parental influences were of greater importance than instinctual factors as causes of emotional or behavioral disorder, but also believed strongly that an underlying genetically determined biochemical abnormality was present in schizophrenia. His work fit in with the development of ego psychology, influenced by Anna Freud, Heinz Hartmann, and David Rappoport.

FIGURE 6.3-7 Sandor Rado. (Courtesy of New York Academy of Medicine.)

Adaptational Psychodynamics Rado viewed the psychic apparatus as an organ of adaptation. Effective adaptation is psychological health. Psychological illness or maladaptive behavior is a failure of this adaptive mechanism. There are hierarchical levels of human mental integration: hedonic, emotional, emotional thought, and unemotional thought. These levels parallel the phylogenetic and ontogenetic evolution of the brain with the gradually increasing influence of the neocortex over the more primitive parts of the brain. The hedonic level is the realm of pain and pleasure. Pain signals impending damage and causes flight. Pleasure indicates benefit and elicits clinging or moving toward. Persistent pain has the secondary effect of reducing self-esteem through a sense of failure; pleasure has the secondary effect of enhancing self-esteem through a sense of success. More than most other psychoanalysts, Rado stayed true to Freud's belief that a neurobiological basis for the psychic apparatus would be discovered. He integrated Walter B. Cannon's theory of emotions with James Papez' identification of the limbic system as the neuroanatomical locus of emotional and hedonic mechanisms. He also anticipated the recognition by modern evolutionary biology and neuropsychology of hedonic mechanisms being the keys to memory.

The emotional level of psychic integration consists of the emergency emotions (e.g., fear, rage) and the tender or welfare emotions (e.g., love, pride). The emergency emotions are responses to real or anticipated pain and result in flight or fight. The welfare emotions are responses to actual or anticipated pleasure or benefit and prepare the organism to embrace the pleasurable stimulus. Rado's observation that ordinary human mental activity is primarily at the hedonic and emotional levels was borne out by George Vaillant's later studies that followed a group of successful college undergraduates into their middle years.

Emotional thought justifies and reinforces the emotions from which it springs and is not reality-based; it corresponds to Freud's primary process. Examples of emotional thought include dreams, fantasies, delusions, and the hallucinations of psychiatric illness. Rational or unemotional thought operates on Freud's reality principle, making it possible to delay action and gratification, to forgo present pleasure for future gain, and to control emotional responses. Rado concluded that much effort is required by ordinary adults to combat emotional thinking, and that the emergency emotions are far stronger than the welfare emotions.

Conscience Conscience is the part of the mature psychic apparatus, operating unconsciously, that rewards good behavior, raising self-esteem, and punishes bad behavior by guilt, thus lowering self-esteem. Unlike Sigmund Freud's conception of the superego, conscience is seen as primarily constructive and adaptive. It facilitates cooperation with others and reduces destructive forms of competition. Although conscience stimulates adaptive behavior, it can also produce pathological behavior.

Anticipating the work of Lawrence Kohlberg on the developmental stages of moral thought in both children and adults, Rado described the development of conscience as originating in the dependent child's wish and need to remain in its parents' good graces. Initially, the child projects its own omnipotent feelings onto the parents. Believing that parents see and know all causes the child to fear punishment that it feels is inescapable. A child who misbehaves experiences guilty fear, which has the positive effect of stimulating expiatory behavior. The child admits wrongdoing and accepts punishment to restore the parents' positive feelings. This relieves the painful guilty fear and restores the child's self-esteem. Pain-dependent or masochistic behavior is a form of expiatory behavior based on guilty fear. Fear motivates the person to seek punishment in advance, which may at times permit satisfaction of a formerly proscribed desire. Although not a complete explanation of masochistic behavior, Rado provided an important link in its understanding.

Both discipline and conscience formation are complicated by the buildup of rage that occurs when the child obeys or submits to its parents. The rage, which is repressed, constantly seeks discharge and is a potentially serious problem for both the individual and society.

Treatment According to Rado, psychological health is a predominance of welfare emotions in a reasonably independent, self-reliant person. Such a person creates a self-reinforcing interpersonal field by stimulating pleasurable feelings in others. By recognizing the interpersonal field within which emotions occur, Rado reflects the influence of Sullivan and Kurt Lewin. The emergency responses are still active, but are released nondestructively through various types of play or constructive activity and through dreaming.

The aim of psychotherapy is to increase the influence of the welfare emotions on behavior and to reduce the influence of the emergency emotions. Patients are encouraged to relinquish the dependency that both causes and results from maladaptive behavior and to become reasonably self-reliant. The past is explored in therapy primarily to increase understanding of the present rather than to reconstruct the past. Rado focused more on examining patients' present-day behavior than on the recovery and analysis of memories and the development of insight based on those reconstructions. He preferred to educate patients directly instead of developing and analyzing the transference, as was done in classical analysis.

Rado would have framed Mr. A.'s difficulties as failures of adaptation to his mental illness and adult life and as a regression to the hedonic level of adaptation in which pleasure is sought and pain avoided. As Mr. A.'s self-esteem was reduced through his failure to face his obligations, he avoided the resultant pain through flight into mania and the use of drugs. Activation of his emergency emotions led him to behave and to fail in ways that provoked punishment. He failed at graduate school and was arrested for disturbing the peace. Rado would have emphasized reducing the self-defeating behaviors that alienated Mr. A. from others and on increasing

behaviors that would result in support and positive input from others. Rado would have encouraged Mr. A. to examine his dependence on his family and to decide if he wanted to continue in this dependent role. Mr. A. would have been helped to uncover his welfare emotions and to relate to his family in a positive way that would allow him to reduce his rage toward his parents for being controlling of him, reduce his acting out against them, and enable him to develop into his own person.

FRANZ ALEXANDER

Alexander (1891–1964) was one of the second generation of psychoanalysts. He was born in Budapest and attended medical school there, graduating in 1912 (Fig. 6.3-8). He conducted research at the Institute for Experimental Pathology in bacteriology until World War I when he practiced clinical microbiology on the Italian front, primarily combating malaria. Following the war he joined the department of psychiatry at the University of Budapest Medical School as a brain researcher. This led to an encounter with Freud's work and in 1919, he became the first student of the Berlin Psychoanalytic Institute. In 1930 Alexander became a visiting professor of psychoanalysis at the University of Chicago and in 1932 founded the Chicago Psychoanalytic Institute. He established the Institute independent of the psychoanalytic societies, leading the Chicago Institute to be one of the most creative sources of psychoanalytic thought. During the same time he began his interest in psychosomatic illnesses, helping to found the journal *Psychosomatic Medicine*. A guiding principle of his work was to make psychoanalysis an integral part of medicine.

FIGURE 6.3-8 Franz Alexander. (Courtesy of Franz Alexander.)

In 1946 Alexander became Professor of Psychoanalysis at the University of Southern California, where he continued his work on psychosomatic medicine and became interested in the interface of learning theory, the psychophysiology of stress, and psychoanalysis.

Theory of Personality Alexander did not develop a unique overarching theory of personality; his contribution was his application of psychoanalytic thought to pathophysiological processes. In this he laid the ground work for the burgeoning fields of psychosomatic medicine, behavioral medicine, and psychophysiology. Alexander created the basis for the biopsychosocial model and studied the mind and body at a time when American psychiatry, despite Freud's original ideas, had become purely psychological in its orientation. In studying and treating many patients with serious physical illness, he was also forced to consider creative modifications of therapeutic technique.

Alexander and his group began by intensively studying, by means of clinical interviews, patients who had one of seven illnesses that had been identified by general practitioners as having strong psychological components. Out of these clinical studies emerged the *specificity hypothesis*, which proposed that certain illnesses are the product of a complex interaction of specific constitutional predispositions, specific unconscious conflicts, and specific types of stressors that activate such conflicts. The group then tested these hypotheses in a series of clinical studies of patient populations. The independent variable was usually the ability of skilled clinicians to predict a patient's illness from disguised case reports. The model he proposed remains the fundamental psychosomatic conceptualization; a variety of illness situations are now studied by controlling for genetic influences while measuring and modifying intrapsychic conflict and external stressors.

In the field of psychotherapy Alexander was one of many who sought to shorten the analytic process. He hypothesized that intellectual insight was not the central curative factor in therapy. Rather, he emphasized the role of corrective emotional experience. This led him to experiment with variations in technique that might facilitate such experiences. This, his most controversial stand, nearly ruptured his relations with the psychoanalytic movement.

Theory of Psychopathology The seven diseases that Alexander studied were peptic ulcer disease, ulcerative colitis, essential hypertension, Grave's disease, neurodermatitis, rheumatoid arthritis, and bronchial asthma. Alexander and his group identified what they believed to be the single, specific core conflict which, in interaction with constitutional predispositions and in the circumstance of a particular stressor, would activate the disease.

The core conflict identified in peptic ulcer disease is hyperindependence as a defense against unacceptable dependency needs. The stressor that would result in an acute attack could be any situation that demanded that the afflicted individual openly acknowledge or ask that dependency needs be met. Ordinarily, patients would use dominance and control to intimidate others into meeting their dependency needs. Thus, Alexander was the first to describe the “little boy” business executive prone to peptic ulcers. The groundwork was laid for John J. Brady's executive monkey experiments. This particular illness model has received more confirmatory evidence than any of the others, especially from a study of army inductees in whom psychological profiles as described by Alexander, in combination with measurements of serum pepsinogen, were extraordinarily successful in predicting the development of duodenal ulcer. It is not clear what the impact of the recent recognition of a specific pathogen means for this particular hypothesis. On the one hand, there is ample evidence that Alexander was right. Perhaps the identified pathogen, *Helico pylori*, is merely the constitutional factor that had previously been thought to be serum pepsinogen concentration, or perhaps Alexander's hypothesis applies to some subgroup of patients with duodenal ulcer. However, today peptic ulcer disease is treated primarily as an infectious ailment.

Alexander's theory of ulcerative colitis also implicated dependency conflicts; however, rage at unmet needs was seen as the defining feature. This rage provokes guilt and the urge to make restitution towards the object of anger by means of gifts of achievements and successes. The model here is clearly the angry child seeking to placate a parent by means of performance. The precipitating event for reactivation of the illness is the perception that the efforts at placation will be unsuccessful. Alexander claimed that this resulted in excess parasympathetic activity, leading to diarrhea. Alexander's own study in which skilled internists and psychiatrists reviewed case descriptions in which each case had one of the seven identified psychosomatic illnesses, and then predicted which illness each patient had, resulted in correct identification of over half of the patients with ulcerative colitis from their dynamics alone—well beyond chance. However, most clinicians now regard George Engel's object-relations based formulation to be the more accurate.

Alexander's hypothesis about essential hypertension focused on inhibited anger and suspiciousness in an outwardly compliant, cooperative individual. A hypertensive patient often goes for long periods with blood pressure under good control and then, seemingly inexplicably, experiences a dramatic rise in blood pressure. Alexander attributed these episodes to incidents when the chronic anger is exacerbated and intense defenses must be used, causing chronic sympathetic, fight-or flight activation. Several psychophysiological studies have suggested that this theory has some validity at least in terms of short-term changes in the blood pressure of a subset of hypertensives. Labile hypertensive patients appear to fit this model best.

The proposed specific conflict for rheumatoid arthritis postulates conflicts over rebellion against overprotective parents. A compromise formation in which the conflict is discharged via physical activity, especially sports, works for a time, but the anger eventually is expressed in self-sacrifice designed to control others. Failure of this pattern results in increased ambivalent tension, directly expressed via muscular contractions that lead to joint degeneration. A remarkable amount of confirmatory evidence for this constellation appeared in a series of psychological test studies; however, more recent research has failed to replicate much of this and has implicated more general stress issues, life change, and psychoneuroimmunological mechanisms.

Alexander proposed that the wheeze of bronchial asthma represented a symbolic cry. The specific conflict, according to him, was the wish for protection versus the fear of envelopment. This conflict leads to sensitization to separation issues, which become the events that provoke the suppressed cry of the asthma attack. In recent years it has become clear that the population with asthma is far more heterogeneous both psychologically and physiologically (in terms of vulnerability to allergens) than was recognized in Alexander's time. The role of a vicious physiological-psychological cycle in which asthma stimulates panic, which in turn triggers pathological pulmonary psychophysiological responses, has been a focus of more recent research.

Although the role of conflicts in neurodermatitis remains widely accepted by clinicians, the specific conflict proposed by Alexander, that early deprivation leads to wishes for closeness that are opposed by a fear of it, is no longer accepted. Finally, Grave's disease (thyrotoxicosis) is no longer widely accepted as a psychosomatic illness. Alexander's hypothesis was that premature responsibility led to a martyr-like denial of dependency.

Treatment The specificity hypothesis led Alexander to focus his psychotherapeutic efforts in a way that other analysts of his time did not. He reasoned that if he could help patients to resolve their core conflict without necessarily addressing other parts of the personality structure, the medical illness would improve. Indeed, he published numerous case studies suggesting just this kind of success. In addition, he was among the first to question the value of intellectual insight as the curative agent in psychotherapy. He proposed that a corrective emotional experience is the central agent of change. A corrective emotional experience involved disconfirmation within the transference relationship of previous assumptions and projections.

Thus, Alexander felt justified in introducing a variety of techniques that would initially induce and heighten the emotional experience of the transference, and, subsequently, challenge the underlying unconscious assumptions. These techniques included manipulating the frequency and length of sessions, making direct suggestions about the patient's life, self-conscious alteration of the therapist's behavior according to the patient's conflict and behavior therapy techniques. In many ways this was the most controversial aspect of Alexander's work. Serious questions about the validity of his suppositions and the ethics of his “manipulative stance” were raised. He was impatient with the slow, methodical process of convincing his colleagues, his intellectual energy led him to embark on ever-newer experiments while other analysts were still struggling to digest his previous suggestions. Yet today few would quarrel with the concept that emotional learning is at least as important as intellectual insight in psychotherapeutic success. Alexander's efforts to modify and shorten the analytic process have come closer to the norm in psychiatric practice than in classical analysis.

Ironically, although Alexander's therapeutic innovations seem prescient today, his specificity hypothesis, which was in his time far less controversial, seems simplistic, naive, and forced. He did not have available the sense of how complicated illness causation truly is. The dominant model of the time was infectious diseases: one organism, one illness. In addition, the complexity of social phenomena and stressors was unknown in his time. Yet, Alexander was the first to postulate a multicausal etiology for disease: a specific constitutional defect, a specific conflict, and a specific stressor. He was also the first to study mind-body interactions in a systematic way. Thus, he laid the very basis upon which his own formulations seem so limited today.

In the absence of a clear psychosomatic illness, Alexander would have little to say specifically about Mr. A's dynamics. However, like any good dynamically oriented psychotherapist, he would readily recognize Mr. A's core conflict about dependence-independence. Mr. A would be understood as rebelling against his mother's nurturant control, while secretly desiring to continue basking in it. He has serendipitously found that illicit drugs and his mental illness can be harnessed such that his mother has to rescue him and take over, while he can avoid becoming aware of the conflict. Alexander's unique contribution to Mr. A's case would be to propose that he could be treated with 40 or fewer sessions, focusing exclusively on this conflict (and ignoring others such as oedipal components and issues with his father). Further, Alexander's agenda would be to structure the therapeutic relationship to intensify Mr. A's experiencing of this conflict. He might do this by taking a somewhat intrusive, directive stance, thus creating the opportunity for a rapidly, intense transference around this issue. Alexander might then spread out the sessions to elicit expression of the patient's yearning for that control. The key therapeutic effort would then be to respond constructively to attempts by Mr. A to assert himself appropriately and take over control of his life—for example by learning more about his illness and taking responsibility for his treatment.

WILHELM REICH

Reich (1897–1957) was one of Freud's most controversial disciples; his latter years were marred by mental illness (Fig. 6.3-9) Reich fixed on and elaborated Freud's early, but later discarded view, that neurosis results from the damming up of sexual energy. Blockage of normal orgasm can lead to partial conversion of sexual energy into aggression, but residual tension manifests in the form of characteristic physical tensions that reflect the underlying character armor of each individual. In so doing, Reich laid the groundwork for a psychoanalytic theory of personality. Prior to Reich, the focus was on symptoms and psychopathology. There was occasional mention of “character” in psychoanalytic work, but there was no focus on personality as an entity unto itself. This shift is of great importance because as neurobiology has come to explain more and more Axis I mental disorders it has become clear that psychodynamic theories' enduring strengths involve understanding and treating personality traits and disorders. Together with Anna Freud and Heinz Hartmann, Reich is regarded as the originator of ego psychology.

FIGURE 6.3-9 Wilhelm Reich. (Courtesy of New York Academy of Medicine.)

Personality Theory Reich did not disagree with Freud's notions concerning personality development, including character types based on fixation at specific levels of psychosexual maturation. Reich elaborated the interpersonal and physical behavior of these personality types. Specific behavioral traits constitute *character armor* that defends against internal and external dangers. Character armor is composed of involuntary, repetitive, ego-syntonic behaviors that prevent the emergence of repressed impulses. For instance, the trait of ingratiation frequently defends against hostile impulses, just as the traits of hostility or self-assertion may defend against wishes to be dependent and passive. These traits manifest physically in the voluntary musculature as characteristic postures (clenched jaw or fist, rigid or bowed back) or in excessive stiffness or fluidity of movement. Although Reich's ideas of the behavioral and postural components being central features of character armor are no longer accepted, his ideas successfully transformed the focus of psychodynamic thought from individual wishes defended against by individual defenses, expressed in specific transferential forms into an emphasis on patterns, organization, and intertwining of all of these elements.

Hysterical Character The hysterical character has the least body armoring, hence the most lability of function. Body movements tend to be soft, rolling, and sexually suggestive. These individuals are superficial, excitable, flighty, fearful, highly suggestible, and easily disappointed. Their armor helps to defend against easy sexual arousability by flushing out potential sexual stimuli in the environment and then reacting to them with anger.

Compulsive Character These individuals are tense and restrained, walk stiffly and sit rigidly. They are overconcerned about orderliness, tend to ruminate, and are indecisive and distrusting. They experience a blockage between their thoughts and feelings. Because they have little access to their feelings, they have little ability to prioritize their actions, to make decisions, or to sense others' reactions to them. The compulsive character avoids expression of repressed impulses by rigid overcontrol. Because of this, these individuals are very threatened by trivial changes in routine.

Phallic-Narcissistic Character Phallic-narcissistic persons appear cold, reserved and prickly. They are outspoken, provocative, and seek positions of power. Frustrated at the genital-exhibitionist stage of development, the men are identified with the penis and the women with the fantasy of having a penis. The men have strong erective potency, but little capacity for intimacy; the women actively dominate men.

Masochistic Character Masochistic individuals suffer, complain, and damage and deprecate themselves in ways that provoke and torture others. Reich differed with the analytic interpretation that these persons enjoy suffering. He thought the opposite—that pleasure is painful for the masochist because of an enormous need, excessive guilt, and the resultant low tolerance for love or pleasure. Suffering allows the masochist to then indulge in a certain amount of self-gratification. Sexual intercourse can be enjoyed, for example, if the partner is inconsiderate or if intercourse is accompanied by the fantasy of rape.

Treatment Reich's major contribution was in the realm of treatment. He was the first to recognize the need to deal with character resistances before attempting to

recover repressed material: that interpersonal resistances need to be dealt with before free association is possible. He did this by analyzing patients' character armor—their characteristic behaviors (including tone of voice, posture, and physical movements) in the analytic setting—before proceeding to an analysis of the unconscious. Reich worked face to face with patients and sought to relax their character armor by physical manipulation. This type of therapy, called *vegetotherapy* by Reich, is still practiced by Reich's followers as *bioenergetics*.

A Reichian analyst might have seen Mr. A.'s seeming rebellion against conformity as a defense against his fear of being away from his mother's protection and domination. The Reichian might have pointed out that the more Mr. A. rebelled, the more tightly he was binding himself to his mother. The therapist might also have suggested that the behaviors that seemed so much under Mr. A.'s control were in fact compulsive behaviors; his prolongation of his manic states and use of drugs were being dictated by forces outside of his awareness to protect him from internal and external dangers. Among them, the internal danger of recognizing his rage at his parents for their overprotection and the external danger of failing to deal with a world with which he was poorly equipped to deal. As these behaviors became more ego-alien, the Reichian would have begun a classical analysis of the conflicting unconscious forces involved, including the patient's fear of engulfment by his mother and his wish to be engulfed; fear of his attachment and the wish to be attached, and the harshness of his own superego. The patient's identification with his father as one who superficially placates, but with contempt, would have been explored. The analysis would have included Mr. A.'s dreams. His dream of flying might have been interpreted as his fear that to separate from his mother he would need to become invulnerable to all life's potential injuries, a superman. His lobby dream might have been interpreted as his wish to return to the womb as a defense against the injuries he might experience away from his mother.

OTTO RANK

Rank (1884–1939) was a 21-year-old student when he met Freud (Fig. 6.3-10). Rank later earned a doctorate in psychology and eventually became a peer of Freud's before ultimately breaking with him. Rank saw each person as an artist whose ultimate task is the creation of an individual personality. In Rank's view the neurotic was an *artiste manque*, a person whose strong creative urge was stultified by the negative use of will.

FIGURE 6.3-10 Otto Rank. (Courtesy of New York Academy of Medicine.)

Rankian Dialectic The basis for his break with Freud was Rank's view that the birth trauma was more important than the oedipal conflict. According to Rank, the physical and psychological experience of birth gives rise to a primal anxiety that is dealth with by primal repression. The crucial intrapsychic conflict that occurs in all developmental phases is the conflict between maintaining the primal bliss of attachment and experiencing the excitement and fear associated with separation.

Union stands in contrast to separation; likeness stands in contrast to difference. In adulthood, movement toward another person is possible only if one knows who one is, which can come about only through having experienced separation. Movement toward autonomy is possible only after having established the sense of belonging and self-worth that derive from the experience of fitting in or belonging.

Moving toward union or separation is not an innate biological process but an act of will. In moving toward and engaging with another person, all individuals experience their need for belonging. Moving away from others allows individuals to experience their uniqueness. Maturity is the triumph of will over the forces that inhibit movement both toward and away from others—guilt, death-fear, and life-fear.

Rank saw guilt as the price to be paid for any act of will. Moving toward union causes guilt over being needy; moving away causes the guilt of abandoning another person. Death fear is the fear of losing one's identity by fusing with another person. The weaker one's personal identity, the stronger the death-fear. Life-fear, by contrast, is fear of losing all ties in the process of becoming separate. Every person experiences the cycle of movement from union to separation and back again as part of the life process. This movement takes place at various levels: family, societal, artistic, and spiritual. At each level, there is one or more movement toward union and rebirth. Each person, for example, usually yields to a love experience in which personal differences are set aside to experience unity with another, to experience self-worth, and to be relieved of the sense of difference. The yielding to another ends when the will asserts its separateness and a new affirmation of individuality occurs.

Will, the prime mover in the Rankian dialectic, is an irreducible creative force. It is not solely an agency for the expression of freudian sexual or aggressive impulses, nor is it the will to power in the Adlerian sense. The beginning of will is in the child's “no,” an assertion of what the child will not do. In maturation, will becomes a positive force. Neurotic persons, however, deny will, because of guilt over what they will. They deal with that guilt by using defense mechanisms such as projection and rationalization. Viewed from this perspective, neurotics are strong-willed persons who cannot acknowledge what they will or even that they will. As a result, they cannot use their will constructively in the service of their greatest potential artistic creation, their own personalities.

Treatment Rankian psychotherapy is a here-and-now interaction with the therapist that mobilizes the patient's will and results in a rebirth experience. The treatment, which is time-limited, focuses on the relationship with the therapist. In the therapist-patient relationship are reenacted earlier life struggles, especially struggles involving intimacy. After patients are strengthened through the therapist's acceptance, they begin a process of negative will assertion that is seen as resistance in classical analysis. Rank regarded this negative will assertion as indicative of growth and supported it. Now able to provide self-affirmation on their own, patients free themselves of the therapist and begin to individuate. They overcome the life-fear by living up to their fullest potential. Therapy is not aimed at reconstructing personal history, it is a struggle in the here and now between the therapist as a representative of transference objects and of reality.

The therapeutic process parallels the process of personality growth. At first, the therapeutic relationship recapitulates prototypical early relationships. The first rebirth experience for patients is claiming their own individual personalities and their uniqueness as human beings. The second phase is their discovery of the physical universe and their likeness to it. Later, they claim their distinctness as creators of themselves. With the emergence of the self, individuals unite with ideological, philosophical, and spiritual reality and experience the final birth of the ideal person, a self-fulfilled person who no longer needs to create to justify his or her existence. In Rank's theories the beginning of the shift to what would now be called pregenital issues in psychoanalytic thought can be seen. Very early on, many of Freud's associates sensed that something was missing from the theory. Klein, Rank, and others searched for this understanding in the earliest phases of development. Although rarely cited these days, Rank's work clearly anticipated that of the object relations therapists and the self psychologists.

A Rankian therapist would have seen Mr. A. as attempting to achieve separation from his family of origin, but as lacking the will to achieve real separation. The prolongation of his manic episodes and his use of drugs helped him to avoid the pain of separation and individuation, and his failure to complete graduate school allowed him to maintain his dependence on his family. Initially, the Rankian therapist might have approved of Mr. A.'s attempts to sustain his manic episodes and his marijuana use as a negative assertion of will against his mother's efforts to run his life. The therapist would then have suggested that developing more constructive and positive means might enable Mr. A. to become more free in reality. The Rankian therapist might challenge Mr. A. to develop attachments outside his family and to develop skills to help free himself. Resistance to the development of these skills would have been interpreted as fear of separation and aloneness. Mr. A. would then be encouraged to develop his will in a positive direction; following a path of action that he desired instead of rebelling against a path of action seemingly dictated by others. Finally, Mr. A. would work toward separation from the therapist as a prototypical separation-individuation experience.

KAREN HORNEY

Physician-psychoanalyst Horney (1885–1952), who emphasized the preeminence of social and cultural influences on psychosexual development, focused her attention on the differing psychology of men and women, and explored the vicissitudes of marital relationships (Fig. 6.3-11). She was one of three women whom Freud trained so as to have female analysts who would contribute their unique perspective to the psychoanalytic theory about women; the other two were Anna Freud and Helene Deutsch. Horney's view that repression and sublimation of biological drives are not the primary determinants of personality development led to her removal as an instructor in the New York Psychoanalytic Institute and her founding, in 1941, of the American Psychoanalytic Institute.

FIGURE 6.3-11 Karen Horney. (Courtesy of the Association for the Advancement of Psychoanalysis, New York.)

Personality Theory Horney believed that personality development results from the interaction of biological and psychosocial forces that are unique for each individual. At the core of each personality is an enduring *real self*. Partially equivalent to the Freudian ego and partly to Eric Berne's child ego state, the real self combines choice, will, responsibility, and identity with spontaneity and aliveness. A natural unfolding process of *self-realization* leads to the development of human potential in three basic directions: toward others—to express love and trust, against others—to express healthy opposition, and away from others toward self-sufficiency.

Although conditions during childhood may block psychological development healthy growth is always possible if the internal blockages can be removed. Children whose family situation leads them to feel endangered concentrate on psychological survival, and may do so at the cost of developing stereotyped coping mechanisms. All human beings suffer from *basic anxiety*, which Horney saw as the normal response to the infant's helplessness and separateness. How families respond to this fundamental situation, guided by cultural norms, determines whether individuals spend their lives struggling with basic anxiety or pursuing self-realization.

Horney believed that the attributes of passivity and suffering were not biologically specific to women, as taught by the analysts of her day, and that male and female personalities are in fact culturally determined.

Theory of Neurosis Horney defined neurosis in both intrapsychic and interpersonal terms. She noted that her patients complained, not of the symptomatic neuroses such as phobias and compulsions, but of unhappiness, blockage, lack of fulfillment in their work, and inability to establish or maintain relationships. She saw these individuals as having a complex system of self-perpetuating defensive patterns against basic anxiety—a character neuroses.

Safety Seeking Children move psychologically in three directions to relieve their anxiety, make life safe and predictable, and achieve satisfaction. They seek affection and approval, become hostile, or withdraw. Children eventually employ the coping strategy that best meets their needs, but if only one basic strategy is used children become limited in their coping repertoire and in their experience of themselves and their world. Their sense of safety is tenuous because there is now danger from within, from suppressed or repressed feelings and impulses. Given continued unfavorable environmental conditions, conflicting feelings are driven into the unconscious, and such children are left with a sense of discomfort, anxiety, apprehension, and an insecure sense of self. At this juncture their point of reference is externalized; patterns of behavior rigidify and increasing blockages to growth develop. Horney designates these complex, relatively fixed attitudes toward self and others as neurotic trends.

Character Types Horney's three main character types are based on the predominant mode of relating to others. The *compliant, self-effacing type* results from the defensive operation of clinging to others. These individuals try to curry the favor of others, subordinate themselves to others and are reluctant to disagree for fear of losing favor. The *aggressive, expansive type* results from moving against others, and relies heavily on power and mastery as a means to achieve security. The *detached resigned type* results from moving away from others in an attempt to avoid both dependency and conflict. These are very private individuals who, while refusing to compete openly, see themselves as rising above others.

Supplemental Means to Relieve Inner Tension The overdevelopment of any one of the three basic interpersonal styles suppresses the two others. In a manner analogous to Jung's complexes, repressed impulses continue to be active and to produce conflict. An artificial harmony is achieved by the use of mental mechanisms such as blind spots, compartmentalization, rationalization, and coping techniques such as excessive self-control, arbitrariness, elusiveness, cynicism, and externalization.

Idealized Image As teenagers, individuals who will grow up to be neurotic create an ideal image that, if achieved, promises to end their painful feelings and provide self-fulfillment. The idealized image counterbalances the alienation from their core selves that developing neurotic individuals undergo because the survival techniques they adopted earlier force them to override their genuine wishes, feelings, and thoughts. The idealized image covers over all the contradictions, conceals the defensive nature of their behavior, and restores a sense of wholeness. Energy formerly available for self-realization is now used in efforts to become like the idealized image. For example, an individual who has adopted the strategy of moving toward others and is consequently dependent on others for affection and approval experiences the fear of reasonable self-assertion as saintly humility and considerateness of others.

Because the ideal self is imaginary, neurotic persons are readily bruised by confrontation with reality, and work excessively hard to prove they are in fact their ideal selves. This results in a type of perfectionism that insists on flawless excellence in which “I should” replaces “I want” or “I need.” It also results in the neurotic ambition to be first and in a strong drive for revenge against those perceived as having interfered with their efforts to match the ideal self. This aspect of Horney's theories anticipates some of Heinz Kohut's earliest ideas on the origins of narcissism.

Claims, Shoulds, and Self-Hatred Despite their frequent self-disparagement, neurotic individuals expect to be treated as though they were their ideal selves. These claims to special treatment, when frustrated, produce anger, righteous indignation, and resentment. The “shoulds,” or self-imposed demands that they live up to their idealized selves, are irrational and unrelated to the realities of daily life. They are projected, experienced as demands made by others, and are also demanded of others. This results in neurotic persons being critical of others and very sensitive to criticism themselves.

Self-hatred results when the threat arises that neurotic individuals may be unable to achieve their idealized selves. If support were not needed for the idealized self, claims, shoulds, and self-hatred would not be such an important part of the psychic apparatus.

Neurotic Pride and the Pride System Glorifying aspects of the idealized self, *neurotic pride*, substitute for healthy self-confidence. Thus, when their pride is injured by others, neurotic individuals become enraged and seek to avenge their injury and to conceal their self-deception by achieving a vindictive victory over the offending person. Together with supporting claims and shoulds, neurotic pride and self-hatred form a defensive network or *pride system* that protects the idealized self. Any attempt to reduce elements of the pride system is experienced as an attack on the person. Despite the armoring of their defensive network, such individuals are not at peace because they are in inner conflict with the forces that protect them. The conflict between the forces driving toward healthy self-realization and the pride system is the *central inner conflict*.

There is also conflict within the pride system itself. Neurotic pride and claims are associated with the glorified idealized image; self-hatred and shoulds are associated with the unacceptable aspects of the self. When attempts are made to satisfy both forces simultaneously, conflict arises. Attempts to avoid these conflicts involve

further alienation from the real self.

Alienation Alienation from self is one of the most serious consequences of neurotic development. It results from the combination of repeated denial of external reality and the repression of genuine thought, feelings, and impulses. As the process of alienation continues, neurotic individuals lose touch with the core of their being and can no longer determine or act on what is right for them. Their feelings may range from uncertainty and confusion to inner deadness and emptiness.

Analytic Treatment Horney did not regard adult neurotic persons as recapitulating childhood experiences and thus did not focus on the recovery of childhood memories; she dealt instead with the self-perpetuating neurotic process. She stressed the importance of dreams in analysis and later, the exploration of the patient-analyst relationship. She was one of the earliest analysts to recognize and make constructive use of her own feelings toward patients. To Horney, psychoanalysis was a cooperative venture that enabled patients to free themselves from their neurotic structures and mobilize themselves toward self-realization. The analyst's responsibility was to assist in liberating patients from *blockages*, the forces that impede healthy growth.

Early in therapy, termed the *disillusioning process*, the two types of blockages are identified and examined. The first group of safety-oriented blockages, *protective blockages*, help to avoid the anxiety caused by self-awareness. They include silence, lateness, depreciating the analyst, the use of drugs, and even the use of self-accusation as a means to avoid further exploration.

Positive-value blockages reinforce patients' satisfaction with themselves and support their idealized self. In the disillusioning process the analyst identifies both types of blockages, exposing the protective blockages before exposing the blockages that defend the idealized image. Analyzing the positive-value blockages first would arouse too much fear.

Qualities of the Analyst These qualities, later described by Carl Rogers as therapist-offered conditions, include maturity, belief in constructive conflict resolution, and the ability to communicate hope and respect. Analysts listen, clarify, provide directions and suggest alternative resolutions to conflicts. Horney emphasized the need for the analyst to help move patients out of their alienation and suggested that therapists be flexible, tailoring their interventions to patients' present needs. She did not recommend using the couch or a fixed number of sessions per week.

Therapeutic Process Horney believed that fundamental attitudinal changes were the best means to change self-defeating, self-alienating behaviors. She created a setting in which patients were able to assess themselves as individuals, free to discover and choose personal values that fit with their real self. This type of reorientation begins after the disillusioning phase of treatment. As patients begin to question their present values and their idealizing process abates, they can revise their values and develop more flexible values consonant with their inner self. Dreams are used in all phases of treatment to bring patients into better contact with their real self. As unconscious attempts to solve conflicts, dreams can show constructive forces at work that are not yet discernible in patients' conscious thoughts and behavior.

As patients mobilize their constructive forces, they experience the struggle between the pride system and the real self. In the process they experience uncertainty, psychic pain, and self-hatred. As the central conflict is resolved successfully, patients move into the final phase of treatment: the discovery and use of their real inner self.

From the standpoint of Karen Horney, Mr. A.'s process of self-realization has been blocked in all three directions. He has failed to develop the ability to love and to trust, he expresses opposition in an unhealthy way, and he has made self-defeating moves toward independence. He would be seen as having developed a detached style of relating to others, having substituted hypomania and the use of drugs for real relatedness. He justifies his illness by taking pride in it. His goal of becoming a writer is part of the development of an ideal self that is a detached observer. He supported his ideal self by having fantasies of involvement with other writers, by convincing his professor that he was able to write a play as his term paper, and by procrastination. He has become more and more isolated from his real self by denying the reality of the facts that his mental illness and cannabis abuse posed a danger to him, and that he was failing in school.

A therapist in the Horney tradition would begin by pointing out that Mr. A.'s drug use and his sustaining of manic episodes were blocking his ability to learn about himself, and would work with Mr. A. toward abstinence from substances of abuse and appropriate use of mood-stabilizing agents. The therapist would later begin to point out that Mr. A.'s pride in his manic episodes served the purpose of sustaining a false self of boundless energy and creativity. Mr. A. would be encouraged to decide who he really wanted to be—a writer in fantasy or a persons able to obtain real satisfaction from real accomplishments and real relationships. His dreams would be examined and their themes explored. His omnipotent, messianic dreams would be interpreted as evidence of his ideal self and viewed in the light of reality. His dreams of being exposed would be interpreted as the fear engendered by his ideal self as a means of defending itself against exposure: “If you expose me, you will be embarrassed and humiliated.” Mr. A. would be supported through his fear of humiliation and the pain of realizing that he had been deceiving himself. His self-loathing would be interpreted as the activity of the pride system in defending his ideal self. As he begins to relinquish his ideal self, he will begin to discover and to mobilize his real inner self in directions that might not have been predictable at the beginning of treatment.

ERICH FROMM

Psychoanalyst Fromm (1900–1980) was often thought of as the archetypical neo-Freudian, the leader among those who emphasized that culture and social setting influence an individual's dynamics as much as instincts do (Fig. 6.3-12). Neither physician nor biologist, Fromm a native German received his doctorate in philosophy, sociology, and psychology from the University of Heidelberg in 1922. There he was exposed to a Marxist emphasis on how history shapes societies and how societies in turn, shape individuals according to economic needs. He was trained as a psychoanalyst at the Berlin Psychoanalytic Institute and then founded, with his wife, Frieda Fromm-Reichmann, the Frankfurt Psychoanalytic Institute. In 1933 he immigrated to the United States and in 1949 he moved to Mexico City to found an institute. In 1974 he moved to Switzerland where he died in 1980. As much a social critic as a personality theorist, he was later claimed by the existential and humanistic psychoanalysts. Fromm's intellectual agenda was the integration of Freud's theory of a dynamic unconscious with Karl Marx's theory of history and social criticism.

FIGURE 6.3-12 Erich Fromm. (Courtesy of Erich Fromm.)

Personality Theory For Fromm, two central facts dominate human behavior: the inevitability of separateness and the historical and social moment into which each person is born. He argued that every person struggles to recapture the state of blissful union that existed prenatally. From the moment the baby begins to recognize itself as a separate human being a titanic struggle begins, pitting the desperate anxiety of loneliness against the urge to fully express and actualize oneself, ultimately transcending the self. Most individuals find the loneliness too painful to bear and they suppress their striving for individuation in the service of maintaining the illusion of connectedness. They are socialized by their parents into the roles defined by the society into which they are born. Fromm actually used the term *symbiosis* years before Margaret Mahler used it to describe the universal human yearning for fusion, safety, and security.

Facing aloneness and choosing individuation leads to freedom and a productive life. However, true freedom is too terrifying for many persons who instead construct a series of illusions that engender a feeling of safety and security. They create a pseudoself, think pseudothoughts, and experience pseudofeelings in support of these illusions, thereby cutting themselves off from the fullness of their own inner lives. Fromm saw Freud's theory as a special case of his own more general ideas. The

illusions that Victorian society offered involved sublimation of sexuality and aggression in the service of social respectability. Social respectability, in turn, provided the illusion of acceptance and security. In other places and at other times, different solutions might be offered. Early in World War II, shortly after his escape from Nazi Germany, Fromm wrote about the willingness of people to give up their own freedom to serve an authoritarian society. In the 1950s he wrote of the pursuit of material acquisitions in the service of postwar productivity, leading to self-satisfied conformity. Fromm's most direct application of Marxism was in his hypothesis that individual development has paralleled historical development since the time that humankind freed itself from symbiosis with nature and embarked upon a unique path, evolving inevitably towards the Marxist utopia: the end of history in the universally humane society. However, Fromm departed from other Marxists who saw revolution as the only healthy response to an inevitably repressive society. He believed that even within an imperfect culture, individuals could face their terror, give up their pseudoselves and pseudothoughts, and choose to become themselves—encountering others who had made similar choices with love and mutuality. In order to achieve this, Fromm said four basic human needs must be met: *relatedness, transcendence, identity*, and a *frame of orientation*. Relatedness is the need to feel connected to other humans. Transcendence refers to rising above basic instincts. Identity is the need to feel accepted yet unique. Emphasis on the need for a frame of orientation led Fromm late in his career to an exploration of the constructive and destructive roles that religion may play in individual lives.

Theory of Psychopathology As a social philosopher and critic, Fromm did not really develop a systematic theory of psychopathology. He identified three major mechanisms of retreat from individuation. Some individuals, he said, may seek an authoritarian solution, trying to live through someone or something external to themselves, relying on that for their sense of adequacy. Others may become destructive, attacking anything that confronts them with their separateness and aloneness. Most individuals develop a conformist attitude, warding off the anxiety of experiencing their own intentionality by accepting socially offered thoughts, roles, and attitudes.

These mechanisms result in four different unproductive orientations or characters typical of twentieth-century society: receptive, exploitative, hoarding, and marketing. Fromm's economic perspective on available social roles is made evident here. The *receptive character* often appears to be cooperative and open; however, the primary agenda is to establish a passive relationship with a leader who solves problems magically. *Exploitative characters* are likewise interested in filling themselves up from the outside; however, they aggressively manipulate and usurp whatever will reduce their terror. *Hoarders* collect, store, and close in upon themselves, often being cold and aloof in their efforts to feel secure. *Marketers* treat themselves as a plastic commodity to be manipulated as needed in order to achieve externally validated success.

Treatment Fromm wrote nothing at all on the practice of psychotherapy; therefore, what is known is derived from anecdotal reports by those who studied with him or were treated by him. They report his emphasis on a tender and empathic inquiry into the self-deceptions and illusions created by patients in their efforts to ward off the anxiety of separateness and to maintain some sense of connectedness to significant others. He placed great emphasis on the tendency of unloved children to identify intensely with parental values in order to capture the magical safety they seem to offer. At a time when most psychoanalysts were preoccupied with detailed examinations of the instincts and defenses, Fromm contributed a sense of the range and richness of inner experience that underlies superficial adaptation. He contributed the sense that a new authenticity could be found by those willing to confront the truth about themselves with all its terror of aloneness.

In some respects, Fromm's ideas are uniquely applicable to Mr. A. who is clearly maintaining the illusion of his separateness by conforming to the socially sanctioned role of rebellious artist. His bipolar I disorder becomes an unexpected weapon in his battle with his terror of aloneness. The dread in the hotel lobby dream is probably the closest this fear comes to consciousness. Fromm would gently probe this dream and Mr. A.'s self-image. He would point out the chains created by Mr. A.'s seemingly rebellious independence. The self-destructive quality of the patient's lifestyle would also be confronted and investigated. Ultimately, Mr. A. would have to experience his own loneliness and face his terror of it in order to find true freedom, true intimacy, and an authentic self-definition.

HARRY STACK SULLIVAN

Sullivan (1892–1949) is generally acknowledged as the most original and distinctive American-born theorist in dynamic psychiatry (Fig. 6.3-13). Although rarely acknowledged explicitly since the late 1970s, most American psychiatrists make significant use of concepts and approaches that he developed. For many years the primary theoretical dispute within dynamic psychiatry circles was between the classical Freudians and the Sullivanians (or interpersonal psychoanalysts). When psychiatrists use the terms *parataxic distortion*, apply the concept of self-esteem, consider the importance of preadolescent peer groups in development, or view a patient's behavior as an interpersonal manipulation, they are applying concepts Sullivan first proposed.

FIGURE 6.3-13 Harry Stack Sullivan. (Courtesy of the New York Academy of Medicine.)

Sullivan graduated from medical school in Chicago in 1917. He spent from 1921 to 1930 in the Washington, D.C. area working with schizophrenia patients at St. Elizabeth's and then Sheppard and Enoch Pratt Hospitals where he developed a reputation as a remarkable clinician with an uncanny ability to communicate with floridly psychotic patients. He initiated the first of what would now be called therapeutic communities. Later, he entered private practice in New York, and eventually returned to the Washington area where he was involved in clinical, consulting, and teaching activities. In the 1920s and 1930s he wrote a number of papers on schizophrenia, later collected in *Schizophrenia as a Human Process*. His other books were compiled from his lectures by his students; most were published posthumously, which explains some of the density and seeming disorganization of his written work.

Personality Theory Sullivan rejected the Kraepelinian dogma of his day that dominated psychiatric thinking about schizophrenia. Sullivan would elucidate the meaning of passages of patient speech that Kraepelin presented as nonsensical. In searching for alternative understandings of psychosis he turned initially to Freud, but rejected his theories as increasingly rigid and dogmatic. Thus, he developed his own working theory of personality, psychopathology, and therapy.

Sullivan was very concerned that language could be misleading. He was very wary of self-reifying conceptualizations that led to rigid theories and tried to emphasize the psychiatrist as participant-observer in the clinical situation. By emphasizing this aspect of the role, he sought to keep observations as objective as possible, although he recognized the difficulty this presented in dealing with private emotional experience. What can be observed is the social interaction of patients; thus, he defined personality as the “relatively enduring pattern of interpersonal relations which characterize a human life.” From the outset, his focus was very different from the intrapsychic emphasis of psychoanalysis. By approaching psychopathology in this way, he necessarily created a field theory rather than a structural theory, characterized by temporal and interactive processes. Sullivan defined a “dynamism” as “the relatively enduring pattern of energy transformations,” that is, recurrent interpersonal behavior patterns.

Sullivan's theory is fundamentally one of needs and anxiety. Needs are defined as needs for satisfaction and needs for security. Anxiety occurs when fundamental needs are in danger of not being met and is the primary motivator of human behavior. Needs for satisfaction include physical needs (e.g., air, water, food, warmth, etc.); emotional needs include needs especially for human contact and for expressing one's talents and capacities. Because infants are utterly unable to meet their own needs, interpersonal relationships are a central concern. Decades before Margaret Mahler wrote of a symbiotic stage in infant development, Sullivan spoke of the “empathic linkage” between caretaker and infant, and described the complicated interaction of infants communicating tension and anxiety, arousing anxiety in the caretaker, leading to tender responses to the infant's needs. Failure to meet these needs results in loneliness and anxiety.

Sullivan defined security as the absence of anxiety. Thus, needs for security are defined as the need to avoid, prevent, or reduce anxiety. Because there is no such thing as a perfect mother or parent, anxiety is inevitable and becomes the primary driver for personality development. The self-system is defined by Sullivan as the

dynamism that is responsible for avoiding or reducing anxiety. Sullivan equated the self, identity, or ego with the individual's developed patterns for avoiding the discomforts that arise from the failure of others to meet one's fundamental needs. It exists, like all else, purely within an interpersonal framework. The self-system develops a set of mechanisms, called *security operations*, which effect this goal.

Security operations function within Sullivan's theory much as defense mechanisms do within psychoanalytic theory. The specific security operations, however, were defined interpersonally and Sullivan tried to link them closely to actual observation or experience. Some bore the same labels and definitions as Anna Freud's but Sullivan is best known for three contributions that bore his distinct stamp: apathy, somnolent detachment, and selective inattention. These were drawn from observing the way infants and young children reacted to painful interactions, such as scolding, with their parents.

The self-system accrues from ever-evolving interpersonal experiences, that is, fulfillment of needs for satisfaction as a result of the emphatic linkage with the mother. The most difficult experiences are not necessarily those involving failure to meet the child's needs, but the child's sensing of the caretaker's anxiety in the process of responding to those needs. This arouses anxiety in the child, promotes the need to establish a sense of security, and leads to evolution of the self-system and the development of security operations. The self-system is divided into three parts. The “good me” is a set of images, experiences, and behaviors associated with an unanxious, tender, empathic, and approving on accepting response from the environment. The “bad me” comes to be associated with ideas, actions, and perceptions that provoke anxiety and disapproval from caretakers. Some situations, however, provoke such intense anxiety that they are entirely disavowed and disowned; they become part of the “not me.” Eventually, the empathic linkage becomes unnecessary and the self-system operates autonomously within the individual, developing ever more subtle and complex ways to manage the person's anxiety.

Developmental Theories Sullivan had two theories of development, one cognitive, the other social. He postulated three developmental cognitive modes of experience whose degree of persistence into adulthood would be important in understanding psychopathology. The *prototaxic* mode, characteristic of infancy and early childhood, involves a series of disconnected, brief states experienced as totalities with no temporal relationship. In later life, mystical experiences and schizophrenic fusion represent persistent prototaxic experiences. *Parataxic* experience begins early in childhood as the self-system begins its more independent functioning. It too involves a series of momentary experiences; however, they are now recorded in sequence and with apparent connection to one another. They may be given symbolic meanings, but rules of logic are absent and coincidence plays a major role in how the world is perceived. The self-system uses this mode to seek effective anxiety-reducing behaviors and to repeat them, seeking sameness and predictability. Sullivan used this mode to explain transference, slips of the tongue, and paranoid ideation. The *syntactic* mode of experiencing is based upon the development of language and consensual validation. The world and the self is perceived within rules of logic, temporal sequencing, external validity, and internal consistency. Thinking about oneself as well as others becomes testable and modifiable based upon rigorous analysis of experiences in a variety of different situations. Maturity may be defined as extensive predominance of the syntactic mode of experiencing.

Social development is somewhat based on these evolving cognitive modes. However, disturbed interpersonal relationships may cause persistence of the more primitive (prototaxic or parataxic) ways of experiencing the world. Social development is characterized by the satisfaction needs, which are predominant, and the interpersonal sphere in which these and their resulting security needs are sought to be fulfilled. Each stage is also characterized by the primary “zone of interaction”—bodily areas through which the individual channels needs, anxiety, and relief—in interactions with the environment. These aspects of Sullivan's theory bear a superficial resemblance to Freud's genetic theory; however, Sullivan accorded them far less importance; they are mere conduits when compared to psychoanalytic libido theory.

Infancy spans birth to the onset of language and is characterized by the primary need for bodily contact and tenderness. The prototaxic mode predominates and the primary zones of interaction are oral and, to some extent, anal. Insofar as needs are fulfilled with a minimum of anxiety, the infant experiences euphoria and a sense of well-being. To the extent that some anxiety is commonly present in the caretakers, apathy and somnolent detachment will be regularly used as security operations, persisting into adult life as a basic detached and passive stance. If anxiety and inconsistency are severe, intense experiences of dread will persist, presenting in later life as the eerie, uncanny, bizarrely disruptive internal states seen in individuals with schizophrenia.

Childhood begins with the onset of usable language, continues until the beginning of school, and is characterized by the child's focus on the parents as other from whom praise and acceptance are sought. The primary mode of experience shifts to the parataxic and the most common zone of interaction is anal. The child needs an approving adult audience. This leads to a variety of learning—of language, behavior, self-control, and so on. It can also be observed in a variety of trial-and-error efforts by the child to find what pleases. Gratification leads to an expansive self-system with many facets of life associated with the “good me” and positive self-esteem. Moderate anxiety leads to chronic anxiety, uncertainly, and insecurity. Extreme anxiety results in giving up known successful behavior in favor of self-defeating patterns that fulfill others' expectations.

The juvenile era covers ages 5 to 8. The shift to syntactic cognitive modes begins and the interpersonal focus spreads to the peer group and outside authority figures. Peers and teachers have the opportunity to approve and accept behavior previously frowned upon within the family (e.g., talking dirty with one's friends). Interpersonal cooperation, competition, play, and compromise become the gratifying experiences. Juveniles learn to negotiate their own needs with a legitimate social concern without sacrificing their self-esteem in the process. The risks of excessive anxiety are either too great a need to control and dominate social situations or internalization of restrictive, prejudicial social attitudes.

Preadolescence, ages 8 to 12, marks the child's movement from peer group cooperation and competition based upon rules towards genuine intimacy with a chum. Sullivan saw this phase as a particularly important stage in which the give and take of the special friend could repair and undo distortions that resulted from excessive anxiety at earlier stages. This is the point at which the individual truly moves outside the family and engages in a free give and take with another person unfettered by the same dynamics. During this stage the major shift towards syntactic thinking takes place, although some distortions may persist into adolescence. The preteen years see the initiation of a capacity for attachment, love, and collaboration or their failure to develop in the face of excessive anxiety. Although sexual exploration may be a part of the chum relationship, Sullivan did not see sexuality as a central element in this developmental phase.

Adolescents, beginning at puberty, are seen to have concerns similar to those of preadolescents, except that lust is added to the interpersonal equation. Thus, the same needs for a special sharing relationship persist but shift to the other sex for their outlet, whereupon a major opportunity for learning or severe anxiety begins. As the person faces culturally defined stereotyping, many new opportunities for social experimentation may lead to consolidation of self-esteem or self-ridicule. The struggle to integrate lust with intimacy is accomplished by painful trial and error. If this is completed with the self-system relatively intact, the later years of adolescence are an opportunity to expand the syntactic mode to such areas as a consensual view of interpersonal relations, values and ideals, career decisions, and social concerns.

Theory of Psychopathology Sullivan abhorred diagnostic labeling for being unhelpful, overly restrictive, dehumanizing, and used primarily to impress patients and colleagues. In discussing schizophrenia, he said, “We are all much more simply human than otherwise.” Thus, he sought to understand the fundamental human process within his patients, especially his sickest ones. He saw psychopathology as resulting from excessive anxiety arresting development of the self-system and thereby limiting both opportunities for interpersonal satisfaction and available security operations. He viewed psychiatric patients as struggling to maintain their self-esteem with very limited means. To understand them, the developmental phase at which they operate has to be gauged and the interpersonal needs they express have to be understood.

Sullivan believed that several different factors could play a role in the particular form that these disturbances might take. The level of anxiety at particular developmental stages could lay the groundwork for a developmental arrest. Basic cognitive capacity might play a role in the choice of security operations relied upon or retained. The degree of success achieved interpersonally combined with whatever capacities are used would affect later success. Finally, the chance occurrence of stresses encountered during life was deemed a factor. Thus, Sullivan theorized that anyone might develop schizophrenia, even persons with relatively successful developmental histories, should their chosen defenses fail dramatically and their life stresses mount in the extreme. However, it was more likely that schizophrenia patients would be highly vulnerable along all four dimensions whereas others with greater developmental strengths might become obsessive, hysteroid, schizoid, or paranoid.

Interpersonal Psychotherapy Sullivan emphasized that the psychiatrist is a participant-observer in all interactions with patients. He thought deeply and extensively about the nuances and opportunities involved in this unique situation. By interacting actively with patients, verbal and nonverbal expressions of recurrent interpersonal patterns become apparent. These observations then inform the therapist's further behavior, thereby creating the opportunity for change. This process occurs over seconds and over months and years as the psychotherapy unfolds. Sullivan saw this perspective as an antidote to what he perceived as the wrongheaded emphasis on objective neutrality embodied by the “blank screen” model of psychotherapist behavior. He argued that parataxic distortions emerge in all interactions, not only in the classical analytic situation. This differing view of transference, and of it being a universal human process were among the core debates for

decades between classical analysts and interpersonal analysts.

Sullivan saw therapy as elucidating the patient's interpersonal patterns, exploring their utility in the service of the patient's needs, and considering alternative, more favorable possibilities. Thus, he shared the ego psychologists' understanding that even the sickest behavior was the best adaptation available at a given moment to the patient. He emphasized the experiencing of the distortions, the needs, the patterns, and the potential changes within the ongoing interaction with the therapist. He saw great power in the very entanglement of the therapist with the patient and recognized the ability of a skilled therapist to manage the interpersonal process to reveal patterns and to shape the patient's emotional experience. However, he constantly emphasized and respected the ultimate autonomy of his patients, who could still in the end choose not to reshape their approach to the world.

Sullivan viewed psychotherapy as divided into four distinct stages: inception, reconnaissance, detailed inquiry, and termination. Inception involves the very beginning, often only a part of the first interview, during which the contract and roles are stipulated. Reconnaissance might go on for as many as 10 to 15 sessions, during which the therapist identifies the patient's recurring patterns and assesses their adaptive and maladaptive qualities. The detailed inquiry is a very lengthy process of exploring the patient's thoughts, feelings, and memories; evaluating and re-evaluating data from earlier stages seeking to recognize, clarify, and change persistent parataxic distortions. The recurrent patterns are discussed within the context of the person's developmental history, needs, anxieties, failures, and successes. There is often much ongoing interchange between patient and psychiatrist as feelings and perceptions are validated or questioned within the context of mutual emotional interchange within each session. Termination is a product of the evolving contract and understanding between the patient and therapist and may reflect either extensive or limited goals. Sullivan emphasized the constant reassessing of goals by the therapist and the power of the ongoing negotiation and renegotiation of the therapeutic contract as a means to reveal and change parataxic distortions. The ultimate goal of psychotherapy is to experience as much as possible within the syntactic mode and to broaden the repertoire of the self-system. To the extent that this is achieved, individuals are in a position to become responsible for their ongoing growth through subsequent interpersonal interactions.

Sullivan would see Mr. A. as probably arrested in childhood, when his fear of displeasing his mother led him to give up healthy self-esteem strivings for independence in favor of a distant yet dependent position. He used drugs and psychosis as escapes in order to maintain some degree of self-esteem. His “good me” consists of his debating and his intelligence. His “bad me” is expressed in his rebelliousness whereas the “not me” seems to encompass issues of closeness, independence, and constructive engagement with others and with life's tasks. In therapy, the reconnaissance phase would be extremely important, identifying the moment-to-moment interactions through which Mr. A.'s security operations interfere with his attempts at constructive independence. His bemusement and detachment would be identified, as would his escape into prototaxic thought through drugs and noncompliance with medications. His acceptance of his mother's management of his trust fund would be noted as well. A Sullivanian therapist would actively interact, empathically identifying and confronting Mr. A.'s ways of avoiding authenticity and constructive interaction. Once identified, their meanings would be gently probed as would the feelings associated with them. The terror of displeasing his mother and its effects on Mr. A.'s ongoing interactions with the therapist would be addressed. Interactions would be examined for their consequences, with the assumption that the outcome had bearing on the motivation. Finally, Mr. A. would be encouraged to try out different ways of relating to the therapist as a prelude to restructuring his outside relationships. He would be encouraged to think rationally about his circumstances, to seek a good peer group, develop close friendships, and gradually move away from his near-exclusive focus on his immediate family for all of his interpersonal needs.

ERIC BERNE

Berne (1910–1970) was an American original in both style and substance. He worked in the San Francisco area most of his career, breaking with psychoanalysis in the mid-1950s, but never becoming antianalytic as did many of his followers (Fig. 6.3-14). Like many others, he felt the need to develop briefer treatments than were offered at the time.

FIGURE 6.3-14 Eric Berne. (Courtesy of Wide World Photos.)

A group gathered around him that came to be known as the San Francisco Transactional Analysis seminars. Through weekly discussions of clinical cases and social and political issues, Berne gradually refined his theory. Berne was wry and provocative in his approach to human behavior. His approach contributed much to what would now be called “pop psychology,” although his particular popularity and faddishness has long since faded. Few clinicians now call themselves transactional analysis therapists; nonetheless, Berne's ideas remain useful in grasping hidden agendas in human interactions.

Personality Theory For Berne the primary motivator of all human behavior is the need for “strokes”—attention, recognition, and response from others. Early survival depends on adequate physical contact, stimulation, and nurturance. This need remains strong, but later becomes more symbolic and interpersonal. Children learn rapidly what works within their family and practice it extensively. This led to one of Berne's more widely quoted observations: “Negative strokes are better than no strokes at all.” People evolve ways of interacting with their world to obtain regular strokes in whatever way possible, and in whatever way they have been taught to define a stroke (e.g., sympathy in response to chronic depression may provide such gratifying attention that the depression cannot be given up). So great is the need for regular stroking that blatantly destructive actions persist in the face of insight, recognition, and enduring psychological or physical pain. Like Adler and Sullivan, Berne suggested that hidden social needs motivate human behavior to the extent that, with rare exceptions, there is an interpersonal hidden agenda in all human activity. For Berne the unit of observation was the *transaction*, the short-term process of individuals interacting with each other. He spent much energy analyzing transactions to try to discover patients' definitions of strokes and their preferred mechanism for obtaining them. He noted that most people engage in very predictable, stereotyped, repetitive transactions—the content may vary from situation to situation, but the form tends to be quite rigid. He called these transactions “games,” and his bestseller, *Games People Play*, captured the imagination of the American public in 1964. Some games are harmless, some are socially encouraged, many have destructive elements or at least limit opportunities for more gratifying relationships (intimacy), and some are highly destructive. A common, socially accepted game is cocktail party flirtation, which is ordinarily pleasant and harmless, but depending on the intensity, frequency, and seriousness with which it is played may result in inability to experience intimacy, disrupted marriages, or even physical harm.

Berne divided the human psyche into three primary parts; *child, parent*, and *adult*, with two of those further subdivided (Fig. 6.3-15). He called these *ego states*. An ego state consists of characteristic body language, voice qualities, verbal productions, and affective experience. The Child represents the persistence of childlike experience and expression in all persons. It is divided into the *natural child*, the ego state in which the spontaneity, joy, and intuitive perceptiveness of young children persists in all adults; the *adapted child*, the part that is compliant and cooperative, and the *rebellious child*, the repository of that part of each person prone to fight authority, challenge accepted wisdom, and struggle for autonomy. The *Parent* is the residue of internalized parental messages and injunctions. It is divided into two parts, the *critical parent* and the *nurturing parent*. The Critical Parent bears some resemblance to Freud's superego, embodying rules, values, instruction, criticism, and restrictions. The Nurturing Parent is the internalization of positive caring experience, the memory of loving interactions. The Adult is a purely rational, data-processing element that is objective, calculating, and weighs options and estimates probabilities.

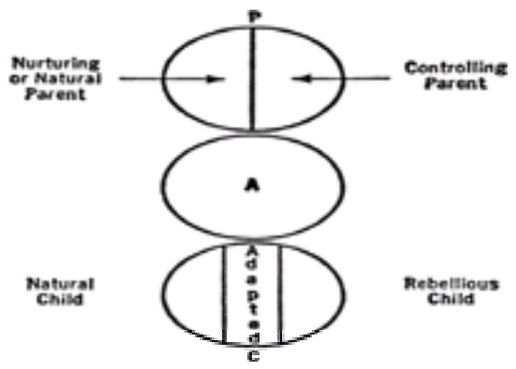

FIGURE 6.3-15 Eric Berne's descriptive model of the personality. (Reprinted with permission from Berne E: *What Do You Say After You Say Hello?* Random House, New York, 1972.)

Mental health is the flexible availability of all ego states with no one predominating. Excess critical parent produces guilt and depression, but insufficient Critical Parent produces sociopathy. Excess nurturing parent produces a narcissistic laziness whereas insufficient Nurturing Parent causes an inability to soothe oneself and to maintain self-esteem. Excess Adult results in a cold, overly rational person, but insufficient adult leaves individuals unable to balance the various internal forces in their lives. Too much natural child may result in irresponsible, behavior but not enough depletes the ability to experience joy in living. An overabundance of rebellious child results in self-destructive battles with no constructive purpose; however, insufficient rebellious child may result in an overly conformist stance. Excess adapted child prevents appropriate striving toward autonomy; insufficient adapted child prevents participation in group or hierarchical efforts.

Berne placed great emphasis on the child's ability to intuit parental messages and instructions, especially those communicated nonverbally and unconsciously. In this way the growing child might encode a conscious verbal instruction in the critical parent (e.g., be strong, be perfect, be smart), while internalizing a more powerful, unconscious message in the child (don't grow up, don't leave me, don't surpass me). Usually, a model for carrying out the instruction is preserved in the adult (be passive, drink alcohol, run around with women, act crazy). Together these make a *script*. Berne emphasized the active role of the child in searching for the messages and in accepting them. This was crucial for him because it emphasized the individual's responsibility in deciding to follow the script, even though this might have occurred at a very young age. The entire basis for psychological change lies within the person's capacity, having once accepted the script, to later reject it. However, Berne was impressed by the force with which individuals would play out the script throughout their lives. Scripts come in all varieties, ranging from the successful to the utterly self-destructive. Thus, for Berne, the business of human living involved carrying out one's script according to the transactions learned as a child.

Theory of Psychopathology Given this view of human nature, Berne's understanding of psychopathology is based on the adaptiveness of a person's games and script, and the capacity for adaptive use of all ego states. Written at a time when American psychiatry rejected phenomenological diagnosis, Berne's theories are difficult to accept. However, he did describe numerous clinical situations and case histories that are familiar to all psychiatrists and he analyzed them according to their scripts. He was very impressed with the role of fantasy and fairy tales in children's development and would often ask patients what their favorite story was as a child. He would then search for the hidden identifications and messages within the story to help discern the patient's script. Insofar as he developed a nosology it lay in the scripts encouraged by culturally sanctioned fairy tales. For example, Cinderella is about a young girl waiting to be rescued by a fairy godmother or a charming prince and unwilling to assert herself against unjust authority. Little Red Riding Hood is about a girl who likes to chat with wolves and gets into trouble trying to please everybody. Men, he found, would identify with the wolf or Prince Charming, and were unable to regard women as real persons.

A separate nosology of games was also developed and catalogued by Berne. Examples from this nosology include cocktail party flirtation (Rapo I), which concludes with the man and woman, having experienced the stroke of feeling attractive and attracted, move on to other conversations, seeking other persons to attract. Rapo II usually ends with a line like, “What kind of girl do you think I am?” and Rapo III ends up with a painful, destructive love affair. Another game is “Why Don't You... Yes, But.” In its mild, socially acceptable form one person presents a problem, successfully enlisting the advice of many others; however, for every proferred solution the person has a ready explanation of why it will not work. The stroke (or payoff) is the attention and the feeling of superiority achieved by defeating others. The second-degree version of this game produces a chronic, helpless depression; the third-degree version may lead to the back wards of a state hospital.

There was no rigid mapping of his nosology of games and of scripts and Berne never claimed that either was exhaustive. They were intended as clinical examples to guide the psychotherapist's thinking in evaluating each patient who would present unique script and transactional issues. Individuals were assessed according to the straightness of their transactions (Were there hidden communications and invitations emanating from different ego states?), the adaptiveness of their scripts, and the predominant ego state and its effects. Thus, Berne might describe a compulsive person as having too much critical parent, or a histrionic person as having too much natural child. In the final analysis, however, he saw his catalogues and nosology as poor substitutes for careful, clinical study of individual patients.

Treatment Having been trained in psychoanalysis, Berne used many psychoanalytic techniques but he applied them very differently from traditional psychoanalytic methods. Much of his clinical work was done in groups with the therapist being very active in confronting and interpreting the interpersonal behavior of the patient. He emphasized the initial contract with the patient and the setting of clear, concrete goals. He encouraged therapists to inquire very early on in therapy, “How will we know when our work is finished?” If a clear and specific answer was not forthcoming, he would inform the patient that therapy had not yet begun. The focus would then be on the patient's difficulty in defining a recognizable endpoint for the work. Berne's approach was heavily informed by the psychoanalytic concept of resistance but he applied it in a very different manner. Berne would examine any treatment goal defined by patients for evidence that it was a way to continue playing games or pursuing scripts more effectively, rather than giving them up entirely.

Like Sullivan, Berne regarded the ongoing interaction between patient and therapist as the key ingredient of psychotherapy. The therapist's job was to interact actively with the patient; recognize the ego states, games, and scripts being enacted; and to counter them using confrontation, interpretation, or various other interpersonal maneuvers that were designed to thwart the enactment and to confront the patient with a choice and an opportunity to relate differently. He emphasized simple direct statements using everyday words to stimulate affective experience and interaction. Commonly, interactions would begin with his inquiry, “What do you want to work on today?” Observers sometimes described his work as individual therapy within a group context because the other patients, who were silent observers, would often respond emotionally to the interaction.

Berne placed great emphasis on the role of individual responsibility for one's life and experience, which he felt had been obscured by the emphasis on psychic determinism. He constantly communicated the opportunity for choosing to continue or discontinue the patterns or habits encouraged as a child. He also showed patients how they invite others to behave in various ways, thereby creating their own interpersonal reality. Health lies in recognizing one's option for deciding which invitations to offer to others and which to accept from others, even though these invitations may be communicated unconsciously and nonverbally.

Mr. A. is clearly playing a very destructive game and living out a particularly unproductive script. The script seems to contain the messages: “Don't ever leave me,” and “Act crazy.” The message: “You must be special” may also exist. The games used might be called “I'm out of control,” “Helpless,” “Artiste,” and “I'm better than you are.” A transactional analyst would place Mr. A. in a group and work with him on these games and scripts. Many such therapists have productively combined Berne's theories with Fritz Perls' empty chair technique. Mr. A. might be asked to carry on imaginary dialogues with his mother, professor, his illness, marijuana, and admired writers of fiction, with Mr. A. assuming both sides of the dialogue. The transactions he employs would be elucidated in this way and he would be invited to recreate the transaction more constructively. His internalization of his mother as an excessively Critical Parent would be identified, and the fantasy dialogues would be used to enable more internalization of both his mother's and father's nurturing qualities, which would contain other script messages such as “grow up” or “be successful.” Similarly, his excess Rebellious Child would be demonstrated and efforts to bring it under more adult control would be made.

In looking at the formulations and treatment plans each of these theorists might propose for Mr. A. significant, even contradictory interpretations become apparent. Commonalities exist but are expressed in different words and concepts. Each of these great thinkers has contributed to the overall understanding of the psychodynamics of psychopathology and treatment.

SUGGESTED CROSS-REFERENCES

Many sections of this textbook refer to the theorists described above. Some are included in Section 55.1 on the history of psychiatry. Others are referred to Section

6.1 on psychoanalysis, Section 6.4 on theories of personality and psychopathology: schools derived from philosophy and psychology, and Section 25.1 on the history and current theoretical concepts in psychosomatic medicine.

SECTION REFERENCES

Adler A: In *The Individual Psychology of Alfred Adler: A Systematic Presentation in Selections from His Writings*, HL Ansbacher, RR Ansbacher, editors. Basic Books, New York, 1956.

Adler A: In *Problems of Neurosis: A Book of Case Histories*, P Mairet, editor. Harper & Row, New York, 1964.

Alexander F: *Psychosomatic Medicine*. Norton, New York, 1950.

Alexander F, French T, Pollock GH: *Psychosomatic Specificity*. University of Chicago Press, Chicago, 1968.

Berne E: *Games People Play*. Grove, New York, 1964.

Berne E: *What Do You Say After You Say Hello? The Psychology of Human Destiny*. Bantam, New York, 1972.

Fromm E: *Escape From Freedom*. Avon Books, New York, 1965.

*Greenberg JR, Mitchell SA: *Object Relations in Psychoanalytic Theory*. Harvard University Press, Cambridge, MA, 1983.

Horney K: *The Neurotic Personality of Our Time*. WW Norton, New York, 1937.

Horney K: *Neurosis and Human Growth*. Norton, New York, 1950.

*Jung CG: *Two Essays on Analytic Psychology*. Princeton University Press, Princeton, NJ, 1966.

Jung CG: *The Practice of Psychotherapy*, ed 2. Princeton University Press, Princeton, NJ, 1966.

Lieberman EJ: *Acts of Will: The Life and Work of Otto Rank*. Free Press, New York, 1985.

McGuire W, editor: *Analytical Psychology: Notes of the Seminar Given in 1925 by C.G. Jung*. Princeton University Press, Princeton, NJ, 1989.

McLynn F: *Carl Gustav Jung*. New York, St. Martin's Press, 1997.

*Meyer A: *Collected papers of Adolf Meyer*, 4 vols. Johns Hopkins University Press, Baltimore, MD, 1948–1952.

Meyer A: *Psychobiology: A Science of Man*. Charles C Thomas, Springfield, IL, 1957.

*Mullahy P: *Psychoanalysis and Interpersonal Psychiatry: The Contributions of Harry Stack Sullivan*. Science House, New York, 1970.

Noll R: *Aryan Christ: The Secret Life of C.G. Jung*. Random House, New York, 1997.

Rado S: *Psychoanalysis of Behavior*, 2 vols. Grune & Stratton, New York, 1962.

Rank O: *The Trauma of Birth*. Harper & Row, New York, 1973.

*Reich W: *Character Analysis*. Farrar, Straus & Young, New York, 1949.

Sharaf M: *Fury on Earth: A Biography of Wilhelm Reich*. St. Martin's Press/Marek, New York, 1983.

Sullivan HS: *The Interpersonal Theory of Psychiatry*. WW Norton & Co, New York, 1953.

Stevens A: *On Jung*. Routledge, London, 1990.

Wagner J: *USATAA Articles—Violence in America*. Available at: http://usataa.org/articles_violence.html. Accessed July 15, 1999.

Walz JD: *Carl Jung: Anthology*. Available at http://www.enteract.com/~jwalz/Jung/. Accessed August 11, 1998.

6.4 APPROACHES DERIVED FROM PHILOSOPHY AND PSYCHOLOGY

PAUL T. COSTA, JR., PH.D. AND ROBERT R. MCCRAE, PH.D.

Few topics are of more interest to people than people, but an understanding of human nature is not inborn. Children must learn how to interpret others' behaviors and motives, and they do this so well that by adulthood they have acquired the use of hundreds of words to describe themselves and others, and have developed implicit theories of personality. Formal, explicit theories—accounting for the universals of human nature and the reasons for individual differences—have a long intellectual history in philosophy and theology. In the twentieth century the study of personality was undertaken by psychologists, and after an initial period of vigorous but speculative theorizing a body of sound empirical findings has emerged. No single theory of personality is universally accepted today, but there is increasing agreement on the facts that must be explained.

After a sampling of philosophical views of the person, this chapter provides an overview of psychological theories, grouped under the traditional rubrics of behavioral, humanistic, and trait approaches. Because of its importance in contemporary personality research, special attention is given to the trait perspective.

PHILOSOPHY

The recognition of individual differences is probably as old as human culture, but differences in personality were long confounded by differences in status, class, or caste. Individuals from a higher status were assumed to be superior human beings, more sensitive, honorable, and wise. They were endowed with these characteristics either by education, bloodline, or—in the view of Indian philosophy—by moral behavior in a previous life.

Plato The first major account of personality in Western thought was provided by Plato (circa 428–348 BC). In *The Republic*, Plato made extended comparisons between the constitutions of different states and the constitution of the soul. Just as every state must have peasants and artisans, soldiers, and rulers, so each individual has appetites (for food, sex, and so on), passions (for honor and advancement), and reason. The relative strength of these three determines character as well as fitness for a particular place for the individual in society. Intelligent and thoughtful people ought to rule, passionate people should be chosen to defend the state, and dull and spiritless individuals, lacking reason and passion, should be given the menial chores of agriculture and industry.

Plato assumed that these psychological characteristics, like physical strength or musical talent, were largely inborn, but he regarded them as a property of the individual, not the individual's social status. Although he assumed that high-status citizens would normally bear children of the greatest potential, he specifically acknowledged that there would be exceptions, and in his ideal state children of the lower classes would be promoted and those of the higher classes demoted on the basis of their own merit. Here personality would be the basis of the social order, not vice versa. His most radical extension of this idea was to argue that women should be given social equality and allowed to become soldiers and rulers if they possessed the necessary mental and physical qualifications.

One of the recurring questions in personality theory has been the relative importance of nature versus nurture. Trait psychologists—particularly those interested in the study of temperament—have frequently pointed to innate differences in personality whereas behaviorists and psychoanalysts have emphasized the formative influences of the environment and early childhood experiences. Although Plato acknowledged the importance of inborn potential, he was also keenly aware of the influence of education. Much of *The Republic* is devoted to his views on the effects of physical exercise, mental instruction, and poetry and music on the development of personality. Present-day concerns about the influences of television and rock music on children follow in this tradition.

Like most philosophers, Plato was more concerned with understanding virtue and vice than psychopathology, but because he considered vice to be the result of weak or corrupted nature rather than free but evil choice, his discussions of character can be viewed as early descriptions of what might now be viewed as personality disorders. Just as there are better and worse forms of government for states, so there are also better and worse configurations of reason, passion, and appetite. Plato described five types, corresponding to five forms of government. The ideal of mental health, corresponding to government by wise rulers, is one in which reason holds passions and appetites in check. The arrogant and ambitious type has an excess of passion or pride; in the avaricious type there is an excess of appetite. However, in both these types reason still retains some authority; for example, avaricious individuals can control most of their appetites in order to indulge their desire for money. In the fourth, self-indulgent, type, appetites are undisciplined, and in the debauched fifth type (corresponding to a despotic government), reason is completely disregarded, and a clearly psychopathological state is reached: “Thus, when nature or habit or both have combined the traits of drunkenness, lust, and lunacy, then you have the perfect specimen of the despotic man.”

Aristotle Plato's vivid but rough typology was succeeded by Aristotle's (384–322 BC) detailed analysis of human character in the *Nicomachean Ethics*. Courage, temperance, generosity, pride, ambition, irascibility, friendliness, boastfulness, and shame were all defined and distinguished. Aristotle attributed pathological variations on these traits to innate defects or to disease processes: “Among all the excesses of foolishness, cowardice, intemperance and irritability some are bestial, some diseased. If, e.g., someone's natural character makes him afraid of everything, even the noise of a mouse, he is a coward with a bestial sort of cowardice.” Aristotle considered variation within the normal range the result of training and habit, and thus subject to praise or blame.

Aristotle's basic moral precept is the “golden mean.” He argued that extreme standing on either end of a trait dimension should be avoided. Thus, both stinginess and extravagance are vices whereas generosity is a virtue; similarly, vanity and humility represent excessive or insufficient self-esteem. This theory continues to influence some modern notions of psychopathology, in which marked deviation from the norm in either direction is considered pathological. Individuals excessively concerned with social attention may be regarded as having a histrionic personality disorder; those insufficiently concerned with social attachments may have a schizoid disorder.

Aristotle carried conceptual analysis to a level that has seldom been surpassed, distinguishing, for example, between the superficially similar qualities of temperance and self-control: individuals are temperate if they have healthy and moderate impulses that require little control whereas individuals have self-control only if they have immoderate appetites that are nevertheless held in check. Such considerations allowed Aristotle to form rational taxonomies of traits, which anticipate the empirical taxonomies proposed by twentieth-century factor analysts.

Immanuel Kant and Arthur Schopenhauer The last great period of the Western philosophical tradition begun by Plato and Aristotle was inaugurated at the end of the eighteenth century by Immanuel Kant (1724–1804) whose critical philosophy forms a transition between purely rational approaches to human nature and the empirical sciences that followed. One of Kant's last works, *Anthropology from a Pragmatic Point of View*, considered both natural and moral variations in character, and reintroduced the Roman physician Galen's taxonomy of choleric, phlegmatic, sanguine, and melancholic types to modern psychology.

Sciences are usually distinguished from philosophy by their reliance on empirical data, but it should not be imagined that philosophers made no use of experience. On the contrary, like many of the clinicians who offered psychological theories of personality, philosophers based their ideas heavily on their observations of human nature. A striking instance of this is provided by Arthur Schopenhauer (1788–1860), a nineteenth-century follower of Kant whose dark view of the world as a place of purposeless striving had an extraordinary influence on early personality theorists, including Freud and Jung.

One of the central beliefs of Western thought has been that human happiness or misery is the result of external conditions—what we now call “quality of life.” But Schopenhauer's acute observations, reported in *The World as Will and Representation*, led him to propose “the paradoxical but not absurd hypothesis that in every individual the measure of pain essential to him has been determined once and for all by his nature.... His suffering and well-being would not be determined at all from without, but only by... what is called his temperament.”

In support of this view Schopenhauer noted that wealth and power do not make us happy, “for we come across at least as many cheerful faces among the poor as

among the rich.” He also argued that the effects of great misfortunes or successes are short-lived, and that for the most part our evaluations of the external causes of our state of mind are illusory attributions.

> We often see our pain result only from a definite external [cause] and... believe that, if only this were removed, the greatest contentment would necessarily ensue. But this is a delusion.... [The pain] would appear in the form of a hundred little annoyances and worries over things that we now entirely overlook.

Recent scientific research on psychological well-being has confirmed this account in every detail: well-being is chiefly a function of enduring personality dispositions; wealth, social class, and other markers of the objective quality of life are virtually unrelated to subjective happiness; and processes of adaptation quickly return individuals to their own characteristic baseline of happiness after favorable or unfavorable life events. Schopenhauer's observation is also consistent with recent evidence on the heritability and lifelong stability of many mental disorders.

However, reason and insight are not enough. On the basis of his own experience and the examples of history, Schopenhauer also concluded that “man inherits his moral nature, his character, his inclinations, his heart from the father, but the degree, quality, and tendency of his intelligence from the mother”—a conclusion not supported by the findings of modern behavioral genetics. Psychology broke from philosophy over precisely this need to seek empirical verification of hypotheses, but it carried with it concepts and insights accumulated over two millennia of profound thought about human nature.

BEHAVIORAL AND SOCIAL LEARNING APPROACHES

Theories of Personality Behaviorism as a school of psychology grew up in reaction to the prevailing mentalistic model, in which introspection was used to determine the contents and operations of consciousness. John B. Watson (1878–1958) proposed that a scientific psychology should confine itself to an examination of observable behavior and explain all human conduct in terms of stimuli and learned responses. Ivan Pavlov's experiments with conditioned responses offered hope that such a science could be successful, and theorists such as Clark L. Hull (1884–1952) provided elaborate mathematical models of learning.

Radical Behaviorism Certainly the most influential behaviorist, and perhaps the most influential psychologist of the century, was B.F. Skinner (1904–1990). Skinner's basic concept was *operant conditioning*, in which behaviors are a function of the organism's history of reinforcement. The observation that animals can be taught tricks by giving them rewards and punishments is nothing new. However, behaviorists like Skinner refined and systematized this idea, using elegant experimental designs to determine the effects of the amount and schedule of reinforcements, the use of positive and negative rewards, and the difficulty of the discriminations required. Behaviors could be shaped, maintained, or eliminated by the judicious use of these principles.

Skinner was a radical behaviorist, a purist who denied not only the scientific value but even the existence of mind. Further, he avoided any neurophysiological or psychophysiological theorizing, preferring to study the “empty organism.” Individual differences were ignored in understanding basic phenomena and were explained in individuals as being the result of different histories of reinforcement. Even differences between species were neglected: Skinner believed that the pigeon provided an adequate model for the study of learning in all organisms, and he and his followers were in fact able to replicate many of their animal findings using human subjects. Much of Skinner's success can be attributed to his single-minded pursuit of a highly circumscribed set of variables.

Skinner's view of personality was, predictably, a reductionistic one. As he stated in *About Behaviorism:*

> a self or personality is at best a repertoire of behavior imparted by an organized set of contingencies. The behavior a young person acquires in the bosom of his family composes one self; the behavior he acquires in, say, the armed services composes another.

This position is rejected by humanistic psychologists, who attribute more choice and control to the individual, as well as by trait psychologists, who see consistencies of behavior that appear to transcend the consistencies of the reinforcing environment. Many personality psychologists have argued that controlled laboratory experimentation is a poor basis for theories of personality, because individuals play a large role in selecting and shaping their own environments. Skinner's radical behaviorism was rejected or modified by many later learning theorists who acknowledged the power of conditioning but also recognized differences among species and among individuals within a species.

Social Learning Theory One of the most distinctive features of human organisms is their use of speech, which makes possible elaborate thinking and planning as well as complex social interactions. In recent decades learning theorists have increasingly emphasized social and cognitive processes. Among the most important have been Julian Rotter and Albert Bandura both of whom have offered versions of social learning theory.

Rotter's theory proposes that human behavior is guided not only by the actual history of reinforcement, but also by plans, goals, and expectations of success. Individuals will perform a behavior if they believe it is likely to lead to a valued goal, based upon their past experiences in general and in similar situations. Individuals with a history of success are likely to have a generalized expectancy that they can control their lives; they are described as having an *internal locus of control*. At the opposite extreme are those whose prior efforts have been generally unsuccessful; they come to believe that rewards and punishments are a matter of luck or the arbitrary decisions of powerful others. Such individuals are said to have an *external locus of control*. Locus of control has been one of the most popular variables in personality research, employed in numerous studies that generally support Rotter's theory.

Bandura's version of social learning theory also acknowledges the importance of internal cognitive processes. Individuals learn not only on the basis of their own experience, but also through vicarious reinforcement from observing others. Bandura's demonstration of modeling effects in experiments conducted in the 1960s gave scientific legitimacy to the social learning perspective.

Both Rotter's and Bandura's theories are general theories of behavior, not specifically theories of personality. However, for them personality is more than a collection of learned behaviors. The total pattern of experience will lead to a generalized expectation of reinforcement or to a general sense of self-efficacy that can be considered the central individual difference variable. People are to be characterized primarily on the basis of their beliefs in their own ability to control their lives, because these beliefs powerfully determine the effort they make to adapt to their surroundings.

Social-cognitive approaches to personality are among the most influential for current research in personality. These approaches focus on individuals understanding of themselves and how these self-appraisals shape goals, plans, and behaviors. Because of their origins in social learning theory, these theories tend to emphasize the role of the environment, pointing out that an individual's sense of self varies from setting to setting. Because of their ties to social theory, these theories usually explain personality in terms of the effects of social interactions. For example, Hazel Markus has suggested that individuals have a number of *possible selves*—conceptions of what one is or could be—which result from the messages significant others provide. For such theorists, concern has moved beyond the social learning of specific behaviors to the learning of entire identities.

Theories of Psychopathology Psychoanalytic theories of personality grew out of attempts to understand psychopathology, and thus the two are intimately connected. By contrast, behavioral approaches have focused on the general principles by which behavior is acquired and maintained, and psychopathology is usually treated as an area of application. Learning theories have had much more influence on methods of psychotherapy than on theories of psychopathology itself.

Behavioral approaches might suggest two different classes of explanations for psychopathology: psychopathology might be related to the mechanisms of learning or it might be considered the result of learning behaviors that are maladaptive or socially unacceptable.

In the 1950s Hans Eysenck (1916–1997), a psychologist who has figured prominently in both learning and trait schools of personality, proposed that individual differences in the dimension of introversion-extraversion determined the ease with which individuals could acquire conditioned responses, which in turn determined the form of psychopathology to which they were prone. Very introverted individuals, he proposed, were easily conditioned and thus acquired many inhibitions. They were predisposed to the development of depressive, anxious, and obsessive-compulsive disorders. By contrast, extreme extraverts were considered to be resistant to conditioning and were likely to develop hysterical and psychopathic disorders. In later versions of Eysenck's theory, psychopathic disorders were grouped with psychotic disorders and linked to a different dimension of personality, called *psychoticism*.

Most behaviorists have viewed the laws of learning as universal processes and considered psychopathology to be the result of normal learning processes. In the 1920s Irena Shenger-Krestounika taught a dog to discriminate between a circle and an ellipse as a cue for food. When the ellipse was made increasingly circular, the dog's ability to discriminate between them was taxed, and the dog began to struggle, squeal, and bite. Ivan Pavlov dubbed this an “experimental neurosis,” and

proposed that human neuroses might have parallel causes.

Probably the most famous attempt to explain psychopathology in learning theory terms was provided by John Dollard and Neal Miller. Dollard, an anthropologist, and Miller, an experimental psychologist, shared an interest in psychoanalysis. Their goal was to translate psychoanalytic concepts into the more testable terminology of learning theory. Consider, for example, the central psychoanalytic notion of *repression*. Sexual behaviors in the child might be punished by parents, and the child might learn to associate the behaviors with pain. By stimulus generalization even the thought of the behaviors would elicit anxiety, and cognitive processes that blocked these thoughts would lessen anxiety and thus be reinforced. Eventually, the thoughts would be effectively barred from consciousness.

Many behaviorists who did not share Dollard and Miller's enthusiasm for psychoanalysis followed their lead in attempting to explain psychopathology in terms of principles of learning. Phobias in particular were easily explained as conditioned responses reinforced by avoidant behavior. Similarly, compulsions could be understood as a kind of self-reinforcing behavior: each time the compulsive act is performed, the anxiety associated with not performing the act is reduced, increasing the probability that the behavior will be repeated.

Social learning theorists have also noted the self-perpetuating nature of some maladaptive behavior. Individuals who lack a strong sense of self-efficacy in social situations may avoid them. As a consequence, they fail to learn the social skills that would enhance their self-confidence. Self-defeating behaviors, which may appear irrational, can often be understood in terms of the dynamics of learning.

Application of Theory to Therapy Behaviorists have a rather rudimentary view of personality, seeing it as an assemblage of learned behaviors. They also tend to see psychopathology in superficial terms. Psychological maladjustment is considered to be the result of learned behaviors, called *symptoms*, but there is no underlying disorder of which they are symptomatic. Curing the symptoms cures the disorder. At worst, this position is naive and simplistic, equating the patient's presenting problem with the real source of difficulty. At best, however, it focuses attention on a specific problem that can be concretely addressed.

A large number of techniques for behavior modification have been used with considerable success in treating symptoms of psychopathology. Joseph Wolpe developed *systematic desensitization* as a treatment for phobias. Patients were instructed to relax and were then presented with increasingly vivid cues of the phobic object. Eventually they were able to face the object itself without anxiety. A more dramatic technique is *implosive therapy*, in which the individual is confronted directly with the feared object (e.g., a room full of snakes) without an opportunity to escape. Because the object itself is harmless and because avoidant behavior cannot be performed and is, therefore, not reinforced, the phobic reaction is swiftly extinguished.

Therapeutic interventions may be based on eliminating the reinforcements that sustain behavior, on punishments for unwanted behaviors, or on modeling or shaping of more desirable behaviors. Any variable known to affect the acquisition or extinction of behaviors may provide an opportunity for behavior change, and behavioral techniques have been applied to physiological responses as well as to voluntary behaviors through techniques of biofeedback. Behavior therapies have been used extensively in treating phobias, in controlling addictive behavior, in reducing the self-destructive behavior of autistic children, and in improving classroom discipline—for the behaviorist, the distinction between psychopathology and bad behavior is generally unimportant. Behavior therapies are most effective when the problem can be clearly traced to a particular set of behaviors or conditioned responses; they are much less effective in dealing with vague complaints of confusion and distress, although these are frequently the problems the patient presents to the clinician.

HUMANISTIC APPROACHES

Theories of Personality Psychoanalysis and behaviorism are mechanistic theories that trace human behavior and experience to the gratification of instinctual impulses or to the acquisition of learned responses. Many of the most influential personality theorists of the twentieth century have defined themselves in terms of their opposition to these two approaches. Although they vary widely in terms of their explanations of personality, humanistic approaches share a positive evaluation of human nature and emphasize its unique and distinctively human aspects. Personality produces as well as reflects organization, rationality, consistency, future orientation, planfulness, self-expression, cognitive complexity, and adaptability. Human reason and freedom of will, the capacity for growth and change, the need for love and self-transcendence are prized by most humanistic theorists.

Social learning theories might be seen as a humanized form of behaviorism because they recognize the role of complex symbolization and language in human learning. In a similar way, many now-classic humanistic theories were intended as modifications of psychoanalysis. Indeed, the first major psychoanalytic revisionist was Carl Jung, who argued that human beings had spiritual as well as sexual needs. Henry Murray also made major modifications to psychoanalysis in his view of personality. For example, Murray credited the mature ego with much more autonomy than Freud granted it, and he argued that the individual's sense of morality was not fixed by the superego instilled in childhood but could continue to develop into more rational and altruistic forms. Erich Fromm and Karen Horney minimized the instinctual origin of personality development and suggested that culture played a large role in shaping the individual. Erik Erikson proposed stages of psychosocial development to parallel Freud's stages of psychosexual development, emphasizing such distinctly human characteristics as identity, intimacy, and generativity; Erikson also theorized that personality development continued throughout the lifespan, giving encouragement to research on aging and personality.

Gordon Allport Although he is usually classified as a trait psychologist, in his general orientations to theorizing Gordon Allport (1897–1967) was clearly a humanist. For him, man's behavior is proactive, reflecting internal, self-initiating characteristics more than situational forces. In his view, human personality possesses psychological coherence and both momentary (*cross-sectional*) and long-term (*longitudinal*) organization. Allport considered personality functioning to be characteristically rational, organized and influenced by such conscious characteristics as long-range goals, plans of action, and philosophies of life.

Perhaps the most salient and controversial feature of his approach was the extreme emphasis that Allport placed on the uniqueness of the individual personality. Allport viewed the major task of *personology* (or personality psychology) as the understanding and prediction of the individual case. In order to grasp the real personality, personal dispositions must be assessed, which requires intensive study of an individual's past, present, and anticipated future functioning through the use of such techniques as the case history and content analysis of personal documents. In *Becoming: Basic Considerations for a Psychology of Personality*, Allport championed the view that concepts and laws must be developed to fit the individual case; Allport coined the terms *idiographic* and *morphogenic* to symbolize his conviction that “each person is an idiom unto himself, an apparent violation of the syntax of the species.”

Allport was deeply concerned with identifying personality functions, which he discussed under the concept of the *proprium*, the superordinate concept in his system. Propriate functioning not only organized and integrated actions and experience but also provided the impetus to psychological growth. Allport described the functions of the proprium as *sense of body, self-identity, self-esteem, self-extension, rational coping, self-image*, and *propriate striving*. These propriate functions were vital to personality, and although they were ongoing they were by no means considered unchanging. Allport theorized that the propriate functions are modified throughout life, predominantly in the direction of greater differentiation and integration, or growth. The development of selfhood, away from the undifferentiated, opportunistic functioning of infancy and early childhood towards propriate functioning and striving is part of human nature for Allport. Individuals guide or direct their lives by attempting to fulfill their sense of self or proprium. Development continues into adulthood, with increasing signs of maturity and personal lifestyle.

Abraham Maslow Abraham Maslow interpreted personality in motivational terms. The individual's whole life—perceptions, values, strivings, and goals—is focused on the satisfaction of a set of needs, and the needs themselves are arranged in a universal hierarchy. These needs serve an organizing and integrating role in life and are not to be understood as a simple and invariant set of responses to environmental pressures; they organize and create action possibilities and external reality.

At the lowest level of the motivational hierarchy are physiological needs for food, water, sex, and sleep. The second level consists of safety needs—needs for protection and security. The third level consists of needs for love and belongingness, and the fourth for self-respect and esteem from others. Above these lie the higher needs—for beauty, truth, justice, and *self-actualization*, the development of one's full potential as a unique human being.

Individuals live at the lowest level of motivation that is problematic for them. That is, if needs for food and shelter are not routinely met, these needs become the overriding concern in life: the hungry individual will risk danger and social ostracism to find food. Those who have always been well fed, however, learn to take the satisfaction of physiological needs for granted, and their attention is dominated by higher needs. Instead of examining specific behaviors and their reinforcements, Maslow's theory concerns the long-term satisfaction of broad classes of needs, and thus gives a much broader depiction of the individual.

Maslow devoted much of his writing to a characterization of higher needs including self-actualization, a drive to fulfill one's unique potentials. He believed that personality psychology had become obsessed with psychopathology and felt that a corrective emphasis on positive mental health was needed. His biographical studies of such exemplary people as Eleanor Roosevelt and Abraham Lincoln suggested a number of distinctive characteristics of self-actualizers, including accurate perception of reality, creativity, a need for privacy, and the frequent experience of mystical or “peak” experiences. As an exception to his general theory of motivation,

he also noted that such individuals often skip the lower levels and proceed directly to self-actualization. The most creative artists and musicians never seemed to care about poverty or lack of social acceptance.

George Kelly One of the most unconventional theories of personality was offered by George Kelly (1905–1967). It is in some respects a purely cognitive approach, but one with few ties to traditional learning theory. Instead of seeing human beings as organisms that are conditioned by their environment, Kelly argues that they should be seen as scientists trying to make sense of their world. In *The Psychology of Personal Constructs* he states the fundamental postulate of his theory: "A person's processes are psychologically channelized by the ways in which he anticipates events."

The basic unit for understanding personality is the *personal construct*, a schema for classifying and interpreting experiences. For example, an individual might characterize other people in terms of the contrast *strong* versus *weak*. Each new acquaintance would be categorized as either a strong or a weak person, and subsequent interactions with this person would be guided by the original construct. In the course of experience it would probably be necessary to reclassify some individuals initially thought to be strong as weak, and vice versa; more importantly, some people might act in ways that are neither strong nor weak, leading the individual to develop new constructs (say, *friendly* versus *hostile*) that are more useful in predicting other people's behavior.

This rather abstract and bloodless theory is made relevant to psychopathology by Kelly's ingenious reconstruing of some basic emotional reactions. *Anxiety* is defined by him as the awareness that one's construct system is inadequate for construing important events. *Guilt* is the recognition that one's behavior is inconsistent with the ways in which one construes oneself. *Hostility* is viewed as the attempt to force experience to fit one's existing constructs. Such definitions are remote from common sense as well as from clinical notions of anxiety, guilt, and hostility, but that is precisely why they offer the prospect of novel ways of treating them.

Carl Rogers Carl Rogers (1902–1987) is probably the most influential humanistic personality theorist. He articulated a formal theory of personality, pioneered a major school of therapy, *client-centered therapy*, and encouraged rigorous research on his theory and therapy. Rogers held that all organisms tend toward their own actualization—that mental health and personal growth are the natural condition of humankind. Psychopathology is a defensive distortion of this actualization process, and psychotherapy consists of creating conditions in which defense is unnecessary. Given these conditions, patients (or *clients*, as Rogers called them) essentially cure themselves.

Under ideal conditions people's needs, desires, and goals emerge naturally as part of self-actualization and are recognized as part of the self; individuals are fully open to experience. In real life, however, one person's needs and desires often conflict with persons', in particular, children find themselves in conflict with their parents, who are perceived as withholding love when the child (from their perspective) misbehaves. Because love is so essential, children internalize these *conditions of worth* and believe that they are good and worthwhile individuals only when their self is consistent with the ideals imposed by significant others. To maintain their sense of worth, children may distort their experience—this leads to anxiety and to self-defeating behavior.

In some respects Rogers' theory is much like Freud's: both see psychopathology as the result of defensive distortions and see the ideal state as one in which individuals can accept conflicts and deal with them rationally. Rogers is a humanistic theorist because he assumes that human nature is essentially good and that defenses are ultimately unnecessary. For Freud the impulses of the id are eternally primitive and selfish, and their full actualization would be socially catastrophic.

Dan McAdams Since the mid-1980s a number of scholars in the humanities and social sciences have turned their attention to the *life narrative* as a focus of research. These writers argue that the consciousness of self that distinguishes human beings from other animals is not a static list of personal characteristics; it takes the form of a story. In telling their life stories, people explain themselves in the context of their history and their significant relationships. Life narratives are not objective life histories; they are subjective interpretations that give personal meaning to past, present, and future events.

The personality psychologist most closely associated with this perspective is Dan McAdams. McAdams has postulated that personality can be understood on three levels: level 1 consists of *traits*, abstract and enduring tendencies seen in general styles of action and experience; level 2 is defined by *personal concerns*, the goals, plans, and strategies that preoccupy the individual at a certain time and place; and level 3 is the *life narrative*, the story the person tells to himself or herself and to others in an effort to give a sense of unity and purpose to life.

Psychiatrists and psychotherapists have listened to patients' life stories for many decades, usually seeking clues to the origins of current problems. McAdams' approach is different; he wishes to understand the narratives as stories. Stories can be analyzed in terms of such features as narrative tone (e.g., tragic, comic, ironic), imagery, theme (what goals the characters pursue), ideological setting, and nuclear episodes (crises, turning points). From this perspective key life events are not construed as causes of development, but as symbols of identity. A patient's vivid recollection of a childhood humiliation may be important chiefly because it conveys the patient's sense of being a victim to whom fate has always been and always will be unfair.

Much of the scholarship on life narratives has been in a humanistic tradition that eschewed empirical rigor: a narrative reduced to a set of numbers is no longer a narrative. McAdams and his colleagues, however, have begun to conduct psychometric studies on such basic issues as the inter-rater reliability and temporal stability of life story variables. These studies will form the basis for a scientific evaluation of narrative perspectives on personality.

Theories of Psychopathology Humanistic theories of personality tend to emphasize the positive aspects of human nature and discuss maladjustment in terms of failures of and blocks to the full growth and development of the individual. It is of some interest to note that both Rogers and Kelly formulated their theories in the context of counseling students—individuals who presumably had relatively minor maladjustments and considerable personality strengths. The applicability of humanistic theories to patients suffering from schizophrenia or dementia is certainly questionable, but the theories have had a profound effect on routine clinical practice, where clearly diagnosable psychiatric disorders seldom account for all the patient's problems.

Humanistic psychologists differ tremendously in their views of the origins of maladjustment. Rogers pointed to internalized conditions of worth acquired chiefly during childhood. Erich Fromm, who was influenced by Karl Marx, blamed society as a whole for instilling nonproductive orientations in individuals, such as *hoarding* or *marketing* orientations. Kelly said little about the origins of maladjustment, but thought its essence was an ineffective construct system too rigid to be corrected by experience.

Other personality theorists such as Rollo May have been influenced by existential philosophy and have argued that the essential characteristic of human nature is freedom. Freedom, however, implies responsibility, and it is what Fromm called the attempted "escape from freedom" that often leads to psychopathology. In this view, individuals are the ultimate source of their own problems. The debate about responsibility and mental illness continues today, perhaps most conspicuously in questions about whether alcoholism should be considered a disease or a failure of self-discipline.

Application of Theory to Therapy Some humanistic theories of personality—for example, Allport's theory—have had little impact on psychotherapy; others, like Rogers' theory, have been tremendously influential. It is helpful to recall that for decades the dominant form of therapy was psychoanalysis, a process that could take years and in which treatment was focused on dreams, childhood memories, and the ongoing relation with the therapist (*transference*) rather than on the immediate problems of the patient. Many of the standard techniques of contemporary counseling, clinical psychology, and psychiatry rest upon a very different set of assumptions about human nature made scientifically respectable by the work of humanistic psychologists.

Brief psychotherapies often consist of opportunities for individuals to express their feelings and rethink their problems in a supportive atmosphere. The therapist may provide advice and guidance, or at least offer new ways in which patients can think about their problems. This general approach is used by many psychiatrists as an adjunct to medication even to treat serious mental disorders. The process implicitly assumes that individuals, even those requiring psychotherapy, are basically rational and able with some help to solve their own problems; it also assumes that given the right conditions they will move toward mental health: patients are seen as scientists and self-actualizers.

Noting that Freud's "psychoanalytic cure" was based on a conscious understanding of one's life experiences, McAdams has argued that having a satisfying life narrative is a requirement for complete mental health. A good life story has coherence, credibility, a capacity for growth and change, and a generative orientation toward the world. One goal of psychotherapy may be to help patients rewrite their life narratives along these lines. This is a modest but realistic goal. Psychotherapists cannot easily alter patients' personality traits nor can they undo traumatic events of the past, but they may be able to help patients make sense of their dispositions and their life history. Human tragedy can bring great suffering, but in the hands of an artist it can also be the source of great beauty.

The humanistic emphasis on freedom and responsibility has often clashed with the psychiatric tradition of regarding mental disorders as diseases. Labeling individuals as "schizophrenics" or "phobics" is dehumanizing (and is discouraged by the editorial style of this textbook), and critics like Thomas Szasz have argued

that mental disorders are social and ethical judgments, not matters of medical fact. There is abundant evidence that some mental disorders have a biological basis, but the criticisms of humanistic psychologists make the point that the disorder occurs in a human being who in many respects may be like any other person. Behaviors, experiences, or relationships may be normal or abnormal, but people are not.

TRAIT AND FACTOR MODELS

Individual differences are peripheral concerns in many social learning and humanistic theories of personality; they are the central focus of trait theories. The study of variations in human character and temperament goes back at least to Theophrastus, a Greek whose *Characters* depicted thirty different types. The *Morose type*, for example, he described as follows:

> A malignant temper sometimes vents itself chiefly in ferocity of language. The man whose tongue is thus at war with all the world, cannot reply to the simplest inquiry except by some such rejoinder as—'Trouble not me with your questions:‘ nor will he return a civil salutation.... He has no pardon for those who may unwittingly shove or jostle him, or tread upon his toe.... He will neither wait for, nor stay with anyone long: nor will he sing, or recite versus, or dance in company. It is a man of this spirit who dares to live without offering supplications to heaven.

Theories of Personality The scientific study of individual differences in personality can be traced to Sir Francis Galton in England, who laid the foundations of psychometrics, and to G. Heymans in the Netherlands, who undertook the first large-scale study of rated personality traits. The first major trait theorist in the United States was Allport, whose 1937 volume *Personality: A Psychological Interpretation* spelled out the basic issues in trait psychology. He defined a *trait* as “a neuropsychic structure having the capacity to render many stimuli functionally equivalent, and to initiate and guide equivalent (meaningfully consistent) forms of adaptive and expressive behavior.” In this view something in the brain of the morose man makes him see even simple questions or greetings as a personal affront, and his sullen attitude is expressed in a variety of social situations.

Allport believed that traits were concrete features of individuals that uniquely described them and that might be understood by a case study of a single individual. A contrasting view is that traits are dimensions of individual difference that can only be discovered by comparing and contrasting different individuals; individuals are then described in terms of their standing on a set of common traits. The two definitions are closely related; people who are more anxious than 99 percent of the population presumably have a neuropsychic structure that makes them so anxious.

Characteristics of Traits Although there are many different trait theories, there is general agreement on several key features of traits: (1) Traits are tendencies to show consistent patterns of thoughts, feelings, and actions. Behaviors that are specific to a single setting or situation may better be considered habits rather than traits; some evidence of cross-situational consistency is necessary to infer a trait. However, concrete instances of behavior have many determinants, including learned habits, aroused needs, social contexts, role requirements, and the influence of many different and potentially conflicting traits, so the influence of a specific trait on any particular behavior may be quite small. It is usually only by viewing behavior across many different situations that a consistent pattern can be detected. (2) Traits are relatively enduring features that characterize the individual. In this respect they are to be distinguished from transient moods or episodes of mental disorder that affect the individual. The fact that traits are relatively enduring does not mean that they cannot change; traits are not immutable even if they are durable. (3) Traits are continuously distributed, usually approximating a normal or bell curve. Although it is convenient to speak about “introverts” and “extraverts,” most individuals are “ambiverts,” showing some of the characteristics of introverts and some of the characteristics of extroverts. With the possible exception of masculinity and femininity, there is no consistent evidence of discrete personality types.

From time to time there has been controversy over the reality of traits, fueled by the fact that human beings easily and readily ascribe traits to others on the basis of little or no information, with correspondingly limited accuracy. These personality ascriptions may be triggered by stereotypes of age or physical appearance or may be quite idiosyncratic. Demonstrations of this fact in laboratory experiments by social psychologists, together with the relatively loose cross-situational consistency of most traits, led a generation of psychologists in the 1970s to conclude that traits were cognitive fictions. Subsequent work—particularly demonstrations that judges who knew the individual well agreed with each other and with the self-reports of individuals about their standing on a variety of traits—restored faith in the consensual validity of traits. However, the controversy does make the crucial point that some trait ascriptions are more accurate than others and that first impressions may be quite misleading. Psychiatrists ought not to assume that their clinical judgments of a patient's personality are correct; validated personality questionnaires and rating forms completed by knowledgeable others may be needed to portray and understand personality accurately.

Personality Structure and Factor Analysis The most important differences among trait theorists are in the specific traits they have conceptualized and measured. Jung identified introversion and extraversion as basic personality variables, Bandura emphasized self-efficacy, Rogers was concerned with openness to experience. Over the years, literally thousands of scales have been developed to measure traits that psychologists considered important in understanding personality. As early as 1936 Allport pointed to another source for identifying personality traits: the natural language. A monograph he published with Henry Odbert listed some 18,000 terms extracted from an unabridged dictionary that could be used to describe people; some of them were mere evaluations (e.g., *swell, awful*), but he regarded about 4000 as legitimate trait terms.

The problem for trait psychologists was how to choose a manageable set of traits from among the many possible constructs. It was obvious that trait terms were highly redundant—for example, *anxious, worrying, nervous, apprehensive*, and *fearful* reflect very similar, almost identical characteristics; what was needed was a procedure for identifying major groups of traits that covaried. Factor analysis, a statistical technique that reduces the complexity of a set of correlations among variables, was first used in personality research by J.P. Guilford, and has remained one of its basic tools. The factors, or dimensions, identified in this process correspond to groups of closely related traits; the set of basic dimensions identified by the factor analysis constitutes a model of the structure of personality traits.

Raymond Cattell developed one of the first and most influential factor models. He reasoned that in the course of cultural evolution any personality trait important in human social interaction would have been noticed and named; the 4000 trait terms identified by Allport and Odbert could thus be assumed to represent an exhaustive listing of personality characteristics—this has become known as the *lexical hypothesis*. Cattell grouped synonyms and near-synonyms together to obtain a set of 35 personality variables, and asked respondents to rate acquaintances on each of these sets of terms. He intercorrelated the ratings and factored the correlations, identifying 12 factors. Together with four more factors found in research using self-report questionnaires, these became the basis for the 16 Personality Factor Questionnaire (16PF), a self-report instrument that has been widely used in personality research and clinical psychology for 40 years.

It was originally hoped that factor analysis would provide an objective solution to the question of personality structure, but for many years there was little agreement among factor analysts. Eysenck believed that Cattell's model was unreplicable and needlessly complex. He proposed a simple and powerful two-dimensional model that identified extroversion–introversion *(E)* and neuroticism–emotional stability *(N)* as superfactors, and showed that if the scales of the 16PF were themselves factored, the two largest factors resembled his *E* and *N*. He and his wife, Sybil Eysenck, developed a series of instruments to measure these factors (and later a third superfactor called *psychoticism*) that have also been widely used, particularly in the United Kingdom. Eysenck's stature as a learning theorist and a critic of psychoanalysis contributed to the importance of these dimensions in psychiatric contexts.

Five-Factor Model Eysenck's two factors were widely replicated, but they seemed to omit many important characteristics, such as curiosity, aggression, and achievement striving. An alternative solution was offered in 1961 by two U.S. Air Force psychologists, Ernest Tupes and Raymond Christal. They began with the 35 clusters of traits identified by Cattell and obtained ratings on these clusters in eight different samples. They found that a five-factor solution fit the data in all eight samples. Two of the factors resembled Eysenck's *E* and *N*, but three other factors were new.

A small group of lexical researchers, including Warren Norman and Lewis R. Goldberg, continued to study personality structure as represented by natural language trait adjectives, and after 20 years came to the conclusion that the five-factor structure proposed by Tupes and Christal was essentially correct. Renewed interest in this model showed that the five factors could be recovered in analyses of self-reports as well as observer ratings, in ratings of children, college students, and older adults, and in several different languages, including non–Indo-European languages like Chinese, Hebrew, and Filipino. The factors appeared in analyses of trait adjectives, descriptive phrases, and questionnaire scales. Contemporary five-factor theorists differ somewhat on their conceptualizations of the factors and consequently give them somewhat different labels. The terms *neuroticism (N), extraversion (E), openness to experience (O), agreeableness (A)*, and *conscientiousness (C)* are used here. The five factors and some of the traits that define them are presented in Table 6.4-1.

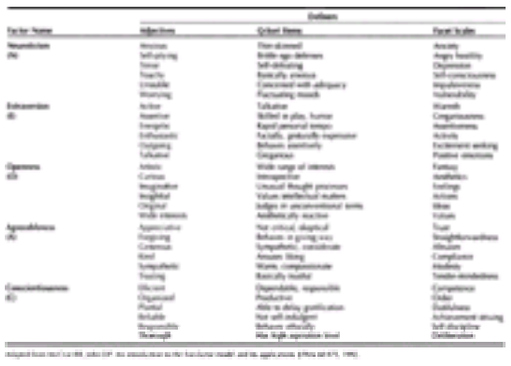

Table 6.4-1 Examples of Adjectives, California Q-Sort Items, and NEO-PI-R Facet Scales Defining the Five Factors

Many psychologists were skeptical of the lexical hypothesis that had led to the five-factor model. These critics believed that personality theory and clinical experience would lead to the identification of important traits for which no lay terms existed, and they continued to offer alternative models. Katharine C. Briggs and Isabel Briggs Myers operationalized Jung's psychological functions in the Myers-Briggs Type Indicator, which classifies individuals in terms of the dichotomies introversion versus extroversion, intuition versus sensing, thinking versus feeling, and judging versus perceiving. Timothy Leary argued that the traits that influence social interactions were better represented in a circular order than as a set of factors, and many instruments have been developed to measure this model.

A particularly important system was suggested by Theodore Millon, who was interested in personality traits associated with psychiatric disorders. His reviews of clinical literature led to a theory of personality and psychopathology that specified 11 personality disorders as extreme variants of normal traits. For example, the histrionic personality disorder is supposed to be related to the trait of gregariousness, the schizoid personality disorder to the trait of detachment. Millon's theory had a profound impact on the formulation of Axis II in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and in subsequent editions. His instrument, the Millon Clinical Multiaxial Inventory, has been widely used by psychiatrists and clinical psychologists.

The fact that trait theories of personality are usually tied to assessment inventories makes it relatively easy to make empirical comparisons, and since the late 1980s researchers in several countries have undertaken the task of relating different trait models. There is a growing consensus among these researchers that virtually all the traits measured by theory-based personality questionnaires—including those derived from psychodynamic, behavioral, and humanistic theories—are related to one or more of the five factors of Tupes and Christal. However, instruments vary in comprehensiveness. For example, the four scales of the Myers-Briggs Type Indicator correspond to four of the five factors, whereas measures of the interpersonal circumplex represent only two factors (extroversion and agreeableness). Table 6.4-2 lists some scales that have been shown empirically to be linked to each of the five factors. Thus, at the broadest level the problem of personality structure appears to have been resolved: personality is described by five basic factors.

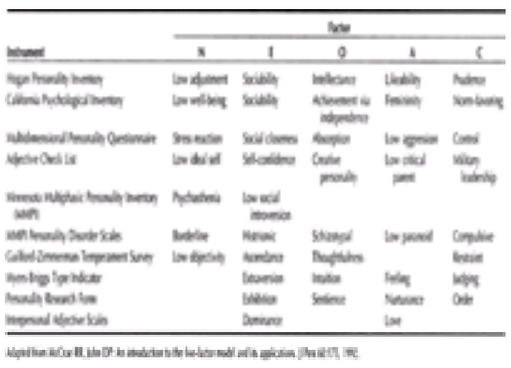

Table 6.4-2 Some Scales Empirically Related to Each of the Five Factors

However, personality cannot be exhaustively described in this way. Most trait psychologists adopt hierarchical models of trait structure and assume that the broadest factors are composed of more specific traits, which in turn are defined by discrete behaviors. In the Revised NEO Personality Inventory (NEO-PI-R), for example, six specific traits or facets are measured for each of the five factors (or domains) or personality. The facet scales are listed in Table 6.4-1; the hierarchical organization is illustrated in Figure 6.4-1. Assessment on the level of specific facets provides a more detailed and personalized portrait of the individual.

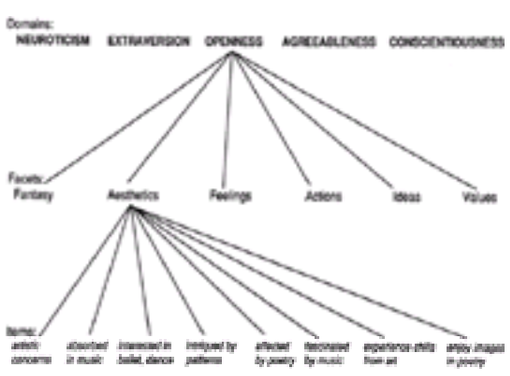

FIGURE 6.4-1 An example of hierarchical organization in the Revised NEO Personality Inventory: Domains, facets of openness, and items measuring openness to aesthetics.

Lifespan Development of Personality Traits Psychoanalytic, learning, and humanistic theories have usually offered causal explanations for individual differences. Thus, differences in self-efficacy are supposed to be caused by different histories of success in pursuing goals; variations in openness to experience might be the result of differences in the conditions of worth imposed by parents. The factor analysts who scoured the dictionary for trait terms usually bypassed this issue, being content to describe the personality differences found in adults. In principle, some traits might be inherited, some instilled by parents, some learned from experience; whatever their origin, they can be measured and used to predict important life criteria.

Until very recently it was generally assumed that personality was shaped primarily by a variety of environmental influences, including parental love and discipline, social and economic opportunities, and life experiences through childhood and adolescence. Surprisingly, this assumption has rarely been tested: there has been very little prospective longitudinal research documenting links between early childhood experiences and subsequent adult personality. A different research design using the techniques of behavior genetics can answer these questions by comparing personality measures in adults with different degrees of genetic and environmental similarity. For example, similarity between identical twins reared apart can be attributed to genetic influences, whereas similarity between adopted siblings reared together must result from environmental influences.

Since the 1970s behavior genetics studies using many different samples, personality measures, and methods of data analysis have concluded that personality traits are to a considerable extent heritable and to a considerable extent the result of unknown and idiosyncratic causes, but that they are hardly attributible to shared

environmental causes at all. Socioeconomic status, family diet, religious training, parental modeling, and all the other influences that children growing up in the same household would normally share seem to have little or no influence on adult personality. This conclusion applies as much to character traits such as achievement striving and modesty as to temperament traits such as anxiety and energy level. This dramatic and counterintuitive finding will doubtless reshape theories of personality. As Sandra Scarr commented, “Psychology has no adequate theories to account for individual variation in behavior because our theories address the wrong sources of variation.” Future theories will ascribe much more importance to genetic influences, and developmental psychology may become more concerned with explaining how heritable traits come to be expressed in different family and social settings.

For many years developmental psychologists assumed that personality development ended with adolescence. A 21-year-old is legally an adult and is unlikely to grow much in either physical height or intellectual capacity. However, recent research has shown consistent changes in personality between college age and middle adulthood. There is a dramatic decline in excitement seeking over that interval and cross-sectional and longitudinal studies of American college students have shown that *N, E*, and *O* decline during the decade of the 20s, whereas *A* and *C* increase. As a result, 30-year-olds are less emotional and better socialized than 20-year-olds.

There is reason to believe that these changes are not the product of the relatively indulgent American experience of adolescence. The same pattern of age differences is seen in less affluent countries like Croatia and in non-Western countries like Korea. Such cross-cultural studies suggest that there may be intrinsic maturational processes in personality development.

There have also been important developments in understanding personality change after age 30. Traits are defined as relatively enduring patterns of behavior, but most psychologists assumed that traits would be modified by life experiences. Jung postulated that the process of individuation required each individual to express all his or her potentials, and thus the young introvert would normally become an extravert in old age, and vice versa. Lay stereotypes of aging held that as they age, people become cranky, conservative, and depressed. Gerontologists rebutted these myths of aging, arguing that old age was more likely to bring maturity, wisdom, or detachment. Theories of adult development, popularized in Gail Sheehy's best-selling *Passages*, suggest that personality changed in stages, and in particular that men (and perhaps women) went through a midlife crisis in their 30s or 40s.

In sharp contrast to these theories are the findings from a number of independent longitudinal studies of personality in adulthood. These studies present a clear picture of predominant stability of the full range of personality traits. That is, the average levels of most traits neither increase nor decline much with age, and individuals tend to maintain the same rank order. The 30-year-old who is outgoing, curious, and hardworking is likely to become an outgoing, curious, and hardworking 80-year-old. This conclusion has been supported by studies using self-reports as well as observer ratings of all five factors of personality and appears to apply to both men and women. There are often dramatic changes in personality as a result of dementing disorders in old age, but normal life experience seems to have little impact on personality after age 30.

There is distressingly little data on the long-term effects of psychopathology on personality. Individuals in recovery from an episode of major depressive disorder generally score high on measures of *N*, and it is sometimes argued that this is a result of the experience of depression. However, the few prospective studies of the first onset of major depressive disorder typically show that these individuals scored high on *N* prior to the first episode, suggesting that high *N* may be an enduring feature of people who are prone to clinical depression.

Theories of Psychopathology The links between psychopathology and trait psychology are at once intimate and complex. For decades psychiatrists and clinical psychologists have administered psychological tests such as the Minnesota Multiphasic Personality Inventory (MMPI) and the Cattell's measure of personality, the 16PF, to inform the diagnosis of mental disorders; these same tests have been used routinely by trait psychologists to examine relations between personality and such criteria as vocational preference, creative potential, and coping with life stress. The dividing line between normal variations in personality dispositions and psychopathology is frequently unclear.

This poses more of a problem to psychiatric nosology than it does to trait theories of personality. Psychiatric diagnoses are supposed to represent the presence of a discrete disorder, and the fourth edition of DSM (DSM-IV) specifies the criteria that must be met to confer each diagnosis. This categorical model is appropriate for some disorders, but not others. For example, it may be very difficult to separate social phobia from extreme shyness and self-consciousness, or a schizoid personality disorder from marked introversion. This is entirely consistent with trait models of personality, which regard individual differences as continuously distributed variables.

Neuroticism and Psychopathology One theory of psychopathology states that some psychiatric disorders reflect extreme standing on personality traits, particularly those related to *N*. Among the traits that covary in normal populations to define this factor are predispositions to experience chronic levels of negative affects such as fear, anger, shame, and sadness. Individuals with very high standing on these traits may qualify for a diagnosis of generalized anxiety disorder, borderline personality disorder, social phobia, or dysthymic disorder.

Neuroticism may also be considered a risk factor for the development of psychiatric disorders that are not themselves trait-like. Several recent studies have demonstrated that individuals who score high on measures of *N* are at increased risk of subsequently receiving a diagnosis of major depressive disorder. Poor control of urges and impulses and excessive concern with physical functioning are also characteristics associated with *N*, and it might be hypothesized that individuals high in *N* are predisposed to develop eating disorders and conversion disorder (hypochondriasis). Neuroticism is a generalized disposition to experience psychological distress, and individuals seeking psychiatric treatment are almost always distressed. It is therefore not surprising that virtually all clinical populations, from drug abusers to schizophrenia patients, score high on measures of *N*.

Historically, the term *neurosis* or *psychoneurosis* was coined to identify psychiatric disorders of functional origin that were closely related to anxiety and its control. It survived in the revised third edition of DSM (DSM-III-R) in many secondary labels (*dysthymia*, for example, was also called *depressive neurosis*), but is not used in DSM-IV. Many personality psychologists object to the use of the term *neuroticism* as a label for a dimension of normal personality because of its suggestion of psychopathology. Many psychiatrists object to the term because it is tied to an outdated psychiatric nosology. However, the term serves a useful purpose: it is a reminder to clinicians that patients with a wide variety of diagnoses share many features related to chronic psychological distress, and a reminder to personality psychologists that the difference between normal and abnormal functioning is often only one of degree. The cliché that people are all more or less neurotic has some scientific basis.

Personality Traits and Personality Disorders Axis II of the DSM-IV is used for the diagnosis of personality disorders, which are defined as inflexible and maladaptive personality traits. It is reasonable to ask whether these traits are the same as or different from those encountered in nonpsychiatric populations. Several recent studies on this question have concurred in finding strong and replicable links between scales measuring personality disorders and the five factors in both normal and clinical populations.

Only high scores of *N* are associated with psychiatric impairment, but both poles of the other factors appear to be associated with specific forms of psychopathology. Individuals high in *E* tend to have histrionic and narcissistic personality disorders; those who are low in *E* have avoidant and schizoid personality disorders. High *C* is associated with obsessive-compulsive personality disorder; low *C* with antisocial personality disorder. Low *A*, or antagonism, is characteristic of individuals with paranoid, antisocial, and narcissistic personality disorders, as well as with the proposed sadistic personality disorder. High *A* is associated with dependent personality disorder. The hypothesized relations between personality traits and DSM-IV personality disorders are presented in Table 6.4-3.

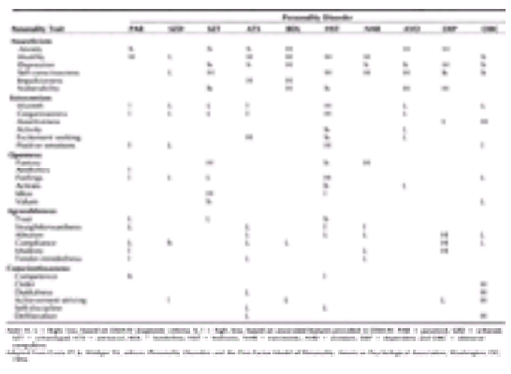

Table 6.4-3 Hypothesized Relations Between DSM-IV Personality Disorders and Five-Factor Model Personality Traits

It is a matter of current controversy just how these personality traits are related to the disorders: are normal personality traits carried to an extreme inherently maladaptive, or do these traits merely predispose individuals to develop disorders under certain circumstances? Some research has also called into question the meaningfulness of the syndromes recognized by DSM-IV, by showing that the symptoms used to define the disorders, when separately assessed, do not covary in ways that match the DSM-IV syndromes. Instead, the symptom clusters that do emerge are interpretable in terms of the five-factor model. As a result of such evidence, proposals have been made to replace the current categorical model by dimensional models that relate problems in adjustment to standing on basic dimensions of personality. This seems to be a logical extension of trait theories of psychopathology.

Application of Theory to Therapy Psychodynamic, behavioral, and humanistic theories all specify the causal mechanisms by which psychopathology is created and maintained and thus imply points of intervention. If maladjustment is caused by rigidly internalized conditions of worth, then unconditional positive regard in the therapeutic setting may provide a cure; if the problem is a learned behavior, then altering reinforcements may extinguish it. Trait models do not emphasize causal or developmental explanations and thus may seem to have no implications for psychotherapy; however, they have many.

Recent evidence on the substantial heritability of most personality traits clearly indicates that Allport was right: traits do have some underlying neuropsychic structure. Research on the psychophysiology of traits is a topic of growing interest and parallels the extensive work done on the neurophysiological basis of many forms of psychopathology. In this broad sense psychopharmacological approaches to psychotherapy are in principle consistent with trait theories of personality: if personality traits and disorders reflect brain processes, drugs that affect the brain may offer useful instruments for psychotherapy. However, there is still much to learn; for example, individuals with dysthymic disorder who score high on measures of trait depression often do not respond to antidepressant medication.

Personality Assessment Historically the chief role of trait psychology in psychotherapy has been in diagnosis and assessment. A number of self-report measures, such as the MMPI and the Millon Clinical Multiaxial Inventory, were designed specifically to measure psychopathology and include primarily items that tap psychiatric symptoms. They can be regarded as measuring psychopathological traits and states although they have often been used in college and volunteer samples as measures of personality per se. These instruments are chiefly of value as aids to psychodiagnosis. General personality questionnaires, including the 16PF, the Guilford-Zimmerman Temperament Survey, and the Edwards Personal Preference Schedule, have also been used for decades in clinical psychology and psychiatry as part of a complete psychological assessment. Several measures of the five-factor model have recently appeared, including the Hogan Personality Inventory and the NEO Personality Inventory (NEO-PI), and there has been considerable interest in the clinical application of the five-factor model. The primary advantage of this model is its comprehensiveness; by assessing traits from all five dimensions, the clinician can efficiently obtain a full portrait of the individual.

Personality psychologists and psychometricians have devoted years to the development of self-report inventories, and the utility of this approach to assessment is beyond doubt. Self-reports, however, are by no means infallible. Patients may not understand their own personalities or may deliberately misrepresent themselves. Concerns about defensiveness and socially desirable responding have led to the use of projective tests, the development of special validity scales to detect and correct for distorted responses, and reliance on the clinical judgment of the psychiatrist. Each of these possible solutions introduces problems of its own, however, and most clinicians rely on multiple sources of information.

One source of information that has been underutilized in clinical settings is informant ratings. Research since the 1980s has confirmed that ratings on standardized instruments by significant others—usually spouses or parents—can provide reliable and valid assessments of personality. These findings appear to hold for clinical as well as volunteer samples. Personality ratings may be particularly valuable when patients are incapacitated or are strongly motivated to present an overly favorable or unfavorable picture of themselves.

Uses of Trait Profiles Because traits are enduring and stable features of the patient's behavior and experience assessing a broad range of traits gives the clinician a sense of what the person is like, which can be useful for many purposes beyond the formulation of a diagnosis. A complete personality profile can point to the patient's strengths as well as weaknesses, can help predict the course of therapy, and can aid in the selection of an optimal mode of treatment.

Humanistic theories of personality stress human potentials for altruism, creativity, and commitment and argue that psychotherapy must use these assets. Trait theorists believe that some people are altruistic, creative, or committed but others are not. Assessing traits allows the clinician to identify and capitalize on the particular strengths that characterize each individual patient.

Standing on trait dimensions can also give clues to the probable course of therapy. Patients who are low on agreeableness are distrustful and uncooperative, and it may prove very difficult to form a therapeutic alliance with them; by contrast, patients who are extremely high on *A* may be excessively compliant and become dependent on the therapist. Conscientiousness involves commitment and self-discipline. Patients high in *C* will probably adhere to treatment recommendations and work hard to solve their problems; those low in *C* are less persistent and dedicated, and the therapist may have to work to motivate them.

Finally, a few controlled studies and a good deal of clinical experience suggests that personality traits influence the effectiveness of various kinds of treatment interventions. Interpersonal therapies, in which patients are required to speak a great deal about themselves, appear to be most effective for extraverts; pharmacological management may be superior for introverts. Similarly, individuals who are high in openness to experience benefit from such techniques as guided imagery, whereas those who are low in openness prefer biofeedback. Ideally, the choice of therapies should be guided not only by the nature of the disorder, but also by the enduring characteristics of the patient.

As an illustration of profile interpretation, Figure 6.4-2 provides data on a well-known public figure, Richard M. Nixon. As part of a project on personality traits of U.S. Presidents, two experts independently described Nixon using the observer rating form of the NEO-PI-R. Agreement between the two experts was very high, and their ratings were averaged.

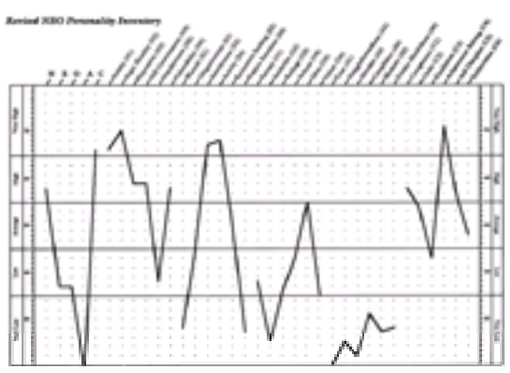

FIGURE 6.4-2 NEO-PI-R profile for Richard M. Nixon, based on combined ratings from two experts. (Data provided courtesy of Rubenzer S, Faschingbauer T, Ones D: In *Personality and the Presidency: A Scientific Inquiry*. Symposium presented at the Annual Convention of the American Psychological Association, August, 1996, Toronto. Profile form reproduced by special permission of the publisher, Psychological Assessment Resources, Inc, Lutz, FL, from the Revised NEO Personality Inventory by Paul T. Costa, Jr, and Robert R. McCrae © 1978, 1985, 1989, 1992 by PAR, Inc. Further reproduction is prohibited without permission of PAR, Inc.)

The left side of the profile sheet shows Nixon's standing on the five factors: high on *N*, average on *E*, low on *O*, very low on *A*, and very high on *C*. Specific traits, grouped by domain, are given toward the right of the profile sheet. Among the most notable features are Nixon's very high *N1:* Anxiety and *N2:* Angry Hostility and his very low scores on all six facets of agreeableness, reflecting both his personal distress and his antagonism toward others. However, the profile also shows some clear strengths, such as very high levels of *E3:* Assertiveness, *E4:* Activity, and *C4:* Achievement striving, that help explain his political successes.

Table 6.4-3 suggests traits that are likely to be associated with each of the DSM-IV personality disorders, and computer interpretations can compare observed profiles with these patterns. For example, Nixon's high scores on *N2:* Angry Hostility and *N4:* Self-Consciousness, and his low scores on *A2:* Straightforwardness, *A4:*

Compliance, *A5:* Modesty, and *A6:* Tender-mindedness are consistent with the diagnosis of narcissistic personality disorder. The profile in Figure 6.4-2 also suggests the possibility of paranoid, schizotypal, antisocial, and obsessive-compulsive personality disorders. Whether Nixon in fact would have met formal criteria for the diagnosis of any of these disorders is questionable, but he clearly shared some features of the disorders.

In addition to suggesting possible diagnoses, personality profiles can be used to anticipate the course of psychotherapy. For example, Nixon's generally low *O* scores mean that he would probably have been uncomfortable in therapy, and, among therapies, probably would have preferred directive techniques and behavior modification over free association and guided imagery. His low *A* scores suggest the likelihood of problems in forming a therapeutic alliance, but his high *C* scores mean that he would have worked hard at assigned tasks.

Psychotherapy and Personality Change If psychopathology is an expression or an outcome of personality traits, and if, as longitudinal studies demonstrate, personality traits change little over time, how can psychotherapy be effective? The pessimistic answer is that it cannot. Many psychiatric disorders are lifelong or recurrent, and treatment consists of management rather than cure. It is well known that those who have the best prognosis for recovery are those who are initially least impaired. In these cases the disorder may represent a transient adjustment reaction among individuals who have relatively healthy personality traits. Individuals very high in *N* and low in *A* and *C* may be poor candidates for psychotherapy.

However, so pessimistic a conclusion is unwarranted. Longitudinal studies show that people change little in the course of normal experience, but they do not rule out the possibility that direct interventions can make changes. It is unlikely that psychotherapy will make dramatic and lasting changes in basic personality traits, but more modest improvements may be enough to allow the patient to function adequately in daily life. Interventions may be particularly effective in early adulthood, when traits are not yet fully developed. Even without changing personality, therapy may help individuals to adapt to their own nature. The chronically anxious person may learn techniques of relaxation; the disagreeable individual may learn social skills that improve interpersonal relationships.

In some sense, most forms of psychotherapy can be seen as learning experiences. Psychoanalysts promote insight into unconscious conflicts; behaviorists help their clients understand the contingencies that reinforce troubling behavior; humanistic psychologists encourage patients to discover their true potential. Trait psychology can also be a source of insight. All human beings are in some respects alike, and it is comforting to learn that one is not alone in one's suffering. However, in other respects people are different from one another, and it can also be therapeutic to learn this fact. Helping patients to understand their own enduring dispositions can prepare them for some of the problems they face in life.

SUGGESTED CROSS-REFERENCES

Psychodynamic perspectives on personality and psychopathology are treated extensively in Section 6.1 on psychoanalysis and Section 6.3 on other psychodynamic schools. Chapter 24 discusses personality disorders. Chapter 7 on diagnosis and psychiatry deals with issues of assessment; Section 7.4 on personality assessment is particularly relevant. Section 3.3 on learning theory provides technical background for the discussion of behavioral approaches to personality, and Chapter 30 on psychotherapies—particularly Section 30.2 on behavior therapy, Section 30.6 on cognitive therapy, and Section 30.8 on brief psychotherapy—gives details about the application of theories of personality to therapeutic processes. A different view of adult development is explained in Chapter 50 on adulthood.

SECTION REFERENCES

Allport GW: *Personality: A Psychological Interpretation*. Holt, New York, 1937.

Bandura A: *Social Learning Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

Costa PT Jr, McCrae RR: Influence of extraversion and neuroticism on subjective well-being: Happy and unhappy people. J Pers Soc Psychol *38:*668, 1980.

Costa PT Jr, McCrae RR: Personality stability and its implications for clinical psychology. Clin Psychol Rev *6:*407, 1986.

Costa PT Jr, McCrae RR: Domains and facets: Hierarchical personality assessment using the Revised NEO Personality Inventory. J Pers Assess *64:*21, 1995.

*Costa PT Jr, Widiger TA: *Personality Disorders and the Five-Factor Model of Personality*. American Psychological Association, Washington, DC, 1994.

Digman JM: Personality structure: Emergence of the five-factor model. Annu Rev Psychol *41:*417, 1990.

Dollard J, Miller NE: *Personality and Psychotherapy: An Analysis in Terms of Learning, Thinking, and Culture*. McGraw-Hill, New York, 1950.

Eysenck HJ, Eysenck M: *Personality and Individual Differences*. Plenum, London, 1985.

Halverson CF, Kohnstamm GA, Martin RP, editors: *The Developing Structure of Temperament and Personality from Infancy to Adulthood*. Erlbaum, Hillsdale, NJ, 1994.

Heatherton TF, Weinberger J: *Can Personality Change?* American Psychological Association, Washington, DC, 1994.

Kelly GA: *The Psychology of Personal Constructs*. Norton, New York, 1955.

*Maddi SR, Costa PT Jr: *Humanism in Personology: Allport, Maslow and Murray*. Aldine, Chicago, 1972.

Maslow AH: *Motivation and Personality*. Harper & Row, New York, 1954.

McAdams DP: Personality, modernity, and the storied self: A contemporary framework for studying persons. Psychol Inq *7:*295, 1996.

McCrae RR, Costa PT Jr: *Personality in Adulthood*. Guilford, New York, 1990.

*McCrae RR, Costa PT Jr: Toward a new generation of personality theories: Theoretical contexts for the five-factor model. In *The Five-Factor Model of Personality: Theoretical Perspectives*, JS Wiggins, editor. Guilford, New York, 1996.

McCrae RR, John OP: An introduction to the five-factor model and its applications. J Pers *60:*175, 1992.

*Miller T: The psychotherapeutic utility of the five-factor model of personality: A clinician's experience. J Pers Assess *57:*415, 1991.

*Nathan PE, Langenbucher JW: Psychopathology: Description and classification. Annu Rev Psychol *50:*79, 1999.

Ozer DJ, Reise SP: Personality assessment. Annu Rev Psychol *45:*357, 1994.

Pervin LA: A critical analysis of current trait theory. Psychol Inquiry *5:*103, 1994.

*Phares EJ: *Introduction to Personality*, ed 2. Scott, Foresman, Glenview, IL, 1988.

Rogers CR: *On Becoming a Person: A Therapist's View of Psychotherapy*. Houghton Mifflin, Boston, 1961.

Scarr S: Distinctive environments depend on genotypes. Brain Behav Sci *10:*38, 1987.

Schopenhauer A: *The World as Will and Representation*, ed 3. Dover, New York, 1969.

Skinner BF: *About Behaviorism*. Knopf, New York, 1974.

Strack S, Lorr M: *Differentiating Normal and Abnormal Personality*. Springer, New York, 1994.

Watson D, Clark LA, editors: Special issue on personality and psychopathology. J Abnorm Psychol *130:*1, 1994.

Widiger TA, Trull TJ: Personality and psychopathology: An application of the five-factor model. J Pers *60:*363, 1992.

Wiggins JS, Pincus AL: Conceptions of personality disorders and dimensions of personality. Psychol Assess *1:*305, 1989.

CHAPTER 7. DIAGNOSIS AND PSYCHIATRY: EXAMINATION OF THE PSYCHIATRIC PATIENT

7.1 PSYCHIATRIC INTERVIEW, HISTORY, AND MENTAL STATUS EXAMINATION

MYRL R. S. MANLEY, M.D.

The purpose of a psychiatric diagnostic interview is to gather information that will enable the examiner to make a diagnosis. Having established a diagnosis, the clinician can then make predictions about the future course of a disorder and the likely response to treatment. As with all areas of medicine, treatment decisions are guided by diagnosis. Unlike most disciplines of physical medicine, however, psychiatry has no external validating criteria, no laboratory tests to confirm or refute diagnostic impressions. Consequently the diagnosis is wholly a product of the skills and knowledge of the individual psychiatrist and can never be better than the judgment made by individual clinicians.

Because of the absence of external validating criteria or biologic markers, diagnostic reliability is an intrinsic problem in clinical psychiatry. Before 1980 the problem was compounded by official diagnostic descriptions that were narrative and impressionistic. Schizophrenia was overdiagnosed in the United States because the description failed to distinguish it sufficiently from mood disorders with psychotic symptoms. Starting with the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) of the American Psychiatric Association (APA) in 1980 and through subsequent editions and revisions, diagnostic criteria have been based on descriptive phenomenology of clinical symptoms and clinical course. The move away from interpretive, intuitive, and impressionistic diagnoses has succeeded in improving diagnostic reliability. It has also strengthened the medical model of psychiatry.

Medical Model and Psychodynamic Formulation The medical model has become the dominant mode of psychiatry in the United States and much of the rest of the world. In this model psychiatrists are seen as physicians who specialize in the treatment of psychiatric disorders. The function of the diagnostic interview is to gather sufficient information to establish a categorical diagnosis or diagnoses. The diagnosis is used to predict the future course of a disorder and its likely response to treatment, and it is the foundation for all subsequent therapeutic decisions.

In contrast, the adaptational model sees psychiatrists as specialists in behavior and adaptation whose expertise can benefit people whether or not they have a diagnosable psychiatric disorder. Psychoanalytic psychiatry is an example of this model. A psychoanalytic interview is less concerned with establishing a diagnosis than with surveying psychological functions as they have evolved over an individual's lifetime.

A psychodynamic formulation draws from the principles of psychoanalytic theory. It describes personality structure in terms of ego strengths (including principal defense mechanisms, regulation of drives, relationships with other people, and reality testing), principal psychological conflicts, and developmental history, with particular emphasis on early childhood. The psychodynamic formulation is not intended to produce a diagnosis but rather to describe an array of psychological and adaptive capacities. These descriptions allow the analytic psychiatrist to formulate a theoretical model that explains current symptomatic behavior and interpersonal or functional limitations. It serves as the template for the conduct of a psychoanalytic therapy by anticipating unconscious intrapsychic conflicts and unacknowledged developmental arrests or delays.

The psychodynamic formulation differs from a diagnostic interview using the medical model in several significant aspects. It is more concerned with the unique characteristics of the particular individual than with commonalties of a diagnostic class. The data of a psychodynamic formulation are overwhelmingly interpretive and inferential. Defense mechanisms cannot be seen; their existence can only be inferred from observed (or described) behavior. Moreover, the psychodynamic formulation is inextricably tied to a single theoretical school. Prior understanding of psychoanalytic theory is prerequisite to making the formulation, and experienced analysts may offer different formulations from the same data.

In clinical practice psychiatric assessments are likely to draw from both analytic and medical models and are shaped by the unique circumstances of individual patients. There are patients for whom a psychodynamic formulation is relatively more important and will thus take up a greater portion of the interview. Among these are persons being considered for a psychodynamic therapy or those with a constellation of complaints not subsumed in a conventional categorical diagnosis. There are also persons for whom a psychodynamic formulation may be of negligible importance, for example, an individual with a diagnosis of obsessive-compulsive disorder or simple phobia in which the planned therapies are pharmacological and behavioral rather than psychodynamic. However, even in these circumstances many psychiatrists prefer to have some sense of a person's psychological makeup and developmental history to avoid focusing on symptom relief to the exclusion of other areas of potential concern and to deal with resistances to therapy as they arise. The principles of psychoanalytic theory may offer a workable model for organizing and using these concepts.

PSYCHIATRIC EVALUATION

The psychiatric evaluation comprises two sections. The first, a section of histories (e.g., psychiatric, medical, family) includes the patient's description of how symptoms of the current episode have evolved, a review of past episodes and treatments, a description of current and past medical conditions, a summary of family members' psychiatric problems and treatments, and the patient's personal history, which reveals interpersonal and adaptive functioning over time. Information for the history will come from the patient but may be supplemented by collateral information from family members, social referral agencies, previous treating physicians, and old hospital records. The second section of the psychiatric evaluation, the mental status examination, systematically reviews emotional and cognitive functioning of the patient at the time the interview is conducted. Diagnostic criteria in DSM-IV draw from both the history and the mental status examination. An outline of the psychiatric assessment is presented in Table 7.1-1.

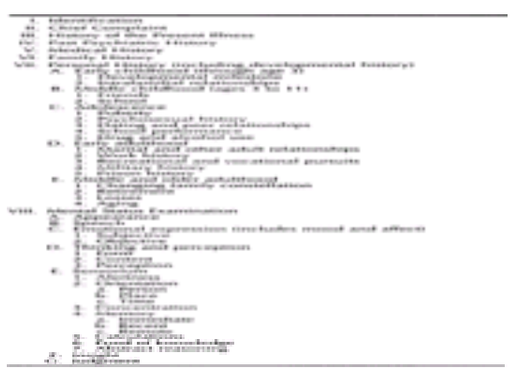

Table 7.1-1 Psychiatric History and Mental Status Examination

PSYCHIATRIC HISTORY

Identification The identification establishes the basic demographics of the patient. It includes age and sex, frequently includes racial or ethnic information, and occasionally includes religious affiliation. It is not necessary to list family, social, occupational, and educational information covered more fully in the personal history unless such information has direct diagnostic relevance. For example, although military service would not routinely be mentioned, it would be helpful to mention that a

patient is an excombatant when posttraumatic stress disorder is being considered. (The case from which the following examples are taken was provided by Richard Nathanson.)

> This is the first psychiatric hospital admission for Mr. A., a 21-year-old bisexual male employed part-time as a veterinarian's assistant. He currently lives with his boyfriend and has never been married.

Chief Complaint The chief complaint is a verbatim recording of the patient's reason for seeking treatment or evaluation. Putting the chief complaint in the patient's own words, even if implausible or nonsensical, conveys valuable information about the person's capacity for insight and self-observation.

> “I took an overdose of some pills but I'm fine now.”

History of the Present Illness This section is a chronological description of how symptoms in the current episode have unfolded over time. The psychiatrist needs to determine not only the nature of symptoms, but also when they emerged and how they have progressed. The characteristics of symptoms should be described in detail; small distinctions may be diagnostically useful. Stating that a patient suffers from insomnia is less useful than describing the insomnia. Difficulty in falling asleep, difficulty in maintaining sleep, and a decreased need for sleep are each associated with different disorders. Attention should be paid to pertinent negatives as well as pertinent positives. In a patient complaining of depression, for example, the absence of vegetative symptoms is significant and should be mentioned. Whether the patient has been in treatment, has been taking any psychotropic medication, and has (or has not) been compliant are essential elements of the history of the present illness. If a patient has stopped taking a prescribed medication the reasons should be determined. Noncompliance is a symptom that needs to be investigated and not simply dismissed as poor judgment or character weakness. Noncompliance has many possible causes: unpleasant adverse effects, failure to understand the necessity for chronic medication despite symptomatic improvement, insufficiently treated symptoms such as the fear of being poisoned by medication, a reluctance to see oneself as psychiatrically impaired, or simply lacking the transportation and money to get a prescription refilled. Any current alcohol or other substance use should be described, including amounts, frequency, and last use. It is also useful to ask why the person came for treatment at this time, and what the patient believes to be causing the present symptoms.

> The patient is a 21-year-old male with a history of one prior suicide attempt who was brought to the hospital emergency room by ambulance, accompanied by his boyfriend and roommate, after having taken an overdose of atenolol (Tenormin) (twenty-five 25-mg tablets), zolpidem (Ambien) (twenty 10-mg tablets), and possibly fluoxetine (Prozac) pills (number unknown) in a suicide attempt. All of these medications had been prescribed for the patient's boyfriend.
>
> The evening before, the patient had had a fight with his boyfriend. He believed that his boyfriend wasn't giving him enough time and was not committed to the relationship. They slapped and punched each other. The patient went to a bar with another friend where he consumed four beers and a shot of vodka. When he arrived back at home, his boyfriend told him to go to sleep, which he interpreted as a continued lack of interest on the boyfriend's part. He locked himself in the bathroom and took many of his boyfriend's pills. The patient filled the empty pill bottles with water and left them in plain view on the sink.
>
> The patient then unlocked the door and went to bed, telling his boyfriend where to find syrup of ipecac. (Mr. A. carries the ipecac in his knapsack to induce vomiting when he drinks too much alcohol.) His boyfriend saw the pill bottles, realized the patient had overdosed, and tried to induce vomiting by administering syrup of ipecac. Unable to get the patient to vomit, he called an ambulance.
>
> Mr. A currently drinks alcohol once or twice each week, usually on the weekends. A usual amount of alcohol for him consists of two vodka tonics and eight or nine beers. Often, he drinks until he blacks out. He used cocaine this past New Year's Eve (“a few lines”) and occasionally uses heroin intransally (unknown amounts). He denies any history of intravenous drug use.
>
> He remembers always speaking and thinking very quickly, because he was always “so bright and talented and good looking and smart.” In addition, he has needed little sleep since age 15. The racing thoughts, pressured speech, and decreased need for sleep have become more pronounced since September, when he started feeling “very up.” Since then he has been getting at most 5 hours of sleep each night without feeling tired. He says he can be very influential, but that he has no special powers. He reports that he gets angry quickly and that his mood can change very easily. Since September he's also felt more depressed and physically restless. He considers suicide frequently, “just to escape the boredom of life.” He has lost 6 pounds in the last 2 months. There is no history of hallucinations or of delusional thought.

Past Psychiatric History The past psychiatric history describes all previous episodes and symptoms whether treated or not. The history should begin with the first onset of symptoms and progress chronologically to the current episode. It describes symptoms in detail and clearly delineates their longitudinal progress. Disorders that are chronic and relapsing are distinguished from isolated episodes of disturbance. It is particularly important to obtain the fullest possible information on prior treatments. The best predictor of treatment response is past experience. If a person has taken psychiatric medication before, it is essential to determine not only which drug, but the dosage and length of treatment, to distinguish nonresponse from a subtherapeutic drug trial. Therapeutic benefits and adverse effects should be noted. Similarly, if a patient has received psychotherapy it is important to establish which modality of therapy, at what frequency, for what length of time, and with what benefit.

> The patient has a history of violent outbursts. His mother reports that as an infant, he would bang his head against the floor. Throughout his childhood, he got into fights with other children and would even attack family members and teachers. In seventh grade, he threw a chair at a teacher. Once he attacked his older brother and kicked him in the head repeatedly until he lost consciousness and required medical attention. During his senior year in high school, the patient was forced to see the school therapist because of a heated argument with a teacher. The teacher claimed that the patient tried to hit him, and though the patient denied this, he was expelled. He met with the psychologist five or six times and stopped treatment when he graduated. There is a history of one previous suicide attempt 2 years ago, precipitated by the infidelity of his first boyfriend. The patient fashioned a noose and began to hang himself. When he began to feel pain, he stopped. He never told anyone about this attempt and never sought or received treatment. Psychotropic medication has never been prescribed.

Medical History The importance of a thorough, accurate medical history is difficult to overstate. The occurrence of major illness or surgery is likely to be of considerable significance in a person's life and may be the precipitant of psychiatric disturbance. For example, in response to having a heart attack, a middle-aged man might develop anxiety, depression, and a fear of sex. In addition, many medical conditions and their treatments cause psychiatric symptoms that are clinically indistinguishable from primary psychiatric disorders. Hypoglycemia can cause panic and anxiety; hypercalcemia, depression and lethargy; and acute porphyria, psychotic symptoms. Moreover, the presence of underlying medical conditions will inform treatment decisions: tricyclic antidepressants will be avoided in patients with cardiac conduction abnormalities, and bipolar I disorder patients with a history of renal disease are more likely to be treated with an anticonvulsant than with lithium. The names and dosing schedules for all currently prescribed nonpsychiatric drugs should be obtained to avoid possible adverse interactions with any new prescription.

In many instances the findings from a current physical examination and laboratory studies will be important in the diagnostic assessment. This is always true of hospitalized psychiatric patients, and may be necessary for selected outpatients. Few primary psychiatric disorders include physical signs. (One exception is panic disorder, which may include tachypnea, tachycardia, hypertension, and dilated pupils.) Their presence always warrants medical evaluation. Aspects of the clinical history may also suggest the need for medical investigation: abrupt onset of symptoms in an older adult with no prior psychiatric history, symptoms atypical for purely psychiatric disorders (e.g., vomiting and diarrhea or shaking chills), or a history of recent illness or treatment. In addition, abnormalities of the sensorium section of the mental status examination are most typical of delirium and dementia and indicate the need to look for underlying medical conditions.

> The patient has had streptococcal pharyngitis five times in the past 4 years. At age 15, he fractured his right wrist in a fight. He denies any other medical conditions. He takes no medications and has no known drug allergies. He has never been hospitalized or undergone any surgical procedures.

Family History Many psychiatric disorders are familial, and many of those appear to have a genetic component to the cause. Knowing who is in the patient's family and which, if any, psychiatric disorders have been diagnosed may help in diagnosis and treatment planning. Caution is urged however against overinterpretation of such data. Experienced clinicians are good at establishing the presence of a family psychiatric history but are less able to identify specific disorders in family members not personally interviewed. Even when specific disorders can be clearly identified, the information is of limited use because our knowledge of spectrum

disorders—that is, which disorders occur together in families and the relative risk of their occuring together—is still embryonic. On the other hand, the family history can clearly show who is in the family, who is available for support, who may be exacerbating symptoms, whether a general vulnerability to psychiatric disorders exists, and what stresses have been caused by a family member's illness. It is also useful to establish which treatments have been attempted and which have been successful; treatment response is often familial.

> The patient's immediate family consists of his father (age 51), his mother (age 49), one older brother (age 26), and one younger sister (age 19). His father suffers from alcohol dependence. Family history is otherwise negative for psychiatric disorders, medical disease, dementia, addiction, suicide attempts, and violence.

Personal History The personal history is intended to describe events of major significance throughout a person's life, to highlight those that may be etiologically significant, and to describe functioning over time. Which elements of the personal history are important will vary from patient to patient and cannot easily be prescribed. The information presented in the personal history will be shaped by other information from the interview. The psychiatrist will make ongoing clinical judgments about what is important and what is not. Major items commonly include early childhood friendships, education and any changes in school performance, romantic involvements, work history, military or jail experiences, and leisure activities.

The personal history may or may not include a detailed developmental history. A full development history begins with significant events and complications during pregnancy and delivery. Was the pregnancy full term? Did the patient's mother use drugs or alcohol during the pregnancy? Were there prenatal complications? Milestones of development in infancy and childhood (walking, talking, bowel and bladder control) can be described, although the information may be lacking or unreliable if the adult patient is the only source of information. Social development in childhood is revealed through information about friendships, schooling, and extracurricular activities. Often of considerable use is the patient's recollections of intrafamilial relationships including their relationships with each other and each separately with the patient.

A landmark of adolescence is the onset of puberty. The age of menarche, the circumstance of its onset, and preparations are likely to be significant emotional events in the lives of women. The growth of pubic and axillary hair signals the onset of puberty for boys. Late-onset puberty often has significant emotional and social consequences for boys: they are often less confident, more self-conscious, and less likely to be widely popular or leaders in school. Early experiences in dating, first sexual experiences, and any confusion or discomfort about sexual orientation are all important aspects of adolescent development.

Adolescence is also often a time of first drug experimentation. Information about which drugs, how much, how often, and under what circumstances should be obtained as well as changing patterns of drug or alcohol use. The development of career goals, undertaking advanced education, starting a first job, entering the military, and establishing a partnership (including marriage) all signify the transition from adolescence to adulthood.

Development continues through an individual's lifetime. The changes in adulthood are less likely to be as clearly demarcated or as universal as the developmental milestones of childhood or adolescence, but for many people the achievement of career and family goals will predominate. The examiner should inquire about current relationships, whether the patient is currently married and has ever been married before. Qualitative descriptions of interpersonal relations are important in diagnosing personality disorders and assessing suitability for some kinds of psychotherapy. In addition, recent research has drawn attention to the importance of strong, emotional, personal bonds in maintaining both physical and mental health.

The review of a patient's work history includes a summary of the jobs held, the length of time in each, and the reasons for leaving. Of considerable importance is the possible discrepancy between aspiration and achievement. The examiner should determine the extent to which psychopathology has interfered with the capacity for sustained productive work.

It is useful to ask patients what they do with their free time. Avocational, recreational, and humanitarian pursuits are important in the lives of many adults, and their absence may be diagnostically significant. For some people, military or prison history is important. As adults grow older, new issues such as children leaving home, the death of one's parents, retirement, and the loss of friends will emerge.

Some writers have advocated obtaining a detailed developmental history for all patients (at least one writer argues for inclusion of a detailed prenatal history), but such information is rarely of use in formulating a diagnosis or planning treatment for adult patients. Unlike a psychoanalytic assessment, which may be stretched over several sessions and which attempts to be comprehensive in its survey of developing character traits and ego strengths, a clinical diagnostic interview is more focused. A solid knowledge of psychopathology shapes the interview as it progresses. It is difficult to justify the time and expense of obtaining a detailed developmental history for a person with, for example, a simple phobia, obsessive-compulsive disorder, or panic disorder. Such information is necessary neither for establishing the diagnosis, nor for prescribing and implementing effective treatment.

On the other hand, describing functional capacity over time is necessary and useful. Deteriorating school performance, an irregular work history with failure to progress to higher levels of responsibility, premature discharge from the military, an inability to sustain friendships or romantic involvements for any period of time, may all have diagnostic and prognostic significance. The distinction between schizophrenia and bipolar disorder, both chronic, relapsing conditions with the possibility of psychotic symptoms, was first made on the basis of deterioration in the former and stability of function in the latter. Personality disorders that by definition attempt to identify core *trait* features will present with characteristic difficulties in interpersonal relationships or work capacity relatively unchanged over the course of a lifetime. The personal history may contain information helpful in making a prognosis as well as diagnosis. For example, a good premorbid adjustment reflected in school and work history indicates a good prognosis in patients diagnosed with schizoaffective disorder.

The personal history also helps identify key events that may have helped precipitate current symptoms: divorce, loss of work, death of a family member, serious financial setbacks. However, with the exception of posttraumatic stress disorder, identification of a precipitating event is not required to make a diagnosis. It may be useful although not necessary. Sexual functioning should be reviewed in the personal history. Basic screening questions include whether the patient is currently sexually active, who are preferred partners, whether there has been a change in sexual functioning, and the extent to which sex is pleasurable. If sexual difficulties emerge, more detailed questions can follow. When complaints of sexual disturbances do occur, the information is likely to be more appropriate for the history of the present illness than for the personal history.

The patient grew up in a medium-size city. His relationships with family members have always been difficult. While he often got into physical fights with his brother and his father, there is no history of physical, verbal, or sexual abuse.

In school, he had difficulty controlling his behavior, getting into fights and cursing a great deal. He was a good student when he worked, receiving As and Bs and honors in middle school and in the ninth grade. He failed all of his classes in the sophomore year and needed to repeat the year. He attributed this to ignoring school and partying too much. During his last year of high school, he attended a special arts school, studying drama and jazz. He graduated from high school after 5 years.

He had trouble making friends. He had a girlfriend for 2½ years in high school, but they broke up because they argued too much. He met his first boyfriend at age 17, but this boy cheated on him, precipitating the previous suicide attempt. In the past 2 or 3 years, the patient has found himself primarily attracted to men but still considers himself bisexual.

After high school, he lived with his parents for almost 2 years. During that time, he played drums in a band and held part-time jobs as a sales clerk. These jobs never lasted more than a few months, usually ending in verbal fights with his employer. He applied to music school but was not accepted because of his poor high school performance. Six months before admission, he moved to a large city to live with his uncle, hoping to take classes without credit. Lack of financial resources required him to work instead. He worked for 24 hours each week, off the books, as a veterinarian's assistant. He felt that his uncle was too controlling and domineering and moved in with another friend in October. In November, he met his current boyfriend, Mr. B., at a bar. A week and a half later, he moved into the boyfriend's house in the suburbs. Mr. B. reports that his relationship with the patient has been characterized by extreme suspiciousness and neediness. The patient has recurrent thoughts that his boyfriend doesn't care about him and is cheating on him, even though there is no evidence.

The patient has no savings and is supported by his boyfriend. His parents have severe financial difficulties and are unable to help him. Contact with family members is presently limited to his mother and his aunt because he doesn't get along with any other family members.

The patient has had about 25 different sexual partners, about half male and half female. He does not always practice safe sex and has engaged in high-risk behaviors such as receptive anal sex without using a condom. Because of this, he has been tested for human immunodeficiency virus (HIV) infection four times. The result was negative each time, most recently 2 months before admission.

There is a significant history of substance abuse. During his second and third years in high school, he smoked marijuana two to three times each week. He stopped after graduation and has since smoked marijuana only two or three times. He used cocaine once in high school and only occasionally in the past 2 years, snorting a few lines each time. He has used lysergic acid diethylamide (LSD) and phencyclidine (PCP) but never regularly. He has smoked cigarettes since age 13, ranging from half a pack to 2½ packs a day. He stopped smoking for 9 months 3 years ago but resumed and currently smokes half a pack a day.

MENTAL STATUS EXAMINATION

The mental status examination of psychiatric patients is analogous to the physical examination in physical medicine. It provides a format for the systematic observation and recording of information about a person's thinking, emotions, and behavior. These data combined with information from the history are the basis for formulating a differential diagnosis. As is true for the physical examination, a physician conducting a mental status examination notes only those findings present at the time of interview. Historical information is excluded. A patient may report having had auditory hallucinations the day before, but unless they are present when the examination is conducted, hallucinations are not recorded in the mental status examination.

The physician must also be as objective as possible in making mental status observations. Since 1980 the various editions of DSM have emphasized descriptive phenomenology in making psychiatric diagnoses to the exclusion of inferred, intrapsychological processes, in an explicit attempt to enhance diagnostic reliability. Observed data are always more reliable than inferred data. Correspondingly, the mental status examiner should strive to record findings that are as free of interpretation or inference as possible.

The formal organization of the mental status examination ensures completeness. In the actual interview of a psychiatric patient it is seldom necessary to proceed with an inflexible, prescribed series of questions. Much of the mental status examination is observational and can be made in the course taking the history. There are several specific tests of cognitive function, but much of this information can be obtained simply by talking with a patient. The experienced clinician does several things simultaneously in conducting a psychiatric interview: establishing rapport, eliciting important historical information, recognizing areas of greater or lesser emotional intensity, and making ongoing mental status observations.

ORGANIZATION OF THE MENTAL STATUS EXAMINATION

Numerous variations on the mental status examination have been described. The specific format matters less than ensuring that the observations are complete and logically organized to facilitate diagnosis. The outline presented here and summarized in Table 7.1-1 is widely used.

Appearance A brief description is given of the patient's appearance, behavior, and manner of relating to the examiner, with particular attention paid to abnormalities. Is the patient overdressed or underdressed? Is the patient wearing excessive, garish make-up? Is the patient disheveled, unkempt, or ungroomed? Is the patient cooperative, oppositional, hostile, seductive, or impassive? Are there unusual movements? Is the patient making smacking or chewing motions? Is there a tremor? Is the patient pacing? Although a comprehensive psychiatric assessment always includes a physical examination, obvious signs of physical illness (e.g., pallor, jaundice, labored breathing, or dilated pupils) are also mentioned under “appearance.”

The patient is a muscular young man appearing his stated age, wearing jeans, a white t-shirt, and sneakers. He wears several rings on his fingers and bracelets on both wrists. There is an obvious healing cut on his upper lip, which is slightly swollen. He is unshaven, but has an overall neat appearance and adequate hygiene. He sits with his arms crossed in a chair that swivels and uses his feet to swivel through roughly 90 degrees back and forth throughout the interview. He maintains good eye contact.

Speech The speech section of the mental status examination describes the physical production of speech, not the ideas being conveyed. Observations may be made about volume, rate, spontaneity, syntax, and vocabulary. Any speech abnormality such as dysarthria or aphasia is described. The speech of a manic patient may be loud and pressured. Conversely, the speech of a depressed patient may be soft and hesitant.

He speaks spontaneously and very rapidly, becoming pressured at times, but he is interruptible. Volume is occasionally loud. Rhythm and expressive intonation are normal. Speech is understandable, but some words are poorly articulated because of the high rate of speech production.

Emotional Expression It has been a convention for many years to describe emotional expression in terms of mood and affect, and those terms are still used extensively. Mood has commonly been described as the prevailing emotional state, and affect as the expression and expressivity of a patient's emotions. The term *affect* derives from the psychoanalytical literature and was originally intended to describe the feeling tone accompanying ideas or mental representations of external objects. *Mood* in turn was believed to derive from the summation of affects. By definition, affect would fluctuate with an individual's changing thoughts. Mood was more constant over time.

In this author's opinion, there are compelling reasons to abandon the distinction between mood and affect and no longer include a description of affect in the mental status examination. In its original psychoanalytic meaning, affect could be inferred but not directly observed because it was an intrapsychological phenomenon.

Moreover, the neurobiology of mood disorders, as now understood, is far more complex than the product of feeling tones accompanying ideas. This was implicitly acknowledged in 1987 when the revised third edition DSM (DSM-III-R) replaced the category of affective disorders with mood disorders. In a diagnostic scheme that is phenomenologic-descriptive and atheoretical with regard to etiology, this author believes the concept of affect is an anomaly.

Equally important is the fact that *affect* has acquired multiple and sometimes contradictory meanings. The distinction between mood and affect tends to focus on two parameters: constancy versus change and internal versus external. There has been a tendency to link the two, so that what is constant and internal (mood) is distinguished from what is transitory and external (affect). This coupling does not conform to clinical reality. It is quite possible for an individual's subjective experience of emotion to shift rapidly (e.g., in some manic states or frontal lobe syndromes). It is also quite possible for the external expression of emotion to be unvarying despite changes in subjective experience, as for example is sometimes seen in the residual phase of schizophrenia. In addition to its different definitions, no research has shown that affect can be reliably rated even by experienced clinicians.

Rather than attempting to distinguish mood and affect, the author's position is that in the mental status examination it is preferable to describe subjective and objective components of emotional expression separately. The subjective component is how individuals describe their inner emotional state: I feel happy; I feel sad, anxious, hopeless, exhilarated, etc. The objective component describes the way in which emotion is communicated through facial expression, vocal tone, and body posture. The two may be discordant. A patient whose eyes are filling up with tears may describe himself as feeling “fine.” Both objective and subjective components of emotion may fluctuate rapidly or remain unvarying. Both may be intense or blunted, and both may be appropriate or inappropriate to the topic being discussed. The long-term predisposition to jollity, melancholy, exuberance, or restraint is *temperament* rather than *mood*. Because the mental status examination describes only what is observed at the time of interview, an evaluation of temperament is not possible. As mentioned above however, the terms *mood* and *affect* are in common use and are to be found in most outlines of the psychiatric report and mental status.

Subjectively he reports feeling angry and depressed because he is being kept on a locked ward. Objectively he appears tense, angry, and sad at different times. His emotional expression is labile, of full range, and appropriate to content. His eyes fill with tears at times.

Thinking and Perception If psychotic symptoms exist, they are most likely to be described in this section. *Thinking* is subdivided into two subcategories: form and content.

Thought Form *Thought form* refers to the way in which ideas are linked, not the ideas themselves. Thoughts may be logically associated and goal directed. If they are not, a disorder of thought form (also formal thought disorder or sometimes, thought disorder) may exist. A number of different formal thought disorders have been described and are listed in Table 7.1-2. No thought disorder is pathognomonic for a particular disorder. However, a specific disorder of thought form is sometimes more characteristic of one diagnosis than another and may thereby convey diagnostic significance. For example, clang associations and flight of ideas are most closely associated with manic states, derailment and thought blocking with schizophrenia.

Table 7.1-2 Formal Thought Disorders

His thoughts are generally logical and goal directed, although he is quite circumstantial, launching into emotional accounts of relevant ideas but including many irrelevant details. There is no evidence of flight of ideas, loosening of associations, perseveration, tangentiality, or thought blocking.

Thought Content *Thought content* describes a patient's ideas. Abnormalities of content include delusions, ideas of reference, and obsessions. *Delusions* are fixed, false beliefs that are not shared by others as part of a religious or subcultural group. They are rigidly held regardless of evidence to the contrary. Except for delusional disorders, the type of delusion is not pathognomonic but may be associated closely enough with a particular disorder to have diagnostic implications. For example, delusions of guilt and somatic delusions are characteristic of (but not unique to) major depression with psychotic features. Delusions of persecution may be seen in schizophrenia and mania.

The patient who believes that everyday neutral occurrences carry specific, unique, and personal significance is said to have *ideas of reference*. A person may believe, for example, that a television announcer is attempting to convey a hidden message or that a stranger passing by on the street is signaling something of significance by brushing his hair or blowing his nose. Depending on the fixity and details of the belief, some ideas of reference may also be delusional.

Obsessions are unwanted, intrusive thoughts experienced by patients as symptomatic and beyond their control. The content of an obsession may be virtually anything but is often a disturbing thought of doing something embarrassing, hurtful, or dangerous. For example a young father may have thoughts of his daughter being sexually molested, a middle-aged woman of shouting obscenities during a church service. Because of the effort to control their thinking and because patients with obsessions are often deeply chagrined by their content, it is necessary to inquire specifically about their presence and not rely on voluntary reporting. *Preoccupations* are thoughts that predominate a person's thinking but are usually not experienced as unwanted or symptomatic. Examples include preoccupations with health, with money or social status, or with injustices.

A careful psychiatric examination always includes an assessment of suicide potential—even if there is no evidence of suicidality in the history—and of the potential for violence toward others. It is best to ask simple and direct questions; for example, Do you think about hurting yourself or about taking your life? The evaluation of suicidality and violence are discussed more fully below.

Mistrustful, suspicious thought is evident: He is preoccupied with thoughts that his boyfriend may have cheated on him. He also expresses extreme mistrust of the staff's motives, believing that the staff overanalyzes and carelessly misinterprets his statements and actions. He threatens to elope from the unit, claiming to know several ways to escape. He has inflated self-esteem, claiming to be extremely talented in a lot of areas, conceding that there are people who are better than he is, but that with a little practice, for example, he can be the best musician ever. He denies current suicidal or homicidal thoughts, intent, or plan.

Perception Perceptual abnormalities include hallucinations and illusions. Hallucinations are sensory perceptions generated wholly within the central nervous system (CNS) in the absence of any external stimulus. They can occur in any sensory modality: auditory, visual, tactile, olfactory, or gustatory. Auditory and visual hallucination are the most common. The modality of hallucination has no diagnostic significance, with the exception of *formication*, a tactile hallucination of insects crawling over or under the skin, which is strongly associated with withdrawal from alcohol and other central nervous system (CNS) sedatives. Illusions originate with true sensory stimuli, which are then misprocessed or misinterpreted. A patient looking at the shadows created on a wall by a rustling curtain may actually see threatening monsters. Illusions are widely believed to be more common in delirium than in other psychiatric disorders, despite the absence of empirical confirmation. Depersonalization and derealization (the sense that oneself or the world are not real) may also be recorded as perceptual abnormalities in the mental status examination.

He described hearing a man's voice, muffled, but at times intelligible, saying his name or short phrases such as “they're wrong.” There was no evidence of hallucinations in any other modality.

Sensorium This section includes assessment of several cognitive functions that collectively describe the overall intactness of the CNS. Cognitive disorders such as the syndromes of delirium and dementia and psychiatric disorders caused by drugs or general medical conditions are particularly likely to result in abnormalities in the

sensorium. The set of cognitive functions described in this section are subserved by different brain regions and, taken as a whole, provide a survey of whole brain functioning.

Alertness Alertness describes the degree of wakefulness and may range from fully awake and alert to comatose and nonresponsive. The degree of alertness may be stable or fluctuating.

Orientation Orientation is conventionally described in three spheres: person, place, and time. Orientation to person reflects an understanding of who one is and one's relationship to others. Orientation to time and place exists in multiple dimensions. If a patient is disoriented it is important to establish the degree. Is a patient aware of being in a hospital but not know which hospital? Does the patient believe it is a hotel instead of a hospital? Does the patient know the city in which the interview is being conducted? The date, day of the week, and time of day? The calendar year? If not, can the patient describe the season or distinguish morning from afternoon? It is common for hospitalized patients who are removed from normal environmental cues to be mildly disoriented to time.

Concentration Concentration describes the ability to sustain attention over time. Concentration is one of the cognitive functions most easily assessed simply by talking with a patient. Patients who forget the examiner's question, are distracted by extraneous stimuli, or lose track of what they are saying have impaired concentration. Concentration may be more formally tested in several ways. One of the most commonly taught and frequently misused tests is “serial sevens” in which a patient is asked to count backward from 100 by 7s. This is a valid test of concentration only if the person can comfortably perform the mental subtractions and if it is carried out for a substantial period of time. It is not intended to test the ability to perform calculations; the ability to concentrate and the ability to perform calculations should be evaluated separately. Alternative tests of concentration include counting backward by 3s, reciting the alphabet backward, spelling *world* backward, and naming the months of the year backward.

Memory Memory must be evaluated across the spectrum of immediate to remote. The brain substrates for long-term memory are different from those for immediate recall and short-term memory. This is illustrated clinically by patients with an anterograde amnesia such as Korsakoff's syndrome, in which long-term and immediate recall may be intact but recent memory is grossly impaired. As with concentration, much information about memory will be revealed in the course of the general interview. One test of immediate recall is to say (without inflection or verbal spacing) a series of numbers and have the patient repeat the series. A progressively longer sequence of numbers is presented, and both forward and backward recall are tested. Most adults can easily recall five r six numbers forward and three or four in reverse. Recent memory is for events several minutes to hours old and may be evaluated by giving patients the names of three or four unrelated objects and asking them to repeat them after 5 to 10 minutes. Remote memory describes events 2 or more years old. It is usually revealed in the course of obtaining patients' histories, although it may be necessary to confirm facts through collateral sources.

Calculations Calculations describes the ability to manipulate numbers mentally. Simple addition, subtraction, or multiplication questions may be used. Problems of money and change are often helpful with patients with limited educational background. For example, if a magazine costs $3.50 and you pay with a ten dollar bill, how much change should you be given? As noted above, the person should not be asked to perform serial subtractions to test calculating ability since it also requires concentration.

Fund Of Knowledge Fund of knowledge must be tailored to the unique circumstances and educational level of the individual. Patients are often asked to name presidents of the United States, starting with the incumbent and proceeding backward as far as can be remembered. This is not appropriate for everyone; recent immigrants to the United States may have difficulty with this even though they could give a detailed political history of their home country. Questions about current events, key geographical facts (what ocean lies between South America and Africa?), and sports may further help in the assessment.

Abstract Reasoning Abstract reasoning describes the ability to mentally shift back and forth between general concepts and specific examples. The capacity for abstract reasoning is usually not achieved before ages 12 to 13, and for some people is never achieved. The patient's use of jokes, metaphors, or aphorisms during the interview often reveals this ability. Of all the frequently used ways to test abstract reasoning, asking proverb interpretation is probably the least useful. For example, a clinician might ask the patient, “What does it mean, when someone says, “People who live in glass houses shouldn't throw stones'?” A conventional response, one that is able to generalize from the specifics of the proverb to the generalization might be “Don't criticize others of what you are guilty yourself.” A nonabstract response would address the concrete particulars without grasping the larger meaning, for example, “You would break the glass.” (Some answers will be idiosyncratic and difficult to classify as either abstract or concrete: “The police would see you and would come to arrest you.”)

One difficulty in using proverbs to assess abstract reasoning is that people tend to recite meanings learned over the years rather than reasoning through the proverb when it is given in the mental status examination. Indeed, interpreting an unfamiliar proverb can be difficult even for well-educated people whose abstract reasoning is unimpaired. In addition, proverbs tend to be culture bound, and their interpretation may be better evidence of cultural literacy than of the capacity to reason abstractly. (Consider, for example, the East African proverb, “You can't cover a horned animal with a canvas cloth.” This means that “truth will eventually be revealed despite efforts to suppress it.”) One alternative to proverb interpretation in assessing abstract reasoning is to ask for the similarity between two or more items, which really means asking to what conceptual abstraction do both belong. (A dog and a spider are both living things.) Asking for differences seldom gives useful information about abstract reasoning because the only plausible answers tend to describe concrete physical properties (e.g., “One has four legs, the other eight”). In another test the patient is given three objects, two of which are conceptually related, and asked which object does not belong (e.g., shirt, sock and hammer).

Insight This portion of the mental status examination describes patients' capacity to recognize and understand their own symptoms and illness. It does not measure the severity of illness. Patients with mild somatoform disorders may fail to recognize the emotional origins of their physical symptoms. On the other hand, some psychotic patients understand that their hallucinations are a symptom of a psychiatric disturbance that needs better control.

Judgment Observations about judgment in the mental status examination address two issues: can the person recognize prevailing social norms of behavior and comply, and will this person be able to cooperate with medical evaluation and treatment. Of all areas in the mental status examination, this is the least descriptive and most inferential. The psychiatrist must often draw from information in the history to supplement mental status findings. Some writers have advocated posing hypothetical situations to patients, asking, for example, “What would you do if you found a sealed, stamped, addressed envelope lying on the sidewalk?” Problems arise in using these kinds of data, particularly to the exclusion of other information. The presumed correct response to such scenarios is often obvious, and the answer may be very different from the patient's actual behavior. Moreover such questions often miss the complexity of variables shaping behavior and are simplistic in assuming a single correct response. An indigent homeless person who would open the envelope to see if there was money inside may be demonstrating good judgment in the context of his or her circumstances. Judgment may be more usefully assessed by observing the patient's behavior during the interview and by asking for elaboration on true incidents in the recent history—for example, “Why did you stop taking the medication?” or “Tell me what you were thinking when you gave away your car keys and registration to a stranger. Does it seem like a good idea now? Would you do it again?” Some psychiatrists advocate describing intelligence in the mental status examination. This cannot be done with any validity or reliability without the use of standardized instruments, and even then it may be difficult to distinguish between intelligence and education. Rather than record an impressionistic hunch, the examiner will do better to present the data of the evaluation without interpretation. Areas that loosely correlate with intelligence are vocabulary (under *speech*) and fund of knowledge and abstract reasoning in the *sensorium*.

Alertness: Alert and awake throughout the interview

Orientation: Intact to person, place, and time

Concentration: Spelled *word* backward correctly; serial 7s performed correctly and without hesitation

Memory: Registration and recent memory (5 minutes) intact for 3/3 phrases (blue rose, 37, happiness); long-term memory appears intact as evidenced by his detailed recall of past events in the history.

Calculations: 6 × 12 = 72; $2.00 – 65 cents = $1.35

Fund of knowledge: Good. He knew presidents back to Carter. He said WWII started around 1940 and then spontaneously added, “Hitler and Normandy.” He knew that Einstein was responsible for the theory of relativity.

Abstract thinking: Somewhat concrete; similarities: apple/orange—“both fruits”; poem/statue—“both have form”; fly/tree—“both are nature, both are iridescent green, flies fly around crap, which is brown, the same color as tree bark”

Insight: Poor. The patient does not recognize the presence of any illness or that his behavior is dangerous, stating, “Maybe I have a very mild case of mania, but if I need to be here, then 90 percent of everyone in the world needs to be locked up.” He initially refused to take medication and repeatedly says he doesn't need to be “locked up,” that he can take care of his minor relationship problems as an outpatient. He calls his drinking “minimal” and doesn't realize that it precipitates dangerous, self-destructive behavior.

Judgment: Fair. He cooperates with staff even though he doesn't think he needs hospitalization because he fears that a history of involuntary commitment would make it difficult for him to realize his goal of becoming a teacher. He says that the next time he is angry with his boyfriend, he will “work it out,” and not try to kill himself.

TECHNIQUES FOR THE PSYCHIATRIC ASSESSMENT

The psychiatrist wants not only to ask the questions necessary for formulating a differential diagnosis, but also to establish rapport and create an atmosphere of confidence and trust. Relaxed and trustful patients are more likely to provide useful information than those who are nervous or on guard.

Time and Setting The initial psychiatric assessment usually lasts between 45 and 90 minutes, with the length of time agreed upon in advance. Sometimes additional time is necessary to complete the evaluation, in which case it is better to schedule an additional session. Extending the length of the first session unilaterally is a discourtesy to the patient whose time may be scheduled with other demands, and it risks fatigue for both parties. In addition seeing the patient at different times helps to determine more accurately which presenting features are purely situational and how they have changed with time.

The evaluation should be conducted in a comfortable room with pleasant lighting. While there is no reason to make the room impersonal, dramatic paintings, panoramic views, or expensive antiques may be distracting to a patient during a first visit. The psychiatrist should budget the full amount of time allotted for the interview and attempt to ensure that there are no interruptions during the session. Routine phone calls and messages should be intercepted by an answering machine or secretary. A comfortable waiting area should be provided for patients who arrive early. Many psychiatrists prefer to have the patient's and examiner's chairs of relatively equal size and height to minimize any sense of intimidation.

Interview The psychiatrist should have in mind the categories of information needed and their structural organization in the summary evaluation. However, it is seldom useful to proceed with a prescribed check-list of questions. Rather, the interview should be shaped as it progresses, with the psychiatrist's knowledge of psychopathology used to ask detailed or probing questions about significant issues as they emerge. It is often best to begin with a general, open-ended question (e.g., “How can I help you?” or “What brings you here today?”) and to allow the patient to talk freely for several minutes before interposing further questions. However, the psychiatrist must keep in mind the information needed to formulate a diagnosis and treatment plan and be prepared to structure the interview more tightly if the patient appears to be vague and rambling. If a patient talks on with no pauses or natural endings, it may be necessary to interrupt and redirect. This can be done with courtesy and minimal disruption. For example, the psychiatrist might say, “Excuse me for interrupting, but I'd like to come back for a moment to the trouble you mentioned having had with sleep. How many hours of sleep did you get last night?”

Open-Ended and Closed-Ended Questions Open-ended questions ask the patient to speak spontaneously with relatively little structure or organization imposed by the examiner (e.g., “Tell me about your growing up”). Closed-ended questions on the other hand ask for factual answers to specific questions. (e.g., “How far did you go in school?”) Open-ended questions are commonly associated with psychodynamic interviewing and closed-ended questions with phenomenologic-descriptive diagnostic interviewing; this is an unhelpful oversimplification. A skillful diagnostic interview uses both types of questions. It is often useful to begin the interview with broad open-ended questions and to become more closed-ended and directive as the interview progresses.

There are a number of advantages in starting with open-ended questions: the content is less limited by the examiner's preconceptions, disorders of thought form are more likely to be revealed in spontaneous speech than in two- or three-word answers, and emotional responses may be more obvious. In addition, many patients prefer to tell their stories with their own words and emphases without interruption. Other patients, such as those who are psychotic, depressed, or paranoid, may need more structured questions. Closed-ended questions are particularly useful in clarifying information, in gathering factual data efficiently, and in describing the absence of key symptoms. Patients are unlikely to spontaneously describe what does not exist.

Supportive and Obstructive Interventions Psychiatrists do much more during an interview than ask questions. They provide feedback and information, offer reassurances, and respond emotionally to what the patient is saying. The psychiatrist's facial expression and body posture (correctly or erroneously) conveys information to the patient. Interventions may be classified as supportive or obstructive depending on the extent to which they increase the flow of information and enhance or diminish rapport. Interventions classified as supportive include

Encouragement. Patient: I am not very good at putting things into words.

Doctor: I think you have described the situation very well.

Reassurance. Doctor: I can understand how those experiences must have frightened you, but I think it is very likely they'll respond to treatment.

Acknowledging emotion. Doctor: Even now it brings tears to your eyes when you talk about your mother.

Nonverbal communication. Body posture and facial expression that convey interest, concern, and attentiveness.

Interventions classified as obstructive include

Compound questions. Doctor: Have you experienced any change in your appetite and sleeping?

Judgmental questions. Doctor: How do you think your wife felt when she found out about your affair?

Why questions. Doctor: Why do you feel anxious when you go outside?

Not following the patient's lead. Patient: I have trouble sleeping through the night.

Doctor: Any change in appetite?

Patient: I keep waking up out of nightmares about my daughter.

Doctor: Do you have less energy than usual?

Minimization or dismissal. Patient: I'm not able to keep my checkbook balanced the way I need to.

Doctor: Oh, I wouldn't worry about it. Lots of people don't even try.

Premature advice. Patient: Work is almost unbearable. My supervisor watches me like a hawk and criticizes the tiniest little mistake I make.

Doctor: Why not write her a memo and outline your grievances?”

Nonverbal communication. Yawning, checking one's watch. Patients can often detect an interviewer's inattention by the *absence* of facial expression or body movement.

Psychiatric interviewing is a complex, multifaceted task that is shaped by the personalities and circumstances of the interview. The concept of supportive and obstructive interventions has broad, general use but cannot be applied rigidly. There are circumstances in which an intervention that would be obstructive with other patients may be helpful or even necessary. For example, although it is usually most helpful to follow the patient's lead, the psychiatrist interviewing a hypochondriasis patient or a depressed patient with somatic concerns may need to ignore or interrupt perseveration about physical symptoms and redirect the interview to other topics. In addition, at times the psychiatrist may appropriately ask patients how their misdeeds are perceived by others (a judgmental question) to test empathy.

Interpreting Behavior During an Initial Diagnostic Interview The psychoanalytic techniques of interpretation and clarification should be used minimally if at all during an initial diagnostic assessment. The psychiatrist is unlikely to have enough information to make accurate interpretations. Moreover, the context of a trusting long-term relationship that facilitates acceptance of interpretations will be lacking. Patients who arrive late for the session may well be manifesting anxiety or ambivalence or may equally well have been delayed by circumstances beyond their control. Rather than interpreting the lateness, the psychiatrist would do better to express regret at the late start and offer sympathetic understanding to the patient for the problems causing the lateness. Separately, the psychiatrist can ask whether the patient was nervous or had mixed feelings about coming for the evaluation.

An exception to the rule against interpretations in the first session is the patient for whom a psychodynamic psychotherapy may be recommended. The psychiatrist may then want to evaluate the patient's response to a gentle test probe to assess suitability for psychodynamic work.

A 28-year-old woman who asked to start therapy with a psychoanalytic psychiatrist missed her first session. At the second session she reported that she simply forgot the appointment. The psychiatrist asks, “Is it possible that you let yourself forget on purpose?”

The woman laughs and says, “Well to tell you the truth, it's really my boyfriend whose been pushing me to start therapy; I'm not so sure. Do you think it's a good idea?”

Recording and Note Taking Electronic recording of sessions is seldom necessary and is often detrimental. Patients become self-conscious and guarded when their every word is being recorded. Many patients are concerned about the uses to which the recordings will be put and are rightfully concerned about potential abuses. Few patients are likely to feel comfortable knowing that their sometimes critical comments about another person or intimate discussion of private issues might be heard by anyone other than the evaluating psychiatrist. Recordings must never be made without the patient's knowledge and consent.

Psychiatrists vary in the extent to which handwritten notes are taken during a session. Some examiners take no notes at all during the evaluation and then write a summary after the patient has gone. They argue that note taking is a distraction to both psychiatrist and patient, that it becomes a barrier to subtle emotional observation and understanding, and that some patients become preoccupied by what is being written down, thus contaminating what they say. By not taking notes during the interview they train themselves to remember better the details of the session. They point out that when they do rely on notes they are unlikely to remember anything that was not written down.

Other psychiatrists take almost continuous notes. They believe it is possible to do so unobtrusively and that it provides subtle reassurance to patients about the seriousness with which the examiner is taking their statements. The result, they believe, is a more accurate record, not vulnerable to the distortions of memory.

Still other psychiatrists take notes selectively, when they believe that accurate and detailed documentation is necessary (e.g., a complicated history of previous medication treatments and variable responses). Individual psychiatrists develop techniques that work well for them, but they must remain mindful of the impact their individual decisions have on the ability to get useable information. Flexibility and common sense must not be lost. It would be destructively unfeeling for a psychiatrist to continue note taking in the middle of an intense emotional moment with the patient sobbing. Requests by the patient that notes not be taken may be explored but should always be respected.

Whether or not notes are taken during the session, psychiatrists have a medical and legal obligation to maintain a written record of every patient encounter. Such records document that the encounter occurred and that the assessment was complete. The record contains the historical and mental status data on which a diagnosis and treatment recommendations are based. The physician describes in detail all treatment recommendations and other advice given. (Psychiatrists should also routinely compile written descriptions of all telephone exchanges with patients. This is particularly important if the telephone exchange includes a change in treatment.) It is helpful to keep in mind the legal rule of thumb: “If it isn't written down, it didn't happen.”

Special Problems in Interviewing

PSYCHOTIC PATIENTS Patients with psychotic symptoms have difficulty thinking clearly and reasoning logically. Their ability to concentrate may be impaired, and

they may be distracted by hallucinations and delusional beliefs. Psychotic patients are often frightened and may be quite guarded. Quite often, the evaluation of a patient with psychotic symptoms needs to be more focused and structured than that of other patients. Open-ended questions and long periods of silence are apt to be disorganizing. Short questions are easier to follow than long ones. Questions calling for abstract responses or hypothetical conjectures may be unanswerable.

For patients with hallucinations, the full phenomenology of the hallucination should be explored. The patient is asked to describe the sensory misperception as fully as possible. For auditory hallucinations this includes content, volume, clarity, and circumstances; for visual hallucinations, this includes content, intensity, the situations in which they occur, and the patient's response. The evaluator should distinguish between true hallucinations on the one hand and illusions, hypnagogic and hypnopompic hallucinations, and vivid imaginings on the other. Hallucinations are perceived as real sensory stimuli and should not be dismissed as fanciful; however the psychiatrist should ask questions about their fixity and the patient's level of insight. “Does it ever seem that the voices are coming from your own thoughts?” or “What do you think is causing the voices?”

Delusions by definition are fixed, false beliefs. Delusional patients often come to psychiatric evaluation having had their beliefs dismissed or belittled by friends and family. They will be on guard for similar reactions from the examiner. It is possible to ask questions about delusions without revealing belief or disbelief (e.g., “Does it seem that people are intent on hurting you?” rather than, “Is there a plot to hurt you?”). Careless use of psychiatric jargon should be avoided, particularly in evaluating delusions. Words such as *grandiose, paranoid*, and indeed the word *delusion* itself will seem harsh and judgmental and are unlikely to be helpful in eliciting information. Many psychiatrists have found that patients can speak more freely when asked to talk about the accompanying emotions rather than the belief itself (“It must be frightening to think there are people you don't know who are plotting against you.”) Although the psychiatrist does not attempt to reason them away, a gentle probe may determine how tenaciously the beliefs are held (“Do you ever wonder whether those things might not be true?”).

Patients with paranoid delusions (and patients with high levels of nondelusional suspiciousness) are best evaluated with a respectful, but somewhat distant, formality and with scrupulous honesty. Efforts to reassure or ingratiate often increase suspicion. The psychiatrist must keep in mind the possibility of being incorporated into a delusional belief and should ask about it directly (“Are you concerned that I might try to hurt you?”).

Disorders of thought form can seriously impair effective communications. The evaluating psychiatrist should note formal thought disorders while minimizing their adverse impact on the interview. When derailment is evident, the psychiatrist typically proceeds with questions calling for short responses. For a patient experiencing thought-blocking, the psychiatrist needs to repeat questions, remind the patient of what was already said, and in general provide an organization for thinking that the patient is unable to provide.

DEPRESSED AND POTENTIALLY SUICIDAL PATIENTS Severely depressed patients may also have difficulty concentrating, thinking clearly, and speaking spontaneously. The intensity of mood disturbance can seem all consuming and may well lead to distortions in thinking and perception. Some depressed patients will have psychotic symptoms in addition to cognitive difficulties. The psychiatrist evaluating a depressed patient may need to be more forceful and directive than usual. It will sometimes seem that the examiner must provide all the emotional and intellectual energy for both participants. Although depressed patients should not be badgered, long silences are seldom useful, and the examiner may need to repeat questions more than once. Ruminative patients—for example, those who continually repeat how worthless or guilty they are—will need to be interrupted and redirected.

All patients must be asked about suicidal thoughts. Depressed patients may need to be questioned more fully. A thorough assessment of suicide potential addresses intent, plans, means, and perceived consequences as well as history of attempts and family history of suicide. Many patients will mention their thoughts of suicide spontaneously. If not, the examiner can begin with a somewhat general question such as, “Do you ever have thoughts of hurting yourself?” or “Does it ever seem that life isn't worth living?” These can then be followed up with more specific questions. The examiner must feel comfortable enough to ask simple, straightforward, noneuphemistic questions. Asking about suicide does not increase the risk. The psychiatrist is not suggesting a course of action the patient has not already contemplated. Specific, detailed questions are essential for prevention.

Intent The examiner must determine the seriousness of the wish to die. Some patients report that they wish they were dead but would never intentionally do anything to take their own lives. This level of intent is often referred to as *passive suicidal ideation*. Other patients express greater degrees of determination. Near the other end of the spectrum of intent is the patient who says, “I've decided I have to kill myself and nothing you can say or do will change that.” At the most extreme level of determination are the patients most difficult to help, those who tell no one about their suicidal plans and proceed in a deliberate, systematic manner. It is also useful to ask about restraining influences, both internal and external (e.g., “Do you worry that you might not be able to resist those impulses?” or “How have you been able to keep from hurting yourself so far?”). Patients with auditory hallucinations commanding then to kill themselves often describe the hallucinations as irresistible despite any real desire to die.

Plans Patients with well-formulated plans are generally at greater risk than patients who don't know what they would do, but the method of suicide is not always a reliable indication of the risk. Even though some actions such as jumping or shooting are much more likely to be fatal than others, patients make mistakes. A pill overdose taken at the time a spouse is expected to arrive home may become deadly if the spouse is delayed in traffic. The psychiatrist should also ask about preparatory actions such as giving away goods or putting one's estate in order.

Means Asking patients about the intended means of suicide is helpful in two ways. First it clarifies the urgency of the situation; persons wanting to shoot themselves who have a loaded gun at home are more dangerous than those who have no idea where to find a gun. Second, the understanding of intent is sharpened by knowing whether a patient has thought through the steps necessary to carry out the action.

Perceived Consequences Patients who see something desirable resulting from their deaths are at increased risk for suicide. A reunion fantasy, the belief that a person will be reunited with a deceased loved one, may be a powerful motivating force toward suicide. On the other hand, some potentially suicidal patients are restrained by what they see as negative consequences (e.g., “My children need me too much; they'd never be able to get along without me,” or “I couldn't hurt myself. My parents would never get over their grief.”). The psychiatric history and the family history for all patients, even those not currently suicidal, should mention any previous suicide attempt or suicides by family members. Both circumstances are recognized to increase the current risk, even if previous attempts were thought to be superficial.

At times treatment must take precedence over evaluation. In rare circumstances the threat of suicide is so imminent that immediate action must be taken to hospitalize the patient. Even during a first evaluation session, the psychiatrist must be prepared to make whatever professional response is necessary to safeguard the well-being of the patient.

AGITATED AND POTENTIALLY VIOLENT PATIENTS Whether in a private office or a psychiatric emergency room, psychiatrists sometimes find themselves interviewing potentially violent patients. In these circumstances, the task is twofold: to conduct an assessment but also to contain behavior and limit the potential for harm.

Most unpremeditated violence is preceded by a prodrome of accelerating psychomotor agitation. The patient may begin pacing and pounding the fist in a hand. Speech may become loud, abusive, obscene, and threatening. The temporal arteries may begin to throb. Researchers and clinicians in emergency psychiatry suggest that the prodrome may last from 30 to 60 minutes before erupting into physical violence. Thus the psychiatric evaluator has both early signals of impending violence and a period of time in which the agitation may be quieted.

Several steps can be taken to minimize the agitation and potential risk. The interview should be conducted in a quiet, nonstimulating environment. There should be enough space for the comfort of both patient and psychiatrist, with no physical barrier to leaving the examination room for either of them. During the interview, the psychiatrist should avoid any behavior that could be misconstrued as menacing: standing over the patient, staring, or touching.

The psychiatrist must ask the questions necessary to complete an adequate evaluation but must attempt not to be provocative. It is certainly appropriate to allow the patient to drink water, use a bathroom, and, for extended evaluations in an emergency room, eat food. However, these should never be offered as bargaining chips (“I'll let you get a drink of water if you'll tell me what happened just before the police brought you here.”) The examiner must also avoid promising outcome in exchange for cooperation (“If you'll just talk with me for another half hour, I'll make sure you don't have to go into the hospital.”)

The psychiatrist should ask whether the patient is carrying weapons and may ask the patient to leave the weapon with a guard or in a holding area. The psychiatrist should not request that patient hand over any weapons. Dangerous mishaps can occur during the transfer; moreover, the sudden shift in power created by an armed psychiatrist may feel extremely threatening to paranoid patients. If the patient's agitation continues to increase, the psychiatrist may need to terminate the interview.

Depending on the setting, assistance from security personnel, physical, or chemical restraints may be appropriate. The physician's own subjective sense of comfort or fear should be heeded. A frightened, intimidated examiner may be incapable of an accurate professional evaluation.

PATIENTS FROM DIFFERENT CULTURES AND BACKGROUNDS Differences in race, nationality, and religion and other significant cultural differences between patient and interviewer can impair communication and lead to misunderstandings. Despite its widespread use throughout the world, the possible cultural biases of DSM-IV are still being debated; for example, the distinctions between mood disorders and somatoform disorders appear less valid in some countries than in the United States.

In addition, it may be difficult for a culturally naive psychiatrist to evaluate symptoms that are relative rather than absolute. There is usually no difficulty in documenting the presence of auditory hallucinations regardless of cultural differences. However, assessing whether or not a delusion is “bizarre” (as required by DSM-IV for delusional disorder) is more difficult because “bizarre” has meaning only in reference to cultural norms. The belief by East Africans in the direct intervention of ancestral spirits in the day-to-day life of individuals is commonplace. The chief executive officer of an American corporation who announces that he will divest the company of two subsidiaries because of signals he received that morning from ancestral spirits will be thought exceedingly bizarre by colleagues and shareholders. Personality disorders, whose criteria are preponderantly relative rather than absolute (e.g., “shows arrogant, haughty behaviors or attitudes”), are notoriously difficult to diagnose cross-culturally.

Apart from diagnostic categories, the vocabulary used to describe emotional distress varies from culture to culture. European-Americans commonly describe symptoms in terms of named emotions (“I've been feeling anxious and depressed all week”). Hispanic-Americans are more likely to describe physical symptoms (“I've had a headache all week, and I'm so tired I can hardly move”). Sometimes symptoms that are commonplace within a culture are unheard of to outsiders. Residents of Anglophonic countries in East and West Africa often describe the sensation of a snake crawling under their skin, moving from one part of the body to another. This appears to be a symptom of general emotional distress without particular diagnostic significance. Heard by a Western physician, the symptom may be misinterpreted as a somatic delusion or ignored altogether, because it does not register in the examiner's conceptual understanding of disorders.

Additional problems are encountered when doctor and patient speak different languages. When an interpreter is needed, the person should be a disinterested third party, unknown to the patient. Using family or friends to translate inevitably invites distortions in what the patient is said to report. Translators must be instructed to translate verbatim what the patient says—a difficult task for even the most experienced professional translators. Some words and expressions are simply untranslatable. It may be impossible to convey a formal thought disorder through translation.

An additional difficulty may arise in establishing rapport between doctor and patient of different ethnic or cultural groups. The use of honorifics, the extent of direct eye contact considered appropriate, or whether it is acceptable for men and women to shake hands, all vary considerably among different groups. Patients from minority groups may be quite guarded in speaking with a doctor from the majority group. Some groups such as traditional Chinese-Americans strongly believe that family problems should not be discussed outside the family, including with physicians. The evaluating psychiatrist must proceed with humility and respect. Rather than offer reassurances of understanding and acceptance, it is usually better to ask, “Have I understood this in the way you meant it?”

SEDUCTIVE PATIENTS The warmth, openness, acceptance, and understanding that are helpful to most psychiatric interviews may engender feelings of romantic longing in some patients, especially (but not exclusively) those who are lonely and socially isolated. Other patients may have flirtatious and seductive ways as their habitual style of relating with other people. Seductiveness may be manifested in a patient's dress, behavior, and in what is said. It runs the gamut from gentle suggestion to explicit proposition. A young man may sit with his legs spread wide apart, a young woman may wear a low-cut revealing dress, or a middle-aged woman, when shaking hands, may hold the psychiatrist's hand a few seconds longer than appropriate for the situation.

Of course sex is not the only enticement with which psychiatrists can be seduced. Patients may offer insider information for profitable trading in the stock market, promise an introduction to a movie star friend, or suggest that they will dedicate their next novel to the psychiatrist. While it is easy to understand that some offers by patients such as the possibility of a sexual involvement cannot be acted on without considerable harm to the patient, others may seem more innocuous. However, because they nearly always introduce a different agenda into the therapy than that originally contracted for and because they create additional, more ambiguous levels of obligation between therapist and patient, any psychiatric work is inevitably contaminated, and the ability to help the patient is compromised. Consequently, gaining material or social benefit from the patient other than the agreed upon fee is unethical.

Whether to offers of sex, money, or celebrity, the psychiatrist's response is the same. In the course of ongoing psychotherapy and in the context of an established relationship, seductive behavior is discussed and examined in an effort to understand its meaning. Is it for example, a way of distancing, of gaining control, or of compensating for feelings of vulnerability and inferiority? To what extent are the feelings being expressed by the patient part of the transference? The psychiatrist should make it clear that what is being offered will not be accepted, in a way that preserves good rapport and does not unnecessarily assault the patient's self-esteem.

Seductive behavior during an initial psychiatric assessment must be handled somewhat differently. When the behavior is mild and indirect, it may be best to ignore it; commenting on a woman's exposed cleavage only makes it clear that the psychiatrist is picking up sexual cues and is most unlikely to facilitate the interview. More-explicit propositions call for more-direct responses and may afford the psychiatrist the chance to explain the nature of the therapeutic relationship and the need to establish boundaries. The psychiatrist should also make clear that it is the violation of those boundaries that is being rejected and not the patient. For example, to the patient who offers a celebrity introduction, the interviewer might reply, “That's very nice of you to propose, but I think I will best be able to help you if we pretty much stick to the issues that brought you in to see me.”

PATIENTS WHO LIE A fundamental stance in psychiatric interviewing is recognizing that what is being heard may not be literal truth. The unreliability of memory and the vagaries of psychopathology through which a patient's narrative is processed will distort and falsify. The interviewer understands that what is historically untrue may nevertheless be emotionally true and is therefore a meaningful part of the diagnostic assessment or psychotherapy.

At times patients lie consciously with the explicit intent of deceiving the therapist. The purpose may be secondary gain (e.g., exemption from jury duty, a supply of psychoactive drugs, a leave of absence from graduate school), in which case the person is malingering. Malingering is not a mental disorder in DSM-IV. More rarely a patient will explicitly lie not for any obvious external advantage but simply for whatever psychological benefit is conferred by assuming the sick role, in which case the person may have a factitious disorder, which is a DSM-IV mental disorder.

Because psychiatrists do not have recourse to biologic markers or other external validating criteria, the patient's report must be accepted as an honest statement of experience. There is no way to establish whether a person is experiencing auditory hallucinations other than through self-report. Nevertheless, an experienced clinician may detect subtle discrepancies, internal inconsistencies, or suspiciously atypical symptoms; these can certainly be queried without necessarily assuming that the patient is lying.

> A 29-year-old woman describes almost unremitting migraine headaches and is asking for narcotic pain medication.
>
> Patient: I really need your help. The pain is unbearable. I can't do anything anymore. I just want to lie in bed in a dark room with the cover pulled over my head.
>
> Doctor: That does sound miserable. But I'm struck by the fact that you obviously care about your appearance and have given some time and attention to your hair, makeup, and the way you are dressed. Was that despite the pain you have been describing?

Of course the examiner is more likely to be deceived during the initial diagnostic assessment than in an ongoing psychotherapy in which the therapist has much more knowledge of a patient's background, thinking, and functioning over time. It may be difficult to catch a practiced liar in an initial session. Arguably the interviewer should not try. Being lied to angers most people, certainly no less psychiatrists who must depend on trust to perform their work. However, believing a patient's lies is not a professional failure. Psychiatrists are trained to detect, understand, and treat psychopathology, not to function as lie detectors. While a certain level of suspicion is essential in the practice of psychiatry, the clinician determined never to be taken in by deceitful patients will approach patients with such exaggerated suspiciousness that therapeutic work is not possible.

Finally, not all patients' untruths are conscious lies. Patients with somotoform disorders such as conversion disorder or pain disorder are presumably unaware of the

emotional bases of their physical complaints. In describing their somatic symptoms they are stating a psychological reality, not attempting to deceive the interviewer.

Empathy A diagnostic interview often provides considerable relief to patients. Puzzling and sometimes frightening symptoms are framed in the context of medical understanding. Bizarre experiences can be rationally understood and intelligently organized in meaningful ways that allow us to make informed predictions about treatment response and recovery. Of equal importance to an intellectual understanding is our capacity to understand emotionally the experiences of our patients.

Empathy is an essential characteristic of psychiatrists, but it is not a universal human capacity. An incapacity for normal understanding of what other people are feeling appears to be central to the disturbance of certain personality disorders such as antisocial and narcissistic personality disorders. While empathy can probably not be created, it can be focused and deepened through training, observation, and self-reflection. It manifests in clinical work in a variety of ways. An empathic psychiatrist may anticipate what is felt before it is spoken and can often help patients articulate what they are feeling. Nonverbal cues such as body posture and facial expression are noted. Patients' reactions to the psychiatrist can be understood and clarified.

Patients sometimes say, “How can you understand me if you haven't gone through what I'm going through?,” but clinical psychiatry is predicated on the belief that it is not necessary to have other people's literal experiences to understand them. The shared experience of being human is often enough. Whether in an initial diagnostic setting or in an ongoing therapy, patients draw comfort from knowing that we are not mystified by their suffering.

SUGGESTED CROSS-REFERENCES

Section 9.3 deals with the typical signs and symptoms of psychiatric illness, Section 9.5 deals with neuropsychological and intellectual assessment of adults, and Section 9.8 deals with psychiatric rating scales. Similarly, Chapter 10, on the clinical manifestations of psychiatric disorders, is an essential correlate to interviewing and examining the patient. More-specialized focus is provided in Section 2.1, which deals with the clinical assessment and approach to diagnosis in neuropsychiatry. Section 3.1 on perception and cognition and Section 3.4 on the biology of memory amplify points made in this section. Section 29.1 includes more-detailed information on suicide, and Section 29.2 includes information on other psychiatric emergencies. Additional relevant information is found in Chapter 45, which deals with mood disorders and suicide in children and adolescents. Taking a developmental history implies familiarity with the aspects of normal and abnormal development; readers may find the following sections of special interest: Section 6.2 deals with Erik H. Erikson and his ideas about child and adult development; Chapter 32 deals extensively with normal development in children and adolescents; adult development is covered at great length in Chapter 50; and normal aging is the focus of Section 51.2c.

SECTION REFERENCES

Akiskal H, Van Valkenburg C: Mood disorders. In *Diagnostic Interviewing*, ed 2, M Hersen, SM Turner, editors. Plenum, New York, 1994.

Akiskal HS, Akiskal A: Mental status examination: The art and science of the clinical interview. In *Diagnostic Interviewing*, ed. 2. M Hersen, SM Turner, editors. Plenum, New York, 1994.

Alarcon RD: Culture and psychiatric diagnosis: Impact on DSM-IV and ICD-10. Psychiatr Clin North Am *18:*449, 1995.

Andreasen NC, Black DW: Interviewing and assessment. In *Introductory Textbook of Psychiatry*, ed 2, American Psychiatric Press, Washington, DC, 1995.

Chikara F, Manley MRS: Psychiatry in Zimbabwe. Hosp Community Psychiatry *42:*943, 1991.

Dunner DL: Diagnostic assessment. Psychopharmacology *16:*431, 1993.

Faraone SV, Tsuag MT: Measuring diagnostic accuracy in the absence of a “gold standard.” Am J Psychiatry *151:*650, 1994.

Horowitz MJ: Personality disorder diagnoses (Letter). Am J Psychiatry *155:*1464, 1998.

Goodwin DW, Alderson P, Rosenthal R: Clinical significance of hallucinations in psychiatric disorders. Arch Gen Psychiatry *24:*76, 1971.

Gabbard GO: Finding the “person” in personality disorders (Editorial). Am J Psychiatry *154:*891, 1997.

Kandel ER: A new intellectual framework for psychiatry. Am J Psychiatry *155:*457, 1998.

Kaplan HI, Sadock BJ, editors: *Kaplan and Sadock's Synopsis of Psychiatry*, ed 8. Williams & Wilkins, Baltimore, 1998.

Lazare A, Alonso A: The mental status examination: III. Psychodynamic dimensions. In *Outpatient Psychiatry: Diagnosis and Treatment*, ed 2, A Lazare, editor. Williams & Wilkins, Baltimore, 1989, p 200.

Lewis JM: For better or worse: interpersonal relationships and individual outcome. Am J Psychiatry *155:*582, 1998.

Lieberman JA, Rush AJ: Redefining the role of psychiatry in medicine. Am J Psychiatry *153:*1388, 1996.

Lyketsos CG, Aritzi S, Lyketsos GC: Effectiveness of office-based psychiatric practice using a structured diagnostic interview to guide treatment. J Nerv Ment Dis *182:*720, 1994.

Mahl PC, Warrick McLaughlin GD: Listening to the patient. In *Psychiatry*, A Tasman, J Kay, JA Lieberman, editors. Saunders, Philadelphia, 1997.

Mahrer AR, Edwards HP, Durak GM, Sterner I: The psychotherapy patient and the initial session: What to do with the emotional state. Psychother Patient *1:*39, 1985.

Manschreck TC, Keller MB: The mental status examination: I. General appearance and behavior, emotional state, perception, speech and language. In *Outpatient Psychiatry: Diagnosis and Treatment*, ed 2, A Lazare, editor. Williams & Wilkins, Baltimore, 1989.

Perry SW, Stein SP: Common misconceptions about the mental status exam (Letter). Am J Psychiatry *142:*1391, 1985.

Remick RA, Sadovnick AD, Gimbarzevsky B, Lam RW, Athanasios PZ, Huggins MJ: Obtaining a family psychiatric history: Is it worth the effort? Can J Psychiatry *38:*590, 1993.

Risen CB: A guide to sexual history taking. Psychiatr Clin North Am *18:*39, 1995.

Rush AJ: Problems associated with the diagnosis of depression. J Clin Psychiatry *51:*15, 1990.

Sandler J, Dare C, Holder A: *The Patient and the Analyst: The Basis of the Psychoanalytic Process*. International Universities Press, New York, 1976.

Schreiber SC: The psychiatric interview, psychiatric history, and mental status examination. In *Synopsis of Psychiatry*, RE Hales, SC Yudofsky, editors. American Psychiatric Press, Washington, DC, 1996.

Schwartz E: The mental status examination. In *Clinical Diagnostic Interviewing*, R Craig, editor. Jason Aronson, Northvale, NJ, 1989.

Shea SC: Contemporary psychiatric interviewing: Integration of DSM-III-R, psychodynamic concerns, and mental status. In *Handbook of Psychological Assessment*, vol 131, Goldstein G, Hersen M, editors. Pergamon, New York, 1990.

Silberman EK, Certa K: Psychiatric interview: Settings and techniques. In *Psychiatry*, A Tasman, J Kay, JA Lieberman, editors. Saunders, Philadelphia, 1997.

Strauss GD: The psychiatric interview, history, and mental status examination. In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Swartz S: Sources of misunderstanding in interviews with psychiatric patients. Professional Psychol Res Pract *23:*24, 1992.

Tucker GJ: Putting DSM-IV in perspective (editorial). Am J Psychiatry *155:*159, 1998.

*Westen D, Shedler J: Revising and assessing axis II, part I: Developing a clinically and empirically valid assessment method. Am J Psychiatry *156:*258, 1999.

*Westen D, Shedler J: Revising and assessing axis II, part II: Toward an empirically based and clinically useful classification of personality disorders. Am J Psychiatry *156:*273, 1999.

Textbook of Psychiatry

7.2 PSYCHIATRIC REPORT AND MEDICAL RECORD

BENJAMIN J. SADOCK, M.D.

PSYCHIATRIC REPORT

The psychiatric report is a written document that details the findings obtained from the psychiatric history and mental status examination. For those who prefer to use it, the report may follow the outline described in the previous section; however, it may also be formatted in other ways, providing all the pertinent data are recorded.

The psychiatric report includes a final summary of both positive and negative findings and an interpretation of the data. It has more than descriptive value; it has meaning that helps provide an understanding of the case. The examiner addresses critical questions in the report: Are future diagnostic studies needed and if so which ones? Is a consultant needed? Is a comprehensive neurological workup needed including an electroencephalogram or computerized tomography scan? Are psychological tests indicated? Are psychodynamic factors relevant? The report includes a diagnosis made according to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), which uses a multiaxial classification scheme consisting of five axes, each of which should be covered (see Figure 9.1-1). A prognosis is also discussed in the report, with both good and bad prognostic factors listed. Finally, a treatment plan discusses and makes firm recommendations about management issues.

A detailed outline of the psychiatric report follows. It is one of the most comprehensive available, developed jointly by the author and Harold I. Kaplan.

I. **Psychiatric history**
   A. *Identification:* name, age, marital status, sex, occupation, language if other than English, race, nationality, and religion if pertinent; previous admissions to a hospital for the same or a different condition; with whom the patient lives
   B. *Chief complaint:* exactly why the patient came to the psychiatrist, preferably in the patient's own words; if that information does not come from the patient, note who supplied it
   C. *History of present illness:* chronological background and development of the symptoms or behavioral changes that culminated in the patient's seeking assistance; patient's life circumstances at the time of onset; personality when well; how illness has affected life activities and personal relations—changes in personality, interests, mood, attitudes toward others, dress, habits, level of tenseness, irritability, activity, attention, concentration, memory, speech; psychophysiological symptoms—nature and details of dysfunction; pain—location, intensity, fluctuation; level of anxiety—generalized and nonspecific (free floating) or specifically related to particular situations, activities, or objects; how anxieties are handled—avoidance, repetition of feared situation, use of drugs or other activities for alleviation
   D. *Past psychiatric and medical history:* (1) emotional or mental disturbances—extent of incapacity, type of treatment, names of hospitals, length of illness, effect of treatment; (2) psychosomatic disorders: hay fever, arthritis, colitis, rheumatoid arthritis, recurrent colds, skin conditions; (3) medical conditions: follow customary review of systems; sexually transmitted diseases; alcohol or other substance abuse; at risk for acquired immune deficiency syndrome (AIDS); (4) neurological disorders: headache, craniocerebral trauma, loss of consciousness, seizures or tumors
   E. *Family history:* Elicited from patient and from someone else, since quite different descriptions may be given of the same people and events; ethnic, national, and religious traditions; other people in the home, descriptions of them—personality and intelligence—and what has become of them since patient's childhood; descriptions of different households lived in; present relationships between patient and those who were in family; role of illness in the family; family history of mental illness; where does patient live—neighborhood and particular residence of the patient; is home crowded; privacy of family members from each other and from other families; sources of family income and difficulties in obtaining it; public assistance (if any) and attitude about it; will patient lose job or apartment by remaining in the hospital; who is caring for children
   F. *Personal history (anamnesis):* history of the patient's life from infancy to the present to the extent it can be recalled; gaps in history as spontaneously related by the patient; emotions associated with different life periods (painful, stressful, conflictual) or with phases of life cycle
      1. Early childhood (through age 3)
         a. Prenatal history and mother's pregnancy and delivery: Length of pregnancy, spontaneity and normality of delivery, birth trauma, whether patient was planned and wanted, birth defects
         b. Feeding habits: breast-fed or bottle-fed, eating problems
         c. Early development: maternal deprivation, language development, motor development, signs of unmet needs, sleep pattern, object constancy, stranger anxiety, separation anxiety
         d. Toilet training: age, attitude of parents, feelings about it
         e. Symptoms of behavior problems: thumb sucking, temper tantrums, tics, head bumping, rocking, night terrors, fears, bed wetting or bed soiling, nail biting, masturbation
         f. Personality and temperament as a child: shy, restless, overactive, withdrawn, studious, outgoing, timid, athletic, friendly patterns of play, reactions to siblings
         g. Early or recurrent dreams or fantasies
      2. Middle childhood (ages 3 to 11); early school history—feelings about going to school, early adjustment, gender identification, conscience development, punishment; social relationships, attitudes toward siblings and playmates
      3. Later childhood (prepuberty through adolescence)
         a. Peer relationships: number and closeness of friends, leader or follower, social popularity, participation in group or gang activities, idealized figures; patterns of aggression, passivity, anxiety, antisocial behavior
         b. School history: how far the patient went, adjustment to school, relationships with teachers—teacher's pet or rebellious—favorite studies or interests, particular abilities or assets, extracurricular activities, sports, hobbies, relationships of problems or symptoms to any school period
         c. Cognitive and motor development: learning to read and other intellectual and motor skills, minimal cerebral dysfunction, learning disabilities—their management and effects on the child
         d. Particular adolescent emotional or physical problems: nightmares, phobias, masturbation, bed wetting, running away, delinquency, smoking, drug or alcohol use, weight problems, feeling of inferiority
         e. Psychosexual history
            i. Early curiosity, infantile masturbation, sex play
            ii. Acquiring of sexual knowledge, attitude of parents toward sex, sexual abuse
            iii. Onset of puberty, feelings about it, kind of preparation, feelings about menstruation, development of secondary sexual characteristics
            iv. Adolescent sexual activity: crushes, parties, dating, petting, masturbation, wet dreams and attitudes toward them
            v. Attitudes toward same and opposite sex: timid, shy, aggressive, need to impress, seductive, sexual conquests, anxiety
            vi. Sexual practices: sexual problems, homosexual and heterosexual experiences, paraphilias, promiscuity
         f. Religious background: strict, liberal, mixed (possible conflicts), relationship of background to current religious practices
      4. **Adulthood**
         a. Occupational history: choice of occupation, training, ambitions, conflicts; relations with authority, peers, and subordinates; number of jobs and duration; changes in job status; current job and feelings about it
         b. Social activity: whether patient has friends or not; is he or she withdrawn or socializing well; social, intellectual, and physical interests; relationships with same sex and opposite sex; depth, duration, and quality of human relations
         c. Adult sexuality
            i. Premarital sexual relationships, age of first coitus, sexual orientation

ii. Marital history: common-law marriages, legal marriages, description of courtship and role played by each partner, age at marriage, family planning and contraception, names and ages of children, attitudes toward raising children, problems of any family members, housing difficulties if important to the marriage, sexual adjustment, extramarital affairs, areas of agreement and disagreement, management of money, role of in-laws
iii. Sexual symptoms: anorgasmia, impotence, premature ejaculation, lack of desire
iv. Attitudes toward pregnancy and having children; contraceptive practices and feelings about them
v. Sexual practices: paraphilias such as sadism, fetishes, voyeurism, attitude toward fellatio, cunnilingus; coital techniques, frequency

d. Military history: general adjustment, combat, injuries, referral to psychiatrists, type of discharge, veteran status
e. Value systems: whether children are seen as a burden or a joy; whether work is seen as a necessary evil, an avoidable chore, or an opportunity; current attitude about religion; belief in heaven and hell

II. **Mental status**

Sum total of the examiner's observations and impressions derived from the initial interview

A. *Appearance*
1. Personal identification: May include a brief nontechnical description of the patient's appearance and behavior as a novelist might write it. Attitude toward examiner can be described here: cooperative, attentive, interested, frank, seductive, defensive, hostile, playful, ingratiating, evasive, guarded
2. Behavior and psychomotor activity: gait, mannerisms, tics, gestures, twitches, stereotypes, picking, touching examiner, echopraxia, clumsy, agile, limp, rigid, retarded, hyperactive, agitated, combative, waxy
3. General description: posture, bearing, clothes, grooming, hair, nails; healthy, sickly, angry, frightened, apathetic, perplexed, contemptuous, ill at ease, poised, old looking, young looking, effeminate, masculine; signs of anxiety—moist hands, perspiring forehead, restlessness, tense posture, strained voice, wide eyes; shifts in level of anxiety during interview or with particular topic

B. *Speech:* rapid, slow, pressured, hesitant, emotional, monotonous, loud, whispered, slurred, mumbled, stuttering, echolalia, intensity, pitch, ease, spontaneity, productivity, manner, reaction time, vocabulary, prosody

C. *Mood and affect*
1. Mood (a pervasive and sustained emotion that colors the person's perception of the world): how does patient say he or she feels; depth, intensity, duration, and fluctuations of mood—depressed, despairing, irritable, anxious, terrified, angry, expansive, euphoric, empty, guilty, awed, futile, self-contemptuous, anhedonic, alexithymic
2. Affect (the outward expression of the patients inner experiences): how examiner evaluates patient's affects—broad, restricted, blunted or flat, shallow, amount and range of expression; difficulty in initiating, sustaining, or terminating an emotional response; is the emotional expression appropriate to the thought content, culture, and setting of the examination; give examples if emotional expression is not appropriate

D. *Thinking and perception*
1. Form of thinking
   a. Productivity: overabundance of ideas, paucity of ideas, flight of ideas, rapid thinking, slow thinking, hesitant thinking; does patient speak spontaneously or only when questions are asked, stream of thought, quotations from patient
   b. Continuity of thought: whether patient's replies really answer questions and are goal directed, relevant, or irrelevant; loose associations; lack of cause-and-effect relationships in patient's explanations; illogical, tangential, circumstantial, rambling, evasive, perseverative statements, blocking or distractibility
   c. Language impairments: impairments that reflect disordered mentation, such as incoherent or incomprehensible speech (word salad), clang associations, neologisms
2. Content of thinking
   a. Preoccupations: about the illness, environmental problems; obsessions, compulsions, phobias; obsessions or plans about suicide, homicide; hypochondriacal symptoms, specific antisocial urges or impulses
3. Thought disturbances
   a. Delusions: content of any delusional system, its organization, the patient's convictions as to its validity, how it affects his or her life; persecutory delusions—isolated or associated with pervasive suspiciousness; mood-congruent or mood-incongruent
   b. Ideas of reference and ideas of influence: how ideas began, their content, and the meaning the patient attributes to them
4. Perceptual disturbances
   a. Hallucinations and illusions: whether patient hears voices or sees visions; content, sensory system involvement, circumstances of the occurrence; hypnagogic or hypnopompic hallucinations; thought broadcasting
   b. Depersonalization and derealization: extreme feelings of detachment from self or from the environment
5. Dreams and fantasies
   a. Dreams: prominent ones, if patient will tell them; nightmares
   b. Fantasies: recurrent, favorite, or unshakable daydreams

E. *Sensorium*
1. Alertness: awareness of environment, attention span, clouding of consciousness, fluctuations in levels of awareness, somnolesence, stupor, lethargy, fugue state, coma
2. Orientation
   a. Time: whether patient identifies the day correctly; or approximate date, time of day; if in a hospital, knows how long he or she has been there; behaves as though oriented to the present
   b. Place: whether patient knows where he or she is
   c. Person: whether patient knows who the examiner is and the roles or names of the persons with whom in contact
3. Concentration and calculation: subtracting 7 from 100 and keep subtracting 7s; if patient cannot subtract 7s, can easier tasks be accomplished—4 × 9; 5 × 4; how many nickels are in $1.35; whether anxiety or some disturbance of mood or concentration seems to be responsible for difficulty
4. Memory: impairment, efforts made to cope with impairment—denial, confabulation, catastrophic reaction, circumstantiality used to conceal deficit; whether the process of registration, retention, or recollection of material is involved
   a. Remote memory: childhood data, important events known to have occurred when the patient was younger or free of illness, personal matters, neutral material
   b. Recent past memory: past few months
   c. Recent memory: past few days, what did patient do yesterday, the day before, have for breakfast, lunch, dinner
   d. Immediate retention and recall: ability to repeat six figures after examiner dictates them—first forward, then backward, then after a few minutes' interruption; other test questions; did same questions, if repeated, call forth different answers at different times
   e. Effect of defect on patient: mechanisms patient has developed to cope with defect
5. Fund of knowledge: level of formal education and self-education; estimate of the patient's intellectual capability and whether capable of functioning at the level of his or her basic endowment; counting, calculation, general knowledge; questions should have relevance to the patient's educational and cultural background
6. Abstract thinking: disturbances in concept formation; manner in which the patient conceptualizes or handles his or her ideas; similarities (e.g., between apples and pears), differences, absurdities; meanings of simple proverbs, such as, “A rolling stone gathers no moss,”; answers may be concrete (giving specific examples to illustrate the meaning) or overly abstract (giving generalized explanation); appropriateness of answers
7. Insight: degree of personal awareness and understanding of illness
   a. Complete denial of illness
   b. Slight awareness of being sick and needing help but denying it at the same time
   c. Awareness of being sick but blaming it on others, on external factors, on medical or unknown organic factors
   d. Intellectual insight: admission of illness and recognition that symptoms or failures in social adjustment are due to irrational feelings or disturbances, without applying that knowledge to future experiences
   e. True emotional insight: emotional awareness of the motives and feelings within, of the underlying meaning of symptoms; does the awareness lead to changes in personality and future behavior; openness to new ideas and concepts about self and the important people in his or her life
8. Judgment
   a. Social judgment: subtle manifestations of behavior that are harmful to the patient and contrary to acceptable behavior in the culture; does the patient understand the likely outcome of personal behavior and is patient influenced by that understanding; examples of impairment
   b. Test judgment: patient's prediction of what he or she would do in imaginary situations; for instance, what patient would do with a stamped, addressed letter found in the street

III. **Further diagnostic studies**

A. Physical examination

B. Neurological examination
C. Additional psychiatric diagnostic interviews
D. Interviews with family members, friends, or neighbors by a social worker
E. Psychological, neurological, or laboratory tests as indicated: electroencephalogram, computed tomography scan, magnetic resonance imaging, tests of other medical conditions, reading comprehension and writing tests, test for aphasia, projective or objective psychological tests, dexamethasone-suppression test, 24-hour urine test for heavy metal intoxication, urine screen for drugs of abuse

IV. **Summary of findings**

Mental symptoms, medical and laboratory findings, and psychological and neurological test results, if available, are summarized. Include medications patient has been taking, dosage, duration.

Clarity of thinking is reflected in clarity of writing. When summarizing the mental status, for example, the phrase “Patient denies hallucinations and delusions” is not as precise as “Patient denies hearing voices or thinking that he is being followed.” The latter indicates the specific question asked and the specific response given. Similarly in the conclusion of the report one would write “Hallucinations and delusions were not elicited.”

V. **Diagnosis**

Diagnostic classification is made according to DSM-IV, which uses a multiaxial classification scheme consisting of five axes, each of which should be covered in the diagnosis

Axis I: clinical syndromes (e.g., mood disorders, schizophrenia, generalized anxiety disorder) and other conditions that may be a focus of clinical attention

Axis II: personality disorders, mental retardation, and defense mechanisms

Axis III: any general medical conditions (e.g., epilepsy, cardiovascular disease, endocrine disorders)

Axis IV: psychosocial and environmental problems (e.g., divorce, injury, death of a loved one) relevant to the illness

Axis V: global assessment of functioning exhibited by the patient during the interview (e.g., social, occupational, and psychological functioning); a rating scale with a continuum from 100 (superior functioning) to 1 (grossly impaired functioning) is used

VI. **Prognosis**

Opinion about the probable future course, extent, and outcome of the disorder; good and bad prognostic factors; specific goals of therapy

VII. **Psychodynamic formulation**

Causes of the patient's psychodynamic breakdown—influences in the patient's life that contributed to present disorder, environmental, genetic, and personality factors relevant to determining patient's symptoms; primary and secondary gains; outline of the major defense mechanism used by the patient

VIII. **Comprehensive treatment plan**

Modalities of treatment recommended, role of medication, inpatient or outpatient treatment, frequency of sessions, probable duration of therapy; type of psychotherapy; individual, group, or family therapy; symptoms or problems to be treated. Initially, treatment must be directed toward any life-threatening situations such as suicidal risk or risk of danger to others, which require psychiatric hospitalization. Danger to self or others is an acceptable reason (both legally and medically) for involuntary hospitalization. In the absence of the need for confinement, a variety of outpatient treatment alternatives are available: day hospitals, supervised residences, outpatient psychotherapy or pharmacotherapy among others. In some cases, treatment planning must attend to vocational and psychosocial skills training and even legal or forensic issues. Comprehensive treatment planning requires a therapeutic team approach using the skills of psychologists, social workers, nurses, activity and occupational therapists, and a variety of other mental health professionals, with referral to self-help groups (e.g., Alcoholics Anonymous [AA]) if needed. If either the patient or family members are unwilling to accept the recommendations of treatment and the clinician thinks that the refusal of the recommendations may have serious consequences, the patient, parent, or guardian should sign a statement to the effect that the recommended treatment was refused.

MEDICAL RECORD

The psychiatric report is a part of the medical record; however, the medical record is more than the psychiatric report. It is a narrative that documents all events that occur during the course of treatment, most often referring to the patient's stay in the hospital. Progress notes record every interaction between doctor and patient, reports of all special studies including laboratory tests, and prescriptions and orders for all medications. Nurses' notes help describe the patient's course: Is the patient beginning to respond to treatment? Are there times during the day or night when symptoms get worse or remit? Are there adverse effects or complaints by the patient about prescribed medication? Are there signs of agitation, violence, or mention of suicide? If the patient requires restraints or seclusion are the proper supervisory procedures being followed? Taken as a whole, the medical record tells what happened to the patient since first making contact with the health care system. It concludes with a discharge summary that provides a concise overview of the patient's course with recommendations for future treatment, if necessary. Evidence of contact with a referral agency should be documented in the medical record to establish continuity of care if further intervention is necessary.

Use of the Record The medical record is not only used by physicians but is also used by regulatory agencies and managed care companies to determine length of stay, quality of care, and reimbursement to doctors and hospitals. In theory, the inpatient medical record is accessible to authorized persons only and is safeguarded for confidentiality. In practice however, absolute confidentiality cannot be guaranteed. Guidelines for what material needs to be incorporated into the medical record are given in Table 7.2-1.

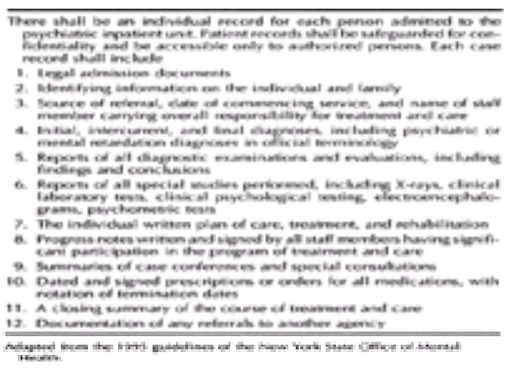

There shall be an individual record for each person admitted to the psychiatric inpatient unit. Patient records shall be safeguarded for confidentiality and be accessible only to authorized persons. Each case record shall include

1. Legal admission documents
2. Identifying information on the individual and family
3. Source of referral, date of commencing service, and name of staff member carrying overall responsibility for treatment and care
4. Initial, intercurrent, and final diagnoses, including psychiatric or mental retardation diagnoses in official terminology
5. Reports of all diagnostic examinations and evaluations, including findings and conclusions
6. Reports of all special studies performed, including X-rays, clinical laboratory tests, clinical psychological testing, electroencephalograms, psychometric tests
7. The individual written plan of care, treatment, and rehabilitation
8. Progress notes written and signed by all staff members having significant participation in the program of treatment and care
9. Summaries of case conferences and special consultations
10. Dated and signed prescriptions or orders for all medications, with notation of termination dates
11. A closing summary of the course of treatment and care
12. Documentation of any referrals to another agency

Adapted from the 1981 guidelines of the New York State Office of Mental Health.

Table 7.2-1 Medical Record

The medical record is also crucial in malpractice litigation. Robert I. Simon summarized the liability issues as follows:

> Properly kept medical records can be the psychiatrist's best ally in malpractice litigation. If no record is kept, numerous questions will be raised regarding the psychiatrist's competence and credibility. This failure to keep medical records may also violate state statutes or licensing provisions. Failure to keep medical records may arise out of the psychiatrist's concern that patient treatment information be totally protected. Although this is an admirable ideal, in real life the psychiatrist may be legally compelled under certain circumstances to testify directly about confidential treatment matters.

Outpatient records are also subject to scrutiny by third parties under certain circumstances, and psychiatrists in private practice are under the same obligation to

maintain a record of the patient in treatment as the hospital psychiatrist. Table 7.2-2 lists documentation issues of concern to third-party payers.

Table 7.2-2 Documentation Issues*

Personal Notes and Observations According to laws relating to access to medical records, some jurisdictions (such as in the Public Health Law of New York State) have a provision that applies to a physician's personal notes and observations. Personal notes are defined as “a practitioner's speculations, impressions (other than tentative or actual diagnosis) and reminders.” The data are maintained only by the clinician and cannot be disclosed to any other person. Psychiatrists concerned about material that may prove damaging or otherwise hurtful to the patient if released to a third party may consider using this provision to maintain doctor-patient confidentiality.

Patient Access to Records Patients have a legal right to access their medical records. This right represents society's belief that the responsibility for medical care has become a collaborative process between doctor and patient. Patients see many different physicians, and they can be more effective historians and coordinators of their own care with such information.

Psychiatrists must be careful in releasing their records to the patient if, in their judgment, the patient can be harmed emotionally as a result. Under these circumstances, the psychiatrist may choose to prepare a summary of the patient's course of treatment, holding back material that might be hurtful—especially if it were to get into the hands of third parties. In malpractice cases, however, it may not be possible to do so.

> An 18-year-old man claimed that he was refused admission to college because he indicated on his application that he had once seen a psychiatrist. In fact, that was not the reason admission was denied; nevertheless, he sued the school on those grounds. At trial it was revealed that he had been hospitalized on several occasions for severe psychotic episodes, some of which were associated with suicide attempts. The defendant school psychiatrist suggested to the judge that it might be harmful to the patient if he were present when excerpts of his psychiatric record were read aloud in the courtroom. The judge agreed, but the man's attorneys objected, and he was permitted to remain. Extensive narratives of his life and complex diagnostic evaluations of psychosis by various psychiatrists from different institutions were read. He became visibly upset during this process and began to attack his attorneys verbally. The court proceeding had to be interrupted.

Problem-Oriented Medical Record In 1969 Lawrence L. Weed published *Medical Records, Medical Education and Patient Care* in which he described the problem-oriented medical record. The problem-oriented medical record lists all health problems discovered in the initial workup. Problem areas are added to and corrected over time. Active problems are listed in one column, and as they are resolved, they are transferred to an inactive column. Progress notes are dated, titled, and numbered according to the problem list. As a final check, the record is audited for thoroughness, reliability, efficiency, and standards of treatment and outcome.

Problem-Oriented Record in Medical Education Many medical schools are using educational techniques based upon the problem-oriented medical record. Teaching and testing organizations such as the United States Medical Licensing Examination (USMLE) are relying on the ability of a student to deal with a problem-oriented medical record as an evaluation tool. In such exercises, students are provided with general patient information, which may include a summary of the physical examination and positive elements in the psychiatric report. Students must delineate the problem areas that need attention, determine preventive or treatment options for each problem, and understand the patient from biological, social, and psychological points of view.

Although the problem-oriented medical record is not the standard format used by psychiatrists, it has influenced how patients are viewed. For example, in DSM-IV, specific psychosocial and environmental problems that may affect diagnosis of mental disorders are recorded on Axis IV. The problems are divided into nine categories that can each affect the person adversely: (1) problems with primary support group (e.g., death of family member), (2) problems related to the social environment (e.g., absence of friends), (3) educational problems (e.g., discord with teachers or classmates), (4) occupational problems (e.g., stress at work), (5) housing problems (e.g., unsafe neighborhood), (6) economic problems (e.g., excess debt), (7) problems with access to health care services (e.g. no health insurance), (8) problems with the legal system (e.g., litigation), and (9) other problems (e.g., floods, earthquakes).

> A 55-year-old married man complains of being fearful that he will be forced to resign from an administrative job because his company is being downsized. He complains of anxiety, insomnia, and irritability and gives no history of previous dysphoric states. After a thorough evaluation it becomes apparent that the proximate cause of his symptoms is the probability of losing his job. Symptom removal is relatively simple with the aid of anxiolytic or hypnotic drugs; however, a more comprehensive approach is required. The therapist has to attend to the occupational problem that might include having the patient confront his superiors, applying for reassignment to another area within the company, seeking other employment, evaluating his assets for early retirement, job retraining, or other approaches directed toward the vocational crisis.

EVALUATION AND MANAGEMENT (E/M) SERVICE GUIDELINES

In addition to the classical psychiatric report described above, physicians (including psychiatrists) are being asked to follow documentation guidelines for evaluation and management services (E/M) developed jointly by the American Medical Association (AMA) and the Health Care Financing Administration (HCFA). The E/M guidelines were developed to help doctors record and document one of the more than 2000 diagnostic codes listed in International Classification of Diseases, 9th Revision, Clinical Modifications (ICD-9-CM) and DSM-IV. In 1989 physicians and other health care professionals were required to report ICD codes when billing Medicare, Medicaid, and private insurance carriers for reimbursement to doctors, hospitals, and patients.

Reason for E/M Guidelines Guidelines were developed to correct purported deficiencies in the reporting of health care services by hospitals and physicians. Audits in 1995 and 1996 by the Health Care Financing Administration (HCFA) alleged that over 23 billion dollars was overpaid to teaching hospitals and related facilities, which amounted to 14 percent of Medicare Part B fee-for-service payments.

Levels of Service There are four levels of E/M service depending on the types of history and type of examination. The types of history and examination that relate to each level are (1) *problem focused*—brief history of present illness, limited examination of the affected body area or organ systems; (2) *expanded problem focused*—brief history of present illness, review of systems pertinent to problem, limited examination of the affected body area or organ system and related organ system; (3) *detailed*—extended history of present illness, extended review of systems, pertinent past family and social history, extended examination of the affected body area and related organ systems; (4) *comprehensive*—extended history of present illness, extended review of systems, complete past family and social history, and a general multisystem examination or complete examination of a single organ system. Table 7.2-3 summarizes the four levels of service and their requirements.

Level of E/M Service	Chief Complaint	History of Present Illness	Review of Systems	Past/Family/Social History	Physical Examination*
1. Problem focused	Yes	Brief	N/A†	N/A	Limited
2. Expanded problem focused	Yes	Brief	Pertinent problem	N/A	Limited
3. Detailed	Yes	Extended	Extended	Pertinent	Extended
4. Comprehensive	Yes	Extended	Complete	Complete	Extended

*For psychiatrists, the mental status examination is equivalent to the physical examination; however, vital signs should be taken, if indicated.
†N/A, not applicable.

Table 7.2-3 Level of Evaluation and Management (E/M) Services

E/M and Psychiatry Psychiatrists are expected to use the E/M guidelines with some modifications. The guidelines assume appropriately that psychiatrists are expert in, and will spend most of their time performing, a psychiatric history and mental status examination. The E/M requirements for psychiatry are listed in Table 7.2-4. Psychiatrists are not expected to perform an extensive physical examination. The physical examination is relevant only as it pertains to the psychiatric problem and may include measurement of blood pressure and pulse, especially if the patient is taking medication. It also includes an assessment of the patient's general appearance including nutritional status, body habitus, presence of gross physical deformities, and attention to grooming. Demented patients, for example, often are unkempt. Depressed patients may exhibit stooped posture and a furrowed brow. Abnormal movements such as tics may reflect dystonic reactions to dopaminergic receptor antagonists such as chlorpromazine (Thorazine). Psychiatrists are not expected to examine organ systems, nor are they expected to perform a neurological examination. The patient should be referred to an internist or specialist (e.g., gynecologist) if special physical examinations are indicated on the basis of the history or review of systems.

- Description of speech including rate, volume, articulation, coherence, and spontaneity with notation of abnormalities (e.g., perseveration, paucity of language)
- Description of thought processes including rate of thoughts, content of thoughts (e.g., logical vs. illogical, tangential), abstract reasoning, and computation
- Description of associations (e.g., loose, tangential, circumstantial, intact)
- Description of abnormal or psychotic thoughts including hallucinations, delusions, preoccupation with violence, homicidal or suicidal ideation, and obsessions
- Description of the patient's judgment (e.g., concerning everyday activities and social situations) and insight (e.g., concerning psychiatric condition)

Complete mental status examination including

- Orientation to time, place, and person
- Recent and remote memory
- Attention span and concentration
- Language (e.g., naming objects, repeating phrases)
- Fund of knowledge (e.g., awareness of current events, past history, vocabulary)
- Mood and affect (e.g., depression, anxiety, agitation, euphoria, lability)

Content and documentation requirements

Level of examination	Perform and document
Problem focused	One to five elements identified by a bullet†
Expanded problem focused	At least six elements identified by a bullet†
Detailed	At least nine elements identified by a bullet†
Comprehensive	Perform all elements identified by a bullet†

† The areas covered in this examination have been established by many practitioners as being the most frequently used and most valuable for beginning medical students that for experienced clinicians who can rule in positive and negative findings without rigidly following a formal outline.
From American Medical Association and Health Care Financing Administration (1995).

Table 7.2-4 Elements of Examination in Psychiatry*

Psychiatrists may be expected to conduct a review of systems that seeks to identify signs or symptoms that the patient may be experiencing. In a brief problem-focused examination this is not required; however, in a comprehensive examination all bodily systems are reviewed. Table 7.2-5 lists a symptom review with examples about which the psychiatrist may inquire.

- Constitutional symptoms: Fever, weight loss, nutritional status
- Eyes: Visual acuity, diplopia
- Ears, nose, mouth, throat: Hearing, tinnitus, tonsillitis, twitching, mouth movements
- Cardiovascular: Dyspnea, palpitations, edema, leg or chest pain
- Respiratory: Hemoptysis, wheezing
- Gastrointestinal: Nausea, vomiting, melena, constipation, diarrhea, use of laxatives
- Genitourinary: Frequency, pain on urination, hematuria, difficulty starting stream, sexual problems
- Musculoskeletal: joint pain, limitation in range of motion, balance
- Integumentary (skin and/or breast): Moles, lumps, rashes
- Neurological: Vertigo, paralysis, paresthesia, syncope
- Endocrine: Changes in weight, nervousness, fatigue, menstrual history
- Hematologic and lymphatic: Bleeding, bruising, nosebleeds
- Allergic and immunological: Asthma, drug reactions, frequency of colds or infections

Table 7.2-5 Review of Systems

Critique of E/M Guidelines Many physicians complain that the E/M guidelines are extremely complex and that proper compliance with documentation takes time away from patient care. One 50-page HCFA publication explaining how to use the guidelines has been criticized as being useful for medical students just learning how to interview patients but unnecessary for experienced physicians who can document positive and negative findings accurately and succinctly. A work group of the American College of Physicians working with the HCFA and the AMA concluded in 1998:

> All of us recognize the need for Medicare carriers to....examine what we do. There has got to be a way to realistically record what you do in the office—but not to do it in the bureaucratic monstrosity whose only aim is documentation.

As a result of such criticism, the HCFA has placed the guidelines on hold; however, some private insurers are currently using them. With some modifications they will be mandated for use eventually in all specialties, including psychiatry.

CASE REPORT

The case report (also known as the case study) has an honored place in the history of psychiatry and in the advancement of medical knowledge. Single case studies are included in most medical and psychiatric journals, and they continue to provide a wealth of data about both common and uncommon conditions. Critics believe that case studies suffer from methodological flaws, since one cannot generalize from a single case study to a population at large, but authors of case reports rarely, if ever, make that claim. Rather, they limit themselves to objective, reliable reporting and the presentation of facts. Any hypothesis offered is usually tentative and conditional upon further testing.

With the rise of pharmacotherapy, case studies have assumed growing importance. The single case is of little value in the development of drugs; large samples studied over long periods of time are necessary to determine the efficacy and safety of a drug; however, the literature is replete with single case studies that signaled adverse drug reactions or new uses for an established drug.

The literature is also replete with reports by physicians and other researchers who used themselves as subjects of experiments. The many reports by doctors of their personal experience with illness are sobering reminders to the rest of the profession of what it means to be a patient. Sigmund Freud developed his theory of dream interpretation on the basis of his own dreams, and Jean Piaget derived the theories of cognition in children from observing his own children. William Beaumont demonstrated the relationship between human gastric functions and various emotional states in his patient Alexis St. Martin who had a gastric fistula that resulted

from a gunshot wound.

Below is a chronology of 86 famous “named” cases in psychiatry that date from 1577 to 1980. These single cases were culled from the literature by David Davis, Professor Emeritus in Psychiatry at the University of Missouri–Columbia School of Medicine. Some cases are of greater interest or importance than others, but all contributed to the slow, steady advance of psychiatric knowledge over the centuries.

1577 *Barbara Kremers*—10-year-old girl alleged to have fasted for 6 months; a possible instance of malingering. She was treated by Johann Weyer and reported in detail.

1622 *William Perry, the Boy of Bilson*—A 13-year-old boy who produced black urine. He eventually admitted to mixing his urine with ink, which illustrates how the concept of simulated disease was confused with witchcraft. Described by Richard Baddeley.

1774 *Fraulein Oesterlin*—Anton Mesmer treated her with magnets and developed the notion that the magnetic effect came from magnetic fluid emanating from his own person, an effect he called animal magnetism.

1775 *Maria-Theresa Paradis*—A blind musician treated by Mesmer whose sight was restored only in his presence then was lost again, presumably related to the effects of secondary gain.

1774–1824 *Katherina Emmerich*—A woman who had visions and displayed the stigmata of Christ's Passion. Described in detail by Clemens Brentano.

1785 *Victor Race*—Treated by Armand-Marie-Jacques de Chastenet, Marquis de Puységur, who magnetized him into an artificial somnambulism in which the patient could diagnose his own disease and describe its course. Puységur believed that cure resulted from the effect of the magnetizer's will rather than from magnetic fluid.

1788 *King George III*—A detailed review of his five episodes of illness documented over 58 years tends to substantiate the view that he suffered from a classic case of porphyria.

1799 *Victor, the Wild Boy of Aveyron*—A feral child discovered in a tree when he was 12 years old. Considered by Pinel to be incurably retarded but treated by Jean Marc Gaspard Itard.

1811 *Wild Peter of Hanover*—Was found wandering alone at age 12. Considered to be a specimen of true “natural man.” Reported by Blumenbach.

1815 *Mary Reynolds*—A person with dual personality known as “La dame de MacNish” said to have been described by John Kearsley Mitchell, the father of Weir Mitchell.

1826 *Friederiche Hauffe*—Also known as “The Seeress of Prevorst,” she was treated by Justinus Kerner with magnetic trances. During one trance she devised a “nerve-tuner” that Kerner then built and used in his treatment.

1832 *Kaspar Hauser*—A “feral” boy who appeared at the gates of Nuremberg in 1828. Described by Anselm Ritter von Feuerbach.

1836–1837 *Estelle L'Hardy*—Known as “La petite ressuscitée,” she was treated by Antoine Despine, Sr. During magnetic treatment, another personality emerged that apparently was the healthy one, which eventually became the dominant personality.

1842–1843 *Gottliebin Dittus*—A woman treated with exorcism by Johann Christoph Blumhardt. Comparisons have been drawn between his method and that used in the psychotherapy of patients with severe schizophrenia.

1843 *Daniel M'Naghten*—Glasgow paranoia patient who was tried for the willful murder of Edward Drummond, the private secretary to Sir Robert Peel. The questions put to the judges by the House of Lords were the basis for the M'Naghten Rules.

1858–1893 *Félida X*.—A patient with dual personality studied by Eugéne Azam and described as “dedoublement de la personnalité” in whom the first personality was not aware of the second personality while the second personality was aware of both.

1868 *Phineas Gage*—The first description of “frontal lobe syndrome,” presented by J. M. Harlow to the Massachusetts Medical Society. During a blasting operation in 1848 a long iron rod was driven through Gage's skull pushing out a cylindrical mass of gray matter from the left hemisphere. He survived for 20 years after the original accident.

1885 *Leonie*—A patient with multiple personality studied by Pierre Janet who found that she could be hypnotized from a distance and that the personality that emerged under hypnosis was a reenactment of previous hypnotic experiments.

1886 *Lucie*—Patient in the first case of cathartic cure that Pierre Janet published. He used hypnosis and automatic writing and described the nature of the rapport.

1888 *Blanche Wittman*—One of Charcot's most renowned subjects, known as “la reine des hysteriques.” She demonstrated Charcot's three stages of hypnosis and was later treated by Jules Janet (the brother of Pierre), at which time another personality emerged under hypnosis.

1889 *Marie*—Patient in Pierre Janet's second case of cathartic cure in which he demonstrated the importance of fixed subconscious ideas and the role they played in illness.

1889 *Emmy Von N*.—Sigmund Freud's first use of cathartic treatment. It was his first attempt to work with Breuer's method, differing only in having the patient recall the initial traumatic event under hypnosis, after which the patient had to suggest that the symptom disappear.

1890 *Achilles*—A patient with demoniacal possession cured by Pierre Janet by unraveling the fixed subconscious ideas.

1891–1893 *Madame D*.—The patient around whom Charcot developed his concept of dynamic amnesia as distinct from organic amnesia. She was also treated by Pierre Janet who encouraged her to talk freely under hypnosis to dissolve her fixed ideas.

1891 *Marcelle*—A patient illustrating psychological analysis and synthesis using hypnosis and automatic writing, described by Pierre Janet.

1891 *Ansel Bourne*—An early multiple personality patient with mutual amnesia for each personality. Bourne was examined by William James and described by Richard Hodgson.

1892 *Lucie R*.—A girl with an olfactory hallucination treated by *Sigmund Freud* by means of catharsis.

1892 *Katherina*—A girl with difficulty breathing, headache, and anxiety attacks related to early sexual trauma, described by Freud.

1892 *Elizabeth Von R*.—The patient who suggested Freud's technique of “free association,” which is described in the report.

1894 *Hélene Smith*—A medium whose real name was Catherine-Elise Muller. She was described by Theodore Flournoy, who used her to demonstrate the importance of forgotten childhood memories and the reversion to various stages of childhood in the patient's fantasies and suggested that these expressed secret wishes.

1894 *Justine*—A celebrated patient with a morbid fear of cholera cured by Janet with his method of analysis of fixed ideas.

1895 *Mr. N*.—A patient with hysterical amnesia treated under hypnosis by Auguste Forel.

1895 *Anna O*.—The patient, whose real name was Bertha Pappenheim, treated by Josef Breuer for a variety of hysterical symptoms including dual personality in 1881. This case convinced Freud of the power of unconscious memories and suppressed affects in hysteria. Reputedly her production of pseudocyesis was responsible for Breuer's lack of further exploration of the unconscious with Freud.

1895 *Irma*—The patient about whom Freud had his “dream of Irma's injection.” This is the first dream that he analyzed fully with his new technique of associations and recognized as a wish fulfillment.

1896–1904 *Madeleine*—A patient followed by Janet for 25 years. She walked on tiptoe, was subject to hysterical delusions, and showed the stigmata of the Passion. Based on her material Janet developed his theory of emotions and concepts of religious psychology.

1898 *Miss Beauchamp*—A patient whose eventual emergence of four personalities may have been unconsciously suggested by the investigator, Morton Prince. During treatment he succeeded in synthesizing one personality from them.

1899 *Hélène Preiswerk*—Carl Gustav Jung's cousin, a medium on whom he made observations that became the subject of his dissertation in 1902. During semisomnambulistic states a personality named Ivenes emerged, which Jung considered to be the adult personality of the medium.

1900 *Meb*—A patient with hysterical hallucinations with mystical and erotic themes treated by Janet with hypnosis, somewhat similar to the case of Achilles.

1902 *Irène*—Another patient with hysterical disturbance, somnambulistic crisis, hallucinations, and amnesia. During her treatment Janet found that hypnotic treatment had to be supplemented with mental stimulation and reeducation.

1903 *Nadia*—A girl obsessed by the fear of becoming fat (related to fear of rejection) treated by Janet and showing some resemblance to Binswanger's Ellen West.

1905 *Dora*—Patient with a complex hysterical cough treated by Sigmund Freud in 1900 using dream interpretation. The intricacies of the interpersonal relationships are described as well as mental mechanisms, including the therapeutic importance of transference.

1906 *Miss Frank Miller*—A patient prone to hypnagogic experiences who was reported by Flournoy. Jung later based his interpretations of mythology and history of religion on this account. By applying these ideas to his own fantasies he developed his notions of anima and self.

1908 *Little Hans*—A 4-year-old boy with horse phobia who was treated by Freud through the boy's father. This is the first case of child analysis and the first control analysis. It elucidated the mechanism of displacement and contributed to libido theory. Hans was later reputed to have become the famous opera impresario Rudolph Bing.

1909 *The Rat Man*—A patient with obsessional neurosis described by Freud. The patient was preoccupied with the idea of rats in the East that bored into his anus as punishment.

1909 *Hersilie Rouy*—The subject of two autobiographies depicting delusional claims of royal birth. Described by Paul Serieux and Joseph Capgras.

1911 *Daniel Paul Schreber*—Freud attempted a pathography based on Schreber's description of his delusion and on legal reports about him, and developed a theory of repressed homosexuality to account for Schreber's paranoid illness.

1916 *Doris Fisher*—A patient with five personalities described by Walter Franklin Prince.

1918 *The Wolf-Man*—A patient who suffered from lack of will-power treated by Freud and later by Ruth Mack Brunswick. He had a terrifying dream of wolves at age 3½. This case helped Freud understand countertransference.

1920 *Albert B*.—A 9-month-old child who was conditioned by J. B. Watson and R. Raynor to fear a rat.

1924 *Erna*—A 6-year-old child analyzed by use of play therapy. Reported by Melanie Klein.

1925 *Gregor Adamsberger*—A patient used by Theodore Reik to elucidate the unconscious facts traceable in the psychopathology of judicial error.

1925 *Elena F*.—A patient with dual personalities, one French, one Italian, which G. E. Morselli attempted to fuse. As a result of treatment, her psychotic symptoms disappeared. Described by Henri F. Ellenberger as perhaps the most remarkable case of multiple personality ever published.

1925 *Peter*—A boy with fear of a playroom rabbit who J. B. Watson treated by reconditioning.

1927 *Amala and Kamala*—Two feral children raised together in a den of wolves. First described by P. C. Squires.

1933 *Ikara*—A patient who reputedly described memories from a previous life; reported by Max Bircher-Benner.

1933 *Charles Poultney*—A patient with mutually amnestic multiple personality described by S. I. Franz.

1935 *Bill W*.—An autobiography of the founder of AA.

1941 *Isabelle*—A child who learned to speak after 6½ years of silence. Reported by Marie Mason.

1942 *Herbert Bryan*—Subject of a complete transcript of a phonographically recorded counseling process with additional therapist's comments, as described by Carl R. Rogers.

1942 *Clare*—A patient with compulsive modesty, dependence, and the compulsive need to force others to recognize her superiority. The analysis by Karen Horney illustrates her formulation of the social and cultural etiology of neurosis.

1943 *Tom*—A patient with gastric fistula on whom observations were made concerning the effect of normal situational stresses and various emotional states on his gastric function by Steward Wolf and Harold Wolff.

1944 *Ellen West*—A girl with a fear of becoming fat, treated by Ludwig Binswanger who applied the principles of existential analysis to clinical psychiatry and psychopathology.

1944 *Harold*—A criminal psychopath hypnoanalyzed by Robert M. Lindner.

1951 *Renée*—Patient is perhaps the best documented study of a single patient. The therapist, M. A. Sechahaye, entered the psychotic world of the patient, accepted

the symbolism involved, and attempted to materialize the patient's fundamental needs at a symbolic level, the so-called symbolic realization.

1953 *Marisa*—A patient with dual personality, reported by G. E. Morselli, in which electroencephalograms were found to differ in the two personalities.

1954 *Monte Durham*—Defendant in case in which Judge D. L. Bazelon in the Federal Appellate Court in Washington, D.C., ruled that a man should not be held responsible for an otherwise criminal act when his behavior was the "product of mental disease or deficiency."

1956 *Monica*—An infant girl with congenital atresia of the esophagus in whom a gastric fistula was established in the fourth day of life. George L. Engel and his colleagues made detailed observations of behavior and gastric secretion were made until she reached age 22 months, when the fistula was closed. She has been followed into adult life, and observations of her interaction with her own daughter have been described.

1957 *Eve*—A patient with multiple personality whose real name was Chris Sizemore, reported by Corbett Thigpen and Hervey Cleckley.

1960 *S. V. Shereshevski*—A man with hypermnesia studied by A. R. Luria for over 30 years. He could effortlessly memorize sequences or tables containing 50, 100, and more figures presented to him aurally or visually, some after an interval of 20 years.

1964 *Catherine Lake*—A patient was committed to St. Elizabeth's Hospital in 1962 with chronic brain syndrome due to arteriosclerosis with a psychotic reaction. She appealed on the grounds that a suitable alternative to "total confinement" in a mental hospital was warranted. This was the first major judicial concern over the treatment and rights of hospitalized patients.

1966 *Charles Rouse*—Patient who was committed to St. Elizabeth's involuntarily as criminally insane after being acquitted of an offense by reason of insanity but was not receiving treatment. This is a case of treatment and rights of hospitalized patients. The decision stated that "the purpose of involuntary hospitalization is treatment not punishment."

1969 *Rose R*.—A 63-year-old woman (whose real name was Sylvia) in whom forced reminiscence was induced by levodopa (Larodopa). She was the earliest patient described in a series treated by Oliver Sachs and originally published as "Incontinent Nostalgia induced by L-Dopa in 1970." (This case and others were depicted in a book *Awakenings* in 1973; a documentary TV film in 1973; three plays including A *Kind of Alaska* by Harold Pinter in 1982, with the character Deborah based on Rose R; a radio drama on CBC 1987; and a movie in 1987.)

1971 *Mary Barnes*—A patient treated by R. D. Laing who illustrates his acceptance of regression and rebirth in psychosis and his view of psychosis as a necessary stage in the achievement of health.

1971 *Sandra*—A 6-month-old infant whose behavior modification used lemon juice to control a life-threatening rumination.

1971 *Ricky Wyatt*—A mentally retarded patient at Alabama's Bryce Hospital whose guardian challenged the adequacy of his treatment, raising the issue of whether civilly or criminally committed mentally ill and mentally retarded patients have the right to adequate individual treatment. The class action was brought by patients and staff.

1972 *Archie W. Brawner*—Defendant who drank wine and went to a party. When several fights broke out, he left, returned with a gun, and shot and killed Billy Ford. He was diagnosed variously as having epileptic personality disorder or explosive personality. In his appeal against a conviction of second-degree murder the court decided to discard the Durham Rule and adopt instead the test recommended in 1962 by the American Law Institute in its model penal code.

1973 *Sybil*—Patient with dissociative identity disorder (multiple personality disorder) treated by Cornelia Wilbur. She exhibited 16 "personalities" and alleged that she had suffered sexual child abuse by her mother. Later researchers have questioned whether the therapist interpreted the patient's state correctly. Book of the same name was written by Flora Rheta Shreiber, and the story was also made into a movie.

1973 *Mercedes*—A black woman whose treatment deals with themes of social injustice, racism, and the ability to bear children. Example of Rollo May's existential therapy.

1974 *Martha*—An illustration of rational emotive therapy by Albert Ellis.

1975 *Kenneth Donaldson*—Man who had been civilly committed because of schizophrenia and confined at the Florida State Hospital for 14 years. He sued state hospital psychiatrists for money damages pursuant to the Civil Rights Act (i.e., deprivation of liberty), and appealed to the Supreme Court—right to liberty had been violated.

1975 *Rubie Rogers*—A chronic schizophrenic 39-year-old woman treated with haloperidol (Haldol) and one of seven patients at Boston State Hospital who filed a civil rights action against the superintendent and psychiatrists. After 74 days of trial and argument, an 8000-page transcript covering testimony of 50 witnesses, and a further 2000 pages of posttrial briefs, this case established the absolute right of patients to refuse treatment, although after a judicial finding of incompetence a guardian may authorize treatment.

1976 *Prosenjit Poddar*—Originally an untouchable from India, a student, and a voluntary outpatient at the Mental Health Clinic at the University of California who told his therapist he intended to kill a student, Tatiana Tarasoff. The therapist notified campus police that he was dangerous and should be committed under a 72-hour emergency observation law. However, the therapist's supervisor vetoed the recommendation. Two months later, he killed Tarasoff. Her parents sued the University of California for negligence. California Supreme Court ruled that a physician or psychotherapist who believes a patient may injure or kill someone must warn the potential victim, victim's relatives or friends, or authorities (known as the Tarasoff I decision—named for the victim). In 1982 the California Supreme Court issued a second ruling in the case of *Tarasoff v Regents of University of California* and broadened the duty to warn to include the duty to protect—known as Tarasoff II.

1977 *Nadia*—An autistic child, born in 1967 to a family of Ukraine emigres, who exhibited extraordinary drawing ability from age 3½ to 10.

1979 *John Rennie*—An involuntary patient with a bipolar illness at a New Jersey state hospital who objected to the forced administration of medications. This case concerning the right to refuse treatment established the use of an appeal process to determine a patient's dangerousness and competency and whether less restrictive treatment existed.

1980 *John W. Hinckley, Jr*.—A person with paranoid schizophrenia who shot President Ronald Reagan and Press Secretary James Brady. Hinckley was found not guilty by reason of insanity in 1982. The court testimony spawned an outcry about the unreliability of psychiatric diagnoses, which contributed to the refining of DSM criteria. This case also led to the Brady Bill, gun-control legislation (requiring a 5-day waiting period to cut down on the sale of handguns to felons and fugitives) that became law in 1993.

1980 *Michelle*—A Canadian housewife, Michelle Smith, treated by Lawrence Pazder, a psychiatrist who helped her "recover" memories of torture by a satanic cult at age 5. Their book *Michelle Remembers* set off the satanic-ritual-abuse craze of the ensuing decade.

See reference for D.A. Davis for citations for the cases above.

ETHICAL ISSUES

Psychiatrists continually make judgments about what is or is not appropriate material to include in the psychiatric report, the medical record, the case report, and other written communications about the patient. Such judgments often involve ethical issues. In a case report, for example, the patient should not be identifiable, a position made clear in the American Psychiatric Association's *Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry*, which states that published case reports must be suitably disguised to safeguard patient confidentiality without altering material to provide a less-than-complete portrayal of the patient's actual condition. In some instances obtaining a written release from the patient that allows the psychiatrist to publish the case may also be advisable, even if the patient is

appropriately disguised.

Psychiatrists sometimes include material in the medical record that is specifically directed toward warding off future culpability if liability issues are ever raised. The following vignette illustrates one such instance.

> A psychiatrist noted in the chart that he discussed with the patient the important side effects of a particular medication he was about to prescribe. In fact, he only discussed the two adverse effects most commonly observed. He chose not to discuss others because he was concerned that the patient would develop them via suggestion and wrote that in the chart. When the patient developed an adverse effect (even though minor and not life threatening), he sued the psychiatrist for not advising him of that possibility. The case was settled in the psychiatrist's favor on the basis of the note in the medical record.

Psychiatrists in the military face unique ethical problems because confidentiality does not exist under the military code of conduct.

> A 19-year-old white single male, new to military service, presented with a history of periodic episodes of anxiety when taking showers in groups with other men. He identified himself as gay and recognized that his anxiety was related to his fear of acting out his sexual impulses, thus risking court-martial and dishonorable discharge should he ever be discovered. The psychiatrist was on the horns of a dilemma: whether to report the soldier to his commanding officer (as he was obliged to do under the military code) or to protect the soldier from acting on his impulses that would place him in danger (in keeping with the medical ethic to do no harm). After discussing various options, he and the patient agreed on the latter option. A diagnosis of anxiety disorder was made, which allowed the patient to receive an honorable discharge on medical grounds, based on a recognized psychiatric disorder. No record of his homosexual orientation was made.

With the advent of managed care and the need to send periodic progress reports and documentation of signs and symptoms to third-party reviewers to pay for treatment, some psychiatrists may diminish or exaggerate symptomatology. The following case report and discussion illustrates the ethical difficulties psychiatrists face in dealing with managed care.

> Mrs. P. admitted herself to the hospital because she was afraid she might kill herself. She was suffering from a major depressive episode, but she improved markedly during the first weeks on Dr. A.'s ward. Although Dr. A. believed that Mrs. P. was no longer suicidal, he thought she would benefit greatly from continued hospitalization. Because he knew that Mrs. P. could not afford to pay for hospitalization and that the insurance company would pay only if the patient was suicidally depressed, he decided not to document Mrs. P.'s improvement. He noted in the chart that "the patient continues to have a risk of suicide."
>
> This case illustrates one of the difficulties posed by cost constraints and managed care. Dr. A. responds to this difficulty by tailoring the chart. Does he engage in a form of deception? Yes, he intentionally misleads by what he writes and what he omits writing in the chart. Although what he writes is true in some literal sense, his statement is misleading in the context of treatment. Mrs. P. is not suicidally depressed in the way she was.
>
> What Dr. A. omits from the chart is also deceptive. Whether a particular omission is deceptive depends, in part, on the roles and expectations of the people involved. Not telling your colleague that you dislike his tie is not a deception. It is simply tact, unless your role or relationship involves the expectation that you offer your candid opinion. Dr. A.'s case is different. His professional role is to document the patient's course, and the expectation is that he will note any significant improvement. Thus his failure to document Mrs. P.'s progress accurately is a kind of deception.
>
> The second and more difficult question is whether deception is justified in this instance. The answer to that question depends on the reasons for the deception, the reasons against it, and the alternatives available.
>
> The reasons for this deception are obvious. Dr. A.'s aim and primary obligation is to help the patient. He believes that Mrs. P. would benefit greatly from continued hospitalization that she cannot afford. He may also believe that it is unfair for the insurance company to refuse to pay for inpatient treatment of nonsuicidal depression and that his deception rectifies that unfair practice.
>
> There are also important reasons against this deception. The first reason concerns honesty and social trust. It is a good thing if people can rely on what others say and write. Without some honesty and trust, many social exchanges and practices would be impossible. Deception, even for beneficent purposes, has real potential to damage social trust. A risk exists that deception may damage people's trust in the profession of psychiatry and even patients' trust in their psychiatrists. Damage to trust may, in turn, compromise treatment.
>
> The second reason concerns future medical treatment. If Mrs. P. seeks medical treatment in the future, the physicians who attend her will read the misleading notes. If they believe that the notes are an accurate account of the previous treatment, they may suggest an inappropriate treatment for the present problem. Even if they have doubts about the accuracy of the notes in her chart, they are deprived of an accurate history and report. In either case, the prior deception can hinder treatment.
>
> The third reason concerns obligations and coverage policies. Dr. A. seems to ignore the obligation he has to the population that is covered by the insurance policy. He shifts a burden onto this population by forcing the insurance company to pay for treatment it did not agree to cover. Perhaps the insurance company should pay for inpatient treatment in cases like Mrs. P.'s; perhaps its policies are unreasonable and unfair. But Dr. A.'s deception does not challenge the insurance company and pressure it to change its policy; nor does his deception encourage patients and their families to contest the company's policies. The use of deception simply circumvents, in an ad hoc way, a policy that should be challenged and discussed.
>
> Dr. A. also seems to ignore his obligation to future patients. By introducing an inaccuracy into the chart, he compromises the value of medical records research. His deception works, in a small way, to deprive future patients of the benefit of research that relies on medical records.
>
> Whether the deception is justified depends not only on the weight of the reasons for and against the deception but also on the available alternatives. One alternative is to tailor the chart. Another alternative is to describe Mrs. P.'s response accurately and discharge her to outpatient care. But a third alternative exists. Dr. A. can accurately document the patient's course and recommend continued hospitalization. He can petition the insurance company for coverage. If the insurance company decides not to approve further inpatient care for the patient, Dr. A. can appeal that decision. This alternative is more time consuming, and there is no guarantee it will succeed, but it avoids all the problems associated with the use of deception. (Reprinted with permission from Dwyer J, Shih A: The ethics of tailoring the patient's chart. Psychiatr Serv *49:*1309, 1998.)

FUTURE DIRECTIONS

The psychiatric report, which includes the psychiatric history and mental status examination, has remained fairly stable over the years, but diagnostic terms have undergone change and will continue to do so. In part, this represents true advances in our understanding of psychiatric illness, but as Karl Menninger stated: "Defining new syndromes and reordering them is always an avocation—a veritable addiction—of psychiatrists."

For example, in 1952 the American Psychiatric Association's Committee on Nomenclature and Statistics published the first edition of DSM (DSM-I). Four editions have been published since then, the last being DSM-IV in 1994. Many psychiatrists have been critical of the many versions of DSM that have appeared, and some such as Mark Zimmerman believe that any new revisions to DSM should follow a "restrained, sober and deliberate pace." In their book *The Selling of DSM*, Stuart Kirk and Herb Kutchins supported that position. They stated:

First, none of the revisions are stimulated by clinical practitioners demanding a new classification system. Second, the writing of each revision has become more time consuming, elaborate, and politically complex. Third, the numerous changes made with each revision are justified explicity or implicity as improving the scientific credibility of the classification system, although the scientific basis for them is often questionable. Finally, the process of revision inevitably begins by attacking the current system, even when it has only recently been adopted, and ends by claiming the superiority of the new one.

With others, the author believes that the publication of the various editions of DSM has been one of the most important developments in psychiatry in the past 50

years. The *Comprehensive Textbook of Psychiatry* has always relied on DSM as the template for deriving the table of contents and the organizing psychopathology. At the same time, contributors to this textbook are encouraged to critique DSM and to suggest changes as they see fit. The reader will find many such critiques throughout this book.

The issues of documentation and access to medical records will receive much more study in future years. Legislative patient privacy bills are before Congress as of this writing to protect and limit third-party access to medical records. In spite of this, however, the trend toward less patient privacy continues. Proposed legislation would give each person in America a unique medical identifier that could encode on microchip one's medical record from the cradle to the grave. Should this or similar legislation be enacted into law, medical privacy will continue to erode.

Finally, critics of current documentation processes claim that filling out forms developed by governmental agencies—more necessary for fiscal than for medical needs—takes time away from direct patient care. In the future, society will have to determine how far the process of documentation must proceed before the need to archive material unrelated to direct patient care interferes with the care of the patient.

SUGGESTED CROSS-REFERENCES

Psychiatric rating scales appear in Section 7.8 on psychiatric rating scales and Section 9.1 on classification of mental disorders. A detailed discussion of DSM-IV's multiaxial system appears in Section 9.1 on the classification of mental disorders. The psychiatric interview, history, and mental status examination are discussed in Section 7.2. Medical assessment and laboratory testing are discussed in Section 7.7. Computer-based testing of the psychiatric patient is covered in Section 7.9.

SECTION REFERENCES

Callis R, Polmantier PC, Roeber EC: *A Casebook of Counselling*. Appleton-Century-Crofts, New York, 1955.

Coleman M, Gillberg C: *The Schizophrenias; A Biological Approach to the Schizophrenia Spectrum Disorders*. Springer, New York, 1996.

*Davis DA: A chronology of seventy famous “named” cases in psychiatry. J Oper Psychiatry *5:*17, 1974.

*Dwyer J, Shih A: The ethics of tailoring the patient's chart. Psychiatr Serv *49:*1309, 1998.

Evans J: *Three Men*. Knopf, New York, 1954.

*Evidence-based medicine on the wards. Editorial. ACP Journal Club *130:*15, 1999.

*Ginzberg E: The uncertain future of managed care. N Eng J Med *340:*144, 1999.

Greist H: Computer-based assessment of patients. J. Clin Pharmacol *18:*359, 1998.

Lewis NDC: *Outline for Psychiatric Exam*. New York, State Department of Mental Hygiene, Albany, NY, 1943.

*Mark D: Effects of physician awareness of symptom-related expectations and mental disorders: A controlled trial. Arch Fam Med *8:*135, 1999.

Martin S: Indefinite delay in E & M. AMA News *41:*18, 1998.

Shea SC: *Psychiatric Interviewing*, ed 2. Saunders, Philadelphia, 1998.

*Simon RI: *Clinical Psychiatry and the Law*, ed 2. American Psychiatric Press, Washington, DC, 1992, p 89.

*Spitzer RL, Gibbon M, Skodol AE, Williams JBW, First MB: *DSM-IV Casebook: A Learning Companion to the Diagnostic and Statistical Manual of Mental Disorders*, ed 4. American Psychiatric Press, Washington, DC, 1994.

Stephenson W: Methodology of single case studies. J Oper Psychiatry *5:*3, 1974.

*Weed LL: Medical records that guide and teach. N Engl J Med *278:*593, 1968.

Wedding D, Corsini RJ: *Case Studies in Psychotherapy*. Peacock, Itaska, IL, 1989.

Zilboorg G, Henry GB: *A History of Medical Psychology*. Norton, New York, 1941.

7.3 SIGNS AND SYMPTOMS IN PSYCHIATRY

BENJAMIN J. SADOCK, M.D.

William Osler (1898–1919) said of medicine in general: “It is learned only by experience; it is not an inheritance; it cannot be revealed. Learn to see, learn to hear, learn to feel, learn to smell, and know that by practice alone can you become expert.” And so it is with psychiatry. One sees the posture of depression, hears the neologisms in schizophrenia; smells the odor of alcoholism, and feels the violent patient's anger. Eventually, with practice, the psychiatrist learns the full range of signs and symptoms. It is the rare psychiatrist, however, who encounters them all.

Signs are objective; symptoms are subjective. Signs are the clinician's observations, such as noting a patient's agitation; symptoms are subjective experiences, such as a person's complaint of feeling depressed. In psychiatry signs and symptoms are not so clearly demarcated as in other fields of medicine; they often overlap. Because of this, disorders in psychiatry are often described as syndromes—a constellation of signs and symptoms that together make up a recognizable condition. Schizophrenia, for example, is more often viewed as a syndrome than as a specific disorder. This concept is expressed in the terms “schizophrenic spectrum” or “the group of schizophrenias.”

NORMALITY

The concept of normality is ambiguous, has multiple meanings, and is influenced by diverse and complex value systems. Normality is sometimes equated broadly with health. The World Health Organization (WHO) defines normality as a state of complete physical, mental, and social well-being. Most psychiatrists consider normality to be the absence of psychopathology; but that concept is limited by defining mental health as simply the absence of mental disease.

Psychiatry has been criticized over the years by certain groups for its portrayal of normality. The psychology of women, for example, has been criticized as sexist because it was formulated initially by men; similar criticism comes from other groups who believe the portrayal of their psychological issues is biased by placing undue emphasis on psychopathology rather than healthy attributes. A much discussed issue is the change in psychiatry's view of homosexuality from abnormal to normal that took place 25 years ago, an evolution shaped by cultural norms, society's expectations and values, professional biases, individual differences, and the political climate of the time. Many attempts have been made to define normality in psychiatry; however, none has been entirely satisfactory. Table 7.3-1 outlines the divergent views of several psychoanalytic thinkers and clinicians.

Theorist	Concept
Sigmund Freud	Normality is an ideal fiction.
Karl Menninger	Normality is the ability to adjust to the external world with contentment and to master the task of existence.
Alfred Adler	The person's capacity to develop social feeling and to be productive is related to mental health; the ability to work heightens self-esteem and makes one capable of adaptation.
Clara Kirk	Normality is the capacity to live without fear, guilt, or anxiety and to take responsibility for one's own actions.

Table 7.3-1 Concepts of Normality

MENTAL HEALTH

The concept of mental health is continually evolving out of new experience and new research. Major theoreticians include, among others, Daniel Offer and Melvin Sabshin, who formulated four perspectives of normality.

Normality as Health The first perspective is basically the traditional medical psychiatric approach to health and illness. Normality is equated with health, and health is an almost universal phenomenon. Behavior is assumed to be within normal limits when no manifest psychopathology is present. If all behavior were to be put on a scale, normality would encompass the major portion of the continuum, and abnormality would be the small remainder.

Normality as Utopia The second perspective conceives of normality as a blend of the diverse elements of the mental apparatus that culminates in optimal functioning. This approach seeks perfection. When discussing normality however, Sigmund Freud stated, “A normal ego is like normality in general, an ideal fiction.”

Normality as Average The third perspective is based on a mathematical principle of the bell-shaped curve. This approach considers the middle range to be normal and both extremes, deviant. Variability is described only within the context of total groups, not of one individual.

Normality as Process The fourth perspective stresses that normal behavior is the result of interacting systems. The normality-as-process perspective stresses changes over time rather than a cross-sectional definition of normality.

Investigators who subscribe to this approach are found in the behavioral and social sciences. Most typical are Erik Erikson's conceptualization of epigenesis of personality development and the seven developmental stages he described as essential to the attainment of mature adult functioning.

Karl Jaspers Karl Jaspers (1883–1969), the German psychiatrist and philosopher, described a “personal world”—the way the person thinks or feels—that could be either normal or abnormal. According to Jaspers, the personal world is abnormal when it (1) springs from a condition that is recognized universally as abnormal such as schizophrenia, (2) when it separates the person from others emotionally, and (3) when it does not provide the person with a sense of “spiritual and material” security.

Jaspers was a proponent of phenomenology, in which the clinician studies psychological signs and symptoms with the goal of understanding the internal experience of the patient. By carefully listening to the patient, the psychiatrist temporarily enters the mental life of the patient. Jaspers believed that to fully understand the signs and symptoms observed in the patient, the clinician must have no prior assumptions. A person who reports a hallucinatory experience, for example, must not be judged thereby as being abnormal or psychotic. To be used diagnostically, the phenomenon must occur repeatedly and be characteristic of a known disorder.

Some investigators are developing a research strategy that measures a person's mental state at various times during the day in different life settings to validate diagnosis. Thus a sign or symptom is examined within the context of its appearance. What is abnormal in one setting or at one time of day may be normal in another.

Finally, there is the commonly accepted and widely used definition by Robert Campbell, adapted from his Psychiatric Dictionary: Psychically normal persons are those

who are in harmony with themselves and with their environment. They conform with the cultural requirements or injunctions of their community. They may possess medical deviation or disease, but as long as this does not impair their reasoning, judgment, intellectual capacity, and ability to make harmonious personal and social adaptation, they may be regarded as psychically sound or normal.

Many psychiatric signs and symptoms can be understood as various points on a spectrum of behavior ranging from normal to pathological. It is extremely rare to have a pathognomonic sign or symptom in psychiatry, although in some cases he disturbance is specific to a neurological deficit. As John Nemiah wrote: “Psychiatry is a science of inexhaustible complexity. It is as infinite as the range of human emotions and behavior. One cannot possibly learn it all.” Because of that, psychiatry still remains as much art as science.

GLOSSARY OF SIGNS AND SYMPTOMS

abreaction A process by which repressed material, particularly a painful experience or a conflict, is brought back to consciousness; in this process, the person not only recalls but relives the repressed material, which is accompanied by the appropriate affective response.

abstract thinking Thinking characterized by the ability to grasp the essentials of a whole, to break a whole into its parts and to discern common properties. To think symbolically.

abulia Reduced impulse to act and think, associated with indifference about consequences of action. Occurs as a result of neurological deficit, depression, schizophrenia.

acalculia Loss of ability to do calculations; not caused by anxiety or impairment in concentration. Occurs with neurological deficit, learning disorder.

acataphasia Disordered speech in which statements are incorrectly formulated. Patients may express themselves with words that sound like the ones intended but are not appropriate to the thoughts, or they may use totally inappropriate expressions.

acathexis Lack of feeling associated with an ordinarily emotion-charged subject; in psychoanalysis, it denotes the patient's detaching or transferring of emotion from thoughts and ideas. Also called *decathexis*. Occurs in anxiety, dissociative, schizophrenic, and bipolar disorders.

acenesthesia Loss of sensation of physical existence.

acrophobia Dread of high places.

acting out Behavioral response to an unconscious drive or impulse that brings about temporary partial relief of inner tension; relief is attained by reacting to a present situation as if it were the situation that originally gave rise to the drive or impulse. Common in borderline states.

aculalia Nonsense speech associated with marked impairment of comprehension. Occurs in mania, schizophrenia, neurological deficit.

adiadochokinesia Inability to perform rapid alternating movements. Occurs with neurological deficit, cerebellar lesions.

adynamia Weakness and fatigability, characteristic of neurasthenia and depression.

aerophagia Excessive swallowing of air. Seen in anxiety disorder.

affect The subjective and immediate experience of emotion attached to ideas or mental representations of objects. Affect has outward manifestations that may be classified as restricted, blunted, flattened, broad, labile, appropriate, or inappropriate. *See also* **mood**.

ageusia Lack or impairment of the sense of taste. Seen in depression, neurological deficit.

aggression Forceful, goal-directed action that may be verbal or physical; the motor counterpart of the affect of rage, anger, or hostility. Seen in neurological deficit, temporal lobe disorder, impulse-control disorders, mania, schizophrenia.

agitation Severe anxiety associated with motor restlessness.

agnosia Inability to understand the import or significance of sensory stimuli; cannot be explained by a defect in sensory pathways or cerebral lesion; the term has also been used to refer to the selective loss or disuse of knowledge of specific objects because of emotional circumstances, as seen in certain schizophrenic, anxious, and depressed patients. Occurs with neurological deficit. For types of agnosia, see the specific term.

agoraphobia Morbid fear of open places or leaving the familiar setting of the home. May be present with or without panic attacks.

agraphia Loss or impairment of a previously possessed ability to write.

ailurophobia Dread of cats.

akathisia Subjective feeling of motor restlessness manifested by a compelling need to be in constant movement; may be seen as an extrapyramidal adverse effect of antipsychotic medication. May be mistaken for psychotic agitation.

akinesia Lack of physical movement, as in the extreme immobility of catatonic schizophrenia; may also occur as an extrapyramidal effect of antipsychotic medication.

akinetic mutism Absence of voluntary motor movement or speech in a patient who is apparently alert (as evidenced by eye movements). Seen in psychotic depression, catatonic states.

alexia Loss of a previously possessed reading facility; not explained by defective visual acuity. *Compare* **dyslexia**.

alexithymia Inability or difficulty in describing or being aware of one's emotions or moods; or elaboration of fantasies associated with depression, substance abuse, and posttraumatic stress disorder.

algophobia Dread of pain.

alogia Inability to speak because of a mental deficiency or an episode of dementia.

ambivalence Coexistence of two opposing impulses toward the same thing in the same person at the same time. Seen in schizophrenia, borderline states, obsessive-compulsive disorders.

amimia Lack of ability to make gestures or to comprehend those made by others.

amnesia Partial or total inability to recall past experiences; may be organic (*amnestic disorder*) or emotional (*dissociative amnesia*) in origin.

amnestic aphasia Disturbed capacity to name objects, even though they are known to the patient. Also called *anomic aphasia*.

anaclitic Depending on others, especially as the infant on the mother; anaclitic depression in children results from an absence of mothering.

analgesia State in which one feels little or no pain. Can occur under hypnosis and in dissociative disorder.

anankasm Repetitious or stereotyped behavior or thought usually used as a tension-relieving device; used as a synonym for obsession and seen in obsessive-compulsive (anakastic) personality.

androgyny Combination of culturally determined female and male characteristics in one person.

anergia Lack of energy.

anhedonia Loss of interest in and withdrawal from all regular and pleasurable activities. Often associated with depression.

anomia Inability to recall the names of objects.

anorexia Loss or decrease in appetite. In *anorexia nervosa* appetite may be preserved but patient refuses to eat.

anosognosia Inability to recognize a physical deficit in oneself (e.g., patient denies paralyzed limb).

anterograde amnesia Loss of memory for events subsequent to the onset of the amnesia common after trauma. *Compare* **retrograde amnesia**.

anxiety Feeling of apprehension caused by anticipation of danger, which may be internal or external.

apathy Dulled emotional tone associated with detachment or indifference; observed in certain types of schizophrenia and depression.

aphasia Any disturbance in the comprehension or expression of language caused by a brain lesion. For types of aphasia, see the specific term.

aphonia Loss of voice. Seen in conversion disorder.

apperception Awareness of the meaning and significance of a particular sensory stimulus as modified by one's own experiences, knowledge, thoughts, and emotions. *See also* **perception**.

appropriate affect Emotional tone in harmony with the accompanying idea, thought, or speech.

apraxia Inability to perform a voluntary purposeful motor activity; cannot be explained by paralysis or other motor or sensory impairment. In *constructional apraxia*, a patient cannot draw two- or three-dimensional forms.

astasia abasia Inability to stand or walk in a normal manner, even though normal leg movements can be performed in a sitting or lying down position. Seen in conversion disorder.

astereognosis Inability to identify familiar objects by touch. Seen with neurological deficit. *See also* **neurological amnesia**.

asyndesis Disorder of language in which the patient combines unconnected ideas and images. Commonly seen in schizophrenia.

ataxia Lack of coordination, either physical or mental. 1. In neurology, refers to loss of muscular coordination. 2. In psychiatry, the term *intrapsychic ataxia* refers to lack of coordination between feelings and thoughts; seen in schizophrenia and in severe obsessive-compulsive disorder.

atonia Lack of muscle tone. *See* **waxy flexibility**.

attention Concentration; the aspect of consciousness that relates to the amount of effort exerted in focusing on certain aspects of an experience, activity, or task. Usually impaired in anxiety and depressive disorders.

auditory hallucination False perception of sound, usually voices but also other noises such as music. Most common hallucination in psychiatric disorders.

aura 1. Warning sensations such as automatisms, fullness in the stomach, blushing, and changes in respiration, cognitive sensations, and mood states usually experienced before a seizure. 2. A sensory prodrome that precedes a classic migraine headache.

autistic thinking Thinking in which the thoughts are largely narcissistic and egocentric, with emphasis on subjectivity rather than objectivity, and without regard for reality; used interchangeably with autism and dereism. Seen in schizophrenia, autistic disorder.

behavior Sum total of the psyche that includes impulses, motivations, wishes, drives, instincts, and cravings, as expressed by a person's behavior or motor activity. Also called *conation*.

bereavement Feeling of grief or desolation, especially at the death or loss of a loved one.

bizarre delusion False belief that is patently absurd or fantastic (e.g., invaders from space have implanted electrodes in a person's brain). Common in schizophrenia. In nonbizarre delusion content is usually within range of possibility.

blackout Amnesia experienced by alcoholics about behavior during drinking bouts; usually indicates reversible brain damage.

blocking Abrupt interruption in train of thinking before a thought or idea is finished; after a brief pause, person indicates no recall of what was being said or was going to be said (also known as *thought deprivation*). Common in schizophrenia and severe anxiety.

blunted affect Disturbance of affect manifested by a severe reduction in the intensity of externalized feeling tone; one of the fundamental symptoms of schizophrenia, as outlined by Eugen Bleuler.

bradykinesia Slowness of motor activity, with a decrease in normal spontaneous movement.

bradylalia Abnormally slow speech. Common in depression.

bradylexia Inability to read at normal speed.

bruxism Grinding or gnashing of the teeth, typically occurring during sleep. Seen in anxiety disorder.

carebaria Sensation of discomfort or pressure in the head.

catalepsy Condition in which persons maintain the body position into which they are placed; observed in severe cases of catatonic schizophrenia. Also called *waxy flexibility; cerea flexibilitas. See alsc* **command automatism**.

cataplexy Temporary sudden loss of muscle tone, causing weakness and immobilization; can be precipitated by a variety of emotional states and is often followed by sleep. Commonly seen in narcolepsy.

catatonic excitement Excited, uncontrolled motor activity seen in catatonic schizophrenia. Patients in catatonic state may suddenly erupt into excited state and be violent.

catatonic posturing Voluntary assumption of an inappropriate or bizarre posture, generally maintained for long periods of time. May switch unexpectedly with catatonic excitement.

catatonic rigidity Fixed and sustained motoric position that is resistant to change.

catatonic stupor Stupor in which patients ordinarily are well aware of their surroundings.

cathexis In psychoanalysis, a conscious or unconscious investment of psychic energy in an idea, concept, object or person. *Compare* **acathexis**.

causalgia Burning pain that may be either organic or psychic in origin.

cephalagia Headache.

cenesthetia Change in the normal quality of feeling tone in a part of the body.

cerea flexibilitas Condition of a person who can be molded into a position that is then maintained; when an examiner moves the person's limb, the limb feels as if it were made of wax. Also called *catalepsy* or *waxy flexibility*. Seen in schizophrenia.

chorea Movement disorder characterized by random and involuntary quick, jerky, purposeless movements. Seen in Huntington's disease.

circumstantiality Disturbance in the associative thought and speech processes in which a patient digresses into unnecessary details and inappropriate thoughts before communicating the central idea. Observed in schizophrenia, obsessional disturbances, and certain cases of dementia. *See also* **tangentiality**.

clang association Association or speech directed by the sound of a word rather than by its meaning; words have no logical connection; punning and rhyming may dominate the verbal behavior. Seen most frequently in schizophrenia or mania.

claustrophobia Abnormal fear of closed or confining spaces.

clonic convulsion An involuntary, violent muscular contraction or spasm in which the muscles alternately contract and relax. Characteristic phase in grand mal epileptic seizure.

clouding of consciousness Any disturbance of consciousness in which the person is not fully awake, alert, and oriented. Occurs in delirium, dementia, and cognitive disorder.

cluttering Disturbance of fluency involving an abnormally rapid rate and erratic rhythm of speech that impedes intelligibility; the affected individual is usually unaware of communicative impairment.

cognition Mental process of knowing and becoming aware; function closely associated with judgment.

coma State of profound unconsciousness from which a person cannot be roused, with minimal or no detectable responsiveness to stimuli; seen in injury or disease of the brain, in such systemic conditions as diabetic ketoacidosis and uremia, and in intoxications with alcohol and other drugs. Coma may also occur in severe catatonic states and in conversion disorder.

coma vigil Coma in which a patient appears to be asleep but can be aroused (also known as *akinetic mutism*).

command automatism Condition associated with catalepsy in which suggestions are followed automatically.

command hallucination False perception of orders that a person may feel obliged to obey or unable to resist.

complex A feeling-toned idea.

complex partial seizure A seizure characterized by alterations in consciousness that may be accompanied by complex hallucinations (sometimes olfactory) or illusions. During the seizure, a state of impaired consciousness resembling a dreamlike state may occur, and the patient may exhibit repetitive, automatic, or semipurposeful behavior.

compulsion Pathological need to act on an impulse that, if resisted, produces anxiety; repetitive behavior in response to an obsession or performed according to certain rules, with no true end in itself other than to prevent something from occurring in the future.

conation That part of a person's mental life concerned with cravings, strivings, motivations, drives, and wishes as expressed through behavior or motor activity.

concrete thinking Thinking characterized by actual things, events, and immediate experience, rather than by abstractions; seen in young children, in those who have lost or never developed the ability to generalize (as in certain cognitive mental disorders), and in schizophrenic persons. *Compare* **abstract thinking**.

condensation Mental process in which one symbol stands for a number of components.

confabulation Unconscious filling of gaps in memory by imagining experiences or events that have no basis in fact, commonly seen in amnestic syndromes; should be differentiated from lying. *See also* **paramnesia**.

confusion Disturbances of consciousness manifested by a disordered orientation in relation to time, place, or person.

consciousness State of awareness, with response to external stimuli.

constipation Inability to defecate or difficulty in defecating.

constricted affect Reduction in intensity of feeling tone less severe than that of blunted affect.

constructional apraxia Inability to copy a drawing, such as a cube, clock, or pentagon, as a result of a brain lesion.

conversion phenomena The development of symbolic physical symptoms and distortions involving the voluntary muscles or special sense organs; not under voluntary control and not explained by any physical disorder. Most common in conversion disorder, but also seen in a variety of mental disorders.

convulsion An involuntary, violent muscular contraction or spasm. *See also* **clonic convulsion** *and* **tonic convulsion**.

coprolalia Involuntary use of vulgar or obscene language. Observed in some cases of schizophrenia and in Tourette's disorder.

coprophagia Eating of filth or feces.

cryptolalia A private spoken language.

cryptographia A private written language.

cycloplegia Paralysis of the muscles of accommodation in the eye; observed at times as an autonomic adverse effect (anticholinergic effect) of antipsychotic or antidepressant medication.

decompensation Deterioration of psychic functioning caused by a breakdown of defense mechanisms. Seen in psychotic states.

déjà entendu Illusion that what one is hearing one has heard previously. *See also* **paramnesia**.

déjà pensé Condition in which a thought never entertained before is incorrectly regarded as a repetition of a previous thought. *See also* **paramnesia**.

déjà vu Illusion of visual recognition in which a new situation is incorrectly regarded as a repetition of a previous experience. *See also* **paramnesia**.

delirium Acute reversible mental disorder characterized by confusion and some impairment of consciousness; generally associated with emotional lability, hallucinations or illusions, and inappropriate, impulsive, irrational, or violent behavior.

delirium tremens Acute and sometimes fatal reaction to withdrawal from alcohol, usually occurring 72 to 96 hours after the cessation of heavy drinking; distinctive characteristics are marked autonomic hyperactivity (tachycardia, fever, hyperhidrosis, dilated pupils), usually accompanied by tremulousness, hallucinations, illusions, and delusions. Called *alcohol withdrawal delirium* in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). *See also* **formication**.

delusion False belief, based on incorrect inference about external reality, that is firmly held despite objective and obvious contradictory proof or evidence and despite the fact that other members of the culture do not share the belief.

delusion of control False belief that a person's will, thoughts, or feelings are being controlled by external forces.

delusion of grandeur Exaggerated conception of one's importance, power, or identity.

delusion of infidelity False belief that one's lover is unfaithful. Sometimes called *pathological jealousy*.

delusion of persecution False belief of being harassed or persecuted; often found in litigious patients who have a pathological tendency to take legal action because of imagined mistreatment. Most common delusion.

delusion of poverty False belief that one is bereft or will be deprived of all material possessions.

delusion of reference False belief that the behavior of others refers to oneself; that events, objects, or other people have a particular and unusual significance, usually of a negative nature; derived from idea of reference, in which persons falsely feel that others are talking about them (e.g., belief that people on television or radio are talking to or about the person). *See also* **thought broadcasting**.

delusion of self-accusation False feeling of remorse and guilt. Seen in depression with psychotic features.

dementia Mental disorder characterized by general impairment in intellectual functioning without clouding of consciousness; characterized by failing memory, difficulty with calculations, distractibility, alterations in mood and affect, impaired judgment and abstraction, reduced facility with language, and disturbance of orientation. Although irreversible because of underlying progressive degenerative brain disease, dementia may be reversible if the cause can be treated.

denial Defense mechanism in which the existence of unpleasant realities is disavowed; refers to keeping out of conscious awareness any aspects of external reality that, if acknowledged, would produce anxiety.

depersonalization Sensation of unreality concerning oneself, parts of oneself, or one's environment that occurs under extreme stress or fatigue. Seen in schizophrenia, depersonalization disorder, and schizotypal personality disorder.

depression Mental state characterized by feelings of sadness, loneliness, despair, low self-esteem, and self-reproach; accompanying signs include psychomotor retardation or at times agitation, withdrawal from interpersonal contact, and vegetative symptoms such as insomnia and anorexia. The term refers to either a mood that is so characterized or a mood disorder.

derailment Gradual or sudden deviation in train of thought without blocking; sometimes used synonymously with *loosening of association*.

derealization Sensation of changed reality or that one's surroundings have altered. Usually seen in schizophrenia, panic attacks, dissociative disorders.

dereism Mental activity that follows a totally subjective and idiosyncratic system of logic and fails to take the facts of reality or experience into consideration. Characteristic of schizophrenia. *See also* **autistic thinking**.

detachment Characterized by distant interpersonal relationships and lack of emotional involvement.

devaluation Defense mechanism in which a person attributes excessively negative qualities to self or others. Seen in depression, paranoid personality disorder.

diminished libido Decreased sexual interest and drive. (Increased libido is often associated with mania.)

dipsomania Compulsion to drink alcoholic beverages.

disinhibition 1. Removal of an inhibitory effect, as in the reduction of the inhibitory function of the cerebral cortex by alcohol. 2. In psychiatry, a greater freedom to act in accordance with inner drives or feelings and with less regard for restraints dictated by cultural norms or one's superego.

disorientation Confusion; impairment of awareness of time, place, and person (the position of the self in relation to other persons). Characteristic of cognitive

disorders.

displacement Unconscious defense mechanism by which the emotional component of an unacceptable idea or object is transferred to a more acceptable one. Seen in phobias.

dissociation Unconscious defense mechanism involving the segregation of any group of mental or behavioral processes from the rest of the person's psychic activity; may entail the separation of an idea from its accompanying emotional tone, as seen in dissociative and conversion disorders. Seen in dissociative disorders.

distractibility Inability to focus one's attention; the patient does not respond to the task at hand but attends to irrelevant phenomena in the environment.

dread Massive or pervasive anxiety, usually related to a specific danger.

dreamy state Altered state of consciousness, likened to a dream situation, that develops suddenly and usually lasts a few minutes; accompanied by visual, auditory, and olfactory hallucinations. Commonly associated with temporal lobe lesions.

drowsiness State of impaired awareness associated with a desire or inclination to sleep.

dysarthria Difficulty in articulation, the motor activity of shaping phonated sounds into speech, not in word finding or in grammar.

dyscalculia Difficulty in performing calculations.

dysgeusia Impaired sense of taste.

dysgraphia Difficulty in writing.

dyskinesia Difficulty in performing movements. Seen in extrapyramidal disorders.

dyslalia Faulty articulation caused by structural abnormalities of the articulatory organs or impaired hearing.

dyslexia Specific learning disability syndrome involving an impairment of the previously acquired ability to read; unrelated to the person's intelligence. *Compare* **alexia**.

dysmetria Impaired ability to gauge distance relative to movements. Seen in neurological deficit.

dysmnesia Impaired memory.

dyspareunia Physical pain in sexual intercourse, usually emotionally caused and more commonly experienced by women; may also result from cystitis, urethritis, or other medical conditions.

dysphagia Difficulty in swallowing.

dysphasia Difficulty in comprehending oral language (*reception dysphasia*) or in trying to express verbal language (*expressive dysphasia*).

dysphonia Difficulty or pain in speaking.

dysphoria Feeling of unpleasantness or discomfort; a mood of general dissatisfaction and restlessness. Occurs in depression and anxiety.

dysprosody Loss of normal speech melody (*prosody*). Common in depression.

dystonia Extrapyramidal motor disturbance consisting of slow, sustained contractions of the axial or appendicular musculature; one movement often predominates, leading to relatively sustained postural deviations; acute dystonic reactions (facial grimacing, torticollis) are occasionally seen with the initiation of antipsychotic drug therapy.

echolalia Psychopathological repeating of words or phrases of one person by another; tends to be repetitive and persistent. Seen in certain kinds of schizophrenia, particularly the catatonic types.

ego-alien Denoting aspects of a person's personality that are viewed as repugnant, unacceptable, or inconsistent with the rest of the personality. Also called *ego-dystonia. Compare* **ego-syntonic**.

egocentric Self-centered; selfishly preoccupied with one's own needs; lacking interest in others.

ego-dystonic *See* **ego-alien**.

egomania Morbid self-preoccupation or self-centeredness. *See also* **narcissism**.

ego-syntonic Denoting aspects of a personality that are viewed as acceptable and consistent with that person's total personality. Personality traits are usually ego-syntonic. *Compare* **ego-alien**.

eidetic image Unusually vivid or exact mental image of objects previously seen or imagined.

elation Mood consisting of feelings of joy, euphoria, triumph, and intense self-satisfaction, or optimism. Occurs in mania when not grounded in reality.

elevated mood Air of confidence and enjoyment; a mood more cheerful than normal but not necessarily pathological.

emotion Complex feeling state with psychic, somatic, and behavioral components; external manifestation of emotion is *affect*.

emotional insight A level of understanding or awareness that one has emotional problems. It facilitates positive changes in personality and behavior when present.

emotional lability Excessive emotional responsiveness characterized by unstable and rapidly changing emotions.

encopresis Involuntary passage of feces, usually occurring at night or during sleep.

enuresis Incontinence of urine during sleep.

erotomania Delusional belief, more common in women than in men, that someone is deeply in love with them (also known as *De Clérembault's syndrome*).

erythrophobia Abnormal fear of blushing.

euphoria Exaggerated feeling of well-being that is inappropriate to real events. Can occur with drugs such as opiates, amphetamines, and alcohol.

euthymia Normal range of mood, implying absence of depressed or elevated mood.

evasion Act of not facing up to, or strategically eluding, something; consists of suppressing an idea that is next in a thought series and replacing it with another idea closely related to it. Also called *paralogia; perverted logic*.

exaltation Feeling of intense elation and grandeur.

excited Agitated, purposeless motor activity uninfluenced by external stimuli.

expansive mood Expression of feelings without restraint, frequently with an overestimation of their significance or importance. Seen in mania, grandiose delusional disorder.

expressive aphasia Disturbance of speech in which understanding remains but ability to speak is grossly impaired; halting, laborious, and inaccurate speech (also known as *Broca's, nonfluent*, and *motor aphasia*).

expressive dysphasia Difficulty in expressing verbal language; the ability to understand language is intact.

externalization More general term than *projection* that refers to the tendency to perceive in the external world and in external objects elements of one's own personality, including instinctual impulses, conflicts, moods, attitudes, and styles of thinking.

extroversion State of one's energies being directed outside oneself. *Compare* **introversion**.

fantasy Daydream; fabricated mental picture of a situation or chain of events. A normal form of thinking dominated by unconsciousness material that seeks wish fulfillment and solutions to conflicts; may serve as the matrix for creativity. The content of the fantasy may indicate mental illness.

false memory A person's recollection and belief by the patient of an event that did not actually occur. In *false memory syndrome* persons erroneously believe that they sustained an emotional or physical (e.g., sexual) trauma in early life.

fatigue A feeling of weariness, sleepiness, or irritability following a period of mental or bodily activity. Seen in depression, anxiety, neurasthenia, and somatoform disorders.

fausse reconnaissance False recognition, a feature of paramnesia. Can occur in delusional disorders.

fear Unpleasurable emotional state consisting of psychophysiological changes in response to a realistic threat or danger. *Compare* **anxiety**.

flat affect Absence or near absence of any signs of affective expression.

flight of ideas Rapid succession of fragmentary thoughts or speech in which content changes abruptly and speech may be incoherent. Seen in mania.

floccillation Aimless plucking or picking, usually at bedclothes or clothing, commonly seen in dementia and delirium.

fluent aphasia Aphasia characterized by inability to understand the spoken word; fluent but incoherent speech is present. Also called *Wernicke's, sensory*, and *receptive aphasia*.

folie à deux Mental illness shared by two persons, usually involving a common delusional system; if it involves three persons, it is referred to as *folie à trois,* etc. Also called *shared psychotic disorder*.

formal thought disorder Disturbance in the form of thought rather than the content of thought; thinking characterized by loosened associations, neologisms, and illogical constructs; thought process is disordered, and the person is defined as psychotic. Characteristic of schizophrenia.

formication Tactile hallucination involving the sensation that tiny insects are crawling over the skin. Seen in cocaine addiction and delirium tremens.

free-floating anxiety Severe, pervasive, generalized anxiety that is not attached to any particular idea, object, or event. Observed particularly in anxiety disorders, although it may be seen in some cases of schizophrenia.

fugue Dissociative disorder characterized by a period of almost complete amnesia, during which a person actually flees from an immediate life situation and begins a different life pattern; apart from the amnesia, mental faculties and skills are usually unimpaired.

galactorrhea Abnormal discharge of milk from the breast; may result from the endocrine influence (e.g., prolactin) of dopamine receptor antagonists, such as phenothiazines.

generalized tonic-clonic seizure Generalized onset of tonic-clonic movements of the limbs, tongue biting, and incontinence followed by slow, gradual recovery of consciousness and cognition; also called *grand mal seizure*.

global aphasia Combination of grossly nonfluent aphasia and severe fluent aphasia.

glossolalia Unintelligible jargon that has meaning to the speaker but not to the listener. Occurs in schizophrenia.

grandiosity Exaggerated feelings of one's importance, power, knowledge, or identity. Occurs in delusional disorder, manic states.

grief Alteration in mood and affect consisting of sadness appropriate to a real loss; normally, it is self limited. *See also* **depression; mourning**.

guilt Emotional state associated with self-reproach and the need for punishment. In psychoanalysis, refers to a feeling of culpability that stems from a conflict between the ego and the superego (conscience). Guilt has normal psychological and social functions, but special intensity or absence of guilt characterizes many mental disorders, such as depression and antisocial personality disorder, respectively. Psychiatrists distinguish shame as a less internalized form of guilt that relates more to others than to the self. *See also* **shame**.

gustatory hallucination Hallucination primarily involving taste.

gynecomastia Femalelike development of the male breasts; may occur as an adverse effect of antipsychotic and antidepressant drugs because of increased prolactin levels or anabolic-androgenic steroid abuse.

hallucination False sensory perception occurring in the absence of any relevant external stimulation of the sensory modality involved. For types of hallucinations, see the specific term.

hallucinosis State in which a person experiences hallucinations without any impairment of consciousness.

haptic hallucination Hallucination of touch.

hebephrenia Complex of symptoms, considered a form of schizophrenia, characterized by wild or silly behavior or mannerisms, inappropriate affect, and delusions and hallucinations that are transient and unsystematized. Hebephrenic schizophrenia is now called *disorganized schizophrenia*.

holophrastic Using a single word to express a combination of ideas. Seen in schizophrenia.

hyperactivity Increased muscular activity. The term is commonly used to describe a disturbance found in children that is manifested by constant restlessness, overactivity, distractibility, and difficulties in learning. Seen in *attention-deficit/hyperactivity disorder*.

hyperalgesia Excessive sensitivity to pain. Seen in somatoform disorder.

hyperesthesia Increased sensitivity to tactile stimulation.

hypermnesia Exaggerated degree of retention and recall. It can be elicited by hypnosis and may be seen in certain prodigies; also may be a feature of obsessive-compulsive disorder, some cases of schizophrenia, and manic episodes of bipolar I disorder.

hyperphagia Increase in appetite and intake of food.

hyperpragia Excessive thinking and mental activity. Generally associated with manic episodes of bipolar I disorder.

hypersomnia Excessive time spent asleep. May be associated with underlying medical or psychiatric disorder, narcolepsy, be part of the Klein-Levin syndrome, or be primary.

hyperventilation Excessive breathing, generally associated with anxiety, which can reduce blood carbon dioxide concentration and produce lightheadedness, palpitations, numbness, and tingling periorally and in the extremities, and occasionally syncope.

hypervigilance Excessive attention to, and focus on, all internal and external stimuli; usually seen in delusional or paranoid states.

hypesthesia Diminished sensitivity to tactile stimulation.

hypnagogic hallucination Hallucination occurring while falling asleep, not ordinarily considered pathological.

hypnopompic hallucination Hallucination occurring while awakening from sleep, ordinarily not considered pathological.

hypnosis Artificially induced alteration of consciousness characterized by increased suggestibility and receptivity to direction.

hypoactivity Decreased motor and cognitive activity, as in psychomotor retardation; visible slowing of thought, speech, and movements. Also called *hypokinesis*.

hypochondria Exaggerated concern about health that is based not on real medical pathology but on unrealistic interpretations of physical signs or sensations as abnormal.

hypomania Mood abnormality with the qualitative characteristics of mania but somewhat less intense. Seen in cyclothymic disorder.

idea of reference Misinterpretation of incidents and events in the outside world as having direct personal reference to oneself; occasionally observed in normal persons, but frequently seen in paranoid patients. If present with sufficient frequency or intensity or if organized and systematized, they constitute delusions of reference.

illogical thinking Thinking containing erroneous conclusions or internal contradictions; psychopathological only when it is marked and not caused by cultural values or intellectual deficit.

illusion Perceptual misinterpretation of a real external stimulus. *Compare* **hallucination**.

immediate memory Reproduction, recognition, or recall of perceived material within seconds after presentation. *Compare* **long-term memory; short-term memory**.

impaired insight Diminished ability to understand the objective reality of a situation.

impaired judgment Diminished ability to understand a situation correctly and to act appropriately.

impulse control Ability to resist an impulse, drive, or temptation to perform some action.

inappropriate affect Emotional tone out of harmony with the idea, thought, or speech accompanying it. Seen in schizophrenia.

incoherence Communication that is disconnected, disorganized, or incomprehensible. *See also* **word salad**.

incorporation Primitive unconscious defense mechanism in which the psychic representation of another person or aspects of another person are assimilated into oneself through a figurative process of symbolic oral ingestion; represents a special form of introjection and is the earliest mechanism of identification.

increased libido Increase in sexual interest and drive.

ineffability Ecstatic state in which persons insist that their experience is inexpressible and indescribable, that it is impossible to convey what it is like to one who never experienced it.

initial insomnia Falling asleep with difficulty; usually seen in anxiety disorder. *Compare* **middle insomnia; terminal insomnia**.

insight Conscious recognition of one's own condition. In psychiatry, it refers to the conscious awareness and understanding of one's own psychodynamics and symptoms of maladaptive behavior; highly important in effecting changes in the personality and behavior of a person.

insomnia Difficulty in falling asleep or difficulty in staying asleep. It can be related to a mental disorder, can be related to a physical disorder or an adverse effect of medication, or can be primary (not related to a known medical factor or another mental disorder). *See also* **initial insomnia; middle insomnia; terminal insomnia**.

intellectual insight Knowledge of the reality of a situation without the ability to use that knowledge successfully to effect an adaptive change in behavior or master the situation. *Compare* **true insight**.

intelligence Capacity for learning and ability to recall, integrate constructively, and apply what one has learned; the capacity to understand and think rationally.

intoxication Mental disorder caused by recent ingestion or presence in the body of an exogenous substance producing maladaptive behavior by virtue of its effects on the central nervous system. The most common psychiatric changes involve disturbances of perception, wakefulness, attention, thinking, judgment, emotional control, and psychomotor behavior; the specific clinical picture depends on the substance ingested.

intropunitive Turning anger inward toward oneself. Commonly observed in depressed patients.

introspection Contemplating one's own mental processes to achieve insight.

introversion State in which a person's energies are directed inward toward the self, with little or no interest in the external world.

irrelevant answer Answer that is not responsive to the question.

irritability Abnormal or excessive excitability, with easily triggered anger, annoyance, or impatience.

irritable mood State in which one is easily annoyed and provoked to anger. *See also* **irritability**.

jamais vu Paramnestic phenomenon characterized by a false feeling of unfamiliarity with a real situation that one has previously experienced.

jargon aphasia Aphasia in which the words produced are neologistic; that is, nonsense words created by the patient.

judgment Mental act of comparing or evaluating choices within the framework of a given set of values for the purpose of electing a course of action. If the course of action chosen is consonant with reality or with mature adult standards of behavior, judgment is said to be *intact* or *normal*; judgement is said to be *impaired* if the chosen course of action is frankly maladaptive, results from impulsive decisions based on the need for immediate gratification, or is otherwise not consistent with reality as measured by mature adult standards.

kleptomania Pathological compulsion to steal.

la belle indifférence Inappropriate attitude of calm or lack of concern about one's disability. May be seen in patients with conversion disorder.

labile affect Affective expression characterized by rapid and abrupt changes, unrelated to external stimuli.

labile mood Oscillations in mood between euphoria and depression or anxiety.

laconic speech Condition characterized by a reduction in the quantity of spontaneous speech; replies to questions are brief and unelaborated, and little or no unprompted additional information is provided. Occurs in major depression, schizophrenia, and organic mental disorders. Also called *poverty of speech*.

lethologica Momentary forgetting of a name or proper noun. *See* **blocking**.

Lilliputian hallucination Visual sensation that persons or objects are reduced in size, more properly regarded as an illusion. *See also* **micropsia**.

localized amnesia Partial loss of memory; amnesia restricted to specific or isolated experiences. Also called *lacunar amnesia; patch amnesia*.

logorrhea Copious, pressured, coherent speech; uncontrollable, excessive talking; observed in manic episodes of bipolar disorder. Also called *tachylogia; verbomania; volubility*.

loosening of associations Characteristic schizophrenic thinking or speech disturbance involving a disorder in the logical progression of thoughts, manifested as a failure to communicate verbally adequately; unrelated and unconnected ideas shift from one subject to another. *See also* **tangentiality**.

macropsia False perception that objects are larger than they really are. *Compare* **micropsia**.

magical thinking A form of dereistic thought; thinking similar to that of the preoperational phase in children (Jean Piaget), in which thoughts, words, or actions assume power (e.g., to cause or prevent events).

malingering Feigning disease to achieve a specific goal, for example, to avoid an unpleasant responsibility. *Compare* **factitious disorder**.

mania Mood state characterized by elation, agitation, hyperactivity, hypersexuality, and accelerated thinking and speaking (flight of ideas). Seen in bipolar I disorder. *See also* **hypomania**.

manipulation Maneuvering by patients to get their own way, characteristic of antisocial personalities.

mannerism Ingrained, habitual involuntary movement.

melancholia Severe depressive state. Used in the term *involutional melancholia* as a descriptive term and also in reference to a distinct diagnostic entity.

memory Process whereby what is experienced or learned is established as a record in the central nervous system (registration), where it persists with a variable degree of permanence (retention) and can be recollected or retrieved from storage at will (recall). For types of memory, see the specific term.

mental disorder Psychiatric illness or disease whose manifestations are primarily characterized by behavioral or psychological impairment of function, measured in terms of deviation from some normative concept; associated with distress or disease, not just an expected response to a particular event or limited to relations between a person and society.

mental retardation Subaverage general intellectual functioning that originates in the developmental period and is associated with impaired maturation and learning, and social maladjustment. Retardation is commonly defined in terms of intelligence quotient (I.Q.): mild (50–55 to 70), moderate (35–40 to 50–55), severe (20–25 to 35–40) and profound (below 20–25).

metonymy Speech disturbance common in schizophrenia in which the affected person uses a word or phrase that is related to the proper one but is not the one ordinarily used; for example, the patient speaks of consuming a “menu” rather than a “meal,” or refers to losing the “piece of string” of the conversation, rather than the “thread” of the conversation. *See also* **paraphasia; word approximation**.

microcephaly Condition in which the head is unusually small as a result of defective brain development and premature ossification of the skull.

micropsia False perception that objects are smaller than they really are. Sometimes called *Lilliputian hallucination. Compare* **macropsia**.

middle insomnia Waking up after falling asleep without difficulty and then having difficulty in falling asleep again. *Compare* **initial insomnia; terminal insomnia**.

mimicry Simple, imitative motion activity of childhood.

monomania Mental state characterized by preoccupation with one subject.

mood Pervasive and sustained feeling tone that is experienced internally and that, in the extreme, can markedly influence virtually all aspects of a person's behavior and perception of the world. Distinguished from affect, the external expression of the internal feeling tone. For types of mood, see the specific term.

mood-congruent delusion Delusion with content that is mood appropriate (e.g., depressed patients who believe they are responsible for the destruction of the world).

mood-congruent hallucination Hallucination with content that is consistent with either a depressed or manic mood (e.g., depressed patients hearing voices telling them that they are bad persons; manic patients hearing voices telling them that they have inflated worth, power, or knowledge).

mood-incongruent delusion Delusion based on incorrect reference about external reality, with content that has no association to mood or is mood inappropriate (e.g., depressed patients who believe that they are the new Messiah).

mood-incongruent hallucination Hallucination not associated with real external stimuli, with content that is not consistent with either depressed or manic mood (e.g., in depression, hallucinations not involving such themes as guilt, deserved punishment, or inadequacy; in mania, not involving such themes as inflated worth or power).

mood swings Oscillation of a person's emotional feeling tone between periods of elation and periods of depression.

motor aphasia Aphasia in which understanding is intact but the ability to speak is lost. Also called *Broca's expressive* or *nonfluent aphasia*.

mourning Syndrome following loss of a loved one, consisting of preoccupation with the lost individual, weeping, sadness, and repeated reliving of memories. *See also* **bereavement; grief**.

muscle rigidity State in which the muscles remain immovable; seen in schizophrenia.

mutism Organic or functional absence of the faculty of speech. *See also* **stupor**.

mydriasis Dilation of the pupil; sometimes occurs as an autonomic (anticholinergic) or atropine-like adverse effect of some antipsychotic and antidepressant drugs.

needle phobia The persistent, intense, pathological fear of receiving an injection.

negativism Verbal or nonverbal opposition or resistance to outside suggestions and advice; commonly seen in catatonic schizophrenia in which the patient resists any effort to be moved or does the opposite of what is asked.

negative signs In schizophrenia: flat affect, alogia, abulia, apathy.

neologism New word or phrase whose derivation cannot be understood; often seen in schizophrenia. It has also been used to mean a word that has been incorrectly constructed but whose origins are nonetheless understandable (e.g., “headshoe” to mean “hat”), but such constructions are more properly referred to as *word approximations*.

neurological amnesia 1. Auditory amnesia: loss of ability to comprehend sounds or speech. 2. Tactile amnesia: loss of ability to judge the shape of objects by touch. *See also* **astereognosis**. 3. Verbal amnesia: loss of ability to remember words. 4. Visual amnesia: loss of ability to recall or recognize familiar objects or printed words.

nihilism Delusion of the nonexistence of the self or part of the self; also refers to an attitude of total rejection of established values or extreme skepticism regarding moral and value judments.

nihilistic delusion Depressive delusion that the world and everything related to it have ceased to exist.

noeisis Revelation in which immense illumination occurs in association with a sense that one has been chosen to lead and command. Can occur in manic or dissociative states.

nominal aphasia Aphasia characterized by difficulty in giving the correct name of an object. *See also* **anomia; amnestic aphasia**.

nymphomania Abnormal, excessive, insatiable desire in a female for sexual intercourse. *Compare* **satyriasis**.

obsession Persistent and recurrent idea, thought, or impulse that cannot be eliminated from consciousness by logic or reasoning; obsessions are involuntary and ego-dystonic. *See also* **compulsion**.

olfactory hallucination Hallucination primarily involving smell or odors; most common in medical disorders, especially in the temporal lobe.

orientation State of awareness of oneself and one's surroundings in terms of time, place, and person.

overactivity Abnormality in motor behavior that can manifest itself as psychomotor agitation, hyperactivity (hyperkinesis), tics, sleepwalking, or compulsions.

overvalued idea False or unreasonable belief or idea that is sustained beyond the bounds of reason. It is held with less intensity or duration than a delusion, but is usually associated with mental illness.

panic Acute, intense attack of anxiety associated with personality disorganization; the anxiety is overwhelming and accompanied by feelings of impending doom.

panphobia Overwhelming fear of everything.

pantomime Gesticulation; psychodrama without the use of words.

paramnesia Disturbance of memory in which reality and fantasy are confused. It is observed in dreams and in certain types of schizophrenia and organic mental disorders; includes phenomena such as déjà vu and *déjà entendu*, which may occur occasionally in normal persons.

paranoia Rare psychiatric syndrome marked by the gradual development of a highly elaborate and complex delusional system, generally involving persecutory or grandiose delusions, with few other signs of personality disorganization or thought disorder.

paranoid delusions Includes persecutory delusions and delusions of reference, control, and grandeur.

paranoid ideation Thinking dominated by suspicious, persecutory, or grandiose content of less than delusional proportions.

paraphasia Abnormal speech in which one word is substituted for another, the irrelevant word generally resembling the required one in morphology, meaning, or phonetic composition; the inappropriate word may be either a legitimate one used incorrectly, such as “clover” instead of “hand,” or a bizarre nonsense expression, such as “treen” instead of “train.” Paraphasic speech may be seen in organic aphasias and in mental disorders such as schizophrenia. *See also* **metonymy; word approximation**.

parapraxis Faulty act, such as a slip of the tongue or the misplacement of an article. Freud ascribed parapraxes to unconscious motives.

paresis Weakness or partial paralysis of organic origin.

paresthesia Abnormal spontaneous tactile sensation, such as a burning, tingling, or pins-and-needles sensation.

perception Conscious awareness of elements in the environment by the mental processing of sensory stimuli; sometimes used in a broader sense to refer to the mental process by which all kinds of data, intellectual, emotional, as well as sensory, are meaningfully organized. *See also* **apperception**.

perseveration 1. Pathological repetition of the same response to different stimuli, as in a repetition of the same verbal response to different questions. 2. Persistent repetition of specific words or concepts in the process of speaking. Seen in cognitive disorders, schizophrenia, and other mental illness. *See also* **verbigeration**.

phantom limb False sensation that an extremity that has been lost is in fact present.

phobia Persistent, pathological, unrealistic, intense fear of an object or situation; the phobic person may realize that the fear is irrational but, nonetheless, cannot dispel it. For types of phobias, see the specific term.

pica Craving and eating of nonfood substances, such as paint and clay.

polyphagia Pathological overeating.

positive signs In schizophrenia: hallucinations, delusions, thought disorder.

posturing Strange, fixed, and bizarre bodily positions held by a patient for an extended time. *See also* **catatonia**.

poverty of content of speech Speech that is adequate in amount but conveys little information because of vagueness, emptiness, or stereotyped phrases.

poverty of speech Restriction in the amount of speech used; replies may be monosyllabic. *See also* **laconic speech**.

preoccupation of thought Centering of thought content on a particular idea, associated with a strong affective tone, such as a paranoid trend or a suicidal or homicidal preoccupation.

pressured speech Increase in the amount of spontaneous speech; rapid, loud, accelerated speech, as occurs in mania, schizophrenia, and cognitive disorders.

primary process thinking In psychoanalysis, the mental activity directly related to the functions of the id and characteristic of unconscious mental processes; marked by primitive, prelogical thinking and by the tendency to seek immediate discharge and gratification of instinctual demands. Includes thinking that is dereistic, illogical, magical; normally found in dreams, abnormally in psychosis. *Compare* **secondary process thinking**.

projection Unconscious defense mechanism in which persons attribute to another those generally unconscious ideas, thoughts, feelings, and impulses that are in themselves undesirable or unacceptable as a form of protection from anxiety arising from an inner conflict; by externalizing whatever is unacceptable, they deal with it as a situation apart from themselves.

prosopagnosia Inability to recognize familiar faces that is not due to impaired visual acuity or level of consciousness.

pseudocyesis Rare condition in which a nonpregnant patient has the signs and symptoms of pregnancy, such as abdominal distention, breast enlargement, pigmentation, cessation of menses, and morning sickness.

pseudodementia 1. Dementia-like disorder that can be reversed by appropriate treatment and is not caused by organic brain disease. 2. Condition in which patients show exaggerated indifference to their surroundings in the absence of a mental disorder; also occurs in depression and factitious disorders.

pseudologia phantastica Disorder characterized by uncontrollable lying in which patients elaborate extensive fantasies that they freely communicate and act upon.

psychomotor agitation Physical and mental overactivity that is usually nonproductive and is associated with a feeling of inner turmoil, as seen in agitated depression.

psychosis Mental disorder in which the thoughts, affective response, ability to recognize reality, and ability to communicate and relate to others are sufficiently impaired to interfere grossly with the capacity to deal with reality; the classical characteristics of psychosis are impaired reality testing, hallucinations, delusions, and illusions.

psychotic 1. Person suffering from psychosis. 2. Denoting or characteristic of psychosis.

rationalization An unconscious defense mechanism in which irrational or unacceptable behavior, motives, or feelings are logically justified or made consciously tolerable by plausible means.

reaction formation Unconscious defense mechanism in which a person develops a socialized attitude or interest that is the direct antithesis of some infantile wish or impulse that is harbored either consciously or unconsciously. One of the earliest and most unstable defense mechanisms, closely related to repression; both are defenses against impulses or urges that are unacceptable to the ego.

reality testing Fundamental ego function that consists of tentative actions that test and objectively evaluate the nature and limits of the environment; includes the ability to differentiate between the external world and the internal world and to accurately judge the relation between the self and the environment.

recall Process of bringing stored memories into consciousness. *See also* **memory**.

recent memory Recall of events over the past few days.

recent past memory Recall of events over the past few months.

receptive aphasia Organic loss of ability to comprehend the meaning of words; fluid and spontaneous but incoherent and nonsensical speech. *See also* **fluent aphasia; sensory aphasia**.

receptive dysphasia Difficulty in comprehending oral language; the impairment involves both comprehension and production of language.

regression Unconscious defense mechanism in which a person undergoes a partial or total return to earlier patterns of adaptation; observed in many psychiatric conditions, particularly schizophrenia.

remote memory Recall of events in distant past.

repression Freud's term for an unconscious defense mechanism in which unacceptable mental contents are banished or kept out of consciousness; important in both normal psychological development and in neurotic and psychotic symptom formation. Freud recognized two kinds of repression: (1) repression proper, in which the repressed material was once in the conscious domain and (2) primal repression, in which the repressed material was never in the conscious realm. *Compare* **suppression**.

restricted affect Reduction in intensity of feeling tone less severe than in blunted affect but clearly reduced. *See also* **constricted affect**.

retrograde amnesia Loss of memory for events preceding the onset of the amnesia. *Compare* **anterograde amnesia**.

retrospective falsification Memory becomes unintentionally (unconsciously) distorted by being filtered through a person's present emotional, cognitive, and experiential state.

rigidity In psychiatry, a person's resistance to change, a personality trait.

ritual 1. Formalized activity practiced by a person to reduce anxiety, as in obsessive-compulsive disorder. 2. Ceremonial activity of cultural origin.

rumination Constant preoccupation with thinking about a single idea or theme, as in obsessive-compulsive disorder.

satyriasis Morbid, insatiable sexual need or desire in a male. *Compare* **nymphomania**.

scotoma 1. In psychiatry, a figurative blind spot in a person's psychological awareness. 2. In neurology, a localized visual field defect.

secondary process thinking In psychoanalysis, the form of thinking that is logical, organized, reality oriented, and influenced by the demands of the environment; characterizes the mental activity of the ego. *Compare* **primary process thinking**.

seizure An attack or sudden onset of certain symptoms, such as convulsions, loss of consciousness, and psychic or sensory disturbances; seen in epilepsy and can be substance induced. For types of seizures, see the specific term.

sensorium Hypothetical sensory center in the brain that is involved with clarity of awareness about oneself and one's surroundings, including the ability to perceive and process ongoing events in light of past experiences, future options, and current circumstances; sometimes used interchangeably with *consciousness*.

sensory aphasia Organic loss of ability to comprehend the meaning of words; fluid and spontaneous but incoherent and nonsensical speech. *See also* **fluent aphasia; receptive aphasia**.

sensory extinction Neurological sign operationally defined as failure to report one of two simultaneously presented sensory stimuli, despite the fact that either stimulus alone is correctly reported. Also called *sensory inattention*.

shame Failure to live up to self-expectations; often associated with fantasy of how person will be seen by others. *See also* **guilt**.

simultanagnosia Impairment in the perception or integration of visual stimuli appearing simultaneously.

somatic delusion Delusion pertaining to the functioning of one's body.

somatic hallucination Hallucination involving the perception of a physical experience localized within the body.

somatopagnosia Inability to recognize a part as one's own (also called *ignorance of the body* and *autotopagnosia*).

somnolence Pathological sleepiness or drowsiness from which one can be aroused to a normal state of consciousness.

spatial agnosia Inability to recognize spatial relations.

speaking in tongues Expression of a revelatory message through unintelligible words; not considered a disorder of thought if associated with practices of specific Pentecostal religions. *See also* **glossolalia**.

stereotypy Continuous mechanical repetition of speech or physical activities; observed in catatonic schizophrenia.

stupor 1. State of decreased reactivity to stimuli and less than full awareness of one's surroundings; as a disturbance of consciousness, it indicates a condition of partial coma or semicoma. 2. In psychiatry, used synonymously with *mutism* and does not necessarily imply a disturbance of consciousness; in *catatonic stupor*, patients are ordinarily aware of their surroundings.

stuttering Frequent repetition or prolongation of a sound or syllable, leading to markedly impaired speech fluency.

sublimation Unconscious defense mechanism in which the energy associated with unacceptable impulses or drives is diverted into personally and socially acceptable channels; unlike other defense mechanisms, it offers some minimal gratification of the instinctual drive or impulse.

substitution Unconscious defense mechanism in which a person replaces an unacceptable wish, drive, emotion, or goal with one that is more acceptable.

suggestibility State of uncritical compliance with influence or of uncritical acceptance of an idea, belief, or attitude; commonly observed among persons with hysterical traits.

suicidal ideation Thoughts or act of taking one's own life.

suppression Conscious act of controlling and inhibiting an unacceptable impulse, emotion, or idea; differentiated from repression in that repression is an unconscious process.

symbolization Unconscious defense mechanism in which one idea or object comes to stand for another because of some common aspect or quality in both; based on similarity and association; the symbols formed protect the person from the anxiety that may be attached to the original idea or object.

synesthesia Condition in which the stimulation of one sensory modality is perceived as sensation in a different modality, as when a sound produces a sensation of

color.

syntactical aphasia Aphasia characterized by difficulty in understanding spoken speech, associated with gross disorder of thought and expression.

systematized delusion Group of elaborate delusions related to a single event or theme.

tactile hallucination Hallucination primarily involving the sense of touch. Also called *haptic hallucination*.

tangentiality Oblique, degressive, or even irrelevant manner of speech in which the central idea is not communicated.

tension Physiological or psychic arousal, uneasiness, or pressure toward action; an unpleasurable alteration in mental or physical state that seeks relief through action.

terminal insomnia Early morning awakening or waking up at least 2 hours before planning to. *Compare* **initial insomnia; middle insomnia**.

thought broadcasting Feeling that one's thoughts are being broadcast or projected into the environment. *See also* **thought withdrawal**.

thought disorder Any disturbance of thinking that affects language, communication, or thought content; the hallmark feature of schizophrenia. Manifestations range from simple blocking and mild circumstantiality to profound loosening of associations, incoherence, and delusions; characterized by a failure to follow semantic and syntactic rules which is inconsistent with the person's education, intelligence, or cultural background.

thought insertion Delusion that thoughts are being implanted in one's mind by other people or forces.

thought withdrawal Delusion that one's thoughts are being removed from one's mind by other people or forces. *See also* **thought broadcasting**.

tic disorders Predominantly psychogenic disorders characterized by involuntary, spasmodic, stereotyped movement of small groups of muscles; seen most predominantly in moments of stress or anxiety, rarely as a result of organic disease.

tinnitus Noises in one or both ears, such as ringing, buzzing, or clicking; an adverse effect of some psychotropic drugs.

tonic convulsion Convulsion in which the muscle contraction is sustained.

trailing phenomenon Perceptual abnormality associated with hallucinogenic drugs in which moving objects are seen as a series of discrete and discontinuous images.

trance Sleeplike state of reduced consciousness and activity.

tremor Rhythmical alteration in movement, which is usually faster than one beat a second; typically, tremors decrease during periods of relaxation and sleep and increase during periods of anger and increased tension.

true insight Understanding of the objective reality of a situation coupled with the motivational and emotional impetus to master the situation or change behavior.

twilight state Disturbed consciousness with hallucinations.

twirling Sign present in autistic children who continually rotate in the direction in which their head is turned.

unconscious 1. One of three divisions of Freud's topographic theory of the mind (the others being the conscious and the preconscious) in which the psychic material is not readily accessible to conscious awareness by ordinary means; its existence may be manifest in symptom formation, in dreams, or under the influence of drugs. 2. In popular (but more ambiguous) usage, any mental material not in the immediate field of awareness. 3. Denoting a state of unawareness, with lack of response to external stimuli, as in a coma.

undoing Unconscious primitive defense mechanism, repetitive in nature, by which a person symbolically acts out in reverse something unacceptable that has already been done or against which the ego must defend itself; a form of magical expiatory action, commonly observed in obsessive-compulsive disorder.

unio mystica Feeling of mystic unity with an infinite power.

vegetative signs In depression, denoting characteristic symptoms such as sleep disturbance (especially early morning awakening), decreased appetite, constipation, weight loss, and loss of sexual response.

verbigeration Meaningless and stereotyped repetition of words or phrases as seen in schizophrenia. Also called *cataphasia. See also* **perseveration**.

vertigo Sensation that one or the world around one is spinning or revolving; a hallmark of vestibular dysfunction, not to be confused with dizziness.

visual agnosia Inability to recognize objects or persons.

visual amnesia *See* **neurological amnesia**.

visual hallucination Hallucination primarily involving the sense of sight.

waxy flexibility Condition in which person maintains the body position into which they are placed. Also called catalepsy.

word approximation Use of conventional words in an unconventional or inappropriate way (metonymy or of new words that are developed by conventional rules of word formation) (e.g., “handshoes” for gloves and “time measure” for clock); distinguished from a *neologism*, which is a new word whose derivation cannot be understood. *See also* **paraphasia**.

word salad Incoherent, essentially incomprehensible mixture of words and phrases commonly seen in far-advanced cases of schizophrenia. *See also* **incoherence**.

xenophobia Abnormal fear of strangers.

zoophobia Abnormal fear of animals.

SIGNS AND SYMPTOMS REVISITED

Descriptions of signs and symptoms in psychiatry have remained fairly constant over the years; however some terms fall in and out of favor. In the various editions of *Diagnostic and Statistical Manual of Mental Disorders* (DSM), for example, some terms have been retained and others omitted, and some terms are not common to DSM and the *International Classification of Diseases* (ICD)

The fourth edition of DSM (DSM-IV) eliminated the diagnosis of organic mental disorder in an attempt to indicate that all mental disorders may have a biological basis,

or medical cause. The diagnosis of organic mental disorder is called “delirium, dementia, and amnestic and other cognitive disorders” in DSM-IV. The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), however, retains the diagnostic category of organic mental disorders.

Neurasthenia is omitted from DSM-IV but is retained in ICD-10. Although a significant number of patients diagnosed with neurasthenia can also be classified as having a depressive or anxiety disorder, many patients cannot, and ICD-10 reflects that fact.

DSM-IV eschews the term “psychogenic.” Nevertheless, it appears in ICD-10 to refer to the fact that life events or difficulties play an important role in the genesis of many psychiatric disorders. Similarly, DSM-IV has eliminated the term “neurosis,” which is used in ICD-10. The author regards both as useful terms that should be retained.

The grouping of signs and symptoms with diseases, disorders, or syndromes may be influenced by social traditions, prevalent customs or philosophies, and even unconscious determinants such as the classifier's view of the world. Over 35 years ago, Karl Menninger anticipated the structure of the mathematical device currently in use in DSM-IV. He wrote: “If the patient has, let us say, five symptoms, one can look up each of these symptoms and find which disease is so characterized under all five headings. Then, *voila*! the diagnosis!” Menninger suggested that the trend toward tabulating disease states was antithetical to understanding the person experiencing the illness and deemphasized the compassionate approach toward the patient that is the hallmark of psychiatry. That mathematical approach is reflected in the algorithms and decision trees used in DSM-IV and in the various computer programs that record signs and symptoms to provide a diagnosis.

All those who work with the mentally ill must consider that a description of a sign and symptom is a fact that can be validated by a group of observers. It is not emotive. A description of signs and symptoms is the science of psychiatry; the skill of the observers and their creative imagination and ability to empathize is the art of psychiatry.

SUGGESTED CROSS-REFERENCES

Section 7.2 describes the psychiatric report and medical record. Chapter 8 on clinical manifestations of mental disorders, provides examples of signs and symptoms in psychiatric disorders. Section 9.1 describes the classification of mental disorders.

SECTION REFERENCES

*Andreasen NC: The clinical assessment of thought, language, and communication disorders: I. The definition of terms and evaluation of their reliability. Arch Gen Psychiatry *36:*1315, 1979.

*Ayd FJ: *Lexicon of Psychiatry, Neurology, and the Neurosciences*. Williams & Wilkins, Baltimore, 1995.

Bruno FJ: *Psychological Symptoms*. Wiley, New York, 1993.

*Campbell RJ: *Psychiatric Dictionary*, ed 7. Oxford University Press, New York, 1995.

de Vries MW: Recontextualizing psychiatry: Toward ecologically valid mental health research. Transcult Psychiatry *34:*185, 1997.

*Etter HS: Doctor discontent. N Engl J Med *340:*649, 1999.

*Ginzberg E: The uncertain future of managed care. N Engl J Med *340:*144, 1999.

*Jaspers K: *General Psychopathology*. Johns Hopkins University Press, Baltimore, 1997.

*Kaplan HI, Sadock BJ: *Comprehensive Glossary of Psychiatry and Psychology*. Williams & Wilkins, Baltimore, 1991.

Kaplan HI, Sadock BJ: Typical signs and symptoms of psychiatric illness. In *Comprehensive Textbook of Psychiatry*, ed 6. HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Kutchins H, Stuart KA: *Making Us Crazy—DSM: The Psychiatric Bible and the Creation of Mental Disorders*. Free Press, New York, 1997.

Meninger K: *The Vital Balance*. Viking, New York, 1963.

World Health Organization: *Lexicon of Psychiatric and Mental Health Terms*, ed 2. World Health Organization, Geneva, 1994.

7.4 CLINICAL NEUROPSYCHOLOGY AND INTELLECTUAL ASSESSMENT OF ADULTS

REX M. SWANDA, PH.D., KATHLEEN Y. HAALAND, PH.D., AND ASENATH LARUE, PH.D.

Clinical neuropsychology is a specialty in psychology that examines the relationship between behavior and brain functioning in the realms of cognitive, motor, sensory, and emotional functioning. In general, the clinical neuropsychologist integrates the medical and psychosocial history with the reported complaints and the pattern of performance on neuropsychological procedures to determine whether results are consistent with a particular area of brain damage or a particular diagnosis. Although neurological syndromes are often the focus of referrals, the neuropsychological examination also has a valuable place in diagnosing and treating behavioral symptoms that are associated with other medical, psychological, and psychiatric conditions.

NEUROPSYCHOLOGY AND OTHER DISCIPLINES

The unique contributions of neuropsychology can be clarified by examining its relationship to the closely allied disciplines of clinical psychology, behavioral neurology, and neuropsychiatry.

Clinical Psychology Clinical neuropsychology is recognized by the American Psychological Association as a distinct specialty area, with board certification through the American Board of Clinical Neuropsychology under the auspices of the American Board of Professional Psychology. It is primarily differentiated from general clinical psychology by its focus on thorough and extensive evaluation of a broad range of cognitive and emotional factors and their potential relationship to brain damage. There is considerable overlap between the two areas in the approach to assessment, chiefly characterized by reliance on the psychometric foundations of reliability, validity, and normative standards to define behavioral symptoms and complaints objectively. Emotional factors are the province of both. Both evaluate cognition, although the clinical psychologist generally focuses on issues involving general intellectual, academic, and vocational skills rather than neurological factors. Both clinical psychologists and neuropsychologists are involved in treatment, which can include psychotherapy as well as cognitive retraining to remediate deficits. Psychoeducational functions are also served by both fields in discussing symptoms, assessment results, and their implications with patient, caregivers, and other health professionals.

Behavioral Neurology and Neuropsychiatry The medical specialties of behavioral neurology and neuropsychiatry overlap considerably with clinical neuropsychology. However, the neuropsychologist's unique contribution to the assessment of cognitive deficits lies in the theoretical background in cognitive psychology as well as psychometrically based methods, which use standardized instruments with age- and education-based normative data whenever possible. In contrast, the behavioral neurologist and neuropsychiatrist are more likely to use a mental status examination to identify cognitive deficits, relying on internal norms, which are based upon the individual clinician's extensive experience with the examination procedures or pathognomonic signs (e.g., dysfluent speech, visual neglect). The behavioral neurologist is trained to diagnose a broad spectrum of neurological disorders, the neuropsychiatrist's focus lies in diagnosing neurological features of psychiatric patients, and the neuropsychologist focuses on the cognitive and behavioral manifestations of a broad range of disorders that can produce cognitive impairment. Ideally, all three work together to provide complementary perspectives to a patient's workup.

HISTORY

Clinical neuropsychology has its roots in psychology, neurology, and psychiatry. The groundbreaking work of Pierre Broca and Carl Wernicke in the nineteenth century first suggested that complex functions such as speech and auditory comprehension could be localized to particular areas in the left hemisphere. In 1909, Karl Brodmann published a “cytoarchitectonic” map of the cerebral cortex (Fig. 7.4-1), based on different histological patterns of cells in various parts of the cortex. Over the years, this representational map has proved to have value in identifying functional differences across distinctive cortical regions and has become a standard reference for identifying cortical areas. Views of functional localization were repopularized in the 1960s by Norman Geschwind, who emphasized the importance of connections among different parts of the brain in producing complex behavior. Arthur Benton's Iowa school developed a series of carefully constructed psychometric tests, based on concepts that had originally been identified by behavioral neurology to assess specific deficits relative to the normal population. At about the same time, Hans-Lukas Teuber defined the concept of “double dissociation,” which is regarded as the strongest evidence for localization of a particular function. This notion is based on observations of mutually exclusive brain-behavior relationships, such that damage affecting region 1 produces performance deficits in test A but not in test B, while damage affecting region 2 produces performance deficits in test B but not in test A. Aleksandr Luria, the Russian neuropsychologist, also developed notions of functional localization but used a theoretical framework that associated component cognitive processes with complex skills and their neuroanatomical correlates. In the United States, Ralph Reitan applied the psychometric standards of North America to a number of instruments for the express purpose of assessing brain damage, resulting in the widely used Halstead-Reitan Neuropsychological Test Battery.

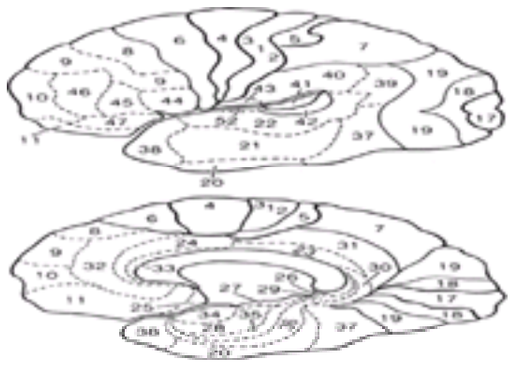

FIGURE 7.4-1 Brodmann areas of the human cortex, showing convex surface **(top)** and medial surface **(bottom)**. (Reproduced with permission from Elliott HC: *Textbook of Neuroanatomy*. Lippincott, Philadelphia, 1969.)

NEUROANATOMICAL CORRELATES

The early history of neuropsychology was driven in large part by the goal of linking behavioral deficits to specific neuroanatomical areas of dysfunction or damage. Although this emphasis served the important function of validating neuropsychological tests that are commonly used today, the localizing function of neuropsychological assessment is now less important in light of recent advances in neuroimaging techniques. Increasing knowledge in the neurosciences has also led to a more sophisticated view, in which complex cognitive, perceptual, and motor activities are controlled by neural systems rather than single structures within the brain. For instance, neuroanatomical studies have identified multiple parallel circuits that interconnect various cortical and subcortical sites, and each of these different circuits appears to subserve different behavioral functions. While the complexity of these relationships is important to appreciate, it is also necessary to understand that certain areas within both the cerebral cortex and subcortical areas, such as the basal ganglia, influence some behaviors more than others. An understanding of these brain-behavior relationships is particularly helpful when evaluating patients with focal damage, to ensure that the neuropsychological

evaluation adequately assesses relevant behavior that is likely to be associated with that area and its interconnecting pathways.

HEMISPHERIC DOMINANCE AND INTRAHEMISPHERIC LOCALIZATION

Many functions are mediated by both the right and left cerebral hemispheres. However, important qualitative differences between the two hemispheres can be demonstrated in the presence of lateralized brain injury. Various cognitive skills that have been linked to the left or right hemisphere in right-handed people are listed in Table 7.4-1. Although language is the most obvious area that is largely controlled by the left hemisphere, the left hemisphere is also generally considered to be dominant for limb praxis (i.e., performing complex movements, such as brushing teeth, to command or imitation) and has been associated with the cluster of deficits identified as Gerstmann syndrome (i.e., finger agnosia, dyscalculia, dysgraphia, and right-left disorientation). In contrast, the right hemisphere is thought to play a more important role in controlling visuospatial abilities and hemispatial attention, which are associated with the clinical presentations of constructional apraxia and neglect, respectively.

Left Hemisphere	Right Hemisphere
Aphasia	Visuospatial deficits
Right-left disorientation	Impaired visual perception
Finger agnosia	Neglect
Dysgraphia (aphasic)	Dysgraphia (spatial, neglect)
Dyscalculia (number alexia)	Dyscalculia (spatial)
Constructional apraxia (details)	Constructional apraxia (Gestalt)
Limb apraxia	Dressing apraxia
	Anosognosia

Table 7.4-1 Selected Neuropsychological Deficits Associated With Left or Right Hemisphere Damage

Although lateralized deficits such as these are typically characterized in terms of damage to the right or left hemisphere, it is important to keep in mind that the patient's performance can also be characterized in terms of preserved brain functions. In other words, it is the remaining intact brain tissue that drives many behavioral responses following injury to the brain—and not only the absence of critical brain tissue.

Language Disorders The special role of the left hemisphere in the control of language functions in most right-handed individuals has been validated in many studies. These include the results of sodium amytal testing in epilepsy surgery patients as well as the incidence of aphasia following unilateral stroke to the left versus right hemisphere. Although it is rare for right-handers to be right hemisphere dominant for language, it does occur in about 1 percent of the cases. Hemispheric dominance for language in left-handed people is less predictable. About two thirds of left-handers are actually left hemisphere dominant for language, while about 20 percent each are right hemisphere dominant or bilaterally dominant.

A number of classification systems have been developed over the years to describe various patterns of language breakdown. A common method takes into account the presence or absence of three key features: (1) fluency, (2) comprehension, and (3) repetition (i.e., intact ability to repeat verbally presented words or phrases).

Broca's Aphasia Broca's aphasia (also called nonfluent, or expressive aphasia) has traditionally been characterized by impaired verbal fluency, intact auditory comprehension, and somewhat impaired repetition. It has long been thought to be associated with damage to Broca's area (i.e., left inferior frontal convolution) or Brodmann area 44 (Fig. 7.4-1). However, more-recent neuroimaging data in stroke patients has shown that the full syndrome of Broca's aphasia, including agrammatism (telegraphic speech), is found only in the presence of more extensive damage that encompasses the suprasylvian area from Broca's area to the posterior extent of the sylvian fissure.

Wernicke's Aphasia Wernicke's aphasia (also called fluent or receptive aphasia) is characterized by intact verbal fluency, impaired comprehension, and somewhat impaired repetition. It has been associated with damage to Wernicke's area in the region of the superior temporal gyrus. The impaired ability to comprehend has a direct impact on the individual's ability to self-monitor language output and may be related to a breakdown of the syntactic structure of language. It would not be unusual for an affected patient to fluently produce unintelligible strings of utterances that might potentially be confused with the so-called jargon aphasia which is associated with schizophrenia.

Conduction Aphasia Patients with conduction aphasia demonstrate relatively intact auditory comprehension and spontaneous speech, because of the preservation of Wernicke's and Broca's areas. However, the ability to repeat words and phrases is specifically impaired, traditionally attributed to damage to the arcuate fasciculus, which interconnects Wernicke's and Broca's areas.

Global Aphasia Another common classification, global aphasia, is characterized by impairment in all three dimensions of fluency, comprehension, and repetition. In reality many patients with aphasia cannot be neatly classified within a specific system because the pattern of deficits does not exactly fit clear descriptive categories. In fact, detailed language assessment of most aphasic patients typically demonstrates deficits in all three areas, although the degree of deficit among the three areas varies.

Limb Apraxia Limb apraxia and other cognitive-motor skills deficits are more commonly seen with left than with right hemispheric damage. There are data that show that the difference in the incidence of limb apraxia after left or right hemispheric damage is not as great as with language. This suggests that left hemisphere dominance for disorders of complex movement is not as strong as for language. While limb apraxia has not traditionally been considered to be of substantial functional importance, recent data reviewed by Leslie Rothi and Kenneth Heilman also suggest that limb apraxia significantly affects rehabilitation outcome. For instance, ideomotor limb apraxia can be associated with impaired spatiotemporal execution of complex movements, which results in orientation errors, such as carving a turkey by moving the knife up and down rather than back and forth. Conceptual apraxia might result in using the wrong object to perform a movement, such as attempting to use a toothbrush to eat. Finally, sequencing errors and ideational errors can lead to disrupted activities, such as trying to light a candle before striking the match.

Arithmetic Arithmetic skills can be impaired after either left or right hemispheric damage. Left hemispheric damage, especially of the parietal lobe, produces difficulty in reading and appreciating the symbolic meaning of numbers (number dyslexia). Left hemispheric damage can also be associated with impaired conceptual understanding of the problem (anarithmetria). In contrast, the deficits in arithmetic computation that can accompany right hemispheric damage are more likely to be observed in written problems. These emerge as problems with the spatial aspects of arithmetic, such as errors resulting from hemispatial neglect, poor alignment of columns, or visual misperceptions and rotations that can result in confusion of signs for addition and multiplication.

Spatial Disorders Right hemispheric damage in right-handers is frequently associated with deficits in visuospatial skills. Common assessment techniques include drawings and constructional or spatial assembly tasks.

Visuospatial Distinctive qualitative errors in constructing block designs and in drawing a complex geometric configuration (e.g., Rey-Osterreith Complex Figure) can be seen with either right or left hemispheric damage. In the presence of lateralized damage to the right hemisphere, impaired performances often reflect the patient's inability to appreciate the gestalt or global features of a design. In the example shown in Figure 7.4-2, this is seen in the patient's failure to maintain the 2 × 2 matrix of blocks, which is instead converted into a column of four blocks. In contrast, damage to the left hemisphere commonly results in inaccurate reproduction of internal details of the design, including improper orientation of individual blocks, but the 2 × 2 matrix (i.e., the gestalt) is more likely to be preserved. Similar differences can be seen with drawings, as in the example of the Rey-Osterreith Complex Figure shown in Figure 7.4-3. The patient with right parietal damage draws isolated details of the design, while failing to convey the interrelationship of the details in the overall design configuration. In contrast, the patient with left hemisphere damage tends to maintain the global framework of the design but lose the details. Therefore, many neuropsychologists emphasize that a neuropsychological understanding of the impairment depends not just on a set of test scores, but also on a qualitative description of the type of error. This often allows the impairment to be linked to a specific

neuroanatomical region as well as enabling a better understanding of the mechanisms of the deficit for rehabilitation purposes.

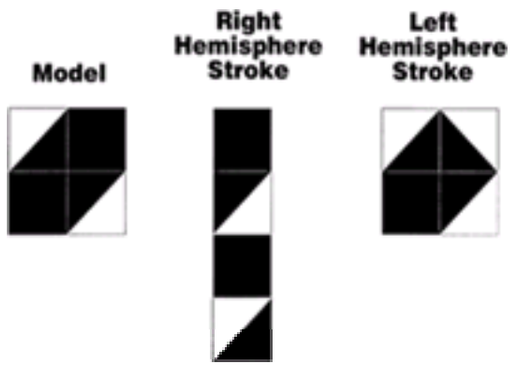

FIGURE 7.4-2 Examples of Block Design constructions seen in a right hemispheric stroke patient and a left hemispheric stroke patient.

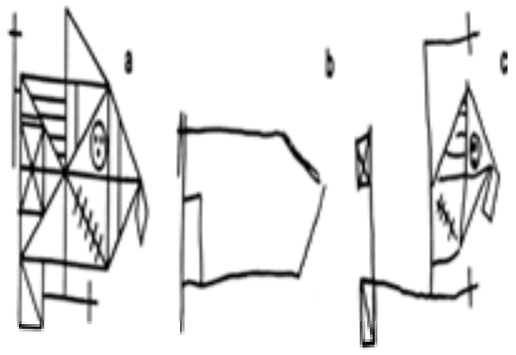

FIGURE 7.4-3 Rey-Osterreith Complex Figure **(a)** model, and drawings from memory by **(b)** a patient with left hemispheric injury and **(c)** a patient with right hemispheric injury. (Courtesy of Academic Press.)

In another example, damage to the right hemisphere tends to be associated with decreased appreciation of global features of visual stimuli, while left hemispheric damage tends to be associated with decreased analysis of local features and detail. This notion is illustrated in Figure 7.4-4, where a left hemisphere–damaged patient reproduces ambiguous figures as a simple triangle or letter “M” with no regard for the internal characters that actually make up the designs. In contrast, the local approach of a patient with right hemispheric damage emphasizes the internal details (small rectangles or letter “Z”) without appreciation for the gestalt that is formed by the internal details. This example also illustrates the important point that behavioral responses (including errors) are driven as much by preserved regions of intact brain functioning as by the loss of other regions of brain functioning.

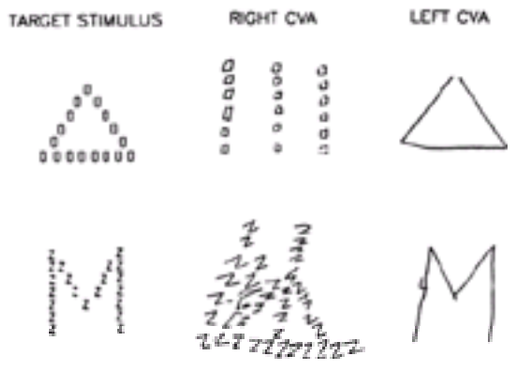

FIGURE 7.4-4 Global local target stimuli with drawings from memory by a patient with right hemispheric cerebrovascular accident (CVA) and by a patient with left hemispheric CVA. (Courtesy of Academic Press.)

Neglect Failure to detect visual or tactile stimuli or to move the limb in the contralateral hemispace is most commonly associated with right hemispheric damage. Visual neglect can be assessed at bedside with line cancellation and line bisection tasks, in which the paper is placed at the patient's midline, and the patient is asked to either cross out all the lines on the page or to bisect the single line presented. The method of double simultaneous stimulation is another standard procedure for demonstrating the deficit. Neglect has devastating functional effects and should be a standard consideration in the evaluation process. It is most frequently associated with right parietal damage, but damage to other areas within the cerebral cortex and subcortical areas can also produce this problem.

Dressing Apraxia This syndrome tends to arise in association with spatial deficits following right hemispheric damage. The resulting difficulty in coordinating the spatial and proprioceptive demands of dressing can be seen in the patient's difficulty in identifying the top or bottom of a garment, as well as right-left confusion in inserting the limbs into the garment. As a result, dressing time can be painfully protracted, and the patient may actually display more functional dependence than might otherwise be expected from assessment of simple motor or spatial skills alone.

Memory Disorders Memory complaints constitute the most common referral to neuropsychology. Thorough neuropsychological examination of memory considers the modality (e.g., verbal versus spatial) in which the material is presented and uses presentation formats that systematically assess different aspects of the information processing and storage system that forms the basis for memory. Accumulated research indicates that specialized processing of verbal and spatial memory material tends to be differentially mediated by the left and right hemispheres, respectively. In addition to interhemispheric differences in functional localization, specific memory problems can be associated with breakdown at any stage in the information processing model of memory. These stages include (1) registration of the material through *attention*, (2) initial processing of the material within *short-term memory*, (3) encoding and storage of material in *long-term memory*, and (4) *retrieval* processes, in which material moves from long-term memory storage back into consciousness. A great advantage of neuropsychological assessment is that these various types of memory problems can be readily isolated and described during the examination. Once identified, the specific nature of the deficit can then have important implications for diagnosis, treatment, and prognosis.

Encoding The initial encoding of material can be influenced by a variety of factors, including deficits in attention, language, and spatial processing abilities. Encoding is usually measured by immediate recall of information (e.g., narrative stories, designs) or by demonstrating learning of new material across multiple trials (e.g., word lists). Because attention itself can be affected by many factors, including neurologically based disorders (e.g., head injury, acute confusional state) and psychiatric disorders (e.g., depression, anxiety), this is a particularly important aspect of a proper assessment of memory.

Retrieval and Storage Deficits in recall can be associated with impaired retrieval, in which case the material is still present but not readily accessible, or it can be due to impaired storage of information. The best way to differentiate these problems is to assess recognition, typically by using test conditions that offer some version of a multiple-choice format. If the patient can demonstrate intact recognition, the problem lies in poor retrieval, but if recognition is impaired, the problem is better attributed to impaired storage of new information. This distinction is important because the functions of retrieval and storage are subserved by different neuroanatomical

structures. Impaired storage is associated with dysfunction of the medial temporal lobe–diencephalic systems, while impaired retrieval can be associated with a variety of structures, including the frontal lobes. Since it is always easier to recognize than to recall memory material, these techniques also offer a reasonable basis for assessing the possibility of malingering.

Executive Function The prefrontal lobes and their interconnections are thought to play an important role in controlling executive functions. As conceptualized by Muriel Lezak, these executive functions include volition (i.e., formulation of a goal, motivation to achieve the goal, and awareness of one's own ability to achieve the goal), planning, purposive action (response initiation, maintenance, switching, and stopping), and execution (which requires self-monitoring and self-correction as well as control of the spatiotemporal aspects of the response). Damage to the frontal lobes has also been associated with personality changes as was historically exemplified by the famous nineteenth century case of Phineas Gage, who became irresponsible, socially inappropriate, and unable to carry out plans after a tamping iron was blown through his frontal lobes. Although hemispheric differences in the control of executive functions by the frontal lobes have not been as well documented as in the parietal and temporal lobes, the Montreal Neuropsychology Group led by Brenda Milner has identified asymmetry in the control of verbal and design fluency.

The prefrontal lobes and their interconnections have been separated into the dorsolateral, orbitofrontal, and anterior cingulate divisions. Jeffrey Cummings has identified neurobehavioral roles for each of these divisions and linked them to differential neuroanatomical relationships between cortical and subcortical areas. Though this conceptual organization has not been examined in great detail, it provides a useful working model. All of these syndromes are associated with specific corticostriate circuits and are identified here by their cortical sites.

The dorsolateral prefrontal lobe circuit is associated with higher cognitive functions, such as working memory, hypothesis generation, retrieval and initation of unique words and designs, organization of information (e.g., clustering word lists on the basis of semantic category), and development of alternating and sequential motor programs. The orbitofrontal syndrome is characterized by marked personality changes, such as instability, lability, and impulsivity. The anterior cingulate circuit is associated with impaired response initiation (i.e., akinetic mutism in the most extreme case) and response inhibition. These patients have difficulty inhibiting prepotent responses on “go–no go” tests (e.g., “When I say red, squeeze my hand; when I say green, do nothing”).

PSYCHOMETRIC ISSUES

One of the strengths of neuropsychological assessment lies in the detailed quantitative record it provides for different areas of cognitive performance as well as the foundation of normative standards on which interpretations are based. Considering the normative database for any given test is particularly important to ensure that deficits seen on individual test performances cannot be explained by such demographic characteristics of the patient as age, education, or cultural considerations. Of course, this approach also depends on the care that has been expended in developing and validating specific testing techniques. Ethical guidelines for psychological assessment established by the American Psychological Association require that tests have demonstrated validity and reliability for the particular clinical application in which they are used. These guidelines serve as the basis for valid judgments to be made about change over time, based on a given individual's performance over successive examinations. This foundation also ensures that test scores can be regarded as comparable even when administered by different examiners and in different settings.

Validity Validity refers to a test's ability to measure what it purports to measure. Validity can be ascertained by (1) correlating scores from a particular test with other, well-established measures of the same skill, (2) relating test scores to diagnostic classifications or to measures of brain structure and function, or (3) relating test outcomes to aspects of everyday functions that are thought to depend on that skill (e.g., comparing pencil-and-paper tests of visuospatial skill with ability to find one's way around a shopping center). Historically, neuropsychological tests have been validated by examining their ability to differentiate brain-damaged patients from non–brain-damaged patients. More recently, they have been validated against neuroradiological or neurosurgical data and diagnostic categories (e.g., patients with multiple sclerosis vs. medical controls). Less attention has been paid to relation between neuropsychological test outcomes and everyday function, although modest correlations have been documented between test outcomes, self-care activities, and selected complex skills.

Reliability Reliability refers to the consistency of test scores. Tests are judged by their (1) internal consistency (i.e., the extent to which all items appear to be measuring the same thing), (2) by their ability to yield similar scores across multiple test occasions in normal individuals, and (3) by the consistency in scores obtained when different persons administer and score a test. Individual cognitive measures often have lower reliabilities than composite scores derived from several measures. Within a particular ability domain, some measures exhibit higher reliabilities than others. For example, within the Wechsler Memory Scale–Revised, some tests of Attention and Delayed Recall have fairly low test-retest reliabilities (e.g., r = .68 and .45, for Digits Backward and delayed recall of word pairs, respectively), but the composite index scores for Attention/Concentration and Delayed Recall have acceptably high test-retest stability (r = .93 and .84, respectively). This underscores the importance of using more than one test to measure each type of ability in a thorough neuropsychological evaluation.

Predictive Utility In selecting tests for diagnostic applications, it is important to estimate in advance the predictive utility of particular tests for particular diagnostic discriminations. An estimation of predictive utility considers the sensitivity, specificity, and positive and negative predictive values of a test. A test that yields positive findings for most persons with disease and negative outcomes for most persons without disease has both high sensitivity and specificity. Often, published values for sensitivity and specificity are based on ideal comparisons (e.g., clear cases of Alzheimer's disease versus carefully screened normal controls) and do not permit estimation of how useful a test will be in a general clinical population.

Base Rates Computation of positive and negative predictive values addresses different and potentially more practical questions: given a positive test outcome, what is the probability of having disease? And, given a negative test outcome, what is the probability of not having disease? Positive and negative predictive values are computed as the ratio of accurate positive findings to total positive outcomes and the ratio of accurate negative findings to total negative outcomes, respectively. Predictive values are judged relative to estimated base rates of illness within the referral setting. For example, if the base rate of dementia among older adults referred for psychiatric assessment is 50 percent, then any test that purports to ascertain dementia must have a predictive value above 50 percent to warrant its application.

Another aspect of base rate is related to the frequency with which abnormal neuropsychological findings are obtained in normal individuals. Figure 7.4-5 illustrates the importance of appreciating the fact that non–brain-damaged individuals without psychiatric problems may perform in the impaired range. In fact, important normative standards developed by Robert Heaton, Igor Grant, and Charles Matthews indicate that few healthy individuals complete a neuropsychological protocol without any impaired scores, while as many as 38 percent of “normals” perform in the impaired range on 6 or more discrete scores in a 40-score battery. Given these data, a naive approach to neuropsychology that considers an individual to be impaired by simply counting the number of test scores in the impaired range would be very misleading. This information is particularly important when assessing for subtle deficits in mild head injury. These data also underscore the importance of basing conclusions upon a pattern of deficits. In other words, the strongest evidence of brain injury would be a pattern of consistent impairment on several measures within the same cognitive domain, as opposed to impairment on isolated tasks that represent several different cognitive domains.

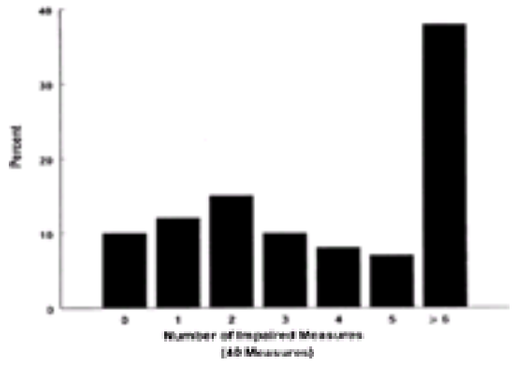

FIGURE 7.4-5 Graph showing percentages of normal individuals who score in the impaired range on 0 to 6 or more neuropsychological measures in a battery of 40 measures. (Adapted by special permission from Heaton RK, Grant I, Matthews CG: *Comprehensive Norms for an Expanded Halstead-Reitan Battery*. Psychological Assessment Resources. (PAR), Lutz, FL, 1991. Further reproduction is prohibited without permission of PAR.)

The predictive value of neuropsychological tests varies considerably for different diagnostic questions. Certain performance impairments are highly specific to deficits

in specific brain regions. However, in many neurological diagnoses, multiple brain systems can be affected, and there is overlap in affected systems across disease entities. In testing for dementia, for example, neuropsychological outcomes are more sensitive than specific, and they must be used in combination with other medical and historical data to arrive at a more specific diagnosis. This underscores the importance of interpreting neuropsychological data in the context of the patient's history.

Normative Data Considerable research in recent years has been directed toward expanding the normative database for clinical neuropsychological tests with respect to important demographic factors. Normative data are now available for many cognitive measures across a broad age range. On average, persons in their early 70s score about a standard deviation unit lower than those in their 20s on tasks that combine cognitive processing demands with requirements for speed of performance, so it is important to compare a given individual's performance on these tasks with that of age peers. Education can have an even stronger effect on many aspects of cognitive performance. For subtests of the Wechsler Adult Intelligence Scale–Revised, for example, education accounts for an average of 26 percent of the variance in performance, compared with just under 9 percent of the variance that is attributable to age. Cultural factors and linguistic background can affect familiarity with particular stimuli, such as the use of clocks to tell time, or can influence the manner in which common tasks are routinely performed, such as remembering telephone numbers. Gender differences are relatively small on most clinical neuropsychological tests, but there are a few types of tasks, such as those measuring motor speed and strength, for which it is important to consider gender. Among healthy individuals, the combined effects of age, education, and gender account for about 45 percent of the variability in intelligence test performance and for about 64 percent of the variance in the average impairment rating from the Halstead-Reitan Neuropsychological Test Battery.

Estimates of Premorbid Functioning For the neuropsychologist to conclude that deficits are present, premorbid levels of functioning must be estimated, usually on the basis of educational and occupational history as well as performance on tests that are usually minimally affected by brain damage (e.g., fund of vocabulary and general information). Impairment is usually defined on the basis of any statistically significant deviation from that level of estimated premorbid ability. As illustrated in the normal bell curve shown in Figure 7.4-6, statistically significant findings on any given set of tests would be at least one standard deviation unit above or below the mean (or average) performance established for that individual, either in comparison to established normative standards or in relation to that individual's own performance across the other tests. For example, a 32-year-old Anglo patient from the Midwest who has only 4 years of education and has worked as a laborer, may have a premorbid level of functioning estimated to lie in the borderline to low-average range, which represents the population that performs at least one standard deviation unit below average. Therefore, most test scores would be expected to hover around that level. To demonstrate the likely presence of brain injury, such a patient would have to have test scores at least two standard deviation units below average or pathognomic signs of definite impairment consistent with the medical history, such as signs of lateralized neglect or aphasia. This is another example of the importance of considering the context of other demographic and cultural factors in the interpretation of neuropsychological evaluation. How would this same patient be considered at age 81 years instead of 32 or with an advanced academic degree? Would the considerations differ if a patient with that same age, education, and work experience had only recently emigrated to the United States from Haiti?

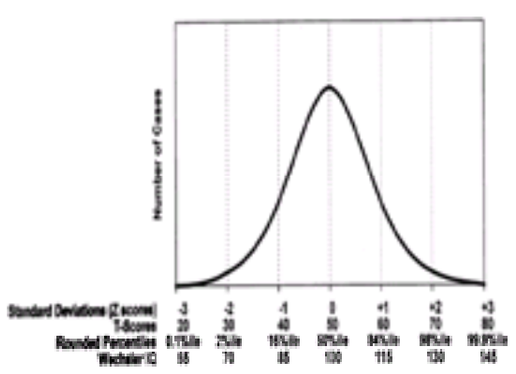

FIGURE 7.4-6 Normative (bell) curve comparing *Z*-scores, *T*-scores, percentiles, and Wechsler I.Q. scores.

Moderating Factors Performance on neuropsychological tests can be influenced by transient or situational factors. Alcohol and other substance use can impair attention, memory, and speed of cognitive processing, as can many prescription medications, especially in elderly and brain injured persons. Psychiatric disorders such as bipolar I disorder, major depressive disorder, or psychotic disorders can undermine cognitive performance, rendering formal testing invalid in severe conditions or reducing scores below expected levels in mild cases. For many clinical neuropsychological tests, the impact of drug use and psychiatric state on neuropsychological performance has been documented through research. However, the entire performance profile, combined with clinical observation and input from the patient and collaterals, must be considered in interpreting the impact of these moderating variables. An experienced neuropsychologist should be expected to specify relevant factors that could be expected to undermine or contribute to lower-than-expected performance, such as impaired vision or hearing loss, acute illness, or recent significant stressors.

GENERAL REFERRAL ISSUES

Most neuropsychological referrals are made for diagnostic purposes, to ascertain if brain impairment is present or to differentiate among different neurological or psychiatric disorders. Other important uses of testing include establishing a baseline of performance for assessing future change and planning for rehabilitation or management of behaviors affected by brain impairment. The specific methods of neuropsychological assessment reflect the individual's unique presentation of symptoms and complaints, history and development, the perspective of the neuropsychologist, and the referral question.

Level of Functioning A common referral issue involves documentation of level of functioning for a variety of purposes, including assessment of change or competence, especially in the presence of diagnoses such as dementia, stroke, and head injury.

Differential Diagnosis Like those of any other diagnostic procedure, the results of a neuropsychological examination must be interpreted in light of all available information, including the history and any associated medical factors that are documented or reported for the individual. Many neurological, psychiatric, and functional complaints have clusters of symptoms in common, with complaints of concentration or memory problems being among the most frequently reported problems. For example, impaired attention and concentration is commonly found among many groups of patients who have a history of mild closed-head injury, psychiatric issues (e.g., depression, posttraumatic stress disorder, anxiety), or any of several types of neurological disorders that are associated with subcortical involvement (e.g., Parkinson's disease, multiple sclerosis, human immunodeficiency virus [HIV]-related cognitive decline). Therefore, a pattern of cognitive impairment across several critical areas of functioning does not necessarily provide a sufficient basis for a specific diagnostic conclusion; it must be integrated with the medical and relevant psychosocial factors to arrive at a pecific differential diagnosis.

Age- or Stress-Related Cognitive Change Many middle-aged and older adults have concerns about everyday concentration and memory failures, and with heightened public awareness about conditions such as Alzheimer's disease, an increasing number of these individuals seek evaluations for these concerns. Neuropsychological testing provides a detailed, objective picture of different aspects of memory and attention, which can be helpful in reassuring healthy persons about their abilities. It also provides an opportunity for assessing undetected mood or anxiety disorders that may be reflected in cognitive concerns and for offering suggestions about mnemonic strategies that can sharpen everyday function.

Mild Traumatic Brain Injury A significant proportion of persons who have suffered a mild traumatic brain injury complain of problems with attention, memory, and mood in addition to headache or other forms of pain for many months after the injury. Many of these individuals show no clear evidence of brain injury on magnetic resonance imaging (MRI) or other neuroradiological tests. Neuropsychological testing can be a crucial method for determining the extent of objective cognitive deficit, examining the role of psychological factors in perpetuating cognitive problems, and assessing for possible residual brain impairment. Many such patients receive medications for mood or anxiety problems, possibly in combination with pain medications, and many could benefit from psychological help in managing chronic pain and psychological distress. In addition, many of these patients are in litigation, which can complicate the neuropsychologist's ability to identify the causes for impairment. Although outright malingering represents a relatively infrequent complication, more-subtle presentations of chronic illness behavior must be a prominent consideration when potential legal settlements or disability benefits are in question. This is particularly important in the case of mild head injury, when subjective complaints may be disproportionate to the objectively reported circumstances of the injury, especially since most follow-up studies of mild head injury indicate return to

neuropsychological baseline with no objective evidence of significant cognitive sequelae after 3 to 12 months following injury.

Poststroke Syndromes After the recovery from acute stroke, patients may be left with residual deficits, which can affect memory, language, spatial skills, reasoning, or mood. Neuropsychological testing can help identify areas of strength, which can be used in planning additional rehabilitation and can provide feedback on the functional implications (e.g., for work or complex activities of daily living) of residual deficits. Such assessment of functional skills can also be helpful to a psychiatrist who is managing mood and behavioral symptoms or dealing with family caregivers.

Detecting Early Dementia When concerns about a person's memory functioning are expressed by relatives instead of the patient, there is a higher probability of a neurological basis for the functional problems. Neuropsychological testing combined with a good clinical history and other medical screening tests is highly effective in distinguishing early dementia from the milder forms of declining memory and executive functioning that can be seen with normal aging. Neuropsychological evaluation is particularly helpful in documenting cognitive deterioration and differentiating among different forms of dementia, such as Alzheimer's disease and frontal lobe dementia. In addition, an incentive for early diagnosis of dementia now lies in the fact that a portion of patients with early dementia may be candidates for memory-enhancing therapies (e.g., tacrine [Cognex] or donepezil [Aricept]), and testing can provide an objective means for monitoring treatment efficacy. Other conditions that warrant neuropsychological assessment for early detection and potential treatment include HIV-related cognitive deficits and normal-pressure hydrocephalus.

Distinguishing Dementia and Depression A substantial minority of patients with severe depression exhibit serious generalized impairment of cognitive functioning. In addition to problems with attention and slowing of thought and action, there may be significant forgetfulness and problems with reasoning. By examining the pattern of cognitive impairments, neuropsychological testing can help to identify *pseudodementia*, a dementia-like condition seen in elderly patients with depressive disorders. Perhaps more common is a mixed presentation, in which depression coexists with various forms of cognitive decline, increasing the severity of cognitive dysfunction beyond what would be expected from the neurological impairment alone. Neuropsychological testing can provide a baseline for measuring the effectiveness of antidepressant therapy in alleviating cognitive and mood symptoms.

Change in Functioning Over Time In many cases, it may be most productive and clinically essential to reexamine a given patient with follow-up neuropsychological assessment after 6 months to a year, because many neurological diagnoses carry clear expectations regarding normal rates of recovery and decline over time. This is illustrated by the population of patients who complain of cognitive sequelae following mild head trauma, for whom the current literature indicates that the greatest proportion of recovery of function is likely to occur over the initial 6 months to a year postinjury. Although subtle signs of recovery can continue after that period, failure to improve following the injury—or worsening of complaints—would clearly indicate the probability of contributing psychological factors, the existence of a preexisting condition, such as dementia, or outright malingering.

Assessment of Competence Neuropsychologists are often asked to consult in determining individuals' competence to make decisions or to manage personal affairs. Neuropsychological testing can be useful in these cases by documenting areas of clear and significant impairment and identifying areas of strength and well-preserved skill. However, opinions about competence should not be based on test findings alone but must include other, more direct observations (e.g., in-home assessment, collateral interviews) of everyday function. Standards of competence vary from state to state but are generally defined by state statute. As a general rule of thumb, however, in the interests of protecting individual rights, consideration of competence is usually best approached in the narrowest possible sense, so as to infringe as little as possible on individuals' freedom to represent their own interests. Therefore, requests for neuropsychological assessment for purposes of competency should identify as specifically as possible those aspects of decision making and behavior that are of concern. Frequent concerns about competence arise with regard to the individual's ability to make (1) financial and legal decisions, (2) health care treatment decisions, and (3) decisions regarding living situations or placements, which often require institutional care arrangements. Other issues that involve higher standards of competence include ability to drive or ability to work or practice in a given profession (e.g., air traffic controller, surgeon, financial advisor). Although decisions regarding competence in any of the above areas must ultimately be made by judicial or legislatively empowered bodies (e.g., state licensing boards), the behaviorally based recommendations from a neuropsychological examination can provide crucial data on which decisions of competency are likely to be based.

Forensic Evaluation Neuropsychological evaluation of individuals in matters pertaining to criminal or civil law requires specialized knowledge beyond expertise in neuropsychology. Although neuropsychological evidence may be pertinent in criminal cases, the involvement of neuropsychologists is frequently elicited in matters involving head injury, especially in the case of mild head injury associated with motor vehicle accident. As a distinct subspecialty, this area of practice requires integration of knowledge of statutes, laws, precedents, and legal procedures as well as expertise in identifying and describing the impact of an injury or event on cognitive, emotional, and behavioral functioning.

COGNITIVE SCREENING BY OTHER HEALTH CARE PROFESSIONALS

A wide range of health care professionals are involved in assessment of cognitive and emotional status. Although many informally estimate a patient's cognitive state, research has shown that systematic use of a structured mental status examination greatly increases the accuracy of detecting cognitive impairment, much as depression rating scales increase the accurate identification of mood disorder. One of the most widely used screening instruments for documenting gross changes in mental status is the Mini-Mental State Examination (MMSE). It is widely used in general medical and geriatric settings as an initial assessment of mental status and can serve, along with a careful history, as an indicator for more precise neuropsychological evaluation. This is particularly true if performance on the MMSE deteriorates over time. While the MMSE is likely to underestimate the prevalence of cognitive deficits in well-educated older persons with early Alzheimer's disease or in younger adults with focal brain injury, it will probably overestimate the presence of cognitive deficits in persons with little education. Therefore, cutoff scores for concluding that impairment is present must be adjusted for age and education.

Although cognitive mental status examinations can be useful in screening for gross signs of cognitive impairment, they do not suffice for distinguishing specific causes of cognitive impairment and are not interchangeable with neuropsychological testing. The in-depth analysis and standardized testing approach offered by neuropsychological assessment is most useful in cases of questionable or mild impairment, when there is evidence or complaint of persistent focal impairments, or when psychiatric and neurological symptoms coexist.

DOMAINS OF FORMAL NEUROPSYCHOLOGICAL ASSESSMENT

Neuropsychological examination systematically assesses functioning in the realms of attention and concentration, memory, language, spatial skills and sensory and motor abilities, as well as executive functioning and emotional status. Overall intellectual abilities are not only described as a reflection of current abilities, but subtest variability can also be used to indicate any differences from documented or estimated levels of intellectual functioning at some earlier point in time. Psychological contributions to performance are also considered with regard to personality and coping style, emotional lability, presence of thought disorder, developmental history, and significant past or current stressors. The expertise of the neuropsychologist lies in integrating findings obtained from many diverse sources, including the history, clinical presentation, and several dozen discrete performance scores that make up the neuropsychological data. The overall summary of the examination can be expected to describe distinctive patterns in the individual's ability to process efficiently and integrate materials that differ in structure and complexity as well as modality of presentation and response.

The actual practice and procedures of clinical neuropsychology have developed in two general directions, which are effectively described in greater detail in texts by Lezak and by Grant and Kenneth Adams.

Battery Approach The battery approach, exemplified by the Halstead-Reitan Neuropsychological Test Battery, grew directly out of the psychometric tradition in psychology. This battery includes a large variety of tests that measure most cognitive domains as well as sensory and motor skills. Typically, all parts of the test battery are administered regardless of the patient's presenting problem. This approach has the advantage of identifying problems that the patient has not mentioned and that the medical history may not necessarily predict. It has the disadvantage of being time consuming (i.e., 6 to 8 hours) and originally did not include a very thorough assessment of memory or attention.

Hypothesis-Testing Approach The qualitative hypothesis-testing approach is historically best exemplified by the work of Luria and more recently developed as the Boston process approach by Edith Kaplan and her colleagues. It is characterized by detailed evaluation of areas of functioning that are related to the patient's complaints and predicted areas of impairment, with relatively less emphasis on aspects of functioning less likely to be impaired. The hypothesis-testing approach has been particularly helpful in illuminating the differential roles of the two hemispheres, as discussed above. This approach has the advantage of efficiently honing in on areas of impairment and producing a detailed description of the deficits from a cognitive processing standpoint, but it has the shortcoming of potentially overlooking

unexpected areas of deficits.

Integration of Quantitative and Qualitative Methods Historically, these two approaches were diametrically opposed, but the two approaches have converged increasingly in the 1990s so that the differences are not as apparent. These changes in clinical practice are particularly driven by the growing impact of economic factors that make it less practical to evaluate patients for long periods of time. Neuropsychologists who use either approach increasingly use screening evaluations as a first step in efficiently determining whether a diagnosis can be made with less information or whether additional testing is necessary to identify more-subtle problems. Therefore, even neuropsychologists who emphasize a hypothesis-testing approach are likely to begin with a screening protocol that efficiently assesses the major areas of neuropsychological functioning. This may or may not be followed by additional testing in selected areas that might clarify the reasons for the deficits demonstrated on the screening evaluation.

Neuropsychological Examination Techniques The past decade has seen a virtual explosion in the growth of more-sophisticated and better-standardized tests and procedures for neuropsychological evaluation. Although a comprehensive listing of tests and techniques is beyond the scope of this chapter, excellent reviews of current techniques are found in texts by Lezak as well as Otfried Spreen and Esther Strauss. Asenath LaRue also has published detailed observations about pertinent assessment issues in aging and neuropsychology. A list of examples of common neuropsychological tests and techniques is provided in Table 7.4-2.

Table 7.4-2 Selected Tests of Neuropsychological Functioning

Interview The clinical interview provides the single best opportunity for identifying the patient's concerns and questions, eliciting a direct description of current complaints from the patient, and understanding the context of the patient's history and current circumstances. Although the patient is typically the primary interview source, it is important to seek corroborating information from interviews with caregivers or family members and thorough review of relevant records, such as medical and mental health treatment and educational and employment experiences.

Intellectual Functioning Assessment of intellectual functioning serves as the cornerstone of the neuropsychological examination. The Wechsler intelligence scales have represented the traditional gold standard in intellectual assessment for many years, based on carefully developed normative standards. The scope and variety of subtests on which the summary intelligence quotient (I.Q.) values are based also provide useful benchmarks against which to compare performances on other tests of specific abilities. The latest revision of this instrument, the Wechsler Adult Intelligence Scale–III, offers the additional advantage of greatly extended age norms (ages 16 to 89) and direct normative comparisons to performances on the Wechsler Memory Scale–III. In general, the Wechsler intelligence scales use a broad set of complex verbal and visuospatial tasks that are normatively summarized as a Verbal I.Q., a Performance I.Q., and Full Scale I.Q. In the context of a neuropsychological examination, the patient's performance across the procedures provides useful information regarding longstanding abilities as well as current functioning. Most neuropsychologists recognize that the summary I.Q. values provide only a ball-park range for characterizing an individual's general level of functioning. Therefore, it is usually more appropriate and meaningful to characterize an individual's intellectual functioning in terms of the range of functioning (e.g., borderline, low average, average, high average, superior) represented by the I.Q. value rather than the specific value itself.

Careful examination of the individual's performance across the various verbal and performance subtests is crucial for observing the patient's pattern of strengths and weaknesses, as well as the degree to which these performance characteristics are consistent with the history and performance on other aspects of the neuropsychological examination. Estimation of longstanding (premorbid) intellectual abilities is as important as the measurement of current functioning, to gauge the degree to which an individual may have deteriorated. However, it may be difficult to demonstrate subtle cognitive decline in individuals who fall into either extreme of the intellectual spectrum, with premorbid abilities in the superior or borderline ranges of functioning.

With regard to the question of differential hemispheric representation, the Verbal I.Q. and Performance I.Q. have historically been reported to be associated with left and right hemispheric functioning, respectively, over the past few decades. However, more-recent views indicate that in addition to language and spatial skills, the subtests of the Wechsler intelligence scales reflect other contributions such as speed and sustained concentration. Therefore, well-trained neuropsychologists do not simply assume that a discrepancy between Verbal I.Q. and Performance I.Q. is due to unilateral hemispheric damage. Important clues to the nature of the contributing problem can often be gleaned by considering the pattern of performance across other aspects of the examination and by carefully analyzing the types of errors observed. For example, in addition to visual-spatial demands, the Performance I.Q. subsets (e.g., Block Design, Picture Arrangement) tend to be timed and depend more on novel problem-solving skills than the Verbal subtests. Clearly then, low Performance I.Q. relative to Verbal I.Q. might reflect more-general executive problem-solving deficits or generalized slowing, perhaps in association with depression or any of a number of neurological syndromes with motor symptoms (e.g., Parkinson's disease, multiple sclerosis).

Attention Attention underlies performance in virtually all other areas of functioning and should always be considered as a potential contributor to impaired performances on any tests that require vigilance or rapid integration of information. Measures of attention and concentration have traditionally been included in the Wechsler intelligence and memory scales to assess orientation and freedom from distractibility. These procedures also provide a useful basis for anticipating the individual's ability to comprehend, process information, and otherwise engage in the assessment process. Digit Span requires patients to repeat increasingly longer strings of digits as a way of assessing ability to process relatively simple information, while Digit Span Backward reflects more-complex simultaneous processing and cognitive manipulation demands. The Mental Control subtest includes measures that require serial addition and recitation of automatic (or overlearned) sequences of numbers or letters. More-demanding specialized techniques have been developed that are highly sensitive to subtle deficits in rapid simultaneous processing. For example, in Dorothy Gronwall's Paced Auditory Serial Addition Test patients successively sum pairs of numbers at increasingly faster rates of presentation, adding each new digit to the immediately preceding digit. Other procedures for assessing attention and sustained concentration can be found in the form of pencil-and-paper measures of digit, letter, or symbol cancellation, and various versions of computer-assisted measures of attention and vigilance.

Memory Complaints of memory problems constitute one of the most common reasons for referral to neuropsychology. As described above, the neuropsychologist uses an information-processing approach to assess memory problems that might involve difficulty with encoding, retrieval, or storage of new information. The Wechsler Memory Scale–III is the latest revision of a widely used battery of subtests with several measures of attention, memory, and new learning ability. The original version, developed by David Wechsler in 1945, was adapted in the 1970s by Elbert Russell to include immediate and delayed recall comparisons and was supplemented by the development of older-age norms from several laboratories. A revised version of the battery, published in 1987, formalized these modifications and added five summary indexes: Verbal and Visual Memory, General Memory, Attention/Concentration, and Delayed Memory. The current version, published in 1997, offers additional advantages by extending the normative comparisons to age 89, adding procedures for standardized assessment of recognition (to determine if impaired recall performance is due to impaired retrieval or storage), and linking normative comparisons to performances on the Wechsler Adult Intelligence Scale–III.

Many clinicians choose not to administer the entire Wechsler Memory Scale and instead use particular procedures for which standardized normative data are available. For basic assessment of memory encoding and recall, Logical Memory assesses the individual's ability to recall two different narrative stories immediately after presentation and again following a 30-minute delay. The difference between the amount of material recalled immediately and following the delay constitutes a sensitive measure of memory retention. Visual Reproduction requires immediate and delayed recall of increasingly complex geometric designs, each of which is presented for 10 seconds prior to being drawn by the patient. Retention of the material following a similar 30-minute delay interval can be measured against standardized normative data. The Wechsler Memory Scale–III also includes a working memory index, which may be useful for further differentiating memory impairment due to encoding problems from those due to retention problems.

A well-standardized example of a verbal list-learning technique is the California Verbal Learning Test. This procedure uses five learning trials of a 16-item shopping list that is made up of four exemplars from each of four semantic categories (fruits, tools, spices, and clothing). Aside from measuring several aspects of learning and memory, this test also yields rich qualitative data about the patient's use of higher-order conceptual strategies and the efficiency of verbal learning. Proactive interference is examined by a single free-recall trial of a second list of words. Free and cued recall of the original list after both short and long delays, as well as a recognition trial, further examine the memory processes of storage, retention, and retrieval. Published norms extend to age 80 for males and females. Administration in clinical practice typically requires 20 to 25 minutes, not counting a 20-minute delay interval, so this task may be too demanding for more-impaired individuals. However, recent findings suggest that it might serve as a sensitive early cognitive marker for dementia of the Alzheimer's type.

An alternative list learning procedure to consider for older patients is Paula Fuld's Object Memory Test, in which the patient must identify (tactually or visually) 10 common objects that are hidden in a bag, then attempt to learn the list over a series of five “selective reminding” trials. This is a specialized technique for measuring memory retention and storage, in which patients are asked to learn a list of items over five trials. At the end of each recall trial, subjects are selectively reminded of only those words that were omitted. This test is especially useful for patients with hearing impairment, and the normative standards afford limited comparisons with older individuals who are either community-residing or in assisted living arrangements. The Weschler Memory Scale–III also includes a list-learning task that may be useful for screening purposes, especially for elderly persons or patients whose premorbid functioning is estimated to be below average.

Language Assessment of language examines both expressive abilities and comprehension. However, most neuropsychologists screen for language impairment rather than administer an extensive formal language assessment battery, such as the Boston Diagnostic Aphasia Examination. Expressive language is commonly assessed by measures of verbal fluency that require rapid generation of words within semantic (e.g., names of animals) and phonetic categories (e.g., words beginning with specified letters of the alphabet). Another measure of expressive language, the Boston Naming Test–Revised, provides a standard 60-item measure of visual confrontation naming ability. Standardized evaluation of comprehension often uses procedures that systematically assess one-, two-, and three-step commands in a format such as the Token Test or by examining comprehension of more-complex narrative material, as in the Complex Ideational subtest of the Boston Diagnostic Aphasia Examination.

Visuospatial Functions Complex visuospatial abilities can be assessed through procedures that were developed in Benton's laboratory, such as Facial Recognition and Judgment of Line Orientation. Visual constructional ability, another important aspect of the examination, taps into the person's ability to draw spatial designs or assemble two- or three-dimensional figures (Fig. 7.4-2 and Fig. 7.4-3). In addition to the significant visuospatial component, these tasks reflect contributions of executive-planning and organizational abilities. More-impaired individuals can be asked to copy simple geometric forms, such as a Greek cross or intersecting pentagons.

The widely used technique of Clock Drawing provides a surprisingly sensitive measure of planning and organization, especially for older individuals suspected of dementia. While problems involving poor organization, perseveration, and possible neglect are obvious in the drawing that is illustrated in Figure 7.4-7, more-subtle difficulties can also be detected, especially when a patient's performance is evaluated in light of premorbid expectations. Like virtually all assessment techniques, even this relatively simple procedure must be used in a standard manner. Most neuropsychologists use the same set of instructions each time, offering the patient a single sheet of unlined paper and saying, “Now I would like you to draw a clock. Put all the numbers on the clock, and set the time to 10 after 11.” Encourage the patient to use the whole page and draw a relatively large clock, so that evidence of neglect or executive deficits in organization and planning can be easily seen. Although the rich qualitative information that can be gleaned from this task may not suffice alone for diagnostic purposes, this measure provides a useful screening technique—especially for older or more seriously impaired individuals—and can serve as a basis for deciding whether to refer the patient for more-detailed neuropsychological examination.

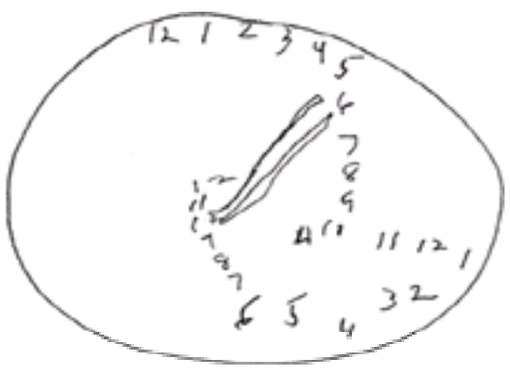

FIGURE 7.4-7 Clock drawing by a patient with vascular dementia, showing poor planning and organization, perseveration, and possible neglect.

The Rey-Osterreith Complex Figure Test is a more demanding constructional task that requires the individual to copy a highly complex geometric configuration and then draw the figure from memory. Important qualitative information having to do with problem-solving strategies can be extracted by carefully monitoring and documenting the strategy that is used in drawing the figure. Figure 7.4-2 and Figure 7.4-3 illustrate different spatial deficits seen on the Block Design and Complex Figure constructional tasks in the presence of lateralized brain damage in either the right or left hemispheres.

Sensory and Motor Functions Double simultaneous stimulation in the visual, tactile, and auditory modalities is a standard component of the Halstead-Reitan Neuropsychological Test Battery and can be useful for assessing the integrity of basic sensory functions. Grip Strength and rapid Finger Tapping are commonly used measures of motor strength and speed that are sensitive to lateralized brain dysfunction. A related measure, the Grooved Pegboard, assesses the individual's ability to efficiently integrate tactual motor dexterity, motor speed, and eye-hand coordination, but performance on this measure may be bilaterally impaired following unilateral brain injury because of its high demands for cognitive integration.

Executive Functions One of the most important aspects of the neuropsychological examination lies in the assessment of higher executive functions, which play an important role in the planning and initiation of independent activities, self-monitoring of performance, inhibition of inappropriate responses, switching between tasks, and planning and control of complex motor and problem-solving responses. Although the prefrontal lobes have long been considered important in mediating these functions, more-recent developments in the neurosciences have led to increased appreciation of the essential role played by extensive cerebral interconnections between subcortical and cortical regions of the brain.

An underlying executive component is implied to some degree in many of the techniques discussed above. For example, complex attention, as assessed on a serial addition task, depends on the ability to flexibly shift focus and sustain a cognitive set. Memory functions rely on the ability to process information efficiently, including the use of strategies to organize incoming information (e.g., clustering words in a memory list on the basis of semantic categories). In fact, memory deficits that can be attributed to retrieval failure in the information-processing model of memory are often associated with other executive deficits. However, some techniques are often used specifically as primary measures of executive functions. In the Wisconsin Card Sorting Test an individual sorts a long series of cards with designs according to various principles that can only be gleaned by incorporating feedback from the examiner. This measure, then, assesses the individual's ability to generate and test various solutions, to flexibly respond to changing circumstances, and to integrate new information with past experience to maximize the efficiency of the response. However, recent data show that deficits on this test do not depend solely on the frontal lobes. Other techniques based largely on the work of Luria are more qualitative and assess the individual's ability to inhibit inappropriate motor responses as well as the tendency to perseverate when asked to draw a series of repetitive graphomotor sequences. Figure 7.4-8 contains several examples of such sequences. In each case, the sample is presented with instructions for the patient to reproduce the sample and continuously repeat the segment across a page. Each prior sequence should be covered or removed from the patient's view before going on to the next, so that any tendency to reproduce elements from a prior sequence can properly be interpreted as evidence of perseveration.

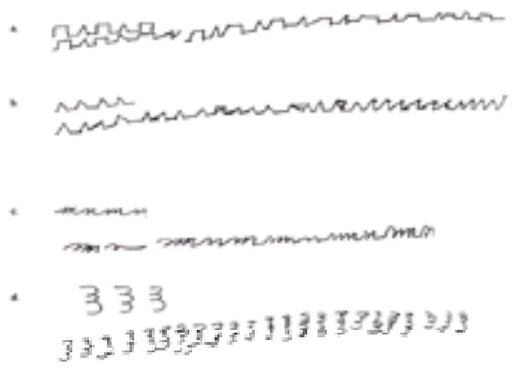

FIGURE 7.4-8 Patient reproductions of Repetitive Graphomotor Sequences, showing perseveration of elements across designs (from **a** to **b**) and within designs (**c** and **d**).

Therapeutic Discussion of Results A key component of the neuropsychological examination process is the opportunity to discuss results of the examination with the patient and family or other caregivers. This meeting can be a powerful therapeutic opportunity to educate and clarify individual and relationship issues that can affect the patient's functioning. If the patient's active cooperation in the initial examination was appropriately enlisted, the patient will be optimally prepared to invest value and confidence in the findings of the examination. At the time of the results discussion, it is useful to review the goals of the examination with the patient and supportive family or caregivers and to clarify the expectations of those who are present. Typically, these sessions include information about the patient's diagnosis, with emphasis upon the natural course and prognosis as well as compensation and coping strategies for the patient and family. Given the impact of chronic neurological disease upon the family system as well as the patient, explicit discussion of these issues is critical in maximizing adjustment to brain injury. It is equally important to relate the impact of the results to the patient's current living circumstances, future goals, and course of adjustment. It is not unusual for strong emotions and underlying tensions within family relationships to come to light in the context of honest discussion, and so the results discussion can be an important therapeutic opportunity to model effective communication and problem-solving techniques.

SUGGESTED CROSS-REFERENCES

The neuropsychological and intellectual assessment of children is discussed in Section 7.6. Section 7.5 addresses personality assessment. The biology of memory is discussed in Section 3.4. Attention-deficit disorders are discussed in Chapter 39. Communication disorders are the subject of Section 37.

SECTION REFERENCES

American Psychological Association: Ethical principles of psychologists and code of conduct. Am Psychol *47:*1597, 1992.

Benton AL, Sivan AB, Hamsher K deS, Varney NS, Spreen O: *Contributions to Neuropsychological Assessment*, ed 2. Oxford University Press, New York, 1983.

Cummings J: Frontal-subcortical circuits and human behavior. Arch Neurol *30:*873, 1993.

*Cummings J, Trimble MR: *A Concise Guide to Neuropsychiatry and Behavioral Neurology*. American Psychiatric Press, Washington, DC, 1995.

Goodglass H, Kaplan E: *Assessment of Aphasia and Related Disorders*. Lea & Febiger, Philadelphia, 1972.

*Grant I, Adams KM: *Neuropsychological Assessment of Neuropsychiatric Disorders*, ed 2. Oxford University Press, New York, 1996.

Haaland KY, Harrington DL: Hemispheric asymmetry of movement. Curr Opinion Neurobiol *6:*796, 1996.

Heaton RK: *Comprehensive Norms for an Expanded Halstead-Reitan Battery: A Supplement for the Wechsler Adult Intelligence Scale–Revised*. Psychological Assessment Resources, Odessa, FL, 1992.

Heaton RK, Grant I, Matthews CG: *Comprehensive Norms for an Expanded Halstead-Reitan Battery*. Psychological Assessment Resources, Odessa, FL, 1991.

*Heilman KM, Valenstein E: *Clinical Neuropsychology*, ed 3. Oxford University Press, New York, 1993.

LaRue A: *Aging and Neuropsychological Assessment*. Plenum, New York, 1992.

*Lezak M: *Neuropsychological Assessment*, ed 3. Oxford University Press, New York, 1995.

*Loring DW, editor: *INS Dictionary of Neuropsychology*. Oxford, New York, 1999.

Luria AR: *Higher Cortical Functions in Man*, B Haigh, translator. Basic Books, New York, 1966.

Mace NL, Rabins PV: *The 36-Hour Day*. Johns Hopkins University Press, Baltimore, 1981.

Mitrushina MN, Boone KB, D'Elia LF: *Handbook of Normative Data for Neuropsychological Assessment*. Oxford University Press, New York, 1998.

Nussbaum PD: *Handbook of Neuropsychology and Aging*. Plenum Press, New York, 1997.

*O'Reilly RC, Farah MJ: Simulation and explanation in neuropsychology and beyond. Cogn Neuropsychol *16:*49, 1999.

Rizzo M, Tranel D: *Head Injury and Post Concussive Syndrome*. Churchill Livingstone, New York, 1996.

Rothi JG, Heilman KM: *Apraxia: The Neuropsychology of Action*. Psychology Press Publishers, East Sussex, UK, 1997.

*Shea CS: *Psychiatric Interviewing*, ed 2. Saunders, Philadelphia, 1998.

*Spar JE, LaRue A: *A Concise Guide to Geriatric Psychiatry*, ed 2. American Psychiatric Press, Washington, DC, 1997.

Spreen O, Strauss E: *A Compendium of Neuropsychological Tests*, ed 2. Oxford University Press, New York, 1998.

Snyder PJ, Nussbaum PD: *Clinical Neuropsychology: A Pocket Handbook for Assessment*. American Psychological Association, Washington, DC, 1998.

Storandt M, VandenBos GR: *Neuropsychological Assessment of Dementia and Depression*. American Psychological Association, Washington, DC, 1994.

Wechsler D: *Wechsler Adult Intelligence Scale, ed 3*. The Psychological Corporation, San Antonio, TX, 1997.

Yudofsky SC, Hales RE: *The American Psychiatric Press Textbook of Neuropsychiatry*, ed 2. American Psychiatric Press, Washington, DC, 1992.

7.5 PERSONALITY ASSESSMENT: ADULTS AND CHILDREN

RUSSELL L. ADAMS, FH.D., AND JAN L. CULBERTSON, FH.D.

Numerous books have been written about personality assessment and many focus solely on one particular approach to personality assessment. *Personality* refers to an individual's enduring and pervasive personal motivation, emotion, interpersonal style, attitudes, and traits. *Personality assessment* is the systematic measurement of an individual's personality.

REASONS FOR PSYCHOLOGICAL REFERRAL

Psychological tests can be expensive and require a considerable amount of time to administer, score, and interpret. Testing is not routinely done on all psychiatric patients, as the costs would be prohibitive. Even in today's cost-containment environment, psychological testing of selected patients can be helpful from both a clinical and a cost-benefit-analysis perspective.

Assisting in Differential Diagnosis Some patients present diagnostic dilemmas, yet a diagnosis can be essential for both treatment and understanding the cause and prognosis of the disorder.

> A previously high-functioning 35-year-old female with no psychiatric family history was driving in the country with her husband. All of a sudden, without provocation and with no known drug use, she asked her husband to stop the car. She got out of the car and told him that she had received a message that that particular spot was holy and had great significance. A voice told her to get out of the car and run around in a circle approximately 120 feet in diameter. Following this episode she did not sleep for 24 hours. She became hypomanic, talked loudly, neglected personal hygiene, and experienced grandiose delusions. She felt that she was a messenger from Andre, a person who lived on that land some 3000 years ago. Her grooming and personal hygiene deteriorated, and she got violent toward others who questioned her emotional stability. The differential diagnosis includes schizophrenia and bipolar disorder. Testing could help in determining the differential diagnosis and in formulating a treatment plan.

Aiding Psychotherapy Psychological tests can be useful in psychotherapy, especially short-term, problem-centered therapy in which understanding patients and their problems must be accomplished quickly. Psychological assessment can be used in pretreatment planning, assessing the progress of therapy, and evaluating its effectiveness. Patients need to have objective information about themselves at the time of therapy if they are to go about changing themselves productively. Personality tests, particularly objective tests, allow patients to compare themselves with objective norms and evaluate the extent and magnitude of their problem. Testing also can reveal areas of the patient's life that may be problematic, but for which the patient may not have a full appreciation. Information about the patients' ability to reveal information about themselves can also be helpful. Sometimes a test report can be the subject of an early treatment session. Psychological tests may reveal considerable information concerning the patient's inner life, feelings, and images that may facilitate therapy. Psychological testing can provide baseline information at the beginning of therapy, so that repeat testing can be used to assess change during the course of therapy.

Need for Specific Information Psychological tests can often assist in specific problem areas. For example, a young college student may be unsure what to major in (“I don't know what to do the rest of my life.”) Vocational interest inventories such as the Strong Vocational Interest Inventory or the Kuder Occupational Interest Survey may help to define the student's vocational interest. The Strong Vocational Interest Inventory compares a given individual's responses with those of a large group of successful individuals in a variety of occupations. Testing results can show that a student's interests are similar, for example, to those of persons successful in occupations such as social work or public administration and dissimilar to those of engineers or scientists.

EVALUATING THE USEFULNESS OF PERSONALITY ASSESSMENT INSTRUMENTS

The quality of personality tests varies widely, from well-constructed, empirically validated instruments to psychological tests in the Sunday supplement of the newspaper or on the Internet. Evaluating the usefulness of particular psychological instruments can challenge even the well informed.

Just as the *Physician's Desk Reference* helps the physician evaluate the usefulness of medications, the Buros *Mental Measurement Yearbook* helps the psychologist evaluate the usefulness of personality instruments and other psychological and achievement tests. The *Mental Measurement Yearbook* is a reference source of critical reviews of individual tests. In these reviews, the scientific underpinnings of each test are evaluated by an expert in the field. The reviews generally address such issues as the reliability and validity of the tests, the normative group used in construction, and various empirical studies used to check the reliability and validity of the instrument. The advantages and disadvantages of the test are described. The reviews generally describe situations in which use of the test is appropriate. Practical information concerning the test publisher, length of time of administration, and other considerations is also included. (The *Mental Measurement Yearbook* is available in most university libraries and can be ordered through the book's publisher.)

Normative Sample When one evaluates a personality assessment instrument, certain factors concerning the samples used to construct the test should be considered. Basic issues such as the size and representativeness of the sample used to construct the test must be evaluated. A sample of 30 persons with a given diagnosis in a single location is seldom appropriate, because so many moderating factors affect results, such as age, education, sex, and racial or ethnic grouping. These factors frequently affect the norms and appropriate controls must be included for these variables in many cases.

For example, the Minnesota Multiphase Personality Inventory–2 (MMPI-2), a well-constructed instrument, had approximately 2900 subjects tested initially. However, approximately 300 subjects were eliminated because of test invalidity or incompleteness of needed information. The final group comprised 2600 community volunteers, 1462 females and 1138 males. The MMPI-2 normative group was selected using census data as a guide and was broadly representative in terms of region of the country, racial composition, age, and formal education. Despite efforts to control for moderating variables, the sample was slightly overrepresentative in higher education levels. Obviously, the cost of obtaining a large census-based normative sample is enormous, and few tests can meet these challenging criteria. Nevertheless, many tests that fall short of the national sample (and most do) can be useful in the hands of an experienced clinician who is sensitive to the problem and makes appropriate clinical adjustments in interpreting the results.

Test Characteristics To be useful, any psychological test must be completed, in its entirety, by the intended test taker. If the questions are offensive or difficult to understand, the person taking the test may not complete all items. These omissions can create problems, especially when normative tables are used to interpret results. An instrument should be short enough to complete in a relatively short period of time. The Beck Depression Inventory or the Geriatric Depression Scale can easily be completed in 5 to 10 minutes. Because of its short length, the test can be given repeatedly to a depressed patient in treatment to monitor changes over time. The MMPI-2, on the other hand, takes approximately 1 to 1½ hours to complete. This test is not appropriate for repeat testing over a short period of time. Most tests are easily scored and compared with normative tables.

Test Validity Perhaps the most important characteristic in evaluating the scientific merit of a given personality test is the validity of the instrument. Does the test measure what it purports to measure? If a test is designed to measure depression, does it indeed measure depression? Although validity may seem to be a simple issue to address, it can become complex, especially when one is trying to measure such characteristics as self-esteem, assertiveness, hostility, or self-control. The test developer must first operationally define the personality characteristic in question. A number of different approaches have been used to measure validity,

including face validity, logical content validity, construct validity, criterion validity, factor validity, among others.

Face Validity Face validity refers to the content of the test items themselves. In other words, do the items appear to measure what they purport to measure? For example, does a depression scale ask questions concerning sadness, guilt, sleep disturbance, weight loss, helplessness, hopelessness, or suicidal thoughts? One problem with face validity is that professionals differ in their subjective appraisal of individual items. One psychologist may think a given item does indeed measure the concept under consideration, while other professionals may not. To address these differences in definition, the test developers may use an approach known as *logical content validity*. With this approach, in which the manual describes the procedure used to select items for the test, results will not depend significantly on the subjective judgment of any one professional. For example, in developing the Taylor Manifest Anxiety Scales, a number of different psychologists were asked to review items from the Minnesota Multiphasic Personality Inventory (MMPI) that seemed to measure a formal definition of manifest anxiety, and only those items agreed on by this group of psychologists were included. Many creative approaches have been used to demonstrate the logical content validity of various tests.

Criteria and Construct Validity While face validity refers to the degree that test items appear on the surface to measure what the instrument, as a whole, purports to measure, criterion validity uses data outside the test itself to measure validity. For example, if a test was designed to measure hypochondria, one would expect a patient with a high score to have more visits to the physician's office, complain of more physical symptoms, and use prescribed and over-the-counter medications more extensively. These criteria are objective, are external to the test itself, and can be independently measured. Scores on the hypochondria instrument can be statistically correlated with a number of separate objective criteria such as those mentioned above to determine whether the test measures the theoretical construct of concern (*construct validity*).

Concurrent and Predictive Validity To determine test *concurrent validity*, these external measures are obtained when the test is given to the sample of subjects. Thus, the concurrent validity of the test reveals that at a given time, those with high scores on the test were more likely to manifest the behavior than those with low scores. Occasionally, however, a test developer is interested in predicting future events. For example, a test designed to predict marital satisfaction among engaged couples may use as an external criterion the couples' future marital satisfaction and marital status. Thus the external criterion must be obtained a number of years later for the test to have predictive validity. Because of the difficulty in obtaining such information several years later, predictive validity is seldom available in most personality tests, although when such data are available they can be valuable.

The *discriminant validity* of a test tells whether or not the test can discriminate between known groups of patients at a given time. For example, can a test of depression statistically discriminate nondepressed subjects from subjects who suffer from independently diagnosed major depressive disorder, adjustment disorder with depressed mood, and dysthymic disorder? Can a measure of depression statistically discriminate mild, moderate, and severe major depressive disorder?

Factor Validity Factor validity uses a multivariate statistical technique known as factor analysis to determine whether certain major groups of items on a given test cluster together empirically. For example, on a personality test measuring depression, do items concerning vegetative symptoms tend to covary? Are there other clusters of items on the depression test measuring cognitive problems, mood disturbance, and social withdrawal?

Reliability To be useful, a test must not only be valid, but must also be reliable. *Reliability* refers to the degree to which a test consistently measures what it purports to measure. The key word here is *consistently*. For example, a patient with agoraphobia takes a new agoraphobia test today and takes the same test tomorrow. However, on second testing the scores change markedly, although the patient had not received any treatment or undergone any change in the underlying agoraphobia. This test is not useful because it does not measure the trait in question consistently. The test has poor reliability. Several means of checking reliability exist, including test-retest reliability, internal consistency reliability, and parallel form reliability.

Test-Retest Reliability Test-retest reliability is obtained by administering the same test to a group of subjects on two occasions and statistically correlating the results. For a test to be useful, the correlation coefficient should be at least .80 if the two tests were administered within 2 weeks of each other and if the trait in question is stable.

Internal Consistency Reliability Another way to determine internal consistency reliability is to divide a given test into two equal parts and correlate the two halves for the test with each other statistically. This technique determines the *split-half reliability* of a test. The first half of the test should be highly correlated with the second half of the test if the test is measuring what is purportedly being measured consistently. Alternatively, the odd-numbered items could be correlated with the even-numbered items (odd-even consistent reliability). When a test is split into two parts and the scores obtained correlate with each other, the resulting correlation coefficients statistically underestimate the true reliability of the test because the magnitude of the correlation is influenced by the number of items in the test. Hence, if there were half that number of items in the original test, the results would be underestimated. The Spearman-Brown formula is used to correct for this underestimation. Again, a reliability coefficient of .80 to .85 is needed to demonstrate usefulness in most circumstances. However, the higher the reliability as measured by the correlation coefficient, the better the test instrument. No test is perfectly reliable. The highest possible correlation coefficient is 1.0 and the lowest is 0, assuming no negative correlation.

To determine how internally consistent a given test is, a test developer also may use a statistical technique called Cronback's Coefficient Alpha. With this statistic, one can determine the average correlation among all the items on the test. The higher the average correlation, the better the reliability.

Parallel Form Reliability Sometimes two separate forms of the same test are needed. For example, if taking a test at one time by itself influences patients scores the second time they took the same test, then parallel forms of the tests are needed. *Parallel forms* of a test measure the same construct with different items. The correlation coefficient between the two parallel forms of the same test is computed to ensure that the test does in fact measure the same construct. Such parallel form reliability should be .90 or higher.

Use of Standard Error of Measurement to Assess Reliability The usefulness of a given test can be assessed by examining the test's standard error of measurement, which should be included in the test's manual. The *standard error of measurement* (SEM) is a single statistic used to estimate the score of a given patient if the patient took the same test again within a short period of time. For example, a patient takes a given test today and obtains a score of 90. The SEM on the test is 3. An SEM of 3 together with an individual's score of 90 on the test means that if that individual taking that same test again tomorrow would score within ± 1 SEM of the original test performance (e.g., + or – 3 points) 68 percent of the time. Thus, the patient in question should score between 87 (90 – 3) and 93 (90 + 3) on the test 68 percent of the time when taking the same test again. The same individual would score ± 2 SEM units (e.g., 2 × 3 = 6) 95 percent of the time on second testing.

In the above example, the patient with an original score of 90 on a given test would score between 84 (90 – 6) and 96 (90 + 6) on second testing 95 percent of the time. Other things being equal, the smaller the SEM relative to the standard deviation of the tests, the better. SEM is calculated by the test developer based on the tests' reliability.

ASSESSMENT OF PERSONALITY IN ADULTS

Objective Personality Tests Objective personality tests have a rather straightforward approach. Patients are usually asked specific standard questions in a structured written or oral format. All patients are typically asked the same questions. The data obtained from a given patient are compared with similar data obtained from the normative group. The degree to which the patient deviates from the norm is used in the interpretation. The patient's responses are scored according to agreed-upon criteria, and the scores are then compared with normative tables and often converted to standardized scores, percentiles, or both. The MMPI-2 is an example of an objective personality test.

Literally hundreds of different objective personality instruments exist. Table 7.5-1 contains a representative sample of objective personality tests, with a concise description of the test and its strengths and weaknesses. The MMPI-2 is the most frequently used objective personality test by far.

Table 7.5-1 Objective Measures of Personality in Adults

Focusing on the MMPI-2 should orient the reader to both objective personality assessment and objective personality development. This is not meant to imply that the MMPI-2 is the only valid and reliable objective personality instrument or that other objective personality tests are not as useful for a particular purpose. Many useful objective personality instruments exist.

MMPI-2 Two excellent books on the MMPI-2 are James N. Butcher's *The MMPI-2 in Psychological Treatment* and John R. Graham's *MMPI-2 Assessing Personality and Psychopathology*. The MMPI-2 consists of 568 true-or-false questions concerning a wide variety of issues. The patient completes the test with a pencil. The MMPI-2 is relatively easy to administer and score and takes approximately 1 to 1½ hours for most patients to complete. It requires only an eighth grade reading comprehension. The test provides a number of internal checks for measuring the patient's cooperation. The patient is simply asked to answer true and false to such questions as “I have a good appetite” or “No one seems to understand me.”

Scoring the MMPI-2 involves adding up the number of responses on numerous scales and comparing the results with certain normative information. Interpretation of the MMPI-2 is more straightforward than that of many other tests because of the vast amount of empirical information available concerning the MMPI-2. More than 12,000 articles and books have been published on this test to date.

There are over 115 translations of the MMPI-2. The test is also cost-effective because relatively little professional time is involved in the administration and scoring. The patient is simply given instructions and then completes the test independently in the professional's office. In certain situations, the MMPI-2 can be administered on the computer. A computer can also plot test profiles and compute various supplemental scales. The number of available specialized scales increases the usefulness of the MMPI-2 with various populations (e.g., patients with posttraumatic stress disorder, college students who are maladjusted, or prisoners with violent tendencies).

Questions cover a wide range of different areas. When a patient takes the MMPI-2, questions are not grouped in any particular order. Various items in the test can be selected, sorted, and analyzed according to various criteria as an aid in interpretation. The most common scales of the MMPI-2, including the validity scales, clinical scales, content scales, supplemental scales, and critical items are discussed below.

VALIDITY SCALES Patients can approach a mental status interview in a variety of ways. They can be painfully honest and straightforward with the interviewer, revealing their innermost aspirations and fears. They may be somewhat distrusting of the interviewer and may try to slant the content of the interview to obtain the outcome they want. They may be disingenuous or even attempt to be deceitful by malingering. Obviously, the information obtained by psychiatrists or other mental health professionals in an interview may well be colored by the patient's approach to the interview. Some patients by their very nature or reserve may use denial successfully or have little insight into their feelings and thoughts.

When patients are administered personality tests such as the MMPI-2, they approach the test as they approach an interview. One advantage of the MMPI-2 is that the test has been devised to assess different test-taking attitudes (e.g., fake bad, fake good, social desirability responding). Use of the validity scales yields information that helps the clinician interpret the remaining scales of the MMPI-2. Is the patient overly defensive? Is the patient attempting to fake good or fake bad? Does the patient maintain a reasonable balance between self-disclosure and excessive defensiveness?

Lie Scale The Lie Scale consists of 15 items designed to measure general frankness. The items include questions concerning personal characteristics that are highly desirable in our culture but infrequently practiced (e.g., “I have never failed to tell the truth,” “Once in a while I think of things too bad to talk about” or “I do not always tell the truth”). Patients who score high on the lie scale may be psychologically naive or defensive. They may want to be seen in a positive light or they may have difficulty admitting common human frailties.

F-Scale The F-Scale is one of the longest scales, consisting of 60 items that received positive responses from 10 percent or fewer of the normative samples. This scale includes such items as “Evil spirits possess me at times” and “Most any time I would rather sit and daydream than do anything else.”

Items on the F-Scale have quite diverse content and may indicate severe psychopathology such as hallucinations or delusions. Most people in the normal sample responded to fewer than five of these items. Individuals who obtain exceptionally high scores on this scale are often either extremely pathological or trying to claim mental illness falsely. As with all personality tests, interpretation depends greatly on the setting in which the test was given, as base rates for different settings and circumstances differ. For example, patients who have been seen for court sanity evaluations have a different base rate of symptom magnification than patients seen for adjustment problems in an outpatient clinic.

Backpage Infrequency (Fb) Scale A few patients, particularly those with passive aggressive tendencies, will start to complete the MMPI-2 and toward the end of the test may respond to the items in a random fashion. The patient may simply want to finish the test quickly and leave the testing situation. The Fb scale for the MMPI-2 is designed to pick up this tendency.

Variable Response Inconsistency (VRIN) Scale The VRIN Scale consists of 67 pairs of items. Each question of the pair probes for the same or similar information in a slightly different way. For example, the test questions asked are similar to the following: (1) I sleep very poorly at night and (2) I wake up fully refreshed each morning. A patient who responds “true” to question 1 would normally respond “false” to the question 2. Failure to do so reflects inconsistency.

CLINICAL SCALES The clinical scales are generally considered the substance of MMPI-2 interpretations. The original clinical scales have an empirical base in the sense that the items were selected because they differentiated between specific clinical groups and the normative control group. The control group for the original MMPI developed in the 1940s consisted of 1500 persons who were visitors to the University of Minnesota hospitals and other normal groups. Clinical groups varied in size, but for the most part contained at least 50 subjects. For example, the schizophrenic scale was developed with patients with different types of schizophrenia, and the depression scale was developed with a sample of psychiatric patients with depression, including some with manic-depressive illness (bipolar I disorder). Table 7.5-2 briefly describes the original sample groups. Although most clinical scales were developed by using a specific clinically defined group, researchers with the MMPI soon learned, for example, that a given patient with schizophrenia did not simply score high on the schizophrenia scale and low on all other scales. A schizophrenic patient might score high on many scales but the highest score might well be on the schizophrenic scale. If all patients who scored high on the schizophrenic scales were given diagnoses of schizophrenia, a large number of false positives would result. To address this confusion, clinicians began to refer to the clinical scales by a number (0 to 9) rather than by the diagnostic entity that underlies its development.

Table 7.5-2 MMPI Scale Development

COMBINING SCALE ELEVATIONS IN CODE TYPES As has been illustrated, interpreting single-scale evaluations on the MMPI-2 is problematic; therefore, most clinicians use a combination of scale elevations in interpretation. This procedure usually involves using the patient's highest two clinical scales that exceed a *t*-score of 65 to provide a configured interpretation that takes into account the relation between the scales. In this approach, for example, a 1-2 or 2-1 code type would be the same. In other words, a patient who scored highest on scale 1 (Hs) and second highest on scale 2 (D) would be considered to have the same basic profile as an individual who scored highest on scale 2 (D) and second highest on scale 1 (Hs).

A 3-point code, likewise, would list the three highest scales, all of which normally would be above a T-score of 65. These scales also would usually be significantly higher than all the other clinical scales.

On the MMPI-2, uniform *t*-scores are used on clinical scales except in scales 5 and 0. *Uniform* t-*scores* are standardized scores in which the percentile equivalent and distribution are matched among the various clinical scales in terms of skewness and kurtosis. Thus a uniform *t*-score of 65 on scale 7 is roughly equivalent to a uniform *t*-score of 65 on scale 9. The use of uniform *t*-score rather than linear *t*-scores was one of the improvements of the MMPI-2 over the original MMPI.

A review of the interpretation of common two- and three-code types can be found in most standard textbooks on the MMPI, some of which are included in the References. Certain MMPI authorities have used slightly more complex means of interpreting the MMPI-2. This approach emphasizes specific relationships about multiple scales. For example, Harold Gilderstadt and Jan Duker described criteria in their book *A Handbook for Clinical and Actuarial MMPI Interpretation* for what is referred to as a 1, 2, 3, 4-type scale as follows:

1. Hs, D, Hy, and Pd greater than *t*-score 70 and higher than all other scales.
2. Si less than *t*-score of 70.
3. L, F, and K less than *t*-score 70 unless two or more scales are greater than 100, in which case F, less than *t*-score 80.

The problem in using this complex method is that in many settings only a relatively small percentage of patients match these criteria. In other words, if one uses this method rigidly, a substantial number of patients seen in an inpatient unit would not precisely match one of these complex criteria.

CONTENT SCALES FOR MMPI-2 Several efforts have been made to group items according to certain specified dimensions. With respect to the MMPI-2, Butcher and colleagues developed the method used most frequently. These 15 narrow-band scales were developed by use of a combination of rational and statistical procedures. Development of the procedure resulted in internally consistent, relatively independent scales in which item overlap between scales is minimal. Scales include such content areas as anxiety, anger, cynicism, low self-esteem, fears, obsessiveness, and depression, among many others. Norms for both males and females were simply derived using the same uniform *t*-scores used in the MMPI-2 clinical-scale formation. Reliability data on the scales are acceptable, given the relatively few items in some of the content scales.

Scoring guidelines have been presented by Graham in *MMPI-2 Assessing Personality and Psychopathology*. Validity data for the new scales are limited at the current time; however, data accumulation is expected over the years. Since efforts were made to develop reasonably homogeneous items, interpretation of these scales consists of examining high scores and low scores. For example, patients with high scores on the low self-esteem scale tend to have poor self-concept, to anticipate failure, to be overly sensitive to criticism, and to have trouble making decisions. Low scorers, on the other hand, are self-confident and decisive. A given content scale such as the anger scale, considered alone, is similar to rapid assessment instruments.

HARRIS & LINGOES CONTENT SCALES Most of the MMPI clinical scales were developed using empirical key approaches. For example, an item was included in the depression scale because that item discriminated patients in the reference group from persons in the normal group. Although this method is scientifically sound, the approach results in the inclusion of subtle items that seemingly may have little to do with depression. For example, items such as “I sometimes tease animals” (false) and “It takes a lot of argument to convince most people of the truth” (false) are included in the depression scale (D) even though on the surface they have little to do with depression. Such items are called subtle items. Obvious items are those with strong face validity, such as “I wish I could be as happy as others” (true) and “I brood a great deal” (true). The depression subscale reflects a heterogeneous item group coming from a variety of aspects of depression such as subjective depression, psychomotor retardation, mental dullness, etc. The same total score might reflect a quite different combination of items. Knowing which items in combination actually contribute to the total score can be helpful in interpretation. The Harris & Lingoes subscale can help in that process.

The Harris & Lingoes scales were developed logically by grouping items in a given clinical scale into standard scales judged to reflect a more homogeneous trait. Thus, the Harris & Lingoes scales for depression reflect separate scales for subjective depression (D1), psychomotor retardation (D2), physical malfunction (D3), mental dullness (D4), and brooding (D5).

The Harris & Lingoes scales serve two major purposes: (1) to see which group of items were elevated for a patient scoring high on a given clinical scale and (2) to see why a marginally elevated scale was elevated. For example, one patient who scored marginally high on the depression scale may be evidencing psychomotor retardation and no subjective depression, while another may be just the opposite.

SUPPLEMENTAL SCALES The MMPI contained 566 items, and numerous investigators have used the items to develop special-purpose or supplemental scales. By 1972 over 450 supplemental scales were developed, and the number has kept growing. Scales have been developed to predict success in psychotherapy (Es) and to differentiate alcoholic patients from nonalcoholic patients (MACR) and assaultive prisoners from others not committing a crime (overcontrolled hostility, Oh). Many doctoral dissertations describe research conducted to uncover a new supplemental scale. The success of supplemental scales in accomplishing their stated purpose varies considerably. Many such scales were used for only one study; others such as the MacAndrew Alcoholism (MAC) Scale have been used in many studies over the years.

COMPUTER SCORING AND INTERPRETATION The MMPI-2 can be scored by a computer program. A typical computer scoring program not only scores individual items into validity, clinical, content, and supplemental scales, but also prints out certain positive critical items marked by the patient. These programs save considerable time because the program scores all the material, which may take hours to score by hand. Few clinicians have the time or patience to score routinely the vast number of scales found in the output of a typical computer program. The MMPI-2 computer programs can also be used to interpret the test clinically or at least give clinical hypotheses concerning the patient. These computer-generated reports are in narrative format. Since the report comes from a computer, one may get the impression that the information is more valid or accurate than a clinical interpretation. Although attempts have been made to validate narrative statements made by the computer, efforts have only begun in this regard. One problem with computer interpretation of narrative reports is that the algorithms underlying the narrative reports are proprietary information owned by the company producing the program and are not available to the general user.

Rapid Personality-Assessment Instruments Rapid assessment instruments are narrow-spectrum, self-report, objective instruments that measure some unidimensional aspect of personality. The instrument is written in clear, simple language and provides face validity. Typically, these short instruments can be completed in less than 15 minutes. The scoring is simple and straightforward, usually just adding up the number of positive responses. Many instruments have cutting

scores above which an individual is considered to have a clinically significant problem in that area.

Literally hundreds of such measures exist that can assess a wide variety of personality states and such traits as geriatric depression, obsessive-compulsiveness, impulsivity, and state-trait anger. Table 7.5-3 presents a list of such measures. The overall score typically provides an index of the magnitude of the patient's problems. Most of these tests can be used repeatedly to assess change over time. Most of the instruments were developed with relatively small samples and have just been reported in a few studies; however, a few instruments are widely used.

Table 7.5-3 Rapid Personality Instruments

Some rapid assessment instruments were developed out of the behavioral assessment tradition. One advantage of short personality-assessment procedures is high face validity for the domain of the behavior in question. Most instruments have standardized instructions and small normative groups. Performances of patients with the psychiatric problem in question (anger dyscontrol, or geriatric depression) are frequently reported in terms of mean performance and standard deviation. One can find a rapid assessment instrument to assess many of the common psychiatric problems seen today. Another advantage of these instruments is the ease of administration and scoring. The major disadvantage is that they differ greatly in their reliability and validity. Normative groups are frequently small and are obtained in one location. Most do not have national examples. Many of these short instruments have few if any studies of their construct validity. Most do not systematically assess response style such as faking bad, faking good, or social desirability responding. Examples of the hundreds of rapid assessments include measures of bulimia, obsessive-compulsive behavior, sexual functioning, fear and panic, and personal beliefs.

Projective Personality Test Projective personality tests, in contrast to objective personality instruments, are more indirect and unstructured. Unlike objective tests, in which the patient may simply mark “true” or “false” to given questions, the variety of responses to projective personality tests is almost unlimited. Instructions are usually general, allowing the patient's fantasies to be expressed. Patients generally do not know how their responses will be scored or analyzed, making it difficult to obtain a desired result. Projective tests typically do not measure one particular personality characteristic such as Type A personality (i.e., narrow band measurement) but instead are designed to assess a personality as a whole (i.e., broad-band measurement).

Projective tests often focus on latent or unconscious aspects of personality. Obviously, psychologists and others differ in the degree to which they rely on unconscious information. In many projective techniques the patient is simply shown a picture of something and asked to tell what the picture reminds them of. Projective techniques (objective hypothesis) assume that when presented with an ambiguous stimulus such as an inkblot, for which there are an almost unlimited number of responses, the patients' responses will reflect fundamental aspects of their personality. The ambiguous stimulus is a sort of screen on which individuals project their own needs, thoughts, or conflicts. Different persons have different thoughts, needs, and conflicts and hence widely different responses. A schizophrenic patient's responses often reflect a rather bizarre, idiosyncratic view of the world.

Hundreds of different projective techniques have been developed—most of which are not widely used today. Table 7.5-4 contains a list of commonly used projective tests with a brief description, and the strengths and weaknesses of each. This Chapter focuses on the more widely used projective instruments, including the Rorschach test, Thematic Apperception Test (TAT), and sentence completion test.

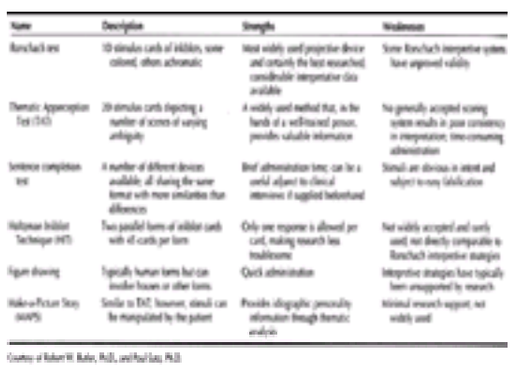

Table 7.5-4 Projective Measures of Personality

The literature on projective techniques is vast. The history of projective techniques goes back many years. Arthur Carr pointed out in the fourth edition of the *Comprehensive Textbook of Psychiatry* that Shakespeare appreciated the importance of one's association to ambiguous stimuli. In an effort to feign insanity, Hamlet associated to a cloud formation certain images that presumably reflected his personality.

Rorschach Herman Rorschach, a Swiss psychiatrist, developed the first major use of projective techniques around 1910. The Rorschach test is the most frequently used projective personality instrument. The test consists of 10 symmetrical, ambiguous inkblots. The inkblot card appears as if ink was poured onto a piece of paper and folded over—hence the symmetry. These 6½ by 9½ inch inkblot cards are the standard stimuli and by convention are referred to by Roman numerals I to X. The cards are presented on thick cardboard, often referred to as plates. Half of the inkblot cards (I, IV, V, VI, and VII) are achromatic (black and white and shades of gray), and the remaining are chromatic. An example of a Rorschach card (card I) can be seen in Figure 7.5-1.

FIGURE 7.5-1 Plate I of the Rorschach Test. (Reprinted with permission from Hermann Rorschach, Rorschach®-Test.© Verlag Hans Huber AG, Bern, Switzerland,

1921, 1948, 1994.)

Using Samuel Beck's approach, the patient is given in essence the following instructions: “You'll be given a series of cards one by one. The cards have on them designs made up of inkblots. Look at each inkblot and tell the examiner what you see on each card and anything that may be represented there. Look at each card as long as you like, only be sure to tell the examiner everything that you see on the card as long as you look at it. When you have finished with these cards give the examiner a sign that you are through with them.”

Minimal interaction between the examiner and the patient occurs while administering the Rorschach, which ensures that standardization procedures are upheld. The examiner writes down verbatim what the patient says during the free-association or response-proper phase. If the patient rotates the card during the response, appropriate notation is made on the test protocol. After the patient has responded to all 10 cards, an inquiry phase begins. During this phase, the patient is asked, “What about the card reminds you of the response you gave?” An almost unlimited range of responses is possible with the Rorschach test and most projective tests. Six different scoring and interpretation systems for the Rorschach tests are available for the clinician. A structural summary of the various measures derived from the patient's response is usually computed.

Scoring and interpretation of the Rorschach test can be time consuming. Administering and scoring the Rorschach test usually takes about 1½ to 2 hours. Typically, responses are scored according to the location on the blot where the response was seen. What about the blot caused the patient's response is referred to as the determinants. The actual content of the patients' response (e.g., animals, human anatomy) is also scored. The scoring system is elaborate and can result in dozens of different indexes being computed; however, the reliability of some of these measures has been questioned. Although the reliability and validity of the Rorschach test as a personality instrument are limited, many studies do support its usefulness as a personality assessment instrument in the hands of a qualified and experienced psychologist.

The limited scientific support of the Rorschach test, however, has resulted in critical reviews by some psychologists. A number of graduate schools of clinical psychology no longer offer formal coursework in the Rorschach test, although most still do.

Table 7.5-5 gives sample responses of four patients to Rorschach card I. Notice how different the responses are to this same card and how the responses reflect the patients' underlying personality.

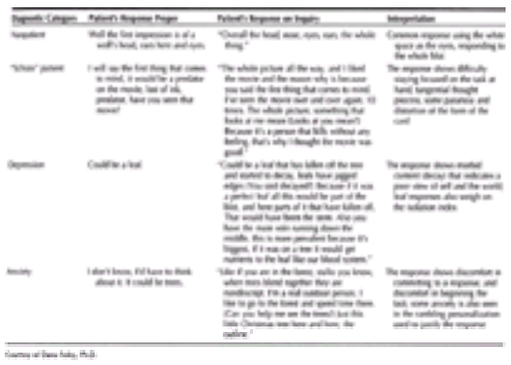

Table 7.5-5 Sample Rorsachach Responses and Interpretation to Card I by Diagnostic Category

Table 7.5-6 shows certain cluster variables that are derived from John Exner's scoring system. The interpretive utility of these cluster variables is also described in the Table.

Cluster Variable	Interpretive Value
Control stress tolerance	Can assess a person's stress tolerance, ability to control behavior, and responses in the environment; can assess a person's style of processing information
Situation-related stress	Can assess the amount of situational stress affecting a person's capacity for control
Affective features	Can assess how a person manages emotional stimuli, types of emotional features (e.g., depression and anger), and ability to modulate their affect
Self-perception	Can assess how persons view themselves and ability to look at themselves
Interpersonal perception	Can assess how a person views others, ways a person interacts with others, and how well a person identifies with others
Information processing	Can assess how persons organize information in their world and how they process it
Cognitive mediation	Can assess the conventionality of a person's perceptions and degree of reality distortions
Ideation	Can assess aspects of ideation such as processing style, use of defenses such as fantasy and intellectualization, and whether thoughts proceed in a logical, coherent fashion or have some thought slippage

Courtesy of Dana Foley, Ph.D.

Table 7.5-6 Cluster Variables (Exner System) and Basic Interpretive Utility

Thematic Apperception Test Although the Rorschach test is clearly the most frequently used projective personality test, the TAT is probably in second place. Many clinicians include both the TAT and the Rorschach test in a battery of tests for personality assessment. The TAT consists of a series of 10 black and white pictures that depict individuals of both sexes and of different age groups who are involved in a variety of different activities. For example, on card 1, a young boy is shown sitting at a table, looking at a violin. Card 2 depicts a farm scene in which a young woman in the foreground is carrying books in her hands; a man is working in the fields nearby, and an older woman is seen in the background. Typically, a patient is shown 10 TAT cards and asked to make up stories about them. The patient is asked to tell what is going on in the picture; what was going on before the picture was taken; what the individuals in the picture are thinking and feeling; and what is likely to happen in the future. An example of a TAT card is presented in Figure 7.5-2. Some examples of typical sample responses to the TAT appear in Table 7.5-7.

FIGURE 7.5-2 Card 12F of the Thematic Apperception Test. (Reprinted with permission from Henry A. Murray, Thematic Apperception Test, Harvard University Press, Cambridge, MA. Copyright © 1943 President and Fellows of Harvard College, © 1971 Henry A Murray.)

Diagnostic Category	Indicated Response
Nonpatient	This lady has a serious look on her face. She had been talking with the older woman in the background. The older woman had been trying to persuade her to do something that she did not feel was quite right. The younger lady is thinking "I can't do what she wants." She is feeling resentful put upon by this woman. The young lady is frustrated by the request. She is trying to think of a response to the request. What happens in the future is she ignores the old woman's request.
"Schizo" patient	This young lady is about to be killed by that witch in the background. The witch is getting ready to poison her. You have to be careful nowadays, the witch wants all of her money and her car. I know I have to watch out. Everybody will get you if they can. I don't trust anyone. Better safe than sorry. . . . I'm not sorry for what I've done, sorry is not something a person should be. Love means you don't have to say you are sorry.
Depression	I hate pictures like this, I am not really good at it anyway. The lady in front is thinking about her mother. Her mother was not an evil lady but the young lady feels like her mother didn't care about her, didn't love her and was very selfish. Her mother didn't provide support. The young woman is thinking that she has been deprived and emotionally abused. The young lady will think for a few minutes and will then go into the bedroom to be by herself.
Anxiety	The young lady is getting ready to go to work but she doesn't want to go. She rented a room from the old lady in the background. The young lady is about to get fired at work. She does not like her job but she needs the money. She doesn't know how she is going to support herself when she loses her job. She goes out to work but has a miserable day. She eventually loses her job and the old landlady kicks her out of her room.

Table 7.5-7 Sample Thematic Apperception Test Responses to TAT Card 12F

Henry Murray developed the TAT at the Harvard Psychological Clinic in 1943. The stories the patient make up concerning the pictures, according to the projective hypothesis, reflect the patient's own needs, thoughts, feelings, stresses, wishes, desires, and view of the future. According to the theory underlying the test, a patient identifies with a particular individual in each picture. This individual is called the hero. The hero is usually close to the patient in age and frequently of the same sex, although not necessarily. Theoretically, patients attribute their own needs, thoughts, and feelings to this hero. The forces present in the hero's environment represent the *press* of the story, and the *outcome* is the resolution of the interaction between the hero's needs and desires and the press of the environment. The TAT, in the hands of a skilled interpreter, can reveal considerable information about the patient's personality that can then be combined with information obtained from the interview and other testing.

The TAT by itself is not an effective diagnostic instrument, nor was it designed to be. Moreover, as measured by traditional methods of assessing reliability and validity, the TAT lacks the scientific merit of objective instruments. The fact that it remains reasonably popular today, however, speaks to its usefulness in clinical practice.

Sentence Completion Test Although a projective instrument, the sentence completion test is much more direct in soliciting responses from the patient. The patient is simply presented with a series of incomplete sentences and asked to complete the sentence with the first response that comes to mind. Examples of possible incomplete sentences are as follows:

> My father seldom....
>
> Most people don't know that I'm afraid of....
>
> When I was a child, I....
>
> When encountering frustration, I usually....

The purpose of the test is to elicit, somewhat indirectly, information about the patient that cannot be elicited from other measures. Since the patient responds in writing, the examiner's time is limited. The length of time it takes to complete this test varies greatly, depending on the number of incomplete sentences. Tests can range from fewer than 10 sentences to more than 75.

There are many variations of sentence completion tests. Some clinicians have developed their own. One form developed by Julian Rotter has some established validity and reliability, but most sentence completion tests do not. Special-purpose sentence completion tests have been developed to measure different problem areas. For example, one sentence completion test is used with patients who have chronic pain, and another to assess issues concerning transsexual patients. The sentence completion test is seldom if ever used alone but is combined with other appropriate instruments.

Advantages of the sentence completion are short administration time, ease of administration, variety of instruments, and ease of construction. Disadvantages are lack of reliability and validity studies and ease of fabrication and deception.

Behavioral Assessment Behavioral assessment involves direct measurement of a given behavior. Rather than focus primarily on human characteristics, such as repression, ego strength, or self esteem (vague terms to a behaviorist), strict behavioral measurement concentrates on the direct measurement that can be observed, such as the number of temper tantrums per unit time; duration, intensity, and number of hyperventilation episodes; or the number of cigarettes smoked per 24-hour period.

Although an early strict behaviorist would count only behaviors that were observable, a broader definition of behavior has emerged, under which just about anything people do, whether overt (e.g., crying, swearing, or handwashing) or covert (e.g., feeling and thinking), is considered behavior. To measure covert behavior, the patient is typically the informant. For example, to measure dysfunctional thinking, patients may simply carry a mechanical counter and press it every time they have a particular target self-disparaging thought during a given period of time.

Direct Counting of Behavior Measuring overt behavior is direct and can be done by the patient, a family member, or an impartial observer. Cognitive behavior therapists use these measurements to establish baselines for a given undesirable behavior (e.g., violent thoughts that the patient may wish to reduce). Similarly, behavior that the patient wants to increase can also be measured (time spent studying, time out of bed, distance walked on a treadmill). Follow-up measures of the behavior monitor progress and quantify improvement.

Different aspects of a target behavior can be counted or measured, such as frequency (e.g., number of behavioral occurrences), duration (length of time behavior occurred), and intensity. Occasionally a behavior occurs so frequently that counting each occurrence becomes problematic. In such instances, one can simply have the individual indicate whether or not the behavior occurred during a given period of time.

The advantage of counting specific behavior is that it is direct, is objective, is easily understood, and quantifies behavioral change. The problem with counting specific behavior is that many patients have so many different symptoms it would not be possible to quantify them all. Sometimes counting a target behavior itself alters the behavior. Covert behavior can be measured by use of a number of different approaches; for example, asking patients to keep a log of dysfunctional thoughts concerning their employer during a given 18-hour period.

Rating Scales Rating scales have been developed to detect various aspects of personality. On some of these scales, patients simply rate themselves on a given scale dimension such as sadness: 1 = I am not sad, 7 = I wish I were dead. On other scales a significant other person is asked to rate the patient: 1 = the patient seldom mentions being sad, 7 = the patient says he or she would be better off dead. Some of the more commonly used rating scales are designed to quantify information obtained by the mental health professional during a mental status interview.

Brief Psychiatric Rating Scale One of the more commonly used interview rating scales is the Brief Psychiatric Rating Scale (BPRS). To use the BPRS, the psychologist or psychiatrist completes a mental status interview with the patient and then rates that patient on a series of 18 psychiatric symptoms such as motor retardation, blunted affect, conceptual disorganization, anxiety, and guilt. Expanded definitions of each of these terms are provided to the examiner. The interviewer rates each domain on a seven-point Likert scale from “not present,” the lowest rating, to “extremely severe,” the highest rating. An experienced interviewer can complete the ratings in 2 or 3 minutes. The BPRS has been used extensively in drug-outcome and other studies. The advantage of the BPRS is that the interrater reliability is reasonably high for a rating scale of this nature. A summary of over 300 studies using the BPRS found an interrater reliability correlative of .80 or above on the total score in most studies. The first author has used this instrument to teach medical students to observe patient's behavior, and relatively inexperienced medical students can establish reasonably good reliability. Other advantages of the BPRS include the ease and speed of rating and the well-defined symptom

description.

One disadvantage of this scale is examiner subjectivity. Another disadvantage is that interviewers can only rate patients on what they personally observe during the interview. Examiners differ greatly in their style of conducting a mental status interview; thus an examiner may ask many questions in one given area, (e.g., somatic concern) while relatively overlooking another area (e.g., suspiciousness). Another examiner might do just the opposite.

Schedule for Affective Disorders and Schizophrenia (SADS) The BPRS ratings are made on the basis of a mental status interview and do not require that the examiner ask any specific questions of the patient. This approach allows flexibility but results in considerable variation interview style. A flexible interview results in different information being gathered by different interviewers. The SADS, on the other hand, is a highly structured interview instrument. The interviewer is required to ask each patient a series of prescribed questions to ensure that all relevant areas are addressed. However, the structured interview does vary somewhat from patient to patient because of the branching quality of the interview schedule. For example, a patient is asked a question similar to the following: “Have you ever heard voices or other things that weren't there or that other people couldn't hear or see?” Based on the patient's response, the examiner asks other detailed, prescribed follow-up questions concerning hallucination, or if the response was negative, the interviewer moves on the next question in a different area. This approach ensures that all areas are covered in a comprehensive fashion. The SADS is especially helpful for establishing a reliable diagnosis. The SADS can also be used as an index of behavioral severity. Behavioral changes can be determined by repeated administration.

The SADS is quite time consuming, mainly because it is comprehensive. After an initial evaluation using the complete SADS, a condensed SADS (SADS/C) can be used for follow-up.

The advantage of the test is that it is comprehensive and has reasonably good reliability. One disadvantage is its length; another is that in the structured interview the interviewer must read questions from a lengthy booklet, which makes it difficult to establish eye contact and rapport with the patient. The SADS has been used for both research and clinical purposes, probably more frequently for the former.

ASSESSMENT OF PERSONALITY IN CHILDREN AND ADOLESCENTS

Assessment of emotional and interpersonal characteristics in children presents many challenges to the clinicain because of the discontinuities in development that exist throughout childhood, adolescence, and adulthood. Many clinicians are reluctant to assign the diagnosis of a personality disorder before the ages of 16 to 18 years because of the rapid changes that occur during childhood. However, assessment of children and adolescents can often reveal antecedent symptoms, behaviors, or traits associated with emotional disorders at an age when these problems are quite amenable to intervention. For this reason, facility in assessment of emotional disorders in children and adolescents is important for mental health professionals.

Special Considerations in Assessment of Children Children with symptoms of emotional or behavioral disorders are best assessed within developmental and ecological contexts—both of which help one to interpret the child's symptoms from the perspective of developmental influences on behavior, with consideration of the risk and protective factors in the child's social environment. Indeed, the balance of risk and protective factors often provides important clues to the etiology of the child's current problems and the prognosis for effective intervention.

Developmental Context Knowledge about the normal sequence and transitions of development forms a fundamental backdrop from which to view children's suspected psychopathology. The major developmental transition of infancy related to formation of a secure attachment relationship with significant caregivers gives way to the movement from dependence to greater self-reliance in the toddler years. The developmental tasks most salient during the preschool years involve development of a growing capacity for empathy and self-control, while showing a desire to master developmental tasks. Within the early to middle elementary years, youngsters strive for greater mastery of knowledge and intellectual and academic skills, leading to feelings of productivity and competence. The developmental tasks of adolescence center around separation-individuation, resolving conflicts with authority figures, peer group identification, and realistic appraisal and evaluation of self-qualities. Although development is not in linear stages, familiarity with the primary developmental tasks and transitions of each age period provides an important context in which to view current symptoms.

Decisions about appropriate assessment methods are also based on developmental factors. Before children participate in projective testing procedures such as story-telling tasks, the clinician must have developmental information about their level of expressive language, receptive language, and conceptualization ability. Knowledge of children's reading proficiency is critical when presenting them with self-report measures. If children are asked to complete projective drawings, information about their level of visual-motor development is important for interpretation. Young children often do not have the motor or language abilities to provide meaningful responses to projective testing procedures, but they may reveal much about their socialization abilities, fears, anxieties, and significant relationships through play. Therefore, play observation techniques can be a useful alternative to more-formal projective measures. Likewise, some adolescents may resist providing responses on projective measures that require verbal disclosure to a clinician but willingly complete objective paper-and-pencil personality measures that require a less direct response. Choosing an approach to assessment with the developmental context in mind enhances the validity of the information obtained.

Ecological Context The broad social and ecological context of children's family, peer group, and social relationships and the culture in which they live can influence the interpretation of assessment information. From a developmental psychopathology perspective, most psychopathology is expressed as an interaction among various factors that are operating at the levels of the individual (including developmental and personality attributes), the family (e.g., parenting skill, the security of primary attachment relationships, marital stability, extended family support), the community (e.g., influences of work, school, informal social networks, socioeconomic factors, family social isolation), and the larger cultural context of society (e.g., cultural values and beliefs that govern behavior).

Two 4-year-old girls were referred by their Head Start teacher because of concerns about the recent onset of regressive behavior (e.g., enuresis during the school day, immature speech patterns). The girls were interviewed separately but were reluctant to talk with the clinician. A play interview was set up with each, using dolls and a dollhouse with a variety of furniture. The first girl assumed the role of “mother,” and played out the scenario of feeding and diapering the baby doll in a nurturing manner. The second girl was aggressive in her play, with enactment of the adult dolls hitting the child dolls and making them “die.” The child dolls were described as having blood on them. The family context for the first girl revealed that a new baby sibling had been born just prior to the onset of regressive symptoms; the infant was born prematurely, and the mother spent much time with the infant in the hospital. Both the arrival of the new sibling and the mother's separation from her 4-year-old daughter created the social context for the emergence of the child's regressive behavior. In the second case, the girl's mother was interviewed and initially denied any recent stresses in the family. After the clinician questioned her further and provided a description of her daughter's play, the mother revealed that she had a new boyfriend who had just moved into the house. She ultimately revealed that she had noted her daughter's fear of the boyfriend and her frequent tearfulness at home. She reported suspecting that her boyfriend might be sexually molesting her daughter and agreed to call Protective Services in the presence of the clinician to make a report.

As these cases illustrate, there can be vastly different explanations for similar presenting symptoms, and often the projective assessment procedures only suggest concerns, without providing enough specific information about the nature and etiology of the problems. The social context can reveal both risk and protective factors that are important in conceptualizing the child's problems. For instance, in the case of the 4-year-old girl who was sexually abused, the presence of the mother's boyfriend in the home was a risk factor, as were the mother's isolation from her extended family, poverty, and living in a dangerous, violent neighborhood. However, protective factors included the mother's strength in wanting to protect her daughter by admitting the problem, making a report to Protective Services, and forcing the boyfriend to leave the home. Another protective factor was a concerned and observant Head Start teacher who noted the girl's behavioral change and sought help immediately. The ecological approach allows one to examine the possible multiple determinants of emotional psychopathology in children and elucidates the interaction between the risk and protective factors in the child's life.

Use of Informant Information Children and adolescents are usually referred for assessment because of concerns of their parents or caregivers. Teachers may also be the source of specific concerns. For this reason, information relevant to the diagnosis is typically obtained from these significant adults who can provide important information regarding the child's behavior in various settings. Reliance on persons other than the client as reporters of the primary symptoms is a fundamental difference in the diagnostic process with adults and children.

The validity of the information presented about children's symptoms is often a concern to clinicians. During intake, parents often express feelings of anxiety, frustration, or both regarding their child's problems, and their descriptions of the child may be exaggerated or vague (e.g., “she never minds” or “he always acts like a monster”). It is not uncommon for depressed parents to report an increase in number and severity of symptoms in their children. When one suspects that the informant's perceptions may be distorted, collateral information must be obtained from teachers or others familiar with the child's current problems. A primary task is helping informants translate imprecise complaints to specific descriptions of behaviors of concern, using methods that help the clinician ascertain the nature,

frequency, and severity of symptoms. Several behavioral assessment procedures are useful in providing age- and sex-referenced ratings of symptom characteristics.

Specialized Training Clinicians who conduct personality assessment of children need training not only in clinical assessment methods, but also in developmental psychology and child psychopathology. Presentation of many emotional disorders in prelatency years differs from postlatency presentation. For instance, young children with depression may exhibit predominant symptoms of conduct problems, tantrums, oppositional behavior, and hyperirritability rather than the presentation most often associated with older children and adults (depressed mood, diminished interest in activities, vegetative symptoms, anhedonia, etc.). Training and experience in assisting the child to meet the demands of the testing situation are also critical. Children's ability to participate in testing depends on their attention and concentration ability, anxiety regarding separation from significant others during the testing, fatigue or hunger states, motivation and persistence ability, and the relatively greater influence of familial, cultural, and environmental variables. A clinician with specialized training in working with children will both understand these influences on child test-taking behavior and have the skills to work with the children to achieve more valid results.

ASSESSMENT OF CHILDREN AND ADOLESCENTS

Like assessment of adults, personality assessment of children can be accomplished via three primary methods: projective, objective, and behavioral tests and procedures. The projective methods involve direct interaction with the child and adolescent, whereas objective and behavioral methods often involve obtaining information from significant adults in the child's life as well as from direct interaction with the child. With the evolution of more sophisticated statistical methodology and psychometric science in recent years has come the development of new objective and behavioral measures of personality. Improved validity indexes and psychometric procedures that take into account informant reporting are now routinely included. Many of the projective procedures have changed less, though improvements in developmental norms for interpretation have increased the diagnostic validity of measures such as the Rorschach test.

Projective Assessment Procedures As stated in the adult section, objective tests of personality present the patient with a structured set of questions and a finite range of answers. Projective tests, on the other hand, present more ambiguous stimuli and ask the adult or child to make up something (i.e., story, percept, or drawing) from the stimulus. The most common projective assessment procedures for children and adolescents are the Rorschach test, various projective story-telling measures (e.g., Roberts Apperception Test, Children's Apperception Test [CAT]), projective drawings (e.g., human figure and Kinetic Family drawings), and incomplete sentence procedures (Table 7.5-8).

Name	Age Range	Description
Rorschach inkblot test	5 years–adulthood	Consists of 10 inkblots, some colored and others achromatic, used as basis for eliciting associations that reveal disrupted personality development.
Children's Apperception Test (CAT)	3–10 years	Two versions of the CAT—animal and human—depict characters in various social situations and are used to elicit stories from children. Younger children are felt to identify more readily with the animal figures, while older children are usually presented the human figures. Scoring and interpretation are based on psychodynamic theory.
Adolescent Apperception Cards	11–19 years	Eleven picture cards focus on parent–peer–and sibling–adolescent interaction, pulling for themes of physical and sexual abuse, neglect, peer acceptance, loneliness, depression, drug use, and domestic violence. Two versions are available: one depicting white teenagers and the other, black teenagers.
Roberts Apperception Test for Children (RATC)	6–15 years	Sixteen picture cards—with parallel male and female versions depicting white or black characters—are designed to elicit information about both adaptive and maladaptive functioning. The interpersonal futures links the adaptive and clinical scales in order to specific people in the child's life and facilitates a better understanding of the interaction with family members, peers, teachers, and others. Has standardized scoring system and normative data for cut interpretation.
Tell-Me-A-Story (TEMAS)	5–18 years	Multicultural apperception test with 23 color picture cards (15 of which are sex specific) depicting minority (Hispanic, black) or nonminority characters. Measures 10 personality functions (e.g., aggression, interpersonal relations, self-concept), 10 cognitive functions (e.g., time, theory), impulses, and 7 affective functions (e.g., happy, sad, angry, fearful). Has an objective scoring system and normative data based on a diverse cultural and ethnic sample.
Projective drawings	3 or 4 years through adolescence	Various versions exist, from individual drawings of human figures, a house, and tree to kinetic drawings of family. Simple, cost-effective measures that provide information on children's perception of self and relationships with others. Especially useful with children who have difficulty with verbal expression. Objective scoring available for some drawings (e.g., human figures), but interpretation of other types of projective drawings is often subjective.
Sentence- and story-completion tests	8 or 9 years through adolescence	A number of different formats available, each providing a sentence stem or initiating a story, after which the child is asked to complete the sentence or story. Provides information on such factors as interpersonal relationships and dynamics, self perception, wishes and worries.

Table 7.5-8 Projective Assessment Procedures for Children

Rorschach Inkblot Test Projective instruments such as the Rorschach test allow the clinician to explore dynamics of the child's personality by gathering information on both the child's perceptual-cognitive world and the inner fantasy world. The Rorschach test ideally is used as part of a more comprehensive battery that includes an interview with the child and significant adults, expressive (play) techniques, and perhaps story-telling techniques, to allow the child maximum freedom and spontaneity of expression.

The Rorschach test with children has a long research and clinical history of examining developmental norms and symbolic interpretations. Various normative studies have added to its interpretive validity, with the comprehensive studies of John E. Exner and Irving B. Weiner discussed in *Assessment of Children and Adolescents*, the third volume of their book *The Rorschach: A Comprehensive System*, providing the most recent and detailed normative data available for children ages 5 to 16 years. Exner's descriptive statistics for the normative data are provided in tables for each year of age; computerized programs for obtaining the structural summaries in the context of these developmental norms are also available. The clinician using the Rorschach test to evaluate children and adolescents must take care to analyze the structural summary within the context of appropriate age norms, as a given result interpreted as normal for a young child could be of concern in an adolescent. Children's Rorschach responses have been examined as a function of their cognitive functioning, academic performance, and behavioral problems within the school setting. The underlying conceptual framework for this work hypothesizes a direct relation between secondary process development and school achievement. A published review of projective tests suggests that the child who can deal successfully with drive-laden impulses through play and fantasy is seen as more capable, open, and flexible. According to Exner and Weiner, the Rorschach test is not particularly good at discriminating one specific personality disorder from another, as is, for instance, a structured clinical interview. However, the Rorschach test can provide useful information regarding the way children with personality disorder traits view the world and can shed light on manifestations of flawed or disrupted personality development.

As with adults, there are numerous systems for administering and scoring the Rorschach test with children, but all ask children to say what they see on the blot (i.e., the percept), followed by an inquiry referring back to each response. Whether the inquiry should be done following the child's free association responses to all 10 blots or whether the inquiry is best accomplished after each individual blot is controversial. Proponents of the latter approach suggest that young children may have difficulty remembering their reasoning behind the original free associations or may become fatigued by the end of the test, thus limiting their cooperation and responsiveness to the inquiry. Research by Exner and Weiner found the latter approach to inquiry to be unnecessary in their large child sample, except for a small percentage of children who were seriously disturbed. Of most concern is that inquiry immediately following each free association response may contaminate subsequent responses; thus reserving the inquiry until free association responses have been obtained for all 10 blots is preferred. Clinicians must also be aware of state anxiety as a potential confounding variable in children's responses to the Rorschach test. Care in building rapport and an explanation of the purpose and process of testing can ease the situational anxiety.

As with adults, scoring is done on the basis of response characteristics, or determinants, such as form, color, shading, texture, and dimensionality. The content and form quality of the child's responses is also used in scoring and interpretation.

Projective Story-Telling Procedures In projective story-telling approaches, the child is presented with a picture stimulus of human or animal figures in rather ambiguous situations. The child is asked to make up a story about the figures—a story that has a beginning and an end and includes the thinking and feeling of the persons represented in the pictures. A fantasy response is evoked, and the resulting projective information is a combination of the perceptual and the imaginative. Stories are typically analyzed for repetitive, unique, intense, or problematic themes, beliefs, or affects. This procedure closely resembles the TAT approach used in adults.

Children's Apperception Test The initial CAT, developed in 1949, used animal figures and was developed for children ages 3 to 10 years. Animal figures were thought to be more culture-free than human characters. In 1965, the human figures version (CAT-H) was produced, showing human figures in situations as analogous as possible to those pictured in the animal version. During administration, the cards are presented individually in the numbered order of the card (because certain cards were designed for sequential impact). The child is asked to tell a story about each picture (e.g., what is going on, what happened before, what will happen next). There is some debate about the use of prompts with young children, and whether such prompts (e.g., How did the story end?) may contaminate important projective information. Generally, prompts are often necessary to help the young child understand what is expected. Young children have a tendency to merely describe or label portions of the picture and may not understand the concept of telling a story with a beginning, middle, and conclusion. However, the clinician must always guard against overly intrusive or helpful prompts that guide the child's responses in a particular direction or suggest a specific format for the story. The various scoring protocols for CAT have focused on analysis of ego functions and evaluating the relative use of various defense mechanisms. However, qualitative interpretation is also made on the basis of recurrent or sequential themes and determination of identification figures, while taking into consideration the child's family and case history

information.

Comparisons have been made between CAT and the Rorschach test as to the types of information obtained and the overlap between them. These tests have been found to complement rather than duplicate each other, as they tap different aspects of personality functioning. The CAT generally can be expected to reflect interpersonal and familial relationships, ego functioning, sources of discomfort, and conflicts and past traumas. In contrast, the Rorschach test, according to M. R. Haworth, explores deeper levels of personality, emotional responsiveness, mechanisms for control, the extent of reality contact, and available defenses.

Although the CAT continues to be used, its picture stimuli are quite outdated in appearance and often do not appeal to children today. Further, clinicians are not as apt to be trained formally in the psychodynamic theories of interpretation that are used with the CAT. Newer apperception tests, such as the Roberts Apperception Test for Children (RATC), are being used with increasing frequency because of their modern appearance and because they use scoring procedures that are less reliant on psychodynamic theory.

Roberts Apperception Test for Children The manual for the RATC offered several criticisms of existing picture apperception tests as a rationale for the development of the RATC. It reported that most of the previously developed techniques either were not developed specifically for children or were appropriate for a limited age range. Second, the nature of the stimulus characteristics of the cards, rather than the child's unique personality, often was the primary determinant of the child's responses. Finally, the existing tests lacked a standardized system for scoring the thematic content and structural characteristics of the child's responses. Therefore, the interpretation of thematic apperception tests was subjective and often based on the clinician's own experiences, theoretical orientation, and personality style. The RATC was designed to overcome these limitations by using stimulus cards that (1) depict children and are designed for use with children ages 6 through 15 years, (2) emphasize everyday interpersonal events of contemporary life, (3) are consistent in their presentation, and (4) are easily scored, objective measures that yield high interrater agreement. In addition, the RATC provides normative data for 200 well-adjusted children ages 6 through 15 years to aid in clinical interpretation. Interpretation of the RATC, like that of other projective measures, is based on the assumption that children presented with ambiguous drawings of children and adults in everyday interaction will project their typical thoughts, concerns, conflicts, and coping styles into the stories they create.

The RATC uses 16 stimulus cards, with parallel male and female versions, that represent important interpersonal themes (e.g., parental disagreement and affection, peer conflicts, observation of nudity, and sibling rivalry). A supplemental version of the RAT has African-American figures. The responses are scored on a number of scales measuring both adaptive and maladaptive functioning. Table 7.5-9 illustrates the Profile Scales and Indicators.

Adaptive scales:	Reliance on Others
	Support-Other
	Support-Child
	Limit Setting
	Problem Identification
	Resolution 1
	Resolution 2
	Resolution 3
Clinical scales:	Anxiety
	Aggression
	Depression
	Rejection
	Unresolved
Indicators:	Atypical Response
	Maladaptive Response
	Refusal

Table 7.5-9 Profile Scales and Indicators for Roberts Apperception Test for Children

In addition, the child's responses may be scored on the Interpersonal Matrix, which summarizes the relations between the scales and indicators, and the figures identified in the child's stories. In general, the standardized scoring procedures aid interpretation and clinical utility of this scale.

Projective Drawing Techniques Projective drawing techniques are often helpful in establishing rapport and engaging shy or negative children who are reluctant to become involved in more-verbal interaction with the clinician. The nondirective nature of the technique allows the child time to adjust to the testing environment and to the clinician without being propositioned for a verbal response. The nonverbal nature of the task also makes it amenable to younger children and those who are cognitively impaired or non-English–speaking.

Two common approaches to projective drawings are the House-Tree-Person and Kinetic Family Drawing. In the former, the child is first asked to draw a person and, after completing that drawing, is asked to draw a person of the sex opposite to that of the person in the first drawing. When finished, the child is asked to draw a house and then a tree. Interview questions asking the child to identify the humans, tell what they are doing in the picture, or how they are feeling can shed light upon the child's thoughts about the drawings. With the house, the child may be asked who lives in the house, what they are doing, etc. Kinetic Family Drawing is accomplished by telling the child to “Draw a picture of your family doing something.” The drawing can elicit verbal comments that concern family cohesiveness or conflict, the perceived role of the child within the family system, relationships with significant others within the family and the degree of interaction versus isolation of various family members, and family structure and the hierarchy of power.

In addition to the verbal information elicited from the drawings, graphic analysis can include placement of the drawing on the page, overworking or sketchiness of the lines, size and relative placement of the drawings, pencil pressure used, amount of detailing, symmetry, indicators of dissociation (i.e., incongruities between the graphic drawing and the verbal description of it), and use of shading or color.

As with other projective techniques, projective drawings are not intended to be used alone to make interpretations of personality functioning. Corroborating information to support any interpretations of the drawings must be obtained through interview or other test procedures that could shed light on the meaning of the child's performance.

Story- and Sentence-Completion Techniques A story-completion technique often begins with a sentence (or a couple of sentences) that represents the beginning of a story plot to be completed by the child. An illustrative example from the Madeline Thomas Stories follows:

A boy goes to school. During recess he does not play with the other children. He stays by himself in a corner. Why?

Story-completion techniques are useful in adding clinical information about rather focal problems and possibly the child's problem-solving skills in addressing these problems. Although there is little research on diagnostic utility and standardized scoring and interpretation schemata, the qualitative information obtained is often worth the time expended in administering the technique.

Sentence-completion procedures can be administered in various formats. One approach is to read the sentence stems to the child rapidly while asking the child to respond with the first information that comes to mind. This approach may reduce response latencies and loosen the child's defenses to response, especially if the items are rather repetitive. Often, information about family dynamics and relationships, and the child's self-esteem, wishes, and worries are revealed more directly than with other projective techniques. Valuable clinical information can be obtained by noting the child's response latencies, the positive or negative tone of the responses, and the predominant mood endorsed in the responses. Another administration approach allows children to write the answers to the sentence stems on a paper, working at their own pace. This is often the best approach to use with adolescents, who value autonomy. Another advantage of sentence-completion techniques is their economy and flexibility of administration.

Objective Personality Measures Objective approaches to child personality assessment typically have straightforward test stimuli and clear instructions regarding completion of the tests, as opposed to projective approaches that typically use more-unstructured, ambiguous test stimuli. Objective tests typically have good standardization, reliability, and validity, and they often are norm-referenced to provide comparisons with a particular criterion group.

The advantages of using objective measures with children are similar to those used with adults. Disadvantages include the length of the measures (most have several hundred questions to which the informant must respond), the reading level required for completion (which could place children and adolescents at a disadvantage),

and the initial expense of purchasing either computer administration or computer scoring software. Despite the disadvantages, objective personality measures remain an important part of a comprehensive personality assessment by providing a broad survey of major areas of psychopathology at the initial stages of the evaluation. Descriptions of some of the major objective personality measures follow, and a more complete listing is provided in Table 7.5-10.

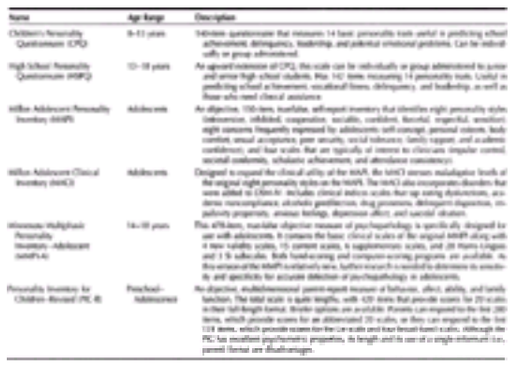

Table 7.5-10 Objective Personality Measures for Children

Robert P. Archer states in his book *MMPI-A: Assessing Adolescent Psychotherapy* that the Minnesotta Multiphasic Personality Inventory–Adolescent (MMPI-A) serves two important functions in assessment of adolescents: (1) it provides an objective evaluation and description of an adolescent's level of functioning related to selected standardized dimensions of psychopathology, and (2) repeated administration can provide a means of assessing ongoing changes in psychopathology, perhaps as a result of intervention. The MMPI-A should be administered in most cases only to adolescents from 14 to 18 years. On occasion, 12- and 13-year-old adolescents who meet the administration guidelines can be evaluated with the MMPI-A, but it should not be administered to anyone younger than 12 years of age. Above age 18, the MMPI-2 is the most appropriate version to administer. This lengthy measure demands that an adolescent of any age be able to read and understand the items (which are written at an average seventh-grade level) and tolerate the time and persistence required to complete the measure. Trained personnel must administer the test in a supervised environment. The MMPI-A is geared toward a population at risk for psychopathology and is not particularly useful for exploring personality styles of nonclinical adolescents.

One concern about the use of the MMPI-A and other objective personality measures during adolescence is the stability of the symptomatology. Most would agree that the MMPI-A is best used to obtain a description of current adolescent psychopathology rather than to make long-term predictions regarding future functioning. The ongoing maturational changes in adolescent personality development most likely account for the lack of stability in predictions of future psychopathology; this underscores the need for the clinician to be sensitive to the developmental issues in this population.

Millon Clinical Multiaxial Inventories Among the most widely used objective measures of child personality is the Millon series of assessment instruments. Two of the scales are specific to adolescents: the Millon Adolescent Personality Inventory (MAPI) and the Millon Adolescent Clinical Inventory (MACI). The original MAPI was designed to be used in both clinical and nonclinical (e.g., vocational, educational) settings to provide an appraisal of adolescent personality and concerns. The goal of Millon and his colleagues was to create a broad instrument that would be valid and reliable in assessing the inherent traits and current state of adolescents. The MAPI is written at a sixth-grade reading level and provides information on eight personality styles that mirror Millon's theory of personality (Table 7.5-10 provides a listing of the MAPI scales). At maladaptive levels, the personality styles correspond to Axis II disorders found in the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R). In addition, the MAPI contains 12 additional scales: 8 that focus on worries that many teens experience, and 4 that address specific behaviors of concern. Because of its nonclinical nomenclature and design, the MAPI can be used successfully to identify personality styles and concerns in a population of normal adolescents. Prior to the development of the MACI, the MAPI was used with clinical populations as well. However, as Mark Marvish stated in *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*, the MAPI shares the same disadvantage as the MMPI-A: the stability of prediction based on its scores varies.

The MACI was designed to assess maladaptive levels of the original eight MAPI personality styles. In addition, it added items to correspond to the three new personality disorders added to the fourth edition of DSM (DSM-IV), and it identifies adolescents who have marked tendencies toward borderline functioning. The 12 personality patterns assessed by the MACI are given descriptive rather than diagnostic labels, partly because of the problems of predictive stability within the adolescent population. The MACI also added several new clinically oriented scales to the original MAPI format; whereas the MAPI had 4 scales that address behavioral adjustment problems, the MACI expanded this section to include 11 scales that reflect more extremely maladjusted behaviors (see Table 7.4-9).

Given the recent development of the MACI, empirical support for its validity, reliability, and interpretation is needed. However, it is well grounded in theory, and it benefits from the empirical information gained with its predecessor, the MAPI.

Personality Measures for Specific Disorders in Children In contrast to the multidimensional personality measures discussed above, several measures address more-specific disorders in children, such as depressive and anxiety disorders. Several of these measures are described in Table 7.5-11.

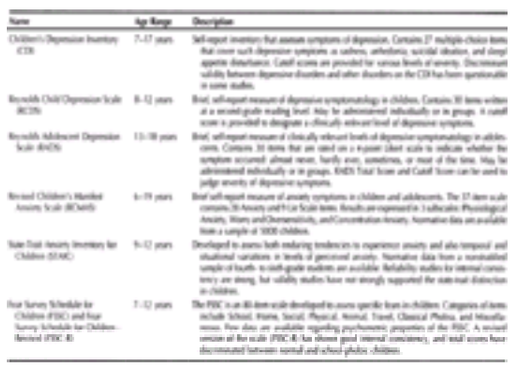

Table 7.5-11 Personality Measures for Specific Disorders in Children

Often clinicians use the multidimensional personality measures to obtain a broad overview of risk for psychopathology and then use the narrower-band, more specific measures to explore a particular set of symptoms in greater detail. Neither type of personality inventory is used to confirm a diagnosis, but both provide valuable information about the nature and severity of symptoms that can be combined with other approaches to arrive at a diagnosis.

Advantages of the specific personality inventories include their brevity, low cost in terms of time to administer, and ease in scoring and interpreting. However, as with similar adult measures, caution is needed in reviewing the psychometric qualities of these personality measures, particularly with regard to discriminant validity for the disorder under study versus other disorders versus results for children without disorders. Many of the scales were developed for children in a narrow age range, with a small normative sample that is not representative or stratified, and they have questionable discriminant validity. As long as the clinician uses these measures as part of a more comprehensive approach to diagnosing emotional problems and understands their strengths and weaknesses, the measures can be useful.

Behavioral Assessment Procedures Behavioral assessment procedures offer a highly structured method of obtaining information about behavioral and emotional functioning and social competencies of children and adolescents. These procedures include direct observations and informant ratings on age- and gender-normed scales. The popularity of these measures has grown in recent years, in part because of their improved psychometric properties, their cost-effectiveness, and their

utility in multitrait-multimethod diagnostic procedures (Table 7.5-12).

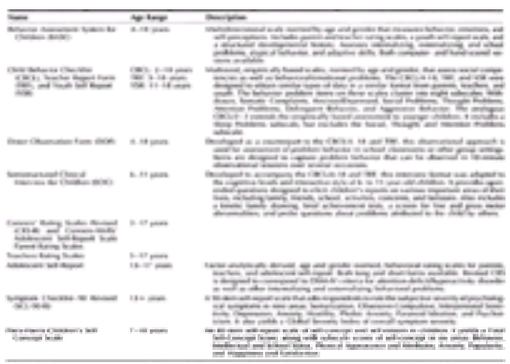

Table 7.5-12 Behavioral Assessment Procedures for Children

Validity of Informant Reports Use of behavioral rating scales raises questions about the validity of informant information. The research on agreement among various raters of child behaviors is consistent in showing greater agreement between raters who interact with a child in similar situations (e.g., between mothers and fathers) than between raters who interact with the child in different situations (e.g., between parents and teachers, or parents and children). Thomas Achenbach and colleagues found that agreement tends to vary across age and types of problems rated: agreement was greater for 6- to 11-year-olds than for adolescents, and for externalizing behavior problems than for internalizing them. Sex differences are also apparent when the number of reported behavioral problems is examined. Achenbach reports that both parents and teachers observe more behavioral and emotional problems for referred and nonreferred males than for females in childhood and adolescence. Achenbach says that studies of youth self-report of problems reveal that 11- to 18-year-old females report more problems than males of the same age. Differences in number of problems reported are also a function of the type of population (inpatient versus outpatient versus nonreferred). According to Achenbach for nonreferred males and females, children's self-ratings of total problems were higher than parent or teacher ratings. In an inpatient sample, Alan Kazdin and colleagues reported that children rated their own symptoms of depression and aggression lower than their mothers did but not lower than their fathers. Yet the magnitude of the children's self-ratings relative to those of other children accurately discriminated diagnostic groups. Catherine Stanger and Michael Lewis, in a study of interrater (i.e., parent, teacher, youth) agreement with 13-year-old youths, demonstrated that the youths rated themselves significantly higher than all other raters on all scales and that mothers and fathers rated more behavioral concerns than teachers, especially for internalizing problems. Stanger and Lewis say that the research to date suggests the importance of collecting information from multiple informants regarding child and adolescent behaviors and that ratings of no one informant can be substituted for those of another.

Advantages and Disadvantages of Behavioral Approaches There are several advantages of the behavioral approaches to assessment of behavior and emotional functioning in children and youths. These procedures are cost-effective in that they maximize the amount of information obtained with little clinician time. They often have convenient hand-score or computer-scoring methodology, another cost-effective aspect. Use of behavioral assessment increases the likelihood of obtaining information from multiple sources (e.g., teachers, parents) across multiple settings (e.g., school, home, day care). These sources of information are necessary for some diagnoses, such as attention deficit/hyperactivity disorder. Many of the scales are empirically derived, factor-analytical scales that are normed for age and sex and generally possess good psychometric properties.

Disadvantages of behavioral rating methods for children include questions about the validity of informants' reports and concerns about informant reading level. The behavioral ratings are filtered through the perceptions of the informant, and the degree of frustration, emotional pathology (e.g., depression) and intellectual and academic skills of the informant are critical to understanding the report. There is much debate about how to handle discrepant ratings across informants. Although perfect correlation is not expected, the issue of how to weigh one person's observations against those of another remains unresolved.

Behavioral Assessment System for Children (BASC) The BASC is a relatively new (1992) scale developed as a multimethod, multidimensional approach to evaluating the behavior and self-perceptions of children from 4 to 18 years of age. It is *multimethod* in that it has five components that can be used individually or in any combination: a self-report scale on which children or adolescents can describe their own emotions and self-perceptions, two rating scales (one for parents and one for teachers) that gather descriptions of the child's observable behavior, a structured developmental history, and a form for recording and classifying direct observations of the child's classroom behavior. It is *multidimensional* in that it measures many aspects of behavior and personality, including both positive and negative dimensions. The BASC allows comparison of information from multiple sources and offers various types of validity checks that allow the clinician to evaluate the consistency and veracity of the informants. Scales are consistent across age, gender, and forms (i.e., parent, teacher). For the Parent and Teacher Rating Scales, the Externalizing Problems Scale includes subscales of aggression, hyperactivity, and conduct problems; the Internalizing Problems Scale includes subscales of anxiety, depression, and somatization; and the School Problems Scale includes subscales of attention and learning problems. The Other Problems Scale includes two subscales: atypicality and withdrawal. Finally, the Adaptive Skills Scale includes adaptability, leadership, social skills, and study skills. The Self-Report of Personality (SRP) contains four primary scales that differ from the parent and teacher versions. The Clinical Maladjustment Scale includes subscales of anxiety, atypicality, locus of control, social stress, and somatization. The School Maladjustment Scale includes attitude toward school, attitude toward teachers, and sensation seeking. The Other Problems Scale includes depression and sense of inadequacy subscales; and the Personal Adjustment Scale includes subscales of relations with parents, interpersonal relations, self-esteem, and self-reliance. Information obtained on each of the scales can be referenced against national age norms (general, male, and female versions) or clinical norms. Special indexes are incorporated into the BASC to assess the validity of each respondent's responses (F index). Critical items may be interpreted individually on the Parent and Teacher Rating Scales. On the SRP, patient responses are also viewed on an L (“fake good”) index and the V index designed to detect invalid responses caused by poor reading comprehension, failure to follow directions, or poor contact with reality.

The BASC offers the clinician a psychometrically well-developed behavioral measure that takes only 10 to 20 minutes to complete. An advantage of the BASC, compared with many other rating scales, is the clear distinction between attention and hyperactivity symptoms for children suspected of having attention-deficit/hyperactivity disorder. The BASC thus provides information in a format to support clinical subtyping of attention-deficit/hyperactivity disorder. A disadvantage is the amount of time necessary to score and develop the profiles for each rating scale by hand. Computer scoring is likely the most cost-effective way to make use of this scale, to avoid spending undue time on tedious scoring procedures.

Achenbach Child Behavior Checklists The checklists developed by Achenbach have been perhaps the most widely used behavioral rating scales in child and adolescent clinics in recent years. Like the BASC, the Achenbach scales include a parent rating (the Child Behavior Checklist, [CBCL]), a teacher rating (Teacher Report Form, [TRF]), and a self-report (Youth Self-Report, [YSR]). The CBCL is appropriate for children 4 to 18 years of age; the TRF, for children 5 to 18 years old; and the YSR, for those 11 to 18 years old. Each scale is interpreted in comparison with a large normative sample stratified by age and sex. A cross-informant computerized scoring paradigm is provided to assist with comparisons of the CBCL, TRF, and YSR measures for a given subject.

A version of the CBCL for toddlers (CBCL/2–3) was developed in 1992 and has gained in popularity since that time. It includes five of the subscales from the CBCL and adds a subscale on Sleep Problems. A separate computerized scoring system is available for this version of the CBCL.

Other Behavioral Personality Approaches Many other behavioral approaches to assessment are available in addition to behavior rating scales, as discussed above. Direct observations of child and adolescent behavior can be a useful adjunct to other assessment procedures, whether the observation is unstructured or structured according to a specific format. For instance, Russell Barkley and colleagues developed a structured observational paradigm, the Restricted Academic Playroom Situation, designed to provide an analog situation for observing and recording symptoms of attention-deficit hyperactivity disorder during individual academic work. In this paradigm, the child or adolescent doing independent math problems is observed in a clinic playroom with toys or age-appropriate materials present. The child is told to complete as many math problems as possible, to remain on the chair at the table, and not to touch the toys or materials. Interval ratings are made in five behavioral categories: off task behavior, fidgeting, being out of seat, vocalizing, and playing with objects. Many structured observational paradigms offer normative data against which to compare the clinical behaviors of interest, thus helping the clinician to place the observational information within a broader context.

REFERRAL FOR PSYCHOLOGICAL ASSESSMENT

Interpretation of personality tests requires specialized training and experience in psychology. Although certain rapid and narrow band personality instruments (e.g., Beck Depression Inventory) can be administered and interpreted without extensive training, more wideband objective tests and most projective personality instruments require considerable training in psychology. For this reason, psychiatrists and other physicians should refer to psychologists for interpretation of personality tests. Referral is especially important in medicolegal situations.

SUGGESTED CROSS REFERENCES

Chapter 2 discusses clinical assessment and approaches to diagnosis in neuropsychiatry. Chapter 5 presents statistics and experimental design. Chapter 8 discusses theories of personality that are derived from philosophy and psychology. Chapter 7 discusses the psychiatric interview, history, and mental status examination. Section 7.6 and Section 7.7 present neuropsychological assessment of adults and children respectively. Cognitive therapy is presented in Chapter 30.6.

SECTION REFERENCES

Achenbach TM: *Manual for the Child Behavior Checklist and 1991 Child Behavior Profile*. University of Vermont, Department of Psychiatry, Burlington, 1991.

Achenbach TM: *Manual for the Teacher's Report Form and 1991 TRF Profile*. University of Vermont, Department of Psychiatry, Burlington, 1991.

Achenbach TM: *Manual for the Youth Self-Report and 1991 YSR Profile*. University of Vermont, Department of Psychiatry, Burlington, 1991.

Achenbach TM, McConaughy SH, Howell CT: Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychol Bull *101:*213, 1987.

*Anastasi A: *Psychological Testing*. MacMillan, New York, 1988.

*Archer RP: *MMPI-A: Assessing Adolescent Psychopathology*. Erlbaum, Hillsdale, NJ, 1992.

Beck SJ, Beck AG, Levitt EE, Molish HB: *Rorschach's Test, 1. Basic Processes*. Grune & Stratton, New York, 1961.

Butcher JN: *Computerized Psychological Assessment*. Basic Books, New York, 1987.

*Butcher JN: *The MMPI-2 in Psychological Treatment*. Oxford University Press, New York, 1990.

Butler RW: Personality assessment of adults and children, In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Caldwell AB: *MMPI Supplemental Scale Manual*. Caldwell Report, Los Angeles, 1988.

Corcoran K, Fischer J: *Measures for Clinical Practice*. The Free Press, New York, 1987.

Cushman LA, Scherer MJ: *Psychological Assessment in Medical Rehabilitation*. American Psychological Association, Washington, DC, 1995.

Dahlstrom WG, Walsh GS, Dahlstrom LE: *An MMPI Handbook*. University of Minnesota Press, Minneapolis, 1975.

*Exner JE, Weiner IB: *The Rorschach: A comprehensive system*, vol 3. Wiley, New York, 1995.

Gilberstadt H, Duker J: *A Handbook for Clinical and Actuarial MMPI Interpretation*. Saunders, Philadelphia, 1965.

Graham JR: *The MMPI: A Practical Guide*. Oxford University Press, New York, 1978.

Graham JR: *MMPI-2 Assessing Personality and Psychopathology*. Oxford University Press, New York, 1990.

Groth-Marnat G: *Handbook of Psychological Assessment*. Wiley-Interscience, New York, 1990.

Haworth, MR: Children's Apperception Test. In *Projective Techniques for Adolescents and Children*, AI Rabin, editor. Springer, New York, 1986.

Hersen M, Bellack AS: *Dictionary of Behavioral Assessment Techniques*. Pergamon, New York, 1988.

Kazdin A, Esveldt-Dawson K, Unis A, Rancurello M: Child and adolescent evaluations of depression and aggression in psychiatric inpatient children. J Abnorm Child Psychol *11:*401, 1983.

*Kazdin A, French N, Unis A, Esveldt-Dawson K: Assessment of childhood depression: Correspondence of child and parent ratings. J Am Acad Child Adolesc Psychol *22:*157, 1983.

Maris RW, Berman AL, Mattsburger JT, Yufit RI: *Assessment and Prediction of Suicide*. Guilford, New York, 1992.

Marvish ME: *The Use of Psychological Testing for Treatment Planning and Outcome Assessment*. Erlbaum, Hillsdale, NJ, 1994.

McArthur DS, Roberts GE: Roberts Apperception Test for Children Manual. Western Psychological Services, Los Angeles, 1991.

*Mitrushina MN, Boone KB, D'Elia LF: *Handbook of Normative Data for Neuropsychological Assessment*. Oxford University Press, New York, 1999.

Stanger C, Lewis M: Agreement among parents, teachers, and children on internalizing and externalizing behavior problems. J Clin Child Psychol *22:*107, 1993.

*Steer RA, Ball R, Ranieri WF, Beck AT: Dimensions of the Beck Depression Inventory-II in Clinically Depressed Outpatients. Clin Psychol *55:*1, 1999.

Zlotogorski Z: Recent research on the Rorschach Test with children. In *Projective Techniques for Adolescents and Children*, AI Rabin, editor. Springer, New York, 1986.

7.6 NEUROPSYCHOLOGICAL AND INTELLECTUAL ASSESSMENT OF CHILDREN

IDA SUE BARON, PH.D. AND EILEEN B. FENNELL, FH.D.

Neuropsychological and intellectual assessment of children involves a systematic application of standardized procedures. The testing produces results that enable the neuropsychologist to make an informed opinion about brain-behavior relationships or about behavioral interrelationships when direct inference about the status of the central nervous system cannot be made. The testing is generally requested as part of a multidimensional evaluation that includes other medical, educational, and psychological data. The contribution of neuropsychological testing is that it offers unique data that sharpen appreciation of the bases for a variety of deficits or dysfunction. These may range widely (e.g., from mild developmental delay of uncertain etiology to pronounced cognitive dysfunction as a result of disease or trauma). Neuropsychological data also allow temporal judgments, because long-standing problems present with a different pattern or profile of functioning than those attributable to immediate or recent events. Additionally, obtaining comprehensive baseline data enables future documentation of interval change consequent to a specific treatment regimen, such as the late effects of systemic therapy for an oncology patient, the protracted recovery course following closed head injury, the prognosis for a child whose seizures are well controlled with antiepileptic drugs, the pattern of preserved and residual dysfunction following surgical intervention for a brain tumor.

It is useful to have a framework to understand the significant contribution of neuropsychological and intelligence assessment for children with various disorders and conditions. This increases the probability that appropriate referrals are made, that the nature of the evaluation is understood, that results are integrated appropriately, and (most importantly) that treatment recommendations are implemented in concert with other relevant data acquired about the child. The neuropsychological evaluation is likely one of several consultations or procedures requested. Its strength is in the meaningful recommendations that are individually based and considered with respect to the child's environment (e.g., school, hospital, rehabilitation center), the resources available, the child's chronological age and developmental level, and whether it serves clinical or research purposes.

Importantly, neuropsychological testing determines both strengths and weaknesses in cognitive and emotional functioning. Understanding both capabilities and deficits is necessary to delineate how neuropsychological data can contribute practically to treatment recommendations and cognitive rehabilitation (i.e., have ecological validity). The concluding interpretation of a neuropsychological assessment depends on the psychologist's ability to combine clinical observations with the results of a structured objective examination to make a clinical judgment about a child's distinctive cognitive, behavioral, and emotional functioning.

Testing also provides an opportunity to appraise intractable behavior that is not likely to be ameliorated despite intervention. For example, a teenager whose sloppy, dysgraphic handwriting continues to be a source of personal frustration and low grades is unlikely to make significant gains in a treatment program designed to improve fine motor control, as might be expected for a younger child. Teachers need to reduce their emphasis on the quality of written production and instead judge the teenager on the content of what has been learned. At this age, recommendations for improving keyboard skills and teacher acceptance of computer-assisted written output, for instance, may be more advantageous to a teenager's academic performance and self-image than continued insistence on a mode of production never mastered sufficiently compared with peers despite years of attempts.

FACTORS UNIQUE TO TESTING CHILDREN

Children bring a number of complicating factors to the evaluative process that makes their evaluation quite unlike an evaluation of an adult. Among these are the influences of maturational change, brain reorganization, behavioral variability, differing diagnostic considerations, different brain-behavior rules, and the highly relevant influence of environment, personality, and psychosocial factors.

Maturational Change For the adult, deficit is commonly expressed as a loss of function. Maturation is complete, and a neurological insult has the effect of making an aspect of the adult's behavioral repertoire dysfunctional. For example, an aphasia may follow a cerebral vascular accident in the cerebral hemisphere principally mediating language function. In contrast, a child with a vascular lesion in the left hemisphere who has not reached developmental maturity may show neurobehavioral effects that are far more variable.

If a child sustains a cerebral insult, a number of different outcomes may be observed. There may be no appreciable effects of the insult; that is, the effects of interference in normal development are not immediately discernible even with careful assessment. However, dysfunction may be detected when the child “grows into the deficit” at a later developmental stage. For example, poor academic achievement in a highly visual subject such as geometry may result from head trauma at a young age (impaired visuoperceptual function) or from an inherited learning disability that is only manifested when the academic task requirements are complex enough to make this disability evident. A second possible outcome is delayed development. For example, language acquisition may slow after a viral illness. A third possibility is that development is halted. No further progress occurs in a particular functional area. A fourth possibility is that behavior regresses. Loss of previously acquired skills because of a loss of functional ability is analogous to loss of function in the adult. For example, motor developmental milestones accomplished at age-appropriate times may be lost as intracranial pressure increases due to hydrocephalus, or language may regress or disappear in the presence of an acquired epileptiform aphasia.

Brain Reorganization The potential for brain organization (plasticity) is often dramatic in childhood, and a large literature attests to the potential for recovery of function after childhood disorder or dysfunction. However, cognitive recovery in childhood, while impressive, is not necessarily complete. The full implications of a neurological insult may not be apparent for years. Further, it is not unusual for a child's recovery to follow a different course than that of an adult. For example, a child's prognosis following severe head trauma and prolonged coma, hemispherectomy or lobectomy, or early lateralized deficit can differ predictably and dramatically from that for an adult.

Behavioral Variability Children express more behavioral variability during the test session than is typical for adults. The constraints, rules, and limit setting required during the child's test session far exceed that usually required when testing an adult. Determination of when a best performance is being obtained may be more difficult, and evaluating motivation and appropriate concern about performance is not easy. This becomes apparent, for example, when reevaluation of a child previously tested elsewhere reveals a deliberate withholding of information with the previous examiner because of personal dislike for the examiner or the testing situation or in defiance of parental pressure for testing.

Different Diagnostic Considerations Evaluation of a child requires diagnostic considerations different from those for an adult. Children present with a different set of symptoms from those typical of adults with the same condition. For example, brain tumor in a child more often presents acutely with headache, nausea, and vomiting rather than with gradual cognitive changes or psychiatric symptoms, or a child's seizure type may change with increasing chronological age, whereas seizure type in an adult is more stable. Expectations about cognitive functioning based on a typical adult presentation may not be accurate for a child.

Different Brain-Behavior Rules Established brain-behavior rules derived from empirical study of adults do not readily translate to children. The effects of interference in brain development may be appreciable but do not follow the rules associated with adult brain-behavior relations. For example, in some circumstances a small focal lesion may result in diffuse disorder rather than being limited to a discrete behavioral presentation. Unlike in adults, one cannot identify specific brain regions in children as dysfunctional. It is even difficult to make direct reference to lesions in the substrate of the brain. This has been the source of much misinformation and confusion in interpretation of test results. For example, the adult assessment literature often refers to tests of frontal lobe function; however, these same tests and other tests that superficially appear to be measuring the same behavioral indexes cannot reasonably ascribe frontal lobe dysfunction to a child. Specificity of a test to a brain region is more problematic for the very young child but in some instances may be considered if the child is over 11 years of age. Further, children have

different clinical manifestations of underlying neurological compromise from adults. Therefore, the same cause may not yield the same profile of functioning for the child and the adult.

Environmental, Personality, Psychosocial Factors Environmental, personality, and psychosocial factors strongly influence a child's behavior. These factors may affect the expression of a deficit, may confound the assessment of cognitive sequelae of brain impairment, and may influence the choice of treatment recommendations. A consideration of context and how it affects cognitive, behavioral, and emotional functioning is increasingly recognized to be extremely important.

Thus, the conclusions one can reach after a neuropsychological evaluation also differ for a child and an adult. Rather than identify neural system failure or focal abnormality or address the brain substrate specifically, the intent in a child evaluation is to describe the child's full range of capabilities so that needed intervention and remediation of targeted problems can be planned. These interventions must be appropriate for the child's specific developmental level, and neuropsychological data are useful in establishing where the child is on the developmental continuum.

H. Gerry Taylor and Jack Fletcher discuss four common fallacies that lead to misinterpretation of the capacity of neuropsychological procedures to evaluate cerebral status in children. These are related to the differences in assessing children and adults. The fallacies are that (1) procedures from adult neuropsychology are differentially sensitive to childhood brain disorders, (2) tests developed for adults measure similar abilities in children, (3) the nature or pattern of deficits associated with learning or behavioral problems can be assumed to signal abnormal brain status, and (4) dysfunction describes the brain rather than the behavior. Careful attention to these potential interpretive fallacies is critical when one requests or receives the results of a child's neuropsychological evaluation.

DEVELOPMENTAL MODELS

No life-span model of neuropsychological development exists. Such a model ideally would provide continuity in understanding developmental progression or plateau in neurological disorders affecting normal development, encourage a broader consideration of function and methodologies, and encourage definition and differentiation of functions during active stages of developmental gain or delay rather than relying on post hoc analyses. Such a model would also allow formulation of age-related specific recommendations that might positively influence developmental course and provide a basis for outcome measurement. Further, such a model would encourage expansion of knowledge about the natural history of normal and abnormal brain-behavior relationships in childhood and encourage innovations or new techniques that better science and practice.

A few developmental models have been tentatively proposed in the child neuropsychology literature despite the inherent obstacles to their development. These have generally been directed toward older children, adolescents and, more rarely, infants and young children. A functional organizational approach has been suggested to separate the different types of variables underlying a developmental neuropsychology focused on behavioral theory and research. An emphasis was rightly placed on the importance of development and the active processes of change. The neuropsychologist's role was conceptualized as extending beyond simplistic inferences about brain status based on behavioral test results or observations toward a focus on relations among the manifest disabilities and basic competencies, and on relations between behavioral variables and central nervous system factors. A later assessment model by Taylor and Fletcher provided an overview of the behavioral and biological levels of analysis common in the neuropsychological assessment of children. Four components of assessment were presented: (1) manifest disability, a description of the child's failure to master age-appropriate behavior; (2) assessment of the child's cognitive or psychosocial traits that affect the manifest disability; (3) evaluation of relevant environmental variables that relate directly to the psychosocial traits; and (4) evaluation of biological variables that relate to, and influence, the cognitive traits. As the field of child neuropsychology matures, the usefulness of a developmental model to guide assessment and intervention practices will likely be demonstrated.

REFERRALS

Referral questions are important because they determine the focus of the initial stages of a psychological assessment. The psychologist needs clarity and completeness in the referral question to understand the purpose of the neuropsychological consultation and for presumptive diagnostic considerations. Since questions raised by the referral source affect the early phases of the test session, they should be stated in some detail. For example, a referral that states “child with heart disease and probable attention-deficit/hyperactivity disorder, please evaluate” provides some direction. But it is clearly insufficient compared with a referral that is more complete—for example, “5-year-old white male with a history of coarctation of the aorta repair requiring neonatal resuscitation on several occasions. Currently he has school issues (e.g., spaciness) yet academically is doing well. He has difficulty remembering routine tasks. Please evaluate for possible neuropsychological deficit.” The difference between these two statements is obvious; the former provides minimal direction, while the second identifies a number of serious medical problems with potential clinical significance. The direction one should take with the patient based on the first referral is indeterminate. The second referral raises immediate concerns about late effects of serious cardiac abnormality and early oxygen deprivation, the possibility of epileptiform activity, executive function deficit associated with early brain insult, and involvement of subcortical structures important for learning and retrieval of information.

Referral Questions in Psychiatry Referral questions about psychiatry patients may originate for varying reasons. As pointed out by Rebecca Rieger and the first author of this section, four types of questions are typically asked. The first question is whether the observed behavior is primarily a consequence of cerebral-processing dysfunction (i.e., whether a contributing organic element confounds or contributes to the clinical presentation). This question asks whether the observed behavior has a basis in neuropsychological impairment that was not fully appreciated in other neurodiagnostic investigations. For example, impaired social interactions and inappropriate responses to social cues characterize some children's behavior, including that of the child with Asperger's disorder. Also, hallucinations may result from primary psychiatric cause or an undiagnosed complex partial epilepsy. Thus, the observed behavior may be illuminated by consideration of a strong neurological or neuropsychological basis.

A second question is whether the clinical picture involves comorbidity. One must understand the impact of one disorder on another—for example, how attention–deficit/hyperactivity disorder affects a child with a mood disorder or whether there is documentation of neuropsychological impairment related to chronic substance abuse in a teenager with a conduct disorder. Coexisting conditions have obvious differential treatment considerations, but understanding their interaction is essential.

The third question is whether data can confirm that development is normal or whether the data suggest that the child's development has been delayed or inhibited as a consequence of neurological impairment. The neuropsychologist must be familiar with the stages of normal child psychological and neuropsychological development and the dynamic transitions that characterize movement from one developmental level to the next. The neuropsychologist must also recognize truly abnormal development and not ascribe a deficit to behavior that is normal for the child at that developmental stage.

A fourth question is how the neuropsychological data influence treatment and rehabilitation planning. For example, does the child have sufficient expressive language, abstract reasoning, and judgment for highly verbal (e.g., psychoanalytic) therapy? Recommendations may suffice after one evaluation. However, neuropsychological data also allow fine-grained serial testing. This allows assessment of the effects of a specific treatment regimen or clinical population for research purposes or serial reevaluation of a single child for clinical purposes.

Review of the referral question is only a first step in the exploratory process associated with child testing. Adjustments are made as the testing session progresses. For example, if tests requiring speed and nonverbal perception are less well performed than verbal tests that have no time constraints, evidence of more-specific types of visuoperceptual or visuospatial deficit may be sought. Alternatively, one may ask whether the speeded psychomotor performance and organization of action required by these tasks determine the low scores, whether periods of inattention vary within a brief time interval and within a test independent of the sensory modality being assessed, or whether the acquisition of new information and its retrieval over acute and long-chronic conditions are impaired. Thus, a continuous testing of alternative hypotheses occurs in response to behaviors observed clinically as well as to those revealed by formal test instruments. The formal neuropsychological evaluation provides an opportunity to establish whether evidence of dysfunction exists that was not detected by other neurodiagnostic procedures or never elicited in clinical interviews or clarified by other data obtained in a careful history taking and record review. A neuropsychological examination attempts to learn how the brain functions by asking the child to demonstrate many different types of behavior. Doing so yields a dynamic picture of the brain as it works to accomplish these functions, and observing such dynamic behavior, one can identify problems the child is experiencing in accomplishing these tasks as well as determining how well the child can perform a wide variety of cognitive tasks.

Referral Populations Certain patient populations are appropriately referred for neuropsychological evaluation, while others benefit less from this consultation. A child with a history of developmental delay in language or visuomotor functioning should be assessed in the preschool years (ages 3 to 5 years) to determine the type and degree of delay, the specific recommendations needed (e.g., speech and language, occupational, or physical therapy) and the risk for problems with learning and achievement on school entry. A child with a significant delay in mental or cognitive development may need to be assessed for appropriate school placement according

to state guidelines. Children who have sustained direct insult to the central nervous system (e.g., brain tumor, vascular malformation, viral or bacterial infection, or cerebrovascular infarction) should be assessed to determine the effects of these cerebral insults on their intellectual and neuropsychological development. Children who are born with malformations of the brain or spinal cord or who have genetic or metabolic disorders that can affect normal development need to be assessed for the degree to which these developmental disorders have affected brain growth and functional maturation. Many types of seizure disorders of childhood and antiepileptic drugs can directly disrupt development. Children with medical disorders such as acute lymphocytic leukemia that require prophylactic chemotherapy or radiation are at risk for developing neuropsychological dysfunction secondary to the treatment regimen. Chronic medical conditions such as renal failure, diabetes, or chronic heart disease may affect a child's brain functions. In a number of developmental disorders (e.g., learning disabilities, attention-deficit/hyperactivity disorder, Tourette's disorder, childhood anxiety, and depression), the question of a comorbid disorder may arise. A variety of other instances may lead to a request for neuropsychological evaluation, such as the child with disruptive or aggressive behavior or conduct disorder, as well as those who manifest some features of a pervasive developmental disorder. A less common (and likely less helpful) referral is the moderately to severely retarded child. In the extreme low range of intellectual functioning, differential identification of strengths and weaknesses may be difficult because of the child's limited cognitive abilities. The usefulness of traditional child neuropsychological assessment in pervasive developmental disorders, childhood schizophrenia, or mood disorders awaits proved demonstration. Research has attempted to identify specific potential brain dysfunctions in these disorders with neuropsychological research methods. The usefulness of a typical neuropsychological examination for these clinical populations requires further investigation and clarification.

Once the child is evaluated, a written report summarizing the neuropsychological evaluation is prepared. This report addresses the concerns stated in the reason for referral and confirms or disconfirms the original considerations. The conclusions may suggest new possibilities or amend incomplete or erroneous preliminary assumptions. The conclusions should elaborate on the child's full range of abilities, providing detail about how the intact or impaired abilities affect diverse cognitive functions and extrapolating to predictions about the child's behavior in real-life situations.

TEST APPROACH

A variety of acceptable approaches to neuropsychological test selection exist. Most commonly encountered are those that involve (1) a fixed battery, (2) a battery and supplemental tests, (3) a process approach, and (4) an eclectic or flexible approach. Proponents of the fixed battery administer a specific, unchanging set of tests that generally have been normed on the same population. The range of tests within the battery is assumed to sample sufficient behavior for valid conclusions. The widely recognized Halstead-Reitan Neuropsychological Test Battery is one such battery. Scores obtained on a battery or within a test containing multiple subparts may be comparable (i.e., they are scaled similarly and standardized on the same population) or they may need to be converted for commonality of meaning. However, inflexible test batteries are not useful for the neuropsychologist who is concerned with individualizing an approach on the basis of specific referral questions. In general, most fixed batteries have been designed to answer a specific set of questions. For example, the Halstead-Reitan battery is designed to answer the following questions: Is there an organic brain lesion? Is it focal or diffuse? Is it chronic or acute? Is it lateralized? Is it localized?

Proponents of administering a test battery supplemented by additional tests choose complementary tests that cover weaknesses in the core battery. A core battery may be particularly insensitive to the condition being assessed, and the addition of other test measures successfully expands the database. For example, the Halstead-Reitan Neuropsychological Test Battery fails to sufficiently tap functions often compromised by closed head injury. The addition of tests of new learning and memory for different modalities (e.g., auditory and verbal and visual encoding, storage, and retrieval) and of attention and executive function reduces the likelihood that important cognitive sequelae of a head injury will be missed. Neuropsychologists who endorse a process approach emphasize the utility of considering qualitative aspects of a performance in addition to quantitative aspects. They also value testing-the-limits procedures to better understand how someone solves a problem or reacts to additional specific probes or information about the task. The usual quantitative data are then supplemented by alternative scoring and careful recording of the process that the child used to achieve the desired end point. Systematically highlighting qualitative behavioral features of the child's behavior may add valuable information that clarifies how a child integrates information and produces a specific behavioral response.

The eclectic, (or flexible), approach is perhaps the most increasingly common procedure endorsed by neuropsychologists. Tests are chosen on the basis of the specific functions they are likely to assess. Tests may be selected from established fixed batteries, but complete batteries are generally not used. Often the neuropsychologist uses a personalized selection of core tests as an efficient screening battery, which is then elaborated on during the test session, in response to the child's behavior during the session and the particular profile that emerges as the session progresses.

Both the process and the eclectic or flexible approaches have special utility in child assessment. Both approaches encourage the neuropsychologist to attempt to evaluate functional areas for which clinical instruments are not available or are not well standardized. Reliance is encouraged on a creative application of key developmental concepts through informed observation, directed interview with significant adults, and the use of more naturalistic activities.

No matter which approach is preferred, testing should allow the neuropsychologist to highlight specific indexes of impairment and ascribe a useful treatment plan to ameliorate the underlying problem that generated the referral. With data from multiple sources, including real world observations or reports, hypotheses about functional strengths and weaknesses are more easily translated into intervention recommendations.

TEST SELECTION

Neuropsychological testing involves administering instruments that range widely in the specificity or generality of what they purport to assess. The tests also range in their applicability to different clinical populations. Some tests that are sensitive to impairment resulting from one condition may be insensitive to the effects of another. For example, tests that may reveal a reading disorder may not prove as useful for the child who has had a traumatic brain injury. In the former case, extensive evaluation of aspects of receptive and expressive language take precedence over tests that target functions especially vulnerable to change. In the latter case, tests of functions that are most vulnerable to change as a consequence of an acquired insult are needed, such as those that evaluate mental control, attentional components, new learning, and flexible problem-solving behavior and information processing.

Because a single test result rarely confirms either normality or dysfunction, conclusions are reached only after consideration of all available data. Current test results are weighed in relation to historical information including prior testing, data from other allied health evaluations, and data from appropriate statistical sources that serve as a comparison group. Intact functioning may be suggested by normative sample comparisons, but impaired functioning may be documented when intraindividual comparisons are made. For example, a child may obtain an average verbal intelligence quotient (I.Q.) score and average scores on tests of receptive and expressive language. However, other test data may document above-average functioning on tests that do not have a strong language component, which suggests that comparison only with the normative group is insufficient and that average function for this child indicates relative weakness. Thus, average function is an insufficient basis for concluding there is no impairment. Tests of right and left upper extremity motor and sensory-perceptual function may find evidence of right-sided impairment, supporting a hypothesis of compromised left cerebral function. These data attest to differences in functioning between the child's right and left sides and are independent of conclusions reached by comparing the child with other children.

Further, tests have different applicability depending on whether their principal use is for clinical evaluation or research investigation. Research test batteries may include some commonly used intellectual and neuropsychological measures that may be modified, shortened, or adapted to answer a specific research question. For example, many research protocols do not use the complete Wechsler Intelligence Scale for Children (WISC) to assess intellectual functions but instead include two or three subtests that are highly correlated with Verbal, Performance, or Full Scale I.Q. Similarly, selected aspects of a particular function may be emphasized rather than the full range of measures (e.g., verbal list learning). A research protocol may include experimental tasks which are neuropsychological in nature but are designed to answer a limited or specific question, such as how much does verbal recitation of a word list disrupt speed of motor tapping. Finally, a research battery may be specifically designed to be administered serially and may, for example, only measure speed, tracking, or attention. As a result, a research battery may not cover the full range of assessment measures that may be covered in a typical clinical examination.

DOMAINS OF TESTING

The field of psychological assessment is expanding; more well-normed tests exist and more behaviors can be assessed. There is an increasing transition to looking at subdomains of some cognitive areas and their importance and relevance for different clinical populations. Cognitive domains are partitioned in an effort to understand preservation of function and identify functions that are vulnerable. As a result, it is no longer acceptable to obtain only general intelligence data (i.e., mental age compared with chronological age) or administer a few broad tests that have limited or no validity with respect to specific brain functions. Current procedure is to examine discrete functions not routinely examined in the recent past, such as attention and alertness, executive function, learning and retrieval, affective social interaction, and communication and language readiness. The neuropsychologist chooses to assess a number of cognitive domains of function with varying degrees of

completeness. Table 7.6-1 presents a list of the common domains or broad areas of function that are commonly assessed in a clinical neuropsychological evaluation.

General intelligence
Academic achievement
Executive functioning
Attention, concentration, orientation
Receptive and expressive language
Sensory-perceptual functioning
Motor functioning
Visuospatial analysis and constructional skills
Learning and retrieval
Personality and behavioral assessment
General development and adaptive behaviors

Table 7.6-1 Domains Assessed in a Neuropsychological Evaluation

Areas of function assessed in a neuropsychological evaluation are listed in Table 7.6-2 along with a sampling of tests appropriately considered within each broad neuropsychological domain and a summary description of what they measure. (The tests listed are intended only as examples of tests that measure an aspect of the specific domain within which it falls.)

Table 7.6-2 Selected Neuropsychological Tests and Function Assessed by Cognitive Domain

Tests of Intelligence and Achievement Generally, a comprehensive evaluation screens each relevant domain to establish integrity of function. The decision to administer more detailed testing within a domain is based on the referral question, the response to screening measures, and behaviors observed during the test session. A growing number of tests are available for inclusion in an individually designed test session. The choice depends on examiner experience and preference.

Table 7.6-3 lists commonly used intelligence and academic achievement tests. These tests vary, and the choice of instrument is based on age of the child, available standardization data, conceptual basis, and range of subtests available to address the referral question. For example, the Stanford-Binet Intelligence Scale is appropriate for children beginning at 2½ years of age. It yields a composite I.Q. score and scores for four scales: verbal reasoning, abstract visual reasoning, quantitative reasoning, and short-term memory. The Kaufman Assessment Battery for Children is applicable for children 2½ to 12½ years of age. It makes a conceptual distinction between simultaneous and sequential processing. A composite score and scores for each of these processing concepts are derived. The McCarthy Scales of Children's Abilities has normative data for children between 2½ and 8½ years of age. It allows for computation of a General Cognitive Index, based on results of three other indexes: verbal, perceptual-performance, and quantitative. It also yields a memory index and a motor index. The Test of Nonverbal Intelligence is a language-free measure of intelligence, aptitude, abstract reasoning, and problem solving, appropriate for children with communication problems or those who do not read or write well. It is similar to the Raven's Matrices Test but was normed in the United States. It is appropriate for children aged 6 years or older.

Intelligence tests
Bayley Infant Scale of Development–2nd edition (1–42 months)
McCarthy Scales of Children's Intelligence (2.5–8.5 years)
Kaufman Assessment Battery for Children (2.5–12.5 years)
Stanford-Binet Intelligence Test, 4th edition (2.5 years–adult)
Wechsler Preschool and Primary Scale of Intelligence–Revised (4–6.5 years)
Wechsler Intelligence Scale for Children–III (6–16 years)
Test of Nonverbal Intelligence–3 (6–89 years)
Raven's Progressive Matrices (Coloured or Standard) (5.5 years–adult)
Academic achievement tests
Wide Range Achievement Test–3rd edition
Woodcock Johnson Psycho-Educational Test Battery–Revised
Peabody Individual Achievement Test–Revised
Wechsler Individual Achievement Tests
Bracken Basic Concept Scales (school screener)

Table 7.6-3 Selected Intelligence and Academic Achievement Tests

Intelligence and achievement tests are not validated with respect to brain function, but they provide valuable data for a neuropsychological evaluation. The Wechsler Intelligence Scale for Children–III (WISC-III) is the most commonly administered general intelligence test for children at least 6 years of age. It is a battery comprising 13 subtests (10 core tests and 3 alternates). Three summary I.Q.s are obtained: (1) the Verbal I.Q., based on five verbal subtests; (2) the Performance I.Q., based on five performance subtests; and (3) the averaged index of general intellectual functioning, the Full Scale I.Q. Language expression, comprehension, and application of verbal skills in problem solving are assessed within the Verbal Scale. Nonverbal visuoperceptual analysis, synthesis, expression, psychomotor speed, and visuomotor efficiency are assessed within the Performance Scale. Neither scale assesses several important factors that influence successful performance, such as motivation, creativity, organizational skill, or study habits. Four additional index scores can be obtained. The Verbal Comprehension Index is based on performances on the information, similarities, vocabulary, and comprehension subtests. The Perceptual Organization Index is based on performances on the picture completion, picture arrangement, block design, and object assembly subtests. The Freedom from Distractibility Index is based on arithmetic and digit span performances. The Processing Speed Index is based on coding and symbol search performances.

The individual subtests may be particularly useful because each depends on a variety of capabilities, and dysfunction of any one could result in a low score. Possible deficit is then suggested that can be more directly assessed with specific neuropsychological test measures. For example, a low score on the picture arrangement subtest that requires sequencing nonverbal material may suggest visual organization problems. This could be explored further with tests of visuomotor organization such as copying complex designs, tests of the motor and sensory integrity of the right side of the brain, graphomotor sequencing and upper extremity motor sequencing, and other tests with a high sensitivity to visual planning, organization, and production deficits. Thus, the pattern, or profile, of the WISC-III subtest scatter may illuminate a neuropsychological problem, but it does not confirm the problem. It is not common for an individual subtest score to convincingly highlight neuropsychological dysfunction in the child, as it might in the adult. Additional confirmation that these data provide accurate evidence of a problem is necessary.

There are six subtests within the Verbal Scale:

Information requires that the child provide factual information learned over time, both at school and at home, about events, people, objects, and places.

Similarities requires abstract verbal reasoning and verbal expression of the underlying concept connecting word pairs.

Arithmetic requires mental manipulation of numbers (i.e., numerical reasoning) to respond orally and attention and concentration.

Vocabulary assesses word knowledge and verbal fluency.

Comprehension requires common sense and socially appropriate solutions for a variety of practical social situations.

Digit Span assesses immediate recall of number strings (forward and backward) and attention and concentration for auditorily presented stimuli.

There are seven subtests within the Performance scale:

Picture Completion requires attention to essential visual detail and visual discrimination.

Coding involves copying in response to a key set of number-symbol pairs, visuomotor coordination, speed, and attention and concentration.

Picture Arrangement requires appreciation of social nuance in a series of pictures and planning a logical sequence.

Block Design requires visuospatial analysis and synthesis and abstract visual problem solving.

Object Assembly requires visuoperceptual analysis and synthesis of pieces of recognizable objects into a meaningful whole.

Symbol Search requires visuomotor speed, visuoperception, attention and concentration, and vigilance.

Mazes requires visuomotor planning, attention and concentration, and vigilance.

Self-Report and Parent-Report Tests A large number of tests and procedures exist that the neuropsychologist may administer to the child or the parents to determine the type, frequency, and severity of behavioral problems or symptoms that the child may manifest. Table 7.6-4 provides a list of commonly used self-report and parent-report instruments and tests and procedures administered to parents to identify the child's specific developmental attainments and adaptive behaviors.

Personality, behavior, and symptom report
Child Behavior Checklist
Personality Inventory for Children–Revised
Reynolds Depression Scale (adolescents)
Children's Depression Inventory (6–13 years)
Children's State-Trait Anxiety Scale
Revised Children's Manifest Anxiety Scale
Harter Scales of Perceived Competence
Eyberg Child Behavior Inventory (oppositional-defiant/conduct disorder)
Conners Rating Scales (hyperactivity, inattention, conduct disorder)
General development and adaptive behaviors
Child Development Inventory
Vineland Adaptive Behavior Scales

Table 7.6-4 Personality, Behavior, and Symptom Report Measures and General Development and Adaptive Behaviors

Two commonly used parent measures include the Child Behavior Checklist and the Personality Inventory for Children–Revised. Both allow the parent to report a variety of social, academic, physical, intellectual, and psychiatric problem behaviors that their child may exhibit. Each measure contains a variety of subscales that assess the presence of specific problematic behaviors as endorsed by the parent. Scores are standardized on the basis of fairly large normative samples. The child's score is compared with scores of the age-appropriate normative group. While both scales were originally developed in a mental health setting, each instrument has been used in a variety of clinical settings among different pediatric diagnostic groups, and clinical sample norms are available.

Other self-report measures have been developed to assess specific childhood diagnoses or symptoms, such as depression or anxiety disorders. A similar parent-report format has also been developed for other childhood disorders such as oppositional defiant disorder, attention-deficit/hyperactivity disorder, and conduct disorder. As with many other child diagnostic instruments, it is helpful to refer to the normative sample data to determine the utility of the test in certain specific medical conditions or among particular ethnic or socioeconomic minority groups. The typical neuropsychological report includes the reference group used for specific test findings.

The effect of parent distress on parent-report measures must be considered. It is insufficient to have a parent complete these questionnaires without an adjunctive interview. The clinical interview permits the neuropsychologist to evaluate the degree to which parental distress about their child's behavior may have led them to overendorse problems. Alternatively, some parents consistently minimize or fail to report symptoms. They may be unaware of or in denial about their child's behavioral difficulties. This may have a negative impact on their willingness to institute treatments recommended from the neuropsychological evaluation.

Tests of General Development and Adaptive Functioning Tests of general development and adaptive functioning are also useful in a neuropsychological evaluation (see Table 7.6-4). These tests are designed to elicit specific information about behaviors that a parent or other caregiver has observed. These behaviors are grouped according to developmental progress, and the child's attainments are compared with those of a same-age, normative group. The Child Development Inventory, normed through age 6 years but applicable to the older, but delayed child, has subscales that assess social skills, gross and fine motor skills, language, self-help skills, and general development. The Vineland Adaptive Behavior Scales include four subdomains (communication, social skills, daily living, and motor skills) as well as a composite Adaptive Behavioral Score. The child's scores are converted into scores that identify variations from typical age-appropriate attainments, and the child's development across the subdomains can be compared with age norms. Unlike the Child Development Inventory, which has a yes-no format, the Vineland scales require extensive training in parent interviewing for accurate scoring.

Tests of Attention and Executive Functioning Sometimes tests assess aspects of functioning related to more than one domain. For example, the Trail Making Test requires the child to draw a line between numbers in sequence for Part A (i.e., 1-2-3-4-...) and to alternate between numbers and letters in sequence for Part B (i.e., 1-A-2-B-3-C-4-D...). This test has been considered both a test of attention, because it requires focus and sustained attention to complete the sequencing, and a test of executive function, because it requires simple mental shifting between the two automatic language sequences. Because it is timed, it is also used to index the speed with which these tasks are accomplished. It is in the final interpretation that the assignment of a test to a function is justified, based on the full range of behavior sampled during the assessment. As a result, a test-by-test interpretation of brain dysfunction is not acceptable clinical practice. Instead, the total performance across a variety of domains must be integrated in the summary interpretation of the examination.

Subdomains of function are receiving greater attention in the neuropsychology literature. Attention and executive functioning are two domains that have been discussed and empirically tested from the perspective of the child neuropsychologist.

The data sources for attention in infants and young children are largely the examiner's direct observations, the indirect observations reported by a reliable caretaker,

and rating of behaviors related to attention. For older children, specific components of attention have been hypothesized along with tests that most likely measure these components. Table 7.6-5 presents fractionated subcomponents of attention and of executive function that have been considered in the recent literature.

Attention	Executive function	
Focus/execute	Plan	Inhibit
Sustain	Organize	Shift flexibly
Encode	Initiate	Monitor
Shift	Sustain	Evaluate

Adapted from Mirsky AE, Anthony BJ, Duncan CC, Ahearn MB, Kellam SG: Analysis of the elements of attention: A neuropsychological approach. Neuropsychol Rev 2:109, 1991, and Denckla MB: Executive function, the overlap zone between attention-deficit/hyperactivity disorder and learning disabilities. Int Pediatr 4:155, 1989.

Table 7.6-5 Subdomains of Attention and Executive Function

Four factors were empirically identified by Allan Mirsky and colleagues. They proposed a four-factor model and associated neuroanatomical regions and tests for each factor. The first factor, *focus execute*, refers to the child's ability to scan the stimulus field and respond in a meaningful way. A deficit related to focus is suggested if the child gives the stimulus material cursory attention or cannot perform a task under distraction conditions. Tests sensitive to this perceptual motor speed factor are the Wechsler Digit Symbol and Coding subtests, the Stroop Test, the Trail Making Test, and letter and number cancellation tasks. Implicated brain regions are the inferior parietal lobe, superior temporal cortex, and corpus striatum. The second factor, *sustain*, refers to the child's ability to be vigilant and attend for a time interval. Observation of the child's ability to complete tasks on a consistent basis or noting the degree to which the child must be redirected back to a task after having been engaged by the task, provide clinical indexes of the sustain factor. The computerized Continuous Performance Test was cited for sustain. Critical brain regions include the tectum, mesopontine regions of the reticular formation, and midline reticular nucleus of the thalamus. The third factor, *encode*, refers to sequential registration, recall, and mental manipulation of information. Tests included the Wechsler Digit Span and Arithmetic subtests. Involvement of the amygdala, hippocampus, and superior temporal region was proposed. The fourth factor, *shift*, refers to the child's ability to be flexible and adaptive. The Wisconsin Card Sorting Test is sensitive to shifting capacity. Critical brain regions are the prefrontal cortex and anterior cingulate.

The literature on the development of regulatory functions across the age span and the importance of the domain of executive function have been growing. A variety of abilities are subsumed under the heading executive function. For example, the ability to initiate, sustain, inhibit, and shift was proposed by Martha Denckla. Initiate or drive refers to goal-setting and deciphering the means to an end. Sustain refers to how well one can fixate on a task and organize oneself. Inhibit is the capacity to monitor, evaluate, and adapt. Shift is the ability to change ideas in response to environmental factors.

Executive control functions are thus a collection of varying abilities that involve regulatory control over thought and behavior in the service of goal-oriented problem solving. Besides formal tests, clinical data about executive function can be obtained by observing the child's ability to problem solve in the natural environment and initiate play, by parental interview regarding the child's inhibitory abilities, by assessing how flexible a child is when faced with a changing routine, and on ratings on specific behavioral scales. Children have differential developmental trajectories for executive function subcomponents. The neuropsychologist must consider these when focusing an assessment toward those that are relevant for a particular developmental period. The problems of executive regulation of cognition and behavior are potentially relevant to children who have acquired brain injury, lead poisoning, or in utero toxic exposure and to many other diagnostic groups. Increasing attention is also being directed toward the interplay of executive functions and attention-deficit/hyperactivity disorder.

STATISTICAL CONCEPTS

A variety of statistical terms are commonly used by neuropsychologists in the test reports written after an examination. It is useful to have a familiarity with these terms and the statistical bases of test interpretation.

The expectation that a test has good reliability and good validity is not always satisfied in neuropsychology testing. *Reliability* refers to the likelihood that an individual will obtain a test score at one point in time that is similar to the score obtained if the test is readministered at another point in time. It is a reference to consistency. *Validity* refers to whether the test is measuring what it is intended to measure; for example, whether a test that seems to measure auditory attention is in fact a good test of auditory attention.

Published psychological tests are accompanied by test manuals that describes their statistical features. Included among these is the *standardization sample*, the population of subjects on whom the test was normed. The psychologist compares the examinee with the standardization sample to determine if the published normative data are a fair and representative sample for the examinee. For example, normative data may be provided for different levels of education, different cultural backgrounds, different sex, and different ages. There may also be tables that enable conversion of test scores to mental age equivalents, to relate the child's actual functioning to that of younger or older children. Not all tests have sufficient normative data to allow ideal comparisons. Therefore, psychologists also rely on a substantial research literature that addresses the usefulness of different tests with different clinical populations. Together, test manuals and empirical studies provide a more comprehensive basis, ensuring the most reliable and valid assessment of a child and appropriate analyses of the test results.

The actual score on a test is the *raw score*. The raw score conveys no comparative information and must be converted to a score that has meaning across individuals and across tests. Psychological test scores are commonly converted to standardized scores; then two or more scores can be compared on the basis of their standard deviation from a mean (i.e., average) score. The *standard deviation* (SD) is an index of the average amount of deviation of the scores from the mean. It is calculated by squaring and averaging individual deviations (the variance) and then extracting the square root of the variance. Test manuals provide the needed standard deviations, facilitating interpretation of specific scores. For example, the Wechsler series of intelligence tests are constructed so that subtest scores have a mean of 10 and a standard deviation of 3. A score deviation greater than 1 SD in either direction is significant. Allowance is made for the *standard error of measurement*, the range of how much a score can vary. This range is determined by the standard error of measurement (i.e., the actual error band around each obtained score or how a test score will vary if the test is readministered). These ranges are presented in tabular format for many standardized tests. By referring to the appropriate table within the test manual, the neuropsychologist can determine that an I.Q. of 100 falls between these two scores with either a 5 percent level of confidence (the probability that the score will normally fall outside this range is less than 5 percent) or a more restrictive 1 percent level of confidence.

For example, if a child's I.Q. is reported to be 100, the likelihood is that the child's true score falls somewhere between two scores. This range is determined on the basis of the standard error of measurement (i.e., the actual error band around each obtained score) or, practically, how a test score will vary if the test is readministered. These ranges are presented in tabular format for many standardized tests. Thus, standardized scores and the nonstandardized percentile score enable comparison of one person's results with the results of others and make intraindividual comparisons possible for performances on different tests if the standardization sample is similar. These conversions are essential for valid interpretation. Therefore, a score two points lower cannot be interpreted as indicating a poorer performance than the higher score if they are within the same acceptable range of variability.

For a difference between two subtest scores to be considered statistically significant, it must have a specific magnitude (e.g., more than 1 SD). However, because there is allowance for the standard error of measurement, a lower score cannot be interpreted as clearly indicating a poorer performance than the higher score. It is common for the lay reader of psychological reports to misinterpret a “lower” score as indicating poorer performance or impairment when both scores may fall within the same acceptable range of variability and thus be accurately interpreted by the neuropsychologist as equivalent.

Certain standard scores are more commonly referred to in psychological reports than others. One standard score that is frequently used is a deviation I.Q. The summary score obtained on the Wechsler intelligence tests is a deviation I.Q. These scores have a mean of 100 and a standard deviation of 15. Sometimes scores are reported as T-*scores*. A *T*-score has a mean of 50 and a standard deviation of 10. A conversion to a Z- *score* means that the mean is zero and the standard deviation is 1. Reporting standard scores allows the reader to consider the degree to which a child exceeds expected average performance or fails to reach the mean

level. Table 7.6-6 presents selected WISC-III I.Q. scores, their standard deviations from the mean, and their percentile rank equivalents.

I.Q. Score	Number of SDs from the Mean	Percentile Rank Equivalent
145	+ 3	99.9
130	+ 2	98
125	+ 1 2/3	95
120	+ 1 1/3	91
115	+ 1	84
110	+ 2/3	75
105	+ 1/3	63
100	0 (Mean)	50
95	− 1/3	37
90	− 2/3	25
85	− 1	16
80	− 1 1/3	9
75	− 1 2/3	5
70	− 2	2
55	− 3	0.1

Table 7.6-6 Relation of Selected WISC-III I.Q. Scores to Standard Deviations From the Mean and Percentile Rank Equivalents

A hypothetical normal distribution curve approximates score distributions. A range from –1 to +1 standard deviation (e.g., I.Q. score between 85 and 115, *T*-score between 40 and 60, *Z*-score between –1.0 and +1.0) is the range within which 68 percent of the population falls, with 34 percent falling on either side of the mean. Approximately 95 percent of the scores will fall within 2 standard deviations of the mean, and approximately 98 percent of the scores will fall within 3 standard deviations of the mean. Performance that falls 2 standard deviations below the mean or lower indicates significant negative deviation from average performance and is often assigned as a critical cutoff point. Extreme performance (either extremely low or extremely high on the normal distribution) is represented by a score that falls 3 standard deviations away from the mean in either a positive or negative direction.

Each standard score can be assigned to a nonstandard score, a percentile. For example, a score that falls at the 50th percentile is the score that can be expected to be obtained by 50 percent or fewer of the population; the person does as well or better than 50 percent of the normative population or, alternatively, 50 percent of the normative population do better.

Descriptive ability ranges are often referred to in psychological reports to summarize test performance for both the lay person and the statistically sophisticated reader. Over time, these terms have achieved a commonality among readers that makes it advantageous to report them. They are based on a statistical foundation as well, however, and should be interpreted accurately. For example, the Wechsler I.Q.s are assigned to descriptive ability ranges based on their standard deviation from the mean of 100. The ranges and comparable standard scores and percentiles for the Wechsler descriptive ability ranges are summarized in Table 7.6-7. For example, a score from 90 to 109 falls within the average range, and therefore between the 25th and 75th percentiles.

	Standard Score	Percentile Range	Population (%)
Very superior	>129	≥98	2.2
Superior	120–129	91–97	6.7
High average	110–119	75–90	16.1
Average	90–109	25–74	50
Low average	80–89	9–24	16.1
Borderline	70–79	2–8	6.7
Deficient	<70	≤2	2.2

Table 7.6-7 Descriptor Ranges, Intelligence Quotients and Theoretical Population Percentages

The reader of psychological reports may see reference to two additional terms, the sensitivity and specificity of a test. *Sensitivity* refers to how often a test correctly identifies as impaired an individual who is truly impaired. Sensitivity is given as a percentage (i.e., true positive for correct identification). *Specificity* refers to how often a test identifies a normal individual as normal. Specificity is given as the percentage of true negative, or percentage that correctly identifies normal controls.

A final type of score refers to the presence of a test behavior as a sign (pathognomonic sign) of impairment. This type of score is analogous to findings on physical examination for mental status examination that are evidence of pathology. For example, just as decreased limb tone suggests a long motor tract sign, mirror movements on a tapping test in an older child suggest the failure to develop normal suppression of motor movements across the corpus callosum. Similarly, the child whose spontaneous speech at age 8 years is notable for sound substitutions (e.g., /tw/ for /th/) is showing evidence of a developmental articulation disorder. Finally, the concept of delay versus hard signs is also useful. Typically, the *delay* implies that the behavior is normal for a younger child, whereas a hard sign implies a lateralized or focal finding that often points to pathology of a specific brain region. The concept of “delay” has sometimes been used to suggest that with time the behavior will develop to normal for age. In reality, many delayed behaviors become functional deficits as the child ages but the behavior is maintained.

SUGGESTED CROSS-REFERENCES

Supplemental approaches to testing are described in different chapters of this book. Section 7.4 describes the neuropsychological and intellectual assessment of adults. Chapter 33 details the psychiatric examination of children. Information relevant to the syndromes discussed in this section may be found in Section 2.5 on neuropsychiatric aspects of head trauma, Chapter 35 on learning disorders, Chapter 38 on pervasive developmental disorders, Chapter 39 on attention-deficit disorders, and Section 48.6 on pharmacotherapy.

SECTION REFERENCES

*Baron IS: Interpreting psychological reports. In *Current Management in Child Neurology*, BL Maria, editor. Decker Periodicals, Hamilton, Ontario, 1999.

Baron IS, Fennell EB, Voeller K: *Pediatric Neuropsychology in the Medical Setting*. Oxford University Press, New York, 1995.

*Baron IS, Gioia GA: Neuropsychology of infants and young children. In *Handbook of Human Brain Function: Assessment and Rehabilitation*, vol III: *Neuropsychology*, G Goldstein, PD Nussbaum, SR Beers, editors. Plenum, New York, 1998.

Benton AL: Behavioral indices of brain injury in school children. Child Dev *33:*199, 1992.

Bernstein JH, Waber D: Developmental neuropsychological assessment: The systemic approach. In *Neuropsychology*, AA Boulton, GB Baker, M Hiscock, editors. Humana, Clifton, NJ, 1990.

Denckla MB: Executive function, the overlap zone between attention deficit hyperactivity disorder and learning disabilities. Int Pediatr *4:*155, 1989.

Dennis M: Capacity and strategy for syntactic comprehension after left or right hemidecortication. Brain Lang *10:*287, 1980.

Dennis M, Lovett M, Wiegel-Crump CA: Written language acquisition after left or right hemidecortication in infancy. Brain Lang *12:*54, 1981.

*Fennel EB: Issues in child neuropsychological assessment. In *Clinician's Guide to Neuropsychological Assessment*, R Vaderploeg, editor. Erlbaum, Hilldale, NJ, 1994.

Fennell EB, Bauer RM: Models of inference in evaluating brain-behavior relationships in children. In *Handbook of Child Clinical Neuropsychology*, CR Reynolds, editor. Plenum, New York, 1989.

Fletcher JM: Attention in children: Conceptual and methodological issues. Child Neuropsychol *4:*81, 1998.

Fletcher JM, Taylor HG: Neuropsychological approaches to children: Towards a developmental neuropsychology. J Clin Neuropsychol *6:*39, 1984.

Frith U: *Autism and Asperger syndrome*. Cambridge, Cambridge University Press, 1991.

*Gathercole S: The development of memory. J Consult Clin Psychol *39:*3, 1998.

Lezak M: *Neuropsychological Assessment*, ed 3. Oxford University Press, New York, 1995.

*Loring D, editor: *INS Dictionary of Neuropsychology*. Oxford University Press, New York, 1999.

Mahoney WJ, D'Souza BJ, Haller JA, Rogers MC, Epstein MH, Freeman JM: Long-term outcome of children with severe head trauma and prolonged coma. Pediatrics *71:*756, 1983.

Marcotte AC, Stern C: Qualitative analysis of graphomotor output in children with attentional disorders. Child Neuropsychol *3:*147, 1997.

Mirsky AF, Anthony BJ, Duncan CC, Ahearm MB, Kellam SG: Analysis of the elements of attention: A neuropsychological approach. Neuropsychol Rev *2:*109, 1991.

Passler MA, Isaac W, Hynd GW: Neuropsychological development of behavior attributed to frontal lobe functioning. Dev Neuropsychol *1:*349, 1985.

Plomin R: Environment and genes: Determinants of behavior. Am Psychol *44:*105, 1989.

Rieger R, Baron IS: Psychological and neuropsychological testing. In *Textbook of Child and Adolescent Psychiatry*, ed 2, J Wiener, editor. American Psychiatric Press, Washington, DC, 1997.

Rourke BP: *Nonverbal Learning Disabilities: The Syndrome and the Model*. Guilford, New York, 1989.

Rourke BP, Bakker D, Fisk H, Strang J: *Child Neuropsychology: An Introduction to Theory, Research and Clinical Practice*. Guilford, New York, 1983.

*Rutter M: Developmental neuropsychiatry: concepts, issues, and prospects. J Clin Neuropsychol *4:*91, 1982.

*Stefanatos GA, Kollros, P, Rabinovich H, Stone JJ: Acquired epileptiform aphasia (Landau-Kleffner Syndrome): Current concepts and controversies. Dev Learn Disord *2:*3, 1998.

Taylor HG, Fletcher JM: Neuropsychological assessment of children. In *Handbook of Psychological Assessment*, ed 2, G Goldstein, M Herson, editors. Wiley, New York, 1990.

Taylor HG, Schatschneider C: Child neuropsychological assessment: A test of basic assumptions. Clin Neuropsychol *6:*259, 1992.

Thal DJ, Marchman V, Stiles J, Aram D, Trauner D, Nass R, Bates E: Early lexical development in children with focal brain injury. Brain Lang *40:*491, 1991.

Welsh MC, Pennington BF: Assessing frontal lobe functioning in children: Views from developmental psychology. Dev Neuropsychol *4:*119, 1988.

Welsh MC, Pennington BF, Grossier DB: A normative-developmental study of executive function: A window on prefrontal function in children. Dev Neuropsychol *7:*131, 1991.

Wilson B: Neuropsychological assessment of preschool children. In *Handbook of Clinical Neuropsychology*, vol 2, S Filskov, TJ Boll, editors. Wiley, New York, 1986.

Wilson BC, Risucci DA: A model for clinical-quantitative classification. Generation 1: Application to language-disordered preschool children. Brain Lang *27:*281, 1986.

Yeates KO, Taylor HG: Neuropsychological assessment of older children. In *Handbook of Human Brain Function: Assessment and Rehabilitation*, vol 3, *Neuropsychology*, G Goldstein, PD Nussbaum, SR Beers, editors. Plenum, New York, 1998.

Ylvisaker M, Szekeres SF, Hartwick P, Tworek P: Cognitive intervention. In *Educational Dimensions of Acquired Brain Injury*, RC Savage, GF Wolcott, editors. Pro-Ed, Austin, TX, 1994.

7.7 MEDICAL ASSESSMENT AND LABORATORY TESTING IN PSYCHIATRY

RICHARD B. ROSSE, M.D., LYNN H. DEUTSCH, D.O. AND STEPHEN I. DEUTSCH, M.D., IH.D

Traditionally the medical assessment of psychiatric patients has been designed to detect underlying medical conditions thought to contribute to psychiatric symptomatology. In some studies the rates of physical illness in certain psychiatric populations (e.g., the chronic mentally ill) are reported to be quite high. Because of the high incidence of physical problems and morbidity in some psychiatric populations, it has even been argued that qualified psychiatrists should assume the role of primary care physicians for such patients. Regardless of whether the psychiatrist is functioning as a medical specialist or primary care provider, psychiatrists are now expected to be able to direct the medical evaluations of their patients.

Psychiatric manifestations of a medical illness (e.g., anxiety, depression, hallucinations, delusions, delirium) are not specific to any one medical condition. Hence, the clinician needs to be able to consider a list of different organic possibilities for the patient's psychiatric symptoms. Table 7.7-1 outlines a list of medical conditions that can present with psychiatric symptoms. Each of these possibilities can argue for a different set of laboratory or diagnostic tests. The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) requires as a criterion that the patient's condition not be attributable to a known medical condition or substance. This implies that possible organic causes for the patient's psychiatric complaints have been actively excluded. When an organic factor is identified as being contributory to a patient's psychiatric presentation, the patient's diagnosis indicates the presumed organic etiology. The discovery of an organic cause for a psychiatric presentation can have profound treatment implications for directing the therapy away from mere symptomatic treatment towards a therapeutic intervention appropriate to the underlying medical problem. The psychiatrist needs to remain current with the psychiatric literature that describes the medical mimics of psychiatric disease.

Table 7.7-1 Some Medical Conditions That May Present With Neuropsychiatric Symptoms

Laboratory and diagnostic testing has also become increasingly important for the safe and maximally effective use of certain psychotropic medications. Such testing may involve assessment of certain baseline (pretreatment) laboratory and other diagnostic tests (e.g., electrocardiogram [ECG]) to establish a patient's capacity to tolerate a particular medication. The laboratory is also used during medication use to monitor therapeutic concentrations in blood, to confirm compliance, and to detect unrecognized toxicity. The specific laboratory and diagnostic tests used during the baseline and follow-up evaluations depend on the medication being used and the patient's age and medical condition. Clinicians need to be knowledgeable about the potential toxicities of the medications they use. They need to understand which organ systems are affected adversely by the psychotropic agents they employ and to be knowledgeable about the tests that best probe for potential toxicity. Finally, the laboratory is also essential in the assessment and management of patients with substance use disorders.

MEDICAL HISTORY AND PHYSICAL EXAMINATION IN PSYCHIATRY

A medical history and review of systems are important in the evaluation of psychiatric patients. When it comes to arriving at an appropriate diagnosis for a patient, some studies suggest that the history is the most helpful, the laboratory testing is the least helpful, and the physical examination is of intermediate value. Unfortunately, it is often difficult to get a reliable medical history and review of systems from certain neuropsychiatric patients (e.g., those with dementia, delirium, psychosis, and depression or those who are agitated and uncommunicative). In such patients the clinician must rely on alternative sources of medical history (e.g., medical records, family members). Unfortunately, this is an era in which economic pressures limit the amount of time available or scheduled for examining patients and their charts, and often history taking suffers the most under such pressures.

The traditional parameters of the medical history start with the patient's current physical complaints, the history of their onset, and current medications (including over-the-counter medications, vitamins, and home remedies). The medical history should include an assessment of the patient's past overall health; investigation into past injuries (such as motor vehicle accidents, head trauma, periods of unconsciousness); surgeries; occupational and toxic exposures; illicit drug use; alcohol, tobacco and caffeine use; previous hospitalizations for medical reasons; and any allergies to medications. A traditional review of systems may elicit symptoms suggestive of underlying medical disease. Many psychiatrists think that the therapeutic alliance is strengthened by the process of being involved with the medical evaluation of patients, including performing a physical examination. However, the psychiatric practitioner must use judgment in determining who can best complete the medical evaluation of particular patients, especially those with paranoia and sexual concerns. When a physical examination is performed, a chaperone should be present when appropriate. If the psychiatrist defers performing a physical examination, it should be for as short a period of time as possible (and only after careful consideration of the risks and benefits of deferring such an examination). The reasons for deferring the examination should be carefully documented, and the time frame for completing the physical examination provided.

LABORATORY TESTING

Although some clinicians and researchers support a frugal and selective approach for ordering screening tests for psychiatric patients, others claim that important medical problems are missed by restrictive screening protocols. Consensus is lacking on which tests should be included in a screening battery for psychiatric patients. However, all agree that the relevance of any test that is ordered should be clear to the physician who orders it. Different patient populations may be best served by different screening test batteries (some less selective than others). For instance, broader laboratory screening evaluations are often justified in the following patient populations: (1) patients over the age of 65, (2) patients living in poverty, (3) patients with histories of drug or alcohol abuse, (4) institutionalized patients, and (5) patients with evidence of cognitive impairments or self-neglect. In the future improvements in the information available to the clinician at the clinician-computer interface at the time of order entry should help mental health clinicians make more appropriate decisions regarding which tests to order. Expert computer systems that allow for input of appropriate patient history and physical examination findings should be able to further cue the psychiatrists about which tests should be ordered.

Choosing the Tests The patient's history and physical examination typically dictate which tests are ordered. Laboratory abnormalities are typically useful when they optimize clinical outcomes; that is, if the test results will contribute to the detection of a previously unrecognized medical condition or otherwise influence treatment. Diagnostic testing can also serve a therapeutic function by reassuring the patient or family that other serious medical problems do not appear to be present. Table

7.7-2 lists some diagnostic tests used in the evaluation of a wide range of psychiatric patients; Table 7.7-3 lists numerous psychiatric indications for various tests.

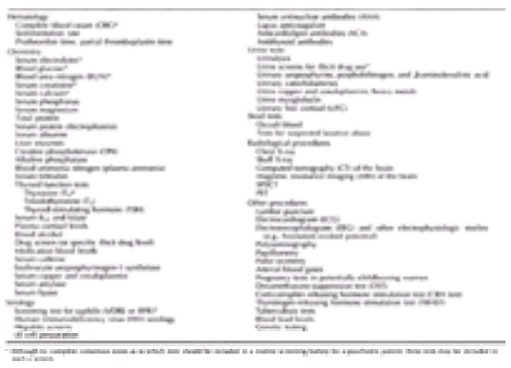

Table 7.7-2 Laboratory Diagnostic Tests of Potential Relevance to the Psychiatric Patient

Table 7.7-3 Psychiatric Indications for Diagnostic Tests

Follow-Up Tests Follow-up laboratory testing is important whenever a significant test abnormality has been found. Additional tests may help to determine the cause or seriousness of an abnormality, or simply represent a repeat of a previously ordered test to determine the reproducibility of the result. Certain experts in the field of diagnostic testing are concerned that extensive testing strategies increase the chances of irrelevant and clinically insignificant "abnormal" results and create a cascade of further unnecessary testing. In addition to some of this further testing being unnecessary and expensive, it has been argued that some follow-up testing can even be dangerous for the patient. Consultation with medical colleagues in the relevant subspecialty can help guide the psychiatrist in the most appropriate use of follow-up tests.

Sensitivity, Specificity, and Predictive Value Three terms that help the clinician to interpret the significance of test results are *sensitivity, specificity*, and *predictive* value. *Sensitivity* is the proportion of subjects with a positive test result who have the disease out of all the subjects affected with the disease. Hence, sensitivity reflects the ability of an abnormal test result to correctly identify persons who actually have the disease. The more false negatives a test has, the lower its sensitivity. *Specificity* refers to the proportion of negative test results among a population of persons who do not have the disease. The more false positives a test has, the lower the test's specificity. Tests with low sensitivity or specificity are inadequate as screening tests. *Predictive value* can be described as *positive or negative.* A positive predictive value is the probability of having the disease given a positive test whereas negative predictive value is the probability of not having the disease given a negative test result. Table 7.7-4 provides the formulas for sensitivity, specificity, and predictive value.

Sensitivity (probability of a positive test result in one who has the disorder)
Sensitivity = true positives/true positives + false negatives × 100%

Specificity (probability of a negative test result in one who does not have the disorder)
Specificity = true negatives/true negatives + false positives × 100%

Predictive value of a positive test result (percentage of all positive test results that are true positives)
Predictive value = true positives/true positives + false positives × 100%

Predictive value of a negative test result (percentage of all negative test results that are true negatives)
Predictive value = true negatives/true negatives + false negatives × 100%

Table by Darrell G. Kirch, M.D.

Table 7.7-4 Sensitivity, Specificity, and Predictive Value

Laboratory and diagnostic tests do not have absolute sensitivities, specificities, and predictive values. Also, different laboratories use different cut-off values to provide the critical values that determine whether a result is abnormal or normal (i.e., positive or negative). In general, "normal" laboratory values for healthy persons are often described as those that fall within two standard deviations of the mean of a normal population. However, few laboratory values have been demonstrated to fall within a true Gaussian distribution. Varying cut-off values will change the sensitivity and specificity of a test. Changes in cut-off values that increase the sensitivity of a test typically decrease its specificity and vice versa. Moreover, the predictive value of a test varies with the actual prevalence of the disorder in the population being studied. Hence, tests with high sensitivity could be of little value as screening tests if the prevalence of the disorder in the population being screened is very low. Patients with many signs, symptoms, and risk factors for a particular disease increase the predictive value of a positive test result suggestive of the disease. Patients with few signs, symptoms, or risk factors for a particular disease are less likely to belong to a population afflicted with the disease; thus, the predictive value of a negative test result is increased in such patients. Therefore, it is important to integrate all aspects of the patient's presentation (e.g., history, review of systems, physical examination, other test results) as well as an understanding of a particular test's sensitivity, specificity, and predictive value when interpreting the significance of a laboratory result.

SPECIAL POPULATIONS

Geriatric Patients Geriatric patients are more likely to have at least one chronic medical disease or disability, therefore the diagnostic assessment of the geriatric patient requires a rigorous search for organic factors that may be related to the presenting psychiatric symptoms. Screening laboratory tests are considered especially important in the psychiatric evaluation of a geriatric patient with a recent onset of psychiatric symptoms, change in behavior, or symptoms resistant to therapy. A more extensive laboratory screening battery is generally justified in older neuropsychiatric patients. The specific set of screening tests is typically determined by the history and clinical presentation, but often includes a complete blood count (CBC) with differential; a full chemistry panel (e.g., "chem 20"); thyroid function tests (e.g., at least a serum thyroid-stimulating hormone [TSH]), serological tests for syphilis, B_{12}, and folate; a midstream urine test; chest X-ray, and ECG.

Urinary Tract Infection Urinary tract infection is a common medical problem in older patients. In addition to symptoms of dysuria, increased frequency, abdominal pain, nausea, and vomiting, these infections can present with mental status changes. Geriatric patients may not become febrile in the presence of serious infections. Urinalysis usually demonstrates bacturia, enteric rods, and pyuria, but these may be absent in the geriatric patient. Urine culture and sensitivity should follow for a

definitive diagnosis. Males with urinary tract infections unrelated to catheterization and females with recurrent infections should have a urological referral.

Anemia Anemia is a laboratory abnormality found in more than 30 percent of geriatric outpatients. Whereas younger patients display the more classic signs of pallor, weakness, and fatigue, the older anemic patient may present with confusion, apathy, or behavioral change. In the elderly an anemia workup should be pursued because symptoms may improve with only a small change in the serum hemoglobin. Laboratory workup for anemia typically begins with a CBC with differential, platelet count, reticulocyte count, and peripheral smear. Based on these studies, the anemia can usually be classified as microcytic, macrocytic, or normocytic. In the elderly the anemia may present as a combination of several pathological processes. Additional laboratory tests in the workup of a microcytic anemia include a serum iron level, total iron-binding capacity (TIBC), serum ferritin, and stool exam for occult blood. The presence of chronic disease in the elderly may lead to a normal TIBC and mask the more typical laboratory findings of low iron and elevated TIBC. Therefore, in the presence of low serum ferritin and negative stool for occult blood, colonoscopy may be indicated to identify the source of blood loss. Macrocytic anemia may be secondary to vitamin B_{12} or folate deficiency. Red cell folate concentrations are a more accurate reflection of folate than serum folate concentrations but may not be available in all laboratories. Pernicious anemia is the most frequent cause of vitamin B_{12} deficiency in the elderly and may present with psychiatric symptoms such as delirium, hallucinations, personality changes, delusions, depression, anxiety, or mania. Neurological exam may reveal paresthesias and abnormal proprioception. The neuropsychiatric symptoms can occur prior to the hematological changes, making serum B_{12} and folate levels important laboratory tests to consider in any elderly patient with psychiatric symptoms. Certain clinical features are more common in patients with a history of heart failure or atherosclerotic heart disease: specifically, fatigue, shortness of breath, worsening angina pectoris, and peripheral edema. Mental status changes of confusion, depression, agitation, and apathy may occur in older anemic patients with no previous psychiatric history. Dizziness is a common complaint in these patients, and pallor may be detected in the mucosa of the oral cavity and conjunctiva.

Thyroid Disease Thyroid disease in the elderly is common and can present with an atypical picture. Nonspecific symptoms such as fatigue and constipation or a change in functional ability are associated with thyroid disease. Hypothyroidism in the elderly may present with symptoms of lassitude, constipation, cold intolerance, fatigue, and cognitive impairment. These symptoms may be attributed to depression or degenerative dementia. Myxedema coma is a medical emergency that occurs primarily in patients over age 50. Older patients with hyperthyroidism may present with typical symptoms (e.g., heat intolerance, weight loss, tremor, palpitations, and atrial fibrillation) or with apathetic hyperthyroidism (manifested by anorexia and fatigue).

Hearing and Visual Loss Over 25 percent of patients 65 years and older manifest hearing impairment, which can lead to social isolation, depression, and confusion. The hearing impairment, usually in the high-frequency range, may interfere with conversation and the psychiatrist needs to consider this in the diagnostic process. The workup for hearing impairment includes an otoscopic examination of the external ear for impaction of cerumen. If the ear is clear, patients can be screened with a handheld audioscope or referred to audiology.

Visual impairment is present in 15 percent of persons 65 and older and twice that in patients over age 85. Cataracts and macular degeneration account for the majority of these cases, but glaucoma and diabetic retinopathy are other important causes of visual deficits. Screening tests of visual acuity include a visual acuity chart. Patients with visual problems should be appropriately referred to rule out the presence of serious eye disease. Persons used to adequate visual functioning in their youth who have to adjust to failing vision in old age can be at risk of considerable depression.

Dementia The neuropsychiatric condition most closely associated with increasing age is dementia. All geriatric patients presenting to a psychiatrist should be screened with a brief cognitive screening examination such as the Mini-Mental State Examination (MMSE). If new cognitive deficits are present that are not accounted for by limited education, a laboratory workup for reversible causes of dementia is indicated. The National Institutes of Health Consensus Development Conference Statement on the Differential Diagnosis of Dementing Diseases recommends careful screening for a patient with new onset of dementia in an attempt to identify treatable causes of dementia. Regarding the choice of diagnostic tests, the conference recommendation is as follows—"the best diagnostic test is a careful history and physical and mental status examination by a physician with a knowledge of and an interest in dementia and dementing disease." The conference report goes on:

> the laboratory tests that are used should be individualized based on the history and physical and mental status examination. Overtesting may expose the patient to discomfort, inconvenience, excess cost, and the likelihood of false-positive tests that may lead to additional unnecessary testing. Undertesting also has hazard; for example, in elderly persons, medical diseases may have nonspecific presentations such as dementia.

Besides a thorough medical history and physical and neurological examinations, laboratory tests typically used in a dementia evaluation include a CBC with differential, sedimentation rate, blood chemistry panel (with liver and kidney function tests), serum electrolytes (some include magnesium and zinc blood levels), thyroid function tests, B_{12} and folate levels, and a serological test for syphilis. It has been argued that human immunodeficiency virus (HIV) testing should be considered in the laboratory evaluation of the patient with dementia because some patients may be reluctant to admit to a history that would have put them at risk of HIV infection, and neurological involvement is a common sequela of HIV infection. Chest X-ray and ECG are also usually included, and a brain imaging study such as computed tomography (CT) or magnetic resonance imaging (MRI) are now considered standard; electroencephalograms (EEG) are also frequently obtained. Some clinicians are even ordering single photon emission computed tomography (SPECT) scans to further their understanding of their patient's dementing condition. On SPECT and positron emission tomography (PET) scans, single focal perfusion defects, or multiple areas of patchy hypoperfusion are suggestive of ischemic brain changes (vascular dementia) whereas dementia of the Alzheimer's type typically shows decreased perfusion in temporoparietal areas bilaterally. Lumbar puncture is obtained only if meningitis is suspected.

Patients Taking Tacrine Tacrine (Cognex) is a reversible cholinesterase inhibitor used in the treatment of dementia of the Alzheimer's type. Because this medication is associated with liver toxicity, baseline (i.e., pretreatment) and follow-up serum transaminase levels (specifically serum glutamic pyruvic transaminase [SGPT], now known as alanine aminotransferase [ALT]), need to be obtained. SGPT levels are monitored every other week for at least the first 16 weeks of treatment. Treatment typically continues according to the recommended titration schedule as long as the SGPT concentration remains less than two times the upper limit of normal (ULN). During tacrine treatment, the drug manufacturer recommends that ALT monitoring be done weekly if blood levels of this serum transaminase rise to above two times ULN (but remain less than three times ULN). If ALT concentrations rise to between three and five times ULN, the medication is reduced by 40 mg a day. Blood level monitoring is relaxed to every other week only after the ALT levels return to normal limits. It is recommended that tacrine treatment be halted if ALT levels rise to greater than five times ULN. In patients whose serum liver enzymes return to normal levels and who can be rechallenged, weekly blood level monitoring is recommended for at least the first 16 weeks after the initiation of the rechallenge. If serum ALT levels remain in the normal range, monitoring can be reduced to monthly for 2 months and every 3 months thereafter. Patients who become jaundiced with total bilirubin levels greater than 3 mg/dL in association with ALT elevations are immediately and permanently discontinued from tacrine.

Self-Abusers Various diagnostic assessments can be useful in the evaluation of the patient who engages in deliberately self-injurious behavior. Clues to such behavior include wrist or arm lacerations or scars; the presence of open wounds, bruises, burns, or abrasions; or evidence of swelling or deformity secondary to injury. Self-damaging behavior might even cause skull deformities. Blood toxicology is important for the evaluation of the patient who has overdosed on prescribed or abused drugs. The degree of organ involvement in the overdose can be assessed by the appropriate diagnostic tests. For instance, in the case of an overdose of a tricyclic antidepressant an ECG is used to assess the degree of cardiac toxicity. In the context of acetaminophen (Tylenol) overdoses, liver function tests are important, and in the context of aspirin overdose, serum electrolyte determinations are needed to assess anion gap. Deformity in a region of the body would require an X-ray evaluation of the appropriate area. Usually, consultation with medical or surgical colleagues is necessary to resolve issues related to self-inflicted injury in the psychiatric patient.

Restrained Patients Certain patients are restrained to prevent them from inflicting further harm on themselves or others. Because such individuals cannot attend to their own needs, they must be carefully monitored. The assessment of these patients includes the frequent monitoring of vital signs and vigilance for signs of dehydration (e.g., dry mouth, loss of skin turgor) and improper restraint technique (e.g., edema, swelling, or abrasions). The extremities need to be checked regularly to ensure adequate circulation. Very agitated patients might need to have their urine checked for evidence of myoglobinuria secondary to muscle breakdown. For patients in whom the etiology of the agitated behavior is unclear, the laboratory is vital in assessing possible "organic" causes (e.g., serum and urine toxicology screens, serum blood screens for evidence of infection or electrolyte disturbance, CT scans if there is the suggestion of possible stroke). Lumbar puncture with analysis of cerebrospinal fluid (CSF) would be necessary in a patient with signs of meningeal irritation or other clinical findings suggestive of infection.

Psychiatric Patients With Weight Loss Certain psychiatric conditions are associated with patients experiencing weight loss. These conditions include anorexia nervosa, melancholic depression, severe psychosis, catatonia, and severe alcohol or substance dependence. In such patients a laboratory assessment is needed to determine the degree of malnutrition and starvation and to identify a potential medical emergency (e.g., due to low serum potassium). It is important to monitor vital signs and orthostatic blood pressure changes. Tests that are ordered frequently in these patients include a CBC with differential (e.g., to determine the degree of anemia, although hemoconcentration might give an artificially elevated result prior to rehydration), serum electrolytes (especially potassium and blood urea nitrogen), and a urinalysis (e.g., the specific gravity would be high in a dehydrated patient). Other tests that might be ordered include serum iron studies, serum total protein,

and albumin.

In patients with anorexia nervosa, in addition to the tests just described, an erythrocyte sedimentation rate (ESR) is often recommended. The ESR is usually low in patients with anorexia nervosa, hence this test can be somewhat useful in differentiating anorexia nervosa from other medical conditions causing anorexia. An ECG is important to assess for the impact of low potassium and starvation on the heart. Some researchers have recommended CT on MRI to exclude the presence of brain tumors. Endocrine status is frequently evaluated, including tests of thyroid function and other pituitary hormones. Stool and urine tests for laxative abuse might also be useful for patients suspected of such abuse. Certain patients might abuse appetite suppressants (note the presence of a relationship between certain appetite suppressants and pulmonary hypertension—about 23 to 46 cases per million persons per year). Bone density studies can be useful to further assess the physical impact of the patient's eating disorder.

First Episode of Psychosis Experts recommend a careful medical evaluation for a patient with a first episode of psychosis (or subsequent episode in a patient who has never had a medical workup). The evaluation should include a complete physical examination, and the core laboratory screening battery should include a urine toxicology screen for drugs of abuse, a general chemistry screen, a CBC with differential, and urinalysis. The clinician should carefully consider a range of supplemental tests as indicated, which includes a pregnancy test, ECG, brain MRI or CT, and EEG.

New Patients With Manic or Major Depressive Episodes Experts also recommend a complete physical examination, as well as a core laboratory screening battery that includes thyroid function tests (if not recently obtained), CBC, general chemistry screen, urinalysis (especially if the patient is on or might be started on lithium [Eskalith]), urine toxicology test for drugs of abuse, and ECG in patients over the age of 40 (or younger in patients with a history of cardiac disease). Serum concentrations of mood stabilizers (e.g., lithium, valproic acid [Depakene] or carbamazepine [Tegretol] are also obtained if the patient is already on any of these medications. A pregnancy test should also be ordered if appropriate. Brain imaging (e.g., MRI, CT) and EEG studies are often also employed.

Substance Abusers The laboratory has become increasingly important in the physician's assessment and monitoring of the substance-abusing patient. The laboratory is useful for detecting substances of abuse and for evaluating the impact the substance use is having on the patient's body. Often the laboratory detection of abused substances and certain diagnostic test abnormalities related to the substance abuse (e.g., abnormal liver function tests in alcohol-abusing patients) is used therapeutically to confront the denial of a patient with a substance abuse disorder and to help engage the patient in treatment.

The most commonly used specimen for the detection of drugs of abuse is urine, although toxicological analyses can also be performed on blood specimens. The period of time that the clinician can detect drugs in blood specimens is typically shorter than the length of time drugs can be detected in urine specimens because drugs and their metabolites are excreted and detectable in the urine for longer periods of time than they are detectable in blood. However, the length of time that a particular drug of abuse can be detected in the urine is somewhat variable, depending on the specific drug, the duration and amounts of the substance used, and concomitant medical problems (e.g., liver or kidney disease). Nevertheless, Table 7.7-5 provides a listing of some common drugs of abuse that can be detected in urine specimens, along with a typical length of time after recent use that the substance can be detected. Other specimens that have been studied to detect substance abuse include saliva and hair samples. A commercially available test for the measurement of alcohol in saliva exists.

Drug	Length of Time Detected in Urine
Alcohol	7–12 hours
Amphetamine	48 hours
Barbiturate	24 hours (short-acting) 3 weeks (long-acting)
Benzodiazepine	3 days
Cocaine	6–8 hours (metabolites 2–4 days)
Codeine	48 hours
Heroin	36–72 hours
Marijuana	3 days to 4 weeks (depending on use)
Methadone	3 days
Methaqualone	7 days
Morphine	48–72 hours
Phencyclidine (PCP)	8 days
Propoxyphene	6–48 hours

Table 7.7-5 Drugs of Abuse That Can Be Tested in Urine

Numerous methodologies exist for the detection of substances of abuse in the potentially substance-abusing patient. The most common strategy involves the use of a screening test and, in the context of a positive result, a more sensitive, confirmatory test (e.g., a screening enzyme-multiplied immunoassay technique [EMIT] followed up by confirmatory gas chromatography-mass spectrometry [GC-MS] testing).

Alcohol-Related Disorders In patients with alcoholism, the extent of liver enzyme elevations and abnormalities for other laboratory tests of liver function reflect the degree of liver damage in these patients. The liver enzymes examined include gamma-glutamyl transaminase (GGT), AST, and ALT. Also reflecting the degree of liver damage in these patients are abnormalities of other liver function tests (LFTs), such as reduced serum total protein and albumin, and prolonged prothrombin time (PT). Laboratory abnormalities in these patients can be used to confront their denial of their alcoholism. The clinician should note that patients with severe alcohol-related hepatic damage might no longer be able to generate liver enzyme elevations despite advanced liver disease. The alcohol-abusing patient should also be assessed for possible head trauma (e.g., subdural hematomas); hence, a CT scan might be needed (and a skull X-ray if there is evidence of skull injury).

Patients with alcohol use disorder should have a CBC to evaluate potential anemia (classically megaloblastic in nature but can be mixed megaloblastic-microcytic). Serum concentrations of magnesium, electrolytes, blood urea nitrogen (BUN), and creatinine are also important in the laboratory assessment of this patient population.

Patients with signs and symptoms of alcohol withdrawal (e.g., elevated blood pressure, pulse, tremor [6 to 8 contractions per second]) will typically have rapidly decreasing or undetectable blood alcohol concentrations. Alcohol withdrawal delirium is a medical emergency requiring careful attention to the medical needs of these patients. Hence, a negative test for alcohol in a patient with a history of alcohol abuse is not inconsistent with an ensuing medical emergency related to alcohol withdrawal.

Intravenous Substance Use A number of substances of abuse are delivered by the addict through the intravenous (IV) route (e.g., heroin, cocaine, amphetamines). The needles have often been used by other addicts and inadequately sterilized prior to use. These needles can be potentially infected with viral agents capable of causing hepatitis and acquired immune deficiency syndrome (AIDS).

Hepatitis Testing All forms of viral hepatitis (including hepatitis A, B, C, and D) have been associated with the drug-addict lifestyle. Only hepatitis B, C, and D are transmitted by the intravenous route. Hepatitis A is transmitted by the fecal-oral route, and has been associated with the poor sanitary conditions to which addicts are exposed. Serum hepatitis panels for antigens and antibodies associated with hepatitis-causing viruses are indicated in patients at high risk for these hepatitides or with evidence of hepatic damage (e.g., jaundice, elevated liver enzymes, elevated serum bilirubin, prolonged serum coagulation tests, urobilinogen on urinalysis).

HIV Testing The spread of HIV amongst the intravenous-drug–abusing population continues to be a major health care disaster. Such patients are driven by their addiction to engage in high-risk behaviors that expose them to the virus. When HIV testing is ordered, false-positive tests in a high-risk population are extremely rare. Patients undergoing HIV testing need to have appropriate pretest and posttest counseling. HIV brain infection can result in a myriad of neuropsychiatric manifestations, including dementia and psychosis.

Medical Conditions in Psychiatry Multiple medical conditions can present with psychiatric symptoms (Table 7.7-1).

Thyroid Disorders Both hyperthyroidism and hypothyroidism can mimic a wide variety of psychiatric disorders, including anxiety, depression, and psychosis. Psychiatrists are understandably interested in excluding thyroid disorder as a cause of their patients' psychiatric symptoms. Although the onset of symptoms in hyperthyroidism is often rapid, the onset of symptoms in hypothyroidism can be insidious. When “routine” screening tests are advocated for psychiatric patients, some degree of thyroid testing is usually included in the laboratory screening battery. With highly sensitive TSH assays now available, many cost-conscious clinicians and

researchers are advocating the use of a serum TSH alone for routine screening purposes in psychiatric populations. Other thyroid tests include serum thyroxine (T_4), serum triiodothyronine (T_3), T_3 resin uptake (T_3RU), free T_4, and the free thyroxine index (FTI), sometimes referred to as T_7 or T_{12}. Other thyroid tests include antithyroglobulin antibodies and antithyroid microsomal antibodies and the thyrotropin-releasing hormone stimulation test (TRHST). Patients who demonstrate goiter, ophthalmopathy, evidence of a hypermetabolic state, or have a history or symptoms suggestive of thyroid disease (e.g., heat or cold intolerance) or risk factors for thyroid problems (e.g., history of neck or chest irradiation, thyroid surgery or iodine-131 treatment, or treatment with lithium) should have a laboratory thyroid evaluation. Hyperthyroidism that results from the accidental or deliberate intake of excessive amounts of thyroid supplements is called *thyrotoxicosis factitia*. Finally, although generalized resistance to thyroid hormone has been associated with an increased incidence of attention-deficit/hyperactivity disorder, because of the rarity of the thyroid hormone disorder in children with attention-deficit/hyperactivity disorder, it is not recommended that thyroid function be routinely measured in cases of nonfamilial attention-deficit/hyperactivity disorder.

The thyroid gland can be inspected by observing the lower half of the neck in the anterior triangles and noting the presence of an ascending mass as the patient swallows. Next, the gland can be palpated from behind or in front. alpation of the gland from behind involves standing behind the patient and exploring the region behind the sternocleidomastoid muscles. The thyroid isthmus can be palpated on the anterior aspect of the tracheal rings. Next, the lower and upper poles of the lateral lobes are palpated. The inclination of the patient's head should be shifted to relax the neck muscles in addition to instructing the patient to swallow. If the thyroid gland undergoes hyperplasia and a goiter is palpated, a thyroid bruit may be auscultated from accelerated blood flow. A small diffuse goiter may be present before the menstrual period, during pregnancy, or throughout the menstrual cycle in females 12 to 20 years old. It can also be present with thyrotoxicosis. In addition to palpation of the thyroid to detect enlargement, palpation also involves a search for nodules and tenderness.

Adrenal Diseases Adrenal disease can be associated with a wide variety of psychiatric symptoms, including depression, anxiety, and psychosis. Low plasma cortisol is seen in Addison's disease (adrenal insufficiency) and elevated blood cortisol levels are seen in Cushing's disease. Evidence of hypercortisolism can be detected in urine; in some patients with major depressive disorder urinary cortisol concentrations approach or are in the range associated with Cushing's disease. An elevated urinary 17-hydroxycorticosteriod concentration has been proposed as a marker of potential suicidality by some psychiatric researchers. Patients with major depressive disorder can have alterations in the normal diurnal variation of plasma cortisol concentrations (measurable with the diurnal cortisol test [DCT], as well as a decreased suppression of serum cortisol by the exogenous administration of dexamethasone (measurable with the dexamethasone suppression test [DST]). Recent studies have reported evidence of increased adrenal gland volume in patients with major depressive disorder compared to controls. Patients with a history or physical examination suggestive of pheochromocytoma (e.g., paroxysmal or persistent hypertension, flushing, anxiety, panic attacks) might need a battery of tests including the following: a 24-hour urine for measurement of catecholamines and metabolites (such as metanephrine or vanillymandelic acid [VMA]), plasma levels of catecholamines, and a CT scan of the abdomen. Nuclear medicine scans using iodine-131-labeled meta-iodobenzy guandine are employed to locate the catecholamine-secreting tumors in this condition.

Pernicious Anemia The neuropsychiatric symptoms associated with vitamin B_{12} deficiency (e.g., dementia, depressive disorder, psychotic disorder) might not always be associated with megaloblastic anemia or other classic laboratory findings typically connected with pernicious anemia (e.g., abnormal Schilling tests). It has been suggested that elevated serum homocysteine and methylmalonic acid concentrations in patients with vitamin B_{12} deficiency might be predictive of those patients whose neuropsychiatric manifestations might respond to parenteral vitamin B_{12} therapy. Serum folate concentrations are usually measured with B_{12} concentrations because folate deficiency can result in a similar constellation of neuropsychiatric impairments. In patients with pernicious anemia, folate therapy without vitamin B_{12} administration can result in a normalization of the hematological abnormalities but continued progression of the neuropsychiatric manifestations of the disease.

Acute Intermittent Porphyria Acute intermittent porphyria can be associated with psychiatric symptoms (psychosis) even when the physical manifestations (i.e., abdominal pain, neuropathy, autonomic dysfunction) of the disorder are not present. During acute attacks there are typically elevations of uroporphyrin, urine porphobilinogen (PBG), and d-aminolevulinic acid (ALA) concentrations in the urine. Elevations of these substances are much less common during asymptomatic periods. Diminished erythrocyte uroporphyrinogen-1-synthetase can be seen during both symptomatic and asymptomatic periods of the illness.

Wilson's Disease Wilson's disease is an autosomal recessive disorder of copper metabolism that typically becomes symptomatic (with movement disorder, psychosis, personality changes) in the second to third decades of life. There is an elevation of urine copper after a 24-hour collection, decreased serum copper, and low serum ceruloplasmin. In patients with Wilson's disease it is important to monitor laboratory abnormalities related to copper deposition in the liver or kidney. Kayser-Fleischer rings adjacent to the cornea can be seen during a slit-lamp examination.

Syphilis With central nervous system (CNS) involvement, this infectious disease has classically been related to a wide variety of neuropsychiatric complications, including psychosis and dementia. The rapid plasma reagin test (RPR) or Venereal Disease Research Laboratories (VDRL) test is the most common screening test for syphilis. Positive results on these tests are followed up with fluorescent treponemal antibody absorption (FTA-ABS) tests. The test that is diagnostic of tertiary or CNS syphilis is the CSF VDRL. A positive CSF VDRL is typically associated with other CSF findings such as abnormal CSF protein and white blood cell counts.

Systemic Lupus Erythematosus This rare clinical condition can be associated with psychiatric manifestations including psychosis, delirium, depression, and dementia. It is important to detect it as the cause of psychiatric symptoms to ensure proper treatment (e.g., treatment of the lupus-associated cerebritis with steroids). Laboratory tests for lupus include the lupus prep (phenothiazines can cause false-positive results of this test) and tests for serum antinuclear and antideoxyribonucleic acid (DNA) antibodies. The antiphospholipid antibody known as lupus anticoagulant has been described in patients with lupus and rarely in patients treated with antipsychotic medications, especially chlorpromazine (Thorazine). Although the presence of lupus anticoagulant is associated with a lengthening of serum coagulation tests such as the prothrombin time, it is clinically associated with an increased incidence of thrombotic events such as stroke.

Pediatric Autoimmune Neuropsychiatric Disorder Associated With Streptococcal Infections This condition is associated with an exacerbation of neuropsychiatric symptoms (e.g., obsessions, compulsions, tics) in children and adolescents. It has been hypothesized to be associated with the development of antibodies against streptococcal infection that cross-react with basal ganglia structures; acute basal ganglia enlargement has been demonstrated in conjunction with this syndrome. The pathological process of this disorder is thought to be analogous to Sydenham's chorea. The syndrome has also been referred to as *pediatric, infection-triggered, autoimmune neuropsychiatric disorders*.

Environmental Contamination Poisoning with certain contaminants common in many workplace environments, such as heavy metals, is described as having potential neuropsychiatric manifestations. For instance, lead poisoning can result in varying degrees of lead encephalopathy, a problem that affects children and adults. Poisoning with other heavy metals, such as mercury and manganese, can also result in neuropsychiatric symptoms. Exposure to various solvents can also result in various neuropsychiatric symptoms (e.g., *solvent encephalopathy* or *painter's syndrome*). Appropriate laboratory testing for these substances is typically based on a careful occupational history.

MONITORING PSYCHOTROPIC MEDICATIONS

Antidepressant Medications The role of the laboratory for therapeutic drug monitoring (TDM) was more prominent in the previous era when tricyclic drugs were the main class of medications used to treat depression. TDM has not been established for the newer antidepressants such as the selective serotonin reuptake inhibitors (SSRIs). With these new nontricyclic antidepressants now in widespread use, the role of the laboratory in managing patients on antidepressant medications has diminished. The tricyclic drugs were also associated with a host of other potential toxicities (e.g., cardiac), which required other types of diagnostic assessments (e.g., ECG); the enhanced safety profile of the newer agents relative to the tricyclic drugs has precluded much of this other diagnostic testing as well.

Baseline Studies There are no standardized recommendations for the pretreatment evaluation of a patient about to begin antidepressant medication. It is important for the clinician to have a clear understanding of the potential adverse reactions to the antidepressant agent that is going to be used, as well as an understanding of the medical conditions and physiological vulnerabilities of the patient who will be taking the medication. For instance, a liver enzymes evaluation might be ordered in a patient with a past history of liver disease because antidepressant agents are metabolized by the liver and liver disease, as evidenced by elevated liver enzymes, might suggest extra caution when using these medications; it may also suggest the need for periodic follow-up tests of liver function to assess the effect of the antidepressant drug on liver functioning. Women who are potentially childbearing might have a pregnancy test performed so that the use of the antidepressant agent in such a person will be fully informed. An obvious aim in treating such patients is to try to expose the fetus to as few medications as possible, even to medications that are not clearly teratogenic. Furthermore, patients with a history of cardiac disease about to start treatment with a tricyclic drug should have a baseline ECG; this is especially true for patients over the age of 50. Tricyclic medications can affect intracardiac conduction and lengthen PR, QT, and QRS intervals, and patients with baseline abnormalities of these measures might need periodic follow-up ECGs. Serious and fatal arrhythmias could theoretically result from lack of attention to ECG changes that are related to the tricyclic agents used. Baseline or follow-up ECG abnormalities might suggest that the clinician choose an antidepressant medication less associated with alterations in ECG parameters. Changes in cardiac conduction induced by a tricyclic drug might be especially problematic in the following

patients: the elderly; those with pre-existing cardiac conditions; those on concomitant Type I antiarrhythmia agents such as quinidine (Cardioquin) and lidocaine (Xylocaine); those on digoxin (Lanoxin) or other cardiac glycosides that affect atrioventricular node conduction; and those with low serum potassium or magnesium.

Therapeutic Drug Monitoring Relationships between blood concentrations and therapeutic response have been established for some of the tricyclic medications. It has been recommended that blood-concentration determinations of some of these agents be employed in certain situations, such as: (1) patients with questionable medication compliance, (2) patients who are having a poor response while on a typical dose for a reasonable period of time, (3) patients who have adverse effects at very low dosages, (4) patients with medical problems (or the elderly) in whom it is important to use as low a potentially therapeutic level of the drug as possible, and (5) patients who need to obtain a potentially therapeutic blood level as rapidly as possible because of their clinical condition (e.g., if the patient is severely suicidal). Wide variations between different patients' blood levels on standard dosages of tricyclic medications have been described, presumably reflecting differences in drug absorption and metabolism. Therapeutic levels have been described for: (1) imipramine (Tofranil)—plasma imipramine plus desmethylimipramine concentrations should exceed 200 ng/mL, (2) nortriptyline (Pamelor)—a therapeutic window between 50 to 150 ng/mL), and (3) desipramine (Norpramin)—therapeutic concentrations are described as exceeding 125 ng/mL). In the literature, slightly different therapeutic concentrations are sometimes described. Clinicians should check with their laboratory for slight variations in these therapeutic concentrations that may have been established for their laboratory. Therapeutic blood concentrations for other tricyclic medications are sometimes provided (e.g., amitriptyline [Elavil], 75 to 175 ng/mL). The *therapeutic range* describes a range of blood concentrations at which it is more likely for the patient to experience a therapeutic response (although this does not preclude individual patients from experiencing antidepressant effects outside described ranges).

Clinicians should also be aware of what the toxic blood concentrations are for the tricyclic agents because higher blood concentrations of these medications are associated with greater degrees of toxicity. For the nontricyclic drug bupropion (Wellbutrin), a therapeutic blood concentration between 20 and 75 ng/mL has been described. Therapeutic blood concentrations for other antidepressant drugs (tricyclic and nontricyclic) have been less clearly established but are sometimes obtained to assess compliance, poor response to a conventional antidepressant dose, adverse effects experienced at a very low dosage of the antidepressant, and possible toxic ranges (e.g., in overdose). The utility of blood concentrations of the SSRI agent fluoxetine (Prozac) for the assessment of compliance is limited because of the long half-lives of the parent compound and its metabolite—only more dramatic cases of noncompliance would be detectable. With the monoamine oxidase inhibitors (MAOIs), when laboratory tests are sought for assessing optimal dosing reports suggest that one should try to achieve a level of inhibition of at least 80 percent of platelet monoamine oxidase activity.

In the medical assessment of the patient on tricyclic medications or MAOIs, the clinician should also be alert to clinically significant orthostatic hypotension. If clinical symptoms develop while the patient is standing (e.g., accompanied by 30 mm Hg or greater decrease in systolic and 15 mm Hg or more decrease in diastolic pressure and pulse changes greater than 20 beats per minute), a switch to a different antidepressant drug associated with less orthostatic hypotension might be indicated. Tachycardia is also associated with the more anticholinergic tricyclic drugs. Additionally, weight gain can be associated with many of the antidepressant agents, including tricyclics and MAOIs, and should be monitored in patients who appear vulnerable to such weight gain. Men with prostatic hypertrophy and patients with fecal impaction or constipation are at some risk from the anticholinergic adverse effects of the tricyclic medications. Priapism has been rarely associated with the antidepressant agent trazodone (Desyrel) and the appearance of such priapism typically represents a medical emergency. The clinician should remember that doses of buproprion above 450 mg daily are associated with an increased risk of seizure.

Monitoring of Children Various psychiatric drugs are being used in children, and monitoring of potential toxicities from these agents is important. For instance, about 30 to 60 percent of outpatients in child psychiatry fulfill criteria for major depressive disorder, and when tricyclic drugs are used, monitoring of drug concentrations in plasma plays a role in the management of these patients. Weight-corrected oral dosing of tricyclic agents in children has not been shown to reliably predict plasma levels in individual patients. Limitations to the upward adjustment of dosage of tricyclic medications include severity of “nuisance” adverse effects and significant ECG and cardiovascular changes (i.e., heart rate >130/min; PR interval >0.18 sec; QRS width >30 percent above baseline; blood pressure >140/90 mm Hg). Indeed, ECG monitoring has been shown to be an important laboratory tool for assessing compliance and assessing cardiac toxicity. A significant relationship has been shown between maintenance plasma levels of some tricyclic drugs and clinical response. For instance, for imipramine, the authors of one study recommended a therapeutic blood concentration discriminating cut-off value of 150 ng/mL. In another study with nortriptyline, plasma levels between 60 and 100 ng/mL were associated with a positive clinical response in prepubertal children with depression. However, more research is necessary to determine meaningful therapeutic blood concentrations for antidepressant medications in children. The clinical laboratory will be essential to the development of effective pharmacotherapy for prepubertal depression and management of these children. At a minimum, blood concentration tests will help to determine compliance and dosage regulation.

LITHIUM Although lithium is clearly effective in the treatment of bipolar I disorder, the therapeutic and toxic blood levels of this agent are very close to one another, and in certain individuals even seem to overlap. Additionally, lithium has effects on a number of organ systems of which the clinician should be aware. Lithium therapy is associated with a benign elevation of the white blood cell count (WBC), which can range to about 15,000 cells per mm_3. This WBC elevation can sometimes be mistaken for signs of infection, or wrongly be attributed to lithium in the context of other signs of infection (e.g., fever, cough, discomfort on urination, malaise). Furthermore, lithium can have adverse effects on electrolyte balance (especially in patients on thiazide diuretics), thyroid function, the kidney, and the heart. Hence, common lithium pretreatment tests include serum electrolytes, BUN serum creatinine, urinalysis, thyroid function tests (TFTs) (e.g., TSH, T_4, T_3RU), and an ECG. In patients with a history suggestive of possible kidney problems, a 24-hour urine test for creatinine and protein clearance is recommended, although some clinicians routinely order this test in patients about to begin lithium therapy. Some have argued that antithyroid antibody testing is helpful in assessing the potential of lithium-induced hypothyroidism. Because of the potential cardiac teratogenicity of lithium, a pregnancy test in potentially child-bearing women should be ordered. Periodic follow-up of serum electrolytes, BUN, creatinine, TFTs, ECG, and 24-hour urine for creatinine and protein clearance are recommended. The frequency and exact makeup of the follow-up testing battery should be dictated by the patient's medical condition. Suggested therapeutic blood concentrations for the lithium ion in acute mania range from 0.8 to 1.5 mEq/L, although a range where potential toxicity might be manifest is 1.2 to 1.5 mEq/L (a “warning range” where the risk of developing toxicity rises rapidly). The serum lithium concentrations of patients with lithium levels above 1.0 mEq/L should be carefully monitored. Maintenance therapeutic blood concentrations of lithium are around 0.6 to 0.9 mEq/L. Lithium concentrations for therapeutic blood-level monitoring are drawn as close to 12 hours after the last dose of lithium as possible. Steady-state concentrations in patients without renal dysfunction are reached about 5 to 8 days after initiation of lithium treatment or after a dosage change. Patients with evidence suggestive of lithium toxicity should have a concentration measured immediately. These patients might need close medical attention, including hemodialysis when the lithium levels exceed 2 to 4 mEq/L (especially in the patient in poor medical condition with poor lithium excretion).

Lithium levels are often measured every 1 to 2 weeks during the first 2 months of treatment. Unless there are indications to do otherwise, follow-up testing for patients on lithium includes lithium concentrations obtained every 3 to 6 months, yearly thyroid function testing (e.g., T_4 and TSH), serum BUN, and creatinine. Serum electrolytes are measured every 6 to 12 months, with a 24-hour urine for volume and glomerular filtration rate (GFR) in patients with suspected renal disease. Changes in lithium dosing would necessitate more frequent blood level monitoring.

Antipsychotics Except for the antipsychotic clozapine (Clozaril), no clear pretreatment and follow-up laboratory and diagnostic evaluation strategy exists. Additionally, no specific therapeutic blood levels for these agents have emerged although some suggested therapeutic blood concentrations for several antipsychotic agents exist in the psychiatric literature. For instance, some studies have suggested that a therapeutic window for treatment efficacy exists for haloperidol (Haldol), and clinically meaningful responses to clozapine have been associated with blood concentrations between 200 and 450 ng/dL; one study noted the greatest benefit with blood concentrations greater than 350 ng/dL. For haloperidol, gross toxic adverse effects, including neuroleptic malignant syndrome, confusion, seizures, or catatonia, have been associated with blood concentrations greater than 30 ng/mL; blood-level reductions were associated with a reduction in these toxic adverse effects. There is also significant dose-to-blood-level variability between individuals. Blood concentrations for antipsychotic agents have been described as useful for assessing medication noncompliance or perhaps in treatment-refractory patients who might be very rapid metabolizers of the antipsychotic agent being used. Blood concentrations testing can also be utilized to determine if a drug interaction is lowering the blood level of an antipsychotic drug. Additionally, some clinicians have advocated the use of at least a single blood level in patients who do not respond to an antipsychotic agent.

Clinicians need to be aware of the potential toxicities of the antipsychotic agents that they use, and order laboratory and diagnostic tests accordingly. For instance, clinicians using clozapine need to be aware of its potential to cause fatal agranulocytosis and seizures (although these risks exist with all antipsychotic agents, albeit to a lesser extent). With clozapine the risk of agranulocytosis (defined as an absolute neutrophil count [ANC] of less than $500/mm_3$) appears to be about 1.3 percent per year; the risk of seizures reaches approximately 5 percent of persons treated with higher dosages. Other potential adverse effects of antipsychotic medication include melanism from phenothiazines (especially chlorpromazine [Thorazine] related), abnormal lactation, and gynecomastia. It is also probably advisable to establish the childbearing status of at-risk women prior to the use of antipsychotic medications.

Serious cardiac arrhythmias have been reported associated with antipsychotic use (e.g., thioridazine [Mellaril]); hence, patients with preexisting cardiac disorder might need careful ECG follow-up when these agents are used. In patients about to start on antipsychotic agents with known cardiac effects, obtaining a baseline ECG is prudent. Note that in patients over the age of 50, the frequency of ECG abnormalities is considered significant enough to warrant such screening. ECG screening is

indicated in persons with a history of cardiac disease, regardless of age. Clinicians should also be aware that significant orthostasis can be associated with some antipsychotic agents. Patients with concomitant hepatic or renal disease require periodic monitoring of these organ systems (e.g., with periodic physicals, LFTs, BUN, and creatinine tests).

Neuroleptic Malignant Syndrome The diagnosis and follow-up of this potentially fatal adverse reaction to antipsychotic medications can be assisted by the laboratory. This disorder typically consists of varying degrees of hyperpyrexia, autonomic instability (e.g., pulse greater than 100), severe extrapyramidal dysfunction (95 percent with “lead-pipe” rigidity) and delirium. Laboratory abnormalities often include creatine phosphokinase, WBC elevations, myoglobinuria, and liver enzyme elevations. Patients with neuroleptic malignant syndrome are at risk for serious medical complications including renal failure, pneumonia, respiratory arrest, and cardiovascular collapse. Laboratory and diagnostic testing for these conditions should be ordered as indicated.

Clozapine-Induced Agranulocytosis The agranulocytosis that develops in clozapine-treated patients often occurs within the first 18 weeks of treatment. Although the drop in the WBC is usually gradual, the progression to agranulocytosis can be rapid and can occur within a few days. Fever, flu-like symptoms, sore throat, or petechiae and ecchymosis (i.e., from thrombocytopenia) on any part of the body may be a sign of granulocytopenia in patients taking clozapine. At this time there is no way to predict who will develop this potentially fatal adverse event.

Prior to initiating clozapine therapy, a WBC count should be obtained. It has been recommended that patients with a WBC count of less than 3500/mm $_3$ not be treated with clozapine. If during treatment with clozapine the WBC falls below 3500/mm $_3$, or there is a significant decrease from a higher WBC baseline that does not fall below 3500 (i.e., a single drop of 3000 or more or a cumulative drop of more than 3000 over a 3-week period), or immature white cell forms are detected, a repeat WBC with differential should be obtained immediately. If the count is between 3000-3500/mm $_3$, the ANC should be above 1500/mm$_3$ for therapy to continue (and follow-up WBC counts should be done twice weekly). WBCs below 3000/mm $_3$ or ANCs below 1500/mm$_3$ are typically indications to discontinue clozapine therapy, and WBC counts with differentials should be obtained daily. Clozapine therapy can be resumed when WBC counts are over 3000/mm $_3$ and ANC concentrations are over 1500/mm$_3$. Twice-weekly WBC counts with differentials should be obtained until the WBC reaches 3500/mm $_3$. At WBCs below 2000/mm$_3$ and ANCs below 1000/mm$_3$, bone marrow studies are recommended. It is now also recommended that patients who have total WBC counts that fall below 2000 mm $_3$ during clozapine therapy or ANCs that drop below 1000/mm$_3$ should not be rechallenged with clozapine (Table 7.7-6).

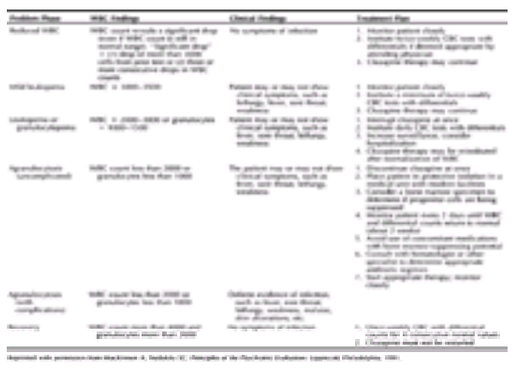

Table 7.7-6 Clinical Management of Reduced White Blood Cell (WBC) Count, Leukopenia, and Agranulocytosis in Patients Taking Clozapine

Clozapine-Related Seizures The risk of seizure from clozapine rises with increasing dosage, and reaches approximately 5 percent of persons treated with high doses of clozapine (between 600 to 900 mg a day). Although most of the clozapine-related seizure events are of the generalized tonic-clonic type, myoclonic seizures and cataplectic-like events have also been described. Before raising clozapine doses above 600 mg a day, the clinician should consider obtaining an EEG; however, according to some reports an abnormal EEG might predict therapeutic response to clozapine. In treating suspected or confirmed clozapine-related seizures, phenytoin (Dilantin) is avoided because of its ability to lower serum clozapine levels and its association with agranulocytosis. Carbamazepine can also cause agranulocytosis and should not be used in patients taking clozapine.

Anticonvulsants

Valproate Valproic acid and divalproex (Depakote) are now commonly used in the management of bipolar I disorder. They have also been used to treat agitation in patients with dementia, depression, and borderline personality disorder. In adults, serum concentrations above 45–50 µg/mL are described as the most effective in managing bipolar I disorder. Serum concentrations above 125 µg/mL are associated with a greater frequency of adverse effects, including thrombocytopenia. Acutely manic patients appear to have the greatest therapeutic response and the fewest adverse effects with valproic acid blood concentrations between about 45 and 100 µg/mL (although therapeutic concentrations up to 125 µg/mL are described). Caution should be exercised when using valproic acid and divalproex sodium in patients with liver disease. These medications are typically discontinued when liver enzyme elevations exceed three times the upper limit of normal. Serious adverse reactions that have been reported during valproic acid therapy include hepatitis, liver failure, hematological abnormalities (e.g., decreased platelets, bone marrow suppression) and acute hemorrhagic pancreatitis. Finally, these agents should not be used during the first trimester of pregnancy because of their association with birth defects.

During the first 2 months, serum valproic acid levels can be obtained as often as every 2 weeks; CBC and LFTs should be obtained monthly. During long-term use, serum valproic acid levels can be obtained every 3 to 6 months, and CBC and LFTs every 6 to 12 months. Any changes in valproic acid or divalproex dosing would necessitate more frequent monitoring of blood levels.

Carbamazepine Carbamazepine has been used as the sole or adjunctive medication in various psychiatric disorders, especially bipolar I disorder. When utilized for the treatment of psychiatric disorders, therapeutic blood levels for carbamazepine have not been clearly established. However, many clinicians aim for carbamazepine blood concentrations between 4 to 15 µg/mL (the blood concentrations with established therapeutic efficacy in the treatment of seizures). Carbamazepine has been associated with aplastic anemia in up to 1 in 10,000 patients treated with the agent, although more recent research has revised the occurrence of severe hematological adverse effects to about 1 in 125,000. However, decreases in the WBC count are common in patients taking this medication. In addition, a benign drop in the red blood cell count may occur within the first week of treatment, which reverts to normal without the drug having to be discontinued. Recommendations for the discontinuation of carbamazepine treatment have included WBC counts less than 3000 per mm $_3$, erythrocyte counts less than 4.0 × 10$_6$/mm$_3$, hemoglobin less than 11 mg/dL, platelet counts less than 100,000 per mm$_3$, or reticulocyte counts less than 0.3 percent. A conservative plan for the laboratory monitoring of patients taking carbamazepine is provided in Table 7.7-7.

Test	Frequency
1. Complete blood count	Before treatment and every 2 weeks for the first 2 months of treatment; thereafter, once every 3 months
2. Platelet count and reticulocyte count	Before treatment and yearly
3. Serum electrolytes	Before treatment and yearly
4. Electrocardiogram	Before treatment and yearly
5. AST, ALT, LDH, alkaline phosphatase	Before treatment and every month for the first 2 months of treatment; thereafter, every 3 months
6. Pregnancy test for women of childbearing age	Before treatment and as frequently as monthly in noncompliant patients

Reprinted with permission from MacKinnon A, Yudofsky SC: *Principles of the Psychiatric Evaluation*. Lippincott, Philadelphia, 1991.

Table 7.7-7 Laboratory Monitoring of Patients Taking Carbamazepine

Other less stringent recommendations for laboratory follow-up have suggested that during the first 2 months of carbamazepine treatment, serum carbamazepine concentrations be obtained every 1 to 2 weeks, and a CBC and LFTs be obtained monthly. Clinicians should be aware that carbamazepine induces liver enzymes that increase its own metabolism and lower blood concentrations during the initial 2 to 3 weeks of its use. Hence, some clinicians recommend more careful monitoring of carbamazepine blood levels during the first month of treatment. During long-term treatment, serum concentrations can be obtained every 3 to 6 months, and CBC and LFTs every 6 months.

Carbamazepine can also have effects similar to the tricyclic antidepressant drugs on the ECG. Like valproic acid and divalproex sodium, carbamazepine is potentially teratogenic; hence, pretreatment pregnancy testing is advised in women of childbearing potential. Carbamazepine has also been reported to reduce haloperidol blood levels with clinical deterioration of the patient. Concurrent use of carbamazepine and clozapine has been associated with reductions in plasma clozapine concentrations approaching 50 percent. Finally, carbamazepine has also been associated with lowering serum sodium concentrations, potentially progressing to hyponatremia and even water intoxication (syndrome of inappropriate antidiuretic hormone secretion). Hence, some clinicians recommend a baseline serum sodium determination in patients who are going to be started on carbamazepine, as well as periodic follow-up determinations.

BRAIN IMAGING

Computed Tomography CT is used by neuropsychiatrists to exclude the presence of brain lesions in their patients. There is no evidence in support of making functional psychiatric diagnoses based on CT findings. There has been some controversy in the medical literature regarding the use of this diagnostic procedure as a screening tool. Most physicians would agree that psychiatric patients with focal neurological findings, a past history suggestive of brain insult (e.g., a history of head trauma or seizures), an abnormal EEG, or delirium should be seriously considered for a CT scan of the head. Others have proposed that CT scans should be obtained in psychiatric patients who evidence either confusion, dementia, delirium, movement disorder, anorexia nervosa, prolonged catatonia, or who are psychotic for the first time. Some also recommend a CT scan in patients over the age of 50 who show a personality change or who are having their first episode of mood disorder. A normal CT scan result oes not entirely rule out the possibility of organic brain syndrome or insult.

Magnetic Resonance Imaging MRI visualizes brain structures with remarkable clarity and offers better resolution of brain structure than CT. The indications for using MRI in a psychiatric patient are similar to those for CT scans. MRI is perhaps superior to CT scans when the suspected CNS disease process includes white matter demyelinating disease, nonmeningeal neoplasm, vascular malformation, degenerative disease such as Huntington's disease, and when a seizure focus is suspected. The MRI technique is also useful in patients unable to tolerate iodine-based contrast materials or intravenous procedures that might be needed during a CT scan evaluation. Additionally, the MRI can be used to clarify ambiguous CT scan findings.

The MRI technique is typically contraindicated in patients with pacemakers, aneurysm clips, pregnant women, and patients with potentially magnetic foreign bodies. Additionally, the CT scan is often recommended over MRI when the disease process suspected involves the pituitary gland, calcified brain lesions, meningeal tumors, and acute parenchymal infarction or hemorrhage. The clinician should be aware that a variety of anxiety reactions can occur during MRI scanning, requiring that some patients be carefully screened prior to scanning in the more claustrophobic MRI scanners and prepared for the experience (perhaps with friends and family on hand, with prescanning behavioral desensitization, or premedication with a short-acting antianxiety agent).

ELECTROPHYSIOLOGY

Electroencephalogram In patients with a possible organic CNS component to their psychiatric disturbance, an EEG can be useful. The clinician should realize that a single normal EEG result does not rule out organic pathology. Diffuse slowing is perhaps the most common EEG abnormality in psychiatric patients, constituting about 40 percent of abnormal EEGs in psychiatric populations. EEG slowing in psychiatric patients has been associated with diminished cognitive functioning, being on a larger number and higher doses of medications, increased length of hospital stay, and greater illness severity. If a seizure disorder is strongly suspected, “sampling errors” can be corrected by repeat EEGs or 24-hour ambulatory recording (often accompanied by video recordings of the patient's behavior to further document potential seizure activity). Sleep-deprived EEG recordings further increase the likelihood of unmasking latent abnormal EEG activity. Some clinicians use nasopharyngeal (NP) leads to increase the diagnostic yield. However, NP leads can be quite uncomfortable for patients and can keep them from falling asleep during EEG recording. The failure to sleep during the EEG recording can significantly diminish the usefulness of the EEG as a diagnostic tool because it is during sleep that latent abnormal EEG activity is often best detected. Clinicians should also remember that a seizure diagnosis is typically made on clinical grounds and is not entirely dependent on abnormal EEG findings. Finally, EEG monitoring is also used by some clinicians and researchers during the administration of electroconvulsive therapy (ECT). Such EEG monitoring assures the clinician that a fully therapeutic seizure has been induced.

Polysomnography This procedure is used in the evaluation of various sleep disorders (e.g., insomnia, sleep apnea, parasomnias, narcolepsy, male erectile disorder). Polysomnography involves the assessment of multiple physiological parameters while the patient sleeps (or attempts to sleep); the parameters include EEG, electro-oculogram (EOG), electromyogram (EMG), ECG, blood oxygen saturation, blood pressure, respiratory effort, and body temperature. A special daytime polysomnographic evaluation called the multiple sleep latency test is used in the evaluation of narcolepsy.

Evoked Potentials Evoked potentials refer to brain electrical activity elicited by stimuli. Exogenous, or sensory, evoked potentials (visual, auditory, and somatosensory) have great utility in detecting peripheral and central neural conduction abnormalities. Evoked potential testing can be used in the differentiation of certain organic versus functional complaints (e.g., the visual evoked potential [VEP] can be used to evaluate hystorical blindness). Other forms of evoked potential studies include the brainstem auditory evoked potential (BAEP). When neuropsychiatric symptoms of unclear etiology are present, certain evoked potential findings can be useful for detecting possible underlying neurological dysfunction (e.g., evoked potential testing might suggest an underlying demyelinating condition such as multiple sclerosis as a cause of what was previously regarded as “functional” symptoms).

Genetic Testing Clinical chromosome analysis usually takes place in the context of a prenatal analysis of genetic disorders. Possible clinical indications for ordering a cytogenetic analysis in a psychiatric patient who is not in a gynecologic or obstetric setting might include the psychiatric patient with congenital anomalies or nonspecific mental retardation. If a patient has cytogenetic analysis performed for some reason (e.g., as part of the workup of infertility, habitual abortion, amenorrhea, or ambiguous external genitalia), careful attention needs to be paid to the potential adverse psychological impact of the result on the patient and family.

AMOBARBITAL INTERVIEW

The amobarbital interview involves the slow intravenous infusion of sodium amytal, with careful attention to avoid oversedation and respiratory depression. Reports exist describing the amobarbital interview as a useful aide in differentiating certain functional versus organic conditions. For instance, in certain stuporous patients with schizophrenia or major depressive disorder, amobarbital has been described as making the patient more verbal and less guarded. Patients whose stuporous state is secondary to a neuromedical condition (e.g., brain tumor, stroke) typically become more confused, with clear deterioration on cognitive examination. An alternative medication that has been proposed as perhaps better than amobarbital for drug-assisted interviewing is the high-potency benzodiazepine midazolam (Versed); for safety, pulse oximetry has been used in conjunction with midazolam administration.

NEW CLINICAL AND RESEARCH TOOLS

Biological Markers There has been a great deal of interest in finding neurophysiological markers of psychiatric disorders. Such biological markers would not only reveal underlying pathophysiologies for psychiatric disorders, but they could also serve as diagnostic tests. Biological markers could assist in: (1) making accurate psychiatric diagnoses, (2) identifying patients at risk of developing a psychiatric disorder (e.g., so preventive steps could be implemented), and (3) predicting treatment response. Biological markers currently being investigated encompass a wide range of procedures, some of which are listed in Table 7.7-8. Difficulties that have arisen for many of the proposed markers include problems with sensitivity, specificity, reliability, and contamination from artifactual influences (e.g., concurrent medical illnesses, medication effects, and normal individual variation). None of these markers have clearly established themselves as useful for routine clinical practice. More precise biological markers for idiopathic psychiatric conditions of greater clinical and research utility will be most forthcoming when the underlying pathophysiologies of these conditions are determined. Because the etiologies and genetic contributions of psychiatric disorder largely remain unknown, biological marker research in psychiatry often takes on the character of a fishing expedition with better fishing spots suggested by earlier encouraging findings or intriguing hypotheses.

Table 7.7-8 Some Biologic Markers Under Investigation

Brain Imaging Techniques An exciting area of biological marker research involves several brain imaging techniques that have become available to neuropsychiatrists and it is understandable that tests that directly measure and image brain structure and function are of interest in psychiatry.

Electroencephalography Computerized topographic mapping of electrophysiological data represents an enhancement of conventional EEG and evoked potential (EP) testing. In this brain-imaging tool, a computer is used to amass and process large quantities of EEG and EP data from a patient. The computers analyze the data in various ways and graphically present the data as two-dimensional, typically color-coded maps of brain electrical activity. The newer technology of computerized topographic mapping of EEG and EP data has not found a clear niche in the diagnostic assessment of the psychiatric patient. There has also been the hope that the computerized EEG would help in the development of the EEG as a functional measure of the therapeutic activity of various psychiatric drugs.

Recent EEG work in patients with schizophrenia treated with clozapine has suggested that clozapine “responders” demonstrate increased EEG photic driving compared to nonresponders. The increase in photic driving was positively correlated with clinical improvement. Indeed, patients with schizophrenia who develop abnormal EEGs while on clozapine tend to evidence a better clozapine response compared to those patients whose EEGs are unaffected by clozapine. Even when not taking any psychiatric medications, patients with schizophrenia tend to display EEG changes compared to normals, such as lower alpha frequencies (which have been shown to correlate negatively with total positive symptom scores in these patients).

Sleep EEG (Polysomnography) A potentially powerful biological marker of psychiatric illness uses EEG obtained during sleep. Sleep EEG abnormalities described in major depressive disorder include an increase in the overall amount of rapid-eye movement (REM) sleep and a shortened period before the onset of REM sleep (shortened REM latency). Neuromedical conditions giving rise to psuedodepressive presentations are typically associated with decreased REM sleep (e.g., patients with dementia usually have increased amounts of non-REM sleep). In schizophrenia, decreased slow-wave sleep and increased sleep latency and fragmentation have been reported, especially during relapse. Computerized analyses of the sleep EEG, with the generation of topographic maps of sleep EEG activity, are also available.

Computed Tomography Researchers continue to study possible useful subcategories of psychiatric diagnoses that could be based on CT findings (e.g., schizophrenia with or without ventricular enlargement). For instance, there has been the suggestion that certain patients with schizophrenia who have enlarged ventricles might be more resistant to treatment with antipsychotic agents. For instance, increased prefrontal cortical sulcal widening has been associated with a poor response to clozapine. In patients with chronic alcohol dependence, a relationship between severity of frontal atrophy and negative symptoms (as measured with the Scale for the Assessment of Negative Symptoms) has been reported, although the ability of such atrophy to predict future abstinence and treatment response has not been examined. However, there is currently no evidence that supports making functional psychiatric diagnoses based on CT scan abnormalities.

Magnetic Resonance Imaging Various neuropsychiatric studies using the MRI technique have yielded some interesting findings. Patients with schizophrenia have been reported to have smaller frontal lobes, cerebrums, craniums, and hippocampi (but larger lenticular nuclei and thalami) than normal controls. Differences in the symmetries of various brain structures have also been noted when subjects with schizophrenia and normal control subjects are compared. In elderly depressed patients, MRI signal hyperintensities in subcortical gray matter have been reported to be more common than in nondepressed comparison subjects. The full utility of MRI in neuropsychiatry, with its many different scanning modes, has not yet been completely explored.

Functional MRI Although traditional MRI provides anatomical images of the brain, the functional MRI (fMRI) technique might ultimately provide the researcher and clinician images of brain activity that rival the clarity of MRI anatomical images. fMRI can be directly correlated with high-resolution, three-dimensional anatomical MRI images. In a recent fMRI study it was reported that patients with schizophrenia demonstrated less frontal lobe activation and greater temporal lobe activation on the left during a word fluency task compared to comparison subjects. The utility of this technique for informing researchers about brain functioning in psychiatric illnesses is in its infancy, but it is expected that there will be many indications for this procedure in the future.

Magnetic Resonance Spectroscopy Magnetic resonance spectroscopy (MRS) is related to MRI but uses more powerful magnetic fields to evaluate certain aspects of brain function and metabolism. Information about brain phospholipid, carbohydrate, protein, amino acid, and high-energy phosphate metabolism, brain intracellular pH, as well as information about lithium and fluorinated psychopharmacological agents, can be obtained by this technique. For instance, using phosphorus-31 nuclear magnetic resonance ($_{31}$P NMR) spectroscopy, metabolism of brain high-energy phosphate and membrane phospholipids was studied in the dorsal prefrontal cortex of drug-naive patients with schizophrenia. Significant differences were found between the patients and controls, with dorsal prefrontal cortex hypoactivity in patients suggested by decreased adenosine triphosphate and inorganic orthophosphate levels. Alterations in membrane phospholipid metabolism were suggested by lower levels of phosphomonoesters and higher levels of phosphodiesters compared to the control subjects. Technological improvements in MRS should further the contributions of this functional brain-imaging technique to neuropsychiatric research.

Positron Emission Tomography PET allows for the direct visualization of cortical and subcortical brain functioning. Depending on the type of positron-emitting isotope used, different aspects of brain functioning can be studied, including brain glucose metabolism, cerebral blood flow, brain oxygen use, and binding parameters of specific neurotransmitter receptors. PET findings include abnormalities of the anteroposterior gradient of glucose utilization and higher subcortical-to-cortical glucose metabolism ratios in patients with schizophrenia compared to normal controls. PET findings in schizophrenia, as well as in other psychiatric disorders such as bipolar I disorder, substance use disorders, obsessive-compulsive disorders, and panic disorders, have not always been replicated by other investigators. PET remains a research tool available at only a few research centers.

Single Photon Emission Computerized Tomography Like PET, SPECT visualizes both cortical and subcortical brain activity. Because of certain technical differences, the image resolution is typically not as good with SPECT as compared to state-of-the-art PET scanners. Depending on the radioisotope employed, different aspects of brain function can be studied, such as brain blood flow. SPECT may also be utilized to image the distribution of neurotransmitter receptors in vivo. SPECT scanning defects have been described in a wide variety of neuropsychiatric conditions. For instance, one study reported an average of 7.3 SPECT defects in patients with chronic fatigue syndrome versus a mean of 0.4 defects in controls. Additionally, SPECT changes were reported to correlate with changes in clinical status. Another study reported a similar number of SPECT scan defects in a group of patients with chronic fatigue syndrome and a group of patients with major depressive disorder (6.5 defects per patients versus 1.7 in healthy controls).

Magnetoencephalography Conventional and computerized EEG gather data based on measurements made of cortical brain activity. Magnetoencephalography (MEG) detects the magnetic fields associated with neuronal electrical activity in both cortical as well as deeper brain tissues. MEG is noninvasive and does not expose the patient to harmful radiation; MEG is available in only a few research centers.

Regional Cerebral Blood Flow Regional cerebral blood flow (rCBF) is a nuclear medicine technique involving the introduction of metabolically inert radioactive substances into the body. These substances arrive at the brain via the blood, and radiation emanating from the brain is picked up by detectors surrounding the skull. rCBF can delineate blood flow in cortical structures. Using rCBF, there have been reports of decreased blood flow to certain frontal regions of the brain in patients with schizophrenia as well as in patients with other neuropsychiatric disorders.

Other Neurophysiological Markers

Eye Movement Deficits (EMD) A deficit in smooth pursuit eye movement (SPEM) performance has been described in up to 85 percent of patients with schizophrenia (versus approximately 10 percent in control populations). The SPEM deficit has been proposed to be a biological marker of schizophrenia. Another EMD involves an antisaccade deficit, which has been described in a number of neuropsychiatric conditions, including in patients with frontal cortical and basal ganglia lesions, schizophrenia, and obsessive-compulsive disorder. The antisaccade performance deficit includes a decreased ability to inhibit reflexive glances toward stimuli that subjects are instructed to look away from. Schizophrenia patients with greater smooth pursuit deficit tend to show more antisaccade performance deficits. EMDs can be measured using a number of different techniques, including electrooculography (EOG) and infrared scleral reflectance techniques.

Electroretinogram The electroretinogram (ERG) measures electrical changes produced in the retina by flashes of light. ERG amplitudes are reduced in certain diseases (e.g., retinitis pigmentosa and myotonic dystrophy), and also during the prodromes for these disorders. Decreased ERG amplitude has been reported in cases of thioridazine toxicity. In neuropsychiatric conditions, ERG amplitude reductions (specifically beta-wave amplitude) have been described in some patients with autistic disorder and those with various types of dementia, such as dementia of the Alzheimer's type and Creutzfeldt-Jakob disease.

Pupillometry It has been reported that persons with dementia of the Alzheimer's type have an exaggerated pupil dilation response to the ophthalmic administration of the anticholinergic medication tropicamide (Mydriacyl) compared to persons without the disease. The pupil dilation response of individuals with forms of dementia other than of the Alzheimer's type was reported as similar to that of the healthy controls, suggesting that this pupillary marker might be more specific for Alzheimer's disease than for other forms of dementia. The pupil dilation response was also reported as more sensitive for the presence of premorbid and morbid Alzheimer's disease than conventional neuropsychological testing. However, these findings based on pupil size measures have recently been challenged. One recent report suggests that increases in pupillary oscillation might accompany prodromal mania, supporting other suggestions that pupillometry might be useful for clinical psychophysiological diagnosis and management of psychiatric patients.

Quantitative Measures of Visual Scanning Such quantitative visual scanning procedures are performed on data obtained while patients view various visual stimuli. One method of quantitative visual scanning assessment is based on eye movement research suggesting that it takes at least approximately 40 to 50 milliseconds (ms) to fully scrutinize and cognitively process a single image element. Hence, visual fixations of shorter duration are thought to subserve more automatic or preattentive cognitive processes. In schizophrenia, such quantitative assessments of eye movements while patients viewed slides of faces were reported to correlate with symptom severity. In cocaine addicts, quantitative visual scanning measures while viewing cocaine cues have been shown to correlate with measures of cocaine craving.

Endocrine Stimulation Techniques

Dexamethasone Suppression Test (DST) There was the hope that this laboratory test would prove useful for the diagnosis and management of psychiatric patients. While a version of this test is used routinely by endocrinologists in the evaluation of Cushing's disease, the test was found by research psychiatrists to be a possible marker of major depressive disorder. There was initial enthusiasm that the test would be useful in assisting with psychiatric differential diagnosis (e.g., differentiating dementia from the pseudo-dementia of depression; differentiating schizoaffective disorder from schizophrenia) and in the prediction of treatment response and suicidal risk. Although many of the research findings have proved interesting, it has been argued that overall they do not support the routine clinical use of this test. However, although pretreatment baseline DST results are of questionable value, many agree that posttreatment DST cortisol nonsuppression predicts a poor outcome in depressive disorders, with a higher risk of early relapse.

A common version of the DST as employed by psychiatrists involves the administration of a 1-mg dose of dexamethasone (Decadron) at 11 PM with a serum cortisol concentration determination prior to dexamethasone administration (to assess baseline cortisol concentrations) and other serum cortisol determinations at various points over the next 24 hours (typically 8 AM, 4 PM, and 11 PM). The DST result is considered abnormal if the postdexamethasone serum cortisol concentration exceeds about 5 µg/dL. However, the reliability of many commercial assays at about the 5 µg/dL cut-off value is questionable; hence, there is a recommendation that plasma cortisol concentrations between 4 and 7 µg/dL be interpreted cautiously. Additionally, many other limitations to the routine use of the DST have emerged, including significant potential artifactual contamination from concomitant medical conditions (e.g., diabetes mellitus), medications (e.g., carbamazepine) and the potential that the test results will direct the clinician's attention away from other more important clinical issues for the patient. If the DST is used, the psychiatrist should be attentive to the possible causes of false-positive or false-negative DST results outlined in Table 7.7-9.

False Positives	False Negatives
Cushing's syndrome	Addison's disease
Weight loss or malnutrition	Hypopituitarism
Obesity	Slow dexamethasone metabolism
Bulimia nervosa	Drugs
Pregnancy	Synthetic corticosteroids
Alcohol abuse and withdrawal	Indomethacin
Anorexia nervosa	High doses of benzodiazepines
Temporal lobe epilepsy	High doses of cyproheptadine
Dementia	
Diabetes mellitus	
Infection	
Trauma	
Recent surgery	
Advanced age	
Fever	
Carcinoma	
Renal or cardiac failure	
Renovascular hypertension	
Cerebrovascular disorder	
Antipsychotic withdrawal	
Tricyclic withdrawal	
Drugs	
High doses of estrogens	
Narcotics	
Sedative-hypnotics	
Anticonvulsants	

Adapted from Carroll [1, 2nd, 3rd]

Table 7.7-9 Causes of False-Positive or Negative Results on the Dexamethasone-Suppression Test (DST)

Thyrotropin-Releasing Hormone Stimulation Test Endocrinologists use a version of this test in the evaluation of hypothyroidism. The hypothyroid patient has an augmented TSH response to an intravenous (IV) injection of thyrotropin (Thytropar). The test has also been proposed as useful in grading different forms of hypothyroidism. In Grade 1 hypothyroidism the diagnosis is obvious based on clinical and biochemical grounds (e.g., elevated serum TSH and low serum T_4). Grade 2 hypothyroidism is less obvious and without clear clinical features of hypothyroidism (except perhaps depression), but serum TSH is elevated. Grade 3 hypothyroidism is characterized by normal baseline TSH and thyroid hormone values, but the results of the TRHST are abnormal. It has been argued that Grade 3 hypothyroidism can also present with or accompany depression in some patients, and would be missed if only the conventional serum thyroid hormone assays are used (i.e., serum TSH and T_4). Furthermore (and perhaps confusingly so) a blunted TRHST result (i.e., a result more common to hyperthyroidism) has been proposed as a biological marker of major depressive disorder with diagnostic and prognostic meaning. Although a blunted TRHST result has been reported in up to 30 percent of depressed patients, the result does not appear specific to major depressive disorder (e.g., blunted results have been reported in patients with other affective spectrum disorders such as alcoholism, panic disorder, bulimia nervosa, and borderline personality disorder).

Provocative Tests of Panic Disorder These tests challenge patients with various agents such as carbon dioxide (via inhalation), lactate (via IV infusions) or IV infusions of such substances as caffeine, isoproterenol (Isuprel) b-carboline and flumazenil (Romazicon)—all of which have been noted to be capable of inducing panic attacks in patients so predisposed (and less so in individuals without a history of panic disorder). Lactate infusion is the most extensively studied provocative test for panic disorder, and up to 72 percent of panic disorder patients have been reported as experiencing a panic attack after the IV infusion of sodium lactate. Anxiogenic responses to lactate or fenfluramine provocation have been described as greater in patients with higher panic attack frequencies. In patients with posttraumatic stress disorder, IV infusions of lactate have been demonstrated as capable of triggering flashbacks of the traumatic event, with at times dramatic affective displays in response to the flashback.

Panic attacks precipitated by IV lactate infusion have been shown to be inhibited by tricyclic antidepressant medication and alprazolam (Xanax), which are psychotherapeutic agents of demonstrated efficacy in the treatment of panic disorder. Responses to challenge tests for anxiety disorders might lead to useful subtyping paradigms for these disorders. The specificity of lactate infusions for panic disorder is unclear, as patients with primary depressive disorder with secondary panic attacks have similar rates of lactate-induced panic attacks as those patients with primary panic disorder. Simple hyperventilation procedures have also been reported as capable of precipitating panic attacks in patients so predisposed, although hyperventilation appears to be a less potent stimulus of panic attacks than IV infusions of lactate. The clinician should note that hyperventilation procedures should be used with caution because of their ability to trigger seizures. At this time, these provocative tests are more useful as research tools to study various paroxysmal anxiety disorders than they are as diagnostic tests.

Biochemical Markers From blood, CSF, and urine, the clinician and researcher can obtain many potential biochemical markers, including neurotransmitter substances and their metabolites (e.g., dopamine, homovanillic acid [HVA], norepinephrine, its metabolite 3-methoxy-4-hydroxyphenylglycol [MHPG], serotonin and its metabolite 5-hydroxyindoleacetic acid (5-HIAA), and amino acids such as tryptophan, tyrosine, glycine, and glutamate). Studies involving most of these markers and many other proposed markers have yielded mixed results. What follows are some interesting recent highlights regarding this area of research.

Plasma Homovanillic Acid The measurement of homovanillic acid in plasma (pHVA), a major acidic metabolite of dopamine, may reflect changes in the activity of dopamine in the brain. The measurement of pHVA may have a practical clinical application in the identification of schizophrenia patients responsive to antipsychotic medications. For instance, some reports suggest that responders to these drugs are characterized by antipsychotic-induced reductions in pHVA. The time-dependent decreases in pHVA in treatment responders further suggest that the ability to dampen presynaptic dopaminergic activity is related to the therapeutic efficacy of antipsychotic medications.

3-Methoxy-4-Hydroxyphenylglycol When measured in 24-hour urine collections, MHPG has been reported to be lower in patients with bipolar I disorder than in patients with bipolar II disorder. Low urine MHPG might predict imipramine (Iofranil) response in depressed patients with bipolar disorder but not in depressive disorder (unipolar) patients. High urinary MHPG concentrations have also been associated with a subtype of depressed patients more likely to show cognitive features of learned helplessness.

Low CSF MHPG (along with decreased 24-hour urinary norepinephrine-epinephrine ratios) has been proposed as a measure of a decreased noradrenergic activity that could suggest an increased predisposition to suicidal behavior.

5-Hydroxyindoleacetic Acid Associations have been reported between low CSF 5-HIAA and suicidal behavior, aggression, poor impulse control and impulsive suicidal acts, disturbed behavior in childhood, violent suicide attempts, and depression in patients with diagnoses including major depressive disorder, schizophrenia, alcohol use disorders, and adjustment disorder. Hence, although this marker might not have nosological specificity, it might mark problematic behaviors across a wide variety of psychiatric diagnoses. There have also been reports of an association between elevated CSF 5-HIAA concentrations and anxious, obsessional, or inhibited behaviors.

Markers of Alzheimer's Disease SPECT and PET are emerging as useful tools in the detection and evaluation of patients with dementia of the Alzheimer's type. When SPECT and PET are used, these patients typically show decreased perfusion in temporoparietal areas bilaterally, while patients with ischemic brain changes (vascular dementia) more often demonstrate single focal perfusion defects or multiple areas of patchy hypoperfusion.

The presence of an *Apo*E allele has also been associated with an increased risk of Alzheimer's disease. Individuals with two copies of the *Apo*E-e4 allele appear to have an especially increased risk. By late middle age, persons who are cognitively normal but who are homozygous for the e4 allele for *Apo*E have been reported to show evidence of reduced glucose metabolism on PET in the regions of the brain as do patients with dementia of the Alzheimer's type. These findings have been interpreted as being supportive of the notion that *Apo*E-e4 is a risk factor for this disorder.

A variety of other biological markers of Alzheimer's disease are currently under examination. One test for neural thread protein (NTP) from CSF is now being marketed for assisting in the diagnosis of these patients. The test measures the brain protein 21 kD NTP that is reported to be increased in patients with dementia of the Alzheimer's type as compared to normal controls. Other potential CSF tests for this disorder include measurement of CSF-tau (increased) and CSF-amyloid (decreased). One proposed biochemical marker of blood-brain barrier damage in dementia caused by cerebrovascular pathology is the CSF albumin-to-serum albumin ratio. The CSF albumin-to-serum albumin ratio is normal in patients with dementia of the Alzheimer's type but elevated in vascular dementia. The hope is that all these emerging diagnostic tests will be able to diagnose the disease early in its course. Clinicians would then be able to start drug treatments early, when such interventions would have the greatest potential of being effective. Persons with vascular dementias can be identified by CT, MRI, SPECT, or PET scan.

Inflammatory mechanisms have been implicated in the pathogenesis of dementia of the Alzheimer's type, and certain biochemical and genetic markers of inflammation may serve as diagnostic aids in the identification of persons affected with the disease or at risk of developing it. For instance, there have been reports of elevations of acute-phase reactive proteins in sera and CSF (e.g., a_1-antichymotrypsin) of patients and unaffected first-degree relatives of patients with this disorder. These markers of inflammation may become targets for novel pharmacotherapy and laboratory indices to monitor therapeutic response to treatment with anti-inflammatory medications. Additionally, knowledge of the normal handling of proteolytic products of amyloid precursor protein (APP) and factors mediating inflammation in brain of these patients may help to identify targets for novel pharmacotherapy and genetic markers of the disease. For instance, measurements of (1) apolipoprotein J and allelic variants (apolipoprotein J has been shown to be the predominant binding protein for the b-amyloid split product in CSF), (2) complexes of apoJ with soluble b-amyloid and soluble complement membrane-attack complex (MAC), and (3) low-density lipoprotein receptor (LRP-2) and allelic variants, may serve as potential risk markers for Alzheimer's disease.

Genetic Testing In theory, markers for the defective genes in the heritable neuropsychiatric disorders ultimately will be identified; these markers should prove to be powerful diagnostic tests for these diseases. At this time no functional (idiopathic) psychiatric disorder has been associated clearly with a specific chromosome or gene location. Using new molecular biology techniques, studies are beginning to suggest chromosomal loci for a host of neuropsychiatric problems.

For instance, independent groups have provided evidence for a vulnerability locus for schizophrenia and associated disorders on chromosome 8p. This vulnerability locus on 8p was likely to segregate in about 10 to 25 percent of the multiplex families studied. Data from this series have also shown potential linkage between schizophrenia and a vulnerability locus on chromosome 6p in about 10 to 30 percent of the multiplex families. Thus, there is an interesting possibility that epistasis or gene-gene interaction can take place between the 8p and 6p regions in at least some patients. Another interesting allelic marker whose absence in patients and mothers of offspring with schizophrenia may be associated with schizophrenia is the locus (DRB1*04) encoding the class II human leukocyte antigen (HLA) DR4 on chromosome 6p. There have also been reports of an association between schizophrenia and the absence of DQB1*0602, another class II HLA gene, in African-Americans and Chinese men from Singapore.

Evidence has been collected suggestive of a susceptibility gene for bipolar I disorder in the pericentrometric region of chromosome 18. Candidate genes in this region that may have some etiological or pathological association with bipolar I disorder include Golf, the a subunit of a heterotrimeric GTP-binding protein, and a corticotropin receptor gene. The lithium ion influences the affinity of the a subunit for GTP, antidepressant medications modulate expression of GTP-binding proteins, and the function of the hypothalamic-pituitary-adrenal axis is altered in mood disorders.

In Alzheimer's disease, genetic defects on chromosomes 21, 14, and 19 have been implicated. The presence of an *Apo*E allele has also been associated with an increased risk of the disease, but the presence of the allele is not considered diagnostic. *Apo*E maps to a region on chromosome 19 that has been implicated in linkage studies of the disease. Recent evidence suggests that genetic screening of mutant presenilin alleles (presenilin genes located on chromosomes 14 and 1 have been designated as *PS*1 and *PS*2, respectively) might also lead to possible diagnostic genetic tests of persons at risk. Ultimately, batteries of tests might be needed to fully determine the at-risk profile of an individual for a particular neuropsychiatric disease.

The further extension of genotypic analyses, including restriction fragmentation length polymorphism, polymerase chain reaction amplification, and sequence-specific oligonucleotide probe analysis, and specialized statistical procedures to assess linkage to complex traits will undoubtedly result in the identification of genetic risk factors for neuropsychiatric illnesses such as schizophrenia, bipolar I disorder, and dementia of the Alzheimer's type. The discovery of genetic defects responsible for neuropsychiatric disease, along with the uncovering of their roles in the diseases and biological products, will revolutionize laboratory testing of these conditions. At this time these genetic procedures and markers remain only research tools. However, there is a substantial effort to identify genes responsible for specific diseases sponsored by the federal government's human genome project, which involves the mapping of the entire human genome (composed of over 100,000 genes, of which approximately 80,000 are represented in the brain). The mapping of the human genome should enhance the clinician's ability to diagnose neuropsychiatric disease and will result in new conceptualizations of psychiatric disorders. It is hoped that this project will lead to better preventive and therapeutic measures for these disorders. Nevertheless, a host of clinical and ethical issues pertaining to the interpretation of these genetic risk factors and the counseling of patients and families will be raised by this promising area of research.

SUGGESTED CROSS-REFERENCES

Assessment is also discussed in Section 2.1, the other sections of Chapter 7, Chapter 32, Chapter 33, Chapter 34, Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39, Chapter 40, Chapter 41, Chapter 42, Chapter 43, Chapter 44, Chapter 45, Chapter 46, Chapter 47, Chapter 48 and Chapter 49 on child psychiatry, and

Chapter 51 on geriatric psychiatry. Neuroimaging is discussed in Section 1.15 and Section 1.16. Substance-related disorders are discussed in Chapter 11, schizophrenia in Chapter 12, mood disorders in Chapter 14, and anxiety disorders in Chapter 15. Endocrine and metabolic disorders are discussed in Section 25.6; dementia of the Alzheimer's type is discussed in Chapter 10.

SECTION REFERENCES

Ananth J, Gamal R, Miller M, Wohl M, Vandewater S: Is the routine CT head scan justified in psychiatric patients? A prospective study. J Psychiatry Neurosci *18:*69, 1993.

*Anfinson TJ, Kathol RG: Screening laboratory evaluation in psychiatric patients: A review. Gen Hospital Psychiatry *14:*248, 1992.

Bonne O, Krausz Y, Lerer B: SPECT imaging in psychiatry: A review. *14:*296, 1992.

Bouman WP, Pinner G, Johnson H: Incidence of selective serotonin reuptake inhibitor (SSRI) induced hyponatraemia due to the syndrome of inappropriate antidiuretic hormone (SIADH) secretion in the elderly. Int J Geriatr Psychiatry *13:*12, 1998.

Boutros NN: Diffuse electroencephalogram slowing in psychiatric patients: A preliminary report. J Psychiatry Neurosci *21:*259, 1996.

Bowden CL, Janicak PG, Orsulak P, Swann AC, Davis JM, Calabrese JR, Goodnick P, Small JG, Rush AJ, Kimmel SE, Risch SC, Morris DD: Relation of serum valproate concentration to response in mania. Am J Psychiatry *153:*765, 1996.

Burke MJ, Preskorn SH: Short-term treatment of mood disorders with standard antidepressants. In *Psychopharmacology: The Fourth Generation of Progress*, Bloom FE, Kupfer DJ, editors. Raven Press, New York, 1995.

Denburg SD, Carbotte RM, Ginsberg JS, Denburg JA: The relationship of antiphospholipid antibodies to cognitive function in patients with systemic lupus erythematosus. J Int Neuropsychol Soc *3:*377, 1997.

*D'Ercole A, Skodol AE, Struening E, Curtis J, Millman J: Diagnosis of physical illness in psychiatric patients using Axis III and a standardized medical history. Hosp Comm Psychiatry *42:*395, 1991.

Elia J, Gulotta C, Rose SR, Marin G, Rapoport JL: Thyroid function and attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry *33:*169, 1994.

Expert Consensus Panel for Bipolar Disorder: Treatment of bipolar disorder. J Clin Psychiatry *57:*12A, 1996.

Expert Consensus Panel for Schizophrenia: Treatment of Schizophrenia. J Clin Psychiatry *57*(Suppl):12B, 1996.

Freeman DJ, Oyewumi LK: Will routine therapeutic drug monitoring have a place in clozapine therapy? Clin Pharmacokinet *32:*93, 1997.

Giedd JN, Rapoport JL, Leonard HL, Richter D, Swedo SE: Case study: Acute basal ganglia enlargement and obsessive-compulsive symptoms in an adolescent boy. J Am Acad Child Adolesc Psychiatry *35:*913, 1996.

Goldenberg DL: Fibromyalgia, chronic fatigue syndrome, and myofascial pain syndrome. Curr Opin Rheumatol *7:*127, 1995.

Greenwald BS, Kramer-Ginsberg E, Krishnan RR, Ashtari M, Aupperle PM, Patel M: MRI signal hyperintensities in geriatric depression. Am J Psychiatry *153:*1212, 1996.

Grunberger J, Linzmayer L, Grunberger S, Saletu B: Pupillometry in clinical psychophysiological diagnostics: Methodology and proposals for application in psychiatry. Isr J Psychiatry Relat Sci *29:*100, 1992.

*Hollister LE: Electrocardiographic screening in psychiatric patients. J Clin Psychiatry *56:*26, 1995.

Keshavan MS, Kapur S, Pettegrew JW: Magnetic resonance spectroscopy in psychiatry: Potential pitfalls and promise. Am J Psychiatry *148:*976, 1991.

Hughes JR: A review of the usefulness of the standard EEG in psychiatry. Clin Electroencephalogr *27:*35, 1996.

Kahn EM: Imaging of brain electrophysiologic activity: Applications in psychiatry. Gen Hosp Psychiatry *14:*99, 1992.

Kronig MH, Munne RA, Szymanski S, Safferman AZ, Pollack S, Cooper T, Kane JM, Lieberman JA: Plasma clozapine levels and clinical response for treatment-refractory schizophrenic patients. Am J Psychiatry *152:*179, 1995.

Levin JM, Ross MH, Renshaw PF: Clinical applications of functional MRI in neuropsychiatry. J Neuropsychiatry Clin Neurosci *7:*511, 1995.

Maier M: In vivo magnetic resonance spectroscopy. Applications in psychiatry. Br J Psychiatry *167:*299, 1995.

Marcum JM: The use of midazolam with pulse oximetry in the drug-assisted interview. J Clin Psychiatry *57:*111, 1996.

Mookhoek EJ, Sterrenburg VD, Nieuwegiessen IM: Screening for somatic disease in elderly psychiatric patients. Gen Hosp Psychiatry *20:*102, 1998.

Moseley ME, deCrespigny A, Spielman DM: Magnetic resonance imaging of human brain function. Surg Neurol *45:*385, 1996.

*National Institutes of Health Consensus Development Panel: Differential diagnosis of dementing disease. Nat Inst Health Consens Dev Conf Consens Statement *6:*1, 1987.

Nilsson K, Gustafson L, Faldt R, Anderson A, Vaara I, Nilsson R, Alm B, Hultberg B: Plasma methylmalonic acid in relation to serum cobalamin and plasma homocysteine in a psychogeriatric population and the effect of cobalamin treatment. Int J Geriatr Psychiatry *12:*67, 1997.

Perry JC, Jacobs D: Overview: Clinical applications of the amytal interview in psychiatric emergency settings. Am J Psychiatry *139:*552, 1982.

Ribeiro SCM, Tandon R, Grunhaus L, Greden JF: The DST as a predictor of outcome in depression: A meta-analysis. Am J Psychiatry *150:*1618, 1993.

Rosse RB, Giese AA, Deutsch SI, Morihisa JM: *A Concise Guide to Laboratory and Diagnostic Testing in Psychiatry*. American Psychiatric Press, Washington, DC, 1989.

Sachdev P, Mason C, Hadzi-Pavlovic D: Case-control study of neuroleptic malignant syndrome. Am J Psychiatry *154:*1156, 1997.

Scinto LFM, Daffner KR, Dressler D, Ransil BI, Rentz D, Weintraub S, Mesulam M, Potter H. A potential noninvasive neurobiological test for Alzheimer's disease. Science *266:*1051, 1994.

Siegler EL, Tamres D, Berline JA, Allen-Taylor L, Strom BL, et al: Risk factors for the development of hyponatremia in psychiatric inpatients. Arch Intern Med *155:*953, 1995.

Sloan EP, Fenton GW, Kennedy NS, MacLennan JM: Electroencephalography and single photon emission computed tomography in dementia: A comparative study. Psychol Med *25:*631, 1995.

Small GW, Rabins PV, Barry PP, Buckholtz NS, DeKosky ST, Ferris SH, Finkel SI, Gwyther LP, Khachaturian ZS, Lebowitz BD, McRae TD, Morris JC, Oakley F, Schneider LS, Streim JE, Sunderland T, Teri LA, Tune LE: Diagnosis and treatment of Alzheimer disease and related disorders. Consensus statement of the American Association of Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society. JAMA *278:*1363, 1997.

Swedo SE, Leonard HL, Mittleman BB, et al: Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. Am J Psychiatry *154:*110, 1997.

Tiihonen J, Vartiainen H, Hokola P: Carbamazepine-induced changes in plasma levels of neuroleptics. Pharmacopsychiatry *28:*26, 1995.

*Wahlsten D: Single-gene influences on brain and behavior. Annu Rev Psychol *50:*599, 1999.

*Weinberger DR: Brain disease and psychiatric illness: When should a psychiatrist order a CT scan? Am J Psychiatry *141:*1521, 1984.

Work Group on Bipolar Disorder: Practice guidelines for the treatment of patients with bipolar disorder. Am J Psychiatry *151*(Suppl):1, 1994.

Zametkin AJ, Ernst M, Silver R: Laboratory and diagnostic testing in child and adolescent psychiatry: a review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*464, 1998.

7.8 PSYCHIATRIC RATING SCALES

DEBORAH BLACKER, M.D., SC.D.

A variety of questionnaires, interviews, checklists, outcome assessments, and other instruments are available to inform psychiatric practice, research, and administration. These instruments, which are grouped here under the term *psychiatric rating scales*, are used with increasing frequency in the practice of psychiatry. Psychiatrists must be aware of rating scales for several reasons. Most critically, many such scales are useful in practice for monitoring patients over time or for providing information that is more comprehensive than that generally obtained in a routine clinical interview. In addition, scales are sometimes required administratively to justify the need for services or to assess quality of care. Last, but equally important, these scales are used in the research that informs the practice of psychiatry, so familiarity with them provides a deeper understanding of the results of that research and the degree to which it applies to psychiatric practice.

POTENTIAL BENEFITS AND LIMITATIONS OF RATING SCALES IN PSYCHIATRY

Rating scales in psychiatry serve to standardize the information collected across time or observers. This standardization ensures a comprehensive evaluation that may aid treatment planning by establishing a diagnosis, ensuring a thorough description of symptoms, identifying comorbid conditions, and characterizing other factors affecting treatment response. In addition, it can establish a baseline for follow-up of the progress of illness over time or in response to specific interventions. This is particularly useful when several clinicians are involved, for instance in a group practice or clinic setting, or in the conduct of psychiatric research.

In addition to standardization, most rating scales also offer the user the results of a formal evaluation of their performance characteristics. This means that the clinician can know to what extent a given scale produces reproducible results (*reliability*) and how it compares with more definitive or established ways of measuring the same thing (*validity*).

Rating scales also offer some practical advantages. First, they can save valuable physician time: self-administered rating scales can be administered in the waiting room, or a nurse or technician can administer an interview prior to a session with the physician. In addition, rating scales may make it easier to obtain information about sensitive areas such as cognitive decline or sexual functioning in which direct questioning is sometimes experienced as more intrusive.

However rating scales are not a panacea. They can provide erroneous measurements because of difficulties in administration or limitations in the underlying construct. In this respect they do not differ from clinical assessments, but they may appear to provide more definitive information and thus give a spurious sense of security. At the practical level, they take time that might better be devoted to other pursuits. The critical decision about using a formal assessment tool in clinical practice is whether on balance it contributes useful information in an efficient manner. This decision depends on the specific clinical setting and goal, the practical attributes of the scale, and its psychometric properties.

TYPES OF SCALES AND WHAT THEY MEASURE

Scales are used in psychiatric research and practice to achieve a variety of goals. They also cover a broad range of areas and use a broad range of procedures and formats.

Measurement Goals Most psychiatric rating scales in common use fall into one or more of the following categories: making a diagnosis (e.g., the Structured Clinical Interview for DSM-IV [SCID] or the Diagnostic Interview Schedule for Children [DISC]); measuring severity and tracking change in specific disorders (e.g., the Hamilton Rating Scale for Anxiety [HAM-A] or the Mini-Mental State Examination [MMSE]) or general symptoms (e.g., the Symptom Checklist-90) or in overall outcome (e.g., the Behavior and Symptom Identification Scale [BASIS-32]); screening for conditions that may or may not be present (e.g., the CAGE or the Zung Self-Rating Depression Scale).

Constructs Assessed Psychiatric practitioners and investigators assess a broad range of areas, referred to as *constructs* to underscore the fact that they are not simple, direct observations of nature. These include diagnoses, signs and symptoms, severity, functional impairment, quality of life, and many others. Some of these constructs are fairly complex and are divided into two or more domains (e.g., positive and negative symptoms in schizophrenia, or mood and neurovegetative symptoms in major depression). Many scales yield separate scores, or *subscales*, for each domain. Especially when these domains are seen as substantially independent, they may be referred to as dimensions (e.g., Axis I and Axis II in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV], multidimensional personality traits).

Categorical Versus Continuous Classification Some constructs are viewed as *categorical*, or classifying, while others are seen as *continuous*, or measuring. Categorical constructs describe the presence or absence of a given attribute (e.g., competency to stand trial) or the category best suited to a given individual among a finite set of options (e.g., assigning a diagnosis). Continuous measures provide a quantitative assessment along a continuum of intensity, frequency, or severity. In addition to symptom severity and functional status, multidimensional personality traits, cognitive status, social support, and many other attributes are generally measured categorically.

The distinction between categorical and continuous measures is by no means absolute. *Ordinal* classification, which uses a finite, ordered set of categories (e.g., unaffected, mild, moderate, severe), stands between the two. In addition, a cutpoint is frequently used with a continuous or ordinal scale to indicate a threshold for membership in a corresponding category. For instance, individuals with Mini-mental State Examination scores below 24 may be considered to have a dementia, or those with Hamilton Rating Scale for Depression (HAM-D) scores above 8 may be considered to have an episode of major depression.

Measurement Procedures Rating scales differ in measurement methods. Issues to be considered include format, raters, and sources of information.

Format Rating scales are available in a variety of formats. Some are simply checklists or guides to observation that help the clinician achieve a standardized rating. Others are self-administered questionnaires or tests. Still others are formal interviews that may be *fully structured* (i.e., specifying the exact wording of questions to be asked) or *partly structured* (i.e., providing only some precise wording, along with suggestions for additional questions or probes). Whether fully structured or not, instruments may be written so that all questions are always included, or they may have formal skip-out sections to limit administration time.

Individual items also vary in their format. Most commonly, scales use yes-no or multiple choice questions. Often, answers are graded on a *Likert scale*, an ordinal scale with three to seven points that measures severity, intensity, frequency, or other attributes. Likert scales are most often partially or fully *anchored*, assigning a meaning to each numeric level. The same anchors can apply to all items or the instrument may provide specific anchors for each. Occasionally, questionnaires include open-ended questions, especially at the beginning, which may be used to help establish rapport. In semistructured or unstructured interviews, this information also serves to guide the rest of the interview and aids in forming a clinical impression about the patient.

Raters Some instruments are designed to be administered by doctoral-level clinicians only, while others may be administered by individuals such as psychiatric nurses or social workers with more limited clinical experience. Still other instruments are designed primarily for use by lay raters with little or no experience with psychopathology. In general, more training is required to administer less-structured scales. In addition, some scales require extensive training, even for experienced clinicians, to master the appropriate procedures and achieve a good result. Virtually all scales perform better when raters are familiar with their format and specific content.

Source of Information Instruments also vary in the source of information used to make the ratings. Information may be obtained solely from patients, who generally know the most about their condition. In some instruments, some or all of the information may be obtained from a knowledgeable informant. When the construct involves limited insight (e.g., cognitive disorders or mania) or significant social undesirability (e.g., antisocial personality, substance abuse), other informants may be preferable. Informants may also be helpful when the subject has limited ability to recall or report symptoms (e.g., delirium, dementia, or any disorder in young children). Some rating scales also allow or require inclusion of information from medical records or patient observation.

ASSESSMENT OF RATING SCALES

In clinical research, rating scales are mandatory to ensure interpretable and potentially generalizable results and are selected on the basis of coverage of the relevant constructs, expense (based on the raters, purchase price, if any, and necessary training), length and administration time, comprehensibility to the intended audience, and quality of the ratings provided. In clinical practice, one considers these factors and also whether a scale would provide more or better information than would be obtained in ordinary clinical practice or contributes to the efficiency of obtaining that information. In either case, the assessment of quality is based on *psychometric*, or mind measuring, properties.

Psychometric Properties The two principal psychometric properties of a measure are reliability and validity. Although these words are used almost interchangeably in everyday speech, in the context of evaluating rating scales they are distinct. To be useful, scales should be *reliable*, or consistent and repeatable even if performed by different raters, at different times, or under different conditions, and they should be *valid*, or accurate in representing the true state of nature.

Relation Between Reliability and Validity Establishing a measure's reliability is generally considered primary, since it is difficult to reach valid judgments without first achieving consistency. However, problems with reliability can be overcome to an extent by combining information from several assessments. Unfortunately, improved reliability does not guarantee improved validity, and some efforts to improve reliability may actually limit validity. For example, a personality disorder instrument might focus on overt behaviors rather than inner thoughts and feelings to achieve higher reliability but at the cost of losing some of the most valid information about personality. Even with clinically trained raters, it is particularly difficult to achieve reliability on items requiring subjective clinical judgment (e.g., feelings evoked in the examiner): nonetheless, when used by experienced diagnosticians, such items may contribute substantially to valid diagnoses.

Reliability Reliability refers to the consistency or repeatability of ratings and is largely empirical. In the categorical context, it refers to whether agreement can be reached on the classification of each individual. In the continuous context, it refers to whether agreement can be reached on the assignment of a given score. It can also be seen as precision; that is, whether a measure yields a ballpark estimate or a finely graded score. An instrument is more likely to be reliable if the instructions and questions are clearly and simply worded and the format is easy to understand and score. There are three standard ways to assess reliability: internal consistency, interrater, and test-retest.

INTERNAL CONSISTENCY Internal consistency assesses agreement among the individual items in a measure. This provides information about reliability because each item is viewed as a single measurement of the underlying construct; thus, the coherence of the items suggests that each of them is measuring the same thing (and hence all of them are). Internal consistency is measured most often with coefficient alpha (also known as Cronbach's alpha), which ranges between 0 and 1; values of .75 or above are considered good. However, the internal consistency of a measure depends on the internal consistency of the construct that the measure purports to assess and is higher for unidimensional constructs than those with two or more relatively independent domains.

INTERRATER AND TEST-RETEST RELIABILITY Interrater (also called *interjudge*, or *joint*) reliability is a measure of agreement between two or more observers using the same information to evaluate the same subjects. Estimates may vary with assessment conditions; for instance, estimates of interrater reliability based on videotaped interviews tend to be higher than those based on interviews conducted by one of the raters. Interrater reliability tends to be higher than *test-retest reliability*, a measure of agreement between evaluations at two points in time, in which the information obtained may differ (e.g., be associated with differences in interviewer skill, interviewer mood, room conditions, or subject's attitude). In addition, test-retest evaluations measure reliability only to the extent that the subject's true condition remains stable in the time interval, which is problematic for many conditions but virtually impossible for rapidly fluctuating conditions like state anxiety. However, because the test-retest situation more closely reflects the clinical problems associated with serial evaluations by multiple clinicians, to the extent that concerns about interval change can be eliminated, it is generally a more useful indicator of reliability in practice.

Interrater reliability and test-retest reliability of continuous constructs are measured with the intraclass correlation coefficient (ICC), while those of categorical constructs are measured with the kappa coefficient (k). A weighted version of k is available to penalize large disagreements more than small ones (e.g., between schizophrenia and psychotic depression compared with schizophrenia and schizoaffective disorder). Both k and the ICC are measures of agreement corrected for the agreement expected by chance alone and both range from 0 to 1. As a rule of thumb, a k or ICC above .8 is considered excellent, those in the .7 to .8 range are considered good, and those in the .5 to .7 range are considered fair. However, the degree of reliability required varies with the clinical purpose; extremely reliable ratings are required before administering potentially dangerous treatments, while more modest reliability may suffice for estimating rates in a population.

ISSUES IN INTERPRETING RELIABILITY DATA When interpreting reliability data, remember that reliability estimates published in the literature may not generalize to other settings. Factors to consider are the nature of the sample, the training and experience of the raters, and the test conditions. Issues regarding the sample are especially critical. In particular, reliability tends to be higher in samples with high variability, in which it is easier to discriminate among individuals. Thus, for continuous measures, reliability tends to be higher when the sample includes individuals with a wide range of scores. For categorical measures, reliability tends to be higher when the prevalence of the attribute being measured is fairly high. Reliability estimates also depend on the fraction of difficult cases (e.g., individuals near a diagnostic threshold or those resistant to being interviewed), since large numbers of these tend to diminish observed reliability.

Validity Validity refers to conformity with truth or a gold standard that can stand for truth. In the categorical context, it refers to whether an instrument can make correct classifications. In the continuous context, it refers to accuracy, or whether the score assigned represents the true state of nature. While reliability is empirical, validity is partly theoretical; many constructs measured in psychiatry have no absolute truth. Even so, some measures yield more useful and meaningful data than others. Validity assessment is generally divided into face and content validity, criterion validity, and construct validity.

FACE AND CONTENT VALIDITY *Face validity* refers to whether the items appear to assess the construct in question. Although a rating scale may purport to measure a construct of interest, a review of the items may reveal that it embodies a very different conceptualization of the construct. For instance, an “insight” scale may define insight in either psychoanalytic or neurological terms. However, items with a transparent relation to the construct may be a disadvantage when measuring socially undesirable traits such as substance abuse or malingering. *Content validity* is similar to face validity but describes whether the measure provides good balanced coverage of the construct and is less focused on whether the items give the appearance of validity. Content validity is often assessed with formal procedures such as expert consensus or factor analysis.

CRITERION VALIDITY *Criterion validity* (sometimes called predictive or concurrent validity) refers to whether or not the measure agrees with a gold standard or criterion of accuracy. Suitable gold standards include the long form of an established instrument for a new shorter version, a clinician-rated measure for a self-report form, and blood or urine tests for measures of drug use. For diagnostic interviews, the generally accepted gold standard is the longitudinal, expert, all data (LEAD) standard, which incorporates expert clinical evaluation, longitudinal data, medical records, family history, and any other sources of information.

When comparing continuous measures with a gold standard, a correlation coefficient is the statistic most often reported. For categorical variables such as diagnoses (or continuous measures with a cutpoint), sensitivity and specificity are the statistics of choice. *Sensitivity* refers to the test's ability to identify true cases, or its true positive rate. *Specificity* is the test's accuracy in identifying noncases, or one minus the false positive rate. In general, the more sensitive a test, the less specific it is. If the threshold for diagnosis, for example, is lowered, more cases are detected but at the expense of some false positives; if the threshold is raised to decrease the number of false positives, true cases are inevitably missed. The optimal threshold depends on the consequences of false positives and false negatives.

CONSTRUCT VALIDITY When an adequate gold standard is not available—a frequent state of affairs in psychiatry—or whenever additional validity data are desired, construct validity must be assessed. To accomplish this, one can compare the measure with *external validators*, attributes that bear a well-characterized relation to the construct under study but are not measured directly by the instrument. External validators used to validate psychiatric diagnostic criteria and the diagnostic instruments that aim to operationalize them include course of illness, family history, and treatment response. For example, when compared with schizophrenia measures, mania measures are expected to identify more individuals with a remitting course, a family history of major mood disorders, and a good response to lithium.

Two special cases of assessing validity using external validators have particular relevance for clinical psychiatry. One is *discriminant validity*, which examines a

measure's ability to discriminate between populations that are expected to differ on the construct of interest. For example, does a sociopathy measure correctly separate individuals in jails from those living in the community? Although such discriminations are important in clinical practice, the true test of a measure is its ability to discriminate at the margins. A study of discriminant validity is more clearly relevant if it includes the types of cases encountered in clinical practice (e.g., psychotic depression versus schizoaffective disorder) rather than more-easily discriminated populations (e.g., psychotic depression versus normal). Another special case is *sensitivity to change*; the fact that a measure shows expected changes (e.g., an improvement with an efficacious treatment or a decrement with a progressive disease) can be a strong validator.

When assessing validity in areas with few established measures and no gold standard or criterion of accuracy can be established, the assessment of the validity of the measure is limited by the validity of the construct itself. Nonetheless, by triangulating between a better definition of the construct, better ways to measure it, and better exploration of how it operates in clinical practice and research, the field moves to greater validity over time.

SELECTION OF PSYCHIATRIC RATING SCALES

The rating scales used in psychiatric practice and research presented below are grouped by topic, beginning with such general issues as diagnosis, functioning, symptom severity, and side effects and then proceeding to specific diagnostic groups, organized according to the section of DSM-IV. The selection was made on the basis of coverage of major areas and common use in clinical research, current (or potential) use in clinical practice, or both. A brief discussion of measurement issues for each area is followed by a description of each instrument, its psychometric properties, and its potential uses. Whenever possible, a brief, clinically useful instrument is provided in each area. References for each measure, organized by topic, are listed in Table 7.8-1. These references include more-detailed information about each measure and its psychometric properties and may also provide either the measure itself or instructions for obtaining a copy.

Table 7.8-1 Key References for Measures Included

Functional Status, Impairment, and General Symptom Severity The broad area of functional status, impairment, and general symptom severity cuts across a variety of diagnoses and is thus useful for grading patients by functional status or overall severity without reference to specific symptomatology. The instruments presented here have a strong mental health focus and often include items on psychiatric symptomatology. Instruments focused on more-global functioning or domains such as mobility and self-care are not generally included here.

Global Assessment of Functioning (GAF) Scale and Social and Occupational Functioning Scale (SOFAS) The GAF, shown in Table 7.8-2 was developed in the early 1990s to rate Axis V of DSM-IV and provides a measure of overall functioning related to psychiatric symptoms. The GAF is extremely similar to the Global Assessment Scale (GAS) used for the same purpose in the third edition of DSM (DSM-III) and the revised third edition (DSM-III-R), from which it was derived. A related instrument is the SOFAS, proposed as a new axis in Appendix B of DSM-IV, which focuses only on functioning and not on symptoms and does not try to discriminate between functional changes related to psychiatric and nonpsychiatric causes (Table 7.8-3). Both scales are clinician rated on a 100-point scale based on all available information, with clear descriptions of each 10-point interval. Ratings are generally made for the past week, but longer intervals (e.g., highest during the past year) can be used. Instructions for rating the GAF and SOFAS are included in DSM-IV; clinician raters do not require additional training to use these scales. The GAS has received more extensive evaluation and shows fair-to-good reliability and good validity judged against clinician ratings of the degree of impairment. GAF or GAS ratings are often required for billing purposes. In addition, the scales have been used to track change with treatment in inpatient and outpatient practice and in multiple research studies. The major criticism of the GAS and GAF is that they tend to confound symptoms and functioning, so that individuals with significant symptomatology (e.g., fixed delusional system) score low even when their social and occupational functioning is relatively spared.

Table 7.8-2 Global Assessment of Functioning (GAF) Scale

Table 7.8-3 Social and Occupational Functioning Assessment Scale (SOFAS)

Global Assessment of Relational Functioning The GARF (Table 7.8-4) was developed in the late 1980s to provide a measure of the quality of functioning in relationships analogous to the measure of individual functioning provided by the GAF or SOFAS. It was subsequently included in Appendix B of DSM-IV as an additional axis for further consideration. It provides a global rating on a 100-point scale based on a review of three major areas: problem-solving, organization, and emotional climate. Anchors are provided for each quintile of each domain. The GARF is focused on the particular needs of family and couple therapists but can be

rated by any clinician. Ratings are generally based on the present, but alternate periods (e.g., the past year, or in the period following a major stressor) can be implemented. The GARF has not been extensively evaluated, but preliminary evidence suggests that clinician and even nonclinician raters can achieve good to excellent reliability with only minimal training. The validity of the GARF is supported by expected correlations with other measures of family and couple distress and functioning. The GARF shows promise for rating relational functioning but only time will tell whether it will prove useful in clinical or research practice.

Table 7.8-4 Global Assessment of Relational Functioning (GARF)

Behavior and Symptom Identification Scale The BASIS-32 (Table 7.8-5) was developed in the early 1990s to provide a broad but brief overview of psychiatric symptoms and functional status from the patient's point of view for use in assessing the outcome of psychiatric treatment. The instrument assesses a wide range of areas, including family and work relationships; ability to complete regular tasks at home, work, or school; and symptoms of anxiety, depression, psychosis, and substance abuse. Each item is rated on a five-point scale focused on the degree of difficulty during the preceding week. BASIS-32 can be completed as a paper and pencil test (requiring 5 to 20 minutes) or the questions can be read aloud with the patient selecting the best answer from a laminated card (requiring 15 to 20 minutes). It can be scored readily by hand. A computerized scoring system is also available. BASIS-32 generates an overall score and five subscales: relation to self and others, daily living and role functioning, depression and anxiety, impulsive and addictive behavior, and psychosis. Good reliability and validity have been demonstrated. Its simple administration, brevity, and broad coverage make it well suited to its original task, and it is frequently used at baseline, during, and after treatment to monitor progress. It can provide valid ratings across a wide range of psychiatric impairment but is not generally suitable for individuals with substantial cognitive impairment. It is also not suitable for children under age 14.

Table 7.8-5 BASIS-32

Symptom Checklist-90 Revised (SCL-90-R) and Brief Symptom Inventory (BSI) The SCL-90-R was developed in the mid-1970s from the older Hopkins Symptom Checklist, a multidimensional measure of the severity of psychopathology. (*SCL-90* refers to a very similar earlier version but is frequently used to denote the current version as well.) The BSI was developed in the early 1980s as a short form of the SCL-90-R. Both cover the following domains: depression, anxiety, phobia, psychoticism, paranoia, obsessive-compulsive, hostility, somatization, and interpersonal sensitivity. Even the longer SCL-90-R fits on two sides of a single sheet of paper and can be completed in 20 minutes or less. The SCL-90-R is a self-report measure with 90 items Likert-scaled from 0 to 4 on the basis of the distress caused over the past week. The BSI is very similar, but has 53 items. Hand scoring is relatively simple, and computerized scoring is available. Both yield *t* scores based on extensive normative data for each subscale, a Global Severity Index, a Positive Symptom Distress Index, and a Positive Symptom Total. Reliability is fair to good, depending on the subscale, and most subscales appear reasonably valid when assessed against more specific measures (e.g., the depression scale against a HAM-D). The principal use of the SCL-90-R and BSI has been in characterizing psychopathology in treatment and other studies, but they have sometimes been used in primary care clinics as a screening tool for psychopathology. However, caution may be warranted in this setting because their sensitivity may be limited in some areas. In addition, the subscales bear only a modest relation to the corresponding DSM-IV disorders.

Side Effects The instruments described below are used to detect and quantify side effects from psychiatric medications, specifically motor effects of antipsychotics. Because of the focus on motor symptoms, these scales include a brief, focused physical examination as well as questions posed to the patient.

Abnormal Involuntary Movement Scale (AIMS) The AIMS (Table 7.8-6) is a clinical examination and rating scale that was developed in the 1970s to measure dyskinetic symptoms in patients taking antipsychotic drugs. The AIMS has 12 items, each of which is rated on an item-specific five-point severity scale ranging from 0 to 4. Total scores are not generally reported. Instead, changes in global severity and individual areas can be monitored over time. Ten items cover the movements themselves, divided into sections rating global severity and those related to specific body regions; two items concern dental factors that can complicate the diagnosis of dyskinesia. In the presence of extended neuroleptic exposure and the absence of other conditions causing dyskinesia, mild dyskinetic movements in two areas or moderate movements in one area suggest a diagnosis of tardive dyskinesia. The AIMS was developed for clinician raters, but lay raters can be trained to use it. It can be completed in under 10 minutes. Excellent reliability has been demonstrated, especially for experienced raters, and the instrument appears valid. In many clinical settings, the AIMS is considered standard clinical practice for patients receiving long-term neuroleptic drugs and is useful in clinical practice and research, both for monitoring patients for the development of tardive dyskinesia and for tracking changes in tardive dyskinesia over time.

Table 7.8-6 Abnormal Involuntary Movement Scale (AIMS) Examination Procedure

Simpson-Angus Rating Scale for Extrapyramidal Side Effects The Simpson-Angus scale was developed to monitor the effects of antipsychotic drugs. It has 10 items, each of which is rated on an item-specific, five-point severity scale ranging from 0 to 4. Scores are reported as the mean on all 10 items, with 0.3 considered the upper limit of normal. It is strongly focused on parkinsonian symptoms, particularly rigidity, but includes one akathisia item. It is designed for clinician use but can be administered by trained lay raters and takes about 10 minutes to administer. Good reliability has been reported, and validity is supported by the correlation of scores with antipsychotic drug dose. The scale is useful in a wide variety of clinical settings to monitor parkinsonian adverse effects and the impact of interventions to treat these effects.

Psychiatric Diagnosis Instruments assessing psychiatric diagnosis are central to psychiatric research and may have utility in clinical practice as well. However, they tend to be rather long, especially for individuals reporting many symptoms, who may require many follow-up questions. When evaluating such instruments, one must be sure that they implement current diagnostic criteria and cover the diagnostic areas of interest. For instance, few cover personality disorders (with the exception in some cases of antisocial personality), and not all cover disorders that typically begin in childhood.

Structured Clinical Interview for DSM-IV The SCID was developed in the early 1990s to provide a standardized DSM-III-R Axis I diagnosis based on an efficient but thorough clinical evaluation. It has since been updated for DSM-IV. The semistructured diagnostic interview begins with a section on demographic information and clinical background. Then there are seven diagnostic modules focused on different diagnostic groups: mood, psychotic, substance abuse, anxiety, somatoform, eating, and adjustment disorders. Both required and optional probes are provided, and skip-outs are suggested when no further questioning is warranted. All available information, including that from hospital records, informants, and patient observation, should be used to rate the SCID. The SCID is designed to be administered by experienced clinicians and is generally not recommended for use by lay interviewers. In addition, formal training in the SCID is required, and training books and videos are available to facilitate this. In individuals without symptoms, the interview takes approximately 1 hour, but it may take up to 3 hours in individuals with extensive symptomatology. Although the primary focus is research with psychiatric patients, a nonpatient version (with no reference to a chief complaint) and a more clinical version (without as much detailed subtyping) are also available. Reliability data on the SCID suggest that it performs better on more-severe disorders (e.g., bipolar I disorder, alcohol dependence) than on milder ones (e.g., dysthymia). Validity data are limited, as the SCID is more often used as the gold standard to judge other instruments. It is considered the standard interview to verify diagnosis in clinical trials and is extensively used in other forms of psychiatric research. It can also be used to ensure a systematic evaluation in psychiatric patients; for instance, on admission to an inpatient unit or at intake into an outpatient clinic. It is also used in forensic practice to ensure a formal and reproducible examination.

Diagnostic Interview Schedule (DIS) and Composite International Diagnostic Instrument (CIDI) The DIS and CIDI are fully structured diagnostic interviews designed for lay administration. The DIS was developed in the 1980s for use in the Epidemiologic Catchment Area (ECA) study in the United States, which aimed to assess rates of current and lifetime psychiatric illness according to DSM-III in a large and diverse community sample, and has since been updated for DSM-III-R and DSM-IV criteria. In 19 diagnostic modules, it covers a broad range of Axis I conditions in adults, plus several childhood disorders and antisocial personality. The most recent version also includes more information about symptoms, impairment, and treatment. The CIDI was developed from the DIS for international use and covers both ICD and DSM criteria in 11 diagnostic modules; the CIDI does not cover antisocial personality or childhood disorders. The instruments are fairly similar, and both involve verbatim reading of questions with little or no rewording allowed; only specified probes may be used for follow-up. The DIS takes 90 minutes to 2 hours; the CIDI may be somewhat shorter. Both can be scored by computer, yielding diagnoses and symptom profiles. A computerized, self-administered version of the CIDI is also available. The instruments are designed for use by lay interviewers with extensive training in their use: formal training is recommended. Reliability appears to be good for both, at least for more-severe disorders. Validity appears problematic for the DIS; studies of agreement with clinician diagnoses have yielded inconsistent results, with marked discrepancies often observed for psychotic disorders. The validity of the CIDI is still being evaluated. Both instruments have been used extensively in psychiatric research, particularly in epidemiological settings and provide valuable data. However, some caution is warranted in interpreting these data, given concerns about the instruments' validity.

Primary Care Evaluation of Mental Disorders (PRIME-MD) The PRIME-MD was developed in the mid-1990s to provide an efficient screening and evaluation tool for common mental disorders seen in the primary care setting. The instrument has two parts: a 25-item patient questionnaire that screens for a range of symptoms and a structured interview designed to follow up on any symptoms identified in the patient questionnaire or through other means. The interview has five modules covering mood, anxiety, alcohol, somatoform, and eating disorders. The patient questionnaire is very brief and can be completed in under 5 minutes. If follow-up is required, a primary care practitioner can complete the structured questionnaire in about 10 minutes. Training in the use of the instrument or careful review of the instruction manual is recommended. A self-report version of the structured interview is in development. Reliability appears to be fair to good, better for more severe diagnoses. Validity judged against psychiatrist evaluations is quite good. The screening instrument has appropriately high sensitivity, and the follow-up interview provides reasonable specificity. The PRIME-MD appears to be useful for primary care settings and may also be useful in psychiatric practice when a quick screen is desired. However, its utility for the latter purpose is limited by its lack of coverage of the more-severe psychopathology not typically seen in primary care (e.g., psychotic symptoms, mania).

Psychotic Disorders A variety of instruments are used for patients with psychotic disorders. Those reported here are symptom severity measures. A developing consensus suggests that distinguishing positive and negative symptoms in schizophrenia is worthwhile, and more-recently developed instruments implement this distinction. Because patients with psychotic disorders often lack insight and are sometimes agitated, patient observation is required in addition to direct questioning. Thus, most instruments in this domain must be administered by psychiatrists or others with clinical training.

Brief Psychiatric Rating Scale (BPRS) The BPRS (Table 7.8-7) was developed in the late 1960s as a short scale for measuring the severity of psychiatric symptomatology. Developed primarily to assess change in psychotic inpatients, it covers a broad range of areas including thought disturbance, emotional withdrawal and retardation, anxiety and depression, and hostility and suspiciousness. Its 18 items are rated on a seven-point item-specific Likert scale from 0 to 6, with the total score ranging from 0 to 108 (in some scoring systems, the lowest level for each item is 1, and the range is 18 to 126). Because the ratings include observations as well as patient reports of symptoms, the BPRS can be used to rate patients with very severe impairment. It is intended for use by experienced clinicians and can be administered in 30 minutes or less, including patient interview and observation. Reliability of the BPRS is good to excellent when raters are experienced but is more difficult to achieve without substantial training; anchored versions and a semistructured interview have been developed to increase reliability. Validity is also good as measured by correlations with other measures of symptom severity, especially those assessing schizophrenia symptomatology. The principal use of the BPRS is as an outcome measure in treatment studies of schizophrenia, and it functions well as a measure of change in this context. However, it has been largely supplanted in more-recent clinical trials by the newer measures described below. In addition, given its focus on psychosis and associated symptoms, it is only suitable for patients with fairly significant impairment. Its use in clinical practice is less well supported, in part because considerable training is required to achieve the necessary reliability.

Table 7.8-7 Brief Psychiatric Rating Scale

Positive and Negative Syndrome Scale (PANSS) The PANSS was developed in the late 1980s to remedy perceived deficits in the BPRS in the assessment of positive and negative schizophrenia and other psychotic disorders by adding additional items and providing careful anchors for each. The PANSS includes 30 items on three subscales: 7 items covering positive symptoms (e.g., hallucinations and delusions), 7 covering negative symptoms (e.g., blunted affect), and 16 covering general psychopathology (e.g., guilt, uncooperativeness). Each item is scored on a seven-point item-specific Likert scale ranging from 1 to 7; thus the positive and negative subscales each range from 7 to 49, and the general psychopathology scale from 16 to 112. The PANSS requires a clinician rater because considerable probing and clinical judgment are required. A semistructured interview guide is available. The ratings can be completed in 30 to 40 minutes. Reliability for each scale

is fairly high, with excellent internal consistency and interrater reliability. Validity also appears good based on correlation with other symptom severity measures and factor analytic validation of the subscales. The PANSS has become the standard tool for assessing clinical outcome in treatment studies of schizophrenia and other psychotic disorders and is sensitive to change with treatment. Its high reliability and good coverage of both positive and negative symptoms make it excellent for this purpose. It may also be useful for tracking severity in clinical practice, and its clear anchors make it easy to use in this setting.

Scale for the Assessment of Positive Symptoms (SAPS) and Scale for the Assessment of Negative Symptoms (SANS) The SAPS and SANS were designed to provide a detailed assessment of positive and negative symptoms of schizophrenia and may be used separately or in tandem. The domains assessed include hallucinations, delusions, bizarre behavior, and thought disorder for the SAPS and affective flattening, poverty of speech, apathy, anhedonia, and inattentiveness for the SANS. Each instrument consists of 30 fully anchored items each scored 0 to 5; thus the total score ranges from 0 to 150 for each. Each must be rated by an experienced clinician and requires approximately 30 minutes to complete. Good-to-excellent interrater reliability exists if trained interviewers are used, and each scale has high internal consistency as well. Validity is supported by correlation with other symptom severity instruments. The SAPS and SANS are principally used to monitor treatment effects in clinical research and have also been used to help characterize positive and negative symptoms in studies of schizophrenia phenomenology. The comprehensive characterization of symptomatology provided might also be useful in clinical practice, but there is little experience in this area at the present time.

Mood Disorders The domain of mood disorders includes both depressive and bipolar disorders. The issues for mania are similar to those for psychotic disorders, in that limited insight and agitation may hinder accurate symptom reporting, so clinician ratings including observational data are generally required. Rating depression, on the other hand, depends substantially on subjective assessment of mood states, so interviews and self-report instruments are both common. Because depression is common in the general population and involves significant morbidity and even mortality, screening instruments, especially those using a self-report format, are potentially quite useful in primary care and community settings.

Hamilton Rating Scale for Depression (HAM-D) The HAM-D was developed in the early 1960s to monitor the severity of major depression, with a focus on somatic symptomatology. The version in most common use has 17 items, although versions with different numbers of items, including the 24-item version (Table 7.8-8), have been used in many studies as well. Most versions do not include some of the symptoms used to diagnose depression in DSM-III and its successors, most notably increased sleep and increased appetite. Items on the HAM-D are scored 0 to 2 or 0 to 4, with total scores on the 17-item version ranging from 0 to 50: scores of 7 or less may be considered normal; 8 to 13, mild; 14 to 18, moderate; 19 to 22, severe; and 23 and above very severe. The HAM-D was designed for clinician raters but has been used by trained lay administrators as well. Ratings are completed by the examiner on the basis of patient interview and observations. A structured interview guide has been developed to improve reliability. The ratings can be completed in 15 to 20 minutes. Reliability is good to excellent, including internal consistency and interrater assessments. Validity appears good based on correlation with other depression symptom measures. The HAM-D has been used extensively to evaluate change in response to pharmacological and other interventions. It is more problematic in elderly and medically ill persons, in whom somatic symptoms may not indicate major depression.

Table 7.8-8 Hamilton Rating Scale for Depression

Beck Depression Inventory (BDI) The BDI was developed in the early 1960s to rate depression severity, with a focus on behavioral and cognitive dimensions of depression. The current version, the Beck Depression Inventory–II (Beck-II), has added more coverage of somatic symptoms to be compatible with DSM-IV and covers the most recent 2 weeks. Earlier versions focus on the past week or even shorter intervals, which may be preferable for monitoring treatment response. The BDI includes 21 self-report items, each of which has four statements describing increasing levels of severity; the total score ranges from 0 to 84. Scores of 0 to 9 are considered minimal; 10 to 16, mild; 17 to 29, moderate; and 30 to 63, severe. The scale can be completed in 5 to 10 minutes. Internal consistency has been high in numerous studies. Test-retest reliability is not consistently high, but this may reflect changes in underlying symptoms. Validity is supported by correlation with other depression measures. The principal use of the BDI is as an outcome measure in clinical trials of interventions for major depression, including psychotherapeutic interventions. Because it is a self-report instrument, it is sometimes used to screen for major depression, for instance in medical outpatients. Various cutoffs have been suggested for a diagnosis of major depression, but even a cutoff of 9 has only fair sensitivity, at a cost of considerable nonspecificity, suggesting that the instrument has limited use for screening. The instrument's strength lies in measuring the depth of depression; it is not suitable for making a diagnosis.

Zung Self-Rating Depression Scale The Zung scale was developed in the 1960s to provide a self-report measure of major depression with broad coverage of depression symptomatology. It has 20 items, each of which is scored from 1 to 4, based on the fraction of time in which it occurs. Half of the items are scored positively, and half negatively; positive items must be reversed to obtain a total severity score, which is then converted by formula to a scaled score. Scaled scores under 50 are considered normal; 50 to 59, minimal depression; 60 to 69, moderate depression; and over 70, severe depression. Most individuals can complete the Zung scale in 5 to 10 minutes. The Zung scale has been in use for many years but has not been extensively evaluated. However, reliability is good based on split-half and internal consistency studies. Validity also appears good based on correlations with other depression measures and the ability to discriminate between depressed and nondepressed outpatients. The Zung scale has been used to follow depressed patients in treatment studies; however, there is less variation in Zung scale scores than some other measures, which limits its utility as a change measure. It has also been used to screen for depression in medical outpatients or community interventions, including National Depression Screening Day, in which a threshold score of 50 was used to identify potential cases of depression requiring follow-up by a clinician.

Young Mania Rating Scale (YMRS) The YMRS is a checklist developed in the late 1970s to provide a brief but thorough evaluation of the severity of mania that could be used to monitor treatment response or detect relapse. It consists of a checklist of 11 items rated either 0 to 4 (seven items) or 0 to 8 (four items). Each item has five item-specific anchors. The total score ranges from 0 to 60. Ratings include clinical observation, so it must be rated by a clinician, but reliable ratings have been obtained by nurses on inpatient units and beginning psychiatric residents. Reliability is good, based on interrater reliability and internal consistency studies. Validity also appears good, based on correlation with other mania measures. The YMRS is useful for evaluating response to treatment in clinical research, and it is sensitive to change in this setting. It might also be used to assess treatment response or monitor for relapse in treated or untreated patients, although extensive experience with this use has not been reported.

Anxiety Disorders The anxiety disorders addressed by the measures below include panic disorder, generalized anxiety disorder, and obsessive-compulsive disorder. When examining anxiety measures, one must be aware that their definitions have changed significantly over time. Both panic and obsessive-compulsive disorder are relatively recently recognized, and the conceptualization of generalized anxiety disorder has shifted over time. Thus, older measures have somewhat less relevance for diagnostic purposes, although they may identify symptoms causing considerable distress. Whether reported during an interview or on a self-report rating scale, virtually all measures in this domain, like the measures of depression discussed above, depend on subjective descriptions of inner states.

Hamilton Rating Scale for Anxiety The HAM-A (Table 7.8-9) was developed in the late 1950s to assess anxiety symptoms, both somatic and cognitive. Because conceptualization of anxiety has changed considerably, the HAM-A provides limited coverage of “worry” required for DSM-IV diagnosis of generalized anxiety disorder and does not include the episodic anxiety found in panic disorder. There are 14 items, each of which is rated 0 to 4 on an unanchored severity scale, with the total score ranging from 0 to 56. A score of 14 has been suggested as the threshold for clinically significant anxiety, but scores of 5 or less are typical in individuals in the community. The scale is designed to be administered by a clinician, and formal training or the use of a structured interview guide is required to achieve high reliability. A computer-administered version is also available. Reliability is fairly good, based on internal consistency, interrater, and test-retest studies. However, given the lack of specific anchors, reliability should not be assumed to be high across different users in the absence of formal training. Validity appears good, based on correlation

with other anxiety scales, but is limited by the coverage of domains critical to the modern understanding of anxiety disorders. Even so, the HAM-A has been used extensively to monitor treatment response in studies of generalized anxiety disorder and may also be useful for this purpose in clinical settings.

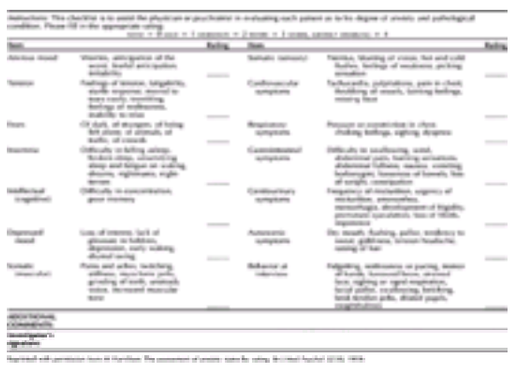

Table 7.8-9 Hamilton Rating Scale for Anxiety

Panic Disorder Severity Scale (PDSS) The PDSS is a recently developed, brief rating scale aimed at measuring the severity of panic disorder. It was based on the Yale-Brown Obsessive Compulsive Scale (YBOCS) and has seven items, each of which is rated on an item-specific five-point Likert scale. The seven items address frequency of attacks, distress associated with attacks, anticipatory anxiety, phobic avoidance, and impairment. The items are scored 0 to 4, and the total score ranges from 0 to 28. The instrument was designed for use by clinicians, but a patient-scored computerized version is in development. Reliability is excellent, based on interrater studies, but, in keeping with the small number of items and multiple dimensions, internal consistency is limited. Validity is supported by correlations with other anxiety measures, both at the total and item level, and lack of correlation with the HAM-D. Because the PDSS has been available for a fairly short period of time, there is limited experience with its use. However, it appears to be sensitive to change with treatment and is thus likely to prove useful as a change measure in clinical trials or other outcome studies for panic disorder, as well as for monitoring panic disorder in clinical practice.

Yale-Brown Obsessive Compulsive Scale (YBOCS) The YBOCS (Table 7.8-10) was developed in the late 1980s to measure the severity of symptoms in obsessive-compulsive disorder. It has 10 items rated on the basis of a semistructured interview. The first five items concern obsessions: the amount of time they consume, the degree to which they interfere with normal functioning, the distress they cause, the patient's attempts to resist them, and the patient's ability to control them. The remaining five items ask parallel questions about compulsions. Each item has a set of item-specific anchors scored 0 to 4, so total scores for obsessions and compulsions each range from 0 to 20, and overall total score ranges from 0 to 40. Typical scores for patients with obsessive-compulsive disorder are in the 16 to 30 range, and a threshold of 16 is typically used for inclusion in drug trials. The semistructured interview and ratings can be completed in 15 minutes or less. A self-administered version has recently been developed and can be completed in 10 to 15 minutes. Computerized and telephone use also provide acceptable ratings. Prior to the first use of the YBOCS, an associated 64-item checklist is administered to provide a more detailed assessment of the specific content of the patient's obsessions and delusions. Reliability studies of the YBOCS show good internal consistency, interrater reliability, and test-retest reliability over a 1-week interval. Validity appears good, although data are fairly limited in this developing field. The YBOCS has become the standard instrument for assessing obsessive-compulsive disorder severity and is used in virtually every drug trial. It may also be used clinically to monitor treatment response.

Table 7.8-10 Yale-Brown Obsessive-Compulsive Scale

Substance Use Disorders Substance use disorders include both abuse and dependence on both alcohol and drugs. These disorders, particularly those involving alcohol, are common and debilitating in the general population, so screening instruments are particularly helpful. Because these behaviors are socially undesirable, underreporting of symptoms is a significant problem. Validation against drug tests or other measures is of great value, particularly when working with patients who have known substance abuse.

CAGE The CAGE was developed in the mid-1970s to serve as a very brief screen for significant alcohol problems in a variety of settings, which could then be followed up by clinical inquiry. CAGE is an acronym for the four questions that make up the instrument: (1) Have you ever felt you should *c*ut down on your drinking?; (2) Have people *a*nnoyed you by criticizing your drinking?; (3) Have you ever felt bad or *g*uilty about your drinking?; (4) Have you ever had a drink first thing in the morning to steady your nerves or to get rid of a hangover (*e*ye-opener)? Each "yes" answer is scored as 1, and these are summed to generate a total score. Scores of 1 or more warrant follow-up, and scores of 2 or more strongly suggest significant alcohol problems. The instrument can be administered in a minute or less either orally or on paper. Reliability has not been formally assessed. Validity has been assessed against a clinical diagnosis of alcohol abuse or dependence, and these four questions perform surprisingly well. Using a threshold score of 1, the CAGE achieves excellent sensitivity and fair to good specificity. A threshold of 2 provides still greater specificity, but at the cost of a fall in sensitivity. The CAGE performs well as an extremely brief screening instrument for use in primary care or in psychiatric practice focused on problems unrelated to alcohol. However, it has limited ability to pick up early indicators of problem drinking that might be the focus of preventive efforts.

Alcohol Use Disorders Identification Test (AUDIT) The AUDIT (Table 7.8-11) was developed by the World Health Organization in the late 1980s as a brief screening instrument designed for the early detection of hazardous (i.e., involving the risk of harm) and harmful (i.e., involving the presence of harm) alcohol use in a variety of settings. It focuses on both the past year and current drinking. It includes a 10-item core screening instrument covering alcohol consumption, drinking behaviors, and alcohol-related problems. Each item is rated using item-specific anchors scored 0 to 4 and summed for a total score of 0 to 40. The AUDIT can be administered and scored in less than 5 minutes and does not require professional training. The AUDIT also offers a Clinical Screening Procedure involving a physical examination and blood tests, which adds no more than 5 to 10 minutes to a routine medical examination. Reliability of the AUDIT appears good, based on internal consistency data. Validity judged against a clinical diagnosis of alcoholism is also good; using a threshold score of 8 is quite sensitive but somewhat nonspecific. A score of 10 is more specific but at a cost in specificity. Validity also appears good, based on correlation with other alcohol self-report measures and risk factors for alcoholism. The AUDIT provides an excellent brief screen for alcohol problems and is particularly good for detecting problem drinking at a fairly mild stage. However, its focus on early detection of hazardous and harmful drinking makes it less suited for use as a diagnostic instrument.

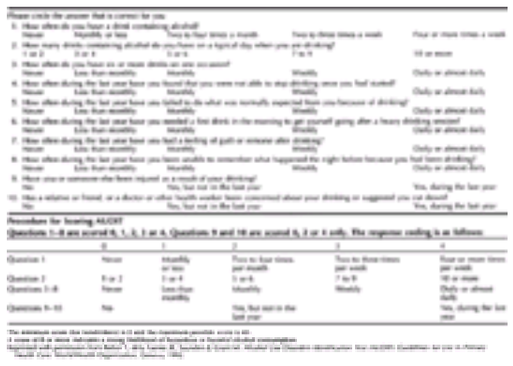

Table 7.8-11 AUDIT

Drug Abuse Screening Test (DAST) The DAST was developed in the early 1980s to serve as a screening and assessment instrument for drug abuse. The DAST is an adaptation of the Michigan Alcohol Screening Test (MAST), used to screen for alcoholism. It focuses on lifetime drug use, so it is not designed to measure changes over time. The current version of the DAST has 20 items, all of which are answered “yes” or “no,” and can be given orally or as a paper and pencil questionnaire. An earlier version had 28 items, so the 20-item version is sometimes called the Brief DAST. The positive items can be summed to form a 20-point scale. The DAST can be administered and scored in less than 10 minutes. Reliability is very good, based on internal consistency. Validity based on ability to detect drug abuse disorder also appears high, with excellent sensitivity and fairly good specificity using a threshold score of 5. The DAST is useful as a screening device for drug abuse problems in patients with other mental disorders, particularly alcohol abuse. It also provides an overview of problem severity that may be useful in guiding treatment choices.

Addiction Severity Index (ASI) The ASI was developed in the early 1980s to serve as a quantitative measure of symptoms and functional impairment caused by alcohol or drug disorders. It covers demographics, alcohol use, drug use, psychiatric status, medical status, employment, legal status, and family and social issues. Frequency, duration, and severity are assessed. It has 142 items in varying formats including yes-no, multiple choice, and scaled items. They include both subjective and objective items reported by the patient and observations made by the interviewer. In each area, the ASI yields the rater's global assessment of severity along with a computer score on a 0 to 1 scale. The 142-item version includes information on the past 30 days and lifetime status, but a shorter version is available for use at follow-up. The instrument is designed for clinician administration but has been used successfully by trained lay raters. Training is recommended, and both manuals and formal training programs are available. A computerized version is also available. The standard ASI requires 45 to 75 minutes to complete, but the follow-up version can be completed in 15 to 20 minutes. Very good to excellent reliability has been demonstrated for the overall composite score, with somewhat lower reliability for severity ratings in each area. Validity has also been demonstrated, based on correlation with other measures and discrimination of patient and nonpatient populations. Normative data are available for a range of populations of alcohol and drug abusers, including alcohol clinic patients, drug abusers, homeless persons, and prisoners. The principal use of the ASI is as an aid to treatment planning and the assessment of treatment outcome in clinical and law enforcement settings. It is relatively time consuming to administer but performs well for this purpose. It is also used in clinical research as a sensitive indicator of baseline severity and change over time, which allows comparison between clinical research and clinical practice.

Eating Disorders Eating disorders include anorexia nervosa, bulimia nervosa, and binge-eating disorder, which is included in Appendix B of DSM-IV and is gaining acceptance among eating-disorders clinicians and researchers. A wide variety of instruments, particularly self-report scales, are available. Because of the secrecy that may surround dieting, bingeing, purging, and other symptoms, validation against other indicators (e.g., body weight for anorexia, dental examination for bulimia) may be very helpful. Such validation is particularly critical for patients with anorexia, who may lack insight into their difficulties.

Eating Disorders Inventory (EDI) The EDI was developed in the early 1980s to provide a multidimensional self-report assessment of eating disorder symptomatology and related psychological attributes. The current version of the EDI, the EDI-2, has 91 items in 11 subscales: Drive for Thinness, Bulimia, Body Dissatisfaction, Ineffectiveness, Perfectionism, Interpersonal Distrust, Interoceptive Awareness, Maturity Fears, Asceticism, Impulsive Regulation, and Social Insecurity. Each item is rated on a six-point frequency-based Likert scale. The two ratings at the symptomatic end of the scale are scored 2; the middle two, scored 1; and the last two, scored 0; then the items in each subscale are added to generate a subscale score. Instrumentwide total scores may also be obtained but are not considered meaningful. Norms for each subscale are available for a variety of eating-disordered and nonclinical populations. The EDI can be completed in less than 20 minutes. Easy-reading and childhood versions are available. A computerized version is also available. Reliability data indicate very good internal consistency and test-retest reliability for virtually all EDI subscales. Validity of the EDI subscales is supported by correlation with related eating disorder measures and discrimination between patient and nonpatient samples. The EDI subscales also correlate moderately with ratings of these domains by trained clinicians. The EDI has several uses that apply in both clinical and research practice. Its principal use is in providing a range of data that may help in treatment planning. For instance, body dissatisfaction is an important predictor of prognosis and treatment response in bulimia nervosa. Some of the subscales, particularly the Bulimia scale, have also been shown to be sensitive to change with treatment and thus may be used to monitor patients over time. The EDI has also been used for screening purposes in primary care or other settings to identify individuals at high risk for eating disorders. The Drive for Thinness, Bulimia, and Body Dissatisfaction scales are probably the most useful in this regard. For instance, scores of 14 or above on the Drive for Thinness scale suggest an increased risk for anorexia nervosa and warrant further evaluation.

Bulimia Test-Revised (BULIT-R) The BULIT-R was developed in the mid-1980s to provide a categorical and continuous assessment of bulimia nervosa. The current version, while designed for DSM-III-R criteria, has been validated for DSM-IV as well. The BULIT-R has 36 self-report items, each scored on an item-specific five-point Likert scale. Of these items, 8 provide descriptive information, and the remaining 28 are summed to provide the total score, which ranges from 28 to 140. Young women with bulimia nervosa typically score above 110, while young women without disordered eating typically score below 60. The instrument can be completed in about 10 minutes. The BULIT-R shows high reliability, based on studies of internal consistency and test-retest reliability in multiple studies. Validity is supported by high correlations with other bulimia assessments. The recommended cutoff of 104 suggested to identify probable cases of bulimia shows high sensitivity and specificity for a clinical diagnosis of bulimia nervosa. Using cutoffs between 98 and 104, the BULIT-R has been used successfully to screen for bulimia nervosa. As with any screening procedure, follow-up by clinical examination is indicated for individuals scoring positive; clinical follow-up is particularly critical because the BULIT-R does not distinguish clearly between different types of eating disorders. The BULIT-R may also be useful to track symptoms over time or in response to treatment, in both clinical and research practice, although more detailed measures of the frequency and severity of bingeing and purging may be preferable in research settings.

Cognitive Disorders Dementia is becoming an increasing focus of psychiatric practice. A wide variety of measures are available. Most involve cognitive testing and provide objective, quantifiable data. However, scores vary by educational level in subjects without dementia, so these instruments tend to be most useful when the patient's baseline score is known. Other measures focus on functional status, which can be assessed on the basis of a comparison with a description of the subject's baseline function; these types of measures generally require a knowledgeable informant and may thus be more cumbersome to administer, but they tend to be less subject to educational biases.

Mini-Mental State Examination The MMSE (Table 7.8-12) is a 30-point cognitive test developed in the mid-1970s to provide a bedside assessment of a broad array of cognitive functions including orientation, attention, memory, construction, and language. It can be administered in under 10 minutes by a busy doctor or a technician and scored rapidly by hand. The MMSE has been extensively studied and shows excellent reliability. Validity appears good, based on correlations with a wide variety of more comprehensive measures of mental functioning and clinicopathological correlations. One common use of the MMSE is in screening for dementia, in both office practice and epidemiological or clinical research. For this purpose, a cutoff of 24 for identifying cases of dementia has been suggested, but it is probably more accurate to use age- and education-adjusted norms to interpret the results. For patients with extensive education, who may score 30 out of 30 despite clear evidence of functional decline, a more difficult cognitive test, full neuropsychological battery, or clinical interview may be required to detect dementia. The other principal use of the MMSE is in following the progression of dementia over time. As a rule of thumb, mild dementia ranges from a score of 20 to 24, moderate from 11 to 19, and severe from 0 to 10. However, these figures do not take the educational differences noted above into account. In addition, the MMSE does not do as well tracking progression of dementia in the lower ranges, as many patients become untestable.

Table 7.8-12 Mini-Mental State Examination

Blessed Information Memory Concentration Test (IMC) The IMC, sometimes called the Blessed IMC after its developer, was developed in the late 1960s for studies of the relationship between dementia severity and neuropathological changes. The original version of the scale, developed in Britain, had 29 items, the current American version has 26. Areas assessed include information (date, time, place, name, age, remote personal information, dates of the world wars, name of governmental leaders), memory (a name and address for 5-minute recall), and concentration (counting forward and backward from 1 to 20). A six-item version, sometimes called the short form of the Blessed or the Orientation Memory Concentration (OMC) test, is also available: it asks only the time of day, month, year, the 5-minute recall of the address, and counting backward from 20 and is highly predictive of total IMC score. IMC scores on individual items are weighted; on the 26-item IMC they range from 0 (no errors) to 33 and on the six-item version, scores range from 0 to 28. The IMC can be administered in person or over the phone by a trained clinician or lay rater. Based on internal consistency and test-retest studies in demented subjects over a 1- to 6-week interval, reliability of the 26-item and 6-item IMC both appear very good. Validity also appears good, based on clinicopathological correlations and correlations with other dementia severity measures. Studies of changes with time in patients with Alzheimer's disease show an average annual increase of 3 to 4 points on the 26-item version and 2.5 points on the 6-item version. The principal use of the IMC is assessing dementia severity over time, either through the natural course of the disease or in response to treatment interventions. The IMC is also sometimes used as a screening instrument in clinical practice or community research studies. A cutoff score of 10 has been recommended for the 6-item version, but no standard cutoff is recommended for the 26-item version. In any case, very limited population data exist, and norms are not available by age or education, which are very likely to affect test results. Because of its focus on memory items, the IMC may perform better as a severity or screening measure in patients with Alzheimer's disease than in those with other dementing illnesses.

Global Deterioration Scale (GDS), Brief Cognitive Rating Scale (BCRS), and Functional Assessment and Staging Tool (FAST) The GDS, BCRS, and FAST are a group of measures designed to provide ordinal staging of cognitive and functional status in patients with dementia, particularly those with Alzheimer's disease. The three instruments use a consistent seven-point scale. The GDS is a simple rating scale that describes seven stages from normal aging to severe dementia: (1) normal, (2) subjective complaints only, (3) subtle deficits with little or no functional decline except in very demanding tasks (e.g., managerial tasks at work, or preparing an elaborate social event like a holiday meal), (4) definitive deficits that interfere with complex activities of daily living (ADLs) (e.g., balancing a checkbook), (5) deficits that interfere with independent living in the community, (6) deficits that interfere with basic ADLs (e.g., dressing, toileting), and (7) profound deficits leading to the need for continuous assistance. The BCRS focuses on cognitive issues and describes the same seven levels in five different domains referred to as axes: (I) concentration, (II) recent memory, (III) remote memory, (IV) orientation, and (V) self-care. The FAST focuses on functional status, again in seven stages, but it adds substages within stages 6 and 7. All three scales should be completed by an individual with clinical experience (physician, psychologist, nurse, or trained technician) after a review of all available information from the patient, informants, and medical records. The FAST can generally be completed in as little as 10 to 15 minutes, but the GDS and especially the BCRS may require 30 to 45 minutes. Reliability of these scales is excellent, based on interrater and test-retest studies. Validity of all three measures is supported by correlations with other cognitive and functional status scales. GDS and BCRS stages have also been validated against neuropathological data. FAST stages correspond closely to typical progression in Alzheimer's disease. The GDS, BCRS, and FAST are useful in staging dementia, especially Alzheimer's disease, which is much more likely to follow the described ordinal stages closely. Such staging may be used to provide a concise description for patients referred to other clinicians or settings or to track changes over time or in response to treatment. The FAST is especially useful in staging severe dementia and has been used extensively to assess the need for services. The GDS and BCRS are both sensitive to change, with average declines of approximately 0.5 GDS or BCRS subscale points per year in Alzheimer's disease patients.

Personality Disorders and Personality Traits Personality may be conceptualized categorically as personality disorders or dimensionally as personality traits, which may be viewed as normal or pathological. The focus here is on personality disorders, and the traits are generally viewed as their milder forms. DSM-IV defines 10 personality disorders in three clusters, and an additional two disorders (passive-aggressive and depressive personality) are proposed in Appendix B for further study. Patients tend not to fall neatly into DSM-IV personality categories; instead, most patients who meet criteria for one personality disorder also meet criteria for one or more others, particularly within the same cluster. This and other limitations in the validity of the constructs themselves makes it difficult to achieve validity in personality measures. Personality measures include both interviews and self-report instruments. Self-report measures are appealing in that they require less time and may appear less threatening to the patient. However, they tend to overdiagnose personality disorders. Because many of the symptoms suggesting personality problems are socially undesirable, and because patients' insight tends to be limited, clinician-administered instruments, which allow for probing and patient observation, may provide more accurate data.

Structured Clinical Interview for DSM-IV Axis-II Personality Disorders (SCID-II) The SCID-II is the counterpart of the SCID for making DSM diagnoses of personality disorders. The initial version was developed for DSM-III-R in the mid-1990s, and the current version makes diagnoses according to DSM-IV. The SCID-II is organized by disorder and includes all 10 DSM-IV personality disorders plus the two proposed in Appendix B. A 119-item self-report screening questionnaire is generally given first to eliminate sections not needing further exploration: each of the items corresponds to a specific criterion for a DSM-IV personality disorder. The SCID-II proper includes one or two yes-no items for each criterion, with each affirmative answer to be followed-up by examples from the person's life. Based on these answers each criterion is scored 1 for false, 2 for subthreshold, and 3 for present, allowing criteria scores to be summed for a dimensional measure of each disorder or combined following the DSM-IV diagnostic rules for a categorical approach. The screening questionnaire can be completed by the patient in about 20 minutes; the interview generally requires about an hour. The SCID-II must be administered by doctoral-level clinicians, and training in the SCID-II is also required. A computerized administration and scoring program is available. Reliability is good for the presence or absence of any disorder but only fair for specific personality disorders; the reliability of dimensional assessment is somewhat better. Validity is somewhat harder to determine, as agreement with clinician assessment tends to be modest, but given its comprehensiveness and strict adherence to DSM-IV criteria, the SCID-II may actually be more valid. The SCID-II is most useful to provide a standardized, comprehensive assessment of personality disorders, whether in research, forensic, or clinical settings.

Personality Disorder Questionnaire (PDQ) The PDQ was developed in the late 1980s as a self-report questionnaire designed to provide categorical and dimensional assessment of DSM-III-R personality disorders and was subsequently revised for DSM-IV. An alternate version includes the two disorders in Appendix B as well. Another alternate version allows for ratings within the last few weeks and is designed to serve as a change measure. The current PDQ, the PDQ-IV, includes 85 yes-no items, designed primarily to assess the diagnostic criteria for DSM-IV personality disorders. Within the 85 items are embedded two validity scales to identify underreporting, lying, or inattention. There is also a brief clinician-administered Clinical Significance Scale to address the impact of any personality disorder identified by the self-report PDQ. The PDQ can provide categorical diagnoses with a scaled score for each or an overall index of personality disturbance based on the sum of all the diagnostic criteria. Overall scores range from 0 to 79; patients with personality disorders generally score above 30, psychotherapy outpatients without such disorders tend to score in the 20-to-30 range, and normal controls tend to score below 20. The PDQ can be completed in under 30 minutes. Computerized administration and scoring are available. Reliability is fair to good for dimensional assessment and quite variable for categorical assessment, with good reliability for obsessive-compulsive and antisocial personality, and inadequate reliability for many disorders. Validity judged against semistructured clinician-administered interviews is also variable. The PDQ, like other self-report instruments, tends to overdiagnose personality disorders, with many false positives and few false negatives. Its brevity, excellent sensitivity, and poor specificity make it most useful as a screening device, with a follow-up semistructured interview for patients screening positive.

Childhood Disorders A wide variety of instruments is available to assess mental disorders in children. Despite a rich array of instruments, the evaluation of children remains difficult for several reasons. First, the child psychiatric nosology is at an earlier stage of development, and construct validity is often problematic. Multiple changes in diagnostic criteria from DSM-III to DSM-III-R to DSM-IV complicate the choice of measures. Second, because children change markedly with age, it is virtually impossible to design a measure that covers children of all ages. Finally, because children, particularly young children, have limited ability to report their symptoms, other informants are necessary. This often creates problems because child, parent, and teacher reports of symptoms frequently disagree, and the optimal

way to combine information is unclear.

Child Behavior Checklist (CBCL) The CBCL is a family of self-rated instruments that survey a broad range of difficulties encountered in children from preschool age through adolescence. One version of the CBCL, designed for completion by parents of children aged 4 to 18, is shown in Table 7.8-13. Another version is available for parents of children ages 2 to 3. The Youth Self Report is completed by children ages 11 to 18, and the Teacher Report Form is completed by teachers regarding school age children. The scale includes not only problem behaviors, but academic and social strengths as well. Each version includes approximately 100 items scored on a 3-point Likert scale. Scoring can be done by hand or computer, and normative data are available for each of the three subscales: problem behaviors, academic functioning, and adaptive behaviors. A computerized version is also available. The CBCL does not generate diagnoses, but instead suggests cutoff scores for problems in the "clinical range." Parent, teacher, and child versions each show high reliability on the problem subscale, but the three informants frequently do not agree with one another. The CBCL may be useful in clinical settings as an adjunct to clinical evaluation: they provide a good overall view of symptomatology and may also be used to track change over time. They are used frequently for similar purposes in research involving children and thus can be compared with clinical experience. The instrument does not, however, provide diagnostic information, and its length limits its efficiency for tracking purposes.

Table 7.8-13 Child Behavior Checklist for Ages 4–18

Diagnostic Interview Schedule for Children (DISC) The DISC was originally developed in the early 1990s as a fully structured diagnostic interview for making DSM-III diagnoses in children. It has since been revised for DSM-III-R and DSM-IV. The current DISC, the DISC-IV, covers a broad range of DSM-IV diagnoses, both current and lifetime. It has nearly 3000 questions but is structured with a series of stem questions that serve as gateways to each diagnostic area, with the remainder of each section skipped if the subject answers "no." Subjects who enter each section have very few skips, so both symptom scales and diagnostic information can be obtained. Child, parent, and teacher versions are available. Computer programs are available to implement diagnostic criteria, to generate severity scales based on each version, or to combine parent and child information. A typical DISC interview may take more than an hour for a child, plus an additional hour for a parent, but because of the stem question structure, the actual time varies widely with the number of symptoms endorsed. The DISC was designed for lay interviewers. It is fairly complicated to administer, and formal training programs are highly recommended. Reliability of the DISC is only fair to good and generally is better for the combined child and parent interview. Validity judged against a clinical interview by a child psychiatrist is also fair to good, better for some diagnoses and better for the combined interview. The DISC is well tolerated by parents and children and can be used to supplement a clinical interview to ensure comprehensive diagnostic coverage. Because of its inflexibility, some clinicians find it uncomfortable to use, and its length makes it less than optimal for use in clinical practice. However, it is used frequently in a variety of research settings.

Children's Depression Inventory (CDI) The CDI is a 27-item self-report measure of mood symptoms in children aged 7 to 17. A 10-item screening version is also available. The instrument may be administered as a simple pen-and-pencil test using a special easily scored form or by computer. The CDI has a first-grade reading level, so even young children can generally complete it on their own. Good reliability has been demonstrated, and good validity is suggested by its ability to distinguish currently depressed children from those with partially remitted depression or with other psychiatric disorders, along with its correlation with other measures of childhood depression. The principal uses of the CDI are in screening for depression in psychiatric patients or epidemiological surveys: it works fairly well for this purpose but tends to miss a substantial fraction of cases because children are not always good reporters. The other use is as a change measure in office practice or clinical trials. It appears to have adequate sensitivity to change for this purpose but has not been subjected to extensive evaluation.

Conners Rating Scales The Conners Rating Scales are a family of instruments designed to measure a range of childhood and adolescent psychopathology, but are most commonly used in the assessment of attention-deficit/hyperactivity disorder. There are teacher, parent, and self-report (for adolescents) versions and both short (as few as 10 items) and long (as many as 80 items, with multiple subscales) forms. Extensive normative data drawn from an ethnically diverse population are available for each sex across a broad age range. Even the longer forms can be completed in 15 to 20 minutes, and scoring can be accomplished rapidly. Rater training is not required. Reliability data are excellent for the Conners Rating Scales. However, teacher and parent versions tend to show poor agreement. Validity data suggest that the Conners Rating Scales are excellent at discriminating attention-deficit/hyperactivity disorder patients and normal controls. It has more difficulty separating attention-deficit/hyperactivity disorder from other disruptive behavioral disorders such as conduct disorder, but this may substantially relate to the genuine clinical difficulties separating these syndromes. Newer versions of the Conners have been developed that aim to improve these discriminations, but they have not yet been subjected to extensive testing. The principal uses of the Conners Rating Scales are in screening for attention-deficit/hyperactivity disorder in school or clinic populations and following changes in symptom severity over time; sensitivity to change in response to specific therapies has been demonstrated for most versions of the Conners.

SUGGESTED CROSS-REFERENCES

Section 5.2 discusses statistics and experimental design. Section 7.1 discusses the psychiatric interview, history and mental status examination; 7.5 the personality assessment of adults and children; 7.4 neuropsychological assessment of adults; and 9.6 neuropsychological assessment of children. Section 11.1 discusses the classification of mental disorders.

SECTION REFERENCES

*Carmines E, Zeller R: *Reliability and Validity Assessment*, publ no. 17. Sage, Beverly Hills, CA, 1979.

Hulley SB, Cummings SR: Planning the measurements: Precision and accuracy. In *Designing Clinical Research: An Epidemiologic Approach*, SB Hulley, SR Cummings, editors. Williams & Wilkins, Baltimore, 1987.

Kendler KS: Toward a scientific psychiatric nosology: Strengths and limitations. Arch Gen Psychiatry *47:*969, 1990.

McDowell I, Newell C: *Measuring Health: A Guide to Rating Scales and Questionnaires*. Oxford University Press, London, 1996.

Murphy J: Diagnostic schedules and rating scales in adult psychiatry. In *Textbook in Psychiatric Epidemiology*, MT Tsuang, M Tohen, GEP Zahner, editors. Wiley, New York, 1995.

Murphy JM, Berwick DM, Weinstein MC, Borus JF, Budman SH, Klerman GL: Performance of screening and diagnostic tests: Application of receiver operating characteristic (ROC) analysis. Arch Gen Psychiatry *44:*550, 1987.

*Nunnally JC: *Psychometric Theory*, ed 2. McGraw Hill, New York, 1978.

Regier DA, Kaelber CT, Rae DS, Farmer ME, Knauper B, Kessler RC, Norquist GS: Limitations of diagnostic criteria and assessment instruments for mental disorders: Implications for research and policy. Arch Gen Psychiatry *55:*109, 1998.

*Robins E, Guze SB: Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. Am J Psychiatry *126:*983, 1990.

Robins L: Epidemiology: Reflections on testing the validity of psychiatric interviews. Arch Gen Psychiatry *42:*918, 1985.

Schutte NS, Malouff JM: *Sourcebook of Adult Assessment Strategies*. Plenum, New York, 1995.

*Sederer L.L, Dickey B: *Outcomes Assessment in Clinical Practice*. Williams & Wilkins, Baltimore, 1996.

Shrout PE, Spitzer RL, Fleiss JL: Quantification of agreement in psychiatric diagnosis revisited. Arch Gen Psychiatry *44:*172, 1987.

*Spitzer RL: Psychiatric diagnosis: Are clinicians still necessary? Compr Psychiatry *24:*399, 1983.

Spitzer RL, Fleiss JL: A re-analysis of the reliability of psychiatric diagnosis. Br J Psychiatry *125:*341, 1974.

Sudman S, Bradburn NM, Schwarz N: *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass, San Francisco, 1996.

Winokur G, Zimmerman M, Cadoret R: 'Cause the Bible tells me so. Arch Gen Psychiatry *45:*683, 1988.

Zarin DA, Earls F. Diagnostic decision making in psychiatry. Am J Psychiatry *150:*197, 1993.

7.9 COMPUTER-BASED TESTING OF THE PSYCHIATRIC PATIENT

MARVIN J. MILLER, M.D.

Twenty-five years of advances in personal computers and software have begun to affect how patients are evaluated and treated and how information about them is stored. Few clinicians have managed to avoid the impact of computers entirely.

New tools will be widely used only if they are tested, validated, confirmed, and publicized throughout the professional community. Computer applications have slowly gone through this process and have gained acceptability. Insurance and managed-care companies lead the way in implementing computers for accounting, monitoring, and practice-shaping tools. Hospitals generally introduce computers through the business office and bring them no further into the health care process.

A few computer applications, however, have gained rapid acceptance and use. The Internet is a worldwide communication network of cables, software, and computers that has rapidly gained popularity in psychiatric education, communication, and program distribution. The Internet is a decentralized tool and thus amenable to rapid growth unfettered by rigid rules and constraints. Considerable caution is therefore required on the part of the clinician because there is little regulation or supervision. Individual clinicians browsing the Internet can explore applications and Web sites of value and can benefit from them immediately if they are of demonstrated quality.

Computers equipped with appropriate cameras and software can transmit video and audio over telephone lines at low cost. This opens rich opportunities for the mental health clinician to consult with colleagues, receive education, and evaluate and treat patients. It also raises many new questions of reimbursement, supervision, licensure, and efficacy. However, the power of the technology is setting the pace and more than fifteen states have some telemedicine projects in place.

Some computerized evaluation tasks (quantitative electroencephalogram (EEG) and magnetic resonance imaging) are readily accepted as computerized tools because they would be impossible or impractical without the computational power of computers. Other tasks such as computer-based testing of patients have experienced slow but gradual acceptance. Computer education, therapy, or both of patients has proceeded even more slowly. The area of mental health computing with the most untapped potential is the computerized evaluation of patients.

COMPUTER-BASED TESTING

In the past 30 years of computerized testing in the mental health field there has been a steady increase in the quality and quantity of test instruments available and a steady decrease in the costs of these tests. The earliest computerized interviews were done at the University of Wisconsin on a LINC computer provided by the National Institute of Mental Health. The initial computerized interviews gathered a medical history that was then printed out for the physician to use. The interview was branching and would select a maximum of 320 questions out of a library of more than 500 questions in the program.

Progress in the development of computerized testing remained very slow until the arrival of the microcomputer in 1980 changed the rate of development. Individual clinicians could more easily justify the cost of the equipment and could dedicate it to a task of computer testing. The hardware capabilities available expanded rapidly to include color graphics, photos, sound, and even physiological modalities.

Early attempts at computerizing psychiatric tests involved slavish attempts to mimic the traditional tests exactly. One early investigator devised a computerized version of an intelligence test on a PDP-8 computer with a concrete replication of the original formats. When different sections of the tests were presented, drawers would open and close to present the proper test material to a patient. The same investigator also developed one of the earliest computerized therapy programs; a group of patients eager to lose weight were treated with either a computerized program, a peer group, or a control group—all lost significant amounts of weight at an equal rate.

Dimensions of Computerized Testing The gradual realization of the many advantages of computer-based testing prompted the gradual move toward computerized testing of the psychiatric patient. Answers can be gathered from the patient more quickly and more accurately by the computer. One question presented on the screen at a time helps the patient to focus and answer that question more quickly than if presented on paper. Accuracy is enhanced because there is less chance that the patient will record answers incorrectly and there is less chance that the psychometrist will record or key punch the answers incorrectly. Computer-based testing also is less threatening and encourages patients to reveal sensitive material more readily. The computer summarizes the responses to the questions more quickly and more accurately than humans do.

Interpretation of the raw data is still a difficult task. Various studies show that a combination of computerized summary and human interpretation provides the most reliable and most comprehensive evaluation of the test results. Neither the human nor the computer can do as well individually as the two combined. However, acceptance of computerized testing remains mixed. Various evaluations show that patient acceptance (even among seriously ill patients) is higher than clinician acceptance.

Some clinicians reject computer-based testing entirely. Others accept the testing results in an uncritical way, as if the computer were infallible. Proper use of computer-based testing avoids both these extremes; the use of these tests is recommended only after careful development and assessment of test results in conjunction with other tests in the battery and other clinical information gathered in more traditional ways.

Early Developments John Greist was an early pioneer in computerized testing in mental health. He developed an early program to assess suicidality that showed that patients will reveal sensitive information about suicidal impulses to the computer. The computer program could predict the probability of a suicide gesture as well as clinicians could. His group went on to develop a variety of computerized tests for diagnostic purposes and symptom assessment in a variety of areas. The first psychological test to be widely computerized was the Minnesota Multiphasic Personality Inventory (MMPI). The volume of test questions and the large number of experimental scales dictated the use of the computer as a time-saving measure for test administration and test scoring.

Ethical and Professional Concerns Computer-based testing is only one component of a professional evaluation of mental illness. The mental health professional must be the main source of quality assurance during the evaluation and needs to integrate the computer-based testing with all the other aspects of the evaluation. Psychiatrists and other professionals are bound by strict guidelines about initial specialty training, certification in the specialty, continuing medical education, and general ethical conduct. The highly trained professional signing the evaluation of a patient assumes the ultimate responsibility for the content and quality of an evaluation.

Persons constructing a computerized test need to be concerned about various aspects of reliability and validity. They need to ensure that the test will produce similar results at various points in time. Use of the computer can often enhance standardization of a test because the conditions under which the test is administered can be more precisely regulated.

The validity of a test also needs to be addressed by the test designer or developer. Tests need to demonstrate content validity by proving that adequate aspects of the topic in question are sampled. Computerized tests also need to show criterion-related validity. A test to measure attention-deficit/hyperactivity disorder should show a high correlation with an expert clinician making the same diagnosis; computer-based testing should also show predictive validity if possible. The computer evaluation of attention-deficit/hyperactivity disorder should show change and correlation with medication used to treat the disorder. The computer-based testing should also demonstrate construct validity by showing that a carefully defined theoretical construct is measured and described.

For some tests the issue of reliability can be addressed by comparing the paper-and-pencil form of the test with the computerized version; these comparisons

generally show a very high correlation. Some tests, however, are designed specifically for the computer and measure things that cannot be easily measured without a computer. Testing for reliability and validity becomes more complex for these tests.

The results of reliability and validity studies should be published in the professional literature and should be reviewed by clinicians attempting to use a specific test in their practice. Some tests enjoy wide popularity among the public and among some clinicians despite the lack of evidence in the literature of concurrent or predictive validity. Careful clinicians will ask for a bibliography regarding a test before purchasing and using that test.

Taking clinical action on a set of numbers or phrases derived from computerized testing is similar to getting numbers back from a clinical laboratory report of a blood test. The clinician must first have a measure of confidence in the clinical laboratory or the person developing the computerized test. The individual numbers, however, mean nothing unless they are evaluated within the context of the total patient condition and are integrated with other complementary or contradictory findings. The clinician's skill in using computer-based testing is the most central part of the overall ability to evaluate a patient.

Does Computerized Testing Work? Patient acceptance of computerized testing has been very rapid and clinician acceptance is gradually increasing. Studies that have compared old (usually paper-and-pencil) versions of tests with computer-based tests have generally shown very good equivalence. However, when a group of 188 psychiatric patients were given a standard verbal digit span test as well as a computerized (visual) version of the verbal digit span the patients performed more poorly on the computerized test. The State Trait Anxiety Inventory and the Beck Depression Inventory (BDI) were administered in paper-and-pencil and computerized formats. The study also evaluated the amount of computer anxiety or computer phobia present among the students and found that those with greater computer anxiety also had high BDI scores. The study was not designed to find out why the depression scores were higher or whether they more accurately revealed the amount of a depression present in the subject. A study of alcohol consumption seemed to show that patients were more willing to divulge sensitive information about alcohol consumption to the computer than they were to the human interviewer.

There are also valid questions concerning the tendency of computerized test subjects to respond in socially desirable ways. One study of 162 college students tested with the MMPI, the Research Evaluation Form, and the Drinking Habits Questionnaire supported the hypothesis that subjects reveal less damaging information to a human interviewer. Another study used the Paulhus Balanced Inventory of Desirable Responding (BIDR) on a group of 241 university students. The subscale evaluating responses in the socially desirable direction was higher in the computer-tested group as compared with the group administered this test by paper and pencil. However, a more refined version of the socially desirable subtest revealed no significant differences between computer and paper-and-pencil administration. Students who took these tests anonymously reported more pathology than they did in the nonanonymous situation.

Using a later version of the BIDR test in an effort to replicate these results in a sample of 246 naval recruits revealed similar effects for anonymity but different effects for method of administration, which supports the idea that the initial version of the social desirability scale contains some confounding factors that did not properly evaluate the question of truthful responding.

Deliberate Faking An evaluation of 100 college students taking a test by computer compared with 100 students taking the same test by paper and pencil revealed that 10 percent of the former group gave random responses and 4 percent of the latter group gave random responses. However, the random responders could be identified in both groups by inclusion of a scale that consisted of statements that could not possibly be true. Response latency also identifies faking. *Latency* is the number of milliseconds required to answer each question and the shortest response latencies are noted in random responding; this is followed by faking good responses, followed by honest responses, followed by faking bad patterns. There is a suggestion, however, that latency scores regarding questions evaluating medical problems are not as predictable.

Another effort evaluated malingering. The addition of response latency as a measure of malingering increased the correct classification of malingerers by 30 percent as compared with traditional methods of detection. An extensive evaluation of patient acceptance in computerized testing situations revealed very good acceptance. The three main dimensions of this study compared traditional testing versus keyboard testing versus touch screen testing. Three hundred and sixty-seven psychiatric patients were tested on a variable number of eight computerized tests available. The standardized test was preferred in one case and the computerized test was preferred in another. In one other test the standard approach was preferred over the keyboard approach but was no different from the touch screen approach in terms of patient preference. There were, however, frequent differences between the touch screen approach—four tests were more easily or enjoyably performed on the touch screen as compared with the keyboard and only in one situation was the keyboard preferable to the touch screen to a significant degree. The researchers concluded that computerized testing was widely accepted.

Cost Impact of Computerized Testing Potential cost savings exist in many aspects of computerized testing but surprisingly little has been done to document this. Administration of a test by a microcomputer, if amortized over the life of the computer, will probably be less than $1 per hour. This compares very favorably with what it costs when a psychometrist or a psychologist administers the same test. Test scoring by computer saves additional time and money. The computerized printout of test scores takes less time than the human typing of test results would require. Test royalties, if applicable, would be the same in both cases.

Clinicians who have computerized testing readily available tend to perform more tests and to do a broader evaluation of the patient. This does have implications for quality of care. Computerized testing can be summarized and available for the clinician to view much earlier than testing done by conventional means. In some hospital settings this can make a difference of 1 day of hospital care. More thorough assessment and more focused treatment also has the potential to reduce the length of stay for a patient or to reduce the number of outpatient visits, which can reduce total health care costs. More studies need to be done to show the potential savings. It is estimated that more than 500 computerized tests are now available for use within the mental health field.

Tests for Children and Adolescents A community center offering crisis and counseling services for children and adolescents can administer a screening battery to help with evaluation. Self-administered tests can evaluate family structure, family satisfaction, and drug use. This system can help to summarize the initial evaluation, provide a storage medium for problem lists, maintain notes of treatment sessions, and keep records for billing purposes. Diagnosing and monitoring attention-deficit disorders in children and adolescents is a very important task. Many different investigators have computerized this task and it does seem sensitive to treatment with methylphenidate (Ritalin). Some have started programs to measure these disorder via a microcomputer or on dedicated equipment.

The computerized continuance performance test has been significantly correlated with teacher ratings of school performance, subtests of the Wechsler Intelligence Scale for Children-Revised that relate to attention, and some scales from the Conner Parent and Teacher Rating Scales.

Parents of children with problems fill out the computerized version of the Child Behavior Checklist and report preferring it to the paper-and-pencil version. Respondents further displayed this preference by providing significantly more information about their child's difficulties on the computerized version.

An evaluation of the paper-and-pencil versus the computer-administered version of the Piers Harris Children's Self Concept Scale revealed no significant differences between the means of the scales derived from each version. The scale summarizing behavioral problems did show significantly more variance on the paper-and-pencil test as compared with the computerized version.

Personality Tests The Minnesota Multiphasic Personality Inventory (MMPI) are the most widely used assessments of personality and have good equivalence between the computer version and the traditional version. One study compared 80 community volunteers under one of four testing conditions: computer-computer, traditional-traditional, computer-traditional, or traditional-computer. These four sequences revealed no significant differences on any of the traditional scales and on 27 special scales. Another study used 150 college students to compare computer versus traditional administration of the MMPI. This study revealed more differences in the time of testing variable as compared with the format of testing. An evaluation of computer versus traditional testing of the MMPI in a group of 77 substance abusers also found no significant differences between formats.

The computerized version versus the paper-and-pencil version of another popular test of personality, the Personality Research Form-E, was compared and showed good test-retest stability. Surprisingly, the infrequency scale (used to detect careless responding) was the only scale that did not show good levels of reliability. The sample population of 55 subjects was quite small and probably not sufficient to warrant discarding the Infrequency scale as has been suggested.

The Jenkins Activity Survey helps to detect the type-A personality style associated with specific physiological responses to stress. Computerized and paper-and-pencil versions were compared in a group of 60 college students. There were no significant differences in the two forms when comparing method of administration or time (first or second) of testing. Reaction time for each question was also measured and was found to be significantly associated with the Speed and Impatience Scale on the second testing but not on the initial testing.

Obsessive-Compulsive Disorder Computerized tests assess the severity of symptoms of obsessive-compulsive disorder very well. The Yale-Brown Obsessive Compulsive Scale (Y-BOCS) evaluates the presence and severity of symptoms of this disorder and was computerized with high reliability results. The total Y-BOCS score correlated at $r = .88$ between the clinician-administered and the computer-administered forms and the subscales correlations were almost as high. Preference of the patient regarding form of testing was evenly split between the computer and the clinician versions. A version of the Y-BOCS has also been designed for administration by touch-tone telephone hooked up to a talking computer. This method was compared with a live telephone interview and a paper-and-pencil version. The agreement among the three methods was very high ($r = 0.99$) but more of the patients (44 percent) preferred the live interview and none preferred the talking computer.

Depression The most widely used self-rating scale for depression is the BDI. This was computerized and administered to 330 inpatients. There was no paper-and-pencil comparison group but total scores and coefficient alpha were very similar to other published studies of the BDI. The scores on the BDI also correlated strongly with the clinical diagnosis of depression. The most frequently used depression rating for use by clinicians is the Hamilton Rating Scale for Depression (HAM-D). A self-administered version of the Hamilton scale that was developed for computer administration required 37 questions to cover the original 17 items. The mean difference in total score for the total sample was only 0.1 (17.29 for the clinical version and 17.19 for the computerized version). The correlation with the BDI was $r = 0.925$. Using a cutoff score of 17, the clinician-administered version showed a specificity of 100 percent and a sensitivity of 100 percent in identifying patients with major depressive disorder whereas the computer version showed a specificity of 100 percent and a sensitivity of 94 percent.

Anxiety The Hamilton Rating Scale for Anxiety (HAM-A) has been one of the instruments most frequently used to evaluate anxiety, which is among the most frequent of the mental disorders in the United States. The HAM-A contains 14 items designed to be administered by a trained clinician. The computerized version of the HAM-A designed for self-administration contains 31 questions and was tested in a group of 292 adults including persons diagnosed as having a generalized anxiety disorder, panic disorder with or without agoraphobia, social phobia or obsessive-compulsive disorder, major depressive disorder, and community controls with no current diagnosis. The computer and the clinician versions of this test were administered in a counterbalanced form 1 week apart. The scores in the patients with anxiety disorder correlated very highly between the computer and clinician versions (coefficient alpha $r = .92$). The computer and the clinician versions of the test in a group of anxiety disorder patients were different at a significance of $P < .05$. This, however, translated to a difference of about one point in the total scale score, which may not translate to a clinically meaningful difference. On each occasion the scores on the computer were slightly higher than the scores of the clinician rater.

Neuropsychology and Psychophysiology One large category of computer programs exploits the unique abilities of the microcomputer to administer tests consistently and in a very carefully timed pattern. Neuropsychological tests are very expensive to administer, score, and interpret. The digit span test, used to evaluate attention and memory, is simple enough so that the computerized version is fairly straightforward. The digit substitution test is a more complex evaluation of logic in addition to memory and can help to evaluate the effects of research medications on cognition.

Patients with schizophrenia who were tested on a computerized version of the Tower of London test showed a slow and inefficient pattern of problem solving. The patients were slower in moving the pieces of the puzzle and required more moves to solve a particular problem. Schizophrenia patients also tend to do poorly in a specially modified version of the continuous performance test. These patients show a greater drop in performance over a 10-minute task than do normal controls.

The evaluation of memory is an important part of the psychiatric evaluation. Recently, a comparison of eight different implementations of a memory test was conducted. The Halstead category test is a test that measures abstracting ability and is very sensitive to various kinds of brain damage. The test has been successfully computerized with good replication of the traditional testing method. The Cognitive Neuromotor Testing Battery includes five different tasks designed to measure speed of processing, vigilance, memory, and visuomotor skills. This test was originally designed to measure skills needed to drive but it was also found to be sensitive to the adverse effects of medications such as those accompanying treatment with benzodiazepines. The test was found to have a good correlation with an existing neuropsychological test.

Investigators have also noted the sensitivity of computerized testing in detecting neurobehavioral toxicology and a series of computerized tests are available for this purpose and are reviewed in the literature.

Psychiatric Diagnosis The Diagnostic Interview Schedule is a diagnostic instrument used by psychiatrists and lay interviewers in several large epidemiological studies. Its reliability as a valid diagnostic tool is established and several investigators have computerized this tool. The reliability of these computerized tests has generally been modest—the kappa values for the Levitan study ranged from .17 to .92 with a mean of .46. These values show the need to regard the computerized diagnostic instrument as suitable for screening only; they also point out the need for correlation of computer testing with the remainder of the clinical evaluation.

Anorexia Nervosa A self-administered test for anorexia nervosa was computerized and is correlated highly with the standard paper-and-pencil version according to a recent sample of 47 subjects.

Posttraumatic Stress Disorder Approximately 15 to 20 percent of war veterans report this disorder; civilians who experience a severely traumatic event are also at high risk. A semi-structured clinical interview was computerized for more efficient administration; the computerized version includes 34 questions and takes about 15 minutes to administer. The kappa for diagnosis of the disorder was .90 with a sensitivity of .95 and a specificity of .95, indicating excellent validity.

Alcoholism Because treatment costs for alcohol and substance abuse problems have surpassed treatment costs for all the other mental illnesses since the mid-1980s, careful and efficient detection of these disorders of abuse is extremely important. Computerized screening and follow-up instruments have shown good validity and reliability. The self-administered Alcoholism Screening Test is a 35-item test in its long form and a 9-item test in its short form and it has demonstrated validity. The computerized version showed excellent correlation with the paper-and-pencil version. The differences in regard to format and order of administration were both nonsignificant. Agreement between the two forms was 95 percent. A follow-up instrument, the Substance Use Disorder Diagnostic Schedule, showed a similarly high agreement with the interviewer-administered format. k values were .86 for presence or absence of alcohol abuse and .86 for presence or absence of other drug abuse. These two computerized tests reported mild but nonsignificant increases in the amount of alcohol consumption as compared with reports from the interviewer-administered forms.

Vocational Testing This testing can advance the process of making a career choice by identifying particular interests and aptitudes. The array of interests can then be compared to the responses of persons in particular fields and can help to achieve a compatible career choice. Computerized batteries have been developed to test an individual's aptitudes and interests and each of these tests has shown increased efficiency and reliability in the testing process.

Behavioral Assessment Careful evaluation of potential reinforcers for a behavior therapy program conducted in a computerized evaluation was found to save time in identifying the optimal set of behavior reinforcers.

Adaptive Testing Adaptive testing consists of a computerized method for selecting the optimal questions to be presented to a patient. These consist of branching patterns that depend on a screening question to decide the particular branches taken for a particular patient. The advantages of this method are that it saves the patient testing time and energy. Application of this method to lengthy tests such as the MMPI show substantial correlation with the full administration of the MMPI. Shortening the testing time is a dubious achievement because the computer-administered testing is extremely inexpensive. Development and validation of subscales of the MMPI or other tests is also hindered when only a subset of the test is administered to any particular patient, making it very difficult to develop norms. Questions that serve to screen patients and determine the choice of branches are sometimes too brief and too superficial to make a sensitive determination of the branching, which reduces the overall validity of the test; it is therefore wise to be cautious when using adaptive testing methods.

Legal Aspects of Computerized Testing Computerized testing is similar to any other kind of psychological evaluation in the legal arena. The expert clinician on the witness stand is the one who is under scrutiny and must use these evaluation tools within the proper context and must have information about their development and validity. The expert must then consolidate all the available structured and clinical evaluation in order to form an opinion with “reasonable certainty.” Because the computerized testing can be subjected to measures of validity and reliability, the expert witness must be prepared to document knowledge of the development of a test. The clinical opinion of an expert witness is also subject to cross-examination and will almost certainly have less documented evidence of accuracy or validity.

Computer-Based Test Interpretation Simple computerized scoring of test results is rapid, precise, and efficient. Computer scoring of tests such as the MMPI-2 with its many research scales is almost a necessity and is well accepted.

Computer-based test interpretation of the test results is much more controversial. Validation of computer-based test interpretation has lagged behind the validation of

the test in question. Many computer-based test interpretation systems had attempted to generate interpretive statements for the MMPI-2. These statements were generally compared with statements by expert clinicians who looked at their own interpretation of the MMPI-2 but these were rarely compared with direct clinical observation of patients for cross-validation. Some interpretive systems make public the decision rules or the mathematical formulas that govern the generation of specific statements appearing on the interpretive summary. Other systems, however, keep the decision-making process secret and thus the clinician has no specific basis on which to judge the legitimacy of the derived interpretation, which makes it very difficult to use interpretive statements in court when the basis of those interpretations has not been revealed. Users of computer-based test interpretation products deserve to know the research underlying interpretive statements.

Legitimate arguments exist about the optimal mathematical approaches to use in modeling the process of psychiatric diagnosis. The cookbook approach used in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* is easy to understand conceptually but is too simplistic to reflect accurately how human beings actually make diagnostic decisions and act on them. Some computerized systems use the logical decision tree, the Bayesian model, cluster analysis, multiple discriminate functions, or even neural networks. In fact, different clinicians may use different models successfully. Some clinicians are able to define the models they use whereas others cannot reduce these models to specific algorithms, which helps to explain why the field of computer-based test interpretation continues to evolve and must be approached cautiously.

FUTURE DIRECTIONS

The rapid proliferation of the Internet has attracted the attention of those interested in public education and in the assessment of psychiatric illnesses. Several Internet sites now offer the public a chance to answer a select group of questions on a particular illness and receive feedback about the presence or absence of the illness in themselves. Referral sources are offered so that the individual can follow up on the situation. A recent World Wide Web search using the words “test” and “mental” found three Web sites (among the first 30 listed) offering online testing. One additional site offered a free test for depression that could be downloaded and used. Clinicians can also obtain tools to construct their own computerized questionnaires. This, along with the availability of numerous commercial testing programs, allows clinicians to use computerized testing of psychiatric patients in more extensive and comprehensive ways than ever before.

SUGGESTED CROSS-REFERENCES

Statistics and experimental design are covered in Section 5.2. Psychiatric rating scales are discussed in Section 7.8. The psychiatric report and medical record is covered in Section 7.2.

SECTION REFERENCES

*Allen CC, Ellinwood EH, Logue PE: Constructive validity of a new computer-assisted cognitive neuromotor assessment battery in normal and inpatient psychiatric samples. J Clin Psychol *49:*784, 1993.

Baer L, Brown-Beasley MW, Sorce J, Henriques A: Computer-assisted telephone administration of a structured interview for obsessive-compulsive disorder. Am J Psychiatry *150:*1737, 1993.

Booth-Kewley S, Edwards JE, Rosenfeld P: Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? J Appl Behav Sci *77:*562, 1992.

Bucholz KK, Robins LN, Shayka JJ, Przybeck TR, Helzer JE, Goldring E, Klein MH, Greist JH, Erdman HP, Skare SS: Performance of two forms of a computer psychiatric screening interview: Version I of the DISSI. J Psychiatr Res *25:*117, 1991.

Choca J, Morris J: Administering the Category Test by computer: Equivalence of results. Clin Neuropsychol *6:*9, 1992.

Conoley CW, Plake BS, Kemmerer BE: Issues in computer-based test interpretive systems. Comp Hum Behav *7:*97, 1991.

Davis LJ, Hoffman NG, Morse RM, Luehr JG: Substance Use Disorder Diagnostic Schedule (SUDDS): The equivalence and validity of a computer-administered and an interviewer-administered format. Alcohol Clin Exp Res *6:*250, 1992.

Davis LJ, Morse RM: Self-administered alcoholism screening test: A comparison of conventional versus computer-administered formats. Alcohol Clin Exp Res *15:*155, 1991.

Faust D, Ziskin J: Computer-assisted psychological evaluation as legal evidence: Some day my prints will come. Comput Hum Behav *5:*23, 1989.

Fekken GC, Holden RR: Psychometric evaluation of the microcomputerized personality research form. Educ Psychol Measure *49:*875, 1989.

*Finger MS, Ones DS: Psychometric equivalence of the computer and booklet forms of the MMPI: A meta-analysis. Psychol Assess *11:*58, 1999.

Flowers JV, Booraem CD, Schwartz B: Group therapy client outcome and satisfaction as a function of the therapists' use of rapid assessment instruments. Small Group Res *24:*116, 1993.

French CC, Beaumont JG: The reaction of psychiatric patients to computerized assessment. Br J Clin Psychol *26:*267, 1987.

French CC, Beaumont JG: Microcomputer version of a digit span test in clinical use. Interacting Comput *4:*163, 1992.

*Glover G: Mental health informatics and the rhythm for community care: Information systems in psychiatry must be released from the asylums and updated. BMJ *311:*1038, 1995.

Goodman WK, Price LH, Rasmussen SA, Mazure C, Fleischmann RL, Hill CL Heninger GR: The Yale-Brown Obsessive Compulsive Scale: I. Development, use and reliability. Arch Gen Psychiatry *46:*1006, 1989.

Greist JH, Klein MH, Erdman HP, Bires JK, Bass SM, Machtinger PE, Kresge DG: Comparison of computer- and interviewer-administered versions of the Diagnostic Interview Schedule. Hosp Community Psychiatry *38:*1304, 1987.

Harrell TH, Honaker LM, Hetu M, Berwager B: Computerized versus traditional administration of the Multidimensional Aptitude Battery—Verbal scale: An examination of reliability and validity. Comput Hum Behav *3:*129, 1987.

Hedlund JL, Vieweg BW: Automation in psychological testing. Psychiatr Ann *18:*217, 1988.

Honaker LM, Harrell TH, Buffaloe JD: Equivalency of microtest computer MMPI administration for standard and special scales. Comput Hum Behav *4:*323, 1988.

Kane RL, Kay GG: Computerized assessment in neuropsychology: A review of tests and test batteries. Neuropsychol Rev *3:*1, 1992.

Kobak K, Reynolds WM, Greist JH: Development and validation of a computer-administered version of the Hamilton Anxiety Scale. Psychol Assess *5:*487, 1993.

Kobak KA, Reynolds WM, Rosenfeld R, Greist JH: Development and validation of a computer-administered version of the Hamilton Depression Rating Scale. Psychol Assess *2:*56, 1990.

Kobak KA, Taylor L, Dottl SL, Greist JH, Jefferson JW, Burroughs D, Katzelnick DJ, Mandell M: Computerized screening for psychiatric disorders in an outpatient community mental health clinic. Psychiat Serv *48:*22, 1997.

Levitan RD, Blouin AG, Navarro JR, Hill J: Validity of the computerized DIS for diagnosing psychiatric inpatients. Can J Psychiatry *36:*728, 1991.

Mattila MJ, Aranko K, Mattila ME, Paakkari I: Effects of psychotropic drugs on digit substitution: Comparison of the computerized symbol digit substitution and traditional digit symbol substitution tests. J Psychopharmacol *8:*81, 1994.

Miller MJ: Computerized models of psychiatric diagnosis. In *Using Computers in Clinical Practice*. EM Schwartz, editor. Haworth Press, New York, 1984.

*Miller MJ, Hammond KW, Hile MG, editors: *Mental Health Computing*. Springer-Verlag, New York, 1996.

Neal LA, Busuttil W, Herapath R, Strike PW: Development and validation of the computerized clinician administered Post Traumatic Stress Disorder Scale-1-Revised. Psychol Med *24:*701, 1994.

Nestor PG, Faux SF, McCarley RW, Shenton ME, Sands SF: Measurement of visual sustained attention in schizophrenia using signal detection analysis and a newly developed computerized CPT task. Schizophr Res *3:*329, 332, 1990.

Rosenfeld R, Dar R, Anderson D, Kobak KA: A computer-administered version of the Yale-Brown Obsessive Compulsive Scale. Psychol Assess *4:*329, 1992.

Steer RA, Rissmiller DJ, Ranieri WF, Beck AT: Use of the Computer Administered Beck Depression Inventory and Hopelessness Scale with psychiatric inpatients. Comput Hum Behav *10:*223, 1994.

Vansickle TR, Kapes JT: Comparing paper-pencil and computer based versions of the Strong-Campbell Interest Inventory. Comput Hum Behav *9:*441, 1993.

*Vaterott M, Callier J, Hile M: The development of the Missouri Automated Reinforcer Assessment (MARA): An update. In *Computer Applications in Mental Health: Education and Evaluation*, MJ Miller, editor. Haworth Press, New York, 1992.

Vieweg BW, DiFranco B: The use of automated assessment with seriously mentally ill clients. Behav Health Tomorrow *4:*37, 1995.

CHAPTER 8. CLINICAL MANIFESTATIONS OF PSYCHIATRIC DISORDERS

JOEL YAGER, M.D., AND MICHAEL J. GITLIN, M.D.

Like other medical disorders, psychiatric disorders express themselves in characteristic ways. Deviations from normal, from mild to severe, may occur in the intensity, duration, timing, and content of thoughts, emotions, and behaviors. Inquiring about subjective complaints and clinical symptoms and observing and eliciting the clinical signs of psychiatric disorders parallels taking a history and conducting a physical examination in general medicine. Many psychiatric complaints and disorders have to be understood in a broad context, requiring a more thorough evaluation and comprehension of the patient's interpersonal world, work role, family life, and culture than is typical in general medical practice. The nature and expression of psychiatric signs and symptoms are profoundly altered by the patient's strengths, coping capacities, and psychological defenses, so that the clinical picture ultimately represents a balance between psychopathology and psychological strengths.

The most important distinction between typical presentations of medical diseases and those of psychiatric disorders is the greater importance in psychiatric disorders of the patients' sometimes idiosyncratic descriptions of their qualitative internal states, subjective experiences that are often difficult to describe in words. Poets and novelists are often more capable than clinicians at characterizing and delineating the precise quality and experience of many psychiatric symptoms. Many patients and clinicians often find it difficult to accurately communicate a fully comprehensible and reliable description of even familiar, somewhat universal feeling states.

> An usually articulate woman, a published writer, tried to relate the differences she felt between the fatigue she associated with her disabling Lyme disease and the fatigue she experienced during a previous bout of atypical depression. Although unusually gifted in her capacity to form images in words, she found that her capacity to express exactly what she meant was not up to the task. She remained frustrated, knowing that she had not been able to adequately describe or convey the subtle but real distinctions in how she had experienced these two states.

Subjective descriptions of psychiatric symptoms are inherently less reliable, or at least less objective, than more directly measurable and quantifiable data such as blood pressure, temperature, and laboratory test results. A great deal of the research in psychiatric diagnosis over the last 25 years has been concerned with increasing the reliability of observer-rated clinical symptom assessments. In many ways this research has had the desired impact—clinicians and researchers using a variety of structured interviews can come to reasonable agreement on what symptoms patients are experiencing and whether these patients meet the criteria for most of the specific psychiatric disorders in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). However, one of the costs of this increasing reliability has, in many instances, been the narrowing of the field of clinical vision. Clinicians who rely predominantly on structured interviews and checklists may become somewhat closed-minded and risk ignoring clinical phenomena that are very important, but that may not be part of the structured interview framework or mental set. Furthermore, the quest for reliability can go only so far in describing phenomena that are difficult to express in words.

Despite these difficulties, a thorough assessment of the clinical history and description of the psychopathology and a detailed account of the patient's subjective experiences are important because:

1. Significant diagnostic distinctions are made primarily on the basis of the historical information and elicited phenomenology. The more detailed, complete, and correct the diagnosis, the more rational and precise the treatment planning and the more reliable the prognosis. Consider, for example, the importance of accurately distinguishing between antipsychotic-induced akathisia and anxiety symptoms related to psychotic thinking. On the basis of the diagnosis made, opposite therapeutic strategies may be attempted.
2. The clinician's capacity to fully hear and communicate a comprehensive understanding of a patient's internal experiences helps to diminish the patient's sense of isolation, so characteristic of many of these disorders, and fosters the growth of a therapeutic alliance, increasing the likelihood of compliance with treatment.

PREDISPOSING VULNERABILITIES

Genetic and Intrauterine Factors Genetic vulnerabilities play an important role in the expression of many psychiatric disorders. Prominent among these are dementia of the Alzheimer's type, schizophrenia, mood disorders, anxiety disorders, and alcohol dependence. For virtually all these disorders, the nature of the inherited vulnerabilities is still largely unknown.

Intrauterine processes contribute to many psychiatric disorders. For example, maternal starvation and influenza infections during the second trimester of pregnancy have been implicated in the pathogenesis of schizophrenia. Maternal smoking and low birth weight may be risk factors in the pathogenesis of attention-deficit disorders in children. Maternal alcohol abuse or dependence may lead to the fetal alcohol syndrome, a major cause of developmental disability.

Constitutional Factors Considerable research demonstrates that by birth and shortly afterward infants differ widely in temperament—in their spontaneous activity levels and thresholds, intensity, and duration of their reactions to external stimuli; the regularity or irregularity of certain biological rhythms such as sleep; tendencies to approach or withdraw from new stimuli; the speed and degree of adaptation; attention span and distractibility; the persistence of behavior; and qualities of mood. Based on such early behaviors, children may be described as having easy or difficult temperaments, as being quick or slow to warm up. Temperament, however, is not immutable. There are discontinuities over time, and the development of temperament and its lasting impact on personality development is at least in part a function of the goodness of fit with a child's family. Nevertheless, these temperamental qualities correlate somewhat with behavioral problems, especially through early childhood.

Aside from temperament, other persistent normal variations in personality development seem to be constitutionally related, and may influence subsequent resilience or vulnerability. Traits such as introversion, extroversion, and neuroticism appear to be relatively enduring and stable personality dimensions. Other temperamental qualities that endure include tenacity, novelty-seeking, and being relatively open to new experiences. Subtypes of intelligence, such as those related to conceptual, mathematical, musical, kinesthetic, and interpersonal abilities have been postulated as having separate genetic determinants and patterns of development. The type A and B personality patterns, hardy and resilient personalities, highly strung, sensitive, fussy, irritable, and pessimistic characteristics have all been described as generally lifelong qualities that originate in early childhood.

Physiological Stressors Physiological vulnerability may result from long-standing problems or from newly acquired ones. All the metabolic, toxic, infectious, and other causes of physical illness produce increased vulnerability to psychiatric disturbance. Studies have shown higher utilization of psychiatric services by those who are physically ill, and higher than expected prevalence of physical disease among the psychiatrically impaired.

Some children with prepubertal onset, obsessive-compulsive disorder and tic disorders that have an episodic symptom course have been found to have pediatric autoimmune neuropsychiatric disorders associated with streptococcal (group A b-hemolytic streptococcal) infections (PANDAS). Accompanying symptoms during episodes of exacerbation are emotional lability, separation anxiety, nightime fears and bedtime rituals, cognitive deficits, oppositional behaviors, and motoric hyperactivity. In PANDAS patients flareups of behavioral problems are commonly associated with documented group A b-hemolytic streptococcal infections or symptoms of pharyngitis and upper respiratory infections.

Human immunodeficiency virus (HIV) infection leading to seropositivity and acquired immune deficiency syndrome (AIDS) vividly illustrates the multiple and complex ways in which stressors can lead to psychiatric disturbances. These patients' psychiatric symptoms may represent organic changes that are the direct effects of the

virus on the central nervous system (CNS), producing changes in cognition, personality, and mood; understandable psychological adjustment responses of patients in response to an overwhelming life-threatening disorder; or the emergence of latent or quiescent primary psychiatric problems provoked by the psychological stress of the viral illness.

A 55-year-old successful businessman with a strong family history for unipolar depression started to feel mentally sluggish, and developed depressed mood, psychomotor retardation, and ruminative self-doubts. A medical work-up revealed high serum calcium that led to a diagnosis of hyperparathyroidism, which was ultimately treated surgically. Once the parathyroid problem was corrected, the mood disorder disappeared without further treatment.

Environmental Stressors Complex relationships exist between various life events, particularly threatening, unpredictable, and uncontrollable negative events, and the development of psychiatric symptoms. In general, such undesirable life events predispose individuals to develop psychiatric symptoms, especially if they already have a preexisting psychiatric disorder. After exposure to the same negative stressors such as a serious accident or act of violence, individuals who previously had anxiety disorders are more prone than those without such histories to subsequently develop symptoms of posttraumatic stress disorder. Although individual responses vary widely, truly catastrophic events, such as incarceration in a concentration camp, cause enduring psychiatric disturbances in a high percentage of survivors whether or not they had prior psychiatric problems. Similarly, the stress-related consequences of combat also vary widely, so that some veterans exposed to combat develop long-lasting posttraumatic stress disorder whereas others develop very few persistent symptoms. The death of a parent or spouse, divorce, and major physical injury affect some people profoundly and others hardly at all in the long run. Significant stressors are likely to be more traumatic during early development rather than later, or at certain critical developmental periods compared to other times. For example, the loss of a parent at a very young age is likely to be more traumatic and have more profound and lasting effects than the loss of a parent as an adult.

The combined impact of negative life events and poor emotional and practical social supports is important in the pathogenesis of at least some psychiatric disturbances. One British study found that women who were depressed were much more likely to have lost a parent at an early age, to be relatively housebound with three or more young children, and to lack a good confiding relationship with a spouse or other confidant. In that study biological vulnerability to depression seemed less important than the accumulation of negative life circumstances in the development of the disorder. People who are usually competent in all role functions may fall apart completely when a supportive spouse who has bolstered them and taken care of many of their needs dies suddenly. Patients presenting with a major depressive episode have experienced more uncontrollable actual and threatened losses such as the death of a spouse in the year prior to onset. Nevertheless, not all psychiatric disturbance is attributable to easily identified provoking negative life events; indeed, some major negative life events that seem to have preceded the onset of a serious psychiatric disturbance may in fact have occurred only after the psychiatric disturbance actually began. For example, someone who attributes the onset of depression to having been fired from a job several months previously may already have been functioning suboptimally and may have been fired as a consequence of a depression-induced decline in role function.

Certain environmental features can counter the effects of environmental stressors and protect against breakdowns. Stable families and friends, good financial circumstances, and supportive churches and communities offer some protection. Research has shown that individuals with psychiatric disturbances have fewer social supports than normal controls. This may be caused by friends' and relatives' withdrawal from deviant behaviors or by the disturbed individual's withdrawal from deleterious family and social relationships. In contrast, physically ill persons have more social supports than others, perhaps reflecting their ability to recruit help in times of need. The quality as well as the quantity of social supports is important too. As has been demonstrated in schizophrenia and mood disorders, for example, even in close families negative relationships may have deleterious effects in initiating and in sustaining psychiatric disturbance.

The negative impact of a physiological or environmental stressor is closely related to its personal meaning to the individual in question. For example, the loss of a spouse who has been chronically demented, disabled, and burdensome will have a very different impact than the loss of vital, supportive, loving spouse.

CHARACTERISTICS OF PSYCHIATRIC SIGNS AND SYMPTOMS

Signs and symptoms form the two major categories of clinical phenomena. Classically, as for most medical disorders, the distinction between the two is clear. *Symptoms* are subjectively experienced disturbances that are not necessarily observable by others. Patients complain of symptoms—for example, chest pain, headache, and tingling sensations. *Signs* are abnormalities that are observable by an examiner, including those that are easily evident in the course of a routine encounter with the patient as well as those elicited only through specific physical, mental status, or laboratory examinations.

In psychiatry the line between symptoms and signs is often more blurred than in general medicine. For instance, many phenomena often considered to be symptoms of psychiatric disorders may not be experienced as psychiatric problems by patients. Hearing an angel's voice may represent a manifestation of a psychotic disorder, yet the patient may vigorously dispute that the experience is a psychopathological symptom. Additionally, auditory hallucinations are often considered to be signs of a psychotic disorder even though, by their very nature, they are subjective internal experiences (symptoms). Further complicating the distinction, some psychiatric phenomena, such as the classical psychological defense mechanisms, may only be inferred from speech and behaviors but are not directly observable.

Signs and symptoms are said to be present when the limits of normal variability are surpassed. Abnormalities may manifest as alterations in amplitude (e.g., excesses or deficits), duration, intensity, timing, and modifiability of physiological events, perceptions, emotions, thoughts, and motor activities. These limits are often arbitrary. Examples include the number of hours of sleep, the intensity of anger, or the extent of mood lability. However, for other experiences the distinction between normal and abnormal is qualitative, not quantitative. However, for some phenomena, any is too much. In mainstream American culture, for instance, any experience of thoughts being broadcast out loud is considered pathological. These signs and symptoms must all be considered in context because exactly what constitutes “normal” varies from culture to culture and from situation to situation. A behavior or subjective experience that may be defined as symptomatic in one context may be perfectly acceptable and within normal bounds in another.

A phenomenon should be considered abnormal only if it seems deviant within the patient's unique culture after its full physiological and environmental context is taken into account, and if it causes personal or interpersonal impairment. Too often, phenomena prematurely mislabeled as psychopathological turn out to be perfectly understandable and nonpathological once the whole situation is appraised. Conversely, some examiners are reluctant to label certain phenomena as psychopathological even when they clearly are, for fear of stigmatizing the patient.

Within cultures, most interpersonal interactions are carefully regulated by tight sets of rules and controls, and constrained by reasonably well defined sets of expectations and acceptable limits. Even slight deviations from these acceptable limits are quickly perceived by lay persons as well as professionals, because behavioral deviances are often experienced as threats. Deviations in amplitude, duration, and intensity can occur in facial expressions, gestures, postures, vocalizations, language, and other expressions of emotion and thought. A small increase in the rate of speech, an intrusion into one person's conversation by another who does not allow proper pauses, a gesture that comes just a bit too close to a face, an excessively rigid or distant stance, or a gaze that is too staring or too avoidant—each signals social insensitivity and alerts the observer to deviant behavior.

Reliability Problems Among the core difficulties in psychiatric evaluation has been that multiple observers may note different symptoms or interpret signs differently when interviewing the same patient. These discrepancies may be caused by differences in the patient's status or in information imparted by the patient from examination to examination, in the observers' definitions of the symptoms or signs in question, and differences in perceiving and interpreting the patient's responses to general presentation or questions within the interview. These three types of reliability problems are called *information variance, criterion variance*, and *observation bias*.

Although good interrater reliability can be achieved for most symptoms of Axis I disorders, this may not hold true for personality disorders or for some specific symptoms. Furthermore, good interrater reliability may occur consistently only under optimal circumstances and may not be as common in clinical practice.

Even when simply responding to direct questions about symptoms, patients may answer differently depending on the interviewer's manner, how the questions are asked, their personal sense of trust or safety, whether they have answered these questions before, the amount of cuing that may signal the desired response fatigue, or a host of other variables.

Most clinicians still rely heavily on their own clinical intuition and subjective responses to patients as part of a diagnostic assessment. Unfortunately, whether accurate or not, these clinical judgments are often based on nonconscious assumptions, comparisons with other patients not well remembered, or distortions based on the clinician's own personal experiences. When the bases for these intuitions can be identified and described clearly, they may prove to be reliable and valid. However, intuitions are often wrong—simple trust in intuition alone is not sufficient. Thus, a clinician's sense that a patient is angry and potentially violent may result from the patient's subtle (but verifiable) body language and tone of voice—or it may represent a countertransference distortion that is not prompted by any observable patient

behavior.

Clinicians often too quickly label behaviors as inappropriate when they fail to appreciate and understand contextual or cultural considerations. Appropriateness depends heavily on context, and definitions of what is proper in a given context may also be highly subjective. Appropriate behavior or clothing in some parts of California may be inappropriate in Boston. A low intensity of emotional expression leading to a clinical description of “constricted affect” may reflect cultural norms or a psychopathological state.

Nonspecific Nature of Signs and Symptoms Until psychiatry discovers reliable diagnostic laboratory tests to define clinical syndromes, the field will continue to construct diagnostic categories based on the clustering of signs and symptoms within specific time frames. Unfortunately, pathognomically specific signs or symptoms rarely exist in psychiatry; virtually all psychiatric symptoms are nonspecific and are usually seen in a many different disorders. Depressed mood, for example, occurs in a wide variety of diagnostic groups—including major depressive disorder, schizophrenia, some personality disorders, and mood disorder due to a general medical condition. Even the so-called first-rank symptoms of schizophrenia described by Kurt Schneider are diagnostically nonspecific and are seen with some frequency in otherwise classic depressive and bipolar disorders.

In general medicine, symptoms not recognized as part of a clearly defined syndrome are often described as being of unknown origin. Thus, a fever that cannot be ascribed to a known disorder, such as pneumonia, is described as a *fever of unknown origin*. Given the nonspecific nature of psychiatric symptoms, it would seem wise to use similar conventions, referring to *hallucinations of unknown origin* or *depressed mood of unknown origin* when a symptom cannot be clearly linked to a well-described syndrome.

Even though individual signs and symptoms may be organized into syndromes and disorders, they often have courses of their own. Thus, in the appearance or the resolution of a disorder, certain associated signs and symptoms may appear very early or may persist after all the others have waned. For example, in the restricting form of anorexia nervosa, excessive exercise is often the first symptom to appear and the last to abate even after dieting has stopped. In some cases, certain signs and symptoms that are commonly associated with a given disorder may fail to appear. Each sign and symptom may have its own pattern and variable response to treatment. In the treatment of schizophrenia, for example, some patients experience rapid resolution of hallucinations but have persistent delusions without ever having any other thinking disorders, whereas others may have no residual hallucinations or delusions but still have prominent thinking disorders.

Sign and Symptom Categories Signs and symptoms have been categorized in a variety of ways: state versus trait, primary versus secondary, and form versus content. The state versus-trait distinction refers to whether the sign or symptom is an enduring characteristic of the person (‘traits”) or time-limited. Phenomena associated with specific Axis I disorders are usually state-associated phenomena. However, some enduring traits may also be symptoms. A person who always worries a great deal, chronically exhibits catastrophic thinking, and feels subjectively nervous in many different circumstances since early childhood may have *trait anxiety*. However, if such symptoms of anxiety are present only during a specific time frame, for example, over a 9-month period in conjunction with a full depressive syndrome, they are best described as *state-related symptoms*. During the acute stages of psychiatric disorders marked by dramatic state characteristics, it is unwise to infer that any of the prominent signs or symptoms are enduring traits, even those usually associated with personality. Thus, a diagnosis of dependent personality *traits* based on an acutely depressed patient's behavior is often incorrect. Similarly, manipulative behavior in the midst of a hypomanic or manic episode should not be considered evidence for enduring manipulative traits unless these behaviors are also present when the mania has clearly resolved.

Distinctions between primary and secondary symptoms have been blurred by varying definitions of these terms. The distinction may refer to causal relations between what is primary and secondary, temporal sequence between the two symptom sets, or inability to more clearly understand the origin of the various symptoms. Basing the distinction between *primary* and *secondary* upon causality implies that the terms cause and effect are clearly understood. In attention-deficit/hyperactivity disorder, for instance, the attention deficit is thought to be primary whereas the hyperactivity is thought to be secondary, caused by the inability to attend. Patients who develop severe dependent personality traits and chronic demoralization only after experiencing several incapacitating psychotic mood episodes might be described as having primary mood disorders and secondary personality disorders. Conceptual models of psychopathology in which some signs and symptoms are seen as restitutive albeit ineffective attempts to cope with more fundamental psychopathological deficits use a primary-secondary model. For example, Eugen Bleuler viewed thought disorder as a primary symptom in schizophrenia, but viewed hallucinations and delusions as secondary symptoms that developed to help the patient cope with the chaos of the primary symptoms. These models must be viewed as hypothetical constructs only and applied with great caution because in the vast majority of clinical phenomena little evidence indicates that one symptom is more primary than another.

Temporal sequence in the appearance of certain symptoms is regularly used as the basis for deciding the primacy of certain symptoms, behaviors, or disorders, such as in trying to determine what is primary and what is secondary when substance abuse occurs in conjunction with depression or anxiety. These differences are not trivial but often have strong treatment implications, because for example, treating a primary mood disorder in a substance-abusing patient (with a long course of medication) is quite different from expecting that with prolonged sobriety a secondary mood disorder will resolve on its own. Furthermore, it is becoming increasingly clear that the presence of certain preexisting psychiatric conditions, such as personality disorders, increases one's vulnerability for the subsequent development of other psychiatric disorders, such as major depressive disorders.

Unfortunately, establishing temporal sequence with any certainty is usually difficult. Although there is a high comorbidity between bulimia nervosa and major depressive disorder, in excess of 50 percent according to some studies, attempts to establish which disorder is primary have been inconclusive. Even with careful historical analysis, major depressive disorder precedes bulimia nervosa, bulimia nervosa precedes major depressive episodes, and the two conditions start concurrently in about equal percentages.

Making simple categorical distinctions between primary and secondary signs, symptoms, and disorders is less important than understanding the contribution of each element as a thread in the evolution and development of a given clinical presentation. From this perspective, each element can be viewed as dynamically affecting the appearance, manifestations, and course of the others and as exerting its own influence on the pathogenesis and treatment of the specific syndromes and associated disorders. For example, the onset, specific nature, and independent time course of sleep disturbances and insomnias may prove to be important in understanding the appearance, manifestations, and sequence of other prominent signs and symptoms of the mood disorders of which they are a part, and in their treatment.

This view is particularly important because despite the excellent conceptual contributions made by categorical diagnostic systems such as the American Psychiatric Association's DSM-IV, distinctions are often fuzzy in clinical practice and comorbidity among so-called categorically distinct disorders is often the rule rather than the exception. For example, data from the National Comorbidity Study show that 14 percent of the population experience three or more comorbid psychiatric disorders. In such individuals the dynamic interactions and mutual influences of various signs and symptoms, and their biological underpinnings, becomes impossible to disentangle.

Furthermore, the categories that currently comprise DSM-IV are not going to be the last word in the evolving history of psychiatric diagnosis. Recent studies show that psychiatric signs and symptoms may be usefully grouped into psychotic syndromes that differ in some respects from current DSM-IV categories. In a large family study of probands with broadly defined schizophrenia and affective illness and their first degree relatives, using a sophisticated statistical technique called latent class analysis Kenneth Kendler and his colleagues found six classes of psychosis including classic schizophrenia, major depression, schizophreniform disorder, bipolar-schizomania, schizodepression, and hebephrenia. While these classes bore substantial resemblance to current or historical nosological constructs, several of them differ from DSM-IV nosological constructs. In another study the three factors ordinarily associated with symptoms of schizophrenia, representing positive, negative and disorganized symptom domains, were found not to be specific to schizophrenia but were found in other schizophrenia-spectrum psychoses and in nonschizophrenia-like psychotic conditions as well.

Context Signs and symptoms are usually not static entities; depending on the context, they often vary in intensity or even in their existence. The depressed mood of a melancholic depression may persist regardless of the external situation whereas the depressed mood of milder depression may vanish completely during certain situations—including a psychiatric interview—only to reappear at other times. Signs and symptoms that occur only in specific settings or with certain internal states are referred to as *state-dependent*. For example, certain hallucinations or memories may be present only during states of drug or alcohol intoxication; in some patients, hives may erupt as a psychophysiological response only during states of anger. Interpersonal context is also important. Some persons become violent only when involved in sadomasochistic relationships or in certain group settings, such as within adolescent gangs. In gangs, social pressures for conformity and expectations for aggressive behavior may provoke or release pathological behaviors that might otherwise never be expressed by gang members individually.

Problems and Impairments Beyond the classic signs and symptoms of psychiatric disorders, recent attention has focused on the problems and impairments that psychiatric signs, symptoms, and disorders generate in affecting specific role functions and in causing social and economic burdens for the patient and others. These problems and impairments often cut across traditional sets of signs and symptoms of which categorical diagnoses are composed, affecting, for example, basic abilities to care for oneself and one's family, marital functioning, child rearing, wage earning, school performance, and social behavior. They constitute the issues with which

patients and families contend, and they need to appear on the problem lists that treatment plans and specific interventions target. Studies reveal that the impairments imposed by major depression are considerable with regard to physical functioning, role limitations, and social functioning. Problems such as violent temper outbursts, sexual aggression, or lack of job skills, which may impair role functioning in several spheres, must be directly addressed in their own right regardless of the associated DSM-IV diagnoses. These impairments enter determinations of ratings for Axis V of DSM-IV, which addresses the global assessment of functioning, and are of considerable importance in evaluating treatment outcomes. Table 8-1 lists some examples of critical impairments that have been recognized as often requiring urgent or intensive levels of care.

Anxiety
Assaultive behavior
Compulsions
Concomitant medical condition
Delusions (nonparanoid)
Delusions (paranoid)
Dissociative states
Dysphoric mood
Eating disorder
Fire setting
Hallucinations
Homicidal thought/behavior
Inadequate health care skills
Manic thought/behavior
Medical risk factor
Medical treatment noncompliance
Mood lability
Obsessions
Phobia
Physical abuse perpetrator
Psychomotor retardation
Psychotic thought or behavior
Running away
Self-mutilation
Sexual trauma perpetrator
Substance abuse
Suicidal thought or behavior

Adapted from Goodman AL, Brown JL, Delia PH: Managing Managed Care II: A Handbook for Mental Health Professionals, ed 2. American Psychiatric Press, Washington, DC, 1996.

Table 8-1 Signs, Symptoms, and Impairments Requiring Immediate Care

Need for a Comprehensive Perspective A psychiatric disorder may be characterized by disturbances involving a wide variety of areas in the patient's life, including the biological, psychological, behavioral, interpersonal, and social spheres. In practice, common psychiatric syndromes often manifest in each of these dimensions. (Table 8-2). Viewing the patient from multiple perspectives, using the so-called biopsychosocial model (similar to the multiaxial approach of DSM-IV) enables clinicians to consider psychopathology and its effects on a patient's life in the broadest possible manner. To illustrate, Figure 8-1 lists some clinical hypotheses commonly used by clinicians as they link collections of signs and symptoms into syndromes and consider the treatment options that would logically follow.

Biologically derived hypotheses
1. As an organic mental disorder
2. As a mood disorder
3. As a nonaffective functional psychosis
4. In relation to the abuse of drugs or alcohol
5. As a disorder other than those listed in 1 through 4, above
Psychodynamically derived hypotheses
1. As being related to personality style
2. As being related to a precipitating event and its dynamic meaning
3. As a manifestation of unresolved grief
4. As a developmental crisis
5. As being related to ego functioning and related psychodynamic issues
Sociocultrually derived hypotheses
1. As resulting from the nature and the social effects of stressful life events
2. As being related to the extent, the nature, and the accessibility of social support
3. As being related to definitions of and responses to breakdown in the sociocultural grouping
4. As being related to the patient's motivation, treatment goals, and the dynamics of the entry process to help seeking
5. As social communication
Behaviorally derived hypotheses
1. As disordered thinking, feeling, or acting that is the result of specific antecedent events
2. As disordered thinking, feeling, or acting resulting from reinforcing consequences of the behaviors
3. As disordered thinking, feeling, or acting in response to sociocultural and biological events
4. As a deficit of behaviors (rather than as disordered behaviors) in the areas of thinking, feeling, and acting
5. By an analysis of areas of effective functioning

Adapted from Lazare A: Hypothesis testing in the clinical interview. In Outpatient Psychiatry: Diagnosis and Treatment. Lazare, editor. Williams & Wilkins, Baltimore, 1979.

Table 8-2 Common Current Clinical Hypotheses Used to Assess Signs and Symptoms: Ways of Understanding the Patient's Problems

FIGURE 8-1 Biological, psychological, and social forces interact and effect the psychiatric health of a person. (Adapted from Richmond JB, Lustman SL: Total health: A conceptual visual aid. J Med Educ *29:*23, 1954.)

Because the amount of information gathered in a thorough assessment of a psychiatric disorder is potentially overwhelming, clinicians often tend to limit their fields of vision and appreciate only part of the available information; the clinician's theoretical orientation and other personal and cultural factors also limit what is perceived. Research has demonstrated that clinicians tend to perceive primarily those signs and symptoms that are most in accord with their own opinions and with the tools they have to treat psychiatric disorders, a phenomenon known as *concept-driven perception*. The theoretical biases of clinicians seem to be related both to the microcultures of their training programs and to their own personality traits. Such differences may lead one clinician to see a major mood disorder, to be treated with medication, where another sees a pervasive personality problem with dysthymic disorder, to be treated with psychotherapy, and to use different technical terms to label the same phenomenon. A psychodynamic psychiatrist might see psychomotor retardation where a neuropsychiatrist sees bradykinesia; a psychodynamicist might see depressed mood and muted speech where a neuropsychiatrist sees mask-like facies and aprosodic speech; the psychodynamicist might see ruminative thought where a neuropsychiatrist sees forced thinking; a psychodynamicist might see a grimace where a neuropsychiatrist sees a tic. Given the extent to which words themselves shape our concepts of reality, the consequences of using these different labels for very similar phenomena may be significant. Figure 8-1 illustrates concept-driven perception in which each clinician who adheres to a prominent contemporary point of view perceives only some of the potentially available phenomena related to a psychiatric disorder. Although there is overlap, each observer also perceives information that other observers do not notice. Also, some information that may be highly relevant to diagnosing or treating the disorder may be overlooked by all the observers.

The intermittent nature of many psychiatric signs and symptoms; the potential unreliability, selective recall, and false recall of patients and others in reporting symptoms and events; differing interpretations of elicited information or observations; and subjective theoretically driven biases that influence the clinician's perception of signs and symptoms, all contribute to potential errors in data collection. To help guard against misinformation and simplistic understandings and formulations, wherever possible complete assessment of a psychiatric patient requires consultation with family, friends, co-workers, and other professional observers to enrich the history and to provide supplemental observations of the patient over time.

SOMATIC MANIFESTATIONS OF PSYCHIATRIC DISORDERS

Virtually all Axis I, symptom-based disorders and most other psychiatric disorders are characterized by disturbances in at least some basic physiological functions. Although frequently nonspecific in nature, the severity of these somatic signs and symptoms provides markers as to the amount of biological disruption seen in the disorders that cause them. Furthermore, somatic symptoms can also cause exacerbations of some disorders. If untreated, these processes can create destructive feedback loops in which the disorder causes symptoms, which then exacerbates the disorder, which increases the symptoms, and so on. For example, the insomnia of manic or hypomanic episodes will, if untreated, cause a marked worsening of the mania; similarly, the weight loss of anorexia nervosa causes starvation effects such

as a preoccupation with food, thus exacerbating one of the hallmark features of the underlying disorder.

Sleep Disturbances Abnormalities of sleep may manifest in the amount, quality, and timing of sleep, as well as by the presence of abnormal events during sleep.

Insomnia is usually defined by its subjective component as the sensation of sleeping poorly. Most but not all patients complaining of insomnia will demonstrate some sleep abnormality if examined in a sleep laboratory. Insomnia is a common, often chronic symptom or sign of many different psychiatric disorders and conditions including substance abuse, depressive disorder, generalized anxiety disorder, panic attacks, manic episodes (in which the diminished sleep does not always provoke a complaint), and acute schizophrenia. It may also occur as a consequence of aging or as a symptom or disorder not associated with other psychopathology. Insomnia may also result from the ingestion of substances that alter the normal sleep-wake cycle, including alcohol or stimulants, and by the discontinuation of sedative-hypnotics. Although much attention is often paid to distinguishing patterns of insomnia, such as difficulty falling asleep versus middle or terminal insomnia (early-morning awakening), or linking specific patterns linked to specific conditions (e.g., major depressive episode with melancholic features and associated terminal insomnia) the clinical utility of these distinctions is unclear.

> A 62-year-old woman presented with new-onset primary insomnia. In spite of an extensive medical work-up by her primary physician, and equally extensive psychiatric examination, no other medical, psychiatric, iatrogenic, or substance-related problems usually causing or associated with insomnia could be identified. She was treated for her symptoms with hypnotic agents and experienced partial improvement. The insomnia persisted for more than a year before spontaneously remitting as mysteriously as it initially appeared.

Hypersomnia, characterized by either excessive nighttime sleep or excessive sleepiness during the day, is less common than insomnia. However, it too may reflect a number of different pathological states. Some depressed patients, especially those with a history of mania or hypomania, may exhibit hypersomnia. Hypersomnia may also be seen during amphetamine withdrawal, excessive use of sedatives or anxiolytics, or in conjunction with a variety of medical disorders. In *narcolepsy*, the patient has sudden attacks of irresistible sleepiness, a symptom that may be part of a broader syndrome that includes *cataplexy* (sudden attacks of generalized muscle weakness leading to physical collapse in the presence of alert consciousness), *sleep paralysis* (waking from sleep with a sensation of being totally paralyzed that may persist for minutes), and *hypnagogic hallucinations* (vivid visual hallucinations that occur at the point of falling asleep). Narcoleptic attacks are often precipitated by unusual states of arousal (e.g., cataplexy may immediately follow unrestrained laughter or orgasm). Daytime sleepiness may reflect *sleep apnea*. In this disorder, middle-aged patients demonstrate severe snoring—often first reported by their bed partners—and periods when breathing stops. The condition results from soft-palate abnormalities that cause intermittent airway obstruction throughout the night; patients awaken repeatedly to find themselves gasping for air. Associated daytime fatigue is common in sleep apnea. Periodic hypersomnia also occurs in the *Kleine-Levin syndrome*, a condition that typically affects young men, in which periods of sleepiness alternate with confusional states, ravenous hunger, and protracted sexual activity. Intervals of days, weeks, or months may pass between these episodes. *Sleep drunkenness* is characterized by excessive sleep and great difficulty awakening completely, with confusion and motor incoordination soon after arising. Excessive daytime sleepiness may also occur secondary to abnormalities to the brainstem, hypothalamus, or thalamus.

Sleepwalking disorder (somnambulism) and *sleep terror disorder* are two sleep disorders characterized, respectively, by aimless wandering with incomplete arousal, and by acute anxiety and physiological arousal without awakening. Although both disorders typically begin in childhood, sleepwalking may be also be initially precipitated by some psychotropic medications.

Nightmares are a common complaint, often associated with traumatic events, anxiety disorders, and mood disorders, but are also not uncommon as an occasional event in otherwise healthy individuals. Vivid dreams and nightmares may also be an adverse effect of medications.

> A 24-year-old Peace Corps volunteer started to experience frequent, vivid, and disturbing nightmares shortly after arriving at her placement in South America. She mentioned these symptoms to other volunteers with whom she was traveling and learned that several of them were also having similar nightmares. All had recently started courses of mefloquine (Lariam) for malarial prophylaxis, and one had been forewarned to expect vivid nightmares as a commonly occurring adverse effect of this medication.

Sensory symptoms during sleep, typically described by patients as peculiar feelings in their legs causing an irresistible need to move around, are characteristic of *restless legs syndrome*. The motor abnormality of repetitive myoclonic jerking of the legs, awakening both patients and their partners, is known as *nocturnal myoclonus*.

Appetite and Weight Disturbances Aside from the anorexia of medical illnesses, especially in their later stages, loss of appetite is most commonly seen in depressive disorders, grief, and anorexia nervosa, and is also commonly seen in conjunction with significant anxiety. Anorexia is often accompanied by changes in taste (e.g., foods begin to taste bitter or flat or have an unpleasant aroma). Patients with eating disorders may resist hunger to restrict food intake in order to achieve a physiologically unrealistic low weight. *Hyperphagia* (increased appetite) occurs in some depressed patients, both with and without a history of mania bipolar I or bipolar II disorder. Binge eating, of up to several thousand calories per episode, may occur as an attempt to self-soothe and emotionally self-regulate during times of increased tension and anxiety, and as a key feature of bulimia nervosa or of binge-eating disorder. Increased appetite may be seen, albeit rarely, in some hypothalamic disorders or in bilateral temporal lobe dysfunction, such as the Kluver-Bucy syndrome, in which it occurs in association with emotional placidity, hypersexuality, hyperorality, and other symptoms.

Other Ingestive Disturbances—Substance Abuse The use of chemicals such as alcohol, opioids, marijuana, cocaine, amphetamines, hallucinogens, and other drugs to change one's inner emotional and tensional state is well known. Used at inappropriate times and in excess, such ingestions may culminate in various syndromes of abuse and dependence and may complicate other psychiatric disorders. Use of nonnutritious substances for their presumed psychoactive effects has even been reported to occur in the wild among elephants and baboons.

Energy Disturbances Normal energy levels vary considerably among people. Some people fatigue easily and are perceived by themselves and others as having weak constitutions, whereas others appear to have almost boundless energy and much less need for sleep.

Fatigue is a common nonspecific symptom that occurs in both medical and psychiatric disorders. It is also frequently seen as an unexplained complaint in primary-care practices; in one study 24 percent of patients complaining of fatigue received no medical or psychiatric diagnosis. Historically, fatigue not caused by another disorder, typically in association with “nervousness,” has been described by terms such as *asthenia, neurocirculatory asthenia, neurasthenia*, and *psychasthenia*. Consistent with this tradition, many fatigued patients, having been labeled depressed or neurotic by their physicians, are referred to psychiatrists after routine work-up has ruled out anemia, hypothyroidism, sleep apnea, and other frequent somatic causes.

Recently, patients with primary complaints of tiredness have been most commonly diagnosed as having *chronic fatigue syndrome*, incorrectly labeled *Epstein-Barr viral syndrome*. This disorder is characterized by fatigue lasting months to years, typically beginning soon after a viral syndrome. In addition to the fatigue, chronic fatigue syndrome is characterized by myalgias and cognitive changes, such as forgetfulness and poor concentration. Although controversy still exists on the extent to which cases of chronic fatigue syndrome represent discrete postviral diagnostic syndromes, mislabeled cases of depression, or modern versions of neurasthenia, evidence continues to mount to suggest that this syndrome is a discrete postinfectious entity, and is not simply a variant of or a disguised mood or anxiety disorder.

> A highly accomplished, energetic 40-year-old woman was referred for psychiatric consultation after her physicians were unable to offer a definitive physiological diagnosis despite extensive medical work-up following the acute onset of profound fatigue occurring in the wake of a mild viral illness. This fatigue caused her to be totally incapacitated and bedridden for many months, and left her feeling helpless and distraught. Her condition was exacerbated by even small amounts of alcohol. Two independent psychiatric evaluations concluded that the patient had no psychiatric disorder. The patient developed this syndrome after jogging near Lake Tahoe. Epidemiological studies were found that reported several other cases of profound fatigue originating in this particular area after vigorous exercise. She was treated for symptoms of fatigue and gradually made an almost full recovery after about 2 years.

Disturbances in Sexual Drive As with energy, the normal range of sexual drives is great. Some individuals are naturally lusty, whereas others have limited sexual desire. Diminished sexual drive with impotence or decreased libido is seen in a wide variety of neurological, metabolic, and other somatic syndromes. Among neurological disorders, complex partial seizures are commonly associated with hyposexuality, occurring in 50 percent of patients. Psychiatric disorders known for diminished sexual drive include depressive disorders, schizophrenia, substance-related disorders, and marital conflict. Diminished libido, impotence, and anorgasmia are also common sequelae of many pharmaceutical agents, including psychotropic medications.

A 24-year-old woman, treated for depression with a selective serotonin reuptake inhibitor by her primary physician, developed profoundly diminished sexual sensation and anorgasmia in the course of treatment. In order to increase her sexual feelings as her dating and sex life resumed, she stopped taking her medication without first discussing this decision with her physician. Several months later, when her depression returned, she revealed what she had done and was effectively started on a different class of medication that lacked these adverse effects.

Increased sexual activity may be seen in some neurological, substance-induced, and psychiatric disorders. Manic patients frequently exhibit hypersexual interests and behaviors to an unusual degree compared to their euthymic interests and behaviors. Hypersexuality is occasionally seen in conjunction with epileptic syndromes or in patients who have suffered diencephalic injuries.

Altered sexuality, including fetishism, sadomasochism, and pedophilia, may be seen as isolated psychiatric syndromes. In individuals whose previous sexual behaviors were within the bounds of social propriety for their groups, inappropriate sexual behaviors may signal early brain disease or psychosis. Cross-dressing may occur in transvestites, transgenderists, transsexuals, or occasionally in psychiatric conditions such as schizophrenia.

A 36-year-old former Marine pilot became psychotic several years after being discharged from active duty. During the next several years while being successfully treated for schizophrenia, he started to cross-dress in a garish, almost cartoonish manner, and sought hormone treatment in Mexico to reduce his secondary male sex characteristics and to foster the development of breasts. Although he considered undergoing a genital amputation and having a plastic surgeon build a vagina-like vault, he hesitated to do so, describing how he really enjoyed shocking the men he took home from bars by revealing that he had a penis. On several occasions these men, outraged and humiliated, physically attacked him and hurt him badly.

Appearance Studies show that clinicians often formulate an initial psychiatric diagnosis within 30 seconds of seeing a patient. Although approximately half of such initial impressions prove to be incorrect, the remainder are validated by psychiatric history and mental status examination, revealing just how much information is communicated by appearance and body language.

Among the physical disorders whose appearance suggest coexistent psychiatric conditions are acromegaly, Cushing's disease, Down syndrome, systemic lupus erythematosus, fetal alcohol syndrome, Klinefelter's syndrome, and Wilson's disease. The general appearance of the skin may suggest the presence of occult psychiatric problems. Skin condition and tone may reveal hypervascularity and ruddiness suggestive of alcoholism, abscesses are indicative of hypodermic needle abuse, tattoos are indicative of certain group affiliations, or weathering and wasting indicative of self-neglect and malnutrition. Healed scars on the wrists and arms suggest a pattern of self-mutilation from depression, personality disturbance, or both. Patchy baldness, especially in conjunction with torn or infected cuticles, indicate *trichotillomania*, a syndrome of compulsive hairpulling. Psychophysiological symptoms reflecting psychiatric disturbance include urticarial reactions and neurodermatitis, the latter resulting in part from self-excoriation, destructive scratching secondary to compulsions, and unrelenting sensations of discomfort.

Examination of the head and neck may reveal exophthalmos or puffy eyelids suggesting thyroid disease, the marked pupillary dilation of anxiety or amphetamine abuse, the miosis of narcotic abuse, the abnormal pupillary pigments of Wilson's disease, the salivary gland enlargement of bulimia nervosa, or the necrosis of the nasal septum seen in cocaine abuse, among other signs. Frequent sighing is a common respiratory sign in depression. Simple sighing must be distinguished from respiratory dyskinesia in psychotic patients who have been treated with neuroleptic agents. The latter may occur as an acute dyskinesia resulting from antipsychotic medication or it may be a late manifestation and component of tardive dyskinesia.

Deviant appearance is quickly perceived by lay persons as well as professionals, and may contribute to the frightened and stigmatizing social withdrawal by strangers and acquaintances so often experienced by psychiatric patients. Akathisia and dystonic movements and Parkinsonian shuffling gait in patients taking neuroleptic agents, as well as the dilapidated, unkempt appearance of some psychiatric patients can immediately signal psychiatric patient status to observers. The term *Diogenes syndrome* has been used to describe old people who have a filthy personal appearance that demonstrates severe self-neglect about which they have no shame.

DISTURBANCES IN THINKING

Normal Thinking *Thinking* refers to the ideational components of mental activity, processes used to imagine, appraise, evaluate, forecast, plan, create, and will. Most thought involves complex rules that are best approximated by *fuzzy logic* decision-making algorithms that use neural net technology, increasingly applied by scientists and engineers in situations where all-or-none, black-or-white thinking does not apply, and where multifaceted, contradictory, and competing possibilities and biases are the rule. Most of what is known about thinking derives from the study of language as the product and reflection of thought; yet a great deal of thinking takes place preverbally and nonverbally. Thinking occurs in images, music, kinesthetic sensations, and in symbols other than linguistic ones. Attempts to transmit preverbal and nonverbal thought using only words are frustrating and unsatisfactory. Creative artists have considerable difficulty describing the inner states of tension and inchoate awareness from which ideas are distilled.

Ordinary thought is far from logical. Streams of conscious thought are intruded upon by competing thoughts and associations, and by outside stimuli, and attention is easily distracted. Ordinary conversation is marked by recurring asides, interruptions, delays, and the loss of ideas. Decisions are often made on the basis of very few cues and inadequate evidence: people jump to conclusions and beliefs that are not supported by evidence are zealously held. Thinking in stereotypes is more common than thinking in logical categories; from an evolutionary perspective, thinking in stereotypes and by approximation has probably been more adaptive than thinking in strictly defined categories. This tendency helps to account for clinicians' tendencies to make diagnoses by approximation and intuition based on prototypes, and to feel less comfortable using formal lists of criteria found in statistical manuals, such as DSM-IV.

Individuals vary greatly in their predominant *cognitive styles*, and a given individual's style of thinking will also shift considerably from time to time. *Cognitive style* is one's predominant manner of information processing and decision making, the particular biases and distortions that thinking processes make by means of augmenting, elaborating, or minimizing incoming information, and the extent to which careful and deliberate logic versus intuition versus thoughtless, anxiety-induced impulsivity is used to guide the decision-making process. A particular cognitive style may come to dominate a person's repertoire so completely as to interfere with the flexible, adaptive responses required to deal with the usual variety of daily needs. An *obsessional style* of thinking is marked by attention to detail and hypervigilance concerning the possible implications of a particular thought or event. This may take the form of preoccupation with strict adherence to established rules, values, or beliefs. An obsessional style may be highly adaptive in certain situations, as in professions requiring meticulous detail such as library science, computer programming, and surgery. However, excessively rigid obsessionality may be maladaptive, as when someone scrupulously sticks to the rules even when such adherence is self-destructive and short-sighted. A *hysterical* style of thinking is characterized by global, diffuse, impressionistic, emotionally laden evaluations of situations where lack of attention is given to details and nuances. This style is poorly adaptive to detail-oriented work, but may be useful in the arts, certain aspects of marketing and sales, and in some social situations.

Two brothers watched *Kramer vs. Kramer*, a film about the dissolution of a marriage and the struggles of caring for the couple's only child after the divorce. In discussing the movie afterwards, one brother, a business lawyer specializing in writing complex contracts, remembered the plot in detailed linear sequence. However, he had great difficulty explaining the underlying feelings and motivations of any of the characters. The other brother, an artist, could not remember details of the plot and confused the sequence of scenes. However, he could easily and accurately explain the relationships between the characters and their feelings.

Types of Thinking Because of the different ways in which normal and abnormal thinking expresses itself, attempts have been made to subtype thinking by the extent to which logical versus nonlogical thought is utilized. Freud's division of thought into *primary* and *secondary process thinking* is still widely used.

PRIMARY PROCESS Primary process thinking, the more primitive type, is typically seen in dreams, but also prominent in young children and in psychotic states. This type of thinking disregards logic, permits contradictions to exist simultaneously, disregards the linear notion of time, and is dominated by wish and fantasy. It uses symbol, metaphor, imagery, condensation, displacement, and concretism in its organization, creating the jumbled and incoherent style of thinking characteristic of dreams. Primary process thinking represents what has been loosely and metaphorically called *right brain thinking*, associated with visual images and creative thought.

SECONDARY PROCESS Secondary process thinking is characterized by logic. Unlike primary process thinking, the secondary process uses linear notions of time, clearly delineated abstract categories, and deductive rules of logic. The ability to think abstractly and to think in detail about future plans is characteristic of secondary process thinking. Normal secondary process thinking is also characterized by predictability, coherence, and redundancy. Words, vocal inflections, and gestures provide important contextual cues and create a sense of overall coherence to the communication. Ideas follow one another in a sequence that is understandable to

the listener.

A non-Freudian typology of thought divides thinking into three types: *fantasy thinking, imaginative thinking*, and *rational* or *conceptual thinking*. Fantasy thinking allows the person to escape from, or deny, reality, and can be seen in normal as well as pathological thinking. Everyone occasionally uses fantasy thinking when daydreaming. Some dissociative and psychotic phenomena illustrate the most pathological manifestations of fantasy thinking. Imaginative thinking merges fantasy and memory in order to generate plans for the future. Rational or conceptual thinking uses logic to solve problems.

Regardless of how one categorizes thought, people can fluidly shift from linear, secondary process or rational thought to fantasy, primary process or nonlogical thought, as in the free associative method used in psychoanalysis. During this process, individuals willfully surrender the controls that maintain secondary process thinking and switch to the less controlled modes of primary process thinking in which thoughts are loosely associated by emotional associations or based on peripheral, concrete, coincidental, loosely similar, or trivial aspects of a thought. Additionally, the fact that increases in primary process thinking can be induced in normal people under experimental conditions or with fatigue suggests that more primitive thought processes, such as those seen in psychosis, are usually inhibited by higher-order processes, and that their appearance may be release phenomena; that is, nonlinear or psychotic thinking may indicate the functional absence of those overriding control systems that ordinarily sift, evaluate, and regulate the form and flow of thought before it reaches consciousness.

THOUGHT DISTURBANCES

Flow and Form Disturbances Because the underlying processes that govern thought are not understood, current systems for classifying thought abnormalities are primarily descriptive. Conventional classification separates form and flow from the content of thought. Yet, many types of abnormal thinking include both form and content abnormalities. Thus, whereas delusions are usually classified as thought content disturbances, they are also marked by formal abnormalities such as rigidity and imperviousness of thought to external influence or to information that clearly contradicts the delusional idea. Although *formal thought disorder* typically refers to marked abnormalities in the form and flow or connectivity of thought, some clinicians use the term broadly to include any psychotic cognitive sign or symptom.

As with energy and sexuality, normal variations in the flow and form of thought are considerable. For some people, thinking appears to be effortless—rapid and productive, exhibiting linear, goal-directed thoughts and creativity, with digressions and occasional leaps but always controlled and comprehensible. For others thinking is a difficult exercise—a slow, painstaking process with low output compared to other people, or scattered, with difficulty staying with a topic or finishing a single thought. Most people experience mixtures of these extremes. Disturbances in the flow and form of thought occur with regard to rate, continuity, control, and complexity.

Thinking can be unusually slow or accelerated. Slowed (or retarded) thought, (e.g., as seen in depression), is typically goal directed but characterized by little initiative or planning. Patients experiencing retarded thought often describe feeling that even simple thought requires monumental effort, as if molasses were cluttering their thinking. These difficulties are expressed as slowness in decisionmaking and as long latency of response, increased pause times when speech is initiated and during speech. *Thought blocking*, seen in schizophrenia, is experienced as the snapping off or as a sudden break in a train of thought, as if a wall suddenly came down interrupting thinking (and speaking) in midsentence. To an outside observer, without further explanation from the patient, thought blocking may appear identical to *thought withdrawal*, a disturbance in the control of thought in which the patient feels as if some alien force has intentionally withdrawn the thoughts from consciousness. The patient's further description and explanation of the inner experience is necessary to distinguish these two symptoms.

> A 26-year-old man with paranoid schizophrenia frequently broke off his conversation in mid-sentence. To the puzzled examiner he explained that the mysterious force that controlled him with a computer chip in his brain closely monitored his thought and speech, and would shut him down whenever it was concerned that he might inadvertently say something that was classified information.

Accelerated rates of thinking, typically accompanied by fast talking, can be seen as a normal variant. Rapid rates of speech, influenced heavily by cultural and situational factors, only sometimes reflects truly rapid thought. (For example, it is not at all clear that New Yorkers, who characteristically speak more quickly than people from some other cities, actually think at a faster rate. Similarly, auctioneers and some radio and television announcers can speak with astonishing rapidity, probably reflecting both innate capacities as well as learned psychomotor skills.) *Pressure of speech*—speech that is rapid, excessive, and typically loud—is characteristic of mania (or hypomania), stimulant intoxication and, occasionally, anxiety. *Flight of ideas* occurs when the flow of thought increases to the point where the train of thought switches direction frequently and rapidly. The associative links between conceptual topics during flight of ideas are comprehensible to the listener, sometimes with considerable effort! Listening to a flight of ideas that is not overwhelmingly fast can be both a dizzying and enjoyable experience for the listener, as exemplified by the successful performance style of certain contemporary comedians, notably Robin Williams.

Continuity Disturbances in the continuity of thought may take several forms. In *circumstantiality* the flow of thought includes many digressive turns and associations, often including a great deal of unnecessary detail. Transcripts of circumstantial thought or speech are marked by multiple commas, subclauses, and parenthetical asides. Nonetheless, in circumstantial thought or speech the speaker eventually returns to the point that was initially intended without having to be prompted by the listener.

In contrast, in *tangentiality*, the person's thought wanders further and further away from the intended point, without ever returning, so that the person may not even remember what the original point was supposed to be. In *vorbeireden*, a form of tangentiality, the person talks past the point and never quite gets to the central idea. Tangentiality is a mild form of *derailment*, in which there is a breakdown in associations. *Loose associations* exemplify more severe derailment, in which the flow of ideas is no longer comprehensible to the listener because the individual thoughts seem to have no logical relation to one another. Loose associations are classically a hallmark feature of schizophrenia. In extreme cases, the associations of phrases and even individual words are incomprehensible, and syntax—the rules of grammar by which phrases are organized into sentences and words into phrases—may be disrupted. *Word salad* describes the stringing together of words that seem to have no logical association, and *verbigeration* describes the disappearance of understandable speech, replaced by strings of incoherent utterances.

Clang association refers to a sequence of thoughts stimulated by the sound of a preceding word. For example, a patient with mania said, “I'll kill with a drill or a pill—God, I'm ill—what swill.” In *echolalia* the patient repeats a sentence just uttered by the examiner. Repetition of only the last uttered word or phrase is called *palilalia*, a symptom found most often in patients with chronic schizophrenia.

Perseveration and *stereotypy* are two other associative abnormalities in which the flow of thought or speech appears to get stuck. In *perseveration*, a sentence or phrase is repeated, sometimes several times over, after it is no longer relevant; perseveration is commonly seen in delirium and other organic mental disorders. *Stereotypy* refers to the constant repetition of a phrase or a behavior in many different settings, irrespective of context.

Disturbances in the control of thought include *delusional passivity* and *obsessional thinking*. In *delusional thought passivity*, patients experience their own thoughts as being under the control of other forces. Thought passivity may take several forms: in *thought insertion* thoughts are experienced as having been placed within the patient's mind from the outside; in *thought withdrawal* thoughts are whisked out of the mind; in *thought broadcasting* patients experience their thoughts as escaping their minds to be heard by others. These experiences are often combined with specific delusions of control, seemingly to explain the passivity experiences. Several of these phenomena were included by Kurt Schneider among the first-rank symptoms of schizophrenia. Today, these symptoms are viewed more broadly as nonspecific psychotic symptoms, and are no longer considered to be pathognomonic of schizophrenia.

> A 40-year-old man who had been living in a state hospital for many years described how he was the outer shell in a set of nested beings. A homunculus-like figure in the center controlled another being surrounding him, and the patient himself was simply the outer wrapping. The inner homunculus made all the decisions, and pulled all the strings, so that the patient was simply a passive recipient of his thoughts and of the instructions that ordered him to carry out each and every act in his life.

Obsessional thinking is stereotyped, repetitive, persistent thinking that is recognized as one's own thoughts. In contrast to patients with delusional thought passivity, obsessional patients do not experience their thoughts as being controlled by outside forces. Nonetheless, they experience only partial control over the obsessional thoughts. They can, with great effort, stop thinking the obsessional thoughts but cannot prevent them from recurring. Thus, characteristic of obsessions is the subjective experience of compulsion, the resistance to it, and the preservation of insight. As bizarre as some obsessions are, patients know that these thoughts are irrational and their own. At times, obsessions may be pervasive enough to dominate the patient's consciousness. Obsessions may be simple—a sequence of words—or elaborate—such as enumerating the possible consequences of a past behavior and elaborating a cascading sequence of typically catastrophic events. Typical obsessional themes in obsessive-compulsive disorder involve preoccupations with dirt and contamination, fear of harming others, symmetry, and those related

to health and appearance.

> A 24-year-old woman was preoccupied with the fear that she would be contaminated by germs that were all around her. These thoughts were inescapable, and led her to narrow her range of activities considerably, to the point of being nearly housebound. She had to comply with a series of ritualistic acts to ward off contamination in her house.

Obsessional thoughts are usually seen in conjunction with *compulsive behaviors*, which are rituals linked to the obsessions, typically constructed to undo the effects of the thought.

The most prominent disturbance of thinking complexity is an impaired capacity to think abstractly. Abstract thinking is the ability to assume a mental set, to keep simultaneously in mind all the aspects of a complex situation, to move from feature to feature as indicated by the situation, and to abstract common properties. Complex thinking also concerns the ability to simultaneously consider many different, vague, and subtle aspects of situations; to appreciate differing and contradictory points of view; and to integrate these multiple dimensions to form opinions that are marked by differentiatedness and nuance. Normal individuals vary greatly in their abilities to engage in abstract thinking—geniuses in mathematics and theoretical physics leave most mortals far behind. *Concrete thinking* is a disturbance in the ability to form abstract concepts, generally illustrated by literal-mindedness and the inability to abstract the commonality of members of a group, for example, the fact that a flea and a tree are similar in that they are both living things. Concrete thinkers seem unable to free themselves from the literal or superficial meanings of words. Concrete thinkers may be more prone to prejudice and stereotypic thinking, more likely to manifest unidimensional or all-or-none reactions to complex situations. Concrete thinking can be seen in individuals with lower intelligence, organic mental disorders, and schizophrenia. Patients with schizophrenia may also exhibit highly selective disturbances of abstraction.

THOUGHT CONTENT

The normal content of thought, the buzzing, booming stream of consciousness that constitutes the stuff of everyday life, is composed of awareness, concerns, beliefs, preoccupations, wishes, and fantasies occurring with various degrees of clarity, vividness, differentiation, imagination, and strength. Normal thought is often illogical, containing many beliefs and prejudices that, although clearly contradictory, are nevertheless held with passion and conviction.

Belief systems are the scaffolding of thought, chains of impressions, and expectations around which plans and behaviors are organized. Belief systems may be attitudinal, setting general expectations and biases about the world that inform how incoming information is processed; examples of belief systems are optimism, pessimism, and paranoia. Some beliefs are effervescent and fleeting, whereas others are pervasive, tenacious, enduring, and influential. Some beliefs are unique and private, whereas many are shared by others.

Imaginative fantasy is an important component of normal thought. The vivid, eidetic imaginations of young children can produce fantasies in which children become fully immersed, almost as if in hypnotic states. During latency many children develop imaginary companions as playmates. In later years, imaginative thinking in which previously separate streams of thought playfully interact with one another to produce new ideas may be the essence of the creative reverie. Artists, writers, and creative scientists may retain access to these forms of thinking more readily than others. Meditative states of mind may facilitate the emergence of imaginative insights. Such thinking may also occur in dreams. Intrusive reveries are normal and common components of the usual adult stream of consciousness. During periods of specific deprivation, such as starvation or sexual deprivation, elaborate wish-fulfilling daydreams frequently occur.

Ideas are the contents of the stream of thought. Those that are consistent with one's sense of self, compatible with the individual's self image, are called *ego-syntonic*. Other thoughts that conflict with one's central values are called *ego-alien* or *ego-dystonic*. An ego-dystonic impulse to kill someone, inconsistent with one's predominant value systems, may generate a counteractive ego-syntonic thought such as, “You really don't mean it.”

Disturbances in Thought Content Abnormal beliefs and convictions form the core of thought content disturbances. Considerations of abnormality regarding beliefs and convictions must take the person's culture into account. Beliefs that may seem abnormal in one culture or subculture may be commonly accepted in another. For example, religious hallucinations, attributed to psychological or biological factors by contemporary Western societies, are routinely attributed to religious and spiritual causes by many other cultures. With regard to intensity of conviction, distorted beliefs range on a continuum from overvalued ideas to the determined, unshakable belief that is characteristic of fixed delusions. Abnormal beliefs and delusions are, in most circumstances, diagnostically nonspecific. Delusions are commonly seen in mood disorders, schizoaffective disorder, delirium, dementia, and substance-related disorders, as well as in schizophrenia and delusional disorders.

Overvalued ideas are unreasonable and sustained abnormal beliefs that are held beyond the bounds of reason. Patients with overvalued ideas have little or no insight into the fact that their ideas are very unlikely to be valid; however, the ideas themselves are not as patently unbelievable as most delusions. The distorted body images of body dysmorphic disorder exemplify overvalued ideas. Morbid jealousy and preoccupation with a spouse's possible infidelity may constitute an overvalued idea if no real evidence has ever existed to warrant such suspicion.

> A 32-year-old woman, fatigued for many months, complained of being “allergic to everything.” She initially associated her fatigue with eating certain foods, then with using certain cosmetics and soap products, then with wearing certain types of clothing, and then with being around certain types of housepaints, carpets, and draperies. These beliefs resulted in severe restrictions and functional limitations in her work and social life.

Ideas of reference are false personalized interpretations of actual events in which individuals believe that occurrences or remarks refer specifically to them, when in fact they do not. Ideas of reference may be less firmly held than delusional beliefs.

A psychotic young man came to the emergency room in a state of severe water intoxication. He had been sitting at home drinking gallons and gallons of ginger ale, after having seen a television advertisement that said “Drink Gingerale!” He believed this message was specifically addressed to him, and that dire consequences would result if he disobeyed.

Delusions *Delusions* are fixed, false beliefs, strongly held and immutable in the face of refuting evidence, that are not consonant with the person's education, social, and cultural background. Thus, delusional thoughts can only be understood or evaluated with at least some knowledge of patients' interpersonal worlds, such as their involvements with religious or political groups. One of the mind's primary functions is to generate beliefs, including myths and meaning systems. These beliefs provide the individual with a sense of personal and group identity and with ways of understanding reality. They are most noticeable when shared untestable beliefs form the basis for group cohesion as in religions and cults. Some groups adhere to their cherished beliefs despite the abundance of plausible contrary evidence, for example, some fundamentalist sects take the biblical creation story literally. In the face of contrary evidence or grave personal threat, individuals often cling to their primary beliefs as matters of faith (i.e., alternative, nonrefutable bases for understanding). The strong faith with which religious, political, and nationalistic convictions are held, even at the cost of death, shows the power that untestable beliefs can have on behavior. Potential mental health advantages of religious beliefs have been demonstrated in epidemiological studies showing that those with a sense of personal devotion report fewer depressive symptoms.

Subjectively, delusions are indistinguishable from everyday beliefs. Therefore, the subjective experience of a delusion is no different from the subjective experience of believing that the earth is round or that my spouse is the same person I married on my wedding day. Because of the identical experience of delusions and other strongly held beliefs, it is generally impossible to argue a patient out of a delusional belief. The content of delusions is highly influenced by culture. Whereas centuries ago delusions of persecution often concerned persecution by the devil and had religious connotations, persecutory delusions today often take on contemporary political and social perspectives.

> A 42-year-old Native American Vietnam veteran fled to a remote area of the Rocky Mountains to escape a world-wide conspiracy that he believed was trying to control each and every individual, including him. He was aware of this conspiracy because when he lay on the ground at night he could see countless stars and knew that they, and everything on earth, were all connected by “the Web.” Confirming this belief, he heard the murmurings of all computer messages, radio and television transmissions, phone calls, and even face-to-face conversations as part and parcel of this web.

Although delusions are diagnostically nonspecific, some specific types of delusions are more prevalent in one disorder than another. For example, although delusions of control and delusional percepts are often seen in schizophrenia, they also occur, albeit less frequently, in psychotic mood disorders. Similarly, classic

mood-congruent delusions with grandiose themes seen in mania or delusions of poverty characteristic of depression may also be seen in schizophrenia.

Table 8-3 lists some characteristics by which delusions have been classified. *Simple delusions* contain relatively few elements, whereas *complex delusions* may contain extensive elaborations of people, spirits, motives, and situations.

Sample versus complex
Complete versus partial
Systematized versus nonsystematized
Primary (autochthonous) versus secondary
How they affect behavior

Table 8-3 Characteristics of Delusions

Systematized delusions are usually restricted or circumscribed to well-delineated areas, and are ordinarily associated with a clear sensorium and absence of hallucinations. They are often isolated from other aspects of behavior. In contrast, *nonsystematized delusions* usually extend into many areas of life, and new data—new people and situations—are constantly incorporated to further support the delusion. The patient usually has concurrent mental confusion, hallucinations, and some affective lability. Whereas the patient with a closed systematized delusional system may go about life relatively unperturbed, the patient with a nonsystematized delusion frequently has poor social functioning and often behaves in response to the delusional beliefs.

Complete delusions are those held utterly without doubt. In contrast, *partial delusions* are those in which the patient entertains doubts about the delusional beliefs. Such doubts may be seen during the slow development of a delusion, as the delusion is gradually given up, or intermittently throughout its course.

Delusions have also been categorized into *primary* and *secondary* forms. Unfortunately, these terms have been used in a variety of ways, such that the distinctions are confused in the literature. According to one definition primary delusions are those that are not further understandable in terms of the patient's specific context, such as culture or mood, whereas delusional beliefs that could be understood or appreciated as an extension of a cultural belief or of a mood are secondary. In this framework, a mood-congruent psychotic delusion with themes of worthlessness or guilt would be a considered a secondary delusion. According to a different definition, a primary delusion refers to and is synonymous with an *autochthonous delusion*, one that takes form in an instant, without identifiable preceding events, as if full awareness suddenly burst forth in an unexpected flash of insight, like a bolt from the blue. These delusions may be quite elaborate.

> An 18-year-old young man from a middle-class Jewish family who had become increasingly socially isolated and withdrawn over a 2-year period presented to an emergency room in an utter panic, with his penis taped to his thigh, believing that his penis had started to shrink into his body and would disappear. No other delusional beliefs or hallucinations were evident.

Aside from the autochthonous types, three other types of delusions have been described as primary. *Delusional percept* refers to the experience of interpreting a normal perception with a delusional meaning, one which has enormous personal significance to the patient. *Delusional atmosphere* or *delusional mood* is a state of perplexity, a sense that something uncanny or odd is going on that involves the patient, but in unspecified ways. Ordinary events may take on heightened significance but the delusional interpretations are fleeting whereas the uncanny feeling lingers. Typically, after a period of time full-blown delusions develop, replacing the delusional mood. *Delusional memory* is the memory of an event that is clearly delusional. As an example, a patient “remembered” his fourth-grade teacher slipping lysergic acid diethylamide (LSD) into his apple juice; this memory served to explain his psychotic disorder. The elaboration of false memories and their subsequent fixed beliefs may assume delusional proportions.

> A young woman with schizophrenia attended a trauma group with her roommate and gradually came to believe that she had been repeatedly sexually assaulted by her father, from the time she was in the crib. What started out as vague dream-like images gradually coalesced into a series of “sensed” memories that then took on specific visual images of her father's fingers penetrating her, and of his leering down at her in the crib. Her parents were horrified by these accusations, and there was not a shred of evidence to corroborate her increasingly venomous accusations.

Patients vary considerably in the extent to which they take action in response to delusional thoughts. Just as patients can experience delusions of their thoughts being controlled *(thought passivity)*, they may similarly experience their feelings, behaviors, and will as controlled by outside forces. These *delusions of control* (or *passivity experiences*) occasionally, albeit uncommonly, result in dramatic self-destructive or aggressive behaviors, as illustrated by the murderer who called himself Son of Sam. This psychotic killer murdered several people in New York and claimed that he was the powerless agent of a force that required him to commit the acts. To defend themselves and others against delusional anticipated events, some patients may take bold and occasionally destructive actions.

> A 36-year-old man with schizophrenia was brought to the hospital after enucleating one of his eyes with a pencil. He believed that he had to sacrifice his eye in order to avert world destruction. In an elaborate delusional scheme, he was convinced that the evil he saw in the world was going to lead God to bring destruction upon all mankind, and he hoped that by sacrificing his eye God would see less of the evil in the world.

Table 8-4 lists some classic types of delusions. Although less common than those involving paranoia, grandiosity, and influence, *delusions of misidentification* are prominently reported because of their inherently intriguing nature. In *Capgras's syndrome* the patient believes that someone close to him has been replaced by an exact double. In *Fregoli's phenomenon* strangers are identified as familiar persons in the patient's life. In the *delusion of doubles*, patients believe that another person has been physically transformed into themselves.

Delusions of persecution
Delusions of grandeur
Delusions of influence
Delusion of having sinned
Nihilistic delusions
Somatic delusions
Delusion of doubles (*Doppelgänger*)
Delusional jealousy (Othello syndrome)
Delusional mood
Delusional perception
Delusional memory
Delusions of erotic attachment (Clérambault's syndrome)
Delusions of replacement of significant others (Capgras's syndrome)
Delusions of disguise (Frégoli's phenomenon)
Shared delusions (folie à deux, folie à trois, folie à famille)

Table 8-4 Some Classic Types of Delusions

Delusions are not only seen in isolated individuals. Shared delusional disorder may occur in couples *(folie a deux)* and in families *(folie en famille)*. Many psychiatrists consider group delusions to be present in some cults, however, the exact distinction between delusions and other zealous beliefs held by larger, more traditional, and

well-organized religious, political, and other groups, is arguable.

Self-Mutilatory, Suicidal, Aggressive, and Homicidal Preoccupations Other disturbances in the content of thought include maladaptive, self-destructive, and other-directed destructive thoughts. In some individuals, such thoughts occur only during times of acute distress, in reaction to specific stressful situations. In others, these thoughts may be chronically present, like background music in the stream of consciousness, varying in the frequency and intensity of their presence, in the specificity and vividness of their accompanying images, and in how compelling they are.

> A 36-year-old unemployed man with alcoholism became increasingly preoccupied with the idea that his wife was unfaithful. He accused and threatened her, stalking her whenever she left the house, filled with rage and entertaining fantasies of killing her and her presumed lover if he found them together. After months of trying to appease and reason with her husband, and after several instances of being beaten, the wife finally left him. She took their child to a shelter, hoping that he would leave them alone, but constantly feared for their safety.

Disturbances of Judgment Judgment involves a complex and diverse group of mental functions that includes analytic thinking, social and ethical action tendencies, and depth of understanding or insight. Analytic thinking includes the capacity to discriminate and to weigh the pros and cons of alternatives. Social and ethical action tendencies are closely related to culture and upbringing. The evidence for genetic factors in antisocial personality disorder (defined primarily by judgments that lead to criminal behaviors) points to the additional role of constitutional factors. Insight may reflect intelligence, learning, cognitive style, and the capacity to integrate intellectual knowledge with emotional awareness.

Impairments of judgment occur in many psychiatric disturbances. Anxiety states, intoxications, fatigue, and even group pressures may cause temporary impairments of judgment in otherwise normal individuals. Organic brain damage, cognitive disorders, and psychotic disorders may chronically impair any aspect of judgment in any person regardless of premorbid character. Poor role models and deviant social backgrounds may lead to social and ethical action tendencies quite different from those of the examiner. Thus, someone raised in a criminal environment may have superb analytic judgment and self-awareness, which are, however, put to illegal use.

Judgment may be impaired in one dimension and spared in others. Individuals may retain sound ethical judgment when their analytic capacities fail or they may retain excellent analytic abilities for nonpersonal matters while lacking insight into personal situations or behaviors. Thus, some people who can provide socially appropriate responses to traditional mental status examination questions such as what one would do in a movie theater if a fire broke out, or what one would do with a stamped and sealed addressed envelope found in the street, might at the same time be incapable of accurately assessing crucial clinical information or more personal matters specifically related to the capacity to provide informed consent, such as the pros and cons of receiving a medication or electroconvulsive therapy; or regarding judgments necessary to provide oneself with food, clothing, and shelter; or insight into one's state of health or illness. The apocryphal story about the delusional patient who was able to accurately evaluate and fix a broken-down car that had stymied the mechanics, ending with the patient's declaring “I may be crazy, but I'm not stupid” indicates the selective nature of poor judgment within psychiatric disorders.

The term *insight*, usually applied in the context of self-awareness, has been used in a variety of ways. Basic insight refers to a superficial awareness of one's situation, (e.g., that one is ill). A deeper level of insight is operating when the patient has an intellectual appreciation of what is going on (e.g., “I have hallucinations and delusions, and my doctors have told me that I have schizophrenia and must take medication”). Still deeper levels of insight reflect more complete cognitive and emotional appreciation of a situation (e.g., “I realize that I have schizophrenia, that it impairs my judgment and social function at times, and that I will have to take medications if I am to minimize my symptoms and try to make the most of my life. I feel profoundly disappointed about this affliction because it prevents me from achieving some of the goals I've always wished for. Nevertheless, I have do my best to get over my disappointment and hurt feelings so that I can get whatever I can out of life.”)

Judgment may be impaired by several factors, including cognitive clouding (as in disturbances of consciousness, such as intoxication that impairs one's usual analytic abilities), self-deception, and impulsivity.

Self-deception refers to the almost universal tendency to hide certain issues about the external world or about ourselves from various levels of one's own awareness. Self-deception functions as a coping strategy, fostering or maintaining comfortable perspectives about the world, and avoiding confrontation with issues and realities that would inevitably stir up painful conflicts or the need for difficult actions; it thereby helps an individual to preserve emotional calm. In addition, studies suggest that self-deception enables one to act and to be perceived as more convincing in the service of particular goals, as in romantic relationships or business dealings. Therefore, while kidding oneself may sometimes reflect impaired judgment, it may also yield certain important strategic advantages.

Impulsive judgment describes a tendency to avoid taking the time to fully understand and integrate all the facts and levels of awareness required for optimal decisionmaking. Impulsive judgment may occur only with certain issues or situations (such as how one picks investments), may signal an impaired state (such as substance intoxication), or may reflect a pervasive character trait.

Rapidly made judgments, and even so-called “snap” judgments, are not all maladaptively impulsive, even when they involve very important areas of life. Rapid decisions can be very accurate, highly adaptive, and even life-saving, especially if made against a background of great experience, wisdom, and forethought concerning the area requiring the decision.

Disturbances of Consciousness *Consciousness* can be defined as subjective awareness of the self and environment. Biologists increasingly believe that a continuum of consciousness exists, extending from lower animals through *Homo sapiens*. However, consciousness is subject to conflicting definitions and conceptualizations, and exactly where consciousness begins in evolution remains unclear. Philosophers agree that it is the subjectivity of experience, the so-called *qualia* of consciousness, that clearly distinguishes living consciousness from the 1990s' best versions of self-regulating automata, elegant computers, and robots. All current attempts even to approach an understanding of consciousness are very unsatisfying; consciousness remains unexplained and as yet unexplainable from a scientific point of view. Consciousness has been viewed as an emergent property of complex biological nervous systems, as a poorly understood general property of an even more mysterious and complex universe, or as a phenomenon to be understood only in religious and spiritual terms. One of the best analyses to date of possible relationships between the construction of a brain and the possibility of consciousness has been set forth by Gerald Edelman, who believes that reflective consciousness cannot occur until complex higher-order brain systems evolve whose major functions are to monitor the experiences, activities, and results of activities of those lower-order brain systems that deal directly with appraising and responding to the external and internal environments. Such higher-order meta-systems require the presence of memory, so that current and immediate impressions can be checked and compared against past experiences. These meta-systems may use a variety of sensor mechanisms to detect and signal their sensations or perceptions of various events. Some of these sensors may correspond to feeling states and some may correspond initially to preverbal thought-like mechanisms that contain the capacity to develop and recognize abstract categories, and ultimately conceptual language-based thought.

Clinically, consciousness can be considered from both qualitative as well as quantitative viewpoints. Qualitatively, consciousness does not seem to be an all-or-none phenomenon. Rather, conscious experiences may gradually and phasically shift in focus, intensity, and clarity; altered states of consciousness may occur in which some aspects of consciousness, such as sensation, perception, memory, orientation, and judgment are enhanced or impaired relative to other aspects. Quantitatively, crude divisions can be made between states depending on the relative presence, impairment, or total absence of consciousness. Even within a single individual consciousness is not a unitary phenomenon. Multiple streams of thought, operating at multiple levels of preconsciousnesses, appear to exist in all of us almost all the time, with various elements in these coexisting streams constantly shifting into higher or lower levels of conscious awareness. In pathological states, even more remarkable properties of consciousness are seen; for example, the existence of co-consciousness in humans who have had commissurotomies, and of seemingly multiple discrete consciousnesses in patients with dissociative identity disorders.

Experiments involving patients with commissurotomies of the corpus callosum have shown the existence of two virtually separate systems of consciousness that seem to operate simultaneously. For example, when in the course of an experiment the picture of a nude woman was flashed only to the right brain (the left visual field) of a commissurotomized patient, the subject verbally denied being aware of anything unusual (i.e., the left brain—the verbal brain—was unaware). However, the patient started to squirm and blush, blurting out, “Oh, you have some machine!” Similarly, when a cup was presented to the right brain (left visual field) only, the patient denied seeing anything (the left brain was unaware and the language output of the left brain indicated no awareness), but he was able to pick out the cup from an assortment of objects with his left hand (right brain control). This literal splitting of verbal awareness from visual-spatial awareness in the brain produces behavior that is at least superficially similar to that of patients who deny being consciously upset by an event, but who react with strong visceral responses. Although this formulation is simplistic, the separate consciousness for logical-verbal and for spatial-visual awareness demonstrated in split-brain experiments may be crude analogues for more highly differentiated and discrete types of awareness and modes of information processing. Furthermore, the very fact that there are separate and

even competing modes of consciousness may increase the likelihood of psychological distress, because the various modes are capable of yielding internally conflicting views of reality.

Psychological and Physiological Factors In ordinary states of alert consciousness individuals are able to pay an adequate amount of attention to their surroundings and to reflective thought. Normal people vary enormously in their ability to pay careful attention in different settings without being distracted; individual variations may reflect temperamental and cognitive style differences as well as physiological shifts within the individual. Many functions of attentive consciousness including attention, planning, and the capacity to switch appropriately between mental tasks have recently been linked to the activity of specific neurons in area 46 of the prefrontal cortex.

A sense of increased consciousness with heightened alertness, awareness, and sharper thinking may be experienced in highly aroused emotional states such as threat, sexual attraction, falling in love, or other high-stakes events such as hunting (in primitive peoples), sporting competitions, or performing in front of an important audience. High levels of arousal do not necessarily guarantee effective attention because optimal consciousness depends on optimal arousal. Too little arousal, caused by illness or fatigue, may result in insufficient stimulation and mental lethargy, diminishing the sense of alertness and attentiveness; too much arousal may result in hyperintense alertness but distractibility and scattered attention.

Consciousness involves, among other things, the experience of a continuous sense of self and of the environment, existing coherently in time and space. The experience of time and its passage may be altered by shifts in the level of awareness and by emotional states such as boredom, concentration, pain, and discomfort. The experience of time and space may be altered by hypnosis, cannabis, psychoactive and psychedelic drugs, and other events that directly affect brain physiology.

Disturbances in the Level of Consciousness Levels of consciousness (i.e., alertness, awareness, and attentiveness) may be pathologically increased or decreased. Such changes are diagnostically nonspecific and can occur in many different disorders. When levels of arousal and alertness are mildly elevated, as in hypomania or with the ingestion of small amounts of psychostimulants, subjective experiences are typically positive. In these situations the person experiences intense alertness, prolonged concentrating ability, and hyperesthesias in which perceptual vividness is heightened: colors are brighter, sounds are sharper, and touch is more intense than usual. With further increases in arousal and consciousness as seen in mania, more severe intoxications with amphetamines and cocaine, and catatonic excitement, attention deteriorates. Heightened alertness transforms into hypervigilance and paranoia, and hyperesthesias become unpleasant.

Diminished levels of consciousness can be described on a continuum. *Clouding of consciousness* is marked by diminished awareness of sensory cues and diminished attentiveness to the environment and to the self. Secondary-process thinking is most notably compromised, and more primary-process thinking emerges into consciousness. In this state, one's ability to appreciate subtleties and to think in a nuanced manner is diminished, and is replaced by more dichotomous all-or-none, stereotypic thinking. The level of consciousness may fluctuate rapidly in relation to the internal physiological state or to the degree of external stimulation. In alterations of consciousness, confusion may occur with disorientation to time, place, or person. The patient is usually highly distractible and unable to pay sustained attention to a single stimulus.

Torpor is a condition in which the patient is drowsy, falls asleep easily, and shows a narrowed range of perception and slowed thinking. *Stupor* is a state of diminished consciousness in which the patient remains mute and still although the eyes are open and may follow external objects. In the most extreme impairment of consciousness, *coma*, there is no evidence of mental activity at all. The patient appears essentially to be functioning on a decorticate or decerebrate level. In *akinetic mutism* or *coma vigil*, patients with profound brainstem lesions appear to be awake with their eyes open, but there is in fact no evidence of consciousness.

Delirium, an acute confusional state, is usually characterized by a relatively abrupt onset and a short duration of clouded, reduced, and fragmented attention; impaired memory and learning; perceptual and cognitive abnormalities such as hallucinations and delusions; disrupted sleep; and other autonomic dysfunction. The level of consciousness may be consistently diminished or may fluctuate. The electroencephalogram (EEG) usually shows diffuse slowing. Typical motor abnormalities include an increase in general restlessness, fine and coarse tremors, and myoclonic jerks. Autonomic disturbances commonly include tachycardia, fever, elevated blood pressure, diaphoresis, and pupillary dilatation. The causes of delirium are legion, including systemic medical disorders such as metabolic imbalances or infections; intracranial disorders caused by traumatic, structural, and electrical causes; and substance intoxication and withdrawal states.

Attentional difficulties are manifest by impairments in the person's ability to deploy, focus, and sustain attention. Some attentional difficulties first appear in early childhood as developmental problems of uncertain cause, and are described as *attention-deficit disorder*. Secondary attention-deficit difficulties may appear de novo in adulthood from a variety of exogenous agents, psychiatric disorders, and late-life developmental and degenerative factors.

In *narcolepsy*, which is characterized by sudden lapses into sleep, one's usual ability to stay alert and maintain consciousness is impaired. At times the onset of profound sleepiness is gradual, accompanied by hypnogogic phenomena, in which dreamlike images invade the consciousness; at other times the shift in consciousness appears to be almost instantaneous. This syndrome, occurring in about 1 in 10,000 persons, is thought to be the second most frequent cause of automobile accidents after alcohol intoxication.

Altered States of Consciousness Consciousness may also be qualitatively changed with the production of altered states. Drugs such as scopolamine (Transderm) with strong central anticholinergic properties, some seizures, and on occasion other conditions associated with delirium can induce *twilight states*, dream-like states of wakeful consciousness in which attention is poor, a mixture of primary and secondary process thinking appears, and patients fade in and out of alertness. Dream-like experiences intrude into the stream of conversation. Emotional outbursts or violent acts may occur during twilight states.

Mystical states of consciousness may occur in normal and pathological conditions. Intense meditation and peak or epiphanic experiences, reported by more than 10 percent of normal individuals in community surveys, may produce a sense that the self dissolves or expands, that the self fuses mystically with the cosmos, that time stops, and that universal meaning becomes clear. These perceptions may be accompanied by a sense of rejuvenation and renewed personal identity, ineffability, intense emotionality, and concurrent perceptual changes. Such experiences do not ordinarily last more than a few minutes. Many people have achieved these states through the use of psychedelic agents such as mescaline and LSD. Reports of a white light at the end of a tunnel, described by individuals using psychedelics and during near-death experiences, have been linked to specific neurophysiological pathways thought to be stimulated under these conditions.

Although *hypnosis* lacks a consensually accepted definition, its hallmarks are selective attention, suggestibility, and dissociation. Most people, but not all, can be hypnotized to some degree. Up to 90 percent of people are capable of achieving a light trance, whereas 10 to 20 percent are capable of entering a deep trance and exhibiting remarkable hypnotic phenomena. Hypnosis occurs when the subject is in a state of heightened, not diminished, attention. EEG studies have shown hypnotized subjects to be fully awake and alert. The heightened concentration probably accounts for the unusual levels of sensory and motor performance often seen under hypnosis and self-hypnosis.

Hypnotic phenomena include hypnotically induced hallucinations (including negative hallucinations in which the subject selectively fails to perceive sights, sounds, or other stimuli), anesthesia, sustained motor behaviors and acts of strength ordinarily beyond the individual's capacity, and distortions of memory (both hypermnesias and amnesia). Several phenomena that reveal the multiple nature of consciousness, for example, co-consciousness, are also demonstrable. Experiments have shown that even when a subject in a deep trance has achieved profound hypnotic anesthesia and can, for example, keep a hand submerged in ice water for longer periods of time than usual, part of the hypnotized subject's consciousness continues to register exactly how painful the experience actually is and can signal the researcher about the pain by finger movements without the subject having any conscious awareness or disturbance. This phenomenon, called the *hidden observer*, has also been seen in postsurgical patients who, under hypnotic trance after surgery, have been able to accurately recall conversations in the operating room that occurred while they were under general anesthesia. Dissociative and psychosomatic phenomena have also been induced with hypnosis. With posthypnotic suggestion, subjects may carry out complex actions without any hint that they are actually doing so because they were previously instructed to act that way under hypnosis. When asked why they are carrying out these activities, such subjects will usually make up reasons, while seemingly unaware of the real reasons for their actions. It has been suggested, not entirely facetiously, that many normal daily activities are conducted in a trance-like posthypnotic state, and although these activities are attributed to conscious intention, they may in fact be carried out as a result of previous suggestion. Advertisers know this well. Urticaria (hives) can be hypnotically induced and hypnotically made to disappear. Plantar warts have been successfully treated with hypnosis, and in these conditions diminished blood supplies to their bases have been demonstrated. It has recently been appreciated that yoga masters can exert remarkable control over basic bodily functions through self-hypnosis. As yet, little is known of the full extent to which heightened concentration may influence physiological regulation.

Suggestibility Pathological suggestibility may be seen in several clinical conditions. Automatic obedience has been described in *echolalia* (the automatic repetition of a sentence or phrase just uttered by another person), *echopraxia* (the automatic mimicking of a movement performed by another person), and *waxy flexibility* (maintaining for a prolonged period of time a posture in which one is placed), symptoms common in catatonic states. In situations of group delusions, and sometimes

in cults, passive individuals adopt the delusional beliefs of stronger ones. In epidemic hysteria, as described so beautifully among young women at the Salem witch trials in Arthur Miller's *The Crucible*, distorted and even delusional perceptions and beliefs may sweep over a group that has been highly aroused by a charismatic leader.

Autosuggestibility can be seen in the constructions of false memories, in which an individual progressively comes to believe that something that never happened in fact occurred. Such false memories may be held with such great conviction that they are indistinguishable from the memories of real events. Various types and degrees of self-deception may be more common in individuals who are more suggestible.

DISSOCIATIVE PHENOMENA *Dissociation* refers to the splitting off from one another of what are ordinarily closely connected behaviors, thoughts, or feelings. *Dissociative states* are those in which there is a disturbance or alteration in the normally integrated functions of identity, memory, or consciousness and include trances, fugues, blackouts, multiple personalities (*dissociative identity disorder*), and dissociative frenzies. Although dissociative states are ordinarily thought to be functional in nature, arising as an adaptive defense in individuals subjected to a great deal of trauma, particularly at early ages, they occur regularly with a variety of neurological disorders, particularly with partial complex seizures. In one series, one third of patients with complex partial seizures had dissociative phenomena including multiple personality. In these patients the dissociative phenomena were not related to the seizure activity, but to interictal alterations.

As in posthypnotic amnesia, elaborate activities can occur in dissociative states for which the subject will have no conscious memory. This amnesia is functional in nature and may be reversed by hypnosis or drug-facilitated disinhibition, for example, with amobarbital (Amytal) infusion. In many of the functional dissociative states, amnestic episodes may occur for years or decades before the patient seeks medical or psychiatric attention. *Blackouts* are periods of amnesia in alcoholism, other intoxications, or following head trauma. An alcoholic blackout period may last for hours to days, after which the person has no recollection of what transpired although other observers will attest to the fact that during this period the individual carried out multiple complicated behaviors. Although memory of the blackout is lost to the predominant consciousness, during subsequent reintoxication memories of events occurring during the previous blackout may be reawakened. This phenomenon, known as *state-dependent memory*, occurs in many other conditions as well, signifying that one's ability to retrieve specific memories may be highly influenced by specific physiological alterations caused by external intoxicants or other unusual physiological states.

When he was questioned during an initial psychiatric interview, a 28-year-old euthymic man with a history of severe bipolar I disorder could not recall the content of his thinking during any of his manic episodes. However, at the beginning of his next manic period he reminded his psychiatrist about this question, and was able to describe the grandiose delusions and hallucinations he had experienced during his previous manic states.

Dissociative fugue is characterized by prolonged periods in which individuals carry out very complex activities without having any recollection of their previous lives, identities, or even names. They often travel away from customary locales and assume entirely new identities. By definition, psychogenic fugue cannot result from a neurological disorder. In comparison, the discontinuity of experience in *dissociative amnesia* is typically more circumscribed and does not involve assuming an entirely new identity. The dissociated memories and moods often reveal themselves in disguised form as nightmares, intrusive visual images, and conversion symptoms. Typically, in psychogenic amnesia an individual may not be able to recollect what transpired during a specific period of time, for example, before the age of 9 or 10 in the context of a traumatic childhood, or catastrophic events such as traumatic, gruesome combat, or less momentous events that a person prefers to forget in order to preserve self-esteem by denying shameful, immoral, or illegal activities.

A 42-year-old man who had been in an automobile accident several months previously, without any clear report of head injury at that time, was seen in a hospital claiming that he was unable to remember any of the events of the accident or any events for a week prior to or subsequent to the accident. After several interviews conducted after he had been administered amobarbital he was able to recall that although formal police charges had never been filed against him, the accident had been his fault and that a passenger in the other car had been killed as a result of the accident. His memories evoked a great deal of psychological pain, grief, and remorse.

Dissociative identity disorder is a chronic, dissociative state in which two or more separate ongoing identities or personalities alternate in consciousness. It usually occurs in persons who as young children were severely and repeatedly brutalized. The number of identities is variable, with some cases reporting 25 or more identities. The development of dissociated alter personalities is thought to be a last-ditch, primitive psychological defense against inescapable and unbearably traumatic situations. The personalities may be of different ages and even different sexes. Typically, the presenting identity is dysphoric, anxious, and constricted, may suffer headaches and periods of blackout or amnesia, and is not aware of the other personalities. A second identity is typically vivacious and uninhibited. Another identity may know all about the other personalities and has a wise perspective on the life events leading to the problems and on possible solutions. A classic case is described in a popular book and film, *The Three Faces of Eve*. In so-called channeling, dissociated complex part-personalities are produced in trance states, in which fictitious past lives or spirit lives are created.

In *Ganser's syndrome* the patient responds to questions by giving approximate or patently ridiculous answers, for example, in response to the question “What sound does a dog make?” the patient answers “moo.” Additional features of the syndrome include alterations in consciousness, hallucinations (or pseudohallucinations), conversion phenomena, and amnesia for the episode during which these symptoms were manifest. This syndrome has most commonly been reported in prisoners, and is generally thought to be a dissociative state, although organic features may contribute.

Depersonalization refers to an alteration in one's experience and awareness of the self, leading to feelings of being unreal or detached from one's own body, of feeling like an automation; it is often accompanied by the complaint that the individual lacks all feelings and sensory experiences. Those experiencing depersonalization frequently fear that they are going crazy. Because of this fear, patients often endure depersonalization experiences for long periods of time before describing them to a mental health professional. Depersonalization is also characterized by frequent internal nonaudible dialogues between the participating self and the observing self, but with full awareness that both parties are the same person (a feature that distinguishes it from hallucinations.) Mild sensory distortions, but not hallucinations, are commonly associated with the experience. Depersonalization is seen in a variety of neurological and psychiatric disorders and is common in complex partial seizures. It may occur in the context of depression, anxiety disorders, or certain personality disorders or as an entity by itself (depersonalization disorder).

In *derealization*, individuals know themselves to be real, but feel that the world around them has suddenly become unreal. Derealization often but not always accompanies depersonalization. Transient episodes of depersonalizaiton and derealization occur frequently in normal persons, particularly during states of fatigue, sleep deprivation, or stressful situations such as bereavement, learning of a terminal diagnosis, or sudden awareness that one is about to be in an inescapable vehicle accident.

A 27-year-old man described driving in his car when he realized that he was about to be hit by a huge tractor trailer and had no way of avoiding the collision. During the seconds prior to impact, at the moment of impact itself, and for minutes following the impact when he sat, dazed and thankful to be alive, in his totally destroyed car, he felt completely estranged from events, saying to himself, “This can't be real; this isn't happening to me; I'm going to wake up and things will be normal.”

Disturbances of the Self At the most basic level, the key components of self-awareness are the reality and integrity of the self (that I am one person), the continuity of self (that I am the same person now that I was in the past and that I will be in the future), the boundaries of self (that I can distinguish between myself and the rest of the world as not-self), and activity of self (that it is *I* who is thinking, doing, feeling). Additional components of a sense of self include body image, and various self-evaluations including self-esteem and ego ideal (ideal self). *Body image* is an individual's mental representation of his or her own body. *Self-esteem* is thought to reflect how one measures up to the desired self-image. To the extent that what one sees in oneself approximates what one would like to be, elf-esteem is positive. *Ego ideals* are fantasies of the optimum person one could ever wish to be. Any of these qualities may be disturbed in psychiatric disorders.

Within each individual is a group of *social selves* composed of the roles and identities that the person assumes and that are evoked in various contexts. The presenting “self” varies depending on the persons with whom we are interacting, for example, parent, romantic partner, child, friend, or employer, and depending on what role we are assuming, for example, child, parent, colleague, or lover. Accompanying each of these “selves” are various levels of objective and subjective self-awareness and self-understanding.

Disturbances in Sense of Self Disturbances of the basic elements of self-awareness may be seen in a variety of disorders. Discontinuity phenomena are characteristic of dissociative disorders, such as dissociative amnesias and dissociative fugue. Depersonalization reflects a mild disturbance in the awareness of self as the agent of activity. More severe disturbance is characteristic of the psychotic passivity phenomena seen in schizophrenia. Boundary disturbances may be

considered characteristic of all psychotic states, regardless of diagnosis.

Disorders of self-integrity are characteristic of dissociative identity disorders as well as severe borderline personality disorders in which a person's self-concept and expression of this concept to others is erratic, leading to a sense of unstable identity. The *as-if personality* typically adopts characteristics of those who are particularly important to the individual, like a personality chameleon. In adopting these characteristics, the as-if personality does not appear to be acting, but seems to experience and manifest the assumed traits in a very genuine manner, at least for a while. A change in relationships and situations usually prompts the as-if personality to summarily discard previously held traits and to assume new ones that better fit with the new circumstances.

False self describes a persona or a faulty and limited superficial aspect of the personality that an individual builds up as a mechanism for adapting to a hostile world—to please, control, or negotiate with others and with himself or herself. However, the false self fails to incorporate, integrate, or validate important fundamental needs, wants, values, and beliefs. Through self-deception and denial the individual may consciously believe that this "self" constitutes his or her entire being. However, the false self is a relatively fragile construction that has usually warded off and denied fundamental strivings, which may include needs for autonomy, acting with integrity, expressing certain desires, beliefs or talents, or other unacknowledged aspects of the self. When these warded-off needs finally break through and demand expression at various points in development, the defective false self may collapse, leading to a period of distress and identity confusion which individuals sometimes describe as a "nervous breakdown."

Patients with pseudologia fantastica and the impostor syndrome demonstrate extreme examples of inconsistency in the sense of self. Patients with *pseudologia fantastica* compulsively spin out webs of lies, ordinarily self-aggrandizing ones, and also appear to be trying very hard to deceive themselves into believing that they are true. In the *impostor syndrome*, such fantasies are acted out by liars and impostors who seem to fervently wish that these fantasies were their reality, as if they cannot accept themselves and would be overwhelmingly ashamed to be known for who they actually are. The impostor compulsively adopts the identities of others and may, for example, show up properly attired at diplomatic functions and society galas and interact with the other guests under the assumed identity. Some famous impostors have repeatedly insinuated themselves into inner circles of high society and government.

Transsexualism is a syndrome characterized by the feeling that one was born into a body of the wrong sex and marked by the desire, starting at an early age, to be a person of the opposite sex. Male-to-female transsexualism is reported more often than female-to-male transsexualism. Both psychodynamic and biological theories have been advanced to explain these phenomena. Such persons seek procedures to change their anatomy with hormones or surgery.

Self-esteem is a measure of one's self-appraisal in relation to one's values and ego-ideals. Negative self-esteem is characteristic of depressive disorders, many personality disorders, and situational failures. Superficially inflated self-esteem may be seen in mania (or hypomania), or, in a fluctuating manner, in narcissistic and other personality disorders.

Although some individuals regard their ego-ideals as unattainable and are content to live as imperfect human beings, others strive to approximate their ideals. People who feel driven to achieve unattainable ideal goals or to become their unrealistically perfect ideal selves are likely to be chronically dysphoric and have poor self-esteem, since their attempts to become their ego-ideals are doomed to failure.

Disorders of the Will Central to the sense of self is the concept of will or volition. Psychologically, will is linked to the concepts of intentionality and of transforming awareness and knowledge into initiating action, as the bridge between desire and act. To manifest normal will, individuals must be aware and feel desires, and these desires must arise from within themselves. Concepts related to will that may become the focus of clinical attention when disturbed include motivation and decision making (i.e., the capacity to make choices).

Pathologically *heightened will*, seen primarily in manic states, is characterized by excessively intense desires and an overly facile capacity to make decisions, with complex questions being decided on in an instant. With heightened psychological energy, these individuals can start new courses of action with astonishing rapidity. Closer examination of these actions in more extreme cases, however, reveals that they share much in common with decreased will, in that the intense desires and quick decisions often reflect impulsiveness, which can be considered an escape from true willing and decision-making rather than enduring desires or thoughtful decision-making.

The term *abulia* has been used to describe the loss, lack, or impairment of the power to will or to execute what is in mind. Individuals with abulia show a diminished sense of motive or desire and impairment in making the transition from motive and desire to execution of action. Deficiencies in will may be seen in a variety of psychiatric disorders, and at the end of life when patients have surrendered their will to live and are simply waiting to die. In schizophrenia a diminished sense of will can be seen in passivity phenomena, as well as in other negative (or deficit) symptoms that may affect thoughts, feelings, and behaviors. These include lack of drive, impersistence at tasks, and a general inner flatness. Depressed patients also describe volitional disturbances, as in their general apathy and anhedonia. Patients who chronically inhale solvents (e.g., glue, gasoline, toluene), smoke marijuana very heavily, and chronically use hallucinogens have a characteristic *amotivational syndrome*. The extent to which this lack of motivation results from or contributes to the chronic substance abuse is a matter of debate.

In obsessive-compulsive disorder both the obsessional thoughts and the compulsive rituals are experienced as ego-dystonic and not consonant with the patient's conscious desires and will. Similarly, although patients with anorexia nervosa initially have the conscious experience of willing and controlling their intake of food, during the course of the disorder the sense of willfulness is replaced by passivity, of being subjugated by obsessional thoughts and compulsive behaviors that assume control of the eating behavior.

> A 35-year-old woman with chronic anorexia nervosa felt compelled to cut her food into pieces no larger than rice grains, fearing that larger pieces would make her fat, and that she might choke. She permitted herself no foods that contained any fat. Other compulsions required her to swim eight or sixteen laps in the local pool each day. Failing to do so would result in a masochistic punishment, usually burning her arm with a lit cigarette.

Disturbances of volition are among the more common complaints of patients with personality disturbances who request psychotherapy. Individuals with dependent personalities are characterized by difficulties in making decisions by themselves and often engage in courses of action contrary to their own desires. Similarly, individuals with passive-aggressive personality disorder obscure their own desires by being excessively involved in the demands made upon them by others. Their courses of action do not reflect their own decisions so much as the thwarting of others' desires. People with compulsive personalities use inflexible rules, thereby precluding courses of action based on independent evaluation, individual desires, and decisions. In other situations, they are indecisive, sometimes making impulsive decisions at the last minute when forced to decide. Finally, many individuals seek treatment because of self-designated disturbances of willing: they do not know what they want, they are unable to make choices among several options, or they procrastinate excessively. Often these problems may mask other fears—of wanting, commitment, taking initiative, hard work, success, making a mistake, being criticized, angering others, and of all the consequences related to such actions.

Disturbances of Orientation Orientation refers to one's awareness of time, place, and person. Accurate orientation requires the integrity of attention, perception, memory, and ideation. Impairments occur primarily in organic mental disorders (i.e., structural and toxic metabolic brain abnormalities) and occasionally in dissociative and psychotic states.

Normal individuals vary tremendously in their attention to the details of time, and in the extent to which their bodies automatically keep time. Some people have reliable built-in clocks by which they can awaken themselves at precise times or gauge the passage of time with uncanny accuracy, even in the absence of external cues, as during a psychotherapy session, for example. Others have difficulty making judgments about time and may develop pathological lateness or habitually schedule more activities than could ever be accomplished in the available time. Benign disorientation to time is common. After a few days in a hospital bed, most people do not know exactly what the day or date is because they are not attending to or receiving their usual cues.

Pathological time disorientation can be mild or severe, with inaccuracies of estimation ranging from days to years. The dates reported by disoriented individuals may have personal significance, such as those of important births, marriages, or deaths.

Because spatial cues are generally more available and obvious than temporal cues, disorientation to place often signifies a greater degree of cognitive impairment than disorientation to time, and therefore rarely occurs in the absence of time disorientation. Disoriented persons may know, more or less, the type of place they are in without knowing the specific place—patients may recognize that they are in *a* hospital without being able to name the hospital.

A 56-year-old former banking official with advanced Alzheimer's disease demonstrated preserved superficial social graces and was generally in a pleasant mood. When anyone asked him where he was, however, he appeared to become momentarily aware of his profound deficits and became very agitated, developing a marked tremor of his left arm and hand and fearfully looking about. This would last for a few minutes, until he was calmed and reassured that he was all right.

Disorientation to *person*, a lack of awareness of one's own identity, is typically seen only in advanced dementias, such as primary degenerative dementia of the Alzheimer's type or in dissociative states. In postconcussion amnesia, transient global amnesia, and dissociative fugue, knowledge of one's own identity may disappear and a person may remain unidentified for an indefinite period until the memory for self returns.

Disturbances of Memory Memory is not a unitary phenomenon. Capacities to remember vary for the different senses and perceptions. One person may have prodigious musical memory, with the capacity to remember and reproduce whole musical pieces after one hearing, but be incapable of remembering people's names or telephone numbers. Exceptionally detailed verbal memories have been associated with obsessional cognitive styles. When individuals with extraordinary memories complain of memory loss, ordinary memory tests may be inadequate to detect their deficits because their relative memory loss may have reduced their capacities to a point within the range of most normal people.

Memory functions have been divided into three stages: registration, retention, and recall. *Registration* or acquisition refers to the capacity to add new material to memory. The material may be sensory, perceptual, or conceptual and may come from the environment or from within the person. In order for new material to be acquired, the person must attend to the information presented, it must then be registered through the appropriate sensory channels, and then be processed or cortically organized. *Retention* is the ability to hold memories in storage. Large numbers of neurons are thought to be involved in the storage of a specific memory, and it is believed that reverberating circuits are formed in which memory traces are held by means of changes in proteins or synaptic connectivity or both. *Recall* is the capacity to return previously stored memories to consciousness.

Newly registered material is transferred incrementally from immediate to short-term memory to long-term memory. Immediate memory lasts for 15 to 20 seconds; short-term memory (or recent memory) for several minutes up to 2 days (the time involved in new learning and its early consolidation); and long-term (or remote) memory for longer periods of time. Different physiological processes mediate each of these stages of memory. Because of this, processes that affect immediate or short-term memory often spare long-term memory. The processes by which memories are transferred from short-term to long-term stores are unknown.

Cognitive scientists now refer to short-term memory as *working memory*, the system that briefly stores and processes information needed for planning and reasoning. Recent studies suggest that the working memory system consists of at least two short-term memory buffers, one for verbal and another for visual memories, plus a central executive that manipulates and coordinates information stored in the two buffers for problem-solving, planning, and organizing activities. Parallel processing systems involving specific areas of the prefrontal cortex and other brain areas appear to operate separately with respect to various processes concerned with working memory. For example, separate prefrontal areas appear to be involved in working memory functions concerned with object identity and spatial locations.

Other studies suggest that different types of memories are stored and retrieved by different brain systems, so that there is at least a dual memory system. The first system, sometimes called a *conditioned-emotional system*, or system for *implicit memory*, or *perceptual memory*, or *nondeclarative memory*, is present from birth, operational through life, and is addressable by situational, sensory, or affective cues. Past experiences are expressed through images, behaviors, or emotions. These memories need not involve any conscious memories of a past experience. Conditioned fear responses represent examples of memories elicited in this system. The second system, sometimes called *narrative-biographical memory, explicit memory, reflective memory*, or *declarative memory*, emerges during the preschool years, and includes information significant to the self. Memories are addressable through intentional retrieval efforts, apart from the original learning conditions. They are identified as representing personally experienced events, and compose the individual's life history, roughly equivalent to memory with consciousness or memory with awareness. Clinical studies suggest that in at least some amnesias implicit and explicit memory functions may become dissociated. Disturbances in memory occur through the interruption of registration, retention, or recall.

Disturbances in Registration Registration and short-term memory retention are usually impaired in disorders that affect vigilance and attention such as head trauma, delirium, intoxication, psychosis, spontaneous or induced seizures, anxiety, depression, and fatigue. A variety of other metabolic and structural brain disturbances can affect short-term memory as well, particularly lesions affecting the mammillary bodies, hippocampus, fornix, and closely associated areas. Patients with impaired attention and concentration who are able to demonstrate immediate recall may not be able to retain or recollect these items from short-term memory. Benzodiazepine use has been associated with working memory difficulties, especially in the elderly. Some short-acting high-potency benzodiazepines used as sleeping pills may be particularly troublesome in this regard.

Disturbances in Retention The retention of memories is impaired in posttraumatic amnesia as well as in a number of cognitive disorders, such as dementia of the Alzheimer's type and the Wernicke-Korsakoff syndrome. The latter, which ordinarily results from chronic thiamine deficiency seen with alcoholism, is associated with pathological alterations in the mammillary bodies and thalamus.

Disturbances in Recall Disturbances in recall can occur even when memories have been registered and are in storage. Research has shown that memories are not passively retrieved but are actively reconstructed. Each act of recollection requires an act of putting the memory together, not simply lifting it ready made from a file. Because memories are often retrieved for specific purposes to meet the individual's particular needs and agendas, this act of reconstruction is often subject to the introduction of distortions and falsification. As a result, memories may fail to truly represent past events. At times, failure to recall may signify that the memory traces themselves have disappeared and are no longer retrievable. However, difficulties in recall can occur separately, as in the everyday event of forgetting the name of a person or object, only to spontaneously remember it hours or days later. In normal forgetting, more remote events are less well remembered than recent ones, and important events are most vividly retained in memory. Some patients with dementia may lose memories for all events occurring after a specific date or event, as if the slate had been wiped clean, but retain earlier memories. Some individuals may progressively erase memories, so that they recall only earlier and earlier events.

An 80-year-old patient with Alzheimer's disease initially had no memory for events that occurred after he was 70, a few months later he was unable to recall events that had occurred after he was 60, and several years later, although he could not remember his children who had been born in his 30s, he could still recall his spouse and siblings, from earlier periods of life. At each examination it seemed as if his memories for months' or years' worth of events were simply being erased along the time line of his life.

Under usual conditions, forgotten events can be recalled with prompting, associative memories, or other forms of stimulation such as hypnosis. As described earlier, state-dependent memories are recall failures, reversed by reinstituting the context in which the memory was originally formed.

Amnestic disorders are syndromes in which short-term and long-term memory is impaired within a state of normal consciousness. Thus, memory disturbances in delirium should strictly speaking not be considered amnestic syndromes. *Anterograde amnesia* is the inability to register or learn new information (and therefore to form new memories) from a specific event onward; it typically follows head trauma, states of cerebral physiological imbalance, or drug effects. Patients who receive electroconvulsive therapy (ECT) frequently have anterograde amnesias during the course of the treatments; the amnesia gradually fades over numbers of weeks. *Retrograde amnesia* is an impairment in recalling memories that were established before a traumatic event, extending backwards in time for variable periods. As memory is regained, the more remote memories usually return first. A patient originally amnestic for the 3-month period prior to an accident may ultimately be left with amnesia for events only a day or an hour just prior to the accident. In organically caused retrograde amnesias, remote memories are usually intact while amnesia may exist for more recent events. This contrasts with *dissociative amnesia*, in which the time periods of forgotten events may be more spotty or selective.

Hyperamnesia, unusually detailed and vivid memory, may occur in gifted persons, in association with obsessive-compulsive and paranoid personality traits, and in hypnotic trances. *Intrusive memories* may occur in posttraumatic stress disorder, signaling failure of the mechanisms that usually keep unwanted memories and information out of working memory.

A 36-year-old Hispanic woman who had been brutally raped 3 years before, after which the perpetrator savagely cut her genitals and anus with a razor, experienced frequent intrusive memories of the event, associated with marked anxiety and occasional panic. These memories were most likely to intrude during times of repose, when she was trying to fall asleep, and when she was feeling anxious about other ongoing issues in her life.

Although many forgotten memories can be recalled in hypnotic trance, retrospective falsification and distortion may also occur under hypnosis. (Memories recalled under hypnosis usually are not accepted as evidence in court.) Retrospective falsification of memory, the development of false memories, is called *paramnesia*, also

known as *fausse reconnaissance. Confabulation* is another common form of paramnesia in which the patient fills in memory gaps with inaccurate information. The responses given to questions by patients who confabulate may reflect past experiences or constitute bizarre, fantastic stories. Confabulation correlates poorly with memory deficit and is thought to reflect frontal lobe dysfunction and a failure of self-monitoring. Confabulation is prominent in certain alcohol amnestic syndromes such as Wernicke-Korsakoff syndrome as well as other disorders of the mammillary bodies, thalamus, or frontal lobes.

Dêja vu is the sense that one has previously seen or experienced what is transpiring for the first time; it is a false impression that the current stream of consciousness has previously been recorded in memory. Related phenomena are *dêja entendu*, a sense that one has previously heard what is actually being heard for the first time, and *dêja pensé*, a feeling that one has at an earlier time known or understood what is being thought for the first time. Experiences of *jamais vu, jamais entendu*, and *jamais pensé* involve feelings that one has never seen, heard, or thought (respectively) things that in fact one has. These phenomena are all common in everyday life but may increase in states of fatigue or intoxication and in association with complex partial seizures or other psychopathological states.

Dementia is a syndrome in which the essential feature is an acquired impairment of short- and long-term memory with associated impairments of abstract thinking and judgment, personality changes, and other cortical disturbances. The symptoms always involve more than one sphere of function. In later stages, patients with dementia may become helpless, too confused to use a stove, and incapable of remembering the names of close relatives. They may wander into dangerous situations, oblivious of their surroundings. Dementias are caused by a variety of pathogenic processes, some of which are reversible, such as hypothyroidism and subdural hematoma; others are irreversible, such as dementia of the Alzheimer's type and vascular dementia. Although the characteristic cognitive disturbances seen in severe major depressive episodes are usually called *pseudodementias*, many neuropsychiatrists believe that profound cognitive dysfunction meeting criteria for dementia associated with depression should properly be labeled a reversible dementia syndrome.

DISTURBANCES IN PERCEPTION

Normal perception first requires that the individual be capable of receiving information as sensations. The data must then be organized to make them meaningful and comprehensible, such as distinguishing figure from ground, or focusing attention selectively on some part of the sensory field. The organized entities are called *percepts*. In states of sensory deficit such as blindness, deafness, and anesthesia perception is impaired but is still possible because individuals generally perceive information about an object through several sensory modalities concurrently. The intensity of sensation and perception is affected by vigilance and attention. Highly focused attention, as in intense concentration or hypnosis, may result in unusually acute sensation and perception—hyperesthesia, hyperacusis, or extraordinary visual acuity. Focused attention may also result in failure to sense or perceive: deep anesthesia and negative hallucinations induced by hypnosis are simply induced failures to perceive what exists in the world.

Humans usually operate in an average expectable environment in which certain types and levels of sensory input are expected, and for which the nervous system is primed. Excessive or inadequate stimulation in any sensory modality, levels of input that are extraordinarily intense, or the presentation of novel stimuli that are entirely different from anything previously experienced by the individual can provoke distorted perceptions in most normal people. For example, total sensory deprivation produced in carefully controlled artificial environments may elicit visual and auditory illusions and hallucinations.

Individuals generally exhibit *selective perception* of the world, depending on what is salient at the moment and on their individual memories, emotions, fantasies, and values. Pregnant women are more likely to perceive babies around them than are people who are not as preoccupied with childbearing.

The intensity of perceptions depends on individual sensitivities as well as on mood, anxiety, and substance use. Depressed patients often describe that colors look faded, that the world looks washed out or gray, even though their capacity to recognize specific colors is unchanged. Similarly, mania is often characterized by heightened perceptions, *hyperesthesia*. When extreme, these intense perceptions are uncomfortable. Hyperesthesia can also be seen during benzodiazepine withdrawal, hallucinogen intoxication, and occasionally as part of an epileptic aura.

The intensity of perception may vary with cognitive style and other psychological and neurological factors. Some individuals tend to be *augmenters* and others *minimizers* of bodily experiences. Chronic pain and some hypochondriacal syndromes may occur more commonly among somatic augmenters.

Selective deficits may occur in the perception of emotions. *Emotional aprosodies* have been described in which patients with specific neurological deficits or depression are selectively unable to recognize the expression of facial emotion. These have been linked by position emission tomography (PET) scan to blunted activity in the right prefrontal cortex and insula.

Illusions Perceptual distortions in estimating size, shape, and spatial relations are common even in the absence of psychiatric disorders, especially when one is fatigued or excessively aroused. *Illusions* are misinterpretations of real sensory stimuli, as when a child in a dark bedroom at night sees monsters emanating from shadows on the walls. *Pareidolias* are playful and whimsical voluntary illusions that can be seen when one looks at ambiguously defined or evanescent images, such as flames in a fireplace or clouds. Both the onset and termination of these perceptions are entirely voluntary. *Trailing*, another visual illusion, is the perception that an object moving steadily in space is followed by temporally distinct, after-images of itself. The effect is that of a series of stroboscopic photos. This phenomenon may occur with fatigue and is typically seen with marijuana and mescaline intoxication.

Hallucinations are perceptions that occur in the absence of corresponding sensory stimuli. Phenomenologically, hallucinations are ordinarily subjectively indistinguishable from normal perceptions. Hallucinations are often experienced as being private, so that others are not able to see or hear the same perceptions. The patient's explanation for this is typically delusional. Hallucinations can affect any sensory system and sometimes occur in several concurrently. When perception is altered, illusions and hallucinations, and often delusions as well, are frequently experienced together. Some studies have found that 90 percent of patients with hallucinations also have delusions, and about 35 percent of patients with delusions also have hallucinations. About 20 percent of patients have mixed sensory hallucinations (mostly auditory and visual) that may accompany functional as well as organic conditions. A given external stimulus may evoke very different perceptual distortions in different persons. For example, of three scientists who floated in sensory deprivation tanks for long periods of time one experienced a few illusions and no hallucinations; the second had many illusions and a few faint auditory and visual hallucinations; the third had vivid, dramatic, and complex visual and auditory hallucinations.

Hallucinations are experienced by many normal people under unusual conditions. It has been estimated that between 10 to 27 percent of the general population have experienced memorable hallucinations, most commonly visual hallucinations. *Hypnagogic* and *hypnopompic* hallucinations are common, predominantly visual hallucinations that occur during the moments immediately preceding falling asleep and during the transition from sleep to wakefulness, respectively. Hypnagogic and hypnopompic hallucinations both occur in normal persons and are also characteristic symptoms of narcolepsy. In acute bereavement, up to 50 percent of grieving spouses have reported hallucinating the voice or presence of the deceased, and following amputations, phantom limb hallucinations are common. Patients who become visually impaired often develop pseudohallucinations (i.e., visual hallucinations with preserved insight) with preserved cognitive status, called's *Bonnet syndrome*. These observations suggest a supersensitivity deprivation hypothesis, that when deprived of important and anticipated perceptual stimuli, the mental apparatus may overinterpret any sensory stimulation as evidence of the presence of the needed objects.

A perceptual release theory suggests that hallucinations emerge from the combined presence of intense states of internal arousal and diminished sensory input (including poor attention and poor capacity to sort out relevant from irrelevant input). Thus, diminished input from the environment (as in sensory deprivation) or reduced capacity to attend to and take in the input (as in delirious states) heighten the likelihood that internal sensations, images, and thoughts will be interpreted as originating in the outside environment.

Hallucinations vary according to sensory modality, degree of complexity of the hallucinated experience, the levels of conviction about their reality, the clarity of their contents, the location of their sources of origin, the degree of volitional control over them, and the degree to which the hallucination influences the person's behavior.

Auditory hallucinations range in complexity from hearing unstructured sounds such as whirring noises or muffled whispers to ongoing multiperson discussions about the patient. Simple auditory hallucinations are more commonly associated with organic psychoses, such as delirium, complex partial seizures, and toxic and metabolic encephalopathies. Deafness can produce hallucinations consisting of noises or of formed music. Auditory hallucinations are classically associated with schizophrenia (seen in 60 to 90 percent of patients) but are also frequently seen in mood disorders with psychotic features; 20 percent of manic patients and less than 10 percent of depressed patients experience auditory hallucinations.

Three types of auditory hallucinations commonly associated with schizophrenia (also seen less commonly in patients with psychotic depressions and mania) are: audible thoughts described as hallucinated voices that speak aloud what the patient is thinking; voices that give a running commentary on the patient's actions; and

hearing two or more voices arguing with each other, often about the patient who is referred to in the third person.

> A 23-year-old woman with schizophrenia heard several choruses of angels and “higher beings” who intermittently argued with each other about how she should be spending her time, and what she should do to hasten the arrival of the Messiah on earth. The multitudes of voices also addressed her directly, but the cacophony was often so great that she could distinguish only one or two voices, belonging to the more powerful or influential angels. She ordinarily took their advice and recommendations to heart, but she was quite perplexed by the fact that the angels often could not agree.

Although auditory hallucinations in schizophrenia are frequently mood-neutral, hallucinations in patients with mood disorders are characteristically consistent with their mood. In psychotic depression, the voices may be unrelievedly critical and sadistic; in mania the voices often refer to the patient's specialness.

> A 50-year-old former schoolteacher with bipolar disorder had characteristic auditory hallucinations during each of her episodes of mania and of melancholia. During manias she heard celestial voices praising her and instructing her to start elaborate international businesses. When melancholic she heard accusatory voices telling her that she had deeply hurt, offended, and harmed many of her students by not grading them accurately, and that as a result the FBI was searching for her and was certain to jail and torture her for the rest of her life.

Command hallucinations order patients to do things. Often the commands are benign reminders about everyday tasks: “Pick up your shoes” or “Clean the table.” However, the voices may also be frightening or dangerous, commanding acts of violence toward the self or others, such as “Jump off the roof, you're not worth anything,” or “Pick up the knife and kill your mother.” These voices vary in insistence and persistence, and patients differ in their capacities to ignore these commands. Patients with marked passivity may be helpless in the face of command hallucinations, and may feel impelled to carry out the orders. Even though one study did not find command hallucinations to be associated with a higher risk of harm to the patient or others, the presence of command hallucinations and the patient's ability to resist must be assessed carefully.

> A young man with schizophrenia heard an insistent voice ordering him to attack his mother with a kitchen knife because she was really an agent of the Devil. He was terrified, and told his mother and his psychiatrist about the voices, assuring them that he was aware that the voices were bad, and that he could resist them. When he stopped taking his medications for a few weeks, he felt that the voices become stronger, more insistent, and was less able to resist obeying them. At one point, immediately after telling his mother about his great anguish in fending off the voices, he grabbed a large kitchen knife and started to slash his own arm in an effort to deflect an attack on her. He was hospitalized and re-medicated, as a result of which the intensity of the voices abated, although they remained constantly in the background.

Visual hallucinations occur in a wide variety of neurological and psychiatric disorders, including toxic disturbances, drug withdrawal syndromes, focal CNS lesions, migraine headaches, blindness, schizophrenia, and psychotic mood disorders. Although visual hallucinations are generally assumed to characteristically reflect organic disorders, they are seen in one quarter to one half of schizophrenia patients, often but not always in conjunction with auditory hallucinations.

Visual hallucinations range from simple and elemental, consisting of flashes of light or geometric figures, to elaborate visions, such as a flock of angels.

Stimulation of one sensory modality sometimes evokes perceptual distortions in another. Marijuana and mescaline intoxication, for example, have been associated with *synesthesia*, an experience in which sensory modalities seem fused. This is also a normal experience for many people. Music may be experienced visually, the sound fusing with visual illusions; a tactile sensation may be experienced as a color (e.g., a hot surface may feel “red”).

In certain religious subcultures visual hallucinations may be experienced as normal. In one fundamentalist Pentecostal Church, worshipers danced themselves into a frenzy and, without using any drugs, several participants shared visions of the Virgin Mary at the altar.

During a period of great personal turmoil, a 24-year-old Hispanic woman with great religious conviction and cluster B personality traits, was praying in church when she noticed the Madonna and a host of female angels all smiling at her. She felt as if she were being graced, and experienced a profound sense of peace and relief. On subsequent visits to the same church, these visions returned and were always comforting to her.

Autoscopic hallucinations are hallucinations of one's own physical self. Such hallucinations may stimulate the delusion that one has a double *(Doppelgänger)*. Reports of near-death out-of-body experiences in which individuals see themselves rising to the ceiling and looking down at themselves in a hospital bed may be autoscopic hallucinations. In *Lilliputian hallucinations*, the individual sees figures in very reduced size, like midgets or dwarfs. They may be related to the perceptual distortions of *macropsia* and *micropsia*, respectively the perceptions of objects as much bigger or smaller than they actually are.

Haptic hallucinations involve touch. Simple haptic hallucinations, such as the feeling that bugs are crawling over one's skin (*formication*) are common in alcohol withdrawal syndromes and in cocaine intoxication. When unkempt and physically neglectful patients complain of these sensations, they may be caused by the presence of real physical stimuli such as lice. Some tactile hallucinations, having intercourse with God, for example, are highly suggestive of schizophrenia, but may also occur in tertiary syphilis and other conditions, and may in fact be stimulated by local genital irritation. *Olfactory* and *gustatory hallucinations*, involving smell and taste respectively, have most often been associated with organic brain disease, particularly with the uncinate fits of complex partial seizures. Olfactory hallucinations may also be seen in psychotic depression, typically as odors of decay, rotting, or death.

The term *pseudohallucination* has been used in two ways. First, pseudohallucination refers to perceptions experienced as coming from within the mind (i.e., not at the boundary or outside the mind). Using this definition, loud voices that are alien, ascribed to other beings, but that the patient knows are actually within the mind rather than out in space, are pseudohallucinations. The term has also been used to describe hallucinatory experiences whose validity the patient doubts. A better term for this second phenomenon is *partial hallucination*, analogous to *partial delusion. Functional hallucinations* are rare hallucinations that occur only in connection with a specific external perception, for example, in the presence of a sound such as running water, or a color, or a particular place. However, unlike illusions, the hallucinated sounds are not elaborations of the perception but are simply triggered only in that specific context.

> A 25-year-old farmer with schizophrenia told of a talking tree on his property. During previous episodes he had experienced a variety of auditory hallucinations that were generally well controlled with medication. However, each time he came near to this large, old tree, he would hear a profound, wise voice—as if the tree were one with the earth and the universe, and had important guidance for him. He often came to the tree when he was troubled, seeking the hallucinatory experiences.

Ictal hallucinations, occurring as part of seizure activity, are typically brief, lasting only seconds to minutes, and sterotyped. They may be simple images—such as flashes of light—or elaborate ones, such as visual recollections of past experiences. During the hallucinations the patient ordinarily experiences altered consciousness or a twilight sleep.

Migrainous hallucinations are reported by about 50 percent of patients with migraine. Most are simple visual hallucinations of geometric patterns, but fully formed visual hallucinations, sometimes with micropsia and macropsia, may also occur. This complex has been called the *Alice in Wonderland syndrome* after Lewis Carroll's descriptions of the world in *Through the Looking Glass*, which mirrored some of his own migrainous experiences. In turn, these phenomena closely resemble visual hallucinations induced by psychedelic drugs such as mescaline.

A *flashback* is an intense visual reexperience of highly charged past events, which are often replays of hallucinations. They are typically associated with heavy use of hallucinogens such as LSD and mescaline and often occur months after the last drug ingestion. The images may be simple or complex geometric patterns or they may consist of previously experienced elaborate drug-induced hallucinations. Flashback phenomena may be state-dependent. For example, visual hallucinations initially experienced with hallucinogens are more likely to be subsequently experienced as flashbacks when the subject is smoking marijuana. In posttraumatic stress disorder, some complex intrusive flashback-like images may attain an hallucinatory vividness. Images often include horrifying memories of traumatic events that may force themselves repeatedly into consciousness until they are acknowledged and worked through.

A 35-year-old man with a history of polysubstance abuse and who constantly smoked marijuana estimated that he had used hallucinogens including LSD and mescaline more than 100 times before having a series of devastatingly frightening hallucinatory experiences of devils, of his body being consumed and eaten by wild animals, and of burning in hell-fires. These were accompanied by such profound paranoia and panic attacks that he swore off “heavy drugs,” but continued to use alcohol and marijuana. Several months later, during a period of personal crisis during which he smoked an unusually large amount of powerful marijuana he suddenly reexperienced the worst devil-filled flashback; this experience lasted for several hours in spite of the efforts of several of his friends to talk him down.

Hallucinosis is a state of active hallucination occurring in someone who is alert and well oriented. This condition is seen most often in alcoholic withdrawal, but it may also occur during acute intoxications and other drug-mediated states.

A 30-year-old woman being treated for a depressive disorder with a monoamine oxidase inhibitor snorted cocaine at a party. For the next 3 days she described vivid hallucinatory experiences while in an alert state. She managed to drive her car throughout this time, although with some difficulty. In her psychiatrist's office she alternated between relating coherently to the psychiatrist and responding to her dreamlike complex visual and auditory hallucinations. These phenomena abated within 4 days.

Body Image Distortions Body image includes both perceptual and ideational components, and may reflect primarily perceptual distortions or combinations of disturbed perception and self-appraisal. Body image disturbances can occur as normal responses to abrupt changes in the body (e.g., following amputation), in brain disease, and in psychiatric disorders. Phantom-limb phenomena are classic body image problems in which an amputated limb is still felt to be present. The sensation may diminish gradually over time; the patient feels as if the phantom is receding into the stump

Agnosias, lack of awareness of some parts of the body, may accompany brain damage, most often of the nondominant parietal lobe. Patients with obvious motor or sensory deficits may deny that any deficit exists at all *(anosognosia)*, or the denial may be limited to half of the body *(hemiagnosia)*, usually the left side. In *hemidepersonalization syndromes*, a less common disorder *(hemisomatognosia)*, patients feel that one of their limbs is missing, again usually on the left side. Body image distortions in which a limb feels too heavy *(hyperschemazia)* or weightless *(hyposchemazia)* can occur as a consequence of neurological conditions such as infarction of the parietal lobe. In *duplication phenomena*, patients feel as if part or all of them has doubled (e.g., that they have two heads or two bodies). These rare phenomena may occur in schizophrenia, complex partial seizures, and migraine.

Dysmorphophobia refers to conditions in which patients distortedly perceive and intensely dislike the shape of a particular body part. As such, these symptoms are misnamed because there is no true phobic component, such as fear or avoidant behavior. Fine lines exist between perceptual distortions and realistic but unhappy appraisals of one's body, given the high social value placed on physical appearance. Dysmorphophobia may occur in the context of some personality disorders or as an isolated disorder, called *body dysmorphic disorder*. In some ways, dysmorphophobia resembles an overvalued idea. Patients may develop dysmorphophobias in relation to any body part; common concerns are hair, breasts, penis, nose, or the entire body. For some, changing the body part, as in rhinoplasty for those who do not like their noses, seems to effect a lasting positive change in body image, with patients becoming happier with themselves and feeling more attractive for years or a lifetime. Patients with severe dysmorphophobia may undergo multiple plastic surgeries and feel dissatisfied with every result. At times, the condition forms part of a larger and more pervasive syndrome, such as anorexia nervosa.

A 24-year-old engineering student from a rigid, devout, and loving home was convinced that his mild pectus excavatum condition was an atrocious deformity that accounted for his never having had a girlfriend. A mild deformity did exist, but his reaction to it was far in excess of the actual problem. He was embarrassed to take showers in the dorm, afraid that other students would see him and make fun of his deformity. He sought the services of a surgeon to fix the deformity; the surgeon sent him for psychiatric consultation prior to performing the surgery. No other psychiatric difficulties were evident, and the results of psychological testing were nonrevealing. His father, a rather literal-minded man, was in full agreement with the son's desires to have the corrective surgery. With no clear contraindication, the surgeon agreed to perform the operation. A 6-month follow-up revealed that the student was much happier, and was now dating for the first time in his life.

Hypochondriacal complaints also combine perceptual and ideational distortions. Selective hypervigilance to bodily sensations may result in a higher likelihood of perceptions of unpleasant and potentially pathological body experiences among the worried well, hypochondriacal populations, patients with somatization disorder (Briquet's syndrome), and some patients with a panic disorder.

Body image distortions may at times be severe or bizarre. Some psychotic patients with schizophrenia or depression develop somatic delusions. In depression, this often expresses itself as a delusion that part of the body or the entire body is rotting or cancerous. Some culture-bound syndromes in non-Western culture express themselves with body image distortions, such as koro, in which the man fears that his penis is shrinking into his abdomen.

DISTURBANCES OF MOOD

Defining, describing, understanding, and categorizing moods has long been among the most important and difficult tasks in psychiatry. The language of feelings is filled with terms that seem to have mostly idiosyncratic meanings as patients, phenomenologists, and psychiatrists all struggle to describe inner emotions and to correlate them with external behavior. Even basic terms such as *mood, affect, emotion*, and *feelings* lack universal definition. The most common convention, used here, defines *mood* as a sustained or prevailing subjective feeling tone or range of tones. *Affect* is the moment-to-moment feeling state, sometimes rapidly shifting in response to a variety of thoughts and situations, which the clinician can observe. *Emotions* have been defined as moods and affects that are connected to specific ideas, or to the physical concomitants of moods and affects. *Feelings* are the most poorly defined of all, leading Karl Jaspers to ultimately describe them as everything for which there is no other name. In common parlance and often professionally as well, these words are sometimes used interchangeably.

Moods, affects, and emotions can be described by a number of important qualities: intensity (shallow to deep); range (broad to narrow or flat); stability (rigid to labile); reactivity to external events (none to much); periodicity (periodic to aperiodic); congruence with thought content (congruent or appropriate to incongruent); speed of resolution (rapid to slow); and viscosity (evanescent to persistent). The individual's lifelong predominant mood is one component of temperament. Thus, for example, one may be described as having a calm, buoyant, irritable, depressive, anxious, or sensitive temperament.

Moods, affects, and emotions serve as internal and external signal systems. They signal the state of the individual to others, and often elicit the necessary help and support from the environment. A baby's face communicates its state of need, tension, or contentment, thereby recruiting appropriate parental interventions. As adults, much of the most important interpersonal communication is transmitted nonverbally through cues that signal the observer about our moods. Positive words communicated by a scowling or sullen face will lead listeners to perceive an angry message, regardless of the spoken words. Moods also have an infectious quality and serve as important ways of influencing others. A cheerful mood towards others influence their moods toward cheerfulness; they in turn are more likely to reciprocate that cheerfulness.

Internally, moods, affects, and emotions let individuals know how well or how poorly they are doing, allowing them, for instance, to gauge the distance between actual self-appraisal and desired self-expectations. For example, individuals who desire to master important goals and feel that they have a reasonably good chance of doing so will ordinarily experience pleasant emotional states in relation to these goals. If something intervenes to prevent them from reaching these goals, so that there is an insurmountable gap between their desires and the likelihood of success, they may feel hopeless. In addition to serving as signal systems, emotional states of nonspecific tension, arousal, or anger usually imply that some action will be necessary to secure their discharge or release.

Emotional states and their expression are regulated by biological, psychological, and cultural influences. For example, *emotional lability*, characterized by rapidly shifting emotions that seem unrelated to the situation, typically occurs premenstrually in some women, with varying periodicity in cyclothymic individuals and in those with cluster B personality disorders, and in relation to need states such as hunger, sleepiness, and sexual frustration. Mood shifts have also been related to environment-related physiological influences such as seasonal changes in light. Psychological regulation of emotions may be related to specific coping mechanisms and the ability to self-soothe, which are developmentally determined. Conscious and preconscious psychological mechanisms, including varieties of self-talk, may help to calm or to inflame the emotions. Cultural factors significantly regulate emotional expression. Although the facial expressions for basic emotions are similar in all cultures studied, the range and style of emotional expression permitted in relation to specific contexts varies greatly from culture to culture, and from family to family. Some cultures and families are stiff-lipped and inhibit the open expression of emotion; others encourage emotional display. Marked differences exist among cultures in the emotional expression of acute grief, fear, pain, and affection.

Depression The term *depression* has been used variously to describe an emotional state, a syndrome, and a group of specific disorders. When seen as part of a syndrome or disorder, depression has autonomic, visceral, emotional, perceptual, cognitive, and behavioral manifestations, as illustrated in Table 8-1. As a nonpathological ubiquitous mood state lasting from hours to days, but sometimes longer, feelings of depression are synonymous with feeling *sad, blue, down in the dumps, unhappy*, and *miserable*. Depressed mood is common and appropriate following a disappointment or loss. For most people, innate psychological resilience, coping options, and supportive social networks help to alleviate these brief depressive states and prevent them from becoming chronic. Some individuals suffer from chronically depressed mood, tend to view the world as a difficult place, filled with obstacles and burdens, see themselves as victimized, and lack hope for the future. The extent to which constitutional, developmental, and ongoing aversive life events contribute to this pervasive worldview is unknown. Persons who in early life were deprived and traumatized may be less resilient and more prone to chronic depressive features than are others. Repeated failures and the impact of unrelenting, uncontrollable, and unpredictable negative life events may set the stage for learned helplessness in humans just as they do in animals. A subset of chronically depressed individuals may also suffer from temperamental, biologically driven depression, often seen in conjunction with strong genetic loading for severe mood disorders.

Some depressive states are normal and common reactions to major, unwelcome, and undesirable life events. Normal *bereavement* best exemplifies this. In bereavement following major losses such as the death of a parent, spouse, or child, people experience sadness, pining, and yearning, but do not ordinarily have the feelings of guilt, unworthiness, and self-reproach that characterize depressive disorders. Feelings of helplessness and hopelessness may be temporarily present in bereavement, but they ordinarily pass with time. In uncomplicated cases, the process of bereavement takes 3 to 6 months in the acute phase, and up to a year for complete resolution. Bereaved persons are more likely to feel physically ill and seek general health care than at other times, and older widowers are more liable to die than age-matched nonbereaved controls. *Pathological grief reactions*, bereavements that last more than a year, may be seen when the surviving spouse was excessively dependent on the deceased and is unable to obtain emotional and practical (e.g., financial) support elsewhere, or when the survivor is unable to grieve fully because of markedly ambivalent feelings towards the deceased. The inadequate expression of grief because of incomplete bereavement is thought to be pathogenic in many subsequent psychiatric disorders. For example, impulsive acting-out behavior among adolescents who have lost a parent is often assumed to result from unresolved grief.

A variety of medical disorders may cause depressive syndromes. Most common among these are endocrine abnormalities such as hypothyroidism and hyperparathyroidism and CNS disorders such as cerebrovascular diseases and Parkinson's disease. Depressions are more common in strokes affecting left anterior lesions than other locations. Some medications, especially antihypertensive agents affecting adrenergic tone such as reserpine (Serpasil) and possibly beta blockers may also trigger depressions. The importance of a genetic diathesis in these iatrogenic depressions is not yet known. Depressive syndromes and disorders in general, however, are unquestionably familial and are likely to have genetic contributions, especially in depressions associated with bipolar I disorder.

Cognitive features of depression are prominent. Characterizing the exact nature of the memory impairment using standardized tests has been difficult. Cognitive tasks requiring sustained effort and elaborate cognitive processing may be more disrupted in depression than tasks that can be accomplished more automatically. The so-called *cognitive triad of depression* consists of pervasive cognitive schema related to feelings of worthlessness, helplessness, and hopelessness—expectations that no one and nothing can or is likely to help now or in the future: “I'm not OK, the world is not OK, and it's never going to get any better.” In geriatric populations the effect of depression on cognition may be so profound as to produce a true dementia syndrome, often called *pseudodementia*, a misnomer because the dementia is real but reversible.

Suicidal phenomena are of particular concern. Suicide is common in severe depressive disorders, with 15 percent of untreated depressed patients ending their lives in suicide. Depressed patients comprise the largest diagnostic group of all completed suicides. However, suicide occurs at high rates in many other conditions as well, notably substance-abuse disorders, schizophrenia, and personality disorders. Suicide may occur in these conditions with or without a diagnosable comorbid depressive disorder. Depressed patients with co-morbid alcohol abuse may be at particularly high risk for suicide.

Although consistent, useful, validated predictors of suicide do not exist, certain demographic features are associated with higher risk. These include being white, male, older, and living alone. The single most important factor in the psychiatric history is that of past suicide attempts. A history of violent behavior may also predict suicide. Murderers have a very high suicide rate, especially those who murder family members during episodes of domestic violence. Among clinical signs, hopelessness, anhedonia, and severe anxiety may predict increased suicide risk. Serious physical illness in association with other risk factors such as depression may place a patient at higher risk. A genetic predisposition towards suicidal behavior cuts across diagnostic lines and plays a role in suicide risk. This may reflect a tendency towards impulsive behavior, correlating with low CNS concentrations of 5-hydroxyindoleacetic acid (5-HIAA), the major metabolite of serotonin.

Suicidal gestures are also common among impulsive, dependent, and self-hating depressed persons, for whom they serve as tension-releasing behaviors and as cries for help that may enlist desired social support. Because such gestures have been associated with an increased risk for subsequently completed suicide, they should not be taken lightly. Nonsuicidal self-destructive behaviors such as self-mutilations and repeated unnecessary risk-taking are also common in depressive syndromes and in personality disorders. Subintentional suicide may result when suicidal gestures go awry or when reckless behavior, such as taking unnecessary risks in combat or driving while drunk, prove fatal.

Elated Moods Elated moods include euphoria, elation, exaltation, and ecstasy. They are marked by feelings of well-being and expansiveness, optimism, capability, pleasure, and grace. Such moods are normally experienced when life is going very well, when long-sought-after goals are achieved, and in states of love, religious fervor, and spiritual transcendence. Peak experiences and experiences of mystic fusion are often accompanied by feelings of exaltation and ecstasy. Sexual pleasure and some chemically mediated states of altered consciousness may also induce these feelings.

Abnormally elated moods are primarily seen as part of manic states and from the effects of certain medications and street drugs. When subtle, as in *hypomania*, the mood can be ebullient, and brimming with self-confidence, but with occasional irritability. Other characteristic symptoms of hypomania are increased energy, decreased need for sleep, rapidly flowing thoughts, excessive talking, inflated self-esteem with a demanding nature toward others, and diminished judgment. *Mania* is a more extreme state in which judgment and sleep are impaired to the point of marked functional disruption. As the mania exacerbates, irritability and anger increase, alternating rapidly with a brittle expansiveness. Cognitions become increasingly disorganized. Psychotic symptoms, usually involving themes of grandiosity or specialness, occur in 50 percent or more of patients with mania. With increasing escalation of the manic state, thinking becomes very fragmented, psychotic symptoms are more prominent, and the syndrome may appear indistinguishable from schizophrenia. These three manic states—hypomania, mania and the psychotic mania, fragmented manic state are often referred to as stage I, II, and III mania, respectively. Manic states occur in bipolar disorders, substance-induced mood disorders, and mood disorders due to a general medical condition. Such secondary manias may follow specific cerebral insults, accompany systemic disorders, or occur following ingestion of some drugs including amphetamines, antidepressants, bromocriptine (Parlodel), decongestants, isoniazid (Nydrazid), and corticosteroids. Mania is the second most common neuropsychiatric disturbance induced by steroids, occurring in 30 to 35 percent of patients who develop steroid-induced behavioral disorders. Up to 12 percent of patients treated with levodopa (Dopar) and bromocriptine for parkinsonism develop mania. Right hemispheric brain lesions are specifically associated with secondary mania.

Anxiety Like depression, the term *anxiety* refers to a number of different entities—a normal transient feeling, often with adaptive functions, a symptom seen in a wide variety of disorders, and a group of disorders in which the symptom of anxiety forms a dominant element. As a transient, disagreeable emotional state, anxiety may be adaptive, signaling anticipated or impending threat and motivating necessary action. In contrast to *fear*, the emotional state that exists when a source of threat is precise and well known, anxiety occurs when the threat is not well defined.

Patients often find it difficult to describe feelings of anxiety precisely; at its core, however, anxiety is characterized by intense negative affect, associated with an undefined threat to one's physical or psychological self. Patients will use words such as *tense, panicky, terrified, jittery, nervous, wound-up, apprehensive*, and *worried* to describe their sensations. Anxiety is additionally characterized by somatic, cognitive, behavioral, and perceptual symptoms. The somatic symptoms of anxiety are legion and often dominate the subjective symptoms: a partial list includes twitching, tremors, hot and cold flashes, sweating, palpitations, chest tightness, difficulty swallowing, nausea, diarrhea, dry mouth, and decreased libido. Cognitively, anxiety is characterized by hypervigilance, poor concentration, subjective confusion, fears of losing control or of going crazy, and catastrophic thinking. Behavioral symptoms include fearful expressions, withdrawal, irritability, immobility, and hyperventilation. Perceptual disturbances, including depersonalization, derealization, and hyperesthesia (especially hyperacusis), are also common.

Trait anxiety refers to a lifelong pattern of anxiety as a feature of temperament. Individuals with trait anxiety are skittish, hypersensitive to stimuli, and psychophysiologically more reactive than others. In contrast, *state anxiety* refers to episodes of anxiety that are tightly bound to specific situations and that do not persist after the provoking situation has abated. *Free-floating anxiety* is characterized by a persistently anxious mood in which the cause is unknown and in which large numbers of diverse thoughts and events all seem to trigger and compound the anxiety. In contrast, *situational anxiety* occurs only in relation to specific

occasions or external stimuli, as in phobias.

Anxiety symptoms can result from numerous physical conditions as well as from other psychiatric disorders. Many endocrine, autoimmune, metabolic, and toxic disorders, as well as medication adverse effects, are known to generate anxiety. The psychiatrist must differentiate the response of the patient to an underlying condition (i.e., secondary anxiety) from symptoms generated by the primary disorder itself. In psychiatric populations, anxiety symptoms are prevalent among patients with psychotic disorders, cognitive disorders, depression, and substance-related disorders, as well as in the specific anxiety disorders. In patients with schizophrenia, anxiety must be differentiated from *akathisia*, a common and often overlooked syndrome of subjective restlessness, anxiety, and agitation resulting from antipsychotic medication. The coexistence of anxiety symptoms and depression in major depressive disorder is substantial; anxiety symptoms such as anxious mood and irritability are seen in the majority of depressed patients. Additionally, half to two thirds of patients with a panic disorder will experience a major depressive episode during their lifetime. Medication and drug effects—from intoxication, adverse effects, or as part of withdrawal—are also common causes of anxiety. Many patients with severe anxiety become dependent on anxiolytic drugs (e.g., benzodiazepines, and other sedatives) or alcohol for symptom relief. During attempts to discontinue these substances and sometimes during their ongoing use, confusing admixtures of anxiety symptoms, medication effects, and withdrawal symptoms may occur. Although all the anxiety symptoms caused by drug use are also seen in primary anxiety disorders, perceptual disturbances such as depersonalization and hyperaesthesia may be more common in sedative-hypnotic withdrawal than in primary anxiety disorders.

Despite the general observation that anxious patients demonstrate increased startle responses, specific consistent differences have not been found in the physiological hyperreactivity of patients with anxiety disorder versus controls. In part, this reflects individual differences in reactivity among anxious patients; one person may respond to a specific stimulus with increased pulse and blood pressure while another might show changes in the opposite direction. Similarly, although patients who suffer panic attacks but not controls tend to experience panic attacks in response to sodium lactate infusion, other biological measures, especially those reflecting the catecholamine system and thought to reflect central noradrenergic activity, have failed to elucidate the biological underpinnings of panic attacks. Despite the lack of consistent findings, a great deal of evidence suggests that biological factors strongly contribute to the appearance of anxiety disorders.

Psychological Causes From a psychological point of view, anxiety may signal conflict between opposing desires, wishes, or beliefs on the one hand, and major disequilibria generated by negative life events on the other hand. *Role strains*, conflicts between the major social roles that form a person's identity—spouse, parent, child, wage earner, professional, community member—are common sources of anxiety. The more important the conflict and the less obvious the resolution, the greater is the associated anxiety. For example, anxiety symptoms may first emerge when an individual is confronted with an unavoidable unhappy choice, such as between sustaining a marriage or accepting a career advancement requiring a major move that is unacceptable to the spouse. At times, these conflicts may escape conscious awareness: the person may feel anxious but not know why. Anxiety disorders frequently result from a combination of several factors. A person in a work conflict facing an important deadline may try to alleviate initial anxiety symptoms by overworking or ingesting caffeine or amphetamines to keep alert, then become exhausted and fatigued, and ultimately use alcohol excessively to calm down, with each of these elements contributing separately to an anxiety state.

Certain developmental life situations are associated with anxiety: *Stranger anxiety* develops when infants 6 to 8 months old begin to recognize the difference between their mother and others. When children first go to school, mild anxiety symptoms are common; if the anxiety is excessive, *separation anxiety* or *school phobia* may result. During adult life, anxiety often centers around issues of mastery and accomplishment, both in personal and work life. *Performance anxiety*, or *stage fright*, is a specific type of pathological anxiety in which anxiety escalates to panic when public performance is required. In later life, the deterioration of one's body may engender anxiety related to feelings of helplessness and *death anxiety*.

Panic *Panic attack* is a circumscribed episode of severe state anxiety lasting minutes to hours, with symptoms escalating in a crescendo pattern. The subjective experience is one of utter terror, fears that one will die, go crazy, or lose control, accompanied by many of the somatic symptoms of anxiety mentioned above, including severe chest pains, marked shortness of breath, and exhausting fatigue. Individual isolated panic attacks are common, with up to 30 percent of the general population experiencing at least one attack each year. Panic attacks occur more regularly, and typically more severely, as part of panic disorder or in association with other anxiety disorders. Patients with other psychiatric disorders may experience limited-symptom panic attacks, with episodes characterized by less intense anxiety and by fewer and milder physical symptoms, such as isolated paresthesias or difficulty breathing. These limited-symptom attacks may represent aborted full-blown panic attacks that are not further exacerbated by secondary psychological reactions to the initial symptoms.

A 43-year-old Navy Captain began to experience episodes of severe chest pains, profound sweating, and fear that he was about to die. He had gone to the emergency room on several occasions, and each time had a normal electrocardiogram (ECG) and other tests. His capacity to work was impaired, and he started to stay at home rather than drive to work, since all his episodes began while he was driving. Although he had no prior history of significant alcohol abuse, he also began drinking large amounts of wine in an effort to stave off the attacks. The emergency room physician diagnosed panic attacks. A psychiatric interview revealed that the patient's father had died of a heart attack in his mid-40s, and that the patient had never adequately dealt with issues surrounding the death.

Although American psychiatry has segregated panic attacks from other forms of anxiety, assuming categorical, phenomenological, and biological differences, these distinctions are far from universally accepted. Some psychiatrists view panic as simply an extreme form of anxiety, to be understood as part of a continuum of intensity.

Phobias are irrational fears. In an effort to reduce the intense anxiety attached to phobic objects and situations, patients do their best to avoid the feared stimuli. Thus, phobias consist both of the fears and the avoidance components. The fear itself may include all the symptoms of extreme anxiety, up to and including panic. In *specific phobias*, persistent, irrational fears are provoked by specific stimuli. Table 8-5 lists some illustrative phobias. Common specific phobias include fear of dirt, excreta, snakes, spiders, heights, and blood.

Acrophobia	Fear of heights
Agoraphobia	Fear of open spaces
Amathophobia	Fear of dust
Apiphobia	Fear of bees
Astrapophobia	Fear of lightning
Blennophobia	Fear of slime
Claustrophobia	Fear of enclosed spaces
Cynophobia	Fear of dogs
Decidophobia	Fear of making decisions
Electrophobia	Fear of electricity
Eremophobia	Fear of being alone
Gamophobia	Fear of marriage
Gatophobia	Fear of cats
Gephyrophobia	Fear of crossing bridges
Gynophobia	Fear of women
Hydrophobia	Fear of water
Kakorraphiophobia	Fear of failure
Katagelophobia	Fear of ridicule
Keraunophobia	Fear of thunder
Musophobia	Fear of mice
Nyctophobia	Fear of night
Ochlophobia	Fear of crowds
Odynophobia	Fear of pain
Ophidiophobia	Fear of snakes
Pnigerophobia	Fear of smothering
Pyrophobia	Fear of fire
Scholionophobia	Fear of school
Sciophobia	Fear of shadows
Spheksophobia	Fear of wasps
Technophobia	Fear of technology
Thalassophobia	Fear of the ocean
Traumatophobia	Fear of the number 13
Tropophobia	Fear of moving or making changes

Table 8-5 Specific Phobias

Behavioral, psychodynamic, and biological theories have all been advanced as causes of phobias. Some well-known phobias, such as fear of animals, may result either from early traumatic events (developing along the paradigm of classical Pavlovian conditioning), or from displacements of early psychodynamic conflicts. Genetic influences may also play a role in the development of phobias. For example, some individuals with blood-injection-injury phobias, which strongly clusters among biological relatives, may be genetically predisposed by vagal responses to certain stimuli. Animal models also indicate possible biological vulnerability. Some monkeys that have never previously been exposed to snakes panic when placed in the presence of a snake. Because such fear responses obviously have adaptive value, it has been suggested that some human phobic responses also represent exaggerations of adaptive behaviors shaped by evolutionary biology.

Complex phobias, more elaborate than specific phobias, involve fears related to a broader range of situations. *Agoraphobia*, the best known, refers to a fear of open spaces. Current thinking suggests that agoraphobia is usually a secondary reaction to panic attacks. According to this view, individuals who have become terrified of having panic attacks in public retreat to the safety of their own homes, hoping to reduce the likelihood of panic attacks by avoiding places where they were once triggered and where they may feel exposed and embarrassed. Patients with *social phobia* become overwhelmingly anxious and fear situations in which they may be observed. In the limited type, only a few specific situations evoke the fear, such as speaking in public or using a public lavatory. In the general type, broad-based

fears of social situations globally hamper the person's interpersonal life.

Aggression, Hostility, Impulsiveness, and Violence The spectrum of aggressive emotions and behaviors is characterized by heightened vigilance in response to a sense of threat and enhanced readiness to attack. Physiological tone may be geared for a fight. Assertiveness, the adaptive aspect of these emotions, includes sensing that something needs to be done and feeling willing and competent to take constructive action. The manner and extent to which aggressive emotions can be expressed varies from society to society and situation to situation. These emotions are among the most carefully regulated because of their potential destructiveness. Acts of aggression may begin with verbal threats and intimidation and extend from physical bullying and assault to homocide, sometimes including acts of calculated violence and sadism.

Irritability is an unpleasant feeling state characterized by inner unease. Although minor irritability may abate when the cause is eliminated, escalating irritability has been compared to a psychic itch, begging for discharge via anger outbursts, sometimes leading to temper flareups with verbal or physical lashing out. Unlike anger, irritability does not lessen after an outburst. It is diagnostically nonspecific, seen in a variety of physiological states, anxiety, and mood disorders and as a lifelong temperamental quality. Hunger, sleepiness, sexual frustration, and pain are among the physiological triggers commonly associated with irritability.

A 36-year-old woman with bulimia nervosa but no marked mood or anxiety disorder was successfully treated with a selective serotonin reuptake inhibitor (SSRI). During a follow-up visit during which she and her husband were seen conjointly, her husband remarked about her previous irritability, particularly premenstrually, which he saw as a long-standing personality feature that had abated considerably as a result of the treatment. The patient, her husband, and her psychiatrist all attributed the decrease in irritability to the SSRI. The husband, only half-jokingly, remarked “Doc—whatever you decide to do— *never* take her off that stuff—or if you do, warn me a few days ahead of time and I'll clear out of the house!”

Individual differences in the tendency toward experiencing and expressing anger and violence are biological, developmental, and cultural in origin. Some infants are irritable from birth. Subtle early birth injuries and brain anoxia may increase the susceptibility of some persons to be violent. Furthermore, studies of EEG patterns in violent persons show increased abnormalities, especially in those with repeated violence and violence with little or no obvious motive. Soft neurological signs are also seen in violent criminals. Biochemically, low concentrations of 5-hydroxyindoleacetic acid (5-HIAA) in cerebral spinal fluid (CSF) has been associated with a variety of impulsive behaviors such as violent crimes, recurrent fire-setting, and violent suicide attempts.

In support of the hypothesis that an inverse relationship exists between central serotonergic system function and impulsive-aggressive behavior, a recent double-blind placebo-controlled trial demonstrated that pharmacological enhancement of serotonin activity with fluoxetine (Prozac), an SSRI, reduced impulsive aggressive behavior and irritability in subjects in whom such behavior was prominent.

The pathological childhood triad of bedwetting past the age of 6, setting fires, and torturing animals has been associated with subsequent violent behavior in adults. Interpersonally, studies show that violence-prone individuals require more personal space around themselves than other people do. Violent individuals feel threatened when approached too closely, particularly from the rear.

Psychological and social contributions are also strong. Violence in families breeds violence, and battered children often grow up to be battering adults. Cultural norms for the expression of violence differ considerably. In some socioeconomic and ethnic groups violent gangs organize the energies of many adolescent youth. For some, violent behavior is an adolescent socialization pattern necessary to prove one's manhood. Like other social organizations, violent gangs have detailed rules that inhibit and govern the expression of violence. Some unpredictable and unsocialized violent persons, loners, are too violent to be contained even in gangs.

Aggressive and violent behavior is diagnostically nonspecific. Violence in schizophrenia may occur as a consequence of paranoid delusions, in response to command auditory hallucinations, or secondary to passivity experiences. Manic patients and those in mixed states may be violent, often in response to minimal provocation. Violent behavior commonly occurs in patients with antisocial and borderline personality disorders; in the latter the violence is often self-directed as well as other-directed. Violent behavior may occur in epilepsy—although rarely during true ictal periods; in frontal lobe syndromes as a release phenomenon; and in association with abused substances, particularly disinhibiting sedatives such as alcohol or stimulants such as amphetamines and cocaine, which increase irritability, aggressiveness, and paranoia.

Impulsive violence may be provoked by a number of stimuli and situations. Alcohol is perhaps the most common disinhibitor of violence. Intrafamilial violence, the most common setting for homicide, is frequently related to alcohol intoxication. In *intermittent explosive disorder*, violent behaviors typically erupt after a person has ingested alcohol, a phenomenon known as *pathological intoxication*. In these often ferocious outbursts the individual may confront or provoke any potential target for violence, including total strangers and police, but girlfriends, wives, and parents are frequent victims. Patients with episodic dyscontrol commonly have histories of violent sexual behavior including rape and, often while intoxicated, speeding and reckless driving, sometimes chasing down, stopping, and attacking other motorists who they feel “get in their way.”

A 35-year-old man who had a long history of truancy and a poor school record, fighting in school, bar fights, DWIs, and being assaultive while in jail, was seen in a mental health center. Since release from jail, in spite of a diligent attempt on his part to stay out of trouble for the sake of his wife and children, he found himself getting increasingly edgy. He had acquired a job as a short-haul truck driver and had been doing well until the previous day when he felt that someone cut him off on the freeway. In a blind rage he chased down the driver of the other car, attacking him with a tire jack. Luckily, he was restrained by several bystanders and no one was physically hurt.

Temper Tantrums Immature individuals with persistent personality problems may fail to develop mechanisms to inhibit the temper tantrums they displayed as children. Particularly if childhood tantrums produced the desired result, learned tantrum behaviors may persist into adult life. Although such individuals may be pleasant and sociable when life is going well, they lack the capacity to tolerate frustration, and are easily provoked by threats to self-esteem and self-image, and by not having their own way. When frustrated or threatened they may act like bullies, glare, snarl, yell, shout, intimidate, pout, sulk, and sometimes be physically violent.

Displaced Rage When circumstances prevent the expression of rage directly against the persons or institutions provoking frustration, other outlets for aggression are often found. Acts of violence that are either calculated or wanton may result. Cruelty to animals and fire setting may persist as adult forms of destructive behavior. Rape, an act of control, intimidation, terror, and humiliation, may also displace frustrations that are not expressed more adaptively.

Sadism may occur with or without explicit sexual gratification. Calculated cruelty conducted seemingly without anger or emotional arousal may reflect inadequate development of social morality or individual conscience, as in the conduct of torturers and some cold-blooded murderers. In some societies and under specific circumstances at certain times in history, such activity has been socially sanctioned, suggesting at least that some people lack inborn inhibitions against cruelty or violence.

A 45-year-old veteran hospitalized for chronic depression and barely contained aggressive outbursts described how as a teenager he had been employed as an enforcer by a drug ring. He and his partners took pleasure in torturing and then killing men they were assigned to murder who failed to make their drug payments—as warnings to others who might try to avoid paying. What bothered him the most was that after a while he too started to enjoy participating in these activities.

Self-Mutilation For a variety of reasons, in many different cultures, and in many different disorders people commit acts of violence against themselves ranging from body piercing to cutting and burning to auto-amputation. Psychotic patients may perform extremely self-destructive acts short of actual suicide that often have symbolic import, such as enucleating their eyes or castrating themselves. Patients with borderline personality disorders may cut themselves repeatedly with broken glass or razor blades or burn their arms, legs, breasts, or other body parts with cigarettes. Patients typically deny that these acts are meant to be suicidal, but describe the need to feel external pain to mirror internal suffering, to release tension, or to counteract dissociative-like numbness.

A 25-year-old man presented to the emergency room in a deeply suicidal state. Examination revealed his arms to be riddled with cigarette burns of varying ages, from old scars to fresh burns, and he had multiple healing and healed razor and knife cuts on all his extremities. He described how these cuts were not the result of suicidal acts, which he carefully and precisely differentiated from his current mood state, but were the results of rages of self-hatred—he meant to inflict pain upon himself during periods of particular self-hate and frustration.

Trichotillomania is a syndrome of compulsive hair pulling, resulting in bald patches. It is often associated with other self-mutilatory behavior, such as picking the face, nails, or cuticles to the point of infection and bleeding. Trichotillomania may sometimes be related to obsessive-compulsive disorder.

Children with *Lesch-Nyhan syndrome*, a developmental disability syndrome caused by a congenital metabolic abnormality, bite and pick at themselves so compulsively as to do themselves great harm, and routinely require restraint. Occasionally, patients with Tourette's disorder demonstrate compulsive self-harming behavior.

Other Disturbances of Feelings Diminished levels of emotional intensity may be seen in anxiety disorders, mood disorders, and schizophrenia. Mild emotional flattening with blunted ability to feel joy is common in dysthymia. Some patients with narcissistic and borderline personality disorders complain of inner emptiness and pervasive boredom and ennui, without demonstrating diminished affect in interviews. Similarly, patients with prominent depersonalization describe numbed emotions. Pathological levels of *blunt* or *flattened affect*, indicating markedly diminished affective expression in relation to specific thought content, may be seen in chronic schizophrenia (as part of the *deficit syndrome*), some organic mental syndromes, and severe depressions. Although the term *blunted affect* is not classically used to describe the affective flatness of severe depression, it is not always easy to distinguish between schizophrenic and depressive flatness on phenomenological grounds. *Anhedonia*, the lack of pleasurable feelings from activities that ordinarily provide pleasure, is also seen as part of severe depressions or schizophrenia. Chronically psychotic patients often exhibit emotional deterioration in which affective experience and expression is entirely unrelated to thought content. *Inappropriate affect* is incongruency of affective expression and thought content. The patient may display loud and raucous laughter or giggling in relation to bland or sad thoughts, or may show grief without apparent reason. Inappropriate affect sometimes indicates that the thoughts have private meaning for the patients; the emotional expression might make better sense if the private meaning was understood. Inappropriate affect must be distinguished from affective expressions that may actually be appropriate in a given subculture or ethnic group that is unfamiliar to the observer, and from defensive affect, such as the nervous laughter used to alleviate tension or ward off crying. *Affective lability* is characterized by rapid emotional shifts, often within seconds to minutes. It is commonly seen during hypomanic states, late luteal phase dysphoric disorder (premenstrual syndrome), postpartum blues, other states of physiological instability, and in certain personality disorders.

Alexithymia is difficulty identifying, describing, and differentiating between feelings or distinguishing between feelings and physical sensations. Alexithymic individuals often have constricted imaginations and fantasies, are preoccupied with objects and events in the outside world, and have a limited private, personal internal life. When distressed, these patients are simply aware of not feeling well and usually complain of somatic symptoms, leading to frustrating interactions with their physicians who are unable to find physical causes for the presenting physical complaints. Some view alexithymia as a condition in which affect is communicated through somatic language.

DISTURBANCES IN MOTOR ASPECTS OF BEHAVIOR

Motor behavior is normally finely coordinated, purposeful, and adaptive, and necessary activities are usually carried out efficiently. In psychiatric disturbances motor abnormalities can involve generalized overactivity or underactivity or manifest in a wide range of specific disorders of movement.

Overactivity *Restlessness* and *agitation* are diffuse increases in body movement, usually noted as fidgeting, rapid and rhythmic leg- or hand-tapping, and jerky start-and-stop movements of the entire body, accompanied by inner tension. Restlessness accompanies psychiatric conditions of high emotional arousal or confusion such as toxic states, deliria, mania, agitated depressive disorders, and anxiety disorders, as well as many medical disorders such as hyperthyroidism. In some depressive states, agitation is often accompanied by pacing and hand wringing.

Generalized overactivity, in which patients seem to have increased physical energy, is distinguished from agitation by its lack of inner tension and by more purposeful movements. It is commonly seen in mania, hypomania, anorexia nervosa, and as part of attention-deficit/hyperactivity disorder.

In *catatonic* excitement, less common now than in the era before antipsychotic agents, patients exhibit disorganized and overactive behaviors including frantic jumping, thrashing of limbs, and seemingly senseless menacing or attacking behaviors. Such excitement is seen in mania, periodic catatonia, catatonic forms of schizophrenia, and some culture-bound syndromes such as *amok. Confusional excitement* is a state of restlessness and generalized purposeless activity seen in ictal states, some acute intoxications, and deliria.

Decreased Motor Activity Global reductions in motor activity—motor retardation—are seen in a variety of physical disorders such as hypothyroidism, Addison's disease, some infectious and postinfectious conditions including chronic fatigue syndrome and post-polio syndrome, and other fatiguing conditions, as well as in some organic mental disorders, intoxications, schizophrenias, and depressive disorders. Poverty of movement (*akinesia*, or more properly, *hypokinesia*) may occur in schizophrenia and as an adverse effect of antipsychotic drugs. Changes in the voice frequently accompany the reduced motor activity in schizophrenia and depression, with normal inflection replaced by monotonous tone and prolonged speech latency. In stuporous states patients remain immobile although their eyes are open and they are apparently awake.

Conversion reactions are functional, nonphysiological, psychogenic impairments in sensory or motor functions. Common motor forms include various paralyses and pareses, including limb paralyses, ataxias, and aphonias. In *globus hystericus* the patient is unable to swallow. Patients with *astasia-abasia* have marked unsteadiness of gait. Sensory conversion reactions include blindness, deafness, anesthesia, and analgesia. Some hyperesthesias and pain syndromes may also originate as conversion symptoms.

Mutism may result from a variety of peripheral muscle and CNS conditions and from functional disorders. Mutism may occur in profound depression, catatonic states, and conversion reactions. *Selective mutism* is occasionally seen in adjustment disorders and some personality disorders.

Motor Disturbances Many motor disturbances are seen in psychiatric disorders. Some form part of the core symptoms of the disorders; some occur in disorders that, by their nature, bridge neurology and psychiatry (such as Tourette's disorder); others are acute or chronic adverse effects of medications.

Tremor *Tremors*, involuntary oscillating movements of the limbs or head, may occur at rest or with movement. *Physiological tremors*, which are minimal at rest and increase with activity, are characterized by small amplitude and high frequency. They are characteristic of anxiety, fatigue, toxic or metabolic disorders such as caffeinism or hyperthyroidism, and are commonly seen in patients taking a number of different psychiatric medications including lithium (Eskalith), valproate (Depakene), and stimulating antidepressants. *Coarse tremors* with larger amplitude and lower frequency are seen in Parkinson's disease and cerebellar disease. *Asterixis* is a large-amplitude flapping tremor of the hands seen in hepatic disease.

Parkinsonism Parkinsonian symptoms and signs may be seen in psychiatric disorders, particularly in patients taking antipsychotic medications. Symptoms include akinesias with marked decrease in normally spontaneous fidgeting, stiff gait with diminished arm swing, pill-rolling nonintention tremors (which seem to be less common in neuroleptic-induced parkinsonism compared to Parkinson's disease), expressionless soft and monotonous speech, micrographic handwriting, and cogwheel rigidity.

Dystonia Although dystonic movements are seen in many neurological disorders, in psychiatric patients they are almost always secondary to the use of antipsychotic medications (neuroleptic-induced acute dystonia). Dystonic reactions consist of intermittent or sustained muscle spasms, typically of the head or neck. Common varieties include tongue spasms causing dysarthria, torticollis (neck spasm), and oculogyric crisis in which there is a forced upward gaze. Opisthotonus (spasms of paraspinal muscles leading to an arched posture) is seen less often. These reactions are most common in young males and typically occur soon after beginning or increasing the dose of a conventional antipsychotic medication.

Akathisia *Akathisia* is a syndrome of motor restlessness seen predominantly in the context of antipsychotic medications and some antidepressant drugs (neuroleptic-induced acute akathisia). It has subjective as well as motor components. Subjectively, patients experience muscle tension, difficulty finding a comfortable body position, and inability to stop moving; they feel as though they were “jumping out of their skin.” Objectively, akathisia classically manifests by rocking from foot to foot while standing, frequently crossing and uncrossing the legs when seated, and pacing. Sleep may be disturbed because of physical discomfort. Subjective components of akathisia may be difficult to distinguish from anxiety caused by the primary disorder (typically schizophrenia). Rarely, the restlessness and inner agitation becomes sufficiently uncomfortable to provoke acts of violence. In *pseudoakathisia* objective signs of akathisia are present, but the patient denies feeling restless.

A 24-year-old hospitalized man with schizophrenia was given 10 mg of haloperidol (Haldol) daily. Two days after beginning the medication, he became increasingly agitated—pacing around the ward, muttering to himself, and rapidly alternating between sitting and getting up. It was impossible to conduct an interview, in part because he would neither sit nor stand in one place. The staff thought his behavior indicated an escalation of his psychosis, and the dosage of the antipsychotic drug was doubled. Within the next few days the patient became even more agitated and told the staff that he was convinced that a motor inside him would not stop running. (Whether the motor was metaphorical or delusional was unclear.) One of the ward psychiatrists suspected akathisia, lowered the dose of the medication and prescribed anticholinergic medication and a b-adrenergic receptor antagonist to combat the akathisia. Within 2 days the patient became far more relaxed, less restless, and less psychotic.

Tardive Dyskinesia Neuroleptic–induced *tardive dyskinesia* is a movement disorder that occurs only in the context of antipsychotic medication use, occasionally after many months, more commonly after years. The abnormal movements may persist with or without continued medication use or may diminish or disappear over time. The dyskinetic movements occur at rest and can usually be temporarily suppressed volitionally or by purposeful action, distraction, or sleep. The movements are varied. In the most common type, which affects the face, especially the mouth and lips, tongue-thrusting, chewing, lip-smacking, and eye-blinking movements are seen. Another common type is characterized by choreoathetoid movements such as writhing finger motions. In the less common but more severe truncal dyskinesias, the torso moves in thrusting motions, and respiratory dyskinesia is characterized by grunting and irregular breathing patterns. Other tardive (late) syndromes include tardive akathisia and tardive dystonia in which the abnormal movements emerge late in treatment or upon medication discontinuation.

Neuroleptic Malignant Syndrome *Neuroleptic malignant syndrome*, a potentially fatal complication of antipsychotic medication, is characterized by muscle rigidity, fever, diaphoresis, delirium, mutism, and blood pressure abnormalities. Some view neuroleptic malignant syndrome as the most severe end of a spectrum that starts with neuroleptic-induced Parkinsonism, progresses to extrapyramidal syndrome with fever, and then to fulminant neuroleptic malignant syndrome.

Rabbit Syndrome This uncommon medication-induced extrapyramidal syndrome is often misdiagnosed as tardive dyskinesia. It most closely resembles a limited expression of a Parkinsonian tremor. Patients make rapid chewing movements similar to those made by rabbits, ordinarily faster and more regular than the orofacial tic of tardive dyskinesia. The tongue is spared.

Blepharospasm is a rapid and violent repetitive, spasmodic movement of the eyelids. These movements are often an adverse effect of antipsychotic or other medications but are also common in a variety of neurological disorders, including Meige's syndrome and Tourette's disorder.

Tics *Tics* are rapid, repetitive often spasmodic, jerking involuntary movements that serve no apparent purpose. The person may try to disguise or hide the tic in a seemingly purposive movement, and the movement may ultimately be shaped into a mannerism. Tics are the central feature of tic disorders, are associated with other disorders, and may occur as a consequence of stimulant (e.g., amphetamines) use.

Tourette's disorder is characterized by a chronic shifting array of motor and vocal tics. The tics may include grunts, coughs, clicks, or sniffs, while motor symptoms may include eye blinking, tongue protrusions, facial grimacing, hopping, and twitches. Complex tics may merge into complex compulsive behaviors, such as squatting, deep-knee bends, and retracing steps. *Coprolalia*, characterized by sudden verbal outbursts of obscenities, occurs in fewer than a third of Tourette's disorder patients. Mental coprolalia is an associated feature in which obscene words or phrases suddenly intrude into consciousness in an ego-dystonic manner. Obsessive-compulsive symptoms as well as attention-deficit symptoms are also common in Tourette's disorder.

Serotonin Syndrome *Serotonin syndrome* is a disorder typically caused by the combination of two or more medications with serotonergic properties. It is characterized by restlessness, myoclonus, hyperreflexia, diaphoresis, shivering, tremor, and mental status changes, such as confusion.

Motor Disturbances of Schizophrenia Many of the abnormal movements ascribed to tardive dyskinesia and other antipsychotic-induced extrapyramidal syndromes had been described in chronically psychotic patients before the introduction of antipsychotic medications. In one series of 100 patients, the large majority of whom were diagnosed as having schizophrenia, a review of medical records prior to 1955 revealed that abnormal purposive movements were found in 83 percent, mannerisms and tics in 71 percent, abnormal eye movements in 27 percent, abnormal postures or facial movements in 42 percent, and gait abnormalities in 10 percent. A recent study that comprehensively assessed hard and soft neurological signs, including motor coordination, involuntary movements, integrative sensory functions, cognitive functions, and primitive reflexes, found abnormalities in each of these areas in 67 percent of schizophrenia patients and 19 percent of their never medicated sibs but not in a matched comparison group. Levels of abnormalities were correlated in patient-sib pairs. In another study spontaneous dyskinesias were present in 12 percent of schizophrenia spectrum subjects (who met 3 of 4 criteria from the revised third edition of DSM [DSM-III-R] for paranoid, schizotypal, or schizoid personality disorder) but not in controls. These findings suggest that many patients with schizophrenia have neurological symptoms not caused by medications, and that severe psychiatric disorders may have a neurological component as well.

Catatonia refers to a broad group of movement abnormalities usually associated with schizophrenia, but also found in other disorders such as mania, depression, many neurological disorders (especially those involving the basal ganglia, limbic system, diencephalon, and frontal lobes), systemic metabolic disorders, toxic drug states, and periodic catatonia. Catatonic stupor and excitement have already been noted. *Stereotypies* are repetitious, bizarre, seemingly non-goal-directed complex organized gestures or postures that are thought to have private meanings to the patient. Examples include continuously and repeatedly crossing oneself or blessing others in a religious gesture, waving in a stylized manner, and making profane gestures. The stereotypic behaviors commonly seen in children with autistic disorder (constant spinning or rocking) may provide self-soothing, steady sensory input that help the patients reduce the degree to which they are disturbed by the ordinarily unpredictable and uncontrollable stimulation coming from the environment. *Bizarre posturing* may also be seen in catatonia. One patient with chronic catatonic schizophrenia routinely stood for hours on one leg with his arms in the air like a crane. In *echopraxia*, the patient imitates the examiner's movements, and in *echolalia* imitates speech, as if in mimicry. Some catatonic patients exhibit *waxy flexibility*, maintaining unusual postures into which they have been placed for prolonged periods of time. *Negativism* may take the form of refusing to behave in a prescribed manner, or resisting passive movement.

Gait disturbances in patients with psychiatric disorders include a variety of neurogenic gaits consistent with brain disease, intoxications, and medication adverse effects. These include the festinating gait of parkinsonism, spastic and ataxic gaits of neurological disease and psychiatric medications, waddling and reeling gaits associated with intoxications, and the nonphysiological gait disturbances seen in astasia-abasia, a form of conversion disorder. Gait mannerisms include clowning, prancing, military, and effeminate gaits.

Bruxism, chronic teeth gnashing, may occur involuntarily during tension states, or as an isolated occurrence during delta sleep in which it has sometimes been associated with benzodiazepine or alcohol use. In severe cases serious damage to dental enamel and temporomandibular joint pain may occur.

Myoclonus, characterized by focal muscle jerking, can be caused in psychiatric patients by certain medications, such as serotonin-reuptake inhibitors or monoamine oxidase inhibitors. Myoclonic jerks may be difficult to distinguish from tics, but the latter often represent larger muscle groups and more highly organized motor patterns. Myoclonus may be seen at rest but is more obvious during motor activity.

SEIZURE-LIKE BEHAVIORS In addition to the generalized, petit mal, and complex partial seizures seen in some psychiatric patients, a number of nonepileptic seizure-like behaviors must be distinguished. *Breath-holding spells*, generally innocuous impulsive and tantrumlike phenomena, usually occur in small children who hold their breaths during moments of oppositional rage, and who may faint as a result. Jerking or twitching motor movements may occur. *Temper tantrums* in young children may look like seizures, especially to the uninformed observer. The children may lie on the floor, screaming and kicking, and fail to respond to the environment. *Conversion seizures* (*hysterical seizures, pseudoseizures*) must be differentiated from genuine epileptic seizures. Patients retain consciousness, lack abnormal reflexes, and are not incontinent. However, because so many conversion seizures occur in patients who have genuine epilepsy and who know a good deal about the condition, the differential diagnosis is sometimes difficult.

COMPULSIVE BEHAVIORS Compulsive behaviors may occur in relation to everyday activities such as gambling, sexual conquest, shopping, and watching television, or in relation to substances such as alcohol, cocaine, and opioids or food. Other compulsions involve reckless risk-taking behaviors that provide stimulation and dispel dysphoric moods. Sexual compulsive perversions such as exhibitionism and sadomasochism may serve similar purposes. Compulsions are seen in a variety of psychotic and nonpsychotic psychiatric disorders. The cravings that underlie compulsive behaviors are strong motivating forces, and the compulsive behaviors may regulate emotions. As yet unknown similarities may underlie all compulsive and addictive mechanisms.

In obsessive-compulsive disorder the compulsions are ritualized, repetitive behaviors that are performed with the goal of satisfying, neutralizing, and undoing obsessional thoughts. Although intended to decrease anxiety, rituals are never more than transiently successful. The most common compulsions involve checking to make certain that gas jets and faucets have been turned off and that windows and doors are locked, hand washing, repeating certain phrases, counting objects, and

placing objects in a prescribed order.

> A 20-year-old Japanese-American man compulsively touched every electrical socket he passed, washed his hands several times each hour, and spent countless time leaving and returning to his house to check doors and windows. His mother described that these behaviors began when he was 8 or 9, and had been unrelenting ever since.

A variety of other subtle neuropsychological deficits have been reported in obsessive-compulsive disorder, including impairments on measures of spatial working memory, spatial recognition, and motor initiation and execution. In one controlled study these impairments were not seen in comparison patients with panic disorder or a depressive disorder.

LANGUAGE DISORDERS

Communication difficulties may result from disorders of thinking as previously described, abnormal speech patterns in mood disturbances and schizophrenia; or from primary speech-fluency disorders such as stuttering and stammering; disorders of the articulation and speech apparatus; and CNS disturbances involved in hearing and speech generation (aphasias).

Manic patients typically exhibit *pressured speech* in which the speed of word stream is accelerated. If severe, the speech may be garbled, imprecise, and difficult to understand. Patients with psychomotor retardation depression speak slowly, monotonously, and have a long speech latency in response to questions. Patients with schizophrenia may be difficult to understand because of their disorder or because of the dysarthric effect of antipsychotic medication.

Allusory speech is vague, imprecise, and hard to comprehend because too few cues and details are provided for the listener. Such speech may be heard from some patients with schizophrenia, certain personality disorders, or even normal individuals who wish to convey a sense of mystery by just being suggestive, whose suspiciousness causes them to be reluctant to spell things out clearly, or who believe that the listener is more aware of their private codes, meanings, and allusions than is the case.

Speech Disorders *Stuttering* and *stammering* (ordinarily synonymous), refer to disturbances in the rhythm and fluency of speech due to blocking, convulsive repetition or prolongation of sounds. This disorder affects males two to three times as often as females, and there is a high rate of familial transmission.

Aphasias *Aphasias*, impairments of language produced by brain dysfunction, are ordinarily described as being fluent and nonfluent. In *fluent aphasias*, which generally reflect dysfunction in the left temporal and parietal area, patients have a normal or even elevated verbal output, sometimes with logorrhea, but they ignore the social conventions of conversation. They produce many well-articulated phrases with normal prosody, but there is little informational content. The fluent aphasias are further divided according to the extent of comprehension by the patient and the ability of the patient to repeat what the examiner says. The principal fluent aphasias are *Wernicke's aphasia, conduction aphasia, anomic aphasia*, and *transcortical sensory aphasia*.

Nonfluent aphasias are characterized by slow and poor verbal output, difficulty with spontaneous speech, omission of grammatical connecting words, and poor prosody. Patients may produce one-word replies or very short phrases. Brain lesions that cause nonfluent aphasias typically tend to occur in the anterior left hemisphere. The principal nonfluent aphasias are *Broca's aphasia, transcortical motor aphasia, global aphasia*, and the *mixed transcortical aphasias*.

In *aprosodias* the nonverbal aspects of speech, the melody, pauses, timing, stress, accent, and intonation, are impaired. Damage to the right prefrontal region has been associated with *expressive aprosodias*, and damage to the right temporal region and insula has been associated with *receptive aprosodias* impairments in the ability to understand and interpret the prosodic communications of another person. They are analogous to *receptive aphasias*, in which the individual is unable to understand another person's words.

DISTURBANCES IN INTERPERSONAL RELATIONSHIPS

Normal interpersonal relationships include relationships with parents, children, spouses, lovers, siblings, extended-family members, friends, colleagues, coworkers, and members of the larger community. These relationships ordinarily help provide for the satisfaction of basic drives, for affiliative needs, and for finding purpose and meaning in life. Through stable and satisfying relationships, human needs are met for intimacy, including love, sex, and affection; to be cared for and nurtured, provide care, learn, play, relax, dominate, and be productive through mutual effort. Interpersonal relationships are carefully regulated by means of interpersonal signs and signals. The extent to which deviance from these patterns is tolerated in a given relationship varies from behavior to behavior, relationship to relationship, family to family, and culture to culture.

Disturbances in interpersonal relationships may be viewed as characteristics attributable to a single person or as characteristics of an interpersonal system. Individual disturbances are considered to be undesirable or maladaptive personality traits. When these traits are present to a significant extent and interfere with social functioning or cause distress, they may comprise a personality disorder. Disturbances of interpersonal relationships have also been described at a systems level (e.g., as dyadic and family patterns of system disturbance).

Personality Traits and Disorders Personality, variably defined, is the characteristic pattern of an individual's attitudes, behaviors, beliefs, feelings, thoughts, and values—the sum of a person's emotional, cognitive, and interpersonal attributes. Personality traits are the prominent and characteristic features of an individual's personality and do not imply psychopathology. Aspects of personality are present from early life, and personality traits are relatively stable from adolescence onward, consistent across different environments, and recognizable by friends and acquaintances. The term *personality disorder* should be reserved for those consistent patterns of thought, feeling, and behavior that are inflexible and maladaptive. Personality disturbances manifest primarily in interpersonal contexts and can be viewed as interpersonal behavior disorders.

The determinants of personality are multiple and varied, and include innate and early biological, developmental, and environmental factors inside and outside the home. Through learning and the environment, temperamental factors (genetic or constitutional) are shaped into character.

The dimensional approach to personality and personality pathology characterizes individuals along a continuum of traits. Five dimensions of temperament have been described which appear to be somewhat independent and to have strong genetic contributions: *Neuroticism* (highly emotional, reactive, and thin-skinned, contrasting with emotional stability), *extraversion* (contrasting with intraversion), *openness* (contrasting with discomfort with novel experiences), *agreeableness* (contrasting with contrariness), and *conscientiousness* (contrasting with fickleness). These temperamental attributes may have implications for the course of psychotherapies that cut across diagnostic categories.

Another dimension of personality not adequately dealt with in the DSM-IV concerns moral behaviors such as honesty and integrity. The extent to which individuals behave honestly and with integrity differs considerably across individuals and in different situations. Deception and lying are common behaviors that occur in benign forms (e.g., in white lies) and in pathological forms, psychiatrically important in antisocial personality disorder and sociopathic behavior, pathological liars, and malingerers. Deception and lying may be difficult to assess clinically in the absence of additional informants. Studies of nonhuman primates indicate that at least among chimpanzees, deception (equivalent to lying and dishonesty) is relatively common and in some situations adaptive.

Another recently proposed personality typology characterizes personality along three dimensions related to temperamental characteristics presumed to be strongly influenced genetically—*harm avoidance, novelty seeking*, and *reward dependence*. High scores on the three dimensions characterize inhibition and pessimism, impulsive and exploratory behavior, and dependency and sentimentality, respectively. Different personality types can be described according to patterns of scores on the three dimensions. For example, antisocial personalities are characterized by high novelty seeking, low harm avoidance, and low reward dependence, whereas dependent characters have low novelty seeking, high harm avoidance, and high reward dependence.

DSM-IV uses a categorical approach to personality. The large overlap among the DSM personality disorders and the clustering of these personality disorders into three broad groups imply a lack of clear boundaries to the currently defined categories. The three DSM-IV clusters describe odd or eccentric types (*Cluster A*); dramatic, emotional, and erratic types (*Cluster B*); and anxious and fearful types (*Cluster C*).

The odd or eccentric group includes paranoid, schizoid, and schizotypal personality disorders. Patients with these personality disorders have the core traits of being

interpersonally distant and emotionally constricted. People with paranoid personality disorder are quick to feel slighted and jealous, carry grudges, and expect to be exploited and harmed by others. People with schizoid personality disorder lack friendships or close relationships with others and are indifferent to praise or criticism by others. People with schizotypal personality disorder display odd beliefs, engage in odd and eccentric gestures and practices, and exhibit odd speech.

The dramatic, emotional, and erratic group includes borderline, histrionic, narcissistic, and antisocial personality disorders. Patients with these personality disorders characteristically have chaotic lives, emotions, and relationships. People with borderline personality disorder are impulsive, unpredictable, angry, temperamental, unstable in relationships, compulsively interpersonal, and self-damaging with regard to sex, money, and substance use. People with histrionic personality disorder are attention-seeking, exhibitionistic, seductive, and self-indulgent; exhibit exaggerated expressions of emotions; and are overconcerned with physical appearance. People with narcissistic personality disorder tend to be hypersensitive to criticism, exploitative of others, egocentric with an inflated sense of self-importance, feel entitled to special treatment, and demand constant attention. People with antisocial personality disorder are described almost exclusively in behavioral rather than affective or relational terms. They commit truancy, lie, steal, start fights, break rules, are unable to sustain work or school, and shirk day-to-day responsibilities.

The anxious and fearful group includes patients with avoidant, dependent, and obsessive-compulsive personality disorders. Patients with these disorders are characterized by constricting behaviors that serve to limit risks. People with avoidant personality disorder avoid relationships, people with dependent personality disorder avoid being responsible for decisions, and people with obsessive-compulsive personality disorder use rigid rules that preclude new behaviors. People with avoidant personality disorders are hypersensitive to rejection and are reluctant to enter close relationships in spite of strong desires for affection. Those with dependent personality disorders show excessive reliance on others to make major life decisions, stay trapped in abusive relationships for fear of being alone, have difficulty initiating projects on their own, and constantly seek reassurance and praise. Individuals with obsessive-compulsive personality disorders exhibit restricted expressions of warmth, tenderness, and generosity, and also exhibit stubbornness with a need to be right and to control decisions; indecisive at times, they often apply rules and morals too rigidly, to the point of being inflexible.

A characteristic personality disturbance seen with frontal lobe damage is referred to as *organic personality disorder* in the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), and as *personality change due to a general medical condition* in DSM-IV. Its features include irritability, inappropriate jocularity with euphoria, inappropriate socially disinhibited behavior, and impulsiveness. Other patients, with damage to different areas of the frontal lobe, in contrast, exhibit apathy and indifference.

Interpersonal Systems Couples and families have been studied as systems in their own right, and many qualities of these systems have been identified as being clinically important. A scheme for categorizing relational disorders has been proposed for future editions of DSM, but so far no single generally accepted typology of family psychopathology or interactional types has been established. However, elements of marital discord and harmony have been operationalized in several standard marital inventories. Characteristics of couples and families that have received the most attention include the rules of communication, such as those governing the directness or indirectness with which disagreement and conflict are addressed; the manner (organized or chaotic) in which communications are conducted; taboo topics and secrets about which no one can openly communicate; the nature and degree of emotional expression including affection and anger; the cohesiveness, loyalty, and compatibility of members; the nature of the members' shared identities on the one hand and their autonomous development and separateness on the other; the extent to which members treat one another respectfully or take one another for granted and use one another; the distribution of power and decision making among members; the maintenance of generational boundaries (e.g., age-appropriate performance of life roles); and the members' orientation, concurrence, and disagreement about important values involving moral, religious, intellectual, cultural, financial, occupational, and child-rearing issues, as well as aspirations, health practices, leisure activities, and other belief systems.

Despite the lack of an accepted system for describing disturbances in family systems, some common patterns have been identified. Many imbalanced relationships in which one partner largely dominates the other, may remain stable for years (*skewed* relationships). Some couples have chronically unstable relationships with constant overt conflict (*schismatic* relationships).

A characteristic family environment, called *high expressed emotion*, has been identified that defines a relapse-prone family environment in which one individual has schizophrenia, bipolar I disorder, anorexia nervosa, or major depressive disorder. This interactional pattern includes demeaning, intense personal criticism (“You are rotten and lazy”) and emotional overinvolvement with the identified patient. Aspects of overinvolvement can be measured by quantifying the numbers of hours of face-to-face contact and by the extent to which relatives' categorically assert how the patients feel without ever bothering to ask the patients.

Couple and family system difficulties are most likely to erupt during predictable stressful events in the normal family life cycle, such as during the newlywed period; pregnancy and childbearing; difficult or contentious child-rearing; difficulties with parents, in-laws, and other extended family; insurmountable and unanticipated financial or career problems; serious illness or death of a child or relative; the childrens' adolescence; departure of children from the home; infidelity; and separation.

Interpersonal Disturbances in Illness Behavior *Abnormal illness behavior (dysnosognosia)* is a persistently pathological mode of experiencing, evaluating, and responding to one's own health status despite lucid and accurate appraisal and management options provided by a health professional. These behaviors can be considered as interpersonal disorders between patients and health care professionals. Central to all of these behaviors is the adoption of the sick role by the patient, who then engages in characteristic interactions with health care providers—which typically leave both the provider and the patient dissatisfied. Patients with abnormal illness behavior typically seek repeated medical evaluations from a multitude of physicians, often undergoing a series of expensive laboratory tests. At times, the level of complaints provokes unnecessary invasive laboratory examinations or surgeries which, in turn, place the patient at genuine medical risk.

Abnormal illness behaviors may be unconscious or conscious. Unconscious abnormal illness behaviors are those in which the patient believes the symptoms reflect some genuine illness. These behaviors may occur in somatization disorder (in which multiple symptoms and organ systems are affected), conversion disorders, somatoform pain disorder (in which no cause can be found for the subjective level of pain), and hypochondriasis (in which the primary fear is of having a serious disorder). Abnormal illness behaviors in which patients act sick when they are fully aware that they are not include malingering (in which external incentives—usually financial—are the motivating factors) and factitious disorder with physical or psychological symptoms (Munchausen's syndrome). In *Munchausen's syndrome (factícious disorder with predominantly physical signs and symptoms)*, patients repeatedly and compulsively present themselves for medical care with feigned or self-induced illness. These self-induced conditions may be so serious as to ultimately cause death: some patients inject themselves with feces to cause systemic infections that then warrant hospitalization and intensive care. When the self-induced nature of the illnesses is discovered, medical staff members often become outraged at these patients. The patients rarely accept or cooperate with psychiatric care, so few have been adequately studied. Most do not appear to be psychotic, but seem to have a disturbance in personality structure. In *factícious disorder by proxy*, a caregiver, usually a parent, induces illness in a child.

FUTURE DIRECTIONS

Like psychiatric diagnostic classifications, fashions among psychiatric signs and symptoms change. Characteristics once given prominence, such as the bony protuberances of the skull studied by phrenologists a century ago, are no longer accorded much importance, whereas only in the past few decades have newly described clinical phenomena such as family-expressed emotion and alexithymia been appreciated. Because of the shifts in what is considered relevant and the current dominance of biological research, it would be easy to assume that the nuances of clinical, descriptive psychopathology are mostly of historical interest. As long as the ultimate goals of clinical psychiatry are to help patients feel better and function better, attending to patients' subjective complaints with a firm knowledge of clinical descriptors will continue to be vital aspects of the skills of psychiatry.

SUGGESTED CROSS-REFERENCES

The psychiatric interview, history, and mental status examination are discussed in Section 7.2. Additional definitions of typical signs and symptoms of psychiatric illness are included in Section 7.3. Preception and cognition are discussed in Section 3.1, memory in Section 3.4, and classification of mental disorders in Section 9.1.

CHAPTER REFERENCES

Berrios GE, Gili M: Abulia and impulsiveness revisited: A conceptual history. Acta Psychiatr Scand *92:*151, 1995.

Cassad SL, Adami H, Moran M, et al: Spontaneous dyskinesia in subjects with schizophrenia spectrum personality. Am J Psychiatry *155:*70, 1998.

Cloninger RC, Svrakic DM, Przybeck TR: A psychobiological model of temperament and character. Arch Gen Psychiatry *50:*975, 1993.

Coccaro EF, Kavoussi RJ: Fluoxetine and impulsive aggressive behavior in personality-disordered subjects. Arch Gen Psychiatry *54:*1081, 1997.

Committee on the Family—Group for the Advancement of Psychiatry: A model for the classification and diagnosis of relational disorders. Psychiatric Serv *46:*926, 1995.

Costa PT Jr, McCrae RR: Stability and change in personality assessment: The revised NEO personality inventory in the year 2000. J Personality Assess *68:*86, 1997.

Crichton P: First-rank symptoms or rank-and-file symptoms? Br J Psychiatry *169:*537, 1996.

*Cummings JL: *Clinical Neuropsychiatry*. Grune & Stratton, New York, 1985.

D'Esposito M, Grossman M: The physiological basis of executive function and working memory. The Neuroscientist *2:*345, 1996.

Flaum M, Arndt S, Andreasen NC: The reliability of "bizarre" delusions. Compr Psychiat *32:*59, 1991.

George MS, Parekh PI, Rosindky N, Ketter TA, Kimbrell TA, Heilman KM, Herscovitch P, Post RM: Understanding emotional prosody activates right hemisphere regions. Arch Neurol *53:*665, 1996.

Goodman M, Brown JA, Deitz PM: *Managing Managed Care II: A Handbook for Mental Health Professionals*. American Psychiatric Press, Washington, DC, 1996.

Goodwin FK, Jamison KR: *Manic-Depressive Illness*. Oxford University Press, New York, 1990.

Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritze K: Functioning and well-being outcomes of patients with depression compared with chronic general medical illnesses. Arch Gen Psychiatry *52:*11, 1995.

Hilgard ER: *Divided Consciousness: Multiple Controls in Human Thought and Action*. John Wiley, New York, 1977.

Ismail B, Cantor-Graae E, McNeil TF: Neurological abnormalities in schizophrenic patients and their siblings. Am J Psychiatry *155:*84, 1998.

Jaspers K: *General Psychopathology*. University of Chicago Press, Chicago, 1963.

Kendler KS, Karkowski LM, Walsh D: The structure of psychosis: latent class analysis of probands from the Roscommon Family Study. Arch Gen Psychiatry *55:*492, 1998.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. Arch Gen Psychiatry *51:*8, 1994.

Koenigsberg HW, Handley R: Expressed emotion: From predictive index to clinical construct. Am J Psychiatry *143:*1361, 1986.

Lazare A, editor: *Outpatient Psychiatry: Diagnosis and Treatment*, ed 2. Williams & Wilkins, Baltimore, 1989.

*Lyketsos CG, Steele C, Glaik E, Rosenblatt A, Steinberg M, Warren A, Sheppard JM: Physical aggression in dementia patients and its relationship to depression. Am J Psychiatry *156:*66, 1999.

*Nakaya M, Suwa H, Komahashi T, Ohmori K: Is schizophrenic symptomatology independent of the phase of the illness? Psychopathology *32:*23, 1999.

Nemiah J: Alexithymia: Present, past—and future? Psychosom Med *58:*217, 1996.

Oulis PG, Mavreas VG, Mamounas JM, Stefanis CN: Clinical characteristics of auditory hallucinations. Acta Psychiatr Scand *92:*97, 1995.

Pilowsky I: The concept of abnormal illness behavior. Psychosomatics *31:*207, 1990.

Purcell R, Maruff P, Kyrios M, Pantelis C: Neuropsychological deficits in obsessive-compulsive disorder. Arch Gen Psychiatry *55:*415, 1998.

*Rapaport D, editor: *Organization and Pathology of Thought*. Columbia University Press, New York, 1951.

Ratakonda S, Gorman JM, Yale SA, Amador XF: Characterization of psychotic conditions. Use of the domains of psychopathology model. Arch Gen Psychiatry *55:*75, 1998.

Sachdev P, Loneragan C: The present status of akathisia. J Nerv Ment Dis *179:*381, 1991.

*Schneider K: *Clinical Psychopathology*. Grune & Stratton, New York, 1959.

Shapiro D: *Neurotic Styles*. Basic Books, New York, 1965.

Sims A: *Symptoms in the Mind: An Introduction to Descriptive Psychopathology*. Bailliere Tindall, London, 1988.

*Simon GE, Gureje O: Stability of somatization disorder and somatization symptoms among primary care patients. Arch Gen Psychiatry *56:*90, 1999.

Snaith P: Anhedonia: A neglected symptom of psychopathology. Psychol Med *23:*957, 1993.

Sobin C, Sackeim HA: Psychomotor symptoms of depression. Am J Psychiatry *154:*4, 1997.

Stone MH: *Abnormalities of personality: Within and beyond the realm of treatment*. WW Norton, New York, 1993.

Swedo SE, Leonard HL, Garvey M, Mittleman B, Allen AJ, et al: Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections. Clinical description of the first 50 cases. Am J Psychiatry *155:*264, 1998.

Taylor CB, Arnow B: *The Nature and Treatment of Anxiety Disorders*. Free Press, New York, 1988.

*Trabert W: Shared psychotic disorder in delusional parasitosis. Psychopathology *32:*30, 1999.

Yalom I: *Existential Psychotherapy*. Basic Books, New York, 1980.

Yudofsky SC, Hale RE, editors: *Textbook of Neuropsychiatry*, ed 3. American Psychiatric Press, Washington, DC, 1997.

9.1 CLASSIFICATION OF MENTAL DISORDERS

MICHAEL P. BOGENSCHUTZ, M.D., AND H. GEORGE NURNBERG, M.D.

Advances in scientific psychiatry are to a great extent shaped by its system of classification. Systems of classification are fundamental to all sciences, containing the concepts upon which theory is based, and influencing what can and can not be seen. The classification of illnesses *(nosology)* has always been an integral part of the theory and practice of medicine. Classification is necessary to provide a conceptual framework within which to place what is observed, to communicate efficiently about illness states, to allow decisions to be made regarding treatment, to predict outcomes to measure change, and to keep records. Medical diagnostic systems are predominantly open systems of hypothetical compound constructs that are subject to falsification and further scientific inquiry. These systems allow meaningful research questions to be posed, the answers to which in turn affect the diagnostic system.

In medical nomenclature the primary categories of interest are diagnoses. A *diagnosis* may refer to a consistent syndrome whose cause is unknown or varied, a specified disorder of structure or function whose cause remains unknown, a deviation from a specified physiological norm, or the effects of a specified etiological agent or process. This underlying ambiguity has important implications.

Psychiatry is not different from other branches of medicine in these respects; however, psychiatric classification has evolved independently of medicine. At present, the official psychiatric nosology is viewed as separate from the rest of medical nosology, with disorders conceptualized as *psychiatric* or *nonpsychiatric* although it is understood that medical illness can cause psychiatric symptoms and that psychiatric disorders can have biological bases. To the extent that mental disorders are generally typological in nature, with less emphasis on clear boundaries between disorders than on the central core of each disorder, the classification system and specialty of psychiatry are often held as less “medical” or scientific than other branches of medicine. Extensive efforts to correct this perception resulted in a paradigmatic shift from hermeneutic to empirically based approaches, and the development of a nosology intended to increase diagnostic reliability and facilitate research efforts. These changes are embodied in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and its current successor, the fourth edition of DSM (DSM-IV).

HISTORY

Psychiatric illnesses were widely recognized in the ancient world. Melancholia and hysteria were identified in Egypt and Sumeria as early as 2600 BC. In India a psychiatric nosology was contained within the medical classification system of the Ayur-Veda, written about 1400 BC. Both Hippocrates and Plato created systems to classify mental disorders in classical Greece. Although the Hippocratic system was based primarily on empirical observation, the Platonic system was philosophically rooted in rational idealism. However, this distinction between empiricism and rationalism is not absolute. Hippocrates' system included a humoral theory of temperament that could be considered idealist, and Plato's rational thought was partly aimed at explaining observed reality. Classification in classical Rome was based primarily on the system of Hippocrates. The Galenic, humoral system persisted into the Middle Ages in Europe.

During the European Renaissance and Enlightenment, systems of classification came to reflect the belief in the ordered and uniform character of Nature. Classification was an important part of the largely descriptive science of the day. Thomas Sydenham was a leading proponent of the notion of discrete and uniform disease categories. Carolus Linneaus and Francois Boisser de Sauvages attempted to apply the taxonomic methods of biology to medical and psychiatric illnesses, with categories based on observed signs and symptoms. In contrast, a separate current of thought in late–eighteenth-century France urged the development of a theoretical framework for medical nosology.

The tension between paradigms based on observation and theory continued in the nineteenth century. The establishment of asylums allowed more prolonged and intensive observation of patients than had previously been possible. Autopsies became increasingly common and were widely accepted, and attempts were made to find a neuropathological basis for psychiatric symptoms. Emil Kraepelin aimed to develop a “natural” classification in which cause, symptomatology, and course would converge; however, his categories were based largely on similarity of symptoms.

At the beginning of the twentieth century, no universally accepted system of psychiatric diagnosis existed in Europe or the United States, creating confusion and credibility problems. In the United States the initial impetus for developing a classification of mental disorders was the need to collect statistical information for the census. By the 1880 census, seven categories of mental illness were distinguished: mania, melancholia, monomania, paresis, dementia, dipsomania, and epilepsy. The increasing role of government in health care created a greater push for diagnostic uniformity. The first standard psychiatric nosology was produced in this country in 1918 by the American Medico-Psychological Association (forerunner of the American Psychiatric Association [APA]) and the National Commission on Mental Hygiene. It consisted of 22 disorders to be used primarily for gathering uniform statistics in all mental institutions, and was meant to be revised every 5 years. In 1935 the APA collaborated with the New York Academy of Medicine to develop a nationally acceptable psychiatric nomenclature to be incorporated in the American Medical Association's (AMA's) *Standard Classified Nomenclature of Disease*. Because the AMA system focused on severe mental illness, the U.S. military produced a broader nomenclature to incorporate outpatient presentations of World War II veterans in the 1940s. The sixth revision of *Manual of International Statistical Classification of Diseases, Injuries, and Causes of Death* (ICD-6), developed in 1948 by the World Health Organization (WHO), was not entirely satisfactory to American psychiatrists, so the APA developed a variant that was published in 1952 as the first edition of DSM (DSM-I). Heavily influenced by Adolf Meyer's psychobiological view of reactions, this was the first official manual of mental disorders to focus on clinical utility for classification. DSM-II followed in 1968 without substantial change, other than the elimination of the term *reaction* from the diagnoses.

As with DSM-I and II, the development of DSM-III was coordinated with the development of the ninth revision of *International Statistical Classification of Diseases* (ICD-9) and published in 1980 by the APA. It represented a return to a descriptive system of diagnosis, based on explicit operational diagnostic criteria, theoretically neutral, and multiaxial in format. The revised third edition of DSM (DSM-III-R) (1987) and DSM-IV (1994) refined the diagnostic categories based on available empirical data, and proceeded to make the current diagnostic system compatible with that of the current 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) system.

PSYCHOSIS, PSYCHOPATHY, AND NEUROSIS

As eighteenth-century psychiatry widened the boundaries of insanity and encouraged the development of a descriptive psychopathology, the terms *neurosis*, *psychosis*, and *psychopathic* were conflated. Although these terms have essentially disappeared from the current names of individual diagnostic categories, they reflect the intellectual underpinnings of earlier psychiatric thinking, and influenced how disturbed thinking, moods, and behaviors came to be classified. Their usage persists, and more often than not is applied incorrectly in vernacular parlance. The word “psychopathic” in particular, conjures up a range of negative applications.

Psychosis The term *psychosis* was introduced by Ernst Feuchtersleben in 1845 to describe mental illness as a complex interaction of the psychic and the physical, and to denote a subclass of the nervous diseases referred to at that time as the *neuroses*. Confusion soon arose when *psychosis*, *psychoneurosis*, and *psychopathy* all came to mean the same thing. Efforts to clarify the confusion in terminology continued over the remaining half of the nineteenth century. By the beginning of the twentieth century, under the strong influence of Kraepelin, the meanings of *neurosis* and *psychosis* had become reversed. Freud proposed a nonorganic etiology for psychoneuroses. Karl Jaspers and Kurt Schneider continued in Kraepelin's tradition in Germany, so that the psychoses came to represent severe constitutional or endogenous psychiatric illness. The neuroses and psychopathies (now called *personality disorders*) came to represent exogenous reactions or static deviations on a

continuum from normal.

Perhaps more than any other psychodiagnostic construct, the term *psychosis* has been repeatedly misused for political ends. In the social-political context of Germany in the late 1930s, this classification served the ends of the Nazi regime by providing a pseudoscientific distinction or dividing line that could be used for the eugenics program that systematically killed institutionalized patients with mental retardation, schizophrenia, and manic-depressive illness (bipolar I disorder). Soviet psychiatry would later also be used to promote political agendas based on psychiatric nosological constructs.

Although the traditional meaning of the term *psychosis* emphasized loss of reality testing and impairment of mental functioning—manifested by delusions, hallucinations, confusion, and impaired memory—two additional meanings have evolved over the past 100 years. One common use of the term *psychosis* denotes severe impairment of social and personal functioning characterized by social withdrawal and an inability to perform usual household and occupational roles. The second use of the term specifies the extent of ego regression as a criterion for psychosis. As a consequence of various and multiple meanings, the term has lost its precision in current clinical and research practice. It is possible that the term will fail to appear in future official nosological systems, following the trend set in DSM-III for the class of neuroses.

According to the glossary of the American Psychiatric Association, *psychotic* means grossly impaired in reality testing. The term may be used to describe the behavior of a person at a given time or a mental disorder in which all persons with the disorder have grossly impaired reality testing at some time during its course. Gross impairment in reality testing is defined as existing when individuals incorrectly evaluate the accuracy of their perceptions and thoughts, and make incorrect inferences about external reality, even in the face of contrary evidence. *Psychotic* does not apply to minor distortions of reality that involve matters of relative judgment. For example, depressed persons who underestimate their achievements are not described as psychotic, whereas those who incorrectly believe that they have caused actual catastrophes are so described.

Direct evidence of psychotic perceptions, thoughts, and behavior is the presence of delusions or hallucinations not accompanied by insight into their pathological nature. The term *psychotic* is also appropriate when behavior is so disorganized that it is reasonable to infer that reality testing is grossly disturbed. Examples include markedly incoherent speech without apparent awareness by the person that the speech is not understandable, and the agitated, inattentive, and disoriented behavior seen in a phencyclidine psychotic disorder.

The psychotic disorders are those that are predominantly characterized by psychosis. In DSM-IV the psychotic disorders include schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, brief psychotic disorder, shared psychotic disorder, psychotic disorder due to a general medical condition, substance-induced psychotic disorder, and psychotic disorder not otherwise specified. In addition, some severe mood disorders can have psychotic features. A person with a nonpsychotic mental disorder may exhibit transient psychotic thoughts and behaviors. For example, a person with borderline personality disorder may, when under the influence of substantial stress, loss, or substance abuse, come to believe in the reality of hearing voices, of conspiracies, or of government persecution.

Psychopathy As late as the end of nineteenth century, the adjective *psychopathic* meant psychopathological and applied to any form of mental disorder. However, Koch, Gross, Morel, and others were narrowing the concept to apply to less severe forms of pathology that would eventually evolve into the contemporary concepts of personality disorders. Kurt Schneider's publication of “Psychopathic” charted the course by moving the subject away from the realm of the nonspecific “moral” of James Cowles Prichard to abnormal personalities. He conceived of psychopathic personalities not as pathological in a medical sense, but as deviations from the mean or normal condition, although with a somatic basis that might modulate the form of psychoses. Thus, outside the disease model, they could be enduring, episodic, or reactive.

Psychopathic personality became a subclass of the larger group of abnormal personalities. In the conceptualization of discordant adaptation, Eugen Kahn used the term *psychopathic* to designate these conditions as complex states that lay intermediately between mental health and illness. Although Sigmund Freud was less interested in the issue, other psychodynamic and psychoanalytic writers like Daniel Stern and Wilhelm Reich contributed significantly to the understanding of character pathology. Carl Gustav Jung reinvigorated views of personality by shifting the focus away from archaic views of stereotyped behavioral forms to the contemporary perspective of combinations of dimensions and typologies. Ernst Kretschmer proposed typology based on biological speculation that followed Kraepelin's approach to the psychoses. These perspectives led incrementally to the current concepts of personality disorder that, like the current nosological paradigm, are primarily descriptive and directed towards neurobiological explanatory hypotheses. Although *psychopathy* does not appear in the current official nosology, a residue of the concept is retained in DSM-IV's antisocial personality disorder and ICD-10's dissocial personality disorder.

Neurosis William Cullen coined the term *neurosis* in 1784 to signify a broad category of disorders affecting the nervous system. The term is now defined as a chronic or recurrent nonpsychotic disorder, characterized mainly by anxiety, that is experienced or expressed directly, or is altered through defense mechanisms. It appears as a symptom, such as an obsession, a compulsion, a phobia, or a sexual dysfunction. Although not used in DSM-IV, the term is still found in ICD-10.

DSM-III redefined *neurotic disorder* as a mental disorder in which the predominant disturbance is a symptom or group of symptoms that is distressing to the individual and is recognized by him or her as unacceptable and alien (ego-dystonic); reality testing is grossly intact. Behavior does not actively violate gross social norms although it may be quite disabling. The disturbance is relatively enduring or recurrent without treatment, and is not limited to a transitory reaction to stressors. Earlier psychoanalytic developmental explanatory theories have been stripped from the construct. Because of DSM-III's atheoretical stance, there is no defining etiologic factor.

In ICD-10 a class called “neurotic, stress-related, and somatoform disorders” encompasses the following: phobic anxiety disorders, other anxiety disorders (including panic disorder, generalized anxiety disorder, and mixed anxiety and depressive disorder), obsessive-compulsive disorder, adjustment disorders, dissociative (conversion) disorders, and somatoform disorders. ICD-10 also includes neurasthenia as a neurotic disorder characterized by mental and physical fatigability, a sense of general instability, irritability, anhedonia, and sleep disturbances. Many cases diagnosed as neuroses outside the United States fit the descriptions of anxiety disorders as diagnosed by American psychiatrists.

There are no diagnoses called “neuroses” in DSM-IV. For those who still adhere to the term, *neurosis* encompasses a broad range of disorders with various signs and symptoms. Beyond signifying that gross reality testing is intact, neurosis has lost the precision currently necessary for a diagnostic category. In terms of an individual's functioning, it can reflect an intermediate level of impairment in a number of areas.

FEATURES OF NOSOLOGIC SYSTEMS

Nosology is the study and practice of classification in medicine. The basic purpose of classification is data reduction or condensation of information. The “natural” classification is that which predicts the maximum possible number of facts. Although seemingly straightforward, taxonomy raises a number of philosophical and methodological issues.

Nosological systems generally consist of conjectured disorders (diagnostic categories) as well as rules for making the diagnosis of each category. Because they are not isomorphic concepts, the disorder should not be confused with the diagnosis. A disorder (disease, syndrome, or complex) is a hypothetical entity with ontological status. The process of diagnosis is the attribution of the disorder to particular cases. Once this attribution has been made, the individual can be said to “have” the diagnosis but not the disorder. Unless there is a known cause or an infallible definition or both, diagnosis requires the use of a fallible rule or algorithm. Thus, one can have a disorder without having the corresponding diagnosis, and one can have a diagnosis without the corresponding disorder.

The fundamental question of the ontological status of the diagnostic categories is rarely addressed directly in psychiatry. The metatheoretical choice is between realism and nominalism. *Realism* assumes that the entity (e.g., the disorder) exists independently of its being named. *Nominalism* asserts that the entity has no reality independent of being named; diagnostic categories are cultural constructs, not independently existing entities.

Every diagnostic system, including DSM-IV, also has an explicit or implicit epistemological position. The epistemological question is, “On what is valid knowledge based?” Depending on the paradigm chosen, (e.g., empirical-descriptive or etiologically based), different answers with different implications will be given to this question.

Value judgments involved in choosing a nosological system concern the primary purposes for which the system is designed. Important considerations include the following: clinical utility; research utility; diagnostic reliability; diagnostic validity; other statistical features; administrative, political, forensic, and economic

consequences; theoretical relevance; and universal acceptability. The choice and evaluation of nosologic systems depends in large measure on the priority given these various factors. For example, a system designed solely for clinical use would emphasize convenience of use and coverage (i.e., what proportion of patients presenting clinically can be given a relevant diagnosis), whereas a system designed primarily for research purposes would place more emphasis on reliability, diagnostic validity, and theoretical relevance.

Various systems of medical diagnosis refer to constructs (i.e., syndromes, disorders, diseases) of different epistemological status. A *syndrome* is a set of signs and symptoms that co-occur at a greater than chance frequency. Diagnoses based on groups of symptoms should at minimum represent syndromes. A *disorder* represents the conjunction of a syndrome with a clinical course. At this level, something is known about the mechanism of the illness as well as its symptoms, although the underlying causes remain obscure. A *disease*, the optimal medical category, is a conjunction of etiology and pathology. In the case of a true disease, the symptoms, pathology and pathophysiology, and underlying causes are all known, as are the relations between them. In medicine, criteria used in classification belong to all three levels: symptoms, mechanisms, and causes. Causes are considered definitional and are the ultimate basis for a natural nosology. Progress in medical understanding is the process of moving from the superficial symptoms to the underlying "real" disease. A precise understanding of syndrome and course furthers the discovery of mechanisms and causes. For example, the separation of Down syndrome from the more general category of mental deficiency facilitated LeJeune's discovery of the underlying etiology: trisomy 21. A similar relation is seen between the disorder dementia paralytica and the discovery of the causative agent of the disease syphilis: *Treponema pallidum*.

The official diagnostic nomenclature in DSM-IV uses the term *disorder* because most of the entities lack the features necessary to warrant the term *disease*. Cause is not specified except for cases of posttraumatic stress disorder, mental disorders due to a general medical condition, and substance-induced mental disorders. Other than some of the dementias, even these disorders lack a specified mechanism; most psychiatric illnesses and many medical illnesses are not diseases in the strict sense of the word.

THEORY OF CLASSIFICATION

Classification Versus Typology In a traditional system of classification, *categories* are defined by a small number of individually necessary and jointly sufficient criteria that can be applied unambiguously. Any case can be determined to belong or not to belong to a given category by applying the rules defining the category to the case. Thus, the boundaries are crisply defined. Individuals within such a category are homogeneous with respect to the defining characteristics of that category. For example, all vertebrates have a spinal column; no other creature has this characteristic. *Hierarchies* are composed of clearly nested sets. All the classes of vertebrates have a backbone, but are distinguished by other characteristics (e.g., the presence of hair and mammary glands for mammals). Every mammal and bird is a vertebrate. The periodic table of elements is another example of such a hierarchical, categorical system. This system works to the extent that the classes of interest have such a structure. Ludwig Wittgenstein showed that many common object categories (e.g., "game") do not have clear defining features, which is true for many psychiatric illnesses as well.

In the typological view, cases are assigned to categories if they sufficiently resemble a typical member (or prototype) of the category. Not all characteristics of the category are necessary for membership; some may be merely correlated with membership. This view permits the existence of borderline cases and heterogeneity within the category. Hierarchical sets need not be perfectly nested; members of a subset need not have all of the (correlated) characteristics of the more inclusive set. The typal category has a clearly defined center but the boundary is indistinct, dependent in borderline cases on judgment or the use of a chosen cut-off point. Typology has been called "the opposite of true classification" because it puts similar things together rather than taking different things apart. It is argued that the prototypal model more closely resembles clinical reality in psychiatry, where borderline cases are common and present the most diagnostic difficulty; studies have shown that this is the way that clinicians conceive diagnostic categories. DSM-IV is basically a prototypal system.

Categorical and Dimensional Systems A fundamental distinction is made between categorical and dimensional systems. Modern psychiatry and medicine aspire to the categorical system of classification, that is, the disorder, and generally the diagnosis, is either present or not present. The patient has or does not have a discrete diagnosis that differs qualitatively from normal. In a pure categorical system all diagnostic criteria are necessary and sufficient to make the diagnosis. Patients with the diagnosis are a homogeneous group in respect to that entity. Categorical systems differ in the extent to which hierarchy is recognized, and the extent to which categories are allowed to overlap. Categories may overlap in two senses. Two different categories may share features (criteria) and they may share members (both diagnoses in the same individual). DSM-IV is a modified categorical system in that diagnosis is dichotomous (either present or absent), but individuals who have a given diagnosis may be quite heterogeneous with respect to the defining criteria. Both forms of category overlap occur in the current psychiatric nosology.

In a dimensional classification system, there are no discrete categories. Individuals are described in quantitative terms along continuous factors that have a (usually normal) distribution throughout the whole population. Because the dimensions are continuous, various intermediate measures between the two extreme poles of the dimensions can be expressed quantitatively. Each individual can be represented on a point in *N*-dimensional hyperspace (*N* represents the number of factors or dimensions in the model) defined by the degree to which each factor is expressed. Pathology represents a statistical deviation from the quantitatively defined mean or norm. Multidimensional systems of personality classification have been used extensively in psychology, and found to be useful in predicting external validity variables. However, because of the long-held medical tradition of conceptualizing disease entities as discrete categories, dimensional approaches have met with less clinical acceptance in psychiatry. Dimensional approaches are found in portions of DSM-IV, such as Axis V and the three clusters of Axis II personality disorders (*odd*, *dramatic*, and *anxious*).

There are other systems, such as fuzzy set theory, that combine features of categorical and dimensional systems. Fuzzy set theory allows cases to belong to sets (e.g., diagnostic categories) to varying degrees. Diagnostic groupings are maintained, but membership varies in degree (from 0 to 1).

Dimensional and categorical constructs can coexist within a system of classification. However, the structure of each needs to be recognized so that it can be used appropriately. For example, in cancer the diagnosis is either present or absent, but when present, the disease is graded along a severity dimension (staging) that has important implications for treatment and prognosis. Similarly, rating scales can be used to measure the severity or change in an illness that is conceived as categorical, for example, the Hamilton Rating Scale for Depression or Brief Psychiatric Rating Scale. Distinguishing categorical and dimensional structures is at present a major challenge for psychiatry, because categorical disease entities may be manifested dimensionally and dimensional entities may appear categorical.

Diagnostic Hierarchy Another important structural consideration pertains to diagnostic hierarchy. Are categories mutually exclusive, or may an individual simultaneously belong to more than one category? In most cases DSM-IV has abandoned the hierarchical structure of the earlier edition, but exceptions still exist. A second issue in hierarchy is whether larger categories of a superordinate level of abstraction are recognized (e.g., "affective disorder"). Such categories are analogous to the genera, families, and so on of biological classification of animal and plant species. Various diagnostic systems differ in the emphasis placed on interrelationships of the categories.

DSM-IV uses as many as three levels of grouping above the individual disorder. DSM-IV states that disorders are grouped into the larger sections "based on their shared phenomenological features in order to facilitate differential diagnosis." In many cases there are also different codes for different types of a single disorder.

Operational Methods of Diagnosis (Algorithms) In practice, illness categories are inseparable from the means of diagnosing them and the diagnostic process should reflect the true nature of the latent structure of the disorder as exactly as possible. For example, the diagnostic algorithm for a categorical construct should return a categorical diagnosis. Such a disorder is conceived as homogeneous, and the individuals assigned to the category should be homogeneous with respect to the diagnostic criteria.

The simplest and optimal diagnostic algorithm is the use of a gold standard. The gold standard is an infallible criterion or test that is pathognomonic for a disorder. For example, Klinefelter's syndrome may be diagnosed unequivocally by histological demonstration of the XXY karyotype. However, there are few true gold standards in psychiatry. Biochemical or genetic etiological markers of psychiatric illness, if found, are possible gold standards of the future.

Lacking a gold standard, categorical diagnoses are often made using a number of fallible diagnostic criteria, each one of which is thought to be correlated with the disorder. Diagnostic heterogeneity is introduced when not all, but a minimum number, of all criteria are necessary and sufficient to make the diagnosis. In a formal prototypal model of classification, the diagnosis is made if the individual represents the prototype closely enough. Although several forms of diagnostic algorithms are used in DSM-IV (Figure 9.1-1), it is predominantly a modified prototypal system. Thus, in DSM-IV's borderline personality disorder where 5 or more of 9 criteria are required to make the diagnosis, a total of 256 possible combinations of criteria can make the diagnosis. This diagnostic heterogeneity affects the validity of the diagnosis. Heterogeneity is inevitable and to be expected in research exploring heritability, biological test measures, treatment outcome, clinical course, and

comorbidity correlates of diagnoses made by such algorithms.

FIGURE 9.1-1 Differential diagnosis of psychotic disorders. (Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistic Manual of Mental Disorders*, ed 4. ©American Psychiatric Association, Washington, DC, 1994.)

This procedure can be thought of as using a scale with a certain cutoff point to make the diagnosis. In the preceding example, the criteria for borderline personality disorder can be thought of as being scored 0 (absent) or 1 (present), and the diagnosis made if the individual scores 5 or higher on the scale. When such a scale is used, by making a number of assumptions it is also possible to use the score as a dimensional measure if the diagnosis is conceptualized in dimensional terms. The score contains more information (in the mathematical sense of the word) than the categorical diagnosis (e.g., knowing that a subject has exactly 7 of 9 criteria for borderline personality disorder is more informative than knowing that he or she meets the criteria for borderline personality disorder). However, if the latent structure is actually categorical, then the dimensional score might contain considerable “noise” that could obscure the true dimensional structure. Dimensional scoring procedures can be much more complicated than this, and individuals may be scored on a number *(N)* of dimensions of interest. Such dimensional categories can always be reduced to categorical diagnoses corresponding to particular regions of *N*-dimensional space.

Psychometric Evaluation of Diagnostic Procedures Where there is a gold standard, it is possible directly to measure the relation of the results of the diagnostic test procedure to the actual presence or absence of the disorder. The *sensitivity* of the diagnosis is the probability that the diagnosis will be positive given the presence of the disorder. The *specificity* is the probability that the diagnosis will be negative given the absence of the disorder. The *predictive value positive* is the probability that the disorder is present given that the diagnosis is present. The *predictive value negative* is the probability that the disorder is absent given that the diagnosis is absent. It must be emphasized that these relationships are not inherent to the diagnostic procedure but must be measured in the context of particular populations, and can vary greatly between different populations.

Reliability refers to the degree to which the results of a diagnostic procedure remain stable across different raters and at different times. Inter-rater reliability is high when two or more raters have a high probability of reaching the same diagnosis for the same individual tested at approximately the same time. *Test-retest reliability* refers to the situation in which rater bias is not an issue (e.g., self-report questionnaires or laboratory tests), but results still may differ when the same test is repeated in the same individual. *Longitudinal reliability* is a measure of the stability of the diagnosis over time. Pearson's *r* (correlation coefficient) and k are frequently used statistical measures of reliability. One reason that reliability is important is that it is a necessary although by no means sufficient condition for diagnostic validity.

Diagnostic categories and algorithms are constructed in the hope that they will be meaningful and useful; this is true for dimensional as well as categorical measures. *Diagnostic validity*, a compound term derived from various realms of experience, is used to describe such meaningfulness. *Face validity* refers to the extent that a category appears to represent a real phenomenon, as judged by experts. *Descriptive validity* refers to the existence of reliably assessed criteria that can identify a category and distinguish it from other conditions, particularly those that are phenomenologically similar. *Predictive validity* is the ability of a conjectured entity to predict external variables. In the prevailing psychiatric epistemology, this is operationalized by the procedures described by John Feighner for establishing predictive validity. These include the study of familial transmission, course of illness, treatment response, and biologic markers for the construct in question. *Construct validity* is the ideal or hierarchically superordinate standard of diagnostic validity. It is defined as the experimental confirmation of hypotheses concerning the etiology and pathophysiology of an illness construct, demonstrating that the category represents a real and naturally occurring entity with a specific pathological mechanism. Construct validity has both convergent (sensitivity) and divergent (specificity) aspects.

Mathematical Models The development of sets of operational criteria and assessment procedures has resulted in improved diagnostic reliability. Because most of the constructs in psychiatric nosology are latent (i.e., not directly observable) and without an infallible diagnostic criterion or definition (i.e., gold standard), refinements to establish diagnostic validity must rely on statistical methods. Consequently, as computers have become more powerful, statistical and other computational models have acquired increasing popularity and acceptance as methods to identify and confirm diagnostic categories and dimensions. Methods for isolating psychiatric disorders and assigning individuals to them are promoted under the concept of numerical taxonomy, the methodological sequel to typology. Exploratory analyses use the results to suggest divisions of patients into groups or taxa. Confirmatory analyses examine the syndromal homogeneity and underlying structure of existing diagnostic groups. If the classification system developed by numerical taxonomy conforms to the existing groups, the nosology can be considered to be mathematically valid.

Numerical taxonomy involves the grouping of items according to their evaluated affinity. These procedures involve the mathematical analyses of a data matrix. *Factor analysis* used to be a common technique to identify component dimensions. It has been largely supplanted by *cluster analysis*, which is complementary and attempts to identify homogenous subgroups of individuals on the basis of some measure of similarity. Testing for multimodality or bimodality in the distribution of scores on a typological variable is another approach to distinguishing groups; for example, bimodality or a point of rarity in distribution between groups is proposed to indicate categorically distinct groups.

To overcome some of the inherent limitations of these approaches, a variety of mixture models have come forward. They essentially posit that patients come from a mathematically describable mixture of probability distributions that can be characterized as true groups. The general mixture model assumes that each subject is described by a number of measurements that are normally distributed within groups. Categorical or qualitative group distinction is demonstrated by the presence of more than one normal component distribution in the distribution of canonical coefficient scores derived from discriminant function. The latent class model is another advanced type of mixture model that assumes that the observed data reflect a mixture of distributions in which indicants are not of normal distribution. For example, grade-of-membership typology utilizes the distribution of the presence and absence of probabilistic indicants in individual subjects to identify degrees of membership in pure types.

All the numerical taxonomic methodologies produce groups or taxa on typological grounds. The critical test of their validity is their ability to predict important clinical variables in external domains of familial transmission, course of illness, treatment response, and biologic markers. Although these mathematical methods may produce taxa with strongly defined centers, further inquiry must proceed to determining clear boundaries to be ultimately compatible with the current idealized medical classification paradigm.

DEVELOPMENT OF DSM-IV

In 1988 the APA announced its decision to revise DSM-III-R, which had been published the previous year. Scheduled for 1992, DSM-IV publication was delayed until 1994. Many believed its publication was premature because there would be insufficient time to accumulate a sufficient database on which to develop DSM-IV. Objections were raised in various sectors. Clinicians and recently graduated resident trainees were resistant to changes in the criteria in which they were trained. Patient advocacy groups were concerned that there would be confusion when diagnoses were changed because of a different diagnostic system. Researchers were concerned about the implications of a new nosological system on the relevance of findings obtained with a different system. The American Psychological Association and other organizations of nonpsychiatric mental health professionals, such as social workers and psychiatric nurses, were critical because they had little, if any, input into the final product that they were required by law to adhere to.

Nevertheless, DSM-IV was published in 1994. The criticism of a lack of replicated research data on which to base DSM-IV was addressed in the introduction to the

manual. It states that the Task Force on DSM-IV conducted a three-stage empirical process prior to publication. The stages were:

1. Literature reviews, conducted to provide comprehensive and unbiased information on which to base DSM-IV diagnostic criteria. DSM-IV notes, however, that for some issues insufficient data were available, and in those cases existing data were reanalyzed.
2. Data reanalysis that consisted of “analyses of relevant unpublished data sets” of criteria included in DSM-III-R. According to the DSM-IV Task Force, this approach made it possible for work groups to “question several criteria sets that were then tested in the DSM-IV field trials.”
3. Field trials that compared DSM-III, DSM-III-R, and ICD-10, and proposed DSM-IV criteria sets. The field trials collected information on the reliability and performance characteristics of each criteria set as a whole and on specific items within each criteria set. Twelve field trials were conducted at more than 70 sites, evaluating more than 6,000 subjects. Whether that is a large enough database on which to base a revision is open to question. Many psychiatrists in various sectors (e.g., clinical, research) believe that insufficient time has elapsed between revisions of DSM to allow replicated research on which to base DSM-IV.

BASIC FEATURES OF DSM-IV

DSM-IV is the current classification of mental disorders; it is used by mental health professionals of all disciplines and is cited in insurance reimbursement, disability deliberations, statistical determinations, and forensic matters. Although there has been substantial criticism of each consecutive version of the DSM, DSM-IV is the official nomenclature and is used throughout this textbook. Table 9.1-1 lists all the disorders as they appear in DSM-IV.

Table 9.1-1 DSM-IV Classification of Mental Disorders*

DSM-IV strives to be neutral or atheoretical with regard to causes. Thus, it attempts to describe what the manifestations of the mental disorders are; only rarely does it attempt to account for how the disturbances come about. The definitions of the disorders usually consist of descriptions of the clinical features.

Specified diagnostic criteria are provided for each mental disorder. Those criteria include a list of features and, in most cases, how many must be present for the diagnosis to be made. The use of specific criteria tends to increase the reliability of the diagnostic process among clinicians.

DSM-IV also systematically describes each disorder in terms of its associated features: specific age-, culture-, and gender-related features; prevalence, incidence, and predisposing factors; course; complications; familial pattern; and differential diagnosis. In cases where many of the specific disorders share common features, that information is included in the introduction to the entire section. Laboratory findings and associated physical examination signs and symptoms are described when relevant. DSM-IV explicitly states that it is not a textbook. No mention is made of causal theories, management, or treatment; nor are the controversial issues surrounding particular diagnostic categories discussed.

DSM-IV provides explicit rules to be used when the information is insufficient (diagnosis to be deferred or provisional), or the patient's clinical presentation and history do not meet the required criteria of a prototypical category (atypical type, residual, or not otherwise specified).

Multiaxial Evaluation DSM-IV is a multiaxial system that comprises five axes and evaluates the patient along each. Axis I and Axis II comprise the entire classification of mental disorders: 17 major groupings, more than 300 specific disorders, and almost 400 categories. In many instances the patient has one or more disorders on both Axis I and II. For example, a patient may have major depressive disorder noted on Axis I and borderline and narcissistic personality disorders on Axis II. In general, multiple diagnoses on each axis are encouraged (Table 9.1-2 and Table 9.1-3).

Table 9.1-2 Multiaxial Evaluation Report Form

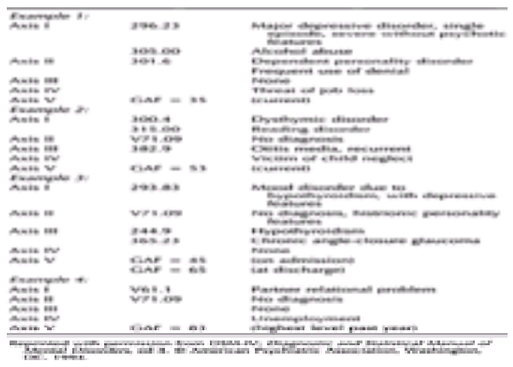

Table 9.1-3 Examples of How to Record the Results of a DSM-IV Multiaxial Evaluation

Axis I Axis I consists of all mental disorders except those listed under Axis II, and other conditions that may be a focus of clinical attention.

Axis II Axis II consists of personality disorders and mental retardation. The habitual use of a particular defense mechanism can be indicated on Axis II.

Axis III Axis III lists any physical disorder or general medical condition that is present in addition to the mental disorder. The identified physical condition may be causative (e.g., hepatic failure causing delirium), interactive (e.g., gastritis secondary to alcohol dependence), an effect (e.g., dementia and human immunodeficiency virus [HIV]-related pneumonia), or unrelated to the mental disorder. When a medical condition is causally related to a mental disorder, a mental disorder due to a general condition is listed on Axis I and the general medical condition is listed on both Axis I and III.

Axis IV Axis IV is used to code psychosocial and environmental problems that contribute significantly to the development or the exacerbation of the current disorder (Table 9.1-4). The evaluation of stressors is based on the clinician's assessment of the stress that an average person with similar sociocultural values and circumstances would experience from psychosocial stressors.

Problems with primary support group
Problems related to the social environment
Educational problems
Occupational problems
Housing problems
Economic problems
Problems with access to health care services
Problems related to interaction with the legal system/crime
Other psychosocial and environmental problems

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 9.1-4 Axis IV: Psychosocial and Environmental Problems

Axis V Axis V is the Global Assessment of Functioning (GAF) scale (Table 9.1-5) with which the clinician judges the patient's overall level of functioning during a particular time period (e.g., the patient's level of functioning at the time of the evaluation or the patient's highest level of functioning for at least a few months during the past year). Functioning is conceptualized as a composite of three major areas: social functioning, occupational functioning, and psychological functioning. The GAF scale, based on a continuum of severity, is a 100-point scale with 100 representing the highest level of functioning in all areas.

Table 9.1-5 Global Assessment of Functioning (GAF) Scale*

Nonaxial Format DSM-IV also allows clinicians who do not wish to use the multiaxial evaluation format to list the diagnoses serially, with the principal diagnosis listed (Table 9.1-6).

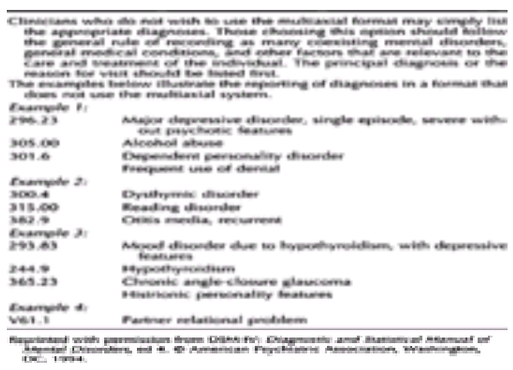

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 9.1-6 Nonaxial Format

DSM-IV's Relation to ICD-10 The DSM-IV was designed to correspond with ICD-10, developed in 1992. There was a strong consensus that diagnostic systems used in the United States must be compatible with the International Classifications of Diseases (ICD) to ensure uniform reporting of national and international health statistics. In addition, Medicare requires that billing codes for reimbursement follow ICD.

ICD-10 is the official classification system used in Europe and many other parts of the world. All categories used in DSM-IV are found in ICD-10, but not all ICD-10 categories are in DSM-IV. In the United States, ICD-10 codes can be used on insurance forms and other documents requiring diagnoses. The codes in DSM-IV are fully compatible with ICD-10 and are listed in the appendix; however, some terms and diagnostic categories used in ICD-10 are not used in DSM-IV.

CRITIQUE OF DSM-IV

In formulating a critique of the DSM-IV, it is important at the outset to establish what the DSM-IV is and is not intended to be. The introduction of the DSM-IV states:

> The utility and credibility of the DSM-IV require that it focus on its clinical, research, and educational purposes and be supported by an extensive empirical foundation. Our highest priority has been to provide a helpful guide to clinical practice. We hoped to make DSM-IV practical and useful for clinicians by striving for brevity of criteria sets, clarity of language, and explicit statements of the constructs embodied in the diagnostic criteria. An additional goal was to facilitate research and improve communication among clinicians. We were also mindful of the use of DSM-IV for improving the collection of clinical information and as an educational tool for teaching psychopathology.

An official nomenclature must be applicable in a wide diversity of contexts. DSM-IV is used by clinicians and researchers of many different orientations (e.g.,

biological, dynamic, cognitive, behavioral, interpersonal, family, systems). It is used by psychiatrists, other physicians, psychologists, social workers, nurses, occupational and rehabilitation therapists, counselors, and other health and mental health professionals. DSM-IV must be usable across settings—inpatient, outpatient, partial hospital, consultation liaison, clinic, private practice, and primary care, and with community populations. It is also a necessary tool for collecting and communicating accurate public health statistics. Fortunately, all these many uses are compatible with one another.

These goals are similar to those of DSM-III and DSM-III-R. A strong emphasis is placed on clinical utility across a variety of settings, consensus, and utility in education and research. Although laudable goals, they are not the only ones that could have been chosen as priorities, and they are not necessarily compatible in all cases.

It is also important to note what the DSM-IV does not state as goals. It does not intend to be a definitive or permanent taxonomy. It does not claim that all the categories are necessarily valid, nor that they be clearly distinguished from each other in all cases. It uses the term *disorder* exclusively, rather than *disease*, indicating that it does not assume that the entities described are diseases.

A common criticism of DSM-IV is that it seems to place consensus and universal acceptability above validity. This is essentially a value judgment supported by the goals quoted above. Although DSM-IV remains unpopular among some clinicians (especially those who are psychodynamically oriented), it appears to have been successful in this goal as evidenced by its position as the standard nomenclature in the United States and in much of the world.

However, criticisms relating to validity remain. The rates of comorbidity, or more accurately co-occurrence, between many of the DSM-IV diagnostic categories are higher than those found in any other branch of medicine. This brings into question whether the categories actually represent discrete entities or rather different aspects of larger symptom complexes. It is also not clear whether many of the criteria sets actually represent syndromes in the statistical sense (i.e., criteria that occur together significantly more often than chance). Although the DSM-IV Task Force made a major effort to render the nosology consistent with existing empirical data, most categories lacked the adequate empirical database for that determination, even at the level of internal (face and descriptive) validity. In addition, the extensive heterogeneity allowed by the polythetic criteria sets and their algorithms limits the utility of many of the diagnoses.

Another criticism of DSM-IV concerns its atheoretical stance. Beyond the inherent paradox of such a position, should a nosological system be atheoretical (i.e., purely empirical)? The goal was to create a document that would be acceptable to health professionals regardless of theoretical outlook; explicit etiological and pathological theories are for the most part eschewed. The structure of the DSM-IV implies that mental disturbances can be characterized as disorders and categorized in a meaningful way based on signs and symptoms, particularly related to cognition and behavior. This in itself constitutes a theoretical stance. Although it may be the best that can be done at this time, the failure to involve developments in theory and cause is a significant impediment to further advances.

The categorical approach adopted by the DSM-IV is familiar and comfortable to those with medical training. However, other than an appeal to tradition, there is little evidence that such a system is more useful or valid than a dimensional or other taxometric system. There is an equally long tradition of evaluating personality traits in a dimensional fashion. Doubts have been raised regarding the categorical nature of personality disorders, psychotic disorders, and mood disorders. DSM-IV clearly acknowledges the potential contributions of the dimensional approach, and contains polythetic and dimensional as well as categorical constructs. Whether it is reasonable to apply a common form to all psychiatric disorders or whether different models, representing different structures, should be applied to different classes of illness remains an open question.

The DSM-IV multiaxial system implies the existence of qualitative differences between Axis I psychiatric disorders, Axis II personality disorders, and Axis III medical disorders but there is little evidence to support this position. The distinction of Axis I and Axis II was originally made in DSM-III to encourage use of the latter diagnoses out of concern that they could be missed because of the established hierarchical rules of the nosology. Hierarchical diagnosis was essentially abandoned in DSM-III-R, and DSM-IV follows that revision. The categorical distinction between psychiatric and medical conditions makes less and less sense as knowledge accumulates concerning the biological bases of psychiatric disorders and the effects of psychosocial variables on medical illnesses. In the absence of supporting evidence, the distinction between Axes I, II, and III appears anachronistic and possibly counterproductive.

In summary, DSM-IV is a practically-oriented, consensus system that incorporates much of current clinical belief and research data. Many of its limitations are direct consequences of its intentionally descriptive approach; others stem from an across-the-board adoption of a quasi-categorical system using polythetic criteria. To its credit, DSM-IV has achieved an acceptable level of validity, primarily in face validity and more modestly in descriptive validity, which has enabled advances to be made in external predictive realms. Most importantly, as an open system, it provides a medium of discourse for clinicians and researchers, and is intended to be replaced in the future by more definitive, research-based nosological systems that incorporate etiology and pathology.

SUGGESTED CROSS-REFERENCES

The psychiatric report is discussed in Section 7.2, typical signs and symptoms in Section 7.3, neuropsychological assessment in Section 7.4, clinical manifestations of psychiatric disorders in Chapter 8, and international perspectives on psychiatric diagnosis in Section 9.2.

SECTION REFERENCES

American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. American Psychiatric Association, Washington, DC, 1994.

*Andrews G, Slade T, Peters L: Classification in psychiatry: ICD-10 versus DSM-IV. Br J Psychiatry *174:*3, 1999.

Beer MD: Psychosis: A history of the concept. Comprehensive Psychiatry *37:*273, 1996.

Berrios GE: European views on personality disorders: a conceptual history. Compr Psychiatry *34:*14, 1993.

Bryant KJ, Rounsaville B, Spitzer RL, Williams JB: Reliability of dual diagnosis: Substance dependence and psychiatric disorders. J Nerv Ment Dis *180:*251, 1992.

Cloninger CR, Martin RL, Guze SB, Clayton PJ: Diagnosis and prognosis in schizophrenia. Arch Gen Psychiatry *42:*15, 1985.

*Frances A, Mack AH, First MB, Widiger TA, Ross R, Forman L, Davis WW: DSM-IV meets philosophy. J Med Philosophy *19:*207, 1994.

Grayson DA: Can categorical and dimensional views of psychiatric illness be distinguished? Br J Psychiatry *151:*355, 1987.

Grove WM, Andreasen NC: Multivariate statistical analysis in psychopathology. In *Contemporary Directions in Psychopathology: Toward DSM-IV*, T Millon, GL Klerman, editors. Guilford, New York, 1986.

Kendall RE, Brockington IF: The identification of disease entities and the relationship between schizophrenic and affective psychoses. Br J Psychiatry *137:*324, 1980.

*Mack AH, Forman L, Brown R, Frances A: A brief history of psychiatric classification. Psychiatr Clin North Am *17:*515, 1994.

*Nathan PE, Langenbucher JW: Psychopathology: Description and classification. Annu Rev Psychol *50:*79, 1999.

*Pichot P: Nosological models in psychiatry. Br J Psychiatry *164:*232, 1994.

*Robbins E, Guze SB: Establishment of diagnostic validity in psychiatric illness: Its application to schizophrenia. Am J Psychiatry *126:*983, 1970.

Scadding JG: The semantic problems of psychiatry. Psychol Med *20:*243, 1990.

Widiger TA, Frances A: The DSM-III personality disorders: Perspectives from psychology. Arch Gen Psychiatry *42:*615, 1985.

Widiger TA: The categorical distinction between personality and affective disorders. J Pers Disord *3:*77, 1989.

Wilson M: DSM-III and the transformation of American psychiatry: A history. Am J Psychiatry *150:*197, 1993.

Woodbury M, Manton KG: A new procedure for analysis of medical classification. Meth Inform Med *21:*210, 1982.

World Health Organization: *The ICD-10 Classification of Mental and ehavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization, Geneva, 1992.

World Health Organization: *International Statistical Classification of Diseases and Related Health Problems*, rev 10. World Health Organization, Geneva, 1992.

9.2 INTERNATIONAL PSYCHIATRIC DIAGNOSIS

JUAN E. MEZZICH, M.D., FH.D., ANGEL A. OTERO-OJEDA, M.D. AND SING LEE, M.D.

The concept of “diagnosis” at any point in time, defines the field of medicine in general and psychiatry in particular. It does so by delineating the informational base necessary for clinical care. With various degrees of systematization and explicitness, diagnostic schemas, as consensual notions and formats for describing clinical conditions, have emerged since the dawn of mankind. In every case, these notions have been embedded within their time and culture.

Building on conceptual contributions over the past two centuries in various parts of the world, and having the 100-year-old International Classification of Diseases (ICD) as general reference, the emphasis for advancing psychiatric diagnosis during the past several decades has been on more-systematic and clearer formulations of psychopathology and nosology. This has led to gains in interdiagnostician agreement (diagnostic reliability) and universal communicability of diagnostic statements. These developments have been widely regarded as propitious (although not guarantees) for diagnostic validity or usefulness and for the broader advancement of the field.

Among recent efforts to update diagnostic validity, clinically and epidemiologically, are developments to enhance existing universalistic diagnostic systems by paying attention to both local realities and the uniqueness of the individual. The first type of these developments involves adaptations of the international classification system to regional or national clinical patterns and needs. The second corresponds to idiographic or personalized formulations.

CONCEPTUAL AND METHODOLOGICAL DEVELOPMENTS

Through proposals arising in different parts of the world, the following have emerged in recent decades as key conceptual and methodological developments for psychiatric diagnosis. They refer to the following major themes: systematic psychopathological description and comprehensive diagnostic formulation.

Systematic Description of Psychopathology Early roots of explicit and systematized psychopathological description can be found in nineteenth-century France when symptoms were first used as units of analysis of abnormal behavior. Current concepts of psychiatric nosology can be traced back to the end of the nineteenth century, highlighted by Valentin Magnan's notion of clinical evolution in France and Emil Kraepelin's dichotomy of the major psychoses in Germany. Other significant contributions to the nosology of severe mental disorders of recent impact on psychiatric classification are the German description of cycloid psychoses, the Scandinavian concept of psychogenic psychosis, and the French delineation of *boufée délirant*. More recently psychopathologists from Asia, Africa, and Latin America have offered informative reports on acute transient psychoses and somatically and psychologically textured characterizations of the neuroses.

The manifestations of mental disorders constitute the focus of the so-called phenomenological description of psychopathology. That approach encouraged careful observation of clinical presentations, particularly symptom profiles, while minimizing etiological inferences. Many questions remain regarding the organization of standard nosologies, for example, the number of major classes of mental disorders, the arrangement of subclasses, and the hierarchical relationships among diagnostic categories. Furthermore, etiopathogenic perspectives—from genetics, to psychodynamics, to general systems—may in the future contribute enriched and more valid formulations of mental disorders.

Explicit or operational diagnostic criteria, a mainstay of modern diagnostic methodology, were persuasively proposed by the British psychiatrist Edward Stengel as a step in dealing with the widespread confusion in classification documented in his international survey, commissioned by the World Health Organization (WHO). The actual development of operational criteria was pioneered, chronologically, by José Horwitz and Juan Marconi in Latin America, Peter Berner in Austria, and John Feighner and associates in the United States. Operationalized diagnostic criteria probably represent the most conspicuous response to the need for clarity in psychiatry and are considered essential for its progress as a scientific discipline. On the other hand, the limitations of using the criteria sets include the arbitrariness often involved in setting thresholds between cases and noncases, the cumbersomeness of their use in daily practice, and the burden they impose on meaning and usage across cultures. An approach that promises to enhance categorical definition by accommodating graded typicality and flexible boundaries involves the use of the prototypical categorization model. This is connected to mathematical *fuzzy set theory*, and its use is being explored on particularly problematic areas of psychopathology, such as personality disorders and even schizophrenia and obsessive-compulsive disorder.

Multiaxial Diagnostic Formulation Attempts to represent more fully the complexity of a patient's condition have emerged predominantly under the generic term *multiaxial approach*. Multiaxial diagnosis intends to portray the intricacies of the clinical condition through the systematic and separate assessment and formulation of highly informative aspects or domains. Standardized measurements, either typologies or dimensional scales, have been proposed with which to appraise each domain.

The first multiaxial schemata in both psychiatry and general medicine were aimed at articulating key components of an illness. Starting in 1949, the pioneers of the field who independently proposed a methodical and almost graphical assessment of syndromes and etiology were Erik Essen-Möller and Snorre Wohlfahrt in Sweden, Maurice Lecomte and associates in France, Tadeusz Bilikiewicz in Poland, and José Leme-Lopes in Brazil.

The first multiaxial schema in general medicine was the Systematized Nomenclature of Pathology, which accommodated axes on topography, morphology, etiology, and symptoms. One of the latest is the International Classification of Diseases for Oncology (ICD-O), which focuses separately on neoplastic topography and morphology (the latter including tumor behavior and differentiation).

A more recent and far-reaching purpose of the multiaxial model is to furnish a biopsychosocial description of the patient's entire clinical condition. This encompasses not only pathologies (mental and nonmental), but also psychosocial environmental factors and the consequences of illness on the individual's functioning and quality of life. The work of John Strauss, Michael Rutter, and associates in psychiatry and of Alvin Feinstein, J. S. House, and associates in general medicine are pertinent to this objective. Among the challenges for further development of the multiaxial model are the need for greater simplicity and ease of use as well as empirical appraisal of its reliability and validity across the world.

The first national diagnostic system to incorporate a multiaxial approach was the Swedish classification of mental disorders, which was based on the previously referred proposals by Essen-Möller and Wohlfahrt. More recently, the methodological developments outlined previously (phenomenological description, explicit diagnostic criteria, and multiaxial formulation) have structured the third and fourth editions of the American Psychiatric Association's (APA's) *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III and DSM-IV), as well as the WHOs 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), which has a central role for diagnosis across the world.

ICD-10

Origins of the International Classification The origins of the ICD can be traced back to the taxonomic work of the Swedish biologist Carolus Linnaeus in the eighteenth century. In *Genera Plantarum* he stated: “All the real knowledge which we possess depends on methods by which we distinguish the similar from the dissimilar ... We ought therefore by attentive and diligent observation to determine the limits of the genera, since they cannot be determined a priori. This is the great work, the important labor, for should the general be confused, all would be confusion.” Linnaeus stimulated scholars from all walks of life. For example, Francois Boissier de Sauvages and William Cullen developed formidable nosologies encompassing thousands of species of disease, organized into classes, orders, and

genera.

Bringing these leads to fruition, the first International Statistical Congress held in Brussels in 1853 commissioned the Englishman William Farr and the Italian Marc d'Espine to prepare a “uniform nomenclature of causes of death applicable to all countries” as a way to obtain comparable health status information across the world. The proposals that ensued were based primarily on either topographical or etiological principles. Accommodating both, Jacques Bertillón of Paris prepared the First International Classification of Causes of Death, adopted at the International Statistical Congress of 1893.

Since then there have been revisions of the ICD approximately every 10 years. From its foundation the WHO, created in 1948, assumed the preparation of these revisions as a constitutional responsibility. The sixth revision contained a critical expansion of the scope of the international classification by covering morbidity in addition to mortality. Correspondingly, psychiatric illness appeared for the first time in the classification with one category: mental illness and deficiency. The ninth revision had as one of its innovations the presentation of a glossary for the capsular definition of mental disorders. While modest in informational detail, this glossary signified recognition by the WHO that the intricacy of mental problems required more than the labels used in all other chapters of the ICD.

Outline of ICD-10 Work toward the preparation of ICD-10 started in 1979, the same year in which the ninth revision of *International Statistical Classification of Diseases* (ICD-9) was put into effect. Its developmental process involved the participation of the eight Collaborating Centers for the Classification of Diseases, specialty divisions (such as Mental Health) at both the headquarters and regional offices of WHO, nongovernmental organizations such as the World Psychiatric Association (WPA), and a miscellaneous panel of interested groups and individuals, all working under the coordination of the WHO Unit on the Development of Epidemiological and Health Statistical Services.

First to be noted in ICD-10 is its expanded scope, as indicated by its title, *International Statistical Classification of Diseases and Related Health Problems*. This expression continues the trend that starting with an original set of causes of death, added morbidity in its fifth revision, and more recently added problems such as disabilities and factors that influence health status, recognizing that more information is needed to deal effectively with the evolving and complex issues of health care and health promotion.

ICD-10 uses an alphanumeric code composed of a letter followed by several digits. That arrangement more than doubles the number of available categories. Splitting, rather than lumping, of categories has marked the progression of ICD revisions, which has increased the need for categorical slots. The first four characters of the code are internationally official. The fifth- and sixth-character fields are available for regional and special purpose adaptations. This arrangement maintains international communication while accommodating local diversity.

Another powerful and innovative concept is ICD-10 as a family of disease and health-related classifications. At the core of the family are the 21 main chapters coded at the official three-character and four-character levels and the short tabulation lists of causes of death and morbidity. Peripherically located are the following classifications: (1) specialty-based adaptations (e.g., for oncology) where the chief difference from the core classification lies in the further extension of the ICD codes; (2) classifications for primary care and general medical practice, characterized by the condensation of categories and emphasis on less rigorous diagnostic terminology and more immediate therapeutic utility; and (3) classifications of information outside the core ICD, such as that corresponding to disabilities and medical procedures. Also part of the family is the International Nomenclature of Diseases, which encompasses a list of recommended names for all diseases as well as their definitions. In contrast to the concept of nomenclature, a classification, in the words of the ICD pioneer Farr, “groups diseases that have considerable affinity or that are lable to be confounded with each other, and therefore is likely to facilitate the deduction of general principles.”

Main forms of human illness and related conditions constitute the 21 chapters as the core of ICD-10 (Table 9.2-1). New chapters structure the enlarged lists of disorders of the nervous system (Chapter VI), eye and adnexa (Chapter VII), and ear and mastoid process (Chapter VIII). The expanded chapter on neoplasms covers one full letter and shares another with blood disorders, which encompasses immunological conditions such as acquired immune deficiency syndrome (AIDS). Also the classification of neoplasms is multiaxial (one axis denotes topography and another morphology, i.e., histological type, and tumor invasiveness and differentiation).

Chapter	Title
I	Certain infectious and parasitic diseases
II	Neoplasms
III	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV	Endocrine, nutritional, and metabolic diseases
V	Mental and behavioral disorders
VI	Diseases of the nervous system
VII	Diseases of the eye and adnexa
VIII	Diseases of the ear and mastoid process
IX	Diseases of the circulatory system
X	Diseases of the respiratory system
XI	Diseases of the digestive system
XII	Diseases of the skin and subcutaneous tissue
XIII	Diseases of the musculoskeletal system and connective tissue
XIV	Diseases of the genitourinary system
XV	Pregnancy, childbirth, and the puerperium
XVI	Certain conditions originating in the perinatal period
XVII	Congenital malformations, deformations, and chromosomal abnormalities
XVIII	Symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified
XIX	Injury, poisoning, and certain other consequences of external causes
XX	External causes of morbidity and mortality
XI	Factors influencing health status and contact with health services

Reprinted with permission from World Health Organization. International Statistical Classification of Diseases and Related Health Problems, ed 10, vol 1. World Health Organization, Geneva, 1992.

Table 9.2-1 List of Core Chapters of ICD-10

ICD-10 Classification of Mental Disorders Mental and behavioral disorders are housed within Chapter V of ICD-10 and are coded with the letter F (Table 9.2-2). The use of the sixth letter of the Gregorian alphabet to denote chapter V is explained by the assignment of two letters to a very lengthy list of conditions in chapters on infectious and parasitic diseases. After the letter F, the first digit of the Chapter V diagnostic codes denotes 10 major classes of mental and behavioral disorders: F0 through F9. The second and third digits (third and fourth characters) identify progressively finer categories. For example, the code F30.2 sequentially denotes the mental chapter, mood disorders class, manic episode, and the presence of psychotic symptoms. In this manner, 1000 four-character mental disorder categorical slots are available in ICD-10.

Table 9.2-2 ICD-10 Classification of Mental Disorders

F0-Organic, Including Symptomatic, Mental Disorders This class is etiologically based on physical disorders or conditions involving or leading to brain damage or dysfunction. The first clusters have disturbances of cognitive functions as prominent features and include the dementias (Alzheimer's, vascular, associated with other diseases, and unspecified), organic amnestic syndrome, and delirium not induced by psychoactive substances. The second cluster has as its most conspicuous manifestations alterations in perception (hallucinations), thought (delusions), mood (depressed or manic), various emotional domains (such as anxiety and dissociation), and personality.

F1-Mental and Behavioral Disorders Due to Psychoactive Substance Use In contrast to earlier classifications, this class subsumes all mental disorders related to

psychoactive substance use, from patterns of dependence and harmful use to various organic brain syndromes induced by substances. The diagnostic process and coding starts with identification of the substance involved (i.e., alcohol, opioids, cannabinoids, sedatives, or hypnotics, cocaine, other stimulants, hallucinogens, tobacco, volatile solvents, and other substances and combinations of them). Identified next in the code is the involved clinical condition: acute intoxication, harmful use (previously known as abuse and characterized by a pattern of use causing damage to physical or mental health), dependence syndrome, withdrawal state (with or without delirium), psychotic disorder, amnesic syndrome, residual and late-onset psychotic disorder, and other and unspecified mental disorders.

F2-Schizophrenia, Schizotypal, and Delusional Disorders This class has schizophrenia as its centerpiece, a disorder characterized by fundamental and distinctive distortions of thinking and perception and by inappropriate or blunted affect. The remaining categories of nonorganic, nonaffective psychoses are considered somewhat related, phenomenologically or genetically, to schizophrenia. Particularly interesting is the cluster of acute and transient psychotic disorders, which encompasses a heterogeneous set of acute-onset and relatively short-lived psychoses (polymorphic with or without schizophrenic symptoms, acute schizophrenia-like, and others) reportedly frequent in industrially developing countries (where most of the world population lives).

F3-Mood (Affective) Disorders The fundamental disturbance in this class is a change in mood or affect, usually involving depression or elation, often accompanied by a change in level of activity. Included here are manic episode, bipolar affective disorder (characterized by recurrent episodes involving both depression and elation), depressive episode, recurrent depressive disorder, persistent mood disorder (cyclothymia, dysthymia), and other and unspecified mood disorders.

F4-Neurotic, Stress-Related, and Somatoform Disorders This grouping is based on a historical concept of neurosis that presumes a substantial role played by psychological causation and that mixtures of symptoms are common, particularly in less severe forms often seen in primary care. Included in this book are phobic anxiety and other anxiety disorders, obsessive-compulsive disorder, reactions to severe stress and adjustment disorders, dissociative and conversion disorders, somatoform disorders, and other neurotic disorders (e.g., neurasthenia and depersonalization-derealization syndrome).

F5-Behavioral Syndromes Associated With Physiological Disturbances and Physical Factors Included here are eating disorders, nonorganic sleep disorders, and sexual dysfunction, mental disorders associated with the puerperium and not elsewhere classified, psychological factors influencing physical disorders, and abuse of non-dependence-producing substances (e.g., antidepressants, hormones, analgesics, and many folk remedies).

F6-Disorders of Adult Personality and Behavior This class includes clinical conditions and behavioral patterns that tend to persist and the expression of an individual's characteristic lifestyle and mode of relating to self and others. The main subclass involves personality disorders, which are deeply ingrained and enduring behavior patterns, manifesting as inflexible responses to a broad range of personal and social situations. An innovative category is that of enduring personality change, neither developmental nor attributable to brain damage or disease, and usually emerging after catastrophic experiences or another psychiatric illness. The broad class also includes impulse, gender identity, sexual preference, and sexual development and orientation disorders.

F7-Mental Retardation Mental retardation, one of the oldest in the history of psychiatric classifications, involves arrested or incomplete mental development, characterized by impaired cognitive, language, motor, and social skills evidenced during the person's formative period and contributing to the overall level of intelligence. Its subcategories correspond to various levels of severity: mild, moderate, severe, and profound mental retardation. Extent of behavioral impairment is also coded.

F8-Disorders of Psychological Development Disorders of psychological development are characterized, as a class, by the following attributes: onset during infancy or childhood, impairment or delay of functions connected to the maturation of the central nervous system, and a steady course unlike the remissions and relapses usual in many mental disorders. The functions affected most frequently include language, visuospatial skills, and motor coordination. A major subclass encompasses a variety of specific developmental disorders, classified by the abilities involved: speech and language, scholastic skills, and motor function. The other major subclass corresponds to pervasive developmental disorders, many of which are more saliently characterized by deviance rather than delay in development but always involving some degree of delay. Most conspicuous here are childhood and atypical autistic disorder and Rett's syndrome and other childhood disintegrative disorders.

F9-Behavioral and Emotional Disorders With Onset Usually Occurring in Childhood and Adolescence This complex class complements F7 and F8. Child-onset disorders included first are hyperkinetic disorders characterized by early onset, overactive and poorly modulated behavior associated with marked inattention, lack of persistent task involvement, and pervasiveness over situations and time. Conduct disorders are defined by a repetitive and persistent pattern of dissocial, aggressive, or defiant behavior. Also included in this class are emotional, social-functioning, tic, and other disorders usually starting in childhood or adolescence.

The full ICD-10 classification of mental disorders has three presentations corresponding to various degrees of definitional detail, aimed at serving different purposes and uses:

1. An abbreviated glossary containing the principal features of each disorder, for the use of statistical coders and medical librarians, published within the ICD-10 general volume
2. Clinical descriptions and diagnostic guidelines, containing widely accepted characterizations of an intermediate level of specificity, intended for regular patient care and broad clinical studies
3. Diagnostic criteria for research, characterized by more-precise and rigorous definitions

Multiaxial Presentation of ICD-10 Over the past two decades, the WHO advanced some important initiatives on multiaxial diagnosis. One was the Multiaxial Classification of Child Psychiatric Disorders, first designed in 1969 and revised and expanded several times since then. Its 1975 pentaxial version encompassed the following axes: psychiatric disorders, physical disorders, developmental disorders, intellectual level, and abnormal psychosocial situations. Another was the Triaxial Classification of Health Problems for Primary Care, which contained axes on physical, psychopathological, and social problems.

On these precedents and many others from the international literature, in the late 1980s the WHO Mental Health Division began preparing the Multiaxial Presentation of ICD-10. The conceptual bases of this development included a critical analysis of over 20 published multiaxial proposals originating in countries spanning three continents, which revealed important commonalties in the clinical domains covered. A second developmental principle was simplicity in the multiaxial schema, to enhance the prospects of its effective use across the world. The third principle was to base the instruments for axial assessment on components of the ICD-10 family of classifications, which had benefited from wide international consultations and field trials.

The Multiaxial Presentation of ICD-10 is composed of three axes: I. Clinical Diagnoses; II. Disablement; and III. Contextual Factors. The number of axes is lower than the four or five usually included in published multiaxial schemata and constitutes a condensation of those axes most frequently included in multiaxial proposals published in the international literature, which affords a measure of content validity to the schema. The value of the simplicity of the schema is enhanced by its potential for generalization beyond psychiatric practice. Information on clinical pathology, disablement, and contextual factors appears to be relevant to all health care.

Axis I—Clinical Diagnoses This axis accommodates both mental and nonmental (general medical) disorders, underlining a fundamental commonality among all illnesses. All significant disorders identifiable in a given individual are to be listed and coded according to Chapter I,Chapter II,Chapter III,Chapter IV,Chapter V,Chapter VI,Chapter VII,Chapter VIII,Chapter IX and Chapter X (the disease chapters) in the core classification of ICD-10 (Table 9.2-2).

Axis II—Disablements This axis appraises the consequences of illness in terms of impairment in the performance of basic social roles. The assessment instrument is a shortened version of the WHO Disability Assessment Scale, whose structure was condensed into four dimensions or areas: (1) personal care, (2) occupational functioning (as remunerated worker, student, or homemaker), (3) functioning with family (assessing both the regularity and quality of interactions with relatives and household members), and (4) broad social behaviors (interaction with other individuals and the community at large and leisure activities).

Axis III—Contextual Factors This axis attempts to portray the context of illness in terms of several ecological domains. These include problems related to the family or primary support group, general social environment, education, employment, housing and economic circumstances, legal issues, family history of illness, and personal life management and lifestyle. Assessment involves identifying problematic broad categories and recording specific factors. This structure is based on ICD-10 Chapter XXI (factors influencing health status and contact with health services).

Primary Health Version The ICD-10 Primary Health Version is a simple, brief classification arrangement compatible with and translatable into the ICD-10 standard classification of mental disorders (Table 9.2-3). It is linked with management aids prepared for use by primary care practitioners. The short list of categories was selected principally on the basis of importance to public health and the availability of effective and acceptable management. The centerpiece of the package is a set

of pocket-sized flip-cards, one for each selected category. One side of the flipcard exhibits assessment information, such as presenting complaints, diagnostic features, and differential diagnosis. The other side displays management guidelines, such as essential information for patient and family, specific counseling for the patient and family, medication, and specialist consultation. Additional elements of the package include flow charts, symptom indexes, and a computerized version.

F00*	Dementia
F05	Delirium
F10	Alcohol use disorders
F11*	Drug use disorders
F17.1	Tobacco use
F20*	Chronic psychotic disorders
F23*	Acute psychotic disorders
F31	Bipolar disorder
F32*	Depression
F40*	Phobic disorders
F41.0	Panic disorder
F41.1	Generalized anxiety
F41.2	Mixed anxiety and depression
F43*	Adjustment disorder
F44*	Dissociative disorder (conversion hysteria)
F45	Unexplained somatic complaints
F48.0	Neurasthenia
F50*	Eating disorders
F51*	Sleep problems
F52	Sexual disorders
F70	Mental retardation
F90	Hyperkinetic (attention deficit) disorder
F91	Conduct disorder
F98.0	Enuresis

* Means that some ICD-10 codes is included.

Table 9.2-3 ICD-10 Primary Health Care Categories

ADAPTATIONS OF ICD-10

The ICD-10 Classification of Mental and Behavioral Disorders is being accepted by most countries and by the WPA as the international standard in the field for both statistical reporting and for clinical care and research. Emerging now is the need to harmonize international communication with recognition of cultural diversity and specific local requirements. To express this perception, several national adaptations, versions, and annotations of ICD-10 are being developed and published.

The rationale and arguments for the development of local glossaries include the following:

1. Local glossaries from across the world can serve as the fundamental bases for the preparation of bottom-up international diagnostic systems.
2. Local glossaries, through its attention to various aspects of clinical reality, can facilitate the implementation of a comprehensive biopsychosocial framework, away from reductionisms of different types.
3. Local glossaries can reflect the cultural integrity of different countries and human groups and promote the value of their health-related concepts and practices.
4. Local glossaries can embody and transmit the intellectual contributions of national and regional scientists and professional leaders for the benefit of the field around the world.
5. Local glossaries can facilitate the effective use of international diagnostic systems, by adapting the various components of these systems to national patterns and needs.

The best known of these national versions is the APA's DSM-IV. In 1980 the APA published the innovative DSM-III characterized by a phenomenological emphasis on the conceptualization and organization of mental disorders, the use of explicit diagnostic criteria, and a multiaxial formulation. It acquired wide international visibility and significantly influenced the field of psychiatric classification. After preparations for ICD-10 were started by WHO, the APA initiated the development of DSM-IV, attempting this time to keep close to the international standard (which was already incorporating much of the methodological features advanced by DSM-III). Perhaps the principal attribute of DSM-IV is the scholarly emphasis in its development on the basis of critical literature reviews, reanalysis of existing databases, and focused field trials.

Cuban Glossary of Psychiatry Development of the various editions of the *Cuban Glossary of Psychiatry* started in 1975. They have attempted to reflect the realities and needs of Cuba in particular, within the general framework of Latin American culture. Emphasis on this perspective was presented first as a responsibility to the local population and second as an effort to contribute to the bottom-up building of international classification. The latter implies that carefully prepared technical contributions are as worthy of consideration for constructing an international reference as those coming from Europe, North America, or any other part of the world.

The Cuban glossaries have furthermore attempted to harmonize the general, represented by the existing ICD, with the local. Therefore, its authors consider the glossary to be the basic ICD-10 with only the changes and additions needed to ensure its applicability and usefulness for psychiatric care in Cuba.

As a broader conceptual framework for the preparation of the Cuban glossaries, their authors considered a syncretism of local historical traditions, the existing social and clinical reality of the region, and international scientific contributions. Among the many specifically Latin American contributions they were attentive to the following. Honorio Delgado, a Peruvian psychiatrist, published the influential *Curso de Psiquiatría* in the middle of this century, with a masterful presentation of psychopathological phenomenology and nosology. In 1954 in Rio de Janeiro Leme Lopes published the innovative *Dimensoes do Diagnóstico Psiquiátrico*, with a pioneering multiaxial formulation of clinical conditions. The Cuban psychiatrist José Bustamante published in 1975 with anthropologist A. Santa Cruz, what can be arguably considered the first textbook of transcultural psychiatry in the world. Carlos Alberto Seguín from Peru, with his volume on folkloric psychiatry, is emblematic of the rich Latin American contributions to the description of popular or culture-bound syndromes in the region.

Early Editions of *Cuban Glossary of Psychiatry* The Cuban glossaries of psychiatry have been the most complete and distinctive Latin American effort to annotate and adapt the existing ICD to local reality. The general principles guiding the preparation of the Cuban Glossaries of Psychiatry include the following:

1. Compatibility with ICD by incorporating modifications and additions through the fifth digits of the diagnostic code and the use of qualifying phrases
2. Consideration of all suggestions offered at provincial and national seminaries on the classification of mental disorders
3. Deliberate effort to harmonize different viewpoints
4. Systematic participation of most psychiatrists in the country and many clinical psychologists and other professionals concerned with mental health (e.g., teachers, social workers, and forensic doctors).

Its first edition (GC-1), published in 1975, constitutes an adaptation of the eighth revision of the *International Statistical Classification of Diseases, Injuries, and Causes of Death* (ICD-8). Its architect was Carlos Acosta-Nodal, professor of psychiatry at Havana University, working under the auspices of Eduardo Ordaz, director of Havana Psychiatric Hospital. The GC-1 was composed of 69 adaptations of ICD-8. The second edition of GC (GC-2), published in 1983 (Fig. 9.2-1), was an adaptation of ICD-9. Its development was again chaired by Acosta-Nodal. It contained 90 modifications of ICD-9 in addition to substantive chapters devoted to historical, theoretical, and clinical aspects of psychiatry.

FIGURE 9.2-1 Cover of the *Third Cuban Glossary of Psychiatry, 1998* (Tercer Glosario Cubano de Psiquiatria [GC-3]).

Third Edition of *Cuban Glossary of Psychiatry* The current third edition *Cuban Glossary of Psychiatry* (GC-3) was developed under the chairmanship of Angel Otero-Ojeda, from Havana University and the Havana Psychiatric Hospital. It includes a number of contributions on the diagnosis of mental disorders as experienced and presented in Cuba. These encompass both general psychiatry and child psychiatry. A basic principle in the development of GC-3 was to be similar to ICD-10, with minimal differences. In line with this, the coding system of ICD-10 was faithfully followed. Contributions and changes were incorporated through the employment of fifth digits in the diagnostic code or through the utilization of codes not used in ICD-10. The ICD-10 diagnostic guidelines for the various psychiatric disorders were also respected to the largest extent possible. In some cases, supplemental text was added. References to DSM-IV and to the ICD-10 criteria for research were often made. Of note, the GC-3 encourages diagnostic formulations based on the judicious use of all information available and allows experienced clinicians to formulate diagnoses without strictly adhering to standard diagnostic criteria.

GC-3 starts with an introductory chapter that outlines the principles of the diagnostic process and the positive aspects and problems of the current classifications. One such problem is the hospital-focus of most systems, which may limit their relevance to the emerging community-based psychiatry.

Another interesting chapter of GC-3 is that on “Syndromes of Difficult Placement” (sometimes referred to as culture-based syndromes). This includes both widely known folk syndromes such as amok, brain fag, and *rusto* as well as some syndromes and idioms of distress reported by Cuban psychiatrists. Illustrative of the latter is *abríu*, which refers to certain children believed to have the power to exercise a malign supernatural influence on their relatives, particularly siblings, who as a consequence can experience various illnesses and even die.

Additions made in the main body of GC-3 include (1) “neurotic behavior” as qualifier in broad categories such as mood disorders and neurotic, stress-related, and somatoform disorders, (2) diagnosis based on premorbid features for disorders of chronic or episodal source in broad categories such as schizophrenia and related disorders and mood disorders, (3) risky behaviors such as suicide or substance use, which are of particular relevance to community psychiatry. An illustrative deletion was dementia in children.

Some of the most innovative contributions of GC-3 involve its multiaxial schemes, which build on the standard multiaxial presentation of ICD-10. The GC-3 multiaxial scheme follows:

Axis I—Clinical Diagnosis. Both mental and nonmental disorders are included as in ICD-10's multiaxial system.

Axis II—Disabilities. Disablements in personal care, occupational functioning, functioning with family, and broader social functioning are included following the guidelines of ICD-10's multiaxial presentation.

Axis III—Psychoenvironmental Factors (Adverse). Included here are contextual problems listed under the third axis of WHO's multiaxial presentation of ICD-10.

Axis IV—Psychoenvironmental Factors (Other). Illustrative of factors considered here are to live alone and to be particularly “practical” or “romantic.”

Axis V—Maladaptive Behavior and Psychological Needs. Included here are conditions such as hypertrophic affective needs, indecisiveness, and difficulties managing hostility.

Axis VI—Other Significant Factors. Included here is miscellaneous information such as that resulting from laboratory tests and responses to therapeutic interventions.

Work on the GC-3 stimulated the Section on Diagnosis and Classification of the Latin American Psychiatric Association, under the leadership of Carlos Berganza (San Carlos University, Guatemala), Miguel Jorge (Escola Paulista de Medicina, Brazil), Otero-Ojeda (Havana University, Cuba) and Juan E. Mezzich (Mount Sinai School of Medicine in New York and Cayetano Heredia Peruvian University in Lima) to organize preparation of the first *Latin American Glossary of Psychiatry* as a Latin American annotation of ICD-10. Its components include the conceptual bases of psychiatric diagnosis, the hierarchical organization and diagnostic criteria of psychiatric nosology, culture-bound syndromes and idioms of distress, and multiaxial and comprehensive diagnostic formulations.

Chinese Classification of Mental Disorders In China attempts to classify mental disorders began around 1958. The first classificatory scheme was published in 1979 and named the *Chinese Classification of Mental Disorders*, first edition (CCMD-1) in 1981. Under the influence of the DSM and ICD systems, a number of subsequent revisions culminated in the publication of the second edition (CCMD-2) in 1989 and the revised second edition (CCMD-2-R) (Fig. 9.2-2) in 1995. For the first time in China operationalized criteria for a broad range of diagnostic categories have become available. Since China has over one fifth of the world's population, knowledge of the CCMD-2-R is central to the knowledge base of psychopathology in general. Chinese people also constitute one of the fastest growing ethnic minority groups in Western societies, so understanding the CCMD-2-R may attune clinicians to certain Chinese forms of distress in an intercultural treatment context.

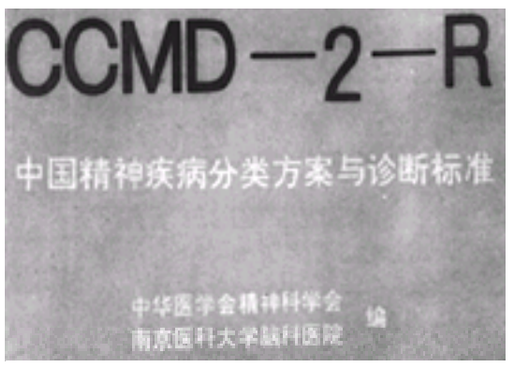

FIGURE 9.2-2 Cover of the *Chinese Classification of Mental Disorders*, ed 2, Revised, (CCM-D-R), 1995.

In devising CCMD-2-R, Chinese psychiatrists sought to harmonize with the international classification on one hand and to sustain a nosology that is useful in a huge and heterogeneous country on the other. As a result, CCMD-2-R and the ICD-10 share a broadly comparable architecture. However, inasmuch as symptom recognition and taxonomic strategy in psychiatry reflect the cultural norms and values of the society in which they are embedded, this blending is legitimately incomplete. The CCMD-2-R is a concise handbook of 238 pages that contains the equivalent or closest ICD-9 and ICD-10 codes (crosswalks) alongside the diagnostic headings. It contains a number of cross-culturally salient features, including particular additions (*gigong*-induced mental disorders, traveling psychosis), deletions (somatoform disorders, pathological gambling, a number of personality and sexual disorders), retentions (unipolar mania, neurosis), and variations (neurasthenia, depressive neurosis, anorexia nervosa) of diagnostic categories. A number of field trials have indicated adequate reliability and validity.

Addition of Culture-Specific Diagnostic Categories Both ICD-10 and DSM-IV lack a coded classification of culture-bound syndromes, which are often cited in the argument against an international classification of mental disorders. By contrast, the CCMD-2-R includes *qigong*-induced mental disorder as a culture-related mental disorder. *Qigong* is a popular trance-based healing system that consists of meditational or diverse styles of movement exercise or both. It is induced by using a culture-syntonic set of suggestions based on the Chinese concept of *qi* (vital energy). *Qigong*-induced mental disorder is believed to arise from inappropriate or excessive application of *qigong* or inability to terminate the *qigong*. This causes the flow of *qi* to deviate from the *jing luo* conduits and become *fire*, and may result in a variety of symptoms that do not fit into one coherent syndrome recognized in ICD-10 or DSM-IV. These include *qi*-related somatic discomforts, uncontrolled motor activity, anxiety, fright, weepiness, irritability, delusions, identity disturbance, hallucinations, mania, depression, and suicidal, bizarre, and violent behavior. The condition is usually brief. Treatment consists of a short course of tranquilizers and education about the proper way of practicing *qigong*. Relapse is uncommon.

In the Western psychiatric literature, psychosis related to traveling is uncommonly reported and is usually confined to air travel among subjects with a preexisting history of mental disorder. By contrast, the category of traveling psychosis arose from reports of acute psychosis developing among thousands of rural migrants who traveled in severely crammed trains over long distances in search of jobs in the richer regions of China. Its principal manifestations include an acute onset, perplexity,

disturbed consciousness, anxiety, persecutory delusions, horrifying illusions or hallucinations, motor excitement, impulsive and suicidal jumping off the train, and injuring others. In most cases, the termination of travel, rest, and renourishment lead to spontaneous recovery within a few hours to a few days. A personal or family history of mental disorder is rare. Even if traveling psychosis has an organic cause (e.g., hypercapnia, sleep deprivation, fatigue, and dehydration), its origin must be seen in the context of China's rapid market reforms, which result in marked economic regionalism and massive domestic migration. Granting traveling psychosis a special nosological status in CCMD-2-R is expected to promote research into its prevention (e.g., improved conditions for traveling and regulation of migration). It also serves judicial functions, since offenders with this diagnosis may be granted a verdict of diminished responsibility or acquittal.

Deletion of Culturally Inappropriate Categories Despite the common belief that Chinese people are prone to somatization, almost the whole block of somatoform disorder in ICD-10 and DSM-IV is excluded in the CCMD-2-R. According to ICD-10, the main features of such conditions are the repeated presentation of physical symptoms and persistent requests for medical investigations despite negative findings and reassurances by doctors that the symptoms have no physical basis. The patient usually resists attempts to discuss the possibility of psychological causation even in the presence of precipitating psychosocial stressors. This definition embodies a radical mind-body epistemology that is exotic to traditional Chinese or Ayurvedic medicine. Several factors have made it difficult for Chinese psychiatrists to apply the category of somatoform disorders. Chinese patients, if given the opportunity, readily communicate dysphoria and relate somatic symptoms to psychosocial stressors. Rather than being mutually exclusive, their somatic and emotional symptoms are highly intercorrelated. Clinically, their somatization may be considered a context-dependent strategy of engaging the concern of physicians who often work at overcrowded clinics. Moreover, the hybrid (half Greek and half Latin) word “somatoform” is a terminological puzzlement to Chinese psychiatrists, who find that neuroses (including neurasthenia) are clinically more useful categories for engaging patients in treatment.

N. S. was a 35-year-old lower social class housewife who presented to a psychiatric clinic with a mixture of symptoms associated with significant impairment. These included fullness in the head, weakness, worries, insomnia, cold intolerance, and difficulty in breathing for 2 years. She had previously worked as a clerk but quit to become a full-time housewife after marriage. Exploration revealed that she was burdened with caring for a teenage daughter and a 10-year-old son with childhood autistic disorder. She was also worried about the fidelity of her emotionally disengaged husband when he started to work in a nearby city for 1 to 2 days a week. Because of the absence of a persistently depressed or anxious mood, she did not meet the definitional thresholds for DSM-IV mood or anxiety disorders. Instead, she could be given the diagnosis of undifferentiated somatoform disorder according to DSM-IV or neurasthenia according to CCMD-2-R. During clinical interview, N. S. did not think she had major depression. She was particularly puzzled by the diagnosis of undifferentiated somatoform disorder (*weifenhua quti zhangai*), which was both unintelligible and experience-distancing to her. In her view, she merely had neurasthenia as a result of multiple stressors in the family. Although the Chinese psychiatrist wrote down the DSM-IV diagnosis in her case notes, he used the term *neurasthenia* to enhance clinical communication. N. S. declined pharmacotherapy but accepted counseling readily.

From a cultural constructionist perspective, personality disorders are based on Anglo-American conceptions of personhood and codes of appropriate behavior and owe their existence to the medicalization of disvalued social behavior. Accordingly, transformations in the values of a society determine whether they are called disease, sin, or crime. Given the disparity between the Oriental and Occidental conceptions of personhood, queries over the contextual validity of personality disorders are to be expected. Personality disorders are neither common clinical diagnoses nor popular research topics in China. Published studies indicate that a high percentage of Chinese subjects fail to fall into the subtypes recognized in DSM-IV and ICD-10. Thus, anxious (avoidant) and dependent personality disorders are excluded in CCMD-2-R. This is because many of their defining features (e.g., excessive preoccupation with being criticized or rejected in social situations, and subordination of one's own needs to those of others on whom one is dependent) are normative in the Chinese culture, which defines personhood not by autonomy and assertiveness, but by intergenerational dependence and selfeffacement. Likewise, since people who gamble immoderately and ruin their lives are considered bad rather than mad in Chinese society, the category of pathological gambling is excluded in the CCMD-2-R.

Retention of Diagnostic Categories Unipolar mania is no longer found in ICD-10 or DSM-IV, according to which a patient with two episodes of mania is assumed to have bipolar disorder. However, longitudinal studies indicate that recurrent mania continues to be seen in China and question the obligatory labeling of such Chinese patients as having bipolar disorder. The possibility that they may exhibit particular biological correlates, treatment response, and outcome supports a separate nosological status in CCMD-2-R.

The words “neurosis” and “neurotic” have completely disappeared from the DSM-IV, while the ICD-10 no longer retains neurosis as a major organizing principle. Instead, neurosis is finely partitioned into a variety of anxiety and depressive disorders, resulting in a very high prevalence of comorbidity being found in recent Western epidemiological studies. Among Chinese psychiatrists, the *neurosis (shenjing zheng)* has been used as a popular descriptive and etiological concept since the 1950s. It is therefore preserved in CCMD-2-R, which emphasizes as its main characteristics chronicity (at least 3 months), the presence of predisposing personality and social factors, and the preservation of insight.

Variation of Diagnostic Categories By tradition, Chinese psychiatrists have confined the diagnosis of mood disorder to bipolar disorder and psychotic depression. Chronic mild depression such as dysthymia is considered a form of neurosis and has been subsumed under neurasthenia and, more recently, depressive neurosis. Clinically, depressive neurosis encompasses a highly heterogeneous group of conditions that overlap with anxiety, mood, and personality disorders. Affected Chinese patients usually present with insomnia, headache, worries, and other indigenous forms of dysphoria rather than depression *(yiyu)*, which is an uncommon term in the everyday life of Chinese people. Among Chinese psychiatrists, whether depressive neurosis should be classified as a mood or neurotic disorder remains a subject of debate that is unlikely to be solved by empirical research alone.

Since CCMD-2-R has tightened the concept of neurasthenia according to Western epistemological assumptions (i.e., it should not be diagnosed if mood or other neurotic disorders are present), the use of the diagnosis is becoming less common among Chinese psychiatrists. But neurasthenia remains a common clinical diagnosis among doctors in China as a whole. According to CCMD-2-R, it is a type of neurosis composed of any three of five nonhierarchical groups of weakness, dysphoria, excitement, nervous pain, and sleep symptoms. This flexible symptom configuration is based on ample clinical experience in China and has been supported by culturally sensitive studies of the variegated illness experience of Chinese neurasthenic patients. By contrast, DSM-IV and ICD-10 representations of neurasthenia (as an undifferentiated somatoform disorder and a chronic fatigue disorder, respectively) lack contextual validity in Chinese society.

Because anorexia nervosa is still a rare condition in China, CCMD-2-R criteria for anorexia nervosa closely follow ICD-10 and DSM-IV. However, recent studies indicate that the core diagnostic construct of intense fear of fatness even when underweight is overrestrictive in a cross-cultural context. Such a fear is often absent among Chinese anorectic patients, who use other rationales (e.g., abdominal bloating, loss of appetite, or no hunger) for voluntary noneating. More culture-flexible diagnostic criteria that take account of the local meanings of food refusal will be adopted in CCMD-3.

N. F. aged 28, height 1.51 m (4 ft, 11 in), ideal body weight 47 kg (103 lb), was a single clerk who was referred to the psychiatric clinic from a gynecological unit for 6 years' history of unexplained secondary amenorrhea and weight loss from 43 kg (95 lb; body mass index, 18.9 kg/m_2) to 32 kg (70 lb; body mass index, 14.0 kg/m_2). She came from a working-class family in which the father had sexually molested her when she was 12. She experienced chronically low self-esteem and a strong sense of loss of control over her life. Since graduation from high school, she had worked continuously in a Protestant church in the hope that it would provide security for her. N. F. had always been slim and had never thought of dieting to pursue beauty. She attributed her poor food intake to loss of appetite during family dinner, when her father scolded her ruthlessly. She no longer cared how much and what she ate but sought refuge by fleeing the dining table as quickly as possible. She regarded this as a form of silent protest against him. Physical examination revealed a pubescent look and the usual signs of emaciation. Extensive investigation revealed no physical cause for the weight loss. When interviewed with the Eating Disorders module of the Structured Clinical Interview for DSM-III-R (SCID), she failed to fulfill the criteria for anorexia nervosa, while her depressive symptoms fell short of those for a major depressive episode. N. F. received individual psychotherapy for 2 years, during which family conflicts, powerlessness, and the lack of meaning in life constituted the organizing theme of her illness experience. She never mentioned the fear of fatness and sometimes challenged why the therapist, instead of understanding her lack of meaning of life, was so preoccupied with her eating behavior and body weight. She fulfilled culture-flexible diagnostic criteria for anorexia nervosa.

Regional Adaptations of ICD-10 in Other Asian Countries The ICD-10 has been tested and is increasingly used in other Asian countries such as India, Japan, and Korea. Available evidence indicates that it is a reasonably feasible scheme that promotes reliability and international comparisons across most diagnostic categories. Nonetheless, local modifications are required to enhance its contextual validity and make it user friendly. For example, the Japanese Society for International Diagnostic Criteria in Psychiatry (JSIDCP) has also decided to retain a unitary concept of neurosis, which should be connected with predisposing personality traits and life events. As in CCMD-2-R, mild depression is considered a form of neurosis rather than a mood disorder, and the term “somatoform disorder” is avoided. Further examination of the discrepancies between Asian and Western nosological systems will encourage reflective self-criticism in one hand and contribute to an internationally valid system of psychiatric classification on the other.

TOWARD COMPREHENSIVE DIAGNOSIS MODELS

While the national adaptations of the ICD represent attempts at resolving the tension between universalistic and local perspectives and needs, another emerging cluster of efforts in the field is aimed at integrating different informational domains and the perspectives of different evaluators in the construction of comprehensive diagnostic formulations.

Conceptual Framework The conceptual framework for these developments includes the articulation of previously divergent historical and philosophical traditions, new notions on health and health care being advanced by the WHO, and innovative developments in general and psychiatric epidemiology. From a historical and philosophical viewpoint, the longitudinal evolution of diagnostic systems can be seen as unfolding on three parallel lines. One is represented by a synthetic, bold, and abstract Platonic conceptualization of a disease entity as a sufficient descriptor of a patient's clinical condition. The second involves an analytical, textured, and experiential Aristotelian viewpoint. The third is an empathetic Hippocratic approach. The Platonic tradition has informed the long-standing international effort to classify illnesses as reflected in the various versions of the ICD and national versions such as DSM-IV. The Aristotelian perspective is reflected in the descriptions of the patient's contextualized clinical condition, using standardized typologies and scales, which has appeared under the term “multiaxial diagnostic evaluations” The Hippocratic tradition is related to recent efforts to focus on the individuality of the patient.

The WHO, through recent meetings of its executive board, is expanding its 50-year definition of health, proposing that it is not merely an absence of illness but a dynamic (or interactive) state of complete physical, emotional, social, and spiritual wellbeing. Additionally, WHO's explorations on the description of health status is expanding its focus from disease to functioning and other positive aspects of health. One of these is quality of life, which is to be evaluated predominantly by the individual involved. Furthermore, in the formulation of its policy in *Health for All for the 21st Century*, WHO is incorporating ethics, equity, and human rights as new, important considerations.

Epidemiology, as the basic science of public health, is undergoing a substantial conceptual revision. Newer formulations go beyond the infectious disease model and the chronic disease model, calling for approaches that incorporate multiple and interactive levels of analysis as suggested by Mervin Susser, editor of the *American Journal of Public Health*, under the term “ecological epidemiology.”

Integrating Standardized and Idiographic Formulations Building on the traditions and developments outlined above, the WPA is designing a more comprehensive diagnostic model as part of a project on international guidelines for diagnostic assessment. The model has two key components:

1. A standardized multiaxial formulation that covers nosology, disabilities, and contextual factors. The multiaxial formulation is aimed at statistically reliable measurement of key aspects of the clinical condition to facilitate sharing of diagnostic and treatment information among clinicians and across the world.
2. An idiographic or personalized formulation that focuses on the individuality of a particular patient. The idiographic formulation is aimed at providing complementary descriptive information, engaging the patient more fully in the process of clinical care and fulfilling ethical aspirations of respect to the dignity of the patient and attending to his or her expectations in dealing with health problems and enhancing quality of life.

Table 9.2-4 presents schematically the components of this diagnostic model. The standardized multiaxial formulation of this model is basically organized according to the multiaxial presentation of ICD-10.

A. Standardized multiaxial formulation
- I. Clinical diagnoses (mental and nonmental disorders)
- II. Disablements
  Difficulties in fulfilling social roles in regard to personal care, occupational functioning, functioning with family, and broader social functioning
- III. Contextual factors

B. Idiographic formulation
- I. Clinician perspectives
- II. Patient/family perspectives
- III. Integration of clinician and patient/family perspectives

Table 9.2-4 Comprehensive Diagnostic Model Incorporated Into the WPA International Guidelines for Diagnostic Assessment

A brief description of the components of the idiographic formulation follows:

I. Clinician's perspectives. This involves an effort to synthesize the information available including a summary of the patient's problems and assets, biological explanations, psychological explanations, social explanations, and biopsychosocial integration, that is, a contextualized articulation of all explanatory models pertinent to the patient's clinical condition.
II. Patient/family's perspectives. These may include a biographical history; an understanding of the nature, causes, and context of the illness experienced; self-appraisal of quality of life (e.g., physical and emotional health, functioning, social support and resources, personal and spiritual fulfillment); and attitudes and expectations toward clinical care (acute and rehabilitative treatment and preventive efforts).
III. Integration of clinician and patient or family's perspectives. This integration is expected to be based on empathetic rapport, reflecting mutual respect, interest, and human feelings between clinician and patient. The clinician and patient attempt to reach the best possible joint understanding of the patient's clinical condition in a context that includes family factors. They negotiate a care plan, considering personal issues (e.g., patient's and clinician's preferences and disposition toward treatment, sense of autonomy, time available, travel requirements), social factors (e.g., social supports, particularly availability and interest of the family), and financial circumstances (e.g., insurance and managed care constraints). Finally, they jointly monitor the progress of care and its outcome.

FUTURE DIRECTIONS

The developments on comprehensive diagnostic models and regional adaptation of ICD-10 reveal the ebullience of the diagnostic field, especially when appraised from a broad international perspective. It seems likely that the ongoing tension between universality and diversity in diagnostic systems will continue to yield innovative solutions. Emerging proposals are increasingly involving integrated assessments of health status and according pointed attention to the ethical requirements of psychiatric diagnosis. These proposals must be carefully formulated and thoughtfully and widely evaluated if they are to contribute effectively to the fulfillment of diagnosis as a conceptual and practical tool for clinical care, health promotion, and epidemiology.

SUGGESTED CROSS-REFERENCES

Section 4.1 discusses anthropology and psychiatry, while Section 4.4 discusses cross-cultural psychiatry. Psychiatric diagnosis is covered in Chapter 7, and the DSM-IV classification is covered in Section 9.1. Neurasthenia is discussed in Chapter 16 on somatoform disorders.

SECTION REFERENCES

Acosta-Nodal C, editors: *Glosario Cubano de la Clasificación Internacional de Enfermedades Psiquiátricas*, (GC-1), ed 1. Hosp Psiquiátr Habana, Havana, Cuba, 1975.

Acosta-Nodal C, editors: *Glosario Gubano de la Clasificación Internacional de Enfermedades Psiquiátricas* (GC-2), ed 2. Editorial Científico-Técnica, Havana, Cuba, 1983.

*Andrews G, Slade T, Peters L: Classification in psychiatry: ICD-10 versus DSM-IV. Br J Psychiatry *174:*3, 1999.

Badley EM: An introduction to the concepts and classifications of the International Classification of Impairments, Disabilities, and Handicaps. Disability Rehabil *15:*161, 1993.

Berner P: Der Lebensabend der Paranoiker. Wien Z Nervenheilkd *27:*115, 1969.

Berrios GE: The history of descriptive psychopathology. In *Psychiatry Epidemiology: Assessment Concepts and Methods*, JE Mezzich, MR Jorge, JIM Salloum, editors. Johns Hopkins University Press, Baltimore, 1994.

Bramer GR: International Statistical Classification of Diseases and Related Health Problems—Tenth Revision. World Health Stat Q *41:*32, 1988.

Bustamante JA, Santa Cruz A: Psiquiatría Transcultural. Editorial Cientifico-Técnica, Havana, Cuba, 1975.

Calles Bajo N, Valdés López G, Ceiro Valcarcel E: *Estudio Comparativo en Cuba y Africa acerca de los Conceptos Populares sobre la Enfermedad Mental*. La Habana Empresa Bibliográfica, MINSAP, Havana, Cuba, 1980.

Chinese Medical Association: *Chinese Classification of Mental Disorders*, ed 2, (CCMD-II). Hunan Medical University Press, Hunan, China, 1989.

Chinese Medical Association and Nanjing Medical University: Chinese Classification of Mental Disorders, ed 2, revised. Dong Nan University Press Nanjing, China, 1995.

Cooper JE: The presentation of psychiatric classifications. In *International Classification in Psychiatry. Unity and Diversity*, JE Mezzich, M von Cranach, editors. Cambridge University Press, Cambridge, 1988.

Delgado H: *Curso de Psiquiatría*. Editorial Científico-Médica, Barcelona, 1963.

Echazabal-Campos MA, Otero-Ojeda AA: Uso de sistemas taxonómicos por los psicólogos en Cuba. Rev Hosp Psiquiátr Habana *39:*151, 1998.

Essen-Möller E, Wohlfahrt S: Suggestions for the amendment of the official Swedish classification mental disorders. Acta Psychiatr Scand *47:*551, 1947.

*Fabrega H: International systems of diagnosis in psychiatry. J Nerv Ment Dis *182:*256, 1994.

Fujinawa A: Overview of Japanese experiences on diagnostic classification: Past and present of the classification of mental disorders in Japan. In *Psychiatric Diagnosis. A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, New York, 1994.

González Menéndez R: *El Médico ante el Transtorno Psiquiátrico Menor*, ed # 1. Editorial Oriente, Santiago de Cuba, Cuba 1989.

Horwitz J, Marconi J: El problema de las definiciones en el campo de la salud mental. Definiciones aplicables en estudios epidemiológicos. Bol Oficinia Sanit Panam *60:*300, 1966.

House JS, Landis KR, Umberson D: Social relationships and health. Science *241:*540, 1988.

Huttenlocher J, Hedges LV: Combining graded categories: Membership and typicality. Psychol Rev *101:*157, 1994.

*James A, Kastrup M, Katsching A, Lopéz-Ibor JJ, Mezzich JE, Sartorius N: *Multiaxial Presentation of ICD-10 for Use in Adult Psychiatry*. Cambridge University Press, Cambridge, 1997.

Kastrup MC: Psychosocial domains in comprehensive diagnostic models. In *International Review of Psychiatry*, A Costa e Silva, CC Nadelson, editors. American Psychiatric Press, Washington, DC, 1993.

Kleinman A: *Rethinking Psychiatry: From Cultural Category to Personal Experience*. Free Press, New York, 1988.

Kraepelin E: *Psychiatrie*, ed 5. Barth, Leipzig, 1896.

Laín Entralgo P. *El Diagnostico Medico: Historia y Teoria*. Salvat, Barcelona, 1982.

Lee S: The vicissitudes of neurasthenia in Chinese societies: Where will it go from the ICD-10? Trancult Psychiatry *31:*153, 1994.

Lee S: Self-starvation in contexts—towards the culturally sensitive understanding of anorexia nervosa. Soc Sci Med *41:*25, 1995.

*Lee S: Culture in psychiatric nosology: The CCMD-2-R and international classification of mental disorders. Cult Med Psychiatry *20:*421, 1996.

Leme Lopes J: *As Dimensões do Diagnostico Psiquiátrico*. Agir, Rio de Janeiro, 1954.

Leonhard K: *Afteilung der Endogenen Psychosen*. Akademie, Berlin, 1957.

Li XJ, Li RR, Ren HX, Ren BH, Gu HX, Wu TH, Shi CW: An investigation of inducing factors associated with sudden psychiatric disorders of train passengers. Chin J Psychiatry *29:*47, 1996.

Mezzich JE: Ethics and comprehensive diagnosis. *Psychopathology*, in press.

Mezzich JE: The World Psychiatric Association and the development of ICD-10. In *Psychiatry: A World Perspective*, C Stefanis, AD Rabavilas, CR Soldatos, editors. Excerpta Medica, Amsterdam, 1990.

Mezzich JE, Honda Y, Kastrup MC, editors: *Psychiatric Diagnosis: A World Perspective*. Springer-Verlag, New York, 1994.

Mezzich JE, Kleinman A, Fabrega H, Parron D: *Culture and Psychiatric Diagnosis*. American Psychiatric Press, Washington, DC, 1996.

Mezzich JE, Schmolke MM: Quality of life and comprehensive clinical diagnosis. In *Quality of Life in Mental Disorders*, H Katsching, H Freeman, N Sartorius, editors. John Wiley, New York, 1997.

Orley J, Kuvken W: *International Quality of Life Assessment*. Springer-Verlag, Heidelberg, 1994.

Otero-Ojeda AA: *Adaptación Cultural del Sistema Multilaxial de la CIE-10 A Través De Ejes Complementarios*. Hospital Psiquiatrico de la Habana, Editorial Científico-Técnica, Havana, Cuba, 1994.

Otero-Ojeda AA, Acosta Nodal C: *Características y Aportaciones Fundamentales del Tercer Glosario Cubano de Psiquiatría*. Hospital Psiquiátrico de La Habana, Havana, Cuba, 1996.

*Otero-Ojeda AA, editor: *Tercer Glosario Cubano de Psiquiatría*. Hospital Psiquiátrico de La Habana, Havana, Cuba, 1998.

Percy C, van Holten V, Muir C: *International Classification of Diseases for Oncology* (ICD-0). World Health Organization, Geneva, 1990.

Pull C, Chaillet G: The nosological views of French-speaking psychiatry. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, New York, 1994.

Sartorius N, Kaelber CT, Cooper JE, Roper MT: Progress towards achieving a common language in psychiatry. Results of the field trials of the clinical guidelines accompanying the WHO Classification of Mental and Behavioral Disorders in ICD-10. Arch Gen Psychiatry *50:*115, 1993.

Seguin CA: The concept of disease. Psychosom Med *8:*252, 1946.

Strauss JS: The person—key to understanding mental illness: Towards a new dynamic psychiatry, III. Br J Psychiatry *161:*19, 1992.

Survey and Test Group for the CCMD-2-R: A report on the national field of the second revised edition of the Chinese classification and diagnostic criteria of mental disorders. Chin J Psychiatry *29:*27, 1996.

Üstün TB, Goldberg DA, Sartorius N: ICD-10. Classifying primary care mental disorders. Proc 146th annual meeting of the American Psychiatric Association, May, 1993.

World Health Organization: *The ICD-10 Classification of Mental and Behavioral Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization, Geneva, 1992.

World Health Organization: *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. World Health Organization, Geneva, 1993.

World Health Organization: *International Statistical Classification of Diseases and Related Health Problems*, rev 10, vol 1. World Health Organization, Geneva, 1992.

Wu CY: A clinical analysis of seventy six cases of qigong induced psychotic disorders. J Clin Psychol Med *2:*7 (in Chinese), 1992.

Xu WY, Chen ZJ: An eight to ten year outcome study of unipolar mania. Arch Psychiatry *4:*88 (in Chinese), 1992.

Textbook of Psychiatry

CHAPTER 10. DELIRIUM, DEMENTIA, AND AMNESTIC AND OTHER COGNITIVE DISORDERS

ERIC D. CAINE, M.D. AND JEFFREY M. LYNESS, M.D.

Psychiatry is in the midst of a profound transformation, at once struggling to incorporate a dynamic understanding of neuroscience and molecular biology while maintaining a view of unique persons or individuals as the central focus of therapeutic intervention. To date it has been beyond the scope of knowledge to effectively integrate research data regarding individual differences with more abstract findings regarding fundamental aspects of brain development or aging-related neurodegeneration. Discovering the bases for the major neuropsychiatric diseases can be expected to provide powerful clues for defining the nature of how neurobiological processes are expressed as emotions, thoughts, or actions, or how life events and daily experiences alter and shape brain growth and development.

Since the late 1980s, a major conceptual transition has occurred in the way clinicians and researchers view the relation between mental disorders and brain function. For much of the past century psychiatry was trapped in an either-or dilemma—either a condition was viewed as a symptomatic manifestation of structural cerebral or systemic pathology (organic), or it was considered psychological or emotional in nature (functional). However, clinicians recognized that there are no behaviors that do not involve the brain, and that the transmission of culturally derived processes from individual to individual is influenced by each person's central nervous system (CNS). Behaviors defined by some cultures as abnormal may be mediated by normal neurophysiology; in contrast, patients with damaged brains may develop compensatory strategies reflective of CNS plasticity to ameliorate the effects of disordered neural systems. The same behavior (e.g., suicide) may reflect normal or abnormal physiology. For many diseases the CNS develops normally but acquires its dysfunction later in life whereas other diseases may reflect aberrant wiring patterns or connections that function in a neurochemically or neurophysiologically normal fashion. Despite such complexity, clinicians often depended on a single criterion for defining organic disorders—either the detection of a structural lesion or, less often, the diagnosis of a known disease process; however, such an approach is no longer satisfactory.

Diagnostic decisions have depended largely on available technology. For more than a century gross postmortem examination and microscopic histopathological examination were the primary tools. Working in a tradition of clinical-pathological correlation, psychiatrists, neurologists, and others used a dichotomous approach to both diagnosis and classification—that is, a lesion was either present or absent. This approach proved heuristically limiting and became increasingly unrewarding. Indeed, histopathology is now no longer the gold standard for defining brain-based diseases. It was only marginally useful for psychiatry, and in the near future molecular biological methods will replace it for neurology. The advent of new technologies already has undermined the pseudocertainty of earlier years. Many patients with functional syndromes are found to have CNS abnormalities when studied with magnetic resonance imaging (MRI), positron emission tomography (PET), or single photon emission computed tomography (SPECT). Should these syndromes be reclassified as organic? Most would argue that such changes would be premature because the pathophysiological significance of newer findings remains obscure. Such arguments also are pertinent to the entire question of organic versus functional.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV) establishes a new approach to those questions. The categories of organic and functional have been abandoned. When a psychopathological syndrome is known to be a symptomatic manifestation of a systemic medical or cerebral disorder, it is designated as “due to . . .” (secondary), with a designation of the specific disease process. When it is considered to be an idiopathic psychiatric disturbance, it is designated primary. A clinician should follow a careful process of case reasoning before settling on the primary or secondary status of a disorder. To appropriately diagnose a patient as having an idiopathic (i.e., primary) condition, the clinician must necessarily exclude all definable, potentially etiological disease processes.

The clinician must exercise equal caution before diagnosing a disorder as secondary or symptomatic. To date there are no widely accepted guidelines for establishing probable causal relations between psychopathological conditions and detected cerebral abnormalities. Traditionally, such assignment of probable causality has been left to clinical judgment. DSM-IV outlines such guidelines; they have the effect of encouraging the clinician to undertake a thorough evaluation and to postulate causal connections conservatively.

DEFINITION

The primary-secondary classification, like similar classifications, reflects the thinking of its time. The change in terminology in DSM-IV from *organic* to *due to* . . . is more than cosmetic in that it captures the conceptual shift away from structure and lesion and toward active disease process and etiology. The broad group of cognitive disorders includes dementia, delirium, amnestic disorder, and other syndromes in which disordered cognition caused by known (or presumed) disease entities is the central characteristic featuret (Table 10-1). Specific secondary syndromes are scattered through the nosology, classified along with other phenomenologically similar clinical conditions (e.g., mood disorders due to general medical conditions are grouped among the mood disturbances). Such groupings are intended to foster differential diagnostic consideration; the changes in DSM-IV are intended to enhance rigorous clinical reasoning. Use of more specific designations (e.g., mood disorder due to thyroid deficiency, with major depressive-like episode) strengthens diagnostic specificity when contrasted to the previously used organic mood disorder and lays the foundation for more meaningful comparative research.

Delirium
Delirium due to a general medical condition
Substance-induced delirium
Delirium due to multiple etiologies
Delirium not otherwise specified
Dementia
Dementia of the Alzheimer's type
Vascular dementia
Dementia due to other general medical conditions
Dementia due to HIV disease
Dementia due to head trauma
Dementia due to Parkinson's disease
Dementia due to Huntington's disease
Dementia due to Pick's disease
Dementia due to Creutzfeldt-Jakob disease
Dementia due to other general medical conditions
Substance-induced persisting dementia
Dementia due to multiple etiologies
Dementia not otherwise specified
Amnestic disorders
Amnestic disorder due to a general medical condition
Substance-induced persisting amnestic disorder
Amnestic disorder not otherwise specified
Cognitive disorder not otherwise specified

Table 10-1 DSM-IV Cognitive Disorders

Throughout this chapter the term “neuropsychiatry” is used in reference to the field of medicine that considers the brain bases of mental disorders. In the United Kingdom this field is sometimes called organic psychiatry. At one time nearly all of psychiatry was neuropsychiatry; at another time, few would have chosen that label. Considering the brain substrates of behavior necessarily forces clinicians and researchers also to recognize the experiential, psychological, social, and cultural aspects of the patients and the problems they encounter.

HISTORY

The development of neuropsychiatry and the growth of general psychiatry coincided with competition and ultimately cooperation between public psychiatric asylums (now called hospitals or centers) and clinical practice in universities and private offices. Different ideologies or dogmas developed, depending on whether the clinician was seeing principally institutionalized psychotic patients, for whom there was little hope for improvement or recovery, or ambulatory patients, whose apparent psychological accessibility gave rise to therapeutic optimism. Additionally, psychiatrists in the asylums (often called alienists) had different needs than the nerve

doctors or neurointernists who saw the walking wounded in their offices.

Table 10-2 presents a brief categorization of historical periods in neuropsychiatry. It is probably not presumptuous to state that Wilhelm Griesinger (1817–1868) created neuropsychiatry with the publication of his book in 1845, crafted after practicing 2 years in Winenthal, one of the leading German asylums. An advocate of physiological medicine, Griesinger attempted to steer German medical practice away from both the romantic and somatic schools of that time. He asserted that psychiatry was part of medicine and that “psychological diseases are diseases of the brain.” He also advocated knowing one's patients well, understanding their life course, and appreciating how their mental disease affected their overall functioning. He advanced a specific notion of the ego that attempted to explain all disease under a single conceptual view based on a gradual pathological erosion of ego integrity. He supported the idea of careful neuropathological observation, although he never pursued such work in the later fashion of Theodor H. Meynert, Karl Wernicke, or Alois Alzheimer.

1845–1865	Griesinger—Integration I
1865–1900	Neuropathological and descriptive psychiatry
1900–1930	Gradual transition I
1930–1950	Ferment
1950–1965	Psychodynamic era (United States)
1965–present	Gradual transition II
The future	Integration II

Table 10-2 Historical Periods in Neuropsychiatry

Despite Griesinger's attempted integration, many of his notions now seem simplistic or misleading, especially the idea that all mental illnesses reflected one basic pathological process that could be divided into stages. The first of Griesinger's disease stages involved an assault on the ego by the basic disease, although no frank pathological disruption was apparent. In the second and third stages, ego disintegration was completed and permanent brain changes took place. Griesinger believed that therapeutic intervention would be successful only during the first stage. Griesinger and his contemporaries made no particular distinction between psychiatric and neurological problems, and patients with progressive neurological diseases were seen in asylums like Winenthal. He proposed joint psychiatry and neurology clinics, and founded one in Berlin in 1861. Most importantly he catalyzed the development of university neuropsychiatry in contrast to the asylum psychiatry that was prevalent in his day, and thus provided the means for developing academic, research-based approaches to questions that had largely been outside rigorous medical scrutiny.

Meynert (1833–1893), the next major player on the neuropsychiatric scene, steered away from Griesinger's integrative center toward an extreme of neuropathological determinism. Meynert's 1874 book, titled *Psychiatry: Diseases of the Forebrain*, largely dealt with neuroanatomy. He is probably best remembered for his histopathological studies and has deservedly been called a pioneer of neuropsychiatric pathophysiology. Meynert also consolidated within universities what proved to be both a sterile theoretical position and a form of clinical psychiatry that had little benefit for either patients in the asylum or the walking wounded. Perhaps unwittingly, Meynert and his intellectual colleagues placed neuropsychiatry in a position where it would decline. It was presumptuous to believe that all clinically significant behavioral disturbances had a demonstrable cerebral substrate, especially given available laboratory techniques. Such work took place in a context of minimal understanding of the basic aspects of neuronal or regional cerebral functioning. It was ironic that the driving investigative force, a search for pathologically defined brain abnormalities, was to become a basis for undoing the field. Psychiatry in general and neuropsychiatry in particular have been plagued by a sense of intellectual exclusivity—the either-or dilemma—in their intellectual conceptions. This sense of exclusivity and the related tendency to decry integrative (multidetermined) theoretical approaches may have reflected the ultimate conceptual complexity of the research and clinical tasks that have confronted those who would understand the cerebral bases of behavior and mental disorders: “the brain we are studying is more complex than the brain that is studying it.”

In the context of these limitations of neuropsychiatry, two fundamentally different paths emerged. The first was exemplified by the work of Emil Kraepelin (1856–1926). Although Kraepelin supported the neuropathological work of Alzheimer, he spent considerable effort developing a rigorous clinical classification of psychiatric disorders, particularly those observed in asylum settings. The classification was largely atheoretical, based as it was on form and course. Kraepelin hoped that clinical description and classification would ultimately lead to pathological correlation, a hope that has yet to be realized fully. The second path was developed by nerve doctors, neurointernists who saw their patients in offices or on the wards of neurological hospitals. Jean Martin Charcot (1825–1893) and Sigmund Freud (1856–1939) were notable among those practitioners. The idea of looking at an individual's development in the context of early life experience did not originate with Freud (Griesinger had also advocated it), but Freud pushed farthest the notion of defining the meaning of particular behaviors in terms of real and imagined life events. The different paths blazed by Freud, Kraepelin, and Meynert coexisted during the early decades of the twentieth century, with no one route of investigation clearly predominant. Clinical-pathological correlation had its greatest triumphs with the recognition of the causes of general paresis and pellagra. However, the large asylums remained full, and there were no specific therapies for clinicians to use.

In the United States, by contrast, the period from 1930 to 1950 was a time of great ferment and change, with an examination of new ideas and therapies. Shortly before World War II clinicians experimented with a variety of somatic interventions and opined about the cerebral bases of the major psychiatric disorders. New treatments, including barbiturate coma, insulin shock, and the convulsive therapies, were developed. During this time of novel therapeutics, clinicians undertook what seemed a logical step to many—ablative neurosurgical intervention, ultimately dubbed psychosurgery. Although it was based on poorly substantiated notions of cerebral functions and how they went awry in the major psychiatric disorders, frontal lobotomy spread rapidly in the United States following World War II, fueled by a desire to empty large state mental hospitals and reduce public expenditures for patients with chronic mental disorders. Psychosurgery offered the prospect of instituting a definitive medical procedure that either cured or markedly improved previously intractable syndromes. However, psychosurgery and its practitioners failed to fulfill their promises and neuropsychiatry eventually became a term of opprobrium. By the mid-1950s brain-oriented views of behavior were widely considered to offer few clinically or theoretically fruitful insights, and brain-oriented psychiatrists were seen as useless or even clinically harmful to those they treated. Juxtaposed with the 1930s' plunge into organic psychiatry and its therapies was the growth of psychoanalysis, sparked by revolutionary theories brought from Europe by analysts fleeing Adolf Hitler. Young neuropsychiatrists, neurologists, and neuroscientists proved a receptive audience for these ideas, as they discovered far greater explanatory power in the notions of Freud than in those of Charles Sherrington and the doctrine of nerve transmission. Enthusiasts found analytic insights filling unmet needs: Freud's theories and techniques supplied both tools for data collection through the free-associative interview and a coherent system to organize these findings. More importantly, these techniques directed specific interventions and provided the physician with something to do beyond watching impaired patients remain unchanged or become progressively worse.

Stanley Cobb (1887–1968) was among the leaders in the attempt to integrate psychiatry and neurology. Cobb trained in neuropathology and taught the basic neuropathology course at Harvard Medical School for several generations. A student of Adolf Meyer, he espoused a dynamic life course view. Although he maintained an appreciation for neuropathology, he moved away from a primary interest in cerebral circulation to a consideration of psychiatric disorders during the late 1920s and the 1930s. Cobb developed a conceptual pyramid as an integrative device to illustrate his views, deliberately leaving a gray, uncharted zone between pathology and clinical psychiatry. Despite subsequent advances in neuroscience, the uncharted zone seems no less opaque now than 60 years ago, when Cobb first published the conceptual pyramid in his textbooks.

The years immediately after World War II were a time of rapid change, away from neuropathology-based psychiatry and toward psychodynamic and psychoanalytic psychiatry. The growth of psychiatry departments and medical schools was spurred by federal initiative, as was the deinstitutionalization of the seriously ill. Economic motives contributed to the latter, but more important was the sense that a therapeutic triumph might be at hand. This sense was associated with the optimism following World War II of psychiatrists and psychiatrists-to-be (often physicians from other disciplines who were assigned to wartime psychiatry services), coupled with the hope for the successful use of psychoanalysis in a wide variety of disorders and the development, during the 1950s, of a more specific psychopharmacology. Notably, psychopharmacology did not reflect a greater degree of neuroscientific understanding; rather, serendipity, clinical acumen, and innovative thinking served as guiding beacons. Later developments of new compounds did result from attempts at pharmacological modeling. Indeed, more recent understandings of CNS function were catalyzed by having specific agents that could reliably alter brain activity. Central to this activity were industry-sponsored initiatives to discover and develop novel pharmacotherapeutic medications.

Formal psychiatric classification and nomenclature evolved during the post-World War II era. Prior to the adoption of the first edition of DSM (DSM-I) in 1952, psychiatric hospitals used the *Statistical Manual for the Use of Hospitals for Mental Disease*, first published under a slightly different title in 1918. Nearly all the categories in the manual were used for classifying patients with brain-related mental disturbances. As noted in DSM, that approach proved suitable for only about 10 percent of the cases seen by the Armed Forces during World War II. In contrast to the *Statistical Manual for the Use of Hospitals for Mental Disease*, DSM outlined two broad categories, one for disorders caused by or associated with impairment of brain tissue function (divided into acute and chronic), and the second for those of psychogenic origin without clearly defined physical cause or structural change in the brain. The brain disorder section classified conditions by duration and defined etiology (e.g., infection, intoxication, and tumor) with no attention to clinical phenomena whereas the psychogenic section began the move toward a more clinically specific categorization. For the latter, difficulties adjusting to internal and external stresses were the key pathogenic factors. Thus, a psychological theory officially supplanted the dominant brain view of earlier classification manuals. Despite the change in dominant explanatory theory, DSM also maintained and further codified an either-or philosophy set forth by the earlier post-Griesinger neuropsychiatrists.

A similar stance was taken in the second edition of DSM (DSM-II), with a separation of organic brain syndromes from psychoses not attributed to physical conditions listed previously. Brain syndromes were said to result from diffuse impairment of brain tissue and to be manifested by the following symptoms: impairment of orientation, memory, all intellectual functions (e.g., comprehension and calculation), and judgment, as well as lability and shallowness of affect. The organic brain syndromes were divided into psychotic and nonpsychotic conditions, the former also including senile and presenile dementia, depending solely on the severity of functional impairment. Beyond that crude separation, there were no specifying clinical features; further classification depended on defining a cause. *Acute* and *chronic* were indicated as diagnostic subcodes. Psychodynamic psychiatry was not successful in treating the more seriously impaired residents of the state hospitals, and often was found wanting among ambulatory populations. Competing approaches sprang up that claimed similar or greater effectiveness. Overall, it has been difficult to definitively demonstrate treatment success when using psychotherapeutic modalities, although recent efforts at treatment evaluation have proved both more enlightening and more promising. A neurochemically oriented biological psychiatry took hold and became pre-eminent in the research laboratory, if not always in the clinic setting. However, there has been no successful jump from synapse to behavior, integrating understanding anew. This dearth of explanation has made a fertile soil for the reemergence of neuropsychiatry.

COMPARATIVE NOSOLOGY

DSM-III and DSM-III-R The third edition of DSM (DSM-III), published in 1980, and the 1987 revised third edition (DSM-III-R) moved to discard the theoretical underpinnings based on stress-related psychological reactions and emphasized phenomenology as part of an innovative multiaxial system of classification. Nonetheless, the organic versus nonorganic dichotomy was maintained. Organic mental disorders were clearer in their clinical typology, with a greater array of subtypes and causes, related either to Axis III physical disorders or conditions or to use of psychoactive substances. Importantly, categories for mood, personality, anxiety, and hallucinatory and delusional disorders were added. These changes were made in an effort to increase recognition of clinical variations, but they lacked sufficient descriptive detail to allow reliable comparisons with idiopathic Axis I syndromes. Unfortunately, like DSM-I and DSM-II, DSM-III provided no guidelines or discussion on the question of causal connection between systemic medical or cerebral diseases and secondary psychiatric manifestations. Thus, there has been no consideration of the clinical reliability of the organic designation or of its validity.

Scientific Developments Since the 1970s a period of transition characterized by the absence of a dominant theoretical view within psychiatry has been under way. Scientific developments outside the field have profoundly shifted the direction of psychiatric thought and have come largely from behavioral neurology, clinical neuropsychology, and basic laboratory neuroscience. Challenged by the rehabilitative needs of many patients returning from World War II with focal cerebral lesions, a small number of neurologists and quantitative psychologists began to study the effects of those injuries just as psychiatrists were shifting their attention away from cerebral processes. This work quickly expanded to include patients with vascular lesions. By the late 1960s behavioral neurologists and clinical neuropsychologists were recognized specialists, although few in number, and a growing literature examined the intellectual and behavioral consequences of specific regional cerebral lesions. These writings stimulated a modern resurgence of the clinical-pathological correlative tradition first developed during the late 1800s. Since the late 1980s psychiatric researchers have drawn increasingly from the lessons of focal lesion models.

Simultaneously, basic laboratory neuroscience burgeoned, with important findings reported at a fast pace in recent years. Researchers have moved from a focus on synapses and neurons to subneuronal molecular biological processes. Investigative techniques have changed extremely quickly, facilitating the detection of variations in complex neurobiological systems while outstripping the ability to define the terms *normal* and *abnormal*.

Even as psychiatric researchers studied focal lesion syndromes with greater enthusiasm, the shortcomings of those models became more apparent. Most importantly, they involve cerebral substrates distinctly different from those of the major idiopathic psychiatric syndromes. Whereas strokes reflect vascular anatomy, major psychopathology is based on dysfunction in interacting and widespread neurochemical systems. Many clinical psychiatric disorders reflect long-term (perhaps developmental) abnormalities that affect psychological growth and interpersonal events across the life span. Focal cerebral lesions on the other hand are acquired later in life in the context of a developmentally intact CNS; thus, analogies must be drawn with caution.

Psychiatric researchers face another, more daunting obstacle when considering the brain bases of mental disorders. Since the 1970s many have chosen to compare specific diagnoses with putative CNS alterations. The exclusive use of categorical diagnoses, while beneficial for enhancing clinical rigor, has yet to prove rewarding when applied to diagnosis-brain correlative research paradigms. Although empirically defined clinical syndromes form the bedrock of modern psychopharmacology, they may not be amenable to fundamental neurobiological characterization.

The historic lessons of neuropsychiatry include the realization that a defined or unified etiology may be associated with a striking diversity of clinical presentation (e.g., general paresis and Huntington's disease); conversely, specific syndromes are often the manifestation of heterogeneous etiologies. Here again, maintenance of a rigid either-or view is likely to prove disappointing. Rather, future researchers will be required to understand how experiences in the midst of normal development permanently change the brain. Too often neuropsychiatry has viewed the universe along the trajectory of brain to behavior. It will be just as important to understand it from behavior to brain and back again. Psychiatry as a broad field, and neuropsychiatry in particular, must address an array of key questions regarding brain-behavior and behavior-brain relationships.

1. How do developmental abnormalities that occur in utero or in the early years of life lead to later emerging psychopathology? What are the genetic and environmental causes of these defects, and what are their molecular biological modes of expression?
2. How does environment and life experience change brain structure and function? In essence, what are the mechanisms of brain plasticity? How do normative and abnormally stressful life events lead to permanent changes in neural function?
3. How is brain and organism homeostasis maintained or lost? What are the neural underpinnings of cyclical or episodic changes in behavioral, emotional, and cognitive functioning?
4. What are the genetic, molecular, and environmental factors associated with brain aging and normative cognitive decline, and neurodegenerative diseases?

The limits of neuroscientific inquiry need to be recognized. Neuroscience will not improve the interviewing skills of an individual psychiatrist; moreover, it will be many years or decades before there is a clear understanding of the neural bases for individual differences: When the field progresses to the point where clinicians and scientists can merge an appreciation of a person's life story with knowledge regarding laws of the nervous system, it will finally reach a level of true integration.

DSM-IV The change in terminology in DSM-IV from *organic* to *due to* represents more than a simple semantic alteration—it emphasizes the need to define etiology, not site or structure. The term "*organic*", as used for many years pointed to defined pathological lesions and was contrasted with the term " *functional*" or physiological abnormalities that could not be detected by existing laboratory procedures. In the absence of sensitive and specific diagnostic laboratory tests, descriptive laboratory technology was often misapplied, giving a false impression of diagnostic validity. Similarly, the presence of a definable abnormality was considered sufficient to establish an organic diagnosis even though no standards were available for setting a threshold of evidence or data needed to attribute the cause of a symptom to an observed lesion.

DSM-IV takes a conservative approach to the problem. Establishing a secondary diagnosis should, whenever feasible, follow a chain of reasoning that etiologically connects a psychopathological syndrome with a systemic medical or primary cerebral disorder. The coexistence of Axis I and Axis III diagnoses in an individual case is not sufficient to infer a causal relationship, even when an apparent association or correlation is present. To more confidently determine whether an association is causal, the clinician should attempt to define the strength (relative risk), consistency of form, specificity, coherence of association, and temporal relation of clinical manifestations to the proposed disease process. Defining each attribute may not be feasible for all disorders or in every case, but it does provide stronger ground for

advancing an etiological link. When insufficient data are available to establish a causal relation, it is preferable to provide unlinked Axis I and Axis III diagnoses.

It is notable that the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) maintained “organic” as a superordinate category. Although many specific examples of syndromic diagnostic criteria are similar to those in DSM-IV, ICD-10 retained the approach favored by earlier editions of the DSM. Where DSM-IV strives to highlight steps necessary for establishing a primary general medical diagnosis, ICD-10 states: “Rather, the clinical manifestations resemble, or are identical with, those of disorders not regarded as ‘organic' in the specific sense restricted to this block of the classification. Their inclusion here is based on the hypothesis that they are directly caused by cerebral disease or dysfunction rather than resulting from either a fortuitous association. . . or a psychological reaction to its symptoms. . .”

DIAGNOSIS

Thorough clinical evaluation forms the basis for diagnosing secondary disorders. Beyond a detailed personal history and mental status examination, the clinician often must depend on supplementary laboratory evaluation, including such procedures as cerebral imaging, neuropsychological testing, and electroencephalography (EEG). Four steps form the basis for establishing a secondary (“due to . . .”) diagnosis with greater confidence: (1) definition of the specific psychopathological syndrome, (2) delineation of other manifestations of the primary disorder, (3) demonstration of active cerebral or systemic disease, and (4) demonstration of an elevated prevalence between the proposed etiological disorder and the described psychopathological picture. These steps may not always occur sequentially, as both syndrome and disease may be recognized.

Definition of the Specific Psychopathological Syndrome It is essential to describe the clinical disorder as precisely as possible. Subtyping should be undertaken when feasible, particularly with the specification of target symptoms for treatment. Use of broader or not otherwise specified terminology is available for less phenomenologically specific cases. Ideally, the clinician seeks to establish etiological relationships between definable disease processes and specific clinical presentations. The multiple presentations of general paresis, however, underscore that one pathogenic agent may cause multiple syndromic forms.

Patients with many secondary psychiatric disturbances present with symptoms that are atypical of primary (idiopathic) psychiatric disorders. Other clinical features, such as older age at onset, may serve to raise the index of suspicion. Syndrome definition involves severity as well as form. Severity implies a continuum, and the application of a diagnosis implies that the disorder has exceeded a threshold of severity. For example, although behavioral changes often arise following a cerebral lesion, a categorical diagnosis is not warranted when symptoms have not had a measurable impact on a person's functional integrity. Researchers may wish to study mildly symptomatic phenomena, but clinicians typically reserve diagnoses for conditions that cause substantially disordered behaviors, those interfering with the patient's daily life and personal well-being.

Delineation of Other Manifestations of the Primary Disorder Secondary psychopathological syndromes rarely occur alone but typically keep company with other symptoms and signs of the primary systemic or cerebral disorder. Thus, it is essential to define those cognitive, neuropsychological, peripheral, or other clinically ascertained manifestations of the disease process. For example, human immunodeficiency virus (HIV)-induced manic symptoms typically are accompanied by signs of testable cognitive impairment whereas depression due to Huntington's disease can be diagnosed with confidence only in the setting of a defined movement disorder. Identifying co-occurring manifestation provides an overall clinical context for more confidently establishing a secondary diagnosis.

Demonstration of Active Cerebral or Systemic Disease The clinician should seek nonbehavioral confirmation of the primary disease process. Such confirmation typically involves laboratory testing, including the full array of medical diagnostic procedures. One must be cautious, however, in the interpretation of many tests. An example is the use of cerebral imaging in psychiatric patients. Detection of a structural abnormality on computed tomography (CT) or MRI is not equivalent to demonstrating active cerebral disease because such imaging studies provide static (i.e., nonphysiological, nonfunctional) information in most applications. Much remains unknown regarding the link between MRI findings, definable cerebral pathology, and specific pathophysiologies or diseases.

Elevated Prevalence Rate Between Proposed Etiological Disorder and Described Pathological Picture This guideline cannot always be fulfilled, but argues for utilizing data-based conclusions that can be applied to clinical practice. Simply recording that a change in behavior occurs after the emergence of a particular cerebral disorder, for example, is insufficient proof. A specific syndrome should occur with a prevalence in association with an etiological disorder that is above the base rate in the general population.

Many clinicians recommend as the principal criterion for establishing causality the demonstration of a close temporal association of onset and course of the primary disorder and the secondary psychiatric syndrome. Although frequently useful, this criterion is not always applicable. For example, symptomatic psychosis due to epilepsy may gradually emerge 10 to 15 years after the onset of seizures. Conversely, psychiatric symptoms and signs may be the first clues to a systemic or cerebral disease, and detection of the primary pathological process may follow the emergence of psychiatric symptoms by months or longer. Many secondary psychiatric conditions also may persist after the primary disease process has resolved; examples are the secondary conditions consequent on thyroid deficiency, long-term alcohol use, or long-term exposure to neurotoxic compounds. Secondary syndromes may remit quickly, slowly, or incompletely, depending on the specific disease and whether lasting cerebral changes are present. Also, secondary syndromes may be amenable to symptomatic treatment even while the primary disorder remains without a cure.

ICD-10 recommends four criteria for classifying a syndrome as organic: (1) evidence of cerebral disease, damage, or dysfunction, or of systemic physical disease, known to be associated with one of the listed syndromes; (2) a temporal relationship (weeks or a few months) between the development of the underlying disease and the onset of the mental syndrome; (3) recovery from the mental disorder following removal or improvement of the underlying presumed cause; and (4) absence of evidence to suggest an alternative cause of the mental syndrome (such as a strong family history or precipitating stress).

ICD-10 also lists conditions known to increase the relative risk for the syndromes classified here, including epilepsy; limbic encephalitis; Huntington's disease; head trauma; brain neoplasms; extracranial neoplasms with remote CNS effects (especially carcinoma of the pancreas); vascular cerebral disease, lesions, or malformations; lupus erythematosus and other collagen diseases; endocrine disease (especially hypothyroidism and hyperthyroidism, Cushing's disease); metabolic disorders (e.g., hypoglycemia, porphyria, hypoxia); tropical infectious and parasitic diseases (e.g., trypanosomiasis); toxic effects of nonpsychotropic drugs (propranolol [Inderal], levodopa [Dopar], methyldopa [Aldomet], steroids, and antihypertensive and antimalarial agents).

Ultimately the clinician must make an informed judgment as to whether the psychiatric condition is primary or secondary. Prevalence data, for example, reflect group trends whereas the clinician has to make a decision regarding an individual. Two approaches are available. The clinical decision is relatively uncomplicated if a previously demonstrated elevated prevalence links a specific syndrome with a specific etiology, in the presence of additional supporting clinical features and consistent laboratory tests. Probabilistic reasoning in such cases leads to the conclusion that there is a cause-and-effect relation. A temporal association, when meaningfully present, further confirms the connection.

When a clinical research data base is less well established, however, it becomes even more critical to document rigorous clinical reasoning, in effect demonstrating how the detected historical, clinical, and laboratory features are not consistent with what is known about idiopathic psychiatric conditions. Again, establishing the causal connection should reflect the clinician's effort to undertake conscientious probabilistic case reasoning. Attribution of a secondary designation implies a link that is more probable than not: a standard that exceeds 50 percent is recommended, although absolute certainty may be possible. Such a standard does not require that the systemic medical or primary cerebral disease be the sole factor contributing to symptom expression; rather, *due to* . . . connotes a predominant pathogenic role. When causal probability is considered less certain (i.e., possible but not probable), the clinician should not define a syndrome as secondary in nature.

Diagnostic decisions based on incomplete data will be inevitable, and thoughtful clinical judgment remains the abiding rule. *Due to* . . . should be used conservatively; attribution of cause invites a careful consideration of the factors that contribute to disease formation. When doubt remains, the provision of a primary (idiopathic) psychiatric diagnosis will best serve the interests of the patient and avoid the premature closure of clinical evaluation.

Neuropsychiatric Assessment Neuropsychiatric assessment follows the principles of all comprehensive clinical evaluations: it is based on thorough acquisition of the current and past medical history, family history, developmental and social history, and a review of personal habits. The neuropsychiatric clinician seeks to integrate the data on unique individual development, the signs and symptoms of disease, and an understanding of behavior-brain relationships into a meaningful appraisal of functional integrity.

Clinical reasoning should entail a time-oriented view, with the clinician noting how the patient has progressed or failed to develop across his or her life course. The temporal perspective is buttressed by an understanding of normative development as well as by an appreciation of the natural history of disease processes. In

particular, the clinician should be mindful of the unique characteristics of primary cerebral disorders, whether inborn, acquired early in life, or of later onset. Fundamental to neuropsychiatric evaluation, diagnosis, and prognosis is an understanding of disease evolution at psychological and neurobiological levels of analysis. Frequently the clinician must tolerate the uncertainty of not knowing (in an absolute sense) the mechanisms by which brain diseases cause behavioral problems; the clinician then has the task of developing practical and effective solutions to problems that may not have specific scientific answers. Despite having a recognized pathological basis, most neuropsychiatric disorders do not have specific cures and continue to require empirical, symptomatic treatment approaches.

Clinicians repeatedly face a dual dilemma: etiological specificity often is related to variable clinical expression (e.g., general paresis and Huntington's disease). There is no biological law such as “one pathogen, one clinical presentation.” Clinical variability is the rule rather than the exception. Conversely, there are relatively few final common pathways for the expression of a wide variety of disease processes. These pathways include (1) alterations in arousal, attention, and concentration; (2) alterations in affective state, including both the expression of emotion and the feeling of mood; (3) alterations in perception, including ideational or physical and internal or external; (4) alterations in intellectual function (such as memory, language, or the organization of thought processes); (5) alterations in personality; and (6) alterations in motor function. Thus, behavioral abnormalities tend to be nonspecific, and despite substantial evidence from behavioral neurology that focal lesions may lead to distinctive patterns of intellectual deficit, there are insufficient data to confidently support such assertions for major psychopathological syndromes. Moreover, idiopathic or primary psychiatric disorders may mimic symptomatic psychopathological conditions that are secondary to specific systemic medical or cerebral disease processes and vice versa. What confounds the situation is that there has been insufficient research to establish how often cerebral lesions lead to discrete psychopathological syndromes or specifically how any defined psychopathological disorder is related to a particular localized cerebral abnormality.

Thus, the clinician must use an empirical method, based on careful clinical reasoning, that allows the development of a preliminary diagnosis and an initial treatment plan. The clinician should specify in advance what possible therapeutic benefits might be derived and should understand how the natural history of the disorder will unfold if proposed treatment options prove ineffective. The clinician should also be ready to undertake further evaluation if an unanticipated outcome arises. By establishing a future-oriented or outcome-oriented clinical perspective, the clinician can reduce the degree of uncertainty and establish a structural approach for systematically and self-critically scrutinizing treatment interventions.

Neuropsychiatric Case Reasoning The approach to neuropsychiatric case reasoning required for such formulation and planning entails blending the disparate traditions that developed in psychiatry in the past century. It draws from Meynert and John Hughlings Jackson (1835-1911), as well as from the behavioral neurologists of recent decades, an appreciation of brain-behavior and behavior-brain relations, with an attempt to understand the laws that govern the CNS. Such a pathobiological method, through lesion location and an appreciation of probabilistic generalities about brain function and neuropathology, in effect argues that all nervous systems are created equal. It benefits from a thoroughly documented array of case studies that seek to define the specific behavioral expression of focal cerebral lesions. It suffers from the fact that nervous systems are not identical and that personal circumstances powerfully influence the expression of disease. Nonetheless, it has taught clinicians much about what to assess and expect when dealing with disordered brain function.

The second approach to case reasoning is derived from the Kraepelinian tradition that continues to be expressed in DSM-IV. This method argues for the precise elaboration of symptoms and signs, the definition of specific syndromes, the identification of target symptoms amenable to therapeutic intervention, and the use of diagnoses for prognostic purposes. The strengths of such an approach lie in the rigorous case definition based on thorough observation and data collection, with the derived ability to generalize from one case to another. Shortcomings, akin to the problems with lesion localization, include the substantial degree of variability that exists within the boundaries of stereotypic diagnostic descriptions and the difficult-to-quantify influences of personal life circumstances.

The third method of case reasoning evolved from dynamic psychiatry and recognizes the individual as having unique personal and developmental attributes that are expressed throughout the course of life. The clinician using this method of case reasoning seeks to understand meaning as well as event and to appreciate disease process in the patient's broader social and cultural context. The neuropsychiatric clinician must view illness in all its complexity.

The different modes of case reasoning are brought together for clinical purposes through understanding how psychological meaning, symptoms, and disease process and socially defined aspects of illness each affects the patient's ability to function autonomously. Although function is not a direct measure of pathology or disease process, assessing how each person has undertaken specific developmentally important tasks is useful for appraising the interaction of those factors. Depending on the individual case, it may be possible to more clearly state which method of case reasoning is most effective for developing a treatment plan and understanding aspects of prognosis. Ultimately, the clinician is as much interested in the patient's return to prior functional integrity as in symptom remission. Treatment success cannot be proclaimed, for example, in the resolution of psychotic symptoms associated with epilepsy if the patient continues to be socially withdrawn or isolated and no longer capable of independent living.

Aging Age and its relation to the expression of illness must be recognized as a changing backdrop for all neuropsychiatric disorders. Age may be used as a convenient indicator for locating the patient in an evolving biological, psychological, and social matrix; a consideration of aging effects cannot await the last stages of the assessment, at which point aging is viewed solely as a factor modifying disease expression. Rather, thoughtful understanding of the aging-related context of a patient's illness is essential to obtaining the fullest view of the relevant life factors contributing to disordered behavior.

Data Acquisition The patient's history is an essential feature of neuropsychiatric evaluation because it provides the clinician the opportunity to develop the equivalent of a serial mental status examination across the patient's life course and to identify target symptoms that may respond to treatment. The clinician seeks to discern when, if ever, the patient functioned autonomously and effectively and to define the personal, social, psychological, symptomatic, and medical factors related to primary disease that contributed to a decline in function or to a failure of normal development. The history provides the opportunity to view the unfolding or evolution of signs and symptoms. The clinician strives to develop a variety of corollary information sources when assessing the patient's history so that the most complete view of the illness may be obtained. Corollary information sources may be particularly important for evaluating the history of patients who lack the cognitive capacity to relate their own life stories effectively and they are especially important for understanding the social and cultural context of specific symptoms.

During the history taking, the clinician seeks to elicit the functional anatomy of an illness. Subtle cognitive disorders, fluctuating symptom pictures, and progressing disease processes may be effectively tracked in a detailed rendition of changes in the patient's daily routine involving such factors as self-care, job responsibilities, and work habits; meal preparation; shopping and personal support; interactions with friends; hobbies and sports; reading interests; religious, social, and recreational activities; and ability to maintain personal finances. Understanding the fabric of life for each patient provides an invaluable source of data regarding many of the final common behavioral pathways cited previously, including attention and concentration, intellectual abilities, personality, and motor skills, and more typical symptomatic psychiatric features such as mood state and perception. The examiner seeks to find the particular pursuits that the patient has identified as most important or central to his or her lifestyle and attempts to discern how those pursuits have been affected by the emerging clinical condition. Such a method provides the opportunity to appraise both the impact of the illness and the patient-specific settings for monitoring the effects of future therapies.

Mental Status Examination Following a thorough history acquisition, the neuropsychiatrist's primary tool is the assessment of mental state. Formal mental status examination fell into disrepute when descriptive psychiatry was seen as irrelevant to the effective implementation of dynamically oriented psychotherapies. Its value is now undisputed. Like the physical examination, the mental status examination is a means of surveying predetermined functions and abilities to allow a definition of personal strengths and weaknesses. It is a repeatable, structured view of symptoms and signs; uniformity of approach assists in the reliable definition of findings and promotes effective communication between clinicians. It also establishes the basis for future comparison, essential for documenting therapeutic effectiveness, and it allows comparisons between different patients, with a generalization of findings from one to another. Table 10-3 lists the components of a comprehensive neuropsychiatric mental status examination.

Table 10-3 Neuropsychiatric Mental Status Examination

General Description Often, teachers and texts place the so-called sensorium as one of the last items for reporting when describing the mental status examination; the term is too broad, but consideration of arousal and responsiveness to the environment should be one of the first domains of assessment. If the patient has a significant disorder of attention or arousal, other aspects of the examination may be invalid. Together, attention and comprehension are the pillars of the mental status examination. Problems of arousal and inability to comprehend the fundamental aspects of the examination tend either to invalidate many findings or to warrant caution in their interpretation.

Language and Speech The clinician may use language function, particularly when assessing output, to estimate the patient's level of education and intelligence. It is essential, whenever possible, to estimate the patient's premorbid intellectual abilities. Definition of educational attainment during acquisition of the history aids in this process, but further appraisal during mental status assessment is valuable. However, this method must be used carefully because low educational attainment, a different language or cultural background, or acquired brain damage may confound any estimation.

Thought Assessment of thought processes involves appraising form and content. Thought form relates closely to language; for example, the clinician must distinguish between fluent aphasia or other disorders of word output and formal thought disorders related to psychosis (such as tangential responses or derailment). There are no ideational or perceptual manifestations that exclusively reflect neuropsychiatric disorders. Although some investigators have emphasized that olfactory hallucinations, for example, indicate brain disease specifically, such assertions have not been supported by well-designed epidemiological studies. Moreover, whereas the major primary psychiatric disorders have no known etiologies, there is no doubt that they involve abnormalities of brain functioning that result in the widest array of symptoms.

Mood and Affect When assessing affective and mood state, the examiner should appraise the congruence between expressed mood and demonstrated emotion. Patients with cerebral lesions occasionally demonstrate pseudobulbar affect or affective incontinence. The signs of pseudobulbar affect often include affective overshoot or disconnected affect, in which the patient responds to an appropriate stimulus but the expression is exaggerated or the emotional expression is unrelated to any defined mood. Although such behaviors can be observed in patients with idiopathic or primary psychiatric disorders, careful observation over an extended period often demonstrates that they are distinguishable from behaviors encountered in patients with mood disorders.

Insight and Judgment *Insight* denotes looking in while *judgment* reflects looking out. Both entail processes of appraisal or assessment, of one's own state of mind, one's motivations and actions, or one's relationships to others. Discussing the events leading to a clinical evaluation and comparing the patient's version with data gleaned from key informants (family, friends, other clinicians) provide an opportunity to define the congruity of the patient's understanding with that of others. Comparing examination-derived findings with the patient's insight (self-appraisal of mental state) serves as a direct or first-hand assessment.

Cognition When testing cognitive functions the clinician should evaluate memory; visuospatial and constructional abilities; and reading, writing, and mathematical abilities. Abstraction ability is also valuable to assess, although a patient's performance on tasks, such as proverb interpretation, may be difficult to evaluate when abnormal. Proverb interpretation may be a useful bedside projective test in some patients, but the specific interpretation may result from a variety of factors, such as poor education, low intelligence, and failure to understand the concept of proverbs, as well as a broad array of primary and secondary psychopathological disturbances. Although testing similarities are also education-sensitive, similarities may be more easily understood by patients.

A variety of standardized assessments have been developed in recent decades to assist with mental status evaluation. These include psychopathological rating scales that depend on self-report as well as examiner administration and brief evaluations of cognitive function that have proved helpful in examining individuals with developing cerebral diseases. Clinicians who use brief evaluations, however, must be cautious when interpreting their findings, which are subject to both false-negative and false-positive errors. For example, many tests use single cutoff points as thresholds for establishing abnormality. However, patients with focal lesions who experience discrete intellectual impairments may remain within the normal range of performance. Patients with idiopathic psychiatric disorders, such as major depressive disorder, may perform at abnormal levels on standardized cognitive protocols, inviting unwary clinicians to diagnose them as having dementia. Such assessments may also be susceptible to systematic differences among the elderly and are sensitive to lower educational level. Because they are tools for screening a large number of persons, bedside cognitive tests tend to be least helpful at the extremes, either when appraising highly intelligent individuals who are suffering intellectual declines but remain above the top rung of the test or when testing those who show substantial cognitive decline. The latter may continue to have residual intellectual abilities, some of which may prove helpful for maintenance care, but tests may prove insensitive to assessing those abilities.

PATHOLOGY AND LABORATORY EXAMINATION

Like all medical tests, psychiatric evaluations such as the mental status examination must be interpreted in the overall context of thorough clinical and laboratory assessment. Psychiatric and neuropsychiatric patients require careful physical examination, especially when there are issues involving etiologically related or comorbid medical conditions. When consulting internists and other medical specialists, the clinician must ask specific questions in order to focus the differential diagnostic process and use the consultation most effectively. In particular, most systemic medical or primary cerebral diseases that lead to psychopathological disturbances also manifest with a variety of peripheral or central abnormalities. Assignment of a patient's behavioral disturbance to a symptomatic or secondary status reflects, in part, the definition of other nonbehavioral manifestations of the primary disease.

An important element in the description of secondary psychiatric disorders is the use of laboratory assessment procedures to further define the characteristics of the systemic medical or cerebral process that is etiologically related to the psychiatric symptoms in question. This requires that psychiatrists understand the range of disorders that can lead to behavioral abnormalities. A screening laboratory evaluation is sought initially and may be followed by a variety of ancillary tests to increase the diagnostic specificity. Table 10-4 lists such procedures.

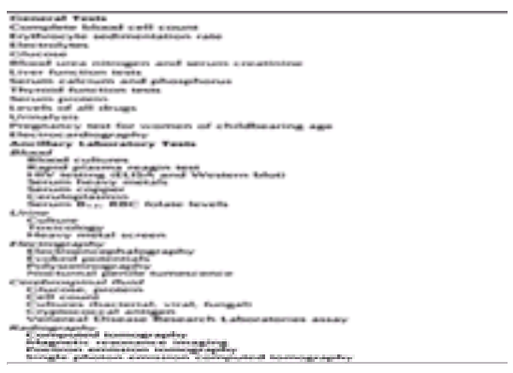

Table 10-4 Screening Laboratory Tests

A clinician requesting specific laboratory tests should be led by informed clinical suspicion as well as by an appreciation of the relative costs and benefits of each test. With the exception of low-cost screening procedures, few tests should be requested without a clearly defined rationale. Different approaches are taken for inpatients versus outpatients and for those with regular medical care versus those who have none. Repetition of recently performed tests is often without value.

Electroencephalography

EEG is an easily accessible, noninvasive test of brain dysfunction that has a high sensitivity in many disorders but relatively low specificity. Beyond its recognized uses in epilepsy, EEG's greatest utility is in detecting altered electrical rhythms associated with mild delirium, space-occupying lesions, and continuing complex partial

seizures where the patient remains conscious although behaviorally impaired. EEG is also sensitive to metabolic and toxic states, often showing a diffuse slowing of brain activity. Focal slowing, when present, may be indicative of a variety of causes such as space-occupying lesions (tumors, cerebral abscesses) or subdural hematomas. However, a superficial EEG (one that is recorded through the skull) is often insufficient for source localization and may prove insensitive to a variety of abnormal processes, necessitating nasopharyngeal recording to better define abnormalities generated by the temporal lobes or direct cortical (surface) recording to localize seizure foci. The EEG findings change with aging, with a general reduction in alpha wave activity, and with increases in the relative amounts of theta and delta wave activity. Early in the course of disorders such as Alzheimer's disease the standard EEG finding usually remains normal and therefore is often unrevealing. As part of sleep polysomnography, recent studies have suggested that the EEG may aid in the future in the distinction between elderly subjects with major depressive disorder associated with cognitive impairment and those with a primary neurodegenerative process underlying their dementia.

Computed Tomography and Magnetic Resonance Imaging

CT scanning and MRI have proved to be powerful neuropsychiatric research tools. Recent developments in MRI allow the direct measurement of structures such as the thalamus, basal ganglia, hippocampus, and amygdala, as well as temporal and apical areas of the brain and the structures of the posterior fossa. MRI has largely replaced CT as the most utilitarian and cost-effective method of imaging in neuropsychiatry. Patients with acute cerebral hemorrhages or hematomas must continue to be assessed using CT, but these patients present infrequently in psychiatric settings. MRI better discriminates the interface between gray and white matter and is useful in detecting a variety of white matter lesions in the periventricular and subcortical regions. The pathophysiological significance of such findings, designated by such terms as *rims, caps, unidentified bright objects*, and *leukoaraiosis*, remains to be defined. Such abnormalities are detected in younger patients with multiple sclerosis or HIV infection and in older patients with hypertension, vascular dementia, or dementia of the Alzheimer's type. However, their prevalence is also increased in healthy, aging individuals who have no defined disease processes. At present, those types of findings should be viewed in the same light as one would consider atrophic changes; namely, they are detected in a highly sensitive fashion but are usually nonspecific or nondiagnostic in meaning. White matter hyperintensities are more extensive and more frequent in individuals with disease, particularly those with disorders involving cognitive dysfunction, but they are too variable to contribute to the diagnosis or prognosis in an individual case. Like CT, the greatest utility of MRI when used in the evaluation of patients with dementia arises from what it may exclude (tumors, vascular disease) rather than what it can demonstrate specifically.

Because of MRI's ability to delineate brain anatomy and its sensitivity to white matter changes, these guidelines remain utilitarian when modified appropriately. Indications for ordering MRI in psychiatric patients include (1) delirium or dementia of unknown etiology; (2) a first episode of psychosis of unknown etiology; (3) a movement disorder of unknown etiology; (4) the initial evaluation of anorexia nervosa; (5) prolonged catatonia; (6) the initial onset of a major mood disorder or personality change after age 50 years; (7) the presence of unanticipated behavioral, intellectual, or functional decline in an already diagnosed psychiatric patient in whom the clinician would normally expect long-term stability or, at worst, a relapsing-remitting course with a return to baseline between episodes; and (8) the presence of any new behavioral or intellectual disorder in a patient infected with HIV.

Imaging studies have been overused in the periodic monitoring or reassessment of patients with suspected dementia of the Alzheimer's type in whom earlier examinations showed characteristic cerebral changes. Unless one suspects a missed diagnosis of normal pressure hydrocephalus, or perhaps failure to detect microinfarctions on CT when such a finding on MRI might have ruled out Alzheimer's type, repeated scans are not warranted.

Occasional patients may become agitated in the MRI tube; premedication with a benzodiazepine can minimize the problem. The magnetic field prohibits use of MRI in patients with pacemakers or metal implants, including metallic surgical clips, although many patients now receive MRI-compatible clips at surgery.

Positron Emission Tomography, Single Photon Emission Computed Tomography, and Functional Magnetic Resonance Imaging

Physiologically based techniques for imaging the brain, such as PET and SPECT, involve the injection of radioactively labeled, naturally occurring compounds or a radiopharmaceutical, with subsequent demonstration of cerebral blood flow or the incorporation of the labeled compounds into specific metabolic pathways. Such imaging methods have shown promise in studying the neurochemical and physiological bases of a variety of neuropsychiatric disorders. However, the cost of PET currently precludes its use as a routine diagnostic procedure, and there are insufficient data to project its ultimate utility for routine clinical evaluation. SPECT can be performed more readily and more cheaply, but whether it will have specific diagnostic utility in general psychiatry remains to be determined. Functional MRI (fMRI) holds great promise as a research tool to explore the physiological bases of complex behavioral processes. However, its potential utility as a clinical diagnostic tool remains to be defined.

Neuropsychological Testing

Neuropsychological testing provides a standardized, quantitative, reproducible evaluation of a patient's cognitive abilities. Such procedures may be useful for initial evaluation and periodic assessment. Tests are available that assess abilities across the broad array of cognitive domains, and many offer comparative normative groups or adjusted scores based on normative samples. The clinician seeking neuropsychological consultation should understand enough about the strengths and weaknesses of selected procedures to benefit fully from the results obtained. For example, many tests do not have appropriate aging-related norms (because they have been used primarily in young and middle-aged adults who are better educated) and therefore are less useful when used in children or the elderly. In general, clinicians should understand that a variety of distinct, competing neuropsychological schools of thought have developed different views regarding methods of individual evaluation, use of the tests, and interpretation of the data. Because neuropsychological evaluation is evolving rapidly and provides a remarkable array of tools for assaying disordered behavior, sophistication in the use and interpretation of those tests will benefit the clinician.

ETIOLOGY AND DIFFERENTIAL DIAGNOSIS

Factors Affecting Disease Presentation Neuropsychiatric evaluation and diagnosis are based on a fundamental understanding of the mechanisms by which pathobiological processes, both systemic and cerebral, express themselves through altered CNS function. The factors that influence symptom expression can be approached from several perspectives.

The first perspective relates to what might be called mode of action. Systemic disorders typically express themselves indirectly, through as yet undefined centrally active substances, defined endocrine disruptions, or fundamental metabolic alterations. Their effects tend to be generalized but often include delirium, dementia, or mood disturbance. In contrast, selective destruction of specific brain regions is more frequently associated with decrements in discrete cognitive tasks or behaviors. One must be cautious with such generalizations, however, as focal lesions in key brain regions (such as those involving brainstem structures) may cause delirious states. Moreover, the clinician may encounter substantial variability in the range of behavioral abnormalities caused by specific focal damage.

A second perspective relies on knowing the natural history of particular pathological processes. Diseases tend to progress or unfold in characteristic fashions, thus allowing for continuing differential diagnostic consideration over time. Also, meaningful prognosis depends on a thorough appreciation of natural history.

A third perspective derives from recognizing the timing of an insult within a neurodevelopmental framework, where the long-term impact of any process or event will depend in part on the compensatory or recovery capacities of the brain. Such capacities change as part of the aging process (they may be fundamental to aging), but much remains unknown.

A final perspective has to do with the types of cells and regions damaged by specific diseases. Degenerative disorders (Huntington's disease, Parkinson's disease) often lead to destruction of neurochemical systems. Hypoperfusion or pulmonary insufficiency both cause hypoxia, which in turn affects regions with especially vulnerable cell populations (such as the hippocampus) or regions near the ends of vascular trees (the so-called watershed zones, including many brain association areas). Lesions due to ischemic and hemorrhagic cerebrovascular disease reflect vascular anatomy rather than the pathoanatomy associated with degeneration of functionally significant neurochemical systems. Brain toxins may act by binding to specific neurochemical receptors, causing differential damage in direct proportion to regional variations in receptor concentration. Knowledge of the cell populations and anatomical regions affected, when integrated with the other perspectives, assists in understanding or anticipating the full effects of the primary disorder.

A variety of disorders can lead to behavioral abnormalities. They can be subsumed under the following broad categories: trauma, tumor, infection, immune and autoimmune disorders, cardiovascular disease, congenital and hereditary conditions, physiological disorders, primary psychiatric disorders, metabolic disorders, demyelinating disorders, degenerative diseases, substance-induced disorders and disorders due to toxins, and malingering.

Trauma Head trauma leading to brain injury is a possible cause for delirium, dementia, and amnestic disorder, as well as all of the secondary psychiatric syndromes.

Traumatic brain injury is largely a disease of modernity, with the majority of injuries resulting from motor vehicle accidents, gunshot wounds, or occupational mishaps. Estimates point to an annual incidence of 400 to 600 cases per 100,000 population, but such figures must be viewed with caution in light of variable definitions at the less severe (mild) end of the injury spectrum.

Pathophysiology Head trauma can cause brain injury through multiple mechanisms, both direct and indirect. Table 10-5 lists the factors that contribute to brain injury after head trauma. The clinician must recognize that brain injury from head trauma often results in pathology in areas beyond the site of direct impact. In addition, certain areas of the brain are more susceptible to injury regardless of the site of impact (Fig. 10-1). Those areas include the orbitofrontal and frontal pole convexities as well as the anterior temporal lobes, which lie close to bony skull prominences. Rotational and horizontal movements can produce shearing in areas of the brain that are relatively immobile, such as central white matter fiber pathways. Shearing forces can produce diffuse and extensive damage, also unrelated to the actual site of impact. Thus, frontal, subcortical, and limbic structures are especially vulnerable to traumatic head injury. This may explain the diversity of neuropsychiatric sequelae and the occasional occurrence of disproportionate disruptions in personality, behavior, and affect when cognitive and motor functions are largely spared. Penetrating head injuries or injuries in which the head has not been able to rotate or move may spare patients from the extensive injuries associated with indirect effects, despite significant direct damage. Bullet or penetrating missile injuries, however, may disrupt neuronal function beyond the site of impact through the effects of high-frequency vibratory waves.

Direct Effects
Contusion underlying point of trauma (coup)
Contusion linearly opposite point of trauma (contrecoup)
Compression from overlying depressed skull fracture
Compression from overlying hematoma
Indirect Effects
Diffuse impact damage
Widespread damage in cerebral white matter
Discrete lesions in corpus callosum
Discrete lesions in rostral brainstem
Common cortical contusions independent of direction of impact
Orbitofrontal
Anterior temporal
Frontopolar
Secondary cerebral processes
Cerebral edema
Increased intracranial pressure with cerebral herniation (Duret's hemorrhages)
Increased intracranial pressure with uncal herniation and posterior cerebral artery entrapment (medial occipital infarction)
Multifocal ischemic changes
Contributing secondary systemic processes
Shock (blood loss, ruptured viscera, sepsis, etc.) with hypotension
Pulmonary failure with anoxia
Long bone fracture with fat emboli

Adapted from Alexander MP: Traumatic brain injury. In *Psychiatric Aspects of Neurologic Disease*, vol 2, DF Benson, D Blumer, editors. Grune & Stratton, New York, 1982.

Table 10-5 Pathophysiological Mechanisms of Brain Injury After Head Trauma

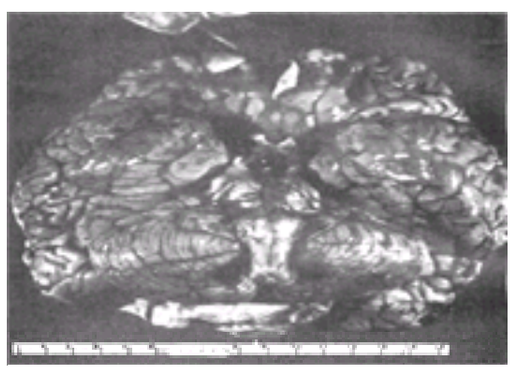

FIGURE 10-1 Severe contusion of the frontal poles has resulted in their atrophy and distortion. (Courtesy of H. M. Zimmerman, M.D.)

Psychiatric Symptoms Delirium is the acute manifestation of all head injuries that are likely to produce long-lasting sequelae. In severe head injury there is an initial loss of consciousness (coma), followed by gradual recovery, with the delirium taking the form of progressive stages of semiwakefulness, distractibility, and confusion, and finally a stable level of consciousness. The entire process may be brief or may take hours to weeks. In milder injuries there may be a brief absence of consciousness, such as momentary dazing, passing out, or transient confusion. A brief lapse or alteration of consciousness occurring after head trauma is defined as *concussion*. Table 10-6 includes frequently observed features of concussion and Table 10-7 lists common symptoms. Table 10-8 divides concussion cases into three grades of severity. Grades 2 and 3 cases require neurological evaluation; Grade 3 cases warrant immediate transport to an emergency department for assessment.

Vacant stare (befuddled facial expression)
Delayed verbal and motor responses (slow to answer questions or follow instructions)
Confusion and inability to focus attention (easily distracted and unable to follow through with normal activities)
Disorientation (walking in the wrong direction; unaware of time, date, and place)
Slurred or incoherent speech (making disjointed or incomprehensible statements)
Gross observable incoordination (stumbling, inability to walk tandem/straight line)
Emotions out of proportion to circumstances (distraught, crying for no apparent reason)
Memory deficits (exhibited by repeatedly asking a question that has already been answered, or inability to memorize and recall 3 of 3 objects in 5 minutes)
Any period of loss of consciousness (paralytic coma, unresponsiveness to arousal)

Reprinted with permission from Quality Standards Committee, American Academy of Neurology (chief authors, Kelly JP, Rosenberg JH): The management of concussion in sports (summary statement). Neurology 48:581, 1997.

Table 10-6 Frequently Observed Features of Concussion

Early (minutes and hours):
Headache
Dizziness or vertigo
Lack of awareness of surroundings
Nausea or vomiting
Late (days to weeks):
Persistent low-grade headache
Light-headedness
Poor attention and concentration
Memory dysfunction
Easy fatigability
Irritability and low frustration tolerance
Intolerance of bright lights or difficulty focusing vision
Intolerance of loud noises, sometimes ringing in the ears
Anxiety and/or depressed mood
Sleep disturbance

Reprinted with permission from Quality Standards Committee, American Academy of Neurology (chief authors, Kelly JP, Rosenberg JH): The management of concussion in sports (summary statement). Neurology 48:581, 1997.

Table 10-7 Symptoms of Concussion

Grade 1	Transient confusion No loss of consciousness Concussion symptoms or mental status abnormalities resolve in 15 minutes
Grade 2	Transient confusion No loss of consciousness Symptoms resolve in >15 minutes
Grade 3	Any loss of consciousness

Table 10-8 Gradations of Concussion

Cognitive disorders are frequent after traumatic brain injuries. Global impairment may be seen after extensive head injury or prolonged coma, although those deficits may improve dramatically in the months following injury. Dementia or a persistence of global cognitive impairment is less common, reflecting the high mortality associated with more severe injuries. When dementia is seen, it is usually associated with hemiparesis, aphasia, or other indicators of severe and extensive injury.

Persisting dementia with gradually progressive deficits may be associated with multiple recurrent head traumas. The condition has been termed chronic traumatic encephalopathy and has been noted to occur after even minor multiple head traumas. Dementia pugilistica, or boxer's dementia, is an example. Onset usually occurs at the end of a boxer's career but chronologically earlier than the onset of the degenerative dementias. A subcortical pattern of dementia (discussed later under degenerative diseases) is typically present, with prominent parkinsonian features as well as dense memory impairment. Neuropathological studies have demonstrated global atrophy with specific involvement of the midbrain and mesial temporal lobe, presumably reflecting the direct and indirect effects of multiple injuries. Plaques and tangles are often noted, but the pathophysiological mechanism remains unknown.

Memory disturbance is nearly always present with any trauma severe enough to cause a concussion. *Posttraumatic amnesia* occurs invariably after concussive brain injury and refers to the inability to register new memory. The duration of posttraumatic amnesia, which may be very brief, is a significant indicator of severity but can be assessed only after the patient has regained a stable level of consciousness. *Retrograde amnesia* is the inability to recall events prior to the injury. It can be assessed by asking patients about their last memories before the injury. Retrograde amnesia generally shrinks with recovery whereas a postinjury memory deficit tends to remain constant; patients do not recover memories from the period of posttraumatic amnesia. Additionally, patients may suffer persisting impairment of new learning and recall (an amnestic disorder) as a result of permanent pathological changes incurred because of the traumatic event. Persisting specific deficits in the context of overall robust recovery can be disabling and frustrating for the patient, who appears normal to others although still impaired cognitively and functionally. Depending on the specific nature of any deficits, these patients would be diagnosed according to DSM-IV as having cognitive disorder not otherwise specified or amnestic disorder due to traumatic brain injury.

Postconcussional disorder is a disabling cluster of symptoms of uncertain pathophysiology. It emerges within hours to days (or a few weeks) of a mild head injury and is characterized by headache, dizziness, fatigue, poor concentration and mild memory impairment, problems sleeping, irritability, anxiety, and often significant problems with mood regulation or frank clinical depression. Diminished spontaneity, apparent apathy, and other personality changes are noted also. The cluster of symptoms is remarkably consistent from patient to patient. Table 10-9 presents the proposed research diagnostic criteria for postconcussional disorder that are included in DSM-IV.

A. A history of head trauma that has caused significant cerebral concussion.
Note: The manifestations of concussion include loss of consciousness, posttraumatic amnesia, and less commonly, posttraumatic onset of seizures. The specific method of defining this criterion needs to be established by further research.
B. Evidence from neuropsychological testing or quantified cognitive assessment of difficulty in attention (concentrating, shifting focus of attention, performing simultaneous cognitive tasks) or memory (learning or recalling information).
C. Three (or more) of the following occur shortly after the trauma and last at least 3 months.
(1) becoming fatigued easily
(2) disordered sleep
(3) headache
(4) vertigo or dizziness
(5) irritability or aggression on little or no provocation
(6) anxiety, depression, or affective lability
(7) changes in personality (e.g., social or sexual inappropriateness)
(8) apathy or lack of spontaneity
D. The symptoms in criteria B and C have their onset following head trauma or else represent a substantial worsening of preexisting symptoms.
E. The disturbance causes significant impairment in social or occupational functioning and represents a significant decline from a previous level of functioning. In school-age children, the impairment may be manifested by a significant worsening in school or academic performance dating from the trauma.
F. The symptoms do not meet criteria for dementia due to head trauma and are not better accounted for by another mental disorder (e.g., amnestic disorder due to head trauma, personality change due to head trauma).

Reprinted with permission from American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, ed 4. (© American Psychiatric Association, Washington, DC, 1994.

Table 10-9 DSM-IV Research Criteria for Postconcussional Disorder

ICD-10 defines postconcussional syndrome as follows:

> The syndrome occurs following head trauma (usually sufficiently severe to result in loss of consciousness) and includes a number of disparate symptoms such as headache, dizziness (usually lacking the features of true vertigo), fatigue, irritability, difficulty in concentrating and performing mental tasks, impairment of memory, insomnia, and reduced tolerance to stress, emotional excitement, or alcohol. These symptoms may be accompanied by feelings of depression or anxiety, resulting from some loss of self-esteem and fear of permanent brain damage. Such feelings enhance the original symptoms and a vicious circle results. Some patients become hypochondriacal, embark on a search for diagnosis and cure, and may adopt a permanent sick role. The etiology of these symptoms is not always clear, and both organic and psychological factors have been proposed to account for them. The nosological status of this condition is thus somewhat uncertain. There is little doubt, however, that this syndrome is common and distressing to the patient.

At least three features of the syndrome are necessary for diagnosis, according to ICD-10 (Table 10-10). Laboratory tests may be helpful for corroboration. Some observers have argued that the requirement in DSM-IV to wait 3 months before a definitive postconcussional diagnosis may lead to a delay in establishing a proper diagnosis and initiating therapy for some patients. Indeed, postconcussional conditions transiently or spontaneously resolve for most affected individuals, with symptom remission during the first 3 to 6 months following injury. Although occasional individuals develop posttraumatic migraine, patients with postconcussional disorder more commonly describe symptoms reminiscent of muscle tension headaches arising frontally or posteriorally and occasionally involving temporal regions as well. Some report tenderness persisting at the site of impact, but that is less frequent and its pathophysiological basis is unknown. Major depression in the context of postconcussional disorder may not remit unless specific antidepressant treatment is initiated. Postconcussional symptoms that persist beyond 12 weeks should raise suspicions of additional brain pathology, such as an undetected subdural hematoma or a chronic cognitive impairment syndrome. A thorough evaluation is warranted. Postconcussional headaches can persist and prove disabling, and patients may benefit from the judicious use of analgesic as well as antidepressant agents. However, clinicians also must be vigilant regarding the development of headaches caused by analgesic rebound, a paradoxical but common cause of apparent chronic posttraumatic headaches. Secondary mood disorders are commonly seen with severe injury, although they may be more common after minor injury as part of the postconcussional syndrome. All forms of psychotic symptoms that are seen in idiopathic schizophrenia can be seen after traumatic injury. They are most common in the immediate delirious period but can persist once a stable level of consciousness has been obtained.

Table 10-10 ICD-10 Diagnostic Criteria for Personality and Behavioral Disorders Due to Brain Disease, Damage and Dysfunction

Personality Change Due to a General Medical Condition This disorder is a frequent concomitant of traumatic brain injury, owing to the vulnerability of the frontal lobes and the important role those structures play in the expression of personality. Two personality syndromes have been described with frontal lobe injury: the *orbitofrontal syndrome*, characterized by disinhibition, explosiveness, and jocularity; and the *frontopolar syndrome*, characterized by apathy, behavioral inertia, and indifference. Patients may appear indifferent to their incapacities or may confabulate regarding their injury and hospitalization. Less marked personality changes, such as irritability and a so-called short fuse, are common, especially as part of the postconcussional syndrome.

Adjustment disorders can occur at any point once a stable level of consciousness has been attained. Patient and family must adjust to loss of capacity, increased irritability and fatigue, a possible change in family roles, absence from work, financial constraints, and legal entanglements. As in all adverse circumstances, premorbid personality heavily influences the patient's adaptive capacities. Unfortunately, clinicians have at times seen the presence of an adjustment disorder or a prior history of maladaptive personality functioning as a reason to conclude that patients are not suffering from behavioral or cognitive impairments arising from brain injury. The evaluation and treatment of head trauma require clinical flexibility to address the broadest range of symptoms and syndromes.

Course and Prognosis The course of recovery from posttraumatic syndromes depends on the severity of the initial injury and the location of damage. The duration of coma and of posttraumatic amnesia may be useful prognostic indicators. Dramatic improvements can occur within days and continue for up to 6 months. Overall recovery may continue up to 24 months, with motor and physical improvement often preceding behavioral and cognitive restoration; less frequently, recovery continues beyond 2 years after injury. The neurobiological mechanisms leading to recovery are unknown.

Treatment There are no specific treatments for the cognitive abnormalities associated with head trauma. Life-sustaining and life-supportive short-term therapies may be needed initially, and the psychopathological conditions resulting from head trauma may warrant symptomatic therapies. Despite a boom in institutions and companies offering cognitive rehabilitation, it remains unproved scientifically whether those methods significantly augment natural recovery processes.

Litigation The high frequency of closed head injuries and posttraumatic complaints, together with the ready availability of psychiatrists, neurologists, and psychologists willing to testify as “experts,” have combined with lawyers in today's litigation-prone culture to bring postconcussional disorder center stage in American courtrooms. While there are numerous instances of claims related to bona fide cognitive, emotional, and behavioral deficits resulting from brain injury, it also is clear that many are unsubstantiated. There appear to be four often interacting factors that contribute to the latter situation: (1) Absence of concussion—Suits frequently request damages for postconcussion syndrome (or a related term) when a careful review of medical records reveals no evidence of pertinent symptoms or signs, such as altered consciousness, posttraumatic amnesia, nausea and vomiting, photophobia, or headache: (2) Nonspecific symptoms—Postconcussional disorder is diagnosed due to fatigue, headache, and dysphoria in the absence of symptoms or signs of concussion; complaints may have developed weeks or even months after an accident or injury. (3) Diagnostic mythology—Once a clinician labels a condition postconcussional without substantiation, this label is then promulgated and other clinicians accept it at face value without an independent review of all necessary data. The diagnosis soon assumes mythic proportions. (4) Lack of common sense—[A patient is diagnosed as having specific posttraumatic psychopathology or cognitive decline without the earlier medical, psychiatric, educational, and vocational records having been reviewed, which reveals the presence of the claimed symptoms or signs before the purported brain injury.]

In all litigation-related evaluations, the possibility of malingering needs to be considered. Relevant hints or clues include: a substantial discrepancy between mild clinical findings and severe or dramatic subjective complaints; relatively intact personal and vocational functioning and markedly abnormal neuropsychological test scores; vague complaints without objective test or functional correlates; disparity between complaints of vocational limitations and continued vigorous recreational activities (e.g., hunting, weight training, volleyball, or tennis); a history of legal difficulties or multiple accident claims; evidence of a clinically significant personality disorder.

An evaluator must also be vigilant to assess whether there is evidence of an undiagnosed mood disorder. Frequently an individual with severe physical injuries gradually develops complaints of dysphoria, headaches, poor concentration, and memory dysfunction. Diagnosed as suffering a postconcussional disorder, the plaintiff or patient believes he has permanent brain damage rather than an eminently treatable mood disorder.

Tumor Intracranial tumors, whether of primary CNS or metastatic origin, can cause behavioral disturbances by directly affecting brain function. They may do so by destroying or compressing brain parenchyma (from mass effect or edema), through obstructive hydrocephalus, or by disrupting brain vasculature (Fig. 10-2). The nature of the ensuing behavioral disturbance depends on factors already discussed, such as time course and injury location.

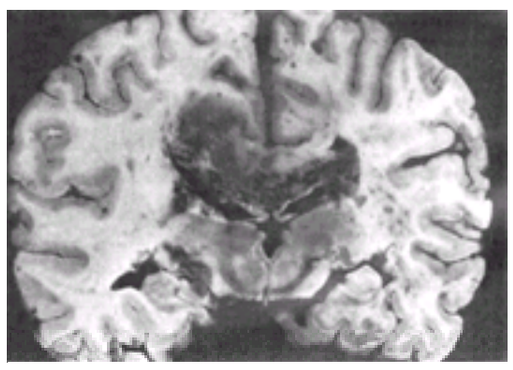

FIGURE 10-2 Glioblastoma multiforme. The massive tumor crosses the midline in the corpus callosum. (Reprinted with permission from Hirano A: *A Guide Neuropathology*. Igaku-Shoin, New York, 1981.)

Extracranial nonbrain neoplasms may indirectly alter brain function and cause psychiatric symptoms by any of several pathways. The cancer may disturb one or more organ systems known to affect brain function. For example, lung cancer may cause hypoxemia and metastatic prostate carcinoma may lead to obstructive uropathy with consequent renal failure. Paraneoplastic syndromes may lead to metabolic abnormalities (e.g., hypercalcemia) commonly associated with behavioral changes. Intriguingly, cancer may cause psychiatric symptoms without any known metabolic or other organ system disturbance; a commonly cited example is the onset of a major depressive disorder as the first clinical manifestation of occult pancreatic carcinoma. The mechanisms of such phenomena are unknown, although it has been speculated that blood-borne humoral factors secreted by the tumor are centrally active.

Infection Infections can produce any of the range of cognitive impairments or secondary syndromes that are sudden or insidious in onset. Acute infectious processes involving the CNS often produce delirium as a component of fulminant deterioration. Chronic psychopathology can result either from a chronic infectious process, such as neurosyphilis or Creutzfeldt-Jakob disease, or from persisting structural brain damage incurred as a result of an acute infection, as in the long-term sequelae

of herpes simplex encephalitis.

Syphilis Syphilis is a chronic infection resulting from inoculation with the spirochete *Treponema pallidum*. It is transmitted through sexual contact. Primary syphilis is a local disease manifested by a lesion at the site of inoculation, usually the penis, vagina, or mouth, within 2 to 3 weeks after inoculation. Secondary syphilis, manifested by a recurrent rash occurring anywhere on the body but especially on the palms and soles, has its onset 6 weeks to 6 months after initial exposure. After the rash resolves, syphilis may enter a latent stage that lasts 2 to 10 years after inoculation; serology remains positive throughout the latent stage. Tertiary syphilis may involve skin, bone, and the aorta, as well as the CNS. Neurosyphilis can occur 5 to 35 years after the initial inoculation. Neurosyphilis is divided into four stages: (1) an asymptomatic stage, without symptoms but with abnormal cerebrospinal fluid (CSF); (2) meningovascular syphilis, characterized by headache, nuchal rigidity, irritability, and delirium; (3) tabes dorsalis, with signs of posterior column degeneration, such as ataxia (due to loss of proprioception resulting in a slapping or high-stepping gait and trophic joint changes—Charcot's joints), areflexia, paraesthesias (described as lightning pains and typically involving the extremities), incontinence, impotence, and abnormal pupillary findings (the classic Argyll Robertson pupil, which accommodates but does not respond to direct light response); and (4) general paresis, also known as general paralysis of the insane, dementia paralytica, or paretic neurosyphilis, the classic neuropsychiatric disorder of tertiary syphilis (Fig. 10-3).

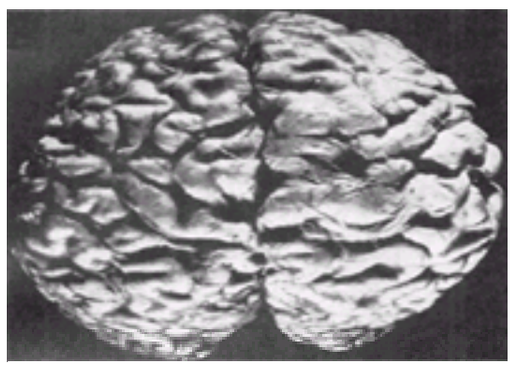

FIGURE 10-3 Paretic neurosyphilis. Thickening of the meninges and atrophy of the cerebral convolutions. (Reprinted with permission from Merritt HH, Adams RD, Solomon HC: *Neurosyphilis*. Oxford University Press, New York, 1946.)

General paresis has great significance for the history of psychiatry because it was one of the first instances in which severe behavioral and cognitive disturbances could be attributed directly to an etiologically definable brain disease. General paresis can present as almost any form of psychiatric disturbance or dementia syndrome. The classically described grandiose presentation has become rare whereas depressive presentations have become more common. Often a general change in personality is the initial presentation, with apathy, lability, and coarsening of behavior. Dementia is of a mixed pattern, with prominent impairment of memory, language, and judgment, as well as loss of initiative and psychomotor slowing. Neuropathologically the brain demonstrates diffuse degeneration with marked lymphocytic infiltration throughout.

Creutzfeldt-Jakob Disease *Creutzfeldt-Jakob disease* is an infection that causes a rapidly progressive cortical-pattern dementia. The infectious agent, a *prion*, is a subviral replicative protein that is now known to cause a variety of so-called spongiform diseases in animals and humans. The 1998 Nobel Prize for Medicine was awarded to Stanley Prusiner for his work describing this novel biological entity. The age at onset of Creutzfeldt-Jakob disease is usually in the sixth or seventh decade, although onset can occur at any age. The incidence is 1 in 1,000,000. The clinical symptoms vary with progression of the illness and depend on the regions of the brain that become involved. Patients may present initially with nonspecific symptoms, including lethargy, depression, and fatigue. Within weeks, however, more fulminant symptoms develop, including progressive cortical pattern dementia, myoclonus, and pyramidal and extrapyramidal signs. Although blood, CSF, and imaging studies are unremarkable, the EEG can demonstrate a characteristic pattern of diffuse symmetric rhythmic slow waves. A presentation with rapid deterioration, myoclonus, and the characteristic EEG pattern should raise suspicion of Creutzfeldt-Jakob disease. The definitive diagnosis is made by postmortem microscopic examination, which demonstrates spongiform neural degeneration and gliosis throughout the cortical and subcortical gray matter; white matter tracts are usually spared. Prion disease can incubate for decades before the emergence of clinical symptoms and subsequent rapid progression. Reported routes of transmission include invasive body contacts, such as direct tissue transplantation (e.g., corneal transplants) or hormonal extracts (e.g., human growth hormone, before synthetic supplies were developed). Familial patterns have also been reported, which suggests that there may be genetic susceptibility to infection or vertical transmission of the disease agent. No antiviral agents have been shown to be effective in retarding or slowing disease progress, although amantadine (Symmetrel) has been reported occasionally to have had some success. Death usually ensues within 6 months to 2 years of onset. During the past several years, a pathologically similar condition, bovine spongiform encephalopathy, has been described. Diagnosed primarily in the United Kingdom, this disease underscores the effects of modern animal husbandry methods on the amplification of rare diseases and the continuing threat of zenobiotic transmission of these to humans.

Viral Encephalitis Viral encephalitis varies in severity, depending on the specific etiological agents. Mild disease is more common with mumps, and enteroviral infections can be limited to headache and malaise. Severe disease is characteristic of infections such as rabies and herpes simplex. Herpes simplex encephalitis is the most common of the severe nonepidemic encephalitides. It is of interest to neuropsychiatry because of the preferential involvement of the orbitofrontal and medial-temporal regions of the brain. A typical presentation consists of severe encephalitis of rapid onset, high fever, headache, nuchal rigidity, focal neurological signs, and delirium. Rarely, a sudden, transient psychosis may herald the onset. Occasionally the onset is more insidious, with the clinical picture at presentation limited to personality change or memory impairment. Necrosis of the frontal and temporal lobes can occur rapidly. Mortality is high: approximately 70 percent. Whenever herpetic encephalitis is suspected, a definitive diagnosis should be made as rapidly as possible by brain biopsy, with the subsequent urgent initiation of antiviral therapy. Survivors may sustain deficits related to temporal and frontal lobe damage, including a dense amnesia disproportionate to the degree of other intellectual impairment; hallucinations in all spheres, including olfactory and gustatory; components of a Kluver-Bucy syndrome; partial complex seizures; aphasia; and anosmia.

Human Immunodeficiency Virus HIV-1 has created a late–twentieth-century epidemic parallel in severity and pervasiveness to the scourges of bygone eras. Acquired immune deficiency syndrome (AIDS), the later stages of HIV infection, has been recognized since the 1980s. In recent years its neuropsychiatric manifestations have become a focal point for diagnosis and therapy, as patients live longer through the use of partially effective antiviral therapies and a variety of second-line medications employed for treating opportunistic infections. The following discussion focuses on neuropsychiatric phenomena that appear to result from HIV-1 action in the brain.

Mr. Zeigler is a 55-year-old, married, Latin American businessman who is hospitalized with an 8-month history of diarrhea, fatigue, and weight loss. He has sought help from several institutions both in the United States and Europe, but his illness remains undiagnosed. A psychiatric consultation is requested because both the patient and his physician think he is depressed and wonder what role this might play in his weight loss and overall condition.

Mr. Zeigler gives a detailed history of his family's emigration from Europe when he was a child, his personal success in business, and the progressive difficulty he has been experiencing because of his weight loss and fatigue. He has lost 85 pounds over the 8 months and now has to force himself to eat. In the past, eating had been a great pleasure for him and he considered himself a gourmet cook. Although he complains of some difficulty with his memory and concentration, he continues to manage a multinational business and to conduct complex financial deals. He says he feels sad but is hopeful that the diagnosis can be made quickly. He conducts himself in the same autocratic manner in the hospital that he is accustomed to displaying in business and with his family. He has many interests, including an active sex life, which he wishes to resume once he regains his strength.

Mrs. Zeigler confirms her husband's history and speaks of his complete control of his business and of the family's financial affairs. She describes how this has created conflicts with her sons, who resent their father's unyielding control, even though they work in the family company. It is her opinion that her husband is depressed and that this is the cause of most of his symptoms. In response to questions about his activities, she agrees that his fatigue seems the only obstacle to pursuing his interests. She cannot answer any questions about his sex drive because she stopped having sex with him 10 years before this illness. He accepted this, and she presumed he frequented prostitutes.

Over the next few days Mr. Zeigler's condition deteriorates markedly, and he is thought to have had a stroke because of some slurred speech and a slight weakness of the right side of his body. He then becomes short of breath and is admitted to the intensive care unit. A chest X-ray suggests *Pneumocystis carinii* pneumonia, which is confirmed by bronchoscopy. He does not respond to co-trimoxazole (Bactrim) and is started on pentamidine (NebuPent). While in the intensive care unit Mr. Zeigler is delirious, frequently hallucinating, and often incoherent in both Spanish and English. His children fly to the United States because he is not expected to survive. This prediction proves to be incorrect, and his pneumonia resolves after several weeks of treatment. A CT scan of the brain suggests a CNS infection with toxoplasmosis, and examination of the stomach by endoscopy leads to the diagnosis of gastrointestinal *Isospora*. Surprisingly, all these infections respond to treatment.

It is now clear that Mr. Zeigler has AIDS, and his physician presents this diagnosis to him along with an inquiry about his sexual experiences. Mr. Zeigler is enraged by his doctor's implication of homosexuality and adamantly denies any homosexual activity. He discusses his adaptation to his wife's decision to cease sexual activity with him. He has frequented prostitutes in the Far East, where he traveled regularly on business. It seems impossible to him that he might have AIDS, although he admits to having contracted syphilis 4 years before his current illness. On discharge, Mr. Zeigler is given the diagnosis of AIDS, with the only clear risk factor his sexual contact with prostitutes.

Six months later Mr. Zeigler and his wife return to the United States for further evaluation of his mental status. His wife is concerned that he has become depressed because he is no longer able to handle his financial affairs. She feels his personality has undergone a radical change in that he no longer seems to care about anything, in spite of the fact that his appetite has returned to normal and he has regained much of his lost weight. Much of his time is now spent sitting idly in their garden.

When examined, Mr. Zeigler appears to be in good physical health. However, his mental condition has obviously deteriorated; it is not possible to conduct an interview in English although previously he had spoken several languages fluently. He smiles pleasantly, but is both disoriented and confused, even when speaking in Spanish. This surprises even his wife, since she had not been aware of this change in his cognitive functioning. He has poor short-term memory and cannot perform simple calculations. His remote memory is intact, although his wife feels that he has confused some historical events. Mr. Zeigler seems unaware that there are any deficits in his intellectual functioning. Medical evaluation does not reveal any active infections. (Reprinted with permission from *DSM-IV Casebook*.)

Clinicians began to recognize the variety of neuropsychiatric manifestations of HIV in the mid-1980s. Most prominent were major mood disturbances (major depressive, dysthymic, and less commonly, bipolar disorders) and a characteristic progressive cognitive impairment that was labeled AIDS dementia complex. Rarely, patients developed psychoses, at times with a schizophrenic presentation, as well as alterations of personality. Later, carefully conducted epidemiological studies revealed that persons at highest risk for HIV infection suffered elevated rates of mood and substance use disorders before contracting the disease. Intertwined with syndromes that were thought to be direct results of primary HIV infection of the CNS or secondary complications from other infections or tumors were a variety of adjustment and mood disturbances, reflecting responses to a progressive, inevitably terminal disease.

The psychopathological manifestations of HIV cover the major symptom clusters, as well as AIDS dementia complex and delirium. Clinicians have used empirical treatments, many with substantial symptomatic response. Intervention with antiviral therapies also has shown beneficial behavioral effects, especially when CSF indices of CNS infective activity have suggested that the primary disease has increased in its activity. However, there have been few carefully conducted therapeutic trials to establish the overall efficacy of any symptomatic psychopharmacotherapy.

AIDS dementia complex is characterized predominantly by a subcortical presentation, with prominent psychomotor slowing and difficulties with concentration and memory. Early associated motor deficits include ataxia, leg weakness, tremor, and loss of fine motor coordination. Patients commonly become apathetic or withdrawn. The course is steadily progressive, at times punctuated by abrupt acceleration. Like other dementing disorders, AIDS dementia complex progresses to a late stage characterized by severe dementia, mutism, incontinence, paraplegia, and in some cases, myoclonus.

During the latter part of the 1980s, controversy developed regarding the temporal sequence of the emergence of cognitive abnormalities versus other symptoms reflecting the advance of HIV infection to full-blown AIDS. There is no dispute that AIDS dementia complex may be the predominant feature of AIDS for some patients, but there is uncertainty regarding the presence of cognitive abnormalities in patients who are both clinically asymptomatic and without laboratory evidence of encroaching immune suppression. Many patients with HIV develop a mild (minor) cognitive disorder that has many of the same features of AIDS dementia complex but that is as not cognitively severe or as impairing functionally. Recently, an American Academy of Neurology AIDS Task Force developed a set of standard nomenclature for neurological manifestations of HIV-1 infection, including both cognitive and peripheral neurological findings.

In autopsy series, 75 to 90 percent of brains of patients dying from HIV infection show neuropathological alterations. In addition to changes due to secondary or opportunistic infections, there is widespread subcortical white matter pathology with relative sparing of cortical structures. Those diffuse, noninflammatory changes are now subsumed under *HIV leukoencephalopathy*. Microscopic examination also reveals foamy macrophages and multinucleated giant cells invading both white matter and subcortical nuclei, particularly basal ganglia structures. Such focal inflammatory findings are characterized as HIV encephalitis. Also, there may be pathology in the spinal cord associated with paraparesis, particularly vacuolar myelopathy. Insofar as many other patients have significant cognitive deficits in the context of relatively little pathological alterations, it is clear why investigators have encountered difficulty when attempting strict clinical-pathological correlation.

The mechanism by which HIV causes its functional effects remains unknown. Current data point to (1) neurons and supporting cellular structures, through the actions of the virus itself or from its coat proteins; (2) the undesirable effects of activated immune components (such as activated macrophages); and (3) possible excitatory neurotoxic effects of endogenous neurotransmitters that have been dumped into surrounding interstitial fluids (e.g., quinolinic acid affecting glutamate receptor subtypes, leading to the toxic accumulation of intracellular calcium).

The diagnosis of HIV-related neuropsychiatric syndromes requires a high index of suspicion and a sensitivity to possible demographic risk factors, including homosexual behavior, sexual promiscuity, intravenous substance abuse, and sexual relations with high-risk partners. In addition, there is a gradual movement of the HIV virus into the broader heterosexual population. Psychopathological changes may precede frankly defined cognitive abnormalities. The clinician also must be alert to early, subtle intellectual decline: the patient may remain within the normal range on standard neuropsychological tests but may perform at a level lower than was attained previously. In addition to neuropsychological assessment, neuroimaging may demonstrate abnormalities in subcortical periventricular and deep white matter.

AIDS dementia complex has become a major target in pharmacotherapeutic trials to cure or ameliorate the effects of HIV infection. Preventing its emergence or prolonging the time it takes to appear have become possible end points for some studies. Others are considering AIDS dementia complex as a direct target for intervention. Future antiviral pharmacotherapies may be targeted specifically to the brain to eradicate any possible reservoirs of HIV, in a fashion similar to the use of irradiation or antitumor agents in children with leukemia.

Immune and Autoimmune Disorders Three broadly defined pathophysiological mechanisms involving the immune system can be associated with neuropsychiatric

disorders: (1) hypofunction of the immune system may contribute to infectious diseases and possibly to neoplastic illnesses; (2) definite or putative autoimmune diseases may cause behavioral disturbances by affecting the function of organ systems in a way that compromises brain activity (e.g., the hyperthyroidism of Graves' disease or hepatic failure from primary biliary cirrhosis); (3) autoimmune illnesses may also affect brain function more directly by causing cerebral ischemia due to vasculitis or by direct CNS parenchymal inflammation. Some primary neurological diseases (e.g., multiple sclerosis) may involve autoimmune pathophysiology (Fig. 10-4). A major example of an autoimmune disorder involving the CNS is systemic lupus erythematosus.

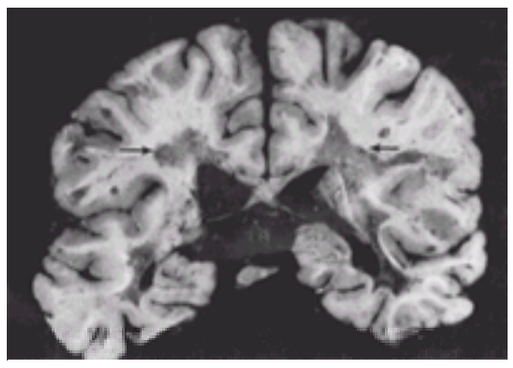

FIGURE 10-4 Multiple sclerosis. Coronal section of cerebral hemispheres showing large, sharply demarcated plaques adjacent to the bodies of the lateral ventricles (arrows). Other plaques are found adjacent to the temporal horns and smaller plaques are present in the subcortical white matter and centra semiovales. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

Cathy Jarvis, a 25-year-old mother with a 3-year history of systemic lupus erythematosus was admitted to a university hospital in an acute confusional state with inability to maintain attention or to carry on a coherent conversation and with marked disorientation to time and place. Before her hospitalization she had become progressively more confused over a number of days, and had started to believe that the neighbors were watching her. On the day of admission she had run out of her house and into the street in a state of uncontrollable agitation.

On admission to the hospital emergency room, Cathy was given intramuscular haloperidol (Haldol), but by the next morning her clinical picture had worsened dramatically. She was now rigid, mute, uncommunicative, and unresponsive to all questions and she exhibited facial grimacing. Her course fluctuated so that at times she became excited, screamed continuously, and seemed to be responding to auditory and visual hallucinations; at other times she was mute and rigid. She required total nursing care, with intravenous feeding, catheterization, and four-point restraint. She received frequent sedation with lorazepam (Ativan), a short-acting benzodiazepine. Because she was thought to have lupus cerebritis, intravenous methylprednisolone (Depo-Medrol), a steroid, was begun, but there was no improvement in her clinical condition.

During the next 3 weeks Cathy's condition deteriorated. She lost considerable weight, was unable to stand, and continued to require total nursing care. On day 28 she was referred for electroconvulsive therapy (ECT). After seven treatments she gradually responded, sought to feed herself and to stand, was more alert, and recognized her family. Rigidity was now only occasionally present. A lumbar puncture demonstrated the presence of immunoglobulin-G (IgG) antineuronal antibodies in high titer, consistent with a diagnosis of CNS involvement with lupus.

Over the next few weeks periods of lucidity alternated with rigidity, mutism, negativism, and staring. By day 90 of Cathy's hospitalization, a second course of ECT was begun. After 10 treatments, she was verbal, euthymic, and cooperative. (Reprinted with permission from *DSM-IV Casebook*.)

Often lupus is considered in the differential diagnosis of new-onset psychopathological syndromes. Although the etiology of lupus is not known, evidence implicates immunological mechanisms in its pathogenesis. Numerous organ systems may be involved. The disease may affect brain function (thereby producing psychiatric symptoms) indirectly, through such mechanisms as fever, renal failure, or pulmonary disease. In a minority of patients it may cause pathology directly, most likely from vasculitis affecting cerebral vessels. Patients with CNS disease may experience seizures, transverse myelopathies, or behavioral abnormalities, including delirium, psychotic syndromes, and affective lability. Clinicians evaluating patients with psychiatric symptoms of recent onset (particularly women in the second through fifth decades of life) should carefully consider the medical history, review of systems, physical examination findings, and routine laboratory screens, to look for evidence of systemic organ system involvement. The erythrocyte sedimentation rate (ESR), while nonspecific, is substantially elevated during acute CNS lupus and provides a useful screen. More specific laboratory tests (e.g., antinuclear antibody assay and antibodies to double-stranded deoxyribonucleic acid [DNA]) may be pursued when indicated. Neuroimaging scans may show cerebral infarctions but are often normal early in the disease. Glucocorticoids are the mainstay of treatment for acute CNS systemic lupus erythematosus. Psychotropic medications may be needed to treat specific behavioral symptoms (e.g., antipsychotic agents for severe agitation during delirium).

Cardiovascular Disease Because the extremely high metabolic activity of the brain is obligatorily aerobic, the brain is exquisitely sensitive to relatively minor perturbations in blood flow. Thus, alterations in cardiac function ranging from grossly obvious (cardiogenic shock) to relatively subtle (compensated congestive heart failure, chronic low-output states) often manifest with CNS dysfunction. The consequent psychiatric phenomenology may vary due to largely unknown factors, but delirium, dementia, and depressive episodes are especially common. Perfusion failure, depending on its cause, may lead to insidious, gradual changes or dramatic decrements in function. Transient profound drops in blood pressure, typically associated with major cardiac events (including surgery), may lead to mental status alterations that are difficult to pinpoint initially. Clinicians are faced with distinguishing soon-to-remit symptoms, such as postoperative delirium arising from metabolic imbalances, from subtle persisting intellectual and behavioral alterations caused by hypofusion leading to cell death.

Intrinsic cardiac illness, such as mural thrombus or valvular disease, may also be a source of embolic cerebral infarction (Fig. 10-5). Cardiovascular disease may also lead to brain dysfunction by serving as a risk factor for cerebrovascular disease. Identified risk factors for stroke include hypertension, diabetes mellitus, cigarette smoking, atrial fibrillation, left ventricular hypertrophy, and coronary artery disease.

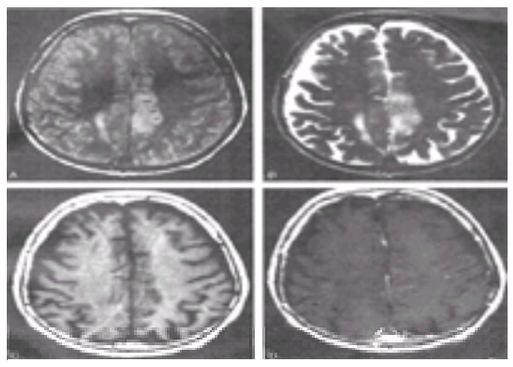

FIGURE 10-5 Acute cortical infarction. **A** and **B**, Proton-density and T2-weighted axial MRI scans show increased signal intensity within the medial cortex of left frontal and parietal lobes. Note swelling of gray matter and prominence of blood vessels within this lesion. **C** and **D** T1-weighted axial MRI scans before and after gadolinium administration demonstrate several linear foci of contrast enhancement within the area of infarction in the left frontal and parietal lobes, most consistent with enhancing arterial branches. Contrast enhancement of arterial branches is consistent with static blood flow within the infarct, and generally is seen only within the first few hours to 5 days after the onset of acute infarction. (Reprinted with permission from Rowland LP, editor: *Merritt's Textbook of Neurology*, ed 9. Williams & Wilkins, Baltimore, 1995.)

Cerebrovascular disease of any cause and pathophysiology—thrombotic, embolic, or hemorrhagic—will affect brain function. Psychiatric symptoms may develop

either suddenly (presumably in relation to abrupt neuronal losses and dysregulation) or gradually (perhaps in relation to cumulative infarcted brain tissue and to longer-term adaptations of neurochemical systems). Table 10-11 presents an overview of the variety of cerebrovascular events.

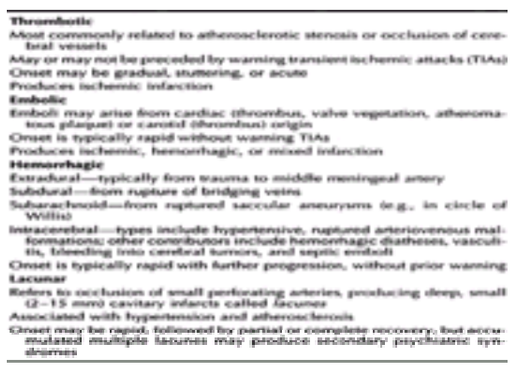

Thrombotic
Most commonly related to atherosclerotic stenosis or occlusion of cerebral vessels
May or may not be preceded by warning transient ischemic attacks (TIAs)
Onset may be gradual, stuttering, or acute
Produces ischemic infarction
Embolic
Emboli may arise from cardiac (thrombus, valve vegetation, atheromatous plaque) or carotid (thrombus) origin
Onset is typically rapid without warning TIAs
Produces ischemic, hemorrhagic, or mixed infarction
Hemorrhagic
Extradural—typically from trauma to middle meningeal artery
Subdural—from rupture of bridging veins
Subarachnoid—from ruptured saccular aneurysms (e.g., in circle of Willis)
Intracerebral—types include hypertensive, ruptured arteriovenous malformations; other contributors include hemorrhagic diatheses, vasculitis, bleeding into cerebral tumors, and septic emboli
Onset is typically rapid with further progression, without prior warning
Lacunar
Refers to occlusion of small perforating arteries, producing deep, small (2–15 mm) cavitary infarcts called *lacunes*
Associated with hypertension and atherosclerosis
Onset may be rapid, followed by partial or complete recovery, but accumulated multiple lacunes may produce secondary psychiatric syndromes

Table 10-11 Types of Cerebrovascular Events

Recent data have undercut the long-held notion that vascular disease due to a vascular etiology always progresses in stair-step fashion, as has been described for the course of vascular dementia. Rather, progression appears to vary in rate and form, depending both on the type of vasculature affected (large versus small vessels) and on basic pathobiology (perfusion insufficiency versus occlusive disease).

It is important to recognize that psychiatric phenomena may manifest without other clinical evidence of a neurological event. For example, patients with right parietal infarction have presented with delirium but no other neurological symptoms or signs. Alternatively, an infarction in the distribution of the left middle cerebral artery may cause Wernicke's aphasia, characterized by fluent paraphasic or jargon-filled speech, poor comprehension and repetition, and suspicious or aggressive behavioral responses. No other signs of cerebral disease may be evident, and such patients have been misdiagnosed as having paranoid schizophrenia. It is also important to understand that pathology does not specify phenomenology. For example, patients with multiple infarcts on MRI are often labeled as having vascular dementia; however, multiple infarctions may manifest with dementia, with other behavioral syndromes (mood disorder, psychosis, anxiety, personality change), or with no definable neuropsychiatric syndrome.

Congenital and Hereditary Conditions Patients with congenital and hereditary conditions often present for psychiatric evaluation because of the frequency of associated behavioral disturbances.

> A 15-year-old boy was brought to the emergency room by his mother, who, clutching the on-call resident's arm, pleaded, “You've got to admit him; I just can't take it anymore.” The patient had been brought home from a special school by his mother 6 months previously. The mother showed the resident papers from the school that indicated that the patient's IQ was 45. He had had several placements, beginning at age 8. On visiting days, the boy always pleaded with his mother, “Mommy, take me home.” After a year or so away, the patient would be brought home by his mother, who had always been wracked by guilt because of his retardation and her inability to manage him in the home. The patient was an only child whose parents had been divorced for the past 4 years. The father had moved to another city.
>
> During the last 6 months at home, the patient had increasingly become a behavior problem. He was about 5'9" tall and weighed almost 200 pounds. He had become destructive of property at home—breaking dishes and a chair during angry tantrums—and, more recently, physically aggressive. He had hit his mother on the arm and shoulder during a recent scuffle that began when she tried to get him to stop banging a broom on the apartment floor. The mother showed her bruises to the resident and threatened to call the mayor's office if the hospital refused to admit her son.
>
> On examination, the boy was observed to have the typical signs of Down syndrome, including thick facial features, slightly protruding tongue, epicanthic fold of the eyelids, and simian crease of the palms of the hands. With indistinct and slurred speech, the boy insisted that he “didn't mean to hurt anybody.” (Reprinted with permission from *DSM-IV Casebook*.)

Congenital and hereditary conditions are of great significance in the understanding of brain-behavior relationships because they are often seen in biologically homogeneous populations with relatively specific behavioral and neuropsychiatric syndromes. Congenital conditions are caused either by genetic abnormalities affecting autosomes, sex chromosomes, or single genes or by fetal insults during the prenatal, perinatal, or immediate postnatal periods. Table 10-12 lists several developmental and hereditary disorders with significant neuropsychiatric manifestations. In DSM-IV these conditions are classified according to the age at which symptoms manifest, what symptoms are present, and whether the symptoms are progressive or static.

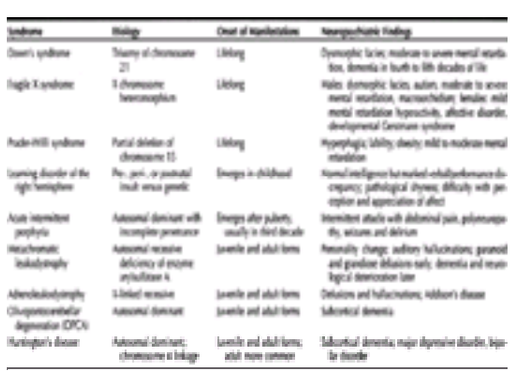

Syndrome	Etiology	Onset of Manifestations	Neuropsychiatric Findings
Down's syndrome	Trisomy of chromosome 21	Lifelong	Dysmorphic facies; moderate to severe mental retardation; dementia in fourth to fifth decades of life
Fragile X syndrome	X chromosome heteromorphism	Lifelong	Males: dysmorphic facies, autism, moderate to severe mental retardation, macroorchidism; females: mild mental retardation hyperactivity, affective disorder, developmental Gerstmann syndrome
Prader-Willi syndrome	Partial deletion of chromosome 15	Lifelong	Hyperphagia; lability; obesity; mild to moderate mental retardation
Learning disorder of the right hemisphere	Pre-, peri-, or postnatal insult versus genetic	Emerges in childhood	Normal intelligence but marked verbal/performance discrepancy; pathological shyness; difficulty with perception and appreciation of affect
Acute intermittent porphyria	Autosomal dominant with incomplete penetrance	Emerges after puberty, usually in third decade	Intermittent attacks with abdominal pain, polyneuropathy, seizures, and delirium
Metachromatic leukodystrophy	Autosomal recessive deficiency of enzyme arylsulfatase A	Juvenile and adult forms	Personality change; auditory hallucinations; paranoid and grandiose delusions early; dementia and neurological deterioration later
Adrenoleukodystrophy	X-linked recessive	Juvenile and adult forms	Delusions and hallucinations; Addison's disease
Olivopontocerebellar degeneration (OPCA)	Autosomal dominant	Juvenile and adult forms	Subcortical dementia
Huntington's disease	Autosomal dominant; chromosome 4 linkage	Juvenile and adult forms; adult more common	Subcortical dementia; major depressive disorder, bipolar disorder

Table 10-12 Examples of Developmental and Hereditary Disorders With Neuropsychiatric Manifestations

Down Syndrome Down syndrome results in static phenotypic abnormalities, including characteristic facies and mental retardation, and has been associated with a progressive decline in functioning beginning in the third or fourth decade of life. Alzheimerlike pathological changes are frequently detected at autopsy, even in patients who did not exhibit functional decline before death.

Fragile X Syndrome Fragile X syndrome is the second most common cause of mental retardation in men and one of the few known causes of the autism syndrome. Female heterozygotes also manifest significant psychiatric pathology, including mood disorders, difficulties with behavioral control, and a neuropsychological profile of dyscalculia, right-left disorientation, and constructional dyspraxia similar to the Gerstmann syndrome, described in patients with acquired lesions in the dominant parietal lobe. Learning disorders usually become obvious when the child begins school.

Huntington's Disease *Huntington's disease*, also known as *Huntington's chorea*, has been the focus of intensive neuropsychiatric, genetic, and pharmacological research since the 1970s. First described by George Huntington on Long Island in 1872, the disorder has received intense scrutiny, and its site on chromosome 4 has been determined. The disorder is related to an unstable trinucleotide repeat (CAG), associated with more than 36 copies of the specific sequence. Onset typically occurs in middle life, usually between ages 25 and 50 years. The juvenile form, with onset occurring during adolescence, is somewhat different phenomenologically, with a greater degree of dystonia early in the disease process and a faster rate of disease progression. A greater number of CAG trinucleotide repeats is associated with an earlier age of onset of clinical disease. Lower cognitive performance is detectable in those clinically asymptomatic at-risk individuals who have greater numbers of repeats. A greater number of repeats is associated with more severe neuronal loss in both caudate and putamen at postmortem examination, and perhaps a faster rate of pathological deterioration. Although the exact neurobiological role of the abnormal protein (*huntingtin*) encoded in Huntington's disease remains

unknown, other findings point to defects in mitochondrial energetics as a possible contributing mechanism.

Huntington's disease is not diagnosed formally until the typical movements appear, although both psychiatric and neuropsychological manifestations may precede the emergence of motor abnormalities. The psychopathology associated with Huntington's disease has a wide range of manifestations, commonly including affective presentations (typically depression, but mania as well); psychoses, often with a schizophrenic appearance; personality changes, and anxiety disorders. Some individuals, however, may proceed through the entire course of the illness with no evident psychopathology. Often the psychopathology is most florid during the early and middle stages of the disease, but as the characteristic subcortical dementia proceeds patients begin to exhibit less characteristic behavior and thus appear less symptomatic. The suicide rate is higher in patients with Huntington's disease than in the general population but suicidal ideation may be difficult to detect because patients tend to be less spontaneous and forthcoming as a result of the cognitive difficulties associated with the disorder. Interviewers must take an active or probing approach; patients who quickly pass off inquiries when presented with open-ended questions may provide more information when queried with specifically structured interview methods.

The cognitive disorder of Huntington's disease is more consistent in its presentation than the psychopathological picture, although it too evolves over the course of the disorder. Patients usually experience mild memory difficulties, and the first symptoms may be subtle problems with organizing, planning, and sequencing. Spontaneity and verbal elaboration may be diminished relatively early, although that appears to be somewhat more variable in its time of onset. As the disease progresses, psychomotor slowing progresses relatively rapidly, with concomitant difficulty with complex tasks, while recall of old knowledge and factual information remains less affected. Unlike patients with dementia of the Alzheimer's type, many patients with Huntington's disease remain insightful long into the course of their disease. Thus, their mood disturbances and potential suicidality may be tied in part to a clear realization of their situation. Indeed, even as patients respond to standard antidepressant therapy with enhanced sleep, energy, appetite, and improved overall mood state, they may remain realistically pessimistic about their long-term situation.

Patients with Huntington's disease begin to develop an apathetic appearance as the disease progresses. Early in the course, they continue to show interest and responsiveness when presented with structured situations in which they can take part; frank apathy and disinterest develop later and persist even in the context of prompted or structured assistance. Although some degree of verbal learning impairment is an early feature of the dementia, it is more prominent later in the disease course. Similarly, subtle visuospatial processing problems may occur early but do not become prominent clinically until later.

Just as the cognitive disturbance of Huntington's disease evolves slowly, so too there is a gradual change in the associated movement disorder. In most affected adults the movement disorder is typically choreiform at the outset but becomes more dystonic and bradykinetic as the disease progresses. Toward the end of the disease course, patients are bedridden, mute, and overcome by a severe dystonic state.

The pathology and neurobiology of Huntington's disease have been studied intensively in recent years (Fig. 10-6). The striatum bears the brunt of the pathology, with interruption of crucial corticostriatothalamocortical relays. Although there are no immediate, reciprocal corticostriatal connections, that multineuronal pathway similarly modulates motor function, cognition, and perhaps mood. Recent theories suggest abnormal function of excitatory neurotransmitters, most apparently acting on glutamate receptors, that serve as endogenous neurotoxins. Although efforts have been made to use symptomatic pharmacological treatments for both the psychiatric and motor symptoms, more recent pharmacotherapeutic trials have aimed at preventing progression of the disease by employing potential glutamate receptor blockers. Such efforts have provided models for similar therapeutic approaches to Parkinson's disease and Alzheimer's disease.

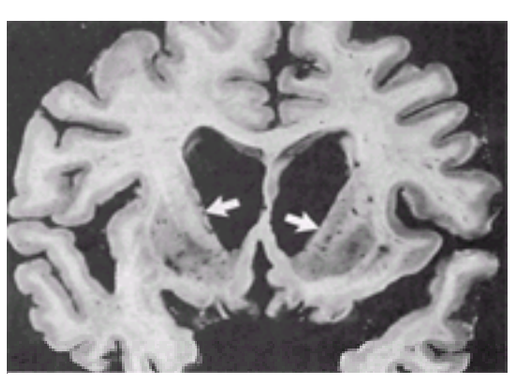

FIGURE 10-6 Huntington's chorea. There is marked atrophy of the caudate nuclei (arrows) and mild dilatation of the lateral ventricles. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

The mood disorders associated with Huntington's disease have proved amenable to symptomatic treatments. Standard doses of antidepressant medications may be needed, although patients often respond sensitively to rapid changes in medication and experience unwanted adverse effects. ECT has been beneficial for severe major depressive symptoms, especially in high-risk suicidal patients. The schizophrenia-like presentations of Huntington's disease appear less responsive to antipsychotic therapy than phenomenologically similar idiopathic disturbances. Patients with Huntington's–disease–related anxiety disorders have shown sufficient benefit from available medication regimens to warrant empirical trials. Psychotherapy, usually with the patient and family treated together, may lead to substantial therapeutic gains. Clinician commitment to the long haul may prove especially reassuring and stabilizing.

Like other hereditary neuropsychiatric disorders, Huntington's disease illustrates the need for all psychiatric evaluations to include a careful documentation of family history. Patients with Huntington's disease may present with mood or psychotic disturbances and no apparent abnormal involuntary movements and may be treated symptomatically with pharmacotherapeutic agents, only to evince the characteristic motor disorder later. Ignorance of the family history has led some to misinterpret that progression as evidence of tardive dyskinesia. The patient may remain incorrectly diagnosed until cognitive impairment becomes unmistakable. In the meantime, patients and families have lost the opportunity to clarify their life plans and develop support for future needs. Psychiatrists must remain vigilant in taking the family history whenever evaluating a new patient.

Other Conditions Learning disorders involving left hemisphere functions such as reading, writing, or mathematics are well known clinically. A learning disorder of the right hemisphere has been described that is characterized by intact linguistic and academic skills; left-sided soft (nonlocalizing) neurological signs; and profound impairments in functions dependent on the right hemisphere, including visuospatial skills, modulation of affect, and the paralinguistic aspects of communication. The etiology of the learning disorder is unknown, although a retrospective history of prenatal or perinatal insults is common, as is a family history of similar impairments. Acute intermittent porphyria is a hereditary disorder that is intermittent. In between episodic attacks, most patients maintain normal development. The leukodystrophies and degenerative hereditary disorders listed in Table 10-12 can produce symptoms during childhood or not until adulthood. Development until the appearance of symptoms is normal. However, in each of these disorders psychiatric symptomatology can precede other evidence of the disease process and lead to an erroneous diagnosis of an idiopathic psychiatric disorder.

Physiological Disorders—Epilepsy Epilepsy is the prototype of a physiological disease process that manifests psychiatric symptoms. It has long held the interest of neuropsychiatry and has been studied intensively, if not always fruitfully. The complexities of defining brain-behavior relationships in epilepsy merit extended discussion.

Definition Epilepsy is defined as a condition of recurrent seizures due to CNS disease or dysfunction. Seizures are behavioral alterations of abrupt onset and termination that are associated with sudden electrical discharges of the brain. Although the essential paroxysmal form remains constant and in fact defines a seizure, the content of the behavioral disturbance can vary widely. Seizures can be generally classified into two broad categories, generalized and focal. In generalized seizures the electrical abnormality usually originates from subcortical structures (primarily the brainstem) and then spreads simultaneously to all areas of the cortex. Loss of consciousness is invariable, and the seizure phenomenology is symmetrical and bilateral. Focal seizures originate from a specific brain locality, usually the temporal lobe. The abnormal electrical discharge may remain at the site of origin, proceed gradually to adjacent areas, or spread to include the entire cortex (secondary generalization). The clinical phenomenology of a focal seizure depends on the site from which that seizure originates and may be unilateral and restricted to a particular muscle group, sensation, affect, and so on. The epilepsies are classified based on the type of seizure and the inferred anatomical substrate (Table 10-13). Seizure type and phenomenology are usually constant within the course of a particular patient's disorder. The stereotyped presentation is a major feature of

evaluation, diagnosis, and assessment of treatment efficacy.

Primary generalized epilepsy
Tonic-clonic (grand mal)
Absence (petit mal)
Myoclonic
Other
Partial (focal) epilepsy
Simple (elementary) symptomatology
Focal motor
Focal sensory
Vegetative
Mixed
Complex symptomatology
Partial complex (psychomotor)
Secondary generalized
Unclassifiable

Table 10-13 Classification of the Epilepsies

Clinical Features Seizures can proceed in stages and may include a prodrome, aura, ictus, and a postictal period. Psychopathology may manifest during any of these stages as well as during the interictal (between-seizure) period (Fig. 10-7). A prodrome can be seen in generalized epilepsy, although it is more common in focal epilepsy, particularly temporal lobe epilepsy. A prodrome may consist of irritability, apprehension, sullenness, or a sense of discomfort or disease that builds up gradually over hours to days before a seizure. The prodromal state remits abruptly with the onset of the seizure. The pathophysiological basis for the prodromal state is unknown.

FIGURE 10-7 Progression of phases in epileptic seizure disorders.

Auras are focal seizures or the initial focal onset of a seizure and are associated with definable abnormal electrical discharges. Auras are abrupt in onset, last for seconds to minutes, may progress to a generalized seizure, or may terminate as the seizure ends. The type of clinical phenomenon depends on the site of origin and can include motor, sensory, autonomic, perceptual, cognitive, and affective abnormalities. Table 10-14 lists a number of common clinical manifestations of auras or focal seizures based on the anatomical site of origin. The auras accompanying seizures originating in the temporal lobe are the most varied. In general, auras may comprise a variety of symptoms and may have unique, individual-specific features, such as the crying out of a particular phrase in a particular language. Despite the great variety of auras, in any individual auras tend to be stereotyped and consistent from seizure to seizure.

Sensory	Cognitive
Headache	Transient dysphasia
Focal pain	Speech automatisms
Paresthesia	Subjective confusion
Motor	Obsessional thinking
Nystagmus	Thought blocking
Head turning	Distortion of time sense
Scanning	Affective
Posturing	Fear
Stuttering	Anxiety
Autonomic	Sadness
Facial flushing	Embarrassment
Hot flashes	Placidity
Pallor	Guilt
Dizziness	Anger
Chest pain	Joy
Tinnitus	Elation
Perceptual	Other
Micropsia, macropsia	Genital sensations
Heightened auditory acuity	Sexual behaviors
Derealization, depersonalization	
Déjà vu, jamais vu	
Hallucinations (visual, auditory, olfactory, gustatory, and tactile)	

Table 10-14 Neuropsychiatric Manifestations (in the Aura and Ictus) of Focal Epilepsy

The *ictus*, the epileptic attack, may be generalized or focal. *Primary generalized tonic-clonic epilepsy (grand mal epilepsy)* is characterized by a behavioral arrest or sudden loss of consciousness. This is followed by tonic extension of the upper and lower extremities, then clonic, rhythmic jerking of the extremities. The jerking gradually decreases in frequency, leading to muscle flaccidity. The total duration of the ictus is usually 2 to 5 minutes. Associated features may include urinary and bowel incontinence, sweating, and tachycardia. Generalized absence or petit mal epilepsy is characterized by brief lapses of consciousness lasting 3 to 30 seconds. There are no associated tonic-clonic movements, nor is there loss of postural tone. There may be a slight rhythmical twitching of the mouth. Seizures can occur numerous times during the day. Absence seizures are common and occur primarily in children ages 4 to 12 years. In both types of generalized epilepsies there is amnesia for the epileptic event. Myoclonic epilepsy is characterized by nonrhythmic, brief jerks of the limbs, trunk, and head. Myoclonic jerks are asynchronous, with body parts jerking at different times and in different sequences, and there is no loss of consciousness. Myoclonic epilepsy is frequently of familial etiology, although it may be associated with brain injury or systemic disease that affects brain function, such as chronic renal disease or hepatic insufficiency.

Partial or focal seizures are distinguished by their localized site of origin. The symptomatology can be simple (elementary) or complex, with the latter characterized by some degree of impairment of consciousness. Partial seizures with elementary symptomatology include focal motor symptoms, focal sensory symptoms, autonomic symptoms, or mixed symptomatology; consciousness is retained throughout the episode, although the seizure discharges can spread to other areas of the brain (jacksonian march) and can also develop into a generalized seizure (secondary generalization). The postictal period may be characterized by a residual focal deficit, such as motor weakness (Todd's paralysis) or dysphasia. Secondary generalization may occur rapidly, giving the false impression of an immediate generalized seizure; videotaped monitoring with simultaneous EEG may be necessary for differentiation.

Partial complex seizures, also known as psychomotor seizures, temporal lobe seizures, or temporal-limbic seizures, are perhaps of greatest interest to psychiatry. The great majority of partial complex seizures originate in the temporal lobes, but the frontal lobes and other sites have also been recorded as seizure foci. The range of presenting symptomatology varies from patient to patient and may include a broad spectrum of disturbances in behavior, cognition, and affect. Auras are frequent in partial complex seizures, representing the focal onset, and it may be difficult to distinguish the aura from the ictus. They are generally associated with clouding of consciousness but retention of posture and muscle tone. The patient may exhibit simple or complex movements, such as pulling on clothing, buttoning or unbuttoning clothing, purposeless hand movements, and fumbling with objects, or may continue with the behavior initiated prior to the seizure, such as closing a window. There may be staring, lip smacking, and wandering. The actual ictus cannot be distinguished from the aura.

Psychiatrists should be familiar with the range of symptomatology associated with partial complex or temporal lobe epilepsy because that disorder is an important diagnostic consideration in adult patients presenting with the new onset of behavioral disturbances. In light of the protean possible manifestations of partial complex epilepsy, the physician must keep in mind that there is a general consistency to the form of partial complex seizures: They have a definite and observable onset and termination; they are always associated with impairment in consciousness, such as confusion or inability to perform cognitive tasks; and they are relatively stereotyped for an individual from episode to episode.

The postictal period may also be characterized by severe disturbances of behavior. Primary generalized tonic-clonic seizures are usually followed by a period of sleep, sometimes headache, and nausea. Focal or partial seizures may have residual focal deficits of varying duration. In partial complex seizures, recovery of consciousness may lag behind recovery of motor function. Frequently automatic behavior, such as repetitive mouth movements, arm movements, and pacing, can be observed. As the postictal period is essentially a delirious state, confusion and cognitive impairment remain. Any disturbance of mood is possible, including anger, lovingness, and the epileptic furor (random, typically nondirected displays of violence and property destruction). The postictal period usually lasts only minutes, although it may last hours to days. A very rare but disturbing psychiatric postictal complication has been termed subacute postictal aggression. This is associated with directed violence that occurs in well-documented cases of epilepsy, where a patient stereotypically becomes increasingly psychotic, paranoid, and combative following a bout of uncontrolled seizures. Treatment requires both anticonvulsant and antipsychotic medications, although reported cases have occurred in patients who have poorly controlled epilepsy.

Course Epilepsy has an annual incidence of approximately 20 to 50 new cases per 100,000 population. The prevalence is 0.2 to 1.0 percent. The majority of cases of epilepsy in patients older than 15 years are of the partial or focal type. Only approximately 25 percent of adolescents or adults over the age of 15 years with seizures have generalized epilepsy. Seizures with onset in childhood are more commonly generalized, particularly absence seizures. Primary generalized tonic-clonic seizures usually occur for the first time before the age of 35 years, although they can occur at any age; absence seizures usually first manifest between the ages of 4 and 12 years. Focal seizures also have their onset commonly before the age of 20 years.

The natural history of seizure disorders has not been defined clearly. Up to one third of all seizures may remit spontaneously without treatment. Absence seizures are generally outgrown by the age of 20 years, although many patients do develop another form of generalized epilepsy as adults.

The etiological considerations for seizures vary with the age at onset. Early-onset epilepsy is usually a concomitant of genetic factors or an insult to the developing neural system in utero or in childhood; the latter can include trauma, infections, or toxic exposures. For seizures starting in adulthood, the etiological considerations include alcohol or drug withdrawal, trauma, infection, and tumors. The tumors are the primary causes of seizure disorders during the middle adult years; cerebrovascular disease is the most common etiology among the elderly.

Treatment Anticonvulsant pharmacotherapy first developed empirically, without specific knowledge of either the pathophysiology of seizures or the neurochemical mechanisms of therapeutic agents. Although the primary prescribed anticonvulsants (carbamazepine [Tegretol], phenytoin [Dilantin], valproic acid [Depakene], and phenobarbital) have remained consistent since the 1980s, recent years have seen the rapid emergence of new compounds (Table 10-15). Most medications work through one of three mechanisms of action: potentiation of g-aminobutyric acid (GABA)-mediated neuronal inhibition, inhibition of glutamate-mediated excitatory postsynaptic receptors, or control of sodium and calcium voltage-gated ion channels; Table 10-16 summarizes these actions. It is particularly important to note that anticonvulsant medications have proven to be a fruitful source of novel psychiatric compounds.

Conventional Drugs	
Phenobarbital	1912
Phenytoin	1938
Trimethadione	1946
Phensuximide	1953
Methsuximide	1957
Ethosuximide	1960
Benzodiazepines	1965
Carbamazepine	1974
Valproate	1978
New Drugs*	
Felbamate	1993
Gabapentin	1994
Lamotrigine	1995
Topiramate	1997
Tiagabine	1997
Vigabatrin	In Europe
Oxcarbazepine	In Europe
Loreclezole	In trials
Stiripentol	In trials
Zonisamide	In trials

* Dates refer to U.S. introduction; lamotrigine was marketed earlier in Europe.
Adapted from Rogawski MA: Mechanisms of antiepileptic drugs. Epilepsia 40(suppl 10):S3, 1999.

Table 10-15 Chronology* of Antiepileptic Drugs

	Effects on Voltage-Gated Channels			Effect on GABA Receptor	Effects on Excitatory Amino Acid Receptors	Indications
	Sodium	Calcium	Potassium			
Barbiturates	+/?	+/?	?	+ +	+/?	Broad-spectrum anticonvulsants
Benzodiazepines	+/?	-	?	+ +	-	Broad-spectrum anticonvulsants
Carbamazepine	+ +	+/?	+/?	-	-	Partial seizures
Ethosuximide	-	+ + (T-type)	-	-	-	Absence seizures
Felbamate	+	?	?	+/?	+ + (NMDA)	Broad-spectrum anticonvulsant*
Gabapentin	+/?	-	?	+	+/?	Partial seizures
Lamotrigine	+ +	-	?	-	+/?	Partial seizures/broad-spectrum
Oxcarbazepine	+	+/?	?	-	-	Partial seizures
Phenytoin	+ +	+/?	-	-	-	Partial seizures
Tiagabine	-	+	+	+ +	-	Partial seizures
Topiramate	+/?	-	?	+/?	+ (Non-NMDA)	Partial seizures/broad-spectrum
Vigabatrin	-	-	?	+ +	-	Partial seizures/infantile spasms
Valproate	+	+/?	+	+/?	-	Broad-spectrum anticonvulsant

* Felbamate is not a first-line treatment because of risk of aplastic anemia and hepatic failure.
+/? Shown but not defined
+ Established
+ + Well-established
- Negative finding
? Unestablished data
Reprinted with permission from Ives M: Molecular mechanisms of antiepileptic drugs. Epilepsia 40:S3, 1999.

Table 10-16 Antiepileptic Drugs

Psychopathology Psychopathology, namely disturbances in behavior, cognition, perception, or mood, can occur at any point in the seizure process. The prodrome may be characterized by a sense of irritability or apprehension. Families often report that they know when a relative is going to have a seizure on the basis of a change in temperament or disposition. Auras may include a variety of psychopathology, including dissociative experiences, hallucinations in all spheres, derealization, depersonalization, and disturbances of mood or affect. The disturbances of mood, such as a subjective sense of fear, anxiety, or depression, can be distinguished from normal expressions of the same emotion in that generally they are more crude, stereotyped, and brief emotional states. Joy, elation, or euphoria is less common. Ictal states can manifest striking changes in behavior that are likely to be coarse and disorganized. The list of ictal manifestations in Table 10-14 includes a variety of hallucinations and dissociative experiences, as well as sudden and unpredictable shifts in mood. Other sensory or psychic experiences can occur out of their usual context (a classic case is that of a woman who had spontaneous orgasms in church). The postictal state is a delirium that can display the full range of disturbances in level of arousal, ranging from stupor to hypervigilance. Partial seizures may be followed by a milder delirium that is detectable only from disorganization in behavior or difficulty with simple cognitive tasks such as registration and repetition. The issue of interictal psychopathology has been much studied and debated, in general, the psychiatric symptomatology associated with the prodrome, aura, ictus, and postictal state is remarkably broad—virtually any thought, feeling, movement, or perception that the brain can produce may be seen. The clinician must be attentive to the form and course of these symptoms whenever considering epilepsy as a possible etiological explanation for abnormal behavior.

Violence and Aggression The issue of violence or aggression as a neuropsychiatric manifestation of an ictus has provoked much controversy. In the legal arena epilepsy is occasionally invoked as a defense to mitigate culpability for a violent or even murderous act. Irritability or agitation can be a component of the prodrome, the aura and ictus can encompass angry affect and striking out, and the postictal state can manifest with fear and confusion with intact motor function. Although this might suggest the possibility of violent acts as a component of seizures, there is limited potential for such actions. Automatic acts of violence during epileptic seizures are short-lived, fragmentary, undirected, and most often occur in response to actions (such as attempts at restraint) that provoke or irritate the seizing individual. Examples include spitting, swearing, and striking out in a flailing fashion. For violence to be considered a manifestation of epilepsy, it must conform to the known

temporal sequence and symptomatology of a seizure; namely, there must be a clear onset and termination, together with other clinical signs (e.g., confusion, incontinence, and impairment of consciousness) and stereotypy. A special 1981 epilepsy task force of the National Institute of Neurological and Communicative Disorders and Stroke (NINCDS), after studying videotapes of selected violent patients with epilepsy, recommended criteria to determine if a particular act of violence is ictal: (1) a clear diagnosis of epilepsy; (2) documented automatisms, preferably on videotape; (3) documented aggression during the automatisms that corresponds to an EEG-proved ictus; (4) demonstration that the aggressive act is characteristic of the patient's usual seizure form; and (5) clinical consensus that the act was related to the actual seizure.

People with epilepsy have long been thought to display psychopathology during the interictal period as well. Epilepsy was thought to be a subtype or complication of insanity, and so was included in most psychiatric nosologies of the past two centuries. Epilepsy patients typically were housed in asylums for the insane. In 1791 Philippe Pinel included among his recommendations for asylums the suggestion that other patients be shielded from epilepsy patients because of their “almost always incurable” status and his sense that “few objects are found to inspire so much horror and repugnance . . . than the sight of epileptic fits.” Griesinger stated that “a very great number of epileptics are in a state of chronic mental disease even during the intervals between the attacks.” Interictal psychopathology can be grouped into psychotic disorders, mood disorders, personality abnormalities, cognitive disorders, and secondary repercussions.

Psychotic Disorders Jean Etienne Esquirol, in his 1845 description of female institutionalized epilepsy patients, reported frequent psychotic symptoms, including hallucinations in all spheres:

> They have hallucinations most varied . . . they think they see luminous bodies by which they fear they might be embraced . . . they smell odors the most fetid . . . they hear sounds like the bursting of a thunderbolt, the roll of drums, the clash of arms in the din of combat. Karl Jaspers, in *General Psychopathology*, classified epilepsy as one of the three major psychoses, along with schizophrenia and manic-depressive illness. He defined *genuine epilepsy* as “convulsive disorders which are not due to any known somatic process.” Numerous studies have evaluated psychosis among epilepsy patients; unfortunately, most have been hampered by the lack of clear or standardized definitions for the symptoms being investigated. In addition, many studies have not distinguished between psychosis occurring in the context of the ictus, the prodrome, or the postictal state, and many have not indicated whether such symptoms were detected specifically during periods of interictal electrical stability. Few studies have discriminated between symptoms occurring in clear consciousness versus those occurring coincident with impaired consciousness. Nonetheless, clinicians generally encounter a higher incidence of psychosis among epilepsy patients, particularly those with temporal lobe foci, than in the general population. Some studies have specifically referred to paranoid ideation, delusions, ideas of reference, visual hallucinations, and first-rank auditory hallucinations as being common. In 1963 Eliot Slater and A.W. Beard identified an atypical schizophrenia, characterized by visual and auditory hallucinations, ideas of reference, and persecutory delusions, occurring in the context of preservation of affect and a level of social adaptation better than that of comparably psychotic schizophrenia patients. They further noted that the psychotic symptoms did not occur until many years after the onset of seizures (a mean of 14 years) and then occurred with an apparent periodicity. The occurrence of the psychotic episodes was unrelated both to the frequency of seizures and to measured anticonvulsant efficacy.

Although the true incidence or prevalence of psychotic disorders in epilepsy patients is unknown, there is general consensus that the psychotic disorder seen in epilepsy is distinct in form from idiopathic psychiatric diseases and is characterized by the features described by Slater and Beard. Other associated findings may include an association between psychosis and left hemispheric seizure focus (especially in the temporal lobe), female sex, sinistrality, or tissue abnormality (alien tissue, such as hamartomas or focal dysplasia, is more commonly found at autopsy or after surgical excision in the psychotic patients). These psychotic disturbances have been referred to as the schizophrenia-like psychosis of epilepsy and the interictal psychosis of epilepsy. Standard antipsychotic medications are beneficial symptomatically but not as efficacious as in idiopathic psychotic disorders. Some studies have noted marked improvement in the psychotic symptoms with improved seizure control following either pharmacotherapy or surgical excision of the seizure focus. A smaller number of studies have reported an increase in psychotic symptoms occurring with improved seizure control, prompting a theory of antagonism between symptoms of psychosis and seizure control.

The observation of a consistent psychotic disorder of increased prevalence in epilepsy patients has led to an intense search for the underlying mechanism in the hope of describing a more general explanation for psychotic processes. Three hypothetical mechanisms for interictal psychoses have been advanced. The first hypothesis suggests that the schizophrenia-like illness in epilepsy is epileptic in origin or related to abnormal electrical brain discharges. Kindling, an experimental animal model for the spread of epileptic foci, has been suggested as a paradigm for the development of psychosis in epilepsy. Chronic stimulation of the brain in animals can lower the electrical threshold for the development of electrical or clinical seizures. Over time, abnormal discharges develop at previously subthreshold levels of stimulation or even spontaneously. Abnormal behaviors associated with these experimentally induced brain discharges can also persist, even after stimulation has ceased and there are no motoric convulsions. It has been suggested that human kindling occurs at various brain foci, particularly the temporolimbic structures, resulting in psychotic and behavioral disturbances that may manifest only after years of seizure activity.

A second hypothesis for the development of interictal psychosis focuses on a proposed antagonistic relationship between seizure frequency (or, more accurately, EEG abnormality) and psychotic symptoms: forced normalization is the putative process by which a psychosis of sudden onset can manifest with the achievement of seizure control and associated with a normal cortical EEG. Studies of forced normalization have primarily involved case reports, and findings have been difficult to replicate. Clinical treatment regimens based on the antagonism theory, such as allowing episodic seizures or performing ECT on psychotic epilepsy patients, have proved ineffective, further weakening that proposal.

A third hypothesis to explain interictal psychosis suggests that it may not be related specifically to abnormal electrical activity but may instead reflect a common brain dysfunction that causes both epilepsy and psychosis. This hypothesis stresses the dysfunctional or broken brain inferred in epilepsy and regards the psychotic disorder as yet another symptomatic manifestation. Recent quantitative brain-imaging studies in epilepsy patients with psychotic disorders have not revealed consistent structural abnormalities within that patient group. However, specific symptom correlations have been reported, among them an increased frequency of temporal lobe structural abnormalities in epilepsy patients with auditory hallucinations. Similar findings associating specific psychotic symptoms with defined cerebral abnormalities have also been reported in other psychiatric conditions (e.g., schizophrenia). Further symptom-based research utilizing imaging and physiologically sensitive techniques may better delineate the brain regions where dysfunction can lead to particular psychotic symptoms.

Mood Disorders Affective changes can occur as part of the seizure prodrome, aura, ictus, or postictal state. Irritability is a common prodromal manifestation. Temporal lobe auras may be accompanied by mood abnormalities, most commonly fear and anxiety, although a depressive affect is possible and, more rarely, elation or euphoria. Descriptions of postictal sadness are common. All of those affective changes are generally brief, lasting minutes to hours. They differ from normally experienced vacillations in mood, in that they occur independent of any particular context, and from the pervasive and enduring affective changes found in primary mood disorders.

Mood disorders in the interictal period have not been studied as comprehensively as psychotic or personality disorders. Many authors have noted that epilepsy patients have a strong tendency to endorse items of sadness and anxiety on self-report inventories. Few studies have used clinical examinations or standardized interviews to determine the presence of mood disorders such as major depressive disorder or bipolar I disorder. One study of epilepsy patients diagnosed with major depressive disorder found that at least half of the patients had family histories of mood disorder and that no clear relation existed between severity of depression and seizure type, seizure frequency, seizure focus, or age at seizure onset. Although definitive studies of the incidence and prevalence of clinically defined mood disorders in epilepsy are needed, there is less support overall (compared with psychosis) for an elevated prevalence of mood disorders in epilepsy patient populations. Many patients do express persisting dysphoria, perhaps reflecting the dissatisfaction and maladjustment associated with a chronic disease.

Regardless of the cause of the dysphoria or dissatisfaction in epilepsy patients, there is an increased prevalence of suicide attempts and completed suicides. The incidence of suicide in patients with epilepsy is fivefold greater than in the general population. In patients with temporal lobe epilepsy the incidence of suicide increases to 25 times that of the general population, but the underlying psychopathology remains to be defined.

Personality Change There has been a long-standing misperception that an epileptic personality is distinguishable and common. Esquirol noted in his studies of 385 female epilepsy patients that “only one fifth were free from intellectual derangement, but nearly all of these were irritable, peculiar, and easily enraged.” Griesinger commented on the “dominant, suspicious, discontented, misanthropic perversion of sentiment . . . observed in many epileptics.” Eugen Bleuler spoke of the “epileptic excess of emotion . . . easily aroused, remarkably persistent . . . difficulty in abandoning any particular thought . . . fixation to a single theme . . . precise attention to detail.” Karl Jaspers described “viscosity, slowing down, explosiveness and dementia” as characteristic of epilepsy patients. These characterizations were frequently based on chronically institutionalized patient populations representing a selection of the most severely impaired patients, in whom the effects of brain injury (especially related to repeated seizures, status epilepticus, and recurrent hypoxia), the deprivations of institutionalization, and toxic treatments undoubtedly confounded clinical observation. More recent attempts at detecting a recognizable, diagnosable personality disorder in community samples of epilepsy patients resulted neither in the description of a discrete personality syndrome nor in a higher prevalence of known personality disorders.

Because standard personality inventories have not uncovered specific abnormalities in epilepsy patients some researchers have focused on particular traits or behaviors. An interictal behavior syndrome of temporal lobe epilepsy has been described that encompasses four traits or behaviors: (1) altered sexuality, usually a decreased interest in sexual matters but at times involving hypersexuality or deviant sexual interests; (2) hyperreligiosity, described as an unusually deepened interest in moral affairs and matters of global importance, with vivid case descriptions of multiple religious conversions and intrusive polemicizing; (3) hypergraphia, with patients maintaining voluminous writings, including journals, essays, and novels; and (4) viscosity or stickiness, a characteristic described for more than a century, including a preoccupation with detail, digressive or overly inclusive speech, and resulting impairments in social discourse.

Although the literature supporting such a personality syndrome is rich with clinical case histories, systematic study to define such a syndrome has been difficult to replicate. Investigators using an 18-point inventory of those behaviors found they could not distinguish patients with epilepsy from other psychiatric populations or patients with temporal lobe epilepsy from those with generalized epilepsy. Thus, the bulk of data suggests that the clinical complex of overinclusiveness in speech, interpersonal action, and writing; alteration of sexuality; and intensified emotion and cognition (hypercosmiscity) is rare and not specific for temporal lobe epilepsy. However, isolated features of this cluster may be more common among patients with temporal lobe epilepsy. When the entire picture is encountered clinically in the absence of a readily apparent seizure disorder, the clinician may wish to pursue a more intensive evaluation if the patient fails to respond to standard psychiatric therapies.

Cognitive Disorders An early view of epilepsy considered it a degenerative disorder with a progressive deterioration in cognition, similar to that seen in the degenerative dementias. Modern prospective studies, however, have disproved this belief and have demonstrated no progressive decline in cognitive skills in a general population of epilepsy patients. A subpopulation of patients with epilepsy does demonstrate a lower intelligence quotient (I.Q.) spread than the normal distribution. This probably results from a combination of factors, including the original brain damage or dysfunction responsible for the epilepsy, occasional disruption attributable to the seizure disorder, and drug effects. Numerous anticonvulsants, including ethosuximide (Zarontin), phenytoin, phenobarbital (Solfoton), and carbamazepine, have been demonstrated to lower performance on tasks of concentration, memory, and motor speed; motor speed is least impaired.) Epileptic dementia, while certainly uncommon, has been described in patients with defined CNS lesions or uncontrolled seizures. It probably reflects the cumulative effect of frequent seizure-induced hypoxic episodes and perhaps toxicity from long-standing treatment with high dosages of anticonvulsants. Phenytoin, in particular, has been demonstrated to cause cerebellar degeneration with long-term use.

Behavioral and Secondary Repercussions Despite a less glamorous research appeal, the more compelling and clinically demanding aspects of interictal function may be the behavioral, interpersonal, and social problems arising from irritability, agitation, or aggression. There are no well-defined, systematic, or tested approaches to treating those behavioral difficulties when they arise. It was long believed that maladaptive behavior, particularly aggressive behavior, as characterized by physical assaultiveness, destructiveness, and self-injury, was more common in patients with epilepsy. Recent methodologically rigorous studies, however, demonstrated that maladaptive behavior does not correlate with the presence of epilepsy when epilepsy populations are compared with appropriately matched controls. Maladaptive behavior is, however, related to the overall extent of brain damage in both types of populations. This important point cannot be overemphasized: Epilepsy is a florid manifestation of a physiologically abnormal brain, and in most instances, the fundamental CNS dysfunction that causes seizures also causes associated neuropsychiatric abnormalities.

Especially among severely afflicted epilepsy patients, obstacles to effective social functioning and personal autonomy pose the greatest therapeutic challenges for clinicians and families. Patients may not receive adequate education because seizures may interfere with daily school attendance. Patients may be restricted from many occupations owing to employers' fears of patients sustaining injury in the workplace. Limitations on driving can markedly diminish the independent functioning of these individuals and many patients are compelled to remain dependent on family, even into adulthood.

Pseudoseizures Pseudoseizures simulate the motor behavior of true seizures but do not involve abnormal electrical discharges. Pseudoseizures can be distinguished from true seizures by the form of the seizure and by the lack of the usual associated features. The form does not fit the known patterns for epileptic attacks and can consist of random flailing about. Furthermore, the form can be variable from seizure to seizure, lacking the stereotypy typical of true seizures. Incontinence and tongue biting are rare. There is minimal confusion at the conclusion of the episode, and no abnormalities are detected on neurological examination. EEG can be helpful in distinguishing pseudoseizures, especially if one can be obtained during an event and then studied for evidence of electrical discharges that would correspond with the motor behavior. An EEG obtained after a true seizure should demonstrate areas of slowing that would not be seen after a pseudoseizure. Serum prolactin levels increase markedly immediately after a seizure and can be helpful in distinguishing true seizures from pseudoseizures. Video EEG telemetry is the definitive means of determining whether an observed seizure is epileptic in origin or a pseudoseizure.

Pseudoseizures are more likely in patients who suppress emotion or express emotion through somatic means. Pseudoseizures most often are conversion disorders in which the patient does not have conscious volitional control of the behavior; rarely are pseudoseizures the result of faking or malingering. Pseudoseizures frequently occur in patients who have true epileptic disorders, confounding the diagnosis. W. Alwyn Lishman has aptly noted that the diagnostic error of interpreting epilepsy as pseudoseizures is probably much more common than the reverse and is far more detrimental to the patient's well-being. The diagnosis of pseudoseizures rests not on the presence of any particular personality traits or identifiable psychosocial stressors, but rather on the form of the seizure, associated features, and EEG confirmation.

Primary Psychiatric Disorders The presence of intellectual deficits, whether identified with bedside procedures or on standardized neuropsychological tests, does not automatically warrant the diagnosis of a cognitive impairment disorder. Neuropsychological abnormalities occur frequently in many patient populations. Once neglected or considered epiphenomena of more central emotional disturbances, cognitive processing deficits are now known to be key components of clinical disorders such as major depressive disorder (especially in the elderly), acute and chronic schizophrenia, chronic alcohol dependence, and perhaps obsessive-compulsive disorder. Cognitive impairment disorders, such as dementia or delirium, or secondary psychiatric syndromes are all caused by specific disease processes; vigilant diagnostic evaluation usually leads to detection of a primary systemic or cerebral disturbance.

Difficulties may arise when the cause is presumed but cannot be proved, as in the case of dementia of the Alzheimer's type, where the definite diagnosis must await postmortem brain examination. When a patient has both a major depressive syndrome and clinical findings consistent with incipient dementia of the Alzheimer's type, it may not be possible to determine immediately the fundamental disturbance being evaluated. Such confusing situations typically arise with the near-simultaneous onset of both symptom clusters or in patients who have experienced major depressive disorder previously. Such patients require careful definition of symptoms, initiation of therapy for all potentially treatable conditions, and serial monitoring of the patient's responses. Documentation of the longer-term course also assists in disentangling and recognizing separately contributing disease processes.

Despite careful observation and follow-up evaluation, the clinician may remain uncertain whether a syndrome is idiopathic or secondary to other detected diseases. In such instances, it is preferable to diagnose a primary psychiatric condition on Axis I, define all systemic or cerebral conditions on Axis III, and thereafter maintain a high order of vigilance while monitoring the course of the disorder longitudinally. It is important to note questions or uncertainties in the medical record for later scrutiny, for that practice avoids premature diagnostic closure.

Clinicians must also guard against willingness to provide a psychiatric diagnosis when specialists from other medical disciplines have ruled out specific disease processes after laboratory tests have been unrevealing. The failure to define an organic disease does not warrant a functional diagnosis by default. As emphasized in DSM-IV, specific clinical signs, symptoms, and course are needed to establish the presence of a primary psychiatric disorder.

Metabolic Disorders Because most systemic medical conditions can directly or indirectly affect brain function, any list of illnesses that may cause a secondary psychiatric syndrome or cognitive disorder must be incomplete. Table 10-17 lists some frequently described potential causes. The precise pathophysiological mechanisms by which the disease process alters brain function are poorly understood in most cases. More than one process may be involved. For example, a patient with acute myelogenous leukemia may have altered brain function resulting from the neoplastic process itself, anemia (with decreased oxygen delivery to the brain), brain hemorrhages (caused by thrombocytopenia), and infections.

Endocrine	
	Hypothyroidism or hyperthyroidism
	Hypoparathyroidism or hyperparathyroidism
	Sex hormones—too little, too much, or cycle-related
	Hypocorticolemia or hypercorticolemia
	Pheochromocytoma
	Diabetes mellitus
	Carcinoid syndrome
Pulmonary	
	Hypoxemia
	Hypercarbia
	Infections (pneumonia, bronchiectasis, abscess)
	Restrictive lung disease
	Obstructive lung disease
	Tumors
Hematologic	
	Anemia
	Polycythemia
	Neoplasms
Hepatic	
	Liver failure due to any etiology
Renal	
	Renal failure due to any etiology
	Infections
	Neoplasms
Nutrition	
	Marasmus
	Kwashiorkor
	Vitamin deficiencies (thiamine, niacin, B_{12}, folate)
	Dehydration
Miscellaneous	
	Hyperelectrolyte and hypoelectrolyte disturbances (especially so- dium, potassium, phosphate, calcium, magnesium)
	Hypoglycemia and hyperglycemia
	Acidemia
	Alkalemia
	Porphyria

Table 10-17 Metabolic and Other Systemic Disturbances

Secondary psychiatric syndromes may be the first, most prominent, or only clinical phenomena to call attention to the underlying condition (e.g., depression due to occult pancreatic carcinoma and cognitive deficits due to vitamin B_{12} deficiency even in the absence of other neurological or hematological signs). Secondary behavioral changes may also result from multiple etiologies, of which any one alone might or might not be sufficient to produce the psychiatric disturbance (e.g., delirium due to mild anemia, mild hyponatremia, and marginal hypoxemia). The rate of change may also be important with certain etiologies. For example, a sudden drop in serum sodium to 125 mEq/L is more likely to produce behavioral changes than a chronic hyponatremia of 125 mEq/L.

Etiologies particularly identified with specific psychiatric syndromes are discussed under each syndrome. However, most etiologies can produce more than one syndrome (e.g., hypothyroidism is most often associated with a depressive state but may also cause mania, delirium, or dementia). How a specific etiological factor causes varied behavioral changes presumably results from both trait-dependent and state-dependent brain diatheses that largely are not understood.

Demyelinating Disorders With regard to secondary psychiatric syndromes, multiple sclerosis is the most important demyelinating disorder. Although it may cause delirium, dementia, and nonaffective psychoses, mood disturbances have been described frequently. Isolated, persistent euphoria has long been noted and is thought to be physiologically related to demyelinated lesions in the limbic system, frontal lobes, and basal ganglia (Fig. 10-8). *Emotional incontinence*, also termed *pathological laughing or weeping*, is a state of labile affective expression that is apparently disconnected from underlying mood. Although pathophysiological mechanisms remain uncertain, it is speculated that interruption of pathways between the telencephalon and lower regions is responsible.

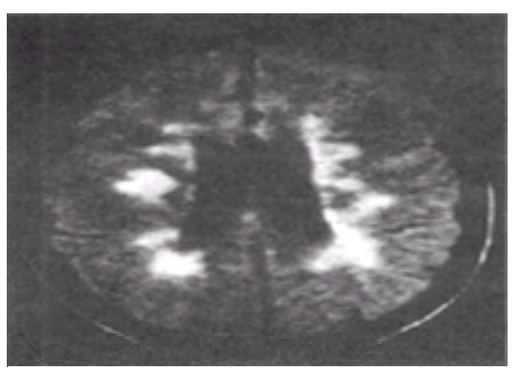

FIGURE 10-8 MRI study of the head. Extensive confluent periventricular lesions are seen in a patient with multiple sclerosis. The corresponding CT scan was normal. (Reprinted with permission from Lukes SA, Crooks LE, Aminoff MJ, Kaufman L, Panitch HS, Mills C, Norman D: Nuclear magnetic resonance imaging in multiple sclerosis. Ann Neurol *13*:596, 1983.)

Depressive syndromes are especially prevalent in patients with multiple sclerosis. Most studies have focused on major depressive disorder or have not used specific diagnostic criteria; therefore, little is known about minor or other subsyndromal depressions. Some data suggest that depressive syndromes can be caused directly by the demyelinating process in specific CNS regions; for example, patients with cerebral disease have higher rates of depression than patients with spinal multiple sclerosis. However, there is also evidence implicating psychological and social factors in the pathogenesis of depression in these patients, and it is likely that many of their depressive syndromes are not secondary to or specifically symptomatic of cerebral disease.

Bipolar disorders are less well studied, although they appear to be more prevalent in multiple sclerosis populations than in the general population. Whether bipolar disorders in multiple sclerosis patients represent true secondary bipolar disorders, mania induced by treatments with corticosteroids, or manifestations of a shared genetic diathesis is not known.

Other demyelinating disorders include variants of acute encephalomyelitis (including postinfectious encephalomyelitis and acute posttraumatic demyelinization). These disorders generally manifest initially with coma or delirium, and survivors may have lasting cognitive and behavioral disturbances.

The treatment of psychiatric syndromes in demyelinating disorders, as for most secondary syndromes, is empirical and based on target symptoms (e.g., antipsychotic agents for psychosis). Studies suggest that antidepressant medications, when used in conjunction with supportive psychotherapy, are helpful in patients with multiple sclerosis. Tricyclic drugs, and possibly other classes of antidepressant drugs, reduce emotional incontinence associated with multiple sclerosis even in the absence of a full depressive syndrome.

Degenerative Diseases Degenerative disease processes involve deterioration of brain function resulting in the loss of previously attained capacities. Some degree of CNS degeneration occurs as part of the aging process and is reflected in alterations in gross brain structure; neuronal cell number, morphology, and function; and neurotransmitter synthesis, metabolism, and function. The degeneration manifests clinically with psychometrically definable cognitive declines, particularly in secondary or long-term memory, speed of mental processing, visuospatial processing, divided attention, and cognitive flexibility. A range of decrements is associated with normal aging, and clinicians and investigators are developing neuropsychological procedures that distinguish such normative processes from those associated with degenerative diseases. Indeed, DSM-IV includes age-related cognitive decline as a clinical condition that may be a focus for clinical attention although it is not a mental disorder. It allows physicians to explain to concerned healthy patients that their aging-associated decrements in cognitive processing are distinctive from incipient dementia. Other similar conditions include problems such as borderline mental functioning or bereavement.

However, the border between normative cognitive changes and incipient dementia is ill defined. Patients who present with acquired deficits below the normal range are at higher risk statistically of developing progressive problems. However, cross-sectional cognitive tests cannot reveal time of onset, and in some individuals test results are always below normal. Conversely, people with above-average intellectual abilities may experience cognitive decline without exhibiting objective deficits because their performance remains within normal limits despite obvious functional impairments. Degenerative CNS diseases can produce disturbances in cognition, mood, behavior, personality, and motor and perceptual function. In recent years investigators have more clearly defined that the major degenerative diseases (whether hereditary or idiopathic) reflect deterioration in specific or discrete neurochemical systems where there is as well a regional specificity that reflects the location of cell bodies and their ultimate terminal zones. Such diseases are best considered neurochemical system diseases and their general pattern is one of insidious onset with a gradual progression of deficits. Dementia is the most common syndromic presentation and worsens as a reflection of the long-term progressive process; it ultimately reflects widespread CNS disease. Secondary mood, psychotic, and personality syndromes are also seen but are more likely to manifest earlier in the course, when the degeneration may be more localized.

Cortical and Subcortical Dementia Degenerative CNS diseases can be distinguished clinically from one another by the relative impairment and sparing of various cognitive and behavioral functions. Two basic clinical patterns of dementia have been characterized clinically: *cortical* and *subcortical*. The cortical pattern of

dementia is characterized by impairments in memory (primarily a storage and recall deficit) and gnostic-practic abilities (primarily involving language, visuospatial abilities, calculation, and motor praxis). Executive or managerial functions such as organization, judgment, abstraction, emotional control or modulation, and insight and social judgment are similarly affected. Fine and gross motor movements are generally preserved until later in the disease course. Personality often remains intact or displays subtle variations, with patients becoming more passive or less spontaneous, or becoming coarse and crude in their interactions. With disease progression the changes in personality become more common and pronounced. Affective expression is generally preserved, although again a coarsening may be noted in the form of emotional lability. Early in the disease, patients frequently discern and express dismay about their intellectual decline.

The subcortical pattern is characterized by a generalized slowing of mental processing. Specific cognitive skills, such as calculation, naming, or copying are less affected initially, in contrast to their early decline in the cortical degenerative processes. Verbal and visual memory impairment may be present early in the course, although such impairment more often takes the form of forgetfulness or a failure of retrieval that is initially amenable to prompting, in contrast to the more severe recall deficits of cortical dementia. Patients also show deficits in learning new motor movements or complex psychomotor procedures. Planning and organizational skills are disrupted. Abnormal movements are common and manifest as a slowing and awkwardness in normal movement or as the intrusion of such extraneous movements as chorea or tremor. In contrast to the early impairment of language function in cortical disease, language is relatively spared, although the motor production of speech may be abnormal. The personality change is often marked, with striking patterns of apathy, inertia, and diminished spontaneity. Mood disorders, including major depression and mania, occur frequently. The presenting symptoms in subcortical degenerative processes may be those of a personality change or a mood disorder at a time when cognitive impairment or motor dysfunction is not yet obvious. In the cortical processes, by contrast, the presenting symptoms more often reflect cognitive impairment, particularly memory and language dysfunction. As the dementia and the degenerative process progress, the clinical presentations of cortical and subcortical diseases become nearly indistinguishable from one another.

The term “subcortical dementia” was first used to describe the cognitive and behavioral deficits seen in patients with Huntington's disease. A similar clinical pattern was soon described for other subcortical diseases, such as progressive supranuclear palsy and Parkinson's disease. Although the term was initially used in reference to a clinical picture that could be localized to the subcortex, *subcortical dementia* is now considered a pseudoanatomical designation. It is clear from imaging and neuropathological studies that cortical dementia (e.g., dementia of the Alzheimer's type) is not restricted pathologically to the cortex; major affected cholinergic fiber pathways are subcortical in origin. Subcortical diseases similarly affect regions outside the subcortex, especially the frontal lobes, because of the brain's robust frontal-subcortical connections. Moreover, failure of subcortical nuclei that directly receive cortical efferent pathways can lead to clinical symptoms whose cerebral level of origin cannot be differentiated. Nonetheless, the cortical-subcortical distinction has been of clinical utility in defining patterns of cognitive, behavioral, mood, personality, and motor impairment, especially in the early stages of the degenerative disease process.

Alzheimer's Disease Alzheimer's disease is the prototype of a cortical degenerative disease. Alzheimer's original description in 1906 detailed most of the familiar clinical and neuropathological features. Of note, his patient suffered from paranoia in addition to cognitive decline. Currently, the diagnosis of Alzheimer's disease requires neuropathological confirmation, and the diagnosis is used clinically for cases identified antemortem. Age at onset s earlier in patients with a family history of the disease. Despite some data to suggest distinctive age-related clinical patterns, no phenomenological separation between early-onset and late-onset cases has been found consistently enough for age to substitute for detailed clinical description; however, early-onset dementia of the Alzheimer's type may have a more rapidly progressive course. A major component of the presenting symptoms is usually subjective complaints of memory difficulty, language impairment (“I can't find the word”), and dyspraxia (e.g., difficulty driving). Diagnosis at this juncture is primarily based on exclusion of other possible etiologies for dementia. No features of the physical examination or laboratory evaluation are pathognomonic for dementia of the Alzheimer's type. Some studies have apparently discriminated patients with dementia of the Alzheimer's type from patients with dementia of other etiologies and from normal controls by using techniques such as EEG, MRI, and SPECT. These studies have been difficult to replicate consistently, and at present, brain-imaging studies are best used to exclude other identifiable causes. Indeed, available technological diagnostic methods have not proved more sensitive and specific than astute clinical evaluation in comparisons of patients with dementia of the Alzheimer's type and healthy control subjects. PET holds promise but currently is too expensive for clinical diagnostic use.

A variety of diagnostic criteria sets have been developed for dementia of the Alzheimer's type. Clinical criteria have been verified prospectively in autopsy studies and have been found to be highly specific although only moderately sensitive. Implementation of the criteria requires extensive evaluation, including an informant-based history, neurological examination, neuropsychological testing, and laboratory and neuroimaging data. Studies using clinical samples collected in research centers tend to show the highest correlation between premorbid diagnosis and postmortem histopathology. In part this reflects the sophisticatmon of the evaluators; also, research cohorts can exclude subjects who show signs of other confounding conditions during the evolving disease course. Recent studies have shown substantial inter-center diagnostic variation, as well as variation when using different diagnostic criteria. Variability is greatest in population studies, in contrast to clinical samples. Factors found to be protective in epidemiological studies include higher education, larger head circumference, and cigarette smoking. The latter factor likely reflects some type of neuroprotective effect of nicotine; the former two point to a “brain reserve” hypothesis, where the unknown deleterious effects of the basic Alzheimer's disease pathobiology are mitigated by initially having either more brain substance or greater associative connections because of the stimulator effects of education.

Alzheimer's disease is characterized pathologically by generalized atrophy of the cerebral cortex (Fig. 10-9) and by neurofibrillary tangles, neuritic (amyloid) plaques, and granulovacuolar degeneration (Fig. 10-10). Although plaques and tangles may be detected in the brains of the nondemented elderly, they are more numerous in patients with dementia. In recent years investigators have attempted to circumvent the qualitative overlap in symptoms by developing stricter quantitative, age-adjusted pathological criteria for Alzheimer's disease. Controversy remains whether brains with plaques from individuals without dementia were “normal variations” or early pathological signs of incipient disease. A definitive diagnosis ultimately requires both the characteristic dementia in life and the characteristic pathology after death.

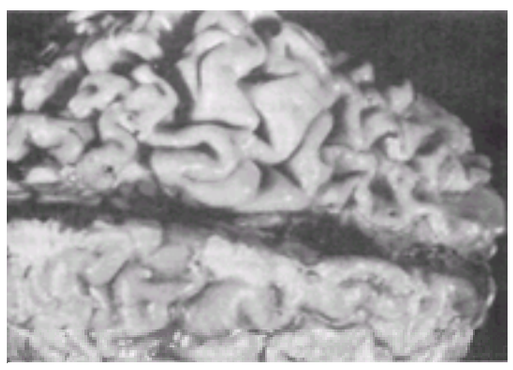

FIGURE 10-9 Alzheimer's disease. View of exposed left cortex showing severe atrophy. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

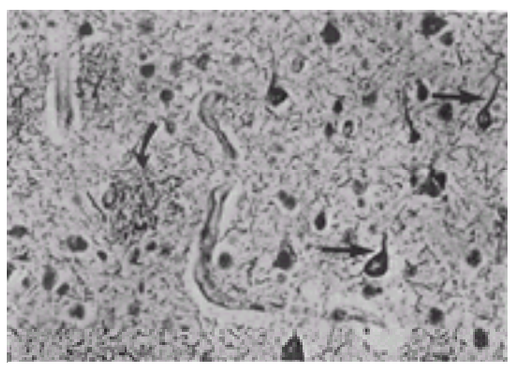

FIGURE 10-10 Light micrograph of the cerebral cortex showing neurofibrillary tangles (arrows) and senile plaque (curved arrow) in Alzheimer's disease. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

During recent years substantial effort has been devoted to the study of the molecular pathobiology of Alzheimer's disease, with the identification of at least four chromosomal loci associated with familial cases; the degeneration of central neurochemical systems, especially basal forebrain structures related to acetylcholine-mediated neurotransmission; factors associated with the formation of plaques and tangles; and exogenous (e.g., infectious and toxic) processes that may contribute to the development of sporadic cases. Molecular biologists have sought to understand the formation of the abnormal amounts of amyloid that constitute the cerebral plaques characteristic of the disease. Although amyloid itself is a normal brain product, it has been suggested that excessive amounts may be neurotoxic. Others continue to see amyloid accumulations solely as a disease byproduct. Attention has recently turned to amyloid precursor protein and the intriguing possibility of regulating amyloid production pharmacologically. The discovery of an association between apolipoprotein E4, controlled by a gene located on chromosome 19, suggests further avenues for investigating risk factors and pathogenetic mechanisms. Taken together, these recent findings point to a heterogeneous array of pathobiological processes contributing to the final clinical and histological picture known as Alzheimer's disease. The postmortem and antemortem presentations appear to be relatively generalized (i.e., nonspecific) outcomes of widely divergent etiologies.

The natural course of dementia of the Alzheimer's type, as of all the degenerative disorders, is exacerbation and progression of clinical symptomatology. Brain degeneration as measured by in vivo imaging techniques such as MRI has not been found to correlate closely with the state of clinical disease. The final common clinical picture is of a bedridden patient, wholly dependent on others for all basic functions, even for turning in bed. Nutrition can often be provided only by nasogastric or gastrointestinal tubes. Death usually results from aspiration or from infectious processes associated with prolonged recumbency.

Parkinson's Disease Described by James Parkinson in 1817, Parkinson's disease is a prototype of a subcortical degenerative disease. It is idiopathic and must be distinguished from parkinsonian syndromes that arise from a variety of causes.

Parkinson's disease is the result of the degeneration of subcortical structures, primarily the substantia nigra but also the globus pallidus, putamen, and caudate (Fig. 10-11). Cells containing dopamine are predominantly affected, although serotonergic and other systems are disrupted as well. Just as the appellation “cortical pattern” is pseudoanatomical, so in subcortical Parkinson's disease there can be significant degeneration of cortical structures. The parkinsonian syndrome manifests with structural damage that reflects the underlying process or insult. Medication-induced parkinsonism presumably involves only a dysfunction of the basal ganglia structures, without any obvious pathoanatomical abnormality. The typical age at onset of Parkinson's disease is between 50 and 60 years but may vary widely, with onset sometimes occurring one to two decades earlier. The clinical course is chronic and progressive, with severe disability attained after approximately 10 years. A smaller proportion of patients have a more rapidly progressive disease, and a yet smaller group has a slowly progressive disorder in which deterioration plateaus or remains minimal for two to three decades.

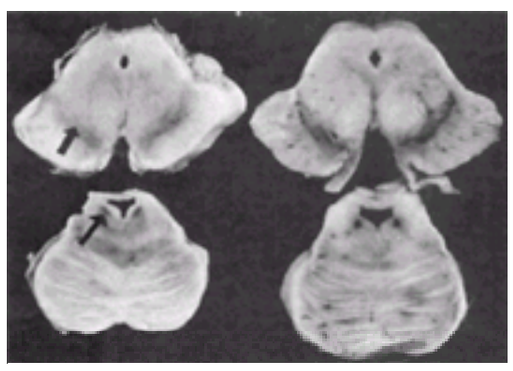

FIGURE 10-11 Parkinson's disease. Section of midbrain and pons showing depigmentation of substantia nigra and locus ceruleus in Parkinson's disease on left (arrows) and normal substantia nigra and locus ceruleus on right. (Reprinted with permission from Golden A, Powell DE, Jennings CD: *Pathology: Understanding Human Disease*, ed 2. Williams & Wilkins, Baltimore, 1985.)

In general, subcortical diseases are thought to impinge on the three Ms—movement, mentation, and mood. In Parkinson's disease all three of these areas are affected, although not always uniformly. The movement abnormalities are characterized by the triad of tremor, rigidity, and bradykinesia. The tremor and rigidity can be unilateral or bilateral. Bradykinesia is manifested by slowness in the initiation and execution of movement. The typical presentation, with a masklike facies, minimal blink, and monotonic speech, is a concomitant of the rigidity and slowness of movement. Other prominent characteristics include postural changes such as chin-to-chest flexion and gait abnormalities. The gait is characteristically slow and shuffling, and the patient has difficulty turning (en bloc turning) and trouble initiating and stopping walking. Seborrhea, sialorrhea, excessive fatigue, and constipation are also common.

Maurice Rosen was 69 when he made an appointment for a neurological evaluation. He had recently noticed that his memory was slipping and he had problems with concentration that were beginning to interfere with his work as a self-employed tax accountant. He complained of slowness and losing his train of thought. Recent changes in the tax laws were hard for him to learn, and his wife said he was becoming more withdrawn and reluctant to initiate activities. However, he was still able to take care of his personal finances and accompany his wife on visits to friends. Although mildly depressed about his disabilities, he denied other symptoms of depression, such as disturbed sleep or appetite, feelings of guilt, or suicidal ideation.

Mr. Rosen has a long history of treatment for episodes of depression, beginning in his 20s. He has taken a number of different antidepressants, and once had a course of electroconvulsive therapy. As recently as 6 months before this evaluation, he had been taking an antidepressant. Two years ago he developed an intermittent resting tremor in his left hand and a shuffling gait. Although the diagnosis of Parkinson's disease was considered by his psychiatrist, it was not confirmed by a neurologist, and therefore no additional treatment was given.

The neurologist who was now evaluating him found that his spontaneous speech was hesitant and unclear (dysarthric). Cranial nerve examination was normal. Motor tone was increased slightly in the neck and all limbs. Alternating movements of his hands were performed slowly. He had a slight intermittent tremor of the left arm at rest. Reflexes were symmetrical. A diagnosis of idiopathic Parkinson's disease was made, and he was placed on a low dose of carbidopa (Atamet), a medication that alleviates the symptoms of Parkinson's disease.

A neuropsychological examination performed 3 weeks later revealed average performance on the Wechsler Adult Intelligence Scale-Revised (full scale I.Q. = 104), but a verbal I.Q. of 118 and a performance I.Q. of 84. Memory as assessed by a 12-item, 10-trial, selective reminding task was poor, with no more than 7 items recalled on any trial, and only 3 words recalled after a 15-minute delay, although the patient could recognize the remaining words. He showed marked difficulty in drawings of overlapping figures and parallel lines. He was unable to draw three-dimensional figures. In language testing he demonstrated impaired naming. In summary, Mr. Rosen displayed evidence of impairment in memory, naming, and constructional abilities. These may have been secondary to slowness, poor planning, and perseveration. The deficits were believed to result from Parkinson's disease. Additional evaluation included an MRI, which revealed only generalized atrophy, and an EEG, which was significant for background generalized slowing. (Reprinted with permission from *DSM-IV Casebook*.)

Mentation or cognition in Parkinson's disease is an area of controversy. Most patients complain of slowed thinking, sometimes called bradyphrenia. In general, approximately 20 to 30 percent of patients with Parkinson's disease are found to have dementia, with the likelihood greater in those with late-onset disease (after age 70 years). Approximately 40 percent of nondemented patients with Parkinson's disease, however, demonstrate some neuropsychological impairment in most studies. The impairments are primarily in visuospatial capacities, as measured by copying, tracing, and tracking tasks, and in the shifting of cognitive sets, as measured by the Wisconsin Card Sorting Test or the Stroop Test. Such deficits have been noted in the absence of cognitive-based functional decline or other evidence of cognitive impairment. Controversy has emerged over whether these two patterns represent a single continuum of dementia integral to the process of Parkinson's disease or are two separate processes indicative of two distinct diseases. Neuropathologically, cases intermediate between Parkinson's disease and Alzheimer's disease exist, with the characteristic microscopic features of the latter and Lewy bodies in the substantia nigra suggesting the former. There is no clear line of division as yet between a process resembling dementia of the Alzheimer's type on which abnormal parkinsonian movements are superimposed and a clinical presentation of Parkinson's

disease in which the patient slowly develops a global progressive dementia.

Mood disorders have been frequently reported in association with Parkinson's disease. Depression is the most common; mania is virtually unreported. The mean frequency of depression is approximately 40 percent, with a reported range of 4 to 70 percent. No relation has been demonstrated between the frequency and severity of depression and the patient's current age, the age at onset of symptoms of Parkinson's disease, the duration of those symptoms, the severity of motor signs, or the response to medication. No relation has been demonstrated among mood, rigidity, bradykinesia, or tremor. Although depression has been found more commonly in patients with Parkinson's disease who display prominent gait and postural changes, the relation between mood and the severity of the disability is limited. There may be some association between depression and laterality of disease, for patients with left brain disease appear to have a higher frequency of depression than patients with right brain disease. This pattern suggests that the mood disturbance is a primary manifestation of brain deterioration and not a reactive psychological response to chronic illness and disability. Although the evidence relating lateralization to a higher frequency of depression is preliminary, it does recall data regarding poststroke depression and its putative relationship to left hemisphere localization but not to the extent of disability. The phenomenology of depression in Parkinson's disease is similar to that of idiopathic major depressive disorder, for it includes subjective dysphoria, pessimism, irritability, and suicidality, but perhaps less self-disparagement and self-blame. Some patients present with anxiety or panic attacks. Anergia, psychomotor retardation, and early-morning awakening are three symptoms that have been found to be nonspecific for depressive disorder in those patients, as they overlap considerably with the manifestations found in nondepressed patients with Parkinson's disease. Some data support the view that the depression of Parkinson's disease is associated specifically with decreased CNS serotonin concentrations. The on-off syndrome, in which patients experience severe fluctuations in mobility ranging from normal movement to a frozen state, has also been associated with changes in mood. On-off phenomena usually occur after years of chronic treatment with levodopa and can manifest as a between-dosage effect or randomly throughout the day. Many studies have reported changes in mood coincident with changes in motoric function, namely, subjective and objective dysphoria in the off period and, less frequently, abnormal elation and euphoria during the on period.

Psychosis as a concomitant of Parkinson's disease has been reported in the context of mood disorders (e.g., psychotic depression) or as a consequence of treatment. There are no reports of a specific personality change characteristic of Parkinson's disease except for the apparent apathy and lack of initiative that are often subsumed under bradykinesia and bradyphrenia.

The pharmacological treatment of Parkinson's disease addresses mood and movement because there is no known regimen for the improvement of cognition. For movement dysfunction dopamine precursors, such as levodopa or levodopa-carbidopa, (Sinemet), are a mainstay of treatment. Gait, posture, rigidity, and akinesia are generally more responsive to levodopa than is tremor. Anticholinergic agents and the dopamine agonists (bromocriptine [Parlodel] and pergolide [Permax]) are second-line agents. The monoamine oxidase (MAO) type B inhibitor selegiline (Eldepryl) has been demonstrated to apparently slow the progression of motor dysfunction, although in the low doses used it did not have significant antidepressant efficacy. Its effects on the development of cognitive impairment are unknown. Currently a new generation of dopamine receptor agonists is emerging; their long-term utility will become clear in the coming years. All symptomatic antiparkinsonian agents can cause delirium, a common iatrogenic concomitant of the disease. Levodopa has also been reported to cause visual hallucinations in some patients, even in the absence of delirium. Surgical treatments—stereotactic lesioning of the thalamus or globus pallidus—were used in the past to alleviate the motor dysfunction of Parkinson's disease. Although that approach had largely been replaced by pharmacological treatment, newer and more precise operative procedures have been developed in the past few years. Transplantation of fetal neural tissue into the caudate of the adrenal medulla also has been attempted, but no data from well-controlled studies are available regarding the effects of the procedure on mood and mentation. Recent findings suggest that fetal tissue transplantation may dramatically alleviate severe motor symptoms. Many differences exist in surgical protocols and transplantation methods, underscoring the highly experimental nature of these procedures.

Treatment of the mood disorder associated with Parkinson's disease involves the same agents that have proved valuable in treating idiopathic major depressive order. Antidepressant medication from all categories have proved efficacious. ECT is of value for treating both the mood component and the motor dysfunction; dramatic improvement in all aspects of movement has been demonstrated on standardized neurological examinations. Several studies have reported sustained improvement in motor function for as long as 6 months after treatment; however, most detected a short-lived improvement of days to weeks. ECT is recommended for patients with Parkinson's disease and the on-off syndrome, particularly when significant mood changes are present.

In summary, Parkinson's disease is a prototypical subcortical pattern degenerative disease. The overlap of clinical phenomena between a basal ganglia disease, such as Parkinson's disease, and major depressive disorder can be striking. Both are characterized by qualitatively similar impairments in the realms of movement, mentation, and mood. Differing terminologies have arisen to describe similar signs and symptoms in each. *Psychomotor slowing (psychomotor retardation)* a term used to encompass both the motoric and cognitive slowing seen in depression, is quite similar to the bradykinesia and bradyphrenia described early in the course of Parkinson's disease. A recent study that used a nonmotor measure of bradyphrenia demonstrated close correlations between cognitive slowing and severity of the mood disorder both in depressed patients with Parkinson's disease and patients with idiopathic major depressive disorder, suggesting a close phenomenological relation between the bradyphrenia of Parkinson's disease and so-called psychomotor slowing. It underscored the idea that basal ganglia disorders are fertile ground for research and insight into the neurobiological bases of idiopathic mood disorders.

Dementia With Lewy Bodies Since the late 1980s research has revealed that, beyond dementia of the Alzheimer's type and vascular dementia, a common cause of progressive dementia may be related to the presence of Lewy bodies in the brainstem and cerebral cortex. Lewy bodies—intracytoplasmic, spherical, eosinophilic neuronal inclusion bodies—are scattered through the brainstem, subcortical nuclei, limbic cortex (cingulate, entorhinal, amygdala), and neocortex (temporal > frontal = parietal). Parkinson's disease, in contrast, manifests Lewy bodies in subcortical nuclei, in addition to degeneration of dopamine cell bodies in substantia nigra. Table 10-18 lists the pathological features of dementia with Lewy bodies; Table 10-19 includes recently developed consensus guidelines for clinical diagnosis. Neuropsychiatric features, including visual hallucinations, delusions, fluctuating attention, and executive or managerial cognitive deficits, are prominent; although not specific, mood disturbances are common.

Essential for diagnosis
Lewy bodies
Associated but not essential
Lewy-related neurites
Plaques (all morphological types)
Neurofibrillary tangles
Regional neuronal loss—especially brainstem (substantia nigra and locus coeruleus) and nucleus basalis of Meynert
Microvacuolation (spongiform change) and synapse loss
Neurochemical abnormalities and neurotransmitter deficits

Reprinted with permission from McKeith IG, Galasko D, Kosaka K, et al: For the Consortium on Dementia with Lewy Bodies. Neurology 47:1113, 1996.

Table 10-18 Pathological Features Associated With Dementia With Lewy Bodies

1. The central feature required for a diagnosis of dementia with Lewy bodies is progressive cognitive decline of sufficient magnitude to interfere with normal social or occupational function. Prominent or persistent memory impairment may not necessarily occur in the early stages but is usually evident with progression. Deficits on tests of attention and of frontal-subcortical skills and visuospatial ability may be especially prominent.
2. Two of the following core features are essential for a diagnosis of probable dementia with Lewy bodies, and one is essential for possible dementia with Lewy bodies:
a. Fluctuating cognition with profound variations in attention and alertness
b. Recurrent visual hallucinations that are typically well formed and detailed
c. Spontaneous motor features of parkinsonism
3. Features supportive of the diagnosis are:
a. Repeated falls
b. Syncope
c. Transient loss of consciousness
d. Neuroleptic sensitivity
e. Systematized delusions
f. Hallucinations in other modalities
4. A diagnosis of dementia with Lewy bodies is less likely in the presence of:
a. Stroke disease, evident as focal neurologic signs or on brain imaging
b. Evidence on physical examination and investigation of any physical illness or other brain disorder sufficient to account for the clinical picture

Reprinted with permission from McKeith IG, Galasko D, Kosaka K, et al: For the Consortium on Dementia with Lewy Bodies. Neurology 47:1113, 1996.

Table 10-19 Consensus Criteria for the Clinical Diagnosis of Probable and Possible Dementia With Lewy Bodies

Substance-Induced Disorders Pharmacological compounds are potent and frequent causes of psychopathology. This effect, especially when caused by environmental and occupational neurotoxins, has been little studied; more attention has been paid to peripheral or motor effects than to the less easily quantified behavioral alterations. Although alcohol-induced neuropsychiatric syndromes have long been known of, the CNS consequences of abuse of recreational drugs received attention only in recent decades, in part because of societal disapproval of recreational drug use, disinterest among investigators, and the inherent difficulty of separating drug effects from confounding person effects.

Broadly speaking, there are four classes of chemically induced psychopathology: (1) that due to environmental contamination, both natural (e.g., venoms and poisonous foods) and human-made (e.g., gasoline contamination of well water); (2) that due to occupational exposure; (3) that due to recreational use, abuse, or dependence on substances causing transient or lasting CNS toxic effects; and (4) the iatrogenic complications of prescribed or over-the-counter medications. Compounds must be considered from the perspectives of (1) acute or immediate effects (e.g., behavioral symptoms of acute intoxication), (2) longer-term responses to persistent exposure, and (3) lasting consequences that persist after the cessation of any direct pharmacological action. The last may be especially complex because of the extremely prolonged retention (months to years) of some compounds within the body.

When considering the possibility of drug- or toxin-induced psychopathology, the clinician must undertake a careful chain of reasoning, akin to deciding on any secondary diagnosis but differing in several respects. Initially, the clinician must ascertain whether an exposure occurred and at what level; for example, an industrial hygienist may have been exposed in a possible occupational incident. Next, it is critical to understand the toxicity of a substance (especially as it might relate to different chemical forms), its mode of action (when known), its effects in various animal species, and its clinical manifestations. Typically, these issues are within the realm of toxicologists. Subsequently, the clinician seeks to define the systemic clinical manifestations of the exposure in the particular patient. Although behavioral changes may be the only exposure-related findings, more often there is a variety of consistent symptoms, signs, and laboratory findings that together make a coherent clinical picture. It is within that larger context that the clinician views any presumptively related psychopathology.

When attempting to establish a neurotoxic diagnosis, physicians also must recognize other cardinal features of toxic exposures that influence clinical reasoning: (1) neurotoxic chemicals often cause nonfocal/nonspecific syndromes; (2) standard laboratory tests may have limited diagnostic utility; (3) there usually are strong dose-response relationships; (4) a single toxin may produce multiple syndromes; (5) with a few specific exceptions (e.g., asbestos, therapeutic cerebral radiation exposure), toxin-induced illnesses occur concurrently with exposure or following a short latency period; (6) the chemical formula may not predict toxicity; (7) a compound without known neurotoxic activity can interact or potentiate the effects of known neurotoxins; and (8) neurotoxic disease often may be asymptomatic.

Ideally, one would like to know the neuropsychiatric effects of all CNS toxic compounds. In the absence of such information, the clinician must describe symptoms and signs in detail and compare them with available data. The clinician defines the temporal course of exposure and assesses how the emergence of specific psychiatric manifestations relates to known actions of the compounds in question. Simultaneously, the clinician must consider the form of the disorder and establish whether it suggests a pathological CNS process or is more consistent with primary (idiopathic) psychopathology. The clinician must decide whether the syndrome in question might reflect other unrelated disease processes as well. Any measurable clinical and laboratory manifestations of CNS disease should be identified. Although no single measurement can be considered proof, taken together such measurements offer the possibility of establishing a diagnosis with a high degree of clinical certainty.

The array of environmental and occupational compounds to which people may be exposed is large. Except for patients exposed to recreational and iatrogenic agents, until recently most psychiatrists did not treat patients with toxic exposures. This is changing rapidly as a result of late twentieth-century technology and increasing societal awareness. The clinician should remember that toxic exposures are used by some patients to explain a pantheon of personal ills, many of a psychiatric nature. It is neither appropriate to treat those complaints lightly nor clinically sound to accept such pathogenic explanations without firm clinical support.

Recreational drugs may be used or abused intentionally to cause dose-dependent behavioral changes, including anxiolysis with nicotine, intoxication with alcohol or marijuana, or psychosis with mescaline. There may be additional unwanted psychiatric phenomena with drug intoxication. Unwanted secondary syndromes commonly described include anxiety and insomnia due to caffeine; paranoid psychosis with cocaine; mood alterations (dysphoria, anxiety, euphoria) accompanying the perceptual disturbances induced by hallucinogenic substances; agitated (often violent) psychotic states with phencyclidine (PCP); and depressive symptoms and seemingly paradoxical disinhibition of aggressive impulses with sedative-hypnotic agents. However, there may be considerable variability and therefore lack of specificity in syndromic association with particular substances. Also, many substances in sufficient doses can cause delirium, which may itself have associated psychiatric symptoms ranging from mood disturbance to psychosis.

Withdrawal syndromes are also commonly encountered with drugs of abuse and tend to be characteristic of the class of drug. Withdrawal from nicotine may produce anxiety or irritability; withdrawal from stimulants produces a hypersomnic dysphoric crash; and withdrawal from opiates produces a well-described state that includes malaise, anxiety and irritability, drug craving, insomnia, psychomotor agitation, anorexia, and a variety of physical symptoms (e.g., diarrhea, piloerection, mydriasis, hypertension, and tachycardia). Delirium often follows withdrawal from alcohol and sedative-hypnotic medications but is not a component of other drug withdrawal syndromes (e.g., opioid-related).

Numerous medications have been implicated in causing psychiatric phenomena. Prescription drugs and over-the-counter preparations may cause physiologically induced behavioral changes, either through intoxication (which may involve use at therapeutic or supratherapeutic levels) or withdrawal. As with other causes of secondary psychiatric disorders, combinations of medications and medical illnesses may cause behavioral changes even when each medication or illness alone does not.

Among prescribed medications, psychotropic drugs are designed to effect behavioral changes. Unwanted psychiatric syndromes, such as antipsychotic-induced depressive syndromes or the delirium of lithium toxicity, occur often. Countless medications have been implicated in secondary psychiatric syndromes; it is rare indeed for a medication to be listed in the *Physicians' Desk Reference* without an accompanying description of some potential neuropsychiatric adverse effect. However, this information must be interpreted with caution. Behavioral adverse effects are also noted with placebos; therefore, distinguishing physiological from psychological symptoms may be difficult. Psychiatric symptoms may also reflect the clinical manifestations of the primary illness being treated (e.g., delirium in a patient receiving a new parenteral antibiotic may be due to the antibiotic or the targeted infection). Finally, numerous other factors may complicate the process of establishing the etiological significance of a particular medication. For example, b-adrenergic-receptor antagonists have been postulated for many years to cause depressive syndromes, yet a recent large study, carefully controlled for patient demographics, medical illness, medications, and other factors, was unable to find a significant independent association between b-adrenergic receptor antagonists and depression. Despite these caveats, many medications clearly can cause secondary psychiatric syndromes.

COGNITIVE DISORDERS

Delirium Delirium, a transient disorder of brain function manifested by global cognitive impairment and other behavioral phenomena, is a common disease state that has been described for centuries. Nevertheless, it is frequently missed or misdiagnosed, with the potential for substantial attendant morbidity and mortality. Recognition and appropriate evaluation and treatment of delirium should be an imperative, not just for psychiatrists but for all physicians.

Definition DSM-IV includes delirium under cognitive disorders. Delirium is a syndrome, with core features of impairment of consciousness with attentional deficit, other cognitive alterations, and a relatively rapid onset of the disorder with a characteristically fluctuating course. Frequently there are other associated clinical phenomena, which may appear more prominent to the uneducated observer than the core features.

History

Physicians have long recognized states of altered behavior, including changes in level of consciousness, of acute onset that were associated with fever, poisons, or other medical or neurological diseases. There are references to such presentations in the writings of Hippocrates and in much subsequent Greco-Roman literature. Descriptions of the syndrome similar to modern definitions appear from the late Middle Ages through the eighteenth century. The history, however, is obscured by an etymological web that to this day impedes communication and education about the disorder. Numerous terms have been used to describe the syndrome of delirium, including *phrenitis, frenzy*, and *febrile insanity*; conversely, the term “delirium” has also been applied to other psychiatric states that led to insanity.

By the nineteenth century emphasis was placed on disordered consciousness as the hallmark of delirium. The phrase “clouding of consciousness” dates to that time

and is still used in many quarters today despite lack of clarity as to what it means. Similarly, the term “confusion” was used frequently, despite the lack of a specific relation to delirium.

The work of George Engel and John Romano in the 1940s, summarized by them in publications from the 1950s, indicated that attentional and other cognitive disturbances were best viewed as the core features of the syndrome and that the state was associated with acute brain failure, as demonstrated by slowing on the EEG. Subsequent work on the pathophysiology of delirium has been relatively scant. Zbigniew Jerzy Lipowski, beginning in the 1960s and continuing to the present, has been instrumental in raising clinical and research awareness of delirium, defining the syndrome according to strict criteria and popularizing (especially in the psychiatric community) the use of the term “delirium.” Recent years have seen alterations in diagnostic criteria, as evidenced by the removal of associated clinical features such as psychomotor changes from the required criteria. There has also been increasing study of epidemiology, clinical course, and risk factors for onset or poor outcome.

Unfortunately, etymological confusion remains. Numerous synonyms remain in common use, especially in nonpsychiatric medical fields; some of them are encephalopathy, acute confusional state, and acute organic brain syndrome. Some neurologists maintain a distinction between delirium, which they reserve to describe extremely agitated delirious states with frank thought process disorganization, perceptual disturbances, and autonomic hyperactivity, and acute confusional states, which they use to describe all other, often less severe delirious states. Most psychiatrists, and many other workers in the field, believe that such distinctions are premature at best (because of a lack of evidence of differing causes or pathophysiologies between the two) and misleading at worst, obscuring the commonality of core clinical features, potential etiologies, and management approaches.

Epidemiology There have been relatively few studies of the incidence and prevalence of delirium. Little is known about the epidemiology of delirium in community or other nonpatient, noninstitutionalized populations. An estimated 10 to 15 percent of general medical inpatients are delirious at any given time, and studies indicate that as many as 30 to 50 percent of acutely ill geriatric patients become delirious at some point during their hospital stay. Rates of delirium in psychiatric and nursing home populations are not well established but are clearly substantial. Risk factors for the development of delirium include increased severity of physical illness, older age, and baseline cognitive impairment (e.g., due to dementia).

Delirium is frequently unrecognized by treating physicians. Because of its wide array of associated symptoms, it may be detected but misdiagnosed as depression, schizophrenia, or other psychiatric disorder. Delirium is a frequent cause for psychiatric consultation in the general hospital but often is not recognized as such by the referring physician.

Etiology The syndrome of delirium reflects brain dysfunction that is almost always due to identifiable systemic or cerebral disease or to drug intoxication or withdrawal. A partial list of frequently encountered causes is given in Table 10-20. Often delirium is due to multiple simultaneous causes, each one of which may or may not be enough to cause delirium by itself. On rare occasions a syndrome nearly indistinguishable from delirium may manifest as part of the course of another Axis I disorder such as bipolar I disorder.

Drug intoxication
Anticholinergics
Lithium
Antiarrhythmics (e.g., lidocaine)
H_2-receptor blockers
Sedative-hypnotics
Alcohol
Drug withdrawal
Alcohol
Sedative-hypnotics
Tumor
Primary cerebral
Trauma
Cerebral contusion (as an example)
Subdural hematoma
Infection
Cerebral (e.g., meningitis, encephalitis, HIV, syphilis)
Systemic (e.g., sepsis, urinary tract infection, pneumonia)
Cardiovascular
Cerebrovascular (e.g., infarcts, hemorrhage, vasculitis)
Cardiovascular (e.g., low-output states, congestive heart failure, shock)
Physiological or metabolic
Hypoxemia, electrolyte disturbances, renal or hepatic failure, hypo- or hyperglycemia, postictal states (as examples)
Endocrine
Thyroid or glucocorticoid disturbances (as examples)
Nutritional
Thiamine or vitamin B_{12} deficiency, pellagra (as examples)

Table 10-20 Causes of Delirium

Diagnosis and Clinical Features The syndrome of delirium is almost always caused by one or more systemic or cerebral derangements that affect brain function.

A 74-year-old African-American woman, Ms. Richardson, was brought to a city hospital emergency room by the police. She is unkempt, dirty, and foul smelling. She does not look at the interviewer and is apparently confused and unresponsive to most of his questions. She knows her name and address, but not the day or the month. She is unable to describe the events that led to her admission.

The police reported that they were called by neighbors because Ms. Richardson had been wandering around the neighborhood and not taking care of herself. The medical center mobile crisis unit went to her house twice, but could not get in and presumed she was not home. Finally, the police came and broke into the apartment, where they were met by a snarling German shepherd. They shot the dog with a tranquilizing gun, and then found Ms. Richardson hiding in the corner, wearing nothing but a bra. The apartment was filthy, the floor was littered with dog feces. The police found a gun, which they took into custody.

The following day, while Ms. Richardson was awaiting transfer to a medical unit for treatment of her out-of-control diabetes, the supervising psychiatrist attempted to interview her. Her facial expression was still mostly unresponsive, and she still didn't know the month and couldn't say what hospital she was in. She reported that the neighbors had called the police because she was “sick,” and indeed she had felt sick and weak, with pains in her shoulder; in addition, she had not eaten for 3 days. She remembered that the police had shot her dog with a tranquilizer, and said the dog was not in “the shop” and would be returned to her when she got home. She refused to give the name of a neighbor who was a friend, saying, “he's got enough troubles of his own.” She denied ever being in a psychiatric hospital or hearing voices, but acknowledged that she had at one point seen a psychiatrist “near Lincoln Center” because she couldn't sleep. He had prescribed medication that was too strong, so she didn't take it. She didn't remember the name, so the interviewer asked if it was Thorazine. She said no, it was “allal.” “Haldol?” asked the interviewer. She nodded. The interviewer was convinced that was the drug, but other observers thought she might have said yes to anything that sounded remotely like it, such as “Elavil.” When asked about the gun, she denied, with some annoyance, that it was real and said it was a toy gun that had been brought to the house by her brother, who had died 8 years ago. She was still feeling weak and sick, complained of pains in her shoulder, and apparently had trouble swallowing. She did manage to smile as the team left her bedside. (Reprinted with permission from *DSM-IV Casebook*.)

DSM-IV gives separate diagnostic criteria for delirium due to a general medical condition (Table 10-21), for delirium related to systemic medical conditions or primary cerebral conditions, substance intoxication delirium (Table 10-22), substance withdrawal delirium (Table 10-23), delirium due to multiple etiologies (Table 10-24), and delirium not otherwise specified (Table 10-25) for a delirium of unknown cause or due to causes not listed, such as sensory deprivation. However, the core syndrome is the same, regardless of cause.

A. Disturbance of consciousness (i.e., reduced clarity of awareness of the environment) with reduced ability to focus, sustain, or shift attention.
B. A change in cognition (such as memory deficit, disorientation, language disturbance) or the development of a perceptual disturbance that is not better accounted for by a pre-existing, established, or evolving dementia.
C. The disturbance develops over a short period of time (usually hours to days) and tends to fluctuate during the course of the day.
D. There is evidence from the history, physical examination, or laboratory findings that the disturbance is caused by the direct physiological consequences of a general medical condition.
Coding note: Include the name of the general medical condition on Axis I, e.g., delirium due to hepatic encephalopathy; also code the general medical condition on Axis III.

Table 10-21 DSM-IV Diagnostic Criteria for Delirium Due to a General Medical Condition

A. Disturbance of consciousness (i.e., reduced clarity of awareness of the environment) with reduced ability to focus, sustain, or shift attention.
B. A change in cognition (such as memory deficit, disorientation, language disturbance) or the development of a perceptual disturbance that is not better accounted for by a pre-existing, established, or evolving dementia.
C. The disturbance develops over a short period of time (usually hours to days) and tends to fluctuate during the course of the day.
D. There is evidence from the history, physical examination, or laboratory findings of either (1) or (2):
(1) the symptoms in criteria A and B developed during substance intoxication
(2) medication use is etiologically related to the disturbance
Note: This diagnosis should be made instead of a diagnosis of substance intoxication only when the cognitive symptoms are in excess of those usually associated with the intoxication syndrome and when the symptoms are sufficiently severe to warrant independent clinical attention.
Note: The diagnosis should be recorded as substance-induced delirium if related to medication use.
Code: (Specific substance) intoxication delirium (Alcohol; amphetamine (or amphetamine-like substance); cannabis; cocaine; hallucinogen; inhalant; opioid; phencyclidine (or phencyclidine-like substance); sedative, hypnotic, or anxiolytic; other (or unknown) substance (e.g., cimetidine, digitalis, benztropine))

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-22 DSM-IV Diagnostic Criteria for Substance Intoxication Delirium

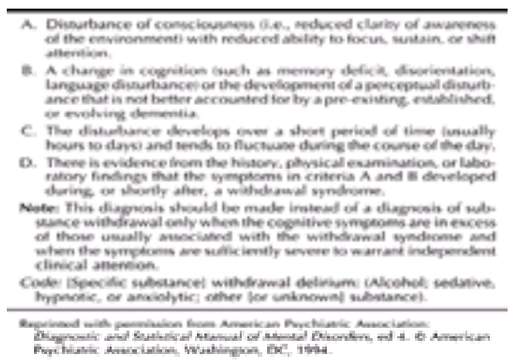

A. Disturbance of consciousness (i.e., reduced clarity of awareness of the environment) with reduced ability to focus, sustain, or shift attention.
B. A change in cognition (such as memory deficit, disorientation, language disturbance) or the development of a perceptual disturbance that is not better accounted for by a pre-existing, established, or evolving dementia.
C. The disturbance develops over a short period of time (usually hours to days) and tends to fluctuate during the course of the day.
D. There is evidence from the history, physical examination, or laboratory findings of either (1) or (2):
(1) the symptoms in criteria A and B developed during, or shortly after, a withdrawal syndrome.
(2) medication use is etiologically related to the disturbance
Note: This diagnosis should be made instead of a diagnosis of substance withdrawal only when the cognitive symptoms are in excess of those usually associated with the withdrawal syndrome and when the symptoms are sufficiently severe to warrant independent clinical attention.
Code: (Specific substance) withdrawal delirium; (Alcohol); sedative, hypnotic, or anxiolytic; other (or unknown) substance).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-23 DSM-IV Diagnostic Criteria for Substance Withdrawal Delirium

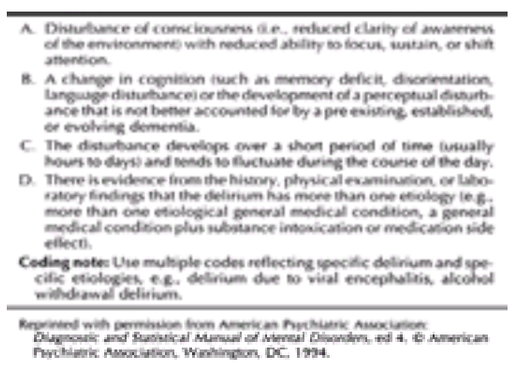

A. Disturbance of consciousness (i.e., reduced clarity of awareness of the environment) with reduced ability to focus, sustain, or shift attention.
B. A change in cognition (such as memory deficit, disorientation, language disturbance) or the development of a perceptual disturbance that is not better accounted for by a pre existing, established, or evolving dementia.
C. The disturbance develops over a short period of time (usually hours to days) and tends to fluctuate during the course of the day.
D. There is evidence from the history, physical examination, or laboratory findings that the delirium has more than one etiology (e.g., more than one etiological general medical condition, a general medical condition plus substance intoxication or medication side effect).
Coding note: Use multiple codes reflecting specific delirium and specific etiologies, e.g., delirium due to viral encephalitis, alcohol withdrawal delirium.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-24 DSM-IV Diagnostic Criteria for Delirium Due to Multiple Etiologies

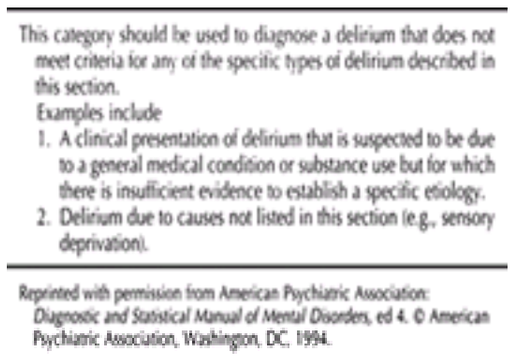

This category should be used to diagnose a delirium that does not meet criteria for any of the specific types of delirium described in this section.
Examples include
1. A clinical presentation of delirium that is suspected to be due to a general medical condition or substance use but for which there is insufficient evidence to establish a specific etiology.
2. Delirium due to causes not listed in this section (e.g., sensory deprivation).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-25 DSM-IV Diagnostic Criteria for Delirium Not Otherwise Specified

The core features of delirium include altered consciousness, such as decreased level of consciousness; altered attention, which may include diminished ability to focus, sustain, or shift attention; impairment in other realms of cognitive function, which may manifest as disorientation (especially to time and space) and decreased memory; relatively rapid onset (usually hours to days); brief duration (usually days to weeks); and often marked, unpredictable fluctuations in severity and other clinical manifestations during the course of the day, sometimes worse at night (sundowning), which may range from periods of lucidity to quite severe cognitive impairment and disorganization.

Associated clinical features are often present and may be prominent. They may include disorganization of thought processes (ranging from mild tangentiality to frank incoherence), perceptual disturbances such as illusions and hallucinations, psychomotor hyperactivity and hypoactivity, disruption of the sleep-wake cycle (often manifested as fragmented sleep at night, with or without daytime drowsiness), mood alterations (from subtle irritability to obvious dysphoria, anxiety, or even euphoria), and other manifestations of altered neurological function (e.g., autonomic hyperactivity or instability, myoclonic jerking, and dysarthria). The EEG usually shows diffuse slowing of background activity, although patients with delirium due to alcohol or sedative-hypnotic withdrawal have low-voltage fast activity.

ICD-10 takes a somewhat different approach to establishing the diagnosis of delirium (Table 10-26). It requires the concurrent presence of disturbances in consciousness and attention, perception, thinking, psychomotor behavior, emotion, and the sleep-wake cycle. All features must be present to some degree for a definite diagnosis, thus making for a more restrictive classification than DSM-IV. The key to diagnosing delirium is to maintain a heightened suspicion for the syndrome whenever a patient experiences a relatively rapid change in or the new onset of any psychiatric symptom or sign. Once the diagnosis is suspected, the history (usually obtained from informants such as family, nursing staff, and prior treaters) and mental status examination can elucidate the cognitive disturbances at the core of the syndrome and uncover associated clinical phenomena that may affect management or suggest the etiology.

Table 10-26 ICD-10 Diagnostic Criteria for Delirium, Not Induced by Alcohol and Other Psychoactive Substances

The presence of delirium should prompt careful investigation for contributing causes. Delirium may be the first, most prominent, or only clinical manifestation of the new onset of a medical condition or the worsening of a previously diagnosed illness. A careful medical history (including medication and drug history), physical examination, and neurological examination must be undertaken, and various laboratory tests, neuroimaging procedures, lumbar puncture, and EEG may be useful. In searching for cause, physicians should remember that relatively minor abnormalities (e.g., a mild anemia plus slight hyponatremia plus slight hypercalcemia) may additively produce delirium even if each abnormality alone would not normally do so. Sometimes the search for an etiology does not yield a clear cause of the delirium; the patient still has the syndrome of delirium, however, and vigilance for clinical contributing factors must be maintained. Physicians should avoid the common mistake of believing delirium to be ruled out by the lack of obvious etiology, thereby proving that the behavioral disturbance is functional in origin. EEG may be helpful in such cases by demonstrating diffuse brain dysfunction, although that in itself does not demonstrate etiology.

Current understanding of the pathophysiology of delirium is limited. Dysfunction of the reticular activating system has been speculated, given its role in arousal. There is evidence for hypofunction of cholinergic systems, particularly in the basal forebrain and pons. There is some evidence for dysfunction of several other neurochemical systems, including noradrenergic, GABAergic, and serotonergic; more undoubtedly await investigation. Earlier speculation about globally decreased cerebral metabolism has not been confirmed, but also has not been carefully studied in delirium despite the availability of techniques such as PET and SPECT. Even more obscure is the pathophysiological link between specific systemic conditions and delirium. The classic model of such a link is anticholinergic drug toxicity, which has been presumed to cause delirium as a direct consequence of hypoactivity of cholinergic systems. Some recent work has demonstrated increased GABAergic transmission, putatively because of increased concentrations of endogenous benzodiazepine-like substances, in patients with delirium and fulminant liver failure.

Differential Diagnosis Much attention has been given to differentiating delirium from dementia. Usually that distinction can be easily made by noting temporal factors (course of onset and progression of the disturbance) and by recognizing that level of consciousness and attention are affected prominently and early in delirium. Dementia by definition does not involve an alteration of consciousness, although attentional dysfunction develops as the syndrome progresses in severity. However, delirium is often superimposed on a pre-existing dementia. If the history is unknown and mental status examination results are lacking in a patient with severe dementia, it can be difficult to tell if there is a new delirium, or if the delirium has resolved and the patient is back to baseline (which may be a new baseline reflecting deterioration from a previous level of function). In such cases it is prudent to assume that the patient has delirium and to proceed with a careful clinical evaluation.

Although thought process disorganization, perceptual disturbances, or mood symptoms may lead the uninitiated to diagnose idiopathic psychiatric illness, the constellation of altered level of consciousness, prominent attentional and other cognitive deficits, and temporal course usually makes the differentiation of delirium from mood, psychotic, and anxiety disorders straightforward. The previous psychiatric history can be helpful, but the clinician must use care in interpreting it because patients with chronic psychiatric illness are also at risk for developing delirium due to medications, drug abuse, or other conditions. Rarely, patients with other Axis I illnesses (particularly the schizophrenias and bipolar disorders) may develop flagrantly disorganized, incoherent states with obvious attentional impairment and the examiner may be unable to test their other cognitive functions. (When found in the course of bipolar disorder, this state has been incorrectly called manic delirium.) Such states cannot be reliably distinguished phenomenologically from delirium due to the more usual medical causes, and they warrant the same thorough search for contributing etiologies accorded other deliria.

Course and Prognosis By most definitions, although not by DSM-IV criteria, delirium is a transient condition. For most patients the syndrome resolves within days to a few weeks. However, in sicker populations the mortality associated with delirium is high in the short term (acute hospitalization) and increases with several months of follow-up. It is not clear if increased mortality is independently associated with delirium or if it can be accounted for by known medical pathology. In some patients an apparently new dementia becomes evident on resolution of the delirium; the dementia may not have been present or may have been present but unrecognized prior to the delirium.

Treatment The primary treatment of delirium is to identify and ameliorate any causal or contributing medical conditions. As part of that effort, the dosages of all sedatives and other CNS-active medications should be minimized as much as possible. (The exception is sedative-hypnotic or alcohol withdrawal delirium, in which treatment of the underlying problem requires the administration of a cross-tolerant agent such as a benzodiazepine.) Delirious patients may need extra supportive physical care; maintenance of basic functions such as food and fluid intake is crucial to rapid recovery. Keeping the patient in an environment that is quiet and free of unnecessary stimulation may help reduce agitation. Frequent cues to orientation may also be helpful. Supportive contacts with the patient, family, and sometimes staff members are necessary to reassure the patient that the new, often frightening behavioral state reflects physical illness and that the patient is not going crazy. Attention may need to be paid to the patient's legal capacity to participate in informed clinical care decisions.

The patient with a quiet, hypoactive delirium needs no specific pharmacotherapy. However, many delirious patients show persistent or intermittent psychomotor agitation that may interfere with nursing care or necessary tests and procedures. Control of the agitation is essential to prevent inadvertent self-damage and allow appropriate evaluation and treatment. Physical restraints may be used transiently when necessary. If sedation is desired, the drug of choice is a high-potency antipsychotic agent in relatively low dosages (e.g., haloperidol 0.5 to 1 mg orally or parenterally, up to several mg a day). Low-potency agents, benzodiazepines, and other sedatives (antihistamines, barbiturates) should generally be avoided because they are likely to worsen the delirious state. At times of severe, life-threatening agitation (e.g., if a patient in the intensive care unit is removing the endotracheal tube, arterial lines, and so forth), sedation at nearly any cost becomes necessary, and combinations of antipsychotic agents, benzodiazepines, and opioids have been used, as have neuromuscular-blocking agents, such as pancuronium (Pavulon), use of which depends on the availability of adequate ventilatory support).

There have been case reports of improvement in or remission of delirious states due to intractable medical illnesses with ECT. Although ECT may rarely be advised by a consultant with expertise in the procedure, routine consideration of ECT for delirium is not advised.

Dementia Interest in the study and care of patients with dementia has increased, coincident with the proportional increase of the elderly in the population. Although dementing disorders are defined by their multiple cognitive deficits, patients can present with the full array of psychiatric symptoms. And although dementia is most often associated with progressive processes, it does not by itself denote a deteriorating course. Thus, the clinician must seek any curable or treatable causes of dementia whenever it is recognized clinically, before irreversible CNS changes supervene.

Definition Dementia is a diminution in cognition in the setting of a stable level of consciousness. Dementia denotes a decrement of two or more intellectual functions, in contrast to focal or specific impairments such as amnestic disorder or aphasia. The persistent and stable nature of the impairment distinguishes dementia from the altered consciousness and fluctuating deficits of delirium. Dementia must also be distinguished from long-standing mental subnormality, as the former represents an acquired loss of or decline in prior intellectual and functional capacities.

History

Dementia has long been understood as describing an acquired cognitive and behavioral decline associated with brain disease. Jean Étienne Dominique Esquirol, in his classic, early nineteenth-century nosological work, *Mental Maladies: A Treatise on Insanity*, provided perhaps the first modern definition of dementia: “A cerebral affection usually chronic . . . and characterized by a weakening of the sensibility, understanding, and will.” In his study of over 300 patients, Esquirol described the noncognitive symptoms of dementia, reporting hallucinations, delusions, aggressive behavior, and motor abnormalities in many of the patients. Interestingly, however,

he included among the causes of dementia not only aging, head trauma, syphilis, and alcohol abuse, but also conditions such as “menstrual disorders, . . . onanism, . . . disappointed affections, . . . and political shocks.” Later investigators described neuropathological correlations for the dementia syndromes, firmly establishing the relation between brain disease and dementia. Contemporary interest has focused again on an etiological basis for the observed pathological and pathophysiological abnormalities and on risk factors, preventive measures, and specific treatments for dementia.

Comparative Nosology In DSM-III and DSM-III-R, dementia was listed as both a syndrome and a disorder. The development of specific criteria for the symptom constellation was a major departure from all previous nosologies. It proved to be a major conceptual advance for clinical practice and research. The dementia syndrome was one of the possible presentations of psychoactive substance-induced organic mental conditions and of organic mental conditions associated with Axis III physical disorders. Dementia was also listed as a group of specific disorders, including primary degenerative dementia of the Alzheimer's type, multi-infarct dementia, dementia associated with alcoholism, and dementia not otherwise specified. A severity scale of mild, moderate, or severe was provided. DSM-IV eliminates the distinction between dementia as a syndrome and dementia as a disorder. Instead, it delineates those dementing disorders that are related to specific systemic medical or cerebral conditions (e.g., dementia of the Alzheimer's type and vascular dementia). DSM-IV criteria emphasize the defining features of dementia, namely the multiple deficits that represent a decline from a previously attained level of functioning, and incorporate specific information for distinguishing the etiological subcategories from each other, relying on course of the disease, the presence or absence of focal neurological signs and symptoms, laboratory evidence of neurological damage, a history of significant substance abuse, or other evidence of a contributing medical condition. Dementia of the Alzheimer's type is a diagnosis of exclusion, requiring that other potentially etiological CNS or systemic medical conditions be ruled out.

Beyond DSM-IV, there are alternative, conceptually overlapping systems for diagnosing dementia. ICD-10, in contrast to DSM-IV, maintains the approach adopted in DSM-III and DSM-III-R, with a general syndromic definition, which is then applied to specific disorders; for example, dementia in Alzheimer's disease or dementia in Huntington's disease. ICD-10 defines *dementia* as a syndrome in which:

> there is a disturbance of multiple higher cortical functions, including memory, thinking, orientation, comprehension, calculation, learning capacity, language, and judgement. Consciousness is not clouded. Impairments of cognitive function are commonly accompanied, and occasionally preceded, by deterioration in emotional control, social behaviour, or motivation . . . In assessing the presence or absence of a dementia, special care should be taken to avoid false-positive identification: motivational or emotional factors, particularly depression, in addition to motor slowness and general physical frailty, rather than loss of intellectual capacity, may account for failure to perform.

ICD-10 also states:

> Dementia produces an appreciable decline in intellectual functioning, and usually some interference with personal activities of daily living, such as washing, dressing, eating, personal hygiene, excretory and toilet activities. How such a decline manifests itself will depend largely on the social and cultural setting in which the patient lives. Changes in role performance, such as lowered ability to keep or find a job, should not be used as criteria of dementia because of the large cross-cultural differences that exist in what is appropriate, and because there may be frequent, externally imposed changes in the availability of work within a particular culture.

This latter statement differs fundamentally from DSM-IV. It underscores an unresolved controversy in the clinical and scientific literature where, in the absence of diagnostic tests for specific disease entities, clinicians and researchers look for sensitive indicators of onset and disease impact. Although ICD-10 explicitly eschews a functional performance criterion, its diagnostic guidelines for dementia then state: “The primary requirement for diagnosis is evidence of a decline in both memory and thinking which is sufficient to impair personal activities of daily living. The impairment of memory typically affects the registration, storage, and retrieval of new information, but previously learned and familiar material may also be lost, particularly in the later stages. Dementia is more than dysmnesia: there is also impairment of thinking and of reasoning capacity, and a reduction in the flow of ideas. The processing of incoming information is impaired, in that the individual finds it increasingly difficult to attend to more than one stimulus at a time, such as taking part in a conversation with several persons, and to shift the focus of attention from one topic to another. If dementia is the sole diagnosis, evidence of clear consciousness is required.” Thus the central component outlined in ICD-10 is “a decline . . . sufficient to impair personal activities of daily living.” Many investigators view this level of decrement as overly severe or too far progressed, potentially precluding use of newly available therapeutic agents until the degenerative process has advanced unnecessarily. Also, these distinctive approaches to diagnostic criteria underscore the potential for substantial variation between the results of studies employing one set versus the other. In the absence of a ‘gold standard' test to externally validate one approach or the other, clinicians and investigators comparing research results must maintain a high degree of caution.

ICD-10 includes four dementia categories: (1) dementia in Alzheimer's disease; (2) vascular dementia; (3) dementia in diseases classified elsewhere in the ICD (e.g., dementia in Pick's disease, Huntington's disease, Parkinson's disease, Creutzfeldt-Jakob disease); and (4) unspecified dementia.

Another set of research criteria for the diagnosis of dementia of the Alzheimer's type, established by the National Institute of Neurological Communicative Diseases and Stroke (NINCDS) and the Alzheimer's Disease and Related Disorders Association (ADRDA), now the Alzheimer's Association, has become known as the NINCDS-ADRDA criteria. Several studies have shown that a diagnosis of probable dementia of the Alzheimer's type according to NINCDS-ADRDA criteria selects patients similar to those diagnosed using DSM-III criteria. Depending on the case series, both criteria sets have been capable of identifying cases of Alzheimer's disease confirmed postmortem with a 70 to 90 percent specificity. The DSM-IV criteria share many features for the diagnosis of probable dementia of the Alzheimer's type but go beyond them to more clearly define important behavioral subtypes, akin to ICD-10, that may help guide symptomatic treatment interventions.

Recent critics regard the DSM-IV specifying phrase “with behavioral disturbance” as inadequate. Some advocate the enumeration of discrete subtypes of noncognitive neuropsychiatric syndromes to further classify the behavioral and psychological signs and symptoms found in patients with neurodegenerative dementing disorders. Whether such an approach will prove useful will depend upon defining more completely: (1) the presentation (i.e., form) of the clinical signs and symptoms that require treatment; (2) whether specific noncognitive neuropsychiatric syndromes exist as a central feature (although neither necessary nor sufficient diagnostically) of Alzheimer's disease or other diseases that cause dementia; (3) the prognostic significance of such symptoms, signs, or syndromes; (4) their pathobiological and causal substrates; (5) psychological, social, and environmental factors that affect their expression; and (6) their response(s) to specific therapeutic interventions.

Epidemiology The prevalence of dementia rises exponentially with age. The estimated prevalence of moderate to severe dementia in a population aged 65 years or older is consistently reported at approximately 5 percent. Within that age group the exponential curve is pronounced so that the prevalence in the subgroup aged 65 to 69 years is 1.5 to 2 percent; in the subgroup aged 75 to 79 years it is 5.5 to 6.5 percent; and in the subgroup aged 85 to 89 years it is 20 to 22 percent. Dementia of the Alzheimer's type is the most common dementing disorder in clinical and neuropathological prevalence studies reported from North America, Scandinavia, and Europe. Prevalence studies from Russia and Japan show vascular dementia to be more common in those countries. It remains unclear whether those apparent clinical differences reflect true etiological distinctions or inconsistent uses of diagnostic criteria. Dementia of the Alzheimer's type becomes more common with increasing age; among persons older than 75 years, the risk is six times greater than the risk for vascular dementia. There is a suggestion of higher rates of dementia of the Alzheimer's type in females and higher rates of vascular dementia in males. In geriatric psychiatric patient samples, dementia of the Alzheimer's type is a much more common etiology (50 to 70 percent) than vascular dementia (15 to 25 percent).

Studies of the incidence of dementia have been plagued by widely differing methodology and results. Again, there is an exponential increase in incidence with age, although some reports have noted a leveling off starting around age 75 years.

Etiology Table 10-27 lists common causes of dementia. Alzheimer's disease, the most common type of degenerative dementia, was discussed in an earlier section. Huntington's disease and Parkinson's disease were also discussed earlier in the chapter as paradigmatic examples of subcortical degenerative processes, with clinical and neuropathological descriptions separating them from cortical dementias. There may be clinical and neuropathological overlap between Alzheimer's disease and Parkinson's disease, especially among older patients. The significance of this finding remains unknown.

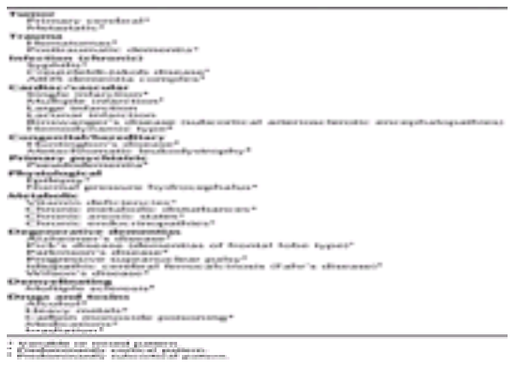

Table 10-27 Causes of Dementia

Frontal Lobe Degeneration In recent years several authors have sought to distinguish dementias of the frontal lobe from other disorders. The uncertain status of dementias of the frontal lobe as distinct clinical and neuropathological entities has not yet warranted their formal inclusion in DSM-IV or ICD-10. They are described as cortical dementias that are found in as many as 10 to 20 percent of cases in some neuropathological series. Age at onset is apparently between 50 and 60 years for the majority, but the reported range is broad—20 to 80 years. The early clinical features of frontal lobe dementias are typified by damage to the frontal lobes and include prominent changes in personality and behavior. The personality changes include disinhibition, social misconduct, and lack of insight; these changes progress to apathy, mutism, and repetitive behaviors. A variant of the Kluver-Bucy syndrome, a condition originally described in monkeys that had undergone surgical ablation of the temporal lobes, is also described in the early stages of frontal lobe dementias and is characterized by combinations of disrupted eating behavior, hyperorality, mood disturbances, and sensory agnosias. Language, praxis, and gnosis are relatively spared early in the disease course, in contrast to dementia of the Alzheimer's type. However, dementias of the frontal lobe are described as progressive conditions that may in some cases involve memory as well as other cognitive functions. To date no studies have attempted to prospectively discriminate dementia of frontal lobe origin from dementia of the Alzheimer's type, with subsequent neuropathological confirmation to determine clinical diagnostic accuracy.

Neuropsychological testing in patients suspected of having dementia of frontal lobe origin may demonstrate disproportionate impairment in tasks related to frontal lobe function, such as deficiency in abstract thinking, attentional shifting, or set formation. Structural neuroimaging, such as CT or MRI, may reveal prominent atrophy of the frontal lobe, especially early in the disease process. Functional neuroimaging may prove more reliable for distinguishing dementia of frontal lobe origin from dementia of the Alzheimer's type. Regional cerebral blood flow studies using radioactively labeled xenon and SPECT studies have demonstrated disproportionate decreases in blood flow, radio tracer uptake, and glucose metabolism in the frontal lobes in patients with suspected or autopsy-confirmed frontal lobe dementia.

At present, the definitive diagnosis of any degenerative dementia rests on postmortem neuropathological examination. Only one type of frontal lobe dementia, Pick's disease, is associated with distinctive histopathological abnormalities that allow for certain diagnosis. Swollen neurons known as Pick cells and intraneuronal inclusions known as Pick bodies define the disorder neuropathologically (Fig. 10-12). Demyelination and gliosis of the frontal lobe white matter may also be found. Other frontal lobe dementias have been referred to as dementia of the frontal lobe type or frontal lobe degeneration of non-Alzheimer's type. They have been distinguished from Alzheimer's disease by their marked gross morphological involvement of frontal and anterior temporal lobes, with relative sparing of the postcentral and temporoparietal areas mostly affected in Alzheimer's disease, and by the absence of amyloid plaques and neurofibrillary tangles microscopically. The lack of positive neuropathological inclusion criteria leaves many of these clinical conditions as disease entities of uncertain status, defined histopathologically by the absence of specific features. Whenever the hallmark findings of Alzheimer's disease are present, that diagnosis has been applied, irrespective of prior clinical findings. Thus, there are no data available to determine how many clinically diagnosed cases of frontal lobe dementia have been recast as Alzheimer's disease after death.

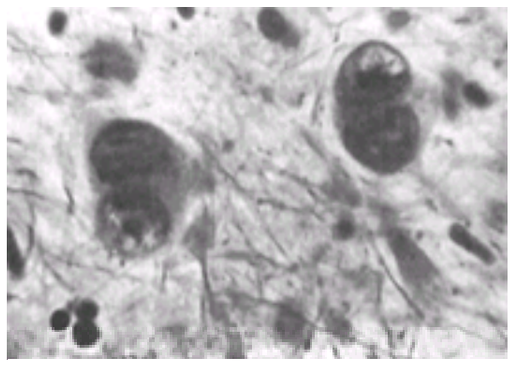

FIGURE 10-12 Intraneuronal inclusions in Pick's disease. Silver stain. (Reprinted with permission from Rowland LP, editor: *Merritt's Textbook of Neurology*, ed 9. Williams & Wilkins, Baltimore, 1995.)

Of the potentially multiple forms of dementia associated with progressive frontal lobe dysfunction, only one type can be distinguished from Alzheimer's disease neuropathologically; the others show no defining postmortem signs. They may also be difficult to distinguish clinically in life. In the early stages of disease, the predominance of behavioral and personality disturbance, the presence of primitive reflexes, and neuropsychological and neuroimaging evidence of disproportionate frontal lobe involvement can help with a more confident premortem diagnosis of frontal lobe dementia. Some authors have assumed that there are many variants of dementia of frontal lobe origin that cannot be distinguished from each other clinically; at present, only Pick's disease has definitive neuropathological features.

Subcortical Degeneration Huntington's disease and Parkinson's disease were discussed earlier as examples of degenerative disorders with a subcortical pattern of deficits. *Progressive supranuclear palsy*, first described in 1964, is a degenerative disease involving the brainstem, cerebellum, and basal ganglia. The presenting history is usually notable for a gait disturbance, particularly spontaneous toppling. The clinical examination is notable for supranuclear paralysis of extraocular movements, particularly in the vertical plane. Dysarthria and dystonic rigidity of the neck and trunk are also common. Onset is usually after age 50 years, with progressive muscular rigidity. Neuropathology is notable for cell loss and gliosis of various nuclei in the brainstem, basal ganglia, and cerebellum, with striking preservation of the cortex. Progressive supranuclear palsy and Huntington's disease were the two disorders to which the label “subcortical dementia” was originally applied. In progressive supranuclear palsy a marked slowing of cognitive processes, apathy, and lack of initiative have been described, associated with relative sparing of language, memory, and praxis. *Fahr's disease* involves idiopathic calcification of the basal ganglia. A subcortical dementia with a parkinsonian syndrome has been described. (Mild basal ganglia calcification is frequently observed incidentally on neuroimaging studies. The clinical significance of that finding is unknown.) Basal ganglia calcification can also be seen in patients with disorders of calcium metabolism, with the expected patterns of subcortical dementia and movement disorder.

Vascular Etiologies Cerebrovascular diseases together comprise the second most common cause of dementia. This category of dementia was referred to in the past as arteriosclerotic dementia, reflecting the belief that vascular insufficiency was responsible for the cognitive degeneration. That has now been supplanted by the belief that tissue damage or infarction underlies the vascular dementias. Cerebral infarction can be the result of a number of processes, of which thromboembolism from a large vessel plaque or cardiothrombus is the most common (Fig. 10-13). Anoxia due to cardiac arrest, hypotension, anemia, or sleep apnea can also produce ischemia and infarction. Cerebral hemorrhage related to hypertension or an arteriovenous malformation accounts for approximately 15 percent of cerebrovascular disease (Fig. 10-14).

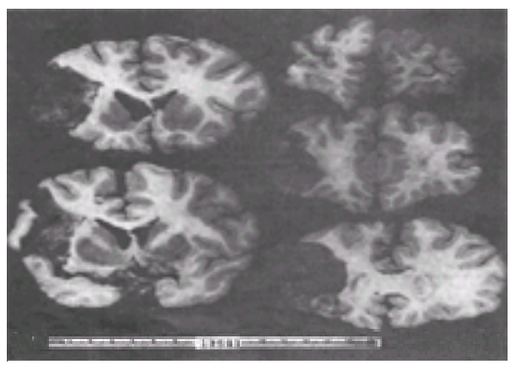

FIGURE 10-13 Hemorrhagic infarct in the territory of the middle cerebral artery. (Reprinted with permission from Hirano A: *A Guide to Neuropathology*. Igaku-Shoin, New York, 1981.)

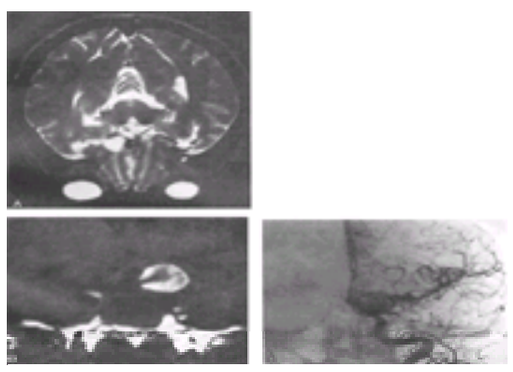

FIGURE 10-14 Giant aneurysm. **A**, T2-weighted axial MRI scan shows a large, hyperintense mass in the left suprasellar region with medial displacement of the distal left internal carotid artery. **B**, T1-weighted coronal MRI scan demonstrates the heterogeneous parasellar mass; the areas of increased signal intensity indicate thrombus. A small curvilinear focus of flow void is seen medially most consistent with small residual patent vascular lumen. These findings are suggestive of partially thrombosed giant aneurysm. **C**, Anteroposterior view of arterial phase of left common carotid arteriogram illustrates a giant aneurysm of the left internal carotid artery. (Reprinted with permission from Rowland LP, editor: *Merritt's Textbook of Neurology*, ed 9. Williams & Wilkins, Baltimore, 1995.)

The clinical characteristics of a vascular dementia depend on the area of infarction. As such, there is a wide variability in the possible presenting features of a vascular dementia. Single infarctions may result in the discrete loss of one particular function (e.g., language) without dementia per se. However, some strategically located infarctions can affect more than one domain of cognitive function and mimic the clinical picture of a global dementia. An example is the angular gyrus syndrome that can occur with large posterior lesions in the dominant hemisphere. It has been characterized as manifesting with alexia with agraphia, aphasia, constructional disturbances, and Gerstmann syndrome (acalculia, agraphia, right-left disorientation, and finger agnosia). Although the findings are similar to those of dementia of the Alzheimer's type, angular gyrus syndrome can be distinguished by its abrupt onset, the presence of focal neurological, EEG, and imaging abnormalities, and preservation of memory and ideomotor praxis.

Vascular dementia is more commonly associated with multiple infarctions. The infarctions may take the form of numerous large infarctions accompanied by widespread cognitive and motor deficits. Tiny, deep infarctions, *lacunes*, result from disease of the small arteries that usually involves subcortical structures, such as the basal ganglia, thalamus, and internal capsule. The neurological and cognitive deficits may resolve quickly after each of the small strokes; however, the deficits may accumulate, leading to a persisting functional and intellectual decline. In the past a stepwise pattern of deterioration was described for that type of vascular dementia, but it was dropped from the DSM-IV criteria, as no specific pattern of deterioration has been reliably demonstrated for vascular dementias. Similarly, the description of patchy deficits has been deleted, in light of the marked variability in presentation of vascular dementia, depending on the type of vasculature and the site and extent of infarction.

Binswanger's Disease Also called subcortical arteriosclerotic encephalopathy, this is characterized by microinfarctions of the white matter with sparing of the cortex. It was originally believed to be a rare form of dementia that could be diagnosed only at autopsy. With the advent of sophisticated neuroimaging techniques such as CT and MRI and the common observation of white matter hyperintensities, there is renewed interest in the disease. Binswanger's disease produces a subcortical pattern of dementia, as the neuropathology is restricted to white matter. However, the mere presence of white matter hyperintensities on MRI is not adequate for diagnosis, as those areas may represent small infarctions, focal demyelination, or simply dilated perivascular spaces. Some studies have found no postmortem pathological correlate to white matter hyperintensities detected on MRI in vivo. Recently, criteria have been proposed for the diagnosis of Binswanger's disease that include clinical and neuropsychological confirmation of dementia, the presence of vascular risk factors, evidence of focal cerebrovascular disease, evidence of subcortical dysfunction, and bilateral white matter abnormalities greater than 2 mm in size on CT or T2-weighted MRI scans. Vascular dementia of the hemodynamic type is a classification that has been used to refer to cognitive impairments that arise secondary to hypotensive episodes, such as those due to cardiac dysrhythmias or hypotension. They may overlap phenomenologically with other conditions that result from chronic hypoxia.

Wilson's disease (*hepatolenticular degeneration*) is an inherited disorder involving abnormal metabolism of copper. Copper accumulates in both the liver and the CNS, particularly in the striatum, caudate, and putamen. Onset usually occurs during childhood or adolescence, although it may be delayed until middle age. Personality change and behavioral disturbance are the most common neuropsychiatric manifestations (and frequently the presenting symptoms of the disease), but cognitive impairment may also be present. The latter takes the form of a subcortical dementia, with psychomotor slowing and loss of initiative, in the presence of relatively spared language functions, memory, and praxis. Motor symptoms are prominent in a parkinsonian pattern and include rigidity, tremor, and, at times, athetosis. The diagnosis is confirmed by assay of serum copper levels and urinary copper excretion. Treatment with chelating agents—dimercaprol (BAL); in the past and penicillamine (Cuprimine) more recently—can retard the progression of the disease and in some instances can result in improvement in clinical features. Neuropsychiatric symptoms are treated symptomatically.

Other Causes Primary psychiatric disorders can present with cognitive impairment. The term “pseudodementia” has been used to describe cognitive deficits that can be seen in the presence of idiopathic psychiatric illness, especially major depressive disorder. The deficits are usually subcortical in nature, involving attention, speed of mental processing, memory retrieval, and verbal fluency and elaboration. Patients may register new material but have difficulty with spontaneous recall that typically improves when they are presented with recognition cues. Pseudodementia was originally thought to be simply another expression of the depressed patient's lack of energy and unwillingness to attend to tasks. More recently, it has become clear that the deficits of pseudodementia represent fundamental cognitive deficits related to the same brain dysfunction that is responsible for the depressive symptoms. *Pseudodementia syndrome of depression* is one current term that is synonymous with *pseudodementia* and may more accurately reflect the nature of the pathobiological process. Recent studies have indicated that it may have a poorer prognosis, especially in the elderly, and several investigators have described a persistent mild anomia in the same patient population.

Schizophrenia was viewed at first as a disorder in which cognitive impairment was a prominent feature (dementia precox). Negative symptoms such as paucity of speech, poverty of ideas, blunting of affect, and functional deterioration contributed to that perception. Contemporary studies have demonstrated consistent cognitive deficits in certain subgroups of schizophrenia patients, primarily involving neuropsychological tasks thought to be sensitive to frontal lobe function. However, it is unclear whether those deficits are acquired over the course of the illness or represent cognitive skills that have never developed, consistent with the neurodevelopmental hypothesis of schizophrenia.

Normal pressure hydrocephalus is an idiopathic disorder caused by partial obstruction to the flow of CSF into the subarachnoid space. Onset typically occurs after age 60 years. The pathophysiology is thought to be related to disruption of neural function, either through stretching of periventricular fibers or through disruption of the pressure differential between the ventricular and subdural spaces, compromising neuronal function by altering cerebral blood flow. The classic clinical triad of dementia, incontinence, and gait disturbance is not present uniformly in all patients with normal pressure hydrocephalus, especially early in the course, although it

nearly always emerges if the condition goes unrecognized or untreated. The diagnosis is based on clinical findings, neuroimaging evidence of ventricular dilation in the absence of sulcal widening (Fig. 10-15), and normal CSF pressure measurements on lumbar puncture. The dementia can be of a subcortical or cortical pattern and may at times be reversed with CSF shunt surgery. Specific indicators of a positive outcome remain to be established, although identification of the etiology and a short disease course favor improvement in the dementia. Rarely, case reports have documented marked improvements up to 4 years after the onset of progressive dementia.

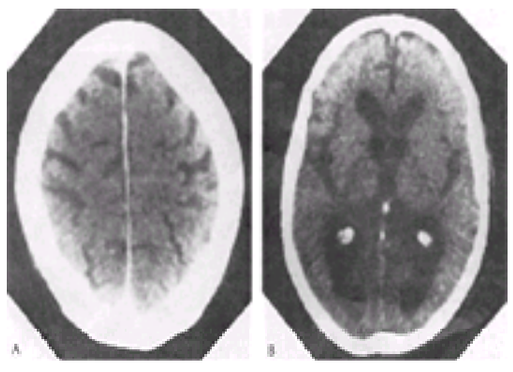

FIGURE 10-15 Brain CT scans. Marked ventricular dilatation (**A**) and widening of cortical sulci (**B**) indicative of hydrocephalus ex vacuo in a 64-year-old woman with dementia. (Reprinted with permission from Rowland LP, editor: *Merritt's Textbook of Neurology*, ed 9. Williams & Wilkins, 1995.)

Irradiation-induced dementia is an iatrogenic concomitant of cranial radiation treatment that has been reported with greater frequency as the posttreatment survival time for patients with intracranial tumors has lengthened. Although transient cognitive deficits can be observed coincident with treatment or soon after treatment, a progressive irreversible dementia can begin 6 to 24 months after the termination of treatment. White matter is particularly sensitive to the deleterious effects of irradiation, and the dementia is predominantly subcortical in nature, reflecting the preferential white matter degeneration. The pathophysiology has been hypothesized to involve arteriolar leakage and localized edema.

DIAGNOSIS AND CLINICAL FEATURES

DSM-IV has eliminated the general syndrome of dementia that was included in DSM-III-R. The dementia diagnoses in DSM-IV are dementia of the Alzheimer's type (Table 10-28), vascular dementia (Table 10-29), dementia due to other general medical conditions (Table 10-30), substance-induced persisting dementia (Table 10-31), dementia due to multiple etiologies (Table 10-32), and dementia not otherwise specified (Table 10-33). The ICD-10 diagnostic criteria for dementia are presented in Table 10-34, Table 10-35, Table 10-36, and Table 10-37.

Table 10-28 DSM-IV Diagnostic Criteria for Dementia of the Alzheimer's Type

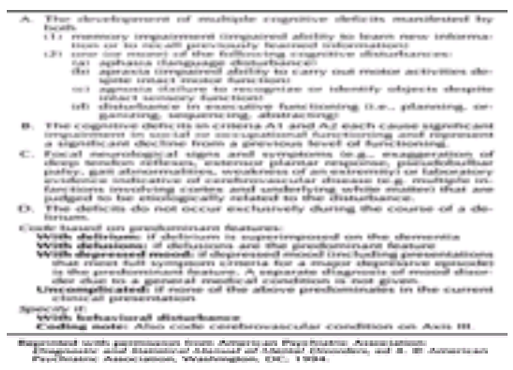

Table 10-29 DSM-IV Diagnostic Criteria for Vascular Dementia

Table 10-30 DSM-IV Diagnostic Criteria for Dementia Due to Other General Medical Conditions

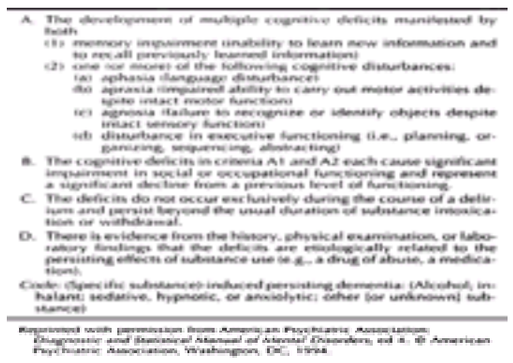
A. The development of multiple cognitive deficits manifested by both
(1) memory impairment (inability to learn new information and to recall previously learned information)
(2) one (or more) of the following cognitive disturbances:
(a) aphasia (language disturbance)
(b) apraxia (impaired ability to carry out motor activities despite intact motor function)
(c) agnosia (failure to recognize or identify objects despite intact sensory function)
(d) disturbance in executive functioning (i.e., planning, organizing, sequencing, abstracting)
B. The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning.
C. The deficits do not occur exclusively during the course of a delirium and persist beyond the usual duration of substance intoxication or withdrawal.
D. There is evidence from the history, physical examination, or laboratory findings that the deficits are etiologically related to the persisting effects of substance use (e.g., a drug of abuse, a medication).
Code: (Specific substance)-induced persisting dementia: (Alcohol); inhalant; sedative, hypnotic, or anxiolytic; other (or unknown) substance)
Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-31 DSM-IV Diagnostic Criteria for Substance-Induced Persisting Dementia


A. The development of multiple cognitive deficits manifested by both
(1) memory impairment (inability to learn new information and to recall previously learned information)
(2) one (or more) of the following cognitive disturbances:
(a) aphasia (language disturbance)
(b) apraxia (impaired ability to carry out motor activities despite intact motor function)
(c) agnosia (failure to recognize or identify objects despite intact sensory function)
(d) disturbance in executive functioning (i.e., planning, organizing, sequencing, abstracting)
B. The cognitive deficits in criteria A1 and A2 each cause significant impairment in social or occupational functioning and represent a significant decline from a previous level of functioning.
C. There is evidence from the history, physical examination, or laboratory findings that the disturbance has more than one etiology (e.g., head trauma plus chronic alcohol use, dementia of the Alzheimer's type with the subsequent development of vascular dementia).
D. The deficits do not occur exclusively during the course of delirium.
Coding note: Use multiple codes based on specific dementias and specific etiologies, e.g., dementia of the Alzheimer's type, with late onset uncomplicated; vascular dementia, uncomplicated.
Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-32 DSM-IV Diagnostic Criteria for Dementia Due to Multiple Etiologies

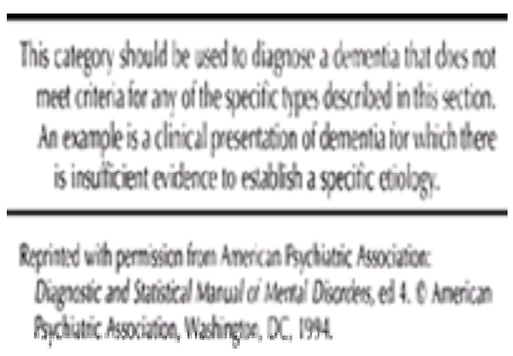
This category should be used to diagnose a dementia that does not meet criteria for any of the specific types described in this section. An example is a clinical presentation of dementia for which there is insufficient evidence to establish a specific etiology.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-33 DSM-IV Diagnostic Criteria for Dementia Not Otherwise Specified

Table 10-34 ICD-10 Diagnostic Criteria for Dementia

Table 10-35 ICD-10 Diagnostic Criteria for Dementia in Alzheimer's Disease

Table 10-36 ICD-10 Diagnostic Criteria for Vascular Dementia

Table 10-37 ICD-10 Diagnostic Criteria for Dementia in Other Diseases Classified Elsewhere

Dementia of the Alzheimer's Type The DSM-IV diagnostic criteria for dementia of the Alzheimer's type emphasize the presence of memory impairment and the associated presence of at least one other symptom of cognitive decline (aphasia, apraxia, agnosia, or abnormal executive functioning). The diagnostic criteria also require a continuing and gradual decline in functioning, impairment in social or occupational functioning, and the exclusion of other causes of dementia. DSM-IV suggests that the age of onset be characterized as early (at age 65 or below) or late (after age 65) and that a predominant behavioral symptom be coded with the diagnosis, if appropriate.

Vascular Dementia The general symptoms of vascular dementia are the same as those for dementia of the Alzheimer's type, but the diagnosis of vascular dementia requires the presence of either clinical or laboratory evidence supportive of a vascular cause of the dementia.

Dementia Due to Other General Medical Conditions DSM-IV lists six specific causes of dementia that can be coded directly: HIV disease, head trauma, Parkinson's disease, Huntington's disease, Pick's disease, and Creutzfeldt-Jakob disease. A seventh category allows the clinician to specify other nonpsychiatric medical conditions associated with dementia.

Substance-Induced Persisting Dementia The primary reason that this DSM-IV category is listed both with the dementias and with the substance-related disorders is to facilitate the clinician's thinking regarding differential diagnosis. The specific substances that DSM-IV cross-references are alcohol; inhalant; sedative, hypnotic, or anxiolytic; and other or unknown substances.

Clinical Diagnosis and Evaluation The first step in the diagnosis of dementia is to establish that the cognitive deficits have occurred in a patient with a stable level of consciousness, without fluctuation or waxing and waning. It must also be demonstrated that the patient has multiple deficits rather than a focal disturbance such as that seen in amnestic disorder or primary progressive aphasia (the insidious onset of a slowly progressive language disturbance with relatively preserved memory, reasoning, judgment, and comportment). Once the basic criteria for the diagnosis of dementia have been met, the task is to determine which etiology is responsible by using the standard means of history, clinical examination, and laboratory evaluation.

A 61-year-old high-school science department head, who was an experienced and enthusiastic camper and hiker, became extremely fearful while on a trek in the mountains. Gradually, over the next few months he lost interest in his usual hobbies. Formerly a voracious reader, he stopped reading. He had difficulty doing computations and made gross errors in home financial management. On several occasions he became lost while driving in areas that were formerly familiar to him. He began to write notes to himself so that he would not forget to do errands. Very abruptly, and in uncharacteristic fashion, he decided to retire from work, without discussing his plans with his wife. Intellectual deterioration gradually progressed. He spent most of the day piling miscellaneous objects in one place and then transporting them to another spot in the house. He became stubborn and querulous. Eventually he required assistance to shave and dress.

When examined 6 years after the first symptoms had developed, the patient was alert and cooperative. He was disoriented with respect to place and time. He could not recall the names of four or five objects after a 5-minute interval of distraction. He could not remember the names of his college and graduate school or the subject in which he had majored. He could describe his job by title only. In 1978 he thought that John Kennedy was president of the United States. He did not know Joseph Stalin's nationality. His speech was fluent and well articulated, but he had considerable difficulty finding words and used many long and essentially meaningless phrases. He called a cup a vase, and identified the rims of glasses as “the holders.” He did simple calculations poorly. He could not copy a cube or draw a house. His interpretation of proverbs was concrete, and he had no insight into the nature of his disturbance.

An elementary neurological examination revealed nothing abnormal, and routine laboratory tests were also negative. A CT scan, however, showed marked cortical atrophy. (Reprinted with permission from *DSM-IV Casebook*.)

For dementia of the Alzheimer's type, a family history of the dementia is probably the most important risk factor after advanced age. A family history of Down syndrome or of hematological malignancies, such as leukemia, myelolymphoma, or Hodgkin's disease, is also associated with an increased risk for Alzheimer's disease. There is some evidence for a familial predisposition to vascular dementia, but it has not been demonstrated as clearly as for dementia of the Alzheimer's type. The family history is of greatest significance in the heredity dementias, such as Huntington's disease, which is transmitted via a single autosomal dominant gene with nearly 100 percent penetrance. A history of a parent or grandparent with a movement disorder and dementia should alert the clinician to that diagnostic possibility. Huntington's disease does not skip generations, although family members may have died from other causes prior to the emergence of definable symptoms. A familial pattern has been established for Wilson's disease, with a presumptive autosomal recessive gene responsible for abnormal copper metabolism. Metachromatic leukodystrophy similarly is inherited in a recessive pattern with incomplete penetrance.

Degenerative dementias as a group do not have well-established risk factors other than old age and familial patterns. For dementia of the Alzheimer's type, other risk factors identified tentatively in recent years include female sex, a past history of head trauma, and lower education. Vascular dementias are highly associated with the risk factors for cerebrovascular disease. Those factors include hypertension (especially with systolic pressures greater than 160 mmHg), cardiac disease, transient ischemic attacks, diabetes mellitus, carotid bruits, and sickle cell disease. Obesity, a sedentary lifestyle, tobacco use, alcohol consumption, and elevated serum cholesterol and lipid levels are less well established as risk factors for cerebrovascular disease.

A history of severe head trauma or multiple traumas over a period of time (such as in boxers) should raise the suspicion of dementia related to brain trauma. Although severe head trauma earlier in life increases the risk of dementia of the Alzheimer's type, its mechanism of action is unknown. A history of an untreated or partially treated sexually transmitted disease should raise the suspicion for neurosyphilis. The presence of risk factors for HIV infection, namely homosexuality, multiple sexual partners, and intravenous drug use, similarly increase the risk for dementia due to HIV disease. Patients with chronic medical illnesses, especially if poorly controlled, such as epilepsy, renal failure, or hepatic cirrhosis, are also at greater risk for developing dementias. A history of occupational exposure to heavy metals or other toxins should be obtained as part of any evaluation for dementia.

Pathology and Laboratory Examination A general physical examination is a routine component of the workup for dementia. It may reveal evidence of systemic disease causing brain dysfunction, such as an enlarged liver and hepatic encephalopathy, or it may demonstrate systemic disease related to particular CNS processes. The detection of Kaposi's sarcoma, for example, should alert the clinician to the probable presence of AIDS and the associated possibility of AIDS dementia complex. Focal neurological findings, such as asymmetrical hyperreflexia or weakness, are seen more often in vascular than in degenerative diseases. Frontal release signs and primitive reflexes, while suggesting pathology in the frontal lobe, are present in many disorders and often point to a greater extent of

progression.

Laboratory evaluation can assist in definitive identification of the etiological agent. The range of possible etiologies of dementia mandates selective use of laboratory tests. The evaluation should follow informed clinical suspicion, based on the history and physical and mental status examination results. Table 10-4 lists a number of laboratory tests useful in evaluating specific diseases presenting as dementia.

Differential Diagnosis The first step in the diagnosis of dementia is to exclude delirium. Delirium can mimic every possible psychiatric disorder and symptom. It is most common in the same populations in which dementia is most common, namely the elderly and the brain-injured. It can be distinguished from dementia by its cardinal feature, disturbance of consciousness. Level of consciousness or arousal must be determined to be stable before a diagnosis of dementia can be made with confidence. Dementia must also be distinguished from focal or specific cognitive impairments, such as those seen in aphasic or amnestic patients. Mood disorders can present with cognitive symptoms, particularly in the dementia of depression or pseudodementia. A history of a mood disorder or a current disturbance in neurovegetative function should alert the clinician to the possibility of a major depressive disorder.

Course and Prognosis The course and prognosis of a dementia syndrome vary with its cause. Dementia does not in itself imply a progressive deterioration, although many of the pathobiological processes underlying dementia are degenerative, and there is no known means of altering the progressive clinical deterioration. The rate of progression may vary within families or from individual to individual. Occasionally, progression can be halted or slowed in the vascular dementias if contributing risk factors for further vascular events can be reduced. Some dementias, such as those related to endocrine or metabolic processes or drug intoxications, may resolve entirely with the treatment or with removal of the basic disorder. However, a long-standing cerebral insult often leads to chronic clinical deficits that persist even when the insult has been removed. Dementias related to tumor and infection usually follow a similar pattern.

Age at onset is an important feature of any illness. Alzheimer's disease is the most common cause of dementia in the United States. Onset usually occurs after age 60 years and the prevalence increases exponentially with each successive decade, although cases have been reported in patients as young as 30 years. Familial forms of dementia of the Alzheimer's type appear to have an earlier age at onset. Cerebrovascular disease, the second most common cause of dementia, is associated with an earlier age at onset overall. Dementia secondary to other medical conditions usually arises only after the disease has progressed for some time. This observation is true of the dementias associated with infectious, physiological, metabolic, and toxic processes. The age at onset of Huntington's disease is usually between 30 and 50 years, but onset may occur earlier or later.

The dementias can be distinguished to some extent by their course, especially earlier in the disease process. Degenerative dementias are insidious in onset and gradually progressive. Despite the clinical rule of a steadily progressive course in dementia of the Alzheimer's type, some individuals may reach a plateau for several years in the overall functional impairment before progression resumes and continues on to death. Vascular dementias may follow a stepwise pattern, in which new deficits appear abruptly and associated with new vascular events, but the vascular dementias also often have an insidious onset and a slow but steadily progressive course. Dementias related to infection are usually acute, although syphilis and cryptococcal meningitis can have an indolent course. Metabolic dementias may begin rapidly or slowly, depending on the underlying systemic disease; correction of the basic deficiency or disturbance may result in improvement, although the cognitive deficits often persist. Drug- or toxin-related dementias may improve once the insult has been discontinued, although radiation-induced dementia is an exception: It first manifests many months after radiation exposure has ceased, and a progressive course ensues.

Treatment The first step in the treatment of dementia is verification of the diagnosis. Accurate diagnosis is imperative, for the progression may be halted or even reversed if appropriate therapy is provided. Preventive measures are important, particularly in vascular dementia. Such measures might include changes in diet, exercise, and control of diabetes and hypertension. Pharmacological agents might include antihypertensive, anticoagulant, or antiplatelet agents. Blood pressure control should aim for the higher end of the normal range, as that has been demonstrated to improve cognitive function in patients with vascular dementia. Blood pressure below the normal range has been demonstrated to result in further impairment of cognitive function in the patient with dementia. The choice of antihypertensive agent can be significant in that beta-blocking agents have been associated with exaggeration of cognitive impairment. Angiotensin-converting enzyme (ACE) inhibitors and diuretics have not been linked to the exaggeration of cognitive impairment and are thought to lower blood pressure without affecting cerebral blood flow (cerebral blood flow is presumed to correlate with cognitive function). Surgical removal of carotid plaques may prevent subsequent vascular events in carefully selected patients.

For the degenerative dementias, no direct therapies have been demonstrated conclusively to reverse or retard the fundamental pathophysiological processes. The search for such an agent has been exhaustive and fraught with frustration. Such studies are constructed on a growing foundation of knowledge regarding brain neurochemistry and the derangements found in dementia. Numerous neurotransmitters, including acetylcholine, dopamine, norepinephrine, GABA, and serotonin, and several neuropeptides, including somatostatin and substance P, are decreased in dementia. Alzheimer's disease has been studied the most extensively, but similar decreases in neurotransmitters have been found in Huntington's disease, alcohol–induced persisting dementia, vascular dementia, Parkinson's disease, and (rarely) in normal aging. Multiple neuropharmacological strategies have been devised in the hope of replenishing the deficient neurotransmitters. Replacement therapy for acetylcholine has been the most common and widely publicized strategy. Efforts at replenishment have included the use of acetylcholine precursors (e.g., example, choline [Anthropan] and lecithin [Phoschol]), cholinergic agonists (e.g., pilocarpine [Salagen] and arecoline), and cholinesterase inhibitors. Treatment with physostigmine (Antilirium, Eserine), a short-acting cholinesterase inhibitor, has consistently resulted in small but statistically significant improvements in memory in patients with dementia of the Alzheimer's type and in healthy control subjects during brief-duration infusion studies. New, longer-acting forms now are being investigated. Tacrine (Cognex), became the focus of public debate after a 1986 study reported alleged marked improvements in 16 patients with dementia of the Alzheimer's type. That study, however, was criticized for substantial methodological limitations and was not replicated in several subsequent attempts. Two multicenter studies of varying design were published in late 1992. One study, with an enriched population, aimed to maximize detection of beneficial effect, but found only marginal improvement and no overall evidence of clinically meaningful change. The second reported statistically significant but still modest improvements in cognition. The Food and Drug Administration (FDA) eventually approved the use of tacrine as a therapeutic agent for dementia of the Alzheimer's type. Clinicians must be aware of both its limited demonstrated benefit and its hepatotoxic potential.

Recently the FDA approved the cholinesterase inhibitor, donepezil (Aricept), for symptomatic treatment of mild to moderate cognitive deficits in patients with presumed Alzheimer's disease. Therapeutic effects have been modest. Dosages of 5 to 10 mg daily were given in experimental trials; common adverse effects have included nausea, diarrhea, and vomiting. Insomnia, muscle cramps, and anorexia have occurred occasionally, but unlike tacrine, so far there has been no reported hepatotoxicity. In summary, it has become clear that there are therapies available that may improve the function of patients with dementia of the Alzheimer's type without incurring severe toxicity. Thus it now seems reasonable to declare “When in doubt, treat!” This reflects a fundamental shift in the care of these individuals, moving beyond long-held nihilism to a more optimistic view of clinical intervention. It is the first step in a treatment revolution that will reach full force during the next 10 to 15 years.

Other experimental approaches to treating cognitive impairment or disease progression include a plant extract of *Ginkgo biloba*, estrogens, vitamin E, and prednisone (Deltasone, Orasone). Many researchers have concluded that the notion of a single or selective neurotransmitter defect for any specific dementing illness is simplistic and that future research efforts should be directed toward neuronal protection and regeneration. Selegiline (Eldepryl), a monoamine oxidase (MAO) type B (MAOB) inhibitor, has apparently slowed the progression of Parkinson's disease, presumably by limiting endogenous generation of destructive oxidative products. Similar antioxidant treatments are being used experimentally with other dementias, including Huntington's disease and vascular dementia. Naloxone (Narcan), an opiate antagonist, is thought to have possible application in vascular dementia based on animal studies in which it was demonstrated to decrease the sequelae of cerebral ischemia. Nerve growth factor is being studied as a means of promoting neural regeneration or sprouting.

The absence of curative therapies does not preclude efforts to ameliorate disturbing clinical problems. Symptomatic measures are the rule for behavioral management of most dementia syndromes. Programs that emphasize a high degree of regularity and consistency in daily schedule and environment can mitigate the risk of development of catastrophic reactions or explosive outbursts. All pharmacological agents that are used for the idiopathic psychiatric disorders can be used in patients with dementia, although usually at dosages one half to two thirds lower. Antidepressant medications and ECT are safe and effective for significant depressive symptoms. The use of antipsychotic should be restricted to patients with defined psychotic symptoms because patients with dementia are more susceptible to the parkinsonian adverse effects inherent in these agents. Clinicians and researchers are now cautiously using newer atypical antipsychotic drugs, seeking to avoid these adverse effects. Benzodiazepines may be used briefly and judiciously for emergency sedation but otherwise should be avoided because they can produce delirium and tend to further compromise residual cognitive capacities. Lithium (Eskalith), centrally active b-adrenergic blockers, carbamazepine (Tegretol), and valproate (Depakene) have been used empirically in the treatment of affective lability and aggressive outbursts. Empirical management therapies should be used in conjunction with environmental modifications. Individual psychotherapy may have benefit for patients in the early stages of dementia, especially to assist them in coping with their losses. The positive effects of a therapeutic relationship can still be felt at later stages when patients have more severe cognitive deficits. Family education and support are vital components of any treatment approach, as all members benefit from extensive knowledge about course and prognosis, as well as needing

assistance when assuming new roles in their relationships with the patient.

Amnestic Disorders The inclusion of amnestic disorders in the psychiatric nosology reflects the classification's roots as a manual for state hospital or asylum patients. The number of individuals given amnestic diagnoses due to nutritional deficiency, often related to chronic alcohol dependence, has declined. In contrast, traumatic causes have increased dramatically during recent decades.

Definition The essential feature of amnestic disorders is the acquired impaired ability to learn and recall new information, coupled variably with the inability to recall previously learned knowledge or past events. The impairment must be sufficiently severe to compromise personal, social, or occupational functioning. The diagnosis is not made if the memory impairment exists in the context of reduced ability to maintain and shift attention, as encountered in delirium, or in association with significant functional problems due to the compromise of multiple intellectual abilities, as seen in dementia. Amnestic disorders are secondary syndromes caused by systemic medical or primary cerebral diseases, substance use disorders, or medication adverse effects, as evidenced by findings from clinical history, physical examination, or laboratory examination.

History and Comparative Nosology Although amnestic disorder has been long described, its specific recognition has been relatively recent. It was most clearly elaborated by Sergei Korsakoff and was included among the alcoholic psychoses in DSM-I and DSM-II, as well as in earlier editions. In DSM-I it was classified under chronic brain syndrome associated with intoxication. Understanding that *psychosis* was the term used to denote more severe disturbances of mental status, the authors of DSM-I stated: “The latter [severe alcohol-related brain damage] may manifest itself by the type of chronic delirium formerly diagnosed as Korsakoff's psychosis.” Specific discussion of the amnestic syndromes was absent. Like DSM-I, DSM-II provided little clinical description of amnestic disorders, although a slightly longer definition was presented in the text.

DSM-III and DSM-III-R, in contrast, provided an in-depth discussion and more specific diagnostic criteria. However, both volumes failed to underscore the essential quality of amnestic disorder as characterized by a specific cognitive deficit in the realm of memory, while dementia syndromes were reflective of multiple failures, including memory and other impaired intellectual abilities. DSM-III and DSM-III-R required “demonstrable evidence of impairment in both short- and long-term memory,” whereas the key feature of the disorder is an inability to learn and later recall new information. In addition, neither DSM-III nor DSM-III-R provided for the separation of transient from persistent amnesia. ICD-10 maintains the approach of DSM-III.

Epidemiology Data are not available for estimating the point or lifetime prevalence, incidence, or lifetime risk of persistent amnestic disorder. One recent study indicated that transient global amnesia may have an incidence of 5.2 cases per 100,000 population per year. There are no specific data available on age at onset or culture- or sex-related aspects beyond those relating to the genesis of primary etiological disease processes. For example, transient global amnesia typically occurs after age 50 years.

Etiology Amnestic disorder often occurs as the result of pathological processes that cause damage to specific diencephalic and middle temporal lobe structures (e.g., mammillary bodies, the hippocampus). The pathology is commonly bilateral, but deficits may arise from unilateral lesions. Pathogenic processes include closed-head trauma and penetrating missile wounds, focal tumors, surgical intervention, encephalitis due to infection from herpes simplex virus, infarction of the territory of the posterior cerebral artery, and hypoxia. A common cause of amnestic disorder is the chronic use of alcohol and associated thiamine deficiency.

Transient amnestic disorder, when encountered as a transient global amnesia, is typically associated with cerebrovascular disease and pathology in the vertebrobasilar system. Transient amnesia may also arise from episodic physiological or metabolic disorders, such as acute intoxications or seizures.

DIAGNOSIS AND CLINICAL FEATURES

Diagnosis The differentiation between amnestic syndrome and amnestic disorder made in DSM-III-R has been eliminated in DSM-IV. For the diagnosis of amnestic disorder, DSM-IV requires the “development of memory impairment as manifested by impairment in the ability to learn new information or the inability to recall previously learned information,” and the “memory disturbance causes significant impairment in social or occupational functioning.” A diagnosis of amnestic disorder due to a general medical condition (Table 10-38) is made when there is evidence of a causatively relevant specific medical condition (including physical trauma). DSM-IV further categorizes the diagnosis as being transient or chronic. A diagnosis of substance-induced persisting amnestic disorder is made when there is evidence that the symptoms are causatively related to the use of a substance (Table 10-39). DSM-IV refers the clinician to specific diagnoses within substance-related disorders: alcohol-induced persisting amnestic disorder; sedative, hypnotic, or anxiolytic-induced persisting amnestic disorder; and other (or unknown) substance-induced persisting amnestic disorder. DSM-IV also provides for the diagnosis of amnestic disorder not otherwise specified (Table 10-40). The ICD-10 diagnostic criteria for organic amnesia syndrome not induced by alcohol and other psychoactive substances are listed in Table 10-41.

A. The development of memory impairment as manifested by impairment in the ability to learn new information or the inability to recall previously learned information.
B. The memory disturbance causes significant impairment in social or occupational functioning and represents a significant decline from a previous level of functioning.
C. The memory disturbance does not occur exclusively during the course of a delirium or a dementia.
D. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition (including physical trauma).

Specify if:
Transient: if memory impairment lasts for 1 month or less.
Chronic: if memory impairment lasts for more than 1 month.

Coding note: Include the name of the general medical condition on Axis I, e.g., amnestic disorder due to head trauma; also code the general medical condition on Axis III.

Table 10-38 DSM-IV Diagnostic Criteria for Amnestic Disorder Due to a General Medical Condition

A. The development of memory impairment as manifested by impairment in the ability to learn new information or the inability to recall previously learned information.
B. The memory disturbance causes significant impairment in social or occupational functioning and represents a significant decline from a previous level of functioning.
C. The memory disturbance does not occur exclusively during the course of a delirium or a dementia and persists beyond the usual duration of substance intoxication or withdrawal.
D. There is evidence from the history, physical examination, or laboratory findings that the memory disturbance is etiologically related to the persisting effects of substance use (e.g., a drug of abuse, a medication).

Code: (Specific substance)-induced persisting amnestic disorder: (Alcohol; sedative, hypnotic, or anxiolytic; other [or unknown] substance)

Table 10-39 DSM-IV Diagnostic Criteria for Substance-Induced Persisting Amnestic Disorder

This category should be used to diagnose an amnestic disorder that does not meet criteria for any of the specific types described in this section.

An example is a clinical presentation of amnesia for which there is insufficient evidence to establish a specific etiology (i.e., dissociative, substance induced, or due to a general medical condition).

Table 10-40 DSM-IV Amnestic Disorder Not Otherwise Specified

A. There is memory impairment, manifest in both
   1. A defect of recent memory (impaired learning of new material) to a degree sufficient to interfere with daily living
   2. A reduced ability to recall past experiences
B. There is no
   1. Defect in immediate recall (as tested, for example, by the digit span)
   2. Clouding of consciousness and disturbance of attention. Delirium, not induced by alcohol and other psychoactive substances
   3. Global intellectual decline (dementia)
C. There is objective evidence from physical and neurological examination, laboratory tests and/or history of an insult to, or a disease of the brain (especially, involving bilaterally the diencephalic and medial temporal structures but other than alcohol encephalopathy) that can reasonably be presumed to be responsible for the clinical manifestations

Comments: Associated features, including confabulations, emotional changes (apathy, lack of initiative), and lack of insight are useful additional pointers to the diagnosis but are not invariably present.

Table 10-41 ICD-10 Diagnostic Criteria for Organic Amnesic Syndrome, Not Induced by Alcohol and Other Psychoactive Substances

Clinical Features The inability to learn and recall new information, the cardinal feature of the disorder, is most apparent on spontaneous, unstructured recall tasks but is also evident on tasks that provide recall cues or recognition paradigms where the stimulus is presented again, often among mnemonically equivalent distractor items. Depending on lesion localization, deficits may be predominantly related to verbal or visual stimuli. (Studies have demonstrated repeatedly that individuals with amnestic disorder may learn how to perform novel procedures that are not mediated verbally, such as motor tasks, even though they later fail to recall having had those learning experiences.)

Problems remembering previously learned materials are present variably among amnestic patients. For example, a patient who suffered traumatic brain damage and who continues to exhibit deficits in new learning may remember events up to a time shortly before the injury. In some cases the interval of preinjury recall impairment may diminish as the patient recovers (shrinking retrograde amnesia), where inaccessible memories from several years before the injury are gradually produced and the extent of the amnesia diminishes in the context of clinical improvement. Recall deficits due to other causes may involve memory for knowledge and events gained over many years' duration.

For some forms of amnestic disorder, events from the remote past may be better remembered than more recent events. However, such a gradient of recall is not present uniformly among individuals with amnestic disorders. Typically, the ability to immediately repeat a sequential string of information (e.g., a digit span) is not impaired in amnestic disorder; when such impairment is evident, it suggests the presence of attentional dysfunction that may be indicative of delirium. Amnestic disorders may be transient, lasting for several hours to a few days, as in transient global amnesia, or persistent, lasting at least 1 month. In the context of a newly developed but unresolved memory impairment, the term *provisional* should be added to a diagnosis of transient amnesia.

Transient global amnesia is a form of transient amnestic disorder associated with episodes that are characterized by a dense, transitory inability to learn new information (i.e., to form sustained memories), with a variable (ultimately shrinking on recovery) inability to recall events that occurred during the duration of the disturbance. The episode is followed by restoration to a completely intact cognitive state. There are no data to suggest that the memory impairment is associated with disturbed or abnormal behavior beyond the mild confusion or perplexity that may be manifest during the episode.

Depending on the cause of the disorder, the onset of amnesia may be sudden or gradual. Head trauma, vascular events, or specific types of neurotoxic exposure (e.g., carbon monoxide poisoning) may lead to acute mental status changes. Prolonged substance abuse, chronic neurotoxic exposure, or sustained nutritional deficiency exemplify conditions that may lead to an insidious memory decline, eventually causing a clinically definable cognitive impairment.

Amnestic disorder may develop as a result of alcohol dependence, associated with dietary and vitamin deficiency. Alternatively, it may be the primary clinical deficit arising from traumatic head injury and may present as the major feature of a postconcussional state. When memory dysfunction exceeds other features of a postconcussional syndrome, it is preferable to diagnose the condition as amnestic disorder due to head trauma.

Although persons with amnestic disorders may manifest other features of the primary systemic or cerebral disease that cause the development of the memory impairment, disordered mental status may be the sole presenting feature. Thus, a clinician may misconstrue the history of a blandly confabulating person unless other corroborating persons are available. When amnestic disorder is the result of alcohol dependence and vitamin deficiency, other neurological complications of alcohol ingestion and malnutrition such as peripheral neuropathy and cerebellar ataxia, may be observed.

A 46-year-old house painter is admitted to the hospital with a history of 30 years of heavy drinking. He has had two previous admissions for detoxification, but his family states that he has not had a drink in several weeks, and he shows no signs of alcohol withdrawal. He looks malnourished, however, and on examination is found to be ataxic and to have a bilateral sixth cranial nerve palsy. He appears confused and mistakes one of his physicians for a dead uncle.

Within a week the patient walks normally, and there is no longer any sign of a palsy. He seems less confused and can now find his way to the bathroom without direction. He remembers the names and birthdays of his siblings, but has difficulty naming the past five United States presidents. More strikingly, he has great difficulty in retaining information for longer than a few minutes. He can repeat a list of numbers immediately after he has heard them, but a few minutes later does not recall being asked to perform the task. Shown three objects (keys, comb, ring), he cannot recall them 3 minutes later. He does not seem worried about this. Asked if he can recall the name of his doctor, he replies, “Certainly,” and proceeds to call the doctor “Dr. Masters” (not his name), whom, he claims, he first met during the Korean War. He tells a long untrue story about how he and “Dr. Masters” served as fellow soldiers.

The patient is calm, alert, and friendly. Because of his intact immediate memory and spotty but sometimes adequate remote memory, one can be with him for a short period and not realize he has a severe memory impairment. Although treated with high doses of thiamine, the short-term memory deficit persists and appears to be irreversible. (Reprinted with permission from *DSM-IV Casebook*.)

Pathology and Laboratory Examination Laboratory findings diagnostic of the disorder may be obtained using quantitative neuropsychological testing. Standardized tests also are available to assess recall of well-known historical events or public figures, to characterize the nature of an individual's inability to remember previously learned information. Performance on such tests varies among individuals with amnestic disorder. Subtle deficits in other cognitive functions may be noted in individuals with amnestic disorder. However, memory deficits constitute the predominant feature of the mental status examination and account largely for any functional deficits. No specific or diagnostic features are detectable on imaging studies such as MRI or CT. However, damage of middle-temporal lobe structures is common and may be reflected in enlargement of third ventricle or temporal horns or in structural atrophy detected on MRI.

Differential Diagnosis The central feature of amnestic disorder is the inability to learn and recall new information, in the context of variable difficulties recalling

previously learned factual knowledge. Less efficient memory is a component of normatively defined age-related cognitive decline but is neither functionally impairing nor below the statistically normal range when assessed with quantitative procedures. Patients with amnesia uniformly show significant abnormalities on cognitive or neuropsychological tests. Disordered memory is also a feature of delirium and dementia. When memory dysfunction occurs in the context of impaired consciousness, with reduced ability to focus, sustain, or shift attention, delirium predominates. The coexistence of memory impairment and multiple cognitive deficits (e.g., aphasia, apraxis, agnosis, and disturbance in executive functioning) warrants the diagnosis of dementia. Confabulation is a mental status finding encountered in patients with dementia as well as amnesia.

Amnestic disorder may emerge from an evolving clinical picture that includes confusion and disorientation, occasionally with attentional problems that suggest delirium. For example, classically described Korsakoff's syndrome has been associated most often with the delirium of Wernicke's encephalopathy. The latter typically clears quickly with appropriate treatment. Confabulation may be noted during the early stages of the disease process and is often indicated by the recitation of imaginary events to fill gaps in memory, but that sign tends to disappear with time. Profound amnesia typically is associated with disorientation to place and time but rarely to person. Disorientation to self may be encountered in patients with severe dementing disturbances characterized by multiple cognitive deficits but is atypical of pure amnestic disorder. Many patients with severe amnestic disorder lack insight into their deficits, and they explicitly deny its presence despite evidence to the contrary. The lack of insight may contribute to accusations or agitation in rare instances. More commonly, apathy, lack of initiative, emotional blandness, or other changes suggestive of altered personality function may be encountered. Individuals may be superficially friendly or agreeable, but they frequently have a shallow or diminished range of affective expression. Patients with transient global amnesia most often appear bewildered or befuddled. Although they have been described participating in complex activity or conversations in the course of an episode, that is a much less common presentation.

Occasionally, patients may demonstrate intact abilities to learn new information associated with profound memory loss for a circumscribed period of time. That pattern occurs in the setting of a discrete (time-limited) process that temporarily interferes with the patient's ability to establish new memories. Such processes include acute intoxication, transient delirium or encephalopathy (e.g., a seizure), or some other transient disruption of cerebral functioning (e.g., a transient ischemic attack). Such transient amnestic episodes must be defined clinically in the context of the primary disease processes; failure to establish a primary systemic or cerebral etiology suggests a psychogenic origin when that symptom pattern is encountered.

Dissociative amnesia typically does not involve deficits in learning and recalling new information; rather, patients present with a circumscribed inability to recall previously learned information while they continue to function normally in the present.

Psychiatric consultation is requested by an emergency room physician on an 18-year-old male who has been brought into the hospital by the police. The youth appears exhausted and shows evidence of prolonged exposure to the sun. He identifies the current date incorrectly, giving it as September 27 instead of October 1. It is difficult to get him to focus on specific questions, but with encouragement he supplies a number of facts. He recalls sailing with friends, apparently about September 25, on a weekend cruise off the Florida coast, when bad weather was encountered. He is unable to recall any subsequent events and does not know what became of his companions. He has to be reminded several times that he is in a hospital, as he expresses uncertainty as to his whereabouts. Each time he is told, he seems surprised.

There is no evidence of head injury or dehydration. Electrolytes and cranial nerve examination are unremarkable. Because of the patient's apparent exhaustion, he is permitted to sleep for 6 hours. Upon awakening, he is much more attentive, but is still unable to recall events after September 25, including how he came to the hospital. There is no longer any doubt in his mind that he is in the hospital, however, and he is able to recall the contents of the previous interview and the fact that he had fallen asleep. He is able to remember that he was a student at a southern college, maintained a B average, had a small group of close friends, and has a good relationship with his family. He denies any previous psychiatric history and says he has never abused drugs or alcohol.

Because of the patient's apparently sound physical condition, a sodium amytal interview is performed. During this interview he relates that neither he nor his companions were particularly experienced sailors capable of coping with the ferocity of the storm they encountered. Although he had taken the precaution of securing himself to the boat with a life jacket and tie line, his companions had failed to do this and had been washed overboard in the heavy seas. He completely lost control of the boat and felt he was saved only by virtue of good luck and his lifeline. He had been able to consume a small supply of food that was stowed away in the cabin over a 3-day period. He never saw either of his sailing companions again. He was picked up on October 1 by a Coast Guard cutter and brought to shore, and subsequently the police had brought him to the hospital. (Reprinted with permission from *DSM-V Casebook*.)

Patients with resolved transient amnesia (e.g., transient global amnesia) may have a superficially similar history retrospectively. They manifest failure of recall for matters or events that occurred during the discrete episode in question. Thorough clinical investigations of patients with amnestic disorder typically reveal a primary cerebral or systemic medical condition that is etiologically related to the genesis of the mental status abnormality. During an episode, patients with transient amnesia generally have a confused or bewildered demeanor and exhibit marked difficulty with new learning tasks. Episodes of psychogenic amnesia end abruptly, typically associated with an expressed awareness of having no memories for the time period of the amnestic or fugue state. In contrast, the retrograde memory defect of transient global amnesia gradually shortens as the patient recovers; when recovery is complete, the memory gap spans only the period of the episode.

Course Although the mode of onset is typically abrupt, data suggest that individuals with alcohol-induced amnestic disorder may develop deficits insidiously over many years as a result of repeated toxic and nutritional insults before the emergence of a final, dramatically impairing episode of illness apparently related to thiamine deficiency. Transient amnesia due to a cerebrovascular etiology may be recurrent, with episodes lasting from several hours to several days. Amnestic disorders due to head trauma, for example, may last variable amounts of time, with the greatest deficit apparent immediately after injury and improvement occurring during the ensuing 2 years (further improvement beyond 24 months has been noted, but less commonly). Full recovery may occur, although severe injuries are typically characterized by residual deficits. Disorders due to destruction of middle-temporal lobe structures, such as infarction, encephalitis, surgical ablation, or malnutrition in the context of alcohol dependence, may cause densely persisting impairments.

Treatment Whenever a primary systemic or cerebral disorder is causally tied to the amnestic syndrome, initial treatment (with thiamine, antiviral medication, aspirin) must be directed toward the underlying pathological process. Presently there are no known, definitively effective treatments for amnestic disorder that are specifically aimed at reversing apparent memory deficits. A variety of pharmacotherapeutic trials have been to no avail. Recently, centers for cognitive rehabilitation have been established whose rehabilitation-oriented therapeutic milieu is intended to promote recovery from brain injury, especially from traumatic causes. Despite the high cost of extended care at these sites, which provide both long-term institutional and daytime services, no data have been developed to define therapeutic effectiveness for the heterogeneous groups of patients who participate in such tasks as memory retraining. Persons with amnestic disturbances worthy of diagnosis experience major impediments in their social and vocational functioning. They may require supervised living situations to ensure appropriate feeding and care.

Other Cognitive Disorders Disorders such as dementia and amnesia are specific categorical designations that are intended to define disease states. However, intellectual functioning can also be viewed from a dimensional perspective, ranging from optimal to grossly deficient. Dementia represents an abnormal decline from a previous level of attainment; mental retardation reflects the failure to develop adequate intellectual function.

Within this broad framework multiple domains of intellect are recognized that involve a wide variety of brain-related cognitive processes. The determination of normal and abnormal usually is made by comparing a person's performance on a variety of neuropsychological tests with predetermined normative standards. Ideally, the clinician would like lifelong (i.e., premorbid) serial cognitive testing to aid with diagnosis; occasionally, school, military, or vocational records provide an acceptable alternative. Usually one must compare a patient's results against published norms. Those norms may vary in quality, and the clinician should be aware whenever possible of factors such as the education, sex distribution, socioeconomic status, and age distribution of normative samples.

Cognitive Disorder Not Otherwise Specified DSM-IV includes a new diagnostic category, *cognitive disorder not otherwise specified,* to deal with patients whose clinical presentation does not conform to a diagnosis of delirium, dementia, or amnesia. The designation is useful for patients with mild deficits in cognitive functioning that result from conditions such as head trauma, chronic alcohol dependence, or HIV infection. In the recovering alcoholic, for example, or the patient with a significant but resolving posttraumatic amnesia, intellectual abnormalities may be detectable objectively and noted subjectively, although they may be only minimally impairing functionally. Those deficits may disappear over time or remain as subtle residua. HIV infection may cause a mild decline in cognition; current research has demonstrated such decrements repeatedly. Of note, the performance of many patients has remained within the normal range even as the test scores have decreased significantly. The diagnostic criteria for cognitive disorder not otherwise specified appear in Table 10-42.

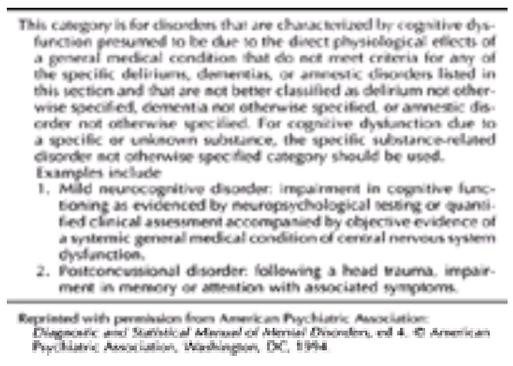

This category is for disorders that are characterized by cognitive dysfunction presumed to be due to the direct physiological effects of a general medical condition that do not meet criteria for any of the specific deliriums, dementias, or amnestic disorders listed in this section and that are not better classified as delirium not otherwise specified, dementia not otherwise specified, or amnestic disorder not otherwise specified. For cognitive dysfunction due to a specific or unknown substance, the specific substance-related disorder not otherwise specified category should be used.

Examples include

1. Mild neurocognitive disorder: impairment in cognitive functioning as evidenced by neuropsychological testing or quantified clinical assessment accompanied by objective evidence of a systemic general medical condition of central nervous system dysfunction.
2. Postconcussional disorder: following a head trauma, impairment in memory or attention with associated symptoms.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-42 DSM-IV Diagnostic Criteria for Cognitive Disorder Not Otherwise Specified

Mild Neurocognitive Disorder To define those conditions with greater specificity, the World Health Organization developed the ICD-10 diagnostic category of mild cognitive disorder (Table 10-43). A similar DSM-IV construct (mild neurocognitive disorder) is included in an appendix as an example of cognitive disorder not otherwise specified. Table 10-44 lists the DSM-IV research criteria for mild neurocognitive disorder. To date, the interface between amnestic disorders or dementing disorders and mild neurocognitive disorder has not been defined reliably or validly.

Table 10-43 ICD-10 Diagnostic Criteria for Other Mental Disorders Due to Brain Damage and Dysfunction and Due to Physical Disease

A. The presence of two (or more) of the following impairments in cognitive functioning, lasting most of the time for a period of at least 2 weeks (as reported by the individual or a reliable informant):
   (1) memory impairment as identified by a reduced ability to learn or recall information
   (2) disturbance in executive functioning (i.e., planning, organizing, sequencing, abstracting)
   (3) disturbance in attention or speed of information processing
   (4) impairment in perceptual-motor abilities
   (5) impairment in language (e.g., comprehension, word finding)
B. There is objective evidence from physical examination or laboratory findings (including neuroimaging techniques) of a neurological or general medical condition that is judged to be etiologically related to the cognitive disturbance.
C. There is evidence from neuropsychological testing or quantified cognitive assessment of an abnormality or decline in performance.
D. The cognitive deficits cause marked distress or impairment in social, occupational, or other important areas of functioning and represent a decline from a previous level of functioning.
E. The cognitive disturbance does not meet criteria for a delirium, a dementia, or an amnestic disorder and is not better accounted for by another mental disorder (e.g., a substance-related disorder, major depressive disorder).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 10-44 DSM-IV Research Criteria for Mild Neurocognitive Disorder

In addition to conditions such as HIV infection, head trauma, or alcohol dependence, mild cognitive decline with neuropsychological performance below the level of age-matched peers may be encountered as an early sign of a progressive degenerative disease. The use of cognitive disorder not otherwise specified as a diagnosis can serve to describe provisionally a patient who the physician suspects will develop a more malignant dementia of the Alzheimer's type and in whom a definitive diagnosis is premature owing to the relative mildness of the symptoms and an associated lack of clarity regarding clinical course. The label of not otherwise specified demands maximum clinical scrutiny and vigilance when employed in this fashion.

Other Cognitive Conditions Clinical investigators and geriatric psychiatrists have recently joined cognitive psychologists in studying aging-related cognitive decline involving such functions as spontaneous verbal memory, cognitive flexibility and abstracting ability, visuospatial processing, divided attention, speed of mental processing, and naming. Aging-related decrements in those functions do not relate to any specific or defined neuropathology, although they may reflect underlying neurobiological deterioration. Of note, objective documentation of individual decline in test performance may be impossible. Although experimental comparisons of groups of healthy older subjects with comparably educated younger groups show consistent changes with aging, there are no data to suggest that the overall decline is a harbinger of disease.

Many persons with normal (i.e., normatively defined) aging-related intellectual decrements seek clinical evaluation for forgetfulness, especially out of fear that they may be developing Alzheimer's disease. Their complaints often include inability to recall names or words spontaneously, absent-mindedness, the need to use reminder lists, or subtle problems with concentration. Careful interviewing typically reveals mild anxiety about minor intellectual problems, the use of effective compensatory mental strategies, and intact personal and social functioning, with little evidence of definable interference from perceived cognitive inadequacies in their daily lives. The absence of significant functional decline, together with performance within the normative (i.e., based on similarly aged samples) range on neuropsychological testing, in the context of an unrevealing general medical evaluation points to aging-associated cognitive alterations.

Because of ample data on the phenomenon and clinicians' need to provide concerned patients with an understandable terminology to define their perceived difficulties, DSM-IV groups age-related cognitive decline among those conditions not attributable to a mental disorder that are a focus of attention or treatment. A variety of other common problems are included in that class, among them borderline intellectual functioning, academic problems, adult antisocial behavior, and marital problems.

Figure 10-16 presents schematically in a dimensional perspective the relations between increasing age and cognitive performance, depicting changes in the normative range, mild cognitive impairment, and dementia. The aging-related decline in normative performance underscores the difficulty of establishing an absolute standard of cognitive deficit that is indicative of impairment due to a categorical disease process. The figure also suggests that there will always be patients detected in the range of mild impairment. As long as there are few (or no) pathobiologically exact laboratory tests to determine with certainty specific cognitive impairment disorders, thoughtful clinical judgment will remain a central part of the diagnostic process.

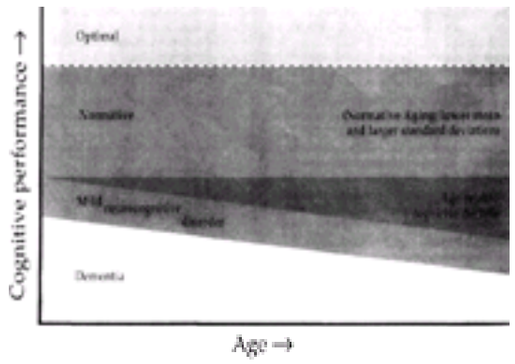

FIGURE 10-16 Aging-associated changes in ranges of cognitive performance.

Mental Disorders Due to a General Medical Condition DSM-IV has taken a different approach to categorizing the mental disorders due to a general medical condition than did DSM-III-R. In DSM-III-R the disorders were classified under the broader category of organic mental disorders. In DSM-IV each mental disorder due to a general medical condition is classified within the category that most resembles its symptoms (Table 10-45). For example, the diagnosis psychotic disorder due to a general medical condition is found in the DSM-IV section on schizophrenia and other psychotic disorders. The symptom-based organization of DSM-IV is meant to facilitate clinical decision making regarding the differential diagnosis of symptoms. For example, the clinician who is evaluating a patient with depression can refer to the DSM-IV section on mood disorders and find mood disorder due to a general medical condition as one of the diagnoses. That diagnosis should help to clarify the importance of considering the possibility of a mental disorder due to a general medical condition for almost all psychiatric presentations.

DSM-IV Category	Mental Disorders due to a General Medical Condition
Delirium, dementia, amnestic, and other cognitive disorders	Delirium due to a general medical condition Dementia due to other general medical conditions Amnestic disorder due to a general medical condition
Schizophrenia and other psychotic disorders	Psychotic disorder due to a general medical condition
Mood disorders	Mood disorder due to a general medical condition
Anxiety disorders	Anxiety disorder due to a general medical condition
Sexual disorders	Sexual dysfunction due to a general medical condition
Sleep disorders	Sleep disorder due to a general medical condition
Mental disorders due to a general medical condition not elsewhere classified	Catatonic disorder due to a general medical condition Personality change due to a general medical condition Mental disorder not otherwise specified due to a general medical condition

Table 10-45 Mental Disorders Due to a General Medical Condition

Mood Disorder Due to a General Medical Condition Secondary mood syndromes are characterized by a prominent mood alteration that is thought to be the direct physiological effect of a specific medical illness or agent. These disorders are often difficult to define and have not been extensively researched; therefore, only limited information can be provided.

Definition The key feature is prominent, persistent, distressing, or functionally impairing depressed mood (anhedonia) or elevated, expansive, or irritable mood, judged to be caused by either an Axis III condition or by substance intoxication or withdrawal. Cognitive impairment is not the predominant clinical feature; otherwise, the mood disturbance would be viewed as part of delirium, dementia, or other cognitive deficit disorder. The diagnostician is asked to specify if the mood syndrome is manic, depressed, or mixed, and if criteria for a fully symptomatic major depressive or manic syndromic are fulfilled.

History and Comparative Nosology Mood disturbances secondary to medical conditions have long been described, but attention was rarely paid to the presence or absence of coexisting intellectual deficits. DSM-III introduced the term and the formal concept of organic affective syndrome, which required both mood alteration and two associated symptoms (as found in primary affective illnesses) to be present and thought due to specific medical etiologies. DSM-III-R eliminated the requirement for associated symptoms. DSM-IV marks the first explicit diagnostic criteria to denote whether or not the disturbance meets full major depressive or manic syndromic criteria. There has been much disagreement in the literature about primary depressive disorders and whether minor depressive disorders exist and how best to define them. Similar arguments might apply to lesser depressive syndromes of secondary origin. Terminology aside, there has been little research in the area of secondary mood disorders; what data exist are hampered by differing (or absent) operationalization of what constitutes sufficient evidence for defining causality.

Epidemiology There are no clear data on which to base statements of incidence or prevalence of secondary mood disorders in any clinical or community population. It is clear that depressive symptoms and a wide array of systemic and primary cerebral conditions coexist to a far higher degree than can be explained by chance. Unfortunately, establishing a causal relation between depressive symptoms and a specific medical entity is difficult; therefore, the percentage of those coexisting symptoms that can be called secondary remains unknown. Further, many reported studies did not assess a range of syndromic criteria (i.e., major versus minor depression), and many simply quantified depressive symptoms by using rating scales without determining if the symptoms attained a threshold level of clinical significance. One noteworthy point is that depression in the medically ill appears to be equally prevalent by sex, or possibly slightly higher in men than in women. This disparity, when compared with the preponderance of females with primary depressive disorders, is often cited as an indicator of the importance of viewing secondary mood disturbances separately.

Rates of mood disorder in the medically ill have been carefully described in several neurological diseases. For example, at least one research group has documented high rates of criteria-defined major and minor depressive syndromes in patients shortly after cerebrovascular accidents. Correlation of stroke lesion location and size with presence (and possibly with type) of depressive syndrome suggests the role of direct disruption of brain physiology as a causal mechanism. Also, the presence or severity of depression does not correlate highly with physical impairment due to hemiparesis, for example, and may be higher than rates found in patients with similar levels of overall disability due to nonneurological conditions, suggesting that depression in neurological patients is not simply a psychological reaction to illness and disability. Similarly, less extensive descriptions of increased rates of depressive symptoms and syndromes have been reported for populations with Parkinson's disease, Huntington's disease, HIV infection (with presumed direct CNS involvement), and multiple sclerosis.

Determination of the secondary or symptomatic nature of these mood syndromes is problematic and is further complicated by the fact that at least some of the patients with these neurological illnesses had substantial cognitive impairment. Although DSM-IV attempts to address these issues directly, uncertainties remain. For example, if a patient with Huntington's disease has dementia and a depressive syndrome, the clinician might choose the diagnosis of dementia due to Huntington's disease with depressed mood. However, early in the disease process a syndromically depressed Huntington's patient with few cognitive impairments would warrant a diagnosis of mood disorder due to Huntington's disease with major depressive episode. As the intellectual impairment progresses, does the clinician abandon one diagnosis for another or add a second Axis I diagnosis of dementia? Such borderline situations are expected to generate uncertainty, which can be ameliorated by careful documentation of one's clinical reasoning.

Secondary mania appears to be much less prevalent in most neurological illnesses, with the exceptions of multiple sclerosis and possibly Huntington's disease. Case reports abound of putative secondary mania due to a variety of other causes, but the prevalence is not known. Finally, patients with secondary mood syndromes may have increased rates of prior mood disorders and higher rates of family history of mood disorder. Therefore, secondary syndromes may reflect an interaction between a precipitating agent or illness and the patient's diathesis toward mood pathology.

Etiology The list of potential causes for both depressive and manic syndromes is long. Table 10-46 lists some of the causes most commonly considered.

Table 10-46 Causes of Secondary Mood Disorders

Diagnosis and Clinical Features The depressive or manic symptoms found in secondary mood disorders are phenomenologically similar to those found in primary (idiopathic) mood disorders (see Table 14.6-17). It is not known if certain symptoms occur more commonly in the secondary disorders; presumably the prevalence may vary depending on the specific etiology of the secondary disorder. For example, anxiety has been described as prominent in major depressive syndromes seen in patients with Parkinson's disease; however, no studies have compared depressed patients with Parkinson's disease with similarly aged patients experiencing idiopathic major depressive disorder.

Associated clinical phenomena may include other manifestations of the cause of the secondary mood disorder, such as soft neurological signs or contributing cognitive impairment; indeed, those findings are used to support the assessment of the mood symptoms as secondary in origin. There are no specific tests to confirm the diagnosis of a secondary mood disorder. In addition, little is known about how neurobiological abnormalities seen in idiopathic mood disorders differ from secondary syndromes. Moreover, despite recent interest in secondary mood disorders, there is still little definitive understanding of the neuroanatomical substrate of those disorders. Physical (including neurological) examination and specific laboratory tests or procedures may be crucial to establishing the presence of the primary disease state.

Differential Diagnosis There are two broad domains of differential diagnosis to consider when establishing the presence of a secondary mood disorder. The first is phenomenological: Does the patient have clinically significant manic or depressive symptoms in the absence of evidence of a predominant cognitive deficit? That assessment requires attention to symptoms and function in the history and mental status examination. As part of the process, the clinician is also establishing whether there is a clearly defined mood syndrome sufficient to warrant an empirical treatment trial with antidepressant medications.

The second domain is etiological: Does the patient have an Axis III condition or a state of substance intoxication or withdrawal that is causing the mood disturbance? Establishing the presence of the relevant condition depends on standard psychiatric and medical-neurological assessments; establishing the causal relation to the mood disorder may be difficult.

Course and Prognosis Although the literature on the course of clearly defined secondary mood disorders is scant, substantial work has demonstrated that all depressive conditions that are comorbid with general medical illnesses or substance-related disorders have poorer prognoses than those that have no demonstrated associations, whether measured by means of symptomatic expression, functional disability, or mortality. It is therefore likely that secondary depressive illness is most often a chronic disease that is sometimes characterized by periods of remission followed by recurrences and sometimes by continuous illness. The prognosis varies, depending on the etiological disease state; depression secondary to a readily treatable disease (e.g., hypothyroidism) has a better outcome than depression associated with a terminal, essentially untreatable condition (e.g., metastatic pancreatic carcinoma). Little is known about the relation of specific depressive syndromes to outcome even within a specific etiological group. For example, there have been attempts to distinguish outcome in poststroke patients based on the presence of minor or major depressive syndromes; this work has had little replication by other investigator groups and the results may not generalize to depression secondary to other conditions. There is even less information available on the course of secondary mania, although many case reports suggest it is chronic and refractory to treatment.

Treatment Treatment response has received limited study. Standard antidepressant medications (e.g., tricyclic antidepressants, MAO inhibitors, selective serotonin reuptake inhibitors (SSRIs), and ECT are effective in many depressed patients with medical and neurological illnesses or substance use disorders. However, the utility of studies of antidepressant drugs has been limited by the scant clinical definition of study patients in many publications; a designation of secondary mood disorder, with depressive features, provides little information on specific target symptoms and their response. The efficacy of newer somatic agents (e.g., SSRIs, bupropion [Wellbutrin]) and of newer psychotherapeutic approaches has received little systematic study. A comparison of response to specific treatment modalities in syndromically similar primary and secondary mood disorders has not been done. However, the comorbid pathology found in secondary syndromes may limit treatment trials, either because of contraindications or because of increased susceptibility to adverse effects.

Given the severely limited data, the clinician treating a patient with a secondary mood disorder must rely on the following general guidelines. The underlying causes should be treated as effectively as possible. Persisting manic or major depressive syndromes are likely to require somatic therapies; standard treatment approaches as used for the corresponding primary mood disorder should be employed, although the risk of toxic effects may require more gradual dosage increases. At a minimum, psychotherapy should focus on psychoeducational issues (in particular, the concept of a secondary behavioral disturbance may be new or difficult for many patients and families) and support. More specific intrapsychic, interpersonal, and family issues are addressed as indicated. Approaches to subsyndromal secondary mood disorders are even less well established, but clinically significant disturbances warrant empirical trials of the same treatments used in fully syndromic patients.

Psychotic Disorder Due to a General Medical Condition *Psychosis* has been a term of inconsistent definition, used in recent years to refer exclusively to symptoms of a striking nature, such as hallucinations and delusions, but in the past to severe affective syndrome (e.g., psychotic versus neurotic depression); to cognitive symptoms, such as confusion, disorientation, or altered memory (such as Korsakoff's psychosis); or as a means of describing the severity of any psychiatric symptom (“of psychotic proportions”). DSM-III-R in its appended glossary defined psychosis as “gross impairment in reality testing and the creation of a new reality.” Hallucinations, delusions, bizarre behavior, and incoherent speech were considered direct evidence of psychosis. Psychotic symptoms have been recognized as nonspecific, as they can be seen in any of the major psychiatric illnesses, such as dementia, schizophrenia, or bipolar I disorder, as well as in many systemic medical or cerebral disorders.

Definition In order to establish the diagnosis of psychotic disorder due to a general medical condition the clinician first must exclude syndromes in which psychotic symptoms may be present in association with cognitive impairment (e.g., delirium and dementia of the Alzheimer's type) but not the defining feature, and the clinician must determine with confidence the causal link. In turn, secondary psychotic disorders must be excluded before a diagnosis of a primary (idiopathic) psychotic disorder is entertained.

History Psychotic symptoms, including delusions, hallucinations, incoherent speech or formal thought disorder, and bizarre behavior, have been recognized for centuries. In medical prehistory they were perceived in a theological light as evidence of demonic possession or punishment for a moral failing. In the eighteenth century, at the threshold of the modern brain disease model, psychosis was sometimes attributed to a bodily dysfunction, either systemic or related specifically to the CNS. However, tension has persisted among clinicians since the 1800s as to whether psychotic symptoms are the manifestations of a dysfunctional brain or are volitional or psychological reactions to life circumstances. Reports of secondary psychosis have abounded for centuries and have contributed to an understanding of those behaviors as specific symptoms of brain disease. Many descriptions, however, have had limited use because of vagaries in the definition of psychosis and the lack of any uniform means for establishing a correlation between psychotic symptoms and the associated systemic or cerebral medical condition under study.

Recently, three types of investigations have emerged for the study of secondary syndromes, including secondary psychosis. One type of study compares psychopathological symptoms in patient groups with and without known primary medical conditions. This comparison has been done strictly on a retrospective basis for secondary psychosis, with only one available prospective study of secondary delusional disorders. Another type of investigation has examined patient groups with known CNS disease, such as cerebrovascular disorder or Huntington's disease, with a careful description of any associated psychopathology. The third investigative track has selected patients with known psychopathology, such as hallucinations in schizophrenia, and sought evidence to correlate symptoms with CNS dysfunction using a variety of measures (e.g., neuroimaging of the temporal lobe). The secondary psychotic disorders are a window through which insights into the

neurobiological basis for psychotic processes may be obtained.

Comparative Nosology Secondary psychotic syndromes were categorized in DSM-II as psychoses associated with organic brain syndromes. The syndromes included in that category were the dementias, deliria, and psychoses associated with other cerebral and systemic conditions. Entry into the category depended on cognitive symptoms, such as disturbances of orientation, memory, judgment, and lability of affect. The term “psychosis” continued to be used for the sake of historic continuity, with the acknowledgment that “many patients for whom these diagnoses are clinically justified are not in fact psychotic.” DSM-III improved on the nosology by establishing the general rubric of organic brain syndromes, with six specific syndromes, including organic hallucinosis and organic delusional syndrome. In DSM-IV, psychotic disorder due to a general medical condition (with its available subtypes) has been moved out of the organic group to the phenomenological cluster to which it is related. This shift underscores the need for differential diagnosis, the clinical importance of defining etiology whenever possible, and the idea that primary psychopathology is idiopathic—that is, without known cause.

Epidemiology The incidence and prevalence of secondary psychotic disorders in the general population are unknown. The prevalence of psychotic symptoms is increased in selected clinical populations, such as nursing home residents with dementia of the Alzheimer's type, but it is unclear how to extrapolate these findings to other patient groups.

Etiology Virtually any cerebral or systemic disease that affects brain function can produce psychotic symptoms. Table 10-27 lists examples within each of the broad categories of diseases that can produce dementia; each of those diseases is also capable of producing psychotic symptoms, both in the presence and in the absence of cognitive impairment. Degenerative disorders, such as Alzheimer's disease or Huntington's disease, may present initially with new-onset psychosis, with minimal evidence of cognitive impairment at the earliest stages.

Diagnosis and Clinical Features To establish the diagnosis of a secondary psychotic syndrome (see Table 13.3-13). The clinician first determines that the patient is not delirious, as evidenced by a stable level of consciousness. A careful mental status assessment is conducted to exclude significant cognitive impairments, such as those encountered in dementia or amnestic disorder. The next step is to search for systemic or cerebral diseases that might be causally related to the psychosis. Psychotic symptomatology per se is not helpful in distinguishing a secondary from a primary (idiopathic) cause.

Comparative studies have not demonstrated any distinguishing phenomenological features in secondary psychosis or any difference in frequency or severity of the psychosis when compared to idiopathic psychosis. Olfactory and auditory hallucinations, although claimed anecdotally to suggest a secondary or symptomatic etiology, have proved unreliable. Some studies have suggested that exclusively positive psychotic symptoms, in the absence of negative symptoms and personality change, reflect a secondary cause; this suggestion has not been tested prospectively. Age at onset is a factor that should alert clinicians to the possible emergence of a secondary psychotic disorder, reflecting both the age-related increased prevalence of diseases affecting brain function and the natural history of primary psychotic syndromes, with their markedly diminished incidence after ages 40 to 45 years.

All patients who present with the new onset of psychotic symptoms should undergo a thorough clinical evaluation emphasizing personal medical history, family medical history, and medical review of systems. A systematic physical and neurological examination should be performed. (The examiner should bear in mind, however, that nonlocalizing, soft neurological signs and a variety of dyskinesias can be present in idiopathic schizophrenia, even in the drug-naïve patient.) A neuroimaging evaluation with MRI for any new-onset psychosis, irrespective of patient age, is recommended. The detection of a systemic or cerebral abnormality does not automatically lead to the determination of secondary psychosis; establishing a secondary status requires thoughtful clinical reasoning.

Differential Diagnosis The differential diagnosis involves first establishing that the symptoms and signs encountered are in fact psychotic, according to the more specific modern definition. Confabulation may be mistaken for delusions. *Confabulation* is the spontaneous or prompted production of inconsistent and fabricated statements, often in response to questions or environmental stimuli. Although memory impairment is present in those who confabulate, the more salient cognitive deficit involves an inability to suppress or self-analyze the automatic fabrications and responses. Confabulation differs from delusions in that the fabricated beliefs are quite transient and varying. A behavioral response to the confabulated belief is usually absent. The presence of confabulation is also suggestive of brain disease, often involving the anterior temporal lobe (memory impairment) and the frontal lobes (loss of self-analysis). Perceptual disturbances that result in illusions or other misinterpretations of environmental stimuli must be distinguished from hallucinations, which are experienced as true perceptual experiences but without an actual stimulus.

Agnosias, or deficit syndromes, such as *prosopagnosia, topographic agnosia,* or *phonagnosia* (inability to recognize familiar faces, places, or sounds, respectively), can occur in the context of intact peripheral perception and can be mistaken for delusional beliefs as well as hallucinations. It is important to distinguish these deficit syndromes and to recognize that they point to parietal lobe dysfunctions that are not associated with other psychotic symptoms.

The phenomenology or type of psychotic symptom does not help distinguish idiopathic from secondary etiologies. However, once the suspicion of a secondary etiology has arisen, the specific psychotic presentation may suggest a particular brain region or direction for further investigation. Table 10-47 lists a number of specific psychotic symptoms that have been consistently associated with disease in particular brain regions. *First-rank symptoms,* originally described by Kurt Schneider as pathognomonic symptoms of schizophrenia, are now accepted as nonspecific psychotic symptoms occurring in all psychotic disorders. Although nonspecific for diagnosis, they have been associated with abnormalities in the left temporal lobe. Complex delusions have been associated with lesions in subcortical regions. Simple persecutory ideas are more common than complex or systematized delusions in patients with significant cognitive deficits. Patients apparently require a variety of intact intellectual abilities (and presumably underlying brain substrate) in order to produce psychotic symptoms of greater complexity. *Anton's syndrome* refers to denial of blindness, classically described in patients with acquired cortical blindness arising from bilateral occipital cortex damage. More recently, it has been described in patients with peripheral optic neuropathy, suggesting that the syndrome may be a variant of the other denial-of-deficit syndromes, such as anosognosia. *Misidentification syndromes* have been described primarily in idiopathic psychotic disorders, although recent studies have pointed to nondominant parietal and frontal lesions as the basis for many. One recent neuropsychological theory proposes that the right hemisphere plays a role in the appreciation of the individuality or uniqueness of people, places, and objects and that lesions in the right hemisphere can result in delusions of misidentification.

Symptoms	Site	Laterality
First-rank symptoms Thoughts spoken aloud Voices commenting Third-person voices arguing Made actions Made feelings Thought withdrawal Thought diffusion Delusional perception	Temporal lobe	Dominant hemisphere
Complex delusions	Subcortical or limbic	
Anton syndrome	Occipital lobe, optic tract	Bilateral
Anosognosia	Parietal lobe	Nondominant hemisphere
Misidentification syndromes Capgras syndrome Reduplicative paramnesia Fregoli syndrome Intermetamorphosis syndrome	Parietal, temporal, frontal lobes	Nondominant hemisphere, bilateral

Table 10-47 Psychotic Symptoms Associated With Abnormality of Specific Brain Regions

Course and Prognosis The course and prognosis of secondary psychotic syndromes depend on their etiology. Vivid psychotic symptoms arising from head trauma may improve dramatically during recovery. Delusions associated with degenerative diseases may diminish as the disease worsens, for the capacity to generate those more complex cognitions is gradually lost. Some secondary psychotic disorders improve with treatment of the underlying disorder, such as the interictal psychosis of epilepsy, which often improves with the pharmacological or surgical control of seizures. Psychotic disorders secondary to infectious disease may not improve, despite eradication of the infectious organism, because of irreversible tissue damage sustained during the acute infection.

Treatment The principles of treatment for a secondary psychotic disorder are similar to those for any secondary neuropsychiatric disorder, namely, rapid identification of the etiological agent and treatment of the underlying cause. Antipsychotics medications afford empirical symptomatic treatment for the psychotic symptoms, although secondary psychotic disorders often prove more refractory than idiopathic disorders to such treatment. Patients with primary systemic or cerebral diseases frequently are more vulnerable to the untoward adverse effects of antipsychotic drugs. To date, there has been insufficient use of serotonin-dopamine antagonists,

such as clozapine (Clozaril), to judge their utility with these conditions.

Anxiety Disorder Due to a General Medical Condition Secondary anxiety syndromes are characterized by prominent anxiety symptoms that are thought to be the direct physiological effect of a specific physical illness or agent. Those disorders have received even less careful scrutiny than secondary mood disorders; therefore, the qualifications made in the section above apply equally or more so to the following discussion.

Definition The key feature of anxiety disorder due to a general medical condition is the presence of prominent anxiety symptoms, which may include generalized anxiety, panic attacks, obsessions, compulsions, or phobias and which are judged to be caused by either an Axis III condition or by substance intoxication or withdrawal. In addition, the anxiety symptoms are not thought to be better explained by another mental disorder (e.g., the anxiety that can be seen in delirium or adjustment disorder with anxious mood). The diagnostician is asked to specify if the anxiety syndrome includes generalized anxiety, panic attacks, obsessive-compulsive symptoms, or phobias.

History and Comparative Nosology For centuries clinicians have described anxiety symptoms as prominent features in a variety of conditions; for most of the twentieth century many of those descriptions focused on patients with endocrinopathies, neurological illnesses, mitral valve prolapse, and substance-related states. The formal concept of any organic mental disorder other than cognitive disorders was introduced by DSM-III; however, organic anxiety disorder was not presented as a distinct entity until DSM-III-R. DSM-III-R limited the diagnosis to either generalized anxiety or panic attacks; DSM-IV broadens the possible related phenomena to include obsessions and compulsions.

Secondary anxiety syndromes have received little study. There are numerous descriptions of anxiety symptoms associated with medical illness or substance-related states, but the operationalization of “secondariness” is generally absent. Further, most studies have included patients with generalized anxiety or panic symptoms; reports of secondary obsessive-compulsive phenomena are few.

Epidemiology The prevalence of anxiety symptoms is high in general medical patients and in patients with many of the specific medical illnesses that are putative potential causes for secondary anxiety syndromes. However, the incidence and prevalence of secondary anxiety disorders, obtained from well-operationalized criteria for syndromic and etiological diagnosis, are not known. Similarly, rates of prior anxiety disturbances or of a family history for anxiety disorders are not known.

Etiology The list of potential causes for anxiety syndromes is long, with nearly complete overlap with the potential causes for mood syndromes. Causes most commonly described in anxiety syndromes include substance-related states (intoxication with caffeine, cocaine, amphetamines, and other sympathomimetic agents; withdrawal from nicotine, sedative-hypnotics, and alcohol), endocrinopathies (especially pheochromocytoma, hyperthyroidism, hypercortisolemic states, and hyperparathyroidism), metabolic derangements (e.g., hypoxemia, hypercalcemia, and hypoglycemia), and neurological disorders (including vascular, trauma, and degenerative). Many of these conditions are either inherently transient or easily remediable. Whether that reflects the pathophysiology of secondary anxiety or is an artifact of reporting (e.g., anxiety with subacute onset and complete resolution after removal of a pheochromocytoma is more likely to be reported as an example of anxiety due to a medical illness than is chronic anxiety in the context of chronic obstructive pulmonary disease) is not known. Much attention has been paid to the association of panic attacks and mitral valve prolapse. The nature of that association is unknown, and therefore the diagnosis of panic attacks secondary to mitral valve prolapse currently is premature. Interestingly, several recent reports have sought to tie obsessive-compulsive symptoms to the development of pathology in the basal ganglia.

Diagnosis and Clinical Features The symptoms of secondary anxiety disorders are by definition phenomenologically similar to those found in the corresponding primary anxiety disorder (e.g., panic attacks and obsessions) (see Table 15.6-17). It is not known if certain symptoms are seen more commonly in the secondary variety; presumably the rate of co-occurrence may vary, depending on the specific etiology of the secondary disorder. As with all secondary syndromes, associated clinical phenomena may include other manifestations of the cause of the secondary anxiety disorder, such as soft neurological signs or subtle cognitive impairment (which may have been used to support the assessment of the anxiety symptoms as being secondary in origin).

There are no specific tests to confirm the diagnosis of secondary anxiety disorder, and little is known about how neurobiological abnormalities seen in primary anxiety disorders differ in secondary syndromes. Physical (including neurological) examination and specific laboratory tests or procedures may be necessary to establish the presence of the etiological disease state.

Differential Diagnosis As for other secondary disorders, two broad domains of differential diagnosis must be considered to establish the presence of a secondary anxiety disorder. The first is phenomenological: does the patient have clinically significant anxiety, panic attacks, obsessions, or compulsions, along with an absence of evidence for another primary or secondary psychiatric syndrome? The second is etiological: does the patient have an Axis III condition, or a state of substance intoxication or withdrawal, that is causing the phenomenology? As always, establishing the causal relationship may be difficult.

Course and Prognosis Little information is available on the course of secondary anxiety disorders. The outcome presumably depends on the specific etiology; thus, anxiety due to hyperthyroidism may well remit with treatment of the hyperthyroid state whereas anxiety due to cardiomyopathy with a low-output state may run a more chronic course.

Treatment Well-designed treatment studies of carefully described patients with secondary anxiety disorders are lacking. Aside from treating the underlying causes, clinicians have found benzodiazepines to be helpful in decreasing anxiety symptoms; supportive psychotherapy (including psychoeducational issues focusing on the diagnosis and prognosis) may also be useful. The efficacy of other, more specific therapies in secondary syndromes (e.g., antidepressant medications for panic attacks, SSRIs for obsessive-compulsive symptoms, behavior therapy for simple phobias) is unknown.

Sleep Disorder Due to a General Medical Condition Sleep disorders can result from a diversity of causes, among them stressful life circumstances, crossing time zones, pulmonary or laryngeal structural abnormalities, systemic diseases (e.g., renal failure), or primary cerebral pathology. However, many sleep disorders, such as narcolepsy, sleep terrors, and enuresis, are idiopathic and occur without known systemic or central abnormalities. The epidemiology of secondary sleep disorders has not been studied systematically.

Definition and Diagnosis Sleep disorders can manifest in four ways: by an excess of sleep (*hypersomnia*), by a deficiency of sleep (*insomnia*), by abnormal behavior or activity during sleep (*parasomnia*), and by a disturbance in the timing of sleep (*circadian rhythm sleep disorders*). Primary sleep disorders occur unrelated to any other medical or psychiatric illness. The DSM-IV nosology is deliberately simple and nondetailed. The patient is assigned to broad categories based on presenting symptoms and the etiological consideration of primary versus secondary disorder (see Table 21-17). The *International Classification of Sleep Disorder* is a more comprehensive and detailed nosology that requires the usage of polysomnography for many of the diagnoses (see Table 21-3).

Etiology and Differential Diagnosis Table 10-48 lists a number of conditions in which a disturbance of sleep has been frequently and characteristically described, allowing conditions to be designated as causes of secondary sleep disorder. *Parkinsonism,* related to either idiopathic Parkinson's disease, medications, or head trauma, frequently results in a secondary sleep disorder. As many as 75 percent of patients with Parkinson's disease complain of sleep disturbance, usually frequent awakenings during sleep. The difficulty maintaining sleep can have a number of causes. Sleep is fragmented owing to the brain degeneration that disrupts the neurophysiological and neurochemical pathways of sleep. In addition, the symptoms of Parkinson's disease can disrupt sleep. Although tremor is diminished during sleep, muscular rigidity is increased and can prevent the patient from turning or finding a comfortable position, resulting in arousal and awakening. Medications used to treat Parkinson's disease can disrupt sleep. Levodopa preparations frequently cause disruptive dreams and nightmares and may also increase nocturnal myoclonus, repetitive, brief leg jerks that awaken the patient and fragment sleep. Levodopa can be stimulating and may prevent the initiation of sleep if taken close to bedtime. Dementia due to degenerative disease can impinge on sleep in a manner similar to parkinsonism, with the degeneration of pathways vital for normal sleep. *Sundowning*, the emergence of severely disruptive behavior, such as agitation and paranoia, at night, associated with the inability to maintain sleep, is a major management issue in the home care of patients with dementia. The pathophysiology is unknown at present, although some have speculated that sundowning is a nocturnal delirium secondary to degeneration of the suprachiasmatic nucleus. Alternatively, sundowning is viewed as a disruption of circadian rhythms, rapid eye movement (REM) parasomnias, or simply postawakening confusion during which the patient with dementia is unable to distinguish between dreams and current reality. Dementia of the Alzheimer's type is accompanied by an exaggeration of the sleep changes associated with normal aging, with a decrease in total sleep time as well as in slow wave and REM sleep. The sleep disturbances worsen as the disease progresses.

Condition	Sleep Symptoms
Parkinsonism	Frequent awakenings, disturbance of circadian rhythms
Dementia	Sundowning, frequent awakenings
Epilepsy	Difficulty initiating sleep, frequent awakenings, parasomnias
Cerebrovascular disease	Difficulty initiating sleep, frequent awakenings
Huntington's disease	Frequent awakening
Kleine-Levin syndrome	Hypersomnia
Uremia	Restless legs, nocturnal myoclonus

Table 10-48 Medical Conditions Commonly Associated With a Secondary Sleep Disorder

Epilepsy can be a true sleep disorder. Most seizure disorders are activated by sleep or arousal from sleep. Both local and generalized epilepsy can occur during sleep, resulting in difficulties maintaining sleep. Seizures may manifest as parasomnias, such as night terrors, sleepwalking, or head banging, although most parasomnias are not related to epilepsy.

Cerebrovascular disorders can impinge on the initiation and maintenance of sleep. No specific lesions have been consistently correlated with a particular sleep disturbance, although brainstem lesions in general are apt to disrupt sleep architecture.

In *Huntington's disease* patients experience frequent awakenings and decreased total sleep time, a pattern common to many subcortical dementia syndromes. With the progression of the disease the movement disorder may manifest during sleep, further disrupting sleep.

Chronic renal failure, anemia, and diabetes mellitus can cause nocturnal myoclonus and the *restless legs syndrome*. The latter is characterized by the experience of deep pains in the lower calf, prompting the patient to keep the legs in constant motion and impinging severely on the ability to initiate sleep.

Kleine-Levin syndrome is a rare disorder characterized by hypersomnia, compulsive eating, sexual disinhibition, personality change, and psychosis. There is a 3 to 1 male-to-female predominance, with onset of symptoms typically occurring in adolescence. Hypersomnia is marked and is the most consistent feature. Compulsive eating and sexual disinhibition, such as public masturbation or propositioning of strangers, complete the syndrome. Incomplete or atypical variants are more common than the full syndrome. Irritability is frequent, and hallucinations or affective symptoms may be present. Symptoms last hours to weeks and are cyclical, with a full return to baseline on many occasions. Symptoms recur in a varying frequency of 1 to several months. The syndrome can be preceded by flulike symptoms or head trauma, although the precise etiology and pathophysiology are unknown. Presumably, there is hypothalamic system dysfunction with the manifest disturbances in sleep, eating, and sexual behavior. In most patients the episodes decrease in frequency and eventually disappear entirely.

Treatment The diagnosis of a secondary sleep disorder hinges on the identification of an active disease process known to exert the observed effect on sleep. Treatment first addresses the underlying neurological or medical disease. Symptomatic treatments focus on behavior modifications, such as improvement of sleep hygiene. Pharmacological options may also be used, such as benzodiazepines for restless legs syndrome or nocturnal myoclonus, stimulants for hypersomnia, and tricyclic antidepressant medications for manipulation of REM sleep.

Sexual Dysfunction Due to a General Medical Condition Specific syndromes characterized by sexual dysfunction thought to be physiologically caused by a general medical condition are female or male hypoactive sexual desire disorder, male erectile disorder, dyspareunia, and other male or female sexual dysfunction.

History and Comparative Nosology Numerous medical conditions, medications, and drugs of abuse can affect sexual desire and performance. However, despite the attention psychiatry has paid to presumed psychologically mediated sexual dysfunction, the role of physiological diseases was downplayed in earlier psychiatric diagnostic systems. DSM-III listed only functional sexual dysfunctions. DSM-III-R allowed sexual dysfunctions to be classified as psychogenic only or as due to both biogenic and psychogenic causes, but required purely biogenic syndromes to be coded on Axis III. The inclusion of secondary sexual disorders as Axis I diagnoses in DSM-IV is consistent with that edition's inclusive approach to behavioral syndromes.

Epidemiology Although surveys have repeatedly demonstrated a high prevalence of sexual dysfunctions in the general population, valid data on secondary dysfunctions are lacking. Similarly, certain medications may be associated with specific rates of sexual symptoms, but the percentage of patients with truly secondary syndromes is not known.

Etiology Potential causes of sexual dysfunctions are listed in Table 10-49. The type of sexual dysfunction is affected by the etiology, but specificity is rare; that is, a given etiology may manifest as one (or more than one) of several syndromes. General categories include medications and drugs of abuse, local disease processes that affect the primary or secondary sexual organs, and systemic illnesses that affect sexual organs via neurological, vascular, or endocrinological routes.

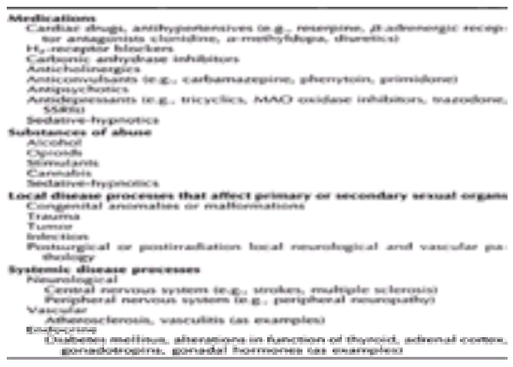

Table 10-49 Causes of Secondary Sexual Dysfunctions

Diagnosis and Clinical Features The clinical features of the sexual dysfunction resemble those of the various primary dysfunctions (see Table 19.1a-17. There may be additional findings due to the underlying disease process. For example, in male erectile disorder due to diabetic autonomic neuropathy, the patient may have symptoms of bowel and bladder autonomic dysfunction as well as evidence of diabetes mellitus itself.

Differential Diagnosis Phenomenology determines the syndromic diagnosis (e.g., erectile dysfunction versus orgasmic disorder). Medical history, physical examination, and relevant laboratory testing are required to demonstrate the presence of physical conditions that are potentially etiological for the sexual dysfunction. However, presence alone does not establish an etiological link. Clinical judgment is necessary and is based on temporal association, assessment of potentially contributory psychosocial factors (or more gross psychopathology), and other factors; the determination of secondary status is often difficult. One exception to that difficulty is male erectile dysfunction. Patients with secondary erectile dysfunction are unable to sustain erections under any circumstances whereas those with primary (i.e., psychogenic) disorders may give a history of variable erectile ability, depending on environment, partner, or other circumstances. If in doubt, a nocturnal penile tumescence study may be helpful because only males with secondary erectile dysfunction will fail to demonstrate tumescence during sleep.

Course and Prognosis The course and prognosis of secondary sexual dysfunctions vary widely, depending on the etiology. Drug-induced syndromes generally remit with discontinuation (or dosage reduction) of the offending agent. Endocrine-based dysfunctions also generally improve with restoration of normal physiology. By

contrast, dysfunctions due to neurological disease may run protracted, even progressive, courses.

Treatment The treatment approach varies widely, depending on the etiology. When reversal of the underlying cause is not possible, supportive and behaviorally oriented psychotherapy with the patient (and perhaps the partner) may minimize distress and increase sexual satisfaction (e.g., example, by developing sexual interactions that are not limited by the specific dysfunction). Support groups for people with specific types of dysfunction are available. Other symptom-based treatments may be used in certain conditions; for example, sildenfanil (Viagra) administration or surgical implantation of a penile prosthesis may be used in the treatment of male erectile dysfunction.

Mental Disorders Due to a General Medical Condition Not Elsewhere Classified DSM-IV has three additional diagnostic categories for clinical presentations of mental disorders due to a general medical condition that do not meet the diagnostic criteria for specific diagnoses. The first of the diagnoses is catatonic disorder due to a general medical condition (Table 10-50). The second diagnosis is personality change due to a general medical condition. The third diagnosis is mental disorder not otherwise specified due to a general medical condition (Table 10-51).

A. The presence of catatonia as manifested by motoric immobility, excessive motor activity that is apparently purposeless and not influenced by external stimuli, extreme negativism or mutism, peculiarities of voluntary movement, or echolalia or echopraxia.
B. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (e.g., a manic episode).
D. The disturbance does not occur exclusively during the course of a delirium.

Coding note: Include the name of the general medical condition on Axis I, e.g., catatonic disorder due to hepatic encephalopathy; also code the general medical condition on Axis III.

Table 10-50 DSM-IV Diagnostic Criteria for Catatonic Disorder Due to a General Medical Condition

This residual category should be used for situations in which it has been established that the disturbance is caused by the direct physiological effects of a general medical condition, but the criteria are not met for a specific mental disorder due to a general medical condition (e.g., dissociative symptoms due to a complex partial seizure).

Coding note: Include the name of the general medical condition on Axis I, e.g., mental disorder not otherwise specified due to HIV disease; also code the general medical condition on Axis III.

Table 10-51 DSM-IV Mental Disorder Not Otherwise Specified Due to a General Medical Condition

Personality Change Due to a General Medical Condition Personality refers to the constellation of enduring traits and behavioral style that essentially defines the person. Personality develops through adolescence and achieves a degree of stability in early adulthood. Both biological disposition as well as environmental factors influence the development of personality. In adults, behavioral style can be described in terms of interests, activities, pleasures, social relations, predominant mood and temperament, standards, usual outlook on life, range of coping mechanisms, and so forth. There is a robust theoretical and clinical literature delineating specific traits, such as self-consciousness, impulsivity, gregariousness, excitement-seeking, openness, and so forth, along dimensions or continua. Standardized measures are available to determine where along the spectrum for each trait a particular patient lies. This provides a personality profile that can be considered relative to standardized norms. The process is quite similar to the dimensional perspective used to assess intelligence and the determination of an I.Q.

The past concept of organic personality syndrome focused on identifying a generic category of particular traits and behaviors associated with brain injury or dysfunction. This conceptual approach has been maintained in DSM-IV, although it sought to base its classification of personality changes solely upon consistently reported behavioral alterations. Suggestions to classify disorders upon anatomical localization (such as frontal lobe syndrome) were rejected. To date there has been little theoretical work attempting to integrate the dimensional perspectives used in the description of normal personality with the categorical approach used in the study of CNS disease and related personality disturbances.

Definition Personality change means that the person's fundamental means of interacting and behaving have been altered; that is, traits that had been regular and consistent over a lifetime have changed. Personality change must be distinguished from the transient disturbances of behavior that frequently occur in reaction to environmental circumstances. When a true personality change occurs in adulthood, the clinician should always suspect brain injury or insult.

History The impact of brain insults on personality has long been recognized. John M. Harlow's description of personality change in Phineas Gage, who sustained a penetrating head injury, remains the classic description:

> He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom), manifesting but little deference for his fellows, impatient of restraint or advice when it conflicts with his desires, at times pertinaciously obstinate, yet capricious and vacillating, devising many plans of future operation, which are no sooner arranged than they are abandoned in turn for others appearing more feasible. A child in his intellectual capacity and manifestations, he has the animal passions of a strong man. . . . In this regard his mind was radically changed, so decidedly that his friends and acquaintances said he was “no longer Gage.”

The frequency of association between brain injury and personality change prompted a search for a generic personality disorder applicable to all brain injury, as well as brain–locale-specific or disease-specific personality disorders. An example of the former is the organic personality disorder found in the earlier versions of DSM. Organic personality disorder was defined as a persistent disturbance of personality due to a specific organic factor involving affective instability, recurrent aggression or rage, impaired social judgment, apathy and indifference, or suspiciousness or paranoid ideation. The interictal personality disorder of temporal lobe epilepsy, characterized by hyperreligiosity, overinclusive speech and behavior, and sexual deviance, was originally presented as a disease-specific personality disorder that was thought to be of high validity. Subsequent studies did not find these traits specific for temporal lobe epilepsy or any other epilepsy. Attempts at defining locale-specific personality disorder have been hampered by difficulties in finding naturalistic human lesions that are indeed localized: Strokes, head trauma, and degenerative diseases, for example, rarely are confined to neat anatomical lobar boundaries.

Nonetheless, the most fruitful approach to delineating personality change disorders has come from the study of frontal lobe injury, where consistent and well-defined traits and behaviors have been associated with particular areas of brain injury. At least two distinct but overlapping secondary personality changes have been identified after injury to the orbitofrontal and frontal convexity areas. Frontal lobe dysfunction may play a key role in all personality and behavioral disturbances because there are vast networks of neural connections between specific areas of the frontal lobe and various limbic and subcortical structures. A similarly complex neuropsychological system suggests that the frontal lobe (more specifically, the prefrontal cortex) modulates many of the basic cognitive, linguistic, attentional, and perceptual processes that originate in other brain areas. Injury to the frontal lobes results in dysfunction in how basic cognitive functions, such as language or memory, are expressed.

Comparative Nosology DSM-I included a category of acute and chronic brain syndromes, defined as disorders due to a diffuse impairment of brain tissue function

from any cause. DSM-II provided basic symptoms for a generic organic brain syndrome, such as impairments in orientation, memory, calculation, learning, and judgment, and lability and shallowness of affect. Although there was no specific category for secondary personality change, it would have been included in the nonpsychotic organic brain syndromes. DSM-III eliminated the unitary organic mental syndrome and allowed for a variety of organic syndromes in which an organic factor was judged etiologically related. Organic personality syndrome nonetheless required at least one of four specific characteristics, including lability, impulsivity, apathy, or suspiciousness; DSM-III-R added recurrent aggression to the list of criteria.

The limitations of the nosology are clear. Personality encompasses a broad range of traits and behaviors not limited to those specified in the organic personality disorder category. The disturbance of personality is identified not from the presence of any particular behavior or trait, but rather as a change from premorbid personality. DSM-IV has dropped the category of organic personality disorder and replaced it with personality change due to a general medical condition. The specific phenomenological criteria were dropped in favor of a general persistent personality disturbance that represents a change from the individual's previous characteristic personality pattern. Subtypes based on the particular phenomenology evident include labile type, disinhibited type, aggressive type, paranoid type, apathetic type, other type, combined type, and unspecified type.

Epidemiology The epidemiological difficulties in ascertaining cases of secondary personality changes are clear: No one particular behavior or trait is diagnostic; rather, a change in a patient's personality structure must be documented. Such documentation often requires recourse to an external informant because patients with personality change are frequently unreliable self-informants. The overinclusive range of personality traits enumerated in previous editions of DSM allowed researchers to pick and choose traits; in addition, the means of measuring them were not consistent from study to study. As a result, reliable incidence and prevalence figures for secondary personality change are not available. Specific personality trait changes for particular brain diseases—for example, passive and self-centered behaviors in dementia of the Alzheimer's type—have been reported; the studies reporting those results, however, have not been replicated, and it remains uncertain how the findings should be applied to other disorders.

Etiology The range of etiologies of secondary personality change is vast and diverse and may involve any of the basic pathological processes described in the previous section. Diseases that preferentially affect the frontal lobes or subcortical structures are more likely to manifest with prominent personality change. Head trauma is a common cause. Strokes involving the anterior communicating or middle cerebral arteries selectively damage frontal lobe structures, often resulting in personality change. The anterior communicating artery is also a common site for aneurysms, which can result in secondary personality change. Frontal lobe tumors, such as meningiomas and gliomas, can grow to considerable size before coming to medical attention, as they may be neurologically silent (i.e., without focal signs). Degenerative disorders affecting the frontal lobes can present with personality change long before cognitive symptoms are evident. Among progressive dementia syndromes, especially those with a subcortical pattern of degeneration, such as AIDS dementia complex, Huntington's disease, or progressive supranuclear palsy, significant personality disturbance manifests often. Multiple sclerosis can impinge on the personality, reflecting subcortical white matter degeneration. Exposures to toxins with a predilection for white matter, such as irradiation, may also produce significant personality change disproportionate to the cognitive or motor impairment.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for personality change due to a general medical condition are listed in Table 10-52. The diagnosis of a secondary personality change rests entirely on the history. A clear and detailed description of the patient's premorbid personality must be obtained. This history usually is collected from an external informant who knew the patient at baseline as well as currently. The first task in evaluating the history is to determine whether a change in personality has indeed occurred or whether the current disruptive behaviors represent long-standing traits that have been exacerbated by a change in circumstance. In addition, delirium must be ruled out.


A. A persistent personality disturbance that represents a change from the individual's previous characteristic personality pattern. (In children, the disturbance involves a marked deviation from normal development or a significant change in the child's usual behavior patterns lasting at least 1 year.)
B. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (including other mental disorders due to a general medical condition).
D. The disturbance does not occur exclusively during the course of a delirium and does not meet criteria for a dementia.
E. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify type:
Labile type: if the predominant feature is affective lability
Disinhibited type: if the predominant feature is poor impulse control as evidenced by sexual indiscretions, etc.
Aggressive type: if the predominant feature is aggressive behavior
Apathetic type: if the predominant feature is marked apathy and indifference
Paranoid type: if the predominant feature is suspiciousness or paranoid ideation
Other type: if the predominant feature is not one of the above, e.g., personality change associated with a seizure disorder
Combined type: if more than one feature predominates in the clinical picture
Unspecified type

Coding note: Include the name of the general medical condition on Axis I, e.g., personality change due to temporal lobe epilepsy; also code the general medical condition on Axis III.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. Washington, DC, American Psychiatric Association, Washington, DC, 1994.

Table 10-52 DSM-IV Diagnostic Criteria for Personality Change Due to a General Medical Condition

Once a diagnosis of a personality change has been established, the search for an etiological agent begins. An insidious and progressive course is suggestive of a degenerative process or a neoplasm. An abrupt onset of personality change is more suggestive of a vascular event or trauma. Risk factors for HIV infection should raise the suspicion of HIV or neurosyphilis infection. A complete history of toxic exposures, including alcohol and recreational drug use, environmental or occupational toxin exposures, and medications, should be obtained. The search for a causative agent can be aided by the presence of other evidence of brain dysfunction, such as motor abnormalities and cognitive impairment.

The particular form of a personality change may be helpful in determining the locus of injury or brain dysfunction, although much research remains to be done. The prefrontal cortex is often implicated in secondary personality change disorders. Two frontal lobe personality syndromes have been described, correlating with injury to the orbitofrontal and dorsolateral frontal cortical regions. Table 10-53 outlines the behavioral and personality changes associated with each. The anatomical designations for those syndromes may be misleading because frontal regions form rich neuronal networks with subcortical and limbic structures. Subcortical dementia also is characterized by significant personality deteriorations, such as apathy, aspontaneity, and slowing. Patients with multiple sclerosis sometimes present with a euphoric personality, which probably reflects a disruption of the orbitofrontal subcortical network.

Orbitofrontal	Frontopolar
Disinhibition	Apathy
Inappropriate jocularity	Indifference
Affective lability	Psychomotor slowing
Impulsivity	Inaction

Table 10-53 Frontal Lobe Personality Change Syndromes

Global degenerative processes, such as dementia of the Alzheimer's type, involve significant personality change that has been less well characterized. In general, there is a coarsening of the personality with loss of subtlety and finesse. An exacerbation of premorbid traits is possible, with a suspicious patient becoming paranoid or a flamboyant patient becoming histrionic. Agitation or aggression is a common concomitant of brain disease. When they occur in a patient with a premorbid history of violence and a short temper, it may be difficult to determine if a secondary personality change has occurred, even when CNS dysfunction is evident.

Laboratory evaluation for secondary personality change is the same as for other secondary disorders. The most important element is informed clinical suspicion

regarding specific disease processes.

Differential Diagnosis Secondary personality change must be differentiated from adjustment disorders occurring, for example, in response to environmental stressors or major medical disorders. Apathetic and amotivational symptoms in patients with dorsolateral frontal lesions may be mistaken for major depressive disorder. The former can be distinguished by a lack of pervasive dysphoria, intact neurovegetative function, and the absence of self-disparagement and hopelessness. Euphoria and disinhibition with the orbitofrontal syndrome may be ascribed to mania. The orbitofrontal syndrome, however, does not display heightened motor activity, excessive energy, and disrupted sleep; neither does it follow the cyclical course of bipolar I disorder but rather produces a persistent and consistent clinical picture.

Course and Prognosis The course of and prognosis of secondary personality syndromes depend on the course of the etiological systemic or cerebral disorder. Personality change secondary to mass lesions or hydrocephalus can improve dramatically with surgery, chemotherapy, or radiation therapy. However, each of these treatments may result in a different personality change syndrome. Personality change secondary to head trauma may improve slowly and gradually over the course of months or years, although residual disturbances may remain. Personality change due to degenerative processes can be most disruptive early in the disease process when the patient retains a measure of volition and control of motor capacities. Ironically, management of such patients may ease as the disease progresses, when the personality evolves into greater apathy, unresponsiveness, and akinesia. Personality change associated with epilepsy can improve dramatically with seizure control by pharmacotherapy or surgery.

Treatment Treatment for secondary personality syndromes is first directed toward correcting the underlying etiology. Symptomatic treatments as a group have been marginally effective at best. Lithium carbonate, carbamazepine, and valproic acid have been used for the control of affective lability and impulsivity. Aggression or explosiveness may be treated with lithium, anticonvulsant medications, or a combination of lithium and an anticonvulsant agent. Centrally active b-adrenergic receptor antagonists, such as propranolol (Inderal), have some efficacy as well. Antipsychotic medications are no more effective in the dampening of aggression than the previously mentioned agents, induce greater discomfort, and introduce the risk of tardive dyskinesia. Apathy and inertia have occasionally improved with psychostimulant agents. Because cognition and verbal skills may be preserved in patients with secondary personality changes, they may be candidates for psychotherapy. Families should be involved in the therapy process, with a focus on education and understanding the origins of the patient's inappropriate behaviors and coarsening. Issues such as competency, disability, and advocacy are frequently of clinical concern in those patients in light of the unpredictable and pervasive behavior change.

SUGGESTED CROSS-REFERENCES

A discussion of psychiatric clinical manifestations of specific neurological and systemic disorders appears in Chapter 2 on neuropsychiatry and behavioral neurology. Neuropsychological and intellectual assessment of adults is presented in Section 7.3, assessment of children in Section 7.5, and medical assessment and laboratory testing in Section 7.6. Discussions of substance-related disorders appear in Chapter 11, schizophrenia in Chapter 12, psychotic disorders in Chapter 13, anxiety disorders in Chapter 15, factitious disorders in Chapter 17, dissociative disorders (including dissociative amnesia) in Chapter 18, sexual dysfunctions in Section 19.1a on normal human sexuality and sexual dysfunctions, sleep disorders in Chapter 21, and personality disorders in Chapter 24. Primary care psychiatry is presented in Section 28.1, and the psychiatric aspects of HIV infection and AIDS in Section 2.8. Physiological aspects of normal aging (including age-related cognitive decline) is discussed in Section 51.2c and dementia of the Alzheimer's type and other dementing disorders of late life are discussed in Section 51.3e.

CHAPTER REFERENCES

Alexander MP: Traumatic brain injury. In *Psychiatric Aspects of Neurologic Disease*, vol II, D Benson, D Blumer, editors. Grune & Stratton, New York, 1982.

American Academy of Neurology: Nomenclature and research case definitions for neurologic manifestations of human immunodeficiency virus–type 1 (HIV-1) infection. Neurology *41:*778, 1991.

Avoli M: Molecular mechanisms of antiepileptic drugs. Sci Med *4*(4):54, 1997.

*Bradford Hill A: The environment and disease: Association or causation? Proc R Soc Med *58:*295, 1965.

Burns A, Levy R: *Dementia*. Chapman & Hall Medical, London, 1994.

Caine ED, Joynt RJ: Neuropsychiatry. . . again. Arch Neurol *43:*325, 1986.

*Collins S, Law MG, Flecher A, Boyd A, Kaldor J, Masters CL: Surgical treatment and risk of sporadic Creutzfeldt-Jacob disease: A case-control study. Lancet *353:*693, 1999.

Cummings JL: *Clinical Neuropsychiatry*. Grune & Stratton, New York, 1985.

Damasio H, Grabowski T, Frank R, Galaburda AM, Damasio AR: The return of Phineas Gage: Clues about the brain from the skull of a famous patient. Science *264:*1102, 1994.

Evans AS: Causation and disease: A chronological journey. Am J Epidemiol *108:*249, 1975.

Gerard EM, Spitz MC, Tobin JA, Shantz D: Subacute postictal aggression. Neurology *50:*384, 1998.

*Grant I, Adams K: *Neuropsychological Assessment of Neuropsychiatric Disorders*, ed 2. Oxford University Press, New York, 1996.

Grant I, Martin A: *Neuropsychology of HIV Infection*. Oxford University Press, New York, 1994.

Harlow JM: Recovery after severe injury to the head. Publ Mass Media Soc *2:*327, 1868.

Huntington's Disease Collaborative Research Group: A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. Cell *72:*971, 1993.

*Janicki MP, Dalton AJ, editors: *Dementia, Aging, and Intellectual Disabilities: A Handbook*. Brunner/Mazel, Philadelphia, 1999.

Jaspers K: *General Psychopathology*, J Hoenig, MW Hamilton, translators. University of Chicago Press, Chicago, 1963.

*Jorm AF: *The Epidemiology of Alzheimer's Disease and Related Disorders*. Chapman & Hall, London, 1990.

Joseph AB, Young RR, editors: *Movement Disorders in Neurology and Neuropsychiatry*. Blackwell Scientific, Boston, 1992.

Kelly JP, Rosenberg JH: Diagnosis and management of concussion in sports. Neurology *48:*575, 1997.

Krauthammer C, Klerman GL: Secondary mania. Arch Gen Psychiatry *35:*1333, 1978.

*Lipowski ZJ: *Delirium—Acute Confusional States*, ed 2. Oxford University Press, New York, 1990.

Kremer B, Goldberg P, Andrew SE, Theilmann J, Telenius H, Zeisler J, Squittieri F, Lin B, Bassett A, Almqvist E, Bird TD, Hayden MR: A worldwide study of the Huntington's disease mutation: The sensitivity and specificity of measuring CAG repeats. N Engl J Med *330:*1401, 1994.

Lezak MD: *Neuropsychological Assessment*, ed 3. Oxford University Press, New York, 1995.

Liptzin B, Levkoff SE, Gottlieb GL, Johnson JC: Delirium. J Neuropsychiatry Clin Neurosci *5:*154, 1993.

*Lishman WA: *Organic Psychiatry: The Psychological Consequences of Cerebral Disorder*, ed 2. Blackwell Scientific, London, 1987.

Lyness JM, editor: *Affective Disorders from Medical Conditions*. Semin Clin Neuropsychiatry *2:*228, 1997.

Marx OM: Nineteenth-century medical psychology. Isis *61:*355, 1970.

McAllister TW, Green RL, editors: *Neurobehavioral Consequences of Traumatic Brain Injury*. Semin Clin Neuropsychiatry *3:*160, 1998.

Miller NE, Lipowski ZJ, Lebowitz BD, editors: Delirium: Advances in research and clinical practice. Int Psychogeriatr *3:*97, 1991.

Minden SL, Schiffer RB: Affective disorders in multiple sclerosis: Review and recommendations for clinical research. Arch Neurol *47:*98, 1990.

O'Donoghue JL, editor: *Neurotoxicity of Industrial and Commercial Chemicals*. CRC Press, Boca Raton, FL, 1985.

Popkin MK: “Secondary” and drug-induced mood, anxiety, psychotic, catatonic, and personality syndromes: A review of the literature. J Neuropsychiatry Clin Neurosci *4:*369, 1992.

Restak R, editor: *Neuropsychiatry of Minor Head Injury*. Semin Clin Neuropsychiatry *2:*160, 1997.

Reynolds EH: Structure and function in neurology and psychiatry. Br J Psychiatry *157:*481, 1990.

Robinson RG, Starkstein SE: Current research in affective disorders following stroke. J Neuropsychiatry Clin Neurosci *2:*1, 1990.

Salloway S, Malloy P, Cummings J, editors: *The Neuropsychiatry of Limbic and Subcortical Disorders*. J Neuropsychiatry Clin Neurosci *9:*313, 1997.

Schaumburg HH, Spencer PS: Recognizing neurotoxic disease. Neurology *37:*276, 1987.

Shorter E: *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac* . Wiley, New York, 1997.

Silver JM, McAllister TW: Forensic issues in the neuropsychiatric evaluation of the patient with mild traumatic brain injury. Neuropsych Pract Opin *9:*102, 1997.

Slater E, Beard AW: The schizophrenia-like psychoses of epilepsy. Psychiatric aspects. Br J Psychiatry *109:*95, 1963.

Terry RD, Katzman R, Bick KL: *Alzheimer Disease*. Raven, New York, 1994.

White BV: *Stanley Cobb: A Builder of the Modern Neurosciences*. Francis A. Countway Library of Medicine, Boston, 1984.

Zegans LS, Coates TJ, editors: *Psychiatric Manifestations of HIV Disease*, vol 17. Saunders, Philadelphia, 1994.

CHAPTER 11. SUBSTANCE-RELATED DISORDERS

11.1 INTRODUCTION AND OVERVIEW

JEROME H. JAFFE, M.D.

Whether a society views substance use primarily as a moral or a legal problem, when it creates difficulties for the user or ceases to be entirely volitional it becomes the concern of all the helping professions, including psychiatry. This chapter on substance-related disorders is made up of separate sections organized around the syndromes engendered by the use of each of the major groups of pharmacological agents that are commonly misused (abused). This section deals with issues that are common across categories of drugs—the nomenclature and diagnostic schemes of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of *International Classification of Diseases and Related Health Problems* (ICD-10), the history of substance use and dependence, epidemiology, and the etiological factors and treatment principles that appear to be common to these syndromes.

General Organization of DSM-IV and ICD-10 DSM-IV includes two broad categories of substance-related disorders: substance use disorders (substance dependence and substance abuse), and a diverse grouping of substance-induced disorders (such as intoxication, withdrawal, psychotic disorder, and mood disorders). Thus, in DSM-IV the topic of substance-related disorders goes beyond substance dependence and abuse and closely related problems to include a wide variety of adverse reactions not only to substances of abuse, but also to medications and toxins. The medications associated with substance-induced disorders range from anesthetics to over-the-counter medications and include such diverse drug categories as anticholinergics, antidepressants, anticonvulsants, antimicrobial drugs, antihypertensive agents, corticosteroids, antiparkinson agents, chemotherapeutic agents, nonsteroidal anti-inflammatory drugs, and disulfiram (Antabuse). In addition, several categories of substance-induced disorders can be associated with a wide range of nonmedicinal toxic materials, ranging from heavy metals and industrial solvents to insecticides and household cleaning agents. DSM-IV groups the diagnostic criteria for substance dependence, abuse, intoxication, hallucinogen persisting perception disorder, and withdrawal syndromes in a section titled “Substance-Related Disorders,” whereas the other substance-related disorders (e.g., substance-induced mood disorders and substance-induced delusional disorders) are described in the sections covering the disorders that they most closely resemble phenomenologically (Table 11.1-1).

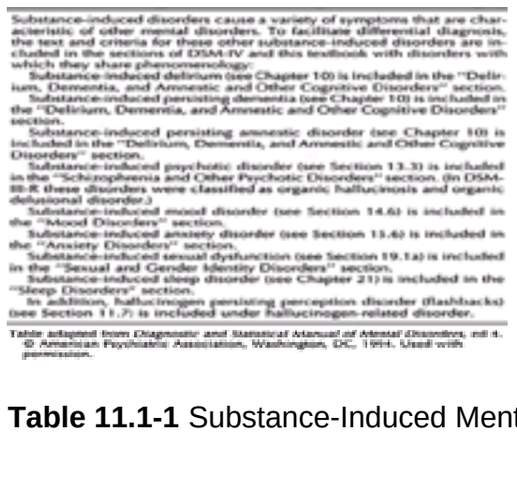

Substance-induced disorders cause a variety of symptoms that are characteristic of other mental disorders. To facilitate differential diagnosis, the text and criteria for these other substance-induced disorders are included in the sections of DSM-IV and this textbook with disorders with which they share phenomenology:

Substance-induced delirium (see Chapter 10) is included in the "Delirium, Dementia, and Amnestic and Other Cognitive Disorders" section.

Substance-induced persisting dementia (see Chapter 10) is included in the "Delirium, Dementia, and Amnestic and Other Cognitive Disorders" section.

Substance-induced persisting amnestic disorder (see Chapter 10) is included in the "Delirium, Dementia, and Amnestic and Other Cognitive Disorders" section.

Substance-induced psychotic disorder (see Section 13.3) is included in the "Schizophrenia and Other Psychotic Disorders" section. (In DSM-III-R these disorders were classified as organic hallucinosis and organic delusional disorder.)

Substance-induced mood disorder (see Section 14.6) is included in the "Mood Disorders" section.

Substance-induced anxiety disorder (see Section 13.4) is included in the "Anxiety Disorders" section.

Substance-induced sexual dysfunction (see Section 19.1a) is included in the "Sexual and Gender Identity Disorders" section.

Substance-induced sleep disorder (see Chapter 21) is included in the "Sleep Disorders" section.

In addition, hallucinogen persisting perception disorder (flashbacks) (see Section 11.3) is included under hallucinogen-related disorder.

Table adapted from *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994. Used with permission.

Table 11.1-1 Substance-Induced Mental Disorders Included Elsewhere in the Textbook

The DSM-IV section dealing with substance dependence and substance abuse presents descriptions of the clinical phenomena associated with the use of 11 designated classes of pharmacological agents: alcohol, amphetamines or similarly acting agents; caffeine; cannabis; cocaine; hallucinogens; inhalants; nicotine; opioids; phencyclidine (PCP) or similar agents; and sedatives, hypnotics, and anxiolytics. A residual twelfth category includes a variety of agents, such as anabolic steroids and nitrous oxide, that are not in the 11 designated classes.

ICD-10 considers the disorders due to psychoactive substance use within the confines of an alphanumeric system that allows only nine categories of pharmacological agents, with one residual category to cover both multiple drug use and use of psychoactive substances not included in the nine designated categories. DSM-IV and ICD-10 categorize substances comparably, with the following exceptions. Caffeine and PCP are considered distinct categories in DSM-IV; whereas in ICD-10, problems related to caffeine are included in the category of other stimulants such as amphetamine, and phencyclidine must be included with hallucinogens or in the residual category. Also, ICD-10 has a special category for abuse of non–dependence-producing substances (Table 11.1-2). Specifically mentioned are antidepressants, analgesics, antacids, vitamins, and steroids or hormones.

A wide variety of medicaments and folk remedies may be involved, but the particularly important groups are psychotropic drugs that do not produce dependence, such as antidepressants, laxatives, and analgesics that may be purchased without medical prescription, such as aspirin and paracetamol. Although the medicines may have been originally prescribed or recommended in the first instance, prolonged, unnecessary, and often excessive dosage develops, which is facilitated by the availability of the substances without medical prescription.

Persistent and unjustified use of these substances is usually associated with undue energy expense, often involves unnecessary contacts with medical professionals or supporting staff, and is sometimes marked by the harmful physical effects of the substances. Attempts to discourage or forbid the use of the substance are often met with resistance, for laxatives and analgesics this may be in spite of warnings about or even the development of physical harm such as renal dysfunction or ophthalmic disturbance. Although it is usually clear that the patient has a strong motivation to take the substance, no dependence or withdrawal symptoms develop as in the case of the psychoactive substances specified in mental and behavioral disorders due to psychoactive substance use.

Identify the type of substance involved:

Antidepressants
(such as tricyclic and tetracyclic antidepressants and monoamine oxidase inhibitors)

Laxatives

Analgesics
(such as aspirin, paracetamol, phenacetin, not specified as psychoactive substance use)

Antacids

Vitamins

Steroids or hormones

Specific herbal or folk remedies

Other substances that do not produce dependence
(such as glycerol)

Unspecified

This material is reproduced from *International Statistical Classification of Diseases and Related Health Problems*, 10th revision, Volume 1. Geneva, World Health Organization, 1992.

Table 11.1-2 ICD-10 Diagnostic Criteria for Abuse of Non-Dependence-Producing Substances

DEFINITIONS AND DIAGNOSIS

Substance Dependence The revised third edition of DSM (DSM-III-R), DSM-IV, and ICD-10 formulations for substance abuse and dependence closely follow the concepts and terminology developed in 1980 by an International Working Group sponsored by the World Health Organization (WHO) and the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) of the United States, which defined substance dependence as follows:

> A syndrome manifested by a behavioral pattern in which the use of a given psychoactive drug, or class of drugs, is given a much higher priority than other behaviors that once had higher value. The term “syndrome” is taken to mean no more than a clustering of phenomena so that not all the components need always be present or not always present with the same intensity. . . . The dependence syndrome is not absolute, but is a quantitative phenomenon that exists in different degrees. The intensity of the syndrome is measured by the behaviors that are elicited in relation to using the drug and by the other

behaviors that are secondary to drug use. . . . No sharp cut-off point can be identified for distinguishing drug dependence from non-dependent but recurrent drug use. At the extreme, the dependence syndrome is associated with “compulsive drug-using behavior.”

That central notion is continued in DSM-IV, which states:

> The essential feature of dependence is a cluster of cognitive, behavioral, and physiological symptoms indicating that the individual continues substance use despite significant substance-related problems.

The central notion in ICD-10 is virtually the same:

> a cluster of behavioural, cognitive, and physiological phenomena that develop after repeated substance use and typically include a strong desire to take the drug, difficulties in controlling its use, persisting in its use despite harmful consequences, a higher priority given to drug use than to other activities and obligations, increased tolerance, and sometimes a physical withdrawal state.

The DSM-IV and ICD-10 criteria for substance dependence are presented in Table 11.1-3 and Table 11.1-4. DSM-IV uses seven criteria to describe a generic concept of dependence that applies across 11 classes of pharmacological agents. ICD-10 requires that three of six criteria be met and also applies across classes of drugs.

Table 11.1-3 DSM-IV Diagnostic Criteria for Substance Dependence

Table 11.1-4 ICD-10 Diagnostic Criteria for Mental and Behavioral Disorders Due to Psychoactive Substance Use

DSM-IV and ICD-10 use a polythetic syndrome definition, in which no one specific criterion is required so long as three or more are present. However, DSM-IV asks the clinician to specify whether physiological dependence (evidence of criterion 1, tolerance, or criterion 2, withdrawal) is present or absent. Evidence indicates that physiological dependence is associated with a more severe form of the disorder.

In addition to requiring the clustering of three criteria in a 12-month period, DSM-IV includes a few other qualifications. It states specifically that the diagnosis of dependence can be applied to every class of substances except caffeine. That point is admittedly controversial, and some researchers believe, on the basis of the same DSM-IV generic criteria, that caffeine produces a distinct form of dependence, although it is relatively benign for most persons.

Some persons use several categories of drugs and are clearly drug dependent, according to the generic criteria, but it may not be possible to ascertain whether they are dependent on any one specific class of drugs. When at least three groups of substances are involved, DSM-IV calls the condition polysubstance dependence (Table 11.1-5). DSM-IV also makes provision for classifying substance-related disorders that cannot be classified in any of the previous categories (e.g., nitrous oxide, anticholinergics, anabolic-androgenic steroids) or for an initial diagnosis of dependence or abuse when the specific substance is not known. A similar residual category is included in ICD-10, but steroids are given a distinct code. The DSM-IV diagnostic criteria for other (or unknown) substance-related disorders are listed in Table 11.1-6.

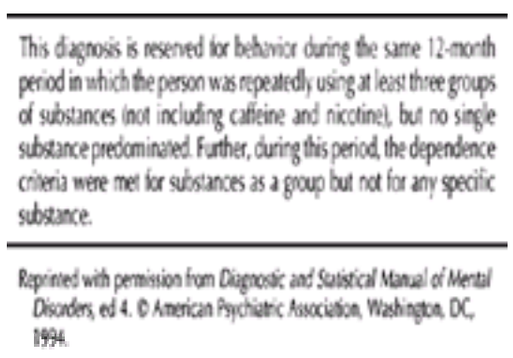

Reprinted with permission from *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.1-5 DSM-IV Diagnostic Criteria for Polysubstance Dependence

Table 11.1-6 DSM-IV Diagnostic Criteria for Other (or Unknown) Substance-Related Disorders

Patterns of Remission and Course Specifiers DSM-IV and ICD-10 deal with remission by providing distinct modifying terms that can be appended to a diagnosis of substance dependence. DSM-IV terms are more varied than those of ICD-10 (Table 11.1-7). The DSM-IV course specifiers require a period of at least 1 month, after a period of active dependence, during which no criteria of dependence are present. If a patient has not met any criteria for dependence for at least 1 month but for less than 12 months, the course specifier to use is early full remission. If the period during which no criteria of dependence are met exceeds 12 months, the specifier of sustained full remission can be used. If the full criteria for dependence or abuse have not been met for less than a year, but one or more criteria have been present, early partial remission may be used. If the period exceeds 12 months, sustained partial remission may be used. Two additional remission specifiers should be used when appropriate: “on agonist therapy (includes partial agonists)” and “in a controlled environment.” Several factors, such as duration of remission and duration of period of dependence, must be considered in deciding that a person has fully recovered and no longer warrants a diagnosis of dependence. The modifiers that describe the course of dependence in ICD-10 are similar, but specific criteria for selecting them are not provided (Table 11.1-4).

Table 11.1-7 DSM-IV Course Modifiers for Substance Dependence

Substance Abuse DSM-IV defines the essential features of substance abuse as follows:

> A maladaptive pattern of substance use manifested by recurrent and significant adverse consequences related to the repeated use of substances. . . . These problems must occur recurrently during the same 12-month period. . . . [T]he criteria for Substance Abuse do not include tolerance, withdrawal, or a pattern of compulsive use and instead include only the harmful consequences of repeated use. A diagnosis of Substance Abuse is preempted by the diagnosis of Substance Dependence if the individual's pattern of substance use has ever met the criteria for Dependence for that class of substances.

The DSM-IV criteria for substance abuse are shown in Table 11.1-8.

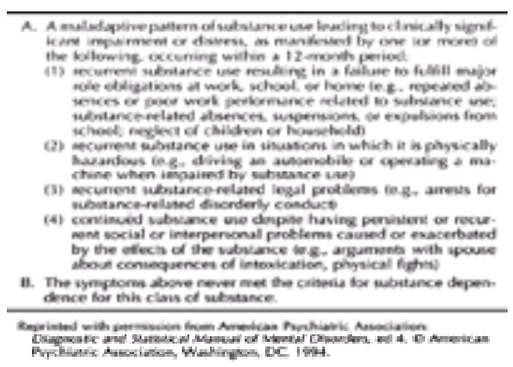

A. A maladaptive pattern of substance use leading to clinically significant impairment or distress, as manifested by one (or more) of the following, occurring within a 12-month period:
(1) recurrent substance use resulting in a failure to fulfill major role obligations at work, school, or home (e.g., repeated absences or poor work performance related to substance use; substance-related absences, suspensions, or expulsions from school; neglect of children or household)
(2) recurrent substance use in situations in which it is physically hazardous (e.g., driving an automobile or operating a machine when impaired by substance use)
(3) recurrent substance-related legal problems (e.g., arrests for substance-related disorderly conduct)
(4) continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance (e.g., arguments with spouse about consequences of intoxication, physical fights)
B. The symptoms above never met the criteria for substance dependence for this class of substance.

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.1-8 DSM-IV Diagnostic Criteria for Substance Abuse

A major difference exists between DSM-IV and ICD-10 with regard to the diagnosis of substance abuse. ICD-10 does not use the term “abuse.” Instead, it includes a category of *harmful use*, which substantially differs from the DSM-IV concept of “abuse.” The concept of “harmful use” is limited to mental and physical health (e.g., hepatitis and overdose, or episodes of depressive disorder resulting from heavy alcohol use). The concept specifically excludes social impairment, stating: “The fact that a pattern of use of a particular substance is disapproved of... or may have led to socially negative consequences such as arrest or marital arguments is not in itself evidence of harmful use.” Four diagnostic criteria must be met to make the ICD-10 diagnosis of harmful use.

Substance Withdrawal *Substance withdrawal*, as used in DSM-IV, is a diagnostic term rather than a technical term. Thus minor symptoms that technically are due to cessation of substance use (e.g., the coffee drinker's early morning precoffee lethargy or minor headache) would not by themselves fulfill the criteria for substance withdrawal, unless they are accompanied by a maladaptive behavior change and cause some clinically significant distress or impairment in social, occupational, or other important area of functioning. DSM-IV does not recognize withdrawal from caffeine, cannabis, or PCP, although some observers believe that specific signs and symptoms can be observed when those agents are abruptly discontinued after a period of heavy use. ICD-10 does describe a cannabinoid withdrawal state.

Withdrawal is commonly, but not invariably, associated with substance dependence. The signs and symptoms of withdrawal vary with the specific class of drug. In general, the severity of withdrawal is related to the amount of substance used and the duration and patterns of use. Withdrawal is seen not only when substance use is stopped but also when reduced use of a substance or a change in metabolism results in lower tissue levels. The DSM-IV generic criteria for substance withdrawal are shown in Table 11.1-9; the ICD-10 general criteria are shown in Table 11.1-4. Specific diagnostic criteria for withdrawal from each category of drugs, to be used when the general criteria have been met, are also provided.

A. The development of a substance-specific syndrome due to the cessation of (or reduction in) substance use that has been heavy and prolonged.
B. The substance-specific syndrome causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Table 11.1-9 DSM-IV Diagnostic Criteria for Substance Withdrawal

Substance Intoxication Substance intoxication is defined more narrowly in DSM-IV than it might be in a pharmacology text. A variety of substances may produce unwanted physiological or psychological effects that could be construed as substance intoxication effects (e.g., excessive sleepiness following use of an antihistamine), but unless the symptoms are associated with maladaptive behavior, the effects would not constitute substance-induced intoxication as defined in DSM-IV. Furthermore, whether a behavioral effect is maladaptive depends on the social and environmental context in which it occurs. If alcohol makes a person unusually sociable, a bit garrulous, and a little uncoordinated at a family celebration this is probably not maladaptive drinking behavior, the same behavior at a formal business meeting probably is. Similarly, ICD-10 specifies that intoxication must produce disturbances in the level of consciousness, cognition, perception, affect, or behavior that are of clinical importance. However, it requests clinicians to further specify which of several common complications of intoxication (e.g., trauma, delirium, convulsions) are also present. The DSM-IV general criteria for substance intoxication are shown in Table 11.1-10. In addition, ICD-10 provides specific sets of diagnostic criteria for each of the drug categories and for multiple drugs, to be used once the generic criteria for intoxication have been met. Also shown are the additional specifiers for complications of intoxication (Table 11.1-4).

A. The development of a reversible substance-specific syndrome due to recent ingestion of (or exposure to) a substance. **Note:** Different substances may produce similar or identical syndromes.
B. Clinically significant maladaptive behavioral or psychological changes that are due to the effect of the substance on the central nervous system (e.g., belligerence, mood lability, cognitive impairment, impaired judgment, impaired social or occupational functioning) and develop during or shortly after use of the substance.
C. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Table 11.1-10 DSM-IV Diagnostic Criteria for Substance Intoxication

Substance-Induced Disorders In addition to dependence, abuse, intoxication, and withdrawal, the use of certain psychoactive drugs can induce syndromes that used to be called organic mental disorders. To avoid implying that other psychiatric disorders do not have an organic basis, DSM-IV designates these syndromes substance-induced disorders and recognizes the following categories: substance intoxication, substance withdrawal, substance-induced withdrawal delirium, substance-induced intoxication delirium, substance-induced persisting dementia, substance-induced persisting amnestic disorder, substance-induced mood disorder, substance-induced anxiety disorder, substance-induced psychotic disorder, substance-induced sexual dysfunction, and substance-induced sleep disorder.

In recording a diagnosis of a substance-related disorder, the clinician should indicate the specific agent causing the disorder, if known, rather than the broad drug category; that is, substance-induced intoxication, pentobarbital (Nembutal) rather than substance-induced intoxication, sedative-hypnotics. However, the diagnostic code should be selected from the list of classes of substances provided in sets of criteria for the substance-induced disorder being recorded. For each of the substance-induced disorders (other than intoxication and withdrawal), the clinician is asked to specify whether the onset was during intoxication or during withdrawal. Thus, a specific substance-induced disorder would have a three-part name delineating (1) the specific substance, (2) the context (whether the disorder occurred during intoxication or during withdrawal or occurs or persists beyond those stages), and (3) the phenomenological presentation (e.g., diazepam [Valium]-induced anxiety disorder with onset during withdrawal).

Table 11.1-11 shows the various disorders induced by the major categories of substances recognized by DSM-IV and indicates which disorders are seen during intoxication and during withdrawal. Although they are not included specifically in the table, anabolic-adrenergic steroids can also induce psychotic mood, anxiety, and sleep and sexual disorders, and their withdrawal can also be associated with mood and sleep disorders. ICD-10 has a distinctly different approach to recording these drug-related disorders. With the first and second digits after the letter committed to designating the drug category, additional psychiatric syndromes are indicated by the use of the third and fourth digits. For example, persistent mood disorder associated with hallucinogens is designated F16.72. For the diagnosis to be made, the mood disorder would need to meet the criteria listed for mood (affective) disorders.

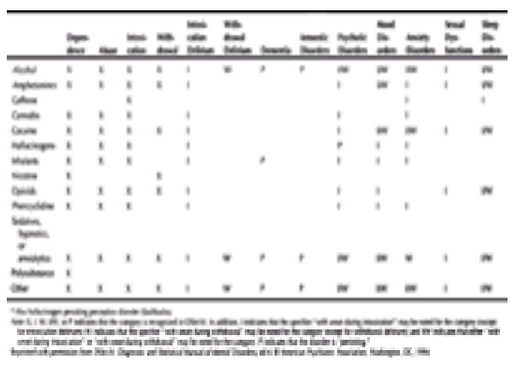

Table 11.1-11 DSM-IV Diagnoses Associated With Class of Substances

Evolving Terminology The terminology used to describe the substance-related disorders has been repeatedly revised as concepts about the nature of drug-using behavior have evolved. In the 1980 third edition of the DSM (DSM-III) drug use disorders were divided into two major categories, drug abuse and drug dependence, and specific criteria for diagnosis were given. In DSM-III-R, adopted in 1987, the two categories were retained, but the diagnostic criteria were modified. Further revisions were made for DSM-IV, which adopted the terms “substance abuse” and “substance dependence,” probably to eliminate the use of the more cumbersome term “alcohol and drug dependence including tobacco.” For similar reasons, ICD-10 adopted the term “psychoactive substance dependence.”

In much of the world literature on drug dependence, the term “dependence” is used to convey two distinct ideas: (1) a behavioral syndrome and (2) physical or physiological dependence. *Physiological dependence* can be defined as an alteration in neural systems that is manifested by tolerance and the appearance of withdrawal phenomena when a chronically administered drug is discontinued or displaced from its receptor. Because the dual use of the word causes confusion, the

1980 ADAMHA-WHO working group recommended restricting use of the term “dependence” to describe the behavioral syndrome and substituting the term “neuroadaptation” for physical dependence. Such a substitution would have emphasized several points. First, the continued use of many drugs, including tricyclic antidepressants and b-adrenergic receptor antagonists, causes neuroadaptive changes followed by withdrawal phenomena, but not by drug-seeking behavior, on their discontinuation. Second, neuroadaptive changes begin with the first dose of an opioid or sedative drug, and therefore, such changes in and of themselves are not a sufficient cause (or definition) of drug dependence as a behavioral syndrome.

Why Use “Addiction”? The words “addict” and “addiction” often have pejorative connotations; they are also frequently trivialized and used to refer to ordinary activities, such as exercising and solving crossword puzzles. However, the term “addiction” continues to have the core connotation of decreased control, and some chapters in this book have retained such terms as “opioid addict” because they are less awkward to use than terms such as “severely opioid-dependent person” when referring to persons who are dependent on drugs to a severe degree. Here the word “dependent,” unmodified, is used to mean behaviorally dependent. The term “physiological dependence” or “physical dependence” is used to refer to the physiological changes that result in withdrawal symptoms when drugs are discontinued.

COMPARATIVE NOSOLOGY

DSM-IV and ICD-10 The generic concept of dependence is virtually identical in DSM-IV and ICD-10. By requiring the clinician to specify whether tolerance and withdrawal are present, DSM-IV appears to recognize a special significance for tolerance and physiological dependence. Some data indicate that among alcoholics the presence of physical dependence and, to a lesser degree, tolerance is associated with a more severe variety of the syndrome. In practice, however, requiring evidence of these criteria would not substantially reduce the number of cases meeting the criteria for dependence in most drug categories, with the exception of hallucinogens, a class of drugs for which DSM-IV does not list physiological dependence as a criterion. There is generally a high level of agreement between DSM-IV and ICD-10 for making a diagnosis of dependence, although the descriptions of the criteria for determining the presence and severity of the syndrome differ. They both require that three elements of the syndrome have been present in a 12-month period. The DSM-IV categorization of drug classes differs somewhat from the one used by ICD-10, which, constrained by a new alphanumeric system, uses only nine drug categories by including caffeine with amphetamine-like stimulants and PCP with other psychoactive agents.

The word “abuse” is also commonly used in ways that differ significantly from the definitions developed for use in DSM-IV. In popular and legislative contexts *drug abuse* means any use of an illicit substance or any nonprescribed use of a drug intended as a medicine, as well as the harmful or excessive use of legally available substances, such as alcohol and tobacco.

Despite the reliability of DSM-IV and ICD-10 criteria for dependence in many European and Anglo-American cultures, several criteria (e.g., narrowing of drinking repertoire, time spent obtaining the drug, and even tolerance for the drug) have posed difficulties in other cultures, especially when dealing with alcohol. Tolerance is often understood when applied to drugs, but not to alcohol; in some cultures, holding one's liquor is a sign of manhood. Clinicians are more likely to make a diagnosis of drug dependence than alcohol dependence even when behavioral signs are comparable. In several cultures, little or no distinction is recognized between use, abuse, and harmful use of illicit drugs.

Other Perspectives The criteria for diagnosis in DSM-IV and ICD-10 were developed from what is essentially a biopsychosocial model of substance dependence. In such a model multiple factors—genetic, psychological, sociological, and pharmacological—contribute to the observed clinical syndromes. Such apparent unanimity about drug dependence should not obscure the existence of dissenting perspectives, which take several forms. In one the biopsychosocial model is accused of giving too much weight to biological factors and too little recognition to the notion of human will and responsibility, of medicalizing deviant behavior for the benefit of treatment professionals, and of creating universal exculpation for all those who fail to live up to reasonable societal expectations. But some professionals have implicitly criticized the same biopsychosocial model for not giving sufficient weight to the ideas that substance dependence is a specific primary disease (i.e., not a symptom of other psychiatric difficulties), that those who develop the disease have no control over their intake of certain substances, and that denial of the presence of a problem is a major characteristic of the disease.

Concepts about substance dependence can be arrayed along several dimensions that are not entirely independent or orthogonal: broad versus narrow, disease versus learned behavior, and social versus medical. The narrow concept of substance dependence accepts as disorders those maladaptive behaviors associated primarily, if not exclusively, with the ingestion of substances generally accepted as pharmacological agents. Compulsive eating, gambling, running, hair pulling, and repetitive excessive sexual activities are not included among the dependence disorders, although those problems may share certain features that resemble a decreased ability to choose and are sometimes ameliorated by participation in support groups founded on principles similar to those of Alcoholics Anonymous (AA). A broad approach would create a superclass of disorders that would include a number of such behaviors not involving pharmacological agents.

At the disease end of the disease-versus-behavioral syndrome dimension is a belief that dependence is not a learned behavior that can be modified or ameliorated with relearning but is a primary disorder caused by an interaction between a substance and a person with some genetic vulnerability and that only total abstinence can arrest the progression of the disease. The medical-versus-social dimension typically describes a range of views on how best to respond to problems with substances, rather than differences about the essential nature of the problems. The medical model stresses issues of assessment—treatment, planning, and record keeping—and sometimes treatment that can be rendered only by those with professional training (not necessarily physicians). The social model emphasizes the importance of social supports and integrating the person with a problem into a network of recovering persons who can offer continuing support. The assessment and recording of progress and outcome as generally practiced by credentialed professionals is minimized.

HISTORY

The most commonly abused drugs have been in use for hundreds, if not thousands, of years. For example, opium has been used for medicinal purposes for at least 3500 years, references to cannabis (marijuana) as medicinal can be found in ancient Chinese herbals, and wine is mentioned frequently in the Bible. The indigenous people of the Western Hemisphere were smoking tobacco and chewing coca leaves generations before the arrival of the Spaniards. Some of the problems caused by alcohol and other drugs, such as drunkenness, are described in the Bible and in the writings of the ancient Greeks and Romans. As new and more concentrated forms of drugs were discovered or invented or new routes of administering them were developed, new problems related to their use emerged. For instance, when cheap gin was introduced into England in the eighteenth century, the alcohol-related problems that emerged were considered more serious than those associated with beer and wine. Although opium smoking was a major problem in Asia in the eighteenth and nineteenth centuries, new problems were seen after morphine, the most active opium alkaloid, was isolated in 1806. Morphine was subject to misuse by injection from the late nineteenth century on, and intravenous morphine and heroin use began to spread in the early part of the twentieth century. Tobacco use and its associated problems did not become widespread until the nineteenth century, when new methods of curing the leaves produced a mild smoking tobacco and cigarettes were introduced, which made common the practice of inhaling tobacco smoke deeply into the lungs. By the early twentieth century, cigarette smoking was a popular practice.

Medicalizing Excessive Drug Use In 1810 Benjamin Rush, who is often credited as the first American physician to suggest that excessive use of alcohol was a disease rather than exclusively a moral defect, proposed the establishment of a sober house; in 1835 Samuel Woodward, a pioneer in the establishment of asylums for the insane, advocated similar asylums for inebriates. Contemporaneous with those early moves to involve medicine in dealing with excessive alcohol use was the emergence of the temperance movement and the Washingtonians—groups of reformed drunkards concerned with helping others to adopt and maintain sobriety. In the process the Washingtonians developed many of the principles of self-help that were rediscovered by AA almost a century later. When the ideas of voluntarism and self-help as exemplified by Washingtonian societies failed to eliminate the problem of drunkenness, physicians began to debate more seriously the idea of coerced treatment in inebriate asylums supported by public funds. In 1870 advocates of the approach established the American Association for the Cure of Inebriates (AACI), dedicated to setting up hospitals for such persons, conducting research, and teaching medical students and physicians how to treat inebriety. At first those physicians who believed in a more spiritual, voluntary approach to the problem (neo-Washingtonians) were part of the AACI, but gradually the more somatically oriented factions, which advocated medically supervised asylums (and compulsory treatment when needed), gained ascendancy. Furthermore, the focus of concern was no longer limited to those who abused alcohol. Thomas Crothers, the secretary of AACI, saw inebriate asylums as places to treat all those who used any variety of intoxicant or narcotic to excess. However, very few publicly supported inebriate asylums ever opened.

Early Attitudes The closing years of the nineteenth century saw growing concern about the excessive and inappropriate use of drugs, including alcohol and tobacco as well as opiates and cocaine. First isolated from the coca leaf in 1860, cocaine came into widespread use in 1885 when pharmaceutical companies began selling it in the United States and Europe. In 1884 Sigmund Freud had published a review of the potential therapeutic uses of cocaine. Some medical authorities in the United States shared his enthusiasm, and cocaine was recommended by the Hay Fever Association as a remedy for that malady. Within a few years, however, it was recognized that cocaine had the capacity to induce toxic psychosis as well as gain control over behavior. It was also recognized that long-term opiate use had dependence-inducing effects. Nevertheless, in the United States, until the beginning of the twentieth century, both the opium alkaloids and cocaine were still found in

patent medicines that were sold over the counter for a wide variety of indications, and their labeling often did not reveal their contents.

Although achieving long-term cure of morphinism was reported to be exceedingly difficult, until the turn of the twentieth century neither the public nor the medical profession saw the habitual user of opium or morphine as invariably suffering from a moral deficit. Those who had developed the morphine habit represented the entire socioeconomic spectrum, with women outnumbering men by about two to one. Various political and literary figures were known to use opiates but to lead otherwise productive and exemplary lives. However, cocaine use and the morphine habit were also common among gamblers, petty thieves, prostitutes, and other disreputable members of society. Persons with emotional problems and those who had formerly used alcohol to excess were probably also overrepresented among opium users, since it was not unusual at the time for physicians to prescribe opiates to control emotional problems and alcoholism.

The problem of using the same institution for treatment of drug users who had antisocial tendencies and those who led more conventional lives was as vexing to early advocates of medical treatment as it is to present-day practitioners. Many proponents of inebriate asylums did not want to take responsibility for persons who had frequent or serious encounters with the police because it was thought that such persons would make it impossible to create an atmosphere conducive to recovery. Partly to cope with the problem, even some of the proponents of a disease model of inebriety maintained the distinction between “inebriety the disease” and “intemperance the vice.”

Early Control Efforts: Evolution of the Criminal Model By the late 1890s the public and the medical community were no longer indifferent to drug use and habituation. In 1893 the Anti-Saloon League was founded, reinvigorating a temperance movement that advocated the total prohibition of alcohol. Medical texts in England, Europe, and the United States contained descriptions of morphinism, theories of its causation, and recommendations for withdrawal and postwithdrawal treatment. Some texts also described problems of cocainism. Medical authorities in the United States cautioned against overly liberal prescribing of cocaine and opiates by physicians and expressed great concern about the presence of those drugs in unlabeled proprietary over-the-counter medicines. State laws were passed aimed at controlling the sale of opiates and cocaine, especially in patent medicines. In 1900 the cocaine in Coca-Cola was replaced by caffeine.

Partly to support the efforts of the Chinese government to control opium use in China, representatives of the United States government led the movement to negotiate an international treaty to control traffic in opium, cocaine, and related drugs. The first such treaty was signed in The Hague in 1912. Negotiators from the United States were also interested in the international control of cannabis but could not get other nations to view the substance as sufficiently problematic to warrant it. (Such control was achieved in 1925 at the Second Geneva Convention.) The Hague Convention required the signatories to pass domestic legislation controlling opiates and cocaine. The Harrison Act of 1914, the first federal legislation to regulate opiates and cocaine, was designed to restrict access to opiates and cocaine to doctors, dentists, pharmacists, and legitimate importers and manufacturers and brought the United States into compliance with the convention.

State regulations concerning the sale of opiates and cocaine, the introduction of aspirin and the barbiturates, and the Pure Food and Drug Act of 1906, which required labeling of patent medicines, were already having an impact on the use of opiates in medicine when the Harrison Act was passed in 1914. Although many medical and political leaders in the United States believed that much of the problem of drug dependence resulted from careless prescribing by physicians, the Harrison Act was not originally intended to interfere with the legitimate practice of medicine or to cause special hardship for those already dependent on opiates. For several years after the Harrison Act was passed, a few cities operated clinics that prescribed morphine to persons with established morphine habits. Most of those dependent on opiates before the Harrison Act became abstinent within a few years after it was passed, although generally not as a result of treatment at the clinics.

Fluctuating Attitudes Major changes had taken place in American attitudes and practices by the 1920s. The Eighteenth Amendment to the U.S. Constitution, which prohibited the sale of alcohol, became law in 1920 and radically changed drinking behavior in the United States. Within a year after alcohol prohibition was enacted, 14 states also passed cigarette prohibition laws. Even less popular than alcohol prohibition, those laws were all repealed by 1927, and by the mid-1920s Americans were smoking 80 billion cigarettes a year. However, cocaine use, so prevalent at the turn of the century, was no longer widespread.

Disillusioned by the reluctance of morphine addicts at clinics to detoxify and by repeated relapses among those who did, doctors began to recommend (not for the first time) compulsory treatment with confinement until cure. As the new laws curtailed legitimate supplies of opiates, an illicit traffic developed to provide them to morphine addicts who could not or would not use the clinics. Increasingly, the drug sold was heroin, which had been introduced for medical use in 1898 but was quickly found by drug users to have effects quite similar to those of morphine. Many who patronized the illicit traffickers and used the clinics had histories of delinquency and criminal activity, and eventually that subgroup came to predominate. Reformers, moralists, and the popular press found in the opiate habit, and in the reputation of those who continued to use morphine, proof of the evils inherent in those drugs.

Negative publicity, lurid stories, medical disillusionment, and pressure from law enforcement agents combined to label the morphine clinics as medical folly and brought about their closing, the last in 1923. At the same time a series of United States Supreme Court decisions implied that prescribing even small amounts of opiates or cocaine to an addict for treatment of addiction was not proper medical practice and was thus an illegal sale of narcotic drugs. Several physicians were imprisoned, and numerous others were tried, reprimanded, or otherwise harassed. By the early 1920s persons addicted to opiates were not welcome in doctor's offices, and they were often refused treatment at hospitals. *Dope addict* and *dope fiend* had become common terms, and the average layperson, as well as some otherwise well-informed members of the medical profession, appeared to believe that the opiate molecule was inherently evil. In the late 1930s cannabis acquired a similar reputation, and in 1937 the United States Congress passed legislation prescribing criminal penalties for its use, sale, or possession. Alcohol prohibition had been repealed in 1933.

New Drug Problems The first of the barbiturate sedatives, barbital, was introduced into clinical medicine in 1903, followed over the next 30 years by scores of congeners that differed primarily in their duration of action. Within a few years after the introduction of each new compound, the first case reports of abuse, dependence, and withdrawal appeared in the medical journals, a pattern that was repeated with the nonbarbiturate sedatives, such as glutethimide (Doriden), ethchlorvynol (Placidyl), and meprobamate (Miltown) in the 1950s.

Amphetamine, first synthesized in 1887, was put into clinical use in 1932 as a drug to shrink mucous membranes. By 1935 its central stimulant effects had been recognized and found useful for treating narcolepsy, and dozens of other suggested uses soon followed. Reports that amphetamine was being used as a euphoriant began to appear in the late 1930s, but the full significance of its abuse potential was not appreciated until the post-World War II epidemic of intravenous methamphetamine addiction in Japan. That epidemic, precipitated by the sale of surplus methamphetamine tablets intended for combat troops, involved millions of people. Other amphetamine-like drugs, which have also been subject to abuse, were introduced during the 1950s and early 1960s.

The psychological effects of mescaline were already known and written about at the end of the 19th century. However, public concern about hallucinogens did not reach a high level until the 1960s, when the use of a newly discovered and exceedingly potent compound, lysergic acid diethylamide (LSD), evolved from experimentation by a few college students to more widespread use by even younger people. Phencyclidine, a general anesthetic developed in the 1950s, also became a drug of abuse in the 1970s.

Despite repeated reports of abuse and dependence associated with barbiturates, barbiturate-like sedatives, and amphetamines and related stimulants, and in spite of concerns about experimentation with LSD and related hallucinogens, there were no federal criminal sanctions related to these drugs until 1964, when authority for their control was assigned to the Food and Drug Administration (FDA). In contrast, in the 1950s, concern about heroin addiction had led to ever harsher criminal penalties for its sale or possession. Although law enforcement efforts aimed at controlling heroin use were increased, both the number of new heroin addicts and the crime rates continued to rise throughout the late 1960s. At about that time there was also a sharp increase in the nonmedical use of other substances, such as cannabis and LSD, and a major epidemic of amphetamine abuse and dependence. In addition to amphetamines diverted from medical channels, supplies came from clandestine laboratories. Drug use, especially cannabis, became linked to antiestablishment attitudes, politics, and lifestyles.

Evolving Treatment Approaches Treatment for substance-related problems underwent several dramatic changes during the twentieth century. The large specialized asylums that were advocated in the nineteenth century never materialized. Toward the end of the nineteenth century physicians were primarily concerned about how to manage withdrawal syndromes and whether or not longer compulsory treatment was needed. With the advent of prohibition, the impetus to develop treatments for alcoholism declined sharply. Interest in treating opioid-dependent patients also declined as physicians became discouraged by their tendency to relapse after being detoxified and as opioid use and dependence came to be seen more as criminal behaviors than as medical disorders. A few private sanitoriums continued to provide treatment for opioid dependence. By 1930, as drug-addicted prisoners began to fill the penitentiaries, the federal government saw the need to establish two hospitals, at Lexington, Kentucky, and Fort Worth, Texas, to provide treatment for that population and also to conduct research on the problem of opiate addiction. Treatment of barbiturate and amphetamine dependence took place largely in the mainstream of medical practice and in state hospitals, but there was no consensus on what constituted effective posthospital care.

In the mid-1930s two recovering alcoholics rediscovered the principles of the Washingtonians, added some new principles, and initiated the self-help movement now known as AA. By the 1950s, this movement had begun to inspire analogous self-help efforts among other types of substance abusers.

The situation changed again in the early 1960s. With new outbreaks of heroin use by young people and increasing crime, the federal government and individual states attempted to respond to the problem. California initiated a civil commitment program for addicts under the administrative control of the Department of Corrections; New York City reopened Riverside Hospital to treat juvenile heroin addicts. The first follow-up studies of patients treated at the federal hospital at Lexington revealed exceedingly high rates of relapse after treatment. Both the medical community and the general public demanded new ideas and solutions, including a reconsideration of providing addicts with legitimate opioids through medical channels.

From 1958 to 1967 several major new approaches to treating opioid dependence were developed. Synanon, the prototype therapeutic community, was started in California in 1958 and was soon replicated in New York with the establishment of Daytop Village and Phoenix House. Vincent Dole and Marie Nyswander showed that maintaining selected long-term heroin addicts on large daily doses of methadone (Dolophine) was effective in reducing crime and heroin use. Several research groups demonstrated that heroin addicts would voluntarily try treatment with narcotic antagonists. In the mid-1960s, New York State and the federal government legislated civil commitment programs modeled after the program in California, with an initial period of prolonged institutional care as a key element. Although many treatment programs initiated in the early 1960s continued to focus on the treatment of opioid dependence, others, especially the therapeutic communities, viewed all nonmedical drug use as stemming from similar defects in character structure and offered a generic approach to treating drug dependence.

Alcohol and Nicotine In the 1950s clinicians at Wilmar State Hospital in Minnesota developed a treatment program for alcoholism built on a synthesis of the medical model and the experiences of recovering alcohol abusers using the 12-step principles of AA. That treatment approach was refined and expanded at the Johnson Institute and Hazelden Foundation, also in Minnesota. The modified programs, widely adopted by others, are often referred to as 28-day programs, 12-step programs, or the Minnesota model. In the early 1970s the effort to recognize alcoholism as a disease gained momentum, and the decision of medical insurance carriers to provide coverage for detoxification and inpatient treatment fueled an unprecedented growth of private-sector facilities offering treatment for alcoholism. Almost without exception, they were residential programs using the Minnesota model. The decriminalization of public intoxication spurred a parallel increase in alcohol treatment programs supported by the public sector.

The Surgeon General's Report of 1964 linked cigarette smoking to lung cancer and concluded that tobacco smoking was a form of dependence, although not an addiction. By the 1970s, tobacco dependence was more widely accepted as a valid clinical entity, and various treatments for it were developed. By the late 1980s, as smoking was becoming socially unacceptable, many buildings were declared smoke free, smoking was banned on most airplane flights and in many hospitals, and pharmaceutical companies began to market new products for delivering nicotine (e.g., nicotine chewing gum and transdermal patches) as aids for smoking cessation. By the late 1990s the tobacco companies were negotiating settlements in multiple civil law suits by states and by individuals who had been injured by their tobacco use, and Congress had unsuccessfully debated major tax increases on tobacco and regulation by the FDA.

Two-Tiered System When the cocaine epidemic of the early 1980s struck the middle class, much of the large, private-sector system for treating alcoholism evolved into chemical dependency units offering similar treatments to persons with alcohol problems and those with other varieties of substance dependence. By 1990 it was estimated that more than 8000 recognized programs existed that deal with alcoholism and other substance dependence. The treatment methods used varied widely in terms of settings, costs, philosophical underpinnings, and populations served. New categories of substance-abuse professionals had emerged, and psychiatrists who once had considered the problems to be a low-status area successfully lobbied for the creation of a recognized subspecialty in addiction psychiatry. Treatment capacity was described as a two-tiered system with private and public sectors, in which the private sector served 40 percent of the population but received 60 percent of the total expenditures for treatment. One response to the escalating cost of substance abuse services among those with private medical insurance was the rise of a managed care industry created to control costs on behalf of employers who pay for health insurance, generally by severely limiting the length of stay in hospital settings. Managed care, by refusing to recognize (and pay for) the medical necessity of inpatient treatment for most cases of substance dependence, largely dismantled the rest of the “28-day” inpatient alcohol and drug treatment programs that had serviced patients with insurance. By the mid-1990s managed care principles were routine in the public sector as well, and little remained of the two-tiered system.

Legislation and National Strategies In 1969 Congress recognized the need to give greater attention to the problem of alcoholism and established the National Institute on Alcohol Abuse and Alcoholism (NIAAA) in the National Institute of Mental Health (NIMH). In 1970 legislation was passed, reorganizing the jumble of drug regulatory statutes that had evolved since the passage of the Harrison Act, increasing the resources for controlling the availability of illicit drugs, and assigning the task of enforcement to a new agency, the Drug Enforcement Agency (DEA), which incorporated elements of the FDA and the Bureau of Narcotic and Dangerous Drugs. All drugs subject to special controls were included in one of several categories of the Controlled Substances Act.

In 1971 when United States troops in Vietnam were reported to be using heroin heavily, the Special Action Office for Drug Abuse Prevention (SAODAP) was established in the Executive Office of the President to coordinate government activities and policies relating to drug abuse and to develop and publish an overall national drug strategy. The creation of that office and the associated legislation marked a turning point in United States policy. The notion that opioid dependence was an incurable disorder, which justified the harshest of penalties in the name of prevention, was superseded by a policy that recognized that a substantial proportion of opioid addicts (as well as those with other varieties of drug dependence) could eventually reenter the mainstream of society. New commitments were made to basic research, epidemiology, development of new treatment methods, and evaluation of existing treatment approaches. Methadone maintenance was moved, by executive fiat, from the legal limbo of experimental status to a category that recognized its legitimacy. Regulations intended to prevent inappropriate prescribing of opioids were developed. Federal support for the expansion of community treatment programs was also greatly increased. By 1973 about 200,000 substance users, most of them opioid users, were in treatment in community programs. Those programs were repeatedly and intensively evaluated over the subsequent decade. The legislation that established SAODAP also provided the legislative framework for the National Institute on Drug Abuse (NIDA) in the Department of Health, Education and Welfare (HEW). When it was established in 1974 NIDA became the lead agency for implementing federal policy on treatment, research, and prevention.

By the early 1980s treatment for opioid dependence was generally accepted to have demonstrable impact. However, for most patients in treatment programs, the primary drugs of abuse were no longer opioids but more typically, cannabis, stimulants, or sedatives. During the early and mid-1970s some groups had argued for the decriminalization or legalization of cannabis. The arguments lost much of their force when it was found that in 1979 almost 10 percent of high school students were using cannabis on a daily basis. In response to what they perceived as tolerance toward cannabis use, a number of parents' organizations were formed that were committed to making all drug use unacceptable. Those groups forced NIDA to review and remove from all its publications any statements that could be interpreted as tolerating drug use. This decreased tolerance for drug use grew in parallel with a more general conservative shift in public attitudes. For example, in the 1970s the public and the courts had rejected the use of urine testing as a means of detecting drug use in an effort to interrupt the heroin epidemic; but starting in 1986, federal employees were required by presidential order to undergo such tests. Similar drug testing was encouraged in private industry, giving rise to new industries for detecting the presence of drugs, interpreting test results, and placing drug users in treatment.

By the 1970s it was obvious that the major drug abuse problems in the United States in terms of social and economic impact and health costs were alcoholism and tobacco dependence. Although the Surgeon General's Report of 1964 linking cigarettes to cancer had not produced any dramatic decrease in smoking, the rate of increase in cigarette consumption among men had begun to level out. In 1988 the *Surgeon General's report on the Health Consequences of Smoking* officially defined tobacco dependence as analogous to other varieties of drug dependence. In 1994 the FDA held hearings on the appropriateness of regulating the nicotine in tobacco as an addictive drug. Shortly thereafter, with backing from the president, the FDA assumed authority to regulate advertising of tobacco products; the White House lobbied Congress to pass legislation that would limit advertising and raise federal taxes on tobacco.

In the early 1980s rising demand for the treatment of cocaine dependence, the sudden cocaine-induced deaths of several prominent athletes, and concern about the spread of the human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS) among intravenous drug users led to the Anti-Drug Abuse Act of 1986, which authorized the government to spend nearly $4 billion to intensify efforts against drugs and drug abuse. Although most of that money was allocated to law enforcement activities, federal resources for the treatment of drug dependence and research were also substantially increased. Recognizing the need to do more to prevent drug dependence and provide more treatment, the federal government created a series of offices that by 1992 evolved into the Substance Abuse and Mental Health Services Administration (SAMHSA), with several constituent centers, including the Center for Substance Abuse Treatment (CSAT) and the Center for Substance Abuse Prevention (CSAP). The 1988 Anti-Drug Abuse Act and the 1989 Emergency Supplemental Appropriation created the Office of National Drug Control Policy (ONDCP) in the White House. While still devoting more than two thirds of federal resources available for drug problems to controlling drug supply, this legislation also increased funding for treatment and prevention.

Critics of the emphasis on supply-control gained public attention when they were supported by several prominent conservative writers and economists and garnered the financial support of several well-endowed foundations. While the more thoughtful of these critics have stopped calling for outright legalization of drugs, they have

called for greater emphasis on reducing the harm related to drug use by medically prescribing heroin and other psychoactive drugs and more support for needle-exchange programs. Despite some evidence suggesting that availability of sterile needles can reduce HIV transmission, the federal government continues to ban the use of federal money for such programs.

EPIDEMIOLOGY

A number of distinct methods have been developed to gauge the extent and medical consequences of substance use, abuse, and dependence in the United States. The major recurring surveillance instrument are the National Household Survey on Drug Abuse (Household Survey), the Drug Abuse Warning Network (DAWN), Arrestee Drug Abuse Monitoring System (ADAM—formerly known as the Drug Use Forecasting [DUF] program), and the Monitoring the Future Study (better known as the High School Survey). In addition, data on street availability and purity of illicit drugs, drug seizures, and arrests for drug offenses are collected nationally from the DEA and the Federal Bureau of Investigation (FBI) and locally from municipal police departments. Each of these data sources has strengths and limitations. For example, the Household Survey annually interviews a representative sample of individuals age 12 and older living in households, college dormitories, homeless shelters, and rooming houses. It oversamples minority populations and certain large urban areas, and focuses in detail on drug-using behaviors. It does not interview military personnel or individuals who are living on the street or in institutions (jails or hospitals). It does not attempt to determine whether respondents need treatment or meet formal criteria for drug dependence. In addition, some respondents may be reluctant to admit to certain types of drug use.

The ADAM system interviews, and obtains anonymous urine specimens from, a sample of arrestees in moderate-size cities in the United States. By design, persons charged with sale or possession of drugs cannot make up more than 25 percent of the sample. Although it does not depend on self-reports to measure use, the ADAM results cannot be easily extrapolated to a national population, and the information that can be derived from a single urine test is limited.

In 1989 the DAWN system, which obtains data on drug-related episodes from medical examiners and hospital emergency rooms, was modified so that the reporting emergency rooms constitute a representative sample of such facilities in the continental United States. The DAWN data provide useful information on trends in the morbidity associated with various illicit drugs; but these data need to be interpreted with caution because the DAWN system reports only episodes in which a drug is part of the presenting clinical picture. For example, a rising number of emergency room episodes associated with heroin could mean that more heroin users with AIDS-related problems are seeking primary medical care, rather than that more individuals are using heroin. Similarly, reports by medical examiners of more violent deaths associated with cocaine may signal an escalation of competition among drug dealers, rather than more people using cocaine. The analytical methods do not reveal the nature of the linkage between drug use and the presenting problem, which drugs (if any) played a causal role in the episode, or whether the user was a novice or a chronic user.

The High School Survey has obtained information each year since 1975 from forms returned anonymously by high school seniors. It now includes former seniors now in college and students in the eighth and tenth grades. Although the survey depends on self-report, the trend information it provides is exceedingly useful.

In addition to the recurring data-gathering efforts, important epidemiological information is available from two national studies that systematically interviewed representative samples of the population and used DSM-III or DSM-III-R criteria to develop estimates of current and lifetime prevalence of psychiatric disorders, including substance abuse and substance dependence. These studies are the NIMH Epidemiological Catchment Area (ECA) Study, conducted in the early 1980s, and the National Comorbidity Survey (NCS), conducted between 1990 and 1992. The ECA interviews in five areas of the United States included individuals in institutions (mental hospitals, jails, nursing homes, etc.) and used DSM-III criteria to develop estimates of prevalence. The NCS interviews of a nationally representative sample of noninstitutionalized people used DSM-III-R criteria. Although the ECA was conducted before the cocaine epidemic of the 1980s crested and criteria for diagnosis used were altered somewhat in DSM-III-R, it nevertheless remains a landmark study of the extent of drug abuse and dependence and co-occurring psychiatric disorders.

The ECA study found that 16.7 percent of the U.S. population ages 18 and older met the DSM-III criteria for a lifetime diagnosis of either abuse or dependence on some substance, with 13.8 percent meeting the criteria for an alcohol-related disorder, and 6.2 percent meeting the criteria for abuse or dependence of a drug other than alcohol or tobacco. The NCS found a 26.6 percent lifetime prevalence of substance abuse and dependence, substantially higher than the 16.7 percent found in the ECA. Some of this is probably due to questions in the NCS about prescription drugs that were posed when a patient reported symptoms of dependence, and on differences in criteria (DSM-III versus DSM-III-R). However, there may also have been real increases in prevalence. For illegal drugs and the nonmedical use of prescription drugs, the lifetime rate for dependence in the NCS was 7.9 percent, a figure much closer to the 6.2 percent found for such drugs in the ECA study. The NCS found a 12-month prevalence estimate for any addictive disorder (including dependence and abuse) of 8.2 percent; 4.5 percent alcohol dependence, and 1.8 percent drug dependence. Except for tobacco, men are far more likely than women to use drugs and alcohol and are correspondingly more likely to develop dependence. For example, lifetime and 12-month prevalence rates of alcohol dependence are 20.1 percent and 6.6 ercent for men, but only 8.2 percent and 2.2 percent for women.

Among the major achievements of the NCS analyses were the findings on the proportions of people who had used drugs at any time in their lives (*lifetime users*) who became dependent (overall and for each drug category); the demographic factors that predicted use, dependence, and persistence of dependence; and the prevalence and significance of multiple psychiatric diagnoses. Dependence cannot develop if a drug is never used; thus, presenting data on the prevalence of dependence in the population as whole, including those who never used, can obscure the likelihood of dependence developing among those who do use a particular drug. In the NCS, prevalence of lifetime dependence on the broad range of illicit and nonprescribed medications was 14.7 percent, with male users only slightly more likely (16.4 percent) than female users (12.6 percent) to develop dependence. In a similar analysis of the 12-month prevalence of dependence on these drugs, the rate for the population as a whole was 1.8 percent. However, the 12-month prevalence was 3.5 percent for those who had used any of these drugs at any time in their lives; 10.3 percent for those who had used them in the past 12 months, and 23.8 percent among those who had a lifetime history of dependence. The likelihood of being drug dependent within the past 12 months, given a lifetime history of dependence, was similar for men (24.9 percent) and women (22.2 percent). Lower educational and lower income levels predicted a lifetime history of dependence (odds ratios greater than 2), but race, ethnicity, or living in an urban environment did not. There were also differences in the likelihood that users of a particular drug would become dependent on it. For example, for heroin, the lifetime opioid dependence rate was 23 percent; for tobacco, 32 percent; for cocaine, 16.7 percent; for alcohol, 15.4 percent but only 4.9 percent for psychedelics. Men who used alcohol were more likely to become dependent (21.4 percent) than women (9.2 percent), possibly because they drink more than women but genetics may also play a role.

Table 11.1-12 shows data from the 1996 Household Survey on percentage of respondents who reported using various drugs. The data are shown for four age groups. Persons aged 18 to 25 years reported the highest level of use of illicit drugs during the 30 days preceding the interview; those ages 26 to 34 had the next highest rate and reported a higher lifetime experience with cocaine. Illicit drug use during the 30 days preceding the interview is far more prevalent among young adults (ages 18 to 34, and particularly those 18 to 25 years old) than among those above age 35 or below age 18. Also, whereas recent use is more common in large metropolitan areas than in rural areas, regional, racial, and ethnic differences vary with the age group considered. With the exception of tobacco dependence, all forms of substance abuse or dependence are more common among men than among women. However, recent data indicate that when adjustment is made for differences in rates of use and experimentation with illicit drugs, women are about as likely as men to become dependent. Current illicit drug use (past 30 days) was more common among male (8.1 percent) than female (4.2 percent) respondents, and among the unemployed. Among other demographic subgroups, it was slightly more common among blacks and in the western states.

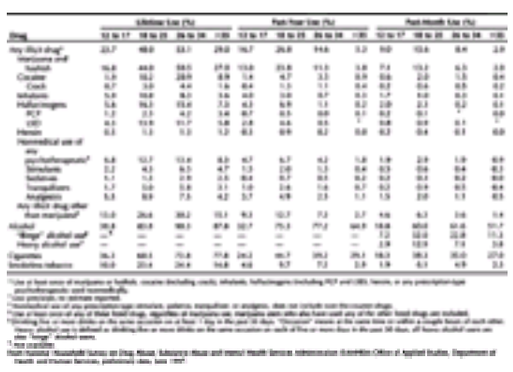

Table 11.1-12 Use of Illicit Drugs, Alcohol, and Tobacco in the U.S. Population by Age Groups

The High School Survey found that self-reported use of cannabis and illicit drugs in general (mostly cannabis) in the past 30 days declined sharply from the high levels (38 to 40 percent) reported in 1977 through 1979 to much lower levels (16 percent) in 1991. The decline in cocaine use began in 1987. However, 30-day prevalence rates for cannabis increased from 1992 through 1997. Cocaine and crack cocaine 30-day prevalence rates also increased slightly from a low of 1.3 percent in 1993 to 2.3 percent in 1997. Other substance use increased also, but levels were still below the peaks observed a decade earlier. In 1997 the annual prevalence rate among high school seniors for use of any illicit drug was 42.4 percent, and for an illicit drug other than cannabis, it was 20.7 percent.

The ADAM system obtains data from a population in which illicit drug use is high and thus provides trend data not readily available from other sources. In general, current drug use among arrestees is several times higher than that among those sampled by national surveys, even though urine tests detect drug use for only a few days, whereas surveys typically ask about drug use over the preceding 30 days. For example, in 1988, the peak of the cocaine epidemic, more than 60 percent of arrestees tested positive for cocaine (80 percent among male arrestees in Manhattan). More-recent data (1995) show a decline in cocaine use and low levels of heroin use.

Epidemics Several major overlapping drug abuse epidemics have occurred over the past 30 years, affecting somewhat different populations. Cannabis use, which had been endemic among certain minority groups and jazz musicians, began to increase in the 1960s, especially among young people, and then spread to other segments of the population. At its peak, in 1978 to 1979, 10 percent of high school seniors were using marijuana on a daily basis. Daily use declined to 5 percent by 1984, to 2 percent by 1991, and then reversed direction and again rose. Similar changes in use rates were reflected in the Household Survey.

An epidemic of heroin use also began in the early 1960s, and incidence peaked between 1969 and 1971. The population of active heroin users reached its highest levels in the early 1970s, but periodic upsurges have occurred as supplies became more available, law enforcement activity waxed and waned, and relapse rates increased among former users. In 1977 the United States government estimated that there were 500,000 opioid abusers and dependent users, and more recently, it revised the estimate to 320,000 occasional users and 810,000 chronic users. In general, the heroin-using population is an aging one, with a high and still growing prevalence of HIV in some areas. The 1996 Household Survey estimated that about 2.3 million people had tried heroin at least once and that 245 thousand had used it in the past year. However, it is believed that a large percentage of heroin users are outside the population interviewed by the survey.

The cocaine epidemic began in the 1970s and reached its peak around 1985, when it was estimated that 5.8 million people in the United States (2.9 percent of the population) had used cocaine in the month prior to survey. The epidemic seems to have passed its peak in most segments of society, with current (past 30 days) use rates in 1996 at about 1.5 to 2 percent among those 18 to 34 (0.8 percent for ages 17 and older). Cocaine use among the heaviest users (weekly or almost weekly) did not decline significantly, but rates decreased among arrestees in 1995.

In the early 1990s, fueled by abundant supplies of cheap illicit methamphetamine produced in many small laboratories, methamphetamine use began to increase in a number of cities in western, southwestern and northwestern parts of the United States. By 1996, gauged by drug tests on arrestees, that epidemic had passed its peak in those areas.

ETIOLOGY

The model of drug dependence from which the DSM-IV and ICD-10 criteria were derived conceptualizes dependence as a result of a process in which multiple interacting factors influence drug-using behavior and the loss of flexibility with respect to decisions about using a given drug. Although the actions of a given drug are critical in the process, it is not assumed that all persons who become dependent on the same drug experience its effects in the same way or are motivated by the same set of factors. Furthermore, it is postulated that different factors may be more or less important at different stages of the process. Thus, drug availability, social acceptability, and peer pressures may be the major determinants of initial experimentation with a drug, but other factors such as personality and individual biology probably are more important in how the effects of a given drug are perceived. Still other factors, including the particular actions of the drug, may be primary determinants of whether drug use will progress to drug dependence, whereas still others may be important influences on the likelihood that drug use will lead to adverse effects or the likelihood of successful recovery from dependence.

Figure 11.1-1 illustrates how various factors might interact in the development of drug dependence. The central element is the drug-using behavior itself. The decision to use a drug is influenced by immediate social and psychological situations, as well as by the person's more remote history. Use of the drug initiates a sequence of consequences that can be rewarding or aversive, and which, through a process of learning, can result in a greater or lesser likelihood that the drug-using behavior will be repeated. For some drugs, use also initiates the biological processes associated with tolerance, physical dependence, and (not shown in the figure) sensitization. In turn, tolerance can reduce some of the adverse effects of the drug, permitting or requiring the use of larger doses, which then can accelerate or intensify the development of physical dependence. Above a certain threshold, physical dependence is generally a distinct recurrent motive for further drug use. Sensitization of motivational systems may increase the salience of drug-related stimuli.

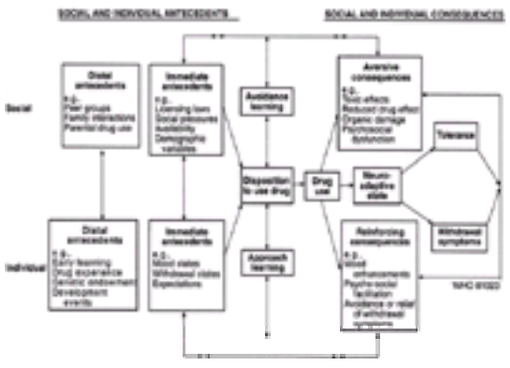

FIGURE 11.1-1 WHO schematic model of drug use and dependence. (Reprinted with permission from Edwards G, Arif A, Hodgson R: Nomenclature and classification of drug- and alcohol-related problems. A WHO memorandum. Bull WHO *99:*225, 1981.)

For simplicity Figure 11.1-1 shows drug use alone as initiating that chain of consequences, but the choices a person makes over and over again are more complex. The decision is whether to use one drug or another or to engage in some behavior that does not involve drug use. Each of those decisions can initiate positive and negative consequences. Changes in the availability, costs, and consequences of alternative behaviors can also influence what appears to be compulsive use of a pharmacological agent. For example, patients in a methadone maintenance program who were using cocaine despite negative consequences (no take-home methadone) reduced their cocaine use when vouchers for goods and services were awarded for clean (negative for cocaine) urine specimens.

Social and Environmental Factors Cultural factors, social attitudes, peer behaviors, laws, and drug cost and availability all influence initial experimentation with substances, including alcohol and tobacco. These factors also influence initial use of more socially disapproved drugs such as cocaine and opioids, but personality factors assume a more important role. Social and environmental factors also influence continued use, although individual vulnerability and psychopathology are probably more important determinants of the development of dependence. In general, the use of the less socially disapproved substances (alcohol, tobacco, and cannabis) precedes the use of opioids and cocaine, and those antecedent substances are sometimes referred to as gateway drugs.

Substantial evidence indicates that consumption of alcohol and tobacco in a population can be altered by changes in their price and availability. When alcohol availability is increased by increasing the number of sales outlets or extending sale hours, consumption tends to rise. When the cost of either alcohol or tobacco is increased in relation to disposable income (e.g., by increased taxes), consumption falls. These factors even influence the behavior of dependent persons, although perhaps not to the same degree as for those who are not dependent. Availability can be altered independently of cost, and alterations can be limited to selected populations (e.g., prohibiting sale of alcohol and tobacco to those under a specific age).

Social, cultural, and economic factors do not always operate synergistically but may sometimes influence consumption in opposite directions. For example, in the late

1980s increased public awareness of how alcohol use adversely affects health resulted in a decline in its consumption. That decline occurred even though alcohol was more freely available, its cost relative to income remained constant or actually decreased, and social pressures against women drinking (unless pregnant) also decreased.

Illicit Drugs Social and cultural factors, including beliefs about the effects of a drug, frequently exert more influence on drug-use patterns than the laws that supposedly reflect such factors. For example, cannabis use increased among high school students from the early 1970s to 1979 and then fell steadily over the next decade, although use and possession were illegal throughout the entire 18-year period, and nothing indicates that it became more expensive or less available during the 1980s. An upward trend in use was noted from 1993 to 1997, although it never reached the peak levels of 1979. Some experts believe the decline in use seen during the 1980s was linked to changing perceptions about the toxic effects of cannabis on health. The rise beginning in the 1990s was correlated with a decline in the perception of the risk of harm from regular use. Similarly, cocaine use increased in the late 1970s, despite high prices for the drug and high risk of criminal penalties; but following several well-publicized deaths from cocaine in the mid-1980s, its use declined among high school seniors and in the general population, even as the price of the drug declined.

Social and cultural factors profoundly influence the availability of illicit drugs, which in turn influences which groups in a society are most likely to become users. Currently, illicit opioids and cocaine are more available in the inner cities of large urban areas than in other parts of the country. Such availability not only influences initial and continued use but also affects relapse rates among those who seek treatment but must live in high-availability areas. When a significant number of users of illicit drugs live in one area, a subculture evolves that supports experimentation and continued use. Many of the areas in which illicit drugs are readily available are also characterized by a high crime rate, high unemployment, and demoralized school systems—all of which serve to reduce the sense of hope and sense of self-esteem associated with resistance to use and good prognosis once dependence develops. Social and educational factors also affect the likelihood for successful recovery from drug dependence; those who find satisfying alternatives are more likely to abstain from drug use.

VIETNAM The experience of United States service personnel who used heroin in Vietnam provided a unique natural experiment in which the influences of availability, vulnerability, and social norms could be observed. From 1970 to 1972 high-grade heroin at very low cost was readily available to young persons separated from their families and usual social norms. Among Army enlisted personnel, about half of those who tried heroin became dependent (at least they developed withdrawal symptoms when they attempted to stop using heroin). Of those who used heroin at least five times, 73 percent became dependent. The background factors that predicted heroin use in the general civilian population—early deviant behavior, such as fighting, drunkenness, arrest, and school expulsion—also predicted drug use in Vietnam, but they were not the best predictors of relapse after the soldiers returned to the United States. Relapse was related to being white, being older, and having parents who had criminal histories or were alcoholic.

Availability and Health Professionals The important role of availability is also illustrated by the repeated observation that physicians, dentists, and nurses have far higher rates of dependence on DEA-controlled substances, such as opioids, stimulants, and sedatives, than other professionals of comparable educational achievement (e.g., accountants or lawyers) who do not have such easy access to the drugs. Compared with controls, physicians appear to be four to five times as likely to take sedatives and minor tranquilizers without supervision by a professional other than themselves. Yet even in that situation other factors play a role. Physicians who had unhappy childhoods are more likely to self-prescribe than those who are healthier psychologically.

Drugs as Reinforcers The belief that persons take drugs because of the subjective effects the drugs produce can be traced to antiquity. Different drugs produce distinctive subjective states, and extensive laboratory evidence shows that persons with experience can distinguish one drug class from another and can even rank different classes and doses on the basis of how much they like the effects. Yet the hold that drugs can eventually exert on a user's behavior is not entirely a function of its initial likeable or euphorigenic actions. For example, the effects of cocaine are typically described as powerfully euphorigenic, producing increased self-esteem, alertness, energy, and well-being; the effects of nicotine are more subtle, producing some mixture of alerting and relaxing; and the subjective effects of alcohol are more likely to be described as relaxing, are more variable, and appear to be more dependent on personality. Despite those differences, dependence (or addiction) can occur with each, and they appear to have shared or overlapping neural substrates for their reinforcing properties.

Almost all of the drugs that are used for their subjective effects and are associated with the development of dependence induce some degree of tolerance. In some cases the tolerance to the toxic and aversive effects is more pronounced than the tolerance to the reinforcing and mood-elevating effects. For example, most opioid users quickly develop tolerance to opioid-induced nausea and vomiting. This may allow users to increase the dose and thus experience greater euphoric effects. Conversely, those who continue to experience aversive drug effects (such as severe flushing with alcohol) may be less likely to persist in using the drug and are at lower risk for developing dependence. Tolerant opioid users do not continue to self-administer opioids solely to prevent the highly aversive withdrawal phenomena. Interviews with heroin users have indicated that despite some tolerance to many of the drug's effects, they continue to experience a brief euphoric effect immediately after an intravenous injection. Among nonalcoholic sons of alcoholic fathers, intrinsic tolerance may be a marker of biological vulnerability to developing alcohol dependence. Sons of alcoholic fathers who were more tolerant to a test dose of alcohol were far more likely to have developed alcohol dependence at 8-year follow-up than those who were less tolerant.

With a few notable exceptions, animals in experimental situations will self-administer most of the drugs that humans tend to use and abuse. Included among the drugs are μ and d opioid agonists, cocaine, amphetamine and amphetamine-like agents, alcohol, barbiturates, many benzodiazepines, a number of volatile gases and vapors (e.g., nitrous oxide and ether), and PCP. Nicotine is also self-administered, although under more specialized conditions; cannabinoid self-administration has been difficult to demonstrate; and LSD-like drugs are not generally found to be reinforcing.

Biological Substrates Knowledge about the neurobiology of drug reinforcement and the mechanisms underlying tolerance and dependence has increased substantially. For opioids (and probably for other drugs as well) the neural systems involved in drug reinforcement and self-administration are distinct from those responsible for some of the other actions (e.g., opioid-induced analgesia) as well as from those that mediate the more visible signs of the withdrawal syndrome characteristic for that drug class. The pathways critical for the reinforcing actions of a number of dependence-producing drugs, such as opioids, amphetamine, cocaine, and to some degree nicotine and alcohol, have their origins in dopaminergic neurons with cell bodies in the ventral tegmental area and projections to the nucleus accumbens and the related structures that make up the “extended amygdala.” This comprises several neural structures receiving input from the limbic cortex, hippocampus, lateral amygdala and midbrain, and projecting axons to the ventral pallidum, the medial ventral tegmental area and the lateral hypothalamus. The medial part of the nucleus accumbens is a particularly important site; dopamine release here is critical for the reinforcing effects of cocaine and amphetamines. It is also important for the reinforcing effects of opioids, but there are opioid receptors on neurons in the nucleus accumbens, and opioids can exert reinforcing effects at that site even when the dopaminergic terminals are destroyed. Evidence suggests that such drugs as nicotine, cannabinoids, and alcohol also activate dopaminergic pathways linked to the nucleus accumbens. Some researchers have proposed that all positive reinforcement, including the reinforcement associated with food reward and sex, critically depends on this dopaminergic circuit.

Dopamine release from mesolimbic dopaminergic neurons may play more than one role in the genesis of drug seeking and drug dependence. Dopamine release has been postulated to facilitate learning which events and behaviors lead to important consequences for the organism and to alert the organism to pay greater attention to such events. In this way, drug-induced dopamine release leads to a greater salience of drug-using opportunities and is linked to wanting and craving.

However, the diverse categories of drugs that activate the mesolimbic dopaminergic system do so by distinct mechanisms, and most have actions on many other neural systems. Reinforcing mechanisms are briefly described in the chapters devoted to specific drugs; however, only a few examples are given. The ventral tegmental area dopaminergic neurons have both nicotinic and g-aminobutyric acid (GABA) receptors. These neurons normally are inhibited by GABAergic activity. The GABAergic neurons acting on the ventral tegmental area express μ- and d-opioid receptors. When these receptors are activated by μ opioids, GABAergic transmission is inhibited and the dopaminergic ventral tegmental area neurons become more active and release dopamine in the nucleus accumbens. However, opioids can also act directly on neurons in the nucleus accumbens, independent of dopamine action.

As a reinforcing drug, cocaine acts primarily at the nerve endings of the serotonergic, dopaminergic, and noradrenergic neurons. When transmitters are released from these neuron into the synapse, they are transported back into the nerve endings by transporter proteins. By occupying these transporter sites, cocaine prevents the reuptake of the transmitters, thus increasing their concentration in the synapse. Cocaine's binding to the dopamine transporter is primarily responsible for its reinforcing effects, but the actions on other neurotransmitters also influence its subjective effects. Amphetamine too, increases dopamine levels at the synapse and binds to the dopamine transporter to some degree. But amphetamine actions at the transporter are not as important as its major action, which is to displace dopamine and norepinephrine from their storage sites in the neuron and thereby lead to their release.

Alcohol is no longer believed to act like a general anesthetic, altering neuronal membranes. Instead, at clinically relevant concentrations its actions may be exerted more selectively on specific receptors and neurotransmitter systems. At clinically relevant concentrations these actions include enhancing the inhibitory action of GABAergic neurotransmitters (by increasing the sensitivity of the GABA receptor) and reducing the excitatory actions of glutamatergic neurotransmitters (by altering

the response of the *N*-methyl D-aspartate [NMDA] receptors). By its blocking actions at the NMDA receptor, ethanol can indirectly alter the release of other neurotransmitters (e.g., serotonin, dopamine, norepinephrine, glutamate, aspartate, and GABA). Low doses of alcohol increase dopamine levels in the nucleus accumbens and elevate brain serotonin concentration. Various regions of the brain differ in their sensitivity to these actions of ethanol. The endogenous opioid system may be involved in some aspects of the mood-elevating effects of alcohol, since the opioid antagonist naloxone reduces alcohol self-administration in animals and the antagonist naltrexone reduces relapse rates in treated alcoholics.

Mesolimbic dopaminergic neurons have multiple nicotinic cholinergic receptors on their cell bodies and terminals in the nucleus accumbens. When activated, these receptors increase dopamine release. Interestingly, regular exposure to tobacco smoke containing nicotine may be more reinforcing than nicotine itself because other chemical entities in tobacco inhibit brain monoamine oxidase type A (MAO_A) and MAO_B, which are involved in the regulation of intraneuronal stores of dopamine. This inhibition increases the amount of dopamine available for release when the dopaminergic neurons are activated.

Drugs can also be reinforcers by terminating aversive states; some of these actions involve dopaminergic systems, but others do not. Some researchers argue that compulsive drug use can be explained on the basis of the positive reinforcing effects of drugs without any need to invoke alleviation of withdrawal distress or any obvious source of antecedent pain or dysphoria. Furthermore, they argue, craving is primarily associated not with cues that evoke withdrawal but with those that evoke memories of positive reinforcement (euphoria). However, evidence now indicates that even when there are no obvious and dramatic withdrawal symptoms (e.g., cocaine, nicotine), adaptive changes in the reward system result in a relative dopaminergic deficiency state (measurable as decreased dopamine levels in the nucleus accumbens) when drug use is stopped or its action ceases. This deficiency state is experienced as dysphoria or anhedonia. Quite often the same drug-using behavior that terminates this dysphoria moves the system to a hyperdopaminergic state associated with euphoria. In short, the behaviors associated with chronic drug use are typically driven by both the avoidance of dysphoria (negative reinforcement) and the pursuit of euphoria (positive reinforcement).

The sensitivity of neural systems to reinforcing drugs such as cocaine and opioids is enhanced by corticosteroids. In animal models, a variety of stresses acting through release of corticotropin-releasing factor (CRF) and the hypothalamic pituitary-adrenal axis can sensitize neural systems and trigger reinitiation of drug taking. There is ample clinical evidence that such stresses can act similarly in drug-dependent individuals immediately following withdrawal and for long periods thereafter. In addition, some drugs may sensitize neural systems to the reinforcing effects of the drug.

Learning and Conditioning Drug use, whether occasional or compulsive, can be viewed as behavior maintained by its consequences. Any event that strengthens an antecedent behavior pattern can be considered a reinforcer of that behavior. In that sense certain drugs reinforce drug-taking behavior. Drugs can also reinforce antecedent behaviors by terminating some noxious or aversive state, such as pain, anxiety, or depression. In some social situations the use of the drug, quite apart from its pharmacological effects, can be reinforcing if it results in special status or the approval of friends. Social reinforcement can maintain drug use until the effects of primary reinforcement or reinforcement by alleviation of withdrawal symptoms come into play. Each use of the drug evokes rapid positive reinforcement, either as a result of the *rush* (the drug-induced euphoria), alleviation of disturbed affects, alleviation of withdrawal symptoms, or any combination of these effects. In addition, some drugs may sensitize neural systems to the reinforcing effects of the drug. With short-acting substances, such as heroin, cocaine, nicotine, and alcohol, such reinforcement occurs several times a day, day in and day out, creating powerfully reinforced habit patterns. Eventually, the paraphernalia (needles, bottles, cigarette packs) and behaviors associated with substance use can become secondary reinforcers as well as cues signaling availability of the substance, and in their presence, craving or a desire to experience the effects increases. With socially acceptable substances, such as tobacco, use becomes so woven into the matrix of daily functioning that some users are reminded of the substances when performing ordinary tasks. Stresses can also act as cues that induce drug taking, particularly in the postwithdrawal period.

Classical Conditioning In addition to the operant reinforcement of drug-using and drug-seeking behaviors, other learning mechanisms probably play a role in dependence and relapse. Opioid and alcohol withdrawal phenomena can be conditioned (in the Pavlovian or classic sense) to environmental or interoceptive stimuli. Such conditioning has been demonstrated in both laboratory animals and abstinent and methadone-dependent human volunteers. For a long time following withdrawal (from opioids, nicotine, or alcohol), the addict exposed to environmental stimuli previously linked with substance use or withdrawal may experience conditioned withdrawal, conditioned craving, or both. The increased feelings of craving are not necessarily accompanied by symptoms of withdrawal. The most intense craving is elicited by conditions associated with the availability or use of the substance, such as watching someone else use heroin or light a cigarette or being offered some drug by a friend. Some workers now believe that the cues that induce memories of drug-induced euphoria are more important for stimulating craving and in predisposing to relapse than either protracted or conditioned withdrawal. Those learning and conditioning phenomena can be superimposed on any preexisting psychopathology, but preexisting difficulties are not required for the development of powerfully reinforced substance-seeking behavior.

Withdrawal Syndromes and Negative Reinforcement Although positive reinforcement is a powerful etiological factor in the genesis of cocaine, amphetamine, and (in some cases) opioid dependence, aversive withdrawal phenomena and negative reinforcement may be equally important influences for a number of other drugs and dominant influences for others. One example of this is seen in most persons who become dependent on benzodiazepines taken in the course of treatment for anxiety syndromes. When drug use is interrupted, some seem to experience a reappearance of the original symptoms; others have new distressing symptoms indicating withdrawal. The use of benzodiazepines alleviates both kinds of aversive states. In either case the drug is acting as a negative reinforcer in perpetuating drug use. Benzodiazepines can induce euphoria in alcoholic patients or in persons with histories of sedative abuse, but they are not reliably euphorigenic in normal, nonalcoholic persons. Benzodiazepine anxiolytic agents may induce euphoria in nondependent, nonanxious persons, but such instances are rare relative to the number of those who experience only relief of anxiety.

In most clinical situations, even among users of highly euphoric illicit drugs, the distinction between positive and negative reinforcing effects does not exist. The alcoholic, the heavy smoker, and the heroin user may experience, simultaneously or sequentially, relief of withdrawal, a sense of ease, and perhaps alleviation of dysphoria and depression. With intravenous drugs there may also be a sudden rush of intense pleasure.

Long-Lasting Changes Associated With Chronic Drug Use After long-term use, most drugs of abuse produce adaptive changes in the brain that are manifested as acute and chronic withdrawal syndromes when drug use ceases. How these changes are produced, how long they persist after cessation of drug use, and how they contribute to relapse are still being explored. But much progress has occurred, as is illustrated by several examples of recent developments.

Tolerance and dependence on opioids involves several mechanisms. Opioid agonist binding to the opioid receptors results in an inhibition of adenylyl cyclase and lower intracellular cyclic adenosidine monophosphate (cAMP) concentrations. Long-term exposure elicits compensatory upregulation of the cAMP pathway, internalization of μ- and d-receptors, and a decrease in the number of G proteins, which couple the receptors to the second messengers and ion channels. Upregulation of adenylyl cyclase is mediated by the transcription factor cAMP response element–binding protein (CREB), which also plays a role in the generation of distinct and persistent Fos-like proteins, which are also thought to be involved in tolerance. As a result of upregulation of cAMP, GABAergic neurons innervating the ventral tegmental area become hyperactive when opioids are withdrawn, thus inhibiting dopaminergic neurons. Such a mechanism may account, in part, for the dysphoria and anhedonia of opioid withdrawal. In addition, chronic opioid use reduces the size of dopamine neurons in the ventral tegmental area; increased production of dynorphin may also serve to inhibit dopaminergic activity at the ventral tegmental area and nucleus accumbens. The glutamatergic system is also involved in opioid adaptation, since NMDA receptor sensitivity is altered by opioids and NMDA antagonists can alter the development of opioid tolerance and physical dependence.

With chronic alcohol use, affected neurons develop adaptive changes that include, among a number of others, supersensitivity or increased numbers of NMDA receptors. When the alcohol is withdrawn the actions of excitatory neurotransmitters at supersensitive NMDA receptors are postulated to produce the hyperexcitability of alcohol withdrawal, including hyperactivity in noradrenergic systems and glutamate-induced neuronal excitotoxicity. Alcohol-dependent patients tested 1 week and 1 month after cessation of alcohol use had cerebrospinal fluid with substantially lower concentrations of GABA and substantially higher concentrations of the excitatory transmitters glycine aspartate, glutamate, and *N*-acetylaspartylglutamate (NAAG) than that of healthy controls. Although these changes may be trait markers rather than consequences of long-term alcohol use, they are what might be expected to result from withdrawal of alcohol after adaptive changes have occurred. Most agents currently used to treat alcohol withdrawal act directly or indirectly at GABA receptors, and perhaps those that act at NMDA sites may provide alternative or even superior therapeutic agents.

Nicotine tolerance may involve desensitization of nicotinic receptors. However, chronic nicotine use increases the number of nicotinic cholinergic receptors, and the mechanisms underlying the nicotine withdrawal syndrome remain unclear. From the symptoms, which include craving, inability to concentrate, irritability, increased appetite, dysphoria, and (sometimes) depression, some hypoactivity in dopaminergic systems is likely. Given the many other actions of nicotine on learning, attention, arousal, and appetite, changes in systems in addition to the mesolimbic are probably also involved.

Cocaine and amphetamines can induce tolerance, dependence, and sensitization, depending in part on whether exposure is continuous or intermittent.

One way to determine the contribution of negative reinforcement to the motivation to continue using a substance or to relapse after withdrawal is to introduce agents that can modify withdrawal syndromes or aversive states. Abundant evidence shows that when psychological interventions are held constant, noninhaled nicotine (delivered by transdermal patches [Nicoderm] or nicotine gum [Nicorette]) significantly increases the probability that smokers trying to quit will be successful. Neither nicotine gum nor transdermal patches produce positive reinforcing effects, but they do alleviate aspects of the nicotine withdrawal syndrome. Thus, it is reasonable to infer that although the symptoms may not be life threatening, the avoidance of nicotine withdrawal plays a significant role in continued smoking and relapse. However, evidence suggests that for some, nicotine (or some other component of tobacco) controls negative effects other than those usually associated with withdrawal. Persons with histories of major mood disorder are more likely to become regular smokers if they try cigarettes and may experience symptoms of depression when they try to stop smoking; those symptoms are suppressed by returning to smoking. Heroin addicts treated with oral methadone or sublingual buprenorphine (Subutex) experience a reduction in opioid withdrawal symptoms but little or no euphoric effects from those agents. Yet, such treatment dramatically reduces elf-administration of heroin. Such findings support the view that acute and protracted opioid withdrawal (or opioid suppression of aversive affects) is an important factor in the perpetuation of heroin use and relapse after withdrawal. Similarly, acamprosate, a structural analogue of glutamate is postulated to reduce relapse in alcoholics following alcohol withdrawal by dampening the hyperexcitability in the glutamatergic system.

Conditioned Withdrawal and Stress Sensitivity In addition to the direct contribution of withdrawal phenomena to the perpetuation of drug use are the indirect effects exerted through learning mechanisms. The regular recurrence of withdrawal-induced aversive states provides ample opportunity for those states to become linked through learning to environmental cues and other mood states, and the rapid relief of withdrawal by drug use results in repeated reinforcement of drug-taking behavior. Long after there are measurable manifestations of acute withdrawal, certain moods or environmental cues can evoke components of the original withdrawal state along with urges to use the drug again. Considerable evidence shows that in former opioid addicts, stress can trigger both craving and relapse, and dysregulation of the hypothalamic-pituitary-adrenal axis persists for long periods after drug cessation.

How long withdrawal phenomena, stress sensitivity, or both continue to contribute to risk of relapse is not clear. Substantial evidence supports a withdrawal syndrome period for alcohol, opioids, and certain sedatives with subtle disturbances of mood, sleep, and cognition that persists for many weeks or months after the acute syndrome subsides. Whether the dysregulation of the hypothalamic-pituitary-adrenal axis is causally related to protracted withdrawal or has a similar time course is still uncertain.

Biological Factors—Vulnerability The children of alcoholic parents are at higher risk for developing alcoholism and drug dependence than are children of nonalcoholic parents. Dependence on other drugs also shows a familial pattern. The increased risk is partly due to environmental factors (parental modeling, neglect, early child abuse), but genetic factors are also important. Numerous studies of laboratory animals have revealed genetically transmitted differences in the reinforcing effects of alcohol and various drugs such as cocaine and opioids and show that genetic factors powerfully influence sensitivity to toxic effects. The evidence for genetic factors in human vulnerability to alcoholism and other drug dependence is derived most convincingly from twin and adoption studies, but family studies are also revealing. Several studies of twins have found a higher concordance rate for alcoholism among identical twins than among fraternal twins. Although identical twins are generally believed to have more social contact than fraternal twins, when the effects of environmental factors are adjusted statistically, genetic factors are still found to have a major influence on the likelihood of becoming dependent. Indeed, in one population-based twin study 48 to 58 percent of the variation in liability to dependence was attributable to genetic factors; the remainder was due to general environmental influences not shared by family members.

In studies of 3372 Vietnam-era veteran twin pairs, the concordance rates for dependence on at least one illicit drug were higher for monozygotic twins (26.3 percent) than for dizygotic (16.5 percent) twins. Generally, overall rates of dependence did not differ among these veterans and contemporary civilians. Biometric modeling identified both common (shared) and drug-specific genetic vulnerability factors as well as general and drug-specific effects of family and nonfamily environment. In the common vulnerability models, 31 percent of the variance for common (shared) vulnerability was due to additive genetic factors, 25 percent to family environmental effects, and 44 percent to nonfamily environmental effects. The importance of common (shared) genetic factors versus drug-specific genetic factors varies considerably for different categories of drugs. For marijuana, stimulant, and sedative abuse, common genetic vulnerability factors accounted for most of the genetic variance, with unique specific genetic factors accounting little. For psychedelics, no specific genetic influence was found. For heroin, 54 percent of the total variance was due to genetic factors, with 38 percent (70 percent of total genetic variance) contributed by unique genetic factors and only 16 percent by common (shared) ones. Another analysis of data from this group of veterans showed that both genetic and environmental factors influenced the initiation of cigarette smoking, but genetic effects accounted for 70 percent of the variance in the persistence of smoking for those who became regular smokers. This study, which is consistent with other genetic studies of smoking, found the genetic contribution to the persistence of smoking to be as great as or greater than the genetic contribution in the genesis of other psychiatric disorders, including alcoholism.

Twin studies in women have revealed strong genetic contributions to the use of caffeine and development of caffeine tolerance, dependence, and withdrawal.

Family studies also point towards general and drug-specific vulnerability factors. In a study of alcoholic probands and their siblings, about 50 percent of the brothers and 25 percent of the sisters met lifetime criteria for alcohol dependence. Compared with controls, these siblings also showed higher rates of tobacco, cocaine, and marijuana use, but the siblings of subjects who were dependent on alcohol and another drug, (presumably a more severe form of dependence), were not more likely to develop alcoholism than siblings of subjects who were dependent only on alcohol. However, siblings of probands who were dependent on both alcohol and marijuana had an elevated risk for marijuana dependence; siblings of probands dependent on alcohol and cocaine were more likely to become cocaine dependent. Statistical analysis that controlled for access to the drugs still showed specific family clustering.

Studies of boys adopted soon after birth have shown higher rates of alcoholism among those whose biological fathers were alcoholics than among those whose biological fathers were not. Some adoption studies pointed toward subtypes of alcoholism among men: one is a later-onset disorder that is less severe and far more sensitive to environmental factors (type I) and the other is associated with early onset, antisocial behavior and criminality in the biological fathers, and a stronger genetic basis for the increased vulnerability (type II). The hypothesis that two genetically distinct types of alcoholism (type I and type II) exist has been criticized on the grounds that it is essentially a relabeling of the older primary-secondary categorization. In the latter, alcohol-dependent persons who do not have antisocial personality disorder are designated as having primary alcoholism; those who first exhibit antisocial personality disorder and later develop alcoholism are designated antisocial personality disorder with secondary alcoholism. Also, several groups have been unable to use the type I and type II criteria to categorize patients with alcohol dependence accurately in clinical studies. However, arguments about the validity of the type I–type II categorization do not diminish the importance of genetic factors in vulnerability to developing alcohol dependence. The results of a large-scale efforts to identify the genes that contribute to vulnerability to alcoholism are now emerging.

As many as one third of alcohol-dependent persons have no family history of the disorder. Men are more likely to develop alcoholism than are women (fourfold to fivefold in the United States). This is true across every culture studied, probably reflecting, in part, social sanctions on drug use and deviant behavior by women. But it is also postulated that women are less likely to drink heavily because they are less tolerant to alcohol. Women who do drink heavily run the same risk of developing alcoholism as men who drink heavily, and women who use illicit drugs are about as likely to develop dependence as men who use such drugs.

In some, but not all studies alcohol-dependent persons are at far higher risk for developing other varieties of drug dependence. A more consistent finding is that drug-dependent persons also are at high risk for alcoholism and often have a family history of alcoholism. Such findings are consistent with data from the twin studies that have found general as well as drug-specific vulnerability factors.

Most researchers believe that no single gene will be found to account for the complexities of inherited risk for drug and alcohol dependence. Some genetic factors may not increase vulnerability to alcoholism but decrease it. A genetically determined variation in the activity of enzymes that metabolize alcohol (alcohol dehydrogenase and aldehyde dehydrogenase [ALDH]), common among some Asian groups, results in high levels of acetaldehyde in response to alcohol ingestion. The effect is to cause alcohol flush reaction and to exert some deterrent effect on alcohol ingestion. Alcoholism is lower among many Asian groups than among whites. Further, Asians with alcoholism are much less likely to have the inactive form of the ALDH enzyme.

Biological and Behavioral Differences Studies exploring how persons with and without family histories of alcoholism might differ have involved measures of personality, drug-use and alcohol-use patterns, psychomotor and cognitive performance, electrical activity of the brain, endocrine responses to challenges with alcohol and other substances, as well as measures of receptor numbers and affinities and enzyme activities (e.g., MAO) in peripheral tissues (e.g., blood platelets and lymphocytes). One finding that has been replicated is that under some conditions, the electrical response of the brain that occurs about 300 ms after a sensory stimulus (the P300 wave) has a smaller amplitude in nondrinking sons and daughters of alcoholic fathers than in control subjects without family histories of alcoholism. The decreased amplitude is believed to reflect a decreased capacity to recognize and interpret complex environmental stimuli. Most studies have found

no differences in intelligence among subjects with and without family histories of alcoholism. However, the results of personality studies are conflicting; some find no differences and others find greater impulsivity, adventurousness, and sensation seeking among those with a positive family history. Studies of the drinking patterns of adolescent and young adult sons of alcoholic persons also have not yielded consistent results; some (but not all) studies show that sons of alcoholic parents are heavier drinkers. Other studies have compared the subjective, motoric, and endocrine responses of young men with and without family histories of alcoholism following challenge exposures to alcohol and other potentially euphoriant drugs (such as benzodiazepines). Sons of alcoholic fathers seem to be more tolerant to the intoxicating effects of modest doses of alcohol, and in some (but not all) studies, higher doses of alcohol produced smaller changes in their prolactin and cortisol concentrations. Furthermore, one study found that sons who had smaller responses to test doses of alcohol at age 20 (i.e., were more tolerant) were fourfold more likely to have developed alcoholism 8 years later. Another study of sons of alcoholic parents found that those who had exhibited smaller electroencephalographic (EEG) alpha frequency responses to alcohol were more likely to be alcohol dependent at 10-year follow-up.

The results of studies using benzodiazepine challenges are also not consistent; one showed a greater euphoric response to alprazolam (Xanax) in sons of alcoholic parents, and another showed no difference between positive and negative family-history groups after a dose of diazepam.

A number of studies have shown that conduct disorder and early childhood aggression are associated with a substantial increase in the likelihood of early involvement with illicit drug use and development of dependence on alcohol and illicit drugs. Considerable evidence supports a role for both genetics and environmental factors in the development of conduct disorder. Antisocial personality disorder represents an independent additional risk factor for addictive disorders. The effects of antisocial personality disorder and family history of an substance-related disorder appear to be additive rather than synergistic. It seems possible that in some of the studies of children and young people at high risk for later drug dependence, the electrophysiological differences, cognitive deficits, and personality differences reflected the presence of conduct disorder or antisocial personality disorder rather than a family history of alcoholism per se.

Psychodynamic Factors and Psychopathology Early psychoanalytic formulations postulated that drug users, in general, suffered from either a special form of affective dysregulation (tense depression) that was alleviated by drug use or from a disorder of impulse control in which the search for pleasure was dominant. More-recent formulations postulate ego defects, which are evinced by the addict's inability to manage painful affects (guilt, anger, anxiety) and to avoid preventable medical, legal, and financial problems. The newer formulations postulating ego defects are to some degree the older formulations with a modest change in terminology that gives greater weight to the inability to cope with painful affects than to the intensity or abnormality of the affects per se. It is postulated that some substances pharmacologically and symbolically aid the ego in controlling those affects and that their use can be viewed as a form of self-medication. For example, it has been suggested that opioids help users control painful anger, that alcohol helps alcoholics control panic, and that nicotine may help some cigarette smokers control symptoms of depression. Although it is conceded that some of those observations may reflect problems produced by long-term use, the psychodynamic perspective is that the psychopathology is the underlying motivation for initial use, dependent use, and relapse after a period of abstinence. However, traditions of passivity and uncovering techniques derived from the psychoanalysis of neurosis are poorly suited to the treatment of most drug addicts. Further, some addicts have great difficulty differentiating and describing what they feel, a difficulty that has been called alexithymia (i.e., no words for feelings).

Family Dynamics One family member's substance abuse is often influenced by substance-using behaviors of others in the family, and these complex interrelationships can profoundly affect their lives. An understanding of the relationships among substance-using patients and their families is relevant for understanding the etiology of substance dependence and its treatment and for helping other family members to cope with problems associated with the substance-using behavior.

More has been written about the families of alcohol-dependent persons and heroin users than about families affected by users of other drugs. Similarities between the family dynamics in these two prototypical dependencies have led researchers and clinicians to assume that certain general principles apply to all varieties of substance dependence. The observation that alcoholism is commonly found in the families of those seeking treatment for other types of dependence, that alcohol-dependent persons are often dependent on other substances as well, and that those addicted to illicit drugs are often alcoholic suggests that there are common features among families with an addicted member. However, there are few data to suggest that the families of those dependent on tobacco or benzodiazepines are as dysfunctional as those affected by alcohol, opioids, or cocaine.

It is not always clear to what degree one family member's behavior causes the substance-using behavior of another or is primarily a response to that behavior. Some writers emphasize that the addiction is a symptom that provides a displaced focus for conflict among other family members and that the user (the designated patient) may be playing a role in maintaining the homeostasis of a dysfunctional family. At the same time, addiction often arises in families in which one or both parents (and sometimes grandparents) have drug or alcohol problems and other psychopathology. Some characteristics commonly observed both in families of persons who are alcohol dependent and of those addicted to illicit drugs are multigenerational drug dependence; a high incidence of parental loss through divorce, death, abandonment, or incarceration; overprotection or overcontrol by one parent (usually the mother), whose life is inordinately dependent on the behavior of the addicted offspring (symbiotic relationships); distant, cold, disengaged, or absent father (when the father is alive); defiant drug-using child, who appears to be engaged with peers but remains unusually dependent on the family well into adult life (pseudo-independence). The actual family dynamics are difficult to characterize because the family members' self-reports about their relationships do not reliably correspond to what outsiders observe. Such families typically do not describe themselves in the way that family therapists see them. Some workers have proposed that unresolved family grief plays a role in the genesis of drug addiction in a family member and that such families cannot deal effectively with separation because of previous losses. Despite the pathological interdependence between the addict and other family members, the addict is often described as passive, dependent, withdrawn, and unable to form close relationships.

Despite all the apparent pathology found in families, in many instances the family brings the substance user into treatment, and the patient often believes that it is the family that is most likely to be helpful in recovery. Furthermore, clinicians now generally believe that involving families in treatment is important, if not essential, to effective intervention. One aspect of treating families is dealing with the tendency of some members to shield the patients from the consequences of their substance use, a behavior usually labeled by clinicians as “enabling” but usually experienced by the family member as loving, supporting, accepting, and protecting. A variation on family therapy, sometimes called network therapy, involves enlisting family members and close friends as allies of the therapist to provide social support and reinforcement of drug-abstaining behaviors. The persons selected to fulfill this role function as part of a treatment team rather than as patients.

CODEPENDENCE The terms “coaddiction,” “coalcoholism,” or more commonly “codependency” or “codependence” have recently come into vogue to designate the behavioral patterns of family members who have been significantly affected by another family member's substance use or addiction. The terms have been used in various ways, and there are no established criteria for codependence, a concept that some writers have expanded far beyond its origins to encompass any personality disorder that involves difficulty in expressing emotions. However, many have criticized the expanded concept of codependence as a largely invalid notion based solely on anecdote. The following summary of some characteristics frequently described as aspects of codependence is not meant to imply the validity of a unitary syndrome.

Enabling Enabling was one of the first and more agreed upon characteristics of codependence or coaddiction. Sometimes family members feel that they have little or no control over the enabling acts. Either because of the social pressures for protecting and supporting family members or because of pathological interdependencies, or both, enabling behavior often resists modification. Other characteristics of codependence include an unwillingness to accept the notion of addiction as a disease. The family members continue to behave as if the substance-using behavior were voluntary and willful (if not actually spiteful) and the user cares more for alcohol and drugs than for the members of the family. This results in feelings of anger, rejection, and failure. In addition to those feelings, the family members may feel guilty and depressed because the addict, in an effort to deny loss of control over drugs and to shift the focus of concern away from their use, often tries to place the responsibility for such use on the other family members, who often seem willing to accept some or all of it.

Denial Family members, like the substance users themselves, often behave as if the substance use that is causing obvious problems were not really a problem; that is, they engage in denial. The reasons for the unwillingness to accept the obvious vary. Sometimes denial is self-protecting, in that the family members believe that if there is a drug or alcohol problem, then they are responsible.

Like the addicts themselves, codependent family members seem unwilling to accept the notion that outside intervention is needed and, despite repeated failures, continue to believe that greater will power and greater efforts at control can restore tranquility. When additional efforts at control fail, they often attribute the failure to themselves rather than to the addict or the disease process, and along with failure come feelings of anger, lowered self-esteem, and depression.

Other Problems Some clinicians have reported high levels of somatic disorders, such as ulcers, colitis, and migraine, among family members of alcoholic persons and addicts and have attributed those illnesses to stress or a somatic expression of the feelings engendered by trying to cope with the family member's addiction. However, in light of the findings that there may be a genetic basis for somatization disorders among the daughters of certain subtypes of alcoholic persons, it is not clear that all of the illnesses seen among the family members of substance users are responses to the stresses of living with an addict.

Other Factors There are other factors that influence the pattern of use and cessation of any given substance. For example, the decision not to use a substance also has consequences that can be aversive or reinforcing, and evidence indicates that when the rewards of not using the substance are high, the likelihood of use is reduced. In addition, many of the substances associated with dependence act directly on systems that subserve both motivation and decision making, raising questions about whether use is always influenced solely by its consequences (learning processes). The cognitive processes and skills that would ordinarily subserve decision-making appear to be impaired by alcohol, barbiturates, cannabis, and several other categories of self-administered agents. Thus, whereas substance use is influenced by learning, the substances also alter the brain itself. This suggests additional problems and possibilities for intervention. Evidence is accumulating that limited cognitive skills reduce the likelihood of successful recovery from substance use and that coping skills can help a person avoid or deal with aversive affective states, environmental stresses, and situations that are associated with a high risk for substance use.

Other factors that influence the course of substance use and dependence are difficult to operationalize or teach or prescribe, but they deserve mention. Studies of the natural history of substance use indicate that recovery is powerfully influenced by the support of family and friends. Many persons report that hope, faith, formal religious affiliation, or the sustaining love of some significant person was more important to their recovery than any specific treatment.

Multiple Factors The biopsychosocial general model of substance dependence presented here does not attempt to assign a weight or special significance to any one factor or interaction. The implication is that for different categories of drugs, different factors may play more or less powerful causal roles in perpetuating substance use or facilitating relapse. For example, positive reinforcing effects may be more important for the development of cocaine dependence, whereas acute and protracted withdrawal phenomena may be more important in the return to opioid use following withdrawal. Even with the same substance, different factors may be more or less important for different persons. Thus, the emergence of depressive symptoms may make it difficult for some cigarette smokers to quit, particularly those with a history of major depressive disorder, and those persons may be helped by antidepressants. Such a multifactorial model implies that certain treatments or interventions may be more effective for one substance category than another and that even among persons using the same substances, different treatments may be indicated.

Figure 11.1-1 also implies that the notion of dependence is not a property of any one element but an abstraction inferred from the relations among the elements of the system. While it is convenient (and required by DSM-IV) to see dependence as a disorder located within a person, any interpretation that overemphasizes one part of the system, whether the biology of the person, social influences, or behavior, is missing part of the nature of dependence.

Comorbidity *Comorbidity* is the co-occurrence of two or more psychiatric disorders in a single patient. A high prevalence of additional psychiatric disorders is found among persons seeking treatment for alcohol, cocaine, or opioid dependence. Although opioid, cocaine, and alcohol abusers with current psychiatric problems are more likely to seek treatment, it should not be assumed that those who do not seek treatment are free of comorbid psychiatric problems; such persons may have social supports that enable them to deny the impact that drug use is having on their lives. Two large epidemiological studies have shown that even among representative samples of the population, those who meet the criteria for alcohol or drug abuse and dependence (excluding tobacco dependence) are far more likely to meet the criteria for other psychiatric disorders also. In the NCS, 51 percent of those who met the criteria for a lifetime addictive disorder received at least one additional mental disorder diagnosis; in the earlier ECA study, the comparable figure was 38 percent. In the ECA study, among those diagnosed with drug dependence the most common additional diagnosis was alcohol abuse-dependence, followed in frequency by antisocial personality disorder, phobic disorders, and major depression for men and phobic disorders, major depression, and dysthymia for women. Almost every psychiatric diagnosis was more common among those who met the criteria for drug dependence, with notable increases in odds ratios for alcoholism, antisocial personality disorder, and mania among women, and for mania, antisocial personality disorder, and dysthymia among men. Both men and women with drug abuse-dependence are at a substantially higher risk for schizophrenia. The extent of comorbidity among individuals in the ECA study is illustrated in Figure 11.1-2.

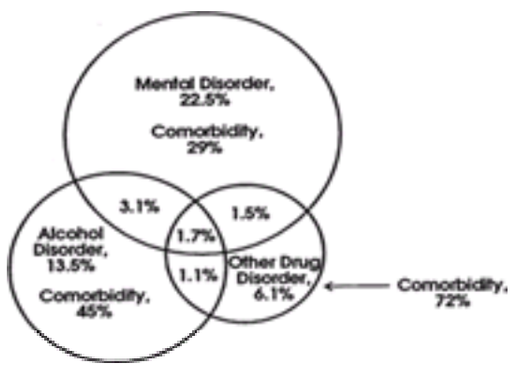

FIGURE 11.1-2 Lifetime prevalence of comobid mental and addictive disorders in the United States, combined community and institutional five-site Epidemiologic Catchment Area data, standardized to the U.S. population. (Reprinted with permission from Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK: Comorbidity of mental disorders with alcohol and other drug abuse. JAMA *264:*2511, 1990.)

In general, the probability of comorbidity is higher for those with a lifetime diagnosis of an opioid or cocaine disorder than for those with a diagnosis of cannabis abuse. Among people in prison the comorbidity rates were even higher than in the general population; addictive disorders were found in 92 percent of prisoners with schizophrenia, 90 percent of those with antisocial personality disorder, and 89 percent of those with bipolar disorders. Among persons with mental disorders seeking treatment in psychiatric specialty settings, 20 percent have a current substance abuse disorder diagnosis.

The findings from the NCS largely confirm the observations of the ECA study that those with substance use disorders are substantially more likely to experience other mental disorders and that those with other mental disorders are far more likely to develop substance use disorders. The NCS also underscored the finding that although 52 percent of respondents had never experienced any DSM-III-R disorder and 21 percent had one such disorder, 13 percent had two disorders and 14 percent had three or more disorders. Furthermore, the 12-month prevalence of a disorder was more likely among those with more than one disorder: 59 percent of all of 12-month disorders occurred in the 14 percent with a lifetime history of three or more disorders, and 89 percent of severe 12-month disorders occurred in the same group.

These findings describe rather than explain comorbidity. They do not shed much light on the question of whether, or in which cases, drug use is at least initially an adaptive effort at self-medication, or whether those with a variety of psychiatric disorders are less able to cope with the effects of substance use and so are more likely to become dependent. It is also not clear whether psychiatric disorders increase the vulnerability to drug abuse and drug dependence or whether some common factor contributes to both. In some cases, however, there does appear to be a causal link between drug use and some psychiatric disorders. For example, evidence indicates that substance abuse (especially alcohol) can cause or increase the risk for depressive disorder; cocaine can increase the frequency of panic disorder; and cannabis, cocaine, and amphetamine use can aggravate or precipitate schizophrenic symptomatology. Some of these are drug-induced disorders (particularly some of the depressive symptoms seen in alcoholics) and clear with cessation of alcohol use. However, some psychiatric disorders (e.g., mood disorder and antisocial personality disorder) often antedate substance use and can be viewed as risk factors or predictors for substance abuse and dependence. This is particularly true of conduct disorder adult antisocial behavior, in which the symptoms often begin before the onset of problematic drug use. The NCS found that the odds of developing alcohol or drug dependence increased fivefold in the presence of conduct disorder without adult antisocial behavior and 10- to 14-fold if only adult antisocial behavior or both conduct disorder and antisocial behavior were present. Of the Axis I disorders, bipolar I disorder is more strongly related to dependence on alcohol or drugs than any other mood or anxiety disorder. In general, about 24.5 percent of those with a 12-month addictive disorder had a mood disorder as well, and 35.6 percent had an anxiety disorder. Overall, 42.7 percent of those with a 12-month addictive disorder had at least one 12-month Axis I mental disorder. In terms of lifetime disorders, 41 to 65.5 percent of those with a lifetime addictive disorder have a lifetime history of at least one Axis I mental disorder, while 51 percent of those with one or more lifetime mental disorders (Axis I or II) have a history of one or more addictive disorders. For lifetime conduct disorder or adult antisocial behavior, the rate of lifetime substance use disorder rises to 82 percent.

Although the possibility of recall bias exists, those with both an affective and an addictive disorder usually report that depression began earlier than substance use. However, temporal relationship between two disorders does not prove causality, even when the development of the first disorder is a predictor of both the likelihood and course of the subsequent disorder. There is the possibility, as has been suggested for smoking and depression, that both disorders are linked to some third common factor. In the NCS, a more chronic course of an addictive disorder was found for those who reported earlier development of primary anxiety disorder, conduct

disorder, or adult antisocial behavior but was not found with earlier onset of other mental disorders.

In the NCS, co-occurring mental disorders also influenced the likelihood of seeking treatment and the treatment sector from which service would be sought. As mentioned, those who had a substance dependence problem were far more likely to seek and receive treatment if they also had a co-occurring mental disorder. About one-third of people with a 12-month history of affective disorder received some treatment; but those who also had an addictive disorder were more likely to have received it in a specialty addiction treatment program.

A collaborative study of the genetics of alcoholism used extensive structured interviews to separate independent mood and anxiety disorders from those that occurred only within the context of active drinking or withdrawal. This study found that over a lifetime, independent mood disorder was less common in alcoholics (14 percent) than in controls (17.1 percent), although more than twice as many alcoholics (2.3 percent) as controls (1.0 percent) met criteria for bipolar disorder. Panic disorder and social phobia were also substantially more common as independent disorders among alcoholics. In general, in this study the large majority of alcohol-dependent men and women did not have independent mood or anxiety disorders. This suggests that the higher rates of co-occurrence of most anxiety and affective disorders found in epidemiological studies or clinical populations probably reflect substance (alcohol)-induced anxiety and mood disorders that will resolve without special intervention once drug use ceases.

TREATMENT

Many people who develop substance-related problems recover without formal treatment. For those who do seek help or advice, particularly those patients with less severe disorders, relatively brief interventions are often as effective as more intensive treatments. Since these brief interventions do not change the environment, alter drug-induced brain changes, or provide new skills, a change in the patient's motivation (cognitive change) probably best explains their impact on the drug-using behavior. For those individuals who do not respond or whose dependence is more severe, a variety of interventions appear to be effective. Although each section in this chapter discusses treatment relevant to the particular substance use disorder, the clinician sees few drug-dependent people who use only one drug. (Nicotine dependence may be an exception.) For example, among patients using an illicit drug, the most common additional diagnosis is alcohol dependence.

It is useful to distinguish among specific procedures or techniques (e.g., individual therapy, family therapy, group therapy, relapse prevention, and pharmacotherapy) and treatment programs. Most programs use a number of specific procedures and involve several professional disciplines as well as nonprofessionals who have special skills or personal experience with the substance problem being treated. The best treatment programs combine specific procedures and disciplines to meet the needs of the individual patient after a careful assessment. However, there is no generally accepted classification either for the specific procedures used in treatment or for programs making use of various combinations of procedures. This lack of standardized terminology for categorizing procedures and programs presents a problem, even when the field of interest is narrowed from substance problems in general to treatment for a single substance, such as alcohol, tobacco, or cocaine. Except in carefully monitored research projects, even the definitions of specific procedures (e.g., individual counseling, group therapy, and methadone maintenance) tend to be so imprecise that one usually cannot infer just what transactions are supposed to occur. Nevertheless, for descriptive purposes, programs are often broadly grouped on the basis of one or more of their salient characteristics: whether the program is aimed at merely controlling acute withdrawal and consequences of recent drug use (detoxification) or is focused on longer-term behavioral change; whether the program makes extensive use of pharmacological interventions; and the degree to which the program is based on individual psychotherapy, AA or other 12-step principles, or therapeutic community principles. For example, government agencies recently categorized publicly funded treatment programs for drug dependence as either methadone maintenance (mostly outpatient), outpatient drug-free programs, therapeutic communities, or short-term inpatient programs. However, these broad descriptions mask as much as they reveal, tend to confuse the setting with the procedures, and obscure differences in the etiological models underlying the treatments used in different programs.

Selecting a Treatment Not all interventions are applicable to all varieties of substance use or dependence, and some of the more coercive interventions used for illicit drugs are not applicable to substances that are legally available, such as tobacco. Changes in addictive behaviors do not occur abruptly, but rather through a series of stages. Five stages in this gradual process has been proposed: precontemplation, contemplation, preparation, action, and maintenance. For some types of addiction the therapeutic alliance is enhanced when the treatment approach is tailored to the patient's stage or readiness to change. For some drug use disorders, a specific pharmacological agent may be an important component of an intervention; for example, disulfiram, naltrexone (ReVia) or acamprosate for alcoholism; methadone, levomethadyl acetate (ORLAAM) (also called L-a-acetylmethadol [LAAM]) or buprenorphine (Buprenex) for heroin addiction; nicotine delivery devices or bupropion (Zyban) for tobacco dependence. Not all interventions are likely to be useful to health care professionals. For example, many youthful offenders with histories of drug use or dependence are now remanded to special facilities (boot camps), other programs for offenders (and sometimes for employees) rely almost exclusively on the deterrent effect of frequent urine testing, and a third group are built around religious conversion or rededication in a specific religious sect or denomination. In contrast to the numerous studies suggesting some value for brief interventions for smoking and for problem drinking, there are few controlled studies of brief interventions for those seeking treatment for dependence on illicit drugs.

In general, for those persons who are severely dependent on illicit opioids, brief interventions (such as a few weeks of detoxification, whether in or out of a hospital) have limited effect on outcome measured a few months later. Among patients dependent on cocaine or heroin, substantial reductions in illicit drug use, antisocial behaviors, and psychiatric distress are much more likely following treatment lasting at least 3 months. Such a time-in-treatment effect is seen across very different modalities, from residential therapeutic communities to ambulatory methadone maintenance programs. Although some patients appear to benefit from a few days or weeks of treatment, a substantial percentage of users of illicit drugs drop out (or are dropped) from treatment before they have achieved significant benefits. Some of the variance in outcome of treatment can be attributed to differences in the characteristics of patients entering treatment and by events and conditions following treatment. However, programs based on similar philosophical principles and using what seem to be similar therapeutic procedures vary greatly in effectiveness. Some of the differences among programs that seem to be similar reflect the range and intensity of services offered. Programs with professionally trained staffs that provide more-comprehensive services to patients with more severe psychiatric difficulties are more likely to be able to retain those patients in treatment and to help them to make positive changes. Differences in the skills of individual counselors and professionals can powerfully affect outcomes. Such generalizations concerning programs serving illicit drug users may not hold for programs dealing with those seeking treatment for alcohol or tobacco or even cannabis problems uncomplicated by heavy use of illicit drugs. In such cases, relatively brief periods of individual or group counseling can produce long-lasting reductions in drug use. The outcomes usually considered in programs dealing with illicit drugs have typically included measures of social functioning, employment, and criminal activity, as well as decreases in drug-using behavior. Treatment for alcoholism and other mental health problems generally has more limited expectations (e.g., reduction in alcohol use and symptoms of psychiatric disorders), although changes in the use of health care resources subsequent to treatment is sometimes an additional measure of treatment efficacy.

Measuring Treatment Outcome The latest published large multisite study of treatment, the Drug Abuse Treatment Outcome Study (DATOS, carried out from 1991 to 1993), interviewed patients at intake and 1, 3, and 12 months after treatment. As in previous multisite studies, sites selected were stable representatives of four major program types: drug-free outpatient, methadone maintenance, short-term residential (chemical dependency), and long-term residential (therapeutic community). Except at the methadone programs, which used group and individual counseling about equally, group counseling was the common element of the other treatments. Some antidepressants and antipsychotic agents were used in the nonmethadone programs, but they were incidental.

This study found a lower level of services available to patients seeking treatment then were available decade earlier. Also, the patients were older and more likely to have a variety of special medical problems (e.g., HIV positive, concurrent psychiatric disorders) and social needs (homelessness). Treatment outcomes were generally consistent with those of previous studies of drug treatment in the public sector. One year after treatment there were substantial decreases in drug use. Levels of weekly or daily cocaine use at 1 year were about 50 percent of pretreatment levels, with greater reduction for those who participated in treatment for 3 months or more. Daily heroin use was lower among patients who remained in methadone maintenance treatment than among those who left. Although cocaine use among patients treated with methadone was somewhat lower, the reduction could not be attributed to treatment. Alcohol and marijuana use did not decline significantly. There was also no apparent decrease in suicidal thoughts or increase in employment, and in contrast to a number of previous multisite studies, multivariate analysis in this study did not confirm the widely reported reduction in predatory and or high-risk sexual behaviors for those in methadone programs. Those who stayed in long-term residential treatment for 6 months or longer showed a major decrease in drug use from preadmission levels for all categories of drugs—66.4 to 22 percent for cocaine; 17.2 to 5.8 percent for heroin; alcohol and marijuana use reduced by more than half. These individuals also reported a 50 percent decrease in illegal activities and about a 10 percent increase in full-time employment.

In DATOS, outpatient drug-free and short-term inpatient programs had very few admissions in which the major drug problem was heroin; the most common presenting drug problem for both was cocaine, followed by alcohol and marijuana. Participation in the outpatient drug-free programs for 3 months or more was associated with a greater decrease in cocaine use at 1 year, (about 50 percent compared with those who stayed 3 months or less). But even 58 percent of those who stayed less than 3 months reported some decrease in cocaine use over preadmission levels. Patients who entered short-term inpatient programs also reported major decreases in drug use at 1 year, but there was no difference between those who stayed more than 2 weeks and those who stayed less than 2 weeks. Since the decision to enter any of the programs studied in DATOS was made by the patient, the study does not give much guidance to a clinician weighing a recommendation for a specific patient.

More guidance comes from a large-scale, random-assignment study of the treatment of alcoholics, which found that three distinct methods of delivering individual therapy over a 12-week period—12-step facilitation, cognitive-behavioral coping skills, and motivational enhancement (four sessions only)—produced comparable and generally quite favorable outcomes. Patient characteristics interacted significantly with the treatment in only one area, alcoholics with low-level psychiatric problems had better outcomes in terms of days of abstinence if assigned to 12-step facilitation rather than cognitive behavioral therapy. Patients who received individual therapy after a brief period of inpatient and intensive day-care treatment (aftercare) had better 1-year outcomes than those who began individual treatment as outpatients.

Currently, entry into treatment rarely reflects a truly informed choice aimed at matching the characteristics and needs of the patient with the capacities and skills of a provider. Findings from studies of public-sector programs serving drug users with relatively few social supports show that more-intensive services, such as vocational, health, and mental health services, increase retention and produce better outcomes at follow-up.

Influence of Philosophical Orientation The kinds of therapeutic procedures deemed valuable or essential by treatment professionals are profoundly affected by their philosophical orientation. For example, one study found that many professionals who adhere to a disease model of substance dependence view reduction of denial, acceptance of disease, need for lifelong abstinence, commitment to recovery, and affiliation with AA as the most important elements of intervention. In contrast, dealing with responsibility, instilling motivation and confidence, teaching relapse prevention, and avoiding high-risk situations were rated highest by psychologists espousing a behavioral model of dependence. Until quite recently, even physicians were unlikely to view pharmacological interventions as having significant value in treating alcoholism or most other forms of drug dependence, although some physicians did prescribe various forms of nicotine for tobacco dependence.

Many controlled studies over many years have shown that the use of illicit opioids (heroin) can be markedly reduced by supervised administration of oral methadone or LAAM. Because of government regulations, the use of these agents is currently limited to practitioners and programs who have obtained special licenses; such programs and practices are rigidly regulated. Buprenorphine, a partial opioid agonist, is also effective. Data also show that naltrexone can reduce relapse rates for alcoholics following withdrawal. Controlled studies conducted in Europe show that acamprosate, a drug believed to act via actions on the glutamatergic system, can also reduce alcoholism relapse rates. However, to date, the pharmacological agents available to treat substance-related disorders have not been widely used, even when there are few regulatory barriers. The relatively indifferent or negative attitudes of physicians toward the use of pharmacological agents in the treatment of alcoholism and drug dependence may change if new and more effective medications become available at reasonable cost and unencumbered by burdensome government regulations. However, at present there seems to be only a modest correlation between the evidence showing that a given intervention or procedure is effective and the likelihood that it will be widely used.

Treatment of Comorbidity—Integrated Versus Concurrent The treatment of the severely mentally ill (primarily those with schizophrenia and schizoaffective disorders) who are also drug dependent continues to pose problems for clinicians. Although some special facilities have been developed that use both antipsychotic drugs and therapeutic community principles, for the most part specialized addiction agencies have difficulty treating these patients. Generally, integrated treatment in which the same staff can treat both the psychiatric disorder and the addiction is more effective than either parallel treatment (a mental health and a specialty addiction program providing care concurrently) or sequential treatment (treating either the addiction or the psychiatric disorder first and then dealing with the comorbid condition.)

Services and Outcome The extension of managed care into the public sector has produced a major reduction in the use of hospital-based detoxification and virtual disappearance of residential rehabilitation programs for alcoholics. Unfortunately, managed-care organizations tend to assume that the relatively brief courses of outpatient counseling that are effective with private-sector alcoholic patients are also effective with patients who are dependent on illicit drugs and who have minimal social supports. For the present, the trend is to provide the care that costs least over the short term and to ignore studies showing that more services can produce better long-term outcomes.

Treatment is often a worthwhile social expenditure. For example, treatment of antisocial illicit drug users in outpatient settings can produce decreases in antisocial behavior and reductions in rates of HIV seroconversion that more than offset the treatment cost. Treatment in a prison setting can produce favorable decreases in postrelease costs associated with drug use and rearrests. Despite such evidence there are problems maintaining public support for treatment of substance dependence, both in the public and private sectors. This lack of support suggests that these problems continue to be viewed, at least in part, as moral failings rather than as medical disorders.

SUGGESTED CROSS-REFERENCES

Individual sections discuss in detail the relevant substances and treatment for their related disorders: alcohol-related disorders in Section 11.2; amphetamine-related disorders, Section 11.3; caffeine-related disorders, Section 11.4; cannabis-related disorders, Section 11.5; cocaine-related disorders, Section 11.6; hallucinogen-related disorders, Section 11.7; inhalant-related disorders, Section 11.8; nicotine-related disorders, Section 11.9; opioid-related disorders, Section 11.10; PCP-related disorders, Section 11.11; sedative-hypnotic–related disorders, Section 11.12, and anabolic-androgenic steroid abuse, Section 11.13. Brief psychotherapy is covered in Section 30.8; alternative therapies, in Section 30.10; and methadone (and other maintenance therapies) in Section 31.23. Drug and alcohol abuse among elderly persons is discussed in Section 51.3h.

SECTION REFERENCES

Akil H, Owens C, Gutstein H, Taylor L, Curran E, Watson S: Endogenous opioids: Overview and current issues. Drug Alcohol Depend *51:*127, 1998.

Anglin MD, Hser Y-I, Grella CE: Drug addiction and treatment careers among clients in the Drug Abuse Treatment Outcome Study (DATOS). Psychol Addict Behav *11:*308, 1997.

Anthenelli RM, Smith TL, Irwin MR, Schuckit MA: A comparative study of criteria for subgrouping alcoholics: The primary/secondary diagnostic scheme versus variations of the type 1/type 2 criteria. Am J Psychiatry *151:*1468, 1994.

Anthony JC, Warner LA, Kessler RC: Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. Clin Exp Psychopharmacol *2:*244, 1994.

Baumohl J, Jaffe JH: History of alcohol and drug abuse treatment in the United States. In *Encyclopedia of Drugs and Alcohol*, vol 3, JH Jaffe, editor. Macmillan, New York, 1995.

Beirut LJ, Dinwiddie SH, Begleiter H, Crowe RR, Hesselbrock V, Nurnberger JI Jr, Porjesz B, Schuckit MA, Reich T: Familial transmission of substance dependence: Alcohol, marijuana, cocaine, and habitual smoking. Arch Gen Psychiatry *55:*982, 1998.

Edwards G, Arif A, Hodgson R: Nomenclature and classification of drug- and alcohol-related problems. A WHO Memorandum. Bull WHO *99:*225, 1981.

*Gerstein DR, Harwood HJ, editors: *Treating Drug Problems*, vol 1. Committee for the Substance Abuse Coverage Study, Division of Health Care Services, Institute of Medicine. National Academy Press, Washington, DC, 1990.

Goldman D, Bergen A: General and specific inheritance of substance abuse and alcoholism. Arch Gen Psychiatry *55:*964, 1998.

Harrison PA, Fulkerson JA, Beebe TJ: DSM-IV substance use disorder criteria for adolescents: A critical examination based on a statewide school survey. Am J Psychiatry *155:*486, 1998.

Hubbard RL, Craddock SG, Flynn PM, Anderson J, Etheridge RM: Overview of 1-year follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). Psychol Addict Behav *11:*261, 1997.

*Institute of Medicine: *Broadening the Base of Treatment for Alcohol Problems*. National Academy Press, Washington, DC, 1990.

Institute of Medicine: *Pathways of Addiction*. National Academy Press, Washington, DC, 1996.

Inturrisi CE: Preclinical evidence for a role of glutamatergic systems in opioid tolerance and dependence. Semin Neurosci *9:*110, 1997.

Jaffe JH: Current concepts of addiction. In *Addictive States*, CP O'Brien, JH Jaffe, editors. Research Publications: Association for Research in Nervous and Mental Disease, vol 70. Raven, New York, 1992.

Jaffe JH, Knapp CM, Ciraulo DA: Opiates: Clinical aspects. In *Substance Abuse: A Comprehensive Textbook*, ed 3, JH Lowinson, P Ruiz, RB Millman, JG Langrod, editors. Williams & Wilkins, Baltimore, 1997.

*Johnston LD, O'Malley PM, Bachman JG: National survey results on drug use from the Monitoring the Future Study. College Students and Young Adults. National Institute on Drug Abuse, Rockville, MD, 1999.

Kaufman E: The family in drug and alcohol addiction. In *Comprehensive Handbook of Drug and Alcohol Addiction*, NS Miller, editor. Marcel Dekker, New York, 1991.

Kessler RC, Crum RM, Warner LA, Nelson CB, Schulenberg J, Anthony JC: Lifetime co-occurrence of DSM-III-R alcohol abuse and dependence with other psychiatric disorders in the national comorbidity survey. Arch Gen Psychiatry *54:*313, 1997.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry *51:*8, 1994.

Kessler RC, Nelson CB, McGonagle KA, Edlund MJ, Frank RG, Leaf PJ: The epidemiology of co-occurring addictive and mental disorders: Implications for prevention and service utilization. Am J Orthopsychiatry *66:*17, 1996.

Koob GF, Le Moal M: Drug abuse: Hedonic homeostatic dysregulation. Science *278:*52, 1997.

Kreek MH, Koob GF: Drug dependence: Stress and dysregulation of brain reward pathways. Drug Alcohol Depend *51:*23, 1998.

McLellan AT, Grissom GR, Zanis D, Randall M, Brill P, O'Brien CP: Problem-service 'matching' in addiction treatment. Arch Gen Psychiatry *54:*730, 1997.

Miller WR: Why do people change addictive behavior? The 1996 H. David Archibald lecture. Addiction *93:*163, 1998.

Musto DF: *The American Disease. Origins of Narcotic Control*. Oxford University Press, New York, 1987.

Nesse RM, Berridge KC: Psychoactive drug use in evolutionary perspective. Science *278:*63, 1997.

Nestler EJ, Aghajanian GK: Molecular and cellular basis of addiction. Science *278:*58, 1997.

Nyman DJ, Cocores J: Coaddiction: Treatment of the family member. In *Comprehensive Handbook of Drug and Alcohol Addiction*, NS Miller, editor. Marcel Dekker, New York, 1991.

*Prescott CA, Kendler KS: Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry *156:*34, 1999.

Prochaska JO, DiClemente CC, Norcross JC: In search of how people change. Applications to addictive behaviours. Am Psychol *47:*1102, 1992.

Project MATCH Research Group: Matching alcoholism treatment to client heterogeneity: Project MATCH posttreatment drinking outcomes. J Stud Alcohol *58:*2, 1997.

*Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK: Comorbidity of mental disorders with alcohol and other drug abuse. JAMA *264:*2511, 1990.

Robinson TW, Berridge KC: The neural basis of drug craving: An incentive-sensitization theory of addiction. Brain Res Rev *18:*241, 1993.

Rounsaville BJ, Bryant K, Babor R, Kranzler H, Kadden R: Cross system agreement for substance use disorders: DSM-III-R, DSM-IV and ICD-10. Addiction *88:*337, 1993.

Schuckit MA, Smith TL: An 8-year follow-up of 450 sons of alcoholic and control subjects. Arch Gen Psychiatry *53:*202, 1996.

Schuckit MA, Tipp JE, Bucholz KK, Nurnberger JI Jr, Hesselbrock VM, Crowe RR, Kramer J: The life-time rates of three major mood disorders and four major anxiety disorders in alcoholics and controls. Addiction *92:*1289, 1997.

Schulteis G, Gold LH, Koob GF: Preclinical behavioral models for addressing unmet needs in opiate addiction. Semin Neurosci *9:*94, 1997.

Self DW, Nestler EJ: Relapse to drug-seeking: Neural and molecular mechanisms. Drug Alcohol Depend *51:*49, 1998.

Simpson DD, Joe GW, Brown BS: Treatment retention and follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). Psychol Addict Behav *11:*294, 1997.

Substance Abuse and Mental Health Services Administration Office of Applied Studies: *Preliminary Results from the 1996 National Household Survey on Drug Abuse*. National Household Survey on Drug Abuse Series: H-3. DHHS publ no. (SMA) 97-3149. SAMHSA, Office of Applied Studies, Rockville, MD, 1997.

True WR, Heath AC, Scherrer JF, Waterman B, Goldberg J, Lin N, Eisen SA, Lyons MJ, Tsuang MT: Genetic and environmental contributions to smoking. Addiction *92:*1277, 1997.

Tsai G, Gastfriend DR, Coyle JT: The glutamatergic basis of human alcoholism. Am J Psychiatry *152:*332, 1995.

Tsai GE, Ragan P, Chang R, Chen S, Linnoila MI, Coyle JT: Increased glutamatergic neurotransmission and oxidative stress after alcohol withdrawal. Am J Psychiatry *155:*726, 1998.

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L: Co-occurrence of abuse of different drugs in men. Arch Gen Psychiatry *55:*967, 1998.

*Uhl GR: Molecular genetics of substance abuse vulnerability: A current approach. Neuropsychopharmacology *20:*1, 1999.

Wise RA: Drug-activation of brain reward pathways. Drug Alcohol Depend *51:*13, 1998.

11.2 ALCOHOL-RELATED DISORDERS

MARC A. SCHUCKIT, M.D.

An understanding of the effects of alcohol and the clinical importance of alcohol-related disorders is essential for the practice of psychiatry. Alcohol intoxication is capable of causing irritability, violent behavior, feelings of depression, and in rare instances, hallucinations and delusions. Longer-term, escalating levels of consumption of alcohol can produce tolerance as well as such intense adaptation of the body that cessation of use can precipitate a withdrawal syndrome usually marked by insomnia, evidence of hyperactivity of the autonomic nervous system, and feelings of anxiety. Therefore, the adequate evaluation of life problems and psychiatric symptoms in a patient requires that the clinician consider the possibility that the clinical situation reflects the effects of alcohol.

The clinical importance of the relation between alcohol use patterns and psychiatric symptoms is amplified by the high prevalence of alcohol intake and related problems in the general population. Ninety percent of people in most Western societies consume alcohol at some time during their lives, and perhaps 30 percent or more of drinkers develop temporary alcohol-related life problems. Severe, repetitive alcohol-related life impairment (i.e., alcohol dependence) is observed at some time during the lives of approximately 10 percent of men and 3 to 5 percent of women, with an additional 5 to 10 percent of each sex developing persistent but less intense alcohol-related life problems that are diagnosed as abuse. Because high levels of alcohol intake can cause diverse medical and psychiatric problems, it has been estimated that 20 to 35 percent of people seeking help from a health care provider have alcohol abuse or dependence. Thus, alcohol-related problems are very common in society.

Despite the diverse nature of the symptomatology and the high prevalence of alcohol-related disorders, the alcoholic is rarely easy to identify. Alcohol abuse and dependence are seen in both genders, all races and ethnic groups, and in all socioeconomic strata. The alcoholic individual usually has a job, functions moderately well in a family setting, and often appears in the psychiatrist's office with general complaints that rarely point directly toward the alcoholism. The clinician establishes the contribution of alcohol to the clinical complaints by avoiding the false stereotype of the alcoholic as a nonfunctional street person, recognizing the high prevalence of this disorder, and considering the diagnosis in all psychiatric patients.

DEFINITION AND COMPARATIVE NOSOLOGY

Alcohol Use Disorder In all diagnostic systems, the definition of alcoholism indicates evidence of repeated impairments from alcohol in multiple areas of life functioning, despite which the person returns to drinking. These basic elements were present in the third edition (1980) and the revised third edition (1987) of the American Psychiatric Association's *Diagnostic and Statistical Manual* (DSM-III, DSM-III-R) and have continued into the fourth edition (DSM-IV) published in 1994. In the most recent manual, dependence is diagnosed as the presence of at least three of seven major areas of life impairment related to alcohol occurring repeatedly and clustering together in the same 12-month period. These difficulties include tolerance, evidence of a withdrawal syndrome when the drug is discontinued or intake is decreased, potential interference with life functioning associated with spending a great deal of time using the substance, and returning to use despite evidence of physical or psychological problems. It is the syndrome of dependence for which the best data are available regarding the usual clinical course of problems, appropriateness of treatment, and potential importance of genetic factors.

DSM-III-R and DSM-IV do not require evidence of tolerance or withdrawal for the definition of dependence. Recognizing the large change in the concept of “alcoholism” associated with the absence of the centrality of these physiological components of the disorder, DSM-IV allows clinicians to subtype dependence into those syndromes with evidence of a physiological component (i.e., tolerance or withdrawal) and those without these phenomena. Some recent data support the clinical importance of withdrawal, and it is hoped that the request for subtyping will facilitate the accrual of data to determine if it is appropriate to continue to emphasize these physiological conditions in DSM-V.

Thus, all patients with a possible alcohol use disorder should first be evaluated for the presence of alcohol dependence. For those who do not meet the criteria for this disorder, however, there is a second potential syndrome to consider, *abuse*. Here an individual who is not dependent on alcohol demonstrates repeated problems within any 12-month period in any one or more of four potential areas of difficulties. These include repeated legal, interpersonal, social, or occupational impairments related to alcohol, as well as use of alcohol in physically hazardous situations. DSM-IV reformulated the concept of abuse in order to identify criteria that were independent of those noted for dependence. However, it is likely that if they continue to drink, the majority of individuals with abuse will go on to develop dependence in the future.

A similar definition of dependence occurs in the tenth revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). Here, however, the threshold for diagnosis is any three of six (rather than seven) items. The items in ICD-10 dependence include all the concepts in DSM-IV, although they are expressed and numbered differently and some concepts are combined into one criterion. ICD-10 also lists a second and less intense alcohol use disorder, known as *harmful use*. The definition of this second syndrome is quite different from DSM-IV because the ICD-10 approach is based on evidence of repeated interference with psychological and physical health functioning, and does not include social impairment, legal problems, or use in physically hazardous situations. Thus, although the definitions of dependence are quite similar in the United States and ICD-10 systems, there is less obvious overlap between *abuse* and *harmful use*.

Attempts have been made to further divide alcohol-dependence into clinically meaningful subgroups. One such approach is demonstrated in DSM-IV where persons with dependence syndromes are to receive an additional notation of whether a physiological component (i.e., tolerance, withdrawal, or both) is present. Other authors have called for the recognition of a more severe early-onset alcohol dependence syndrome, often accompanied by criminality and dependence on other drugs, which has been labeled as *type II* or *type B alcoholism*. These latter approaches are consistent with the recognition that an earlier-onset alcohol dependence syndrome, like most medical and psychiatric disorders, is likely to have a more severe course, but it appears as if some of the prognostic significance of type II or B alcoholism rests with an elevated risk for a concomitant antisocial personality disorder in the early-onset group.

Severity and Remission The DSM-IV definition of dependence also attempted to better clarify the concepts of severity and remission. Regarding the former, no reliable criteria could be developed, and the manual offers the clinician the possibility of incorporating the relatively imprecise divisions of *mild* (with few symptoms), *moderate* (with functional impairment intermediate between mild and severe), and *severe* (with many symptoms); ICD-10 has no formal notation of severity.

Remission is a more complex phenomenon, and the definition was created from data gathered in a field trial that was incorporated into the DSM-IV process. The diagnostic criteria distinguish between the high-risk period in the first 12 months of recovery and at later time points and ask the clinician to specify whether the patient is totally free of substance-related problems. The criteria also consider whether or not the individual is living in a controlled environment such as prison or a hospital; ICD-10 makes similar but not identical distinctions.

EPIDEMIOLOGY

Psychiatrists need to be concerned about alcoholism because this condition is common, and the usual alcoholic person resembles most other patients (i.e., does not fit a stereotype). Thus, it is always important to evaluate the drinking patterns of all patients because alcohol can interact with medications and intensify pre-existing major psychiatric disorders. Clinicians also need to recognize that a high proportion of their patients have temporary but potentially important alcohol-induced

psychiatric syndromes.

Prevalence of Drinking At some time during their lives, 90 percent of the population in the United States drinks, with most persons beginning their alcohol intake in the early to mid-teens years (Table 11.2-1). At any time two out of three men are current drinkers, with a ratio of persisting alcohol intake of approximately 1.3 men to 1 woman. A *current drinker* is defined most commonly as anyone who has used alcohol during the preceding 1 to 3 months, and is differentiated from persons with alcohol problems. The age of highest prevalence of drinking and of greatest alcohol intake is from the middle or late teens years to the mid 20s.

Condition	Population (%)
Ever had a drink	90
Current drinker	60–70
Temporary problems	40+
Abuse*	Male: 10+ Female: 5+
Dependence*	Male: 10 Female: 3–5

* 20–30 percent of psychiatric patients.

Table 11.2-1 Alcohol Epidemiology

Different groups in the United States have different proportions of drinkers. Generally, those who have high education and high socioeconomic status have the highest proportion who are current drinkers. Among religious groups, Jews have the highest proportion who consume alcohol, but the lowest number of persons with alcohol problems. Conservative Protestants and Catholics use alcohol less frequently than liberal Protestants and Catholics. Other groups, such as the Irish, have higher rates of severe alcohol problems, but they also have significantly higher rates of abstention. High rates of alcohol problems are also found among Native Americans and Eskimos.

In the United States in the early 1990s the average person over the age of 14 years consumed 2.31 gallons of absolute alcohol a year. This amount sounds substantial, but it is considerably less than the over 5 gallons of absolute ethanol consumed each year at the time of the American Revolution. The current figure also represents a significant decrease from the amounts consumed during the mid 1970s, and the 2.7 gallons per capita in 1981.

Alcohol Problems Because a high proportion of persons are drinkers, especially in their middle teens to mid-20s, and because the per-capita consumption of alcohol is high, it is not surprising that a large proportion of persons have alcohol-related problems sometime in their lives. A recent 10-year follow-up study of almost 500 men evaluated at age 33 found that during the preceding decade between one quarter and one third had alcohol-related blackouts, approximately one third admitted to driving after consuming enough alcohol to be impaired, and 20 percent reported missing school or work because of either a hangover or a desire to party with alcohol rather than work.

As common as these problems are and as much as they contribute to lost work time and to physical morbidity and mortality, most people appear to mature out of alcohol problems with the passage of time. Thus, people probably experience fewer alcohol-related difficulties during their 30s than during their 20s, and even fewer difficulties in their 40s and 50s. Both per-capita consumption and the proportion of persons with problems related to alcohol appear to decrease with increasing age beyond middle adulthood.

Alcohol Abuse or Dependence The lifetime risk for alcohol dependence is approximately 10 percent for men and 3 to 5 percent for women. The rate of alcohol abuse and dependence combined may be as high as 20 percent for men and 10 percent for women. Those figures translate to perhaps a total of 200,000 deaths a year in the United States from accidents (perhaps 25,000 persons a year alone), suicide, cancer, and heart disease—the leading causes of death among alcoholic men and women. Cirrhosis is also found at increased rates; 15 percent of alcoholic persons meet the criteria for cirrhosis. Because alcoholism is associated with numerous medical and psychiatric problems, alcoholic persons are overrepresented in psychiatric settings, where they make up one quarter to one third of the usual patient load, even in facilities that serve the affluent.

The age of peak onset of alcohol problems severe enough to lead to a diagnosis of alcohol dependence is probably in the middle 20s to approximately age 40. Despite multiple difficulties in social relationships, families, and jobs, high functioning in some areas is likely to remain. Thus, the stereotypical alcoholic person who is a homeless bum is very much the exception rather than the rule, representing only 5 percent of all persons with severe, recurring alcohol-related difficulties.

Age-related differences are found in the pattern of alcohol-related problems. As is true with almost all psychiatric and many medical disorders, the earlier the onset of alcoholism, the greater the chance that the disorder is severe and that another psychiatric disorder pre-existed. Therefore, when alcohol dependence is noted in a teenager, the person probably has another problem, usually conduct disorder (i.e., early antisocial personality disorder). In that instance, the alcohol-related problems are likely to be associated with severe drug difficulties and antisocial problems in school and with family or peers that occurred before the onset of alcohol dependence. At the other extreme, although most alcoholic persons have their problems early in life, perhaps 10 percent or so have an onset of recurring difficulties after the age of 55. The late onset of the disorder tends to be associated with less severe social difficulties and more subtle signs and symptoms, but a greater likelihood of associated medical problems than among younger alcoholic persons.

Comorbid Conditions The alcohol-related disorders are highly prevalent conditions, and psychiatric symptoms are common during intoxication and withdrawal. Therefore, if diagnoses were to be used to indicate prognosis and optimal treatment, an algorithm had to be developed to help the clinician disentangle temporary substance-related psychopathology from independent psychiatric syndromes.

Alcohol is an organic agent, and most diagnostic algorithms formulated from the time of Emil Kraepelin through DSM-IV have recognized the potential dangers of labeling psychiatric disorders when the symptoms develop during a condition strongly influenced by an organic cause. This hierarchical approach has many parallels in medicine. For example, pneumonias that develop *de novo* are recognized as having a cause and a long-term prognosis different from similar clinical conditions superimposed on congestive heart failure, a bronchus blocked by a carcinoma, or during the course of a severe immune deficiency. The symptoms of pneumonia must be recognized, but the treatment and the prognostic implications of the condition are quite different in the diverse situations. Similarly, in psychiatry depressive episodes that are observed during hypothyroid states; tremor and symptoms of anxiety seen in hyperthyroidism; and psychotic symptoms observed in connection with a brain tumor must be recognized and are likely to require intervention. However, in each instance the prognostic and treatment implications of the syndrome are quite different from major depressive episodes, panic disorder, or schizophrenia that developed in the absence of those major preexisting disorders.

This is also true of the alcohol-related disorders. All the diagnostic manuals since DSM-III have warned the clinician that psychiatric syndromes developing only during intoxication or withdrawal from substances do not necessarily indicate an independent psychiatric disorder. Recognizing that those caveats are easy to overlook in clinical pratice, DSM-IV has added an overall statement about the inadvisability of labeling independent psychiatric disorders based on symptoms observed only during alcohol intoxication or within 4 weeks of abstinence from alcohol. Depressions, panic attacks, and psychotic thought processes occurring in the context of alcohol problems usually improve rapidly and then disappear. Thus, they do not usually carry the same prognostic implications as actual major depressive episodes, panic disorders, and schizophrenia. Similarly, DSM-IV reminds clinicians that such symptoms of mood, anxiety, or psychotic disorders that are documented before severe life problems from alcohol or that remain beyond 4 weeks of total abstinence should be carefully evaluated as possible indicators of true comorbidity with the occurrence of two or more independent psychiatric syndromes. These issues are discussed in greater detail in sections on alcohol-induced mood, anxiety, and psychotic disorders.

Finally, three psychiatric syndromes—bipolar I disorder, schizophrenia, and antisocial personality disorder—carry well established heightened risks for subsequent alcohol-related disorders. Individuals with panic disorder or generalized social phobia might also carry a small but statistically significant risks for alcohol abuse or dependence.

PHARMACOLOGY

Including a discussion of pharmacology in a section on alcohol-related disorders is not meant to imply that a disorder can be defined solely through the use of a substance. Most drinkers do not have serious problems related to alcohol and only a minority of drinkers have difficulties severe and pervasive enough to be labeled abuse or dependence. However, all alcoholic persons have a problem with that potent substance. As a result, clinicians cannot understand the disease or its syndromes without knowing something about alcohol itself.

Ethanol (beverage alcohol) is a simple molecule that is well absorbed through the mucosal lining of the digestive tract in the mouth, the esophagus, and the stomach. The most prominent area of absorption, however, is in the proximal small intestine, which is also the site of absorption of many of the B vitamins. Ethanol rapidly enters the bloodstream and, as a result of its high solubility in water, is distributed to almost every body system. As a consequence of its modest fat solubility, alcohol is likely to have effects on body membranes rich in fat, including neurons.

Wine, beer, and such distilled spirits as whiskey, gin, and vodka differ in their content of congeners. *Congeners* are responsible for much of the characteristic taste of the beverage, and consist of combinations of methanol, butanol, aldehydes, phenols, tannins, lead, cobalt, iron, and other substances. Under certain circumstances congeners can have physiological effects, but their potency pales in comparison with the potency of alcohol.

A drink of an alcoholic beverage is usually defined as containing 10 to 12 grams of ethanol. In round figures this is the amount of alcohol contained in approximately 12 ounces of beer (which in the United States has approximately 3.6 percent ethanol), 4 ounces of table wine (containing about 12 percent ethanol), and between 1 and 1.5 ounces of 80-proof spirits (containing 40 percent ethanol). For the average 70-kg (155-pound) person who has an average amount of body fat, one drink is likely to raise the blood alcohol level by approximately 15 to 20 mg/dL. The body subsequently metabolizes and excretes approximately one drink an hour. The rate of absorption of alcohol from the digestive tract is likely to be faster on an empty stomach than after a full meal, especially one rich in fats and carbohydrates.

After absorption into the bloodstream from the small intestine, between 2 and 10 percent of the alcohol is then excreted unchanged from the lungs or the kidneys or through sweat; the majority is metabolized in the liver. Liver metabolism occurs mostly through four pathways, with each resulting in the production of acetaldehyde. Most of the process occurs through the actions of alcohol dehydrogenase (ADH) in the cytosol of hepatic cells. Especially at high blood alcohol levels, some of the alcohol is also broken down in the microsomes of the smooth endoplasmic reticulum (the microsomal ethanol oxidizing system [MEOS] system). The ADH process is the usual rate-limiting metabolic step, occurring relatively slowly because of the liver's need to handle the produced hydrogen ions through use of a cofactor that is in relatively short supply, nicotinamide adenine dinucleotide (NAD).

The acetaldehyde produced by ADH and MEOS is then destroyed by the enzyme aldehyde dehydrogenase (ALDH) in both the liver cell cytosol and mitochondria. This step occurs rapidly, with the result that the average person does not have substantial levels of this substance; this is fortunate because at high levels acetaldehyde can produce histamine release, which through a variety of mechanisms contributes to falling blood pressure, nausea, and vomiting.

The ALDH isoenzyme pattern of an individual is related to the risk for developing alcoholism. Approximately 10 percent of Asian (e.g., Japanese, Chinese, Korean) men and women fully lack the low k_m ALDH form responsible for metabolizing low blood concentrations of acetaldehyde—that is they are homozygotes for the inactive form of this specific isoenzyme. An additional approximately 40 percent of Asian men and women are heterozygotes, being able to produce low concentrations of this most efficient ALDH form. In homozygote individuals, even one half of a standard drink is likely to cause a severe reaction, resulting in intense facial flushing, rapid heartbeat, nausea, and vomiting. Consequently, most such people rarely drink, and have an associated alcoholism risk that is close to zero. Individuals who are heterozygotes regarding this low k_m ALDH form are more likely to drink but demonstrate relatively high levels of sensitivity to alcohol, and are likely to consume relatively low amounts of alcohol compared to other individuals. Thus, the genetically controlled ALDH isoenzyme pattern is an important factor in the alcoholism risk, although one that applies only to Asians because the absence of the low k_m ALDH is not found in other racial groups.

ALDH activity is also important in another context. A variety of medications, such as disulfiram (Antabuse), inhibit the functioning of this low k_m ALDH form. Thus, when alcoholic individuals regularly take disulfiram, even relatively small amounts of alcohol are likely to produce facial flushing, a rapid heartbeat, and nausea within 20 minutes or so of having consumed the alcoholic beverage. Such drugs, known as alcohol-sensitizing agents, have been considered for the treatment of alcohol dependence.

Neuropharmacology All substances of abuse share the ability to produce changes in feeling states and subsequently increase the likelihood that a person will have a psychological drive to continue to take the substance despite potentially severe adverse consequences (psychological dependence). That effect is distinct from the physical dependence that produces the withdrawal or abstinence syndrome that characterizes drugs like alcohol. However, the 300 or so diverse psychoactive drugs differ in many important ways. For example, only a few produce physiological tolerance and clinically relevant levels of withdrawal symptoms when the substance is discontinued. Some drugs markedly increase the chances that a person will have temporary psychoses or depressions; other drugs do not. Some are likely to be lethal in overdose; others appear to be relatively safe at high levels. Clinicians therefore are presented with a daunting challenge if they attempt to memorize all the attributes for each of the hundreds of psychoactive substances.

A useful shortcut is to place drugs of abuse into categories on the basis of their most prominent effects at the usual doses at which they are taken. In this scheme, substances that have as their most prominent usual effects the production of somnolence and decreased neuronal activity but that are not powerful in attenuating pain are labeled as depressants or sometimes as sedative-anxiolytics. They include alcohol, all the benzodiazepines, all the barbiturates, and the carbamate antianxiety drugs, such as meprobamate (Miltown). These substances produce a similar profile of symptoms during intoxication, are potentially lethal in overdose (especially when multiple depressant drugs are taken at the same time), are cross-tolerant, are physically addicting, and produce similar withdrawal syndromes.

The behavioral and physiological changes observed with any substance differ with the dose, the patient's prior history of exposure to the drug, and clinical conditions, including physiological disorders and the patient's state of fatigue. With a drug like alcohol, the effects also change over time after intake, with more pronounced symptoms observed while the blood alcohol levels are rising then when the blood alcohol levels are falling, a phenomenon called *acute tolerance* or the *Mallenby effect*.

Debate continues about the most important mechanisms of action of alcohol on the brain. One of the problems occurs because the drug has a major effect on most neurochemical systems, demonstrating different effects at different doses and sometimes opposite effects during intoxication and withdrawal. One series of theories on the mechanisms underlying intoxication relates to the effects that alcohol has on the cell membrane: alcohol tends to fluidize or decrease the levels of rigidity of the membrane, with subsequent impairments in the cell's ability to control the influx and the efflux of electrolytes. Other research focuses on changes in dopamine, attempting to tie in the effects of alcohol to the pleasure centers of the limbic system, or on serotonin, a neurotransmitter related to appetitive behaviors. Still other investigators point to the potential importance of the neurochemical compounds that may, at least theoretically, be formed between acetaldehyde and the neurotransmitters serotonin and dopamine, producing alkaloids that have properties resembling opioids. Another set of investigations point out the indirect effects that alcohol can have on the benzodiazepine-receptor complexes in the brain. Finally, alcohol has potent effects on glutamate-gated ionophoric receptors, especially those that bind *N*-methyl-D-aspartate (NMDA). With such a diverse range of effects and the absence of an obvious receptor system reacting specifically to alcohol, many leads are promising but few answers are definitive regarding the most clinically relevant effects of alcohol on the nervous system or the way in which the alterations may relate to abuse or dependence on alcohol.

Tolerance With repeated administration of alcohol, larger and larger doses of the drug are required to produce the desired effect. That phenomenon, called *tolerance*, is the ability to tolerate higher and higher doses of the substance and is the result of at least three processes.

Behavioral tolerance reflects the ability of a person to learn how to perform tasks effectively despite the effects of alcohol. It is a learned behavior and the result of repeated practice. *Pharmacokinetic tolerance* is an adaptation of the metabolizing system to rid the body of alcohol rapidly. After several weeks of daily drinking, the liver produces more ADH than usual and expands the MEOS system, with a resulting increase of up to 30 percent in the rate of breakdown of ethanol. Finally, *pharmacodynamic* or *cellular tolerance* is an adaptation of the nervous system so that it can function, despite high blood alcohol concentrations, by resisting the actions of alcohol on the cell. Thus, persons have been observed to be awake, relatively alert, and relatively coordinated despite blood alcohol concentrations of 250 mg/dL, and some people have been awake at blood alcohol concentrations above 600 mg/dL.

Once tolerance has developed for one of the brain depressants, a person often shows a similar reaction to a second drug of that class (*cross-tolerance*). Therefore, a person who has been drinking heavily, has tolerance for alcohol, and then stops drinking can be expected to require a higher dose of benzodiazepines for sleep

induction. If the person took two depressant drugs at the same time, tolerance is not likely to be observed and the mixing of the two substances can have lethal effects. Just as tolerance requires a period of days or weeks to develop, the phenomenon is likely to disappear within a similar period of time after the intake of any depressant drug has ceased.

Some clinicians and researchers have described a phenomenon of *reverse tolerance* or *sensitization*. This is a complex situation that might involve multiple physiological components. For example, whether alcoholic or not, as persons grow older they have increasing levels of sensitivity to most brain depressants, including alcohol. This is because of several age-related changes in the body, including a decrease in the rate at which alcohol is metabolized in the liver and a relative decrease in body water as a consequence of an increasing percentage of body fat. The result of these changes is that higher blood alcohol levels develop in older people, whose neurons also have an enhanced sensitivity to the effects of alcohol. Even more dramatic examples of increased sensitivity to alcohol are seen after severe brain damage (e.g., the consequence of an auto accident or alcohol-related brain deterioration) and after impairment in any of the major alcohol-metabolizing systems, as occurs in cirrhosis.

Effects on the Body Some data indicate that alcohol use is not always harmful and may under certain circumstances even have some beneficial effects. However, any amount of alcohol is considered harmful to the developing fetus, to recovering alcoholics, to people taking medications that may adversely interact with alcohol, and to individuals with certain medical disorders or psychiatric syndromes (such as major depressive disorder or schizophrenia) that might be intensified by alcohol. For other people a maximum of one or two drinks a day appears to be associated with a decreased risk of cardiovascular disease. This association may be caused by an alcohol-induced decrease in platelet adherence or to an increase in at least one portion of high-density lipoprotein (HDL) cholesterol, although the fraction affected does not appear to be the one most potent in protecting against heart disease. Unfortunately, the intake of more than two drinks a day is likely to increase low-density lipoprotein (LDL) cholesterol and triglycerides and to raise blood pressure, with the overall result of increasing the risk of cardiac disorders. Also, even low levels of alcohol intake may increase the risk for breast cancer.

Central and Peripheral Nervous Systems

BLACKOUT *Blackout* indicates a memory impairment (*anterograde amnesia*) for the period of time when the person was drinking heavily but remained awake. This common difficulty is related to the ability of any brain depressant to interfere with the acquisition of memory at high enough doses. Perhaps 40 percent of teenaged and young adult males have blackouts, and memory loss does not by itself indicate a high likelihood of alcohol abuse or dependence. The blackout, which is temporary and limited to memory problems involving a short period of time, is not a DSM-IV diagnosis, and is distinct from *alcohol-induced persisting amnestic disorder*, formerly known as *Wernicke-Korsakoff syndrome*.

SLEEP IMPAIRMENT Alcohol intoxication can help a person fall asleep more quickly, but it tends to depress rapid eye movement sleep and inhibit stage 4 sleep. It is likely to be associated with frequent alternations between sleep stages, a process sometimes referred to as *sleep fragmentation*.

PERIPHERAL NEUROPATHY A more serious and potentially permanent problem is seen in perhaps 10 percent of alcoholic persons after years of heavy drinking. The deterioration of nerve functioning to the hands and feet, called *peripheral neuropathy*, arises through an apparent combination of vitamin deficiencies and the direct effects of alcohol or its metabolites. The symptoms include numbness of the hands and feet, often bilateral, frequently accompanied by tingling and paresthesias. Although the condition is usually relatively mild and often improves with abstinence, the pain and the numbness can result in a permanent impairment.

CEREBELLAR DEGENERATION Characterized by unsteadiness of gait, problems with standing, and mild nystagmus, cerebellar degeneration is probably caused by a combination of the effects of ethanol and acetaldehyde, along with vitamin deficiencies. Treatment usually consists of total abstinence and vitamin supplementation, although complete recovery is not common.

OTHER EFFECTS ON THE CENTRAL NERVOUS SYSTEM A series of temporary but intense psychiatric symptoms are likely to be observed during alcohol intoxication and withdrawal, including depressed mood, severe anxiety, and psychoses. These symptoms often mimic psychiatric disorders but are likely to disappear within weeks of abstinence. Severe amnestic disorders and dementias may also occur.

Gastrointestinal Problems Second only to the nervous system, the gastrointestinal system is most severely affected by heavy drinking. Probably the most common gastrointestinal problem associated with alcohol intake is an acute and at times severe inflammation of the esophagus or the stomach, with stomach inflammation often accompanied by vomiting and bleeding. If gastritis occurs in the presence of dilated esophageal veins seen with cirrhosis, it can induce potentially lethal bleeding.

The liver and the pancreas are especially vulnerable to alcohol. In the liver, increasing alcohol doses result in the accumulation of fats and proteins in the cells, producing a reversible swelling often described as a *fatty liver*. Inflammation of the liver cells accompanied by a subsequent intense rise in some liver function tests and other signs of alcoholic hepatitis can lead to the deposition of excessive amounts of hyalin and collagen near blood vessels, an early stage of cirrhosis. As damage progresses, the normal flow of blood through the liver is impaired, dilated veins or varices develop from the increased abdominal venous pressure, and fluid seeps from the liver capsule, accumulating in the abdomen as ascites. As liver failure progresses, secondary cognitive impairment can develop as various levels of hepatic encephalopathy.

Perhaps 15 percent of alcoholic persons respond to large doses of alcohol with an inflammation of the pancreas that can present as the abdominal emergency of acute pancreatitis. This can lead to a chronic irreversible condition of pancreatic destruction, with associated signs of insufficiency in sugar metabolism (a form of diabetes) and digestive enzymes.

Cerebrovascular and Cardiovascular Problems Heavy intake of alcohol increases the blood pressure and elevates both LDL cholesterol and triglycerides, thus enhancing the risk for myocardial infarction and thrombosis. At high doses, alcohol is also a striated-muscle toxin with a resulting production of what is usually but not always a reversible deterioration in the heart muscle that manifests itself as beating irregularities and signs of heart failure *(alcoholic cardiomyopathy)*. Similar levels of swelling of muscle cells and subsequent muscle pain can be observed in the skeletal muscles.

Blood-Producing Systems High levels of alcohol intake, often in the range of four to eight drinks a day, decrease the production of white blood cells and impairs the ability of those cells to migrate to sites of infection. Such drinking can also affect the stem cells that produce the red blood components, significantly increasing the average size of the red cell (the mean corpuscular volume [MCV]), and can impair the production and the efficiency of blood platelets.

Cancer High rates of many types of cancer are seen in alcoholic persons, especially cancers of the head, neck, esophagus, and stomach; additional areas of enhanced risk include the liver, the colon, and the lungs. The risks probably reflect alcohol-related immune system suppression, but might also be a result of the direct effects of ethanol on mucosal membranes. The heightened rates of malignant tumors in alcoholic persons remain significant even when the possible effects of smoking and poor nutrition are considered.

Fetal Alcohol Effect Alcohol and acetaldehyde can have deleterious effects on the developing fetus. Both substances cross the placenta with ease, and in high enough doses can produce fetal death and spontaneous abortion. Surviving infants of heavy-drinking mothers can evidence any mixture of the components of a syndrome that in its full-blown form can include severe mental retardation, a small head, a diminished physical size, facial abnormalities (including a flat bridge of the nose, an absent philtrum, and an epicanthal eye fold), an atrial septal heart defect, and syndactyly. None of these problems is reversible; once present, the cognitive defects, physical irregularities, and behavioral problems remain throughout life. Because the exact amount of alcohol required and the most vulnerable periods of pregnancy have not been definitively established all pregnant women are advised to abstain from any use of alcohol.

ETIOLOGY

Many factors affect the decision to drink, the development of temporary alcohol-related difficulties in the teenage years and the 20s, and the development of alcohol dependence. The initiation of alcohol intake probably depends on social, religious, and psychological factors, although the high rate of persons who have tried alcohol at some time during their lives indicates that drinking is an almost ubiquitous phenomenon in most Western societies. However, it is important to remember that the factors that influence the decision to drink or those that contribute to temporary problems might be different from those that contribute to the severe, recurring problems of alcohol dependence.

Psychological Theories A variety of theories relate to the use of alcohol to reduce tension, increase feelings of power, and decrease the effects of psychological

pain. Perhaps the greatest interest has been paid to the observation that persons with alcohol-related problems often report that alcohol decreases their feelings of nervousness and helps them cope with the day-to-day stresses of life. The psychological theories are built in part on the observation among nonalcoholic persons that the intake of alcohol in a tense social setting or after a difficult day can, especially in low doses, be associated with an enhanced feeling of well-being and an improved ease of interactions. However, data indicate that in high doses, especially at falling blood alcohol levels, most measures of muscle tension and psychological feelings indicate that heavy drinking is likely to be associated with increased nervousness and tension. The theories that focus on alcohol's potential to enhance feelings of being powerful and sexually attractive and to decrease the effects of psychological pain are difficult to evaluate definitively.

Psychodynamic Theories Perhaps related to the disinhibiting or anxiety-lowering effects of alcohol, at least at rising blood alcohol concentrations, is the hypothesis that some persons may use alcohol to help them deal with self-punitive harsh superegos as a way of decreasing unconscious stress levels. Also, classical psychoanalytical theory hypothesizes that at least some alcoholic persons have become fixated at the oral stage of development and use alcohol to relieve their frustrations by taking the substance by mouth. Alcoholic persons may also use the drug as part of a need for enhanced feelings of power. However, hypotheses regarding arrested phases of psychosexual development, although heuristically useful, have had little effect on the usual treatment approaches and are not the focus of extensive ongoing research. Similarly, several hypotheses have questioned the potential importance of addictive personality attributes because they may reflect levels of impulsiveness and sensation-seeking behavior. However, careful studies have failed to identify a unique personality profile that is prone to addictions, with the exception of antisocial personality disorder.

Behavioral Theories Expectations about the rewarding effects of drinking and subsequent actual reinforcement after alcohol intake contribute to the decision to drink again after the first experience with alcohol. These issues are important in efforts to modify drinking behaviors in the general population, and they contribute to some important aspects of alcoholic rehabilitation.

Sociocultural Theories Sociocultural theories are often based on observations of social groups that have high and low rates of alcoholism. Theorists hypothesize that ethnic groups such as Jews that introduce children to modest levels of drinking in a family atmosphere and that eschew drunkenness have low rates of alcoholism. Some other groups such as Irish men, with high rates of abstention but a tradition of drinking to the point of drunkenness among drinkers, are thought to have high rates of alcoholism. However, these theories often depend on stereotypes that are frequently erroneous and there are several exceptions to these rules. For example, some theories based on observations of the Irish and the French would have predicted high rates of alcoholism among the Italians, although alcohol problems are not generally observed at a high level in this group.

In the final analysis, social and psychological theories probably have more than heuristic value. They outline factors that contribute to the onset of drinking, the development of temporary alcohol-related life difficulties, and even alcoholism. The problem is how to find a way to gather relatively definitive data to support or refute the theories.

BIOLOGICAL THEORIES

Genetic Theories The best supported of the biological theories of alcoholism centers on genetics (Table 11.2-2). One finding supporting the genetic conclusion is the threefold to fourfold increased risk for severe alcohol problems in *close relatives of alcoholic persons*. The rate of alcohol problems increases with the number of alcoholic relatives, the severity of their illness, and the closeness of their genetic relationship to the person under study. The family investigations do little to separate the importance of genetics and environment, but twin studies take the data a step further. The rate of similarity or concordance for severe alcohol-related problems is significantly higher in identical twins of alcoholic individuals than in fraternal twins in most studies. The adoption-type studies have all revealed a significantly enhanced risk for alcoholism in the offspring of alcoholic parents, even when the children had been separated from their biological parents close to birth and raised without any knowledge of the problems within the biological family. The risk for severe alcohol-related difficulties is not further enhanced by being raised by an alcoholic adoptive family.

Close family members have a fourfold increased risk.
The identical twin of an alcoholic person is at higher risk than is a fraternal twin.
Adopted-away children of alcoholic persons have a fourfold increased risk.

Table 11.2-2 Data Supporting Genetic Influences in Alcoholism

These data not only support the importance of genetic factors in alcoholism, but also highlight the complexity of the phenomenon. The absence of evidence of a single major locus indicates the possibility that a limited number of genes operate with incomplete penetrance, or that a combination of genes is required before the disorder expresses itself (a polygenic mode of inheritance). Making matters even more complex is the likelihood that the disorder is solely an expression of environmental events in some families, and the probability that different genetic factors operate in different families to produce a picture of genetic heterogeneity.

Despite these problems, studies have identified men and women at high future risk of alcoholism, usually defined as having an alcoholic parent. Persons at high and low future risk are then compared on psychological and biological parameters. Some protocols then expose young adults to an alcohol challenge, and several investigations have subsequently followed them over time.

Related Biological Theories Different approaches have highlighted several potential leads to biological factors that may affect the risk of developing alcoholism. Investigations of early-teenage children of alcoholic parents, usually including inner-city families or children of persons with antisocial personality disorder, have shown the potential importance of several neurocognitive test results as possible predictors of a risk for alcoholism. Additional leads have come from electrophysiological evaluations of children of alcoholic persons, including the finding that perhaps one third of the sons of severely alcoholic men may have a decreased amplitude of the positive wave observed 300 milliseconds after a rare but unexpected sensory stimulus, the P3 wave of *event-related potential* (ERP). Other studies have shown a potential decrease in the amount of power in the slow alpha range on the background cortical electroencephalogram (EEG), or relative deficiencies in beta waves.

An additional potential phenotypic marker involves the intensity of a person's reaction to an alcohol challenge. An ongoing study involves 453 sons of alcoholics and controls who were originally evaluated at approximately age 20, at which time they were all drinkers but none was alcohol dependent. At the time of initial evaluation, despite matching the two family history groups on drinking and drug use histories, and despite identical blood alcohol concentrations during the drinking experiment, 40 percent of the sons of alcoholics but less than 10 percent of the controls showed remarkably low levels of response to alcohol. The intensity of the reaction was measured by a combination of subjective feelings of intoxication, changes in motor performance while under the influence of alcohol, as well as alcohol-induced changes in blood hormones and electrophysiological functioning of the brain. An average of 8.2 years later, all 453 subjects were located, and information about functioning during the follow-up period was obtained from the subject and an additional informant for 450 individuals (99.3 percent). The low level of response to alcohol at approximately age 20 was a potent predictor of future alcoholism, with 60 percent of the sons of alcoholics who had a low response developing alcoholism by approximately age 30, whereas the same was true for only 15 percent of the sons of alcoholics who showed high levels of response to the alcohol challenge. Sons of nonalcoholics showed a similar pattern of relationship between the low response to alcohol and the future risk of alcoholism. The data revealed that for this relatively highly functional, often white-collar population the low response to alcohol was responsible for a large proportion of the ability of the family history to predict the risk of alcoholism.

These follow-up data underscore some of the important issues related to genetic influences in this disorder. First, it is unlikely that anyone inherits a predestination for alcoholism. Rather, there are likely to be a variety of different genetically influenced characteristics that interact with the environment to increase the risk for severe

and repetitive alcohol-related life problems. Second, identification of traits that enhance the alcoholism risk, and possibly the genes related to these characteristics, might help to pinpoint the social and cultural factors that interact with a biological predisposition to produce the final level of risk for this disorder. It is possible that more knowledge about environmental and potentially changeable influences might help to develop more precise prevention approaches that can be applied to individuals who carry the highest risk.

DIAGNOSIS AND CLINICAL FEATURES

Alcohol use disorders are probably the most common of the serious diagnosable behavioral or psychiatric disorders, and the diagnosis of alcohol dependence or abuse requires a high index of suspicion for the disorder in any patient. The average man or woman presenting with severe and repetitive alcohol problems is likely to be neatly dressed, to show no signs of severe alcohol withdrawal, to have a job and a family, and to complain of a variety of physical conditions or temporary but potentially severe psychiatric complaints. Thus, the clinician must gather a history of alcohol-related life problems from the patient and, whenever possible, a resource person, and must try to determine whether alcohol has caused or contributed to the psychiatric or physiological syndrome. Table 11.2-3 lists the alcohol-related disorders in DSM-IV, and also presents a comparable listing from ICD-10.

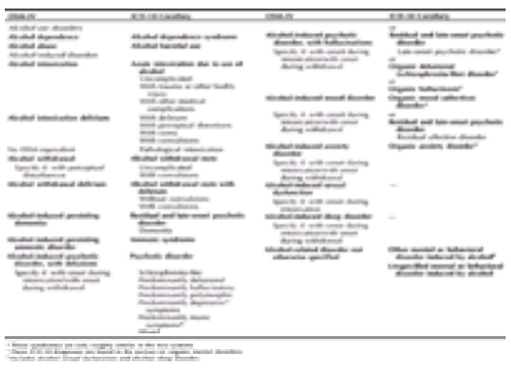

Table 11.2-3 DSM-IV Alcohol-Related Disorders and Corresponding ICD-10 Disorders

For patients presenting with psychiatric symptoms (e.g., anxiety, depression, or psychoses) as well as evidence of alcohol-related problems, the first step is to obtain a careful history from both the patient and a resource person who knows the patient well. Second, in taking the information the clinician must emphasize syndromes that meet diagnostic criteria for major depressive disorder or full anxiety syndromes or other disorders, not just symptoms such as sadness or nervousness. Third, the clinician must establish a time line from birth to the present, noting (1) the approximate age of onset of alcohol problems severe and repetitive enough to justify a diagnosis of alcohol dependence; (2) periods of abstinence of several months or more; and (3) the ages at which the patient met the criteria for any major psychiatric disorders, taking care to emphasize full-blown psychiatric clinical conditions, not isolated symptoms. If a review of the time line reveals no evidence that the additional psychiatric syndromes either antedated the severe alcohol problems or persisted for 4 or more weeks during a period of abstinence, alcoholism is the major disorder. Under these conditions the other psychiatric syndromes are important but temporary conditions that occurred during alcohol intoxication or withdrawal.

Depressive, anxiety, and psychotic syndromes are often seen in people with alcohol-related disorders. However, even if the psychiatric symptoms are intense, they do not indicate a separate psychiatric syndrome when seen only during intoxication or withdrawal (i.e., they can be temporary alcohol-induced conditions). In an effort to encourage clinicians and researchers to consider the entire span of clinical conditions that might be relevant to any syndrome being observed, in DSM-IV all important diagnostic entities related to a specific phenomenon (e.g., depressive disorders, anxiety disorders, psychotic disorders) are now listed within the clinically relevant sections (e.g., the mood disorder section). For the sake of clarity, disorders associated with substances are now labeled as *substance-induced disorders*.

To be termed a substance-induced condition, the syndrome must be clinically meaningful and must resemble the type of disorder described within that DSM-IV section (e.g., a mood disorder). There must be evidence indicating a likelihood that the clinical condition developed during, or within a month of, substance intoxication or withdrawal from a specific substance (such as alcohol) that would be capable of producing a relevant temporary clinical condition (such as a severe mood disturbance). The clinician and researcher are warned that the substance-induced condition should only be diagnosed when the psychiatric symptoms (e.g., depression) are in excess of those usually associated with intoxication or withdrawal. The diagnostic criteria further list the specific substances involved and ask that, if possible, the clinician specify whether the condition had an onset during intoxication or withdrawal. These latter modifiers are important to indicate to the clinician when additional medical and psychiatric treatment might be required. For alcohol-induced mood disorders, diagnoses can also be subtyped regarding the presence or absence of depressive, manic, or mixed features.

DSM-IV offers similar information regarding a substance-induced anxiety disorder, which is listed in the section on anxiety disorders. Again, the condition must be clinically relevant; there must be evidence that a substance, such as alcohol, capable of producing a severe temporary anxiety condition was involved; and a distinction regarding an onset during intoxication or withdrawal is encouraged. The anxiety conditions can be further subdivided regarding the relevance of generalized anxiety symptoms, repetitive panic attacks, obsessive-compulsive symptoms, or phobic symptoms.

The documentation of hallucinations and delusions associated with intoxication or withdrawal from relevant substances is covered in DSM-IV section on psychotic disorders. When the condition is clinically relevant and when evidence exists that a substance (such as alcohol) capable of causing the psychotic symptoms was involved, a diagnosis of a substance-induced psychotic disorder (in this instance alcohol-induced psychotic disorder) can be made. Additional criteria have been developed for alcohol-induced sexual dysfunction (see Table 19.1a-20) and alcohol-induced sleep disorders (see Table 21-18).

Alcohol Dependence DSM-IV provides general criteria for all substance use disorders (see Table 11.1-3). Dependence concerns a history of a broad array of problems, including compulsive intake of alcohol, an increasingly important place in life occupied by the substance, and possibly evidence of physical withdrawal symptoms. Dependence criteria also concern life impairment related to the substance.

Physical dependence or is a phenomenon that appears to be related to tolerance. As the body changes to resist the effects of alcohol, it is likely to reach a condition in which it cannot function optimally unless the brain depressant is present. That condition takes days or weeks to develop.

DSM-IV substance dependence criteria include seven items that are subsets of the nine originally listed in DSM-III-R. These seven items are similar to the ICD-10 dependence syndrome criteria, although ICD-10 deals more directly with evidence of a compulsion to use (see Table 11.1-4). In addition, while maintaining the broad concept of dependence that appeared in DSM-III-R, DSM-IV asks the clinician to use the two items that deal with tolerance or withdrawal to further classify dependent persons into those with and those without evidence of physiological symptoms. This division allows clinicians and researchers to determine the treatment and the prognostic implications of tolerance and withdrawal. The framers of the fifth edition of DSM (DSM-IV) in the twenty-first century will then have data to help them decide whether to return to the emphasis on physiological symptoms that characterized earlier manuals (e.g., DSM-III).

> A 23-year-old graduate student in physics was referred for evaluation by her adviser who was concerned about tardiness at work and recent problems with a lack of clarity of thinking. As he discussed these difficulties with her, the student admitted to being concerned about her drinking, which had been dramatically emphasized to her in a recent intervention carried out by her father and mother. She related that for the last 5 years or so she has regularly consumed 1.5 to two or three bottles of wine each evening (approximately 9 to 18 drinks). In the last 2 years she has noted a marked increase in the need for more alcohol to get the same effects, reported giving up activities with her family in order to drink, spending a great deal of her time drinking, and driving long distances to obtain alcohol. She has repeatedly tried to cut down, often setting a limit of two drinks in an evening, but regularly going on to nine or more standard drinks before stopping. Despite her high general level of functioning, her active participation in a graduate education program, and close interpersonal relationships, this history of alcohol dependence is fairly typical among alcohol-dependent individuals.

Alcohol Abuse The DSM-IV diagnostic criteria for abuse focus on the impairment of social, legal, interpersonal, and occupational functioning in a person who is not alcohol dependent (see Table 11.1-8). ICD-10 presents a diagnosis of *harmful use* that is only roughly similar to DSM-IV. The major difference is the restriction of the international system to issues of physical or psychological problems. The DSM-IV criteria were developed through careful comparisons of DSM-III and DSM-III-R

concepts, and were refined through a series of reanalysis of existing data sets and a large field trial of the criteria for abuse and dependence. The field trial involved comparisons of clinical coverage and demographic correlates of DSM-III-R and possible DSM-IV criteria, as applied to more than 1000 men and women from diverse groups and representing persons from the general population as well as those with diverse substance problems in six centers in the United States and four locations in other countries.

Alcohol Intoxication The DSM-IV diagnostic criteria for alcohol intoxication are based on evidence of recent ingestion of ethanol, maladaptive behavior, and at least one of six possible physiological correlates of intoxication (Table 11.2-4). The ICD criteria for acute alcohol intoxication are generally similar to the DSM-IV criteria. There are seven physiological signs of intoxication listed in the international system, some of which, such as conjunctival injection, are not listed in the DSM-IV criteria.

A. Recent ingestion of alcohol.
B. Clinically significant maladaptive behavior or psychological changes (e.g., inappropriate sexual or aggressive behavior, mood lability, impaired judgment, impaired social or occupational functioning) that developed during, or shortly after, alcohol ingestion.
C. One (or more) of the following signs, developing during, or shortly after, alcohol use:
(1) slurred speech
(2) incoordination
(3) unsteady gait
(4) nystagmus
(5) impairment in attention or memory
(6) stupor or coma
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Table 11.2-4 DSM-IV Diagnostic Criteria for Alcohol Intoxication

As a conservative approach to identifying blood levels that are likely to have major effects on driving abilities in the majority of people, the legal definition of intoxication in most states in the United States requires a blood concentration of 80 or 100 mg ethanol per dL of blood (mg/dL), which is the same as 0.08 to 0.10 grams per deciliter (g/dL). For most people, a rough estimate of the levels of impairment likely to be seen at various blood alcohol concentrations can be outlined. Evidence of behavioral changes, a slowing in motor performance, and a decrease in the ability to think clearly occurs at doses as low as 20 to 30 mg/dL, as shown in Table 11.2-5. Blood concentrations between 100 and 200 mg/dL are likely to produce a progression of the impairment in coordination and judgment to severe problems with coordination (ataxia), increasing lability of mood, and progressively greater levels of cognitive deterioration. Anyone who does not show significant levels of impairment in motor and mental performance at about 150 mg/dL probably has significant pharmacodynamic tolerance. In that range most persons without significant tolerance also experience relatively severe nausea and vomiting. With blood alcohol concentrations in the 200 to 300 mg/dL range, the slurring of speech is likely to become more intense, and memory impairment *(anterograde amnesia* or *alcoholic blackouts)* becomes pronounced. Further increases in blood alcohol concentration result in the first level of anesthesia, and the nontolerant person who reaches 400 mg/dL or higher risks respiratory failure, coma, and death.

Level	Likely Impairment
20–30 mg/dL	Slowed motor performance and decreased thinking ability
30–80 mg/dL	Increases in motor and cognitive problems
80–200 mg/dL	Increases in incoordination and judgment errors Mood lability Deterioration in cognition
200–300 mg/dL	Nystagmus, marked slurring of speech, and alcoholic blackouts
>300 mg/dL	Impaired vital signs and possible death

Table 11.2-5 Impairment Likely to Be Seen at Different Blood Alcohol Concentrations

Prior editions of DSM also described *alcohol idiosyncratic intoxication*. Currently termed *pathological alcohol intoxication* in ICD-10, it was characterized by extreme aggressive behavior occurring within minutes of ingesting relatively small amounts of alcohol, such as two drinks. According to DSM-III-R the person was usually amnestic for the episode, and the aggressive behavior was atypical of the person's usual sober comportment. However, a literature review before the publication of DSM-IV revealed little convincing evidence that such a disorder really exists, and it was deleted from the manual. The status of the condition in ICD-10 is being examined and must be regarded as tentative.

Alcohol Withdrawal In persons who have been drinking heavily over a prolonged period of time, a rapid decrease in the amount of alcohol in the body might produce a variety of physical symptoms. This *withdrawal* or *abstinence syndrome* is characterized by a group of symptoms that are the opposite of what was initially experienced with intoxication. Therefore, after a person is physically dependent on alcohol, abstinence is likely to be accompanied by a coarse tremor of the hands, insomnia, anxiety, and increased blood pressure, heart rate, body temperature, and respiratory rate—a condition labeled in DSM-IV as *alcohol withdrawal* and described in Table 11.2-6. In ICD-10 the criteria for alcohol withdrawal are similar to those listed in DSM-IV, although there are differences in the items listed as physiological correlates of withdrawal as well as in the number of signs required (i.e., three) to make a diagnosis. The DSM-IV criteria for alcohol withdrawal also require that the symptoms must cause clinically significant distress or impairment in an important area of functioning.

A. Cessation of (or reduction in) alcohol use that has been heavy and prolonged.
B. Two (or more) of the following, developing within several hours to a few days after criterion A:
(1) autonomic hyperactivity (e.g., sweating or pulse rate greater than 100)
(2) increased hand tremor
(3) insomnia
(4) nausea or vomiting
(5) transient visual, tactile, or auditory hallucinations or illusions
(6) psychomotor agitation
(7) anxiety
(8) grand mal seizures
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and not better accounted for by another mental disorder.

Specify if:
With perceptual disturbances

Table 11.2-6 DSM-IV Diagnostic Criteria for Alcohol Withdrawal

Although 95 percent or more of withdrawals are limited to these mild or moderate symptoms, for 3 to 5 percent the symptoms include convulsions or delirium. Withdrawal phenomena are likely to begin within 8 hours of abstinence, reach a peak intensity on the second or third day, and markedly diminish by the fourth or fifth day. The symptoms can persist in a mild form for 3 to 6 months or more as part of a protracted withdrawal syndrome.

This 23-year-old graduate student in physics with alcohol dependence was unable to establish even 24 hours of sobriety as an outpatient. Therefore, reflecting her continued drinking, she was referred for inpatient care. Following approximately 10 hours of abstinence, and with a documented blood alcohol concentration of 0 mg/dL, she was noted to be mildly diaphoretic, with a respiratory rate of 25 breaths per minute, blood pressure of 130/90, a mild bilateral tremor of the hands, and a pulse rate of 85 beats per minute. She had a history of jogging 2 to 5 miles a day and these figures represented moderate elevations in her usual vital signs. Treated with multiple vitamins, good nutrition, oral fluids, and benzodiazepines, the symptoms rapidly improved, and her vital signs were close to normal by day four of abstinence.

Alcohol Intoxication and Alcohol Withdrawal Delirium For the small proportion of intoxications and withdrawals that are accompanied by severe cognitive symptoms, both DSM-IV and ICD-10 contain the diagnoses of *alcohol intoxication delirium* and *alcohol withdrawal delirium* (see Table 10-22 and Table 10-23).

When the symptoms of withdrawal are accompanied by a state of severe agitated confusion or delirium, sometimes associated with tactile or visual hallucinations, the diagnosis of *alcohol withdrawal delirium* (also called *delirium tremens* [*DTs*]) can be made. During withdrawal, some alcoholic persons show one or several grand mal convulsions, sometimes called *rum fits*.

A 73-year-old professor emeritus at a university was thought to be in good health when he entered the hospital for an elective hernia repair. Perhaps reflecting his status in the community, the relatively brief history contained no detailed notes of his drinking pattern, and made no mention that his g-glutamyltransferase (GGT) values of 55 units per liter, along with the mean corpuscular volume (MCV) of 93.5 cubic microns. Eight hours postsurgery the nursing staff noted a sharp rise in the pulse rate to 110, an increase in blood pressure to 150/100, prominent diaphoresis, and a tremor to both hands, following which the patient demonstrated a brief but intense grandmal convulsion. He awoke after approximately 15 minutes extremely agitated and disoriented to time, place, and person. A reevaluation of the history and a separate interview with the wife documented alcohol dependence with a consumption of at least six to nine standard drinks per night. Over the following 4 days, the patient's autonomic nervous system dysfunction decreased as his cognitive impairment disappeared. His condition is classified as *alcohol withdrawal delirium* in DSM-IV.

Alcohol-Induced Persisting Amnestic Disorder One of the most intensely studied alcohol-related central nervous system (CNS) syndromes is the relatively rare DSM-IV diagnosis of *alcohol-induced persisting amnestic disorder* (see Table 10-39), which is the result of a relatively severe deficiency in the B vitamin thiamine. Similar criteria are offered in ICD-10 as an amnesic syndrome. Some persons are at higher risk for this syndrome than are others because of a genetically influenced transketolase deficiency. The condition has been historically subdivided into (1) *Wernicke's encephalopathy*, with prominent ataxia and palsy of the sixth cranial nerve, a condition that tends to reverse fairly rapidly with vitamin supplementation, and (2) *Korsakoff's syndrome*, which is permanent in at least a partial form in perhaps 50 to 70 percent of the persons affected. Korsakoff's syndrome is characterized by a pronounced anterograde and retrograde amnesia and potential impairment in visuospatial, abstract, and other types of learning. In most cases the level of recent memory is out of proportion to the global level of cognitive impairment. The 25 percent or so of patients with Korsakoff's syndrome who are likely to recover fully and the 50 percent or so who recover partially appear to respond to 50 to 100 mg of oral thiamine a day, usually administered for many months.

Alcohol-Induced Persisting Dementia An alcohol-related CNS diagnosis relevant to psychiatry is the relatively heterogeneous and poorly studied long-term cognitive problem that can develop in the course of alcoholism, namely *alcohol-induced persisting dementia* (see Table 10-31). Similar syndromes are described in ICD-10 as *residual and late-onset psychotic disorder* or as *other persisting cognitive impairment*. Global decreases in intellectual functioning, cognitive abilities, and memory are observed. Recent memory difficulties are consistent with global cognitive impairment, an observation that helps to distinguish the syndrome from alcohol-induced persisting amnestic disorder. The decreased brain functioning, including problems with psychomotor performance, tends to improve with abstinence, but perhaps half of all affected patients have long-term and even permanent memory and thinking disabilities. Perhaps 50 to 70 percent of these patients evidence increased size of the brain ventricles and shrinkage of the cerebral sulci, although these changes appear to be partially or completely reversible during the first year of complete abstinence. In the final analysis, it is unlikely that there is a single alcoholic dementia syndrome; rather, the problem seems to represent the combined effects of trauma, vitamin deficiencies, and the direct actions of alcohol and acetaldehyde.

Alcohol-Induced Mood Disorder In the context of heavy and repetitive intake of any brain depressant, such as alcohol, symptoms of severe depression are common and may be labeled as an *alcohol-induced mood disorder* (Table 14.6-18). Like DSM-III-R, ICD-10 retains this and most related substance-induced syndromes in the section on organic mental disorders labeled as an *organic mood [affective] disorder*. For long-lasting mood disturbances, ICD-10 also has the labels of *other persistent mood [affective] disorders* and *persistent mood [affective] disorder*. The diagnosis in DSM-IV or ICD-10 focuses on either sadness or mania-like symptoms severe enough to impair functioning that occur only in the context of repeated heavy drinking and continue for several days to 4 weeks after abstinence. Heavy intake of alcohol over several days results in many of the symptoms observed in major depressive disorder, but the intense sadness improves within days to weeks of abstinence. Consistent with the theory that intoxication with brain depressants can cause severe symptoms of depression is the documentation that 80 percent of alcoholic persons report histories of intense depression; 30 to 40 percent were depressed for 2 or more weeks, during which they had symptoms that resembled a major depressive episode. However, when information from patients and resource persons was carefully evaluated, only 5 percent of alcoholic men and 10 percent of alcoholic women ever had depressions that met the criteria for major depressive disorder when they had not been drinking heavily.

Clinical data reveal that when even severe depression develops in alcoholic persons, they are likely to improve fairly rapidly without medications or intensive psychotherapy aimed at the depressive symptoms. A recent study of almost 200 alcoholic men found that 40 percent had severe levels of depression after 1 week of abstinence. However, the percentage with pervasive depressive symptoms decreased to about 5 percent after 3 additional weeks of abstention from alcohol, even though no treatment was given for the mood symptoms.

At the end of several weeks, most alcoholic patients are left with mood swings or intermittent symptoms of sadness that can resemble cyclothymic disorder or dysthymic disorder. Even those mild and intermittent depressive symptoms are likely to diminish and disappear with time. The presence of the dysthymic symptoms usually indicates the normal course of a withdrawal syndrome and not an independent mood disorder.

A consultation was requested on a 42-year-old woman with alcohol dependence who complained of persisting severe depressive symptoms despite 5 days of abstinence. In the initial stage of the interview she noted that she had “always been depressed” and felt that she “drank to cope with the depressive symptoms.” Her current complaint included a prominent sadness that had persisted for several weeks, difficulties concentrating, initial and terminal insomnia, and a feeling of hopelessness and guilt. In an effort to distinguish between an alcohol-induced mood disorder and an independent major depressive episode, a time-line based history was obtained. This focused on the age of onset of DSM-IV alcohol dependence, periods of abstinence that extended for several months or more since the onset of dependence, and the ages of occurrence of clear major depressive episodes lasting several weeks or more at a time. Despite this patient's original complaints, it became clear that there had been no major depressive episodes prior to her mid-20s when alcohol dependence began, and that during a 1-year period of abstinence related to the gestation and neonatal period of the birth of her son, her mood had significantly improved. A provisional diagnosis of an alcohol-induced mood disorder was made. The patient was offered education, reassurance, and cognitive therapy to help her to deal with the depressive symptoms, but no antidepressant medications were prescribed. The depressive symptoms remained at their original intensity for several additional days, and then began to improve. By approximately 3 weeks abstinent the patient no longer met criteria for a major depressive episode, although she demonstrated mood swings similar to dysphemia for several additional weeks. This case is a fairly typical example of an alcohol-induced mood disorder in an individual with alcohol dependence.

Alcohol-Induced Anxiety Disorder Anxiety symptoms fulfilling the diagnostic criteria for *alcohol-induced anxiety disorder* are also common in the context of acute and protracted alcohol withdrawal (see Table 15.6-18). In ICD-10 these are listed as organic anxiety disorders resembling generalized anxiety or panic disorders. Almost 80 percent of alcoholic persons report panic attacks during acute withdrawal; their complaints can be intense enough for the clinician to consider diagnosing a panic disorder. Similarly, during the first 4 to 6 weeks of abstinence, persons with severe alcohol problems are likely to avoid some social situations for fear of being overwhelmed by anxiety (i.e., they have symptoms resembling social phobia); their problems can at times be severe enough to resemble agoraphobia. The symptoms of nervousness during acute and protracted withdrawal can also include many of the problems seen in generalized anxiety disorder. However, when psychological or physiological symptoms of anxiety are observed in alcoholic persons only in the context of heavy drinking or within the first several weeks or months of abstinence, the symptoms are likely to diminish and subsequently disappear with time alone. If one correctly identifies temporary substance-induced syndromes, alcoholic persons are no more likely than people in the general population to have most independent major anxiety disorders. They are, however, much more likely to have temporary but intense symptoms of anxiety. Two anxiety disorders that might be more closely tied to alcoholism are panic disorder and social phobia.

A 52-year-old African-American male lawyer entered treatment for alcohol dependence with the chief complaint of “panic attacks” as well as alcohol problems. In light of evidence of concomitant panic disorder and alcohol dependence, a time-line history was established. This individual first met criteria for panic disorder by approximately age 18, and an intensification of occasional panic attacks occurred as he went off to college. After several visits to an emergency room during his freshman year, the diagnosis of panic disorder was made, and the current retrospective history revealed no evidence of dependence on alcohol or any other substance at that time. His panic disorder was treated with behavioral therapy and antidepressant medication, and was under good control when, at approximately age 33, he experienced an intensification of his alcohol intake and developed associated problems. He appears to have fulfilled criteria for alcohol dependence. During the course of his alcoholism, his panic disorder symptoms increased and decreased in intensity over time, with obvious exacerbations whenever he attempted to cut back on his alcohol intake. The diagnosis was alcohol dependence along with an independent panic disorder. Treatment focused on a continuation of the cognitive-behavioral therapy, antidepressant medications (when needed) for the panic disorder, as well as alcohol rehabilitation.

Alcohol-Induced Psychotic Disorder About 3 percent of alcoholic persons have psychotic symptoms in the context of heavy drinking and withdrawal. In DSM-III-R those problems were labeled *organic hallucinosis* or *delusional disorders*. In ICD-10 they are presented as organic delusional disorders in the organic section and as a psychotic disorder in the substance use disorders section. Many of the symptoms resemble those seen in schizophrenia, but when the psychotic features develop only in the context of alcohol problems they are likely to clear spontaneously. The syndromes are likely to recur only if heavy alcohol intake resumes.

A 39-year-old male letter carrier was brought to an emergency room by the police after he behaved in an unusual fashion at home and complained that his neighbors were trying to kill him. The history obtained from the patient and his wife revealed that his psychotic thinking developed slowly over the preceding 3 weeks; he began with feelings that people were looking at him at work, progressed to vague feelings that people were against him, and went on to frank auditory hallucinations that people at work and in the neighboring houses were talking about their plans to kill him. He had no insight into those paranoid delusions and auditory hallucinations. The relatively abrupt onset of the syndrome—he was in his late 30s—pointed to a potential organic cause, and further probing documented his daily drinking of between 6 and 18 beers for at least the preceding 10 weeks. A diagnosis of alcohol-induced psychotic disorder with onset during intoxication was made, and both hallucinations and delusions disappeared after 3 weeks of abstinence. After alcohol treatment, the man stayed sober for the next 8 months. Unfortunately, he later resumed heavy drinking and had a recurrence of both hallucinations and delusions.

Alcohol-Related Disorder Not Otherwise Specified DSM-IV allows for the diagnosis of *alcohol-related disorder not otherwise specified* for alcohol-related disorders that do not meet the diagnostic criteria for any of the other diagnoses (Table 11.2-7). ICD-10 offers the listings of *other or unspecified mental and behavioral disorders induced by alcohol*.

The alcohol-related disorder not otherwise specified category is for disorders associated with the use of alcohol that are not classifiable as alcohol dependence, alcohol abuse, alcohol intoxication, alcohol withdrawal, alcohol intoxication delirium, alcohol withdrawal delirium, alcohol-induced persisting dementia, alcohol-induced persisting amnestic disorder, alcohol-induced psychotic disorder, alcohol-induced mood disorder, alcohol-induced anxiety disorder, alcohol-induced sexual dysfunction, or alcohol-induced sleep disorder.

Table 11.2-7 DSM-IV Diagnostic Criteria for Alcohol-Related Disorder Not Otherwise Specified

Laboratory and Physical Examination Establishing the diagnosis for alcohol abuse or dependence centers on obtaining from the patient and a resource person a history of the patient's life problems and the possible role played by alcohol. Up to one third of all psychiatric patients are likely to have an alcohol problem that either caused or exacerbated the presenting clinical condition.

The process of identification can also be facilitated by a series of blood tests, outlined in Table 11.2-8. Those state markers of heavy drinking reflect physiological alterations likely to be observed if the patient regularly ingests four or more drinks a day over many days or weeks. One of the most sensitive and specific of the markers (perhaps 60 to 80 percent sensitivity and specificity) is a level of 30 or more units per liter of g-glutamyltransferase (GGT), an enzyme that aids in the transport of amino acids and that is found in most areas of the body. Because this enzyme is likely to return to normal levels after 2 to weeks of abstinence, even 20 percent increases in enzyme levels above those observed after 4 weeks of abstinence can be useful in identifying patients who have returned to drinking after treatment. Equally impressive results have been reported for the measure of a deglycosylated form of the protein transferrin, known as *carbohydrate-deficient transferrin* (*CDT*). Using a commercially available assay, CDTect, and employing a cutoff of 20 mg/L, this test has both a sensitivity and a specificity of 65 to 80 percent for the identification of the heavy consumption of alcohol (e.g., 5 to 8 drinks per day for a week); these figures might be slightly lower for women. With a biological half-life of about 16 days, this test can also be useful in monitoring abstinence in alcoholics. It appears that patients not identified by higher GGT values might still have elevations in CDT so that both tests should be used for identification and abstinence-monitoring functions in alcoholics.

Test	Relevant Range of Results
Gamma glutamyl transferase (GGT)	>30 U/L
Carbohydrate-deficient transferrin (CDT)	>20mg/L
Mean corpuscular volume (MCV)	>91 μm,
Uric acid	>6.4 mg/dL for men >5.0 mg/dcL for women
Serum glutamic-oxaloacetic transaminase (aspartate aminotransferase) [SGOT (AST)]	>45 IU/L
Serum glutamic-pyruvic transaminase (alanine aminotransferase) [SGPT (ALT)]	>45 IU/L
Triglycerides	>160 mg/dL

Table 11.2-8 State Markers of Heavy Drinking Useful in Screening for Alcoholism

The MCV blood test, with perhaps 70 percent sensitivity and specificity, is a state marker when the size of the red blood cell is 91 or more cubic micrometers. The 120-day life span of the red cell does not allow the test to be used as an indicator of a return to drinking after about 1 month of abstinence. Other tests that can be helpful in identifying patients who are regularly consuming heavy doses of alcohol include those for high normal concentrations of uric acid (greater than 6.4 mg/dL, with a range that depends on the sex of the person); mild elevations in the usual liver function tests, including aspartate aminotransferase and alanine aminotransferase; and elevated levels of triglycerides or LDL cholesterol.

A number of physical findings can also be useful in identifying the alcoholic patient. These findings include modest elevations in blood pressure, frequent bruising, cancer of the head and neck and upper digestive tract, an enlarged liver, evidence of cirrhosis, and symptoms consistent with pancreatitis.

A 41-year-old mechanic consulted a physician for a physical examination related to his application for an insurance policy. The history revealed a prior divorce along with some problems in the current marriage, difficulties getting along with his foreman at work who complained about the patient's occasional absences, and some difficulties with insomnia and restless sleep. The patient denied symptoms of intense depression, but did admit to moodiness. The physical examination revealed a mild elevation in blood pressure (140/90), but was otherwise within normal limits. The panel of laboratory results included a MCV of 94.8 cubic microns, a CDT value of 25 mg/L, and a GGT of 55 units per liter. Although the patient had initially reported himself to be a “moderate social drinker,” a more detailed history revealed that he drank 6 to 8 beers on weeknights, and 12 to 20 beers a day on weekends. The physician shared her impressions regarding the pattern of life problems, noted the elevation in the state markers of heavy drinking, and told the patient that she felt that he had reached a point where alcohol was causing repeated problems, while noting in the chart the diagnosis of alcohol dependence. The objective data available to the physician regarding laboratory values helped her efforts to optimize this patient's motivation for abstinence.

DIFFERENTIAL DIAGNOSIS

Once the pattern of alcohol-related life problems has been established, the diagnosis of alcohol abuse or dependence is fairly obvious. A substantial proportion of the information presented in this chapter also helps the clinician to take the next logical step: determining whether an independent major psychiatric disorder exists. Briefly, individuals who present with clinically significant levels of depression, anxiety, or psychotic symptoms in addition to their alcoholism should be evaluated using the time-line approach in order to determine whether the psychiatric symptoms are likely to have been substance-induced (and are thus temporary) or represent independent and longer-term psychiatric disorders. In addition, the clinical course of the psychiatric symptoms should be closely observed during the subsequent several weeks to a month or so of abstinence, in order to determine whether the depression, anxiety, and other symptoms decrease in intensity over time. Although some symptoms might remain as part of a protracted withdrawal syndrome, if the disorder is substance-induced the individual should no longer fulfil criteria for the full diagnostic syndrome after a month or so of abstaining from alcohol.

Antisocial Personality Disorder When the emphasis on the chronological development of symptoms is used, at least three diagnoses— *antisocial personality disorder, schizophrenia*, and *bipolar I disorder*—are likely to predate alcohol abuse or dependence and to be true comorbid conditions. Antisocial personality disorder, listed on Axis II, begins early in life and has major effects on many aspects of life functioning. The diagnosis is based on evidence of severe antisocial behaviors in many areas beginning before the age of 15 years and continuing into adulthood. Persons with antisocial personality disorder are described as impulsive, frequently violent, highly likely to take risks, and unable to learn from their mistakes or to benefit from punishment. A person who carries these characteristics into adolescence, the typically time for experimentation with alcohol and drugs, can be expected to have difficulty controlling substance use. Thus, perhaps 80 percent or more of persons with antisocial personality disorder are likely to have severe secondary alcohol problems in the course of their lives. A diagnosis of pre-existing antisocial personality disorder with subsequent alcohol abuse or dependence indicates someone who is more likely than the average alcoholic person to have severe coexisting drug problems, to be violent, to discontinue treatment prematurely, and to have a much less than optimistic prognosis.

Debate continues on the optimal manner of viewing the co-occurrence of antisocial personality disorder and alcoholism, but most researchers agree that the personality disorder is a separate entity worthy of diagnosis. The genetic factors that increase the risk for antisocial personality disorder may be separate from those that affect the development of alcoholism. In most treatment programs, perhaps 5 percent of alcoholic women and between 10 and 20 percent of alcoholic men have preexisting antisocial personality disorder. Other Axis II-type symptoms are often observed during intoxication and as part of the acute and protracted abstinence syndromes, but they have not been documented to predate the alcohol-related disorders.

Schizophrenia A second disorder in which secondary alcohol problems are more common than in the general population is schizophrenia. Characterized by what is usually a slow onset of paranoid delusions and auditory hallucinations in a clear sensorium and typically beginning in the mid-teens to the 20s, schizophrenia is likely to be severe and debilitating. Possibly because of a lack of long-term treatment facilities, persons with schizophrenia are likely to live in inner-city areas and to spend a great deal of time on the streets. Perhaps because they use alcohol to decrease feelings of isolation or to self-medicate their symptoms, persons with schizophrenia are more likely than those in the general population to go on to have severe alcohol-related life problems. Their alcohol intake is likely to undercut the effectiveness of appropriate antipsychotic medications, to increase mood swings and signs of psychoses, and to contribute to a downward course of schizophrenia that entails repeatedly revolving into and out of inpatient care. Because most alcohol treatment programs exclude actively psychotic patients, schizophrenia persons rarely appear in inpatient alcohol treatment programs. However, severe alcohol-related disorders are observed in 30 percent or so of schizophrenic persons being treated in public mental health facilities.

A 34-year-old unemployed, divorced white male entered treatment for alcohol dependence. His history revealed seven prior hospitalizations since the age of 23, with most discharge diagnoses including schizophrenia, and some noting alcohol dependence. The patient was a poor historian, and no additional informants were available. However, the information from prior hospitalizations revealed that at the time of the initial inpatient stay the patient had experienced approximately 1 year of auditory hallucinations and paranoid delusions. During that time he had not yet demonstrated severe enough problems related to alcohol or other drugs to fulfill criteria for dependence. Subsequently, a hospitalization at age 28 was followed by a 6-month stay in a recovery home during which the records indicated that there had been no alcohol or illicit drug use, despite which the patient continued to demonstrate auditory hallucinations and paranoid delusions without insight. The current hospitalization involved a thorough physical examination and initiation of treatment of alcohol withdrawal while re-evaluating and stabilizing the antipsychotic medication level for this individual who demonstrated both alcohol dependence and an independent schizophrenia disorder.

Bipolar I Disorder The third disorder in which severe alcohol problems are overrepresented is bipolar I disorder. In a manic episode, the patient is hyperexcited and impulsive, carries out most activities to excess, has poor judgment, and is likely to develop temporary alcohol problems. The severity of the manic symptoms usually precludes inpatient alcohol rehabilitation while the patient is actively manic. However, alcohol-related difficulties must be evaluated in histories taken from persons with manic features entering mental health facilities.

Major Anxiety Disorders Finally, there are data from recent studies that support a small but statistically significant association between independent (i.e., not alcohol-induced) panic disorder and perhaps independent social phobia and alcohol dependence. One large investigation involved over 3000 personal interviews carried out across six centers in different parts of the United States. While about 90 percent of alcohol-dependent men and women did not have a major anxiety disorder there was no evidence for a significant increased risk for most major anxiety disorders, the rates of independent panic disorder and independent social phobia were significantly higher than in controls.

Other Disorders Debate in the literature continues about whether major depressive disorder, agoraphobia, obsessive-compulsive disorder, and other major psychiatric diagnoses are overrepresented in the histories of alcoholic persons. Several studies indicate that, when the time-line method is used and a history is obtained from many informants, little evidence is found for very high rates of most independent psychiatric disorders among alcoholic persons, other than the three disorders noted above. Therefore, although the majority of alcoholic persons have temporary psychiatric symptoms, they are not more likely than are persons in the general population to carry an independent psychiatric syndrome other than the three exceptions discussed above.

There are interesting and complex relations between alcohol dependence and dependence on other drugs. Men and women with the antisocial personality disorder demonstrate a marked increased risk for dependence on multiple substances, including alcohol. It is also probable that individuals with dependence on opiates and stimulants (such as cocaine and amphetamines) exhibit an increased risk for alcohol dependence even in the absence of antisocial personality disorder. However, most alcohol-dependent people do not meet the criteria for dependence on other drugs. Several recent investigations of children of alcohol-dependent men and women, as well as the large Collaborative Study on the Genetics of Alcoholism (COGA) indicated that once the effects of the antisocial personality disorder were controlled, alcohol dependence appeared to run relatively true within families, without evidence of a marked crossover between alcoholism and other dependencies. An exception to this general rule is nicotine dependence, which has long been noted to be elevated among alcohol-dependent individuals, a finding that has been hypothesized to relate to either attempts to use nicotine to try to moderate some of the effects of high doses of alcohol or withdrawal, or a possible genetic relationship between nicotine and alcohol-dependence syndromes.

COURSE AND PROGNOSIS

Details of a clinical course that fit the great majority of persons with any disorder are difficult to describe. However, sufficient data regarding alcohol-related disorders are available to offer a general outline of the typical pattern of problems.

Several recent evaluations of large numbers of alcohol-dependent individuals suggest that most subgroups of alcoholics are more similar than different on the time course and prevalence of alcohol-related life difficulties. The differences that do exist reflect characteristics of individuals in society in general and do not appear to

indicate any unique aspects of their alcoholism as such. Thus, for example, similar to women in the general population, compared to male alcoholics, female alcoholics have higher rates of independent depressive episodes, and slightly lower proportions of women who have ever engaged in alcohol-related violence or have had severe alcohol-related driving problems. However, in general, the clinical courses of alcohol-dependent men and women are relatively similar. Older as compared to younger alcohol-dependent individuals are more likely to have medical problems, to take multiple medications, to experience more severe withdrawal syndromes, probably to have a less extensive social support system. Again, these characteristics reflect differences between older versus younger individuals in general more than they indicate potential unique aspects of alcoholism in the geriatric population. Thus, while the reader is advised to take the following information as general guidelines that must be applied with common sense to subgroups of individuals, it is possible to present generalities regarding the usual clinical course of alcoholism and its treatment.

Early Course Patients with antisocial personality disorder who later go on to develop alcoholism have an early onset of drinking, intoxication, and alcohol-related problems, but that scenario is not applicable to the other 80 to 90 percent of alcoholic men and 95 percent of alcoholic women. Usually, alcoholic persons have their first drink (other than taking a sip from a parent's glass) between the ages of 13 and 15 years, the first intoxication is likely to occur at 15 or 16 years, and the first evidence of a minor alcohol-related problem is usually observed in the late teenage years. These milestones do not differ significantly from what is expected for people in the general population who do not later go on to develop alcohol abuse or dependence.

For the average person the pattern of severe difficulties becomes apparent in the middle 20s to the middle 30s when a constellation of symptoms of relatively great severity is likely to be observed: an alcohol-related breakup of a significant relationship, a second alcohol-related driving or public intoxication arrest, evidence of alcohol withdrawal, being told by a physician that alcohol has harmed the person's health, or significant interference with functioning at school or work. This pattern probably does not vary much with the type of beverage used—beer, wine, or spirits.

The landmarks in Table 11.2-9 are only rough estimates and can differ greatly between people and among various groups. Women, for example, are likely to begin drinking later than men, but their subsequent escalation of symptoms is likely to be slightly more rapid than that seen in men.

Age at first drink*	13–15 years
Age at first intoxication*	15–17 years
Age at first problem*	16–22 years
Age at onset of dependence	25–40 years
Age at death	60 years
Fluctuating course of abstention, temporary control, alcohol problems	
Spontaneous remission in 20 percent	

* Same as general population.

Table 11.2-9 Clinical Course of Alcohol Dependence

Later Course Once alcohol's interference with life functioning has become apparent, the future is likely to include periods of drinking problems that repeatedly alternate with periods of abstinence and periods of alcohol intake unassociated with problems *(temporary controlled drinking)*. Abstinence often develops in response to some interpersonal, social, or legal crisis and is likely to produce only mild withdrawal symptoms. Average alcoholic persons are then likely to use the temporary cessation of drinking problems to convince themselves that alcohol is not really a cause for concern after all. Those periods of abstinence, lasting days to months, are common in the course of most persons with alcoholism and are usually followed by periods during which drinking rules are established and are temporarily followed. The person is likely to consume only beer or wine (forgetting that a glass of beer, a glass of wine, and a shot of whiskey have similar amounts of alcohol) and tries to drink only at certain times of the day and under certain conditions. This period of temporary control soon leads to an escalation of alcohol intake, the accumulation of a new set of problems, and a subsequent crisis. These events, in turn, are likely to precipitate a new period of abstinence, and the cycle begins again. Thus, controlled drinking is a common but temporary condition for most alcoholic persons. Those who have less severe alcohol problems, such as those who may fulfill the diagnostic criteria for alcohol abuse in DSM-IV, are probably more likely to have long-term and even permanent periods of control. However, several research projects have indicated that long-term continued control is not likely to be seen once a person meets the diagnostic criteria for alcohol dependence.

An additional attribute important in the course of alcohol dependence is the phenomenon of spontaneous remission. Perhaps in response to nonspecific events or to a crisis, the alcoholic person promises to abstain and keeps the promise forever. Whatever the cause of the abstinence, about 20 percent or more of alcoholic persons, if followed over a long enough period of time, probably do achieve permanent abstinence, even without formal treatment or participation in such self-help groups as Alcoholics Anonymous (AA).

Even in the average blue- or white-collar alcohol-dependent person, the life span is likely to be decreased. This foreshortening of 10 to 15 years is associated with many causes, including the marked increased risks for heart disease, cancer, accidents, and suicide among alcoholic individuals.

Prognosis Between 10 and 40 percent of alcoholic persons enter some kind of formal treatment program during the course of their alcohol problems. A number of prognostic signs are favorable. First is the absence of preexisting antisocial personality disorder or a diagnosis of other substance abuse or dependence. Second, evidence of general life stability with a job, continuing close family contacts, and the absence of severe legal problems also bodes well for the patient. Third, if the patient stays for the full course of the initial rehabilitation (perhaps 2 to 4 weeks), the chances of maintaining abstinence are good. The combination of these three attributes predicts at least a 60 percent chance for 1 or more years of abstinence. Few studies have documented the long-term course, but researchers agree that 1 year of abstinence is associated with a good chance for continued abstinence over an extended period of time. However, alcoholic persons with severe drug problems (especially intravenous drug use or cocaine or amphetamine dependence) and those who are homeless may have only a 10 to 15 percent or so chance of achieving 1 year of abstinence.

Accurately predicting whether any specific person will achieve or maintain abstinence is impossible, but the prognostic factors listed above are associated with an increased likelihood of abstinence. However, the factors reflecting life stability probably explain only 20 percent or less of the course of alcohol use disorders. Many forces that are difficult to measure have significant effects on the clinical course; they are likely to include such intangibles as levels of motivation and the quality of the patient's social support system.

In general, alcoholic persons with preexisting independent major psychiatric disorders—such as antisocial personality disorder, schizophrenia, and bipolar I disorder—are likely to run the course of their independent psychiatric illness. Therefore, for example, clinicians must treat the patient with bipolar I disorder who has secondary alcoholism with appropriate psychotherapy and lithium (Eskalith), use relevant psychological and behavioral techniques for the patient with antisocial personality disorder, and offer appropriate antipsychotic medications on a long-term basis to the patient with schizophrenia. The goal is to keep the symptoms of the independent psychiatric disorder as minimal as possible in the hope that a greater level of life stability will be associated with a better prognosis for the patient's alcohol problems.

TREATMENT

The elements of treatment appropriate for patients with severe alcohol problems are fairly straightforward. Much of the clinical challenge comes in recognizing how prevalent the alcohol-related disorders are, how often those conditions present with symptoms of other psychiatric syndromes, and how to use clinical clues, physical findings, and laboratory tests to identify alcoholism.

Three general steps are involved in treating the alcoholic person once the disorder has been diagnosed—intervention, detoxification, and rehabilitation. Those approaches assume that all possible efforts have been made to optimize medical functioning and to address psychiatric emergencies. Thus, for example, the alcoholic person with symptoms of depression severe enough to be suicidal requires inpatient hospitalization for at least several days until the suicidal ideation disappears.

Similarly, the person presenting with cardiomyopathy, liver difficulties, or gastrointestinal bleeding first needs adequate attention paid to the medical emergency.

The patient with alcohol abuse or dependence must then be brought face-to-face with the reality of the disorder *(intervention)*, be detoxified if needed, and begin rehabilitation. The essentials of these three steps for an alcoholic person with independent psychiatric syndromes are quite similar to the approaches used for the primary alcoholic person without independent psychiatric syndromes. However, in the former case the treatments are applied after the psychiatric disorder has been stabilized to the maximum degree possible.

Intervention The goal in this step, which has also been called *confrontation*, is to break through feelings of denial and to help the patient recognize the adverse consequences likely to occur if the disorder is not treated. Intervention is as a process aimed at increasing to as high a level as possible the levels of motivation for treatment and for continued abstinence.

This step often involves convincing patients that they are responsible for their own actions while reminding them how alcohol has created significant life impairments. The psychiatrist often finds it useful to take advantage of the person's chief presenting complaint, whether it is insomnia, difficulties with sexual performance, an inability to cope with life stresses, depression, anxiety, or psychotic symptoms. The psychiatrist can then teach the patient how alcohol has either created or contributed to these problems, and can reassure the patient that abstinence can be achieved with a minimum of discomfort.

> A physician was consulted by a 43-year-old businessman who was concerned about his wife. He had recently been confronted by their 21-year-old daughter who felt that her mother was an alcoholic. The daughter noted her mother's slurred speech on several recent occasions when the daughter called home, times during the day when the mother was apparently home but did not answer the telephone, and observed high levels of alcohol consumption. A more detailed history revealed that the husband had been concerned about the wife's drinking pattern for at least 5 years, relating her practice of staying up after he went to bed, retiring later with alcohol on her breath. He also noted her consumption of 10 to 12 drinks at parties, with the resulting tendency to isolate herself from the remaining guests, her panic-like behavior regarding the need to pack liquor when they go on trips where alcohol might not be readily available, and what he observed to be a tremor of her hands some mornings during breakfast. The husband was given several potential courses of action, including the possibility of referring the spouse for treatment with the physician. The husband was advised to share his concern with his wife at a time when she was not actively intoxicated, emphasizing specific times and events where her impairment with alcohol was noted. He was also asked to consider whether a close friend of many years and the adult daughter might be included in this intervention, and it was suggested that a tentative appointment might be made with the clinician (or with an alcohol and drug treatment program) so that a next step could be established if the intervention was successful.

A physician intervening with a patient can use the same nonjudgmental but persistent approach each time an alcohol-related impairment is identified. It is the level of persistence rather than exceptional interpersonal skills that usually gets results. A single intervention is rarely enough. Most alcoholic persons need a series of reminders of how alcohol contributed to each developing crisis before they seriously consider abstinence as a long-term option.

Family The family can be of great help in the intervention. Family members must learn not to protect the patient from the problems caused by alcohol otherwise the patient may not be able to gather the energy and the motivation necessary to stop drinking.

During the intervention stage, the family can also suggest that the patient meet with persons who are themselves recovering from alcoholism, perhaps through AA, and they can meet with groups, such as Alanon, that reach out to family members. Those support groups for families meet many times a week and help family members and friends see that they are not alone in their fears, worry, and feelings of guilt. Members share coping strategies and help each other find community resources. The groups can be most useful in helping family members rebuild their lives, even if the alcoholic person refuses to seek help.

Detoxification Most persons with alcohol dependence have relatively mild symptoms when they stop drinking. If the patient is in relatively good health, adequately nourished, and has a good social support system the depressant withdrawal syndrome usually resembles a mild case of the flu. Even intense withdrawal syndromes rarely approach the severity of symptoms described by some early textbooks in the field.

The essential first step in detoxification is a thorough physical examination. In the absence of a serious medical disorder or combined drug abuse, severe alcohol withdrawal is unlikely. The second step is to offer rest, adequate nutrition, and multiple vitamins, especially those containing thiamine.

Mild or Moderate Withdrawal Withdrawal develops because the brain has physically adapted to the presence of a brain depressant and cannot function adequately in the absence of the drug. Giving enough of a brain depressant on the first day to diminish symptoms and then weaning the patient off the drug over the next 5 days offers most patients optimal relief and minimizes the possibility that a severe withdrawal will develop. Any depressant—including alcohol, barbiturates, or any of the benzodiazepines—can work, but most clinicians chose a benzodiazepine for its relative safety. Adequate treatment can be given with either short-acting drugs, such as lorazepam (Ativan), or long-acting substances, such as chlordiazepoxide (Librium) and diazepam (Valium).

An example of treatment is the administration of 25 mg of chlordiazepoxide by mouth three or four times a day on the first day, with a notation to skip a dose if the patient is asleep or feeling sleepy. An additional one or two 25-mg doses during the first 24 hours can be used if the patient is jittery or shows signs of increasing tremor or autonomic dysfunction. Whatever the dosage required on the first day, the benzodiazepine can be decreased by 20 percent of it each subsequent day, with a resulting need for no further medication after 4 or 5 days. When using a long-acting agent, such as chlordiazepoxide, the clinician must avoid producing excessive sleepiness through overmedication; if the patient is sleepy, the next scheduled dose should be omitted. When taking a short-acting drug, such as lorazepam, the patient must not miss any dose because rapid changes in blood benzodiazepine concentrations may precipitate a severe withdrawal.

A social model program of detoxification saves money by avoiding medications while using social supports. This less expensive regimen can be helpful for mild or moderate withdrawal syndromes. Some clinicians have also recommended b-adrenergic receptor antagonists, such as propranolol (Inderal), or a-adrenergic receptor such as agonists, clonidine (Catapres), although these medications do not appear to be superior to the benzodiazepines. Unlike the brain depressants, these other agents do little to decrease the risk of seizures or delirium.

Severe Withdrawal For the approximately 1 to 3 percent of alcoholic patients with extreme autonomic dysfunction, agitation, and confusion—that is, those with alcoholic withdrawal delirium, also called delirium tremens—no optimal treatment has yet been developed. The first key step is to ask why such a severe and relatively uncommon withdrawal syndrome has occurred; the answer often relates to a severe concomitant medical problem that needs immediate treatment. The withdrawal symptoms can then be minimized either through the use of benzodiazepines (in which case high doses are sometimes required), or through antipsychotic agents, such as haloperidol (Haldol) and thioridazine (Mellaril). Once again, doses are used on the first or second day to control behavior, and the patient can be weaned off the medication by about the fifth day.

Another 1 to 3 percent of patients may have a single grand mal convulsion; the rare person has multiple fits and the peak incidence is on the second day of withdrawal. Such patients require a neurological evaluation, but in the absence of evidence of a seizure disorder they do not benefit from anticonvulsant drugs.

Rehabilitation For most patients, rehabilitation includes three major components: (1) continued efforts to increase and maintain high levels of motivation for abstinence, (2) work to help the patient readjust to a life-style free of alcohol, and (3) relapse prevention. Because these steps are carried out in the context of acute and protracted withdrawal syndromes and life crises, treatment requires repeated presentations of similar materials that remind the patient how important abstinence is and that help the patient develop new day-to-day support systems and coping styles.

No single major life event, traumatic life period, or identifiable psychiatric disorder is known to be a unique cause of alcoholism. In addition, the effects of any causes of alcoholism are likely to have been diluted by the effects of alcohol on the brain and the years of an altered life-style, so that the alcoholism has developed a life of its own. This is true even though many alcoholic persons believe that the cause was depression, anxiety, life stress, or pain syndromes. Research, data from records, and resource persons usually reveal that the alcohol contributed to the mood disorder, accident, or life stress, not vice versa.

The same general treatment approach is used in inpatient as well as outpatient settings. The selection of the more expensive and intensive inpatient mode often depends on evidence of additional severe medical or psychiatric syndromes, the absence of appropriate nearby outpatient groups and facilities, and the patient's history of having tried but failed in outpatient care. The treatment process in either setting involves intervention, optimizing physical and psychological functioning, enhancing motivation, reaching out to family, and using the first 2 to 4 weeks of care as an intensive period of help. Those efforts must be followed by at least 3 to 6 months of less frequent outpatient care. Outpatient care uses a combination of individual and group counseling, the judicious avoidance of psychotropic medications

unless needed for independent disorders, and involvement in such self-help groups as AA.

Counseling Counseling efforts in the first several months should focus on day-to-day life issues to help patients maintain a high level of motivation for abstinence and to enhance their levels of functioning. Psychotherapy techniques that provoke anxiety or that require deep insights have not been shown to be of benefit during the early months of recovery and, at least theoretically, may actually impair efforts at maintaining abstinence. Therefore, this discussion focuses on the efforts likely to characterize the first 3 to 6 months of care.

Counseling or therapy can be carried out in an individual or group setting; few data indicate that either approach is superior to the other. The technique used is not likely to matter greatly, and usually boils down to simple day-to-day counseling or almost any behavioral or psychotherapeutic approach focusing on the here and now. To optimize motivation, treatment sessions should explore the consequences of drinking, the likely future course of alcohol-related life problems, and the marked improvement that can be expected with abstinence. Whether in an inpatient or an outpatient setting, individual or group counseling is usually offered for a minimum of three times a week for the first 2 to 4 weeks, followed by less intense efforts, perhaps once a week, for the subsequent 3 to 6 months.

Much time in counseling deals with how to build a life-style free of alcohol. Discussions cover the need for a sober peer group, a plan for social and recreational events without drinking, and approaches for reestablishing communication with family members and friends.

The third major component, relapse prevention, begins with identifying situations in which the risk for relapse is high. The counselor must help the patient to develop modes of coping to be used when the craving for alcohol increases or when any event or emotional state makes a return to drinking likely. An important part of relapse prevention is reminding the patient about the appropriate attitude toward slips. Those short-term experiences with alcohol can never be used as an excuse for returning to regular drinking. The efforts to achieve and maintain a sober life-style are not a game in which all benefits are lost with that first sip. Rather, recovery is a process of trial and error; patients use slips when they occur to identify high-risk situations and to develop more appropriate coping techniques.

Most treatment efforts recognize the effects that alcoholism has on the significant people in the patient's life and an important aspect of recovery involves helping family members and close friends to understand alcoholism and how rehabilitation is an ongoing process that lasts for 6 to 12 or more months. Couples and family counseling and support groups for relatives and friends help the persons involved to rebuild relationships, to learn how to avoid protecting the patient from the consequences of any drinking in the future, and to be as supportive as possible of the alcoholic patient's recovery program.

Medications If detoxification has been completed and the patient is not one of the 10 to 15 percent of alcoholic persons who have an independent mood disorder, schizophrenia, or anxiety disorder, there is little evidence in favor of prescribing psychotropic medications for the treatment of alcoholism. Lingering levels of anxiety and insomnia as part of a reaction to life stresses and protracted abstinence should be treated with behavior modification approaches and reassurance. Medications, including benzodiazepines, for these symptoms are likely to lose their effectiveness much faster than the insomnia disappears; as a result the patient may increase the dose and have subsequent problems. Similarly, sadness and mood swings can linger at low levels for several months. However, controlled clinical trials indicate no benefit in prescribing antidepressant medications or lithium to treat the average alcoholic person who has no independent or long-lasting psychiatric disorder. The mood disorder will clear before the medications can take effect, and patients who resume drinking while on the medications face significant potential dangers. With little or no evidence that the medications are effective, the dangers significantly outweigh any potential benefits from their routine use.

One possible exception to the proscription against the use of medications is the alcohol-sensitizing agent disulfiram. Disulfiram is given in dosages of 250 mg a day before the patient is discharged either from the intensive first phase of outpatient rehabilitation or from inpatient care. The goal is to place the patient in a condition in which drinking alcohol precipitates an uncomfortable physical reaction, including nausea, vomiting, and a burning sensation in the face and stomach. Unfortunately, few data convincingly prove that disulfiram is more effective than a placebo, probably because most people stop taking the disulfiram when they resume drinking. Many clinicians have stopped routinely prescribing the agent, partly in recognition of the dangers associated with the drug itself: mood swings, rare instances of psychosis, the possibility of an increase in peripheral neuropathies, the relatively rare occurrence of other significant neuropathies, and a potentially fatal hepatitis. Moreover, patients with pre-existing heart disease, cerebral thrombosis, diabetes, and a number of other conditions cannot be given disulfiram because an alcohol reaction to the disulfiram could be fatal.

Two additional promising pharmacological interventions have recently been studied. The first involves the opiate antagonist naltrexone (ReVia), which is at least theoretically believed to possibly decrease the craving for alcohol or blunt the rewarding effects of drinking. In any event, two relatively small (i.e., approximately 90 patients on the active drug across the studies) and short-term (i.e., 3 months of active treatment) investigations using 50 mg per day of this drug had potentially promising results. However, it is difficult to evaluate the full impact of this medication unless longer-term studies of relatively large groups of more diverse patients are evaluated.

The second medication of interest, acamprosate (Campral), has been tested in over 5000 alcohol-dependent patients in Europe, this drug is not yet available in the United States. Used in dosages of approximately 2000 mg per day, this medication was associated with an approximately 10 to 20 percent higher rate of positive outcome than placebo when used in the context of the usual psychological and behavioral treatment regimen for alcoholism. Once again, the mechanisms of actions of acamprosate are not known but might relate to an impact on g-aminobutyric acid (GABA) or on the NMDA system craving or the rewarding effects of alcohol.

Another medication with potential promise in the treatment of alcoholism is the nonbenzodiazepine antianxiety drug, buspirone (BuSpar), although the effect of this drug on alcohol rehabilitation is inconsistent between studies. However, at the same time, there is no evidence that antidepressant medications such as the SSRIs, lithium, or antipsychotic medications, are significantly effective in the treatment of alcoholism.

Self-Help Groups Clinicians must recognize the potential importance of self-help groups like AA. Members of AA have help available 24 hours a day, associate with a sober peer group, learn that it is possible to participate in social functions without drinking, and are given a model of recovery by observing the accomplishments of sober members of the group.

Learning about AA usually begins during inpatient or outpatient rehabilitation. The clinician can play a major role in helping patients understand the differences between specific groups. Some groups are composed only of men or women, and others are mixed; some meetings are composed mostly of blue-collar men and women whereas others are mostly for professionals; some groups place great emphasis on religion, and others are eclectic. Patients with coexisting psychiatric disorders may need some additional education about AA. The clinician should remind them that some members of AA may not understand their special need for medications and should arm the patients with ways of coping when group members inappropriately suggest that the required medications be stopped.

SUGGESTED CROSS-REFERENCES

Classification of mental disorders is discussed in Chapter 9, epidemiology in Section 5.1, and the sociocultural sciences in Chapter 4. Delirium and amnestic disorders are discussed in Chapter 10. Other substance-related disorders are discussed in Chapter 11; mood disorders are discussed in Chapter 14, anxiety disorders are discussed in Chapter 15, personality disorders are discussed in Chapter 24, and schizophrenia is presented in Chapter 12. Psychotherapies are discussed in Chapter 30.

SECTION REFERENCES

Bierut LJ, Dinwiddie SH, Begleiter H, Crowe RR, Hesselbrock V, Nurnberger JI, Jr Schuckit MA, Reich T: Familial transmission of substance dependence: Alcohol, marijuana, cocaine, and habitual smoking: A report from the Collaborative Study on the Genetics of Alcoholism. Arch Gen Psychiatry *55* (11): 982, 1998.

*Blane HT, Leonard KE, editors: *Psychological Theories of Drinking and Alcoholism, ed 2*. Guilford, New York, 1999.

Brown SA, Gleghorn A, Schuckit MA, Myers MG, Mott MA: Conduct disorder among adolescent alcohol and drug abusers. J Stud Alcohol *57:*314, 1996.

*Brown SA, Irwin M, Schuckit MA: Changes in anxiety among abstinent male alcoholics. J Stud Alcohol *52:*55, 1991.

*Carmelli D, Swan GE, Page WF, Christian JC: World War II-veteran male twins who are discordant for alcohol consumption: 24-year mortality. Am J Public Health *85:*99, 1995.

Council on Scientific Affairs: Alcoholism in the elderly. JAMA *275:*1, 1996.

Helander A, Carlsson AV, Borg S: Longitudinal comparison of carbohydrate-deficient transferrin and gamma-glutamyl transferase: Complementary markers of excessive alcohol consumption. Alcohol

Alcohol *31:*101, 1996.

Hillbom M: Alcohol and cardiovascular disease. Stroke *26:*40, 1995.

Kabel DI, Petty F: A placebo-controlled, double-blind study of fluoxetine in severe alcohol dependence: Adjunctive pharmacotherapy during and after inpatient treatment. Alcohol Clin Exp Res *20:*780, 1996.

Kendler KS, Walters EE, Neale MC, Kessler RC, Heath AC, Eaves LJ: The structure of the genetic and environmental risk factors for six major psychiatric disorders in women. Arch Gen Psychiatry *52:*374, 1995.

Kessler RC, Nelson CB, McGonagle KA, Edlund MJ, Frank RG, Leaf PJ: The epidemiology of co-occurring addictive and mental disorders: Implications for prevention and service utilization. Am J Orthopsychiatry *66:*17, 1996.

*Koob GF, Roberts AJ. Brain reward circuits in alcoholism. CNS Spectrums *4:*23, 1999.

Kozaric-Kovacic D, Folnegovic-Smalc V, Folnegovic Z, Marusic A: Influence of alcoholism on the prognosis of schizophrenic patients. J Stud Alcohol *56:*622, 1995.

Kranzler HR, Burleson JA, Del Boca FK, Babor TF, Korner P, Brown J, Bohn MJ: Busipirone treatment of anxious alcoholics: A placebo-controlled trial. Arch Gen Psychiatry *51:*720, 1994.

Martin PR, McCool BA, Singleton CK: Genetic sensitivity to thiamine deficiency and development of alcoholic organic brain disease. Alcohol Clin Exp Res *17:*31, 1993.

Nathan PE: Psychoactive substance dependence. In *The DSM-IV Source Book*, T Widiger, A Frances, editors. American Psychiatric Press, Washington, 1994.

Pfefferbaum A, Lim KO, Desmond JE, Sullivan EV: Thinning of the corpus callosum in older alcoholic men: A magnetic resonance imaging study. Alcohol Clin Exp Res *20:*752, 1996.

*Prescott CA, Kendler KS. Genetic and environmental contributions to alcohol abuse and dependence in a population-based sample of male twins. Am J Psychiatry *156:*34, 1999.

*Project MATCH Research Group: Matching alcoholism treatments to client heterogeneity: Project MATCH posttreatment drinking outcomes. J Stud Alcohol *58:*7, 1997.

Schuckit MA: Recent developments in the pharmacotherapy of alcohol dependence. J Consult Clin Psychol *64:*669, 1996.

Schuckit MA: Are the costs of alcoholism treatment justified? Drug Abuse Alcohol News *25:*1, 1996.

Schuckit MA: Biological, psychological, and environmental predictors of the alcoholism risk: A longitudinal study. J Stud Alcohol *59:*485, 1998.

Schuckit MA: *Educating Yourself About Alcohol and Drugs*. Plenum, New York, 1998.

*Schuckit MA, Daeppen J-B, Danko GP, Tripp ML, Smith TL, Li T-K, Hesselbrock VM, Bucholz KK: Clinical implications for four drugs of the DSM-IV distinction between substance dependence with and without a physiological component. Am J Psychiatry *156:*41, 1999.

Schuckit MA, Daeppen J-B, Tipp JE, Hesselbrock M, Bucholz KK: The clinical course of alcohol-related problems in alcohol dependent and nonalcohol dependent drinking women and men. J Stud Alcohol *59:*81, 1998.

Schuckit MA, Smith TL, Daeppen J-B, Eng M, Li T-K, Hesselbrock VM, Nurnberger JI Jr, Bucholz KK: Clinical relevance of the distinction between alcohol dependence with and without a physiological component. Am J Psychiatry *155:*733, 1998.

Schuckit MA, Hesselbrock V: Alcohol dependence and anxiety disorders: What is the relationship? Am J Psychiatry *151:*1723, 1994.

*Schuckit MA, Smith TL: An 8-year follow-up of 450 sons of alcoholic and control subjects. Arch Gen Psychiatry *53:*202, 1996.

Schuckit MA, Tipp JE, Bergman M, Reich W, Hesselbrock VM, Smith TL: Comparison of induced and independent major depressive disorders in 2945 alcoholics. Am J Psychiatry *154:*948, 1997.

Schuckit MA, Tipp JE, Bucholz KK, Nurnberger JI Jr, Hesselbrock VM, Crowe RR, Kramer J: The lifetime rates of three major mood disorders and four major anxiety disorders in alcoholics and controls. Addiction *92:*1289, 1997.

Schuckit MA, Tipp JE, Reich T, Hesselbrock VM, Bucholz KK: The histories of withdrawal convulsions and delirium tremens in 1648 alcohol dependent subjects. Addiction *90:*1335, 1995.

Sullivan EV, Rosenbloom MJ, Deshmukh A, Desmond JE, Pfefferbaum A: Alcohol and the cerebrum: Effects on balance, motor coordination, and cognition. Alcohol Health Res World *19:*138, 1995.

Thun MJ, Peto R, Lopez AD, Monaco JH, Henley J, Heath CW Jr, Doll R: Alcohol consumption and mortality among middle-aged and elderly U.S. adults. N Engl J Med *337:*1705, 1997.

Tsai G, Gastfriend DR, Coyle JT: The glutamatergic basis of human alcoholism. Am J Psychiatry *152:*332, 1995.

*Vaillant GE: A long-term follow-up of male alcohol abuse. Arch Gen Psychiatry *53:*243, 1996.

Van den Brandt PA, Goldbohn A, van't Veer P: Alcohol and breast cancer. Am J Epidemol *141:*907, 1995.

Volpcelli JR, Alterman AI, Hayashida M, O'Brien CP: Naltrexone in the treatment of alcohol dependence. Arch Gen Psychiatry *49:*876, 1992.

Winokur G, Coryell W, Akiskal HS, Maser JD, Keller MB, Endicott J, Mueller T: Alcoholism in manic-depressive (bipolar) illness. Am J Psychiatry *152:*365, 1995.

Yeastedt J, La Grange L, Anton RF: Female alcoholic outpatients and female college students: A correlational study of self-reported alcohol consumption and carbohydrate-deficient transferrin levels. J Stud Alcohol *59:*555, 1998.

11.3 AMPHETAMINE (OR AMPHETAMINE-LIKE)-RELATED DISORDERS

JEROME H. JAFFE, M.D.

Amphetamines are the most widely used illicit drugs, second only to cannabis, in Great Britain, Australia, and several countries of western Europe. In the United States, lifetime and current cocaine use still exceeds the nonmedical use of amphetamines; but in some parts of the country methamphetamine use increased significantly in the 1990s and became a matter for serious concern.

Despite the important pharmacological differences between amphetamine and amphetamine-like drugs and cocaine, the patterns of use, dependence, and toxicity associated with them are similar, as are the treatment approaches currently used. Among the drugs that produce subjective effects quite similar to those of amphetamine and methamphetamine and also have abuse potential, are methylphenidate (Ritalin) and phendimetrazine (Preludin), which are included in Schedule (Control Level) II of the Controlled Substance Act (CSA), and diethylpropion (Tenuate), benzphetamine (Didrex), and phentermine (Ionamin), which are included in Schedules III or IV of the CSA.

DEFINITIONS

Amphetamine use may be associated with a number of distinct disorders, of which dependence and abuse are but two. In the case of amphetamine and amphetamine-like agents, at least 10 other substance-related disorders have been described.

Amphetamine dependence is defined in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as a cluster of physiological, behavioral, and cognitive symptoms that, taken together, indicate that the person continues to use amphetaminelike drugs despite significant problems related to such use (see Table 11.1-3). This brief definition emphasizes the drug-using behavior itself, its maladaptive nature, and how the choice to engage in that behavior shifts and becomes constrained as a result of interaction with the drug over time. *Amphetamine abuse* is a term used to categorize a pattern of maladaptive use of amphetamine or an amphetaminelike drug leading to clinically significant impairment or distress and occurring within a 12-month period in which the symptoms have never met the criteria for amphetamine dependence (see Table 11.1-8).

The amphetamine-induced disorders include amphetamine intoxication, amphetamine withdrawal, amphetamine-induced psychotic disorder with delusions and amphetamine-induced psychotic disorder with hallucinations, amphetamine intoxication delirium, amphetamine-induced mood disorder, (amphetamine-induced anxiety disorder, amphetamine-induced sleep disorder, amphetamine-induced sexual dysfunction, and amphetamine-related disorder not otherwise specified. The coding scheme of DSM-IV provides distinct numbers for amphetamine dependence and amphetamine abuse, but the codes for the other amphetamine-induced disorders are common to several other substance-related disorders.

HISTORY

Amphetamines were introduced into clinical use in the early 1930s. By late in the decade, there was some concern about amphetamine dependence, and in 1938 the first reports of amphetamine psychosis appeared. Nevertheless, between 1932 and 1946 almost three dozen clinical uses of amphetamine were proposed and tried by the medical profession, and some amphetamines were available in over-the-counter nasal inhalers until as late as 1971.

Immediately following World War II, Japan experienced an epidemic of intravenous methamphetamine abuse and dependence, but until the end of the 1960s there was reluctance in the United States to believe that amphetamine and related drugs could cause addiction. However, because of growing concern over their misuse and overuse, the Food and Drug Administration (FDA) placed them under regulatory control in the mid-1960s. Despite these controls the quantity of drugs smuggled into the country or produced illegally in clandestine laboratories increased. There were enough drugs on the street, (which up to that time had come primarily from diversion of legitimately produced drugs), to fuel a major epidemic of amphetamine and methamphetamine abuse in the late 1960s. This epidemic made clear the potential toxicity of the amphetamines, especially when used intravenously, and such terms as “speed freaks” and “speed kills” left an enduring legacy in the popular vocabulary. Over the next decade regulatory controls on legitimately produced amphetamines were progressively tightened. Some misuse of amphetamines and amphetamine-like drugs persisted in the United States, with much of the supply coming from illicit laboratories. When it became illegal to obtain the commonly used precursor phenyl-2-propanone, (P2P), illicit manufacturers found ways to produce methamphetamine from ephedrine and/or pseudoephedrine widely available in over-the-counter medications for colds and asthma. The new method of synthesis actually yields a higher percentage of the active d-isomer of methamphetamine and was adopted both by criminal organizations using large-scale laboratories and by independent producers whose small laboratories, usually located in remote rural areas, are more difficult to detect and eliminate.

In the late 1980s there were reports that smoking of crystalline methamphetamine (“ice”) was on the rise, especially in Hawaii, but through the mid-1990s the use of amphetamine-like stimulants continued to be overshadowed by cocaine abuse in most parts of the United States. Over this same period in the United Kingdom, Australia, and Western Europe amphetamine use always exceeded cocaine use. In the mid-1990s, methamphetamine use rose sharply in several areas of the United States, especially in California and several states in the southwest and northwest. This increased use was evident from surveys, drug testing of arrestees, and emergency room visits for methamphetamine toxicity.

Amphetamines are used legitimately almost exclusively for the treatment of narcolepsy and attention-deficit/hyperactivity disorder, although methylphenidate is more widely prescribed for the latter indication. Some amphetaminelike agents are still prescribed as appetite suppressants, but the use of amphetamine itself for that purpose has been discouraged and is illegal in some states. Amphetamines may be useful in the treatment of atypical depression, but concern about abuse potential has discouraged the controlled clinical studies that would be necessary to define just which patients (if any) might benefit more from amphetamine-like agents than from tricyclic antidepressants or selective serotonin reuptake inhibitors (SSRIs).

COMPARATIVE NOSOLOGY

The DSM-IV diagnostic criteria for amphetamine dependence are the same generic criteria applied to other substances ranging from opioids and cocaine to alcohol. The notion of a generic concept of dependence is shared with the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). In making a diagnosis of dependence there is generally a high level of agreement between DSM-IV and ICD-10: they use similar concepts (the dependence syndrome varying in severity), although the wording of the criteria for determining the presence and severity of the syndrome differ. Both require that three elements of the syndrome occur within a 12-month period. Although DSM-IV appears to lay greater stress on tolerance and physiological dependence by asking clinicians to specify if these elements are present, it is not certain whether patients who exhibit these phenomena have a distinct form of the disorder. Patients diagnosed with alcohol dependence who exhibit tolerance or withdrawal have a more severe syndrome. One study of cocaine users found that even if tolerance and physical dependence had been required to make the diagnosis there would have been about the same number of patients meeting the criteria for dependence.

There is a major difference between ICD-10 and DSM-IV in the classification of what is called substance abuse in DSM-IV. ICD-10 does not use the term “abuse.” Instead, it includes a category of harmful use, which is substantially different from the concept of “abuse” used in DSM-IV. The concept of “harm” is limited to physical

and mental health (e.g., hepatitis, cardiac damage, episodes of depression, or toxic psychosis). It specifically excludes social impairments as follows:

Harmful patterns of use are often criticized by others and frequently associated with adverse social consequences of various kinds. The fact that a pattern of use of a particular substance is disapproved of by another person or by the culture, or may have led to socially negative consequences such as arrest or marital arguments, is not in itself evidence of harmful use.

Another difference between ICD-10 and DSM-IV is in the coding systems, which limit the number of distinct drug categories that can be recorded. ICD-10 separates cocaine-related disorders from those caused by other stimulants. Because of the limits of the system the code for stimulants includes caffeine with the amphetamines and amphetamine-like stimulants. It is not clear whether 3,4-methylenedioxymethamphetamine (MDMA)-related disorders would be included under stimulants or other drugs.

EPIDEMIOLOGY

The National Household Survey on Drug Abuse (NHSDA) conducted in 1997 found that 4.5 percent of adults (ages 12 and older) reported lifetime nonmedical use of stimulants, a slight decline from the previous year. Past-30-day use was 0.3 percent in 1997. The highest rates of use in the past year (1.5 percent) were among 18- to 25-year-olds, followed by 12- to 17-year-olds. The sample was not large enough to detect increases in methamphetamine use in the western and southwestern states that are reflected in emergency room visits for amphetamine-related toxicity or drug testing programs for arrestees

The Monitoring the Future Study (High School Survey) considers amphetamines and similar drugs together as stimulants. Among high school seniors self-reported use of stimulants has been consistently higher than use of cocaine and crack cocaine. Despite the recent increases in methamphetamine use in some parts of the United States, stimulant use for the country as a whole has fallen from the high past-30-days rate of 15.8 percent seen in 1981 to a past-30-days rate of 4.8 percent in 1997.

Two population surveys that used accepted diagnostic criteria to measure the extent of drug abuse and dependence were the Epidemiologic Catchment Area (ECA) study, carried out in the early 1980s using criteria from the third edition of DSM (DSM-III) Survey (NCS), carried out from 1990 to 1992 using criteria from the revised third edition of DSM (DSM-III-R). The ECA report combined categories of dependence and abuse for amphetamine and amphetamine-like drugs. The 1-month, 6-month, and lifetime prevalences of amphetamine abuse or dependence were 0.1, 0.2, and 1.7 percent, respectively. The NCS lifetime dependence rate for 15- to 54-year-olds was 1.7 percent; about 15 percent of respondents gave a history of some nonmedical use of stimulants. Among those who reported nonmedical use about 11 percent met criteria for dependence.

ETIOLOGY

Drug dependence, including amphetamine and amphetamine-like substance dependence, is viewed as resulting from a process in which multiple interacting factors (social, psychological, cultural, and biological) influence drug-using behavior. This process, in some cases, leads to the loss of flexibility with respect to drug use that is the hallmark of drug dependence. According to this biopsychosocial perspective, the actions of the drug are seen as critical. However, not everyone who becomes dependent experiences the effects of a given drug in the same way or is influenced by the same set of factors. Even with the same class of pharmacological agents, different factors, (social, biological, cultural) may be more or less important at different stages of the process.

As with most substances, largely social and cultural factors influence availability and initial use of amphetamines and amphetamine-like drugs; however, pharmacological factors are believed to be very important in perpetuating use and progressing to dependence on these drugs. Amphetamines have potent mood-elevating and euphorigenic actions in humans and are powerful reinforcers in animal models, particularly when the drug effects have rapid onset, as when they are injected or inhaled. Although some physical dependence develops, in contrast to the opioids and sedatives, an aversive withdrawal syndrome probably plays a less prominent role in perpetuating the use of amphetamines and amphetamine-like drugs.

Comorbidity Additional psychiatric diagnoses are quite common among those dependent on amphetamines and amphetamine-like drugs. How this comorbidity is linked etiologically to amphetamine dependence is not always clear, but epidemiological evidence shows that the presence of psychiatric disorders not related to substance abuse (e.g., mood disorders, schizophrenia, and antisocial personality disorder) substantially increases the odds of developing substance abuse or dependence. Those with conduct disorder or antisocial personality disorder are more likely to take risks and to disregard social prohibitions against using illicit drugs. Amphetamines and amphetamine-like drugs may alleviate various psychiatric disorders or dysfunctional states in some persons. For example, some users (a relative few) may find relief from adult attention-deficit disorders. The drugs may alleviate a persistent dysthymic disorder in others, and for such users the anhedonic state following amphetamine cessation may be experienced as more intense. Still others may have found that the drug facilitated sexual activity. Although such factors may explain drug use on more than one occasion, they do not account for progression to dependence or abuse.

In the United States more is known about the characteristics of young people who experiment with, and become dependent on, cocaine and “crack” than about those who use amphetamines. High school juniors and seniors who use illicit drugs in general perform less well in school, have poorer family relationships, report more psychological symptoms and health problems, and exhibit more delinquent behavior. Those who use cocaine and crack are the most delinquent, but reported anxiety and depression is no greater among this group than among those whose illicit drug use does not include cocaine or crack.

Research on the temporal appearance of the syndromes indicates that in some instances and for some syndromes drug use antedates the psychiatric disorders. In one component of the ECA study subjects were reinterviewed 1 year later. Those who reported using amphetamines, amphetamine-like drugs, or cocaine in the time between interviews were almost 8 times more likely than nonusers to have developed a depressive disorder and 14 times more likely to have experienced a panic attack. Amphetamine users report a wide range of psychiatric symptoms, most of which are correlated with high levels of drug use. These are discussed more fully below.

Genetic Factors A study of Vietnam era twins found higher concordance rates for stimulant dependence among monozygotic twins than dizygotic twins. The analyses indicated that genetic factors and unique (unshared) environmental factors contributed about equally to the development of dependence. In this study, cocaine, amphetamines, and amphetamine-like drugs were all considered stimulants.

Other Factors Social, cultural, and economic factors are powerful determinants of initial use, continued use, and relapse. Excessive use is far more likely to occur where amphetamines are readily available; this is amply demonstrated by the epidemics of amphetamine use in Japan and the United States and by more recent sharp increases in use that have followed the emergence of illicit large-scale and “kitchen” laboratories synthesizing cheap, relatively pure methamphetamine.

Since in both human and animal studies alternative positive reinforcers compete with drugs as reinforcers, the absence of such nondrug alternatives can be seen as a factor contributing to their use, especially in communities where drugs are available and the social pressures against using them are not strong. Alternative positive reinforcers are not limited to material rewards but include the kinds of psychological rewards associated with satisfactory interpersonal relationships and the self-esteem that derives from achievements in socially acceptable roles.

Learning and Conditioning Learning and conditioning are also believed to be important in perpetuating amphetamine use. Each ingestion, inhalation, or injection of the drug reinforces prior drug-taking behavior. In addition, the environmental cues associated with amphetamine use become associated with the euphoric state so that long after cessation, such cues (e.g., paraphernalia, friends who use drugs) can elicit memories that reawaken a craving for the drug. Also, other drug effects may become cues, so that the effects of alcohol, often consumed with amphetamines, can become a cue eliciting an urge to use amphetamines.

Pharmacological Factors The reinforcing and toxic effects of amphetamines and amphetamine-like drugs play an important role in the genesis of amphetamine dependence and other amphetamine-related disorders. Amphetamines produce subjective effects very similar, if not identical, to those produced by cocaine. Both categories of drugs can produce a sense of alertness, euphoria, and well-being. Performance impaired by fatigue is usually improved. There may be decreased hunger and decreased need for sleep. Patterns of toxicity are also similar, although not identical. Both the amphetamines and cocaine can induce paranoia, suspiciousness, and overt psychosis that can be difficult to distinguish from paranoid-type schizophrenia; both can produce major cardiovascular toxicities. However, the amphetamines and cocaine differ distinctly in their mechanisms of action at the cellular level, their duration of action, and their metabolic pathways.

Amphetamines enhance talkativeness, self-confidence, and sociability. Some people's beliefs about the capacity of these drugs to increase sexual drive and performance also play an important, if indirect, role in their reinforcing effects. Evidence for the enhancement of sexual performance by amphetamines is still largely

anecdotal but seems convincing to some well-trained observers. Amphetamine users, both heterosexual and homosexual, report more frequent sexual activity with more partners than heroin users.

Mechanisms of Action Although amphetamines inhibit reuptake of monoamines to a small degree, their major action is the release of monoamines from storage sites in axon terminals, which in turn increases monoamine concentrations in the synaptic cleft. The release of dopamine in the nucleus accumbens and related structures is thought to account for their reinforcing and mood-elevating effects; the release of norepinephrine is probably responsible for the cardiovascular effects. In contrast to cocaine, which binds to neurotransporters and inhibits reuptake of the neurotransmitters released into the synapse, amphetamine-like drugs are taken into the neurons where they are transported into the neurotransmitter storage vesicles. By changing the internal environment of the vesicles, the drugs cause the neurotransmitters to leak out into the cytoplasm and into the synaptic cleft. The dopamine released into the cytoplasm may undergo oxidation, which results in the production of several highly toxic and reactive chemicals (oxygen radicals, peroxides, and hydroxylquinones). Some of the neuronal toxicity of methamphetamine is due, therefore, not to the drug per se, but to the intracellular accumulation of dopamine.

Methylphenidate, widely used for the treatment of attention-deficit/hyperactivity disorder, has a mechanism of action quite distinct from that of the other amphetamine-like drugs, but is generally grouped with them. Like cocaine, methylphenidate produces actions in the central nervous system (CNS) largely by blocking the dopamine transporters responsible for the reuptake of dopamine from synapses following its release. Recent studies suggest that the relatively low abuse potential of orally administered methylphenidate is due to slow occupation of dopamine transporters in the brain: it takes about 60 minutes for an oral dose to produce peak concentrations in the brain. In clinical studies only one of seven normal adults reported a “high” after doses that produced blockade of 50 percent of the dopamine transporters—a degree of blockade comparable to that achieved with intravenous doses of cocaine. These studies and studies on other drugs show that reinforcing effects depend critically on the rate of change in dopamine concentrations in relevant brain circuits. Furthermore, unlike cocaine, which leaves the brain relatively rapidly, methylphenidate occupies the transporter sites for a much longer time.

Common Routes of Administration Amphetamines and amphetamine-like drugs can be taken orally, by injection, by absorption through nasal and buccal membranes, or by heating, inhalation of the vapors, and absorption through the pulmonary alveoli. As with nicotine, opioids, freebase cocaine, and phencyclidine (PCP), inhaled amphetamine or methamphetamine is almost immediately absorbed with a rapid onset of effects. Unlike cocaine, amphetamine and methamphetamine salts can be vaporized without much destruction of the molecule, thus obviating the need for preparing a freebase form for smoking.

As with the opioids the rapid onset of amphetamine effects from intravenous injection or inhalation produces an intensely pleasurable sensation referred to as a “rush.” The duration of the amphetamine rush has not been studied in the laboratory, but it is presumed to be shorter than the duration of elevated mood. Despite the rapid onset of action following smoked amphetamines, some users, particularly young users in Australia and Great Britain, make a transition from oral to intravenous use. Amphetamine injectors seem to be more likely than injectors of other drugs, such as heroin, to share injection equipment.

Metabolism Amphetamine and methamphetamine are extensively metabolized in the liver, but much of what is ingested is excreted unchanged in the urine. The half-lives of amphetamine and methamphetamine (weak bases) are considerably shortened when the urine is acidic. The half-life of amphetamine after therapeutic doses ranges from 7 to 19 hours and that of methamphetamine appears slightly longer. Thus, after toxic dosage, resolution of symptoms may take far longer (up to several days) with amphetamines than with cocaine, depending on the pH of the urine.

Tolerance and Sensitization Most amphetamine users who seek treatment report needing progressively more amphetamine to get the same euphoric effect; they have developed *tolerance*. Some tolerance also develops to the cardiovascular effects of amphetamine.

In animal models, chronic administration of amphetamine or amphetamine-like drugs (as well as cocaine) also produces a form of *sensitization* in which the response to a given dose is actually enhanced. One theory holds that sensitization to drug effects is attributable to a variety of kindling in the CNS. In the classic studies of kindling, electrical stimulation of the limbic system, which initially has little effect, is applied repeatedly; after a matter of days the threshold for effects decreases and major, long-lasting seizures appear. Animals show similar effects with CNS stimulants, so that repeated doses of amphetamine eventually elicit seizures or stereotyped behaviors not seen with initial doses. The sensitization can be long lasting.

The paranoid states and toxic psychoses that chronic amphetamine users commonly develop are believed to be phenomena to which sensitization develops. Those who have experienced amphetamine psychosis may do so more rapidly with subsequent exposures.

Withdrawal States Although the amphetamine withdrawal syndrome has aversive qualities (e.g., dysphoria and anhedonia), it is generally not deemed as aversive as opioid withdrawal. In most cases it is probably not as critical in perpetuating amphetamine use, although withdrawal anhedonia and fatigue may contribute to an urge to use after brief withdrawal. Postdrug-use anhedonia and dysphoria may be more important for users who have come to depend on the drugs for high energy or helping to project a confident persona, who may be temporarily unable to function without them. For others, withdrawal dysphoria may exaggerate the intensity of an antecedent mood disorder. There does not appear to be a protracted amphetamine withdrawal syndrome.

Mechanisms of CNS Changes Chronic administration of amphetamines results in several adaptive changes in the brain. For example, stimulation of dopamine receptors activates cyclic adenosine monophosphate (cAMP) within neurons in the nucleus accumbens and striatum. This activation initiates a chain of intracellular events that results in altered expression of a number of genes, some of which is mediated by phosphorylation of the transcription factor cAMP response element binding protein (CREB). One of the actions of CREB is to increase transcription of dynorphin in ribonucleic acid (RNA). This action is significant because dynorphin is a selective k-opioid agonist. k-receptor agonists inhibit release of dopamine. Recurrent collateral axons from neurons in the nucleus accumbens are thought to release dynorphin on k-receptors at dopaminergic terminals, thus dampening excessive dopaminergic activity. However, when amphetamine use is stopped and the excessive dopamine release ceases, the compensatory high levels of dynorphin persist and further diminish dopaminergic activity, thus exaggerating the anhedonia and dysphoria of amphetamine withdrawal.

Additionally, neurons of the nucleus accumbens exhibit decreases in the concentration of G_i protein (which inhibits adenylyl cyclase) and increases in levels of cAMP-dependent protein kinase. Both of these changes may persist for weeks and would be expected to upregulate the cAMP pathway. In animal models, manipulations that upregulate the cAMP pathway produce increased self-administration of cocaine (and probably of amphetamine). The persistent changes in the cAMP pathway appear to represent one mechanism of tolerance to the reinforcing effects of stimulants.

Repeated administration of amphetamine results in induction and accumulation of Fos-like proteins, chronic *fos*-related antigens (FRAs) (mediated by phosphorylation of CREB). These chronic FRAs are long-lived and are distinct from Fos-like proteins seen after a single drug exposure. In addition to persistent changes in gene transcription, repeated amphetamine administration produces persistent morphological changes in neurons of the nucleus accumbens. Glutamate transmission, which appears to play an important role in modulating the rewarding and behavior-sensitizing effects of cocaine, does not appear to be involved in these actions of amphetamine. This difference may be important, distinguishing the adaptive changes induced by these two classes of stimulants.

Other Actions Neither the actions of amphetamine nor those of amphetamine-like drugs are selective for dopamine. Amphetamine-like drugs release norepinephrine and serotonin. Some of those actions are relevant to the toxic actions of amphetamine, especially its cardiovascular toxicity.

DIAGNOSIS AND CLINICAL FEATURES

DSM-IV lists a number of amphetamine (or amphetamine-like)-related disorders (Table 11.3-1) but specifies diagnostic criteria only for amphetamine intoxication (Table 11.3-2), amphetamine withdrawal (Table 11.3-3), and amphetamine-related disorder not otherwise specified (Table 11.3-4) in the section on amphetamine (or amphetaminelike)-related disorders. The diagnostic criteria for the other amphetamine (or amphetaminelike)-related disorders are contained in the DSM-IV sections dealing with the primary phenomenological symptom (e.g., psychosis).

Amphetamine use disorders
Amphetamine dependence
Amphetamine abuse
Amphetamine-induced disorders
Amphetamine intoxication
Specify if: with perceptual disturbances
Amphetamine withdrawal
Amphetamine intoxication delirium
Amphetamine-induced psychotic disorder, with delusions
Specify if: with onset during intoxication
Amphetamine-induced psychotic disorder, with hallucinations
Specify if: with onset during intoxication
Amphetamine-induced mood disorder
Specify if: with onset during intoxication/with onset during withdrawal
Amphetamine-induced anxiety disorder
Specify if: with onset during intoxication
Amphetamine-induced sexual dysfunction
Specify if: with onset during intoxication
Amphetamine-induced sleep disorder
Specify if: with onset during intoxication/with onset during withdrawal
Amphetamine-related disorder not otherwise specified

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.3-1 DSM-IV Amphetamine (or Amphetamine-like)-Related Disorders

A. Recent use of amphetamine or a related substance (e.g., methylphenidate).
B. Clinically significant maladaptive behavioral or psychological changes (e.g., euphoria or affective blunting; changes in sociability; hypervigilance; interpersonal sensitivity; anxiety, tension, or anger; stereotyped behaviors; impaired judgment; or impaired social or occupational functioning) that developed during, or shortly after, use of amphetamine or a related substance.
C. Two (or more) of the following, developing during, or shortly after, use of amphetamine or a related substance: (1) tachycardia or bradycardia (2) pupillary dilation (3) elevated or lowered blood pressure (4) perspiration or chills (5) nausea or vomiting (6) evidence of weight loss (7) psychomotor agitation or retardation (8) muscular weakness, respiratory depression, chest pain, or cardiac arrhythmias (9) confusion, seizures, dyskinesias, dystonias, or coma
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.
Specify if: With Perceptual Disturbances

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.3-2 DSM-IV Diagnostic Criteria for Amphetamine Intoxication

A. Cessation of (or reduction in) amphetamine (or a related substance) use that has been heavy and prolonged.
B. Dysphoric mood and two (or more) of the following physiological changes, developing within a few hours to several days after criterion A: (1) fatigue (2) vivid, unpleasant dreams (3) insomnia or hypersomnia (4) increased appetite (5) psychomotor retardation or agitation
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounting for by another mental disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.3-3 DSM-IV Diagnostic Criteria for Amphetamine Withdrawal

The amphetamine-related disorder not otherwise specified category is for disorders associated with the use of amphetamine (or a related substance) that are not classifiable as amphetamine dependence, amphetamine abuse, amphetamine intoxication, amphetamine withdrawal, amphetamine intoxication delirium, amphetamine-induced psychotic disorder, amphetamine-induced mood disorder, amphetamine-induced anxiety disorder, amphetamine-induced sexual dysfunction, or amphetamine-induced sleep disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.3-4 DSM-IV Diagnostic Criteria for Amphetamine-Related Disorder Not Otherwise Specified

Amphetamine Use Disorders The DSM-IV generic criteria for dependence and abuse are applied to amphetamine and related substances (see Table 11.1-3 and Table 11.1-8). Depending on the dose, the route of administration, and the pattern of use, amphetamine dependence has quite variable effects on behavior, the capacity to work, and toxic consequences. With relatively low doses taken orally, behavior may be within normal limits and dependence is manifested only by the fatigue and depressive symptoms that ensue when drug use is interrupted and by the effort devoted to ensuring a supply. With higher doses, in addition to preoccupation with getting the drug, there is often hyperactivity, restlessness, bruxism, hypertalkativeness, irritability and short-tempered behavior, decreased sleep, and decreased appetite often accompanied by weight loss. Generally, mood is elevated; the amphetamine user is gregarious and may express confidence, even some grandiosity. With very high doses and intravenous or pulmonary routes behavior and judgment can be severely disrupted, dependence can develop quickly, and the likelihood of developing toxic paranoid states is high. There may also be repetitive behaviors that appear to have no rational basis, such as taking objects apart or rearranging objects. Such behaviors are probably analogous to the stereotypy seen when animals are repeatedly dosed with amphetamine. Severe aggressive behavior is uncommon, but it may occur during episodes of intoxication or during amphetamine-induced psychosis.

Likelihood of Progression Patients with narcolepsy and children with attention-deficit/hyperactivity disorder can take amphetamine-like drugs or methylphenidate daily for many years without developing significant tolerance to their therapeutic effects and with little escalation of dose or toxicity. When amphetamine and amphetamine-like drugs were more widely used in the treatment of obesity, relatively few patients who took them daily developed dependence. Even when amphetamine-like drugs are taken for nonmedical reasons (e.g., to reduce fatigue or for euphorigenic effects), not all users progress to abuse or dependence. Although the absolute risk of such progression is not precisely known, all estimates suggest that it is high enough to justify a policy that discourages experimentation. One estimate of risk comes from a classic study, carried out in 1974 and published in 1976, of drug use among a representative sample of young men. Seventy-three percent reported having had no experience with amphetamines, but of the 27 percent who had some experience, almost 10 percent (3 percent of the total) reported daily use. Findings from the NCS conducted in the early 1990s were remarkably similar. About 15 percent of interviewees had used a stimulant other than cocaine for extramedical reasons. Of these users, 11.2 percent had become dependent on them (DSM-III-R criteria) by the time of the interview.

Varied Patterns of Use There are several patterns of abuse of amphetamines and similar agents. Some persons may use the drugs intermittently in relatively low doses; for example, truck drivers or students may use them to overcome fatigue or the need for sleep or to derive some positive mood effects. Some intermittent users become dependent and find it difficult to stop; some may eventually escalate the dosage. Since the drugs are no longer available legitimately for these purposes, persons with that pattern of use are likely to obtain them from illicit sources.

Some persons use amphetamines primarily to induce euphoria. Such users often progress to high dosages, especially if they use the drugs intravenously or by

inhalation. These are obviously the most dangerous patterns of use, and they commonly lead to compulsive use or toxic effects. Although intravenous use initially may be intermittent, with days or weeks elapsing between episodes, such high-dose use often progresses to sprees or speed runs, during which several grams of amphetamine might be smoked or injected. The runs can last for days or weeks and are commonly punctuated by episodes of toxicity (amphetamine-induced psychotic disorder with delusions or amphetamine intoxication delirium) or by brief periods of abstinence (crashing), generally precipitated by an interruption in the supply of the drug or exhaustion. Some clinicians have observed that in contrast to cocaine users who prefer to smoke cocaine and use in binges interrupted by periods of cocaine abstinence, methamphetamine users are more likely to use on a daily basis and tend to change routes of administration because the drug is irritating to the nasal mucosa and lungs.

High-dose amphetamine users often combine amphetamine with sedatives, benzodiazepines, or opioids to modulate the stimulant effects. Alcohol use and alcohol abuse are common concomitants of high-dose amphetamine abuse and dependence. Methamphetamine is sometimes used to reduce the sedating effects of alcohol and facilitate and prolong socializing and sexual activity. Some observers believe that methamphetamine use increases the likelihood of multiple sex partners and the transmission of human immunodeficiency virus (HIV). In a study of gay and bisexual men who were methamphetamine injectors (and were not seeking treatment), 54 percent reported sharing needles within the preceding 30 days, and 74 percent reported exchanging sex for money or drugs. The recent increase in the use of methamphetamine in California and several western states has been predominantly among white men 25 to 34 years old.

Comorbidity The frequent co-occurrence of other psychiatric disorders and amphetamine dependence was first noted in the 1950s. The presence of other psychiatric disorders sharply increases the odds of drug dependence in general, and drug-dependent persons are more likely than the general population to meet the criteria for additional psychiatric disorders.

Patients with schizophrenia commonly use amphetamine or cocaine and develop both dependence and toxic syndromes. It has been suggested that schizophrenia patients use stimulants to alleviate negative symptoms or adverse effects of antipsychotic agents. Special programs involving peer-based support groups seem to be effective in linking drug-using schizophrenia patients with outpatient treatment programs.

Amphetamine-Induced Disorders All of the disorders listed in DSM-IV for cocaine (intoxication, psychotic disorder, intoxication delirium, mood disorder, anxiety disorder, sleep disorder, and sexual dysfunction) may occur in association with the use of amphetamine or amphetamine-like drugs. The clinical pictures are similar, if not identical; the DSM-IV diagnostic criteria and codes are identical except for substitution of the word “amphetamine” for the word “cocaine.”

Amphetamine Intoxication The intoxication syndromes of cocaine and amphetamines are similar. Amphetamine intoxication can occur as a result of single doses in non-tolerant individuals, but it is most commonly seen in those who are amphetamine-abusers or are dependent. Some of the manifestations are exaggerated effects of the drug, including euphoria, grandiosity, restlessness, hypervigilance, talkativeness, and stereotyped repetitive behaviors. The patient is generally oriented to time, place, and situation. However, intoxication may be accompanied by visual and tactile hallucinations or illusions. Generally, patients recognize that the symptoms are drug induced. When they do not, a diagnosis of amphetamine-induced psychotic disorder should be considered. Symptoms of amphetamine intoxication usually resolve as the drug is excreted over a period of 24 to 48 hours.

In DSM-IV, the diagnostic criteria for amphetamine intoxication (Table 11.3-2) and cocaine intoxication (see Table 11.6-2) are separated but are virtually the same. DSM-IV allows for noting the presence of perceptual disturbances as a symptom of amphetamine intoxication.

Amphetamine Withdrawal The severity of the amphetamine withdrawal syndrome is presumably related to the intensity and duration of the antecedent drug use. Some elements of the syndrome (dysphoria and fatigue) can be seen after relatively brief binges or “runs” of only a few days, with some less severe aspects of “crashing” reported to occur even after 24 hours of use. During phases of the amphetamine withdrawal syndrome users may experience severe depression that tends to resolve without special treatment when sleep normalizes.

Amphetamine users stabilized on amphetamine prior to withdrawal have been studied. Among the findings noted as early as 1963 were a marked shortening of time to first rapid eye movement (REM) sleep and a marked rebound in total REM sleep. A return to normal levels in some cases required several weeks.

The DSM-IV criteria for amphetamine withdrawal and those for cocaine withdrawal are identical; the ICD-10 criteria are virtually identical. Less is known about the later stages of amphetamine withdrawal, but it is likely that there are periods of increased vulnerability when stimuli previously associated with use elicit memories of drug effects and craving.

Amphetamine-Induced Psychotic Disorder and Intoxication Delirium Although amphetamine-induced psychotic disorder or intoxication delirium are usually seen only when high doses are used for a long time, such syndromes have been reported in apparently vulnerable persons even after therapeutic doses given for a short time. Haloperidol (Haldol) and phenothiazines have been used to treat the psychotic syndrome. Although with cocaine the delusional syndrome is typically of short duration, with the amphetaminelike drugs it may not resolve for many days after drug cessation. Following recovery from either a psychotic or delirium syndrome there may be amnesia for the entire episode or some part of it.

Psychiatrists in Japan have presented data showing that amphetamine-induced psychosis may persist for several years and that in the acute stage there may be disturbance of consciousness (confusion, disorientation) in addition to the more typical mood and delusional symptoms. Following recovery persons who have experienced an amphetamine-induced psychosis seem to be sensitized and will experience acute paranoid psychosis on reexposure to small doses of amphetamines, and some have exacerbations in response to stress.

Rhesus monkeys show hallucinatory-like behaviors and stereotypies in response to low-dose amphetamine challenge as long as 28 months after a 12-week course of low-dose amphetamine exposure.

Amphetamine-Induced Mood Disorder According to DSM-IV, the onset of amphetamine-induced mood disorder can occur during intoxication or withdrawal (see Table 14.6-18). In general, intoxication is associated with manic or mixed mood features, whereas withdrawal is associated with depressive mood features. The manic and hypomanic symptoms often seen during amphetamine use rarely (if ever) persist beyond the period of drug use, but hypophoria, depressive, and anhedonic symptoms that persist well beyond the period of withdrawal are not uncommon. Patients may seek treatment for such persisting symptoms. In such situations the clinician should consider a diagnosis of amphetamine-induced mood disorder. However, it is often difficult to distinguish a substance-induced mood disorder from a primary mood disorder, especially in patients who have a history of depressive symptoms antedating the onset of amphetamine use. Given the pharmacology of amphetamine it is possible that drug-induced changes could aggravate and intensify a primary depressive disorder.

Amphetamine-Induced Anxiety Disorder In DSM-IV, the onset of amphetamine-induced anxiety disorder can also occur during intoxication or withdrawal (see Table 15.6-18). Amphetamine, like cocaine, can induce symptoms similar to those seen in obsessive-compulsive disorder, with repetitive, stereotyped behaviors. However, these ordinarily do not persist beyond the period of drug intoxication and rarely merit a distinct diagnosis. Amphetamine-like drugs can also induce panic attacks in individuals with no previous history of panic attacks. When such episodes persist well beyond the period of drug use and require clinical attention a distinct diagnosis should be entertained.

Amphetamine-Induced Sexual Dysfunction Although amphetamine is often used to enhance sexual experiences, high doses and long-term use are associated with impotence and other sexual dysfunctions. These dysfunctions are classified in DSM-IV as amphetamine-induced sexual dysfunction with onset during intoxication (see Table 19.1a-20).

Amphetamine-Induced Sleep Disorder The diagnostic criteria for amphetamine-induced sleep disorder with onset during intoxication or withdrawal are found in the DSM-IV section on sleep disorders (see Table 21-18). Amphetamine use can produce insomnia and sleep deprivation; persons undergoing amphetamine withdrawal can experience hypersomnolence and nightmares. However, unless these disturbances persist beyond the period of drug use or well beyond withdrawal and are severe enough to merit clinical attention they do not require a separate diagnosis.

Disorder Not Otherwise Specified If an amphetamine (or amphetamine-like)-related disorder does not meet the criteria of one or more of the previous discussed categories, it can be diagnosed as an amphetamine-related disorder not otherwise specified (Table 11.3-4). With the increasing illicit use of designer amphetamines, syndromes may arise that do not meet the criteria outlined in DSM-IV and that necessitate the frequent use of the not otherwise specified category.

Toxicity and Complications Subjects in a survey of amphetamine users in Australia reported various physical and psychological problems that they attributed to

amphetamine use. Commonly reported physical symptoms were tiredness (89 percent), loss of appetite (85 percent), dehydration (73 percent), and jaw clenching (73 percent). Also reported were headaches, muscle pains, shortness of breath, and tremors. The most frequently reported psychological symptoms were mood swings (80 percent), sleep problems (78 percent), anxiety, difficulty concentrating, depression, and paranoia (each about 70 percent), hallucinations, and episodes of aggression and violence (each about 45 percent). Daily use, heavier use, and being male and unemployed were correlated with reporting more symptoms.

Amphetamines produce their most dramatic toxic effects on the CNS and the cardiovascular system. In animal models (rodents and primates), chronically administered high doses of amphetamines produce long-lasting depletion of brain norepinephrine, and more selective but even longer-lasting depletion of dopamine, alterations in dopamine uptake sites, and reduction in serotonergic activity. Because these effects involve damage to both axons and axon terminals, largely sparing the cell bodies, it is not known to what degree these long-lasting effects are permanent. Methamphetamine in particular seems capable of inducing damage to serotonergic fibers, but the noradrenergic system is largely unaffected. These effects may be due to toxic biotransformation products of excessive dopamine within the neuron. The long-lasting dopaminergic changes probably account for the altered, elevated threshold for self-stimulation in animals and the anhedonia reported by chronic amphetamine users for prolonged periods following cessation. It is not known to what degree methamphetamine use causes neuronal damage in humans, but data from patients with Parkinson's disease suggest that there must be considerable damage to the dopaminergic pathways before it becomes evident in function and behavior. In monkeys the toxic effects of chronic amphetamine use include damage to cerebral blood vessels, neuronal loss, and microhemorrhages. In humans high doses of amphetamine have also been associated with lethal hyperpyrexia and with destructive deterioration of arterioles. High doses can also produce convulsions and ultimately coma and death.

Amphetamine-like drugs can cause catastrophes of the cardiovascular system (e.g., intracranial hemorrhage, arrhythmias, and acute cardiac failure) because of their capacity to release norepinephrine, dopamine, and serotonin, and to raise blood pressure. With amphetamines, considerable tolerance develops to the effect on blood pressure. The likelihood of such cardiovascular effects is related to dose and the rapidity with which the drug is absorbed. The use of methamphetamine by smoking or intravenous injection is likely to result in greater cardiovascular toxicity. Amphetamine-induced hyperthermia and free radical formation are believed to be involved in causing rhabdomyolosis and the consequent renal tubular obstruction that is occasionally reported. Because amphetamine use can be associated with increased sexual activity, often accompanied by poor judgment, amphetamine users are at increased risk for venereal diseases, including infection with HIV.

PATHOLOGY AND LABORATORY EXAMINATIONS

Amphetamine and amphetaminelike drugs can be detected for varying lengths of time in urine (several days, depending on dose and sensitivity of the method). Metabolites can also be detected in blood, saliva, and hair. Blood and saliva furnish a better index of current levels, whereas urine provides a longer window of opportunity for detecting use over the previous few days. Hair analysis can reveal drug use over a period of weeks to months but has little applicability in clinical situations.

Such procedures as positron emission tomography (PET) and single photon emission computed tomography (SPECT) have not yet been used during the immediate postamphetamine cessation period, but given the findings seen at autopsy and the pharmacology of amphetaminelike agents, it would not be surprising if there were arteriolar pathology and alterations in dopamine systems.

DIFFERENTIAL DIAGNOSIS

The disorders associated with the use of amphetamine and amphetamine-like drugs need to be distinguished from both primary mental disorders and disorders induced by other classes of drugs. A history of the drug ingestion is important for making these distinctions. However, given the unreliability of self-reports about drug use and the likelihood that many users will deny any drug use at all, laboratory testing for drugs in body fluids and histories from collaterals are very important. Disorders associated with amphetamine use cannot be easily distinguished from those associated with cocaine except by a reliable history or laboratory tests. Users of amphetamine or cocaine and related drugs may exhibit inappropriate optimism, euphoria, and expansiveness; excessive talkativeness; and a decreased need for sleep sometimes associated with irritability in the context of a clear sensorium, a pattern that is also observed in manic and hypomanic episodes of bipolar I disorder and bipolar II disorder, respectively. Those symptoms, however, may not be obvious enough to suggest their relation to drug use, and the first indication of drug dependence may be financial difficulties, an arrest for selling drugs or possessing them, or some drug-induced toxicity.

Intoxication Amphetamine intoxication is diagnosed when the effects of the drug exceed the mood-elevating effects that users typically seek when they use amphetamines. The diagnosis of intoxication would be appropriate when the drug effects are problematic enough to require differentiation from hypomanic or manic behavior.

Amphetamine and amphetamine-like drug intoxication can also be confused with PCP intoxication, although the latter is usually associated with nystagmus, motor incoordination, and some cognitive impairment. Endocrine disorders (such as Cushing's disease) and the excessive use of steroids should also be considered.

Psychotic Disorders Amphetamine-induced toxic psychosis can be exceedingly difficult to differentiate from schizophrenia and other psychotic disorders characterized by hallucinations or delusions. Paranoid delusions occur in about 80 percent of patients, and hallucinations in 60 to 70 percent. Consciousness is clear and disorientation is uncommon. The presence of vivid visual or tactile hallucinations should raise suspicion of a drug-induced disorder. In areas and populations where amphetamine use is common it may be necessary to provide only a provisional diagnosis until the patient can be observed and drug test results are obtained. Even then, there may be difficulties because in some urban areas a high percentage of persons with established diagnoses of schizophrenia also use amphetamines or cocaine. Typically symptoms of amphetamine psychosis remit within a week, but in a small proportion of patients, psychosis may last for more than a month.

Anxiety Disorders Amphetamine-induced anxiety disorder must also be distinguished from panic disorder and generalized anxiety disorder.

Other Symptoms The symptoms that may emerge during withdrawal—depression, dysphoria, anhedonia, disturbed sleep—need to be distinguished from those of primary mood disorders and primary sleep disorders. Unless the symptoms are more intense or more prolonged than is typical of amphetamine withdrawal and require independent treatment, the diagnosis should be limited to withdrawal rather than amphetamine-induced mood disorder. When a diagnosis of amphetamine-induced mood disorder is made, one must specify whether its onset was during intoxication or in withdrawal. It is also possible to specify the subtype of mood disorder (with depressive, manic, or mixed features). In differentiating amphetamine-induced mood disorder from the primary mood disorder the critical factor is the clinician's judgment that the mood disorder was caused by the drug. An amphetamine-induced mood disorder or mood disorder with onset during intoxication or withdrawal usually remits within a week or two. It is appropriate, therefore, to withhold judgment about the diagnosis during the early phase of withdrawal. If depressed mood and related symptoms persist beyond a few weeks, the possibility of alternative causes should be entertained. In considering diagnostic possibilities the clinician should consider the age of the patient at the onset of symptoms and a history of episodes of mood disorder that developed before the onset of drug use or during any long intervals when there was no significant drug abuse.

COURSE AND PROGNOSIS

The natural history of amphetamine dependence in the United States is less well documented than that of opioids or cocaine. Some researchers believe that some intravenous amphetamine users in the 1960s moved on to heroin use in the 1970s. However, it seems likely that many whose use was less severe simply stopped or recovered, whereas others intensified their use of alcohol. Japanese clinicians believe that some amphetamine users may develop persistent psychosis and that those who recover remain at high risk of reexperiencing psychosis if they use amphetamines again (sensitization).

A 3- to 8-year follow-up study of 110 methamphetamine users hospitalized for drug-related problems in Japan in the 1980s found that 12 former patients had died, a mortality rate 11 times that of age- and sex-matched general population controls. However, 56 percent of those still alive had not used amphetaminelike drugs in the year before interview, and most of them also showed improvements in work and family relationships. Twenty-five percent were thought to have highly or moderately unfavorable outcomes in terms of drug use, work, and family relationships. Findings from treatment programs in California suggest that the course and prognosis for amphetamine dependence are probably similar to those for cocaine dependence. The prognosis for Japanese convicted and imprisoned for crimes related to stimulant drug use seems as bleak as that in the United States; 58 percent committed crimes within 1 year after release and 98 percent committed crimes within 5 years.

TREATMENT

There are no specific, well-established treatments for dependence on amphetamine or amphetamine-like drugs and few controlled studies on the treatment of amphetamine dependence. Most casual users do not need or seek treatment. Those with moderately severe dependence obtain treatment in a variety of settings (mostly outpatient drug free) that were not designed specifically to treat amphetamine dependence. The most severe cases, those within the criminal justice system, and the homeless generally drop out of outpatient treatment because of their complex needs or are unable to access treatment at all. Some data on treatment seeking and outcome are available for the state of California where amphetamine abuse began to increase in the early 1990s. In general, methamphetamine users in the state gave reasons for entering treatment similar to those of other substance users—personal motivation (69 percent) and pressure from the criminal justice system (31 percent) and from family or other significant persons (22 percent). Treatment received ranged from residential and ambulatory detoxification to day treatment, 12-step activities, and case management. Although those leaving 12-step programs were more likely to have been considered treatment completers, at follow-up, patients reported similar reductions in drug use (unverified by urine tests) regardless of the treatment received and were neither more nor less successful in this respect than those receiving treatment for heroin, cocaine, or marijuana use. One program that provided a highly structured and manualized cognitive-behavioral treatment making use of a combination of group and individual counseling initially developed for crack cocaine users found that this type of treatment produced equal levels of participation in treatment and equally good outcomes for those dependent on methamphetamine.

A wide variety of pharmacological agents have been explored as adjuncts to, or major elements in, the treatment of amphetamine dependence. Some have been studied in controlled trials. Virtually all of these agents had been previously tried in the treatment of cocaine dependence and produced comparably disappointing results. For example, although imipramine (Tofranil) (150 mg a day) improved treatment retention, it had no significant effect on methamphetamine use. Although an open-label trial of fluoxetine (Prozac) (20 mg a day) was reported to be useful in amphetamine dependence, success in dozens of open-label trials with cocaine-dependent patients has rarely been confirmed when the same agents were studied in double-blind controlled trials.

In Europe and Australia the ethics and efficacy of prescribing oral amphetamines for amphetamine users are hotly debated. This practice is allowed in Great Britain, although it varies from region to region, and virtually no safeguards exist against diversion of prescribed amphetamines to the illicit market. No outcome studies exist that evaluate the values and risks of this practice.

Selection of Treatment Setting The general principles of treatment for amphetamine dependence are not very different from those for cocaine and opioid dependence, but there are fewer replicated studies on the efficacy of any particular treatment approach. As with cocaine and opioid dependence, amphetamine dependence severe enough to require formal treatment is often associated with other psychiatric diagnoses.

Patient heterogeneity requires thoughtful selection among available alternatives. Not all amphetamine users require extensive treatment; some users who are not dependent respond to external pressures, as when employers insist on careful monitoring of drug use. The executive with little history of psychopathology, a supportive social network, economic assets, and personal skills has a different prognosis and a wider range of options than does a patient who is unemployed, alienated from the family, and perhaps also using opioids. Severe depression, psychotic manifestations beyond the initial withdrawal period, and drug use that is completely out of control (i.e., repeated failure to respond to outpatient efforts) seem to be the major accepted criteria for hospitalization.

OTHER AGENTS

Substituted Amphetamines MDMA (“Ecstasy,” “Adam”) is one of a series of substituted amphetamines that also includes 3,4-methylenedioxyethylamphetamine (MDEA, “Eve”), 3,4-methylenedioxyamphetamine (MDA), 2,5-dimethoxy-4-bromoamphetamine (DOB), paramethoxyamphetamine (PMA), and others. These drugs produce subjective effects resembling those of amphetamine and lysergic acid diethylamide (LSD), and in that sense, MDMA and similar analogues may represent a distinct category of drugs (“entactogens”).

A methamphetamine derivative that came into use in the 1980s, MDMA was not technically subject to legal regulation at the time. Although it has been labeled a designer drug in the belief that it was deliberately synthesized to evade legal regulation, it was actually synthesized and patented in 1914. Several psychiatrists used it as an adjunct to psychotherapy and concluded that it was of value. At one time it was advertised as legal and was used in psychotherapy for its subjective effects. However, it was never approved by the FDA. Its use raised questions of both safety and legality, since the related amphetamine derivatives MDA, DOB, and PMA had caused a number of overdose deaths, and MDA was known to cause extensive destruction of serotonergic nerve terminals in the CNS. Using emergency scheduling authority, the Drug Enforcement Agency made MDMA a Schedule I drug under the CSA, along with LSD, heroin, and marijuana. Despite its illegal status MDMA continues to be manufactured, distributed, and used in the United States, Europe, and Australia. Its use is common in Australia and Great Britain at extended dances (“raves”) popular with adolescents and young adults.

Mechanisms of Action The unusual properties of the drugs may be a consequence of the different actions of the optical isomers: the *R*(–) isomers produce LSD-like effects and the amphetamine-like properties are linked to *S*(+) isomers. The LSD-like actions, in turn, may be linked to the capacity to release serotonin. The various derivatives may exhibit significant differences in subjective effects and toxicity. Animals in laboratory experiments will self-administer the drugs, suggesting prominent amphetamine-like effects.

Subjective Effects After taking usual doses (100 to 150 mg), MDMA users experience elevated mood and, according to various reports, increased self-confidence and sensory sensitivity; peaceful feelings coupled with insight, empathy, and closeness to people; and decreased appetite. Difficulty in concentrating and an increased capacity to focus have both been reported. Dysphoric reactions, psychotomimetic effects, and psychosis have also been reported. Higher doses seem more likely to produce psychotomimetic effects. Sympathomimetic effects of tachycardia, palpitation, increased blood pressure, sweating, and bruxism are common. The subjective effects are reported to be prominent for about 4 to 8 hours, but they may not last as long or may last longer, depending on the dose and route of administration. The drug is usually taken orally, but it has been snorted and injected. Both tachyphylaxis and some tolerance are reported by users.

The acute adverse effects reported include precipitation of episodes of panic and anxiety. More-severe brief psychiatric disturbances can also occur, and preexisting pathology does not appear to be a requisite for severe reactions. A healthy drug-free subject, known to be without personal and family psychiatric illness was given a 140-mg dose of the drug and developed a psychosis lasting 2½ hours that included vivid auditory and visual hallucinations and a belief that people were making noise to annoy him intentionally.

Following the acute effects of MDMA there may be a combination of some diminishing residual effects gradually superseded by feelings of drowsiness, fatigue, depression, and difficulty concentrating, somewhat comparable to the crash after cessation of amphetamine use. When young adults who were Saturday night MDMA users were compared with alcohol-only users who frequented the same club, the MDMA users reported elevated mood on the following day but feelings of depression (Beck Depression Inventory scores of about 12) by the fifth day. In contrast, alcohol-only users showed relatively little mood change over the 5-day period; their highest Beck depression scores (about 8) occurred on the second day. In a double-blind placebo-controlled study of normal volunteers given 1.7 mg per kg of body weight of MDMA, some subjects continued to report symptoms typical of MDMA actions (suppressed appetite, jaw clenching, restlessness, heaviness in the legs, difficulty concentrating) 24 hours later. More-persistent neuropsychiatric adverse effects associated with MDMA use include anxiety, depression, flashbacks, irritability, panic disorder, psychosis, and memory disturbance.

Toxicity Although it is not as toxic as MDA, various somatic toxicities attributable to MDMA use have been reported, as well as fatal overdoses. It does not appear to be neurotoxic when injected into the brain of animals, but it is metabolized to MDA in both animals and humans. In animals MDMA produces selective, long-lasting damage to serotonergic nerve terminals. It is not certain if the levels of the MDA metabolite reached in humans after the usual doses of MDMA suffice to produce lasting damage. Nonhuman primates are more sensitive than are rodents to MDMA's toxic effects and show more prolonged or permanent neurotoxicity at doses not much higher than those used by humans (Fig. 11.3-1). Users of MDMA show differences in neuroendocrine responses to serotonergic probes, and studies of former MDMA users show global and regional decreases in serotonin transporter binding, as measured by positron emission tomography. Although psychological assessment of a small sample of users did not reveal evidence of current anxiety or a mood disorder, eight of nine subjects had at least some impairment on at least one test of neuropsychological function.

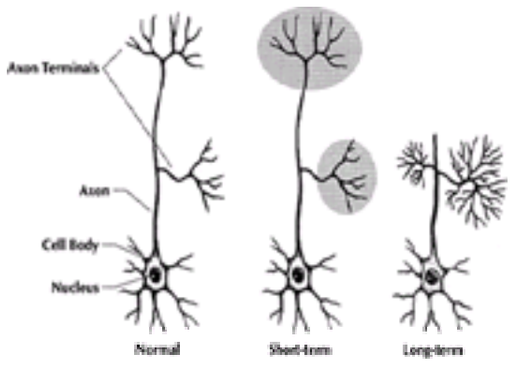

FIGURE 11.3-1 Neurotoxic effects of MDMA. MDMA damages serotonin-producing neurons in the brains of nonhuman primates. The *left illustration* shows a normal neuron. The shaded area in the *middle illustration* shows the axon terminals of the neurons that are damaged by MDMA. The *right illustration* shows how, 12 to 18 months after being damaged by MDMA, serotonin-producing nerve fibers have regrown excessively in some areas and not at all in others. (Reprinted with permission from Mathias R: Like methamphetamine, "ecstasy" may cause long-term brain damage. NIDA Notes *11:*7, 1996.)

Other reported toxicities include arrhythmias, cardiovascular collapse, hyperthermia, rhabdomyolysis, disseminated intravascular coagulation, acute renal failure, and hepatotoxicity. The role that contaminants in illicit MDMA played in the toxic reactions is uncertain, but significant elevations of blood pressure and temperature have been observed after administration of pure MDMA.

MDMA dependence does not appear to be a significant problem in the United States, but cases of dependence have been reported in England. Among a sample of MDMA users in Australia, 28 percent reported that problems related to the use of the drug were mostly acute reactions, such as panic, paranoia, loss of reality, and hallucinations. Only 2 percent reported feeling dependent (needing to use it every day to cope), but 22 percent claimed that they knew someone who had been dependent, and 47 percent believed it was possible to become addicted. MDA, MDMA, PMA, and MDEA have all been linked to psychosis and overdose deaths. The toxic manifestations of overdose include restlessness, agitation, sweating, rigidity, high blood pressure, tachycardia, hyperpyrexia, and convulsions. Chlorpromazine (Thorazine) prevented lethality in dogs, but there are no clinical reports of its use for this purpose in humans.

There are currently no established clinical uses for MDMA, although before its regulation, there were several reports of its beneficial effects as an adjunct to psychotherapy.

Khat The fresh leaves of *Catha edulis*, a bush native to East Africa, have been used as a stimulant in the Middle East, Africa, and the Arabian peninsula for at least 1000 years. Khat is still widely used in Ethiopia, Kenya, Somalia, and Yemen. The amphetamine-like effects of khat have long been recognized, and although efforts to isolate the active ingredient were first undertaken in the nineteenth century, only since the 1970s a has cathinone (*S*(–)a-aminopropriophenone or *S*-2-amino-1-phenyl-1-propanone) been identified as the substance responsible. Cathinone is a precursor moiety that is normally enzymatically converted in the plant to the less active entities norephedrine and cathine (norpseudoephedrine), which explains why only the fresh leaves of the plant are valued for their stimulant effects. Cathinone has most of the CNS and peripheral actions of amphetamine and appears to have the same mechanism of action. In humans it elevates mood, decreases hunger, and alleviates fatigue. Like amphetamine, it is self-administered by laboratory animals and produces increased locomotor activity and stereotypy. At high doses it can induce an amphetamine-like psychosis in humans. Because it is typically absorbed buccally after chewing the leaf and because the alkaloid is metabolized relatively rapidly, high toxic blood levels are not frequently reached. Concern about khat use is linked to its dependence-producing properties rather than to its acute toxicity. It is estimated that five million doses are consumed each day, despite prohibition of its use in a number of African and Arab countries.

In the 1990s several clandestine laboratories began synthesizing methcathinone, a drug with actions quite similar to those of cathinone. Known by a number of street names (e.g., CAT, goob, and crank), its popularity is due primarily to its ease of synthesis from ephedrine or pseudoephedrine, which were readily available until placed under special controls. Methcathinone has been moved to Schedule (Control Level) I of the CSA. The patterns of use, adverse effects, and complications are quite similar to those reported for amphetamine.

SUGGESTED CROSS-REFERENCES

The neural sciences are presented in Chapter 1 and neuropsychiatry and behavioral neurology in Chapter 2. A classification of mental disorders appears in Chapter 9. An introduction to and overview of substance-related disorders is presented in Section 11.1, cocaine-related disorders in Section 11.6, and various drugs in Chapter 31 on biological therapies, particularly sympathomimetics in Section 31.27. Schizophrenia is discussed in Chapter 12, other psychotic disorders in Chapter 13, and attention-deficit disorders in Section 39.1

SECTION REFERENCES

Anthony JC, Warner LA, Kessler RC: Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. Exp Clin Psychopharmacol *2:*244, 1994.

Battaglia G, Napier TC: The effects of cocaine and the amphetamines on brain and behavior: A conference report. Drug Alcohol Dependence *52:*41, 1998.

*Castner SA, Goldman-Rakic PS: Long-lasting psychotomimetic consequences of repeated low-dose amphetamine exposure in rhesus monkeys. Neuropsychopharmacology *20:*10, 1999.

*Center for Substance Abuse Treatment: *Proceedings of the National Consensus Meeting on the Use, Abuse and Sequelae of Abuse of Methamphetamine with Implications for Prevention, Treatment and Research*. DHHS publ. no. SMA 96-8013. Substance Abuse and Mental Health Services Administration, Rockville, MD, 1997.

Curran HV, Travill RA: Mood and cognitive effects of +/–3,4-methylenedioxymethamphetamine (MDMA, "ecstasy"): Weekend "high" followed by mid-week low. Addition *92:*821, 1997.

*Gawin FH, Ellinwood EH: Cocaine and other stimulants. N Engl J Med *318:*1173, 1988.

*Gorelick DA: Pharmacologic therapies for cocaine and other stimulant addiction. In *Principles of Addiction Medicine*, ed 2, AW Graham, TK Schultz, editors. American Society of Addiction Medicine, Chevy Chase, MD, 1998.

Gruber AJ, Pope HG Jr: Psychiatric and medical effects of anabolic-androgenic steroid use in women. Psychother Psychosom, in press.

Hall W, Hando J: Patterns of amphetamine use in Australia. In *Amphetamine Misuse. International Perspectives on Current Trends*, H Klee, editor. Harwood Academic Publishers, Reading, Australia, 1997.

*Hall W, Hando J, Darke S, Ross J: Psychological morbidity and route of administration among amphetamine users in Sydney, Australia. Addiction *91:*81, 1996.

Hyman SE, Nestler EJ: Initiation and adaptation: A paradigm for understanding psychotropic drug action. Am J Psychiatry *153:*151, 1996.

*Jaffe JH: Drug addiction and drug abuse. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, ed 8, AG Gilman, TW Rall, AS Nies, P Taylor, editors. Pergamon, New York, 1990.

Jansen KLR: Ecstasy (MDMA) dependence. Drug Alcohol Depend *53:*121, 1999.

Johnston LD, Bachman JG, O'Malley PM: *Monitoring the Future Study*. University of Michigan, Ann Arbor, 1997.

*Kalant OJ: *The Amphetamines*. Charles C Thomas, Springfield, IL, 1973.

Kalix P: Pharmacological properties of the stimulant khat. Pharmacol Ther *48:*397, 1990.

Kandel DB, Davies M: High school students who use crack and other drugs. Arch Gen Psychiatry *53:*71, 1996.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshelman S, Wittchen H-U, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry *51:*8, 1994.

McCann UD, Szabo Z, Scheffel U, Dannals RF, Ricaurte GA: Positron emission tomographic evidence of toxic effects of MDMA ("Ecstasy") on brain serotonin neurons in human beings. Lancet *352:*1433, 1998.

*Meng Y, Dukat M, Bridgen DT, Martin BR, Lichtman AH: Pharmacological effects of methamphetamine and other stimulants via inhalation exposure. Drug Alcohol Depend *53:*111, 1999.

Office of Applied Studies: *Preliminary Results from the 1997 National Household Survey on Drug Abuse*. National Household Survey on Drug Abuse Series: H6, DHHS publ. no. SMA 98-3251. Substance Abuse and Mental Health Services Administration, Rockville, MD, 1998.

Pope HG Jr, Kouri EM, Hudson JI: The effects of supraphysiologic doses of testosterone on mood and aggression in normal men: A randomized controlled trial. Arch Gen Psychiatry, in press.

Seiden LS, Sabol KE, Ricaurte GA: Amphetamine: Effects on catecholamine systems and behavior. Annu Rev Pharmacol Toxicol *33:*639, 1993.

*Self DW, Nestler EJ: Relapse to drug-seeking: Neural and molecular mechanisms. Drug Alcohol Dependence *51:*49, 1998.

Solowij N, Hall W, Lee N: Recreational MDMA use in Sydney: A profile of "ecstasy" users and their experiences with the drug. Br J Addict *87:*1161, 1992.

*Steele TD, McCann UD, Ricaurte GA: 3,4-Methylenedioxymethamphetamine (MDMA, "ecstasy"): Pharmacology and toxicology in animals and humans. Addiction *89:*539, 1994.

Strang J, Sheridan J: Prescribing amphetamines to drug misusers: Data from the 1995 national survey of community pharmacies in England and Wales. Addiction *92:*833, 1997.

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L: Co-occurrence of abuse of different drugs in men. Arch Gen Psychiatry *55:*967, 1998.

Volkow ND, Wang G-J, Fowler JS, Gatley SJ, Logan J, Ding Y-S, Hitzemann R, Pappas N: Dopamine transporter occupancies in the human brain induced by therapeutic doses of oral methylphenidate. Am J Psychiatry *155:*10, 1998.

Vollenweider FZ, Gamma A, Liechti M, Huber T: Psychological and cardiovascular effects and short-term sequelae of MDMA ("ecstasy") in MDMA-naive healthy volunteers. Neuropsychopharmacology *19:*241, 1998.

White FJ, Kalivas PW: Neuroadaptations involved in amphetamine and cocaine addiction. Drug Alcohol Dependence *51:*141, 1998.

11.4 CAFFEINE-RELATED DISORDERS

ERIC C. STRAIN, M.D. AND ROLAND R. GRIFFITHS, PH.D.

The most widely consumed psychoactive substance in the world is caffeine. It is estimated that over 80 percent of adults in the United States consume caffeine regularly, and throughout the world caffeine consumption is well integrated into daily cultural practices (e.g., the coffee break in the United States, tea time in the United Kingdom, and kola nut chewing in Nigeria). Because caffeine use is so pervasive and widely accepted, disorders associated with caffeine use may be overlooked. However, it is important to recognize that caffeine is a psychoactive compound that can produce a wide variety of syndromes, and the fourth edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)* recognizes several caffeine-related disorders (such as caffeine intoxication, caffeine-induced anxiety disorder, and caffeine-induced sleep disorder). In addition, other caffeine-related disorders, such as caffeine withdrawal and caffeine dependence, are not official diagnoses in *DSM-IV*, but can be of clinical interest.

HISTORY

Caffeine-containing foods and beverages have been consumed for hundreds if not thousands of years. Tea has been cultivated and consumed in China since at least 350 AD, and coffee cultivation spread from Ethiopia to Arabia in the fifteenth century and subsequently became a popular beverage in Arabic cultures during the sixteenth century. However, most caffeine consumption occurred in restricted geographical regions, and it was not until the seventeenth century that caffeine use began to spread, eventually becoming nearly universally available, as it is today. The Dutch imported caffeine in the form of coffee from Arabic countries to Europe, where it first became popular around the middle to end of the seventeenth century (along with tea, tobacco, and chocolate). Medical literature during this time period described coffee as a useful beverage for a wide variety of conditions, although it is notable that coffee's use to induce sobriety was a particularly attractive feature. However, considerable controversy about whether the use of coffee was beneficial or detrimental to health was associated with its expansion into Western cultures—a controversy that is mirrored in the contemporary concerns over the use of caffeine.

The expansion of coffee consumption during the seventeenth and eighteenth centuries can be demonstrated by the estimate that there were between 2000 and 3000 coffeehouses in London by the early part of the seventeenth century (about 1 coffeehouse for 200 to 300 people). The spread and popularity of coffeehouses in Europe highlights how coffee was initially a beverage consumed in public; it was only later that coffee became a beverage consumed in the home. Perhaps the most famous of the coffeehouses in London was operated by Edward Lloyd. Like most coffeehouses, Lloyd's was a center of business, especially for insurance agents, and by the end of the eighteenth century Lloyd's actually became the well-known Lloyd's of London insurance company. However, in England, coffee use was eventually supplanted by tea consumption, with the shift in beverage preference occurring in the first half of the eighteenth century.

In the United States a shift from tea to coffee use occurred in 1773 when colonists protesting British taxes threw cargoes of tea overboard in the Boston harbor—the Boston tea party. The repercussions of this event continue to the present day; the United States is the major consumer of coffee in the world.

In the late nineteenth century caffeine began appearing in various soft drink beverages. While some caffeine contained in soft drink beverages is derived from the cola nut, most caffeine in sodas is added. In the United States caffeine is also added to noncola sodas and to some bottled waters, and consumption of soda has been increasing markedly in the past 25 years.

In the contemporary world caffeine is integral to the economic activity of several countries. For example, only crude oil produces more foreign exchange earnings than coffee for developing countries. In the United States coffee is the major agricultural import and is second only to oil among all imports. This economic and trade activity underscores the extent to which caffeine use has spread over the past four centuries, so that caffeine is now available and accepted virtually everywhere in the world.

COMPARATIVE NOSOLOGY

There is no mention made of caffeine-related disorders in the first edition of DSM (DSM-I), which was published in 1952. In the second edition of DSM (DSM-II), published in 1968, caffeine (and interestingly, tobacco) are explicitly excluded from consideration for a diagnosis of drug dependence, and no other mention of caffeine use disorders is included in this edition. The American Psychiatric Association first included caffeine (and tobacco) use disorders in the third edition of DSM (DSM-III) in 1980. At that time *caffeine intoxication*—the only caffeine-related disorder in DSM-III—was included as a discrete syndrome with specific criteria for its diagnosis. Over successive editions of the DSM, caffeine intoxication has remained and DSM-IV has included criteria for caffeine withdrawal in the appendix. The text suggests that further research is needed to establish caffeine withdrawal as a discrete syndrome. The *DSM-IV* has also included diagnoses of caffeine-induced anxiety and caffeine-induced sleep disorders—that is, conditions in which specific caffeine-induced symptoms (anxiety, sleep disturbance) require clinical attention. Caffeine abuse and caffeine dependence are not included in DSM-IV, although there is evidence that some patients can exhibit a caffeine dependence syndrome.

The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) contains criteria for caffeine intoxication ("Acute intoxication due to the use of other stimulants, including caffeine"), and also includes several other diagnostic categories that can be applied to caffeine such as "harmful use," "dependence syndrome," and "withdrawal states from other stimulants, including caffeine." Thus, while DSM has tended to restrict the clinical syndromes associated with caffeine use, it is important to note ICD-10, another prominent diagnostic system for psychiatric disorders, has a more inclusive set of conditions associated with caffeine use.

EPIDEMIOLOGY

It is difficult to determine the average amount of caffeine used on a daily basis, partly because of the multiple sources of caffeine available to consumers. Caffeine is found in beverages (coffees, teas, sodas), foods (chocolate), and medications (both prescription and over-the-counter drugs), although most caffeine consumed is derived from coffee and tea. Some sources of caffeine such as coffees and colas, are readily identifiable, others are less easily recognized. Thus, for example, noncola sodas can contain caffeine, and certain over-the-counter analgesics also contain caffeine.

Estimates of caffeine consumption require knowledge of the caffeine content of these different sources of caffeine. Also, there can be variability within categories, so that different colas, for example, may have different caffeine contents, and the procedure used to prepare a beverage (i.e., brewed versus instant coffee) can also influence the caffeine content. Table 11.4-1 provides a list of the typical caffeine content of selected caffeine-containing products, although these values represent estimates and amounts may vary.

Substance	Caffeine content
Brewed coffee	100mg/6 oz
Instant coffee	70mg/6 oz
Decaffeinated coffee	4mg/6 oz
Leaf or bag tea	40mg/6 oz
Instant tea	25mg/6 oz
Caffeinated soda	45mg/12 oz
Cocoa beverage	5mg/6 oz
Chocolate milk	4mg/6 oz
Dark chocolate	20mg/1 oz
Milk chocolate	6mg/1 oz
Caffeine-containing cold remedies	25–50 mg/tablet
Caffeine-containing analgesics	25–65 mg/tablet
Stimulants	100–350 mg/tablet
Weight-loss aids	75–200 mg/tablet

Adapted from DSM-IV and Barone JJ, Roberts HR: Caffeine consumption. Food Chem Toxicol *34*:119, 1996.

Table 11.4-1 Typical Caffeine Content of Foods and Medications

Caffeine consumption also varies by age. Figure 11.4-1 shows estimates of per capita caffeine consumption, by those who consume caffeine, for different age groups in the United States. These estimates demonstrate the wide variability in caffeine consumption for different ages. As shown in Fig. 11.4-1, the average daily caffeine consumption for all ages of caffeine consumers is 2.79 mg/kg in the United States. It is worth noting that there is substantial caffeine consumption even by young children (i.e., over 1 mg/kg for children between the ages of 1 to 5 years). Worldwide it is estimated the average daily per capita caffeine consumption is about 70 mg.

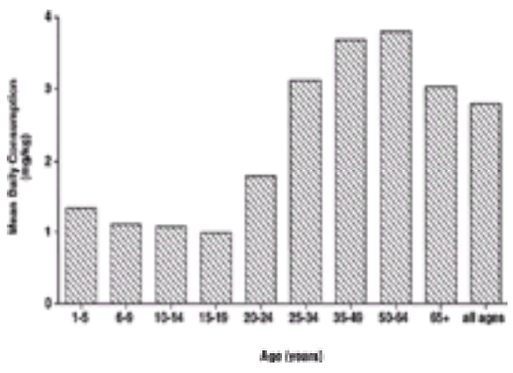

FIGURE 11.4-1 Mean daily caffeine consumption (mg/kg) for different age groups and all ages, in the United States of America. (Adapted from Barone JJ, Roberts HR: Caffeine consumption. Food Chem Toxicol *34:*119, 1996.)

In the United States, the major source of caffeine is coffee, although coffee use peaked in 1962 with a gradual decline throughout the 1980s and 1990s (Fig. 11.4-2). Fifty-two percent of the United States population 10 years of age or older drank coffee in 1993. While coffee is the most popular form in which caffeine is consumed, caffeine use in children was equally divided between tea and coffee until the late 1980s, when there was a decrease in tea use and a compensatory increase in coffee use. Estimates of soda-related caffeine use are more difficult to determine, although it is notable that soda use (predominantly caffeinated) has been steadily increasing in the United States (Fig. 11.4-2). It is estimated that approximately 70 percent of the soda consumed in the United States is caffeinated.

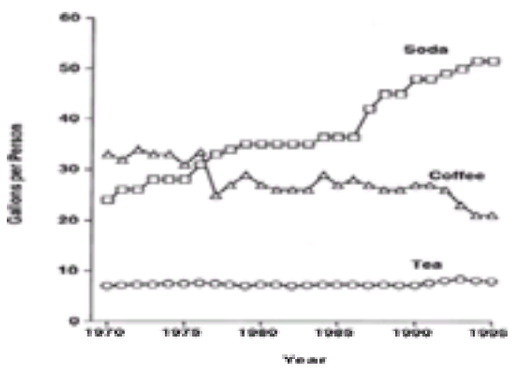

FIGURE 11.4-2 Consumption by year for soda, coffee, and tea in the United States of America (gallons per person). Values represent both caffeinated and noncaffeinated products. (Adapted from Liebman B: The changing American diet. Nutr Action Healthlett *24*(3):8, 1997.)

The average daily caffeine consumption for all adults in the United States is generally estimated to be 3 mg/kg, and for adult consumers of caffeine it is estimated to be 4 mg/kg (the equivalent of about 3 or more cups of brewed coffee in a man of average weight). Among the heaviest adult consumers of caffeine, caffeine intake is at least 5 to 7 mg/kg per day (the equivalent of about 6 cups of brewed coffee in a man of average weight).

ETIOLOGY

It is not uncommon for the first exposure to caffeine to occur during childhood. Following exposure to caffeine, continued caffeine consumption may be influenced by several different factors, such as the pharmacological effects of caffeine, caffeine's reinforcing effects, genetic predispositions to caffeine use, and personal attributes of the consumer (e.g., age).

PHARMACOLOGY AND EFFECTS IN THE CENTRAL NERVOUS SYSTEM

Caffeine is a methylxanthine, as are theobromine (found in chocolate) and theophylline (Theo-Dur), typically used in the treatment of asthma. It is well absorbed from the gastrointestinal tract, with peak plasma concentrations typically occurring within 1 hour after ingestion. Caffeine crosses the blood-brain barrier and is metabolized by the liver. Caffeine metabolism is increased by smoking, and it is not uncommon to find higher rates of caffeine consumption in people who smoke tobacco.

Caffeine (and the methylxanthines in general) exert several different effects, including the relaxation of smooth muscle (hence their therapeutic application in the treatment of asthma, where they relax the smooth muscle of bronchi), complex but probably only modest actions on heart rate and blood pressure (which depend in part upon whether or not the person is tolerant to the effects of caffeine), and increased production of urine (although the mechanism of these diuretic effects is not clear).

Numerous investigations have examined the effects of caffeine in the central nervous system in order to understand the molecular basis for caffeine's effects. Mechanisms proposed as mediating caffeine's effects have included the inhibition of phosphodiesterases and effects on intracellular calcium (although these effects may only occur with higher doses of caffeine), and antagonism at the adenosine receptor. This latter action—adenosine receptor antagonism—appears to be the primary site of action for caffeine's effects at doses typically ingested by humans. Recent studies also suggest an important role of dopamine in mediating some of the behavioral effects of caffeine.

Subjective Effects and Reinforcement Single low to moderate doses of caffeine (i.e., 20 to 200 mg) can produce a profile of subjective effects in humans that is generally identified as pleasurable. Thus, studies have shown that such doses of caffeine result in increased ratings on measures such as well-being, energy and concentration, and motivation to work. In addition, these doses of caffeine produce decreases in ratings of feeling sleepy or tired. Doses of caffeine in the range of 300 to 800 mg (the equivalent of several cups of brewed coffee ingested at once) produce effects that are rated as being unpleasant, such as anxiety and nervousness. Although animal studies have generally found it difficult to demonstrate that caffeine functions as a reinforcer, well-controlled studies in humans have shown that people choose caffeine over placebo when given the choice under controlled experimental conditions. Thus, the profile of caffeine's subjective effects and its ability to function as a reinforcer can contribute to the regular use of caffeine.

Genetics and Caffeine Use Studies examining the possibility of genetic influences on caffeine use have generally focused only on coffee use (rather than on all sources of caffeine), and thus conclusions from these studies need to be viewed with caution. However, several investigations comparing coffee use in monozygotic versus dizygotic twins have shown higher concordance rates for monozygotic twins, suggesting that there may be some genetic predisposition to continued coffee use following exposure to coffee.

Age, Sex, and Race The relation between long-term chronic caffeine use and demographic features such as age, sex, and race has not been widely studied. There is some evidence that suggests that older people may use more caffeine, although caffeine use in adolescents is not uncommon. There is no known evidence that caffeine use differs for men versus women, and there is no data that specifically addresses caffeine use for different races. There is some evidence suggesting that in the United States whites consume more coffee and tea than African-Americans although the survey that found this difference did not distinguish between caffeinated and noncaffeinated products. This survey found no difference in the rates of soda consumption by race.

Special Populations Caffeine's metabolism is increased in people who smoke tobacco, and several studies have shown higher amounts of daily caffeine consumption in people who smoke. Caffeine metabolism is decreased in women who take oral contraceptives, so that lower daily caffeine consumption in this population may simply reflect compensation because of decreased metabolism.

There have also been several studies showing high daily amounts of caffeine use in psychiatric patients. For example, several studies have found such patients consume the equivalent of an average of five or more cups of brewed coffee each day. Finally, high daily caffeine consumption has also been noted in prisoners.

Personality Although attempts have been made to link preferential use of caffeine to particular personality types, results from these studies do not suggest that any particular personality type is especially linked to caffeine use.

DSM-IV DISORDERS

Caffeine use can be associated with five discrete syndromes. Three of these conditions—caffeine intoxication, caffeine-induced anxiety disorder, and caffeine-induced sleep disorder—are described in DSM-IV (Table 11.4-2). A fourth syndrome—caffeine withdrawal—is included in the appendix of DSM-IV, and a final condition—caffeine dependence—is not included in DSM-IV but will be discussed here. These latter two conditions can be diagnosed using the category caffeine-related disorders not otherwise specified, which is included in DSM-IV as a substance-related disorder.

Caffeine-Induced Disorders
Caffeine intoxication
Caffeine-induced anxiety disorder
Specify if:
With onset during intoxication
Caffeine-induced sleep disorder
Specify if:
With onset during intoxication
Caffeine-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.4-2 DSM-IV Caffeine-Related Disorders

Caffeine Intoxication Caffeine intoxication has long been recognized as a discrete syndrome associated with the use of a significant amount of caffeine (Table 11.4-3). Intoxication can be the result of an ingestion of a large amount of caffeine in a person who has not regularly consumed caffeine (a nontolerant individual), and thus may represent an overdose of caffeine. Alternatively, caffeine intoxication can occur in the context of a person who chronically consumes large amounts of caffeine, which produces a more complicated clinical picture. Interestingly, while intoxication by some psychoactive substances, such as alcohol, is sought out by some people, it appears that most individuals do not actively seek repeated episodes of caffeine intoxication.

A. Recent consumption of caffeine, usually in excess of 250 mg (e.g., more than 2–3 cups of brewed coffee).
B. Five (or more) of the following signs, developing during, or shortly after, caffeine use:
(1) restlessness
(2) nervousness
(3) excitement
(4) insomnia
(5) flushed face
(6) diuresis
(7) gastrointestinal disturbance
(8) muscle twitching
(9) rambling flow of thought and speech
(10) tachycardia or cardiac arrhythmia
(11) periods of inexhaustibility
(12) psychomotor agitation
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder (e.g., anxiety disorder).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.4-3 DSM-IV Diagnostic Criteria for Caffeine Intoxication

Epidemiology There have been few studies examining the prevalence of caffeine intoxication, and most of these studies have looked at selected populations (e.g., psychiatric inpatients, college students) and used ambiguous criteria. Notably, one random-digit telephone survey of caffeine use in the general community found that 12 percent of respondents had met the revised third edition of DSM (DSM-III-R) criteria for caffeine intoxication in the past year.

Comorbidity Given the limited number of studies of caffeine intoxication, it is not surprising that there is even less known about comorbid disorders associated with this condition. However, it is notable that there appear to be some associations between the amount of caffeine typically consumed by individuals and certain psychiatric conditions (although it should be stressed that these are associations between the amount of caffeine consumed, not a specific diagnosis of caffeine intoxication). Thus, for example, high caffeine consumption has been noted in patients with psychiatric disorders such as bipolar I disorder, schizophrenia, and personality disorders. Excessive caffeine consumption has also been noted in people who smoke tobacco and may be due to increased metabolism of caffeine in smokers. Finally, although rare, it may be possible for a patient to develop delirium from ingesting an acute, extremely high dose of caffeine, and rare cases of suicide have been noted with excessive caffeine consumption.

There are some patient populations that tend to not consume caffeine—patients with anxiety disorders. It has been shown that patients with generalized anxiety disorder and patients with panic disorder can be more sensitive to the effects of caffeine, and that patients with panic disorder tend to have lower caffeine

consumption. Thus, these may be patient populations in whom there is less likelihood that caffeine intoxication will be observed.

Diagnosis and Clinical Features Caffeine intoxication can present with a variety of clinical features, as shown in the criteria from DSM-IV (Table 11.4-3). In addition to these signs and symptoms, reports have found patients with caffeine intoxication can have fever, irritability, tremors, sensory disturbances, tachypnea, and headaches. It appears that the most common features of caffeine intoxication are anxiety, nervousness, insomnia, gastrointestinal disturbances, tremors, tachycardia, and psychomotor agitation.

Differential Diagnosis Caffeine intoxication severe enough to come to clinical attention is probably a relatively rare condition, and several other disorders should be considered for patients who present with features suggesting caffeine intoxication. These include other substance-related disorders, such as intoxication with other stimulants (e.g., cocaine or amphetamines), and withdrawal from drugs such as sedative-hypnotics (e.g., benzodiazepines), alcohol, or nicotine. In addition, medication adverse effects such as akathisia can present with features suggestive of caffeine intoxication. The final diagnosis of caffeine intoxication depends upon demonstrating the ingestion of a significant quantity of caffeine prior to the onset of symptoms.

Course and Prognosis Caffeine has a relatively short halflife (3-6 hours), so caffeine intoxication will typically resolve quickly and with no significant sequelae. Patients with caffeine intoxication generally have a good prognosis, although there have been isolated reports of people who have completed suicide by ingesting a massive amount of caffeine.

Treatment The first step in treating a patient presenting with evidence suggesting caffeine intoxication is to confirm the diagnosis. It is important to consider all possible sources of caffeine when considering a diagnosis of caffeine intoxication, including medications and beverages that may not be initially recognized as containing caffeine (e.g., some noncola sodas). Because caffeine has a relatively short half-life, short-term management of the patient can be supportive while the syndrome resolves spontaneously. For the patient who habitually consumes large amounts of caffeine and repeatedly experiences episodes of caffeine intoxication, it may be helpful to aid the patient to recognize their total daily caffeine consumption through the use of daily diaries, and then to teach the patient about the adverse effects associated with caffeine use.

Caffeine-Induced Anxiety Disorder A caffeine-induced anxiety disorder can be panic disorder, generalized anxiety disorder, social phobia, or obsessive-compulsive disorder, although the patient does not need to fulfill all the criteria for one of these disorders to quality for a diagnosis of caffeine-induced anxiety disorder (see Table 15.6-18).

There has been no work examining this caffeine-related disorder, although there have been investigations examining the relationship between caffeine use and anxiety. Patients with anxiety disorders generally consume less caffeine than control patients, and experience greater self-reports of anxiety after consuming caffeine. In addition, caffeine has been used as a pharmacological probe in subjects without an anxiety disorder, and a sufficiently high dose can produce panic attacks in subjects.

The diagnosis of caffeine-induced anxiety disorder depends upon linking the use of caffeine to the anxiety symptoms of concern. For a patient with a suspected caffeine-induced anxiety disorder, a trial of caffeine abstinence may aid in clarifying the diagnosis.

Mr. B. was a 28-year-old single African-American male graduate student who was in good health and had no history of previous psychiatric evaluation or treatment. He took no medications, did not smoke or consume alcohol, and had no current or past history of illicit drug use.

His chief complaint was that he had begun feeling mounting “anxiety” when working in the laboratory where he was pursuing his graduate studies. His work had been progressing well, he felt his relationship with his advisor was good and supportive, and he could not identify any problems with staff or peers that might explain his anxiety. He had been working long hours, but found the work interesting and had recently had his first paper accepted for publication.

Despite these successes, he reported feeling a “crescendoing anxiety” as his day would progress. He noted that by the afternoon he would be experiencing palpitations, bursts of his heart racing, tremors in his hands, and an overall feeling of “being on the edge.” He also noted a nervous energy in the afternoons. These experiences were occurring daily, and seemed confined to the laboratory (although he admitted he was in the laboratory every day of the week).

When reviewing Mr. B.'s caffeine intake, it was found he was consuming excessive amounts of coffee. Staff made a large urn of caffeinated coffee each morning, and Mr. B. routinely started with a large mug of coffee. Over the course of the morning he would consume 3 to 4 mugs of coffee (the equivalent of about 6 to 8 5-ounce cups of coffee), and continued this level of use throughout the afternoon. He occasionally had a single can of caffeinated soda, and used no other forms of caffeine on a regular basis. Mr. B. estimated he drank a total of 6 to 8 or more mugs of coffee per day (which was estimated to be at least 1200 mg of caffeine per day). Once pointed out to him, he realized this level of caffeine consumption was considerably higher than at any other time in his life. He admitted he liked the taste of coffee, and felt a burst of energy in the morning when he drank coffee that helped him start his day.

Mr. B. and his physician developed a plan to decrease his caffeine use by tapering off caffeine. Details of such a tapering schedule can be found in the section on treatment of caffeine dependence. Mr. B. was successful in decreasing his caffeine use, and had good resolution of his anxiety symptoms once his daily caffeine use had been markedly decreased.

Caffeine-Induced Sleep Disorder Forms of a caffeine-induced sleep disorder can be insomnia, hypersomnia, parasomnia, or mixed. However, caffeine use is most frequently associated with insomnia (although there are case reports of patients who have hypersomnia in response to caffeine). Caffeine-induced sleep disorder should be diagnosed in patients with caffeine intoxication only if the sleep disturbance is in excess to that which would be expected from intoxication (see Table 21-18).

Although there are essentially no studies directly examining caffeine-induced sleep disorder, there has been work on the relationship between caffeine use and sleep. In general, ingestion of caffeine prior to bedtime results in a delay in sleep onset and poorer sleep quality. For some people, it has been shown that consuming 200 mg of caffeine before bedtime (the equivalent of about 2 cups of brewed coffee) can delay the onset of sleep for up to 4 hours.

Sleep disturbances secondary to caffeine use are more likely in people who are not regular consumers of caffeine. For people who consume caffeine on a daily basis, there is evidence to suggest that some tolerance occurs to the sleep-inhibiting effects of caffeine.

Caffeine-Related Disorder Not Otherwise Specified This category is included in DSM-IV as a substance-related disorder, and is to be used when the patient has a condition that is caffeine-related but not diagnosable as caffeine intoxication, caffeine-induced anxiety disorder, or caffeine-induced sleep disorder (Table 11.4-4). Thus, for example, caffeine withdrawal or caffeine dependence could be coded as a caffeine-related disorder not otherwise specified.

The caffeine-related disorder not otherwise specified category is for disorders associated with the use of caffeine that are not classifiable as caffeine intoxication, caffeine-induced anxiety disorder, or caffeine-induced sleep disorder. An example is caffeine withdrawal.

Table 11.4-4 DSM-IV Diagnostic Criteria for Caffeine-Related Disorder Not Otherwise Specified

Caffeine Withdrawal Caffeine withdrawal was not included in DSM-III-R, apparently because the symptoms of caffeine withdrawal were believed to be mild. Caffeine withdrawal has been included in the appendix of DSM-IV with a set of diagnostic criteria to be used for further evaluation of the disorder (Table 11.4-5). It is not included in the body of DSM-IV, despite evidence for a discrete caffeine withdrawal syndrome, because of concerns that there were only a limited number of symptoms (three), some symptoms were similar, and the symptoms have a high prevalence in the general population and can be frequently associated with other circumstances besides caffeine withdrawal.

A. Prolonged daily use of caffeine.
B. Abrupt cessation of caffeine use, or reduction in the amount of caffeine used, closely followed by headache and one (or more) of the following symptoms:
   (1) marked fatigue or drowsiness
   (2) marked anxiety or depression
   (3) nausea or vomiting
C. The symptoms in Criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to the direct physiological effects of a general medical condition (e.g., migraine, viral illness) and are not better accounted for by another mental disorder.

Table 11.4-5 DSM-IV Research Criteria for Caffeine Withdrawal

Epidemiology Several studies have examined caffeine withdrawal under conditions where it is experimentally induced in a population of subjects. In a study of normal community volunteers whose average daily caffeine consumption was similar to that of the general population, one half of the subjects had moderate or severe headache when they had experimentally induced caffeine withdrawal; 8 to 11 percent of subjects in that study also experienced anxiety, fatigue, and depression while in withdrawal. Other similar studies have found that one third to one half of subjects will experience symptoms of caffeine withdrawal when it is experimentally induced (i.e., subjects undergo a double-blind discontinuation of their daily caffeine use).

There have been relatively few attempts to conduct population-based studies on caffeine withdrawal. Such studies have usually been surveys, and found a wide range of reports of caffeine withdrawal (i.e., from 8–42 percent). In the United States it is estimated that over 80 percent of the adult population are regular consumers of caffeine, and even if the incidence of caffeine withdrawal is low, this still would suggest there are large numbers of people who could experience caffeine withdrawal.

Comorbidity There have been no studies that have specifically sought to determine conditions that are comorbid with caffeine withdrawal. However, patients with higher daily consumption of caffeine may be at increased risk for developing caffeine withdrawal if they miss consuming their typical dose of caffeine.

Diagnosis and Clinical Features A withdrawal condition associated with the cessation of caffeine use has been long recognized, and the most common feature noted with caffeine withdrawal has been headache. DSM-IV also includes marked fatigue or drowsiness, marked anxiety or depression, and nausea or vomiting in the diagnostic criteria. However, several other signs and symptoms of caffeine withdrawal have been noted, such as impaired concentration, yawning, decreased sociability, lassitude, work difficulty, irritability, muscle aches, stiffness, and other flu-like symptoms (Table 11.4-6). In addition, performance on psychomotor tasks has shown to be impaired when subjects are experiencing caffeine withdrawal.

Headache
Sleepiness or drowsiness
Impaired concentration, lassitude, work difficulty
Anxiety
Depression
Irritability
Nausea or vomiting
Muscle aches or stiffness
Yawning
Flu-like symptoms

Table 11.4-6 Clinical Features of Caffeine Withdrawal

CAFFEINE WITHDRAWAL HEADACHE Headache is the most common feature associated with caffeine withdrawal. It is usually described as a generalized, throbbing headache that is worsened by exercise and Valsalva's maneuver, and responds best to caffeine consumption. It typically begins 12 to 24 hours after the last ingestion of caffeine, although the onset can occur as long as 40 hours after the last dose. It usually resolves in 2 to 4 days, although sporadic headaches can continue for up to 11 days.

CAFFEINE WITHDRAWAL HEADACHE AND OTHER SIGNS AND SYMPTOMS OF CAFFEINE WITHDRAWAl There are several other signs and symptoms associated with caffeine withdrawal. Patients can experience these signs and symptoms even without the concurrent presence of headache, suggesting that the other features of caffeine withdrawal are not simply secondary to the headache.

CAFFEINE WITHDRAWAL AND POSTOPERATIVE HEADACHE Headache is a frequent postoperative symptom. Patients undergoing operative procedures are required to abstain from all oral intake prior to the procedure, including daily dietary caffeine they would typically consume, and several studies have examined whether postoperative headache could reflect caffeine withdrawal. Interestingly, patients with higher daily caffeine consumption are at increased risk for developing postoperative headache, and the consumption of a caffeinated beverage on the day of an operative procedure has been shown to lower the rate of postoperative headache.

CAFFEINE WITHDRAWAL AND DOSAGE OF CAFFEINE The risk of developing caffeine withdrawal, and the severity of the withdrawal syndrome are both related to the daily dosage of caffeine consumed (with risk and severity increasing as the dose of caffeine increases). However, it should be noted that caffeine withdrawal can occur with relatively low dosages of caffeine, such as 100 mg per day (the equivalent of about one cup of brewed coffee).

CAFFEINE WITHDRAWAL AND AGE Most studies of caffeine withdrawal have determined the presence and features of caffeine withdrawal in adults. However, caffeine withdrawal has also been shown to occur in children. Like caffeine withdrawal in adults, headache and other caffeine withdrawal symptoms can occur in children.

Differential Diagnosis There are a wide variety of medical conditions and psychiatric disorders that can overlap with the symptoms of caffeine withdrawal. In part, it was this relative nonspecificity of the criteria for caffeine withdrawal that contributed to the decision to not include it in the body of DSM-IV. When considering a patient who may have caffeine withdrawal, disorders as diverse as viral illnesses, sinus conditions, drug withdrawal states (e.g., from amphetamines, cocaine), other types of headaches such as tension or migraine, and medication reactions should be considered in the differential diagnosis. The final determination of caffeine withdrawal should rest upon a determination of the pattern and amount of caffeine consumption, the time interval between the last ingestion of caffeine and the onset of symptoms, and the particular clinical features presented by the patient. Resolution of symptoms by a dose of caffeine may also be useful in clarifying the diagnosis.

Course and Prognosis Caffeine withdrawal typically begins 12 to 24 hours after the last dose of caffeine, and usually resolves in 2 to 7 days. Most symptoms reach their maximal intensity within the first 48 hours after cessation of caffeine use, with subsequent decreasing intensity over the following days. However, it should be noted that there can be considerable variability between people in the manifestations, severity, and time course of caffeine withdrawal. Furthermore, it has been shown that the same person can experience different symptoms and different degrees of severity of symptoms across different episodes of caffeine withdrawal.

Treatment There has been little work systematically investigating the optimal treatment for caffeine withdrawal. The symptoms of caffeine withdrawal typically resolve with the ingestion of a dose of caffeine, and a person with a provisional diagnosis of caffeine withdrawal who continues to be symptomatic after caffeine consumption has been terminated for more than 2 weeks should be carefully re-evaluated and considered for other possible diagnoses. Caffeine withdrawal headaches may also respond to aspirin.

If caffeine withdrawal occurs because a person is attempting to abruptly stop all caffeine use, it may be better to attempt a tapering caffeine dosage schedule.

> Ms. E. was a 32-year-old single white woman employed full-time at a local factory. She occasionally used nonsteroidal anti-inflammatory drugs, but was taking no regular prescription medications. She had a history of alcohol dependence, in remission for 9 years, and was otherwise in good health.
>
> She first began consuming caffeine when she started college, and her current beverage of choice was coffee. She typically drank 4 to 5 mugs of coffee each day, and preferred to drink it without cream, milk, or sugar. She estimated that 5 minutes elapsed between the time she got up in the morning and the time she had her first cup of coffee; her roommate made a pot before Ms. E. got up, and Ms. E. immediately poured a mug when she got out of bed. She spaced her mugs over the course of the day, with her last mug either after lunch or with dinner.
>
> Physicians had recommended she cut down or stop her coffee use because of complaints of mild indigestion, but she had been unable to do so. Her roommate had also complained about her coffee use at times. Ms. E. routinely drank hot coffee in her car, and had spilled it and burned herself on one occasion.
>
> When she had stopped caffeine abruptly, Ms. E. experienced marked irritability; poor concentration; and a severe, generalized headache. When asked to rate the severity of the headache, she replied that “on a scale of 1 to 10 it's a 12.” She also had muscle aches, low energy, lethargy, and a craving to drink a mug of coffee. On the day that she had stopped coffee use abruptly, she left work 2 hours early because of problems with concentrating on the job, and went to bed several hours earlier than usual. She then returned to her usual pattern of coffee use.

Caffeine Dependence Caffeine dependence is not included in DSM-IV, and it is explicitly stated that “A diagnosis of Substance Dependence can be applied to every class of substance except caffeine.” Despite the absence of caffeine dependence in DSM-IV, there is evidence supporting a diagnosis of caffeine dependence in some people with problematic caffeine consumption.

Before reviewing the features of caffeine dependence, it is important to clarify the use of the word “dependence.” Dependence is sometimes used to indicate the presence of physical dependence—a condition characterized by physiological adaptation to the effects of a drug, and usually indicated by the presence of a withdrawal syndrome when drug ingestion is discontinued. There is considerable evidence that caffeine can produce physical dependence, as indicated by the presence of caffeine withdrawal.

The term *dependence* is also used in a second way—to indicate a clinical diagnosis of dependence. Clinical diagnoses of dependence typically have included a constellation of diagnostic criteria that are loosely linked by the theme of problematic use of the drug. Included in these criteria can be evidence of physical dependence, although physical dependence is usually neither necessary nor sufficient to make a diagnosis of a clinical syndrome of dependence. The DSM criteria are used to make a clinical diagnosis of dependence.

Although there are numerous studies and reports on caffeine's ability to produce physical dependence, as evidenced by caffeine withdrawal, there is little research on whether or not some people who consume caffeine can develop a clinical syndrome of dependence. This is a question that only a few studies have addressed.

Epidemiology Caffeine use is common throughout the world. Over 80 percent of adults in the United States consume caffeine regularly, and their average daily caffeine consumption has been estimated to be 280 mg (the equivalent of about 3 cups of brewed coffee). Given the substantial number of people who consume caffeine regularly, and their use of daily dosages that produce psychoactive effects, even a low prevalence rate of caffeine dependence could represent substantial numbers of people.

However, there has only been one study that has examined the prevalence of caffeine dependence in the general population—a random-digit telephone survey that used the generic criteria from DSM-III-R. That study found that 3 percent of respondents fulfilled criteria for severe dependence (7–9 criteria), 14 percent fulfilled criteria for moderate dependence (5–6 criteria), and 27 percent fulfilled criteria for mild dependence (3–4 criteria).

Comorbidity The only study that has reported on comorbid conditions in patients with caffeine dependence was a selected series of case reports and thus it is difficult to generalize to the general population using those results. The study found that about two thirds of patients had a psychiatric disorder in remission, most commonly another substance abuse disorder. However, nearly one half had a mood disorder in remission, one quarter had an anxiety disorder in remission, and one fifth had an eating disorder in remission. Notably, there was a clustering of nicotine dependence, caffeine dependence, and a history of an alcohol use disorder in several subjects, and this clustering of these three disorders has been noted in other studies.

CAFFEINE USE AND NONPSYCHIATRIC ILLNESSES There has been considerable interest in determining whether caffeine use is associated with other physical illnesses (analogous to findings of the relationship between nicotine use and conditions such as heart disease and cancer). Numerous studies have sought to determine whether caffeine use might be associated with heart disease, cancer, breast disease, osteoporosis, and other physical illnesses. In general, no such associations have been found, although large epidemiological studies continue to seek such relationships. When patients question a clinician about the possible association of caffeine with various such physical conditions, they can generally be reassured that there is no evidence of a strong relationship between caffeine use and physical illnesses. Interestingly, despite this lack of association between caffeine use and physical illnesses, a survey of medical specialists found that over three quarters recommended that patients decrease or eliminate their caffeine use for certain conditions (such as anxiety, arrhythmias tachycardia, esophagitis or hiatal hernia, and fibrocystic disease).

It may be that there is a mild association between higher daily caffeine use in women and delayed conception and slightly lower birth weight. However, there are studies that have not found such associations, and effects when found are usually with relatively high daily dosages of caffeine (e.g., the equivalent of 5 cups of brewed coffee per day). For a woman who is considering pregnancy, especially if there is some difficulty in conceiving, it may be useful to counsel eliminating caffeine use. Similarly, for a woman who becomes pregnant and has a high daily caffeine consumption, a discussion about decreasing her daily caffeine use may be warranted.

Diagnosis and Clinical Features DSM-IV provides a generic set of diagnostic criteria that are to be used for determining the presence of a substance dependence syndrome (see Table XX-X). There is little work examining the particular clinical features of caffeine dependence, although there is one report of a series of 16 cases of subjects who fulfilled the DSM-IV criteria for caffeine dependence. Only 4 of the 7 criteria were used in that study, and most of the cases had evidence of the 4 criteria (withdrawal, use continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance use, persistent desire or unsuccessful efforts to cut down or control substance use, and tolerance). In a telephone survey of caffeine users, the most common DSM-III-R dependence criterion reported by respondents was a persistent desire or unsuccessful efforts to cut down or control substance use.

Differential Diagnosis When considering a possible diagnosis of caffeine dependence, it is useful to include other substance dependence syndromes in the differential diagnosis. A clinical syndrome of dependence upon caffeine probably goes overlooked by most clinicians because it is not included in DSM-IV and it is not widely recognized and acknowledged as a substance-related abuse disorder. Most people have no problems associated with being dependent upon caffeine so long as their supply is available on a daily basis. Even if they are unable to obtain caffeine for some reason, the withdrawal syndrome is relatively short-lived and is not life

threatening.

Course and Prognosis No studies have examined the course and prognosis for patients with a diagnosis of caffeine dependence. Subjects with caffeine dependence have reported continued use of caffeine despite repeated efforts to discontinue their caffeine use.

Treatment There have been case reports describing the treatment of patients with problematic caffeine use, although there have been no systematic studies of treating patients with a confirmed diagnosis of caffeine dependence. In general these approaches have used a combination of three techniques to aid patients to decrease or eliminate their caffeine use: gradual tapering of the daily dose, self-monitoring of daily use, and reinforcement for decreased use.

The first step in reducing or eliminating caffeine use is to have patients determine their daily consumption of caffeine. This can best be accomplished by having the patient keep a daily food diary. It is important for the patient to recognize all sources of caffeine in the diet, including different forms of caffeine (e.g., beverages, medications) and to accurately record the amount consumed. After several days of keeping such a diary, the clinician can meet with the patient, review the diary, and determine the average daily caffeine dose in milligrams.

The patient and clinician should then decide upon a fading schedule for caffeine consumption. Such a schedule could involve decreases in increments of 10 percent every few days. Since caffeine is typically consumed in beverage form, the patient can use a substitution procedure in which decaffeinated beverage is gradually used in place of caffeinated beverage. A diary should continue to be maintained during this time, so that the patient's progress can be monitored. The fading schedule should be individualized for the patient so that the rate of decrease in caffeine consumption minimizes withdrawal symptoms. Abruptly stopping all caffeine use should probably be avoided because withdrawal symptoms are likely to develop with sudden discontinuation of all caffeine use.

Ms. G. was a 35-year-old married white homemaker with three children ages 8,6, and 2 years old. She took no prescription medications, took a multivitamin and vitamins C and E on a daily basis, did not smoke, and had no past history of psychiatric problems. She drank moderate amounts of alcohol on the weekends, had smoked marijuana in college but had not used it since, and had no other history of illicit drug use.

She had started consuming caffeinated beverages while in college, and her current beverage of choice was caffeinated diet cola. Ms. G. had her first soda early in the morning, shortly after getting out of bed, and she jokingly called it her “morning hit.” She spaced out her bottles of soda over the course of the day, with her last bottle at dinner time. She typically drank 4 to 5 20-ounce bottles of caffeinated diet cola each day.

She and her husband had argued about her caffeinated soda use in the past, and her husband had felt she should not drink caffeinated soda while pregnant. However, she had continued to do so during each of her pregnancies. Despite a desire to stop drinking caffeinated soda, she was unable to do. She described having a strong desire to drink caffeinated soda, and if she resisted this desire she found that she could not think of anything else. She drank caffeinated soda in her car, which had a manual transmission, and noted that she would fumble while shifting and holding the soda, and spill soda in the car. She also noted that her teeth had become yellowed, and she suspected this was related to her tendency to swish soda in her mouth before swallowing it. When asked to describe a time when she stopped using soda, she reported that she had run out of it on the day one of her children was to have a birthday party, and she did not have time to leave her home to buy more. In the early afternoon of that day, a few hours before the scheduled start of the party, she felt extreme lethargy, a severe headache, irritability, and craving for a soda. She called her husband and told him she planned to cancel the party. She then went to the grocery store to buy soda, and after drinking two bottles she felt well enough to host the party.

While initially expressing interest in decreasing or stopping her caffeinated soda use, Ms. G. failed to attend scheduled follow-up appointments after her first evaluation. When finally contacted at home, she reported she had only sought help initially at her husband's request, and she had decided to try to cut down on her caffeine use on her own.

SUGGESTED CROSS-REFERENCES

Chapter 11 discusses substance-related disorders, anxiety disorders are discussed in Chapter 15, and sleep disorders are discussed in Chapter 21.

SECTION REFERENCES

*Barone JJ, Roberts HR: Caffeine consumption. Food Chem Toxicol *34:*119, 1996.

Bernard ME, Dennehy S, Keefauver LW: Behavioral treatment of excessive coffee and tea drinking: A case study and partial replication. Behav Therapy *12:*543, 1981.

Boulenger JP, Uhde TW, Wolff EA, Post RM: Increased sensitivity to caffeine in patients with panic disorder. Arch Gen Psychiatry *41:*1067, 1984.

Bruce M, Scott N, Shine P, Lader M: Anxiogenic effects of caffeine in patients with anxiety disorders. Arch Gen Psychiatry *49:*867, 1992.

Charney DS, Heninger GR, Jatlow PI: Increased anxiogenic effects of caffeine in panic disorder. Arch Gen Psychiatry *42:*233, 1985.

Comer SD, Haney M, Foltin RW, Fischman MW: Effects of caffeine withdrawal on humans living in a residential laboratory. Exp Clin Psychopharmacol *5:*399, 1997.

*Dager SR, Layton ME, Strauss W, Richards TL, Heide A, Friedman SD, Artru AA, Hayes CE, Posse S: Human brain metabolic response to caffeine and the effects of tolerance. Am J Psychiatry *156:*229, 1999.

Dreisbach RH, Pfeiffer C: Caffeine-withdrawal headache. J Lab Clin Med *28:*1212, 1219, 1943.

Evans SM, Griffiths RR: Dose-related caffeine discrimination in normal volunteers: Individual differences in subjective effects and self-reported cues. Behav Pharmacol *2:*345, 1991.

*Evans SM, Griffiths RR: Caffeine withdrawal: A parametric analysis of caffeine dosing conditions. J Pharmacol Exp Ther *289:*285, 1999.

Garrett BE, Griffiths RR: The role of dopamine in the behavioral effects of caffeine in animals and humans. Pharmacol Biochem Behav *57:*1, 1997.

Garriott JC, Simmons LM, Poklis A, Mackell MA: Five cases of fatal overdose from caffeine-containing “look-alike” drugs. J Anal Toxicol *9:*141, 1985.

Gilbert RM: Caffeine consumption. In *The Methylxanthine Beverages and Foods: Chemistry, Consumption, and Health Effects*, GA Spiller, editor. Alan R. Liss, New York, 1984.

Goldstein A: Wakefulness caused by caffeine. Naunyn-Schmiedebergs Arch Pharmacol *248:*269, 1964.

Goldstein A, Kaizer S, Whitby O: Psychotropic effects of caffeine in man. IV. Quantitative and qualitative differences associated with habituation to coffee. Clin Pharmacol Ther *10:*489, 1969.

Goldstein A, Wallace ME: Caffeine dependence in schoolchildren? Exp Clin Psychopharmacol *5:*388, 1997.

Greden JF, Fontaine P, Lubetsky M, Chamberlin K: Anxiety and depression associated with caffeinism among psychiatric inpatients. Am J Psychiatry *135:*963, 1978.

Greden JF, Victor BS, Fontaine P, Lubetsky M: Caffeine-withdrawal headache: A clinical profile. Psychosomatics *21:*411, 1980.

Griffiths RR, Bigelow GE, Liebson IA: Human coffee drinking: Reinforcing and physical dependence producing effects of caffeine. J Pharmacol Exp Ther *239:*416, 1986.

*Griffiths RR, Woodson PP: Caffeine physical dependence: A review of human and laboratory animal studies. Psychopharmacology *94:*437, 1988.

Griffiths RR, Woodson PP: Reinforcing effects of caffeine in humans. J Pharmacol Exp Ther *246:*21, 1988.

Griffiths RR, Evans SM, Heishman SJ, Preston KL, Sannerud CA, Wolf B, Woodson PP: Low-dose caffeine physical dependence in humans. J Pharmacol Exp Ther *255:*1123, 1990.

*Griffiths RR, Mumford GK: Caffeine—A drug of abuse? In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Hughes JR, Amori G, Hatsukami DK: A survey of physician advice about caffeine. J Subst Abuse *1:*67, 1988.

Hughes JR, Higgins ST, Bickel WK, Hunt WK, Fenwick JW, Gulliver SB, Mireault GC: Caffeine self-administration, withdrawal, and adverse effects among coffee drinkers. Arch Gen Psychiatry *48:*611, 1991.

*Hughes JR, Oliveto AH, Helzer JE, Higgins ST, Bickel WK: Should caffeine abuse, dependence, or withdrawal be added to DSM-IV and ICD-10? Am J Psychiatry *149:*33, 1992.

Hughes JR, Oliveto AH, Helzer JE, Bickel WK, Higgins ST: Indications of caffeine dependence in a population-based sample. In *Problems of Drug Dependence, 1992: Proceeding of the 54th Annual Meeting The College on Problems of Drug Dependence, Inc*., L Harris, editor. NIDA Research Monograph 132, US Dept of Health and Human Services, National Institute on Drug Abuse, Rockville, MD, 1993.

Hughes JR, Oliveto AH: A systematic survey of caffeine intake in Vermont. Exp Clin Psychopharmacol *5:*393, 1997.

Infante-Rivard C, Fernández A, Gauthier R, David M, Rivard G: Fetal loss associated with caffeine intake before and during pregnancy. JAMA *270:*2940, 1993.

James JE, Stirling KP, Hampton BAM: Caffeine fading: Behavioral treatment of affeine abuse. Behav Therapy *16:*15, 1985.

*James JE: *Understanding Caffeine: A Biobehavioral Analysis.* Sage, Thousand Oaks, CA, 1997.

Liguori A, Hughes JR, Grass JA: Absorption and subjective effects of caffeine from coffee, cola and capsules. Pharmacol Biochem Behav *58:*721, 1997.

Nehlig A, Daval J, Debry G: Caffeine and the central nervous system: Mechanism of action, biochemical, metabolic and psychostimulant effects. Brain Res Brain Res Rev *17:*139, 1992.

Parsons WD, Niems AH: Effect of smoking on caffeine clearance. Clin Pharmacol Ther *24:*40, 1978.

Pedersen N: Twin similarity for usage of common drugs. In *Twin Research 3, Part C: Epidemiological and Clinical Studies*, L Gedda, P Parisi, W Nance, editors. Alan R. Liss, New York, 1981.

Schuh KJ, Griffiths RR: Caffeine reinforcement: The role of withdrawal. Psychopharmacology *130:*320, 1997.

Silverman K, Evans SM, Strain EC, Griffiths RR: Withdrawal syndrome after the double-blind cessation of caffeine consumption. N Engl J Med *327:*1109, 1992.

Stanton CK, Gray RH: Effects of caffeine consumption on delayed conception. Am J Epidemiol *142:*1322, 1995.

Strain EC, Mumford GK, Silverman K, Griffiths RR: Caffeine dependence syndrome: Evidence from case histories and experimental evaluations. JAMA *272:*1043, 1994.

Weber JG, Ereth MH, Danielson DR: Perioperative ingestion of caffeine and postoperative headache. Mayo Clin Proc *68:*842, 1993.

11.5 CANNABIS-RELATED DISORDERS

WAYNE MACFADDEN, M.D. AND GEORGE E. WOODY, M.D.

Known in central Asia and China for at least 4000 years, the Indian hemp plant *Cannabis sativa* is a hardy, aromatic annual herb (Fig. 11.5-1). The bioactive substances derived from it are collectively referred to as cannabis. By most estimates, cannabis remains the world's most commonly used illicit drug, with approximately 200 to 300 million regular users. It occupies fourth place in worldwide popularity among psychoactive drugs, after caffeine, nicotine, and alcohol.

FIGURE 11.5-1 Marijuana *(Cannabis sativa)*.

Cannabis sativa is widely cultivated for its fiber, which is used to make rope and cloth; for its seeds, which are used to make oil; and for its psychoactive resin. This resin contains over 60 structurally similar compounds called cannabinoids, of which D_9-tetrahydrocannabinol (THC) is responsible for most of its psychoactive effects. The term *marijuana* most commonly refers to the upper leaves, flowering tops, and stems of the plant, which are cut, dried, and chopped and usually formed into cigarettes. Hashish is the dried black-brown resinous exudate from the tops and undersides of the leaves of the female plant. Other names for cannabis or its products include *bhang, charas, dagga*, and *ganja*; common slang terms are “grass,” “pot,” and “weed.”

The THC content and concentration of different parts of the plant and between species varies greatly. The effects of THC depend on dosage, frequency, and route of administration; setting; and the experience and expectations of subjects. Most marijuana purchased illicitly varies from about 1 to 5 percent THC content; sinsemilla, a more potent variety made from just the buds and flowering tops of female plants may contain 7 to 14 percent THC. Hashish, a sticky resin obtained from the female plant flowers, contains up to 10 percent THC, and hashish oil, a concentrated distillate of hashish, has been assayed at up to 15 to 70 percent THC. In addition to genetic variability, the amount of THC produced by the plant is related to environmental conditions such as the amount of sunlight received, humidity, and soil condition. Other factors that affect the amount of THC actually consumed are smoking technique, the amount destroyed by pyrolysis (about 75 percent), and how quickly the drug is used, because THC deteriorates by about 5 percent per month at room temperature. Assertions that the marijuana available today is much more potent than it was 15 to 20 years ago are difficult to validate. THC content in some specimens confiscated in the mid-1970s was reported to be 5 percent or higher, which most investigators consider relatively potent.

Cannabis is most commonly smoked in marijuana cigarettes, or joints, and experienced users inhale deeply and hold their breath for as long as possible to extract the THC. Amounts ranging from 2 to 50 percent of the THC in the smoke may be absorbed in this manner. Its intoxicating effects may be increased by mixing other drugs or chemicals into the cigarette, such as opium, cocaine paste, or phencyclidine (PCP). It may also be eaten and is often baked in lipid-rich foods, such as brownies.

HISTORY

Archeological evidence suggests that cannabis was introduced to western Europe in approximately 500 BC. There are reports of its widespread use for pleasure-seeking purposes in the early 19th century by Napoleon's soldiers in Egypt. Although previous knowledge of its use existed in the United States, it may have been more widely introduced by Mexican immigrants in the 1920s. Soon afterward Western states pressured the federal government to control cannabis use because it was linked to violence from foreign (often unwelcome) laborers, who were allegedly growing the plant. The Marijuana Tax Act of 1937 established government control over sale and transfer of the drug, and no stamps or licenses were available for its public use. During that time its use was associated with jazz musicians or minority groups who were not well assimilated into the overall culture.

Marijuana's image shifted during the 1960s when the youthful counterculture rediscovered cannabis. It became associated with social protest, and its use spread rapidly throughout the general society. Occasional use, even by adolescents, became common. The use of cannabis became “normalized” in many parts of the United States culture, a fact contributing to the initial difficulty many investigators encountered in associating cannabis with health problems. The favorable attitude toward cannabis reached a peak in about 1978. Since that time its use declined yearly in the overall population until 1993 when it began rising again. In attempts to suppress its use, penalties have risen against users and dealers, but enforcement varies widely among states and municipalities. As possible medicinal uses of cannabis are being explored, with strong advocates for its use in acquired immune deficiency syndrome (AIDS) patients and patients receiving chemotherapy, attempts to provide compassionate use have been undertaken by various constituencies. In 1996, referendums in California and Arizona approved ballot measures allowing the use of marijuana as a medical treatment. In California, this measure allowed possession and use of marijuana by patients or caregivers if recommended by two medical doctors. Controversy continues, as some lawmakers and law enforcement agencies have attempted to curb this practice by enacting preventive legislation to prosecute physicians who recommend use of substances such as cannabis that lack approval of the federal Food and Drug Administration (FDA).

The World Health Organization (WHO) and other governmental advisory and regulatory agencies have consistently maintained that cannabis is a drug requiring close monitoring and stringent control. In 1965 WHO declared that “the harm to society derived from abuse of cannabis rests in the economic consequences of the impairment of the individual's social functions and his enhanced proneness to asocial and antisocial behavior.” In 1969 WHO considered cannabis not physically habit forming but a drug of dependence and recommended keeping it under legal control. The classification of cannabis as a highly controlled (class I) substance by the United States Drug Enforcement Administration is consistent with that view.

Comparative Nosology The diagnostic criteria for cannabis intoxication in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) are similar. Both require at least one of the following four signs of intoxication: conjunctival injection, increased appetite, dry mouth and tachycardia, though DSM-IV requires them to develop within 2 hours of cannabis use. (1) Maladaptive or dysfunctional behavior or (2) psychological or perceptual changes are also required for both manuals. However, ICD-10 identifies 12 specific symptoms for the examiner to choose (one) from, where DSM-IV does not specify but notes it should be “clinically significant.”

DSM-IV does not recognize a clinically significant cannabis withdrawal state. Withdrawal from cannabis is identified in ICD-10, but definite diagnostic criteria are

omitted. It is noted as lasting several hours to up to 7 days, and describes several nonspecific signs and symptoms.

EPIDEMIOLOGY

Prevalence and Recent Trends The Monitoring the Future survey of adolescents in school indicates recent increases in lifetime, annual, current (use within the past 30 days) and daily use of marijuana by eighth and tenth graders, continuing a trend that began in the early 1990s. In 1996, 23.1 percent of eighth graders and 39.8 percent of tenth graders reported lifetime marijuana use. By 1996, past-month marijuana use had increased over 250 percent since 1991 for eighth graders and over 150 percent since 1992 for 10th graders. Lifetime use for high school seniors peaked at 60 percent in 1979, then began decreasing yearly until 1993, when rates began to rise again. Some investigators believe that softening attitudes about the hazards associated with cannabis may be responsible for these recent increases. In evaluating these trends, one should note that occasional use does not imply abuse or dependence. The percentage of users who fulfill criteria for either of these diagnoses is unclear at this time, though it is likely to be relatively low, somewhat analogous to the case with alcohol.

Another measure of the prevalence of marijuana use comes from the National Household Survey on Drug Abuse, a population-based random sample of households throughout the United States. Marijuana was the most commonly used illicit drug in the 1995 study, which found that approximately 77 percent of current illicit drug users were marijuana or hashish users. About one third (31 percent) of the population reported that they had used marijuana one or more times in their lifetime, 8 percent had used it in the past year, and 5 percent had used it in the past month. Those percentages translate to approximately 66 million persons who had used marijuana in their lifetime, 18 million in the past year, and 10 million in the past month. Lifetime prevalence of marijuana use increased with each age group until 34 years, then decreased gradually. Those aged 18 to 21 were the most likely to have used marijuana in the past year (25 percent) or the past month (14 percent), and use was lowest among those age 50 or older, where it was at or below 1 percent.

Demographic Correlates The rate of past year and current marijuana use by males was almost twice the rate for females overall among those age 26 and older. This gap between the sexes narrows with younger users; at ages 12 to 17, there are no significant differences.

Race and ethnicity were also related to marijuana use, but the relationships varied by age group. Among those ages 12 to 17, whites had higher rates of lifetime and past-year marijuana use than blacks. Among 17- to 34-year-old adults, whites reported higher levels of lifetime use than blacks and Hispanics. But among those 35 and older, whites and blacks reported the same levels of use. The lifetime rates for black adults were significantly higher than those for Hispanics. There were no differences in past-month use between whites and blacks among those 12 to 34 years old. Among females, whites reported the highest levels of lifetime use (30 percent), followed by blacks (20 percent) and Hispanics (15 percent). Variations in patterns of marijuana use by geographic region were small as were variations in use between metropolitan and nonmetropolitan areas.

For younger adults (18 to 25), marijuana use was similar across all educational levels, except for lower rates of current use among college graduates. For those ages 26 to 34, current use decreased significantly at higher educational levels. Individuals with less than a high school diploma had higher rates of past-year and current marijuana use than those all in other education categories. Among adults 35 and older, however, this trend was reversed for lifetime use; 33 percent of college graduates reported having ever used marijuana versus 12 percent of those with less than a high school diploma.

Among the total adult surveyed population, past-year and past-month marijuana use were substantially higher among the unemployed than in those in any other employment category.

Current users of marijuana were more likely than nonusers to drink alcohol, smoke cigarettes, and use other illicit drugs. This finding was most marked in younger adults and adolescents; in every age group up to age 35, current marijuana users were 10 to 12 times more likely to use other illicit drugs than those who were not current users.

PHARMACOLOGY

The THC dose needed to produce pharmacological effects in humans from smoking range from 2 to 22 mg. THC is lipid soluble and rapidly absorbed after inhalation. It is highly protein bound and quickly redistributed from blood into other tissues. The percentage of ingested THC that reaches the brain is small; about 1 percent penetrates the blood-brain barrier. The THC that reaches the liver is almost completely metabolized, primarily into active 11-hydroxy-THC and inactive 9-carboxy-THC. More-extensive metabolism in the liver converts 11-hydroxy-THC to many inactive metabolites, including 11-norcarboxy-D-THC, which is detected minutes after smoking. It is the most abundant metabolite in plasma and urine, and also the primary cannabinoid metabolite excreted in the urine; typically it can be detected 2 to 3 days after smoking a single cannabis cigarette. Blood levels of THC peak in about 30 minutes and then decline precipitously as it redistributes through the body to lipid-rich tissues, reaching nearly undetectable levels within 3 to 4 hours.

After some equilibrium between concentrations in the blood and other tissues is established, THC slowly and unevenly reenters the bloodstream from its tissue stores. Its concentration declines slowly in this second phase, with a half-life of 3 days. This low concentration of THC or its active 11-hydroxy-THC metabolite is believed to be below the threshold to produce an effect and is thus clinically insignificant. However, the half-life of these metabolites is at least 50 hours. About two thirds of the drug is excreted via the enterohepatic circulation into the feces, and the remaining one third is removed through the kidney. Most of the metabolites of THC are produced by the liver. Among these compounds, 11-norcarboxy-THC has the highest concentration in the urine and is the metabolite usually screened for in routine toxicological analyses. This and other cannabinoid metabolites can be detected in the urine for 2 to 3 days after casual use; for daily heavy users, detectable levels can persist for up to 4 weeks.

Cardiovascular and central nervous system (CNS) effects, such as mood-altering properties, begin less than 1 minute after inhalation. Peak clinical effects may be delayed for 20 to 30 minutes and persist for at least 2 to 3 hours. Immediately after smoking marijuana, plasma concentrations of cannabinoids are high while effects are low; this reverses at later times. Peak blood concentrations of THC are reached rapidly, within 10 minutes of smoking and decline to 5 to 10 percent of their initial level within 1 hour. Because of this delay between peak blood concentrations and peak drug effects, establishing a relationship between blood levels of THC or its metabolites and the degree of impairment has been difficult. The precise method and time of cannabis use must be known before any useful prediction of impairment from THC can be made. The apparent half-life of THC in both plasma and body fat is about 4 days. However, concentrations of THC in blood correlate only modestly with cannabis intoxication, and there is no consistently demonstrated correlation between concentration of THC in blood and effects on performance.

The pharmacological effects of orally ingested cannabis begin after 30 minutes, peak in 2 to 3 hours, and last 3 to 6 hours. Residual effects, such as subtle changes in mood and fine motor control, can be measured for a longer period if very sensitive testing procedures are used.

An oral dose of approximately 20 mg of marijuana or smoking a cigarette containing about 0.5 to 2 percent THC usually produces intoxication. Orally ingested marijuana requires about three times as much THC as smoked marijuana to produce equivalent effects because only 3 to 6 percent of ingested THC is absorbed.

THC demonstrates weak barbiturate-like actions such as anticonvulsant activity and opioid-like effects, including analgesia, increased catecholamine synthesis, hypothermia, and antidiarrheal activity. Increased limbic system activity has been noted with cannabis use, suggesting that THC may stimulate pleasure-reward mechanisms in the brain.

Cannabis is often used in combination with other drugs. It may alter the effects of amphetamines, atropine, barbiturates, clomipramine (Anafranil), cocaine, ethanol, nicotine, opiates, and phencyclidine. Because of shared hepatic metabolic systems, ethanol and phenobarbital (Donnatal, Quadrival) can inhibit metabolism of THC. Similarly, THC can slow the metabolism of a variety of drugs, including theophylline (Theo-Dur), ethanol, and pentobarbital (Nembutal). Some limited cross-tolerance exists between THC and CNS-depressant drugs. However, there is no evidence that THC is useful for detoxification from sedatives, although they both can enhance or prolong the other's behavioral and psychological effects.

Mechanism of Action The creation and radiolabeling of certain cannabinoids led to identification of a receptor in rat brain membranes. A single binding site was identified that displayed saturable and reversible binding and selectivity for cannabinoids. The pharmacological potency of cannabinoids correlates with their affinity for the cannabinoid binding site. Receptor binding was also found in the peripheral B lymphocyte–enriched areas including the marginal zone of the spleen, the nodular corona of Peyer's patches, and the cortex of the lymph nodes.

Another advance has been the recent isolation of an endogenous cannabinoid-like ligand within the brain, named anandamide, from a Sanskrit word meaning “bliss.” Soon after, a cannabinoid antagonist was discovered that antagonizes cannabinoid-induced inhibition of adenyl cyclase and smooth muscle contraction. This recent

progress suggests the presence of a cannabinoid neurochemical pathway. The function of such a system is unclear, but cannabinoids exert many of their actions by influencing several neurotransmitter systems and their neuromodulators. These include acetylcholine, dopamine, g-aminobutyric acid (GABA), histamine, serotonin, norepinephrine, opioid peptides, and prostaglandins. Cannabinoids enhance the formation of norepinephrine, dopamine, and serotonin. Notably, they stimulate the release of dopamine from rat brains by activating opiate receptors; after THC infusions, dopamine levels in the nucleus accumbens increase as much as twofold, analogous to the surge seen with other addictive drugs. GABA turnover is enhanced by cannabinoids. Catalepsy can result from the interaction of THC with neurotransmitter systems in the basal ganglia. Studies of catalepsy and depression of spontaneous locomotor activity caused by cannabinoids suggest that these effects are mediated by acetylcholine, GABA, and prostaglandins.

Despite these advances, interpretation of the actions of cannabinoids on neurotransmitter systems is often unclear, as evidence suggests that cannabinoids both inhibit and stimulate uptake of transmitters. Recent evidence suggests that few, if any, irreversible effects on brain chemistry are caused by THC administration.

Cannabinoid receptor location and density in animal models has corresponded with its clinical effects in humans. The highest density of receptors occurs in the basal ganglia and molecular layer of the cerebellum, which correlate with its interference in motor coordination. Intermediate levels of binding were found in the hippocampus, the dentate gyrus, and layers I and IV of the cortex, consistent with cannabinoid effects on short-term memory and cognition. Low receptor density is noted in the brainstem areas controlling cardiovascular and respiratory functions, which is consistent with the cannabinoids' lack of lethality.

After binding to receptors, cannabinoids also produce effects through second-messenger systems including inhibition of adenyl cyclase and calcium channels, and by possibly enhancing potassium channels activity. In summary, the biological bases for the multiple effects of cannabis are beginning to be understood, but as expected, they are complex and not yet fully elucidated.

DIAGNOSIS AND CLINICAL FEATURES

Patterns of use vary widely; the most common is intermittent use of marijuana cigarettes, such as smoking one or two joints on a weekly or monthly basis, often on social occasions. As with alcohol, a small proportion of cannabis users develop the pattern of intermittent or daily use of high doses that is typically associated with abuse or dependence. The proportion of users who progress to dependence is unknown but is probably similar to alcohol.

Marijuana has mild-to-moderate reinforcing effects, and benign experiences with it may lead users to try more reinforcing drugs in the belief that drug effects are pleasurable and not to be feared. In fact, initial marijuana usage is a common behavioral pattern in patients who eventually progress to so-called harder drugs such as opiates, stimulants, and stronger psychedelics. It is unlikely that the relationship is causal, however. The purchase or use of marijuana may imply a willingness to use other illicit substances and may also put the user in contact with people who distribute them. From self-report data, marijuana use has been associated with problematic alcohol use and a pattern of general deviance that leads to poor workplace performance.

New studies have postulated a neurobiological basis for the “gateway hypothesis,” in which smoking marijuana is thought to cause some people to abuse harder drugs. Marijuana and heroin both produce dopamine surges in rat brains by activating opiate receptors; marijuana, however, presumably does so indirectly by causing the release of an endogenous opiate. Some speculate that marijuana may thus prime the brain to seek substances like heroin that act in a similar way. Additionally, long-term cannabis use increases release of corticotropin-releasing factor during antagonist-precipitated withdrawal in rats. As corticotropin-releasing factor is believed to be a key element of withdrawal syndromes from alcohol, opiates, and cocaine, causing stress reactions and anxiety, some also speculate that marijuana users may seek these harder drugs to provide relief from the similar neurochemically induced stress and anxiety noted in cannabis withdrawal syndromes.

DSM-IV lists the cannabis-related disorders (Table 11.5-1) but has specific criteria only for cannabis intoxication (Table 11.5-2). The diagnostic criteria for the other cannabis-related disorders are general and are contained in the DSM-IV sections that focus on the major phenomenological symptom—for example, cannabis-induced psychotic disorder in the schizophrenia and other psychotic disorders.

Cannabis use disorders
Cannabis dependence
Cannabis abuse
Cannabis-induced disorders
Cannabis intoxication
Specify if:
With perceptual disturbances
Cannabis intoxication delirium
Cannabis-induced psychotic disorder, with delusions
Specify if:
With onset during intoxication
Cannabis-induced psychotic disorder, with hallucinations
Specify if:
With onset during intoxication
Cannabis-induced anxiety disorder
Specify if:
With onset during intoxication
Cannabis-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.5-1 DSM-IV Cannabis-Related Disorders

A. Recent use of cannabis.
B. Clinically significant maladaptive behavioral or psychological changes (e.g., impaired motor coordination, euphoria, anxiety, sensation of slowed time, impaired judgment, social withdrawal) that develop during, or shortly after, cannabis use.
C. Two (or more) of the following signs, developing within 2 hours of cannabis use:
(1) conjunctival injection
(2) increased appetite
(3) dry mouth
(4) tachycardia
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.
Specify if:
With perceptual disturbances

Reprinted with permission from American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.5-2 DSM-IV Diagnostic Criteria for Cannabis Intoxication

Cannabis Dependence and Cannabis Abuse DSM-IV includes the diagnoses of cannabis dependence and cannabis abuse (see Table 11.1-3 and Table 11.1-8). The experimental data clearly show tolerance to many of the effects of cannabis with sustained use of high doses. Tolerance develops to most of the physical effects of cannabis, including tachycardia, decreased skin temperature, increased body temperature, decreased intraocular pressure, sleep disturbance (decreased rapid eye movement [REM] sleep), electroencephalogram (EEG) changes (increased alpha waves), and impairment of performance on psychomotor tests. There is less agreement about tolerance to the common mood and behavioral changes, but progressive loss of the so-called high has been reported.

Cannabis Intoxication DSM-IV formalizes the diagnostic criteria for cannabis intoxication (Table 11.5-2). These criteria specify that the diagnosis can carry the phrase “with perceptual disturbances.” If perceptual disturbances are present but reality testing is impaired, the diagnosis becomes cannabis-induced psychotic disorder.

Acute intoxication due to use of cannabinoids as outlined in ICD-10 closely mirrors the descriptive themes in DSM-IV. Cannabis intoxication commonly heightens the user's sensitivity to external stimuli, reveals new details, makes colors seem brighter and richer than before, and makes time seem to pass more slowly. With high doses, the user may also experience depersonalization, derealization, illusions, hallucinations, suspiciousness or paranoid ideation.

Cannabis Intoxication Delirium Cannabis intoxication delirium is a DSM-IV diagnosis (Table 10-22). The delirium associated with cannabis intoxication is

characterized by marked impairment on cognition and performance tasks. Even modest doses of cannabis impair memory, reaction time, perception, motor coordination, and attention. High doses that also impair the user's level of consciousness have even more marked effects on these cognitive measures. However, a cannabis-induced, longer-lasting toxic-organic delirium characterized by confusion with disorganization of thought processes, affective lability, delusions, and hallucinations has been reported. This reaction, similar to the delirium produced by other psychomimetics, hallucinogens, or toxins, may last up to 10 days. Whether cannabis is the primary etiological agent in these cases has been questioned by some investigators, however.

Cannabis-Induced Psychotic Disorder High doses of cannabis are more likely than low doses to induce brief psychotic symptoms such as persecutory delusions or auditory and visual hallucinations, especially in persons with underlying psychiatric disorders. Such cases may fulfill the DSM-IV criteria for cannabis-induced psychotic disorder (see Table 13.3-4). It is uncertain whether persons with unstable character structures are more susceptible to these brief psychotic episodes.

Cannabis-induced psychotic disorder is rare; transient paranoid ideation is more common. Florid psychosis is somewhat common in countries where persons have long-term access to cannabis of particularly high potency. The psychotic episodes are sometimes referred to as “hemp insanity.” Cannabis use is rarely associated with a bad-trip experience, such as that associated with hallucinogen intoxication. No persistent cannabis psychosis has been identified, even in chronic heavy users. However, it appears likely that cannabis can exacerbate schizophrenia, as is the case with other drugs that have hallucinogenic properties, though there is no conclusive evidence that cannabis is a causative factor in the development of schizophrenia. Thus, despite worldwide reports associating cannabis with mental illness, chronic psychosis has not yet been reliably and consistently demonstrated to result from cannabis use. Chronic psychotic disorders that may be precipitated by cannabis appear to be related to premorbid vulnerability or psychopathology.

Cannabis-Induced Anxiety Disorder Cannabis-induced anxiety disorder is a common diagnosis (see Table 15.6-18) for acute cannabis intoxication, which in many persons induces short-lived anxiety states that are often provoked by paranoid thoughts. In such circumstances, panic attacks may be induced, based on ill-defined and disorganized fears. The appearance of anxiety symptoms is correlated with the dose and is the most frequent adverse reaction to the moderate use of smoked cannabis. Some cannabis users report occasional unpleasant adverse experiences, most often described as anxiety reactions of mild-to-moderate intensity. Inexperienced users are much more likely to experience anxiety symptoms than are experienced users.

Cannabinoid Withdrawal State ICD-10 notes cannabinoid withdrawal state; DSM-IV does not. It is described as an “ill-defined syndrome for which definitive diagnostic criteria cannot be established at present,” substantiating the lack of systematic nosological research in this area. Symptoms and signs noted in ICD-10 include anxiety, irritability, tremor of outstretched hands, sweating, and muscle aches.

Cannabis-Related Disorder Not Otherwise Specified DSM-IV does not formally recognize cannabis-induced mood disorders; therefore, such disorders are classified as cannabis-related disorders not otherwise specified (Table 11.5-3). Cannabis intoxication can be associated with depressive symptoms, although such symptoms may suggest long-term cannabis use. Hypomania, however, is a common symptom in cannabis intoxication.

The cannabis-related disorder not otherwise specified category is for disorders associated with the use of cannabis that are not classifiable as cannabis dependence, cannabis abuse, cannabis intoxication, cannabis intoxication delirium, cannabis-induced psychotic disorder, or cannabis-induced anxiety disorder.

Table 11.5-3 DSM-IV Diagnostic Criteria for Cannabis-Related Disorder Not Otherwise Specified

DSM-IV also does not formally recognize cannabis-induced sleep disorders or cannabis-induced sexual dysfunction; therefore, both are classified as cannabis-related disorders not otherwise specified. When either sleep disorder symptoms or sexual dysfunction symptoms are present and related to cannabis use, they almost always resolve within days or a week after the cessation of cannabis use.

Flashbacks Flashbacks, in which feelings and perceptions experienced in the intoxicated state are suddenly thrust into consciousness in the nondrugged condition, have also been reported with cannabis use, although not as often as with lysergic acid diethylamide (LSD). It has been suggested that flashbacks result from intermittent release of psychoactive components from the CNS, where they are stored during periods of active use, but that explanation remains highly speculative. A few clinical reports suggest that marijuana use may precipitate flashbacks in persons who have previously used LSD.

Amotivational Syndrome An amotivational syndrome associated with chronic cannabis use was described in the older clinical literature from the Middle East, the Orient, and the United States. The syndrome is marked by apathy, poor concentration, social withdrawal, and loss of interest in achievement. Those features may correlate with the reversible decrement in cerebral blood flow that has been documented as an effect of marijuana. However, most of the reports are not rigorously scientific and lack controls that distinguish between the effects of cannabis and preexisting psychological and social conditions. Subsequent reports using different populations and better scientific methods have failed to demonstrate the syndrome. Several authors have noted that it is difficult to determine which came first, the drug or the amotivation. Most plausible perhaps is the suggestion that in certain persons the pharmacological effects of the drug interact with psychological and social factors to retard motivation and productivity.

Thus, the direct causal role of marijuana in the amotivational syndrome has been seriously questioned. Symptoms may indicate ongoing intoxication or represent normal psychosocial variants that predispose to the use of cannabis and other substances. However, because persistent functional and structural changes in hippocampal neurons in animals subjected to long-term THC administration have been observed, the concept that a developing personality can be altered by chronic intoxication should not be entirely dismissed. In any event, cessation may lead to gradual improvement. Despite those potential adverse effects, many regard cannabis as a relatively safe drug because lethal doses are unknown in humans.

Physical Dependence Although DSM-IV provides no criteria for a cannabis withdrawal syndrome, withdrawal signs elucidated in rats after administration of the newly developed cannabinoid antagonist provide strong evidence that cannabinoids produce physical dependence. In humans, the withdrawal signs and symptoms that result from abrupt cessation of chronic high-dose cannabis abuse are not severe. They may reach their peak about 8 hours after last use and persist for 2 to 3 days. The most prominent symptoms reported are increased irritability and restlessness. Other less reliable signs and symptoms include anxiety, sleep disturbances, anorexia, perspiration, nausea, and muscle pain. Objective signs include increased body temperature, diarrhea, vomiting, weight loss, and hand tremors. No specific treatment is generally required.

Recent animal research has linked corticotropin-releasing factor to these stress-like negative affective states typical of cannabis withdrawal. Corticotropin-releasing factor concentrations are markedly elevated in the brains of chemically treated rats after antagonist-induced withdrawal. This stress hormone may mediate a similar final common pathway of anxiety and emotional stress in withdrawal states produced by various drugs of abuse (opiates, cocaine, and alcohol).

Tolerance Animal species develop tolerance to a variety of pharmacological effects, including antinociception, anticonvulsant activity, catalepsy, depression of locomotor activity, hypothermia, hypotension, corticosteroid release, and ataxia. Tolerance may reach 100-fold. Other psychoactive cannabinoids including nabilone also produce tolerance. Neuronal changes such as downregulation, conformational change, and internalization of receptors have been postulated to produce this tolerance.

Because of tolerance and learned behavior, detecting cannabis intoxication in an experienced user by motor performance may be difficult. In an inexperienced user, intoxication can be detected by many performance tests.

Several factors contribute to tolerance, including potency, expectations, environmental influences, individual differences, and frequency of use. There is evidence for the development of tolerance in humans to THC-induced decreases in intraocular pressure, sleep disturbances, and mood changes. High doses over long periods of time are necessary to produce behavioral tolerance.

ADVERSE EFFECTS

CNS THC exerts prominent effects on the CNS that are highly variable and depend on the user, the dose, and the environment. An initial stimulant effect commonly produces an increased sense of well-being and euphoria and is associated with spontaneous laughter, disinhibition, or quiet reverie. Some experience heightened imagination and creative thinking. Mood changes vary, and anxiety and depression may be induced. These initial effects are often followed by relaxation or lethargy and drowsiness, especially at higher doses. Many users report perceptual and sensory changes, such as a sense that time is passing slowly, accentuation of auditory and visual perceptions, or actual sensory distortions, occasionally involving hallucinations. Because of these perceptual changes (usually only at very high doses), some consider marijuana a hallucinogen with very low potency. Dry mouth and throat and increased hunger are also common.

Most psychoactive adverse marijuana-induced effects disappear when the acute intoxication has ended, depending on individual vulnerability, environmental factors (whether the user is in a nonthreatening area), and dose. Less common adverse effects are changed sensations of bodily perceptions, depersonalization, derealization, acute panic, and frank paranoia.

THC has well-known deleterious effects on higher cognitive functions, related to the duration of cannabis use. It reliably impairs short-term memory but not retrieval of previously learned facts. Attention span, recall, the ability to store knowledge, and the ability to perform tasks requiring multiple mental steps are also adversely affected. In addition, the ability to verbalize is often diminished. The term “temporal disintegration” has been coined to characterize THC's effects on the CNS. Intoxicated persons may have difficulty integrating earlier experiences, expectations, and current perceptions into goal-directed action (e.g., learning new material is usually impaired). College students who used marijuana regularly had impaired skills related to attention, memory, and learning 24 hours after they had last used the drug. Acute neurophysiological changes include suppression of REM sleep and diffuse slowing of background activity on the EEG. Those acute effects last for minutes up to a few hours, depending on dose and individual sensitivity to the drug. However, they can measurably impair performance if higher levels of cognitive or psychomotor skills are required.

THC increases cerebral blood flow in experienced users, correlating with intoxication. Acute THC also increases cerebral metabolic rate, primarily in the cerebellum and prefrontal cortex. Chronic effects have been more elusive, but long-term use has been hypothesized to impair the frontal lobe, which functions in the temporal organization of behavior. Frontal lobe impairment is consistent with the altered perception of time and also with cerebral blood flow studies that demonstrate greatest effects in the frontal lobe region. Some studies suggest that impairment assessed by sensitive measures of brain function can be detected after 5 years of use, but research has yet to demonstrate conclusively that chronic heavy marijuana use results in cognitive deficits that persist after prolonged abstinence. Studies of cerebral morphological changes in humans have not been compelling, though animal studies have shown structural damage to the hippocampus, a structure critical in learning and memory.

Quantitative EEG studies have reported increased alpha power, decreased alpha frequency, and decreased beta activity following acute exposure, which is consistent with a state of drowsiness. Quantitative EEG changes such as increased frontal-central theta wave activity in association with increased alpha wave activity may occur following long-term (5 years or more) exposure.

Psychomotor A variety of motor skills are affected by acute use, including decreased muscle strength, hand steadiness, and performance of simple motor tasks and reflex responses. At higher doses coordination and balance are impaired in a dose-related way. These effects may last several hours after subjective effects have subsided, and they are additive to those produced by alcohol or other sedatives.

Simple reaction time has not been shown to be adversely affected, but response to complex and unforeseen situations is impaired. As these skills are functional components important to driving, they may be related to marijuana-induced impairment of automobile driving. Some data suggest that marijuana produces a significant decrease in speeding and other risk-taking behavior, effects opposite those that often result from alcohol consumption.

Few studies are available on marijuana's effect in causing auto or airplane accidents. In one report of roadside sobriety tests, trained observers found that 94 percent of subjects were intoxicated 90 minutes after smoking moderate amounts of marijuana, and 60 percent after 2 hours. Despite numbers like these, several studies indicate that marijuana use by itself is a relatively minor risk factor in fatal traffic accidents. Although some observations suggest an increased accident risk for THC-positive drivers, interpretation of the data is confounded by concurrent alcohol use, which may synergistically worsen impairment. Some authors speculate that because persons with substance use disorders commonly use cannabis, marijuana testing may be of value as a monitor for more serious types of drug use—particularly, heavy drinking.

Several controlled studies indicate that cannabis intoxication impairs the ability to fly airplanes more than driving ability. That finding can be explained by the increased complexity of the tasks involved in flying. For both flying and driving, no correlation has been established between the degree of impairment and blood or urine levels of THC, rendering it difficult to establish reliable legal levels of intoxication as has been done for alcohol.

Cardiovascular Effects One of the first effects of the drug is a predictable dose-dependent increase in heart rate, probably due to an inhibition of vagal tone. Along with conjunctival reddening from dilation of blood vessels, the heart rate increase correlates with the appearance and duration of psychic effects as well as plasma concentrations of the drug. Myocardial oxygen demand is increased, and exercise tolerance decreased. These effects can lead to myocardial ischemia if a person with coronary artery disease exercises when intoxicated. Cannabinoids may exacerbate preexisting cardiovascular conditions such as angina and congestive heart failure. Orthostatic hypotension can also occur, especially at high doses. Tolerance appears to develop to all these cardiovascular effects.

Pulmonary Effects Smoking is by far the most popular route of cannabis administration. Although not yet subject to the same large-scale, long-term epidemiological analyses that identified tobacco as a carcinogen, cannabis is known to induce pulmonary pathology. Chronic smokers often experience bronchitis, pharyngolaryngitis, and asthma, most likely from marijuana smoke's highly irritating effect on the bronchial epithelium. Marijuana cigarettes contain far more tar (particulates) and respiratory irritants than tobacco. Tar produced by burning marijuana is more carcinogenic to animals than that derived from tobacco, and some workers predict that marijuana smoking will result in malignancies. Adding to the long-term toxic potential is the fact that marijuana smoke is usually inhaled longer and more deeply than cigarette smoke, delivering up to four times more tar to the lungs.

Heavy marijuana smoking causes mild, but significant, large airway obstruction, probably through chronic irritation or inflammation of the bronchial lining. THC itself causes bronchodilation to which little tolerance develops.

Immunological and Carcinogenic Effects Marijuana smoke inhibits pulmonary antibacterial defense systems, primarily alveolar macrophages, neutrophils, and lymphocytes. Cannabinoids may also suppress cellular and humoral immune responses in animals. The clinical significance of these effects has not been demonstrated.

In vitro and in vivo animal studies using relatively high doses of cannabinoids have shown them to have mutagenic and carcinogenic effects and to impair the synthesis of nucleic acids and proteins. While it is impossible to dismiss the potential clinical significance of those reports, no cytogenic abnormalities have yet been consistently documented in human marijuana smokers.

Hormonal and Reproductive Effects In animals, cannabis disrupts all phases of reproductive functioning by direct action on both the hypothalamic-pituitary axis and the gonads. In humans, all aspects of these effects have been suspected but are difficult to confirm. Cannabis has been reported to reversibly inhibit spermatogenesis, with a reduction in the number of sperm cells and an increased prevalence of abnormal cells. Decreased levels of testosterone and decreased size of the testes and prostate after heavy use are also seen but are believed to be reversible. A single marijuana cigarette can suppress plasma leutinizing hormone during the luteal phase of the menstrual cycle. This may account for the higher frequency of anovulatory cycles often associated with marijuana smoking. These phenonema may be due to THC's central effect of interfering with the release of gonadotropin-releasing hormone at a suprapituitary site, thereby disrupting the release of circulating luteinizing and follicle-stimulating hormones.

Cannabinoids cross the placental barrier and appear in maternal milk. Experimentally, cannabis is teratogenic at high doses in some species of animals. However,

cannabis's deleterious effects on the human fetus have been difficult to assess because of the frequent concurrent use of other drugs, cigarettes, and alcohol. Studies that have attempted to control for these variables have frequently associated marijuana with low birth weight. Its effects on other types of human fetal abnormalities have been reported, but the correlation is not well documented.

LABORATORY EXAMINATION

Urine testing for marijuana and other drugs has become common in many settings, including drug treatment programs and places of employment. Most laboratories use the enzyme-multiplied immunoassay (EMIT), although a radioimmunoassay is also commonly used. Both tests, while relatively sensitive and inexpensive, provide an unacceptable level of false-positive results. Thus confirmation by gas chromatography–mass spectrometry is routinely used. Cannabis and its metabolites may be detected in urine at the usual cutoff level of 100 ng/mL for 42 to 72 hours after the psychological effects subside. Passive inhalation that occurs under unusually crowded conditions may also reveal cannabis metabolites in the urine but only if the cutoff level used in the urine test is decreased to 20 to 25 ng/mL, a procedure that increases the frequency of false-positive results.

No clear linear relation between psychoactive effects and urine levels has been demonstrated. Urine that contains cannabis metabolites only implies that cannabis exposure occurred at an indeterminate time prior to testing. To avoid problems that may be associated with identifying very low levels of metabolites in the urine, such as can occur with passive inhalation, most laboratories use a cutoff point of 100 ng/mL or above.

TREATMENT

Persons who use marijuana are often referred for treatment. Referrals are made for persons with widely varying use patterns and treatment needs. At one extreme is the person who uses cannabis intermittently at low doses who was identified by a random drug screening test. At the other extreme is the person who uses high doses daily and meets criteria for dependence. The first person may need only periodic urine testing and infrequent supportive counseling. The second probably requires referral to a specialized, intensive drug rehabilitation program. Thus, as in other clinical situations involving substance use, treatment should begin only after a complete history is taken and a diagnosis is established. A psychiatric examination helps reveal any underlying psychopathology and determine the relation of drug use to mood states and psychiatric symptoms.

Cannabis dependence and abuse are usually treated by the psychosocial methods typically used in drug-free rehabilitation programs. These include attempts to promote realistic and rewarding alternatives to the drug and the associated lifestyles along with a commitment to abstinence from self-administered or unprescribed psychotropic drugs. Treatment usually involves a combination of interventions, including urine testing, participation in 12-step programs, education about drug effects, drug counseling, psychotherapy, and family therapy. Drug-focused group therapy is perhaps the most common treatment for all substance-related disorders, including cannabis-related disorders. Common strategies used by the group are social pressure to reinforce abstinence, teaching socialization and problem-solving skills, reducing stress and the sense of isolation often associated with drug use, relapse-prevention exercises, and varying degrees of confrontation. A trial comparing group-based relapse prevention and social support yielded results comparable to those obtained for treatment of alcohol or nicotine dependence.

Relative to those seeking treatment for alcohol or other substance use disorders, few adults seek treatment for cannabis abuse and dependence alone. Other drug abuse usually precipitates treatment and cannabis dependence is treated along with the other drug problems. However, cannabis use is one of the more common presenting problems among youth seeking treatment for substance abuse. Denial of a drug problem appears commonly in many persons with cannabis abuse or dependence, as it does in persons with other substance use disorders.

Treatment of unpleasant adverse reactions, usually anxiety, consists of calm and gentle reassurance in a warm and supportive atmosphere. Short-term use of anxiolytic agents, such as benzodiazepines, is necessary in some instances when anxiety symptoms are prominent or severe. Short-term use of low doses of antipsychotic medication may be justified if the patient has more-persistent and more-troubling symptoms, such as delusional ideas or frightening flashbacks. Treatment of toxic-delirious states is similarly supportive, symptomatic, and short-term because of their self-limited nature.

Therapeutic Uses Cannabis has been tried as a therapeutic agent for a variety of ailments. Data in this field are exceedingly difficult to collect for a variety of reasons including obtaining regulatory approval and support for research and difficulty in matching placebo against smoked marijuana. Synthetic THC (e.g., drobinol) which can be orally administered, contains the principal psychoactive component of the cannabis leaf. But cannabis contains over 400 chemicals, and it is difficult to ascertain with certainty which have therapeutic properties.

THC has antiemetic effects on the nausea and vomiting caused by chemotherapy. A synthetic oral cannabinoid (dronabinol [Marinol]) has been shown to be generally equal or superior to prochlorperazine (Compazine) but inferior to metoclopramide (Reglan). The relative efficacy of cannabinoids versus newer antiemetics such as ondansetron (Zofran) combined with dexamethasone (Decadron) has not been evaluated. The data on smoked marijuana versus these older and newer agents are limited, but reports suggest that it may be effective in a certain percentage of patients who had no benefit from older agents.

Dronabinol can also reduce the nausea and diarrhea associated with AIDS or with chemotherapy used to treat human immunodeficiency virus (HIV) infection. It also increases appetite and produces weight gain in AIDS and cancer patients, although the weight gain is not in lean body mass. As a result, THC has gained orphan drug status from the FDA to treat nausea and vomiting from chemotherapy and to stimulate appetite in AIDS patients. However, there are no controlled studies of smoked marijuana for this condition, nor any systematic studies of the effects of smoked marijuana on immunological status in HIV-infected patients. With repeated smoking, patients may be expected to experience the typical symptoms of intoxication, such as mood changes and decreases in concentration, coordination, and the ability to estimate time. Separating the undesired side effects of cannabis from its therapeutic effects has been difficult. More controlled studies are needed.

Smoked marijuana lowers intraocular pressure. This has stimulated interest in its use as a treatment for glaucoma. The mechanism of action for this feature is unknown, which limits development of similar agents. This effect is mostly short-lived (3 to 4 hours) and there is no evidence that cannabis is more effective than several other agents in treating patients suffering from glaucoma.

Preclinical and clinical studies of the use of cannabinoids in neurological and movement disorders have been reported. Currently, there is no evidence that cannabinoids are superior to available therapies for multiple sclerosis and partial spinal cord injury. Case studies have reported some benefit of smoked marijuana in treatment of dystonic states, and some evidence suggests a possible role for cannabinoids in the treatment of the epilepsies, but studies are lacking.

Oral THC has analgesic efficacy for cancer pain, but there is a narrow therapeutic margin between the doses that produce useful analgesia and those producing unacceptable adverse CNS effects. Mild hypothermia, prolongation of barbiturate anesthesia, and hyperglycemia are other disorders for which cannabis has not yet demonstrated clinical utility.

SUGGESTED CROSS-REFERENCES

An overview of the substance-related disorders, including substance abuse and substance dependence, appears in Section 11.1. Substance intoxication delirium is discussed in Chapter 10 on cognitive disorders and mental disorders due to a general medical condition. Substance-induced psychotic disorder is discussed in Section 13.3 on acute and transient psychotic disorders and culture-bound syndromes. Substance-induced anxiety disorder appears in Section 15.6.

SECTION REFERENCES

*Adams IB, Martin BR: Cannabis: Pharmacology and toxicology in animals and humans. Addict *91:*1585, 1996.

Andreasson S, Allebeck P, Engstrom A, Rydberg U: Cannabis and schizophrenia: A longitudinal study. Lancet *2:*1483, 1987.

Chait LD, Pierri J: Effects of smoked marijuana on human performance: A critical review. In: *Marijuana/Cannabinoids: Neurobiology and Neurophysiology*, A Murphy, J Bartke, editors. CRC Press, Boca Raton, FL, 1992.

Compton DR, Dewey WL, Martin BR: Cannabis dependence and tolerance production. Adv Alcohol Subst Abuse *9:*129, 1990.

Devane WA, Dysarz FA III, Johnson MR, Melvin LS, Howlett AC: Determination and characterization of a cannabinoid receptor in the rat brain. Mol Pharmacol *34:*605, 1988.

Gieringer DH: Marijuana, driving and accident safety. J Psychoact Drugs *20:*93, 1988.

Gruber AJ, Pope HG: Cannabis psychotic disorder: Does it exist? Am J Addict *3:*72, 1994.

*Hall W, Solowij N, Lemon J: The health and psychological consequences of cannabis use. In *National Drug Strategy Monograph Series*, no. 25. Australian Government Publishing Services, Canberra, 1994.

*Hall W, Solowij N: Adverse effects of cannabis. Lancet *352:*1611, 1998.

*Haney M, Ward AS, Comer SD, Foltin RW, Fischman MW: Abstinence symptoms following smoked marijuana in humans. Psychopharmacology *141:*395, 1999.

Hollister LE: Cannabis—1988. Acta Psychiatr Scand *345:*108, 1988.

Imade AGT, Ebie JC: A retrospective study of symptom patterns of cannabis induced psychosis. Acta Psychiatr Scand *83:*134, 1991.

*Johnson BA: Psychopharmacological effects of cannabis. Br J Hosp Med *43:*114, 1990.

Miller NS, Gold MS: The diagnosis of marijuana dependence. J Subst Abuse Treat *6:*183, 1989.

Millman RB: Cannabis abuse and dependence. In *Treatments of Psychiatric Disorders*, BT Karasu, editor. American Psychiatric Association, Washington, DC, 1989.

*Millman RB, Sbriglio R: Patterns of use and psychopathology in chronic marijuana users. Psychiatr Clin North Am *9:*533, 1986.

Musto DF: Opium, cocaine and marijuana in American history. Sci Am *265:*20, 1991.

O'Brien CO: Drug addiction and drug abuse. In *The Pharmacological Basis of Therapeutics*, ed 9, AG Gilman, TW Rall, AS Nies, P Taylor, editors. McGraw-Hill, New York, 1996.

Pope HG, Yergelen-Todd D: The residual cognitive effects of heavy marijuana use in college students. JAMA *275:*521, 1996.

Rinaldi L: Marijuana: A research overview. Alaska Med *36:*107, 1994.

Rodriguez de Fonseca F, Carrera MRA, Navarro M, Koob GF, Weiss F: Activation of corticotropin-releasing factor in the limbic system during cannabinoid withdrawal. Science *276:*2050, 1997.

Roffman RA, Stephens RS, Simpson EE, Whitaker DL: Treatment of marijuana dependence: Preliminary results. J Psychoact Drugs *20:*129, 1988.

Seldon BS, Clark RF, Curry SC: Marijuana. Emerg Med Clin North Am *8:*527, 1990.

Seth R, Sinha S: Chemistry and pharmacology of cannabis. Prog Drug Res *36:*71, 1991.

Substance Abuse and Mental Health Services Administration: National survey results on drug use from the monitoring the future study. U.S. Department of Health and Human Services, U.S. Government Printing Office, Washington, DC, 1997.

Tanda G, Pontieri FE, Di Chiara G: Cannabinoid and heroin activation of mesolimbic dopamine transmission by a common mu-1 opioid receptor mechanism. Science *276:*2048, 1997.

Thomas H: Psychiatric symptoms in cannabis users. Br J Psychiatry *163:*141, 1993.

Thornicroft G: Cannabis and psychosis. Br J Psychiatry *157:*25, 1990.

Tunving K: Psychiatric effects of cannabis use. Acta Psychiatr Scand *72:*209, 1985.

Tunving K: Psychiatric aspects of cannabis use in adolescents and young adults. Pediatrician *14:*83, 1987.

Voth EA, Schwartz RH: Medicinal applications of delta-9-tetrahydrocannabinol and marijuana. Ann Intern Med *126:*791, 1997.

Weinrieb RM, O'Brien CP: Persistent cognitive deficits attributed to substance abuse. Neurol Clin *11:*663, 1993.

Wert RC, Raulin ML: The chronic cerebral effects of cannabis use. Int J Addict *21:*605, 1986.

Wickelgren I: Marijuana: Harder than thought? Science *276:*1967, 1997.

11.6 COCAINE-RELATED DISORDERS

JEROME H. JAFFE, M.D.

More than 25 million people in the United States used cocaine at least once in the 1980s and 1990s. For many of those people, use progressed to abuse and dependence. In the early 1990s it was more common to have a lifetime history of cocaine dependence than of bipolar disorder (2.7 percent versus 1.6 percent). While the epidemic appears to have passed its peak, cocaine use is still prevalent, and people with cocaine abuse and dependence disorders continue to come for treatment. A wealth of information now exists on the effects of cocaine on the brain and behavior, and on cocaine toxicity, cocaine dependence, and the efficacy of treatment.

DEFINITIONS

Substance use may be associated with a number of distinct disorders of which dependence and abuse are but two. In the case of cocaine, the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) describes 10 other substance-related disorders. Cocaine dependence is defined in DSM-IV as a cluster of physiological, behavioral, and cognitive symptoms that, taken together, indicate that the person continues to use cocaine despite significant problems related to such use. It is defined in the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) as a cluster of physiological, behavioral, and cognitive phenomena in which the use of cocaine takes on a much higher priority for a given individual than do other behaviors that once had a greater value. Central to these definitions is the emphasis placed on the drug-using behavior, its maladaptive nature, and how over time the voluntary choice to engage in that behavior shifts and becomes constrained as a result of interactions with the drug.

Cocaine abuse is a term used in DSM-IV to categorize a pattern of maladaptive cocaine use leading to clinically significant impairment or distress within a 12-month period, but one in which the symptoms have not met criteria for cocaine dependence. ICD-10 does not use the term.

Other cocaine-related disorders include cocaine intoxication, cocaine withdrawal, cocaine-induced psychotic disorder with delusions or with hallucinations, cocaine intoxication delirium, cocaine-induced mood disorder, cocaine-induced anxiety disorder, cocaine-induced sleep disorder, cocaine-induced sexual dysfunction, and cocaine-related disorder not otherwise specified (Table 11.6-1). The DSM-IV coding scheme provides distinct code numbers for cocaine dependence and cocaine abuse, but the codes for the other substance-related disorders do not differentiate cocaine-induced disorders from the other substance-induced disorders, with the exception of those related to alcohol use.

Cocaine use disorders
Cocaine dependence
Cocaine abuse
Cocaine-induced disorders
Cocaine intoxication
Specify if: with perceptual disturbances
Cocaine withdrawal
Cocaine intoxication delirium
Cocaine-induced psychotic disorder, with delusions
Specify if: with onset during intoxication
Cocaine-induced psychotic disorder, with hallucinations
Specify if: with onset during intoxication
Cocaine-induced mood disorder
Specify if: with onset during intoxication with onset during withdrawal
Cocaine-induced anxiety disorder
Specify if: with onset during intoxication/with onset during withdrawal
Cocaine-induced sexual dysfunction
Specify if: with onset during intoxication
Cocaine-induced sleep disorder
Specify if: with onset during intoxication/with onset during withdrawal
Cocaine-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. ID American Psychiatric Association, Washington, DC, 1994.

Table 11.6-1 DSM-IV Cocaine-Related Disorders

HISTORY

Purified cocaine first became commercially available in 1884. Reports of compulsive cocaine use and cocaine psychosis appeared in the European medical literature within the decade. By the beginning of the twentieth century, cocaine use and dependence were not uncommon in the United States. Cocaine was an ingredient of Coca-Cola until 1900; nonprescription proprietary nostrums containing cocaine were widely promoted until the Harrison Act was passed in 1914; and 100 mg of illicit cocaine could still be bought for a quarter in the 1920s. With growing public awareness in the United States of the physical and legal risks of drug use, cocaine use and dependence declined gradually. It remained fairly common in Europe, however. Hans Maier's classic *Der Kokainismus*, published in 1926, included descriptions of relatively contemporaneous clinical cases.

There appears to have been little cocaine use and dependence from the late 1930s to the early 1970s. Although some heroin users also used cocaine, virtually no cocaine-dependent patients entered the U.S. Public Health Service Hospital at Lexington, Kentucky, in the 1960s. Starting in the 1970s and continuing throughout the 1980s, the availability of cocaine increased noticeably. For the first few years of its renewed popularity among affluent young adults, there were few reports of cocaine toxicity or of persons seeking treatment. Some observers who were apparently unaware of previous epidemics, in which the compulsive nature of cocaine use and its serious toxicity had been repeatedly documented, declared cocaine to be a relatively benign drug. Its use spread from those affluent young adults (who tended to use it intranasally) throughout all economic and age levels of society. More-hazardous routes of administration, including injection and inhalation of freebase forms such as “crack,” became common.

By 1985, owing to its increasing availability and declining price, 20 million people had tried cocaine. Its toxicity became quite apparent as the number of emergency room visits for cardiovascular, neurological, and psychiatric complications rose sharply; and its capacity to induce dependence was apparent from the escalating number of requests for treatment. Increasing numbers of drug users, overdose deaths, crime, and the images of “crack babies” damaged in utero by cocaine-using pregnant women gave national visibility to the drug problem, particularly to cocaine use. Federal expenditures for law enforcement escalated. Penalties for drug selling and possession were increased, and national prevention campaigns were initiated. Drug (urine) testing in the workplace became more common. Toward the end of the 1980s casual use declined, as did the number of cocaine-related medical emergency cases seen in hospital emergency rooms. The number of heavy users did not decline as sharply, however, and urine tests of people who were arrested showed that a substantial number of criminals were still using cocaine. The drug continued to be relatively available, less costly than in the 1970s, and by the early 1990s emergency room visits began to increase slightly. However, by the mid-1990s arrestee drug testing indicated that in some large cities people who got in trouble with the law, particularly the younger ones, were using cocaine less.

COMPARATIVE NOSOLOGY

The DSM-IV criteria for cocaine dependence are the same generic criteria as are applied to other psychoactive drugs. The notion of a generic concept of dependence is shared with the revised third edition of DSM (DSM-III-R) and ICD-10. Despite some changes in wording, the syndromes and criteria for making the diagnosis of

dependence are similar in DSM-III-R and DSM-IV. A generally high level of agreement also exists between DSM-IV and ICD-10: they use similar concepts (the dependence syndrome varying in severity), although the wording of the criteria for determining the presence and severity of the syndrome differs. Both require that three elements of the syndrome be noted within a 12-month period. Although DSM-IV appears to place greater stress than ICD-10 on tolerance and physiological dependence (because it asks clinicians to specify if these criteria are present), in practice this has little impact on the proportion of patients seeking treatment who meet diagnostic criteria for dependence. Most patients who meet current DSM-IV criteria for dependence report some tolerance, withdrawal, or both.

ICD-10 and DSM-IV differ in the classification of what is called substance abuse in DSM-IV. ICD-10 does not use the term “abuse” but includes instead the category of harmful use, which differs substantially from the concept of “abuse” used in DSM-IV. However, the concept of harm is limited to physical and mental health (e.g., hepatitis, cardiac damage, episodes of depression, or toxic psychosis). It specifically excludes social impairments, as follows:

> Harmful patterns of use are often criticized by others and frequently associated with adverse social consequences of various kinds. The fact that a pattern of use or a particular substance is disapproved of by another person or by the culture, or may have led to socially negative consequences such as arrest or marital arguments is not in itself evidence of harmful use.

EPIDEMIOLOGY

Cocaine use has fluctuated dramatically over the past three decades, not just in the United States, but also in South America and in western Europe. In the United States the various activities aimed at estimating the extent and consequences of psychoactive drug use are the annual Monitoring the Future Study (MTF) (formerly known as the High School Senior Survey); the National Household Survey on Drug Abuse (NHSDA); the Drug Abuse Warning Network (DAWN), which obtains reports from a selected group of hospital emergency rooms and medical examiners' offices on drug-related adverse effects and deaths; and the Arrestee Drug Abuse Monitoring (ADAM) program (formerly DUF), which obtains its data from urine tests of arrestees at selected jails. All of those estimating techniques have sampling limitations, and none apply standardized diagnostic criteria to substance-use patterns or adverse effects. Consequently, although they provide a picture of use over time, these methods do not reveal changes in the incidence and prevalence of specific substance-related disorders, such as dependence and abuse

In the annual MTF study all indicators of self-reported cocaine use among high school seniors (lifetime, past-year, and past-month use) declined substantially from the high annual prevalence level of 13.1 percent in 1985 to 3.1 percent in 1992, the lowest since the survey began in 1975. Cocaine use then gradually rose again, reaching an annual prevalence rate of 5.5 percent in 1997. The NHSDA found steady declines in the annual use of cocaine, from peak levels of more than 5 percent, from 1982 to 1985, to 2.1 percent in 1992, and then a slower decline to 1.9 percent by 1997. There was relatively little decline in the number of heavy (weekly) cocaine users between 1985 and 1993, but by 1997 past-30-day use was down from an estimated 3.0 percent in 1985 (5.7 million users) to 0.7 percent (1.5 million users). The use of crack cocaine and the incidence of new crack users also declined.

Two population surveys used accepted diagnostic criteria to measure the extent of substance abuse and dependence: the Epidemiologic Catchment Area (ECA) Study, carried out in the early 1980s, used third edition of DSM (DSM-III) criteria; the National Comorbidity Survey (NCS), carried out from 1990 to 1992, used DSM-III-R criteria. The ECA report combined categories of dependence and abuse for cocaine. The 1-month and 6-month prevalence rates for cocaine abuse-dependence were too low to be measurable; the lifetime rate was 0.2 percent. The NCS carried out just after the peak of the cocaine epidemic reported a lifetime prevalence rate for cocaine dependence of 2.7 percent among 15- to 54-year-olds; 16 percent of respondents reported a history of some use of cocaine.

ETIOLOGY

Substance dependence is currently viewed as the result of a process in which social, psychological, cultural, and biological factors influence substance-using behavior. The actions of the drug are seen as critical, but it is recognized that not everyone who becomes dependent experiences the effects of a given drug in the same way. Further, depending on the individual, different factors may be more or less important at different stages of the process, even with the same class of pharmacological agents.

Social and cultural factors largely influence the availability and initial use of cocaine and other substances. In the case of cocaine, pharmacological factors are believed important in perpetuating use and progression to dependence. Cocaine has potent mood-elevating and euphorigenic actions, especially when its effects have rapid onset, as when cocaine is injected or inhaled. Although some physical dependence develops, an aversive withdrawal syndrome probably is less prominent in perpetuating cocaine use than that of opioids and sedatives.

Comorbidity Additional psychiatric diagnoses are quite common among those dependent on cocaine. It is not always evident how this comorbidity is linked etiologically to cocaine, but the epidemiological evidence clearly shows that the presence of a psychiatric disorder not related to substance abuse (e.g., mood disorders, schizophrenia, and antisocial personality disorder) substantially increases the odds of developing substance abuse and dependence. For some persons cocaine may serve to alleviate various psychiatric disorders or dysfunctional states. Some users, for example, may find relief from dysthymic disorder. Others may find that cocaine facilitates sexual activity, permits extended socializing, or counteracts the sedative effects of alcohol. However, while such factors may explain substance use on more than one occasion, they do not account for progression to dependence or abuse.

Cocaine use may induce psychiatric syndromes (e.g., panic disorders) that may persist even after drug use is stopped. Persons with certain types of psychiatric disorders may be prone to experiment with cocaine or other substances, and factors that predispose to psychiatric disorders may also predispose cocaine users to become cocaine dependent.

Research on the temporal appearance of the syndromes indicates that in some instances and for some syndromes, substance use antedates the psychiatric disorder. In one component of the ECA study subjects were reinterviewed 1 year later. Those who reported cocaine or stimulant use in the interval were almost eight times more likely than nonusers to experience depression and 14 times more likely to have had a panic attack. Cocaine users were almost 12 times more likely to experience a manic episode.

The ECA data also show a relation between the extent of cocaine use and other psychiatric disorders. Among men 18 to 44 years of age, those who had never used cocaine or had used it fewer than five times had a lifetime prevalence of major depression of 7.6 percent; it was 11 percent for users who were never daily users and almost 26 percent for those who met the DSM-III criteria for cocaine abuse. Similarly, the lifetime prevalence of panic disorder was related to the extent of cocaine use.

Genetic Factors Laboratory animal strains differ greatly in their willingness to self-administer psychoactive drugs, including cocaine, and strains can be developed that differ even more markedly. The most convincing evidence to date of a genetic influence on cocaine dependence comes from studies of twins. A study of male twins who served in the U.S. military between 1965 and 1975 found higher concordance rates for stimulant dependence (cocaine, amphetamines, and amphetamine-like drugs) among monozygotic than dizygotic twins. The analyses indicated that genetic factors and unique (unshared) environmental factors contributed about equally to the development of stimulant dependence. Other studies have shown genetic contributions to attention-deficit/hyperactivity syndrome, conduct disorder, and antisocial personality disorder. Since these disorders are important risk factors for drug use and dependence, these findings also support genetic involvement in the etiology of drug dependence in general.

Other Factors Social, cultural, and economic factors are powerful determinants of initial use, continuing use, and relapse. Excessive use is far more likely in countries where cocaine is readily available. Different economic opportunities may influence certain groups more than others to engage in selling illicit drugs, and selling is more likely to be carried out in familiar communities than in those where the seller runs a high risk of arrest.

Since in both human and animal studies alternative positive reinforcers compete with drugs as reinforcers, the absence of such nondrug alternatives can be seen as a causal factor for use, especially where drugs are available and the social pressures against using them are not strong. Alternative positive reinforcers are not limited to material rewards but include psychological rewards associated with satisfying interpersonal relationships and the self-esteem that derives from achievements in socially acceptable roles. In animal models, chronic stress mediated by high levels of cortisol increases sensitivity to the reinforcing effects of cocaine and induces relapse to drug self-administration in withdrawn animals.

Learning and Conditioning Learning and conditioning are also considered important in perpetuating cocaine use. Each inhalation or injection of cocaine yields a rush and a euphoric experience that reinforce the antecedent drug-taking behavior. In addition, the environmental cues associated with substance use become associated with the euphoric state so that long after a period of cessation, such cues (e.g., white powder and paraphernalia) can elicit memories of the euphoric state

and reawaken craving for cocaine.

In cocaine abusers (but not in normal controls), cocaine-related stimuli activate brain regions subserving episodic and working memory and produce electroencephalographic (EEG) arousal (desynchronization). Increased metabolic activity in the limbic-related regions such as amygdala, parahippocampal gyrus, and dorsolateral prefrontal cortex correlated with reports of craving for cocaine, but the degree of EEG arousal did not.

Pharmacological Factors As a result of actions in the central nervous system (CNS), cocaine can produce a sense of alertness, euphoria, and well-being. There may be decreased hunger and less need for sleep. Performance impaired by fatigue is usually improved. Some users believe that cocaine enhances sexual performance.

Mechanisms of Action Cocaine inhibits the normal reuptake of monamines from the synaptic cleft by binding to transporter proteins. Its reinforcing effects are primarily due to its actions at the dopamine transporter, producing high levels of dopamine in the synapse. Evidence suggests that stimulation of both dopamine (D_1) and D_2 receptors plays some role in dopamine's reinforcing and salience-enhancing actions. Cocaine also inhibits reuptake of norepinephrine and serotonin. The increase in norepinephrine concentration is important for some of cocaine's toxic effects.

In animals cocaine is considered the most powerful pharmacological reinforcer of drug-taking behavior known. Given free access, animals choose to self-administer cocaine rather than have food, water, or access to other animals. Death from starvation or drug toxicity is the typical consequence of unlimited cocaine access. With limited access (2 to 6 hours a day), cocaine does not gain such control over behavior, and animals may select food in preference to cocaine, depending on the dose, the amount of work they must do to get the dose, and the type and amount of food offered as alternative reinforcers.

Common Routes of Administration Cocaine can be taken orally, by injection, by absorption via nasal and buccal membranes, or by inhalation and absorption through the pulmonary alveoli. Cocaine hydrochloride, the water-soluble form typically used for snorting or injection, is largely destroyed by the heat of burning and so is not well suited for smoking. Cocaine as freebase sublimates before it is destroyed by heat. The hydrochloride salt can be converted to the freebase form by treatment with alkali and extraction with organic solvents. Inhalation of freebase cocaine produces almost immediate absorption and a rapid onset of effects. In the 1980s users learned to avoid the fire hazard of extracting organic solvents and still produce a crude form of freebase cocaine by heating the cocaine with sodium bicarbonate to yield crack, a hard, white mass that is freebase plus impurities. When smoked, this material gives off a crackling sound. In cocaine-producing countries some users may smoke a crude intermediate product, cocaine sulfate (coca paste, pasta basica, basuca), which is usually contaminated with solvents.

As with the opioids, the rapid onset of cocaine's effects after intravenous injection or freebase inhalation produces an intensely pleasurable sensation, or rush. The cocaine rush lasts only a few minutes, whereas other psychological and physiological effects tend to decline more slowly, in parallel with declining concentrations in plasma.

Metabolism The half-life of a single dose of cocaine in the blood is only about 30 to 90 minutes. It is typically hydrolyzed by butyrylcholinesterase (plasma pseudocholinesterase) and liver esterase into inactive metabolites, most significantly benzoylecgonine and ecgonine methylester. The metabolite is generally detectable in urine for 24 to 72 hours after brief periods of use. With repeated high dosages (e.g., 1 to 2 grams daily), cocaine or its metabolites may accumulate in body compartments (e.g., fat and the CNS), from which it is then slowly released. Consequently, using sensitive measures, cocaine may be detectable in the urine of heavy users for 2 weeks. In a study of cocaine-dependent patients admitted to an inpatient treatment unit, the average time from last reported cocaine use to first negative urine test (using a cutoff of 300 ng/mL) was 105 hours, and 20 percent had positive tests for 120 hours or longer.

The concurrent use of cocaine and alcohol may result in the accumulation of a distinct metabolite, cocaethylene. This metabolite is active and longer lasting than cocaine itself and may account for the enhancement of subjective effects and toxicity when the two are used simultaneously.

Tolerance and Sensitization Patients seeking treatment often report needing progressively more cocaine to get the same effect. In laboratory studies, occasional users of cocaine have more marked cardiovascular, subjective, and endocrine responses to intravenous challenge doses of cocaine than subjects who are cocaine dependent. Despite the evidence for some tolerance to blood pressure–elevating effects, even experienced users may sustain significant cardiovascular toxicity.

Chronic use of cocaine also produces a form of sensitization in which the response to a given dose is actually enhanced. In animals repeated doses of CNS stimulants such as cocaine or amphetamine eventually elicit seizures or stereotyped behaviors not seen with initial doses. Sensitization is produced more reliably by intermittent dosing than by continuous dosing. The sensitization can be long lasting.

The paranoid states and toxic psychoses that commonly develop among chronic cocaine users are believed to be among the phenomena to which sensitization develops. Cocaine psychosis develops more rapidly in those who have been chronic users or had developed psychoses previously.

Withdrawal States The cocaine withdrawal syndrome has aversive qualities (e.g., dysphoria and anhedonia). Although withdrawal anhedonia and fatigue are not generally reported to be the most important reasons for relapse after brief withdrawal, some users who have come to depend on cocaine for high energy or to project a confident persona may be temporarily unable to function without it. For others, withdrawal dysphoria may exaggerate the intensity of an antecedent mood disorder. If a protracted cocaine withdrawal syndrome exists, it is more subtle than the syndrome associated with opioid withdrawal. It is in some ways puzzling that patients do not usually attribute craving for cocaine and relapse to withdrawal, since there is considerable evidence that chronic cocaine use produces significant long-lasting changes in many parts of the brain.

COCAINE-INDUCED BRAIN CHANGES The repeated finding of perfusion deficits in the brains of cocaine-dependent subjects recently withdrawn from cocaine is probably not related to tolerance or withdrawal; however, several other findings probably are. Many, but not all, studies using positron emission tomography (PET) and single positron emission computed tomography (SPECT) to study the brains of cocaine-dependent subjects have found an increased number of dopamine transporters in the striatum, a finding consistent with postmortem studies. Within a few days of withdrawal, cocaine abusers show higher than normal cerebral metabolic rates in orbitofrontal cortex and basal ganglia that correlate with craving. At 1 to 4 weeks and at 3 to 4 months postwithdrawal, cocaine abusers have lower metabolic rates in the frontal cortex that correlate with symptoms of depression, and decreased availability of D_2 receptors that correlates with decreased cerebral metabolic rates and years of cocaine use. An increase in μ-opioid receptor binding after 4 weeks of cocaine abstinence also correlated with severity of cocaine craving.

Chronic cocaine use induces a wide range of changes in the brains of animal models; many of these changes appear to be adaptive responses, while others may be linked to sensitization. After a period of chronic, binge-like cocaine administration, the threshold for reinforcement increases, and dopaminergic and serotonergic transmission in the nucleus accumbens decreases. In addition, the density of D_1 receptors increases. Messenger ribonucleic acid (mRNA) for corticotropin-releasing factor (CRF) increases, μ- and k-opioid receptors are upregulated, and the concentration of mRNA encoding for prodynorphin in the striatum and nucleus accumbens increases. Dynorphin, acting on k-receptors, probably serves as a negative feedback mechanism, dampening excessive dopaminergic activity. When cocaine use ceases, enhanced dynorphin activity could contribute to reduced dopaminergic activity manifested in dysphoria and anhedonia. Neurons bearing dopamine receptors in these same brain areas show upregulation of cyclic adenosine monophosphate (cAMP)-dependent kinase and decreased concentrations of Gi protein. Both of these changes contribute to upregulation of the cAMP pathway and activation of various transcription factors, (e.g., CRE-binding protein [CREB]), which results in the production of long-lasting Fos-like proteins that are distinct from those seen after acute cocaine administration. The persistent changes in the cAMP pathway probably represent one mechanism underlying tolerance.

DIAGNOSIS AND CLINICAL FEATURES

Table 11.6-1 lists the DSM-IV cocaine-related disorders.

Patterns of Use and Abuse There are several patterns of cocaine use and abuse. For example, Indians in the Andes chew coca leaves daily, but apparently very few progress to excessive use or toxicity. Although some cocaine users can use it intermittently without becoming dependent, it is not clear how long such intermittent, nondependent use can continue and for what proportion of users. Cocaine use that does not cause problems for the user does not meet the DSM-IV criteria for either dependence or abuse.

Most cocaine users seeking treatment report initial intermittent use. However, at some stage the use escalated, with episodes of high-dose usage becoming more

frequent. Unlike opioid dependence, daily use is not the most common pattern among persons seeking treatment for cocaine dependence. A small percentage of such patients report using high doses, but for only a few days a month over a long period; such persons may still meet the criteria for dependence. That pattern is atypical, but intermittent use is not. Intermittent use consists of episodes or binges of use, often starting on weekends and paydays and lasting until the drug supply is exhausted or toxicity develops. The runs, or binges, during which the drug may be used every 15 to 30 minutes, can last 7 or more consecutive days but typically are shorter. Although there appears to be little tolerance between binges, changes in the response to the drug occur during the binge. Euphoric effects seem less prominent, and anxiety, fatigue, irritability, and depression increase. Any pause in the drug use causes blood concentrations to drop; typically, there is dysphoria rather than a return to normal mood. If cocaine is still available, it is used to dispel the dysphoria. When the binge is interrupted or supplies have been depleted, a cocaine crash quickly follows. Patients report the sense of needing more cocaine to get the same effect (tolerance) more commonly than the experience of pronounced withdrawal. Some users distinguish between a brief crash and withdrawal. A substantial proportion of cocaine users seeking treatment report daily or almost daily use, often associated with daily heroin use.

In the early stages, cocaine use may cause little interference with normal activities. Some persons may even find that the sense of energy and heightened sense of self-confidence facilitate productive activity. Others may find that the cocaine facilitates social interaction, particularly enhancing sexual arousal and enjoyment, at least initially. The development of sexual dysfunction later in the course of use is better documented than is the enhancement.

In addition to feelings of euphoria, cocaine use may also induce concurrent feelings of anxiety, irritability, and suspiciousness. Users may commit crimes to obtain money to buy cocaine, and such crimes may involve violence. In addition, cocaine can induce paranoid ideation, and there are numerous reports of homicide and attempted homicide during such cocaine-induced toxic states.

Cocaine is an especially powerful reinforcer when it is taken in ways that produce a rapid onset of effects. Not only do intravenous and intrapulmonary routes of administration produce a rapid rise in blood and brain drug concentrations and an intense rush, but especially with the smoking of freebase cocaine, an almost equally rapid decline in blood and brain drug concentration occurs as the cocaine is redistributed and metabolized. Compared with those who use it intranasally, users who inhale freebase cocaine or inject the salt intravenously seem more likely to move from experimentation to regular, compulsive use, limited only by the availability of the drug or the money to buy it. Even in the laboratory setting it can be shown that craving for cocaine is briefly intensified a few minutes after intravenous use when brain and blood concentrations are falling. However, while intravenous and pulmonary cocaine use are far more likely to result in compulsive use and dependence, the intranasal route can also lead to dependence and to the full range of cocaine toxicity (including fatalities).

Cocaine abusers frequently use sedatives or opioids to modulate the stimulant and toxic effects of the cocaine, a practice that can lead to concurrent dependence on sedatives or opioids. Sometimes an opiate, such as heroin, and cocaine are injected intravenously simultaneously; the mixture (speedball) is reportedly especially euphorigenic. Similar synergistic effects are seen when cocaine and buprenorphine (Buprenex) are taken simultaneously. Alcohol is probably the substance most commonly used in conjunction with cocaine, and its use may become associated with cocaine use and can trigger cocaine craving in former users trying to abstain from cocaine.

Cocaine Dependence As the drug use progresses, greater priority is often given to obtaining and using cocaine than to meeting other social obligations or avoiding toxicity or arrest. The user may engage in illegal activities to raise money for cocaine or trade sex for it. At this stage the use of cocaine is considered maladaptive and probably meets the DSM-IV criteria for cocaine abuse or dependence. The DSM-IV criteria for cocaine dependence are the same generic criteria applied to other substances (see Table 11.1-3). A diagnosis of dependence requires a maladaptive drug-use pattern that leads to clinically significant impairment or distress, as indicated by at least three of seven criteria presented in the table. DSM-IV instructs the clinician to specify whether physiological dependence is present (i.e., evidence of either tolerance or withdrawal as defined in the diagnostic criteria).

Drug use to prevent withdrawal is not as dominant with cocaine dependence as with opioid dependence. However, the other criteria for dependence are common among heavy users of cocaine. Tolerance to some drug actions (e.g., euphorigenic effects) can coexist with increased sensitization to other actions (e.g., anxiogenic and psychotogenic effects).

Cocaine Abuse Some cocaine users develop problems or adverse effects related to their drug use (i.e., their use is maladaptive), even though such use does not meet the three-criteria requirement for the diagnosis of dependence. Examples of such recurrent maladaptive patterns include use that leads to multiple legal problems; failure to meet major social, school, or work-related obligations; and continued use despite social or vocational difficulties caused by, or aggravated by, cocaine use. When one or more such substance-related problems occur in a 12-month period but the pattern has never met the criteria for dependence, the diagnosis of cocaine abuse (see Table 11.1-8) should be made.

Cocaine Intoxication Among those who meet the criteria for cocaine abuse or dependence, certain psychiatric toxicities are common. Just as alcohol-dependent persons are frequently intoxicated, cocaine users commonly develop the symptoms of cocaine intoxication during the course of a single binge. The euphoria may be accompanied by increasing suspiciousness, hypervigilance, anxiety, hyperactivity, talkativeness, and grandiosity. Users may engage in stereotyped and repetitive behaviors (e.g., disassembling and reassembling the same object). Typically other signs and symptoms of central stimulation occur, such as tachycardia, cardiac arrhythmias, blood pressure changes, pupillary dilation, perspiration, or chills. Hallucinations may occur, including tactile hallucinations. Judgment is impaired, and confusion may occur, but insight into the drug-induced nature of the hallucinations is retained.

Any of these symptoms following the recent use of cocaine should invoke consideration of cocaine intoxication, provided they are not better accounted for by some other medical or mental disorder and there are at least two of a number of physiological signs commonly seen with cocaine use (e.g., tachycardia and elevated blood pressure). The DSM-IV diagnostic criteria for cocaine intoxication (Table 11.6-2) are identical to the criteria for amphetamine intoxication except for the substitution of the word “cocaine” for the words “amphetamine or a related substance.” Any perceptual disturbances should be specified. In one study of cocaine abusers in the community, just over half reported experiencing paranoia or hallucinations at some time; among those who sought treatment, 63 percent reported those symptoms. Cocaine intoxication may occur in occasional users who do not meet the criteria for abuse or dependence.

A. Recent use of cocaine.
B. Clinically significant maladaptive behavioral or psychological changes (e.g., euphoria or affective blunting; changes in sociability; hypervigilance; interpersonal sensitivity; anxiety, tension, or anger; stereotyped behavior; impaired judgment; or impaired social or occupational functioning) that developed during, or shortly after, use of cocaine.
C. Two (or more) of the following, developing during, or shortly after, cocaine use:
(1) tachycardia or bradycardia
(2) pupillary dilation
(3) elevated or lowered blood pressure
(4) perspiration or chills
(5) nausea or vomiting
(6) evidence of weight loss
(7) psychomotor agitation or retardation
(8) muscular weakness, respiratory depression, chest pain, or cardiac arrhythmias
(9) confusion, seizures, dyskinesias, dystonias, or coma
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.
Specify if:
With perceptual disturbances

Reprinted with permission from American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.6-2 DSM-IV Diagnostic Criteria for Cocaine Intoxication

The development of paranoia does not seem to be closely related to cocaine dose. Some cocaine users develop the syndrome at far lower doses than are used by others who do not develop it. Furthermore, a person who has experienced cocaine-induced paranoia is more likely to have it recur with subsequent cocaine use. It is postulated that a change in threshold represents a form of sensitization. Cocaine use has also been linked to development of a panic disorder that outlasts the cocaine use; here too sensitization has been postulated.

Cocaine Intoxication Delirium and Cocaine-Induced Psychotic Disorder Whereas some paranoia or hypervigilance is typical of cocaine intoxication and tactile and other hallucinations may also occur, cocaine use can also induce a toxic delirium and a more persistent toxic psychotic disorder characterized by suspiciousness, paranoia, visual and tactile hallucinations, and loss of insight. The hallucination of bugs (cocaine bugs) or vermin crawling under the skin (formication) is sometimes reported and is often associated with excoriation of the skin. A paranoid syndrome can develop within 24 hours after the beginning of a cocaine binge. When the syndrome develops in the presence of a clear sensorium and the person retains insight into the drug-induced nature of the symptoms, it is called cocaine intoxication,

even when there are hallucinations. When insight is lost but the sensorium is clear, the syndrome is called cocaine-induced psychotic disorder with delusions or with hallucinations. If consciousness is disturbed (i.e., the ability to focus, sustain, or shift attention is reduced) and deficits in memory and orientation exist, the diagnosis is cocaine intoxication delirium.

Cocaine Withdrawal Cocaine withdrawal phenomena have not been as thoroughly studied as those associated with opioids or alcohol. No experimental studies have been conducted in which patients with known baseline characteristics have been stabilized solely on large doses of cocaine and then abruptly withdrawn. Consequently, most data have been derived from interviews and patients' recollections or from observations of hospitalized patients whose level of drug ingestion and prior baseline characteristics can only be estimated.

Emil Erlenmeyer reported in 1886 that depression was likely to be seen when cocaine was stopped, and in *Der Kokainismus* Hans Maier noted depression and apathy upon cessation of cocaine. During the cocaine epidemic of the 1980s about 50 percent of cocaine users reported experiencing some type of withdrawal when drug use was interrupted.

An early description of withdrawal, based on interviews of outpatients, described a three-phase syndrome in which the first phase, the crash, was characterized by agitation, depression, anorexia, and high cocaine craving. This cluster of symptoms was followed by a decrease in cocaine craving, fatigue, depression, and a desire for sleep; followed in turn by exhaustion and hypersomnia, with intermittent awakening, and hyperphagia. The second phase was reported to be heralded by normalized sleep, improved mood, and low levels of craving, but that relatively benign phase was succeeded by a return of anergia, anhedonia, anxiety, and increased cocaine craving, especially in response to stimuli previously associated with cocaine use. A third phase—extinction (which appears to represent a period of extended vulnerability to relapse rather than a phase of an extended withdrawal syndrome)—was also described.

A complex phasic withdrawal has not been reported by others who observed cocaine-dependent patients admitted to clinical and research units. Instead, symptoms of depression and craving for cocaine declined steadily over several weeks. After 3 weeks, sleep, weight, and appetite were mostly comparable to those of normal controls on the same unit. Hypersomnia, disturbed sleep, hyperphagia, and excessive weight gain were not seen, nor was a severe crash observed. The phases and fluctuations in craving previously reported might have been related to environmental stimuli.

Some of the inconsistencies in the findings and symptoms associated with cocaine cessation are probably attributable to differences in the dose and duration of use and to vulnerability factors. In interviews with almost 400 cocaine abusers, including about 100 who were not seeking treatment, some 83 percent reported tolerance to cocaine effects (needing more to get same effect), and 52 percent reported having undergone some type of withdrawal. Those seeking treatment were more likely to report experiencing withdrawal. Available data show no convincing evidence that a protracted cocaine withdrawal syndrome follows resolution of the signs and symptoms associated with abrupt cessation. However, abnormalities of brain function appear to persist for at least 12 weeks, and possibly, subtle withdrawal phenomena increase vulnerability to relapse.

Drug craving, often part of cocaine withdrawal, is not included among DSM-IV diagnostic criteria. While not commonly observed during recent clinical studies, severe depression, sometimes associated with suicidal ideation, is reported in the older literature on cocaine withdrawal and in occasional contemporary clinical reports. To what degree the more severe depressive features are a part of withdrawal or represent the emergence of primary mood disorder is unclear.

The DSM-IV diagnostic criteria for cocaine withdrawal Table 11.6-3 specify that the syndrome follows the cessation (or reduction) of heavy, prolonged cocaine use. Further, the dysphoric mood and other symptoms (e.g., fatigue and sleep disturbances) must be intense enough to cause significant distress or impairment. Thus, the criteria are structured so that the brief dysphoria and fatigue (crash) that follow a single short binge by an occasional user do not lead to a diagnosis of withdrawal.

A. Cessation of (or reduction in) cocaine use that has been heavy and prolonged.
B. Dysphoric mood and two (or more) of the following physiological changes, developing within a few hours to several days after criterion A:
   (1) fatigue
   (2) vivid, unpleasant dreams
   (3) insomnia or hypersomnia
   (4) increased appetite
   (5) psychomotor retardation or agitation
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.6-3 DSM-IV Diagnostic Criteria for Cocaine Withdrawal

Animal Models of Withdrawal Although there is no easily observable animal model of cocaine withdrawal comparable to that for the syndromes seen with alcohol or opioid withdrawal, animal analogues of the postuse dysphoria and anhedonia often seen in humans have been proposed. In rats cocaine typically lowers the threshold for intracranial electrical self-stimulation. After 24 hours of cocaine self-administration, the thresholds for such self-stimulation are elevated above baseline for several days, which suggests a relative dopaminergic deficiency or insensitivity. Rats administered cocaine in a binge pattern had elevated dopamine concentrations in the nucleus accumbens during cocaine administration and below-normal concentrations during withdrawal; after 14 days of drug administration, recovery to pretreatment levels was prolonged.

Other Cocaine-Induced Disorders Other psychiatric syndromes that may develop in the course of cocaine use include cocaine-induced mood disorder, cocaine-induced anxiety disorder, and cocaine-induced sleep disorder. With each of those disorders the clinician should specify whether the onset occurred during intoxication or during withdrawal. DSM-IV also describes cocaine-induced sexual dysfunction and a category of cocaine-related disorder not otherwise specified (Table 11.6-4).

The cocaine-related disorder not otherwise specified category is for disorders associated with the use of cocaine that are not classifiable as cocaine dependence, cocaine abuse, cocaine intoxication, cocaine withdrawal, cocaine intoxication delirium, cocaine-induced psychotic disorder, cocaine-induced mood disorder, cocaine-induced anxiety disorder, cocaine-induced sexual dysfunction, or cocaine-induced sleep disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.6-4 DSM-IV Diagnostic Criteria for Cocaine-Related Disorder Not Otherwise Specified

Cocaine-induced mood disorder can occur during use, intoxication, or withdrawal. During use and intoxication, the disorder is more likely to simulate a manic, hypomanic, or mixed episode; during withdrawal, it is more likely to involve a depressed mood. Such diagnoses are difficult to make during periods of active drug use or during the first week or two of withdrawal. Because sexual dysfunction, anxiety, and disturbed sleep are seen so commonly during cocaine use and withdrawal, the diagnoses should be made only when the disturbances or dysfunctions are judged to be in excess of that usually associated with intoxication and withdrawal and only when severe enough to require independent treatment or attention. Panic episodes that develop during cocaine use may persist for many months following cessation.

Lasting vulnerability to panic attacks may be linked to sensitization phenomena.

Comorbidity The frequent co-occurrence of other psychiatric disorders and cocaine dependence was noted during the cocaine epidemic in the early part of the twentieth century. The presence of other psychiatric disorders sharply increases the odds of substance dependence, and substance-dependent persons are more likely than the general population to meet the diagnostic criteria for additional psychiatric disorders.

Among cocaine users seeking treatment, the rates of additional current and lifetime diagnoses are regularly found to be elevated. In one study, about 300 patients (69 percent men, average age 28, mostly lower socioeconomic class) were interviewed using the Schedule for Affective Disorders and Schizophrenia (SADS). Symptoms occurring within 10 days after the last drug use were not used in making any diagnoses. The additional psychiatric diagnoses are shown in Table 11.6-5. The most common additional lifetime diagnoses were alcoholism (62 percent), antisocial personality (33 percent), and major depression (30 percent). In this sample, depression preceded the onset of drug abuse in about one-third of the patients, whereas alcoholism preceded the onset of drug abuse in 21 percent.

Psychiatric Diagnosis	Current Disorder	Lifetime Disorder
Major depression	4.7	30.5
Cyclothymia/hyperthymia	19.9	19.9
Mania	0.0	3.7
Hypomania	2.0	7.4
Panic disorder	0.3	1.7
Generalized anxiety disorder	3.7	7.0
Phobia	11.7	13.4
Schizophrenia	0.0	0.3
Schizoaffective disorder	0.3	1.0
Alcoholism	28.9	61.7
Antisocial personality disorder–DSM-III	32.9	32.9
Attention-deficit disorder	—	34.9

Adapted from Rounsaville BJ, Anton SI, Carroll K, Budde D, Prusoff BA, Gawin F: Psychiatric diagnoses of treatment-seeking cocaine abusers. Arch Gen Psychiatry 48:43, 1991.

Table 11.6-5 Additional Psychiatric Diagnoses Among Cocaine Users Seeking Treatment (New Haven Cocaine Diagnostic Study Results, Percentages)

Some studies have found that cocaine users who seek treatment have higher rates of depression and adverse consequences of drug use than those who do not. Another study found that those not seeking treatment had comparably severe cocaine use and lifetime and current psychiatric disorders, and higher rates of polysubstance use and involvement with the law, but they also tended to minimize the adverse consequence of substance use and lacked pressure to seek treatment.

The prevalence of schizophrenia has generally been reported to be low among patients admitted to cocaine treatment programs, probably largely because people with schizophrenia are excluded from such programs. In fact, persons with schizophrenia commonly use cocaine or amphetamine and develop both dependence and toxic syndromes, although the diagnosis is not routinely made. Depending on the geographical area, an estimated 12 to 30 percent of persons with schizophrenia also abuse cocaine. It has been suggested that they use cocaine and stimulants to alleviate negative symptoms, postpsychotic depressive disorder of schizophrenia, and the side effects of antipsychotics. One nonblinded study did find fewer negative signs and more anxiety and depression among cocaine users with acute schizophrenia who had used cocaine prior to admission. A substantial proportion of schizophrenic patients admit having used cocaine during the months before hospitalization, but many are less candid about recent drug use, and urine tests frequently reveal recent cocaine use unsuspected by clinicians. Patients with schizophrenia who use cocaine tend to be younger and are more likely to be homeless and unemployed than psychotic patients who are not abusing drugs. Special programs involving peer-based support groups seem to be effective in linking substance-using schizophrenic patients to outpatient treatment programs.

Toxicity and Complications High doses of cocaine can cause a wide variety of toxic effects, including cardiac arrhythmias, coronary artery spasms, myocardial infarction, and myocarditis. Other reported cardiovascular toxicities include headache, ischemic cerebral or spinal infarction, and subarachnoid or cranial hemorrhage. Toxic effects on the CNS may include seizures, hyperpyrexia, respiratory depression, and death. Cocaine-related seizures and loss of consciousness are often reported on questionnaires given to heavy users (up to 27 percent); most episodes do not lead to emergency room visits. Rhabdomyolysis, not uncommon after large doses of cocaine, may contribute to renal complications, although vasoconstriction alone may suffice to account for renal damage. Sniffing cocaine can cause ulcers of the mucosa in the nose and perforation of the nasal septum from persistent vasoconstriction. Inhaled cocaine freebase is thought to induce lung damage. Gastrointestinal necrosis, caused by vasoconstriction, has been associated with the rupture of swallowed condoms containing large amounts of cocaine. By producing placental vasoconstriction, cocaine may contribute to fetal anoxia. A list of medical complications associated with cocaine intoxication and abuse is shown in Table 11.6-6.

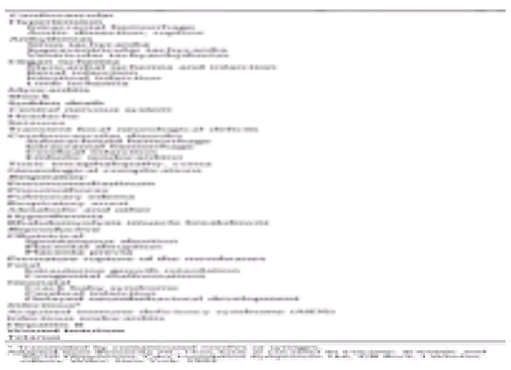

Table 11.6-6 Medical Complications of Cocaine Intoxication and Abuse

Seizures and respiratory depression may be related to cocaine's actions as a local anesthetic, and although the cardiovascular complications are primarily due to its effects on the reuptake of catecholamines in the peripheral nervous system, local anesthetic effects may contribute to myocardial depression. Animal studies reveal significant genetic vulnerability to various kinds of cocaine toxicities, suggesting that some of the observed toxicity in humans may not be predominantly dose dependent and predictable. Furthermore, cocaine elimination is nonlinear at high doses. Probably as a consequence of altered sensitivity of dopaminergic systems in the brain, chronic cocaine users may exhibit abnormal movements such as tics, choreoathetoid movements, and dystonic reactions. They may be particularly sensitive to neuroleptic-induced dystonias.

Cocaine use is frequently associated with increased sexual activity, and sometimes the exchange of sex for cocaine. Such behaviors put cocaine users at elevated risk for venereal diseases, including infection with the human immunodeficiency virus (HIV).

Treatment of Toxicity The treatment of acute cardiac emergencies is aimed at blocking the sympathomimetic effects of the drug and correcting arrhythmias. Some clinicians have recommended using combined a- and b-adrenergic receptor antagonists; however, others have advised against using adrenergic or dopaminergic blockers. Also suggested for myocardial ischemia are calcium channel blockers and nitroglycerine. Grand mal seizures may respond to diazepam (Valium). μ-Agonist opioids reduce cocaine lethality in animals. The fatal hyperpyrexia seen with cocaine shares some features with neuroleptic malignant syndrome and might respond to similar therapy (e.g., dantrolene [Dantrium]). Some researchers advise using ambient cooling.

Pathology and Laboratory Examinations Cocaine metabolites can be detected for varying lengths of time in urine, depending on the dose of cocaine and sensitivity of the assay. They can also be detected in blood, saliva, sweat, and hair. Blood and saliva provide a better index of current concentrations, whereas urine provides a longer window of opportunity for detecting use over the previous few days. Hair analysis can reveal drug use over weeks to months but has little applicability in clinical situations.

Some of the insults to the CNS (e.g., cerebral infarction) are detectable by computed tomography (CT) scans or magnetic resonance imaging (MRI), but most chronic cocaine abusers who do not also abuse alcohol show no evidence of CNS structural damage when examined by these methods. However, studies using PET or SPECT have revealed a variety of functional abnormalities in the brains of recently abstinent cocaine users. Compared with controls, heavy cocaine users showed perfusion defects in the cortex. Considerable improvement in cortical perfusion occurs after several weeks of abstinence, although blood flow in many instances still does not match that of normal controls. Carefully controlled studies have found that brain volumes of cocaine users are slightly smaller (show atrophy) than those of controls who do not use drugs, but do not differ from those of other drug abusers.

Compared with normal controls and patients withdrawn from alcohol, patients withdrawn from cocaine exhibited persistent resting tremor (4 to 6 Hz, similar to that of Parkinson's disease) lasting at least 12 weeks. The tremor is subtle and would not ordinarily be detected by clinical testing; no other cerebellar signs are present. They also exhibit slower reaction times in divided attention tasks that also persisted. Compared with age-matched and education-matched controls, chronic cocaine users are more likely to score in the impaired range on a neuropsychological screening battery. Impairment seems most obvious in concentration and memory, with less impairment in users who had been abstinent longer. To what extent the abnormalities in brain function are causally related to the signs and symptoms associated with cocaine cessation is uncertain.

A few early studies of cocaine addicts reported that almost all patients exhibited hyperprolactinemia lasting several weeks, which seemed consistent with a dopaminergic deficiency. However, several subsequent studies found either no evidence of hyperprolactinemia or a much lower incidence of that effect, and no apparent correlation between high prolactin levels and either cocaine craving or the extent of cocaine use.

DIFFERENTIAL DIAGNOSIS

The disorders associated with the use of cocaine need to be distinguished from both primary mental disorders and disorders induced by other classes of substances. A history of substance ingestion is important in making those distinctions. However, given the unreliability of self-reports about substance use and the likelihood that many users will deny any substance use at all, laboratory testing for drugs in body fluids and histories from collaterals are important. Disorders associated with cocaine use cannot be distinguished from those associated with amphetamines and related substances except by reliable history or laboratory tests. Users of cocaine (and amphetamine and related substances) may exhibit inappropriate optimism, euphoria, expansiveness, excessive talkativeness, and a decreased need for sleep sometimes associated with irritability in the context of a clear sensorium, a pattern also observed in manic and hypomanic episodes of bipolar disorder. However, these symptoms may not be obvious enough to suggest their relation to substance use, and the first indication of substance dependence may be financial difficulties, an arrest for drug sale or possession, or substance-induced toxicity.

Intoxication Cocaine intoxication is diagnosed when the effects of cocaine exceed the mood-elevating effects its users typically seek. The diagnosis of intoxication is appropriate when the effects are problematic enough to require differentiation from hypomanic or manic behavior. Cocaine intoxication can also be confused with amphetamine intoxication and phencyclidine (PCP) intoxication, although the last is usually associated with nystagmus, motor incoordination, and some cognitive impairment. Endocrine disorders (such as Cushing's disease) and excessive use of steroids should also be considered.

Toxic Psychosis Cocaine-induced toxic psychosis can be exceedingly difficult to differentiate from schizophrenia or other psychotic disorders characterized by hallucinations or delusions. The presence of vivid visual or tactile hallucinations should raise suspicion of substance-induced disorder. In areas and populations where cocaine use is common it may be necessary to provide only a provisional diagnosis until the patient can be observed and substance test results are obtained. Even then there may be difficulties because in some urban areas a high percentage of persons with established diagnoses of schizophrenia also use cocaine.

Cocaine-Induced Anxiety Disorder Cocaine-induced anxiety disorder must also be distinguished from generalized anxiety disorder and panic disorder. Panic disorder that has its onset associated with the use of cocaine may persist well beyond the period of cocaine use.

Other Symptoms The symptoms that may emerge during withdrawal—depression, dysphoria, anhedonia, disturbed sleep—need to be distinguished from those of primary mood disorders and primary sleep disorders. Unless the symptoms are more intense or more prolonged than is typical of cocaine withdrawal and so require independent treatment, the diagnosis should be limited to withdrawal, rather than cocaine-induced mood disorder. When a diagnosis of cocaine-induced mood disorder is made, it is important to specify whether the onset was during intoxication or withdrawal. One can also specify the subtype of mood disorder (i.e., with depressive, manic, or mixed features). In differentiating cocaine-induced mood disorder from the primary mood disorders the critical factor is the clinician's judgment that the mood disorder was caused by the cocaine. Generally, a cocaine-induced mood disorder, with onset during intoxication or withdrawal, remits in a week or two. It is appropriate, therefore, to withhold judgment about the diagnosis during the early phase of withdrawal. If depressed mood and related symptoms persist beyond a few weeks, alternative causes should be entertained. In reviewing diagnostic possibilities the clinician should consider the age at which symptoms began and a history of mood episodes that developed before the onset of cocaine use or during any long intervals without significant drug abuse.

COURSE AND PROGNOSIS

Not all cocaine users develop cocaine-related disorders. However, even occasional users can experience cocaine toxicity. Among those who do develop dependence, the time from first use to problematic use ranges from a few months to 6 or more years. The course of cocaine use is often marked by shifts from intranasal to intravenous use and inhalation of freebase forms. In the United States, since most persons who tried cocaine did not become dependent, the decreased cocaine use in the general population in the early 1990s following peak rates of self-reported use in the 1980s does not shed much light on the natural history of cocaine dependence.

Presently little information exists on untreated cocaine dependence, but there are findings on the course of cocaine use among those seeking treatment. A number of short-term (6-month to 2-year) follow-up studies seem to indicate that the course of dependence is more favorable for persons using cocaine who seek treatment than for heroin addicts who seek treatment.

Among veterans on the East Coast randomly assigned to either an inpatient program or a day-hospital program lasting 28 days, 60 percent reported abstinence at 4 months, which was largely maintained at 7 months. About 56 percent of urine specimens were negative for each group at 7 months. In a 1-year follow-up of almost 300 cocaine users, half treated as outpatients and half treated initially as inpatients, both groups showed reductions in self-reported cocaine use during the 30 days before the interview: from an average 17 days per month at admission to 1.1 days for inpatients; from 10 days per month to 5 days for outpatients. Although the data do not give the percentage of those who were entirely abstinent, the improvement levels were substantial and differed from those typically found among heroin-dependent patients seeking treatment. The prognosis appears to be even better for persons with social support.

DRUG ABUSE TREATMENT OUTCOME STUDY (DATOS) In DATOS, the largest recent study of drug users seeking treatment, about 3000 patients from 81 programs were interviewed at entry and 1 year after completion of index treatment. Cocaine dependence was the most frequent primary drug problem, but cocaine was also used by some whose major drug problem involved heroin or alcohol. Among the entire sample of clients 39 percent met criteria for antisocial personality disorder, and 14 percent met criteria for some other DSM-III-R Axis I disorder. At follow-up, weekly or more frequent cocaine use was reported by 35 percent of clients who stayed in long-term residential treatment for less than 3 months and 14 percent of those who stayed longer. For outpatient drug-free programs, the rates were 25 percent for those who stayed less than 3 months and 13.6 percent for those who stayed longer. Weekly or more frequent cocaine use decreased to about 20 percent for short-term inpatient treatment, but the drop was similar for those who stayed more than 2 weeks and those who stayed for a shorter time.

Since baseline levels of cocaine use differed, the results across program types are easier to compare when expressed as percentage reduction from pretreatment levels. Reduction in weekly or more frequent cocaine use for the long-term residential clients was 54 percent for those who stayed less than 3 months and 82 percent for who stayed longer than 3 months; for outpatient drug-free clients, it was 57 percent (less than 3 months) and 87 percent (longer than 3 months); for short-term inpatient clients, 79 percent (less than 2 weeks), and 74 percent (longer than 2 weeks). A substantial proportion of patients were referred for additional social support and treatment after discharge, and many participated in self-help programs. The investigators concluded that there was little difference among program types, but that very different types of patients self-select different treatments.

COCAINE-RETRO This retrospective study identified about 1000 charts of patients treated primarily for cocaine dependence at three types of programs—short-term inpatient, long-term residential, and drug-free outpatient. Some 772 of those patients were interviewed after discharge. Patients in all three types of programs reported substantial reductions not only in cocaine use but also in the use of other drugs. While the self-reports were generally confirmed by drug tests, cocaine use was underreported. About 37 percent of the drug tests were positive for cocaine metabolites (32 percent of the long-term residential patients, 33 percent of the drug-free

outpatients, and 46.4 percent of the short-term inpatients.) The researchers concluded that no significant differences in outcome existed across these three approaches. By self-report, weekly or daily cocaine or crack use was about one third of pretreatment levels for each of the modalities. However, different programs serve different groups. Patients entering these three modalities of treatment differed significantly in terms of gender, age, criminal justice pressure to enter treatment, and the extent and types of support and treatment received after discharge.

VARIETIES OF REMISSIONS Treatment of cocaine dependence may have various outcomes, including, at the extremes, complete relapse to cocaine dependence or total abstinence from cocaine and related drugs for a prolonged period—more than 12 months (sustained full remission). However, sustained partial remissions occur in which, after at least 1 month when no criteria of dependence have been present, one or more criteria of abuse or dependence are again met, but over the course of 12 months fewer cocaine-dependence criteria have been met than the three required for full relapse. There are also situations in which these patterns are observed, but the period of observation is not a full year (early full remission and early partial remission). Any pattern of remission may be observed while the person is in a controlled environment, and that fact should be specified. DSM-IV criteria for both abuse and dependence require maladaptive use associated with distress or impairment. Technically, a person can be in sustained full remission from cocaine dependence despite occasional use, provided the drug use causes no problems or distress and does not escalate. How often such a return to occasional nonproblematic use takes place is unknown.

Cocaine and Crime The typical interactive relation between the use of opioid drugs and crime generally holds true for cocaine users, but some significant differences exist. As with opioid users, considerable heterogeneity exists among cocaine users. Although a history of delinquency or antisocial behavior is often an antecedent to cocaine use, not everyone who uses cocaine or develops cocaine dependence engages in crime, even though the cost of using the drug may create serious financial problems for them. Sometimes, however, a person with no previous criminal behavior will engage in a variety of illegal activities ranging from fraud and white-collar crime to drug selling, prostitution, and predatory crime just to get enough money to buy cocaine. Among persons seeking treatment for any variety of substance abuse, use of cocaine is most highly correlated with income-generating crime.

In a nationwide sample of adolescents 40 percent of serious crimes committed by the entire sample were committed by the 1.3 percent who reported using cocaine. In the late 1980s, when cocaine use declined in the general population in the United States, it rose or merely stabilized among those arrested for a variety of serious offenses. In the late 1990s, however, cocaine use began to decline, especially among younger arrestees.

COCAINE, AGGRESSION, AND VIOLENCE One conceptual framework for thinking about the links between violence and substance (cocaine) use involves three major causal categories: psychopharmacological effects (effects of the substances), economic compulsion (violent crimes committed to obtain money for drugs), and systemic violence (associated with the business methods and lifestyle of drug dealers). Cocaine can induce states of paranoia and aggressive behavior, a common reason why cocaine users are brought to emergency rooms. However, pharmacologically induced aggression is not the major reason why cocaine and crime, and more specifically, cocaine and violence, are linked. Among those arrested for violent crime, the primary predictors of such crime are past arrests for violent crime, poor education, and poor intellectual ability. Past arrest for violence is also associated with antisocial personality disorder. Studies of violent predatory offenders indicate that most had histories of heavy involvement with multiple substance use and with serious crime as juveniles. Among predatory offenders high-frequency substance users were likely to use many substances, particularly heroin and cocaine, and to engage in a variety of crimes, including violent crimes, at high rates. Cocaine also has a nonpharmacologically based link to violent crime; many drug dealers who may not use cocaine routinely themselves resort to violence to protect or expand their customer base.

TREATMENT

Selection of Treatment Setting The general principles of treatment for cocaine dependence do not differ much from those for other varieties of drug dependence. Patient heterogeneity requires careful assessment of the patient and thoughtful selection among alternative treatment approaches. Cocaine dependence severe enough to require formal treatment is often associated with other psychiatric diagnoses. Not all cocaine users require extensive treatment; some who are not severely dependent respond to external pressures, as when employers insist on carefully monitoring substance use. Among the factors influencing selection are the severity of dependence, other drugs being used concurrently, comorbid medical and psychiatric disorders, as well as the preferences of the patient and the alternatives available. Availability, in turn, is often influenced by the policies of managed care companies, the patient's resources, and the types of therapy provided locally.

In general, treatment can be initiated in intensive outpatient settings, although often third party payers do not authorize, and public sector programs cannot provide, the duration of treatment or the intensity shown to be most effective. Some early studies found that inpatient treatment is associated with better outcomes at 1 year than outpatient treatment, even though patients initially treated as inpatients had more-severe cocaine problems. A study using random assignment found that at 4 months, working class veterans treated in a day-hospital program were about as successful in reducing their cocaine use and improving social functioning as those treated in a 28-day inpatient program. However, a somewhat higher proportion of those assigned to the inpatient setting completed the 28-day program. Currently, severe depression with suicidal ideation, psychosis, or substance use that has repeatedly failed to respond to outpatient efforts are the indications for hospitalization. A retrospective study of individuals treated for cocaine dependence in various settings found no advantage in outcome for inpatient treatment lasting longer than 2 weeks.

In many instances, the selection of the setting and type of treatment is made by neither the patient nor the clinician. Patients are often referred (mandated) to treatment by the criminal justice system, which often prefers long-term residential programs (therapeutic communities). The intensity and specificity of services for particular problems (i.e., medical, psychiatric, and vocational) are now considered important determinants of outcome in the specific problem areas.

Detoxification The cocaine withdrawal syndrome is distinct from that of opioids, alcohol, or sedative-hypnotics, since there are no physiological disturbances that necessitate inpatient or residential drug withdrawal. Thus it is generally possible to engage in a therapeutic trial of outpatient withdrawal before deciding whether a more intensive or controlled setting is required for patients unable to stop without help in limiting their access to cocaine. Patients withdrawing from cocaine typically experience fatigue, dysphoria, disturbed sleep, and some craving; some may experience depression. No pharmacological agents reliably reduce the intensity of withdrawal, but recovery over a week or two is generally uneventful. It may take longer, however, for sleep, mood, and cognitive function to recover fully.

Treatment Methods A number of psychological and pharmacological approaches to the treatment of cocaine dependence have been explored. More than 20 different pharmacological agents have been tested in the search for drugs to facilitate withdrawal, reduce postwithdrawal craving, or prevent relapse. In general, no drug with robust therapeutic efficacy has emerged. Psychosocial approaches have included various forms of individual and group psychotherapies, drug counseling, and self-help groups.

It is generally held that total and permanent abstinence from cocaine must be the goal of treatment for those who have developed symptoms of dependence; any use at all is seen as a prodrome to relapse. However, this perspective may underestimate the benefits that accrue from treatment that results in a substantial and prolonged reduction in drug use but falls short of total abstinence. In most studies of treatment effectiveness a significant proportion of patients report substantial reductions in use, even though they are not completely abstinent.

Psychotherapy and Behavior Modification Psychological treatment approaches have used cognitive-behavioral, psychodynamic, and general supportive techniques. One cognitive-behavioral method uses contingency contracting in which it is agreed in advance that for a specified period of time (e.g., 3 months), if the patient uses cocaine (as detected by supervised urine testing) the therapist will initiate actions that will result in serious adverse consequences for the patient, such as informing an employer or a professional credentials board. In one such study 48 percent of potential patients accepted such a contractual arrangement, and 80 percent of those patients successfully abstained from cocaine during the period covered by the contract; many of the successful patients relapsed when the contract expired. Although the technique is not widely used by individual therapists, the general principle of linking drug use and aversive contingencies is central to many employment-based programs and criminal justice programs that use drug (urine) testing. No adequate studies exist comparing such adverse contingency contracts with alternatives.

In contrast, several studies have compared relapse prevention and contingency contracts using positive rewards with 12-step type programs. In these studies, ambulatory cocaine abusers (mostly white, male, white-collar workers) were randomly assigned to either experienced therapists who used 12-step principles or therapists who used behavioral approaches that emphasized contingency management, community reinforcement, and positive rewards (such as vouchers) for cocaine-free urine samples. Retention rates were higher and cocaine use was significantly lower with contingency management and positive rewards. The same researchers found that positive reinforcement was more effective in terms of retention and abstinence than were otherwise identical behavioral treatment methods without positive reinforcement.

Another research group randomly assigned cocaine users (typically white men with more than 12 years of education) to relapse prevention or to a 12-step treatment

that used group techniques. Relapse prevention involved teaching the patient how to recognize high-risk situations and ways of dealing with negative emotions, but it did not offer positive material reinforcement such as vouchers or lottery tickets for cocaine-free urine samples. The 12-step treatment group was led by a man and woman cotherapy team using 3 of the first 12 steps, but it was not Alcoholics Anonymous (AA) or Narcotics Anonymous (NA). The groups did not differ at 6-month follow-up or at end of treatment in retention in treatment or reduction in cocaine use.

In another study, relapse prevention using cognitive-behavioral coping skills training was compared with clinical management in the context of a pharmacotherapy trial comparing desipramine (Norpramin) with placebo. Clinical management was intended to foster a supportive doctor-patient relationship, retention in the protocol, and compliance with medication. Unlike the other two studies, there were more women (27 percent), more minorities (54 percent), fewer high school graduates (24 percent), and fewer patients who were gainfully employed (53 percent). About 40 percent of patients completed the proposed 12-week protocol (mean of 7.2 weekly sessions). Overall, about 70 percent of patients improved, but cognitive-behavioral treatment did not appear better than clinical management. Although there was a trend for more patients receiving relapse prevention or desipramine to complete treatment, the differences were not significant. Relapse prevention appeared to be more helpful with those with more severe cocaine dependence. Desipramine seemed useful primarily for those with less-severe dependence and only early in the course of treatment. There were no differences between the response to drug and placebo at 12 weeks.

Supportive Therapy The specific methods of supportive therapy overlap the techniques used by behaviorists. Patients are helped to separate themselves from drug-using friends and from situations where cocaine is available and which increase drug craving. They are urged to abstain from other substances, such as alcohol and cannabis, because those substances have been reported to increase cocaine craving and the probability of relapse. Patients are also helped to repair the areas of their lives that once provided satisfaction and may have been damaged by the behaviors associated with cocaine use. In addition, patients may be encouraged to participate in Cocaine Anonymous (CA), AA, or NA as a means of gaining control over other substance use.

Psychodynamic, Interpersonal, and Combined Approaches Psychodynamically oriented clinicians emphasize the patient's unconscious motives for using cocaine (e.g., to relieve an inner sense of emptiness or depression). However, experienced clinicians with a wide range of skills believe that a combination of psychological approaches, with the emphasis tailored to the needs of the individual patient, is more effective than treatments that emphasize the principles of only one approach.

Group Psychotherapy Techniques Several distinct approaches to group psychotherapy with cocaine users have been described. Interpersonal group therapy focuses on relationships and uses the group interactions to illustrate the interpersonal causes of individual distress and to offer alternative behaviors. Modified dynamic group therapy is described as emphasizing character, as it manifests itself individually and intrapsychically, and in the context of interpersonal relationships with a focus on affect, self-esteem, and self-care. Both approaches share the view that the group should serve as an interpersonal anchor that leads first to more-stable emotional status and enables members to face unresolved life issues. Both approaches recognize the vulnerability of the patients to narcissistic injury and the need for a supportive, empathetic environment. Some psychotherapists emphasize that the focus in the early months of treatment must be exclusively on the disease and on achieving sobriety and recovery, but modified dynamic group therapy asserts that even early in the process, those goals are not incompatible with attention to characterological problems. Dynamic group psychotherapy assumes that substances are used as self-medication and that the persons most likely to use cocaine include those whose depression, anergia, or boredom is alleviated by it. However, those who place exaggerated value on assertiveness and self-sufficiency may also find cocaine alluring. Since patients must sometimes be abstinent for at least 2 weeks before participating in this type of group therapy, the technique may be more accurately described as relapse prevention rather than treatment to induce initial cessation. Few studies bear on the effectiveness of such group therapy.

Group Counseling Perhaps the most widely used form of psychosocial treatment for cocaine dependence is group counseling, in which the group is open ended with rolling admissions, the group leaders are drug counselors, many of whom are recovering from addiction, and the emphasis is on providing a supportive atmosphere discussing problems in recovery and encouraging participation in 12-step programs. It is unclear if this treatment is properly considered group therapy since the group is continually changing.

Intensive and Eclectic Treatment Most private practitioners who typically work with employed patients of middle and upper socioeconomic class probably use approaches best described as intensive and eclectic and consider the patient's motivation for treatment (stage of change). The goals of treatment and the techniques shift over time, with the initial work focused on forging an alliance with the patient and increasing the motivation to achieve abstinence. Treatment may initially involve several group and individual sessions per week focused on interrupting the substance-use cycle and on developing detailed plans for avoiding exposure to the substances and high-risk situations. Experienced clinicians advise taking supervised urine samples at least weekly, so that patients know that they are accountable for their actions. Results are generally considered measures of progress, not tests designed to catch the patient in lies. Emphasizing shared goals and instilling hope, trust, and confidence may foster the engagement of even resistant patients. While arguing about whether or not the patient is an addict may drive a patient from treatment, since some may not be ready to give up cocaine entirely, clinicians need to be able to work with such patients despite their own doubts that they may be enabling substance-using behavior.

Once abstinence is achieved, the goals of treatment shift to relapse prevention. Slips (occasional use of cocaine), especially during the first 60 days of treatment, should be used as learning experiences, with the focus on how to prevent reoccurrence. In the later stages of treatment a balance must be struck between enabling (tolerating continued use) and being so rigid about use that the patient leaves treatment. Sometimes a temporary suspension from the group, with continued individual sessions, is a therapeutically useful. Generally, the longer the retention in treatment, the better the long-term outcome. Self-help groups can fill the void in the patient's life once occupied by cocaine, but other drug-free alternatives exist.

Cocaine use is often linked to compulsive sexual activity. Some practitioners believe that it is important to ascertain what kinds of sexual behaviors and fantasies are associated with being high on cocaine, since sexual feelings can trigger a craving for cocaine. It may sometimes be appropriate to ask patients to refrain from sex for the first weeks of treatment.

Comparison of Psychotherapy Methods and Programs A large-scale collaborative multisite cocaine treatment study sponsored by the National Institute on Drug Abuse (NIDA) compared different psychosocial treatments. Following a brief period of stabilization, 487 cocaine-dependent patients were randomly assigned to one of four groups: weekly group drug counseling, group counseling plus individual drug counseling based on 12-step principles, group counseling plus individual cognitive therapy, or group counseling plus individual supportive expressive therapy. Group drug counseling was provided for 6 months; individual therapies were provided twice weekly for 3 months, then once weekly for 3 months. Therapy was manual guided, and cognitive therapy and supportive expressive therapists were fully trained professionals. Drug counselors had extensive experience with drug dependence treatment; about one third were in recovery from drug dependence. All patients reported substantial reduced cocaine use, whether measured by Addiction Severity Index composite, days of cocaine use in the past month, or number of months abstinent. Follow-up occurred 1 year after treatment entry (which for some patients was only 6 months after completion). Patients assigned to group drug counseling plus individual drug counseling reported significantly better outcomes; patients assigned to either cognitive therapy or supportive expressive therapy stayed in treatment longer, but outcomes in terms of cocaine use or dependence were not significantly better than those for group drug counseling alone. Among patients assigned to individual counseling, 73 percent achieved 1 month of complete abstinence, and 36 percent achieved 3 consecutive months of abstinence. In the other groups, 17 to 25 percent achieved 3 months of abstinence. By 6 months posttreatment, no delayed benefits of psychotherapy had emerged. Psychiatric severity and the presence of antisocial personality disorder did not significantly affect treatment outcome.

Patients assigned to once-weekly group counseling reported substantially reduced cocaine use starting from the first month of treatment. Although only 23 percent of patients continued for the 6 months of treatment available, the reported improvement in drug use was sustained through the 1 year follow-up. Adding individual sessions of either cognitive therapy or supportive expressive therapy twice weekly (provided by professionally trained therapists) to the weekly group counseling session increased retention in treatment but did not increase the proportion of patients who reported being abstinent or using substantially less cocaine.

In evaluating these outcomes one must know that all patients participated in a stabilization phase lasting 1 to 2 weeks during which they were required to attend one group session and two case-management visits before being assigned to a specific treatment. During that time there was attrition of less motivated patients. The therapists were highly qualified, carefully trained, used a manual to guide therapy, and were supervised. The study population was 77 percent male, 58 percent white and 60 percent employed, with a mean age of 34 and generally low psychiatric severity. Patients taking psychotropic medication, and those with schizophrenia, bipolar disorder, polysubstance dependence, or opioid dependence were excluded. However, 33 percent met criteria for alcohol dependence, 28 percent those for cocaine-induced mood disorder, 14 percent met full criteria for antisocial personality disorder, and 32 percent for adult antisocial personality disorder with history of conduct disorder.

Pharmacological Adjuncts Presently no pharmacological treatments produce decreases in cocaine use that compare with the decreases in opioid use seen when heroin users are treated with methadone, levomethadyl acetate (ORLAAM) (commonly called L-a-acetylmethadol [LAAM]) or buprenorphine. However, a variety of pharmacological agents, most of which are approved for other uses, have been and are being tested clinically for the treatment of cocaine dependence and relapse. Some of these agents are being used routinely by clinicians although little solid evidence exists for their efficacy. The most common premises on which

pharmacological interventions are based are as follows: (1) chronic cocaine use alters dopaminergic systems, so that giving up the drug is associated with a hypodopaminergic state characterized by dysphoria or anhedonia; (2) some cocaine users are using the drug to ameliorate a preexisting psychiatric disorder, such as major depressive disorder, dysthymic disorder, attention-deficit disorder, or cyclothymic disorder; (3) cocaine produces a sensitization, or kindling, effect that somehow predisposes to continued use; and (4) relapse is related to memories of the reinforcing and euphoric effects of cocaine, craving for which can be elicited by stress, other drugs, or environmental stimuli.

Cocaine users presumed to have preexisting attention-deficit/hyperactivity disorder or mood disorders have been treated with methylphenidate (Ritalin) and lithium (Eskalith), respectively. Those drugs are of little or no benefit in patients without the disorders, and clinicians should adhere strictly to maximal diagnostic criteria before using either of them in the treatment of cocaine dependence. In patients with attention-deficit/hyperactivity disorder, slow-release forms of methylphenidate may be less likely to trigger cocaine craving, but the impact of such pharmacotherapy on cocaine use remains to be demonstrated.

Many pharmacological agents have been explored on the premise that chronic cocaine use alters the function of multiple neurotransmitter systems, especially the dopaminergic and serotonergic transmitters regulating hedonic tone, and that cocaine induces a state of relative dopaminergic deficiency. Although the evidence for such alterations in dopaminergic function has been growing, it has been difficult to demonstrate that agents theoretically capable of modifying dopamine function can alter the course of treatment. This has been so even when studies in animal models and open-label studies suggested that they would be successful. In well-designed, controlled trials that obtained objective evidence of drug use, the following agents are among those that have not been found to reduce cocaine use: neurotransmitter precursors, (e.g., dopa; tyrosine); dopaminergic agonists, (bromocriptine [Parlodel]; lisuride; pergolide [Permax]); and antiparkinson drugs that may also affect the dopaminergic system (amantadine [Symmetrel]).

Tricyclic antidepressant drugs such as desipramine and imipramine (Tofranil) have also been tried. Although some double-blind studies that relied heavily on self-reports of drug use yielded some positive results, other studies have not found them significantly beneficial in inducing abstinence or preventing relapse. There is no consensus that the effects of desipramine are robust or reliable enough to justify routine use, but used early in treatment, it may have some transient benefit for patients who are less severely dependent.

Also tried in pilot or open-label studies but not confirmed effective in controlled studies, are other antidepressants, such as bupropion (Wellbutrin), monoamine oxidase (MAO) inhibitors (selegiline [Eldepryl]); selective serotonin uptake inhibitors (SSRIs) (e.g., fluoxetine [Prozac]); mazindol (Sanorex); pemoline (Cylert); antipsychotics (e.g., flupenthixol); lithium; several different calcium channel inhibitors, anticonvulsants (e.g., carbamazepine [Tegretol] and valproic acid [Depakene]). One double-blind study not yet replicated found that 300 mg a day of phenytoin (Dilantin) reduced cocaine use.

Several agents are being developed but have not been tried in human studies. These include agents that would selectively block or stimulate dopamine receptor subtypes (e.g., selective D_1 agonists) and drugs that can selectively block the access of cocaine to the dopamine transporters but still permit the transporters to remove cocaine from the synapse. Another approach is aimed at preventing cocaine from reaching the brain by using antibodies to bind cocaine in the bloodstream (a so-called cocaine vaccine). Such cocaine-binding antibodies do reduce the reinforcing effects of cocaine in animal models. Also under study are catalytic antibodies that accelerate the hydrolysis of cocaine; and butyrylcholinesterase (pseudocholinesterase), which appears to hydrolyze cocaine selectively and is normally present in the body.

Acupuncture Use of auricular acupuncture to treat cocaine and other varieties of dependence behavior has become popular among some groups, including some drug courts and prison-based programs. Controlled studies of its efficacy for treating cocaine dependence (using sham acupuncture) have been conducted but are subject to varying interpretations; large differences in cocaine use (as measured by urine tests) have not been shown. In programs offering treatment in a drop-in outpatient setting, clients typically are instructed to stay as clean and sober as they can and to come in daily for treatment; clinic rules are minimal; treatment usually takes place in groups, with patients sitting in comfortable chairs for about 45 minutes. Dropout rates are generally high. Herbal teas are often consumed as part of the treatment.

Also used for treatment of cocaine dependence are several forms of transcranial electrical stimulation (neuroelectric therapy [NET]). A comparison of NET with sham NET revealed no differences in successful detoxification over a 12-day hospitalization.

Special Populations and Special Considerations

Mixed Addictions

PATIENTS MAINTAINED ON METHADONE Behavioral techniques and pharmacological agents have been used to help patients maintained on methadone reduce their use of cocaine (as measured by urine tests). Some methadone programs use progressive sanctions, such as decreased take-home privileges, a decreased methadone dosage, and finally in some cases, discharge from the program. However, in one comparison study, decreasing the methadone dosage proved far less effective than giving small (5 mg) increases (up to 120 mg a day in some cases) for each cocaine-positive urine test. However, an analysis of the relation between cocaine use and methadone dose at another clinic did not find less cocaine use among patients maintained on higher doses of methadone. Providing small rewards (such as vouchers for goods and services) contingent on submitting a urine specimen negative for cocaine does reduce the frequency of cocaine use.

Other pharmacological agents that have been tried for methadone-maintenance patients who also use cocaine include bromocriptine, amantadine, desipramine, bupropion, carbamazepine, and buprenorphine. In a controlled comparison of amantadine, desipramine, and placebo, self-reports of cocaine use were lower for both medication groups at 4 weeks, but there were no significant differences in cocaine-positive urine tests, and there were no significant group differences at 8 or 12 weeks.

Early animal studies and open-label clinical studies suggested that buprenorphine might help decrease cocaine use among patients dependent on both opioids and cocaine. In three double-blind controlled studies comparing buprenorphine and methadone (one of which was not specifically designed to test effects on cocaine use), cocaine use was not affected, although both buprenorphine and methadone significantly decreased heroin use. In animal models buprenorphine appears to antagonize some of the lethal effects of cocaine and to increase its reinforcing effects. Carbamazepine, which substantially stimulates the metabolism of methadone, has not shown any specific effects on cocaine use in controlled studies. In summary, no pharmacological agent has been shown to have reliably robust effects on cocaine use among patients maintained on methadone.

COCAINE AND ALCOHOL In patients dependent on both cocaine and alcohol, the opioid antagonist naltrexone (Revia) had no effect on cocaine use. In an open-label trial, disulfiram (Antabuse) seemed useful for reducing cocaine use, perhaps because it discouraged the use of alcohol which is often used with cocaine.

Women, Pregnant Women, and Their Children Data suggest that although women who seek treatment tend to be more severely drug dependent, they respond as well to treatment as do men. Women dependent on cocaine have a number of special needs, especially with respect to their physical health. Cocaine use by pregnant women represents a hazard to the fetus.

At the peak of the cocaine epidemic, 10 to 45 percent of women who received obstetrical care in some urban hospitals reported using cocaine at some time during pregnancy. There is some controversy about the frequency and permanence of any damage sustained by the fetus, but there is little question that maternal cocaine use can be associated with some perinatal morbidity and mortality. Separating cocaine effects from the effects of other substances and of other maternal behavior is exceedingly difficult but some toxicity may be due to cocaine-induced hypertension, tachycardia, and vasoconstriction, which lead to impaired placental blood flow and decreased transfer of nutrients and oxygen to the fetus.

Some toxicity also results from direct effects of cocaine on the fetus. Depending on the severity of the placental and fetal effects and when they occur during gestation, the result may be teratogenic, with destruction of developing tissues or overall retardation of fetal growth. Commonly reported abnormalities in fetuses exposed to cocaine are microcephaly and structural abnormalities in brain and urinary tract development. Ischemic and hemorrhagic lesions in the newborn brain have also been reported. Spontaneous abortions, premature birth, placenta previa, and abruptio placenta are complications of pregnancy that are more common among women who use cocaine than among nonusers; low-birth-weight babies are also common.

Despite the risks, only a small percentage of the infants exhibit what might be called a neonatal cocaine exposure syndrome, which consists of poor feeding, irritability, tremor, and abnormal sleep patterns. Those abnormalities are most evident on the second day after birth and last for less than a week or two. Sudden infant death syndrome (SIDS) is reported to be more common among infants exposed to cocaine in utero, but since there are no controls, the evidence for this is not

conclusive. The long-term neurological, cognitive, and developmental consequences of intrauterine cocaine exposure are still not clear, but after the first few months, most of these children appear to be developmentally within normal limits. A follow-up study tested children ages 6 to 9 who were exposed prenatally to cocaine and then compared them with unexposed controls matched for gender, birth weight, ethnicity, and socioeconomic status. Intelligence quotient (I.Q.) scores did not differ and were unchanged when adjusted for caregiver I.Q. and home environment.

There appears to be no contraindication to discontinuing cocaine abruptly during pregnancy (unlike opioids), and prompt abstinence from cocaine should be the goal of treatment.

Patients With Other Psychiatric Disorders Persons with cocaine-dependence who also have mood or anxiety disorders are generally managed in programs that focus on the substance-use problem. Several clinical reports indicate that cocaine users with bipolar disorders generally are not compliant with prescribed lithium.

Patients with a history of attention-deficit/hyperactivity disorder are also likely to have antisocial personality disorder. Although some studies found that patients with antisocial personality disorder and cocaine abuse responded relatively poorly to treatment, more recent studies found that in terms of their cocaine use, improvement was comparable to that of patients who did not have antisocial personality disorder. Up to 10 percent of persons meeting criteria for cocaine abuse also met DSM-IV criteria for adult attention-deficit/hyperactivity disorder. Clinical trials using methylphenidate on the assumption that cocaine use was an attempt to self-medicate attention-deficit/hyperactivity disorder have led to the caution that in most cases, the patient's demand for methylphenidate escalates and craving for cocaine is stimulated. A slow-release form of methylphenidate may be less likely to elicit craving, but its efficacy in reducing cocaine use has not been demonstrated.

PATIENTS WITH DEPRESSION Symptoms of depression are common among patients seeking treatment for cocaine dependence. In studies of tricyclic drugs and SSRIs in patients who were using cocaine, outcomes of those who initially had depressive symptoms and patients who were initially without such symptoms did not differ. In a large-scale study of psychosocial treatment, cocaine-dependent patients who met criteria for any Axis I disorder were less likely to drop out once they became engaged in treatment. The relation of symptoms of depression and of concurrent major depression to treatment outcome needs further clarification. The most sensible course is to treat significant depression with antidepressants only if they persist after cessation of drug use.

PATIENTS WITH SCHIZOPHRENIA Persons with schizophrenia and other psychotic disorders who use cocaine have been managed within either primary drug treatment or psychiatric facilities. In all settings, concurrent use of alcohol and cocaine further complicates treatment. There is a growing consensus that parallel treatment (two separate programs, one treating substance dependence/abuse and the other treating schizophrenia) is less effective than a comprehensive integrated program that deals with both disorders concurrently.

Intensive case management that gives patients access to social services makes it possible to treat patients with schizophrenia who abuse cocaine in the same day-hospital setting as non-substance-abusing patients. However, requiring abstinence for admission or retention may be unrealistic for such patients, and some of the traditional rules concerning substance abuse and poor attendance may need to be relaxed. Most patients are not initially motivated to participate in abstinence-oriented programs, but when attention is paid to a patient's level of motivation, most can be engaged and moved toward active treatment.

The use of cocaine, amphetamines, and cannabis exacerbates schizophreniform disorder, and such use is not an uncommon problem. Among patients receiving public assistance or disability payments, such use seems to increase substantially shortly after they receive monthly checks. Some cocaine (or stimulant) use may represent an attempt to alleviate negative symptoms, depression, or the side effects of antipsychotic agents. The last problem might be dealt with by using newer antipsychotic agents that have fewer extrapyramidal adverse effects.

SUGGESTED CROSS-REFERENCES

See Chapter 1 for discussion of the neural sciences and Chapter 2 for a presentation of neuropsychiatry and behavioral neurology. A classification of mental disorders appears in Chapter 9. An introduction and overview of substance-related disorders is presented in Section 11.1, and amphetamine-related disorders in Section 11.3. Various drugs are discussed in the chapter on biological therapies (Chapter 31), particularly sympathomimetics in Section 31.21. Schizophrenia is discussed in Chapter 12, and other psychotic disorders in Chapter 13. Animal research and its relevance to psychiatry is discussed in Section 5.4. Cognitive-behavioral therapy is discussed in Section 48.3

SECTION REFERENCES

Alterman AI, McLellan AT: Inpatient and day hospital treatment services for cocaine and alcohol dependence. J Subst Abuse Treatment *10:*269, 1993.

Anthony JC, Warner LA, Kessler RC: Comparative epidemiology of dependence on tobacco, alcohol, controlled substances, and inhalants: Basic findings from the National Comorbidity Survey. Exp Clin Psychopharmacol *2:*244, 1994.

Benowitz NL: How toxic is cocaine? In *Cocaine: Scientific and Social Dimensions*, Ciba Foundation Symposium 166, GR Bock, J Whelan, editors. Wiley, New York, 1992.

Brown RA, Monti PM, Myers MG, Martin RA, Rivinus T, Dubreuil ME, Rohsenow DJ: Depression among cocaine abusers in treatment: Relation to cocaine and alcohol use and treatment outcome. Am J Psychiatry *155:*220, 1998.

*Bullock ML, Kiresuk TJ, Pheley AM, Culliton PD, Lenz SK: Auricular acupuncture in the treatment of cocaine abuse: A study of efficacy and dosing. J Subst Abuse Treatment *16:*31, 1999.

Carroll KM, Rounsaville BJ: Contrast of treatment-seeking and untreated cocaine abusers. Arch Gen Psychiatry *49:*464, 1992.

Carroll KM, Rounsaville BJ, Gordon LT, Nich C, Jatlow P, Bisighini RM, Gawin FH: Psychotherapy and pharmacotherapy for ambulatory cocaine abusers. Arch Gen Psychiatry *51:*177, 1994.

*Childress AR, Mozley PD, McElgin W, Fitzgerald J, Reivich M, O'Brien CP: Limbic activation during cue-induced cocaine craving. Am J Psychiatry *156:*1, 1999.

Daras M: Neurologic complications of cocaine. NIDA Research Monogr *163:*43, 1996.

Drake RE, Mueser KT, Clark RE, Wallach MA: The course, treatment and outcome of substance disorder in persons with severe mental illness. Am J Orthopsychiatry *66:*42, 1996.

Fletcher BW, Tims FM, Brown BS: Drug Abuse Treatment Outcome Study (DATOS): Treatment evaluation research in the United States. Psychol Addict Behav *11:*216, 1997.

Foltin RW, Fischman MW: Effects of “binge” use of intravenous cocaine in methadone-maintained individuals. Addiction *93:*825, 1998.

Gatley SJ, Volkow ND: Addiction and imaging of the living human brain. Drug Alcohol Depend *51:*97, 1998.

Gawin FH, Ellinwood EH: Cocaine and other stimulants. N Engl J Med *318:*1173, 1988.

Gold MS, Miller NS: Cocaine (and crack): Neurobiology. In *Substance Abuse: A Comprehensive Textbook*, ed 3, JH Lowinson, P Ruiz, RB Millman, JG Langrod, editors. Williams & Wilkins, Baltimore, 1997.

*Gorelick DA: Pharmacologic therapies for cocaine and other stimulant addiction. In *Principles of Addiction Medicine*, ed 2, AW Graham, TK Schultz, editors. American Society of Addiction Medicine, Chevy Chase, MD, 1998.

Higgins ST, Budney AJ, Bickel WK, Foeng FE, Donham R, Badger GJ: Incentives improve outcome in outpatient behavioral treatment of cocaine dependence. Arch Gen Psychiatry *51:*568, 1994.

Hubbard RL, Craddock SG, Glynn PM, Anderson J, Etheridge RM: Overview of 1-year-follow-up outcomes in the Drug Abuse Treatment Outcome Study (DATOS). Psychol Addict Behav *11:*261, 1997.

Hyman SE, Nestler EJ: Initiation and adaptation: A paradigm for understanding psychotropic drug action. Am J Psychiatry *153:*151, 1996.

Jaffe JH, Cascella NG, Kumor KM, Sherer MA: Cocaine-induced cocaine craving. Psychopharmacology *97:*59, 1989.

*Kreek MJ, Koob GF: Drug dependence: Stress and dysregulation of brain reward pathways. Drug Alcohol Depend *51:*23, 1998.

Levin FR, Evans SM, Kleber HD: Prevalence of adult attention-deficit hyperactivity disorder among cocaine abusers seeking treatment. Drug Alcohol Depend *52:*15, 1998.

Little KY, McLaughlin DP, Zhang L, McFinton PR, Dalack GW, Cook EH Jr, Cassin GJ, Watson SJ: Brain dopamine transporter messenger RNA and binding sites in cocaine users. Arch Gen Psychiatry *55:*793, 1998.

McLellan AT, Grossman DS, Blaine JD, Haverkos HW: Acupuncture treatment for drug abuse: A technical review. J Substance Abuse Treatment *10:*569, 1993.

Mendelson JH, Sholar M, Mello NK, Teoh SK, Sholar JW: Cocaine tolerance: Behavioral, cardiovascular, and neuroendocrine function in men. Neuropsychopharmacology *18:*263, 1998.

*Mulvaney FD, Alterman AI, Boardman CR, Kampman K: Cocaine abstinence symptomatology and treatment attrition. J Subst Abuse Treatment *16:*129, 1999.

Musto D: Opium, cocaine and marijuana in American history. Sci Am *265:*40, 1991.

*Ness RB, Grisso JA, Hirschinger N, Markovic N, Shaw LM, Day NL, Kline J: Cocaine and tobacco use and the riskof spontaneous abortion. N Engl J Med *340:*333, 1999.

*Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LJ, Goodwin FK: Comorbidity of mental disorders with alcohol and other drug abuse. JAMA *264:*2511, 1990.

Rounsaville BJ, Anton SI, Carroll K, Budde D, Prusoff BA, Gawin F: Psychiatric diagnoses of treatment-seeking cocaine abusers. Arch Gen Psychiatry *48:*43, 1991.

Rounsaville BJ, Bryant K: Tolerance and withdrawal in the DSM-III-R diagnosis of substance dependence. Am J Addictions *1:*50, 1992.

Rounsaville BJ, Bryant K, Babor T, Kranzler H, Kadden R: Cross system agreement for substance use disorders: DSM-III-R, DSM-IV, and ICD-10. Addiction *88:*337, 1993.

Schottenfeld RS, Pakes JR, Oliveto A, Ziedonis D, Kosten TR: Buprenorphine versus methadone maintenance treatment for concurrent opioid dependence and cocaine abuse. Arch Gen Psychiatry *54:*713, 1997.

Self DW, Nestler EJ: Relapse to drug-seeking: Neural and molecular mechanisms. Drug Alcohol Depend *51:*49, 1998.

Shaner A, Khalsa E, Roberts L, Wilkins J, Anglin D, Hsieh S-C: Unrecognized cocaine use among schizophrenic patients. Am J Psychiatry *150:*758, 1993.

Silverman K, Higgins ST, Brooner RK, Montoya ID, Cone EJ, Schuster CR, Preston KL: Sustained cocaine abstinence in methadone maintenance patients through voucher-based reinforcement therapy. Arch Gen Psychiatry *53:*409, 1996.

Siqueland L, Crits-Christoph P, Frank A, Daley D, Weiss R, Chittams J, Blaine J, Luborsky L: Predictors of dropout from psychosocial treatment of cocaine dependence. Drug Alcohol Depend *52:*1, 1998.

Tsuang MT, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L: Co-occurrence of abuse of different drugs in men. Arch Gen sychiatry *55:*967, 1998.

van den Bree MB, Svikis DS, Pickens RW: Genetic influences in antisocial personality and drug use disorders. Drug Alcohol Depend *49:*177, 1998.

*van Gorp WG, Wilkins JN, Hinkin CH, Moore LH, Hull J, Horner MD, Plotkin D: Declarative and procedural memory functioning in abstinent cocaine abusers. Arch Gen Psychiatry *56:*85, 1999.

Wasserman GA, Kline JK, Bateman DA, Chiriboga C, Lumey LH, Friedlander H, Melton L, Heagarty MC: Prenatal cocaine exposure and school-age intelligence. Drug Alcohol Depend *50:*203, 1998.

Weddington WW, Brown BS, Haertzen CA, Cone EJ, Dax EM, Herning RI, Michaelson MA: Changes in mood, craving and sleep during short-term abstinence reported by male cocaine addicts. Arch Gen Psychiatry *47:*861, 1990.

11.7 HALLUCINOGEN-RELATED DISORDERS

HENRY DAVID ABRAHAM, M.D.

The discovery of lysergic acid diethylamide (LSD) by Albert Hofmann in 1943 proved to be a double-edged sword that simultaneously widened the understanding of central mechanisms of neural regulation and led to the widespread abuse of hallucinogenic drugs. The emergence of this semisynthetic hallucinogen raised important questions about synaptic transmission, the mechanisms of hallucinations, and the phenomenology of the functional psychoses. However, the possibilities of the use and misuse of such hallucinogens was greater than that of botanical hallucinogens such as psilocybin mushrooms and peyote cacti because synthetic hallucinogens are readily made, easily distributed, cheaply sold, and thousands of times more potent than botanical preparations. These factors, among others, led to the abuse of synthetic hallucinogens and to the development of several disorders now seen in psychiatric practice.

DEFINITION

An hallucinogenic drug has been defined by the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as “any agent which has alterations in perception, cognition and mood as its primary psychobiological actions in the presence of an otherwise clear sensorium. Most commonly this includes indoleakylamines and phenethylamines, and excludes, inter alia, the anticholinergics, the arylcyclohexylamine dissociative anesthestics such as phencyclidine, stimulants such as amphetamine and cocaine, bromism and heavy metal intoxication.” Excluded from this class are agents that produce hallucinations in the context of a delirium. The term *hallucinogen* emphasizes perceptual effects, but the literature supports ample evidence of the effects of hallucinogens on mood and cognition as well. Table 11.7-1 illustrates the representative hallucinogens.

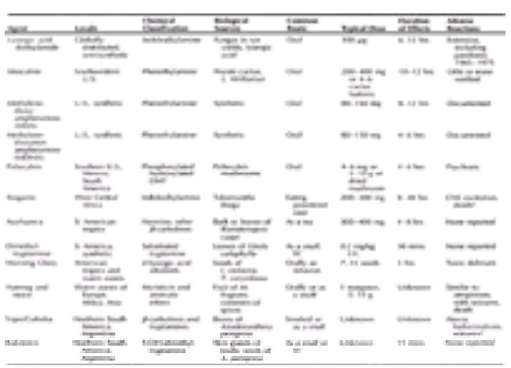

Table 11.7-1 Overview of Representative Hallucinogens

By definition these drugs are intoxicants. Clinically the use of hallucinogenic drugs is associated with panic attacks, hallucinogen persisting perception disorder (flashbacks), psychosis, delirium, and mood and anxiety disorders. Empirical evidence more strongly supports the first three as discussed below. Although DSM-IV defines hallucinogen abuse and dependence with criteria applicable to other agents of abuse, hallucinogens differ from addictive drugs in that cessation after long-term use is not associated with a distinct withdrawal syndrome.

HISTORY

The omnivorous diet of primeval man no doubt led to the discovery of certain plants which, while meager in nutrition, possessed extraordinary abilities to alter consciousness. Thus was introduced to history a class of plants that assumed importance in ritual, religion, and recreation. One of the earliest plants used for its mind-altering properties is the fly agaric mushroom, *Amanita muscaria*, which is thought to have been discovered by aboriginal Siberian hunters who observed intoxicated reindeer. The same plant is believed to have been described as *soma* in the 3500 year old Sanskrit text, the Rigveda. The Eleusian Mysteries of ancient Greece were thought to have used ergot alkaloids in the holy potion, *kykeon*, to induce mystical states. In the Middle Ages women used brooms as medicinal applicators for the vaginal insertion of hallucinogenic ointments, a probable origin of the contemporary association of brooms with witches. In 944 AD the ingestion of ergot-laden rye, produced by the ergot fungus *Claviceps purpurea*, is said to have killed 40,000 people in Europe.

In the nineteenth century the mycologist Mordecai Cubit Cooke discriminated opiates from hallucinogens, a distinction that was more precisely drawn by Ernest Bosc DeVeze and Louis Lewin decades later. A Texas physician, John Briggs, described his self-experimentation with muscale buttons in 1887, the alkaloids of which Lewin described in the following year. Field work began in earnest in the 1920s, leading to the identification of 120 plants possessing hallucinatory properties, 80 percent of which are found in the western hemisphere. In 1928 Heinrich Klüver used mescaline to analyze the formal structure of hallucinations. The ethnobotanical work of Richard Schultes and others documented the use of hallucinogens among contemporary cultures such as the Waiká of Brazil and the Huichol of Mexico. Psychoactive agents were described in the bark of Banisteriopsis (b-carbolines), seeds of the common Morning Glory (lysergic acid), and skin glands of toad Bufo bufo (bufotenine). In 1943 Albert Hofmann discovered LSD by tracking down its identity after he accidentally ingested the drug and suffered a transient psychosis. Over the next 20 years psychiatrists in England and the United States attempted to use LSD as a therapeutic agent for a variety of illnesses, with unconvincing results. By the 1950s the drug was used by academicians, theologians, and the military for a variety of purposes. In the 1960s fascination with the drug, fueled by the American media, exploded in an epidemic of hallucinogen use involving much of Europe and the United States. The residuum of this epidemic continues today as an endemic among the young.

COMPARATIVE NOSOLOGY

DSM-IV lists a number of hallucinogen-related disorders (Table 11.7-2), but contains specific diagnostic criteria only for hallucinogen intoxication and hallucinogen persisting perception disorder (flashbacks). The diagnostic criteria for the other hallucinogen-related disorders are contained in the DSM-IV sections that are specific to each symptom. For example, hallucinogen-induced mood disorder is discussed with other mood disorders. Table 11.7-2 compares several diagnostic systems.

Hallucinogen use disorders
Hallucinogen dependence
Hallucinogen abuse
Hallucinogen-induced disorders
Hallucinogen intoxication
Hallucinogen persisting perception disorder (flashbacks)
Hallucinogen intoxication delirium
Hallucinogen-induced psychotic disorder, with delusions
Specify if:
With onset during intoxication
Hallucinogen-induced psychotic disorder, with hallucinations
Specify if:
With onset during intoxication
Hallucinogen-induced mood disorder
Specify if:
With onset during intoxication
Hallucinogen-induced anxiety disorder
Specify if:
With onset during intoxication
Hallucinogen-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. ©American Psychiatric Association, Washington, DC, 1994.

Table 11.7-2 DSM-IV Hallucinogen-Related Disorders

Differences between the revised third edition of DSM-III-R and DSM-IV are as follows: (1) psychoactive substance abuse was a residual category in DSM-III-R; DSM-IV provides specific diagnostic criteria for substance abuse; (2) the DSM-IV criteria for substance dependence are also modified from those of DSM-III-R; (3) DSM-IV generally permits subtyping of dependence with (or without) physiological dependence, but because hallucinogens produce no withdrawal, that option does not apply here; DSM-IV expands the list of hallucinogen mental disorders to include (4) hallucinogen intoxication delirium and (5) hallucinogen-induced anxiety disorder. Hallucinogen-induced psychotic disorder in DSM-IV is coded as a substance-induced psychotic disorder, developed concurrently with drug use, and is characterized by a longer-than-expected period of psychotic symptoms; as a rule of thumb for hallucinogens this means beyond 48 hours. Hallucinogen persisting perception disorder in DSM-IV applies to intermittent flashbacks, recurring days to years after the last hallucinogen use. This disorder now appears in certain subjects to be both permanent and constant. The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) makes provision for flashbacks, but confuses the issue by considering them a psychotic disorder, which most authorities do not. DSM-IV provides no diagnosis for persisting hallucinogen-induced psychosis, although the literature now supports the existence of this disorder.

In general, ICD-10 does not draw helpful distinctions between hallucinogen-related disorders. ICD-9 provides for seven hallucinogen-induced diagnoses, but summarily lumps four of them under “nondependent abuse.” The *International Classification of the Diseases, 9th Revision, Clinical Modification* (ICD-9-CM), a U.S. government classification for Medicare and Medicaid billing, calls intoxication “abuse” and provides no separate diagnosis of hallucinogen-induced anxiety disorder or flashbacks. ICD-10 names abuse “harmful use,” and has neither hallucinogen-induced mood disorder nor hallucinogen-induced anxiety disorder diagnoses; in part, these omissions reflect the lack of systematic research of these entities.

EPIDEMIOLOGY

Surveys in the United States and Europe report that the use of hallucinogenic drugs among the young exceeds that of cocaine and heroin use. In the United Kingdom from 1989 to 1993 use of LSD rose from 7 to 11 percent. A survey of 781 German adolescent drug abusers found that 14.1 percent used LSD. A survey of 458 Danish students found that 7.2 percent experimented with a European hallucinogenic mushroom, *P. semilanceata*. Survey data in the United States show higher trends. In an analysis of data from two large universities, 13.7 percent of the students surveyed reported using hallucinogenic mushrooms. Trend data, gathered annually from approximately 15,000 U.S. high school seniors, reveal a precipitous rise in hallucinogen use from 10.3 to 15.1 percent in the decade from 1987 to 1997, beginning in 1991 (Fig. 11.7-1). Demographic data from 18,054 American householders in 1990 shows that LSD use is most likely to occur between the ages of 18 and 25. Use is more common among males, whites, and Hispanics and is more likely to occur in the American Northeast and West. Use is associated with lower levels of employment and education, although the parents of LSD users are more often of a white-collar socioeconomic status. Among American undergraduates the prevalence of the use of 3, 4-methylenedioxy-N-methylamphetamine (MDMA, “ecstasy”) has ranged widely from 2.3 to 24 percent in the first half of the 1990s.

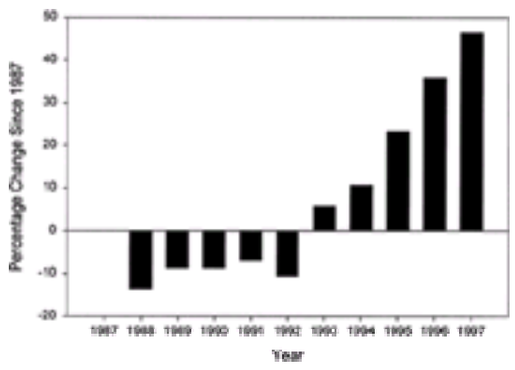

FIGURE 11.7-1 Percentage change in lifetime use of hallucinogen use in U.S. high school seniors since 1987.

PHARMACOLOGY

Several hundred known plants possess hallucinogenic properties. They vary widely in potency, but are orders of magnitude less potent than synthetic agents. Alexander Shulgin has synthesized and screened 179 phenethylamines for these effects in humans. The commonly used hallucinogens are the peyote cactus, containing mescaline, which is indigenous to the American Southwest and northern Mexico; mushrooms producing psilocybin and its active metabolites; and chemically substituted tryptamines such as dimethyltryptamine (DMT) and LSD. Hallucinogens have been classified according to their course of action assuming their usual clinical route of administration. By this schema, parenteral DMT is ultra–short-acting, with an onset when given parenterally of 1 to 5 minutes. Diethyltryptamine is short-acting (15–30 minutes), oral psilocybin is intermediate (30–60 minutes orally), LSD is long-acting (4–10 hours), and oral 2, 5-dimethoxy-4-methyl-amphetamine (DOM) and ibogaine are considered ultra–long-acting, lasting as long as 24 hours.

Of the hallucinogens the pharmacology of LSD has been studied most extensively. The drug is active orally in doses of 50 to 100 µg. It is readily absorbed through the mucous membranes and exerts its psychological effects within 20 minutes. The drug is marketed illicitly as single crystals, drops deposited on blotter paper, or in sugar cubes. A drop of LSD when placed on a piece of blotter paper is for all practical purposes invisible. To mark its location, the spot is often stamped with an image of a cartoon character or New Age design. The drug's potency is underscored by the report of a drug dealer who was apprehended after he accidentally became psychotic from sitting on a wet surface with a sheet of “blotter acid” in his back pocket. Because of such potency the drug is seldom injected. Although it can cause a toxic psychosis lasting 6 to 12 or more hours, only 1 percent of the ingested drug reaches the central nervous system. Hallucinogenic drugs bind to multiple synaptic receptors in the brain. It has been shown that hallucinogenic potency is associated with binding affinities at the serotonin type 2 (5-hydroxytryptamine [5-HT $_2$]) receptor where these drugs act as partial agonists. Repeated administration of LSD in rats downregulates the density of 5-HT $_2$ binding sites, further implicating this receptor in the drug's mechanism of action. In monkey brain LSD is anatomically distributed maximally in the pituitary, pineal, the visual and auditory cortex, the hypothalamus, and the limbic cortex. This parallels the finding of high concentrations of 5-HT $_2$ receptors in human cerebral cortex.

Studies in animals have produced mixed results on toxicity. Animals as a rule do not self-administer hallucinogens as they do reinforcing agents such as heroin and cocaine, possibly reflecting the drugs' predilection for affecting higher cortical centers that are less developed in nonhuman mammals. The LD $_{50}$ of LSD in mice is 150,000 µg/kg, whereas an elephant was killed by an LSD injection of 100 µg/kg. Accidental ingestion of between 10,000 and 100,000 µg LSD by eight persons resulted in confusion, hallucinations, and hemorrhage, but no deaths.

However, fatalities have been documented with MDMA and 3,4-methylenedioxy-N-ethylamphetamine (MDEA, “Eve”). These drugs are ring-substituted isopropylamines. Drug discrimination studies in rats suggest that MDMA is “amphetamine-like.” Although MDMA was first synthesized in 1914, the drug achieved prominence in 1986 when clinicians claimed that the drug enhanced psychotherapy and improved social function. The following year widespread use was reported on

college campuses. In the following decade serotonergic neurotoxicity was described in numerous animal models for both MDMA and 3,4-methylenedioxy-N-amphetamine (MDA). MDMA deaths have been associated with drug-induced compulsive dancing, hyponatremia, and hyperthermia. Autopsy studies reveal massive hepatic necrosis and changes of heat stroke. Serotonergic cells appear particularly sensitive to MDMA; for these cells the drug is cytotoxic in rodents and monkeys, with effects lasting as long as 18 months. The coadministration of either selective serotonin uptake inhibitors or 5-HT $_2$ antagonists in animals appear to attenuate MDMA neurotoxicity.

LSD is readily metabolized by the liver through hydroxylation and conjugation. Although tolerance occurs after 2 to 3 days of serial use, there is no withdrawal syndrome from hallucinogens, and no physiological dependence. Psychological dependence has been clinically observed. Cross-tolerance occurs between LSD, mescaline, and psilocybin, but not between these agents and amphetamines or marijuana. Phenylalkylamines, however, may elicit both hallucinogenic and amphetamine properties in varying ratios depending on the nature of the substituted moieties.

PSYCHOPHARMACOLOGY

The classic description of the psychological effects of LSD remains that by Albert Hofmann, recalled in retrospect after an attempt to record his experiences ended with the entry into his journal, “Beginning dizziness, feeling of anxiety, visual distortions, symptoms of paralysis, desire to laugh—”

In addition to these symptoms, there are signs of sympathetic arousal, including mydriasis, hypertension, hyperreflexia, tachycardia, and tremor. Following this period of arousal, visual illusions and hallucinations may occur. They can be in any sensory modality, but visual events predominate. Synesthesia, in which one sensory modality stimulates a second one, (seeing sounds or hearing colors) is clinically rare, and probably reflects heightened connectivity between adjacent cerebral regions. Mood changes, by comparison, can be profound or can or reflect preexisting emotional conflicts. Feelings of terror, depression, and panic, described by Hofmann, may present to the clinician as a psychiatric emergency. As these perceptual and affective components of the trip subside, the user may be left with cognitive states of transcendence or paranoia. The entirety of the trip seldom lasts more than 12 hours (Table 11.7-3).

DSM-IV	DSM-III-R	ICD-9	ICD-9-CM	ICD-10
Hallucinogen abuse	Hallucinogen abuse	Nondependent abuse of hallucinogens	Hallucinogen abuse	Hallucinogen harmful use
Hallucinogen dependence	Hallucinogen dependence	Hallucinogen dependence	Hallucinogen dependence	Hallucinogen dependence syndrome
Hallucinogen intoxication	Hallucinogen hallucinosis	Nondependent abuse of hallucinogens	Hallucinogen abuse	Hallucinogen acute intoxication
Hallucinogen intoxication delirium		Acute confusional state	Drug-induced delirium	Hallucinogen acute intoxication
Hallucinogen-induced mood disorder	Hallucinogen mood disorder	Nondependent abuse of hallucinogens	Drug-induced organic affective syndrome	
Hallucinogen-induced anxiety disorder		Nondependent abuse of hallucinogens	Hallucinogen abuse	
Hallucinogen-induced psychotic disorder	Hallucinogen delusional disorder	Hallucinogen paranoid and/or hallucinatory state	Drug-induced organic delusional syndrome	Hallucinogen psychotic disorder
Hallucinogen persisting perceptual disorder	Posthallucinogen perception disorder			Hallucinogen residual psychotic disorder (flashbacks)
Hallucinogen-related disorder NOS	Hallucinogen organic mental disorder NOS		Unspecified drug-induced mental disorder	Other or unspecified mental or behavioral disorder induced by hallucinogen

Table 11.7-3 Hallucinogen Diagnoses, DSM and ICD

Initial studies compared the above hallucinogen-induced psychotic disorder to schizophrenia, but clinical and single photon emission tomography (SPECT) studies now clearly differentiate hallucinogen-induced psychotic states from schizophrenic states. The type of hallucinogenic experience is strongly influenced by the personality of the user, the instructional set, the setting in which the drug is used, and the chemical structure of the drug. More severe reactions occur when the patient is given the drug surreptitiously.

ETIOLOGY

The rising trend of hallucinogen use among high school students in the United States since the late 1980s attests to the interplay of several factors in their use. Longitudinal studies show that hallucinogenic drugs like LSD occupy a midrange position in the progression of many drug use careers, in that many adolescents experiment first with legal drugs (alcohol and tobacco), fewer progress to misdemeanor drugs (marijuana and hallucinogens), and fewer still to felony drugs (cocaine and heroin). Accessibility is a key factor in drug use, and is high for LSD. A student can purchase a single dose of LSD for a few dollars, a fraction of the cost of a first-run Hollywood movie. The ease with which the synthetic hallucinogens can be manufactured in illegal laboratories keeps the price down. Facilitating the distribution of hallucinogens is their efficacy in microgram quantities, permitting them to be smuggled at will into any quarter of society, including schools, hospitals, and prisons, and almost as a routine, into rock concerts. The drugs are also attractive because of their prolonged effects and the absence of withdrawal symptoms and other stigmata of addiction.

PSYCHIATRIC DISORDERS

With the excitement that accompanied the discovery of LSD in 1943 came a spate of experiments attempting to glean therapeutic benefits from this new agent. A review of all human studies published from 1960 to 1994 listed in *Index Medicus* revealed that favorable reports of LSD's effects in humans predominated from 1960 to 1965. But beginning in 1968, the Summer of Love in the United States that arguably was the modal peak in the LSD epidemic of the 1960s, adverse reports began to outnumber positive ones, initiating a legal, psychiatric, and social reaction against the use of this class of drugs; this reaction has persisted into the present (Fig. 11.7-2). The figure describes a curve that can be applied not only to the trends found in the literature of LSD, but to any number of enthusiasms that mark the history of ideas, each characterized by initial excitement followed by sober reassessment. DSM-IV lists ten disorders associated with hallucinogenic drugs. (Table 11.7-2). Of these, the commonest adverse reactions to hallucinogens are intoxication, hallucinogen-induced anxiety disorder, hallucinogen persisting perception disorder, and hallucinogen-induced psychotic disorders. Clinical reports also support the diagnosis of hallucinogen-induced mood disorder.

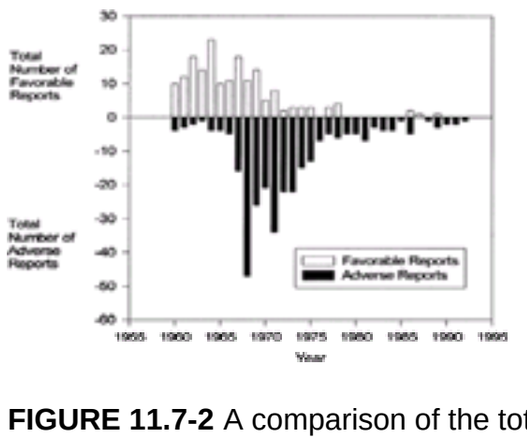

FIGURE 11.7-2 A comparison of the total number of clinically favorable and adverse reports on LSD appearing in the scientific literature between 1960 and 1995.

Hallucinogen Abuse and Dependence

Diagnosis and Clinical Features The diagnosis of an abuse of hallucinogens is determined by DSM-IV criteria used for other forms of substance abuse, namely, the recurrent use of an hallucinogen in a maladaptive pattern associated with physical risks or adverse impact on psychosocial function. Dependence is diagnosed when three or more of the following are present: tolerance, increasing drug consumption, unsuccessful efforts to cut down, drug-seeking behaviors, reduction in psychosocial function, and continued use despite knowledge of adverse consequences. The latter diagnosis is uncommon, given the time-limited course of

hallucinogen use in most patients.

Differential Diagnosis The patient presenting with a clinical picture of the sudden onset of inappropriate affect, visual hallucinations, and paranoid ideation suggests hallucinogen toxicity. Confusing this picture may be toxicity arising from other agents, including phencyclidine, anticholinergics, inhalants, and numerous other drugs. Laboratory tests screening for amphetamines, tetrahydrocannabinol, opiates, cocaine, benzodiazepines, and barbiturates are widely available. The presence of visual hallucinations or pseudohallucinations almost invariably points to a toxic, metabolic, vascular, epileptic, or neoplastic etiology in the central nervous system, rather than to schizophrenia. Hallucinations may also arise from within the eye as entoptic phenomena, but the patient with cataracts or retinal disease is likely to be older than the hallucinogen-abusing person.

Course, Prognosis, and Treatment The lifetime character of hallucinogen abuse is commonly described as a bell curve characterized by initial experimentation with the drug in adolescence, followed by a rising frequency of use, and finally a decline and end to use because of disinterest or the onset of chronic psychopathology. Of note is that few persons appear for treatment of hallucinogen abuse or dependence, but do present with the comorbid symptoms of anxiety, depression, psychosis, and suicidal ideation following the use of hallucinogens. Although clinical observation relates the onset of these disorders to the use of hallucinogens, the etiological role of hallucinogen abuse in mood and psychotic disorders has not been established by controlled studies. However, data suggest that substance use in general and hallucinogen use in particular is associated with adolescent suicide. The risk of suicide in the patient with psychosis following hallucinogen use is not inconsiderable. A clinical rule of thumb is that any patient suffering a chronic illness following hallucinogenic drugs, such as a perceptual, mood, or psychotic disorder, should be considered at higher than general risk for suicide for the duration of time that the disorder is present.

Hallucinogen Intoxication

Diagnosis and Clinical Features The characteristic feature of hallucinogen intoxication is the rapid onset of alterations in mood, cognition, and perception in the presence of a clear sensorium and following the ingestion of a drug in a commonly distributable form. Because memory is preserved, the details of the trip may be retained. This has led to a number of extraordinary personal descriptions of drug-altered states of consciousness, including those by William James, Aldous Huxley, and others. Psychological distress is more likely the hallmark of the hallucinogen user who presents for psychiatric help. Such cases may present with a variety of perceptual and conceptual phenomena leading to anxiety and panic. The DSM-IV diagnostic criteria for hallucinogen intoxication are listed in Table 11.7-4.

A. Recent use of a hallucinogen.
B. Clinically significant maladaptive behavioral or psychological changes (e.g., marked anxiety or depression, ideas of reference, fear of losing one's mind, paranoid ideation, impaired judgment, or impaired social or occupational functioning) that developed during, or shortly after, hallucinogen use.
C. Perceptual changes occurring in a state of full wakefulness and alertness (e.g., subjective intensification of perceptions, depersonalization, derealization, illusions, hallucinations, synesthesias) that developed during, or shortly after, hallucinogen use.
D. Two (or more) of the following signs, developing during, or shortly after, hallucinogen use:
(1) pupillary dilation
(2) tachycardia
(3) sweating
(4) palpitations
(5) blurring of vision
(6) tremors
(7) incoordination
E. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Table 11.7-4 DSM-IV Diagnostic Criteria for Hallucinogen Intoxication

Differential Diagnosis The patient presenting to an emergency service with a possible adverse reaction to hallucinogens should routinely be screened for other toxic agents capable of presenting with psychosis or delirium. These include the abuse of substituted amphetamines, cocaine, and PCP, as well as withdrawal from alcohol and benzodiazepines. Screens of urine can eventually rule out abuse of other drugs, which may present with psychosis, but depend on cooperation from the patient. Although drug detection technology can identify LSD in such cases, tools including gas chromatography and mass spectroscopy are not universally available to clinicians. The single best diagnostic tool remains a careful history and physical examination. Often, the history must be obtained from a less disturbed friend. A nettlesome issue in the differential diagnosis of hallucinogen intoxication is the possibility of PCP intoxication. Discriminating LSD from PCP effects is important because the treatment for one disorder may exacerbate the other disorder. One useful technique is the palm test. The examiner holds up his or her own hand and asks the patient to name all the colors visible in the palm. If the patient is using LSD, the response will be a stream of improbable colors and occasional images; if the user has taken PCP, he or she will remain mute or attack the examiner's hand. Examiner agility is helpful in the latter case.

Other features in the differential diagnosis include historical evidence of ingestion of an hallucinogen in an epidemiologically common form, the time of onset of toxic symptoms, and consensual validation of the drug history from companions. A physical examination will reveal marked mydriasis and other signs of autonomic arousal consistent with the acute drug state, in contrast to the miosis observed in PCP intoxication. Cases may be complicated by adulteration of the hallucinogen, mistaken attribution, or supplementation of the hallucinogen with marijuana, alcohol, cocaine, or other drugs.

Course and Prognosis The natural history of the “bad trip” is benign, tending to end, depending on the nature of the hallucinogen used, with a resolution of psychotic symptoms 6 to 12 hours following the time of ingestion.

Treatment Persons have historically been treated for hallucinogen intoxication by psychological support for the remainder of the trip. This is a time-consuming and potentially hazardous undertaking given the lability of a patient with hallucinogen-related delusions. Accordingly, treatment of hallucinogen intoxication is the oral administration of 20 mg of diazepam (Valium). This medication brings the LSD experience and any associated panic to a halt within 20 minutes, and is to be considered superior to “talking down” the patient over a period of hours, or to administering antipsychotic agents. The marketing of lower doses of LSD and a more sophisticated approach to treatment of casualties by drug users themselves have combined to reduce the appearance of this once-common disorder in psychiatric treatment facilities.

Hallucinogen-Induced Mood Disorder DSM-IV provides a diagnostic category for hallucinogen-induced mood disorder (see Table 14.6-18). Unlike cocaine-induced mood disorder and amphetamine-induced mood disorder, in which the symptoms are somewhat predictable, mood disorder symptoms accompanying hallucinogen abuse can be variable. Abusers may experience manic-like symptoms with grandiose delusions or depression-like feelings and ideas or mixed symptoms. As with the hallucinogen-induced psychotic disorder symptoms, the symptoms of hallucinogen-induced mood disorder usually resolve once the drug has been eliminated from the person's body.

However, depression, and suicide have been reported following the use of MDMA. Weekend use of MDMA has been shown to be followed by depression midweek, similar to the crashing syndrome of amphetamines and cocaine. The suggestion that hallucinogens may induce a chronic mood disorder is controversial, because patients may use mood-altering drugs to self-medicate a preexisting depressive disorder.

Hallucinogen-Induced Anxiety Disorder One feature of this class of drugs that invites use is the unmasking and amplification of pleasurable affect. Feeling states appear unbidden, but are seldom under the user's control. When the affects are frightening, they may cause the user to fly into a drug-induced panic. Such patients are likely to be found terrified on the fringes of the rock concert or in the waiting rooms of psychiatric emergency centers.

> A 20-year-old man had a 7-year history of polysubstance abuse, including having used LSD an estimated 400 times. While driving with his girlfriend he ingested an unknown quantity of LSD and became intoxicated; he reported using no other drugs at this time. Within minutes after ingestion, he began to experience visual hallucinations that intensified as he drove. When he attempted to speak to his girlfriend, he saw that she had become a giant lizard. He became terrified and attempted to kill her by crashing the car, injuring himself and his passenger. By the time of discharge from the hospital 3 days later, his panic had resolved.

Treatment is the use of oral benzodiazepines as described for hallucinogen intoxication.

Hallucinogen Persisting Perception Disorder Prolonged visual disturbances following LSD have been described for over 40 years. They may occur contiguously

with the trip or spontaneously recur days to weeks following cessation of drug use. Visual symptoms that occur sporadically after drug use and last for a second or two are usually designated “flashbacks.” Those that linger continuously, with minor fluctuations in intensity, are more properly considered to be hallucinogen persisting perception disorder (Table 11.7-5).

A. The reexperiencing, following cessation of use of a hallucinogen, of one or more of the perceptual symptoms that were experienced while intoxicated with the hallucinogen (e.g., geometric hallucinations, false perceptions of movement in the peripheral visual fields, flashes of color, intensified colors, trails of images of moving objects, positive afterimages, halos around objects, macropsia, and micropsia).

B. The symptoms in criterion A cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The symptoms are not due to a general medical condition (e.g., anatomical lesions and infections of the brain, visual epilepsies) and are not better accounted for by another mental disorder (e.g., delirium, dementia, schizophrenia) or hypnopompic hallucinations.

Table 11.7-5 DSM-IV Diagnostic Criteria for Hallucinogen Persisting Perception Disorder (Flashbacks)

Such recurrences may involve somatic and emotional symptoms, but perceptual ones predominate. These may include geometric hallucinations, false fleeting perceptions of movement in the peripheral fields, flashes of color, and afterimagery. Common complaints are the persistence of trailing images as an object moves through the patient's visual field. Brighter objects such as auto tail lights induce trails more easily, although these may be induced simply by moving the examiner's hand across the patient's visual field, a maneuver with high sensitivity and specificity for hallucinogen persisting perception disorder. The entire visual field in many cases may be described as reticulated, grainy, or filled with numerous vibrating pinpoint-sized dots, to the extent that patients complain that they can see the air. This phenomenon of “aeropsia” also appears to be highly specific for hallucinogen persisting perception disorder. Comorbidly occurring secondary conditions include depressive and anxiety disorders; alcohol dependence may occur secondarily. Less frequently, psychosis may be present although the vast majority of patients with flashbacks have preserved reality testing, and understand that their new perceptual burden is “not real” and is more properly considered a form of pseudohallucinations.

A 16-year-old girl presented with a complaint of “never being happy.” She had used LSD for the first time at the age of 13, and used the drug a total of 30 times until 3 months prior to her consultation. Occasional LSD trips had been notable for panic attacks, but she denied spontaneous panic. At the age of 14 she noted the onset of depression, which coincided insidiously with the onset of continuous visual symptoms. These included images of faces on the floor and walls (*pareidolias*), dots too numerous to count hovering in the air, flashes of color and white light, trails of moving objects, positive afterimages, boxes on walls, and occasionally, a face or its outline floating in space. At 15 she attempted suicide; major depression was diagnosed. A sleep electroencephalography (EEG) with sphenoidal leads was unremarkable although a year later a quantitative EEG showed abnormal sensory evoked potentials.

A 19-year-old college student presented with a complaint of “feeling high all of the time.” His past history was positive for daily marijuana use from ages 15 to 17. In that period he also used LSD 13 times. He was asymptomatic for 3 years until he experienced an attack of anxiety, visual disturbances, and hyperacusis. The initial spell lasted several hours, and was characterized by the experience of continuous visual trails and auditory reverberations in passing stimuli. He felt distant from his body, and had a sense of impending doom. The anxiety attack subsided, but the visual disturbances continued without abatement. Among them was a complaint of seeing “millions” of clear dots in the air, “like I can see the molecules.” A computerized tomogram of the cerebrum and an EEG were negative. A brief trial of thioridazine (Mellaril) made the visual symptoms worse, but chlordiazepoxide (Librium) was effective in reducing but not obliterating his visual symptoms. The visual symptom of “seeing” the air was intensified on gazing at the sky, but seemed to remit spontaneously in the following 2 years. Visual and auditory trailing persisted, however, as have occasional episodes of panic attacks.

A 26-year-old male was referred for an evaluation of possible LSD-related disorder. His past medical history was negative for preexisting disorders of the central nervous system. His family history was negative for psychiatric illness. The patient's first use of LSD occurred at age 14 with an alleged dose of 100 micrograms. In the next year he used the drug on approximately 30 occasions, and used marijuana frequently as well. Drug use coincided with family conflict, truancy, and running away from home. At the age of 15 on the day following a dysphoric LSD experience, he awoke with a sense that “something snapped” on the left side of his head. He suffered progressive dysphoria, depression, and withdrawal. Peering into a friend's eyes generated LSD-like feelings and panic. He hallucinated yellow geometric forms and the trails of passing objects. He became irrationally afraid of people, was unable to attend school, and 2 months later was hospitalized in a psychotic state. He was treated with chlorpromazine (Thorazine) with partial relief from paranoid delusions. Over the next 9 years his major symptom was depression in a context of constant visual pseudohallucinations. He saw trails of objects and auras around people. When listening to music he generated internal geometric imagery. He was able to visualize “the air,” which he described as yellow dots hovering in space. These images persisted daily for 12 years.

Course and Prognosis Roughly half the patients with hallucinogen persisting perception disorder recover completely in 5 years, but others may irreversibly continue to have symptoms. The appearance of this disorder is not dose dependent; it may arise from a single LSD ingestion. The disorder is exacerbated by autonomic arousal, including psychostimulants such as amphetamines, cocaine, cold tablets containing pseudoephedrine, and caffeine; also by marijuana use, including passive inhalation; excessive fatigue, overexercise, and intercurrent infections. Symptoms are commonly precipitated by entering a dark environment. The disorder is associated with vivid visual hallucinations during alcohol withdrawal. Psychophysical evidence is consistent with the hypothesis that this disorder is associated with disinhibition of visual information processing; that is, once stimulated, the visual system is slow to return to a resting baseline. Quantitative electroencephalography also reveals evidence of visual disinhibition in these patients. Auditory responses are delayed (Fig. 11.7-3). It is noteworthy that these findings closely parallel those EEG findings described during the acute administration of LSD to a variety of species, including humans. Thus, the patient's claim that he is “still tripping” is borne out neurophysiologically.

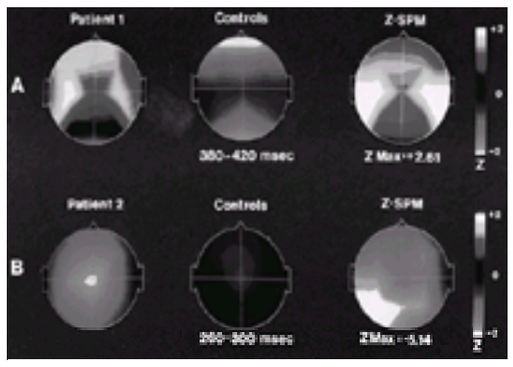

FIGURE 11.7-3 Illustrative cases of quantitative EEG abnormalities in hallucinogen persisting perreotion disorder (flashback). **Patient 1:** A 26-year-old computer programmer used LSD at the age of 18 on 15 occasions; 31 months later the patient experienced the abrupt onset of intense, LSD-like set of visual and affective disturbances lasting all night. At 25 he suffered the spontaneous onset of hourly flashing white lights centrally and black dots in his peripheral fields, which have continued for the past 10 years. Topographic brain maps are shown during the 380-420 msec epoch of the visually evoked potential in row **A**. The upper left map represents the subject's data. The upper middle map shows control subjects for the same poststimulus latency epoch. The right upper image is a significance probability map (SPM) showing *Z*-scores resulting from a comparison of the data from the left and middle upper maps. The patient shows an enhancement in the visually evoked signal involving both temporal regions of the cerebrum. **Patient 2:** A 23-year-old musician used LSD on 16 occasions over a 4-month period at the age of 20. Within 2 months he began to notice a progressive, continuous visual disorder characterized by flashes of color, persisting afterimages, haloes around objects, a grainy texture to the sky, and the lingering trails of objects as they passed through his visual field. The graininess in the visual field interfered with night vision. Topographic brain mapping is illustrated during the 260-300 msec epoch of auditory evoked potentials in row **B**. Note the region of reduced electrical activity in

the left posterior temporal region in the lower right map. (Reprinted with permission from Abraham HD, Duffy FH: Stable quantitative EEG difference in post-LSD visual disorder by split-half analysis: Evidence for disinhibition. Psychiatry Res *67:*173, 1996.) (See Color Plate 6.)

Treatment Treatment for hallucinogen persisting perception disorder is palliative. The first step in the process is correct identification of the disorder; it is not uncommon for the patient to consult a number of specialists before the diagnosis is made. Pharmacological approaches include long-lasting benzodiazepines such as clonazepam (Klonopin), and to a lesser extent anticonvulsants including valproic acid (Depakene) and carbamazepine (Tegretol). No drug is yet completely effective in ablating symptoms. Antipsychotic agents should only be employed in the treatment of hallucinogen-induced psychoses. It is not uncommon for these agents to exacerbate the disorder within the first 72 hours of administration. Case reports suggest sertralin and naltremene may be beneficial. Serotonin type 2 antagonists such as risperidone (Risperdal) have also been shown to exacerbate the intensity of visual symptoms. A second dimension of treatment is behavioral. The patient must be instructed to avoid gratuitous stimulation in the form of over-the-counter drugs, caffeine, alcohol, and avoidable physical and emotional stressors. Marijuana smoke is a particularly strong intensifier of the disorder, even when passively inhaled. Finally, three comorbid conditions are associated with hallucinogen persisting perception disorder—panic disorder, major depression, and alcohol dependence. All these conditions require primary prevention and early intervention.

Hallucinogen-Induced Psychotic Disorders

Diagnosis and Clinical Features If psychotic symptoms are present in the absence of retained reality testing, a diagnosis of hallucinogen-induced psychotic disorder may be warranted (see Table 13.3-4). DSM-IV also allows clinicians to specify whether hallucinations or delusions are the prominent symptoms.

Because the ingestion of an hallucinogen induces a toxic mental state analogous to a psychosis, it is useful operationally to define a posthallucinogen psychotic disorder as one that continues for more than 48 hours following ingestion. Attack rates of such disorders among those using hallucinogens are reported to occur in a range of 0.08 to 4.6 percent, with a trend of higher rates occurring in psychiatric patients and lower ones in healthy volunteers. Patients may suffer the onset of a psychosis immediately on ingestion of the drug, or may have a lucid interval of days to months before the onset. In the former case the hallucinogen is linked more convincingly as the cause of the psychosis than in the latter situation. The patient presenting with a prolonged psychosis following the verified ingestion of LSD or other hallucinogen should raise a strong suspicion of a substance induced disorder.

Differential Diagnosis Prior to diagnosis the clinician must consider preexisting psychotic disorders like schizophrenia; mistaken attribution in samples of LSD that are in fact amphetamines, phencyclidine, or other drugs with psychotic potential; or sources of delirium such as alcohol withdrawal. Posthallucinogen psychotic symptoms include mood swings with euphoria and grandiosity, multimodal hallucinations, and hyperreligiosity. When compared to non–drug–using schizophrenia patients, patients with posthallucinogen psychosis have earlier ages of onset, more visual hallucinations, more depression, and a higher prevalence of families with mood disorder. Clinically they are more likely to suffer the positive symptoms of psychosis in the context of schizoaffective disorder. Finally, individuals claiming to have developed a lifelong psychosis following a single dose of LSD without being able to describe the symptoms of a typical trip are more likely to be suffering from a paranoid rather than a postdrug disorder.

> A 19-year-old musician used LSD on three occasions. Following the third trip, he failed to regain his pre-drug mental state. Instead, he experienced auditory and visual hallucinations, rapid and incoherent speech, sleeplessness, zoophilic delusions, and suicidal ideation. He acted out his psychotic fears aggressively, showering with his clothes on, ripping towel racks from the wall, and requiring four-point restraints. This disorder continued for 2 months during which time he was unsuccessfully treated with lithium carbonate (Eskalith), lorazepam (Ativan), propranolol (Inderal), haloperidol (Haldol), benztropine (Cogentin), and diphenhydramine (Benadryl). His symptoms slowly abated with a combination of lithium and perphenazine (Trilafon).
>
> A 22-year-old female photography student presented to the hospital with inappropriate mood and bizarre thinking. She had no prior psychiatric history. Nine days prior to admission she ingested one or two psilocybin mushrooms. Following the immediate ingestion, the patient began to giggle. She then described euphoria, which progressed to auditory hallucinations and belief in the ability to broadcast her thoughts on the media. Two days later she repeated the ingestion, and continued to exhibit psychotic symptoms to the day of admission. When examined she heard voices telling her she could be president, and reported the sounds of “lambs crying.” She continued to giggle inappropriately, bizarrely turning her head from side to side ritualistically. She continued to describe euphoria, but with an intermittent sense of hopelessness in a context of thought blocking. Her self-description was “feeling lucky.” Haloperidol was begun, 10 mg twice a day, along with benztropine 1 mg three times a day and lithium carbonate 300 mg twice a day. On this regimen her psychosis abated after 5 days.

Another variety of posthallucinogen psychosis presents with catatonia, confusion, and multimodal hallucinations, supporting the concept of hallucinogenic delirium.

> A 23-year-old undergraduate was premorbidly a sociable, outgoing individual with no prior history of drug use. Over a 3-day period he used LSD on three occasions, and continued to act in a bizarre manner. He described auditory and visual hallucinations, including images of his estranged father on the ceiling, which fixated his gaze upwardly for long periods. He was disoriented and his gait was reduced to 3-inch steps. Speech was reduced to rare one- and two-word sentences. He could not follow simple commands. Generalized apraxia was present. Pauses of 30 to 60 seconds were noted between an examiner's question and his verbal response. A screen of his urine was negative for toxins. An awake EEG showed generalized slowing in the theta range. The catatonia, delusions, and hallucinations were partially reduced with antipsychotic medications, fluoxetine (Prozac), and electroconvulsive therapy, but the patient continued to exhibit psychotic symptoms chronically.

This uncommon case shares features of hallucinogen-induced psychotic disorder, hallucinogen intoxication delirium, and schizophrenia. The continuity of a drug state with an enduring psychotic disorder suggests the first, though whether this class of drugs activates a pre-existing schizophrenogenic trigger remains to be shown.

Treatment Treatment of hallucinogen-induced psychosis does not differ from conventional treatment for other psychoses. However, in addition to antipsychotic medications, a number of agents have been reported to be effective including lithium carbonate, carbamazepine, and electroconvulsive therapy. Antidepressant drugs, benzodiazepines, and anticonvulsant agents may each have a role in treatment as well. One hallmark of this disorder is that, as opposed to schizophrenia, in which negative symptoms and poor interpersonal relatedness may commonly be found, patients with hallucinogen-induced psychosis exhibit the positive symptoms of hallucinations and delusions while retaining the ability to relate to the psychiatrist. Medical therapies are best applied in a context of supportive, educational, and family therapies. The goals of treatment are the control of symptoms, a minimal use of hospitals, daily work, the development and preservation of social relationships, and the management of comorbid illnesses including alcohol dependence, depression, and suicide. The preservation of insight in these patients permits a painful comparison of their lives to those of their healthy contemporaries as the latter reach milestones of the life cycle while the former do not. Optimizing psychological and social development of the patient with psychosis is the central strategy for treatment.

Hallucinogen Intoxication Delirium DSM-IV allows for the diagnosis of hallucinogen intoxication delirium (see Table 10-22), a relatively rare disorder beginning during intoxication in those who have ingested pure hallucinogens. Hallucinogens are often mixed with other substances, however, and the other components or their interactions with the hallucinogens can produce a clinical delirium.

Hallucinogen-Related Disorder Not Otherwise Specified Chronic users of hallucinogenic drugs may present a number of symptoms and behaviors that do not easily conform to the preceding classification. Cerebral edema, has been reported following the use of MDMA. Not uncommonly, patients who have used LSD in the past may describe a constellation of symptoms suggestive of a disorder of the temporal lobes, including hyperreligiosity, sexual dysfunction, and occasionally hallucinations (Table 11.7-6).

The hallucinogen-related disorder not otherwise specified category is for disorders associated with the use of hallucinogens that are not classifiable as hallucinogen dependence, hallucinogen abuse, hallucinogen intoxication, hallucinogen persisting perception disorder, hallucinogen intoxication delirium, hallucinogen-induced psychotic disorder, hallucinogen-induced mood disorder, or hallucinogen-induced anxiety disorder.

Table 11.7-6 DSM-IV Diagnostic Criteria for Hallucinogen-Related Disorder Not Otherwise Specified

This syndrome must be differentiated from schizophrenia because the global, social, and occupational dysfunction of schizophrenia is absent in this patient. Not uncommonly hyperreligiosity may be coupled with organizational abilities so that the patient may present as a minister of a church of his own creation, a benign enough species of psychological diversity. On rare occasions, however, persons such as Charles Manson, combining drug use with quasireligious dogma, may form cults with lethal outcomes.

SUGGESTED CROSS-REFERENCES

A general discussion of substance use disorders (intoxication, withdrawal, abuse, dependence, and persisting disorders) is found in Section 11.1. Section 12.7 on schizophrenia highlights differences in the clinical presentation and course between schizophrenia and hallucinogen abuse and hallucinogen, dependence, and Section 11.10 on phencyclidine (and related substances) clarifies differences from disorders involving those compounds.

SECTION REFERENCES

*Abraham H: Visual phenomenology of the LSD flashback. Arch Gen Psychiatry *40:*884, 1983.

*Abraham HD, Fava M: Order of onset of substance abuse and depression in a sample of depressed outpatients. Comprehensive Psychiatry *40:*44, 1999.

Abraham H, Wolf E: Visual function in past users of LSD: Psychophysical findings. J Abnorm Psychol *97:*443, 1988.

*Abraham HD, Aldridge A, Gogia P: Psychopharmacology of the hallucinogens. Neuropsychopharmacology *14:*285, 1996.

Abraham HD, Duffy FH: Stable qEEG differences in post-LSD visual disorder by split half analyses: Evidence for disinhibition. Psychiatry Res Neuroimag *67:*173, 1996.

Abramson H, Jarvik M, Kaufman M, Kronetsky C, Levine A, Wagner M: Lysergic acid diethylamide (LSD-25): Physiological and perceptual responses. J Physiol *39:*3, 1955.

Bowers M: Acute psychosis induced by psychotomimetic drug abuse, I: Clinical findings. Arch Gen Psychiatry *27:*437, 1972a.

Bowers M: Acute psychosis induced by psychotomimetic drug abuse, II: Neurochemical findings. Arch Gen Psychiatry *27:*440, 1972b.

*Bowers M: Psychoses precipitated by psychotomimetic drugs. Arch Gen Psychiatry *34:*832, 1977.

Brawley P, Duffield J: The pharmacology of hallucinogens. Pharm Rev *24:*31, 1972.

Breakey W, Goodell H, Lorenz P, McHugh P: Halluconigenic drugs as precipitants of schizophrenia. Psychol Med *4:*225, 1974.

Cohen S, Ditman K: Prolonged adverse reactions to lysergic acid diethylamide. Arch Gen Psychiatry *8:*475, 1963.

Dewhurst H: Differential diagnosis and treatment of lysergic acid diethylamide induced psychosis. The Practitioner *209:*327, 1972.

Elkes C, Elkes J, Mayer-Gross W: Hallucinogenic drugs. Lancet *268:*719, 1955.

Favazza A, Domino E: Recurrent LSD experience (flashbacks) triggered by marijuana. University of Michigan Medical Center J *35:*214, 1969.

Fink M, Simeon J, Haque W, Itil T: Prolonged adverse reactions to LSD in psychotic subjects. Arch Gen Psychiatry *15:*450, 1966.

Freedman D: On the use and abuse of LSD. Arch Gen Psychiatry *18:*330, 1968.

Glennon R, Teitler M, McKenney J: Evidence of 5-HT2 involvement in the mechanism of hallucinogenic agents. Life Sci *35:*2505, 1984.

Halpern JH, Pope HG Jr: Do hallucinogens cause residual neuropsychological toxicity? Drug Alcohol Depend *53:*247, 1999.

Hatrick J, Dewhurst K: Delayed psychosis due to LSD. Lancet *2:*742, 1970.

Hermle L, Spitzer M, Gouzoulis E: Arylalkamine-induced effects in normal volunteers: On the significance of research in hallucinogenic agents for psychiatry. In *Fifty Years of LSD: Current Status and Perspectives of Hallucinogens*, A Pletscher, D Ladewig, editors. Parthenon Publishing Group, Parthenon, NY, 1994.

*Hofmann A: *LSD: My Problem Child*. McGraw Hill, New York, 1980.

Hollister L: Drug-induced psychoses and schizophrenic reactions: A critical comparison. Ann NY Sci *96:*80, 1962.

Horowitz M: Flashbacks: Recurrent intrusive images after LSD. Am J Psychiatry *126:*556, 1969.

Isbell H: Tolerance to LSD. Fed Proc *14:*354, 1955.

Klüver H: *Mescal and the Mechanisms of Hallucinations*. University of Chicago Press, Chicago, 1966.

Krystal JH, Price LH, Opsahl C, Ricaurte GA, Heninger GR: Chronic 3,4-methylenedioxymethamphetamine (MDMA) use: Effects on mood and neuropsychological function? Am J Drug Alcohol Abuse *18:*331, 1992.

Lake R, Stirba A, Kinneman R, Carlson B, Holloway H: Mania associated with LSD ingestion. Am J Psychiatry *138:*1508, 1981.

Lang R, Barr H: Lysergic acid diethylamide (LSD-25) and schizophrenic reactions. J Nerv Ment Dis *147:*163, 1968.

Lerner AG, Oyefe I, Isaacs G, Sigal M: Naltrexone treatment of hallucinogen persisting perception disorder. Am J Psychiatry *154:*437, 1997.

Lewin L: *Phantastica. Die Betaubenden und Erregenden Genussmittel*. Verlag von Georg Stike, Berlin, 1924.

Muller D: ECT in LSD psychosis: A report of three cases. Am J Psychiatry *128:*351, 1971.

Ricaurte GA, LS Forno, Wilson MA, De Lanney LE, Irwin I, Molliver M, Langston JW: (+/–)3,4-Methylenedioxymethamphetamine selectively damages central serotonergic neurons in nonhuman primates. JAMA *260*(1):51, 1988.

Sankar D: *LSD, A Total Study*. PJD Publications, Westbury, NY, 1975.

Schultes R, Hofmann A: *The Botany and Chemistry of Hallucinogens*. Charles C. Thomas, Springfield, IL, 1980.

Shulgin A: *Pihkal, A Chemical Love Story*. Transform Press, Berkeley, CA, 1991.

*Siegel R: The natural history of hallucinogens. In *Hallucinogens: Neurochemical, Behavioral, and Clinical Perspectives*, B Jacobs, editor. Raven, New York, 1984.

Strassman R, Qualls C, Uhlenhuth E, Kellner R: Dose-response study of N,N-dimethyltryptamine in humans. II. Subjective effects and preliminary results of a new rating scale. Arch Gen Psychiatry *51:*98, 1994.

Teitler M, Leonhardt S, Appel NM, Defouza EB, Glennon RA: Receptor pharmacology of MDMA and related hallucinogens. Ann NY Acad Sci *600:*626, 1990.

Young CR: Sertraline treatment of hallucinogen persisting perception disorder. J Clin Psychiatry *58:*85, 1997.

11.8 INHALANT-RELATED DISORDERS

THOMAS J. CROWLEY, M.D.

Inhalant drugs (sometimes called “volatile substances”) are widely available and frequently misused, especially by adolescents. About 20 percent of American eighth-grade students report that they have used these substances for psychoactive effects, more than the number who have tried marijuana. Most adolescents who try inhalants apparently discontinue them after one or a few times. However, for a smaller group of adolescents, especially those with comorbid conduct disorder, inhalant use may foreshadow many years of polysubstance abuse or dependence, including drug injections. Most of these persons eventually shift to other drugs, although some continue daily use of inhalants themselves for many years, suffering major behavioral and organ pathology from the drugs' chronic toxicity. A still smaller number of adolescents die from acute inhalant toxicity, often during their first use of the drugs. Indeed, in Great Britain where detailed records are available, inhalants have become a leading cause of adolescent death because, although such deaths are infrequent, adolescent deaths generally are uncommon.

Inhalants are volatile hydrocarbons, such as toluene, *n*-hexane, methylbutyl ketone, trichloroethylene, trichloroethane, dichloromethane, gasoline, and butane. They make up four commercial classes: (1) solvents for glues and adhesives; (2) propellants for aerosol paint sprays, hair sprays, frying pan sprays, and shaving cream; (3) thinners (e.g., for paint products and typing correction fluids); and (4) fuels.

DEFINITION

The section on inhalant-related disorders in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) “includes disorders induced by inhaling the aliphatic and aromatic hydrocarbons... Less commonly used are halogenated hydrocarbons... and other volatile compounds containing esters, ketones, and glycols.” DSM-IV provides two broad categories of inhalant-related disorders (Table 11.8-1). The first category is inhalant use disorders (inhalant abuse and inhalant dependence), which are characterized by maladaptive patterns of inhalant use (e.g., frequency, dose, danger). The second category, inhalant-induced disorders (such as inhalant intoxication), result from the toxic effects of inhaled substances. Because of epidemiological and pharmacological differences, DMS-IV excludes from the inhalant-related disorders conditions related either to anesthetic gases or to amyl and butyl nitrites, classifying these as other (or unknown) substance-related disorders; some disorders associated with those compounds are briefly discussed at the end of this chapter. Although fumes of such combustible drugs as crack cocaine and tobacco also are inhaled, DSM-IV similarly places disorders related to those drugs in separate categories.

Inhalant use disorders
Inhalant dependence
Inhalant abuse
Inhalant-induced disorders
Inhalant intoxication
Inhalant intoxication delirium
Inhalant-induced persisting dementia
Inhalant-induced psychotic disorder, with delusions
Specify if:
With onset during intoxication
Inhalant-induced psychotic disorder, with hallucinations
Specify if:
With onset during intoxication
Inhalant-induced mood disorder
Specify if:
With onset during intoxication
Inhalant-induced anxiety disorder
Specify if:
With onset during intoxication
Inhalant-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.8-1 DSM-IV Inhalant-Related Disorders

COMPARATIVE NOSOLOGY

The section on inhalant-related disorders in DSM-IV lists three major categories: inhalant abuse, inhalant dependence, and inhalant intoxication. The other inhalant-related disorders have their diagnostic criteria specified in the DSM-IV sections that specifically address the major symptoms. For example, inhalant-induced psychotic disorder is included with other psychotic disorders.

The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) refers to inhalants as “volatile substances.” ICD-10 does not use the term “abuse,” which is in DSM-IV, offering instead the term “harmful use.” DSM-IV provides no diagnosis of inhalant withdrawal (which clinicians describe, but which probably is rare), whereas ICD-10 includes that diagnosis, but gives no diagnostic criteria.

EPIDEMIOLOGY

The 1995 United States National Household Survey estimated that 5.7 percent of Americans had used an inhalant at some time in their lives. The prevalence was 8.0 percent among males and 3.5 percent among females. The figures were highest for males 18 to 34 years of age; almost 15 percent of them reported using an inhalant at some time. The overall prevalence rates for use of inhalants at any time in one's life were 6.7 percent among whites, 3.5 percent among Hispanics, and 1.8 percent among blacks.

A large annual survey has examined trends in drug use among American high school students. It shows that many adolescents have used inhalants. Among eighth-grade students inhalants are the most commonly used drug (except for alcohol and tobacco); about 20 percent of eighth-graders report having used an inhalant, while only about 17 percent report having used marijuana. Native-American students on reservations may be especially vulnerable to inhalant use; one large survey found that 34 percent of reservation-dwelling Native-American eighth-grade students had used an inhalant.

In 1976 about 10 percent of interviewed high school seniors reported some use of inhalants. The figure peaked at 18 percent in 1990, and remained at about 17 percent in 1996. The number of students using inhalants more frequently is reflected in the proportion of high school seniors who report using an inhalant in the last 30 days. That figure gradually rose from about 1 percent of students in 1976 to 2.5 percent in 1996.

In relation to other abused drugs inhalants are not a major cause of morbidity or mortality in the United States. Inhalants accounted for only 0.3 to 0.4 percent of drug-related episodes in federally monitored emergency rooms between 1991 and 1994. Similarly, inhalants contributed to only about 1.4 percent of federally

reviewed medical-examiner death reports for 1994.

For inhalant males and whites (in comparison to females and blacks) were overrepresented in both medical-examiner reports and emergency-room visits. However, age distributions differed between patients in emergency rooms and those deceased. In 1994 the peak prevalence for inhalant-related emergency-room episodes was among patients 12 to 17 years of age (40 percent of all inhalant-related episodes), whereas the peak prevalence of deaths occurred in the age range 35 to 44 years (35 percent of inhalant-related death reports).

Medical examiners considered over 40 percent of inhalant-related deaths to be suicides. Similarly, a suicide attempt was part of the presentation in 38 percent of inhalant-related emergency visits. Thus, it appears that inhalant intoxication often is associated with suicidality.

In Great Britain inhalants are among the leading causes of death in adolescents and are the single most common cause in those 15 years old. First-time users there are the most likely to die, perhaps because they are inexperienced at this dangerous pastime. Comparable data are not available for the United States. In several developing countries heavy use of inhalants is common in groups of street kids, who live with no ties to adults.

American data show widespread experimentation with inhalants, relatively few current users, and still fewer inhalant emergencies or deaths, suggesting that most users try the drugs a few times and stop without mishap. Most of these users probably will not develop further drug problems, but nevertheless the odds of such problems are much greater among them. For example, inhalant use is associated with increased odds of later injection-drug use, a risk factor for infection with the human immunodeficiency virus (HIV). In a general population survey, persons reporting any use of inhalants were 45 times more likely, and those who had used both inhalants and cannabis were 89 times more likely, than others to have injected drugs. Similarly, a prospective study of inner-city youths found that those who had used inhalants in adolescence were nine times more likely to use heroin later.

The available data suggest that most adolescents who survive a brief experimentation with inhalants and who do not have conduct disorder soon abstain from inhalants and avoid pathological outcomes. However, among adolescents with conduct disorder, perhaps half will develop adult antisocial personality disorder. Inhalant problems in these youths often herald serious alcohol and polysubstance use in adulthood, and a few of these youths will become chronic, deteriorated, inhalant-dependent adults.

PHARMACOLOGY AND TOXICOLOGY

Data on inhalant choice in the United States are limited, but of some 20 abused compounds, toluene and gasoline may be the most popular. Various authorities recommend industrial exposure limits for toluene of 50 to 100 parts per million (ppm), although a recent study showed that a 6-hour exposure to 100 ppm produced a temporary neuropsychological performance decrement of about 10 percent. By comparison, inhaled concentrations from a glue-containing bag may reach 10,000 ppm, and vapors from several tubes of glue may be inhaled each day. About 15 to 20 breaths of 1 percent gasoline vapor produce several hours of intoxication. Sniffing vapor through the nose or huffing (taking deep breaths) through the mouth leads to transpulmonary absorption with very rapid drug access to the brain. Breathing through a solvent-soaked cloth, inhaling fumes from a glue-containing bag, huffing vapor sprayed into a plastic bag, or breathing vapor from a gasoline can are common.

Toluene concentrations in blood in hospitalized intoxicated persons reportedly range from 0.8 to 8 µg/g. Exposure at the industrial maximum of 100 ppm produces blood levels around 0.5 µg/g, although moderate exercise may triple these levels through increased respiration. Brain and fat achieve higher concentrations than blood because lipophilic compounds preferentially distribute there. The coadministration of alcohol dramatically raises toluene concentration in blood, increasing toxicity, probably through competition for hepatic metabolizing enzymes.

About 20 percent of a toluene dose is excreted unchanged in the breath, but most is metabolized in the liver to hippuric acid before urinary excretion. Breath concentrations of toluene fall by half within a few minutes after the end of a prolonged exposure. Blood concentrations fall more slowly, becoming undetectable 4 to 10 hours after exposure. Urinary hippuric acid remains measurable somewhat longer; hippurate-creatinine ratios above 1 gram per gram suggest toluene use, but benzoic acid food preservatives may generate false-positive hippuric acid concentrations.

The cellular mechanisms of inhalant action are unclear. Hypotheses include cell membrane fluidization or interactions at g-aminobutyric acid (GABA)-gated chloride channels, but data are very sparse. Behavioral actions in animals suggest that inhalants act like alcohol, barbiturates, and other depressants of the central nervous system (CNS). Like depressants they produce motor stimulation at lower doses and motor suppression at higher doses, as well as ataxia and loss of righting reflex. Inhalants also have anticonvulsant actions and show depressant-like effects in certain behavioral paradigms. Animals will work to self-administer inhalants, and animals trained to press one lever when injected with alcohol or pentobarbital and another when injected with saline will press the depressant-appropriate lever after exposure to toluene vapor, suggesting that the subjective experience after toluene or a depressant is similar. Moreover, alcohol and benzodiazepines potentiate inhalant effects.

Rodents develop withdrawal seizures after several days' exposure to trichloroethane, a frequently abused inhalant. The seizures are blocked by toluene, ethanol, pentobarbital, and midazolam (Versed), a benzodiazepine. Thus, inhalants can produce physical dependence, and they show cross-dependence with familiar CNS depressants.

In human subjects low-dose (0, 75, or 150 ppm) toluene exposures for several hours produce dose-related decrements in tests of perception, memory, and manual dexterity, with increased headaches, mucosal irritation, thirst, and sleepiness.

Organ Pathology Inhalants are associated with many potentially serious adverse effects. The most serious of these is death, which can result from respiratory depression, cardiac arrhythmias, asphyxiation, aspiration of vomitus, or accident or injury (e.g., driving while intoxicated with inhalants). Placing an inhalant-soaked rag and one's head into a plastic bag, a common procedure, may cause coma and suffocation.

Chronic inhalant users may have numerous neurological problems. Computed tomography and magnetic resonance imaging reveal diffuse cerebral, cerebellar, and brainstem atrophy with white-matter disease, a leukoencephalopathy. Several studies of house painters and factory workers who have been exposed to solvents for long periods also have found evidence of brain atrophy on computed tomography scans, with decreased cerebral blood flow. Neurological signs and symptoms may include hearing loss, peripheral neuritis, headache, paresthesias, cerebellar signs, persisting motor impairment, parkinsonism, and lead encephalopathy. The combination of organic solvents with high concentrations of copper, zinc, and heavy metals has been associated with the development of brain atrophy, temporal lobe epilepsy, decreased intelligence quotient (I.Q.), and a variety of electroencephalographic (EEG) changes.

Other serious adverse effects associated with long-term inhalant use include irreversible hepatic disease or renal damage (tubular acidosis) and permanent muscle damage associated with rhabdomyolysis. Additional adverse effects include cardiovascular and pulmonary symptoms (e.g., chest pain and bronchospasm) as well as gastrointestinal symptoms (e.g., pain, nausea, vomiting, and hematemesis). Several reports suggest that in utero exposure to toluene may produce an embryopathy similar to fetal alcohol syndrome. Fortunately, there is no convincing evidence that toluene, the best-studied inhalant, produces genetic damage in somatic cells.

ETIOLOGY

Multiple factors contribute to the etiology of inhalant-related disorders. First, availability is important in determining the prevalence of abuse or dependence on a drug. Inhalants are cheap, available in several forms in most households, easily concealed, legal to possess, and simple to take. Second, inhalant use apparently is rewarding, both through direct pharmacological action and through the drugs' social effects. As mentioned, under certain circumstances animals repeatedly self-administer inhalants, showing that these substances have innate reinforcing properties. In addition adolescents usually gather in small groups to use inhalants and being a user gains entry to the group, socially reinforcing the use. Third, inhalant users often can evade detection or punishment by parents or school authorities, since the drugs quickly produce a high that passes within a few hours. Fourth, inexpensive inhalants may be one of the few exciting and novel experiences available to youths in impoverished communities providing few other reinforcers. This may help to explain the high prevalence of inhalant use on some Indian reservations.

In addition to those extrinsic factors at least one intrinsic factor contributes to inhalant problems. A risk-taking propensity may lead some persons to the at-the-brink excitement and danger of inhalant intoxication. Persons with adolescent conduct disorder or adult antisocial personality disorder are prone to taking extreme risks, and many inhalant users have those disorders. Several studies suggest an association of inhalant use and conduct problems. Among youths in grades 7 through 12, inhalant users (compared with others who used no drugs or who used only cannabis or alcohol) had many characteristics suggesting conduct disorder. They accepted

cheating more readily, admitted to more stealing, perceived less objection to drug use from their families, liked school less, and reported more sadness, tension, anger, and a feeling of being blamed by others. In addition, school surveys showed that solvent users were more likely to be involved with other drugs. Similarly, among youths referred to court-mandated education for minor alcohol offenses, those who also had used inhalants reported fewer school honors and more expulsions, truancy, academic failures, criminal offenses, running away, and associations with troubled peers, as well as many more drug and alcohol problems. More of them also had mothers or siblings with alcohol- or drug-related problems. Some families are burdened by antisocial personality disorder and substance dependence in the adults and by conduct disorder and substance use disorders (often including inhalant abuse or dependence) in the adolescent children, and there is growing evidence that genetics plays a role in these familial disorders.

A group home referred a 16-year-old single Hispanic female to a university substance-treatment program for evaluation and recommendations regarding inhalant problems. The patient had been ordered to the group home for auto theft, menacing with a weapon, and being out of control by her family. By age 15 she had regularly been using inhalants and drinking alcohol heavily. She had tried typewriter erasing fluid, bleach, tile cleaner, hairspray, nail polish, glue, and gasoline, but preferred spray paint. She had huffed paint many times each day for about 6 months at age 15, using a maximum of eight paint cans per day. The patient said, “It blacks out everything.” Sometimes she had lost consciousness, and she believed that the paint had impaired her memory and made her “dumb.”

The patient reported sexual abuse by an older, nonparental male relative beginning at age 3 and continuing for many years. By fifth grade she had begun showing extensive conduct problems, eventually including fighting, truancy, multiple runaways, gang involvement, and bringing weapons to school. Her family reportedly permitted gang meetings in their home. The patient reported stabbing one person with a screwdriver, another with a knife, and beating another unconscious with a bat. She said that her violence was greatest when she was intoxicated. The patient listed as her strengths and abilities as drawing, cooking, staying clean, fighting, and giving good tattoos. In formal testing her thinking seemed slow, and she had some difficulty understanding questions. Her I.Q. scores were Verbal, 72; Performance, 87; and Full Scale, 77. She met diagnostic criteria for inhalant dependence, alcohol abuse, and conduct disorder.

The evaluating program recommended (1) individual and group substance treatment, emphasizing the adverse cognitive and health effects of huffing; (2) urine monitoring; (3) further neurological and neuropsychological assessment; (4) family evaluation and treatment addressing the patient's anger about sexual abuse and her rebelliousness; (5) specific attention in treatment to the patient's anger and aggression; (6) psychoeducation concerning contraception and protection from sexually transmitted diseases; and (7) active support for continued schooling, with consideration of placement in special education. The patient returned to the group home for several months, and it is unclear which of these recommendations were implemented. She then rejoined her parents in a distant community. One year after the evaluation the patient and two others died when their speeding car hit a tree. An investigating officer said, “It appears that all of them had been sniffing or huffing paint.”

INHALANT DEPENDENCE AND INHALANT ABUSE

Diagnosis and Clinical Features A diagnosis of inhalant abuse or inhalant dependence should be considered in persons showing intermittent changes compatible with substance intoxication, together with an odor of organic solvents, inhalation paraphernalia, or the occasionally present perioral or perinasal papular glue-sniffers' rash. The cardinal feature of inhalant abuse is repeated use of inhalants in ways that produce a physical hazard or adverse social consequences for the user (see Table 11.1-8). Inhalant dependence is characterized by repeated use resulting in some combination of adverse consequences, loss of control of the drug use, and tolerance or withdrawal (see Table 11.1-3). Although DSM-IV provides no diagnosis for inhalant withdrawal, 17 percent of inhalant users in DSM-IV's substance field trial complained of withdrawal symptoms after inhalant use. Thus, despite the absence of a separate inhalant withdrawal diagnosis, patients' complaints of withdrawal symptoms probably should be counted toward a diagnosis of inhalant dependence. ICD-10 uses the category of volatile substance dependence syndrome for inhalant dependence. ICD-10 does not include the diagnostic category of inhalant abuse, offering instead the category of harmful use of volatile substances, which is defined as a pattern of use causing damage to health.

Differential Diagnosis Three conditions should be considered. First, most adolescents who experiment with inhalants stop spontaneously after one or a few episodes of use, never meeting criteria for diagnoses of inhalant abuse or inhalant dependence. Although such use is dangerous and occasionally fatal (most often on the first try), the many spontaneous resolutions support recommendations of minimalist, nonalarmist interventions for adolescent users who do not meet criteria for diagnoses of inhalant use disorders; parents also should be involved. Second, polysubstance use is common in adolescent patients, and abuse or dependence on drugs other than inhalants is established through history, physical findings, and toxicological screens. Such disorders may exist in addition to, or instead of, inhalant abuse or dependence. Third, uncontrolled and impulsive behavior during repeated inhalant intoxications may mimic aspects of, or be comorbid with, conduct disorder or antisocial personality disorder. Antisocial behavior before the onset of inhalant abuse or dependence, or in periods of abstinence, suggests the presence of these disorders.

Course and Prognosis The relatively high prevalence of inhalant use in high school surveys, and its relatively low prevalence in adulthood, led one expert to state that inhalant use “should be regarded as a passing phase or fad.” However, although most inhalant users probably do not progress to serious adult disorders, the risk of such progression is much greater for those who have used inhalants than for those who have not. Studies indicate that inhalant use is associated with increased risk for future diagnoses of antisocial personality disorder and other substance use disorders. Among adult substance-dependent patients, a history of inhalant use indicated a significantly enhanced risk for antisocial personality disorder, social phobia, polysubstance use, and injection drug use. Among a group of adult heroin addicts, those who had been inhalant users appeared to be a “marginal group with particularly unfavorable developmental conditions and a specific course of addition.”

Most adolescents who progress from inhalant use to inhalant abuse or dependence eventually shift to other drugs, but some continue active inhalant dependence into adulthood. Such chronic patients may use the drugs for extended periods each day for many years; they demonstrate moderate criminal activity, weight loss, medical disease, slow and slurred speech, impaired attention and memory, and often are both dirty and louse-ridden.

Tolerance occurs and, less commonly, mild withdrawal involving sleep disturbance, irritability, shakiness, sweating, fleeting illusions, and nausea. One observer reported tachycardia, delusions, and hallucinations during withdrawal.

Medical problems in chronic users may include (1) muscle weakness, sometimes with myoglobinuria and rhabdomyolysis; (2) gastrointestinal problems, such as pain, nausea, vomiting, or hematemesis; (3) renal dysfunction, often with severe electrolyte imbalance; (4) cardiomyopathy; (5) hepatotoxicity; (6) pulmonary disorders (pulmonary hypertension, increased airway resistance, and acute respiratory distress); and (7) hematopoietic disorders (including elevated carboxyhemoglobin levels, methemoglobinemia, hemolytic anemia, aplastic anemia, and even acute myelocytic leukemia). Neurological problems include (1) headache, (2) paresthesias with peripheral neuropathy, (3) reversible cerebellar signs or cerebellar degeneration, (4) radiological abnormalities of widened sulci and basal cisterns, and (5) dementia (e.g., lead encephalopathy from leaded gasoline or white matter dementia from toluene).

Researchers have examined individual patients or small series of mothers who regularly used toluene during pregnancy. The studies, although they need large-scale replication, strongly suggest that such inhalant use, often with accompanying distal renal tubular acidosis in the mother, has devastating effects. Mothers may experience nausea, vomiting, abdominal pain, elevated blood pressure, and early contractions. Preterm delivery is common, and even after correction for gestational age, the infants show intrauterine growth retardation. Growth retardation continues postnatally. Dysmorphic facies, similar to those of the fetal alcohol syndrome, may occur. Perinatal infant deaths are not infrequent. The management of pregnancy in women with inhalant abuse or inhalant dependence should aim at abstinence with attention to the early detection of renal tubular acidosis, preterm labor, and fetal growth retardation.

Treatment No controlled studies guide the treatment of adults or adolescents who meet criteria for inhalant abuse or inhalant dependence. Obviously, appropriate medical care is required for the disorders' medical sequelae. In addition vigorous treatment is needed for adolescent patients who progress from experimentation to inhalant abuse or dependence. Most of these youths have comorbid conduct disorder and are at serious risk for adverse outcomes.

One authority recommends that a comprehensive treatment plan include eight aspects: detoxification; a peer advocate system; assessment of physical, cognitive, and neurological deficits; building on existing strengths; developing new strengths; therapists trained in solvent abuse; attention to personal and family issues; and assistance in returning back to the community.

The author and his colleagues provide day-treatment and nonhospital residential programs for adolescents 13 to 19 years of age with combined substance dependence and conduct disorder. About 20 percent of males and 6 percent of females admitted have diagnoses of inhalant abuse or dependence. However,

treatment for these polysubstance-dependent, delinquent youths is not specific by drug category (e.g., inhalants), but instead targets substance use and conduct problems generally. Most referrals are from social-service and juvenile-justice agencies, which pay for the treatment. Currently suicidal youths and those with recent fire-setting are excluded, although many patients have past histories of these problems, together with considerable violence and gang involvement.

Treatment in these programs begins with detailed interviews addressing use and establishing diagnoses of abuse or dependence for each drug category in DSM-IV. Most patients meet criteria for dependence on several substances. Interviews also address diagnoses of disorders commonly comorbid in this group: conduct disorder attention-deficit/hyperactivity disorder, major depressive disorder, dysthymic disorder, and posttraumatic stress disorder. Interviews also address experiences of abuse or neglect, which are very common in these patients. Group and individual therapy is behaviorally oriented, with immediate rewards for progress in treatment and punishments for lapses to previous behaviors. Patients attend onsite schools with special education teachers, together with planned recreational activities. The programs provide birth control consultations with Planned Parenthood as well as a nursery for the babies of adolescent female patients. The patients' parental families, often very chaotic, are engaged in a modification of structural family therapy. Participation in 12-step programs is required. Treatment interventions are coordinated closely with interventions by community social workers and probation officers. No medications are prescribed for inhalant abuse or dependence per se, but a child-and-adolescent psychiatrist often prescribes antidepressants for depression, disulfiram (Antabuse) for comorbid alcohol dependence, or pemoline (Cylert) for attention-deficit/hyperactivity disorder. Progress is monitored with urine and breath samples at intake and frequently during treatment. Patients who fail to abstain in day treatment may transfer to residential care. Treatment usually lasts 3 to 12 months. Termination is considered successful if the youth has practiced a plan to stay abstinent in a supportive, drug-free environment; to interact with the family in a more productive way; to work or attend school; and to associate with drug-free, nondelinquent peers. In many cases, of course, these goals are only partially accomplished. As with all other treatments for inhalant abuse or dependence, controlled studies of long-term outcomes are lacking.

Laboratory Examinations Therapy for substance use disorders often uses repeated tests of biological samples to validate patients' reports of abstinence or use. However, with inhalants such tests may be difficult to interpret. First, these volatile compounds have a relatively brief sojourn in the body and may be detected in urine for only a few hours after use. Second, even if the compounds occur in urine, they may volatilize out of samples during transfer or storage. Third, although hippuric acid, a toluene metabolite, can be detected longer than toluene, hippuric acid also may be produced from foods, raising a question of false-positive findings. Fourth, inhalants may bind to, or pass through, the plastic of urine cups or breath collection bags, reducing concentrations and making the compounds undetectable. Thus, the most careful monitoring of inhalant use would involve frequent urine samples (e.g., two or three per week) at random times, collected in tightly sealed glass containers with little or no air space, and carefully refrigerated until analysis. Analyses would be both for inhalants themselves and for the ratio of hippuric acid to creatinine. However, even under these ideal conditions, the short half-life of inhalants makes inhalant monitoring much less valuable clinically than monitoring for many other substances.

Similarly, breath samples may be collected in specially designed glass traps or in Tedlar (not Mylar) bags. However, the half-life of inhalants in alveolar air apparently is a matter of minutes, and so breath samples may be useless for monitoring treatment progress in patients who show no current signs of intoxication.

INHALANT INTOXICATION

Diagnosis and Clinical Features Inhalant intoxication should be considered in persons showing an acute onset of behavioral disturbance, coupled with the characteristic odor of organic solvents or the presence of inhalation paraphernalia. Inhalant intoxication is an inhalant-related, clinically significant maladaptive behavioral disorder that develops during or immediately after inhalant use and (assuming survival) clears a few hours later. Intoxication signs initially may include vomiting and motor stimulation, followed by slowing, ataxia, depressed reflexes, slurred speech, disorientation, impaired judgment, lethargy, or coma. Bronchospasm, chest pain, cardiac arrhythmias or arrest, trauma, accidental burns, seizures, aspiration of vomitus, or suffocation in a plastic bag may result. Users often show slowed speech, elated mood, fearfulness, illusions, auditory and visual hallucinations, delusions, and perceptions of altered body size. The DSM-IV diagnostic criteria are listed in Table 11.8-2. ICD-10 provides a comparable diagnostic category of acute intoxication due to use of volatile substances.

A. Recent intentional use or short-term, high-dose exposure to volatile inhalants (exc halling anesthetic gases and short-acting vasodilators).
B. Clinically significant maladaptive behavioral or psychological changes (e.g., belligerence, assaultiveness, apathy, impaired judgment, impaired social or occupational functioning) that developed during, or shortly after, use of or exposure to volatile inhalants.
C. Two or more of the following signs, developing during, or shortly after, inhalant use or exposure:
(1) dizziness
(2) nystagmus
(3) incoordination
(4) slurred speech
(5) unsteady gait
(6) lethargy
(7) depressed reflexes
(8) psychomotor retardation
(9) tremor
(10) generalized muscle weakness
(11) blurred vision or diplopia
(12) stupor or coma
(13) euphoria
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Reprinted with permission from American psychiatric Association: Diagnostic and Statistical manual of mental Disorders, edit 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.8-2 DSM-IV Diagnostic Criteria for Inhalant Intoxication

Laboratory Examinations As noted above, random, intermittent monitoring of biological samples has modest value for confirming self-reported abstinence in nonintoxicated patients with inhalant abuse or dependence, because inhalants are so briefly detectable in the body. However, patients showing behavioral signs of inhalant intoxication do, in most cases, have detectable concentrations of inhalants in urine and breath. Confirming their presence may have later clinical or forensic value. Urine samples should be collected in glass vessels with little or no air space, sealed tightly, and refrigerated until analysis for inhalants themselves and for the ratio of hippuric acid (a toluene metabolite) to creatinine. Alternatively, breath samples may be collected in specially-designed glass traps or in Tedlar (not Mylar) bags, and they should be analyzed for inhalants within a few hours.

Differential Diagnosis Differentiation from other intoxications is aided by a history of inhalant use, the presence of inhalant odor and residues on the skin or clothing, a characteristic perioral rash from contact with organic solvents, and toxicological examination of body fluids. Polysubstance use is common among solvent users, and concurrent intoxications with other drugs may be assessed by history and toxicological examinations. Despite evidence of inhalant intoxication, in comatose patients other explanations (e.g., closed head injury) must be sought. Dextrose 50 percent for injection (50 grams) and naloxone (Narcan) 2 mg intravenously help rule out coma of diabetic or narcotic origin. If delirium develops in the course of an intoxication with inhalants, the diagnosis is inhalant intoxication delirium, rather than inhalant intoxication. If a mood disturbance, anxiety, or psychosis appear very prominently during an intoxication and if those symptoms are severe enough to warrant independent clinical attention, the diagnosis should be inhalant-induced mood disorder, inhalant-induced psychotic disorder, or inhalant-induced anxiety disorder, respectively.

Course and Prognosis The onset of intoxication is almost instantaneous after the inhalation of volatile hydrocarbons, given the rapid absorption of those inhalants across pulmonary membranes and their quick distribution into the brain and other lipids. Inhalant drugs are rapidly metabolized and excreted, and inhalant intoxication usually lasts a few hours or less. Unless trauma, hypoxia, cardiac arrest, burns, or other problems ensue, there probably are no lasting effects from one or a few intoxications, except that each use of these reinforcing drugs increases the probability of further use. Prolonged, repeated use causes persisting effects.

Treatment Inhalant intoxication, like alcohol intoxication, usually receives no medical attention and resolves spontaneously. However, effects of the intoxication, such as coma, bronchospasm, laryngospasm, cardiac arrhythmias, trauma, or burns, need treatment. Otherwise, care primarily involves reassurance, quiet support, and attention to vital signs and level of consciousness. Some evidence suggests that physical agitation during inhalant intoxication may precipitate cardiac arrhythmias or cardiac arrest, so the environment should be calming and reassuring. However, sedative drugs, including benzodiazepines, are contraindicated, since they may potentiate inhalant effects. Because of the short half-life of inhalants, inhalant intoxication usually improves considerably after about 30 minutes of abstinence, unless other drugs were also consumed. Following resolution of the intoxication a careful evaluation is needed with appropriate intervention or referral for inhalant abuse or dependence, other substance use disorders, conduct disorder, or antisocial personality disorder.

INHALANT INTOXICATION DELIRIUM

DSM-IV provides a diagnostic category for inhalant intoxication delirium (see Table 10-22). Inhalant intoxication delirium is a disturbance of consciousness and a

change in cognition that results from intoxication with inhalants and is not better explained by dementia. The course and treatment are like those of inhalant intoxication but the additional confusion requires special attention to patient safety. If the delirium results in severe behavioral disturbances, short-term treatment with a dopamine receptor antagonist—for example, haloperidol (Haldol)—may be necessary. Benzodiazepines should be avoided because of the possibility of adding to the patient's respiratory depression. ICD-10 provides a comparable diagnostic category, “acute intoxication due to use of volatile substances, with delirium.”

INHALANT-INDUCED PERSISTING DEMENTIA

Diagnosis and Clinical Features Studies of inhalant-caused cognitive impairment have been equivocal and have been beset by numerous methodological problems, including a focus on adolescent users with briefer lifetime exposures. But clinical and some research evidence suggests that some inhalant-using adults develop inhalant-induced persisting dementia. For example, among toluene users (average age, 29 years) studied with magnetic resonance imaging, the neuropsychological deficits correlated strongly with the severity of cerebral white matter abnormalities.

The cardinal feature of the disorder is dementia resulting from the use of inhalants (see Table 10-31). Nearly all of these persons also meet the criteria for inhalant dependence. Patients with inhalant-induced persisting dementia have memory impairment and at least one of the following: aphasia (language disturbance), apraxia (impaired ability to carry out motor activities despite intact motor function), agnosia (failure to recognize or identify objects despite intact sensory function), and disturbed executive functioning (planning, organizing, sequencing, abstracting). The symptoms must significantly impair social or occupational functioning, represent a decrement from earlier functioning, not occur exclusively in the course of a delirium, and persist beyond the usual duration of inhalant intoxication. The ICD-10 category of dementia in other specified diseases classified elsewhere includes inhalant-induced persisting dementia.

Differential Diagnosis Nearly all of these patients have inhalant dependence and many are dependent on alcohol, which also produces dementia. Moreover, histories of head injury are very common among such patients. Thus, despite clear evidence of prolonged inhalant use, this disorder requires a full evaluation for the multiple causes of dementia.

Course and Prognosis Few of these patients have been studied prospectively. Despite some reports of improvement when patients abstained from inhalants, it seems likely that most neuropsychological deficits that persist for days or weeks after intoxication will continue or worsen. Moreover, as dementia progresses, patients become more difficult to treat, and each relapse adds to their cerebral toxicity.

Treatment There is no established treatment for the cognitive and memory problems of inhalant-induced persisting dementia. Low-key street outreach and extensive social service support have been offered to severely deteriorated inhalant-dependent homeless adults. Patients may require extensive support within their families or in foster or domiciliary care.

INHALANT-INDUCED PSYCHOTIC DISORDER

The essential features of inhalant-induced psychotic disorder are prominent hallucinations or delusions judged to be due to the direct physiological effect of inhalant substances. Such psychotic symptoms sometimes develop during intoxication with inhalants, so this diagnosis applies to patients who meet criteria for inhalant intoxication but who also have psychotic symptoms in excess of those usually associated with inhalant intoxication. The psychotic symptoms must be severe enough to warrant independent clinical attention. This diagnosis is not made in the presence of inhalant intoxication delirium. The clinician can specify whether hallucinations or delusions predominate (Table 13.3-4).

The course and treatment of inhalant-induced psychotic disorder are like those of inhalant intoxication. The disorder is brief, lasting a few hours to (at most) a very few weeks beyond the intoxication. Vigorous treatment of such life-threatening complications as respiratory or cardiac arrest, together with conservative management of the intoxication itself, is appropriate. Confusion, panic, and psychosis mandate special attention to patient safety. Severe agitation may require cautious control with haloperidol (5 mg per 70 kg intramuscularly, repeated once in 20 minutes if needed). Sedative drugs, including benzodiazepines, may potentiate and worsen inhalant intoxications.

Despite similarities in name, DSM-IV's inhalant-induced psychotic disorder differs from ICD-10's residual and late-onset psychotic disorder due to volatile substance use. The former is a variant of inhalant intoxication, while the later is conceptualized as persisting long after direct psychoactive substance effects abate. Controversy continues as to whether inhalants produce persisting psychotic states. In ICD-10, acute intoxication due to use of volatile substances includes patients with marked psychotic symptoms arising in the course of the intoxication.

INHALANT-INDUCED MOOD DISORDER

The essential feature of inhalant-induced mood disorder (see Table 14.6-18) is a prominent disturbance of mood judged to be due to the direct physiological effect of inhalant substances. Such mood symptoms sometimes develop during intoxication with inhalants, so this diagnosis applies to patients who meet criteria for inhalant intoxication but who also have mood symptoms in excess of those usually associated with inhalant intoxication. The mood symptoms must be severe enough to warrant independent clinical attention. The clinician can specify one of the following subtypes: with depressive features (probably the more common subtype), with manic features, or with mixed features. This diagnosis is not made in the presence of inhalant intoxication delirium. In ICD-10, acute intoxication due to use of volatile substances includes patients with marked affective symptoms arising in the course of the intoxication.

The course and treatment of inhalant-induced mood disorder are like those of inhalant intoxication. Inhalant-induced mood disorder is brief, lasting a few hours to (at most) a very few weeks beyond the intoxication. Although antidepressant or antimanic drugs are seldom appropriate for these relatively brief disorders, a history and psychosocial attention to suicide are important. Suicide has been implicated in 40 percent of inhalant-related medical examiners' death reports and in 38 percent of inhalant-related visits to hospital emergency departments.

INHALANT-INDUCED ANXIETY DISORDER

The essential features of inhalant-induced anxiety disorder (see Table 15.6-18) are prominent anxiety symptoms judged to be due to the direct physiological effect of inhalant substances. Such anxiety symptoms sometimes develop during intoxication with inhalants, so this diagnosis applies to patients who meet criteria for inhalant intoxication but who also have anxiety symptoms in excess of those usually associated with inhalant intoxication. The anxiety symptoms must be severe enough to warrant independent clinical attention. This diagnosis is not made in the presence of inhalant intoxication delirium. The clinician can specify one of the following subtypes: with generalized anxiety, with panic attacks, with obsessive-compulsive symptoms, or with phobic symptoms; generalized anxiety and panic attacks are probably most common. In ICD-10, acute intoxication due to use of volatile substances includes patients with marked anxiety symptoms arising in the course of the intoxication.

The course and treatment of inhalant-induced anxiety disorder are like those of inhalant intoxication. Sedative and antianxiety drugs are contraindicated, since they worsen inhalant intoxication, which precipitates inhalant-induced anxiety disorder.

INHALANT-RELATED DISORDER NOT OTHERWISE SPECIFIED

The diagnosis of inhalant-related disorder not otherwise specified is reserved for inhalant-related disorders that do not fit into one of the above diagnostic categories (Table 11.8-3).

The inhalant-related disorder not otherwise specified category is for disorders associated with the use of inhalants that are not classifiable as inhalant dependence, inhalant abuse, inhalant intoxication, inhalant intoxication delirium, inhalant-induced persisting dementia, inhalant-induced psychotic disorder, inhalant-induced mood disorder, or inhalant-induced anxiety disorder.

Table 11.8-3 DSM-IV Diagnostic Criteria for Inhalant-Related Disorder Not Otherwise Specified

NITROUS OXIDE–RELATED DISORDERS

DSM-IV includes nitrous oxide–related disorders among other substance-related disorders because of differences in modes of action and associated problems. Nitrous oxide was introduced for clinical practice in 1844 and is still a widely used inorganic gas anesthetic. It also is a propellant in whipped-cream dispensers. Not surprisingly, nitrous oxide misuse seems to appear most commonly among the health-care and food-service workers who use these preparations, although there is little scientific information on its epidemiology. The drug has a rapid onset and offset of action. It is mostly excreted in the breath with little or no biotransformation. Some users experience euphoria and a pleasant dreamlike state with nitrous oxide, but several studies show that it is not a reinforcer for most research subjects who have been exposed only a few times. Chronic use may produce diffuse polyneuropathy and myelinopathy with extensive, although sometimes reversible, neurological symptoms mimicking those of vitamin B_{12}–related pernicious anemia. Active vitamin B_{12} requires reduced cobalt, but nitrous oxide irreversibly oxidizes cobalt, suppressing the activity of an important enzyme, methionine synthase.

A 35-year-old male dentist with no history of other substance problems complained of problems with nitrous abuse for 10 years. This had begun as experimentation with what he had considered a harmless substance. However, his rate of use increased over several years, eventually becoming almost daily for months at a time. He felt a craving before sessions of use. Then, using the gas while alone in his office, he immediately felt numbness, a change in his temperature and heart rate, and an alleviation of depressed feelings. “Things would go through my mind. Time was erased.” He sometimes fell asleep. Sessions might last a few minutes, or up to 8 hours; they ended when the craving and euphoria ended. He had often tried to stop or cut down, sometimes consulting professionals about the problem.

Cases in the literature strongly suggest that nitrous oxide intoxication, dependence, and abuse do occur. Considering the drug's brief duration of action and users' intermittent inhalation patterns, nitrous oxide appears unlikely to produce clinically significant withdrawal. It is not clear, however, whether nitrous oxide produces other substance-related disorders. There are no clinically practical urine or breath tests for the presence of nitrous oxide. Misusers of the drug who develop neurological symptoms may have low serum concentrations of vitamin B_{12}, and remission of neurological symptoms has followed combined administration of B_{12} and folate, together with abstinence from nitrous oxide. In the absence of controlled studies of treatment for nitrous oxide use disorders, general principles for treating other substance use disorders should guide the treatment of these patients.

AMYL AND BUTYL NITRITE–RELATED DISORDERS

DSM-IV includes amyl and butyl nitrite–related disorders among other substance-related disorders. The vapors of amyl nitrite, a volatile liquid, produce vasodilation and smooth muscle relaxation. Supplied in easily “popped” glass vials, amyl nitrite found wide use as acute inhalation therapy for angina pectoris between 1867 and about 1980. Sublingual nitroglycerin tablets now have superseded that use. Amyl nitrite and its close relative, butyl nitrite, enjoyed a flurry of use as recreational drugs during the 1970s and 1980s. Reports from that time suggest that the use mainly was among adolescents, users of other drugs, and homosexual men. The latter group especially reported using the drugs to enhance orgasm, and these nitrites also were considered to produce a high, a feeling of wild abandon, or an altered state of consciousness. Nonmedical users called amyl nitrite vials “poppers.” Those vials' availability declined with their declining medical use. Butyl nitrite was then increasingly sold as a “room odorizer” (but really for inhalation) under suggestive names such as Rush or Locker Room.

In one study of patients who had misused these drugs, most experienced dizziness, lightheadedness, cardiac palpitations, blurred vision, and a feeling of warmth immediately after the inhalation. Others complained of immediate headache, nasal burning, nausea, cough, dyspnea, or syncope. Nearly half found the experience not at all pleasant, and the others rated it only as “fair to good.”

A few case reports document severe, but nonfatal methemoglobinemia after nitrite inhalation by persons genetically deficient in methemoglobin reductase. There also are reports of rapidly fatal methemoglobinemia after oral ingestion (not inhalation) of these compounds by genetically normal persons. Some epidemiological studies further suggested that among homosexual men nitrite use might increase the risk of HIV infection or of developing Kaposi's sarcoma after HIV infection. However, those observations were heavily confounded, since nitrite users tended to have more partners and to pursue unsafe sexual practices more frequently.

Considered together, the lack of a compelling instantaneous high, the immediate unpleasant feelings, and the longer-term risks from amyl and butyl nitrite seem to have reduced the prevalence of their use. In 1979 an annual national survey of high school seniors found that 6.5 percent reported using these drugs in the previous year, but that percentage fell to 1.2 percent by 1997. Similarly, among Americans over 12 years of age living in households, only 2 percent said in 1996 that they ever had used amyl nitrite “for kicks or to get high.”

No studies have examined users of amyl or butyl nitrite with modern diagnostic procedures, and it is unknown how many (if any) meet criteria for intoxication, abuse, or dependence. Some persons with sustained industrial exposures to nitrites experience withdrawal headaches on weekends or vacations, but it seems unlikely that intermittent, brief exposures to these short-acting drugs would produce clinically significant withdrawal in recreational misusers of nitrites. There are no studies of the treatment of amyl or butyl nitrite–related disorders, and so such treatment should follow general principles for the treatment of other substance-related disorders.

SUGGESTED CROSS-REFERENCES

An overview of substance-related disorders is given in Section 11.1, hallucinogen-related disorders are discussed in Section 11.7, and phencyclidine-related disorders are discussed in Section 11.11. Mood disorders are discussed in Chapter 14. Chapter 15 reviews anxiety disorders, and Chapter 13 covers psychotic disorders.

SECTION REFERENCES

Altenkirch H, Kindermann W: Inhalant abuse and heroin addiction: A comparative study on 574 opiate addicts with and without a history of sniffing. Addict Behav *11:*93, 1986.

Boutros NN, Bowers MB: Chronic substance-induced psychotic disorders: State of the literature. J Neuropsychiatry Clin Neurosci *8:*262, 1996.

Chang AP, England JD, Garcia CA, Sumner AJ: Focal conduction block in n-hexane polyneuropathy. Muscle Nerve *21:*964, 1998.

Compton WM, Cottler LB, Dinwiddie SH, Spitznagel EL, Mager DE, Asmus G: Inhalant use: Characteristics and predictors. Am J Addict *3:*263, 1994.

Cottler LB, Schuckit MA, Helzer JE, Crowley T, Woody G, Nathan P, Hughes J: The DSM-IV field trial for substance use disorders: Major results. Drug Alcohol Dependend *38:*59, 1995.

*Dinwiddie SH: Abuse of inhalants: A review. Addiction *89:*925, 1994.

Dinwiddie SH, Reich T, Cloninger CR: The relationship of solvent use to other substance use. Am J Drug Alcohol Abuse *17:*173, 1991.

Dinwiddie SH, Reich T, Cloninger CR: Solvent use as a precursor to intravenous drug abuse. Compr Psychiatry *32:*133, 1991.

Echeverria D, Fine L, Langolf G, Schork A, Sampaio C: Acute neurobehavioral effects of toluene. Br J Ind Med *46:*483, 1989.

Esmail A, Meyer L, Pottier A, Wright S: Deaths from volatile substance abuse in those under 18 years: Results from a national epidemiological study. Arch Dis Child *69:*356, 1993.

*Evans EB, Balster RL: CNS depressant effects of volatile organic solvents. Neurosci Biobehav Rev *15:*233, 1991.

Evans EB, Balster RL: Inhaled 1,1,1-trichloroethane-produced physical dependence in mice: Effects of drugs and vapors on withdrawal. J Pharmacol Exp Ther *264:*726, 1993.

Filley CM, Heaton RK, Rosenberg NL: White matter dementia in chronic toluene abuse. Neurology *40:*532, 1990.

*Flanagan RJ, Ives RJ: Volatile substance abuse. Bull Narc *46:*49, 1994.

Johns A: Volatile substance abuse and 963 deaths. Br J Addict *86:*1053, 1991.

*Johnson EO, Schutz CG, Anthony JC, Ensminger ME: Inhalants to heroin: A prospective analysis from adolescence to adulthood. Drug Alcohol Depend *40:*159, 1995.

Johnston LD, O'Malley PM, Bachman JG: *National Survey Results on Drug Use from the Monitoring the Future Study, 1975–1994*, vol 1. National Institute on Drug Abuse, Rockville, MD, NIH publ no. 95-4026, 1995.

Kozel N, Sloboda Z, De La Rosa M: *Epidemiology of Inhalant Abuse: An International Perspective*. National Institute on Drug Abuse research monogr 148, Rockville, MD. NIH publ no. 95-3831, 1995.

*Levinthal CF: Drugs, behavior and modern society. Allyn & Bacon, Boston, 1999.

Morton HG: Occurrence and treatment of solvent abuse in children and adolescents. Pharmacol Ther *33:*449, 1987.

Pearson MA, Hoyme E, Seaver LH, Rimsza ME: Toluene embryopathy: Delineation of the phenotype and comparison with fetal alcohol syndrome. Pediatrics *93:*211, 1994.

Ramsey J, Anderson HR, Bloor K, Flanagan RJ: An introduction to the practice, prevalence and chemical toxicology of volatile substance abuse. Hum Toxicol *8:*261, 1989.

Rosenberg NL, Kleinschmidt-DeMasters, Davis KA, Dreisbach JN, Hormes JT, Filley CM: Toluene abuse causes diffuse central nervous system white matter changes. Ann Neurol *23:*611, 1988.

*Schutz CG, Chilcoat HD, Anthony JC: The association between sniffing inhalants and injecting drugs. Compr Psychiatry *35:*99, 1994.

Sharp CW, Rosenberg NL: Volatile substances. In *Substance Abuse: A Comprehensive Textbook*, ed 2, JH Lowinson, P Ruiz, RB Millman, JG Langrod, editors. Williams & Wilkins, Baltimore, 1992.

Spiller HA, Krenzelok EP: Epidemiology of inhalant abuse reported to two regional poison centers. Clin Toxicol *35:*167, 1997.

Substance Abuse and Mental Health Services Administration: *Annual Medical Examiner Data 1994*. Rockville, MD, DHHS publ no. (SMA) 96-3078, 1996.

Substance Abuse and Mental Health Services Administration: *Annual Emergency Department Data 1994*. Rockville, MD, DHHS publ no. (SMA) 96-3104, 1996.

Substance Abuse and Mental Health Services Administration: *National Household Survey on Drug Abuse: Population Estimates 1995*. Rockville, MD, DHHS publ no. (SMA) 96-3095, 1996.

Westermeyer J: The psychiatrist and solvent-inhalant abuse: Recognition, assessment, and treatment. Am J Psychiatry *144:*903, 1987.

Wilkins-Haug L, Gabow PA: Toluene abuse during pregnancy: Obstetric complications and perinatal outcomes. Obstet Gynecol *77:*504, 1991.

Young SE, Mikulich SK, Goodwin MB, Hardy J, Martin CL, Zoccolillo MS, Crowley TJ: Treated delinquent boys' substance use: Onset, pattern, relationship to conduct and mood disorders. Drug Alcohol Depend *37:*149, 1995.

11.9 NICOTINE-RELATED DISORDERS

JOHN R. HUGHES, M.D.

Nicotine dependence is the most prevalent, most deadly, most costly, yet most treatable of the substance dependencies. In the past psychiatrists have often not participated in the diagnosis and treatment of nicotine dependence, perhaps because of the incorrect notion that most smokers do not need treatment or that psychiatric patients are unlikely to respond to smoking cessation treatment because most health insurers or organizations do not reimburse for treatment of smoking. There is increasing recognition that smoking is usually a form of substance dependence and that treatment for smoking is highly cost effective. In addition, persons who continue to smoke despite significant social pressure appear to be those who are severely nicotine dependent or who have significant psychiatric problems that interfere with cessation. These two trends—increasing legitimization of treatment and a selection bias toward highly dependent smokers with comorbid mental disorders—suggest that psychiatrists will play a larger role in treating smoking cessation.

DEFINITION

Nicotine dependence and withdrawal are the two defined nicotine-related disorders in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). The essential feature of any substance dependence disorder in DSM-IV is that "the individual continues use of the substance despite significant substance-related problems." Because 50 percent of smokers die of a smoking-related illness, this definition clearly is applicable to nicotine use. The essential feature of withdrawal is "a substance-specific maladaptive behavioral change... that is due to the cessation of, or reduction in, heavy and prolonged substance use." Because nicotine withdrawal produces an observable, well-defined, time-limited syndrome in over half of smokers, this definition also appears appropriate.

Nicotine abuse is not included in DSM-IV because abuse is confined to significant psychosocial but not physical problems and the former is rare with nicotine use. The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) includes harmful use, a category similar to abuse but that includes continued use that causes physical problems; thus, harmful use from the nicotine-containing products often occurs. Nicotine intoxication is very rare; it is included in ICD-10 but not DSM-IV.

HISTORY AND COMPARATIVE NOSOLOGY

Tobacco use in the New World dates back to at least 600 AD and was introduced into European culture in the sixteenth century. Early on, most tobacco use was via pipes, smokeless tobacco, or cigars. The cigarette became popular beginning in the early 1900s with the invention of the cigarette-making machine and the use of acidifying agents to permit nicotine to enter the lower respiratory tract where it could be rapidly absorbed into the arterial circulation. Cigarette use grew dramatically in the first half of the twentieth century. The first reports of the association of smoking and disease began in the 1950s and culminated with the 1964 *Surgeon General's Report on Smoking and Health*. Use in the United States fell slightly after this but then began to decline dramatically in the 1970s but has plateaued recently. However, in the 1990s use has stabilized plus noncigarette products such as cigars and smokeless tobacco use have increased. Although tobacco use in other Western nations has declined somewhat, use in developing countries is actually increasing.

Nicotine dependence was widely accepted but not codified until the 1980 third edition of DSM and (DSM-III) and the ninth revision of the International Statistical Classification of Diseases (ICD-9) included tobacco dependence and withdrawal as disorders. The revised third edition of DSM (DSM-III-R) and the 1988 U.S. Surgeon General's *The Health Consequences of Smoking—Nicotine Addiction* concluded that smoking was a form of substance dependence.

EPIDEMIOLOGY

In 1998 25 percent of Americans smoked, 25 percent are former smokers, and 50 percent have never smoked cigarettes. The prevalence of pipe, cigar, and smokeless tobacco use is less than 2 percent. The prevalence of smoking in the United States was decreasing about 1 percent a year but it has not changed in the last 4 years because of increased initiation and decreased cessation.

The mean age of onset of smoking is 16 and few persons start after 20. Dependence features appear to develop quickly. Classroom and other programs to prevent initiation are only mildly effective and increased taxation does decrease initiation.

Over 75 percent of smokers have tried to quit, about 40 percent try to quit each year. On a given quit attempt only 30 percent remain abstinent for even 2 days and only 5 to 10 percent stop permanently. However, most smokers make 5 to 10 attempts such that eventually 50 percent of eversmokers quit. In the past, 90 percent of successful quit attempts involved no treatment. However, with the advent of over-the-counter (OTC) and nonnicotine medications in 1998, about one third of all quits were due to use of medication.

In terms of the diagnosis of nicotine dependence per se, about 20 percent of the population develops nicotine dependence at some point, making it the most prevalent psychiatric disorder. Approximately 85 percent of current daily smokers are nicotine dependent. Nicotine withdrawal occurs in about 50 percent of smokers who try to quit.

Smoking is now as common in women as in men. Smoking is more prevalent in those with lower education and income, in most ethnic groups, and is especially high in psychiatric patients (50 percent), including those with other substance use disorders (80 percent).

ETIOLOGY

Both nicotine and acetylcholine interact with nicotinic-cholinergic receptors. Originally these receptors were thought to be confined to the ganglia and neuromuscular junction but have now been found in several areas of the central nervous system (CNS). These receptors are unusual in that they quickly desensitize. This phenomena plus the fact that repeated use of nicotine increases, not decreases, the number of receptors suggests that nicotine may actually act much as an antagonist as an agonist at this receptor.

The dependence-producing effects of nicotine appear to be modulated by dopamine (e.g., nicotinic-cholinergic receptors lie on dopamine neurons, nicotine increases dopamine, and dopamine blockers attenuate nicotine self-administration). Nicotine also increases norepinephrine, epinephrine, and serotonin and these increases may modulate some of the reinforcing effects from cigarettes.

Nicotine via cigarettes is rapidly absorbed directly into the arterial circulation and reaches the CNS in less than 15 seconds. Peak behavioral and cardiovascular effects occur within a few minutes. Nicotine is metabolized via the liver and has a half-life of about 2 hours. Nicotine levels from smoking typically rise in the morning, plateau in the evening, and fall to near zero in the night. This pattern causes an acute tolerance such that the first cigarettes of the day are more potent than later cigarettes.

Nicotine improves performance on long, fatiguing, boring tasks. It decreases anger and stabilizes mood. It decreases hunger and food intake and increases metabolic

rate. When a smoker experiences these effects, it is often not clear how much of them result from nicotine combating withdrawal and bringing the smoker back to normal and how much is actual improvement above the norm. Some of nicotine's effects (e.g., performance enhancement) appear to occur independent of withdrawal relief.

Nicotine use, like most substance use, begins because of social reinforcement. However, with repeated exposure, many young users find the pharmacological effects of nicotine well suited to help them with the demands of adolescence. In addition, a physical dependence on nicotine begins so that periods of nonuse become uncomfortable.

Children more likely to start smoking are those who have a high need to conform, low academic performance, rebelliousness, depressive symptoms, and poor self-esteem. Both peer and family influences are paramount. Attention deficit, conduct, and alcohol or drug use disorders increase the risk of initiation and maintenance of smoking.

Although not widely known, initiation and cessation of tobacco use are as heritable as alcoholism (alcohol dependence or abuse). Some of the genetic effects are shared with alcohol and some are specific to tobacco (Table 11.9-1). The biological and behavioral mechanisms for genetic effects on tobacco use are not known.

	Genetic Variance		
Substance	Common to Both Substances	Substance-Specific	Total
Tobacco	36%	20%	56%
Alcohol	17%	32%	49%

Table 11.9-1 Percent of Variation in Cigarette Smoking and Alcohol Consumption, Attributable to Genetic Sources

DIAGNOSIS AND CLINICAL FEATURES

Table 11.9-2 lists the DSM-IV nicotine-related disorders. Most of the generic criteria for substance dependence readily apply to nicotine (e.g., tolerance, a withdrawal syndrome, use to avoid withdrawal, inability to stop despite repeated attempts, and continued use despite knowing that use is harmful) (Table 11.9-3). Because nicotine is legal and easily available, spending a great deal of time to procure the drug and giving up activities to use the drug are rare.

Nicotine use disorder
Nicotine dependence
Nicotine-induced disorder
Nicotine withdrawal
Nicotine-related disorder not otherwise specified

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.9-2 DSM-IV Nicotine-Related Disorders

Criteria	Examples
Tolerance	Absence of nausea, dizziness, etc; most smokers escalate use to 1 pack/day or more by age 25
Withdrawal	See Table 11.9-4
The substance is often taken in larger amounts or over a longer period than was intended	Most smokers do not intend to smoke 5 years later, but over 70% continue to use
Persistent desire or unsuccessful effort to stop	5–10% of self-quitters are successful on each attempt; 50% of smokers have not been able to stop despite repeated attempts
A great deal of time is spent to obtain, use, or recover from the drug	Leaving work site to smoke
Important activities are given up or reduced because of substance use	Not taking a job because of smoking restrictions in the workplace
Use continues despite knowledge of problems caused by substance	Many smokers have heart disease, chronic obstructive pulmonary disease, or ulcers and continue to smoke

Table 11.9-3 Examples of DSM-IV Nicotine Dependence Criteria

There are several reasons smoking can produce such a severe dependence: (1) nicotine produces many different effects that can be used in many different situations (2) with pulmonary; absorption, nicotine reaches the brain in a few seconds producing a rapid effect; (3) smoking allows the user to titrate the dose; (4) smoking is an intense habit >200 puffs/day × 20 years); (5) there are many environmental cues eliciting smoking (e.g., others smoking and advertisements); and (6) nicotine use almost never impairs the user via intoxication.

Nicotine withdrawal (Table 11.9-4) is manifested by changes in mood, insomnia, difficulty concentrating, restlessness, decreased heart rate (average decline is 8 beats per minute [bpm]), and weight gain (average is 2 to 3 kg). The insomnia appears to be specific to increased awakenings and intense dreaming. Postcessation weight gain is due both to increased eating and the loss of nicotine stimulation of metabolism. Craving is common and increased coughing and poor performance on vigilance tasks can occur. The syndrome is typically worse in cigarette smokers, intermediate in users of smokeless tobacco, and mild in users of nicotine replacement products. Most withdrawal symptoms peak at 1 to 3 days and last 3 to 4 weeks; however, 40 percent of smokers have withdrawal that lasts for more than 4 weeks. In addition, craving and weight gain often persist for 6 months or more.

A. Daily use of nicotine for at least several weeks.
B. Abrupt cessation of nicotine use, or reduction in the amount of nicotine used, followed within 24 hours by at least four of the following signs: (1) dysphoric or depressed mood (2) insomnia (3) irritability, frustration, or anger (4) anxiety (5) difficulty concentrating (6) restlessness (7) decreased heart rate (8) increased appetite or weight gain
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.9-4 DSM-IV Diagnostic Criteria for Nicotine Withdrawal

Abstinence can also have pharmacokinetic effects. Nonnicotine chemicals in tobacco smoke activate cytochrome P450 enzymes, thereby decreasing the levels of several medications. As a result, smoking cessation increases the concentrations of these medications. Many of these medications are psychiatric medications and often the increase can be clinically significant (e.g., haloperidol [Haldol], clozapine [Clozaril], and fluvoxamine [Luvox] concentration increase 30 to 40 percent with abstinence) (Table 11.9-5).

Abstinence Increases Blood Concentrations			
Clomipramine	Desmethyldiazepam	Haloperidol	Nortriptyline
Clozapine	Doxepin	Imipramine	Propranolol
Desipramine	Fluvoxamine	Oxazepam	
Abstinence Does Not Increase Blood Concentrations			
Amitriptyline	Ethanol	Midazolam	
Chlordiazepoxide	Lorazepam	Triazolam	
Effects of Abstinence Unclear			
Alprazolam	Chlorpromazine	Diazepam	

Table 11.9-5 Effect of Abstinence From Smoking on Blood Concentrations of Psychiatric Medicines

Smoking (and particularly nicotine dependence) is two to three times more prevalent among patients with mood, substance use, and other psychiatric disorders. Conversely, these psychiatric disorders are two to three times more common among current smokers than among never or exsmokers. There are several possible reasons for this association—for example, shared genetic influences on smoking and psychiatric disorders, modeling other psychiatric patients, and boredom. Among adolescents the number of depressive symptoms predicts both the likelihood of starting to smoke and the probability of becoming dependent. Recent work suggests that more than nicotine may be responsible for this association. Tobacco smoke and extracts of smokeless tobacco contain substances other than nicotine that inhibit monoamine oxidase type A (MAO_A) and MAO_B, and recent work using positron emission technology has shown less MAO_B activity in the brains of cigarette smokers. The nature of the chemical entity is not known but in in vitro models, this inhibition, like that of most clinically available MAO inhibitors, is irreversible. This suggests that MAO inhibition would persist until the body synthesizes new enzymes, a process that usually takes several weeks. Psychiatric patients may have a special need for the anxiolytic, anorexic, antiaggression, antidepressant, and improved concentration effects of nicotine. Nicotine also increases the self-administration of alcohol, which may predispose smokers to alcoholism. Finally, smokers with a current or past history of a depressive disorder have more withdrawal upon cessation and a lower rate of smoking cessation. DSM-IV also includes a residual category of nicotine-related disorders not otherwise specified (Table 11.9-6).

The nicotine-related disorder not otherwise specified category is for disorders associated with the use of nicotine that are not classifiable as nicotine dependence or nicotine withdrawal.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.9-6 DSM-IV Diagnostic Criteria for Nicotine-Related Disorder Not Otherwise Specified

DIFFERENTIAL DIAGNOSIS

Many of the symptoms of nicotine withdrawal can mimic, exacerbate, or mask the symptoms of psychiatric disorders or the adverse effects of psychiatric medications (e.g., akathesia, anxiety, depression, irritability, insomnia, and weight gain). Some recent data suggests that cessation of smoking can reinitiate a psychiatric (including substance use) disorder in a subgroup of smokers; however, other data refute this.

COURSE AND PROGNOSIS

About half of smokers finally stop smoking; however, half do not and many of those who stopped had unfortunately done so having already developed smoking-related diseases. Twenty percent of all mortality in the United States is due to smoking. About 45 percent of smokers will die of a smoking-related disease resulting in over 410,000 deaths a year in the United States. Smoking is a huge risk for lung cancer, accounting for over 90 percent of all deaths from lung cancer. Smoking doubles the risk for cardiovascular disease deaths, but because this disease is more prevalent, it accounts for more smoking deaths than lung cancer. Other common smoking-related diseases include chronic obstructive pulmonary disease, low–birth-weight, perinatal complications, other cancers (e.g., throat, breast, and pancreas) and ulcers. Second-hand smoke increases the risk of cancer and heart disease in spouses and the incidence of respiratory and ear problems in children.

The tar in cigarette smoke is responsible for the cancers. Irritants and ciliotoxins appear to be responsible for lung diseases. Carbon monoxide and clotting factors appear to be the most likely causes of cardiovascular disease. The role of nicotine in cardiovascular disease is actually not well defined and how many perinatal problems are caused by carbon monoxide or by nicotine is debatable.

Cessation of smoking almost eliminates the risk of heart disease in 5 years and of lung cancer in 20 years. Although cross-sectional epidemiological data suggest that switching to low-tar cigarettes should decrease cancer, this is still not firmly established. Whether decreasing the number of cigarettes decreases risk has also not been directly tested.

Nicotine intoxication causes abdominal pain, dizziness, headaches, nausea, pallor, palpitations, sweating, vomiting, and weakness. Intoxication is rare although mild

symptoms can occur with nicotine replacement; treatment is supportive.

Epidemiological, biochemical, and clinical trial data suggest that nicotine may be beneficial for dementia of the Alzheimer's type, Parkinson's disease, Tourette's disorder, and ulcerative colitis disorders.

TREATMENT

Psychiatrists are urged to read the two guidelines for the treatment of smoking published in 1996 by the Agency for Health Care Policy Research and by the American Psychiatric Association. All patients should be assessed for smoking status, motivation to quit, and motivators for and barriers to quitting. Smoking status includes current, exsmoker, or neversmoker; type of tobacco used; and frequency of use. Motivation to quit can be classified as *precontemplation* (no plans to quit), *contemplation* (thinking about quitting but no plans), and *preparations* (plans to quit in near future). Common motivators to quit are health concerns, effects of smoking on others, and social pressure. Common barriers to cessation are withdrawal, fear of failure, and fear of weight gain.

Psychiatrists should advise all patients who are not in crisis to quit smoking. Many psychiatric patients are precontemplators and thus the psychiatrist's role is often to use the patient's concerns as motivators for cessation and to suggest ways to decrease barriers to cessation. The psychiatrist should also reintroduce cessation at later visits.

Among patients who are ready to stop smoking, it is best to set a quit date. Most clinicians and smokers prefer abrupt cessation but because there is no good data that abrupt cessation is better than gradual cessation, patient preference for gradual cessation should be respected. Brief advice should focus on the need for medication or group therapy, weight gain concerns, high-risk situations, making cigarettes unavailable, and so forth. Because relapse is often rapid, the first follow-up phone call or visit should be 2 to 3 days after the quit date. These strategies have been shown to double self-initiated quit rates (Table 11.9-7). The psychiatrist should also inform the patient about existing therapies (Table 11.9-8), which can be done by giving the patient the APA booklet *Treatment Works*.

Therapy	Rate (%)
Self-quit	5
Self-help books	10
Physician advice	10
Over-the-counter patch or gum	15
Medication plus advice	20
Behavior therapy alone	20
Medication plus group therapy	30

Table 11.9-7 Typical Quit Rates of Common Therapies

Psychosocial Therapy
Behavior therapy
Pharmacological Therapies
Nicotine Gum
Nicotine Patch
Nicotine gum + patch
Nicotine nasal spray
Nicotine inhaler
Bupropion
Bupropion + nicotine patch
Clonidine*
Nortyptiline*

* Not an FDA-approved use

Table 11.9-8 Scientifically-Proven Treatments for Smoking

Ms. H. was a 55-year-old patient with schizophrenia who smoked 35 cigarettes a day and smoked each cigarette very intensely. She began her cigarette use at around the age of 20 during the prodromal stages of her first psychotic break. Over the next 35 years she had several psychotic breaks and was treated with conventional antipsychotic agents. During the first 30 years of treatment, no psychiatrist or physician advised her to stop smoking, largely because they believed she could not stop.

At age 53 she was diagnosed with diabetes and early ischemic heart disease. At that time her primary care physician recommended smoking cessation. The patient attempted to stop on her own but lasted only 48 hours, partly because her housemates and friends smoked. After a second failure on her own, she became discouraged and concluded that she could not stop smoking. During a routine medication check, her psychiatrist recommended that she stop smoking and the patient described her prior attempts. The psychiatrist and the patient discussed ways to avoid smokers and had the patient announce her intent to quit and request her friends to try not to smoke around her and to offer encouragement for her attempt to quit. The psychiatrist also noted that she became irritable, slightly depressed, restless, and insomniacal during prior cessation attempts, and thus recommended medications and used the APA brochure *Treatment Works* to help the patient decide which medication was best for her. The patient chose a nicotine patch.

The psychiatrist had the patient call 2 days after her quit attempt. At this point the patient stated that the patch and gum were helping but she still felt “left out” when her friends smoked and talked. One week later the patient returned after having relapsed back to smoking. The psychiatrist praised the patient for not smoking for 4 days. He suggested that the patient contact him again if she wished to try to stop again. Seven months later during another medication check the psychiatrist again asked the patient to consider cessation but she was reluctant.

Two months later the patient called and said she wished to try again. She met with the psychiatrist and this time the psychiatrist and patient listed several activities the patient could do to avoid being around friends who smoked, phoned the patient's boyfriend to ask him to assist the patient in stopping, asked the nurses on the inpatient ward to call the patient to encourage her, plus decided to enrol the patient in a support group for the next 4 weeks. The nicotine patch was used again but this time, the nonnicotine medication buproprion (Zyban) was added. The patient was followed with 15-minute visits for each of the first 3 weeks and two phone calls thereafter. She had two “slips” when she became angry with her boyfriend but did not go back to smoking and remained an exsmoker. An unexpected result of her successful cessation was an improvement in the therapeutic alliance between the patient and her psychiatrist.

Discussion Most psychiatrists fail to diagnose and treat nicotine dependence. Unfortunately, sometimes psychiatrists find that although they have adequately treated the disorder that brought the patient into treatment, the patient has such morbidity or mortality from smoking that he or she cannot reap the benefits of psychiatric treatment. Ms. H.'s psychiatrist was correct in using pragmatic plans to help the patient overcome specific problems, in following the patient with short visits or phone calls, and in recommending nicotine replacement. The total amount of time spent with the patient on smoking was about 3 hours. Although this was not reimbursed, the psychiatrist knew that his intervention was an important contribution to the patient's health and was unlikely to be given by other care providers.

Psychosocial Therapies Behavior therapy is the most widely accepted and well-proven psychological therapy for smoking. Behavior therapy consists of several techniques, three of which are supported by good evidence to support them. *Skills training* and *relapse prevention* identifies high-risk situations and plans and practices behavioral or cognitive coping skills for these situations. *Stimulus control* refers to eliminating cues for smoking in the environment. *Rapid smoking* has smokers repeatedly smoke to the point of nausea, etc. in sessions to associate smoking with unpleasant rather than pleasant sensations. This last therapy appears

effective but requires a good therapeutic alliance and patient compliance.

Often behavior therapy is conducted in group settings to promote social support and save on costs. Typically, several sessions are conducted prior to the quit date and several are scheduled for early after the quit date. Although group therapy for alcohol dependence is widely accepted, most smokers believe that they should be able to quit without help. Even those who seek treatment find that smoking cessation treatments are generally not reimbursable, only voluntary agencies and wellness programs offer treatment and, in most locales, treatment is offered only a few times a year. In addition, these programs are only in larger cities and require patient payment. As a result of these barriers, less than 5 percent of smokers enrol in behavior therapies.

Although written materials are the most common self-help format, telephone systems that tailor treatment to patient concerns have recently been found to be more effective and more acceptable. These systems are available free of charge from pharmaceutical firms and some state health departments and industries.

Psychopharmacological Therapies

Pharmacotherapy The AHCPR and the APA guidelines recommends that all smokers be offered medication to aid in smoking cessation. All the medications approved by the Food and Drug Administration (FDA) for smoking cessation appear equally effective, and they all double the quit rate (similar to results obtained with behavior therapy) (Table 11.9-7). With all these medications few produce side-effects; less than 5 percent of patients stop medication due to adverse effects. Although there are many theories, there are no empirically-verified methods to match smokers to specific therapies. Most clinicians believe that patients should be informed of the various therapies and allowed to choose the therapy they believe will be most helpful. In other substance use disorders, psychosocial therapy is thought to be essential and pharmacotherapy is used as an aid to psychosocial therapy. In nicotine dependence the opposite is true: pharmacotherapy is the treatment to be offered to all, and psychosocial therapy should be added when desirable and feasible.

Nicotine Replacement Therapies All nicotine replacement therapies double cessation rates, presumably because they reduce nicotine withdrawal. These therapies can also be used to reduce withdrawal in patients on smoke-free wards. Replacement therapies use a short period of maintenance (6 to 12 weeks) often followed a gradual reduction period (6 to 12 weeks).

Nicotine gum (Nicorette) is an over-the-counter product that releases nicotine via chewing and buccal absorption. A 2-mg (for <25 cigarettes/day smokers) and a 4-mg gum variety (for ³25 cigarettes/day smokers) are available. Smokers are to use 1 to 2 pieces of gum per hour after abrupt cessation. Venous blood concentrations from the gum are $^{1}/_{3}$ to ½ that of between-cigarette levels. Acidic beverages (coffee, tea, soda, and juice) should not be used before, during, or after gum use because these decrease absorption. Compliance with the gum has often been a problem. Adverse effects are minor and include bad taste and sore jaws. About 20 percent of those who quit use the gum for long periods but 2 percent use gum for longer than a year, plus long-term use does not appear to be harmful. The major advantage of nicotine gum is its ability to provide relief in high-risk situations.

Nicotine patches, also sold over the counter are available in a 16-hour no-taper preparation (Habitrol) and a 24- or 16-hour tapering preparation (Nicoderm CQ). Patches are administered each morning and produce blood concentrations of about half those of smoking. Compliance is high and the only major adverse effects are rashes and, with 24-hour wear, insomnia. Long-term use does not occur. Using gum and patches in high-risk situations increases quit rates by another 5 to 10 percent. Whether 24-hour versus 16-hour or taper versus no-taper patches are better has not been tested.

Nicotine nasal spray (Nicotrol), available only by prescription, produces blood concentrations more similar to a cigarette and appears to be especially helpful for heavily dependent smokers. However, the spray causes rhinitis, watering eyes, and coughing in over 70 percent of patients. Although there was initial data suggesting abuse liability, further trials have not found this.

Nicotine inhaler a prescription product, was designed to deliver nicotine to the lungs but the nicotine is actually absorbed in the upper throat. Resultant nicotine levels are low. The major asset of the inhaler is that it provides a behavioral substitute for smoking. The inhaler also provides double quit rates. These devices require frequent puffing, which can cause minor adverse effects.

Nonnicotine Medications Nonnicotine medications may be helpful to smokers who object philosophically to the notion of replacement therapy and to smokers who fail replacement therapy. Buproprion (Zyban) (marketed as Wellbutrin for depression) is an antidepressant medication that has both dopaminergic and adrenergic actions. Dosages of 300 mg a day reliably double quit rates in smokers with and without a history of depression. In one study, combined buproprion and nicotine patch had higher quit rates than either alone. Adverse effects include insomnia and nausea, but these are rarely significant. Seizures have not occurred in smoking trials. Interestingly, nortriptyline (Pamelor) appears to be effective for smoking cessation.

Clonidine (Catapres) decreases sympathetic activity from the locus ceruleus and is thereby thought to abate withdrawal symptoms. Whether given as a patch or orally, 0.2 to 0.4 mg a day of clonidine appears to double quit rates; however, the scientific database for the efficacy of clonidine is not as extensive nor as reliable as that for nicotine replacement; also, clonidine can cause drowsiness and hypotension.

Combined Psychosocial and Pharmacological Therapy Several studies have shown that combining nicotine replacement and behavior therapy increases quit rates over either therapy alone (Table 11.9-8).

Use in Special Populations Use of nicotine replacement during pregnancy is debatable. On the one hand, the concentrations of nicotine from these medications are much lower than those from smoking, plus no carbon monoxide and tar exposure occurs. However, nicotine has been implicated in low-birth–weight and in sudden infant death syndrome. Nicotine replacement is effective in the elderly without excess complications; it has not been tested in smokers under the age of 18. There are no known effective psychosocial or pharmacological treatments for adolescents.

Smokers who have been recently diagnosed with cancer or heart disease often quit on their own; thus, more intensive therapy probably should be reserved for those who fail; however, recently diagnosed smokers should be followed closely so that such therapy can be offered if they do fail.

Psychiatric patients who are trying to stop smoking should be encouraged to undertake both psychosocial and pharmacological therapies because these patients suffer greater withdrawal symptoms and have a poorer ability to stop. In addition, these patients should be followed weekly to detect any possible exacerbation of their psychiatric problem.

Whether alcoholics should stop smoking along with or soon after stopping drinking is debatable. Current evidence suggests that for the large majority of patients, stopping smoking does not threaten sobriety. Whether smokers with a history of alcoholism or other substance abuse need more intense therapy or different content therapy for smoking cessation is unclear.

Patients on smoke-free psychiatric wards can have withdrawal symptoms that require treatment. Because wards often lack cues for smoking, withdrawal is usually not as problematic as anticipated; thus, treatment should usually be reserved until the emergence of symptoms. If the smoker will be taking passes and smoking, nicotine gum is preferred because it can be used intermittently. If the smoker will be confined to the ward or is using the hospitalization to stop smoking, nicotine patches should be used because they are associated with improved compliance. All smokers should be told that the rationale for the smoke-free ward is not a moral one, but rather to protect nonsmokers from second-hand smoke and to prevent tempting exsmokers under stress from returning to smoking.

SUGGESTED CROSS-REFERENCES

Section 11.1 provides an overview of substance abuse and substance-related disorders. Section 11.2 is devoted to alcohol-related disorders; Chapter 20 discussed eating disorders, and Chapter 21 on sleep disorders and Chapter 39 on attention-deficit disorders are of related interest. Anxiety disorders are the subject of Chapter 15, and mood disorders are presented in Chapter 14. Bupropion is discussed in Section 31.11.

SECTION REFERENCES

Benowitz NL: Pharmacologic aspects of cigarette smoking and nicotine addiction. New Engl J Med *319:*1318, 1988.

Dani JA, Heinemann S: Molecular and cellular aspects of nicotine abuse. Neuron *16:*905, 1996.

Escobedo LG, Reddy M, Giovino GA: The relationship between depressive symptoms and cigarette smoking U.S. adolescents. Addiction *93:*433, 1998.

Fiore MC: Cigarette smoking. Med Clin North Am *76:*289, 1992.

Fiore MC, Jorenby DE, Baker TB, Kenford SL: Tobacco dependence and the nicotine patch: Clinical guidelines for effective use. JAMA *268:*2687, 1992.

Fowler JS, Volkow ND, Wang GJ, Pappas N, Logan J, MacGregor R, Alexoff D, Shea C, Schiyer D, Wolf AP, Warner D, Zezulkova I, Cilento R: Inhibition of monoamine oxidase B in the brains of smokers. Nature *379:*733, 1996.

Giovino GA, Henningfield JE, Tomar SL, Escobedo LG, Slade J: Epidemiology of tobacco use and dependence. Epidemiol Rev *17:*48, 1995.

Glassman AH: Cigarette smoking: Implications for psychiatric illness. Am J Psychiatry *150:*546, 1993.

*Glynn TJ, Manley MW: *How To Help Your Patients Stop Smoking*. US Government Printing Office, Washington, DC, 1989.

Heath AC, Madden PAF: Genetic influences on smoking behavior. In *Behavior Genetic Approaches in Behavioral Medicine*, JR Turner, LR Cardon, JK Hewitt, editors. Plenum Press, New York, 1995.

Henningfield JE: Nicotine medications for smoking cessation. N Engl J Med *333:*196, 1995.

*Hughes JR: Possible effects of smoke-free inpatient units on psychiatric diagnosis and treatment. J Clin Psychiatry *54:*109, 1993.

Hughes JR: Overview of nicotine use disorders for alcohol/drug abuse clinicians. Am J Addiction *5:*262, 1996.

*Hughes JR, Fiester S, Goldstein MG, Resnick MP, Rock N, Ziedonis D: American Psychiatric Association practice guidelines for the treatment of nicotine dependence. Am J Psychiatry *153:*S1, 1996.

*Hughes JR, Goldstein MG, Hurt RD, Shiffman S: Recent advances in pharmacotherapy of smoking. JAMA *281:*72, 1999.

Hughes JR, Zarin DA, Pincus HA: Treating nicotine dependence in mental health settings. J Pract Psychictris Behav Health *24:*250, 1997.

*Jorenby DE, Leischow SJ, Nides MA, Rennard SI, Johnston JA, Hughes AR, Smith SS, Muramoto ML, Daughton DM, Doan K, Fiore MC, Baker TB: A controlled trial of sustained-release bupropion, a nicotine patch, or both for smoking cessation. N Engl J Med *340:*685, 1999.

Nicotine Dependence Task Force: Position statement on nicotine dependence. Am J Psychiatry *152:*481, 1995.

*Orleans CT, Slade J: *Nicotine Addiction: Principles and Management*. Oxford University Press, New York, 1993.

Prochaska JO, Goldstein MG: Process of smoking cessation: Implications for clinicians. Clin Chest Med *12:*727, 1991.

Smoking Cessation Clinical Practice Guideline Panel and Staff. The Agency for Health Care Policy and Research Smoking Cessation Clinical Practice Guideline. JAMA *275:*1270, 1996.

United States Department of Health and Human Services: *The Health Consequences of Smoking: Nicotine Addiction: Report of the U.S. Surgeon General*. U.S. Government Printing Office, Washington, DC, 1988.

United States Department of Health and Human Services. *Preventing Tobacco Use Among Young People: A Report of the Surgeon General*. U.S. Government Printing Office, Washington, DC, 1994.

11.10 OPIOID-RELATED DISORDERS

JEROME H. JAFFE, M.D. AND ARI B. JAFFE, M.D.

Throughout the world more than 20 chemically distinct opioid drugs are in clinical use. In the developed countries the opioid drug most frequently associated with abuse and dependence is heroin—a drug that is not used for therapeutic purposes in the United States. Dependence on opioid drugs other than heroin is seen mostly in persons who have become dependent in the course of medical treatment, in health care professionals who have easy access to opioids, and in persons who use drugs that have been diverted from treatment programs. Virtually all the opioid dependence and abuse seen clinically is associated with prototypical μ-agonist opioids. All the μ-agonists produce similar subjective effects. However, the patterns of opioid use and some aspects of toxicity are powerfully influenced by the route of administration and the metabolism of the specific opioid, as well as by the social conditions that determine its price and purity and the sanctions attached to nonmedical use.

DEFINITIONS

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) divides *opioid-related disorders* into *opioid-use disorders* (opioid abuse and opioid dependence) and nine other *opioid-induced disorders* (e.g., intoxication, withdrawal).

Opioid dependence is a cluster of physiological, behavioral, and cognitive symptoms, which taken together indicates repeated and continuing use of opioid drugs despite significant problems related to such use. Drug dependence in general has also been defined by the World Health Organization (WHO) as a syndrome in which the use of a drug or class of drugs takes on a much higher priority for a given person than other behaviors that once had a higher value. These brief definitions each have as their central features an emphasis on the drug-using behavior itself, its maladaptive nature, and on how the choice to engage in that behavior has shifted and becomes constrained as a result of interaction with the drug over time.

Opioid abuse is a term used to designate a pattern of maladaptive use of an opioid drug leading to clinically significant impairment or distress and occurring within a 12-month period, but one in which the symptoms have never met the criteria for opioid dependence.

The *opioid-induced disorders* as defined by DSM-IV include such common phenomena as opioid intoxication, opioid withdrawal, opioid-induced sleep disorder, and opioid-induced sexual dysfunction. Opioid intoxication delirium is occasionally seen in hospitalized patients. Opioid-induced psychotic disorder, opioid-induced mood disorder, and opioid-induced anxiety disorder, by contrast, are quite uncommon with μ-agonist opioids, but have been seen with certain mixed agonist-antagonist opioids acting at other receptors. DSM-IV also includes *opioid-related disorder not otherwise specified* for situations that do not meet the criteria for any of the other opioid-related disorders.

HISTORY

Opiates have been used for at least 3500 years, mostly in the form of crude opium or in alcoholic solutions of opium. Morphine was first isolated in 1806 and codeine in 1832. Over the next century, pure morphine and codeine gradually replaced crude opium for medicinal purposes, although nonmedical use of opium (as for smoking) still persists in some parts of the world. The first semisynthetic opium derivative—diacetylmorphine or heroin—was introduced into medicine in 1898. The first purely synthetic drugs with morphine-like opioids, meperidine (Demerol) and methadone (Dolophine), were introduced into medical practice in the 1940s. The term “opioid” was coined to include the naturally occurring drugs derived from opium (morphine and codeine), the semisynthetic drugs produced from opium derivatives, and a wider range of totally synthetic agents bearing little chemical resemblance to morphine.

Opioid dependence, or at least opioid withdrawal, was first recognized in 1700. Opioid dependence was common by the middle of the nineteenth century, but it was not until later in the century that it came to be seen as an important medical problem. There was public concern about opium smoking by Chinese immigrants and the severe forms of dependence associated with the newly introduced hypodermic needle and syringe. Growing awareness of the problems created by opioid-containing patent medicines sold over the counter or casually dispensed by practitioners with minimal training generated media attention and public debate. In the United States this debate, combined with international political considerations, led to legislation at the state and federal levels restricting opioid use to medically recognized purposes and requiring legitimate prescriptions for most use. The Harrison Act of 1914 had a profound impact on persons who were already addicted because it was interpreted to exclude the provision of opioids to addicts as a legitimate use. Clinics that had been established specifically to provide morphine to addicts were closed, the last in 1923. Doctors were encouraged to avoid opioid addicts entirely. Detoxification efforts were a disappointment to both physicians and patients because relapse was typical. An illicit traffic in opioids arose, providing drugs (mostly morphine and heroin) to persons who could no longer get them through medical channels. Addicts were arrested both for possession and sale. In the early 1930s the United States Public Health Service established two federal hospitals—at Lexington, Kentucky, and Fort Worth, Texas—to deal with the growing number of federal prisoners who were addicted. These hospitals provided long-term residential treatment for prisoners and for some voluntary patients. Later follow-up studies found that relapse rates were very high despite long periods of treatment. Although increasingly harsh penalties for the sale or possession of opioids were enacted, hero in addiction persisted and its prevalence rose following World War II. By the early 1960s some thoughtful observers recommended remedicalizing heroin distribution as a way to reduce crime associated with heroin addiction.

However, several new developments in treatment techniques sharply altered the general perception of opioid dependence as an essentially untreatable disorder. They included the development of therapeutic communities based on the Synanon model (an organization that began in California in 1958); the creation of large-scale civil commitment programs in California (1961) and New York (1965); demonstration of the effectiveness of maintenance on oral methadone in decreasing crime and heroin use; and the development of long-acting opioid antagonists such as cyclazocine, along with the finding that some addicts were willing to take them.

Starting in the late 1960s federal and state sources increased support for research and treatment. In response to the outbreak of heroin addiction among United States military personnel in Vietnam, federal support for treatment was greatly expanded and accelerated in the early 1970s. This support was not merely monetary, but included legislation providing for the legitimate use of methadone and for protection of the confidentiality of patient records.

In the mid-1970s the four dominant treatment modalities were brief detoxification, methadone maintenance, therapeutic communities, and a heterogeneous category generally designated as drug-free outpatient care. Civil commitment declined in influence and support. In the early 1980s, with the expansion of private insurance to cover treatment for alcohol abuse and drug dependence, in-hospital treatment programs based on the 12-step model pioneered in Minnesota proliferated. Although these programs were initially developed as treatments for alcohol abuse, they were gradually broadened to deal with a wider range of dependence, including opioid dependence. For a brief period, “chemical dependence” programs became an additional significant option in the array of treatments available to persons who were dependent on opioids. By the late 1980s, however, the rising cost of treatment and growing government deficits stimulated the emergence of managed care. The impact of managed care on the support for the treatment of opioid dependence has not yet been fully felt, but the availability even of short-term hospital-based treatment has virtually disappeared.

COMPARATIVE NOSOLOGY

The DSM-IV criteria for opioid dependence are the same generic criteria as are applied to other psychoactive drugs. The notion of a generic concept of dependence is shared with the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). In the diagnosis of opioid dependence, there generally is a high level of agreement between DSM-IV and ICD-10: they employ similar concepts (the dependence syndrome varying in degree of severity), although the wording of the criteria for determining the presence and severity of the syndrome differs. Both require that three elements of the syndrome occur within a 12-month period.

There is a major difference between DSM-IV and ICD-10 in how substance abuse is defined. ICD-10 does not use the term “abuse.” Instead, it includes a category of harmful use that is substantially different from the concept of abuse in DSM-IV. But the concept of harmful use is limited to physical and mental health (e.g., hepatitis, overdose, skin abscess) and specifically excludes social impairments. ICD-10 states: “Harmful patterns of use are often criticized by others and frequently associated with adverse social consequences of various kinds. The fact that a pattern of use or a particular substance is disapproved of by another person or by the culture, or may have led to socially negative consequences such as arrest or marital arguments is not in itself evidence of harmful use.”

DSM-IV and ICD-10 also have distinctly different coding systems. ICD-10 separates for record-keeping purposes mental and behavioral disorders due to use of opioids from those caused by other categories of drugs. DSM-IV limits the number of distinct drug-induced syndromes that can be recorded (except under the categories *other* and *unspecified*) as disorders induced by opioids.

EPIDEMIOLOGY

Opioid use in the United States experienced a resurgence in the early 1990s, with emergency department visits related to heroin abuse doubling between 1990 and 1995. This increase in heroin use was associated with an increase in heroin purity and a decrease in its street price. Methods of administration other than injecting, such as smoking and snorting, have been increasing in popularity. The number of current heroin users has been questionably estimated to be between 600,000 and 800,000. The number of people estimated to have used heroin at any time in their lives (‘lifetime users') is estimated at approximately 2 million.

There are a number of activities aimed at estimating the extent and consequences of psychoactive drug use in the United States. All the regularly recurring estimating techniques have sampling limitations, and none apply standardized diagnostic criteria to substance-use patterns or adverse effects. Consequently, although they provide a picture of substance use over time, the methods do not reveal changes in the incidence and prevalence of specific substance-related disorders, such as substance dependence and substance abuse.

The annual Monitoring the Future project (“High School Senior Survey”) has found lifetime rates of heroin use increasing significantly from the low levels reported at the start of the 1990s. Lifetime use rates for heroin peaked at 2.8 percent in 1975, fell to 0.9 percent in 1991, and rose again to 2.1 percent in 1997. The lifetime use rate for other opiates (which has always been higher than for heroin) fell from a peak of 10.3 percent in 1977 to 6.6 percent in 1991, and rose again to 9.7 percent in 1997. For both categories of opioids, past-year and past–30-days use rates were substantially lower. For the population as a whole ages 12 and older, the 1997 National Household Survey on Drug Abuse (NHSDA) found the use of heroin within the past year to be 0.3 percent, and use within the past 30 days was 0.2 percent; recent (past 30 days) nonmedical use of analgesics was 0.7 percent.

Two population surveys have been conducted using accepted criteria to measure the extent of drug abuse and dependence: the Epidemiologic Catchment Area (ECA) Study carried out in the early 1980s using criteria from the third edition of DSM (DSM-III), and the National Comorbidity Survey (NCS) carried out from 1990 to 1992 using the revised third edition of DSM (DSM-III-R) criteria. The NCS found the lifetime prevalence of heroin use to be 1.5 percent overall, but with a prevalence of 2.7 percent among 35-year-olds to 44-year-olds, probably reflecting the peak of the heroin epidemic among adolescents and young adults in the late 1960s and early 1970s. Heroin dependence (lifetime) was 0.4 percent overall, but 0.8 percent among 35-year-olds to 44-year-olds. Those findings indicate that about 32 percent of those who used heroin at the peak of the epidemic became dependent at some time in their lives.

Lifetime history of extramedical use of opioid analgesics (other than heroin) was 9.7 percent, with the highest prevalence among 15-year-olds to 34-year-olds, suggesting a different pattern from that of heroin. Overall, only 7.5 percent of those who used opioid analgesics outside of a medical context developed dependence as defined by DSM-III-R. Obviously, the 6-month and current prevalence rates of dependence would be lower than these lifetime rates.

The use and dependence rates derived from national surveys do not accurately reflect fluctuations in drug use among opioid-dependent and previously opioid-dependent populations. When the supply of illicit heroin increases in purity or decreases in price, as seemed to be the case in 1994, use among that very vulnerable population tends to increase, with subsequent increases in adverse consequences (emergency room visits) and requests for treatment.

Recent increases in the purity of street heroin and users' concern about the risks of HIV transmission have led to a resurgence of heroin snorting as well as smoking (“chasing the dragon”). Casual users may persist in such routes of administration; however, with long-term use and increasing pharmacological tolerance, economic imperatives often lead users to switch to the more efficient parenteral routes. Injecting drug users are also more likely to experience a “rush” than those using other routes of administration.

ETIOLOGY

Opioid dependence is currently seen as a biopsychosocial disorder in which multiple factors interact to influence initiation of use, continued use, and relapse after periods of abstinence. Those factors—pharmacological, social, environmental, personality, psychopathology, genetic, and familial—are the same that must be considered when looking at abuse and dependence on other categories of drugs. What changes in the case of the opioids is the balance of the various factors. For example, as reinforcers of drug-taking behavior, benzodiazepine anxiolytic agents are not as powerful as either opioids or cocaine. Consequently, medical attitudes, prescribing patterns, and pre-existing psychopathology are more dominant as causative factors in the dependence on benzodiazepines whereas pharmacological factors (i.e., intense euphorigenic and reinforcing effects, the avoidance of aversive withdrawal states, and perhaps the persistence of some low-level protracted abstinence) play a larger role in dependence on opioids.

For opioids, as for most substances, it is largely social and cultural factors that influence availability and initial use. In the case of opioid drugs, however, pharmacological factors—the initial effects and their consequences—are believed to play very important roles in the perpetuation of use and of progression to dependence. Opioids have potent mood-elevating and euphorigenic actions in humans and are powerful reinforcers in animal models. This is particularly true when the effects are rapid in onset, such as when the opioids are injected or inhaled. Perhaps more than any other category of drugs, the opioids can induce a variety of physical dependence that results in an aversive withdrawal syndrome when brain opioid levels decline. This aversive syndrome appears to play a key role in perpetuating opioid use and in relapse after brief periods of withdrawal.

Pharmacological Factors Because the reinforcing and physical-dependence–inducing effects of opioid drugs play such an important role in the genesis of dependence (as well as in opioid-induced disorders), the pharmacology of the opioids is briefly summarized here.

Opioids and Opioid Receptors The discovery of multiple, stereospecific opioid receptors and endogenous ligands for those receptors made it easier to understand the multiple actions of morphine-like drugs and of some of the drugs that only partially resembled morphine. Those discoveries, however, also made it necessary to redefine the term “opioid.” An opioid agonist is now defined as any exogenous substance that binds specifically to any of several subtypes of opioid receptor and produces some action. While many opioids produce actions very similar to that of morphine (a prototypical μ-agonist), others may bind to various receptor subtypes in a pattern distinct from that of morphine, producing a dissimilar profile of actions, and may not suppress the morphine abstinence syndrome. Drugs that bind to any of the subtypes of receptors but initiate no actions are opioid antagonists at those receptors.

Several opioid receptor subtypes have been described and characterized. These include: μ, where classic opioids such as morphine bind preferentially and produce actions; k, where the endogenous opioid dynorphin exerts some of its effects; and d, which appears to be the preferential binding site for the endogenous pentapeptide met-enkephalin, as well as for several synthetic peptides. With the development of more specific ligands and modern molecular techniques, subtypes of the μ-, d-, and k-receptors have also been identified. These major receptor subtypes have been cloned in the laboratory.

Cloning studies have recently identified a new receptor with a high degree of homology to traditional opioid receptors. An endogenous ligand to this initially “orphan” receptor has been identified and termed orphanin FQ or nociceptin. The peptide has been demonstrated to have both analgesic and pronociceptive effects in vivo,

depending on site of administration; additional effects on the vasculature and fluid homeostasis have also been noted.

The s receptor was named for the benzomorphan derivative SKF-10, 047 which, in dogs induced excitation and hallucinatory effects but little or no analgesia. Because binding to the s receptor is not antagonized by naloxone (Narcan), it is no longer considered an opioid receptor. As is the case with the opioid receptors, several subtypes of s receptor have been identified. Since phencyclidine (PCP) may exert some of its actions through this family of receptors, they are discussed at greater length in connection with that drug.

The actions of μ-agonist opioids are exerted primarily at receptors on neural tissues in the central nervous system (CNS), the autonomic nervous system, and to some degree on opioid receptors on white blood cells. The actions of opioids include analgesia, respiratory depression, changes in mood (euphoria in some persons), indifference to anticipated distress, drowsiness, decreased ability to concentrate, changes in endocrine and other functions regulated by the hypothalamus, and increased tone of smooth muscle in the gastrointestinal tract. μ-Agonists also induce tolerance and neuroadaptive changes in the CNS that result in distressing withdrawal phenomena when the agonist is stopped after days or weeks of continuous use. In contrast, drugs that act at k receptors, such as U-50, 4884, produce some dysphoria and no significant pupillary change but still induce analgesia. Also in contrast to μ-agonists, which produce dopamine release by neurons originating in the ventral tegmental area, k-agonists inhibit dopamine release in this pathway.

Most of the opioid drugs associated with opioid abuse and dependence are typical μ-agonists, having pharmacological profiles that are quite similar to morphine and differing primarily in terms of metabolism and pharmacokinetics. Thus heroin is more potent and more lipid soluble than morphine, thereby crossing the blood-brain barrier more rapidly, producing a more rapid onset of subjective effects and less general pruritus an adverse effect of morphine. However, heroin is hydrolyzed quite rapidly (half-life of about 3 minutes) to 6-monoacetylmorphine and morphine. Its actions are probably exerted primarily through those metabolites binding to μ receptors, although 6-monoacetylmorphine may bind to d receptors as well.

Codeine (3-methoxymorphine) occurs naturally (0.5 percent) in opium. Codeine is probably a prodrug; after absorption, it is transformed to some degree into morphine, which accounts for its opioid effects. This biotransformation is dependent on the cytochrome P450 isoenzyme CYP 2D6. This enzyme is relatively inactive in 7 percent of Causasians, who probably get minimal analgesia from codeine. The activity of CYP 2D6 can also be inhibited by such drugs as quinidine (Cardioquin), fluoxetine (Prozac), and paroxetine (Paxil). Codeine itself does not bind with great affinity to μ receptors, but it does cause some toxicity, and that probably accounts for its relatively low abuse potential. Hydrocodone and oxycodone (Roxicodon) may also be prodrugs that require metabolic conversion for full activity. Methadone appears to be a typical μ-receptor agonist, but with an extended duration of action after repeated administration. Meperidine (Demerol), has numerous μ-receptor actions, but it probably has some actions at other receptors as well and may enhance the actions of dopamine, most likely by a nonopioid action at the dopamine transporter. One of its metabolites, normeperidine, has convulsant properties, and addicts who use excessive amounts may achieve high enough levels of the metabolite to experience delirium and frank seizures. Similar toxicity may be seen in patients receiving meperidine for pain when the excretion of normeperidine is reduced by impaired renal function. In patients who are taking monoamine oxidase inhibitors (MAOIs) meperidine can cause serious toxicity, including fatal hyperthermia.

In the United States most of the μ-agonists drugs are included in either Schedule (Control Level) I (prohibited for clinical use) or Schedule II (clinical use under the most restrictive prescribing controls) of the Controlled Substances Act (CSA), although some mixtures (e.g., codeine and aspirin) are in the less restrictive Schedules III and IV. Buprenorphine (Buprenex), a partial μ-agonist now included in Schedule V, was misused by opioid addicts in Europe when it was introduced there without restrictions. Abuse or dependence is occasionally seen with opioids that are not prototypical μ-agonists but also have actions at other opioid receptors, such as the mixed agonist-antagonists pentazocine (Talwin) and nalbuphine (Nubain). Such agents, if they are included in the CSA, may be in Schedules III, IV, or V.

Opioid receptors are typical G-protein coupled receptors. Receptors can be linked either to second messenger systems or directly to ion channels. μ-Receptors, for example, can act via G_i or G_o to directly increase potassium flux or to inhibit the action of adenylate cyclase. Subsequent decreases in the activity of cyclic adenosine monophosphate (cAMP)-dependent protein kinases can have immediate effects on phosphorylation-dependent cellular proteins as well as long-term effects on gene expression via decreased phosphorylation of cAMP-dependent transcription factors, such as cAMP response element binding protein (CREB).

Several analgesics now available have actions at more than one receptor type. Some have antagonist actions at one type and agonist actions at another. For example, pentazocine has reinforcing properties and is self-administered by animals and some addicted persons, but it does not appear to exhibit a significant degree of cross-tolerance with μ-agonists and does not suppress μ-agonist withdrawal to any significant degree. It may be a very weak μ-agonist and k-agonist.

Buprenorphine is a partial agonist at the μ-receptor. It supports μ-receptor physical dependence when the degree of physical dependence is low to moderate but precipitates withdrawal when the degree of dependence is high. Because most opioid-dependent persons have a low to moderate level of dependence, buprenorphine typically suppresses withdrawal and maintains dependence. It generally produces neither typical μ-agonist effects nor precipitated withdrawal in patients maintained on doses of methadone below 30 mg. Some withdrawal may occur when methadone doses are above 60 mg, but rapid transfer from lower doses to buprenorphine is usually easily accomplished.

Endogenous Opioid Substances Three distinct neurobiological opioid peptide systems or families have now been well described. A recently discovered fourth family (orphanin FQ) is a topic of active research. Each of the three well-described systems has a distinct genetic basis, separate biosynthetic pathways, and distinct precursor molecules. The anatomical distributions of the cells that produce and release the respective endogenous peptides are also distinct, but there is sometimes considerable overlap. These three systems are usually referred to as (1) the pro-opiomelanocortin (POMC) system, (2) the proenkephalin system, and (3) the prodynorphin system. Each precursor protein produces more than one active peptide that can be detected in body tissues.

The POMC precursor molecule is a 265–amino-acid protein that contains the 91 amino acid peptide b-lipotropin (b-LPH), as well as adrenocorticotropin hormone (ACTH) and melanocyte-stimulating hormone (MSH). Peptides 61 to 91 in b-lipotropin make up beta-endorphin, one of the active opioid fragments produced by the POMC family. The enkephalin system consists primarily of the pentapeptides met-enkephalin (tyr-gly-gly-phe-met) and leu-enkephalin (tyr-gly-gly-phe-leu). In proenkephalin, a 263-amino-acid protein, met-enkephalin or met-enkephalin extended peptides are six times as prevalent as leu-enkephalin. Several additional opioid peptides are also derived from proenkephalin. From the parent pro-dynorphin, the dynorphin system produces the 17-amino-acid peptide dynorphin and several other active dynorphin peptides, all of which are C-terminal extensions of leu-enkephalin. They include dynorphin A (1 to 17), dynorphin A (1 to 8), dynorphin B (1 to 13), a-neoendorphin, b-neoendorphin, and others.

The processing of the precursor molecules to smaller peptides is tissue specific. For example, in the rat the POMC precursor is processed to different peptides by the anterior and intermediate lobes of the pituitary.

The various endogenous peptides tend to bind preferentially to one or more of the opioid receptor subtypes. For example, met-enkephalin appears to prefer d-receptors, and most members of the dynorphin family of peptides display their highest affinity for k-receptors. However, b-endorphin binds to both μ-receptors and d-receptors and does not appear to be as selective as the other endogenous ligands are. Preferential binding is not the same as exclusive binding. Peptides that bind preferentially to one set of receptors can, in high enough concentrations, exert actions at receptors for which they have lower affinities. Thus far, no endogenous peptide has been found that binds as preferentially to the μ-receptor as do such drugs as morphine. Several laboratories have confirmed the presence of morphine and codeine in mammalian brain and adrenal gland and in human cerebrospinal fluid (CSF). Highest concentrations are found in adrenal glands and spinal cord. Since no exogenous sources could be identified, the morphine and codeine so identified are thought to be of endogenous origin, although the synthetic pathways and possible biological function of those nonpeptide substances are still uncertain. Other findings not yet fully explored are the isolation from brain of natural cleavage products of beta-endorphin, which are more potent than naloxone as opioid antagonists; the capacity of certain dynorphin fragments to modify opioid withdrawal; and the likelihood that some of those dynorphin fragments act at nonopioid receptors.

Tolerance and Physical Dependence (Opioid Neuroadaptation) Tolerance does not develop uniformly to all the actions of opioid drugs. There can be high levels of tolerance to some actions of opioids (such that it requires a hundredfold increase in dose to produce the original effect) when responses to other drug actions show only modest tolerance. With μ-agonist opioids there can be remarkable tolerance to their analgesic, respiratory-depressant, and sedative actions. Patients in Swiss heroin clinics have injected as much as 300 mg intravenously as a single dose. Markedly less tolerance develops to the miotic and constipating actions of opioids on the bowel. Intermediate degrees of tolerance to endocrine actions develop, and there appears to be less tolerance to the capacity of opioids to lower the threshold for electrical self-stimulation of the brain. Opioids occupying the same receptor types exhibit a considerable degree of cross-tolerance, but tolerance and physical dependence appear to be specific for each major receptor type.

When tolerance to a given action develops to a μ-agonist, such as morphine, some cross-tolerance will be seen with other μ-agonists. However, when tolerance

develops to a selective k-agonist, such as the investigational drug U-50,488, there is no cross-tolerance to μ-agonists. Similarly, opioid withdrawal phenomena can be suppressed by any opioid that acts at the same receptor, but not by one that acts at another receptor; k-agonists do not suppress withdrawal from μ–agonist-induced physical dependence. Furthermore, physical dependence induced by k-agonists has distinct characteristics and a different pattern of withdrawal signs and symptoms.

Physical (Physiological) Dependence Physical dependence is a substance-induced change in a biological system that becomes manifest by a characteristic response pattern, the withdrawal syndrome, when the drug is removed from the body or displaced from its receptor. In general, responses are opposite in direction to the acute agonistic effects of the drugs, that is, they are rebound hyperexcitabilities. The neuroadaptive changes induced by the repeated administration of opioids occur in cells bearing opioid receptors and in neural systems widespread throughout the organism. In humans the changes begin with the first few doses. For example, if single intramuscular doses of 15 to 18 mg of morphine are given to opioid naive subjects or abstinent former opioid users, large doses of naloxone (10 to 30 mg) given within 24 hours will precipitate a mild μ-agonist withdrawal syndrome. However, some period of continuous receptor occupation is required before the syndrome produced by abrupt cessation reaches an intensity level high enough to be obvious to the clinical observer. Withdrawal phenomena are more intense and more readily detectable when the opioid is rapidly removed from its receptor, as happens with opioid antagonist administration.

A distinction between opioids and other classes of pharmacological agents is that physical dependence not only develops rapidly, but there also appears to be a protracted period of physiological abnormality that follows the acute opioid withdrawal syndrome. This long-lasting *protracted abstinence* syndrome, characterized by hypophoria, irritability, mood instability, and recurrent urges to use opioids, is postulated to result from opioid-induced alterations in endogenous opioid peptide systems, opioid receptors, or other intracellular proteins that are produced in neurons exposed to opioids over long periods. However, long-lasting neuronal changes can also result from indirect actions of opioids. A reduction in the size of dopaminergic neurons in the ventral tegmental area (VTA) has been observed in animals administered morphine over long periods. It is postulated that the protracted abstinence syndrome is responsible for the high rate of relapse observed when well-motivated patients on high doses of methadone are withdrawn from the drug.

Mechanisms of Tolerance and Physical Dependence A variety of mechanisms have been put forth to account for the general observation that opioid tolerance and physical dependence tend to develop in parallel and that withdrawal phenomena tend to be opposite in direction to the acute effects produced by the drugs. Among the mechanisms that have been proposed are variations in receptor number or affinity, alterations in intracellular second messengers and ion concentrations, and changes in the levels of endogenous opioid peptide agonists and antagonists. These are not mutually exclusive but appear to be complementary, and supportive data exist for several of the mechanisms. Opioid drugs can alter the expression of the genes encoding the opioid neurotransmitters. For example, the repeated administration of morphine results in downregulation of proenkephalin expression in striatal neurons whereas administration of opioid antagonists up-regulates proenkephalin expression. There is also considerable evidence that the cAMP second messenger system plays an important role in the development of opioid tolerance. Acute administration of μ or d opioids results in inhibition of adenylyl cyclase and a decrease in cAMP concentrations. With chronic administration, there appears to be a compensatory increase in the expression of adenylate cyclase and the cAMP-dependent protein kinase probably mediated by effects on cAMP-responsive transcription factors such as CREB. When the opioid is removed, these changes result in transiently higher cAMP concentrations. Some aspects of withdrawal, such as excess sympathetic nervous system activation, have been attributed to the hyperexcitability that results from such an upregulation of the cAMP system in the noradrenergic neurons of the locus coeruleus. These neurons express opioid receptors and their activity is inhibited by opioid administration. Other effects of repeated opioid administration may be attributable to the functional decoupling of opioid receptors from their mediating G-proteins. Long-term administration of opioids can upregulate G protein receptor kinases, which in turn can produce a densensitization of opioid receptors (Fig. 11.10-1).

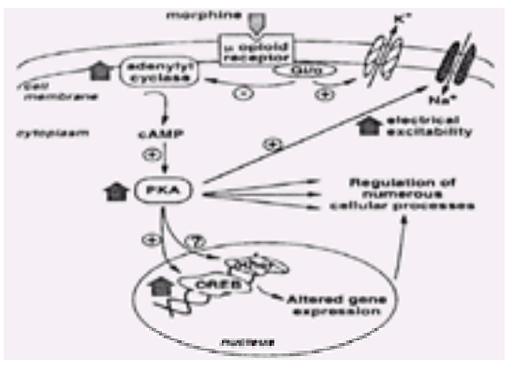

FIGURE 11.10-1 Scheme illustrating opioid actions in the locus coeruleus (LC). Opiates acutely inhibit LC neurons by increasing the conductance of a K^+ channel (light cross-hatch) via coupling with subtypes of G_i and/or G_o and by decreasing a Na^+-dependent inward current (dark cross-hatch) via coupling with $G_{i/o}$ and the consequent inhibition of adenylyl cyclase. Reduced levels of cAMP decrease PKA activity and the phosphorylation of the responsible channel or pump. Inhibition of the cAMP pathway also decreases phosphorylation of numerous other proteins and thereby affects many additional processes in the neuron. For example, it reduces the phosphorylation state of CREB, which may initiate some of the longer-term changes in LC function. Upward bold arrows summarize effects of repeated morphine administration in the LC. Repeated morphine administration increases levels of adenylyl cyclase, PKA, and several phosphoproteins, including CREB. These changes contribute to the altered phenotype of the drug-addicted state. For example, the intrinsic excitability of LC neurons is increased via enhanced activity of the cAMP pathway and Na^+-dependent inward current, which contributes to the tolerance, dependence, and withdrawal exhibited by these neurons. This altered phenotypic state appears to be maintained in part by upregulation of CREB expression. (Reprinted with permission from Nestler EJ: Molecular mechanisms underlying opiate addiction: Implications for medications development. Semin Neurosci *9:*84, 1997.)

Chronic treatment with opioids induces supersensitivity in several distinct transmitter circuits, including the dopaminergic, noradrenergic, cholinergic, and serotonergic systems. The effects of opioids on the noradrenergic neurons of the locus coeruleus have already been noted. The observation that certain a_2-adrenergic agonists such as clonidine (Catapres) can, like opioids inhibit the activity of neurons in the locus ceruleus formed the background for the clinical trials of clonidine in opioid withdrawal. The finding that *N*-methyl-D-aspartate (NMDA) antagonists can suppress a number of opioid withdrawal symptoms has generated interest in the interaction between opioid neurotransmitter systems and the glutaminergic system. NMDA antagonists, most of which are limited in their clinical utility by their phencyclidine-like side-effects, appear to attenuate changes induced by opioid withdrawal in the amygdala and basal forebrain, but not in the locus coeruleus. Further, NMDA antagonists as well as agents that inhibit nitric oxide synthase appear to retard the development of opioid tolerance. It has been suggested that D-methadone, which is not active as a μ-agonist, actually delays the development of tolerance to the active levorotatory isomer component of racemic methadone by actions at the NMDA receptor. Dextromethorphan, commonly used as a cough suppressant, also acts at the NMDA receptor. It has been combined with morphine for use as an analgesic in clinical studies. The mixture enhances the analgesic but not the euphorigenic effects of morphine.

Opioids as Reinforcers In laboratory experiments animals will self-administer μ-opioid and d-opioid agonists, but not k-agonists. Former heroin addicts given opioids report reduced anxiety, increased self-esteem, a better ability to cope with everyday problems, and a decreased sense of boredom. Given intravenously, opioids produce a “rush” or “flash,” a sudden, brief sensation that is reported to be exceedingly pleasurable. Although the rush was customarily described as being much like an orgasmic sensation felt in the abdomen, in more recent interviews addicts have described it in much more varied terms, although still as a much desired experience. The rush, a far shorter phenomenon than a general sense of euphoria, lasts only a minute or two, and is experienced only with rapid drug intake, as with intravenous or intrapulmonary routes. Heroin addicts who self-administered heroin in a research setting seemed to develop tolerance to the anxiety-relieving and mood-elevating effects of opioids and, over a period of several weeks, developed various somatic complaints and reported feeling increasingly anxious and dysphoric. Nevertheless, they were able to experience brief periods of mood elevation for 30 to 60 minutes each time they received single injections. The loss of mood-elevating effects and the appearance of hypophoria and hypochondriasis have also been observed with the long-term administration of methadone in a research setting.

Tolerant opioid users do not continue to self-administer opioids solely to prevent the highly aversive withdrawal phenomena. Interviews with heroin users indicated that they continue to experience a brief euphoric effect immediately following an injection, despite some tolerance to many of the drug effects.

Opioids are synergistic with either amphetamine or cocaine in lowering self-stimulation thresholds in animals and in inducing euphoria in humans. The neurobiological mechanisms of reinforcement of these drugs also appear to overlap. The mesolimbic dopaminergic system, which is thought to play a role in regulating responses to natural forces such as sex and food, is affected by both acute and repeated drug administration. Acutely, opioids increase dopaminergic signals along this pathway from the ventral tegmental area to the nucleus accumbens, in the basal forebrain. This is an indirect action that results from opioid inhibition of GABA-ergic neurons,

which normally act to inhibit the dopaminergic neurons in the ventral tegmental area; opioids can also act directly on neurons in the nucleus accumbens.

Repeated administration of opioids can upregulate the cAMP system in the nucleus accumbens and ventral tegmental area, just as has been described for the locus coeruleus. Just as rebound effects of the cAMP system in the locus coeruleus have been linked to some of the physical symptoms of withdrawal, analogous effects in the nucleus accumbens and ventral tegmental area have been hypothesized to play a role in the aversive motivational states (dysphoria, anhedonia) associated with the withdrawal of drugs. Further evidence for the convergence of drug reinforcement systems is provided by studies of brain glucose metabolism and cerebral blood flow in human volunteers. These studies have found striking similarities in the way that different drugs that induce euphoria alter brain function. When subjects experienced euphoria induced by either opioids or cocaine, euphoria was correlated with decreased brain glucose metabolism, especially in orbito-frontal cortex and thalamus.

Psychological and Social Factors Social attitudes, peer pressure, and drug availability are the major determinants of experimentation with the less socially disapproved drugs. Generally, the use of tobacco, alcohol, and cannabis precedes the use of cocaine and opioids. Since in most cases the earlier substance use continues, most opioid or cocaine users are really multiple drug users. The presence of a significant number of opioid users in a neighborhood, town, or city creates a subculture supportive of experimentation, and continuing use of the drug usually follows. In the United States, areas where availability of illicit opioids is high also have high crime rates, high unemployment rates, and demoralized school systems. These factors all contribute to the sense of hopelessness and the low self-esteem that reduce resistance to drug use and militate against a good prognosis once dependence develops. Along with family factors, these factors may contribute to disproportionately high rates of heroin addiction currently seen among African-American and Hispanic minorities.

A unique natural experiment in which the influence of vulnerability, social norms, and particularly of availability could be observed, was provided by the experience of U.S. service personnel who used heroin in Vietnam. During the period between 1970 and 1972, high-grade, low-cost heroin was readily available to young men separated from their families and usual social norms. Among U.S. Army enlisted men, about 42 percent tried heroin and about half of those who did became physically dependent. Very few of these men, however, continued to use heroin after their return to civilian life. The importance of availability in the genesis of addiction to opioids is also illustrated by the high rates of opioid addiction among physicians, dentists, and nurses compared to other professionals with comparable educational achievement but less easy access to opioids.

Psychodynamic Factors and Psychopathology The psychodynamic perspective is that psychopathology is the underlying motivation for initial drug use, drug dependence, and relapse after a period of abstinence. Recent psychoanalytic formulations postulate ego defects, which are manifest in the addict's inability to manage painful affects (guilt, anger, anxiety), and to avoid preventable medical, legal, and financial problems. Some addicts also appear to have great difficulty in differentiating and describing what they feel, a difficulty that has been aptly called *alexithymia* (i.e., no words for feelings). It is postulated that both pharmacologically and symbolically opioid use helps the ego control those affects and that drug use can be viewed as a form of self-medication.

Epidemiological studies find that persons who use illicit drugs, especially those who use opioids, tend to place more value on independence and less on academic achievement. They are also more tolerant of deviance, and a very substantial number showed significant signs of delinquency before their first experimentation with opioids. A significant proportion of opioid users meet the criteria for antisocial personality disorder, even when those items that are related to illicit drug use are not applied. In Vietnam, where heroin was readily available, the use of heroin by American service personnel was powerfully predicted by the same factors that predicted use in the United States: the frequency and severity of preservice fighting, truancy, drunkenness, arrest, and school expulsion. The ECA Study and the National Comorbidity Study both found that persons with a diagnosis of either drug dependence or alcohol abuse were much more likely than were those without such a diagnosis to also have at least one other mental disorder that was present in the absence of drugs or alcohol. The rates of coexistent psychiatric disorders and substance abuse and findings of high rates of coexistent psychiatric disorders among opioid users seeking treatment do not prove causality; drug use could and does increase the risk of psychopathology. However, these data strongly support the argument that many persons will be left at higher risk for initial treatment failure or relapse if treatment efforts are aimed solely at the drug-using behavior itself and do not address comorbidity.

Family Factors More than 50 percent of urban heroin addicts come from single-parent families, and poor parental functioning has been consistently reported as a risk factor for opioid addiction. A retrospective study of urban males in an East Coast city compared heroin addicts, never-addicted peer controls who were associates of the addicts at age 11, and never-addicted members of the community who were not associates of the addicts. Family risk factors involving deviant behaviors among family members and disruption of family structure (biological parents who had never lived together or were divorced, separated, or died before the subject reached 11 years of age) was 50 percent higher among the heroin addicts before they were age 11.

Typically, even in two-parent families there are disturbed family relationships, with one parent, usually of the opposite sex, intensely involved with the addict and the other parent distant, absent, or punitive. Cross-generational alliances between the drug user and one parent against another parental figure are common. The disability of the drug-using member of the family often serves as a focus for communication among other members, and sometimes may be the main motive for their remaining together; thus, the family equilibrium may be threatened by the addict's recovery.

Despite their seeming rebelliousness and precocious efforts to be independent, opioid users often remain dependent on and in close communication with families of origin well into adulthood. Interestingly, both male and female heroin addicts believe that members of their families of origin or their in-laws would be the most helpful to them in their efforts to give up drugs. Rates of alcohol and drug abuse, mental illness, and antisocial personality disorder are higher in the families of heroin users. Relatives of depressed opioid addicts tend to have higher rates of depression and anxiety, but not of other disorders.

Other Factors It has been postulated that some antecedent metabolic deficiency, such as endogenous opioid dysregulation or one induced by long-term opioid use, may increase vulnerability to becoming opioid dependent but no evidence for such defects has yet emerged. There is now evidence for common and drug-specific, genetically transmitted, vulnerability factors that increase the likelihood of developing drug dependence. Individuals who abuse a substance from any category are more likely to abuse substances from other categories. The common vulnerability factor is influenced by family and nonfamily environmental factors. In a study of Vietnam-era veterans, monozygotic twins were more likely than dizygotic twins to be concordant for opioid dependence. Multivariate modeling techniques indicated that not only was the genetic contribution high for heroin abuse in this group, but also a higher proportion of the variance due to genetic factors was not shared with the common vulnerability factor—that is, it was specific for opioids.

Learning and Conditioning Opioids are positive reinforcers of drug self-administration. They can also reinforce drug-seeking behavior by terminating noxious or aversive states, such as pain, anxiety, or depression (negative reinforcement). In some social situations the use of the drug can also be reinforcing if it results in gaining special status among friends, and such social reinforcement can serve to maintain drug use until the effects of primary reinforcement or reinforcement by alleviation of withdrawal symptoms come into play. Typically, each time the drug is used, reinforcement occurs. It may be the rush, drug-induced euphoria, alleviation of disturbed affect or of withdrawal symptoms, or any combination of those effects. Use of a short-acting opioid, such as heroin, causes such reinforcement to occur several times a day, day after day, creating a powerfully reinforced habit pattern. Eventually the paraphernalia and hustling associated with drug use can become secondary reinforcers as well as cues that signal drug availability, and in their presence craving or desire to experience drug effects increases.

In addition to this operant reinforcement of drug-using and drug-seeking behaviors, classical or Pavlovian conditioning probably plays a role in relapse. In both laboratory animals and human volunteers, opioid withdrawal phenomena can be conditioned to environmental or interoceptive stimuli. Animals that exhibit marked tolerance to the effects of opioids given repeatedly in one situation may exhibit toxicity when the same dose is administered in a novel environment. For long periods following withdrawal, former opioid addicts may experience conditioned withdrawal or conditioned craving when exposed to environmental stimuli previously linked to drug use or withdrawal. Conditions such as watching someone else use heroin or being offered a drug by a friend, rather than conditions associated with withdrawal, elicit the most intense craving.

DIAGNOSIS AND CLINICAL FEATURES

Table 11.10-1 lists the DSM-IV opioid-related disorders.

Opioid use disorders
Opioid dependence
Opioid abuse
Opioid-induced disorders
Opioid intoxication
Specify if: With perceptual disturbances
Opioid withdrawal
Opioid intoxication delirium
Opioid-induced psychotic disorder, with delusions
Specify if: With onset during intoxication
Opioid-induced psychotic disorder, with hallucinations
Specify if: With onset during intoxication
Opioid-induced mood disorder
Specify if: With onset during intoxication
Opioid-induced sexual dysfunction
Specify if: With onset during intoxication
Opioid-induced sleep disorder
Specify if: With onset during intoxication/with onset during withdrawal
Opioid-related disorder not otherwise specified

Adapted from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.10-1 DSM-IV Opioid-Related Disorders

Opioid Intoxication DSM-IV criteria for opioid intoxication are shown in Table 11.10-2. Intoxication can vary in severity. In severe cases of opioid overdose there is usually coma, severely depressed respiration, and pinpoint pupils. There may be gross pulmonary edema with frothing at the mouth, but X-ray evidence of pulmonary changes is seen even in less severe cases. Pulmonary edema is an opioid effect and is sometimes seen with overdoses of oral opioids that have been medically prescribed. Depending on when the patient is seen, there may also be cyanosis, cold clammy skin, and decreased body temperature. Blood pressure is decreased, but only falls dramatically with severe anoxia, at which point the pupils may dilate. Cardiac arrhythmias have been reported and may be related either to anoxia or to the presence of quinine as an adulterant in the opioid.

A. Recent use of an opioid.
B. Clinically significant maladaptive behavioral or psychological changes (e.g., initial euphoria followed by apathy, dysphoria, psychomotor agitation or retardation, impaired judgment, or impaired social or occupational functioning) that developed during, or shortly after, opioid use.
C. Pupillary constriction (or pupillary dilation due to anoxia from severe overdose) and one (or more) of the following signs, developing during, or shortly after, opioid use: (1) drowsiness or coma (2) slurred speech (3) impairment in attention or memory
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.
Specify if: With perceptual disturbances

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.10-2 DSM-IV Diagnostic Criteria for Opioid Intoxication

Opioid Withdrawal The opioid withdrawal syndrome can vary greatly in intensity, depending primarily on the dose of the opioid used, the degree to which the opioid effects on the CNS were continuously exerted, the duration of use, and the rate at which the opioid is removed from the receptors. These generalizations appear to apply as well to other categories of drugs, such as barbiturates and benzodiazepines. The DSM-IV diagnostic criteria for opioid withdrawal are shown in Table 11.10-3.

A. Either of the following: (1) cessation of (or reduction in) opioid use that has been heavy and prolonged (several weeks or longer) (2) administration of an opioid antagonist after a period of opioid use
B. Three (or more) of the following, developing within minutes to several days after criterion A: (1) dysphoric mood (2) nausea or vomiting (3) muscle aches (4) lacrimation or rhinorrhea (5) pupillary dilation, piloerection, or sweating (6) diarrhea (7) yawning (8) fever (9) insomnia
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 11.10-3 DSM-IV Diagnostic Criteria for Opioid Withdrawal

The clinical syndrome observed consists of purposive behavior, which is dependent on the observer and environment (e.g., complaints, pleas, and manipulations directed at getting more drug), and nonpurposive behavior (e.g., piloerection and dilated pupils), which is not goal oriented and is relatively independent of the observer and environment. The opioid withdrawal syndrome, while often exquisitely uncomfortable and distressing, is not, in contrast to withdrawal from alcohol or barbiturates, life-threatening, in healthy adults. Deaths have occurred during abrupt opioid withdrawal in debilitated patients with other medical disorders.

In the case of short-acting drugs, such as morphine or heroin, the first symptoms may be seen within 8 to 12 hours after the last dose of drug. In the least severe cases or very early in withdrawal, symptoms may consist only of dysphoria, irritability, restlessness, and general achiness, with few objective signs. In mild syndromes the signs and symptoms may be limited to craving, anxiety, dysphoria, yawning, perspiration, lacrimation, rhinorrhea, and restless and broken sleep. In more severe cases, as the syndrome progresses, additional signs and symptoms that may be seen include increasingly dilated pupils; piloerection (waves of gooseflesh, from which comes the term “cold turkey” to describe withdrawal); and hot and cold flashes. In severe syndromes, which in the case of heroin and morphine generally reach peak severity about 48 hours after the last dose, the patient may also experience nausea, vomiting, diarrhea, weight loss, fever (usually low grade), and increased blood pressure, pulse, and respiratory rate. Also often observed are twitching of muscles and kicking movements of the lower extremities (whence comes the phrase “kicking the habit.”) It is important to remember that substantial subjective distress can develop before the more obvious physical signs. Also, when so motivated, opioid-dependent patients in early withdrawal may exaggerate the severity of distress in the hope that it will elicit higher doses of drugs for relief.

With short-acting drugs, the acute phase of the syndrome, if untreated, runs its course in 7 to 10 days. In research subjects the acute phase was followed by a more subtle but longer-lasting phase, the protracted abstinence syndrome, that persisted for many weeks.

During this phase many physiological variables reached subnormal values, such as hyposensitivity to the respiratory stimulant effects of carbon dioxide. There was also disturbed sleep, overconcern about bodily discomfort, poor self-image, and a decreased capacity to tolerate stress. It is frequently assumed that the protracted abstinence syndrome plays a role in relapse to opioid use.

With longer-acting drugs, such as methadone or levomethadyl acetate (ORLAAM), which is also known as a-acetylmethadol (LAAM), the onset of withdrawal may be delayed for one to three days following the last dose. Although the syndrome is qualitatively similar, peak symptoms may not occur until the third to eighth day, and the symptoms may persist for several weeks.

Withdrawal from methadone, and presumably from levomethadyl acetate, is also followed by a protracted abstinence syndrome that may last for more than 6 months. If naloxone is given to a patient dependent on methadone, thereby displacing the drug abruptly from the receptors, the withdrawal is immediate in onset, can be quite

severe, and persists until the naloxone is metabolized and the residual methadone reoccupies the receptors.

Withdrawal from analgesics that are presumed to act in part as k-receptor agonists (e.g., pentazocine, nalbuphine) is generally rapid in onset, mild, and lasts a few days. The syndrome bears some similarities to the μ-receptor withdrawal syndrome, but there are also some distinct elements. Withdrawal from pure k-agonists is not associated with drug-seeking behavior. Since drug-seeking behavior is sometimes seen with pentazocine, it is likely that it has some low-level agonist actions at μ-receptors.

Withdrawal symptoms after several weeks of administration of buprenorphine, a partial μ-agonist, at doses of 8 mg daily parenterally or sublingually, are generally not severe in intensity but some patients are uncomfortable enough to request drugs for relief of discomfort and for insomnia. After sublingual administration (8 mg daily), withdrawal symptoms are experienced within a few days after drug use is stopped, and symptoms of the acute phase reach baseline levels by 7 to 10 days. It is usually mild and consists primarily of subjective effects, such as muscle aches, dysphoria, and insomnia, rather than more dramatic autonomic signs such as piloerection or diarrhea. If very large doses of buprenorphine (16 to 32 mg sublingually) have been taken, the drug effect persists for several days, with onset of wthdrawal symptoms at 72 to 96 hours.

Opioid Abuse and Opioid Dependence Opioid abuse (see Table 11.1-8) is a pattern of maladaptive use of an opioid drug leading to clinically significant impairment or distress and occurring within a 12-month period, but one in which the symptoms have never met the criteria for opioid dependence.

Opioid dependence (see Table 11.1-3 and Table 11.1-4) is inferred from behaviors that indicate some decrease in volitional control over the use of an opioid drug. DSM-IV specifies the criteria to be used by the clinician to decide whether the patient exhibits such a decrease in volitional control. These criteria are not specific for opioids, but are believed to apply for all psychoactive agents. DSM-IV does not require that any single criterion be met, and none is given special weight. Thus, the presence of tolerance and physical dependence (withdrawal) is not required. However, according to DSM-IV, if tolerance and physical dependence are present, that should be noted specifically.

Because tolerance develops to many of the actions of opioid drugs after long-term use, opioid effects are not readily detected by even the careful observer. Patients maintained on large oral doses of methadone function quite normally. Physicians, nurses, and other medical personnel who use opioids, even by injection, may go undetected by their colleagues for months or years. Thus, a candid history obtained from the patient or a reliable informant is needed to make a diagnosis of dependence, although evidence of recent and long-term use can be developed by testing urine or hair for the presence of opioids.

Opioid Intoxication Delirium Opioid intoxication delirium is most likely to happen when opioids are used in high doses, are mixed with other psychoactive compounds, or are used by a person with preexisting brain damage (see Table 10-22).

Certain opioids, such as meperidine, have toxic metabolites that can accumulate, causing delirium and sometimes seizures. Impaired renal function increases the likelihood of accumulation.

Opioid-Induced Psychotic Disorder Opioid-induced psychotic disorder can begin during opioid intoxication. The DSM-IV diagnostic criteria are contained in the section on schizophrenia and other psychotic disorders (see Table 13.3-4). Clinicians can specify whether hallucinations or delusions are the predominant symptoms.

Opioid-Induced Mood Disorder Opioid-induced mood disorder (see Table 14.6-18), can begin during opioid intoxication, or withdrawal and can result from chronic use. (see Table 11.10-3). Opioid-induced mood disorder symptoms may be of a manic, depressed, or mixed nature. A person coming to psychiatric attention with opioid-induced mood disorder usually has mixed symptoms, combining irritability, expansiveness, and depression.

Some degree of depressed mood (hypophoria) typically occurs during and for several weeks following withdrawal. Opioid-induced mood disorder should not be diagnosed unless the severity of mood disturbance exceeds what is normally encountered or persists for more than a few weeks and is of sufficient intensity to warrant independent clinical attention.

Opioid-Induced Sleep Disorder and Opioid-Induced Sexual Dysfunction Opioid-induced sleep disorder (see Table 21-18) and opioid-induced sexual dysfunction (see Table 19.1a-20) are diagnostic categories in DSM-IV. Hypersomnia is likely to be a more common sleep disorder among those given opioids therapeutically, but disturbed sleep (insomnia) is a common complaint of patients maintained on opioid agonists such as methadone. The most common sexual dysfunction is likely to be impotence, but patients maintained on methadone may complain of inability to achieve orgasm, rather than impotence.

Opioid-Related Disorder Not Otherwise Specified DSM-IV includes diagnoses for opioid-related disorders with symptoms of delirium, abnormal mood, psychosis, abnormal sleep, and sexual dysfunction. Clinical situations that do not fit into these categories are examples of appropriate cases for the use of the DSM-IV diagnosis of opioid-related disorder not otherwise specified (Table 11.10-4).

The opioid-related disorder not otherwise specified category is for disorders associated with the use of opioids that are not classifiable as opioid dependence, opioid abuse, opioid intoxication, opioid withdrawal, opioid intoxication delirium, opioid-induced psychotic disorder, opioid-induced mood disorder, opioid-induced sexual dysfunction, or opioid-induced sleep disorder.

Table 11.10-4 DSM-IV Diagnostic Criteria for Opioid-Related Disorder Not Otherwise Specified

Psychiatric Comorbidity The high prevalence of additional psychiatric disorders among treated opioid-dependent patients has now been repeatedly confirmed. Currently no subtypology of opioid-dependent patients based on psychopathology has been proposed. However, the type and severity of those additional diagnoses can influence the course of the disorder and the kind of treatment that is most likely to be effective.

Among opioid addicts seeking treatment at a program in New Haven in the 1980s, 87 percent met the research diagnostic criteria (RDC) for a psychiatric disorder, in addition to opioid dependence, at some point in their lives. Other substance-related diagnoses, such as alcohol dependence were considered additional disorders. A study of addicts seeking methadone treatment in Baltimore in the early 1990s found (using DSM-III-R criteria) a 24 percent lifetime prevalence of non–substance-related Axis I disorders, a 35 percent lifetime prevalence of Axis II disorders, and a total prevalence of additional non–substance-related diagnoses of 47 percent. In both studies, the most common diagnoses were mood disorders, alcoholism, antisocial personality disorder, and anxiety disorders (Table 11.10-5, Table 11.10-6, Table 11.10-7). In the Baltimore study, cocaine dependence (65 percent) was actually higher than alcohol dependence (50 percent). Multiple psychiatric diagnoses for a given patient were common. Researchers in Europe and Australia report similar overall distributions of psychiatric disorders among opioid users seeking treatment.

Diagnostic Category*	Lifetime Rates % (Current Rates %)		
	Men (N = 378)	Women (N = 338)	Total
Any Axis I disorder	15.6 (5.0)	33.4 (11.2)	24.0 (8.0)
Mood disorder	11.4 (2.1)	27.5 (5.3)	19.0 (3.6)
Major depressive disorder	8.7 (1.3)	23.7 (5.3)	15.8 (3.2)
Dysthymic disorder	2.4 (2.4)	4.4 (4.4)	3.4 (3.4)
Bipolar I disorder	0.8 (0.8)	0.0 (0.0)	0.4 (0.4)
Anxiety disorder	6.1 (3.4)	10.7 (6.8)	8.2 (5.0)
Simple phobia	1.9 (1.9)	5.3 (3.6)	3.5 (2.7)
Social phobia	1.9 (0.8)	3.6 (2.7)	2.7 (1.7)
Panic disorder	2.1 (0.3)	1.8 (0.9)	2.0 (0.6)
Agoraphobia	0.0 (0.0)	0.6 (0.3)	0.3 (0.1)
Obsessive-compulsive disorder	0.5 (0.5)	0.0 (0.0)	0.3 (0.3)
General anxiety disorder	0.8 (0.8)	0.0 (0.0)	0.1 (0.1)
Eating disorders	0.0 (0.0)	1.5 (0.0)	0.7 (0.0)
Bulimia nervosa	0.0 (0.0)	0.9 (0.0)	0.4 (0.0)
Anorexia nervosa	0.0 (0.0)	0.6 (0.0)	0.3 (0.0)
Schizophrenia	0.0 (0.0)	0.3 (0.3)	0.1 (0.1)

* Multiple disorders possible

Adapted from Brooner RK, King VL, Kidorf M, Schmidt CW, Bigelow GE: Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. Arch Gen Psychiatry 54:71, 1997.

Table 11.10-5 Non–Substance-Related Axis I Psychiatric Disorders in Opioid Users

Diagnostic Category*	Lifetime Rates % (Current Rates %)		
	Men (N = 378)	Women (N = 338)	Total
Opioid dependence	100 (100)	100 (100)	100 (100)
Cocaine			
Dependence	66.1 (39.4)	63.0 (41.1)	64.7 (40.2)
Abuse	13.0 (3.7)	11.8 (3.0)	12.4 (3.4)
Cannabis			
Dependence	58.7 (19.8)	42.0 (12.1)	50.8 (16.2)
Abuse	17.7 (2.6)	12.1 (2.1)	14.9 (2.4)
Alcohol			
Dependence	56.6 (29.4)	43.2 (19.5)	50.3 (24.7)
Abuse	16.1 (2.6)	9.5 (0.9)	13.0 (1.8)
Sedative			
Dependence	47.9 (17.2)	40.8 (16.6)	46.6 (16.9)
Abuse	15.3 (1.3)	10.4 (1.8)	13.0 (1.5)
Stimulant			
Dependence	19.0 (0.0)	19.2 (0.6)	19.1 (0.3)
Abuse	15.6 (0.0)	7.1 (0.0)	11.6 (0.0)
Hallucinogen			
Dependence	21.2 (1.1)	14.2 (0.3)	17.9 (0.7)
Abuse	12.7 (0.0)	5.6 (0.0)	9.4 (0.0)
Other substance†			
Dependence	15.1 (2.6)	12.7 (2.7)	14.0 (2.6)
Abuse	7.9 (0.5)	4.7 (0.6)	6.4 (0.6)

* Multiple diagnoses possible

† Primarily inhalants, sedatives, amphetamine

Adapted from Brooner RK, King VL, Kidorf M, Schmidt CW, Bigelow GE: Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. Arch Gen Psychiatry 54:71, 1997.

Table 11.10-6 Substance Use Disorders in Opioid Users

Diagnostic Category*	Lifetime Rates (%)		
	Men (N = 378)	Women (N = 338)	Total
Any personality disorder	40.5	28.4	34.8
Antisocial	33.9	15.4	25.1
Avoidant	3.4	7.1	5.2
Borderline	1.3	9.5	5.2
Passive-aggressive	3.7	4.4	4.1
Paranoid	4.5	1.8	3.2
Dependent	0.5	3.0	1.7
Histrionic	0.8	2.1	1.4
Narcissistic	1.6	0.0	0.8
Obsessive-compulsive	1.1	0.3	0.7
Schizotypal	0.3	0.3	0.3
Schizoid	0.3	0.3	0.3

* Multiple disorders possible

Adapted from Brooner RK, King VL, Kidorf M, Schmidt CW, Bigelow GE: Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. Arch Gen Psychiatry 54:71, 1997.

Table 11.10-7 Personality Disorders in Opioid Users

Among women, depression, anxiety disorders, and borderline personality disorder were significantly more common than among male patients, and alcoholism, cannabis misuse, and antisocial personality disorder were significantly less common. The rate of current psychiatric illnesses are obviously lower than lifetime rates and vary with the treatment setting. Among the methadone patients in the Baltimore study, 5 percent of men and 11.2 percent of women met criteria for a current diagnosis of a non–substance-related Axis I disorder, most commonly a mood or anxiety disorder. Forty percent of the sample had a current diagnosis of cocaine dependence, and 25 percent suffered concurrently from alcoholism. Other studies have confirmed rates of current major depressive disorder in the range of 17 to 20 percent and concurrent rates of alcoholism (alcohol abuse or dependence) of 20 to 30 percent. Although stimulant dependence (other than cocaine dependence) was not seen among the patients in Baltimore, it occurs frequently in the western United States where methamphetamine use is more prevalent.

The presence of a comorbid psychiatric condition has been shown to be associated with more severe substance abuse in some studies. In the Baltimore study, patients with comborbid psychiatric conditions were more likely to suffer from polysubstance abuse and to have begun abusing drugs at an earlier age. This was particularly true of patients with antisocial personality disorder.

The high rates of psychiatric comorbidity in the methadone maintenance population are by no means unique to this modality of treatment. Similar patterns of additional psychiatric disorders have been found by workers at other public clinics and by clinicians in private practice. Among patients in therapeutic communities, 60 percent reported depressive symptoms during the year before entry, 28 percent had contemplated suicide, and 13 percent had made at least one suicide attempt. And, although opioid abusers who have current psychiatric problems are more likely to seek treatment, it should not be assumed that those who do not seek treatment are free of comorbid psychiatric problems. Among the respondents in the community-based ECA study, about half of those with a substance (other than alcohol) use disorder had one or more additional psychiatric disorders (odds ratio 4:5); anxiety disorder, 28.3 percent (odds ratio 2:5); mood disorder, 26.4 (odds ratio 4:7); antisocial personality, 17.8 percent (odds ratio 13:4); schizophrenia, 6.8 percent (odds ratio 6:2). The probability of comorbidity is higher for those with a lifetime diagnosis of an opioid or cocaine disorder than for those who abuse cannabis. Widespread comorbidity was again found in the similar NCS study: 59 percent of those with a lifetime history of illicit drug abuse or dependence also met DSM-IV criteria for another mental disorder, and 71 percent met the criteria for a lifetime alcohol use disorder.

PATHOLOGY AND LABORATORY EXAMINATION

In opioid abuse and opioid dependence there may be no abnormal laboratory findings at all. Urine tests can usually detect short-acting opioids, such as heroin, for 12 to 36 hours after use. Very potent opioids, such as fentanyl, may not be detected by standard opioid screens. Standard tests for heroin actually test for its main metabolite, morphine. A urine test positive for morphine can also be caused by therapeutic doses of codeine or by the ingestion of modest amounts of poppy seeds (of the type and amount used to flavor bagels and other breads and pastries).

Persons who have shared injection implements often test positive for hepatitis (B and C) and for the human immunodeficiency virus (HIV). Other liver enzymes may be elevated if there is active hepatitis. There may be positive and false-positive tests for syphilis. Chest X-rays may show evidence of pulmonary fibrosis if the person has been using injection materials contaminated with microcrystalline talc or cotton particulates. During withdrawal white blood cell counts and cortisol levels may be elevated.

Physical findings may be unremarkable if opioids are ingested orally; snorting (insufflation) of heroin may irritate nasal membranes. Drug injectors, however, may show widespread evidence of having used unsterile injection equipment. There may be needle tracks over veins on the arms, legs, and, in some cases, the backs of the hands and the femoral and jugular veins. Infections and venous scleroses and lymph obstruction may lead to severe edema of the hands and feet. There may be skin abscesses or scars on accessible skin surfaces as a result of unsterile subcutaneous injections (Fig. 11.10-2). There may be rock-like hardening of subcutaneous and muscle tissue as a result of repeated intramuscular injections of meperidine (often seen among health professionals). Endocarditis may produce fever and heart murmurs. In addition, a variety of neurological sequelae of intravenous heroin use may be detected.

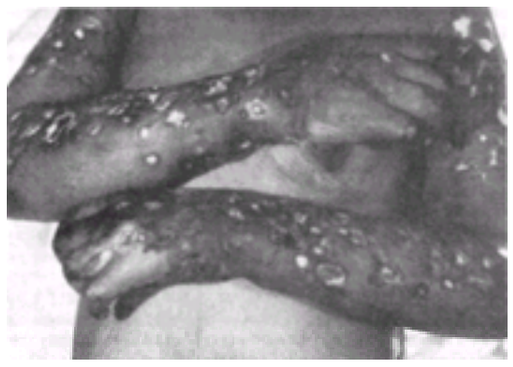

FIGURE 11.10-2 Skin popper. Circular depressed scars often with underlying chronic abscesses, can result from skin popping. (Courtesy of Michael Baden, M.D.)

DIFFERENTIAL DIAGNOSIS

The diagnosis of opioid dependence can be relatively straightforward when the patient is willing to be candid. It can be quite taxing when the patient is motivated to conceal past patterns of opioid use, as is frequently the case among health care professionals or persons who obtain opioids from medical sources by simulating disease or greatly exaggerating the painful nature of disease actually present.

The diagnosis of opioid withdrawal is easier when the opioid is short-acting and the symptoms are accompanied by obvious physiological signs. This is not the typical case, and the clinician is often confronted by a patient complaining of various aches, pains, anxiety, and insomnia in the absence of obvious signs of withdrawal. Patients who are dependent on opioids may withhold information on the use of other classes of drugs. Because opioid withdrawal does not generally cause tremulousness, confusion, delirium, or seizures, the presence of those signs or the failure of insomia and anxiety to respond to reasonable doses of an opioid should raise the possibility of dependence on alcohol, sedatives, or benzodiazepines.

Opioid intoxication must be differentiated from mixed intoxications in which opioids play only a minor role. In general, a failure to respond to modest doses of naloxone suggests that intoxication is caused by a nonopioid. Some patients, more typically the elderly, may respond to therapeutic doses of a μ-agonist with dysphoria and confusion; such reactions are generally short-lived. They are seen more commonly with mixed agonist-antagonists. They should be considered atypical opioid intoxications rather than opioid-induced intoxication delirium or opioid-induced psychotic disorders, which although listed in DSM-IV are quite rare. One possible exception is the state associated with the accumulation of toxic meperidine metabolites. However, even in that case the syndrome does not usually outlast the metabolites and should probably be considered an intoxication.

COURSE AND PROGNOSIS

It was once commonly believed that experimentation with illicit opioids invariably led to dependence, but it is now apparent that only a fraction of those who briefly experiment with illicit opioids develop serious problems. It is still true that those who use opioid drugs heavily—at least once a week—usually go on to use daily, at least for a brief period. Data from national surveys suggest that about one third of those who experimented with heroin during the epidemic of the 1960s and 1970s became dependent on it at some point. It is likely, however, that many who go on to develop some degree of dependence recover without ever seeking formal treatment. In the St. Louis ECA study, only 20 percent of those with a lifetime history of opioid abuse had any symptom during the year prior to the interview.

Some persons apparently can use opioids occasionally (e.g., several times a month) over periods of months or years without becoming drug dependent. It may be that self-imposed limits on the time and place they use drugs may help prevent progression to addiction, but drug injectors are still at risk for death from overdose, as well as for infections and other medical complications. When addiction does develop, the subsequent course of the syndrome depends on environmental factors, the characteristics of the user, the route of administration, and the specific opioid being used.

Heterogeneity of Lifestyle Opioid addicts seeking or entering treatment in the United States and the United Kingdom exhibit a surprising heterogeneity of lifestyles, attitudes toward conventional values, and criminality. Some addicts, except for their drug use, are quite conventional, avoid criminality, work at legitimate occupations, and do not identify with the addict subculture. At the other extreme, some addicts live exclusively by illicit activities and are highly involved with other addicts. A third group appears to identify with both cultures, engaging in some criminal activities and interacting with other addicts, but living primarily on legitimate earnings. A fourth group of addicts seems not to be involved in either the conventional culture or the addict subculture. They tend to be unemployed and to live on welfare rather than on criminal earnings and often have high levels of psychopathology. Health professionals and those who become addicted to opioids in the course of medical treatment probably constitute additional distinct subgroups.

Slips and Relapses In the early stages of opioid use the most typical course for the user of illicit opioids is one of periods of abstinence, either voluntary or forced (by imprisonment or hospitalization) and lasting from a few weeks to many months, followed by relapse to opioid use and readdiction. Among the addicts who enter abstinence-oriented treatments, relapse occurs most often in the first 3 months; a number of studies show that at least two out of three patients relapse within 6 months. Depression and life crises (especially arguments and interpersonal losses) are associated with relapse to illicit drug use; those factors are additive. The impact of those risk factors is reduced by treatment in drug abuse programs.

The theme running through the many specific reasons given for periods of voluntary abstinence is a desire to change life patterns and a weariness with the constant difficulty of trying to obtain illicit opioids. At the time of reentry into treatment, addicts are often less seriously impaired than when first treated, suggesting some residual benefit from intermittent treatment. With successful treatment episodes and with age the duration of each opioid episode tends to shorten.

Repeated relapse is not an inevitable consequence of opioid dependence. Eighty-eight percent of United States Army enlisted personnel who became addicted during a tour of duty in Vietnam did not become readdicted at any time during the 3 years following their return to the United States; 56 percent did not use opioids at all. Of the soldiers who did become readdicted in the first year, 70 percent were no longer addicted in the following 2 years. Only 2 percent of the soldiers who used opioids in Vietnam (6 percent of those who tested positive for opioids at time of departure) entered treatment following their return to the United States. However, the relapse rates for those who did enter treatment were as high as for civilians, with two thirds relapsing within less than a year. It is possible that among those who become dependent on opioids, the ones who cannot stop without formal help have a more severe form of the disorder.

In considering the natural history of opioid addiction, a distinction must be made between the course of all those who report a period of dependence (e.g., among respondents to a household survey) and the course of those who enter treatment. Those who enter treatment may have additional problems or a more severe addiction. Further, opioids are rarely the sole drug class used. During the 1980s a high proportion of opioid-dependent patients also abused or became dependent on cocaine. Depression was found to be a risk factor for continued cocaine use among patients in opioid treatment programs, and cocaine use itself predicted continued heroin use and earlier discontinuation of methadone treatment. Abstinence from opioids or recovery from opioid dependence does not mean that the person has no drug problems. Standard methadone maintenance treatment programs do not appear to have any major effect in preventing cocaine use, and abuse of alcohol and other nonopioid drugs is common among patients in these programs who are no longer using other opioids.

Long-Term Prognosis Studies from both the United States and the United Kingdom support the view that opioid addiction is a disorder that eventually ends for many of those addicts who survive. This seems to be so even for those who seek treatment, a group that may have a more severe variety of the disorder. For a substantial number of these persons, it is prolonged incarceration that ends their opioid use; others stop using opioids but continue to abuse other drugs. Although there are some old opioid addicts in the United States, their numbers are few. Opinion remains divided as to just how many opioid addicts eventually achieve abstinence outside of an institution. One review of long-term follow-up studies found abstinence rates between 10 and 19 percent for drug-free treatment and between 9 and 21 percent following treatment with methadone. A 24-year follow-up of narcotic addicts remanded to the California Civil Commitment Program in the early 1960s interviewed 354 survivors. Seventeen percent had not used narcotics and had no other serious drug problems over the preceding 36 months; 13 percent had been abstinent from narcotics but not from other drugs and had not been incarcerated. The largest group (41 percent), had avoided incarceration but used narcotics and other drugs. The next largest group (29 percent), were incarcerated, and many had used drugs during the 12 months prior to interview. Thus, of those still alive 30

percent were free of opioids, although some used other classes of drugs.

The usual measures of treatment outcome—legitimate work, crime, drug use, family relationships, psychological adjustment—are best predicted by different pretreatment variables. Thus, pretreatment history of high levels of criminal activity most accurately predicts posttreatment criminal activity, and previous stable work history is more predictive of posttreatment gainful employment. Severity of psychological problems at the beginning of treatment, however, is a predictor of outcome on all dimensions. Opioid addicts with the least severe psychological problems appear to respond better to all treatments on all outcome measures.

In long-term follow-ups no important differences in outcome have been found among patients treated initially by different modalities, although over shorter follow-up periods, methadone maintenance, therapeutic communities, and drug-free programs were significantly superior to detoxification programs. Differences—or lack thereof—among different forms of treatment may be affected by pretreatment characteristics of the patient population. For example, studies from the United States and Australia suggest that addicts electing detoxification are younger and have less psychopathology than those entering more lengthy forms of treatment.

Although a period of prolonged abstinence is a good predictor of long-term outcome, a study of heroin users in Texas found that 33 percent of those who reported 3 years of abstinence eventually relapsed. The long-term outcome for opioid dependence among persons with additional major psychiatric disorders, such as schizophrenia or bipolar I disorder, is not generally known. However, one study that specifically examined the efficacy of additional psychiatric care for dual-diagnosis subjects in an outpatient drug dependence clinic found that while they initially had significantly greater percentages of positive urine toxicologies than patients without comorbid psychiatric conditions, they had higher retention rates in treatment and after 6 months of treatment had comparable success rates as the single-diagnosis group. The course of opioid dependence among medical personnel supervised by state-level organizations is far less bleak than that of heroin users who enter public sector treatment programs.

Opioids and Crime The statistical relation between the use of illicit opioids and crime in the United States is unquestionable. Persons who use illicit opioids commit crimes more frequently than do nonusers. There are, however, questions about the degree to which one behavior causes the other. The direct tranquilizing actions of opioids ought to reduce criminal activity rather than increase it.

Throughout much of history in countries where crude opioids (e.g., opium, tincture of opium) were inexpensive and socially acceptable and were either smoked or taken by mouth there was little relation between opioid use and criminal behavior. The association between opioid use and crime emerges primarily in countries, such as the United States, that have tried to restrict the use of opioids to legitimate medical indications but have been unable to eliminate illicit opioid traffic.

Most addicts' crimes are directed at generating income for drugs by shoplifting, petty theft, and selling drugs; however, recently more addicts seem to be engaging in crimes, such as robbery, that involve the potential for violence. However, the relationship between drug use and crime is not one of simple cause and effect. In the United States and the United Kingdom more than 50 percent of heroin addicts interviewed in prisons and jails or treatment programs had been arrested prior to their first opioid use.

In a study in Baltimore, adolescent boys (11 to 14 years old) who later became opioid addicts exhibited substantially more criminal behavior than did peers or age-matched nonpeers from the same community. As noted previously, there is a high incidence of antisocial personality disorder among drug users, even when the diagnostic criteria for this disorder are limited to exclude antisocial behavior engaged in solely because of the need for drugs. Although it might be argued from these data that the criminal behavior seen after the onset of addiction is merely a continuation of a criminal lifestyle, criminal activity does increase sharply after the onset of opioid addiction. Arrest rates for nondrug offenses increase from 1.5-fold to threefold after the onset of addiction, and self-reported property crimes increase to a comparable degree. However, it has also been suggested that crime is a better predictor of opioid use than opioid use is of crime, and that rather than heavy use causing crime, day-to-day success in crime enables the escalation to heavier opioid use.

Further evidence pointing to a causal relationship between opioid use and crime is the sharp reduction in both self-reported criminal activity and arrests during periods of less than daily illicit opioid use. The decrease in crime is seen whether the decrease in opioid use is a result of effective treatment, probation, parole, or spontaneous cessation. When criminal behavior antedates opioid use, however, it is unrealistic to expect the successful treatment of opioid dependence to eliminate criminal behavior entirely. Addicts who were criminally active prior to opioid dependence are more likely to persist in criminal acts when abstinent. It is postulated that only more comprehensive treatment directed at social, psychological, and vocational factors, rather than just at illicit drug use, is likely to alter antisocial lifestyles.

Toxicity, Morbidity, Medical Complications, and Life Expectancy Opioids used orally are relatively nontoxic. Whereas long-term use, as in methadone maintenance, is associated with minor endocrine abnormalities, constipation, and some sleep disturbance, no major damage to internal organs has been noted and no significant impact on longevity would be expected. The cognitive impairment seen with long-term alcohol and sedative use is not generally found with long-term use of oral opioids. Nevertheless, the life expectancy of opioid addicts, especially heroin addicts, is markedly reduced.

Before the emergence of HIV infection among opioid injectors, estimates ranged from a twofold to a threefold increase in the expected mortality rate for older addicts to a twentyfold increase in the expected rate among young addicts. Follow-ups of treated opioid addicts in the United States indicate an overall death rate of 1 to 1.5 percent a year. A substantial proportion of those deaths were due to drug overdose, drug-related infections, and suicide. Homicide is also a common cause of death among urban opioid addicts in the United States. The death rate at British clinics providing clean injectable drugs to young addicts was estimated at 2 to 3 percent a year, at least twenty-fold higher than the death rate of comparably aged contemporaries. Mortality rates far higher than for age-adjusted controls were also observed in Italy during a period when heroin use was not a criminal offense. The suicide rate among opioid addicts is estimated to be three times higher than in the general population. This is probably an underestimation because it is difficult to determine how many overdose deaths are to some degree intentional.

The advent of HIV has changed the patterns of mortality among drug users. In some areas of the United States and Europe, acquired immune deficiency syndrome (AIDS) is now a leading killer of injecting drug users. HIV seroprevalence among users has a wide geographic variability. Although the death rate of heroin addicts in treatment remains higher than that of age-adjusted controls, it is substantially lower than the rate for untreated addicts.

Medical Complications Opioid drugs, properly administered, are associated with few serious medical complications. Most complications associated with opioid abuse are those associated with the route of administration. Because opioid addicts—even physicians who have access to drugs and sterile materials—tend to neglect the hygienic aspects of injecting, infections of skin and systemic organs are quite common. Filtering illicit opioids through cigarette filters or wads of cotton and injecting materials intended for oral use allows starch, talc, and other particulate contaminants to enter into the bloodstream. These particulates can cause pulmonary emboli, which can eventually result in angiothrombotic pulmonary hypertension and right ventricular failure. Staphylococcal pneumonitis may also be related to septic emboli. Endocarditis and septicemia involving lesions either of the tricuspid or of the aortic and mitral valves is a frequent complication. Less frequent but equally serious complications are meningitis and brain abscess. Other frequently seen infections that can be related to injecting the substance or the sharing of needles include viral hepatitis, malaria, tetanus, osteomyelitis, and HIV. Syphilis transmission has also been associated with sharing of needles, although most cases are probably acquired in the usual fashion. False-positive tests for syphilis are also not uncommon among injecting drug abusers. Many opioid addicts who inject have a low-level chronic hepatitis without jaundice and may have abnormal liver function tests. Seroprevalence surveys conducted by the Centers for Disease Control (CDC) found some marker of hepatitis B infection in sera from 60 to 80 percent of drug injectors. Rates of hepatitis C infection are similarly elevated. Abnormal liver function tests, which are found in about two out of three heroin addicts, may persist for long periods after the cessation of injection. Comborbid alcoholism may in some cases contribute to the liver disease.

Other complications associated with poor intravenous injecting technique include chronic edema of extremities (e.g., puffy hands), probably due to lymphatic obstruction caused by contaminants and sclerosis of veins caused by the drugs or their dilutants. Chronic lymphadenopathy was common among addicts even before the advent of HIV and was also thought to be related to particulate contaminants. Subcutaneous or intracutaneous injection (“skin popping”) may cause widespread ulceration and disfigurement as a result of chemical necrosis or infection (see Fig. 11.10-2). These injecting techniques may be used by addicts who have sclerosed their major veins. Some drug users, determined to experience the effects of the drug used intravenously, switch to the use of femoral and jugular veins when the surface veins of the arms and legs have become unusable. Complications associated with injection of particulate contaminated material include pulmonary hypertension—sometimes leading to cor pulmonale—occasionally seen in heroin smokers.

Additional medical complications are likely to be due to contamination of illicit opioids with other chemical substances. A number of changes found at autopsy, such as degernation of the globus pallidus and necrosis of spinal gray matter, may fall into this category. Occasionally there are clinical manifestations in those users surviving overdose experiences. Examples are transverse myelitis, amblyopia, plexitis, peripheral neuropathy, parkinsonian syndromes, intellectual impairment, and personality changes. Pathological changes in muscles and degeneration of peripheral nerves have also been seen. Illicit laboratories sometimes produce opioid-like agents that are extremely toxic or are so potent that even small doses are lethal. For example, 1-methyl,4 phenyl,1,2,3,6-tetrahydropyridine (MPTP), a contaminant of

illegally produced meperidine, produces a severe form of parkinsonism by selective destruction of dopaminergic neurons. While not technically a contaminant, the illicit fentanyl analog, 3-methylfentanyl (“China White”), is 1000 times more potent than morphine and may have been responsible for several hundred overdose deaths.

HUMAN IMMUNODEFICIENCY VIRUS In 1989 it was estimated that from 5 to 33 percent of intravenous drug users in the United States were infected with HIV. However, despite comparable rates of sharing of injection equipment, there is wide regional variation. Seroprevalence rates during the 1990s remained highest in the northeastern United States and Puerto Rico (10 to 65 percent), and lowest (less than 5 percent) in the west, midwest, and south. Risk of infection appeared to be higher among Hispanics and African-Americans. Infection was far lower among former heroin users who had been continuously in methadone treatment. High rates of HIV infection among drug injectors is also found in many countries of Europe and Southeast Asia.

The finding that not all drug users who shared needles were infected with HIV stimulated vigorous prevention efforts aimed at recruiting patients into treatment and teaching them how to avoid infection by cleaning injection equipment properly, not sharing equipment, and not participating in high-risk sex. Outreach by community workers, coupled with specific instruction, reduces self-reported equipment sharing but has far less effect on high-risk sexual behavior. The question of whether providing injecting drug users with sterile equipment at little or no cost would substantially slow the spread of HIV infection is still controversial, but the available evidence indicates that those who inject drugs will use sterile equipment in order to avoid disease if the equipment is available at reasonable cost. Even in the absence of available syringes, many addicts have proven receptive to outreach programs that teach needle sterilization techniques. This recognition has led to the establishment of a number of syringe exchange programs, first in Europe and now in the United States. These programs can be stand-alone sites, incorporated into other kinds of drug treatment centers, operated in mobile units, or via community pharmacies. Additionally, a number of states have begun to liberalize their needle and syringe possession laws, making it possible for addicts to purchase and possess clean needles. In many areas the prevalence of HIV now appears to be declining as those with AIDS succumb and as the noninfected reduce the frequency of sharing injection equipment.

TUBERCULOSIS. Even before the HIV epidemic the incidence of tuberculosis was higher among heroin addicts than in the general population. Now a reemergence of tuberculosis poses a major problem for many treatment programs. Patients with compromised immune systems are far more vulnerable to developing active tuberculosis once infected, and poor compliance with antitubercular medication has led to the emergence of drug-resistant strains of the tubercle bacillus. Studies indicate that 4 to 21 percent of AIDS patients are infected with tuberculosis.

OPIOIDS AND THE IMMUNE SYSTEM. μ-opioid actions at CNS receptors can produce immunosuppressive effects (e.g., decreased natural killer cell activity). Opioid receptors are also found on lymphocytes and naloxone-reversible opioid effects can be demonstrated on white cells in vitro. However, the concentrations required are quite high; it seems unlikely that direct actions on white cells mediate effects on the immune system. In heroin addicts there are changes in the ratio of helper-to-suppressor T-cells and a suppression of cell-mediated immunity. The relation to opioid use is still unclear; the effects are probably more related to unhygienic injection practices. Natural killer cell activity and immunoglobulin G and M are within normal limits in former heroin users maintained for several years on methadone, although abnormalities are typically observed among heroin addicts.

TREATMENT

Treating Intoxication (Overdose) The first task is to ensure an adequate airway. Tracheopharyngeal secretions should be aspirated; an airway may be inserted. The patient should be ventilated mechanically until naloxone, a specific opioid antagonist, can be given. Naloxone is administered intravenously at a slow rate—initally about 0.8 mg per 70 kg of body weight. Signs of improvement—increase in respiratory rate and pupillary dilation—should occur promptly. In opioid-dependent patients, too much naloxone may produce signs of withdrawal as well as reversal of overdosage. If there is no response to the initial dosage, naloxone maybe repeated after intervals of a few minutes. In the past it was thought that if no response was observed after 4 to 5 mg the CNS depression was probably not due solely to opioids. However, buprenorphine is difficult to antagonize with naloxone and higher doses may be required. The duration of action of naloxone is short compared with that of many opioids, such as methadone and levomethadyl acetate, and repeated administration may be required to prevent recurrence of opioid toxicity.

Withdrawal (Detoxification) Currently oral methadone is the opioid drug most often used to ameliorate the severity of withdrawal. Until the recent introduction of levomethadyl acetate, methadone was the only opioid approved for this purpose by federal regulations. Buprenorphine, a partial μ-agonist opioid, is likely to gain approval for treatment of opioid dependence in the near future. Theoretically, however, any opioid can be administered and then gradually reduced. In general, the objective of treating patients hospitalized for detoxification is to make the withdrawal experience tolerable rather than to suppress all symptoms; patients should be told to expect some discomfort.

For patients who have been using street drugs, the initial dosage of methadone is usually 10 to 20 mg orally. If signs of withdrawal persist, the dose can be repeated after about 2 hours. As a general rule, initial stabilization does not require more than 40 mg of methadone during the first 24 hours. Physician addicts and others who had access to pure drugs are exceptions to that general rule and may require higher doses. If the usual daily dose of opioid (heroin, meperidine, hydromorphone) is known, the equivalent withdrawal-suppressing dose of methadone can be calculated. For example, methadone is approximately 3 times as potent as morphine in suppressing withdrawal. Because the objective of treatment is to prevent severe withdrawal, the current tendency is to base dosage on clinical history rather than to wait for obvious withdrawal or to precipitate withdrawal with naloxone. Caution should be used when giving doses above 40 mg of methadone daily because the accumulation of methadone can lead to serious toxicity. Once the patient is stabilized, the dose can be gradually reduced. If patients are hospitalized, this can usually be accomplished within 10 days by reducing the dose by 10 to 20 percent per day.

In some cases of low-level to moderate dependence, the withdrawal process can be accomplished within 1 or 2 days by giving repeated doses of naloxone to precipitate symptoms and ameliorating discomfort with diazepam (Valium) or clonidine. This technique permits rapid transfer to naltrexone (ReVia). If the detoxification process is to be successful with outpatients, the time required may have to be extended considerably. Some clinicians recommend 10 percent a week as the maximum rate of reduction from high dosage and reductions of 3 percent a week when the daily dosage is below 20 mg. Patients may be unwilling to accept such a long period of treatment and financial considerations may preclude this approach.

Despite the best efforts of clinicians and the use of slow detoxification, all studies to date indicate that many patients drop out before completing outpatient detoxification and that relapse rates following successful outpatient or inpatient detoxification are high. Relapse rates are particularly high for users of street heroin who attempt to detoxify on an outpatient basis. For such patients, extending the period of withdrawal from 3 weeks to 6 weeks or using a longer-acting drug, such as levomethadyl acetate, does not appear to alter a pattern of almost universal and rapid return to illicit opioid use. Federal regulations now permit the clinical use of levomethadyl acetate for detoxification. Despite the problems of initial stabilization, the less frequent need for clinic visits may make levomethadyl acetate useful for very slow detoxification schedules.

In an early study in which drug users were detoxified over a 90-day period with either methadone or low doses (2 mg) of sublingual buprenorphine, both groups did equally poorly in terms of dropping out of treatment and reverting to use of illicit opioids. Subsequent studies using higher doses of buprenorphine for detoxification found high patient acceptance, good 30-day retention comparable to that with methadone, reduction in illicit opioid use, and withdrawal symptoms that were generally quite mild. The use of buprenorphine may facilitate the transfer to naltrexone. Buprenorphine has not yet been approved for the treatment of opioid dependence, but studies have shown that patients can be transferred from heroin to buprenorphine even in a primary care setting. Opioid withdrawal can be suppressed for 48 to 72 hours when higher doses of buprenorphine are used. In general, patients who form a good relationship with a therapist or who have been maintained for some time on methadone have a far better prognosis for completing detoxification.

Nonopioid Agents for Detoxification Clonidine, an a_2-adrenergic receptor agonist originally marketed as an antihypertensive agent, has been shown to suppress some elements of the opioid withdrawal syndrome. Patients stabilized on relatively low doses of opioids—for example, 30 to 40 mg of oral methadone a day—can abruptly discontinue the opioids and clonidine can be used to attenuate withdrawal. Clonidine is given orally, starting at doses of 0.1 to 0.3 mg three to four times a day. In outpatient settings a total dosage above 1 mg a day is not recommended, although higher doses (1.5 to 2.5 mg) have been used with hospitalized patients. The major adverse effects are hypotension, which can be quite extreme, and sedation, and the dosage must be carefully individualized. Patients in outpatient clinics who have been stabilized on methadone and have developed a relationship with their therapists were found to be more successful in achieving abstinence when treated with clonidine than patients who were taking illicit heroin when clonidine treatment was begun. Some studies have found that outpatient detoxification with clonidine is almost as successful as detoxification using decreasing doses of methadone. In hospitalized patients clonidine permits more rapid detoxification from opioids than does the gradual withdrawal of methadone and patients are more likely to complete the detoxification process.

Clonidine appears to be least effective in suppressing postwithdrawal muscle aches, lethargy, insomnia, restlessness, and craving. There is no evidence that it is useful in preventing relapse after the completion of detoxification. Clonidine does appear to facilitate the detoxification of patients maintained on methadone and their

subsequent stabilization on naltrexone. Properly used, the combination of clonidine and naltrexone can shorten the period of hospitalization required for detoxification to less than 5 days. The clonidine to naltrexone technique has also been used in a day-hospital outpatient setting. Heroin addicts were given oral clonidine 3 times on the first day, with dosages ranging from 0.1 to 0.3 mg adjusted on the basis of dependence severity and blood pressure response; naltrexone was started on the second day. The initial naltrexone dose of 1 mg (orally) was gradually increased so that patients were receiving 40, 50, and 150 mg by days 3 to 5 respectively.

Lofexidine is another a_2-agonist that is used to ameliorate opioid withdrawal. It is not approved for this use in the United States, but in the United Kingdom it is more widely used than clonidine because it is believed to produce less hypotension.

a_2-Agonists can be used in conjunction with intravenous naloxone to accelerate or precipitate detoxification of heroin-using or methadone-maintained patients. This is accomplished by first briefly stabilizing the patient on buprenorphine and then increasing the doses of naloxone and naltrexone to adjust for the greater affinity of buprenorphine for opioid receptors. Starting patients on oral naltrexone soon after naloxone-precipitated withdrawal symptoms subside does not appear to cause any additional withdrawal symptoms.

Ultrarapid Detoxification The duration of the acute phase of opioid withdrawal can be markedly shortened by using opioid antagonists to remove opioid agonists from opioid receptors. Such a procedure results in a withdrawal syndrome far more intense than that which occurs with mere abrupt discontinuation (cold turkey). Building on the work of researchers who used small doses of naloxone along with benzodiazepine sedation to precipitate withdrawal and initiate naltrexone treatment, several groups of investigators developed methods to make tolerable and further shorten the duration of acute withdrawal. Researchers at Yale used clonidine to reduce withdrawal symptoms associated with starting naltrexone. Others added sedatives to the clonidine-naltrexone regimen. It was reported that the acute withdrawal syndrome could be shortened to less than 2 days if the procedure was initiated under full thiopentone or methohexitone anesthesia lasting 3 to 4 hours. Since these early studies, the procedure of precipitating withdrawal with opioid antagonists under anesthesia (sometimes called *ultrarapid opioid withdrawal* or *detoxification*) has become more widely used and also in many cases commercialized.

The basic technique consists of premedication with several adjunctive medications to reduce vomiting and bowel activity (e.g., octreotide [Sandostatin], ondansetron [Zofran], cimetidine [Tagamet]) and modulate autonomic activity (clonidine), following which the patient is sedated and intubated, and light general anesthesia is induced with midazolam (Versed) or propofol (Diprivan). Patients are then given naltrexone by nasogastric tube. Anesthesia is maintained for 4 to 8 hours, after which patients are allowed to recover. While some patients experience minimal withdrawal upon awakening, many still feel achy and washed out for an additional 24 to 72 hours. Patients are immediately started on a regimen of oral naltrexone.

There have been a number of variations on this general theme, with some practitioners using nalmefine (Revex) instead of naltrexone and others varying the premedications, the anesthetic, or both. Sometimes patients are taken off full μ-agonist opioids and put on buprenorphine, after which withdrawal is precipitated under heavy sedation without use of endotracheal or nasogastric intubation.

There is little reliable follow-up data on people who have undergone this detoxification procedure and there is little reason to believe that ultra-rapid detoxification would by itself be more efficacious than detoxification over 4 to 10 days. However, most commercial organizations offering this service state that they provide more than just detoxification under anesthesia. Most include in the price for the service a number of visits for weekly counseling and naltrexone to be taken orally. Some encourage the family to oversee the ingestion of naltrexone. Since it is typically the family that pays the very high charge for ultra-rapid detoxification, the family is usually motivated to participate in the postdetoxification treatment.

Many clinicians have expressed the view that since opioid detoxification using conventional methods is almost never fatal, ultra-rapid detoxification subjects patients to unnecessary risks. However, it must also be recognized that some patients are extremely reluctant to undergo withdrawal (either inpatient or ambulatory) and many who begin never complete the process. Further research is needed to determine if there is any benefit to the use of this expensive and intensive method.

Other Techniques Because opioid withdrawal is rarely life-threatening in healthy adults, a variety of nonpharmacological approaches to drug abstinence have been and continue to be utilized. Abrupt withdrawal (cold turkey) is used in some countries (e.g., Singapore) as a matter of policy because it is believed that experiencing severe withdrawal is a deterrent to relapse. Abrupt withdrawal, coupled with considerable emotional support, is still used in some therapeutic communities. Because withdrawal stress fatalities can occur in debilitated persons (e.g., those with advanced AIDS, tuberculosis, or heart disease), some caution is indicated in selecting accelerated withdrawal or nontreatment for them.

Acupuncture has also been utilized to alleviate opioid withdrawal. The acupuncture-stimulated release of endogenous opioids provides some rational basis for this approach. Whether opioid withdrawal is substantially alleviated by various electrical devices that purport to stimulate the CNS is still not certain despite efforts at controlled studies.

Herbal medicines aimed at ridding the body of toxic substances are sometimes used in a religious or semireligious context (Fig. 11.10-3). There is no evidence to indicate that when those traditional approaches are used in their own cultural settings the outcome is any better or worse than with the more medically sophisticated approaches typically used in the United States.

FIGURE 11.10-3 Treatment of addicts at Tham Krabok Monastery in Thailand results in a 70 percent success rate, according to its records. The 10-day free treatment begins with a vow to Buddha never to use narcotics again. Then patients are given an herbal medicine that makes them vomit immediately. (Reprinted with permission from White PT, Raymer S: The poppy—For good and evil. Natl Geogr *167:*187, 1985.)

Treatment of Dependence Following detoxification, the treatment process may involve outpatient, inpatient, residential, day-care settings, or prisons. The staff may consist solely of former addicts with considerable personal experience but minimal formal training, formally trained health care professionals, correctional officers, or some combination of these personnel. The program may emphasize group, individual, or family interactions; all drugs may be banned in some programs whereas other programs may be centered on the use of a single drug, such as methadone, or may be designed specifically to target special populations selected by age, sex, culture, race, or special status (e.g., pregnant women). Some of the treatment alternatives to be described, such as opioid maintenance, are not currently available in many other countries and not all options are equally accessible in all parts of the United States. Theoretically, many different combinations of environment, staffing, philosophy, target populations, and pharmacological agents are possible, but in practice there are only a handful of major program varieties, each of which may have a few subvarieties.

Despite a tendency toward a greater diversity of programs for the treatment of addiction in general, most of the opioid users treated annually in the United States are treated in one of four types of programs: detoxification (ambulatory or residential), usually followed by outpatient counseling; opioid maintenance and (methadone, levomethadyl acetate); residential or therapeutic communities; and outpatient drug-free programs. Opioid dependence is also treated in chemical-dependency programs based largely on 12-step principles. Many opioid users benefit from participation in such self-help groups as Alcoholics Anonymous (AA) or Narcotics Anonymous (NA), coupled with participation in a formal program or individual psychotherapy.

For individuals who become dependent on opioids in the context of medical treatment, the subsequent course of the disorder depends largely on the medical

problems that generated the opioid use, the willingness of doctors to continue to prescribe drugs for them, and whether the drug is used orally or by injection. Patients who use opioids orally and whose pain is controlled by adequate doses of legitimately prescribed medication may experience little interference with normal function for many years. If pain is uncontrolled, demands for increased or altered medication schedules sometimes leads to repeated diagnostic workups, surgery, and other procedures, punctuated by attempts at withdrawal—voluntary and coerced—and treatment for depression.

General Principles A person's willingness to accept treatment may change over time as life circumstances, family relationships, and the severity and complications of the dependence change. Consideration should be given not only to the characteristics, wishes, and previous experiences of patients but also to how they are likely to react to the particular treatments that are economically and geographically feasible.

The clinician should make it clear that treatment requires a commitment to a long-term change in life-style, attitude, family dynamics, and sometimes even geographic location, and that the responsibility for making the changes belongs to the patient. If resources permit, the clinician should not rest content with making a diagnosis of opioid dependence, but should make a complete psychiatric assessment and take a thorough drug-use history. Most opioid users have additional psychiatric disorders and other substance-related disorders (Table 11.10-5, Table 11.10-6, and Table 11.10-7). The severity of psychological difficulties (e.g., mood disorders, anxiety disorders, paranoid ideation) and patterns of using drugs other than opioids are predictors of outcome in treatment. Patients should also understand, however, that although antecedent stress, environmental conditions, and underlying psychological difficulties may have played important roles in the genesis of their drug dependence, once it is established, opioid dependence will not resolve spontaneously even if those conditions are improved. Clinicians must communicate the necessity for treating the pathological drug use as a disorder in its own right. Opioid addicts who enter therapeutic communities under pressure from the criminal justice system are almost as likely to benefit from treatment as are those who enter for other reasons. Court pressure enhances retention in treatment, and retention for more than 90 days is a predictor of positive outcome.

Detoxification as Treatment Because brief detoxification alone is generally followed by relapse, some experts believe that detoxification should be provided only as part of a more comprehensive treatment effort. Others believe that brief detoxification should be available for those addicts who want only a limited service. At present it is not clear whether extended ambulatory detoxification (i.e., 180 days) should be viewed as detoxification or a distinct treatment approach.

Counseling Counseling and family and individual psychotherapy following opioid withdrawal are usually used but have not been evaluated satisfactorily. Each of these therapeutic techniques has been evaluated in the context of methadone maintenance treatment programs, where scheduled counseling has been shown to reduce illicit drug use substantially and to increase prosocial behaviors. The efficacy of counseling is further supported by studies showing that the quality of counseling and rapport between patient and counselor have a significant impact on outcome measures. When the dosage of methadone was held constant (e.g., at 60 mg), the addition of drug counseling on a regular basis was distinctly more effective than was emergency counseling in reducing opioid and cocaine use. When patients were provided with on-site medical, psychiatric, employment, and family counseling, the outcome was improved still further. It is reasonable to assume that the quality of counseling, psychiatric, and social services can have an impact on how opioid users not being maintained on methadone respond to treatment. In one study, detoxified heroin addicts who participated in a support group where they learned how to avoid or to cope with situations that provoked drug cravings or feelings of withdrawal (relapse prevention) had lower relapse rates than those assigned to control groups.

Self-Help NA is a self-help group of abstinent drug addicts modeled on the 12-step principles of AA. Such groups now exist in most large cities and can provide useful group support. The outcome for physicians treated in 12-step programs is generally good, but the anonymity that is at the core of the 12-step model has made detailed evaluation of its efficacy in treating opioid dependence difficult.

Opioid Antagonists The use of opioid antagonists to treat opioid dependence was originally based on the assumption that classically conditioned withdrawal symptoms and operantly reinforced drug-seeking behavior contribute to the high relapse rate typically seen after withdrawal from opioids. Theoretically, by blocking the euphoric effects of opioids, treatment with antagonists would lead to the extinction of operantly reinforced drug seeking; by preventing the reestablishment of physical dependence, treatment with antagonists also leads to the eventual extinction of conditioned withdrawal phenomena. These drugs do produce the expected blockade of opioid effects and their toxicity is low. Recent treatment with antagonists has been based on empirical and laboratory observations that patients taking naltrexone experience less craving in the presence of opioid-related cues, presumably because on a cognitive basis they are aware that they are unable to experience the opioid effects.

Naltrexone has been used in a number of clinical trials, but in each of them the dropout rate has been high. Curiously, these very low compliance rates have not been characteristic of alcohol abusers treated with naltrexone.

Naltrexone would appear to be an ideal drug for use in therapy. It is orally effective and when given three times a week (100 mg on weekdays and 150 mg on weekends) it completely blocks the effects of substantial doses of heroin. When naltrexone was tested in a multiclinic, double-blind study, however, very few of those patients who initially expressed an interest in the drug actually took a single dose. At the 6-month follow-up few differences were noted between the naltrexone group and the placebo control group, although the naltrexone group had fewer urine specimens positive for opioids while they were in treatment. In other clinical trials patients continued to take naltrexone for an average of 6 to 8 weeks. A double-blind study in Spain found no differences in drug use or retention in treatment between naltrexone and placebo groups. Former opioid users sometimes report dysphoria, an adverse effect not reported when the drug is given to alcohol abusers. The only consistent side effect reported is a somewhat higher incidence of nausea.

Experienced clinicians do not think that double-blind placebo-controlled trials are the most appropriate way to assess the utility of long-acting antagonists. They believe that a period of 30 to 60 days of treatment with naltrexone immediately after detoxification reduces the probability of relapse to opioid use. In open studies of opioid users on probation who volunteered to take naltrexone, those on active drug were far less likely than those randomly assigned to placebo to violate probation and be reincarcerated. In studies conducted in Israel and Portugal, special efforts were made to involve family members in supporting patients' compliance with naltrexone treatment, resulting in 30 to 40 percent continued abstinence rates at 1-year follow-up. However, in a controlled study conducted in New Haven, Connecticut, neither family therapy nor cash-equivalent vouchers provided contingently improved compliance. New techniques for facilitating the transfer from heroin or methadone to naltrexone (ultrarapid detoxification, clonidine, or buprenorphine and clonidine) may increase the ease of initiating naltrexone treatment, but low compliance and high dropout rates remain unsolved problems. Depot preparations of naltrexone and a somewhat longer-acting antagonist, nalmefine, are under study. In a small number of volunteers a single injection of naltrexone encased in microcapsules has been found to produce blood concentrations adequate to block the effects of street-quality heroin for about 30 days.

Opioid Maintenance In the United States about 750 outpatient methadone maintenance programs operating in 40 states and territories are now providing treatment with methadone or levomethadyl acetate to more than 110,000 patients at any given time. Methadone maintenance is also used in the United Kingdom, Australia, Hong Kong, and a number of European countries. In France, opioid addiction is now being treated with buprenorphine. Thus, opioid maintenance is a major modality for treating opioid dependence throughout much of the world.

Initiated by Vincent Dole and Marie Nyswander in 1964, the maintenance approach postulated that high doses of methadone would alleviate drug hunger and simultaneously block, by means of cross-tolerance, any euphoria produced by self-administered heroin. Thus, opioid users would be freed from the preoccupation with drug-seeking behavior and, with help and rehabilitation, could channel their energies into more productive avenues. Dole and Nyswander further postulated that opioid dependence is unrelated to antecedent psychological difficulties and that most of the traits of instability and unreliability seen in addicts are the consequence rather than the cause of their opioid addiction. Although the results obtained by other programs using methadone were often less dramatic as program criteria for entry were broadened, more disturbed and less motivated patients were admitted, and lower doses of methadone were prescribed, patients still showed substantial decreases in drug use and criminality and increases in health and social functioning.

General Pharmacology Given acutely, methadone is a typical μ-receptor agonist producing euphoria, analgesia, and other typical morphine-like effects. Given long-term by the oral route, however, methadone has several interesting properties that make it unusually useful in maintenance programs. These qualities include its reliable absorption and bioavailability after oral administration, the delay of peak plasma levels until 2 to 6 hours after ingestion, and the apparent nonspecific binding to tissues that creates a large reservoir of methadone in the body. The large reservoir, combined with slow time-to-peak effects, buffers the patients against sharp peaks in subjective effects after ingestion, which in any event are highly attenuated as a result of tolerance. The reservoir of methadone also tends to minimize any sharp declines that would induce withdrawal. Thus, not only is the administration of methadone on a once-a-day schedule possible, but minor variations in dosage over short periods do not induce major changes in biological effects. Although the mean plasma half-life in naive subject is about 15 hours, it ranges from approximately 22 to 56 hours in methadone-maintained patients depending on the measurement technique used. A number of commonly used therapeutic agents can induce liver enzymes and thereby result in more rapid metabolism of methadone and in half-lives significantly shorter than 22 hours; some agents can inhibit metabolism.

Standard Procedures and Government Regulations Since 1972 opioid maintenance programs have been required to operate according to detailed federal, state, and, sometimes, local regulations. Under recently revised federal regulations, opioid-dependent persons who have been dependent (i.e., physiologically addicted) for less than 1 year can be detoxified using opioids, but cannot be maintained beyond the detoxification period without an opioid-free interval. Persons less than 18 years old must have had two prior treatment failures and verified opioid dependence (i.e., physiological dependence) at admission, and have parental consent. Currently, regulations also govern maximum take-home dosage and the duration of daily clinic treatment before take-home dosage may be provided.

These regulations and the entire rationale for special regulations were the subject of a study by the Institute of Medicine (IOM) of the National Academy of Sciences and of a National Institutes of Health (NIH) Consensus Development Conference on the Effective Treatment of Heroin Addiction. The IOM concluded that federal regulations overestimated the danger of diversion of methadone or levomethadyl acetate, overregulated medical practice, and restricted access to treatment. The IOM report recommended sharply reducing the extent of federal regulation and the substitution of guidelines for regulations. The Department of Health and Human Services (HHS) is now proposing to test a new model of federal supervision in which programs and individual practitioners would be accredited by organizations such as the Joint Commission on the Accreditation of Health Care Organizations (JCAHO) and Commission for Accreditation of Rehabilitation Facilities (CARF) using standards based on expert medical opinion. Accredited programs would then be licensed by the Drug Enforcement Agency (DEA), which would retain its independent statutory authority to oversee the security of opioid drugs used by narcotic treatment programs. Since many states and local jurisdictions also have detailed regulations for opioid maintenance programs, the benefit that would flow from a reduction of federal regulations would vary from state to state. It is possible that the accreditation process will provide for opioid maintenance treatment by qualified individual practitioners.

Methadone Dose and Philosophical Differences Among Programs Methadone maintenance programs vary with respect to dosage, attitudes toward continued antisocial behavior, medical and social services provided, and long-term goals. The original programs emphasized methadone dosage sufficient to suppress opioid drug hunger and to induce a cross-tolerance blockade of the effects of illicit opioids, usually 80 to 120 mg a day. Patients were encouraged to remain on methadone indefinitely on the assumption that the return of opioid drug hunger following detoxification would lead to relapse and the loss of any gains achieved during treatment.

Other programs used lower doses of methadone (20 to 60 mg), which were often adequate for partially suppressing drug-seeking behavior, but not for producing adequate cross-tolerance to large doses of heroin. Such programs generally viewed maintenance as a transitional stage to eventual detoxification. Although patients may remain in treatment in these programs indefinitely, the ambience is often more supportive of efforts at gradual withdrawal. Empirically, patients maintained on lower doses of methadone are far more likely to discontinue treatment, and those who do so are unlikely to achieve sustained abstinence from opioids. In a six-program study in the United States the number of patients using intravenous heroin steadily increased after treatment was discontinued. By 12 months 80 percent reported heroin use in the month prior to the interview. Since only a small percentage of these persons were actually discharged as having completed treatment, the figure is perhaps a better illustration of what happens to those who drop out.

Since the publication of the findings on the importance of dose in reducing opioid use and increasing retention in treatment, more programs are prescribing daily doses of methadone of 60 mg or higher. However, the proper dose is not a matter of reaching some arbitrary number. There are wide variations in rates of methadone metabolism, and some experts believe that more attention should be given to using laboratory measures of plasma levels to adjust daily dosage. Data suggest that an average level of 400 ng/mL seems adequate and that trough levels below 150 ng/mL are likely to be associated with some degree of withdrawal or drug hunger. Clinicians should be aware that a variety of therapeutic agents can cause a decrease in plasma levels of methadone by inducing hepatic enzymes (CYP 3A4) that metabolize methadone. Among those are rifampin (Rifadin), phenytoin (Dilantin), barbiturates, carbamazepine (Tegretol), and ethyl alcohol. Valproic acid does not accelerate metabolism. Agents that can inhibit methadone metabolism include erythromycin, cimetidine, and ketoconazole. Zidovudine (Retrovir) does not alter methadone metabolism, but patients on methadone tend to have higher zidovudine levels than do controls. More methadone is excreted when urinary pH is low than when it is high; stress and other factors that lower urinary pH can result in lower plasma levels of methadone.

In both high-dosage and low-dosage methadone programs there is often a continuing struggle with patients over take-home medication. Patients who do not respond to treatment generally continue to use opioid and nonopioid illicit drugs, use alcohol to excess, attend the programs very irregularly, or exhibit combative behavior at the clinic. Such patients pose a dilemma for both types of methadone maintenance programs because they would probably fare worse if discharged. However, permitting them to remain demoralizes the staff and the other patients and often increases drug use among the latter.

Long-Term Effects The relative safety of methadone maintenance in terms of organ toxicity has been firmly established. Tolerance to many of the opioid agonist actions of methadone is incomplete, however, and continuing pharmacological effects are observed. Many of these effects, such as euphoria, drowsiness, and somnolence, are more prominent in the first weeks of treatment; if the dosage level is increased too rapidly, the effects even become manifested at later points in treatment. Some effects, however, persist even after many months of treatment. Among the most common long-lasting effects are constipation, which can sometimes result in fecal impaction and intestinal obstruction; excessive sweating; complaints of decreased libido; and sexual dysfunction, e.g., an inability to sustain an erection. Opioids reduce plasma levels of testosterone and follicle-stimulating hormone for which tolerance is often incomplete, but the correlation between abnormally low plasma levels of hormones and sexual dysfunction is not high.

In general, however, indexes of hypothalamic-pituitary-adrenal axis function are normal in patients maintained on long-term methadone, in contrast with the deranged function typical of active and recently detoxified heroin users. Both sleep abnormalities (insomnia and nightmares) and altered electroencephalogram (EEG) sleep patterns frequently are found during the first months of methadone treatment; the EEG sleep patterns appear to return to baseline, but complaints of sleep abnormalities may persist. For a substantial percentage of patients, even those participating in programs where dosage is adjusted, the major adverse effect is mild to moderate withdrawal symptoms prior to the next dose.

Tolerance does not develop in all patients to the mood-elevating effects of methadone. In double-blind studies some patients regularly report a greater sense of well-being a few hours after ingesting their daily dose; for some patients this effect may be an important factor in retention in treatment.

In-Treatment Outcome The majority of patients treated in methadone programs show significantly decreased opioid and nonopioid drug use, criminal behavior, and symptoms of depression and increased gainful employment. Significant differences in effectiveness across programs are due in some measure to the characteristics of patients treated, but certain program features tend to make some programs more effective than others. In 1980 the average retention rate for a group of methadone clinics participating in a national prospective study was 81 percent at 1 month, 67 percent at 3 months, and 52 percent at 6 months. Higher retention rates are associated with the provision of high-quality social services (especially within the first months), higher doses of methadone (60 mg or more) and allowing patients to know their dosage, the ease of accessibility, no fees or very low fees, and the use of supportive rather than confrontational techniques.

It bears repeating that the effectiveness of opioid maintenance treatment is powerfully influenced by the quality of the additional services provided. Effective individual drug counseling (e.g., one session per week) can result in significantly better outcome in terms of illicit opioid and cocaine use, needle-sharing, crime, employment, and psychological well-being. Treatment outcome can be further improved by the provision of psychiatric services, employment counseling, and family therapy. However, such additional services increase the overall cost of treatment, and in terms of overall improved outcomes it is more cost effective to ensure that all patients have competent standard drug counseling. Providing additional group therapy to patients in maintenance programs does not appear to produce better outcomes.

It has been confirmed repeatedly that positive behavioral change, including decreases in illicit drug use and other criminal behavior, does not typically occur immediately upon entry into treatment but takes place over a period of many months. Treatments lasting less than 90 days usually have little or no impact; consequently, retention in treatment is critical. Several studies have shown a direct relation between methadone dose and the probability of treatment retention.

In Australia patients on doses of 80 mg or higher were twice as likely to remain in treatment as those receiving 60 to 79 mg, who were twice as likely to remain as those receiving less than 60 mg. Dosage has also been shown to have a profound effect on whether or not patients will continue to use illicit heroin. In a study of 6 clinics in the United States there was an inverse relation between methadone dosage (over the range from 20 to £ 80 mg daily) and the percentage of patients using heroin. Although some programs persist in using low doses that have been shown to correlate with both high dropout rates and continued heroin use, their number has decreased as the data on the importance of adequate dosage has become more generally accepted.

Other program factors that influence outcome and retention are the perceived range and quality of the social services provided to the patient early in the course of treatment, whether the program is flexible or confrontational and punitive about occasional illicit drug use, and the competence and quality of the program leadership and the counselors. Outcome correlates with services actually delivered and the degree to which patients perceive that the services are those they believe are important to them. Some of the better-funded programs with better staff-patient ratios did not deliver as many services as did less well-endowed programs. The retention of drug users in treatment programs and the successful reduction of injecting drug use are particularly important in an era when such drug use often results

in transmitting or acquiring HIV infection. Older, black, married, and employed patients tend to remain in methadone programs longer; patients with extensive criminal backgrounds tend to drop out sooner and to perform more poorly while in treatment. The severity or duration of opioid use does not correlate with retention or performance in treatment, and patients who enter treatment under what they perceive as legal coercion show improvement comparable to that of patients who report no such external pressure.

Detoxification and Long-Term Outcome The percentage of patients remaining abstinent from opioids 12 to 36 months after successful detoxification from opioid maintenance has ranged from 12 to 28 percent for unselected samples of patients, some of whom were discharged for violation of clinic rules. When analysis is restricted to those patients who elected to be withdrawn (with staff and patient consensus that treatment was completed), abstinence rates were substantially higher, at least over the first 6 to 12 months. Predictors of retention and positive outcome in treatment do not necessarily predict success in achieving abstinence or long-term positive outcome once withdrawal has been completed.

In general, patients with shorter drug histories, longer time in maintenance treatment but at lower dosages seem to have more success detoxifying. In one national, multiclinic follow-up study 40 percent of former methadone patients interviewed were not using any illicit opioids and did not have any other significant drug problems 6 years after the completion of initial treatment. Other follow-up studies have found a smaller proportion of former maintenance patients who were opioid abstinent.

Although the percentage of heroin addicts maintained on methadone who eventually achieve long-term stable abstinence is not high, there is no convincing evidence that the likelihood of abstinence is higher among those treated in other modalities once adjustment is made for various pretreatment predictors of outcome. In one study where methadone treatment appeared to impair eventual abstinence, that disadvantage was offset by substantially lower rates of incarceration among methadone-treated patients.

Other Opioid Maintenance Drugs

LEVOMETHADYL ACETATE This drug is a μ-receptor agonist. It was approved for use in the treatment of opioid dependence in 1993. Levomethadyl acetate is similar to methadone in its pharmacological actions, and it is converted into the active metabolites nor-acetylmethadol and di-nor-acetylmethadol that have very long biological half-lives, (e.g., 48 to 96 hours for di-nor-acetylmethadol). Figure 11.10-4 shows that the plasma levels of nor-levomethadyl acetate and di-nor-levomethadyl acetate persist for many hours after a single dose. Consequently, levomethadyl acetate can be given as infrequently as three times a week, thereby reducing the inconvenience of attending a clinic daily to ingest the drug and simultaneously reducing concerns about illicit diversion. When levomethadyl acetate is abruptly discontinued, the withdrawal syndrome is slow in onset and relatively mild in intensity but it is at least as protracted as that of methadone. Since the mid-1970s in double-blind as well as large-scale, multicenter, open studies, levomethadyl acetate has been shown to be equivalent to methadone in suppressing illicit opioid use and encouraging productive activity. Prior to its approval by the Food and Drug Administration (FDA) a consistent finding was that retention in treatment was lower with levomethadyl acetate than with methadone; more recently clinicians are reporting that some patients prefer levomethadyl acetate.

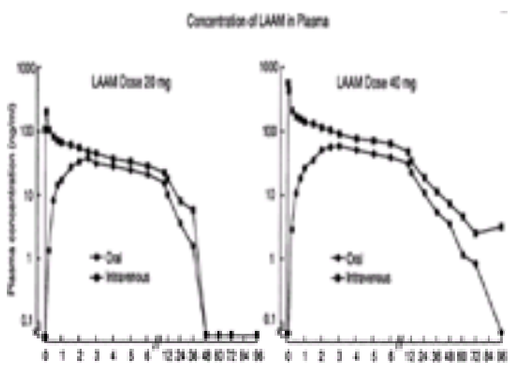

FIGURE 11.10-4 Mean (*N* = 6) concentrations of nor-LAAM and dinor-LAAM after single doses (oral and intravenous) of 40 mg LAAM. The hatch marks on the x-axis indicate a change from 30-minute intervals to 12-hour intervals. (Adapted from Walsh PT, Johnson RE, Cone EJ, Bigelow GE: Intravenous and oral L-alpha-acetylmethadol: Pharmacodynamics and pharmacokinetics in humans. J Pharmacol Exper Ther *285:*71, 1998.)

A random assignment study comparing three dose levels of levomethadyl acetate (25, 25, and 35 mg; 50, 50, and 70 mg; and 100, 100, and 140 mg—all given three times a week) demonstrated that the higher doses were more effective in suppressing heroin use. Patients receiving the medium dose were less likely to drop out of treatment, probably because those on the highest dose felt oversedated. In most cases it is unnecessary to first stabilize heroin users on methadone, and levomethadyl acetate can be used as the initial treatment agent.

Levomethadyl acetate is more complex pharmacologically than methadone and its use demands a more skilled clinician than does methadone. Currently the conditions attached to its use prohibit take-home doses of levomethadyl acetate. This restriction is likely to reduce its utility, except in situations such as long-term detoxification and interim maintenance where no take-home methadone is permitted currently.

BUPRENORPHINE This is a partial μ-receptor agonist that is now available as an analgesic and has been proposed as an alternative to methadone in the treatment of heroin addiction, both for detoxification and for maintenance. It has also been explored in the treatment of concurrent cocaine and opioid dependence. Buprenorphine produces morphine-like effects at low doses, but even when the dose is increased, the intensity of its actions does not seem to exceed that achieved with 30 to 60 mg of morphine. When buprenorphine is given sublingually, morphinelike subjective effects appear to reach a ceiling at about 8 to 12 mg, with only modest increases in effects at 16 and 32 mg. Over this dosage range, significant respiratory depression was not observed. Because of this apparent ceiling, the risk of overdose may be limited. After repeated administration, buprenorphine attenuates or blocks the subjective effects of parenterally administered morphine or heroin. When buprenorphine given sublingually 8 mg a day for several weeks is abruptly discontinued, a generally mild opioid withdrawal syndrome develops.

Buprenorphine has been compared with methadone as a maintenance agent in several double-blind outpatient studies. Although optimal dose ranges have not been established, one study found 8 mg of buprenorphine sublingually to be comparable to 60 mg of methadone orally on measures of retaining patients in treatment and suppressing heroin use. Withdrawal symptoms when buprenorphine was discontinued were mild. In similar studies 8 mg of buprenorphine was significantly less effective than was 80 mg of methadone and was about comparable to 30 mg of methadone; in another study 8.9 mg of buprenorphine was comparable to 54 mg of methadone. In none of those double-blind studies did buprenorphine show superiority in reducing cocaine use. Some patients express a clear preference for buprenorphine over methadone. With higher doses (e.g., 16 to 32 mg sublingually), some patients experience no withdrawal for 48 to 72 hours, permitting a reduction in clinic visits. Buprenorphine may become the treatment drug of choice for most heroin users, although because of its ceiling effects there may be patients who will still require full agonists such as methadone or levomethadyl acetate. In France buprenorphine is now the most widely used maintenance drug for opioid dependence, with about 40,000 patients in treatment.

A formulation combining buprenorphine and naloxone in a 4:1 ratio (SuboxoneR), has completed clinical testing and will be submitted for approval to the FDA. Given sublingually, the naloxone has virtually no effects in opioid-dependent patients; however, if an attempt is made to inject the combintion, the naloxone precipitates withdrawal symptoms. This formulation will have a lower abuse liability than buprenorphine alone.

In a large multisite, random-assignment, double-blind study, 8 mg of sublingual buprenorphine was found to be significantly better than 1 mg. Patients treated with 16 mg were more likely to have 13 weeks of consecutive urine tests negative for illicit opioids than those receiving 1 mg or 4 mg; patients on 16 mg per day were also somewhat less likely to drop out of treatment.

Heroin Maintenance Versus Methadone. Intravenous heroin and oral methadone were compared in a random-assignment study at a London clinic. Most patients given heroin continued to inject opioids and stayed involved with the drug culture. Some of the heroin patients sold part of their prescriptions; others supplemented clinic supplies with opioids from illicit sources. Some of the patients assigned to oral methadone maintenance refused to participate in the study and left treatment immediately; many others left subsequently. At 12 months only 29 percent of oral methadone patients were still at the clinic; of those who left, 40 percent (28 of those initially assigned to oral methadone) were no longer using opioids regularly at the 1-year follow-up. Over the 12-month period more heroin patients than oral

methadone patients died or were admitted to hospitals for drug-related problems. A strong relation was found between criminal activity and continued illicit opioid use, although differences in baseline rates of criminality make it difficult to evaluate the net impact of the two types of treatment on crime.

HEROIN TRIALS. In the early 1990s Switzerland experienced a very sharp rise in heroin use and drug-related crime. This occurred in spite of (or perhaps because of) a policy of toleration begun in 1986 that provided addicts with areas where they could use drugs without fear of arrest (Platzspitz in Zurich and other “needle parks” elsewhere). This seemed only to have acted as a magnet for European drug users. Addiction rates continued to rise along with overdose deaths and the incidence of HIV infection. Efforts to reverse the trend by strict law enforcement and by expanding residential treatment were also less than fully successful in slowing the rise in crime and overdose deaths. In addition, methadone maintenance programs had tripled between 1986 and 1990. With the closing of Platzspitz in 1992, the various Cantons initiated low-threshold methadone prescribing, but many heroin users did not avail themselves of these treatment opportunities.

It was in this context that the Swiss government initiated a research program to test the feasibility of prescribing heroin to confirmed heroin addicts. An initial random-assignment phase indicated that patients strongly preferred intravenous heroin to intravenous morphine or methadone. During the demonstration phase, which began in 1993, patients admitted to 18 treatment centers in several Swiss cities came to specialized clinics several times a day and injected as much heroin as they wished. For most patients heroin dosage first escalated rapidly but after several weeks or months seemed to stabilize, albeit in some cases at up to 900 mg a day intravenously. Most patients came to clinic several times a day. Medical and psychiatric care as well as social work support were readily available. Some patients were given subsidized jobs; all were eligible for welfare assistance. Investigators reported that patients' overall physical and mental health improved. Earnings improved somewhat, and criminal activity declined from 69 to 10 percent, as indicated by police records. Since urine was tested for drugs only every 8 weeks, it is not certain that illicit drug use actually decreased, but patients reported a very sharp decrease in illicit heroin and cocaine use. Reported use of benzodiazepines declined very slowly, and alcohol and cannabis use declined very little. The dropout rate was relatively low, with 69 percent retention at 18 months. Eleven new cases of HIV or hepatitis were recorded, probably due to illicit cocaine use. The annual death rate of 1 percent was far lower than for untreated addicts and lower than for many other treatment programs. Because of the extensive support services and the long clinic hours, the overall cost of operating these clinics, (about $23 per patient a day), was several times higher than the cost of comparable oral methadone maintenance programs. Similar programs would seem to be inappropriate except in areas where there is already easy access to all other forms of treatment. evertheless, the governments of The Netherlands and Australia have been seriously considering the initiation of similar “research” projects.

Therapeutic Communities The underlying philosophy of the more than 300 therapeutic communities that have evolved from Synanon is that the drug addict is emotionally immature and requires total immersion in a specialized social structure in order to modify lifelong, destructive behavioral patterns. The goal is to effect a complete change of lifestyle, including abstinence from drugs, the development of personal honesty and useful social skills, and the elimination of antisocial attitudes and criminal behavior.

To achieve these objectives, the addict was expected to live in the therapeutic community for approximately 12 to 18 months and to participate in frequent group sessions devoted to mutual criticisms of the attitudes and behavior of the participants. The community also acts in many respects as a substitute family. Assumption of responsibility within the community is rewarded with increased personal freedom, material comfort, and the respect of peers. Deviation from community expectations, in terms of behavior or attitude, frequently results in harsh criticisms by staff or sometimes by the entire community. Violence and any form of drug use are totally prohibited and may result in expulsion from the community—the ultimate punishment. Although most therapeutic communities avoid the use of drugs even to ease initial withdrawal, some are more flexible regarding treatment with drugs for heavily dependent new entrants. More recently, some therapeutic communities have recognized that a high percentage of addicts have psychopathology in addition to drug dependence and thus may permit the use of prescribed antipsychotic or antidepressant medication.

Present-day therapeutic communities vary considerably in their attitudes toward professionals and in actual staffing patterns. In every community, however, at least a few exaddicts are employed as key personnel on the staff; the exaddicts serve as role models and visible evidence that recovery and acceptance are possible and expected. Some therapeutic communities, such as Phoenix House, Odyssey, and Second Genesis, are or were directed by psychiatrists and employ a number of health professionals in key positions. Quite recently a number of federally supported programs have begun to develop individualized treatment plans and to use health care professionals, such as physicians, psychologists, or master's degree–level counselors or social workers, to make initial assessments.

Although many therapeutic communities still expect residents to return to the general community after 12 to 18 months, some are experimenting with shorter periods of residence, sometimes because of pressure from funding sources. Because they require so long a period of residence, therapeutic communities obviously have little appeal to those opioid addicts who have stable and gainful employment and satisfactory personal relationships. Currently, a substantial proportion of entrants into therapeutic communities are referred by the criminal justice system. Criteria for entry have thus been modified and external pressure to remain in treatment has been increased. Despite the external pressure, dropout rates are still high: 30 to 40 percent in the first 90 days and 50 percent within 6 months. Furthermore, the percentage of new entrants with severe psychopathology has been gradually increasing. There is substantial variation in drop-out rates among ostensibly similar therapeutic communities, even after adjusting for patient characteristics.

Residence in the therapeutic community results in major reductions in drug-use problems; indicators of depression also decrease significantly. Follow-up studies of graduates and dropouts indicate that patients remaining 90 days or longer exhibit significant decreases in self-reported antisocial behavior, illicit drug use, and recorded arrests, as well as substantial increases in legitimate employment. Improvement was seen in patients who met DSM-III-R criteria for antisocial personality to about the same extent as those who do not. In general, for those without severe psychopathology there is a consistent time-in-program effect up to about 12 months; patients who stay longer exhibit better outcomes along all dimensions at 12-month and 5-year follow-up intervals. In some cases patients found to be doing well have had additional treatment in other programs since leaving the therapeutic community.

Outpatient Drug-Free Programs Such treatment programs often subscribe to the same goals as the residential or therapeutic communities, but they attempt to achieve those goals in a less than 24-hour-a-day setting. The programs vary widely in staffing patterns, philosophy, and program content; they range from highly organized daytime therapeutic communities to drop-in centers offering conversational sessions and recreational activities. Outpatient drug-free programs tend to deal more with multiple-substance abusers than with confirmed heroin addicts, although some programs treat both. Long-term follow-up studies suggest that these programs have an impact beyond what is seen with detoxification alone, but differences in patient characteristics make valid comparisons difficult.

Supplements and alternatives to outpatient drug-free therapies must always be mentioned. They include self-help groups, such as AA and NA, as well as group therapies focusing on cognitive skills for coping with crises and avoiding situations that engender craving (relapse prevention).

Psychotherapies As measured by use of illicit drugs, need for ancillary psychiatric medicines, scores on scales of psychological distress, and amount of legitimate money earned, patients in methadone maintenance programs appear to benefit more when individual, analytically oriented, supportive-expressive, or cognitive-behavioral psychotherapy is added to standard drug counseling than from counseling alone. With such therapy the very bleak prognosis for patients with the most severe psychopathology can be improved. To best engage patients in individual therapy, the therapy should be started early and be an integral part of the program. In general, patients with antisocial personality disorder without depression do not derive additional benefit, but even among that group, those who seem capable of developing an alliance with the therapist appear to do better than those who do not.

Among patients maintained on methadone, one controlled study suggests that skillful family therapy in conjunction with urine monitoring is superior to standard drug counseling in fostering decreased illicit drug use. However, from a public health perspective, such additional psychotherapy for patients treated with adequate doses of methadone does not appear to be as cost effective as providing standard therapy (weekly drug counseling) to more patients.

There are no controlled studies on the efficacy of individual psychotherapy in treating opioid-using patients not stabilized on methadone. Experienced clinicians believe that psychotherapy can influence outcome in opioid users. Most experts now believe that therapists should not be passive but rather should be willing and able actively and empathetically to help patients look at their vulnerabilities. They also need to be flexible and to be able to communicate respect, empathy, and even admiration. At the same time, therapists must be alert to complex countertransferences that drug-using patients typically provoke. Therapists who see drug-dependent patients in office practice settings are increasingly recognizing the need to involve the family and to develop a supportive network to aid in their treatment. The therapeutic interventions in such cases are based less on individual psychoanalytic concepts than on cognitive-behavioral techniques that recognize the role of emotional and environmental cues in triggering craving and predisposing to slips and relapses. Although these techniques have been shown to be of value in controlled studies, their efficacy when transferred to the office practice setting remains unclear.

Treatment of Special Populations

Criminal Justice Clients The criminal justice system has complex interactions with drug-dependence treatment systems. Many patients now enter or remain in treatment because of direct coercion by the courts or correctional system.

Conversely, some patients on probation or parole may enter treatment against the wishes of their probation officers or in violation of some condition of parole. The impact of such coercion on retention in treatment or on outcome has been difficult to estimate because of baseline differences between criminal justice and noncriminal justice patients.

In addition to their relation to the major modalities of treatment, probation, parole, or both can be viewed in their own right as techniques that modify drug-using behavior. Close supervision and drug testing of parolees result in a substantial reduction in daily drug use. The efficacy of such testing and supervision depends in part on the way in which sanctions against use are arranged and the consistency with which such sanctions are exercised. Probation programs that involve use of naltrexone seem to reduce relapse to opioid use.

Civil commitment programs for opioid users involving long periods of institutional care, although still on the law books in some states, have lost popularity. They were exceedingly expensive as compared with all other treatment modalities and there is no evidence of their long-term efficacy. In contrast, there is renewed interest in treating people in jails and prisons who have been drug dependent. Former prisoners who, while incarcerated, participated in programs based on therapeutic community principles tend to have lower rates of recidivism than do matched controls, especially if they receive additional treatment in the community immediately following release.

Pregnant Women In general, opioid use during pregnancy is associated with decreased fetal growth but not with any teratological effects. What is clear, however, is that most women who continue to use heroin while pregnant do not get prenatal care and deliver low-weight babies with an elevated risk for morbidity and mortality. Babies who survive may be infected with HIV and other diseases as a result of maternal high-risk behavior. Most of these risks can be substantially reduced if the mother can be retained in treatment and provided with prenatal care. Sometimes this can be accomplished in drug-free outpatient or residential programs. Utilizing such programs usually requires initial detoxification, which must be undertaken with considerable caution because severe opioid withdrawal early in pregnancy (before 14 weeks) can induce abortion and late in pregnancy (after week 32) can induce fetal distress. If a woman is already maintained on methadone, it is not necessary to undertake withdrawal, but if withdrawal is elected, it should be done slowly (5 mg every other week) and, if possible, in collaboration with clinicians experienced in perinatal addiction. Dose adjustments (increases) may be needed for pregnant women maintained on methadone because of changes in blood volume and metabolism. Babies born to mothers who have been maintained on methadone usually require treatment for neonatal opioid withdrawal. However, if the use of illicit drugs an be eliminated and adequate prenatal care provided, infant mortality and morbidity are reduced, and in the long run the development and cognitive functioning of the baby are not likely to be impaired.

Health Professionals. Health professionals often come to treatment when their drug use is discovered by colleagues or by drug enforcement agencies. Treatment generally consists of a contingency contract under which they agree to frequent drug testing and the certain revocation of their license to practice if they continue to use drugs. This contract is typically supplemented by encouragement to participate in a 12-step program or other therapy. Supervised naltrexone may occasionally be part of the regimen. Under supervision, most (up to 85 percent) of these patients recover and are able to return to practice. Returning to specialties with easy access to opioids can pose a serious hazard to continued abstinence. For example, in one study only 34 percent of anesthesiology residents who were abusing opioids were able to return to their specialty. For a significant percentage of the residents, the first sign of relapse was a fatal overdose.

Other Special Populations. Certain groups of drug users are more likely to enter treatments that seem designed to deal with particular problems, as well as with drug use. For example, some women may feel that only a women's treatment program will deal adequately with issues of childhood sexual abuse, spousal physical abuse, or fears of losing custody of children. Native Americans, Hispanic Americans, or Americans of African or Asian descent may feel more comfortable in programs designed to be sensitive to their cultural values. To what degree such programs are more effective than less specialized programs in retaining patients in treatment or producing long-term behavioral change remains undetermined.

Opioid Dependence With Additional Psychopathology The severity of psychological disturbances—depression, anxiety, paranoid ideation—tends to predict the overall outcome of treatment for opioid dependence. Other predictors of outcome include alcohol abuse and the use of illicit nonopioid drugs, such as cocaine. Furthermore, patients in methadone programs who meet criteria for additional nonsubstance abuse Axis I diagnoses consistently exhibit substantially higher rates of lifetime and current diagnoses of dependence on drugs other than opioids. However, the literature is not entirely consistent on the prognostic significance of additional psychopathology. In a number of studies patients with additional psychopathology, including antisocial personality disorder, appear to do about as well in methadone programs or therapeutic communities as those without such diagnoses, at least in terms of reduced drug use.

Mood Disorders Some form of affective dysregulation is the most common Axis I psychiatric disorder not related to substance abuse found among opioid addicts in treatment (Table 11.10-5). Clinicians agree that those addicts with bipolar I or bipolar II disorders can benefit from treatment with lithium (Eskalith). Lithium has been used in combination with methadone without adverse interactions.

Although clinicians in private practice often prescribe antidepressant medications for opioid-dependent patients with major depressive disorder and minor depression, publicly supported programs often do not. The reasons for this vary. In the past, therapeutic communities were biased against the use of psychoactive agents; they seem to be less so now, and some therapeutic communities have psychiatric consultants and permit psychoactive medication when indicated. Methadone programs do not have a bias against pharmacotherapy, but many still do not commonly assess depressive symptoms, responding only when these symptoms become clinically obvious. Studies of the efficacy of antidepressant agents in the treatment of depressed opioid-dependent patients have yielded inconsistent results. It is possible that adequate antidepressant plasma concentrations may not have been achieved or that patients were not carefully selected.

In methadone-maintained patients whose depression antedated substance use or persisted for more than a month after admission to treatment, a double-blind study found that imipramine (Tofranil) produced significant improvements in mood, as well as self-reported decreases in the use of illicit drugs. However, as measured by urine tests, few patients achieved abstinence. In this study the imipramine dosage was gradually increased, depending on adverse effects and response, to a maximum of 300 mg a day.

Opioid addicts tend to experience a decrease in depressive symptoms after entering treatment, regardless of whether the treatment is an opioid maintenance program or a therapeutic community. Not every patient shows spontaneous improvement, however. Since the severity of psychological impairment is a predictor of overall treatment success, the conscientious clinician will pay particular attention to persistent depression.

Anxiety Disorders Anxiety disorders rarely receive specific treatment in programs devoted to opioid dependence. A very high percentage of addicts with an anxiety disorder also have dysphoria. When treatment seems indicated, antidepressant medications would appear to be the drugs of choice because several of these have antianxiety, antipanic, and antidepressant effects. Also, patients with anxiety disorders may abuse benzodiazepines, and cocaine use can cause cocaine-induced anxiety disorder with panic attacks that persist beyond the period of use. It is not yet clear what effect the use of cocaine by opioid users will have on the prevalence of anxiety disorders among that population.

Alcohol Abuse Alcoholism (alcohol abuse or dependence) is common among opioid addicts in treatment, with lifetime diagnosis rates of 25 to 40 percent. Most opioid addicts who also have alcohol abuse or dependence disorder have some form of mood disorder as well. Several studies suggest that there is an inverse relation between alcohol use and heroin use, with alcohol use increasing as heroin use decreases as a result of treatment—a situation that reverses with relapse. In one methadone program, neither twice-weekly sessions with special AA counselors nor behavioral modification sessions with psychologists were more helpful than a control condition of standard maintenance. However, most alcohol abusers decreased their alcohol consumption as well as their heroin use after entering treatment, and patients and counselors who participated in treatment were enthusiastic about group sessions and expressed the belief that they were beneficial. Disulfiram (Antabuse) can be combined with methadone without adverse effects. In small experimental studies, program privileges, such as take-home doses of methadone, have been made contingent on the ingestion of disulfiram. In controlled studies, however, disulfiram was not superior to placebo in modifying alcohol abuse or dependence among methadone-maintained patients. In some programs methadone itself was made contingent on results of breath alcohol tests. However, making a medicine for one disorder contingent on recovery from another disorder is controversial. Thus, although alcohol abuse or dependence while in treatment is reported to be a major predictor of reversion to opioid abuse following detoxification from maintenance methadone, it appears that there is no specific intervention that is demonstrably effective. In programs that do not use methadone, continued participation in a residential program sharply reduces alcohol use, but data on long-term outcome are lacking. Both outpatient drug-free and residential programs may employ AA and 12-step approaches to controlling alcohol abuse or dependence along

with procedures aimed at opioid use and behavioral change.

Schizophrenia and Other Psychotic Disorders A high percentage of patients (up to 48 percent in some reports) treated in urban hospitals for psychotic disorders, including schizophrenia, also meet the criteria for a substance-use disorder. In most instances, patients are abusing or are dependent on alcohol, cocaine, cannabis, or some combination of those drugs. Even in east coast urban areas where opioid abuse is common, opioid dependence is relatively uncommon among patients hospitalized for schizophrenia. Two hypotheses have been offered to account for that relatively low level. Either people with schizophrenia are not organized enough to be able to deal with the demands and stresses of obtaining and using heroin, or opioids have some ameliorative effects and help persons with schizophrenia avoid hospitalization—in contrast to alcohol, cocaine, and cannabis, which tend to exacerbate schizophrenia. There are case reports but as yet no controlled studies of opioid users whose psychotic or paranoid states responded to opioids but responded only poorly to more traditional antipsychotic agents. The endogenous opioids interact with dopaminergic systems, and it has been postulated that methadone has antimanic and antipsychotic actions. Methadone added to antipsychotic medication produces improvement in antipsychotic-resistant patients with paranoid schizophrenia. Heroin addicts treated with methadone may develop tolerance to these postulated antipsychotic effects of opioids so that after a period on methadone psychosis may break through again. Although lifetime rates for schizotypal features are 6 to 8 percent of patients in drug-abuse clinics, schizophrenia is uncommon—less than 1 percent. For those patients with schizophrenia, dopamine receptor antagonists are probably useful and can be combined with methadone. Some clinicians believe that underlying rage and hostility may play causal roles in opioid dependence and that some addicts use the opioids to control those affects. Clinicians have also observed that, although psychotic patients are reassured by the structure of the usual methadone clinic, patients with borderline personality disorder make the structure and the rules the objects of an all-out war, leading to their rapid and premature discharge from the clinic program.

Opioid Dependence With Other Substance Abuse Many opioid users regularly self-administer some other drug or drugs, (e.g., amphetamines, cocaine, alcohol, barbiturates, benzodiazepines, or cannabinoids). In general, multi-drug users have a greater range and severity of psychiatric problems. In therapeutic communities they create more behavior problems, drop out of treatment sooner, and have poorer posttreatment outcomes. Many polysubstance users are sufficiently dependent on heroin or other opioids to qualify for treatment in methadone programs. Competent counseling has been shown to decrease cocaine use, as has the provision of positive rewards such as vouchers or lottery tickets contingent on urine tests negative for cocaine.

A retrospective study of treatment outcome for polysubstance abuse as a function of the category of drugs abused and the treatment received compared male veterans enrolled in therapeutic community treatment with those in methadone maintenance. All subjects had greater-than-average psychopathology in addition to polysubstance abuse. The groups were further categorized into opioid-stimulant, opioid-depressant, and opioid-only users. At 6-month follow-up the opioid-only group had better outcomes on all measures than either of the other groups, with no clear superiority of either treatment approach. The opioid-stimulant group did significantly better in methadone maintenance and the opioid-depressant group did significantly better after treatment in a therapeutic community. The researchers inferred that recovery from the depression and cognitive impairment that typically accompany the abuse of sedatives was facilitated by the prolonged abstinence afforded by treatment in a therapeutic community. Specialized facilities using therapeutic community approaches have evolved to provide treatment to drug abusers with serious psychiatric disorders, especially those who are also homeless. Their results suggest that major reductions in drug use and rehospitalization can be achieved.

SUGGESTED CROSS-REFERENCES

An overview of substance-related disorders is given in Section 11.1. Alcohol-related disorders are discussed in Section 11.2, amphetamine-related disorders in Section 11.3, and cocaine-related disorders in Section 11.6. Drug and alcohol abuse among the elderly is discussed in Section 51.3h. Posttraumatic stress disorder is discussed in Chapter 15 and sexual dysfunction in Section 19.1a.

SECTION REFERENCES

Akil H, Meng F, Devine DP, Watson SJ: Molecular and neuroanatomical properties of the endogenous opioid system: Implications for treatment of opiate addiction. Semin Neurosci *9:*70, 1997.

Amass L, Bickel WK, Crean JP, Blake J, Higgins ST: Alternate-day buprenorphine dosing is preferred to daily dosing by opioid-dependent humans. Psychopharmacology *136:*217, 1998.

*American Psychiatric Association: Practice guideline for the treatment of patients with substance use disorders: Alcohol, cocaine, opioids. Am J Psychiatry *152*(Suppl):5, 1995.

American Psychiatric Association Commission on AIDS: Policy statement on needle exchange programs. Am J Psychiatry *153:*1660, 1996.

Bailey RC, Hser Y-I, Hsieh S-C, Anglin MD: Influences affecting maintenance and cessation of narcotics addiction. J Drug Issues *24:*249, 1994.

Ball J, Ross A: *The Effectiveness of Methadone Maintenance Treatment*. Springer-Verlag, New York, 1991.

Brewer C: Ultra-rapid, antagonist-precipitated opiate detoxification under general anaesthesia or sedation. Addict Biol *2:*291, 1997.

Brooner RK, King VL, Kidorf M, Schmidt CW, Bigelow GE: Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. Arch Gen Psychiatry *54:*71, 1997.

Cornish JW, Metzger D, Woody GE, Wilson D, McLellan AT, Vandergrift B, O'Brien CP: Naltrexone pharmacotherapy for opioid dependent Federal probationers. J Subst Abuse Treat *14:*529, 1997.

Darke S, Ross J: Polydrug dependence and psychiatric comorbidity among heroin injectors. Drug Alcohol Depend *48:*135, 1997.

De Leon G, Staines G, Sacks S: Passages: A therapeutic community-oriented day treatment model for methadone-maintained clients. J Drug Issues *27:*341, 1997.

des Jarlais DC, Hagan H, Friedman SR, Friedmann P, Goldberg D, Frischer M, Green S, Tunving K, Ljungberg B, Wodak A, Ross M, Purchase D, Millson ME, Myers T: Maintaining low HIV seroprevalence in populations of injecting drug users. JAMA *274:*1226, 1995.

Dyer KR, White JM: Patterns of symptom complaints in methadone maintenance patients. Addiction *92:*1445, 1997.

Eissenberg T, Bigelow GE, Strain EC, Walsh SL, Brooner RK, Stitzer ML, Johnson RE: Dose-related efficacy of levomethadyl acetate for treatment of opioid dependence. JAMA *277:*1945, 1997.

Ernst M, London ED: Brain imaging studies of drug abuse: Therapeutic implications. Semin Neurosci *9:*120, 1997.

Finnegan LP, Hagan TA, Kaltenbach K: Opioid dependence: Scientific foundations of clinical practice. Proc NY Acad Med *67:*223, 1991.

Fudala PJ: Levomethadyl acetate—Pharmacology and pharmacokinetics, developmental history, and therapeutic considerations. Subst Abuse *17:*127, 1996.

Fudala PJ, Jaffe JH, Dax EM, Johnson RE: Use of buprenorphine in the treatment of opioid addiction, II. Effects of daily and alternate-day administration and abrupt withdrawal. Clin Pharmacol Ther *47:*525, 1990.

Grella CE, Wugalter SE, Anglin MD: Predictors of treatment retention in enhanced and standard methadone maintenance treatment for HIV risk reduction. J Drug Issues *27:*203, 1997.

Institute of Medicine: *Federal Regulation of Methadone Treatment*. National Academy Press, Washington, DC, 1995.

*Jaffe JH, Knapp CM, Ciraulo DA: Opiates: Clinical aspects. In *Substance Abuse: A Comprehensive Textbook*, ed 3, JH Lowinson, P Ruiz, RB Millman, JG Langrod, editors. Williams & Wilkins, Baltimore, 1997.

Jansson LM, Svikis D, Lee J, Paluzzi P, Rutigliano P, Hackerman F: Pregnancy and addiction: A comprehensive care model. J Subst Abuse Treat *13:*321, 1996.

Johnson RE, Jaffe JH, Fudala PJ: A controlled trial of buprenorphine treatment for opioid dependence. JAMA *267:*2750, 1992.

Khantzian E, Halliday KS, McAuliffe WI: *Addiction and the Vulnerable Self*. Guilford, New York, 1990.

Kleber HD, Topazian M, Gaspari J, Riordan CE, Kosten T: Clonidine and naltrexone in the outpatient treatment of heroin withdrawal. Am J Drug Alcohol Abuse *13:*1, 1987.

Kraft MK, Rothbard AB, Hadley TR, McLellan AT, Asch DA: Are supplementary services provided during methadone maintenance really cost-effective? Am J Psychiatry *154:*1214, 1997.

*Kreek MJ: Rationale for maintenance pharmacotherapy of opiate dependence. In *Addictive States*, vol 30. CP O'Brien, JH Jaffe, editors. Research Publications, Association for Research in Nervous

and Mental Disease. Raven, New York, 1992.

Ling W, Charuvastra C, Collins JF, Batki S, Brown LS Jr, Kintaudi P, Wesson DR, McNicholas L, Tusel DJ, Malkerneker U, Renner JA Jr, Santos E, Casadonte P, Fye C, Stine S, Wang RIH, Segal D: Buprenorphine maintenance treatment of opiate dependence: A multicenter, randomized clinical trial. Addiction *93:*475, 1998.

Maddox JF, Desmond DP: Ten-year follow-up after admission to methadone maintenance. Am J Drug Alcohol Abuse *18:*289, 1992.

*McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP: The effects of psychosocial services in substance abuse treatment. JAMA *269:*1953, 1993.

Najavits LM, Weiss RD: The role of psychotherapy in the treatment of substance-use disorders. Harvard Rev Psychiatry *2:*84, 1994.

*Nestler EJ: Molecular mechanisms underlying opiate addiction: Implications for medication development. Semin Neurosci *9:*84, 1997.

Nestler EJ, Aghajanian GK: Molecular and cellular basis of addiction. Science *278:*58, 1997.

Neto D, Xavier M, Aguiar P, David M, Sardinha L, de Almeida C: Sequential combined treatment of heroin addicted patients in Portugal with naltrexone and family therapy. Eur Addict Res *3:*138, 1997.

Nunes EV, Quitkin FC, Donovan SJ, Deliyannides D, Ocepek-Welikson K, Koenig T, Brady R, McGrath PJ Woody G: Imipramine treatment of opiate-dependent patients with depressive disorders. Arch Gen Psychiatry *55:*153, 1998.

Nurco DN, Kinlock T, Balter MB: The severity of pre-addiction criminal behavior among urban, male narcotic addicts and two non-addicted control groups. J Res Crime Delinquency *30:*293, 1993.

O'Connor PG, Carroll KM, Shi JM, Schottenfeld RS, Kosten TR, Rounsaville BJ: Three methods of opioid detoxification in a primary care setting. Ann Intern Med *127:*526, 1997.

Pani PP, Trogu E, Contu P, Agus A, Gessa GL: Psychiatric severity and treatment response in a comprehensive methadone maintenance treatment program. Drug Alcohol Depend *48:*119, 1997.

Prendergast ML, Grella C, Perry SM, Anglin D: Levo-alpha-acetylmethadol (LAAM): Clinical, research, and policy issues of a new pharmacotherapy for opioid addiction. J Psychoactive Drugs *27:*239, 1995.

*Reisine T, Pasternak G: Opioid analgesics and antagonists. In *Goodman & Gilman's The Pharmacological Basis of Therapeutics*, ed 9. JG Harman, LE Limbird, RW Molinoff, RW Ruddon, AG Gilman, editors. McGraw Hill, New York, 1995.

Robins LN: Vietnam veterans' rapid recovery from heroin addiction: A fluke or normal expectation? Addiction *88:*1041, 1993.

Rounsaville BJ, Weissman MM, Kleber HD, Wilber C: The heterogeneity of psychiatric diagnosis in treated opiate addicts. Arch Gen Psychiatry *39:*161, 1982.

Schottenfeld RS, Pakes JR, Oliveto A, Ziedonis D, Kosten TR: Buprenorphine vs methadone maintenance treatment for concurrent opioid dependence and cocaine abuse. Arch Gen Psychiatry *54:*713, 1997.

Schulteis G, Gold LH, Koob GF: Preclinical behavioral models for addressing unmet needs in opiate addiction. Semin Neurosci *9:*94, 1997.

*Strain EC, Bigelow GE, Liebson IA, Stitzer ML: Moderate- versus high-dose methadone in the treatment of opioid dependence. A randomized trial. JAMA *281:*1000, 1999.

Strang J, Griffiths P, Gossop M: Heroin smoking by “chasing the dragon": Origins and history. Addiction *92:*673, 1997.

Tsuang MT, Lyons MJ, Lyons MJ, Meyer JM, Doyle T, Eisen SA, Goldberg J, True W, Lin N, Toomey R, Eaves L: Co-occurrence of abuse of different drugs in men. Arch Gen Psychiatry *55:*967, 1998.

Walsh SL, Johnson RE, Cone EJ, Bigelow GE: Intravenous and oral L-alpha-acetylmethadol: Pharmacodynamics and pharmacokinetics in humans. J Pharmacol Ex Ther *285:*71, 1998.

White PT, Raymer A: The poppy—for good and evil. Natl Geogr *167:*187, 1985.

11.11 PHENCYCLIDINE (OR PHENCYCLIDINE-LIKE)-RELATED DISORDERS

STEPHEN R. ZUKIN, M.D.

Phencyclidine (1-1 (phenylcyclohexyl)piperidine; PCP), also known as “angel dust,” entered the illicit street market in 1965. In the late 1970s it became one of the leading illicit drugs. Although its popularity subsequently declined, it remains a prominent drug of abuse in a number of American cities. It acts by a novel mechanism, has a long duration of action, and is very potent by any route of administration. PCP use carries a high risk of behavioral, physiological, and neurological toxicity and is highly reinforcing.

COMPARATIVE NOSOLOGY

Table 11.11-1 illustrates the similarities and differences between the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) and International Classification of Diseases (ICD) with respect to PCP-related disorders. Compared with the revised third edition of DSM (DSM-III-R), the fourth edition of DSM (DSM-IV) is more specific and more comprehensive. PCP-induced anxiety disorder is a category unique to DSM-IV. Because of the unusually complex pharmacodynamic and pharmacokinetic properties of PCP and related drugs, fluctuations among categories over time are common in the clinical situation as are PCP-induced states that fail to correspond to any established category.

Table 11.11-1 Nosology of Phencyclidine-Related Disorders

HISTORY

PCP was developed in the 1950s as a potential general anesthetic by Parke-Davis under the trade name Sernyl. Clinical trials in anesthesia were promising in that PCP did not depress vital cardiovascular and respiratory functions. By contrast to conventional general anesthetics, which induce a state of relaxed sleep, PCP induced a state of apparent catatonia, including flat facies, open mouth, fixed staring, rigid posturing, and (sometimes) waxy flexibility. Patients seemed dissociated from the environment without classical unconsciousness. For this reason, PCP was termed a *dissociative anesthetic*.

As many as half of patients given PCP anesthesia showed severe adverse reactions during emergence, such as agitation and hallucinations. Psychotic reactions persisting for up to 10 days then ensued in many of these patients. For this reason, PCP was abandoned for human anesthesia in favor of ketamine Ketalar), a less potent PCP derivative with a shorter duration of action. PCP was used as an experimental probe of psychotomimetic mechanisms in normal and schizophrenic subjects during the late 1950s and early 1960s. Given PCP's well-documented dysphoric and aversive properties in surgical patients, it seems surprising that it attracted so many illicit users beginning in 1965. In the late 1970s it was one of the most prevalent abused drugs in the United States; relatively infrequent but flamboyant and destructive episodes of PCP-induced aggressive and violent behaviors captured the public imagination and the attention of the press. The manifold and severe medical complications of PCP abuse led to an explosion of medical literature on this topic during the late 1970s and the 1980s. The fundamental mechanism of action of PCP, interaction with a unique high-affinity brain binding site, was discovered in 1979. PCP abuse continues to be a serious public health problem in a number of American cities. In recent years there has been a significant increase in abuse of ketamine, also known as “special K.” Although ketamine is less potent than PCP and has a shorter duration of action, they share the same fundamental mechanism of action. Unlike PCP, ketamine has not been scheduled under the Controlled Substances Act and can be found in hospital pharmacies, from which diversion has frequently occurred.

EPIDEMIOLOGY

PCP, a phenylcyclohexylamine, is easily synthesized from piperazine, cyclohexanone, and potassium cyanide. The synthesis proceeds through the intermediate 1-piperidinocyclohexanecarbonitrile (PC), which is reacted with phenylmagnesium bromide to form PCP. The simplicity of this reaction, which requires almost no training or equipment, suggests that PCP production may have an economic advantage over drugs derived from natural products or drugs that are difficult to synthesize. PCP is often misrepresented as another illicit substance (e.g., tetrahydrocannabinol [THC]), cocaine, methamphetamine, or a lysergic acid diethyamine [LSD]-like psychedelic drug). Marijuana is frequently adulterated with PCP by dealers. Thus PCP intoxication may be seen in patients who are not aware that they have taken PCP.

PCP has had many street names, including angel dust, devil's dust, dust, crystal, cyclones, embalming fluid, wet, killer weed, mintweed, peace pill, goon, jet fuel, surfer, black whack, Illy, crazy Eddie, purple rain, and milk. In combination with other drugs it has been called beam me up Scottie or tragic magic (crack dipped in PCP); love boat (marijuana dipped in PCP), and blunt (marijuana and PCP in a cigar wrapper). Its abuse was first reported in 1965, when it had limited popularity. In the early 1970s PCP was usually mixed with other drugs; by the late 1970s most street samples analyzed were pure PCP. The most common routes of administration are smoking and snorting; every conceivable route of administration has been reported, including injection. Regional differences in typical routes of administration have been reported, with snorting predominating in Philadelphia and Chicago, both smoking and snorting in Miami, and smoking in Seattle.

PCP abuse peaked between 1973 and 1979, with a smaller peak between 1981 and 1984. It is estimated that in 1976, 9.5 percent of persons between 18 and 25 had ever used PCP, increasing to 13.9 percent in 1977 and 14.5 percent in 1979. For 12- to 17-year-olds the corresponding figures were 3.0, 5.8, and 3.9 percent. By 1988, these figures had declined to 1.2 percent for those 12 to 17 and 4.4 percent for those 18 to 25. Since 1985 the highest rates of use have been in those between 26 and 34 years of age. Some 12.8 percent of high-school seniors reported PCP use in 1979; this declined to about 3 percent from 1987 through 1995. The 1994 survey data indicate that 4.3 percent of household residents aged 12 years and older, or about nine million people, had used PCP at least once in their lives, up from 2.9 percent in 1985. In 1994 478,000 people (0.02 percent) were estimated to have used PCP at least once in the prior year. Although this number is dwarfed by annual rates of marijuana (8.5 percent), cocaine (1.7 percent), inhalant (1.1 percent), and crack (0.6 percent) use, it is still significant.

PCP-related emergency room mentions increased more than fivefold between 1974 and 1979, then declined; by 1991 PCP ranked 27th among drugs mentioned. In

order, the leading cities for PCP emergency room mentions were Los Angeles, Chicago, Washington, New York City, San Francisco, Baltimore, and Philadelphia. Legal and policy changes contributed to the decline in PCP use during the 1980s. PCP was reclassified from Schedule III to Schedule II under the Controlled Substances Act. Reporting of production of piperidine, a major PCP precursor, was mandated. Penalties for PCP possession with intent to sell were significantly increased.

For 1994 high school seniors asked when they had begun to use PCP, the mode was within the 2 to 3 years prior to interview (while in grades 10 and 11). The same survey also showed that PCP ranked as the seventh most-often-discontinued drug after inhalants (56 percent), heroin (50 percent), methamphetamine (47 percent), steroids (46 percent), tranquilizers (44 percent), and methaqualone (43 percent).

Although information on use is not available for 8th and 10th graders, questions on their questionnaires concerned perceptions regarding the difficulty in obtaining PCP. In 1994, about 18 percent of the 8th graders, 24 percent of the 10th graders, and 31 percent of the 12th graders thought that PCP was “fairly easy” or “very easy” to obtain. The proportion of 8th and 10th graders who thought PCP was easy to get was similar to those who thought heroin and crystal methamphetamine were easy to obtain.

In 1995 the Community Epidemiology Work Group (CEWG), supported by the National Institute on Drug Abuse (NIDA), issued reports indicating increases in PCP use in Washington, D.C., Philadelphia, and New York City (emergency rooms); in Washington, D.C., and Chicago (juvenile arrestees); and in Miami (treatment admissions, particularly in conjunction with marijuana).

ETIOLOGY AND PSYCHOPHARMACOLOGY

Phencyclidine and N-methyl-D-aspartate Receptors The unique primary central nervous system (CNS) actions of PCP result from binding to high-affinity PCP receptors that are highly selective for drugs (from a variety of chemical classes) that elicit PCP-like behavioral effects. The relative potencies of such drugs in eliciting PCP-like behaviors are proportional to their potencies in competing for radioligand binding to the PCP receptor. By virtue of the location of the PCP binding site within the ionophore of the *N*-methyl-D-aspartate (NMDA) receptor, drugs that bind to PCP receptors also block NMDA-activated channels in electrophysiological assays. The rank order of potency of such drugs as channel blockers parallels their behavioral potencies and their potencies in binding to the PCP receptor. The highest densities of PCP receptors are found in the dentate gyrus and the CA1 and CA2 subfields of the hippocampus and in anterior forebrain areas, including neocortex, consistent with the ability of PCP to disrupt learning, memory, and higher cortical functions.

Unlike the other types of glutamate receptor channels, NMDA channels are permeable to both Ca_{2+} and Na_{+}. Following NMDA receptor activation, NMDA-mediated Ca_{2+} flux may lead to stimulation of calmodulin-dependent kinases with activation of postsynaptic second-messenger pathways. A unique functional property of the NMDA channel is that it is blocked in a voltage-dependent manner by the endogenous Mg_{2+} ion. The dual voltage- and ligand-dependence of NMDA receptors permits them to function in a “Hebbian” manner to integrate information from multiple input streams. One stream is represented as a modulation of presynaptic glutamate release, while additional streams are reflected in modulation of resting membrane potential on the postsynaptic NMDA–receptor bearing dendrites. Ca_{2+} flow through open, unblocked NMDA channels may trigger long-term potentiation, which in turn may represent the neurophysiological substrate underlying learning and memory formation. In rodents, PCP and other NMDA antagonists induce profound memory disturbances linked to inhibition of hippocampal long-term potentiation. In humans, amnesia can be seen in both PCP intoxication and ketamine anesthesia.

NMDA receptors possess two distinct agonist binding sites, one for glutamate and other excitatory amino acids (e.g., aspartate) and another for glycine and other structurally similar amino acids such as D-serine. In vivo, glutamate and glycine/D-serine have distinct roles in receptor activation. Like most classical neurotransmitters, glutamate is released from presynaptic nerve endings in pulsatile fashion and then rapidly deactivated. By contrast, glycine in forebrain is neither concentrated in presynaptic nerve endings nor released in response to electrical stimulation. Furthermore, endogenous glycine concentrations are typically at or above the affinity constant of the NMDA-associated glycine site for these agents, indicating that activity-stimulated release of glycine is not required for neurotransmission. Rather, the local glycine concentration appears to set the tonic level of NMDA excitability, thus determining the degree to which presynaptic glutamate release (triggered by presynaptic glutamate release) leads to postsynaptic excitation. Glycine thus functions more as a neuromodulator than as a classical neurotransmitter. D-Serine is the only D-amino acid present in high concentration in brain. Thus D-serine, like glycine, may help regulate the tonic excitability level of NMDA excitability. The NMDA receptor-associated glycine site is anatomically, pharmacologically, and functionally distinct from the classical strychnine-sensitive glycine receptor found primarily in hindbrain, which binds D-serine with low affinity.

Opening the NMDA channel facilitates access of PCP to its receptor, accelerating the rate at which PCP-induced blockade of NMDA receptor-mediated neurotransmission takes place relative to the rate in the absence of NMDA-receptor activation. In the former case, PCP gains access to its binding site rapidly via the open channel pore; in the latter case, the highly lipophilic PCP molecule gains access to its binding site via slow diffusion through the lipid bilayer.

Molecular Characterization of the PCP Receptor The NMDA receptor complex is heteroligemeric, consisting of combinations of NMDAR1 and NMDAR2 subunits. NMDAR1 is the key subunit in the formation of the receptor complex. All functional NMDA receptor complexes contain an NR1 subunit; complexes may also contain variable numbers of modulatory subunits (NR2A-D). The glycine binding site is contained within the NR1 subunit; thus all NMDA receptors show glycine sensitivity. However, glycine affinity is modulated by NR2 subunits; thus receptors containing 2D subunits appear more sensitive to glycine than those containing 2B subunits, which in turn are more sensitive than those containing 2A subunits. Eight variants of NMDAR1 (NMDAR1a–h) have been identified, which reflect alternative mRNA splicing. Receptors containing such variants may also have distinct sensitivities to agonists, antagonists, Zn_{2+}, and polyamines.

NMDA receptors are primarily found postsynaptically and occur on both projection- and interneurons. However, NMDA subunit composition varies both among brain regions and among cell types within a region. In some brain regions, especially in target areas of the mesolimbic-mesocortical system, NMDA receptors may also be localized presynaptically and thus may regulate release of dopamine from presynaptic terminals. It has been suggested that such receptors consist solely of NR1 subunits and thus may have a different pharmacological profile from postsynaptic receptors.

Additional CNS Actions of PCP At doses at least 10 times those at which it exerts its unique behavioral effects by blocking NMDA receptor–mediated neurotransmission, PCP also blocks presynaptic monoamine reuptake, thus directly increasing synaptic levels of dopamine and norepinephrine. Concentrations sufficient to block monoamine reuptake may be achieved during high-dose intoxication and may contribute to the stimulatory behavioral and physiological effects of PCP seen at those doses. Ketamine, even at high dose, fails to block monoamine transporters. This observation may explain why the incidence of episodes of extreme agitation and violence following ketamine intoxication is lower than that after PCP intoxication. At doses associated with extremely high dose intoxication, PCP blocks neuronal Na_{+} and K_{+} channels as well as muscarinic cholinergic receptors. Such effects may be relevant to the seizures observed following PCP overdose. PCP also interacts with a variety of other CNS receptors including opiate and g-aminobutyric acid (GABA) benzodiazepine receptors. However, most such effects take place at concentrations unlikely to be encountered in clinical situations.

Neurotoxicity of PCP In animals, neurotoxic effects of PCP are observed at doses (e.g., 5 to 50 mg/kg) significantly above those required for behavioral activity. At those doses in the rat, PCP induces neuronal vacuolization, particularly in neurons in rat posterior cingulate/retrosplenial cortex. Similar vacuolization is observed following administration of dizocilpine (i.e., MK-801) or ketamine, indicating that the effect is probably mediated at NMDA receptors. The effect is initially observed in layers III and IV of cortex. At lower doses (e.g., 5 mg/kg), the effect is transient, reaching a peak approximately 12 hours after PCP or dizocilpine administration and then resolving over 12 to 18 hours. Much higher doses of PCP may lead to neuronal necrosis, which is apparent even 48 hours after drug administration and is seen in hippocampus and other limbic areas as well as in the posterior cingulate retrosplenial cortex (PCRS). Administration of high-dose PCP also leads to increased glucose uptake and expression of heat shock and glial fibrillary acidic protein (GFAP). Vacuolization can be inhibited by prior administration of anticholinergic, GABAergic, or antipsychotic agents and is potentiated by administration of pilocarpine. Direct evidence of similar neurotoxic effects in humans is lacking, despite the large number of patients who have undergone ketamine anesthesia for surgery.

Toxicity of 1-Piperidinocyclohexanecarbonitrile (PC) During the 1970s significant percentages of analyzed street samples of PCP were found to be contaminated by dangerously high amounts of residual PC, the intermediate compound in PCP synthesis. In physiological saline, PC decomposes to hydrocyanic acid (HCN). When PC-contaminated PCP is smoked, about 58 percent of the PC breaks down to cyanide and organic byproducts. Although devoid of PCP-like pharmacological activity, PC is thus more toxic and may be implicated in some PCP-related fatalities.

Behavioral Pharmacology of PCP Considerable species variation exists in the behavioral effects of PCP. In rodents, PCP induces a characteristic syndrome of hyperactivity and stereotypical movements, which is only partially reversible by neuroleptics but can be more robustly antagonized by agents such as glycine that

augment NMDA receptor-mediated neurotransmission. Because the serum half-life of PCP is shorter and its volume of distribution is larger in rodents than in primates, rodents typically require higher weight-adjusted doses of PCP than primates to elicit behavioral effects. Sensitization to the behavioral effects of PCP has been detected after daily administration in rodents. PCP also inhibits rodents' social behavior, an effect that is not reversed by either typical or atypical antipsychotic agents.

PCP-induced hyperactivity in rodents appears to reflect increased dopaminergic neurotransmission within nucleus accumbens, since this effect can be selectively inhibited by lesions in that brain area. Other components of the PCP-induced behavioral syndrome (e.g., alterations in social behavior, stereotypy) are not affected by lesions of the accumbens, suggesting that those behaviors are mediated by other brain regions. NMDA receptors are also present in substantia nigra (A9) and ventral tegmental area (A10). Glutamatergic innervation of substantia nigra from prefrontal cortex strongly influences dopaminergic activity levels; NMDA receptors appear to be the primary mediators of glutamate-induced stimulation of midbrain dopaminergic neurons. To the extent that NMDA receptors stimulate A9 and A10 neurons, PCP would, paradoxically, be expected to diminish dopaminergic outflow from striatum and accumbens. However, direct application of PCP to A10 does not inhibit dopamine cell firing or alter dopamine release in accumbens, although it does prevent NMDA-induced neuronal activation. Thus, the behavioral effects of PCP in rodents appear to stem predominantly from its interactions within dopamine terminal fields rather than within dopaminergic midbrain nuclei.

Gross behavioral effects of PCP in nonhuman primates more closely resemble the effects of the drug in humans. In monkeys, a PCP dose of approximately 0.5 mg/kg produces tranquilization in which animals appear awake but unresponsive to the environment. At 1.0 mg/kg, PCP induces a cataleptoid state in which animals show waxy flexibility and rigidity resembling catatonic schizophrenia. Doses of 2.5 mg/kg lead to stupor; 5.0 mg/kg, to surgical anesthesia; and 15 mg/kg, to convulsive seizures.

In all species tested, PCP profoundly disrupts learning and memory. In rodents, PCP-like drugs disrupt spatial delayed alteration performance; this effect is reversible by type 2 dopamine (D_2) receptor antagonists. Thus in cortex as in striatum, the effects of PCP-like drugs in part reflect secondary dysregulation of dopaminergic neurotransmission. In monkeys, PCP-like drugs induce impairments in learning and working memory performance that are not reproduced by amphetamine, demonstrating that the effect cannot be attributed entirely to increased dopaminergic neurotransmission. However, NMDA antagonists potentiate the disruptive effects of amphetamine on learning in monkeys, indicating that interactions between the NMDA and dopamine systems are important.

Pharmacokinetics of PCP PCP is highly lipid soluble and has a volume of distribution of 6.2 L/kg in humans. Significant brain concentrations are rapidly attained after inhaled, smoked, or topical administration (PCP intoxication has been documented after inadvertent wearing of clothing on which a PCP solution had been spilled). A typical street dose is about 5 mg (one pill, joint, or line), which results in a serum PCP concentration between 0.01 and 0.1 µM. Psychotic reactions have been observed in association with undetectably low serum concentrations (<0.02 µmol), while concentrations above 0.4 µmol induce gross impairment of consciousness. Serum concentrations above 1.0 µM are strongly associated with coma, seizures, respiratory arrest, and death. The highest recorded serum and cerebrospinal fluid (CSF) concentrations are in the range of 1.0 to 2.0 µM. Users typically titrate dosage in an effort to maximize the high while avoiding unconsciousness. Failures of judgment or variations in purity of supplies often result in inadvertent overdose, which may lead to severe medical complications. The serum half-life of PCP varies considerably among individuals, averaging about 20 hours. The half-life of ketamine in humans is about 2 hours. The lipophilicity of PCP facilitates accumulation in fatty body tissues, including the brain. PCP concentrations in brain and in adipose tissues may be more than 10 times those in plasma. Mobilization of adipose stores (e.g., during exercise) may release sequestered PCP, leading to flashbacks.

Metabolism of PCP is predominantly hepatic; hydroxylase metabolites are secreted renally. PCP (pK_a = 8.5) is largely ionized in the stomach or the urinary tract. Passing through the pyloric valve, PCP enters a nonacidic environment in the small intestine in which it becomes largely nonionized and readily absorbed across the mucosal membrane, whereupon enterohepatic recalculation can account for the fluctuating clinical course so often observed in PCP intoxication.

REINFORCING EFFECTS OF PCP

Experimental monkeys avidly self-administer large doses of PCP intravenously or orally. Monkeys given unlimited access to PCP maintain nearly continuous intoxication. This pattern resembles that seen with opiates and CNS depressants but differs from that seen with classical stimulants. Furthermore, in contrast to findings on opiate self-administration, monkeys will self-administer doses of PCP high enough to cause marked behavioral effects. Given the similarities between behavioral effects of PCP in monkeys and humans, this research validates the clinical impression of avid human self-administration of PCP. PCP-like drugs stimulate brain reward areas, lowering the threshold for intracranial self-stimulation. Such effects, which define a classical profile for drugs abused by humans, are shared by such other abused compounds as opiates, stimulants, and benzodiazepines, despite their differing mechanisms of action.

TOLERANCE TO PCP

In rats and monkeys, repeated administration of PCP daily or more frequently leads to two- to fourfold rightward shifts in dose-response curves. The major determinant of this moderate tolerance appears to be biodispositional. Much greater tolerance is induced by continuous self-administration than by intermittent dosing. Limited human data exist on PCP tolerance, but tolerance has been reported in burn patients given repeated doses of ketamine for analgesia.

PHYSICAL DEPENDENCE ON PCP

After unlimited-access self-administration of PCP for a month or longer, severe withdrawal reactions were observed in monkeys when the drug was discontinued, including vocalizations, bruxism, oculomotor hyperactivity, diarrhea, piloerection, somnolence, tremor, and seizures. Similarly severe reactions might be expected following PCP binges by human abusers. Despite a paucity of controlled studies of tolerance and dependence in humans, the animal literature suggests that PCP must be considered comparable to classical drugs of abuse in this respect.

DSM-IV DISORDERS

Table 11.11-2 lists the DSM-IV PCP-related disorders. Clinical presentations of PCP-induced disorders frequently fail to fit within any established diagnostic category. Mixtures of features of several categories are quite common. Frequently, the presentation varies over time as serum PCP levels fluctuate. Therefore, even the most careful single examination cannot be definitive. The absence of serious medical complications at a single time does not preclude development of such complications hours later. Delirium and psychotic features often come and go several times during intoxication. Unless serious medical complications cause irreversible cardiovascular, renal, or neurological damage, complete recovery from PCP intoxication is usually seen within 24 to 72 hours. Even in prolonged phencyclidine-induced psychotic disorder, complete recovery is usually seen within 6 weeks.

Phencyclidine use disorders
Phencyclidine dependence
Phencyclidine abuse
Phencyclidine-induced disorders
Phencyclidine intoxication
Specify if:
With perceptual disturbances
Phencyclidine intoxication delirium
Phencyclidine-induced psychotic disorder, with delusions
Specify if:
With onset during intoxication
Phencyclidine-induced psychotic disorder, with hallucinations
Specify if:
With onset during intoxication
Phencyclidine-induced mood disorder
Specify if:
With onset during intoxication
Phencyclidine-induced anxiety disorder
Specify if:
With onset during intoxication
Phencyclidine-related disorder not otherwise specified

Based on American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. Washington, DC, American Psychiatric Association, 1994.

Table 11.11-2 DSM-IV Phencyclidine-Related Disorders

A 17-year-old male was brought to the emergency room by the police, having been found disoriented on the street. As the police attempted to question him, he became increasingly agitated; when they attempted to restrain him, he became assaultive. Attempts to question or examine him in the emergency room evoked increased agitation. Initially it was impossible to determine vital signs or to draw blood. Based upon the observation of horizontal, vertical, and rotatory nystagmus, a diagnosis of PCP intoxication was entertained. Within a few minutes of being placed in a darkened examination room, his agitation markedly decreased. Blood pressure was 170/100; other vital signs were within normal limits. Blood was drawn for toxicological examination. The patient agreed to take 20 mg of diazepam (Valium) orally. Thirty minutes later, he was less agitated and could be interviewed, although he responded to questions in a fragmented fashion and was slightly dysarthric. He stated that he must have inadvertently taken a larger-than-usual dose of "dust," which he reported having used once or twice a week for several years. He denied use of any other substance and any history of mental disorder. He was disoriented to time and place. The qualitative toxicology screen revealed PCP and no other drugs. Results of neurological examination were within normal limits but very brisk deep tendon reflexes were noted. Some 90 minutes after arrival his temperature, initially normal, was elevated to 38°C, blood pressure had increased to 182/110, and he was poorly responsive to stimulation. He was admitted to a medical bed. His blood pressure and level of consciousness continued to fluctuate over the ensuing 18 hours. Results of hematological and biochemical analyses of blood, as well as urinalyses, remained within normal limits. A history obtained from his family revealed that the patient had had multiple emergency-room visits for complications from PCP use during the previous several years. He had completed a 30-day residential treatment program and had participated in several outpatient programs but had consistently relapsed. The patient was discharged after vital signs and level of consciousness had been within normal limits for 8 hours. At discharge, nystagmus and dysarthria were no longer present. A referral to an outpatient treatment program was made.

Phencyclidine Abuse and Phencyclidine Dependence There have been no systematic studies of PCP abuse, dependence (see Table 11.11-2), or withdrawal states in humans, but primate research gives every reason to expect that they are clinically significant. Patterns of PCP use have been documented in specific populations in which users self-administer multiple, high daily doses continuously for weeks, months, or years. The avidity with which many users return to PCP use immediately after treatment for severe PCP intoxication or psychosis suggests that intense craving is experienced after medically supervised withdrawal. No specific treatments have been developed for PCP abuse or dependence. Relapse can reportedly be averted with antidepressant treatment in selected patients, suggesting that some users are attracted to PCP by its antidepressant properties.

Phencyclidine Intoxication, Phencyclidine Intoxication Delirium

Diagnosis And Clinical Features The cornerstone of diagnosis of PCP intoxication (Table 11.11-3) is a history of PCP use, detection of PCP in body fluids, or both. A number of assays for PCP are available; however, for several reasons, it can be dangerous to rely on them. First, the combination of PCP's lipophilicity, long duration of action, and pK_a predict that prominent clinical effects may be present in the absence of measurable concentrations of PCP in blood or CSF. This is particularly true for acute psychotic reactions, which have been observed at serum PCP concentrations undetectable by most assay systems. A positive assay for PCP should generally be used only qualitatively; it is particularly important not to assume that a low measured concentration predicts an uneventful recovery. pK_a considerations in relation to varying urinary and intestinal pH and enterohepatic recirculation indicate that concentrations can be expected to fluctuate. Clinical indications of significant neuronal hyperexcitability, hypertension, hyperthermia, or other physiological dysregulation should be interpreted as indicating a possible impending medical emergency. Reliance upon history may also yield many false negatives, because PCP may have been misrepresented as another drug or used as an adulterant by the dealer. As PCP levels fluctuate, various mixtures of symptoms of intoxication, delirium, and psychosis are seen.

A. Recent use of phencyclidine (or a related substance).
B. Clinically significant maladaptive behavioral changes (e.g., belligerence, assaultiveness, impulsiveness, unpredictability, psychomotor agitation, impaired judgment, or impaired social or occupational functioning) that developed during, or shortly after, use of phencyclidine.
C. Within an hour (less when smoked, "snorted," or used intravenously), two (or more) of the following signs:
   (1) vertical or horizontal nystagmus
   (2) hypertension or tachycardia
   (3) numbness or diminished responsiveness to pain
   (4) ataxia
   (5) dysarthria
   (6) muscle rigidity
   (7) seizures or coma
   (8) hyperacusis
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Specify if:
With perceptual disturbances

Table 11.11-3 DSM-IV Diagnostic Criteria for Phencyclidine Intoxication

Physiological Bases of Intoxication, Delirium and Medical Complications The range of clinical effects of PCP can be correlated with dose, serum PCP concentration, and interaction with several molecular target sites (Fig. 11.11-1). The CNS NMDA receptor complex would be the only system affected significantly at very low PCP doses. Serum PCP concentrations up to about 0.1 µM correspond to a clinical state manifesting psychotomimetic symptoms and impaired cognition without physiological disturbances of vital functions other than nystagmus, hypertension, tachycardia, dysarthria, or hyperacusis. Serum concentrations just above about 0.1 µmol correspond to a state of delirium and at about 0.3 µM to incipient dissociative anesthesia, with prominent numbness or diminished responsiveness to pain. At still higher doses, as additional types of receptors are occupied, acute brain syndrome accompanied by prominent neurological and cardiovascular complications is seen. Deaths from hyperthermia, status epilepticus, and hypertensive crisis have been reported. At high doses, PCP exerts direct excitatory effects upon skeletal muscle endplates. These effects may underlie the many reports of superhuman strength from the late 1970s. Together with the behavioral effects of high-dose PCP, these effects probably account for muscle trauma, which in a number of patients has resulted in rhabdomyolysis, myoglobinuria, and renal failure. Serum concentrations of 1.0 µmol and above are associated with coma and death. PCP-induced delirium (see Table 10-22) and coma result from the combination of noncompetitive inhibition of the PCP-NMDA receptor, blockade of the reuptake sites for catecholamines and indolamines, and blockade of sodium and potassium channels and nicotinic and muscarinic cholinergic receptors.

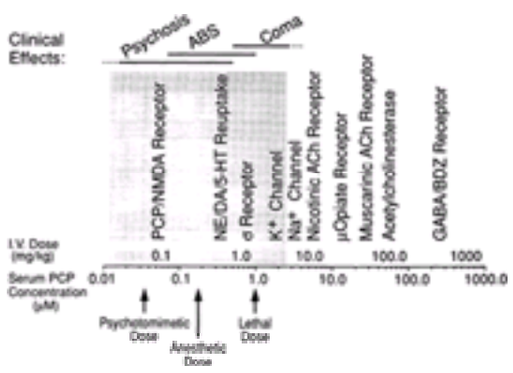

FIGURE 11.11-1 Dose range of PCP effects. The relation of dose, serum concentration, molecular target sites, and clinical effects are shown. The *shaded area* represents clinically relevant interactions. (Reprinted with permission from Javitt DC, Zukin SR: Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry *148:*1301, 1991.)

NEUROLOGICAL MANIFESTATIONS The vast majority of PCP-intoxicated patients manifest horizontal, vertical, or rotatory nystagmus, which can help distinguish PCP intoxication from a naturally occurring psychotic state. Neuronal hyperexcitability is dose dependent, ranging from increased deep tendon reflexes, through opisthotonos, to focal or generalized seizures, to status epilepticus. Focal neurological findings may arise from cerebral vasoconstriction. Coma can occur at any point during intoxication.

BEHAVIORAL MANIFESTATIONS Clinically urgent short-term behavioral complications of PCP abuse stem from behavioral disinhibition, which can be coupled with

severe agitation, panic, rage, and aggression. The disruption of sensory input by PCP can cause unpredictable, exaggerated, distorted, or violent reactions to environmental stimuli. Such reactions are more common with somewhat higher doses, at which some delirium as well as neurological symptoms and other medical complications are observed. The behavioral manifestations can severely compromise the clinician's ability to treat the medical complications.

CARDIOVASCULAR MANIFESTATIONS PCP-induced hypertension is dose dependent. Mild hypertension is seen in all patients with PCP intoxication. With increasing PCP dose, the panoply of sympathomimetic responses is seen; hypertensive crisis is common in association with high doses and may lead to cerebral hemorrhage.

AUTONOMIC MANIFESTATIONS Severe hyperthermia has been observed, which can be delayed, fatal, or both. The anticholinergic properties of PCP can evoke the full spectrum of atropine-like toxicity in a dose-dependent manner and can be managed accordingly.

Differential Diagnosis Even when PCP is detected in body fluids, remember that PCP is frequently taken in combination with another drug or drugs. Depending upon the analytic method used, an assay for PCP may fail to detect an intoxicating concentration of PCP. While not conclusive or totally specific for PCP intoxication, the constellation of nystagmus, ataxia, and mild hypertension is typically observed even in mild, low-dose PCP intoxication.

Treatment Treatment of PCP intoxication aims to reduce systemic PCP levels and address significant medical, behavioral, and psychiatric issues. For both intoxication and PCP-induced psychotic disorder, while resolution of current symptoms and signs is paramount, the long-term goal of treatment is prevention of relapse to PCP use. PCP levels may fluctuate over many hours or even days, especially after oral administration. Therefore, a prolonged period of clinical observation is mandatory before concluding that no serious or life-threatening complications will ensue.

Trapping of ionized PCP in the stomach has led to the suggestion of continuous nasogastric suction as a treatment for PCP intoxication. However, this strategy can be needlessly intrusive and can induce electrolyte imbalances. Administration of activated charcoal is safer, and it binds PCP and diminishes toxic effects of PCP in animals.

Trapping of ionized PCP in urine has led to the suggestion of urinary acidification as an aid to drug elimination. However, this strategy may be ineffective and is potentially dangerous. Only a small portion of PCP is excreted in urine; metabolic acidosis itself carries significant risks; and acidic urine can increase the risk of renal failure secondary to rhabdomyolysis. Because of the extremely large volume of distribution of PCP, neither hemodialysis nor hemoperfusion can significantly promote drug clearance.

No drug is known to function as a direct PCP antagonist. Any compound binding to the PCP receptor, which is located within the ion channel of the NMDA receptor, would block NMDA receptor–mediated ion fluxes as does PCP itself. NMDA receptor mechanisms predict that pharmacological strategies promoting NMDA receptor activation (e.g., administration of a glycinelike drug) would promote rapid dissociation of PCP from its binding sites. Evidence indicates that oral glycine in massive doses can antagonize PCP-induced behaviors in the mouse. Large doses (80 grams a day) of glycine administered experimentally to schizophrenic patients have shown no significant toxicity; however, no clinical trials for PCP intoxication in humans have been carried out to date. Therefore, treatment must be supportive and directed at specific symptoms and signs of toxicity. Classical measures should be used for medical crises including seizures, hypothermia, and hypertensive crisis.

Because PCP disrupts sensory input, environmental stimuli can cause unpredictable, exaggerated, distorted, or violent reactions. A cornerstone of treatment is therefore minimization of sensory inputs to PCP-intoxicated patients. Patients should be evaluated and treated in as quiet and isolated an environment as possible. Precautionary physical restraint is recommended by some authorities, with the risk of rhabdomyolysis from struggle against the restraints balanced by the avoidance of violent or disruptive behavior. Pharmacological sedation can be accomplished with oral or intramuscular antipsychotics or benzodiazepines; there is no convincing evidence that either class of compounds is clinically superior. Because of the anticholinergic actions of PCP at high doses, neuroleptics with potent intrinsic anticholinergic properties should be avoided.

Course and Prognosis Complete recovery from PCP intoxication is the rule, in the absence of major medical complications. However, many patients relapse to PCP use immediately after discharge from treatment, even for severe PCP-related complications. Intoxication usually occurs in the context of abuse, dependence, or both. Unfortunately, no specific behavioral treatments for PCP abuse and dependence have been described. Case reports indicate successful responses to residential and intensive outpatient treatment regimens with long-term follow-up including urine monitoring with or without contingency contracting.

Phencyclidine-Induced Psychotic Disorder

Diagnosis and Clinical Features Minute, single intravenous doses (0.05 to 0.1 mg/kg) of PCP administered to normal volunteers can rapidly induce a psychotic state, lasting 4 to 6 hours, characterized by withdrawal, autism, negativism, inability to maintain a cognitive set, and concrete, impoverished, idiosyncratic, and bizarre responses to proverbs and projective testing. Some subjects show catatonic posturing. These schizophrenia-like alterations in brain functioning extend beyond the symptom level; in formal studies of neuropsychological function, PCP can induce a spectrum of specific disturbances in attention, perception, and symbolic thinking strikingly similar to those seen in schizophrenia. The most severe impairment caused by PCP was observed in tests requiring selective attention and paired-association learning. An important clinical correlate of these data is that any person under the influence of even a small dose of PCP or a similar drug may have profound alterations of higher emotional functions affecting judgment and cognition, even in the absence of gross neurological findings. Well-formed hallucinations and delusions are relatively uncommon.

Challenge experiments involving recompensated schizophrenic subjects found that a single low dose of PCP can rekindle the presenting symptomatology for as long as 6 weeks, without evoking any symptoms or signs not typical of schizophrenia. Thus schizophrenia or preschizophrenia patient abusing a PCP-type drug run an extremely high risk of severe and prolonged psychiatric morbidity, continuing well beyond the time when PCP is present in the system. While in most cases PCP-induced psychosis closely resembles schizophrenia, a number of reports of mania-like psychotic states, both acute and prolonged, induced by PCP have appeared. Patients with an existing or incipient bipolar disorder are probably particularly sensitive to this complication.

In retrospective studies of patients hospitalized for complications of PCP use during the 1970s, the presenting symptoms could not distinguish PCP-intoxicated patients from schizophrenic patients. In the clinical situation one must recognize that acute PCP-induced psychosis can persist longer than the signs required for formal diagnosis of PCP intoxication in DSM-IV, particularly after low drug doses. By the time a patient comes to medical attention, psychosis may be the only notable finding. PCP-induced psychosis may remit after 4 to 6 hours or can persist for up to 6 weeks after signs of acute physiological reactions have cleared. For both acute and prolonged psychotic disorders induced by PCP, DSM-IV criteria for PCP-induced psychotic disorder (see Table 13.3-4) may fail to be met because of the absence of delusions or hallucinations, despite the presence of severe thought disorder.

Differential Diagnosis For several reasons, the differential diagnosis of PCP-induced psychotic disorder may prove challenging. First, the patient as well as other informants may be unaware that PCP has been used, because it may have been misrepresented as another drug or used as an adulterant. Second, PCP-induced psychosis can persist long beyond the time when PCP is detectable in lood and beyond the time when neurological and cardiovascular symptoms and signs of PCP intoxication are present. Third, careful retrospective analyses have demonstrated that frequently PCP-induced psychosis cannot be distinguished from naturally occurring schizophrenia on the basis of presenting symptoms and signs. Detection of PCP still leaves open the possibility that the patient suffers from a preexisting psychotic disorder that has been exacerbated by PCP, a possibility that can be ruled out only by following the patient's future course.

Course and Prognosis In most cases, once any acute medical complications are successfully treated, both peripheral and CNS complications, including psychosis, resolve completely within 24 to 72 hours. In the case of prolonged PCP-induced psychotic disorder, the rule is complete recovery within 4 to 6 weeks, regardless of whether antipsychotics have been administered. However, the rate of subsequent relapse to PCP use is very high. Persistence of a psychotic disorder beyond 8 weeks indicates the possible presence of an underlying psychotic disorder exacerbated, but not caused, by PCP.

Treatment For either acute or prolonged psychotic reactions, neuroleptics are commonly prescribed. The literature suggests that they are ineffective for most patients, but no systematic studies of this issue have been performed. Both neuroleptics and benzodiazepines have been reported to be moderately effective against the agitation and behavioral disinhibition that may be present. However, the psychotic symptoms themselves often fail to respond to pharmacological treatment. As a consequence, inpatient treatment may be required throughout the course of the psychotic reaction

Phencyclidine-Induced Mood Disorder PCP-induced mood disorder (see Table 14.6-18) is a somewhat elusive category. A relatively small percentage of patients display a prominently elevated mood verging on hypomania after exposure to PCP. Symptoms and signs may clear as PCP concentrations decline or may progress to

a PCP-induced manic state with psychotic features. In patients with a bipolar disorder, the intrinsic antidepressant properties of PCP may provoke a hypomanic or manic reaction.

Phencyclidine-Induced Anxiety Disorder The literature provides little or no support for PCP-induced anxiety disorder (see Table 15.6-17) as a distinct clinical entity. Preclinically, PCP and related drugs fit an anxiolytic and antidepressant profile. Distinguishing anxiety from the agitation and behavioral disinhibition observed in many patients with PCP intoxication would be difficult.

PCP-Related Disorder Not Otherwise Specified The diagnosis of PCP-related disorder not otherwise specified is the appropriate diagnosis for a patient who does not fit into any of the previously described diagnoses (Table 11.11-4).

The phencyclidine-related disorder not otherwise specified category is for disorders associated with the use of phencyclidine that are not classifiable as phencyclidine dependence, phencyclidine abuse, phencyclidine intoxication, phencyclidine intoxication delirium, phencyclidine-induced psychotic disorder, phencyclidine-induced mood disorder, or phencyclidine-induced anxiety disorder.

Table 11.11-4 DSM-IV Diagnostic Criteria for Phencyclidine-Related Disorder Not Otherwise Specified

SUGGESTED CROSS-REFERENCES

Other substance-related disorders are treated in Chapter 11. Schizophrenia, a disorder related symptomatically and pathophysiologically to phencyclidine-induced psychotic disorder, is discussed in Chapter 12. Delirium and its differential diagnosis are covered in Chapter 10. Amino acid neurotransmitters are covered in Section 1.5.

SECTION REFERENCES

Aronow R, Miceli JN, Done AK: A therapeutic approach to the acutely overdosed PCP patient. J Psychedelic Drugs *12:*259, 1980.

Baldridge EB, Bessen HA: Phencyclidine. Emerg Med Clin North Am *8:*541, 1990.

Balster RL: Clinical implications of behavioral pharmacology research on phencyclidine. NIDA Res Monogr *64:*148, 1986.

Balster RL, Johanson CE, Harris RT, Schuster CR: Phencyclidine self-administration in the rhesus monkey. Pharmacol Biochem Behav *1:*167, 1973.

Brust JC: Acute neurologic complications of drug and alcohol abuse. Neurol Clin *16:*503, 1998.

Burns RS, Lerner SE: Perspectives: Acute phencyclidine intoxication. Clin Toxicol *9:*477, 1976.

Chen G, Ensor CR, Russell D, Bohner B: The pharmacology of 1-(1-phenylcychohexyl) piperidine-HCl. J Pharmacol Exp Ther *127:*240, 1959.

Cohen BD, Rosenbaum G, Luby ED, Gottlieb JS: Comparison of phencyclidine hydrochloride (Sernyl) with other drugs. Arch Gen Psychiatry *6:*79, 1961.

Crowley TJ: Phencyclidine (or phencyclidinelike)-related disorders. In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Davis BL: The PCP epidemic: A critical review. Int J Addict *17:*1137, 1982.

*Deutsch SI, Mastropaolo J, Rosse RB: Neurodevelopmental consequences of early exposure to phencyclidine and related drugs. Clin Neuropharmacol *21:*320, 1999.

Domino EF, Luby E: Abnormal mental states induced by phencyclidine as a model of schizophrenia. In *PCP (Phencyclidine): Historical and Current Perspectives*, EF Domino, editor. NPP Books, Ann Arbor, MI, 1981.

Ellison G: The *N*-methyl-D-aspartate antagonists phencyclidine, ketamine, and dizocilpine as both behavioral and anatomical models of the dementias. Brain Res Rev *20:*250, 1995.

*Erard R, Luisada PV, Peele R: The PCP psychosis: Prolonged intoxication or drug-precipitated functional illness? J Psychedel Drugs *12:*235, 1980.

Freeman AS, Martin BR, Balster RL: Relationship between the development of behavioral tolerance and the biodisposition of phencyclidine in mice. Pharmacol Biochem Behav *20:*373, 1984.

Huntley GW, Vickers JC, Morrison JH: Cellular and synaptic localization of NMDA and non-NMDA receptor subunits in neocortex: Organizational features related to cortical circuitry, function and disease. Trends Neurosci *17:*536, 1994.

Itil T, Keskiner A, Kiremitci N, Holden JMC: Effect of phencyclidine in chronic schizophrenics. Can Psychiatr Assoc J *12:*209, 1967.

Javitt DC, Zukin SR: Biexponential kinetics of [$_3$H] MK-801 binding: Evidence for access to closed and open *N*-methyl-D-aspartate receptor channels. Mol Pharmacol *35:*387, 1989.

*Javitt DC, Zukin SR: Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry *148:*1301, 1991.

Luby ED, Cohen BD, Rosenbaum F, Gottlieb J, Kelley R: Study of a new schizophrenomimetic drug Sernyl. AMA Arch Neurol Psychiatry *81:*363, 1959.

McCarron MM, Schulze BW, Thompson GA, Conder MC, Goetz WA: Acute phencyclidine intoxication: Incidence of clinical findings in 1000 cases. Ann Emerg Med *10:*237, 1981.

*McCarron MM, Schulze BW, Thompson GA, Conder MC, Goetz WA: Acute phencyclidine intoxication: Clinical patterns, complications, and treatment. Ann Emerg Med *10:*290, 1981.

Michaelis EK: Glutamate neurotransmission: Characteristics of the NMDA receptors in mammalian brain. Neural Notes *2:*3, 1996.

Misra AL, Pontani RB, Bartolemeo J: Persistence of phencyclidine (PCP) and metabolites in brain and adipose tissue and implications for long-lasting behavioural effects. Res Commun Chem Pathol Pharmacol *24:*431, 1979.

Newmeyer JA: The epidemiology of PCP use in the late 1970s. J Psychedel Drugs *12:*211, 1980.

*Petersen RC, Stillman RC, editors: Phencyclidine (PCP) abuse: An appraisal. NIDA Res Monogr *21:*1, 1978.

Toth E, Lajtha A: Antagonism of phencyclidine-induced hyperactivity by glycine in mice. Neurochem Res *11:*393, 1986.

Tsai G, van Kammen DP, Chen S, Kelley ME, Grier A, Coyle JT: Glutamatergic neurotransmission involves structural and clinical deficits of schizophrenia. Biol Psychiatry *44:*667, 1998.

Yesavage JA, Freeman AM III: Acute phencyclidine (PCP) intoxication: Psychopathology and prognosis. J Clin Psychiatry *44:*664, 1978.

*Zukin SR, Sloboda Z, Javitt DC: Phencyclidine (PCP). In *Substance Abuse–A Comprehensive Textbook*, ed 3, JH Lowinson, P Ruíz, RB Millman, JG Langrod, editors. Williams & Wilkins, Baltimore, 1997.

Zukin SR, Zukin RS: Specific ^{3}H-phencyclidine binding in rat central nervous system. Proc Natl Acad Sci USA *76:*5372, 1979.

11.12 SEDATIVE-, HYPNOTIC-, OR ANXIOLYTIC-RELATED ABUSE

DOMENIC A. CIRAULO, M.D., AND OFRA SARID-SEGAL, M.D.

Sedative, hypnotic, or anxiolytic dependence is influenced by several factors: the most important is the pharmacology of drugs. Dependence is also affected by the beliefs and values of physicians, who must be aware of their attitudes toward dispensing drugs of this class. Finally, patients, and society at large, are influenced by the moral, economic, and political views of the time. Taken together these factors influence medical care, sometimes to the detriment of patients, who may be denied appropriate medication because of such biases.

DEFINITION

Determining Abuse Liability Any study of sedative, hypnotic, or anxiolytic dependence is limited by the operational definitions of abuse and dependence. Drug abuse is a behavioral and social phenomenon, albeit with significant pharmacological origins and medical consequences.

Abuse liability is established by asking a series of questions about a drug or class of drugs. Questions that originate in society's values lead to studies of illicit trafficking; population surveys of nonprescription use; audits of prescribing practices; monitoring of diversion from physicians' offices, pharmacies, or manufacturing sites; emergency room treatment; and overdose deaths. Questions regarding pharmacological characteristics also predict abuse liability: drug-induced euphoria and the development of tolerance, an abstinence syndrome, or cross-tolerance with known abusable substances may all predict risk of abuse. Finally, a number of ingenious methods for determining drug preferences in humans and animal models have been developed to determine abuse liability.

The terms “abuse” and “misuse,” refer to the use of a drug in a manner that is not consistent with generally accepted medical practice or social and legal custom (e.g., use without a valid prescription or deliberately to produce intoxication, pleasure, or a high). The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) defines abuse as a “maladaptive pattern of substance use manifested by recurrent and significant adverse consequences related to repeated use.” Some authorities limit *use* to situations in which a drug is taken for a legitimate medical or psychiatric disorder but is used in a way that is not consistent with medical practice. The distinction between *abuse* and *misuse* is not widely accepted, and the terms are used interchangeably here. *Recreational use* refers to use of drugs solely for their hedonic value. According to DSM-IV, substance dependence is “a cluster of cognitive, behavioral, and physiological symptoms indicating that the individual continues use of the substance despite significant substance-related problems. There is a pattern of repeated self-administration that usually results in tolerance, withdrawal, and compulsive drug-taking behavior.” The term *physiological dependence* is used according to the definition of the World Health Organization (WHO): a pathological state brought about by repeated administration of a drug and that leads to the appearance of a characteristic and specific group of symptoms (i.e., an *abstinence syndrome* or *discontinuance syndrome* when the drug is discontinued or, in the case of certain drugs, significantly reduced. The abstinence syndrome has specific characteristics referable to the autonomic nervous system (e.g., tremor, sweating, tachycardia, and startle response) and must be clearly distinguished from recurrence of the features of the underlying disease for which the drug was originally given. It can be reversed or attenuated by readministering the discontinued drug or administering a drug with cross-tolerance. Some authorities reserve the *discontinuance syndrome* for withdrawal syndromes that develop with therapeutic agents, to distinguish physiological dependence that develops in the course of treatment from dependence that develops as part of drug abuse. The terms are used interchangeably here.

Substances The drugs discussed in this section are referred to as anxiolytic or sedative-hypnotic drugs. The terminology for that group of drugs is not clearly established. Their sedative or calming effects are on a continuum with their hypnotic or sleep-inducing effects. Anxiolytic drugs cover a wide spectrum of pharmacological agents, and the terminology implies more specificity than actually exists. For most of these drugs, the differentiation of their sedative, hypnotic, and anxiolytic activities has more to do with marketing than with pharmacology. The broad group of drugs that historically has been included in that class exhibits considerable variation in clinical utility, toxicity, risk for abuse, and potential for diversion or recreational use. The abuse liability of those agents is classified in this section as follows: (1) benzodiazepine receptor agonists (2) barbiturates, and (3) miscellaneous sedative-hypnotic drugs with limited clinical use (meprobamate [Miltown], chloral hydrate [Noctec], ethchlorvynol [Placidyl], glutethimide, and methaqualone). In the practice of both psychiatry and addiction medicine, the drugs that are most important clinically are the benzodiazepines.

H3>ETIOLOGY AND NEUROPHARMACOLOGY

The benzodiazepines, barbiturates, and barbiturate-like substances all have their primary effects on the g-aminobutyric acid (GABA) type A ($GABA_A$) receptor complex, which contains a chloride ion channel, a binding site for GABA, and a well-defined binding site for benzodiazepines. The barbiturates and barbiturate-like substances are also believed to bind somewhere on the $GABA_A$ receptor complex. When a benzodiazepine, barbiturate, or barbiturate-like substance does bind to the complex, it increases the affinity of the receptor for its endogenous neurotransmitter, GABA, and increases the flow of chloride ions through the channel into the neuron. Benzodiazepines have little or no effect in the absence of GABA. At high doses, barbiturates directly activate chloride channels. The influx of negatively charged chloride ions into the neuron is inhibitory, since it hyperpolarizes the neuron relative to the extracellular space.

Although all the substances in this class induce tolerance and physical dependence, the underlying mechanisms are best understood for the benzodiazepines. Long-term benzodiazepine use attenuates the receptor effects caused by the agonist. Specifically, after long-term benzodiazepine use, GABA stimulation of the $GABA_A$ receptors results in less influx of chloride than it did before benzodiazepine administration. Benzodiazepine agonists lead to several processes that result in downregulation of the $GABA_A$ receptor. Benzodiazepine agonists produce acute desensitization of the receptor (tachyphylaxis) and promote $GABA_A$ receptor sequestration in cortical neurons. Chronic administration uncouples the GABA and benzodiazepine binding sites. μ-Receptor mechanisms underlying uncoupling are unknown, although changes in phosphorylation, altered subunit composition, and conformational changes in the receptor have been suggested. Contradictory findings have been reported with respect to alterations in receptor density. Binding affinity is not altered by chronic administration, although receptor alterations leading to uncoupling are almost certainly important in tolerance development. The changes in $GABA_A$ function do not completely explain the clinical phenomenon. For example, although patients develop tolerance to the sedative effects of benzodiazepines, the antianxiety and amnestic effects of these drugs are relatively persistent.

ABUSE LIABILITY

Surveys of prescribing practices, patient-initiated dosage changes, and recreational or nonmedical benzodiazepine use along with reports from emergency rooms, medical examiners, and law enforcement agencies all suggest that benzodiazepine abuse is not a major public health problem. They are rarely used for recreational purposes, and most studies suggest that they are not commonly the sole drug of abuse. The vast majority of medical and psychiatric patients use benzodiazepines appropriately, although they may be abused by patients dependent on alcohol or other drugs.

Abuse liability is evaluated, in part, by measuring the reinforcing properties of the drug. If its pharmacological effects increase a behavior (e.g., self-administration of a dose or the work required to permit self-administration), then the drug is a positive reinforcer and has abuse liability. Both animal and human studies have assessed the reinforcing properties of the benzodiazepines.

Animal Studies Animal studies demonstrated that benzodiazepines have minimal reinforcing effects. Although a few studies show that animals self-administer benzodiazepines more frequently than other drugs, they are consistently less potent in that regard than are drugs of abuse (e.g., cocaine) or even many other

sedative-hypnotics drugs, including most barbiturates.

Animal studies using schedule-induced polydipsia (the use of intermittent food administration to the experimental animal, which promotes ingestion of fluids, including drug solutions) indicate that midazolam (Versed) is ingested more often than water, but self-administration of chlordiazepoxide (Librium) and flurazepam (Dalmane) does not differ significantly from that of water. Prior exposure to ethanol or sedative-hypnotic drugs may increase the reinforcing effects of benzodiazepines, but some studies do not support that finding. Baboons clearly prefer ethanol but not diazepam or triazolam (Halcion) in the schedule-induced polydipsia model.

Another animal model predicting abuse liability consists of drug infusion in response to activity by the animal, such as lever pressing (Fig. 11.12-1). Drug infusion may allow either continuous or intermittent drug availability, and rates of administration of the drug are compared with those of the vehicle, other sedative-hypnotic drugs, and standard drugs of abuse such as cocaine. Studies with continuous drug access suggest that triazolam, diazepam (Valium), midazolam, clobazam, and chlordiazepoxide have reinforcing effects; studies with intermittent drug availability show that diazepam, midazolam, triazolam, alprazolam (Xanax), lorazepam (Ativan), chlordiazepoxide, and bromazepam maintain responses at a level higher than vehicle, which indicates abuse liability. With the possible exception of triazolam and midazolam, in studies using this model, the benzodiazepines consistently demonstrate lower abuse liability than do barbiturates other than phenobarbital, which also has relatively low abuse liability. Partial benzodiazepine receptor agonists, such as abecarnil and bretazenil, have low abuse liability in those models.

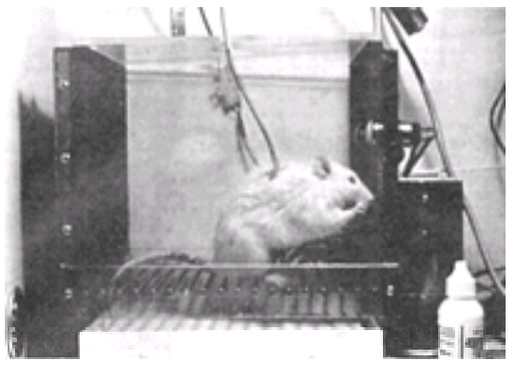

FIGURE 11.12-1 Animal in an operant chamber self-administers a drug by pressing a bar to receive an infusion of a predetermined dose that is delivered from a syringe pump via a cannula implanted in its jugular vein. (Courtesy of Conan Kornetsky, Boston University School of Medicine.)

Animal studies support lower abuse liability for the benzodiazepines than for most barbiturates. Among the benzodiazepines some evidence suggests that triazolam and midazolam may have the greatest potential for abuse, but more studies are required to define differences within the class.

Human Studies The assessment of abuse liability in humans relies on two predictive models: self-administration of benzodiazepines by experimental subjects who have a history of drug abuse and subjective responses that correlate with high abuse potential. In the first paradigm, one type of study allows subjects with a history of sedative-hypnotic dependence to self-administer orally (i.e., sample) different color-coded drugs and placebo under double-blind conditions in a simulated social environment. During subsequent sessions they are permitted to self-administer the drug they like the best. Some studies require that the subjects perform a task (e.g., riding a stationary bicycle) to earn doses. The abuse liability of different drugs is assessed by how often they are chosen for self-administration or by how much work a subject does to earn a dose. Scales measuring drug effect are also administered. The paradigm has the advantage of directly observing the behavior involved with ingesting the drug, albeit under somewhat artificial circumstances. It works best for former sedative-hypnotic abusers without psychiatric illness. It is probably safe to infer that those subjects are administering the drug to get high. For other subjects the reason for self-administration may not be clear. Do patients with anxiety disorder self-administer the drugs because they are effective anxiolytics or because they induce euphoria? Similarly, if subjects are in withdrawal from sedative-hypnotic drugs or ethanol, do they choose the agent that alleviates their withdrawal symptoms most effectively or the drug that produces the greatest euphoria? An additional problem with this design is that it exposes subjects to multiple doses of potentially addictive drugs.

An alternative strategy is to administer only a single dose of the test drugs and then compare subjective responses associated with abuse liability. A set of scales, which differs slightly among laboratories, usually consists of the Morphine Benzedrine Group subscale of the Addiction Research Center Inventory (ARCI-MBG), self-reports of sedation, liking, and intensity of drug effect; monetary street value of the drug; and similarity to reference drugs of abuse. Studies demonstrate that estimates of abuse liability using this paradigm are consistent with assessments based on the self-administration model. One disadvantage of the single-dose paradigm is that it measures only subjective effects, not behavior. It also has some of the same limitations as the self-administration paradigm. Does the subject like the drug because it has a therapeutic effect or because it is intoxicating? Both designs attempt to circumvent the problem by including an adequate battery of scales that measure several different dimensions of drug effect. Dosage equivalency and homogeneity of the subjects (especially with respect to a personal or family history of psychoactive substance abuse and psychiatric illness) are critical factors in the study design and in the interpretation of data from both models.

The predictive models in humans clearly establish that benzodiazepines occupy an intermediate position on the spectrum of abuse liability of sedative-hypnotic drugs. Methaqualone and barbiturates generally produce higher ARCI-MBG and drug-liking scores than do benzodiazepines. Buspirone (BuSpar), a nonbenzodiazepine anxiolytic agent (i.e., it is essentially devoid of benzodiazepine agonist activities), has little or no abuse liability in those models. With respect to differences among the benzodiazepines, the weight of the evidence suggests that diazepam has greater abuse liability than halazepam (Paxipam), oxazepam (Serax), chlordiazepoxide, or clorazepate (Tranxene). Diazepam, lorazepam, alprazolam, and triazolam exhibit similar profiles of abuse potential using those paradigms. The clinical relevance of the findings has not been adequately studied.

Flunitrazepam (Rohypnol), a benzodiazepine agonist, is not marketed in the United States but has entered the country through sources in Central and South America. Reports of abuse have appeared in both the clinical literature and public media. The drug has been implicated in date rape, in which flunitrazepam is added to a woman's alcoholic beverage to reduce her judgment, inhibition, and physical ability to resist sexual activity. Memory for the event may also be impaired by the combination of alcohol and flunitrazepam. A comprehensive review of the abuse liability of flunetazepam published in 1997 concluded that little evidence suggested that the drug had higher abuse liability than other benzodiazepines, except for its preference among opioid abusers. Nonetheless, the manufacturer has reformulated the drug, making it more difficult to disguise when added to ethanol. Also requiring further study is the relative roles of pharmacokinetic and pharmacodynamic facts in determining the abuse liability of various benzodiazepines. Some drugs (e.g., diazepam) may appear to have high abuse liability because of rapid absorption from the gastrointestinal tract after oral doses. Little is known about differences in abuse liability between full and partial agonists or among drugs that have preferential effects on specific subtypes of the benzodiazepine receptor. Zolpidem (Ambien), for example, is an imidazopyridine hypnotic agent with a rapid onset of action and a short elimination half-life. Although some evidence suggests that it is selective for the central benzodiazepine type 1 receptor (BZ 1), other studies suggest that zolpidem and triazolam depress energy metabolism in the same areas of the brain. With respect to subjective, psychomotor, and memory effects, there are no significant differences between a single 10-mg dose of zolpidem and a single 0.25-mg dose of triazolam. High doses of triazolam (0.75 mg) are identified more often as similar in effect to barbiturates, benzodiazepine, or alcohol than are high doses of zolpidem (45 mg). Zolpidem abuse and dependence have been reported.

Research in the 1990s investigating patient characteristics that influence abuse liability shows that when alcoholic persons and their adult children are given a single dose of alprazolam or diazepam, they have greater increases on abuse potential scales than do nonalcoholic controls without a family history of alcoholism. The differences in reinforcing effects are not attributable to pharmacokinetic differences as measured by peripheral plasma concentrations; however, pharmacodynamic differences have been noted using functional magnetic resonance imaging and electroencephalographic measures. A survey of 5000 male college students found that 23 percent of those with a positive family history of alcohol abuse reported using benzodiazepines (versus 0.9 percent of those without such a family history). Findings from another study indicate that moderate drinkers self-administer higher doses of diazepam than do light drinkers. Whether these results are drug specific or apply to the class of benzodiazepines as a whole is not known.

When using a self-administration model, subjects with anxiety do not choose diazepam more frequently than they do placebo, but when anxious subjects seeking treatment are tested, they show a high rate of diazepam selection. As no evidence from survey data or clinical experience suggests that patients with uncomplicated anxiety disorders have high rates of benzodiazepine abuse, these seemingly contradictory findings point out the limitations of self-administration in nondependent subjects (i.e., whether subjects with psychic distress take the drug for its therapeutic or euphoric effects).

EPIDEMIOLOGY

Several surveys have been conducted to determine the extent of benzodiazepine use by prescription and by illicit means. Drug sales to retail pharmacies provide one measure of benzodiazepine use. Sales of benzodiazepines are monitored by IMS Global Services, Ltd. The data indicate that benzodiazepine tranquilizer sales in the United States peaked in 1973 to 1975, when annual sales were approximately 87 million prescriptions. Between 1980 and 1983 sales of benzodiazepines to retail pharmacies in the United States declined from an average maintenance dose for the major indication per 1000 inhabitants per day of 17.5 to 15.8, but by 1985 it was back up to 17.1. Sales of drugs with long half-lives such as diazepam and flurazepam decreased, whereas sales of drugs with shorter half-lives such as alprazolam, triazolam, and temazepam (Restoril) increased. From 1986 to 1989 benzodiazepine tranquilizer sales declined to about 56 million prescriptions.

The prevalence of benzodiazepine use is presented in Table 11.12-1 Alprazolam accounted for 33 percent of the U.S. market share in 1989, diazepam for 25 percent and lorazepam for 19 percent. Of the benzodiazepine hypnotics, triazolam had a 49 percent market share in 1989, temazepam 30 percent, and flurazepam 21 percent. In 1989 the United States showed greatest volume of retail sales of benzodiazepines, but per capita exposure was at the median of the nine countries with the largest number of sales (United States, France, Japan, Italy, United Kingdom, Germany, Spain, Brazil, and Canada). The United States had a per capita exposure of 16.9 dosage units per year for benzodiazepines used as tranquilizers; France had the highest (55 dosage units per year), and Brazil had the lowest of the nine countries (7.3 dosage units per year). Despite having the lowest total sales volume, Canada had a per capita exposure of 30.9 dosage units per year. Rates of per capita exposure to benzodiazepine tranquilizers were three to five times higher than were rates of exposure to benzodiazepine hypnotic agents, except in the United Kingdom, where hypnotic use exceeded tranquilizer use (12.4 versus 8.9 per capita exposure). In 31 countries surveyed, benzodiazepine tranquilizer sales increased by 13 percent and benzodiazepine hypnotic sales increased by 47 percent from 1981 to 1989. Notable exceptions were the United Kingdom, where sales of both tranquilizers and hypnotics declined, and Germany, where a decline in tranquilizer sales was almost offset by an increase in hypnotic sales. In 1989, lorazepam had 20.7 percent of the world market share, diazepam had 17.1 percent, and alprazolam, 10.5 percent.

- Retail sales of benzodiazepines reached their highest level in 1973–1975 at 87 million a year but have declined gradually since then
- The United States has the highest volume of sales in the world, but per capita sales are at the median of the nine countries with the largest sales
- Sales of drugs with short elimination half-lives (e.g., alprazolam) have increased compared with drugs with long half-lives (e.g., diazepam)
- The National Institute on Drug Abuse Household Survey found that the percent of the general population reported nonmedical use of tranquilizers
- A survey of the United States population showed that self-reported medical use of a tranquilizer declined from 10.9 percent to 8.3 percent from 1970 to 1990; hypnotic use declined from 3.5 to 2.6 percent during the same period
- Persons who abuse alcohol and drugs use and abuse benzodiazepines at higher rates than do anxiety disorder patients without substance abuse histories

Table 11.12-1 Prevalence of Benzodiazepine Use

In 1990 over half of the patients surveyed in the United States who were taking tranquilizers took them for a month or less, and 25 percent took them for a year. Seventy percent of hypnotic users took them for less than a month, and 14 percent for a year or longer. Benzodiazepine sales worldwide continued to fluctuate between 1990 and 1996. In 1996 the United States accounted for the greatest volume of retail sales measured in U.S. dollars, followed by Japan, Italy, France, Germany, Brazil, Argentina, Spain, and Belgium. In 1996 alprazolam accounted for 13.2 percent of world market share in U.S. dollars; followed by clonazepam (12 percent) and lorazepam (11.4 percent), diazepam accounted for about 7.7 percent. In the United States clonazepam accounted for 28.2 percent of benzodiazepine market share in U.S. dollars followed by midazolam (20.1 percent) and alprazolam (15.6 percent).

Interview data are important supplements to prescription monitoring. Despite problems with validity and reliability, self-report data offer insight into the appropriateness of benzodiazepine use. In surveys of the U.S. population, self-reported tranquilizer use declined slightly from a 1970–1971 survey in which 10.9 percent of the adult population reported use of anxiolytics during the previous year to 8.3 percent in a 1990 survey. Hypnotic use declined from 3.5 to 2.6 percent during the same time period, with little change since 1979. Long-term users were more likely to be older and to be women, and they had high levels of emotional distress and chronic health problems. Over half of patients surveyed in 1990 took benzodiazepines for a month or less, and 25 percent took them for longer than 12 months, an increase from the 15 percent who were long-term users in 1979. The reasons for the increase are not clear, but they may include chronic benzodiazepine treatment for panic disorder, an aging population that is more likely to take benzodiazepines, or a decrease in the number of short-term users.

Seventy percent of hypnotic users took the drug for less than a month in 1990, and 14 percent used them for 12 months or longer, a slight increase from 1979, when 11 percent of users took the drug for 12 months or longer.

A survey of unsupervised changes in dosage indicated that 12 percent of patients had decreased their anxiolytic dosage, and 6 percent had increased it; 9 percent of patients taking hypnotics decreased their dosage, and 8 percent increased it. To put these unauthorized increases in perspective, the same survey indicated that 13 percent of patients have unauthorized increases in antidepressant dosages, and those drugs have low abuse liability. A prospective study found that patients were more likely to use antidepressants than benzodiazepines on a long-term basis, suggesting that long-term use should not, in itself, be used to measure abuse liability.

The National Institute on Drug Abuse's National Household Survey indicated that nonmedical use of tranquilizers was not a major public health problem. In 1994, the total percentage of the general population reporting nonmedical use of tranquilizers in their lifetime was 4 percent; 1.9 percent of those 12 to 17 years old, 4.5 percent of those 18 to 25 years old, 8.2 percent of those 26 to 34 years old, and 3 percent of those over 35 years of age reported nonmedical use. A total of 0.5 percent of the population reported nonmedical use of tranquilizers in the month prior to the survey, ranging from 0.2 to 0.6 percent among the different age groups.

Surveys of psychiatric patients (except those diagnosed as substance abusers) have demonstrated high rates of benzodiazepine prescription but almost uniformly low abuse. In one study of 2719 outpatients none of the 178 patients who had received benzodiazepines were diagnosed with abuse or dependence on the basis of criteria in the revised third edition of DSM (DSM-III-R). Studies of inpatient psychiatric patients suggest that the rate of benzodiazepine abuse or dependence ranges from 0.4 to 13 percent of admissions. A study from the University of Munich found that 6.7 percent of 9408 admissions had a diagnosis of benzodiazepine abuse or dependence, approximately half of whom were dependent on benzodiazepines alone. Lorazepam was the most commonly abused benzodiazepine, and oxazepam was the least, even though the latter drug was the most commonly prescribed benzodiazepine in the country.

Substantial evidence indicates that benzodiazepines are frequently used by patients who abuse other drugs. A survey of drug abuse treatment centers across the United States in 1979–1981 found that 22.6 percent of patients used minor tranquilizers (primarily benzodiazepines) weekly or more frequently. In a survey of opioid abusers in Sheffield, England, 90 percent reported benzodiazepine use daily or almostly daily. Use began with a prescription in one third of the sample. The most commonly used benzodiazepine was diazepam and the primary goal was relief from such symptoms as insomnia, withdrawal, and anxiety rather than a high. Lower rates of benzodiazepine use were reported in French (50 percent) and German (30 percent) studies. In the last two studies heroin addicts underreported benzodiazepine use, requiring toxicological studies to detect use.

Reports from Scotland since the mid-1980s indicated that intravenous drug abusers were using temazepam, often in combination with buprenorphine (Buprenex). One study reported that 92 percent of intravenous drug abusers had used temazepam in the year preceding the survey, and most of them had injected the drug intravenously. Marketed as a liquid-filled capsule, the fluid was extracted by a syringe and injected. That formulation was changed to a gel-filled capsule, which drug abusers liquefy by boiling with water or heating in a microwave prior to injection. Several cases of ischemia secondary to vasculitis and venous thrombosis were reported as a consequence of intravenous or intraarterial injection of the new formulation. In recent studies in Australia 37 percent of intravenous drug abusers reported benzodiazepine use in a typical drug-using month, paralleling the situation in the United States and Europe. A study in Denmark examined drugs in blood samples from unconscious drug addicts after the intake of an heroin overdose and found one or more benzodiazepines (usually diazepam) in 75 percent of patients. In a study from Canada of inpatients admitted for treatment of severe benzodiazepine dependence, patients had an extensive history of other lifetime and concurrent psychoactive substance use disorders. All patients had a prior history of other psychoactive drug use: 77 percent reported past opioid dependence and 53 percent past alcohol dependence. At admission, 83 percent of the patients were opioid dependent and 67 percent were cocaine dependent. In addition, patients in this study

had a high frequency of lifetime psychiatric disorders (major depression [33 percent] and panic disorder [30 percent]), and current psychiatric diagnoses (antisocial personality disorder [42 percent], avoidant personality disorder [25 percent], generalized anxiety disorder [20 percent], borderline personality disorder [17 percent], and panic disorder [13 percent]).

Several studies examined rates of benzodiazepine use among patients at methadone clinics. Three different U.S. clinics reported 22, 40, and 44 percent of their patients had benzodiazepines in their urine. European clinics report comparable rates. Flunitrazepam (where available) appears to be the preferred benzodiazepine. Usually taken orally, it may also be snorted. In other countries, diazepam, alprazolam, and lorazepam are most commonly abused. In a study from Austria, patients in a methadone clinic rated their preference for the effects of a variety of drugs. Benzodiazepines were rated behind heroin, cocaine, all opiates, cannabis, barbiturates, and stimulants. Flunitrazepam and diazepam were the highest-rated benzodiazepines, but alprazolam was not available in Austria when the survey began. In a study in three U.S. cities, methadone (Dolophine)-maintenance patients rated diazepam as producing the best high, with lorazepam and alprazolam also valued for their effects and considered all three significantly different from clorazepate, oxazepam, and chlordiazepoxide. Concern has been raised about abuse of alprazolam in methadone clinics, and at least one clinic reported that alprazolam use has surpassed diazepam use. Patients taking methadone presumably use benzodiazepines to boost the effects of methadone, but some studies have suggested that self-medication is also an important motivation. Opioid abusers who abuse benzodiazepines are more likely to engage in such other high-risk behaviors as needle sharing.

Cocaine abusers appear less likely to use benzodiazepines than are patients in methadone clinics, with alcohol and opioids the preferred secondary drugs of abuse. Benzodiazepines are used to attenuate the cocaine crash or to alleviate anxiety during cocaine use. These patients generally use therapeutic doses of benzodiazepines on an intermittent basis.

Surveys indicate that many alcoholic persons have taken or are currently taking benzodiazepines. They receive benzodiazepines for treatment of the alcohol withdrawal syndrome or coexisting anxiety or sleep disorders and often present for such treatment without revealing the extent of their alcohol use to the prescribing physician. Physicians must be cautious because evidence suggests an increased risk of benzodiazepine abuse in alcoholics; however, case reports document the efficacy of benzodiazepines in selected alcoholic persons with anxiety disorders.

Experimental studies have shown that moderate drinkers self-administer diazepam and that alcoholic persons have greater euphoric responses to a single dose of alprazolam than do control subjects. Both diazepam and alprazolam increase euphoria and liking scores in alcoholic persons and increase desire for ethanol as measured by self reports on visual analog scales. The changes in euphoria and liking are minimal with halazepam (compared with diazepam), suggesting a lower abuse liability in such patients.

Most survey data indicate that between 12 and 33 percent of alcoholic persons who enter treatment use benzodiazepines concurrently. Alcoholic women may have higher rates of concurrent use than alcoholic men. How many alcoholic patients misuse benzodiazepines is not entirely clear, but one study found that 15 percent of alcoholic outpatients who also used sedatives (mostly benzodiazepines) used them as prescribed, 24 percent had both used them appropriately and abused them, and 61 percent abused them. Another study reported that a little more than half of alcoholic outpatients with positive urine drug screens for benzodiazepines were abusing them. These data should be regarded cautiously because treatment population surveys are biased to select only patients for whom previous therapy has failed.

In Finland drug usage was evaluated in 1993 in drivers suspected of driving under the influence. Benzodiazepines were the most frequent drugs found in body fluid samples. Their usage went up from 6 percent in 1979 to 22.9 percent in 1993. Similarly, in Oslo, the drugs found most frequently in blood and urine samples of suspected drunken drivers were alcohol, tetrahydrocannabinol, benzodiazepines, and amphetamines.

If 33 percent of patients with alcoholism use benzodiazepines, and between 50 and 60 percent of those patients are abusing them, then approximately 15 to 20 percent of all alcoholic persons presenting for treatment may be abusing benzodiazepines. Benzodiazepines are used to self-medicate withdrawal symptoms or anxiety disorders, to produce euphoria, and to enhance the effects of ethanol. Other groups may also be at high risk for benzodiazepine abuse. In England and Wales, a 1995 pilot study in the prisons showed an increase in the number of urine samples that tested positive for opioids or benzodiazepines, from 4.1 percent in 1993–1994 to 7.4 percent in 1995. People with disabilities appear to abuse tranquilizers at up to twice the rate of people without disabilities.

DIAGNOSIS AND CLINICAL FEATURES

Diagnosis DSM-IV lists a number of sedative-, hypnotic-, or anxiolytic-related disorders (Table 11.12-2) but contains specific diagnostic criteria only for sedative, hypnotic, or anxiolytic intoxication (Table 11.12-3) and sedative, hypnotic, or anxiolytic withdrawal (Table 11.12-4). Other sedative-, hypnotic-, or anxiolytic-related disorders have their diagnostic criteria outlined in the DSM-IV sections that are specific for the major symptom (e.g., sedative-, hypnotic-, or anxiolytic-induced psychotic disorder.

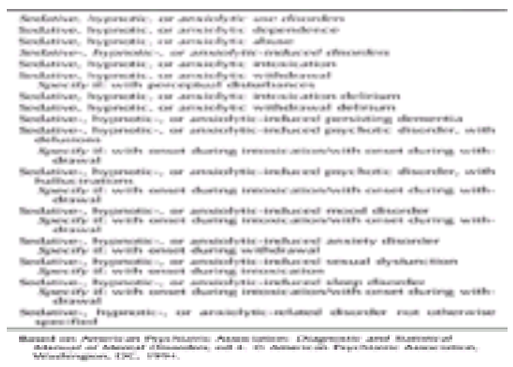

Table 11.12-2 DSM-IV Sedative-, Hypnotic-, or Anxiolytic-Related Disorders

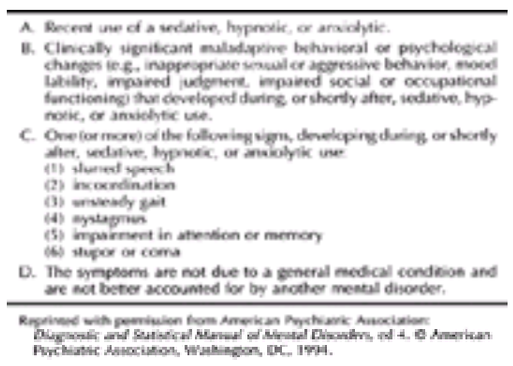

Table 11.12-3 DSM-IV Diagnostic Criteria for Sedative, Hypnotic, or Anxiolytic Intoxication

A. Cessation of (or reduction in) sedative, hypnotic, or anxiolytic use that has been heavy and prolonged.
B. Two (or more) of the following, developing within several hours to a few days after criterion A:
(1) autonomic hyperactivity (e.g., sweating or pulse rate greater than 100)
(2) increased hand tremor
(3) insomnia
(4) nausea or vomiting
(5) transient visual, tactile, or auditory hallucinations or illusions
(6) psychomotor agitation
(7) anxiety
(8) grand mal seizures
C. The symptoms in criterion B cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to a general medical condition and are not better accounted for by another mental disorder.

Specify if:
With perceptual disturbances

Table 11.12-4 DSM-IV Diagnostic Criteria for Sedative, Hypnotic, or Anxiolytic Withdrawal

Dependence and Abuse Sedative, hypnotic, or anxiolytic dependence and sedative, hypnotic, or anxiolytic abuse are diagnosed according to the general criteria in DSM-IV for substance dependence and substance abuse (see Table 11.1-3, Table 11.1-4, and Table 11.1-8).

Intoxication DSM-IV contains a single set of diagnostic criteria for intoxication by any sedative, hypnotic, or anxiolytic substance (Table 11.12-3). Although the intoxication syndromes induced by all those drugs are similar, subtle clinical differences are observable, especially with intoxications that involve low doses. The diagnosis of intoxication by one of that class of substances is best confirmed by obtaining a blood sample for substance screening.

Withdrawal DSM-IV contains a single set of diagnostic criteria for withdrawal from any sedative, hypnotic, or anxiolytic substance (Table 13.11-4). The clinician can specify “with perceptual disturbances” if illusions, altered perceptions, or hallucinations are present but are accompanied by intact reality testing. Benzodiazepines are associated with a withdrawal syndrome, and withdrawal from barbiturates can be life threatening. Withdrawal from benzodiazepines can also result in serious medical complications, such as seizures.

Delirium DSM-IV allows diagnosis of sedative, hypnotic, or anxiolytic intoxication delirium and sedative, hypnotic, or anxiolytic withdrawal delirium (see Table 10-22 and Table 10-23). Delirium that is indistinguishable from delirium tremens associated with alcohol withdrawal is more commonly seen with barbiturate withdrawal than with benzodiazepine withdrawal. Delirium associated with intoxication can be seen with either barbiturates or benzodiazepines if the dosages are high enough.

Persisting Dementia DSM-IV allows diagnosis of sedative-, hypnotic-, or anxiolytic-induced persisting dementia (see Table 10-31). The existence of the disorder is controversial, since whether a persisting dementia is due to the substance use itself or to associated features of the substance use is uncertain. The diagnosis will need further validation by DSM-IV criteria.

Persisting Amnestic Disorder DSM-IV allows diagnosis of sedative-, hypnotic-, or anxiolytic-induced persisting amnestic disorder (Table 10-39). Amnestic disorders associated with sedatives, hypnotics, and anxiolytics may have been underdiagnosed. One exception is the increased number of reports of amnestic episodes associated with short-term use of benzodiazepines with short half-lives (e.g., triazolam).

Psychotic Disorders The psychotic symptoms of barbiturate withdrawal can be indistinguishable from those of alcohol-associated delirium tremens. Agitation, delusions, and hallucinations are usually visual, but sometimes tactile or auditory features develop after about 1 week of abstinence. Psychotic symptoms associated with intoxication or withdrawal are much more common with barbiturates than with benzodiazepines and are diagnosed as sedative-, hypnotic-, or anxiolytic-induced psychotic disorders (see Table 13.3-4). The clinician can further specify whether delusions or hallucinations are the predominant symptoms.

Other Disorders Sedative, hypnotic, and anxiolytic use has also been associated with mood disorders (see Table 14.6-18), anxiety disorders (see Table 15.6-18), sleep disorders (see Table 21-18), and sexual dysfunctions (see Table 19.1a-20). When none of those diagnostic categories are appropriate for a person with a sedative, hypnotic, or anxiolytic use disorder, the appropriate diagnosis is sedative-, hypnotic-, or anxiolytic-related disorder not otherwise specified (Table 11.12-5).

The sedative-, hypnotic-, or anxiolytic-related disorder not otherwise specified category is for disorders associated with the use of sedatives, hypnotics, or anxiolytics that are not classifiable as sedative, hypnotic, or anxiolytic dependence; sedative, hypnotic, or anxiolytic abuse; sedative, hypnotic, or anxiolytic intoxication; sedative, hypnotic, or anxiolytic withdrawal; sedative, hypnotic, or anxiolytic intoxication delirium; sedative, hypnotic, or anxiolytic withdrawal delirium; sedative-, hypnotic-, or anxiolytic-induced persisting dementia; sedative-, hypnotic-, or anxiolytic-induced persisting amnestic disorder, sedative-, hypnotic-, or anxiolytic-induced psychotic disorder; sedative-, hypnotic-, or anxiolytic-induced mood disorder; sedative-, hypnotic-, or anxiolytic-induced sexual dysfunction; or sedative-, hypnotic-, or anxiolytic-induced sleep disorder.

Table 11.12-5 DSM-IV Diagnostic Criteria for Sedative-, Hypnotic-, or Anxiolytic-Related Disorder Not Otherwise Specified

Clinical Features and Patterns of Abuse

Oral Use Sedative, hypnotic, and anxiolytic drugs can all be taken orally, either occasionally to achieve a time-limited specific effect or regularly to obtain a constant, usually mild, intoxication. The occasional-use pattern is associated with young persons who take the substance to achieve specific effects—relaxation for an evening, intensification of sexual activities, and a short-lived mild euphoria. The user's personality and expectations about the substance's effects and the setting in which the substance is taken also affect the substance-induced experience. The regular-use pattern is associated with middle-aged, middle-class people who usually obtain the substance from the family physician as a prescription for insomnia or anxiety. Abusers of this type may have prescriptions from several physicians, and the pattern of abuse may go undetected until obvious signs of abuse or dependence are noticed by the person's family, coworkers, or physicians.

Intravenous Use A severe form of abuse involves the intravenous use of these substances. The users are mainly young adults intimately involved with illegal substances. Intravenous barbiturate use is associated with a pleasant, warm, drowsy feeling, and users may be inclined to use barbiturates more than opiates or opioids because of the low cost of barbiturates. The physical dangers of injection include transmission of the human immunodeficiency virus (HIV), cellulitis, vascular complications from accidental injection into an artery, infections, and allergic reactions to contaminants. Intravenous use is associated with rapid and profound tolerance, dependence, and a severe withdrawal syndrome. One survey reported that 28 percent of heroin addicts injected benzodiazepines, and these patients had higher rates of polydrug use, HIV risk-taking behavior, criminal activity, psychological distress, health problems, and overdose than addicts who did not inject benzodiazepines.

Benzodiazepines Since the introduction of chlordiazepoxide in 1960, benzodiazepines have become the primary drugs used to treat anxiety and insomnia, largely replacing barbiturates and other sedative-hypnotic agents. Zolpidem, an imidazopyridine hypnotic drug, is chemically distinct from the benzodiazepines but has similar clinical effects and acts at the $GABA_A$-benzodiazepine receptor complex. The benzodiazepines have lower abuse liability than most barbiturates, pose a much lower risk when taken in overdose, and have fewer interactions with other drugs. The advantages of the benzodiazepines must be weighed against the risk for abuse and physiological dependence.

Intoxication Benzodiazepine intoxication can be associated with behavioral disinhibition, potentially resulting in hostile or aggressive behavior. The effect is perhaps most common when benzodiazepines are taken in combination with alcohol. Benzodiazepine intoxication is associated with less respiratory depression than

barbiturate intoxication.

Withdrawal Syndrome

SIGNS AND SYMPTOMS Studies in the early 1960s by Leo Hollister established that abrupt discontinuation of high doses of chlordiazepoxide or diazepam could lead to a withdrawal syndrome. More-recent studies show that therapeutic doses given for weeks to months may also be associated with withdrawal syndrome.

The American Psychiatric Association's "Task Force Report on Benzodiazepine Dependence, Toxicity, and Abuse" defined withdrawal as a true abstinence syndrome consisting of "new signs and symptoms and worsening of preexisting symptoms following drug discontinuance that were not part of the disorder for which the drugs were originally prescribed." Many authorities have taken issue with that definition of withdrawal and prefer to distinguish withdrawal only from recurrence, not from rebound symptoms, viewing the true abstinence syndrome as consisting of rebound symptoms plus new signs and symptoms.

The signs and symptoms of the benzodiazepine discontinuation syndrome (Table 11.12-6) have been classified as major or minor, like those of the alcohol withdrawal syndrome. According to that classification, minor symptoms include anxiety, insomnia, and nightmares. Major symptoms (which are extremely rare) include grand mal seizures, psychosis, hyperpyrexia, and death.

The following signs and symptoms may be seen when benzodiazepine therapy is discontinued; they reflect the return of the original anxiety symptoms (recurrence), worsening of the original anxiety symptoms (rebound), or emergence of new symptoms (true withdrawal)

Disturbances of mood and cognition:
- Anxiety, apprehension, dysphoria, pessimism, irritability, obsessive rumination, paranoid ideation

Disturbances of sleep:
- Insomnia, altered sleep-wake cycle, daytime drowsiness

Physical signs and symptoms:
- Tachycardia, elevated blood pressure, hyperreflexia, muscle tension, agitation–motor restlessness, tremor, myoclonus, muscle and joint pain, nausea, coryza, diaphoresis, ataxia, tinnitus, grand mal seizures

Perceptual disturbances:
- Hyperacusis, depersonalization, blurred vision, illusions, hallucinations

Table 11.12-6 Signs and Symptoms of the Benzodiazepine Discontinuation Syndrome

The discontinuance syndrome may also be divided into symptoms of rebound, recurrence, and withdrawal. *Rebound symptoms* are symptoms for which the benzodiazepine was originally prescribed that return in a more severe form than they had before treatment. They have a rapid onset following termination of therapy and a brief duration. *Recurrence* refers to return of the original symptoms at or below their original intensity. The pattern and course of these symptoms will reflect the anxiety disorder for which treatment was originally instituted.

Withdrawal symptoms (Table 11.12-6) are loosely categorized into four types: (1) disturbances of mood and cognition, (2) disturbances of sleep, (3) physical signs and symptoms, and (4) perceptual disturbances. Mood and cognitive symptoms are anxiety, apprehension, dysphoria, irritability, obsessive ruminations, and paranoia. Sleep disturbances include insomnia, altered sleep-wake cycle, and daytime drowsiness. Somatic symptoms are agitation, tachycardia, palpitations, motor restlessness, muscle tension, tremor myoclonus, nausea, coryza, diaphoresis, lethargy, muscle and joint pain, hyperreflexia, ataxia, tinnitus, and seizures. Perceptual disturbances include hyperacusis, depersonalization, blurred vision, illusions, and hallucinations.

The temporal sequence of symptom development is not well established, but upon the abrupt cessation of benzodiazepines with short elimination-half-lives, symptoms may appear within 24 hours and peak at 48 hours. Symptoms arising from abrupt discontinuance of benzodiazepines with long half-lives may not peak until 2 weeks later. Although some investigators suggest that a subgroup of patients has withdrawal syndromes that last for many months, no medical or scientific evidence validates the existence of such a syndrome. Prolonged symptoms are almost certainly attributable to recurrence of the original anxiety or progression of the anxiety disorder itself.

RISK FACTORS Factors influencing the development of the discontinuance or withdrawal syndrome are listed in Table 11.12-7. Studies on both humans and animals have attempted to identify risk factors for the development of physiological dependence. Animal studies demonstrate that the severity of withdrawal is greater with higher dosages and longer periods of drug administration, but that has not been as consistently demonstrated in clinical populations. When patients are stratified into high- and low-dose groups, the withdrawal syndrome is seen to be more severe in the high-dose group. However, two recent studies, one using a daily 21-mg dosage of diazepam equivalent as the cutoff separating high- and low-dosage groups and the other using three groups based on diazepam daily dosage (less than 6 mg, between 6 and 10 mg, and more than 10 mg), indicate no significant differences in withdrawal symptoms between groups. One possible explanation for the failure to demonstrate a consistent relation between dose and withdrawal symptoms is that subjects in clinical studies are being treated for an anxiety disorder, and when the benzodiazepine is discontinued, the reemergence of the original anxiety symptoms may be confused with withdrawal symptoms. Furthermore, the drug dosage is probably higher in patients with the most severe disease, complicating any relation between dosage and the intensity of the discontinuance syndrome.

- Dosage of benzodiazepine
- Duration of benzodiazepine treatment
- Rate of drug taper
- Psychopathology

Table 11.12-7 Factors Influencing the Development of the Benzodiazepine Discontinuation Syndrome

According to one study withdrawal symptoms are related to treatment duration when the length of the treatment is less than 8 months but not when it is a year or longer. The findings of another group indicate that patients taking benzodiazepines for more than 5 years have more withdrawal symptoms than those taking them for less than 5 years.

Mild withdrawal symptoms may occur with abrupt discontinuation of therapeutic doses after 4 weeks of benzodiazepine treatment. There is a risk of rebound insomnia after a few days to 1 week of treatment with benzodiazepine hypnotic drugs with short elimination half-lives. Benzodiazepine hypnotics with longer elimination half-lives are less likely to induce rebound insomnia upon abrupt discontinuation. The likelihood of a serious withdrawal syndrome increases as treatment continues, and many authorities see 4 months of treatment at therapeutic doses as a critical point in the development of clinically significant physiological dependence. That does not imply, however, that 4 months is an upper limit for treatment duration.

The risk of withdrawal is also influenced by the rate at which the drugs are discontinued. Gradual tapering of the benzodiazepines with short elimination half-lives is associated with fewer withdrawal symptoms than is their abrupt discontinuation. Clinical research shows less distinct differences between gradual and abrupt discontinuation of benzodiazepines with long half-lives, either because these drugs have a self-tapering action or because the period of tapering and observation in

many studies is insufficient.

If benzodiazepines are abruptly discontinued, the withdrawal syndrome related to short half-life agents appears earlier and may be more intense than that with long half-life drugs. Differences in intensity have not been proved because most studies do not monitor withdrawal symptoms for 2 weeks, which may be the time of peak symptoms with some drugs with long half-lives. Furthermore, any difference in withdrawal severity between short and long half-life agents is not entirely supported by animal studies. There is no difference in symptom severity of the withdrawal syndrome after animals are treated with either midazolam or chlordiazepoxide, for example, because essentially all benzodiazepines, regardless of their pharmacokinetic properties in humans, are eliminated very rapidly in small-animal species.

The development of physiological dependence differs depending on whether it results from benzodiazepine partial agonists or full agonists. In animal models the partial agonists bretazenil and abecarnil are associated with fewer withdrawal symptoms than are such classic agonists as diazepam or clorazepate (the prodrug for desmethyldiazepam). Clinical studies in anxious patients, however, indicate that abrupt termination of abecarnil is associated with withdrawal symptoms in some patients.

Some studies in animals indicate that periodic administration of the benzodiazepine antagonist flumazenil (Romazicon) during the chronic administration of lorazepam, diazepam, triazolam, or clobazam may attenuate the withdrawal syndrome.

Personality traits may be a risk factor for development of the benzodiazepine withdrawal syndrome. Withdrawal severity is greater in patients with higher scores on the dependence scale of the Minnesota Multiphasic Personality Inventory–2 (MMPI-2), high prewithdrawal levels of anxiety and depression, lower educational level, and passive-dependent personality disorder.

Overdose The benzodiazepines, in contrast to the barbiturates and the barbiturate-like substances, have a large margin of safety when taken in overdoses, a feature that contributed significantly to their rapid acceptance. The ratio of lethal to effective dose is about 200 to 1 or higher because of the minimal respiratory depression associated with the benzodiazepines. Even when grossly excessive amounts (more than 2 grams) are taken in suicide attempts, the symptoms include only drowsiness, lethargy, ataxia, some confusion, and mild depression of the user's vital signs. A much more serious condition prevails when benzodiazepines are taken in overdose in combination with other sedative-hypnotic substances, such as alcohol. In such cases small doses of benzodiazepines can cause death. The availability of flumazenil, a specific benzodiazepine antagonist, has reduced the lethality of the benzodiazepines, since flumazenil can be used in emergency rooms to reverse the effects of the benzodiazepines.

Barbiturates Since the advent of the benzodiazepines, barbiturates use has been limited in modern medicine. Phenobarbital (Solfoton) is still prescribed as an anticonvulsant and as a sedative, especially for children. It is also a common component of many combination products and reduces the stimulating effects of sympathomimetic agents. Butalbital is an intermediate-acting barbiturate found in a widely used combination product that also contains acetaminophen and caffeine (Fiornal) and is approved for the treatment of muscle contraction headaches. Given the limited availability of barbiturates, it is uncommon, at least in the United States, for addiction units to treat patients dependent on barbiturates other than butalbital and phenobarbital. The barbiturates marketed currently in the United States are listed in Table 11.12-8.

The substitution technique for sedative-hypnotic withdrawal requires calculation of equivalent doses of phenobarbital to replace the sedative-hypnotic agent that the patient is taking. The following are doses of various sedative-hypnotic agents for which a 30-mg dose of phenobarbital should provide adequate coverage of a withdrawal syndrome. Daily doses of phenobarbital should rarely exceed 600 mg using this protocol.

Generic Name	Trade Name	Dose (mg)
Amobarbital	Amytal	100
Aprobarbital	Alurate	40
Butabarbital	Butisol	100
Butalbital	Many combination products (e.g., Fiorinal)	100
Pentobarbital	Nembutal	100
Secobarbital	Seconal	100
Chloral hydrate	Generic	500
Ethchlorvynol	Placidyl	500
Glutethimide	Generic	250
Meprobamate	Miltown	400

Table 11.12-8 Barbiturates and Other Sedative-Hypnotic Agents

A major disadvantage of the use of the barbiturates is the development of pharmacokinetic and pharmacodynamic tolerance. *Tolerance* is defined as reduced drug response as a result of either decreased drug concentration at the site of action, usually the result of increased drug metabolism (pharmacokinetic) or of cellular adaptive changes with unchanged or higher drug concentrations at the site of action (pharmacodynamic). Pharmacodynamic tolerance begins after acute doses and continues to develop over weeks to months. Tolerance to the mood-altering and sedative effects develops to a greater extent than does tolerance to the lethal effects, increasing the risk of accidental overdose.

Intoxication When barbiturates and barbiturate-like substances are taken in relatively low doses, the clinical syndrome of intoxication is indistinguishable from that associated with alcohol intoxication. The symptoms include sluggishness, incoordination, difficulty in thinking, poor memory, slowness of speech and comprehension, faulty judgment, disinhibition of sexual and aggressive impulses, a narrowed range of attention, emotional lability, and exaggeration of basic personality traits. The sluggishness usually resolves after a few hours, but the impaired judgment, distorted mood, and impaired motor skills may remain for 12 to 24 hours, depending primarily on the half-life of the abused substance. Other potential symptoms are hostility, argumentativeness, moroseness, and (occasionally) paranoid and suicidal ideation. The neurological effects include nystagmus, diplopia, strabismus, ataxic gait, positive Romberg's sign, hypotonia, and decreased superficial reflexes.

Dependence and Withdrawal Physiological dependence may develop after a daily dose of 400 mg of pentobarbital for 3 months; abrupt discontinuation results in paroxysmal abnormalities on the electroencephalogram (EEG) in about 30 percent of patients. At a daily dosage of 600 mg of pentobarbital for 1 to 2 months, a withdrawal syndrome characterized by anxiety, insomnia, anorexia, tremor, and EEG changes occurs in about half of the patients, and 10 percent may have a single seizure. At higher dosages of 800 to 2200 mg a day for several weeks to months, abrupt discontinuation leads to minor symptoms within 24 hours of the last dose, which include apprehension and uneasiness, insomnia, muscular weakness, twitches, coarse tremors, myoclonic jerks, postural faintness and orthostatic hypotension, anorexia, vomiting, and EEG changes. Minor symptoms may persist up to 2 weeks. At high doses major symptoms develop. As many as 75 percent of patients may have grand mal seizures on the second or third day after withdrawal, and two thirds have more than one seizures. The interictal EEG shows four spike-wave discharges per-second. Two thirds of these patients develop delirium between the third and eighth day of withdrawal, which is sometimes accompanied by hypothermia, which may be fatal. Disorientation, visual hallucinations, and frightening dreams may precede the onset of full delirium. Delirium may be exceedingly difficult to reverse, even with large doses of a barbiturate; thus clinicians should never wait for the appearance of withdrawal symptoms before instituting therapy. The duration of the withdrawal syndrome is between 3 and 14 days; most end by the eighth day.

Overdose Barbiturates are lethal when taken in overdose, because they induce respiratory depression. In addition to intentional suicide attempts, accidental or unintentional overdoses are common. Barbiturates in home medicine cabinets are a common cause of fatal drug overdoses in children. As with benzodiazepines the lethal effects of barbiturates are additive to those of other sedative-hypnotic drugs, including alcohol and benzodiazepines. Barbiturate overdose is characterized by induction of coma, respiratory arrest, cardiovascular failure, and death.

The lethal dose varies with the route of administration and the degree of tolerance for the substance after a history of long-term abuse. For the most commonly abused barbiturates, the ratio of lethal to effective dose ranges between 3 to 1 and 30 to 1. Dependent users often take an average daily dose of 1.5 grams of a short-acting barbiturate, and some have been reported to take as much as 2.5 grams a day for months. The lethal dose is not much greater for the long-term abuser than it is for the neophyte. Tolerance develops quickly to the point at which withdrawal in a hospital becomes necessary to prevent accidental death from overdose.

Miscellaneous Sedative-Hypnotic Drugs The drugs discussed below have little or no role in modern therapeutics. Unfortunately, prescriptions for some have been increasing, especially in New York, where a 1989 triplicate prescription program has led to a 44 percent decline in benzodiazepine prescribing in retail pharmacies, 60 percent in the Medicaid program, and 30 percent in Blue Cross–Blue Shield programs. Corresponding prescription increases of 125 percent for meprobamate, 29

percent for ethychlorvynol, and 136 percent for chloral hydrate have been reported.

The older sedative-hypnotics drugs are similar in effect to the barbiturates, and they lead to tolerance and physiological dependence. The withdrawal syndrome is similar to the barbiturate withdrawal syndrome, and the protocols described for barbiturates should be used for safe withdrawal of patients dependent on those drugs.

Meprobamate Meprobamate, a carbamate derivative, has weak efficacy as an antianxiety agent. At clinical doses in humans, it has minimal muscle-relaxant effects, but it may have a mild analgesic effect in musculoskeletal pain and may potentiate analgesics. Typical daily doses are between 1200 and 1600 mg, with a maximum of 2400 mg, in three or four divided daily doses. A discontinuance syndrome occurs after several weeks of treatment with a daily dosage of 2400 mg and mild symptoms may be seen with long-term therapy at doses of 1600 mg daily. Onset of the discontinuance syndrome occurs within 12 to 48 hours after abrupt discontinuation and lasts for an additional 12 to 48 hours. Seizures may be common in withdrawal from meprobamate and reports published in the 1970s suggest that serious withdrawal symptoms are more common with meprobamate than with barbiturates. Fatalities from overdose have been reported after doses as low as 12 grams. The manufacturer offers the following guidelines for blood concentration: 5 to 20 µg/mL is the normal range with recommended doses; 30 to 100 µg/mL is associated with mild-to-moderate overdosage, with patients in stupor or light coma; 100 to 200 µg/mL is associated with serious overdose, deeper coma, and fatalities; concentrations above 200 µg/mL are associated with more fatalities than survivals. Meprobamate induces microsomal enzymes and may exacerbate intermittent porphyria.

Chloral Hydrate Chloral hydrate is used as a hypnotic in doses of 0.5 to 2.0 grams. After oral administration it is rapidly transformed to trichloroethanol, the metabolite responsible for its pharmacological activity. It is also effective in blocking experimentally induced seizures; however, anticonvulsive and hypnotic doses are similar in humans, making the benzodiazepines and barbiturates better choices for anticonvulsive medications.

At high doses chloral hydrate may produce respiratory depression, hypotension, gastric necrosis, depressed cardiac contractility, and a shortened refractory period. Even at therapeutic doses it is associated with gastric distress and flatulence. Somnambulism, disorientation, and paranoid ideation may occur. Tolerance and physiological dependence develop with long-term use, and the withdrawal syndrome is similar to the barbiturate abstinence syndrome. Several drug interactions occur with chloral hydrate, including displacement of ral anticoagulants and acidic drugs from protein binding sites by the metabolite trichloroacetic acid. When combined with furosemide, flushing, tachycardia, hypotension or hypertension, and diaphoresis have been reported. It potentiates the effects of ethanol, and it should not be used in patients with intermittent porphyria. Fatalities from overdose may occur with as little as 4 grams, although patients have survived after taking 30 grams.

Ethchlorvynol Ethchlorvynol is a rapidly acting sedative-hypnotic agent with anticonvulsant and muscle-relaxant properties. Recommended doses are 500 to 1000 mg. Initial effects (which may include euphoria) occur 15 to 30 minutes after an oral dose and plasma levels peak at 1 to 1.5 hours. The elimination half-life of the parent compound is 10 to 20 hours. The duration of the hypnotic effect is about 5 hours. Long-term use of ethchlorvynol has been associated with toxic amblyopia, scotoma, nystagmus, and peripheral neuropathy, which are usually reversible upon drug discontinuation. Intravenous self-administration may cause pulmonary edema.

Idiosyncratic reactions of excitement and stimulation have been described. Hypersensitivity reactions with urticaria and rare fatal thrombocytopenia may also arise. It should not be used in patients with intermittent porphyria. Tolerance, physiological dependence, and a withdrawal syndrome are seen with long-term administration. Lethal doses are typically between 10 and 25 grams; one person was reported to have died after taking 2.5 grams with ethanol, and one patient survived but was in a coma for 7 days after taking 50 grams.

Glutethimide and Methyprylon Glutethimide and methyprylon (no longer marketed in the United States) are piperidinedione sedative-hypnotic drugs that have high liability for abuse. Glutethimide resembles the barbiturates in most respects, but differs from them in having significant anticholinergic activity. An overdose can cause ileus, bladder atony, mydriasis, hyperpyrexia, myoclonic jerks, and convulsions. A lethal dose is between 10 and 20 grams; intoxication may be seen at a dose of 5 grams. An unusual characteristic of glutethimide is that a withdrawal syndrome can occur in patients taking therapeutic doses (0.5 to 3.0 grams daily) even when they have not stopped taking the drug. The abstinence syndrome includes tremulousness, nausea, tachycardia, fever, tonic muscle spasms, and grand mal seizures. Withdrawal from a combination of glutethimide and antihistamine has caused catatonia and dyskinesia. When glutethimide is taken with codeine (the street name is a “load”) a euphoria similar to that with heroin results. In these patients, detoxification may require both opioid withdrawal and sedative withdrawal.

Methaqualone Methaqualone is no longer available in the United States, and much of what is sold on the street as methaqualone is actually diazepam. Methaqualone is a quinazoline with sedative, anticonvulsive, local anesthetic, and antispasmodic activity. It has antitussive effects comparable to those of codeine and weak antihistaminic activity. During the 1960s and 1970s, those in the drug culture considered it to provide the ultimate high of all sedative-hypnotic agents, producing euphoria comparable to that from heroin. Chronic users may take daily dosages between 75 mg and 2 grams, with an average daily dosage of 775 mg. Adverse effects include peripheral neuropathy, nightmares, somnambulism, gastric discomfort, and urticaria. Death has been reported after 8 grams, but one report attributed most deaths under the influence of methaqualone to accidents due to impairment produced by the drug. Tolerance and physiological dependence are seen with chronic dosing.

An overdose of methaqualone may result in restlessness, delirium, hypertonia, muscle spasms, convulsions, and (in very high doses) death. Unlike barbiturates, methaqualone rarely causes severe cardiovascular or respiratory depression, and most fatalities result from combining methaqualone with alcohol.

TREATMENT

Withdrawal

Benzodiazepines Guidelines for the treatment of benzodiazepine discontinuance syndrome and the clinical management of withdrawal are presented in Table 11.12-9. The most common clinical situations that require medically supervised withdrawal from the benzodiazepines involve patients with anxiety disorders or chronic insomnia who have been maintained on therapeutic doses of benzodiazepines for several months or years, patients taking supratherapeutic doses of benzodiazepines for treatment-resistant types of anxiety or because of inappropriate dosage escalation, and patients who abuse benzodiazepines as part of a mixed abuse pattern (e.g., those who use benzodiazepines to self-medicate alcohol withdrawal or to end a cocaine run).

1. Evaluate and treat concomitant medical and psychiatric conditions
2. Obtain drug history and urine and blood sample for drug and ethanol assay
3. Determine required dose of benzodiazepine or barbiturate for stabilization, guided by history, clinical presentation, drug-ethanol assay, and in some cases challenge dose
4. Detoxification from supratherapeutic dosages:
   a. Hospitalize if there are medical or psychiatric indications, poor social supports, or polysubstance dependence or the patient is unreliable
   b. Some clinicians recommend switching to longer-acting benzodiazepine for withdrawal (e.g., diazepam, clonazepam); others recommend stabilizing on the drug that patient was taking or on phenobarbital
   c. After stabilization reduce dosage by 10 percent on the second or third day and evaluate the response, keeping in mind that symptoms that occur after decreases in benzodiazepines with short elimination half-lives (e.g., lorazepam) appear sooner than with those with longer elimination half-lives (e.g., diazepam)
   d. Reduce dosage further by 10 to 25 percent every few days if tolerated
   e. Use adjunctive medications if necessary (e.g., carbamazepine, β-adrenergic receptor antagonists, valproate, clonidine, and sedative antidepressants have been used but their efficacy in the treatment of the benzodiazepine abstinence syndrome has not been established)
5. Detoxification from therapeutic dosages:
   a. Initiate 10 to 25 percent dose reduction and evaluate response
   b. Dose, duration of therapy, and severity of anxiety influence the rate of taper and need for adjunctive medications
   c. Most patients taking therapeutic doses have uncomplicated discontinuation
6. Psychological interventions may assist patients in detoxification from benzodiazepines and in the long-term management of anxiety

Table 11.12-9 Guidelines for Treatment of Benzodiazepine Discontinuance Syndrome

In planning a rational withdrawal clinicians should remember the factors that influence the development of physiological dependence. The risk of withdrawal is greatest in patients taking high doses over long periods. The onset of withdrawal and its peak severity are earlier with agents with short half-lives, and if they are not tapered, withdrawal from them may be more severe than that from agents with long elimination half-lives, although the last point has not been definitively established. Concurrent use of other sedative-hypnotic agents (e.g., barbiturates or ethanol) may alter both the time course and severity of the benzodiazepine withdrawal syndrome.

TAPERING The basic principle underlying safe withdrawal is gradual tapering of the benzodiazepine dosage. In cases of detoxification from therapeutic doses, the

daily dosage is known, and the tapering protocol is easily calculated. Anyone taking a benzodiazepine for 2 weeks or longer should be tapered from the drug. Initially, the dosage is reduced by approximately 10 to 25 percent, and the patient is observed for any signs and symptoms of withdrawal. Withdrawal signs appear earlier for drugs with short half-lives than for drugs with long half-lives. Subsequent reductions depend on how the patient responds to initial changes and require individualization based on careful observation of the patient's condition following each dosage change.

In some cases patients taking high therapeutic doses of a benzodiazepine for a year or more may require several months or longer to stop the medication completely, not primarily for pharmacological reasons, but so they can learn alternative strategies for coping with anxiety. Most patients taking midrange therapeutic doses for a shorter time tolerate weekly reductions ranging from 10 to 25 percent. Most authorities agree that the last phases of drug discontinuation are the most difficult for the patient, and it may be necessary to slow the taper rate or to stop the taper entirely. The adjunctive use of nonbenzodiazepine medications may be useful in these circumstances.

Detoxification from supratherapeutic dosages of benzodiazepines may require a slightly different approach. Most clinicians hospitalize such patients because of the greater medical risks associated with supratherapeutic dosage withdrawal. Patients may be tapered using the benzodiazepine they have been taking or switched to a drug with a long elimination half-life, such as diazepam or clonazepam (Klonopin), using the dose equivalencies listed in Table 11.12-10 to determine the initial daily dosage. The estimated equivalent dosage is then administered in divided doses on day 1 to ensure that an accurate history and appropriate equivalency are established. Some clinicians stabilize patients on that dosage for 2 to 3 days; others prefer to reduce the dose by 30 percent on the second day, followed by 5 to 10 percent daily reductions. The use of carbamazepine (Tegretol) or valproate (Depakote) in high-dosage withdrawal often permits even larger daily reductions in the benzodiazepine dose. Most high-dose detoxification can be completed in 2 weeks or less using this protocol. For some patients the dose cannot be reduced that rapidly, however, and slower, longer tapers are sometimes necessary. The efficacy of anticonvulsants for the treatment of the benzodiazepine discontinuance syndrome has not been established, although clinical experience to date is encouraging.

Generic Name	Trade Name	Dose (mg)
Alprazolam	Xanax	1
Chlordiazepoxide	Librium	25
Clonazepam	Klonopin	0.5–1
Clorazepate	Tranxene	15
Diazepam	Valium	10
Estazolam	ProSom	1
Flurazepam	Dalmane	30
Lorazepam	Ativan	2
Oxazepam	Serax	30
Prazepam	Paxipam	80
Temazepam	Restoril	20
Triazolam	Halcion	0.25
Quazepam	Doral	15
Zolpidem*	Ambien	10

* An imidazopyridine benzodiazepine agonist.

Table 11.12-10 Approximate Therapeutic Equivalent Doses of Benzodiazepines

P>When the physician knows the patient has been taking supratherapeutic dosages but cannot determine the exact dosage, 20 mg of diazepam may be given to estimate tolerance. The dose should be repeated every 2 hours until mild sedation occurs. The total dose required to induce mild sedation is then considered the initial dose, and the detoxification proceeds as described above.

Some concerns have been raised regarding a lack of cross-tolerance between triazolobenzodiazepines and other benzodiazepines. Those concerns are based entirely on anecdotal reports and are not supported by controlled trials or experimental data. Case reports suggest that diazepam does not adequately treat the withdrawal syndrome from alprazolam and that the combination of chlordiazepoxide and diazepam fails to attentuate withdrawal symptoms from trialozam, although questions have been raised concerning the adequacy of the diazepam doses in those cases. Others report that lorazepam can successfully treat triazolam withdrawal and that clonazepam can be substituted for alprazolam in patients with panic disorder. If the primary drug that the patient is taking is not going to be used in the withdrawal protocol, clonazepam or lorazepam may be the best drugs to use when detoxifying patients who have been taking supratherapeutic dosages of triazolobenzodiazepines. It is not clear whether higher doses of diazepam would have alleviated the alprazolam withdrawal syndrome in the published case reports. Some authorities recommend the use of phenobarbital in such cases.

DETOXIFICATION FROM MULTIPLE DRUGS OF ABUSE Polysubstance abuse may present special problems for safe withdrawal. The most common polysubstance abuse pattern influencing medical management of detoxification from benzodiazepines is the concurrent use of alcohol, opioids, and cocaine. Combination with alcohol alters the time course and increases the severity of the sedative withdrawal syndrome. Adequate dosages of a benzodiazepine usually suffice for safe withdrawal; however, in rare instances even large dosages of diazepam can be ineffective, and switching to a barbiturate becomes necessary.

When supratherapeutic doses of benzodiazepines are abused with large amounts of opioids, it is best to discontinue the former drug first and then stabilize the patient on methadone or some oral opioid. When the period of risk for major symptoms of sedative-hypnotic withdrawal passes, opioid withdrawal may begin. Addiction to low or moderate dosages of benzodiazepines or opioids often permits simultaneous withdrawal from both classes of drugs.

Benzodiazepines are often used to terminate a cocaine or other psychostimulant run or to decrease anxiety induced by cocaine. Benzodiazepines used intermittently in that manner usually do not require detoxification. The chronic use of high dosages of benzodiazepines with cocaine is less common but does require medically supervised benzodiazepine taper. The cocaine abstinence syndrome is characterized by depression, irritability, lethargy, amotivation, hypersomnolence, confusion, and drug craving. The clinical presentation of the sedative-hypnotic withdrawal syndrome may be altered by recent cocaine use and is best monitored by carefully measuring vital signs (e.g., increases in pulse, temperature, and blood pressure).

Some clinicians prefer to use phenobarbital for detoxification of patients with mixed-drug abuse. A dose of phenobarbital equivalent to the benzodiazepine and other sedatives the patient is taking is calculated (Table 11.12-8 and Table 11.12-10) or, preferably, the dose is determined from a challenge test, and the patient is stabilized on that amount in divided daily doses for the period of peak risk of withdrawal. The maximum daily dosage of phenobarbital is 600 mg. After stabilization, phenobarbital dosage is reduced by 30 to 60 mg every 2 or 3 days.

ADJUNCTIVE MEDICATIONS Several adjunctive medications are used to treat the benzodiazepine withdrawal syndrome, but in no case has their efficacy been consistently supported in controlled trials nor is there evidence that any specific adjunctive medication is superior to another.

Carbamazepine, given in doses that achieve the therapeutic concentrations for treatment of seizures, may partially attenuate the withdrawal syndrome from benzodiazepines. Initial doses are 200 mg twice a day, although some clinicians recommend a single dose of 400 mg at bedtime to take advantage of the sedative effect. If initial doses are well tolerated, carbamazepine is increased to 600 mg on the second or third day. Typical total daily doses range from 400 to 800 mg. After several days of treatment benzodiazepine dosage is tapered. Despite reports that tapering can be completed within a week, even from supratherapeutic dosages, some patients require a slow taper that lasts several weeks. Carbamazepine can be tapered quickly 2 to 4 weeks after the last benzodiazepine dose. Carbamazepine may not be successful in withdrawing patients with panic disorder because it is ineffective in blocking panic attacks, but has shown some success even in these patients, as long as a slow benzodiazepine taper is used. The efficacy of other anticonvulsants is undetermined, but clinical experience suggests that valproate may reduce withdrawal symptoms as well. Enteric-coated tablets of divalproex (Depakene) are used in starting dosages of 500 to 1000 mg divided in two or three daily doses. Dosages are increased to achieve serum concentrations of 50 to 120 µg/mL. Some clinicians prefer to use daily loading doses of 20 mg/kg. For patients who abuse alcohol and benzodiazepines, valproate should be avoided when liver enzyme activities are three times their normal values. Gabapentin (Neurontin) may also reduce benzodiazepine withdrawal symptoms, but experience is limited.

The b-adrenergic receptor antagonists may also attenuate withdrawal symptoms. Propranolol (Inderal) in doses of 60 to 120 mg may reduce the symptoms of benzodiazepine withdrawal, although tachycardia and elevated blood pressure are the primary symptoms affected, while the subjective sense of discomfort and dysphoria persists. The b-adrenergic receptor antagonists do not have prophylactic effects against seizures and should not be used in monotherapy for withdrawal. Nevertheless, in some cases propranolol or atenolol (Tenormin), 50 to 100 mg, may be helpful adjuncts to withdrawal treatment.

Clonidine (Catapres), 0.1 mg administered twice a day to 0.2 mg three times a day or as a patch, has been used in the treatment of low-dose benzodiazepine

withdrawal with variable results. Clonidine does not provide protection against withdrawal seizures. It may be effective only when used before withdrawal symptoms develop.

Although antidepressants have not been rigorously studied in the treatment of benzodiazepine withdrawal, many experienced clinicians prescribe them. Agents with well-established antipanic and sedative effects may be useful adjunctive medications for patients whose anxiety recurs during taper or after drug discontinuation. Imipramine (Tofranil), amitriptyline (Elavil), doxepin (Adapin, Sinequan), nefazodone (Serzone), trazodone (Desyrel), selective serotonin reuptake inhibitors, and monoamine oxidase inhibitors may all be useful in that role. Melatonin is a useful treatment for insomnia during withdrawal.

PSYCHOLOGICAL TREATMENTS Specific psychological treatments for withdrawal have not been widely studied. Cognitive behavioral strategies for anxiety management have been applied to patients withdrawing from benzodiazepines with mixed results. Patients who meet the criteria for generalized anxiety disorder are poor candidates for relaxation training.

Many practitioners believe that successful withdrawal requires transmitting to the patient a sense of control over the withdrawal symptoms. In support of this belief, placebo substitution is associated with fewer withdrawal symptoms after drug discontinuation than is withdrawal without placebo. Apparently the belief that one is taking an effective medication is enough to reduce symptoms.

One way to foster a sense of control over withdrawal symptoms is to link the symptoms of anxiety to environmental intrapsychic stressors. Cognitive restructuring teaches patients about the withdrawal syndrome and helps them identify and relabel withdrawal symptoms as anxiety, thus permitting the implementation of adaptive coping strategies, which include systematic desensitization, in vivo graded exposure, and group problem solving. Using a diary to record mood states and their precipitants sometimes helps cognitive restructuring. A diary may also alert the clinician to maladaptive responses to benzodiazepine withdrawal, such as increased alcohol consumption.

The concept of “cognitive coping” should be introduced to patients, and examples of self-statements (“talking to yourself”) that are applicable to withdrawal and anxiety management should be discussed and provided as flash cards. They include such statements as “I feel nervous now but it won't last,” “I feel bad but it's not worse than when I had the flu,” or “I feel uncomfortable, but nothing bad is going to happen to me.” Self-statements should be short, simple, and relevant to the patient.

Patients are encouraged to be active participants in developing their withdrawal schedule. Clinicians should be flexible about the rate of taper and should encourage the patient to join in the withdrawal process, especially those who are withdrawing from therapeutic doses. It is often useful to request that patients choose which doses should be decreased or eliminated according to their perceived need. For some patients, setting a maximum permissible daily dose may be necessary.

LONG-TERM OUTCOME Few adequately designed follow-up studies have been carried out on patients who terminated benzodiazepine treatment. Many studies suffer from sample selection bias, diagnostic heterogeneity of the sample, and failure to assess environmental factors or to control for concurrent psychotherapy or psychoactive medication; but despite those shortcomings, several interacting findings have emerged.

In one study of patients who had discontinued benzodiazepines from 10 months to 3.5 years earlier, 70 percent were no longer taking benzodiazepines and had few or no symptoms, 22 percent were not taking benzodiazepines but required other psychotropic medications, and 8 percent had returned to benzodiazepine use. Another study assessed patients who were evaluated for a discontinuation program 3 to 6 years earlier and found that 45 percent were not using benzodiazepines, although some were using other psychoactive medication. Of those who successfully completed the tapering program, 73 percent were benzodiazepine free at follow-up, compared with 39 percent of patients who entered but did not complete the program, and 14 percent of patients who did not participate in the program. Another study of patients treated for benzodiazepine dependence found that 56 percent of benzodiazepine-free (no use during the preceding month) patients had moderate-to-severe anxiety, and 38 percent had moderate-to-severe depression at 1 to 5 years follow-up. A different study found that even though 66 percent of patients were not taking benzodiazepines at follow-up, 75 percent had taken them at some time during the 5 years after entering a discontinuation program.

The follow-up studies suggest that most patients can successfully discontinue taking benzodiazepines. They also indicate that some patients experience recurrent anxiety that requires alternative psychoactive medication or reinstitution of benzodiazepine therapy. However, a subgroup of patients who discontinue benzodiazepines actually show improvement in anxiety over pretapering levels, which suggests that those who continue to have significant symptoms of anxiety or depression while taking a benzodiazepine should probably undergo a drug-free trial period. The value of a structured program for discontinuation and the substitution of medications with low abuse liability are supported in several studies. There appears to be no relation between long-term outcome (i.e., not using benzodiazepines at follow-up) and the dosage or duration of treatment or the specific benzodiazepine that was discontinued.

Barbiturates There are three common methods of establishing the barbiturate dosage required for safe withdrawal: (1) administration of a test dose of pentobarbital (Nembutal) to determine tolerance, (2) calculation of required doses based on estimated equivalencies, and (3) administration of phenobarbital loading doses.

In the first method the initial step is to determine tolerance. Once intoxication has subsided but before withdrawal symptoms have developed, an oral dose of 200 mg of pentobarbital is administered on an empty stomach and the effects are observed after 1 hour to determine the stabilization dose (Table 11.12-11). If no changes are observed, the test is repeated 3 hours later using 300 mg. If there is still no response, the total requirement may exceed 1600 mg a day. The calculated daily dose is divided and given every 4 to 6 hours for a 2- to 3-day stabilization period. Daily dose reductions are usually 10 percent of the stabilization dose, but withdrawal regimens must be individualized. Phenobarbital may be substituted for pentobarbital for stabilization and withdrawal at one third the dose.

Symptoms after Test Dose of 200 mg Oral Pentobarbital	Estimated 24-Hour Oral Pentobarbital Dose (mg)	Estimated 24-Hour Oral Phenobarbital Dose (mg)
Level I: Asleep but arousable; withdrawal symptoms not likely	0	0
Level II: Mild sedation; patient may have slurred speech, ataxia, nystagmus	500–600	150–200
Level III: Patient is comfortable; no evidence of sedation; may have nystagmus	800	250
Level IV: No drug effect	1000–1200	300–600

Modified from Cicaulo DA, Shader RI, editors: *Clinical Manual of Chemical Dependence*. American Psychiatric Press, Washington, DC, 1991. From data in Ewing JA, Rakerwell WT: Diagnosis and management of depressant drug dependence. Am J Psychiatry 125:1808, 1967.

Table 11.12-11 Pentobarbital Test Dose Procedure for Barbiturate Withdrawal

The second method of barbiturate detoxification requires calculating the equivalent hypnotic dose of phenobarbital on the basis of the dosage of barbiturate or other sedatives the patient reports taking, also taking into consideration any alcohol consumed. A dose of 30 mg of phenobarbital is substituted for an equivalent hypnotic dose of the sedative. The patient is stabilized for the period of peak vulnerability to withdrawal—2 to 3 days for sedative-hypnotic agents with short half-lives—and reductions of 30 to 60 mg of phenobarbital are made every 2 or 3 days. Clinicians using this method report that only rarely are doses of phenobarbital greater than 600 mg required. The procedure is not recommended, however, because published equivalencies are only approximations, and patients vary greatly in their metabolism and pharmacodynamic response to the same dose of the sedative. Furthermore, animal studies indicate that cross-tolerance among the various anxiolytic and sedative-hypnotic agents is not complete. Further, the method relies on the patient's history of drug use, which is often unreliable. Plasma concentrations can establish which sedatives have been ingested and may indicate the degree of tolerance when clinical correlations are made. Estimation of tolerance (and thus determination of dosage requirements) is best made after either a test dose or a loading dose is administered. Table 11.12-8 and Table 11.12-10 list approximate equivalencies of selected barbiturates, benzodiazepines, and sedatives.

The final method for medical withdrawal from barbiturates involves using loading doses of phenobarbital. According to the original protocol, 120-mg doses are administered every 1 to 2 hours until three of five signs are present—nystagmus, drowsiness, ataxia, dysarthria, emotional lability—or if patients are in withdrawal, until abstinence symptoms abate. No additional drug is administered, and because of its long half-life, the drug self-tapers. Patients are assessed before each dose.

Some clinicians have modified the protocol to use doses of 60 mg every 1 to 2 hours and to use a gradual taper rather than abrupt discontinuation. The originators of the protocol reported a mean loading dose of 1440 mg, with hourly doses sometimes required for 15 to 20 hours. They have also administered phenobarbital intravenously at 0.3 mg/kg per minute, with close medical supervision, for medically ill patients.

Overdose The treatment of overdose of the general class of substances involves gastric lavage, activated charcoal, and careful monitoring of vital signs and central nervous system activity. Overdose patients who come to attention while awake should be kept from slipping into unconsciousness. Vomiting should be induced, and activated charcoal should be administered to delay gastric absorption. If the patient is comatose, the clinician must establish an intravenous fluid line, monitor the patient's vital signs, insert an endotracheal tube to maintain a patent airway, and provide mechanical ventilation if necessary. Hospitalization of a comatose patient in an intensive care unit is usually required during the early stages of recovery from such overdoses.

SUGGESTED CROSS-REFERENCES

Substance-related disorders are discussed throughout Chapter 11 with an introduction and overview and relevant general tables in Section 11. Biological therapies are the focus of Chapter 31; benzodiazepines in Section 31.10 , monoamine oxidase inhibitors in Section 31.20 , and serotonin-specific reuptake inhibitors in Section 31.25.

SECTION REFERENCES

Barnes EM: Use-dependent regulation of $GABA_A$ receptors. Int Rev Neurobiol *39:*53, 1996.

Busto UE, Romach MK, Sellers EM: Multiple drug use and psychiatric comorbidity in patients admitted to the hospital with severe benzodiazepine dependence. J Clin Psychopharmacol *16:*51, 1996.

Ciraulo DA, Barnhill JG, Ciraulo AM, Sarid-Segal O, Knapp C, Greenblatt DJ, Shader RI: Alterations in pharmacodynamics of anxiolytic in abstinent alcoholic men: Subjective responses, abuse liability, and encephalographic effects of alprazolam, diazepam, and buspirone. J Clin Psychopharmacol *37:*64, 1997.

*Costa E, Guidotti A: Benzodiazepines on trial: A research strategy for their rehabilitation. Trends Pharmacol Sci *17:*192, 1996.

Ciraulo DA, Sands BF, Shader RI: Critical review of liability for benzodiazepine abuse among alcoholics. Am J Psychiatry *145:*1501, 1988.

*Ciraulo DA, Sarid-Segal O, Knapp C, Ciraulo AM, Greenblatt DJ, Shader RI: Liability to alprazolam abuse in daughters of alcoholics. Am J Psychiatry *153:*956, 1996.

Darke S: Benzodiazepine use among injecting drug users: Problems and implications. Addiction *89:*379, 1994.

De Wit H, Griffiths RR: Testing the abuse liability of anxiolytic and hypnotic drugs in humans. Drug Alcohol Depend *28:*83, 1991.

*Dunne DL: Long term use of sedative and hypnotic medication. Arch Gen Psychiatry *56:*355, 1999.

DuPont RL: Abuse of benzodiazepines: The problems and the solutions. Am J Drug Alcohol Abuse *14*(Suppl):1, 1988.

Elsesser K, Sartory G, Maurer J: The efficacy of complaints management training in facilitating benzodiazepine withdrawal. Behav Res Ther *34:*149, 1996.

Evans SM, Critchfield TS, Griffiths RR: Abuse liability assessment of anxiolytics/hypnotics: Rationale and laboratory lore. Br J Addict *86:*1625, 1991.

Evans SM, Funderburk FR, Griffiths RR: Zolpidem and triazolam in humans: Behavioral and subjective effects and abuse liability. J Pharmacol Exp Ther *255:*1246, 1990.

Evans SM, Griffiths RR, de Wit H: Preference for diazepam, but not buspirone, in moderate drinkers. Psychopharmacology *123:*154, 1996.

Farre M, Teran M-T, Cami J: A comparison of the acute behavioral effects of flunitrazepam and triazolam in healthy volunteers. Psychopharmacology *125:*1, 1996.

Gilson SF, Chilcoat HD, Stapleton JM: Illicit drug use by persons with disabilities: Insights from the National Household Survey on Drug Abuse. Am J Public Health *86:*1613, 1996.

Greenblatt DJ, Harmatz JS, Zinny MA, Shader RI: Effect of gradual withdrawal on the rebound sleep disorder after discontinuation of triazolam. N Engl J Med *317:*722, 1987.

Griffiths RR, Bigelow G, Liebson I: Human drug self-administration. Double-blind comparison of pentobarbital, diazepam, chlorpromazine, and placebo. J Pharmacol Exp Ther *210:*301, 1979.

Griffiths AN, Jones DM, Richens A: Zopiclone produces effects on human performance similar to flurazepam, lormetazepam and triazolam. Br J Clin Pharmacol *21:*647, 1986.

Griffiths RR, Lamb RJ, Ator NA, Roache JD, Brady JV: Relative abuse liability of triazolam: Experimental assessment in animals and humans. Neurosci Biobehav Rev *9:*133, 1985.

Griffiths RR, Lamb RJ, Sannerud CA, Ator NA, Brady JV: Self-injection of barbiturates, benzodiazepines and other sedative-anxiolytics in baboons. Psychopharmacology *103:*154, 1991.

Hayward P, Wardle J, Higgitt A, Gray J: Changes in “withdrawal symptoms” following discontinuation of low-dose diazepam. Psychopharmacology *125:*392, 1996.

Hobbs WR, Rall TW, Verdoorn TA: Hypnotics and sedatives: Ethanol. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, ed 9, AG Gilman, TW Rall, editors. Pergamon, New York, 1996.

Hutchinson MA, Smith PF, Darlington CL: The behavioral and neuronal aspects of the chronic administration of benzodiazepine anxiolytic and hypnotic drugs. Prog Neurobiol *49:*73, 1996.

Ito T, Suzuki T, Wellman SE, Ho IK: Pharmacology of barbiturate tolerance/dependence: $GABA_A$ receptors and molecular aspects. Life Sci *59:*169, 1996.

Kales A, Manfredi RL, Vgontzas AM, Bixler EO, Vela BA, Fee EC: Rebound insomnia after only brief and intermittent use of rapidly eliminated benzodiazepines. Clin Pharmacol Ther *49:*468, 1991.

Lillsunde P, Korte T, Michelson L, Portman M, Pikkarainen J, Seppala T: Drugs usage of drivers suspected of driving under the influence of alcohol and/or drugs. A study of one week's samples in 1979 and 1993 in Finland. Forensic Sci Int *77:*119, 1996.

Malcom R, Brady KT, Johnston AL, Cunningham M: Types of benzodiazepines abused by chemically dependent inpatients. J Psychoactive Drugs *25:*315, 1993.

Mumford GK, Evans SM, Fleishaker JC, Griffiths RR: Alprazolam absorption kinetics affects abuse liability. Clin Pharmacol Ther *57:*356, 1995.

Nutt DJ, Costello MJ: Rapid induction of lorazepam dependence and reversal with flumazenil. Life Sci *43:*1045, 1988.

Piercey MF, Hoffmann WE, Cooper M: The hypnotics triazolam and zolpidem have identical metabolic effects through the brain: Implications for benzodiazepine receptor subtypes. Brain Res *554:*244, 1991.

Rickels K, Case W, Schweizer E, Garcia-Espana F, Fridman R: Long-term benzodiazepine users 3 years after participation in a discontinuation program. Am J Psychiatry *148:*757, 1991.

Rickels K, Case WG, Schweizer E, Garcia-Espana F, Fridman R: Benzodiazepine dependence: Management of discontinuation. Psychopharmacol Bull *26:*63, 1990.

Rickels K, Case WG, Schweizer EE, Swenson C, Fridman RB: Low-dose dependence in chronic benzodiazepine users: A preliminary report of 119 patients. Psychopharmacol Bull *22:*407, 1986.

Rickels K, Fox IL, Greenblatt DJ, Sandler KR, Schless A: Clorazepate and lorazepam: Clinical improvement and rebound anxiety. Am J Psychiatry *145:*312, 1988.

Roache JD, Henningfield JE, Jaffe JH, Klein S, Sampson A: Reinforcing effects of triazolam in sedative abusers: Correlation of drug liking and self administration measures. Pharmacol Biochem Behav *50:*171, 1995.

Roy-Byrne PP, Dager SR, Cowley DS, Vitaliano P, Dunner DL: Relapse and rebound following discontinuation of benzodiazepine treatment of panic attacks: Alprazolam versus diazepam. Am J Psychiatry *146:*860, 1989.

Rush CR, Griffiths RR: Zolpidem, triazolam, and temazepam: Behavioral and subject-rated effects in normal volunteers. J Clin Psychopharmacol *16:*146, 1996.

*Salzman C: Addiction to benzodiazepines. Psychiatr Q *69:*251, 1998.

Sellers EM, Ciraulo DA, Dupont RL, Griffiths RL, Kosten TR, Romach MK, Woody GE: Alprazolam and benzodiazepine dependence. J Clin Psychiatry *54:*64, 1994.

*Simmons MM, Cupp MJ: Use and abuse of flunitrazepam. Ann Pharmacother *32:*117, 1998.

Spiegel DA, Bruce TJ, Gregg SF, Nuzzarello A: Does cognitive behavior therapy assist slow-taper alprazolam discontinuation in panic disorder? Am J Psychiatry *151:*876, 1994.

Steentoft A, Worm K, Pedersen CB, Sprehn M, Mogensen T, Sorensen MB, Nielsen E: Drugs in blood samples from unconscious drug addicts after the intake of an overdose. Int J Legal Med *108:*248, 1996.

Uhlenhuth EH, De Wit H, Balter MB, Johanson CE, Mellinger GD: Risks and benefits of long-term benzodiazepine use. J Clin Psychopharmacol *8:*161, 1988.

Woods JH, Katz JL, Winger G: Benzodiazepines: Use, abuse, and consequences. Pharmacol Rev *44:*151, 1992.

Woods JH, Wigner G: Current benzodiazepine issues. Psychopharmacology *118:*107, 1995.

Woods JH, Winger G: Liability of flunitrazepam. J Clin Psychopharmacol *17*(Suppl):1S, 1997.

Zawertailo LA, Busto U, Kaplan HL, Sellers EM: Comparative abuse liability of sertraline, alprazolam, and dextroamphetamine in humans. J Clin Psychopharmacol *15:*117, 1995.

11.13 ANABOLIC-ANDROGENIC STEROID ABUSE

HARRISON G. POPE, JR., M.D., AND KIRK J. BROWER, M.D.

The anabolic steroids are a family of drugs comprising the natural male hormone testosterone and a group of many synthetic analogues of testosterone synthesized since the 1940s (Table 11.13-1). All these drugs possess various degrees of *anabolic* (muscle-building) and *androgenic* (masculinizing) effects. Thus they should more correctly be called anabolic-androgenic steroids: however, the terms anabolic steroids and simply steroids are used for brevity. Steroids have a number of legitimate medical applications, such as in the treatment of hypogonadal men, and in the treatment of certain diseases, such as muscular dystrophy, various anemias, and the wasting syndrome associated with acquired immune deficiency syndrome (AIDS). However, steroids are widely used illicitly by individuals seeking to gain increased muscle mass and strength, either for athletic purposes or simply to improve personal appearance.

Compounds usually administered orally
Fluoxymesterone (Halotestin, Android-F, Ultandren)
Methandienone (formerly called methandrostenolone) (Dianabol)
Methyltestosterone (Android, Testred, Virilon)
Mibolerone (Cheque Drops*)
Oxandrolone (Anavar)
Oxymetholone (Anadrol, Hemogenin)
Mesterolone (Mesteranon, Proviron)
Stanozolol (Winstrol)
Compounds usually administered intramuscularly
Nandrolone decanoate (Deca-Durabolin)
Nandrolone phenpropionate (Durabolin)
Methenolone enanthate (Primobolan depot)
Boldenone undecylenate (Equipoise*)
Stanozolol (Winstrol-V*)
Testosterone esters blends (Sustanon, Sten)
Testosterone cypionate
Testosterone enanthate (Delatestryl)
Testosterone propionate (Testoviron, Andriolan)
Testosterone undecanoate (Andriol, Restandol)
Trenbolone acetate (Finaject, Finaplix*)
Trenbolone hexahydrobenzylcarbonate (Parabolan)

* Veterinary compound.
† Many of the brand names listed above are foreign, but are included because of the widespread illicit use of foreign steroid preparations in the United States.

Table 11.13-1 Examples of Commonly Used Anabolic Steroids†

HISTORY

Testosterone was first synthesized in 1935, and by the 1940s various synthetic derivatives of testosterone had appeared. Both the anabolic and androgenic actions of these drugs were quickly recognized; they were reportedly first used by Hitler's troops in World War II to increase aggressiveness, although this claim has been difficult to verify. Also in the late 1930s and early 1940s a number of psychiatric investigations suggested that testosterone possessed antidepressant properties. The hormone was believed to be especially effective for the “male climacteric” and was frequently prescribed in Europe and in the United States for depressed middle-aged men. With the advent of electroconvulsive therapy (ECT) and standard antidepressant medications, however, the use of anabolic steroids for psychiatric purposes rapidly waned.

By the 1950s athletes had recognized the value of anabolic steroids. The first reported use of these drugs in athletics was apparently by the Russians in 1954, followed within a few years by elite athletes from many other countries. Over the next three decades the use of steroids rapidly expanded among athletes at all levels, especially in sports requiring muscle mass and strength, such as bodybuilding, powerlifting, football, and field events such as the shot put. During the 1960s and 1970s many placebo-controlled studies tested whether anabolic steroids were truly effective for gaining muscle mass. The results of these studies were frequently equivocal, usually because the studies employed dosages and durations of steroid administration far below those used by actual athletes in the field. On the basis of these studies many reputable textbooks and reviews in pharmacology and endocrinology concluded that steroids were ineffective for athletic purposes. Athletes who had actually used these drugs and who were aware of the enormous muscle gains that they could produce tended to hold the medical profession in great contempt for its ignorance; consequently, these athletes often discounted medical warnings about the possible physical and psychiatric dangers of these agents.

By the late 1980s and early 1990s medical professionals increasingly acknowledged that anabolic steroids were indeed highly effective for achieving muscle gains and that use of these agents had approached epidemic proportions in the United States. Early reports also began to document psychiatric morbidity from steroids as well. These trends led to increasing concern about the public health consequences of anabolic steroid use, and in 1991 the United States Congress passed the Steroid Trafficking Act, adding anabolic steroids to the list of Schedule III controlled substances and placing jurisdiction over these substances in the hands of the Drug Enforcement Administration (DEA). Based on widespread anecdotal experience with anabolic steroids users it seems as if this legislative change has had a marked effect in that it has led to a diminished supply of genuine steroids on the black market and a greatly increased number of counterfeit steroids with no pharmacological effects. However, unlike many of the other drugs of abuse, anabolic steroids may be purchased legally without a prescription in many countries, and can be sent illicitly by mail to the United States (e.g., to post office boxes). This pattern makes interdiction difficult, and ensures that a substantial supply of genuine steroids will be available to illicit users in the United States for the foreseeable future.

COMPARATIVE NOSOLOGY

Psychiatric syndromes associated with use of anabolic steroids are cited in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV) under the general category of other substance-related disorders. The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) classifies steroid-related disorders under abuse of non-dependence-producing substances, along with agents such as antacids and laxatives. A rapidly growing body of literature, however, now supports a syndrome of anabolic steroid dependence and withdrawal, together with several anabolic steroid-induced psychiatric disorders. The best established among the latter are a syndrome of hypomanic or manic symptoms associated with steroid use and a depressive syndrome associated with steroid withdrawal. This chapter will parallel the standard terminology of DSM-IV for other drugs of abuse, discussing anabolic steroid dependence and abuse, anabolic steroid-induced mood disorder, anabolic steroid-induced psychotic disorder, and anabolic steroid-induced disorder not otherwise specified.

EPIDEMIOLOGY

Anabolic steroid use is widespread among men in the United States but is much less frequent in women. Among the most reliable data regarding the prevalence of steroid use are those from the National Household Survey on Drug Abuse (NHSDA). This survey, which has been conducted regularly since 1971, first included questions about anabolic steroids in 1991. Findings from this survey indicated that approximately 900,000 American males and approximately 150,000 American females reported having used anabolic steroids at some time during their lives. Approximately 250,000 males and 50,000 females were estimated to have used steroids within the past year. Among individuals who had used these drugs during the past year, nearly one quarter or 64,000 were between 12 and 17 years old. Various studies of high school students in the United States have produced even higher estimates of the prevalence of anabolic steroid use among adolescents. For example, one study of 3403 twelfth-grade boys in 46 public and private hospitals in the United States found that 6.6 percent of students reported current or past use of anabolic steroids, with two thirds of the users reporting that they had first tried these drugs when they were 16 years of age or less. Similarly high estimates have emerged from other studies of high school and college students in the United States; the prevalence of steroid use among male students has generally been many

times greater than that among female students in these studies. One possible criticism of these results, however, is that some students may have misunderstood the survey questions and claimed that they had used anabolic steroids when in fact they had used only corticosteroids, prescribed to them for conditions such as asthma or dermatological problems. Even allowing for this possible source of error, however, it is evident that rates of anabolic steroid use among American adolescents are high.

The current high rates of steroid use among younger individuals appears to represent an important shift in the epidemiology of steroid use. As recently as the mid-1970s use of these drugs was largely confined to competition bodybuilders, other elite weight-training athletes, and elite athletes in other sports. Currently, however, it appears that an increasing number of young men, and occasionally even young women, may be using these drugs purely to enhance personal appearance rather than for any athletic purpose.

Individuals with one form of substance abuse are often found to abuse other substances as well, and some studies document polysubstance use among steroid users. However, other anabolic steroid users report that they carefully abstain from using drugs such as alcohol and cocaine, because they are concerned with maximizing their muscularity or athletic ability. Anabolic steroid users, especially those involved in competition bodybuilding, may use a wide variety of other *ergogenic* (performance-enhancing) drugs to gain muscle, lose fat, or lose water for bodybuilding competitions. Other drugs commonly used by competitive bodybuilders are thyroid hormones, such as liothyronine (Cytomel) and lerothyroxine (Synthroid); human growth hormone and its daughter compound, somatomedin-C (insulin-like growth factor-1 [IGF-1]); insulin; dehydroepiandrosterone (DHEA); androstenedione; sympathomimetic agents such as amphetamines, ephedrine and pseudoephedrine (Sudafed); b-adrenergic receptor agonists such as clenbuterol; the amino acid derivative gamma hydroxybutyrate (GHB); various laxatives and diuretics; the opioid agonist-antagonist nalbuphine (Nubain), and others.

PHARMACOLOGY

Chemistry All steroid drugs—including anabolic-androgenic steroids, estrogens, and corticosteroids—are synthesized in vivo from cholesterol and resemble cholesterol in their chemical structure (Fig. 11.13-1). Testosterone has a four-ring chemical structure containing 19 carbon atoms (Fig. 11.13-2). Most synthetic derivatives of testosterone are produced by one or more of the following modifications: (1) forming an ester at the 17-b-hydroxyl group, (2) adding an alkyl group to the 17-a position, or (3) altering the structure at a different carbon site.

FIGURE 11.13-1 Physical effects of anabolic steroid use. These photographs compare a “natural” bodybuilder who has never used anabolic steroids (left), with a man who has used large doses of anabolic steroids over several years (right). Both men are 67 inches tall and have 7 percent body fat. The man on the left weighs 170 pounds and represents approximately the maximum degree of muscularity obtainable without drugs. His fat-free mass index is 25.4 kg/m $_2$ by the formula of Elena Kouri et al. (see references). The man on the right weighs 213 pounds and has a fat-free mass index of 31.7 kg/m $_2$. Note that the muscle hypertrophy from steroid use is particularly marked in the upper body in the pectoralis, deltoid, trapezius, and biceps muscles. Any man significantly more muscular than the man on the left has almost certainly abused anabolic steroids.

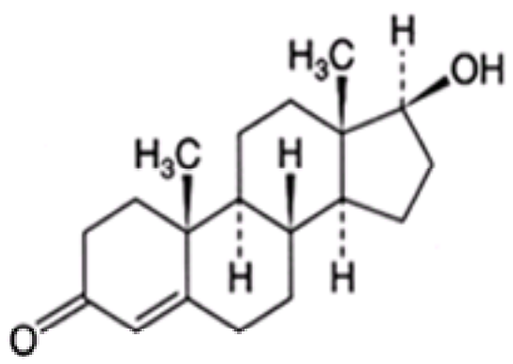

FIGURE 11.13-2 Molecular structure of testosterone.

Aqueous testosterone suspension is essentially inactive when ingested by the oral route because of first-pass metabolism in the liver, and is only briefly active when injected because it is rapidly absorbed and metabolized. Synthetic steroids overcome these problems. Testosterone esters (such as testosterone cypionate (Depo-Testosterone) and testosterone enanthate) confer the advantage of slower absorption than testosterone when injected, and their effects last from 2 to 4 weeks. Testosterone esters are hydrolyzed to testosterone after they are absorbed, so these agents can be monitored by blood concentrations. Testosterone is metabolized in part to dihydrotestosterone, which is up to 10 times more potent than testosterone as an androgen, and in part to the estrogen estradiol, which accounts for some of the adverse effects seen in men, such as gynecomastia. The effects of testosterone depend both on the balance of metabolic enzymes and on the concentration of androgen receptors in a particular tissue. The C-17-alkylated synthetic steroids make oral administration possible, but once-daily use is generally required.

Pharmacodynamics Steroids bind androgen receptors in the cytoplasm of cells. The androgen-receptor complex is then translocated to the cell nucleus where it augments gene transcription and ultimately new protein synthesis. The anabolic steroid effects desired by illicit users result especially from new protein synthesis in muscle tissue. The mechanism of action for psychoactive effects is poorly understood, but may include enzymatic conversion to estrogen in certain brain regions.

Normal testosterone plasma concentrations for men range from 300 to 1000 ng/dL. Generally, 200 mg of testosterone cypionate taken every 2 weeks will restore physiological testosterone concentrations in a hypogonadal male. A eugonadal male who initiates physiological dosages of testosterone will have no net gain in testosterone concentrations, because exogenously administered steroids shut down endogenous testosterone production via feedback inhibition of the hypothalamic-pituitary-gonadal axis. Consequently, illicit users take higher-than-therapeutic dosages in order to achieve supraphysiological effects. The dose-response curve for anabolic effects may be logarithmic, which could explain why illicit users generally take 10 to 100 times the therapeutic dosages.

Effects on Specific Organs and Systems The first effects of endogenous androgens are to induce male sexual differentiation in the fetus. Accumulating evidence suggests that fetal and perinatal androgens cause sexual dimorphism of the brain (at least in nonhuman mammals) as well as differentiation of the sexual organs. During puberty, testosterone levels increase in boys to produce male secondary sex characteristics. The phallus, prostate, and seminal vesicles all grow and develop, and testosterone and gonadotropins are involved in sperm production. Androgens cause the larynx to enlarge, the vocal cords to lengthen and thicken, and the voice to deepen. A male pattern of body hair develops, including hair on the face and chest. Sebaceous glands grow and increase their secretions, predisposing the adolescent to acne. Bones and muscles also grow during puberty as a result of androgen stimulation. Exogenous administration of anabolic-androgenic steroids to a prepubertal male will also induce all these effects.

In adults, physiological levels of testosterone continue to be important for spermatogenesis, although abnormally high levels will suppress spermatogenesis by

inhibiting release of luteinizing hormone, which is also required for spermatogenesis. Exogenous steroids increase muscle growth, promote mineralization of bones, and stimulate erythropoiesis by increasing erythropoietin levels and by direct action on the bone marrow.

Therapeutic Indications Anabolic steroids are primarily indicated for testosterone deficiency (male hypogonadism), hereditary angioedema (a congenital skin disorder), and some uncommon forms of anemia. In women they are given, although not as first-choice agents, for metastatic breast cancer, osteoporosis, endometriosis, and adjunctive treatment of menopausal symptoms. In men they have been used experimentally as a male contraceptive and for treating major depressive disorder and sexual disorders in eugonadal men. Recently, they have been used to treat wasting syndromes associated with AIDS.

Adverse Reactions The most common adverse medical effects of anabolic steroids involve the cardiovascular, hepatic, reproductive, and dermatological systems. In general, the more common adverse effects tend to be cosmetic and are reversible upon cessation of the steroid agent. Serious, life-threatening effects occur rarely but the long-term consequences of nonmedical steroid use have not been studied. A recent report revealed a shortened life span in mice exposed to high-dose anabolic steroids, but the generalizability to humans is unknown. It is established that the orally active C-17-alkylated steroids are generally more toxic to the liver and more likely to adversely affect cholesterol levels than the injectable testosterone esters.

Anabolic steroids produce an adverse cholesterol profile by increasing levels of low-density lipoprotein cholesterol and decreasing levels of high-density lipoprotein cholesterol. High-dose steroid use may also activate hemostasis. Steroids are widely reported to increase blood pressure, although most controlled studies fail to confirm this. Nevertheless, the adverse effects on cholesterol and hemostasis have led to concerns of increased coronary artery disease and atherosclerosis among steroid users. Isolated case reports of myocardial infarction, cardiomyopathy, left ventricular hypertrophy, and stroke among steroid users do exist, but there are no epidemiological studies that assess the risk to the cardiovascular and cerebrovascular systems from using high-dose, illicit steroids; fortunately, changes in cholesterol levels are reversible when steroids are discontinued.

Liver effects include cholestatic jaundice, benign and malignant liver tumors, and peliosis hepatis (blood-filled cysts that may rupture and cause death), although all these effects are rare. At least one case of prostatic cancer has been reported, and benign prostatic hypertrophy is also possible. Steroid-induced effects in males include testicular atrophy and sterility, both reversible after discontinuing steroids, and gynecomastia, which may persist until surgical removal. In females, shrinkage of breast tissue, irregular menses (diminution or cessation), and masculinization (clitoral hypertrophy, hirsutism, and deepened voice) can occur. Masculinizing effects in women may be irreversible, although this issue has not been well studied. Androgens taken during pregnancy could cause masculinization of a female fetus. Dermatological effects include acne and male pattern baldness. Anabolic steroid abuse by children has led to concerns that steroid-induced premature closure of bony epiphyses could cause shortened stature. Other uncommon adverse effects include edema of the extremities due to water retention, exacerbation of tic disorders, sleep apnea, and polycythemia.

Laboratory Findings The use of anabolic steroids can be associated with elevations in liver function tests such as bilirubin, lactate dehydrogenase (LDH), alkaline phosphatase, aspartate amino transferase (AST or SGOT), and alanine amino transferase (ALT). However, elevations in LDH, ALT, and AST, as well as the muscle enzyme creatine phosphokinase (CPK), are common in weight-training athletes, even in the absence of steroid use, as a result of muscle trauma from intensive weightlifting. The enzyme g-glutamyltransaminase (GGT), which is present exclusively in liver and not in muscle, may be more useful to distinguish frank liver toxicity from mere muscle trauma in steroid users.

An altered cholesterol profile—decreased levels of high-density lipoprotein cholesterol and increased levels of low-density lipoprotein cholesterol—is seen especially with oral administration of the C-17-alkylated anabolic steroids. Total cholesterol and triglycerides concentrations may also be elevated. Increases in the hematocrit and hemoglobin levels may occur relative to a patient's baseline values, although the hematocrit is rarely abnormally elevated unless steroids are combined with erythropoietin, as some endurance athletes have done.

Endocrine Tests Steroids inhibit the release of luteinizing hormone (LH) and follicular-stimulating hormone (FSH), resulting in lower serum levels. Serum testosterone and estradiol levels vary with the particular agent and timing of the sample. For example, the use of supraphysiological doses of testosterone esters will increase both testosterone and estradiol levels whereas the exclusive use of most other anabolic steroids will decrease these hormonal levels. Similarly, testosterone and estradiol levels are decreased during steroid withdrawal because of the lingering inhibition of the hypothalamic-pituitary-gonadal axis.

Other Tests Semen analysis may reveal decreased sperm count and motility as well as abnormal sperm morphology. Left ventricular hypertrophy (LVH) can be found on the electrocardiogram (ECG), although steroid non-users can have LVH on the basis on intensive strength training alone. An echocardiogram can rule out impaired diastolic function, which has been reported in some steroid users with LVH.

ETIOLOGY

The major reason for taking illicit steroids is to enhance either athletic performance or physical appearance. Steroid-taking is reinforced because steroids can produce the athletic and physical effects that users desire, especially when combined with proper diet and training. Further reinforcement derives from winning competitions and from social admiration for physical appearance. Steroid users also perceive that they can train more intensively for longer durations with less fatigue and with decreased recovery times between workouts.

The dramatic effects of anabolic steroids on muscle growth are illustrated in Figure 11.13-1, which compares a “natural” bodybuilder who has never used these drugs with a bodybuilder of identical height and body fat who has used steroids extensively.

While the anabolic or muscle-building properties of steroids are clearly important to those seeking to enhance athletic performance and physical appearance, psychoactive effects may also be important in the persistent and dependent use of anabolic steroids. Anecdotally, some steroid users report feelings of power, aggressiveness, and euphoria, which become associated with and can reinforce steroid-taking.

Cross-sectional studies have delineated a number of demographic and psychosocial differences between steroid users and nonusers. In general males are more likely to take steroids than females, and athletes are more likely to take steroids than non-athletes. Among weight lifters, one study found that both steroid users and individuals thinking about using steroids (high-risk group) were more likely than those not thinking about using steroids (low-risk group) to spend increased time each week lifting weights, to be training specifically for a bodybuilding competition, to use nonsteroidal substances for training enhancement (such as vitamins, amino acids, protein supplements, stimulants, and so-called natural testosterone releasers), and to know other steroid users. The study highlights the importance of training characteristics, other substance use, and peer influences as risk factors for initiating steroid use.

Interestingly, the high-risk nonuser group was more likely than either the low-risk group or the steroid users to “feel not big enough.” Indeed, recent investigations have suggested that some male and female weightlifters may suffer from muscle dysmorphia—a form of body dysmorphic disorder in which the individual feels that he is not sufficiently muscular and lean. Individuals with muscle dysmorphia may perceive themselves to be much smaller and weaker than they actually are, causing this disorder to formerly be called reverse anorexia nervosa. Such individuals appear to be particularly likely to experiment with anabolic steroids and other ergogenic drugs as compared to individuals who do not display comparable disorders of body image. By extension, the risk factors for dependent use may overlap with those for initial use. One study found higher steroid doses and more dissatisfaction with body image among dependent users than nondependent users, suggesting that both pharmacological factors and self-perceptions were risk factors for steroid dependence.

ANABOLIC STEROID DEPENDENCE AND ABUSE

The possibility of anabolic steroid dependence was mentioned in the scientific literature as early as 1980. Three case reports appeared in the medical literature between 1988 and 1990, which described young men who initially took steroids to enhance their weight-lifting and bodybuilding activity, and later used the drugs to combat depression when they tried to stop using them. The men reported feeling addicted and sought professional help to stop using illicitly obtained steroids because they were unable to accomplish discontinuation on their own.

Several research groups have utilized the revised third edition of DSM (DSM-III-R) criteria for substance dependence to diagnose steroid dependence in samples of weightlifters taking illicit steroids. One group made presumptive DSM-III-R diagnoses using an anonymous, self-administered questionnaire and found responses suggestive of steroid dependence in 57 percent of 49 men. Self-reported diagnostic symptoms and their frequencies included taking more steroids than intended (51 percent), difficulty cutting down or controlling use (16 percent), continuing to use despite adverse consequences (37 percent), tolerance (18 percent), and withdrawal (84 percent). In addition to a desire to take more steroids (52 percent), and headaches (20 percent), the most frequently reported withdrawal symptoms were

depressive in nature and included fatigue (43 percent), depressed mood (41 percent), restlessness (29 percent), anorexia (24 percent), insomnia (20 percent), and decreased libido (20 percent).

Another research group, using a face-to-face interview and the Structured Clinical Interview for DSM-III-R (SCID), found that 14.3 percent of 77 steroid users met criteria for steroid dependence. A third group reported that approximately 25 percent of 88 steroid users undergoing extensive clinical research interviews and assessment for psychiatric and medical effects appeared to manifest dependence on steroids. Finally, a study of 91 inpatient substance abuse treatment programs identified 68 steroid users of whom 69 percent were judged to meet DSM-III-R criteria for steroid dependence. The wide range of prevalence rates across the four studies (14 to 69 percent) probably reflects differences in sampling and differences in diagnostic methods, so the true prevalence of steroid dependence in steroid users and in the general population is unknown.

Unpublished observations suggest that women may also manifest steroid dependence, but female dependence on male hormones appears to be relatively rare when compared to men. Also, there appear to be no reported cases of anabolic steroid dependence in patients who were prescribed steroids for legitimate medical indications such as hypogonadism or anemia. Thus, physicians need not be concerned about dependence when legitimately prescribing steroids at therapeutic doses.

Etiology The mechanism of steroid dependence is not firmly established. For example, some authors have argued that steroid dependence represents a compulsive attachment to athletic activities—accompanied by feeling high when actively participating in those activities—rather than to any psychoactive properties of the steroids themselves. According to this view, steroids are taken more for their “muscle-active” than for their psychoactive properties. Indeed, steroids do not produce immediate sensations of euphoria following short-term administration, nor do animals self-administer steroids, as happens with drugs like cocaine and heroin. Although users do report euphoria following long-term administration (days to weeks) of steroids, such feelings could result from satisfying workouts or from the social rewards for having a big and muscular body.

Nevertheless, steroids do bind to specific receptors in the brain and alter neurotransmitter functioning, albeit in ways that are poorly elucidated to date. Steroids have also been used effectively to treat major depressive disorder and several controlled studies document the mood-altering properties of steroids. Finally, steroids mimic the electroencephalographic patterns produced by dopaminergic stimulants and they appear to sensitize the brain to the rewarding properties of amphetamine. One recent study demonstrated a steroid-induced increase in endogenous opioid concentrations in the ventral tegmental area, a portion of the brain postulated to be involved in the reinforcing action of other abused drugs such as cocaine. However, direct stimulation of brain reward pathways by anabolic steroids has not been proven.

In addition to positive reinforcement mechanisms, the negative reinforcement mechanisms of avoiding withdrawal symptoms, particularly depression, and avoiding the feeling of being too small may contribute to steroid dependence.

Course and Prognosis No longitudinal studies of dependent steroid users have been published to date. Based on two case reports it appears that steroid dependence can develop within 9 to 12 months after initiating use. In these cases users took supratherapeutic doses, and combined multiple steroid drugs, including both oral and injectable forms. However, dependence has also been described in a male weightlifter who used a single, oral agent at three times the therapeutic dose for nearly 9 years. Severity of dependence is also variable. In one study, 8.2 percent of 49 steroid users reported six or more DSM-III-R dependency symptoms, which is consistent with severe dependence.

Some authors have speculated that steroid withdrawal has two phases, with phase 1 lasting a week or less and consisting of opioid-like withdrawal symptoms. Phase 2 begins within the first week, consists of depressive symptoms, and may last for several months. Other than one case report, however, there is no published evidence of opioid-like withdrawal symptoms following the discontinuation of anabolic steroids.

Few steroid users seek substance abuse treatment. One study estimated that steroid users accounted for less than 0.1 percent of patients entering treatment programs for substance abuse. The low number of documented treatment cases stands in contrast to the large number of Americans estimated to use steroids illicitly (300,000 per year) and to be at risk for developing dependence (14 to 69 percent of users). However, it is not clear that most individuals fulfilling research criteria for steroid dependence require treatment in order to discontinue use.

Treatment Few descriptions and no controlled trials have been published of treatment for steroid-related disorders. Treatment is based on a comprehensive assessment including history, mental status and physical examinations, and laboratory tests. As in the case of other psychiatric disorders, treatment should address both the primary disorder and associated biological, psychological, and social factors. As in the case of other substance use disorders, treatment must motivate steroid users to discontinue use, ameliorate withdrawal-related symptoms, target associated psychiatric disorders, and address psychosocial triggers to relapse.

Abstinence is the treatment goal of choice for patients manifesting steroid abuse or dependence. To the extent that steroid users abuse other addictive substances (including alcohol), traditional treatment approaches for substance-related disorders may be utilized. Nevertheless, steroid users may differ from other addicted patients in several ways that have implications for treatment. First, the euphorigenic and reinforcing effects of steroids may only become apparent after weeks or months of use in conjunction with intensive exercising. When compared to immediately and passively reinforcing drugs such as cocaine, heroin, and alcohol, steroid use may entail more delayed gratification. Second, steroid users may manifest greater commitment to culturally endorsed values of physical fitness, success, victory, and goal directness than users of other illicit drugs. Finally, steroid users are often preoccupied with their physical attributes, and may rely excessively on these attributes for self-esteem. Treatment therefore depends on a therapeutic alliance that is based on a thorough and nonjudgmental understanding of the patient's values and motivations for using steroids.

Steroid Withdrawal Supportive therapy and monitoring are essential for treating steroid withdrawal because suicidal depressions can occur. Hospitalization may be required when suicidal ideation is severe. Patients should be educated about the possible course of withdrawal, and reassured that symptoms are time-limited and manageable. Antidepressant agents are best reserved for patients whose depressive symptomatology persists for several weeks after steroid discontinuation and who meet criteria for major depressive disorder. Selective serotonin reuptake inhibitors (SSRIs) are the preferred agents because of their favorable adverse effect profile and their effectiveness in the only reported case series of treated steroid users with major depressive disorder. Physical withdrawal symptoms are not life-threatening and do not ordinarily require pharmacotherapy. Nonsteroidal anti-inflammatory drugs (NSAIDs) may be useful to treat musculoskeletal pain and headaches.

Endocrine Pharmacotherapies Various endocrine pharmacotherapies for steroid withdrawal have been proposed on theoretical grounds, including testosterone substitution and taper, human chorionic gonadotropin, and estrogen blockers. Endocrine pharmacotherapies have as their goal the restoration of hypothalamic pituitary gonadal axis function. Hypothalamic pituitary gonadal activity can remain depressed for 12 or more weeks following cessation of prolonged, high-dose steroid use and resembles hypogonadotropic hypogonadism. However, there are no studies demonstrating the efficacy of endocrine pharmacotherapies for treating illicit steroid users during withdrawal, so their use remains experimental and cannot be routinely recommended.

ANABOLIC STEROID-INDUCED MOOD DISORDERS

Irritability, aggressiveness, hypomania, and frank mania associated with anabolic steroid use probably represent one of the most important public health issues associated with these drugs. Although athletes using these drugs have long recognized that syndromes of anger and irritability (sometimes called “roid rage”) could be associated with steroid use, these syndromes were little recognized in the scientific literature until the late 1980s and 1990s. Since then a series of observational field studies of athletes has suggested that some steroid users develop prominent hypomanic or even manic symptoms during steroid use. Of the 12 such studies that have appeared to date, 11 have found at least some such effects in association with anabolic steroid use, with an apparent association between the dosage of steroids used and the frequency of psychiatric effects. Specifically, in individuals using the equivalent of 300 mg of testosterone a week or less, psychiatric effects appear rare; in individuals taking intermediate dosages, between 300 and 1000 mg of testosterone equivalent per week, mood syndromes appear more common; and in individuals taking the equivalent of more than 1000 mg a week mood syndromes become quite common and are occasionally severe.

The association of steroid dose with psychiatric effects is illustrated by examining individual studies. At the low end of the dosage range, for example, one study of 53 men examined individuals who were using a mean dose of only 318 mg of testosterone or equivalent a week. The men in this study exhibited essentially no psychiatric changes during steroid use. By comparison, another study examined 71 men and 6 women who were using a mean dose estimated as somewhat more than 500 mg a week. The authors of this study noted that 1 (1.3 percent) of the subjects developed mania and 6 (7.8 percent) developed hypomania. One subject (1.3 percent) described major depressive disorder during steroid use, and 5 (6.5 percent) subjects reported major depressive disorder in association with steroid withdrawal. Three of the five users reported a suicide attempt during a period of withdrawal from steroid use. At the high end of the dosage range, a third study

examined 39 men and 2 women who had used a mean of 750 mg per week. In this study, five (12 percent) of the subjects reported a manic syndrome, and five (12 percent) reported psychotic symptoms at some time in association with steroid use. Interestingly, among the 8 subjects reporting manic or psychotic symptoms or both in this study, the average weekly dose was about 900 mg. A further illustration of the relationship between steroid dose and the prevalence of psychiatric syndromes is provided in Table 11.13-2, which provides the frequency of hypomanic, manic, and major depressive syndromes reported among 88 steroid users in one study where the users are grouped on the basis of the maximum dosage of steroids used. Psychiatric syndromes are almost absent at the lowest dosages, but affected nearly half of the individuals using dosages equivalent to 1000 mg or more of testosterone a week.

Weekly Steroid Dose[a] (No. of Users)	No. (%) of Subjects Displaying Mood Disorders[a]			
	Manic Episode	Hypomanic Episode	Major Depressive Disorder	Total[b]
Low (N = 12)	0	1 (8)	0	1 (8)
Medium (N = 51)	0	5 (10)	3 (6)	7 (14)
High (N = 25)	4 (16)	3 (12)	7 (28)	11 (44)

[a] Significance of relationship between increasing doses of steroids and prevalence of mood disorders by two-sided trend test: manic episode, P = .01, major depressive disorder, P = .003, total, P = .004.
[b] Total is less than sum of individual categories because some individuals displayed both a manic or hypomanic episode and a major depressive episode in association with steroids.
[c] Low represents less than 300 mg/wk of testosterone or equivalent; medium, 300 to 1000 mg/wk; and high, more than 1000 mg/wk.
Reprinted with permission from Pope HG Jr, Katz DL: Psychiatric and medical effects of anabolic-androgenic steroid use: A controlled study of 160 athletes. Arch Gen Psychiatry 51:375, 1994.

Table 11.13-2 Relationship of Weekly Steroid Dosage to Prevalence of Mood Disorders

Observational field studies, however, are subject to a number of methodological limitations common to retrospective studies of this nature. First, subjects are being asked to describe psychiatric symptoms that may have occurred well in the past and that therefore may be colored by recall bias. Second, these subjects were generally taking drugs obtained on the black market that were of unknown potency or authenticity. Thus, assumptions regarding the nature and dosage of the drugs are necessarily crude at best. Third, selection bias may influence which subjects present for interviews in observational studies: subjects who have experienced prominent psychopathology in association with steroid use might be either more likely or less likely to participate. Fourth, expectational effects and influences from the subculture of the gymnasium may greatly influence subjects' reactions to steroids. For example, an individual lifting weights regularly in a “hard core” gymnasium might be primed to expect increases in his irritability and aggression with steroid use, even if such changes do not actually occur.

Laboratory studies using anabolic steroids in normal volunteers avoid most of these limitations, but are subject to problems of their own. Specifically, most laboratory studies, including endocrinological investigations, physiological studies, and studies using anabolic steroids to treat various medical conditions, have used very modest doses of testosterone or other anabolic steroids, rarely exceeding the equivalent of 300 mg of testosterone a week. Furthermore, such studies have rarely attempted to measure psychiatric effects systematically. Not surprisingly, these studies have generally failed to note prominent psychiatric changes in their subjects. As a result, some such studies have erroneously concluded that anabolic steroids lack psychiatric effects. This error is similar to that of many earlier negative studies of the effects of steroids on muscle growth, again because the studies typically used doses far lower than those actually used by athletes in the field.

Recently, four newer laboratory studies have appeared in which normal male volunteers were given dosages of 500 mg a week or more of testosterone or equivalent under placebo-controlled, double-blind conditions. These studies represent the first available controlled investigations that have used doses of steroids more closely approximating those used by actual steroid abusers. Of 109 men receiving such dosages in these four studies combined, 5 exhibited prominent hypomanic or manic syndromes during steroid exposure. These subjects displayed levels of symptomatology that might be expected to seriously impair normal social or occupational functioning; one subject study even requested that he be placed in seclusion on the ward where the study was conducted. By contrast, however, the great majority of the men in these four studies exhibited very few behavioral changes with high dosages of anabolic steroids. Thus, it appears that most individuals taking anabolic steroids, even at fairly high dosages, will experience few psychiatric effects although an occasional individual may develop a severe manic or hypomanic reaction. At present the mechanism of these reactions is unknown; they may represent an idiosyncratic phenomenon unrelated to premorbid psychiatric function.

Figure 11.13-3 shows changes in ratings of manic symptoms in three subjects who received testosterone cypionate in dosages up to 600 mg a week versus inert placebo, each for periods of 6 weeks in a double-blind, crossover design. Subjects varied widely in their response to testosterone. The subject on the left side of the figure exhibited virtually no change in manic ratings during either the testosterone or placebo periods; the subject in the middle displayed a modest rise in manic symptoms in association with testosterone, but not sufficient to significantly affect his social or occupational functioning. By contrast, the subject on the right displayed a marked manic reaction, approaching the range of patients hospitalized for treatment of manic episodes, requiring the investigators to discontinue testosterone administration 1 week earlier than planned. The symptoms of mania resolved promptly within 2 weeks following discontinuation of testosterone.

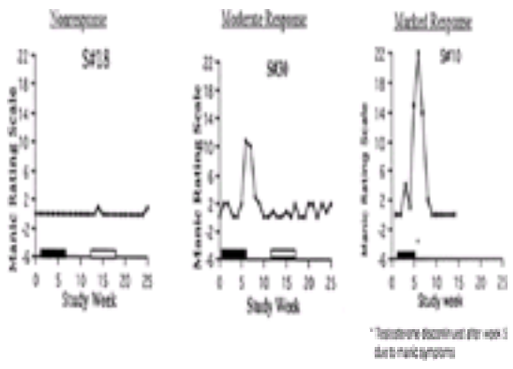

FIGURE 11.13-3 Examples of marked response, moderate response, and nonresponse to testosterone versus placebo on the Young Manic Rating Scale. The black bar represents the period of testosterone cypionate administration in each subject. Testosterone was administered in gradually increasing doses (150 mg/wk for the first 2 weeks, 300 mg/wk for the next 2 weeks, and 600 mg/wk for the final 2 weeks of the period). The white bar represents the period of placebo administration. A 6-week washout period was interposed between the two treatment periods to minimize carryover effects.

It is difficult to extrapolate from these findings to estimate the frequency of prominent hypomanic or manic reactions among steroid users in the field. Illicit users may employ higher dosages than could ethically be administered in the laboratory; such users may take several different steroids together (“stacking”); they may use alcohol or other drugs while taking steroids; and of course they do not screen themselves with the same care as would be the case in a laboratory study. Finally, some users may become more sensitized to the psychiatric effects of steroids after using them for multiple courses over time; since most of the subjects in available laboratory studies were steroid-naive prior to treatment they may not be representative of actual steroid users in the field. For all these reasons the crude rate of approximately 5 percent previously estimated may represent a lower bound for the true prevalence of steroid-induced hypomanic and manic syndromes in the overall population of users.

A possible serious consequence of steroid-induced mood disorders may be violent or even homicidal behavior. Several published reports have anecdotally described individuals with no apparent history of psychiatric disorder, no criminal record, and no history of violence who committed violent crimes, including murder, while under the influence of anabolic steroids. In a number of cases steroid use has been cited in criminal trials as a possible mitigating factor in the defense of such individuals. Although a causal link is difficult to establish in these cases, evidence of steroid use has frequently been presented in forensic settings as a possible mitigating factor in criminal behavior. One recent study of 133 sequential, newly incarcerated male felons found that 3 (2.3 percent) reported committing violent crimes while under the influence of anabolic steroids.

At least one published study, together with anecdotal experience, suggests that violence associated with anabolic steroids is frequently directed at women. In one

study that specifically examined this issue 23 male steroid users were compared with 14 male non-users, using the Conflict Tactics Scales to assess users' relationships with their girlfriends or spouses. Significant differences emerged on several of these scales between users on drugs and nonusers and between users on drugs and the same users when off drugs.

Anabolic steroid-induced depressive syndromes have generally not been reported in laboratory studies, but have been documented in field studies. Case reports of completed suicides associated with steroid withdrawal have also appeared. In some instances it appears that a brief and self-limited syndrome of depression occurs upon steroid withdrawal, probably as a result of the depression of the hypothalamic-pituitary-gonadal axis following exogenous steroid administration. Such syndromes might be expected to occur more often in the field than in laboratory settings because athletes typically take steroids for longer intervals than would occur in the laboratory, resulting in more pronounced neuroendocrine depression and shrinkage of the testes.

Not all depressive syndromes of this nature, however, are self-limited. Some reports have described episodes of depression that continued for months after discontinuing steroids, and which required treatment with fluoxetine (Prozac), tricyclic antidepressant agents, or electroconvulsive therapy (ECT). Again, however, given the uncontrolled and retrospective nature of these observations, it is difficult to judge the true frequency of such syndromes in the field.

ANABOLIC STEROID-INDUCED PSYCHOTIC DISORDER

Psychotic symptoms are rare in association with anabolic steroid use but have been described in a few cases, primarily in individuals who were using the equivalent of more than 1000 mg of testosterone a week. Usually these symptoms have consisted of grandiose or paranoid delusions, generally occurring in the context of a manic episode, although occasionally occurring in the absence of a frank manic syndrome. In most cases reported psychotic symptoms have disappeared promptly (within a few weeks) after the discontinuation of the offending agent, although temporary treatment with antipsychotic agents was sometimes required. As with manic reactions to steroids, the mechanism for these seemingly idiosyncratic psychotic reactions is unknown. It is interesting to note, however, that the chemically related family of hormones, corticosteroids, also produce idiosyncratic manic and psychotic symptoms in occasional individuals while creating few psychiatric effects in the great majority of patients. The mechanism of action of these idiosyncratic effects with corticosteroids is also unknown.

ANABOLIC STEROID-RELATED DISORDER NOT OTHERWISE SPECIFIED

Occasional anecdotal reports have described decreases in symptoms of anxiety disorders, such as panic disorder and social phobia, during steroid use, and exacerbations of these disorders during steroid withdrawal. Conversely, some individuals report increases in anxiety symptoms while taking steroids. However, no systematic data are available on these entities.

Anabolic steroid use may frequently be associated with use of other licit and illicit ergogenic drugs. Thus, athletes may exhibit both physical and psychiatric morbidity associated with the use of these other drugs, and also may develop apparent abuse and dependence of multiple ergogenic drugs, although this syndrome has been poorly characterized in the literature to date. For example, use of ephedrine and other central nervous system stimulants to promote fat loss may lead to insomnia, anxiety, and other features of stimulant abuse and dependence syndromes. Use of large doses of adrenal hormones, such as DHEA and androstenedione, may lead to masculinizing effects in female athletes. DHEA may also possess antidepressant properties, although data in this area remain limited. Use of excessive dosages of thyroid hormones may produce psychiatric effects similar to those seen in thyrotoxicosis. Use of the opioid agonist/antagonist nalbuphine, a drug increasingly popular among anabolic steroid users, may lead to features of opioid dependence, and may even require detoxification.

Mr. A. is a 26-year-old, single white man. He is 69 inches tall and presently weighs 204 pounds with a body fat of 11 percent. He reports that he began lifting weights at age 17, at which time he weighed 155 pounds. Within a year of beginning his weightlifting he began taking anabolic steroids after obtaining them through a friend at his gymnasium. His first cycle of steroids, lasting for 9 weeks, involved methandienone, 30 mg orally a day, and testosterone cypionate, 600 mg intramuscularly a week. During these 9 weeks he gained 20 pounds of muscle mass. He was so pleased with these results that he took five further cycles of anabolic steroids over the course of the next 6 years. During his most ambitious cycle, approximately 1 year ago, he used testosterone cypionate, 600 mg per week; nandrolone decanoate (Deca-Durabolin), 400 mg a week; stanozolol (Winstrol), 12 mg a day; and oxandrolone (Anavar), 10 mg a day.

During each of the cycles Mr. A. has noted euphoria, irritability, and grandiose feelings. These symptoms were most prominent during his most recent cycle, when he felt invincible. During this cycle he also noted a decreased need for sleep, racing thoughts, and a tendency to spend excessive amounts of money. For example, he impulsively purchased a $2700 stereo system when he realistically could not afford to spend more than $500. He also became uncharacteristically irritable with his girlfriend, and on one occasion put his fist through the side window of her car during an argument—an act inconsistent with his normally mild-mannered personality. After this cycle of steroids ended, he became moderately depressed for about 2 months, with hypersomnia, anorexia, markedly decreased libido, and occasional suicidal ideation.

Mr. A. smoked marijuana almost daily during his last 2 years of high school and continues to smoke at least twice a week. He has experimented briefly with hallucinogens, cocaine, opiates, and stimulants, but has rarely used them in the last 5 years. However, he has used a number of drugs to lose weight in preparation for bodybuilding contests. These include ephedrine, amphetamine, triiodothyronine, and thyroxin. Recently, he has also begun to use the opioid agonist-antagonist nalbuphine (Nubain) intravenously to treat muscle aches from weight lifting. He reports that intravenous nalbuphine use is widespread among other anabolic steroid users of his acquaintance.

Mr. A. exhibits characteristic features of muscle dysmorphia. He checks his appearance dozens of times a day in mirrors, or when he sees his reflection in a store window or even in the back of a spoon. He becomes anxious if he misses even one day of working out at the gym, and acknowledges that his preoccupation with weightlifting has cost him both social and occupational opportunities. Although he has a 48-inch chest and 19-inch biceps, he has frequently declined invitations to go to the beach or a swimming pool for fear that he would look too small when seen in a bathing suit. He is anxious because he has lost some weight since the end of his previous cycle of steroids and is eager to resume another cycle of anabolic steroids in the near future.

SUGGESTED CROSS-REFERENCES

An overview of substance-related disorders is given in Chapter 11; mood disorders are discussed in Chapter 14 and psychotic disorders in Chapter 13. Body dysmorphic disorder is covered in Chapter 16.

SECTION REFERENCES

Bahrke MS, Wright JE, Strauss RH, Catlin DH: Psychological moods and subjectively perceived behavioral and somatic changes accompanying anabolic-androgenic steroid use. Am J Sports Med *20:*717, 1992.

Bidwill MJ, Katz DL: Injecting new life into an old defense: Anabolic steroid-induced psychosis as a paradigm of involuntary intoxication. Univ Miami Entertainment Sports Law Rev *7:*1, 1989.

*Bhasin S, Storer TW, Berman N, Callegari C, Clevenger B, Phillips J, Bunnell TJ, Tricker R, Shirazi A, Casaburi R: The effect of supraphysiologic doses of testosterone on muscle size and strength in normal men. N Engl J Med *335:*1, 1996.

Bronson FH, Matherne CM: Exposure to anabolic-androgenic steroids shortens life span of male mice. Med Sci Sports Exerc *29:*615, 1997.

*Brower KJ: Withdrawal from anabolic steroids. Curr Ther Endocrinol Metab *6:*338, 1997.

Brower KJ: Anabolic steroids: Addictive, psychiatric, and medical consequences. Am J Addict *1:*100, 1992.

Brower KJ: Clinical assessment and treatment of anabolic steroid users. Psychiatr Ann *22:*35, 1992.

Brower KJ, Blow FC, Young JP, Hill EM: Symptoms and correlates of anabolic-androgenic steroid dependence. Br J Addict *86:*759, 1991.

Brower KJ, Blow FC, Hill EM: Risk factors for anabolic androgenic steroid use in men. J Psychiatr Res *28:*369, 1994.

Buckley WA, Yesalis CE, Friedl KE, Anderson W, Streit A, Wright J: Estimated prevalence of anabolic steroid use among male high school seniors. JAMA *260:*3441, 1988.

Catlin DH: Anabolic steroids. In *Endocrinology*, ed 3., LJ DeGroot, editor. WB Saunders, Orlando, FL, 1995.

Catlin DH: Effects and complications of anabolic androgenic steroids. In *Handbook on Drug Abuse*, SB Karch, editor. CRC Press, Boca Raton, FL, 1998.

Clancy GP, Yates WR: Anabolic steroid use among substance abusers in treatment. J Clin Psychiatry *53:*97, 1992.

Clark AS, Lindenfeld RC, Gibbons CH: Anabolic-androgenic steroids and brain reward. Pharmacol Biochem Behav *53:*741, 1996.

Evans NA: Gym and tonic: A profile of 100 male steroid users. Br J Sports Med *31:*54, 1997.

Forbes GB, Porta CR, Herr BE, Griggs RC: Sequence of changes in body composition induced by testosterone and reversal of changes after drug is stopped. JAMA *267:*397, 1992.

Gruber AJ, Pope HG Jr: Psychiatric and medical effects of anabolic-androgenic steroid use in women. Psychother Psychosom, in press.

Haupt HA, Rovere GD: Anabolic steroids: A review of the literature. Am J Sports Med *12:*469, 1984.

Isacsson G, Garle M, Ljung EB, Asgard U, Bergman U: Anabolic steroids and violent crime—An epidemiological study at a jail in Stockholm, Sweden. Compr Psychiatry *39:*203, 1998.

Johansson P, Ray A, Zhou Q, Huang W, Karlsson K, Nyberg F: Anabolic androgenic steroids increase beta-endorphin levels in the ventral tegmental area in the male rat brain. Neurosci Res *27:*185, 1997.

Kashkin KB, Kleber HD: Hooked on hormones: An anabolic steroid addiction hypothesis. JAMA *262:*3166, 1989.

Kouri EM, Pope HG Jr, Katz DL, Oliva PS: Fat-free mass index in users and non-users of anabolic-androgenic steroids. Clin J Sport Med *5:*223, 1995.

Lin GC, Erinoff L, editors: *Anabolic steroid abuse* (DHHS Publication No. ADM 91-1720). U.S. Government Printing Office, Washington, DC, 1990.

Madea B, Grellner W: Long-term cardiovascular effects of anabolic steroids. Lancet *352:*33, 1998.

Malone DA Jr: Pharmacological therapies of anabolic androgenic steroid addiction. In *Pharmacological Therapies for Drug and Alcohol Addictions*, NS Miller, MS Gold, editors. Marcel Dekker, New York, 1995.

Malone DA Jr, Dimeff RJ: The use of fluoxetine in depression associated with anabolic steroid withdrawal: A case series. J Clin Psychiatry *53:*130, 1992.

Malone DA Jr, Dimeff R, Lombardo JA, Sample BRH: Psychiatric effects and psychoactive substance use in anabolic-androgenic steroid users. Clin J Sports Med *5:*25, 1995.

Pope HG Jr, Katz DL: Homicide and near-homicide by anabolic steroid users. J Clin Psychiatry *51:*28, 1990.

Pope HG Jr, Kouri EM, Powell KF, Campbell C, Katz DL: Anabolic-androgenic steroid use among 133 prisoners. Compr Psychiatry *37:*322, 1996.

*Pope HG Jr, Katz DL: Psychiatric and medical effects of anabolic-androgenic steroid use. Arch Gen Psychiatry *51:*375, 1994.

Pope HG Jr, Gruber AJ, Choi P, Olivardia R, Phillips KA: “Muscle dysmorphia": An underrecognized form of body dysmorphic disorder. Psychosomatics *38:*548, 1997.

*Schumacher J, Muller G, Klotz KF: Large hepatic hematoma and intraabdominal hemorrhage associated with abuse of anabolic steroids. N Engl J Med *340:*1123, 1999.

*Su TT, Pagliaro M, Schmidt PJ, Pickar D, Wolkowitz OM, Rubinow DR: Neuropsychiatric effects of anabolic steroids in male normal volunteers. JAMA *269:*2760, 1993.

Tricker R, Casaburi R, Storer TW, Clevenger B, Berman N, Shirazi A, Bhasin S: The effect of supraphysiological doses of testosterone on angry behavior in healthy eugonadal men. J Clin Endocrinol Metab *181:*3754, 1996.

Wilson JD: Androgen abuse by athletes. Endocr Rev *9:*181, 1988.

Wroblewska AM: Anabolic-androgenic steroids and body dysmorphia in young men. J Psychosom Res *42:*225, 1997.

Yesalis CE, editor: *Anabolic Steroids in Sport and Exercise*. Human Kinetics, Champaign, IL, 1993.

Yesalis CE, Cowart VS: *The Steroids Game*. Human Kinetics, Champaign, IL, 1998.

*Yesalis CE, Kennedy NJ, Kopstein AN, Bahrke MS: Anabolic-androgenic steroid use in the United States. JAMA *270:*1217, 1993.

CHAPTER 12. SCHIZOPHRENIA

12.1 SCHIZOPHRENIA: INTRODUCTION AND OVERVIEW

ROBERT W. BUCHANAN, M.D. AND WILLIAM T. CARPENTER, JR., M.D.

Schizophrenia is the paradigmatic illness of psychiatry. It is a clinical syndrome of variable but profoundly disruptive psychopathology, which involves thought, perception, emotion, movement, and behavior. The expression of these symptoms varies across patients and over time, but the cumulative effect of the illness is always severe and usually long lasting.

HISTORY

Written descriptions of symptoms commonly observed today in patients with schizophrenia are found throughout recorded history. Early Greek physicians described delusions of grandeur, paranoia, and deterioration in cognitive functions and personality. These behaviors were generally considered to merit social sanction. However, since these symptoms are not necessarily unique to schizophrenia, one cannot be certain whether these behaviors were actually associated with what today would be called schizophrenia. Indeed, several scholars have argued that schizophrenia is of relatively recent origin.

Schizophrenia did not emerge as a medical condition worthy of study and treatment until the eighteenth century. By the nineteenth century the various psychotic disorders were generally viewed as insanity or madness, and the movement to conceptualize the condition as a regrettable affliction replaced the view of insanity as a reprehensible behavior. During the middle to late nineteenth century many clinical categories were described, but a general approach capable of integrating the widely diverse manifestations of mental illness into distinguishable clinical syndromes was lacking.

A major impediment to distinguishing schizophrenia from other forms of psychoses was the existence of another common type of insanity, *general paresis*. The symptom manifestations of general paresis were quite diverse and overlapped extensively with those of schizophrenia. The cause of syphilitic insanity was subsequently traced to a spirochetal infestation, and antibiotics were eventually found to be effective in treatment and prevention. The identification of syphilitic insanity enabled Emil Kraepelin to delineate the two other major patterns of insanity: *manic-depressive psychosis* and *dementia praecox* (or dementia of the young), and to group together under the diagnostic category of dementia precox the previously disparate categories of insanity, such as hebephrenia, paranoia, and catatonia. In differentiating dementia precox from manic-depressive disorder, Kraepelin emphasized what he believed to be the characteristic poor long-term prognosis of dementia precox, as compared to the relatively nondeteriorating course of manic-depressive illness. In *Dementia Praecox and Pathophysiology* (1919) Kraepelin went on to describe what he believed to be the two principal pathophysiological or disease processes occurring in dementia precox:

> On the one hand we observe a weakening of those emotional activities which permanently form the mainsprings of volition. In connection with this, mental activity and instinct for occupation become mute. The result of this part of the process is emotional dullness, failure of mental activities, loss of mastery over volition, of endeavor, and of ability for independent action. The essence of personality is thereby destroyed, the best and most precious part of its being, as Griesinger once expressed it, torn from her . . . The second group of disorders, which gives dementia praecox its peculiar stamp . . . consists in the loss of the inner unity of the activities of intellect, emotion, and volition in themselves and among one another. Stransky speaks of an annihilation of the "intrapsychic co-ordination" . . . [T]his annihilation presents itself to us in the disorders of association described by Bleuler, in incoherence of the train of thought, in the sharp change of moods as well as in desultoriness and derailments in practical work. But further, the near connections between thinking and feeling, between deliberation and emotional activity on the one hand, and practical work on the other is more or less lost. Emotions do not correspond to ideas.

The description of the former process provides the conceptual framework for the avolitional or negative symptom component of the illness, and the description of the latter process provides the conceptual framework for the positive symptoms of schizophrenia.

In 1911 Eugen Bleuler, recognizing that dementia was not a usual characteristic of dementia precox, suggested the term *schizophrenia* (splitting of the mind) for the disorder. Bleuler introduced the concept of primary and secondary schizophrenic symptoms; his four primary symptoms (the four As) were abnormal associations, autistic behavior and thinking, abnormal affect, and ambivalence. Of these four symptoms Bleuler viewed as central to the illness the loss of association between thought processes and among thought, emotion, and behavior. Typical examples of these losses of associations are silly giggling on receiving news of the death of a loved one, the introduction of magical thinking and peculiar concepts into an ordinary discussion, and the sudden display of angry behavior without experiencing anger (or an understandable provocation).

Bleuler's view that a dissociative process is fundamental to schizophrenia and that this process underlies a wide variety of the symptom manifestations of schizophrenia has supported a major paradigm for conceptualizing the illness, namely, that in spite of its various manifestations, schizophrenia is a single disease entity in which there is extensive similarity in cause (etiology) and mechanism (pathophysiology) across all patients with the disorder. In this view, a neurophysiological disturbance of indeterminate origin and nature occurs that is manifest as dissociative processes adversely influencing the development of mental capacities in the areas of thought, emotion, and behavior. Depending on the individual's adaptive capacity and environmental circumstances, this fundamental process could lead to secondary disease manifestations such as hallucinations, delusions, social withdrawal, and diminished drive.

There are many parallels in medicine for this single-disease model. Diabetic patients share an impairment in glucose metabolism, but the secondary manifestations vary considerably depending on which organ systems are involved. Similarly, seizure disorders may share a common pathophysiological mechanism, but different lesion locations lead to marked variability in signs and symptoms. One patient may have full-body convulsions, while another may experience strange sexual sensations and excessive religiosity. The diverse manifestations of syphilitic insanity best illustrate the utility of this disease-entity approach for schizophrenia.

The major alternative etiopathophysiological model conceptualizes schizophrenia as a clinical syndrome rather than a single disease entity. This view holds that although patients with schizophrenia share a sufficient commonality of signs and symptoms to validly differentiate them from patients with other forms of psychosis (e.g., mood disorder with psychotic features, substance-induced psychotic disorder), more than one disease entity will eventually be found within this syndrome. This view is supported by the existence of multiple risk factors and heterogeneity in clinical presentation, treatment response, and clinical course. The emergence over the past 50 years of proof that mental retardation is a clinical syndrome comprised of multiple disease entities rather than a single disease entity best illustrates this construct. Schizophrenia currently maintains the status of a clinical syndrome in the absence of evidence for the existence of a single disease entity.

There are other competing models for conceptualizing schizophrenia, which, although seriously debated in the past, are presently dismissed as demonstrably invalid or so seriously reductionistic as to not account for major observations associated with the illness. Nondisease models, such as the societal reaction theory ("a sane reaction to an insane world") or Thomas Szasz's theory that schizophrenia is a myth enabling society to manage deviant behavior, cannot adequately account for the distribution of schizophrenia among biological relatives, the myriad of associated brain abnormalities, the normalizing effects of drug treatment, and the extensive similarity and lifetime prevalence and clinical manifestations of schizophrenia across widely divergent cultures. Narrow framework disease models that attempt to account for the illness solely at the level of psychological mechanisms are also demonstrably inadequate in accommodating the known facts of the illness. Genetic or immunovirological causal factors cannot be addressed by reductionistic theories operating at the psychological or social levels. The many biological, psychological, and social factors relevant to the understanding and treatment of the person with schizophrenia require a broad medical model and eschew reduction to any single level of the functioning organism.

In summary, schizophrenia is appropriately and accurately conceptualized as a disease process. Although it is possible that a unifying etiology or pathophysiology will eventually be uncovered that will account for all or almost all cases, it seems more likely that more than one disease entity exists within the clinical syndrome of schizophrenia, with each having a distinguishable etiology and pathophysiology. Any reductionistic approach to the description or explanation of the disorder cannot adequately account for the range of relevant information and facts. A broad medical model that integrates factors ranging from the molecular to the psychosocial level of organization is necessary to describe schizophrenia, to account for the range of pathogenic influences, and to provide for treatment and rehabilitation.

EPIDEMIOLOGY

Schizophrenia is a leading public health problem that exacts enormous personal and economic costs worldwide. Schizophrenia affects just under 1 percent of the world's population (approximately 0.85 percent). The number of affected individuals increases if schizophrenia spectrum disorders are included in prevalence estimates. The concept of schizophrenia spectrum disorders is derived from observations of psychopathological manifestations in the biological relatives of patients with schizophrenia. Diagnoses and approximate lifetime prevalence rates (percentage of population) for spectrum disorders are: schizoid personality disorder (fractional), schizotypal personality disorder (1 to 4 percent), schizoaffective psychosis (0.7 percent), and atypical psychoses and delusional disorder (0.7 percent). The relation of these disorders to schizophrenia in the general population is unclear, but in family pedigree studies the presence of a proband with schizophrenia significantly increases the prevalence of these disorders among biological relatives.

Schizophrenia is found in all societies and geographical areas. Although comparable data are difficult to obtain, incidence and lifetime prevalence rates are roughly equal worldwide. The positive symptom component of the disorder usually becomes manifest during late adolescence and early adulthood, although there is a difference in onset associated with gender. In males, the incidence of the onset of positive symptoms peaks during years 17 to 27, whereas in females the peak incidence is a lengthy plateau between the years 17 to 37. Rural and urban incidence figures are probably similar, but there is a greater prevalence of schizophrenia among urban and lower socioeconomic populations. This is generally attributed to the “social drift” phenomenon in which afflicted or vulnerable individuals tend to lose their occupation and social niche and drift toward pockets of poverty and inner-city areas. Occasional geographical areas of increased prevalence of schizophrenia are interesting in terms of illness etiology. For example, a Northern Scandinavian isolated population appears to have a gene pool enriched for schizophrenia vulnerability, probably brought to the region generations ago by two immigrating families.

Because schizophrenia begins early in life; causes significant and long-lasting impairments; makes heavy demands for hospital care; and requires ongoing clinical care, rehabilitation, and support services, the financial cost of the illness in the United States is estimated to exceed that of all cancers combined. In 1990 the direct and indirect costs of schizophrenia were estimated at $33 billion. The locus of care has shifted dramatically over the last 40 years from long-term hospital-based care to acute hospital care and community-based services. In 1955 about 500,000 hospital beds in the United States were occupied by the mentally ill, the majority of whom had a diagnosis of schizophrenia; that figure is now under 250,000.

Deinstitutionalization has dramatically reduced the number of beds in custodial facilities, but an overall evaluation of the consequences of deinstitutionalization is disheartening. Many patients have simply been transferred to alternative forms of custodial care (instead of to treatment or rehabilitative services), including nursing home care and poorly supervised shelter arrangements. Others have been released to communities often unable or unwilling to provide the minimal requirements for clinical care or humane support. For the more fortunate patients the burden of care has shifted to the family, creating an extreme hardship for large numbers of families in this country. The estimated overall financial burden to these families ranges from $2 to 2.5 billion. The less fortunate patient may either have no place to live, be forced to live in circumstances of isolation and hopelessness, or end up in jail. Patients with a diagnosis of schizophrenia are reported to account for 33 to 50 percent of homeless Americans. Managed care places further pressure to reduce bed utilization while communities remain marginally prepared and a relative dearth of alternative care systems exists. Continuity-of-care systems, which include assertive outreach programs and supervised housing and emergency care, provide an effective alternative to hospital-based care for many patients, but costs are substantial, and simply shifting cost from impoverished public hospital sectors has not proved feasible.

ETIOLOGY

The etiological process or processes by which a causal agent creates the pathophysiology of schizophrenia is not yet known. However, a good deal is known about risk factors for developing schizophrenia, which leads to direct inferences regarding possible etiopathophysiologies. Family, twin, and adoptive studies have long since documented a robust contribution of genetic factors to the etiology of schizophrenia, with genetic factors established as relevant to some, perhaps all, cases. However, it is not yet known which genes are involved or how the proteins they produce contribute to the pathophysiology of schizophrenia. Recent linkage analyses have made substantial progress towards identifying a potential location on chromosome 6, and have also provided preliminary indications of additional genetic contributions associated with chromosomes 4, 8, 15, and 22. Other markers of early influences, including gestational and birth complications, exposure to influenza epidemics, Rhesus (Rh) incompatibility, starvation, and an excess of winter births, further suggest a neurodevelopmental pathological process in schizophrenia; however, here, too, the exact pathophysiological mechanisms associated with these risk factors are not known. There are interesting reports that a subgroup of patients with the avolitional component of the illness, as assessed by the deficit syndrome, do not share in the winter birth excess, but rather show a summer birth excess, suggesting the possibility of a separate disease entity within the schizophrenia syndrome. A number of speculations regarding viral and immune mechanisms, sometimes posited as an explanation of the season of birth risk factor, are plausible, but no virus or immune mechanism has yet been established as an etiological factor in schizophrenia. Finally, substance abuse has been identified as a risk factor for developing schizophrenia.

A central conceptual issue in the investigation of the etiology of schizophrenia is whether schizophrenia is a neurodevelopmental or a neurodegenerative disorder. Is the cause of schizophrenia to be found in the failure of the normal development of the brain, or is it to be found in a disease process that alters a normally developed brain? Both these options, or a combination of these options, may be true because the schizophrenia syndrome probably represents more than one disease process, or a developmental abnormality may increase the risk for the subsequent occurrence of the disorder. Although Kraepelin believed that schizophrenia had an early onset and was a chronic deteriorating disorder, the examination of the clinical course of the illness has not been helpful in clarifying this issue. Subtle neurological manifestations, cognitive dysfunction, and disturbances in affect are often present early in the course of illness, usually prior to the onset of hallucinations and delusions, and perhaps from birth. However, it is not clear whether these abnormalities reflect abnormal brain development or are the consequences of an early lesion to a normal brain. Nor is it clear whether the early morbid picture progresses into the full manifestation of psychosis or whether early morbidity represents a vulnerability state susceptible to expressing psychosis in the context of a later lesion, or stressful new demands on cognition and interpersonal skills later in adolescence and early adulthood. It is clear that the illness process usually plateaus within the first 5 to 10 years of psychosis and does not manifest progressive deterioration throughout its course. Late-life improvement, perhaps based on a decrease in the intensity of the psychotic component of the illness, is more typical than continued progression.

An alternative perspective, which has produced somewhat less ambiguous results, is the neuropathological investigation of schizophrenia. Although there are sporadic reports of gliosis in schizophrenic brains, which may indicate the presence of a neurodegenerative disease process and subsequent neuropathological response, the preponderance of current evidence is consistent with the hypothesis that schizophrenia is a neurodevelopmental disorder. Of particular relevance in this regard are the neuropsychological, cognitive psychological, and neuroimaging findings in first-episode cases, which tend to be similar to findings in more chronic cases and, in the few longitudinal studies, tend not to progress. Perhaps even more decisive are abnormalities in morphological features, which are believed to be developmental in nature and are associated with at least some forms of schizophrenia. Such findings range from abnormalities in peripheral development such as finger ridge formation to abnormal cell migration to landmarks of abnormal brain development such as asymmetry of the planum temporale. The consistency with which the known data point to early deviations in the development of the central nervous system has been useful in focusing theory and investigative work.

The explosion of information on the neurobiology of brain development has led to considerable new knowledge on the potential mechanisms of pathogenic influences. It is now clear that subtle deviations in the development of the brain could create dysfunctions associated with specific behaviors. Postmortem findings of abnormalities in neural plate formation, which suggest a deviation in programmed cell migration or reduced cell density, provide intriguing support for the proposition that the developmental process that establishes normal brain cytoarchitecture may have gone awry in schizophrenia. Another view is that the brain has established extensive redundancy during the developing years, and that the fine-tuning necessary for efficient functioning involves eliminating certain nerve cells and many of the synapses connecting cells. A failure to adequately prune nerve cells and synapses, or to err in selection for pruning could, in theory, underlie dysfunctions that later lead to schizophrenia symptoms. Altered nerve cell migration or pruning are speculative, but illustrate plausible mechanisms by which risk factors could alter normal brain development in schizophrenia.

Principal hypotheses regarding causation include altered expression of genes, neuroimmunovirology factors, and birth and pregnancy complications such as hypoxic

or neurotoxic damage.

Altered Expression of Genes Schizophrenia and schizophrenia-related disorders (i.e., schizotypal, schizoid, and paranoid personality disorders; schizophreniform disorder; and other nonaffective psychotic disorders) occur at an increased rate among the biological relatives of patients with schizophrenia. This increased rate is most dramatically illustrated in the case of monozygotic twins, who have an identical genetic endowment and a concordance rate for schizophrenia between 40 to 50 percent. This rate is four to five times the concordance rate in dizygotic twins or the rate of occurrence found in other first-degree relatives (i.e., siblings, parents, or offspring). The role of genetic factors is further reflected in the drop off in occurrence of schizophrenia among second- and third-degree relatives, in whom one would hypothesize a decreased genetic loading. The finding of a higher rate of schizophrenia among the biological relatives of an adopted-away person who develops schizophrenia, as compared to the adoptive, nonbiological relatives who rear the patient has added further support to the overwhelming pedigree and twin study evidence suggesting a significant genetic contribution to the etiology of schizophrenia. However, the data on monozygotic twins clearly demonstrate the fact that individuals who are genetically vulnerable to schizophrenia do not inevitably become schizophrenic; environmental factors must be involved in determining a schizophrenia outcome. If a vulnerability and liability model of schizophrenia is correct in its postulation of an environmental influence, then other biological or psychosocial environmental factors may prevent or cause schizophrenia in the genetically vulnerable individual. Possible environmental factors include the risk factors described, as well as psychosocial factors.

A major obstacle to delineating which genes are involved in schizophrenia is the fact that the mode of genetic transmission in schizophrenia is unknown. No current model (e.g., single gene dominant or recessive, polygenetic, multifactorial, or latent trait) satisfactorily accounts for the data. Determining the mode of transmission in a putative genetic disorder requires a known phenotype and genetic homogeneity across the pedigrees. Neither of these conditions is met in schizophrenia. To understand the etiology of schizophrenia, it will eventually be necessary to identify the actual genes and their products, and to evaluate their expression in the brain. The delineation of the different phenotypic manifestations of the schizophrenic genes or markers of the phenotypes is crucial, both for case ascertainment and in moving genetic inquiry closer to the neuronal effects of schizophrenia-related genes. Measures of smooth pursuit eye movements (SPEM), information processing (e.g., the continuous performance task and forced span of attention test), and sensory gating are the most prominent candidate markers. These measures have been found to distinguish schizophrenic probands and their biological relatives from control groups. Similarly, patients and their biological relatives are more likely than comparison groups to fail to inhibit neuronal response to a repeated stimuli (measured by a peak amplitude in electrical signal at about 50 milliseconds). The P50 sensory gating phenomena marker is of particular interest because it captures a basic neuronal property whose dysfunction could explain schizophrenic pathophysiology. In a recent application a defect in a neuronal mechanism that regulates response to auditory stimuli was used to define a schizophrenia phenotype, and positive linkage was found on an area of chromosome 15 with markers near the site of the a 7 nicotinic receptor. This receptor is thought to mediate normal inhibition of the auditory evoked response to the second of paired stimuli.

It has proven exceedingly difficult to progress from evidence confirming a genetic contribution to the etiology of schizophrenia to evidence implicating specific genes in the disease. Nonetheless, the area of genetic investigation is highly promising because there is unequivocal evidence for a genetic contribution to some, perhaps all, forms of the illness. There is presently an explosion of knowledge and techniques relevant to discovering the genetic basis for human disease. Linkage analysis has quickly moved from a few marker probes to banks of hundreds, and the entire genome will soon be examined with probes spaced along all chromosomes. Analytic techniques have been developed to evaluate polygenetic disorders, and gene substructure techniques now enable investigators to focus on candidate genes found to distinguish schizophrenic brains.

Neuroimmunovirology Immune and viral hypotheses of schizophrenia are as old as scientific knowledge in these areas. That a virus could cause a neuropsychiatric disease was confirmed when Pasteur isolated the rabies virus in 1881. But schizophrenia is not an acute encephalitis or a fulminating infection; it involves more subtle pathophysiological mechanisms, which make it more difficult to establish etiology. Furthermore, the epidemiological data supporting an infectious theory, although interesting, is weak. Schizophrenia may have a north to south prevalence gradient in the Northern hemisphere (south to north in the Southern hemisphere), may be endemic to a few areas (e.g., northern Sweden), has a winter birth excess, and, similar to multiple sclerosis, has monozygotic twin discordance. However, it has been difficult to conduct definitive studies of immunovirological hypotheses because any potential marker of an immune or viral process associated with schizophrenia is applicable to only some cases of schizophrenia and is subject to interpretation as being due to conditions associated with the disease (e.g., crowding of chronically hospitalized patients, exposure of chronic patients living in low socioeconomic circumstances, and poor health habits).

Viral theories remain popular despite the difficulty of validating any particular version. Their popularity stems from the fact that several specific viral theories have the power to explain the particular localization of pathology necessary to account for a range of manifestations in schizophrenia without overt febrile encephalitis. The six general pathogenic models of viral and immune pathophysiology relevant to schizophrenia are described next.

Retroviral Infection A retrovirus can insert itself into the genome and thereby alter it which could initiate a genetic contribution to schizophrenia. It is postulated that the retrovirus inserts itself into the genome and alters the expression of the host's own genes and the genes of the host's offspring toward the development of schizophrenia (the virogene hypothesis). There is no evidence at present to support the retrovirus theory of schizophrenia, and at least one study has failed to find retrovirus-associated enzymes that would be present in an active infection but not in a virogene scenario.

Current or Active Viral Infection Many researchers postulate that viruses with an affinity for the central nervous system are involved in the etiology of schizophrenia. It is hypothesized that either a neurotropic virus infects nerve cells in discrete parts of the brain and causes sustained alterations in the functioning of the involved neural systems, or that byproducts of a viral infection have direct toxic effects on nerve cell functioning. Abnormal immune indices have been reported in schizophrenia and could be indicative of an active infectious process; however, most investigators consider the viral factor to be an early event that results in ensuing brain damage, which in twins has a long-lasting effect.

An alternative formulation of this hypothesis is based on the observation that viruses can infect the brain, with substantive disease manifestations only showing up many years later. In theory, this could account for the subtle early manifestations frequently observed in schizophrenic patients, which are followed by more intense symptom manifestations 10 to 30 years later.

A substantial challenge to either formulation of the current or active viral infection hypothesis is the absence of direct evidence substantiating a viral etiology, including the lack of physical signs of encephalitis (e.g., lymphocytic infiltrate) in postmortem tissue and the failure to recover or isolate a putative agent.

Past Viral Infection This hypothesis posits a virus infecting certain brain tissues either early in life to create a vulnerability to schizophrenia or as a causal mechanism for the initial illness processes that later lead to the picture of classic schizophrenia. The resulting tissue damage produces long-lasting alterations in neural systems, leading to schizophrenia manifestations without persistent viral infection. Gliosis, sometimes observed in postmortem tissue, would support the proposition of an earlier viral infection, and would also help account for the fact that signs of encephalitis are not ordinarily observed in the postmortem brain tissue of schizophrenia patients. A limited number of experiments have been unsuccessful in using brain material from schizophrenic patients as a source for transmitting central nervous system (CNS) viral infection into the brain tissue of other species. Although many viruses do not easily cross species, these studies lend some support to the proposition that even if a virus is relevant to the etiology of schizophrenia, it is not causing an active infection at the time of patient death.

Virally Activated Immunopathology One of two general mechanisms is proposed in this category. The first is based on the observation that viruses are normally endogenous to the human brain and are discontinuously or focally distributed in the brain. Periodic viral reactivation of these foci normally does not result in psychotic symptoms. However, in an individual with a genetically or environmentally determined abnormal immune response to viruses, it is hypothesized that viral reactivation could result in an induction of schizophrenic psychopathology. This theory regards the products of immunoreactivity, such as alpha interferon, as the mediators of the pathogenic influence. There is little direct evidence to support this hypothesis, but it receives indirect support from findings of abnormalities in alpha interferon responsiveness in schizophrenic patients as compared to normal controls.

The second mechanism in this category is that the virus may induce the host to fail to recognize its own tissues as “self” and, as a consequence, mount a destructive immune response against them. The virus may do this by altering some cellular component, such as normally cryptic neural cell surface proteins, causing it to stimulate a host response. A cytotoxic or antibody response would cause direct interference of nerve-cell function by either destruction of the cells or, in the case of receptor proteins, altered neurotransmission.

Autoimmune Pathology The aforementioned viral induction of an autoimmune pathology is an example of this pathogenic model. Schizophrenia has also been hypothesized to be an idiopathic autoimmune disease, such as rheumatoid arthritis or systemic lupus erythematosus; wherein, for reasons that are not entirely clear but probably involve genetics, some tissues are not recognized as self and become the target of immune response.

Secondary Influences: In Utero Exposure to Maternal Infection A number of epidemiological studies have reported that women who are exposed to influenza epidemics during the second trimester of pregnancy are more likely to give birth to offspring who are at increased risk for schizophrenia. This observation raises the possibility that some attribute of maternal infection such as fever or cytokine activation perturbates normal brain development during the period of active neural cell migration. This interesting etiological angle has recently been challenged by studies attempting to assess whether the mother was actually infected, rather than simply being exposed to an epidemic. Such case ascertainment might strengthen the finding, but at present this type of study can only be validly conducted using a prospective design and contemporaneous serological and molecular diagnostic techniques.

Birth and Pregnancy Complications Infants born with a history of pregnancy or birth complications are at increased risk for developing schizophrenia as adults. The reason for this has not been established. The following plausible explanations, which are not mutually exclusive, guide present-day research.

1. The genes that create vulnerability for schizophrenia may also alter early embryonic development in a manner that leads to increased likelihood of gestational and birth complications.
2. Adverse influences on the developing brain during early gestation create a risk for both birth complications and schizophrenia. The potential role of Rh incompatibility as a risk factor for schizophrenia is an interesting example of this proposition.
3. Gestational or birth complications may cause hypoxic damage. Brain regions most frequently implicated as deviant in schizophrenia (e.g., hippocampus) are among the areas in the developing brain that are most sensitive to hypoxia.

PATHOPHYSIOLOGY

Schizophrenia is a disease of the brain. However, it is easier to make this assertion than to document any actual deviations in brain physiology. Since the illness represents a disturbance in some, but not all, brain functions, it is reasonable to suppose that specific areas or neural circuits of the brain are involved and that the manifestations of schizophrenia must necessarily involve altered processing of physiological information; this altered processing would, in turn, be dependent on disturbances of cytoarchitectural, biochemical, or electrophysiological properties of the neural systems.

Throughout most of this century examination of postmortem brain tissue has been the principal source of data with relevance to the neuroanatomy of schizophrenia. Early reference to schizophrenia as “the graveyard of neuropathology” was not because of a lack of neuropathological findings, but rather because of the lack of a discernible pattern in the frequently observed pathological findings and the possibility that deviations were either artifactual in nature or were a consequence, rather than a cause, of the disease. For example, head trauma and viral infections affecting the brain would be more common in crowded custodial hospitals than in typical comparison groups. Moreover, the widespread use of neuroleptic drugs in the treatment of schizophrenia introduced additional artifacts in the investigation of brain pathophysiology. Finally, knowledge of brain-behavioral relations was not sufficiently detailed to guide neuropathological inquiry during much of this century.

Scientists have long been keenly aware of the necessity for the development of noninvasive techniques to study the functioning brain of living patients. This is particularly important in the absence of valid animal models. During the middle third of this century, pneumoencephalography (PEG) provided substantial evidence for enlarged brain ventricles, suggesting diminished tissue in schizophrenia compared to controls. Electroencephalography (EEG) provided nformation on cortical surface electrical activity, but neither PEG nor EEG techniques could provide a comprehensive evaluation of the functioning human brain.

The development of structural (e.g., computerized axial tomography [CT] and magnetic resonance imaging [MRI]) and functional (e.g., positron emission tomography [PET], single photon emission computerized tomography [SPECT], functional MRI, magnetoencephalography, and MR spectroscopy) in vivo imaging techniques have made a more detailed view of brain structure and physiology possible. These techniques have become available at a time when a better understanding of the interconnections between cortical and subcortical structures and their implications for brain-behavior relations is emerging from preclinical studies of the brain. CT studies have replicated the PEG observation of enlarged ventricles and have further shown that a substantial proportion of schizophrenic patients, in comparison to normal controls, exhibit increased sulcal widening. These results suggest that schizophrenic patients may have relatively less brain tissue, a condition that could represent either a failure to develop or a subsequent loss of tissue. With its enhanced gray and white matter resolution, MRI is able to provide a far more detailed assessment of specific brain structures. Studies employing MRI have found evidence in schizophrenic patients for decreased cortical gray matter, especially in the temporal cortex, decreased volume of limbic system structures, (e.g., the amygdala, hippocampus, and parahippocampus), and increased volume of basal ganglia nuclei. These findings are consistent with the findings of neuropathological examinations of postmortem tissue, including ultrastructural examination, which in some cases indicates cell loss, misalignment of cells, altered intracellular structure and protein expression, or gliosis.

Structural findings may help clarify the meaning of altered patterns of function. Functional imaging studies have documented abnormal patterns of glucose metabolism or blood flow during the performance of specific cognitive tasks. These techniques are also able to provide insights into the functional neuroanatomy of the various symptom complexes that characterize patients with schizophrenia, with preliminary evidence suggesting a differential association of functional indexes with the positive psychotic symptoms and primary, enduring negative symptoms (Fig. 12.1-1 and Fig. 12.1-2).

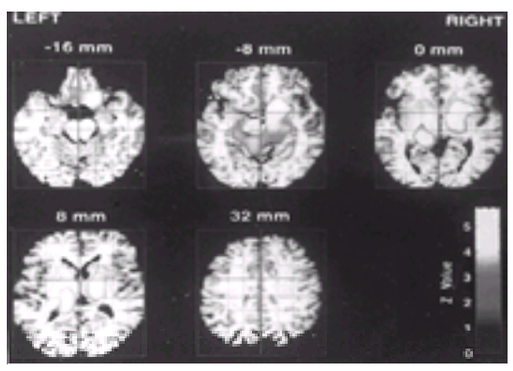

FIGURE 12.1-1 Axial sections demonstrating brain areas with significantly increased activity during auditory verbal hallucinations in the group study. Functional PET results (threshold at $Z > 3.09$, $P < 0.001$, by reference to the unit normal distribution) are displayed in color, superimposed upon a single structural T1-weighted magnetic resonance imaging (MRI) scan that has been transformed into the Talairach space for anatomical reference. Section numbers refer to the distance from the anterior commissure-posterior commissure line, with positive numbers being superior to the line. The areas of activation extend into the amygdala bilaterally, and into the right orbitofrontal cortex. Although these regions of extension are consistent with the limbic paralimbic component of activity during hallucinations, and may contribute to drive and affect in this context, definitive statements cannot be made in the absence of discrete maxima. (Reprinted with permission from Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, Seaward J, McKenna P, Chua SE, Schnorr L, et al: A functional neuroanatomy of hallucinations in schizophrenia. Nature *378:*1769, 1995.)(See Color Plate 7.)

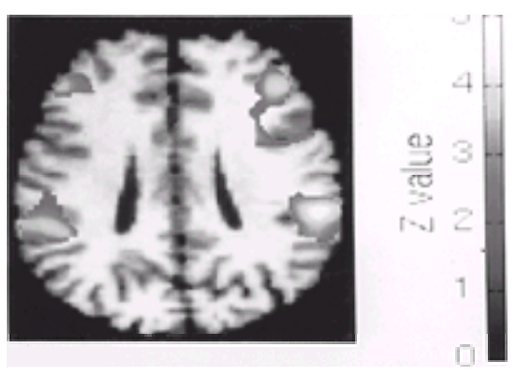

FIGURE 12.1-2 There is a significant difference in O_{15} activity in the prefrontal and parietal cortex during the performance of an auditory discrimination task in deficit and nondeficit patients, with deficit patients having decreased activity in these regions. (Courtesy of A. Lahti, Maryland Psychiatric Research Center, Baltimore, MD.)

(See Color Plate 7.)

Present-day knowledge of the pathophysiology of schizophrenia is acquired from the study of living subjects by using structural and functional imaging, and anatomically relevant symptom assessment and neuropsychological techniques. These technologies are supplemented by advances in postmortem biochemical, molecular, and structural evaluations to test increasingly sophisticated neuroanatomical and biochemical theories of schizophrenia.

Major Neuroanatomical Theories Over the last 20 years there has been a gradual evolution from conceptualizing schizophrenia as a disorder that involves discrete areas of the brain to a perspective that views schizophrenia as a disorder of brain neural circuits. These neural circuit models of the pathophysiology of schizophrenia posit that either a structural or a functional lesion disrupts the functional integrity of the entire circuit. There are several factors that have contributed to this change in perspective. First, the delineation of the neuroanatomy of the different neurotransmitter pathways has led to an increased appreciation of how different brain regions are connected with each other and how cortical and subcortical structures are able to reciprocally regulate the function of each other. For example, the identification of the mesolimbic and mesocortical dopaminergic pathways contributed to the development of neuroanatomical hypotheses implicating the prefrontal cortex and limbic system in the pathophysiology of schizophrenia. The further delineation of the reciprocal regulatory pathways between the prefrontal cortex and the limbic system, particularly the hippocampus, led to more recent formulations of these hypotheses, in which limbic and prefrontal neuroanatomical models of schizophrenia have been integrated into a single unifying neurodevelopmental theory of schizophrenia. These hypotheses propose that an early developmental lesion of the dopaminergic tracts to the prefrontal cortex results in the disturbance of both prefrontal and limbic system function, and leads to the positive and negative symptoms and cognitive impairments observed in patients with schizophrenia.

Prefrontal cortex and limbic system hypotheses are the predominant neuroanatomical hypotheses of schizophrenia. The demonstration of decreased volumes of prefrontal gray or white matter, prefrontal cortical interneuron abnormalities, disturbed prefrontal metabolism and blood flow, decreased volumes of hippocampal and entorhinal cortex, and disarray or abnormal migration of hippocampal and entorhinal neurons provide strong support for the involvement of these brain regions in the pathophysiology of schizophrenia. In the context of neural circuit hypotheses linking the prefrontal cortex and limbic system, studies demonstrating a relation between hippocampal morphological abnormalities and disturbances in prefrontal cortex metabolism or function are particularly interesting.

A second contributing factor to the adoption of a neural circuit conceptual framework has been the increased understanding of how the brain is organized into local microcircuits, which consist of the connections among afferent and efferent neurons and interneurons, and macrocircuits. An example of the latter are the segregated parallel basal ganglia-thalamocortical neural circuits, which connect the cerebral cortex with the thalamus through the basal ganglia. Each of these circuits is hypothesized to subserve a discrete range of functions. Several investigators have used these circuits as a starting point for their hypotheses of schizophrenic pathophysiology. These hypotheses differ from each other primarily on their point of emphasis. For example, integrating data from animal studies, and neurobehavioral, functional, and structural imaging studies in humans, it has been hypothesized that dysfunction of the anterior cingulate basal ganglia-thalamocortical circuit underlies the production of psychotic symptoms (Fig. 12.1-1) and dysfunction of the dorsolateral prefrontal circuit underlies the production of primary enduring negative or deficit symptoms (Fig. 12.1-2). Dysfunction in one of these circuits may be independent from dysfunction in the other.

P>A third factor has been the elucidation of the neural basis of cognitive functions observed to be impaired in patients with schizophrenia. The observation of the relationship among impaired Wisconsin Card Sort Test (WCST) performance and diminished prefrontal cortex blood flow and diminished hippocampal volume provides strong support for the validity of prefrontal cortex or limbic system neuroanatomical models. Similarly, the delineation of the neural circuits for language and attention or information processing have influenced the conceptualization of schizophrenia pathophysiology. The classic language circuit, which includes Wernicke's and Broca's areas and associated cortical and subcortical structures, has been hypothesized to be involved in the production of hallucinations, delusions, and positive formal thought disorder. This hypothesis is the most important alternative to the anterior cingulate hypothesis for positive symptoms. The involvement of this circuit, at least for auditory hallucinations, has been documented in a number of functional imaging studies contrasting hallucinating versus nonhallucinating patients.

Attention and information processing abnormalities are routinely observed in patients with schizophrenia. The type of abnormalities range from disturbances in sensory gating to disturbances in visual information processing. The latter impairments have been argued to be selectively related to negative symptoms. The overlap between brain regions that have been implicated in the production of negative symptoms and the visual information processing neural circuit, which includes inferior and superior parietal and prefrontal cortices, caudate and thalamic nuclei, and the reticular activating system, provides a neuroanatomical rationale for the relationship between these two dimensions of schizophrenia and a conceptual framework for future studies of the neuroanatomy of negative symptoms.

The development of neural circuit hypotheses offers tremendous advantages to the investigation of the neuroanatomy of schizophrenia. First, these hypotheses more accurately reflect the actual organization of the brain. Second, models of neural circuit hypotheses can be developed to investigate how perturbations of circuit function can lead to schizophrenia signs and symptoms. Neural circuit models have been created for both the cognitive and symptom manifestations of schizophrenia. Third, neural circuit hypotheses provide a conceptual framework for hypothesis-testing studies and optimize the interpretation of information derived from current brain imaging and postmortem studies. Finally, the use of neural circuit models implicates brain regions, such as the thalamus and the cerebellum, that are not typically conceptualized as being central to the neuroanatomy of schizophrenia.

Major Biochemical Theories Information is processed in neuronal networks through the transmission of an electrical signal from a nerve cell through its axon and across synapses to postsynaptic receptors on other nerve cell components. Nerve cells generally receive, process, and send signals to and from thousands of other cells. The transmission of the signal across the synapse and the processing of the signal within a cell involve a complex series of biochemical events that require large amounts of energy and include gene expression and the synthesis and degradation of protein. It is evident that physiological function in any brain system involves the chemistry of that system, and that dysfunction can emanate from these biochemical processes. It is therefore natural to assume that the biochemistry of the brain plays a fundamental role in the disruptions of brain function involved in schizophrenia. The move from a general concept of the biochemistry of schizophrenia to specific theories is based on two principal sources of knowledge. The first is an ever-increasing understanding of intracellular communication from the cell membrane to the nucleus and the cell's genetic material and of intercellular communication through the various neurotransmitter systems of the brain. The second source is knowledge of the mechanism of action of drugs that can induce schizophrenia-like behaviors or that alter symptom expression in patients with schizophrenia. Our knowledge of cellular communication and the pharmacological actions of antipsychotic drugs have led to biochemical hypotheses involving dopamine, noradrenalin, serotonin, acetylcholine, glutamate, and several neuromodulatory peptides and their receptors. Because there are many possibilities, it is important to understand the general development of a biochemical hypothesis of schizophrenia, of which the dopamine hypothesis is the most prominent and enduring.

Dopamine and Schizophrenia The hyperdopamineric hypothesis of schizophrenia arose from two sets of observations of drug action relating to the dopaminergic system. Drugs that increase dopamine system activity, such as d-amphetamine, cocaine, levodopa (Larodopa), and methylphenidate (Ritalin), can induce a paranoid psychosis that is similar to some aspects of schizophrenia. When administered to schizophrenic patients, these compounds may produce a transitory worsening of symptoms, especially in the area of hallucinations, delusions, and thought disturbance. In contrast, drugs that share the capacity to block postsynaptic dopamine receptors reduce the symptoms of schizophrenia. Substantial evidence supports the role of postsynaptic dopamine blockade as an initiating factor in a cascade of events responsible for the mode of therapeutic action of antipsychotic drugs. Other mechanisms, such as depolarization blockade, have been implicated as plausible explanations for long-term antipsychotic effects. That these actions are actually corrective for the pathophysiological disturbance in schizophrenia is suggested by the fact that dopamine-stimulating drugs can worsen schizophrenic symptoms or induce psychosis. This rationale for the role of dopamine excess, particularly for the cognitive and positive symptom aspects of schizophrenia, is compelling.

However, despite the compelling evidence for the role of dopamine in schizophrenia, testing the hypothesis has proven problematic. Clinical studies across a broad range of indices of dopamine metabolism have been characterized by marked variability in results. The most decisive clinical testing of the hypothesis has been at the level of observed drug action and symptom manipulation. Studies aimed at measuring abnormal concentrations of dopamine or its metabolites in blood, urine, and spinal fluid are confronted by problems that are almost insurmountable. In large fluid compartments, alterations in dopamine metabolism associated with schizophrenia will represent only a minor contribution to the particular index of dopamine metabolism; spinal fluid necessarily provides a summation of total brain activity, most of which is not considered germane to schizophrenia, and blood and urine provide even more indirect indices.

Functional imaging studies provide indirect evidence of dopamine involvement through the examination of metabolic rates in brain regions where dopamine is an important neurotransmitter. For example, data confirming metabolic alterations in limbic anatomy are consistent with a disturbance in dopamine metabolism, but it is not possible to determine the extent to which this reflects an alteration of dopamine biochemistry versus an alteration of any one of a number of interacting neurotransmitter and neuromodulatory systems. A more informative approach for assessing abnormal dopamine metabolism in patients with schizophrenia is to infuse

subjects with an indirect dopamine agonist and then determine the extent to which radioligand occupancy of postsynaptic dopamine receptors is reduced by competition with the increased endogenous dopamine. The comparison of preinfusion and postinfusion radioligand occupancy provides an index of dopamine release and reuptake rates. PET studies of dopamine receptor distribution and the density of receptor expression may offer an alternative approach for documenting the dopamine hypothesis. The observation of an increased quantity of dopamine type 2 (D_2) receptors in the caudate nucleus of drug-free schizophrenic patients is an example of this approach, but replication has been difficult. The extension of this approach to other dopamine receptor types is an important new direction of research.

Finally, there is the potential for the relatively precise biochemical study of dopamine in postmortem tissue, but here, as with the use of body fluids, sources of artifact and imprecision have been difficult to manage. The concentration of a neurotransmitter in any tissue will be altered as cellular components break down following death and as small differences in dissection from brain to brain take place. The administration of neuroleptic drugs during life almost always confounds the biochemistry of postmortem tissue, and one can rarely be sure of the extent to which any biochemical finding is secondary, rather than primary, to the schizophrenic disease process. In addition, there are a large number of candidate areas for brain dysfunction, so that one may easily examine the wrong location. It is also quite possible that areas of biochemical dysfunction earlier in life are no longer dysfunctional at the time of death or that the biochemistry of death may obscure the biochemistry of life.

Despite these methodological limitations, postmortem studies have reported differences between schizophrenic and control brains. For example, increased concentration of dopamine has been found in the left amygdala (a limbic system structure) in the postmortem brains of patients with schizophrenia. This finding has been replicated and, since it is lateralized, is not likely to be an artifact. There has also been a report of an increase in D_2 postsynaptic receptors in postmortem tissue of schizophrenic patients whose medical records provided a diagnosis of schizophrenia but did not reveal neuroleptic drug use. These results suggest that the increase in binding (receptor) number is not secondary to neuroleptic drugs. The investigation of receptor abnormalities has been extended to other dopamine receptor types, and an increase of D_4 receptors in entorhinal cortex, independent of antipsychotic use, has been reported.

Although conclusive evidence for the dopamine excess theory has been elusive, the hypothesis remains a viable explanation for the positive symptoms of schizophrenia. It is a particularly robust proposition for explaining the antipsychotic effect of neuroleptic drugs. Interestingly, recent studies have suggested the possibility that a dopamine deficiency may also occur in schizophrenic patients. For example, an inverse correlation between cerebrospinal fluid (CSF) and homovanillic acid (HVA) concentrations and negative symptoms has been reported. Also, patients with influenza encephalitis, who were mistaken for being schizophrenic, tended to have emotional dullness and low drive. Similarities in these cases with aspects of Parkinson's disease (which is known to involve loss of dopamine neurons) and the fact that some of these postencephalitic patients developed Parkinson's disease, lends support to a dopamine deficiency hypothesis for the negative symptom aspect of schizophrenia. In addition, neuroleptic drugs, which are dopamine-blocking agents, produce behaviors suggestive of the negative symptoms of schizophrenia in animals and humans free of mental illness. A modification of the dopamine hypothesis, incorporating the possibility of concomitant dopamine excess and deficiency, would restrict dopamine excess to the dopaminergic pathways projecting to the basal ganglia and limbic system and dopamine deficiency to the mesocortical pathways. Hypofunction of the mesocortical neurons would account for the negative symptoms of schizophrenia.

Glutamate and Schizophrenia Glutamate is the major excitatory neurotransmitter in the brain. Interest in the possible role of glutamate in the pathophysiology of schizophrenia has emerged from an increased understanding of the N-methyl- D-aspartate (NMDA) receptor complex, a major glutamate system receptor; an increased understanding of the interactions between glutamatergic and dopaminergic and GABAergic systems; and observations of the acute and chronic effects of phencyclidine (PCP). The consequences of PCP use provide a compelling model of schizophrenia symptomatology. Short-term administration of PCP produces symptoms that have been argued to mimic both the positive and negative symptoms of schizophrenia. Chronic administration produces a hypodopaminergic state in the prefrontal cortex, a state that has been argued to result in negative symptoms. PCP occupies receptors within the open calcium channels of the NMDA receptor complex, thereby blocking ion flow. PCP and the analogue ketamine (Ketalar) interfere with glutamatergic transmission. In addition to the observation of schizophrenia-like symptomatology in humans abusing PCP or ketamine has been used in the laboratory and has been observed to produce transitory mild manifestations of positive and negative symptoms in normal volunteers and a transitory and mild worsening of positive symptoms in patients with schizophrenia. Activation of dopamine receptors inhibiting glutamatergic neurons or decreased NMDA-mediated inhibition of dopamine neurons, either directly or through the actions of GABAergic interneurons, could be associated with a dopamine-excess psychosis (Fig. 12.1-3). These considerations support a hypoglutamatergic hypothesis for schizophrenia pathophysiology and predict a therapeutic effect for compounds activating the NMDA receptor complex. This is a difficult strategy to implement because excessive glutamatergic activity is neurotoxic; however, activation of the NMDA receptor complex via the glycine site with either glycine or d-cycloserine has been reported to alleviate negative symptoms in patients with schizophrenia.

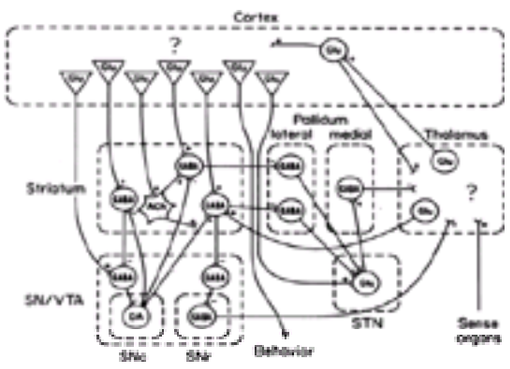

FIGURE 12.1-3 A tentative scheme of interactions between glutamate and dopamine in the basal ganglia. The cholinergic interneurone in the striatum is a large, aspiny cell with a rich collateral network that can be assumed to make synaptic contacts with a large number of other striatal cells. The cholinergic interneurone receives a cortical glutamatergic input on its soma, while its axon terminals are in synaptic contact with medium-sized, spiny GABAergic output neurones. Only two such GABA neurons are shown but in reality it is reasonable to assume that one cholinergic neurone innervates many GABAergic neurones. The cholinergic interneurone also makes contact (although maybe not forming a real synapse) with dopaminergic nerve terminals. From the way the synapses are drawn here, the cortex would be able to control the activity in the GABAergic output neurones projecting to the thalamus via the medial segment of the globus pallidus (partly via the subthalamic nucleus and substantia nigra pars reticulata. In this manner the cortex can selectively suppress impulse flow in one subpopulation of GABAergic projection neurones while facilitating impulse flow in another, thus presumably enabling a meaningful behavior by suppressing irrelevant locomotor programs. The importance of glutamatergic pathways for maintaining a purposeful behavior is revealed by the primitive locomotor pattern that results from treatment with the NMDA antagonist MK-801.
For the sake of simplification, the different thalamic nuclei are not shown. Conceivably, striatopallido-thalamic neurones can influence the entire thalamus via, for example, the reticular nucleus, which communicates with all other thalamic nuclei. Apart from the corticostriatal glutamatergic pathway, there are at least three other corticifugal systems that the cortex can use to protect itself from overstimulation: (1) the corticonigral projection; (2) the corticothalamic projection, which terminates in the thalamic intralaminar nuclei, from which a thalamostriatal projection originates; and (3) the corticosubthalamic projection.

Abbreviations: DA, dopamine; Glu, glutamate; Snc, Substantia nigra pars compacta; Snr, substantia nigra pars reticulata; STN, subthalamic nucleus; VTA, ventral tegmental area. (Reprinted with permission from Carlsson M, Carlsson A: Interactions between glutamatergic and monoaminergic systems within the basal ganglia-implications for schizophrenia and Parkinson's disease. Trends Neurosci *13:*896, 1990.)

The glutamatergic hypothesis exemplifies a major transition that has occurred recently in the biochemistry of schizophrenia. Prior to this transition, observations of drug actions in schizophrenia first led to clinical treatment and then to the advancement of the pathophysiological theory of schizophrenia. With the ever-increasing knowledge of the neural organization of the brain and of the various properties and receptor sites of neurotransmitters, it is now possible to postulate pathophysiological theory first and then attempt to derive new clinical treatment from theory. New treatment approaches will be developed more rapidly in the future, based on a broader range of pathophysiological hypotheses and the availability of animal models for aspects of the illness that are not therapeutically responsive to dopamine blockade-based medications.

Other Neurotransmitters and Neuromodulators Any neurotransmitter involved in neural systems subserving behaviors whose disruption could result in symptoms

of schizophrenia is naturally of interest in schizophrenia theory and research. The rich innervation of the frontal cortex and limbic system with serotonergic neurons, the modulatory effect of these neurons on dopaminergic neurons, and the involvement of these pathways in the regulation of a broad range of complex functions has led several investigators to posit a pathophysiological role for serotonin in schizophrenia. These hypotheses have taken various forms over the course of the last four decades. In the early 1950s a serotonergic deficiency hypothesis was proposed for schizophrenia. Observations of hallucinations in subjects who had ingested lysergic acid diethylamide (LSD), a compound that is chemically similar to serotonin and blocks serotonin receptor sites, furthered the hyposerotonin hypothesis. However, drugs that decrease serotonin activity tend to reduce schizophrenic symptoms (e.g., reserpine [Serpasil], some antipsychotics, clozapine [Clozaril]), and have diminished interest in the deficiency hypothesis.

Of greater current interest are hypotheses positing that a serotonin excess causes positive and negative symptomatology. The robust serotonergic antagonist activity of clozapine and other new-generation antipsychotics, coupled with clozapine's demonstrated effectiveness for positive symptoms in chronic, treatment-resistant patients have contributed to the current emphasis on this proposition. However, several studies have raised questions about the efficacy of serotonin antagonists for either negative symptoms broadly defined or deficit symptoms. Moreover, pharmacological modification of serotonin systems with specific serotonergic agents has not produced impressive clinical results.

As with the dopamine hypothesis, the strength of the support for the serotonin hypothesis is derived from reasoning based on knowledge of brain and behavior relationships, the anatomy of neural transmitter systems, and drug mechanism of actions, and the same weaknesses in the clinical and postmortem studies on dopamine apply to serotonin also.

A similar rationale can be applied to construct hypotheses implicating norepinephrine in the psychopathology of schizophrenia. Anhedonia, (i.e., the impaired capacity for emotional gratification and the decreased ability to experience pleasure), has long been noted to be a prominent feature of schizophrenia. A selective neuronal degeneration within the norepinephrine reward neural system could account for this symptom. However, biochemical and pharmacological data bearing on this proposal are inconclusive. As with dopamine and serotonin, there have been both noradrenergic excess and deficiency pathophysiological hypotheses.

Neuromodulatory hypotheses focus on the fact that neuropeptides, such as substance P and neurotensin, are co-localized with the catecholamine and indolamine neurotransmitters, and influence the action of these neurotransmitters. Alterations in neuromodulatory mechanisms could facilitate, inhibit, or otherwise alter the pattern of firing in these neuronal systems. Explorations of neuromodulator hypotheses are preliminary and inconclusive at this time.

Integrative Hypotheses The natural evolution of pathophysiological hypotheses of schizophrenia is the development of comprehensive models that integrate both neuroanatomical and biochemical hypotheses. The superimposition of the neurotransmitters involved in the connections among cortical, basal ganglia, and thalamic structures that comprise the basal ganglia-thalamocortical neural circuits is a prime example of this approach. Through glutamate projections from the cortex to the basal ganglia, the cerebral cortex facilitates the performance of selected behaviors while inhibiting others. The excitatory glutamatergic neurons terminate on GABAergic and cholinergic neurons, which in turn suppress or excite dopaminergic and other neurons. This regulatory activity can enable the cortex to protect itself from overstimulation from thalamocortical neurons. The elucidation of the neuroanatomy and biochemistry of cortical microcircuits has also served as a starting point for the articulation of pathophysiological hypotheses of schizophrenia. These integrative models provide a framework for identifying potential neurotransmitter targets for drug development, as well as providing explanatory models for the observed effects of pharmacological agents in patients with schizophrenia (e.g., PCP-induced psychotic symptoms mediated through the interactions of glutamate and other neurotransmitter systems in the neocortex, basal ganglia, or limbic system structures).

DIAGNOSIS

The history of the diagnosis of schizophrenia is often misunderstood, which has led to erroneous conclusions about the validity of the diagnostic process. Throughout most of the twentieth century there has been substantial agreement among diagnosticians throughout the world, using seemingly divergent diagnostic approaches, in the recognition of typical cases of schizophrenia. There has also been no difficulty in distinguishing schizophrenia from normality. Although useful refinements have evolved, diagnostic systems in place when effective drug treatment was introduced in 1952 were capable of identifying suitable subjects for therapy.

The major areas of disagreement among diagnostic approaches were how broad the definition of schizophrenia should be; whether positive symptoms, including hallucinations, delusions, and positive formal thought disorder, were required; and whether positive symptoms in the absence of known organic causes always signified schizophrenia. In general, the broader the definition the greater the likelihood that more subtle cases would be included and the greater the likelihood that disagreement would arise regarding the diagnosis of such cases. Even in such cases, there was little disagreement regarding the presence of psychopathology; rather, when present, the disagreement focused on whether the psychopathology observed was part of schizophrenia. This difference in viewpoint did create problems, which became important as different types of drugs were found to be effective for different classes of illness.

The success of the scientific search for more effective drugs for specific disease classes created the urgency to establish an agreed-upon diagnostic approach to schizophrenia and the major affective disorders in order to maximize appropriateness of treatment. The need for such agreement was also highlighted by the results of an influential study comparing diagnostic approaches in the United Kingdom with those in New York City, in which it was convincingly demonstrated that American diagnosticians employed a much broader and less defined construct of schizophrenia than their British counterparts. For a time in North America, especially in the northeastern United States, a broad definition of schizophrenia tended to include two categories of patients ill suited for the standard pharmacological treatment of schizophrenia. The first category was patients with bipolar or major depressive disorders with psychotic features, who, if erroneously considered to have schizophrenia, were administered antipsychotic medication rather than the more specific and effective treatments available for patients with these disorders (i.e., antidepressants, lithium, and electroconvulsive therapy). The second category included patients with schizophrenia spectrum personality disorders, (i.e., schizoid, schizotypal, and borderline personality disorders). These patients were sometimes misdiagnosed as having schizophrenia and were thus likely to be administered drugs designed for the positive symptoms of schizophrenia, which provided them little benefit and subjected them to substantial risk.

A considerable body of research during the 1960s and 1970s clarified many diagnostic issues and set the stage for the development of a diagnostic system implemented in the third edition of the American Psychiatric Association's *Diagnostic and Statistical Manual of the Mental Disorders* (DSM-III). The DSM-III approach, with specified symptom-based diagnostic criteria and demonstrated reliability, is now the accepted diagnostic system in North America and throughout the international research community. The use of this approach has led to the reliable and consistent differential diagnosis of schizophrenia, which has enhanced scientific and clinical communication and substantially increased the likelihood of the effective use of diagnostically specific treatments. The DSM-III approach has been retained in the revised third edition of DSM (DSM-III-R) and the fourth edition of DSM (DSM-IV) and has been incorporated in the International Classification of Diseases (ICD) diagnostic system. The ultimate goal is to standardize the diagnosis of schizophrenia across all diagnostic systems. Substantial progress has been made in this area, with extensive integration between DSM-IV and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10).

Beyond Diagnosis A valid diagnostic system for schizophrenia has considerable utility for clinical and epidemiological purposes. It is now possible to estimate the occurrence of schizophrenia accurately, to identify individuals suffering from the illness process, to guide treatment and rehabilitation considerations, and to differentiate schizophrenia from other illnesses with similar manifestations but with importantly different treatment requirements. However, diagnosis at the syndrome level has not been an adequate guide to the scientific study of either the etiology or pathophysiology of schizophrenia nor to the development of treatments for all key features of the illness.

The traditional approach to reducing the heterogeneity of the schizophrenia syndrome has been to delineate subtypes and attempt to confirm or disprove their validity. The classic subtypes, disorganized (DSM-IV) or hebephrenic (ICD-10), paranoid (DSM-IV and ICD-10), catatonic (DSM-IV and ICD-10) and simple schizophrenia (ICD-10) or simple deteriorative disorder (DSM-IV) represent the most frequently used subtype approach for reducing heterogeneity. Although important differences such as age of onset and pattern of symptom development validate these subtypes, the classical subtypes have not provided a strong heuristic framework for the study of differential etiology and pathophysiology.

In light of the limitations of the classic subtypes, alternative approaches have been sought to reduce syndromal heterogeneity. One approach that has received considerable attention is the proposition that specific symptom complexes define pathological entities that differ from one another in neuroanatomical pathophysiology, in course and onset, in treatment requirements, and possibly in etiology. Interest in the proposition that symptom complexes, or domains of psychopathology, represent unique disease processes has emerged from the extensive study of the longitudinal patterns of symptom manifestations in patients with schizophrenia. A large number of studies have documented that the symptoms of schizophrenia usually segregate into three semi-independent symptom complexes: (1) hallucinations and delusions; (2) disorganized behavior, including positive formal thought disorder, bizarre behavior, and inappropriate affect; and (3) primary, enduring negative or deficit symptoms, including restricted affective experience and expression, diminished drive, and poverty of thought. Longitudinal studies provide

support for the long-term independence and stability of these domains. These results suggest a modification of the central paradigm for the study of the etiopathophysiology and neuroanatomy of schizophrenia: the study of the etiopathophysiology and neuroanatomy of schizophrenia becomes the study of the etiopathophysiology and neuroanatomy of hallucinations and delusions, of disorganized behavior, and of negative symptoms. This approach is also germane for treatment and rehabilitation studies.

The domains approach has been extensively applied to the investigation of deficit symptoms. These symptoms differ from the other symptom complexes in their familial heritability, neuroanatomy as evidenced in both structural and functional neuroimaging studies, and response to antipsychotic treatment. Long-term outcome, season of birth, and age of onset are also distinctive. These results provide strong support for the heuristic value of this approach and raise the hope that this approach to heterogeneity reduction will yield more decisive data in studies of etiopathophysiology and neuroanatomy and will provide explicit information regarding the efficacy profile of pharmacological treatments.

Cognitive Impairment In addition to the three symptom complexes, patients with schizophrenia also manifest a broad array of cognitive impairments, including impaired performance on measures reflecting attention, information processing, executive function, memory, and language capabilities. These manifestations are not used in the diagnosis of schizophrenia, but are a critical component of the disorder. On a theoretical level, attentional and verbal memory impairments are conceptualized as vulnerability markers, which may be useful in defining schizophrenia phenotypes, may be applicable to early detection, and may provide a basis for creating new models for treatment development. Cognitive impairments are also hypothesized to meaningfully determine many aspects of quality of life and functional capacity and adjustment. Moreover, the neuropsychological assessment of cognitive impairments permits probablistic anatomical inferences, and the use of cognitive tasks that assess these impairments has become increasingly important in guiding functional neuroimaging studies.

The relations among cognitive impairments and the symptoms of schizophrenia are unclear. For many years, cognitive impairments were conceptualized as the psychological foundations of symptom manifestations. However, there is a large body of evidence that has documented the relative independence of cognitive impairments and symptoms. For example, clinical trials have repeatedly demonstrated that large changes in symptom status can occur without a corresponding improvement in cognitive function as reflected in cognitive or neuropsychological test performance. Parenthetically, these trials have also revealed the lack of effective pharmacological treatment for these fundamental manifestations of the illness. The use of the three-symptom complex model and less complex cognitive paradigms may lead to the elucidation of possible relationships between the various cognitive impairments and the symptom complexes.

In summary, the manifestations of schizophrenia have been consistently described since the turn of the century. The conceptualization of schizophrenia as a clinical syndrome, importantly distinguished from manic-depressive and other psychoses, has been validated. Research on diagnostic systems has produced modest modifications in classification and has demonstrated the adequacy of the reliability and validity of current approaches. It has also produced a reasonable degree of uniformity in international usage that serves both clinical and scientific purposes. Because the clinical syndrome of schizophrenia probably represents more than one pathological process, specifically addressing the etiology, pathophysiology, and treatment of specific symptom complexes offers important new power to research designs.

COURSE, PROGNOSIS, AND OUTCOME

In his pioneering description of schizophrenia Emil Kraepelin argued that schizophrenia was characterized by an early onset, which was followed by a chronic deteriorating course. Eugen Bleueler suggested that a chronic deteriorating course was a frequent but not a necessary outcome. However, neither of these early workers took into account the extent to which their observations were based on chronic, institutionalized populations. Extensive longitudinal-outcome data on patients who were treated prior to and after the introduction of antipsychotics support a more optimistic prognostic picture. Whereas schizophrenia is always a serious disease, it is now clear that patients suffering from it may follow a variety of courses over the long term, including some that are relatively benign. It remains true that although schizophrenia does not always progress to a deteriorated end state, there are substantial and enduring adverse consequences for most patients.

The course of the illness can be divided into four major epochs: premorbid adjustment, onset of illness, middle course, and late course.

Premorbid Adjustment Premorbid adjustment refers to symptoms that appear prior to the onset of positive symptoms. Twenty-five to fifty percent of patients with schizophrenia have abnormal premorbid adjustment, which may be manifested as poor social and scholastic adjustment or diminished social drive; decreased emotional responsivity; withdrawn, introverted, suspicious, or impulsive behavior; idiosyncratic responses to ordinary events or circumstances; short attention span; and delayed developmental milestones or poor motor and sensorimotor coordination. Childhood asociality, a trait that has been referred to in the past as a poor prognostic indicator, is probably more appropriately conceptualized as the early morbid manifestation of deficit symptomatology. Disturbances in social behavior have been picked up as early as infancy by workers who have noticed a lack of responsiveness and emotional expression in infants who later developed schizophrenia. It is also evident, however, that deficit symptoms may have their onset following psychosis and become part of the progression of the illness during the initial years of psychosis. Subtle forms of positive formal thought disorder may also be manifest before overt hallucinations and delusions occur. Studies that have evaluated the development of the offspring of mothers with schizophrenia have observed cognitive difficulties during the pre-teen and teenage years in these high-risk children.

Onset of Illness The second epoch, onset of illness, typically refers to the onset of positive symptoms (i.e., hallucinations, delusions, and positive formal thought disorder). The onset of positive symptoms is insidious in about half of the patients, with the earliest signs of involvement occurring many years before the appearance of the more blatant manifestations of psychosis. In other cases, onset is relatively sudden or acute, with the onset of positive symptoms marking a sharp deviation in development. Patients with the insidious type of onset are very likely to have a poor intermediate course and a poor long-term outcome. In contrast, patients with normal development and ordinary personality attributes who experience a relatively sudden appearance of hallucinations, delusions, and disorganized thought vary widely in their intermediate and long-term outcomes.

Although the rate of schizophrenia is relatively similar in women and men, there are gender differences in age of onset. In males, the peak age of onset ranges from 17 to 27, whereas females tend to have a wider and flatter window of vulnerability, with age of onset generally between ages 17 and 37.

Middle Course The middle epoch or course of illness may be subdivided into two subepochs. The first 5 to 10 years of illness are frequently characterized by multiple exacerbations of positive symptoms, during which a patient may return to an asymptomatic baseline between episodes, or remain actively psychotic without achieving full recovery. This subepoch is followed by a plateau phase, in which patients experience a stabilization of their symptoms and a decrease in the number of exacerbations. Recent studies have made it evident that the underlying deterioration associated with schizophrenia principally occurs during the onset of illness and the first half of the middle phase, rather than over the remaining course of illness. However, complications caused by the illness lead to ever-increasing impediments to normal existence, so that secondary effects may be progressive even though the primary pathology has plateaued. For example, patients who live in understimulating environments will lose social skills and work capabilities even if their symptoms improve. Effective treatment late in the course of a chronic disease will diminish morbidity, but it will not restore lost experience and opportunity—nor will it overcome stigma. A history of disabling schizophrenia is a serious social and occupational burden regardless of the degree of recovery.

Late Course In the late epoch there is a tendency for the intensity of positive symptoms to diminish, and many patients with long-term impairments regain some degree of social and occupational competence. Although the illness becomes less disruptive and easier to manage, the effects of years of dysfunction are rarely overcome. It would be highly unusual for an individual with a chronic form of the illness to gain the niche in society and the quality of personal life that would have been possible had the illness not been present. More typically, patients continue to manifest direct signs of the illness process throughout their lives. Twenty- to forty-year follow-up studies provide a basis for estimating that approximately 55 percent of patients with schizophrenia have moderately good outcomes and 45 percent have more severe outcomes. These figures are more optimistic than earlier views for at least two reasons. First, sample selection was broader and more representative. Second, effective treatments, which make a considerable difference in the short-term course, also have a modest impact on the long-term course of the illness.

Although no present treatment approach can prevent or cure schizophrenia, some approaches have had remarkable remedial effects on course. Despite not being scientifically verified, there is considerable evidence from a large body of clinical experience that a form of schizophrenia referred to as *devastating schizophrenia*, which represented about 15 percent of the cases before the introduction of antipsychotic medication, now represents only about 5 percent of the cases. This form of the illness had an acute rather than insidious onset, but, paradoxically, had an unrelenting deteriorating course. Another line of evidence suggests that outcome may be related to the time interval between the initial detection of schizophrenia and the initiation of antipsychotic treatment. The more rapidly patients are treated, the more benign is the course of illness. This observation has led to an increased interest in establishing a methodology for early detection and the development of intensive therapeutic interventions, which combine pharmacological and psychosocial treatments, in order to ascertain whether future course can be substantially affected by treatment. There is also considerable evidence suggesting that the prophylactic use of antipsychotic medication reduces the relapse rate by more than

one-half. This fact is largely responsible for making it possible to substantially reduce inpatient care in favor of brief hospital stay, crisis intervention, and community-based treatment. The level of success associated with this major shift in primary treatment setting as well as the serious shortcomings associated with shifting care to unprepared communities are noted in the discussion on treatment and rehabilitation.

Predictors of outcome have been principally found to be related to the already-established pattern of illness, the early developmental pattern, and the emotional qualities of the patient. Patients with limited emotional expression, who demonstrate a lack of social drive and social affiliation during childhood, and who display poor social and occupational functioning in recent years, are quite likely to run a chronic course of the illness. On the other hand, patients who have a normal developmental history with an abrupt onset of psychosis, and who have not established a pattern of social and occupational failure, have a much better prognosis. There is also evidence that prognosis is better in females than males. However, some patients in the good prognosis group will progress to a devastating form of the illness. In general, there are more reliable predictors of poor than good prognosis, with prognosis uncertain until the pattern of illness has been established.

There are interesting results from the World Health Organization study of the social determinants of outcome in different cultures. This work has documented that the course of schizophrenia tends to be more benign in developing countries than in developed countries. This difference in course is generally understood as representing a psychosocial influence on course rather than cultural differences in the causes of schizophrenia. The incidence and lifetime prevalence of the disease appear to be relatively comparable across cultures and societies. One compelling explanation for the observed difference is that the sociocentric structures of developing countries place less demand on individual performance and provide a more broadly supportive interpersonal environment than do the egocentric cultures of more developed nations. With their marked emphasis on individual accomplishment and productivity, the latter nations are more demanding and stressful for those with impaired drive or impaired mental functioning. Rather than finding an appropriate, usually reduced level of functioning, the patient with schizophrenia in Western industrialized societies tends to be isolated, with greatly reduced opportunities for work and meaningful social contacts. Indicative of this lack of involvement, unemployment rates for patients with schizophrenia are over 70 percent in the United States.

TREATMENT AND REHABILITATION

The history of the care and treatment of patients with schizophrenia is replete with instances of both humane and inhumane approaches. From a practical and moral standpoint, the value of humane care is intrinsic and does not rest on scientific evaluation of efficacy. There is a large body of literature and scientific data regarding the pharmacological and psychosocial interventions and the rehabilitation of patients with schizophrenia. The general conclusions of this accumulated information are presented below.

Pharmacological Interventions Prior to 1952, there were no generally applicable, effective pharmacological treatments of schizophrenia. Reserpine had been used with some limited success, and electroconvulsive treatment (ECT) was important in reducing symptoms in the most acutely disturbed cases. This situation changed abruptly with the introduction of chlorpromazine (Thorazine) in France in 1952 and in North America in 1954, which ushered in the modern era of effective pharmacological therapy for schizophrenia.

The antipsychotic drugs used to treat schizophrenia have a wide variety of mechanisms of action, but all share the capacity to occupy postsynaptic dopamine receptors in the brain. Conventional antipsychotics or dopamine-receptor antagonists are often referred to as *neuroleptics* because of their neurological adverse effects. The new antipsychotics are less likely to exhibit these effects and have been referred to as atypical antipsychotics or serotonin-dopamine antagonists. The generally recognized clinical effect of antipsychotic drugs is to diminish symptom expression and reduce relapse rates. Although sedation may be a side effect and diminished anxiety a clinical effect, the primary value of these drugs is their remedial effect on positive symptoms, and not their sedating or tranquilizing properties. In contrast to positive symptoms, conventional antipsychotics have not been shown to be effective for either primary, enduring negative or deficit symptoms or the cognitive impairments observed in patients with schizophrenia.

Antipsychotic drugs are used throughout the world for four primary clinical purposes: (1) to manage acute positive symptomatic disturbances; (2) to induce remission from positive symptom exacerbations; (3) to maintain the achieved clinical effect over prolonged periods of time (maintenance therapy); and (4) to prevent relapses or new episodes of positive symptom expression (prophylactic therapy). A recent emphasis with regard to the use of conventional antipsychotic drugs has been on dose reduction, in the hope of diminishing adverse effects without losing clinical benefit. The intent is to administer the drugs in a manner that will increase patient compliance and avoid illness exacerbations caused by patients' discontinuing their medication.

The first atypical antipsychotic to be available for clinical use was clozapine. Clozapine has a unique mechanism of action, and was shown in the 1970s to have a superior effect on patients resistant to the therapeutic effects of conventional antipsychotics. However, patients on clozapine run an approximately 1 percent risk of agranulocytosis. This potentially lethal cessation in the production of white blood cells was associated with a series of deaths in Finland during the mid-1970s and led to a decreased use of clozapine in Europe and a failure to market the drug in the United States. Interest in clozapine in the United States was rekindled by the results of a large-scale multicenter study in chronic, treatment-resistant patients with schizophrenia. The study yielded convincing evidence of clozapine's effectiveness in ameliorating positive symptoms in approximately one-third of these patients. In addition, the study also showed that clozapine could be used with relative safety within the context of careful monitoring for agranulocytosis. Clozapine represents the first incremental gain in the effectiveness of the pharmacological agents used to treat schizophrenia since the original introduction of chlorpromazine.

The demonstration of clozapine's efficacy for treatment-resistant patients has spawned considerable interest in the development of new pharmacological treatments for schizophrenia. Risperidone (Risperdal), olanzapine (Zyprexa), quetiapine (Serlect), ziprasidone and other compounds have quickly followed. Since the specific mode of clozapine's superior therapeutic efficacy is not known, it is not possible to design new compounds with confidence that they will have superior efficacy. However, each of the new medications appears to be as effective an antipsychotic drug as the conventional antipsychotics, but with a substantially decreased adverse effect burden. This decreased adverse effect burden may result in greater effectiveness based on better patient compliance, and may reduce the incidence of long-lasting adverse effects (e.g., tardive dyskinesia). As each new drug is introduced, it will be important to examine whether the drug shares with clozapine a superior efficacy for positive symptoms. In addition, it will be important to examine if there are any meaningful advantages to using atypical antipsychotic medications in first-episode patients. Finally, future drug development must also recognize the absence of significant therapeutic efficacy of currently available medications for the primary avolitional component of the disease and the fundamental cognitive impairments detected by various psychological performance tasks. The rapidly advancing knowledge of brain biochemistry and brain-behavior relationships has set the stage for the development of new models for psychopathological processes other than positive symptoms. These new models may prove useful for screening potential novel treatments, which would enable a more comprehensive treatment of the varied manifestations of schizophrenia.

Lithium and antiepileptic, antidepressant, and antianxiety drugs have also been used to treat the positive symptoms of schizophrenia. However, these drugs have not proved to be effective alternatives to antipsychotic therapy, nor has there been a consistent demonstration of substantially enhanced benefits when they are used in combination with antipsychotics. A small subgroup of patients may be differentially responsive to a class of drugs other than antipsychotics, but in the absence of the capacity to identify in advance which patients will respond favorably, it is difficult to prove or disprove this proposition. In contrast, these drugs and a series of medications that counteract the side effects of conventional antipsychotics have been effective for co-occurring anxiety and depressive, manic, and aggressive symptoms.

Augmentation strategies have also been used for persistent negative symptoms, including deficit symptoms. These strategies have included the use of dopamine and serotonergic and noradrenergic agents. Perhaps the most promising development is the attempt to treat these symptoms by activating the NMDA receptor at the glycine site. Glycine and d-cycloserine have produced encouraging results in preliminary controlled clinical trials.

ECT was frequently used to treat patients with schizophrenia prior to the introduction of antipsychotic drugs. ECT is particularly effective in the treatment of catatonic stupor and excitement, but generally produces results similar to those obtained with antipsychotics, (i.e., a reduction of positive symptoms rather than a reversal of long-term functional impairments). Although ECT is safe and painless, its use is restricted, in part by litigation and societal attitudes, but also because any therapeutic advantage gained in an initial series of treatments is not easily maintained. Also, there is no indication that ECT is effective in patients who are resistant to conventional antipsychotics. For all of these reasons, drug treatment approaches are generally preferred.

Psychosocial Interventions The debate over whether patients should be administered pharmacological agents or psychosocial treatments has given way to the search for how these treatments should be optimally integrated. Controlled clinical trials have conclusively demonstrated that intensive psychotherapy is less effective than drug treatment; that it is not superior to less expensive, less ambitious psychosocial forms of psychotherapy; and that it should no longer be considered an alternative to the use of antipsychotic drugs. In addition, studies have repeatedly demonstrated that supportive forms of psychosocial treatment are entirely compatible with drug treatment and can increase the effectiveness of overall treatment, reduce the amount of medication necessary, enhance patient participation in

the full range of treatment, and optimize social and occupational functioning. Especially impressive are studies documenting the considerable additional benefit achieved in reducing relapse and hospitalization rates when family therapy and education programs are added to maintenance pharmacological treatment. These studies make it clear that psychosocial and rehabilitative interventions have become essential components of the comprehensive treatment of patients with schizophrenia.

Psychosocial and rehabilitation interventions include supportive, problem-solving, educationally oriented psychotherapy; family therapy and education programs aimed at helping patients and their families understand the patient's illness, reduce stress, and enhance coping capabilities; social and living skills training; vocational training, including job coaching; and the provision of supervised residential living arrangements. The development and increased utilization of psychosocial services have been complemented by the evolution of services designed to decrease the utilization of inpatient hospital services and to maintain the patient in the community. Assertive community treatment teams are designed to provide intensive outreach services to patients who are unable to be maintained in the community with traditional outpatient clinical treatment. Crisis management services, including 24-hour crisis beds and partial hospitalization programs, represent alternatives to hospitalization during periods of symptom exacerbation.

The development of these services reflects the ongoing shift in the treatment of the patient with schizophrenia from a hospital-based to a community-based system of care. When optimal treatment with these services is provided, the rewards of therapeutic accomplishment, reduction in morbidity, and economic cost benefits are profound and rival therapeutic accomplishments found anywhere in medicine. The demonstrated benefits of these services challenge the field to establish an adequate community-based treatment approach prepared to meet the challenge and demands of broad-based integrated treatment.

FUTURE DIRECTIONS

The care and study of the person afflicted with schizophrenia are extraordinarily interesting and promising. Basic brain science has matured, and technological advances permit increasingly sophisticated questions to be addressed regarding the anatomy, ultrastructure, and function of the brain. The field is closer to understanding risk factors at the level of causal mechanism, and new treatments are being developed at an increasing rate. The cadre of schizophrenologists capable of integrating basic and clinical sciences has grown substantially, and new paradigms providing heuristic advantage in the classification of psychopathological phenomena provide and address the problem of heterogeneity, which has undermined so much of the investigative work in schizophrenia. Physiological markers have been validated, and investigators are able to articulate, with ever-increasing specificity, the what and where of brain dysfunction in patients with schizophrenia.

The twenty-first century promises to be a time of fundamental discovery regarding the etiology and pathophysiology of what may be the world's most vexing public health problem. These developments have emerged at a time of decreasing stigma, increasing partnership in clinical care and research with citizen advocacy groups, and the initiation of nationwide private fundraising for research on this disease.

SUGGESTED CROSS-REFERENCES

More detailed discussions of etiology, brain structure and function, clinical features, and somatic and psychosocial treatments are presented in the other sections of Chapter 12. A detailed introduction to areas of neuroscience and cognitive science relevant to schizophrenia is provided in Section 1.2 on functional neuroanatomy, Section 1.3 on neuronal development and plasticity, Section 1.15 and Section 1.16 on brain imaging, and Section 3.1 on perception and cognition.

SECTION REFERENCES

Andreasen NC, Arndt S, Swayze V II, Cizadlo T, Flaum M, O'Leary D, Ehrhardt JC, Yuh WTC: Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. Science *266:*294, 1994.

*Andreasen NC, O'Leary DS, Cizadlo T, Arndt S, Rezai K, Boles Ponto LL, Watkins G ll, Hichwa RD: Schizophrenia and cognitive dysmetria: A positron-emission tomography study of dysfunctional prefrontal-thalamic-cerebellar circuitry. Proc Natl Acad Sci *93:*9985, 1996.

Braff DL: Information processing and attention dysfunctions in schizophrenia. Schizophr Bull *19:*233, 1993.

Breier A, Su T-P, Saunders R, Carson RE, Kolachana BS, De Bartolomeis A, Weinberger DR, Weisenfeld N, Malhotra AK, Eckelman WD, Pickar D: Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: Evidence from a noval positron emission tomography method. Proc Natl Acad Sci *94:*2569, 1997.

Buchanan RW, Brandes M, Breier A: Treating negative symptoms: Pharmacological strategies. In *The New Pharmacotherapy of Schizophrenia*, vol 36, A Breier, editor. American Psychiatric Press, Washington, DC, 1996.

Carlsson A: Neurocircuitries and neurotransmitter interactions in schizophrenia. Int Clin Psychopharmacol *3:*21, 1995.

Carpenter WT, Buchanan RW: Schizophrenia. N Engl J Med *330:*681, 1994.

*Carpenter WT, Buchanan RW, Kirkpatrick B, Tamminga C, Wood F: Strong inference, theory testing, and the neuroanatomy of schizophrenia. Arch Gen Psychiatry *50:*825, 1993.

Cohen JD, Servan-Schreiber D: A theory of dopamine function and its role in cognitive deficits in schizophrenia. Schizophr Bull *19:*85, 1993.

Conley RR, Buchanan RW: Evaluation of treatment-resistant schizophrenia. Schizophr Bull *23:*663, 1997.

*Freedman R, Coon H, Myles-Worsley M, Orr-Urtreger A, Olincy A, Davis A, Polymeropoulos M, Holik J, Hopkins J, Hoff M, Rosenthal J, Waldo MC, Reimherr F, Wender P, Yaw J, Young DA, Breese CR, Adams C, Patterson D, Adler LE, Kruglyak L, Leonard S, Byerley W: Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci *94:*587, 1997.

*Goff DC, Tsai G, Levitt J, Amico E, Manoach D, Schoenfeld DA, Hayden DL, McCarley R, Coyle JT: A placebo-controlled trial of D-cycloserine added to conventional neuroleptics in patients with schizophrenia. Arch Gen Psychiatry *56:*21, 1999.

*Goff DC, Wine L. Glutamate in schizophrenia: Clinical and research implications. Schizophr Res *27:*157, 1997.

*Heresco-Levy U, Javitt DC, Ermilov M, Mordel C, Silipo G, Lichenstein M: Efficacy of high-dose glycine in the treatment of enduring negative symptoms of schizophrenia. Arch Gen Psychiatry *56:*29, 1999.

Howells JG, editor: *The Concept of Schizophrenia: Historical Perspectives*. American Psychiatric Press, Washington, DC, 1991.

Kane JM: Drug therapy: Schizophrenia. N Engl J Med *334:*34, 1996.

Kendler KS, Straub RE, MacLean CJ, Walsh D: Reflections on the evidence for a vulnerability locus for schizophrenia on chromosome 6p24-22. Am J Med Genet (Neuropsychiatr Genet) *67:*124, 1996.

Lawrie SM, Abukmeil SS: Brain abnormality in schizophrenia. Br J Psychiatry *172:*110, 1998.

*Mortensen PB, Pedersen CB, Westergaard T, Wohlfahrt J, Ewald H, Mors O, Andersen PK, Melbye M: Effects of family history and place and season of birth on the risk of schizophrenia. N Engl J Med *340:*603, 1999.

Olney JW, Farber NB: Glutamate receptor dysfunction and schizophrenia. Arch Gen Psychiatry *52:*998, 1995.

Schooler NR, Keith SJ, Severe JB, Matthews SM, Bellack AS, Glick ID, Hargreaves WA, Kane JM, Ninan PT, Frances A, Jacobs M, Lieberman JA, Mance R, Simpson GM, Woerner MG: Relapse and rehospitalization during maintenance treatment of schizophrenia. Arch Gen Psychiatry *54:*453, 1997.

*Silbersweig DA, Stern E, Frith C, Cahill C, Holmes A, Grootoonk S, Seaward J, McKenna P, Chua SE, Schnoor L, Jones T, Frackowiak RSJ: A functional neuroanatomy of hallucinations in schizophrenia. Nature *378:*176, 1995.

Susser E, Neugebauer R, Hoek HW, Brown AS, Lin S, Labovitz D, Gorman JM: Schizophrenia after prenatal famine: Further evidence. Arch Gen Psychiatry *53:*25, 1996.

Weinberger DR: From neuropathology to neurodevelopment. Lancet *346:*552, 1995.

12.2 SCHIZOPHRENIA: EPIDEMIOLOGY

GRAYSON S. NORQUIST, M.D., M.S.P.H., AND WILLIAM E. NARROW, M.D., M.P.H.

An etiologically puzzling and clinically severe disorder, schizophrenia has long held the attention of psychiatric epidemiologists. Until fairly recently, however, efforts to elucidate the epidemiology of schizophrenia were hindered by two critical deficiencies: the lack of reliable case definition and limited generalizability of findings because of reliance on treated samples. Two lines of scientific development have converged since the late 1970s to help resolve these issues and advance epidemiological studies of schizophrenia. First was the development of *Diagnostic and Statistical Manual of Mental Disorders* (DSM), culminating in the publication of operationalized criteria for schizophrenia in the third edition of DSM (DSM-III). Second, modern epidemiological methods, particularly those of chronic disease and genetic epidemiology, have been applied to the study of schizophrenia. These methods have increased the robustness of epidemiological research findings because of their emphasis on precise study designs, representative sampling, and sophisticated techniques of data analysis.

Psychiatric epidemiology is traditionally concerned with patterns of psychopathology in human population groups and the factors that influence these patterns. It examines the occurrence of pathology in terms of time, place, and individual characteristics, in order to elucidate the etiology of illness and its population burden. In terms of schizophrenia, such research includes studies of prevalence and incidence, natural history of illness including risk and protective factors for onset, remission, and relapse; longitudinal followup of populations at high risk for schizophrenia, including children of parents with schizophrenia or relatives of a proband with schizophrenia; and genetic epidemiology, including twin, family, association, and linkage studies in samples that are representative of the persons with schizophrenia, an associated marker of interest, or a population isolate with a high prevalence of the disorder.

In epidemiology, a *population* is a collection of individuals defined by time, place, and characteristics such as age, sex, and race. Although general community populations are often studied, epidemiological study populations may be defined in other ways, including treatment status or exposure to a risk factor. For rare disorders such as schizophrenia, it is often easier to sample study cases from treated populations. However, by excluding untreated cases, the findings are not generalizable to all individuals with schizophrenia. *Point prevalence* is defined as the number of persons in a population who are affected with a disorder at a given point in time. *Incidence* is defined as the number of persons without a disorder at the beginning of a given time period who subsequently develop the disorder in that time period. A “first” or “true” incident case has never had a previous episode of disorder; a *recurrent* case has had a previous episode. *Period prevalence* includes existing cases at the beginning of a given time period (point prevalence), plus all incident cases developing in the time period, both first incidence and recurrence. An important concept in epidemiology is that prevalence is proportional to incidence and duration ($P \sim I \times d$). Thus, in a chronic condition such as schizophrenia, a steady prevalence is maintained by a long duration of illness despite a relatively low incidence rate; cutting short the duration of illness would decrease the prevalence of the condition if incidence remained unchanged.

EVOLUTION OF NOMENCLATURE AND DIAGNOSTIC CRITERIA

The value of operationalized diagnostic criteria as a commonly accepted language for clinicians and researchers cannot be overemphasized. In the United States the DSM system of classification has become the diagnostic standard for both clinical and research purposes.

DSM In 1952 the first edition of DSM (DSM-I) was published by the Mental Hospital Service of the American Psychiatric Association. DSM-I was derived largely from the section on mental disorders in the Standard Classified Nomenclature of Disease developed by the National Conference on Nomenclature of Disease in 1933. In DSM-I, “Schizophrenic Reactions” were classified under “Disorders of Psychogenic Origin or without Clearly Defined Physical Cause or Structural Change in the Brain.” and were described as “synonymous with formerly used term dementia praecox.”

It represents a group of psychotic reactions characterized by fundamental disturbances in reality relationships and concept formations, with affective, behavioral, and intellectual disturbances in varying degrees and mixtures. The disorders are marked by strong tendency to retreat from reality, by emotional disharmony, unpredictable disturbances in stream of thought, regressive behavior, and in some, by a tendency to “deterioration.” The predominant symptomatology will be the determining factor in classifying such patients into types.

The types of schizophrenic reactions were simple, hebephrenic, catatonic, paranoid, acute undifferentiated, chronic undifferentiated, schizo-affective, childhood, and residual.

DSM-II The second edition of DSM (DSM-II) was published in 1968 and was an attempt to achieve uniformity of diagnostic classification at an international level. The nomenclature used in DSM-II was based, with a few exceptions, on the terms used in the eighth revision of the World Health Organization's (WHO's) *International Statistical Classification of Diseases, Injuries, and Causes of Death* (ICD-8). DSM-II had initial attempts to remove from the diagnostic nomenclature implications about the nature or cause of a disorder. Thus, the “schizophrenic reaction” of DSM-I became “schizophrenia” in DSM-II. Attempts were also made to distinguish schizophrenia from the psychotic mood disorders. The term *schizophrenia* was defined as:

> . . . a group of disorders manifested by characteristic disturbances in thinking, mood, and behavior. Disturbances in thinking are marked by alterations of concept formation which may lead to misinterpretation of reality and sometimes to delusions and hallucinations, which frequently appear psychologically self-protective. Corollary mood changes include ambivalent, constricted, and inappropriate emotional responsiveness and loss of empathy with others. Behavior may be withdrawn, regressive and bizarre. The schizophrenias, in which the mental status is attributable primarily to a thought disorder, are to be distinguished from the Major affective illnesses which are dominated by a mood disorder. The paranoid states are distinguished from schizophrenia by the narrowness of their distortions of reality and by the absence of other psychotic symptoms.

Several subtypes remained unchanged in DSM-II: simple, hebephrenic, paranoid, childhood, and residual. Catatonic type was divided into excited and withdrawn subtypes, and schizoaffective type was divided into excited and depressed subtypes. Acute undifferentiated type was renamed acute schizophrenic episode. Latent type was distinguished from the DSM-I chronic undifferentiated type in order to cover patients unofficially diagnosed as having incipient, prepsychotic, pseudoneurotic, pseudopsychopathic, or borderline schizophrenia.

DSM-III and DSM-III-R The publication of the third edition of DSM (DSM-III) in 1980 represented a revolutionary advance in the development of a common diagnostic language for clinicians and researchers. It offered explicit criteria for diagnosing disorders based on observable signs and symptoms, rather than the earlier prose definitions that had a rather wide latitude of clinical interpretation. Beginning with DSM-III and continuing through the revised third edition of DSM (DSM-III-R) and the fourth edition (DSM-IV), the criteria for schizophrenia have followed a general pattern. These are: (1) a listing of the psychotic symptoms, of which one or two are required to be present; (2) a requirement for decline in social functioning and self-care; (3) “exclusion criteria” in which other disorders must be ruled out before assigning a diagnosis of schizophrenia; and (4) a duration and course criterion. Thus, in DSM-III, *schizophrenic disorder* was defined by the presence of six criteria: Criterion A required the presence of one of a list of six psychotic symptoms: three having to do with delusions; two with auditory hallucinations; and one with thought disorder associated with affective disturbance, delusions, hallucinations, or catatonic or grossly disorganized behavior. Criterion B required a deterioration in functioning from a previous level. Criterion C required a 6-month duration of illness with an active phase that included criterion A, and additional prodromal and residual symptoms, which were listed. Criterion D excluded persons for whom psychotic symptoms were preceded by a manic or depressive syndrome or for whom the mood syndrome was not “brief” in relation to the duration of the psychotic syndrome. Criterion E restricted age of onset to under 45 years. Criterion F ruled out syndromes that were “due to” any organic mental disorder or mental retardation. In DSM-III, catatonic, paranoid, and residual types were maintained. Hebephrenic

type was renamed *disorganized type*. The term *chronic* was removed from *chronic undifferentiated type*. Schizoaffective type was removed from the schizophrenic disorders, renamed *schizoaffective disorder* and placed in the “Psychotic Disorders not Elsewhere Classified” category; unlike schizophrenia, no criteria were provided for the diagnosis of schizoaffective disorder in DSM-III. Another innovation in DSM-III was its classification of course of illness into categories of subchronic, chronic, subchronic with acute exacerbation, chronic with acute exacerbation, and in remission.

DSM-III-R, published in 1987, contained several changes in criteria for schizophrenia (no longer “schizophrenic disorder”). The various types of delusions specified in DSM-III criterion A were simplified—DSM-III-R distinguished only bizarre and nonbizarre delusions. In addition criterion A symptoms were expanded to allow nonauditory hallucinations. A minimum duration requirement of 1 week for criterion A symptoms was set. Criterion B was clarified, with a comparison point of the person's highest-ever level of functioning. Criterion C in DSM-III became criterion D in DSM-III-R; little was changed in this criterion except for the addition of “marked lack of initiative, interests, or energy” as a prodromal or residual symptom. Criterion D, the mood disorder exclusion, became criterion C in DSM-III-R; now if a mood syndrome was ever present during an active phase of the illness, schizophrenia was not diagnosed. Significantly, criterion E, the age requirement, was dropped in DSM-III-R. The organic mental disorder exclusion remained in DSM-III-R, and a new criterion was added that dealt with the comorbidity of autistic disorder and schizophrenia.

DSM-IV Few changes were made to the diagnosis of schizophrenia in DSM-IV. Most significantly for case identification, the duration requirement for criterion A symptoms was increased to 1 month. In Criterion A, the concept of negative symptoms was added and the concept of loosening of associations was dropped and replaced with “disorganized speech.” Specific criteria for prodromal and residual phases were dropped. Modifications were made to the organic disorder exclusion to include direct physiological effects of a substance, and the autistic disorder relationship was expanded to include all pervasive developmental disorders. No additional schizophrenia types were added in DSM-IV.

DIAGNOSTIC ASSESSMENT INSTRUMENTS

Identification of persons with mental disorders in the community, regardless of treatment status or severity of disorder, is the ultimate test of a diagnostic classification system. In order to be useful for service planning and research needs, a credible diagnostic classification system must be able to detect and correctly classify untreated cases and cases on the threshold of diagnosis. Unfortunately, for many years the lack of reliably operationalized diagnostic criteria hindered the ability of epidemiologists to identify cases in the community, and prevalence rates were usually based on treated cases only.

Diagnostic Interview Schedule In the 1970s several psychiatric interviews were developed: the Schedule for Affective Disorders and Schizophrenia (SADS), a clinical interview based on Research Diagnostic Criteria (RDC); the Present State Examination (PSE), a nondiagnostic clinical interview that covered present symptoms only; the Psychiatric Epidemiological Research Interview (PERI), another nondiagnostic interview; and the Renard Diagnostic Interview, a diagnostic interview based on Feighner diagnostic criteria. The development of DSM-III in 1980 spurred renewed interest in gathering community epidemiological data based on the new criteria, but none of these interviews were entirely suitable for the purpose. To respond to this need, the Diagnostic Interview Schedule (DIS) was developed for use in the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area (ECA) Program. The DIS covered DSM-III criteria for schizophrenia and schizophreniform disorder; Research Diagnostic Criteria for schizophrenia and schizoaffective disorder, manic and depressed types; and Feighner criteria for schizophrenia. It was a highly structured interview, based on respondent self-report only, which did not allow interviewer discretion in the administration of questions or recording of responses and did not rely on other clinical information to make diagnoses. Diagnoses were generated by computer algorithm, not by the interviewer. Because of these factors, the DIS was suitable for administration by nonclinician interviewers and could be administered in large surveys for relatively low cost. The DIS has undergone subsequent revisions to incorporate DSM-III-R and DSM-IV diagnostic criteria and it has been translated into over twenty other languages. Various reliability and validity studies performed on the DIS have demonstrated inconsistent results for schizophrenia; in such studies, results are often within an acceptable range if corrected for the low prevalence of schizophrenia in the population.

CIDI With the evolution of the DSM and the International Classification of Diseases (ICD) diagnostic systems, the need for a comprehensive diagnostic instrument for use in cross-cultural and comparative studies worldwide became apparent. To address this need, the Composite International Diagnostic Instrument (CIDI) was developed as a joint venture of WHO and the U.S. Alcohol, Drug Abuse, and Mental Health Administration. The structured question-and-probe structure of the DIS served as the template for the construction of the CIDI. Questions from the PSE, an interview widely used in epidemiological studies outside the United States were then added. Updates of the CIDI added criteria from the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), and then DSM-III-R and DSM-IV criteria. The CIDI, like the DIS, remains a highly structured diagnostic interview with diagnoses made via computer algorithm. Throughout its development, efforts were made to ensure its cross-cultural appropriateness.

COMMUNITY SURVEYS

Three community surveys are most frequently cited for data on the prevalence and incidence of schizophrenia.

ECA NIMH Epidemiologic Catchment Area Program The ECA NIMH Program is the largest community survey of mental disorders ever undertaken in the United States. A total of 18,571 household residents and 2290 institutional residents (of nursing homes, jails, psychiatric hospitals) age 18 and over were sampled and interviewed in five areas: New Haven, Baltimore, Durham, St. Louis, and Los Angeles. Two face-to-face interviews were done 12 months apart (Wave I and Wave II). A telephone interview (face-to-face in New Haven) of the household respondents was conducted 6 months after Wave I. The institutional residents were interviewed in Waves I and II only; no telephone interview was conducted for these respondents. DSM-III diagnostic data were obtained at Waves I and II using the DIS. Respondents were asked about their use of health services at each wave. Questions to ECA respondents pertaining to use of mental health services covered use of ambulatory specialty mental and addictive, general medical, and human services, and admissions to hospitals and residential treatment centers for reasons related to mental health or addictions.

National Comorbidity Survey Following the success of the ECA Program, the National Comorbidity Survey (NCS) was conducted in the early 1990s to obtain more detailed information about mental disorders in the community, particularly the relation between co-occurring mental disorders and co-occurring mental and substance use disorders. Conducted by the Institute for Social Research at the University of Michigan, the study used a nationally representative household sample of 15- to 54-year-olds. A modified version of the CIDI (the UM CIDI) for DSM-III-R was used in the NCS. For psychotic disorders, in addition to a CIDI computerized diagnosis, a clinical reinterview was conducted with individuals who screened positive for psychosis on the CIDI. Results pertaining to the schizophrenia and related disorders are usually presented in a summary category called *nonaffective psychosis*, which is made up of schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, and atypical psychosis. Comparisons of the results of the ECA to the NCS are not straightforward because of the differences in the diagnostic instrument, differences in age range of respondents, and temporal differences, among other factors.

WHO Collaborative Study on the Determinants of Outcome of Severe Mental Disorders (DOS) This study was conducted at 12 field centers in 10 countries across the world. At each of the 12 centers, all persons in a catchment area making first contact with a psychiatric, medical, or other agency for symptoms of possible schizophrenia were identified, assessed, and followed for 2 years. Like the NCS, the study selected only individuals between the ages of 15 and 54. The assessment instrument was the PSE. The final cohort size was 1379. Incidence of schizophrenia was obtained in 7 sites and clinical information on diagnosis and course was obtained at all 12 sites.

PREVALENCE

A 1987 review of over 70 prevalence studies of schizophrenia published since 1948 identified point prevalence in various population groups ranging from 0.06 percent to 1.7 percent, with lower rates in developing countries. It was suggested that this difference was not entirely caused by differences in diagnostic procedures and study methods. Rather, it was posited that higher recovery rates in developing countries and etiological heterogeneity, among other factors, could account for at least a tenfold difference in prevalence.

The prevalence of schizophrenia in the ECA is presented in Table 12.2-1. According to convention in the presentation of ECA results, these figures represent the combined prevalence of schizophrenia and schizophreniform disorders. One-month prevalence is conventionally viewed as “current” prevalence, so based on the ECA results, 0.7 percent of the adult population currently has a diagnosis of schizophrenia. The 1-year prevalence is similar, whether measured retrospectively or prospectively. The prospective period prevalence illustrates the chronicity of schizophrenia: more than twice as many persons had the disorder at the start of the year (0.7 percent) as developed the disorder or had a relapse during the following year (0.3 percent).

Length of Time	Prevalence (%) (SE)
1-month	0.7 (0.1)
1-year	1.0 (0.1)
Lifetime	1.5 (0.1)
Prospective	
1-month	0.7 (0.1)
1-year new	0.3 (0.1)
1-year total	1.1 (0.1)

SE, standard error

Table 12.2-1 Prevalence of Schizophrenia Disorders in the NIMH Epidemiologic Catchment Area Program

Based on the UM CIDI DSM-III-R computer diagnosis, the lifetime prevalence of “narrowly defined psychotic illness” (schizophrenia or schizophreniform disorder) in the National Comorbidity Survey was 1.3 ± 0.2 percent, very close to the ECA lifetime prevalence. Based on clinician diagnoses, the lifetime prevalence of narrowly defined psychotic illness in the NCS dropped to 0.16 ± 0.06 percent.

In examining the sociodemographic correlates in Table 12.2-2, it is helpful to keep in mind that an unadjusted prevalence represents the prevalence of schizophrenia as it actually appears in the community. The odds ratio tells us whether this rate is different when corrected for differences in age, sex, race, marital status, and socioeconomic status among the groups. For example, the prevalence of schizophrenia among blacks is 1.2 ± 0.2 percent, twice as high as for non-black, non-Hispanic persons. However, when correcting for differences in age, sex, race, and socioeconomic status between the ethnic groups, the odds of having schizophrenia among blacks (0.86) is actually no different than the odds for non-black, non-Hispanic persons (whose odds ratio as the comparison for this group is set at 1.00). It should also be noted that despite the size of the ECA survey, relatively few cases of schizophrenia were detected so there is limited power to detect significant differences in prevalence.

Correlates	Unadjusted Rate, % (SE)	Odds Ratio*
Age		
18–24	0.8 (0.2)	—
25–44	1.1 (0.2)	2.31
45–64	0.5 (0.1)	0.76
≥65	0.1 (0.1)	0.10†
Sex		
Male	0.7 (0.1)	—
Female	0.7 (0.1)	0.80
Race or ethnicity		
Black	1.2 (0.2)	0.86
Hispanic	0.4 (0.2)	0.57
Non-Black and non-Hispanic	0.6 (0.1)	—
Marital status		
Married	0.5 (0.1)	—
Single	1.1 (0.2)	2.55
Separated or divorced	1.5 (0.3)	2.61†
Widowed	0.4 (0.2)	2.27
Socioeconomic status		
1 (high)	0.3 (0.1)	—
2	0.6 (0.2)	2.28
3	0.9 (0.1)	3.96
4 (low)	1.2 (0.2)	8.14†

* Odds ratios adjust for age, sex, race or ethnicity, marital status, socioeconomic status, and ECA site.
† $P = 0.0011$ Bonferroni-corrected P value of $= 0.055$
SE, standard error

Table 12.2-2 Sociodemographic Correlates of 1-Month Prevalence of Schizophrenia in the NIMH ECA Program

Persons over age 65 were significantly less likely than persons between 18 to 24 years of age to have a diagnosis of schizophrenia. Persons aged 25 to 44 were twice as likely to have the diagnosis, although this difference did not reach statistical significance. The sex ratio was relatively even. Although blacks had twice the rate of schizophrenia in the community compared to other ethnic groups, when adjusted for socioeconomic status, age, and other factors, this difference disappeared. Being unmarried, particularly separated or divorced, was associated with a diagnosis of schizophrenia. There was an eightfold increase in odds of having schizophrenia in the lowest socioeconomic quartile compared to the highest quartile.

Comorbidity of schizophrenic disorders with substance use disorders in the ECA Program is shown in Table 12.2-3. Both the ECA and the NCS have demonstrated extremely high lifetime comorbidity of the psychotic disorders and substance use disorders; 50 to 60 percent of persons with schizophrenia or nonaffective psychosis had a comorbid alcohol or drug use diagnosis. Comorbidity in the NCS was higher, which may be due to methodological differences as well as secular changes in drug and alcohol use in the United States.

Comorbid Disorder	Rate (%) (SE)
Alcohol dependence	24.0
Any alcohol use disorder	33.7
Drug dependence	12.9
Any drug use disorder	27.5
Any drug or alcohol use disorder	47.0

SE, standard error

Table 12.2-3 Comorbidity of Schizophrenia With Substance Use Disorders in the ECA Program

INCIDENCE

In the WHO DOS, incidence (based on service contacts) ranged between 0.016 and 0.042 percent per year across the sites for broadly defined schizophrenia. For narrowly defined schizophrenia, incidence showed less variation, ranging from 0.007 to 0.014 percent, which was not a statistically significant difference. Age- and sex-specific incidence showed a tendency toward earlier onset in males, a consistent finding in both developing and developed countries. Overall disease expectancy was virtually the same for males and females across the overall age range of 15 to 54 years. Because the DOS minimized methods variance through standardized case ascertainment and assessment methods, the consistency in the incidence of narrowly defined schizophrenia across several international sites is a noteworthy finding. The study's investigators raised fundamental questions in the study of risk factors for schizophrenia such as the possibility of a widely distributed genetic liability, the roles of ubiquitous and culturally specific environmental factors in interacting with that liability, and alternatively, the possibility of multiple genetic liabilities with remarkably similar phenotypic expression in various population groups.

Several studies report that the incidence of schizophrenia is on the decline. Interpretation of these findings is not easy, requiring as it does an accounting of the rapid changes in psychiatry that have occurred in the past 30 years, as well as in social policy and in population demographics. For example, incidence based on inpatient admissions, once widely accepted, is no longer a valid indicator of illness onset because of the shift of treatment away from institutionalization. Diagnostic systems have changed dramatically and these changes are likely to be reflected in different estimates. As the populations of developed countries continue to get older, fewer citizens will be at risk for developing schizophrenia. The issue of declining incidence has not yet been satisfactorily resolved and controversy remains as to what degree the findings represent a true decline in new cases versus a methodological artifact.

USE OF SERVICES

In the ECA Program about 64 percent of persons with a current diagnosis of schizophrenia used some form of mental health service in a 1-year period, a relatively high treated percentage compared to the other disorders surveyed and similar to bipolar I and II disorder and somatization disorder. About 17 percent received inpatient treatment at some point in the year. Consistent with the clinical picture of schizophrenia, most treated individuals had contact in a specialty or general medical setting. Among all ECA disorders, individuals with schizophrenia had the highest proportion (46 percent) treated in the specialty sector, and the bulk of treatment visits were made to this sector as well. Despite the large number of people who had visited the general medical sector, relatively few visits were made there. About 14 percent of persons with schizophrenia were seen only in the general medical sector.

RISK FACTORS

The term *risk* refers to the likelihood that a person who does not currently have schizophrenia will develop the disorder after exposure to certain factors. Thus, a *risk factor* for schizophrenia is an inherent or acquired characteristic or an external condition associated with an increased probability of developing schizophrenia. Epidemiological studies in schizophrenia seek to determine the most important risk factors for this disorder.

The concept of risk can be expressed in several ways. The most common is a report of the absolute number of new schizophrenia cases detected in a population exposed to a postulated risk factor. The terms *relative risk (risk ratio)* and *risk difference (attributable risk)*—expressions of the relationship of the incidence in those exposed to the risk factor to that of those not exposed—are also often used. In case control studies, if the disorder is rare, the risk ratio is approximated by an odds ratio.

Significant risk factors are identified through the use of several different study designs. One type, cross-sectional studies, reports descriptive data at a defined point in time, such as the increased presence of a particular factor in a population with a higher prevalence of schizophrenia. Case control studies compare schizophrenia cases with unaffected controls and determine whether those who express the disease were exposed to a given risk factor. The most informative (and expensive) study design is the prospective cohort study, which follows a group over time to determine whether those exposed to certain risk factors have a higher incidence of schizophrenia.

Risk factors are categorized in several different ways: demographic and concomitant factors (such as age, sex, race, social class), precipitating factors that operate immediately before the onset of schizophrenia (such as life events, migration), and predisposing factors that act for a long period of time or during an earlier part of life (such as genes, perinatal complications, infections). Another schema describes risk factors as either familial influences or sociodemographic factors. The latter can be further subdivided into mutable factors (such as social class, marital status, immigration) and immutable ones (such as ethnic group, sex, birthplace); mutable sociodemographic factors could be a result and not a cause of the disease. This latter distinction is particularly difficult to disentangle in cross-sectional studies.

Several cautions are necessary before reports from studies of risk factors for schizophrenia are evaluated. First, a high prevalence of schizophrenia in a particular area may be the result of protracted illness rather than an increased incidence of schizophrenia (i.e., prevalence is roughly equal to incidence × duration). Second, studies that report only the prevalence of schizophrenia may have failed to control other confounding factors, such as socioeconomic status, that might increase prevalence. Third, designating something as a risk factor does not imply that everyone exposed to it is at personal risk of developing schizophrenia. It means that the group of people exposed to the risk factor at some time are likely to show a higher incidence of schizophrenia than a similar group who were not exposed. Risk does not prove causation but rather an association between that risk factor and the development of schizophrenia. Fourth, schizophrenia may be an etiologically heterogeneous disorder involving many risk factors and many protective factors. Earlier studies or risk factors have many methodological problems, the most important being the failure to standardize diagnostic criteria for selection of schizophrenia cases. However, those studies have helped in the continued search to understand this complicated disorder.

Genetic Factors Identification of a genetic influence is a major challenge in the understanding of schizophrenia. The search for a genetic risk factor has been examined through studies of twins, of families, and of adopted-away children of parents with schizophrenia. Twin studies have shown a concordance of 33 to 78 percent among monozygotic twins, but of only 8 to 28 percent in dizygotic twins. Those results may be affected by selection bias if monozygotic twins are more likely to come to the attention of researchers than are dizygotic twins. Also, monozygotic twins may have greater environmental similarity. Family studies reveal that first-degree relatives of a person with schizophrenia have approximately a fivefold to tenfold chance of developing schizophrenia than nonrelatives. Children have about a 35 percent greater chance of schizophrenia if both parents have schizophrenia compared with about a 1 percent lifetime risk if neither parent has schizophrenia. Although the results from family studies are thought to indicate genetic influences, similar environmental factors among relatives cannot be discounted. Adoption studies are conducted in an effort to control environmental influences. Those studies show that the adopted-away offspring of persons with schizophrenia are at increased risk for schizophrenia and schizophrenia-spectrum disorders. More recent studies using narrower, criterion-based definitions of schizophrenia have reported risk figures that are lower than those reported in earlier studies.

Although there are methodological problems with all three study approaches, findings from them suggest some type of genetic influence in schizophrenia, the significance of which has yet to be delineated. Likewise, the mode of transmission has not been found. Recent efforts have focused on linkage analyses and attempts to locate specific genes. Genetic and environmental factors play a role in the development of schizophrenia, and further refinement in methodology should help to identify the environmental and genetic components of schizophrenia.

Ethnicity and Racial Factors Several studies have discovered differences in the prevalence and number of new cases of schizophrenia among various ethnic and racial groups. The findings are not consistent and may result from failure to control for confounding factors such as social class, age, sex, and immigration status. Data from the NIMH ECA study confirm that if potential confounding factors such as socioeconomic status are controlled, the difference in prevalence across races disappears.

Previous studies of different geographical areas have found a higher prevalence and a larger number of new cases in different countries (e.g., Ireland) and within countries (e.g., the Istrian peninsula of Yugoslavia). Most studies comparing geographical areas are usually flawed because they fail to validate diagnostic methods in different ethnic groups and localities. The WHO Determinants of Outcome study reported that the incidence of schizophrenia is similar in various cultures, especially when a restricted definition is used. If true differences in incidence can be shown, perhaps differences in environmental characteristics, genetic characteristics, or both, can be found in these areas.

Age Early studies showed mean ages of onset for schizophrenia well below 45 in men and women. However, recent data indicate that onset after age 45 is not as rare as was previously assumed. Data from the ECA study reveal that schizophrenia may remain undiagnosed in the elderly because the disease has a different presentation in this age group. When compared with younger persons, most elderly people with delusions or hallucinations may not have the typical pattern of chronic progressive schizophrenia and are less likely to be significantly impaired or to be under the care of a mental health specialist.

Sex Studies that do not separate groups by age of onset show a male-to-female ratio of close to 1, but this changes when various age cohorts are examined. Men are most likely to have the onset of symptoms between ages 15 and 25; women are at highest risk at ages 25 to 35; the reasons for this difference are not clear. The disease may manifest differently in the two sexes, hormonal factors may be involved or sociocultural factors may predispose men to earlier case findings.

As data from the WHO DOS show, when different cultures are examined the findings (earlier date for first treatment and first hospitalizations for men) are the same. More asocial premorbid characteristics, birth complications, and cerebral structural changes (especially in the left or dominant hemisphere) have been reported in men than in women, and schizophrenia in men may have a more chronic and disabling course. The findings are not conclusive and are limited by methodological problems such as failure to control for sociocultural factors.

Season and Birth Order Studies have shown that a disproportionate number of persons with schizophrenia are born during winter months (seasonal excess of approximately 10 percent); which, together with a birth pattern in their nonschizophrenic siblings that is similar to that seen in the general population, suggests the presence of a seasonal factor. Proposed explanations for this seasonal effect include deleterious environmental factors in the winter (such as temperature, nutritional deficiencies, infectious agents); a genetic factor in those with a propensity for schizophrenia that protects against infection and other insults and thus increases the likelihood of survival; and more frequent conception in the spring and summer by the parents of persons with schizophrenia.

Although no experimental testing has been conducted, studies appear to favor the harmful-effects hypothesis that schizophrenia involves infectious agents, but the

other hypotheses have not been ruled out conclusively. Although some studies in the southern hemisphere confirm a higher birth rate for schizophrenic persons in winter than in other seasons, further study of that hypothesis is needed. There are a number of methodological problems with previous studies. If there are statistically significant increases of schizophrenic births during the southern hemisphere winter, environmental factors should be favored over sociocultural ones. Whether winter- and summer-born persons with schizophrenia differ is not clear, but that would not necessarily be expected if the causative agent is active all year but more active in the colder months.

Early studies also reported a characteristic birth order pattern for persons with schizophrenia, but the results have not been consistent and family size can affect the findings. For example, some have found schizophrenia to be unusually common in the youngest children of large families and in the first-born sons of small families. Again, methodological problems limit the value of the studies.

Birth and Fetal Complications When compared with controls, persons with schizophrenia as a group, and especially male infants, experience a greater number of birth complications. Some studies have also reported a relationship between perinatal complications and early onset of disease, negative symptoms, and poorer prognosis. The crucial factor appears to be transient perinatal hypoxia, although not all infants so affected later develop a psychiatric disorder. There is, however, a general trend toward psychopathology in persons who have suffered obstetrical complications; such events appear to increase the vulnerability to development of schizophrenia and probably are not a specific cause. Some have proposed that complications at birth may be the result of preexisting fetal neurodevelopmental abnormalities or a vulnerability to such abnormalities. No prospective studies have been done, and retrospective case control studies may be biased if informants interviewed about a relative with schizophrenia try harder to remember birth complications than do informants reporting on healthy controls. Obstetrical records often refer only to severe complications.

Social Class Social class can be specified in various ways using some combinations of income, occupation, education, and place of residence. In previous studies the prevalence and number of newly identified cases of schizophrenia have been reported to be higher among members of the lower than the upper social classes. Two different explanations have been proposed. One explanation is that socioenvironmental factors found at lower socioeconomic levels are a cause of schizophrenia (social causation theory). Those factors include more life event stressors, increased exposure to environmental and occupational hazards and infectious agents, poorer prenatal care, and fewer support resources if stress does occur.

The other explanation is that lower socioeconomic status is a consequence of the disorder (social selection or drift theory). The insidious onset of inherited schizophrenia is believed to preclude elevating one's status or to cause a downward drift in status. Prospective studies have shown that persons with schizophrenia have less upward mobility from generation to generation than do the general population and that there is downward drift after the onset of symptoms. Many continue to argue this unsettled question, but a recent study strongly suggests that social drift processes are more important than social causation.

Marital Status Reports based on first hospital admissions have shown higher rates of schizophrenia for unmarried than for married patients, and some have inferred that single status contributes to the development of schizophrenia. However, the phenomenon may be similar to that described under social class; that is, the disease lessens the chance of marriage and increases the chance of divorce. Studies have not shown marriage to have a protective effect against schizophrenia and have not shown an excess of schizophrenia in widowed persons. Previous research using subjects hospitalized for the first time may have been flawed because unmarried and married men appear to have different hospital utilization patterns.

Immigration A higher risk for schizophrenia among recent immigrants than in native populations has been reported, but no study to date has confirmed that immigration stress leads to schizophrenia. Indeed, the ECA study found a low prevalence of schizophrenia among Mexican-Americans studied in Los Angeles, most of whom were immigrants. The generally reported increased prevalence of schizophrenia among immigrants could result from selection (i.e., persons with schizophrenia may be more likely to leave their families); from the failure to control for such other factors as social class, age, and sex; or from the failure to compare immigrant patients to nonimmigrant controls from the same homeland. These methodological issues limit any conclusions that can be drawn from existing reports.

Urbanization and Industrialization The prevalence of schizophrenia has been reported to be higher in urban environments than in rural areas. This is consistent with widely held beliefs that cities are places of rapid change and social disorganization, whereas rural areas are more socially stable and the inhabitants more integrated. However, data from the ECA study show no difference in the prevalence of schizophrenia between urban and rural areas when such factors as race, sex, and age are controlled.

The assertion that the prevalence and incidence of schizophrenia have increased in the twentieth century has been tested by comparing developing countries with industrialized nations, but such studies are fraught with methodological problems. For example, because infant mortality is lower in industrialized countries, those likely to develop schizophrenia may survive more frequently. Families are smaller and more insular, and ill members may be more obvious. The question of whether schizophrenia is more prevalent in modern times has also been studied by analyzing the reported number of new cases over time. However, it is difficult to control for probable diagnostic or recognition bias across centuries, especially for a disease that was first defined only in the late 1800s.

Life Stressors The association between stressful life events (such as loss of job, divorce) and the etiology and course of schizophrenia has been much studied. Schizophrenia or relapse of a preexisting disorder often follows extraordinary stress, so it has been suggested that such stress might provoke acute schizophrenia in a healthy person. Others argue that stress plays only a marginal role in the pathogenesis of the disorder or simply triggers schizophrenia in vulnerable persons. The few studies that have considered the issue have suffered the usual methodological problems of retrospective case-control studies and have had difficulty in outlining predispositional factors in schizophrenia. The stressor might have triggered the onset of a disorder that would have occurred without the stressor. The issue is not settled and will require further studies, especially prospective ones in which the role and severity of stressors in individual cases can be considered.

Infections Anatomical changes suggestive of viral infection of the central nervous system have been reported in some people with schizophrenia. A viral hypothesis is consistent with seasonal excesses and geographical differences. Viruses could also interact with a genetic predisposition, familial transmission, or both, in complex ways in the development of the disease. Recent studies have reported that exposure to viral infections during the second trimester may increase the risk for development of schizophrenia. As yet no study has conclusively shown an association between viral infection and the onset of schizophrenia. Further studies, especially those that can show evidence of viral transmission, are needed.

Suicide Risk Suicide is a leading cause of mortality in people suffering from schizophrenia. Estimates vary, but as many as 10 percent of people with schizophrenia may die because of a suicide attempt. Although the risk for suicide is greater in people with schizophrenia than in the general population, some risk factors—such as being male, white, and socially isolated—are similar in both groups. Factors such as depressive illness, a history of suicide attempts, unemployment, and recent rejection also increase the risk for suicide in both populations. Previous studies have revealed other risk factors that are unique to this disorder. Among these are being young and male and having a chronic illness with numerous exacerbations. A postdischarge course involving high levels of psychopathology and functional impairment increases the risk for suicide. In addition, people who have a realistic awareness of the deteriorative effects of the illness and a nondelusional assessment of their future are at increased risk for suicide. Other factors such as fear of further mental deterioration, hopelessness, excessive dependence on treatment, or loss of faith in treatment increase the risk of suicide in people with schizophrenia. The risk of mortality is especially high in the young, during the early postdischarge period, and early in the course of illness, although the risk persists across the person's life span. Risk factors identified in previous studies may be helpful in assessing acute suicidal risk in a specific individual. Further research is needed to better understand what risk factors are most predictive of future suicide in people with schizophrenia and what interventions are most helpful in preventing suicide.

Childhood Schizophrenia As with adult-onset schizophrenia, different diagnostic criteria can affect the interpretation of results from studies of childhood-onset schizophrenia. Early definitions of childhood-onset schizophrenia tended to be broad and often included patients with autistic disorder. Recent diagnostic systems have departed from these earlier definitions by using the more restrictive criteria applied to adults that emphasize hallucinations and formal thought disorder. This restrictive definition, however, fails to consider developmental issues, such as the nature of delusions in childhood, and how a formal thought disorder can be diagnosed in a child under 8 years of age whose formal cognitive processes are not fully developed. Others have considered developmental stages in diagnosing childhood-onset schizophrenia, but no consensus has been reached. The accuracy of any reported epidemiological data on childhood-onset schizophrenia is compromised by differences in diagnostic criteria. Therefore, the prevalence of childhood-onset schizophrenia is not clear, but it is probably less than that of early infantile autism and is estimated to be less than that of adult-onset schizophrenia. There does not appear to be a greater incidence in boys than girls, as there is in infantile autism.

The risk factors of childhood-onset schizophrenia are not well known, and many investigators have simply extrapolated from adult findings. However, environmental stressors, perinatal complications, and central nervous system dysfunction have all been reported to occur more frequently in children who are diagnosed with

schizophrenia.

FUTURE DIRECTIONS

Future epidemiological work in schizophrenia should use multisite, prospective, long-term studies. The WHO studies provide some of the foundations for such proposed efforts. However, longitudinal prospective studies of people at risk should be carried out, from near birth, and extending through the ages of major risk (early adult years). Such studies, with appropriate controls, should incorporate opportunities for genetic mapping of families at risk; chromosomal studies; and current laboratory measures of potential psychophysiological vulnerability such as continuous performance and sensory discrimination testing, neuroimaging, and other measures evolving with methodological advances. The expense of such studies would not be greater than that of comparable multisite, long-term studies of risk factors for cardiovascular and other diseases and would be small compared to the extraordinary direct and indirect costs of this most devastating of mental disorders.

SUGGESTED CROSS-REFERENCES

Some of the methods and concepts applicable to this section are discussed in Section 5.2 on statistics and experimental design. The genetics of schizophrenia is discussed in Section 12.5. Other aspects of schizophrenia are presented throughout the other sections of Chapter 12. Other psychotic disorders are reviewed in Chapter 13. Section 11.3 discusses amphetamine-related disorders, Section 11.7 discusses hallucinogen-related disorders, and Section 11.11 discusses phencyclidine-related disorders.

SECTION REFERENCES

Caldwell CB, Gotesman II: Schizophrenics kill themselves too: A review of risk factors for suicide. Schizoph Bull *16:*571, 1990.

Cannon T, Mednick S, Parnas J, Schulsinger S, Praestholm J, Vestergaard A: Developmental brain abnormalities in the offspring of schizophrenic mothers. Arch Gen Psychiatry *50:*551, 1993.

Castle DJ, Murray M: The epidemiology of late-onset schizophrenia. Schizophr Bull *19:*691, 1993.

*Castrogiovanni P, Iapichino S, Pacchierotti C, Pieraccini F: Season of birth in psychiatry. A review. Neuropsychobiology *37:*175, 1998.

Cohen A: Prognosis for schizophrenia in the third world: A reevaluation of cross-cultural research. Cult Med Psychiatry *16:*53, 1992.

Crow TJ: Prenatal exposure to influenza as a cause of schizophrenia. There are inconsistencies and contradictions in the evidence. Br J Psychiatry *164:*588, 1994.

Day R, Nielsen JA, Korten A, Ernberg G, Dube KC, Gebhart J, Jablensky A, Leon C, Marsella A, Olatawura M, Sartorius N, Stromgren E, Takahashi R, Wig N, Wynne LC: Stressful life events preceding the acute onset of schizophrenia: A cross-national study from the World Health Organization. Cult Med Psychiatry *11:*123, 1987.

*Dohrenwend BP, Levov I, Shrout PE, Schwartz S, Noveh G, Link BG, Skodol AE, Stueve A: Socioeconomic status and psychiatric disorders: The causation-selection issue. Science *255:*946, 1992.

Eaton WW: The epidemiology of schizophrenia. In *Handbook of Studies on Schizophrenia*, GD Burrows, TR Norman, G Rubinstein, editors. Elsevier, New York, 1986.

Edgerton RB, Cohen A: Culture and schizophrenia: The DOSMD challenge. Br J Psychiatry *164:*222, 1994.

*Gottesman I, Moldin S: Schizophrenia genetics at the millennium: Cautious optimism. Clin Genet *52:*404, 1997.

Haffner H: What is schizophrenia? Changing perspectives in epidemiology. Eur Arch Psychiatry Neurol Sci *238:*63, 1988.

Jablensky A: Schizophrenia: Recent epidemiologic issues. Epidemiol Rev *17:*10, 1995.

Jablensky A, Sartorius N, Ernberg G, Anger M, Korten A, Cooper JE, Day R, Bertelsen A: Schizophrenia: Manifestations, incidence, and course in different cultures. Psychol Med *20:*38, 1992.

Jablensky A: The 100-year epidemiology of schizophrenia. Schizophr Res *28:*111, 1997.

Jones P, Cannon M: The new epidemiology of schizophrenia. Psychiatr Clin North Am *23:*1, 1998.

Jones PB, Rantakallio P, Hartikainen A, Isohanni M, Sipila P: Schizophrenia as a long-term outcome of pregnancy, delivery, and perinatal complications: A 28-year follow-up of the 1966 North Finland general population birth cohort. Am J Psychiatry *155:*355, 1998.

Karayiorgou M, Kasch L, Lasseter VK, Hwang J, Elango R, Bernardini DJ, Kimberland M, Babb R, Francomano A, Wolyniec PS, Lamacz M, Nestadt G, Meyers D, Ott J, Childs B, Antonarakis S, Kazazian HH, Housman DE, Pulver AE: Report from the Maryland Epidemiology Schizophrenia Linkage Study: No evidence for linkage between schizophrenia and a number of candidate and other genomic regions using a complex dominant model. Am J Med Genetics (Neuropsychiatr Genet) *54:*345, 1994.

Keith SJ, Regier DA, Rae DS: Schizophrenic disorders. In *Psychiatric Disorders in America*, LN Robins, DA Regier, editors. Free Press, New York, 1991.

Kendler KS, Diehl SR: The genetics of schizophrenia: A current, genetic-epidemiologic perspective. Schizophr Bull *19:*261, 1993.

*Kendler KS, Gallagher TJ, Abelson JM, Kessler RC: Lifetime prevalence, demographic risk factors, and diagnostic validity of nonaffective psychosis as assessed in a US community sample. Arch Gen Psychiatry *53:*1022, 1996.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler K: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry *51:*8, 1994.

Kringlen E, Cramer G: Offspring of monozygotic twins discordant for schizophrenia. Arch Gen Psychiatry *46:*873, 1989.

Levinson DF, Mahtani MM, Nancarrow DJ, Brown DM, Kruglyak L, Kirby A, Hayward NK, Crowe RR, Andreasen NC, Black DM, Silverman JM, Endicott J, Sharpe L, Mohs RC, Siever LJ, Walters MK, Lennon DP, Jones HL, Nertney DA, Daly MJ, Gladis M, Mowry BJ: Genome scan of schizophrenia. Am J Psychiatry *155:*741, 1998.

Lewis MS: Age incidence and schizophrenia: Part II. The season of birth controversy. Schizophr Bull *15:*75, 1989.

McGrath J, Castle D: Does influenza cause schizophrenia? A five-year review. Aust NZ J Psychiatry *29:*23, 1995.

Muller HG, Kleider W: A hypothesis on the abnormal seasonality of schizophrenic births. Eur Arch Psychiatry Neurol Sci *239:*331, 1990.

Narrow WE, Regier DA, Rae DS, Manderscheid RW, Locke BZ: Use of services by persons with mental and addictive disorders. Arch Gen Psychiatry *50:*95, 1993.

Norman RM, Malla AK: Stressful life events and schizophrenia. Br J Psychiatry *162:*161, 1993.

O'Callaghan E, Gibson T, Colohon HA, Walshe D, Backley P, Lorkin C, Waddington JL: Season of birth in schizophrenia. Br J Psychiatry *158:*764, 1991.

Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, Goodwin FK: Comorbidity of mental disorders with alcohol and other drug abuse. JAMA *264:*2511, 1990.

Regier DA, Farmer ME, Rae DS, Myers JK, Kramer M, Robins LN, George LK, Karno M, Locke BZ: One-month prevalence of mental disorders in the United States and sociodemographic characteristics: The epidemiologic catchment area study. Acta Psychiatr Scand *88:*35, 1993.

*Regier DA, Narrow WE, Rae DS, Manderscheid RW, Locke BZ, Goodwin FK: The de facto U.S. mental and addictive disorders service system. Arch Gen Psychiatry *50:*85, 1993.

Torrey EF, Bowler AE: Geographical distribution of insanity in America: Evidence for an urban factor. Schizophr Bull *16:*591, 1990.

Torrey EF: Prevalence studies in schizophrenia. Br J Psychiatry *150:*598, 1987.

Torrey EF, Bowler AE, Rawlings R, Terrazas A: Seasonality of schizophrenia and stillbirths. Schizophr Bull *19:*557, 1993.

*Waldo MC: Schizophrenia in Kosrae, Micronesia: Prevalence, gender ratios, and clinical symptomatology. Schizophr Res *35:*175, 1999.

Wing JK, Cooper JE, Sartorius N: *Measurement and Classification of Psychiatric Symptoms: An Instruction Manual for the Present State Examination and CATEGO Program* . Cambridge University Press, Cambridge, 1974.

Textbook of Psychiatry

12.3 SCHIZOPHRENIA: BRAIN STRUCTURE AND FUNCTION

RAQUEL E. GUR, M.D., FH.D. AND RUBEN C. GUR, FH.D.

The past decade has seen a transformation in psychiatric thinking about schizophrenia. Its conception as a complex behavioral disorder that reflects an interplay between biological and psychosocial factors has been changing as developments in the neurosciences have yielded tools for probing neural substrates of behavior. In particular, advances in neuroimaging methods have enabled translational research in which schizophrenia can be studied from the molecular to the behavioral levels using complementary top-down and bottom-up strategies. Such work has identified some consistent aberration in brain structure and function that may help formulate our new conception of schizophrenia as a brain disorder. This is not to dismiss environmental stressors, but rather to put these in the perspective of a brain disorder in evolution.

Two major aspects of brain integrity can be assessed through neuroimaging: structural anatomy and functional activity. In view of the complexity and course of schizophrenia, these measures need to be taken across the life span and longitudinally to document the association between brain changes and behavior. Furthermore, the disorder-related effects are superimposed on healthy individual differences—for example, sex differences and maturational changes—that have to be established in healthy people before we can understand pathological changes. However, despite the complexity of such an approach, its implementation can yield new opportunities for elucidating the neural substrate of schizophrenia in a way that will lead to improved diagnosis and treatment.

Accordingly, the section first describes studies of structural imaging in schizophrenia, which have been linked to neuropathological findings in postmortem research. This is followed by findings obtained by functional imaging studies.

NEUROANATOMICAL STUDIES IN SCHIZOPHRENIA

Neuroanatomical correlates of dysfunctional performance have provided the foundation for current thinking about brain regulation of behavior. The behavioral aberrations manifested in schizophrenia implicate a diffuse abnormality likely to involve several brain systems. Defining the neuroanatomical differences and possible changes associated with schizophrenia is arguably a prerequisite for understanding its neural substrates and for interpreting functional studies of brain physiology and neurochemistry. Structural studies have progressed from reliance on ratings to planimetric measures and, more recently, reliable computerized segmentation methods for obtaining volumetric measures. The improvement in precision of neuroanatomical parameters has yielded some consistency in effects and correlations with clinical and neurobehavioral measures.

Structural Neuroimaging Earlier neuroimaging studies with computerized tomography (CT) applied nonvolumetric measures and suggested enlarged ventricles, implying reduced brain parenchyma. Magnetic resonance imaging (MRI) has advanced the study of the neuroanatomy of schizophrenia. It offers improved sensitivity for examining sulcal changes, better contrast resolution, direct multiplanar imaging, lack of bone artifacts, and no ionizing radiation. The field has progressed from initial studies of small samples, examining multiple regions with low-field scanners, using linear and area measurements, to the application of computerized image analysis in large samples. This has enabled linking neuroanatomical measures to two behavioral dimensions: clinical features of the disorder and neurocognitive deficits.

Whole Brain and CSF Volumes Studies with MRI have replicated earlier findings with CT indicating smaller brain volume and more cerebrospinal fluid in patients with schizophrenia than in healthy people. As can be seen in Figure 12.3-1, a person with schizophrenia shows evidence for widening of CSF spaces in both the ventricles and the sulci. Image segmentation methods have permitted increasingly precise quantitation of brain and CSF volume, and although these studies generally support the notion of increased CSF relative to brain volume in schizophrenia, they also indicate considerable overlap with healthy people (Fig. 12.3-2). This suggests that abnormalities at the level of whole brain may characterize only subtypes of patients with schizophrenia. Some patients exhibit a concomitant decrease in brain and increase in CSF volume, consistent with atrophy, whereas other patients show concomitant decreases in brain and CSF volume (hence cranial volume) which is more consistent with dystrophy. A third group shows neither abnormality.

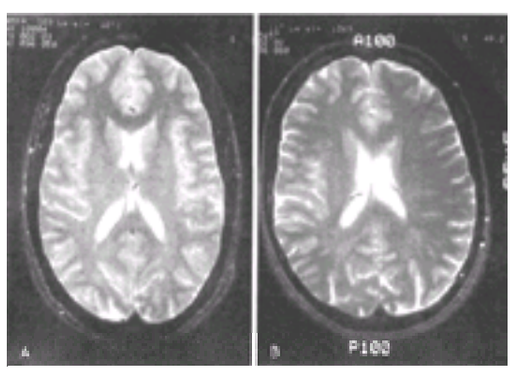

FIGURE 12.3-1 MRI of a young healthy adult **(A)** and a same-aged adult with schizophrenia **(B)**. Radiological examination showed evidence of the patient's increased cerebrospinal fluid in ventricles and sulci.

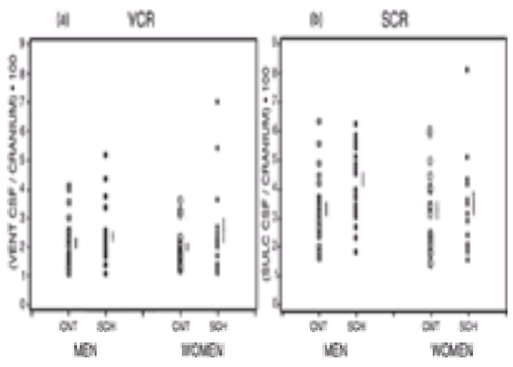

FIGURE 12.3-2 Scatterplot of brain volume of adults with schizophrenia relative to that of healthy adults matched sociodemographically.

These measures have been related to phenomenological and other clinical variables such as premorbid functioning, symptom severity, and outcome. The results

suggest that whole brain measures are related to clinical features. Abnormalities in these measures are likely to be more pronounced in patients with poorer premorbid functioning, more severe symptoms, and worse outcome. The concept of brain reserve that has been suggested in other disorders, such as Alzheimer's disease, may apply to schizophrenia as well. Thus, normal brain and CSF volumes are preliminary indicators of protective capacity. As our understanding of how brain systems regulate behavior in health and disease improves, we can take advantage of neuroimaging to examine specific brain regions implicated in the pathophysiology of schizophrenia.

Regional Volumes Examination of brain regions implicated in schizophrenia has required methodological developments to enable testing specific hypotheses regarding the neuroanatomical basis of the aberrant behavior. Studies have evolved beyond comparing patients with controls, and current work linking regional measures to specific symptoms and subtypes requires large samples and systematic data acquisition. Furthermore, given the subtle changes evident in schizophrenia, findings must be evaluated relative to well-characterized healthy people. For example, the effects of normal aging appear to be sexually dimorphic. Cross-sectional studies reported age-related reduction in frontal and temporal lobe volumes in healthy men but not women (Fig. 12.3-3).

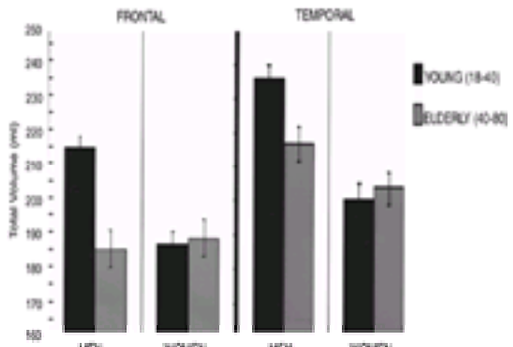

FIGURE 12.3-3 Frontal and temporal lobe volume in younger and older healthy adult men and women. (Reprinted with permission from Cowell PE, Turetsky BT, Gur RC, Grossman RI, Shtasel DL, Gur RE: Sex differences in aging of the human frontal and temporal lobe. J Neurosci *14:*4748, 1994.)

The main regions showing consistent abnormalities in schizophrenia have been frontal and temporal lobe structures. Lower frontal and temporal lobe volume has been observed in patients than in healthy controls matched demographically (Fig. 12.3-4). Reduced temporal lobe volume correlated with both memory impairment and severity of negative symptoms. These findings were observed in first-episode patients, indicating that structural changes are evident at the first clinical presentation, which supports a neurodevelopmental origin. The differences are more pronounced in men than in women with schizophrenia.

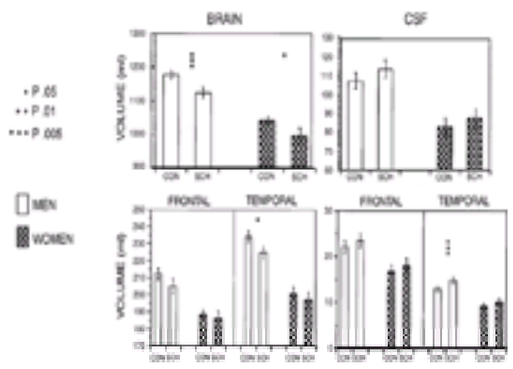

FIGURE 12.3-4 Frontal and temporal lobe volume in adults with schizophrenia compared with that of healthy adults matched sociodemographically.

Reduced volume was reported in multiple regions including the superior temporal gyrus, hippocampus, and thalamus. These structures are critical for maintaining the integrity of the complex behaviors that are impaired in schizophrenia. Figure 12.3-5 illustrates some of the regions that show abnormal volume in schizophrenia, as well as an example of an image that has been segmented into gray and white matter and CSF. Most regions show volume decrease; the exception is basal ganglia regions reported to show increased volume in schizophrenia. This increase seems to be related to the effects of dopamine receptor antagonists.

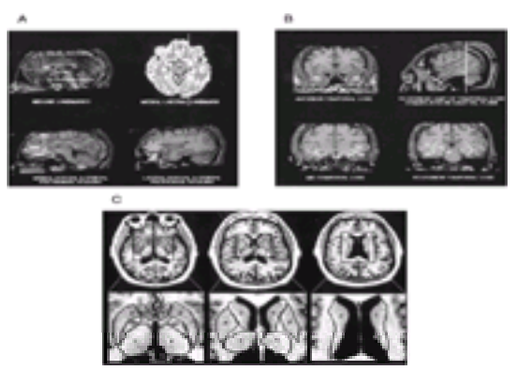

FIGURE 12.3-5 Illustration of regions that have shown volume abnormalities in schizophrenia: **A**, frontal lobe; **B**, temporal lobe; **C**, basal ganglia.

Gray Matter and White Matter Segmentation MRI can yield detailed anatomical information apart from the demarcation of brain and CSF. With sequences that are T1-weighted it is possible to segment gray from white matter (Fig. 12.3-5A, upper right image). Several lines of investigation have demonstrated advantages of applying such methods. Gray matter changes have been found during adolescence and in the course of the normal aging process. Gray-white segmentation is critical in developmental studies in which age-related decreases in gray matter may be obscured by simultaneous increases in total brain and cranial volumes. These improvements in image processing methodology have helped determine whether tissue loss and disorganization in schizophrenia is primarily a gray matter deficit or whether abnormalities in white matter are also involved.

Reduced cortical gray matter was noted in a number of studies that evaluated chronic patients with schizophrenia. More recently, one study of first-episode patients also reported a gray matter deficit in individuals with a recent onset of illness.

Longitudinal Studies Efforts to elucidate the pathophysiology of schizophrenia have focused on the role of the neurodevelopmental relative to progressive neurodegenerative processes. Documenting neuroanatomical aberrations with structural neuroimaging and evaluating their course in relation to clinical and neurobehavioral manifestations can help test such hypotheses. CT and MRI studies have been primarily cross-sectional, and a longitudinal design is necessary to examine the possibility of progressive deterioration suggested by the neurodegenerative hypothesis. The few longitudinal evaluations of structural abnormalities have not been integrated with clinical and neurobehavioral measures.

Results from CT follow-up studies, with planimetric methods and ratings of sulcal enlargement, have varied. Some studies found no significant changes in

ventricle-brain ratio (VBR) in relatively small samples rescanned after a number of years, commonly ranging from 2 to 5. Other studies reported that some patients do show an increased VBR over similar time spans. The investigators noted that these initial studies have limitations related to sample size, patient characteristics, and scanning and measurement procedures.

While most follow-up CT studies evaluated chronic patients with schizophrenia, MRI longitudinal studies have examined first-episode patients. This is an informative population because the design enables prospective follow-up starting early in the course of illness. One group of investigators found no ventricular changes in a follow-up (1–2 years) study of 13 patients and 8 controls. Lyn DeLisi and her colleagues initially evaluated 16 patients and 5 controls, studied 2 years after a first psychotic episode. Patients showed no consistent change in ventricular size with time, although there were individual increases or decreases. With a larger group of 24 patients and 6 controls no significant changes were observed in ventricular or temporal lobe volume at follow-up. Recently, a report on 20 of these patients and 5 controls rescanned over 4 years noted greater decreases in whole-brain volume and enlargement in left ventricular volume in patients and concluded that subtle cortical changes may occur after the onset of illness.

These authors have described a reliable and validated method for obtaining MRI measures of brain volume. In healthy adults, these parameters have been related to sex differences and the effects of aging, and in schizophrenia, they were related to clinical features. This method was applied in a longitudinal study of 40 patients (20 first-episode, 20 previously treated) and 17 healthy controls, rescanned an average of 2.5 years later. Volumes of whole brain, CSF, and frontal and temporal lobes were measured. Severity of negative and positive symptoms was assessed, medications were monitored, and neurobehavioral functioning in eight domains was evaluated. First-episode and previously treated patients had smaller whole-brain and frontal and temporal lobe volumes than controls at intake. Longitudinally, reduced frontal lobe volume was found only in patients, whereas temporal lobe reduction was also seen in controls. The association between volume reduction and symptom change differed between patient groups, but in both first-episode and previously treated patients, volume reduction was associated with decline in some neurobehavioral functions. The existence of neuroanatomical and neurobehavioral abnormalities in first-episode patients indicates that brain dysfunction occurred before clinical presentation. However, the longitudinal studies suggest progression in which anatomical changes may affect some clinical and neurobehavioral features of the illness in some patients.

The limited number of longitudinal MRI studies and small sample sizes leaves the question of progression unresolved and precludes confident distinction of disease-related changes from those associated with normal aging. Furthermore, standard therapeutic interventions need to be included in such longitudinal studies.

FUNCTIONAL NEUROIMAGING

Application of Functional Brain Imaging Methods Evolving technology provides an increasing array of measures of brain function. Some of these measures overlap and others are complementary. For example, the functional integrity of the brain can be examined through measures related to energy metabolism, such as rates of glucose and oxygen utilization and cerebral blood flow. Neuroreceptor function can be assessed through methods for measuring receptor density and affinity at presynaptic and postsynaptic sites. Methods that have been applied in schizophrenia included the Xenon-133 (^{133}Xe) clearance technique for measuring cortical cerebral blood flow; positron emission tomography (PET) for assessing glucose metabolism, cerebral blood flow, and neuroreceptor functioning; single photon emission computerized tomography (SPECT) for studying cerebral blood flow and neuroreceptors; and, more recently, functional MRI (fMRI) for measuring changes attributable to cerebral blood flow. Figure 12.3-6 illustrates the application of such methods in healthy people.

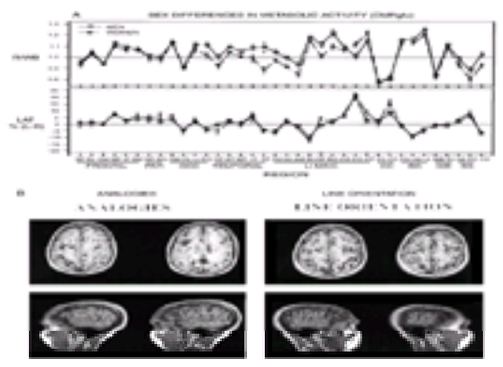

FIGURE 12.3-6 Illustration of functional imaging data obtained in healthy people: **A**, sex differences in local glucose metabolism; **B**, activation with verbal and spatial tasks as seen by functional MRI. Abbreviations: SF, superior frontal; DL, dorsal prefrontal-lateral; DM, dorsal prefrontal-medial; MF, midfrontal; IF, inferior frontal; SM, sensorimotor; SP, superior parietal; SG, supramarginal gyrus; OL, occipital cortex, lateral; OM, occipital cortex, medial; LI, lingual gyrus; FG, fusiform gyrus; OT, occipital temporal; ST, superior temporal; MT, midtemporal; IT, inferior temporal; TP, temporal pole; PH, parahippocampal gyrus; HI, hippocampus; AM, amygdala; IN, insula; OF, orbital frontal; RG, rectal gyrus; CA, cingulate gyrus-anterior; CG, cingulate gyrus; CP, cingulate gyrus-posterior; C1, corpus callosum-anterior; C2, corpus callosum-posterior; CN, caudate nucleus; LM, lenticular-medial (globus pallidus); LL, lenticular-lateral (putamen); MB, mammillary body; TH, thalamus; MI, midbrain; PO, pons; CE, cerebellum. Cortical regions are grouped by lobe in a rostral-caudal order, followed by corpus callosum and subcortical regions. This order heuristically also reflects ontogenic and evolutionary development. Top graph shows means + SEM of region to whole brain (R/WB) ratios, and bottom graph shows laterality differences in percentage, i.e., 100*(LR)/mean(L,R). (See Color Plate 7.)

Links between clinical features of schizophrenia and brain function have been guided by hypotheses relating behavior to specific brain regions and systems implicated in schizophrenia. These links are based on preclinical research and the emergence of symptoms commonly seen in schizophrenia that also occur following brain lesions. Persistent negative symptoms have been observed as a eurobehavioral sequela of frontal lobe damage. Other frontal lobe functions such as abstraction, attention, verbal fluency, mental flexibility, and concept formation are also impaired. Productive positive symptoms of hallucinations and delusions have been related to the temporolimbic system, and this region is also implicated by evidence of impaired learning and memory. The greater impairment in verbal functions, the similarity of some symptoms to those observed in patients with left temporal lobe epilepsy, and the increased frequency of left-handedness (or, rather, “left-sidedness” as measured by a combined index of strength of right-handedness, footedness, and sighting dominance) in schizophrenia have led to the laterality hypothesis stipulating left hemispheric dysfunction. Thus the laterality gradient has been examined in several studies. Subcortical regions have been studied with special emphasis on the basal ganglia (implicated by the dopamine hypothesis) and the thalamus (related to sensory gating).

These early efforts, focusing on brain systems that are likely to modulate normal and pathological psychotic behavior, have generated hypotheses that can be examined with functional brain imaging. In addition to obtaining baseline measures of the resting topography of glucose metabolism and cerebral blood flow, functional imaging is an especially powerful methodology for the probe paradigm. There are two complementary approaches: neurobehavioral probes and neuropharmacological probes. The application of neurobehavioral probes has enhanced our ability to evaluate brain systems that regulate specific processes in healthy people and in those affected by schizophrenia, including attention, learning, memory, and executive functions. Neuropharmacological probe paradigms include examination of neuroreceptor function as well as the effects of pharmacological intervention on cerebral blood flow and metabolism.

CEREBRAL METABOLISM AND BLOOD FLOW STUDIES

Semour Kety and his colleagues have pioneered the measurements of whole-brain metabolism and blood flow in healthy people and reported normal values for patients with schizophrenia. Subsequent studies of regional cerebral metabolism and blood flow can be divided into those that measure the physiological parameters at a resting state and those that introduce a perturbation, or challenge, in the form of a neurobehavioral probe or a pharmacological intervention. Initially, investigators have aimed at assessing whether resting cerebral blood flow and glucose metabolism differ between patients with schizophrenia and healthy controls. The topography of physiological activity was examined along the anterior-posterior, subcortical-cortical, and lateral dimensions.

Resting Baseline The frontal lobes were implicated when early physiological studies of cerebral blood flow, reported that patients did not show the normal pattern of more anterior than posterior flow. This “hypofrontal” disturbance in the anterior-posterior gradient has been supported by some, but not all, studies of resting cerebral blood flow (by ^{133}Xe and SPECT) and glucose metabolism (with PET). The relation between this pattern of metabolic activity and clinical variables has also been examined. Decreased frontal metabolic activity has been associated with duration of illness and negative symptoms. Longer duration of illness and more severe

negative symptoms are related to a relative decrease in frontal lobe metabolism.

Differences in resting values between patients and controls were also found in laterality indexes, suggesting relatively higher left hemispheric values in more severely ill patients. Furthermore, improvement in clinical status correlated with a shift toward lower left hemispheric metabolism relative to that in the right hemisphere. This supports hypotheses derived from behavioral data concerning lateralized abnormalities in schizophrenia.

After assessing global, anterior-posterior, and lateral dimensions, investigators have begun the study of functional changes in brain systems linked to other impaired behavior in schizophrenia. Dysfunction in temporolimbic structures, including the hippocampus as well as temporal cortex, is supported by neuroanatomical and neuropsychological studies. Lateralized abnormalities in these regions, with greater left than right hemispheric dysfunction, are implicated by characteristic clinical features of schizophrenia, such as thought disorder, auditory hallucinations, and language disturbances. PET studies of temporal lobe metabolism show both increased and decreased glucose utilization. Decreased metabolism was also noted in hippocampus and anterior cingulate cortex. Studies in this region have been limited in part by instrument resolution.

Metabolism and flow pattern in temporolimbic regions have also been related to symptoms. An oxygen-15 (^{15}O)-labeled water study with PET described abnormal cerebral blood flow in the parahippocampal gyrus, associated with positive symptoms. Hallucinations were associated with SPECT blood flow changes in the hippocampus, parahippocampus, and amygdala. There are conflicting reports of superior temporal gyrus functional changes in schizophrenia during active auditory hallucinations. While one study suggested that patients with hallucinations have lower relative metabolism in Wernicke's region, another study showed asymmetrical temporal lobe perfusion (lower in the left than the right) in patients with auditory hallucinations. In one PET study the rate of glucose metabolism was greater in the left anterior temporal lobe and was related to the severity of symptoms. This is consistent with another reported association between severity of symptoms and a relative increase in left hemispheric metabolism.

These reports varied in the method used and definition of regional parameters. Most studies used ratios such as region to whole brain or anterior to posterior rather than absolute values of activity. Inconsistencies in findings could also be related to sample size, heterogeneity, analytical approaches, and individual techniques. Most studies included relatively small samples of patients, which varied in important clinical factors such as chronicity, symptom subtypes and severity, level of functioning, and history of treatment. Furthermore, inclusion criteria varied, and some laboratories applied more stringent criteria (e.g., related to history of comorbidity of substance abuse or head trauma with loss of consciousness). Another potential source of variability in results is the definition of resting state. Investigators have been reluctant to include an unstructured resting state because of concern that such measures will be uncontrolled and therefore produce unreliable results. Some studies used reduced sensory input, and others used sensory stimulation to standardize this condition. However, several studies examined the reproducibility of resting baseline measures with relatively unstructured conditions (i.e., eyes open and ears unoccluded, with ambient noise kept to a minimum). These studies found high reproducibility among healthy subjects and patients with schizophrenia.

Given the demonstrated reliability of the standardized resting baseline condition, these authors believe such a condition should be included in physiological neuroimaging studies. This will serve three main purposes. First, it will permit comparison across studies within a center as technology evolves and patient characteristics change. Without a common resting baseline condition it would be impossible to interpret differences in results. Second, it will enable comparability across centers. Imagine the need to explain why two centers using the same or similar tasks find evidence for different regional abnormalities in schizophrenia. If resting baseline values are available and are comparable in the two samples, different task effects could be legitimately attributed to theoretically meaningful sources such as task condition or symptomatic variability. Third, a standardized resting baseline provides a reference point for determining whether a given task or condition has increased neural activity. In studies that have included such a condition, cognitive activation was consistently shown to increase cortical activity in both patients and controls. Using a resting baseline condition enables the investigator to make much stronger statements in interpreting regional effects. Rather than being restricted to statements that a given region has changed in its activation relative to the remainder of the brain, resting baseline data can be used to determine whether the task has induced increased neural activity.

Functional changes in the basal ganglia have been examined with PET and SPECT. Several PET studies implicate basal ganglia dysfunction in schizophrenia. The withdrawal-retardation factor (emotional withdrawal, blunted affect, and motor retardation) of the Brief Psychiatric Rating Scale has been negatively correlated with PET basal ganglia metabolic activity. Neuroleptic-naive patients with schizophrenia were reported to have relatively increased blood flow in the left globus pallidus. Other PET studies report decreased basal ganglia metabolism in schizophrenia, while yet others found increased basal ganglia metabolic rates following administration of neuroleptic medication.

Thus, while the contribution of PET metabolic and blood flow studies so far has been to add to the growing evidence implicating basal ganglia involvement in schizophrenia, the exact nature of the dysfunction remains unclear. In particular, the relation between basal ganglia and frontal lobe activity in schizophrenia needs further scrutiny. Emerging evidence from structural and functional imaging indicates a dynamic interrelationship between the various key regions. One study showed that patients with schizophrenia not only fail to activate dorsolateral prefrontal cortex in response to the Wisconsin Card Sorting Test, but they also fail to inhibit caudate activation. Hence, in schizophrenia, basal ganglia continue to show relatively increased flow in the caudate during performance of the task, while healthy controls seem to demonstrate a reciprocal relationship in which relative blood flow decrease in the basal ganglia is associated with increased perfusion to the frontal region.

Activation Studies Regardless of the debate over the value of obtaining resting baseline measures, measures of cerebral blood flow and metabolism during the performance of cognitive tasks clearly tend to accentuate differences between patients and controls. Perhaps even more importantly, such measures are critical for establishing the link between behavioral deficits and the ability of brain regions to become activated in response to task demands. This expectation has been supported in studies that used neurobehavioral probes.

The general approach in the field has been to work from hypotheses, derived from neurobehavioral data, which associate behavioral measures with regional brain function. Task selection can be made to include a target task (for which patients are expected to have a differential deficit) and control tasks. Patients are then compared with healthy controls in the pattern of task-induced changes in regional brain activity. This has now become the established research paradigm and significant progress has been made since the early studies with ^{133}Xe.

The authors, Daniel Weinberger, and their colleagues applied the ^{133}Xe method during resting measures and while subjects were performing specific tasks. Both groups found no differences in overall or hemispheric cerebral blood flow between patients and controls at resting baseline. However, distinct abnormalities were seen when physiological activity was measured in response to cognitive probes. Pursuing the laterality hypothesis, the first author of this section and coworkers administered tasks with a demonstrated link to left (verbal analogies) and right (spatial line orientation) hemispheric functioning. Healthy controls showed the expected greater left hemispheric increase for the verbal task and greater right hemispheric increase for the spatial task. However, patients with schizophrenia had a bilaterally symmetrical activation for the verbal task and greater left hemispheric activation for the spatial task. Thus, patients failed to show the normal left hemispheric dominance for the verbal task and instead showed left hemispheric overactivation for the spatial task.

Similarly, Weinberger and coworkers found no regional abnormalities in the resting cerebral blood flow of patients with schizophrenia. However, distinct abnormalities were reported in the dorsolateral prefrontal region during activation with the Wisconsin Card Sorting Test of abstraction and mental flexibility, which is sensitive to frontal lobe damage. Application of this paradigm to the study of monozygotic twins discordant for schizophrenia revealed that all affected twins had lower dorsolateral prefrontal cortex cerebral blood flow response than discordant cotwins. Furthermore, negative symptoms, which have been related to frontal lobe dysfunction, showed a negative correlation with frontal blood flow during performance of executive tasks but not control tasks. Probing brain systems with specific tasks has also been advanced in SPECT and in cerebral blood flow studies with PET. These methods have also indicated abnormalities in patients with schizophrenia with a range of tasks including memory, executive, and attentional measures. The consistent finding is a lack of normal regional activation in response to task, and activation in some regions not seen in healthy subjects.

These results suggest that brain systems recruited for the performance of specific tasks in healthy people are not similarly engaged in patients with schizophrenia. What may account for such aberrations? Genetic liability, neurodevelopmental abnormalities in which brain systems fail to achieve maturity, or the impact of a psychotic process that interrupts normally developed structures and processes? Does therapeutic intervention ameliorate the abnormal signature? How specific are the results to schizophrenia? These are some of the questions yet to be answered that can certainly be addressed with neuroimaging.

Functional MRI The introduction of MRI is an exciting, more recent development in functional imaging research. Functional MRI methods offer several potential advantages over PET for imaging brain function, including higher spatial resolution, higher temporal resolution, noninvasiveness, lack of ionizing radiation, direct correlation with anatomical imaging, greater reproducibility, and economy. Disadvantages include the loud background noise generated by the gradients, difficulties in

presenting stimuli and performing tasks in the magnet bore, claustrophobia, low signal-to-noise ratio for most methods, and lack of quantitation in physiological units for most methods. Many of these disadvantages can be overcome by using specialized equipment compatible with the MRI environment. These methods are described briefly because they are recent and hold potential for functional imaging in schizophrenia.

Currently, three main techniques exist for MRI of the brain. Gadolinium bolus-tracking was the first technique to be applied to mapping task-specific regional brain function in animals and humans by use of MRI. In normal brain, gadolinium diethylenetriaminepentaacetic acid (DPTA) is an intravascular tracer, allowing semiquantitative transit time and blood volume images to be calculated with rapid imaging techniques. Because of the accumulation of the intravascular tracer, the number of determinations is limited to two to three per day. Because gadolinium DPTA is an intravascular tracer, dynamic measurements of its passage through brain yield measurements of cerebral blood volume and mean transit time rather than cerebral blood flow, but changes in cerebral blood flow are generally reflected by changes in these other indexes.

Blood oxygenation–sensitive imaging has been most widely applied to fMRI, replicating previous PET studies. The technique relies on magnetic susceptibility effects of deoxyhemoglobin that cause regional signal decreases in imaging sequences that are sensitive to susceptibility (e.g., echoplanar or routine gradient echo sequences). With regional brain activation studies a net increase in signal intensity is observed in regions known to be activated by the task. The increase in image intensity corresponds to a local decrease in deoxyhemoglobin. This finding is attributed to a greater increase in regional blood flow than in regional oxygen consumption, a notion supported by PET measurements of blood flow and oxygen consumption with regional brain activation. A wide variety of pulse sequences can be used to obtain blood oxygenation–sensitive imaging measures. Many simple activation paradigms have been tested, and activation has been observed with both fast and slow imaging. A typical response is a 1 to 25 percent change in regional image intensity, which develops over 3 to 8 seconds following task initiation. Susceptibility effects of deoxyhemoglobin are field dependent. Thus, a scanner with 1.5 tesla field strength would typically record signal changes with functional activation of about 0.25 to 5 percent, while at 4 telsa changes up to 25 percent have been observed. The main advantage of ultrafast imaging is that the time course of signal change can be observed and multislice imaging can be carried out in a reasonable time period.

The third technique, arterial spin tagging (quantitative perfusion imaging) uses magnetization tagging of endogenous arterial water to determine the perfusion of brain parenchyma by comparing images obtained with and without a labeled arterial supply. The method is analogous to steady-state techniques used in PET, since the regional signal intensity depends upon the arterial blood flow (which delivers labeled spins) and the T1 relaxation rate (which causes the labeling to decay). This technique has the important advantage of providing quantitative cerebral blood flow parameters. Furthermore, perfusion is measured in brain parenchyma directly and is thus better localized than measurements obtained by use of an intravascular tracer, which is most sensitive to venous outflow effects. There may also be less motion sensitivity than with blood oxygenation–sensitive imaging.

Application of this technology to the study of schizophrenia is quite new. Perry Renshaw and colleagues measured the relative change in image signal intensity caused by photic stimulation in eight patients and nine controls. The mean signal intensity change in the primary visual cortex was significantly greater in patients than in controls. A subsequent study examined a sample of 12 subjects with schizophrenia and 11 healthy controls performing a word fluency task, associated with left frontal lobe function. Patients showed less left frontal activation and greater left temporal activation than controls. Sensorimotor cortex and supplementary motor area activation were examined in right-handed patients (8) and controls (9) during finger-to-thumb opposition. All subjects showed a significant activation of the supplementary motor area and both ipsilateral and contralateral sensorimotor cortices. Compared with controls, patients showed a decreased activation of both sensorimotor cortices and supplementary motor area as well as a reversed lateralization effect. Increased understanding of the technology and elucidation of neural systems involved in the processing of tasks in healthy people should enhance our ability to apply this methodology to schizophrenia.

Effects of Medication The pharmacological status of patients undergoing metabolic and blood flow studies has varied. Research has ranged from investigations in which antipsychotic agents were considered a variable that needed to be controlled to those in which pharmacological intervention was introduced in a standardized fashion to examine treatment effects on the regional metabolic landscape. The washout period in studies that attempted to control the effects of antipsychotic drugs on cerebral blood flow and metabolism has commonly been short, ranging from 2 to 4 weeks. This period is a compromise between what is feasible and desirable. Monte Buchsbaum and colleagues examined cerebral glucose metabolism in cortical-striatal-thalamic circuits in a large sample of unmedicated men with schizophrenia. They found that patients had low metabolic activity in the medial frontal cortical regions and the basal ganglia, as well as an impaired lateralization pattern in the frontal and temporal regions. More recently in schizophrenia research, antipsychotic drug-naive first-episode patients have been studied. This population is particularly informative when the study is focused on the effects of pharmacological intervention. The study of neuroleptic-naive patients before pharmacological intervention separates the disease state from its treatment. The pattern of abnormalities summarized above is evident in first-episode patients across studies that examined differences between their first episode and episodes of previously treated patients. This suggests that disruption in normal brain processes is apparent at the presentation of illness and cannot be attributed to treatment or chronicity. While this is an informative approach, further progress can be made in metabolic studies using complementary methods to integrate pharmacological probes with metabolic studies.

A repeated-measures longitudinal design has been applied in a limited number of PET studies. In addition to examining symptom severity over time, this paradigm is singularly useful when pharmacological intervention is standardized. One study compared the effects of thiothixene (Navane) and haloperidol (Haldol) in chronic patients who were scanned off medication and after 4 to 6 weeks on medication. A different pattern of global and regional glucose metabolism was seen in the two groups. In another study PET scans were obtained at weeks 5 and 10 of a double-blind crossover trial of haloperidol and placebo in 25 patients with schizophrenia. Low relative metabolism in the striatum on placebo was associated with improved symptomatology. Responders to treatment had increased metabolism in the striatum after treatment. Nonresponders failed to show such a change and had more marked hypofrontality on medication. In a subsequent study, 12 patients were scanned before and 4 to 6 weeks after treatment with clozapine (Clozaril) or thiothixene. The drugs had a differential effect, with clozapine increasing and thiothixene decreasing metabolism in the basal ganglia, right more than left. Henry Holcomb and coworkers used a repeated-measures design to study glucose metabolism in 12 patients on a fixed dose of haloperidol and 5 and 30 days after drug withdrawal. No differences were observed between metabolism on medication and after 5 days of discontinuation. However, at 30 days, metabolism decreased in the caudate, putamen, and anterior thalamus and increased in the frontal cortex and anterior cingulate. The authors concluded that the basal ganglia are the site of the primary antidopaminergic action of haloperidol and that other changes observed are mediated through the cortical-striatal-thalamic pathways. The integration of pharmacological and neurobehavioral probes is a potentially powerful approach. For example, patients exhibited enhanced activation of the anterior cingulate after administration of apomorphine, suggesting a modulating role for dopamine.

Methodological Considerations and Potential Limitations Anxiety has complex effects on regional cerebral blood flow and metabolism, which investigators in a few laboratories have reported. It would seem desirable to measure anxiety carefully by use of complementary behavioral and psychophysiological procedures and to examine the relation of these measures to the regional metabolic and cerebral blood flow values and performance.

Motivation is an important factor in cognitive studies of schizophrenia. Whether poor performance can be improved by providing instructions and monetary reinforcement has been addressed in studies with the Wisconsin Card Sorting Test. One approach to this issue is the calculation of “mental effort” scores by subtracting basal cognitive abilities (e.g., I.Q. measures) from current performance. This difference between current and basal performance provides a measure of how well subjects perform in relation to their inherent ability, which may provide a parameter of motivation that can be related to cognitive and physiological data. This approach has been taken in ^{133}Xe and PET cerebral blood flow studies.

Task selection and choice of stimuli raise several questions. There are reasons to prefer elemental tasks that have been used extensively in cognitive psychology and are applicable across physiological measurements. A continuous presentation format of the tasks provides flexibility and ensures that subjects receive continuous stimulation during the measurement epoch. The importance of examining the issue of epoch has been recently demonstrated in a study by the Iowa group.

In many studies, self-paced task presentation was used with the hope of engaging the subject's utmost mental resources and efforts. This was considered essential for the ^{133}Xe and the PET flurodeoxyglucose studies, which integrate data across long periods of clearance (15 and 40 minutes, respectively). The disadvantage of self-paced administration for the PET cerebral blood flow and functional MRI measurements may, however, outweigh their advantage because of the brief measurement epoch. For brief duration, there could be considerable variability in the number of stimuli processed by subjects, and because patients with schizophrenia have slower initiation and response times, the differences in cerebral blood flow activation between patients and controls could be hard to interpret.

Central to the goal of relating regional cerebral blood flow change to task performance and clinical state variables is the problem of correlating behavioral data with physiologic data that are themselves intercorrelated. Innovative statistical approaches (e.g., Statistical Probability Mapping [SPM]) are used to address the global scaling factors inherent in this area of investigation.

NEURORECEPTOR STUDIES

Study of Neuroreceptors Another critical window for assessing brain function, the study of neuroreceptors, can give insight into the nature of neurochemical abnormalities in schizophrenia. Because advances in elucidating the pathophysiology of schizophrenia require understanding neurotransmitter function, the application of PET and SPECT to the study of receptor occupancy is likely to have an impact in the near future. These efforts are guided both by an extensive psychopharmacological literature and by advances in basic neuroscience on neuroreceptor subtyping. Functional neuroimaging is the meeting ground of preclinical and clinical neuropharmacology. Human neuroreceptor PET studies have built on progress with in vitro binding measurements of receptor density and affinity and neuroreceptor autoradiography. Psychotic symptoms seen in schizophrenia have been associated with dysfunction of dopamine, and the dopamine hypothesis has undergone revisions on the basis of these data.

PET Studies of D_2 Ligands The development of radioligands for PET studies first focused on the dopamine type 2 (D_2) receptor because of its clinical significance in relation to treatment with a neuroleptic agent. The study of antipsychotic drug-naive patients could potentially differentiate effects of the psychotic state before antipsychotic-drug intervention. Two major methodologies for quantitative measurement were developed and applied in the study of schizophrenia. Investigators at Johns Hopkins University applied [^{11}C]*N*-methylspiperone and reported that patients have higher D_2 Bmax values than controls. Studies at the Karolinska Institute, using [^{11}C]raclopride, reported similar B_{max} and K_d values in patients and controls.

These apparent differences have been discussed and summarized extensively and are likely related to multiple factors including patient variables, ligand properties, and PET modeling methods. Because the ligands differ in binding properties and sensitivity to endogenous dopamine, studies permitting a more direct comparison will be particularly helpful. In such an effort, Anna-Lena Nordstrom and coworkers evaluated the reproducibility of the [^{11}C]*N*-methylspiperone finding in a study of seven neuroleptic-naive patients and seven controls, before and after administration of 7.5 mg of haloperidol. Consistent with previous quantitative PET study of [^{11}C]raclopride binding, there were no differences between patients and controls pretreatment, and after haloperidol the specific binding of [^{11}C]*N*-methylspiperone was reduced by 80 to 90 percent. More recently, investigators at Johns Hopkins replicated the initial report in a new sample of drug-naive patients with schizophrenia. Other data reveal D_2 receptor density increases in psychotic, but not in nonpsychotic, patients with bipolar I disorder. The increase is comparable to that reported in schizophrenia. This raises questions regarding the specificity of the dopamine hypothesis to schizophrenia versus other psychotic syndromes.

Jean-Luc Martinot and coworkers measured D_2 striatal dopamine receptors using [^{76}Br]-bromospiperone in a PET study of 12 untreated patients with schizophrenia and found no increase in receptors in patients relative to controls. In a subsequent study, [^{76}Br]bromolisuride was applied to the measurement of striatal D_2 receptors in 19 untreated patients and 14 controls. Again, no differences in striatum-to-cerebellum ratios emerged, and no relation to symptoms or subtypes was evident in either study.

Receptor Function and Clinical Response The study of neuroreceptors can also address issues related to the relationship between receptor function and signs such as akathisia, commonly seen in patients treated with neuroleptic agents. Farde and colleagues determined in four control subjects the activity of [^{11}C]SCH 23390, a selective D_1 receptor antagonist. Two PET studies, at low and high doses of the radioligand, were conducted per subject. Transient akathisia occurred only when binding in the basal ganglia was at a high level with 45 to 59 percent occupancy. The D_2 receptor antagonist [^{11}C]raclopride was measured in 20 controls and 13 patients. Akathisia was associated with maximal ligand binding in the basal ganglia in patients and controls. Adam Wolkin and colleagues found that neuroleptic-resistant patients with schizophrenia did not differ from neuroleptic responders in degree of D_2 receptor occupancy by the antipsychotic agents. The regional distribution and kinetics of haloperidol binding were studied with [^{18}F]haloperidol in a PET study of five patients with schizophrenia examined while on haloperidol and after a drug washout and nine controls. Wide regional distribution of the ligand was evident in the cerebellum, basal ganglia, and thalamus, in contrast to the specific binding to the basal ganglia of [^{18}F]*N*-methylspiroperidol. Thus, small structural differences among butyrophenones are associated with changes in kinetics and distribution.

Typical and Atypical Antipsychotics PET neuroreceptor methods have also been applied in studies comparing atypical (serotonin–dopamine antagonists) and typical (dopamine receptor antagonists) antipsychotic drugs. The properties of clozapine binding to D_1 and D_2 receptors were examined in an open study of 5 patients, relative to 22 patients treated with dopamine receptor antagonists. Clozapine induced lower D_2 occupancy (38 to 63 percent), whereas D_2 receptor occupancy with dopamine receptor antagonists at conventional doses was 70 to 89 percent. Neuroleptic-induced extrapyramidal syndromes were associated with higher D_2 occupancy. In a follow-up study, Nordstrom and coworkers examined the relation between D_2 receptor occupancy and antipsychotic drug effect in a double-blind PET study using [^{11}C]raclopride. Seventeen patients with schizophrenia were randomly assigned to three groups treated with varied dosages of raclopride. A PET study was conducted at steady-state on 13 patients during the third to fourth week of treatment. A curvilinear relation between plasma concentration of raclopride and D_2 receptor occupancy was obtained. A significant relationship was noted between D_2 receptor occupancy and Brief Psychiatric Rating Scale percentage change as a measure of outcome. The D_2 receptor occupancy in patients who had extrapyramidal adverse effects was higher than in patients without. Nordstrom and coworkers examined D_1, D_2, and 5-hydroxytryptamine type 2 (5-HT_2) receptor occupancy in 17 patients treated with clozapine (125 to 600 mg a day) applying [^{11}C]SCH23390, [^{11}C]raclopride, and [^{11}C]*N*-methylspiperone. D_2 receptor occupancy (20 to 67 percent) was lower than for dopamine receptor antagonists (70 to 90 percent); D_1 receptor occupancy (36 to 59 percent) was higher than that reported for dopamine receptor antagonists (0 to 44 percent); and 5-HT_2 receptor occupancy was high (84 to 94 percent). Thus clozapine shows a combination of relatively high D_1, low D_2, and quite high 5-HT_2 receptor occupancy values, and serum concentrations are not predictive of receptor occupancy. In a PET study of [^{11}C]raclopride, Shitis Kapur and coworkers determined D_2 receptor occupancy induced by 2 mg a day of haloperidol for 2 weeks in seven patients. High levels of D_2 occupancy (53 to 74 percent) were noted with substantial clinical improvement. A similar investigation in nine patients receiving 2 to 6 mg a day of risperidone (Risperdal) showed receptor occupancy (66 to 79 percent) similar to that of dopamine receptor antagonists and higher than that of clozapine. When 10 patients with psychoses treated with loxapine were evaluated for D_2 and 5-HT_2 receptor occupancy, the agent differed from serotonin-dopamine antagonists. It has high 5-HT_2 receptor occupancy, which is not higher than D_2 occupancy. These research paradigms illustrate the integration of functional neuroimaging with pharmacological research. Incorporation of these strategies to psychopharmacological studies of schizophrenia with available therapeutic agents can advance the field and guide treatment intervention.

SPECT Studies The D_2 receptor SPECT ligand iodine-benzamide ^{123}I-iodobenzamide (IBZM) has been applied in studying dopamine D_2 receptors in patients with schizophrenia. Fifty-six patients were evaluated and a semiquantitative analysis of D_2 receptor binding was calculated (basal ganglia to frontal cortex ratio of activity). These ratios in patients taking typical neuroleptic agents were significantly lower than those in the neuroleptic-free subjects but not lower than those in the patients taking serotonin-dopamine antagonists (clozapine, remoxipride). No overall elevation of D_2 receptor binding was observed comparing 20 patients off medications and 20 controls, but a left lateralized asymmetry was found in male patients.

METABOLITE STUDIES: MAGNETIC RESONANCE SPECTROSCOPY (MRS)

MRS provides analytical qualitative and quantitative data on cellular metabolism and molecular structure. It has been used to study metabolism in vitro and in vivo in animals and humans. Spectral localization methods permit the measurement of ^{1}H and ^{31}P nuclear magnetic resonance (NMR) spectra from precisely localized volumes of interest, and this provides the basis for applying these techniques to study brain diseases. Because the technology is fundamentally similar to that used in MRI, several groups have begun to develop an approach that integrates these two modalities into a single examination. There are few reports that use this approach to investigate the underlying metabolism of neuropsychiatric disorders. Jay Pettegrew and colleagues pioneered applying phosphorus-31 (^{31}P) methods to the study of several neuropsychiatric diseases, including schizophrenia. They reported hypofunction in the dorsolateral prefrontal cortex in a sample of antipsychotic drug-naive patients with schizophrenia. The patients had significantly lower levels of phosphorus monoesters (PME) and higher levels of phosphorus diesters (PDE) than normal controls. Inorganic phosphate concentration was decreased and ATP concentration was increased in the patients. These latter results were interpreted as reflecting hypofunction of the dorsal prefrontal cortex in the patients. This interpretation is consistent with reports of decreased blood flow and decreased utilization of glucose in this region, as summarized above. A follow-up case report described a patient who exhibited in the ^{31}P MRS spectra PME and PDE levels similar to those reported for schizophrenia, well before the onset of psychotic symptoms. This finding led the authors to suggest that MRS may be of value in examining high-risk subjects such as family members of patients with schizophrenia for the presence of spectral abnormalities. Another group reported ^{31}P MRS results on patients with schizophrenia that support the findings of Jay Pettegrew and coworkers. Thus, a growing body of evidence from several laboratories shows converging findings.

This suggests the possibility of dysfunction in the normal process of programed synaptic pruning. Abnormal pruning could result in neuronal loss as well as upregulation of the postsynaptic dopaminergic receptors. These changes observed with ^{31}P clearly suggest that there should be alterations in the levels of the

compounds routinely detectable by localized proton MRS.

Considerable interest exists in obtaining solvent-suppressed proton spectra in humans. As the technical issues involved in obtaining spectra are being solved, it is important to begin to relate MRS measurements to the underlying biochemistry in the tissue being sampled. The role of *N*-acetylaspartate was reviewed by D.L. Birken and W.H. Oldendorf. This compound was found by NMR in glia and neurons but not on astrocytes. For this reason *N*-acetylaspartate concentration has been proposed as an index of neuronal integrity. The roles of the amino acids present in the brain have been examined. There is about 12 mM glutamate present in the brain, making it by far the most abundant amino acid. The rates of glutamate synthesis and oxidation differ in astrocytes and neurons. Two important products of glutamate are glutamine which is formed from glutamate by glutamate synthetase located in astrocytes and g-aminobutyric acid (GABA), an inhibitory neurotransmitter. Aspects of the metabolism of these compounds and the influence of this metabolism on MRS spectral appearance has received increased attention. Glutamate is largely confined to the tissue in which it is formed by barriers in permeability. Glutamine can be converted to glutamate at the site of neurotransmitter activity by glutaminase, which is present in neurons. The combination of two enzymes, glutamine synthetase (which converts glutamate to glutamine) and glutaminase, acts as a sort of cycle to maintain the concentrations of glutamate and glutamine. GABA and glutamate concentrations were determined by MRS in cultured preparations of cortical neurons and cerebellar granule cells, and colleagues granule cells contained large amounts of glutamate, while the neuronal cells contained large amounts of GABA.

Detection of these compounds in vivo in clinical studies showed increased glutamine concentrations in patients with chronic hepatic encephalopathy. Douglas Rothman and coworkers showed that it is possible to determine the glutamine concentration in human brain by spectral editing methods. The glutamine concentrations reported were in excellent agreement with literature values. These workers referenced their measurement of glutamate to the concentration of total creatine present (9.6 mM) in the brain. The key feature in their methodology was the use of short echo delays (12 milliseconds) to estimate the glutamate present. It seems therefore that MRS can be used to measure the concentrations of these amino acids, which may provide some insights into the activity of the excitatory glutamate and inhibitory (GABA) neurotransmitters present in the tissue being sampled. The studies so far are preliminary, since larger samples and a comprehensive and systematic approach to behavioral assessment are needed to link behavioral dimensions to both neuroanatomy and metabolism.

FUTURE DIRECTIONS

Structural and functional neuroimaging research in schizophrenia has made progress in advancing the understanding of neuroanatomical and neurophysiological substrates of this disorder. Structural imaging studies have identified subtypes of patients with reduced brain volume, and lower regional volumes have also been reported in structures that are key to healthy processing of complex behavior. While it is too early to outline with any precision the network of regions most affected, some consistent evidence has emerged implicating frontotemporal and corticostriatal thalamic regions. By and large, these structural abnormalities are present early in the course of illness and are related to disease features. These research methods can be applied to informative populations such as high-risk individuals and in genetic paradigms. Furthermore, recent evidence for progressive changes in some patients encourages longitudinal studies.

Two areas have been examined in functional imaging studies: energy metabolism and neuroreceptor studies. In a review of this field Goran Sedvall concluded that the major future contribution for understanding the pathophysiology of schizophrenia will be achieved through advanced resolution and development of new ligands for neurotransmitter systems. While the authors agree on the potential of these developments, they believe that metabolic studies can make unique contributions that will prove essential for finding the neural basis of schizophrenia and ultimately for improved treatment. In the context of the overall effort in neurobiological research in schizophrenia, functional neuroimaging studies have advanced the understanding of brain dysfunction related to neurobehavior and neuropharmacology. The field has reached some maturity in developing appropriate paradigms, and there is now a need for adequate sample size in patient and healthy populations, with attention to clinical heterogeneity and variability in brain function in relation to gender and age.

One of the major challenges in this research is the integration of neuroimaging data across anatomical and functional measures with clinical and neurobehavioral variables. A potential strength of functional neuroimaging is the integration of data on neuroreceptor function, metabolites, and metabolic activity. Ultimately, dysfunctional neurotransmitter systems translate to aberrant metabolism. Since cerebral blood flow and metabolism reflect neuronal activity, relating these domains is prerequisite to understanding the neurobiology of schizophrenia. As new receptor subtypes are cloned and radioligands are developed and available for human studies, it will be necessary to know which neuroreceptor measures result in increased neuronal activity and in turn how regional activation relates to behavior.

Thus, while new receptor ligands and improved resolution are welcome and exciting, as is the development of methods for MRS and flow measures, it is unlikely that the neural basis of schizophrenia will simply result from applying the right method with sufficient resolution. Rather, the harder and longer route of understanding the interaction among the brain's structural integrity, regional activity, and neuroreceptors as they affect the clinical and neurobehavioral manifestations of schizophrenia will probably be needed. On the positive side, this examination may yield partial answers of immediate benefit for treatment, and the evolution of this work will systematically improve our ability to articulate a neuropsychiatric perspective of this devastating disorder.

SUGGESTED CROSS-REFERENCES

The neural basis of schizophrenia psychopathology is further discussed in Section 1.1 an introduction and overview of neural sciences, in Section 1.2 on functional neuroanatomy, in Section 1.4 on monoamine neurotransmitters, in Section 1.10 on basic molecular neurobiology, and in Section 3.5 on brain models of mind. Neuroimaging is presented in Section 1.15 and Section 1.16. Typical antipsychotics drugs are presented in Section 31.17 on dopamine receptor antagonists. Atypical antipsychotics are covered in Section 31.26 on serotonin-dopamine antagonists. Other aspects of schizophrenia are discussed throughout Chapter 12.

SECTION REFERENCES

Andreasen NC, Arndt S, Swayze V II, Cizadlo T, Flaum M, O'Leary D, Ehrhardt JC, Yuh WT: Thalamic abnormalities in schizophrenia visualized through magnetic resonance image averaging. Science *266:*294, 1994.

Andreasen NC, Flashman L, Flaum M, Arndt ST, Swayze V II, O'Leary DS, Ehrhardt JC, Yuh WTC: Regional brain abnormalities in schizophrenia measured with magnetic resonance imaging. JAMA *272:*1763, 1994.

Bartha R, Williamson PC, Drost DJ, Malla A, Carr TJ, Cortese L, Canaran G, Rylett RJ, Neufeld RWJ: Measurement of glutamate and glutamine in the medial prefrontal cortex of never-treated schizophrenic patients and healthy controls by proton magnetic resonance spectroscopy. Arch Gen Psychiatry *54:*959, 1997.

Bertolino A, Nawroz S, Mattay VS, Barnett AS, Duyn JH, Moonen CTW, Frank JA, Tedeschi G, Weinberger DR: Regionally specific pattern of neurochemical pathology in schizophrenia as assessed by multislice proton magnetic resonance spectroscopic imaging. Am J Psychiatry *153:*1554, 1996.

Buchsbaum MS, Someya T, Teng CY, Abel L, Chin S, Najafi A, Haier RJ, Wu J, Bunney WE Jr: PET and MRI of the thalamus in never-medicated patients with schizophrenia. Am J Psychiatry *153:*191, 1996.

Cannon TD, van Erp TG, Huttunen M, Lonnqvist J, Salonen O, Poutanen VP, Standerstskjold-Nordenstam CG, Gur RE, Yan M: Regional gram matter, white matter, and cerebrospinal fluid distributions in schizophrenic patients, their sibliings, and controls. Arch Gen Psychiatry *55:*1084, 1998.

DeLisi LE, Tew W, Xie S, Hoff AL, Sakuma M, Kushner M, Lee G, Shedlack K, Smith AM, Grimson R: A prospective follow-up study of brain morphology and cognition in 1st episode schizophrenic patients: Preliminary findings. Biol Psychiatry *38:*349, 1995.

Dolan RJ, Fletcher P, Frith CD, Friston KJ, Frackowiak RS, Grasby PM: Dopaminergic modulation of impaired cognitive activation in the anterior cingulate cortex in schizophrenia. Nature *378:*180, 1995.

Farde L, Nordstrom AL, Wiesel FA, Pauli S, Halldin C, Sedvall G: PET analysis of central D1 and D2 dopamine receptor occupancy in patients treated with chemical neuroleptics and clozapine: Relation to extra-pyramidal side effects. Arch Gen Psychiatry *49:*538, 1992.

Frith CD, Friston KJ, Herold S, Silbersweigh D, Fletcher P, Cahill C, Dolan RJ, Frackowiak RS, Liddle PF: Regional brain activity in chronic schizophrenic patients during the performance of a verbal fluency task. Br J Psychiatry *167:*343, 1995.

Gur RC, Erwin RJ, Gur RE: Neurobehavioral probes for physiologic neuroimaging studies. Arch Gen Psychiatry *49:*409, 1992.

Gur RE, Jaggi JL, Shtasel DL, Ragland JD, Gur RC: Cerebral blood flow in schizophrenia: Effects of memory processing on regional activation. Biol Psychiatry *35:*3, 1994.

Gur RE, Maany V, Mozley D, Swanson C, Bilker W, Gur RC: Subcortical MRI volumes in neuroleptic-naive and treated patient with schizophrenia. Am J Psychistry *155:*1711, 1998.

Gur RE, Mozley PD, Resnick SM, Mozley LH, Shtasel DL, Gallacher F, Arnold SE, Karp JS, Alavi A, Reivich M, Gur RC: Resting cerebral glucose metabolism and clinical features of schizophrenia. Arch Gen Psychiatry *52:*657, 1995.

Gur RE, Mozley PD, Shtasel DL, Cannon TD, Gallacher F, Turetsky BI, Grossman RI, Gur RC: Clinical subtypes of schizophrenia differ in brain and cerebrospinal fluid volume. Am J Psychiatry *151:*343, 1994.

Gur RE, Pearlson GD: Neuroimaging in schizophrenia research. Schizophr Bull *19:*337, 1993.

Holcomb HH, Cascella NG, Thaker GK, Medoff DR, Dannals RF, Tamminga CA: Functional sites of neuroleptic drug action in the human brain: PET/FDG studies with and without haloperidol. Am J Psychiatry *153:*41, 1996.

Kapur S, Remington G, Jones C, Wilson A, DaSilva J, Houle S, Zipursky R: High levels of dopamine D_2 receptor occupancy with low-dose haloperidol treatment: A PET study. Am J Psychiatry *153:*948, 1996.

Kapur S, Zipursky R, Remington G, Jones C, McKay G, Houle S. PET evidence that loxapine is an equipotent blocker of 5-HT_2 and D_2 receptors: Implications for the therapeutics of schizophrenia. Am J Psychiatry *154:*1525, 1997.

Kotrla KJ, Weinberger DR: Brain imaging in schizophrenia. Ann Rev Med *46:*113, 1995.

Lim KO, Tew W, Kushner M, Chow K, Matsumoto B, DeLisi LE: Cortical gray matter volume deficit in patients with first-episode schizophrenia. Am J Psychiatry *153:*1548, 1996.

Nopoulos P, Flaum M, Andreasen NC: Sex differences in brain morphology in schizophrenia. Am J Psychiatry *154:*1648, 1997.

Nordstrom AL, Farde L, Eriksson L, Halldin C: No elevated D_2 dopamine receptors in neuroleptic-naive schizophrenic patients revealed by positron emission tomography and [^{11}C]N-methylspiperone. Psychiatry Res *61:*67, 1995.

Nyberg S, Farde L, Halldin D. Delayed normalization of central D_2 dopamine receptor availability after discontinuation of haloperidol decanoate. Arch Gen Psychiatry *54:*953, 1997.

O'Leary DS, Andreasen NC, Hurtig RR, Kesler ML, Rogers M, Arndt S, Cizadlo T, Watkins GL, Ponto LL, Kirchner PT, Hichwa RD: Auditory attentional deficits in patients with schizophrenia. A positron emission tomography study. Arch Gen Psychiatry *53:*633, 1996.

Pettegrew JW, Keshavan MS, Panchalingam K, Strychor S, Kaplan DB, Tretta MG, Allen M: Alterations in brain high-energy phosphate and membrane phospholipid metabolism in first-episode, drug-naive schizophrenics. Arch Gen Psychiatry *48:*563, 1991.

Petty RG, Barta PE, Pearlson GD, McGilchrist IK, Lewis RW, Tien AY, Pulver A, Vaughn DD, Casanova MF, Powers RE: Reversal of asymmetry of the planum temporale in schizophrenia. Am J Psychiatry *152:*715, 1995.

Rojas DC, Teale P, Sheeder J, Simon J, Reite M. Sex-specific expression of Heschle's gyrus functional and structural abnormalities in paranoid schizophrenia. Am J Psychiatry *154:*1655, 1997.

Sedvall G: The current status of PET-scanning with respect to schizophrenia. Neuropsychopharmacology *7:*50, 1992.

Shenton ME, Kikinis R, Jolesz FA, Pollak SD, LeMay M, Wible CG, Hokama H, Martin J, Metcalf D, Coleman M, McCarley RW: Abnormalities of the left temporal lobe and thought disorder in schizophrenia. A quantitative magnetic resonance imaging study. N Engl J Med *327:*604, 1992.

Siegel BV, Buchsbaum MS, Bunney WE Jr, Gottschalk LA, Haier RJ, Lohr JB, Lottenberg S, Najafi A, Nuechterlein KH, Potkin SG, Wu JC: Cortical-striatal-thalamic circuits and brain glucose metabolic activity in 70 unmedicated male schizophrenic patients. Am J Psychiatry *150:*1325, 1993.

Suddath RL, Casanova MF, Goldberg TE, Daniel DG, Kelsoe JR Jr, Weinberger DR: Temporal lobe pathology in schizophrenia: A quantitative magnetic resonance imaging study. Am J Psychiatry *146:*464, 1989.

Turetsky BI, Cowell PE, Gur RC, Grossman RI, Shtasel DL, Gur RE: Frontal and temporal lobe brain volumes in schizophrenia: Relationship to symptomatology and clinical subtype. Arch Gen Psychiatry *52:*1061, 1995.

*Velakoulis D, Pantelis C, McGorry PD, Dudgeon P, Brewer W, Cook M, Desmond P, Bridle N, Tierney P, Murrie V, Singh B, Copolov D: Hippocampal volume in first-episode psychoses and chronic schizophrenia. Arch Gen Psychiatry *56:*133, 1999.

Weinberger DR, Berman KF, Suddath R, Torrey EF: Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: A magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. Am J Psychiatry *149:*890, 1992.

12.4 SCHIZOPHRENIA: NEUROBIOLOGY

MICHAEL F. EGAN, M.D., AND THOMAS M. HYDE, M.D., Ih.D.

Schizophrenia is a chronic mental illness affecting approximately 1 percent of the population. Beginning in early adulthood, schizophrenia typically causes a dramatic, lifelong impairment in social and occupational functioning. From a public health standpoint, the costs of treatment and lost productivity make this illness one of the most expensive disorders in medicine. Despite the tremendous economic and emotional costs, research on schizophrenia lags far behind that on other major medical disorders. A primary impediment to developing more effective treatment is the limited understanding of the etiology and neurobiology of this disorder. New technologies, such as neuroimaging and molecular genetics, are removing the obstacles that once blocked major progress in the field. Although the stigma associated with the illness has not yet been eliminated, these new techniques have markedly altered the conception of the nature of schizophrenia.

One of the most rapidly changing fields is genetics. Family, twin, and adoption studies have clearly shown that genes play a prominent role in the development of schizophrenia. Estimates of heritability typically range from 50 to 85 percent. Initial attempts to isolate major genes using linkage studies were unsuccessful, but more recent approaches using increasingly sophisticated methods have uncovered several chromosomal regions that may harbor genes of minor effect. It seems likely that schizophrenia is the result of the interaction of many genes, some of which also interact with environmental factors. Investigations of environmental factors have looked at the role of stress, viruses, obstetrical complications, and in utero insults, among others. None of these have been definitively shown to be causative. It is possible that different combinations of genetic and environmental factors affect specific neurobiological systems, leading to a final common pathway of neural dysfunction.

Several neurobiological abnormalities have been found to have major implications for understanding the pathophysiology of schizophrenia. The first are structural brain abnormalities. Initially seen decades ago using pneumoencephalography, structural changes have been more clearly delineated using computerized tomography (CT) and magnetic resonance imaging (MRI). The most commonly reported alterations include enlarged lateral ventricles, enlarged third ventricle, and reduced volume of a number of structures, including hippocampus, amygdala, and frontal and temporal cortices. These abnormalities may predate the onset of illness. Second, functional cortical deficits have been seen with a variety of techniques, such as neuroimaging and neuropsychological testing. Prefrontal and temporal lobe dysfunction is most prominent, and is possibly related to structural abnormalities. Third, neuropathological studies have consistently failed to find any evidence of gliosis to account for the structural deficits. If anything, they tend to find subtle cytoarchitectural alterations. The recurring theme of this research suggests some type of failure in neuronal migration, orientation, or connectivity. Finally, several neurotransmitter systems appear to play a role, particularly in the expression of positive as well as negative psychotic symptoms. Evidence for alterations in the dopamine system is the most compelling. Other neurotransmitters have also been implicated, including glutamate, serotonin, and g-aminobutyric acid (GABA).

Neurochemical, structural, and functional imaging abnormalities can be understood in the context of the neural circuits involved and models of the illness. Cortico-striato-thalamic, limbic, and dopamine systems all appear to play a role. These three interconnected pathways mediate different aspects of higher-level information processing, such as judgment, memory, planning, and motivation. Their involvement could arise in several ways. One model suggests that neurodevelopmental abnormalities occur in utero. The clinical manifestations of schizophrenia appear only after brain development is largely completed, in late adolescence. Although this hypothesis has come to dominate thinking about schizophrenia, the neurodevelopmental model has several weaknesses.

ROLE OF GENES AND ENVIRONMENT

Genetic Factors Family, twin, and adoption studies indicate that there is a major heritable component to schizophrenia. Whereas the incidence in the normal population is approximately 0.5 to 1 percent, the lifetime risk in first-degree relatives is roughly 10 percent, indicating that the risk to first-degree relatives is 10 times that of the general population. This strongly implicates a familial factor in the etiology of the illness. Twin and adoption studies have shown that this is mostly, if not entirely, due to genetic factors. For example, the concordance rate in monozygotic twins is approximately 50 percent, as compared to 10 to 14 percent for dizygotic twins, suggesting that heritability may be as high as 80 percent. Of seven adoption studies, all found an increased incidence of schizophrenia in biological relatives, but not in adoptive relatives. This data convincingly demonstrates that genetic factors rather than shared, familial environmental factors are at work.

Although such epidemiological data implicate a major heritable component, the genetic architecture appears complex. Early attempts at modeling genetic transmission in families (using segregation analysis) suggested that heritability could not be explained by a single, dominant gene. In the early 1990s, increasingly sophisticated modeling indicated that at least several genes were involved, each with incomplete penetrance. One very real possibility is that there are many genes of minor effect. Such genes are difficult to detect using traditional linkage approaches. A triggering role for the environment in those with a genetic predisposition has also been hypothesized. While genetic modeling has been heuristically useful, the lack of a clear genetic mechanism complicates attempts to find the causative genes.

Early linkage studies were based on traditional assumptions that a single dominant gene produced the illness. These were, in general, unsuccessful. The first published study to use restriction fragment length polymorphisms (RFLP) reported linkage between two markers on the long arm of chromosome 5 (5q11-13) and schizophrenia. Subsequently, a number of other groups using separate cohorts were unable to replicate this, and several were able to clearly reject linkage to loci from 5q. While this failure dampened enthusiasm for genetic studies of schizophrenia, the relentless advances in statistical genetics and the molecular biology of the human genome have provided powerful new tools for detecting genes of minor effect. For example, some of the problems with specifying the unknown parameters needed for linkage analysis can be circumvented by using nonparametric approaches. These approaches use large collections of sibling pairs, both affected with the illness. As investigators have begun to use such tools, positive linkage reports for schizophrenia susceptibility genes of minor effect have been reported. Recent examples of putative schizophrenia susceptibility loci yielding some evidence of confirmation include loci on chromosomes 6, 8, and 22.

Despite real advances, several statistical issues continue to complicate interpretation of linkage studies. First, ambiguities persist about what diagnoses should be included. Family and adoption studies have suggested that diagnoses such as schizoaffective disorder, schizotypal personality disorder, and atypical psychosis are genetically related. To hedge their bets, investigators looking for linkage typically test several definitions of “schizophrenia spectrum,” ranging from narrow to very broad inclusion criteria. This means that more family members are included in the analysis as diagnostic criteria broaden. Second, the issue of which genetic model to use continues to plague parametric approaches. Typically, linkage studies include dominant, recessive, and mixed models. Here again, investigators hedge their bets by testing three or four genetic models. Since both problems lead to multiple testing, correction for multiple comparisons is indicated. Unfortunately, it is not entirely clear how to correct for this multiple testing. Currently, the commonly accepted significance level (*p* values) for initial linkage reports is ~10_{-4} to 10_{-5} or a logarithm of the odds (LOD) score of 3.3 and 0.01 for confirmation. Several groups have published putative replications of linkage findings based on these statistical criteria.

The first linkage study with some independent confirmation came from a study of a large Irish cohort. Using microsatellite markers and 186 multiplex schizophrenia families, evidence was found for linkage to the short arm of chromosome 6 (6p22). However, when adjusting for multiple comparisons, a genome-wide significance was estimated at .05 to .08 percent. When the original cohort was extended to 265 pedigrees, an LOD score of 3.51 was obtained, again using a moderately broad definition of illness. The LOD score was highest with a model of intermediate penetrance; of note, only 15 to 30 percent of pedigrees were linked. Supportive evidence for linkage to 6p22 was found in three independent studies. Interestingly, the three replications were different from the original in several ways; one used a recessive model and a narrow definition. In contrast to the original findings, a dominant model and broad disease definition yielded an LOD score of 0.06. A second replication study found suggestive linkage at a marker very close to the one in the original report. Not unexpectedly, a number of studies have failed to replicate the D6p22 linkage. These results illustrate some of the complexities of linkage studies of schizophrenia, but also provide some hope that these methods will uncover the genes involved in schizophrenia.

In addition to 6p22-24, at least two other regions have yielded evidence of linkage to schizophrenia. Ann Pulver and colleagues first described evidence for suggestive linkage to chromosome 8 at 8p22-p21 using 57 multiplex families. Soon after, another group, using a very broad definition of illness, reported confirmatory evidence for linkage using the Irish cohort; again, only 10 to 25 percent appeared to be affected by this putative susceptibility gene. A second attempt at replication by a multicenter collaborative group also found support for linkage. Suggestive evidence for a third potential vulnerability locus was reported for chromosome 22 at q12-q13.1.

Although the evidence for susceptibility loci in these reports does not overlap completely, the differences in location are not large. In other heritable complex diseases, for example, susceptibility genes have been cloned that are 20 centimorgans (about 20 million base pairs) away from the sites initially linked to the illness. As with 6p22-24, the strength of evidence for linkage to both 8p22-p21 and 22q12-q13.1 depends in part on what is acceptable as a significant replication. This in turn is related to how multiple tests are corrected for using a variety of phenotypes and model parameters. It is possible that there are schizophrenia susceptibility genes in a roughly 10 to 20 cM area in these regions that may each affect a small percentage of families. Several other regions have received attention but the evidence is less compelling. These regions include, at present, 5q, 6q, 9p, 18p, and 22q.

Using the candidate gene approach, weak support for involvement of the dopamine type 3 (D_3) receptor gene has emerged. In 1992, an excess of homozygosity was noted in schizophrenia patients compared to controls for a polymorphism in the first exon of the D_3 receptor gene. A few subsequent studies have supported a modest association between schizophrenia and homozygosity of the Ser-9-Gly polymorphism, but a large number of other studies failed to replicate this association. Linkage studies with D_3 receptor gene polymorphisms have not found significant LOD scores. As more functional variants in candidate genes are discovered, focused association studies of these genes will become increasingly common.

It is crucial to determine exactly what is inherited. One possibility is that genes determine susceptibility to certain environmental factors. Another possibility is that specific neurobiological abnormalities are produced by specific genes. Family studies have shown that relatives have an increased incidence of several neurobiological traits associated with schizophrenia. These include structural brain abnormalities, changes in evoked potentials, eye-tracking dysfunction, negative symptoms, and subtle cognitive deficts. These parameters could be more basic phenotypes that are closer to the molecular manifestations of the genes that cause schizophrenia. If so, they may improve the ability to detect these genes.

Environmental Factors Family-based epidemiological studies clearly demonstrate that environmental factors play a role in the pathophysiology of schizophrenia. The contributions of environmental factors have been estimated to be as much as 30 to 50 percent. Genetic modeling indicates that genes could set the threshold for liability to environmental factors. It is sobering to realize that environment can play a crucial role even in disorders that appear to be autosomal dominant. For example, phenylketonuria is an autosomal-dominant disorder that causes mental retardation. The illness is expressed, however, only if individuals with the abnormal gene ingest phenylalanine. Without this critical environmental exposure, the illness does not develop. Environmental factors hypothesized to play a role in schizophrenia range from problems with maternal bonding and early rearing to poverty, immigration status, stress, and viruses. The neurodevelopmental hypothesis has shifted the research focus somewhat from psychosocial variables to those that affect brain development. Several specific insults have been implicated, including pregnancy and birth complications, in utero viral infections (such as influenza), season of birth, and prenatal starvation.

Research into pregnancy, obstetric, and neonatal complications has had a particularly significant impact on the field. These complications include events such as prolonged labor, prematurity, preeclampsia, toxemia, fetal distress, and hypoxia. The majority of studies examining the incidence of such complications find increases in patients with schizophrenia. Positive studies include those that compare patients with matched controls, with their own well siblings, and even with a monozygotic twin discordant for schizophrenia. On the other hand, several impressive negative reports, including prospective, epidemiological surveys have failed to find a significant increase in such complications. However, these studies have been criticized on methodological grounds, thereby leaving the issue in doubt. Some authors have suggested that perinatal complications may increase risk only in persons with a genetic predisposition whereas others assert just the opposite. Although these conflicting findings make definite conclusions tentative, the bulk of the data suggests that perinatal complications are increased somewhat in patients with schizophrenia.

One possibility for how such complications lead to schizophrenia is that they produce some type of brain damage. The hippocampus, for example, is particularly susceptible to perinatal hypoxia and this limbic structure is thought to play an important role in schizophrenia. A number of studies have found that patients with a history of obstetric complications have increased likelihood of structural brain abnormalities, such as enlarged ventricles. A similar relationship has been seen in nonschizophrenic controls with a history of obstetric complications. However, many studies have failed to find any relationship between structural abnormalities and obstetric complications. Some authors have suggested that only nongenetic forms of the illness (sporadic cases) are more likely to have structural problems and obstetric complications, but data on this are very mixed. More problematic, obstetric complications are thought to mediate increased risk by transient hypoxia but hypoxia typically produces gliosis, a finding notably absent from the postmortem literature on schizophrenia. An alternative explanation is that obstetric complications are themselves secondary to abnormal fetal brain development. In any event, if obstetric complications increase the risk of schizophrenia, they are likely to be a minor factor; most persons with these complications do not develop schizophrenia and most patients with schizophrenia do not have an obvious history of obstetric complications.

A second risk factor that has been extensively studied is season of birth. There appears to be an increased incidence of schizophrenia associated with winter and spring birth dates. This finding is controversial, and has been attributed by detractors to a statistical artifact. If there is such a relationship, it could implicate an infectious process, such as a virus; viral infections are more common during winter months. Viral hypotheses have taken several forms, and candidates include slow viruses, retroviruses, or virally activated autoimmune reactions. In a related vein, several large-scale epidemiological studies have reported that the frequency of schizophrenia is increased following exposure to influenza during the second trimester. The effect is slight, however, and some studies have not observed this relationship. Another intriguing risk factor is starvation or poor nutrition. In studies of the effects of starvation during World War II in Holland, researchers found that starvation at the time of conception and in the first trimester increased the risk of developing schizophrenia by a factor of 2. Other factors recently reported to increase the risk of schizophrenia include Rh incompatibility and low intelligence quotient (I.Q.).

At this point, no single major factor has been unambiguously identified as an environmental cause of schizophrenia, and it is likely that none exists. As with genetic loci, environmental effects probably consist of a variety of factors, each having a minor effect at best. These will be difficult to detect, as will their hypothesized interaction with genes of minor effect, without large-scale studies.

STRUCTURAL AND FUNCTIONAL NEUROIMAGING

Neuroimaging studies of schizophrenia have demonstrated alterations in both structural and functional measures. Structural abnormalities include increased volume of the third and lateral ventricles, sulcal widening, and reduced volume of gray matter regions. Functional abnormalities include alterations in blood flow and measures of chemical moieties using MRI spectroscopy. Neuroimaging has also been used to assay receptor density and dynamic parameters related to dopamine release. Neurochemical studies are discussed separately in sections on specific neurotransmitters.

Neuroimaging has had a major impact on the conceptualization of schizophrenia. The notion that patients with schizophrenia have an actual deficit in the volume of brain tissue clearly established that this was a brain disease rather than a purely psychological or biochemical disorder. Functional neuroimaging has implicated the prefrontal and temporal lobes in particular, and has begun to relate activity in these regions to the clinical manifestation of schizophrenia. As critical as these findings have been, important controversies remain.

Structural Abnormalities One of the most widely replicated neurobiological findings in schizophrenia research is that of altered volume of cerebral structures. Increased size of the cerebral ventricles and reduced brain volume were observed early in the twentieth century using pneumoencephalography and postmortem material. These early findings, however, had little enduring impact on the field. The advent of CT technology renewed interest in cerebral volumetric parameters. The earliest CT studies found enlargement of the lateral and third ventricles and cortical sulci. Although these findings were initially viewed with skepticism, over 100 subsequent studies have been published with lateral ventricular enlargement reported in 75 percent, third ventricular enlargement in 83 percent, and cortical changes in 67 percent. Concerns that ventricular enlargement could be secondary to factors such as antipsychotic medications, institutionalization, and diet have generally been ruled out. Furthermore, studies using MRI, with its markedly enhanced resolution, have confirmed the presence of lateral and third ventricular enlargement and provided estimates of tissue loss to be roughly 3 to 10 percent.

The finding of ventricular enlargement dramatically shifted the focus of research on schizophrenia. Subsequently, several critical questions have dominated this landscape. First, is ventricular enlargement caused by focal areas of tissue loss or a more generalized process? Second, do the structural abnormalities predate the

onset of the illness, implicating a neurodevelopmental process, or do they arise concomitantly with the illness, suggesting a neurodegenerative process? Third, are all patients affected or only a subgroup? Finally, what are the functional implications of these abnormalities?

To localize brain abnormalities, researchers have looked at a variety of measures, including cortical sulcal enlargement, ventricular enlargement, and quantitative measures of individual brain structures. Regarding cortical sulcal enlargement the data are split, with some reporting sulcal enlargement in the frontal and temporal lobes whereas others have found more diffuse enlargement. More specific measures of cortical volume typically show reductions of temporal and, less consistently, frontal lobe volume. These reductions involve gray rather than white matter, although some studies have found reductions in white matter as well. Regional volumetric studies of specific brain structures have generally focused on the temporal lobes. Bilateral volume reductions in amygdala-hippocampus, parahippocampal gyrus, entorhinal cortex, and superior temporal gyrus have been reported. In the ventricular system, increased volume in the temporal poles of the lateral ventricles has been found most often; increased volume of the frontal horns and third ventricle is also commonly found. In quantitative studies of subcortical regions, findings have been mixed. Some researchers find no changes in areas such as caudate, putamen, nucleus accumbens, and external segment of the globus pallidus; others have reported increased striatal volume and reduced globus pallidus (internal segment) volume. Increased striatal volume is thought to be an effect of treatment with antipsychotic medications. Reduced volume of the thalamus has also been observed.

The notion that the temporal and frontal lobes may play a particularly important role in schizophrenia has been supported by findings from other areas. For example, neurological damage to the temporal lobes sometimes produces positive psychotic symptoms, such as hallucinations, while damage to the frontal lobes is associated with negative symptoms such as apathy, social withdrawal, and blunted affect. On neuropsychological testing, patients with schizophrenia typically show impaired frontal and temporal lobe function. More recently, magnetic resonance spectroscopy has been used to examine these regions. This new technology can measure in vivo concentrations of a variety of neurochemical moieties. These include *N*-acetyl aspartate (NAA), an intraneuronal amino acid sensitive to mitochondrial energy metabolism and to pathological processes affecting neuronal integrity, choline-containing compounds, creatine plus phosphocreatine, glutamate, glutamine, and high-energy phosphate-containing compounds. Several intriguing findings have emerged. First, specific reductions of NAA have been observed in the dorsolateral prefrontal cortex and hippocampal area, probably reflecting neuronal pathology in these locations. Other areas are, for the most part, unaffected. Second, an imbalance between phosphomonoesters and phosphodiesters has been described in the frontal cortex. These studies, combined with volumetric data, lend support to the theory that there may be selective deficits in frontal and temporal regions.

Attempts to pinpoint when volumetric alterations occur have led to studies of patients at the onset of their illness. This issue is crucial to understanding what neurobiological processes could possibly account for structural abnormalities. In general, first-break studies have found the same alterations seen in prior studies of patients with chronic schizophrenia. These results are supported by the lack of relation between volumetric alterations and duration of illness or age of onset seen in studies of such patients. If an active process produced tissue loss, the loss would be correlated with illness duration, which it is not in most studies. On the other hand, cognitive deficits associated with schizophrenia do not progress but probably develop very early in the illness. Although portions of these deficits may be present even in childhood, a significant component probably develops sometime around the onset of the illness. It is not inconceivable that structural abnormalities could develop at the same time. Such changes would not necessarily be detected by first-break studies. Another approach has been to scan first-break patients when they initially present for treatment and then again several years later. The results have been mixed: some find no changes whereas others have suggested that a subgroup of patients do show slight, but progressive tissue loss. The latter approach has been criticized on methodological grounds and certainly more studies are needed. At present, it seems fairly certain that structural abnormalities are present from very early on in the illness.

A third issue is whether structural alterations are present in all patients or only a subgroup. Several early studies had found associations between ventricular enlargement and a variety of clinical characteristics, including poor premorbid adjustment, age of onset, cognitive impairment, negative symptoms, poor response to antipsychotic medications, and greater incidence of tardive dyskinesia. Such observations have led to suggestions that there are two forms of schizophrenia, one involving a hyperdopaminergic state and the other involving structural abnormalities. Since then, many CT and MRI studies have examined this issue but have generally failed to confirm this schema. Structural abnormalities do appear to be correlated to some degree with cognitive impairment and negative symptoms, but these correlations are not particularly robust. Another approach to subtyping has been to look at the distribution of these deficits. In a meta-analysis of studies that have used CT scans to evaluate ventricular enlargement, the lack of a bimodal distribution in over 1000 patients suggests that a clear subgroup with these abnormalities does not exist.

An elegant attempt to assess the frequency of structural abnormalities was provided by a study of discordant monozygotic twin pairs. Unaffected monozygotic twins serve as an ideal control for assessing illness-related changes. In an MRI study of 15 such pairs, the ill twin had more pronounced deficits for most structural measures in over 85 percent of cases. These findings are similar to those of a prior twin study using CT scans. The data suggest that volumetric abnormalities in schizophrenia are very common, if not ubiquitous; detecting the abnormality may depend on having a perfectly matched genetic control (Fig. 12.4-1) because patients with normal ventricular volume were often seen to have significantly larger ventricles than their unaffected twin. However, when this MRI study of twins was expanded to 27 discordant pairs, lateral ventricular enlargement was only seen in about 63 percent of the affected twins relative to the unaffected twins. This is only somewhat higher than 50 percent, which is what would be expected by chance. In this expanded sample, it appears that ventricular enlargement may not be universal. In contrast, hippocampal measures continued to predict the affected twin in roughly 80 percent of cases, which is significantly higher than the 50 percent chance level; this suggests that hippocampal pathology is common. Although the exact percent of patients having structural abnormalities is not known, it is probably fairly high. An alternative view is that structural abnormalities represent a quantitative trait that is commonly associated with schizophrenia but neither necessary nor sufficient to produce the illness.

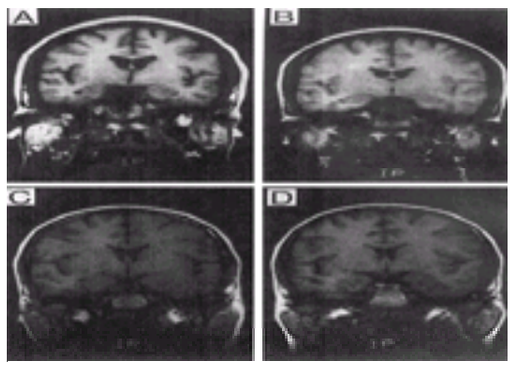

FIGURE 12.4-1 MRI scans (coronal sections) of two sets of discordant monozygotic twins (**A** and **B** = set 1; **C** and **D** = set 2). For each pair, one has schizophrenia (**A** and **C**) while the other does not (**B** and **D**). For both pairs, the affected twin has larger ventricles than the unaffected twin, even though ventricular size appears to be within the normal range for the affected twin (C). (Courtesy of D. Weinberger and E. F. Torrey.)

In summary, structural abnormalities, such as enlarged ventricles and reduced cortical volume, are a prominent feature of schizophrenia. It is unclear whether cortical involvement is multifocal or diffuse. Temporal and frontal lobe regions are certainly involved. These abnormalities are present very early in the illness. It is too early to say, however, whether they are present from birth or develop at a later stage. Structural abnormalities may be present in a majority of patients, although the exact percentage is unknown. The prevalence is most apparent when compared to ideally matched genetic controls. Structural abnormalities are correlated to some degree with clinical aspects of the illness, such as cognitive deficits. A key issue remains unresolved: what neurobiological processes account for these enigmatic changes?

Functional Neuroimaging Functional neuroimaging refers to a group of methods that look at changes in regional neural activity by measuring regional cerebral blood flow (rCBF) or glucose utilization. These two parameters can be measured with several techniques, including positron emission tomography (PET),* (SPECT), and more recently functional MRI (fMRI), each having its own particular advantages and disadvantages. These techniques have been used to explore brain regions that may be dysfunctional in schizophrenia. Several designs have been employed: (1) patients and controls are compared at rest; (2) they are compared during cognitive testing that normally increases activity in a particular brain region; and (3) brain activity is correlated with psychiatric symptoms, either cross-sectionally among patients or within a patient over time. The most consistent finding is reduced activation of the prefrontal cortex (hypofrontality), but other regions, such as the temporal lobes, have also been implicated. Also, correlations have been found between specific symptom clusters and regional activity in both frontal and temporal areas (Fig. 12.4-2).

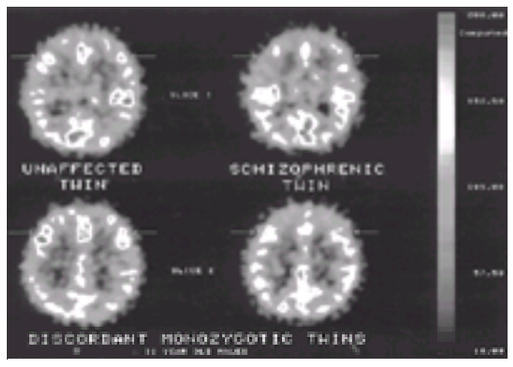

FIGURE 12.4-2 PET scans using H_2O_{15} of two monozygotic twins, one with (right) and one without (left) schizophrenia. Top and bottom scans show two levels through the dorsolateral prefrontal cortex. At the time of scanning, subjects are performing a cognitive task that typically requires prefrontal cortical function. The affected twin blood flow to the dorsolateral prefrontal cortex is markedly reduced compared to the unaffected twin. (Courtesy of R. Berman and D. Weinberger.) (See Color Plate 8.)

Frontal lobe function has been studied most intensively. Initially reported in 1974, the finding of reduced frontal blood flow has been controversial. Many studies, particularly those looking at the resting state, have not found evidence of hypofrontality. Such studies have been criticized, however, because the resting state is an uncontrolled feature and may introduce unnecessary variability. Using cognitive tasks that appear to require prefrontal activation in controls, a number of studies have consistently found that patients with schizophrenia fail to increase blood flow to this region. Although many resting studies have reported hypofrontality, most, if not all, studies using activation tasks have found hypofrontality; this suggests that the use of activation tasks can increase the sensitivity of these procedures to detecting abnormalities by assessing function of regions involved in the illness (Fig. 12.4-2).

The finding of hypofrontality in schizophrenia has often been interpreted as an artifact of poor performance, motivation, clinical state, medications, or other factors. However, studies have not shown that these factors account for differences between patients and controls. For example, poor performance on working memory tasks is not necessarily associated with reduced prefrontal blood flow. Patients with Huntington's disease and groups with low I.Q. who do equally poorly on prefrontal cognitive tasks, are able to activate the dorsolateral prefrontal cortex. Interestingly, hypofrontality appears to be correlated with several structural and neurochemical indices. Prefrontal activation is highly correlated with homovanillic acid (HVA) concentrations in cerebrospinal fluid, possibly reflecting prefrontal dopamine activity. Hypofrontality has also been correlated with hippocampal volume in one study of discordant monozygotic twins, suggesting a dysfunctional circuit. Finally, preliminary reports suggest that reduced prefrontal NAA concentrations, markers of neuronal integrity, are correlated with reduced frontal activation. These data imply that hypofrontality could result from a process that affects neuronal viability in both frontal and hippocampal regions and that these have downstream effects on the regulation of prefrontal dopamine.

The temporal lobe has also been examined with functional neuroimaging techniques. Both elevated and reduced blood flow has been reported. The most common finding is an association between resting blood flow and positive psychotic symptoms. For example, one report found a correlation between increased psychopathology and blood flow to the left parahippocampal gyrus; a second found a similar correlation between positive symptoms and left temporal lobe blood flow. More specific correlations have been seen for auditory hallucinations and activation of Broca's area and medial temporal regions. A potential criticism of this finding is that patients may have simply been responding to auditory hallucinations with their own vocalizations. Activation of Broca's area, in this case, would be expected and trivial. Research into the relation between symptom clusters and blood flow revealed that positive symptoms were associated with increased medial temporal flow, negative symptoms with decreased prefrontal (dorsolateral) blood flow, and disorganization with increased cingulate flow. This parcelation of symptoms with neuroanatomy suggests that separate but related neurophysiological processes may underlie specific types of symptoms. The few studies that examine several regions simultaneously tend to find changes in the coordinated activity between regions, particularly between prefrontal and temporal areas. Typically, increased activation in temporal areas is found in functional connectivity of the two regions.

One report on other brain regions found increased left globus pallidus activity at rest; others have reported both decreased and increased glucose utilization in the striatum (Fig. 12.4-3). Antipsychotic medications appear to increase striatal metabolism, suggesting that medications are an important confound. Reduced cingulate activation has also been described. As newer techniques that do not depend on radioactivity, such as fMRI, are more commonly used, further characterization of these and other brain regions can be expected.

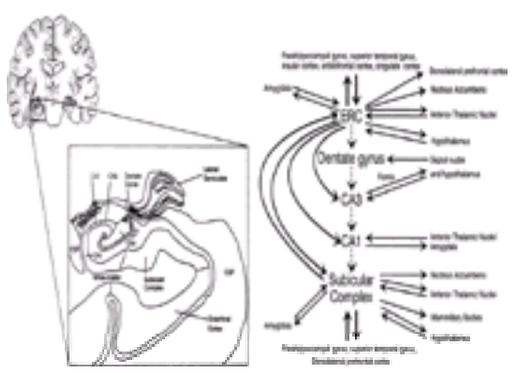

FIGURE 12.4-3 A, Schematic diagram of the mesial temporal lobe at the level of the body of the hippocampus and posterior entorhinal cortex, in coronal section. **B**, The illustrated connections of the coronal section are described in this table. (Drawn by Kyle Christensen.)

In summary, blood flow to several brain regions, including prefrontal and temporal areas, is altered in schizophrenia. These changes may be related to or may underlie positive and negative symptoms as well as some cognitive deficits. Regional abnormalities may also be related to each other, indicating a more global problem with the function of the larger systems or neural networks. Many questions remain. How closely are the changes in temporal and prefrontal activity associated with the clinical features of schizophrenia? Is the activation of other brain regions affected? Can functional brain imaging pinpoint which brain areas cause specific symptoms? What neurobiological processes account for differences in brain function? Correlations with structural abnormalities, dopamine metabolites, and regional NAA levels suggest that these variables could play a role.

NEUROPATHOLOGY

The neuropathological basis of schizophrenia remains obscure despite an increased number of techniques applied to the investigation of this subject. The future appears bright, however, as more laboratories across the world become engaged in this research. Regions that have become the focus of postmortem studies include temporal and limbic structures (hippocampus, amygdala, hypothalamus, nucleus accumbens, and cingulate cortex), and prefrontal and orbitofrontal cortices. Other paralimbic structures recently have been added to the neural network thought to be dysfunctional in schizophrenia, including the ventral tegmental area, substantia nigra, anterior thalamic nuclei, and entorhinal cortex. With this focused approach a number of intriguing findings have emerged; almost all still need independent replication, and the confounds of antemortem exposure to antipsychotic drugs must be considered when reviewing these studies.

Temporal Lobes

Mesial Structures Perhaps the one region that has received the greatest attention in postmortem schizophrenia research is the mesial temporal lobe, which contains the entorhinal cortex, amygdala, and hippocampal formation (Fig. 12.4-3). These structures have been examined in both morphological and neurochemical studies. The entorhinal cortex, which relays cortical input into the hippocampus and distributes output from the hippocampus to a diverse group of brain structures, has been carefully scrutinized. The laminar distribution of neurons in the superficial layers of the rostral entorhinal cortex has been reported to be abnormal and disorganized by

several independent groups of investigators. One study in particular has suggested that the subtle changes in neuronal aggregation may be restricted to layers II and III. Taken together, these data suggest a mild disruption of normal cytoarchitectural features. Although it may not be impossible for this to occur later in life, the findings would strongly support the notion that abnormal neuronal migration may occur during brain development in patients with schizophrenia.

The finding that cytoarchitectural abnormalities are present in the entorhinal cortex have recently been contested by two carefully controlled, anatomically precise studies. Both studies failed to find the abnormal cytoarchitectural features described previously and suggested that earlier reports may have been confounded by incomplete matching of sections from normal controls and individuals with schizophrenia. The normal cytoarchitecture of the entorhinal cortex markedly changes along its rostrocaudal extent, making the issue of appropriate matching critical. However, the entorhinal cortex may not be entirely normal in schizophrenia; one study found a limited reduction in neuronal number and density. This is consistent with other reports of smaller volume, a reduction in the number of neurons, and volumetric measures using MRI.

Neurochemical elements that subserve the anatomic integrity of a given brain region have also been measured, as an indirect assessment of the cytoarchitecture. Microtubule-associated proteins (MAPs) are important elements of the neuronal cytoskeleton. One recent study found a marked loss of MAPS immunoreactivity in the subiculum and the entorhinal cortex in schizophrenia. This finding was interpreted as support for and evidence of cytoarchitectural abnormalities in this mesial temporal lobe. However, given the qualitative nature of most immunostaining techniques, direct replication and additional investigations with more quantifiable strategies are needed.

Synaptophysin is a synaptic vesicle protein, and as such is widely distributed throughout the central nervous system. Levels of synaptophysin or its mRNA on both can be used as indices of synaptic density. Decreased synapsin I, but not synapsin IIb or synaptophysin, has been found in the hippocampus of patients with schizophrenia. A more recent report noted a reduction in synaptophysin messenger ribonuclei acid (mRNA) in CA4, CA3, subiculum, and the parahippocampal gyrus. There were no changes in synaptophysin in these regions, however, suggesting that the loss of synapses may occur at extra-hippocampal sites. Alternatively, local circuits within the hippocampus may be compromised but the ability to detect these changes is limited by the volume of extra-hippocampal input to this brain region. In any event, this finding is another element in the emerging picture of structural alterations in the mesial temporal lobe.

Hippocampus The hippocampus, the predominant structure within the mesial temporal lobe, also may have anatomic abnormalities. Postmortem studies of the hippocampus have proliferated since the mid-1980s. One group found a volume reduction in the whole hippocampal formation in schizophrenia. Others, however, have reported that decreased volume is restricted to the white matter of the left hippocampus, or in the volume of the CA4 subfield. A number of other postmortem studies have found subtle structural abnormalities in the hippocampal formation in schizophrenia, providing a relatively robust body of evidence implicating alterations of the hippocampal formation in schizophrenia.

Within the pyramidal cell layer of the hippocampus, the most recognizable microscopic feature is the orientation of pyramidal cells. While cellular disarray in the CA1-prosubiculum and CA1-CA2 interface has been observed by one group, at least three other groups were unable to replicate this finding. Decreased numbers of pyramidal cells in hippocampal subfields and reduced neuronal size (in left CA1 and CA2, and right CA3) have also been found. These are both consistent with prior MRI findings. Alteration in the density of staining of the mossy fibers in the hilus of the dentate gyrus, and several hippocampal subfields have been seen as well. However, this finding is surprising because cell loss in the adjacent entorhinal complex should lead to an increase in the staining density of the mossy fibers. Finally, decreased polysialic acid-neural cell adhesion molecule (PSA-NCAM) immunoreactivity has been reported in the CA4 subfield of the hippocampus in schizophrenia. PSA-NCAM, a cell adhesion molecule, is thought to be important in synaptic rearrangements in adulthood. Although no clear consensus has emerged on the nature of pathological change within the hippocampus proper, there is abundant evidence of structural abnormalities.

Amgydala The amygdala, located within the mesial temporal lobe, has major interconnections with the entorhinal cortex and hippocampus, as well as many other structures. The amygdala appears to have a smaller volume in schizophrenia patients; this finding is in accordance with postmortem reports.

Prefrontal Cortex Postmortem studies of the prefrontal cortex have been stimulated by the deficits observed with in vivo neuroimaging. One recent study found increased neuronal density in prefrontal area 9; a change of similar magnitude was observed in occipital area 17 as well, suggesting a widespread pathological process. This finding was interpreted as representing a loss of neuropil throughout the cortex in schizophrenia without accompanying gliosis. Area 9 has also been shown to have a smaller average neuronal size and an increased density of smaller neurons, with unchanged glial size and density. The absence of gliosis again suggests that the pathological change in schizophrenia is probably not an active inflammatory process. Area 17, visual cortex, did not show any of these abnormalities, suggesting some anatomic specificity of this finding. In addition to smaller neuronal size, layer 3 pyramidal cells may have diminished dendritic spine density, which in part may explain the abnormalities in neuropil noted by others. Finally, area 46, prefrontal cortex adjacent to area 9, also has increased neuronal density in layers 2, 3, 4, and 6, and a thinning of layer 2. Taken together, these studies suggest a loss of neuropil in the prefrontal cortex, and abnormalities in the cellular constituency of this region.

A somewhat murky picture has emerged from studies of the distribution of neurons in the subcortical white matter underlying the prefrontal cortex. Such neurons are thought to represent a vestige of neuronal migration during early brain development. One group found an increased density of nicotinamide-adenine dinucleotide phosphate diaphorase-positive neurons in the deep white matter and a lower density in the superficial white matter underlying the superior and middle frontal gyri. This is consistent with a developmental arrest in the migration of cortical neurons from deeper white matter areas to superficial cortical layers. A second, similar study looked at MAP2-immunoreactive neuron distribution in the subcortical white matter underlying area 46 and the transition zone between areas 46 and 9 in the prefrontal region. Patients with schizophrenia had a greater density of MAP2-immunoreactive neurons in the superficial white matter compared to controls. In contrast to the first study, no differences are seen in deeper white matter. This was interpreted as either abnormal expression of MAP2, a defect in neuronal migration, a failure of programmed cell death, or a decrease in white matter volume in schizophrenia patients. Although these two studies looked at different neuronal subpopulations, the different findings are contradictory and must be interpreted with caution.

Orbitofrontal Cortex The orbitofrontal cortex has also come under scrutiny, at least in part because of interconnections with a variety of limbic system structures and the efficacy of leukotomy in the treatment of some clinical aspects of schizophrenia. In area 10, orbitofrontal cortex, a decrease in neuronal number, maximal in layers 4 and 5, and in cortical thickness has been observed in a small sample of schizophrenia subjects. A similar reduction in areas 4 (frontal), 24 (cingulate), and 17 (occipital), has also been seen, suggesting a pancortical process. A more recent study found a significant reduction in neuronal density in layer 6 of area 10, but also in layer 5 of area 24 (cingulate cortex) and layer 3 of area 4 (primary motor cortex). The meaning of changes in such disparate layers cannot be easily explained, especially in light of the findings in areas 9 and 17.

Neurochemical analyses also have been performed on the prefrontal cortex as an index of structural integrity. One group examined the concentrations of synaptic vesicle associated protein-25 (SNAP-25) a synaptosomal associated protein involved in neurotransmitter release. Using quantitative Western blots, they found an elevation in SNAP-25 concentrations in area 9, reductions in areas 10 and 20 (temporal cortex), and no change in area 17. Such findings could be due to either a change in synaptic density or to an abnormality in neurotransmitter release; the former interpretation may account at least in part for the decreased neuropil in area 9.

Cingulate Cortex The anterior cingulate cortex (area 24) is part of the neural network subserving the cortical regulation of emotion and attention, both of which appear to be deficient in schizophrenia. In a series of postmortem studies, one group demonstrated an increase in vertical axon number in the cingulate cortex of schizophrenia patients. These researchers have also reported abnormalities in neuronal aggregation in layer 2 of area 24 and a decrease in the number of interneurons in layers 2-6 of this region. Others have seen an abnormality in the usual asymmetry of weight and surface area for the anterior cingulate cortex; independent replication of these findings will be important.

Other Regions Subcortical structures also may have an abnormal anatomy in schizophrenia. Consistent with MRI studies, the mediodorsal nucleus of the thalamus may have fewer neurons in schizophrenia patients in comparison to controls. Studies of the basal ganglia are somewhat limited. Whereas one study did not find any absolute volume differences in the striatum as a whole or individually in the caudate, putamen, or nucleus accumbens, a second group reported an increase in left striatal volume in schizophrenia patients. A third report on the ultrastructure of the caudate nucleus using electron microscopy ound abnormalities in synaptic morphology and dystrophic and reactive changes in astrocytes. Regarding midbrain dopaminergic nuclei, decreased volume of the lateral substantia nigra, and a decrease in the average volume of the nerve cell bodies in the medial segment have been observed. Several other studies have found no significant brainstem pathology or relatively nonspecific findings. Clearly, more research needs to be devoted to the brainstem, given the importance of ascending catecholamine and serotonin systems in regulating the activity of forebrain structures, and the clinical data implicating these neurochemical systems in schizophrenia.

Gliosis Of all these subtle yet potentially important cytoarchitectural findings, one of the most critical observations is the apparent absence of gliosis. The importance of this stems from theoretical implication that reduced volume of brain regions and other abnormalities are not the result of an active pathological process: instead, they are likely to be secondary to very early developmental processes. The issue of whether gliosis is present has been addressed by many postmortem studies over

the past century. Of these, at least a dozen recent studies have used methodologically superior quantitative techniques. While several have noted increased gliosis, the large majority has found no differences between brains from patients with schizophrenia and those from normal controls. These include studies using several different techniques for counting glial cell number, such as the Holzer stain, Nissl stain, and immunoreactivity for glial fibrillary acidic protein. Some methodological questions about the ability of some techniques to detect the effects of chronic gliosis persist; it seems unlikely, however, that clinically relevant gliosis would be obscured.

The wide variety of potentially important findings must be approached with a healthy skepticism. Several common problems plague almost all postmortem volumetric and cell counting studies in schizophrenia. First, standard stereological techniques, using serial sections at regular intervals through the rostrocaudal extent of the mesial temporal lobe, are infrequently applied. Fortunately, more recent studies are employing stereology with greater frequency. Moreover, rarely, if ever, is the time of fixation carefully controlled, so that there is a wide variation within and across studies. Tissue shrinkage, which affects tissue volume and cell density, and maybe quality of cell staining, varies with the duration of fixation. Nevertheless, postmortem studies point to subtle volume reductions in the hippocampal formation in schizophrenia. The precise neuropathological changes that underlie this volume reduction remain controversial.

NEUROCHEMISTRY

Dopamine One of the most important observations in twentieth-century psychiatry is that dopamine antagonists ameliorate symptoms of schizophrenia. The implication that too much dopamine causes psychosis has dominated research for well over two generations and continues to exert a profound impact. In its most basic form, the dopamine hypothesis states that an excess of subcortical dopamine neurotransmission leads to psychotic symptoms. Observations that the prefrontal cortex modulates subcortical dopamine release have established a compelling link between cortical abnormalities and changes in the dopamine system. A current version of the dopamine hypothesis is that dopamine is dysregulated; levels may be reduced in the prefrontal cortex and altered in complex ways in subcortical and limbic regions. Reduced cortical dopamine could explain hypofrontality, impaired cognition, and negative symptoms (such as anhedonia and lack of motivation). Altered subcortical and limbic dopamine, on the other hand, could cause positive symptoms (such as hallucinations and delusions). Theories about the role of dopamine in schizophrenia have advanced in tandem with the increased understanding of the neurobiology of dopamine.

Neurobiology of Dopamine Dopamine (Fig. 12.4-4) is synthesized from tyrosine through dopa. The first step, the conversion of tyrosine to dopa by tyrosine hydroxylase, is the rate-limiting step, and is subject to feedback regulation. The major metabolic product of dopamine catabolism in humans is homovanillic acid, and, to a lesser extent dihydroxyphenylacetic acid and 3-methoxytyramine. Concentrations of these metabolites have been examined in the brain, cerebral spinal fluid (CSF), plasma, and urine of patients with schizophrenia to look for evidence of increased or decreased dopamine neurotransmission.

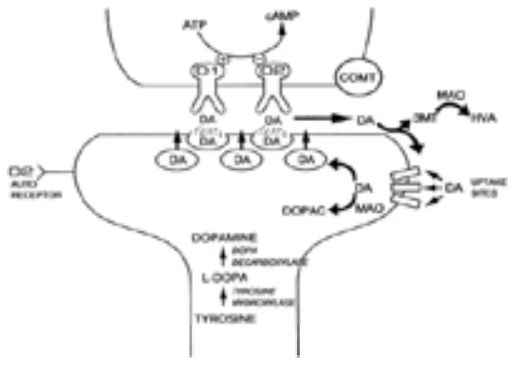

FIGURE 12.4-4 Dopamine metabolism and synaptic structure. In this schematic synapse, dopamine is released into the synaptic cleft where it can act on D_1 or D_2 postsynaptic receptors. Synaptic dopamine is inactivated by reuptake pumps or by catabolism via COMT and MAO. Presynaptic D_2 autoreceptors modulate dopamine synthesis and release in the striatum. (Drawn by Kyle Christensen.)

Dopamine cell bodies are primarily located in two midbrain nuclei: the ubstantia nigra (pars compacta) and ventral tegmental area. Projections from these nuclei have three primary target regions, and are named accordingly. The nigrostriatal tract carries nigral dopaminergic projections to subcortical motor control areas of the striatum (caudate and putamen in humans). The nigrostriatal projections come primarily from the substantia nigra but also, to a lesser extent, from the ventral tegmental area. Mesolimbic dopamine projections from this area target a number of limbic regions, such as the nucleus accumbens and temporal lobes. The mesocortical dopamine pathway projects primarily from the ventral tegmental area to the prefrontal cortex. A fourth dopamine tract is found entirely within the hypothalamus. In addition to different target regions, these separate projection systems function independently to some degree and are regulated by different mechanisms.

Dopamine exerts its effects through at least five receptor types, D_1 through D_5, identified on the basis of their deoxyribonucleic acid (DNA) sequence. Most pharmacological functions of dopamine receptors characterized so far are attributed to D_1 and D_2 receptors. Much less is known about the actions of D_3, D_4, and D_5 receptors. The D_1 family includes D_1 and D_5, while the D_2 family includes D_2, D_3, and D_4 receptors. Genes for the D_2 family have a number of introns, leading to alternative splicing and several isoforms. For example, the D_2 receptor has two common splice variants, a long and short form, usually both expressed in the same cell. The D_4 receptor has numerous polymorphisms, including longer and shorter forms, although these do not arise through alternative splicing. Different isoforms of the D_2 family may have different affinities for second messenger systems, presumably leading to variations in biological effects. Introns or alternative splicing variants for the D_1 family of receptors have not yet been identified.

D_1 and D_2 receptors are found predominantly on the primary efferent neurons of the striatum, and limbic system (e.g., the nucleus accumbens), prefrontal cortex, and other cortical regions. D_2 receptors are also located on the presynaptic dopamine terminals in target regions and dopamine cell bodies in the midbrain. These autoreceptors regulate dopamine synthesis, neuronal firing, and release. The latter two autoreceptors are not on mesocortical nerve terminals in the prefrontal cortex. D_3 receptors are expressed predominantly in subcortical limbic regions, such as the islands of Calleja and nucleus accumbens in the rodent, but are also seen in the hippocampus. D_4 receptors are thought to be presynaptic regulators of glutamate release on projections from cortical areas to the striatum and some limbic regions. D_5 receptors are found in limited distribution in the thalamus, hippocampus, and hypothalamus.

The role of the dopamine system in the overall economy of the brain is not well understood. The relation between dopamine cell loss and Parkinson's disease established its role in regulating motor activity. The link between dopamine and drugs of abuse suggest a critical role in motivation and reward. Increasingly sophisticated electrophysiological studies have shown that activation of subcortical dopamine pathways alert the organism to changes associated with the prediction of future salient and rewarding events. This function is essential for predicting future events, which allows an organism the ability to plan and control interactions with the environment. Furthermore, prefrontal cortical dopamine is critically involved with working memory, a key component for higher-level information processing tasks. Thus, dopamine is involved in motor behavior, motivation, reward, and a variety of higher cognitive tasks, all of which have been implicated in schizophrenia. Clearly, the dopamine system has a complex molecular, cellular, and physiological neurobiology, and this underlies an equally complex functional role in normal brain and behavioral function.

Dopamine and Schizophrenia Evidence for the dopamine hypothesis of schizophrenia comes from a variety of sources. One approach has been to examine the effects of different medications on schizophrenic symptoms. Drugs that block D_2 receptors reduce psychotic symptoms; dopamine agonists worsen symptoms. These observations form the cornerstone of the dopamine hypothesis. A second approach has been to look at various indices of dopaminergic neurotransmission in patients with schizophrenia. Such indices include measures of presynaptic activity, such as the major dopamine metabolites, dihydroxyphenylacetic acid and homovanillic acid, as well as postsynaptic markers, primarily dopamine receptors. Metabolite studies have examined homovanillic acid in urine, plasma, CSF, and autopsied brain. Receptor studies have been performed on postmortem brain tissue and in living patients using PET and SPECT. More recent methods have been used to assess in vivo presynaptic dopamine levels and dopamine release using both PET and SPECT. Dopamine neurotransmission could be altered by changes in any one of a number of neuronal functions, including synthesis, degradation, release, uptake, receptor binding, or effects on second and third messenger systems. Although several decades of research have not provided definitive affirmation of the dopamine hypothesis, increasingly sophisticated methods to assess in vivo dopamine

activity are beginning to yield important clues.

The notion that dopamine neurotransmission is increased in schizophrenia derives its most compelling support from clinical observations on the effects of drugs that impact psychotic symptoms. The introduction of antipsychotic medications in 1954 was a dramatic breakthrough in psychiatry and initiated an intense search for their mechanism of action. In 1963 antipsychotic medications were found to increase the concentrations of dopamine metabolites. It was suggested that increased metabolite concentrations were a compensatory response to the blockade of dopamine receptors by antipsychotic agents and a subsequent reduction in dopamine neurotransmission. The idea that these drugs reduced dopamine neurotransmission was further supported by the observation that they also induced parkinsonian adverse effects, symptoms that had recently been linked to the loss of midbrain dopamine neurons. In 1977, following pharmacological characterization of the D_2 receptor, a striking correlation was reported between the relative clinical potencies of all clinically available antipsychotic medications and their ability to block D_2 receptors. This landmark finding convincingly demonstrated that antipsychotic effects were mediated by D_2 receptor blockade.

While the correlation between clinical potency and D_2 blockade for antipsychotic medications was compelling, several problems emerged. D_2 blockade occurs within hours of administration, but the antipsychotic effects can take days or weeks to develop; this suggests that a secondary process is required. Studies of the chronic effects of neuroleptics then led to the observation that, after several weeks, dopamine neurons themselves stopped firing. After short-term administration of antipsychotic medications there is an initial increase in dopamine neuronal firing as neurons attempt to overcome D_2 blockade; eventually this overexcitation leads to the phenomenon of depolarization block, where depolarized neurons simply stop firing. Reduced neuronal firing was thought to markedly reduce dopamine release, leading to reduced dopamine neurotransmission. For some time, the depolarization block theory was crucial in supporting the view that antipsychotic drugs exert their therapeutic effects by reducing dopamine neurotransmission. Subsequently, a number of studies have not found reduced dopamine release after long-term treatment with antipsychotic medication. While methodological issues are still debated, this suggests that some process other than a simple reduction in dopamine release may underlie the therapeutic effects of these medications.

Other observations have been difficult to reconcile with the dopamine hypothesis. For example, many symptoms such as cognitive deficits, anhedonia, and alogia typically fail to respond to treatment with antipsychotic medications, suggesting that other processes are involved. A second problem relates to the unique clinical effects of clozapine (Clozaril). Clozapine has been shown to benefit patients who do not respond to dopamine receptor antagonists. The dopamine hypothesis, on the other hand, implies that D_2 blockers should be equally efficacious. The unique clinical effects of clozapine suggest that it may have a different mechanism of action. Clozapine's effects have been attributed to several properties, such as its antagonism of serotonin receptors or its combination of D_1, D_2, and D_4 blockade. Drugs developed to mimic different aspects of clozapine's receptor-binding profile, such as risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine (Seroquel), share some of clozapine's “atypical” characteristics.

A second line of evidence supporting the dopamine hypothesis comes from observing the effects of dopamine agonists. Chronic amphetamine abuse, for example, increases dopamine release and can lead to a psychosis similar to paranoid schizophrenia. Amphetamine-induced psychotic disorder, however, lacks other features associated with schizophrenia, such as negative symptoms and cognitive impairment. Furthermore, psychotic symptoms only develop after prolonged use (and typically at high doses), whereas dopamine neurotransmission is increased shortly after a single dose of amphetamine. This suggests that repeated increases in dopamine release produce secondary changes that are more directly responsible for the psychosis.

METABOLITE STUDIES The search for more direct evidence of altered dopamine release in schizophrenia led to investigations of dopamine and its metabolites in urine, plasma, CSF, and postmortem brain tissue. Consistent with the basic dopamine hypothesis, several studies of plasma homovanillic acid have found increases in unmedicated schizophrenia patients compared with controls. These studies sometimes report correlations between concentrations of homovanillic acid and severity of psychosis. Furthermore, antipsychotic medications appear to reduce plasma homovanillic acid over time, correlating with patients' improvement. Methodological problems, however, cloud the interpretation of studies using plasma homovanillic acid. It is unclear whether plasma homovanillic acid correlates with its concentrations in limbic brain regions, areas most likely to underlie the production of psychotic symptoms.

Investigators have also looked at dopamine metabolite levels in CSF. While most studies have failed to find significant changes, several have reported a correlation between concentration of homovanillic acid and severity of psychotic symptoms. Studies of medication-free patients have tended to show a reduction in dopamine metabolites. Negative correlations have been found between concentrations of homovanillic acid in CSF and ventricular enlargement and severity of negative symptoms (e.g., anhedonia and flat affect). Prefrontal cognitive deficits have also been associated with reduced CSF homovanillic acid, perhaps consistent with a model of subcortical dopaminergic overactivity and prefrontal cortical hypoactivity; methodological issues make the interpretation of CSF studies problematic. First, dopamine and metabolite concentrations in the CSF are affected by a number of variables that are not commonly controlled. These include diet, time of day, height, and motor activity. Second, increased ventricular volume itself could affect the concentration of homovanillic acid. Third, CSF monoamine concentrations appear to have little relation to either regional brain levels of dopamine or, more importantly, to more direct measures of dopamine neurotransmission. Certainly if dopamine transmission in the prefrontal cortex is reduced and subcortical transmission is increased, it is difficult to predict what would happen to CSF concentration. Nevertheless, CSF data is often interpreted as supporting the notion that too much dopamine is related to positive symptoms whereas too little underlies negative symptoms.

More direct assessments of dopamine neurotransmission have come from postmortem studies of dopamine metabolites. Increased dopamine or homovanillic acid or both have been reported in a number of brain regions, although reports are often inconsistent. For example, one study found increased dopamine in the left amygdala, a second reported increases in the nucleus accumbens, and a third found increases in the caudate but not the accumbens. Increased homovanillic acid has been found in the cortex, accumbens, and caudate. The latter finding has been attributed to the effects of previous treatment with antipsychotic medications. At this point no clear consensus can be derived from studies of dopamine metabolites.

DOPAMINE RECEPTOR STUDIES A number of studies using postmortem brain tissue have shown increased numbers of D_2 dopamine binding sites in the brains of schizophrenia patients. A major confounding issue is whether this increase is a primary alteration in schizophrenia or secondary to long-term treatment with antipsychotic agents, known to cause rapid D_2 upregulation in animals. Studies in nonmedicated and medication-naive patients are conflicting. A number of studies of patients off medication for at least 1 month have found increased D_2 receptors, although several have not. It has been suggested that treatment with antipsychotic medications cannot account for the marked increase and bimodal distribution of D_2 receptors seen in patients who had been treated. Imbalances between D_1 and D_2 receptors have also been reported. Recent studies of D_3 receptors have suggested that D_3 mRNA may be processed abnormally in cortical neurons of patients with schizophrenia, resulting in reductions in the normal D_3 mRNA transcript. On the other hand, a postmortem study of striatal D_3 receptor binding found a significant increase in patients who were medication free for 1 month. D_4 receptors have been harder to assay because of the lack of specific ligands. While two reports using an indirect method have found evidence of increased D_4 receptor density, assays of mRNA for D_4 using highly specific antisense probes have not found increased levels.

Neuroimaging techniques have been used to measure indices of dopamine neurotransmission in living human patients. Striatal D_2 receptors have been assayed in medication-free patients by several groups using PET; the results, however, have been conflicting. One study found increased receptor numbers while two others did not. These studies used different PET ligands to measure D_2 receptor density, perhaps accounting for the conflicting results. One of the PET ligands binds only to D_2 and D_3 receptors; the second also binds to D_4 receptors. The discrepant PET findings have been attributed to an increase in D_4 receptors, consistent with postmortem studies. More recently, in vivo neuroimaging methods have been refined to assay presynaptic indices of dopamine storage and release. In this paradigm, radioactive D_2 ligand binding is examined at baseline and following a pharmacological challenge with amphetamine. The dramatic increase in dopamine release caused by amphetamine displaces the postsynaptic binding of the D_2 ligand. The washoff of the D_2 ligand can thus be used as an index of dopamine release. Unmedicated patients with schizophrenia show reduced ligand binding after amphetamine, but not at baseline. This suggests that patients with schizophrenia have increased synaptic dopamine following amphetamine. One explanation for this is that presynaptic stores may be increased; another possibility is that synaptic reuptake is reduced. Although methodological issues continue to be refined, this promising lead implies that subtle aspects of dopamine neurotransmission may be altered.

ANIMAL MODELS Animal studies have been invaluable in efforts to understand normal and abnormal function of the dopamine system. Of particular relevance for schizophrenia research are studies that attempt to model dysfunctional dopamine systems in a way that may shed light on the neurobiology of psychosis.

Initial attempts to develop relevant animal models began with repeated, high doses of stimulants (such as amphetamine), based on the association between stimulant abuse and psychosis in humans. Repeated stimulant treatment was also thought to model repeated stress, an apparent trigger of psychotic relapse. Remarkably, stimulants increase the sensitivity of the mesolimbic dopamine system to stress, a process referred to as *sensitization*. Furthermore, in some paradigms stimulants can reduce presynaptic indices of dopamine activity, which has led to speculations that repeated increases in dopaminergic transmission (e.g., from stress) could lead to sensitization in limbic regions and long-term dopamine depletion in prefrontal regions. Thus, long-term administration of stimulant may provide a model to explore the

interactions between known triggers of psychosis and dysfunctional dopamine systems.

Another promising line of animal research suggests that alterations in dopamine neurotransmission in one region may be secondary to primary deficits in another. For example, depletion of dopamine from prefrontal regions can increase dopamine metabolism in the striatum of rats. This suggests that a primary reduction of prefrontal dopamine in humans could theoretically lead to secondary alterations in subcortical dopamine. Reduced prefrontal dopamine could certainly explain the hypofrontality and negative symptoms that characterize schizophrenia. Although no direct evidence has shown that there are dopamine abnormalities in these regions, the indirect evidence reviewed above is suggestive. In a related line of research, structural damage to cortical and limbic regions has been shown to change subcortical dopamine neurotransmission. For example, within the limbic system lesions of the hippocampus or amygdala alter dopamine neurotransmission in the nucleus accumbens and prefrontal cortex. Such observations have been critical in attempts to relate structural and functional changes in frontal, temporal, and hippocampal regions with abnormalities in the dopamine system. They suggest that information-processing deficits in frontal and limbic regions have marked effects on subcortical processes, including dopamine neurotransmission.

The dopamine hypothesis continues to exert a profound effect on research in schizophrenia. The discovery of new subtypes of dopamine receptors along with new neuroimaging approaches offer improved methods to study the function and pathophysiology of this system in humans. Particularly important for schizophrenia research is the finding that dopamine subsystems are interconnected and that damage to different brain regions previously implicated in schizophrenia can have marked effects of dopamine neurotransmission. At present, a variety of indirect data suggest that prefrontal dopamine neurotransmission may be reduced whereas subcortical dopamine is dysregulated in schizophrenia. Whether these changes are real and whether they are secondary to cortical or limbic dysfunction remains to be seen.

Glutamate Interest in glutamate's role in the pathophysiology of schizophrenia has developed relatively recently. This interest was spurred primarily by two observations. First, acute ingestion of phencyclidine (PCP), a glutamate antagonist, produces a syndrome similar to schizophrenia. Second, glutamate is an essential neurotransmitter in those neural networks that may be involved in schizophrenia. Subsequently, a variety of postmortem and clinical data have been garnered in support of a glutamatergic abnormality.

Neurobiology of Glutamate Glutamate is one of the most prevalent neurotransmitters in the brain. Virtually all neurons in the brain are affected when glutamate is applied. A nonessential amino acid that does not cross the blood-brain barrier, it can be synthesized in the brain from glutamine. The dominant mode of inactivation of synaptic glutamate is via reuptake by specific, high-affinity uptake sites.

The four classes of glutamate receptors have been identified and named after their affinity for specific ligands: *N*-methyl-D-aspartate (NMDA), a-amino-3-hydroxy-5-methylisoxazole-4-proprionic acid (AMPA), kainic acid (KA), and L-aminophosphono-butyric acid (AP4). The first three are ionotropic receptors; their effects are mediated by changes in ionic conductance through neuronal membranes, including sodium, potassium, and calcium. The ionotopic receptors have been implicated in neurotoxicity following ischemia, mediated in part by increased intracellular calcium influx and apoptosis. The NMDA receptor is functionally different from the others and has been implicated in long-term potentiation (a process related to memory) in the hippocampus. Paradoxically, NMDA blockade can also result in neurotoxicity, apparently modulated by interneurons and activation of non-NMDA glutamate receptors. The last type of glutamate receptor, labeled by AP4, is a metabotropic receptor, a member of the family of G-protein–linked receptors. The metabotropic receptors modulate activation of second messengers, such as phospoinositide and cyclic adenosine monophosphate (cAMP), which can produce long-term, modulatory effects. Major advances in understanding the molecular biology of these receptors is increasing the understanding of their function.

The NMDA receptor is a complex protein that has particular relevance for schizophrenia research. Blockade of the NMDA receptor by phencyclidine (PCP), a noncompetitive antagonist, produces symptoms similar to those seen in schizophrenia. NMDA receptor activation is excitatory, reducing postsynaptic membrane potential. PCP binds to a site within the open NMDA ion channel, thus blocking ionic flux. The mechanism by which NMDA antagonism produces psychotic symptoms is unclear; one theory is that NMDA antagonists exert their psychotomimetic effects via NMDA receptors' role in regulating striatal and limbic dopamine neurotransmission. Of note, NMDA receptor density is highest in the hippocampus and prefrontal cortex, two areas already implicated in the pathophysiology of schizophrenia. Altered neurotransmission in these regions could also play a role in PCP's effects.

The NMDA receptor has a number of modulatory sites that regulate ionic conductance. Endogenous modulators include glycine, zinc, magnesium, and the polyamine spermidine. The glycine modulatory site has become a target for drug development. Increasing NMDA neurotransmission by increased glycine binding has been hypothesized to reduce symptoms of schizophrenia. Several studies have attempted to do so using glycine agonists such as milacemide or cylcoserine, and the results have been mixed. Another potential pharmacological target is the high-affinity glycine uptake pump. Antagonists of this site should increase synaptic glycine concentrations, enhancing NMDA neurotransmission. Similarly, antagonists of the glutamate reuptake pump could boost NMDA receptor activation. It is unclear whether ongoing efforts to develop antagonists at these sites will lead to therapeutic agents for patients with schizophrenia. A major difficulty with increasing NMDA neurotransmission is its narrow range of physiological responsivity. If NMDA stimulation is too high, seizures or neurotoxicity can result.

Glutamate is relevant to the neurochemistry of schizophrenia because of its role in key neural networks. Projections to and from cortical and hippocampal pyramidal neurons use glutamate as a primary neurotransmitter. These include projections to subcortical structures such as the striatum, nucleus accumbens, and ventral tegmental area; output from these areas is strongly modulated by glutamate. Thalamic projections to the cortex also employ glutamate as the major neurotransmitter. Glutamate neurotransmission is important not only for rapid synaptic transmission between these regions, but also for experience-dependent cortical plasticity and memory. This is particularly true for the voltage-sensitive NMDA receptor, a likely candidate for modulating memory traces at Hebbian synapses. Glutamate's essential role in key neural networks, memory and cortical plasticity, thus make it a likely candidate for involvement in altered information processing in schizophrenia.

Glutamate in Schizophrenia Acute intoxication with the NMDA antagonist, PCP, produces hallucinations, thought disorder, negative symptoms, and cognitive deficits. In comparison, dopamine agonists, such as amphetamine, primarily induce paranoid delusions, and only after long-term use. The differences in these drug-induced psychoses suggest that glutamatergic neurotransmission could be more proximal to the pathological processes mediating psychosis. The search for more direct evidence has focused on CSF and postmortem studies of brain tissue.

Studies of glutamate levels in CSF and brain have been mixed. An initial, pioneering study of CSF found low levels of glutamate in patients compared with controls. Possible methodological problems make this data difficult to interpret, however, and three subsequent studies have been unable to replicate the finding. Two studies have looked at glutamate levels in postmortem brain tissue. One found no differences whereas the other found specific reductions in the hippocampus and prefrontal cortex in patients with schizophrenia. The latter study also looked at a neuropeptide co-localized with glutamate *N*-acetylaspartylglutamate [NAAG]. The NAAG pathway has recently been identified as an important comodulator of glutamate neurotransmission. The reported changes in NAAG and its metabolism in brains of patients with schizophrenia open up a provocative new area to explore possible alterations in glutamate neurotransmission.

Postmortem receptor studies have been more promising. In general, these studies have tended to find increased receptor binding in prefrontal regions and reductions in temporal areas. Two reports have found increased kainate binding in the medial frontal cortex; a third found increases in orbitofrontal NMDA receptors. An increase of prefrontal cortical glutamate uptake sites has also been described. A recent molecular study using in situ hybridization and probes for all five NMDA receptor subunits, while not finding an overall increase in receptor mRNA did find a 53 percent increase in the expression of a subunit (NR2D), suggesting a change in the functional properties of prefrontal NMDA receptors.

In the temporal lobe, several abnormalities of the glutamate system have been published. Autoradiographic studies have reported that KA receptor binding is reduced, particularly in the hippocampus. Consistent with this finding, reduced expression of mRNA for receptor subunits has been found in temporal lobe areas. Reduced density of temporal lobe AMPA receptors has been seen, but less consistently. In a recent extension of this work, mRNA transcripts for *Glu* R1 and *Glu* R2 were assayed; these transcripts code for AMPA receptor subunits. Consistent with receptor studies, reductions were seen in the hippocampus and other temporal lobe areas. Finally, the glutamate reuptake site has been assayed in the temporal lobe as an index of presynaptic glutamate terminal number. Reduced levels of mRNA for the reuptake site suggest a possible reduction in terminal number and thus in axonal projections. Regarding other brain regions, some receptor studies have performed on material from the basal ganglia. Increases in AMPA receptors and reduced NMDA receptors have been reported; some, studies, but not all, have found reduced glutamate uptake sites.

Taken together, the postmortem literature is notable for a myriad of findings implicating alterations in glutamatergic neurotransmission. However, given the typical small number of brains studied and large number of variables, replication of specific findings is critical. Some have theorized that there is a loss of glutamatergic neurons in temporal areas, consistent with structural neuroimaging findings of reduced volume. In this schema, increased glutamate receptors in the cortex and putamen are hypothesized to be secondary to reduced glutamatergic inputs or neurotransmission. The increased focus on glutamate in postmortem studies will bring

increasingly sophisticated assessment of this neural system.

Serotonin The idea that serotonin may play a role in schizophrenia was first postulated when the hallucinogen lysergic acid diethylamide (LSD) was found to block serotonin receptors. Since then, basic studies have begun to unravel the surprising complexity of this system and have provided new targets for investigation. Studies of schizophrenia have looked at a variety of parameters, including plasma serotonin levels, brain receptor levels, and clinical response to serotonergic drugs. Two findings are particularly promising: first, data from postmortem studies have found changes in frontal cortical receptor number; second, new “atypical” antipsychotic medications that are both serotonergic and dopaminergic antagonists appear to have clinical advantages over pure D_2 antagonists. These developments have increased the focus on serotonin in schizophrenia.

Basic Neurobiology Serotonin (5-hydroxytryptamine) is synthesized from tryptophan and is broken down into 5-hydroxyindolic acetic acid (5-HIAA) by monoamine oxidase (MAO). Tryptophan is an essential amino acid; dietary intake of tryptophan can affect CNS synthesis of serotonin. Serotonin synthesis is also modulated by autoreceptors on nerve terminals. Synaptic serotonin is inactivated primarily by reuptake pumps on presynaptic neurons and glia; following uptake, serotonin is repackaged into vesicles or broken down to 5-HIAA. Both serotonin itself and its uptake pumps are found in blood platelets, where they play a role in clotting. In the CNS, serotonin neuronal cell bodies are located in the brainstem in nine separate nuclei. Axons from these cells project through the median forebrain bundle to virtually all regions of the CNS, including the cortex, limbic regions, and the striatum.

The effects of serotonin are mediated by an ever-increasing number of receptor subtypes. Currently, seven classes of serotonin receptors have been characterized: serotonin (5-hydroxytryptamine [5-HT])-type 1 (5-HT_1) through 5-HT_7. Ten subtypes have been described in the 5-HT_1 family (5-HT_{1a} through 5-HT_{1e}), three in the 5-HT_2 family (5-HT_{2a} through 5-HT_{2c}) and one for 5-HT_3. Most relevant for schizophrenia are the 5-HT_2 and 5-HT_3 subtypes. $5HT_2$ receptors are found in the prefrontal cortex, striatum, and nucleus accumbens; 5-HT_3 receptors are found in cortical, limbic, and subcortical areas, such as the amygdala and hippocampus.

The serotonin system subserves a bewildering array of physiological and behavioral functions. For example, somatodendritic 5-HT_2 receptors regulate dopaminergic neuronal firing. Striatal nerve terminal serotonin receptors inhibit dopamine release. Behaviorally, serotonin has effects on cardiovascular, respiratory and motor activity, emesis, sexual behavior, aggression, anxiety, mood, and pain. Frontal serotonin, in concert with dopamine, may play an important role in the modulation of attention and arousal. Recently, basic research using aplysia has shown that serotonin plays a critical role in synaptic mechanisms associated with learning and memory; it may also have important neurotrophic effects during development and in the adult organism.

Serotonin in Schizophrenia The earliest studies to examine serotonin in schizophrenia looked at peripheral measures, such as serotonin concentrations in plasma and uptake in platelets. These studies found increased concentrations in plasma and, less consistently, reduced uptake in platelets. Studies of CSF metabolites have been mixed and suffer from the same methodological confounds described for dopamine. More direct measures of CNS neurotransmission include postmortem assays of serotonin activity, including concentrations in brain tissue; receptor-binding density; reuptake site binding; and levels of mRNA for receptor subtypes, reuptake sites, and synthetic enzymes for serotonin itself.

Although there have been multiple reports of abnormal serotonin levels in a variety of neural structures, only two findings have been replicated: increased levels in the putamen and increased levels in the globus pallidus. One difficulty with this approach is that measurement of neurotransmitters and their metabolites is notoriously unreliable because of their instability in postmortem tissue. In comparison, receptors, reuptake sites, and the mRNA for receptors, reuptake sites, and synthetic enzymes are more stable. Of studies looking at these parameters, the 5-HT_2 subclass has received the most attention. Following an initial report of a reduction in prefrontal cortex in the density of this receptor, two other research groups replicated this finding although a third did not. Whereas this abnormality may be intrinsic to schizophrenia, it is also possible that reduced 5-HT_2 receptor density is a consequence of therapy with antipsychotic drugs. The density of reuptake sites for serotonin also appear to be reduced in schizophrenia, particularly in frontal and anterior cingulate cortices.

Studies looking at the mechanism of action of atypical antipsychotic drugs, such as clozapine, have fueled much of the recent interest in serotonin's role in schizophrenia. Clozapine has a variety of therapeutic properties different from the dopamine receptor antagonists. These could be due to clozapine's ability to block 5-HT_{2a}, 5-HT_{2c}, 5-HT_3, 5-HT_6, or 5-HT_7 receptors, or to increased serotonin release in the prefrontal cortex. When one compares serotonin-dopamine antagonists, which share some of clozapine's properties, such as the reduced liability to produce parkinsonian symptoms, two impressive similarities are their 5-HT_2-binding affinity and the ratio of 5-HT_2 to D_2 binding. This suggests that serotonergic antagonist properties may account for the improved adverse effects profile and perhaps also the enhanced therapeutic efficacy often attributed to the serotonin dopamine antagonists.

In summary, both postmortem studies and drug trials using 5-HT_2–D_2 antagonists suggest that serotonin may play an important role in schizophrenia. Data implicating frontal and anterior cingulate cortices are particularly striking. It is unclear, however, whether alterations in serotonin neurotransmission are primary or secondary and how they may relate to the other neurobiological processes described. Some preliminary investigations suggest that maternal exposure to toxins can produce long-term changes in serotonin neurotransmission. This raises the possibility that neurodevelopmental insults could alter serotonin neurotransmission in adults. Researchers interested in the mechanisms of amphetamine-induced behavioral sensitization have also begun to suspect that serotonin may play a significant role. If sensitization were to be involved in schizophrenia, as has been suggested, serotonin could be a factor. Although research in serotonin has typically taken a backseat to research on dopamine, its relevance for schizophrenia continues to increase as more is revealed about its many neurobiological properties.

Other Neurotransmitters A wide variety of additional neurochemical systems have been studied in schizophrenia, several of which are noteworthy because of potentially interesting findings or because of how extensively they have been studied. These include GABA, norepinephrine, neurotensin, and cholecystokinin. As with other neurotransmitters, studies of these systems have typically looked at transmitter and metabolite levels in brain, CSF, or plasma, as well as receptor protein and mRNA expression in specific brain regions.

GABA Particularly intriguing is research into the role of GABA, which is the major inhibitory neurotransmitter in the brain. Virtually all neurons are inhibited by GABA, and up to 40 percent of neurons use GABA as their major neurotransmitter. Many GABA neurons are local inhibitory interneurons, but GABA neurons in some regions (such as the striatum) are also primary efferent neurons. GABA is synthesized from glutamate via the enzyme glutamic acid decarboxylase (GAD). GABA acts at two receptor subtypes, $GABA_A$ and $GABA_B$, the former being the more important in the CNS. A variety of drugs act at GABA receptors, including alcohol, benzodiazepines, and barbiturates. Findings implicating GABA in schizophrenia include reduced number of GABAergic cortical interneurons, increased $GABA_A$ receptor density in the prefrontal cortex, and reduced GABA uptake sites in the hippocampus. All three findings are consistent with reduced GABA cell number or GABA neurotransmission. Studies of mRNA have found reduction in prefrontal GAD mRNA but not in prefrontal $GABA_A$ receptor mRNA. The former is consistent with reduced GABA neuronal activity; the latter is not. This preliminary effort suggests that GABA cell number or activity is reduced in schizophrenia. As with other postmortem findings, however, further replication is necessary before they can be accepted with confidence.

Norepinephrine Norepinephrine, another monoamine neurotransmitter, has been intensively studied in schizophrenia, although interest has waned recently. Similar to dopamine and serotonin, norepinphrine neurons are located in the brainstem in a group of nuclei (including the locus ceruleus) that project to a variety of cortical and subcortical regions. Norepinephrine acts at two receptor families, adrenergic and b-adrenergic receptors; at least seven a and three b subtypes have been cloned. Both receptor families exert their effects via changes in G-protein–mediated second messenger systems, including cAMP and phosphoinositol. Two neuropeptide transmitters, galanin and neuropeptide Y, are colocalized in noradrenergic neurons. Norepinephrine and its co-transmitters are involved in a number of physiological and behavioral processes including the sleep-wake cycle, arousal, stress, and memory. Both basic and clinical studies support a role for this system in psychiatric disorders such as anorexia nervosa, bulimia nervosa, anxiety disorders, post-traumatic stress disorder, depressive disorders, substance dependence, and substance withdrawal. Many of the behavioral states mediated by the noradrenergic system are markedly altered in schizophrenia, suggesting a role here as well. However, more direct evidence is lacking and any changes in noradrenergic function in schizophrenia may be secondary to the agitation that frequently accompanies psychosis.

Initial studies of norepinephrine examined concentrations in plasma, CSF, and brain tissue. Both plasma and CSF concentrations of norepinephrine and its metabolite appear to be increased in patients with schizophrenia, although this has not been a consistent finding. Concentrations are reduced with treatment with antipsychotic agents and are correlated with clinical improvement. Recently, increased plasma concentrations have been associated with deficit symptoms whereas reduced plasma levels have been associated with depressive symptoms. These two findings seem contradictory and cast doubt on the usefulness of this approach. Furthermore, conclusions from such studies are limited by the same methodological pitfalls described above for other neurotransmitters, including the confounds of treatment with antipsychotic agents and the meaning of peripheral measures. Studies of brain norephinephrine and its receptors have been mixed, with some finding elevations and others finding no changes. The clinical effects of adrenergic agents have generally not been impressive. At least one report found that the presynaptic a_2-adrenergic receptor agonist clonidine (Catapres) reduces psychotic symptoms, presumably by reducing norepinephrine release. On the other hand, several other

studies did not find this effect, and at least one group has reported therapeutic effects for an a $_2$-adrenergic receptor antagonist, idazoxane. Finally, a number of genetic association studies have looked at the incidence of polymorphisms for genes related to norepinephrine neurotransmission, including dopamine beta hydroxylase and the norepinephrine transporter. Although relatively common polymorphisms have been reported for both, no association with schizophrenia has been reported.

Neuropeptides Two other interesting candidate molecules that have been studied in schizophrenia are the neuropeptides cholecystokinin and neurotensin. Both are found in a number of brain regions implicated in schizophrenia, such as the substantia nigra, nucleus accumbens, hippocampus, and various cortical regions. Both are colocalized with dopamine, GABA, glutamate, and other neurotransmitters. Several studies have reported changes in the levels of the peptides themselves, mRNA, or receptors. For example, the following findings have had some degree of replication: reduced temporal lobe cholecystokinin peptide concentrations, reduced cholecystokinin receptor density in both temporal and frontal regions, and reduced cholecystokinin mRNA in the temporal lobe. In general, further replication is required. Drug trials with the cholecystokinin agonist ceruletide have been mixed. Several open trials were promising, but double-blind trials were not. Unfortunately, it is not certain that ceruletide crosses the blood-brain barrier.

Neurotensin's appeal is due in part to its endogenous antipsychotic-like properties. Not only is it colocalized in dopaminergic neurons, but infusions of neurotensin into the nucleus accumbens block the excitatory effects of stimulants and reduce behavioral activation. Neurotensin levels in the nucleus accumbens are markedly increased by treatment with antipsychotic medication. CSF studies have shown reduced neurotensin concentrations and correlations between reduced concentrations and increased psychopathology in drug-free patients with schizophrenia. However, postmortem studies have not shown differences between patients and controls in concentrations of the peptide itself. Such studies are confounded by the pronounced effects of antipsychotic drugs on central nervous system (CNS) neurotensin. One recent report found a 40 percent reduction of neurotensin receptors in the entorhinal cortex in patients with schizophrenia. Further replication and exclusion of effects of treatment with antipsychotic agents will clarify the significance of this finding.

NEURAL CIRCUITS

The variety of structural, functional, and neurochemical abnormalities described implicate disordered information processing in several interconnected neural pathways in patients with schizophrenia. A description of the anatomical components of these pathways and their possible function will provide a basis for integrating the many abnormalities noted in schizophrenia (Figs. 12.4-5).

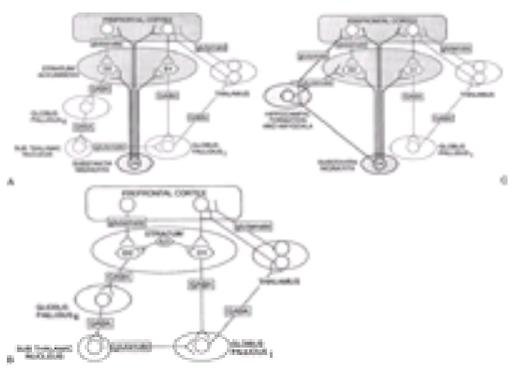

FIGURE 12.4-5 A, Neural networks implicated in the neurobiology of schizophrenia. Cortico-striatal-thalamic pathway. Prefrontal glutamatergic projections synapse on GABAergic striatal neurons that express either D_1 or D_2 receptors. The independent D_1 and D_2 pathways are referred to as the direct and indirect pathways respectively. They have separate efferent pathways projecting to either the globus pallidus, pars externa = E, or pars interna = I. Both pathways ultimately project back to the anterior thalamus. **B**, Ascending dopamine projection pathways modulate circuits in A. Dopamine neurons (DA) from the substantia nigra tend to project mainly to the striatum, while the adjacent ventral tegmental area DA neurons projects primarily to the prefrontal cortex, ventral striatum, and limbic regions. **C**, Limbic projections to circuits from **A** and **B**. The hippocampal formation and amygdala project to the prefrontal cortex and ventral striatum. They receive glutamatergic cortical input and dopamine projections from the VTA. (Drawn by Kyle Christensen.)

As cortical abnormalities have played a dominant role in theories of schizophrenia, understanding the functional connectivity of these areas is important. One of the most intensively studied pathways is the cortico-striato-thalamic loop. The prefrontal cortex, the most highly and recently evolved part of the primate brain, sends a massive glutamatergic projection to subcortical regions, most notably the striatum (putamen and caudate in humans). The striatum in turn sends GABAergic projections through a number of downstream basal ganglia nuclei that ultimately feed into the anterior thalamus. Completing the loop, the anterior thalamus sends a massive glutamatergic projection back to the prefrontal cortex. Several salient features are noteworthy. First, this loop appears to consist of at least five separate but parallel channels processing different types of information (such as cognitive, emotional, and motoric information). Second, output from the striatum is split into two opposing, counterbalancing pathways. The so-called direct and indirect loops are modulated by D_1 and D_2 receptors, respectively. Their coordinated output modulates information returned to the cortex via the anterior thalamus. Third, within the striatum itself, the ventral portion (commonly referred to as the *nucleus accumbens*) receives predominantly limbic inputs, while dorsal regions receive inputs more relevant for motor function. This functional segregation is maintained in downstream projection regions.

A second important system that modulates activity of the cortico-striato-thalamic pathway is the dopamine system. Dopamine neurons in the substantia nigra and ventral tegmental area project to the striatum, nucleus accumbens, and prefrontal cortex. Dopamine modulates cortical output to the striatum via input to glutamatergic pyramidal neurons. In the striatum, dopamine axons synapse on the primary output neurons, the medium-sized, spiny, GABAergic neurons. Coordinated cortical and subcortical dopamine neurotransmission may be important for normal information processing through this loop. Furthermore, dysfunction in one area may produce changes in another. For example, lesions of the prefrontal cortex can induce alterations in subcortical dopamine neurotransmission.

A third neural system interacting with the first two is the limbic system. This complex system involves hippocampus, amygdala, thalamus, hypothalamus, and cingulate gyrus, among others. This immense circuit, subserving functions related to memory and emotional experience, among many others, has direct projections to both prefrontal cortex and ventral striatum. The prefrontal cortex has reciprocal projections back to the mesial temporal lobe and hippocampus. The hippocampus, amygdala, and cingulate have important projections to the ventral (or limbic) aspect of the striatum. This area, in turn, projects to the thalamus via the ventral aspect of the globus pallidus, the pars interna. In this way, three major brain regions—the cortex, limbic system, and basal ganglia—communicate and interact. Information-processing abnormalities in one area, such as in the hippocampus, would have significant downstream effects on other regions, such as prefrontal cortex and striatum.

Structural and functional measures have implicated some abnormality in all three components of these interacting systems. It is uncertain which are primary and which are secondary. It seems very possible that different types of lesions could alter the function of individual components, which could then produce secondary downstream changes in connected circuits.

NEUROBIOLOGICAL MODELS

The essential neurobiological features of schizophrenia may place some constraints on plausible pathophysiological processes. First, there is a major genetic contribution. Many genes are likely to be involved and these may function in part by increasing vulnerability to the deleterious effects of environmental factors. Several environmental factors have been hypothesized to increase the risk of schizophrenia, perhaps by producing subtle brain damage. Structural abnormalities have played an important role in placing theoretical constraints on mechanisms. Since they are present from early in the illness and do not appear to progress, they may predate the onset of illness. Neuropathological data and studies of obstetric and perinatal complications support the idea that an early lesion may account for structural changes. The apparent lack of gliosis in postmortem studies is particularly critical and implicates in utero factors. Structural and functional neuroimaging, as well as neuropsychological data and animal studies present converging evidence for the importance of frontal and temporal regions. Finally, altered dopamine and glutamate neurotransmission are likely to play a part in the expression of psychotic symptoms.

The neurodevelopmental model can account for many of these findings. In short, some process (genetic or environmental) produces damage to selected brain areas

early in life. Temporal lobe regions such as the hippocampus may be particularly vulnerable. Secondary functional abnormalities develop later. As the prefrontal cortex matures in late adolescence, the behavioral and cognitive sequelae of subtle structural deficits become manifest. One result is hypofrontality and cognitive impairment. Alterations in limbic and prefrontal function then produce downstream, secondary alterations in subcortical dopamine, glutamate, and other neurotransmitter systems. Dopamine dysfunction, in particular, may lead to positive psychotic symptoms. The feasibility of this model has received substantial validation from animal studies showing the delayed behavioral and neurobiological effects of minor damage to the hippocampus in neonatal rats. Observations that children at risk for schizophrenia have a number of subtle neuropsychiatric abnormalities, such as deficits in attention, motor control, and social interactions, also support the neurodevelopmental model.

Although the neurodevelopmental hypothesis has been an important organizing heuristic since the mid-1980s several critical issues remain unresolved. First, it remains unclear when structural abnormalities actually develop. Finding such abnormalities in young children who go on to develop schizophrenia would offer strong support for this hypothesis. Alternatively, if these abnormalities develop later in life (e.g., in mid-adolescence), other mechanisms would be implicated. For example, it is unclear whether dendritic “pruning” or an apoptotic mechanism could account for volumetric reductions in areas such as the hippocampus. Observations of reduced neuronal size suggest that factors regulating this parameter could play a role. Second, despite the myriad of findings, the lack of any consistently replicable neurodevelopmental lesion in postmortem studies continues to leave the issue in doubt. It is entirely possible that no single lesion exists. Third, the issue of heterogeneity remains unresolved. Although patients with schizophrenia have structural and functional alterations as a group as compared to controls, it remains unclear whether these are necessary features of the illness. Certainly many patients are in the normal range in some or many of these measures. The same is true for most neurodevelopmental parameters. Many patients have completely normal or even above-average function in childhood and adolescence. Most patients with schizophrenia have no known history of pregnancy, obstetric, or neonatal insults. Is it possible that different patients have abnormalities restricted to differing prefrontal, temporal, or subcortical areas? Such primary lesions could induce secondary dysfunction in connected regions. Fourth, the delayed onset of psychosis presents some problems for the neurodevelopmental model. Although onset is typically in the early 20s, some patients do not develop symptoms until the fourth or even fifth decade of life. It seems most likely that such cases involve mechanisms other than or in addition to neurodevelopmental processes.

Several alternative models have been put forward to deal with some of these problems. For example, structural abnormalities could develop in adolescence, very early in the illness. It is unclear what could account for this, but candidate mechanisms might include reduction in neuronal size or excessive dendritic pruning. Neurotransmitter abnormalities, such as in the dopamine and glutamate systems, may follow. Another possibility is that some cases of schizophrenia are due to increased stress associated with entry into adulthood. This could trigger dopamine abnormalities in genetically vulnerable individuals. Structural abnormalities, in these cases, could be nonspecific vulnerability factors or could be secondary to psychosis itself. A third possibility is that schizophrenia is a heterogeneous illness with several dimensions, none of which is necessary or sufficient. Different domains could involve neurodevelopmental cortical dysfunction, dopamine and glutamate function, cortical regulation of dopamine, and interdependent functioning of a myriad of heteromodal cortical neural networks. In this model, a complex web of genetic and environmental factors could impact on these many neural networks.

One approach toward settling this issue is to examine neurobiological traits associated with schizophrenia. Such traits may be closer to the underlying physiological deficits induced by genes associated with the illness. As such, these traits may have a simpler genetic architecture, making it easier to detect their genes in linkage studies. A number of potential phenotypes have been identified that are clearly familial and thus may have a significant genetic basis. These include impaired sensory gating, eye-tracking dysfunction, perceptual aberrations, schizotypal symptoms, attentional impairment, deficit symptoms, structural brain abnormalities, and cognitive deficits. The feasibility of this approach has been validated by a recent report of linkage using a measure of impaired sensory gating. Suppression of the auditory p50 wave in a sensory gating paradigm has been linked to 15q13–14. This is very close to the a7 nicotinic cholinergic receptor, previously implicated in impaired p50 suppression. Several other preliminary reports have used eye tracking and positive psychotic symptoms. The use of such intermediate phenotypes may also reveal genes that are more important to functional outcome. Unfortunately, the heritability and genetic architecture of most intermediate phenotypes are uncertain, despite a wealth of data showing that many such traits are familial. Studies to assess these parameters and attempt linkage will require phenotyping large numbers of patients.

The underlying neurobiology of schizophrenia remains a mystery. Genetically, the disorder is complex, confounding efforts to locate causative genes. Similarly, the effects of environment are subtle, with no clear major factor emerging. Pregnancy, labor, and delivery complications may play a limited role. Increasingly sophisticated techniques, guided by greater understanding of basic neurobiology, are being used to uncover alterations in a number of brain parameters. Neurobiological abnormalities include reduced volume of several brain structures, sulcal widening, and increased ventricular size. Cortical abnormalities, particularly in the prefrontal and temporal cortices, have also been implicated by cognitive testing and functional neuroimaging. Postmortem studies have failed to find a major lesion or gliosis that could account for structural abnormalities. They have, however, detected a variety of subtle cytoarchitectural changes, perhaps caused by abnormal neurodevelopment. Several neurotransmitters, including dopamine, glutamate, and serotonin, have been implicated. The putative structural, functional, and neurochemical abnormalities can be understood in the context of the neural systems they comprise. These include cortical-striatal-thalamic loops, ascending dopamine projection pathways, and the limbic system. Interconnections between these systems make it difficult to determine which lesions are primary and which are secondary.

The neurodevelopmental model has been a critical organizing heuristic that synthesizes these seemingly disparate observations. This theory suggests that nonspecific lesions in early life, perhaps in utero, produce subtle behavioral anifestations in childhood. The onset of psychosis is delayed until brain maturation reaches later stages in late adolescence. Many questions remain unanswered, however, leaving some aspects of this theory in doubt. Combining techniques such as neuroimaging with molecular genetics provide fertile areas for future research to separate the strands that make up the tangled web of schizophrenia.

SUGGESTED CROSS-REFERENCES

Section 1.2 reviews functional neuroanatomy in greater detail. Section 1.3, Section 1.4, and Section 1.5 contain additional information on dopamine, glutamate, and other neurotransmitters. Section 1.18 describes the basic principles of genetic linkage analysis and Section 1.15 and Section 1.16 provide a more thorough discussion of the principles of neuroimaging.

SECTION REFERENCES

Akbarian S, Bunney WE, Potkin S, Wigal SB, Hagman JO, Sandman CA, Jones EG: Altered distribution of nicotinamide-adenine dinucleotide phosphate-diaphorase cells in frontal lobe of schizophrenics implies disturbance of cortical development. Arch Gen Psychiatry *50:*169, 1993.

*Bachus SE, Kleinman JE: The neuropathology of schizophrenia. J Clin Psychiatry *57:*72, 1996.

Benes FM, Sorensen I, Vincent SL, Bird ED, Sathi M: Increased density of glutamate-immunoreactive vertical processes in superficial laminae in cingulate cortex of schizophrenic brain. Cereb Cortex *2:*503, 1992.

Bertolino A, Nawroz S, Mattay VS, Barnett AS, Duyn JH, Moonen CT, Frank JA, Tedeschi G, Weinberger DR: A regionally specific pattern of neurochemical pathology in schizophrenia as assessed by multislice proton magnetic resonance spectroscopic imaging. Am J Psychiatry *153:*1554, 1996.

Breier A, Su TP, Saunders R, Carson RE, Kolachana BS, de Bartolomeis A, Weinberger DR, Weisenfeld N, Malhotra AK, Eckelman WC, Pickar D: Schizophrenia is associated with elevated amphetamine-induced synaptic dopamine concentrations: Evidence from a novel positron emission tomography method. Proc Natl Acad Sci U S A *94:*2569, 1997.

Buka SL, Tsuang MT, Lipsitt LP: Pregnancy/delivery complications and psychiatric diagnosis. A prospective study. Arch Gen Psychiatry *50:*151, 1993.

Carlsson A, Lindquist M: Effect of chlorpromazine and haloperidol on formation of 3-methoxytyramine and norepinephrine in mouse brain. Acta Pharmacol Toxicol (Copenh) *20:*140, 1963.

*Casey BJ: Brain development: Maturization in brain activation. Am J Psychiatry *156:*505, 1999.

*Creese I, Burt DR, Snyder SH: Dopamine receptor binding predicts clinical and pharmacological potencies of antischizophrenic drugs. Science *192:*481, 1976.

Crow TJ: Positive and negative schizophrenic symptoms and the role of dopamine. Br J Psychiatry *137:*383, 1980.

*Davis KL, Kahn RS, Ko G, Davidson M: Dopamine in schizophrenia: A review and reconceptualization. Am J Psychiatry *148:*1474, 1991.

Degreef G, Ashtari M, Bogerts B, Bilder RM, Jody DN, Alvir JM, Lieberman J: Volumes of ventricular system subdivisions measured from magnetic resonance images in first-episode schizophrenic patients. Arch Gen Psychiatry *49:*531, 1992.

Done DJ, Johnstone EC, Frith CD, Golding J, Shepherd PM, Crow TJ: Complications of pregnancy and delivery in relation to psychosis in adult life: Data from the British perinatal mortality survey

sample. Br J Med *302:*1576, 1991.

Egan MF, Chrapusta S, Karoum F, Lipska BK, Wyatt R: Effects of chronic neuroleptic treatment on dopamine release: Insights from studies using 3-methoxytryamine. J Neural Transm *103:*777, 1996.

Farde L, Wiesel FA, Stone-Elander S, Halldin C, Nordstrom AL, Hall H, Sedvall G: D_2 dopamine receptors in neuroleptic-naive schizophrenic patients. Arch Gen Psychiatry *47:*213, 1990.

Fish B, Marcus J, Hans SL, Auerbach JG, Perdue: Infants at risk for schizophrenia: Sequelae of a genetic neurointegrative defect. A review and replication analysis of pandysmaturation in the Jerusalem Infant Development Study. Arch Gen Psychiatry *49:*221, 1992.

Freedman R, Coon H, Myles-Worsley M, Orr-Urtreger A, Olincy A, Davis A, Polymeropoulos M, Holik J, Hopkins J, Hoff M, Rosenthal J, Waldo MC, Reimherr F, Wender P, Yaw J, Young DA, Breese CR, Adams C, Patterson D, Adler LE, Kruglyak L, Leonard S, Byerley W: Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA *94:*587, 1997.

Goldberg TE, Gold JM: Neurocognitive functioning in patients with schizophrenia: An overview. In *Psychopharmacology, The Fourth General of Progress*, FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Gottesman II, Shields J: A polygenic theory of schizophrenia. Proc Natl Acad Sci USA *58:*199, 205, 1967.

Harrison PJ: On the neuropathology of schizophrenia and its dementia: Neurodevelopmental, neurodegenerative, or both? Neurodegeneration *4:*1, 1995.

Hyde TM, Casanova MF, Kleinman JE, Weinberger DR: Neuroanatomical and neurochemical pathology in schizophrenia. In *American Psychiatric Press Review of Psychiatry*, vol 10, A Tasman, SM Goldfinger, CA Kaufmann, editors. American Psychiatric Association Press, Washington, DC, 1991.

Ingvar DH, Franzen G: Distribution of cerebral activity in chronic schizophrenia. Lancet *2:*1484, 1974.

Javitt DC, Zukin SR: Recent advances in the phencyclidine model of schizophrenia. Am J Psychiatry *148:*1301, 1991.

Kane J, Honigfeld G, Singer J, Meltzer HY: Clozapine for the treatment-resistant schizophrenic: A double-blind comparison with chlorpromazine. Arch Gen Psychiatry *45:*789, 1988.

Krimer LS, Herman MM, Saunders RC, Boyd JC, Hyde TM, Carter JM, Kleinman JE, Weinberger DR: A qualitative and quantitative analysis of the entorhinal cortex in schizophrenia. Cereb Cortex *7:*732, 1997.

Laruelle M, Abi-Dargham A, van Dyck CH, Gil R, D'Souza CD, Erdos J, McCance E, Rosenblatt W, Fingado C, Zoghbi SS, Baldwin RM, Seibyl JP, Krystal JH, Charney DS, Innis RB: Single photon emission computerized tomography imaging of amphetamine-induced dopamine release in drug-free schizophrenic subjects. Proc Natl Acad Sci U S A *93:*9235, 1996.

Liddle PF, Friston KJ, Frith CD, Hirsch SR, Jones T, Frackowiak RS: Patterns of cerebral blood flow in schizophrenia. Br J Psychiatry *160:*179, 1992.

Lipska BK, Weinberger DR: Genetic variation in vulnerability to the behavioral effects of neonatal hippocampal damage in rats. Proc Natl Acad Sci U S A *1292:*8906, 1995.

*McGuffin P, Owen MJ, Farmer AE: Genetic basis of schizophrenia. Lancet *346:*678, 1995.

McNeil TF: Obstetric factors and perinatal injuries. In *Handbook of Schizophrenia*, vol 13, *Nosology, Epidemiology and Genetics*, MT Tsuang, JC Simpson, editors. Elsevier Science, New York, 1988.

Mednick SA, Machon RA, Huttunen MO, Bonett D: Adult schizophrenia following prenatal exposure to an influenza epidemic. Arch Gen Psychiatry *45:*189, 1988.

Meltzer HY, Matsubara S, Lee JC: Classification of typical and atypical antipsychotic drugs on the basis of dopamine D_1, D_2 and serotonin 2 pK_i values. J Pharmacol Exper Therap *251:*238, 1989.

Ohuoh DC, Hyde TM, Kleinman JE: The role of serotonin in schizophrenia: An overview of the nomenclature, distribution, and alterations of serotonin receptors in the central nervous system. Psychopharmacology *112*(Suppl):S5, 1993.

Pettegrew JW, Keshavan MS, Panchalingam K, Strychor S, Kaplan DB, Tretta MG, Allen M: Alterations in brain high-energy phosphate and membrane phospholipid metabolism in first-episode, drug-naive schizophrenics: A pilot study of the dorsal prefrontal cortex by in vivo phosphorous 31 nuclear magnetic resonance spectroscopy. Arch Gen Psychiatry *48:*563, 1991.

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, Housman D, Kazazian HH, Meyers D: Sequential strategy to identify a susceptible gene for schizophrenia: Report of potential linkage on chromosome 22q12-q13.1: Part 2. Am J Med Genet *54:*36, 1994.

Reveley AM, Reveley MA, Clifford CA, Murray RM: Cerebral ventricular size in twins discordant for schizophrenia. Lancet *2:*540, 1982.

Schultz W, Dayan P, Montague PR: A neural substrate of prediction and reward. Science *275:*1593, 1997.

Seeman P, Lee T, Chau-Wong M, Wong K: Antipsychotic drug doses and neuroleptic/dopamine receptors. Nature *261:*717, 1976.

*Selemon LD, Goldman-Rakic PS: The reduced neuropil hypothesis: A circuit based model of schizophrenia. Biol Psychiatry *45:*17, 1999.

Selemon LD, Rajkowska G, Goldman-Rakic PS: Abnormally high neuronal density in the schizophrenic cortex. A morphometric analysis of prefrontal area 9 and occipital area 17. Arch Gen Psychiatry *52:*805, 1995.

Shelton RC, Weinberger DR: X-ray computerized tomography studies in schizophrenia: A review and synthesis. In *Handbook of Schizophrenia* vol 1, *The Neurology of Schizophrenia*, HA Nasrallah, DR Weinberger, editors. Elsevier, Amsterdam, 1986.

Straub RE, McLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS: A potential vulnerability locus for schizophrenia on chromosome 6p24-22: Evidence for genetic heterogeneity. Nature Genet *11:*287, 1995.

Suddath R, Christison GW, Torrey EF, Casanova MF, Weinberger DR: Anatomical abnormalities in the brains of monozygotic twins discordant for schizophrenia. N Engl J Med *322:*7879, 1990.

Torrey EF, Miller J, Rawlings R, Yolken RH: Seasonality of births in schizophrenia and bipolar disorder: A review of the literature. Schizophr Res *28*(1):1, 1997.

van Kammen DP, Kelley M: Dopamine and norepinephrine activity in schizophrenia: An integrative perspective. Schizophr Res *4:*173, 1991.

*Weinberger DR: Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry *44:*660, 1987.

*Weinberger DR, Berman KF: Prefrontal function in schizophrenia: Confounds and controversies. Philos Trans R Soc Lond B Biol Sci *351:*1495, 1996.

Wyatt RJ: Neuroleptics and the natural course of schizophrenia. Schizophr Bull *17:*235, 1991.

12.5 SCHIZOPHRENIA: GENETICS

KENNETH S. KENDLER, M.D.

The goal of this section is to provide an overview of the current state of knowledge of the genetics of schizophrenia. The following key questions are relevant. Is schizophrenia a familial disorder? To what extent is any familial aggregation of schizophrenia the result of genetic versus environmental factors? What kinds of psychiatric disorders are transmitted within families? The author believes that genetic factors play an important role in the familial transmission of schizophrenia, thus the following additional questions are pertinent: What are the most likely kinds of genetic transmission mechanisms? and What is the current status of and future prospects for identifying the chromosomal location of specific genes that predispose to schizophrenia?

IS SCHIZOPHRENIA FAMILIAL?

The most basic question in the genetics of schizophrenia is whether the disorder aggregates (or runs) in families. Technically, familial aggregation means that a close relative of an individual with a disorder is at increased risk for that disorder compared to a matched individual chosen at random from the general population. This chapter reviews family studies of schizophrenia examining primarily first-degree relatives (parents, full siblings, and offspring) because little systematic information on more distant relationships has been gathered in recent years.

In a 1967 review paper Edith Zerbin-Rüdin listed 17 major family studies of schizophrenia involving first-degree relatives. By 1980 at least 9 other major studies had been reported. All these studies consistently showed a substantially greater risk for schizophrenia in the close relatives of persons with schizophrenia than would be expected in the general population.

However, nearly all these studies suffered from three methodological limitations. First, because no control groups were used, the rates of schizophrenia in the general population required for comparison had to be derived from the literature. Second, diagnoses were made nonblind, with the research team always knowing that the individual being evaluated was a relative of a schizophrenic individual. Third, neither structured personal interviews nor operationalized diagnostic criteria were used. In fact in many of the early studies it is unclear how many individuals were personally examined and how many were evaluated from indirect information such as reports of relatives or doctors, or from hospital notes.

In the early 1980s several research groups questioned the validity of earlier family studies of schizophrenia. These researchers suggested that the evidence for the familial aggregation of schizophrenia may result from consistent biases in the previous studies. In addition, they were concerned that the diagnostic approach to schizophrenia in these earlier studies might have been overly broad. They argued that the familial aggregation of schizophrenia might be weak or absent when narrowly diagnosed. Since 1980, 11 major family studies of schizophrenia have been reported that used blind diagnoses, control groups, personal interviews, and operationalized diagnostic criteria. These studies permit a more rigorous evaluation than has hitherto been possible of the degree to which schizophrenia aggregates in families.

The key results from these studies are summarized in Table 12.5-1, which contains the diagnostic criteria used in the study, the nature of the control proband group, and the *p* value (i.e., the probability of observing such a difference in the rates of schizophrenia in the two groups by chance, if the true rates were identical). The term *proband* refers to the individual through whom the family was identified for study. A typical family study of schizophrenia would then begin with two types of probands: those with schizophrenia and a matched group of control probands. Relatives of these probands would then be systematically assessed. Table 1 also presents the *lifetimes at risk* in the assessed relatives of schizophrenia and control probands, and the *morbid risk* for schizophrenia in the two groups. Lifetimes at risk is the sum for all assessed relatives of the proportion of their lifetime risk for schizophrenia they have completed thus far. *Morbid risk* (MR) is a statistic commonly used in genetics and equals the total proportion of individuals who would be expected to be affected with a disorder in a given population if all members of that population have completed their age at risk. Finally, the table includes the correlation of liability. If schizophrenia is caused by several genetic and environmental factors that act approximately additively in influencing an individual's liability or predisposition to schizophrenia, then this figure represents the degree of correlation between first-degree relatives in overall risk of the disease. This is a very useful figure because it combines into a single, easily understood statistic the risk figures for schizophrenia in relatives of schizophrenic and control probands. The *higher* the correlation of liability, the *stronger* is the degree of familial aggregation of schizophrenia.

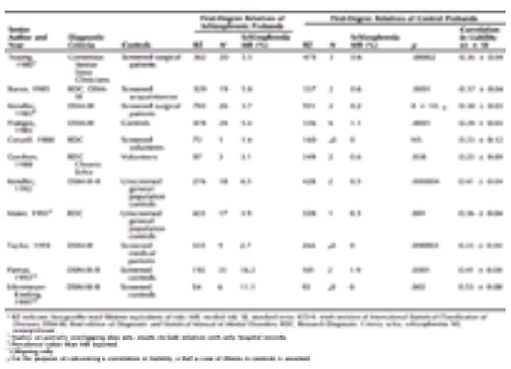

Table 12.5-1 Summary Results of Major Recent Family Studies of Schizophrenia That Included a Normal Control Group, Personal Interviews With Relatives, and Blind Diagnosis of Relatives*

Before turning to the major results summarized in this table, two preliminary comments are in order. First, the sample of relatives studied varies very widely in the different investigations. For example, the lifetimes at risk in relatives of schizophrenic patients range over tenfold, from over 700 to 54. On average, the larger studies will provide more stable statistical estimates for the true risk of schizophrenia in relatives of schizophrenia and control probands. Second, various diagnostic criteria were utilized in the different studies. However, eight studies used criteria from either the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), Research Diagnostic Criteria (RDC) (for chronic schizophrenia), or revised third edition of DSM (DSM-III-R) criteria, all of which require at least 6 months of illness, usually with functional impairment, in addition to specified psychotic symptoms. Third, while nine of the studies examined all available first-degree relatives (parents, full-siblings, and offspring), two studies examined only offspring.

There are five major conclusions that can be drawn from the large body of work summarized in the table. First, the risk for schizophrenia in relatives of schizophrenic probands varies widely across studies, from a low of 1.4 percent to a high of 16.2 percent. Much of this fluctuation may probably be attributable to differences in diagnostic criteria or statistical fluctuations in small samples (the lowest risk is found in the smallest study). However, it remains possible that there are true population differences in the risk for schizophrenia in relatives of schizophrenic probands.

Second, the risk for schizophrenia in the relatives of nonpsychiatric control probands is relatively similar across studies, ranging—with the exception of the study of

Parnas et al—from only 0.2 to 1.1 percent, corresponding closely to the range of risks for schizophrenia found in general population studies.

Third, in every study the risk for schizophrenia was higher in the relatives of schizophrenic probands than in relatives of control probands. Across these studies the risk of schizophrenia was, on average, 11 times greater in relatives of schizophrenic probands than in relatives of matched control probands. Fourth, in all but one study, the difference in risk for schizophrenia in the relatives of schizophrenic and control probands was quite unlikely to be attributable to chance (p-value < 0.05). In a number of studies, the p-values were very low (less than 0.001), indicating that such differences in risk would be extremely unlikely to occur by chance.

Finally, although there was some variation, the correlation in liability for all studies fell in the range from +0.23 to +0.53, with a weighted mean across the 11 studies of +0.35. Most of the largest studies that used relatively narrow diagnostic criteria for schizophrenia obtained correlations of liability in the narrow range of +0.32 to +0.41 because the highest and lowest correlations in the table come from the smallest and next to smallest studies, respectively. These results suggest that most of these studies can be seen as replications of one another because they provide similar results on the observed degree of familial aggregation of schizophrenia. The correlation of liability between first-degree relatives in the range of +0.30 to +0.40 indicates a relatively strong degree of familial aggregation.

In conclusion, the questions raised in the early 1980s about the degree of familial aggregation of schizophrenia can now be addressed satisfactorily. The results of a large number of recent, carefully performed family studies support the conclusions of earlier and less methodologically rigorous investigations in finding that schizophrenia strongly aggregates in families. The familial aggregation of schizophrenia appears to be quite substantial when it is defined using modern, relatively narrow diagnostic criteria such as those found in DSM-III-R and DSM-IV. On average, the risk for schizophrenia in the relatives of controls is between 0.5 and 1.0 percent, compared to between 3 and 7 percent in relatives of schizophrenic probands in most studies. The best estimate of the correlation in liability to schizophrenia in first-degree relatives is probably between +0.3 and +0.4.

FAMILIAL AGGREGATION: GENETIC VERSUS ENVIRONMENTAL FACTORS

Twin Studies Resemblance among relatives can be ascribed to shared environment (nurture) or shared genes (nature). A major goal in psychiatric genetics is to determine the degree to which familial aggregation for a disorder like schizophrenia results from environmental versus genetic mechanisms. Although sophisticated analyses of family data can begin to make this discrimination, nearly all that is known about this problem in schizophrenia comes from twin and adoption studies.

Twin studies are based on the assumption that monozygotic (MZ) and dizygotic (DZ) twins share a common environment to approximately the same degree. However, MZ twins are genetically identical, whereas DZ twins (like full siblings) have on average only half of their genes in common. Although the validity of the second assumption is beyond question, the first (equal environment) assumption has been a focus of considerable controversy.

Several studies have shown that measures of the social environment (for example, common friends, attitudes of parents and teachers) are more highly correlated among young MZ twins than among young same-sex DZ twins. These results at first appear to suggest that the equal environment assumption is false. However, there is another possible interpretation. Similarity in environment might make MZ twins more similar, but it is also plausible that by behaving alike, MZ twins seek out or create more similar environments for themselves. These two alternative hypotheses have been empirically evaluated in a number of studies, nearly all of which suggest that the environmental similarity of MZ twins is the *result* and not the cause of their behavioral similarity. Current evidence from an increasingly wide range of studies supports the general validity of the equal environment assumption of twin studies.

Results are available from 13 major twin studies of schizophrenia (Table 12.5-2). None of these, however, meets all the methodological criteria outlined above for family studies and the additional criterion that zygosity assignment be made blind with respect to psychiatric diagnosis. Some studies come closer to this model than others. For example, a variety of different clinicians made diagnoses from blind case abstracts in the original report from the Maudsley twin series of Irving Gottesman and James Shields. These case records have more recently been examined using modern operationalized criteria with similar overall results. In the study by Kenneth Kendler and Dennis Robinette from the National Academy of Sciences-National Research Council (NAS-NRC) Registry, psychiatric diagnoses were collected from a wide variety of clinical settings in which clinicians could not possibly have been aware of any research hypotheses. Furthermore, it could be shown that zygosity assignment was not biased with respect to psychiatric diagnosis. The new Norwegian and Finnish studies used the high-quality twin and psychiatric registries in Norway and Finland and thus should be representative of all treated cases of illness. Whereas the Norwegian study was based on personal psychiatric assessments and performed with structured instruments and DSM-III-R operationalized criteria, the Finnish study used previously recorded hospital and disability diagnoses. The sample size of the Finnish study was relatively large (253 pairs) while the Norwegian sample was much more modest in size (52 pairs). Unfortunately, both studies relied on self-report zygosity measures and the interviews and diagnoses in the Norwegian study were performed nonblind.

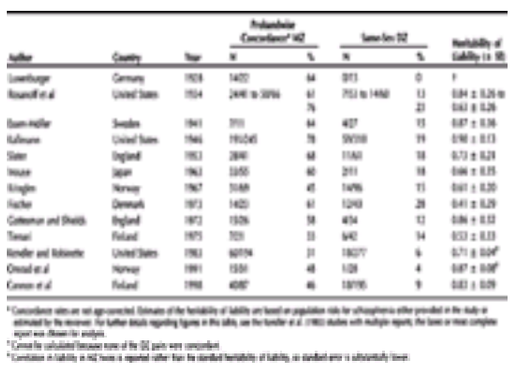

Author	Country	Year	Proband-wise Concordance* MZ		Same-Sex DZ		Heritability of Liability (± SE)
			N	%	N	%	
Luxenburger	Germany	1928	14/22	64	0/13	0	†
Rosanoff et al	United States	1934	24/40 to 33/56	61 76	7/53 to 14/60	13 23	0.84 ± 0.26 to 0.63 ± 0.26
Essen-Möller	Sweden	1941	7/11	64	4/27	15	0.87 ± 0.36
Kallmann	United States	1946	119/245	78	59/310	19	0.90 ± 0.13
Slater	England	1953	28/40	68	11/63	18	0.73 ± 0.23
Inouye	Japan	1963	33/55	60	2/11	18	0.66 ± 0.35
Kringlen	Norway	1967	51/69	45	14/96	11	0.61 ± 0.20
Fischer	Denmark	1973	14/23	61	12/43	28	0.41 ± 0.29
Gottesman and Shields	England	1972	15/26	58	4/14	12	0.86 ± 0.33
Tienari	Finland	1975	7/21	33	6/42	14	0.53 ± 0.33
Kendler and Robinette	United States	1983	60/174	31	18/277	6	0.71 ± 0.04‡
Onstad et al	Norway	1991	13/31	48	1/28	4	0.87 ± 0.08‡
Cannon et al	Finland	1998	40/87	46	13/195	9	0.83 ± 0.09

* Concordance rates are not age-corrected. Estimates of the heritability of liability are based on population risks for schizophrenia either provided in the study or estimated by the reviewer. For further details regarding figures in this table, see the Kendler et al. (1993) review. For studies with multiple reports, the latest or most complete report was chosen for analysis.
† Cannot be calculated because none of the DZ pairs were concordant.
‡ Correlation in liability in MZ twins is reported rather than the standard heritability of liability; no standard error is available for this estimate.

Table 12.5-2 Concordance With Respect to Probands and the Heritability of Liability to Schizophrenia in the Major Twin Studies Reported to Date

All these studies agree that proband-wise concordance for schizophrenia (the risk for schizophrenia in the cotwins of a schizophrenic proband twin) is much higher in MZ than in DZ twins, but the absolute rates of concordance vary widely. Two factors are probably responsible for most of this variation. First, some studies defined schizophrenia more broadly than others. Second, some studies obtained most of their proband twins from chronically hospitalized populations; others used population-based registries where milder cases would commonly occur. Twin studies have often but not always found a positive relationship between concordance and severity of illness.

Heritability of Liability The diagnostic approach to schizophrenia and the method of ascertaining probands should equally affect concordance rates in MZ and DZ twins. Therefore, a better method of comparing results across studies would be a summary statistic based on concordance in both MZ and DZ twins. One of the best of these is the heritability of liability as calculated from the correlations in liability in MZ and DZ twins. This statistic ranges from 0.0 if genetic factors play no role in susceptibility to a disorder to a maximum of 1.0 if genes entirely determine disease risk. Because this statistic is based on the polygenic multifactorial threshold model, which may or may not be appropriate for schizophrenia, these results should be regarded as only one plausible way of approximating reality. Nonetheless, the major twin studies of schizophrenia agree in estimating the heritability of liability of schizophrenia at between 0.6 and 0.9 (Table 12.5-2). These results suggest that genetic factors play a major role in the familial transmission of schizophrenia.

Genetic theory predicts that if all the familial aggregation of schizophrenia were due to genetic factors, then the heritability of liability should be approximately double the correlation in liability found in first-degree relatives (because, on average, first-degree relatives have half of their genes in common). Comparing the results of Table 12.5-1 and Table 12.5-2 indicates that, at least as a rough approximation, this hypothesis is supported. The range of the heritability of liability to schizophrenia calculated from twin studies is approximately twice the range of the correlation in liability to schizophrenia found in first-degree relatives in most family studies.

Nongenetic Familial Transmission Twin studies also provide two powerful tests for the role of nongenetic familial transmission in the liability to schizophrenia. First, one can ask whether the correlation in liability in DZ twins is more than half that which would be predicted in MZ twins if only additive genetic factors were operating. A review of all major twin studies to date suggests that nongenetic factors may play at most a modest role in the transmission of schizophrenia. Second, the risk for schizophrenia in DZ co-twins can be compared with that in siblings of schizophrenic probands. Although having the same degree of genetic relationship to the affected proband, DZ co-twins certainly share more of the familial environment that do ordinary siblings. Several twin studies have suggested that a difference in risk does exist between these two groups. However, such a difference has not been consistently found across all studies and was *not* found in the recent Norwegian

small-sample twin family study of schizophrenia.

Adoption Studies Adoption studies can clarify the role of genetic and environmental factors in the transmission of schizophrenia by studying two kinds of rare but informative relationships: (1) individuals who are genetically related but do not share familial-environmental factors, and (2) individuals who share familial-environmental factors but are not genetically related. Table 12.5-3 summarizes, in the order discussed, the major adoption studies of schizophrenia, reporting raw data and statistical tests. Our summary here will be organized by the kind of adoption design utilized.

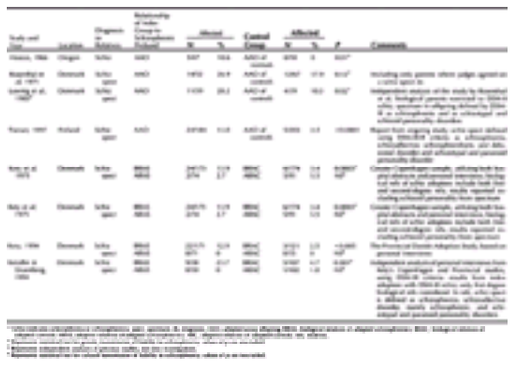

Table 12.5-3 Summary Results of Major Adoption Studies of Schizophrenia*

Affected Biological Parent Design Three studies have compared the adopted-away offspring of schizophrenic parents with the adopted-away offspring of matched controls. In the first of these, Leonard Heston found a significant excess of schizophrenia in adopted-away offspring of schizophrenic versus control mothers. The second such study was performed in Denmark under the direction of David Rosenthal and found similar results, which, however, fell short of statistical significance, particularly when only parents with a consensus diagnosis of schizophrenia or schizophrenia spectrum were included. This study has been the subject of a blind reanalysis using DSM-III criteria, which, when including only biological parents with a consensus diagnosis of schizophrenia from the original investigators, found a significant excess of schizophrenia spectrum in adopted-away offspring of schizophrenic versus control parents. The third and by far the largest such study is still under way in Finland under the direction of Pekka Tienari. Preliminary results indicate a highly statistically significant excess of schizophrenia and schizophrenia spectrum disorders in the adopted-away offspring of schizophrenic mothers compared to the adopted-away offspring of matched control mothers.

Affected Adoptee Design Another major adoption strategy used for studying schizophrenia begins with ill adoptees rather than with ill parents. The full implementation of this design permits two separate experiments: (1) a test for the etiological role of shared environmental factors by comparing the nonbiological adoptive relatives of the schizophrenic with the control adoptees and (2) a test for the etiological role of genetic factors by comparing the biological relatives of the schizophrenic with control adoptees who were raised in households away from their ill relatives. This strategy has been used by Seymour Kety and colleagues in a series of adoption studies carried out in Denmark. The first sample, called the *Copenhagen sample*, began with 34 adoptees located in Copenhagen who received a consensus diagnosis of chronic, borderline, or acute schizophrenia. These adoptees and their matched controls had been separated from their biological parents at an early age and raised by individuals with whom they had no biological relationship. The first report on this series was based on hospital abstracts of all relatives located by the population and psychiatric registries available in Denmark. Schizophrenia and related disorders were significantly concentrated only in the biological relatives of the schizophrenic adoptees. The next phase of this project involved personal interviews of all available and cooperative relatives. After these interviews had been dictated into English and blinded, a diagnostic review also indicated a substantial concentration of schizophrenia spectrum disorders only in the biological relatives of the schizophrenic adoptees.

A second sample beginning with 41 schizophrenia spectrum adoptees from outside of Copenhagen (termed the *Provincial sample*) has also been collected. On the basis of personal interviews with biological and adoptive relatives, the results basically replicated the parallel findings from the Copenhagen sample. Schizophrenia and the spectrum disorders were significantly more common in the biological relatives of schizophrenic versus control probands, while these disorders were equally uncommon in the two groups of adoptive relatives. In an independent review of the interviews from both the Copenhagen and the Provincial samples, using DSM-III criteria, Kendler and Gruenberg replicated and extended all the major earlier findings of Kety and his coworkers.

Vertical Cultural Transmission Because studies of twins contain no parent-offspring pairs, they are not helpful in clarifying whether parents influence their children's risk for schizophrenia in ways that are other than genetic. However, several adoption strategies have been used to clarify the role of parent-offspring environmental transmission (termed *vertical cultural transmission*) in schizophrenia. First, if offspring of schizophrenic persons in part “learn” schizophrenia from their parents, then decreasing the amount of contact between schizophrenic parents and their children should decrease their risk for illness; however, two studies have produced results inconsistent with this hypothesis. Jerry Higgins and coworkers compared the adopted-away offspring of schizophrenic parents with the naturally reared offspring of schizophrenics. Although the sample size was small (23 offspring in each group), follow-up personal interviews indicated a nonsignificant excess of schizophrenia in the adopted-away offspring compared to those children reared by a schizophrenic parent. In a variation of a full adoption study, a similar design was utilized in Israel to compare 25 offspring of schizophrenic parents reared in a kibbutz (where children are raised together in children's houses, yet have considerable contact with their parents) with 25 offspring of schizophrenic parents raised in conventional nuclear-family settings elsewhere in Israel. A follow-up interview in adulthood with these offspring revealed that the risk for DSM-III schizophrenia was actually higher (nonsignificantly) in the kibbutz-reared offspring (13 percent) than in the town-reared offspring (8.7 percent). When the schizophrenia spectrum disorders were included this difference was even greater (26.1 percent versus 13.0 percent) although still short of statistical significance. Both these results are inconsistent with the vertical cultural transmission hypothesis.

A second way to address the vertical cultural transmission hypothesis is to look at the risk for schizophrenia in the adopted-away offspring of normal individuals reared by schizophrenic parents. Although limited by a small sample size and a small number of parents with typical schizophrenia, Paul Wender and coworkers found no evidence for increased rates of illness in such adoptees.

Third, vertical cultural transmission of schizophrenia would predict that among adopted individuals who become schizophrenic, schizophrenia should be overrepresented in their adoptive parents, who would culturally “transmit” schizophrenia. In Kety's Copenhagen and Provincial samples, no excess cases of schizophrenia or related spectrum conditions were seen in the adoptive parents of the schizophrenic adoptees. In separate samples Wender et al twice studied psychopathology in the adoptive parents of schizophrenic adoptees. The first of these studies found evidence for an excess of severe psychopathology in the adoptive parents of schizophrenics. In the second study, which the authors believed was better controlled, no such increase was found.

Fourth, since step-siblings of schizophrenics would be exposed to the same schizophrenogenic rearing environment as the schizophrenic person but would lack the biological relationship to the parents, vertical cultural transmission would also predict an excess of schizophrenia in the step-siblings of schizophrenic persons. Two studies have been unable to find an excess risk for schizophrenia in relatively small samples of step-siblings of schizophrenic probands.

In summary, twin and adoption studies provide strong and consistent evidence that genetic factors play a major role in the familial aggregation of schizophrenia. Evidence for a role for nongenetic familial factors is less clear; some studies suggest that they may contribute modestly to the familial aggregation of schizophrenia, but the majority of studies find no evidence for significant nongenetic familial factors for schizophrenia.

PSYCHIATRIC DISORDERS TRANSMITTED WITHIN FAMILIES

The first systematic family study of schizophrenia, performed by Ernst Rüdin in Emil Kraepelin's newly established Psychiatric Institute in Munich in 1916, found that siblings of schizophrenic patients had increased rates not only of schizophrenia but also of other potentially related psychotic disorders as well. Since that time, a major focus of family, twin, and adoption studies of schizophrenia has been to clarify the nature of the psychiatric syndromes that occur in excess frequency in relatives of schizophrenic patients. This effort has been greatly aided by the emergence of operationalized diagnostic criteria in psychiatry, which permit more precise and reliable diagnoses.

Hypotheses On the level of psychopathological syndromes, four heuristic hypotheses can be articulated about the nature of the liability to schizophrenia that is

transmitted in families: (1) a general (nonspecific) liability to all psychiatric illnesses; (2) a liability to poor psychosocial functioning, oddness, and suspiciousness (schizophrenialike personality disorders); (3) a liability to many forms of psychosis; and (4) a specific liability to typical schizophrenia. These hypotheses are useful because each generates a different prediction about the kinds of psychiatric disorders that should be seen in excess in families of schizophrenic persons.

Nonspecific Liability for All Psychiatric Disorders The first hypothesis predicts that the risk for all major forms of psychiatric illness should be increased in relatives of schizophrenic persons. The hypothesis would be consistent with the unitary hypothesis of mental disorders, which postulates that all psychiatric illness is on a single continuum with schizophrenia at the most deviant end. This hypothesis can be best evaluated in modern family studies and in reanalyses of major adoption studies that have used similar diagnostic criteria and normal control groups.

In the modern family studies that have examined this question, there is nearly uniform agreement that the rates of anxiety disorders and substance dependence disorders are *not* increased in relatives of schizophrenic versus matched control probands. The question of the familial relationship between schizophrenia and affective illness is more controversial. The majority of family and adoption studies that have examined this issue report similar rates for unipolar and bipolar illness in relatives of schizophrenic and control probands. However, at least three recent studies have found a significantly greater risk for unipolar illness in the relatives of schizophrenic probands. The reason for these discrepant findings remain a subject of debate.

Schizophrenia-Like Personality Disorders Both Kraepelin and Eugen Bleuler, the two chief architects of our concept of schizophrenia noted that some close relatives of patients with schizophrenia, although never psychotic, had odd or eccentric personalities that were clinically reminiscent of schizophrenia. Since that time, similar observations have been made by a number of clinicians and researchers. The first and probably the most influential rigorous study of what may be termed these *schizophrenia-related personality disorders* was made by Kety and colleagues in the Danish Adoption Studies referred to previously. Based on a blind diagnostic review with their own diagnostic criteria, Kety and colleagues found a statistically significant excess rate of borderline and uncertain schizophrenia in the biological relatives of schizophrenic versus control adoptees.

More recent applications of operationalized criteria have replicated and extended these earlier findings in support of the second hypothesis. Since 1983 11 family studies have examined the risk for *schizophrenia spectrum*, defined as schizotypal or paranoid personality disorder using DSM-III or DSM-III-R criteria, in relatives of schizophrenic and matched normal control probands (Table 12.5-4). These studies included two reanalyses of different Danish Adoption samples, a Finnish adoption sample, four family studies conducted in the United States, and one family study each conducted in Greece, Ireland, and Germany. The absolute rates of schizotypal and paranoid personality disorder in relatives of both schizophrenic and control probands differ widely across studies. This might be expected because of the quite different approaches used to assess these personality disorder syndromes. However, every study found that schizotypal or paranoid personality was more common in relatives of schizophrenic versus control probands and this difference was statistically significant in 9 of the 11 studies. In aggregate, these results provide strong support for the second hypothesis articulated—that the familial liability to schizophrenia is in part reflected by a set of personality traits related to social isolation, oddness, and suspiciousness.

Author and Year	Study Group	Relatives of Schizophrenic Probands			Relatives of Control Probands			P
		SPD or PPD			SPD or PPD			
		N[a]	N	MR ± SE	N[a]	N	MR ± SE	
Lowing, 1983	Adopted-away offspring	39	6	15.4 ± 5.0	39	3	7.7 ± 4.3	.29
Kendler, 1994[b]	Biological relatives of adoptees	35	6	17.1 ± 6.4	106	4	3.8 ± 1.9	.03
Tienari, 1997[b]	Adopted-away offspring	104	8	4.3 ± 1.2	200	1	0.5 ± 0.5	.03
Baron, 1985	Nuclear family study	329	72	21.9 ± 2.3	337	16	4.7 ± 1.2	$<10 \times 10_{-6}$
Frangos, 1985	Nuclear family study	478	13	2.7 ± 0.7	536	3	0.6 ± 0.3	.006
Coryell, 1988	Nuclear family study	72	3	4.2 ± 2.3	160	4	2.5 ± 1.2	.49
Gershon, 1988	Nuclear family study	108	3	2.8 ± 1.6	380	0	0	.03
Kendler, 1991	Nuclear family study	319	26	8.2 ± 1.5	580	10	1.7 ± 0.5	.000003
Parnas, 1993	Nuclear family offspring only	192	41	21.3 ± 3.0	191	5	3.0 ± 2.2	$<.0001$
Maier, 1994	Nuclear family study	289	6	2.1 ± 0.8	320	1	0.3 ± 0.3	.04
Erlenmeyer-Kimling, 1995[b]	Nuclear family offspring only	44	4	9.1 ± 4.3	90	1	1.1 ± 0.1	.02

[a] MR indicates lifetime morbidity risk; SPD, schizotypal personality disorder; PPD, paranoid personality disorder; MR, morbid risk; SE, standard error.
[b] Prevalence reported rather than MR.

Table 12.5-4 Summary Results of Major Family and Adoption Studies Using Personal Interviews to Examine Risk for DSM-III and DSM-III-R Schizotypal or Paranoid Personality Disorder in First-Degree Relatives of Schizophrenic and Normal Control Probands

General Liability to Psychosis or Specific Liability to Typical Schizophrenia Recent family and adoption studies of schizophrenia have also provided us with substantial data in favor of the third hypothesis presented—that what is transmitted in families of individuals with schizophrenia is a liability to many forms of psychosis. In most but not all the studies that have examined this question, the risk for nonschizophrenic psychotic disorders (e.g., schizophreniform disorder, schizoaffective disorder, delusional disorder, and psychotic disorder not otherwise specified) is increased in relatives of schizophrenic probands compared to the risk seen in relatives of controls. Furthermore, although examined in fewer studies, the risk for schizophrenia appears to be usually increased in the relatives of probands with nonschizophrenic psychotic disorders.

Another specific test of the third hypothesis is to examine the frequency of psychotic affective illness in relatives of schizophrenic patients. In large-scale family studies in Iowa and in Ireland, the author and colleagues found that while relatives of patients with schizophrenia were not at increased risk for affective illness, if affectively ill, they were more than twice as likely to become psychotic as affectively ill relatives of controls. Furthermore, compared to controls, in both studies the relatives of probands with psychotic affective illness were at increased risk for schizophrenia.

Results to date provide strong evidence against the validity of the first and fourth hypotheses. The familial predisposition to schizophrenia is neither completely nonspecific nor highly specific. Results are available to strongly support the second hypothesis and also to provide some evidence in favor of the third hypothesis. Current evidence suggests that the familial liability to schizophrenia increases not only the risk for schizophrenia as it is narrowly defined but also for schizotypal and paranoid personality disorder and probably several nonschizophrenic psychotic illnesses. These findings provide an increasingly complex but informative picture of the nature of the transmitted liability to schizophrenia.

MOST LIKELY MECHANISMS OF KINDS OF GENETIC TRANSMISSION MECHANISMS

The conclusion that genes account for a substantial proportion of the risk for schizophrenia and related psychiatric disorders naturally leads to a desire to understand the mechanisms that underlie such genetic transmission. Is the risk for schizophrenia the result of many genes of small individual effect or can a single gene, acting alone, transmit a major risk of developing schizophrenia? For many years, genetic epidemiological studies have been carried out with the goal of answering these questions. Unfortunately, for reasons we will discuss fourth most of these questions have not yet been definitively answered.

The major genetic-epidemiological approach to this question prior to the advent of linkage studies was statistical modeling either of risk figures for the various classes of relatives of schizophrenia or of the observed pattern of schizophrenia within systematically collected samples of nuclear families or extended pedigrees. This latter technique is called *complex segregation analysis*. In contrast to linkage analysis, complex segregation analysis examines phenotypes only, and *not* genetic marker information. Both these strategies compare the observed patterns of co-occurrence of schizophrenia among family members of close versus more distant genetic relatives with the degree of sharing expected under alternative genetic models.

Studies of the patterns of risk for schizophrenia in major classes of relatives (e.g., monozygotic and dizygotic co-twins, parents, siblings, offspring, nieces/nephews) have usually concluded that the familial transmission of schizophrenia *cannot* be explained solely by a single major locus (SML). In particular, the concordance rate in monozygotic twins (~50 percent) is too high, relative to the risk in siblings and dizygotic twins (5–10 percent). Such a pattern is more consistent with multiple interacting (or epistatic) loci.

With regard to complex segregation analysis, the results to date have been frustratingly inconclusive; however, they have been more frequently inconsistent than consistent with an SML etiology for schizophrenia. Given the genetic complexity of schizophrenia, it is likely that even the most mathematically sophisticated methods of complex segregation analyses are analytic tools too blunt to provide definitive answers.

CURRENT STATUS AND FUTURE PROSPECTS FOR IDENTIFYING SPECIFIC GENES THAT PREDISPOSE TO SCHIZOPHRENIA

Two fundamentally distinct strategies have been employed in attempts to find specific genes that confer susceptibility to schizophrenia: tests of *linkage* and *association*. What follows is a review of the current progress using both these methods.

Linkage Studies of Schizophrenia Linkage has proven to be an immensely powerful method for simple or Mendelian disorders in which a small number of families can usually unambiguously produce strong evidence for linkage to a small chromosomal region. As discussed later, despite much effort such a result has not emerged for schizophrenia. Why? As outlined in Table 12.5-5, schizophrenia differs from Mendelian disorders in at least five critical ways, all of which make successful linkage studies much more difficult.

Characteristics	Mendelian Disorders	Schizophrenia
Penetrance*	Usually complete—MZ concordance ~100%	Incomplete—MZ concordance 30%–70%
Phenocopies	Usually absent	Present
Diagnostic boundaries	Clear	Uncertain
Locus heterogeneity within families	Never	Uncertain, but likely
Locus heterogeneity across families	Variable, but often absent	Uncertain, but likely

* The probability of illness given the disease-predisposing genotype.

Table 12.5-5 Characteristics of Classical Mendelian Disorders Versus Schizophrenia

First, most Mendelian disorders are *fully penetrant*. That is, if you inherit a "disease gene" and live through the period of risk, you will nearly always suffer from the disorder. As outlined above, the pattern of illness in families—in which first-degree relatives of schizophrenic probands have a risk of schizophrenia of ~5 to 8 percent—and the concordance rate in monozygotic twins (40 to 55%) are inconsistent with the action of a highly penetrant SML. Unlike in Alzheimer's disease when a well-recognized series of pedigrees exist in which the disorder segregates as a Mendelian dominant, no one has ascertained pedigrees in which schizophrenia is transmitted as a classical Mendelian disorder. Finally, the offspring of unaffected MZ cotwins have an elevated risk of illness, suggesting that cases of schizophrenia cannot be simply divided into "genetic" and "sporadic" forms. These results all suggest that, in aggregate, genes that exist for schizophrenia have *reduced penetrance*. That is, it is possible to carry a susceptibility gene or genes for schizophrenia and not manifest the illness.

Second, for most Mendelian conditions, in all individuals who manifest typical symptoms of the disease, the symptoms are caused by the disease gene. This is not true for schizophrenia as schizophrenia-like symptoms can be produced by drugs of abuse and more rarely by metabolic or neurological conditions; these are called *phenocopies*.

Third, in nearly all Mendelian disorders, disease development is independent of the environment. This is not the case in schizophrenia, however, when several environmental risk factors, including season of birth, obstetric complications, and intrauterine influenza infections have been shown to increase the risk of illness.

Fourth, in most Mendelian disorders there is an obvious discontinuity between affected and unaffected individuals. Diagnostic boundaries are less clearly delineated for psychiatric illness and are the subject of continued debate in linkage studies. The correct phenotypic boundaries for schizophrenia to use in linkage and association studies are not known.

Fifth, the disease genes that cause Mendelian disorders are sufficiently rare that for all practical purposes the same disease locus will be responsible for all the cases of illness in a pedigree. Schizophrenia is much more common and it is plausible, although unproven, that two or more loci contribute to disease susceptibility in many high-density families.

Across families the situation with Mendelian disorders is variable. In most disorders (e.g., Huntington's disease, cystic fibrosis) mutations at a single locus are responsible for all known cases of illness. However, for some Mendelian syndromes (e.g., limb-girdle muscular dystrophy and retinitis pigmentosa) a number of distinct loci, usually on different chromosomes, have been found in different families. Given that schizophrenia is not a disease but a broad behavioral syndrome, and given the great complexity of the human brain, it is plausible that mutations in many different genes might result in the condition we call schizophrenia.

The tight 1:1 relationship that exists between disease gene and phenotype that is characteristic of Mendelian genetic disorders does *not* apply to schizophrenia. Locus heterogeneity both within and between families is likely to further complicate this picture. These problems all mean that the signal:noise ratio for linkage studies will be much lower for schizophrenia. As is true in any experimental design, the lower the signal:noise ratio the larger the sample size required to detect an effect reliably. This principle is particularly applicable to linkage studies of complex diseases like schizophrenia. As has been demonstrated by a range of formal power analyses, much larger sample sizes are likely be required to detect linkage reliably for schizophrenia than were needed for the genetically simple Mendelian disorders.

Developments and Early Results The field of linkage studies of schizophrenia is in flux and changing rapidly. Certainly the most important development in the last several years has been the emergence of a small handful of replicated findings. Given the vast number of statistical tests that are now performed in most linkage studies (many markers, several diagnostic models, several genetic models, often several different linkage programs), the true error rate emerging from any individual study is nearly impossible to quantify with precision. It remains a major concern that results of quite high apparent statistical significance could occur by chance alone because so many individual tests are performed.

Thus, replication is critical. However, there remain many justifications for nonreplication, including genetic variation between populations and differences in sample size, composition, and diagnostic approach. Given the evidence of replication across several groups for regions on 22q, 6p, and 8p, many would argue that it is increasingly unlikely that all these regions represent false positives. It is difficult to conceive of some inbuilt bias that would produce spuriously positive results across multiple groups in the same chromosomal region.

A further important development in the field has been the recent emergence of results from total genome scans for other complex disorders including juvenile-onset diabetes mellitus, multiple sclerosis, and asthma. These studies may hold important lessons for our struggles with schizophrenia. For juvenile-onset diabetes and multiple sclerosis, a "major" gene appears to exist in the HLA region that has been detected in nearly all studies. However, in other regions, nonreplication across groups is as frequent as replication and this is the pattern seen in initial studies of asthma. A recent study of asthma performed a genome scan in sibling pairs from three different U.S. ethnic groups; almost no replication for putative regions of linkage was seen across these three samples.

These results suggest that the difficulties in detecting replicable linkages for schizophrenia may not be unique to the psychiatric disorders, but rather may reflect a general pattern of problems associated with linkage studies of complex disorders. When the simple and powerful one-to-one relationship between gene and phenotype that is seen in Mendelian disorders breaks down, linkage studies change from being relatively straightforward although arduous tasks to uncertain and murky endeavors. Detecting genes of modest-effect size for complex moderately heritable diseases is likely be a difficult and sometimes frustrating task.

The first linkage study of schizophrenia was reported by Constantinidis in 1958. This study of 36 pairs of siblings was unusual in that in addition to blood groups, it used as "markers" a wide array of phenotypic traits such as hair color, handedness, and hair whorl. Seven further linkage studies of schizophrenia published between 1973 and 1987 examined, in relatively small samples of sib pairs or pedigrees, single markers (dopamine-b-hydroxylase, Gc, and albinism), HLA or a limited range of traditional polymorphisms (e.g., blood groups, proteins, and red cell enzymes).

Although prior evidence for linkage in schizophrenia had been reported (e.g., Turner reported a logarithm of the odds [LOD] score for HLA and "schizotaxia" of 2.57,

which could not be replicated by subsequent investigators), the first widely noted positive evidence for linkage in schizophrenia and the first study to use DNA polymorphisms was reported by Sherrington et al in 1988. Following up on the report of an association of schizophrenia with a partial trisomy of chromosome 5q, these researchers examined the proximal portion of the trisomy, including the 5q11.2 breakpoint, with 2 RFLP polymorphisms in seven British and Icelandic families. The strong evidence they reported for linkage in this region (5q11-13) (LOD scores maximizing over 6 with a broad phenotypic definition), could not, however, be replicated by many other groups nor by the original investigators themselves.

Current State of the Field Space limitations preclude an exhaustive review of linkage studies of schizophrenia here. Instead selective current developments are reviewed. Several complete or nearly complete genome scans for schizophrenia (in which markers are placed at 10- to 20-cM intervals over the entire human genome) have now been completed or published on relatively small high-density family sets. None of these scans have revealed evidence for a large single major locus for schizophrenia. Consistent with the evidence just reviewed, these results suggest that the existence of a single susceptibility locus that accounts for a large majority of the genetic variance for schizophrenia is unlikely.

Within the last 3 years we have seen the first tentative evidence for replicated linkages for schizophrenia susceptibility loci. To date, three regions appear most promising: 22q12-13, 6p24-22, and 8p22-21. It should be noted, however, that the interpretation of these results is quite controversial, particularly as the definition of *replication* for linkage to a complex trait remains uncertain.

22q12-13.1 In 1994 Ann Pulver and colleagues reported that, in a preliminary genome scan of 240 randomly distributed markers in 39 multiplex schizophrenia families ascertained through the Maryland Epidemiology Sample, the strongest evidence for linkage was found for three markers spanning ~23 cM in the 22q13.1 region: D22S268, IL2RB, and D22S307. The maximum homogeneity LOD score (which assumes that all families were segregating a susceptibility gene for schizophrenia in this region) using their initial genetic model was modest: 1.54 for IL2RB. By altering their genetic model across a range of parameters, they were able to maximize the LOD score, again at IL2RB at 2.82. No evidence for linkage heterogeneity was found.

Pulver organized an attempted replication study of this finding with three other research groups from Virginia, the United Kingdom, and France that examined these three markers on 22q in a total of 217 multiplex schizophrenia pedigrees. Overall, the evidence for linkage was not confirmed by the new samples, although some subsequent analyses by the Virginia group produced weakly positive findings.

Quickly following these reports, a number of attempted replications by other individual groups were published. Two of these reports were positive, including a maximum LOD of 2.09 in nine Mormon pedigrees to marker D22S276 (~4 cM telomeric of IL2RB) and a maximum LOD of 1.51 at marker D22S278 (~7 cM centromeric of IL2RB) in 23 high-density British and Welsh pedigrees. Three other reports were largely negative: in 23 British and Icelandic pedigrees, 105 multiplex American families, and 30 German and Israeli families. In addition, Pulver and her colleagues followed up with a report when their family number had increased to 57. They found the 22q region remained most significant in their genome scan, although their most positive marker shifted from IL2RB to D22S278.

These reports were followed by a large collaborative data-pooling effort led by Michael Gill. Eleven groups contributed with the typing of a single marker, D22S278, which they analyzed by the sib-pair method. Using all available data (620 sib pairs), the excess sharing of alleles in affected sib pairs was modest, but significantly deviated from chance expectation ($p = 0.006$). If they used only the sib pairs where both parents were also typed (296 pairs), the evidence in favor of linkage was somewhat stronger ($p = 0.001$). Although 25 percent of random siblings would be expected to share both alleles at a locus identical by descent, the investigators found that 31 to 32 percent of the sib pairs concordant for schizophrenia shared their alleles at the D22S278 marker. The authors calculated that if this locus truly impacts on the liability to schizophrenia, it is likely of rather small effect, being responsible for ~2 percent of the variance in liability.

Another line of research has impacted on the interest in this region on 22q. Microdeletions in this region are responsible for a range of congenital malformation syndromes, including DiGeorge syndrome and velo-cardio-facial syndrome (VCFS). Individuals with VCFS demonstrate excess rates of psychopathology, although some debate remains as to whether these syndromes more closely resemble schizophrenia or bipolar illness. M. Karayiorgou et al found that 2 of 100 schizophrenic patients had microdeletions in this region, a finding that is in excess of the estimated general population frequency of such deletions of 1 of 4000.

6p24-22 The first evidence for linkage for schizophrenia to the 6p region came from the early phases of a genome scan in the 265 small to medium-sized families constituting the Irish Study of High-Density Schizophrenia Families. In the initial publication of this finding in 1995, 16 markers were examined over a 38-cM interval. When a narrow diagnostic model was used, the evidence for linkage was modest (max LOD = 1.18), but increased substantially as the diagnostic definition broadened to include other disorders within the schizophrenia spectrum (max LOD = 3.51 at D6S296). Evidence for linkage fell when the definition was broadened further to include nonspectrum disorders such as anxiety conditions and alcoholism. Using multipoint linkage analysis and a larger number of markers, the relationship between diagnostic breadth and evidence for linkage in this sample is illustrated in Figure 12.5-1.

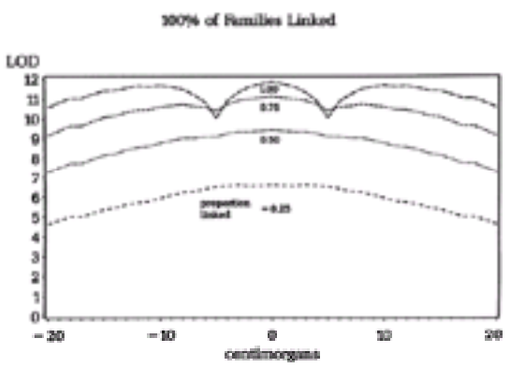

FIGURE 12.5-1 Linkage with homogeneity. LOD scores were obtained from computer analysis of 25 simulated nuclear families (described in the text) in which a disease gene is located at the 0.0 centimorgan map position in all families (genetic homogeneity) and flanked on both sides by genetic marker loci located about 5 centimorgans on either side. Support for four alternative tests of linkage are shown. The hypotheses differ in the hypothesized proportion of families in the collection of families that might contain a major gene conferring susceptibility to schizophrenia. Those hypotheses of linkage with varying levels of genetic heterogeneity are contrasted with the null hypothesis that none of the families contains such a gene (proportion linked = 0.0). When all families are actually linked to the candidate region, as in this example, all four hypotheses show strong support for linkage. However, because LOD scores are based on a $\log_{10}$ scale, the LOD score of 12.0 in support of the true hypothesis (proportion linked = 1.00) provides over one million times (10_6) greater odds of linkage than the hypothesis that receives the lowest support (proportion linked = 0.25), which has an LOD score of 6.0.

Strong evidence for linkage heterogeneity was found—perhaps because of the large sample size. Using several methods, the proportion of families carrying a pathogenic mutation at this putative vulnerability locus was estimated at 15 to 30 percent.

Eleven other published reports on this region have appeared from individual groups. Seven of these studies provide additional evidence of variable strength for a susceptibility locus for schizophrenia in this region; four do not. Briefly, the positive results include: (1) in 43 German and Israeli pedigrees, a maximum LOD of 2.2 near marker D6S274 (~15 cM centromeric of D6S296); (2) in a collaborative sample of 5 large pedigrees and 65 multiplex families, a nonparametric p value for linkage of 0.005 at marker D6S274; (3) in 57 pedigrees from Maryland, a maximum LOD of 1.17 at D6S296 and a sib-pair analysis at the same marker at $p = 0.004$; (4) in 10 German families, modest evidence for linkage (max LODs ranging from 1.41 to 2.08 over a large interval from markers D6S274 to D6S459), which increased substantially when eye-tracking dysfunction was examined (max LOD of 3.51 at D6S271—over 30 cM centromeric to D6S296); (5) an LOD of 2.49 in one mixed French-Canadian family containing cases of both schizophrenia and mood disorder; (6) initially negative results using 5 markers in 19 Southern African Bantu-speaking families which produced mildly positive findings on reanalysis with other statistical methods; and (7) strong evidence for linkage to a quantitative trait reflecting positive symptoms of schizophrenia in 10 moderately large Canadian families of Celtic origin. The negative results include an examination of 4 markers in the region in 45 pedigrees from the United States and Australia, 2 markers in 23 British and Icelandic pedigrees, 17 markers in 211 U.S. families, and 9 markers in 86 Welsh and English families.

In 1996, results of a 14-group collaborative effort on linkage in the 6p region were published. Most of the groups that had previously reported their findings individually

participated in this collaborative effort. Typing 5 markers on 6p on over 400 new families (excluding the Irish sample in which the linkage was first found) and using a narrow diagnostic definition of schizophrenia, a maximum LOD score of 2.19 (which equals a *p* value of ~0.001) was found. It is noteworthy that these modest results for linkage, which were interpreted as suggestive but not definitive, were relatively widely dispersed amongst the groups. For one particular model, for example, modestly positive results (LOD >0.3) were seen in the 6p region by 9 of the 13 new samples examined.

8p22-21 Pulver et al were also the first to report evidence for linkage of schizophrenia to the 8p region. In the 57 high-density schizophrenia families that she and her colleagues ascertained in Maryland, 18 markers were typed over an 81 cM region. Using a relatively narrow definition of schizophrenia, they found evidence for linkage over a ~15 cM interval from markers D8S258 to D8S136. The maximum LOD score was 2.35 at D8S136. Using an analysis based on affected sib-pairs, maximum evidence for linkage (at $p = 0.00004$) was found at D8S258.

Three other reports from individual groups have examined linkage for schizophrenia in this genomic region. A genome scan of five large Icelandic pedigrees found modest evidence for linkage ($p = 0.04$) at marker D8S298 (which maps between D8S258 to D8S136). In 25 Welsh and English families, however, examining 4 markers in this region Kunugi et al could find no substantial evidence for linkage. By contrast, results from the Irish Study of High-Density Schizophrenia Families were quite positive. Examining 15 markers in this region in 265 multiplex families, the strongest evidence for linkage was found for marker D8S1715 (which maps 1-2 cM telomeric of D8S258—maximum LOD = 2.52). Multipoint analysis—which produced the strongest evidence for linkage using a broad definition of the schizophrenia spectrum—demonstrated a 10cM region with LOD scores above 1.90, from markers D8S1715 to D8S1739 (which is ~4 cM centromeric of D8S136). Again, given such a large number of families, strong evidence was found for linkage heterogeneity, with results suggesting that a susceptibility locus in this region was present in only around 10 to 25 percent of families.

The fourteen-group collaborative effort also examined 5 markers in the 8p22-21 region. The maximum LOD score obtained (excluding the families of Pulver et al in which the initial linkage was first found) was 2.22 ($p = 0.001$), which again is suggestive but not definitive. Of the 13 groups that reported new results from this region, positive evidence for linkage (defined as a maximum LOD >0.30) was found in 10.

22q, 6p, and 8p Table 12.5-6 summarizes the current status of linkage findings for these three regions. Results in each of these areas share three features: (1) they have been replicated in at least two other samples in addition to the one in which they were first reported, (2) they have not been replicated in all studies, and (3) a collaborative effort has pooled results across many groups and produced suggestive results.

Replications	22q12–13.1	6p24–22	8p22–21
Initial finding	Pulver et al, 1994 Max LOD = 2.82	Straub et al, 1995 Max LOD = 3.51	Pulver et al, 1995 Max LOD = 2.35
Number of replication attempts:			
Single groups	5	8	3
Number positive	2	4	2
Multigroup	2	1	1
	3 groups, largely negative; 11 groups, 1 marker	14 groups, 5 markers	14 groups, 5 markers
	$p = .006$	$p = .001$	$p = .001$

* April 1997.

Table 12.5-6 Current* Potentially Replicated Linkages in Schizophrenia

Other Linkage Regions of Interest: Dopamine Receptors, X Chromosome, and 5q22-31 Many other linkage reports for schizophrenia have been published since 1990. The largest single collection of such reports has examined the five types of dopamine receptors (D_1 through D_5), with at least ten reports examining the 11q22-23 region containing the D_2 receptor, five reports each examining the 3q13.3 region containing the D_3 receptor and the 11p15 region containing the D_4 receptor, and at least two reports examining the 5q35.1 region containing the D_1 receptor, the 4p16 region containing the D_5 receptor, and the 5p15.3 region containing the dopamine transporter gene. These studies have been consistently negative. Genetic variation in or around the genes coding for the dopamine receptors and transporter do not appear to play a major role in the etiology of schizophrenia. Other negative linkage reports include the $GABA_A$ receptor subunit genes; the glutamate GluR5, GluR6, and NMDAR1 subunit genes; and the $5\text{-}HT_2$ receptor.

Based in part on the observation of an excess of same-sex affected sibling pairs, particularly in families where the illness appears to be paternally transmitted, several investigations have been conducted for linkage of schizophrenia to markers in the pseudo-autosomal region—with mixed results.

Finally, linkage in the 5q21-31 region was recently independently found by two groups. In the genome scan of the Irish study, marker D5S818 in 5q22 produced the second best of the first 223 markers tested. Thirteen additional markers were typed over ~45 cM. The largest two-point LOD score found was with marker D5S393 (LOD = 3.04) and assumed a narrow phenotypic definition (schizophrenia plus poor-outcome schizoaffective disorder.) Multipoint analysis produced substantial evidence for linkage (LOD > 2) over ~20 cM, from markers D5S818 to IL-9, with the peak at D5S804 (LOD = 3.35). Evidence for heterogeneity was found, with estimates that between 10 to 25 percent of the families were segregating a mutation at this locus. At the same time, Schwab and colleagues in Germany, in a study on 44 nuclear families found a maximum multipoint LOD score calculated from affected sibling pairs of 1.8 around marker D5S399, which is only ~3 cM distal to IL-9.

Association Studies of Schizophrenia Association studies examine whether individuals affected by a disease more frequently have a particular allele at some "candidate" genetic locus than individuals not affected by the disease. This association can occur for two reasons. Either the allele being studied directly influences risk for the disorder or, more commonly, the allele is in linkage disequilibrium (LD) with the disease-predisposing mutation. In addition, association can occur for spurious reasons unrelated to disease etiology, such as population stratification. While linkage—the co-segregation of a marker and a disease—is a family-based phenomenon, linkage disequilibrium is a population phenomenon and relies on the specific population history of the marker and disease mutation. It will be seen when a marker is close enough to the disease gene so that the original association created when either the disease or marker gene mutated or entered into a population through a population bottle-neck has not had time to be broken apart by genetic recombination. Association studies have two important advantages when compared to linkage studies. First, association studies can study individual patients. Second, under many circumstances, association studies are considerably more powerful than linkage studies at detecting genes of modest effect. However, association studies have two disadvantages. First, they "scan" much smaller regions of the genome that do linkage studies. The practical effect of this is that association studies must be used for candidates genes only. Currently a genome scan using the association method is not technically feasible. Also, obtaining proper controls for association studies can be difficult and if done improperly can lead to false-positives results. Recently, a new approach to association studies has been adopted that solves this problem. By studying individual patients and their parents, researchers can use as a "control group" the parental genes that did not get transmitted to the patient.

Certainly the longest history of LD studies in schizophrenia have involved the HLA region on chromosome 6p. The earlier literature suggested the possibility of replicated positive associations with HLA A9 and B5 and negative associations with BW35. Furthermore, the positive association with the A9 allele was particularly noted in cases of paranoid schizophrenia. More recent reports using molecular genetic genotyping methods have focused on possible associations of schizophrenia with DQB1 and DRB1 HLA loci.

As polymorphic markers became available near or within neurotransmitter receptors, many reports have examined the association between schizophrenia and serotonin and especially dopamine receptors. Sometimes these studies employed polymorphisms within the gene, such as the Ser_{311}/Cys_{311} structural polymorphism within the D2 receptor gene, a glycine-to-serine missense polymorphism at position 9 in the N-terminal extracellular domain of the D3 receptor and the 48 base-pair sequence repeat in the D4 receptor. Few replicated findings have emerged to date; however, there is a suggestion in some but not all studies of excess homozygosity at the D3 receptor gene in patients with schizophrenia. A European multicenter collaborative group recently reported evidence for a modest association (OR ~ 1.2) between schizophrenia and the T102C polymorphism in the serotonin type 2a receptor. Whereas association studies remain of major interest in studies attempting to clarify the nature of the genetic liability to schizophrenia, it is probably fair to conclude that a powerful, widely replicated finding has yet to emerge with this technique.

Anticipation and Expanded Trinucleotide Repeats In the last decade a new mutational class of expandable trinucleotide repeat (TNR) sequences have been shown to be etiologic in a number of human genetic disorders, including several—such as fragile X and Huntington's disease—which are predominantly

neuropsychiatric in their manifestations. A feature of these disorders is *anticipation*—that the age at onset of the illness decreases and its severity increases across generations. In most of these disorders, anticipation has been clearly related to expansions of the TNR.

Anticipation in neuropsychiatric disorders has been observed and commented upon since the early 1800s, when it formed one of the foundations of the *degeneration theory* of psychiatric illness popular in nineteenth-century France. However, Mott, a student of Pearson, dismissed anticipation as a methodologic artifact early in this century. Motivated by the possibility that schizophrenia might result from TNR expansions, a number of recent studies have examined whether the age at onset of schizophrenia demonstrates anticipation—that is, decreases across generations.

Most studies that have examined this question have found evidence in support of anticipation; the interpretation of these results has been more difficult. A number of potential artifacts could explain earlier ages at onset in children than in their parents, aunts, or uncles; several of these artifacts are very difficult to control for. While opinions differ in this field, there is still cause for skepticism on whether "true" anticipation has been conclusively demonstrated for schizophrenia. This is perhaps an increasingly moot point as investigators are now trying to directly detect expanded TNR in patients with schizophrenia.

A number of studies have attempted to directly assess expanded repeats using Repeat Expansion Detection (RED). Several groups have produced evidence for expanded CAG/CTG repeats using the RED method in schizophrenic versus control subjects. However, the significance of these findings remains uncertain, for at least two reasons. First, no correlation was observed in one of the best studied samples between the presence of expanded repeats and either clinical features of illness or age at onset. Second, the RED technique itself is entirely nonspecific and the significance of the detection of repeat expansions, which may occur anywhere in the entire human genome, remains unclear.

FUTURE DIRECTIONS

The evidence is strong that schizophrenia is a familial disorder and that the familial aggregation of schizophrenia is due largely, although probably not entirely, to genetic factors. Whatever the familial predisposition that operates for schizophrenia, it not only "codes" for the classic, psychotic disorder but also increases liability to "schizophrenia-like" personality disorders and probably for some other nonschizophrenic nonaffective psychoses. Two decades of research using statistical methods have failed to clearly delineate the mode of transmission of schizophrenia, a result that is understandable given its complexity.

Advances in molecular biology have opened up realistic opportunities to localize on the human genome the specific genes that influence the liability to schizophrenia. Association studies have yet to provide convincing evidence for the role of a range of candidate genes in the etiology of schizophrenia. Genome scan strategies have, however, provided at least three regions, on chromosomes 6, 8, and 22 where at least three groups have found evidence for linkage. Although false-positive findings cannot be ruled out, it is likely that one or more of these regions do contain one or more susceptibility genes for schizophrenia. The hope is that more comprehensive mapping studies, coupled with advances in sequencing and positional cloning technology, will in the next few years produce the identification of one or more specific gene defects that predispose to the development of schizophrenia.

Whereas gene identification will itself represent a major advance, it will also represent the beginning of several new lines of research including (1) rational drug design based on knowledge of basic pathophysiology, (2) characterization of genotype-phenotype relationships based on knowledge of specific pathogenic mutations, (3) identification of environmental risk factors that interact with specific genes, and (4) realistic prevention research based on our ability to identify high-risk individuals.

SUGGESTED CROSS-REFERENCES

Population genetics in psychiatry is discussed in Section 1.17, and genetic linkage analysis of the psychiatric disorder is discussed in Section 1.18.

SECTION REFERENCES

Brzustowicz LM, Honer WG, Chow EWC, Hogan J, Hodgkinson K, Bassett AS: Use of a quantitative trait to map a locus associated with severity of positive symptoms in familial schizophrenia to chromosome 6p. Am J Hum Genet *61:*1388, 1997.

Cannon TD, Kaprio J, Lonnqvist J, Huttunen M, Koskenvuo M: Genetic epidemiology of schizophrenia in a Finnish twin cohort: A population-based modeling study. Arch Gen Psychiatry *55:*67, 1998.

Chen WJ, Faraone SV, Tsuang MT: Linkage studies of schizophrenia: A simulation study of statistical power. Genet Epidemiol *9:*123, 1992.

Cloninger CR: Turning point in the design of linkage studies of schizophrenia. Am J Med Genet *54:*83, 1994.

Coon H, Jensen S, Holik J, Hoff M, Myles-Worsley M, Reimherr F, Wender P, Waldo M, Freedman R, Leppert M, Byerly W: Genomic scan for genes predisposing to schizophrenia. Am J Med Genet *54:*59, 1994.

*Gottesman II: *Schizophrenia Genesis: The Origins of Madness*. WH Freeman, New York, 1991.

Kendler KS, Diehl SR: The genetics of schizophrenia: A current, genetic-epidemiologic perspective. Schizophr Bull *19:*261, 1993.

Kendler KS, Gruenberg AM, Kinney DK: Independent diagnoses of adoptees and relatives as defined by DSM-III in the Provincial and National samples of the Danish Adoption Study of Schizophrenia. Arch Gen Psychiatry *51:*456, 1994.

*Kendler KS, McGuire M, Gruenberg AM, O'Hare A, Spellman M, Walsh D: *The Roscommon Family Study:* I. Methods, diagnosis of probands and risk of schizophrenia in relatives. Arch Gen Psychiatry *50:*527, 1993.

Kendler KS, O'Neill FA, Burke J, Murphy B, Duke F, Straub RE, Shinkwin R, NiNuallain M, MacLean CJ, Walsh D: Irish study of high-density schizophrenia families: Field methods and power to detect linkage. Am J Med Genet *67:*179, 1996.

*Kety SS, Wender P, Jacobsen B, Ingraham LJ, Jansson L, Faber B, Kinney DK: Mental illness in the biological and adoptive relatives of schizophrenic adoptees: Replication of the Copenhagen study in the rest of Denmark. Arch Gen Psychiatry *51:*442, 1994.

Kidd KK: Associations of disease with genetic markers: Dejà vu all over again. Am J Med Genet *48:*71, 1993.

Maier W, Lichtermann D, Minges J, Hallmayer J, Heun R, Benkert O, Levinson DF: Continuity and discontinuity of affective disorders and schizophrenia: Results of a controlled family study. Arch Gen Psychiatry *50:*871, 1993.

Maziade M, Bissonnette L, Rouillard E, Martinez M, Turgeon M, Charron L, Pouliot V, Boutin P, Cliche D, Dion C, Fournier JP, Garneau Y, Lavallee JC, Montgrain N, Nicole L, Pirés A, Ponton AM, Potvin A, Wallot H, Roy M-A, le Groupe IREP, Mérette C: 6p24-22 region and major psychoses in the Eastern Quebec population. Neuropsychiatr Genet *74:*311, 1997.

McGuffin P, Sargeant M, Hetti G, Tidmarsh S, Whatley S, Marchbanks RM: Exclusion of a schizophrenia susceptibility gene from the chromosome 5q11-q13 region: New data and a reanalysis of previous reports. Am J Hum Genet *47:*524, 1990.

McInnis MG: *Anticipation:* An old idea in new genes. Am J Hum Genet *59:*973, 1996.

Onstad S, Skre I, Torgersen S, Kringlen E: Twin concordance for DSM-III-R schizophrenia. Acta Psychiatr Scand *83:*395, 1991.

Pulver AE, Karayiorgou M, Wolyniec PS, Lasseter VK, Kasch L, Nestadt G, Antonarakis S, Housman D, Kazazian HH, Meyers D, Ott J, Lamacz M, Liang K-Y, Hanfelt J, Ullrich G, DeMarchi N, Ramu E, McHugh PR, Adler L, Thomas M, Carpenter WT, Manschreck T, Gordon CT, Babb R, Puck J, Childs B: Sequential strategy to identify a susceptibility gene for schizophrenia: Report of potential linkage on chromosome 22q12-q13.1: Part 1. Am J Med Genet *54:*36, 1994.

Pulver AE, Lasseter VK, Kasch L, Wolyniec P, Nestadt P, Blouin J-L, Kimberland M, Babb R, Vourlis S, Chen H, Lalioti M, Morris MA, Karayiorgou M, Ott J, Meyers D, Antonarakis SE, Housman D, Kazazian HH: Schizophrenia: A genome scan targets chromosomes 3p and 8p as potential sites of susceptibility genes. Am J Med Genet *60:*252, 1995.

Risch N, Merikangas K: The future of genetic studies of complex human diseases. Science *273:*1516, 1996.

Schizophrenia Collaborative Linkage Group (Chromosome 22): A combined analysis of D22S278 marker alleles in affected sib-pairs: Support for a susceptibility locus for schizophrenia at chromosome 22q12. Am J Med Genet *67:*40, 1996.

*Schizophrenia Linkage Collaborative Group for Chromosomes 3, 6 and 8: Additional support for schizophrenia linkage on chromosomes 6 and 8: A multicenter study. Am J Med Genet *67:*580, 1996.

Schwab SG, Eckstein GN, Hallmayer J, Lerer B, Albus M, Borrmann M, Lichtermann D, Ertl MA, Maier W, Wildenauer DB: Evidence suggestive of a locus on chromosome 5q31 contributing to susceptibility for schizophrenia in German and Israeli families by multipoint affected sib-pair linkage analysis. Molec Psychiatr *2:*156, 1997.

Sherrington R, Brynjolfsson B, Petursson H, Potter M, Dudleston K, Barraclough B, Wasmuth J, Dobbs M, Gurling H: Localization of a susceptibility locus for schizophrenia on chromosome 5. Nature *336:*164, 1988.

Straub RE, MacLean CJ, O'Neill FA, Burke J, Murphy B, Duke F, Shinkwin R, Webb BT, Zhang J, Walsh D, Kendler KS: A potential vulnerability locus for schizophrenia on chromosome 6p24-22: Evidence for genetic heterogeneity. Nature Genet *11:*287, 1995.

*Straub RE, MacLean CJ, O'Neill FA, Walsh D, Kendler KS: Support for a possible schizophrenia vulnerability locus in region 5q22-31 in Irish families. Mol Psychiatr *2:*148, 1997.

Turecki G, Rouleau GA, Joober R, Mari J, Morgan K: Schizophrenia and chromosome 6p. Am J Med Genet (Neuropsychiatr Genet) *74:*195, 1997.

Zerbin-Rüdin E: Endogene Psychosen. In *Humangenetik: ein kurzes Handbuch in fünf Bäde, vol 2*, PE Becker, editor. Thieme, Stuttgart, 1967, p 446.

12.6 SCHIZOPHRENIA: PSYCHODYNAMIC TO NEURODYNAMIC THEORIES

THOMAS H. MCGLASHAN, M.D., AND RALPH E. HOFFMAN, M.D.

Schizophrenia is unarguably one of the most devastating of the mental illnesses. It often strikes early in life, and onset can signal the presence of an irreversible vulnerability toward psychosis that creates a lifetime of anguish and burden for patients and their loved ones. Equally devastating is the profound alienation that accompanies the disorder's emergence because the symptoms can be strange, unreal, and frequently impossible to connect with empathically. The derailment from normality is profound; in no other disorder is there the sense that one is literally losing his or her mind. Yet it is hard not to conclude that such a travesty is transpiring when one's thoughts become another's voice, one's will becomes enslaved by outside forces, one's capacity to focus and think straight shatters like broken glass, one's wellspring of initiative dries up, and one's apprehension of this tragic change becomes submerged beneath waves of denial, passivity, and nonreactive apathy.

Because schizophrenia is disabling and strange, the disorder has had many explanations and has been the object of more attempts to render it meaningful than any other mental illness. Prior to Galileo, most of these explanations were found in religious texts. From Galileo to Emil Kraepelin, the explanations were found in medical texts alongside neurological disorders and idiocy. Inspired by psychoanalytic thinking, by the early twentieth century the nature and cause of schizophrenia was explained as functional. No longer supernatural or organic in etiology, schizophrenia became a clash of ideas, of wishes, of learned habits (i.e., psychological in its genesis and manifestations). From this milieu came the various psychodynamic theories of schizophrenia. By the late twentieth century, the organic substructures underlying psychological processes received greater attention, leading to models focusing at the mind-body interface, here labeled the neurodynamic theories of schizophrenia.

Neurodynamic theories assert that in addition to psychodynamics, one must look at the neurobiological structures that generate psychology to comprehend symptom formation in schizophrenia more fully. To use the computer metaphor, symptoms arise in schizophrenia because of shifts or defects in the hardware of mentation as well as from conflicts and warps in the software of psychology. This hardware consists of biologically programed neuronal networks in dynamic communication via chemical and electrical connections, hence the term *neurodynamic*.

While purely psychological theories have advanced our understanding of schizophrenia to some degree, this body of thought has not progressed much in the past 30 years. Modern thinking about the psychodynamic contributions to psychopathology has increasingly recognized the central influence of nature as well as experience. The authors feel that schizophrenia, more than any other psychiatric disorder, challenges the boundaries of what is meant by psychodynamic and argue that a broad definition that includes both organic and psychological contributions to mental phenomena is most valid, certainly for schizophrenia and possibly for many other mental disorders as well.

Historically, the 1911 publication of Sigmund Freud's case of Daniel Paul Schreber probably marks the formal beginning of the systematic psychodynamic theories of schizophrenia. For the next 50 years, virtually all thinking in this realm was connected with the various psychanalytic schools, here labeled the classical, interpersonal, and developmental. The next major body of theory emerged from psychoanalytic theories around the middle of the century as family transaction models. Shortly thereafter, following the biological revolution in psychiatry and genetic studies of schizophrenia, the neurodynamic theories emerged, introduced by the stress-diathesis, or vulnerability-stress, hypothesis. Finally, the advent of artificial intelligence and computer simulations of brain functioning gave rise to the most recent parallel distributed processing model of schizophrenic symptom formation.

THEORY

A rigorous and operationally oriented definition of theory envisions it as a set of assumptions and definitions that can generate testable and refutable hypotheses or predictions about a phenomenon. This form of theory constitutes the backbone of modern scientific empiricism. It is not, however, the form of theory that characterized the psychodynamic theories of schizophrenia. The latter are, according to Joseph Lichtenberg, an aesthetically oriented set of assertions explaining something in a manner that is “balanced, logical, and comprehensive while at the same time parsimonious in its assumptions.” Few, if any, of the psychodynamic theories considered here could be operationalized and tested. They are too abstract or based on data collected empathically rather than objectively. None of them can be validated on the basis of empirical data per se. While these theories in general must conform to rules of evidence, their veracity derives from their capacity to help one understand schizophrenia. These theories are compelling in proportion to the degree to which they generate meanings about schizophrenia that make communicable sense and that are useful in one's empathic encounters with afflicted patients. The validity here is face validity and stems from a theory's vividness, connectedness, and depth as well as its parsimonious integration of complexity. Such theory is also useful in alerting doctor and patient as listeners and hunters for what is missing from the latter's experience.

Psychodynamic theory is not, however, useful in posing a series of testable hypotheses that can build a path of knowledge toward a specific causality of schizophrenia. The psychodynamic theories reviewed here ultimately failed at this task and have receded in importance as a result. They remain an important contribution to the total picture of the patient, however, insofar as they provide us with usefully parsimonious metalevel descriptions of both normal and schizophrenic mental phenomena and mechanisms.

CLASSICAL PSYCHOANALYTIC MODELS

Sigmund Freud The classical psychoanalytic model postulates that manifest psychopathology is generated by active and sustained psychological conflict between drive-created wishful impulses and antithetical wishes, reality, or conscience. This conflict generates defenses against the wishful impulse, and these defenses can often be seen in the form of symptoms. Any or all of this drama may be carried on outside of awareness (i.e., unconsciously).

This model finds its most complete elaboration in the structural theory, which postulates the existence of three functional entities in the mind. The *id* is the wellspring of peremptory sexual and aggressive drives and wishes. It is largely unconscious and primitive in structure. The superego, or conscience and ego ideal, is the repository of rules and values learned (internalized) from parents and society during development. It is also largely unconscious but makes its presence known through the affects of guilt and shame. The *ego* is a group of psychological functions that mediate adaptation between the person and the environment (e.g., reality testing) and among conflicting psychological forces within the person (e.g., repression of forbidden impulses). The ego is complex and develops slowly over the course of life. Many of its functions (e.g., defense) are activated by anxiety, the danger signal generated by conflict among psychological forces or with reality. Ego functions, too, operate mostly out of awareness.

Freud postulated that these structures develop during infancy and childhood and are in place by the end of the oedipal period (ages 3 to 5 years). The person at this point has a stable, integrated ego, seen as a sense of self that is enduring and cohesive. Conflict within and among these structures produces the symptomatic and character neuroses. Freud regarded schizophrenia as deriving from psychological development arrested prior to the oedipal stage, prior to development of an integrated ego. According to Freud, this arrest severely compromised the schizophrenic patient's capacity to relate and rendered psychoanalytic treatment problematic, if not impossible.

Although he had virtually no clinical experience with schizophrenic patients, Freud was the first analyst to elaborate a systematic psychodynamic model for this

syndrome. He formulated two models—one emphasizing conflict and defense and the other emphasizing deficiency—as the cause of schizophrenic symptoms.

The conflict-defense model basically explains schizophrenic symptoms by use of the structural model. In this model schizophrenia, like all psychopathology, is the result of conflict and defense. The difference between schizophrenia and the neuroses is purely quantitative, not qualitative. Schizophrenic conflict is more intense and requires frequent use of primitive (i.e., developmentally early) defenses such as denial and projection, which frequently involve a break with reality. The ego functioning of the schizophrenic patient regresses to developmentally earlier stages or levels of organization, the exact level being determined (or fixated) by one or more past psychological traumas. The difference between schizophrenia and neurosis lies in the depth of regression and the point of fixation, which Freud placed in the preoedipal phase of development.

Freud used a deficiency (or deficit) model to explain schizophrenic symptom formation in the case of Schreber. Freud worked from Schreber's published autobiographical account of his paranoid psychosis. Clinically, Schreber's illness began with hypochondriacal preoccupations. These were followed by an apocalyptic panic leading to catatonia, personality change, and symptoms of psychosis, particularly, grandiose and paranoid delusions.

Freud, who was elaborating his libido theory at the time, explained Schreber's psychosis as follows. Conflict initiates the sequence, as it does in all psychopathology. In schizophrenia, however, another process supersedes defense. This process is described as the patient's withdrawal of libidinal or energic investments (cathexes) from the real outside world, especially people (objects, in psychoanalytic parlance). There is a concomitant withdrawal of libidinal investments from the inner, fantasied, mental representations or images of this world and these people. In the developing schizophrenic process, the withdrawn libidinal energy increasingly becomes invested in the patient's self-image, seen clinically as self-aggrandizement or megalomania, or invested in the patient's body image, seen clinically as hypochondriasis. In neurosis, a similar process of withdrawal from real, external relations occurs in response to conflict, but here, the withdrawn libido remains invested in the fantasied objects.

This withdrawal reaches a state so profound as to constitute a break with external reality and relationships and with internal object representations and relationships in fantasy. At this stage, one can see the apocalyptic panic clinically. It represents a projection outward of this internal catastrophe or collapse of psychological investments. This collapse and profound withdrawal constitutes the deficit of schizophrenia. It renders the patient incapable of relationships, including transference, and thus precludes treatment by psychoanalysis. Following this catastrophe, the patient tries to recover and reinvest libido. Since there has been a break with reality, however, these efforts produce the well-known symptoms of schizophrenia, especially hallucinations, delusions, and disordered thinking. The patient has reinvested interest and attention but in objects that are not part of the real world.

In Freud's defense theory, the sequence of conflict, anxiety, and defense is regarded as sufficient to account for schizophrenic psychopathology. In the deficit theory, conflict and anxiety initiate pathogenesis but trigger a withdrawal process that is qualitatively different from defense. Freud never resolved the difference between his two theories. He seemed to say that schizophrenic people are very much like normal and neurotic people in some ways but profoundly different in others. His two theories formed the nidus for much subsequent controversy.

Freud's other theoretical contributions to schizophrenia concern the psychodynamics of delusion formation. Early in his career, he postulated the mechanism of projection, whereby the subject's wish is disavowed and projected onto (or attributed to) another person (the object). Later he suggested that delusions of persecution arise from latent homosexual impulses that undergo reversal and projection. Thus, the situation “I (a man) love him (a man)” is reversed to “I do not love him; I hate him” and projected into “He hates (and persecutes) me.” Later in his career, Freud maintained that the hostility inherent in any form of intense ambivalence toward an object could be projected into feelings of being persecuted by that object.

Paul Federn If Freud was the first major psychodynamic theoretician of schizophrenia, Paul Federn was the first major psychodynamic clinician of schizophrenia. A contemporary of Freud's, he disagreed with his Viennese colleague's pessimism about the schizophrenic patient's capacity to develop transference and to be treated by psychoanalytically informed therapy. He treated many schizophrenic patients and developed techniques that were half a century ahead of his time.

Federn greatly expanded the notion of ego set down by Freud in the structural theory. He was perhaps the first psychoanalytic theoretician to introduce the notion of self. To him, the ego was not just a collection of psychological functions. It also had its own existential being or ego feeling. The various ego functions aggregate into a sum or self that has a feeling of permanence and continuity vis-a-vis time, space, and causality. This is ego feeling, the totality of feeling an individual has of his or her own living being. Ego feeling as subject is “I.” Ego feeling as object is “self.”

Federn also elaborated the concept of ego boundary originally introduced by Viktor Tausk. To Federn, each person possessed an inner and an outer ego boundary. The outer boundary consists of the ego versus the external world; it divides and distinguishes mental phenomena from real phenomena. The inner boundary consists of the repression barrier or the line between conscious and unconscious experience. According to his scheme, which utilized Freud's libido theory, schizophrenia is a disease of the ego. The psychopathological process involves a loss of investments in ego boundaries. Attenuation at the inner boundary means derepression or a reemergence of developmentally earlier (archaic) ego states. Attenuation at the outer boundary means a loss of the distinction between mental and real, seen in the typical schizophrenic symptoms. Mature and archaic ego states, however, can coexist, thus making it possible for the patient to adjust to the real world while still symptomatic and to engage in psychodynamically oriented therapeutic discourse despite illness.

Federn's basically descriptive model follows the theme of defect. Nevertheless, his concepts prefigure the later distinction between psychotic and nonpsychotic aspects of the patient's personality. For him, the schizophrenic process is never total. Furthermore, by highlighting the self phenomenology in the ego of Freud's structural model, Federn anticipated the development of self psychology.

Heinz Hartmann Working within Freud's classical structural theory, Heinz Hartmann was impressed with the ego's complexity and versatility, its strength in opposition to the drives, and its primary aim of serving reality adaptation and survival. Defenses such as intellectualization and sublimation, could also be coping devices. He regarded humans as biological organisms phylogenetically equipped at birth for adaptation to an average expectable environment. This includes primary ego functions like perception, memory, and motility, which are not derived from conflict. Also, ego functions developing later out of conflicts can become autonomous of id and superego or free of conflict to function independently and to serve adaptation. Such functions include language, intellect, thinking, will, judgment, attention, affectivity, reality testing, intention, and object relations in addition to the defenses and primary functions already mentioned. The existence of psychopathology indicates that these ego functions can become reinstinctualized or involved in conflict situations.

The ego, according to Hartmann, also possesses a synthetic function, its aim being to promote homeostasis, or a harmonious equilibrium between the drives of sex and aggression, among the intrapsychic tripartite systems of id, ego, and superego, and between the individual and his or her environment. This supraordinate integrative function carries echoes of Federn's ego feeling. Hartmann, however, regarded the self as an idea (representation) rather than an entity or functioning mental system.

Compared with his extensive contributions to general psychodynamic theory, Hartmann's specific postulates regarding schizophrenia are abbreviated, perhaps because such patients did not constitute a large part of his practice. His theory was a mixture of defense and defect. Like Freud, Hartmann felt that schizophrenic symptoms can result from conflicts secondary to intolerable realities or amplified drive pressures. In addition, he postulated an inborn primary defect in the ego of the preschizophrenic patient that renders the ego incapable of neutralizing certain drive pressures, especially aggression. Aggression generated later in life by conflict or narcissistic injury floods this ego (especially its synthetic functioning) and draws it easily into conflict. The ensuing regression is substantial and mobilizes primitive defenses such as denial and projection, which are viewed as the symptoms of schizophrenia. Not all ego functions regress to the same extent, however, thus accounting for the heterogeneity of the clinical picture.

Hartmann's theory added the importance of aggression to the pathogenesis of schizophrenia. It also placed the source of the syndrome in the preschizophrenic patient's constitution, thus marking such people as qualitatively different from those who later develop normally or neurotically. Finally, Hartmann conceptualized the ego as multidimensional. Schizophrenia affects ego functioning selectively and can therefore be graded in severity. In Hartmann's scheme, it is possible to have greater or lesser degrees of schizophrenia. Overall, he emphasized the biological underpinnings of many psychological functions and thus presaged later neurodynamic theories of schizophrenia.

INTERPERSONAL MODELS

Harry Stack Sullivan While psychodynamic in structure, the interpersonal model of Harry Stack Sullivan, differs fundamentally from psychoanalytic drive theory in content. Drive theory works from the perspective of the person as an individual encountering and shaping the world according to inner arising drives and satisfactions.

Interpersonal theory elaborates the perspective of the individual as a social creature who, from the very beginning, is object related and relationship seeking. Sullivan's model still postulates motivational drives and needs—namely, the need for satisfaction, (mostly biological, in the form of hunger and lust), and the need for security, (mostly psychosocial, in the form of power). All of these needs require interaction with at least one other human being and serve to mediate the interpersonal exchange.

The developmental aspect of Sullivan's theory regards the human infant as being without a psychology separate from the initial mother-infant dyad. Psychological awareness consists of successive discoveries of one's self in relationship with significant others (objects). The first self consists of a we, not an I. Development proceeds according to an increasingly complex hierarchy of needs, all interactional. These are the needs for maternal contact in infancy, parental mirroring in childhood, peer play in latency, chum closeness in early adolescence, and sexual intimacy in late adolescence and beyond.

Anxiety, the affect driving psychopathology, was viewed by Sullivan as external to the infant but imparted to the latter by an anxious parent, usually mother. Anxiety in the interpersonal situation develops three self-states: a good me (low anxiety), a bad me (high anxiety), and a not me (intolerable anxiety). Not-me anxiety is extreme awe, dread, loathing, or panic—so dysphoric as to be experienced rarely, as in nightmares or during severe schizophrenic end-of-the-world panic experiences.

Anxiety leads to the organization of defensive structures, which Sullivan described as self-dynamism or self systems. These function to maximize satisfactions and to maintain security or to minimize anxiety through the use of “security operations” such as selective inattention (dissociation), sublimation, or projection. The self system, in content, is what one takes oneself to be. This is largely secondary to what others take one to be (i.e., it consists mostly of reflected appraisals). The self system security operations establish and protect this content of the self system. In the face of anxiety, this leads to the creation of fantasied defensive self-other constellations such as the self as helpless but deserving and the other as magical and merciful, the self as victimized and hurt and the other as powerful and persecutory, and the self as special and the other as idealizing. Such illusory configurations become superimposed upon and distort a person's here-and-now relationships, a process akin to transference, which Sullivan labeled parataxic distortion.

Sullivan's psychodynamic theory of schizophrenia was informed by extensive clinical experience with acutely and subacutely affected inpatients. These disorders he regarded in the Meyerian tradition as purely functional reactions to encounters between the person and the environment. Central to the psychopathological process is a disturbance in the capacity to relate to others that is not biological in origin but reflects the history of the patient's interactions with significant others, especially with the mother in the formative years. The syndrome itself represents a massive dissociation caused by intense anxiety generated by low self-esteem during interpersonal experience. Sullivan acknowledged the probable existence of hereditary or organic determinants in some disorders, such as chronic process schizophrenia. He did not, however, consider them to be schizophrenia per se, or at least the subtype of schizophrenia to which he felt his theory applied.

The pathogenesis of schizophrenia, according to Sullivan's etiological scheme, begins with a mother who is more anxious than normal and who imparts this tension to her child as excessive not-me experiences. The child's self system, developing around the time of speech acquisition, overcompensates with excessive dissociation and warps its own further development. The adolescent surge of new sexual needs (lust dynamism) assaults this compromised self system. The defensive wall of selective inattention fractures; not-me disorganizing anxiety returns, and panic ensures. This state of terror is characterized not only by the uncanny eruption into awareness of developmentally primitive states of mind, but also by collapse of the integrated self systems into what Sullivan described as “an exceedingly unpleasant form of nothingness.” The afflicted person's primary urgencies at this point are to avoid the not-me menace and to reorganize the self to reestablish meaning and become human again. This reorganization, known as schizophrenia, is effected at the price of reality testing.

According to Sullivan, schizophrenia is more than a disorder. It is also an adaptive strategy for avoiding fragmentation and chaos (panic and terror) and for reconstructing a self with human identity, meaning, and purpose, no matter how fantastic that defensive self-other constellation may be. It is better, for example, to be the helpless victim of tyrannical persecutors than to be nothing at all. One must have character, even if it manifests as caricature. With schizophrenia, the needs for satisfaction and reality are secondary to the needs for security and self meaning.

The descriptive aspects of Sullivan's theory highlight the self, both as a content (idea) and as a functional system. Though Federn may have been the first to describe the self as part of a psychodynamic system (as ego feeling), Sullivan was the first to postulate its functional centrality to human psychology. To him, creating and maintaining the integrity and functional alacrity of a self are primary motivating forces of human beings. With schizophrenia, in fact, the drive for meaning exerts hegemony over all other needs. Sullivan viewed schizophrenia as the result of cumulative experiential traumas during development. His own bias was to regard the preschizophrenic infant as a tabula rasa on which mother's anxieties became etched. The source of pathogenic anxiety is clearly external to the infant, and schizophrenia is seen as an adaptive attempt to cope with that dysphoric milieu.

OBJECT-RELATIONS MODELS

The British object-relations school operated independently of Sullivan but pursued many of the same ideas, regarding humans as inherently social or object related. Their major spokespersons are Melanie Klein and W. R. D. Fairbairn. Latter-day American theorists contributing to a psychodynamic understanding drew heavily on theories of human development. Their major proponents are Margaret Mahler and Ping-Nie Pao.

Melanie Klein For Klein, psychodynamic conflict involved love versus hate in relationships (rather than the tension between wish and reality of classical psychoanalysis). She emphasized the importance of fantasy, both conscious and unconscious, in determining behavior. Such fantasy usually takes the form of a drama involving the self relating with another, constellations that have come to be labeled *internal object relationships*. She added two coping or defense mechanisms to the ego's repertoire: splitting and projective identification. During infancy, these mechanisms promote development and adaptation; during adulthood, they signal trouble. Klein related psychopathology to an overabundance of aggression and hate in relationships. Envy (mostly innate) is especially pathogenic, because it is directed at good objects and their capacity to give, thus destroying hope by devaluing healthy relationships.

Klein conceived of human development as a hierarchy of relational patterns (i.e., positions rather than phases). Two positions, both within the first year of life, are central to normal development—or to later psychopathology: the paranoid position, wherein aggressive, dysphoric interpersonal experiences are split off and projected onto significant others who are then regarded anxiously as persecutory and the depressive position, the infant's guilty recognition of personal responsibility for being the aggressive persecutor at times. While the accuracy of this scheme may be questionable vis-a-vis contemporary infant observation, it is compelling in its description of two mental constellations frequently encountered in patients with severe psychopathology.

Klein's theory of schizophrenia closely followed her developmental scheme. She regarded the potential schizophrenic patient as endowed with strong sadistic and envious impulses that rendered the infant prone to intense paranoid anxieties and, therefore, to the overuse of withdrawal, splitting, and projective identification. Such infants never negotiate the depressed position and remain fixated at the paranoid position, to which they regress in the face of later stress after further development through adolescence.

W.R.D. Fairbairn To Fairbairn, the primary aim of human behavior was contact with another, even if it was unpleasant. He viewed psychopathology entirely from a developmental perspective as the product of failure to establish good object relationships in infancy. Maternal absence or withdrawal during the paranoid-schizoid position leads infants to regard their love as noxious or bad. The resultant schizoid conflict—to love or not to love—sets off a withdrawal from relatedness in reality with compensatory investments in defensive internal object relations. These, like Sullivan's fantasied self-other constellations, provide a sense of security and continuity that is missing in real relationships, especially the earliest ones with parents. Fairbairn conceived of schizophrenia on a continuum with schizoid psychopathology, the difference being one of degree. The schizophrenic patient withdraws loving investments to such an extent that emotional contact with others and with external reality is renounced.

Margaret Mahler Mahler clearly related early developmental experiences to later mental function. Her developmental phases of autism, symbiosis, and separation-individuation captured the attention and imagination of many theorists who saw different forms of psychopathology corresponding to different levels in her developmental progression. Schizophrenia, for example, is regarded as corresponding to Mahler's autistic phase of development. It is assumed or postulated that the preschizophrenic infant fails to form an adequate and stable symbiosis with the mothering object, a developmental failure rendering the child's image of mother inconstant. This developmental failure leaves the individual vulnerable to regression when facing the second and final phase of individuation in late adolescence. The regression itself goes back to the preverbal, presymbiotic stage of autism with loss of ego boundaries, merger experiences, and replacement of reality by autistic fantasy.

Ping-Nie Pao Pao attempted to synthesize the theoretical threads of many schools and forebearers. Based upon extensive clinical experience, he subtyped schizophrenic patients into those with acute cases, for whom conflict plays a more pivotal role, and those with chronic cases, with higher genetic-biological loading.

Pao, like Freud, was impressed with the catastrophic panic experience that signaled the onset of the schizophrenic process and symptom formation. This process is precipitated by psychodynamic conflicts no different in content from those experienced by all people. In the schizophrenia-vulnerable person, however, these conflicts no longer generate neurotic levels of anxiety but at some point catalyze a crisis known as organismic panic, a term modeled after Mahler's developmental observations of states of extreme infantile distress. This panic brings with it paralysis of the ego's integrative capacity and fragmentation of the sense of continuity of self. The latter process constitutes an unbearable loss of a basic sense of safety, leading the ego to mobilize primitive or regressive defenses to reestablish and protect a sense of self, albeit pathologically. This attempt at adaptation or recovery results in postpanic emergence of a different personality, either pieced together with or distorted by psychotic symptoms. Typical delusions, for example, help to construct a new sense of meaningful self and, though often very unpleasant, are clung to tenaciously because their loss threatens return of disorganization and panic.

Pao's etiological-developmental theory attempts to explain the origin of the schizophrenic vulnerability or organismic panic and regression. Like other relationalists, he placed etiology in the experiences of early development. Aberrant constitution and inappropriate mothering combine to generate a series of failed emotional cuings within the dyad, leading to frequent episodes of infantile organismic distress or “pain in being held and pain in being laid down.” Cumulative exposures to such distress bend further development in maladaptive directions, including a tendency toward primitive defensive, impaired capacity for instinctual neutralization; an inability to maintain a sense of reality constancy, heightened aggressive responses to frustration; and heightened wishes for closeness with others, coupled with a dread of self dissolution in symbiosis (the need-fear dilemma originally described by Donald Burnham, Arthur Gladstone, and Robert Gibson). These vulnerabilities lie dormant and do not produce symptoms until the advent of adolescent drive demands and stress.

CRITIQUE OF PSYCHOANALYTIC THEORIES

Psychoanalytic theories of schizophrenia have become increasingly criticized in recent years, especially by analysts such as Martin Willick, mainly because the traumatic-developmental perspective on etiology lacks credibility. Virtually all psychoanalytic theorists postulate an experiential disharmony between the mother and her preschizophrenic infant. Whether this derives from genetic or constitutional factors in the infant or from psychological factors in the parent is secondary, as the purported central pathogenic elements are dysphoric experiences that become internalized as aberrant psychological structures. Explicitly or implicitly, the psychogenic models of schizophrenia regard these experiences as sufficient to explain most, if not all, cases of the syndrome.

Several considerations cast doubt on this postulate. First, recent findings from infant and childhood development challenge many of the assumptions put forth by psychoanalytic developmentalists. For example, normal development is not like pathological stages projected backward. Infants and children are active, stimulus seeking, and socially oriented from day one. Stages such as the narcissistic, autistic, symbiotic, or schizoid-paranoid are not observed, so it is doubtful that schizophrenia could represent regression to one or more of them. Schizophrenia does represent a regression of mental capacity insofar as mentation becomes less complex, especially during active phases. However, as Willick points out, regressive phenomena can result from a primary disturbance in brain function later in life rather than represent a return to experientially determined earlier fixation. Also, infants and children are far more powerfully and intricately preprogrammed for adaptation and survival than psychoanalytic theorists assumed; almost without exception they saw infants as helpless, utterly dependent, and mindless creatures of infinite malleability. The fact that many infants and children survive despite unusually bleak or traumatic rearing suggests that factors orthogonal to nurture may be operative.

Second, clearly some people who develop schizophrenia as adults come from basically healthy families and undergo normal growth and development—a direct challenge to trauma hypotheses of etiology. Furthermore, the reported childhood suffering of schizophrenic patients is often no more severe or profound than that of patients with other forms of mental illness, suggesting the necessary presence of additional, nonexperiential, pathogenic factors. For example, why might similar adverse life circumstances result in anxiety or depression in one person and hallucinations and thought disorder in another?

Finally, the psychogenic theories have difficulty explaining why in most cases of schizophrenia some two decades intervene between the purported pathogenic infantile traumatic experiences and the onset of overt symptoms. If the experiences postulated by these theories do indeed occur, one might expect to see symptom formation at the time followed by predictable and nonrepressible (i.e., observable) deformities in subsequent development, at least in some cases. An infantile catastrophe severe enough to produce an illness of the magnitude of schizophrenia is not likely to go unnoticed, yet such catastrophes and their immediate behavioral consequences have not been documented.

Generally, psychoanalytic theories of schizophrenia (especially those with a more descriptive perspective) continue to inform and help clinicians understand the patients they encounter. In this context, they are worthy of study. Furthermore, while exploring the past with schizophrenic patients may no longer be expected to yield etiological or historical truth, it does provide meaningful metaphors that can be useful in the empathic dialogue between doctor and patient.

FAMILY DYNAMICS AND TRANSACTION MODELS

The family transaction models of schizophrenia represent attempts to understand and explain the syndrome as the transmission of aberrant interactions from the family to the patient. The models are compatible with object-relation-oriented psychoanalytic psychogenic theories in assuming that psychopathology is determined largely by experience and learning within the family during growth and development. The models differ, however, in their respective hypothesis-generating and hypothesis-validating databases. For psychoanalytic theories, the data are the associations of individual patients for family transaction theories, the data are observed interactions in families with one or more schizophrenic members.

These models emerged following World War II, when clinical work with the families of schizophrenic patients revealed with increasing frequency that irrationality was not limited to the identified patient. Unusual and unpredictable interactions were observed between dyads within the family or among family members as an entire unit. Motivated by the idea that these interactions might be schizophrenogenic, several clinical investigators began describing these families and their transaction patterns in some detail.

Gregory Bateson and Donald Jackson Gregory Bateson and Donald Jackson outlined a form of family interaction that they labeled the *double bind*. The interaction usually occurs between a parent and the schizophrenic offspring; it consists of the former giving the latter incompatible (if not antithetical) messages (e.g., stiffly avoiding a physical embrace while asking, “Why don't you show me more affection?”). This sets up an inescapable damned-if-you-do-and-damned-if-you-don't situation, or double bind, in which the offspring feels paralyzed. Bateson and Jackson hypothesized that repeated exposure to such a dilemma generates or aggravates the schizophrenic state.

Ruth and Theodore Lidz Ruth and Theodore Lidz studied the characteristics of families that had a schizophrenic offspring. They looked for and observed disorders in the relationships among family members, especially the triad of mother, father, and schizophrenic child. They described several irrational patterns, such as marital schism between parents who remain married because of pathological interdependence despite considerable overt conflict; marital skew between parents who hide chronic disagreement behind a facade of harmony; permeable generational boundaries, where one parent requires the schizophrenic child to assume a parental role; eroticized parent-child relatedness, wherein one parent treats the schizophrenic child as a peer or contemporary; and emotional divorce, in which family members fail to acknowledge and confirm one another's psychological integrity. The Lidzs asserted that such rrational family functioning sufficed to account for schizophrenia in certain offspring exposed over their formative years.

Lyman Wynne and Margaret Singer Lyman Wynne, Margaret Singer, and colleagues explored communication and cooperation within families with a schizophrenic offspring. From their careful observational work came the concept of communication deviance. This includes parental communications that lack commitment to ideas and percepts; parental communications that are unclear because they are filled with idiosyncratic themes and ideas or have language anomalies, discursive speech, and problems with closure; and parental communications that reflect an inability to establish or maintain a shared focus of attention during transactions with another family member. Among parental communications, they identified two common forms: an amorphous style in which communications are vague, indefinite, and loose and a fragmented style in which communications are easily disrupted, are poorly integrated, and lack closure. They also described familial displays of mutuality, hostility, or both, which serve as facades hiding antithetical themes and conflicts.

Unlike most other family transactional theorists, Wynne and Singer were able to operationalize their concepts into reliable measures, thus allowing their hypotheses to be tested more systematically. They found communication deviance to be more specific to families of schizophrenic patients than to families of patients with depression, personality disorders, neuroses, or no pathology. Furthermore, amorphous patterns of communication deviance correlated more frequently with process schizophrenia, and the fragmented patterns with reactive schizophrenia. They also found significant quantitative correlations between the amount of communication deviance in the parents and the severity of psychopathology in their offspring. For example, schizophrenic offspring were in families in which both parents had high levels of communication deviance; normal and neurotic offspring were in families in which both parents had low levels; and borderline offspring were in families in

which one parent's level was high and the other parent's was low.

Expressed Emotion More recently, family investigators have described several family factors that interact powerfully with schizophrenia, either to precipitate its emergence or to aggravate its course. One factor, called expressed emotion, consists of critical or emotionally overinvolved attitudes and behaviors displayed by parents toward their ill offspring. Another family transactional factor of interest and current study is negative affective style. It includes four kinds of parental behavior: criticism, guilt induction, intrusiveness, and inadequate support. Schizophrenic patients living with families with high expressed emotion or negative affective style relapse significantly more often than those living in families with low expressed emotion or normal affective style.

Critique of Family Transaction School Like the psychoanalytic theories of schizophrenia, the family transaction theories have come under considerable criticism as etiological models. With the exception of communication deviance, few of the family transactions described above are demonstrably specific to schizophrenia. Furthermore, the observed irrational transactions among these families may possibly derive from the necessity of dealing with a child whose deviance is independently determined, thus reversing the hypothesized causal vector. In the absence of hard etiological data, assuming that families transmit and concentrate their irrationalities on a designated family member–victim is a nonproductive assignation of blame that does little to advance understanding but much to undermine working alliances between professionals and afflicted families.

Like the psychoanalytic theories of schizophrenia, the family transaction theories remain viable and useful as descriptive constructs. While irrational behaviors in the families of schizophrenic patients may not cause the illness, these behaviors are nevertheless present and real in their evocative effects. As demonstrated by the expressed emotion and affective style studies, the family's emotional milieu can profoundly influence the onset or course of schizophrenia. Family transactional stress may not be causative. However, strong evidence, indicates that it can be powerfully facilitative in both pathological as well as therapeutic directions. As such, the family theories also fit well into the earliest neurodynamic theory of schizophrenia: the vulnerability-stress, or stress-diathesis, hypothesis.

STRESS-DIATHESIS MODELS

Antecedents The first neurodynamic model, the stress-diathesis hypothesis, views schizophrenia dynamically as a product of interacting forces, some genetic or biological and some psychological, some innate or constitutional and some learned through experience. Unlike the purely psychodynamic theories, nature is more important, as suggested by genetic studies and the efficacy of biological treatments. Both nature and experience, however, are considered necessary to describe and understand schizophrenia.

The Finnish adoption study of Pekka Tienari and colleagues illustrates this model. Comparing adopted-away children of schizophrenic mothers (high genetic-risk probands) with adopted-away children of nonschizophrenic mothers (low genetic-risk controls), they found that schizophrenia developed only in probands with genetic vulnerability who were raised in adoptive families in which the emotional environment was demonstrably stressful. None of the high genetic-risk probands raised in less stressful adoptive families developed psychosis. Likewise, none of the low genetic-risk probands raised in unhealthy adoptive families developed psychosis, although many developed other forms of psychopathology. These results strongly suggest that both a stressful rearing environment and an innate vulnerability to schizophrenia are necessary to generate the syndrome.

The concept of such an interaction began with Freud. Describing the origin of neurosis in *On the History of the Psycho-Analytic Movement*, he wrote:

> Disposition and experience are here linked up in an indissoluble etiological unit. For disposition exaggerates impressions which otherwise have been completely commonplace and have no effect, so that they become traumas giving rise to stimulations and fixations; while experiences awaken factors in the disposition which, without them, might have long remained dormant and perhaps never have developed.

Freud could well have been writing about the origins of psychosis. Certainly, subsequent psychoanalytic theorists took this model seriously in explaining schizophrenia, especially those that emphasized deficit. The true conceptual fathers of today's stress-diathesis model of schizophrenia, however, are Sandor Rado and Paul Meehl.

Rado hypothesized that schizophrenia begins with an inherited disposition, or genotype. Interaction of this genotype with the environment produces the schizophrenia phenotype, a personality type or trait called the schizotype. Central to this trait is an inherent incapacity to experience pleasure. “In the schizotype the machinery of psychodynamic integration is strikingly inadequate, because one of its essential components, the organizing action of pleasure—its motivational strength—is innately defective.” This defect impairs the development of initiative and leads to schizo-adaptations such as compensatory overdependence on others (especially parents) and the elaboration of intricate cognitive processes devoid of affect. Anhedonia results in weak emotional bonds and leads to attenuated relationships. A well-compensated schizotypical person retains a stable schizoid personality. One of poorly compensated schizotype develops exaggerated, bizarre behaviors. Schizophrenia proper represents a decompensated schizotype with adaptive incompetence. The nature and severity of the schizo-adaptation depends on genotypic loading and the degree of familial and environmental stress.

To Meehl, the inherited schizophrenic phenotype (which he labeled schizotaxia) consists of a defect in neural integration. This defect plus social learning (environment) leads to an abnormally organized personality (the schizotype) characterized by cognitive slippage (thought disorder), anhedonia, ambivalence, and aversion to human relationships. Further progression from schizotypy to schizophrenia is a function of the nature and severity of environmental stress versus the availability of help and support.

Model This hypothetically pathogenic interaction between nature and experience came to be known as the stress-diathesis model. As currently conceived, this model accepts that the relative roles of nature and nurture in the etiology of schizophrenia will remain obscure until there are markers for genetic predisposition or constitutional vulnerability. It shifts emphasis from the etiological role of psychodynamic factors to their role in facilitating and preventing the expression of the disease process.

Vulnerability to schizophrenia is seen as an enduring proclivity toward developing clinical symptoms. It is a stable trait independent of nonenduring psychopathological states, meaning that it is present premorbidly, at onset, during symptomatic efflorescence, and during remission. This trait should not, however, be regarded as developmentally static or fixed. Rather, vulnerability may be shaped epigenetically at each developmental phase. Aspects of vulnerability are undoubtedly genetic. Some may be acquired biologically throughout intrauterine, birth, and postnatal complications. Season of birth may also contribute, for reasons yet to be ascertained. The evidence for psychosocially acquired vulnerability is meager at present, but this hypothesis cannot be ruled out.

The stress side of this model postulates that a variety of stressors (i.e. internal or external events requiring adaptation) can convert vulnerability into symptoms. Therefore, coping strengths or supports that diminish stress should minimize or prevent clinical expression of vulnerability.

Following the model, the vicissitudes of schizophrenia are determined by the nature of vulnerability and stress and by the individual's strengths and environmental supports. Interaction of sufficient stress with sufficient vulnerability can lead to transient, intermediate (prodromal) states of dysfunction that amplify existing cognitive, affective-autonomic, and social-coping deficits. This compromised mental state, in turn, interacts negatively with stressors and magnifies their effect in a downward spiraling deterioration that culminates in a full-blown clinical syndrome.

Vulnerabilities to Schizophrenia Finding, mapping, and integrating these vulnerabilities have become a central effort in current schizophrenia research. Virtually all of this investigation has focused on demonstrable phenotypic manifestations of hypothetical genotypic vulnerabilities in children and adolescents at risk for schizophrenia. The list of these vulnerabilities is extensive and grows each year. First are deficits in processing complex information, in maintaining a steady focus of attention, in distinguishing between relevant and irrelevant stimuli, and in forming consistent abstractions. Second are dysfunctions in psychophysiology, suggesting deficits in sensory inhibition and poor control over autonomic responsibility, especially to aversive stimuli. Third are impairments in social competence, such as processing interpersonal stimuli, eye contact, assertiveness, or conversational capacity. These deficits probably reflect both a core disturbance of schizophrenia (vulnerability) and the social outcomes of severe psychopathology. In the past, the source of these difficulties was often attributed to such external elements as neuroleptic drugs or institutions, a perspective that unduly diverted attention from their primacy in the disorder. Fourth are general coping deficits such as overevaluating threat, underestimating internal resources, and extensive use of denial.

Phenotypic manifestations of vulnerability may not be observable premorbidly, especially if the genotypic vulnerability has variable onset of expression. Huntington's disease, for example, is an adult-onset neurological deterioration leading to psychosis and dementia; in a similar fashion, many cases of schizophrenia could result from a genotype whose phenotypic expression is not triggered until late adolescence or early adulthood. This phenotype may be a deficit in the neurophysiological

maturation of self systems during adolescence or a still later-onset neural deterioration or inhibition of these same systems in adulthood. Early-onset genotypes may help to account for patients with easily identifiable phenotypic deviations that begin in childhood as schizotypal aberrations and progress to chronic process schizophrenia later on. Later-onset genotypes may help explain the sudden occurrence of schizophrenia later in life in persons with normal growth and development and healthy premorbid personalities.

Stressors Systematic studies of the stresses that affect the course of schizophrenia, aside from the family environment, have focused on social class and culture, social networks and life events. Socioeconomic and cultural factors have a long history of empirical association with schizophrenia. One of the most replicated findings in the schizophrenia literature is the clustering of schizophrenic patients in the lowest social classes, especially in urban communities. Few now hold that a poor socioeconomic environment causes schizophrenia, but few doubt that it has a major impact on its course. Poverty, ignorance, unemployment, social isolation, poor nutrition, and marginal health care are powerful chronic stressors that lead to frequent breakdown in vulnerable persons.

Schizophrenia and social network are highly interactive, cross-sectionally and longitudinally. Schizophrenic patients usually have social networks that are smaller, less interconnected, simpler, more dependent, casual, nonintimate, and peopled with family as opposed to peers. The interplay between schizophrenia and social networks appears to be circular rather than linear. Initially, the major vector is schizophrenia on social network. Following the appearance of clinical symptoms, however, social network is likely to exert a powerful influence on the subsequent vicissitudes of schizophrenia.

Stressful life events have demonstrated association with schizophrenia, but it may not always be necessary or direct. Questions often arise concerning whether stress has different effects on disease onset and recurrence and whether a stressful event precedes illness or represents a product of symptom exacerbation. Convention dichotomizes stressful events into those that are ambient, nonindependent, or chronic and those that are independent or acute. The former are stresses associated with everyday living, such as family, work, poverty, physical disability, and mental deficit; the latter are stresses associated with largely external or unusual changes, such as loss, death, acute illness, and moves, especially if these changes are unanticipated, undesired, and uncontrolled. Research suggests a high frequency of such events shortly before schizophrenia onset or symptom exacerbation. Furthermore, there appears to be an important interaction between maintenance antipsychotic medication and life-event stress. Patients in the community without medication are vulnerable to both acute and chronic stress. Patients taking medication, however, appear to be protected against either type of stress but are likely to suffer relapse if both types occur concurrently.

NEURONAL NETWORK MODEL

The latest neurodynamic model of schizophrenia hypothesizes with greater specificity the neuroanatomic nature of the vulnerability to schizophrenia and its biophysical contribution to symptom formation. It is build upon postulates that most human mentation arises from neuronal networks organized in parallel distributed processing systems. Furthermore, these systems can, under certain physiological conditions, generate thoughts or feelings in ways orthogonal to the normal processes by which psychology is generated. This pathological development, called “memory parasitism,” can produce mentations that are literally products of aberrant connectedness among the neurons of the brain and are experienced and interpreted psychodynamically as unintended and alien by the bearer of the brain. The physical conditions conducive to such developments are unknown but hypothesized to involve a loss of parallel distributed processing neural circuit density from perinatal insults or from a genetically or developmentally programed excessive pruning of connections between cortical nerve cells.

Parallel Distributed Processing Model Mental events are basically neuronally constructed representations of experience as memory. Memories in general are accessible to consciousness in a manner that is content addressable, i.e., the brain can use part of a memory to access the memory in its entirety. Artificial intelligence researchers such as Donald Hebb, Warren McCulloch, Walter Pitts, John Hopfield, and James McClelland and David Rumelhart have discovered that certain types of computer-assisted computations retrieve content-addressable information in a fashion that partially simulates nature. These systems are composed of large numbers of very simple computing units, generally referred to as “neurons,” which are densely interconnected via synapses. There is no single command unit; the effectiveness of the network as a whole reflects the cooperative interactions of its parts. Each neuron simultaneously receives information from a large number of other neurons and computes its response to these inputs in parallel with the computations of the other neurons of the system. No one-to-one correspondence exists between a memory and activation of a particular neuron. Instead, a memory corresponds to a pattern of activation involving many neurons of the network; memories are created by modifying functional connections among arrays of neurons. Networks that store and retrieve information on the basis of distributed patterns of activation are referred to as parallel distributed processing systems.

An increasing body of neurobiological research suggests that the functional architecture of the mammalian cerebral cortex can be broken down into a number of linked parallel distributed processing-type neuronal assemblies. Of special interest is the long-term memory system, which probably involves the brain's hippocampal and cortical areas. The hippocampus is active during a critical “gestational” period (ranging from minutes to hours), during which input information is distributed to interconnected circuits linking frontal, parietal, and sensory-association cortical areas. If hippocampal functioning is temporarily impaired, memories due to be distributed will be permanently lost. However, once information is functionally linked within the cortex, its retrieval seems not to require an intact functioning hippocampus. Activating a posterior sensory area via an image representation or a frontal area via an abstract concept such as a goal or generalization can immediately “seed” widely distributed cortical circuits to yield complex memories.

Parallel distributed processing systems with content-addressable mnemonic representations probably constitute the nemodynamic underpinning that is normal mental activity and psychology. The psychoanalytic method of free association, for example, relies on the content addressability of mental representations. Freud first observed that human mentation follows patterns that are seldom logical or governed by the rational rules of secondary process. More often than not, mental representations are connected by condensations, displacements, and symbolizations of mnemic content, whether that content is concrete imagery, stories, or abstract theories. These are the primary processes of mentation postulated by Freud to underlie the formation of dreams and neurotic-level symptoms. Content addressability is a new term for an old psychoanalytic postulate that every mental event (thinking, feeling, acting) is overdetermined and can be attained via multiple pathways. This is Robert Waelder's principle of multiple function, wherein a vast array of related factors (e.g., id, ego, superego) must be considered in understanding any behavior. The functional dynamics of parallel distributed processing systems also suggest that an enormous amount of neuronal activity accompanies any single mental event or idea, thus accounting for the dynamic unconscious of psychoanalysis.

Attractor Dynamics in Parallel Distributed Processing Systems Most psychiatric syndromes include some form of repetitively reproduced mental content ranging from obsessional thoughts to the idea fixe of psychotic delusions. Such repetitive memories can be produced in simulated parallel distributed processing systems on the basis of “attractor dynamics.” As noted by Hopfield, these systems tend to be attracted to particular activation patterns, which are low-energy states of the system. *Energy* here is defined on the basis of a branch of physics known as statistical mechanics. Statistical energy corresponds to the degree of disorder in the system. Neurons that interact freely in a parallel distributed processing network will tend to be attracted to activation patterns that minimize its statistical energy. These patterns correspond to the memories of the system, which tend to be reproduced because of these energy minima. Particularly strong memories correspond to very deep energy minima that tend to pull new information to the existing activation pattern to the exclusion of other activation patterns. That is, a memory with a particular deep-energy minimum strives relentlessly to express itself. It is called “parasitic” because it coerces a wide range of input information into a single memory-activation pattern.

One of the authors (R. H.) demonstrated that computer-simulated parallel distributed processing models can provide an account of how such pathological memories are induced. Under varying circumstances described below, a memory can be created that does not reflect any specific prior experience. These spurious memories are not linked to naturally occurring and stored input-activation patterns, but are created de novo by the brain itself and can be triggered by the most irrelevant internal or external cures. These memories link or condense quite disparate contents and may appear at times to be the products of primary process. However, they cannot be grasped by the empathic primary processes of others, i.e., they remain strange and beyond understanding in Karl Jasper's sense of being self-evident or common to human experience. These memories tie together representations or fragments of representations that are ordinarily not linked even by the weakest and most remote primary process pathways, hence the term “loose association.” In a sense, they are violations of content addressability, in which associations are no longer free but are constrained by disturbed attractor dynamics.

Parasitic memories are the product of biophysical energic processes, not primary and secondary psychological processes. At the same time they arise in the midst of other neural networks that are not compromised and that generate normal and abnormal psychology in the usual way (i.e., by primary and secondary processes). As such, persons experiencing parasitic memories may feel as if parts of their brain or mind are functioning out of their control. They may experience these unintended mental events as if someone else were putting thoughts into their minds. Such post hoc reasoning that an outside force is inserting thoughts into one's mind may reflect the difficult task of describing and making sense of the experience of one's mind being repeatedly “captured” by a perseverative attractor seemingly acting according to its own will. In this fashion “parasitic” memories in a parallel distributed processing system can account for many of the seemingly bizarre symptoms of schizophrenia.

Pathophysiology of Memory How might such memories come about? In computer-based parallel distributed processing systems, parasitic memories are known to

derive from two different types of “pathology.” First, parasitic memories can be produced when the system attempts to store too many memories for its capacity. Second, parasitic memories can be produced when an insufficient number of synaptic connections link the neurons of the system. It is further hypothesized that these two types of pathology are “located” in the system of long-term memory storage and recall involving the hippocampus and diffuse cortical circuits of the brain.

There is considerable recent interest in pathology involving the hippocampus in schizophrenia. Neuroradiological and postmortem studies have suggested reduced hippocampal volume and disarray of its pyramidal cells in some schizophrenic patients. Reductions in pyramidal cell number, functional efficiency, or information transfer capacity—perhaps occurring secondary to prenatal complications such as viral infection—could downsize memory storage capability in the hippocampus, thereby rendering it vulnerable, as a parallel distributed processing system, to the induction of memory parasitism.

A second form of schizophrenia may reflect abnormalities in postnatal brain development. Considerable evidence in humans and other primates now indicates that following birth, an overabundance of corticocortical connections exists; these connections are then selectively pruned and shaped during later developmental periods. In humans pruning of cortical synapses seems to extend well into adolescence for connections involving frontal areas. Moreover, the continued elaboration and pruning away of cortical synapses probably continues, albeit at a much reduced rate, throughout adulthood.

These developmental studies, when considered in light of parallel distributed processing simulations, suggest a second mechanism leading to parasitic or spurious memories—reductions in corticocortical connections which are a part of these developmental processes. As suggested by Erwin Feinberg, if excessive developmentally induced pruning of these connections leads to a form of schizophrenia, a ready explanation for this age of onset and involvement in frontal areas is provided; synapse pruning is most prolonged in frontal areas, and the end of adolescence is when the effects of this pruning process would first be fully felt.

An example of a specific model of psychotic symptomatology that is empirically testable has been described by Hoffman and colleagues. A neural network was created to simulate certain aspects of speech perception. Pruning connections between the working memory component of the network caused perceptions to emerge spontaneously. This observation thus provides a model for hallucinated speech (“voices”). The model also permitted simulation of the hypodopaminergic state, and this neuromodulatory state tended to eliminate voices. The testability of the model derived from the fact that hallucinogenic networks also demonstrated subtle but definite speech perception impairments. Empirical study with actual schizophrenic patients demonstrated these speech perception impairments, which were not detected in schizophrenic patients who did not hallucinate. These simulations suggest that hallucinated voices correspond to parasitic memories that usurp working memory circuits of speech perception.

Implications The parallel distributed processing paradigm is neurodynamic because it postulates the creation of symptoms, at least in part, on the basis of defects in the hardware of the brain. Purely psychodynamic paradigms account for pathology solely as a reflection of aberrant “software” or conflicting functional programs in the mind. The parallel distributing processing model regards psychodynamics as necessary but often not sufficient to explain most schizophrenic symptoms. It is a model at the mind-body interface and views psychotic psychopathology as the product of a complex interaction of skewed mentation arising organically with normal or abnormal psychology arising psychodynamically.

The model remains speculative but offers an advantage of being falsifiable. Many of the functional and neuroanatomic predictions of this model may be subject to hypothesis testing and validation, thus bringing theorizing about schizophrenia closer to the realm of scientific empiricism.

In terms of sheer volume, the bulk of psychodynamic theories belong to psychoanalysis. In terms of content, the major issues have not changed much since Freud posited two theories—the structural-conflict theory and the withdrawal-deficit theory. By never integrating these two, Freud seemed to be saying that schizophrenia may be both, in part explained by intrapsychic conflict and in part by something else. Psychodynamic theories became more elaborate over the ensuing years, yet one can still follow the thematic threads of Freud's original part-explanations. Conflict theory has seen a line of development polarized toward object relations, experiential learning within the family, stressors that interact with vulnerability, and the psychological sequelae of organically disturbed neural circuitry. Deficit theory has seen a line of development polarized toward individual drives, complex constitutional inborn factors, the physical vulnerabilities of the stress-diathesis model, and the hard-wiring defects of parallel distributed processing neural networks. Today, these threads are each regarded as valid facets of the overall phenomenon. Both are necessary for a comprehensible and potentially workable theory of schizophrenia.

While none of these psychodynamic models of schizophrenia has solved the mystery of its etiology, each system has offered cogent hypotheses or educated guesses. The neurodynamic models see nature as primary or at least as initiating a process of negative interaction with environmental experience, which becomes pathogenic sooner or later in life, depending on the onset trigger of the genotype. The family transactional model, at least as originally conceived, sees nature or experience as primary and, frequently, as sufficient for generating schizophrenia. Psychoanalytic models have posited one or the other or both. Since Sullivan and Mahler, however, the emphasis has shifted more uniformly toward the etiological primacy of traumatic nurturing experiences during early development, a shift that has come under increasing criticism as outlined.

Overall, the impressive evidence for the existence of genetic and constitutional factors in schizophrenia has raised questions about the etiological hegemony of experience and learning in early development. Hypothetically, some (if not many) cases of schizophrenia may possibly arise from an adolescent-onset neurological dysfunction or deterioration in people who are, up to that point, developmentally normal. Such a process, along the model of overpruned parallel distributed processing neural networks, selectively inhibits or destroys later developmental levels of personality, especially those neuronal networks involved with structures, functions, and representations of the self. In response, the individual falls back on more primitively organized levels of personality and development. Such regression is compensatory and adaptive rather than primary and motivated. It does not occur because of developmental fixation but because simpler developmental levels and patterns may be the only ones left.

IMPLICATIONS FOR TREATMENT

Psychodynamic theories of schizophrenia carry with them distinct implications for treatment. Early proponents of the psychoanalytic conflict model advocated the classical techniques of clarification, confrontation, and interpretation. Early proponents of the psychoanalytic deficit model introduced additional strategies. Federn, for example, felt that the usual psychoanalytic techniques aimed at depression, whereas with schizophrenia, the goal was to foster rerepression. As such, he encouraged positive transference, avoided negative transference, protected patients from undue anxiety and insomnia, taught them to improve their capacities for attention and thinking, exhorted them to give up unrealistic life goals, provided support beyond analytic hours in the form of a skilled nurse-assistant available to the patients at home, and offered consultation to the patient's family (recognizing the importance of the home environment to outcome).

Proponents of the family transactional theories uniformly advocate family therapy in some form. Those that view the family milieu as causing schizophrenia usually regard the entire family as the patients or as the problem and focus interventions accordingly. Those that regard the family as facilitative rather than etiological emphasize the positive and negative effects that domestic tensions can have on the course of the identified patient. Technical strategies in the first instance are more interpretive. In the second instance, they are more psychoeducational.

Proponents of the stress-diathesis and parallel distributed processing models advocate any intervention that enhances strength and support and minimizes stress and vulnerability. This includes psychobiological as well as psychodynamic treatments. The neurodynamic models are the only ones that formally (i.e., theoretically) incorporate biology and endorse it therapeutically. They also define psychodynamic treatment more liberally. Any and all forms of psychosocial intervention, from individual psychotherapy to social skills training, are potentially useful, depending on the modality's track record of efficacy with the specific clinical situation or condition.

Conflict psychodynamic models, in keeping with their bias toward object relations and development in the family, emphasize the therapeutic centrality of the doctor-patient relationship. This relationship is facilitating, parental, soothing, mirroring, and protective, and the patient grows by internalizing the interactions that transpire within the dyad. The patient's actual interpersonal experience of the therapist is crucial: the therapist's reality and benignity serve as reality tests for the patient's transferentially distorted images.

Deficit psychodynamic models, in keeping with their bias toward the patient as an individual with phenotypic abnormalities, emphasize the therapeutic centrality of cognitive perspective and control. The goal of treatment is to enhance the power of the ego by expanding its knowledge and control over the inner drives and psychopathological idiosyncracies. Enlightenment replaces unconscious defense with conscious choice. Therapy from this perspective focuses primarily upon developing the patient's cognitive systems through psychoeducation, training, and rehabilitation. The patient comes to realize that something is wrong, what that something is, and how it can be dealt with.

The conflict psychodynamicists once eschewed deficit theories as therapeutically nihilistic, insisting that there was no way to make up for a biological defect by psychological means. Such an assertion may be literally correct but operationally erroneous. For example, psychological manipulation cannot make paraplegics walk under their own power, but it can train them in prosthetic ambulation, and it can enhance their adaptation and quality of life. Whatever the origin of schizophrenia, its successful psychological treatment involves both the resolution of intrapsychic conflict through insight and the acquisition of psychic structure through affective relationships. If the core of schizophrenia is psychological, then treatment addresses the sick self; if the core is defect, then treatment addresses the healthy self. In the former, it minimizes weakness; in the latter, it maximizes strength. In most cases, it does both.

BIOPSYCHOSOCIAL MODEL

One body of theory encompasses all of the foregoing twentieth-century trends—the biopsychosocial medical model of George Engel. According to this model, each individual patient consists of and participates in multiple systems that are related but also distinct from each other. Common systems are subatomic particles, atoms, molecules, organelles, cells, tissues, organs, organ systems, central nervous system, individual, dyad, family, community, culture and subculture, society and nation, and biosphere. In understanding health and disease, all systems are relevant. Each system of this model has a functional structure, one of its purposes being the reduction of complexity and randomness to protect that system's integrity. The functional structure of the psychological systems in this model consists of meanings that serve to order experience through understanding and explanation.

Psychodynamic approaches to treatment should not ignore biology because the latter exists outside the realm of empathy and meaning. Biological approaches to treatment should not justify psychological retreat from patients because conflict cannot be teased apart by electrophoresis. Finally, treatment advocates of both approaches should be aware of patients' social, cultural, and political need for a place of dignity and safety within society. That is, patients also require adequate attention at the social level of the biopsychosocial system.

Schizophrenia presents most dramatically at the psychological level as a loss or distortion of the self as a meaningful entity. Despite this, schizophrenia is not entirely or even essentially psychological in its nature. Accordingly, proper medical attention to this disorder should be aimed at any and all relevant systems in the biopsychosocial hierarchy. Whatever schizophrenia may be, it is profoundly disabling and usually chronic. Anything therapeutic that works with sufficient safety is relevant, whether it is biological, psychological, or sociological.

SUGGESTED CROSS-REFERENCES

The relevance of brain structure and function in schizophrenia is discussed in Section 12.3; neurobiology is discussed in Section 12.4; and genetics in schizophrenia is discussed in Section 12.5. Somatic treatment is discussed in Section 12.8, psychosocial treatment in Section 12.9, and individual psychotherapy in Section 12.10. Theories of personality and psychopathology are discussed in Chapter 6. Schizophrenia in childhood is discussed in Chapter 38, and schizophrenia in late life is discussed in Section 51.3f.

SECTION REFERENCES

Burnham DL, Gladstone AI, Gibson RW: *Schizophrenia and the Need-Fear Dilemma*. International Universities Press, New York, 1969.

Cohen JD, Servan-Schreiber D: Context, cortex and dopamine: A connectionist approach to behavior and biology in schizophrenia. Psychol Rev *99:*45, 1992.

Engel GA: The need for a new medical model: A challenge for biomedicine. Science *196:*129, 1977.

Federn P: *Ego Psychology and the Psychoses*. Basic Books, New York, 1952.

Feinberg I: Schizophrenia: Caused by a fault in programmed synaptic elimination during adolescence? J Psychiatr Res *4:*319, 1982.

Freud S: The interpretation of dreams. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 5. Hogarth Press, London, 1953.

Freud S: On the history of the psycho-analytic movement. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 14. Hogarth Press, London, 1957.

*Freud S: Psychoanalytic notes on an autobiographical account of a case of paranoid (dementia paranoides). In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 12. Hogarth Press, London, 1958.

Friston KJ: Theoretical neurobiology and schizophrenia. Br Med Bull *52:*644, 1996.

Greenberg JR, Mitchell SA: *Object Relations in Psychoanalytic Theory*. Harvard University Press, Cambridge, MA 1983.

Hebb DO: *The Organization of Behavior*. New York, Wiley, 1949.

*Hinshelwood RD: The difficult patient. Br J Psychiatry *174:*187, 1999.

Hoffman RE: Computer simulations of neural information processing and the schizophrenia/mania dichotomy. Arch Gen Psychiatry *44:*178, 1987.

Hoffman RE: The mechanism of positive symptoms in schizophrenia. Behav Brain Sci *14:*33, 1991.

Hoffman RE, Dobscha SB: Cortical pruning and the development of schizophrenia: A computer model. Schizophr Bull *15:*477, 1989.

Hoffman RE, McGlashan TH: Alterations of speech, thought, perception, and self-experience. In *Psychiatry*, vol 1, A Tasman, J Kay, JA Lieberman, editors. Saunders, Philadelphia, 1997.

Hoffman RE, McGlashan TH: Corticocortical connectivity, autonomous networks and schizophrenia. Schizophr Bull *20:*257, 1994.

*Hoffman RE, McGlashan TH: Parallel distributed processing and the emergence of schizophrenic symptoms. Schizophr Bull *19:*119, 1993.

Hoffman RE, McGlashan TH: Reduced corticocortical connectivity can induce speech perception pathology and hallucinated “voices.” Schizophr Res *30:*137, 1998.

Hoffman RE, McGlashan TH. Synaptic elimination, neurodevelopment, and the mechanism of hallucinated “voices” in schizophrenia. Am J Psychiatry *154:*1683, 1997.

Hoffman RE, Rapaport J, Rezuan A, McGlashan TH, Harcherik D, Servan-Schreiber D: The neural network simulation of hallucinated “voices” and associated speech perception impairment in schizophrenia patients. J Cogn Neurosci *7:*479, 1995.

Hopfield JJ: Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA *79:*2554, 1982.

Jaspers K: *General Psychopathology*. Grune & Stratton, New York, 1959.

Lichtenberg JD: Pao's theory: Origins and future directions. In *Towards a Comprehensive Model for Schizophrenic Disorders*, DB Feinsilver, editor. Analytic Press, Hillsdale, NJ, 1986.

Lidz T: *Schizophrenia and the Family*. International Universities Press, New York, 1965.

McCulloch WS, Pitts W: A logical calculus of the ideas imminent in nervous activity. Bull Math Biophys *5:*115, 1943.

McGlashan TH: Early detection and intervention of schizophrenia: Rationale and research. Br J Psychiatry *172:*3, 1998.

McGlashan TH: The profiles of clinical deterioration in schizophrenia. J Psychiatr Res *32:*133, 1998.

McGlashan TH: Psychosocial treatments of schizophrenia: The potential of relationships. In *Schizophrenia: From Mind to Molecule*, N C Andreasen, editor. American Psychiatric Press, Washington, DC, 1994, p 189.

McGlashan TH, Fenton WS: Subtype progression and pathophysiologic deterioration in the course of early manifest schizophrenia. Schizophr Bull *19:*71, 1993.

*Meehl PE: Toward an integrated theory of schizotaxia, schizotypy, and schizophrenia. J Pers Disord *4:*1, 1990.

Mesulam MM: Large-scale neurocognitive networks and distributed processing for attention, language and memory. Ann Neurol *28:*567, 1990.

Pao P-N: *Schizophrenic Disorders*. International Universities Press, New York, 1979.

Rado S: *Psychoanalysis of Behavior*. Grune & Stratton, New York, 1956.

Rumelhart DE, McClelland JL: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol 1. MIT Press, Cambridge, MA, 1986.

Segal H: *Introduction to the Work of Melanie Klein*. Basic Books, New York, 1973.

*Spring B, Zubin J: Vulnerability to schizophrenic episodes and their prevention in adults. In *Primary Prevention in Psychopathology: The Issues*, vol 1, GW Albee, JM Joffee, editors. University Press of New England, Hanover, NH, 1977.

Stern D: *The Interpersonal World of the Infant*. Basic Books, New York, 1985.

Sullivan HS: *Clinical Studies in Psychiatry*. Norton, New York, 1956.

Tienari P, Sorry A, Lahti I, Narala M, Wahlberg K-E, Ronkko T, Pohjola J, Moring J: The Finnish adoptive family study of schizophrenia. Yale J Biol Med *58:*227, 1985.

Waelder R: The principle of multiple function. Psychoanal Q *5:*45, 1936.

*Willick MS. Psychoanalytic concepts of the etiology of severe mental illness. J Am Psychoanal Assoc *38:*1049, 1990.

Wynne LC, Singer M: Thought disorder and family relations of schizophrenics: II. Classification of forms of thinking. Arch Gen Psychiatry *9:*199, 1963.

12.7 SCHIZOPHRENIA: CLINICAL FEATURES

ROBERT CANCRO, M.D., MED.D.SC., AND HEINZ E. LEHMANN, M.D.

The major public health problem faced by psychiatry is the group of disorders that constitutes the diagnostic category of schizophrenia. These disorders affect approximately 1 percent of the population and most commonly have their onset in early adult life. More importantly, they usually leave the patient with varying degrees of cognitive, affective, and psychosocial impairment. This combination of impairments prevents most patients from achieving their full potential. In all known societies, adulthood is defined in terms of work and procreation. Individuals diagnosed with schizophrenia are frequently unable to perform the complex work tasks seen in the industrialized nations. The more severely impaired are not even able to perform the simpler work tasks associated with developing nations. Interpersonal relationships are frequently impaired enough to prevent courtship and subsequent marriage and procreation. The result is that people with these disorders are denied the social and personal benefits of adulthood to a considerable degree.

No objective criteria exist for the diagnosis of schizophrenia. No characteristic morphological changes in the brains of patients with schizophrenia have been demonstrated; no specific laboratory findings signal its presence; no consistent premorbid history, course, or outcome can be ascertained; and no single cause is known. However, a group of symptoms and signs are basic to and characteristic of schizophrenia; certain drugs, many with a common neurophysiological property, can often eliminate those symptoms almost selectively, and there is enough consensus about its diagnosis, treatment, and prognosis among experienced clinicians to warrant the acceptance of the concept of schizophrenia as a syndrome.

HISTORY

Emil Kraepelin Emil Kraepelin (Fig. 12.7-1) translated Benedict A. Morel's *demence precoce* to *dementia precox*, a term that emphasized the distinct cognitive process (dementia) and early onset (precox) of the disorder. Patients with dementia precox were described as having a long-term deteriorating course and the common clinical symptoms of hallucinations and delusions. Kraepelin distinguished these patients from those classified as having manic-bipolar psychosis who underwent distinct episodes of illness alternating with periods of normal functioning. The major symptoms of patients with paranoia were persistent persecutory delusions, and these patients were described as lacking the deteriorating course of dementia precox and the intermittent symptoms of manic-depressive psychosis. Although Kraepelin had acknowledged that about 4 percent of his patients recovered completely and 13 percent had significant remissions, later researchers sometimes mistakenly stated that he had considered dementia precox to have an inevitable deteriorating course.

FIGURE 12.7-1 Emil Kraepelin, 1856–1926. (Courtesy of National Library of Medicine, Bethesda, MD.)

Eugen Bleuler Eugen Bleuler (Fig. 12.7-2) coined the term *schizophrenia*, which replaced *dementia precox* in the literature. He chose the term to express the presence of schisms between thought, emotion, and behavior in patients with the disorder. Bleuler stressed that, unlike Kraepelin's concept of dementia precox, schizophrenia need not have a deteriorating course. Before the publication of the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), the incidence of schizophrenia increased in the United States (where psychiatrists followed Bleuler's principles) to perhaps as much as twice the incidence in Europe (where psychiatrists followed Kraepelin's principles). After publication of DSM-III, the diagnosis of schizophrenia in the United States moved toward Kraepelin's concept. Bleuler's term *schizophrenia*, however, has become the internationally accepted label for the disorder. This term is often misconstrued, especially by laypeople, to mean split personality. Split personality, now called *dissociative identity disorder*, is categorized in the fourth edition of DSM (DSM-IV) as a dissociative disorder and differs completely from schizophrenia.

FIGURE 12.7-2 Eugen Bleuler, 1857–1939. (Courtesy of National Library of Medicine, Bethesda, MD.)

The Four As Bleuler identified specific fundamental (or primary) symptoms of schizophrenia to develop his theory about the internal mental schisms of patients. These symptoms included associational disturbances (especially looseness), affective disturbances, autism, and ambivalence, summarized as the four As: associations, affect, autism and ambivalence. Bleuler also identified accessory (secondary) symptoms, which included those Kraepelin saw as major indicators of

dementia precox: hallucinations and delusions.

Manfred Bleuler Manfred Bleuler, Eugene Bleuler's son, stressed the essential psychotic nature of the schizophrenic disorders. He recognized that patients might not be psychotic at any given moment but considered that the diagnosis of schizophrenia should only be made when psychosis had been present at some time in the history of the illness. He stated that patients whose primary characteristics were impaired memory or intellectual function were not schizophrenic. Schizophrenia in his conceptualization tended to be chronic and did not show a tendency toward rapid recovery. He felt that a major element of the illness was the coexistence of both psychotic and normal mental processes in the same person at the same time. He emphasized that the mechanisms he described in patients with schizophrenia could be found to some degree in normal people as well. In his conceptualization what was specific for the schizophrenic disorders was the patient's inability to distinguish between the inner and outer realities, as if an individual could not distinguish between dreams and waking experiences. This inability allowed the simultaneous presence of incompatible components to coexist within the patient's personality and consciousness. Bleuler's conceptualization of etiology was that the genetic component was not necessarily pathological but that the particular constellation of genes in these individuals left them more vulnerable to stress.

Adolph Meyer Adolph Meyer emphasized the reactive nature of the schizophrenic disorders. He felt that the illness involved a biological predisposition of the individual and environmental stresses that interacted with it to produce the illness. In this Meyerian vision, neither the stress nor the diathesis alone sufficed to produce illness; vulnerable individuals had to face an adequate and appropriate stress at the right developmental phase of their life. Emphasizing the fortuitous element of life events in the etiopathogeneses of these illnesses led to a more optimistic view of the disorders and raised the possibility that eliminating certain stresses at certain developmental phases could markedly reduce the frequency of the illness. The reaction-pattern approach was carried to an extreme by some workers who believed that enough stress could produce a schizophrenic disorder in anyone. These workers unfortunately failed to recognize the importance placed by Meyer on the preexisting diathesis.

Other Theorists Harry Stack Sullivan, Ernst Kretschmer, Gabriel Langfeldt, Kurt Schneider, and Karl Jaspers added much to the understanding of schizophrenia. Sullivan, who founded the interpersonal psychoanalytic school, emphasized social isolation as a cause and a symptom of schizophrenia. Kretschmer compiled data to support the idea that schizophrenia occurred more often among people with asthenic, athletic, or dysplastic body types rather than among people with pyknic body types; the latter, he thought, were more likely to incur bipolar disorders. These observations may seem strange, but they are not inconsistent with a superficial impression of body types in many homeless persons.

Langfeldt classified patients with major psychotic symptoms into two groups, those with true schizophrenia and those with schizophrenic-like psychosis. In his description of *true schizophrenia*, Langfeldt stressed several factors: insidious onset, feelings of derealization and depersonalization, autism, and emotional blunting. Researchers after Langfeldt gave true schizophrenia other names: nuclear schizophrenia, process schizophrenia, and nonremitting schizophrenia.

Schneider described a number of first-rank symptoms of schizophrenia. They are not identical to Eugene Bleuler's fundamental signs nor do they mean the same thing. The Schneider first-rank symptoms of schizophrenia are not pathognomonic for the disease but are of great pragmatic value in making a diagnosis. When a number of those symptoms are present in a patient in whom other pathology—of toxic or organic origin—can be excluded, then a diagnosis of schizophrenia is indicated. Schneider's first-rank symptoms include hearing one's own thoughts spoken aloud, auditory hallucinations that comment on the patient's behavior, somatic hallucinations, the experience of having one's thoughts controlled, spreading one's thoughts to others, delusions, and the experience of having one's actions controlled or influenced from the outside.

Jaspers, a psychiatrist and philosopher, played a major role in developing existential psychoanalysis. In his view, psychopathology had no fixed concepts or basic principles. Thus his theories of schizophrenia were free of traditional concepts like subject and object, cause and effect, and reality and fantasy, and his philosophic attitude led to an interest in the content of psychiatric patients' delusions. The emphasis on attempting to understand the phenomenology of the schizophrenic experience can be traced to Jaspers.

COMPARATIVE NOSOLOGY

DSM DSM-III and the revised third edition of DSM (DSM-III-R) were major efforts to increase reliability over the earlier versions. It was recognized that independent validation was not yet attainable and therefore the goal of reliability was seen to be of great importance. That emphasis was also reflected in the development of DSM-IV. DSM-IV requires the presence of at least two characteristic symptoms for a significant portion of time during a 1-month period (or less if the patient responded successfully to treatment). The list of characteristic symptoms includes delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and negative symptoms (e.g., anhedonia, mutism). One symptom suffices for a diagnosis if that one symptom consists of bizarre delusions, hallucinations of a voice sustaining a running commentary on the person's behavior or thoughts, or hallucinations of two or more voices conversing with each other.

DSM-IV requires social and occupational deterioration. There must be a diminution in functional level in both social and occupational activities following the onset of illness. These functional disturbances must last for at least 6 months. That 6-month period may include the 1 month of symptoms necessary to fulfill the requirement of characteristic symptoms plus prodromal or residual symptoms or both. Finally, DSM-IV requires diagnostic exclusion of mood disorders with psychotic features and schizoaffective disorder. The disorder may not be a consequence of substance abuse or a general medical disorder.

ICD The definition of schizophrenia in the tenth revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) parallels that in DSM-IV.

> The schizophrenic disorders are characterized in general by fundamental and characteristic distortions of thinking and perception, and affects that are inappropriate or blunted. Clear consciousness and intellectual capacity are usually maintained, although certain cognitive deficits may evolve in the course of time. The most important psychopathological phenomena include thought echo; thought insertion or withdrawal; thought broadcasting; delusional perception and delusions of control; influence or passivity; hallucinatory voices commenting on or discussing the patient in the third person; thought disorders and negative symptoms.
>
> The course of schizophrenic disorders can be either continuous or episodic with progressive or stable deficits, or there can be one or more episodes with complete or incomplete remission. The diagnosis of schizophrenia should not be made in the presence of extensive depressive or manic symptoms unless it is clear that schizophrenic symptoms antedate the affective disturbance. Nor should schizophrenia be diagnosed in the presence of overt brain disease or during states of drug intoxication or withdrawal.

DIAGNOSIS

The DSM-IV diagnostic criteria include course specifiers that offer clinicians several options and describe actual clinical situations (Table 12.7-1). The presence of hallucinations or delusions is not necessary for the diagnosis of schizophrenia; a patient's disorder is diagnosed as schizophrenia when the patient exhibits two of the symptoms listed as symptoms 3 through 5 in criterion A. Criterion B requires that impaired functioning, although not deteriorations, be present during the active phase of the illness. DSM-IV stipulates that symptoms must persist for at least 6 months and that a diagnosis of schizoaffective disorder or mood disorder must be absent. ICD-10 lists certain symptoms as the general criteria for all forms of schizophrenia with the exception of simple schizophrenia. At least one of the following must be present: (1) thought echo, thought insertion or withdrawal, or thought broadcasting; or (2) delusions of control, influence or passivity, clearly referred to body or limb movements or specific thoughts, actions or sensations, delusional perception; (3) hallucinatory voices giving a running commentary on the patient's behavior or discussing the patient between themselves, or other types of hallucinatory voices coming from some part of the body; and (4) persistent delusions of other kinds that are culturally inappropriate and completely impossible, (e.g., being able to control the weather or being in communication with aliens from another world).

Table 12.7-1 DSM-IV Diagnostic Criteria for Schizophrenia

The diagnosis can also be made if at least two of the following are present: (1) persistent hallucinations in any modality, when occurring every day for at least one month, when accompanied by delusions (which may be fleeting or half-formed) without clear affective content, or when accompanied by persistent overvalued ideas; (2) neologisms, breaks or interpolations in the train of thought resulting in incoherence or irrelevant speech; (3) catatonic behavior, such as excitement, posturing or waxy flexibility, negativism, mutism and stupor; and (4) negative symptoms, such as marked apathy, paucity of speech, and blunting and incongruity of emotional responses (it must be clear that these are not due to depression or to antipsychotic medication).

The ICD-10 criteria, used throughout the world, are listed in Table 12.7-2.

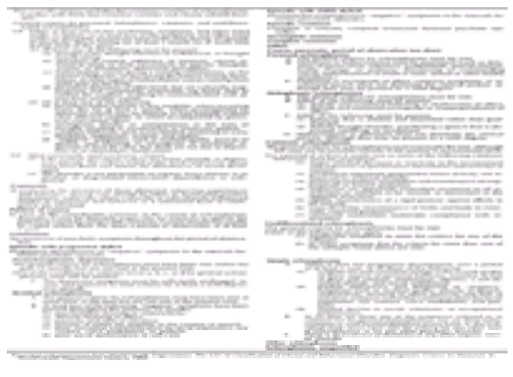

Table 12.7-2 ICD-10 Diagnostic Criteria for Schizophrenia

Key Symptoms The presence of some key symptoms for schizophrenia (e.g., blunting of emotional response or a strikingly inappropriate emotional response) weighs heavily in favor of a diagnosis of schizophrenia. But what is emotional blunting, and what is an inappropriate emotional reaction? For example, is the embarrassed adolescent's sheepish or defying smile an inappropriate emotional reaction? Considerable clinical experience is required to be certain about the presence of such symptoms.

Loosening of Associations The loosening of associations—the specific thought disorder of the schizophrenic—is perhaps one of the most valuable diagnostic criteria, but a good knowledge of psychopathology is required to be sure of its presence and to avoid confusing it with other forms of disturbed thinking, such as manic flight of ideas, disintegration of thought processes due to clouding of consciousness, and impaired reasoning due to fatigue or distraction. It is not sufficient to ask a patient the meaning of a proverb and then, on the basis of one's personal impression, declare that the patient has a pronounced schizophrenic thinking disturbance. It is sometimes impossible to distinguish, on the basis of a proverb test, between the disordered thinking of a schizophrenic and a manic patient, except for the greater verbosity of the manic.

Bizarre Behavior The patient's behavior may furnish significant clues for the diagnosis of schizophrenia. Bizarre postures and grimacing are certainly characteristic of schizophrenic conditions, but identifying a bizarre posture is not always easy. Religious rituals and special positions for meditation or dancing with which the observer is not familiar may be called bizarre. But in a recent case of a withdrawn, suicidal young girl, a possible diagnosis of depression was ruled out in favor of schizophrenia when the girl began eating raw chicken, pouring hot tea over herself, and openly trying to get into bed with her brother-in-law during a weekend home visit.

True catalepsy may be almost pathognomonic of schizophrenia, but it is not a common symptom. A stupor strongly suggests catatonic schizophrenia, but hysteria or a depressive stupor must be carefully ruled out in the differential diagnosis.

The deterioration of social habits, even involving the smearing of feces, does not suffice for the diagnosis of schizophrenia. Such deterioration can occur in various toxic and organic psychoses, temporarily in hysterical twilight states, and even at the peak of a manic episode in bipolar I disorder.

Pronounced social withdrawal also occurs under many conditions, ranging from simple sulking to anxiety and depression (Fig. 12.7-3). Sustained passivity and lack of spontaneity should suggest the diagnosis of schizophrenia only if organic and depressive conditions can be definitely ruled out.

FIGURE 12.7-3 “Schizophrenic Withdrawal.” (Courtesy of Sid Bernstein, Research Facility, Orangeburg, NY.)

Stereotypes and verbigeration strongly suggest schizophrenia, but they occur almost exclusively in chronic, institutionalized patients and are rarely seen today. Frequent and lengthy staring into a mirror and other odd mannerisms also strongly suggest a diagnosis of schizophrenia.

SUBTYPES

DSM-IV classifies the subtypes of schizophrenia as paranoid, disorganized, catatonic, and undifferentied, and residual, predominantly on the basis of clinical presentation (Table 12.7-3). These subtypes are not closely correlated with different prognoses; for such differentiation, specific predictors of prognosis are best consulted. ICD-10, by contrast, uses nine subtypes: paranoid schizophrenia, hebephrenia, catatonic schizophrenia, undifferentiated schizophrenia, simple schizophrenia, other schizophrenia, residual schizophrenia, post schizophrenic depression and schizophrenia, unspecified, with eight possibilities for classifying the course of the disorder, ranging from continuous to complete remission.

Table 12.7-3 DSM-IV Diagnostic Criteria for Schizophrenia Subtypes

Catatonic Type The catatonic type of schizophrenia is dominated by prominent psychomotor disturbances. In addition to meeting the general criteria for schizophrenia, there must be a period of least 2 weeks of catatonic behavior that can be either stuporous or excited. The behavioral disturbance can also involve posturing, negativism, rigidity, waxy flexibility, or command automatism.

Catatonic schizophrenia occurs in two forms: inhibited or stuporous catatonia and excited catatonia. The essential feature of both forms is the marked abnormality of motor behavior.

Stuporous Catatonia Patients with stuporous catatonia may be in a state of complete stupor or may show a pronounced decrease in spontaneous movements and activity. They may be mute or nearly so or may show distinct negativism, stereotypes, echopraxia, or automatic obedience. However, even after standing or sitting motionless for long periods of time, they may suddenly and without provocation have a brief outburst of destructive, unprovoked violence. Occasionally, patients with catatonic schizophrenia exhibit catalepsy or waxy flexibility (Fig. 12.7-4).

FIGURE 12.7-4 Chronic catatonic patient. This patient is immobile, demonstrating waxy flexibility. Her arm is in an uncomfortable position, elevated without support, and her stony facial expression has a *Schnauzkrampf*, or frozen pout. (Courtesy of Heinz E. Lehmann.)

Patients in a state of complete catatonic stupor usually can be aroused from it dramatically by intravenous injection of a short-acting barbiturate. Within minutes the frozen facial expression changes to one of normal animation. They begin to talk and move about normally and in many instances become relatively lucid for an hour or two. The total inhibition of patients in catatonic stupor may result from excessive cerebral excitatory processes. As in a car engine racing so wildly that one cannot put it in gear, that excess prevents the person from performing and behaving in a normal manner until the excessive cerebral functions have been reduced by a chemical agent that depresses brain metabolism and nervous impulses. But, spectacular as the immediate results seem to be, the technique has no significant therapeutic value.

A young, unmarried woman, age 20, was admitted to a psychiatric hospital because she had become violent toward her parents, had been observed gazing into space with a rapt expression, and had been talking to invisible persons. She had been seen to strike odd postures. Her speech had become incoherent.

She had been a good student in high school, then went to business school and, a year before admission to the hospital started to work in an office as a stenographer. She had always been shy, and although she was quite attractive, she had not been dating much. Another girl, who worked in the same office, told the patient about boys and petting and began to exert a great deal of influence over her. The second girl would communicate with her from across the room. Even when they went home at night, the patient would get voice messages telling her to do certain things. Then pictures began to appear on the wall, most of them ugly and sneering. Those pictures had names—one was named shyness, another distress, another envy. Her office friend sent her messages to knock on the wall, to hit the pictures.

The patient was agitated, noisy, and uncooperative in the hospital for several weeks after she arrived, and required sedation. She was given a course of insulin coma therapy, with no significant or sustained improvement. Later she received several courses of electroconvulsive treatment, which also failed to influence the schizophrenic process to any significant degree. Ten years later, when antipsychotic drugs became available, she received pharmacotherapy.

Despite all those therapeutic efforts, her condition throughout her many years of stay in a mental hospital has remained one of chronic catatonic stupor. She is mute and practically devoid of any spontaneity, but she responds to simple requests. She stays in the same position for hours or sits curled-up in a chair. Her facial expression is fixed and stony.

Excited Catatonia Patients with excited catatonia are in a state of extreme psychomotor agitation. They talk and shout almost continuously. Their verbal productions are often incoherent, and their behavior seems to be influenced more by inner stimuli than by their environment. Patients in catatonic excitement urgently require physical and medical control, since they are often destructive and violent to others, and their dangerous excitement can cause them to injure themselves or collapse from complete exhaustion.

An unmarried man, aged 27, had been working as a teacher and was admitted to a psychiatric hospital because he had become increasingly agitated and irrational after several nights of wakefulness. He was extremely talkative and ran about aimlessly. His behavior became very strange; for instance, he tried to clean everything in the house, moved his wristwatch up to his shoulder, stripped his clothes off, chewed large wads of paper in the belief that it was good for him, talked about killing himself, and then said that he might already be dead.

He heard voices ordering him about incessantly, and he frequently laughed without any apparent cause. After chewing the paper he would spit in it and then drink his saliva. He rolled into odd postures on the bed, with his tongue sticking out. He started to jump and dance when taken to the bathroom by a nursing assistant for a shower and destroyed the bathroom furnishings. His gait was manneristic. His speech was utterly incomprehensible. He refused to take any medication and had to be sedated by parenteral medication.

He remained noisy, excited, destructive, and irrational in his behavior for a month; then he improved in response to high dosages of antipsychotic medication and a few electroconvulsive treatments. Three months after admission he was discharged from the hospital, symptom free, with good insight into the nature of his illness. For more than 10 years he has been employed as a teacher.

Periodic Catatonia A rare but intriguing form of catatonia, periodic catatonia, was described by R. Gjessing in 1938. According to Gjessing, patients affected with the disorder have periodic recurrences of stuporous or excited catatonic states. Each recurrence of catatonic behavior is associated with an extreme shift in the patient's metabolic nitrogen balance. Most cases of periodic catatonia seen in recent years responded well to antipsychotic medication, and relapses were usually prevented by maintenance medication.

Disorganized Type The disorganized (formerly called hebephrenic) subtype is characterized by a marked regression to primitive, disinhibited, and unorganized behavior. Hebephrenic patients are usually active but in an aimless, nonconstructive manner. Their thought disorder is pronounced, and contact with reality is extremely poor. Personal appearance and social behavior are dilapidated, emotional responses are inappropriate, and they often burst out laughing for no apparent reason. Incongruous grinning and grimacing are common in these patients, whose behavior is best described as silly or fatuous (Fig. 12.7-5).

FIGURE 12.7-5 A 44-year-old chronic schizophrenic woman showing characteristic mannerism and facial grimacing. (Courtesy of New York Academy of Medicine.)

A 15-year-old girl attended a summer camp where she had difficulties in getting along with the other children and developed animosity toward one of the counselors. On her return home, she refused to listen to her parents, and she heard the voice of a man talking to her, although she could not see him. She rapidly began to show bizarre behavior, characterized by grimacing, violent outbursts, and inability to take care of herself.

Her school record had always been good, and she was fluent in three languages. Her parents described her as having been a quiet, rather shut-in child, with no abnormal traits in childhood. Family relations were reported as having been satisfactory.

When the patient was admitted to a psychiatric hospital, her speech was incoherent. She showed marked disturbances of formal thinking and blocking of thoughts. She was impulsive and seemed to be hallucinating. She stated that she heard voices in her right ear and that a popular singer was running after her with a knife. She also thought that her father was intent on killing her and that she was pregnant because she had hugged one of the residents.

Two months of neuroleptic treatment brought no apparent improvement. She was then given a course of intensive electroconvulsive therapy and continuous sleep treatment. Over a period of a year, she received close to 200 electroconvulsive treatments and 50 subcoma insulin treatments, with little improvement. She was then transferred to another mental hospital, where her behavior has remained very disturbed for almost 20 years.

She is often incontinent and most of the time neglects her physical appearance. Occasionally she spends hours dressing herself, looking in the mirror, and putting on excessive makeup. At times, she has been discovered eating her feces. Occasionally, she adopts the role of a singer or a dancer. She makes statements like “Will I live forever? Nurse, I didn't throw my love away. It is in my stomach, and it hurts.” In the dining room she attempts to grasp the genitals of male patients. High doses of neuroleptics are continuously required to control her behavior. The ultimate prognosis is very poor.

Paranoid Type The paranoid type of schizophrenia is characterized mainly by the presence of delusions of persecution or grandeur. Patients with paranoid schizophrenia are usually older than patients with catatonia or hebephrenia when they break down (i.e., they are usually in their late 20s or their 30s). Patients who have been well up to that age have usually established a place and an identity for themselves in the community. Their ego resources are greater than those of catatonic and hebephrenic patients. Paranoid schizophrenics show less regression of mental faculties, emotional response, and behavior than those with the other subtypes of schizophrenia.

Typical patients with paranoid schizophrenia are tense, suspicious, guarded, and reserved. They are often hostile and aggressive. They usually conduct themselves quite well socially, and their intelligence in areas not invaded by delusions may remain high. Paul Murphy, an American chess champion in the first half of the nineteenth century and one of the greatest chess masters in history, developed paranoid schizophrenia in his middle 20s and was hospitalized for years. But even many years after he had become ill, he played an original and masterful game of chess if he could be persuaded to accept the challenge.

A woman wrote the letter that is reproduced in part below. Like many patients with schizophrenia, this patient is tortured by the experience of being influenced in her bodily functions through fiendish devices used by her enemies.

Dear Dr. P.T.,

It is with the nurse's knowledge that I write this letter to you, regretting at the same to trouble you about a maladjustment that need not occur. I am at a loss to understand why those who are responsible are permitted to indulge in this peculiar pastime. Perhaps those with some authority do not desire to check it, but I realize, in any case, that it is rather difficult to do so inasmuch as when one person is checked, she passes the job on to someone else—I refer to this instrument that they use that completely locks up the intestines and prevents them from elimination at all even with a laxative, which is useless to take in such circumstances. It also twists me between my legs occasioning much discomfort, preventing proper rest in bed when my body is so tightened up that it is impossible to relax. I spoke to you several days ago about unexplainable solutions being injected into my body and my rest disturbed continually during the night. You agreed that if any treatments were ordered they should not take place at these hours. I think you said you would find out about any such requirement or order. Well, the hour of interference has changed—interference is the right word to use, because the one whoever is responsible is doing a great deal of harm. I am awakened every morning to the exact minute about 5:30. About 6:00 I drop off again until 7 AM (breakfast not until 8 AM). During this interval some person interferes with both passages of my body, and I find myself going around the rest of the day like a tank—full of burning Salt Water but quite unable to eliminate. At 5:30 I am quite comfortable but if I rose at this time, I suppose something would take place in the washroom—as these things so often do. The bladder is also interfered with, as is indeed every organ in my body with which I have never had any previous trouble. This has taken place nearly every day this week. If they do not get a chance just before I retire, some person in my room, by arrangement I suppose, awaits for me to get out of bed and as soon as I turn my back to them quickly uses this instrument, so as to insure this locked-up condition of the abdominal region. Because of so much interference, it is sometimes necessary for me to encourage the bladder to completely empty by means of the application of heat in the form of towels rung out in hot water and this is only effective in a standing up position—there is obviously some solution injected and as soon as I have withdrawn it in this manner I am very comfortable, but am not allowed to remain that way long. Something has also taken place this week which has been done once before on this ward and once in West House where it was frowned upon after discovery. I have been fixed up—temporarily, I hope, if the above results are going to continue—with two separate outlets for urination, which almost appears to come through two holes in the pelvic bone. All these months, as the previous nurses know, I have had to endure endless damage by interference with the pelvic. Perhaps they have got tired of that at last—and now it has to be something else, and when they get tired of that what next? . . .. There are other means of displacing the intestines, not only by instruments in the hands of other inmates, but by other mechanical tricks probably operated by the same people.

Undifferentiated Type Frequently, patients who clearly have schizophrenia do not fit easily into one of the other types. DSM-IV classifies those patients as having undifferentiated type.

Residual Type According to DSM-IV, the residual type of schizophrenia is characterized by the presence of continuing evidence of the schizophrenic disturbances in the absence of a complete set of active symptoms or of sufficient symptoms to meet the diagnosis of another type of schizophrenia. Emotional blunting, social withdrawal, eccentric behavior, illogical thinking, and mild loosening of associations commonly appear in the residual type. When delusions or hallucinations occur, they are neither prominent nor accompanied by strong affect.

OTHER TYPES AND RELATED SYNDROMES

Simple Deteriorative Disorder Simple deteriorative disorder (simple schizophrenia) is characterized by a gradual, insidious loss of drive, interest, ambition, and initiative (Table 12.7-4). Hallucinations and delusions are uncommon, and if those symptoms do occur, they do not persist. Patients with simple deteriorative disorder withdraw from contact with other people, tend to stay in their rooms, avoid meeting or eating with other members of the family, stop working, and stop seeing friends. If they are still in school, their marks drop to a low level, even if they were consistently high in the past.

A. Progressive development over a period of at least a year of all of the following:
   (1) marked decline in occupational or academic functioning
   (2) gradual appearance and deepening of negative symptoms such as affective flattening, alogia, and avolition
   (3) poor interpersonal rapport, social isolation, or social withdrawal
B. Criterion A for schizophrenia has never been met.
C. The symptoms are not better accounted for by schizotypal or schizoid personality disorder, a psychotic disorder, a mood disorder, an anxiety disorder, a dementia, or mental retardation and are not due to the direct physiological effects of a substance or a general medical condition.

Table 12.7-4 DSM-IV Research Criteria for Simple Deteriorative Disorder (Simple Schizophrenia)

These patients avoid going out into the street during the day but may go for long walks alone at 2:00 or 3:00 AM. They tend to sleep until noon or later, after staying up alone most of the night. During the early stages of the illness they may have many somatic complaints, variously described as fatigue, nervousness, neurosis, psychosomatic disease, and laziness. Patients are often treated for a year or more before the correct diagnosis is made. In many cases patients with simple deteriorative disorder later become homeless. They become increasingly shallow in their emotional responses and are quite content to drift aimlessly through life as long as they are left alone.

Although patients appear to be indifferent to their environment, they may react with sudden rage to persistent nagging by family members. The immediate reason for admission of patients with simple schizophrenia to a hospital is often an outburst of violence directed against their mothers or fathers for a trivial reason.

An unmarried man, 27 years old, was brought to a mental hospital because he had on several occasions become violent toward his father. For a few weeks he had hallucinations and heard voices. The voices eventually ceased, but he then adopted a strange way of life. He would sit up all night, sleep all day, and become very angry when his father tried to get him out of bed. He did not shave or wash for weeks, smoked continuously, ate very irregularly, and drank enormous quantities of tea.

In the hospital he adjusted rapidly to the new environment and was generally cooperative. He showed no marked abnormalities of mental state or behavior, except for his lack of concern about most things. He kept to himself as much as possible and conversed little with patients and staff. His personal hygiene had to be supervised by the nursing staff; otherwise he would quickly become dirty and very untidy.

Twenty years after his admission to the hospital, he is described as shiftless, and careless, sullen, and unreasonable. He lies on a couch all day. Antipsychotic drug treatment has failed to alter his mental state or behavior. Although many efforts have been made to get him to accept therapeutic work assignments, he refuses to consider any kind of regular occupation. In the summer he wanders about the hospital grounds or lies under a tree. In the winter he wanders through the tunnels connecting the various hospital buildings and is often seen stretched out for hours under the warm pipes that carry steam through the tunnels.

Patients with simple deteriorative disorder may resemble personalities of the schizoid type. The distinguishing feature is the disorder makes its appearance at some time during or after puberty and from then on goes on to definite deterioration; personality deviations usually start earlier and remain the same over the years.

To meet the ICD-10 diagnostic criteria for simple schizophrenia, the individual must show over a period of at least 1 year all of the following manifestations: (1) a significant and consistent change in the overall quality of some aspect of personal behavior such as loss of drive and interest; (2) gradual appearance and deepening of negative symptoms such as marked apathy; and (3) a marked decline in social, scholastic, or occupational performance.

Postpsychotic Depressive Disorder of Schizophrenia The clinical boundaries of the diagnosis are hard to define operationally. The symptoms of postpsychotic depressive disorder of schizophrenia can closely resemble the symptoms of the residual phase of schizophrenia as well as the side effects of commonly used

antipsychotic medications. Distinguishing the diagnosis from schizoaffective disorder, depressive type, is also difficult. The DSM-IV criteria (see Table 14.6-25) specify that the criteria for a major depressive episode be met and that the symptoms occur only during the residual phase of schizophrenia. The symptoms cannot be substance induced or part of a mood disorder due to a general medical condition.

ICD-10 describes a category called postschizophrenic depression. This is a depressive episode that may be prolonged, arising in the aftermath of a schizophrenic illness. The general criteria for schizophrenia must be met, and the depressive symptoms must be sufficiently prolonged or severe to meet the criteria for at least a mild depressive episode. These depressive states are associated with an increased risk of suicide.

Early-Onset Schizophrenia Most patients suffering from schizophrenia develop it in late adolescence and early adulthood. A small minority of patients manifest a similar syndrome in childhood. Such children may at first present diagnostic problems, particularly with differentiation from mental retardation and autistic disorder. Recent studies have established that the diagnosis of childhood schizophrenia may be based on the same symptoms used for adult schizophrenia. What characterizes childhood schizophrenia is not the nature but the dramatic intensity of its symptoms. Its onset is usually insidious, its course tends to be chronic, and the prognosis is mostly unfavorable. Briefly, it resembles the typical Kraepelinian case of dementia precox. What gives childhood schizophrenia unique importance for research is the observation that anatomical features of the brain that are often associated with adult-onset schizophrenia (e.g., enlarged ventricles) are also present in this early-onset form of the disease. Neurobiological studies of children with schizophrenia may therefore provide significant clues to the developmental pathogenesis of adult-onset schizophrenia.

Late-Onset Schizophrenia Late-onset schizophrenia is usually defined as an illness that is clinically indistinguishable from schizophrenia but has an onset after age 45. Since DSM-IV no longer uses an age cutoff, this distinction may no longer be relevant. This condition tends to appear more frequently in women and also tends to be characterized by a predominance of paranoid symptoms. The prognosis is favorable, and these patients usually do well on antipsychotic medication.

Bouffée Délirante (Acute Delusional Psychosis) Since it has been recognized that tardive dyskinesia may develop as a serious complication in a significant minority of patients with schizophrenia on maintenance therapy with antipsychotic drugs, it has become even more important to make a correct initial diagnosis. In French psychiatry the condition known as acute delusional psychosis, or *bouffée délirante*, is not included in the diagnosis of schizophrenia. Rather, *bouffée délirante* is considered a disease entity in its own right, a psychiatric disorder that does not require maintenance pharmacotherapy.

The following are essential criteria for the diagnosis of *bouffée délirante*, not all of which need to be present: (1) frequent background of personality disorder; (2) absence of a schizoid premorbid personality; (3) sudden onset; (4) duration of less than 3 months; (5) spontaneous return to premorbid level of adjustment, even without specific antipsychotic treatment; (6) polymorphous symptoms, a disorderly (kaleidoscopic) succession of differing delusional contents; (7) a fascinating intensity of the delusional experience; (8) oscillations between insight and delusion; (9) mood alterations and fluctuations; (10) increase in delusions in sleep-related states; and (11) sudden termination after days or weeks (rarely months).

Accordingly, *bouffée délirante* episodes belong to the schizophrenic spectrum disorders. In American clinical practice those episodes are usually diagnosed as schizophreniform disorder (see Table 13.1-3, schizoaffective disorder (see Table 13.1-1), or brief reactive psychosis (see Table 13.1-4). French psychiatrists report that about 40 percent of patients with the diagnosis *bouffée délirante* are later reclassified as suffering from schizophrenia.

Schizoaffective Disorder Schizoaffective disorder appears to lie conceptually between schizophrenia and the mood disorders. This category is quite ambiguous because it is the middle ground between two almost arbitrary groupings of patients. DSM-IV requires concurrent symptoms that meet the major criteria for schizophrenia and for a major depressive episode or manic episode or mixed episode. There must be a period of illness of at least 2 weeks, with characteristic delusions and hallucinations and the absence of prominent mood symptoms. Patients who carry this label tend to have a better prognosis.

Oneiroid In the oneiroid state patients feel and behave as though they were in a dream. (*Oneiros* is the Greek word for dream). Patients may be deeply perplexed and not fully oriented in time and place. During the state of clouded consciousness, they may experience feelings of ecstasy and rapidly shifting hallucinated scenes. Illusionary distortions of their perceptional processes (including time perception) and the symptomatic picture may resemble those of a hysterical twilight state. During oneiroid reactions, the observer can most clearly detect the patient's peculiar “double bookkeeping”—patients may be convinced that they are traveling through space on a satellite and at the same time conscientiously follow the regular hospital routine. The patient with oneiroid schizophrenia acknowledges everyday realities but gives priority to contingencies of reality (Fig. 12.7-6). Oneiroid states are usually limited in duration and occur most frequently in acute schizophrenic episodes.

FIGURE 12.7-6 A 25-year-old schizophrenic man produced this eerie-looking mixture of commercial poster and existential quandary about time. (Courtesy of Heinz E. Lehmann.)

OTHER DIAGNOSTIC CRITERIA

A variety of research clinicians, some of whom were mentioned above (e.g., Langfeldt, Schneider, and Jaspers), constructed their own criteria to discriminate for essential features of schizophrenia. However, a number of other diagnostic systems for schizophrenia have been developed (Table 12.7-5).

Table 12.7-5 Essential Features of Various Diagnostic Criteria for Schizophrenia

SIGNS AND SYMPTOMS

General Appearance and Behavior While no specific behaviors or appearances are unique to schizophrenia, experienced clinicians still speak of the precox feeling

(i.e., a failure to emotionally contact with the patient). Schizophrenia patients often give a history of being more sensitive than the average person. This sensitivity involves not only increased responsiveness to sensory stimuli but also increased sensitivity to emotional stimuli and, in particular, critical experiences. Experimental evidence indicates that some individuals who later develop schizophrenia do not screen out stimuli as effectively as normal people, which allows excessive input of stimuli to the nervous system. The failure of selective inhibition can play an important role in symptom production.

Patients with chronic schizophrenia tend to show a neglected personal appearance. Their efforts at grooming tend to be minimal. They show poor regard for the social amenities and act as if they are deliberately turning away from society. As a group, schizophrenia patients are characterized by social withdrawal. They may form emotional attachments with other people, but they cannot communicate the quality and strength of those attachments in a manner that is understandable to other people. Unfortunately, this tends to create a lack of empathy or sympathy toward the patient, which further isolates the patient from family and health care providers.

A common feature in schizophrenia is the loss of ego boundaries. These patients have difficulty determining where they end and where the outside world begins. This leaves them vulnerable to misinterpretation of external events that can be interpreted as affecting them directly.

Many patients show what appears to be an amotivational syndrome. The patient may show a lack of interest in the normal activities of daily living. Nevertheless, a loss of motivation should not be confused with the sudden or gradual intellectual dysfunction that can occur in this disorder. This intellectual dysfunction can lead to failure in school of a young person of good intelligence, which can be misinterpreted as not trying hard enough. This failure may be the earliest diagnostic sign of a developing schizophrenia.

Speech Disorders What has been historically referred to as thought disorder is more correctly identified as a speech disorder. It is assumed that the disorders of language reflect an underlying disorder of thinking. A variety of features have been reported by clinicians for the last 100 years as characteristic of this syndrome. These include the loss of the logical relations between antecedent and subsequent associations that is termed *loosening of associations*. Words can be combined on the basis of sound rather than on meaning called *clang association*. New words may be generated, which are called *neologisms* (Fig. 12.7-7). *Verbigeration* involves the use of words in a stereotypically repetitive fashion. *Echolalia* involves the repetition of the examiner's words. *Thought blocking* involves the sudden and inexplicable blocking of thoughts manifested by the patient's inability to speak.

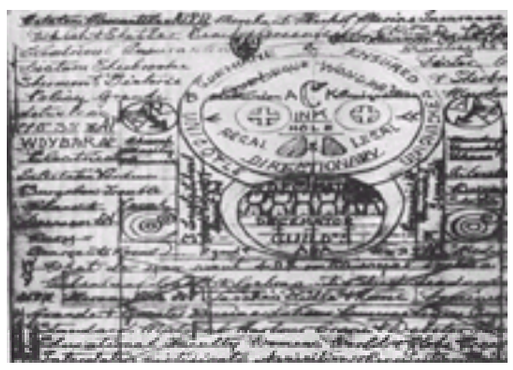

FIGURE 12.7-7 A schizophrenic woman expresses her incoherent thinking combined with neologisms in this drawing. (Courtesy of Heinz E. Lehmann.)

Loosening of associations is based on the late nineteenth-century association theory. According to association theory, language is determined by purpose. This purposefulness is often lost in schizophrenic speech. A sentence completion test illustrates the point. The sentence to be completed was “The man fell on the street . . ..” The patient's response was “because of World War I.” Although the thought of falling might be associated with falling in combat, it was an inappropriate association for the stimulus.

It can be helpful to look at disorders of association as disorders of the word and disorders of the sentence. Disorders of the word range from loss of symbolic meaning of the word as in clang associations to inability to maintain the correct semantic context for a word to approximate use of words, to the creation of new words. Disorders of the sentence include associative failures and failures of system placement. Most words have multiple meanings. Even a simple question such as “Where is your husband?” must be answered in terms of the frame of reference. In one context the question might ask for the physical location of the husband, and in another context it might ask for his identification in his graduating class picture. An example of system shifting was reported by Silvano Arieti. Commenting on the Japanese attack on Pearl Harbor, a patient said, “The next time they may attack Diamond Harbor or Emerald Harbor.” The patient had lost the contextual system of Pearl Harbor as a geographical military base and had substituted a contextual system in which pearls are precious stones.

Incoherence Language appears to be a means of self-expression in schizophrenia rather than a means of communication. Verbal productions are often empty or obscure. Schizophrenic speech contains more words that do not belong than does normal speech. In speech samples, patients with schizophrenia tend to repeat the same words more frequently than do nonschizophrenic people. When normal individuals are asked to fill in the deleted word in a continuous passage they are more likely to be correct when judging normal speech than when judging schizophrenic speech.

The following proclamation was written by a woman with schizophrenia. The repetitive phrases, distorted syntax, and numerous non sequiturs render the text sometimes incoherent. Nevertheless, the paranoid grandiosity, the hostility of the writer, and the content of her delusions are clearly expressed.

> The French Force orders from now on to the German Force to respect the Queen Sacre in Christianity as well as the Queen in France and in other countries, ill treated and destroyed in all countries since the beginning of this century in Europe and allied countries. The Queens are the copartners in masonry of the order of Grand Masters and by doing so the prosperity and balance of the world have been destroyed, they have been destroyed for homosexuality which is the emblem or grand mastery really instead of being distinguished from the criminals who kill the soul and commit the crime of homosexuality of destroying the emblem of grand mastery. The attack on the Queen Sacre in Masonry comes from an inversion of data in the German spying service in 1903 in the class of sorcerers of this organization, deciding that the Chateau de Chambord en France was going to be the Castle not of the saint to be, but of sorceress and killing in soul that child many times without the effect desired obtained.

The following is an example of what may be called concise abstruseness, which sometimes characterizes the communications of patients with schizophrenia and is often used to express undisguished sexual preoccupation. This patient's short apologetic note is to a psychiatrist whom she had bluntly propositioned on frequent occasions. (In this instance she got revenge against him for his rejection by pulling up expensive flowers in his garden.) She had previously inserted a screwdriver into her vagina and later expressed continuing guilt for having done so. The note expresses her sexually laden message briefly and (in her way) to the point. But the letter can be deciphered only with difficulty by the nonschizophrenic person who has not learned to understand the patient's autistic language.

> Dear doctor,
>
> I wasn't thinking too well when I was speaking to you but I do believe you were the postman whom I spent the night with. It is still Dr. David . . . in my heart. Am sick because of the screw driver. Please no hard feelings. Kiss your penis did. I would not harm you . . ..

The following brief transcript from a videotaped interview with a young man with schizophrenia illustrates his autistic preoccupation with sex and death; there seems to be some clang association between “feet” and “foetus.” The patient was puzzled that his interviewer had difficulties following him.

> the fleur de Lys is a castrated ace—you see, the design is the feet—the same as a woman's foetus—now you take five French safes and you put them together between four coffins—that's what it represents

Neologisms Occasionally, patients with schizophrenia create a completely new expression, a neologism, when they need to express a concept for which no ordinary

word exists.

> A woman with schizophrenia who had been hospitalized for several years kept repeating (in an otherwise quite rational conversation) the word "polamolalittersjitterstittersleelitla." Her psychiatrist asked her to spell it out, and she proceeded to explain the meaning of the various components, which she insisted were to be used as one word. "Polamolalitters" was intended to recall the disease poliomyelitis, because the patient wanted to indicate that she felt she was suffering from a serious disease affecting her nervous system; the component "litters" stood for untidiness or messiness, the way she felt inside; "jitterstitters" reflected her inner nervousness and lack of ease; "leelita" was a reference to the French *le lit la* (that bed there), meaning that she both depended on and felt handicapped by her illness. That single neologistic production thus enabled the patient to express—in a condensed, autistic manner—information about her preoccupations and apprehensions that otherwise would have taken a whole paragraph to explain in common language.

Mutism Functional inhibition of speech and vocalization may last for hours or days, but before the use of modern treatment methods, it often used to last for years in patients with catatonic schizophrenia. Many of these patients tend to be monosyllabic and answer questions as briefly as possible. They attempt to restrict contact with the interviewer as much as possible without being altogether uncooperative.

Echolalia Occasionally, patients with schizophrenia exhibit echolalia, repeating in their answers to the interviewer's questions many of the same words the questioner has used.

> Examiner: How did you sleep last night?
>
> Patient: I slept well last night.
>
> Examiner: Can you tell me the name of your head nurse?"
>
> The name of my head nurse is Miss Brown.

Echolalia seems to signal two facts, patients are aware of some shortcomings in their ideation and they are striving to maintain active rapport with the interviewer. They act much like someone learning a new language who answers the teacher's questions with as many of the teacher's words in the strange language as they can possibly manage.

Verbigeration This rare symptom is found almost exclusively in chronic and very regressed patients with schizophrenia. It consists of the senseless repetition of the same words or phrases, and it may go on for days (Fig. 12.7-8). Like neologisms and echolalia, verbigeration is a rare symptom today and is almost restricted to long-term institutionalized schizophrenia patients. Many psychiatrists working with schizophrenia patients in the community may never encounter these manifestations of deterioration.

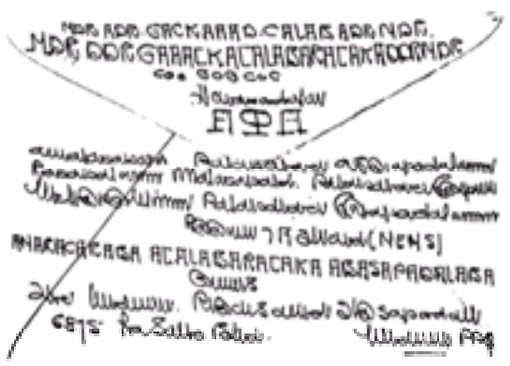

FIGURE 12.7-8 Sample of a chronic schizophrenic's noncommunicative writing. This addressed envelope illustrates manneristic writing, verbigeration, and possibly neologisms. Although the script appears to be exotic, the Arabic numerals and the English street name are recognizable. (Courtesy of Heinz E. Lehmann.)

Stilted Language Some patients with schizophrenia make extraordinary efforts to maintain their social relations, to maintain their relatively stable adjustment. But they may betray their rigidity and artificiality in their interpersonal relations by a peculiarly stilted and grotesquely quaint language.

The following excerpt from a letter written by a physician with schizophrenia who was hospitalized for more than 15 years but is now living by himself in an apartment is an example of such stilted language.

> My dear friend and Professor,
>
> A hearty and cheerful. (Please turn the page over) and a magnanimous good morning to you on this first Wednesday of a glorious New Year: And I do hope that our great and our good Lord and our dearly beloved and kind Shepherd, (kindly read page three now). Will be gracious unto both me and thee. I am sure that He be gracious unto both of us; He has some sound common sense. His being, this morning . . . I have not yet heard (Kindly turn over to 4 now) from any one of my own colleagues when I am leaving noble institution of the healing arts; Nor with whom: Nor through which one of the portals. Though I am sure that you—as much as (Kindly turn to page five, now) one else . . . must be able to enlighten me; very soon, my good old friend . . .

Behavioral Disorders Many patients with schizophrenia show a quantitative change in their activity, most commonly manifested as a reduction in energy, spontaneity, and initiative. In acute stages patients may become excited and show increased activity but usually only in the early phases of the illness. In a qualitative sense their behavior is often poorly coordinated, unpredictable, eccentric, and inappropriate. Even before the development of antipsychotic drugs patients conveyed an awkwardness and stiffness of movement. The great dancer Nijinsky lost his natural gracefulness with the onset of his schizophrenia.

Mannerisms Many patients with schizophrenia exhibit mannerisms of speech and movement. Grimacing is noticeable to varying degrees and at times may appear almost as a tic, particularly in the perioral regions. These perioral movements were reported and described by Kraepelin long before the use of neuroleptic drugs. He considered it a pathognomonic sign for poor prognosis in schizophrenia.

Stuporous States Stuporous states used to be common in the catatonic subtype of schizophrenia. Today they are quite rare and respond quite quickly to modern treatment when they are found. Similarly, catalepsy or waxy flexibility is almost unknown today, whereas 40 years ago it was common. It consists of a waxlike yielding of the movable parts of the body to any effort made to place them in certain positions. Once placed in the position, the patient remains in that position for a long time—even if the position is physically uncomfortable (Fig. 12.7-9). While these extreme examples of stupor and waxy flexibility have essentially disappeared, many chronic patients show a lack of spontaneity and movement that bears an attenuated resemblance to the more severe conditions described above.

FIGURE 12.7-9 A patient exhibiting catatonic posturing. (Reprinted with permission from Davison GC, Neale JM: *Abnormal Psychology: An Experimental Clinical Approach*. Wiley, New York, 1974.)

Echopraxia The motor symptom echopraxia is analogous to echolalia in the verbal sphere. It is the imitation of movements and gestures of the person the patient is observing.

Negativism Negativism refers to a patient's unwillingness to cooperate without any apparent reason for that lack of cooperation. It does not appear to be related to fatigue, depression, suspicion, or anger. Negativism may even take the form of unwillingness to follow a request for a physical movement. It can become so severe that the patient will do the opposite of what is asked. For example, when asked to raise an arm, they may lower it.

Stereotyped Behavior Stereotyped behavior is primarily seen in patients with chronic schizophrenia, including those in the community. At times it may take a motoric form and be expressed in a repetitive pattern of walking or pacing. It may also be demonstrated in repetitive strange gestures, which may or may not have a magical meaning to the patient. Finally, in language one can have the repetition of phrases or comments for long periods. This is separate from preservation and distinct from verbigeration. Interestingly, when schizophrenia patients are engaged psychosocially, this symptom tends to diminish. It appears to be a consequence of psychosocial isolation.

Deteriorated Appearance and Manners Patients with schizophrenia tend to neglect their appearance. This extends not only to elements of personal hygiene such as bathing, but even to changing their clothing. They can appear indifferent to the social amenities, such as returning a greeting. Frequently, they exhibit bad table manners despite the fact that in their premorbid condition they did not do so. Prior to the advent of antipsychotic drugs, regressed patients with schizophrenia could frequently be seen masturbating openly in hospital wards without apparent concern about the public nature of their act.

Affective Features Both quantitative and qualitative changes in affect may occur in patients with schizophrenia.

Reduced Emotional Responses Most commonly, patients show a quantitative change in the intensity of their emotional responses. Many patients with schizophrenia appear indifferent and apathetic. Others show diminished emotional intensity, described as emotional restriction or blunting. This quantitative aspect, which has been emphasized by Bleuler, is common in schizophrenia.

In judging emotional depth, one must consider the cultural background of the patient. A normal emotional expression in one culture may appear to be a reduced emotional response in another. Cultures differ dramatically in their willingness to accept a public display of emotion; therefore, the examiner should have some knowledge of the cultural background of the patient being assessed. Often the patient and family are the best informants; they may be able to describe changes from the premorbid emotional state before they become obvious to even an experienced clinician.

Anhedonia Anhedonia is a particularly distressing symptom. Sandor Rado considered anhedonia to be a cardinal feature of schizophrenia. There is frequently a diminution in the patient's ability to experience pleasure and, in some severe cases, even to imagine a pleasant feeling. Patients may not meet the criteria for the diagnosis of clinical depression but will describe an emotional emptiness or barrenness. Anhedonia can become unbearable enough to contribute to a suicide attempt.

Inappropriate Responses A common finding in schizophrenia is a failure of congruence between an emotional expression and the ideational content. A patient with schizophrenia may talk about the death of a family member with a broad smile. This loss of harmony between the affective display and the ideational content is more common in chronic patients. Loss of congruence creates marked discomfort in the observer and contributes to the tendency of family and friends to remove themselves from contact with the patient. Psychological testing has demonstrated that patients with schizophrenia frequently cannot recognize the emotional state expressed in photographs of faces. This inability to recognize emotional cues in others may be expressed in the patient's inability to show their own emotional experiences in ways that are understood by others. The degree of emotional blunting and inappropriateness of emotional responses are excellent measures of the extent to which the illness has invaded the person's personality and contributed to its deterioration. Severe blunting and inappropriateness are associated with chronic schizophrenia.

Unusual Emotions Schizophrenia is characterized frequently by alterations in emotional reactions to external stimuli and often demonstrates peculiarities of emotions infrequently seen in normal states. Particularly during acute decompensations, patients may describe states of exaltation with feelings of omnipotence, oceanic feelings of oneness with the universe, religious ecstasies, and terrifying apprehensions about the disintegration of their own personality or body. It is not unusual to see intense anxiety about the impending end of the universe. These experiences are rare in normal individuals, except when under the influence of psychotomimetic drugs.

Emotional Sensitivity Many individuals who eventually develop schizophrenia premorbidly demonstrate a hypersensitivity to rejection. Much of the premorbid tendency toward social isolation can be understood as an avoidance mechanism to reduce the risk of rejection. After the onset of the disorder most patients continue to display extreme sensitivity to criticism and rejection. They may react to the tone or content of the therapist's intervention with a marked exacerbation of positive symptoms.

> An intelligent, well-mannered young woman suffering from chronic schizophrenia, who had improved to the point where she could live autonomously in her own apartment, had invited some friends to dinner. The dinner was well prepared and the evening enjoyable, until the friends invaded the patient's kitchen with forceful enthusiasm, and insisted on washing the dishes. Their hostess became enraged and ordered all guests to leave her home immediately. She later explained to her therapist that she felt humiliated and hurt by what she felt was a rude demonstration of her guests' conviction that she would never be able to make order in her own kitchen by herself.

Those who have worked extensively with patients with schizophrenia know that they are extremely sensitive. They are easily hurt by even slightly aggressive or rejecting behavior by others—behavior that in most cases would hardly be noticed by a person of normal sensitivity or, if noticed, would not lead to traumatic experiences.

> For example, a father's refusal to let his son with schizophrenia watch a particular program on television caused the son, who was not known to be a violent or impulsive person, to knife his father to death. In another case a psychiatrist's refusal to see a patient with schizophrenia at the moment the patient requested it (the patient was told he would be seen 3 hours later) caused the patient to commit suicide. Another patient with schizophrenia in remission committed suicide when his parents refused to include him on a 2-week trip to Florida.

Perceptual Disorders Various perceptual disorders occur in schizophrenia. It may be hypothesized that those disorders result from the patient's constant exposure to an overwhelming influx of sensory stimuli. Although normal persons have a fundamental, pervasive feeling of familiarity with the environment to which they have become accustomed and adapted—a necessary background for all normal perception—patients with schizophrenia may experience a haunting unfamiliarity with their environment. That unfamiliarity sometimes comes over them with a sudden jolt; at other times they experience a continuous feeling of strange remoteness, alienation,

and lack of contact (Fig. 12.7-10).

FIGURE 12.7-10 A symbolic representation of alienation that may be very severe in schizophrenia. (Courtesy of Erich Hartmann.)

Spontaneously reported sensory disturbances, mostly optical but also acoustic, olfactory, and gustatory, were observed in 15 percent of a large sample of patients with schizophrenia. The disturbances included hypersensitivity to light, changes in the perception of other people's faces and figures, misperception of movement, hypersensitivity to sound or smell or taste, and other changes in those senses.

Because of the unpredictable variability of the patient's experiences, the gestalt of the visual world are broken into disjointed parts. These patients frequently see objects and people change their dimensions, outlines, and brightness from minute to minute or even from second to second before their eyes. Déjà vu experiences may intrude and produce an uneasy feeling of spurious familiarity. Time may lose any structure or meaning, and the experience of passing time may extend or contract. These changes can be shown in experimental investigations on size and brightness constancy, on critical flicker-fusion frequency, on time estimation, and on many other perceptual functions (Fig. 12.7-11).

FIGURE 12.7-11 Drawings of a house made by a young man during an acute schizophrenic illness. **A** was drawn on admission to the hospital. **B** was drawn after 1 week of phenothiazine treatment. The percept house in the first drawing is fragmented. The windows are displaced, the roof is lying on the ground, beside the walls. In the second drawing, the structure and the perspective of the house are restored. (Courtesy of Heinz E. Lehmann.)

Hallucinations Sensory experiences or perceptions without corresponding external stimuli are common symptoms of schizophrenia. Most common are auditory hallucinations, the hearing of voices. Sometimes the voices are those of God or the devil; sometimes they are the voices of relatives or neighbors. Frequently, the patient can neither recognize nor understand them. Most characteristically, two or more voices discuss the patient in the third person. Frequently, the voices address the patient, comment on the patient's activities and surroundings, or are threatening or obscene and very disturbing to the patient. They may represent some evil outside power over which the patient has no control. Many patients with schizophrenia hear their own thoughts. When they are reading silently, for example, they may be quite disturbed by hearing every word clearly spoken to them.

Patients with schizophrenia experience visual hallucinations less frequently than auditory hallucinations, but they are not rare. Patients suffering from medically caused visual hallucinations experience them primarily at night or during limited periods of the day. They get relief only in sleep. Visual hallucinations that occur in schizophrenia are usually seen nearby, clearly defined, in color, life size, in three dimensions, and moving (Fig. 12.7-12). Visual hallucinations almost never occur by themselves but in combination with hallucinations in one of the other sensory modalities. Tactile, olfactory, and gustatory hallucinations are less common than visual hallucinations. Patients with schizophrenia often experience cenesthetic somatic hallucinations, sensations of altered states in body organs without any special receptor apparatus to explain the sensations (e.g., a burning sensation in the brain, a pushing sensation in the abdominal blood vessels, or a cutting sensation in the bone marrow. Hallucinations may absorb all or much of the patients' attention and may control their behavior to a considerable extent. While they are listening to voices, they may be preoccupied and oblivious to the environment. They may react with laughter or anger or terror and may carry on lengthy conversations with the voices.

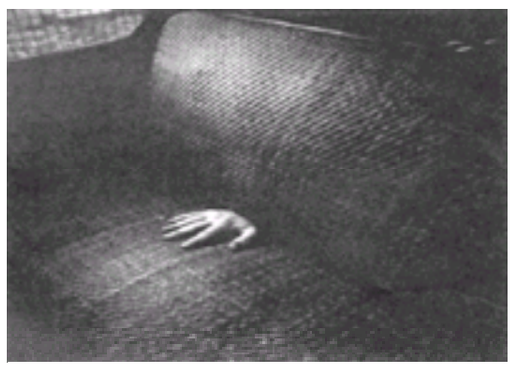

FIGURE 12.7-12 This photograph may symbolize the fantasy world of the schizophrenic. Not all psychiatrists are willing or able to attempt to empathize with that world, which can be a valuable therapeutic technique in some cases. (Courtesy of Arthur Tress.)

Modern treatment methods, particularly pharmacotherapy and social therapies that engage patients in various activities all day, have robbed hallucinations of much of their vividness and persistence. Also, many patients today know what hallucinations are and realize that hearing voices may be considered pathological and they will be considered crazy. Thus, present-day schizophrenia patients are much less likely to discuss their hallucinations openly than they were only 20 years ago.

Dream Content Studies of the dream content of patients with schizophrenia have shown that their dreams are less coherent, less complex, and less bizarre than those of normal persons. The incidence of dreams with color is apparently somewhat higher in patients with schizophrenia than in normal persons. Family members appear more often in dreams, and friends less often. The incidence of pleasant emotions in dreams of patients is similar to that for normal persons, but unpleasant

emotions are more common in the dreams of patients with schizophrenia than in the dreams of normal persons.

Thought Disorders

Delusions By definition, delusions are false ideas that cannot be corrected by reasoning and that are idiosyncratic for the patient (i.e., not part of the patient's cultural environment). They are common symptoms of schizophrenia.

Most frequent are delusions of persecution, which are a key symptom in the paranoid type of schizophrenia. The conviction of being controlled by some unseen mysterious power that exercises its influence from a distance is almost pathognomonic for schizophrenia (Fig. 12.7-13). It occurs at one time or another, in most, if not all, patients with schizophrenia, and for many it is a daily experience. Patients who are convinced that they are being persecuted by powerful agencies often harbor delusions of grandeur; they must be very important if so much effort is spent on their persecution. In connection with their experiences of being physically controlled by unseen forces, many patients with schizophrenia have elaborate delusions that their minds are controlled by telepathy or hypnotism. Modern patients whose delusions have kept up with the scientific times may be preoccupied with atomic power, X-rays, or spaceships that take control over their mind and body. Many patients with schizophrenia have delusional fantasies about the destruction of the world. The radio waves of the 1930s have been replaced by the alien creatures of the new millennium.

FIGURE 12.7-13 In schizophrenia, irrational and idiosyncratic ideas create a fearful world that is difficult for others to experience or understand, as symbolized above. (Courtesy of Arthur Tress.)

> Further to my investigation and research . . .. I would like to inform you that the tadpole in the eyes moves or floats around with the movement of the iris . . .. The tadpole reveals the photographic and its spirit the parabiological matter. From experience the Spirit is more deadly than the vision—the vision could bring on a person a berserk or manic attitude if he is unaware of its tricks—it could also be a danger to schizoid, alcoholic, and neurotic personalities.
>
> Further to the tadpole, it is luminous in the dark at times and flashes rings of light when both eyes are closed.
>
> Have you any idea if science could produce a solution that could cover the iris and eradicate the tadpole and the luminous matter?
>
> I repeat again, this is a diabolical science deliberately done to destroy human nature.
>
> Yours sincerely,
>
> J.M.

A delusional idea may occur with extraordinary rapidity. The patient may experience it as an illumination of the intellect in which the meaning of certain matters suddenly becomes clear. The total certainty is diagnostic of the delusional nature of the belief. There is a variation over time in the degree of certainty about the belief. At times the patient believes but is not absolutely certain, and at other times the patient cannot entertain any doubt in the belief system.

Phases of Cognitive Disturbance K. Conrad studied the development of delusions in patients with schizophrenia. In the first phase of the schizophrenic process, which he called “trema” (German for stage fright), patients become aware that something ominous is happening to them. Somehow, the world around them is changing, and they feel locked in, harassed, and powerless. They may make desperate attempts to regain control through elaborate schemes of body and character building.

Table 12.7-6 contains the schedule that a young man, aged 19, set for himself in a desperate last attempt to achieve a sense of security 3 months before he was admitted to a mental hospital in an acute catatonic stupor.

Time	Activity
7:00–8:00	Cold bath, toiletries, bed, dress
8:00–8:15	Encyclopedia (memorize three facts)
8:15–8:30	Handwriting
8:30–8:45	Brisk walk
8:45–9:00	Breakfast (one apple, one dish of bran, two glasses of milk, two glasses of water)
9:00–10:00	Hearing, sight, and scent
10:00–11:00	European, financial, and sport news
11:00–12:00	Wax floors and clean door knobs
12:00–1:00	Cold bath and exercises
1:00–1:30	Geometry
1:30–1:45	Vegetable lunch (very light)
1:45–2:45	Music
2:45–3:45	Walk as far as Atwater and then back to library
3:45–4:45	Study at library
5:00–6:00	Accounting
6:00–6:20	Cold bath
6:20–6:35	Vegetable supper
6:35–8:05	Accounting
8:05–9:00	Hearing, sight, and scent
9:00–9:30	Strength and Health Magazine
9:30–10:00	Wardrobe
10:00–10:20	Cold bath
10:20–10:40	Study vocabulary (12 words)
10:40–11:00	Undress and toiletries
11:00–11:15	Breathing exercises
11:15–11:30	Prior improvement in mental latitude in diary

Table 12.7-6 Schedule of a Person with Catatonic Schizophrenia

During the trema phase patients are anxious, irritable, and often depressed. That phase may last for only a few days, but it sometimes lasts for weeks or months. In the latter stages of the trema phase, patients may be in a delusional mood that makes them see their environment in a new and strange light, appearances are changed and familiarity is lost. A description of such an experience is given in the following excerpt from a patient's account, after recovery, of his schizophrenic attack.

> I suddenly realized that I just didn't have a clue where I was. I came to consciousness still driving along 401 highway but I had a vague suspicion that I might be nearing Montreal. I made a real effort to stay alert, but I seemed to keep drifting back to my former thoughts. Why had Edith said that I should believe in fairy tales? Was Joan really not dead but in a mental institution? Why had I been so cold that night in my car (the thermostat broke) and what was the significance of coldness, magnetism, and love?
>
> I imagine what I have recorded represents about 10 percent of my thoughts. They flashed through my mind so rapidly and with such turmoil that I became aware my thoughts were running out of control. And I felt that I wanted to see Dr. Wilson . . . Again on the same stretch of Highway 401 I lost the knowledge of location. The trigger again was through theories of cosmic matter in the aurora borealis—the Northern Lights; and the meaning of the Southern Cross.

The trema phase is followed by the epiphany phase, a Greek expression suggesting sudden revelation. Conrad postulated two phases after the apophanous: the apocalyptic fragmented phase and the terminal phase, in which the patient becomes vegetable-like. In the apophanous phase the patient suddenly becomes sure of certain new “facts.”

> A taxi driver this morning touched his cap with his right hand; this meant that he knew where the patient had been last night. A newspaper was lying on the stairs; this meant that his reputation would be ruined before nightfall. A man was feeding two squirrels in the park, indicating that the patient's future would be decided in 2 weeks. A television announcer makes a verbal slip; this broadcasts that the patient is stupid.

Conrad called such delusions *autochthonous*, meaning that they are primary and irreducible. They appear de novo and seem to have little or no connection with the patient's life history or specific stresses and conflicts.

Why do patients believe in the reality of these delusions? What evidence supports the farfetched connections they make? Patients always give the same answer, “I know it.” That direct, immediate, total certainty is the irrational, pathological aspect of the primary delusional experience. That certainty—“I know it”—cannot be explained by analyzing the patient's conscious psychic content. Many leading German psychiatrists, such as Jaspers, have stressed the difference between the primary pathological possibility of having delusions and the contents of delusions, which may have distinct, analyzable meaning based on the patient's psychological conflicts, drives, and needs.

Disturbances of Thinking Disturbances of thinking and conceptualization are one of the most characteristic features of schizophrenia. The feature common to all manifestations of schizophrenia thought disorder is that patients think and reason on their autistic terms according to their own intricate private rules of logic. Schizophrenic patients may be highly intelligent, certainly not confused, and they may be painstaking in their abstractions and deductions. But their thought processes are strange and do not lead to conclusions based on reality or universal logic (Fig. 12.7-14). The first author found that about 70 percent of patients with schizophrenia showed a typical schizophrenic thought disorder, and those who did usually had a withdrawn personality. He hypothesized that premorbid withdrawal characteristics already contained the thought disorder that emerged later.

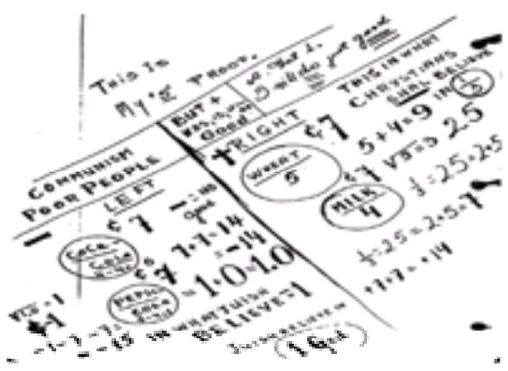

FIGURE 12.7-14 Schizophrenic patient's schema illustrates his fragmented, abstract, and overly inclusive thinking and preoccupation with religious ideologies and mathematical proofs. (Courtesy of Heinz E. Lehmann.)

One study emphasized the fact that the patient with schizophrenia may consider two things identical merely because they have identical predicates or properties. By contrast, in normal logical thought, identity is based on identical subjects and not on identical predicates. The patient with schizophrenia may reason (to quote Silvano Ariete), “The Virgin Mary was a virgin; I'm a virgin; therefore, I'm the Virgin Mary.” However, this particular fallacy is not specific for schizophrenia and is commonly committed by college students who are distracted or fatigued. Arieti believed that schizophrenic cognition uses isolated segments and parts, rather than the whole of the concept.

Patients with schizophrenia may reason: “John is Peter's father; therefore, Peter is John's father.” Such symmetrical reasoning is sometimes justified (e.g., John is Peter's brother, therefore, Peter is John's brother), but at other times such symmetrical conclusions are not justified, and patients not seem to know when they may apply them and when they may not.

Patients with schizophrenia use archaic modes of mystical or magical thinking. Such primitive modes of thinking are closely related to the psychoanalytic concept of primary thought processes that are at work in normal dreaming and allow condensation, reversal, substitution, displacement, and other distortions of conceptual relations impossible in rationally controlled thought. Jung, in fact, compared the psychic processes of schizophrenic patients who are awake to those of normal persons who are dreaming with their eyes open.

Kurt Goldstein described a concretization of thought and a loss of the abstract attitude as typical of schizophrenic thinking. Patients lose their ability to generalize correctly and exhibit in the ordering of their concepts a defect similar to a loss of the figure-ground relation in perceptual performance. That defect is often brought out by the simple clinical test of asking a patient to interpret a well-known proverb. One patient interpreted the saying “A stitch in time saves nine” as “I should sew nine buttons on my coat,” an overly personalized and concrete explanation.

Norman Cameron identified overinclusion as a typical feature of schizophrenic thought disorder. In contrast to patients whose mental functions are impaired by an organic brain lesion and who tend to omit important items in thought and speech, patients with schizophrenia tend to include many irrelevant items in their ideational and verbal behavior. That tendency seems to result from a loosening of associations in the schizophrenia patient. Studies have shown that overinclusive thinking is not a learning defect but an impairment of a central filtering process that normally inhibits external sensations and internal thoughts that are irrelevant to a given focus of attention. Only a well-functioning filtering-inhibiting process makes rational thinking possible. Overinclusive thinking usually develops within the setting of a delusional mood, when things look different, sensations are more intense, and everything seems to have some strange special significance (Fig. 12.7-15).

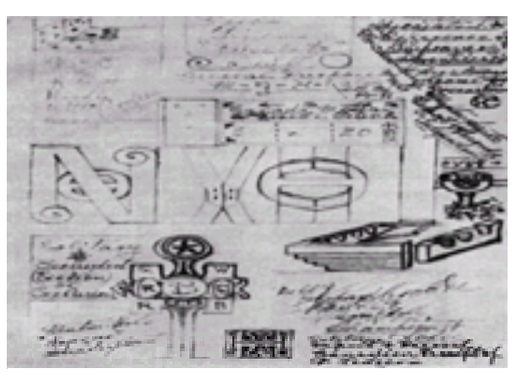

FIGURE 12.7-15 This drawing, carefully executed by a schizophrenic woman, graphically expresses her incoherent thinking and her tendency to perseveration of ideas, combined with an ability to accomplish quite complex drafting. Similar drawings may be produced when normal people doodle while their attention is not focused on what they are doing. (Courtesy of Heinz E. Lehmann.)

David Shakow demonstrated in a series of experiments that patients with schizophrenia cannot hold a set as well as a normal person can. That inability becomes

evident when patients are tested for their reaction time in responding to a stimulus preceded by a ready signal. Introduction of a ready signal shortens the reaction time of a normal person to the stimulus that may follow the signal within 10 to 15 seconds. The patients' reaction time remains the same, whether or not they are warned of the coming stimulus.

Somatic Findings

Physiological Testing A significant proportion of the patients who carry the diagnosis of schizophrenia exhibit neurological test abnormalities. The neurological findings that can be identified in approximately half the population tend to be soft signs. A careful history will also reveal that a significant percentage of patients with the diagnosis of schizophrenia experienced difficult labors and various obstetrical complications at birth. The patient's childhood developmental milestones tend to be delayed. School and work performance tends to be less adequate than that of their siblings.

One of the earliest signs of a developing schizophrenic illness is the loss of the normal gracefulness of body movements. This was reported before the advent of antipsychotic drugs, and it should not be confused with the dystonias seen secondary to medication. Perioral tremors were reported long before the use of dopamine receptor antagonists and should not be confused with tardive dyskinesia. All of these signs are associated with poor prognosis.

During the onset of an acute schizophrenic episode patients show increased autonomic tone, manifested in dilated pupils, moist palms, and moderate tachycardia. Systolic blood pressure tends to be elevated. Interestingly enough, there are few sleep disturbances after the acute stage of illness, although schizophrenic sleep is characterized by a tendency towards reduction of stage 4 sleep.

Water Intoxication Routine laboratory tests of patients sometimes find a low specific gravity of urine and a low sodium concentration in serum. Retrospectively, it may be noted that the patient seems always to be at the water fountain. The syndrome of self-induced water intoxication should then be considered, particularly in the differential diagnosis of seizures in schizophrenic patients. As many as 20 percent of patients with a diagnosis of chronic schizophrenia drink excessive amounts of water, and 4 percent of the chronic population suffer from chronic hyponatremia and episodic water intoxication. The workup for increased water intake should include repeated daily weighing and tests for inappropriate secretion of antidiuretic hormone, whose secretion is sometimes altered by treatment with antipsychotic agents, carbamazepine (Tegretol), lithium (Eskalith), or other drugs.

Constitutional Characteristics The relationship between body build and personality structure has been studied intensively for many years. Early studies showed that the diagnosis of schizophrenia occurred more frequently in persons of asthenic, athletic, and dysplastic body types than in the pyknic type. The pyknic type was believed to be more likely to develop manic-depressive psychosis. Although these observations may seem peculiar today, this perspective was prevalent in the 1930s. In the United States workers such as William Sheldon made extensive studies of the relationship between body type and mental illness. Using more-precise methods for measuring body types, he reclassified people into ectomorphic, mesomorphic, and endomorphic (Fig. 12.7-16). In this classification schizophrenics were more likely to be ectomorphs or mesomorphs.

FIGURE 12.7-16 Sheldon's method of dividing people into distinct body types. **A**, Endomorphic, characterized by roundness and excess subcutaneous fat deposits. **B**, Mesomorphic, characterized by excess muscular tissue. **C**, Ectomorphic, characterized by minimal muscular and subcutaneous tissues.

In urban areas today a high percentage of patients diagnosed with schizophrenia also suffer from concomitant substance abuse—alcohol, drugs, or both. The substance abuse appears to be a form of self-medication, which unfortunately leads to additional problems for the patients.

Two major eye abnormalities occur in schizophrenia. The first is a tendency toward frequent blinking, which is reduced with neuroleptic medication. This increased blink rate may reflect increased dopaminergic tone in the nervous system. The second is that in attempting to follow a moving object smoothly, abnormal saccades occur in approximately one-half to three-quarters of patients studied. Abnormal smooth pursuit may be a neurophysiologic marker for certain aspects of the pathophysiology of some patients with schizophrenia.

Abnormalities of prosody are quite common in patients with schizophrenia. This can be conceptualized as a neurological disorder of the parietal lobe. Considerable evidence from imaging studies indicates that the temporoparietal region is important in the pathophysiology of these disorders.

Imaging studies, including quantitative electroencephalography, reveal problems in having different brain areas work together in a synchronized fashion. This failure of coherent and integrated activity of different cell ensembles may well be the cardinal pathophysiologic finding of this group of disorders.

Suicide and Violence Suicide is surprisingly common in this population. The frequency of attempts varies to up to 40 percent of the population. About 10 percent of people diagnosed with schizophrenia commit suicide during the first 10 years of the illness. The risk of suicide is higher in men, particularly those with the paranoid type. The sudden, otherwise unexplained suicides of university students may be related to their experiencing cognitive dysfunction and positive symptoms that they fear to mention to anyone and that go undiagnosed and untreated. Suicide of an individual with a diagnosis of schizophrenia is much less predictable than it is for someone with depression.

> A 32-year-old man with chronic schizophrenia who lived at the home of his parents and was compliant with treatment committed suicide in a bizarre way while his parents were on a 2-week trip to Florida. The parents had taken such trips previously on several occasions. In a rambling suicide note the patient indicated that he resented not having been asked to accompany them.

Suicide is a significant danger for patients with schizophrenia. Probably more patients with schizophrenia than with bipolar I disorder commit suicide, although the immediate risk of suicide is relatively greater among the latter. Patients with schizophrenia may commit suicide because they are deeply depressed or they may kill themselves in response to the relentless commands they receive from hallucinatory voices.

> A patient with schizophrenia who had jumped to the street from a third-floor balcony sustained several fractures but lived to say that for many days a man's voice had told him persistently to jump out of a window. He did not want to die, and he resisted the voice for as long as he could, but he finally had to yield to its demands.

Patients with schizophrenia are more violent as a group than the general population. This is particularly a problem with patients with the paranoid type who may act quite suddenly and impulsively on a delusional idea. Patients with paranoia tend to be intelligent and capable of forming plans; therefore, they represent a much greater risk than individuals who are disorganized and cannot plan an effective attack. Despite earlier beliefs, command hallucinations do not appear to play a particularly important role in violence. Violence between patients in hospitals frequently results from the attacking patient's mistaken belief that another patient is behaving in a threatening way or getting physically too close. Studies have revealed that violence in a hospital setting can result from undiagnosed neuroleptic-induced acute akathisia. Persistently violent inpatients often do well on special treatment units that provide a more structured program and a less-crowded

environment. The patients who fail to respond to this kind of care usually show neurological signs in addition to their diagnosis.

Unfortunately, it is exceedingly difficult to prevent most schizophrenic homicides, since there is usually no clear warning. Most of the homicides come as a horrifying surprise. Patients who are known to be paranoid with homicidal tendencies should not, as a rule, be allowed to move about freely as long as they retain their delusions and their aggressive tension. But, like the patient who hanged himself without previously manifesting any observable depression, the homicidal schizophrenia patient may appear to be relaxed, even apathetic, and then, within a day or two, kill somebody.

The number of homicides committed by patients with schizophrenia may increase during the next few years. As a result of the gradual reduction in the hospital confinement of patients with schizophrenia, many of them are treated with modern methods of therapy in the community, where it is often impossible to control and supervise their pharmacotherapy and to prevent recurrences of paranoid homicidal behavior.

> A man with schizophrenia who had been going home on weekends for many months was told by his sister that she would no longer ask permission to take him out of the hospital if he would not do his part of the housework—for instance, help with the dishes. On the next weekend visit, the patient killed his sister and his mother. He had shown no signs of disturbance whatsoever during the preceding week, had been sleeping well, and had been attending occupational therapy classes as usual.
>
> A 19-year-old boy was discharged from a mental hospital in what seemed to be a residual state of chronic schizophrenia of the undifferentiated type. He stabbed his father to death when the father, during a state of intoxication, told the patient that he was too much of a bother around the house and that he might as well return to the hospital.
>
> A man with schizophrenia, whose condition had not yet been diagnosed, complained to a general practitioner about various physical ailments. When the physician finally told him that he should not come anymore because there was nothing else he could do for him, the patient quietly left the office. He returned a few hours later and killed the doctor.

Rejection Careful analysis of these unpredictable suicides and homicides leads to the conclusion that the most significant single factor in most of them was a traumatic experience of rejection. The pathological sensitivity of persons with schizophrenia makes them extraordinarily vulnerable to all common life stresses. Rejection, particularly by members of their own family, seems to be more traumatic for them than most other stresses. The act of rejection may seem trivial, and it is often not deliberate on the part of those who reject the patient. In fact, they are practically never aware of it, and the patient may not show any immediate reaction to the rejection at the time.

It may be significant that four of the second author's patients with schizophrenia who committed homicide harbored the delusion that their parents were adoptive parents rather than real parents. That delusion in itself seems to reflect a deep-seated feeling of being rejected by the parents.

Prepsychotic Personality A clinical evaluation must always consider the patient's prepsychotic personality. The clinical, but not typical, history is that of a schizoid personality—quiet, passive children with few friends; daydreaming, introverted, and shut-in adolescents and adults. They are often reported as having been especially good children because they were always obedient and never in any mischief. In school they were good in spelling but poor in arithmetic. They made few friends as children and their deficient friendship was particularly noticeable in adolescence.

Typical schizoid adolescents have few dates, do not usually learn to dance, and have no close boyfriends or girlfriends. They are not interested in petting or other heterosexual or homosexual activities but are often disturbed about masturbation. They avoid competitive sports but like to go to the movies, watch television, or listen to music. They may be avid readers of books on philosophy and psychology.

Less than 25 percent of patients with schizophrenia have a history of the type of schizoid personality described above. The former assumption that schizophrenia is genetically transmitted either as the disease proper or as a schizoid character structure is now changing to the concept of genetic transmission of schizophrenic spectrum diseases (including various personality and neurotic disorders) and a variety of conditions—even valuable traits, such as creative ability. About 90 percent of patients diagnosed with schizophrenia have no known family history of the disease.

Precox Feeling Some clinicians believe they can diagnose a precox feeling. That feeling consists of an intuitive experience by the examiner that determines whether or not it is possible to empathize with the patient. Patients whose emotional distance makes it impossible to establish an empathic rapport are classified as having schizophrenia, providing other criteria are met. The reliability of this approach is questionable and even though important, the use of the examiner's feeling as a diagnostic criteria should be discouraged.

Positive-Negative Distinction The distinction between manifestations of schizophrenia that appear to represent a loss of function (e.g., emotional blunting, poverty of speech) and such symptoms as delusions and hallucinations has been part of the diagnostic process at least since Kraepelin, whose concept of an avolitional syndrome was the predecessor of the concept. Bleuler's division of symptoms into fundamental and accessory types may be seen as supporting that separation, with certain of the negative symptoms assigned diagnostic primacy. However, Kraepelin agreed in principle with Bleuler on the distinction between fundamental and accessory symptoms. In 1913 Kraepelin stated

> the former [fundamental symptoms] constitute the real characteristics of the clinical state and can be demonstrated in each individual case more or less distinctly; the latter [accessory symptoms] may be present but may also be absent; they are not caused by the character of the morbid process but by circumstances which are in loose connection with it . . . [F]rom this point of view the weakening of judgment, of mental activity and of creative ability, the dulling of emotional interest and the loss of energy, lastly, the loosening of the inner unity of the psychic life would have to be reckoned among the fundamental disorders of dementia praecox, while the remaining morbid symptoms, especially hallucinations and delusions . . .. would be regarded more as secondary accompanying phenomena . . ..

As psychiatric nosology has been modified over the past decades, largely in the service of greater reliability, the positive symptoms (designated accessory symptoms by Bleuler) have assumed principal importance in the diagnostic criteria used internationally and in the United States from DSM-III onward.

Two distinct psychopathological processes were postulated in 1980. Type I schizophrenia was characterized by predominantly positive symptoms, good premorbid functioning, sudden onset, normal brain structures by computed tomography (CT), good response to treatment, and a better long-term course. Type II schizophrenia was characterized mainly by negative symptoms, an insidious onset, poor premorbid functioning, abnormalities on CT scans, a tendency to drug resistance, and a poorer long-term course and outcome, often resulting in behavioral deterioration. Other similar groupings include negative and positive schizophrenia (Table 12.7-7) and deficit and nondeficit forms of schizophrenia. The past decade saw a large number of investigations into possible relations between those syndromes and a variety of issues, including course and outcome, neurotransmitter hypotheses, brain imaging findings, and family studies.

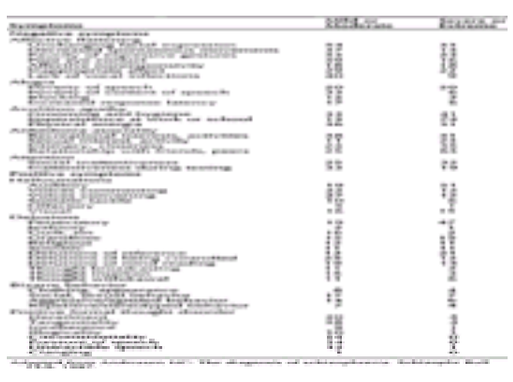

Table 12.7-7 Percentage of Patients With Negative and Positive Symptoms (111 Consecutively Admitted Schizophrenic Patients)

Scales have been developed to measure negative symptoms that have acceptable interrater reliability. Those scales invariably designate flat affect and poverty of speech among the negative symptoms and generally also include anhedonia, apathy, and avolition. Thought disorder, bizarre behavior, and inappropriate affect are more variable in such classifications.

A recent review summarized and compared findings in patients with negative and positive symptoms. Those with negative symptoms experienced an earlier onset of schizophrenia, tended to be male and unmarried, had worse premorbid functioning, had more motor abnormalities, and were more likely to be concordant for illness if an identical twin. In view of those findings, negative symptoms have been reintroduced into the diagnostic classifications as one of the characteristic symptom complexes necessary for the diagnosis of schizophrenia.

Certain manifestations simulate negative symptoms but are a consequence of medication, depression, institutionalization, or other life circumstances. These manifestations must be distinguished from the core negative symptoms of schizophrenia. Moreover, most patients present with a mixture of positive and negative symptoms, which vary in degree over time.

DIFFERENTIAL DIAGNOSIS

Hallucinations and Delusions Another reason for the greater tendency to diagnose schizophrenia in North America than in other countries is that many North American psychiatrists take it for granted that a patient who is hallucinating or who expresses paranoid delusions must be schizophrenic if no organic brain disease can be detected. For example, in a survey of psychiatrists in the United States delusions ranked second of the top 10 symptoms indicating schizophrenia and hallucinations fifth; psychiatrists in Great Britain list delusions in eighth place and hallucinations are not among the first 10 symptoms.

The clinician must not forget that the presence of delusions and hallucinations confirms only the presence of psychosis, not that of schizophrenia. Between 10 and 15 percent of bipolar I disorder patients have hallucinations or delusions. These symptoms indicate a serious loss of contact with reality, a principal criterion for the diagnosis of a psychotic condition, which may or may not be schizophrenic. At times schizophrenia-like symptoms frequently occur in cyclothymic disorder, and delusions and hallucinations may occur in other nonschizophrenic psychiatric conditions (e.g., delusional disorder, psychotic disorder, brief hysterical twilight states, toxic conditions, and mental disorders due to a medical general condition).

Diagnosing schizophrenia simply because delusions and hallucinations are present is like making the diagnosis of a coronary occlusion solely on the basis of pain in the chest or the diagnosis of typhoid fever only on the presence of sustained pyrexia. Single symptoms should certainly suggest conditions in which such symptoms frequently occur, but unless specific, conclusive tests exist for disease processes, a final differential diagnosis must always be based on the complete clinical picture.

Sensory and perceptual disorders, such as hallucinations, may indeed give good diagnostic clues. However, not all types of hallucinations point toward schizophrenia. There are important qualifications regarding the modality, the time, and the content of the hallucinations. Experiences of being controlled by outside forces or having strange, continuous, somatic (cenesthetic) hallucinations or auditory, verbal hallucinations (particularly if the voices are coming from God or the devil or address the patient in the second person or talk about him) may support a diagnosis of schizophrenia. Perceptual distortions of time or objects in space point toward a diagnosis of schizophrenia, but only if they have been present at least several days; otherwise, they may have a toxic origin. The time factor also applies to loss of ego boundaries and the experience of having one's thoughts spread to others.

The presence of delusions provides strong presumptive evidence for schizophrenia only if those delusions have strange, magical, esoteric, or bizarre content. A person who is convinced that he is deliberately discriminated against by his foreman at the factory, that his wife is running around with other men, or that his wife is trying to poison him may have paranoid delusions, but they are not necessarily due to schizophrenia. On the other hand, a man who is convinced that he is the victim of a Pentagon-directed plot to destroy his brain by special death rays beamed at him from space satellites or that he is the Virgin Mary expresses delusions that, by their very bizarre character, point definitely in the direction of schizophrenia.

The diagnosis of schizophrenia cannot be made entirely on the basis of observation, logical reasoning, or objective measurement. It still requires a careful and comprehensive clinical evaluation. Such an evaluation must take into account the presence or absence of certain key schizophrenic symptoms, the patient's prepsychotic personality, the physical findings, the family genetic history, the social environment, the various aspects a good clinical anamnesis may reveal about the natural history of the disease, and any possible precipitating factors.

Catatonia Several investigators have emphasized that catatonia is not a disease entity or exclusively a subtype of schizophrenia but is instead a nonspecific syndrome that occurs quite frequently in other psychiatric conditions such as mania. In one study of 123 patients who satisfied the criteria for manic episode, following catatonic signs were present: sterotypy, echopraxia, and stupor. Others have reported the presence of a typical catatonic syndrome in organic brain diseases (e.g., cerebral aneurysm).

Anxiety Disorders Hysterical symptoms are common in schizophrenic breakdown; thus, the presence of hysterical, dissociative, or even conversion symptoms does not rule out a diagnosis of schizophrenia. Schizophrenic breakdown is preceded by a period of marked tension and anxiety, which may last only a few days or extend over many months. During the acute and subacute stages of a schizophrenic attack, anxiety and depression may color the clinical picture significantly, again without excluding the overruling diagnosis of schizophrenia, which in some cases has not yet diminished the patient's emotional reaction to the onslaught of the psychotic attack. Obsessive symptoms are common in schizophrenia, and what appears to be obsessive-compulsive disorder can develop into schizophrenia.

Bipolar I Disorder The differential diagnosis between schizophrenia and bipolar I disorder should not present many difficulties. The behavior of excited catatonic persons is directed primarily by their own qualitatively disordered mental process. Their actions are unpredictable and appear senseless, affect is difficult to understand, and verbal productions may be irrational and incoherent. Patients with mania, on the other hand, are distractible, and most of their actions are determined by their immediate environment. Their activity resembles that of an excessively busy person, rushing from one superficial job to another. Their affect is clearly one of playful euphoria or angry irritability but always outgoing and expansive. Their verbal productions are accelerated and increased in number, and they reveal a quantitative disorder of association processes, rather than the intrinsic, qualitative thought disorder of the patient with schizophrenia. Nevertheless, several investigators have raised the question, mainly on the basis of a favorable outcome, of whether schizophrenic episodes of relatively short duration are associated with mood disorders.

Depressive conditions should not be diagnosed as schizophrenic unless some unmistakably schizophrenic symptoms are present. In schizoaffective disorders (which may resemble bipolar disorders), the presence of clearly schizophrenic symptoms, such as schizophrenic thought disorder, places the reactions in the diagnostic category of schizophrenia.

Delusional Disorders Since delusions of persecution or grandeur are essential symptoms in both the paranoid type of schizophrenia and in the paranoid type of delusional disorders, the differential diagnosis between the two conditions must be carefully considered when paranoid symptoms prevail. The diagnostic decision must be based on the presence or absence of the essential features of schizophrenia and its paranoid type.

Adolescent Disorders Any psychiatric disorder that occurs during adolescence assumes a certain schizophrenic coloring, since many of the features characteristic of nonschizophrenic adolescent turbulence—exaltation, intense preoccupation with abstract ideas, unpredictable variations of mood, daydreaming, introspection, shyness—are often seen in schizophrenia. Therefore, it is not unusual to misdiagnose a manic or depressive phase of a bipolar disorder as schizophrenia if the patient's first attack occurs in late adolescence. The rule that bipolar disorders do not occur in late adolescence is not always true; later recurrences of that mental disorder may cease to display symptoms resembling schizophrenia, and the correct diagnosis can then be made.

PSYCHOLOGICAL TESTS

There are no psychological tests for schizophrenia comparable to the definitive biological or immunological tests for pregnancy and syphilis. There are only psychological tests that are more or less compatible with a diagnosis of schizophrenia and make the diagnosis more or less probable. Tests can rarely establish a diagnosis by themselves, divorced from the clinical findings.

The clinical psychologist's tests differ from clinical observation and interview; the method of test administration is uniform and the final evaluation of the test findings is based on comparisons with statistical norms, at least with psychometric test instruments. In most clinical centers certain psychological test batteries are routinely used to indicate, confirm, or rule out a diagnosis of schizophrenia. These test batteries are usually composed of projective tests, psychometric tests, and personality inventories. The most frequently used projective tests are the Rorschach test, some drawing tests, and the Thematic Apperception Test (TAT). The most commonly used psychometric tests are the Wechsler Adult Intelligence Scale (WAIS) and some tests that probe concept formation and the organization of thought processes. The most widely used personality inventory is the Minnesota Multiphasic Personality Inventory-(MMP-2), a self-report questionnaire that renders profiles of psychopathology or response styles.

Effects of Medications Many patients with schizophrenia today have received some form of pharmacotherapy before they are seen by a psychiatrist. If they were given adequate dosages of antipsychotic drugs in the early stages of a schizophrenic attack, important key symptoms (e.g. the experience of being controlled, thought hearing, delusions, hallucinations, and inappropriate behavior) may have subsided within 2 or 3 days, and the disorder the psychiatrist now faces may lack all or most of its more specific distinguishing features. The dilemma is similar to that of making an accurate diagnosis of an acute abdominal condition after opioids have been administered or of diagnosing a septicemia after administration of antibiotics. The possible effect of previous antipsychotic pharmacotherapy on key schizophrenic symptoms must always be taken into account when making a differential diagnosis.

Substance-Related Disorders Another drug factor that may render a differential diagnosis of schizophrenia more difficult today is the widespread nonmedical use of amphetamines, crack cocaine, hallucinogens, and drugs with similar effects. Many young people who develop psychotic symptoms have a history of using these drugs. Amphetamines, amphetamine-like drugs, and cocaine in high doses can produce psychotic (usually paranoid) conditions that mimic schizophrenia in their symptoms and course so closely that a differential diagnosis may be impossible in some cases. Fortunately, treatment is the same for amphetamine-induced psychotic disorder and for paranoid schizophrenia. With hallucinogens, the psychotic symptoms may have more of a toxic character, with vivid nonauditory hallucinations and sometimes clouding of consciousness. Differentiating drug-induced psychosis from schizophrenic psychosis may sometimes be difficult. Schizophrenic symptoms seem to develop about 4 years earlier in drug users than in nonusers; thus, drugs may play a precipitating role in the onset of schizophrenia.

Cultural Factors An entirely new set of differential diagnostic problems has been generated by special lifestyles and philosophies, such as beliefs borrowed from Eastern cultures.

> A 20-year-old college dropout was arrested in front of a gas station, where he was meditating and blocking traffic. (Inappropriate and bizarre behavior?) When he was arrested, he responded by laughing. (Inappropriate emotional reaction?) At the police station he expressed his need to “laugh or fuck” to “prevent thinking.” (Bizarre and irrelevant, probably autistic, reasoning?) The police concluded that he was clearly insane and delivered him to a mental hospital.
>
> The young man told the psychiatrist who examined him that he was a Zen Buddhist and that thinking and analyzing things inhibited true growth of personality, according to his philosophy. He was convinced that the two best ways of preventing himself from getting lost in thinking were “laughing and fucking,” because both were incompatible with thinking. “Sort of reciprocal inhibition, if you believe in that stuff.” Under certain circumstances (e.g., in the police station) he could use only the laughing method. He tried to make things “buddhaful;” his smile indicated that that was a pun and not a neologism. Why had he chosen the gas station as a place to meditate? “Well, that's where the winds of Karma blew me.” After 30 minutes he terminated the interview by walking out of the room, remarking that he was becoming upset by the “very bad vibrations” he was getting from the psychiatrist.
>
> After a few day's observation the man was discharged. The psychiatrist suspected that he might develop schizophrenia eventually, but there were no grounds for making the diagnosis at the time.

COURSE AND PROGNOSIS

Course

Natural History Using the term natural history is almost arbitrary in describing a heterogeneous group of disorders with differences in cause, onset, pathogenesis, course, and outcome. It is possible, however, to describe the more common clinical patterns, even though they do not attain the frequency necessary to serve as absolute diagnostic criteria.

Mode of Onset Onset of the schizophrenic disorders ranges from acute to subacute to insidious. Onset usually occurs in late adolescence or early adulthood, somewhat earlier in men and later in women. Commonly following a prolonged period of increasing social withdrawal and turning inward to philosophical or religious interests or both, is the onset of psychotic illness. These individuals frequently show a childhood developmental pattern in which they tend to play by themselves, have few playmates, avoid eye contact, and be somewhat awkward in their motor activities. They tend to achieve developmental milestones later and frequently show poorer school performance than their siblings. About the time of onset of illness there is frequently social deterioriation as well as a loss of interest in personal grooming and hygiene.

The prodromal period may continue for weeks and up to several years before the symptoms suffice to make a diagnosis. When properly recognized this long prodromal period offers an opportunity for early intervention. Studies are under way to evaluate the use of low-dosage antipsychotic treatment in high-risk individuals showing prodromal signs. However, some individuals have very short prodromal periods before the onset of a psychotic state. In these latter individuals the progression to psychosis is more obvious, while in the former cases the progression is almost imperceptible. Nevertheless, despite the variability the first psychotic episode is frequently associated with a stressful life event such as going off to college.

Characteristic of most psychotic states is the loss of reality testing and insight. As a consequence, a psychotic episode is often heralded by an increasing dissonance between persons and their social environment. As the members of the social environment become increasingly concerned it is not unusual for them to use some form of coercion to bring the patient to medical attention.

Episodic Course Some patients with schizophrenia may have five or more psychotic attacks, usually of the catatonic type, without suffering any obvious personality damage, but because the risk of developing a schizophrenic defect increases after each additional schizophrenic attack, today's therapeutic challenge is to use all available measures to prevent relapses.

In Russia, psychiatrists place special diagnostic importance on the occurrence of relapse and base their criteria for the differentiation of various types of schizophrenia on the character of the course of the illness. Three main types of schizophrenia are distinguished in Russia: (1) continuous, with subtypes of sluggish, progressive, and malignant schizophrenia; (2) periodic, not progressive; and (3) shiftlike progressive, again with subtypes mild, progressive, and malignant. Russian psychiatrists assume that these differences in the course of schizophrenia are also related to differences in cause, symptom formation, and response to treatment.

PROGNOSIS

Table 12.7-8 lists features of schizophrenia that weigh toward a good or poor prognosis. It has been clinical knowledge since Kraepelin and Bleuler that the hebephrenic and simple types of schizophrenia have the poorest prognoses, that paranoid reactions have an intermediate prognosis, and that acute catatonic reactions have the best prognosis but that catatonic patients who go on to chronicity usually continue to regress and may become markedly deteriorated.

Good Prognosis	Poor Prognosis
Late onset	Young onset
Obvious precipitating factors	No precipitating factors
Acute onset	Insidious onset
Good premorbid social, sexual, and work histories	Poor premorbid social, sexual, and work histories
Mood disorder symptoms (especially depressive disorders)	Withdrawn, autistic behavior
Married	Single, divorced, or widowed
Family history of mood disorders	Family history of schizophrenia
Good support systems	Poor support systems
Positive symptoms	Negative symptoms Neurological signs and symptoms History of perinatal trauma No remissions in 3 years Many relapses History of assaultiveness

Table 12.7-8 Features Weighing Toward Good to Poor Prognosis in Schizophrenia

Modern pharmacotherapy has changed many of the old prognostic patterns. Today, a paranoid type schizophrenia patient's chances of making a good recovery are at least equal to those with the catatonic type. Even patients with the disorganized type often have good remissions after a few months of pharmacotherapy. The simple deteriorative disorder patient is still the least responsive to modern biological therapies. The more sudden the onset of a schizophrenic attack, the better are the chances for a good remission or complete recovery. If a precipitating event has clearly triggered the breakdown, the chances for a favorable outcome are also relatively better.

As a rule, the younger a patient is at the onset of schizophrenic psychosis, the worse is the prognosis. Patients with onset in childhood or early puberty seldom recover completely. Some investigators believe that there may be two major groups of people with schizophrenia: those with difficulties in childhood before the onset of schizophrenic symptoms and those with a history of normal childhood. The first group is characterized by lower I.Q.s, a higher incidence of difficulties in school, poor peer group adjustment, an earlier age of first psychiatric contact, evidence of possible minimal brain damage, and a poor prognosis.

A history of good adjustment in the important areas of social, sexual, and occupational functioning before the breakdown also indicated a favorable prognosis. Married men with schizophrenia have a better prognosis than do single, divorced, or widowed patients; the fact that they are married is evidence that interpersonal bonds may serve as a bridge for a return to the community. Patients who relate easily to people in their environment and who are capable of emotional warmth and natural emotional reactions have a good chance for reintegration. The presence of depression as in schizophrenia also improves the prognosis. Conversely, sustained emotional withdrawal and aloofness or shallow and inappropriate affective responses are ominous prognostic signs. In 1968 this first author and Felix Sugerman drew attention to the significance of formal thought disorder and of reaction time performance for the prognosis in patients with schizophrenia. They found that a combination of test results for thought disorder and motor reaction time, combined with the presence or absence of depression and marital status, gave the most important clues to the final outcome in their sample of schizophrenia patients. In 1973 a report concluded from an extended follow-up study of patients with schizophrenia that most of them seemed to reach a plateau about 5 years after their first breakdown and that those who had not remitted after 2 to 3 years faced a guarded outlook.

However, one must not assume that the prognostic indicators that are relatively well established for schizophrenia patients in Western culture are universally valid. A different cultural setting may produce different results. For example, one study examined admission data and 12-year outcomes for Moslem Indian, Hindu Indian, and African schizophrenia patients on the island of Mauritius and found that only for the small Moslem Indian sample were prognostic criteria comparable to those of a British sample. In the Hindu Indian and non-Indian groups, the results were no better than chance. Depression was not associated with recovery; sudden onset was a favorable factor in the Indian group but not in the others; flatness of affect was not highly associated with chronicity. Only the absence of a good work history gave the expected results. But two features that are not known to have prognostic significance in the West, psychosomatic symptoms and somatic disease, had a strong association with chronicity in the Hindu Indian and African groups of the Mauritius schizophrenia patients.

Family The family plays an important role in a patient's prognosis. A number of studies in recent years have shown that many patients with schizophrenia come from deeply disturbed families. Before making a prognosis one should determine whether the patient is accepted by the family and whether the dynamic pattern of communication within the family is adequate or characterized by irrationality, deficient sharing of foci of attentions, or the production of double-bind messages.

Relapses A new prognostic factor has emerged in the past few years, the patient's cooperation and conscientiousness in following prescribed drug maintenance therapy. Many patients with schizophrenia can be rendered relatively symptom free within a few weeks or months, but about 60 percent of them can be maintained in that condition only with continued drug therapy after they have been discharged into the community. Placebo-controlled maintenance studies have established this point well. The more often patients neglect taking their maintenance medications (i.e., the worse their compliance), the more likely they are to suffer relapse.

Several important observations have been made about the factors that determine whether a patient in remission will suffer a relapse. The most important protective factor is undoubtedly maintenance therapy with antipsychotic drugs. One study noted a history of sudden social or psychological traumas (e.g., the death of a parent, moving from one apartment to another) during the 3 weeks preceeding schizophrenic relapses in about 60 percent of cases.

The type of home in which the patient in remission resides plays a vital role. In one American study patients with schizophrenia fared better in conjugal homes than in parental homes, but some British investigators made the opposite observation in their sample. Most importantly, clear correlations exist between exposure to expressed negative critical emotions in the household of a patient with schizophrenia in remission and the likelihood of relapse. The critical time limit seemed to be about 35 hours of such exposure a week. When that time limit was exceeded, even maintenance drug therapy was often inadequate in preventing relapse.

Deterioration The risk of personality deterioration increases with each schizophrenic relapse. Schizophrenic recoveries are often called remissions because many of the patients later relapse. Although patients may remit again, each schizophrenic attack carries a greater probability of some permanent personality damage. Risk of personality deterioration increases rapidly after the second relapse. However, chronic schizophrenia does not inevitably lead to intellectual deterioration. In fact, in one sample, patients with chronic schizophrenia retained or improved their mean intelligence scores in spite of old age and prolonged institutionalization over a period of 14 years. In a group of schizophrenia patients who were followed over a 10-year period, schizophrenic symptoms decreased by 15 percent.

Final Outcome A schizophrenia patient's chances for a favorable outcome of the psychosis are estimated today to be about 4 to 5 times better than they were before World War 1. Kraepelin reported in 1913 that nearly 13 percent of his patients with dementia precox recovered from their first attack, but most of them later relapsed. Altogether, only about 15 percent ultimately had passable social remissions, usually with slight-to-moderate personality damage. Today, with good follow-up therapy and well-controlled maintenance drug treatment, only some 10 to 15 percent of patients in remission relapse within a year, compared with about 65 to 70 percent who relapse during the same period without such follow-up treatment. There are five possible outcomes for the patient with schizophrenia: full and permanent recovery; full remission, with one or more future relapses; social remission with personality defect and with the patient either capable of self-care and self-support or dependent on protection and supervision; stable chronicity; and deterioration to a terminal stage.

Deterioration to a Terminal Stage A vegetable-like existence is rare among patients with schizophrenia who have become ill during the past 25 years. Modern biological and social therapies are generally successful in preventing at least the terminal stage of deterioration, which was the most probable outcome of schizophrenia in Kraepelin's time and which was probably due to the ravages of institutionalization more than to schizophrenia. The modern mental hospital has few patients who illustrate this terminal stage of schizophrenia. Almost all patients with the symptoms and signs of extreme regression of behavior, affect, and ideation were admitted 30 or more years ago.

Stable Chronicity Despite all intensive therapeutic efforts, many patients with schizophrenia remain in a state of stable chronicity, although they do not regress to a terminal stage of deterioration. Their psychotic symptoms may make it necessary to keep them hospitalized or, if the symptoms are not severe, they may reside outside the mental hospital. They remain definitely incapacitated, with clearly visible signs and symptoms of active mental disease. In a study of schizophrenia patients from 1913 to 1923, when virtually no treatment was available for chronic schizophrenia, clinicians reported that some patients who had been chronically ill for almost 10 years occasionally had spontaneous remissions that lasted for years.

The combination of institutionalism and poverty frequently works to the detriment of the patient, and it is often difficult to determine how much of the patient's defective

functioning is due to each of these factors. However, many patients do not seem to be much better off in the community than in the hospital. It has been estimated that unemployed patients with chronic schizophrenia living outside the hospital spend about 30 percent of their time doing nothing, which is approximately what one observes in chronic schizophrenic patients residing in a modern hospital. The casual observer in big cities (not only in the poor districts) encounters many of those formerly hospitalized schizophrenic patients along the sidewalks, talking to themselves and passers by, gesticulating, preaching, or just sitting around.

It is difficult to estimate how many patients with schizophrenia today will end up in this category of stable chronicity, but under good therapeutic conditions, the percentage almost certainly will not exceed 30 to 40 percent. The remainder of the patients will either remit or recover.

Remission and Social Recovery Many patients with schizophrenia today fall into the categories of stable remission with personality defect and full remission with relapse. Schizophrenic personality defects and schizophrenic residual states are characterized by a reduction of ambition, initiative, available energy, and emotional responsiveness. Persons in that state may be more withdrawn, more aloof, and more selfish than they were before the onset of illness. They may neglect their personal appearance, and they almost certainly go down on the occupational ladder. Professional persons may, at first, still hold professional positions, but positions with reduced responsibility and less scope for personal initiative. They may eventually end up doing menial work well below their educational level. Persons with schizophrenic personality defect cannot readily assume responsibility. They cannot cope with competitive pressures and cannot tolerate time pressure. They are best suited for quite, routine work they can perform independently from others and at their own pace. Some former schizophrenia patients, therefore, prefer to do night-shift work, because it is less demanding and often permits them to work alone.

The personality defect may be so pronounced that patients cannot take charge of their own affairs and require continuous protective supervision and sheltered work conditions. If the personality defect is less pronounced, patients may be capable of acting independently and supporting themselves, although usually at a lower occupational level than before the illness. Sometimes the personality defect is so slight that only the patient's family and close friends recognize the subtle changes that have taken place—a diminished capacity for enthusiasm, lessened spontaneity, decreased initiative, and a decline in creative imagination. In most daily life situations the patient may even function socially at an apparently normal level.

Full and Permanent Recovery The prospect for full and permanent recovery from a schizophrenic episode is probably considerably brighter today than it was a half century ago, when the chances for such complete recovery were only 2 to 4 percent. Under careful supervision, antipsychotic drugs can now prevent relapse that could not have been prevented before the advent of those drugs. Only with maintenance drug therapy has it become possible to prevent relapses and thereby greatly reduce the risk of personality deterioration in many patients. Nevertheless, the introduction of modern treatment methods over the years has not yet significantly increased the actual number of full, permanent recoveries.

Occupational Rehabilitation Studies found that about 60 percent of schizophrenia patients were employed 75 percent of the time over a 5-year period. A follow-up study of a sample of 108 schizophrenia first-admission patients for 5 years revealed that they had been working for 65 percent of that period. Of a series of 188 discharged patients, 67 percent were employed for more than half of the 1-year follow-up period. Other studies found that more than half of a group of male schizophrenia patients had been employed more than half of the 5-year follow-up period. Among 100 hospitalized patients with schizophrenia followed for 5 years after discharge, 18 percent of the patients had not been employed at all during the follow-up period, but the other patients had been employed, on average, for 57 percent of the follow-up period.

FUTURE DIRECTIONS

Schizophrenia as currently diagnosed does not inevitably lead to severe deterioration but it is usually chronic and psychologically disabling to a significant degree. The better the premorbid social adjustment, the better the prognosis. Individuals who had good interpersonal skills and good psychosocial adaptation premorbidly do better than those who showed premorbid withdrawal. The outcome of schizophrenia tends to be worse in terms of functional level than it is in the affective disorders. Like the mood disorders, it is associated with an increased risk of suicide but a much shorter life expectancy, because suicide tends to occur in young adulthood. While the disorder tends to plateau clinically after about 5 years, it does not tend to remit in symptomatology until after the age of 50. Evidence suggests that patients do better in a rural and less demanding environment than they do in a complex urban culture. While it is possible to be more reassuring today than it was in the preneuroleptic era, the prognosis for full psychosocial recovery is still guarded, and families must be informed honestly of what can be expected.

SUGGESTED CROSS-REFERENCES

The other sections in this chapter on schizophrenia have a bearing on this section. A general discussion of clinical manifestations of psychiatric disorders appears in Chapter 8. Perception and cognition are discussed in Section 3.1. Psychological tests are discussed in Section 7.4 and Section 7.5. DSM-IV is discussed in Section 9.1. Delusional disorders are discussed in Section 13.2, schizoaffective disorders in Section 13.1, mood disorders in Chapter 14, cognitive disorders in Chapter 10, and somatoform disorders in Chapter 16. Personality disorders are discussed in Chapter 24, impulse-control disorders in Chapter 22, and unusual acute and transient psychotic disorders in Section 13.3. Suicide is discussed in Section 29.1. Antipsychotic drugs are covered in Section 31.17 and Section 31.26.

SECTION REFERENCES

Addington J, Addington D: Premorbid functioning, cognitive functioning, symptoms and outcome in schizophrenia. J Psychiatry Neurosci *18:*18, 1993.

*Amador XF, Kirkpatrick B, Buchanan RW, Carpenter WT, Marcinko L, Yale SA: Stability of the diagnosis of deficit syndrome in schizophrenia. Am J Psychiatry *156:*637, 1999.

Andreasen NC, Arndt S. Alliger R, Miller D, Flaum M: Symptoms of schizophrenia: Methods, meanings, and mechanisms. Arch Gen Psychiatry *52:*341, 1995.

Andreasen NC, Flaum M. Arndt S: The Comprehensive Assessment of Symptoms and History (CASH): An instrument for assessing diagnosis and psychopathology. Arch Gen Psychiatry *49:*615, 1992.

Andreasen NC, Olsen S: Negative versus positive schizophrenia: Definition and validation. Arch Gen Psychiatry *39:*789, 1982.

Andreasen NC, Roy M-A, Flaum MA: Positive and negative symptoms. In *Schizophrenia*, SR Hirsch, DR Weinberger, editors. Blackwell Science, Cambridge, England, 1995.

Anpermeyer MC, Katsehning H: Psychotropic medication and quality of life: a conceptual framework for assessing their relationship In *Quality of Life in Mental Disorders*. H Katschnig, HL Freeman, N Sortorious, editors. Wiley, New York, 1997.

Arieti S: *Interpretation of Schizophrenia*. Brunner, New York, 1955.

Bailer J, Brauer W, Rey ER: Premorbid adjustment as predictor of outcome in schizophrenia: Results of a prospective study. Acta Psychiatr Scand *93:*368, 1995.

*Bleuler E: *Dementia Praecox*, or, *The Group of Schizophrenias*, H Zinkin, translator. International Universities Press, New York, 1950.

Crow TJ: Molecular pathology of schizophrenia: More than one disease process? Br Med J *280:*66, 1980.

Fenton WS, McGlashan TH: Natural history of schizophrenia subtypes. Arch Gen Psychiatry *48:*969, 1991.

Freud S: *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Hogarth Press, London, 1953–1966.

*Friedman JI, Temporini H, Davis KL. Pharmacologic strategies for augmenting cognitive performance in schizophrenia. Biol Psychiatry *45:*1, 1999.

Gureje O, Aderibigbe YA, Obikoya O: Three syndromes in schizophrenia: Validity in young patients with recent onset of illness. Psychol Med *25:*715, 1995.

Hwu HG, Tan H. Chen CC, Yeh LL: Negative symptoms at discharge and outcome in schizophrenia. Br J Psychiatry *166:*61, 1995.

Jaspers K: The phenomenological approach in psychopathology. Br J Psychiatry *114:*1313, 1968.

Katshnig H: How useful is the concept of quality of life in psychiatry? In *Quality of Life in Mental Disorders*, H Katschnig, HL Freeman, N Sartorious, editors. Wiley, New York, 1997.

*Kraepelin E: Dementia praecox and paraphrenia. In *The 8th German Edition of the Textbook of Psychiatry*, vol III, part 2, *Endogenous Dementias*, RM Barclay, translator. Livingstone, Edinburgh,

1919.

Maurice G, Stephen B, Abbie L, Waddington JL, Larkin C, O'Callaghan E: Spontaneous abnormal involuntary movements in first-episode schizophrenia and schizophreniform disorder: Baseline rate in a group of patients from an Irish catchment area. Am J Psychiatry *155:*9, 1998.

Schneider K: *Clinical Psychopathology*, MW Hamilton, translator. Grune & Stratton, New York, 1959.

*Solomon LD, Goldman-Rateic PS: The reduced neuropil hypothesis: A circuit based model of schizophrenic. Biol Psychiatry *45:*17, 1999.

Thara R, Eaton WW: Outcome of schizophrenia: The Madras longitudinal study. Aust NZ J Psychiatry *30:*516, 1996.

Wieselgren IM, Lindstrom E, Lindstrom LH: Symptoms at index admission as predictor for 1–5 year outcome in schizophrenia. Acta Psychiatr Scand *94:*311, 1996.

CHAPTER 12. SCHIZOPHRENIA

12.8 SCHIZOPHRENIA: SOMATIC TREATMENT

STEPHEN R. MARDER, M.D.

The somatic treatment of schizophrenia has changed substantially during the 1990s. Until 1990 when clozapine (Clozaril) was introduced in the United States, all available antipsychotic drugs had a similar range of efficacy and were associated with neurological side effects that seriously interfered with their effectiveness. Clozapine was the first of a new generation of antipsychotics that are associated with far fewer extrapyramidal side effects than older drugs and perhaps have better efficacy. Although clozapine's association with agranulocytosis has limited the number of patients who receive it, this agent plays an important role in the treatment of severe psychosis. The introduction of risperidone (Risperdal) in 1994, olanzapine (Zyprexa) in 1996, quetiapine (Seroquel) in 1997, and ziprasidone (Zeldex) in 1998 have given clinicians new alternatives for treating a large number of patients with schizophrenia. The overall impact of these changes on the course of schizophrenia remains to be seen.

HISTORY

The history of somatic therapies in schizophrenia can be divided into two eras: before the discovery of chlorpromazine (Thorazine), the first clearly effective antipsychotic drug, and after. Prior to the introduction of antipsychotics in the early 1950s, several treatments had been administered to individuals with psychotic illness, with results that are difficult to interpret because careful research methods in psychiatry had not been developed. During the late nineteenth and early twentieth century schizophrenia was believed to deteriorate inevitably into dementia. As a result, patients were frequently hospitalized for long periods. Somatic treatments were used to help control the most severe symptoms of the disorder and to make hospitals safer. Sedating agents such as bromides and barbiturates were used to control agitation, and physical treatments such as hydrotherapy and wet sheet packs were also used for their calming effects. In the early 1920s sleep treatment with barbiturates was introduced. This treatment was based on the observation that symptoms tended to improve following an overdose of barbiturates. The method involved maintaining patients in a highly sedated state for days, during which they would awaken only for necessary activities such as eating and personal hygiene.

Insulin coma treatment was introduced during the 1930s. Patients were administered gradually increasing doses of insulin until a coma occurred. After being monitored for an hour the patient was administered glucose, which terminated the coma. Patients were commonly administered as many as 20 comas. Insulin coma was widely used in the treatment of psychosis, suggesting that it may have been somewhat effective. Unfortunately, it never received adequate research trials, and it remains unclear if the treatment was effective. It was abandoned when antipsychotics were introduced.

Prefrontal lobotomy was proposed as a treatment for serious mental illnesses by Egas Moniz in 1935. The support for this treatment came from animal studies in which frontal lobe extirpations in monkeys resulted in an animal that appeared less easily frustrated. The use of frontal lobotomy was common prior to the introduction of effective antipsychotics, although there is a remarkable lack of controlled studies comparing psychosurgery with other treatments. Although reports suggest that lobotomy may have been effective in reducing severe psychotic symptoms, they also resulted in deteriorations in other areas. Following lobotomies patients frequently demonstrated personality deterioration with impulsive and psychopathic behaviors as well as impaired concept formation and ability to plan. Psychosurgery was abandoned as a treatment for schizophrenia after the introduction of effective antipsychotic medications.

Convulsive therapies were developed after it was observed that some patients improved after a seizure. Drugs such as camphor and metrazol were used initially to induce seizures but were abandoned after Ugo Cerletti and Lucio Bini proposed the use of electrically induced convulsions. In its early days electroconvulsive therapy (ECT) was administered without anesthetics or muscle relaxants. The lack of anesthetics inspired fear in many patients, and the lack of muscle relaxants led to injuries from forceful muscle contractions. ECT administered with anesthesia and muscle relaxants continues to have a role in certain types of schizophrenia.

The first effective antipsychotic medications were probably derived from extracts of the rauwolfia plant. Publications from the 1930s and 1940s suggest that these agents were effective for both hypertension and psychosis. Reserpine (Serpasil) the most potent of the rauwolfia alkaloids, was introduced in the early 1950s and was widely prescribed in the United States and elsewhere for schizophrenia and other psychotic illnesses. Studies comparing reserpine with dopamine receptor antagonists suggested that their efficacies were similar. However, reserpine's adverse effects, particularly depression, led most clinicians to prefer the dopamine receptor antagonists. Thus, reserpine is only rarely used for managing psychosis.

The discovery of the phenothiazine chlorpromazine in the early 1950s may be the most important single contribution to the treatment of psychiatric illness. Laborit, a surgeon in Paris, noticed that administering chlorpromazine to patients prior to surgery resulted in an unusual state in which they seemed less anxious regarding the procedure. In 1952 he convinced Jean Delay and Pierre Deniker and other psychiatrists to administer chlorpromazine to psychotic and excited patients—the effects were extraordinary. Chlorpromazine was effective in reducing hallucinations and delusions as well as excitement. It also caused adverse effects that resembled parkinsonism. The use of chlorpromazine spread rapidly through the psychiatric hospitals in Paris and eventually to the rest of the world. Since chlorpromazine was relatively easy to administer to large numbers of patients, it was partially responsible for a substantial reduction in the number of patients in psychiatric hospitals.

Thioridazine and fluphenazine (Permitil, Proloxin) as well as other classes of drugs such as the butyrophenones (e.g., haloperidol [Haldol]) and the thioxanthenes (e.g., thiothixene [Navan]) were developed after the introduction of chlorpromazine. Although these newer agents differed in their potency and their adverse effect profiles, all had similar effectiveness. Clozapine, the first effective antipsychotic with negligible extrapyramidal effects was discovered in 1958 and first studied in the 1960s. However, in 1976 it was found to be associated with a substantial risk of agranulocytosis, which resulted in delays in its introduction. In 1990 clozapine finally became available in the United States, but its use was restricted to patients who responded poorly to other agents. Risperidone, olanzapine, quetiapine, ziprasidone, and other agents with affinity for both dopamine and serotonin receptors cause minimal extrapyramidal side effects and are not associated with a risk of agranulocytosis. These newer agents are replacing older drugs as the standard treatments for schizophrenia.

PHASES OF TREATMENT IN SCHIZOPHRENIA

Somatic treatment varies depending on the phase of a patient's illness. The acute stage is usually characterized by psychotic symptoms that require immediate clinical attention. These symptoms may represent a first psychotic episode or, more commonly, a relapse in an individual who has experienced multiple episodes. Treatment during this phase focuses on alleviating the most-severe psychotic symptoms. Following the acute phase, which usually lasts from 4 to 8 weeks, patients usually enter a stabilization phase in which acute symptoms have been controlled, but patients remain at risk for relapse if treatment is interrupted or if they are exposed to stress. During this phase, treatment focuses on consolidating therapeutic gains with treatments similar to those used in the acute stage. This phase may last as long as 6 months following recovery from acute symptoms. The third stage is the stable, or maintenance, phase when the illness is in relative remission. The goals during this phase are to prevent psychotic relapse and to assist patients in improving their level of functioning.

FOCUS OF TREATMENT

Patients with schizophrenia can demonstrate large differences in the severity of their psychopathology as well as in the type of symptoms they demonstrate. As a result, treatment strategies should be individualized to the characteristics of each patient's illness. Recent studies indicate that psychopathology in schizophrenia can

be classified into three dimensions: psychotic, negative, and disorganized. Psychotic symptoms include hallucinations, ideas of reference, and delusions. These are symptoms that tend to result in hospitalization and to disrupt the lives of patients. Negative symptoms include decreased motivation, emotional blunting, and impoverished speech and thought; these symptoms are associated with the social and vocational impairments of schizophrenia. Disorganized symptoms include disorganized speech and behavior as well as impairments in attention and information processing; these symptoms are also associated with the social and vocational impairments of schizophrenia.

Antipsychotic drugs are most effective in treating the psychotic dimension. As a result, clinicians adjusted their expectations about the goals of pharmacotherapy in schizophrenia. Although some improvement in the other dimensions often occurred with drug treatment, clinicians were usually satisfied when psychotic symptoms were minimized. These expectations changed following the introduction of clozapine. A proportion of patients who improved with clozapine demonstrated changes in other dimensions as well, including better social and vocational adjustments. Similar improvements have been reported on other serotonin-dopamine antagonists. In addition, other studies (discussed later) suggested that both negative and cognitive symptoms improved with these newer medications.

EFFECTIVENESS OF ANTIPSYCHOTIC MEDICATIONS

A large body of evidence supports the effectiveness of antipsychotics for schizophrenia. Many of these studies were carried out in the 1960s when there was skepticism that these agents were truly antipsychotic rather than more effective tranquilizers. An evaluation of these studies by the 1995 Schizophrenia Patient Outcomes Research Team (PORT) found that about 70 percent of patients treated with an antipsychotic achieved remission. In contrast, only about 25 percent of patients treated with placebo remitted. Most studies compared one or more antipsychotic with either a placebo or an agent such as phenobarbital that served as a control; antipsychotic drugs were found to be more effective than either placebo or tranquilizers.

EFFECTIVENESS OF ECT IN SCHIZOPHRENIA

ECT has been studied in both acute and chronic schizophrenia. Studies in patients with recent-onset schizophrenia indicate that ECT is about as effective as antipsychotic medications and more effective than psychotherapy. Other studies suggest that supplementing antipsychotic medications with ECT is more effective than antipsychotic medications alone. Studies of ECT in chronic schizophrenia have been less promising. Anecdotal reports indicate that ECT is effective in patients who respond poorly to antipsychotic medications. Overall these results suggest that ECT probably has a limited role in schizophrenia. Patients should first receive trials of antipsychotic medications; if these medications are ineffective, acutely ill patients can be treated with ECT. Antipsychotic medications should be administered during and following ECT treatment.

TREATMENT OF ACUTE EPISODES

Indications for Somatic Treatment Nearly all patients with acute psychotic symptoms benefit from an antipsychotic medication. Aside from relieving symptoms, evidence indicates that lengthy delays in initiating drug treatment may alter the long-term course of schizophrenia. This evidence is summarized in a scholarly review by Richard J. Wyatt, who found that treatment delays—usually of 6 months or more—were associated with a greater need for hospital treatment and a worse social and vocational outcome. Many of the studies reviewed by Wyatt have important limitations such as lack of randomization and comparing individuals treated during different decades. However, for ethical reasons a definitive study will never be carried out to determine if withholding treatment worsens the long-term course of schizophrenia so it is probably prudent for clinicians to consider the possibility that untreated psychosis can result in a type of permanent damage.

These data do not mean that all patients need to be treated immediately. In some circumstances the management of a patient may be better if drug treatment is delayed for several days. A brief delay may permit clinicians to make a more thorough diagnostic evaluation and rule out causes of abnormal behavior such as substance abuse, extreme stress, medical illnesses, and other psychiatric illnesses.

Assessment Whenever possible, patients should receive a physical examination with a neurological examination, a mental status examination, and a laboratory evaluation before medications are started. A urine screen for drugs of abuse and blood tests for complete blood count (CBC) electrolytes, glucose, and liver, renal, and thyroid function should be ordered. Other evaluations that should be considered are pregnancy tests in women, electrocardiograms (ECGs) when cardiac disease or age is a factor, and human immunodeficiency virus (HIV) and syphilis tests when relevant. The presence of movement disorders—particularly preexisting tardive dyskinesia—should be assessed because they may influence the selection of an antipsychotic.

Antipsychotics are relatively safe drugs so treatment can usually begin before the results of laboratory tests are known. An exception is clozapine treatment, which should only begin after the patient is known to have a normal CBC. Under emergent conditions—for example, when patients refuse to cooperate with an evaluation—antipsychotics can be administered prior to a medical evaluation.

Selection of an Antipsychotic Drug The introduction of new antipsychotic agents has made the selection of an antipsychotic much more complicated. Prior to the development of the new antipsychotics, all the drugs were equally effective for schizophrenia. Many clinicians believed that different subtypes of schizophrenia responded differently to different antipsychotics. For example, it was proposed that more agitated patients responded better to more-sedating drugs whereas more withdrawn patients responded better to less-sedating agents; however, controlled trials failed to support this. The differences among antipsychotics were confined to their side effects, the available formulations, and, to some extent, their cost. The newer antipsychotics challenged this view, suggesting that certain populations of individuals with schizophrenia were likely to do better on a newer antipsychotic.

Antipsychotic drugs can be categorized into two main groups: the older conventional ones, which have also been called *dopamine receptor antagonists* and the newer second-generation drugs which have been called *serotonin-dopamine antagonists (SDAs)*, or more broadly, atypical antipsychotics. This textbook uses the terms *dopamine receptor antagonist* and *SDA*, which refer to the theory that the antipsychotic effects of dopamine receptor antagonists result from the blockade of dopamine type 2 (D2) receptors. The SDAs differ in having effects related to their ratio of D_2 and serotonin (5-hydroxytrptamine [5-HT]) type 2A ($5\text{-}HT_{2A}$) antagonism. The dopamine receptor antagonists are further categorized as being low-, mid- or high-potency, with the higher-potency drugs having a greater affinity for D_2 receptors and a greater tendency to cause extrapyramidal side effects. Low-potency drugs are less likely to cause extrapyramidal side effects, but more likely to cause postural hypotension, sedation, and anticholinergic effects.

A number of factors should be considered in selecting an antipsychotic medication. Perhaps the most important consideration should be the patient's prior experience with drug treatment. This includes both the patient's clinical and subjective response. Regarding the subjective response, studies by Theodore Van Putten and others found that a patient's early response to a query such as “How does this medication agree with you?” was a powerful predictor of whether that patient would comply with taking that particular medication. In other words, if the patient has uncomfortable side effects on a medication, compliance is likely to be poor if that medication is prescribed.

Prior to the introduction of the SDAs few options were available for patients who developed extrapyramidal effects. At times, dosage reduction or changing the patient to a lower-potency dopamine receptor antagonist may be helpful. Unfortunately, many patients experience extrapyramidal or other adverse effects at the lowest effective dosage that is clinically effective. The introduction of the SDAs provides an opportunity for treating these individuals with agents that seldom cause extrapyramidal side effects at their effective dosage.

With these factors in mind, clinicians should consider the factors included in Table 12.8-1. In some cases these recommendations are based on incomplete data. For example, it remains unproven that patients with prominent negative or cognitive symptoms will respond better to an SDA than to a dopamine receptor antagonist. Haloperidol is recommended for pregnant patients because more data support its safety and not because it has proved safer than other drugs.

Factors	Considerations
Subjective response	A dyphoric subjective response to a particular drug predicts poor compliance with that drug
Sensitivity to extrapyrimidal adverse effects	A serotonin-dopamine antagonist (SDA)
Tardive dyskinesia	Clozapine or (possibly another SDA)
Poor medication compliance or high risk of relapse	Injectable form of a long-acting antagonist (haloperidol or fluphenazine)
Pregnancy	Probably haloperidol (most data supporting its safety)
Cognitive symptoms	Possibly an SDA
Negative symptoms	Possibly an SDA

Table 12.8-1 Factors Influencing Antipsychotic Drug Selection

An important nonclinical factor is the cost of the drug; the SDAs are much more expensive. However, evidence indicates that the higher drug costs of these agents may be offset by other factors. Studies from the Department of Veterans Affairs and State Hospitals in Connecticut found that patients treated with clozapine required fewer hospital days than patients treated with a conventional dopamine receptor antagonist. As a result, the reduction in hospital days compensated for the higher drug cost associated with clozapine treatment. Similar results from other studies suggest that the higher costs of risperidone and olanzapine may also be partially offset by a reduced need for hospitalization.

In selecting a drug for first-episode patients, clinicians should give a high priority to minimizing adverse effects. Many of these individuals are ambivalent about drug treatment and may discontinue antipsychotics when they experience relatively mild adverse effects. Unpleasant experiences with medications during this initial episode may be frightening to these individuals and may influence their future attitudes toward pharmacotherapy. These considerations may lead to the selection of an SDA or a relatively low dosage of a high-potency dopamine receptor antagonist. A number of studies indicate that both olanzapine and risperidone are effective for first-episode patients.

Route of Administration The decision regarding route of administration is usually straightforward. Under most conditions, patients should be treated with an oral antipsychotic agent. Most antipsychotic drugs have half-lives that permit a single daily dose. Short-acting intramuscular drugs are useful when the patient refuses oral dosing and when a rapid onset is helpful. Intramuscular administration of most antipsychotics results in peak plasma concentrations in about 30 minutes, with clinical effects emerging within 15 to 30 minutes. Most orally administered dopamine receptor antagonists yield a peak plasma concentration 1 to 4 hours following administration.

Antipsychotic medications can also be administered as long-acting injectable compounds. These drugs differ from short-acting compounds in that they have a very gradual onset of action and are eliminated very slowly. This route of administration is helpful for long-term maintenance therapy, but not for acute treatment because clinicians cannot titrate dosage against adverse effects or clinical effects when the onset of clinical effects may occur weeks or months after a drug or dosage change.

Prescribing Antipsychotics Prior to prescribing an antipsychotic drug, clinicians should describe the medication, its target symptoms, and its possible side effects. It is particularly important to describe adverse effects such as akathisia, which can be misinterpreted as agitation under some circumstances. Patients who are severely disturbed may be unable to participate meaningfully in this discussion. However, most patients benefit from information about the goals of treatment and important risks associated with antipsychotic medication. Patients with schizophrenia may be suspicious so it is particularly important to emphasize that they can participate in interpreting medication effects. Because psychotic individuals may be dependent on the help and support of their families, it is frequently helpful to involve one or more family members in decisions about drug treatment.

In some settings and locations patients must give written or verbal consent prior to receiving an antipsychotic medication. This can be a dilemma for patients who are conceptually disorganized and find it difficult to understand the risks and benefits of drug treatment. Under these circumstances, clinicians should adjust the complexity of the discussion to the patient's state of mind. Thus, it may be appropriate to provide a limited amount of information that focuses on the most common acute adverse effects of the medication when the patient is most seriously impaired. As the patient improves, clinicians may then elaborate on the costs and benefits of medication. For example, detailed discussions about tardive dyskinesia, an adverse effect associated with chronic treatment, may be deferred until the patient has improved and long-term maintenance is being considered.

Psychiatrists must also evaluate whether acutely disturbed patients can participate meaningfully in decisions about their medication. Clinicians should become familiar with local and state laws that affect a patient's right to refuse or accept drug treatment. The most difficult situation is when a patient who desperately needs medication refuses it. Under some conditions, family members who have been educated about schizophrenia may be helpful in convincing patients to accept medication. Every locality has provisions for treating patients against their will under emergency conditions; some areas permit involuntary treatment when certain conditions are met. As patients improve, the great majority eventually accept their own need for medication.

Dosage Selection Finding the best dosage of an antipsychotic is both difficult and important. It is important because these agents, particularly the dopamine receptor antagonists, may cause adverse effects at their effective dosages. Often the clinician must weigh the therapeutic advantages of a particular dosage against uncomfortable or disabling side effects. Figure 12.8-1, from a study by Theodore Van Putten and his colleagues, displays dose-response curves for fluphenazine for both clinical improvement and disabling side effects. These two curves are close together, which indicates that it may be difficult for the clinician to find a dosage that results in clinical improvement without substantial adverse effects.

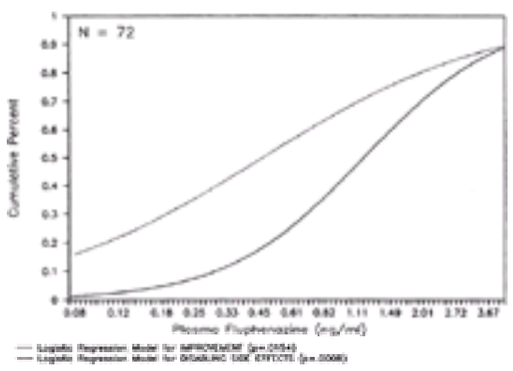

FIGURE 12.8-1 Improvement and disabling side effects as a function of plasma fluphenazine concentration.

Finding the right dosage is difficult because the physician cannot titrate dosage against clinical effects because of the delay between a clinical intervention and the patient's clinical response. Some individuals experience a delay of days or even weeks between the time treatment is started and when the patient eventually responds. These studies are supported by findings indicating that the neurochemical response to an antipsychotic agent is complex and includes an initial blockade of central dopamine receptors, followed by delayed decrease in dopamine turnover.

Although high doses of dopamine receptor antagonists can be associated with extrapyramidal side effects, some patients can tolerate antipsychotics at very high dosages. This is particularly true of nonsedating, high-potency drugs, which has led clinicians to raise the prescribed dosage in hope that higher dosages will lead to greater improvement than moderate dosages. This belief resulted in a substantial increase in the average dosage of antipsychotic drugs prescribed in the United States during the 1970s and 1980s. Many psychiatrists during this period routinely prescribed dosages above 1000 mg a day of chlorpromazine equivalents (20 mg of

haloperidol), whereas others reserved high-dosage treatment for patients who remained symptomatic on lower dosages of medication.

A number of dosage comparison studies have failed to support the routine use of higher doses; that is, when groups of patients are assigned to higher dosages (e.g., more than 2000 mg a day of chlorpromazine or 40 mg a day of haloperidol), the rate of improvement and the amount of improvement are no greater than for those assigned to more moderate dosages. Clinicians are sometimes impressed by individuals who require these higher dosages, suggesting that there is a small group of patients who should be treated with high dosages. However, most patients who receive these high dosages are only partial responders to an antipsychotic and have endured dosage increases that were not associated with improvement.

Dosage comparison studies indicate that dosages below 300 mg a day of chlorpromazine (or 5 mg a day of fluphenazine or haloperidol) are likely to be too low for many psychotic patients. At the same time doses above 1000 mg a day of chlorpromazine (or 20 mg a day of fluphenazine or haloperidol) are seldom necessary and may lead to substantial adverse effects.

Only limited data from controlled trials exist to assist clinicians in finding the best dose of clozapine. The mean dose of clozapine prescribed differs between Europe and the United States, with European physicians commonly prescribing less than 300 mg of clozapine daily and clinicians in the United States often prescribing 500 mg or more. These experiences support the practice of treating most clozapine patients with doses in the range of 300 to 500 mg daily. However, adverse effects, particularly sedation and orthostatic hypotension, are often limiting factors that prevent clinicians from reaching a targeted dosage. Although some patients have an optimal response at dosages between 600 and 900 mg daily, the risk of seizures increases substantially in this dosage range. More-recent studies suggest that patients are more likely to respond to clozapine when plasma concentrations are 350 ng/mL or higher, suggesting that measuring plasma concentrations may be useful for poor responders.

Large multicenter trials indicate that risperidone is most effective at 4 to 8 mg daily. Higher doses may lead to extrapyramidal effects without increased effectiveness. In the United States, the average dosage of risperidone prescribed for schizophrenia is slightly more than 4 mg daily. This suggests that a reasonable practice would be to manage patients with schizophrenia with 4 mg of risperidone and increase the dosage if they fail to respond after 4 to 6 weeks.

Most patients with acute schizophrenia can be managed on dosages between 15 and 25 mg daily of olanzapine. Some individuals respond well to as little as 5 or 10 mg daily. Quetiapine is usually effective when dosages are between 150 and 600 mg daily. Ziprasidone is effective at dosages of 80 to 160 mg daily.

A number of recent findings suggest a reasonable strategy for treating acute schizophrenia. The dose of an antipsychotic that is likely to be effective is the dose that occupies an appropriate number of D_2 receptors. For dopamine receptor antagonists this is approximately 80 percent of receptors. The therapeutic response depends upon processes that occur after these receptors have been occupied for a period of time. This observation is supported by findings from both position emission tomography (PET) scanning and the measurement of plasma homovanillic acid, which suggest that clinical improvement is not associated with the immediate effects of the drug on dopamine receptors, but on processes that occur later. Therefore, the goal in the first days of treatment is to prescribe a drug dosage that occupies an adequate proportion of dopamine receptors and to keep the patient comfortable until the drug is effective. If a patient does not respond in the first week or two, this does not indicate that the current treatment is inadequate. Since most patients on antipsychotic drugs improve during the first 6 weeks, patients should be observed for this interval before a drug is changed. Also, the strategy of using medications on an as-needed basis as a guide to finding the optimal dosage makes very little sense because the immediate and delayed responses to the drug are very different.

A comparison of some antipsychotic drugs is presented in Table 12.8-2.

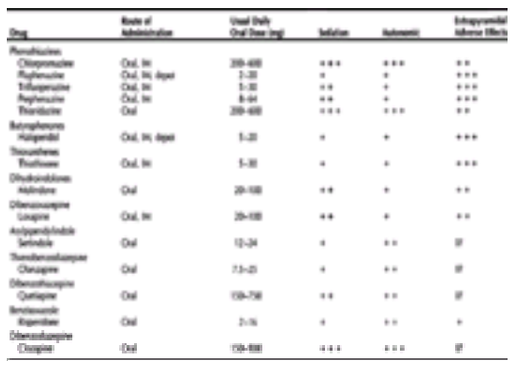

Table 12.8-2 Selected Antipsychotic Drugs

Managing Agitation in Acute Psychosis Agitation in acute schizophrenia can result from disturbing psychotic symptoms such as frightening delusions or suspiciousness or from other causes, including stimulant abuse or extrapyramidal side effects, particularly akathisia. Patients with akathisia can appear agitated when they experience a subjective feeling of motor restlessness. Differentiating akathisia from psychotic agitation can be difficult, particularly when patients cannot describe their internal experience. A trial with an anticholinergic antiparkinson medication or propranolol may be helpful in making the discrimination.

Clinicians have a number of options for managing agitation that results from psychosis. Antipsychotic drugs and benzodiazepines can result in relatively rapid calming when psychotic patients are agitated. An advantage of an antipsychotic agent is that a single intramuscular injection of a high-potency drug such as haloperidol or fluphenazine can result in calming without excess sedation. Low-potency antipsychotics are often associated with sedation and postural hypotension, particularly when they are administered intramuscularly. The disadvantage of high-potency drugs is that extrapyramidal effects can result from a single injection or, more often, from repeated injections. In younger patients, excessive amounts of injected high-potency drugs can lead to dystonia, which may increase the patient's agitation.

A reasonable intervention for the agitated patient is to treat the agitation with either an intramuscular or oral antipsychotic. If the situation is urgent, intravenous or intramuscular drug administration will lead to more rapid calming. The patient should be started on a regimen of oral antipsychotic the same day. If further treatment of agitation is necessary, benzodiazepines may be administered. Lorazepam (Ativan) has the advantage of reliable absorption when it is administered either orally or intramuscularly. The combination of lorazepam with a high-potency antipsychotic agent has been found to be safer and more effective than large doses of antipsychotics in controlling excitement and motor agitation. Moreover, the use of benzodiazepines may reduce the amount of antipsychotic medication that is needed to control psychotic patients.

MANAGING ADVERSE EFFECTS

Patients frequently experience the adverse effects of an antipsychotic agent before they experience clinical improvement. Whereas a clinical response may be delayed for days or weeks after drugs are started, adverse effects often begin almost immediately. For low-potency drugs, these adverse effects are likely to include sedation, postural hypotension, and anticholinergic effects, whereas high-potency drugs are likely to cause extrapyramidal side effects.

This early onset of adverse effects is important because a patient's interpretation of a drug's effectiveness is often associated with how that drug makes them feel. Moreover, one of the challenges of treating the acutely psychotic is maintaining the trust of individuals who may misinterpret experiences and become suspicious. Warning patients about the potential side effects of medication can lead to prompt management and often improves the trust between patient and clinician. Moreover, minimizing the adverse effects may do long-lasting damage to the patient-clinician relationship because one of the powerful predictors of drug reluctance or drug refusal is an earlier experience of adverse effects.

Extrapyramidal Side Effects The most common form of extrapyramidal side effect is neuroleptic-induced acute akathisia, an adverse effect consisting of a subjective feeling of restlessness along with restless movements, usually in the legs or feet. Patients who experience severe akathisia often pace continuously or move their feet restlessly while they are sitting. Some complain that they are unable to feel comfortable, regardless of what they do. Severe akathisia can make patients feel anxious or irritable, and some reports suggest that severe akathisia can result in aggressive or suicidal acts. Researchers have estimated that 25 to 75 percent of patients treated with a high-potency dopamine receptor antagonist experience akathisia. This adverse effect can be difficult to assess and is frequently misdiagnosed as

anxiety or agitation. Akathisia is also thought to be a correlate of poor antipsychotic drug response.

Because patients may experience akathisia as irritability or agitation, asking patients whether they are restless or if they have difficulty sitting still can be helpful in early stages of treatment. At this point, a dosage adjustment, a b-adrenergic receptor antagonist, or an anticholinergic drug may provide considerable relief. Also, patients who have a history of developing severe akathisia that responds poorly to these treatments are likely to do better if they are treated with a new antipsychotic such as clozapine, risperidone, or sertindole.

Neuroleptic-induced acute dystonia is probably the most frightening extrapyramidal side effect. It consists of intermittent or sustained muscular spasms and abnormal postures affecting mainly the musculature of the head and neck, but sometimes the trunk and lower extremities. Common forms of dystonia include abnormal positioning of the neck, impaired swallowing (dysphagia), hypertonic or enlarged tongue, and deviations of the eyes (oculogyric crisis). These reactions usually appear within the first few days of therapy. Dystonias are more likely to occur in younger patients, particularly males in their teens or 20s.

Neuroleptic-induced parkinsonism consists of tremor, muscular rigidity, and a decrease in spontaneous movements, features that resemble the movement disorder in idiopathic parkinsonism. Examination usually reveals a positive glabella tap. This motor disturbance affects about 30 percent of patients who are chronically treated with traditional antipsychotics. The first evidence of drug-induced parkinsonism may be a diminished arm swing or decreased facial expressiveness. Risk factors for antipsychotic-induced parkinsonism include increasing age, dosage, a history of parkinsonism, and underlying basal ganglia damage.

When patients develop neuroleptic-induced parkinsonism, clinicians have a number of alternatives. These include reducing the dosage of the antipsychotic (which is most commonly a dopamine receptor agonist), adding an antiparkinsonism medication, or changing the patient to an SDA that is less likely to cause extrapyramidal adverse effects. The most effective antiparkinsonism medications are the anticholinergic drugs. Although these medications are frequently effective, they also cause their own adverse effects including dry mouth, constipation, blurred vision, and often memory loss. Also, these drugs are often only partially effective, leaving patients with substantial lingering extrapyramidal side effects. Centrally acting b-adrenergic receptor antagonists such as propranolol (Inderal) are frequently effective for treating akathisia; most patients respond to daily dosages between 30 and 90 mg.

Clinicians may consider prescribing prophylactic antiparkinsonism medications for patients who are likely to experience disturbing extrapyramidal effects. These include patients who have a history of extrapyramidal sensitivity or those who are being treated with relatively high dosages of high-potency drugs. Prophylactic antiparkinsonism medications may also be indicated when high-potency drugs are prescribed for young men who tend to have an increased vulnerability for developing dystonias; these patients may also be candidates for the newer drugs.

Some individuals are highly sensitive to extrapyramidal adverse effects at doses that are necessary to control their psychosis. For many of these patients, the adverse effects of the medication may seem worse than the illness itself. These patients should routinely be treated with an SDA because these agents result in substantially fewer extrapyramidal adverse effects than the dopamine receptor antagonists do. These highly sensitive individuals may actually experience extrapyramidal side effects on an SDA. Risperidone may cause extrapyramidal effects at higher dosages, (e.g., above 6 mg) and olanzapine, quetiapine, and ziprasidone may cause akathisia at their higher dosages.

Tardive Dyskinesia and Other Tardive Syndromes Chronic treatment with an antipsychotic—usually for 6 months or more—can result in movement disorders including neuroleptic-induced tardive dyskinesia and other tardive disorders. Tardive dyskinesias commonly consist of abnormal, involuntary movements of the mouth, tongue, trunk, and extremities. The oral-facial movements occur in about three-fourths of patients with tardive dyskinesia and may include lip smacking, sucking, and puckering as well as facial grimacing. Other movements may include irregular movements of the limbs, particularly choreoathetoid-like movements of the fingers and toes and slow, writhing movements of the trunk. Younger patients with tardive dyskinesia tend to develop slower athetoid movements of the trunk, extremities, and neck.

The abnormal movements of tardive dyskinesia are usually reduced by voluntary movements of the affected areas and are increased by voluntary movements of unaffected areas. The abnormal movements of tardive dyskinesia are usually increased with emotional arousal and absent when the individual is asleep. According to the research criteria in DSM-IV the abnormal movements should be present for at least 4 weeks and patients should have been exposed to an antipsychotic agent for at least 3 months (see Table 31.4-5). The onset of the abnormal movements should occur either while the patient is receiving an antipsychotic agent, within 4 weeks of discontinuing an oral agent, or 8 weeks after the withdrawal of a depot antipsychotic drug.

Prevalence surveys indicate that 20 to 30 percent of patients who are chronically treated with a dopamine receptor antagonist exhibit symptoms of tardive dyskinesia. Three to 5 percent of young patients receiving a dopamine receptor antagonist develop tardive dyskinesia each year. The risk in elderly patients is much higher. Although seriously disabling dyskinesia is uncommon, a small proportion of patients have trouble walking, breathing, eating, and talking. Individuals who are more sensitive to acute extrapyramidal effects appear to be more vulnerable to developing tardive dyskinesia. Patients with cognitive disorders and mood disorders may also be more vulnerable to developing tardive dyskinesia than those with schizophrenia.

All dopamine receptor antagonists are associated with a risk of tardive dyskinesia. Evidence from prospective studies indicates that clozapine is associated with a substantially lower risk than dopamine receptor antagonists. Anecdotal evidence indicates that clozapine can decrease abnormal movements in some patients with tardive dyskinesia. At this time it is unclear if the other SDAs are also associated with a lower risk than the dopamine receptor antagonists. The argument has been made that because these drugs are less likely to cause extrapyramidal side effects, they are less likely to cause tardive dyskinesia. For risperidone and olanzapine, prospective studies indicate new cases of tardive dyskinesia develop at a lower rate on these agents than on haloperidol. However, these studies are relatively small and include few cases of tardive dyskinesia. Until these findings are replicated with these agents or other SDAs, clozapine is the drug of choice for individuals who suffer from disabling tardive dyskinesia.

A Task Force on Tardive Dyskinesia of the American Psychiatric Association recently made a number of recommendations for preventing and managing tardive dyskinesia. These include (1) establishing objective evidence that antipsychotic medications are effective for an individual; (2) using the lowest effective dosage of an antipsychotic drug; (3) prescribing cautiously with children, elderly patients, and patients with mood disorders; (4) examining patients on a regular basis for evidence of tardive dyskinesia; (5) considering alternatives to antipsychotic agents, obtaining informed consent, and also considering dosage reduction in patients who develop tardive dyskinesia; (6) if the tardive dyskinesia worsens consider a number of options, including discontinuing the antipsychotic drug, switching to a different drug, or a trial of clozapine.

Regular monitoring for tardive dyskinesia should be a component of management strategies with antipsychotic drugs. The monitoring should be particularly careful for patients who have an increased risk for tardive dyskinesia, including elderly patients, patients who are sensitive to extrapyramidal side effects, and individuals with affective illness. Routine monitoring should include examination every 3 to 6 months, and high-risk groups should be monitored every 3 months.

A summary of extrapyramidal syndromes is presented in Table 12.8-3.

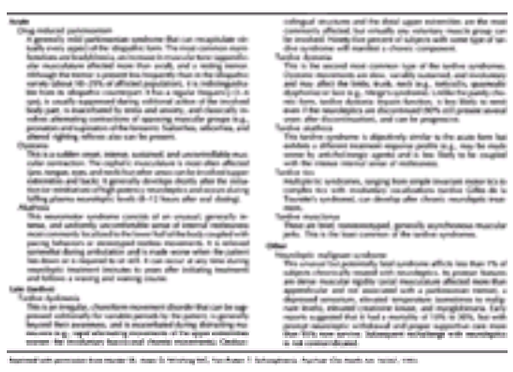

Table 12.8-3 The Drug-Induced Extrapyramidal Syndromes

Other Adverse Effects Sedation and postural hypotension can be important adverse effects for patients who are being treated with low-potency DRAs, such as chlorpromazine and thioridazine, and clozapine. These effects are often most severe during the initial dosing with these medications, so it may be weeks before patients treated with these medications—particularly clozapine—reach a therapeutic dosage. Although most patients develop tolerance to sedation and postural hypotension, sedation may continue to be a problem. Daytime drowsiness may interfere with such patients attempts to return to community life.

All of the dopamine receptor antagonists as well as risperidone increase prolactin concentrations, which can result in galactorrhea and irregular menses. Evidence suggests that prolactin elevation may impair libido in men and women. Fortunately, clozapine, olanzapine, quetiapine, and ziprasidone do not appear to elevate prolactin above normal concentrations. As a result, patients who demonstrate these symptoms on a dopamine receptor antagonist or risperidone and have high prolactin concentration should instead be given an agent that does not increase prolactin concentration.

Many antipsychotic drugs cause disturbances in sexual function including ejaculatory or erectile disturbances in men and decreased libido in women; these effects result in a substantial amount of noncompliance in men. Thus clinicians should discuss issues of sexual functioning with patients and intervene when possible. It is unclear if any of the SDAs are associated with a reduced or greater risk of sexual dysfunction than the dopamine receptor antagonists.

Adverse Effects of Clozapine Clozapine has a number of adverse effects that make it a difficult drug to administer. The most serious adverse effect is a risk of agranulocytosis, a potentially fatal condition that occurs in approximately 1 percent of patients treated with clozapine. As a result, patients who receive clozapine in the United States are required to be in a program of weekly blood monitoring for as long as they receive the drug.

Clozapine is also associated with a higher risk of seizures than other antipsychotics. The risk reaches nearly 5 percent at doses over 600 mg. Patients who develop seizures while on clozapine can usually be managed by reducing the dosage and adding an anticonvulsant, usually a form of valproic acid (Depakene) or divalproex (Depakote). Other adverse effects with clozapine include hypersalivation, sedation, tachycardia, weight gain, fever, and postural hypotension.

NEGATIVE AND COGNITIVE SYMPTOMS

Negative symptoms and cognitive impairment are associated with a substantial number of the social and vocational impairments of schizophrenia. This observation has resulted in a reappraisal of the goals of treatment, with a greater emphasis being placed on treatment strategies for decreasing the severity of these impairments. Most of the attention has focused on negative symptoms.

Carpenter has made an important contribution to this area by classifying negative symptoms into primary and secondary categories. *Secondary negative symptoms* are those that may result from other conditions, such as depression or extrapyramidal side effects. The latter are a common cause of secondary negative symptoms, particularly when patients are experiencing akinesia, an adverse effect that can be manifest in decreased speech, decreased motivation, and decreased spontaneous gestures. In addition, positive or psychotic symptoms may result in secondary negative symptoms. A common example is the patient who is withdrawn or uncommunicative as a result of suspiciousness.

The management of secondary negative symptoms begins with the management of the condition that caused these symptoms. For depression this may include the addition of an antidepressant medication; for extrapyramidal effects this may involve the addition of an antiparkinson medication, a dose reduction, or a change to an antipsychotic—usually an SDA—that is associated with less extrapyramidal adverse effects.

If the previously mentioned causes of secondary negative symptoms have been ruled out, the patient is likely to be demonstrating a type of enduring primary negative symptom. Some evidence suggests that the SDAs are more effective in treating negative symptoms than conventional agents. However, it is unclear if these effects are related to a reduction in secondary negative symptoms. Until this issue is decided by adequate controlled studies, it is reasonable for clinicians to consider prescribing an SDA to patients who have substantial negative symptoms.

Patients with schizophrenia frequently suffer from impairments in attention and information processing. These cognitive impairments can also interfere with their social and vocational rehabilitation, even when their psychotic symptoms have been well controlled. As with negative symptoms, cognitive impairments can be due to other causes such as substance abuse or drug effects of medications. The anticholinergic effects of either an antipsychotic or an antiparkinsonism medication such as biperiden (Akineton) or benztropine (Cogentin) can cause cognitive impairments that are difficult to distinguish from symptoms that are part of the schizophrenia. Decreasing the use of anticholinergic medication by changing to drugs that do not require antiparkinsonism medications—particularly SDAs—may be helpful. Evidence also suggests that clozapine, risperidone, olanzapine and other SDAs may be more effective at treating cognitive impairments than dopamine receptor antagonists. For example, a recent study found risperidone effective in improving verbal working memory. Others have reported that clozapine is effective for improving verbal fluency. Olanzapine has resulted in improvements in a number of cognitive measures. If confirmed, these findings will support the practice of treating cognitively impaired patients with schizophrenia with an SDA.

STRATEGIES FOR POOR RESPONDERS

When patients with acute schizophrenia receive an antipsychotic medication approximately 60 percent improve to the extent that they will achieve a complete remission or experience only mild symptoms; the remaining 40 percent of patients improve, but still demonstrate variable levels of positive symptoms that are resistant to the medications. Rather than categorizing patients into responders and nonresponders, it is more accurate to consider the degree to which the illness is improved by medication. Some resistant patients are so severely ill that they require chronic institutionalization; others respond to an antipsychotic drug with substantial suppression of their psychotic symptoms but demonstrate persistent hallucinations or delusions.

Before considering a patient a poor responder to a particular drug one must be certain that they received an adequate trial of the medication. A 6-week rial on an adequate dosage of an antipsychotic agent is considered reasonable for most patients. If patients demonstrate even mild improvement during this period, it may be reasonable to wait because data indicate that patients may improve at a steady rate for 3 to 6 months. It may be also be helpful to confirm that the patient is receiving an adequate amount of the drug by monitoring the plasma concentration. Information about therapeutic plasma concentrations is available for a number of antipsychotic drugs including haloperidol, clozapine, fluphenazine, trifluoperazine (Stelazine), and perphenazine (Trilafon). A very low plasma concentration may indicate that a patient has been noncompliant or, more commonly, only partially compliant. It may also suggest that the patient is a rapid metabolizer of the drug or that the drug is not being adequately absorbed. Under these conditions increasing the dose may be helpful. If the level is already relatively high, clinicians should consider whether adverse effects may be interfering with therapeutic response.

If the patient is responding poorly, many clinicians will consider raising the dosage above the usual therapeutic level. The use of high dosage in poor medication responders has been studied under a number of circumstances. Nearly all studies found that higher dosages were not associated with greater improvement than conventional dosages, which suggests that changing to another drug is more likely to be helpful than increasing the dosage.

A patient who has responded poorly to a conventional dopamine receptor antagonist is unlikely to do well on another dopamine receptor antagonist. Studies suggest that a poor response to one dopamine receptor antagonist is likely to be followed by a poor response to another; thus, changing to an SDA is more likely to be helpful.

Substantial evidence indicates that clozapine is effective for patients who respond poorly to dopamine receptor antagonists. Double-blind studies comparing clozapine to other antipsychotic agents indicated that clozapine had the clearest advantages over conventional drugs in patients with the most severe psychotic symptoms as well as those who had previously responded poorly to other antipsychotic drugs. The most definitive evidence of clozapine's advantages in this population comes from a multicenter trial reported by John Kane in which clozapine was compared with chlorpromazine. This study was conducted on severely psychotic patients who had failed in trials with at least three antipsychotic drugs. Clozapine was significantly more effective than chlorpromazine in nearly every dimension of psychopathology, including both positive and negative symptoms. This study found that 30 percent of patients treated with clozapine met improvement criteria by the end of the 6-week trial. Studies of longer duration indicate that 60 percent of patients are likely to meet these same improvement criteria when patients are maintained on clozapine for 6 months.

There is also evidence suggesting that risperidone and olanzapine may be helpful when a dopamine receptor antagonist is only partially effective. A Swiss study found that clozapine and risperidone were equally effective in a treatment-resistant population, but risperidone's side effects were more easily tolerated. Another study found that risperidone was somewhat more effective than haloperidol in a similar population and that risperidone was better tolerated. A multicenter comparison of olanzapine and haloperidol in a largely treatment-resistant group of patients found that olanzapine was more effective for both positive and negative symptoms. Taken

together these studies support the practice of trying patients on risperidone or olanzapine when they have responded poorly to a dopamine receptor antagonist.

When switching patients from one antipsychotic to another clinicians should be aware that abrupt changes in drugs and dosage can have serious adverse effects in some individuals. For example, abrupt discontinuation or rapid dosage reduction of low-potency antipsychotics such as clozapine or chlorpromazine can lead to withdrawal adverse effects such as nausea or diarrhea. Anecdotal reports indicate that rapid discontinuation of clozapine can lead to severe psychotic relapse. When treatment is started with some antipsychotics such as clozapine or quetiapine, dosage titration often requires days or even weeks. For individuals who are being changed to treatment with one of these agents, the best strategy usually entails continuing the first agent until the patient is receiving a clinically effective dosage of the second drug. For these reasons, the best methods for switching drugs usually involve a cross-titration in which the patient is gradually changed from one antipsychotic to the other.

MAINTENANCE THERAPY

During the stable or maintenance phase, patients are usually in a relative state of remission, with only minimal psychotic symptoms. The goals during this stage are to prevent patients from suffering psychotic relapse and to assist them to improve their level of functioning. Pharmacotherapy plays an important part in both of these goals. Medications are effective in preventing or delaying psychotic relapse and may also be an important adjunct in managing functional impairments that may interfere with rehabilitation; unfortunately, the adverse effects of medications can undermine these goals.

Drug and Route of Administration for Maintenance Therapy Stable patients maintained on an antipsychotic drug have a much lower relapse rate than patients whose medications are discontinued. Although studies differ, most suggest that 16 to 23 percent of patients a year experience a relapse while receiving medications and 53 to 72 percent relapse without medications. Clinicians are often tempted to discontinue medications in patients who have een well and stable for prolonged periods; unfortunately, these patients also have high relapse rates when their medications are discontinued. Other evidence indicates that patients who experience relapses while they are receiving an antipsychotic drug have milder episodes than patients who relapse on no medication. Donald Johnson has reported that patients whose medications are discontinued are more likely to show dangerous behavior and are more likely to be admitted involuntarily.

These observations about the effectiveness of continuing antipsychotic medication in stable patients have led to the recommendation that most patients with schizophrenia should receive an antipsychotic to prevent relapse. In 1989 the duration of maintenance was considered at an international consensus conference. This consensus group recommended 1 to 2 years of maintenance for patients following a first episode. Although this may be somewhat longer than current practice in many settings, it was recommended because individuals at this stage of their illness may have the most to gain if relapse can be prevented or delayed. First-episode patients may be working or involved in educational programs, both of which can be jeopardized by a second psychotic episode. The consensus conference also recommended that multiepisode patients receive maintenance antipsychotic for at least 5 years. For patients with a history of serious suicide attempts or violent, aggressive behavior, maintenance treatment with neuroleptics may be indicated for longer periods—perhaps indefinitely. The first 3 to 6 months following an acute episode was considered a period of stabilization when patients may not demonstrate acute symptoms but may nevertheless remain more vulnerable to relapse. The consensus conference recommended that following this stabilization period gradual dosage reduction should be implemented at the rate of approximately 20 percent every 6 months until a minimal maintenance dosage level is reached.

There is also evidence that long-acting depot antipsychotic drugs may be the most effective agents for preventing relapse. A number of double-blind and uncontrolled trials have compared oral and depot treatment. The uncontrolled trials usually compared individuals who were assigned to a depot with those assigned to oral medications; patients assigned to depot medication usually demonstrated much lower rates of relapse. The results are less clear for well-designed double-blind studies. However, these studies tended to use highly selected groups of cooperative patients who were carefully monitored, which would tend to undermine the advantages of depot medications by excluding the patients who were most likely to benefit from this form of treatment: that is, patients with compliance problems. Nevertheless, even under these conditions, in a meta-analysis of six studies Janicak found an advantage in favor of depot treatment.

Dosage Reduction Strategies Concerns about the long-term adverse effects of antipsychotic medications—particularly extrapyramidal effects and tardive dyskinesia—have led to a search for methods of treating patients with the lowest effective dose of medication. One strategy proposes using substantially lower antipsychotic dosages during maintenance treatment than those prescribed for initial short-term treatment. Studies indicate that many patients do well when they are treated with dosages that are approximately 20 percent of an initial treatment dosage. The low dosages were in the range of 4 to 10 mg of fluphenazine decanoate administered every 2 weeks. In 1997 Nina Schooler and her coworkers published the results from the Treatment Strategies in Schizophrenia multicenter study. Their 2-year comparison of low and conventional dosages found that low dosages in this range resulted in relapse rates that were slightly higher than those with conventional dosages but within an acceptable range. Moreover, other studies suggest that lower dosages are associated with milder side effects and better patient compliance.

Another strategy, termed targeted or intermittent therapy, proposes gradually reducing and finally discontinuing the medications in stable patients. Patients are then monitored carefully, and medications are reintroduced if early signs of relapse appear. Controlled trials with this strategy have mostly shown discouraging results. Relapse rates were relatively high as were rates of rehospitalization. These results were confirmed in the Treatment Strategies Study in which rates of rehospitalization were significantly elevated in the targeted treatment group.

Still another strategy combines the features of low- and targeted-dose strategies. In this method patients are treated with the same low doses of a depot drug that were used in the low-dose studies. Patients are monitored for early prodromal signs of relapse, and if these symptoms appear, patients are treated with oral medication. Results suggest that this is an effective strategy for making low-dose treatment safer.

Although the SDAs may be excellent drugs for maintenance treatment, few controlled long-term trials have evaluated their effectiveness. Nevertheless, risperidone, olanzapine, quetiapine, and ziprasidone appear to have important advantages. Concern about adverse effects should be substantially less, which would allow clinicians to treat patients with dosages that should be associated with very low relapse rates. If findings suggesting that SDAs are associated with a reduced risk of tardive dyskinesia are confirmed, this advantage will provide a compelling reason to select these newer agents. Also, the improved side-effect profile may result in better medication compliance.

The SDAs also have serious adverse effects that are likely to be important concerns during long-term treatment. Some of the newer agents—particularly olanzapine and clozapine—can result in substantial weight gain. Risperidone can cause some extrapyramidal effects and may cause problems related to increased prolactin concentrations such as irregular menstrual periods and galactorrhea. An important disadvantage of the SDAs is the lack of availability of long-acting formulations that are useful for patients who are unreliable pill takers.

INTEGRATING PHARMACOTHERAPY AND PSYCHOSOCIAL TREATMENT

Most patients with schizophrenia will benefit from a combination of pharmacotherapy and psychosocial treatments. Recent improvements in both domains suggest that the overall outcome of this disorder can be improved if patients receive the optimal forms of both treatments at the appropriate stage of their illness. Research studies and clinical experience suggest that psychosocial treatments are probably most effective when patients have recovered from severe psychotic episodes. During the acute psychotic phase clinical management should emphasize maintaining patient cooperativeness and trust. This is particularly important when there is overt suspiciousness or a tendency to misinterpret the intentions of the treatment team. A successful strategy is likely to include clear explanations of the rationale for treatment and possible drug adverse effects. Family members may be important allies in ensuring cooperation, and family psychoeducation programs have been demonstrated to be helpful during this phase.

It is difficult to generalize about the interactions of drugs and psychosocial treatments for stable patients because psychosocial treatments can vary greatly in terms of content and goals. Nevertheless, a number of important treatment principles can be drawn from the literature on combining treatments. The first is that psychosocial treatments are most likely to be effective when patients have been effectively stabilized on drugs. Early studies by Hogarty ndicated that psychosocial treatments could actually lead to a worse outcome when outpatients with schizophrenia were treated with a placebo. Other studies indicate that patients are most likely to respond to psychosocial treatments when their condition is stable. For example, a recent study with social skills training found that patients who received pharmacotherapy that minimized the proportion of time that they were in a psychotic state also demonstrated the greatest improvements in social adjustment.

Psychosocial treatments may also improve patient response to pharmacotherapy by improving medication compliance. This was suggested in a study in which patients received a form of family treatment that also encouraged medication compliance. Other studies have indicated that psychosocial treatments—particularly family treatment—may decrease the amount of stress that the patient experiences within the family and that this, in turn, decreases the amount of antipsychotic

medication required by the patient.

The introduction of the newer antipsychotics may result in much greater interest in psychosocial interventions. Patients who receive the newer agents may be better candidates for psychosocial treatments when treatment with these agents is associated with improvements in negative and cognitive symptoms as well as reduced adverse effects. Also, patients who improve on clozapine, risperidone, olanzapine, or other drugs may initially appear ready to return to community life. However, these individuals then experience a series of frustrating failures at work, school, or social relationships, which indicate that drug therapy alone may not suffice to prepare them for their new roles.

SUGGESTED CROSS-REFERENCES

For further information related to assessment of the patient with schizophrenia see Section 7.3 on typical signs and symptoms of psychiatric illness and Section 7.8 on psychiatric rating scales. Chapter 12 on schizophrenia is important for a full understanding of the syndrome. To appreciate the antipsychotic medications see Section 31.17 and Section 31.26 on antipsychotic drugs. As other medicines are used to augment antipsychotic medications, see the other sections of Chapter 31 on biological therapies.

SECTION REFERENCES

American Psychiatric Association: Practice Guideline for the treatment of patients with schizophrenia. Am J Psychiatry *154*(Suppl):1, 1997.

Baldessarini RJ, Cohen BM, Teicher MH: Significance of neuroleptic dose and plasma level in the pharmacological treatment of psychoses. Arch Gen Psychiatry *45:*79, 1988.

Baldessarini RJ, Frankenburg FR: Clozapine: A novel antipsychotic agent. N Engl J Med *324:*745, 1991.

Bollini P, Pampallona S, Orza MJ: Antipsychotic drugs: Is more worse? A meta-analysis of the published randomized control trials. Psychol Med *24:*307, 1994.

Carpenter WT Jr, Heinrichs DW, Wagman AMI: Deficit and nondeficit forms of schizophrenia: The concept. Am J Psychiatry *145*:578, 1988.

Conley RR, Carpenter WT Jr, Tamminga CA: Time to clozapine response in a standardized trial. Am J Psychiatry *154:*1243, 1997.

Falloon IRH, Liberman RP: Behavioral family interventions in the management of chronic schizophrenia. In *Family Therapy in Schizophrenia*, WR McFarlane, editor. Guilford Press, New York, 1983.

Goldstein MJ, Rodnick EH, Evans JR: Drugs and family therapy in the aftercare of acute schizophrenics. Arch Gen Psychiatry *32:*1169, 1978.

Hogarty GE, Anderson CM, Reiss DJ, Kornblith SJ, Greenwald DP, Jabna CD, Medonia MJ: Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia: I. One year effects of a controlled study on relapse and expressed emotion. Arch Gen Psychiatry *43:*633, 1986.

Hogarty GE, Anderson CM, Reiss DJ, Kornblith SJ, Greenwald DP, Ulrich RF, Carter M: Family psychoeducation, social skills training, and maintenance chemotherapy in the after-care treatment of schizophrenia: II. Two-year effects of a controlled study on relapse and adjustment. Arch Gen Psychiatry *48:*340, 1991.

Janicak PG, Davis JM, Preskorn SH, Ayd FJ: *Principles and Practice of Psychopharmacology*. Williams & Wilkins, Baltimore, 1993.

Kane JM, Honigfeld G, Singer J, Meltzer H, the Clozaril Collaborative Study Group: Clozapine for the treatment-resistant schizophrenic: A double-blind comparison versus chlorpromazine/benztropine. Arch Gen Psychiatry *45:*789, 1988.

Marder SR, Davis JM, Chouinard G. The effects of risperidone on the five dimensions of schizophrenia derived by factor analysis: Combined results of the North American trials. J Clin Psychiatry *58:*538, 1997.

Marder SR, Hubbard JW, Van Putten T, Midha KK: The pharmacokinetics of long-acting injectable neuroleptic drugs: Clinical implications. Psychopharmacology *98:*433, 1989.

Marder SR, Wirshing WC, Mintz J, McKenzie J, Johnson K, Eckman T, Lebell M, Zimmerman KZ, Liberman RP: Behavioral skills training versus group psychotherapy for outpatients with schizophrenia: Two-year outcome. Am J Psychiatry *153:*1585, 1996.

*Robinson DG, Woerner MG, Alvir JJ, Geisler S, Koreen A, Sheitman B, Chakos M, Mayerhoff D, Bilder R, Goldman R, Lieberman JA: Predictors of treatment response from a first episode of schizophrenia or schizoaffective disorder. Am J Psychiatry *156:*544, 1999.

Rosenheck R, Cramer J, Xu W. A comparison of clozapine and haloperidol in hospitalized patients with refractory schizophrenia. N Engl J Med *337:*809, 1997.

Schooler NR, Keith SJ, Severe JB, Matthews SM, Bellack AS, Glick ID, Hargreaves WA, Kane JM, Ninan PT, Frances A, Jacobs M, Lieberman JA, Mance R, Simpson GM, Woerner MG: Relapse and rehospitalization during maintenance treatment of schizophrenia. The effects of dose reduction and family treatment. Arch Gen Psychiatry *54:*453, 1997.

Tollefson GD, Beasley CM Jr, Tamura RN, Tran PV, Potvin JH: Blind, controlled, long-term study of the comparative incidence of treatment-emergent tardive dyskinesia with olanzapine or haloperidol. Am J Psychiatry *154:*1248, 1997.

12.9 SCHIZOPHRENIA: PSYCHOSOCIAL TREATMENT

JUAN BUSTILLO, M.D., SAMUEL J. KEITH, M.D. AND JOHN LAURIELLO, M.D.

Central to the enthusiasm for psychosocial treatments has been the question of whether applying psychosocial treatment to an illness like schizophrenia implies a belief in a psychosocial origin or causality of the disorder. For the era of the 1960s the belief systems of the time posited such causality as schizophrenogenic mothers, double-binds, and familial pathogenesis. Elements observed in families with a schizophrenic member were linked to causality in a post hoc or *propter hoc* manner. While few hold tenaciously to such concepts now, the pain and suffering inflicted on families during that period of thought still resonates through the professional community. It was a time when families were accused of causing schizophrenia, excluded from the treatment process, and forced to pay the financial and psychological price for both. As an almost predictable result of such accusations and attributions, the families of the severely mentally ill joined forces to fight the remnants of such thinking. However, by 1979, when the National Alliance for Mentally Ill was founded, much of this thinking had receded and the theory that such illnesses were biologically determined and should be biologically treated was proposed. Proposals even surfaced to move research on illnesses like schizophrenia from the National Institute of Mental Health to the National Institute of Neurologic Diseases and Stroke because this would be a more appropriate place for a neurobiological illness to be treated. Psychosocial treatments were frequently seen as merely a means to increase medication compliance and were looked at by some family advocates as psychiatry's new means of blaming the family for causing this illness.

Fortunately, this was also an era of rapid growth in research studies on psychosocial treatments, and it would eventually be the data that preserved and advanced the knowledge and utilization of psychosocial treatments. Psychiatry has been fortunate in possessing one of the richest and most extensive data bases in all of medical science. The role of clinical trials in establishing which treatments are effective for which populations has become one that mental health professionals have thoroughly embraced. The field has moved away from blind belief systems and the great-person hypothesis (i.e., if a great person said to do it a certain way, that was the way it was done) toward testable and tested hypotheses.

GENERAL CONSIDERATIONS

The assumption that psychosocial treatments have the ability to effect change carries with it the possibility that the change can go in either a positive or negative direction. It is important to examine the potential for gains and losses for specific modalities. Intensive individual psychotherapy brought about a level of intensity that may have been too high for people suffering from severe psychosis. The intensity of the environment is a critical factor in modulating treatment impact. There is substantial literature from England that has tried to link an overinvolved, intense, hostile, critical environment (i.e., *expressed emotion*) with poor outcome in schizophrenia. Originally this type of environment was associated with families of origin and was correctly perceived by families as disguised finger pointing. Additional studies were able to point out that others, including care providers, were certainly able to produce similar environments iatrogenically. One leading British investigator has noted that in a total push rehabilitation program in which treatment is very intensive with high expectations, he was able to produce dramatic relapse rates in an otherwise well-stabilized population. From the other side of intensity, however, comes the clear demonstration that institutional environments lacking almost completely in stimulation will produce rather striking negative symptoms of schizophrenia. The psychosocial treatment of schizophrenia includes the role of individual, group, and family therapies; the forms of vocational rehabilitation; and the various service delivery models available.

INDIVIDUAL PSYCHOTHERAPY

It is important to note that psychiatric research has always questioned the efficacy of the treatments for mental illness. The 1960s began with intensive individual psychotherapy serving as the main treatment for schizophrenia with little faith placed in other possible treatments. The first major study to question this belief was conducted at Camarillo State Hospital, California, and involved five modalities: antipsychotic medication alone, antipsychotic medication plus individual psychotherapy, individual psychotherapy alone, electroconvulsive therapy (ECT), and milieu alone. The results of this study showed that modalities that contained the antipsychotic medication did best; the effect of ECT was intermediate; and the modalities that did not use antipsychotic medication (individual psychotherapy alone and milieu alone) did the poorest. Traditional psychodynamic advocates criticized the study for using inexperienced therapists, setting it in a state hospital, not continuing the pharmacology into the community, and because the psychotherapy was not intensive enough (twice a week). A study was conducted to answer these criticisms in the 1980s in Boston at McLean Hospital and Boston University using therapists with more than 3 years of experience with intensive psychotherapy with schizophrenia patients; the pharmacology was continued into the community and the intensive therapy was conducted three times a week. The results of this study, which compared the intensive form of psychotherapy with a supportive form given 60 minutes each week, were quite striking: the more severely ill patients dropped out of the intensive psychotherapy, but not out of the supportive version; despite this differential attrition, in only one outcome measure—ego functioning—intensive treatment had an advantage. Overall any advantages at all favored the supportive group. Intensive individual psychotherapy for schizophrenia would continue to be practiced (for a small group of functional patients who were relatively stable and could afford it), but not based on evidence that it was effective for the general population of schizophrenia patients.

This research on psychosocial treatments moved the field away from the idea that greater intensity was better. Moreover, with the rise of biological psychiatry the field could no longer assume pure psychological causes for schizophrenia. This was now an era that looked toward genetic linkage, enlarged lateral ventricles, and hypofrontality of cerebral functioning. However, in the daily struggles with life, persons suffering from schizophrenia think little of these biological variables but rather about their relationships, their ability to be financially self-supporting, and the skills needed to navigate the complex society they live in. Treatments need to be used that first eliminate symptoms and prevent relapse, after which the skills of daily living and relating need to be achieved.

Researchers at the University of Pittsburgh have recently published the results of a 3-year randomized trial of individual interpersonal therapy for schizophrenia compared to family therapy, combined treatment and supportive therapy. Personal therapy was conducted weekly following an incremental approach individualized for the patients' stage of recovery: The initial phase focused on the relationship between stress and symptoms; the intermediate phase emphasized learning to use relaxation and cognitive reframing techniques when stressed; the advanced phase (which generally started 18 months into treatment) focused on seeking social and vocational initiatives in the community and applying what was learned in interpersonal therapy. Social adjustment (a composite measure including work performance, leisure, and interpersonal relationships) clearly favored the interpersonal therapy group. The greatest differential improvement took place in the last two years of treatment, however there were no advantages in relapse for interpersonal therapy. Relapse rates were low, only 29 percent in 3 years.

GROUP THERAPY

The systematic provision of care for persons with schizophrenia in a group setting has followed the general principles of both psychodynamic and behavioral (learning) theories. The behavioral approach has resulted in well-structured interventions (social skills training, social problem-solving, and cognitive remediation) that are often delivered in a group setting. The psychodynamic group therapies, like the individually based approaches, exist along a continuum between intensive and supportive forms. There have been no studies to date comparable to the Boston Psychotherapy Study on individual therapy of schizophrenia in methodological rigor that test the efficacy of intensive versus supportive group therapy.

A few studies have evaluated the effects of adding group therapy (mostly supportive form) to standard antipsychotic medication in a controlled design. For inpatients, the advantage of adding group therapy appears almost negligible. In one of the better-designed studies, investigators evaluated insight-oriented group therapy during the first 3 weeks of hospitalization. The comparison conditions were a task-oriented group and an unstructured control group. Patients with psychosis in the

insight-oriented group did particularly poorly, with an actual worsening in symptoms ratings. Addition of group therapy to medication for outpatients also suggests very limited effects. The one outpatient study with a large sample (N = 100) that actually compared group with individual therapy for the treatment of schizophrenia found advantages in psychopathology and social adjustment ratings for group treatment after 2 years. Nevertheless, this study was more a naturalistic evaluation of effectiveness in a clinic setting than a true efficacy study. There was no attempt to standardize both interventions or to ensure that the patients received optimal and equivalent pharmacotherapy. In summary, there is little empirical evidence and no compelling theoretical reason to justify submitting an acutely psychotic inpatient to a potentially intrusive and overstimulating group intervention. Similarly, the evidence for a specific psychodynamic effect of group treatment in schizophrenia is quite limited. Supportive outpatient group therapy most likely has a modest effect and will probably continue to be used as a practical setting to monitor symptoms and medication compliance and to provide ongoing education for the person with schizophrenia.

FAMILY THERAPY

The role of the family in the treatment of schizophrenia has gone through a number of changes. The present context sees the family as very important clinically. First, it has been clearly demonstrated that even as passive participants, family members are extraordinarily important in providing the clinical treatment team with accurate information on how the patient is doing. It is clear that families identify impending signs of relapse much earlier in the illness than do patients, who tend to lack insight into the onset of the psychotic process. The scientific investigations on family therapy since the 1970s, however, have gone much further than passive participation in the treatment process; families have been invited to become active members of the clinical treatment team.

Families are perhaps the most consistent factor in a patient's life. Over 60 percent of patients discharged from a psychiatric facility return to their families of origin. Early in the course of the illness, this figure is much higher. Further, even patients who do not reside with their parents frequently remain in close contact with them. With today's shrinking resources for psychiatric care, the family represents a resource that could not be easily replaced. Many families are interested in participating in a treatment program, particularly early in the treatment process and particularly if they are welcomed and valued.

There must, however, be a series of caveats in regard to the involvement of the family in the treatment process. First, mental health systems have looked to many types of facilities to house the mentally ill. It should not be a part of a family's role to turn their home into a psychiatric ward. Many families will want to have their family member at home; some will not. Both positions should be given equal respect. Further, the family should not be seen as the long-term caregiver; their energies and good will should be used in the development of transitional strategies or burnout will occur with rapid demoralization and an increasing likelihood of patient relapse. Third, it cannot be emphasized enough how sensitive families are to criticism. The historical context in which they were disparaged has made them justifiably suspicious of their role vis-a-vis psychiatry; this is compounded by their natural feelings of guilt and responsibility. All efforts must be directed initially at reducing this guilt to its lowest possible level. The treatment focus should be on what families can do positively. The negative factors that need to change must do so in the context of an accepting clinical alliance.

It was first noted in a landmark study in Ventura, California, how the addition of a family component over a brief 6-week period provided significant protection against relapse regardless of the medication status of the patient. The effect of this family involvement was present at the end of 6 months despite the brevity of the therapy. This particular family therapy focused on clarity of communication and limited problemsolving.

The next contribution to establishing the importance of the role of the family in the treatment of schizophrenia came from studies conducted in the United Kingdom. *Expressed emotion* describes a critical, overinvolved environment that led to relapse rates exceeding 50 percent in its presence and 20 percent in its absence. These findings have now been replicated across many cultures and countries worldwide and a number of strategies were developed to take advantage of this new information. These strategies differed in where the treatment was delivered (inpatient setting, clinic, or home), whether the patient was acutely ill or stabilized, whether the patient and family were seen as an individual unit or in groups, and what the family treatment was compared to.

The first generation of studies produced remarkably homogeneous results. Studies were done that added family therapy to energetic social treatments, to medication programs, and to other forms of common treatments. Each of these was able to find a major benefit from the addition of family therapy. The one study done on an inpatient unit with the family therapy added to a well-conducted inpatient program also found significant positive results from this addition. There have also been studies that compared family therapy to individual therapy. One that was conducted in Los Angeles compared individual psychotherapy to a form of behavioral family therapy conducted in the home. The results of this study in terms of symptom exacerbation and relapse were strikingly in favor of the family therapy. Further, although it did not reach significance, there was a strong suggestion that this strategy of in-home sessions was responsible for the use of less medication in those treated with family therapy.

These studies were all of a first-generation type, that is, they were done by people who designed the treatment and were heavily involved in its delivery. A second-generation study in which the developers of the treatment taught others how to deliver it but did not treat patients themselves has now been completed by the National Institute of Mental Health (NIMH) in a five-site collaborative study. The importance of such second-generation studies for public health purposes is critical: the methodology must be transferable or it can never be used in any sort of general manner. The NIMH study compared an in-home (weekly for the first 3 months, biweekly for the next 6 months, and monthly to complete the first year) intensive behavioral family therapy in conjunction with a monthly multiple family and patient group meeting to monthly family and patient group meetings alone. The study was able to demonstrate a number of issues: first, that the family home-based treatment could be taught to clinicians; second, that families and patients would participate; and third, that those families who did participate in the initial component of the treatment (a 4-hour psychoeducational workshop) had better initial outcomes. The overall outcome of the study, however, showed no differences between the intensive, in-home treatment plus the monthly group meeting and the monthly group meeting by itself. One possible interpretation of these results is that there is a general positive outcome from involving families in a therapy program that respects their needs and that provides them with useful clinical information, but that the ultimate refining of the procedure does not appear to add significantly to the outcome.

Consistent in the principles of family therapy is the establishment of a positive alliance with the family. Families should be provided psychoeducational material about schizophrenia that includes basic, clinically useful information such as identification of early signs of relapse, the role of medication, and anticipated adverse effects of these medications. Families and patients need to learn simple communication skills (e.g., how to make a positive request, how to make a negative request, how to use “I” statements). In addition, simple problem-solving skills (e.g., identification of a problem, listing of possible solutions, selecting the solution, practicing the solution, implementing it, and reviewing the outcome) should also be taught. Finally, helping the family expand its social network is extremely important. Each of these components can be delivered in a neutral context of caring and the results achieved are incontrovertible.

PSYCHIATRIC REHABILITATION

Because the majority of persons with schizophrenia, even those with a favorable response to antipsychotic medications, will have residual symptoms, cognitive impairments, and limited social skills, psychosocial interventions aimed at the functional rehabilitation of the patient have been designed and systematically studied since the 1980s. Largely following the work on persons with physical disabilities, psychiatric rehabilitation uses principles of learning and social systems theories to effect three types of changes: (1) direct improvement of the disability, (2) development of a new skill that circumvents the original impairment, and (3) targeted manipulations of the environment to make it more supportive and to improve functioning.

The common goal of these interventions is to improve the person's social competence in self-care, work, leisure, relationships, and parenting. Improvement in any of these areas contributes to more autonomous functioning in the community. Two broad types of clinical studies have looked at the effects of psychiatric rehabilitation: studies of social skills training and studies of particular systems of delivery of mental health services. Both these have been evaluated in several randomized controlled studies, but the studies differ in their methodological rigor and potential for applicability in the community. The clear advantage for the studies of social skills training is the ability to more clearly interpret the findings in terms of the specific teaching techniques used. Studies of social skills training are limited by questions of its generalization to more naturalistic settings. Conversely, studies of systems of delivery of services offer a better opportunity of rapid applicability in the community, although their specific reason for working is difficult to identify. In the current era of managed care and financial restraint it is particularly important to determine which elements of a package of care are essential and which are superfluous. Therefore studies that use specific learning technologies, which have the potential to be assembled in different systems of delivery depending on the needs of particular patients and the resources available, offer the greatest potential.

SOCIAL SKILLS TRAINING

Social skills have been described as those specific response capabilities necessary for effective social performance. By definition all patients with schizophrenia are, at least during a considerable period of time, impaired in one or more social skills. To what extent these limitations are secondary to the positive symptoms, negative symptoms, and cognitive impairments of schizophrenia is still debated. The severity and persistence of symptoms and cognitive deficits undoubtedly contribute to the

poor social competence. The ‘lost opportunities' to acquire the skills, especially during the crucial developmental years of adolescence and early adulthood when the illness first manifests, could also importantly determine the impaired social competence. Social skills training should be distinguished from the activities in other rehabilitation programs in which the acquisition of skills occur incidentally. Three important social skills models are the *basic social skills model*, the *social problem-solving model*, and the *cognitive remediation model*.

Basic Social Skills Model Also known as the motor skills model, the basic social skills model was developed in the 1970s and has been studied in controlled-randomized clinical trials with schizophrenia patients since the 1980s. Typically, dysfunctional complex social behaviors are identified and broken down into more elementary tasks, which are learned through repeated performance; then the elements are combined into a more complete functional repertoire. A socially withdrawn patient may be taught the set of skills necessary to start a casual conversation with a stranger. The patient may first be videotaped while role-playing a conversation with a confederate; then the therapist provides feedback with the videotape focusing on elementary behaviors: Is there avoidance of eye contact or too intense gazing? Are the answers too short and the speech barely audible? Does the patient ask follow-up questions that denote interest and promote the flow of the conversation? Each elementary behaviors (e.g., contact, speech volume, length of response, questions) are repeated until learned. The therapist may model the behavior. Next, the patient role-plays the integrated social repertoire and finally practices it in a natural setting.

The literature is consistent in that patients with schizophrenia can be taught various social skills that can be retained for up to 12 months. The data on whether the social skills learned will lead to an improvement in other important clinical measures, such as symptom severity and relapse rate, are mixed. In one of the most rigorous clinical trials of social skills training to date, researchers at the University of Pittsburgh compared relapse rates for outpatients with schizophrenia assigned to an individually administered program of social skills training plus antipsychotic medications with a group that only received pharmacotherapy. At 1 year of follow-up, there was a 46 percent rate of relapse in the control group compared to 30 percent in the social skills training group. The difference remained statistically significant in favor of social skills training for up to 21 months but not at the end of the study (24 months). The limited durability of a therapeutic effect in schizophrenia is not unique to social skills training and it should be viewed in the context of the total period of time during which the intervention is applied. For antipsychotic medications too, relapse prevention is closely related to length of exposure to treatment. For psychosocial interventions in general and social skills training in particular, booster sessions or continuous treatment may be required in order to maintain a favorable effect on relapse rates.

Even if schizophrenia patients learn social skills, the extent to which these skills generalize in the patients' natural environment and result in improved social competence is a crucial measure of outcome. Generalization has not been clearly demonstrated for the motor skills model. In part to deal with this limitation, problem-solving techniques have been integrated with the more traditional social skills training model.

Social Problem-Solving Model Developed by researchers at the University of California at Los Angeles, this model uses social learning principles and assumes that impairments in information processing underlie the limited social competence present in the patient with schizophrenia. Like traditional social skills training, complex problem behaviors are identified. Patients can then be assigned to each of five target modules for improvement: (1) medication management, (2) symptom management, (3) recreation, (4) basic conversation, and (5) self-care.

For each of these modules an emphasis is placed on learning, receiving, processing, and sending skills, in the hope that these will increase flexibility, durability, and generalization of the module learned. *Receiving skills* are those skills necessary to appropriately identify social cues. *Processing skills* deal with the context-appropriate interpretation of input information and the generation of the various potential responses. *Sending skills* correspond to the behaviors targeted by the traditional motor skills model. A recent study reported the results of a 2-year study comparing problem-solving group therapy with supportive group therapy for relapse prevention and social adjustment in outpatients with schizophrenia. Each group received the same intensity, frequency, and overall length of psychosocial treatment. There was a small but statistically significant advantage for the experimental intervention in two out of six measures of social adjustment after 2 years. There were no differences in relapse rate. Therefore the benefits of the problem-solving therapy were modest at best. It has been argued that for many patients with schizophrenia deliberately evaluating a social situation in terms of information-processing skills may be particularly difficult and could actually limit the acquisition of social skills. The traditional strategy of simple practice so that overlearned responses can be emitted automatically may be more advantageous. The social problem-solving model will require further evaluation but it offers advantages resulting from its potential for applicability: the modules are clearly structured, manuals are available for patients and therapists, and each module stands independently and can be implemented according to the individual patient's needs. A recent report on this model found it to be more effective in a measure of independent living skills than in a control condition. This effect was maintained even 18 months after completing the intervention.

Cognitive Remediation That patients with schizophrenia have a multiplicity of cognitive impairments is well established. These are most likely not just epiphenomena of symptom severity or medication adverse effects because they are present in various subclinical populations (children at high risk for schizophrenia, nonpsychiatric relatives of schizophrenia patients, and persons with schizophrenia-spectrum [Cluster A] personality disorders). The cognitive impairments are usually generalized, but specific functions like attention, memory, and planning may be more impaired than others. The use of cognitive remediation of schizophrenia is relatively new, and the impaired functions that underlie the generalized deficit, as well as the failure in social competence, have yet to be fully established. Researchers have reasoned that the limitations on the durability and generalization of social skills training may be overcome by improving the impairments in basic cognitive functions before teaching social skills. Studies on cognitive remediation of schizophrenia are few in number. There is some evidence that patients can improve their performance through practice in measures of vigilance and planning. Nevertheless, transfer of learning to another test has not been documented, even within the same cognitive domain (vigilance or planning), let alone showing any evidence of generalization to particular social skills. Probably the most comprehensive hierarchical program, which attempts to teach social skills by building upon learning on elementary cognitive functions, is the integrated psychological treatment for schizophrenia devised by researchers in Switzerland. Patients are seen in a group setting three times a week over 3 months and begin training in card sorting and concept formation using computer games (cognitive differentiation subprogram). This is followed by exercises fostering social problem-solving (social perception and verbal communication subprograms). The final subprograms (social skills and interpersonal problem-solving) resemble the more traditional motor skills model. Although the few controlled studies of integrated psychological treatment in schizophrenia show some modest gains in elementary cognitive functions and some learning in social skills, this preliminary literature is not supportive of the view that improved basic cognitive performance will predict acquisition of social skills. The generalization of integrated psychological treatment to measures of symptomatology and social competence in the community remains to be assessed.

It has been argued that until more is known about the specific cognitive dysfunctions in schizophrenia and the way these might determine higher-order problems in social competence, cognitive remediation may be more successful by focusing on cognitive schemas more closely linked to relevant patterns of psychopathology. Schizophrenia is most likely a heterogeneous disorder. Studies of cognitive psychopathology suggest three clinical syndromes within the disorder: psychomotor poverty, disorganization, and reality-distortion, each with an underlying faulty cognitive style. The psychomotor poverty syndrome may be characterized by a failure to initiate a cognitive set, leading to mental poverty and apathy. The disorganization syndrome is a failure to maintain a cognitive set, leading to distraction and disorganized thinking. Finally, the reality-distortion syndrome is an inability to change cognitive sets, resulting in fixed beliefs and misinterpretation of internal stimuli. It has been suggested that initial training in these cognitive styles for patients selected on the basis of their symptomatology may be more fruitful than remediation of nonspecific elementary cognition and may lead to greater flexibility than traditional social skills training.

These techniques focus on improving faulty cognitive processes. Recently there has been an interest in cognitive-behavioral therapy of schizophrenia, which focuses on the content of the symptoms. Researchers in United Kingdom have applied coping strategy enhancement for residual psychotic symptoms. The goal is to teach coping mechanisms that will distract patients and help them ignore their symptoms, hopefully decreasing their distress. Initial reports suggest a beneficial effect on delusions with loss of effect at 6 months and limited generalizability to other symptoms or social functioning. A different approach, and closer to the original cognitive-behavioral therapy for depression, this modality consists of a gentle but systematic verbal challenge of the delusional belief. This is followed by a behavioral experiment that does not support the delusional belief. The outcome of the experiment (reality testing) reinforces the verbal challenge. This technique may help decrease the severity of delusions. Persistent hallucinations seem particularly resistant to cognitive-behavioral therapy, perhaps because of the recurrence of internally generated stimuli that the patient misattributes to an external source. A recent approach attempts to help patients reattribute the voices to themselves through focusing on the meaning and other characteristics of the hallucinations. These content approaches may offer hope for some patients with persistent delusions or hallucinations. Nevertheless, these studies should be viewed as preliminary and awaiting replication in well-controlled studies with larger samples.

Social skills training has been shown to improve social competence in the laboratory and the clinic. The extent to which this learning translates into better role functioning in the community has yet to be determined. Therapeutic effects have been demonstrated for social skills training in clinical measures like relapse rate, but because of the limited durability, booster or follow-up sessions may be required.

SERVICES DELIVERY SYSTEMS

The period of deinstitutionalization of care for the mentally ill in the United States began in the 1950s and peaked in the 1970s. This massive relocation of patients from state hospitals to the community implied but frequently did not deliver a parallel shifting in delivery of services on a long-term basis. These services not only

included the continued provision of psychiatric treatment to prevent rehospitalization, but the availability of stable and secure housing and systematic efforts to rehabilitate and reintegrate patients into the workplace.

Vocational Rehabilitation Work has been viewed as an important element of the treatment of schizophrenia dating back to the era of the large asylums, which often had farms or other protected settings for patients to work in. With the modern goal of maintaining patients to be as functional and autonomous as possible in the community, various programs to help patients find and maintain jobs have been implemented. Despite these efforts, rates of competitive employment for chronically mentally ill persons have been estimated to be less than 20 percent and are more likely to be lower for patients with schizophrenia. With the current average hospital stay being shorter than 2 weeks, there are practically no hospital-based workshop programs and most vocational rehabilitation takes place in the community. These programs are very heterogeneous but can be subsumed under four models: (1) job clubs, (2) Boston University model, (3) transitional employment, (4) and supported employment.

Job clubs were originally used to help unemployed nonpsychiatric persons to find jobs. This approach focuses on teaching individuals how to look for a job and how to apply for it; there is no actual teaching of job skills. Because past history of competitive employment is the best predictor of future employment, this approach is probably not useful for the majority of patients with schizophrenia who never learned particular job skills.

The Boston University Model is an intensive and individualized program. A patient and a counselor first identify a specific paid job that the patient would like to obtain. Then they both proceed through a systematic review of the skills needed for the job, the skills the patient has, and the resources available to improve these skills. A plan is defined that usually takes, 1 or 2 years to implement. This model may be most helpful for younger, less disabled patients who have a prospect for long-term career choices.

The transitional employment program (the train-and-place model) is exemplified by Fountain House, one of the oldest rehabilitation programs in the United States. Initially, the patient's work aptitudes and preferences are assessed. Next, patients are taught general work habits (e.g., appropriate appearance, keeping hours, asking for clarification, following through). Then patients start working in a sheltered program that is usually owned by the rehabilitation agency, learning more specific skills like typing, filing, janitorial work, or maintenance routines. Once a certain level of competency has been consistently maintained for a period of time, they are placed in a regular job. Many of the most severely disabled patients with schizophrenia are incapable of moving beyond the sheltered workshop level. The transitional process of training, sheltered work, and placement appears to reduce stress related to the demands of employment.

Supported employment is the most recent approach. It attempts to improve the opportunities for competitive employment by referring the patient to a regular job that is intended to be permanent. Unlike in the transitional model, the job identified for the patient is not owned by the rehabilitation agency. The patient's skills and a job that matches these skills are identified and the patient begins to work. The goal is for the patient to develop the necessary skills while working, getting on-site support and training from a work coach. This on-site job training was developed for developmentally disabled individuals who require initial high-intensity services that are later phased out. For persons with schizophrenia some form of continuous long-term on-the-job training may be needed.

Outcome research for these vocational rehabilitation programs is particularly limited by the lack of standardization of the experimental intervention and the difficulty in implementing adequate control treatments. Another limitation is that in these studies patients are not usually differentiated diagnostically, but lumped as chronically mentally ill or psychiatrically disabled. Also, the adequacy of pharmacological treatment is generally assumed. Various vocational outcome measures are used: job performance and satisfaction, earnings, percentage of time employed, and maintenance of a full-time or part-time job or competitive employment. Full-time competitive employment is the most desirable outcome measure. A meta-analysis on 19 controlled studies of various vocational rehabilitation interventions found a 51 percent rate of paid employment in the experimental group compared to 29 percent for the control group, a significant difference. There were additional advantages for vocational rehabilitation in terms of job starts and longer duration of employment. In none of the studies did the outcomes favor the control interventions. Nevertheless there were no differences in terms of competitive employment in any of the studies after the patients were discharged from the programs. The success of these interventions has been mostly to help patients to adapt to a specific transitional or sheltered employment, but they have failed to prepare them for future regular employment.

Results from two recent studies involving supported employment provide a more hopeful outlook. In a quasiexperimental study in New Hampshire researchers compared vocational outcomes in two groups: patients involved in a traditional partial hospitalization program and those assigned to an integrated program of supported employment and intensive case management. During the follow-up year the experimental group significantly increased its rate of competitive employment by 14 percent while there was no improvement in the control group. In another study, researchers in Indiana compared two forms of supported employment programs: one with gradual entry and the other with accelerated involvement in the job. The accelerated-entry patients showed significantly superior outcomes in competitive employment at 1-year follow-up (56 versus 29 percent).

Several vocational rehabilitation studies have also measured other clinically relevant outcomes, like rate of hospitalization, medication compliance, and substance abuse; most of these studies find advantages for the experimental treatment. Nevertheless, a causal relationship cannot be inferred from these studies because they all fail to describe other important clinical interventions and patient characteristics.

In summary, vocational rehabilitation programs by definition enhance job-related activities in chronically mentally ill patients (including persons with schizophrenia), but they do not have significant effects on competitive employment once the patient leaves the program. However, there is reason for optimism with some forms of supported employment.

CASE MANAGEMENT

Patients with schizophrenia are often ill prepared to find and maintain proper use of the multiple services they need in order to function in the community. The providers of these services include psychiatrists, nurses, pharmacists, general practitioners, dentists, psychologists, social workers, vocational and recreational therapists, home supervisors, patient advocates, lawyers, and benefits officers. Case managers function at two ends of a continuum: the broker approach and the intensive case management approach.

At a minimum, case managers function as brokers of services and are contacted by other professionals. Case managers identify a new need for the patient and referring the patient to the provider able to deliver these services. This approach may be sufficient for patients with physical disabilities and for some patients with mental illness of moderate severity. Unfortunately, many patients with schizophrenia lack the level of cognitive and social competence to consistently follow through and get their needs met. The polar opposite of the case-manager-as-broker is exemplified by the Assertive Community Treatment (ACT) program. Originally developed by researchers in Madison, Wisconsin, in the 1970s, this is the most carefully defined, well-documented, and successful program for the delivery of services for persons with chronic mental illness. Patients are assigned to one multidisciplinary team (case manager, psychiatrist, nurse, general physician, etc.). The team has a fixed case load of patients and delivers all services when and where needed by the patient, 24 hours a day, 7 days a week. This is a mobile and intensive intervention that provides treatment, rehabilitation, and support activities. These include home delivery of medications, monitoring of mental and physical health, in vivo social skills training, and frequent contact with family members. There is a high staff-to-patient ratio (1:12, compared to 1:30 in traditional case management models). The original study from Wisconsin followed chronically mentally ill patients assigned to the ACT program and compared them with a group discharged from the hospital to standard community care. After 14 months the ACT group showed significant advantages in rates of hospitalization, sheltered employment, independent living, family burden, with essentially no difference in costs. Unfortunately, the advantages were lost after the patients were discharged from the experimental program; there is no present data on the minimum intensity needed for the program to maintain gains or which special population of patients may require continuous services. There have been several replications of the effectiveness of the ACT program in reducing total number of days in the hospital and increasing patient and family satisfaction. Nevertheless, most studies have failed to document improvement in employment, social functioning, and other measures of quality of life. Because of the comprehensive services provided by the ACT program and the lack of adequate controls, it is unknown whether the reduction in hospitalization is due to improved medication compliance, continuity of caregivers, 24-hour coverage, site of service, intensity of services, therapeutic alliance, or a combination of any of these elements. Neither is it known whether ACT is more cost effective than another high-intensity system that does not use multidisciplinary teams with constant availability. Preliminary results are available from a randomized study from Connecticut that compared the ACT program with a program of high-quality brokered case management (two-thirds of the patients were diagnosed with schizophrenia or schizoaffective disorder). In the control intervention, patients had access to mobile crisis services, supported housing, respite services, vocational programs, clubhouses, and outpatient psychiatric clinics. Case managers were responsible for brokering instead of directly providing any of these services. Results show that after 18 months the control group spent twice as much time in the hospital (14.3 percent versus 7.6 percent) as the ACT group. There was also a significant advantage for the ACT group in terms of percentage of time spent in an unstable living situation (11.6 versus 16.9 percent); data on cost effectiveness is not currently available.

In summary, ACT programs can effectively decrease the risk of rehospitalization for chronically mentally ill patients in general and persons with schizophrenia in particular but the beneficial effects do not extend to measures of social functioning. The issue of cost effectiveness will not be settled until a systematic effort is undertaken to identify the essential elements of the ACT program.

RESIDENTIAL TREATMENT AND HOUSING PROGRAMS

Patients with schizophrenia often need supportive housing. They can sometimes reside with their families, but eventually most patients will need other forms of permanent housing. Since deinstitutionalization, housing options for the chronically mentally ill have used a linear continuum paradigm. According to this view, patients are placed in different housing programs following hospitalization, beginning with the most controlled, supervised, and hospital-like setting, and moving down in a progressive fashion to the least restrictive, regular–household-like settings. These housing programs have received numerous names, but they basically vary depending on the level of staff supervision, the length of stay, number of residents, and whether other services in addition to housing are provided (e.g., vocational rehabilitation or recreational activities). For patients who are at imminent risk of relapse or in the initial stage of decompensation, *crisis care centers* may be particularly helpful. These have a 24-hour nursing staff and are supervised by a psychiatrist. Patients are in a more controlled environment where symptom progression and risks can be assessed and medications delivered; no vocational or other rehabilitative services are provided. Length of stay is usually limited to 30 days and disposition is either to a hospital or to a nonclinical residential type of program. *Transitional half-way houses* provide a supportive environment for patients recently discharged from the hospital who require some limited level of clinical supervision, such as dispensation of medication. By definition, length of stay is limited to a few months and patients are then placed in some form of long-term housing. Persons with schizophrenia who have pervasive disruptive symptoms that interfere with daily living skills or are associated with potentially dangerous behavior may require *long-term group residencies*. These often have 24-hour-a-day on-site supervision, usually with minimal clinical services. Their length of stay is between several months and years and no rehabilitative services are provided. *Cooperative apartments* typically house a small number of patients who do not require on-site supervision. *Nursing homes* are a long-term housing and clinical care option for severely disabled individuals who require daily nursing care. *Boarding homes* are usually remodeled apartment buildings that only provide room and board for persons with some form of disability; there are no clinical or rehabilitative services but there is often some minimal supervision. *Foster care homes* are owned by private citizens who agree to house psychiatricaly disabled persons for a fee; the match between patient and provider is particularly important.

Two other well-known centralized programs which provide housing in addition to various rehabilitative services are the Fairweather Lodge and Fountain House. The Fairweather Lodge was conceived in California in the 1960s with the goal of reducing the revolving-door cycle of chronically mentally ill patients who improved while hospitalized but were unable to remain in the community for long periods of time. As originally conceived, a small group of hospitalized patients is taught a variety of skills for daily living. They are encouraged to work together and are eventually discharged to live in a community lodge that serves the functions of housing, socialization, and work; the lodge frequently operates its own business. The basic philosophy is that patients will help rehabilitate themselves while functioning as a family. Although initial controlled studies found advantages in rehospitalization and employment, these have not been replicated—partly due to the difficulties in implementing the original model. With the current limitations on length of hospital stay, this model may be an option only for a minority of patients.

Fountain House was started in New York in the late 1940s not just as a housing option but as a full rehabilitative program in which patient involvement and self-help are emphasized, with support from staff as needed. One component of the program is the clubhouse, which provides opportunities for socialization while functioning as a transitional workshop; the patients learn clerical, maintenance, and food preparation skills while working. There are no controlled randomized studies on the effectiveness of this approach but some uncontrolled studies have found beneficial effects on employment. A new paradigm on housing for the chronically mentally ill, *supported housing*, aims toward further patient reintegration in the community while minimizing stigmatization. This approach is analogous to the concept of supported employment. Patients are encouraged to live in small groups in regular homes in the community. The basic concept is of housing as a normal living environment, not as a setting for services administered by an agency. Patients are not placed but are encouraged to choose a particular home that suits their needs. Long-term support by staff is available but not on site. It is not yet clear for what group of patients and at which point in the course of the illness supported housing will be most beneficial.

SUGGESTED CROSS-REFERENCES

Psychotherapies are covered in Chapter 30. Various aspects of schizophrenia are covered in the other sections of Chapter 12.

SECTION REFERENCES

*Bellack AS, Mueser KT: Psychosocial treatment of schizophrenia. Schizophr Bull *19:*317, 1993.

Bond GR, Drake RE, Mueser KT, Becker DR: An update on supported employment for people with severe mental illness. Psychiatr Serv *48:*335, 1997.

Brown GW, Birlley JLT, Wing JK: Influence of family life on the course of schizophrenic disorders: A replication. Br J Psychiatry *112:*241, 1972.

*Bustillo JR, Lauriello J, Keith SJ: Schizophrenia: Improving outcome. Harv Rev Psychiatry *6:*229, 1999.

Essock SM, Kontos N: Implementing assertive community treatment teams. Psychiatr Serv *46:*679, 1995.

*Falloon IRH, Boyd JL, McGill CW: Family management in the prevention of exacerbations of schizophrenia. N Engl J Med *306:*1437, 1982.

Glick ID, Spencer JH, Clarkin JF, Haas GL, Lewis AB, Peyser J, DeMane N, Good-Ellis M, Harris E, Lestelle V: A randomized clinical trial of inpatient family intervention. IV. Follow-up results for subjects with schizophrenia. Schizophr Res *3:*187, 1990.

Goldstein MJ, Rodnick EH, Evans JR, May PRA, Steinberg MR: Drug and family therapy in the aftercare of acute schizophrenics. Arch Gen Psychiatry *35:*1169, 1978.

*Gunderson JG, Frank AF, Vannicelli JL: Effects of psychotherapy in schizophrenia. II. Comparative outcome of two forms of treatment. Schizophr Bull *10:*564, 1984.

*Hogarty GE, Anderson CM, Reiss D, Kornblith SJ, Geenwald DP, Ulrich RF, Carter M, and the Environmental-Personal Indicators in the Course of Schizophrenia (EPICS) Research Group: Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia, II. Two-year effects of a controlled trial on relapse and adjustment. Arch Gen Psychiatry *48:*340, 1991.

Hogarty GE, Kornblith SJ, Greenwald D, DiBarry AL, Cooley S, Ulrich RF, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, I. Description of study and effects on relapse rates. Am J Psychiatry *154:*1504, 1997.

Hogarty GE, Greenwald D, Ulrich RF, Kornblith SJ, DiBarry AL, Cooley S, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, II. Effects on adjustment of patients. Am J Psychiatry *154:*1514, 1997.

Lehman AF: Vocational rehabilitation in schizophrenia. Schizophr Bull *21:*645, 1995.

Lehman AF, Dixon L, Kerman E, DeForge BR, Postrado LT: A randomized trial of assertive community treatment for homeless persons with severe mental illness. Arch Gen Psychiatry *54:*1038, 1997.

Lieberman RP, Wallace CJ, Blackwell G, Kopelowicz A, Vaccaro JV, Mintz J: Skills training versus psychosocial occupational therapy for persons with persistent schizophrenia. Am J Psychiatry *155:*1087, 1998.

Marder SR, Wirshing WC, Mintz J, McKenzie J, Johnston K, Eckman T, Lebell M, Zimmerman K, Lieberman RP: Two-year outcome of social skills training and group therapy for outpatients with schizophrenia. Am J Psychiatry *153:*1585, 1996.

May PRA: *Treatment of Schizophrenia: A Cooperative Study of Five Treatment Methods*. Science House, New York, 1968.

*Schooler NR, Keith SJ, Severe JB, Matthews SM, Bellack A, Hargreaves WA, Kane JM, Ninan PT, Frances A, Jacobs M, Mance R, Simpson GM, Woerner MG: Relapse and rehospitalization during maintenance treatment in schizophrenia. The effects of dose reduction and family treatment. Arch Gen Psychiatry *54:*453, 1997.

Stanton AH, Gunderson JG, Knapp PH, Frank AF, Vannicelli ML, Schnitzer R, Rosenthal R: Effects of psychotherapy in schizophrenia, I. Designed implementation of a controlled study. Schizophr Bull *10:*520, 1984.

Wing JK, editor: *Schizophrenia: Towards a New Synthesis.* Academic Press, London, 1978.

Textbook of Psychiatry

12.10 SCHIZOPHRENIA: INDIVIDUAL PSYCHOTHERAPY

WAYNE S. FENTON, M.D. AND THOMAS H. MCGLASHAN, M.D.

No single treatment can ameliorate the myriad symptoms and disabilities associated with schizophrenia. As articulated in the American Psychiatric Association's *Practice Guidelines for the Treatment of Patients with Schizophrenia*, therapeutic efforts must be comprehensive, multimodal, and empirically titrated to the individual patient's response and progress. While there is at present no cure for schizophrenia, the skillful application of pharmacological, psychotherapeutic, rehabilitative, and community support interventions can limit illness morbidity and mortality, improve patient outcome, and enhance quality of life.

As with individuals with other long-term medical disorders, many patients with schizophrenia will need comprehensive and continuous care over prolonged periods. To the extent that both biological and psychosocial factors are crucial determinants of the course and outcome of schizophrenia, a psychiatrist may be in the best position to coordinate and integrate the various treatments required and provide continuity of care over time.

Each component of a comprehensive treatment plan for a patient with schizophrenia targets specific aspects of the disorder and its common sequelae. Pharmacological interventions target positive, negative, and disorganized symptom domains; mood symptoms; and cognition. Rehabilitation efforts target deficits in self-care, social or vocational skills, and provide structure in the form of someplace to be and something productive to do. Access to appropriate entitlements, treatment services, housing, and social support are the goals of community support programs. Individual psychotherapy addresses the human aspects of adaptation to a serious psychiatric disorder and targets problems such as denial, demoralization, treatment compliance, personal relationships, and self-esteem. Its focus is on understanding the patient's beliefs, attitudes, aspirations, and experiences. The coordination, timing, and titration of all specific treatment elements is informed by this understanding and by an ongoing assessment of individual patient needs that can often best be achieved within a long-term physician-patient relationship.

Individual psychotherapy is defined most broadly as a professional relationship in which the technical expertise of the physician is directed toward promoting the patient's recovery or toward relieving suffering. At a minimum this physician-patient relationship provides the context in which symptoms and disabilities are assessed, consent and collaboration for treatment obtained, and the effects of interventions are evaluated. More ambitious goals, appropriate for selected patients in settings where time and resources allow, can include the exploration of maladaptive patterns of living through careful scrutiny of relationships with others and the therapeutic relationship itself. Most psychotherapy, as actually practiced, falls somewhere in between, requiring from the therapist a broad base of medical and psychological skills. A psychiatrist providing psychotherapy for schizophrenia patients should probably be prepared to give an intramuscular injection one day, interpret transference the next, and give a patient a ride to work on the third.

Some form of individual psychotherapy in combination with pharmacological treatment is the most common care offered to patients with schizophrenia. Psychodynamic and biological conceptions of schizophrenia have yielded the two distinct therapeutic traditions of investigative and supportive psychotherapy. Current practice represents an amalgamation of these approaches that is best described by *flexible psychotherapy*. This approach draws upon perspectives and techniques derived from each of these traditions to accommodate the heterogeneity of schizophrenia and the individuals who suffer from it.

HISTORY

In the early decades of the twentieth century schizophrenia was viewed as an irreversible and untreatable process. “Organic” psychiatry as represented by Emil Kraeplin saw schizophrenic personality disintegration as an inevitable product of neurological deterioration. Sigmund Freud, representing the mainstream of psychoanalysis, considered dementia precox to be a “narcissistic neurosis” where “libido” was directed inward, away from others. As a result, transference and hence analytic treatment was considered impossible. In this context, the diagnosis of dementia precox most often led to therapeutic nihilism and the recommendation of life-long institutional care.

Despite the misgivings of classical adherents to Freud, individual psychotherapy for schizophrenia in the United States originated as a modification of psychoanalysis. Early psychoanalysts such as A.A. Brill advocated an active effort to promote “rapport” and arouse patients' interest in their own malady. He described providing direct advice about work and relationships, visiting a patient at home, providing didactic reading material, and at times frankly labeling false beliefs as delusions. In time, he observed that confidence in and a “passive attachment” to the physician can develop so that the latter might become a bridge between the patient and reality.

Between 1922 and 1930 Harry Stack Sullivan ran a small treatment unit for male schizophrenia patients at Sheppard and Enoch Pratt Hospital in Towson, Maryland. He was influenced by the early psychobiological perspectives of Adolph Meyer and William Allison White, who emphasized that personality was influenced by life events not only in childhood, but over the entire course of development. Based on his intuition that like cures like, Sullivan staffed his unit with sensitive, shy, and introverted male attendants who possessed a natural proclivity and ease of rapport with withdrawn patients. Stressing that patients' difficulties were similar to those of so-called normals, Sullivan promoted the development of closeness or benevolent intimacy in this milieu. He observed that providing an experience of reciprocal trust—which he hypothesized many patients had lacked during important periods of development—could be beneficial by allowing a “validation of all components of personal worth.”

Careful observation of the difficulties his patients had in maintaining relationships led Sullivan to formulate the paradigm of interpersonal psychiatry. He de-emphasized the prevailing psychoanalytic view that personality was formed and behavior motivated by drives pressing for expression from within. Instead, he recast psychopathology as difficulties in living arising largely from personal and social relations, and as personality warps thought to be the lasting residue of earlier unsatisfactory interpersonal experiences. Over a period of years Sullivan elaborated these ideas in a series of seminars at Chestnut Lodge Hospital in Rockville, Maryland, where, under the leadership of Dexter Bullard, a group of psychoanalysts and social scientists interested in the intensive study of schizophrenia assembled during the 1940s. There the influence of interpersonal patterns among patients and between patients and staff were observed to have a powerful impact on patients' psychopathology. Covert tension and disagreement among staff, for example, often appeared to be associated with worsening of patients' psychotic symptoms; likewise, improvement followed when those tensions were resolved. Such observations drew attention to the influence of psychosocial factors on schizophrenia and raised the notion that the disorder might be caused and potentially cured by psychosocial means.

Drawing upon her European psychoanalytic background, her mentorship with Sullivan, and her clinical work with psychotic patients at Chestnut Lodge, Frieda Fromm-Reichmann integrated the available knowledge concerning the intensive psychotherapy of schizophrenia into a relatively comprehensive body of theory and technique. Her seminal work, *Principles of Intensive Psychotherapy*, articulated a modified form of psychoanalysis that was applicable to patients with severe mental illness, including schizophrenia. The ideas embodied in her writings and clinical work also represent the first elaboration of what became known as *intensive psychodynamic psychotherapy*.

The predecessors of ego and self psychology, interpersonal psychiatry and psychodynamic psychotherapy became dominant paradigms in American psychiatry in the 1940s and 1950s and beyond. Their hopeful and humanistic perspectives were adopted by many influential psychiatric treatment centers. Intensive individual psychotherapy came to be viewed as the treatment of choice and at times the only effective treatment for schizophrenia. Its practitioners kept the field's interest focused on severely ill patients during the decades prior to the widespread availability of effective pharmacological treatments. Exposure to a second generation of charismatic teachers and clinicians (e.g., Otto Will, Steve Fleck, Elvin Semrad, and Harold Searles) kindled interest in treating and studying schizophrenia among countless psychiatric residents and trainees throughout the country.

INVESTIGATIVE PSYCHOTHERAPY

Perspectives derived from the tradition of investigative psychotherapy are of value in understanding the psychological aspects of the disorder and in managing commonly encountered difficulties in forming and maintaining a therapeutic relationship. A review of the voluminous clinical literature in this area suggests that differences in language and terminology notwithstanding, a consistent orientation and approach has been articulated. This can be outlined in terms of the nature of investigative psychotherapy, the nature of schizophrenia from the psychotherapist's point of view, elements of the psychotherapeutic situation, general technical interventions, and general technical attitudes.

INVESTIGATIVE PSYCHOTHERAPY

Sullivan defined the psychiatric interview as a two-person transaction, more or less voluntarily initiated on a progressively unfolding expert-client basis. Its purpose is to elucidate the patient's characteristic patterns of living, the revealing of which is assumed to be useful. The psychotherapeutic encounter is an actual interpersonal experience in which the doctor and patient are both participant observers. According to Fromm-Reichmann, the goal of intensive psychotherapy is alleviation of the patient's emotional difficulties and elimination of symptoms. This is accomplished by undertaking a thorough scrutiny of the patient's life history (especially the history of interpersonal relationships), reviewing in close detail the realities of the patient's current relationships and life situation, and understanding the genetic (historical) roots and current ramifications of maladaptive interpersonal patterns as reflected in the doctor-patient relationship and in daily life. Important emotional experiences related to the patient's difficulties are assumed to have been forgotten and their recovery during the therapeutic process is expected. This process is expected to result in the modification of maladaptive interpersonal patterns and personality growth.

Schizophrenia The literature on intensive psychotherapy emphasizes the influence of the environment and learning in the etiology of schizophrenia. Characteristic difficulties in interpersonal relations among schizophrenia patients are said to include a basic mistrust of and expectation of harm from others, marked ambivalence in relationships with endless oscillations between longing for merger based on intolerance of loneliness and withdrawal and isolation based on terror of closeness, weak or absent ego boundaries with resulting difficulty differentiating one's own thoughts and impulses from those of others, the absence of a sense of self often compensated for by an effort to ascertain the expectations of others and mold oneself accordingly (false self) or alternatively to organize in fixed opposition to the wishes of others, a pervasive posture of passivity (things happen to one and others are the cause of all difficulties), fear that strong emotional arousal of any sort (anger, pleasure, wants, desires) will escalate uncontrollably and lead to panic or catastrophe (with compensating constriction and repression of drives or affects and resultant inability to express affects or desires to others), fragmented or idiosyncratic thinking, frequent misinterpretation of the motivation of others, and an antipathy towards reality with intolerance of frustration and withdrawal into fantasy.

Extrapolating from these aspects of schizophrenic psychopathology, most early proponents of intensive psychotherapy postulated real or fantasized negative first experiences between the infant patient and primary caregivers. This was thought to result in a central unconscious conflict described as “that of a small child dependent on a person by whom he feels persecuted and who is, in his opinion, unstable and uncertain,” as described by Lewis Hill. This position represents the patient's conviction concerning the nature of human relationships and it dominates all thoughts, feelings, and behaviors. Regrettably, psychodynamic theorizing regarding the cause of schizophrenia often devolved into blaming the parents for causing their child's illness. Despite the lack of any credible scientific evidence to support these theories, for a period they were widely and uncritically accepted. Now rejected as invalid, psychodynamic etiological explanations of schizophrenia have left a legacy of mistrust between families, patients, and psychiatry. Clinical descriptions deriving from that era, however, continue to be of value in enriching clinicians' understanding of schizophrenia patients' subjective experience, and may allow clinicians to anticipate the impact of life events and important relationships on the patient's course of illness.

A developmental perspective and hierarchical model of the mind is implicit in the clinical theory underlying intensive psychotherapy. Although a variety of specific developmental schemes has been offered, most suggest some version of the following phases in emotional development: (1) autistic phase during the first weeks of life in which the presence of others is not recognized, satisfaction of biological needs is hallucinated, and only undifferentiated states of anxiety or activation and satiation are present; (2) symbiotic phase in which a boundaryless state of bliss is present with an empathic caregiver who can anticipate and fulfill all needs—successful completion of this phase is thought to form the substrate of basic trust; (3) separation-individuation phase beginning with the ability to ambulate, when the image of the good-enough caregiver becomes progressively internalized, allowing a feeling of security during physical separations. Early in this phase experiences of frustration with the caregiver are emotionally separate from experiences of satisfaction and the respective good object and bad object seem like different people. Aggression and anger deriving from frustrations are projected onto and attributed to the bad object. The good-enough caregiver accepts these projections without responding with excessive retaliatory anger or anxiety and thus contains and soothes the frustrated child. These experiences of containment or holding form the basis for children's later capacity to soothe themselves, modulate affect, and become comfortable with emotional arousal. By the end of this phase cognitive and emotional development allow for recognition that the frustrating and satisfying caregiver are indeed the same person (libidinal object constancy), a recognition that forms the basis of seeing others as separate and complex individuals; (4) Oedipal phase, which is the major focus of psychoanalysis with patients who have neuroses. This involves the mastering of triadic relationships, competitive urges, and identifications with the same-gender parent.

Although no writers about intensive psychotherapy consider the schizophrenia patient's mental functioning as equivalent to that of an infant, more primitive adaptive levels of functioning derived from early phases of development are considered to be ever present, hierarchically underlying more sophisticated adaptive levels acquired later in development. During states of psychotic regression, developmentally primitive states of mind are thought of as gaining ascendancy while higher capacities are temporarily lost. Nonetheless, the retention of some nonpsychotic functioning is assumed, however sick the patient is.

Psychotherapeutic Situation: Participants Although the characteristic difficulties of schizophrenia patients have been elaborated at length in the literature, attributes of the optimal psychotherapist are less well defined. Among those cited as important are an interest in and capacity to tolerate intense affect, dependency, confusion, and ambiguous communication. Basic respect for the patient is a prerequisite, especially respect that stems from a conviction that the patient's problems are not too different from one's own. Aloofness, rigidity, and critical pomposity are especially discouraged. Psychotherapists should be flexible, creative, and willing to admit when they are wrong. The match between patient and therapist is thought to be central, but defies easy categorization. Many authors emphasize that a physician working with schizophrenia patients must possess sufficient self-esteem and sources of satisfaction in his or her nonprofessional life to avoid using the patient to meet personal needs for admiration or prestige.

Setting Intensive psychotherapy must be conducted in a setting of mutual safety. Within an organized care setting, the milieu must be ideologically supportive. Frequency of visits can range from one to five per week. Use of the couch and free associations as in psychoanalysis are discouraged as aggravating disorganization and thought disorder. As part of creating the setting for individual psychotherapy, the therapist endeavors to achieve an early consensus with the patient regarding the nature of the latter's problems, the treatment required, and the rules governing therapy.

Process Process elements in the psychotherapy of schizophrenia refer to expectable developments in the doctor-patient relationship as it evolves over time, including transference and countertransference. The management of these is considered central to the therapeutic endeavor. *Transference* broadly refers to the manner in which the perception of others in the present is shaded or distorted by important past relationships. It is thought of as a natural but often unconscious aspect of all human relationships. Examining transference as it develops and unfolds in the doctor-patient relationship is a major task in investigative (as opposed to supportive) psychotherapy—this examination is expected to be useful in allowing patients to better understand their current difficulties and respond more realistically and productively to people in their current life. It should also facilitate the recall of memories that may be accessible only through their recreation in the transference relationship.

In nonpsychotic transferences the patient perceives or responds to the therapist as if the latter resembled some other important figure from the patient's past. The patient retains the capacity to recognize these as misperceptions, to separate real aspects of the therapist from distorted aspects, and to trace the distortions back to their origin in past experiences.

James, a shy and isolated young man with schizophrenia treated as an outpatient with intensive psychotherapy and medication, was able to begin classes at a local college. Many years earlier he had first become ill and was hospitalized a few days after leaving home to attend college far from his family. An avid reader, he had long dreamed of returning to school to study English literature. As the semester progressed James became increasingly self-conscious and despite a good midterm grade began talking about dropping the course. Wishing to see his patient succeed, James' young male therapist agreed to drive him to class first once, then three times a week. To the therapist's surprise, after several weeks John's resentment erupted in a tirade about not wanting to be forced to go to class so his therapist could look good. Discussion led to recollections of the daily struggles James had had when he became afraid to attend class in junior high—at that time his resentment and school refusal had led to physical altercations with his parents at the front door. The therapist pointed out the similarity between John's past and current reactions and noted that while all parents take pride in their children's success, it was doubtful that this was his parent's principal motivation in pushing him to attend school. Patient and therapist realized that at present it had been James' idea to go to class, and that he would have to decide whether or not to continue.

The capacity to recognize earlier experiences as the source of current distortions is often absent or lost in schizophrenia, leading to transference psychosis in which the patient believes or behaves as if the therapist actually is, or is like, some figure from the past.

Frances, a woman in her 20s with paranoid schizophrenia, had been raised in a neighborhood that she claimed was kept safe and orderly by organized criminals. Furious at her family for having committed her several months earlier she claimed to have no psychiatric problems whatsoever, adamantly refused antipsychotic medications, and complained that her family had bribed the police to lock her up. Despite her extensive denial, she attended scheduled outpatient appointments with complete reliability. After several months, however, Frances began accusing the therapist of spreading the rumor that she was a prostitute and having the police follow and harass her. She called the Mayor's office to report that the therapist was taking bribes and, expecting retaliation, left the therapist's office threatening to harm herself. The therapist, who for weeks had adamantly tried to reassure the patient that he was not having the police follow her, found himself needing to obtain their assistance to arrange for the patient's detention and emergency assessment.

Transference is at times an inevitable and unavoidable development in intensive psychotherapy, including transference psychosis. Nonpsychotic transference can usually be resolved with time, support, and interpretation. Psychotic transference is more difficult to remove and a therapeutic relationship must often be developed that can accommodate this distortion while remaining useful to the patient.

Countertransference refers to all the therapist's thoughts and feelings about the patient. Some of these are distortions arising from the therapist's personal past, but others derive largely from current interaction with the patient. Feelings that arise in work with schizophrenia patients can be particularly intense and uncomfortable, and may include discouragement, fear, worthlessness, hatred, contempt, guilt, rage, envy, or lust. In view of this, awareness of countertransference and the ability through introspection to understand its sources are crucial functions for the psychotherapist.

Countertransference can often serve as an important source of information about the patient's state of mind, particularly in patients who are unable to talk.

John was a recently hospitalized, mute, schizophrenia patient. Moments after seating him for an initial psychotherapeutic interview, the therapist found himself feeling trapped and terrified by the sudden recognition that John, with a paranoid glare, was on the verge of assault. After ruling out a run for the door as too risky, the therapist looked at John and asked: “Are you afraid of me?” “Yes,” indicated a perceptibly relieved patient, “I think something bad is going to happen and you're part of it.” “Well, you're scaring the hell out of me,” said the therapist. Now appreciably calmer John said, “I'm sorry,” and a conversation ensued. Some time later John explained that on the day of his arrival he felt he had been brought to the hospital to be executed and expected at any moment to be shot in the head. Parenthetically, the therapist learned to conduct initial assessments of hospitalized patients in a place where help could be more readily obtained if needed.

The therapist's reaction may also serve as a good barometer for understanding how others typically react and respond to the patient:

Mary, a woman in her 30s, complained bitterly to her female therapist about weakness, fatigue, and an inability to function that she steadfastly attributed to her medications. Try as she might, the therapist's efforts to respond to these complaints, by changing medications, offering advice (such as exercise), and reframing the complaints amounted to nothing. Fed up with complaints that she “didn't understand,” the therapist was herself left feeling useless, angry, and very tempted to tell the patient to quit complaining and take some responsibility for herself. The therapist remembered, however, that for the past year Mary had had no contact whatsoever with her sister after the latter told Mary to quit complaining and try to make something of her life. Many of Mary's relationships with friends had ended with recriminations about their lack of sympathy and understanding.

Finally, successful management of countertransference allows the therapist to create a holding or containing relationship with the patient that is postulated to be central to the mutative action of psychotherapy.

Frank, a schizophrenia patient in his early 30s who lived in the community, had enormously ambivalent feelings about his 20-hours-per-week job as a dishwasher at a nearby hotel. Frank often found his interaction with co-workers stressful and was prone to avoid work; he was also plagued by unrelenting guilt about his inability to move up to full-time work on any sustained basis. Frank's father often reinforced this view by telling him that his illness amounted to little more than laziness, and that with will power and a strong character Frank should certainly be able to handle a full-time job. Over a period of years a clear pattern in Frank's illness was detectable: Frank would work part-time and be stable for a period. Driven by his guilt, he would increase his work hours. The stress and overstimulation of the heavier work schedule would lead to an exacerbation of symptoms and finally a fullblown relapse. He would have to quit work altogether for a period, seek a new part-time job, and the cycle would begin anew. During Frank's relapses, his primary symptom was the delusional fear that he would be “damned” and persecuted for his moral weaknesses.

Over years of outpatient psychotherapy, Frank's interactions with his therapist about work (the therapist himself being something of a workaholic) were of great interest. Whenever Frank contemplated missing a day's work (and there were many such times) he would call the therapist at home the evening before for permission, often presenting the flimsiest of excuses and rationales about why he should not work the next day. Receiving these calls late in the evening (when he himself was still working) the therapist often felt the intense urge to tell Frank to shape up, be a man, show some character, and quit acting like a little work would kill him. Recognizing this urge as countertransference (derived in part from Frank's seemingly unconscious effort to recreate his relationship with his father with the therapist, the therapist was largely able to refrain from such moralistic admonishments, discuss matters calmly, and allow Frank to decide on his own about work the next day.

Over a period of years Frank came to realize that his illness was something real rather than a moral weakness. With this realization he was able to accept the fact that for him, full-time work was probably not going to be possible. The frequent relapses ceased and Frank was able to sustain longer periods of stability. The therapist felt that his recognition of countertransference allowed him to “hold” Frank's self-loathing about his work disability and reflect this back to Frank in a more benign manner—without being overly accusatory or moralistic. To the therapist this seemed to be an important factor leading to Frank's more benign and accepting view of his own disabilities.

Transference and countertransference tend to mirror each other at any given time, and over the course of treatment a range of transference and countertransference configurations are traversed. These may recreate earlier developmental epochs. For example, typical configurations can include (1) an autistic relationship in which the patient does not express the slightest interest or even recognize the existence of the therapist—the therapist in turn feels devalued as a nonhuman object; (2) the idealizing, symbiotic interaction in which the therapist is perceived by the patient as an omnipotent, protective, and loving figure and negative feelings are projected onto others outside the dyad; here the therapist is likely to feel that he or she alone can truly understand this patient, whose problems clearly stem from the insensitivities of others; (3) the hostile, paranoid relationship in which the therapist is perceived as a bad object, untrustworthy, engulfing, and intent on harming the patient; here the therapist often feels hatred and rage at the patient's accusations and is tempted to become defensive or retaliatory, thus fulfilling the patient's expectation of others as untrustworthy.

General Technical Interventions The literature on intensive psychotherapy describes categories of interventions that roughly correspond to different phases of therapy. Although these tasks and strategies may be relevant at any point in treatment, they are often ordered sequentially: (1) establishing a relationship with the patient, (2) elucidating the patient's experience in the here and now, (3) tolerating the mobilized transference and countertransference, (4) integrating the patient's experiences into an expanded perspective of the self, and (5) working through. If therapy progresses, the accomplishment of earlier tasks allows greater attention to

be paid to subsequent ones.

Establishing a Relationship With the Patient Because of suspiciousness, disorganization, indifference, or ambivalence about human attachments, establishing a relationship with the schizophrenia patient can be challenging. Analytic strategies of passive neutrality and anonymity can easily be misinterpreted as disinterest or dislike and are generally discouraged. Consistency, straightforwardness, and an active effort to establish rapport are advocated. Within bounds, a reasonable degree of self-disclosure on the therapist's part can help to counter distortions by allowing the patient to get a fix on the therapist as a person. A relationship should be sought on the patient's terms. If the patient initially wants the therapist only to meet some immediate need (e.g., to secure discharge from a hospital or intervene with the patient with family) this is taken as the starting point and viewed positively as a sign that the therapist is seen as potentially useful. At times, engaging in activity (walking or playing a game), finding a neutral topic of common interest (sports, music), or placidly accepting periods of silence will further promote establishing a relationship. Creativity and patience are the only rules.

If the initial encounters are traversed successfully, a background feeling of security and predictability will increasingly characterize the therapy.

> After summoning the police to have Frances detained and committed, her therapist decided that transfer to a new doctor would probably be best: after all, how could this patient ever trust a therapist who, confirming her worst fears, had her picked up by the police? Wanting some closure before transferring the case, the therapist drove to the state hospital where Frances had been committed to visit her. To his surprise, the patient greeted him warmly and thanked the therapist for going out of his way to visit her. With second thoughts about transferring the patient, the therapist agreed that they would meet at his office after Frances's discharge. Discharged from the hospital on long-acting antipsychotic medication, Frances arrived as scheduled for her outpatient appointment. The therapist acknowledged the awkwardness of meeting again at the site of their last acrimonious encounter. Both patient and therapist agreed that avoiding future hospitalizations would be a common goal. Although, if asked, Frances still denied any psychiatric difficulties, she agreed to continue medication to keep her family and the doctors “off her back.”

At times, a more engaging style can promote the establishment of a relationship. Some have advocated active participation or playing with a patient's communications or symptoms as a means of capturing their attention.

> Despite trials of a variety of treatments Tim, a patient in his mid-30s, often contended that he was in continuous and intensive training by “security forces” to unite world nationalities under one government. His therapist suggested he call a summit of world leaders to further this plan. After an extensive discussion that included consideration of the site of such a meeting, logistics, security, and arrangements for translators, Tim cracked a smile; although he continued to be preoccupied with messages he received from “security forces” the world summit became an inside joke between doctor and patient.

Intellectual conversation can obscure important emotional reactions, but may be promoted to further the goal of establishing a relationship; it can also provide a patient whose self esteem has been damaged the opportunity to exercise areas of competence.

> Donna, a 27-year-old woman, experienced an acute psychotic episode during her second year as a graduate student in comparative religion at a first-rate university. While hesitant to discuss the events leading to her hospitalization, she easily discussed the papers she had to complete to secure readmission to the program. Impressed by Donna's vital engagement with the subject and her ability to present it in a compelling manner, for several weeks the therapist took on the role of a uninformed student glad to receive a survey of comparative religion from such a competent teacher. The therapist felt that acknowledging the patient's intellectual ability and expertise in comparative religion made it considerably easier for the patient to later accept the therapist's expertise in medicine and psychiatry.

Elucidating the Patient's Experience Elvin Semrad viewed the three core tasks of psychotherapy as helping the patient acknowledge, bear, and put into perspective feelings and painful life experiences. Acknowledging the patient's feelings and painful experiences in the present becomes particularly pertinent once a relationship has been established. Acknowledging first requires elucidating affects. Strategies for elucidating include listening, narrowing the focus, seeking concrete detail, acknowledging feelings (especially of loss, anger, sadness), and naming or labeling affects. The therapist may act as a comforter, inquisitor, or teacher, conveying to the patient that experiencing feelings will neither overwhelm the patient nor hurt others. Psychotic symptoms, when expressed, are considered to signal an affective reaction to some actual event that patient and doctor do not yet understand. Examining the patient's day-to-day life in detail will allow the therapist to develop a more vivid picture of the patient's difficulties, frustrations, and characteristic reactions to others. The aim is to help the patient better organize and communicate, to guide the patient into sharpened conceptualizations, and to promote tolerance of life experience as it is.

If successful, therapist and patient will share a common language with which to communicate about the latter's difficulties and increasingly the patient will independently report important life events and emotional reactions to them.

> During a therapy hour, John, a patient in his 20s with treatment-resistant schizophrenia reported that under his control, a nuclear attack had just been launched in Europe and that as he and the therapist sat, millions of people were being killed. With single-minded persistence, despite John's evasive disorganization, the therapist attempted to find out what John had been doing that morning before his therapy appointment. The therapist was able to piece together that immediately preceding the “nuclear attack” John had been listening to the radio and another patient on his unit had changed the station without asking. The therapist suggested, and John acknowledged, that this had made John a bit irritated. The therapist suggested that perhaps there was a connection between John's anger at his peer and his current concern with millions being killed. The therapist wondered whether next time something like that happened it would be possible for John to temperately register his protest to his peer. John indicated this was out of the question—he did not want to be tried for murder. The therapist indicated that in his view, a mild protest was quite different from murder.

Tolerating the Mobilized Transferences and Countertransferences Tolerating affects, transference, and countertransference corresponds to Semrad's concept of “bearing” painful feelings that have been acknowledged. Tolerance is achieved first by the therapist and then, through example, by the patient. Here the concept of “holding” or “containment” is relevant and is thought to be central to the mutative effect of psychotherapy. By means of *projective identification* the patient is thought of as putting negatively valenced affects and self-representations into the therapist who, by processing these feelings in a more mature way, contains, holds, or metabolizes them and makes them palliatively available for reinternalization by the patient. More simply stated, patients experience themselves being accepted, negative emotions and all, and learn from the therapist's example to become better able to accept unwanted aspects of themselves. Thus, patients' identification with the therapist and their functioning is seen as a major factor in the therapeutic action of psychotherapy.

> Having exhausted her ability to respond constructively to Mary's unending complaints, the therapist found herself viewing Mary with contempt and derision, comparing her unfavorably to patients with severe physical disabilities who had struggled courageously to accomplish goals despite their disability. Upon reflection, the therapist became vividly aware of the manner in which Mary's view of herself as the victim of other people's insensitivity had become a self-fulfilling prophesy. Mary's inconsolable complaining was so irritating that these accusations, if not true today, would be true tomorrow: the therapist or anyone else who had to deal with Mary would soon find themselves irritated and would run out of patience. Concluding that Mary's own feelings of incompetence and self-hatred had been projected onto her, the therapist found her contempt for Mary tempered by a better understanding of Mary's suffering. The therapist hoped that if this could be conveyed to Mary, the vicious cycle of Mary's turning nearly everyone against her might be interrupted.

Integrating the Patient's Experience into an Expanded Perspective of Self Broadening patients' understanding of themselves and their situation corresponds to the third part of Semrad's triad: helping patients put into perspective their painful affects, life experiences, and maladaptive solutions. According to Sullivan “No one has grave difficulties in living if he has a very good grasp of what is happening to him.” Thus, providing insight is another way in which psychotherapy is thought to be useful, complementing identification with the therapist. Integrating the patient's experience entails a change in therapeutic relationship and enlists interpretation as its major technical tool.

At this phase in therapy the nature and tenor of the therapeutic relationship changes such that the therapist becomes more demanding, frustrating, and insistent on adaptation, reality testing, and health. While remaining supportive, the therapist increasingly confronts defenses, frustrates wishes, and interprets transference. The major task is to accept the patient but to reject the patient's psychosis and maladaptive interpersonal maneuvers. Insight in this context can occur at several depths: (1) as simple recognition of the fact of illness, (2) as knowledge about the nature of the illness (e.g., that hallucinations come from one's own mind, (3) as recognition of the dynamics of the illness (i.e., that symptoms occur in relation to personal difficulties), and (4) as recognition that symptoms solve problems and conflicts. Emotional insight, gained by direct experience in the doctor-patient relationship, is emphasized. These derive from interpretations pointing out the transferential

nature of the patient's feeling towards the doctor, including their origin in the patient's past experience and their inappropriate application in the patient's current everyday life.

The what, how, and when of interpretation stems from an empathic attunement to the patient's tolerance, although certain recommendations recur repeatedly. Most psychotherapists are reluctant to interpret content (particularly sexual) since patients are often flooded with unwanted thoughts and impulses. Rather, interpretations focus on defensive operations, resistances to therapy, and the link between symptoms and everyday stresses. Especially targeted for interpretation are negative transference, aggressive impulses, depressive concerns, and dependency issues, which are often warded away from consciousness because of the patient's fear that their emergence would be overwhelming. Interpretations are presented in a way that is aimed at helping patients learn to formulate interpretations on their own. They are presented as tentative observations in the spirit of mutual inquiry. Short, simple, nontechnical language should be used and the patient is considered the final judge of an interpretation's validity and usefulness.

If traversed successfully, this phase of therapy will leave patients with a more accepting and complex view of themselves as a people capable of experiencing the full range of human emotions. Patients will no longer view all their difficulties as the product of mistreatment at the hands of others, and will assume more responsibility for their own treatment and health. Passivity and indecision will be seen as an active choice, and patients will recognize that continued progress depends on their willingness to attempt new solutions, both inside and outside of treatment.

Working Through With improvement in psychosis and maturation of the patient's nonpsychotic personality, the phase of integrating evolves into the last phase of working through. Patients become better able to help therapists perform their functions and eventually become capable of performing these functions themselves. Patients see their therapists as real, different, and imperfect in their empathy. Each step towards greater independence and autonomy generates separation anxiety and a “two steps forward and one step back” trajectory can be anticipated. Regressions or symptomatic exacerbations recur but should be shorter, less intense, and more readily influenced by interpretation. The end of treatment may be negotiated, but many authors have noted that not uncommonly patients remain attached and may contact the therapist during crucial junctures many years after therapy has ended.

SUPPORTIVE PSYCHOTHERAPY

Goals Supportive psychotherapy has historically been favored by biologically and pharmacologically oriented clinicians. It is firmly grounded in the medical model in which the patient is seen as suffering from an organically based illness that requires treatment from a physician.

As described by Talcot Parsons nearly 50 years ago, the medical model implies essential elements of expected behavior for both the physician and patient. These elements define the physician's and patient's respective roles, relationship, and responsibilities.

The physician's role is characterized by four key qualities: it is (1) universalistic, (2) functionally specific, (3) affectively neutral, and (4) collectivity oriented. The universalistic norm requires the physician to treat all patients alike according to scientific and medical standards. The role is functionally specific in that the physician is seen as a specialist in health and disease who is expected to limit attention to circumscribed areas involving medical matters. Affective neutrality prevents the doctor from entering too sympathetically into the patient's situation, allowing for steadfastness of judgment and the exercise of emotional control. Finally, collectivity-orientation, as opposed to self-orientation, demands that the doctor treat the patient according to the patient's needs and the health standards of the community.

Like the role of the physician, the role of the patient is defined by expected behavior that involves both rights and obligations. First, the person who is ill is exempt from normal social responsibilities and excused from customary obligations so as to attend to the process of getting well. A second right is exemption from responsibility for illness—the illness is not considered the patient's fault and the patient has the right to receive care. At the same time, the patient has the obligation to want to get well, obtain technically competent help, and cooperate with treatment.

In contrast to the ambitious aim of personality change associated with the intensive therapy tradition, the short- and long-term goals of supportive psychotherapy are comparatively modest. These include (1) relief from the immediate crisis or direct reduction of acute disequilibrium, (2) removal of symptoms to premorbid levels, (3) re-establishment of psychic homeostasis through a strengthening of defenses, (4) sealing over psychotic experiences and conflicts, (5) the circumscribed fostering of adaptation, and (6) mobilization and preservation of healthy aspects of the patient to enable optimal functioning and minimize the impact of persistent deficits. Supportive therapy uses the physician-patient relationship to create a background of adequate clinical care that supports the prescription of effective pharmacological interventions. Functional or social recovery, rather than personality change, is the primary aim of treatment.

Techniques The overall technical approach of supportive psychotherapy is one of pragmatism and management in which based on medical and psychiatric expertise, the physician helps the patient interpret and adapt to reality. As such, the therapist employs techniques that include defining reality, offering direct reassurance, giving advice on current problems of living, urging modification of expectations, and actively organizing the environment for patients who cannot do so themselves. To help stabilize the patient's environment the therapist often maintains close contact with the patient's family or other treaters and may intervene on the patient's behalf with family, employers, and social agencies.

Eliciting and tracking symptomatology and targeting symptoms for psychopharmacological intervention is a major focus for the supportive psychotherapist. Psychopathology is interpreted in a medical context as the unwanted emergence of signs of illness. The basic content of psychotherapy focuses on teaching and relearning—the patient is educated regarding the nature of the illness, taught to monitor symptoms, and act promptly to suppress their exacerbation. The therapist fosters positive transference as a benign authority; positive feelings are treated as real and negative transference is avoided. The therapist may become very active in helping the patient learn new ways of adapting and may use or prescribe cognitive, behavioral or social skill training techniques.

Empirical Studies

Efficacy of Individual Psychotherapy In the decades following the introduction of phenothiazines psychiatry became increasingly divided into adherents of the psychodynamic and biological paradigms. Disagreement concerning the value of intensive psychotherapy became a focal point of often acrimonious ideological and scientific debates. Randomized clinical trial methodology unambiguously demonstrated the value of pharmacological interventions in schizophrenia and came to be seen as the optimal standard for evaluating all treatments. In this context, five studies conducted during the 1960s and 1970s attempted to assess the efficacy of various forms of individual psychotherapy compared to treatment programs that did not specifically feature psychotherapy. Although criticized by proponents of intensive psychotherapy on a number of methodological grounds, together the results of these randomized clinical trials provided little or no evidence for the efficacy of psychotherapy as the sole treatment for schizophrenia. Supporting this conclusion were the results of long-term follow-up studies both at Chestnut Lodge, where many of the techniques of intensive individual psychotherapy were developed, and at Columbia's Psychiatric Institute. Exceptions notwithstanding, these studies found that the majority of patients treated with intensive psychotherapy alone remained seriously and chronically ill.

Reflecting the then-extant ideological rivalry between psychodynamically and biologically oriented clinicians, the Boston Psychotherapy Study was designed in the 1970s to address the methodological weaknesses of earlier clinical trials. This study aimed to evaluate the comparative effectiveness of expressive, insight-oriented individual psychotherapy and reality-adaptive, supportive psychotherapy against a backdrop of high-quality inpatient, outpatient, and pharmacological treatment provided to both patient groups. Contrary to the investigator's expectations, neither treatment emerged as clearly superior, although differential effects across outcome domains were noted: reality-adaptive, supportive therapy was preferentially effective in the areas of recidivism and role performance while expressive, insight-oriented therapy exerted a modest preferential effect on ego functions and cognition.

The disappointing results of randomized clinical trials and follow-up studies contributed substantially to a decline in prestige and influence of the psychodynamic paradigm generally, and intensive individual psychotherapy of schizophrenia in particular. An additional challenge came from infant observational research that indicated that infants were far more active, stimulus seeking, and socially oriented from birth than was suggested by earlier psychodynamic theory. Similarly, stages such as the autistic and symbiotic were not observed and the infants' plasticity, resilience, and capacity for adaptation appeared far greater than earlier thought. These findings challenged the validity of psychosocial theories of etiology, at least as articulated by psychodynamic thinkers. Because of these and other factors, individual psychotherapy research and psychological theorizing concerning schizophrenia slowed to a near halt. The biological paradigm decisively gained ascendancy as the most influential in the field.

Reappraisal of Individual Psychotherapy Significant among the findings from the Boston Psychotherapy Study was the degree to which, despite theoretical differences, the actual techniques employed by the expressive, insight-oriented therapy and reality-adaptive, supportive therapists tended to converge. For example,

both theories were found to employ substantial supportive elements. Sobering but significant were the substantial attrition rates for both types of therapy. Although those who remained in therapy continued to accrue benefits, by the end of 1 year more than half (56 percent) and by the end of 2 years more than two-thirds (69 percent) of all patients had unilaterally dropped out of treatment.

Having found few differences in overall outcome between patients treated with expressive, insight-oriented therapy and reality-adaptive supportive therapy, investigators from the Boston Psychotherapy Study searched for common factors in the treatments associated with positive therapeutic change and good outcome. Results indicated that independent of severity of psychopathology, patients able to form a good alliance with the therapist within the first 6 months of treatment were more likely to remain in therapy and to comply with medication. These patients achieved better outcomes at 2 years and used less medication than those who did not form a therapeutic alliance. Looking at therapist activity, these investigators found across both therapies a strong positive correlation between reductions in patient denial of illness and retardation-apathy, and the therapists' demonstration of a sound dynamic understanding and accurate attunement to the patients' underlying concerns. Directive activity was associated with reductions in anxiety and depression.

Rancorous debates concerning the relative value of drugs versus psychotherapy collapsed in the 1980s under the weight of data in opposition to the view that any single treatment is definitive for schizophrenia. Lack of efficacy weighed against a sole reliance on investigative psychotherapy whereas high rates of medication noncompliance, relapse, and persistent dysfunction defined the limits of a narrow medically oriented approach. The historical distinction between supportive and exploratory therapy was no longer salient.

FLEXIBLE PSYCHOTHERAPY

Flexible psychotherapy refers to a broad and pragmatic approach to psychotherapy that relies on a variety of strategies applied flexibly depending on the individual patient's type of schizophrenia and phase of illness. Such an approach might at various times include supportive, directive, educational, investigative, and insight-oriented activity, provided in the context of an ongoing and stable doctor-patient relationship. The quality of this relationship should be characterized by empathy and a sound dynamic understanding of schizophrenia on the part of the physician. Dogmatic or rigid adherence to a single approach applied to all patients is probably the least likely to be of value. As described here in terms of assumptions, clinical tasks, and interventions, flexible psychotherapy approximates how most patients are treated in current practice. A specific form of flexible psychotherapy, *personal therapy*, has recently been sufficiently operationalized to allow for rigorous empirical evaluations.

Flexible Psychotherapy: Assumptions A flexible approach to psychotherapy is based on a revised set of assumptions about the nature of schizophrenia that recognizes the joint contributions of biological, psychological, and social or environmental factors:

1. The stress-diathesis or vulnerability-stress model represents the best available integration of data pertinent to the etiology, course, and outcome of schizophrenia. The model postulates that schizophrenia results from a dynamic interaction between environmental or experiential stress in a person who is vulnerable to react to this stress with schizophrenic symptom formation.

   The vulnerabilities to schizophrenia are likely to be multiple and heterogeneous. Family and twin studies unambiguously indicate a genetic contribution to vulnerability, although a simple Mendelian mode of inheritance is not evident. Epidemiological investigations point to the contribution of acquired, biologically based environmental risk factors such as maternal influenza, Rh incompatibility, maternal starvation, and obstetrical complications. Although a variety of specific vulnerabilities have been identified, none has been shown to be ubiquitous.

   Biologically based vulnerability may be expressed as relatively enduring psychophysiological, cognitive, or behavioral difficulties among subgroups of patients. Most common findings include (1) deficits in information processing and maintaining a steady focus of attention; (2) dysfunctions in psychophysiology suggesting deficits in sensory inhibition and autonomic responsivity; (3) impairments in social competence; and (4) general coping deficits such as overvaluing threat, underappraising abilities, and extensive use of denial.

   The vulnerability to schizophrenia is seen as a relatively enduring proclivity to developing overt clinical symptoms; vulnerability is likely to be manifest as a set of stable traits present premorbidly, at onset, during acute episodes, and remissions. However, vulnerability is not static but is shaped epigenetically over time by environmental influences; a stress sufficient to precipitate relapse at one time, for example, may be less likely to do so at a later point when new coping strategies or better supports have been acquired.
2. The stress side of the vulnerability-stress model postulates that a variety of stressors (internal or external events requiring adaptation) can precipitate the emergence of symptoms in a vulnerable individual. Given biologically based vulnerability, the onset, course, and outcome of an individual's disorder may be shaped largely by interactions between the person and the environment. Among psychosocial factors, stressful life events, cultural milieu (egocentric versus sociocentric), social class, social network size and density, and emotional quality of the living environment have been demonstrated to be associated with the onset or course of schizophrenia.

   The vicissitudes of illness are best understood as a dynamic product of the affected individual's adaptive assets and vulnerabilities interacting over time with various stresses. In a highly vulnerable individual, sufficient stress can precipitate intermediate (prodromal) states of dysfunction that amplify preexisting cognitive, affective-autonomic, and social coping deficits. In the absence of adaptive strategies or environmental supports these interact negatively with the existing stressors to magnify their effect in a downwardly spiraling process that ends in a full-blown clinical syndrome.

   Clinical experience indicates that the stresses associated with illness onset or exacerbation may be highly individualized, rendering generalization about the typical nature of such stresses difficult. Stresses may be primarily biochemical (as in substance abuse), environmental (as in leaving for college, joining the armed services, breaking up with a girlfriend), or social (poverty, unemployment).
3. Schizophrenia is heterogeneous, as are individuals afflicted with it. The clinical diversity of schizophrenia in relation to vulnerabilities and risk factors, age and type of onset, manifest signs and symptoms, longitudinal course, and long-term outcome suggest that the disorder may be heterogeneous in regard to underlying etiology. This heterogeneity may be partially captured by currently available subtyping systems. Paranoid schizophrenia, for example, is associated with good premorbid functioning, late age of onset, many positive symptoms, intermittent illness over the first several years, and a comparatively high likelihood of good outcome despite a higher risk of suicide; the deficit form of schizophrenia is characterized by poor premorbid functioning, earlier onset, severe and enduring negative symptoms, a low risk of suicide, and often persistent life-long disability. However, an etiology-based subtyping that allows for precise longitudinal prediction for individual patients is not available. At a minimum, schizophrenic illnesses of greater and lesser severity and virulence can be identified and the biological vulnerability of individuals may differ.

   Like the illness itself, individuals afflicted with schizophrenia differ substantially in adaptive capacities, intelligence, and instrumental and verbal competence. Furthermore, the degree of social support available to them varies greatly. In general, the greater the level of instrumental skills acquired prior to the onset of illness (work or educational experience, experience with relationships, experience living independently), the better positioned the individual will be to recover or maintain functioning once the illness has become established. For some patients the failure to acquire much in the way of adaptive skills may represent the product of the life-long vulnerability underlying the tendency to the illness. For other patients a particularly virulent form of schizophrenia may catastrophically erode adequate premorbid skills, resulting in substantial permanent disability. Other patients with good premorbid abilities may be left with their skills and competence largely intact after an acute episode.
4. Schizophrenia is often phasic in course. Systematic investigation of longitudinal course has only recently begun and the understanding of illness phases is preliminary. Phases may include (1) prodromal periods, during which a highly individualized constellation of symptoms that represent early manifestation of clinical decompensation emerges; (2) acute or active phases, often associated with the full-blown emergence of positive symptoms superimposed onto pre-existing deficits; (3) subacute, convalescent, or stabilization phases characterized by gradual restoration of some functioning perhaps associated with postpsychotic depression; (4) moratoriums or adaptive plateaus, characterized by a gradual reconstitution of identity, gathering of support, and strengthening of skills; (5) change points or shifts in functioning over a relatively brief period of time, initiated by the patient's own desires or pressure from others and associated with the potential for either quantum improvement or decompensation; (6) end-state or stable plateaus, relatively enduring periods of stability characterized by greater or lesser fixed deficits or chronic levels of positive symptoms.

Flexible Psychotherapy: Clinical Tasks and Technical Strategies To treat schizophrenia the therapist must use a variety of interventions and strategies. The crucial question is which interventions are of potential value for a particular individual at a particular phase of illness. All interventions aim to minimize the effect of vulnerabilities, bolster adaptive capacities, and reduce the extent and impact of stress. The range of therapeutic tasks and associated goals and interventions can be ordered hierarchically. As outlined in Table 12.10-1, different therapeutic tasks are of particular importance during different illness phases. In addition, although some

tasks are clearly relevant for all patients receiving individual treatment; other tasks, particularly those relating to the goals of intensive psychotherapy, are pertinent for only a small subgroup of patients. The strategic rosetta stone here is the therapist's capacity to shift gears, be flexible, and change roles with all patients based on changing circumstances, always keeping in mind the goal of helping the patient accept, learn about, and self-manage what may often be a chronic and devastating illness.

Table 12.10-1 Flexible Psychotherapy: Therapeutic Tasks, Interventions, and Goals

Consideration of the patient's schizophrenia subtype, current and premorbid functioning, and self-defined treatment goals are all relevant to the determination of appropriate treatment tasks. For patients with severe disorganized or deficit forms of schizophrenia, for example, the most humane and practical goal may be establishing a supportive ongoing treatment within a sheltered setting that minimizes stress and provides for basic human needs for an indefinite period. For the majority of patients who reside in the community amidst varying supportive structures, some degree of psychoeducation and rehabilitative tasks should be planned with the aim of minimizing acute relapses and promoting maximal functioning and quality of life. A primary focus on investigative tasks should be reserved for motivated patients who have established a good working relationship with the therapist and exhibit an interest in and ability to make constructive use of such techniques. These patients are likely to have demonstrated good premorbid functioning, intermittent and less severe forms of schizophrenia, minimal residual deficits, and retention of some capacity for self-observation, curiosity, tolerance of frustration, and humor. Attunement to psychological concerns may be particularly important for patients who have a dramatic response to new medications.

The use of multiple treatment modalities creates the need for someone to orchestrate and coordinate them. In many instances a psychiatrist is best at providing individual therapy and continuity of care over a prolonged period. As is true in medicine generally, the quality of the individual doctor-patient relationship is a major factor in the success of the therapeutic endeavor. Thus, a focus on the skillful use of this relationship usefully informs all tasks at all levels. Here, removed from its outmoded etiological assumptions and overly ambitious aims, the substantial clinical knowledge derived from the tradition of investigative psychotherapy can be applied pragmatically in a contemporary context.

Many of the tasks overlap with the concerns and expertise of other service providers; however, all the tasks should be the concern of the individual psychotherapist and a focus for individual psychotherapy. A common mistake made by trainees is to focus on higher-level psychological tasks while ignoring overwhelming difficulties at the level of basic human services. Thus, the necessary first goal of psychotherapy with a homeless person may be direct assistance in finding housing. Although other professionals may be relied upon to accomplish specific tasks, physician-therapists should consider themselves responsible for ensuring the results of these efforts.

The following general treatment strategies are common to all the specific therapeutic tasks outlined:

1. Evaluation.A thorough evaluation of the patient initiates the treatment process. During medical assessment and stabilization this includes ruling out identifiable physical conditions, assessing competence to consent to treatment and risk, and determining the response of symptoms to short-term pharmacological intervention. Psychosocial assessment inventories available supports and aims to measure the degree to which the patient's adaptive capacities measure up against the stresses and demands of their living environment. Efforts to establish a supportive ongoing treatment test the patient's capacity to trust and rely upon another human being for support and guidance. When applicable, psychoeducational, rehabilitative, and investigative interventions are preceded by an assessment of the patient's cognitive strengths and deficits, allowing interventions to be formulated that match the patient's talents.
2. Continuous reevaluation. The fluid nature of schizophrenia and an individual's adaptation to it over time demands periodic reassessment of course, prognosis, phase of illness, and target problems; as these change, so do treatment goals. Providing concrete support in the form of a ride to work may be helpful early in the effort to promote vocational rehabilitation, but later may promote unwarranted dependency and prolong disability.
3. Timing. The phasic natural history of schizophrenia requires attention to the timing of particular therapeutic tasks. For many patients, in order to minimize stress and forestall relapse, relatively little beyond assessment, stabilization with medication, and establishment of a supportive ongoing treatment should be attempted during the first 6 to 12 months following an acute episode. Once the patient is asymptomatic and shows signs of revitalization, rehabilitation and more complex psychoeducational elements may be gradually introduced.
4. Titration.Treatment interventions should be applied with graded increases of intensity and complexity. Higher-level therapeutic tasks should be attempted and higher levels of work or social functioning expected only after completion and consolidation of earlier gains. Substantial rehabilitation, for example, will rarely be possible until progress has been made in attaining a stable supportive treatment relationship. Similarly, there is evidence that early, active, and ambitious psychologically oriented treatment may be disorganizing or toxic for certain patients. In general, treatment changes should be pursued cautiously, with only one element being modified at a time.
5. Integration with psychopharmacology. Each of these tasks take acute and prophylactic antipsychotic drugs as given for most patients. Control and prevention of psychotic symptoms using the lowest effective dosage of medication is the overall treatment goal. Decisions regarding pharmacological management are often linked to the relative success or failure of accomplishing various psychotherapeutic tasks. Considerable psychoeducation, for example, should be accomplished before attempting maintenance medication dose reduction or the initiation of a targeted (intermittent) medication strategy. Long-acting injectable antipsychotic agents may be useful for patients who are unable to tolerate the daily reminder of illness associated with oral medication or for patients who are unable to maintain a reliable treatment relationship.

Personal Psychotherapy While reflecting clinical practice, flexible psychotherapy has not been formally assessed. Similar in overall outlook to flexible psychotherapy, however, *personal therapy* has recently been developed by Gerry Hogarty and his colleagues at the University of Pittsburgh as a form of individual treatment that is sufficiently operationalized to allow empirical testing. Personal therapy is designed for recently discharged outpatients with chronic or subchronic schizophrenia. Its objective is to enhance personal and social adjustment and forestall late (third-year) relapse. A disorder-specific treatment, it accommodates neuropsychological aspects of schizophrenia and attempts to avoid the adverse effect of poorly timed interventions.

Within a stress-vulnerability model individual-specific stress, often interpersonal, is seen as precipitating affective dysregulation. This loss of control over mood is seen as resulting in poorly reasoned dysfunctional behavior that negatively influences the reciprocal behavior of others in a cycle that may end in relapse. Based on individual patients' needs, personal psychotherapy uses a range of interventions to promote patients' self-awareness and foresight and to equip patients with adaptive strategies that facilitate self-monitoring and self-control of affect.

Personal psychotherapy includes three phases, each with explicitly defined goals and corresponding interventions. The achievement of these goals is carefully assessed before the patient advances to the next level of treatment. Phases of treatment, goals, and operational criteria defining readiness to move to more advanced phases are summarized in Table 12.10-2. Within each phase, the exposure of patients to specific interventions is varied based on individual need. While the therapy was designed to be given over a 3-year time span, patients spend as much time at each level as required to meet advancement criteria and not all patients progress through all three phases. Personal therapy is administered against a backdrop of psychopharmacological treatment that aims to minimize adverse effects by using the lowest medication dose needed to prevent symptom exacerbation.

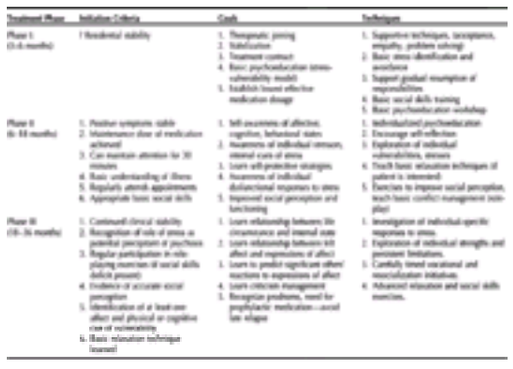

Treatment Phase	Initiation Criteria	Goals	Techniques
Phase I (1–6 months)	↑ Residential stability	1. Therapeutic joining 2. Stabilization 3. Treatment contract 4. Basic psychoeducation (stress-vulnerability model) 5. Establish lowest effective medication dosage	1. Supportive techniques, (acceptance, empathy, problem solving) 2. Basic stress identification and avoidance 3. Support gradual resumption of responsibilities 4. Basic social skills training 5. Basic psychoeducation workshop
Phase II (6–18 months)	1. Positive symptoms stable 2. Maintenance dose of medication achieved 3. Can maintain attention for 30 minutes 4. Basic understanding of illness 5. Regularly attends appointments 6. Appropriate basic social skills	1. Self-awareness of affective, cognitive, behavioral states 2. Awareness of individual stressors, internal cues of stress 3. Learn self-protective strategies 4. Awareness of individual dysfunctional responses to stress 5. Improved social perception and functioning	1. Individualized psychoeducation 2. Encourage self-reflection 3. Exploration of individual vulnerabilities, stressors 4. Teach basic relaxation techniques (if patient is interested) 5. Exercises to improve social perception, teach basic conflict management (role-play)
Phase III (18–36 months)	1. Continued clinical stability 2. Recognition of role of stress as potential precipitant of psychosis 3. Regular participation in role-playing exercises (if social skills deficit present) 4. Evidence of accurate social perception 5. Identification of at least one affect and physical or cognitive cue of vulnerability 6. Basic relaxation technique learned	1. Learn relationship between life circumstances and internal state 2. Learn relationship between felt affect and expressions of affect 3. Learn to predict significant others' reactions to expressions of affect 4. Learn criticism management 5. Recognize prodrome, need for prophylactic medication—avoid late relapse	1. Investigation of individual-specific responses to stress 2. Exploration of individual strengths and persistent limitations. 3. Carefully timed vocational and resocialization initiatives. 4. Advanced relaxation and social skills exercises.

Table 12.10-2 Personal Therapy: Goals, Techniques, and Criteria for Advancement

Investigators at the University of Pittsburgh have recently completed two 3-year randomized controlled trials of personal therapy for newly discharged patients with schizophrenia and schizoaffective disorders. Patients residing with their families were assigned to supportive therapy, personal therapy, families psychoeducation or management, or a combination of the latter two treatments. Patients living alone, who were generally more disabled, were assigned to personal therapy or supportive therapy.

Results indicated that personal therapy was remarkably well accepted by patients participating in the trials. Over 3 years only 8 percent of patients receiving personal therapy and 23 percent of patients in contrasting treatments, were dropped for noncompliance or administrative reasons. The efficacy of personal therapy in relapse reduction was tied to residential status. Patients receiving personal therapy who were living with family experienced fewer relapses. The more impaired group of patients receiving personal therapy who were living alone experienced a greater relapse rate. Consistent with the clinical dictum that psychologically oriented treatments can be futile or harmful when applied before basic human service needs are addressed, personal therapy patients who relapsed were more likely to have unstable housing and difficulty securing food and clothing.

Independent of relapse reduction, personal therapy produced substantial differential improvements in social adjustment and role performance. Whereas improvements in social adjustment among patients receiving supportive and family therapy reached a plateau at 12 months, the personal adjustment of personal therapy patients continued to improve in the second and third postdischarge years with no evidence of a plateau. Relative to supportive and family therapy, individual psychotherapy was superior in promoting a progressive improvement in psychosocial adjustment.

Results from the Pittsburgh clinical trial of personal psychotherapy provide the first rigorous empirical support for the efficacy of disorder-specific and flexible form of individual psychotherapy as part of an overall treatment plan that includes psychopharmacology, attention to social service needs, and rehabilitation. Although research psychotherapies often have a limited impact on clinical practice, personal therapy is not tightly prescriptive. Rather, it provides a general strategy for ordering clinical work that leaves considerable room for individualization. Some form of individual psychotherapy combined with medication is the most common treatment provided to outpatients with schizophrenia. Personal therapy provides a previously lacking empirically supported framework to guide the provision of usual care.

Targeted Individual Psychotherapies Working within an overall clinical framework that takes the necessity of multidisciplinary and individualized treatment as a starting point, several new individual psychotherapies that target specific problems or subgroups of patients have been developed and tested. These psychotherapies have generally been time-limited and administered as an adjunct to usual long-term care that includes medical management and medication. These illness-phase–specific treatments use circumscribed and measurable goals to target improvements in medication-resistant positive symptoms or treatment compliance.

Elizabeth Kuipers and her colleagues at London's Institute of Psychiatry described the result of a randomized trial of cognitive-behavioral therapy for stabilized schizophrenia outpatients with at least one persistent medication-resistant positive psychotic symptom. The cognitive-behavioral intervention began with active efforts to promote engagement and a therapeutic alliance and later used individually tailored skills training, relaxation, and psychoeducational and exploratory techniques. Over a 9-month trial period, individual therapy recipients had a low dropout rate, expressed high levels of satisfaction with treatment and compared to a standard treatment control group showed a 25 percent reduction in symptom severity as measured by the Brief Psychiatric Rating Scale. Patients rated at baseline as showing some cognitive flexibility about delusions were most likely to benefit from this psychotherapeutic intervention.

Nicholas Tarrier and his colleagues at University of Manchester conducted a randomized trial of intensive cognitive-behavioral therapy twice a week for 10 weeks (20 sessions) as an adjunct to usual care for compliant outpatients with medication-resistant positive psychotic symptoms at a residual phase of illness. To control for nonspecific factors, comparison groups included a generic supportive counseling with routine care and routine care–only group. Compared with routine care alone and routine care with supportive counseling, significant improvement in the severity and number of positive symptoms was found in those treated with cognitive therapy 3 months after treatment. Compared with both treatment groups, patients receiving usual care alone experienced more relapses and days hospitalized.

A brief individual psychotherapy termed compliance therapy has been designed for acutely psychotic patients with chronic illness who have recently been readmitted for an illness relapse. This intervention consists of 4 to 6 cognitive-behaviorally oriented sessions lasting 20 to 60 minutes twice a week administered prior to hospital discharge. The sessions focus on (1) a review of symptoms and treatment adverse effects, (2) the patient's actions and beliefs regarding treatment, and (3) reducing stigma comparing pharmacotherapy to the need for medicine in other physical illnesses and defining drug treatment as a freely chosen strategy to enhance quality of life. In a randomized, controlled trial, Anthony David and his colleagues at the Maudsley Hospital reported that inpatients receiving compliance therapy showed greater postdischarge insight and more observer-rated treatment compliance compared with nonspecific counseling. In addition, survival in the community and social functioning were significantly better in the treatment group over an 18-month follow-up period.

Prodromal Phase Intervention Many investigators have noted in recent studies of first-episode schizophrenia that patients typically experience active psychosis for 12 to 24 months before obtaining treatment. The observation that a longer duration of untreated psychosis is associated with a poorer initial pharmacological treatment response and more severe, chronic disability has rekindled interest in early detection and intervention as potential means to prevent the onset of severe symptoms or positively modify the long-term course of schizophrenia. Specialized early intervention programs have been developed in several countries and include individual therapy and case management techniques modified to meet the needs of patients who may be in the prodromal phase. Although no approach to early intervention has undergone rigorous empirical testing, a cognitive-behavioral focus on assisting the person to adapt to the onset of psychotic illness and its effects on self-concept, identity, and self-esteem has been described by Patrick McGorry. Limiting the damage to personal identity, social networks, and role functioning and promoting recovery and adaptation are treatment goals. Providing treatment in a less stigmatizing general youth center, and targeted use of low-dosage antipsychotic medication is a feature of this treatment. In addition, recognizing that many young people who recover from a prodromal state experience a flight into health and a determination to discontinue medication, therapists remain willing to maintain clinical contact without enforcing medication adherence. In this way medication can be quickly reinstituted if required.

Practical Considerations in Outpatient Assessment When scheduling an initial outpatient visit for a patient with schizophrenia the therapist should set aside sufficient time (11/2 to 2 hours) to conduct a thorough preliminary assessment. If the referral is initiated over the phone by someone other than the patient, it is useful to use the phone contact to obtain a cursory outline of the patient's history and current mental status. Information about symptom severity, current medications, and current and past suicidality and aggression should be sought with the aim of determining whether outpatient evaluation can proceed safely. If preliminary contacts suggest the possibility of a need for hospitalization or other immediate short-term care, specific information about what such care resources the patient is eligible to access should be obtained.

It is common for patients with schizophrenia to arrive for a first appointment accompanied by a family member, case manager, or other caregiver. Following introduction to the patient and those arriving with him or her, an initial assessment interview can be conducted with the patient alone. In this interview the therapist may need to make an active effort to promote the patient's comfort. This can be done by, for example, offering coffee, pointing out the specific place to sit, outlining what will be discussed, and if necessary conducting the assessment interview with a specific set of questions. With the patient's permission, it is then often useful to spend some time alone with the accompanying family member or other caregiver. This interview allows the person accompanying the patient to express specific

concerns or worries in private and can provide important additional information about the patient's situation.

Ideally, the outcome of the initial visit will be a mutually agreed upon plan for further assessment or treatment. This plan should include the frequency and duration of visits, payment, medication regime, and arrangements for the patient and caregivers to reach a physician (or other team member) in the event of a crisis between scheduled appointments. Attention should also be given to practical considerations such as transportation to appointments and how and where prescriptions will be filled. If psychotherapy is recommended, a general statement of its methods and goals may be useful ("We will meet so that we can talk together, better understand your difficulties, and work with you on your medications to improve your situation"). In addition, defining some mutually agreed upon area (problem, concern, goal, medication adverse effects) in which the therapist can be seen as potentially useful to the patient will set the stage for a positive therapeutic relationship.

The frequency and duration of psychotherapy are individualized. Weekly psychotherapy for 45 to 50 minutes is most common in an outpatient setting, but the frequency of visits may be increased during periods of clinical instability or if insight-oriented psychotherapy is prescribed. Less frequent visits of shortened duration (15 to 20 minutes) may be negotiated during periods of stability, for patients who have learned to self-manage their illness, or for those who find contact with a physician aversive, disorganizing, or irrelevant.

In addition to setting the stage for establishing a working relationship with the patient, the physician's management of the initial interview should promote collaboration with the patient's family member, case manager, or other caregiver. Both the patient and the caregiver can be told that the patient's confidentiality will be respected, but should the therapist at any time believe a relapse or other dangerous situation is developing, the help of family and other caregivers will be solicited. The family can be encouraged to contact the physician should they develop concerns and the frequency of future family contacts should be agreed upon. Family or caregiver support is crucial to the outcome of treatment and will most likely be extended to a clinician who is felt to be empathic, responsive to concerns, and available.

Therapists should be cognizant of the fact that suicide is the most common cause of premature death in schizophrenia and that patients with schizophrenia are less prone to spontaneously report suicidality. As a result, the therapist should be active in eliciting suicidal ideation; if not directly asked, patients may not reveal their intent. Several studies have shown that two-thirds or more of schizophrenia patients who commit suicide have seen an apparently unsuspecting clinician within 72 hours of death.

Research indicates that less-disabled patients with few negative symptoms who retain the capacity to experience painful affects and can reflect on the seriousness of their situation are at greatest risk of suicide. Periods following discharge from a hospitalization and loss of an important relationship have been identified as high-risk periods. Providing additional therapeutic support for patients during high-risk periods and monitoring for the emergence of potentially treatable syndromal depression may be reasonable strategies to reduce suicide risk.

Medication Compliance Noncompliance with effective psychopharmacological treatments during both short-term and maintenance therapy is a major cause of morbidity among patients with schizophrenia. When prolonged or repeated, noncompliance contributes to a downwardly spiraling cycle of relapse, recidivism, and deterioration of social and instrumental functioning. Empirical correlates of noncompliance include (1) patient-related factors (greater illness severity or grandiosity, lack of insight, substance abuse comorbidity); (2) medication-related factors (dysphoric medication adverse effects, ineffective or excessively high dosages); (3) environmental factors (inadequate support of supervision, practical barriers); and (4) clinician-related factors (poor therapeutic alliance). Available research underscores the multiplicity of explanations for reduced compliance and highlights the necessity of an individualized assessment. Of particular relevance to psychotherapy are patients' health beliefs and the psychological meanings attached to their illness and its treatment.

Among the psychological meanings associated with medication noncompliance, the following have been described: (1) pervasive denial about having an illness and needing treatment; (2) reactive efforts to regain control of one's life and maintain a sense of self-cohesion by organizing in opposition to the will of others; (3) the concrete equation of taking medication with being ill ("If I need drugs, I must be sick; The higher the dose, the sicker I am; I'll stop being ill if I stop taking drugs"); (4) lack of knowledge or incorrect beliefs about medications (taking drugs is a sign of weakness); (5) paranoid views of medication as being poisonous, controlling, or damaging; (6) secondary gain from psychosis—grandiose delusional gratification, escape from normal expectations and responsibilities; (7) pain and anguish accompanying symptom reduction with its attendant recognition that one has been ill and that the illness is severe; (8) displacement from transference—for example, discontinuing medication as an expression of anger towards one's therapist or family; and (9) an expression of unconscious ambivalence or fear of autonomy—as in discontinuing medication immediately before beginning a new job or rehabilitation program.

The variety of possible factors related to noncompliance point to the need for a broad and flexible approach that applies a range of interventions based on current knowledge of psychopharmacology and sound dynamic understanding of the individual patient. General recommendations for improving compliance in the context of psychotherapy have included: (1) conveying interest and concern about medication by asking specific questions about how much medication is being taken, effects, and adverse effects; (2) assuming many patients will at times take more or less medication and creating a therapeutic environment where such experiments are legitimized and can be talked about; (3) involving patients to the greatest extent possible in their own medication treatment—for example, allowing self-regulation of dosage within limits; (4) arranging for the taking of medication under the supervision of family, friends, or others and enlisting their support for medication; (5) direct praise and support for medication compliance; (6) education in the areas of medication adverse effects, relapse prevention, and the biological basis of major mental illness; (7) promoting self-monitoring through record keeping and other behavioral interventions; (8) attending to and building the therapeutic relationship as a lever to change; and (9) helping the patient experience activities that promote self-esteem and compete with psychosis as sources of gratification. The choice of specific interventions should be based on a differential diagnosis that generates hypotheses regarding which specific factors are operative in the individual patient. When lack of knowledge and cognitive deficit are a major factor in noncompliance, specific cognitive and behavioral procedures have been developed to enhance cognitive mastery and skills attainment. When noncompliance represents the unconscious wish to regress or to act out transference, dynamic exploration and interpretation are required. When severe disorganization is a major factor, arrangements for supervised medication administration may be necessary.

Finally, it must be recognized that some patients who appear to be clear candidates for benefiting from medication will continue to refuse this treatment, despite all efforts. These circumstances typically arouse countertransference reactions that require attention, such as telling the patient to seek treatment elsewhere out of anger or wishing to hurt the patient through abandonment. Another unproductive countertransference is withholding advice or support that might be of use to patients in order to see them learn their lesson by experiencing a full-blown relapse. Allowing the noncompliant patient who leaves treatment against medical advice to do so with dignity can set the stage for greater collaboration should the patient return in the future.

Circumstances requiring the therapist to initiate commitment or involuntary administration of medication also typically evoke powerful countertransference. Following such coercive interventions, the psychotherapist may experience considerable guilt about and fear of the patient. At such a time the therapist is tempted to discontinue all contact and turn the patient over to another clinician, rationalizing that a therapeutic alliance will never again be possible after such heavy-handed acts. Experience demonstrates that such an assessment is usually a distortion that more often serves the needs of the therapist than the patient. Following resolution of such an episode, many patients may express gratitude for the therapist's action, and in hindsight recognize that they were in need of treatment.

The results of empirical research have catalyzed a reformulation of psychotherapeutic techniques. An approach to individual therapy that includes appropriate attention to pharmacological, psychotherapeutic, psychosocial, and rehabilitative interventions titrated to the individual patient's response and progress has been articulated in clinical literature. The introduction of new antipsychotic agents with greater efficacy and fewer adverse effects will raise the functional ceiling previously imposed on many patients who experienced persistent symptoms or disabling adverse effects with first-generation agents. The greater efficacy of new medications creates new opportunities to realize the potential therapeutic synergy of blending multiple interventions into an integrated treatment approach.

SUGGESTED CROSS-REFERENCES

General discussions of the psychotherapies appear in Chapter 30. Various aspects of schizophrenia and its treatments are discussed in the other sections of Chapter 12.

SECTION REFERENCES

Alanen YO: *Schizophrenia: Its Origins and Need-Adapted Treatment*. Karnac Books, London, 1997.

American Psychiatric Association: Practice guidelines for the treatment of patients with schizophrenia. Am J Psychiatry *154*(Suppl):1, 1997.

Book HE: Some psychodynamics of non-compliance. Can J Psychiatry *32:*115, 1987.

Brill AA: Schizophrenia and psychotherapy. Am J Psychiatry *9:*519, 1929.

Corrigan PW, Liberman RP, Engel JD: From noncompliance to collaboration in the treatment of schizophrenia. Hosp Community Psychiatry *41:*1203, 1990.

Coursey RD: Psychotherapy with persons suffering from schizophrenia: The need for a new agenda. Schizophr Bull *15:*349, 1989.

Diamond RJ: Enhancing medication use in schizophrenic patients. J Clin Psychiatry *44:*8, 1983.

*Dingman CW, McGlashan TH: Psychotherapy. In *A Clinical Guide for the Treatment of Schizophrenia*, AS Bellack, editor. Plenum, New York, 1989.

Duckworth K, Nair V, Patel JK, Goldfinger SM: Lost time, found hope and sorrow: The search for self, connection, and purpose during “awakenings” on the new antipsychotics. Harv Rev Psychiatry *5:*277, 1997.

Eckman TA, Liberman RP, Phipps CC, Blair KE: Teaching medication management skills to schizophrenic patients. J Clin Psychopharmacol *10:*33, 1990.

Fenton WS, McGlashan TH: We can talk: Individual psychotherapy for schizophrenia. Am J Psychiatry *154:*1493, 1997.

Fenton WS, Blyler CR, Heinssen RK: Determinants of medication compliance in schizophrenia: Clinical and empirical correlates. Schizophr Bull *23:*637, 1997.

Frank AF, Gunderson JG: The role of the therapeutic alliance in the treatment of schizophrenia: Relationship to course and outcome. Arch Gen Psychiatry *47:*228, 1990.

*Fromm-Reichmann F: *Principles of Intensive Psychotherapy*. University of Chicago Press, Chicago, 1950.

Gilbert S, Ugelstad E: Patient's own contributions to long-term supportive psychotherapy in schizophrenia disorders. Br J Psychiatry *164* (Suppl):84, 1994.

Glass LL, Katz HM, Schnitzer RD, Knapp PH, Frank AF, Gunderson JG: Psychotherapy of schizophrenia: An empirical investigation of the relationship of process to outcome. Am J Psychiatry *146:*603, 1989.

*Greenfeld D: *The Psychotic Patient: Medication and Psychotherapy*. The Free Press, New York, 1985.

Gunderson JG, Frank AF, Katz HM, Vannicelli ML, Frosch JP, Knapp PH: Effects of psychotherapy in schizophrenia, II. Comparative outcome of two forms of treatment. Schizophr Bull *10:*564, 1984.

Hogarty GE, Kornblith SJ, Greenwald D, DiBarry AL, Cooley S, Flesher S, Reiss D, Carter M, Ulrich R: Personal therapy: A disorder-relevant psychotherapy for schizophrenia. Schizophr Bull *21:*379, 1995.

Hogarty GE, Kornblith SJ, Greenwald D, DiBarry AL, Cooley S, Ulrich RF, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, I. Description of study and effects on relapse rates. Am J Psychiatry *154:*1504, 1997.

Hogarty GE, Greenwald D, Ulrich RF, Kornblith SJ, DiBarry AL, Cooley S, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, II. Effects on adjustment of patients. Am J Psychiatry *154:*1514, 1997.

Kane JM, McGlashan TH: Treatment of schizophrenia. Lancet *346:*820, 1995.

Kemp R, Kirov B, Everitt B, Hayward P, David A: Randomised controlled trial of compliance therapy. 18-month follow-up. Br J Psychiatry *172:*413, 1998.

*Kendler KS: Long-term care of an individual with schizophrenia: Pharmacologic, psychological, and social factors. Am J Psychiatry *156:*124, 1999.

Kuipers E, Garety P, Fowler D, Dunn G, Bebbington P, Freeman D, Hadley C: London-East Anglica randomised controlled trial of cognitive-behavioral therapy for psychosis. Br J Psychiatry *171:*319, 1997.

Levine IL, Wilson A: Dynamic interpersonal processes and the inpatient holding environment. Psychiatry *48:*341, 1985.

Liberman RP: Psychosocial treatments for schizophrenia. Psychiatry *157:*104, 1994.

McGlashan TH, Keats CJ: *Schizophrenia: Treatment Process and Outcome*. American Psychiatric Press, Washington, DC, 1989.

McGlashan TH: Schizophrenia: Psychosocial treatments and the role of psychosocial factors in its etiology and pathogenesis. In *Annual Review of Psychiatry*, vol 5, A Frances, R Hales, editors. American Psychiatric Press, Washington, DC, 1986.

McGorry PD, Edwards J, Mihalopoulos C, Harrigan SM, Jackson HJ: EPPIC: An evolving system of early detection and optimal management. Schizophr Bull *22:*305, 1996.

*Meehl PE: Schizotaxia, schizotypy, schizophrenia. Am Psychol *17:*827, 1962.

Nuechterlein KH, Dawson ME: Vulnerability and stress factors in the developmental course of schizophrenic disorders. Schizophr Bull *10:*158, 1984.

Parsons T: *The Social System*. Free Press, Glencoe, IL, 1951.

Rako S, Mazer H, editors: *Semrad: The Heart of a Therapist*. Jason Aronson, New York, 1980.

A Recovering Patient: “Can we talk?” The schizophrenic patient in psychotherapy. Am J Psychiatry *143:*68, 1986.

Scott JE, Dixon LB: Psychological interventions for schizophrenia. Schizophr Bull *21:*621, 1995.

Searles HF: *Countertransference*. International Universities Press, New York, 1979.

Stanton AH, Gunderson JG, Knapp PH, Frank AF, Vannicelli ML, Schnitzer R, Rosenthal R: Effects of psychotherapy in schizophrenia, I. Design and implementation of a controlled study. Schizophr Bull *10:*520, 1984.

Strauss JS, Hafez H, Lieberman P, Harding CM: The course of psychiatric disorder, III. Longitudinal principles. Am J Psychiatry *142:*289, 1985.

*Sullivan HS: *The Psychiatric Interview*. Norton, New York, 1970.

Tarrier N, Yusupoff L, Kinney C, McCarthy E, Gledhill A, Haddlock G, Morris J: Randomised controlled trial of intensive cognitive behaviour therapy for patients with chronic schizophrenia. Br Med J *317:*303, 1998.

Weiden P, Havens L: Psychotherapeutic management techniques in the treatment of outpatients with schizophrenia. Hosp Community Psychiatry *45:*549, 1994.

CHAPTER 13. OTHER PSYCHOTIC DISORDERS

13.1 SCHIZOAFFECTIVE DISORDER, SCHIZOPHRENIFORM DISORDER, AND BRIEF PSYCHOTIC DISORDER

JOHN LAURIELLO, M.D., BRENDA R. ERICKSON, M.D., AND SAMUEL J. KEITH, M.D.

There are three disorders in addition to schizophrenia listed in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) in the section "Schizophrenia and Other Psychotic Disorders." The first, schizoaffective disorder, is a complex illness that has changed significantly over time. In its simplest definition, it is presently conceived as an illness with coexisting, but independent, schizophrenic (psychotic) and mood components. Schizoaffective disorder is seen primarily as part of a schizophrenia spectrum rather than an equal hybrid of mood and schizophrenia disorders. Schizophreniform disorder is a diagnosis that assumes another will replace it after 6 months. Most cases of schizophreniform disorder progress to either schizophrenia or schizoaffective disorder, with some cases rediagnosed as a non–schizophrenia spectrum illness (i.e., schizotypal or schizoid personality disorders), while a few resolve completely. Finally, the diagnosis brief psychotic disorder describes an impairment in reality testing that lasts at least 1 day, but less than 1 month. All three disorders have a psychotic component, are often misunderstood, are incorrectly applied, and are not as well studied as schizophrenia, bipolar I disorder, or major depressive disorder.

SCHIZOAFFECTIVE DISORDER

As the end of the century nears great strides have been made in clarifying the diagnostic criteria for many psychiatric illnesses. However, patients often do not fall neatly into set illness criteria. There are several approaches to dealing with such patients. One is to diagnose the patient with two distinct illnesses and treat those illnesses as separate problems. Another possibility is to consider that the patient has a primary illness and symptoms of a second illness that are not as important and might even resolve when the primary illness is treated. A third approach considers that the patient suffers from a distinct blended illness with its own history, diagnosis, and treatment. This last approach best represents the current orthodoxy in the diagnosis and treatment of patients with the DSM-IV diagnosis of schizoaffective disorder. Unfortunately, this approach is not easily applied, often making the diagnosis confusing and convoluted.

History At the beginning of this century, patients with mental illness were grouped together as suffering from the common illness insanity. With the work of Emil Kraepelin, and Eugene Bleuler, distinct diagnostic groups began to emerge. Kraepelin was able to distinguish an unremitting, dementing illness in young patients that became known as schizophrenia, which he contrasted with an episodic illness of affect now known as bipolar I disorder. However, there were patients who did not fit neatly into either category. Bleuler believed that the presence of any symptoms of schizophrenia even when there was an affective component was still schizophrenia. Patients with mixed features of schizophrenia and affective (mood) disorder were first described by George Kirby and August Hoch in the early part of the century. In 1933, Jacob Kasanin introduced the term "schizoaffective psychosis" to describe a group of patients who had symptoms of both affective and schizophrenic illnesses. While he is credited with introducing the term, on subsequent review of these patients, all would now meet the diagnosis of a pure mood disorder. Nevertheless, the term schizoaffective disorder has survived albeit in several different contexts.

Comparative Nosology One of the difficulties in using a diagnosis that depends on not being another diagnosis is that both depend on changes in the other. Schizoaffective disorder is affected by any changes in the diagnostic criteria of schizophrenia, affective disorder, or both. As psychotic affective disorders and schizophrenia have been better distinguished, those who fall through the "diagnostic cracks" have become clearer. In the second edition of DSM (DSM-II) schizoaffective disorder was a subtype of schizophrenia and denoted patients who had any mood symptoms while meeting the criteria for schizophrenia. In contrast, the Research Diagnostic Criteria (RDC) for schizoaffective disorder allowed as few as one symptom of schizophrenia in a patient who met the criteria for a full affective disorder. The third edition of DSM (DSM-III), influenced by studies in the United States and Great Britain, narrowed the diagnosis of schizophrenia and expanded the diagnosis of bipolar disorder. It allowed symptoms of schizophrenia to coexist with a mood disorder as long as these schizophrenic symptoms did not remain when the mood disorder resolved. Moreover, mood-incongruent psychotic symptoms could now exist in bipolar disorder. Finally, schizoaffective disorder moved from its schizophrenia subtype place to stand alone as a "psychotic disorder not elsewhere classified." The revised third edition of DSM (DSM-III-R) expanded this notion by inserting the criterion that a patient with schizoaffective disorder must meet the criteria for schizophrenia for at least 2 weeks independent of any mood syndrome.

DSM-IV has retained most of the DSM-III-R criteria but has stricter diagnostic criteria for schizophrenia. Patients must meet the symptoms of schizophrenia for at least 1 month as opposed to the previous 1-week criterion. Schizoaffective disorder is now listed in the section "Schizophrenia and Other Psychotic Disorders." The 10th revision of *International Statistical Classification of Diseases and Related Problems* (ICD-10) essentially describes the same disorder. The ICD-10 schizoaffective disorders describe single as well as recurrent episodes. Subtypes include manic, depressed, and mixed types. Mixed type includes a cyclic schizophrenia and a mixed schizophrenic-mood psychosis.

Epidemiology There is no psychiatric epidemiological study of the incidence or prevalence of schizoaffective disorder in a general population. Even if there were such studies, older reports might not be useful, because the diagnosis (and therefore the incidence and prevalence) would have changed over time. Prevalence rates for consecutive patients diagnosed in a psychiatric treatment setting are available. These numbers range from 2 to 29 percent, a potentially significant cohort requiring treatment. Several lines of evidence support the idea that one might expect an increased prevalence of schizoaffective disorder in women. Women have a higher prevalence of major depressive disorder than men do, and women with schizophrenia express more affective symptoms than men with schizophrenia do. In family studies of patients with schizoaffective disorder, relatives of females with schizoaffective disorder have a higher rate of schizophrenia and depressive disorders than do relatives of males with schizoaffective disorder.

Etiology It is difficult to determine a cause of a disease that has changed so much over time. One might conjecture that the etiology of schizoaffective disorder as currently defined might be similar to the etiology of schizophrenia. Thus etiological theories of schizoaffective disorder would include some genetic and environmental causation. Molecular genetic studies of schizoaffective disorder have lagged behind recent studies of the genetics of schizophrenia and bipolar I disorder. Available family studies have reported that families of schizoaffective probands have significantly higher rates of relatives with mood disorder than families of schizophrenia probands. Similarly, these schizoaffective probands have more psychotic symptoms than families of mood disorder probands. The results of these family studies have argued that schizoaffective disorder is a unique disorder, separate from schizophrenia and mood disorders.

Possible environmental causes of schizoaffective disorder are similar to those of schizophrenia, including in utero insult (including malnutrition and viral causes) and obstetrical complications. One hypothesis considers schizophrenia to be a developmental and progressing disorder that can be seen in the development of brain dysmorphology. This includes less cortical gray matter and more fluid and fluid-filled spaces; however, no definitive study of patients with DSM-IV schizoaffective disorder has been done. One might assume that schizoaffective patients would have similar brain abnormalities, because the disorder mimics many aspects of schizophrenia.

For nearly a half century the prevailing etiologic theory of schizophrenia was the dopamine hypothesis. In its simplest description it postulates that the underlying abnormality is excess dopamine in areas of the brain, leading to psychosis. Thus, successful treatment with antipsychotics is due to their dopamine-blocking properties. With the successful use of clozapine (Clozaril) and other serotonin-dopamine antagonists, the dopamine hypothesis has been amended. Currently, a critical balance between the neurotransmitters dopamine and serotonin is believed to be important for treating schizophrenia. At the same time it is accepted that there are abnormalities of serotonin and norepinephrine in mood disorders. These theories are particularly interesting when considering underlying causes of schizoaffective disorder. Possibly this balance of dopamine and serotonin is particularly affected in schizoaffective disorder, leading to chronic psychosis and intermittent but substantial mood alterations.

Diagnostic and Clinical Features DSM-IV diagnostic criteria are provided in Table 13.1-1.

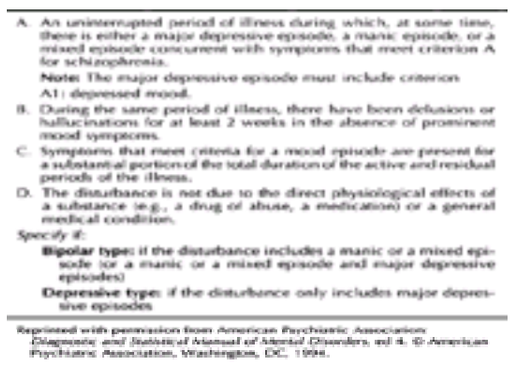

A. An uninterrupted period of illness during which, at some time, there is either a major depressive episode, a manic episode, or a mixed episode concurrent with symptoms that meet criterion A for schizophrenia.
   Note: The major depressive episode must include criterion A1: depressed mood.
B. During the same period of illness, there have been delusions or hallucinations for at least 2 weeks in the absence of prominent mood symptoms.
C. Symptoms that meet criteria for a mood episode are present for a substantial portion of the total duration of the active and residual periods of the illness.
D. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify if:

Bipolar type: if the disturbance includes a manic or a mixed episode (or a manic or a mixed episode and major depressive episodes)

Depressive type: if the disturbance only includes major depressive episodes

Table 13.1-1 DSM-IV Diagnostic Criteria for Schizoaffective Disorder

These criteria are a product of several revisions that have sought to clarify several diagnoses including schizophrenia, bipolar disorder, and major depressive disorder. It was hoped that improving these diagnoses would make schizoaffective disorder begin to stand out apart from them. However, the diagnostic criteria still leave much to interpretation. The diagnostician must accurately diagnose the affective illness, making sure it meets the criteria of either a manic or depressive episode but also determining the exact length of each episode (not always an easy or possible task). The length of each episode is critical for two reasons. First, to meet the B criterion (psychotic symptoms in the absence of the mood syndrome) one has to know when the affective episode ends and the psychosis continues. Second, to meet criterion C the length of all mood episodes must be combined and compared with the total length of the illness. If the mood component is present for a substantial portion of the total illness, then that criterion is met. Calculating the total length of the episodes can be difficult, and it does not help that the term “substantial portion” is not defined. In practice, most clinicians look for the mood component to be 15 to 20 percent of the total illness. Patients who have one full manic episode lasting 2 months but who have suffered from symptoms of schizophrenia for 10 years do not meet the criteria for schizoaffective disorder. Instead, the diagnosis would be a mood episode superimposed on schizophrenia. It is unclear whether the bipolar or depressive type specifiers are helpful, although they may direct treatment options. These subtypes are often confused with earlier subtypes (schizophrenic versus affective type) thought to have implications in course and prognosis. As with most psychiatric diagnoses, schizoaffective disorder should not be used if the symptoms are caused by substance abuse or a secondary medical condition.

The ICD-10 diagnostic criteria for schizoaffective disorder are listed in Table 13.1-2.

Table 13.1-2 ICD-10 Diagnostic Criteria for Schizoaffective Disorders

Ms. A.D. was a 29-year-old white unmarried woman, with a 10-year history of schizoaffective disorder bipolar type. She was first hospitalized after child protection took her son away for alleged child abuse. When the patient was interviewed at that time, she was described as dressed like a “gypsy” with heavy makeup and pressured speech. She told the treatment team her son had been abused by his father, a well-known rock star. During this time she was stabilized on lithium (Eskalith) and haloperidol (Haldol). A.D.'s manic symptoms resolved, but her belief that she was a rock star's girlfriend remained. Since that first hospitalization she has lost custody of her son. She remains delusional about the child's famous father, and in addition, she believes people are out to get her. She has had three distinct episodes of mania during which she needs little sleep and has racing thoughts and pressured speech. She has been intermittently compliant with medications and is currently receiving haloperidol in a long-acting form. In the 10 years of her illness she has never been free of her delusions. She has not been able to work and receives federal disability assistance.

Differential Diagnosis The psychiatric differential diagnosis includes all the possibilities usually considered for mood disorders and for schizophrenia

In any differential diagnosis of psychotic disorders a complete medical workup should be performed to rule out organic causes of the symptoms. A history of substance use with or without a positive toxicology screening test may indicate a substance-induced disorder. Preexisting medical conditions, their treatment, or both may cause psychotic and mood disorders. Any suspicion of a neurological abnormality warrants consideration of a brain scan to rule out anatomical pathology and an electroencephalogram (EEG) to determine any possible seizure disorders (e.g., temporal lobe epilepsy). Psychotic disorder due to seizure disorder is more common than that seen in the general population. It tends to be characterized by paranoia, hallucinations, and ideas of reference. Epileptic patients with psychosis are believed to have a better level of function than patients with schizophrenic spectrum disorders. Better control of the seizures can reduce the psychosis.

The same hypothetical case is used in the 3 cases below with different outcomes to illustrate the diagnostic decision.

Mrs. B. was a 32-year-old married woman with three children. She reported being relatively happy and free of illness until the birth of her third child. She had the usual “baby blues” that resolved after the first month. When her third child was 14 months old, she began to have trouble sleeping, and her husband noticed that she was sometimes irritable and at other times euphoric. She began to talk rapidly and call family members at all hours of the night. One night her husband received a phone call that his wife was in the county jail. She had secretly left the house, gone to a local bar, and instigated a fight with a female patron. The police thought she was acting wildly and suspected some sort of intoxicant. She was taken to the local psychiatric clinic where a urine toxicology screen was negative. She was admitted to the hospital and treated with the benzodiazepine, lorazepam (Ativan), and the mood stabilizer lithium, and after 2 weeks was completely asymptomatic.

Her diagnosis is bipolar I disorder, manic type.

Discussion The patient suffered from an elevated and euphoric mood alone. She did not exhibit any symptoms of schizophrenia and was appropriately treated with a benzodiazepine, lorazepam, to calm her and long-term treatment with a mood stabilizer, lithium. This patient might have future episodes with hallucinations, delusions, or both. These psychotic symptoms may or may not be congruent with her mood state (e.g., a patient who is depressed and has the delusion of being a terrible person who has committed a crime and deserves to suffer and be punished). However, the psychotic symptom might also be very incongruent with the mood. The critical distinction for this patient was that the psychotic symptom existed only during the mood episode. Conceptually, the psychosis was fueled by the mood. Correct the mood and there is no fuel for the psychosis and it also disappears.

Mrs. S. was a 32-year-old married woman with three children. She reports that she has been relatively happy and free of illness until the birth of her third child. She had the usual “baby blues” that resolved after the first month. When her third child was 14 months old she began to have trouble sleeping and her husband noticed that she was becoming increasingly isolated and not able to take care of her children. One night her husband received a phone call that his wife was in the county jail. She had secretly left the house, gone to a local bar and instigated a fight with a female patron. The police thought she was acting wildly and suspected some sort of intoxicant. She was taken to the local psychiatric clinic where a urine toxicology screen was negative. At that time she told the staff that she was sure someone was using her social security number and consuming the benefits she would need when she was older. She had gone to the bar because a man's voice had told her that the person who was using her benefits was there. This voice had been talking to her for over a year and often commented on her looks and actions. The patient was admitted to the hospital, treated with the antipsychotic risperidone (Risperdal), and after 2 weeks of treatment was completely asymptomatic.

Her diagnosis is schizophrenia, paranoid type.

Discussion The patient's primary symptoms were delusions and hallucinations without any accompanying mood abnormality. They were of sufficient severity and duration to give her a diagnosis of the paranoid type of schizophrenia, and she was appropriately treated with an antipsychotic agent. Patients suffering from schizophrenia often have both depressive and euphoric symptoms. A common mistake is assuming that a schizophrenic patient presenting with a full range of affect is a patient with schizoaffective disorder. The presence of euphoria or demoralization alone does not meet the criteria for diagnosis of schizoaffective disorder. Patients must both meet the appropriate criteria for the affective disorder and have the affective disorder for a substantial portion of their chronic illness. That said, a patient with schizophrenia suffering from subsyndromal demoralization or disinhibited behavior might benefit from an antidepressant or mood stabilizer, respectively.

Mrs. S.A. was a 32-year-old married woman with three children. She reports that she has been relatively happy and free of illness until the birth of her third child. She had the usual “baby blues” that resolved after the first month. When her third child was 14 months old she began to have trouble sleeping and her husband noticed that she was becoming increasingly irritable, euphoric, isolated and not able to take care of her children. One night her husband received a phone call that his wife was in the county jail. She had secretly left the house, gone to a local bar and instigated a fight with a female patron. The police thought she was acting wildly and suspected some sort of intoxicant. She was taken to the local psychiatric clinic where a urine toxicology screen was negative. At that time she told the staff that she was sure there was someone using her social security number and consuming the benefits she would need when she was older. She also described herself as being one of the 10 smartest people in the world and was sure that the treatment team did not understand her because of their incompetence and she asked to be seen by the head of the hospital. The patient was treated with the antipsychotic risperidone and the mood stabilizer lithium and was completely asymptomatic after 2 weeks of treatment. A year later her husband brought her back to the psychiatric hospital. He reported she had been doing well and was compliant with her medication, which was now lithium carbonate alone. Her mood has been unremarkable, but in the last month she again began to say that someone had stolen her social security benefits. On interview she was calm and cooperative although a little guarded. She reluctantly admitted that the man's voice had returned recently. Risperidone was added back to her regimen, and after 2 weeks she returned to her usual self.

Her diagnosis is schizoaffective disorder, bipolar type.

Discussion The third case (above) displayed symptoms of both a mania and a delusion. She was appropriately treated with an antipsychotic agent and a mood stabilizer. If the vignette had ended there, one might conclude that she had a manic episode with psychotic features. However, she had an exacerbation a year later. This time her mood was totally normal but her delusions and hallucinations returned. She is restarted on the antipsychotic since it appears the mood stabilizer alone was insufficient. It was very appropriate in this circumstance to first taper the patient off the antipsychotic and try the patient on a mood stabilizer alone, aware of the long-term risks of antipsychotics. However, having had the delusional episode while on the mood stabilizer most likely portends a need for intermittent or maintenance antipsychotic treatment.

These 3 women, while having historical details in common, illustrate the differences between a pure mood disorder, schizophrenia, and schizoaffective disorder.

Course and Prognosis Considering the uncertainty and evolving diagnosis of schizoaffective disorder, determining the long-term course and prognosis is difficult. Given the definition of the diagnosis, one might expect patients with schizoaffective disorder to have either a course similar to an episodic mood disorder, a chronic schizophrenic course, or some intermediate outcome. It has been presumed that an increasing presence of schizophrenic symptoms predicted worse prognosis. Studies using RDC criteria showed that after 1 year patients with schizoaffective disorder had different outcomes depending on whether their predominant symptoms were affective (better prognosis) or schizophrenic (worse prognosis). With the narrower definition of DSM-III-R and DSM-IV, all patients had to have an independent schizophrenic component to meet the diagnosis of schizoaffective disorder. One study that followed patients diagnosed with DSM-III-R schizoaffective disorder for 8 years found that the outcomes of these patients more closely resembled schizophrenia than a mood disorder with psychotic features.

Treatment There are several extensive reviews of the treatment of schizoaffective disorder, but critical evaluation of the results of these studies is not easy. Because the operational definition of schizoaffective disorder has shifted over the last 30 years, comparing or pooling studies is impossible. The efficacy and selection of treatment for a patient under the broader (more mood disorder inclusive) DSM-II criteria may differ from that of the patient diagnosed with the narrower DSM-III-R criteria. However, there are some general recommendations for treatment. The principle rule is to treat the patient's symptoms, not the diagnostic label.

Mood Stabilizers Mood stabilizers are a mainstay of treatment for bipolar disorders and would be expected to be important in the treatment of patients with schizoaffective disorder. Few studies have examined the efficacy of mood stabilizers in schizoaffective disorder, in contrast to the extensive studies of lithium, valproate (Depakote), and to a lesser extent carbamazepine (Tegretol) in bipolar I disorder. A recent study that compared lithium with carbamazepine showed superiority for carbamazepine for schizoaffective disorder, depressive type, but no difference in the two agents for the bipolar type. In practice however, these medications are used extensively alone, in combination with each other, or with an antipsychotic agent. In manic episodes, schizoaffective patients should be treated aggresively with dosages of a mood stabilizer in the middle to high therapeutic blood concentration range. As the patient enters a maintenance phase the dosage can be reduced to low to middle range to avoid adverse effects and potential effects on organ systems (e.g., thyroid and kidney) and to improve ease of use and compliance. Laboratory monitoring of plasma drug concentrations and periodic screening of thyroid, kidney, and hematological functioning should be performed. As in all cases of intractable mania, the use of electroconvulsive therapy (ECT) should be considered.

Psychosis or akathisia must be distinguished from a manic episode. For a psychotic agitation, an antipsychotic agent (often with a benzodiazepine) is indicated. In akathisia, numerous studies have shown that reducing the antipsychotic agent dosage or using benzodiazepine or a b-adrenergic receptor antagonist are helpful.

Antidepressants By definition many schizoaffective patients suffer from major depressive episodes. Treatment with antidepressants mirrors treatment of bipolar depression. Care should be taken not to precipitate a cycle of rapid switches from depression to mania with the antidepressant. The choice of antidepressant should take into account previous antidepressant successes or failures. Selective serotonin reuptake inhibitors (e.g., fluoxetine [Prozac] and sertraline [Zoloft]) are often used as first-line agents because they have less effect on cardiac status and have a favorable overdose profile. However, agitated or insomniac patients may benefit from a tricyclic antidepressant. As in all cases of depression, use of ECT should be considered.

It is very important to try to distinguish psychosis, akinetic syndromes, and primary negative symptoms from depression. Again, psychosis should be adequately treated with an antipsychotic agent. Suspected akinetic treatment can be improved by lowering the dosage of antipsychotic agent, treating with an anticholinergic agent, or switching to a serotonin-dopamine antagonist like clozapine (Clozaril), risperidone, olanzapine (Zyprexa), or quetiapine (Seroquel). Negative symptoms are often difficult to tease out. While there are no definitive studies, these symptoms may improve with the use of serotonin-dopamine antagonists.

Antipsychotic Agents As mentioned above, antipsychotic agents are important in the treatment of schizoaffective disorder. The introduction of chlorpromazine in the 1950s showed antipsychotic agents that block the action of the neurotransmitter dopamine are effective in the treatment of psychosis. Therefore it is not surprising that these compounds are effective in treating the psychotic symptoms that plague patients with schizoaffective disorder. What is not as clear is whether they alone can control both the schizophrenic and the affective symptoms. Because of the complexity and mixed nature of the illness pharmaceutical companies have generally avoided separate studies of antipsychotic agents, with schizoaffective patients. As medications were approved for use in schizophrenia, they were almost immediately used for patients with schizoaffective disorder. With the advent of combined serotonin-dopamine blocking agents, more schizoaffective patients are being recruited for efficacy and safety trials of new antipsychotic agents.

The introduction of the serotonin-dopamine antagonist holds promise for patients with schizophrenia and schizoaffective disorder. While much more data exist on clozapine efficacy in patients with schizophrenia, a few studies have described either comparable or greater efficacy with schizoaffective patients. There may be

several reasons why patients diagnosed with schizoaffective disorder respond favorably to clozapine. First, clozapine has been shown to be superior in treating positive symptoms of schizophrenia and therefore the positive symptoms of schizoaffective disorder. Second, akinetic syndromes that can mimic depressive syndromes are greatly reduced with clozapine use. Third, some evidence indicates that clozapine may have mood-stabilizing properties. This hypothesis is based on several studies showing good results using clozapine in patients with difficult cases of bipolar disorder. Data for risperidone, olanzapine, and quetiapine in treating schizoaffective disorder are minimal. One such study showed superiority of haloperidol and amitriptyline over risperidone in a group of psychotic patients (including schizoaffective disorder patients) with depressive symptoms.

Psychosocial Treatment Considering the present notion that schizoaffective disorder as specified in DSM-IV closely resembles schizophrenia, one can assume that psychosocial treatment of schizoaffective disorder should mimic that of psychosocial treatment of schizophrenia. Therefore, patients should benefit from a combination of family therapy, social skills training, and cognitive rehabilitation. Because the psychiatric field has had difficulty deciding the exact diagnosis and prognosis of schizoaffective disorder, this uncertainty must be explained to the patient. Historically, patients and families have been told that schizoaffective illness has a better prognosis than schizophrenia, but this may no longer be true. Patients and their families must contend with an evolving diagnosis; they may be told the patient is suffering from a treatable mood disorder at first and later told that it is a severe psychotic disorder. The range of symptoms can be quite large as patients contend with both ongoing psychosis and varying mood states. It can be very difficult for family members to keep up with the changing nature and needs of these patients. Medication regimens can be more complicated, with multiple medications frequent, and psychopharmacological education is important. However it is often difficult to explain to patients and their families that new medication treatments have been tested in affective and schizophrenic disorders but not schizoaffective disorder.

SCHIZOPHRENIFORM DISORDER

History Gabriel Langfeldt, first used the term *schizophreniform* in 1939, at the University Psychiatric Clinic in Oslo, Norway. As originally used, this diagnosis relied on a tradition of Scandinavian psychiatry, which had identified a condition that had relatively brief and self-contained psychotic intervals. Patients recovered well and had affective and sometimes hysterical components to their illness, and the diagnosis was used to distinguish a group considered to have little relation to true schizophrenia.

Comparative Nosology In contrast to this rather specific role for schizophreniform disorder, the current DSM-IV diagnosis has relatively little to do with the origin of the term and much more to do with the tradition of Kraepelinian schizophrenia as a chronic illness. Prior to the DSM-IV revision of this diagnostic entity, DSM-III and DSM-III-R had used this diagnosis as a “schizophrenia-in-waiting” diagnosis, with the only difference between the two diagnoses being whether the illness had lasted a total of 6 months including psychotic, prodromal, and residual symptoms. Under the DSM-III and DSM-III-R systems, the psychotic phase of the illness needed to last only 1 week, and less if treated successfully. The remainder of the 6-month-duration criteria for schizophrenia comprised residual or prodromal symptoms. Patients who had an insidious onset with prodromal symptoms preceding the onset of psychotic symptoms by at least 6 months would be given a diagnosis of schizophrenia as soon as the psychotic symptoms lasted 1 week. Those who had limited prodromal symptoms or who had sudden onset of psychosis as the first sign of illness, however, would not be diagnosed as having schizophrenia until the total period of illness reached 6 months. During what for many was a waiting period, the diagnosis of schizophreniform disorder would be used. Because of the relatively brief period of psychosis required (1 week) on the one hand and the similarity with schizophrenia on the other, this category formerly consisted of patients with potentially many types of psychoses—brief reactive psychosis, “schizophrenia-in-waiting,” and true schizophreniform disorder. Unfortunately, the true schizophreniform disorder would be difficult to sort out from this diagnostic system, and relative to the other categories it is probably quite rare, although potentially important as a time-limited psychotic illness that returns to baseline functioning without residual symptoms.

The revisions of DSM-IV have made one of the above overlaps less likely—the one with brief reactive psychosis. To separate these two disorders diagnostically, the DSM-IV diagnosis for schizophreniform disorder requires a month of psychotic symptoms rather than 1 week. Further, brief reactive psychosis has changed to brief psychotic disorder because the diagnostic criteria reaction-to-a-stressor was considered too ubiquitous—but DSM-IV includes the concept as a specifier. From the other side, the diagnosis of schizophreniform has moved much closer to its parent diagnosis of schizophrenia with the requirement for 1 month of psychotic symptoms. Although no data are currently available on the course of schizophreniform illness, the requirement for a greater duration of psychotic illness will probably make it less likely that a given patient will recover before 6 months of total illness comprising both psychotic symptoms and residual or prodromal symptoms (now referred to as attenuated symptoms) is reached.

This category now looks exactly like schizophrenia with an unanticipated full recovery before 6 months. Some data suggest that those who indeed do recover before 6 months have better 5- and 10-year outcomes. Whether this represents a separate disorder category or merely one end of a distribution of outcomes in schizophrenia is yet to be determined. There will always be the unusual patient who appears to have schizophrenia but recovers completely. They are exceedingly rare. Further, this category of illness continues to be severely hampered by a lack of research, and indeed the changing criteria for diagnosis makes it difficult to focus on this “moving target.” Most data will continue to be anecdotal. ICD-10 does not have a designated schizophreniform disorder, although the concept is included in several categories. The diagnosis acute schizophrenia-like psychotic disorder describes a disorder that would otherwise be considered schizophrenia but with symptoms lasting less than 1 month. If the symptoms persist past the month, the ICD-10 diagnosis of schizophrenia should be used. There is also a subclassification for a schizophreniform psychosis manic or depressed type under “schizoaffective disorders”; however, according to DSM-IV, schizophreniform disorder is subsumed order ICD-10's category of other schizophrenia.

Another reason for having this diagnostic category is that it avoids having to use the term *schizophrenia* with all of its negative connotations early in the diagnostic formulation. Many families require considerable time to reconcile the future of their family member. A gradual introduction to the concept of schizophreniform disorder, with a waiting period during which the family can more realistically orient itself and learn about the illnesses in the schizophrenia spectrum may prove helpful to some. Further, because of the negative connotation of schizophrenia and the stigma currently attached to it, a diagnostic system that avoids a false-positive diagnosis of schizophrenia is desirable. A 6-month duration of illness prior to making the diagnosis of schizophrenia will eliminate virtually all false-positive diagnoses.

As noted above, schizophreniform disorder shares an overlap with schizophrenia with two exceptions: the duration of illness is from 1 to 6 months and social or occupational dysfunction is not required to meet the diagnosis, although it may occur at some point in the illness. Given the requirement of 1 month of psychotic symptoms, however, it seems quite unlikely that a person's social and occupational functioning would not be disrupted. DSM-IV describes two possible conditions for this diagnosis: (1) when a person has recovered within the 6-month period (the “pure” form of schizophreniform disorder) and (2) when a person has not had the illness long enough (6 months) to meet the diagnosis of schizophrenia. For this latter condition, the term “provisional” is used. A guide for clinicians is given as a part of the diagnosis, which should be qualified by the presence or absence of good prognostic signs. The following are listed, and two are required for the qualifier of good prognosis: (1) rapid onset of psychotic symptoms, (2) confusion at the peak of psychotic symptomatology, (3) good premorbid social and occupational functioning, and (4) maintenance of a range of affect.

As with most psychiatric diagnoses, schizophreniform disorder should not be used if substance abuse or a secondary medical condition causes the symptoms.

Epidemiology Because of the significant change in the diagnostic criteria for schizophreniform disorder in DSM-IV, there are currently no epidemiological data from community samples. The risks of drawing from treatment samples are well known in terms of the variability introduced by clinic type, socioeducational variables, urban/rural factors, and even treatment philosophy. The elegance of the Epidemiologic Catchment Area (ECA) study with its five sites, using census track data collection, is not likely to be repeated in the near future, and its data were derived using DSM-III criteria. The significant changes from that period would include changing from 1 week of psychotic symptoms to 1 month, adding the concept of “provisional” diagnosis, and adding the good prognostic signs. Clearly, lengthening the requirement for psychotic symptoms was the most significant change, because it eliminated what most certainly were many cases of brief psychotic disorder. The data from the ECA study indicate a lifetime prevalence of 0.2 percent and a 1-year prevalence of 0.1 percent. Even with inclusion of cases of brief psychosis, this is a relatively small category. One could extrapolate that with the further stringency of DSM-IV, the category would become even smaller.

Etiology Because of the change in the duration of illness, most persons who fall in this category will have underlying pathologies similar to those with schizophrenia. This will certainly be true for those who carry this diagnosis provisionally while waiting for the 6-month time period to elapse before changing the diagnosis to schizophrenia. There has been ample speculation about whether “acute” schizophrenia (rapid onset, good premorbid functioning) differs from insidious-onset schizophrenia in anything more than severity of such factors as negative symptoms. A rapid and complete response to a treatment intervention may eventually help to differentiate those in this category from standard antipsychotic nonresponders. The concept that the heterogeneity of the underlying biology may be responsible for differential treatment response is not new, but it has been given increasing credibility with the advent of the serotonin-dopamine antagonists (clozapine, risperidone, olanzapine, and quetiapine). It is now probably safe to say that any set of biological, neurophysiological, psychologic or other tests will find this group of patients looking much more closely like schizophrenia than any other category. In fact, the abnormalities consistent with schizophrenia may already be present in schizophreniform disorder. One such abnormality, decreased gray matter volume, has been seen in MRI studies but to a lesser extent than in patients with chronic

schizophrenia. The cause of pure schizophreniform disorder will probably not be known for a long time, because a patient group that small will be hard to study.

Diagnostic and Clinical Features The DSM-IV criteria for schizophreniform are listed in Table 13.1-3. Schizophreniform disorder in its typical presentation is a rapid-onset psychotic disorder without a significant prodrome. Hallucinations, delusions, or both will be present; negative symptoms of alogia and avolition may be present. Affect may be flattened, which is seen as a poor prognostic sign. Speech may be grossly disorganized and confused, and behavior may be disorganized or catatonic. The symptoms of psychosis, the negative symptoms, and those affecting speech and behavior will last at least 1 month but may last longer. The patient's degree of perplexity about what is happening should be assessed, as this is a differentiating prognostic sign.

A. Criteria A, D, and E of schizophrenia are met.
B. An episode of the disorder (including prodromal, active, and residual phases) lasts at least 1 month but less than 6 months. (When the diagnosis must be made without waiting for recovery, it should be qualified as "provisional.")

Specify if:
Without good prognostic features
With good prognostic features as evidenced by two (or more) of the following:
(1) onset of prominent psychotic symptoms within 4 weeks of the first noticeable change in usual behavior or functioning
(2) confusion or perplexity at the height of the psychotic episode
(3) good premorbid social and occupational functioning
(4) absence of blunted or flat affect

Table 13.1-3 DSM-IV Diagnostic Criteria for Schizophreniform Disorder

Although the above is the typical presentation, a picture exactly resembling that of schizophrenia may also occur. In that case, the onset may be insidious, premorbid functioning may have been poor, and affect is quite blunted. The only differentiation from schizophrenia for this type of presentation will be duration of the total episode of illness. When it has lasted 6 months, the diagnosis becomes schizophrenia. In making the diagnosis in the case with insidious onset, the “attenuated symptoms” of the acute episode may have lasted for some time. If they have been present for at least 5 months and then the acute episode lasts 1 month, the diagnosis of schizophrenia is appropriate, without a prior diagnosis of schizophreniform disorder.

In the typical form of the disorder, the patient returns to baseline functioning by the end of 6 months. Theoretically, repeated episodes of schizophreniform illness are possible, each lasting less than 6 months, but rarely is functioning not lost with repeated episodes of this severe illness, and schizophrenia is a more likely consideration.

Ms. L.J. was a 29-year-old Hispanic second daughter of an intact and stable family. She completed high school without problems and was described as outgoing and friendly. She considered college but opted to work. She spent several years as factory worker and had decided to go back to school and become a teacher. Five months ago she had a sudden “awareness” that God was present and filling the souls of people around her. She became acutely distressed when she realized God was not going to “touch her.” Her family was quite surprised and alarmed by her sudden change in behavior. She was brought to the local emergency room. While she occasionally drank alcohol and had smoked marijuana in the past, the family did not suspect a substance abuse problem. Toxicology screening in the emergency room was negative for substances. She was admitted to the hospital for evaluation. She told the psychiatrist that she felt she had done something wrong and that was why God had abandoned her. She also reported that she felt people on the ward were reading her mind. She was particularly concerned that her critical thoughts about others could be heard and then these angry people would attack her. L.J. was stabilized on haloperidol then switched to risperidone because of side effects. A family meeting was held to discuss her problems. At that time, the psychiatrist recommended a wait-and-see approach. The psychiatrist told the family and patient to follow up with an outpatient doctor and remain on the medication if the outpatient psychiatrist recommended it. Two months after her admission she no longer was distressed by her religious concerns. However, she still felt people could read her mind. Three months after her admission she no longer felt that people could read her mind, and she had returned to her community college. A month later, she stopped taking her antipsychotic agent because she felt she didn't need it. Two weeks ago her family brought her to the emergency room because she was again talking about God and “hiding” from people who could read her thoughts. She initially refused medications but resumed taking them, with some improvement of her psychosis. A family meeting was held to discuss the return of her psychosis and the fact that she may eventually be diagnosed with schizophrenia.

Differential Diagnosis Although the major differential diagnoses are with brief psychotic disorder and schizophrenia, the rapid onset of acute psychosis may be the most important diagnostic point in a patient's course of illness. The clinician should focus on the prior 6 months, taking a detailed history of occupational and social functioning, the pattern of onset, the presence or absence of mood changes, alcohol and substance abuse, and other illness and prescriptive medication. Of special interest will be any family history of psychiatric illness, mood disorders or schizophrenia-like illnesses in particular. A recent study showed a high prevalence of personality disorders after recovery from the psychosis. One could hypothesize that the personality disorder predisposes one to psychosis especially when under stress.

A complete physical examination is always indicated with the presentation of a psychotic illness. Suggestions of endocrinologic involvement, such as thyroid functioning, should be followed up with laboratory studies. If substance abuse is suspected, however remote a possibility, a toxicology screening test should be performed. Changes in sensorium and the rapid onset of symptoms should raise clinical suspicion of substance toxicity. Alcohol may be involved in a number of ways. Certainly, alcohol withdrawal and the onset of delirium may be associated with psychotic symptoms. Further, alcohol abuse leads to unreliable medication taking, even of prescribed medications, which can lead to psychotic features.

The separation of mood disorders with psychotic features from a rapid-onset schizophreniform disorder may be difficult and tests the clinician's skills. Negative symptoms such as alogia, avolition, and blunted affect may be difficult to distinguish from the loss of interest and pleasure seen with major depressive episodes. Appetite, sleep, and other neurovegetative symptoms may also occur with both. The presence of the psychotic features of the illness, in the absence of these mood features, will assist the clinician in making the diagnosis of schizophreniform disorder, but this may take time to evolve.

To differentiate from brief psychotic disorder, a time cutoff has been established, more than 1 day but less than 1 month. During this period, the diagnosis must be brief psychotic disorder. In diagnostic systems prior to DSM-IV, the presence or absence of a stressor was used to differentiate these two conditions further, but it is no longer used in the nosology, except as a descriptor or modifier. Differentiation is based solely on the time line.

Course and Prognosis The course of schizophreniform disorder is for the most part defined in the criteria. It is a psychotic illness lasting more than 1 month and less than 6 months. The real issue is what happens to persons with this illness over time. Most estimates of progression to schizophrenia range between 60 and 80 percent. What happens to the other 20 to 40 percent is currently not known. Some will have a second or third episode during which they will deteriorate into a more chronic condition of schizophrenia. A few, however, may have only this single episode and then are able to continue on with their lives. While this is clearly the outcome desired by all clinicians and family members, it is probably a rare occurrence and should not be held out as likely.

The prognostic features used to characterize the illness are listed above. Their presence will, indeed, be useful in suggesting some likelihood of a favorable outcome. Clinical experience, however, tempers the confidence in these predictors, as many patients with all four of the descriptors have a deteriorating course and outcome.

Treatment Although no available studies have directly addressed the treatment of schizophreniform disorder, the approach should be that for any psychotic disorder of recent onset. The most important initial evaluation is safety, both for the patient and the patient's environment.

Safety Assessment of safety or danger is a complex series of probabilities, not certainties. The best predictor is, of course, past behavior. Someone suffering from the sudden onset of psychosis may not have any past history if this is the first episode. If so, any evidence of prior violence must be seriously considered in forming the initial treatment plan. The evaluation of predictability and hostile affect becomes critical in deciding whether hospitalization is necessary. With someone suffering from an acute psychotic disorder who shows any signs of hostility, anger, and confusion or has a history of explosive or violent activity, hospitalization should be an important consideration. In the absence of these features, hospitalization may be a consideration if the environment itself, usually the family, cannot comfortably

ensure that the treatment plan can be carried out in a safe, stress-reducing manner. For most families this will not be possible.

Inpatient Treatment Plan The inpatient unit is usually a significant part of the initial treatment plan. In addition to pharmacological management, the unit program and philosophy are critical ingredients in helping to stabilize the patient as rapidly as possible. An environment that is critical, intrusive, and overinvolved, with a multistimulus approach to the patient has negative impact on psychosis-prone patients. With this in mind, the patient should not necessarily be required to attend group meetings, therapeutic community, or orientation but should rather be approached in a one-on-one manner with time-limited interactions. Communication should be direct and simple, and the program should be structured with relatively little free time. Visitors should be oriented to this same principle and should be encouraged to visit one at a time.

Outpatient Treatment Plan The patient who has begun to recover from an acute psychotic episode will continue to need a comfortable environment with considerable structured activity. Complex communications and interactions should be kept to a minimum early, although introduction of some simple group work in an attempt to normalize socialization may be carefully planned. Gradual resumption of activities should be attempted one at a time, with mastery achieved before the introduction of new activities. In the case of a student, for example, it would be much better to begin with one course and succeed than with a full course load that would most likely lead to failure. Incremental progress is the goal, and it should extend well beyond the 6 months required for diagnosis.

Role of the Family There is no more significant factor in the successful outcome of a patient with acute onset psychosis than family involvement in the treatment. As reviewed elsewhere, the data are compelling that a clinical treatment program that enlists the family in a positive clinical alliance does better than one that does not, regardless of the other treatment modalities being used. There is no more consistent finding in outcome studies of the late 1980s and 1990s than the positive outcomes found in programs that work with families. In general, most of these programs begin with some form of educational program about schizophrenia, the importance of medication, the expectations of families, and the identification of early signs of impending relapse. Some of the programs have worked elaborately with patients and their families with behavioral paradigms, others have worked with monthly group interactions involving multiple patients and their families. For many of these families this introduction to working as a member of the treatment team enlists them into a long-term positive relationship with the treatment program. For others, it gives them the skills needed to participate in the rehabilitation process. This positive alliance will serve the program, the patient, and the family well. It opens lines of communication and takes a major step toward ensuring that the patient will receive the best monitoring and most appropriate treatment available. Active involvement of a family support group, such as the local chapter of the Alliance for the Mentally Ill, is often quite useful as well. Frequently, however, families experiencing their first episode of psychosis in a family member find association with a group of people who have family members with chronic illnesses to be too threatening. They may wish to believe that their family member will recover, and certainly for those who have true schizophreniform disorder, this will be true.

Pharmacological Therapy The pharmacological approach to the acutely psychotic patient is one of the most challenging and difficult in all of psychiatry. There was an era in psychiatry when there was time to observe the patient to determine whether there was a transient condition that would be self-limiting. The economic forces of today's psychiatry do not permit such an observation period and demand vigorous pharmacological intervention. Perhaps the sole remaining condition in which it would be reasonable to wait before vigorous pharmacological intervention is one that elicits a high index of suspicion of chronic amphetamine abuse, with a positive toxicology screening test result. With these patients it is probably better to wait and treat the agitation with benzodiazepines; the psychosis will usually resolve. Even among these patients will be a small but significant group who will continue with what looks like a schizophreniform or schizophrenia-like illness. Whether this group represent a subgroup of patients who were already at risk for schizophrenia or whether chronic amphetamine abuse sensitizes dopamine receptors in some patients is not known.

Given that it is not economically feasible to wait before initiating treatment, selection of the most appropriate medication becomes a critical decision. The choices basically come down to selection of an antipsychotic agent. For many years this decision involved selecting the antipsychotic agent whose side-effect profile fit the needs of the patient best. If the patient was agitated, a more sedating antipsychotic agent (e.g., chlorpromazine [Thorazine], thioridazine [Mellaril]) would be selected. If not, a less sedating, high-potency compound would be used (e.g., haloperidol, fluphenazine [Prolixin]). Both strategies, however, exposed the patient to extrapyramidal adverse effects initially and to tardive dyskinesia if long-term continuation was needed. With the use of anticholinergic medications, some of the extrapyramidal symptoms could be reduced. However, anticholinergic medications themselves have been associated with decrements in memory, executive functioning, and new learning. Therefore they are used much less frequently than previously and certainly not used routinely unless adverse effects are present. There are now other choices with the advent of the novel serotonin-dopamine antagonists. These antipsychotic agents, while considerably more expensive, hold out the advantage of fewer extrapyramidal adverse effects. They may rapidly become the medications of first choice for psychosis, because they are much better tolerated by the patient and are thus more likely to be taken over a period of time, eliminating the potential for relapse from noncompliance. The expense of readmission of a patient more than makes up for the difference in cost.

Dosage of any antipsychotic agent should be at the lowest possible level, both for adverse effect prevention and for cost. There is a tendency for medication dosages to climb in an effort to shorten the length of the psychosis. Originally, the concept of “rapid neuroleptization was a method of treatment in which a patient was given antipsychotic medication every hour until sedated. Thorough evaluation of this strategy revealed no therapeutic advantages and considerably increased risk for acute dystonic reaction. It is now widely accepted that the full resolution of a psychotic episode may take anywhere from 3 to 6 weeks. Pressure to discharge a patient well before this time certainly places considerable psychological pressure on the physician to increase the medication dosage. It is not clear that there is any advantage to doing this, and maintaining a lower dosage keeps the patient considerably less uncomfortable with adverse effects. If agitation is a problem, addition of a medium- to long-acting benzodiazepine will usually produce the desired results. Benzodiazepines are much better at sedation than are antipsychotic agents. Data suggest that use of a benzodiazepine reduces the amount of an antipsychotic agent that must be used.

A small subgroup of patients present with an acute psychotic episode that rapidly resolves. The more rapid the resolution, the more likely it is that they have a self-limited disorder. These patients will probably not meet the DSM-IV diagnosis of schizophreniform disorder with its requirement of 1 month of symptomatology.

For those whose symptoms do last 1 month or longer and who meet the criteria for schizophreniform disorder, there is a question of how long do they need to be on medication. Although no study has directly addressed this question with patients who met DSM-IV criteria for schizophreniform disorder, strategies have been tested on patients with schizophrenia who have been recruited in an acute episode. Patients who were taken off medication in the first 6 months did much worse than those who were maintained at standard dosages or those who had an 80 percent reduction of dosage. Currently, the low-dosage strategy for maintenance should be considered with the standard antipsychotic agents and probably with the serotonin-dopamine antagonists as well.

With the standard antipsychotic agents, a completely resolved psychotic episode, and full return to premorbid functioning, the usual decision point has been 6 months. This time frame was driven by the finding that almost no cases of tardive dyskinesia occur before 6 months of continuous medication. Going beyond 6 months does increase this risk. With the serotonin-dopamine antagonists, tardive dyskinesia is presumed to be a much lower risk if at all, and thus clinical judgment is needed. If a gradual tapering strategy is selected, the dosage should not be lowered more frequently than every 3 to 4 months if the physician wishes to see the effect of one dosage lowering before initiating the next. Unlike antibiotic use, for example, the infection may well return quickly after premature discontinuation of the medication, psychosis does not immediately reappear even if the medication is completely eliminated. Relapse curves from dosage-discontinuation studies are quite compelling in this regard.

BRIEF PSYCHOTIC DISORDER

History Brief psychotic disorder is a new diagnosis in DSM-IV that subsumes the former diagnostic category of brief reactive psychosis, which first appeared in DSM-II. Brief psychotic disorder is one of the least understood and least studied types of functional psychosis; most research has had methodological flaws and unclear diagnostic criteria. Historically, Karl Jaspers described the concept of a reactive psychosis in 1913. Jaspers described the essential features, which include presence of an identifiable traumatic stressor, close temporal relation between stressor and psychosis, and generally benign course of the psychotic episode.

Jaspers also believed that the content of the psychosis was related to the trauma and served some therapeutic purpose.

Comparative Nosology Over the past century myriad terms have been used to describe psychotic episodes precipitated by stressful events, including good-prognosis schizophrenia, but brief reactive psychosis had gained prominence until the DSM-IV. Compared with the DSM-III and DSM-III-R criteria that required a precipitating stressor and confusion or emotional turmoil during the episode, the new diagnosis of brief psychotic disorder is less restrictive. With its broader definition, brief psychotic disorder will presumably reduce the use of the classification “psychotic disorder not otherwise specified.” Because stressors and reactions to stressors are so ubiquitous and ill defined, reactivity to a stressor is no longer necessary for the diagnosis and is used instead as a descriptor. Scandinavian researchers have been integral in delineating this disorder, which has been gradually gaining international recognition. Conceptual generalization of the disorder is both supported and challenged by culture-bound syndromes such as koro and amok, which demonstrate significant differences while still falling under the rubric of

brief psychotic disorder. The ICD-10 classifies these symptoms as “acute and transient psychotic disorders” (see Table 13.3-1). Subtypes include “acute polymorphic psychotic disorder without symptoms of schizophrenia,” which has an overall picture of unstable, highly emotional symptoms with psychotic features that would not justify a diagnosis of schizophrenia. In contrast, the diagnosis acute polymorphic psychotic disorder with symptoms of schizophrenia also describes an unstable clinical picture, but symptoms of schizophrenia are also present for a major part of the time. If the acute picture is marked by delusions, only a diagnosis of “other acute predominately delusional psychotic disorders” can be used. Finally, any unspecified transient psychotic disorder can be designated as other or unspecified acute and transient psychotic disorders.

Epidemiology Relatively uncommon in DSM field trials, brief psychotic disorder has large discrepancies in reported incidence and prevalence rates because of methodological flaws and diagnostic variability in the literature. Its age of onset is most commonly reported to be in the late 20s or early 30s. Although reliable data on sex and sociocultural determinants are limited, preliminary data suggest a higher incidence in women and persons in developing countries. Such epidemiological patterns are sharply distinct from those of schizophrenia.

Etiology Little is known about the etiology of brief psychotic disorder. The existence of one or many events becomes the identified causative agent in psychotic disorder with marked stressor (brief reactive psychosis). Both the magnitude and the multiplicity of such stressors are posited to be important, but no well-controlled studies assessing the causal role of various types of stressors are available. Severe intrapsychic conflict (an internal stressor) may be the etiological agent for brief psychotic disorder without a marked stressor. Preexisting characterological psychopathology of either cluster A or B variety may predispose a person to development of the disorder. Many explanatory models of this increased vulnerability exist, but most are based on immature defenses and ego development as major contributors. Family studies support a genetic vulnerability to brief reactive psychosis but do not support a genetic link between this disorder and schizophrenia.

Diagnostic and Clinical Features The DSM-IV diagnostic criteria are listed in Table 13.1-4. DSM-IV defines brief psychotic disorder as impairment in reality testing lasting at least 1 day but not more than 1 month. An eventual full return to premorbid levels of functioning is required; if the diagnosis is made without waiting for the anticipated recovery, then the qualifier provision must be added. At least one of the following symptoms is present during the circumscribed illness: delusions, hallucinations, disorganized speech, disorganized behavior, or catatonia. Exclusionary criteria include the presence of a mood disorder with psychotic features, schizoaffective disorder, schizophrenia, and any psychotic disorder secondary to the direct physiological effects of a substance or a general medical condition. If symptoms occur in response to one or more events that would be markedly stressful to almost anyone in similar circumstances and within the same cultural context, then the illness bears the specifier *with marked stressor* (formerly referred to as brief psychotic disorder). Conversely, if symptoms are not in response to such an event, the specifier *without marked stressor* is applied. An additional specifier, *with postpartum onset*, indicates the onset of psychotic symptoms within 4 weeks postpartum.

A. Presence of one (or more) of the following symptoms:
(1) delusions
(2) hallucinations
(3) disorganized speech (e.g., frequent derailment or incoherence)
(4) grossly disorganized or catatonic behavior
Note: Do not include a symptom if it is a culturally sanctioned response pattern.
B. Duration of an episode of the disturbance is at least 1 day but less than 1 month, with eventual full return to premorbid level of functioning.
C. The disturbance is not better accounted for by a mood disorder with psychotic features, schizoaffective disorder, or schizophrenia and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify if:
With marked stressor(s) (brief reactive psychosis): if symptoms occur shortly after and apparently in response to events that, singly or together, would be markedly stressful to almost anyone in similar circumstances in the person's culture
Without marked stressor(s): if psychotic symptoms do not occur shortly after, or are not apparently in response to events that, singly or together, would be markedly stressful to almost anyone in similar circumstances in the person's culture
With postpartum onset: if onset within 4 weeks postpartum

Table 13.1-4 DSM-IV Diagnostic Criteria for Brief Psychotic Disorder

Patients with brief psychotic disorder typically have rapid-onset psychotic symptomatology and often demonstrate emotional turmoil, confusion, or both. Prompt recovery with a full return to premorbid level of functioning within a month is dictated by definition. It is imperative to assess the impact of culture on symptom presentation prior to making the diagnosis. In the case of brief reactive psychosis, the precipitant may be one or a series of life stressors, such as the loss of an important relationship, familial disruption, or combat-related trauma. In such cases, environmental adversity combines with cultural expectations and support systems to manifest symptoms distinctly.

R.S. was a 44-year-old Haitian male admitted for observation at the local emergency room. He was agitated and combative, requiring restraints and several intramuscular doses of droperidol and lorazepam. The psychiatrist could not interview him under these acute circumstances. His mother arrived soon after and was able to give corroborative history. According to his mother the patient had just learned that his wife and two children had died in a natural disaster in Haiti. Several hours after his first evaluation, the patient was calmer. He told staff that he was hearing his wife talking to him and he wished to “join her.” He also believed the Haitian secret police were coming to arrest him. He was admitted to the inpatient ward and began a course of an antipsychotic agent. By the third day of his hospitalization there was no evidence of the previous psychosis. He was discharged from the hospital and given a follow-up appointment in 1 month. When he returned the next month he had been medication free for that time. He was grieving the loss of his family but was not psychotic. He was referred to a grief group, which he attended for the next 6 months. In that time he remained sad, but there were no other episodes of paranoia or hallucinations.

Differential Diagnosis Sharing rapid onset of symptoms, brief psychotic disorder must be differentiated from substance-induced psychotic disorders and psychotic disorders due to a general medical condition. A thorough medical evaluation including a physical examination, laboratory studies, and brain imaging will help rule out many of those conditions. With only cross-sectional information, brief psychotic disorder is difficult to differentiate from other types of functional psychosis.

The relationship between brief psychotic disorder and both schizophrenia and affective disorders remains uncertain. As noted above, DSM-IV has made the distinction between brief psychotic disorder and schizophreniform disorder clearer by now requiring a full month of psychotic symptoms for the latter. If psychotic symptoms are present longer than 1 month, the diagnoses of schizophreniform disorder, schizoaffective disorder, schizophrenia, mood disorders with psychotic features, delusional disorder, and psychotic disorder not otherwise specified need to be entertained. If psychotic symptoms of sudden onset are present for less than a month in response to an obvious stressor, the diagnosis of brief psychotic disorder is strongly suggested. Other diagnoses to differentiate include factitious disorder, malingering, and severe personality disorders, with consequent transient psychosis possible.

Course and Prognosis The course of brief psychotic disorder is found in the diagnostic criteria of DSM-IV. It is a psychotic episode that lasts more than 1 day but less than 1 month, with eventual return to premorbid level of functioning. Approximately half of patients diagnosed with brief psychotic disorder retain this diagnosis; the other half will evolve into either schizophrenia or a major affective disorder. There are no apparent distinguishing features between brief psychotic disorder, acute-onset schizophrenia, and mood disorders with psychotic features on initial presentation. Several prognostic features have been proposed to characterize the illness, but they are inconsistent across studies. The good prognostic features are similar to those found in schizophreniform disorder: acute onset of psychotic symptoms, confusion or emotional turmoil at the height of the psychotic episode, good premorbid functioning, the presence of affective symptoms, and short duration of symptoms. There is a relative dearth of information on the recurrence of brief psychotic episodes, however, so the course and prognosis of this disorder have not been well characterized.

Treatment Although no available studies directly address the treatment of brief psychotic disorder, the treatment approach should focus on the acute onset of psychotic symptoms. In particular, patient safety is of paramount importance. Depending on the danger the patient represents to self and others, psychiatric hospitalization is often warranted. A patient demonstrating acute psychotic symptoms who also displays a hostile affect or has a history of violence is particularly likely to require hospitalization. In addition to providing a safe and structured environment, hospitalization permits observational monitoring and a medical examination investigating potential etiological factors.

If medication is necessary, a high-potency antipsychotic agent in low dosage is typically recommended. An antiparkinsonism agent may be added if extrapyramidal adverse effects occur. A benzodiazepine used in combination with an antipsychotic agent can act synergistically, thereby lowering the necessary doses of each and reducing the risk of side effects. Benzodiazepines can also be used as monotherapy to reduce agitation without obscuring the clinical picture. The role of other

psychotropic medications such as mood stabilizers and antidepressants is not yet clear.

After the acute episode has subsided, long-term treatment is required. An individualized treatment strategy based on increasing problem-solving skills while strengthening the ego structure through psychotherapy, appears to be the most efficacious. Involvement of the family in the treatment process is crucial to a successful outcome and is reviewed elsewhere in this chapter. There is no role for maintenance antipsychotic treatment in brief psychotic disorder; if such treatment is required, the diagnostic assumptions must be questioned.

SUGGESTED CROSS-REFERENCES

A more detailed review of schizophrenia is presented in Chapter 12; acute and transient psychotic disorders are presented in Section 13.3. Mood disorders are covered in Chapter 14. Personality disorders are discussed in Chapter 24.

SECTION REFERENCES

Blacker D, Tsuang MT: Contested boundaries of bipolar disorder and the limits of categorical diagnosis in psychiatry. Am J Psychiatry *149:*1473, 1992.

Goldstein JM, Faraone SV, Chen WJ, Tsuang MT: The role of gender in understanding the familial transmission of schizoaffective disorder. Br J Psychiatry *163:*763, 1993.

Greil W, Ludwig-Mayerhofer W, Erazo N, Engel RR, Czernik A, Giedke H, Muller-Oerlinghausen B, Osterheider M, Rudolf GA, Sauer H, Tegeler J, Wetterling T: Lithium vs. carbamazepine in the maintenance treatment of schizoaffective disorder: a randomized study. European Archives of Psychiatry and Clinical Neurosciences *247:* 42, 1997.

Jorgensen P, Bennedsen B, Christensen J, Hyllested A: Acute and transient psychotic disorder: Comorbidity with personality disorder: Acta Psychiatrica Scandainavica *94:*460, 1996.

Keck PE, McElroy SL, Strakowski SM: New developments in the pharmacological treatment of schizoaffective disorder. J Clin Psychiatry *57:*41, 1996.

Keck PE, McElroy SL, Strakowski SM, West SA: Pharmacological treatment of schizoaffective disorder. Psychopharmacology *114:*529, 1994.

Kendler KS, McGuire M, Gruenberg AM, Walsh D: Examining the validity of DSM-III-R schizoaffective disorder and its putative subtypes in the Roscommon family study. Am J Psychiatry *152:*755, 1995.

Lapensee MA: A review of schizoaffective disorder, I. Current concepts. Can J Psychiatry *37:*335, 1992.

*Lapensee MA: A review of schizoaffective disorder, II. Somatic treatment. Can J Psychiatry *37:*347, 1992.

Lapierre YD: Schizophrenia and manic-depression: Separate illnesses or a continuum? Can J Psychiatry *39*(Suppl):S59, 1994.

Levitt JJ, Tsuang MT: The heterogenity of schizoaffective disorder: Implication for treatment. Am J Psychiatry *145:*20, 1988.

Maier W, Lichtermann D, Minges J, Heun R, Hallmayer J, Benkert O. Schizoaffective disorder and affective disorders with mood-incongruent psychotic features: Keep separate or combine? Evidence from a family study. Am J Psychiatry *149:*1666, 1992.

Migliorini G, Lazzarin E: Acute psychosis: Clinical contribution. Minerva Psychiatr *34:*231, 1993.

Muller-Siechender F, Muller MJ, Hillert A, Szegedi A, Wetzel H, Benkert O, Risperidone versus haloperidol and amitriptyline in the treatment of patients with a combined psychotic and depressive syndrome. J Clin Psychopharmacol *18:*111, 1998.

*Robinson DG, Woerner MG, Ma J, Alivir J, Geisler S, Doreen AM, Sheitman B, Chakos M, Mayerhoff D, Bilder R, Goldman R, Lieberman J: Predictors of treatment response from a first episode of schizophrenia or schizoaffective disorder. Am J Psychiatry *156:*544, 1999.

Rubin P: Neurobiological findings in first admission patients with schizophrenia or schizophreniform disorder. Danish Medical Bulletin *44:*140, 1997.

Smith GN, MacEwan W, Ancill RJ, Honer WG, Ehmann TS: Diagnostic confusion in treatment-refractory psychotic patients. J Clin Psychiatry *53:*197, 1992.

*Susser E, Fennig S, Jandorf L, Amador A, Bromet E: Epidemiology, diagnosis and course of brief psychoses. Am J Psychiatry *152:*20, 1995.

Susser E, Wanderling J: Epidemiology of nonaffective acute remitting psychosis vs. schizophrenia. Arch Gen Psychiatry *51:*20, 1994.

Taylor MA, Amir N: Are schizophrenia and affective disorder related? The problem of schizoaffective disorder and the discrimination of the psychoses by signs and symptoms. Compr Psychiatry *35:*420, 1994.

*Thweatt R: European interest in transient psychotic episodes. Am J Psychiatry *143:*557, 1986.

Tsuang D, Coryell W: An 8-year followup of patients with DSM-III-R psychotic depression, schizoaffective disorder and schizophrenia. Am J Psychiatry *150:*1182, 1993.

*Winokur G, Monahan P, Coryell W, Zimmerman M: Schizophrenia and affective disorder—distinct entities or continuum? An analysis based on a prospective 6-year follow-up. Compr Psychiatry *37:*77, 1996.

*Zipursky RB, Lambe EK, Kapur S, Mikulis DJ: Cerebral gray matter volume deficits in first episode psychosis. Arch Gen Psychiatry *55:*540, 1998.

13.2 DELUSIONAL DISORDER AND SHARED PSYCHOTIC DISORDER

THEO C. MANSCHRECK, M.D., M.P.H.

Once viewed as too rare to warrant a separate classification, delusional disorder has emerged in recent years as a focus of clinical research and treatment innovation. Better definition and a growing literature have revitalized the efforts to characterize, understand, and treat these conditions. Limited but growing evidence supports not only its occurrence, but its distinctiveness from schizophrenia and mood disorder as well as its treatability. Delusional disorder refers to a group of disorders, the chief feature of which is the presence of a nonbizarre delusion. It is the delusion and the relative absence of other psychopathology that unifies these disorders in terms of natural history and impact on functioning. Once called *paranoia*, this condition as defined in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) is easier to recognize and less subject to misdiagnosis.

Despite such advances, clinicians are relatively ill-informed about delusional disorders and many have only seen an occasional example. There are several possible reasons why this is so. Persons with this condition do not regard themselves as mentally ill and actively oppose psychiatric referral. Because they may experience little impairment, they generally remain outside hospital settings, appearing reclusive, eccentric, or odd, rather than ill. If they do have contact with professionals, it is more likely to be with lawyers regarding litigious concerns; with medical specialists regarding health concerns; or with the police regarding complaints of trespass, persecution, or threat, rather than with psychiatric clinicians regarding complaints of emotional disorder. A hallmark of these disorders is that the patient does not believe that he or she is deluded or in need of psychiatric assistance. In the infrequent psychiatric encounter, clinicians tend to diagnose these disorders as other conditions, often as schizophrenia or mood disorders.

Although delusional disorders are uncommon, they are probably not as rare as previously thought. While many individuals with such disorders seek assistance from other medical specialists, judges, or the police, they are increasingly being recognized as psychiatrically ill. The relationship of these disorders to other psychoses remains unclear, and much about them is puzzling. The DSM-IV requirement of excluding other conditions is prudent given the special importance of differential diagnosis. Armed with newer and better criteria, clinical research is ongoing in areas such as natural history, pathogenesis, neuropsychology, neuroimaging, treatment, and even genetics. Although the DSM-IV criteria are not definitive, they have provided a sound basis for clinical and research investigation. Systematic studies based on larger samples of these disorders are needed to anchor classification with sound information; however, such studies may be difficult to conduct because of the large numbers of patients required and their reluctance to participate in research. A biological basis for these disorders is proposed on many grounds, but its definition remains elusive. The study of misidentification syndromes (e.g., the Capgras's syndrome) has led to interesting hypotheses and methods that draw on neuropsychological and clinical insights that may inspire progress in delusional disorders. Treatment remains an obstacle, although recent reports suggest that favorable responses to psychopharmacologic and psychotherapeutic interventions are more common than previously thought.

DEFINITION

Delusional disorder is the current classification for a group of disorders of unknown cause, the chief feature of which is the delusion (Table 13.2-1). Although the specific content of the delusion may vary from one case to the next, it is the occurrence of the delusion, its persistence, its impact on behavior, and its prognosis that unifies these seemingly different disorders. In considerable agreement with Emil Kraepelin's concept of paranoia, the revised third edition of DSM-III-R provides reliable criteria for identifying cases and collecting systematic information about these conditions. This development in classification helped to reestablish the clinical importance of this group of disorders and may have reversed a trend of infrequent diagnosis. The criteria use the term *delusional* to avoid the ambiguity of the term *paranoid* used earlier in the third edition of DSM (DSM-III) classification, *paranoid disorders*, as well as to emphasize that the category includes disorders in which delusions other than those of the persecutory or jealous type are present. Although these changes were initially confusing, especially in terms of comparisons to diagnostic approaches elsewhere, they have gained acceptance and have created a more level playing field for further empirical contributions.

Table 13.2-1 DSM-IV Definition of Delusion and Certain Common Types Associated With Delusional Disorders

In 1994 DSM-IV refined the definition and the boundaries with other disorders, including substance-induced disorders, mental disorders due to general medical conditions, mood disorders, and schizophrenia. No laboratory test exists to assist in diagnosis. The DSM-IV definition, like its predecessors, hinges on the presence of a nonbizarre delusion. DSM-IV acknowledges the difficulty of judging whether a delusion is bizarre, meaning clearly implausible, not understandable, and not derived from ordinary life experiences. In contrast, the nonbizarre delusion involves situations or circumstances that can occur in real life (e.g., being followed, infected, or deceived by a lover). DSM-IV also emphasizes the differential diagnoses of schizophrenia, mood disorders, substance-induced disorders, and mental disorders due to a general medical condition before the diagnosis of delusional disorders can be made. These conceptual refinements and demarcations from other conditions have increased the usefulness of the criteria for delusional disorder.

Delusional Disorder According to DSM-IV, the diagnosis of delusional disorder can be made when a person exhibits nonbizarre delusions of at least 1 month's duration that cannot be attributed to other psychiatric disorders. Definitions of the term *delusion* and types relevant to delusional disorders are presented in Table 13.2-1. *Nonbizarre* means that the delusions must be about situations that can occur in real life, such as being followed, infected, loved at a distance, and so on; that is, they usually have to do with phenomena that, although not real, are nonetheless possible. There are several types of delusions, and the predominant type is specified when the diagnosis is made.

In general, the patient's delusions are well systematized and have been logically developed. The person may experience auditory or visual hallucinations, but these are not prominent features. Tactile or olfactory hallucinations may be present and prominent if they are related to the delusional content or theme, examples are the

sensation of being infested by bugs or parasites, associated with delusions of infestation, and the belief that one's body odor is foul, associated with somatic delusions. The person's behavioral and emotional responses to the delusion appear to be appropriate. Impairment of functioning is not marked and personality deterioration is minimal, if it occurs at all. General behavior is neither obviously odd nor bizarre.

Shared Psychotic Disorder This unusual condition has also been called *folie à deux* and *induced or shared psychotic disorder*. It develops in an individual in the context of a close relationship with another person who has an established delusion that he or she also believes, and requires an absence of psychotic disorder prior to the onset of the induced delusion; it is usually classified with paranoid disorders.

HISTORY

Nineteenth-century psychiatry devoted much attention to the description of paranoid disorders, in which delusions are a cardinal feature. Karl Ludwig Kahlbaum's description of paranoia in 1863 was the first in a series of contributions that culminated in the classification of paranoia, and inspired that of folie à deux, morbid jealousy, the better-known schizophrenias, and mania. His work also led to a recognition that paranoid features are nonspecific characteristics of many medical diseases. Subsequent work has led to refined criteria for paranoid and related disorders and has reestablished awareness of less common paranoid presentations such as delusional disorder.

Many clinicians remember being taught that paranoia is so rare that most would not see a single such patient during an entire career. This widespread belief has compromised interest in paranoid disorders. The fact that most persons with delusional disorder live in the community and do not generally seek psychiatric care has made it difficult to carry out systematic case studies. Indeed, knowledge of these conditions has grown slowly. However, case series such as those of Alistair Munro (for delusional disorder, somatic type) or those of Nils Retterstol have been influential in shaping understanding and awareness. What they reveal is that there are persons with these disorders, that the disorders are complex forms of psychiatric illness, and that much remains to be learned.

A major change in the classification of delusional disorders in DSM-III-R and DSM-IV has been to emphasize the central role of delusions in those disorders and to steer away from the vague label of *paranoid*, which has become synonymous with *suspicious* and has come to apply largely to a personality disorder. Indeed, suspiciousness occurs in only some of these disorders. The history of the concept of paranoia indicates that lack of clarity in its use is not new. The word *paranoia* was coined by the ancient Greeks from roots meaning *beside* and *self*. Hippocrates applied this term to delirium associated with high fever, but other writers used it to describe demented conditions and madness. It sometimes meant thinking amiss, folly, and the like; hence, its meaning was unclear. For centuries the term fell into disuse until a revival of interest in the nineteenth century.

In 1863 Karl Kahlbaum classified paranoia as a separate mental illness: “a form of partial insanity, which throughout the course of the disease, principally affected the sphere of the intellect.” Influenced by the new scientific methods of empirical medicine, Kahlbaum emphasized the importance of natural history in mental illness and restricted the use of the term *paranoia* to a persistent delusional illness that remained largely unchanged throughout its course. He noted that delusions could occur in other medical and psychiatric conditions.

Emil Kraepelin found the paranoid concept troublesome and altered his thinking on it with each edition of his influential textbook. His final view advocated three types of paranoid disorder. Like Kahlbaum, Kraepelin based his conclusions on an analysis of the natural history of mental disorders, particularly on outcome, because etiology was obscure. He restricted the definition of paranoia to an uncommon, insidious, chronic illness (he saw 19 cases) characterized by a fixed delusional system, an absence of hallucinations, and a lack of deterioration of the personality. The types of delusions included persecutory, grandiose, jealous, and possibly hypochondriacal. He considered this illness to derive from defects in judgment, a disorder of personality caused by constitutional factors and environmental stress. Paraphrenia was a second paranoid disorder that developed later than dementia precox and was milder. Hallucinations (auditory in particular) occurred, but there was no mental deterioration (dementia). Finally, there was *dementia paranoides*, an illness that initially resembled paranoia but had an earlier onset and showed a deteriorating course. Because of this latter feature, Kraepelin considered dementia paranoides a form of dementia precox that arose from disorders of thought, cognition, and emotion. Kurt Mayer's follow-up of Kraepelin's 78 paraphrenia cases challenged the validity of this category because the vast majority of patients showed an outcome indistinguishable from that of dementia precox, casting doubt on the separability of this group. Karl Kolle's follow-up of Kraepelin's paranoia cases indicated some overlap with dementia precox. Kraepelin also emphasized that isolated paranoid symptoms occurred in a variety of psychiatric and medical illnesses.

Eugen Bleuler also recognized paranoia; he broadened its definition to include cases with hallucinations—a paranoid form of dementia precox for which he coined the term *schizophrenia*—and an intermediate group. However, he thought that the paranoia described by Kraepelin was so rare that it did not warrant a separate classification. Further, he argued that schizophrenic symptoms must be suspected and carefully sought even in those cases. He believed that paraphrenia and intermediate conditions were forms of schizophrenia linked by “much that was identical,” and particularly by a common disturbance in associative processes. He also emphasized that paranoid symptoms occurred in other conditions and that to label the symptoms schizophrenic required at least one of the fundamental symptoms: loosened associations, ambivalence, inappropriate affect, and autism. Bleuler's contributions reinforced a trend toward the diagnosis of paranoid illness as a form of schizophrenia.

Sigmund Freud used the autobiographical writings of Judge Daniel Schreber to illustrate the role of psychological defense mechanisms in the development of paranoid symptoms. He proposed that Schreber's illness involved a process of denial or contradiction of repressed homosexual impulses toward his father. Persecutory and other delusions result from projecting these denied yearnings onto the environment. Freud did not differentiate subtypes of paranoid disorder, and confused the issue somewhat by proposing that the term *paraphrenia* be substituted for the term *dementia precox* or *schizophrenia*. The major impact of Freud's work was to suggest hypotheses that indicated the relationship between certain delusions and personality.

Ernst Kretschmer's work on the theory of paranoia emphasized that certain sensitive personalities, characterized by depressive, pessimistic, and narcissistic traits, developed paranoid features acutely when key or precipitating experiences occurred at critical moments in their lives. He observed that these individuals did not develop schizophrenia and had a favorable prognosis. A number of other observations, predominantly but not exclusively emanating from European clinicians (e.g., the American concept of *hysterical psychosis*), proposed connections between personality and delusion development. Those efforts, based on various theories of the cause of paranoid disturbance, have persisted despite modest empirical support. Out of such work have come terms, such as *reactive* and *psychogenic psychosis*, which have figured in various classification schemes, undermining the effort to bring international consistency in definition.

Many barriers remain to international agreement on definition. For example, the term *paraphrenia*, unlike *paranoia*, has slipped into near obscurity in North America. In the United Kingdom, however, the diagnosis of late paraphrenia is often used and it is occasionally used in the United States. This term refers to cases of late-onset paranoid symptomatology not characterized by the presence of dementia, confusion, or mood disorder. Interestingly, Kraepelin did not identify a late age of onset in his cases. The potential overlap with late-onset cases of schizophrenia has been a focus of investigation and controversy. With the removal of the DSM-III age criterion for schizophrenia (upper limit of age of onset at 45) in DSM-III-R, cases of late-onset symptoms have tended to be diagnosed as schizophrenia in the United States. Nevertheless, clinical research continues to address the puzzle of whether late-onset cases, despite considerable overlap in clinical features, arise from a variety of causes.

Current controversy is based on these historical antecedents and contemporary practices. DSM-III introduced greater rigor in the assessment by requiring clearer criteria boundaries among the varied disorders with delusions. Increased awareness that delusions result from numerous conditions has had a positive influence on the diagnostic process. Yet much of current clinical and research writing on paranoid conditions has characteristically avoided defining the terms *paranoid* and *delusion*, apparently because everyone was assumed to know what these terms mean. In popular and literary usage the term *paranoid* has come to mean insane, angrily suspicious, distrustful, or irrationally irritable. However vague the concept may be, it continues to be used in clinical work. Because it is necessary to differentiate conditions with paranoid features, a useful concept of the term is fundamental. However, the nature and definition of delusions upon which modern psychopathology and psychiatry are built remain unclear.

Shared Psychotic Disorder Jules Baillarger first described the syndrome in 1860, calling it *folie à communiquée*, although the first description is commonly attributed to Ernest Charles Lasègue and Jules Falret, who described the condition in 1877 and gave it the name of *folie à deux*. The syndrome has also been called *communicated insanity, contagious insanity, infectious insanity, psychosis of association*, and *double insanity*. Marandon de Montyel divided *folie à deux* into three groups (*folie imposée, folie simultanée*, and *folie communiquée*), and Heinz Lehmann added a fourth group, *folie induite*.

PARANOID CONCEPT

Paranoid signs and symptoms are among the most dramatic and serious disturbances in psychiatry and medicine but the term *paranoid* refers to a variety of behaviors that may not be psychopathological nor indicative of schizophrenia; hence, the meaning of the term has become obscure. Some clinicians label ordinary suspiciousness paranoid; others restrict use of the term to persecutory delusions; still others apply the term only to grandiose, litigious, hostile, and jealous behavior, despite the fact that those behaviors may be within the normal spectrum. To make the paranoid concept useful and less vague requires consideration of several points:

1. The term *paranoid* is a clinical construct used to interpret observations, and in order to apply this construct effectively, the clinician must know its meaning and be able to make accurate observations of potentially paranoid behavior.
2. Use of the term *paranoid* means the clinician has judged that the person's behavior is psychopathological. This judgment is usually based on the discovery that the person who displays such features is either disturbed or disturbing to others.
3. Although many contributions to understanding paranoid phenomena have focused on conditions in which paranoid features are central (e.g., schizophrenia for Bleuler, paranoia and dementia paranoides for Kraepelin), those features are not necessarily associated with schizophrenia and can appear in other psychiatric and medical disorders. Hence, paranoid features indicate psychopathology, but no specific cause or outcome (Table 13.2-2)

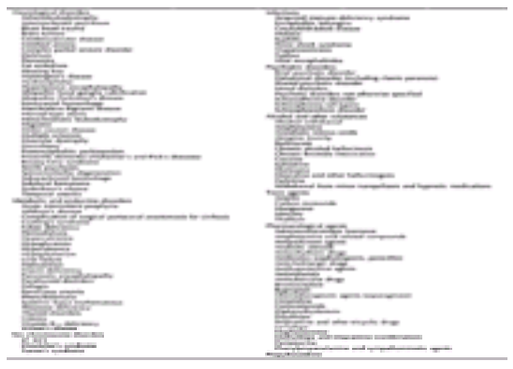

Table 13.2-2 Conditions and Agents Associated With Delusions and Other Paranoid Features

4. The observations that form the basis for judging behavior to be paranoid are of two kinds: subjective (part of the private mental experience of the patient, e.g., a delusion) and objective (observable as a manifest form of behavior, such as litigiousness, guardedness, and grandiosity). Table 13.2-3 is a list of the subjective and objective features that have traditionally been labeled paranoid and that are frequently found in association; some of these features can be manifestations of normal behavior. The judgment that such features are paranoid may rest on how extreme or inappropriate they are, their presence in combination or association with other behaviors on the list, and the presence of delusions.

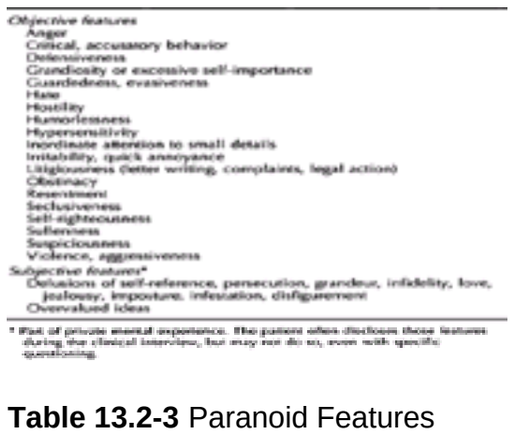

Table 13.2-3 Paranoid Features

5. The term *paranoid delusion* has traditionally referred to a wide variety of delusions, not simply those of grandeur, persecution, or jealousy. Because of recent confusion that term probably should not be used. The term *paranoid* and related terms are defined in Table 13.2-4.

Table 13.2-4 Terminology Connected with Paranoia

Delusions When Karl Jaspers formulated the concept of delusion widely used today, he suggested three criteria: (1) subjective certainty, (2) incorrigibility, and (3) falsity of content. He viewed these criteria as tentative, preferring to consider them as approximations to a definition in that they provided practical suggestions for detecting delusions rather than actually defining them. This and later contributions emphasized a certain humility about the delusion concept that has not been sustained in contemporary formulations of this psychopathological feature. Numerous, often ignored, problems compromise the clinical research utility of the delusion concept. For example, according to DSM-III-R and DSM-IV, delusion is “an incorrect inference about external reality.” This definition has certain implicit and complicating features: (1) there is a process of inference separable from the belief that the process produces, (2) this same process is used by normal persons to generate beliefs about the world, and (3) this process is impaired in delusional patients. As pointed out later in this chapter, the validity of the latter two assertions is questionable. Also, central to the concept of delusional disorder is the distinction between *bizarre* (impossible) and *nonbizarre* (possible) delusions. This distinction has been difficult to apply reliably in clinical assessment, yet on it rests considerable weight in making the diagnosis of delusional disorder. Further examples of difficulties concerning the definition of delusion have been discussed by Manfred Spitzer, who has traced the movement from philosophy to empirical science in the evolution of the definition of delusion. Awareness of the vagaries and imprecise nature of the definition of delusion is essential to clinical and research efforts.

Since the early nineteenth century delusions have been classified by content or theme. Other descriptive dimensions have gained acceptance through clinical use and some empirical research: understandability, degree of certainty, systemization, complexity, relevance to patient's life, plausibility, onset, associated psychopathology, and time course. These features are used to grasp the nature of the delusional experience, translate clinical observations into diagnostic and treatment interventions, and design research.

In clinical encounters delusions are usually easy to detect. Certain features (Table 13.2-3) of the patient's behavior may suggest the presence of delusions or help confirm the impression that the beliefs are delusional. In subtle cases, however, this task is more challenging. The clinician must make a judgment, based on the behavior and reported private mental experience of the patient, of whether or not delusional beliefs are present. Attempts to present counterevidence and argument may be useful to determine whether the patient's views can be influenced by evidence that is usually sufficient to alter the belief of a normal person. This judgment often depends on deciding whether a threshold indicative of psychopathological disorder has been passed, possibly reflected in the inappropriateness or extreme

nature of the patient's behavior, rather than the simple truth or falsity of the belief. In practice, the only effective approach to assessing delusions is to put together as comprehensive a picture as possible regarding the nature of the patient's condition. Lacking laboratory tests for delusions, clinical judgment will be required to some degree in virtually all cases. At the theoretical level, the definition of delusions is moving gradually away from its roots in philosophy and phenomenological description toward a more empirically derived set of features. This process will take considerable time to achieve a satisfactory resolution of the many issues plaguing this aspect of psychopathology.

COMPARATIVE NOSOLOGY

Certain advances have been made in the nosology of delusional disorders, but the variety of current definitions illustrates that consensus has not yet been achieved. The reasons for such differences are multiple: the principal reason is the lack of relevant data—delusional disorders occur infrequently. Typically, patients continue to function and live in the community without ever seeking clinical intervention. When they do, the condition is easily misdiagnosed because patients may have minimal overt identifying characteristics. Limited knowledge, based largely on case reports, exists; systematic, larger-scale studies are uncommon. Most of these studies are European and have employed varied classifications. Also, the fundamental concept that these disorders are distinct from schizophrenia and mood disorders has until recently been unrecognized by many psychiatrists.

Kahlbaum was the first to use the term *paranoia* to designate a diagnostically separate group of disorders. Kraepelin developed this diagnostic concept further by emphasizing the chronic and unremitting nature of paranoia and the lack of other features such as hallucinations that distinguished it from schizophrenia. In 1952 the first edition of DSM (DSM-I) defined paranoid reactions as conditions in which there are persecutory or grandiose delusions, with emotional responses and behavior consistent with the delusions, but generally lacking hallucinations. The subtypes were *paranoia* (a chronic disorder with systematized delusions) and *paranoid state* (a more acute, less persistent condition with less systematized delusions). In 1968 DSM-II largely preserved these concepts.

DSM-III Although new definitions were established in DSM-III in 1980, earlier concepts are still evident. The essential features of paranoid disorders according to DSM-III were persistent persecutory delusions or delusional jealousy not due to any other mental disorder. Included in the group of paranoid disorders were paranoia, shared paranoid disorder, acute paranoid disorder, and a residual category called atypical paranoid disorder. The boundaries between these conditions and other disorders, such as paranoid personality disorder or paranoid schizophrenia, were noted to be vague. Different types of paranoid disorders were classified on the basis of chronicity. The criteria narrowed the bounds of previous classifications by not including cases with marked hallucinations or certain delusions (e.g., hypochondriacal, erotomanic, and others).

DSM-III-R In 1987 DSM-III-R simplified the DSM-III definition, attempted to minimize the confusion associated with the term *paranoid*, and highlighted the view that the formation of delusions in the absence of schizophrenia, mood disorder, or organic disorder is the essential feature of these conditions. In contrast to DSM-III, diagnosis in DSM-III-R and DSM IV requires a month-long duration of symptoms. Subtyping is based on the predominant type of delusion, which is specified (such as jealous, erotomanic, or somatic). This latter feature broadens the category to include a variety of unusual delusions as well as the more common persecutory type. In many respects these criteria are virtually identical to Kraepelin's definition of paranoia. The two exceptions were Kraepelin's reluctance to endorse a subtype of somatic or hypochondriacal paranoia or to permit cases with hallucinations to fall within this diagnosis. Kraepelin believed that cases with hypochondriacal delusions rarely occurred in this pattern.

Shared paranoid disorder was renamed *induced psychotic disorder* in DSM-III-R and was placed in the category of psychotic disorders not elsewhere classified, along with schizophreniform, schizoaffective disorders, and brief reactive psychosis. This represents a fundamental departure from DSM-III, which classified this disorder among the paranoid disorders. The delusional content of patients with this disorder may concern not only persecution or jealousy but virtually any form of delusion, hence the change in terminology. The term *induced* may better describe the nature of the condition, but hardly resolves the puzzle of causation.

DSM-IV In 1994 a revised classification made modest changes in the DSM-III-R criteria in an attempt to refine the definition of delusional disorders. In DSM-III-R the distinction between schizophrenia and delusional disorders had been unclear and controversial. In DSM-III-R this boundary was defined by the nonbizarre qualities of delusions in delusional disorder and the absence of other active-phase symptoms of schizophrenia. Also important was the required absence of other odd or bizarre behavior apart from the delusion. Because the distinction between *bizarre* and *nonbizarre* is difficult to define and therefore to apply reliably, other terms such as *systematized* and *prominent* were suggested. In practice, however, those terms also have limitations. This problem has helped to promote the case for modifying the criteria in another way: specifically, to use the level of impaired functioning as a means of characterizing the distinction between schizophrenia (considerable impairment) and delusional disorders (relatively less impairment). However, the variability of outcomes in both disorders undermines this strategy somewhat. DSM-IV suggests that when poor functioning occurs in delusional disorder, it is the result of the delusional beliefs themselves. For example, a person quits a job because he or she believes that the fumes in the workplace are causing a cancerous growth. That person's financial situation worsens and preoccupation with repeated medical consultations enhances a downward spiral. In contrast, poor functioning in schizophrenia usually results from cognitive compromise and positive and negative symptoms, especially avolition. The resolution of how to make modifications, however, depends on the effectiveness of the criteria in defining homogeneous and valid subsets of psychotic disordered patients. For this purpose, field trials and data analyses have been used to inform the decision scientifically. Although the DSM-IV criteria reflect progress, their validity remains only partly established.

Another unsettled issue that DSM-IV attempts to resolve is the classification of delusional variants of somatoform disorder, specifically body dysmorphic disorder. In this condition the patient suffers from preoccupation with imagined or slight defects in appearance (such as skin blemishes, the size of one or more body parts); it is accompanied by impairment in social and occupational functioning, shame, and repetitive, often ritualistic behaviors. These may include skin picking, mirror checking, requests for reassurance, and attempts to camouflage the supposed deformity. In some cases, the preoccupation appears to be delusional. However, the relationship between nondelusional and delusional variants is unclear; whether the disorders are distinct or overlapping remains unknown. DSM-IV permits dual diagnosis of body dysmorphic disorder and delusional disorder when a delusional belief is present in the former condition. This resolution, in which the same symptoms are given two diagnoses, accurately reflects the available research data on the relationship of these two disorders and also underlines the need for further research to clarify these distinctions. A similar problem arises with respect to delusional variants of hypochondriasis and of obsessive-compulsive disorder, and a similar solution is applied: obsessive-compulsive disorder patients may also be diagnosed as delusional disorder.

Shared Psychotic Disorder DSM-IV renamed the DSM-III-R category *induced psychotic disorder*, calling it *shared psychotic disorder*. This change reflects the attempt to avoid the term *paranoid* and to identify the condition without reference to any presumed cause or mechanism. The goal is to define the boundaries between this condition and more common ones, such as other psychotic disorders, mood disorders with psychotic features, substance-induced psychotic disorders, and psychotic disorders due to a general medical condition.

ICD The ninth revision of *International Statistical Classification of Diseases* (ICD-9) contained more categories for paranoid disorder than the American schemes. Most paranoid disorders fall under the rubric *paranoid state*, including simple paranoid state, paranoia, paraphrenia, and induced psychosis. Additional subcategories include other and unspecified paranoid states. Acute paranoid reactions and psychogenic paranoid psychosis are classified separately. DSM-III, DSM-III-R, and DSM-IV generally reflect an atheoretical position with respect to the causes of these disorders whereas ICD-9 was less neutral. For example, psychogenic paranoid psychosis implies a kind of causal mechanism. The categories of paranoid disorder according to these classifications are summarized in Tables 13.2-5.

ICD-9 (1979)	DSM-III (1980)	DSM-III-R (1987)	ICD-10 (1992)	DSM-IV (1994)
Paranoid state, simple			Delusional disorder	
Paranoia	Paranoia	Delusional (paranoid) disorder	Delusional disorder	Delusional disorder
Paraphrenia (nonclassical paranoid state, late paraphrenia)			Delusional disorder	
Induced psychosis (folie à deux, induced paranoid disorder)	Shared paranoid disorder	Induced psychotic disorder	Induced delusional disorder	Shared psychotic disorder
Other specified states (paranoia querulans, Sensitiver Beziehungswahn)			Delusional disorder	
Unspecified paranoid states	Atypical paranoid disorder		Persistent delusional disorder, unspecified	
Acute paranoid reaction (bouffée délirante)	Acute paranoid disorder		Paranoid reaction	
Psychogenic paranoid psychosis (protracted reactive paranoid psychosis)				

Table 13.2-5 Comparative Nosology of Delusional Disorder

ICD-10 pays more attention to creating classifications similar to DSM-III-R and DSM-IV. Paraphrenia, for example, is subsumed under persistent delusional disorder but delusions must be present for about 3 months in order to diagnose delusional disorder. The subtypes of the disorder overlap with DSM-IV subtypes. For those conditions of less duration, acute and transient psychotic disorder is diagnosed. Induced (shared) delusional disorder is considered a separate designation with a phenomenology similar to persistent delusional disorder.

EPIDEMIOLOGY

Delusional disorder has been considered an uncommon if not rare condition from its earliest descriptions even though epidemiological information is meager. Recent demographic evidence covering a period from 1912 to the 1970s provides an estimate of incidence, prevalence, and related statistics (Table 13.2-6). However, this evidence was assembled using definitions that are not the same as those of DSM-III, DSM-III-R, or DSM-IV. Subsequent data will in all likelihood be somewhat different using the newer criteria. It is clear that the estimates are merely indications, but can be useful guidelines to future appraisals.

Incidence*	0.7–3.0
Prevalence*	24–30
Age at onset (range)	18–80 (mean 34–45 years)
Type of onset	Acute or gradual
Sex ratio	Somewhat more frequently female
Prognosis	Best with early, acute onset
Associated features	Widowhood, celibacy often present; history of substance abuse, head injury not infrequent

Portions of the table adapted from K S Kendler: Demography of paranoid psychosis (delusional disorder). Arch Gen Psychiatry 39: 890, 1982.
* Incidence and prevalence figures represent cases per 100,000 population.

Table 13.2-6 Epidemiological Features of Delusional Disorder

However, certain features of the data are remarkable. For example, the stability of estimated incidence has been striking over extended periods of time in this century. The prevalence of these disorders substantiates the widely held clinical impression that they are uncommon conditions (compared with mood disorders and schizophrenia) but are not rare. Some studies indicate that delusional disorder accounts for a surprising 2 to 8 percent of inpatient psychiatric admission for "functional psychosis." Patients with delusional disorders are somewhat more likely to be women (but this is an inconsistent feature), and to be more socially and educationally disadvantaged as compared to patients with mood disorders. Women tend to be older than men at the time of diagnosis. While the onset age range is wide (18 to 80), most patients are middle-aged. There is suggestive evidence that immigrant status, celibacy among men, and widowhood among women are associated with delusional disorder but all such observations need to be unambiguously replicated.

ETIOLOGY

The cause of delusional disorder is unknown. The epidemiological and clinical literature suggests that certain risk factors may be relevant to etiology and deserve further research elaboration. These risk factors are found in Table 13.2-7. Whether they are risk predictors or simply characteristics or markers of the disorder is unknown. Familial psychiatric disorder, including delusional disorder, is the best documented risk factor at present.

Advanced age
Sensory impairment/isolation
Family history
Social isolation
Personality features (e.g., unusual interpersonal sensitivity)
Recent immigration

Table 13.2-7 Risk Factors Associated With Delusional Disorder

Genetic or family studies that have begun to appear in the literature indicate the possible specific family transmission of delusional disorder. A recent study of genetic variation in deoxyribonucleic acid (DNA) sequence coding for dopamine type 4 (D_4) receptor proteins strongly suggests the involvement of the relevant gene in conferring susceptibility to delusional disorder. The comparison subjects either had schizophrenia or were normal controls.

Paranoid features, including the types of delusions encountered in these disorders, occur in a large and growing number of conditions (Table 13.2-2). Differences in classifying idiopathic delusional disorder add to the problems of understanding causation. Theories and explanations of delusions abound in the literature but empirical evidence to support those theories is limited. With so many uncertainties, any conclusions concerning the cause of delusional disorder must be made cautiously.

Delusional disorder is an uncommon, probably heterogeneous, group of illnesses whose validity has been questioned since Kahlbaum published his views. The major phenomenological feature of these conditions is the formation and persistence of delusions. It is well known that delusions occur in a variety of psychiatric and medical conditions, and that the pathogenesis of delusions is not fully understood. Hence, discussion of etiology in the delusional disorders can proceed along two lines: (1) determining the distinctiveness of the category itself, (2) examining the theories proposed to account for the pathogenesis of delusion formation per se, and (3) integrating the available evidence into testable proposals.

Distinctiveness of Delusional Disorder An issue that is central to attributing causation is whether delusional disorder represents a separate group of conditions or is an atypical form of schizophrenic and mood disorders. The relevant data come from a limited number of studies and is inconclusive. Epidemiological data suggest that delusional disorder is a separate condition; it is far less prevalent than schizophrenic or mood disorders; age of onset is later than in schizophrenia although men tend to experience the illness at earlier ages than women; and the sex ratio is different from that of mood disorder, which occurs disproportionately among women. Findings from family or genetic studies also support the theory that delusional disorder is a distinct entity. If delusional disorder is simply an unusual form of schizophrenic or mood disorders, the incidence of these latter conditions in family studies of delusional disorder patient probands should be higher than that of the general population. However, this has not been a consistent finding. A recent study concluded that patients with delusional disorder are more likely to have family members who show suspiciousness, jealousy, secretiveness, even paranoid illness, than families of controls. Other investigative efforts have found paranoid personality disorder and avoidant personality disorder to be more common in the relatives of patients with delusional disorder than in the relatives of controls or of schizophrenic patients. A recent study documented modest evidence for an increased risk of alcoholism among the relatives of patients with delusional disorder as compared to probands with schizophrenia, probands with psychotic disorder not otherwise specified, and probands with schizophreniform disorder.

Investigations into patient's natural history also lend support to the suggestion that delusional disorder is a distinct category: age of onset appears to be later than in

schizophrenia and outcome generally is better for delusional disorder patients than for schizophrenia patients. Although fraught with methodological shortcomings, premorbid personality data indicate that schizophrenia patients and patients with delusional disorder differ early in life. The former are more likely to be introverted, schizoid, and submissive; the latter extroverted, dominant, and hypersensitive. Delusional disorder patients may have below-average intelligence. Precipitating factors, especially related to social isolation, conflicts of conscience, and immigration, are more closely associated to delusional disorder than schizophrenia. These characteristics support Kraepelin's view that environmental factors may play an important etiological role. Clinical characteristics such as greater intensity of delusions, uncommon occurrence of negative symptoms, and possible association with cerebrovascular disorder in late-onset cases also suggest differences from late-onset schizophrenia. Recent observations of successful treatment with pimozide (Orap) in several subtypes of delusional disorders suggest the possibility of a common pathogenetic mechanism in these disorders. Follow-up studies indicate that the diagnosis of delusional disorder remains fairly stable: only a small proportion of cases (3 to 22 percent) are diagnosed later as having schizophrenia, and even fewer (6 percent) are diagnosed later as having a mood disorder. Outcome in terms of hospitalization and occupational adjustment is markedly more favorable for delusional disorder than for schizophrenia. When social or occupational functioning is poor in delusional disorder, it generally occurs as the result of the delusional beliefs themselves, not because of cognitive impairment or negative symptoms.

The evidence argues in favor of the distinctiveness of delusional disorder, but it is likely that at least some patients diagnosed as having delusional disorder will develop schizophrenia or mood disorders. Hence, current clinical criteria have limitations and need improvement, which may be possible with the use of laboratory techniques or more specified clinical definitions. Furthermore, the data suggest that delusional disorder is relatively chronic and is probably biologically distinct from other psychotic disorders.

PATHOGENESIS

Although a clear understanding of the pathogenesis of delusions remains an unfulfilled hope, several major theories have been advanced. Any adequate hypothesis for delusion formation must deal with certain facts: (1) delusions occur in a variety of medical and psychiatric diseases; (2) not all persons with such conditions develop delusions; (3) the types of delusions are relatively few and strikingly repetitious despite the variety of diseases; (4) delusions can clear rapidly with treatment of the underlying condition or its termination; (5) delusions can persist, and even become systematized; (6) delusions often accompany perceptual changes such as hallucinations or impaired sensory input; (7) delusions may be highly encapsulated features in persons such that their functioning may not be compromised socially, intellectually, or emotionally. Also, any adequate hypothesis must respond to two questions. First, why does the patient have a delusion? This is a question concerning the form of the psychopathology. Second, why does the patient have this particular delusion? This is a question concerning the content of the psychopathology.

There are three categories of theory in delusion formation.

1. Delusions arise in an otherwise intact cognitive system because a deviant pattern of motivational interest is present (psychodynamic mechanism, social attribution theory).
2. Delusions arise as the result of a fundamental cognitive defect that impairs the patient's capacity to draw valid conclusions from evidence (disorder of reasoning).
3. Delusions arise from normal cognitive processes directed at explaining abnormal perceptual experiences (psychobiological mechanism, anomalous experience hypothesis).

These theories need not be mutually exclusive. Delusional beliefs probably are the result of different processes involving one or more of the proposed mechanisms.

Psychodynamic Mechanism In 1911 Freud published “Psychoanalytic Notes Upon an Autobiographical Account of a Case of Paranoia (Dementia Paranoides).” His interpretation of this case, which became the foundation of the psychodynamic theory of paranoia, was based on his reading of the memoirs of the presiding judge of a Dresden appeals court, Daniel Paul Schreber, who had suffered episodes of psychiatric illness in 1884, 1885, and 1893. The second episode led to two prolonged hospitalizations from which the patient obtained discharge in 1902 following legal action, although he was still delusional. Freud asserted that Schreber's 1903 account, *Memoirs of My Nervous Illness*, offered a legitimate basis for theory, as “paranoiacs cannot be compelled to overcome their internal resistances, and ... in any case they only say what they choose to say ....” Freud argued that the written case report can take the place of personal acquaintance, and in the case of Schreber, Freud never saw the patient. Freud asserted that Schreber's case illustrated a general mechanism of delusion formation involving denial or contradiction and projection of repressed homosexual impulses that break out from the unconscious. The forms of delusion in paranoia can be represented as contradictions of the proposition “I (a man) love him (a man).” The following examples illustrate the forms of illogic.

1. Delusion of persecution. In the contradiction ‘I do not love him, I hate him,” a hatred that persons deem unacceptable at the conscious level is transformed and becomes instead: “He hates (elaborated to “persecutes”) me.” Patients can then rationalize their anger by consciously hating those persons whom they perceive to hate them.
2. Delusion of erotomania. The proposition “I do not love him—I love her” is transformed through projection to “She loves me—and so I love her.”
3. Delusional jealousy. To protect against unwarranted, threatening impulses the patient transforms the proposition in this manner: “I do not love him—she (a wife, lover) loves him.” Hence, jealous delusions represent the transformed attractions of the deluded for the lover.
4. Delusion of grandiosity (megalomania). Here the contradiction made is, “I do not love him, I love myself.”

The essence of the theory is that delusions represent attempts to manage the stirrings of unconscious homosexuality. According to the classic theory, the dynamics of unconscious homosexuality are similar for female as well as male patients.

Comment Many theorists have added to the psychodynamic lore on delusion formation from the standpoint of understanding personality factors. For example, some of the vulnerability to delusion formation may be related to deficiently developed trust, narcissistic dynamics, or exaggerated traits such as hypersensitivity.

Critique Freud's mechanism of delusions sidesteps the distinction between form and content in psychopathology. He proposes an inferential process to account for the particular delusion but does not clearly address the issue of why a delusion is formed rather than another symptom, such as hallucination. Verification of the hypothesized mechanism clearly rests on finding evidence that delusions are associated with indications of homosexual tendencies. The theory has been perpetuated in part because an absence of homosexuality can never be proved, and such tendencies can be used as a pillar, even if not a scientifically or empirically demonstrable pillar, in the psychodynamic argument. The few experimental attempts made to test the hypothesis have been inconclusive or equivocal. Although homosexual concerns have been found among some delusional patients, the variety of conditions with similar delusions argues against a common mechanism of unconscious homosexuality in all. Persons who delusional patients say are persecuting them are not always known by them, nor is the persistence of such delusions adequately accounted for in that formulation. Nevertheless, the classic approach has had immense influence and has provided important psychoanalytical concepts, such as projection, and an awareness that developmental experiences may operate to influence the content of delusional thinking. Systematic empirical study would be valuable.

Disordered Reasoning Because the definition of delusion (Table 13.2-1) emphasizes the operation of reasoning processes that have gone haywire, it is not surprising that a number of attempts have been made to establish that disorder of reasoning is related to delusion formation and that such disorders can be observed among deluded patients. Related to the psychodynamic formulation is the proposal that delusions arise on the basis of defects in formal logical reasoning. Popular in the 1950s and 1960s, this view, promulgated by Eilhard von Domarus among others, suggested that errors in logic such as the principle of identity (two subjects are identical on the grounds of identical predicates) have an etiological role. For example, “Charles Manson used drugs; I use drugs, therefore I am Charles Manson.” The empirical assessment of that proposal has failed to establish that deluded patients exhibit more defects in reasoning; rather it appears that normal and deluded persons both make similar errors of reasoning with comparable frequencies.

Two other proposals involving disturbance in reasoning have been studied recently. The first portrays the difficulty underlying delusion formation as a failure in the application of Bayesian reasoning. According to this model of developing beliefs, making choices, and drawing conclusions, delusional patients accept conclusions at levels of probability too low for acceptance by nondelusional persons. However, attempts to demonstrate that failure have had equivocal results. The second proposal suggests that the reasoning processes of delusional patients are influenced by the subject's tendency to assign meaning in a biased manner. The bias arises in making judgments about one's own behavior and that of another person by assigning motives and characteristics to the person involved. Application of this model reflecting motivational and reasoning difficulties (based on social attribution theory) has been tested, but the results do not provide sound support for the formulation.

Other Psychological Mechanisms In *Manic Depressive Insanity and Paranoia*, Kraepelin considered the delusions of paranoia to be the “morbidly transformed expression of the natural emotions of the human heart” and, more specifically, “a kind of psychological compensation for the disappointments of life.” He dismissed

the Freudian psychodynamic mechanism on the grounds that it did not refer to a clear concept of paranoia and that it was not supported by evidence. He also emphasized consitutional factors, especially disturbances of judgment, in his formulation. Other authors have made similar suggestions about the role of need fulfillment in the development of paranoia. For example, delusions of persecution might serve to maintain the self-esteem of the deluded person, according to a social attribution view about delusion formation in which a normal bias—that of assigning blame for negative outcomes to other persons or circumstances—is exaggerated.

Critique These contributions do not address the issue of pathogenesis rigorously. They explain the content of the delusion but not its form. The commonness of the risk factors or antecedent features cited repeatedly as central to delusion formation contrasts dramatically with the uncommonness of delusional disorder.

Psychobiological Mechanisms The French neurologist Gaëtan G. de Clerambault proposed in 1942 that chronic delusions resulted from abnormal neurological events. Infections, lesions, intoxication, and other forms of damage produce automatisms that puzzle or distress the patient initially and eventually demand explanation. The explanations take the form of delusions. Automatisms include hallucinations, a constant parade of memories, feelings of familiarity, false recognition, arresting of thought, disturbances in attention, bizarre tactile sensations, and even kinesthetic sensation. The view that delusions are an explanation for hallucinations is an old concept in psychiatry that has not been well formulated. The fact that hallucinations have been introduced into and retracted from the definition of paranoia over the years also reflects a lack of clarity regarding a possible connection between the two forms of psychopathology.

In 1974 Brendan Maher proposed a similar hypothesis that conceptualized delusions as explanations of anomalous experiences that arise in the environment, the peripheral sensory system, or the central nervous system. A central tenet of his view is that the processes whereby delusional beliefs are formed are similar in their essential nature to those that operate in the formation of normal beliefs and even of scientific hypotheses. Integral to the hypothesis is the assumption that components of this normal operational sequence have a neural substrate that may be activated either by sensory input (as in the hallucinatory effects of drugs) or by the effects of brain damage (as in alcoholism). The activation of any part of the sequence demands explanation and may thus give rise to delusions. The sequence, activated by disturbances in sensory experience, emotional incongruity, or central nervous system abnormalities, has the following stages: (1) anomalous experience, (2) feelings of significance, (3) testing for reality of experience, (4) developing tentative hypotheses, (5) additional observation, (6) exploring insights, and (7) confirmation of the insight by selective observation. In Maher's explanation, the patient is delusional because he or she actually experiences anomalies that demand explanation. The particular content of the delusion is drawn from the past or current circumstances, experience, and the personal and cultural background of the patient. The explanation answers questions such as the following: What is happening? Why? Why do other people deny it is happening? Why is it only happening to me? Who is responsible for it? The delusional explanation offers relief from puzzlement, and that relief works against abandonment of the explanation.

Critique The psychobiological formulation has gone largely unstudied, but there is supporting evidence in the form of studies of altered perception among patients and healthy controls experiencing sensory impairment or sensory deprivation, and among persons taking various drugs of abuse. These studies have demonstrated a high incidence of delusion formation. The failure to detect a fundamental defect in the cognitive process of delusional patients or to identify basic differences in belief formation between persons with delusions and normal controls provides indirect support as well. Indeed, delusions are formed in persons with a range of levels of intelligence and education, further supporting the view that a disturbance in the cognitive processes is not the source of the problem. Also a number of medical conditions show evidence of delusions but no history of cognitive impairment. Clearly, this hypothesis warrants further examination, and it remains to be seen how applicable it is to conditions, such as delusional disorder, in which the occurrence of hallucinations is debated. Sensory impairment and central nervous dysfunction, although apparently likely, have not been firmly established for the disorder. The anomalous experience hypothesis focuses on the psychological mechanisms underlying delusion formation, but a complementary proposal concerns the anatomic loci associated with delusional thinking. Jeffrey Cummings and others have used the growing data on the psychopathological consequences of neurological disease to suggest that delusions occur in diseases involving the limbic system—in particular, temporal lobe structures and caudate nuclei. Diseases characterized by excessive dopaminergic activity or reduced cholinergic activity also carry a heightened risk of delusion formation. Cummings further hypothesizes that the common locus of delusion formation is limbic dysfunction that predisposes the individual to misinterpretation of the environment accompanied by inappropriate perception of threat. Both disease- and patient-related factors influence the content, complexity, and timing of the delusion.

PATHOGENESIS

Although limited by the spareness of research in the area, observations of pharmacological treatment provide complementary insights into the pathogenetic puzzle. Data from treatment reports on delusional disorder suggest that pimozide (Orap) a highly specific dopamine-blocking agent, has greater effectiveness than typical antipsychotic drugs in this condition; some data even suggest that it has a unique role. There are several pharmacological effects of pimozide, in addition to dopamine receptor blockade, that may help explain its effectiveness: (1) relative lack of nonadrenergic blocking action (2) calcium channel antagonism, and (3) opioid receptor blockade. The effect of opiated receptor blockade has been proposed as relevant to reported specific effectiveness in delusional infestation partly based on observations of opiate receptor blocking interventions in delusional disorder somatic type, with delusions of infestation. Intravenous administration of the opioid agonist fentanyl (sublimaze) led to intensified cutaneous sensations whereas administration of naloxone (Narcan) an opioid antagonist, resulted in complete remission of the patient's cutaneous sensation. That pimozide is especially effective in delusional disorder, somatic type, supports the notion that its opiate receptor antagonism blocks central recognition of abnormal peripheral sensation; such a view is consistent with the anomalous experience hypothesis.

Other Relevant Factors Delusions have been linked to a variety of additional factors such as social and sensory isolation, socioeconomic deprivation, and personality disturbance. The deaf, the visually impaired, and possibly immigrants with limited ability in a new language may be more vulnerable to delusion formation than the normal population. Vulnerability is heightened with advanced age. Delusional disturbance and other paranoid features are common in the elderly. In short, multiple factors are associated with the formation of delusions, and the source and pathogenesis of delusional disorders per se have yet to be specified.

Integration The pathogenesis of delusions in general and delusional disorder in particular remains a field of hypotheses with little firm grounding. A variety of theories exist, but empirical support for any theory is markedly limited. Of those available, however, the anomalous experience hypothesis appears to be the best supported and certainly is the most consistent with research findings from other domains. Given the research explosion in neuroscience and psychopathology, this hypothesis should be explored as fully as possible. In delusional disorder, for example, the anomalous experience hypothesis needs to be further specified, for example, on what kinds of anomalous experience could lead to the jealousy delusion, the erotomanic delusion, and so forth. Studies now under way in the misidentification delusions, such as the Capgras's syndrome, provide a model for how such research might be directed. Progress may result from further studies of the neurobiology underlying successful treatment strategies in delusional disorder as well.

DIAGNOSIS AND CLINICAL FEATURES

Delusional Disorder

Diagnosing delusional disorders requires that the clinician match the features of the case to the appropriate criteria (Table 13.2-8). When the clinician has successfully ruled out other disorders, certain features of the case can help to substantiate the diagnosis of delusional disorder. The ICD-10 criteria for delusional disorder are listed in Table 13.2-9.

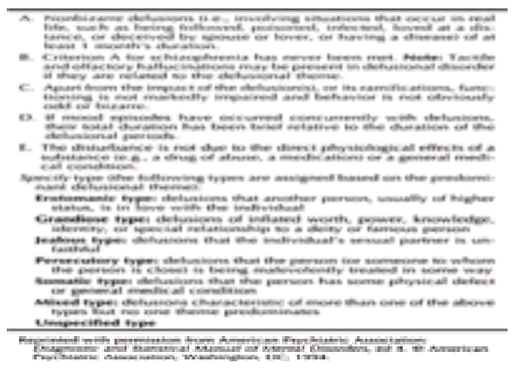

Table 13.2-8 DSM-IV Diagnostic Criteria for Delusional Disorder

Delusional disorder

A. A delusion or a set of related delusions, other than those listed as typically schizophrenic in criterion G (1) (b or d) for paranoid, hebephrenic, or catatonic schizophrenia (i.e., other than commonly impossible or culturally inappropriate), must be present. The commonest examples are persecutory, grandiose, hypochondriacal, jealous (delusional), or erotic delusions.

B. The delusional in criterion A must be present for at least 3 months.

C. The general criteria for schizophrenia are not fulfilled.

D. There must be no persistent hallucinations in any modality that there may be transitory or occasional auditory hallucinations that are not in the third person or giving a running commentary).

E. Depressive symptoms or even a depressive episode may be present intermittently, provided that the delusions persist at times when there is no disturbance of mood.

F. Most commonly used exclusion clause. There must be no evidence of primary or secondary organic mental disorder as listed under organic, including symptomatic, mental disorders, or of a psychotic disorder due to psychoactive substance use.

Specification for possible subtypes

The following types may be specified if desired: persecutory; litigious; self-referential; grandiose; hypochondriacal (somatic); jealous; erotomanic.

Other persistent delusional disorders

This is a residual category for persistent delusional disorders that do not meet the criteria for delusional disorder. Disorders in which delusions are accompanied by persistent hallucinatory voices or by schizophrenic symptoms that are insufficient to meet criteria for schizophrenia should be coded here. Delusional disorders that have lasted for less than 3 months should, however, be coded, at least temporarily, under acute and transient psychotic disorders.

Persistent delusional disorder, unspecified

Reprinted with permission from World Health Organization: The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research. © World Health Organization, Geneva, 1993.

Table 13.2-9 ICD-10 Diagnostic Criteria for Delusional Disorders

DSM-IV defines the core psychopathological feature of delusional disorder as persistent, nonbizarre delusions not explained by other psychotic disorders. Onset can be sudden, following a precipitating event that the patient often reports, or the disorder may emerge gradually and may become chronic. Behavioral and emotional responses are generally appropriate: neither a mood disorder nor the volitional, thinking, and emotional disturbances of schizophrenia are present (including hallucinations, which are quite restricted in delusional disorder). In general, patients with delusional disorder show little disorganization or impairment in their behavior or in the clarity of their thinking.

The delusions are unusual yet they refer to aspects of life that might occur, such as being conspired against, cheated on, physically ill, in love, jealous, and so forth. They are, as George Winokur has suggested, “possible,” rather than totally incredible and bizarre as are many of the delusions of schizophrenia. Delusions are categorized according to their content; the most common are characterized by persecution, disease, and jealousy. The delusions are fixed (persistent) and unarguable. Patients interpret facts to fit the delusion rather than modifying the delusion to fit the facts. There is systematization in the elusional thinking, meaning that a single theme or series of connected themes is present with links to the predominant delusion. Normal life and functioning gradually give way to the dominance of delusional concerns.

Many have proposed that there is a descriptive continuum between paranoid personality disorder, delusional disorder, and the paranoid subtype of schizophrenia in terms of degrees of disorganization and impairment. However, there is little evidence to support the concept that these disorders share more than overlapping psychopathology.

The presence of hallucinations in delusional disorder has been debated, with some theorists arguing that schizophrenia is a more likely diagnosis in such cases and others not, so long as the hallucinations are not marked and persistent. The resolution of this issue remains distant, but it is reasonable to consider infrequent, poorly organized, and simple hallucinations that are not a prominent part of the psychopathology to be a feature of delusional disorder. The hallucinations are usually auditory but may be visual and tend to be more common in acute cases. Other types of hallucinations may occur; however, tactile or olfactory hallucinations may be present and may even be prominent if they are related to the delusional theme.

The person's emotional contact and behavior are generally intact. The emotional response is usually consistent with the delusional concern, and the mood is often appropriately depressed, frustrated, or even intensely angry or elated. Restlessness and agitation may be present. Loquaciousness and circumstantiality, usually accompanying descriptions of the delusions, are found in some patients, but formal thought disorder as sometimes found in schizophrenia is absent. Persons with delusional disorder may behave in a remarkably normal way much of the time; they become strikingly different when the delusion is focused on, at which time thinking, attitude, and mood may change direction abruptly. A. Munro has called this shift in response a characteristic, possibly unique feature of delusional disorder. In the delusional mode the patient is hyperalert, preoccupied, and driven by the delusional concern. In the normal mode, the patients mood becomes calm, the conversation neutral, and the patient finds it easier to focus on other issues. The shift between modes can be difficult for lay persons to comprehend. Social and marital functioning are more likely to be compromised than intellectual and occupational functioning.

Associated features in delusional disorder include those of the paranoid syndrome (Table 13.2-3). The degree of hostility and suspiciousness may be such that violent or aggressive behavior results. Litigious behavior is common among such patients. However, some patients, notably those with somatic delusions, may not display hostility, anger, or even suspiciousness to any considerable degree.

MENTAL STATUS EXAMINATION The patient's complaints are brought to the attention of the clinician by the patient or a third party, such as police, family, neighbors, or a consulted physician or attorney. The patient may have attracted attention by asking for protection, quarreling with neighbors, visiting too many clinics, or similar behavior. The complaint focuses on the distressing behavior and possibly on incidental symptoms. The patient will not complain of a psychiatric condition; in fact, he or she will deny that or the presence of any psychiatric symptoms. Often to the surprise of those expecting to observe a range of mental deviances, examination leads to the discovery that thinking, orientation, affect, attention, memory, perception, and personality are intact. The patient's thinking is so clear and the delusional features are so central to his or her concerns that the clinician begins to anticipate the interview responses of the patient to the point that accurate predictions of specific actions and reactions are possible. Such predictability may distinguish the behavior of the patient with delusional disorder from behavior associated with other psychotic conditions. The patient's behavior and responses to the interview are consistent with the range of features in other paranoid conditions. There may be hostility, anger, lack of cooperation, and a sarcastic or challenging quality in most of what the patient says.

The capacity to act in response to delusions is an important dimension of the evaluation. Level of impulsiveness should be assessed and related to any potential for violent or suicidal behavior. The patient's self-righteousness, the intensity of the delusional experience, and its emotional impact on the patient may be clues to possible violent behavior; any plans for harming others, including homicide, should be inquired about. Suicidal behavior is an equally important concern. Impulses for self-harm arise in settings of frustration, demoralization, and even depression. If such thoughts exist, the patient should be asked how they were handled in the past. Jealousy and erotomania are perhaps especially important concerns in the assessment of possible aggression and violence. Stalking, history of abuse, and arrest records should be inquired about. Careful judgment and diplomatic interviewing are especially important in such presentations.

ASSESSMENT OF DELUSIONS The detection of delusions solidifies the judgment that a paranoid condition is present. Delusions are usually easy to detect. Features of behavior (Table 13.2-3) may suggest their presence. Associated psychopathological symptoms such as hallucinations, disturbed form of thought, and mood disorder may also indicate that delusions are part of the clinical picture.

The clinical challenge is clear in subtle cases. Fundamentally, the clinician must make a judgment based on available observations and the reported private mental experience of the patient. Attempts to dissuade the patient with counterevidence and counterarguments may be useful in determining whether the patient's beliefs can be influenced in ways that are usually sufficient to change a nondelusional person's mind. Spending time in discussion with the patient to grasp the nature of delusional thinking in terms of its themes, impact on the patient's life, complexity, systematization, and related features may be crucial in making the judgment. The most sensible guideline for all cases of suspected delusional thinking is to establish as comprehensive a picture as possible concerning the condition of the patient, including the patient's subjective private experience and evidence of psychopathological symptoms. Such information should reduce much of the uncertainty of the evaluative process.

PERSECUTORY TYPE The delusion of persecution is a classic symptom of delusional disorder; persecutory type and jealousy type delusions are probably the forms seen most frequently by psychiatrists. In contrast to persecutory delusions in schizophrenia, the clarity, logic, and systematic elaboration of the persecutory theme in delusional disorder leave a remarkable stamp on this condition. The absence of other psychopathology, of deterioration in personality, or of deterioration in most areas of functioning also contrasts with the typical manifestations of schizophrenia.

A 56-year old woman, X-ray technician who had emigrated as an adult from Europe, and married late in life, presented to the emergency room. Her complaints were that her husband's business partner of many years intended to get her husband to resign from the business and to destroy their home. Over a number of months she had become gradually aware that a variety of apparently inconsequential incidents (such as unusual cars parked on her isolated residential street, seeing individuals she knew at restaurants, and feeling as if she were being followed each time she drove her car) pointed to a conspiracy to disrupt and ultimately destroy their lives. Her delusion of persecution was remarkably systematized and detailed; her mood in describing this was tense and irritable. There was no evidence of hallucinations, confusion, thought disorder, or mood disorder. Cognition was intact. The patient was quite intelligent and saw the clinical consultation as a means of assisting her husband to deal with the distress of being targeted in such a manner. (The husband had accompanied his wife on these consultations. He also had experienced some delusional thinking in accord with hers.)

The patient showed no evidence that suggested suicidality or potential for violence toward others. She initially refused all medication but gradually over several months of therapy and parallel frequent legal consultations agreed reluctantly to take risperidone (Risperda) and later, for postpsychotic depression, paroxetine (Paxil). She responded within weeks to 0.5 to 1 mg of risperidone administered daily or on alternate days; she refused to take the medication continuously. Within a year, she began to focus on other issues and the emotional intensity of the delusional concerns diminished although they could be aroused with modest stimulation in conversation or from happenings in her home or neighborhood.

JEALOUS TYPE Delusional disorder with delusions of infidelity has been called *conjugal paranoia* when it is limited to the delusion that a spouse has been unfaithful. The eponym *Othello syndrome* has been used to describe morbid jealousy that can arise from multiple concerns. The delusion usually afflicts men, often those with no prior psychiatric illness. It may appear suddenly and serve to explain a host of present and past events involving the spouse's behavior. The condition is difficult to treat and may diminish only on separation, divorce, or death of the spouse.

Richard Krafft-Ebing described the symptom of delusional jealousy in alcoholics in 1891 and believed that extreme jealousy was pathognomonic for alcoholism. Other disorders with this symptom were later described. A recent retrospective analysis of 8134 psychiatric inpatients disclosed a prevalence of delusional jealousy of 1.1 percent among the major diagnostic groups. Among ICD-9 paranoid disorders, a 6.7 percent lifetime point prevalence was determined. Delusional disorder with alcohol dependence frequently shows the single delusion of jealousy, a persistent feature that sometimes remits if alcohol abuse is brought under control. In personality disorders the symptom may be confused with extreme jealousy, but other psychotic features should be absent. The prevalence of delusional jealousy among hospitalized patients with mood disorder was a surprisingly low 0.1 percent. A study of 26,000 psychiatric inpatients using DSM-III-R criteria yielded a 0.17 percent rate of delusional disorder, jealous type. Jealous delusions occur much more frequently in other disorders than in delusional disorder.

Marked jealousy (usually termed *pathological* or *morbid jealousy*) is thus a symptom of many disorders including schizophrenia (where female patients more commonly display this feature), epilepsy, mood disorders, drug abuse, and alcoholism—for which treatment is directed at the primary disorder. Jealousy is a powerful emotion; when it occurs in delusional disorder or as part of another condition it can be potentially dangerous and has been associated with violence, notably both suicide and homicide. The forensic aspects of the symptom have been noted repeatedly, especially its role as a motive for murder. However, physical and verbal abuse occur more frequently than extreme actions among individuals with this symptom. Caution and care in deciding how to deal with such presentations are essential not only for diagnosis, but also from the point of view of safety.

A 47-year-old carpenter was brought for psychiatric examination following complaints by neighbors about his loud yelling and verbal abuse of his girlfriend. The patient resented the psychiatric referral, but was willing to give an account of his concerns. His girlfriend, he complained, was having an affair with someone, but he was not sure who the interloper was. On his own, however, he had begun gathering evidence: strands of hair found in the apartment, photographs of soiled sheets, and suspicious items from the trash—all of which he claimed proved that an affair was ongoing. He revealed plans to tape-record, possibly videotape, his girlfriend's activities while he was on the job. Upon admitting that he had told his girlfriend that he would kill her if the affair persisted, he was admitted to the hospital. He was treated with a serotonin–dopamine antagonist in low dosages and responded with a reduction in the intensity of his rage and preoccupation. Eventually, he left the hospital, but only after his girlfriend had moved away. He still harbored suspicions but accepted the termination of the relationship without voluble opposition.

EROTOMANIC TYPE Patients with erotomania have delusions of secret lovers. Most frequently the patient is a woman, but men are also susceptible to the delusion. The patient believes that a suitor, usually more socially rominent than herself, is in love with her. The delusion becomes the central focus of the patient's existence and the onset can be sudden.

Erotomania, the *psychose passionelle*, is also referred to as *de Clérambault's syndrome* to emphasize its occurrence in different disorders. Besides being the key symptom in some cases of delusional disorder, it is known to occur in schizophrenia, mood disorder, and other organic disorders. There is no mention of erotomania in DSM-III: the condition was termed *atypical psychosis*. DSM-III-R reinstated the condition, and it remains in DSM-IV.

Patients with erotomania frequently show certain characteristics: they are generally but not exclusively women, unattractive in appearance, in low-level jobs, and they lead withdrawn, lonely lives being single and having few sexual contacts. They select secret lovers who are substantially different from themselves. They exhibit what has been called *paradoxical conduct*, the delusional phenomenon of interpreting all denials of love, no matter how clear, as secret affirmations of love. The course may be chronic, recurrent, or brief. Separation from the love object may be the only satisfactory means of intervention. Although men are less commonly afflicted by this condition than women, they may be more aggressive and possibly violent in their pursuit of love. Hence, in forensic populations men with this condition predominate. The object of aggression may not be the loved individual but companions or protectors of the love object who are viewed as trying to come between the lovers. The tendency toward violence among men with erotomania may lead initially to police rather than psychiatric contact. In certain cases resentment and rage in response to an absence of reaction from all forms of love communication may escalate to a point that the love object is in danger.

A 29-year-old male financial analyst, while having lunch in a downtown restaurant observed the arrival of a well-known local media personality, an attractive woman about his age. He experienced several moments of eye contact with the woman and became convinced that she had fallen in love with him. There ensued a barrage of flowers, letters, phone calls, and even several attempts to meet with her at her workplace. The woman rebuffed all such efforts and eventually called the police. The man was arrested on a stalking charge after he was observed following the woman to her residence. He was angry and threatening to the police, finally admitting that he had purchased a handgun but refusing to give a reason for the purchase. He was remanded to a forensic psychiatric unit, treated with pimozide, and eventually discharged on a court-supervised probation.

SOMATIC TYPE Delusional disorder with somatic delusions has been called *monosymptomatic hypochondriacal psychosis*. The condition differs from other conditions with hypochondriacal symptoms in degree of reality impairment. In delusional disorder the delusion is fixed, unarguable, and presented intensely, because the patient is totally convinced of the physical nature of the disorder. In contrast, persons with hypochondriasis often admit that their fear of illness is largely groundless. The content of the somatic delusion may vary widely from case to case. Munro has described the largest series of cases and has used the content of delusions to define three main types: (1) delusions of infestation (including parasitosis); (2) delusions of dysmorphophobia, such as of misshapenness, personal ugliness, or exaggerated size of body parts (this category seems closest to that of body dysmorphic disorder); and (3) delusions of foul body odors or halitosis. This latter category, sometimes referred to as *olfactory reference syndrome*, appears somewhat different from the category of delusions of infestation in that patients with the former have an earlier age of onset (mean 25 years), male predominance, single status, and absence of past psychiatric treatment. Otherwise the three groups, although individually low in prevalence, appear to overlap.

The frequency of these conditions is low, but they may be underdiagnosed because patients present to dermatologists, plastic surgeons, and infectious disease specialists more often than to psychiatrists in the unremitting search for curative treatment. This may partially account for Kraepelin's skepticism about the occurrence of this form of paranoia. Several recent reports indicate that pimozide (a diphenylbutyliperidine and highly specific dopamine blocker) and certain serotonin-specific reuptake inhibitors may be effective in treatment of such disorders, even in cases with a variety of delusional themes. There may be a heightened association of shared psychotic disorder involving primary cases of hypochondriacal delusion; one series reported a quarter of cases with such an association.

Patients with this condition have a poor prognosis without treatment. It affects both sexes roughly equally. A previous history or family history of psychotic disorder is uncommon. In younger patients, a history of substance abuse or head injury is frequent. Although anger and hostility are commonplace, shame, depression, and avoidant behavior are even more characteristic. Suicide, apparently motivated by anguish, is not uncommon.

A 40-year-old single unemployed man is referred by his primary care physician because of repeated consultations related to his complaint of hair loss. A dermatologist evaluated the patient, found no pathology, and told the patient that the minimal hair loss was normal. The patient refused to accept this judgment and demanded a further consultation. Because of managed-care restrictions, the patient consulted two additional specialists with his own (meager) funds with similar results. He had quit his job because of embarrassment about the hair loss and had become increasingly indebted financially. The psychiatric consultation infuriated him but he cooperated because he thought that the hair loss had begun with some “pills” he had been prescribed several years previously for anxiety and insomnia and that a psychiatrist might have something to add to understanding his case, including perhaps an antidote that might relieve the loss of hair. Treatment with an antidepressant agent proved unsatisfactory and the patient was started on an atypical antipsychotic drug with modest success. He complained less frequently about the hair loss and eventually began to express concern about his loneliness and his fear of being a burden to his aging parents, whom he lived with for financial reasons. His insight, however, remained limited and he intermittently voiced his concerns about his appearance and hair loss to his psychiatrist.

GRANDIOSE TYPE Delusions of grandeur (*megalomania*) have been noted for years. They were described in Kraepelin's paranoia and have been associated with conditions fitting the description of delusional disorder. Whether this subtype occurs in clinical practice sufficiently enough to warrant a classification is debatable.

A 51-year-old man was arrested for disturbing the peace. Police had been called to a local park to stop him from carving his initials and those of a recently formed religious cult into various trees surrounding a pond in the park. When confronted, he had scornfully argued that, having been chosen to begin a new townwide religious revival, it was necessary for him to publicize his intent in a permanent fashion. The police were unsuccessful at preventing the man from cutting another tree and made the arrest. Psychiatric examination was ordered at the state hospital, and the patient was observed there for several weeks. He denied any emotional difficulty and had never received psychiatric treatment. There was no history of euphoria or mood swings. The patient was angry about being hospitalized and only gradually permitted the doctor to interview him. In a few days, however, he was busy preaching to his fellow patients and letting them know that he had been given a special mandate from God to bring in new converts through his ability to heal. Eventually, his preoccupation with special powers diminished and no other evidence of psychopathology was observed. The patient was discharged, having received no medication at all. Two months later he was arrested at a local theater, this time for disrupting the showing of a film that depicted subjects he believed to be satanic.

MIXED TYPE The category of mixed type applies to patients with two or more delusional themes. However, this diagnosis should be reserved for cases in which no single delusional type predominates.

UNSPECIFIED TYPE The category of unspecified type is reserved for cases in which the predominant delusion cannot be subtyped within the previous categories. A possible example is certain delusions of misidentification, for example, Capgras's syndrome, named after the French psychiatrist who described the *illusion des sosies* or the illusion of doubles. The delusion in Capgras's syndrome is the belief that a familiar person has been replaced by an impostor or persons. Others have described variants of the Capgras's syndrome, namely the delusion that persecutors or familiar persons can assume the guise of strangers (*Frégoli's phenomenom*) and the very rare delusion that familiar persons could change themselves into other persons at will (*intermetamorphosis*). Each disorder is not only rare but is highly associated with schizophrenia, dementia, epilepsy, and other organic disorders. Reported cases have been predominantly in women, have had associated paranoid features, and have included feelings of depersonalization or derealization. The delusion may be shortlived, recurrent, or persistent. It is unclear whether delusional disorder can appear with such a delusion. Certainly, the Frégoli and intermetamorphosis delusions have bizarre content and are unlikely, but the delusion in Capgras's syndrome is a possible candidate for delusional disorder. The role of hallucination or perceptual disturbance in this condition needs to be explicated.

Shared Psychotic Disorder Shared psychotic disorder (also referred to over the years as *shared paranoid disorder, induced psychotic disorder, folie à deux*, and *double insanity*) was first described by Lasegue and Falret in 1877. It is probably rare, but incidence and prevalence figures are lacking and the literature consists almost entirely of case reports. The disorder is characterized by the transfer of delusions from one person to another. Both persons are closely associated for a long time and typically live together in relative social isolation. In its most common form, *folie imposée* (which is covered by the DSM-IV criteria in Table 13.2-10), the individual who first has the delusion (the primary case) is often chronically ill and typically is the influential member of a close relationship with a more suggestible person (the secondary case) who also develops the delusion. The secondary case is frequently less intelligent, more gullible, more passive, or more lacking in self-esteem than the primary case. If the pair separates, the secondary case may abandon the delusion, but this outcome is not uniformly seen. The occurrence of the delusion is attributed to the strong influence of the more dominant member. Old age, low intelligence, sensory impairment, cerebrovascular disease, and alcohol abuse are among the factors associated with this peculiar form of psychotic disorder. A genetic predisposition to idiopathic psychoses has also been suggested as a possible risk factor. The ICD-10 criteria for induced delusional disorder are given in Table 13.2-11.

A. A delusion develops in an individual in the context of a close relationship with another person(s), who has an already-established delusion.
B. The delusion is similar in content to that of the person who already has the established delusion.
C. The disturbance is not better accounted for by another psychotic disorder (e.g., schizophrenia) or a mood disorder with psychotic features and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Table 13.2-10 DSM-IV Diagnostic Criteria for Shared Psychotic Disorder

A. The individual(s) must develop a delusion or delusional system originally held by someone else with a disorder classified in schizophrenia, schizotypal disorder, persistent delusional disorder, or acute and transient psychotic disorders.
B. The individuals concerned must have an unusually close relationship with one another, and be relatively isolated from other people.
C. The individual(s) must not have held the belief in question before contact with the other person, and must not have suffered from any other disorder classified in schizophrenia, schizotypal disorder, persistent delusional disorder, or acute and transient psychotic disorders in the past.

Table 13.2-11 ICD-10 Diagnostic Criteria for Induced Delusional Disorder

Other special forms have been reported, such as *folie simultanée*, where two people become psychotic simultaneously and share the same delusion. Occasionally, more than two individuals are involved (e.g., *folie à trois, quatre, cinq*; also *folie à famille*), but such cases are especially rare. The most common relationships in *folie à deux* are sister-sister, husband-wife, and mother-child, but other combinations have also been described. Almost all cases involve members of a single family.

There is some question whether patients with such conditions are truly delusional rather than highly impressionable, as frequently there is merely passive acceptance of the delusional beliefs of the more dominant person in the relationship until they are separated, at which point the unusual belief may remit spontaneously. In the DSM-IV criteria the requirement that the secondary case not have a psychotic disorder prior to onset of the induced delusion illustrates the relevance of this question. Also, the psychopathology of secondary cases varies. In DSM-III such patients were required to meet the criteria for paranoid disorder (i.e., show evidence of disturbed personality and perhaps evidence of other psychiatric disorder, mental subnormality, or dementia); some cases may fit the definition of delusional disorder.

A 40-year-old woman consulted physicians to help cure her problem of disagreeable body odor. The physicians failed to satisfy the woman's hopes of diagnosis and treatment, because they found nothing wrong with her. They did occasionally recommend psychiatric consultation, which she refused. Her husband, a quiet, retiring man of 35, accompanied his wife to all medical specialist consultations. When questioned, he shared his wife's concerns about body odor and provided many examples of how distressing this problem had become. When he was told that there really was nothing wrong with his wife, he objected repeatedly and proclaimed that the doctors were incompetent. A psychiatrist was called to the clinic to see the couple and found consistent stories from both. The woman accepted a recommendation for hospitalization on the psychiatry-medical unit, and the husband returned home. After weeks of evaluation and treatment, the woman was discharged. The husband had stopped visiting her, and when informed that his wife would be coming home he said that he thought she had been cured of her problem. However, 3 months later the couple was once again visiting different specialists.

A 52-year-old man was referred by the court for inpatient psychiatric examination after being charged with disturbing the peace. He had been arrested for disrupting a trial, complaining of harassment by various judges. He had walked into a courtroom, marched to the bench, and begun to berate the probate judge. While in the hospital, he told in detail of conspiratorial goings-on in the local judiciary. A target of certain judges, he claimed he had been singled out for a variety of reasons for many years: he knew what was going on; he had kept records of wrongdoings; and he understood the significance of the whole matter. He refused to elaborate on the specific nature of the conspiracy. He had responded to it with frequent letters to newspapers, the local Bar association, and even to a Congressional subcommittee. His mental state, apart from his story and a mildly depressed mood, was entirely normal. A family interview revealed that his wife and several grown children had shared the belief in a judicial conspiracy directed against the patient for a number of years. There was no change in delusional thinking in the patient or the family after 10 days of observation and the patient refused follow-up treatment.

The intensity of conviction is governed by the presence of the primary case in the life of the secondary case. Protection is provided by others who share the delusion and believe that the response is reasonable. Munro has found that shared psychotic disorder is frequently associated with delusional disorder, somatic type. In the second case described for persecutory delusional disorder, the husband, a somewhat passive and isolated man, shared his wife's convictions. With her treatment, he also became less concerned about a conspiracy and began to share his doubts about the whole matter with his therapist.

A recent summary of the Japanese literature indicated that in 97 cases of *fulie à deux*, the phenomenology and epidemiology were similar to those in western reports.

PATHOLOGY AND LABORATORY EXAMINATION

Pathology As in most psychiatric conditions, there is no evidence of localized brain pathology to correlate with clinical psychopathology in patients with delusional disorder. These patients seldom die early and show no consistent abnormalities on neurological examination. Delusions can complicate many disorders and virtually all brain disorders. Certain disorders produce delusions at rates greater than that expected in the general population: for example, epilepsy (especially of the temporal lobe), degenerative dementias (dementia of the Alzheimer's type and vascular dementia), cerebrovascular disease, extrapyramidal disorders, and traumatic brain injury.

Although many types of delusions have been reported in patients with brain disorders, there appear to be particular connections between delusion phenomenology and certain kinds of brain dysfunction. For example, patients with more severe cortical impairment tend to experience more simple, transient, persecutory delusions. This type of delusional experience is characteristic of conditions such as Alzheimer's disease, dementia and metabolic encephalopathy that are also associated with significant cognitive disturbance. More complex (i.e., elaborate and systematic) delusional experiences tend to be more chronic, intensely held, resistant to treatment, and associated with neurological conditions producing less intellectual impairment and strong affective components. Those features occur in patients with neurological lesions involving the limbic system or subcortical nuclei rather than cortical areas. That, coupled with the observation of response of some patients to drug treatment, such as pimozide and other medications, provides a rational basis on which to hypothesize the presence of subcortical pathology, possibly involving systems subserving temporolimbic areas. Available evidence suggests that if there is a lesion, it will be subtle.

Imaging studies have begun to yield subtle findings about delusional disorder. In one study using quantitative volumetry in magnetic resonance imaging, 16 patients with delusional disorder showed lateral ventricle enlargement greater than in subjects with schizophrenia ($N = 31$) and almost twice that of age-matched healthy controls ($N = 35$). Although this study showed no evidence of cortical infarcts (cerebrovascular injury), other studies have suggested that unsuspected cerebral infarction may occur in a high proportion of late-onset cases with delusional disorder. A further examination of these 16 delusional disorder subjects revealed that the degree of physiological right-left asymmetry was significantly greater in the temporal lobes.

Another study has tentatively concluded that eye-tracking dysfunction in the saccadic system is present in delusional disorder, possibly reflecting some attentional impairment related to voluntary saccadic eye movement areas. Despite the subtle nature of such findings, future empirical studies, guided by etiological hypotheses, could lead to breakthroughs. Given the low incidence of delusional disorder, intensive studies of specific cases and of conditions with delusions from known causes (and with identifiable neuropathologies) offer useful beginning points. Recent studies of misidentification syndromes (e.g., Capgras's syndrome) offer the prospect of developing more refined models of neuropathological mechanisms for delusional disorder.

Laboratory Examination A range of assessments is often necessary, but several have a high likelihood of detecting key factors in the case. The use of drug-screening measures is particularly valuable given the marked delusional responses induced by a number of substances, especially alcohol, amphetamines, cocaine, and other central nervous system stimulants.

Neuropsychological assessment may help to disclose evidence of impaired intellectual functioning suggestive of brain abnormalities. The assessment of intelligence may show discrepancies between verbal and performance scores as well as scatter in overall performance. Limited data on delusional disorder (especially the more chronic forms) suggest that average or marginally low intelligence is characteristic of patients with this condition. A preliminary comparison of patients with late-onset delusional disorder and schizophrenia has indicated neuropsychological impairment to be somewhat less for the former group. Projective testing such as the Rorschach test has limited value in making the diagnosis but may confirm features consistent with it. Deviation on the paranoia scale of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) has strong correlations to paranoid features and may help substantiate the diagnosis or raise it as a possibility.

DIFFERENTIAL DIAGNOSIS

Delusional Disorder Because delusional disorders are uncommon, idiopathic, and possess features characteristic of the full range of paranoid illnesses, differential diagnosis has a clearcut logic: delusional disorder is a diagnosis of exclusion. There are many conditions to consider (Table 13.2-4), especially the more common disorders associated with paranoid features (Table 13.2-12). To avoid premature diagnosis, careful evaluation is required.

Alcohol abuse
Drug abuse (especially CNS stimulants)
Anticholinergic toxicity
Sedative-hypnotic withdrawal
Delirium
Dementia
HIV infection
Brain tumor
Epileptic disorder
Mood disorders
Schizophrenia/schizoaffective disorders

Table 13.2-12 More Common Disorders Associated With Paranoid (Delusional) Features

Clinical assessment of paranoid features requires three steps. Initially, the clinician must recognize, characterize, and judge as pathological the presence of paranoid features. Next, the clinician should determine whether they form a part of a syndrome or are isolated. Finally, a differential diagnosis should be developed. The first of

the three steps must be pursued thoroughly. The clinician must be aware that a range of objective traits or behaviors (Table 13.2-3) is often found in paranoid illness and may constitute the only clue that a paranoid illness is present. Patients with paranoid symptoms are frequently unwilling to reveal their subjective experiences to examiners or to cooperate in the clinical investigation. Careful interviewing of the patient and other informants may disclose evidence that the behavior is clearly psychopathological; in other cases, however, that conclusion must await further observations. Sometimes investigation is required to determine whether the belief is indeed delusional or not. Premature acceptance that the patient has delusional disorder has at times been an embarrassment to some clinicians who learn that the patient was not deluded. If the judgment that the patient is delusional seems unassailable, then careful elaboration of the nature of the delusion is called for. The delusional thinking should be examined for its fixity, logic, encapsulation, degree of systematization and elaboration, and its effect on planning and action.

Having determined that a paranoid condition is present, the clinician should attend to premorbid characteristics (personality, adjustment, symptom development, medical problems, and so forth), the course, and associated symptoms to detect patterns of syndromic psychopathology or isolated symptom presentations—this is step two. The discovery of clouded consciousness, perceptual disturbance, other psychopathology, physical signs, or confusing symptoms may suggest different causes for paranoid features. Isolated acute paranoid symptoms, on the other hand, often appear early in medical illness.

Finally, the clinician should resist the temptation to make the diagnosis of schizophrenia or delusional disorder prematurely in cases where paranoid features are present because these features occur regularly in a variety of psychiatric and medical illnesses. Consequently, awareness of the multiple causes of paranoid features (step one) is essential to completing the differential diagnosis (step three).

Certain principles should guide effective assessment. First, it is important to have knowledge of the paranoid features and patterns of the clinical disorders in which they occur. For example, a small percentage (perhaps 10 percent) of schizophrenia cases have their onset after age 40, and most idiopathic psychiatric conditions do not begin after age 50. Second, the premorbid status of the patient should be determined. Generally, a normal premorbid state suggests that acute paranoid features are the consequence of medical disease. Third, an abrupt change in personality, mood, ability to function, and mental state should be noted because this may indicate complications resulting from medical disease. Fourth, in those cases in which there is evidence that the patient has been refractory to psychotropic medication or psychotherapy, the continuing presence of paranoid features should alert the clinician to consider alternative diagnoses.

The final diagnosis in cases where paranoid features are prominent should be made only following: (1) a complete medical and psychiatric history with special attention (because of their high prevalence) to alcohol and other drug substance use (including drugs of abuse, prescribed drugs, and over-the-counter medication use history); (2) a thorough physical examination, including neurological and mental status examinations; (3) appropriate laboratory studies, particularly serological, toxicological, endocrine, microbiological, radiological, and electroencephalographic studies.

There are certain delusional conditions that, because of their frequency and seriousness, should be routinely considered in the differential diagnosis, as among the most likely sources of delusions (Table 13.2-13). For example, delirium, dementia, psychotic disorder due to a general medical condition, and substance-induced psychotic disorder should receive special attention. Awareness of the potential for patients with each of these disorders to present with delusions in a state of clear consciousness prior to the elaboration of the defining syndromal symptoms should be kept in mind.

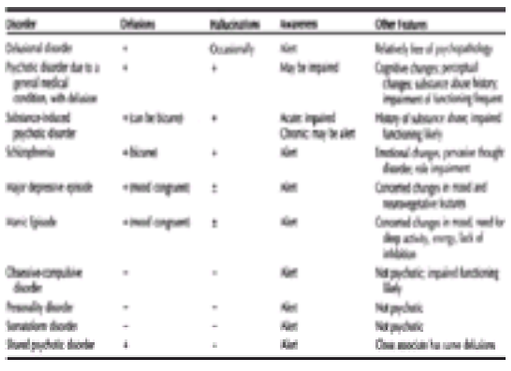

Disorder	Delusions	Hallucinations	Awareness	Other Features
Delusional disorder	+	Occasionally	Alert	Relatively free of psychopathology
Psychotic disorder due to a general medical condition, with delusions	+	+	May be impaired	Cognitive changes; perceptual changes; substance abuse history; impairment of functioning frequent
Substance-induced psychotic disorder	+ (can be bizarre)	+	Acute: impaired Chronic: may be alert	History of substance abuse; impaired functioning likely
Schizophrenia	+ Bizarre	+	Alert	Emotional changes, pervasive thought disorder, role impairment
Major depressive episode	+ (mood congruent)	±	Alert	Concurrent changes in mood and neurovegetative features
Manic Episode	+ (mood congruent)	±	Alert	Concurrent changes in mood; need for sleep; activity, energy, lack of inhibition
Obsessive-compulsive disorder	–	–	Alert	Not psychotic; impaired functioning likely
Personality disorder	–	–	Alert	Not psychotic
Somatoform disorder	–	–	Alert	Not psychotic
Shared psychotic disorder	+	–	Alert	Close association has same delusions

Table 13.2-13 Differential Diagnosis of Delusional Disorder

Psychotic Disorder Due to a General Medical Condition, With Delusions Delusions arise in a number of organic diseases and syndromes, many of which are listed in Table 13.2-2. What they frequently have in common is a disturbance of perception, particularly of visual and auditory functioning. Physical, neurological, and mental status studies as well as laboratory examinations will usually detect the organic causes of delusions. Each evaluation should focus on perceptual disturbance. Medical conditions associated with delusions should be searched for according to the guidelines outlined concerning differential diagnosis.

Substance-Induced Psychotic Disorder, With Delusions Drug intoxications are particularly relevant to this disorder. Substances of abuse, such as amphetamines, hallucinogens, phencyclidine, and cocaine; over-the-counter drugs, such as sympathomimetics; and prescribed drugs, such as steroids, methyldopa (Aldomet) and levodopa (Dopar, Larodopa) can cause psychotic disorder, with delusions, sometimes without prominent cognitive impairment. In acute states, confusion, disorientation, and clouding of consciousness may be evident; in chronic cases the picture may be more difficult to distinguish from delusional disorder because cognitive changes are less pronounced. A careful drug history and screen may establish the diagnosis. A history of alcohol abuse or dependence is so common that it should always be considered; alcoholism is often associated with jealousy, persecutory ideas, and poor impulse control.

Cognitive Disorders Dementia should be considered when paranoid features occur, particularly in older persons. Mental status examination should uncover characteristic cognitive changes absent in delusional disorder. Delirium, with its fluctuating course, confusion, memory impairment, and transient delusions, contrasts with the clarity of mental functioning and the persistence of delusions in delusional disorder and should be considered in acute cases with paranoid features.

Schizophrenia Delusions may be the presenting feature of schizophrenia and this diagnosis should be considered when the delusions are implausible or bizarre, affect is blunted or incongruous with thinking, auditory and possibly visual hallucinations are prominent, thought disorder is pervasive, or role functioning is impaired. Patients with paranoid schizophrenia may have somewhat less bizarre delusions, but role functioning is impaired; also prominent auditory hallucinations are often present, unlike in delusional disorder.

Shared Psychotic Disorder The delusions and symptoms of shared psychotic disorder may resemble those of delusional disorder; however, the delusions arise in the context of a close relationship with a delusional person, are identical in content to the delusions of that person, and diminish or disappear when secondary and primary cases are separated.

Mood Disorders With Psychotic Features The persistent and profound dysphoric mood of patients with depression often points to the proper diagnosis; in delusional disorder, affect may be intense, but is not itself an overwhelming or preoccupying experience to the patient. Delusions in depression, if present, are frequently related to mood (mood-congruent delusions) and usually indicate *severe* depression. For example, patients with feelings of worthlessness or guilt may consider that persecution against them is justified as a punishment for their evil ways. Somatic delusions may be puzzling to differentiate if the clinician fails to consider associated psychopathological features. If delusions occur exclusively during mood episodes, the diagnosis is mood disorder with psychotic features. *Depression* refers to a host of signs and symptoms, and usually has a constellation of neurovegetative features (affecting appetite, sleep, libido, energy, and so forth) that are not part of delusional disorder. Moreover, depression is frequently cyclical and is often associated with a positive family history of mood disorder. Patients with delusional disorder, in contrast, are remarkably free of symptoms other than the delusion. Chronic demoralization may result from repeated failure to obtain the kind of response desired in delusional disorder. Not infrequently, mood symptoms that meet the criteria for a mood episode are present in a delusional condition. Delusional disorder is diagnosed only if the total duration of all mood episodes remains brief relative to the total duration of the delusional disturbance. There is some evidence to suggest that depression is the most common comorbid condition in delusional disorder.

Manic Episode Manic delusions, often grandiose and therefore mood congruent, occur in the severest stages of this illness. This could mislead the diagnostician, but the cyclical nature, the marked change in mood (often euphoric or irritable at a very intense level), the reduced need for sleep, increased energy, easy distractibility, lack of focused concentration ability, lack of social inhibition, and increased activity level of manic episodes should be decisive in distinguishing that condition from

delusional disorder.

Obsessive-Compulsive Disorder Severe forms of this disorder should be considered in the differential diagnosis, especially obsessive-compulsive disorder with poor insight. Preoccupation with fear, unusual rituals, and obsessional beliefs may be puzzling, yet the pervasive effects of the condition on functioning differ from the experience of delusional disorder. Moreover, delusions and hallucinations should be absent. In practice, this differential diagnosis may be difficult to determine without a period of observation. In some cases it may be necessary to make the diagnosis and that of delusional disorder.

Somatoform Disorders Severe forms of body dysmorphic disorder may be difficult to distinguish from delusional disorder. The degree of conviction about imagined physical disfigurement may be the only guide for differential diagnosis. Lack of other features of psychopathology, often present in such cases, may also help to make the distinction.

Hypochondriasis may also be distinguished on the basis of absence of delusions, although many of the behaviors associated with delusional disorders, somatic type, may occur. Usually such patients reveal some doubt or uncertainty about the validity of their health preoccupations. Their overvalued beliefs about disease or affliction may clearly resemble delusional disorder, somatic type; severe cases may require considerable diagnostic effort, and as in obsessive compulsive disorder, also require a second diagnosis of delusional disorder.

Paranoid Personality Disorder Individuals with paranoid personality disorder by definition have abundant paranoid features. They are persistently oversensitive, ready to take offense, suspicious, resentful, rigid, and frequently self-centered. Rather than delusions, such persons tend to report strongly held ideas (overvalued ideas); generally, however, they are believed to be free of delusions, which is the most useful differential feature. There is some evidence that this personality pattern occurs often enough in families of probands with delusional disorder to suggest a possible genetic connection between the two.

Schizoid Personality Disorder and Schizotypal Personality Disorder Paranoid features may occur in patients with these personality disorders as well. The pervasive disturbance in personality functioning and the absence of delusions and other psychotic features are usually definitive distinguishing characteristics. Delusional disorder has generally not been associated with this type of premorbid pattern of personality.

Disorders of Aging Any discussion of differential diagnosis of paranoid features is incomplete unless consideration is given to the occurrence of paranoid features in the elderly. Paranoid features develop frequently in the elderly, and assessment in such cases should be particularly thorough because information about paranoid features among the aged is limited. There are several facts worth knowing: (1) the association of depressive illness with paranoid features is high enough to warrant suspicion of mood disorder in all cases with paranoid features: (2) there appears to be a late-onset syndrome sometimes labeled *late paraphrenia* or *late-onset schizophrenia* in which paranoid characteristics and hallucinosis frequently occur (this diagnosis, however, is warranted only when no other disorder can be diagnosed): (3) the sudden onset of acute paranoid features in the elderly can be a sign of cerebrovascular injury or other medical illness; (4) many of the medical conditions associated with delusions have increased incidence in the elderly population; for example, delusions can arise in the early course of presenile dementia and senile dementia conditions when deficits in clinical examination probes or neuropsychological performance may be inconspicuous; (5) perhaps most important for the general clinician is to recognize sources of increased risk of paranoid disorder among older individuals. It is now known that many factors contribute to the incidence of paranoid features in the aged, including lack of stimulating company, isolation, physical illness, the aging process itself, loss of hearing, and loss of visual acuity, each of which should be carefully assessed. Delusional disorder may be present in the elderly, may even have its onset in the elderly, but the frequency of other causes of paranoid features calls for a prudent, systematic search.

Shared Psychotic Disorder Malingering, factitious disorder with predominantly psychological signs and symptoms, psychotic disorder due to a general medical condition, and substance-induced psychotic disorder need to be considered in the differential diagnosis of shared psychotic disorder. The boundary between shared psychotic disorder and generic group madness, such as among the victims of the Jonestown massacre in Guyana, is unclear.

COURSE AND PROGNOSIS

Delusional Disorder Onset can begin in adolescence but generally occurs from middle to late adulthood with variable patterns of course, including lifelong disorder in some cases. Studies generally indicate that delusional disorder does not lead to severe impairment or change in personality, but rather to a gradual, progressive involvement with the delusional concern. Suicide has been associated with such disorders, although most patients live a normal life span. The base rate of spontaneous recovery may not be as low as previously thought, especially because only the more severely afflicted patients are referred for psychiatric treatment. Retterstol's personal follow-up investigation of a large series of cases has provided much of the viewpoint on the natural history of the disorder.

The more chronic forms of the illness (patients presenting with features for more than 6 months) tend to have their onset early in the fifth decade. Onset is acute in nearly two-thirds of the cases, and gradual in the remainder. In 53 percent the delusion has disappeared at follow-up, is improved in 10 percent, and is unchanged in 31 percent. In more acute forms of the illness the age of onset is in the fourth decade, a lasting remission occurs in over half of patients, and a pattern of chronicity develops in only 10 percent; a relapsing course has been observed in 37 percent.

Thus the more acute and earlier the onset of the illness, the more favorable the prognosis. The presence of precipitating factors signifies a positive outcome, as does female sex and being married. In terms of prognosis, the persistence of delusional thinking is most favorable for cases with persecutory delusions, and somewhat less favorable for delusions of grandeur and jealousy. However, outcome in terms of overall functioning appears somewhat more favorable for the jealousy subtype. Such patients may experience fewer hospitalizations and are less likely to have severe psychotic or schizophrenic deteriorations. Work status at follow-up has indicated that the majority of patients are employed. These observations, although limited to few cases, provide some basis for optimism: perhaps half of cases with delusional disorders may remit, but relapse and chronicity are common.

Comorbidity Depression can be diagnosed as a coexistent disorder in the course of delusional disorder. Evidence indicates that depression is an independent disorder in such cases, that is, the disorders appear to be coincidental in their combination rather than related etiologically. This judgment must be regarded as somewhat tentative, but the clinical value of recognizing comorbid (and often treatable) conditions is straightforward.

Shared Psychotic Disorder The nature of the disorder suggests that separation of the submissive person who has shared psychotic disorder (the secondary case) from the dominant person (the primary case) should result in the resolution and disappearance of the psychotic symptoms in the submissive person. Often, the submissive person requires treatment with antipsychotic drugs, just as the dominant person needs antipsychotic drugs for the psychotic disorder. Because the persons are almost always from the same family, they usually live together after being released from hospital. If separated, the patient will experience a possible remission; if not separated, the patient may have a similar prognosis as the primary case.

TREATMENT

Delusional Disorder Delusional disorder has generally been regarded as resistant to treatment and interventions have often focused on managing the morbidity of the disorder by reducing the impact of the delusion on the patient's (and family's) life. However, in recent years the outlook has become less pessimistic or restricted in planning effective treatment for these conditions. The goals of treatment are to establish the diagnosis, to decide on appropriate interventions, and to manage complications (Table 13.2-14). Fundamental to the success of these goals is an effective and therapeutic doctor-patient relationship, which is far from easy to establish. The patients do not complain about psychiatric symptoms and often enter treatment against their will; even the psychiatrist may be drawn into their delusional nets.

Rule out other causes of paranoid features
Confirm the absence of other psychopathology
Assess consequences of delusion-related behavior
Demoralization
Despondency
Anger, fear
Depression
Impact of search for "medical diagnosis," "legal solution," "proof of infidelity," etc. (i.e., financial, legal, personal, occupational, etc.)
Assess anxiety and agitation
Assess potential for violence, suicide
Assess need for hospitalization
Institute pharmacological and psychological therapies
Maintain connection through recovery

TABLE cellSpacing=0 cellPadding=0 align=left border=0 hspace="10" vspace="5">

Table 13.2-14 Diagnosis and Management of Delusional Disorder

Psychosocial Treatments There is not enough evidence to substantiate the claims for any particular school or approach in talking with the patient. Insight-oriented therapy is usually contraindicated, but a combination of supportive psychotherapeutic approaches and possibly cognitive-behavioral interventions is sensible. It is unlikely that there is any psychiatric condition that requires greater diplomacy, openness, and reliability from the therapist. Considerable skill is required to deal with the profound and intense feelings that accompany these disorders.

Awareness of the fragile self-esteem and unusual sensitivity of these patients is essential for general management and somatic treatment. Clinical experience indicates that direct questioning about the veracity of the delusion, apart from carefully establishing its nature and the evidence to support it during clinical evaluation, is seldom helpful. Although forging an alliance may be especially difficult, responding to the patient's distress rather than to the delusion itself may be effective. Understanding that fear and anxiety serve to stimulate hostility may be the key to adopting a flexible approach that promotes empathy but maintains physical and emotional distance. Patients with the disorder suffer; they often feel demoralized, miserable, isolated, and abandoned. They may face rejection at home, from police or medical specialists, or on the job. However, they can be approached, and their treatment can focus on these experiences.

The goals of supportive therapy are to allay anxiety and initiate discussion of troubling experiences and consequences of the delusion, thereby gradually to develop a collaboration with the patient. In some patients this strategy allows the psychiatrist to suggest means of coping more successfully with the delusional thinking. For example, psychiatrists might encourage patients to keep their delusions to themselves because others might feel surprised, dismayed, or amazed, all at considerable cost to the patient. It may be possible to provide educational intervention to help amenable patients to understand how factors such as sensory impairment, social and physical isolation, and stress contribute to making matters worse. In all such approaches, the overriding aim is to assist in a more satisfying general adjustment.

Cognitive approaches have attempted to reduce delusional thinking through modification of the belief itself, focusing on the associated reasoning or the reality testing of the deluded patient. Unlike noncognitive behavioral approaches that center attention on reduction of verbal behavior (talking about the delusion), this strategy seeks a more lasting and clinically meaningful intervention through multiple techniques that keep the relationship with the patient collaborative. These techniques include distancing, homework, and exploration of emotions associated with various delusions. The effectiveness of cognitive and behavioral therapies has not been studied enough to justify recommendation. Additionally, it is important to determine the long-term as well as the short-term impact of these treatments; nevertheless, they are promising enough to justify continued assessment.

Somatic Treatment Delusional disorder is a psychotic disorder by definition, and the natural presumption has been that the condition would respond to antipsychotic medication. Because controlled studies are limited and the disorder is uncommon, the results required to support this practice empirically have been only partially obtained.

The disparate findings in the recent literature on delusional disorder treatment have been summarized recently, with several qualifications. Of approximately 1000 articles published since 1961, the majority since 1980, 257 cases of delusional disorder (consistent with DSM-IV criteria) of which 209 provided sufficient treatment detail to make comparison, were assessed. Overall treatment results indicated that 80.8 percent of cases either recovered fully or partially. Pimozide (the most frequently reported treatment) produced full recovery in 68.5 percent and partial recovery in 22.4 percent of cases (*N* = 143) treated whereas there was full recovery in 22.6 percent and partial recovery in 45.3 percent of cases (*N* = 53) treated with typical neuroleptic agents [e.g., thioridazine (Mellaril), haloperidol (Haldol), chlorpromazine loxapine (Thorazine), perphenazine (Trilafon), and others]. The remaining cases (*N* = 13) were noncompliant with any treatment, a finding the authors regard as an underestimation (6.2 percent). There were no specific conclusions drawn regarding treatment with selective serotonin reuptake inhibitors, (SSRIs), although a number of such reports have been published. While treatment of the somatic subtype generated the largest number of reports, these authors' meta-analysis indicated that the patterns of response were similar across all subtypes of delusional disorder. Follow-up data and personal experience indicated that long-term, possibly permanent, administration of medication is necessary to maintain remission.

The results of treatment with the serotonin-dopamine antagonists (i.e., clozapine [Clozaril], risperidone olanzapine [Zyprexa], and others) is preliminary. Two known cases of the persecutory subtype have been treated successfully with risperidone and there are published reports of clozapine effectiveness in the persecutory subtype (*N* = 2) and the somatic subtype (*N* = 2), and of risperidone effectiveness in the somatic subtype (*N* = 1). Unfortunately, systematic case series will develop slowly, but these preliminary results suggest that the atypical neuroleptic agents may add to the available treatment options.

Given the limited samples available, case reports are especially valuable; although many authors recommend multisite trials (to augment the small numbers of cases available at any one site), it would be beneficial for further single case reports to be published in the meanwhile. The existing literature could be improved with more attention paid to diagnosis, prior treatments, outcome, and level of compliance, as well as dosage schedules, adverse effects, length of treatment, as well as the reasons for selecting or changing particular agents. Use of (*N* = 1) single case research design strategies might also enhance the generalizability of findings.

The impression is growing that antipsychotic drugs are effective, and a trial, especially with pimozide or a serotonin-dopamine antagonist is warranted. Certainly, trials of antipsychotic medication make sense when the agitation, apprehension, and anxiety that accompany delusions are prominent.

Delusional disorders respond less well generally to electroconvulsive treatment than do major mood disorders with psychotic features. Some cases may respond to SSRIs, especially cases of body dysmorphic disorder with delusional concerns (Table 13.2-15). Where differential diagnosis is unclear between delusional disorder and psychotic depression, a trial of combined therapy with antipsychotic and antidepressant medications therapy may be worthwhile. When standard strategies are unsuccessful, trials of lithium (Eskalith) or of anticonvulsant medication such as carbamazepine (Tegretol) probably should be considered. However, no systematic information to support such approaches is yet available.

Dopamine receptor antagonists (particularly pimozide)
Serotonin-dopamine antagonists
Selective serotonin reuptake inhibitors

Table 13.2-15 Pharmacological Agents With Reports of Successful Use in Delusional Disorder

Somatic treatment is difficult to implement on two levels, The patients' insistence on lack of psychiatric problems may be an insurmountable barrier to initiating treatment, and their sensitivity to all adverse effects may constitute an additional frustrating factor in their care. Noncompliance continues to be a frequent observation in published clinical studies. An open and clear approach to warn patients about and to assist them through possible unpleasant experiences is essential, but the intrinsic nature of active resistance to psychiatric intervention also requires attention. In general, some patients, especially younger patients with delusional disorder, respond to supportive management and somatic treatment. Unfortunately, others, especially the elderly, may be refractory to attempts to reduce their delusional thinking. In all cases goals that are realistic and modest are the most sensible. As most of the difficulty of this disorder results from the effects of the patient's actions concerning the delusions, any preventive approach in that domain has considerable value.

Hospitalization Most delusional disorder patients can be treated effectively in outpatient settings; hospitalization may be necessary when there is potentially dangerous behavior or unmanageable aggressiveness. The patient may show signs of poor impulse control, excessive motor and psychic tension, unremitting anger, brooding, suicidal tendencies, and even threats of self-harm or aggression toward others. Suicidal ideation and planning are also potential grounds for hospitalization. Follow-up studies report suicide above the population base rate; patients with erotomania, jealousy, and persecutory delusions are particularly at risk. Once the psychiatrist decides to hospitalize the patient, it is preferable to inform the patient tactfully that voluntary hospitalization is necessary. If this strategy fails, legal means must be undertaken to commit the patient to a hospital.

Shared Psychotic Disorder The initial step in treatment is minimally the temporary separation of the affected person from the source of the delusions, the dominant partner. This step may not only be therapeutic but diagnostic when evidence of reduced delusional thinking and preoccupation accrue. The patient may need significant support to compensate for the loss of that person. The patient with shared psychotic disorder should therefore be carefully observed for the remission of the delusional symptoms. Antipsychotic drugs can be used if the delusional symptoms have not abated in 1 or 2 weeks.

Psychotherapy with nondelusional members of the patient's family should be undertaken, and psychotherapy with both the patient with shared psychotic disorder and the dominant partner may be indicated later in the course of treatment. In addition, the mental disorder of the dominant partner should be treated. The clinician might use family therapy and social support to modify the family dynamics and to prevent the recurrence of the syndrome. It is often useful to make sure that the family unit is exposed to input from outside sources to decrease the family's isolation. In short, a comprehensive approach emphasizing support and, when necessary, medication is useful.

SUGGESTED CROSS-REFERENCES

Conditions to be differentiated from delusional disorders are discussed in Chapter 12 on schizophrenia, in Chapter 14 on mood disorders, in Chapter 16 on somatoform disorders, in Chapter 24 on paranoid personality disorder, in Chapter 5 on obsessive-compulsive disorder, and in Chapter 10 on mental disorders due to a general medical condition. Aging and psychiatric disorders in the elderly is covered in Chapter 51.

SECTION REFERENCES

Baker PB, Cook BL, Winokur G: Delusional infestation: The interface of delusions and hallucinations. Psychiatr Clin North Am *18:*345, 1995.

Bentall RP, Kaney S, Dewey ME: Paranoia and social reasoning: An attribution theory analysis. Br J Clin Psychiatry *30:*13, 1991.

Catalano M, Nobile M, Norelli E, Nothen M, Smeraldi E: Distribution of a novel notation in the first error of the human dopamine D4 receptor gene in psychotic patients. Biol Psychiatry *34:*459, 1993.

*Copeland JR, Dewey ME, Scott A, Gilmore C, Larkin BA, Cleave N, McCracken CF, McKibbin PE: Schizophrenia and delusional disorder in older age: Community prevalence, incidence, comorbidity, and outcome. Schizophr Bull *24:*153, 1998.

Cummings JL: Psychosis in neurologic disease: Neurobiology and pathogenesis. Neuropsychiatry Neuropsychol Behav Neurol *3:*144, 1992.

de Clerambault GG: *Les Psychoses Passionelles, Oeuvre Psychiatrique*, Presses Universitaires de France, Paris, 1942.

Eastham JH, Jeste DV: Treatment of schizophrenia and delusional disorder in the elderly. Eur Arch Psychiatry Clin Neurosci *247:*209, 1997.

Freud S: Psychoanalytic notes upon an autobiographical account of a case of paranoia (dementia paranoides). In *Standard Edition of the Complete Work of Sigmund Freud*, vol 12. Hogarth Press, London, 1966.

Gambini O, Colombo C, Cavallaro R, Scarone S: Smooth pursuit eye movements and saccadic eye movements in patients with delusional disorder. Am J Psychiatry *150:*1411, 1993.

Herlitz A, Forsell Y. Episodic memory deficit in elderly adults with suspected delusional disorder. Acta Psychiatr Scand *93:*355, 1996.

Howard R. Induced psychosis. Br J Hosp Med *51:*304, 1994.

Howard RJ, Almeida O, Levy R, Graves P, Graves M: Quantitative magnetic resonance imaging volumetry distinguishes delusional disorder from late-onset schizophrenia. Br J Psychiatry *165:*474, 1995.

Howard RJ, Graham C, Sham P, Dennehey J, Castle DJ, Levy R, Murray R. A controlled family study of late-onset non-affective psychosis (late paraphrenia). Br J Psychiatry *170:*511, 1997.

Jaspers K: *General Psychopathology*. University of Manchester Press, Manchester, 1963.

Kashiwase H, Kato M: *Folie a deux* in Japan-analysis of 97 cases in the Japanese literature. Acta Psychiatr Scand *96:*231, 1997.

Kendler KS: Demography of paranoid psychosis (delusional disorder). Arch Gen Psychiatry *39:*890, 1982.

Kendler KS, Walsh D: Schizophreniform disorder, delusional disorder and psychotic disorder not otherwise specified: Clinical features, outcome, and familial psychopathology. Acta Psychiatr Scand *91:*370, 1995.

Kitamura H. A case of somatic delusional disorder that responded to treatment with risperidone. Psychiatry Clin Neurosci *51:*337, 1997.

Kolle K: *Der Wahnkranke in Lichte alter and neuer Psychopathologie*, Thieme, Stuttgart, 1957.

Kraepelin E: *Dementia Praecox and Paraphrenia*, RN Barclay, translator Livingstone, Edinburgh, 1989.

Lo Y, Tsai SJ, Chang CH, Hwang JP, Sim CB: Organic delusional disorder in psychiatric in-patients: comparison with delusional disorder. Acta Psychiatr Scand *95:*161, 1997.

*Maher BA: Delusions: Contemporary etiological hypotheses. Psychiatr Ann *22:*260, 1992.

Maher BA, Spitzer M: Delusions. In *Comprehensive Handbook of Psychopathology*, ed 2, HE Adams, PB Sutker, editors. Plenum, New York, 1993.

*Manschreck TC: Pathogenesis of delusions. Psychiatr Clin North Am *18:*213, 1995.

Manschreck TC: The assessment of paranoid features. Compr Psychiatry *20:*370, 1979.

Manschreck TC, Petri M: The paranoid syndrome. Lancet *2:*251, 1978.

*Manschreck TC: Delusional disorder: The recognition and management of paranoia. J Clin Psychiatry *57*(Suppl):32, 1996.

*McAllister TW: Neuropsychiatric aspects of delusions. Psychiatr Ann *22:*269, 1992.

*Meloy JR, editor: *The Psychology of Stalking*. Academic Press, San Diego, CA, 1998.

Menzies RPD, Federoff JP, Green CM, Isaacson K: Prediction of dangerous behavior in male erotomania. Br J Psychiatry *166:*529, 1995.

Mowat RR: *Morbid Jealousy and Murder*. Tavistock, London, 1966.

*Munro A, Mok H: An overview of treatment in paranoia/delusional disorder. Can J Psychiatry *40:*616, 1995.

*Munro A: *Delusional Disorder*. Cambridge University Press, New York, 1999.

Opler LA, Klahr DM, Ramirez PM: Pharmacologic treatment of delusions. Psychiatr Clin North Am *18:*379, 1995.

Retterstol N: *Paranoid and Paranoiac Psychoses*. Charles C Thomas, Springfield, IL, 1966.

Segal JH: Erotomania revisited: From Kraepelin to DSM-III-R. Am J Psychiatry *146:*1261, 1989.

Schreber D: *Memoirs of My Nervous Illness*. Bentley R, Cambridge, MA, 1995.

*Serretti A, Lattuada E, Cusin C, Smeraldi E: Factor analysis of delusional disorder symptomatology. Compr Psychiatry *40:*143, 1999.

Soares JC, Gershon S. Therapeutic targets in late-life psychoses: Review of concepts and critical issues. Schizophr Res *27:*227, 1997.

Spitzer M: On defining delusions. Compr Psychiatry *31:*377, 1990.

Spitzer RL, First MB, Kendler KS, Stein DJ: The reliability of three definitions of bizarre delusions. Am J Psychiatry *150:*880, 1993.

Winokur G: Familial psychopathology in delusional disorder. Compr Psychiatry *26:*241, 1985.

13.3 ACUTE AND TRANSIENT PSYCHOTIC DISORDERS AND CULTURE-BOUND SYNDROMES

JUAN E. MEZZICH, M.D., FH.D., KEH-MING, LIN, M.D., M.P.H., AND CHARLES CAMPBELL HUGHES, IH.D.

While schizophrenia and mood disorders have commanded attention for many years and are likely to continue doing so for the foreseeable future, several other types of psychotic conditions are emerging as significant. Among these are psychotic disorders associated with general medical diseases and with the use of psychoactive substances as well as a complex group of acute and brief psychotic disorders.

Interest in the latter group stems from several factors. One concerns the phenomenological intricacy of disorders in this group. While all basically share loss of touch with reality or bizarre behavior as core psychopathology, they may diverge extensively in other aspects of their symptomatological profile. Another factor involves geographical epidemiology, with a disproportionately high frequency of acute and brief psychoses reported in the developing countries of the Americas, Asia, and Africa. A third factor involves cultural framework. Most of the syndromic or nosological predecessors of the acute and brief psychoses have been described in defined cultural contexts, whether as *bouffée délirante* in France, psychogenic psychosis in Scandinavia, cycloid psychosis in Germany, or the variety of transient psychoses reported in the traditional societies of the developing world.

The crucial role of the last factor connects the nosological category of acute transient psychoses to the culture-bound syndromes. The importance of cultural framework is certainly relevant to the full range of psychiatric disorders, both from clinical psychopathological and epidemiological perspectives, but it is most distinctive and illustrative in reference to culture-bound syndromes of both psychotic and nonpsychotic types.

DEFINITION AND COMPARATIVE NOSOLOGY

ICD-10 The acute psychoses described in northern European countries and in developing countries have been, for the first time, accommodated and organized in the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), under the category of acute and transient psychotic disorders. The conditions are formulated and arranged according to the following principles, in order of priority:

1. An acute onset (less than 2 weeks) as the key criterion for the whole group. Acute onset denotes a change within 2 weeks or less from a state without psychotic features to a clearly abnormal psychotic state (not necessarily at its peak severity).
2. The presence of typical syndromes. Those include, first, a rapidly changing and variable state called polymorphic, prominent in acute psychoses described in several countries, and, second, the presence of typical schizophrenic symptoms.
3. The presence or absence of associated acute stress (within 2 weeks of the first psychotic symptoms).

P>Complete recovery usually occurs within 1 to 3 months (depending on the specific disorder), often within a few weeks or days. Only a small proportion of patients with these conditions develop persistently disabling states.

DSM-IV The evaluation of a psychotic patient requires consideration of the possibility that the psychotic symptoms result from a general medical condition (e.g., a brain tumor) or the ingestion of a substance (e.g., phencyclidine). Those two situations are classified in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as psychotic disorder due to a general medical condition and substance-induced psychotic disorder, respectively. DSM-IV also includes a diagnosis of catatonic disorder due to a general medical condition to emphasize the special considerations regarding the differential diagnosis of catatonic symptoms.

DSM-IV also includes psychotic disorder not otherwise specified for psychotic disorders that do not meet the criteria for any other specific psychotic disorder. In previous editions of DSM these were called atypical psychoses.

ETIOLOGY

Both psychosocial and biological factors appear to play significant roles in the causation of acute brief psychoses and culture-bound syndromes. More than one factor may be present in any particular case, including a combination of psychosocial and biological factors.

Psychosocial Factors The diagnosis of psychotic disorders depends primarily on an accurate and thoughtful assessment of delusions, hallucinations, and bizarre psychomotor behaviors. Culture profoundly influences the meaning and nature of symptoms in the psychotic disorders reviewed in this chapter as well as the characteristics and meaning of the context and the consequences of the behaviors at hand. Attention to cultural framework can help one interpret behaviors properly and minimize misdiagnosis and diagnostic ambiguity in cross-cultural clinical situations. Lacking adequate information on what constitutes normal behavior patterns or culturally sanctioned idioms of distress, clinicians evaluating patients with different cultural, ethnic, or religious backgrounds are likely to misidentify less severe complaints or behaviors as delusional, hallucinatory, or bizarre. Similarly, they are likely to suspect the existence of major psychopathology in patients with fleeting psychotic manifestations.

Spiritual and religious beliefs can present major diagnostic dilemmas for clinicians. Beliefs in witchcraft and sorcery are common in many societies and may or may not be delusional. Spiritism, Santeria, and various other religious movements and different forms of shamanism practiced in many parts of the world encourage and sanction personal communication and active involvement with the dead, with spirits, and with the various deities. Such supernatural and mystical practices and experiences do not necessarily indicate psychopathology. However, such culturally congruent beliefs often also exert substantial pathoplastic influences on symptom formation in psychotic patients. Similarly, possession and trance phenomena are frequently seen in most non-Western societies, and it is often difficult to determine whether those experiences, in a particular case, are part of an ongoing psychotic process or are culturally and contextually appropriate.

Among the social factors that have been the subject of empirical research in recent years is (particularly among women) departure from the parental village for a number of reasons (e.g., marriage) and return to the parental village, including participation in events such as births and weddings. Investigators have also found a history of job stress (particularly among men) associated with the emergence of clinical manifestations.

Other social factors can also significantly influence symptom formation in psychiatric patients, thereby complicating the diagnosis of psychotic conditions. Sustained exposure to racist and discriminatory behaviors tends to increase vigilance and suspiciousness among members of ethnic minorities, and it may contribute to a higher propensity for paranoid symptoms in such persons. Paranoid symptoms are also more prevalent among those, such as refugees, who are forced to live in an unfamiliar cultural milieu. Fear of political persecution is a reality of life for persons living under oppressive regimes, and it may contribute to a higher prevalence of paranoid ideation in such societies. Because of those complications, it is often difficult to determine whether paranoid experiences among recent immigrants and sojourners are reactive or indicate a more serious and enduring psychotic process.

Biological Factors A variety of biological factors can have etiological involvement in the development of acute grief psychoses and culture-bound syndromes. One is infectious diseases, especially prevalent in developing countries. Empirical research documents the presence of fever in a disproportionately high number of patients with acute transient psychotic disorders.

Physical conditions (e.g., cerebral neoplasms, particularly of the occipital or temporal areas) can induce hallucinations. Sensory deprivation, as occurs in blind and

deaf persons, can also result in hallucinatory or delusional experiences. Lesions involving the temporal lobe and other cerebral regions, especially the right hemisphere and the parietal lobe, are often associated with delusions.

Psychoactive substances are common causes of psychotic syndromes. The most commonly involved substances are alcohol, indole hallucinogens (e.g., lysergic acid diethylamide [LSD], amphetamines, cocaine, mescaline, phencyclidine [PCP], and ketamine. Many other substances, including steroids and levothyroxine (Levoxyr, synthroid) can be associated with substance-induced hallucinations.

DIAGNOSIS AND CLINICAL FEATURES

ICD-10 ICD-10 provides general criteria for acute and transient psychotic disorders as well as criteria for four specific disorders (Table 13.3-1). ICD-10 also includes two residual categories for acute and transient psychotic disorders.

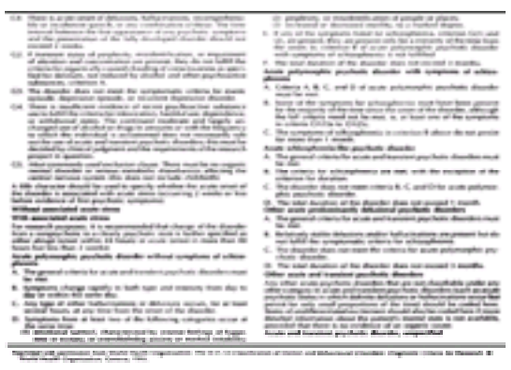

Table 13.3-1 ICD-10 Diagnostic Criteria for Acute and Transient Psychotic Disorders

Acute Polymorphic Psychotic Disorder Without Symptoms of Schizophrenia Acute polymorphic psychotic disorder without symptoms of schizophrenia is characterized by obvious but variable, rapidly changing hallucinations, delusions, and perceptual disturbances, often accompanied by emotional turmoil (happiness and ecstasy or anxiety and irritability). The criteria for manic episode, depressive episode, or schizophrenia are not met. The disorder tends to have an abrupt onset (less than 48 hours) and then a rapid resolution of symptoms. If symptoms persist for more than 3 months, the diagnosis should be changed (e.g., to persistent delusional disorder or some other nonorganic psychotic disorder). The diagnosis accommodates *bouffée délirante* and cycloid psychosis, both either unspecified or without symptoms of schizophrenia, as the following excerpt from the *ICD-10 Casebook* illustrates.

Mrs. Charrière is a 25-year-old Frenchwoman.

Problem Mrs. Charrière was brought by ambulance to a hospital emergency department in the city where she lived. Her husband reported that she had been perfectly normal until the previous evening, when she had come home from work complaining that “strange things were going on” at her office. She had noticed that her colleagues were talking about her, that they had been quite different all of a sudden, and that they had started behaving as if they were acting a part. Mrs. Charrière was convinced that she had been put under surveillance and that someone was listening in on her telephone conversations. All day she had been feeling as if she were in a dream. When she looked in the mirror, she had seemed unreal to herself. She had become increasingly anxious, incoherent, and agitated during the course of the day and had not been able to sleep at all during the night. She had spent most of the night looking out of the window. Several times she pointed at the crows in a nearby tree and told her husband, “The birds are coming.”

In the morning, Mr. Charrière found his wife on her knees as if she were praying. She knocked her head repeatedly against the floor and talked in a rambling way, declaring that she had been entrusted with a special mission, that her boss was a criminal, there were spies everywhere, and something terrible would happen soon. All of a sudden she calmed down, smiled at her husband, and told him that she had decided to convert from Catholicism to Islam. At that stage she became quite elated, started laughing and shouting, and declared that she and her husband could pray to the same god from then on. Shortly afterward she was terrified again and accused her husband of trying to poison her.

History Mrs. Charrière was brought up in a town in the west of France, where her parents owned a small restaurant. She did well in school, went on to college, and trained as an interpreter. During her training she met her future husband, who had come to France from Algeria to train as an interpreter himself. Because both she and her husband were agnostics, the fact that they came from different religious backgrounds had never been a problem. She took a job with an administration related to the European Communities, and her husband found a position with an international interpreting company. The couple were doing well, they had bought a nice house on the outskirts of Mrs. Charrière's home town, and were planning to have a child in the near future.

Mrs. Charrière's parents were in good health. She had a brother and two sisters. At age 18 her younger sister had had a nervous breakdown and in the ensuing years had been hospitalized repeatedly in a psychiatric hospital with a diagnosis of schizophrenia.

Both Mrs. Charrière and her husband refrained from drinking alcohol and were strongly opposed to any kind of drugs, including prescription medicines.

Mr. Charrière described his wife as an outgoing, sociable, and perfectly normal woman. However, he was quite worried about what was happening, all the more since she appeared to have symptoms resembling those he had observed in his sister-in-law.

Findings On admission Mrs. Charrière was frightened and bewildered but was oriented in time, place, and person. She was restless and constantly changed position, standing and sitting, moving about the room, shouting and screaming, weeping and laughing. She talked in a rambling way, shifting from one subject to another without any transition. Something criminal was going on at her office, she said, and she had discovered a secret plot. There were microphones hidden everywhere, she added, and “the birds are coming.” She wondered whether the physician was a real physician or “a spy in disguise.” She went on to speak about “my mission,” declared that Jesus had been a false prophet, that Muhammad was the real prophet, and that she would convince the world of what was right and wrong. She then began to explain that the truth was to be found in numbers. The digit “3” signifies good, she said, and the digit “8” represents evil. Suddenly she started to weep, explaining that her parents had died and that she wished to join them in heaven.

During the first days of hospitalization, Mrs. Charrière continued presenting a rapidly changing symptomatology. Her mood frequently shifted from sadness to elation, and the content of her delusions changed from persecution to mysticism. On several occasions she came out of her room and complained that she had heard people speaking about her, even when there was no one in the vicinity. When asked to describe what she was hearing, she spoke of voices coming from the corridor. She firmly denied that the voices might emanate from within her own body.

The physical examination did not reveal any abnormality. Results of blood tests, including thyroid function, were within normal limits, as were all other special investigations such as an electroencephalogram and brain scan.

Course Mrs. Charrière was treated with 30 mg of haloperidol (Haldol) during the first week, and with half this dose for the following week. After 2 weeks all of her symptoms had disappeared, and she was discharged on medication. She was seen once a week in the outpatient department for another month, during which the medication was progressively reduced and then stopped completely. Two months after the onset of the delusional episode, the patient continued to be free of symptoms.

Discussion The significant features of Mrs. Charrière's disorder were acute polymorphous delusions, rapidly changing mood disturbances, perplexity, depersonalization, and derealization without clouding of consciousness, and occasional auditory hallucinations. The disorder developed to its peak in 24 hours and was resolved in a few weeks, with complete recovery within 6 weeks. The patient had no psychiatric history.

Discussion The significant features of Mrs. Charrière's disorder were acute polymorphous delusions, rapidly changing mood disturbances, perplexity, depersonalization, and derealization without clouding of consciousness, and occasional auditory hallucinations. The disorder developed to its peak in 24 hours and was resolved in a few weeks, with complete recovery within 6 weeks. The patient had no psychiatric history.

The psychiatrist who dealt with this case made a diagnosis of *bouffée délirante*. This concept goes back to the French psychiatrist Valentin Magnan, whose pupil Paul Legrain, proposed the following diagnostic criteria: an acute onset of the disorder “like a bolt from the blue” in the absence of a psychosocial stressor; the presence of unsystematized and rapidly changing “polymorphic” delusions; the presence of emotional turmoil with intense and changing feelings of anxiety, happiness, or sadness; the presence of perplexity, depersonalization, or derealization without clouding of consciousness; and resolution of the disorder with complete recovery within 2 months.

In the ICD-10, the subtyping of acute and transient psychotic disorders rests on the acuteness of onset, the presence of typical syndromes, and the presence of associated stress. In the case of Mrs. Charrière, the onset was abrupt (i.e., the symptoms appeared within less than 48 hours), the syndrome was polymorphic, there were no typically schizophrenic symptoms, and the onset of the disorder was not associated with acute stress. Therefore, Mrs. Charrière's disorder must be coded as acute polymorphic psychotic disorder, without symptoms of schizophrenia, and without associated acute stress. (Reprinted with permission from *ICD-10 Casebook*.)

Acute Polymorphic Psychotic Disorder With Symptoms of Schizophrenia Acute polymorphic psychotic disorder with symptoms of schizophrenia is as polymorphic as the preceding disorders, but is additionally characterized by the consistent presence of typical schizophrenic symptoms. If the schizophrenic symptoms last more than 1 month, the diagnosis should be changed to schizophrenia. The disorder accomodates the concepts of *bouffée délirante* and cycloid psychosis, both with symptoms of schizophrenia.

Acute Schizophrenia-Like Psychotic Disorder Acute schizophrenia-like psychotic disorder is characterized by the consistent and stable presence of typical schizophrenic symptoms, without the polymorphic character of the foregoing disorders. If the schizophrenic symptoms last more than 1 month, the diagnosis should be changed to schizophrenia.

Other Acute, Predominantly Delusional Psychotic Disorders The other disorders are characterized by relatively stable delusions or hallucinations, without fulfilling the criteria for either schizophrenia or the acute polymorphic psychotic disorders. If the delusions persist for more than 3 months, the diagnosis should be changed to persistent delusional disorder and, if only the hallucinations persist, to other nonorganic psychotic disorder. The disorder accommodates the concepts of psychogenic paranoid psychosis and paranoid reaction.

Halime is a 22-year-old student. She is in her first year of medical school in Egypt.

Problem Accompanied by her mother, Halime came to see the physician at the psychiatric outpatient clinic. She was complaining about her nose. For the preceding week, from time to time she could smell a foul odor and was very afraid that it came from herself. She reported hearing voices talking about her behavior and telling her what to do. She had become extremely irritable and was unable to sleep. All these problems began 10 days after returning to her home in Alexandria on summer vacation from the medical school. She could find no obvious reason for the foul smell and the voices, but she thought the condition might be the result of witchcraft. She had developed a friendship with a young man, a fellow student at the medical school, and suddenly, just before the vacation, he had asked her to marry him. She was very surprised, became frightened, and had refused, which had upset him. She now suspected that her boyfriend had caused a spell to be cast on her because of her refusal. Halime's family said her condition was gradually getting worse. The foul smell and the voices seemed to affect her more each day.

History Halime was the first of two children born to a family of average income in Alexandria. Her father was a mechanic and seemed to be a rather shy and gentle person. The mother was ambitious and expressed great concern for her daughter's education. Her family reported that they had great hope that their first child would be a boy. The parents treated Halime as if she were a boy for the first 3 years of her life until the birth of a second child, which was a son. There was no information about mental disorders in the family.

Halime was introverted, thoughtful, and rather stubborn, with only a few friends. She had a high moral standard and had never dated or had sexual relations. She was doing very well at medical school and was determined to become a great physician in a culture where there is still some resistance to women physicians.

She had always been physically strong and had excellent health.

Findings On admission to the clinic, Halime was found to be very self-conscious and tried to avoid being seen by other people. She appeared tense and sad and seemed close to tears. She was reluctant and mentioned with hesitation some “extraordinary experiences.” These included the foul smell, which was like burned meat, and the voices that kept commenting on her behavior. She said the voices described what she was doing “here and now” and added comments. One example she gave of what the voices were saying was, “You are now speaking to the physician. You hope that he can help you, don't you? No hope. We shall overcome.” She explained that she believed this was the result of the witchcraft to which she had been subjected. She seemed able to differentiate in her mind between normal and abnormal perceptual experiences. She stated that she was able to make this distinction but was unable to do anything about the voices.

She showed emotional response of normal modulation, and no abnormalities of speech were observed. She was fully oriented as to time, place, and person and showed no impairment of memory. Her attention seemed sharpened, but her concentration was slightly diminished.

Careful physical and neurological examinations revealed no abnormality. An electroencephalogram with nasopharyngeal electrodes and a computed tomography scan also showed only normal results, and laboratory investigations, including thyroid parameters, were all normal.

Course Halime was prescribed haloperidol (3 mg per day) and a hypnotic. In the course of 4 days, the voices and the foul smell gradually disappeared. At the next visit to the clinic a week later, she complained of drowsiness and fatigue, aching and stiffness in the muscles, and difficulties with concentration. Her haloperidol was reduced to 1 mg per day, and the hypnotic was discontinued. After this she gradually improved, and after an additional 2 weeks she appeared well and was able to manage without medication.

Discussion After a sudden proposal of marriage less than 2 weeks before, Halime developed within just a few days a psychotic disorder with olfactory and auditory hallucinations and with commenting voices mentioning her in the third person. Her explanation about witchcraft cannot for certain be considered delusional in a culture where there is widespread belief in this phenomenon, although someone of her educational level would be expected to consider such belief as a superstition. Otherwise, Halime showed remarkable insight into the nature of her condition, with little or no impairment of her sense of reality. She showed no major disturbance of consciousness. No sign of organic etiology was observed, and psychoactive substance use was not suspected. She did not meet the symptomatic criteria for an affective episode.

On antipsychotic treatment she had a complete remission within 3 weeks. Thus, she had an acute and transient psychotic disorder. The symptomatology was not polymorphic but included a schizophrenic first-rank symptom of commenting voices. The subtype, therefore, will be schizophrenia-like disorder. The psychotic disorder followed, within 2 weeks, an event that may be considered stressful to a young female in Halime's cultural setting.

The full diagnosis therefore will be other acute predominantly delusional psychotic disorder, without associated acute stress. (Reprinted with permission from *ICD-10 Casebook*.)

Mr. Dubois is a 43-year-old Frenchman.

Problem Mr. Dubois was convinced that he was being watched and that someone was listening in on his telephone conversations. He was referred for psychiatric consultation by his family physician because he became increasingly anxious and felt that “strange things” were going on around him.

The problems had begun a week earlier, when Mr. Dubois started having doubts about whether his father, who had died more than 5 years before, was really dead or was still alive and being held hostage by the local municipal council, of which he had been a member until shortly before his death. After all, Mr. Dubois thought, his father had been the most honest man in the world, and it was likely that he had vehemently opposed a shady deal that the council had decided to pursue despite his misgivings.

Mr. Dubois had the impression that people looked at him “in a knowing way” or talked or behaved as if they wanted to convey secret messages. On several occasions, when he was alone, he heard voices whispering something he was not able to understand and noticed that normally static objects seemed to move before his eyes.

History Mr. Dubois came from a small industrial town. His father worked in the local steel factory and was very active in the labor union. The son was a bright student. He went to college and trained as a primary school teacher. He took up a teaching position in his home town, got married, had two children, and led a quiet and rather uneventful life.

The patient's father died of a heart attack at the age of 65. Although he had already retired from his job at that time, he had remained active in local politics. He had been a member of the municipal council until a few months before his death and had left the council only because he was compelled to resign as a result of some disagreements with his colleagues. He had felt very bitter about what he considered an unfair ousting, became morose, and hardly went into town afterward. In the view of his family, the forced resignation from the council contributed greatly to his death.

Mr. Dubois's mother was alive and well, and he had a brother who lived abroad. There was no history of mental illness in the family.

The patient felt quite depressed after his father's death but did not have any other psychiatric symptoms before his current episode. His physical health had always been excellent. He had never smoked, did not take drugs, and drank no alcoholic beverages except for an occasional beer or glass of wine.

Mrs. Dubois described her husband as hard-working, conscientious, and somewhat rigid in his opinions and with a tendency to bear grudges.

Findings On arrival at the psychiatrist's consulting room, Mr. Dubois was anxious, distressed, and bewildered but oriented in time, place, and person. He talked in a coherent way and, although initially suspicious, eventually gave a detailed account of what he had experienced during the previous week. He was convinced that there was a conspiracy against his family and that it had to do with his father's political activities. He remembered his father's funeral, but he was convinced that the coffin must have been empty. In his opinion the conspirators had kidnapped his father and had been holding him as a hostage all these years. Now they had decided to destroy the rest of the family. The patient admitted that he had been under stress for some time, that he had too much work, and that he had let himself get involved in too many activities both at school and outside of school. During the previous week he had not been able to sleep for more than a few hours each night, and he felt exhausted. He did not feel depressed, and in particular he had no feelings of self-reproach or guilt and no thoughts of suicide. He was aware, however, that something was wrong and agreed to stay in the hospital.

The physical examination did not reveal any abnormality. The results of blood tests, including thyroid function, were within normal limits, as were all other special investigations such as an electroencephalogram and brain scan.

Course Mr. Dubois was treated with a haloperidol (30 mg) during the first week, with the dose halved for the second week. After 2 weeks all of the symptoms had disappeared, and Mr. Dubois was discharged on medication. He was seen once a week in the outpatient department for another month, during which the medication was progressively reduced and then stopped completely. Six months after the onset of his delusional episode, the patient continued to be free of symptoms.

Discussion The significant features of Mr. Dubois's disorder were acute delusions, together with occasional auditory hallucinations and disturbance of visual perception. The disorder developed to its peak in a few days and was resolved in a few weeks, with complete recovery occurring within 6 weeks. The patient had no psychiatric history.

The patient was convinced that his father, who had died 5 years before, was in fact alive and held hostage by a group of conspirators. Although extremely improbable, his delusions cannot be considered theoretically and physically completely impossible and thus bizarre (as in a schizophrenia-like disorder).

Mr. Dubois's disorder otherwise meets the general criteria of an acute and transient psychotic disorder. It developed in a few days and could not be attributed to an organic mental disorder or a metabolic disturbance affecting the central nervous system, or to a mood disorder or the use of a psychoactive substance.

The symptoms did not change rapidly in both type and intensity, so they cannot be coded as acute polymorphic disorder. Because Mr. Dubois does not fulfill the symptomatic criteria for schizophrenia, the most probable diagnosis is other acute predominantly delusional psychotic disorder. The onset of the disorder was preceded by a period of overwork, but this did not amount to significant stress.

Other Acute and Transient Psychotic Disorders Other acute psychotic disorders not classifiable under the preceding categories are included in this category provided there is no evidence of an organic cause. Examples include acute psychoses with definite but fleeting delusions or hallucinations, and states of undifferentiated excitement.

Miss Maruyami is Indonesian. She is 30 years old and single and lives with her widowed mother.

Problem Miss Maruyami was brought to the psychiatric hospital by an uncle and two brothers-in-law, who had to hold her tightly to prevent her from running away. They almost had to carry her. She was hissing, spitting, and kicking about with her legs so that she was hardly able to walk.

Four days earlier she had returned from the market in a state of agitation and mild confusion. She claimed that one of the merchants had accused her of shoplifting, and, when she protested, he had further accused her of being an idler and a prostitute who ought to be sent to jail. Since then she had appeared restless, irritable, suspicious, and hardly able to sleep. At night she went about the house checking the locks and peeping through the windows at the neighbor's house in which she could see some shadows and lights move about. She felt that something hostile was going on and that the neighbor was spying on her because he wanted to rape her and kill her. She did not dare leave the house, and she refused to take any food or water; she claimed it had an odd taste because it had been poisoned by the neighbor. She also tried to prevent her mother from eating and drinking, and she threw away the food and vegetables. She became increasingly disturbed, sat staring wildly ahead of her, and hardly responded when her mother talked to her. When a sister and a brother-in-law came to persuade her to see a physician, she recognized them only with difficulty and ran to her room and locked herself up. On the night of her admission to the hospital, she suddenly left the house and attacked a passing neighbor with a large stone, knocking him to the ground. As the man lay there, Miss Maruyami hit him again and would have continued to do so if other neighbors had not restrained her.

History Miss Maruyami grew up in a middle-class neighborhood in Jakarta. She was the youngest of six daughters. At school she was a good student but had only a few friends and usually kept to herself. She did not continue her education after her junior high school year because her parents had financial problems. Since then she had stayed at home and helped her mother look after the house.

Her father, a merchant, died 10 years earlier when he was 60. Her mother is age 65 and is alive and well. Miss Maruyami's sisters are all married, but she is single and lives with her mother in a small house. Her relationship with her parents and sisters was good, although she was not particularly close to them. She was always a quiet person, spending much of her time lost in her thoughts, apparently daydreaming. She seemed shy and self-conscious, especially in public. She was noticeably overweight and often expressed fears that people were staring at her because of her obesity. These fears and suspicions became more marked as she grew older. Physically she was in good health, had never been to the hospital, and received no medication.

Findings On her admission Miss Maruyami was in a state of psychomotor excitement with aggressive and violent outbursts. When left alone she became more quiet and sat staring ahead of her, with sudden startling reactions to minor noises. She did not know the time or the place, but was fairly oriented as to her personal data. Her speech was restricted and somewhat incoherent. She refused to have physical or laboratory examinations, but accepted an injection of 10 mg of haloperidol and 5 mg of biperiden (Akineton). After this she became quiet and relaxed and finally fell asleep. The following day she seemed considerably improved. She was fully oriented but slightly perplexed, with partial amnesia about the previous few days. She only partly remembered her persecutory delusions. Confronted with the incident at the marketplace, she had a prolonged crying spell; became mildly excited, with fluctuations in attention and awareness of her surroundings; and finally again became herself.

She did not remember having hallucinations or other unusual experiences. She no longer believed that the neighbors wanted to kill her, and she wanted to get back home to her mother. She was discharged after 3 days, during which she continued to be quite natural without further medication.

Physical and neurological examinations, an electroencephalogram, and laboratory tests were all normal.

The neighbor, who had suffered a minor concussion of the head as a result of the attack, was not severely wounded. Being married to a distant relative of Miss Maruyami's mother, he did not want to bring the incident to the attention of the legal authorities.

Discussion Miss Maruyami developed a peracute psychotic disorder with persecutory delusions and possibly also with hallucinations, severe psychomotor excitation with violent behavior, and transient states of confusion. The disorder developed immediately after an event that she experienced as severely traumatic. She recovered with complete remission occurring within 1 week. It is a matter for discussion whether her transient states of confusion actually fulfill the criteria for organically caused clouding of consciousness. Otherwise, no signs of organic disorder were observed. Psychoactive substance abuse was not suspected. Miss Maruyami therefore most probably had an acute and transient psychotic disorder but did not fulfill the symptomatic criteria for an affective episode. The symptomatology was mixed and atypical, not pointing to any specific subtypes of acute psychotic disorders. The most likely diagnosis, therefore, is probably other acute and transient psychotic disorder.

It may further be discussed whether the traumatic event described would be considered as severely stressful to most people in similar circumstances within the same culture. In Miss Maruyami's case, she seems to have been made particularly susceptible by the presence of pronounced personality traits of introversive seclusion and hypersensitive self-consciousness, which with the present information is not sufficient for a subsidiary diagnosis of a personality disorder.

The diagnosis therefore would be other acute and transient psychotic disorder, with associated acute stress.

Acute and Transient Psychotic Disorder, Unspecified The residual category accommodates such concepts as brief reactive psychosis not otherwise specified. ICD-10 also includes another residual category for psychoses that do not meet the criteria for any other ICD-10 psychotic disorder (Table 13.3-2).

Psychotic disorders that do not meet the criteria for schizophrenia or for psychotic types of mood (affective) disorders, and psychotic disorders that do not meet the symptomatic criteria for persistent delusional disorder should be coded here (persistent hallucinatory disorder is an example). Combinations of symptoms not covered by the previous categories, such as delusions other than those listed as typically schizophrenic under criterion G1(1)b or d for schizophrenia (ie, other than completely impossible or culturally inappropriate) plus catatonia, should also be included here.

Table 13.3-2 ICD-10 Diagnostic Criteria for Other Nonorganic Psychotic Disorders

DSM-IV

Psychotic Disorder Due to a General Medical Condition The DSM-IV diagnosis of psychotic disorder due to a general medical condition (Table 13.3-3) combines into one diagnosis the two similar diagnostic categories in the revised third edition of DSM (DSM-III-R), organic delusional disorder and organic hallucinosis. The phenomena of the psychotic disorder are defined in DSM-IV by further specifying the predominant symptoms. When the diagnosis is used, both the medical condition and the predominant symptom pattern should be included in the diagnosis (e.g., psychotic disorder due to a brain tumor, with delusions). The DSM-IV criteria further specify that the disorder does not occur exclusively while the patient is delirious or demented and that the symptoms are not better accounted for by another mental disorder.

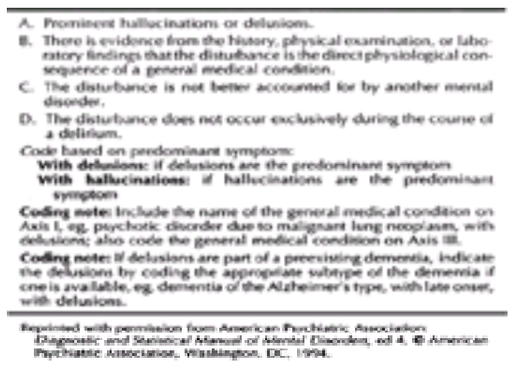

A. Prominent hallucinations or delusions.
B. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder.
D. The disturbance does not occur exclusively during the course of a delirium.

Code based on predominant symptom:
With delusions: if delusions are the predominant symptom
With hallucinations: if hallucinations are the predominant symptom

Coding note: Include the name of the general medical condition on Axis I, eg, psychotic disorder due to malignant lung neoplasm, with delusions; also code the general medical condition on Axis III.

Coding note: If delusions are part of a preexisting dementia, indicate the delusions by coding the appropriate subtype of the dementia if one is available, eg, dementia of the Alzheimer's type, with late onset, with delusions.

Table 13.3-3 DSM-IV Diagnostic Criteria for Psychotic Disorder Due to a General Medical Condition

Substance-Induced Psychotic Disorder DSM-IV has combined the various DSM-III-R diagnostic categories that relate to psychoactive substance–induced psychotic disorders into a single diagnostic category, substance-induced psychotic disorder (Table 13.3-4). The diagnosis is reserved for persons who have substance-induced psychotic symptoms in the absence of reality testing. Persons who have substance-induced psychotic symptoms (e.g., hallucinations) but who have retained reality testing should be classified as having a substance-related disorder (e.g., phencyclidine intoxication with perceptual disturbances). The intent of including the diagnosis of substance-induced psychotic disorder with the other psychotic disorder diagnoses is to prompt the clinician to consider the possibility that a substance is causally involved in the production of the psychotic symptoms. The full diagnosis of substance-induced psychotic disorder should include the type of substance involved, the stage of substance use when the disorder began (e.g., during intoxication or withdrawal), and the clinical phenomena (e.g., hallucinations or delusions).

Table 13.3-4 DSM-IV Diagnostic Criteria for Substance-Induced Psychotic Disorder

Psychotic Disorder Not Otherwise Specified The psychotic disorder not otherwise specified category is used for patients who have psychotic symptoms (e.g., delusions, hallucinations, and disorganized speech and behavior) but who do not meet the diagnostic criteria for other specifically defined psychotic disorders. In some cases the diagnosis of psychotic disorder not otherwise specified may be used when not enough information is available to make a specific diagnosis. DSM-IV has listed some examples of the diagnosis to help guide clinicians (Table 13.3-5).

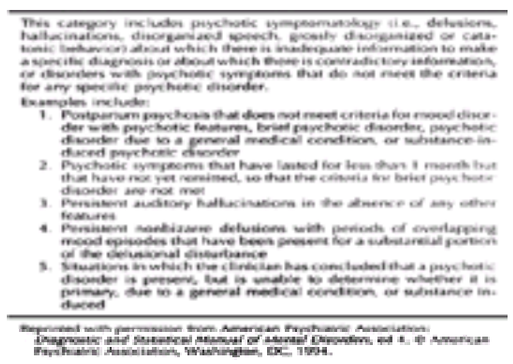

This category includes psychotic symptomatology (i.e., delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior) about which there is inadequate information to make a specific diagnosis or about which there is contradictory information, or disorders with psychotic symptoms that do not meet the criteria for any specific psychotic disorder.

Examples include:
1. Postpartum psychosis that does not meet criteria for mood disorder with psychotic features, brief psychotic disorder, psychotic disorder due to a general medical condition, or substance-induced psychotic disorder
2. Psychotic symptoms that have lasted for less than 1 month but that have not yet remitted, so that the criteria for brief psychotic disorder are not met
3. Persistent auditory hallucinations in the absence of any other features
4. Persistent nonbizarre delusions with periods of overlapping mood episodes that have been present for a substantial portion of the delusional disturbance
5. Situations in which the clinician has concluded that a psychotic disorder is present, but is unable to determine whether it is primary, due to a general medical condition, or substance-induced

Table 13.3-5 DSM-IV Psychotic Disorder Not Otherwise Specified

Culture-Bound Syndromes Perhaps the most dramatic example of the difficulties in applying Western-based nosological concepts and criteria cross-culturally can be found in the ongoing controversy surrounding the culture-bound syndromes. As pointed out by the United States National Institute of Mental Health Culture and Diagnosis Group, the term “culture-bound syndrome” denotes recurrent, locality specific patterns of aberrant behavior and troubling experiences that appear to fall outside conventional Western psychiatric categories. These include categories in folk nosological systems (often organized in relation to perceived cause and symptom clusters) as well as idioms of distress or culturally salient expressions for securing social support and communicating symptoms.

A variety of disorders discussed in the cross-cultural psychiatric and anthropological literature do not conform with conventional DSM-IV or ICD-10 diagnostic categories. In previous diagnostic classifications such disorders were often called atypical psychoses or were placed in the category psychotic disorder not otherwise specified. Taken as a whole, however, labeling these seemingly bizarre and culturally different patterns of disorder psychotic is somewhat problematic. The symptoms displayed include many clinical expressions of lesser severity than those of typical psychotic disorders. Frequently the diagnosis of a psychotic disorder is made simply on the basis of category label rather than a close examination of the symptomatic data. Further discussion of these disorders are discussed with a selection of such syndromes in Table 13.3-6.

Table 13.3-6 Culture-Bound Syndromes

While all psychiatric diagnoses are influenced by their cultural context, perhaps the most dramatic example of the difficulty in applying Western-based nosological concepts and criteria cross-culturally can be found with respect to the so-called culture-bound syndromes. The controversy arises over the issue of whether the assumptions embedded in Western diagnostic schemes (e.g., DSM or ICD) relating to concepts of normality and the assemblages of symptoms characteristic of a discrete disorder are universally applicable in all cultural settings.

The term “culture-bound syndrome” evolved, in fact, to denote recurrent, locality-specific patterns of aberrant behavior and troubling experiences that appear to fall outside conventional Western psychiatric categories. The descriptive phrases formerly used to refer to such phenomena include “cultural and ethnic psychoses and neuroses” and “atypical and exotic psychotic syndromes.” The *culture-bound syndrome* is now generally accepted to refer to culturally based and named patterns of symptoms of mental distress or maladaptive behavior that are prominent in folk belief and practice. Such patterns have etiologies framed by lay cultural assumptions, which in numerous cases are based upon the effects of sorcery, breach of taboo, intrusion of a disease object, intrusion of a disease-causing spirit, or loss of soul. The psychodynamics often represent an exaggerated or pathological reaction to conflicts engendered by central values and behavioral norms in the society. Since they are embedded in the group's ethnomedical practice, institutionalized patterns of diagnoses and societal response typically include treatment by indigenous healers.

Assessment of such syndromes must start with recognition that each human society has an indigenous body of beliefs and practices directed at explaining and treating disease and disorder, and patients internalize that worldview during the process of enculturation. They share their experiences and deal with distress through the currency of commonly understood symbols and meanings. In that light, the diagnostic encounter itself can be used as a point of entry into the patient's world, a “classroom for investigation and discovery.” One cannot become an anthropological expert about each and every possible cultural group but one can try to learn by asking patients to share the cultural norms as they understand them.

PATHOLOGY AND LABORATORY EXAMINATIONS

A large number of general medical problems may cause or exacerbate patients' psychotic conditions, often involving confusing and puzzling presentations. They include such conditions as infections (including human immunodeficiency virus [HIV] infection), head trauma, endocrine disorders (Cushing's and Addison's diseases and disorders of the thyroid and parathyroid glands), autoimmune diseases (systemic lupus erythematosus), vitamin deficiencies, seizure disorder, genetic diseases (Wilson's disease, acute intermittent porphyria), drug and toxin exposures, and the effects of psychoactive drugs. Those conditions are usually included in the differential diagnosis of any psychotic disorder, but they should be given more careful consideration when the patient's symptom profile is polymorphic or inchoate. For such patients laboratory tests should include not only the routine chemistry panels (electrolytes, glucose determination, complete blood counts, renal and liver functions) and urinalysis, but also thyroid function test, syphilis tests, and determination of serum cortisol concentration, vitamin B_{12} and foliate concentrations, and calcium and phosphate concentrations. In addition to a chest X-ray and an electrocardiogram (ECG), an electroencephalogram (EEG) should also be considered. An EEG with sleep deprivation and nasopharyngeal leads also has been recommended. Computerized EEG (brain mapping), magnetic resonance imaging (MRI), single photon emission computered tomography (SPECT), and neuropsychological testing may yield useful information.

Psychosocial assessment should include a careful review of the patient's life history, with special attention to the patient's personality traits and recent stresses. A detailed assessment of family history and dynamics should also be included. Contextual factors, such as psychosocial stressors and supports, should be carefully appraised, along with the ability of the person to perform basic roles (e.g., occupationally, with family and socially).

COURSE AND PROGNOSIS

Patients with an acute and transient psychotic disorder usually experience complete recovery within 1 to 3 months (depending on the specific disorder), often within a few weeks or days, and only a small proportion of patients develop persistently disabling states.

Limited data on the longitudinal course of patients with culture-bound syndromes suggest that some of them eventually develop clinical features compatible with a diagnosis of schizophrenia, bipolar disorder, cognitive disorder, or other psychotic disorders. Thus gathering information from all possible sources is crucial. Since clinical pictures evolve over time, thorough reevaluations should be conducted periodically to refine the diagnosis and improve clinical care.

TREATMENT

The treatment plan for any patient must be individualized, but that principle is particularly important when dealing with cases of acute and transient psychotic disorders and culture-bound psychotic disorders. Because these conditions are intricate and heterogeneous, no standard treatment strategy exists that can be applied to most cases. However, a number of general principles are crucial for the cases of those patients.

Careful evaluation, clinical observation, and comprehensive information gathering are the cornerstones of treatment planning for any psychiatric or general medical disorder. Longitudinal assessments are particularly important in the management of patients who are experiencing acute and transient psychotic disorders and culture-bound disorders. A multiaxial assessment using such schemas as those in ICD-10 and DSM-IV can substantially enhance the validity of diagnosis and the effectiveness of clinical care. A systematic evaluation of the cultural framework of the individual's identity, illness explanations, social context and functioning, and the doctor-patient relationship can be conducted along the lines of DSM-IV.

Because all the disorders discussed in this section share the presence of psychosis, pharmacotherapy frequently involves the use of antipsychotic drugs. Some evidence indicates that the dosage of antipsychotic drugs necessary for acute transient psychotic disorders is significantly lower than that required for other psychotic conditions, especially schizophrenia. It is thus prudent to use the lowest dose that can control the patient's symptoms. Since acute and transient psychotic disorders are often episodic, intermittent use of antipsychotic drugs, guided by the emergence of psychotic symptoms, is worth considering.

Depending on the clinical features of particular patients, many other psychiatric medicines have also been recommended. They include benzodiazepines for controlling agitation, lithium (Eskalith, Lithobid) for modulating mood swings, and antidepressants for ameliorating depressive symptoms. These medicines are often used in conjunction with antipsychotic agents. Anticonvulsants such as carbamazepine (Tegretol) have been reported effective in treating a number of psychotic patients with atypical features.

Limited research has been conducted to date on the efficacy of various psychosocial interventions for managing acute and transient psychotic disorders and culture-bound disorder, but it seems reasonable to consider findings from studies involving other psychotic conditions. These include approaches based on expressed emotion concepts, psychoeducational and skill-competence training, and Thomas McGlashan's phase-specific theory on the need for stimulation in schizophrenic patients (the avoidance of excessive stimuli in the acute phase and the uses of structured activities and stimuli in later phases). It is important to consider involving the family in therapy and to establish a supportive and trusting therapeutic relationship.

The importance of cultural issues in the evaluation and treatment of atypical psychoses can hardly be exaggerated, especially when dealing with patients from non-Western and ethnic minority populations. Cultural information is not only crucial for accurate diagnosis, but also indispensable in the formulation of treatment plans. Treatment approaches that do not take the patient's sociocultural background into account are likely to fail, no matter how well intentioned the therapists may be. For example, in cultures in which family and group harmony and unity are valued over individual independence, the rigid application of Western-based psychotherapeutic techniques may exacerbate, rather than ameliorate, the patient's psychopathological condition. Consideration of the intercultural elements in the clinician-patient relationship is also fundamental for establishment of rapport and effective engagement of the patient and the family in the treatment process.

One promising avenue is collaboration with indigenous healers. Several researchers have reported on their success in the use of indigenous and traditional healers in the treatment of psychiatric patients, especially those whose psychotic conditions are substantially connected to culture-specific beliefs (e.g., fear of voodoo death). Others have mentioned the potential pitfalls and problems in such collaboration. Decisions about involving indigenous healers should be individualized and thoughtfully planned, taking into consideration the setting, the thoughtfulness and flexibility of the available healers, the type of psychopathology, and the patient's characteristics. The World Health Organization (WHO) has long advocated implementation at the local level of a policy of close collaboration between the

conventional health system and traditional medicine, particularly between individual health professionals and traditional practitioners.

Wohl has stressed the implications for therapeutic practice of the need to understand the cultural dynamics of a patient's background. While this is important for therapy with any patient, it is particularly necessary in the treatment of a suspected culture-bound syndrome:

Much of the time in the practice of psychotherapy, culture remains silent, part of a non instrusive background, an invisible yet pervasive feature of the context of psychotherapy ... Psychotherapies differ as the cultures in which they were born and nurtured differ, and each bears the indelible imprint of its culture source. Psychotherapy and the human relationships that comprise both its subject matter and the medium in which it is performed have embedded within them values, rules, assumptions, myths, and rituals of a particular culture. Psychotherapy is thus inescapably bound to a particular cultural framework.

Treatment of a culture-bound syndrome poses several diagnostic challenges, the first of which is determining whether the symptomatology represents a culturally appropriate adaptive response to a situation (although it may be different from the therapist's). What are clinicians to do if confronted with a series of symptoms in a patient that do not fit their conventional diagnostic model? Clinicians are well advised to (1) know or search out the demographics of the local population or catchment area being served; (2) recognize that there is always a local pattern of conceptualization, naming, vocabulary, explanation, and treatment of patterns of distress that afflict a community, including mental disorders; and (3) talk with the family, get instructed in local customs; or search out other modes of documentation.

But while the observed symptom may be familiar in a general sense to the clinician (although pathoplastically different), what separates such an event from conventional understanding is the meaning of the symptom for patients and those who share their cultural background. Determining that meaning and particularly their belief about what has caused the distress, is an important entry point in facilitating therapeutic management and enhancing adherence to the treatment plan. As part of taking the history, ask these patients what they think could have caused the problem, requesting the “patient's explanatory model”: (1) What do you think has caused your problem? (2) Why do you think it started when it did? (3) What do you think your sickness does to you? How does it work? (4) How severe is your sickness? Will it have a short or long course? (5) What kind of treatment do you think you should receive?

Such insight into the dynamics of the patient's world facilitates the clinician's efforts to adapt his or her techniques (e.g., general activity level, mode of verbal intervention, content of remarks, tone of voice) to the cultural background of clients; communicate acceptance of and respect for the patients in terms that make sense within their cultural frame of reference; and be open to the possibility of more-direct intervention in the life of the patient than conventional approaches might suggest.

In conclusion, the treatment of patients experiencing acute transient psychotic disorders and culture-bound syndromes, even more than that of patients with other psychiatric disorders, should be personalized and comprehensive, using judiciously all biological, psychological, and social therapies pertinent to the problem at hand and keeping in mind the cultural framework of the patients and their families.

SUGGESTED CROSS-REFERENCES

Culture-bound syndromes are discussed further in Section 4.1 on anthropology and psychiatry. Cultural psychiatry is also discussed in Section 4.4. The influences of culture on the nature of and responses to psychiatric disorders are discussed in Section 4.2 on sociology and psychiatry. Section 4.3, on evolutionary biology and psychiatry, is also relevant. Section 9.2 covers international perspectives on psychiatric diagnosis. Section 13.1 is devoted to other psychotic disorders, including brief psychotic disorder.

SECTION REFERENCES

Akerele O: The best of both worlds: Bringing traditional medicine up to date. Soc Sci Med *24:*177, 1987.

Bustamante JA: *Psiquiatria Ciencia y Técnica.* Instituto Cubano del Libro, La Habana, 1972.

*Collins PY, Wig NN, Day R, Varma VK, Malhotra S, Misra AK, Schanzer B, Susser E: Psychosocial and biological aspects of acute brief psychoses in three developing country sites. Psychiatr Q *67:*177, 1996.

Cooper JE, Jablensky A, Sartorius N: WHO Collaborative studies on acute psychoses using the SCAAPS schedule. In *Psychiatry: A World Perspective*. CN Stefanis, AD Rabavilas, CR Soldatos, editors. Elsevier, Amsterdam, 1990.

Draguns JG: Dilemmas and choices in cross-cultural counseling: The universal versus the culturally distinctive. In *Counseling Across Cultures*, ed 3, PB Pedersen, JG Draguns, WJ Lonner, JE Trimble, editors. University of Hawaii Press, Honolulu, 1989.

Fabrega H: An ethnomedical perspective on Anglo-American psychiatry. Am J Psychiatry *146:*588, 1989.

Farmer AE, Falkowski WF: Maggot in the set, the snake factor, and the treatment of atypical psychosis in West African women. Br J Psychiatry *146:*446, 1985.

Fisher W, Piazza CC, Page TJ: Assessing independent and interactive effects of behavioral and pharmacological interventions for a client with dual diagnoses. J Behav Ther Exp Psychiatry *20:*241, 1989.

Hughes CC: Culture in clinical psychiatry. In *Culture, Ethnicity, and Mental Illness*, A Gaw, editor. American Psychiatric Press, Washington, DC, 1993.

Hughes CC: The culture-bound syndromes and psychiatric diagnosis. In *Culture and Psychiatry Diagnosis*, JE Mezzich, A Kiemman, H Fabrega, DL Parron, editors. American Psychiatric Press, Washington, DC, 1996.

Indian Council of Medical Research: *Collaborative Study on the Phenomenology and Natural History of Acute Psychosis*. Indian Council of Medical Research, New Delhi, 1989.

Jablensky A, Sartorius N, Ernberg G, Anker M, Korten A, Cooper JE, Day R, Bertelsen A: *Schizophrenia: Manifestations, Incidence and Course in Different Cultures; A World Health Organization Ten-Country Study*. Psychol Med *20*(Suppl):1, 1992.

Johnson FA: African perspectives on mental disorder. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

Jorge MR, Mezzich JE: Latin American contribution to psychiatric nosology and classification. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

Karno M, Jenkins JH: Cultural considerations in the diagnosis of schizophrenia and related disorders and psychotic disorder not otherwise classified. In *DSM-IV Source Book*, TA Widiger, A Frances, HA Pincus, MB First, R Ross, W Davis, editors. American Psychiatric Press, Washington, 1994.

Kirmayer LJ: The place of culture in psychiatric nosology: Taijin Kyofusho and DSM-III-R. J Nerv Ment Dis *179:*19, 1991.

Kleinman A: *Rethinking Psychiatry*. Free Press, New York, 1988.

Lee S: Culture in psychiatric nosology: The OCMD-Z-R and the International Classification of Mental Disorders. Cult Med Psychiatry *20:*421, 1996.

Leonhard K: *Aufteilung der Endogenen Psychosen.* Akademic Verlag, Berlin, 1957.

Lin K-M: Cultural influences on the diagnosis of psychotic and organic disorders. In *Culture and Psychiatric Diagnosis*, JE Mezzich, A Kleinman, H Fabrega, DL Parron, editors. American Psychiatric Press, Washington, DC, 1996.

Lin K-M, Kleinman AM: Psychopathology and clinical course of schizophrenia: A cross-cultural perspective. Schizophr Bull *14:*555, 1988.

Manschreck TC, Petri M: The atypical psychoses. Cult Med Psychiatry *2:*233, 1978.

Mezzich JE, Jorge MR: Psychiatric nosology: Achievements and challenges. In *International Review of Psychiatry*, JA Costa e Silva, CC Nadelson, editors. American Psychiatric Press, Washington, DC, 1993.

*Mezzich JE, Kleinman A, Fabrega H, Parron DL: *Culture and Psychiatric Diagnosis.* American Psychiatric Press, Washington, DC, 1996.

Mezzich JE, Kleinman A, Fabrega H, Parron DL, Good BJ, Lin K-M, Manson S, editors: Cultural issues section. In *DSM-IV Source Book*, TA Widiger, A Frances, HA Pincus, MG First, R Ross, W

Davis, editors. American Psychiatric Press, Washington, DC, 1997.

Otero AA: *Tercer Glorario Cubano de Psiquiatría (GC-3)*. Havana Psychiatric Hospital, Ministry of Public Health, Havana, Cuba, 1998.

Perris C: *A Study of Cycloid Psychoses*. Munksgaard, Copenhagen, 1974.

Pull CB, Chaillet G: The nosological views of French-speaking psychiatry. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

Sartorius N, DeGirolano G, Andrews G, German GA, Eisenberg L: *Treatment of Mental Disorder: A Review of Effectiveness*. World Health Organization and American Psychiatric Press, Washington, DC, 1993.

Seguín CA: *Psiquiatría Folklórica*. Ediciones Errmar, Lima, Peru, 1979.

Shen YC: On the second edition of the Chinese Classification of Mental Disorder. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

Simons RC: *Boo! Culture, Experience and the Startle Reflex*. New York: Oxford University Press, 1996.

*Simons RC, Hughes CC: *The Culture-Bound Syndromes: Folk Illnesses of Psychiatric and Anthropological Interest*. D Reidel, Dordrecht, Holland, 1985.

Strömgren E: Scandinavian contributions to psychiatric nosology. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

*Susser E, Wanderling E: Epidemiology of non-affective acute remitting psychosis vs schizophrenia. Arch Gen Psychiatry *51:*294, 1994.

Takahashi S: Diagnostic classification of psychotic disorders in Japan. In *Psychiatric Diagnosis: A World Perspective*, JE Mezzich, Y Honda, MC Kastrup, editors. Springer-Verlag, Berlin, 1994.

Tanaka-Matsumi J: Taijinkyofusho: Diagnostic and cultural issues in Japanese psychiatry. Cult Med Psychiatry *23:*1, 1979.

Wig NN, Parhee R: Acute and transient psychoses: A view from the developing countries. In *International Classification in Psychiatry: Unity and Diversity*, JE Mezzich, M von Cranach, editors. Cambridge University Press, Cambridge, 1988.

World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders. Clinical Descriptions and Diagnostic Guidelines*. World Health Organization, Geneva, 1992.

Yamashita I: *Periodic Psychosis of Adolescence*. Hokkaido University Press, Hokkaido, Japan, 1993.

13.4 POSTPARTUM PSYCHIATRIC SYNDROMES

RUTA NONACS, M.D., PH.D. AND LEE S. COHEN, M.D.

During the postpartum period up to 85 percent of women experience some type of mood disturbance (Table 13.4-1). For most women the symptoms are transient and relatively mild; however, some women experience a more disabling and persistent form of mood disturbance. Although postpartum mood disorders are relatively common, depressive symptoms that emerge during the postpartum period are frequently overlooked by patients and their caregivers. Puerperal affective illness places both the mother and infant at risk and has been associated with significant long-term effects on child development and behavior. Therefore, prompt recognition and treatment of puerperal mood disorders are essential.

Disorder	Prevalence	Onset	Characteristic Symptoms
Postpartum blues	30 to 85 percent	Within first week	Mood lability, tearfulness, insomnia, anxiety
Postpartum depression	10 to 15 percent	Usually insidious, within first 2 to 3 months	Depressed mood, excessive anxiety, insomnia
Puerperal psychosis	0.1 to 0.2 percent	Usually within first 2 to 4 weeks	Agitation and irritability, depressed mood or euphoria, delusions, depersonalization, disorganized behavior

Table 13.4-1 Classification of Postpartum Mood Disorders

HISTORY

Although Hippocrates is often acknowledged as the first to have recognized postpartum mental illness, historians have noted that what Hippocrates described as a mania related to lactation was more likely delirium associated with puerperal sepsis, which was relatively common in ancient Greece. There was virtually no mention of puerperal mental illness until the 1700s and 1800s, when case reports of “puerperal insanity” began to appear in the French and German medical literature. In 1818 Jean Esquirol was the first to provide detailed, quantitative data on 92 cases of puerperal psychosis drawn from his studies at the Salpetriere during the Napoleonic Wars. However, it is Victor Louis Marce, a French physician, who is best known for his descriptions of postpartum psychiatric illness. In his famous text published in 1856, *Traite de la Folie des Femmes Enceintes,* he laid the foundation for modern conceptualizations of mental illness related to pregnancy and the postpartum period. He was also the first to suggest that physiological changes associated with the puerperium influence maternal mood.

Although puerperal psychosis was familiar to most clinicians by the late nineteenth century, less attention was given to milder forms of puerperal illness. It was not until the 1960s that B. Pitt first described an “atypical” depression (later called “maternity blues”) that affected mothers soon after childbirth and, in contrast to puerperal psychosis, was relatively mild and short-lived. The concept of a more severe form of nonpsychotic depressive illness (i.e., postpartum depression) emerged during the 1970s. Large, population-based studies, which relied upon structured interviews and standardized diagnostic criteria to identify psychiatric illness in new mothers, demonstrated high rates of mild to moderate depression in women during the first 6 months after delivery.

Recent studies have consistently identified the postpartum period as a time of increased risk for the development of psychiatric illness in women. One of the most frequently cited studies on affective illness during the puerperium described a sharp peak in the number of psychiatric admissions during the first 3 months after delivery. Subsequent studies indicate that women who present with significant psychiatric symptoms during the postpartum period suffer most commonly from a mood disorder, either major depressive disorder or a bipolar disorder. During the postpartum period women appear to be at much higher risk for the development of psychiatric illness than at other times in their lives.

Various investigators have argued that postpartum mental illness consists of a group of psychiatric disorders that are specifically related to pregnancy and childbirth and therefore exists as a distinct diagnostic entity. However, recent evidence suggests that affective illness that emerges during the puerperium does not differ significantly from affective illness occurring in women at other times. This opinion is reflected in the fourth edition of *Diagnostic and Statistic Manual of Mental Disorders* (DSM-IV), which includes postpartum psychiatric illness as a subtype of either bipolar disorder or major depressive disorder.

DEFINITION

Postpartum psychiatric illness is typically divided into three categories: (1) postpartum blues, (2) nonpsychotic postpartum depression, and (3) puerperal psychosis (Table 13.4-1). It is helpful to conceptualize these disorders as existing along a continuum, as there may be significant overlap between these three diagnostic subtypes. Although these subtypes vary in severity, it is not clear if they actually represent three distinct disorders.

ETIOLOGY

The puerperium is a period during which significant physiological and psychosocial changes occur. The extent to which a rapidly changing hormonal environment influences the emergence of mood illness has been considered by many. In fact, it was Victor Louis Marce who first suggested, long before the emergence of the modern field of endocrinology, that a physiological transition occurring after delivery may play an important role in the pathogenesis of puerperal illness. Other investigators have emphasized the importance of biological vulnerability to psychiatric illness during the puerperium and have suggested that some individuals may be more susceptible to the physiological changes characteristics of the postpartum period. However, the impact of psychosocial factors in the development of mood disorder during the postpartum period cannot be underestimated. Given the multiplicity of these factors and the complexity of their interactions, it has been extremely difficult to identify risk factors for puerperal psychiatric illness and to reliably predict who will experience postpartum mood disturbance.

Demographic Variables Many groups have investigated the relationship between risk for postpartum blues and depression and various demographic variables including age, marital status, parity, education level, and socioeconomic status; however, there is little consistent evidence to suggest that any particular demographic factor places a woman at increased risk for puerperal affective illness. Although most studies do not find a strong relationship between age and risk for puerperal illness, there is at least one report of high rates (26 percent) of postpartum depression in adolescent mothers.

It has been significantly more difficult to identify risk factors for puerperal psychosis, given the low prevalence of this subtype of postpartum illness. Some reports suggest that primiparous women are more vulnerable to postpartum psychosis than multiparous women. Other studies suggest that various obstetrical complications

(e.g., prolonged labor, caesarean section, stillbirth) may increase the likelihood of postpartum psychosis.

Psychosocial Factors Psychosocial variables appear to play an important role in determining vulnerability to affective illness during the postpartum period. Many studies have sought to link certain personality traits and coping styles with risk for postpartum illness but have yielded inconsistent findings. In contrast, several groups have demonstrated that stressful life events during pregnancy or near the time of delivery appear to increase the likelihood of postpartum depressive illness. One of the most consistent findings is that among women who report marital dissatisfaction or inadequate social supports, postpartum depressive illness is more common.

History of Psychiatric Illness Although it has been difficult to identify specific demographic and psychosocial variables that consistently predict risk for postpartum illness, there is a well-defined association between all types of postpartum psychiatric illness and a personal history of mood disorder (Table 13.4-2). At highest risk are women with a history of postpartum psychosis; up to 70 percent of women who have had one episode of puerperal psychosis will experience another episode following a subsequent pregnancy. Similarly, women with histories of postpartum depression are at significant risk, with rates of postpartum recurrence as high as 50 percent. Women with bipolar disorders also appear to be particularly vulnerable during the postpartum period, with rates of postpartum relapse ranging from 20 to 50 percent.

Disorders	Risk of Relapse at Future Pregnancy
Postpartum Psychosis	70%
Postpartum Depression	50%
Bipolar I Disorder	20–50%
Major Depressive disorder	30%

Table 13.4-2 History of Psychiatric Illness and Risk for Puerperal Relapse

The extent to which a history of major depressive disorder influences risk for postpartum illness is less clear. As compared to women who have experienced only nonpuerperal depressive episodes, women with histories of postpartum depression are clearly at greater risk. Women with histories of mild to moderate affective illness who remain euthymic during pregnancy are probably at lower risk for postpartum depression than women with severe, recurrent depression. For all women (with or without histories of major depression), the emergence of depressive symptoms during pregnancy increases the likelihood of postpartum depression.

Hormonal Factors The postpartum period is characterized by a rapid shift in the hormonal environment. Within the first 48 hours after delivery, estrogen and progesterone concentrations fall dramatically; similarly, cortisol concentrations drop after delivery. As these steroid hormones have been implicated in the pathogenesis of nonpuerperal mood disorders, many investigators have proposed a role for these hormones in the emergence of a mood disorder during the postpartum period.

Progesterone Several anecdotal reports have suggested that mood disturbance during the puerperium may be related to declining concentrations of progesterone and have suggested a beneficial effect of progesterone hormone replacement in the treatment of postpartum psychiatric illness. However, several studies have found no consistent differences in postpartum progesterone concentrations between depressed and nondepressed women.

Estrogen Several studies have explored the relationship between postpartum estrogen levels and risk for postpartum blues and depression and have suggested that postpartum estrogen deficiency may result in postpartum mood disturbance. Although some studies have observed lower estrogen levels in women who developed postpartum blues and depression, most of the studies have yielded negative findings.

Cortisol Concentrations of cortisol, which are high late in pregnancy, peak during labor and delivery. Cortisol concentrations drop rapidly after delivery and then return to baseline levels gradually over the next month. While disturbances in the hypothalamic-pituitary-adrenal axis may play an important role in at least some cases of nonpuerperal major depression, recent studies do not consistently support a relationship between cortisol levels and postpartum blues or depression. The dexamethasone (Dexacidin) suppression test does not appear to distinguish between depressed and nondepressed women during the acute puerperium.

Thyroid Hormones Thyroxine concentrations are high during pregnancy and fall during the postpartum period. Abnormalities in thyroid function tests are relatively common findings during the postpartum period, and clinical hypothyroidism is present in up to 10 percent of women after childbirth. Although thyroid dysfunction, particularly hypothyroidism, may produce psychiatric symptoms, no studies have consistently reported an association between postpartum depression or blues and thyroid dysfunction (either hypothyroidism or hyperthyroidism).

DIAGNOSIS AND CLINICAL FEATURES

Postpartum psychiatric disorders have not been listed separately in recent revisions of the DSM, and no specific criteria for the diagnosis of postpartum psychiatric illness have been provided. According to DSM-IV, postpartum psychiatric illnesses may be indicated with a postpartum onset specifier. The specifier with postpartum onset may be used to describe a major depressive, manic, or mixed episode (in major depressive disorder or bipolar I or II disorder) or brief psychotic disorder, when the episode occurs within the first 4 weeks after delivery (Table 13.4-3). In contrast, the Marce Society, an international scientific organization dedicated to the study of postpartum psychiatric disorders, defines postpartum psychiatric illness as any episode occurring within the first year after childbirth.

Specify if:
With postpartum onset (can be applied to the current or most recent major depressive, manic, or mixed episode in major depressive disorder, bipolar I disorder, or bipolar II disorder, or to brief psychotic disorder)
Onset of episode within 4 weeks postpartum.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. ©, American Psychiatric Association, Washington, DC, 1994.

Table 13.4-3 DSM-IV Criteria for Postpartum Onset Specifier

Given the prevalence of mood disturbance during the puerperium, it is most striking that diagnoses of postpartum mood disorders are so commonly missed. The emergence of mood disorder during the puerperium is often overlooked or ignored by both patients and their caregivers. Some studies report that less than one-third of women with postpartum illness seek professional help. It is common for women to report the persistence of depressive symptoms for many months before the initiation of treatment. Although the symptoms of depression may remit spontaneously, many women are still depressed at 1 year after childbirth. The reasons for this delay in treatment are not well understood. What is clear, however, is the significant impact of untreated depression on both mother and infant. Untreated depression may contribute to the development of a more chronic and refractory mood disorder in the mother. There are also significant data that demonstrate the adverse effects

of maternal depression on the cognitive, emotional, and social development of the child. Given these significant risks, prompt recognition and treatment of postpartum mood disorders are essential.

Postpartum Blues Many women experience mild depressive symptoms during the first week after delivery, which are commonly known as postpartum blues or “baby blues.” Depending on the criteria used to diagnose the blues, prevalence estimates range from 30 to 85 percent. Women with postpartum blues report a variety of symptoms, including dysphoria, mood lability, irritability, tearfulness, anxiety, and insomnia. These symptoms typically peak on the fourth or fifth day after delivery and remit spontaneously by the tenth postpartum day. Postpartum blues are relatively benign and are, by definition, time-limited. While the occurrence of postpartum blues does not necessarily reflect psychopathology in the mother, some women with blues will go on to develop postpartum depression. Women with histories of mood disorder require close monitoring, as some data suggest that blues may herald the development of major depressive disorder in women who have had previous episodes of affective illness. Symptoms of the blues that persist beyond the second postpartum week require further evaluation to rule out the evolution of a more serious affective illness.

Postpartum Depression Major depressive disorder is relatively common during the postpartum period. Both retrospective and prospective community-based studies have revealed rates of postpartum minor and major depression in the range of 10 to 15 percent. These rates of depression reported in puerperal cohorts are similar to those observed in nonpuerperal populations of women.

While some women report the acute onset of symptoms shortly after delivery, depression more commonly develops insidiously over the first 6 postpartum months. A significant proportion of women actually experience the onset of depressive symptoms during pregnancy. The signs and symptoms of postpartum depression are generally indistinguishable from those characteristic of nonpsychotic major depressive disorder that occurs in women at other times. Dysphoric mood, irritability, anhedonia, insomnia, and fatigue are frequently reported; somatic complaints are also common. Ambivalent or negative feelings toward the infant are often reported, and it is common for a woman with postpartum depression to express doubts or concerns about her ability to care for her child. In its most severe form, postpartum depression may result in profound dysfunction. Suicidal ideation is frequently reported; however, suicide rates appear to be relatively low in women who become depressed during the postpartum period.

Although few studies have evaluated the prevalence of comorbid psychiatric illness in this population, severe anxiety and obsessionality are prominent in women with puerperal illness. Symptoms of generalized anxiety, panic disorder, and obsessive-compulsive disorder are often observed in women with postpartum depression.

Puerperal Psychosis Puerperal psychosis is the most severe form of postpartum psychiatric illness. In contrast to postpartum blues and depression, puerperal psychosis is a rare event that occurs in approximately 1 to 2 per 1000 women after childbirth. Its presentation is often dramatic, with onset of psychosis as early as the first 48 to 72 hours postpartum. Most women with puerperal psychosis develop symptoms within the first 2 to 4 weeks after delivery.

In women with this disorder, psychotic symptoms and disorganized behavior are prominent and cause significant dysfunction. Puerperal psychosis resembles a rapidly evolving affective psychosis with manic, depressive, or mixed features. The earliest signs are typically restlessness, irritability, and insomnia. Women with this disorder typically exhibit a rapidly shifting depressed or elated mood, disorientation or depersonalization, and disorganized behavior. Delusional beliefs often center on the infant and include delusions that the child may be defective or dying, that the infant has special powers, or that the child is either Satan or God. Auditory hallucinations that instruct the mother to harm or kill herself or her infant are sometimes reported. Although most believe that this illness is indistinguishable from an affective (or manic) psychosis, some have argued that puerperal psychosis may be clinically distinct in that it is more commonly associated with confusion and delirium than nonpuerperal psychotic mood disorder.

Screening Severe postpartum depression and psychosis are easily recognized; however, milder or more insidious forms of depressive illness are frequently missed. Even severe depressive symptoms that arise during the puerperium may be dismissed by both patients and caregivers as normal or natural consequences of childbirth. Since it is difficult to reliably predict which women in the general population are likely to develop puerperal illness, it is advisable to screen all women for depression during the postpartum period. The greatest obstacle to the diagnosis of postpartum depression is the extent to which clinicians fail to inquire about affective symptoms in women during the postpartum period.

The standard postpartum obstetrical visit at 6 weeks and subsequent pediatric appointments are ideal times to screen for postpartum depressive illness. Screening for mood disorders during the postpartum period may, however, be more difficult than at other times. Many of the neurovegetative signs and symptoms characteristic of major depression (e.g., sleep and appetite disturbance, diminished libido, low energy) are also observed in nondepressed women during the acute puerperium. Various rating scales that have been used to facilitate the diagnosis of depression in nonpuerperal cohorts (e.g., Beck Depression Inventory) have not been validated in puerperal populations. In contrast, the Edinburgh Postnatal Depression Scale (EPDS) is a 10-item, self-rated questionnaire (Table 13.4-4) that has been used extensively for the detection of postpartum depression and has demonstrated satisfactory sensitivity and specificity in women during the postpartum period. Although not commonly employed, the EPDS could easily be integrated into the routine evaluation of women in both obstetrical and pediatric settings and would alert the physician to those women who are in need of a more thorough psychiatric evaluation.

Table 13.4-4 Edinburgh Postnatal Depression Scale (EPDS)

DIFFERENTIAL DIAGNOSIS

Various medical illnesses may mimic psychiatric illness during the postpartum period. Hypothyroidism is relatively common in women after delivery and may cause a constellation of symptoms resembling major depressive disorder. Women with a preexisting psychiatric illness may experience exacerbation of symptoms during the puerperium. Furthermore, any psychiatric illnesses may emerge for the first time during the postpartum period. Schizophrenia or schizoaffective disorder, particularly when characterized by prominent positive symptomatology, may be difficult to distinguish from puerperal psychosis. Whereas mood disorders are the most common type of postpartum psychiatric illness, anxiety symptoms are common during the postpartum period and many present either with or without a coexisting mood disorder. The postpartum period may represent a time of increased risk for the development of panic disorder and obsessive-compulsive disorder.

COURSE AND PROGNOSIS

The duration of postpartum illness appears to be variable. Puerperal episodes are often relatively short-lived and last no more than 3 months. Many women, however, have a more prolonged illness, and several studies suggest that depressive episodes tend to be longer and more severe in those with histories of major depressive disorder; some reports suggest that duration may be related to the severity of illness.

In general, women with postpartum mood disorders have a good prognosis. In about half of the cases, puerperal depression or psychosis represents the first onset of psychiatric illness. Although there appears to be a subpopulation of women who have only puerperal episodes of psychiatric illness, the majority of women with a postpartum mood disorder will go on to have nonpuerperal episodes of psychiatric illness. Rates of recurrence appear to be particularly high in women with bipolar disorders.

Postpartum mood disorders are associated with recurrent psychiatric illness in the mother. The failure to treat may contribute to the emergence of a long-term and more treatment refractory mood disorder. There is data to suggest that the outcome is better in those that receive treatment early during the course of illness.

There is a growing body of literature that demonstrates the detrimental effect of maternal depression on child development. Attachment difficulties are common in new mothers and may be quite severe in women with postpartum depression or psychosis. Long-term follow-up studies have shown that behavioral difficulties were more common in the children of mothers who suffered from postpartum depression. These children also performed worse on structured tests of cognitive ability than children who had mothers who were not depressed. One of the most disastrous outcomes involves harm to the infant. Child abuse and neglect are more common among women who suffer from postpartum psychiatric illness. Infanticide is relatively uncommon; however, it is more likely to occur in women who present with psychotic symptoms.

TREATMENT

Like nonpuerperal depressive illness, postpartum mood disorders present along a continuum. Patients may experience mild or moderate symptoms, or they may present with a more severe depression, characterized by prominent neurovegetative symptoms and marked impairment of functioning. A clinician's approach to the patient should be guided by the type and severity of the symptoms and the degree of functional impairment. However, before initiating psychiatric treatment, medical causes for mood disturbance (e.g., thyroid dysfunction, Sheehan's syndrome) must be excluded. Initial evaluation should include a thorough history, physical examination, and routine laboratory tests.

Postpartum Blues As postpartum blues are usually mild in severity and resolve spontaneously, no specific treatment other than support and reassurance is indicated. Although the symptoms may be distressing, they typically do not affect the mother's ability to function and to care for her infant. Psychiatric consultation is generally not required; however, the patient should be instructed to contact her obstetrician or primary care provider if the symptoms persist longer than two weeks to ensure the early identification of a more severe affective illness. Women with histories of psychiatric illness, particularly postpartum depression, should be monitored more closely, as they are at higher risk for significant puerperal illness.

Postpartum Depression Although postpartum depression is relatively common, few studies have systematically assessed the efficacy of nonpharmacological and pharmacological therapies in the treatment of this disorder. However, there are no data that suggest that postpartum depression should be managed differently than nonpuerperal major depressive disorder. There is, nonetheless, an apparent tendency to treat women with postpartum depression with less urgency than nonpuerperal patients, which places both the mother and infant at risk. In the absence of systematically derived data, depression that emerges during the postpartum period demands the same intensity of treatment as depression that occurs at other times; the earlier the treatment is initiated, the better the prognosis.

Nonpharmacological Therapy Nonpharmacological therapies are frequently employed in the treatment of postpartum depression; however, there are limited data to support the efficacy of these interventions. Although studies that have assessed the use of insight-oriented psychotherapy in the treatment of postpartum depression have yielded inconsistent findings, more structured types of individual psychotherapy have shown promise.

Interpersonal therapy is a time-limited and interpersonally oriented psychotherapy that has been used successfully (in nonpuerperal cohorts) to treat acute episodes of depression. This modality of therapy focuses primarily on interpersonal relationships and has been adapted for the treatment of postpartum depression. In this setting, interpersonal therapy may be used to address the following issues: role transition, disruption of relationships with the spouse and other social supports, and interaction with the infant. In a recent pilot study, interpersonal therapy was shown to be effective for the treatment of women with mild to moderate postpartum major depressive disorder.

Cognitive-behavioral therapy may also be useful in this setting. A recent randomized, placebo-controlled treatment trial demonstrated that short-term cognitive behavioral therapy was as effective as treatment with fluoxetine (Prozac) in women with postpartum depression. Sessions were structured to focus on issues specific to new mothers with postpartum depression: inability to cope with the demands of caring for a child, perceived lack of support, absence of enjoyable activities. A significant reduction in depressive symptoms was observed in women after six sessions of cognitive-behavioral therapy delivered over a 12-week period.

These nonpharmacological interventions may be particularly useful for those patients who are reluctant to use psychiatric medications or for patients with milder forms of depressive illness. These interventions would ideally be performed in the home by a visiting nurse or another trained individual. Further investigation is required to determine the efficacy of this modality in women who suffer from more severe forms of postpartum mood disturbance.

Pharmacological Therapy To date, few studies have assessed the efficacy of antidepressant medications in the treatment of postpartum mood disturbance. The majority of these studies have been open trials, although more recent investigations have employed a double-blind design. Several studies have demonstrated the efficacy of antidepressant medications (e.g., fluoxetine, sertraline [Zoloft] venlafaxine [Effexor]) in the treatment of postpartum major depressive disorder. In all of these studies, standard antidepressant doses were effective and well tolerated.

The choice of an antidepressant drug should be guided by the patient's prior response to antidepressant medication and a given medication's adverse effect profile. Fluoxetine and the other selective serotonin reuptake inhibitors (SSRIs) are ideal first-line agents because they are anxiolytic, generally nonsedating, and well tolerated. Tricyclic drugs are frequently used and, because they tend to be more sedating, may be more appropriate for women who present with sleep disturbance. Given the severity of anxiety symptoms in women with postpartum depression, the adjunctive use of a benzodiazepine (e.g., clonazepam [Klonopin], lorazepam [Ativan]) may be very helpful.

Women who plan to breastfeed must be informed that all psychiatric medications, including antidepressant drugs, are secreted into the breast milk. Concentrations in the breast milk appear to vary widely; however, there is no data to suggest that one antidepressant agent is safer than another for women who are nursing. Available data on the use of tricyclic drugs, fluoxetine, and sertraline during breastfeeding suggest that severe complications related to neonatal exposure to psychotropic medications in breast milk appear to be rare; however, the long-term effects of even trace amounts of medication on the developing brain are not known.

Inpatient Hospitalization In cases of severe postpartum depression, inpatient hospitalization may be required, particularly for patients who are at risk for suicide. In the United Kingdom innovative treatment programs involving joint hospitalization of the mother and baby have been successful; however, mother-infant units are much less common in the United States. In women with severe postpartum illness, electroconvulsive therapy (ECT) should be considered early because it is safe and highly effective. In choosing a treatment strategy, it is important to consider the impact of prolonged hospitalization of the mother on infant development and attachment.

Hormonal Therapy The postpartum period is associated with rapid shifts in the reproductive hormonal environment, most notably a dramatic fall in estrogen and progesterone levels. With increasing evidence to suggest that gonadal steroids modulate neurotransmitter systems involved in the pathogenesis of mood illness, many have proposed a role for hormonal manipulation in the treatment of postpartum mood disturbance. Some authors have suggested that progesterone is helpful in the management of postpartum depression; however, systematically derived data do not demonstrate its usefulness in this setting. A.J. Gregoire and colleagues described a beneficial effect of exogenous estrogen therapy in women with postpartum depression. Although this study was small and was confounded by the inclusion of patients treated with antidepressant medication prior to receiving hormonal therapy, it is the first study to demonstrate that estrogen alone (or possibly when used as an adjunct to an antidepressant agent) may be useful in the treatment of postpartum depression. At this point it is unclear which patients are likely to respond to this type of hormonal therapy. In cases of moderate to severe depression, first-line pharmacological treatment should be an antidepressant medication.

Puerperal Psychosis Puerperal psychosis is a psychiatric emergency that typically requires inpatient treatment; however, systematically derived guidelines for the treatment of this disorder are lacking. Given the well-established relationship between puerperal psychosis and bipolar disorders, some have argued that postpartum psychosis is indistinguishable from a manic psychosis and should be treated similarly. Short-term treatment with an antipsychotic medication as well as a mood stabilizer is appropriate. Most groups have used lithium in the treatment of postpartum psychosis; the efficacy of other mood stabilizers (i.e., valproic acid [Depakene], carbamazepine [Tegretol]) in this setting is not known. Breastfeeding is typically avoided in women treated with lithium, as it is secreted at high levels into the breast milk and may cause neonatal toxicity. ECT (often bilateral) is well tolerated and rapidly effective. Failure to treat puerperal psychosis aggressively places both the mother and infant at increased risk for harm. Rates of infanticide associated with untreated puerperal psychosis have been estimated to be as high as 4 percent; the risk for suicide in this population is also extremely high.

Although some authors recommend the discontinuation of psychotropic medications soon after the psychosis clears, others suggest a longer duration of treatment, arguing that women are at risk for psychiatric illness for up to 1 year after childbirth. Prolonged exposure to conventional antipsychotic agents should be minimized,

given the risk of tardive dyskinesia. When neuroleptic medications are discontinued, they should be tapered slowly and with close observation for early signs of recurrence. Treatment with a mood stabilizer should be maintained beyond the resolution of active symptoms to reduce risk for relapse. The appropriate duration of treatment with a mood stabilizer has not been well established. Whether all patients with puerperal psychosis should subsequently receive maintenance treatment with a mood stabilizer for an indefinite period of time is controversial, although several studies suggest that recurrent mood illness (most commonly bipolar disorder) is the rule following an episode of puerperal psychosis.

Prevention It is not possible to reliably predict which women in the general population will experience postpartum mood disturbance. The identification of women at greatest risk for puerperal illness improves the likelihood of early diagnosis and treatment and provides an opportunity to limit morbidity in the mother and her infant. It is possible to identify certain subgroups of women who are more vulnerable to postpartum mood disorder. Women with histories of affective illness are at particular risk for puerperal worsening of mood. Risk for puerperal relapse as a function of subtype is summarized in Table 13.4-2.

As it is possible to identify certain subgroups of women who are at risk for postpartum illness, many investigators have explored the appropriateness of prophylactic intervention. Several studies have demonstrated that women with histories of bipolar disorders or puerperal psychosis benefit from prophylactic lithium therapy instituted just prior to delivery (at 36 weeks gestation) or no later than the first 48 hours postpartum. Prophylactic lithium appears to significantly reduce relapse rates, as well as to diminish the severity and duration of puerperal illness.

Yet to be adequately investigated is the extent to which other populations of women may also benefit from pharmacological prophylaxis. A pilot study described a beneficial effect of prophylactic antidepressants administered to women with histories of postpartum depression. The efficacy of prophylactic antidepressants in women with recurrent, nonpuerperal depression is currently under investigation.

Psychosocial interventions, such as psychoeducational and supportive groups, are frequently included in the care of women during the postpartum period. The extent to which these interventions are effective in preventing postpartum mood disturbance has not been systematically studied. Several investigators have explored the use of psychoeducational groups during pregnancy and the postpartum period and have demonstrated a significant reduction in the incidence of postpartum depression in women who received this intervention, as compared to untreated controls.

Interpersonal therapy has been used successfully in nonpuerperal cohorts with major depressive disorder to prevent recurrence. Although not yet studied in puerperal populations, IPT and similar techniques may be performed prophylactically during the postpartum period and may be of some value. Those providers who routinely participate in acute and short-term postpartum care (e.g., visiting nurses) may be ideally suited to screen for and intervene in cases of postpartum distress or mood disturbance.

In summary, prophylaxis against postpartum depressive illness may be conceptualized along a continuum where some women are at low risk for puerperal illness whereas others appear to be at high risk for puerperal decompensation. This spectrum of patients at risk for puerperal mood disturbance and the potential role of pharmacological and nonpharmacological prophylaxis is outlined in Table 13.4-5. While a less aggressive “wait-and-see” approach is appropriate for women with postpartum blues or women without histories of psychiatric illness, women at high risk, particularly those with histories of postpartum illness, deserve close monitoring as well as specific prophylactic measures. Further characterization of these subgroups of women and the prophylactic treatments suited to each is clearly warranted.

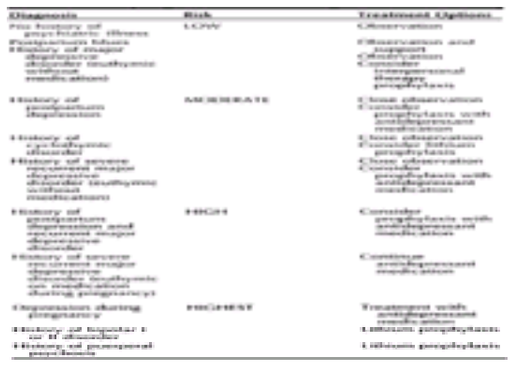

Table 13.4-5 Risk Factors and Treatment Options for Postpartum Mood Disturbance

SUGGESTED CROSS-REFERENCES

Further information on the reproductive endocrinology of pregnancy and the postpartum period and guidelines for the use of psychotropics during pregnancy and lactation can be found in Section 28.2 on psychiatry and reproductive medicine. Mood disorders are discussed at length in Chapter 14, and brief psychotic disorder is discussed in Section 13.1. Psychopharmacology is covered in Chapter 31.

SECTION REFERENCES

Appleby L, Warner R, Whitton A, Faragher B: A controlled study of fluoxetine and cognitive-behavioral counselling in the treatment of postnatal depression. Br Med J *314:*932, 1997.

Brockington IF: *Motherhood and Mental Illness.* Grune & Stratton, New York, 1982.

Brockington IF, Cernik KF, Schofield EM, Downing AR, Francis AF, Kelan C: Puerperal psychosis: Phenomena and diagnosis. Arch Gen Psychiatry *38:*829, 1981.

Cohen LS, Sichel DA, Robertson LM, Heckscher E, Rosenbaum JF: Postpartum prophylaxis for women with bipolar disorder. Am J Psychiatry *152:*1641, 1995.

Cooper PJ, Murray L: The course and recurrence of postnatal depression. Br J Psychiatry *166:*191, 1995.

*Cox JL, Holden JM, Sagovsky R: Detection of postnatal depression: Development of the 10-item Edinburgh Postnatal Depression Scale. Br J Psychiatry *150:*782, 1987.

Cox JL, Murray D, Chapman G: A controlled study of the onset, duration and prevalence of postnatal depression. Br J Psychiatry *163:*27, 1993.

Cummings EM, Davies PT: Maternal depression and child development. J Child Psychol Psychiatry *35:*73, 1994.

Davidson J, Robertson E: A follow-up study of postpartum illness, 1946–1978. Acta Psychiatr Scand *71:*451, 1985.

Gregoire AJP, Kumar R, Everitt B, Henderson AF, Studd JW: Transdermal estrogen for treatment of severe postnatal depression. Lancet *347:*930, 1996.

*Hendrick V, Altshuler LL, Suri R: Hormonal changes in the postpartum and implications for postpartum depression. Psychosomatics *39:*93, 1998.

Kendell RE, Chalmers JC, Platz C: Epidemiology of puerperal psychoses. Br J Psychiatry *150:*662, 1987.

Kennerely H, Gath D: Maternity blues, I. Detection and measurement by questionnaire. Br J Psychiatry *155:*356, 1989.

Klompenhouwer JL, van Hulst AM: Classification of postpartum psychosis: A study of 250 mother- and baby-admissions in the Netherlands. Acta Psychiatr Scand *84:*255, 1991.

*Llewellyn A, Stowe ZN: Psychotropic medications in lactation. J Clin Psychiatry 59 (Suppl):41, 1998.

*Murray L: The impact of postnatal depression on infant development. J Child Psychol Psychiatry *33:*543, 1992.

O'Hara MW, Schlechte JA, Lewis DA, Wright EJ: Prospective study of postpartum blues: Biologic and psychosocial factors. Arch Gen Psychiatry *48:*801, 1991.

O'Hara MW: Postpartum depression: Causes and consequences. Springer-Verlag, New York, 1994.

Pitt B: ‘Atypical' depression following childbirth. Br J Psychiatry *114:*1325, 1968.

Stewart DE, Klompenhouwer JL, Kendall RE, Van Hulst AM: Prophylatic lithium in puerperal psychosis: The experience of three centers. Br J Psychiatry *158:*393, 1991.

Stowe ZN, Casarella J, Landrey J, Nemeroff CB: Sertraline in the treatment of women with postpartum major depression. Depression *3:*49, 1995.

Stuart S, Couser G, Schilder K, O'Hara MW, Gorman L: Postpartum anxiety and depression: Onset and comorbidity in a community sample. J Nerv Ment Dis *186:*420, 1998.

*Stuart S, O'Hara MW: Interpersonal psychotherapy for postpartum depression: A treatment program. J Psychother Pract Res *4:*18, 1995.

Wisner KL, Wheeler RN: Prevention of recurrent postpartum major depression. Hosp Community Psychiatry *45:*1191, 1994.

CHAPTER 14. MOOD DISORDERS

14.1 MOOD DISORDERS: INTRODUCTION AND OVERVIEW

HAGOP S. AKISKAL, M.D.

For nearly 2500 years mood disorders have been described as the most common diseases of mankind, but only recently have they commanded major public health interest. The World Health Organization has ranked depression fourth in a list of the most urgent health problems worldwide. The U.S. Agency for Health Care Policy and Research, a federal agency concerned with medical practice from a public health perspective, devoted 2 volumes to depression among the first 10 it has published on such topics as pain, hypertension, diabetes mellitus, and coronary artery disease. A University of California psychiatrist, Kenneth Wells, demonstrated that the disability induced by depression compares with and often exceeds those of such diseases. Depressive disorders afflict at least 20 percent of women and 12 percent of men at some time during their lives. Despite the availability of effective treatments, many persons with mood disorders are disabled, and rates of suicide (which occurs in about 15 percent of depressive disorders) are high in both young and (especially) elderly men. Although depressive disorders are more common in women, more men than women die of suicide.

The suboptimal outcome of mood disorders documented in recent research reports cannot be ascribed to underdiagnosis and undertreatment alone for several reasons. First, Gerald Klerman and colleagues have suggested that the incidence of mood disorders may be increasing in younger age groups, especially in cohorts born in the 1960s, and may be associated with rising rates of alcohol and substance abuse. Second, mood disorders, once believed to be essentially adult disorders, are increasingly diagnosed in children and adolescents. Third, clinical studies suggest higher rates of chronicity, recurrence, and refractoriness than previously believed. For instance, chronicity, reported by Emil Kraepelin to occur in no more than 5 percent in the early twentieth century in Germany, is now seen in 15 to 20 percent of affectively ill patients in Western countries. Nonetheless, outcome studies coming from university centers tend to overestimate the proportion of cases with less favorable prognosis, and undeniably, many patients seen in private practice experience a favorable outcome. Also, a recent report indicates that depressed patients treated by psychiatrists in private settings receive much better care than those in other settings.

THE SCOPE OF MOOD DISORDERS

In Europe where the concept of mood disorders has historically embraced a broad spectrum of disorders, two British schools of thought have been influential. The Maudsley school—Aubrey Lewis and his followers—has promoted a continuum model, from anxiety disorders to mild neurotic depressions to severe endogenous and psychotic depressions. The Newcastle school, led by Martin Roth, has sharply demarcated those conditions from one another. Although vestiges of both approaches are still influential in clinical and basic research, their significance is presently overshadowed by continental European studies that subdivide mood disorders on the basis of polarity: unipolar (depressive episodes only) and bipolar (depressive episodes plus manic, hypomanic, or mixed episodes). That subdivision, supported by studies in the United States, has served as the basis for much recent research into the classification of mood disorders, as reflected in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). Despite such official sanction, many authorities today continue to see considerable continuity between recurrent depressive and bipolar disorders.

Emerging data also tend to favor a continuum between juvenile and adult mood disorders. This is based on the pioneering contributions by Elva Poznanski at the University of Michigan, as well as the work of Leon Cytrin and colleagues at the National Institute of Mental Health (NIMH), and Joachim Puig-Antich conducted at Columbia University in New York. Childhood bipolarity, too, is receiving increasing clinical attention, thanks to the seminal work of Elizabeth Weller and colleagues, originally conducted at Ohio State University. Finally, the work of Gabrielle Carlson and Michael Strober (when they collaborated at the University of California at Los Angeles) and clinical observations by this author on the juvenile offspring of adult bipolar disorders patients have led to a greater appreciation of the bipolar nature of complex clinical presentations of affective illness in children.

Current concepts of mood disorders in the United States embrace a wide spectrum, including many conditions previously diagnosed as schizophrenia, personality disorder, or neurosis. The diagnostic shift occurred in part as a result of the U.S.-U.K. Diagnostic Project, which demonstrated that schizophrenia was being diagnosed at the expense of mood disorders (Fig. 14.1-1). Conceptual boundaries were further broadened by the availability of new and effective treatments and by the unacceptable risk for tardive dyskinesia and suicide in persons with misdiagnosed mood disorders. More generally, present research interest in mood disorders in the United States emanated from a landmark 1969 NIMH conference on the psychobiology of affective illnesses: the NIMH Collaborative Depression Study—a long-term prospective project deriving directly from recommendations made at the conference—has legitimatized the broader perspective.

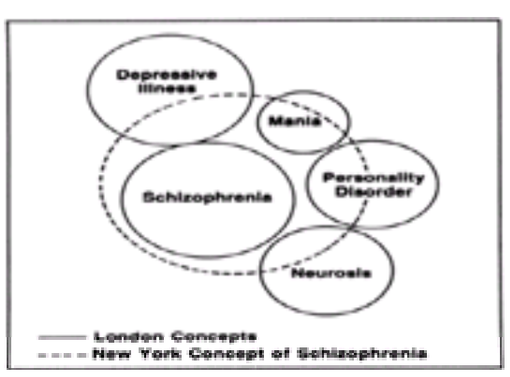

FIGURE 14.1-1 Comparison of British (London) and United States (New York) concepts of schizophrenia. (Reprinted with permission from Cooper JE, Kendell RE, Garland BJ, Sharpe L, Copeland JRM, Simon R: *Psychiatric Diagnosis in New York and London.* Oxford University Press, London, 1972.)

Unfortunately, findings published by Martin Keller and collaborators in the 1980s documenting gross undertreatment of mood disorders continue to describe the current treatment landscape worldwide. Whatever changes have occurred in diagnostic practice do not appear to have significantly affected the morbidity and mortality of mood disorders. This is all the more scandalous because the 1990s have seen new classes of user-friendly antidepressant and mood-stabilizing agents as well as depression-specific psychotherapies. In the author's opinion, this state of affairs results in part from the fact that both specialized and primary care training in mood disorders have failed to keep pace with recent advances.

The situation is analogous to adult-onset diabetes with approximately 50 percent of cases undetected or inadequately controlled, both of which seem to lag behind the field of hypertension, in which early detection and treatment have significantly reduced complications such as stroke. Efforts by patient advocacy organizations—often in concert with national psychiatric organizations and governmental mental health agencies—appear to be increasing public and government awareness of mood disorders. But ultimately, the challenge is to provide all primary care physicians with the requisite hands-on experience in this prevalent group of disorders.

Since mood disorders underlie 50 to 70 percent of all suicides, effective treatment of these disorders on a national level should, in principle, drastically reduce this major complication of mood disorders. A small-scale Swedish study has yielded promising results in this regard. In addition, clinical findings in recurrent mood

disorders have clearly shown the value of lithium prophylaxis in the prevention of suicide as well as overall mortality.

DEFINITIONS

Mood disorders encompass a large group of psychiatric disorders in which pathological moods and related vegetative and psychomotor disturbances dominate the clinical picture. Known in previous editions of DSM as affective disorders, the term *mood disorders* is preferred today because it refers to sustained emotional states, not merely to the external (affective) expression of the present emotional state. Mood disorders are best considered as syndromes (rather than discrete diseases) consisting of a cluster of signs and symptoms sustained over a period of weeks to months, that represent a marked departure from a person's habitual functioning and tend to recur, often in periodic or cyclical fashion.

Major Depressive Disorder and Bipolar Disorder Major depressive disorder (unipolar depression) is reported to be the most common mood disorder. It may manifest as a single episode or as recurrent episodes. The course may be somewhat protracted—up to 2 years or longer—in those with the single-episode form. Whereas the prognosis for recovery from an acute episode is good for most patients with major depressive disorder, three out of four patients experience recurrences throughout life, with varying degrees of residual symptoms between episodes. Bipolar disorders (previously called *manic-depressive psychosis*) consists of at least one hypomanic, manic, or mixed episode. *Mixed episodes* represent a simultaneous mixture of depressive and manic or hypomanic manifestations. Although a minority of patients experience only manic episodes, most bipolar disorder patients experience episodes of both polarity. Manias predominate in men, depression and mixed states in women. The bipolar disorders were classically described as psychotic mood disorders with both manic and major depressive episodes (now termed bipolar I disorder), but recent clinical studies have shown the existence of a spectrum of ambulatory depressive states that alternate with milder, short-lived periods of hypomania rather than full-blown mania (bipolar II disorder). Bipolar II disorder, which is not always easily discriminable from recurrent major depressive disorder, illustrates the need for more research to elucidate the relation between bipolar disorder and major depressive disorder.

Dysthymia and Cyclothymia Clinically, major depressive episodes often arise from a low-grade, intermittent, and protracted depressive substrate known as dysthymic disorder. Likewise, many instances of bipolar disorders, especially ambulatory forms, represent episodes of mood disorder superimposed on a cyclothymic background, which is a biphasic alternating pattern of numerous brief periods of hypomania and numerous brief periods of depression. Dysthymic and cyclothymic disorders represent the two prevalent subthreshold mood conditions roughly corresponding to the basic temperamental dysregulations described by Kraepelin and Ernst Kretschmer as predisposing to affective illness.

It is not always easy to demarcate full-blown syndromal episodes of depression and mania from their subthreshold counterparts commonly observed during the interepisodic periods. The subthreshold conditions appear to be fertile terrain for interpersonal conflicts and postaffective pathological character developments that may ravage the lives of patients and their families. In North America many such patients end up being labeled with borderline personality disorder, which unfortunately often tends to obscure the affective origin of the presenting psychopathology.

Cyclothymic and dysthymic conditions also exist in the community without progression to full-blown mood episodes. As such, they are best considered, respectively, as trait bipolar and trait depressive conditions. Understanding the factors that mediate transition from trait to clinical state is important for preventing manic and major depressive episodes.

Other Subthreshold Mood States Epidemiological studies both in Europe and North America have also revealed other subsyndromal conditions with depressive and hypomanic manifestations with few symptoms (oligosymptomatic mood states) and of short duration (brief episodes). Variously referred to as “minor,” “subsyndromal,” “brief,” or “intermittent,” these descriptions do not merely represent arbitrary lowering of diagnostic thresholds, but herald increasing realization of their importance in early detection of at-risk individuals, as has happened in other medical fields (e.g., diabetes mellitus and essential hypertension). If disabling mood disorders afflict 5 to 8 percent of the general population (Epidemiologic Catchment Area [ECA] study), milder but still clinically significant mood disorders would raise lifetime rates to 17 percent (National Comorbidity Study [NCS]); if subclinical mood states are added, that figure doubles to involve a third of the general population (as reported, for instance, by Kenneth Kendler and colleagues).

Comorbidity Mood disorders overlap considerably with anxiety disorders. As summarized in an NIMH monograph, anxiety disorders can occur during an episode of depression, may be a precursor to the depressive episode, and, less commonly, may occur during the future course of a mood disorder. Those findings suggest that at least some depressive disorders share a common diathesis with certain anxiety disorders. More-recent clinical experience suggests intriguing comorbidity patterns between bipolar II disorder on one hand and panic, obsessive-compulsive, and social phobic states on the other. Furthermore, bipolar I and II disorders are particularly likely to be complicated by use of alcohol, stimulants, or both. In many cases the alcohol or substance abuse represents an attempt at self-treatment of the mood disorder. Finally, physical illness—both systemic and cerebral—occurs in association with depressive disorders with a greater frequency than expected by chance alone. Unless properly treated, such depression negatively impacts on the prognosis of the physical disorder. More provocatively, there is current reawakening in the contribution of cerebral and cardiovascular factors to the origin of late-onset psychotic depressions (previously classified as “involutional melancholia”).

An integrated framework of pathogenesis is necessary for understanding psychopharmacological, somatic, and psychotherapeutic approaches in the clinical management of patients with mood disorders. A historical perspective on current developments is also a valuable lesson in the study of mood disorders.

GRECO-ROMAN DESCRIPTIONS OF MELANCHOLIA AND MANIA

Much of what is known today about mood disorders was described by the ancient Greeks and Romans, who coined the terms *melancholia* and *mania* and noted their relation. The ancients also hypothesized a temperamental origin for those disorders. Much of modern thinking about mood disorders (e.g., the work of the French and German schools in the middle and latter part of the nineteenth century, which influenced current British and American concepts) can be traced back to these ancient concepts.

Melancholia Hippocrates (460–357 BC) described melancholia (“black bile”) as a state of “aversion to food, despondency, sleeplessness, irritability, and restlessness.” Thus, in choosing the name of the condition, Greek physicians (who may have borrowed the concept from ancient Egyptians) postulated the earliest biochemical formulation of any mental disorder. They believed that the illness often arose from the substrate of the somber melancholic temperament which, under the influence of the planet Saturn, made the spleen secrete black bile, ultimately leading to mood darkening through its influence on the brain. Greek descriptions of the clinical manifestations of depression and of the temperament prone to melancholia are reflected in the DSM-IV and in the subdepressive lethargy, self-denigration, and habitual gloom of the person with dysthymic disorder.

One Hippocratic aphorism recognized the close link between anxiety and depressive states: “Patients with fear of long-standing are subject to melancholia.” According to Galen (131–201 AD), melancholia manifested in “fear and depression, discontent with life, and hatred of all people.” A few hundred years later another Roman, Aurelianus, citing the now lost works of Soranus of Ephesus, amplified the role of aggression in melancholia (and its link to suicide) and described how the illness assumed delusional coloring: “Animosity toward members of the household, sometimes a desire to live and at other times a longing for death, suspicion on the part of the patient that a plot is being hatched against him.”

In addition to natural melancholia, which presumably arose from an innate predisposition to overproduce the dark humor and led to a more severe form of the malady, Greco-Roman medicine recognized such environmental contributions to melancholia as immoderate consumption of wine, perturbations of the soul due to the passions (e.g., love), and disturbed sleep cycles. Autumn was considered the season most disposing to melancholy.

Mania A state of raving madness with exalted mood was noted by the ancient Greeks, although it referred to a somewhat broader group of excited psychoses than in modern nosology. Its relation to melancholia was probably noted as early as the first century BC, but according to Aurelianus, Soranus discounted it. Nonetheless, Soranus had observed the coexistence of manic and melancholic features during the same episode, consisting of continual wakefulness and fluctuating states of anger and merriment, sometimes of sadness and futility. Soranus thus seemed to have described what today are called mixed episodes in DSM-IV and ICD-10. Natural melancholy was generally considered a chronic disorder, but Soranus noted the tendency for attacks to alternate with periods of remission.

Although others prior to him hinted at it, Aretaeus of Cappadocia (ca. 150 AD) is generally credited with making the connection between the two major mood states: “It appears to me that melancholy is the commencement and a part of mania.” He described the cardinal manifestations of mania as it is known today:

> There are infinite forms of mania but the disease is one. If mania is associated with joy, the patient may laugh, play, dance night and day, and go to the

> market crowned as if victor in some contest of skill. The ideas the patients have are infinite. They believe they are experts in astronomy, philosophy, or poetry.

Aretaeus described the extreme psychotic excitement that could complicate the foregoing clinical picture of mania:

> The patient may become excitable, suspicious, and irritable; hearing may become sharp... get noises and buzzing in the ears; or may have visual hallucinations; bad dreams and his sexual desires may get uncontrollable; aroused to anger, he may become wholly mad and run unrestrainedly, roar aloud; kill his keepers, and lay violent hands upon himself.

Noting the fluctuating nature of symptoms in the affectively ill, Aretaeus commented

> They are prone to change their mind readily; to become base, mean-spirited, illiberal, and in a little time extravagant, munificent, not from any virtue of the soul, but from the changeableness of the disease.

Aretaeus was thus keenly aware of the characterological distortions so commonly manifested during the different phases of cyclical mood disorders.

Finally, consolidating the knowledge of several centuries, Aretaeus described mania as a disease of adolescent and young men given intermittently to “active habits, drunkenness, lechery” and an immoderate lifestyle (what today might be called cyclothymic disorder). Exacerbations were most likely to occur in the spring.

Affective Temperaments The concept of health and disease in Greco-Roman medicine was based on harmony and balance of the four humors, of which sanguine humor was deemed the healthiest. But even a desirable humor like blood, which made people habitually active, amiable, and prone to jest, could in excess lead to the pathological state of mania. The melancholic temperament, dominated by black bile and predisposed to pathological melancholia, was described as lethargic, sullen, and given to brooding or contemplation; its modern counterparts are depressive personality disorder (now in a DSM-IV appendix) and its clinical expression as dysthymic disorder (included in both ICD-10 and DSM-IV). A long tradition dating back to Aristotle (384–322 BC) attributed creative qualities to the otherwise tortured melancholic temperament in such fields as philosophy, the arts, poetry, and politics. The remaining two temperaments, choleric and phlegmatic, were less desirable, as yellow bile made people choleric (irritable, hostile, and given to rage) and phlegm made them phlegmatic (indolent, irresolute, and timid). The choleric and phlegmatic temperaments would probably be recognized today as borderline personality disorder and avoidant or schizoid personality disorder, respectively.

Many of the original Greek texts on melancholia were transmitted to posterity through medieval Arabic texts such as those of Ishaq Ibn Imran and Avicenna (and their Latin rendition by Constantinus Africanus). In describing different affective states, Avicenna developed the theory of the temperaments to its fullest. He speculated that a special form of melancholia supervened “if black bile be mixed with phlegm” when the illness was “coupled with inertia, lack of movement, and quiet.” Further, mania was not necessarily linked to the sanguine (what today is termed “hyperthymic”) temperament, as many forms of excited madness were believed to represent a mixture of black and yellow bile.

Avicenna further observed that the appearance of anger, restlessness, and violence heralded the transition of melancholia to mania. Those elaborations on Galen's tempermental types might be considered the forerunners of current personality dimensions, deriving mood states from various mixtures of neuroticism and introversion-extroversion. (What both ICD-10 and DSM-IV describe as cyclothymic disorder represents the intense mood lability of high neuroticism coupled with cyclic alternation between extroversion and introversion). Speculation on how diverse depressive phenomena could be understood as a mix of humors anticipated modern multiple-transmitter hypotheses of depression. Ishaq Ibn Imran summarized the existing knowledge of melancholia by considering the interaction of genetic factors (“injured prenatally as the result of the father's sperm having been damaged”) with a special temperament given to “mental overexertion”—though not necessarily physical overactivity—that in turn was associated with “disruption of the correct rhythms of sleeping and waking.” Those views, too, have a very modern ring to them.

MODERN ERA

The first English text (Fig. 14.1-2) entirely devoted to affective illness was Robert Burton's *Anatomy of Melancholy*, published in 1621. A scholarly review of medical and philosophical wisdom accumulated in past centuries, it also anticipated many modern developments. The concept of affective disorder endorsed by Burton was rather broad (as it always has been in the United Kingdom), embracing mood disorders and many disorders today considered somatoform disorders, including hypochondriasis. Although he described “causeless” melancholias, Burton also categorized the various forms of love melancholy and grief. Particularly impressive was his catalogue of causes, culminating in a grand conceptualization:

FIGURE 14.1-2 Frontispiece of Robert Burton's *Anatomy of Melancholy* (1621).

> Such as have Saturn misaffected in their genitures such as are born of melancholy parents as offend in those six non-natural things, are of a high sanguine complexion are solitary by nature, great students, given to much contemplation, lead a life out of action, are most subject to melancholy. Of sexes both, but men more often. Of seasons of the year, autumn is most melancholy. Jobertus excepts neither young nor old.

Burton's six nonnatural things referred to such environmental factors as diet, alcohol, biological rhythms, and perturbations induced by passions such as intense love. Burton himself did not definitively indicate age prevalences. Like nearly all of his predecessors, he favored male (rather than the currently reported female) preponderance. Finally, Burton considered both the melancholic (contemplative) and the sanguine (hot-blooded) temperaments to be substrates of melancholia. Burton's work thus linked certain forms of depression with the softer expressions of the manic disposition, or bipolar II disorder, from which he himself appears to have suffered.

The eighteenth and nineteenth centuries introduced humane hospital care of the mentally ill, thereby permitting systematic clinical observation of the psychopathology and outcome of mood disorders.

Concept of Affective Disorder Although Celsus (ca. AD 30) had described “forms of madness that go no further than sadness,” the French alienist Jean-Philippe Esquirol (1840) may have been the first psychiatrist in modern times to suggest that a primary disturbance of mood might underlie many forms of depression and related paranoid psychoses. Until Esquinol's work, melancholia had been categorized as a form of insanity (i.e., ascribed to deranged reasoning or thought disturbance). Esquirol's observations on melancholic patients led him to postulate that their insanity was partial (dominated by one delusion, a monomania) and that “the symptoms were the expression of the disorder of the affections. The source of the evil is in the passions.” He coined the term “lypemania” (from the Greek, “sorrowful insanity”) to give nosological status to a subgroup of melancholic disorders that were affectively based. Esquirol cited Benjamin Rush (1745–1813), the father of American psychiatry, who had earlier described tristimania, a form of melancholia in which sadness predominated.

Esquirol's influence led other European psychiatrists to propose milder states of melancholia without delusions, which were eventually categorized as simple melancholias and ultimately as primary depressions. Such descriptions culminated in the Anglo-Saxon psychiatric term “affective disorder,” coined by Henry Maudsley

(1835–1918), the renowned British psychiatrist after whom the London hospital is named.

Manic-Depressive Illness and the Question of Psychogenic Depressions Although the connection between mania and depression had been sporadically rediscovered since it was first described 2000 years ago, the clinical work that finally established “circular insanity” (Jean-Pierre Falret's term) as “folie à double forme” (Jules Baillarger's term) was undertaken by these two Esquirol disciples in the 1850s. That accomplishment built on Philippe Pinel's reforms, which championed humane treatment of the mentally ill in Paris around the turn of the eighteenth century and emphasized systematic clinical observations of patients, detailed in case records. French alienists made longitudinal observations on the same patient from one psychotic attack into another. Further, Esquirol had introduced chronicling events in statistical tables. Thus, the Hippocratic approach to defining a particular case by its onset, circumstances, course, and outcome was applied by French alienists in studying the affectively ill. The humanitarian reforms introduced in the nineteenth century ensured that standards of general health and nutrition would improve the outlook for the mentally ill—especially those with potentially reversible disorders such as affective disorders—who could now be discharged from the asylums. The French school then, by segregating the nondeteriorating mood disorders from the dementing types of insanity, paved the way for the Kraepelinian system.

Kraepelin's (1856–1926) unique contribution was not so much grouping together all the forms of melancholia and mania, but his methodology and painstaking longitudinal observations, which established manic-depressive illness as a nosological entity and (he hoped) a disease entity. His rationale was (1) the various forms had a common heredity measured as a function of familial aggregation of homotypic and heterotypic cases; (2) frequent transitions from one form to the other occurred during longitudinal follow-up; (3) a recurrent course with illness-free intervals characterized most cases; (4) the superimposed episodes were commonly opposite to the patient's habitual temperament; that is, mania was superimposed on a depressive temperament and depression was superimposed on a hypomanic temperament; and (5) both depressive and manic features could occur during the same episode (mixed states). Kraepelin's synthesis was developed as early as the sixth (1899) edition of his *Lehrbuch der Psychiatrie* and most explicitly stated in the opening passages of the section on manic-depressive psychosis in the eighth edition (published in four volumes, 1909–1915):

> Manic-depressive insanity includes on the one hand the whole domain of so-called periodic and circular insanity, on the other hand simple mania, the greater part of the morbid states termed melancholia and also a not inconsiderable number of cases of confusional insanity. Lastly, we include here certain slight and slightest colorings of mood, some of them periodic, some of them continuously morbid, which on the one hand are to be regarded as the rudiment of more severe disorders, on the other hand, pass over without boundary into the domain of personal predisposition.

For Kraepelin, the core pathology of clinical depression consisted of lowered mood and slowed (retarded) physical and mental processes. In mania, by contrast, the mood was elated and both physical and mental activity accelerated. His earlier observations on what he termed “involutional melancholia” (referring to 40- to 65-year-old patients with extreme anxiety, irritability, agitation, and delusions) had led him to separate that entity from the broader manic-depressive rubric. But, in the eighth edition of *Lehrbuch der Psychiatrie*, he united melancholia with the manic-depressive group, with the justification that it was a special form of mixed state and that follow-up conducted by his pupil Dreyfus had demonstrated unmistakable excited phases.

The classification of depressive disorders is still evolving. Karl Leonhard in 1957, Jules Angst in 1966, Carlo Perris in 1966, and George Winokur, Paula Clayton, and Theodore Reich in 1969, working independently in four different countries, proposed that depressive disorders without manic or hypomanic episodes (major depressive disorder) that appear in middle age and later are distinct from depressive episodes that begin at earlier ages and alternate with manic or hypomanic episodes (bipolar disorder). The main difference between the two affective subtypes is the greater familial loading for mood disorder—especially for bipolar disorder—among bipolar disorder probands.

Kraepelin had conceded the occurrence of psychogenic states of depression occasioned by situational misfortune. Manic-depressive illness, on the other hand, he believed to be hereditary. Yet he could not document postmortem anatomopathological findings in the brains of manic-depressive patients. Therefore, manic-depression had to be considered a functional mental disorder in which brain disturbances were presumed to lie in altered physiological functions. Such biological factors were deemed absent in the psychogenic depressions. Thus, Kraepelin's classification of mood disorders is both dualistic and unitary. It is dualistic to the extent that he designated them as either psychologically occasioned or somatically caused. It is unitary with respect to disorders in the latter group, which have been termed *endogenous affective disorders* (i.e., due to internal biological causes). In other words, Kraepelin restricted the concept of clinical depression to what DSM-IV terms “major depressive disorder with melancholic features.” Moreover, he postulated a continuum between that condition and what DSM-IV and ICD-10 now term “bipolar disorders.”

As summarized in Table 14.1-1, endogenous depressions have been contrasted with those of exogenous cause (i.e., external and, presumably, psychogenic causes). Transitions between the two groups are so frequent, however, that the two-type thesis of depression has been largely abandoned in official classifications in North American psychiatry. In one study conducted by the author's mood clinic team in Memphis during the 1970s, 100 patients with neurotic depression (the prototype of exogenous depression), prospectively followed over 3 to 4 years, developed episodes with endogenous, psychotic, and even bipolar features (Table 14.1-2). Nonetheless, the endogenous-exogenous dichotomous grouping still has some adherents worldwide who continue to research its potential for clinical predictions. Such research generally attempts to validate the various subtypes on the basis of their clinical characteristics rather than presumed cause. Today most mood disorders experts would probably agree that depressive illness has endogenous and exogenous components in most patients presenting clinically. Consensus would be less likely on how to delimit clinical depressive disorder from comorbid disorders such as the various anxiety disorders, substance use disorders, and personality disorders. Clarifying the boundaries between those disorders has emerged as a principal challenge in the classification of mood disorders.

Manic-depressive	Psychogenic
S (somatic) type	J (justified) type
Autonomous	Reactive
Endogenous	Exogenous
Psychotic	Neurotic
Acute	Chronic
Major	Minor
Melancholic	Neurasthenic
Typical	Atypical
Primary	Secondary
Biological	Characterological

Table 14.1-1 Overlapping Dichotomies of Affective Disorders That Are Not Necessarily Synonymous

Diagnosis and Outcome	N*
Manic episode	4
Hypomanic episode	14
Psychotic depression	21
Endogenous depression	36
Episodic course	42
Unstable characterological features	24
Social invalidism	35
Suicide	3

* The total exceeds 100 because more than one outcome was possible in each patient.

Reprinted with permission from Akiskal H, Bitar A, Puzantian V, Rosenthal T, Walker P: The nosological status of neurotic depression: A prospective 3- to 4-year examination in light of the primary-secondary and unipolar-bipolar dichotomies. Arch Gen Psychiatry 35:756, 1978.

Table 14.1-2 Three- to Four-Year Prospective Follow-up in Neurotic Depressions (N = 100)

Cartesian thinking in seventeenth-century France conceptually separated mind from body, thereby providing physicians autonomy over the somatic sphere, free from interference by the Church. The dichotomous paradigm ensured that study of the two aspects of the human organism would not be confounded by the complexities of mind-body interactions. That is one reason why Kraepelin's descriptive observations have proved valuable to subsequent generations of clinicians. Further, his approach exemplifies the best tradition of scientific humanism in medicine: description and diagnostic categorization of an individual patient are necessary for the physician to apply the knowledge gained from past observation of similarly described and diagnosed patients. One limitation to the Kraepelinian approach is that because of its biological reductionism, it is not sufficiently articulate to account for mind-body interactions in the genesis of mental disorders.

Depressions as Psychobiological Affective Reaction Types Bridging the divide between psyche and soma was the ambition of Swiss-born Adolf Meyer (1866–1950), who dominated psychiatry from his chair at Johns Hopkins University during the first half of the twentieth century. Meyer coined the term "psychobiology" to emphasize that both psychological and biological factors could enter into the causation of depressive and other mental disorders. Because of the nascent state of brain science during Meyer's time, he was more adept at biography than biology and therefore paid greater attention to psychosocial causation. He preferred the term *depression* (pressed down) to *melancholia* because of its lack of biological connotation. He conceived of depressive states in terms of unspecified constitutional or biological factors interacting with a series of life situations beginning at birth or even at conception. From that viewpoint arose the unique importance accorded personal history in depressive reactions to life events.

Meyer's terminological revision left a somewhat confusing legacy in that the term *depression* is now applied to a broad range of affective phenomena ranging from sadness and adjustment disorders to clinical depression and bipolar disorders. Repercussions can be seen in the low threshold for diagnosing major depressive disorder in DSM-IV, which makes it difficult to differentiate major depressive disorder from transient life stresses that produce adjustment disorder with depressed mood. Nosological nuances to which Meyerians paid little attention, such as the difference between melancholic depression and more mundane depressions, are not just a matter of semantics. To the extent that those two forms of depression are seen in different clinical settings, hypotheses based on one population may not apply to the other. For instance, uncontrollable traumatic events may have taught study subjects to feel helpless or to view the world in a negative light, but that does not equate with clinical depression; nor does the process appear to be specific to depression. Failure to make such nosological distinctions further clouds interpretations of the results of trials comparing psychotherapy and pharmacotherapy for depressive disorders.

On the other hand, the Meyerian emphasis on biographical factors for the patient represented a more practical approach to depth psychology. Recent sociological interpretations of depression can also be traced to Meyer's work. But in the final analysis, the Meyerian concern for the uniqueness of the individual has proved heuristically sterile. It deemphasizes what is diagnostically common to different individuals, thereby obscuring the relevance of accrued clinical wisdom for the index patient. For that reason the Meyerian approach, after enjoying clinical popularity for several decades in North America, has given way to neo-Kraepelinian rigor. However, the psychobiological vision of bridging biology and psychology, one of the major preoccupations of psychiatric thought and research today, owes much to Meyer's legacy.

CONTEMPORARY MODELS OF DEPRESSION

From classical times through the early part of the twentieth century, advances in understanding mood disorders involved conceptual shifts from supernatural to naturalistic explanations, from reductionistic, unitarian theories of causation to pluralistic theories, and from dualism to psychobiology. Knowledge of those conceptual developments provides a useful base from which to scrutinize more-recent models and concepts of mood disorder, developed later in the twentieth century. The new approaches, derived from competing theoretical positions, have generated models for understanding various aspects of mood disorders, particularly depressive disorders (Table 14.1-3).

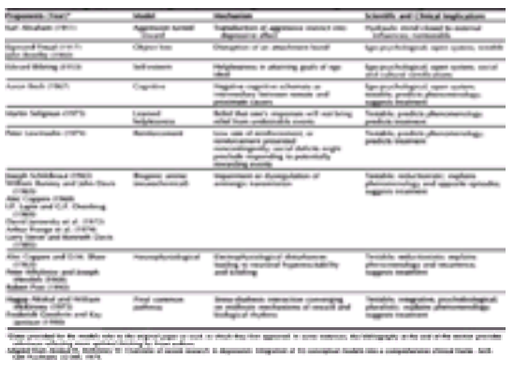

Table 14.1-3 Contemporary Major Models of Depression

The formative influence of early experience as it is dynamically shaped by emerging mental structures during development is the common denominator for the psychoanalytic concepts of psychopathological phenomena. By contrast, behavioral approaches in their more traditional formulations focus on the pathogenetic impact of proximate contexts. The cognitive approaches, which are akin to the behavioral-pathogenetic tradition, nonetheless concede that negative styles of thinking might mediate between proximate stressors and more remote experiences. All three schools—psychoanalytic, behavioral, and cognitive—emphasize psychological constructs in explaining the origin of mood disorders. The biological models, on the other hand, are concerned with defining the somatic mechanisms that underlie or predispose to morbid affective experiences. The schism between psychological and biological conceptualizations is an instance of the mind-body dichotomy that has characterized the Western intellectual tradition since Descartes. After all, psychological and somatic approaches represent merely convenient investigational strategies that attempt to bypass the methodological gulf between mental and neural structures. The ultimate aim is to understand how mood disorders develop within the psychoneural framework of a given person.

Aggression-Turned-Inward Model Sigmund Freud was initially interested in a psychoneural project for all mental phenomena. Limitations of the brain sciences of the day led him to adopt instead a model that relied on a concept of mental function borrowed from physics. The notion that depressed affect is derived from retroflexion of aggressive impulses directed against an ambivalently loved internalized object was actually formulated by his Berlin disciple Karl Abraham and later elaborated by Freud. Abraham and Freud hypothesized that turned-in anger was intended as punishment for the love object that had thwarted the depressed patient's need for dependency and love. Because, in an attempt to prevent the traumatic loss, the object had already been internalized, the patient now became the target of his or her own thanatotic impulses. A central element in those psychic operations was the depressed patient's ambivalence toward the object, which was perceived as a frustrating parent. Aggression directed at a loved object (parent) was therefore attended by considerable guilt. In the extreme such ambivalence, guilt, and retroflexed anger could lead to suicidal behavior.

According to that model, depression was an epiphenomenon of the transduction of thanatotic energy, a reaction that took place in the closed hydraulic space of the mind. Freud's earlier writings had similarly portrayed anxiety as derived from the transformation of dammed-up sexual libido. Although Freud envisioned that psychoanalytic constructs would one day be localized neuroanatomically, the hydraulic mind is a metaphor that does not refer to actual physiochemical space in the brain.

The conceptualization of emotional behavior as an arena of incompatible forces confined to a psyche that is relatively impervious to current influences outside the organism is the major liability of the aggression-turned-inward model and perhaps of orthodox psychoanalysis itself. Although the sexual energy transduction hypothesis of anxiety has been discarded in modern psychoanalytic thought, in modified version the aggression-turned-inward model continues to be used in clinical conceptualization today. The lingering popularity of the model may be due in part to its compatibility with the clinical observation that many depressed patients suffer from lack of assertion and outwardly directed aggressiveness. Yet a substantial number of hostile depressed patients are also encountered in clinical practice (indeed "depression with anger attacks" has been recently described), and clinical improvement in most patients typically leads to decreased, not increased, hostility. Those observations shed doubt on the aggression-turned-inward mechanism as a universal explanation for depressive behavior. Finally, little evidence exists to support the contention that outward expression of anger has therapeutic value in clinical depression.

Outwardly directed hostility in depression is not a new clinical observation: Greco-Roman physicians had noted it. Hostility is best considered a manifestation rather

than a cause of depressive disorder, especially when the disorder is attended by mixed bipolar features. The hostility of the depressed patient can also be understood as an exaggerated reaction to frustrating love objects, as secondary to self-referential attribution, or simply as nonspecific irritability of an ego in affective turmoil; this could in part be a function of a concurrent personality disorder from the erratic cluster. Such common-sense explanations that do not invoke unobservable hydraulic transmutations have greater appeal from heuristic and clinical perspectives.

Object Loss and Depression Object loss refers to traumatic separation from significant objects of attachment. Ego-psychological reformulations of the Abraham-Freud conceptualization of depression have paid greater attention to the impact of such losses on the ego, deemphasizing the id-libidinal and related hydraulic aspects. The depressant impact of separation events often resides in their symbolic meaning for a person rather than in any arbitrary objective weight that the event may have for clinical raters. However, love loss, bereavement, and other exits from the social scene, as defined by the London psychiatrist Eugene Paykel, are presently the concepts most commonly used in practice and research.

Although love melancholy had been described since antiquity, the two affective states were systematically compared for the first time in Freud's 1917 paper on mourning and melancholia. According to current data, the transition from grief to pathological depression occurs in no more than 10 percent of adults and 20 percent of children. These figures suggest that such transition occurs largely in persons predisposed to mood disorders.

John Bowlby of the Tavistock Clinic, London, did a comprehensive clinical investigation of the attachment that the child establishes with the mother or mother substitutes during development, a bond considered the prototype for all subsequent bonds with other objects. Like many psychoanalytic explanations of adult symptom-formation, the object loss model is formulated as a two-step hypothesis, consisting of early breaks in affectional bonds, which provide the behavioral predisposition to depression, and adult losses, which are said to revive the traumatic childhood loss and so precipitate depressive episodes. However, the role of proximate separations in provoking depressive reactions rests on more solid clinical evidence than the hypothesized sensitization resulting from developmental object loss. That realization has led Bowlby to regard childhood sensitization resulting from early deprivation as a generic characterological vulnerability to a host of adult psychopathological conditions.

Compared with aggression turned inward, object loss is more directly relevant to clinical depression; yet it is still pertinent to question whether it is an etiological factor. Studies at the Wisconsin Primate Center have indicated that optimal homeostasis with the environment is most readily achieved when the individual is securely attached to significant others, and the dissolution of such ties appears relevant to the emergence of a broad range of psychopathological disturbances rather than depression per se. A related methodological question is whether object loss operates independently of other etiological factors. For instance, a history of early breaks in attachment may reflect the fact that one or both of the patient's parents had mood disorder, with resultant separation, divorce, suicide, and so forth.

On balance, the ego-psychological object loss model is conceptually superior to its id-psychological counterpart. In postulating an open system of exchange between a person and the environment, the model permits consideration of etiological factors other than separation, such as heredity, character structure, and adequacy of social support—all of which might modulate the depressant impact of adult separation events. Conceptualizing the origin of depression along those lines is in the mainstream of current ideas of adaptation, homeostasis, and disease. An important treatment implication is the value of social support in preventing relapse and mitigating chronicity of depression. That is indeed an ingredient in the interpersonal psychotherapy of depression, which can be conceptualized as a form of brief, focused, and practical psychodynamic therapy.

Loss of Self-Esteem and Depression Reformulation of the dynamics of depression in terms of the ego suffering a collapse of self-esteem represents a further conceptual break with the original id-psychological formulation; depression is said to originate from the ego's inability to give up unattainable goals and ideals. The model further posits that the narcissistic injury that crushes the depressed patient's self-esteem is imposed by the internalized values of the ego rather than the hydraulic pressure of retroflected thanatotic energy deriving from the id. Because the construct of the ego is rooted in social and cultural reality, loss of self-esteem may result from symbolic losses involving power, status, roles, identity, values, and purpose for existence. Thus, the existential and sociocultural implications of depression conceived as a derivative ego state provide the clinician with a far more flexible and pragmatic tool for understanding depressed persons than the archaic hydraulic metaphors related to libidinal vicissitudes. That model represents one of the first attempts to formulate depression in terms that subsequent psychological theory and research could operationalize in more testable form.

Self-esteem is part of the habitual core of the individual and hence is integral to the personality structure. Indeed, low self-esteem conceived as a trait is a major defining attribute of the depressive (melancholic) personality. While it is understandable how such individuals can easily sink into melancholia in the face of environmental adversity, it is not obvious why persons with apparently high self-esteem (e.g., those with hypomanic and narcissistic personalities) also succumb to melancholy with relative ease. To explain such cases, one must invoke an underlying instability in the system of self-esteem that renders it vulnerable to depression. The opposite is also known to occur; that is, manic episodes may develop from a baseline of low self-esteem, as in the case of bipolar disorder patients with antecedent traits of shyness, insecurity, and dysthymia.

The foregoing considerations suggest that the vicissitudes of self-esteem deemed central to the model of depression as loss of self-esteem are manifestations of a more fundamental mood dysregulation. In classical psychoanalysis, such dysregulation is considered of constitutional origin. In general, attempts by psychoanalytic writers to account for bipolar oscillations have not progressed beyond metapsychological jargon, with the possible exception of denial of painful affects as a mechanism in the phenomenology of mania.

Cognitive Model The cognitive model, developed by Aaron Beck at the University of Pennsylvania, hypothesizes that thinking along negative lines (e.g., thinking that one is helpless, unworthy, or useless) is the hallmark of clinical depression. In effect, depression is redefined in terms of a cognitive triad, according to which patients think of themselves as helpless, interpret most events unfavorably vis-à-vis the self, and believe the future to be hopeless. In more recent formulations in academic psychology, these cognitions are said to be characterized by a negative attributional style that is global, internal, and stable and to exist in the form of latent mental schemata that generate biased interpretations of life events.

Because the cognitive model is based on retrospective observations of already depressed persons, it is virtually impossible to prove that causal attributions such as negative mental schemata precede and hence predispose to clinical depression; they can just as readily be regarded as subclinical manifestations of depression. The theoretical importance of the cognitive model lies in the conceptual bridge it provides between ego-psychological and behavioral models of depression. It has also led to a new and widely accepted system of psychotherapy that attempts to alter the negative attributional style, to alleviate the depressive state, and, ultimately, to fortify the patient against future lapses into negative thinking, despair, and depression.

The cognitive model, therefore, has the cardinal virtue of focusing on key reversible clinical dimensions of depressive illness, such as helplessness, hopelessness, and suicidal ideation, while providing a testable and practical psychotherapeutic approach. That approach, however, is less likely to succeed in patients with the full-blown melancholic manifestations of a depressive disorder. It is doubtful that negative cognitions alone could account for the profound disturbances in sleep, appetite, and autonomic and psychomotor functions encountered in melancholic depressions. Further, conceptualizing a multifaceted malady such as depression largely or solely as a function of distorted cognitive processes is reminiscent of pre-Esquirolian notions that emphasized impaired reasoning in the development of depression.

Learned Helplessness Model The learned helplessness model is in some ways an experimental analogue of the cognitive model. The model proposes that the depressive posture is learned from past situations in which the person was unable to terminate undesirable contingencies. The model is based on experiments in dogs that were prevented from taking adaptive action to avoid unpleasant electrical shock and subsequently showed no motivation to escape such aversive stimuli, even when escape avenues were readily available. Armed with evidence from many such experiments, a University of Pennsylvania psychologist, Martin Seligman, postulated a trait of learned helplessness (a belief that it is futile to initiate personal action to reverse aversive circumstances) formed from the cumulation of past episodes of uncontrollable helplessness.

The learned helplessness paradigm is a general one and refers to a broader mental disposition than depression. Thus, it is potentially useful in understanding such diverse conditions as social powerlessness, defeat in sporting events, and posttraumatic stress disorder. In addition, past events might shape a characterological cluster, consisting of passivity, lack of hostility, and self-blame, relevant to certain depressive phenomena. The low hostility observed in some patients during clinical depression could, for instance, be ascribed to the operation of such factors. Learned helplessness could thereby provide plausible links between aspects of personal biography and clinical phenomenology in depressive disorders. Therapeutic predictions for alleviating depression and related psychopathological states capitalize on new cognitive strategies geared to modifying expectations of uncontrollability and the negative attributional style. This illustrates how insights gained from experimental paradigms can be fruitfully combined to address clinical disorders.

Nonetheless, the clinician should be wary of unwarranted clinical extrapolations. For instance, some therapists have argued that the depressed patient's passivity is manipulative, serving to obtain interpersonal rewards. It has also been claimed that such factors have a formative influence on the development of the depressive character. That interpretation appears more relevant to selected aspects of depression than to the totality of the disorder. Depressive behavior and verbalizations clearly have a powerful interpersonal impact, but casting depression as merely a masochistic lifestyle developed to secure interpersonal advantages represents a mechanistic circular argument that could be viewed as disrespectful of the clinical agony of patients with mood disorders. Finally, although most formulations focusing on helplessness have emphasized acquisition through learning, recent experimental research in animals tends to implicate genetic factors in the vulnerability to learning to behave helplessly. The value of the helplessness paradigm may reside in its utility to predict a variety of subthreshold affective disturbances generic to civilian reactions to adversity and trauma.

Depression and Reinforcement Other behavioral investigators, notably Oregon psychologist Peter Lewinsohn, have developed clinical formulations of depression that hinge on certain deficits in reinforcement mechanisms. According to the reinforcement model, depressive behavior is associated with lack of appropriate rewards and, more specifically, with receipt of noncontingent rewards. The model identifies several contributory mechanisms. Some environments may consistently deprive persons of rewarding opportunities, thereby placing them in a chronic state of boredom, pleasurelessness, and, ultimately, despair. That reasoning, however, may offer more insight into social misery than clinical depression. A more plausible postulated mechanism is the provision of rewards that are not in response to the recipient's actions; in other words, the gratis provision of what a person considers undeserved rewards may lead to lowering of self-esteem. Predisposition to depression is formulated in terms of deficient social skills, which are hypothesized to decrease a person's chances of responding to potentially rewarding contingencies in any environment. Indeed, recent research on the relation between personality and depression suggests that such deficits might underlie certain depressive states. Therefore, psychotherapeutic approaches designed to enlarge a patient's repertoire of social skills may prove valuable in preventing some types of depression.

The concepts of depression that have been derived from behavioral methodology and developed in the past several decades are scientifically articulate and therefore testable approaches to clinical depression. Yet, the important distinction between depression on self-report inventories and clinical depression is sometimes overlooked. Further, the behavioral model does not address the distinct possibility that reinforcement deficits may in part represent the psychomotor deficits of depressive illness. Nevertheless, by focusing on reward mechanisms, the behavioral model provides a conceptual bridge between purely psychological and emerging biological conceptualizations of depression.

Biogenic Amine Imbalance Formulation of sophisticated biological explanations of mood disorders had to await development of neurobiological techniques that could probe parts of the brain involved in emotions. Although the complex physiology of the limbic-diencephalic centers of emotional behavior generally cannot be directly observed in humans, much has been learned from animal work. The limbic cortex is linked with both the neocortex, which subserves higher symbolic functions, and the midbrain and lower brain centers, which are involved in autonomic control, hormonal production, and sleep and wakefulness. Norepinephrine-containing neurons are involved in many functions that are profoundly disturbed in melancholia, including mood, arousal, appetite, reward, and drives. Other biogenic amine neurotransmitters that mediate such functions are the catecholamine dopamine, especially important for pleasure, sex, and psychomotor activity, and the indoleamine serotonin, involved in the regulatory control of affects, aggression, sleep, and appetite, among others. Cholinergic neurons, secreting acetylcholine at their dendritic terminals, are generally antagonistic in function to catecholaminergic neurons.

Although the opioid system might, on experimental and theoretical grounds, also serve as one of the neurochemical substrates for mood regulation, in the author's opinion no cogent model of mood disorders involving that system has appeared to date. Likewise, biochemical formulations of mood disorders have paid relatively little attention to the major excitatory brain neurotransmitter glutamate and the inhibitory neurotransmitter g-aminobutyric acid (GABA).

Biogenic Amine Hypotheses Joseph Schildkraut at Harvard University and William Bunney and John Davis at NIMH published the first formal hypothesis connecting depletion or imbalance of biogenic amines (specifically norepinephrine) and clinical depression. The serotonin counterpart of the model was emphasized in the models proposed by Alec Coppen in England and I. P. Lapin and G. F. Oxenkrug in Russia. Both catecholamine and indoleamine hypotheses were essentially based on two sets of pharmacological observations. First, reserpine (which decreases blood pressure by depleting biogenic amine stores) precipitates clinical depression in some patients. Second, antidepressant medications (which alleviate clinical depression) raise the functional capacity of the biogenic amines in the brain. This style of thinking is known as the pharmacological bridge, extrapolating from evidence on mechanism of drug action to the neurotransmitter pathologies presumed to underlie a given psychiatric disorder. Such pharmacological strategies have been of heuristic value in developing research methods for investigation of mood disorders and schizophrenia. Indeed, the research methodology developed by the relatively few investigators working in the area in the past three decades is among the most elegant in the history of psychiatry.

Variations of the biogenic amine model assign somewhat different relative weights to the biogenic amines norepinephrine and serotonin in the development of pathological mood states. Arthur Prange and colleagues at the University of North Carolina formulated a permissive biogenic amine hypothesis in which serotonin deficits permit expression of catecholamine-mediated depressive or manic states. That hypothesis was supported by subsequent animal research showing that an intact serotonin system is necessary for optimal functioning of noradrenergic neurons. Omission of tryptophan from the diet of antidepressant-responsive depressed patients may annul the efficacy of the antidepressant; among healthy volunteers, that special diet also induces sleep electroencephalographic characteristics of clinical depression. Although such findings are provocative, the precursor-loading strategy to increase the brain stores of serotonin (e.g., with L-tryptophan) has not been unequivocally successful in reversing clinical depression. Dietary loading with catecholamine precursors has fared even worse than serotonin-precursor loading in the treatment of depression.

The cholinergic-noradrenergic imbalance hypothesis proposed by David Janowsky and colleagues represents yet another attempt to elucidate the roles of biogenic amines. This hypothesis, along with the related cholinergic supersensitivity hypothesis developed by J. Christian Gillin, has been tested extensively at the University of California at San Diego. Subsequent formulations by Larry Siever and Kenneth Davis at the Mount Sinai Hospital in New York have refocused on noradrenergic dysregulation. The model envisions oscillation from one output mode to the other at different phases of depressive illness. In a provocative extrapolation from that model, bipolar depression would have low noradrenergic output, but many instances of major depressive disorder, just like some anxiety disorders, could be biochemically conceptualized as high-output conditions.

Despite more than three decades of extensive research and indirect evidence, however, no deficiency or excess of biogenic amines in specific brain structures has been shown to be necessary or sufficient for the occurrence of mood disorders. It has been possible neither to confirm the putative role of central norepinephrine in depression nor to discard it altogether. The role of dopamine as formulated by the Italian pharmacologist Gian Luigi Cessa, though studied less extensively than that of norepinephrine, deserves greater recognition: it might have relevance to atypical and bipolar depression as well as mania.

Except for preliminary data from a small brain imaging study showing blunted serotonin responsivity in prefrontal and temporoparietal areas in unmedicated major depressive disorder patients, the evidence for serotonergic disturbance in depression is based on indirect evidence. Moreover, the putative permissive role of serotonin is better documented for aggressive suicide attempts. Serotonergic dysfunction might subserve other conditions characterized by lack of inhibitory control, among them obsessive-compulsive disorder, panic disorders, bulimia nervosa, certain forms of insomnia, alcoholism (alcohol abuse or dependence), and a host of impulse-ridden personality disorders. Such considerations have led Dutch psychiatrist Herman van Praag and his colleagues to postulate a dimensional neurochemical disturbance generic to a large group of disorders within the traditional nosology. This hypothesis might be variously regarded as a challenge to psychiatric nosology or as a statement of the need to supplement clinical classification with biochemical parameters.

It is implied that the foregoing biochemical faults are genetically determined. Although biogenic amine models of mood disorders were developed retrospectively from the pharmacologic action of antidepressant and thymoleptic agents, they have stimulated development of new classes of antidepressants with more selective action on specific neurotransmitter receptors. Their introduction has virtually revolutionized the treatment of depression. Yet the fundamental biochemistry of mood disorders is still far from being understood. Curiously, though selective in action, the new compounds working on the serotonin system have broad effectiveness in a variety of mood-related conditions such as dysthymic disorder, obsessive-compulsive disorder, panic disorder, social phobia, borderline personality disorder, and bulimia nervosa. Such data indirectly favor the hypothesis of an underlying biological commonality to several of these disorders. New antidepressants with dual action on both serotonergic and noradrenergic receptors and emerging data on their possible greater efficacy in melancholic depressions do suggest that the biochemistry of mood disorders involves more-complex dysregulation than is implied in single-neurotransmitter hypotheses. The work of George Henninger and colleagues at Yale University further suggests that monoamines better explain how antidepressants facilitate recovery from depression than the fundamental causes of depression.

Emerging biochemical paradigms are moving away from distal biochemical lesions to focus on molecular perturbations closest to the putative genetic underpinnings of mood disorders. Originally tied to the mechanism of action of mood stabilizers in bipolar disorder, such work is exploring second messenger systems, phosphorylation G proteins, signal transduction, deoxyribonucleic acid (DNA) transcription, and messenger ribonucleic acid (RNA) translation. Again, the search for biochemical

mechanisms is inseparable from the putative mechanism of action of thymoleptic agents.

Neuroendocrine Links Functionally inadequate mobilization of neurotransmitters in the face of continued or repeated stress, as indirectly reflected in pathological modification of noradrenergic and serotonergic receptor function, could represent neurochemical final common pathways of homeostatic failure. Such mechanisms could also provide links with psychoendocrine dysfunction; the hypothesized neurotransmitter deficits may underlie the disinhibition of the hypothalamic-pituitary-adrenal axis, characterized by steroidal overproduction, the most widely studied endocrine disturbance in depressive illness. When challenged with dexamethasone (Decadron), the altered axis resists suppression, thereby offering Bernard Carroll's team (then at the University of Michigan) the possibility of developing the dexamethasone suppression test (DST) for melancholia (subsequently shown to have uncertain specificity for melancholia). That line of research has culminated in the demonstration by the Emory University's Charles Nemeroff of increased concentrations of corticotropin-releasing factor (CRF) in the cerebrospinal fluid of patients with major depressive disorder. CRF also appears relevant to the pathophysiology of anxiety disorders, such as panic disorder.

Another neuroendocrine index of noradrenergic dysregulation, blunted growth hormone response to the a_2-adrenergic receptor agonist clonidine, likewise points to limbic-diencephalic disturbance. However, studies performed in the United States suggest that it is positive in both endogenous depression and severe anxiety disorder (panic disorder). Thyroid-stimulating hormone (TSH) blunting upon thyrotropin stimulation, another common neuroendocrine disturbance in depression, also shows limited specificity.

What is remarkable, however, is that the DST, clonidine (Catapres), and thyrotropin challenge data in aggregate identify most persons with clinical depression. Such evidence of midbrain disturbance argues for considering clinical depression to be a legitimate disease. The disease concept of depression is further buttressed by computed tomography scans showing enlarged pituitary and adrenal glands.

Stress and Depression The concept of a pharmacological bridge implies two-way traffic. The hypothesized chemical aberrations may be primary or biologically induced. Provision should also be made, however, for the likelihood that psychological events that precipitate clinical depression might induce or initiate neurochemical imbalance in vulnerable subjects. That suggestion is supported by studies in animals, in which separation and inescapable frustration effect profound alterations in the turnover of biogenic amines and in postsynaptic receptor sensitivity. Thus, in genetically predisposed persons, environmental stressors might more easily lead to perturbations of limbic-diencephalic neurotransmitter balance. Finally, in vulnerable individuals, especially during the formative years of childhood, psychological mechanisms might more easily perturb midbrain neurochemistry. Traumatic experiences appear particularly potent in this regard.

Neurophysiological Approaches

Neuronal Hyperexcitability Lithium is known to replace intracellular sodium and hyperpolarize the neuronal membrane, thereby decreasing neuronal excitability. Abnormalities in neuronal electrolyte balance (an excess of residual sodium, defined by radioisotope techniques) and hypothesized secondary neurophysiological disturbances were the focus of British investigations by Alec Coppen and colleagues in the early 1960s. The existing data appear compatible with the hypothesized movement of excess sodium into the neuron during an episode of mood disorder and redistribution toward the pre-illness electrolyte balance across the neuronal membrane during recovery. Intraneuronal sodium leakage is postulated in both depressive and manic disorders but deemed more extreme in the latter. Because the harmonious activity of the neuronal cell and, by implication, that of a group of neurons depends on the electrical gradient maintained across its membrane by differential distribution of sodium, abnormalities in sodium concentrations and transport are hypothetically relevant to the production of an unstable state of neurophysiological hyperexcitability. In formulating their thesis of neurophysiologic arousal in melancholic states, Joseph Mendels and Peter Whybrow (both of whom at various times worked at the University of Pennsylvania) have capitalized on the foregoing electrolyte disturbances. The view that mania represents a more extreme electrophysiological dysfunction in the same direction as depression violates the common-sense notion of symptomatological opposition between the two kinds of disorder, yet it may in part account for the existence of mixed states in which symptoms of depression and mania coexist. The NIMH team led by Frederick Goodwin first showed that a substantial minority of depressed patients with a bipolar substrate respond to lithium salts, which further supports the concept of a neurophysiological common denominator to mania and depression. Perturbations of calcium metabolism also appear limited to bipolar patients. Therapeutic implications of this observation (e.g., the use of calcium channel inhibitors in bipolar I disorder) have not yielded consistent results. Finally, rubidium, another alkali metal, has been explored in the depressive phase of bipolar disorders, again with inconclusive results.

Rhythmopathy European studies have shown that depressed patients are phase advanced in many biological rhythms, including the latency to the first rapid eye movement (REM) in sleep. Shortened REM latency, which has been extensively studied by David Kupfer and colleagues at the University of Pittsburgh, has been proposed as another laboratory test for depressive disorder. Formulations of circadian rhythms by Thomas Wehr and Norman Rosenthal, working at NIMH, have focused on abnormalities on brain regulation temperature, activity, and sleep cycles. Others have investigated the role of the pineal hormone melatonin. These neurophysiologic considerations have paved the way for new therapeutic opportunities.

Sleep deprivation and exposure to bright white light can correct phase disturbances and thereby terminate depressive episodes, especially in patients with periodic and seasonal depressions. Daniel Kripke, working at the University of California at San Diego, has shown that the average citizen is light deprived and that phototherapy can benefit many forms of depression. Although the specificity and efficacy of these neurophysiological indexes and manipulations for clinical depression require more extensive research, cumulatively they point to midbrain dysregulation as the likely common neurophysiological substrate of depressive disorders. The foregoing considerations further suggest that the ancient Greeks, who ascribed melancholia to malignant geophysical influences, did not indulge in mere poetic metaphor. The ancients had observed the disturbed circadian patterns and advocated their readjustment to restore euthymia.

Affective Dysregulation A major challenge for research in mood disorders is to characterize the basic molecular mechanisms that underlie the neurophysiological rhythmopathies, which in turn might account for the recurrent nature of the affective pathology as envisioned by Kraepelin. This means that in the most typical recurrent forms of the disorders, the constitutional foundations (manifested as cyclothymic and dysthymic traits) are so unstable that the illness may run its entire course more or less autonomously, with the environment largely serving to turn on and off the more florid phases (episodes). The Parisian psychiatrist Jean Delay, a pioneer in psychopharmacology in the 1950s, has also emphasized affective dysregulation as the fundamental pathology in the spectrum of mood disorders. Robert Post (at NIMH) has hypothesized that the electrophysiological substrates could be so kindled that an oligoepisodic disorder initially triggered by environmental stressors could assume an autonomous and polyepisodic course. He hypothesizes that this phenomenon might occur because neuronal perturbations brought about by stressors in the early course of mood disorders get incorporated into the DNA. This fascinating kindling hypothesis, however, does not seem to pertain to common mood disorders, but those with extreme cyclicity. The monograph on manic-depressive illness by Goodwin and Kay Jamison presents in-depth arguments for this cyclical paradigm of thymopathy.

THEORETICAL SYNTHESIS

Pathophysiological Understanding Modern psychobiology attempts to link experience and behavior to the central nervous system. Building conceptual bridges between the psychological and biological approaches to mood disorders requires sophisticated strategies that go beyond the Cartesian notion of limited mind-body interactions through the pineal gland and the generalizations of the Meyerian school.

In collaboration with William McKinney in 1973, the author developed the conceptual framework that considers the affective syndromes as the final common pathway of various psychological and biological processes. The overarching hypothesis is that psychological and biological etiological factors converge in reversible deficits in the diencephalic substrates of pleasure and reward. Those areas of the brain subserve the functions that are disturbed in melancholia and mania. The integrative model links the central chemistry and physiology of reward mechanisms with the object loss and behavioral models of depression, both of which give singular importance to the depressant role of loss of rewarding interpersonal bonds. A key element of the model is the circadian disturbances observed since ancient times in both depressive and manic syndromes. Both syndromes are conceptualized as clinical manifestations of a disordered limbic system with its subcortical and prefrontal extensions. Brain-imaging studies in melancholic patients by Wayne Drevets at Washington University have tentatively visualized limbic disturbances extending into subcortical structures and occurring primarily in those with a familial diathesis for depression. The amygdala plays a key role in this model. Clinical experience and research data suggest that multiple factors described below converge to produce dysregulation leading to the final common pathway of clinical depression and mania.

Heredity Current evidence indicates a significant genetic role in the causation of bipolar and recurrent major depressive disorders. Although it is not known exactly what is inherited, recent research suggests that heritability involves a broad spectrum of disorders, including milder depressive states. Genetic heterogeneity is likely, and may involve inheritance of a single dominant gene with variable penetrance or in greater likelihood, polygenic inheritance. Different genetic mechanisms will probably involve more than one disorder (e.g., depression and generalized anxiety; bipolar I disorder, stimulant abuse, and alcohol dependence; bipolar II disorder, mood disorder with seasonal patern, bulimia nervosa, and borderline personality).

Developmental Predisposition Parents with mood disorders are often in conflict, which may lead to separation, divorce, and suicide. It can be said that heredity often determines the type of environment into which the child predisposed to mood disorder is born. Developmental object loss, although not specifically involved in causing mood disorder, might modify the expression of the illness, possibly by leading to earlier onset, more-severe episodes, and an increased likelihood of personality disorder and suicide attempts.

Temperament Since ancient times, persons prone to mania and melancholia have been described as possessing certain temperamental attributes, representing variations on the theme of what today is subsumed under cyclothymic, dysthymic, and anxious-inhibited temperaments. Many monozygotic twins discordant for full-blown mood disorders studied by Aksel Bertelsen's Danish research team exhibited affective instability with temperamental moodiness, which strongly suggests that such attributes are genetically determined. Research conducted by Kendler's team at the Medical College of Virginia further suggests that several of the temperamental attributes might be transmitted as part of the genetic liability to mood disorders. The author's research has identified such temperaments in the prepubertal offspring of parents with bipolar I disorders, suggesting that they precede by years to decades the overt onset of major mood disorder episodes. The high expressed emotion atmosphere and the negative critical remarks by relatives and affectively unstable patients documented in the recent psychological literature on mood disorders often reflect the interpersonal clashes between patients and their temperamentally intense relatives. Thus, temperaments appear intimately involved in generating much interpersonal friction, emotional arousal, and sleep loss (just to cite common perturbations) thereby eliciting many of the life stressors that precipitate affective episodes. The use of stimulant drugs either to self-treat lethargy or enhance hypomanic traits could further contribute to episode precipitation.

Life Events Most individuals do not develop clinical depression when exposed to environmental adversity. Such adversity seems to play a pathogenic role primarily in those with an affective diathesis. Actually, the work of Kenneth Kendler at the Medical College of Virginia indicates that genetic factors might underlie the depressive disorder patients' susceptibility to life events. Furthermore, current data suggest that social stressors in the onset of depression are more relevant to the first few episodes of the illness. The evidence linking such events to mania is less robust. At any rate, stressful events often appear to be triggered by the temperamental instability that precedes clinical episodes. Interpersonal losses are common events in the lives of individuals with intense temperaments.

Indeed, a recent study by Peter McGuffin's team at the Institute of Psychiatry, London, raised the possibility that one mechanism by which heredity produces depression is by creating environmental adversities in the lives of individuals predisposed to this illness. This work is now replicated by independent groups of investigators. Whatever the origin of environmental adversity, it is common clinical experience that loss represents an important, perhaps even central, theme in clinical depression. Variables that seem to modulate the impact of adult losses include concurrent life events, resultant changes in lifestyle, lack of interpersonal support, deficient social skills, and the symbolic meaning of the putative loss. The research program of George Brown and his followers in London capitalizes on the foregoing considerations, particularly the importance of early and proximate losses in socioeconomically disadvantaged women who lack supportive relationships. However, that conceptualization downplays the degree to which the social context of the depression reflects the dysthymic temperamental liabilities of those depressed women. Recent research indicates that even social support is determined to a considerable degree by the genetic mechanisms that underlie mood disorders.

Biological Stressors Many physical diseases and pharmacological agents are known to precede the onset of both depressive and manic episodes. Like psychosocial stressors, however, they do not generally seem to cause de novo episodes but mobilize them in persons with a personal and family history of mood disorders. Thyroid disturbances have an important role in practice, because they seem to underlie some cases with rapid cycling in bipolar females.

Sex Clinical and epidemiological studies concur in suggesting that women are at higher risk for mood disorders, with the risk highest for depression. This now appears in part a function of anxious-depressive traits represented by neuroticism. These traits have strong genetic determinants. Women have higher concentrations of monamine oxidase (the enzyme that breaks down monamine transmitters) in the brain and more precarious thyroid status. In addition, low estrogen and high progesterone concentrations have been postulated as possible mediating factors in postpartum depressions, premenstrual accentuation of affective instability, and women's vulnerability to the depressant effect of steroidal contraceptives. Personality factors might also be relevant to the sex differences in depression. In recent collaborative work with University of Pisa psychiatrist Giulio Perugi, the author has proposed the hypothesis that female sex might favor greater expression of dysthymic attributes, whereas hyperthymic traits appear favored by male sex. Those considerations tend to parallel, respectively, the “ruminative” and “active” cognitive response styles reported by Stanford's Susan Nolen-Hoeksema to distinguish the sexes. What specific sex-related biographical factors might interact with sex-related biological factors to produce such trait differences is presently largely unknown. An intriguing possibility is that women, because of their temperamental inclination to depressive cognitions, might more adversely respond to childhood adversities, as well as being more specifically vulnerable to adult stressors related to bonding with men and child rearing. Research by Mark George and colleagues has raised the provocative possibility that women overrespond to sad circumstances over a lifetime, thereby permanently altering anterior limbic and prefrontal brain function in a depressive direction.

The integrative model presented here (Fig. 14.1-3) goes beyond the general provisions of the unified approach developed a quarter century earlier. It is submitted that at least in the highly recurrent forms of the malady, affective temperaments represent the intermediary stage between remote (hereditary) and proximate (stressful) factors and that limbic-diencephalic dysfunction is best characterized as the biological concomitant of the clinical manifestations of the affective syndromes. Like the temperamental dysregulations, these biological disturbances represent a putative stage in the pathogenetic chain. They emerge as temperamental instabilities that react to, provoke, or invite life events, substance use, and alterations in circadian rhythms, which in turn appear to usher in the behavioral, emotional, and cognitive manifestations of the illness.

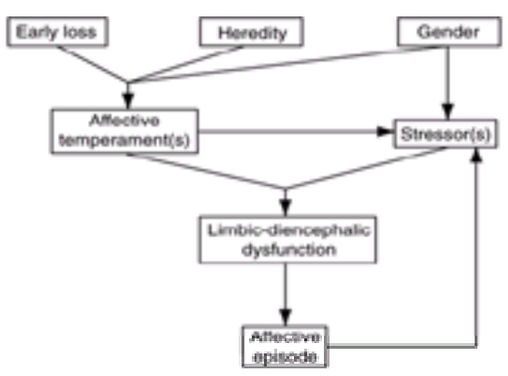

FIGURE 14.1-3 An integrative pathogenetic model of mood disorders.

THERAPEUTIC PERSPECTIVES The foregoing integrative model envisions the joint use of somatic-pharmacological and psychosocial interventions. Although the milder forms of mood disorders can be managed with psychotherapy, somatic treatments are usually required to reverse the biological disturbances in melancholia before the patient can respond to interpersonal feedback. Depressive disorders with psychotic features often necessitate more definitive somatic interventions such as electroconvulsive therapy. Continued psychopharmacological treatment is also effective in decreasing rates of relapse and future recurrence in most.

Psychosocial therapy by skilled clinicians can provide support, combat demoralization, change maladaptive self-attributions, and improve conjugal and vocational functioning. Whether such therapy can also modify personality traits to fortify the patient against new episodes is a future research challenge. In the author's view, it may prove more profitable to attempt to help patients explore professional and object choices that match their temperamental proclivities and assets, which in turn might provide them greater harmony and adaptation in life. Although much needs to be learned about the indications for medication and psychotherapy in different subtypes of mood disorders, research to date not only does not support a negative interaction between the two forms of treatment but on selected parameters suggests additive and even synergistic interaction. There is a great need for patients, their families, and clinicians to understand how a biologically driven illness like depression should be approached from a pragmatic psychotherapeutic perspective.

The challenge for psychiatric research in the decade ahead is to elucidate the basic mechanisms whereby the predisposing, precipitating, and mediating variables reviewed here and others yet to be identified interact to produce the final common path of decompensation in melancholia. Because of the heterogeneity of depressive conditions presenting as a psychobiological final common clinical syndrome and because antidepressant agents, irrespective of specificity to one or another biogenic amine, are about equally effective in two-thirds of those with depressive disorders, the antidepressant agents may be acting not on the primary lesions of these disorders but on a neurochemical substrate distal to the underlying biological faults. The choice of antidepressants is still highly determined by the

side effect profile least objectionable to a given patient's physical status, temperament and lifestyle. That so many different classes of antidepressants—with different mechanisms of action—have been marketed in the 1990s represents indirect evidence for heterogeneity of biochemical lesions. The investigation of central neurotransmitter receptor function continues to occupy much current effort to delineate the putative mechanism of antidepressant action and side effects of classic agents as well as the new compounds which have made the treatment of depression “clinician and patient friendly.” Whether study of specific recepters will unravel the molecular mystery of depression remains to be seen. During the past decade, studies have begun on antidepressant and mood-stabilizing effects on molecular mechanisms believed to be closer to the “genetic underpinnings” of mood disorders. Herein is the promise of the future, a new generation of psychiatrists conversant with both clinical phenomenology and molecular biology. Data suggest that the biological specificity of genetic factors in mood disorders might be translated into distinct temperamental dysregulations, which in turn might predispose to different affective subtypes.

Returning to the therapeutic arena, mounting evidence indicates that in depressed patients with bipolar disorder, antidepressants might provoke mixed episodes, hypomanic episodes, or both, and possibly increase later cycling. The kindling-sensitization model suggests the utility of anticonvulsant medication on episode escalation and might represent yet another example of pathophysiological intervention. Whatever the merit of this model, the last decade has witnessed intense clinical and research interest and Food and Drug Administration approval of clinical introduction of divalproex (Depakene) for bipolar I disorder, and many other promising anticonvulsants are being developed for that disorder. Anticonvulsant mood stabilizers appear to possess a broad spectrum of activity on bipolar disorders, including mixed dysphoric and rapid-cycling forms. Lithium, by contrast, seems more specific to euphoric mania.

Psychoeducational interventions geared to disturbed rhythms of the disorder represent another example of rational therapeutics. Mood clinics should help patients and their significant others to dampen stimulation so that it is kept at an optimal level for depressed patients with cyclothymic traits. All offending drugs (e.g., cocaine, caffeine, and sedative-hypnotic agents) should gradually be eliminated, and circadian disruptions and sleep loss minimized. The greater challenge is learning how to curb the ill-advised actions of patients with cyclical depressions. Psychoeducation and psychotherapy have the task of ameliorating the resulting social problems. Compliance with mood-stabilizer regimens that for many would attenuate episodes and prevent such sequelae is difficult to achieve. Research on compliance-enhancing techniques is needed for the more efficient use of mood stabilizers.

It is tempting to suggest that biogenic amines, the “humors” of modern psychobiology, play the same heuristic role as the ancient humors did for many centuries. The black humor, appropriately evoked in the construct of melancholia in DSM-IV, may not have the same claim for etiological relevance to depressive disorders as norepinephrine and serotonin, but at least has a classical heritage. Dopamine, by contrast, may represent the sanguine humor that drives hypomanic temperaments and manic behavior. When genetic factors contributing to clinical depression and mania are discovered, in all likelihood they will be more linked to temperamental dispositions than to fullblown affective disease phenotypes. The clinician will still need to interpret the myriad of influences that impinge on such inclinations to produce disease in an individual patient. That is, fundamental scientific advances in mood disorders, rather than diminishing the role of practitioners, will actually increase it.

In any discipline, scientific truth is a function of its technology, but understanding the phenomena under consideration is a matter of philosophical temperament that seeks integration and the hope for a unified vision. Research into the causes and treatment of mood disorders has generated abundant recent data suitable for integration into theory and practice, and conceptualizing the origin and treatment of mood disorders can no longer be justified on the grounds of ideological preference alone.

SUGGESTED CROSS-REFERENCES

Anxiety disorders are covered in Chapter 15. The other sections of Chapter 14 cover the various aspects of mood disorders in detail. Epidemiology is the subject of Section 14.2; neurobiologic aspects are the focus of Section 14.4; Section 14.3 is a discussion of genetic aspects; psychodynamic aspects are the subject of Section 14.5. Clinical features are covered in Section 14.6, somatic treatment in Section 14.7 and Section 14.8, and a discussion of psychosocial treatments concludes the chapter in Section 14.9.

SECTION REFERENCES

Abraham K: Notes on the psychoanalytical investigation and treatment of manic-depressive insanity and allied conditions. In *Selected Papers of Karl Abraham.* Hogarth Press, London, 1948.

Ahrens B, Müller-Oerlinghausen M, Schou M, Wolf T, Alda E, Grof P, Grof G, Lenz C, Simhandl K, Thau P, Vestergaard R, Möller HJ: Excess cardiovascular and suicide mortality of affective disorders may be reduced by lithium prophylaxis. J Affect Disord *33:*67, 1995.

Akiskal HS: Toward a temperament-based approach to depression: Implications for neurobiologic research. Adv Biochem Psychopharmacol *49:*99, 1995.

Akiskal HS, Downs J, Jordan P, Watson S, Daugherty D, Pruitt DB: Affective disorders in referred children and younger siblings of manic-depressives: Mode of onset and prospective course. Arch Gen Psychiatry *42:*996, 1985.

*Akiskal HS, McKinney WT: Depressive disorders: Toward a unified hypothesis. Science *182:*20, 1973.

*Andrews G, Slade T, Peters L: Classification in psychiatry: ICD-10 versus DSM-IV. Br J Psychiatry *174:*3, 1999.

Aretaeus of Cappadocia: *The Extant Works of Aretaeus, the Cappadocian*, F Adams, editor-translator. Sydenham Society, London, 1856.

Beck AT: *Depression: Causes and Treatment.* University of Pennsylvania Press, Philadelphia, 1967.

Bertelsen A, Harvald B, Hauge M: A Danish twin study of manic-depressive disorders. Br J Psychiatry *130:*330, 1977.

Bowlby J: Process of mourning. Int J Psychoanal *45:*317, 1961.

Brown GW, Harris T: Social origins of depression: A study of psychiatric disorder in women. Tavistock, London, 1978.

Bunney WE Jr, Davis JM: Norepinephrine in depressive reactions: A review. Arch Gen Psychiatry *13:*483, 1965.

Carlson GA, Strober M: Manic-depressive illness in early adolescence. A study of clinical and diagnostic characteristics in six cases. J Am Acad Child Psychiatry *17:*138, 1978.

Carroll BJ, Feinberg M, Greden JF, Tarika J, Albala AA, Haskett RF, James NM, Kronfol Z, Lohr N, Steiner M, de Vigne JP, Young E: A specific laboratory test for the diagnosis of melancholia: Standardization, validation, and clinical utility. Arch Gen Psychiatry *38:*15, 1981.

Coppen A: The biochemistry of affective disorders. Br J Psychiatry *113:*1237, 1967.

Cytryn L, McKnew D: *Growing Up Sad: Childhood Depression and Its Treatment.* Norton, New York, 1996.

Delay J: *Les Déréglements de L'humeur.* Presses Universitaires de France, Paris, 1946.

Drevets WC, Videen TO, Price JL, Preskorn SH, Carmichael ST, Raichle ME: A functional anatomical study of unipolar depression. J Neurosci *12:*3628, 1992.

Freud S: Mourning and melancholia. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 4. Hogarth Press, London, 1975.

George MS, Ketter TA, Parekh PI, Horwitz B, Herscovitch P, Post RM: Brain activity during transient sadness and happiness in healthy women. Am J Psychiatry *152:*341, 1995.

Gershon ES, Hamovit J, Guroff JJ, Dribble E, Leckman JF, Sceery W, Targum SD, Nurnberger JI, Goldin LR, Bunney WE: A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Arch Gen Psychiatry *39:*1157, 1982.

Gessa GL: Dysthymia and depressive disorders: Dopamine hypothesis. Eur Psychiatry *11:*123s, 1996.

Gilbert P: *Human Nature and Suffering.* Erlbaum, New York, 1989.

Gillin JC, Sitaram N, Duncan WC: Muscarinic supersensitivity: A possible model for the sleep disturbance of primary depression? Psychiatry Res *1:*17, 1979.

*Goodwin FK, Jamison KR: *Manic-Depressive Illness.* Oxford University Press, New York, 1990.

Heninger GR, Delgado PL, Charney DS: The revised monoamine theory of depression: A modulatory role for monoamines, based on new findings from monoamine depletion experiments in humans. Pharmacopsychiatry *29:*2, 1996.

Jackson SW: *Melancholia and Depression: From Hippocratic Times to Modern Times.* Yale University Press, New Haven, 1986.

Janowsky DS, El-Yousef MK, Davis JM, Sekerke HJ: A cholinergic-adrenergic hypothesis of mania and depression. Lancet *1:*632, 1972.

Keller MB, Klerman GL, Lavori PW, Fawcett JA, Coryell W, Endicott J: Treatment received by depressed patients. JAMA *248:*1848, 1982.

Kendler KS, Gardner CO: Boundaries of DSM-IV: An evaluation of DSM-IV criteria. Am J Psychiatry *155:*172, 1998.

Kendler KS, Karkowski-Shuman L: Stressful life events and generic liability to major depression: Genetic control of exposure to the environment? Psychol Med *27:*359, 1997.

Kendler KS: Social support: A genetic-epidemiological analysis. Am J Psychiatry *154:*1398, 1997.

Kendler KS, Eaves LJ, Walters EE, Neale MC, Heath AC, Kessler RC: The identification and validation of distinct depressive syndromes in a population-based sample of female twins. Arch Gen Psychiatry *53:*391, 1996.

*Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: A longitudinal twin study of personality and major depression in women. Arch Gen Psychiatry *50:*853, 1993.

Kessler RC, Zhao S, Blazer DG, Swartz M: Prevalence, correlates, and course of minor depression and major depression in the national comorbidity survey. J Affect Disord *45:*19, 1997.

Klerman GL, Lavori PW, Rice J, Reich T, Endicott J, Andreasen NC, Keller MB, Hirschfield RM: Birth-cohort trends in rates of major depressive disorder among relatives of patients with affective disorder. Arch Gen Psychiatry *42:*689, 1985.

*Kraepelin E: *Manic-Depressive Insanity and Paranoia*, RM Barclay, translator, GM Robertson, editor. Livingstone, Edinburgh, 1921.

Krishman KR: Pituitary size in depression. J Clin Endocrinol Metab *72:*256, 1991.

Kupfer DJ: REM latency: A psychobiologic marker for primary depressive disease. Biol Psychiatry *11:*159, 1976.

Lapin IP, Oxenkrug GF: Intensification of the central serotoninergic processes as a possible determinant of the thymoleptic effect. Lancet *1:*132, 1969.

Leonhard K: *The Classification of Endogenous Psychoses*, R Berman, translator. Irvington, New York, 1979.

Lewinsohn PM, Youngren MA, Grosscup SJ: Reinforcement and depression. In *The Psychobiology of Depressive Disorders; Implications for the Effects of Stress*, RA Depre, editor. Academic Press, New York, 1979.

Lewis A: States of depression: Their clinical and aetiological differentiation. Br Med J *2:*875, 1938.

Mann JJ, Malone KM, Diehl DJ, Perel J, Cooper TB, Mintun MA: Demonstration in vivo of reduced serotonin responsivity in the brain of untreated depressed patients. Am J Psychiatry *153:*174, 1996.

Maser JD, Cloninger CR, editors: *Comorbidity of Mood and Anxiety Disorders.* American Psychiatric Press, Washington, DC, 1990.

McGuffin P, Katz R, Bebbington P: The Camberwell Collaborative Depressive Study: III. Depression and adversity in the relatives of depressed probands. Br J Psychiatry *152:*775, 1988.

Murray CJL, Lopez AD, editors: *The Global Burden of Disease.* World Health Organization, Geneva, 1996.

Nazroo JY, Edwards AC, Brown GW: Gender differences in the onset of depression following a shared life event: A study of couples. Psychol Med *27:*9, 1997.

Nemeroff CB, Widerlov E, Bissette G: Elevated concentrations of CSF corticotropin-releasing factor-like immunoreactivity in depressed patients. Science *226:*1342, 1984.

Nolen-Hoeksema S, Morrow J, Frederickson BL: Response styles and the duration of episodes of depressed mood. J Abnorm Psychol *102:*20, 1993.

Parry B: Reproductive factors affecting the course of affective illness in women. Psychiatr Clin North Am *12:*221, 1989.

Post RM: Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. Am J Psychiatry *149:*999, 1992.

Poznanski E, Zrull JP: Childhood depression: Clinical characteristics of overtly depressed children. Arch Gen Psychiatry *23:*8, 1970.

Prange AJ Jr, Wilson IC, Lynn CW, Alltop LB, Stikeleather RA: L-Tryptophan in mania: Contribution to a permissive hypothesis of affective disorders. Arch Gen Psychiatry *30:*56, 1974.

*Puig-Antich J: Affective disorders in children and adolescents. In *Psychopharmacology: The Third Generation of Progress*, HY Meltzer, editor. Raven, New York, 1987.

Rihmer Z, Rutz W, Pihlgren H: Depression and suicide on Gotland: An intensive study of all suicides before and after a depression-training programme for general practitioners. J Affect Disord *35:*147, 1995.

Schulberg HC: Treating major depression in primary care practice. Eight-month clinical outcomes. Arch Gen Psychiatry *53:*913, 1996.

Seligman MD: *Helplessness: On Depression, Development and Death.* Freeman, San Francisco, 1975.

Siever LJ, Davis KL: Overview: Toward a dysregulation hypothesis of depression. Am J Psychiatry *142:*1017, 1985.

Steffens DC, Krishnan KR: Structural neuroimaging and mood disorders: Recent findings, implications for classification and future directions. Biol Psychiatry *43:*705, 1998.

van Praag HM, Kahn RS, Asnis GM, Wetzler S, Brown SL, Bleich A, Korn ML: Denosologization of biological psychiatry or the specificity of 5-HT disturbances in psychiatric disorders. J Affect Disord *13:*1, 1987.

Wehr TA, Rosenthal NE: Seasonality and affective illness. Am J Psychiatry *146:*829, 1989.

Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam E, Lee CK, Lellouch J, Lépine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen HU, Yeh EK: Cross-national epidemiology of major depression and bipolar disorder. JAMA *276:*293, 1996.

Wilhelm K, Parker G, Hadzi-Pavlovic D: Fifteen years on: Evolving ideas in researching sex differences in depression. Psychol Med *27:*875, 1997.

Weller EB, Weller RA: Bipolar disorder in children: Misdiagnosis, underdiagnosis, and future diagnosis. J Am Acad Child Adolesc Psychiatry *34:*709, 1995.

Wells K, Steward A, Hays R, Burnam M, Rogers W, Daniels M, Berry S, Greenfield S, Ware J: The functioning and well-being of depressed patients: Results from the medical outcomes study. JAMA *262:*914, 1989.

Whybrow P, Mendels J: Toward a biology of depression: Some suggestions from neurophysiology. Am J Psychiatry *125:*45, 1969.

Winokur G, Clayton PJ, Reich T: *Manic-Depressive Illness.* Mosby, St Louis, 1969.

Zis KD, Zis A: Increased adrenal weight in victims of violent suicide. Am J Psychiatry *144:*1214, 1987.

14.2 MOOD DISORDERS: EPIDEMIOLOGY

DAN G. BLAZER II, M.D., FH.D.

INTRODUCTION

Since the 1980s community-based epidemiologic studies of mood disorders throughout the world have been significantly influenced by two large community-based surveys in the United States, the 1981 Epidemiologic Catchment Area Study and the 1991 National Comorbidity Survey. Not only did these surveys provide estimates of the prevalence and distribution of discrete psychiatric disorders (in synchrony with the shifting orientation from symptom burden to operational psychiatric diagnoses reflected in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DMS-III) and its successors, they also demonstrated the feasibility of identifying psychiatric patients by use of assessment procedures that could be administered by nonprofessional interviewers. ECA investigators used the Diagnostic Interview Schedule (DIS) and NCS investigators used a modified version of the Composite International Diagnostic Interview (CIDI). These case assessment procedures have now been used in countries throughout the world, including Canada, Finland, Korea, Taiwan, and New Zealand.

Most of the data presented in this chapter derive from community-based estimates of the frequency, distribution, and correlates of depressive symptoms and major depressive disorder as estimated from these studies. Bipolar disorders receive less attention than they do in clinical studies because community-based epidemiological data are sparse. The lifetime prevalence of mood disorders is high in Western society. Estimates of the lifetime prevalence of mood disorders are presented in Table 14.2-1.

Mood Disorder		Lifetime Prevalence
Depressive disorders		
Major depressive disorder (MDD)	Recurrent, with full interepisode recovery, superimposed on dysthymic disorder	10–25% for women; 5–12% for men
		2–3% of persons with MDD
Dysthymic disorder	Recurrent, without full interepisode recovery, superimposed on dysthymic disorder (double depression)	25–30% of persons with MDD
		3–6%
Bipolar disorders		
Bipolar I disorder		0.4–1.6%
Bipolar II disorder		<1%
	Bipolar I disorder or bipolar II disorder, with rapid cycling	5–15% of persons with bipolar disorder
Cyclothymic disorder		0.4–1.0%

Table 14.2-1 Lifetime Prevalence of Some DSM-IV Mood Disorders

CASE IDENTIFICATION

Case identification (i.e., determining who should be diagnosed as experiencing a mood disorder versus who is only expressing a normal fluctuation in mood) has been at the center of debate regarding the true frequency of mood disorders in the community. This debate derived largely from the significant variance in the estimates of the current and lifetime prevalence of major depressive disorder in the ECA and NCS studies, as demonstrated in Figure 14.2-1, Figure 14.2-2, Figure 14.2-3 and Figure 14.2-4. This debate has relevance to clinicians who treat patients experiencing mood disorders, for they must distinguish normal variations in mood from the mood disorders. The diagnostic criteria for the specific mood disorders in the DSM, which has gone through four editions since its inception, are not easily applied in epidemiological studies. Some diagnostic categories, such as adjustment disorder with depressed mood, cannot be operationalized in standardized interviews because the criteria require a subjective clinical judgment (e.g., the mood disturbance must be related to a specific stressor). Other diagnoses are too inclusive when applied to community samples (e.g., major depressive disorder). Patients identified by current diagnostic criteria are therefore a heterogeneous mix with little clinical relevance beyond symptom severity. In other words, the borderline between clinical depression and normal fluctuation in mood is fuzzy. Even the presence or absence of a symptom may be disputed.

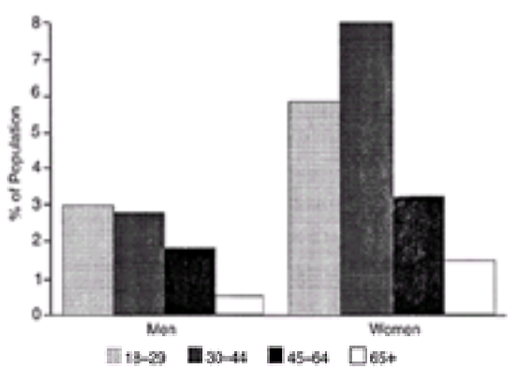

FIGURE 14.2-1 Current prevalence of major depressive disorders by age and gender. (Data derived from the Epidemiologic Catchment Area Study.)

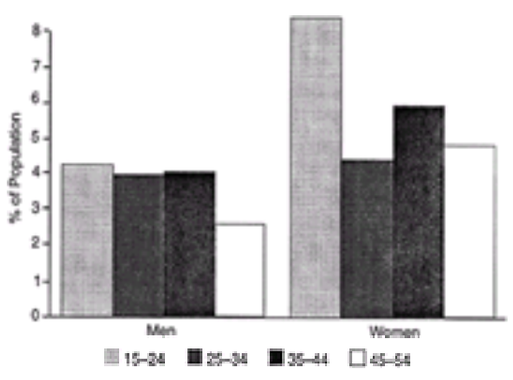

FIGURE 14.2-2 Current prevalence of major depressive disorders by age and gender. (Data derived from the National Comorbidity Survey.)

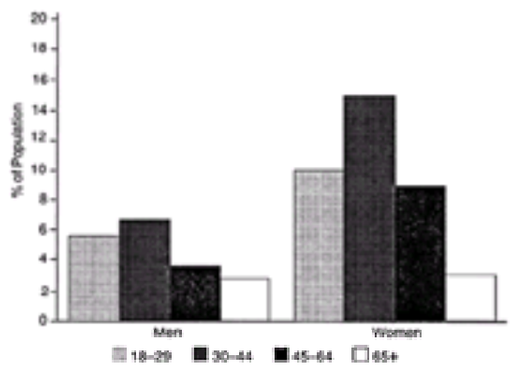

FIGURE 14.2-3 Lifetime prevalence of major depressive disorders by age and gender. (Data derived from the Epidemiologic Catchment Area Study.)

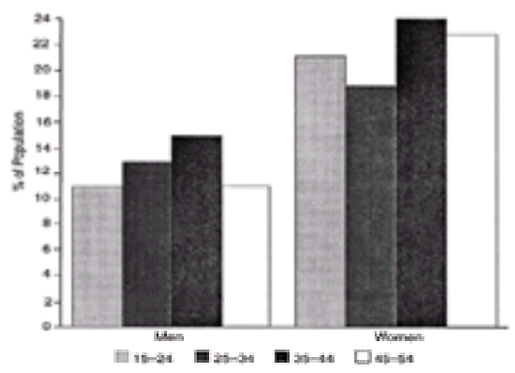

FIGURE 14.2-4 Lifetime prevalence of major depressive disorders by age and gender. (Data derived from the National Comorbidity Survey.)

Some persons in community samples may exhibit depressive syndromes that do not fit the DSM diagnostic system (e.g., for major depressive disorder or dysthymic disorder), but they nevertheless suffer disabling depressive symptoms. Much attention has been focused in recent years on so-called minor or subsyndromal depressive disorder, a syndrome defined by symptoms that are less severe than those of major depressive disorder and of shorter duration than those of dysthymic disorder. The frequency of minor depression was estimated to be nearly 4 percent in the ECA study. Minor depression has generally been divided into two categories: less severe episodes that occur in conjunction with major depressive disorder and symptoms that occur spontaneously. Persons identified in community surveys as experiencing minor depressive disorder have been shown in prospective studies to be at greater risk for time lost at work and increased use of general health services than persons without depressive symptoms. In addition, their risk factor profile is similar to that for major depressive disorder.

Another problem that complicates estimation of the frequency of mood disorders in the community is deterioration of memory over time. Recall of past symptoms is only modestly accurate compared with clinical records of previous depressive episodes. The threshold for reporting a symptom of depression may be higher in a community than in a clinical setting because clinicians often probe for evidence of a symptom that the patient initially denies. Most interview instruments used in epidemiological surveys to identify DSM diagnoses, such as the DIS used in the ECA, were developed in clinical settings and were standardized with classic patients who present to psychiatric treatment settings. In contrast, the modified version of the CIDI used in the NCS (see below) was designed to probe for past history of symptoms more thoroughly than he DIS, because the emphasis of the CIDI was on identifying affected persons in the community who were not experiencing clinically significant depressive symptoms and who had never sought psychiatric treatment. As the threshold for case identification is lower with the CIDI, the estimate of both current and lifetime prevalence of mood disorders is higher. Most persons identified as having mood disorder in the community with the CIDI experience little disruption in function and do not seek psychiatric treatment. Should persons who meet strict fourth edition of DSM (DSM-IV) criteria when interviewed by nonclinicians and asked in detail about past and present symptoms be considered to have true cases of mood disorder, especially major depressive disorder? Therein lies much of the debate.

Comorbidity presents another problem to psychiatric epidemiologists who study mood disorders in community settings. More often than not, symptoms of anxiety and depression overlap. Many subjects receive concurrent diagnoses of major depressive disorder, dysthymic disorder, and generalized anxiety disorder. Most community survey subjects cannot accurately remember whether depression or anxiety was the first syndrome experienced. Do major depressive disorder and generalized anxiety disorder coexist, or is anxiety an epiphenomenon of major depressive disorder? That question remains unanswered.

Nevertheless, development of standardized instruments for case identification in the community, such as the DIS and CIDI, has made it possible to investigate empirically the utility of psychiatric diagnosis. For example, when investigators do not agree upon the frequency of mood disorder with seasonal pattern (i.e., seasonal affective disorder) they can explore their disagreements by using the same criteria and diagnostic instruments for case identification. A recent study demonstrated that though a seasonal pattern of mood swings is common among persons in the community (approximately 8 percent), the frequency of mood disorder with a seasonal pattern, as defined by DSM-IV, is less than 1 percent, even when the most liberal interpretation of the criteria is applied.

DISTRIBUTION OF CASES

Estimates of the prevalence of major depressive disorder by age and gender from the ECA and NCS studies are presented in Figure 14.2-1 and Figure 14.2-2. The DIS was administered for case finding in the ECA study to over 18,000 community and institutionalized subjects (18 + years of age) at five sites throughout the United States—New Haven, East Baltimore, St. Louis, the Piedmont of North Carolina, and Los Angeles. The large numbers of subjects and oversampling of subjects not accurately represented in previous studies, such as African-Americans, Hispanics, and elderly persons, enabled much better estimates of the actual distribution of cases. Over 8000 subjects (15 to 54 years of age) from a nationwide sample were administered a modified version of the CIDI in the NCS study.

Age, Sex, and Residence The most striking finding from the ECA study was a much higher prevalence of all the mood disorders among persons under the age of 45 than in persons 45 years of age and older. Rates were comparable across ECA sites, except for a lower prevalence in North Carolina. The North Carolina sample was composed of both urban and rural residents. Persons in urban areas were as likely to be diagnosed with a mood disorder in North Carolina as in urban areas at other ECA sites. In contrast, rural subjects in North Carolina had much lower rates of major depressive disorder than rural subjects at other ECA sites. In the NCS study, variance by age and geographic location was much less prominent.

The most consistent finding across epidemiological studies of the mood disorders, confirmed by the ECA and NCS studies, is the relatively higher prevalence of major depressive disorder in women than in men. The sex differences are consistent across the life cycle but are much more prominent in young adult and middle-aged persons than in elderly persons and childhood depressive disorders. Because alcohol abuse and mood disorders are often inherited in the same family and alcohol abuse and dependence is more prevalent in men than in women, some have theorized that depressive disorders and alcohol abuse/dependence are phenotypic variants of the same genotype. Little empirical evidence supports this theory.

Sex differences begin in early adolescence and persist at least until midlife. However, women with a previous history of a depressive episode are no more likely to experience a new episode than men with a previous history of a depressive episode. This suggests that the higher risk in women results from women having a higher

risk of experiencing major depressive disorder for the first time. Psychosocial explanations for the higher prevalence of major depressive disorder among women are thus considered the most likely explanation of the sex differences. Epidemiologists have identified stressors that may contribute to increased stress experienced by women, such as maintaining multiple roles as homemaker, professional, wife, and mother.

Race and Ethnicity As illustrated in Figure 14.2-5, the prevalence of the mood disorders does not vary significantly by race or ethnicity. In most epidemiological studies of psychiatric disorders, racial differences in the rates can be explained by socioeconomic and educational differences. The ECA was the first study in Western society that permitted direct comparison of whites, African-Americans, and Hispanics. Previous comparisons, which could not control for geographical differences, were subject to significant bias because prevalence estimates clearly vary by place of residence. The NCS results are similar to the ECA results, though as noted, the overall rates are higher.

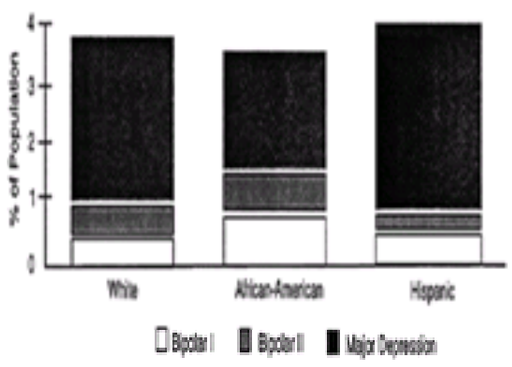

FIGURE 14.2-5 Current (1-year) prevalence of major depressive disorder by race or ethnicity. (Data derived from the Epidemiologic Catchment Area Study.)

International Studies Perhaps the most frequently cited epidemiologic study prior to the ECA and NCS studies was the Stirling County Study from Nova Scotia, Canada. The current prevalence of major depression was estimated to be 4.7 percent in men and 6.0 percent in women. Most studies in developed countries estimate the distribution of major depressive disorder to be greater in women than in men, in young adulthood than in midlife and old age, in urban residents than in rural residents, and among single or divorced persons than among married persons. Few studies document a racial difference when social class and education are controlled. In a recent comparison of population-based epidemiologic studies in 10 countries—the United States, Canada, Puerto Rico, France, West Germany, Italy, Lebanon, Taiwan, Korea, and New Zealand—the lifetime prevalence for major depressive disorder ranged from 1.5 percent in Taiwan to 19 percent in Beirut. Current prevalence ranged from 0.8 percent in Taiwan to 5.8 percent in New Zealand. The difference in prevalence estimates across countries suggests that cultural differences or differences in risk factors may influence the expression of major depression.

Depressive Symptoms The differential frequency of clinically significant depressive symptoms parallels that of major depressive disorder, although the age differences are not nearly so great. In most studies, 8 to 20 percent of community-study participants report depressive symptoms at a level above the cutoff used to screen for major depressive disorder, such as 16 + on the Center for Epidemiologic Studies Depression Scale (CES-D). As with all good screening tests, many potential patients are screened out with more-specific case-identification techniques, such as a standardized diagnostic instrument. Therefore, many persons who appear positive on screening do not meet criteria for specific DSM-IV mood disorders. Those who do appear positive on screening for depressive symptoms, however, experience higher mortality rates, higher disability rates, and poor social functioning.

Incidence Because major depressive disorder is common and tends to remit and recur, the incidence is relatively high. The annual incidence of major depressive disorder in the ECA study was 1.59 percent overall. The distribution by age and gender is presented in Figure 14.2-6. A survey in Lundby, Sweden, revealed an annual first incidence of depression (depression in persons who never experienced depression before) of 0.43 percent in men and 0.76 percent in women. Up to the age of 70, the cumulative probability of a first episode of depression was 27 percent in men and 45 percent in women, making depression one of the most important public health problems (Fig. 16.2-7).

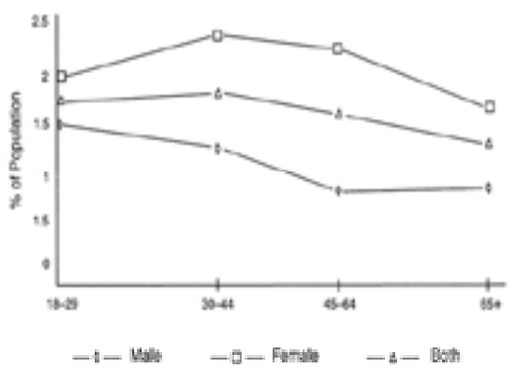

FIGURE 14.2-6 Annual incidence of major depressive disorders by age and gender. (Data derived from the Epidemiologic Catchment Area Study.)

Setting The prevalence of major depressive disorder is much higher in treatment settings than in the community at large. Most investigators find that 10 to 15 percent of persons in acute hospital settings and long-term care facilities meet the criteria for the diagnosis of major depressive disorder. An additional 20 to 30 percent of persons in treatment settings report clinically significant subsyndromal depression (minor depression). The similarities between those patients with major depressive disorder to patients found in psychiatric treatment settings has yet to be documented. Although some depressed medically ill patients respond to antidepressant therapy and brief psychotherapy, many have comorbid conditions that render traditional therapies ineffective.

Depression is also more prevalent in primary care settings than in the general population. Using case-identification methods similar to those in the ECA study, the current prevalence of depression is about twice that found in the general population. In most surveys of primary care clinics, over 20 percent of patients report clinically significant depressive symptoms. Major depressive disorder is diagnosed in one-third to one-half of these outpatients. Young women are at greatest risk for depression in primary care, and most persons who report depressive symptoms to a health care professional report them to a primary care physician.

HISTORICAL TRENDS

The higher prevalence of depression in younger age groups than in older ones has led to the hypothesis that birth cohorts born after World War II are at appreciably greater risk for major depressive disorder than older birth cohorts in advanced Western society. The trend has been observed not only in the United States but also in Sweden, Germany, Canada, and New Zealand.

A number of observations made prior to the ECA study suggest that prevalence rates of depressive disorders are changing. Relevant factors include a progressively lower age of onset of depressive disorders reported in community studies, an increase in childhood mood disorders seen by pediatricians and mental health workers, a decrease in deaths from suicide among the elderly (until about 1980), and a lower average age of onset for depressive disorders in clinical samples since World War II. For example, the risk of first-onset depression was higher for younger birth cohorts than for older birth cohorts in Sweden (Fig. 14.2-7). The trends in suicide data parallel the trends in mood disorders (i.e., suicide rates are much higher in younger persons today than they were in younger persons 30 years ago). Suicide rates in older adults have increased by 25 percent since 1980.

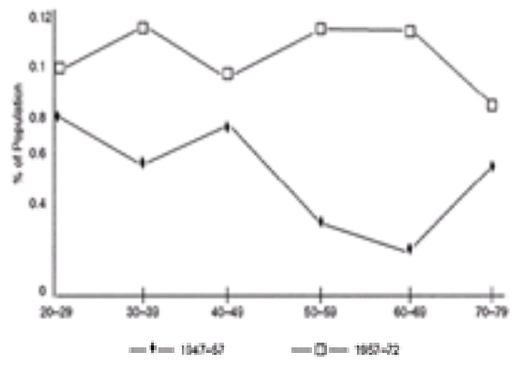

FIGURE 14.2-7 Risk of contracting a first-onset depressive disorder for younger birth cohorts and older birth cohorts in Sweden. (Data derived from Hagnell O, Lanke J, Rorsman B, Ojesjo L: Are we entering an age of melancholy? Depressive illnesses in a prospective study epidemiological study over 25 years: The Lundby Study, Sweden. Psychol Med *12:*279, 1982.)

Factors That Influence Historical Trends Three factors influenced historical trends in the relative prevalence of mood disorders by age: period effect, age effects, and cohort effects.

Period effects are changes in the prevalence of an illness secondary to environmental stressors on the population or particular age groups within the population at a specific period in history. (For example, the uncertainty of employment among college graduates and the trend among younger persons to delay marriage during the 1990s may place young adults at greater risk for depression and suicide because of economic impairment and lack of affiliative relations.)

Age effects are the biological and psychosocial factors that predispose an individual to develop a particular disorder during a specific part of the life cycle. (For example, the genetic predisposition to develop major depressive disorder is probably greatest during the 30s, whereas the predisposition to develop a bipolar disorder is greatest during the 20s.) Age-related changes in the brain, such as increased subcortical hyperintensities on brain magnetic resonance imaging, may also be associated with mood disorders. Perhaps the most consistently observed age effect relevant to mood disorders that has been observed during the 20th century is the positive association between age and suicide among white males in the United States.

Cohort effects are the relative differences in rates of illness across different generations. A cohort is usually defined by the year or decade of birth. Persons born in a given year may be at greater risk for an illness, such as major depressive disorder, throughout their lives. Suicide data reveal marked cohort trends throughout the 20th century. (For example, persons currently 75 to 85 years of age [approximately the birth cohorts of 1915 to 1925] have exhibited lower suicide rates at all ages than either the 1900 or the 1940 birth cohorts.)

Considerable statistical and methodological problems confound sorting out the relative contribution of period, age, and cohort effects upon the prevalence and incidence of mood disorders by age. First, these effects undoubtedly interact. Stressors during a particular period interact with age-related vulnerability. (For example, the current high rate of substance abuse among adolescents may reflect both the vulnerability of adolescents to substance abuse, an age effect, and the greater availability of drugs to adolescents, a period effect.) Second, older persons may not recognize major depressive episodes as such and so do not report them, thus setting the threshold for identifying depression among community-dwelling elders higher. Yet age does not appear to affect the rate of hospitalization for mood disorders. The more severe cases of major depressive disorder are hospitalized, regardless of age, and the relative cohort differences persist in hospitalization rates.

Most investigators have explained the current data as reflecting a period effect. They argue that the risk for depressive disorders increased dramatically for all ages from about 1965 to 1975 but has since stabilized at a higher incidence. Young persons are more vulnerable to that period effect, however, and therefore carry the greatest burden of depressive disorders. A young person who experiences a major depressive episode is likely to exhibit ongoing and severe depressive episodes for many years. Therefore, clinicians can expect to see the current cohort of younger persons bear the burden of major depressive disorder for a long time. Despite being the healthiest and most affluent generation of the 20th century, younger persons may be placed at greater risk for major depressive disorder by a number of environmental risk factors, including increased urbanization, increasing social isolation and anomie, changes in occupational roles and career trajectories for both men and women, increased secularization, and increasing geographic mobility.

IDENTIFICATION OF CAUSES

The risk factors for bipolar I disorder and major depressive disorder identified from community- and clinically based epidemiological studies are summarized in Table 14.2-2. Most have been replicated, yet some most interesting hypotheses about risk for mood disorders have failed to hold up with repeated study. For example, an increased risk for major depressive disorder was discovered in an isolated community of Hutterites who live near the border between the United States and Canada, suggesting that the rigid moral control they exert predisposes community members to depression. Most community studies, however, fail to find that identification with or participation in particular religious groups is associated with an increased risk for major depressive disorder. In contrast, virtually every community survey has demonstrated increased risk for major depressive disorder and depressive symptoms in persons who report negative life events.

Risk Factor	Bipolar I Disorder	Major Depressive Disorder
Gender	No effect	Women at greater risk than men
Race/ethnicity	No effect	No effect
Age	Young at greater risk	Young at greater risk
Socioeconomic status (SES)	Higher SES at somewhat greater risk	Lower SES at greater risk
Marital status	Separated and divorced at greater risk	Separated and divorced at greater risk
Family history	Persons with family history at greater risk	Persons with family history at greater risk
Childhood experiences	Bipolar patients may come from families with low perceived prestige in their community	Evidence that early parental death and disruptive childhood environment leads to major depression
Stressful life events	Negative stressful events associated with increased risk	Negative stressful events associated with increased risk
Chronic stress	No known effect	Chronic stress associated with increased risk
Absence of a confidant	No known effect	Absence of confidant leads to increased risk, especially in women
Residence	Greater risk in suburbs than in inner city	Greater risk in urban areas than in rural areas

Table 14.2-2 Risk Factors for Bipolar I Disorder and Major Depressive Disorder

DEMOGRAPHIC FACTORS

Sex Almost all community-based epidemiological surveys of mood disorders that compare prevalence by sex find that women are twice as likely as men to experience an episode of major depressive disorder. Few investigators discount this finding as an artifact of prejudice in the diagnostic criteria for major depressive disorder or of increased help-seeking behavior among women. Yet female sex has not been demonstrated to be a risk factor per se. The social environment of women and a higher threshold for reporting depressive symptoms in men may account for the increased association. Yet, as discussed above, no mechanism for this apparent increased risk has been established.

Age The average age of onset for both major depressive disorder and bipolar disorders falls between 20 and 40 years. Recent studies also confirm that major depressive disorder can occur in childhood. Bipolar I disorder typically has an earlier age of onset than major depressive disorder, with an average of 30 years. Yet both major depressive disorder and bipolar disorder can first occur at any time during adulthood. Nothing suggests that young age per se places a person at greater risk for the mood disorders (though genetic factors may have their greatest influence at a younger age). Social factors appear to place younger persons at greater risk than elderly persons. Biological predisposition to major depressive disorder may actually increase with age.

Race and Ethnicity Race or ethnicity has not proved to be a significant risk factor for either bipolar I disorder or major depressive disorder. In some community surveys, African-Americans experience a higher prevalence of depressive symptoms. The racial difference usually disappears, however, when other factors, such as socioeconomic status, age, and residence, are controlled. Because treatment for mood disorders is less common for African-Americans than for whites, prevalence studies based on treatment samples usually contain proportionately more whites. Rates for major depression were estimated as higher among Hispanics than among whites and African-Americans in the NCS study in controlled analyses. To date, too few Asians have been included in community-based studies in North America and Europe to permit comparative estimates of risk for Asians. The similarity in overall rates across racial or ethnic groups does not necessarily mean a similarity across all symptoms. For example, older African-Americans have been found more likely to complain of interpersonal problems when depressed and less likely to complain of depressed mood than older whites.

Socioeconomic Status The findings from community-based studies relating to socioeconomic status as a risk factor for-depression are mixed. The overall ECA studies found only a weak correlation between major depressive disorder or bipolar disorders and lower socioeconomic status. The North Carolina ECA study, however, found a consistent relation between socioeconomic status and major depressive disorder, even when multiple potential confounders, such as race and residence, were controlled. In the NCS study, both lower income and education were associated with higher prevalence of major depression. Studies prior to the ECA and NCS found a consistent positive relation between lower socioeconomic status and depression. In one classic study reported by A.B. Hollingshead and Frederick C. Redlich, depressive symptoms were strongly associated with the lower social classes. In a more recent study, working-class women from an eastern suburb of London were much more likely to suffer depressive symptoms than women from higher social classes.

Marital Status Marital status appears to be one of the most consistent risk factors for both depressive symptoms and major depressive disorder. Rates for major depressive disorder are highest among separated and divorced persons and lowest among single and married persons. Recent widowhood is associated with higher rates of major depressive disorder across the life cycle. The risk appears to vary with sex. Single women have lower rates of depression than married women, whereas married men have lower rates than single men. However, the investigator must not confound marital status with the loss of a spouse through death or divorce (a stressful life event). If a woman was widowed during the 6 months prior to the study, then the event not the status is perhaps the causative factor. In addition, cause and effect may be reversed (e.g., depressive illness may place a person at greater risk for divorce). In most studies, however, separated or divorced status places a person at greater risk for depression, even if the marital breakup occurred long before the assessment. The ECA studies, unlike previous studies, also documented a much higher prevalence of bipolar disorders among separated and divorced persons than among those who were single. The highest rates, however, were found among those who were cohabiting, even when adjusting for age, sex, and race or ethnicity. The association of the mood disorders and marital status is also reflected in the association of mood disorders with household size. Major depressive disorder is twice as common among persons living alone than among those who live with others. In persons not living alone household size is not associated with depression.

Marital status may not be the proximal causative factor. The perception of social support and lack of conflict within the social network are critical factors in protecting against mood disorders. Longitudinal studies of the social network and neuroses have shown that the most important predictors of depression are not the objective characteristics of the network but rather the perception of how adequately the network assisted the person. Large-scale community-based investigations of the risk factors for major depressive disorder cannot disentangle the subtleties of the complex interactions between persons and their social network. (For example, dissolution of a difficult marriage may relieve long-standing depressive symptoms.)

Family History Most epidemiological studies of treatment samples have shown a consistent increase in family history of mood disorders among subjects, especially in first-degree relatives. A family history of suicide and alcoholism has also been repeatedly found to be more common among depressed subjects than among controls. Most experts attribute the increased risk for depression when family history is positive to a genetic predisposition; yet shared family environment may also contribute to the increased risk. Genetic transmission is much more firmly established for bipolar I disorder than for major depressive disorder. In family members of bipolar subjects, both bipolar disorder and major depressive disorder are more prevalent.

Early Childhood Experience Much attention has been directed to the association of early childhood experience with onset of mood disorders later in life. Although the complexities of a psychodynamic investigation of childhood traumas cannot be applied in community-based epidemiological studies, even cursory investigation of childhood experiences has revealed correlates. Parental loss before adolescence is a well-documented risk factor for adult-onset depression. A deprived and disrupted home environment also constitutes a risk. Methodological problems make objective study of childhood trauma and deprivation difficult. Some events (e.g., divorce or separation of parents) can be documented reliably, but others (e.g., parental neglect) are quite subjective. The report of parental neglect by a depressed adult may vary depending on the respondent's emotional state at the time of the interview.

Personality Attributes Personality attributes are closely associated with early childhood experience as a risk for mood disorders in later life. Persons predisposed to develop a depressive disorder lack energy, are more introverted, worry, are more dependent, and are hypersensitive. Major depressive disorder is also frequently comorbid with the Axis II disorders; yet the study of the relation of depression and personality is confounded by the time at which personality is studied. Epidemiologists rarely have the opportunity to assess personality before the onset of the first episode of depression. If personality is assessed during an episode of depression, then the depressive symptoms mask certain personality traits and exaggerate others. When a person has experienced and recovered from a depressive episode, its impact on personality makes accurate assessment of premorbid personality difficult. (For example, the personality characteristics associated with depression are exactly those that might emerge in response to the experience of a serious psychiatric disorder.)

Social Stressors Social stressors have received more attention than other risk factors for major depressive disorder across the life cycle, except sex. Three kinds of social stressors can be distinguished: life events, chronic stress, and daily hassles. Life events are the kind most often explored in epidemiological studies. They are identifiable, discrete changes in life patterns that disrupt the usual behavior and threaten the person's well-being. Bereavement, the reaction to loss of a loved one, is the prototypic stressful life event. Chronic stress includes long-term conditions that challenge the person, including financial deprivation, ongoing interpersonal difficulties (e.g., conflict in the workplace), and persistent threat to security (e.g., living in a dangerous neighborhood). Daily hassles are ordinary but stressful occurrences that are ubiquitous in modern life (e.g., managing household finances and unpleasant interactions with neighbors).

Most epidemiological studies reveal a relation between stressful life events, especially negative events, and the onset and outcome of major depressive disorder. Nevertheless, the use of stressful life event scales such as the Schedule of Recent Events introduces many potential biases into the study of stressors and depression. Such scales usually tally the number of events and weight them according to a predetermined algorithm. Most schedules weight events on the basis of normative data from the population. Because the data usually derive from weightings provided by young adults, they do not necessarily apply across the life cycle. (For example, retirement in late life may be a positive event, whereas premature retirement in midlife may present problems that can precipitate a depressive disorder.)

The perception of the event is probably more important than the event itself. More-sensitive measures of stressful events document not only the event itself but the subject's response to it. Was the event perceived to be positive or negative? Even the death of a spouse may be viewed as a positive event if it occurred after a protracted and disabling illness during which the subject was the caretaker. Was the event perceived to be important or unimportant? For some older persons a move may be extremely traumatic, especially if it is the first move in half a century. For others, a move may be a usual and relatively unimportant event, especially in a society in which mobility is becoming more the norm. Was the event expected or unexpected? If income decreases at retirement at a rate expected by the retiree, then the loss of income is much less stressful than if a person is forced to take an unexpected cut in salary while still in the workforce.

The accumulation of stressful negative life events does appear to predispose a person to episodes of major depressive disorder. In a study from New Haven, depressed patients had an average increased frequency of eight life events during the 6 months preceding the onset of depressive symptoms. Those events included marital arguments, marital separation, starting a new type of work, change in work conditions, serious personal illness, death of an immediate family member, serious illness of family members, and a family member leaving home. Stressful events are also associated with the persistence of depressive disorders. In a study from England, adverse events during the year following the initial episode of depression were associated with a poorer outcome of the episode. The adverse effects of life events may be offset by neutralizing events. (For example, if a woman loses her job but soon after finds another job with equal pay and benefits, then the adverse event is neutralized.)

In addition, the impact of a stressful life event may vary depending upon when that event occurs in the natural history of a mood disorder. Persons experiencing a recurrent major depressive disorder are less likely to report a stressful event associated with the onset of episodes after the first two episodes of depression. The phenomena of kindling may become the predominant impetus for recurrent depressive episode.

Chronic stress can place a subject at greater risk for major depressive disorder than specific stressful life events. The stress of service in the Persian Gulf War led to increased frequency of major depressive disorder among military personnel. As long as soldiers are deployed, they have difficulty recovering from the major

depressive episode. In addition, the long-term effects of being deployed also may increase the risk of depression following military service. Persons usually have more difficulty coping with a chronic stressful situation than with specific events.

Few studies document the association of daily hassles with the onset of major depressive disorder. Yet, impulsive acts such as a suicide attempt may be closely associated with daily hassles to which the subject cannot adapt within the context of a stressful life event or chronic stress. That is, daily hassles may be the straw that breaks the camel's back.

Social Support Factors in the social environment that may modify the effects of social stressors have received increased attention in the epidemiological investigation of both physical and psychiatric disorders. One factor is social support, the provision of meaningful, appropriate, and protective feedback from the social environment that enables a person to negotiate environmental stressors. In theory, social support is an attractive concept, for it is potentially more amenable to intervention than environmental stressors. The roots of the construct social support go back at least to the early twentieth century, when Emile Durkheim proposed that persons who are not integrated into society (the condition called “anomie”) are at greater risk for suicide. Social support has four components: the social network, social interaction, perceived social support, and instrumental support. The social network is assessed by identifying those individuals or groups of individuals (e.g., a spouse and children) who are available to the subject. The absence of a spouse is a risk factor for major depressive disorder.

Social interaction is assessed by determining the frequency of interactions between the subject and other network members. A number of studies confirm that social isolation (i.e., a deficit in social interaction) places a subject at greater risk for depression. Yet the quality of the interaction appears to be more important than the frequency of interaction. Perceived social support is assessed by determining the subjective evaluation by the individual of the dependability of the social network, the ease of interaction with the network, the sense of belonging to the network, and the sense of intimacy with network members. The association of major depressive disorder and expressed absence of a confidant exemplifies the relation between perceived inadequate support and depression.

Instrumental support is assessed by determining the concrete and observable services that are provided to the subject by the social network (for example, cooking meals, financial assistance, and nursing services for the physically ill). Although such support is essential to the well-being of the young and the elderly in society, few studies document the association of depression with a deprivation of instrumental support. The physical health of the person is a confounding factor. The need for instrumental support is usually not recognized until the person exhibits an actual need for such services. In addition, the perception of the availability of those services in a time of crisis may not reflect the actual availability.

The construct of social support is strongly influenced by the construct of social integration. An integrated society is a social system that ensures the patterns of interpersonal behavior that are essential to the survival and welfare of its members. Those patterns enable persons to obtain what is needed for subsistence and protection against weather and disease, control hostility and other forms of social disruption, create new members and their education and communication, store information, and permit decision making and united action. Alexander Leighton and his colleagues undertook the most ambitious epidemiological studies of social integration and mental health in a survey of communities in Nova Scotia. Social scientists and anthropologists studied each community to determine its relative integration versus disintegration. At all ages, the rates of depressive disorders (and other psychiatric disorders) were higher in disintegrated communities. Studies of social integration are not as proximal to the individual as studies of social stressors and social support because measures of social integration are not specific to the individual. Studies of integration are ecological, for they document that the overall functioning of the community is associated with the overall level of psychopathology.

Most ecological studies in psychiatric epidemiology have been limited to comparisons of communities by traditional parameters, most commonly, urban versus rural residence. The hypothesis is that rural communities are less stressful than urban communities. In the ECA study of North Carolina, major depressive disorder was twice as common in the urban community, with the largest differences among the young (under 45 years of age) and among women. Those urban-rural differences in prevalence persisted even when the comparison was controlled for race, socioeconomic status, marital status, and age. Nevertheless, most ecological studies fail to demonstrate that after adjusting for other factors, communities determined to be socially integrated protect against mood disorders. The reason, probably, is that investigators are usually studying diverse communities and thus the concept of social integration is probably not relevant in complex societies in which communities are layered upon one another or are patterned together in a mosaic that cannot be easily disaggregated.

Another risk factor for depression is unemployment. At present, most men and women under the age of 65 are in the labor force. Men and women who were unemployed for at least 6 months during the 5 years prior to the ECA survey were more than three times as likely as others to report symptoms of an episode of major depressive disorder during the year prior to the survey. In the NCS study, the rate of major depression was three times higher in persons not working than in those working. Homemakers were nearly three times more likely to experience major depression than the employed as well.

The multiple risk factors for mood disorders form a web of causation. Each factor cannot only affect the subject directly but can interact with other factors. Mathematical models of causative factors are therefore useful for determining the relative importance and complex interaction of those factors. Models include linear and logistic regression analyses.

An example is presented in Table 14.2-3. In this logistic regression analysis, sex, education, income, marital status, employment, and household composition are the control (or adjusted) variables. The net effect of risk factors in the NCS is presented as an odds ratio, an approximation of the relative risk of persons with the characteristic developing a major depressive episode compared with that of persons without the characteristic. For example, the risk for females is 1.36 times greater than that for males.

Characteristic	Odds Ratio	95% Confidence Level
Sex	1.00	–
Male		
Female	1.36*	1.01–1.84
Education (years)		
0–11	1.93*	1.14–3.25
12	1.93*	1.14–3.25
13–15	1.93*	1.14–3.25
16 or more	1.00	–
Marital status		
Married	0.42*	0.28–0.61
Separated/widowed/divorced	1.00	–
Never married	0.58*	0.38–0.88
Employment		
Working	1.00	–
Student	1.00	–
Homemaker	2.40*	1.54–3.75
Other	2.54*	1.51–4.28

Adapted from Blazer DG, Kessler RC, McGonagle KA, Swartz MS. The prevalence of major depression in a national community sample: The National Comorbidity Survey. Am J Psychiatry 151:979, 1994.
* $p < .05$

Table 14.2-3 Adjusted Odds Ratios for Prevalence of Current (30-day) Major Depressive Episode, by Selected Demographic Characteristics

PROGNOSIS

Two recent studies have concentrated on the public health impact of depressive disorders, because of their chronic and disabling nature. In the first, over 1000 patients in a variety of primary care settings were screened for depression. Patients with either depressive disorder or depressive symptoms (without a diagnosis of a specific mood disorder) tended to have worse physical, social, and role functioning. When their objective health status was controlled, they perceived their current health to be worse than did patients who were not depressed, and they reported more physical pain. The poor functioning associated with depressive symptoms, with or without a diagnosis of a mood disorder, was comparable to or worse than that in eight major chronic medical conditions. The number of days in bed with depressive symptoms was significantly greater than with hypertension, diabetes, or arthritis.

In the second study, the ECA sample in North Carolina, persons with a diagnosis of major depressive disorder or dysthymic disorder and with symptoms of minor depressive disorder were followed for 1 year. Compared with asymptomatic individuals, persons with major depressive disorder had a fivefold greater risk of disability, and persons with minor depressive disorder had one-and-one-half times the risk. Persons with minor depressive disorder were at greater risk of developing major depressive disorder at 1-year follow-up.

A number of *natural history* studies of mood disorders have been performed on clinical samples. The most extensively studied cohort derives from the Psychobiology

of Depression Study and consists of over 500 young adult and middle-aged subjects diagnosed with either bipolar I disorder or major depressive disorder. Following diagnosis about 50 percent of subjects recovered during the first year, but fewer than 30 percent of the others recovered during subsequent years. Comorbid dysthymic disorder with a slow onset accompanying psychotic symptoms was associated with less likelihood for recovery. Relapse rates are high for major depressive disorder immediately following recovery. Superimposed dysthymic disorder and a history of three or more major depressive episodes were associated with relapse. Bipolar I disorder patients with only manic episodes had better outcomes than those with major depressive disorder. However, bipolar I patients with a mixed episode (depression and mania) or with rapid cycling had worse outcomes than those with major depressive disorder.

USE OF HEALTH SERVICES

Health services for mood disorders are provided in general health care settings and in specialty settings. Most mental health visits reported by subjects in the ECA study, regardless of disorder, occurred in primary care settings for older persons and in specialty settings for younger persons. Women use mental health services in both settings about twice as often as men. Visits are about equally distributed between general medical providers and mental health specialists. Persons who were depressed used health services more frequently than those with no psychiatric disorder.

SUGGESTED CROSS-REFERENCES

An overview of epidemiology is given in Section 5.1. Social origins of mood disorders are discussed in Section 4.2. Classification of mental disorders is presented in Chapter 9. Specific review of the genetics of mood disorders can be found in Section 14.3. The role of stress in the etiology of psychiatric disorders is discussed in Section 25.9. Suicide is discussed in detail in Section 29.1. The epidemiology of psychiatric disorders in late life is reviewed in Section 51.1b.

SECTION REFERENCES

Blacker CVR, Clare AW: Depressive disorder in primary care. Br J Psychiatry *150:*737, 1987.

Blazer DG, Bachar IR, Manton KG: Suicide in late life: Review and commentary. J Am Geriatr Soc *34:*519, 1986.

Blazer DG, George LK, Landerman R, Pennybacker M, Melville ML, Woodbury M, Manton KG, Jordan K, Locke B: Psychiatric disorders: A rural/urban comparison. Arch Gen Psychiatry *42:*651, 1985.

*Blazer DG, Kessler RC, McGonagle KA, Swartz MS: The prevalence and distribution of major depression in a national community sample: The National Comorbidity Survey. Am J Psychiatry *151:*979, 1994.

*Blazer DG, Kessler RC, Schwartz M: Epidemiology of recurrent major and minor depression with a seasonal pattern. Br J Psychiatry *172:*164, 1998.

*Broadhead WE, Blazer DG, George LK, Tse CK: Depression, disability days, and days lost from work in a prospective epidemiologic survey. JAMA *264:*2524, 1992.

Coryell W, Aksikal HS, Lyon AC, Winokur G, Maqer JD, Mueller TL, Keller MB: The time course of nonchronic major depressive disorder. Arch Gen Psychiatry *51:*405, 1994.

Durkheim E: *Suicide: A Study in Sociology.* Free Press, New York, 1951.

Eaton JW, Weil RI: *Culture and Mental Disorders.* Free Press, New York, 1955.

Eaton WW, Kramer M, Anthony JC, Dryman A, Shapiro S, Locke BZ: The incidence of specific DIS/DSM-III mental disorders: Data from the NIMH Epidemiologic Catchment Area Program. Acta Psychiatr Scand *79:*163, 1989.

Endicott J, Spitzer RL: A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. Arch Gen Psychiatry *35:*837, 1978.

*Hagnell O, Lanke J, Rorsman B, Ojesjo L: Are we entering an age of melancholy? Depressive illnesses in a prospective epidemiological study over 25 years: The Lundby Study, Sweden. Psychol Med *12:*279, 1982.

Hirschfeld RMA, Klerman GL: Personality attributes and affective disorders. Am J Psychiatry *136:*67, 1979.

Hollingshead AB, Redlich FC: *Social Class and Mental Illness.* Wiley, New York, 1958.

Keller MB, Shapiro RW, Lavori PW, Wolfe H: Recovery in major depressive disorder. Arch Gen Psychiatry *39:*905, 1982.

*Kendler KS, Gardner CO, Prescott CA: Clinical characteristics of major depression that predict risk of depression in relatives. Arch Gen Psychiatry *56:*322, 1999.

*Kendler KS, Prescott CA: A population-based twin study of lifetime major depression in men and women. Arch Gen Psychiatry *56:*39, 1999.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Esllerman S, Wittchen H, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. Arch Gen Psychiatry *51:*8, 1994.

Kessler RC, McGonagle KA, Swartz M, Blazer DG, Nelson CB: Sex and depression in the National Comorbidity Survey 1: Lifetime prevalence, chronicity and recurrence. J Affect Dis *29:*85, 1993.

Klewna GL, Weissman MM: Increasing rates of depression. JAMA *261:*2229, 1989.

Koeniq HG, Meador K, Cotlen HJ, Blazer D: Depression in elderly hospitalized patients with medical illness. Arch Intern Med *148:*1929, 1988.

*Leighton AH: *My Name Is Legion.* Basic Books, New York, 1959.

Leighton DC, Harding JS, Macklin DB, Macmillan AM, Leighton AH: *The Character of Danger: Psychiatric Symptoms in Selected Communities.* Basic Books, New York, 1963.

Morris JN: *Uses of Epidemiology*, ed 3. Churchill Livingstone, London, 1975.

Paykel ES: Life stress and psychiatric disorder. In *Stressful Life Events: Their Nature and Effects*, BS Dohrenwend, BP Dohrenwend, editors. Wiley, New York, 1974.

Robins LN, Helzer JE, Croughan J, Ratcliff K: National Institute of Mental Health Diagnostic Interview Schedule: Its history, characteristics, and validity. Arch Gen Psychiatry *38:*381, 1981.

Rorsman B, Grasbeck A, Hagnell O, Lanke J, Ohman R, Ojesjo L, Otterbeck L: A prospective study of first-incidence depression: The Lundby Study, 1951–1972. Br J Psychiatry *156:*336, 1990.

*Rosenthal NE, Mazzanta CM, Barnett RL, Hardin TA, Turner EH, Lam GK, Ozaki N, Goldman D: Role of serotonin transporter promoter repeat length of polymorphism (5-HTTLPR) in seasonality and seasonal affective disorder. Molecular Psychiatry *3:*175, 1999.

Shapiro A, Skinner EA, Kessler LO, Von Korff M, German PS, Tischler GL, Leaf PJ, Benham L, Cottler L, Regier DA: Utilization of health and mental health services: Three Epidemiologic Catchment Area sites. Arch Gen Psychiatry *41:*971, 1984.

Somervell PD, Leaf PJ, Weissman MM, Blazer DG, Bruce ML: The prevalence of major depression in black and white adults in five United States communities. Am J Epidemiol *130:*725, 1989.

Tennant C: Female vulnerability to depression. Psychol Med *15:*733, 1985.

*Weissman M, Merikanges KR, Boyd JH: Epidemiology of affective disorders. In *Psychiatry*, vol 1, section 60, R Michels, AM Cooper, S Guze, LL Judd, GL Klerman AJ Solni, AJ Stunkard, PJ Wilner, editors. Lippincott, New York, 1991.

Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H, Joyce PR, Karam EG, Lee C, Lellouch J, Lepine J, Newman SC, Rubio-Stipec M, Wells JE, Wickrafnaratne PJ, Winchen H, Yeh E: Cross-national epidemiology of major depression and bipolar disorder. JAMA *276:*293, 1996.

Weissman MM, Livingston B, Leaf PJ, Florio L-P, Ilolzer C: Affective disorders. In *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*, LN Robins, DA Regier, editors. Free Press, New York, 1991.

Wells KB, Siewart K, Hays RD, Burnam A, Rogers W, Daniels M, Berry S, Greenfield S, Ware J: The functioning and well-being of depressed patients: Results from the Medical Outcomes Study. JAMA *262:*954, 1989.

Textbook of Psychiatry

14.3 MOOD DISORDERS: GENETICS

JOHN R. KELSOE, M.D.

The familial nature of mood disorders has been a widely observed phenomenon since antiquity. However, only in the last century have systematic studies been conducted that document the degree and nature of this familiality as well as its genetic determinants. In the past two decades molecular genetic technology has brought a new era in the understanding of a wide range of genetic traits and disorders. The application of this technology promises a new, more fundamental understanding of biological etiology of mood disorders and may also produce revolutions in both diagnosis and treatment. Yet mood disorders and other psychiatric disorders are difficult problems for genetic analysis, and early attempts at such analysis using both epidemiological and molecular tools have yielded sometimes contradictory and confusing results. However, as the volume of data increases and more sophisticated analytical methods are used these large problems should become more tractable so the benefits of this approach can be realized.

GENETIC EPIDEMIOLOGY

By determining the rates of illness in different types of relatives, genetic epidemiological studies can provide much information about the familial and genetic nature of a disorder. The questions that can be addressed include: Are mood disorders familial? Are they genetic? What portion of the etiology is genetic? How are the genes for mood disorder transmitted? How do different forms of mood disorder differ in their genetic transmission? How are different forms of mood disorder related to each other?

Numerous such studies conducted over the last century provide much information about the genetic transmission of mood disorders. However, these studies have various methodological limitations. Foremost among these is the range of diagnostic methods used. Many of these studies were conducted before the distinction between depressive (unipolar) and bipolar disorders, and hence results of these studies pool both illnesses. Similarly, many of these studies preceded the introduction of operationalized diagnostic criteria. Therefore, it may not be clear exactly how the diagnoses were made, making it difficult to compare or pool results across studies. Of the studies that distinguish unipolar and bipolar disorder, bipolar disorder has received more attention because of its greater degree of familiality.

Another methodological issue important to such studies is ascertainment bias. If the results of a study are to be meaningfully generalized to the population, the subjects must be selected in a systematic and nonbiased fashion. For example, if *probands* (the first ill subject identified in a family) are selected on the basis of their strong family history, then the results of a family study may inaccurately indicate a strong familial rate of illness in the population. A similar error will be made if ill family members are preferentially selected for study. Systematic ascertainment methods attempt to avoid such bias by studying all patients who present within a certain environment, such as a mood disorders clinic.

Many of these studies are also limited by use of the family history method. In this approach, the rates of illness in family members of probands are determined by systematically questioning probands about their families. Though several excellent standardized instruments for the family history method exist, this method is inherently less accurate than using direct interviews of each family member to make a diagnosis.

Family Studies Family studies address the question of whether a disorder is familial. More specifically, is the rate of illness in the family members of someone with the disorder greater than that of the general population? Typically, all subjects with the disorder in a given environment or population are identified and questioned about illness in their first-degree relatives. The rates of illness are then compared with either the rates in the general population or the rates in first-degree relatives of control subjects. Rates of illness are typically adjusted for age to indicate the morbid risk (i.e., the risk that an individual will develop an illness at some point of his or her life).

Table 14.3-1 illustrates several such studies of bipolar disorders. They indicate a morbid risk of bipolar disorder in first-degree relatives of bipolar disorder probands that ranges between 3 and 8 percent. Compared with a 1 percent rate in the general population, this reflects a substantial familial increase. Similarly, studies of families of probands with depressive disorder (unipolar) reveal morbid risks for depressive disorders among first-degree relatives which are two to three times those of the general population. These data argue strongly for the familial nature of mood disorders. Furthermore, depressive disorders generally occur at a higher rate in the families of probands with bipolar disorders, and the rate of bipolar disorder is elevated in the families of probands with depressive disorders. In fact, depressive disorders are typically the most common mood disorder in families of probands with bipolar disorders. This familial overlap suggests some common genetic underpinnings between these two forms of mood disorder.

Study	Relatives at Risk (N)	Morbid Risk (%)	
		Bipolar	Depressive (Unipolar)
Dunner et al, 1980	1199	4.2	8.2
Gershon et al, 1982	598	8.0	14.9
Rice et al, 1987	557	5.7	23.0*
Sadovnick et al, 1994	1102	3.5	5.7

* Observed rates rather than morbid risk.

Table 14.3-1 Selected Family Studies of Bipolar Disorders

Twin Studies The family study data clearly indicate that mood disorders are familial. However, such studies cannot distinguish whether genetic or environmental factors mediate the familial transmission. Families might share a variety of different environmental factors that could transmit the illness. Such factors might be behavioral but could also be shared exposure to infectious agents, toxins, or other brain insults. Twin studies provide the most powerful approach to separating genetic from environmental factors, or “nature” from “nurture.” Many strategies for twin studies have been used, but most commonly both monozygotic (MZ) and same-sex dizygotic (DZ) twin pairs are identified in which one twin has a mood disorder. The other twins are then examined to determine the proportion of twin pairs in which both twins are affected, termed the *concordance rate.* Typically, twin pairs are selected who have been raised together, so that environmental factors are shared equally. A difference in concordance rate between the MZ and DZ pairs, therefore, reflects the role of heritable genetic factors. An alternative powerful strategy is to study twin pairs raised apart; however, such samples are much more difficult to obtain.

Table 14.3-2 summarizes several twin studies of mood disorders. Considering depressive and bipolar disorders together, these studies find that the concordance rate for mood disorder in the MZ twins is two to four times that in the DZ twins. These are the most compelling data for the role of genetic factors in mood disorders.

Further, the concordance rate for MZ twins is not 100 percent. Thus nonheritable environmental factors also play a significant role in mood disorders. In studies that distinguish bipolar from unipolar disorders, the MZ to DZ concordance ratio for bipolar-bipolar pairs is higher than that for unipolar-unipolar pairs, which indicates greater genetic involvement in bipolar disorders than in unipolar depressive disorders. Furthermore, the rate of depressive disorders is elevated in monozygotic cotwins of probands with bipolar disorders, and to a lesser extent, the rate of bipolar disorders is elevated in the cotwins of probands with depressive disorders. This is consistent with the family data, and it argues for a genetic overlap between bipolar and depressive disorders.

	Monozygotic Twins		Dizygotic Twins	
Study	Twin Pairs (N)	Concordance (%)	Twin Pairs (N)	Concordance (%)
Rosanoff et al, 1935	23	69.6	67	16.4
Kallman, 1954	27	92.6	55	23.6
Bertelsen, 1979	55	58.3	52	17.3
Kendler et al, 1993	154	69.7	326	34.9

Table 14.3-2 Selected Twin Studies of Mood Disorders

Adoption Studies Adoption studies provide an alternative approach to separating genetic and environmental factors in familial transmission. A variety of limitations of twin studies have been raised, including the argument that parents treat monozygotic twins and DZ twins differently, so environment is not equally shared. Adoption studies have been conducted using a variety of experimental designs, but the most common is the adoptee-as-proband strategy. In this approach, probands are identified who have a mood disorder and who were adopted away at birth, thus separating nature and nurture. The rates of psychiatric illness are then determined in both the biological and adoptive parents.

Only a limited number of such studies have been reported, and their results have been mixed. Julien Mendlewicz and John Ranier found a threefold increase in the rate of bipolar disorders in the biological relatives of probands with bipolar disorders. They also observed a twofold increase in the rate of depressive disorders in biological relatives. Similarly, in a Danish sample, Paul Wender and coworkers reported a threefold increase in the rate of unipolar disorder and a sixfold increase in the rate of completed suicide in the biological relatives of mood disorder probands. Other studies, however, have been less convincing. Using affectively ill mothers as probands, Remi Cadoret found a nonsignificant trend toward increased unipolar depressive disorders in the adopted children. In a Swedish sample, Anne-Liis Von Knorring and coworkers found no increase in mood disorders in the biological parents of adoptees with mood disorders. Overall, these results support the role of genetics and are consistent with the twin data. The difficulty of obtaining subjects and the resulting small sample sizes may help explain why these data are weaker than the twin data.

MODE OF TRANSMISSION

If mood disorders are in large part caused by genetic factors, then what is the nature of its genetic transmission? Segregation analysis of the family study data has been used to attempt to answer this question. Are mood disorders the result of one or a few genes transmitted in a mendelian fashion? Or do many genes interact within each individual to predispose to illness? Different modes of transmission result in different patterns of inheritance of illness. By examining these patterns in families, one may attempt to distinguish the different possible modes of transmission. For example, in a simple dominant genetic disorder, one expects to observe that half of the children of an affected parent are also affected. In a recessive disorder, only one-quarter of the children of two nonaffected carriers should be affected. More-complex modes of inheritance involving multiple genes result in other patterns of illness that can be sought in the family data. Typically, in segregation analysis, the predicted patterns of several different models of transmission are tested to see which best fits the observed family data.

The results of such analyses have been mixed. Several such analyses have been inconclusive and excluded all tested models of transmission. Other, more recent analyses using large samples and more-sophisticated genetic models have supported the presence of an autosomal dominant major locus. Most such studies have focused on bipolar disorders because of their greater heritability. However, a recent segregation analysis of early-onset, recurrent depressive disorder has supported the presence of major gene effects with autosomal recessive or codominant modes of transmission. Other studies have supported a multifactorial-threshold model for mood disorders. In these models, the additive effects of multiple genes produce a unitary predisposition common to all mood disorders. Different mood disorders result at different thresholds in this single underlying genetic liability. X-linkage has also been argued based on the observation that female relatives of probands with bipolar disorders have a twofold higher risk for disorders than males. This is also supported by evidence for decreased male to male transmission of bipolar disorders. Drawing any definitive conclusions from such complex and inconsistent data is difficult. The complexity likely results from the presence of multiple genes with multiple modes of transmission. In the face of such heterogeneity, the sample sizes and statistical methods used may have limited power to demonstrate a given mode of transmission consistently. The data do suggest that of the many genes probably operating in mood disorders, some have a major effect on predisposition, and in bipolar disorders these major genes are likely to have autosomal dominant or perhaps X-linked inheritance. Consistent with such heterogeneity, recent analyses have argued for more-complex modes of transmission in which multiple genes interact to predispose to illness.

Several other intriguing results emerge from the family study data. Subjects with mood disorders are more likely to marry spouses who also have mood disorders than is expected by chance. This is termed *assortative mating*, and it leads to a higher rate of families in which the illness can be traced to both the mother's and the father's side of the family than would be expected by chance. Such bilineal families may play an important role in the interaction of multiple genes in the population. Family studies have also indicated that the rate of mood disorders is increasing over time in the population, termed the *cohort effect* (Fig. 14.3-1). Among family members of probands with bipolar disorders, those born more recently have a higher risk for bipolar disorders and an earlier age of onset. The cause of the cohort effect is unknown. It has been speculated to be a result of changing environmental stresses in our society, an artifact of recollection, or possibly an indication of a genetic effect termed *anticipation.*

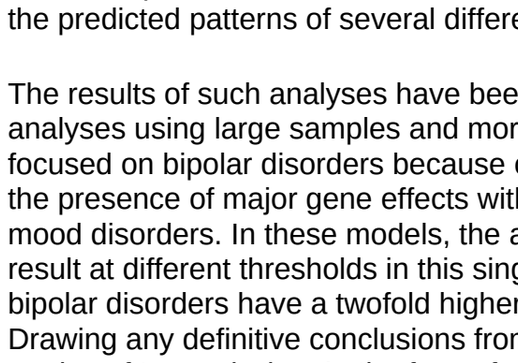

FIGURE 14.3-1 Age-related penetrance curves for bipolar disorder. The risk for bipolar disorder among relatives of bipolar disorder probands is depicted as a function of age. The probability of having bipolar disorder, or penetrance, increases with age. The cohort effect is illustrated by the different age-dependent risk curves for relatives within three different age groups. Relatives born more recently have a higher rate of bipolar disorder and an earlier age of onset. (Reprinted with permission from Rice J, Reich T, Andreasen NC, Endicott J, van Eerdewegh M, Fishman R, Hirshfeld RM, Klerman GL: The familial transmission of bipolar illness. Arch Gen Psychiatry *44:*441, 1987.)

GENETIC RELATIONSHIPS WITHIN THE SPECTRUM OF MOOD DISORDERS

The genetic relation between the various forms of mood disorder has received much study and debate. Bipolar and depressive disorders are widely considered to have some sort of common genetic underpinning, though its exact nature is unclear. The twin and family data reviewed above argue for more depressive disorders occurring in the twins or other relatives of probands with bipolar disorders than is expected by chance. However, it is less clear that bipolar disorders occur in the relatives of probands with depressive disorders at an elevated rate. Twin studies indicate that polarity is usually consistent in MZ twins (i.e., bipolar-bipolar or unipolar-unipolar pairs are much more common than bipolar-unipolar pairs. Yet bipolar-unipolar pairs do occur. These data suggest that bipolar and unipolar disorders are genetically neither completely identical nor completely distinct. Rather, there is a partial genetic similarity. A possible model for the relationship between these genes and disorders is illustrated in Figure 14.3-2. In this model, some or all of the genes for bipolar disorders may also result in depressive disorders. In addition, a larger pool of genes predisposes only to depressive disorders. Therefore, the rate of unipolar disorder would be clearly elevated in families with bipolar disorder. However, because only a minority of cases of unipolar disorder result from bipolar genes, only a small increase in the rate of bipolar disorders would be seen in the relatives of probands with depressive disorders. Such overlapping relationships between genes and disorders may also occur for other forms of mood disorders. A more definitive understanding of these relationships will likely require identification of the specific genes involved.

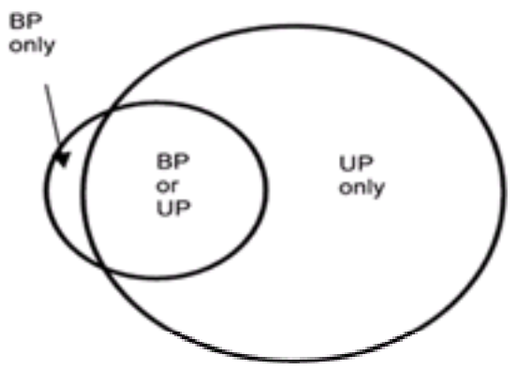

FIGURE 14.3-2 This model of the relation between genes for bipolar and depressive disorders posits that most genes for bipolar disorders can predispose to either bipolar or depressive disorders. A larger set of genes predisposes only to unipolar disorder. Hence, a subset of those with depressive disorders carries genes that may also predispose to bipolar disorders.

This model predicts that a portion of those with unipolar disorder carry genes that may also predispose to bipolar disorder. Such patients have been said to have “bipolar III” disorder by some writers and have been the subject of much discussion and investigation. They are presumably identified by a family history of bipolar disorder or a history of developing hypomania or mania only in response to antidepressant treatment. Similarly, Hagop Akiskal and coworkers have described a hypomanic-like personality style termed *hyperthymic temperament.* Depressive disorder patients with hyperthymic temperament likely carry some bipolar disorder genes. Such hyperthymic-depressive patients are more likely to have a family history of bipolar disorder and to develop mania spontaneously. These depressive disorder patients with a bipolar disorder genetic diathesis may also be more likely to respond to lithium augmentation of antidepressant treatment.

Other forms of mood disorders have also been postulated to be at least somewhat genetically distinct. Several studies have reported that though the risk for bipolar I disorder is similar in the relatives of probands with bipolar I or bipolar II disorder, the risk for bipolar II disorder is greater than that for bipolar I disorder in the relatives of probands with bipolar II disorder. This suggests that bipolar II disorder to some extent breeds true and that a subset of the genes for bipolar disorders predispose preferentially to bipolar II disorder. However, studies of patients with rapid-cycling bipolar disorder find that rapid cycling in the proband does not affect the risk for mood disorders or rapid cycling in relatives. This argues that rapid cycling is not a distinct genetic subform of bipolar disorder.

The genetic relation of schizoaffective disorder to mood disorders is a complex question that involves the role of psychosis in the genetics of mood disorders. The genetic nature of schizoaffective disorder also bears on the genetic relation between schizophrenia and mood disorders. Studies examining familial risks in relatives of schizoaffective disorder patients have led to inconsistent results, some finding an increased risk of schizophrenia and some an increased risk of bipolar disorder. A possible explanation is that patients with schizoaffective disorder represent a mixture, some with bipolar and some with schizophrenia diatheses. This notion is supported by data that show an increased rate of bipolar disorders among relatives of probands with the bipolar type of schizoaffective disorder. The rate of schizophrenia has also been reported to be increased among the relatives of probands with the depressive type of schizoaffective disorder. Alternatively, it has been proposed that the Kraepelinian distinction between the disorders is not valid and that schizophrenia and mood disorders lie at the extremes of a spectrum of a common genetic liability. Recent linkage studies (reviewed below) may be beginning to support the existence of some genetic loci common to both schizophrenia and bipolar disorders.

INTERPRETATIONS OF THE GENETIC FEATURES OF MOOD DISORDERS

The above data argue that mood disorders are not simple genetic traits. No one gene consistently causes illness in all cases in a simple and predictable fashion. In genetic terms, there is not a 1 to 1 relation between the expressed trait (phenotype) and genes (genotype) transmitted in a simple mendelian fashion. Therefore, mood disorders are said to be complex genetic disorders rather than simple mendelian traits. What factors contribute to this complexity?

The twin data argue compellingly that genes account for only 50 to 70 percent of the etiology of mood disorders. Environment or other nonheritable factors must explain the remainder. Thus, a predisposition or susceptibility to disease is inherited. The probability that someone will manifest a trait, given that they have a certain genotype, is termed the *penetrance* of the gene. Mood disorders are said to have “reduced penetrance” (less than 100 percent). Furthermore, the penetrance of the mood disorder genes increases from a low risk for illness in childhood to a maximum in adulthood. The cohort effect (Fig. 14.3-1) further complicates the relation of penetrance to age, causing penetrance to vary with the date of birth. Therefore, families of people with mood disorders likely include individuals who have the genes for mood disorder but do not develop the disease. These are termed *nonpenetrant carriers.* The converse of this situation is individuals who have mood disorder but do not have the genes. Individuals with purely environmentally caused disease are termed *phenocopies.* These factors conspire to produce an indirect relationship between genes and disease.

Variable expressivity refers to the phenomenon of the same gene or group of genes resulting in a variety of different forms of illness. The twin data clearly demonstrate this for mood disorders. Monozygotic twins with identical genomes are observed with one twin exhibiting bipolar and the other unipolar disorder. Nonheritable factors must play a role in the specific manifestation of the predisposition to mood disorder. Such variability in expression is not unique to psychiatric disorders. In neurofibromatosis, for example, ill individuals range in manifestations and severity from those with only pigmented retinal lesions to others with multiple large tumors—a range in expression that results from the same disease gene.

Of the various factors complicating the genetic transmission of mood disorders, the most significant and the most challenging for gene mapping efforts is genetic heterogeneity. *Heterogeneity* refers to the likely occurrence of multiple genes in the etiology of illness. Only the identification of multiple disease genes can convincingly demonstrate genetic heterogeneity. However, the segregation analyses described above strongly suggest its presence. There are several critical questions about the nature of heterogeneity in mood disorders. How many genes are involved? How large an effect does each gene have? How do the genes interact to produce illness? These questions suggest a variety of models of heterogeneity, which can be broadly grouped into those in which disease results from a few genes with major effects (major loci) and those in which disease derives from the combined action of many genes with small effects (polygenic or oligogenic). The answers to these questions are not currently known for mood disorders. However, segregation analyses suggest a mixture of genes of both large and small effect, which are transmitted in a variety of ways.

Evidence for several other complex forms of genetic transmission has also been reported for mood disorders. Some studies of bipolar disorder have indicated that the illness is more likely to be transmitted through mothers than through fathers. Such *parent-of-origin effects* imply a genetic phenomenon called *imprinting.* In imprinting, a genetic locus is processed differently in male and female meioses so that different traits result from maternal and paternal transmission. For example Angelman and Prader-Willi syndromes are two different mental retardation syndromes that result from different maternal or paternal imprinting of the same locus on chromosome 15.

Rather than yielding different phenotypes, the bipolar data suggest that the penetrance of bipolar genes may be affected by imprinting.

Another nonmendelian genetic phenomenon reported in mood disorders is anticipation. In disorders displaying *anticipation*, the severity of the illness increases and the age of onset decreases with successive generations. Anticipation is generally associated with genetic mutations involving trinucleotide repeat expansions. In such disorders (e.g., Huntington's disease or fragile X mental retardation), the defective gene contains a region of deoxyribonucleic acid (DNA) in which a three-nucleotide sequence is repeated a variable number of times. For reasons currently not well understood, the number of repeats increases in successive generations until the gene's function is disrupted and illness results. Anticipation involving both increasing severity of illness and decreasing age of onset has been reported for bipolar disorder. Indirect evidence for the presence of trinucleotide repeat expansions has also been reported; however, no specific gene manifesting such a mutation has been described.

COMPLEX GENETIC DISORDERS

Simple Mendelian Traits Complex genetic disorders are simply those that are not transmitted in classical mendelian patterns. A review of mendelian transmission thus provides a useful background for understanding complex genetic traits. Simple mendelian traits display genetic homogeneity; one gene transmits the trait with complete penetrance. Figure 14.3-3 illustrates the three primary modes of mendelian transmission. In genetic terminology, different forms of a given gene are termed *alleles.* A disease gene may have either a normal, nonmutated allele or a mutated, disease-causing allele. In dominant genetic disorders, only one copy of the disease allele is necessary to cause illness. Dominant illnesses are typically transmitted vertically from grandparent to parent to child. In the simplest case, half of the children of an affected parent will have the disease. In recessive disorders, both copies of the disease gene must be defective for disease to result. Heterozygotes with only one disease allele are nonaffected carriers. Disease typically results from the mating of two nonaffected carriers. One-quarter of the resulting children will be homozygous for the disease allele and hence ill. Recessive illnesses typically appear in a horizontal pattern in families (i.e., in cousins). The nature of the mutation in the gene determines whether it is transmitted in a dominant or recessive fashion. Recessive mutations typically deactivate genes that are expressed in excess. Therefore, an adequate amount of gene product can be produced by only one functioning copy. In dominant disorders, the amount of gene product expressed may be critical. A reduction in gene dosage, resulting from only one functioning copy, leads to illness. Alternatively, a dominant mutation may result in overfunctioning of the gene, which results in illness. In X-linked disorders the disease gene is located on the X chromosome. These may be either dominant or recessive. X-linked dominant disorders are more common in women, while X-linked recessive disorders are more common in men. Father to son transmission is impossible in X-linked disorders, since the father transmits the Y sex chromosome to sons.

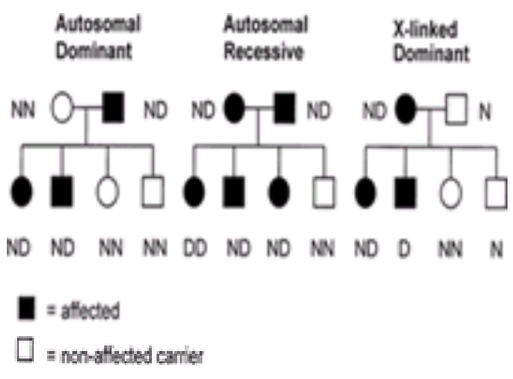

FIGURE 14.3-3 Mendelian transmission. Simple mendelian disorders are transmitted by three different modes. Dominant disorders require only one copy of the disease allele for a family member to be affected. In recessive disorders, both copies of the disease gene must be defective. In X-linked disorders, the disease gene is carried on the X chromosome.

Heterogeneity Models Of the various factors that distinguish complex disorders from mendelian ones, the most important is the presence of multiple genes, or genetic heterogeneity. Multiple genes may combine to produce illness in a variety of different ways, as illustrated in Figure 14.3-4. These heterogeneity models fall into two categories based on genetic effect size. In single major locus models, only one disease gene is necessary to produce illness in a given individual. In an interfamilial heterogeneity model, one gene transmits the predisposition to illness in each family. However, there are different predisposing genes in different families. In an intrafamilial heterogeneity model, any one of multiple genes within the same family can transmit the illness.

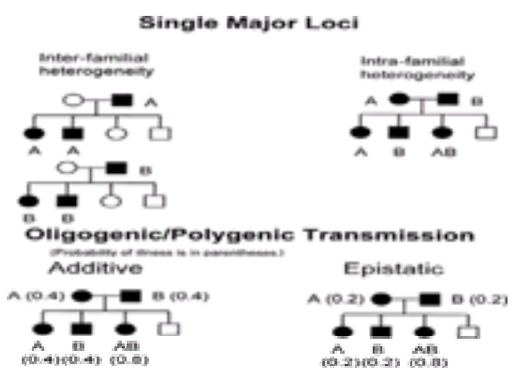

FIGURE 14.3-4 Several different models for the role of multiple genes in genetic disorders. In single major locus heterogeneity models, a single gene is primarily responsible for the predisposition to illness in an individual. However, different single major loci may act in different families or within the same family. In oligogenic models, multiple genes each of smaller effect typically interact to produce the susceptibility to illness. In an additive model, the effects of these genes simply add together. In epistatic models, the overall effect exceeds the sum of each gene acting separately.

In oligogenic or polygenic transmission models, multiple genes of smaller effect interact to predispose to illness. In these models, one gene by itself is unlikely to cause illness. Rather, the probability of illness increases with the number of different genes involved. The terms “oligogenic” and “polygenic” are distinguished simply by whether a smaller or larger number of genes, respectively, are involved. In an additive polygenic model, each gene contributes a certain probability (penetrance) of manifesting the disorder. The total genetic liability to illness is then the sum of the probabilities contributed by all the polygenes. In Figure 14.3-4, under the additive model, the disease alleles for genes A and B alone each convey a 40 percent probability of illness. Individuals who carry the disease alleles for both A and B have an 80 percent risk for illness. The total risk is simply the sum of the risks for each individual polygene. In the epistatic model, the disease alleles at genes A and B each convey a relatively small effect alone. However, individuals with the disease alleles for both A and B have a risk for illness that exceeds the sum of risks for A and B. In the example in Figure 14.3-4, A and B each convey a 20 percent risk alone. The affected daughter with the AB genotype, however, has an 80 percent risk for illness. Epistatic interactions have been observed in many organisms and frequently reflect a “two-hit” effect on a biological system. For example, a neurotransmitter system may be able to tolerate a defect in gene A by increasing the activity of gene B. Similarly gene B may be able to compensate for a defect in A, so that individuals with defects in A or B alone may have only a limited risk for illness. However, if both A and B are defective, the system cannot compensate, and the risk for illness escalates.

Quantitative Traits One of the many difficulties in psychiatric genetics studies is the definition of “affected.” Variable expressivity results in a variety of disorders from similar genotypes, and is frequently not clear which of these disorders should be considered “affected” for the purposes of genetic analyses. An alternative to such dichotomous definitions of phenotype is the use of quantitative phenotypes. Many biological variables are obvious quantitative phenotypes (e.g., blood pressure or height). However, it is not immediately obvious which mood disorder is “worse” or “more” than another. Nevertheless, quantitative phenotypes have been applied with some success to mood disorders. An example is the multiple-threshold model. Though quantitative phenotypes can result from single major loci, the concept has evolved historically in connection with polygenic traits. In such models, each polygene contributes in either an additive or epistatic fashion to the value of the

quantitative phenotype. Quantitative models offer a useful alternative to the dichotomous approach to phenotype. However, the basic problem is that mood disorder phenotypes are more complex than either the dichotomous or quantitative genetic models.

POSITIONAL CLONING OF COMPLEX DISORDERS

Positional cloning refers to the use of molecular genetic methods to identify disease genes based on their chromosomal location. Such studies, directed at the identification of specific disease genes, are the focus of most recent research on the genetics of mood disorders. The methods and strategies of positional cloning are reviewed in more detail elsewhere in this text, but a brief description here prefaces a review of the problems faced in such studies of mood disorders and some of their potential solutions.

DNA markers are segments of DNA of known chromosomal location that are highly variable among individuals. They are used to track the segregation of specific chromosomal regions within families affected with a disorder. When a marker is identified whose alleles consistently cosegregate with disease in families, it is said to be *genetically linked*, which implies that a gene for the disorder is physically near the marker on a chromosome. The Human Genome Project has provided thousands of such markers and precisely mapped them to chromosomal locations. By using several hundred markers, one can systematically survey the genome in search of markers linked to a disease. In this fashion, novel disease genes can be identified on the basis of their chromosomal location rather than their physiological function. This ability to identify novel genes without relying on knowledge of their function or the pathophysiology of the disorder makes positional cloning a powerful approach. This approach has led to the successful identification of genes for numerous diseases such as Huntington's disease and cystic fibrosis.

The statistics of linkage analysis are either parametric or nonparametric. *Parametric analyses* assume a certain model of inheritance (e.g., dominant or recessive) and then test the marker data for the probability of fitting that assumed model. The statistic typically generated is the LOD score. The *LOD score* is the logarithm of the odds for linkage divided by the odds against linkage. A LOD score of 3 represents odds of 10^3 or 1000 to 1 in favor of linkage. The appropriate threshold for statistical significance of LOD scores varies with the nature of the analysis and is currently under some debate, but is in the range of 3.0 to 3.6. Nonparametric analyses do not require an assumption of model of inheritance. Instead, affected family members are tested for significantly increased sharing of marker alleles. The affected sib pair method and the affected pedigree member method exemplify this approach. Parametric methods generally have greater power to detect linkage if the correct model is used. Nonparametric methods have less power but are not vulnerable to error resulting from the use of an incorrect model of inheritance.

When a chromosomal region is identified by genetic linkage studies in families, the disease gene has typically been localized to a region of between 5 and 30 million base pairs of DNA, which might contain several hundred genes. Genetic association studies are then sometimes used to locate the disease gene more precisely within this region. In these studies, markers are tested for significant differences in the frequencies of their alleles between unrelated groups of affected subjects and control subjects. Such differences may be observed if the DNA-sequence variation at the marker is the disease causing mutation itself. Alternatively, the marker allele may have no functional impact, but be so physically close to the disease causing mutation that they have been propagated together through the population. The marker allele and disease mutation are then said to be in *linkage disequilibrium.* In addition to fine mapping, association is also commonly used to test the role of a candidate gene whose known function suggests its possible involvement in the disorder. Physical mapping methods are also used to identify the exact disease gene. These methods involve cloning large regions of DNA and screening them for the presence of genes and possible disease-causing mutations in those genes.

Difficulties of Mapping Complex Disorders Simple mendelian traits show a 1 to 1 correspondence between genotype and phenotype. Everyone in a family who has the disorder has the gene, and everyone who has the gene has the disorder. This allows extraction of maximal information regarding the possible cosegregation of marker and disorder in linkage studies. However, both reduced penetrance and phenocopies loosen the connection between genotype and phenotype and thus reduce the statistical power of linkage methods to identify genes. In complex disorders, it is neither clear that nonaffected individuals did not receive the disease allele nor that affected ones did. Thus one needs to study larger samples of families.

Variable expressivity and difficulty in diagnosis introduce further uncertainty in modeling genetic transmission. It is not clear which of the spectrum of mood disorders should be considered “affected.” If the definition of affected is too narrow, then valuable information from family members with more subtle forms of illness may be lost. However, if the definition of affected is too broad, then error may be introduced by inclusion of too many phenocopies. Most linkage studies currently use a hierarchy of different definitions of illness, ranging from narrow to broad. For example, a narrow model might include as affected only those with bipolar disorder. A broad model would include bipolar disorder and recurrent major depression. The difficulty in making behavioral diagnoses adds further complexities. Psychiatric diagnoses rely on the accuracy of the subject's memories and judgments about behavior that depend on the individual's life and environment.

Other difficulties arise out of the limitations of statistical methods. The most powerful statistical methods currently available are based on mendelian models of transmission. However, it seems clear that mood disorders are not transmitted according to these simple models. Currently most linkage studies are conducted using a variety of genetic models (e.g., dominant and recessive). However, only limited information is available about the robustness of applying these methods to nonmendelian and heterogeneous forms of transmission. Furthermore, recent simulation studies have indicated that even when linkage is established, a much larger chromosomal region is implicated in complex disorders than in mendelian ones. This makes the job of fine mapping and gene identification substantially larger.

The greatest problem facing positional cloning studies is clearly heterogeneity. If mood disorders result from a relatively small number of genes with large effect, then they will be identified soon. However, if a large number of genes with small effect are involved, identifying them will require large sample sizes or a different approach altogether. Current data suggest that the number of genes probably lies somewhere between these extremes. The cost of heterogeneity in terms of sample size is illustrated in Figure 14.3-5. This simulation study, using affected sib pairs, indicates that if a gene is present in less than 25 percent of families under study, the total number of families required to detect linkage increases significantly.

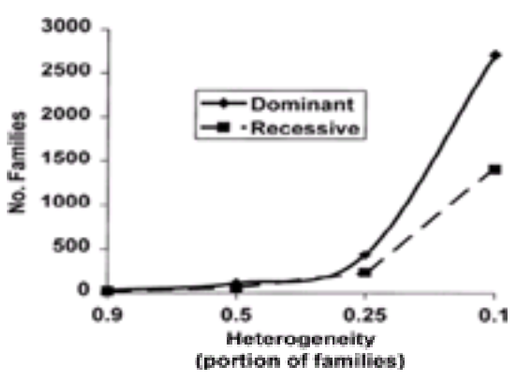

FIGURE 14.3-5 Sample sizes required to detect linkage for different degrees of heterogeneity. The number of families with affected sib pairs required to detect linkage is depicted as a function of the degree of genetic heterogeneity. This power analysis for genetic linkage studies was performed by computer simulation of both an autosomal dominant and a recessive disorder with 50 percent penetrance. If a gene for a disorder occurs in less than 25 percent of families, the number of families required to detect it in a linkage study goes up substantially. (Reprinted with permission from Martinez MM, Goldin LR: The detection of linkage and heterogeneity in nuclear families for complex disorders: One versus two marker loci. Am J Hum Genet *44:*552, 1986.)

Solutions for Mapping Complex Disorders Despite these problems, recently developed tools and strategies for linkage studies of complex traits promise to make the challenges of mood disorders more tractable. Many of these new developments involve using more-sophisticated statistical methods. Parametric methods have recently been developed that can model the interaction of two genes simultaneously. In this way, the power of parametric analytical methods can be applied to more-complex modes of inheritance. However, as more genes are included, the number of possible models for their interaction increases. This may make the choice of the correct model both more difficult and more important.

For these reasons, nonparametric methods have recently received increasing attention. Since they do not depend on assumptions about the mode of transmission, they are less vulnerable to error resulting from use of the wrong model. However, they have lower statistical power to detect genes than do parametric methods using the correct model. Much effort has been focused on developing new nonparametric methods with greater statistical power. There is no current agreement on which

approach is superior. As a result, many investigators use both approaches.

Various approaches to refining the phenotype may improve the odds of success in linkage studies. Studying subforms of illness, whose genetic distinctness is supported by epidemiological data, may allow investigators to reduce the number of genes under study. For example, bipolar II disorder or lithium-responsive illness may represent distinct subsets of bipolar genes. Such subforms of illness may be easier to map. Intermediate phenotypes may also be useful in this regard. *Intermediate phenotypes* are biologic markers associated with illness that segregate in families. Sleep electroencephalographic (EEG) measures and sensitivity to various pharmacological challenges exemplify such possible markers. Intermediate phenotypes can be used to identify psychiatrically well family members who carry susceptibility genes. In this way, these biological markers display higher penetrance than the psychiatric disorder itself and, therefore, retrieve more genetic information from family members. Furthermore, they may also reflect biologically distinct subforms of illness.

The rapid progress of the Human Genome Project will provide new tools for genetic studies. The quality of markers and maps continues to improve. It is likely that within the next several years, all human genes will be identified and mapped, which will be invaluable for the identification of disease genes. The next wave of the Human Genome Project will develop a dense set of markers that will ultimately enable genome-wide association studies. By studying tens of thousands of markers in thousands of ill and control subjects, genes of small effect may be identified. Until recently, such an approach was not considered feasible; however, now both the dense map of markers and the technology for high-throughput screening seem in sight. Methods are now being developed to array genes or DNA markers on microchips allowing the simultaneous detection of thousands of markers. Ultimately, one can expect that most sequence variation in all human genes will be known. This knowledge combined with large samples for association studies should enable detection of virtually all genes involved in mood disorders, even if current linkage strategies are not successful.

RESULTS OF GENETIC LINKAGE AND ASSOCIATION STUDIES

To date the search for genes for mood disorders has focused primarily on bipolar disorder because of the stronger evidence for its genetic basis. In its early years, this search exhibited a series of ups and downs of reported findings and subsequent nonreplications that was likened to the highs and lows of the illness itself. Such high-profile nonreplications led to frustration among investigators and concern that false-positive linkage results might be common. Many investigators argued for highly conservative thresholds for the statistical significance of linkage results to guard adequately against such events. However, subsequent experience has indicated that strongly positive results are relatively uncommon. Furthermore, simulation studies indicate that nonreplications should be expected, even for real loci, until enough data are accumulated to demonstrate linkage convincingly.

Recently a number of positive results have emerged that are quite encouraging in terms of the ability of linkage analysis to identify genes for bipolar disorder. As more of the genome has been surveyed in larger sets of families, several loci have yielded modest evidence for linkage, which is being independently replicated in multiple datasets. It seems likely that at least some of these loci reflect the locations of real susceptibility genes; therefore, it seems appropriate to expect some nonreplications. The road to linkage success may involve neither too-readily dismissing modestly positive results nor too readily accepting strongly positive results but rather seeking independent confirmation. Some of the more prominent of these possible bipolar loci are summarized in Table 14.3-3 and reviewed below. The most notable aspect of the summary in Table 14.3-3 is the number of loci that have support from multiple independent datasets. Only time, further data, and the identification of actual disease genes will determine which of these are real disease loci.

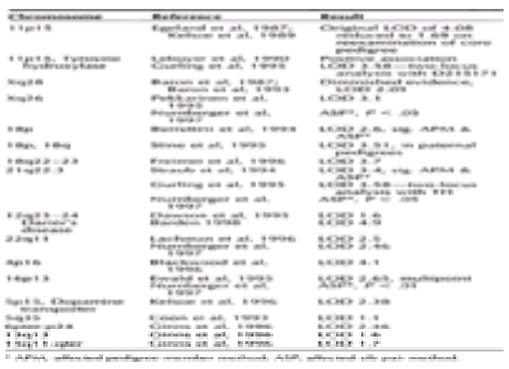

Table 14.3-3 Summary of Loci Potentially Linked to Bipolar Disorders

THE HIGHS AND LOWS OF PAST BIPOLAR LINKAGE STUDIES

Chromosomes 11 and X The first molecular genetic linkage study of mood disorder was conducted in the Old Order Amish population of southeastern Pennsylvania. Bipolar disorders are no more common or different in presentation in the Amish than in the general population. The Amish were chosen for study because they were genetically isolated. The current population of approximately 30,000 Amish are descended from about 50 couples who immigrated nearly 300 years ago from Germany and Switzerland. Since that time, for religious reasons, this group has remained isolated from the surrounding population. The primary advantage of studying such a group is the reduced genetic heterogeneity that results from a founder effect with subsequent genetic isolation. Other advantages of studying this group are its large families and the virtual absence of substance abuse. In 1987, Janice Egeland and coworkers reported evidence for linkage of bipolar disorders to markers at the insulin and Harvey *ras* loci on 11p15. Their data yielded an LOD score of nearly 5, or odds of 100,000 to 1 in favor of linkage. However, numerous other studies in different populations failed to replicate this result. Subsequently, reexamination of an updated and expanded version of the same Amish pedigree substantially reduced the evidence for linkage. Recently, however, association and linkage studies of the nearby tyrosine hydroxylase locus have supported the existence of a bipolar disorder locus in this region.

A similar scenario played out in studies of the X chromosome. Much epidemiological data has suggested a possible X-linked locus for bipolar disorder. In 1987, Meron Baron and coworkers examined two nonmolecular X chromosomes markers, color blindness and glucose-6-phosphate dehydrogenase (G6PD), for linkage to bipolar disorder in a set of Israeli families. They found very strong evidence for linkage to these markers on Xq28. However, as with the Amish study, numerous other studies failed to replicate this result in different samples. Subsequently, the evidence for linkage was substantially reduced when several more-informative molecular genetic markers were examined. More recently, however, several studies have implicated a nearby region, Xq26, in studies of American and Finnish families. Therefore, despite their checkered history, evidence remains for both the 11p15 and Xq26–28 regions, but more data will be required to confirm these possible linkages.

Chromosome 18 One of the most intriguing regions for possible bipolar susceptibility loci is on chromosome 18. Currently this region has the strongest data for linkage, because of the number of independent replications. Evidence for linkage to the pericentrometric region of chromosome 18 was first found by Wade Berrettini and coworkers in a series of 21 North American pedigrees. Subsequently, other studies replicated this finding and identified two more-distal regions on 18q. Some of these reports have found that the evidence for linkage was stronger in families with paternal or mixed transmission than in those with only maternal transmission. These data suggest a parent-of-origin effect at this locus. Other studies have argued that the evidence for linkage to chromosome 18 markers is strongest in families with comorbid panic disorder. Such data begin to suggest that different susceptibility loci may predispose to different forms of bipolar disorder with distinct clinical presentations. It is currently unclear whether these three different reported regions reflect three different susceptibility loci or if one or more reflect false-positive results. However, given the amount of independent data implicating this chromosome, it seems very likely that at least one locus for bipolar disorder resides on chromosome 18.

Other Promising Genetic Loci Several other chromosomal regions have been implicated by multiple studies and may harbor bipolar susceptibility loci. Linkage to the marker PFKL on chromosome 21q was first reported by Richard Straub and coworkers, with most of the evidence coming from one extended pedigree. Subsequently, linkage to this region was replicated in several other studies. A locus for bipolar disorder on chromosome 12 was implicated initially by the serendipitous observation of cosegregation of bipolar disorder and a dermatological disorder, Darier's disease, in a Welsh family. Subsequently, Darier's disease was mapped to chromosome 12. Based on this observation, a set of Welsh families with bipolar disorder and without Darier's disease were studied, and modest evidence for linkage to the same region of chromosome 12 was obtained. Recently, Nicholas Barden and coworkers implicated this same region in a study of a large family from a population isolate in Quebec.

A region on 22q is interesting because it has been implicated in both bipolar disorder and schizophrenia. This region was originally investigated in schizophrenia because of the observation of psychosis in adolescents with the 22q11 deletion syndrome, velocardiofacial syndrome. Several studies have reported evidence of linkage of markers in this region to schizophrenia; other investigators have examined adolescents with velocardiofaciol syndrome and observed rapid-cycling mood disorders. Based on these data, Herbert Lachman and coworkers studied this region in a set of bipolar disorder families and found evidence suggesting linkage. A subsequent study of an expanded version of this family set has further strengthened this result. The National Institute of Mental Health (NIMH) consortium also reported evidence for linkage of bipolar disorders to 22q. Similar to these results on chromosome 22, the same regions on chromosome 13 and 18 have also been implicated in studies of both bipolar disorders and schizophrenia. This raises the intriguing possibility that some genes may play a role in the susceptibility to both disorders. Confirmation of this idea awaits further replication and the definitive identification of disease genes. However, if confirmed, these results may indicate a greater degree of genetic commonality between these two disorders than had been thought.

The dopamine transporter has been implicated in bipolar disorders in a recent linkage study by John Kelsoe and coworkers. Its role as an important regulator of dopamine neurotransmission and the site of action of such stimulants as cocaine and amphetamine, make the dopamine transporter a compelling candidate gene. This gene has also been associated with attention-deficit/hyperactivity disorder in children, which occurs with greater frequency in families with bipolar disorder.

Several complete genome surveys were reported recently that have added substantially to the bipolar linkage data now available. Several additional loci were implicated by these studies including markers on 5q and 4p. A genome survey of the Old Order Amish recently provided data for possible loci on chromosomes 6, 13, and 15. The NIMH Consortium for Linkage Studies of Bipolar Disorder recently reported its initial findings in one of the largest such datasets examined to date, suggesting new loci on chromosomes 1, 6, 7, and 10. Perhaps more importantly, their data support previously reported loci on 16, 21, 22, and X. Such possible replications imply that some genes for mood disorders have relatively large effect and will be detectable by current linkage strategies.

GENETIC COUNSELING

Based on the observation of psychiatric illness in their own families or the increasing public awareness of psychiatric genetics, patients frequently ask clinicians three questions: Are mood disorders genetic? What is the risk to my children or grandchildren? Is there a blood test for the gene? The answer to the last question is the easiest, currently no blood test is available. However, such a test is likely to be available in the future, and it will raise a variety of serious practical and ethical issues. The answer to the first question is also easy: Yes, mood disorders are genetic, a position defended by the large body of epidemiological data summarized in this chapter. However, patients must understand that mood disorders are only partly genetic. The twin studies argue strongly that only 50 to 70 percent of the etiology of mood disorders is genetic. Therefore, a predisposition to illness is inherited that interacts with other, nonheritable factors.

The risk to children and grandchildren is a difficult question that deserves the greatest consideration. The family data indicate that if one parent has a mood disorder, a child's risk for mood disorder is between 10 and 25 percent. If both parents are affected, this risk roughly doubles. A careful family history is needed for more accurate prediction of risk for a specific family. Several factors from the family history should be considered. The more members of the family that are affected, the greater the risk to a child, particularly if the disorder is in both parents' families. The risk is greater if the affected family members are first-degree relatives rather than more distant relatives. A family history of bipolar disorder conveys a greater risk for mood disorders in general and a much greater risk for bipolar disorder specifically. The presence of more-severe illness in the family also probably conveys a greater risk. All these factors should be considered in forming an estimate of risk for the concerned parent.

Equally important to providing the estimate of risk is providing guidance in interpreting and responding to that information. Patients' reactions to risk information vary greatly depending on their own experience with the illness. Some will be relieved to find it so low, others fearful because it is so high. One must emphasize that their child carries a risk or predisposition to illness, not a certainty of illness. One can also describe the range of illness from mild to severe that could result and the availability and efficacy of treatment. Ultimately, the use of such information in family planning is a highly personal decision. Some patients may choose not to have children. For existing children, parents should be told about the typical age of onset, presenting symptoms, and the importance of early recognition and treatment. However, this must be balanced with the goal of not labeling the child or being overly protective.

FUTURE DIRECTIONS

The gradual accumulation of a large body of data led to the successful identification of possible disease loci described above. Therefore, in the near future, gene mapping studies will likely continue to use current strategies. As still more data are accumulated, more reported loci will likely be confirmed, while others will prove false. It currently seems promising that many genes for mood disorders can be identified through such large, though tractable, projects. The development and study of the large samples required will necessitate large collaborative efforts between multiple groups of investigators using common diagnostic and genetic methods. The recent NIMH consortium is a model for such future collaborations.

The variety of methodological advances and alternative approaches detailed above may further advance the project. Larger sets of family and genotypic data will enable application of new, more-sophisticated statistical methods. Larger datasets will also permit the possible subdividing of the disease into subforms based on clinical or biological measures. The rapid advance of the Human Genome Project will greatly facilitate this work by providing a much denser set of markers, a detailed physical map of the genome, and more efficient genotyping and sequencing methods. Ultimately, the identification and mapping of all human genes, the complete sequence of the human genome, and a catalog of human genetic variation will enable development of powerful novel approaches to complex disorders. Such tools and large samples will likely permit identification of most genes for mood disorders in the next several decades.

What will be the implications of this knowledge? The most immediate impact may well be on public opinion. Psychiatrically ill persons have long suffered from the stigma and discrimination that portions of the public impose out of fear and ignorance. Acceptance of major mental illnesses as brain disorders has been slow. The definitive identification of genetic causes may be highly beneficial to public understanding and acceptance.

This knowledge should also have a dramatic impact on our understanding of pathophysiology and approaches to treatment for mood disorders. Most current theories of pathophysiology are based on the mode of action of therapeutic agents, which were for the most part discovered serendipitously. The site of action of therapeutic drugs is not necessarily either the site of the genetic defect or the primary site of the pathophysiology. Identification of disease genes may point to entirely new systems involved in the pathophysiology or components of currently implicated systems that were previously unknown. It is hoped that such an understanding may lead to the rational drug design long sought by patients, clinicians, and pharmaceutical companies. The result may be new drugs that act by completely novel mechanisms, possibly with greater efficacy and specificity. Further down the genetic road lies the prospect of gene therapy for mood disorders. In this approach, DNA is used as a therapeutic agent, delivered to the relevant tissues by engineered viruses or other vectors. Such artificially delivered DNA can provide the correct version of a defective disease gene. Alternatively, a disease process might be ameliorated by changing the expression of other regulatory genes. Such an approach goes straight to the root of biology and pathology and offers the prospect of an extremely powerful and completely new treatment modality. Application to psychiatric disorders faces formidable problems and is likely to occur in the relatively distant future. However, trials of gene therapy for other, simpler disorders are going on today.

The identification of disease genes will probably have a major impact on diagnosis and nosology. Just as the diagnosis of jaundice has given way to a classification scheme based on pathophysiology, in the future, the diagnosis of major mood disorders may be specific to disease mechanism. Mechanism-specific diagnoses may dictate the use of different, more specific and effective treatments. The field of pharmacogenetics promises genetic tests able to predict the best pharmacological treatment for specific patients based on an understanding of genetic differences in drug response. But such knowledge will also carry danger and responsibility. Premorbid DNA testing could become available that would indicate the degree of genetic vulnerability to major mood disorders. Those with psychiatric disorders will then face the same issues of genetic testing currently faced by families with Huntington's disease or breast cancer. Will at-risk family members want such information? How will they use it? How will they cope with it? How can psychiatrists assist them in these decisions? How can patients be protected from discrimination based on such information?

In summary, genetic studies promise a new era of understanding and treatment of mood disorders. Identification of genes will not elucidate pathophysiology. It will merely point the way for the application of modern neuroscience methods in the equally large task of understanding mechanisms. Recent results suggest that such guidance may not be far away.

SUGGESTED CROSS-REFERENCES

A general review of basic molecular genetics is provided in Section 1.10. Principles of population genetics are discussed in Section 1.17. The concepts and methods of linkage analysis and their application to psychiatry are reviewed in Section 1.18. The epidemiology of mood disorders is discussed in Section 14.2.

SECTION REFERENCES

Akiskal HS: The prevalent clinical spectrum of bipolar disorders: Beyond DSM-IV. J Clin Psychopharmacol *16:*4S, 1996.

*Aita VM, Liu J, Knowles JA, Terwilliger JD, Baltazar R, Grunn A, Loth JE, Kanyas K, Lerer B, Endicott J, Wang Z, Penchaszadeh G, Gilliam TC, Baron M: A comprehensive linkage analysis of chromosome 21q22 supports prior evidence for a putative bipolar affective disorder locus. Am J Hum Genet *64:*210, 1999.

Barden N, Morissette J, Shink E, Rochette D, Gagne B, Bordeleau L, Villeneuve A, Sher A, Shaw S, Hopkins P, Sherrington R: Confirmation of bipolar affective disorder susceptibility locus on chromosome 12 in the region of the Darier disease gene. Am J Med Genet *81:*475, 1998.

Baron M, Freimer NB, Risch N, Lerer B, Alexander JR, Straub RE, Asokan S, Das K, Peterson A, Amos J, Endicott J, Ott J, Gilliam TC: Diminished support for linkage between manic depressive illness and X-chromosome markers in three Israeli pedigrees. Nat Genet *3:*49, 1993.

Baron M, Risch N, Hamburger R, Mandel B, Kushner S, Newman M, Drumer D, Belmaker RH: Genetic linkage between X-chromosome markers and bipolar affective illness. Nature *326:*289, 1987.

*Berrettini WH, Ferraro TN, Goldin LR, Weeks DE, Detera-Wadleigh S, Nurnberger JIJ, Gershon ES: Chromosome 18 DNA markers and manic-depressive illness: Evidence for a susceptibility gene. Proc Natl Acad Sci USA *91:*5918, 1994.

Bertelsen A: A Danish twin study of manic-depressive disorders. In *Origin, Prevention and Treatment of Affective Disorder*, M Schou, E Stromgren, editors. Academic Press, London, 1979.

Blackwood DH, He L, Morris SW, McLean A, Whitton C, Thomson M, Walker MT, Woodburn K, Sharp CM, Wright AF, Shibasaki Y, St Clair DM, Porteous DJ, Muir WJ: A locus for bipolar affective disorder on chromosome 4p. Nat Genet *12:*427, 1996.

Coon H, Jensen S, Hoff M, Holik J, Plaetke R, Reimherr F, Wender P, Leppert M, Byerley W: A genome-wide search for genes predisposing to manic-depression, assuming autosomal dominant inheritance. Am J Hum Genet *52:*1234, 1993.

Dawson E, Parfitt E, Roberts Q, Daniels J, Lim L, Sham P, Nothen M, Propping P, Lanczik M, Maier W, Reuner U, Weissenbach J, Gill M, Powell J, McGuffin P, Owen M, Craddock N: Linkage studies of bipolar disorder in the region of the Darier's disease gene on chromosome 12q23-24.1. Am J Med Genet *60:*94, 1995.

Egeland JA, Gerhard DS, Pauls DL, Sussex JN, Kidd KK, Allen CR, Hostetter AM, Housman DE: Bipolar affective disorders linked to DNA markers on chromosome 11. Nature *325:*783, 1987.

Ewald H, Mors O, Flint T, Koed K, Eiberg H, Kruse TA: A possible locus for manic depressive illness on chromosome 16p13. Psychiatr Genet *5:*71, 1995.

Freimer NB, Reus VI, Escamilla MA, McInnes LA, Spesny M, Leon P, Service SK, Smith LB, Silva S, Rojas E, Gallegos A, Meza L, Fournier E, Baharloo S, Blankenship K, Tyler DJ, Batki S, Vinogradov S, Weissenbach J, Barondes SH, Sandkuijl LA: Genetic mapping using haplotype, association and linkage methods suggests a locus for severe bipolar disorder (BPI) at 18q22-q23. Nat Genet *12:*436, 1996.

Gershon ES, Hamovit J, Guroff JJ, Dibble E, Leckman JF, Sceery W, Targum SD, Nurnberger JI, Goldin LR, Bunney WE: A family study of schizoaffective, bipolar I, bipolar II, unipolar, and normal control probands. Arch Gen Psychiatry *39:*1157, 1982.

Ginns EI, Ott J, Egeland JA, Allen CR, Fann CS, Pauls DL, Weissenbachoff J, Carulli JP, Falls KM, Keith TP, Paul SM: A genome-wide search for chromosomal loci linked to bipolar affective disorder in the Old Order Amish. Nat Genet *12:*431, 1996.

Gurling H, Smyth C, Kalsi G, Moloney E, Rifkin L, O'Neill J, Murphy P, Curtis D, Petursson H, Brynjolfsson J: Linkage findings in bipolar disorder [Letter]. Nat Genet *10:*8, 1995.

Kallman F: Genetic principles in manic-depressive psychosis. In *Depression*, PH Hoch, J Zubin, editors. Grune & Stratton, New York, 1954.

Kelsoe JR, Ginns EI, Egeland JA, Gerhard DS, Goldstein AM, Bale SJ, Pauls DL, Long RT, Kidd KK, Conte G, Housman DE, Paul SM: Re-evaluation of the linkage relationship between chromosome 11p loci and the gene for bipolar affective disorder in the Old Order Amish. Nature *342:*238, 1989.

Kelsoe JR, Sadovnick AD, Kristbjarnarson H, Bergesch P, Mroczkowski-Parker Z, Drennan MD, Rapaport MH, Flodman P, Spence MA, Remick RA: Possible locus for bipolar disorder near the dopamine transporter on chromosome 5. Am J Med Genet *67:*533, 1996.

*Kendler KS, Pedersen N, Johnson L, Neale MC, Mathe AA: A pilot Swedish twin study of affective illness, including hospital- and population-ascertained subsamples. Arch Gen Psychiatry *50:*699, 1993.

*Kendler KS, Prescott CA: A twin study of major depression. Arch Gen Psychiatry *56:*39, 1999.

Lachman HM, Kelsoe JR, Remick RA, Sadovnick AD, Rapaport MH, Lin M, Pazur BA, Roe AA, Saito T, Papolos DF: Linkage studies support a possible locus for bipolar disorder in the velocardiofacial syndrome region on chromosome 22. Am J Med Genet *74:*121, 1996.

Leboyer M, Malafosse A, Boularand S, Campion D, Gheysen F, Samolyk D, Henriksson B, Denise E, des Lauriers A, Lepine JP: Tyrosine hydroxylase polymorphisms associated with manic-depressive illness. Lancet *335:*1219, 1990.

MacKinnon DF, Xu J, McMahon FJ, Simpson SG, Stine OC, McInnis MG, DePaulo JR: Bipolar disorder and panic disorder in families: An analysis of chromosome 18 data. Am J Psychiatry *155:*829, 1998.

Marazita ML, Neiswanger K, Cooper M, Zubenko GS, Giles DE, Frank E, Kupfer DJ, Kaplan BB: Genetic segregation analysis of early-onset recurrent unipolar depression. Am J Hum Genet *61:*1370, 1997.

Mendlewicz J, Rainer JD: Adoption study supporting genetic transmission in manic-depressive illness. Nature *268:*327, 1977.

*Nurnberger JIJ, DePaulo JR, Gershon ES, Reich T, Blehar MC, Edenberg HJ, Foroud T, Miller M, Bowman E, Mayeda A, Rau NL, Smiley C, Conneally PM, McMahon F, Meyers D, Simpson S, McInnis M, Stine OC, Detera-Wadleigh S, Goldin L, Guroff J, Maxwell E, Kazuba D, Gejman PV, Badner JA, Sanders AR, Rice J, Bierut L, Goate A: Genomic survey of bipolar illness in the NIMH genetics initiative pedigrees: A preliminary report. Am J Med Genet *74:*227, 1997.

Pekkarinen P, Terwilliger J, Bredbacka PE, Lonnqvist J, Peltonen L: Evidence of a predisposing locus to bipolar disorder on xq24–q27.1 in an extended Finnish pedigree. Genome Res *5:*105, 1995.

*Rice J, Reich T, Andreasen NC, Endicott J, van Eerdewegh M, Fishman R, Hirschfeld RM, Klerman GL: The familial transmission of bipolar illness. Arch Gen Psychiatry *44:*441, 1987.

Rosanoff AJ, Handy L, Plesset IR: The etiology of manic-depressive syndromes with special reference to their occurrence in twins. Am J Psychiatry *91:*725, 1935.

Sadovnick AD, Remick RA, Lam R, Zis AP, Yee IM, Huggins MJ, Baird PA: A mood disorder service genetic database: Morbidity risks for mood disorders in 3,942 first-degree relatives of 671 index cases with single depression, recurrent depression, bipolar I or bipolar II. Am J Med Genet *54:*132, 1994.

Schwab SG, Hallmayer J, Lerer B, Albus M, Borrmann M, Honig S, Strauss M, Segman R, Lichtermann D, Knapp M, Trixler M, Maier W, Wildenauer DB: Support for a chromosome 18p locus conferring susceptibility to functional psychoses in families with schizophrenia, by association and linkage analysis. Am J Hum Genet *63:*1139–1152, 1998.

Stine OC, Xu J, Koskela R, McMahon FJ, Gschwend M, Friddle C, Clark CD, McInnis MG, Simpson SG, Breschel TS, Vishio E, Riskin K, Feilotter H, Chen E, Shen S, Folstein S, Meyers DA, Botstein D, Marr TG, DePaulo JR: Evidence for linkage of bipolar disorder to chromosome 18 with a parent-of-origin effect. Am J Hum Genet *57:*1384, 1995.

Straub RE, Lehner T, Luo Y, Loth JE, Shao W, Sharpe L, Alexander JR, Das K, Simon R, Fieve RR, Lerer B, Endicott J, Ott J, Gilliam TC, Baron M: A possible vulnerability locus for bipolar affective disorder on chromosome 21q22.3. Nat Genet *8:*291, 1994.

*Tsuang MT, Faraone SV: *The Genetics of Mood Disorders.* Johns Hopkins University Press, Baltimore, 1990.

*Vincent JB, Masellis M, Lawrence J, Choi V, Gurling HM, Parikh SV, Kennedy JL: Genetic association analysis of serotonin system genes in bipolar affective disorder. Am J Psychiatry *156:*136, 1999.

14.4 MOOD DISORDERS: NEUROBIOLOGY

MICHAEL E. THASE, M.D.

The role of physical or biological factors in the pathogenesis of depression has been suspected since antiquity. However, it has only been in the latter half of the twentieth century that the technology and experimental methodology have been available to study these processes in mood disorder. During the past 40 years knowledge about the biology of the depressive disorders has grown by leaps and bounds as experimental methods have advanced from studies assaying peripheral specimens to investigations employing more direct measurements of cerebral metabolism, receptor function, and gene activity; regarding mania and bipolar disorders, however, there is comparatively less data.

CLINICAL PHENOMENOLOGY AND THE BRAIN

The signs, symptoms, and subjective experiences of depressed people have long suggested dysfunction of core neurobiological processes. At one level depression involves multiple disturbances of information processing. People who are severely depressed automatically interpret experiences negatively and their access to memories is similarly biased. Cognition and problem-solving skills are further compromised by poor concentration, loss of abstraction, and decreased ability to perform effortful mental tasks. The virtual monologue of negative thoughts and images is often difficult to disrupt and, unlike normal sadness, ventilation to a confidante typically has little beneficial effect. In more extreme cases, delusions or hallucinations grossly distort reality testing. Mental life is thus disturbed by simultaneous excesses and deficits of activity in specific cortical regions, including the hippocampus, prefrontal cortex, and other limbic structures.

Many severe depressions are characterized by a loss of mood reactivity. In essence, something that should elicit an uplift in mood does not do so. One correlate of this phenomenon is an apparent loss of reinforcer salience. Hedonic deficits also are apparent in a loss of gusto for food or sex and there can be significant weight loss, especially in the elderly. These disturbances point to abnormalities in neuroregulatory circuits involving the hypothalamus, nucleus accumbens, and thalamus.

Another correlate of decreased mood reactivity is a fixity or slowing of expression, initiation of activity, and spontaneous movement. This loss of animation, called *psychomotor retardation*, is sometimes confounded by a superimposed state of “driven” psychomotor agitation. These stereotypic behaviors characteristically include biting at the nailbeds or lips, a furrowed brow, tugging at one's hair, and purposeless scratching, in addition to the pacing and frequent postural shifts that accompany most states of dysphoric arousal. Like hedonic deficits, psychomotor disturbances are more commonly observed in older people with depression, with agitation typically seen in the most severe or psychotic depressive syndromes. These observable signs point to the probable involvement of circuits joining prefrontal cortex, thalamus, striatum, and basal ganglia.

Almost all depressed people report fatigue and an inability to feel refreshed by sleep. Some, particularly those under age 40, sleep excessively, whereas others, especially those over age 50, cannot maintain sleep or awaken spontaneously 1 to 2 hours earlier than desired. For those with such “terminal” insomnia, the morning hours are typically the worst of the day (i.e., *diurnal mood variation*). Such circadian disturbances strengthen the likelihood of thalamic dysregulation and also suggest abnormalities in the pons and medulla.

These signs and symptoms often coaggregate, which has formed the basis for several of the classical clinical subtypes of depression. Thus, anhedonia and psychomotor disturbances are more commonly associated with weight loss, middle and terminal insomnia, preservative ruminations, and diurnal mood disturbance, a constellation of the syndrome classified as an endogenous, autonomous, biological, psychotic, or vital depressive subtype. The fact that such depressions are slow to remit spontaneously, yet relatively responsive to electroconvulsive therapy (ECT), clearly reinforce perceptions of underlying biological abnormalities. Such depressions often appear to arise without provocation by a significant stressor, hence the presumption of an autonomous or endogenous etiology. The apparent stability of this syndrome over the centuries is recognized explicitly by the contemporary use of the ancient Greek term *melancholia*, even though black bile is no longer considered to be an etiological factor.

The depressions of early adult life tend to be “nonendogenous” and often seem to be imbedded with manifold psychosocial problems. Of course, an early age of onset may distort personality development and shape one's reportoire of coping skills, which increases the probability of handling stressors poorly. Moreover, such patients seem to elicit, provoke, or stumble into a disproportionately greater number of stressors. Historically, the association of an early age of onset with comorbid anxiety generally further solidified impressions of a conflicted or “neurotic depression.” The presumed lesser role of biological factors in neurotic depression was further suggested by less than gratifying responses to ECT and, when available, to tricyclic drugs. In addition to the nonendogenous and neurotic appellations, these early-onset depressions traditionally have been described as reactive, personal, or exogenous disorders. Many of these depressions run a chronic course that fluctuates between major and minor (dysthymic) levels.

Emerging data on the utility of new antidepressant drugs, particularly serotonergic agents, suggests that biological abnormalities of a different kind might be associated with nonendogenous depression.

A subset of nonendogenous depression was further described by “reverse” neurovegetative symptoms, namely, increased appetite, weight gain, and hypersomnolence. Such symptoms were considered uncommon in the late 1950s and 1960s and hence were referred to as *atypical.* However, as the average age of onset of depression declined into the early 20s and both research and practice has shifted to ambulatory settings, these atypical features are now commonplace. Reverse neurovegetative features are common in several forms of recurrent depression, including seasonal type (winter pattern), bipolar disorders, and major depressive disorder. These subtypes are characterized by an early age of onset and relatively poorer responses to tricyclic medications, as compared to monoamine oxidase inhibitors (MAOIs). Such observations suggest that differences in phenomenology could reflect more specific pathophysiological alterations that affect treatment response.

GENETIC FACTORS

Mood disorders clearly run in families and, as a result, both genetic and environmental influences must be considered. At present, it is well established that bipolar I disorder is more heritable than the other mood disorders, that an early age of onset is associated with greater heritability, and that heritable risk decreases in proportion to the amount of genetic material shared by members of a pedigree. Thus, identical twins have greater heritable risk than fraternal twins, who have a risk comparable to siblings and parents. First-degree relatives, in turn, have greater risk than half-siblings, grandparents, or cousins. Research on fraternal and identical twins or use of the adopted-away paradigm further suggests that heritable risk transcends environmental influences. Thus, vulnerability to mood disorders is, in some fashion, encoded in human deoxynucleic acid (DNA). The next generation of molecular genetic studies should help to identify the specific genes and gene products that influence this risk.

Although the significance of inherited vulnerability cannot be disputed, considerable variability exists both within and across groups with comparable risks. For example, in some at-risk pedigrees, alcohol use disorders and sociopathy are more common among male probands than depression. Among women, there is some evidence that environmental experiences may influence whether a generalized anxiety disorder or a depressive disorder is expressed. The late-onset mood disorders (i.e., those beginning past age 60) have relatively low rates of heritability. Even identical twins do not have 100 percent concordance of bipolar I disorder. Genetics therefore constitutes just one pathway of vulnerability or, perhaps more accurately, a foundation upon which other biopsychosocial risk factors may be expressed.

ETIOLOGY OF EMOTIONAL RESPONSES

The mood disorders are conditions in which normal emotional responses are distorted into more extreme and persistent manifestations. Some consider normal sadness to be an analogue of depressed mood, and elation the normal counterpart of manic euphoria. Continuity between normal and pathological mood states is illustrated by grief or bereavement. Grief is a universal experience, yet it can segue almost inexorably into a state that is so extreme and disabling that professional attention is warranted. Although there are less obvious natural parallels to mania, it is not uncommon for hypomania to go unrecognized. For others, the intensity of new romantic love is associated with changes in perception, cognition, behavior, and judgment that border on manic excitement. The euphorogenic effects of cocaine and other psychostimulants amply document that the “hard wiring” for manic states exists and is expressed if the individual is exposed to the right neurochemical milieu.

Predipositions to various emotional states, mood set points, appear to be partly heritable; this background emotional tone is referred to as *temperament.* Of particular importance to our current topic is a temperament characterized by excessive reactivity and behavioral inhibition, which can be recognized in infancy and is associated with enduring vulnerability to both depression, and anxiety, in adolescence and young adulthood. Similarly, onset of dysthymic traits in childhood presages extremely high rates of depressive and bipolar disorders in adolescence and adulthood. Modern theories of temperament and personality converage around several dimensions, including harm avoidance (neuroticism or behavioral inhibition). In bipolar disorders cyclothymic, extraverted, and novelty-seeking traits are more relevant.

There is now considerable evidence that the basic emotions are manifest across human cultures. The capacity to recognize such emotions in others can be observed so early in infancy that innate processes are undoubtedly implicated. Sadness and crying, for example, may be viewed as universal distress signals. Behaviors consistent with basic emotional states can also be observed in other mammals. For example, certain behavioral predispositions towards aggressivity (or passivity) and social dominance (or subordination) appear to extend across vertebrate species. By virtue of the large neocortex, *Homo sapiens* are distinguished by the capacity to integrate, abstract, and synthesize complex symbolic representations of life experiences; to communicate experiences in direct, abstract, and elaborated forms; and to develop and maintain a self-concept that guides behavior in relation to others and an anticipated future. The domains of competence are similarly broader and diverse. Whereas a highly competent primate may earn a position of dominance if placed in a new group, only the human may seek such a position by drafting a curriculum vitae and lobbying vigorously, and only humans can intentionally misrepresent their competence or motivations, or be affected by the larger social network's knowledge of their past performances.

The importance and intensity of human attachment bonds serves to facilitate the protracted task of childrearing and the necessary advantages of kinship. Indeed, for hundreds of thousands of years *Homo sapiens* adapted in a world in which the average life expectancy was less than 40 years and the infant mortality rate approached 50 percent. Although perhaps politically incorrect, it is not simply social Darwinism to suggest that necessary divisions of labor increased the chances of viable offspring and shaped evolution of gender differences in affectivity, affiliative behavior, and nuturance. Humans are also the slowest to mature of the mammals and functionally the most vulnerable and dependent upon caregivers for the longest period of time.

Having evolved over at least 900,000 years, humans must now deal with the massive changes in environment and social structure that have taken place much too rapidly for natural selection to keep pace. In relative terms, the breathtaking sociocultural and technological changes of the twentieth century have occurred in less than one ten-thousandth of human experience! The resulting complexity of human social systems thus often overtaxes the integrative capacities of the large human neocortex. When such capacities are overwhelmed, humans have to cope and adapt with neural mechanisms that are phylogenetically similar to those of other mammals and vertebrates.

STRESS AND ANIMAL MODELS OF DEPRESSION

Studies of rodents, dogs, cats, and nonhuman primates have confirmed that acute stress responses are characterized by activation of central and peripheral components of two interactive psychoneuroendocrine systems, the cortical-hypothalamic-pituitary-adrenocortical axis and the cortical-sympathomedullary axis. Although such acute stress responses are more akin to fear and anxiety than depression, it is the concepts of uncontrollability and inescapability that link the responses of other mammals to stress to human depression. Whereas stress acutely signals threat, it is loss, the anticipation of loss, or hopelessness that elicits sadness and despair.

Acute stress activates noradrenergic cell bodies in the locus ceruleus, whose axons trigger increased noradrenergic output by the adrenal medulla. At a behavioral level there is heightened arousal, increased perceptual vigilance, and inhibition of consummatory activities such as hunger and sex.

Stress also elicits synthesis and release of corticotropin-releasing hormone from neurons in the hypothalamus and cerebral cortex. This hormone activates the pituitary-adrenocortical components of stress response by causing increased release and synthesis of cortisol, adrenocorticotrophic hormone, and other glucocorticoids, and it synergistically enhances locus ceruleus activity as well.

The acute response to stress is counterbalanced by homeostatic or adaptive mechanisms. These include feedback inhibition by glucocorticoid receptors in the hippocampus and pituitary, down-regulation of postsynaptic noradrenergic receptors, and inhibitory autoreceptors and heteroreceptors on presynaptic norepinephrine neurons. Parallel input from serotoninergic (5-HT) and g-aminobutyric acid (GABA) neurons also exert dampening or inhibitory effects.

Exposure to prolonged, inescapable stress is associated with numerous adaptations and changes in neurobehavioral responses. Although corticosteroid levels may remain elevated, levels of norepinephrine, serotonin dopamine, and GABA in the brainstem and forebrain eventually decrease. Animals trapped in such a state cease trying to resist or escape, and show decreased grooming and appetitive behavior outside of the experimental situation. There are significant individual differences in development of such learned helplessness, as well as differences across pedigrees, or strains, and species. Nevertheless, antidepressant drugs have been shown to prevent, attenuate, or reverse learned helplessness across species.

Studies of the experiences of primates in the wild are more relevant to the stresses faced by people than a rat's response to inescapable painful electric shock. These naturalistic studies demonstrate that a fall from a dominant role within a primate social hierarchy is associated with increased cortical-hypothalamic-pituitary-adrenocortical activity and decreased 5 serotonin neurotransmission. In the wild, subordinate animals with low concentrations of the serotoninergic metabolites 5-hydroxyinoleacetic acid (5-HIAA) are likely to be more aggressive and less sexually active. Moreover, when social status is manipulated by creating a new group from a cohort of subordinate primates, a dominance hierarchy will emerge and the winner will experience a corresponding increase in serotonergic function. Conversely, during times of adversity such as drought or famine, the socially dominant primate experiences an increase in cortical-hypothalamic-pituitary-adrenocortical activity.

Early Adversity Physical, verbal, and sexual abuse have an indelible effect on the life trajectory. Although maltreatment has been well documented in the pathogenesis of posttraumatic stress disorder and borderline personality disorder, it now appears that such a history is also associated with an increased risk of depression.

Studies of animal models confirm that lasting alterations in neuroendocrine behavioral response can result from severe early stress. More recently this vulnerability has been linked to possible enduring changes in gene expression. Animal studies indicate that even transient periods of maternal deprivation can have a similar effect on subsequent response to stress. Early loss and neglect, a correlate of having been raised by parents with a mood disorder, is not uncommon in the history of mood disorder patients. How to integrate these clinical observations with the data on early stress in animals remains a challenge.

MONOAMINE DISTURBANCES

The area of scientific inquiry chosen is dependent on the knowledge base and experimental paradigms available to the investigator. In the early 1960s it was possible to measure catecholamine metabolites in body fluids and the principal indoleamine metabolite 5-HIAA in cerebrospinal fluid. However, the visualization of the functional brain was essentially impossible because waking electroencephalograms (EEGs) provided little useful information aside from documentation of epilepsy or diffusing slowing associated with delirium.

In addition, there were multiple lines of evidence from pharmacological studies implicating perturbations of monoamine systems in both therapeutic and iatrogenic effects of drugs on mood and behavior. Relevant pharmacotherapies of the time included tricyclic antidepressants, MAOIs, dextroamphetamine (Dexedrine), and the

amine-depleting compound reserpine (Serpasil).

Although the early biogenic amine hypotheses have undergone much revision, the critical importance of norepinephrine and 5-HT in the pathophysiology and physiology of mood disorders remains unquestioned. A second important monoamine, dopamine, has received less emphasis and there is also strong evidence of dysfunction in some forms of depressions. Since the mid-1960s there has been a shift in focus away from single neurotransmitters towards neurobehavioral systems, neural circuits, and more intricate regulatory mechanisms. Moreover, postsynaptic receptor families, presynaptic autoreceptors and heteroreceptors, second messengers, and gene transcription factors were not known when the original monoamine hypotheses were formulated.

Noradrenergic Systems Noradrenergic neurons have their cell bodies in the locus ceruleus of the brainstem and project rostrally to the cerebral cortex, limbic system, basal ganglia, hypothalamus, and thalamus (Fig. 14.4-1). This diffuse distribution belies norepinephrine's role in initiating and maintaining limbic and cortical arousal, as well as in modulation of the function of other neurotransmitters. Noradrenergic projections to the hippocampus have recently been implicated in behavioral sensitization to stress, and prolonged activation of the locus cerulus contributes to the state of learned helplessness. The locus ceruleus also is the origin of neurons that project to the adrenal medulla, the principal source of norepinephrine into the peripheral blood circulation.

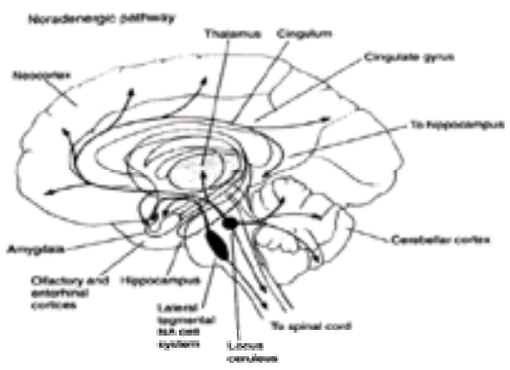

FIGURE 14.4-1 A lateral view of the brain demonstrates the course of the major nonadrenergic pathways emanating from the locus ceruleus and from the lateral brain stem tegmentum. (Reprinted with permission from Kandel ER, Schwartz JH, Jessell TM, editors: *Principles of Neural Science*, ed 3. Appleton & Lange, Stanford, CT, 1991.)

Novel stimuli increase the activity of the locus ceruleus, which in turn is decreased during vegetative functions, such as eating or sleeping. Cognitive processes can amplify or dampen sympathoadrenal responses to internal or external stimuli. Thus, the perception of stress is relayed via the appropriate cortical structures through the thalamus to the locus ceruleus and sympathoadrenal components of the acute stress response.

Stimulation of the medial forebrain bundle, the second major norepinephrine pathway in the brain, elicits increased levels of goal-directed and reward-seeking behavior. Sustained stress also eventually results in decreased levels of norepinephrine in the medial forebrain bundle, which may account for anergia, anhedonia, and diminished libido in depression. Increased noradrenergic output also stimulates inhibitory a_2-adrenergic heteroreceptors on serotonergic neurons.

Serotoninergic Systems Serotoninergic neurons project from the brainstem dorsal raphe nuclei to the cerebral cortex, hypothalamus, thalamus, basal ganglia, septum, and hippocampus (Fig. 14.4-2). Serotonin pathways have both inhibitory and facilitatory functions in the brain. For example, much evidence suggests that 5-HT is an important regulator of sleep, appetite, and libido. Serotonergic neurons projecting to the suprachiasmatic nucleus of the hypothalamus help to regulate circadian rhythms (e.g., sleep-wake cycles, body temperature, and hypothalamic-pituitary-adrenocortical axis function). Serotonin also permits or facilitates goal-directed motor and consummatory behaviors in conjunction with norepinephrine and dopamine. Moreover, serotonin inhibits aggressive behavior across mammalian and reptilian species.

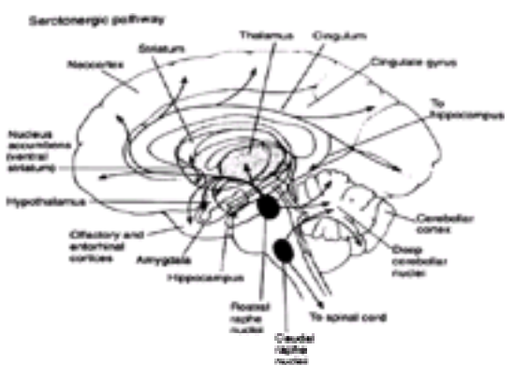

FIGURE 14.4-2 A lateral view of the brain demonstrates the course of the major serotonergic pathways. Although the raphe nuclei form a fairly continuous collection of cell groups throughout the brain stem, they are graphically illustrated here as two groups, one rostral and one caudal. (Reprinted with permission from Kandel ER, Schwartz JH, Jessell TM, editors: *Principles of Neural Science*, ed 3. Appleton & Lange, Stanford, CT, 1991.)

There is some evidence that serotonin neurotransmission is partly under genetic control. Nevertheless, acute stress increases serotonin release transiently, whereas chronic stress eventually will deplete serotonin stores. Chronic stress may also increase synthesis of 5-HT_{1A} autoreceptors in the dorsal raphe nucleus, which further decrease serotonin transmission. Elevated glucocorticoid levels tend to enhance serotonergic functioning and thus may have significant compensatory effects in chronic stress.

Dopaminergic Systems There are four relatively discrete dopamine pathways in the brain (Fig. 14.4-3). The tuberoinfundibular system projects from cell bodies in the hypothalamus to the pituitary stalk, exerting inhibitory control over prolactin secretion. The nigrostriatal system originates from cell bodies in the substantia nigra and projects to the basal ganglia, regulating involuntary motor activity. The cell bodies of the mesolimbic pathway are located in the ventral tegmentum and project to almost all limbic regions: the nucleus accumbens, amygdala, hippocampus, medial dorsal nucleus of the thalamus, and cingulate gyrus. The mesolimbic dopamine pathway modulates emotional expression, learning and reinforcement, and hedonic capacity. The fourth dopamine pathway, also originating in the ventral tegmentum's mesocortical pathway, which projects to the orbitofrontal and the prefrontal cortical regions, helping to regulate motivation, concentration, initiation of goal-directed and complex, executive cognitive tasks. Decreases in mesocortical and mesolimbic dopamine activity have obvious implications in the cognitive, motor, and hedonic disturbances associated with depression. Moreover, dopamine activity appears to be potentiated by nicotinic inputs and glucocorticoids, and dopamine concentrations are correlated with brain serotonin activity.

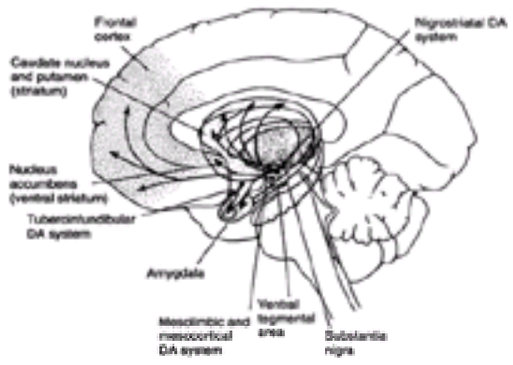

FIGURE 14.4-3 A lateral view of the brain demonstrates the course of the four major dopaminergic tracts. (Reprinted with permission from Kandel ER, Schwartz JH, Jessell TM, editors: *Principles of Neural Science*, ed 3. Appleton & Lange, Stanford, CT, 1991.)

Biogenic Amine Function After nearly 30 years of research it can be concluded that subsets of depressed people manifest one or more abnormalities of monoamine neurotransmission. Decreased central norepinephrine activity can be inferred, in part, from decreased urinary excretion of the metabolite 3-methoxy-4-hydroxyphenylglycol. A partly overlapping subgroup of patients has elevated circulating levels of norepinephrine and its metabolites. This suggests a dissociation of norepinephrine activity in the brain's medial forebrain bundle and the sympathomedullary systems peripheral activities. Increased noradrenergic activity also is reflected by blunted a_1, b, and b-coupled second messenger (i.e., cyclic adenosine monophosphate) responses. Further, an acute response to noradrenergically active antidepressants (i.e., desipramine, nortriptyline, or bupropion) may be transiently reversed by the norepinephrine synthesis inhibitor a-methyl-paratyrosine.

Serotonin dysfunction has been documented in overlapping subgroups of patients using a variety of methods, ranging from low cerebrospinal fluid (CSF) levels of 5-HIAA to decreased cerebral metabolism. Serotonin dysfunction also is reflected by blunted responses to specific serotonin (5-hydroxytryptamine [5-HT]) HT_{1A} subtype 1A ($5\text{-}HT_{1A}$) agonists (e.g., ipsapirone) and nonselective agonists (e.g., L-tryptophan or dex-fenfluramine), decreased neuroendocrine responses, and decreased serotonin uptake sites on blood platelets. Decreased serotonin neurotransmission can be inferred from the findings of cortical-hypothalamic-pituitary-adrenocortical and EEG sleep studies. In functional terms, a state of a relative hypofrontality of cerebral blood flow and glucose metabolism in the brain is fully consistent with decreased neurotransmission by 5-HT neurons projecting from the dorsal raphe nuclei. In support of this observation, dietary depletion of L-tryptophan will induce this abnormality in a subset of vulnerable patients. Tryptophan depletion also reverses acute responses to selective serotonin reuptake inhibitors. Increased sensitivity to dopamine, perhaps mediated by elevated glucocorticoid levels, may contribute to the development of delusions and hallucinations.

Other Neurotransmitters Cholinergic neurons containing acetylcholine are distributed diffusely throughout the cerebral cortex, and have reciprocal or interactive relationships with all three monoamine systems. Abnormal levels of choline, which is a precursor to acetylcholine, are seen in the brains of some depressed patients, perhaps reflecting abnormalities in cell phospholipid composition. Cholinergic agonist and antagonist drugs have differential clinical effects on depression and mania. Agonists can produce lethargy, anergia, and psychomotor retardation in normal subjects; exacerbate symptoms in depression; and reduce symptoms in mania. These effects generally are not sufficiently robust to have clinical applications and adverse effects limit their clinical utility. Via their serotonergic or adrenergic effects, antidepressant drugs may decrease cholinergic function, although direct anticholinergic effects are unrelated to antidepressant activity.

In an animal model of depression, a strain of mice that is supersensitive to cholinergic effects has been found to develop learned helplessness more quickly and cholinergic supersensitivity has been shown to be attenuated by manipulation of adrenergic activity. Conversely, cholinergic agonists can induce changes in cortical-hypothalamic-pituitary-adrenocortical activity and sleep EEG studies that mimic those associated with severe depression. Indeed, some remitted patients with bipolar or depressive disorders, as well as their never-ill first-degree relatives, have a trait-like increased sensitivity to cholinergic agonists.

GABA has an inhibitory effect on ascending monoamine pathways, particularly the mesocortical and mesolimbic systems. Reductions of GABA have been observed in plasma, CSF, and brain GABA concentrations in depression. Animal studies have also found that chronic stress can reduce or deplete GABA levels and, by contrast, GABA receptors can be upregulated by antidepressants.

The amino acids glutamate and glycine appear to be the major excitatory neurotransmitters in the central nervous system. Glutamate and glycine bind to sites associated with the *N*-methyl-D-aspartate (NMDA) receptor and, in excess, can have neurotoxic effects. The hippocampus has a high concentration of NMDA receptors; it is thus possible that glutamate in conjunction with hypercortisolemia mediate the neurocognitive effects of chronic stress. There is emerging evidence that drugs that antagonize NMDA receptors have antidepressant effects.

The binding of a neurotransmitter and postsynaptic receptor triggers a cascade of chemical processes that include the second messenger systems. Receptors interact with the intracellular environment via guanine nucleotide-binding proteins (G proteins). The G proteins, in turn, connect to various intracellular enzymes and effectors (e.g., adenylate cyclase, phospholipase C, and phosphodiesterase) that stimulate the formation of second messengers, such as cyclic nucleotide [e.g., cyclic adenosine monophosphate (cAMP) and cyclic guanine monophosphate (cGMP)], phosphatidylinositols (e.g., inositol triphosphate and diacylglycerol), and calcium-calmodulin. These second messengers regulate the function of neuronal membrane ion channels, neurotransmitter synthesis and release, and protein kinase activity. Protein kinase, for example, catalyzes phosphorylation, an anergy-liberating process involved in synthesis and degradation of neuroreceptors, ion channels, G proteins, and DNA transcription and messenger-RNA translation factors that regulate gene expression. Recent studies have reported abnormalities in platelet adenylate cyclase activity, phosphoinositide hydrolysis, intracellular calcium metabolism, and G-protein function in depressive disorders. Moreover, antidepressant medications may initiate a series of intracellular reactions that “turn down” synthesis of corticotropin-releasing hormone and monoamine receptors and “turn on” peptides such as neuronal growth factors. There is also increasing evidence that mood-stabilizing drugs (e.g., lithium) act upon G proteins or other second messengers.

Alterations of Hormonal Regulation

Cortical-Hypothalamic-Pituitary-Adrenal Cortical Activity Elevated glucocorticoid activity is a hallmark of the mammalian stress response. Evidence of increased cortisol secretion is apparent in 20 to 40 percent of depressed outpatients and 40 to 60 percent of depressed inpatients. Rates are highest among older patients, particularly those with highly recurrent or psychotic depressive disorders. Hypercortisolism is thus one of the most common correlates of melancholic depressions.

A variety of methods can be used to study elevated cortical-hypothalamic-pituitary-adrenal cortical activity: excretion of urinary free cortisol, 24-hour (or shorter time segments) intravenous collections of glucocorticoid plasma cortisol levels, salivary cortisol levels, and tests of the integrity of feedback inhibition. Methods of testing feedback inhibition usually involve administration of the potent synthetic glucocorticoid dexamethasone, which in 0.5-, 1-, or 2-mg doses normally suppresses cortical-hypothalamic-pituitary-adrenal cortical axis activity for 24 hours. Impaired feedback inhibition is reflected by nonsuppression of cortisol secretion at 8 AM the following morning or subsequent escape from suppression at 4 PM or 11 PM. A more recent development is the pairing of dexamethasone suppression with an infusion of corticotropin-releasing hormone. The sensitivity and specificity of these various tests of feedback inhibition are not sufficient for use and adrenocortical hyperactivity is observed in many other psychiatric disorders, albeit usually at a lower prevalence. Nonsuppression usually implicates a premature loss of inhibitory hippocampal glucocorticoid receptors, which also may account for the age dependence of cortisol nonsuppression. Hypercortisolemia associated with early trauma also may permanently decrease synthesis of glucocorticoid receptors or actually lead to atrophy of these vulnerable neurons.

Hypersecretion of cortisol and dexamethasone nonsuppression are imperfectly correlated (about 60 percent concordance). Elevated cortical-hypothalamic-pituitary-adrenal cortical activity in depression is typically not associated with the physical stigmata of Cushing's disease, but it is sufficient to induce a reversible cortical atrophy and is implicated in the genesis of neurocognitive disturbances. Starvation and protracted sleep deprivation also can induce hypercortisolism. Patients with increased cortical-hypothalamic-pituitary-adrenal cortical activity are typically less responsive to attention placebo and psychosocial treatments. However, hypercortisolism does typically resolve with effective treatment and, when persistent, conveys a high risk of relapse. This is presumed to be a consequence of incomplete resolution of the depressive episode at the level of the brain. Dexamethasone and the cortisol synthesis inhibitor ketoconazole (Nizoral Ò) are sometimes used to externally suppress the hypothalamic-pituitary-adrenocortical axis of hypercortisolemic patients with more refractory depressive disorders.

Beyond failure of feedback inhibition, a deficit of 5-HT activity and an increase in norepinephrine or acetylcholine activity have been shown to increase cortical-hypothalmic-pituitary-adrenal cortical activity.

Thyroid Axis Activity About 5 to 10 percent of people evaluated for depression have previously undetected or subclinical thyroid dysfunction, as reflected by an elevated basal thyroid-stimulating hormone (TSH) concentration or an increased TSH response to a 500-µg infusion of the hypothalamic neuropeptide thyrotropin-releasing hormone (TRH). Such abnormalities are often associated with elevated antithyroid antibody levels and, unless corrected with thyroid hormone replacement therapy, may compromise response to treatment. These findings are especially relevant to women with rapid-cycling bipolar disorder.

More commonly, depressed patients receiving a TRH challenge test show a blunted TSH response. This abnormality, which may be state-independent, has been associated with a heightened relapse risk following pharmacotherapy or ECT. The TSH response may represent pituitary downregulation consequent to a prolonged elevation of TRH secretion. In turn, increased TRH secretion could result from a homeostatic response intended to enhance noradrenergic neurotransmission. Some researchers further speculate that the therapeutic benefit of liothyronine (Cytomel) augmentation therapy is the result of correction of this failed homeostatic response.

Growth Hormone Growth hormone secretion from the anterior pituitary is stimulated by norepinephrine and dopamine and inhibited by CRH and somatostatin, a hypothalamic neuropeptide. Somatostatin also inhibits CRH secretion. Secretion of growth hormone follows a 24-hour circadian rhythm, with a characteristic secretory surge during the first few hours of sleep. The most consistent finding in depression is a blunted growth hormone response to clonidine, an a_2 receptor agonist. The onset of sleep and nonselective adrenergic agonists such as desipramine also elicit a blunted growth hormone response.

Somatostatin Although the hypothalamus has the highest concentrations of somatostatin, significant concentrations are also found in the amygdala, hippocampus, nucleus accumbens, prefrontal cortex, and locus ceruleus. In addition to inhibition of growth hormone and release of corticotropin-releasing hormone, somatostatin inhibits GABA, adrenocorticotropic hormone (ACTH) and thyroid-stimulating hormone. Somatostatin levels are lower in the cerebrospinal fluid of people with depression as compared to those with schizophrenia or normals, and increased levels have been observed in mania.

Prolactin Prolactin release from the pituitary is stimulated by serotonin and inhibited by dopamine. Most studies have not found significant abnormalities of basal or circadian prolactin secretion in depression, although a blunted prolactin response to various serotonin agonists has been frequently reported. This response is less likely to be abnormal in premenopausal women, suggesting that estrogen has a moderating effect.

Alterations of Sleep Neurophysiology People prone to depression tend to have a premature loss of deep, slow (delta) wave sleep, and an early onset of the first episode of rapid eye movement (REM) sleep (Fig. 14.4-4). Results of family and twin studies suggest that these related abnormalities are at least partly heritable. Consistent with the expected behavior of a heritable trait, reduced REM latency and deficits of slow-wave sleep typically persist following recovery from a depressive episode. A blunted growth hormone response following sleep onset or administration of adrenergic agonists is correlated with a slow-wave sleep deficit and shows similar state-independent or trait-like behavior.

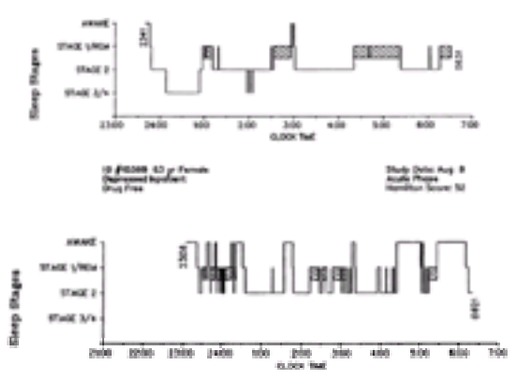

FIGURE 14.4-4 The all night electroencephalographic (EEG) sleep profiles of a healthy young woman and a 63-year-old woman with melancholia. (Reprinted with permission from Thase ME, Howland: Biological Processes in Depression: An Updated Review and Integration. In *Handbook of Depression*, ed 2, EE Beckham, WR Leber, editors. Guilford, New York, 1995.)

More severe depressions are associated with age-dependent decreases in sleep maintenance (i.e., the capacity to sleep without awakenings) and an increase in the phasic intensity of REM sleep, particularly during the first several REM periods. These changes correlate with clinical measures of severity of depression and tend to normalize during times of remission. Other state-dependent abnormalities are hypercortisolemia, dexamethasone nonsuppression, and elevated levels of peripheral catecholamine metabolites.

A deficit of slow-wave sleep and reduced REM latency can be induced and may be caused by a decrease in 5-HT neurotransmission or an increase in central cholinergic activity. Serotoninergic neurons from the dorsal raphe-nuclei project to cholinergic cells in the pons to tonically inhibit REM sleep. During non-REM (NREM) sleep, these cholinergic neurons are inactive. L-tryptophan depletion results in an increase in REM time and a decrease in REM latency.

Serotoninergic neurons projecting rostrally to the thalamus and prefrontal cortex also mediate slow-wave activity, as do drugs that antagonize 5-HT_2 receptors. Low CSF concentrations of 5-HIAA is correlated with diminished slow-wave sleep, an effect mimicked by acute serotonin depletion.

Prefrontal cortical metabolism is normally decreased during nREM sleep, a time of physical and metabolic rest. By contrast, REM sleep is normally associated with an increase in glucose metabolism in the limbic system. In depression there is relatively increased prefrontal glucose metabolism during nREM sleep, corresponding to the diminution of restorative slow wave sleep. REM sleep is associated with an even greater increase in limbic glucose utilization.

The sleep profiles of younger, hypersomnolent patients can be remarkably normal, particularly if laboratory routines do not permit an extended morning sleep period. Such patients may actually have increased slow-wave sleep and an increase in REM time associated with a greater total sleep time. Reduced REM latency, however, has been reported in juvenile depressives by at least one group of investigators.

The combination of reduced REM latency, increased REM density, and decreased sleep maintenance identifies about 40 percent of depressed outpatients, 80 percent of depressed inpatients, and 10 percent of age-matched normal controls. These sleep abnormalities also have been reported in a subset of dysthymic disorder patients. Sleep studies are too expensive and inconvenient to be used routinely for diagnostic purposes in depression and, like dexamethasone nonsuppression, false-positive cases are common in other psychiatric disorders. Nevertheless, current data indicate that patients manifesting this constellation of disturbances are less responsive to psychotherapies and may benefit preferentially from pharmacotherapy. It remains to be seen if those normal individuals with false-positive studies have an increased vulnerability to future episodes of depression. Successful nonpharmacological treatment of the depressive episode results in normalization of this profile in about 50 percent of cases. Pharmacotherapy with most antidepressant agents has an overcorrective effect characterized by prolongation of REM latency and suppression of REM sleep; such medications have variable effects on sleep maintenance. REM suppression may reflect activation of postsynaptic 5-HT_{1A} receptors or enhanced norepinephrine neurotransmission. Both desipramine (Norpramin) and the SSRIs for example, produce a rapid REM suppression in normal patients as well as in patients with depression. The efficacy of antidepressant medications that do not suppress REM sleep, such as nefazodone (Serzone), and bupropion (Wellbutrin) in patients with pathologically increased REM sleep suggests that these formulations may not apply to all depressions.

Alterations of Circadian Rhythms In addition to sleep disturbances, depressed patients often show a blunting of circadian rhythms of cortisol secretion, growth hormone secretion, and body temperature. These changes were originally thought to represent a phase advance of the sleep-wake cycle, although stronger evidence now suggests that circadian rhythms may be disorganized by heightened nocturnal arousal. Increased CRH, increased somatostatin, hypercortisolemia, and decreased 5-HT could all be implicated in this process. By contrast, in some cases of depressive disorders with seasonal (winter) pattern there is evidence of a phase delay of the sleep-wake cycle in relation to the nocturnal onset of melatonin secretion. The role of melatonin in the pathophysiology of mood disorders remains

unclear.

Brain Structure and Function

Structural Lesions Computed axial tomography and magnetic resonance imaging scans provide sensitive, noninvasive methods to assess the brain, including cortical and subcortical tracts, as well as white matter lesions. The most consistent abnormality observed in the depressive disorders is increased frequency of abnormal hyperintensities in subcortical regions, especially the periventricular area, basal ganglia, and thalamus (Fig. 14.4-5). More common in bipolar I disorder and among the elderly, these hyperintensities appear to reflect the deleterious neurodegenerative effects of recurrent mood episodes. Ventricular enlargement, cortical atrophy, and sulcal widening also have been reported in patients with mood disorders as compared to normal controls. In addition to age and illness duration, structural abnormalities are associated with increased illness severity, bipolar status, and increased cortisol levels. Some depressed patients also may have reduced caudate nucleus volumes, suggesting a defect in the mesocorticolimbic system. Cerebrovascular factors, including strong, often involve subcortical frontal and basal ganglia structures, and appear particularly relevant to late-life depression.

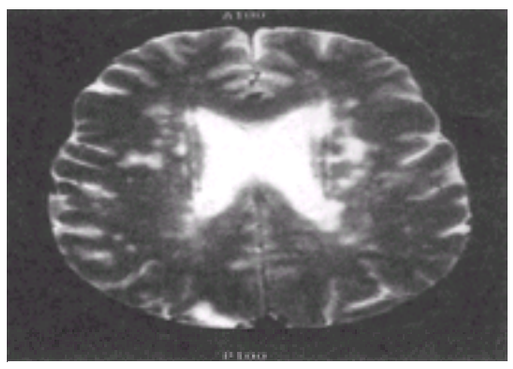

FIGURE 14.4-5 This magnetic resonance imaging (MRI) scan of a patient with late onset major depressive disorder illustrates extensive periventricular hyperintensities associated with diffuse cerebrovascular disease.

Cerebral Metabolic Alterations Positron emission tomography (PET) scanning is currently the most powerful method for visualizing brain metabolism during rest and various states of activation. Normal sadness is associated with an increase in cerebral blood flow to the thalamus and medial prefrontal cortex. This appears to be a nonspecific change associated with diverse emotional responses. More specific activation is seen in the left amygdala, hippocampal formation, and parahippocampal gyrus. Sadness generated by one's own thoughts (as opposed to a video scenario) also is associated with a relative increase in cerebral blood flow to the anterior insular cortex, as is anticipatory anxiety.

Direct activation of limbic structures by intravenous infusions of procaine hydrochloride also has been studied in normal controls. Such activation is characterized by reliable, bilateral increases in cerebral blood flow in the amygdala, parahippocampal gyri, insula, and anterior cingulate cortex. These changes are associated with a wide range of emotions, ranging from euphoria to severe anxiety.

The most widely replicated PET finding in depression is decreased anterior brain metabolism, which is generally more pronounced on the left side. This abnormality appears to be state dependent and has been observed in both depressive and bipolar disorders (Fig. 14.4-6), as well as in depression associated with obsessive-compulsive disorder. There is a reversal of hypofrontality following shifts from depression into hypomania, such that there are greater left-hemisphere reductions in depression compared to greater right-hemisphere reductions in mania.

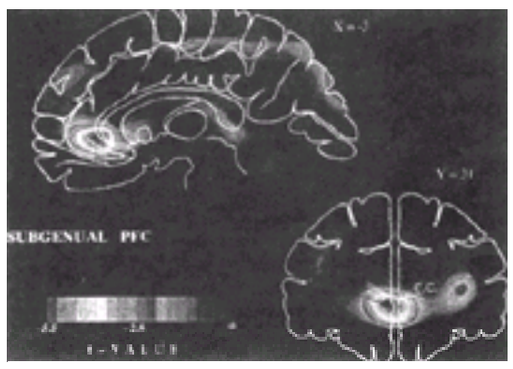

FIGURE 14.4-6 Composite coronal and sagittal sections of positron emission tomography (PET) scans show areas where cerebral glucose metabolism is decreased in depressed patients relative to matched health controls. (Reprinted with permission from Drevets WC: Nature 386:824–827, 1997. Ó 1997 MacMillan Magazine Limited.)

Other studies have observed more specific reductions of reduced cerebral blood flow and metabolism in the dopaminergically innervated tracts of the mesocortical and mesolimbic systems in depression. There is evidence that antidepressant agents at least partially normalize some of these changes.

In addition to a global reduction of anterior cerebral metabolism, increased glucose metabolism has been observed in several limbic regions. The best evidence of this abnormality comes from studies of patients with relatively severe recurrent depression and a family history of mood disorder (Fig. 14.4-7). This abnormality was found to be reversible with effective pharmacotherapy, but persistent when recently remitted patients were studied again when they were off medication. During episodes of depression, increased glucose metabolism is correlated with intrusive ruminations. If truly state-independent, such amygdalar hypermetabolism could represent the emotional “amplifier” that helps to distort the signal of relatively minor stressors in vulnerable people.

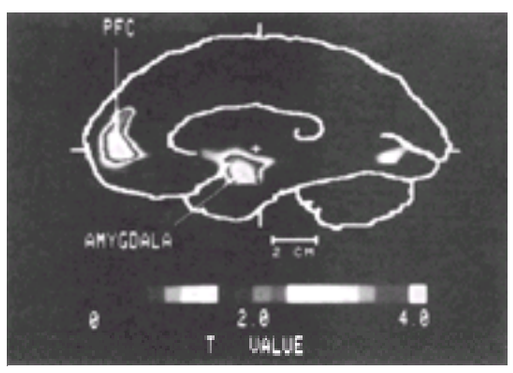

FIGURE 14.4-7 A lateral composite positron emission tomogram (PET) illustrating increased blood flow to the amygdala of patients with pure familial depressive disease as compared to healthy controls. (Reprinted with permission from Drevets WC: Exploring the Functional Anatomy of Depressive Disorder—Part I. In *Masters in Psychiatry*, Cliggott Communications, 1994.)

Magnetic resonance spectroscopy has recently been used to study brain phosphorus metabolism in mood disorders. These studies focus on phosphorus metabolites because they reflect metabolic activity of several second messengers, such as cAMP, cyclic guanine monophosphate, and phosphatidylinositol. An asymmetrical abnormality of phosphorus metabolism has been observed in the frontal lobes of patients with bipolar disorder compared to normal controls, as well as the left frontal lobe and basal ganglia of patients with depression.

IMMUNOLOGICAL DISTURBANCE

Depressive disorders are associated with several immunological abnormalities, including decreased lymphocyte proliferation in response to mitogens and other forms of impaired cellular immunity. These lymphocytes produce neuromodulators such as corticotropin-releasing factor and cytokines, which are peptides known as *interleukins.* There appears to be an association with clinical severity, hypercortisolism, and immune dysfunction and the cytokine interleukin-1 may induce gene activity for flucocorticoid synthesis. The precise clinical relevance of these findings requires further investigation.

SEX DIFFERENCES

There is no compelling evidence that a single, sex-related factor accounts for the increased risk of depression in women. Certainly, alterations of sex hormones or hypothalamic gonadotropins in depression have not been documented consistently. Prior to menopause, estrogen and its metabolites may actually “protect” depressed women from developing hypercortisolemia. Although probably not a specific causal factor in depression, menopause may thus represent a point of transition from less severe depressions characterized by reverse vegetative symptoms to melancholia. There are, however, a multitude of risk factors that may contribute to the “gender gap,” including greater risk of early sexual abuse and current spousal abuse, higher rates of thyroid disease, oral contraceptive use, and premenstrual or postpartum-onset mood disorders.

Alternatively, gender differences in emotional expressivity may interact with sociocultural factors and the texture of modern life to create the increased risk of depression for women. Conversely, use of external coping strategies and greater traditional social role expectations to achieve competence outside of the home (and the resultant economic power) may provide men relatively greater protection against depression. Gender differences in rates of depression are not well documented in nonindustrialized cultures and, in the United States, no differences in men's and women's rates of depression were found in the old-order Amish of Pennsylvania.

FUTURE DIRECTIONS

Major depressive disorder is associated with a myriad of neurobiological disturbances, perhaps as varied as the range of effective treatments and clinical presentations. Broadly viewed as traits (either inherited or acquired) or states (abnormalities only apparent during illness), these disturbances begin to show some psychobiological coherence. State-dependent abnormalities, for example, tend to coaggregate in patients with more severe syndromes and, especially past ages 40 to 50, are associated with the more classic endogenous or melancholic manifestations. These changes include increased phasic REM sleep, poor sleep maintenance, hypercortisolism, impaired cellular immunity, global reductions of anterior cerebral blood flow and glucose metabolism, elevated peripheral noradrenergic metabolites, and possibly increased glucose metabolism in the left amygdala. Such changes suggest, almost without exception, the consequences of an exaggerated and sustained stress response. Once manifest in this fashion, the depressive episode tends to be longer, more disabling, more prone to relapse, and more likely to benefit from pharmacotherapy or ECT.

Trait-like abnormalities include reduced slow-wave sleep, reduced REM latency, blunted nocturnal growth hormone response, and various indicators of decreased serotonin neurotransmission. Heritability of these abnormalities is inferred from family studies and other at-risk paradigms. These abnormalities are associated with an early age of onset and perhaps with increased vulnerability to recurrent illness. They may also increase the likelihood of state-dependent biological changes during depressive episodes, probably by reflecting impairments in the ability to dampen or lower stress responses.

Examples of more persistent but “acquired” abnormalities may include global and focal changes, cortical atrophy, hypertrophy of the adrenal cortex, periventricular hyperintensities, and alterations in CRH synthesis. Blunted response of thyroid-stimulating hormone to an infusion of thyrotropin-releasing hormone and dexamethasone nonsuppression may represent “hybrids,” in that these abnormalities can be slow to normalize and, when persistent after remission, convey a high risk of relapse.

Although specific genes have not yet been identified, vulnerability to mood disorders is heritable for some people. This type of heritability is most likely polygenetic and, in all likelihood, will be best understood through models that include gene-environment interactions. Nevertheless, increased heritability is associated with an earlier age of onset, greater comorbidity, increased risk of recurrent illness, and an increased likelihood of hypomanic or manic episodes.

Two of the more heritable forms of depression, early-onset chronic depression and bipolar depression, are commonly nonmelancholic in clinical presentation and relatively less likely to manifest state-dependent neurobiological disturbances. Both different genetic vulnerabilities and age-dependent changes in the brain's response to depression might explain this apparent paradox. Hypersomnolence and hyperphagia may thus reflect an episodic yet age-dependent homeostatic response to enhance serotoninergic dysfunction. In contrast, the sharp rise of mania post-pubescently may be related to maturational effects of catecholaminergic systems in the brain.

Aging and an accumulating risk of recurrent episodes are inextricably connected, although the diseases of aging that ravage brain function definitely increase the risk of depression, and these syndromes often prove to be resistant to treatment. The late-onset form of depression associated with periventricular hyperintensities also illustrates a more subtle interplay between vascular disease, brain damage, and mood disorder.

Ultimately, depression remains the most human of the Axis I disorders, partly because everyone can relate to sadness, grief, and the heartbreak of lost love and partly because the cognitive world of animals subjected to prolonged, inescapable stress is inaccessible. Even if their mental equivalents of “Why did this happen to me?” “Why bother—nothing I try will help?” and “I wish I was dead” could be identified, however, the incidence, prevalence, disability, and tragedy of depression in humans cannot be generalized to rodents or primates. Strong, sustaining affective bonds and an enduring sense of self-worth and competence are all-important and assaults on these fundamental aspects of human well-being are so frequent that many individuals succumb to depression. Some individuals are more vulnerable than others, and the association between severe depression and numerous, reproducible changes in brain function are well documented. Understanding the mechanisms of adaptation and brain dysfunction that predispose, initiate, distort, and maintain depressive disorders represents our best hope to prevent and relieve the misery and suffering of tens of millions of people.

SUGGESTED CROSS-REFERENCES

Monoamine neurotransmitters are discussed in Section 1.4, and the contributions of the neural sciences in general are the focus of the other sections of Chapter 1. Biological therapies are covered in Chapter 31.

SECTION REFERENCES

Akiskal HS: Interaction of biologic and psychologic factors in the origin of depressive disorders. Acta Psychiatr Scand *319:*131, 1985.

Alexopoulos GS, Meyers BS, Young RC, Campbell S, Silbersweig D, Charlson M: ‘Vascular depression' hypothesis. Arch Gen Psychiatry *54:*915, 1997.

*Anand A, Darnell A, Miller HL, Berman RM, Cappiello A, Oren DA, Woods SW, Charney DS: Effect of catecholamine depletion on lithium-induced long-term remission of bipolar disorder. Biol Psychiatry *45:*972, 1999.

Avissar S, Nechamkin Y, Roitman G, Schreiber G: Reduced G protein functions and immunoreactive levels in mononuclear leukocytes of patients with depression. Am J Psychiatry *154:*211, 1997.

Barden N: Modulation of glucocorticoid receptor gene expression by antidepressant drugs. Pharmacopsychiatry *29:*12, 1996.

Bremner JD, Innis RB, Salomon RM, Staib LH, Ng CK, Miller HL, Bronen RA, Krystal JH, Duncan J, Rich D, Price LH, Malison R, Dey H, Soufer R, Charney DS: Positron emission tomography

measurement of cerebral metabolic correlates of tryptophan depletion-induced depressive relapse. Arch Gen Psychiatry *54:*364, 1997.

*Cooper JR, Bloom FE, Roth RH: *The Biochemical Basis of Neuropharmacology*, ed 3. Oxford University Press, New York, 1996.

Delgado PL, Charney DS, Price LH, Aghajanian GK, Landis H, Heninger GR: Serotonin function and the mechanism of antidepressant action: Reversal of antidepressant-induced remission by rapid depletion of plasma tryptophan. Arch Gen Psychiatry *47:*411, 1990.

Drevets WC, Videen TO, Price JL, Preskorn SH, Carmichael ST, Raichle ME: A functional anatomical study of unipolar depression. J Neurosci *12:*3628, 1992.

*Duman RS, Heninger GR, Nestler EJ: A molecular and cellular theory of depression. Arch Gen Psychiatry *54:*597, 1997.

Higley JD, Thompson WW, Champoux M, Goldman D, Hasert MF, Kraemer GW, Scanlan JM, Suomi S, Linnoila M: Paternal and maternal genetic and environmental contributions to cerebrospinal fluid monoamine metabolites in rheusus monkeys (Macaca mulatta). Arch Gen Psychiatry *50:*615, 1993.

Holsboer F: Neuroendocrinology of mood disorders. In *Psychopharmacology: The Fourth Generation of Progress.* FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Kapur S, Mann JJ: Role of the dopaminergic system in depression. Biol Psychiatry *32:*1, 1992.

Karege F, Bovier P, Rudolph W, Gaillard JM: Platelet phosphoinositide signaling system: An overstimulated pathway in depression. Biol Psychiatry *39:*697, 1996.

Kupfer DJ, Ehlers CL: Two roads to rapid eye movement latency. Arch Gen Psychiatry *46:*945, 1989.

Maes M, Meltzer HY: The serotonin hypothesis of major depression. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

*Manji HK, Potter WZ, Lenox RH: Signal transduction pathways. Arch Gen Psychiatry *52:*531, 1995.

Mayberg HS: Limbic-cortical dysregulation: A proposed model of depression. J Neuropsychiatry *9:*471, 1997.

*Mayberg HS, Liotti M, Brannan SK, McGinnis S, Mahurin RK, Jerabek PA, Silva JA, Tekell JL, Martin CC, Lancaster JL, Fox PT: Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. Am J Psychiatry *156:*675, 1999.

Nemeroff CB: New vistas in neuropeptide research in neuropsychiatry: Focus on corticotropin-releasing factor. Neuropsychopharmacology *6:*69, 1992.

Petty F: GABA and mood disorders: A brief review and hypothesis. J Affect Disord *34:*275, 1995.

Plotsky PM, Owens MJ, Nemeroff CB: Neuropeptide alterations in mood disorders. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

*Rajkowska G, Miguel-Hidaglgo JJ, Wei J, Dilley G, Pittman SD, Meltzer HY, Overholser JC, Roth BL, Stockmeier CA: Morphometric evidence for neuronal and glial prefrontal cell pathology in major depression. Biol Psychiatry *45:*1085, 1999.

Reiman EM, Lane RD, Ahern GL, Schwartz GE, Davidson RJ, Friston KJ, Yun L-S, Chen K: Neuroanatomical correlates of externally and internally generated human emotion. Am J Psychiatry *154:*918, 1997.

Renshaw PF, Lafer B, Babb SM, Fava M, Stoll AL, Christensen JD, Moore CM, Yurgelun-Todd DA, Bonello CM, Pillay SS, Rothschild AJ, Nierenberg AA, Rosenbaum JF, Cohen BM: Basal ganglia choline levels in depression and response to fluoxetine treatment: An in vivo proton magnetic resonance spectroscopy study. Biol Psychiatry *41:*837, 1997.

Ribeiro SCM, Tandon R, Grunhaus L, Greden JF: The DST as a predictor of outcome in depression: A meta-analysis. Am J Psychiatry *150:*1618, 1993.

*Sapolsky RM: Stress, glucocorticoids, and damage to the nervous system: The current state of confusion. Stress *1:*1, 1996.

Schatzberg AF, Schildkraut JJ: Recent studies on norepinephrine systems in mood disorders. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven Press, New York, 1995.

Servan-Schreiber D, Perlstein WM: Pharmacologic activation of limbic structures and neuroimaging studies of emotions. J Clin Psychiatry *58*(suppl):13, 1997.

Soares JC, Mann JJ: The anatomy of mood disorder—review of structural neuroimaging studies. Biol Psychiatry *41:*86, 1997.

*Swann AC, Katz MM, Bowden CL, Berman NG, Stokes PE: Psychomotor performance and monoamine function in bipolar and unipolar affective disorders. Biol Psychiatry *45:*979, 1999.

*Teasdale JD, Howard RJ, Cox SG, Ha Y, Brammer MJ, Williams SCR, Checkley SA: Functional MRI study of the cognitive generation of affect. Am J Psychiatry *156:*209, 1999.

Thase ME, Buysse DJ, Frank E, Cherry CR, Cornes CL, Mallinger AG, Kupfer DJ: Which depressed patients will respond to interpersonal psychotherapy? The role of abnormal electroencephalographic sleep profiles. Am J Psychiatry *154:*502, 1997.

Thase ME, Dubé S, Bowler K, Howland RH, Myers JE, Friedman E, Jarrett DB: Hypothalamic-pituitary-adrenocortical activity and response to cognitive behavior therapy in unmedicated, hospitalized depressed patients. Am J Psychiatry *153:*886, 1996.

*Thase ME, Howland RH: Biological processes in depression: An updated review and integration. In *Handbook of Depression*, EE Beckham, WR Leber, editors. Guilford, New York, 1995.

Weiss JM: Stress-induced depression: Critical neurochemical and electrophysiological changes. In *Neurobiology of Learning, Emotion, and Affect*, JIV Madden, editor. Raven Press, New York, 1991.

14.5 MOOD DISORDERS: PSYCHODYNAMIC ASPECTS

GLEN O. GABBARD, M.D.

A contemporary psychodynamic approach to mood disorders takes into account the strong biological underpinnings of these disorders. The domain of meanings is central to psychodynamic thought. Even when psychiatric illnesses can be traced to alternations in neurotransmitters, the symptomatology may acquire specific meanings for the patient. Moreover, psychosocial stressors and interpersonal events appear to trigger certain neurophysiological and neurochemical changes in the brain that significantly alter the balance of neurotransmitters. In other words, the development of an episode of depression or mania is generally regarded as a final common pathway of a complex mixture of psychosocial, genetic, and biological influences.

Clinically significant depression has long been linked to psychosocial stressors, such as the loss of a spouse. In addition, the loss of a parent before age 11 places adults at a higher than usual risk of depression. Some investigators have linked early childhood losses or separations to changes in neuronal receptor sites in the brain. This mechanism, known as kindling, is postulated to work in the following manner. Environmental stressors in childhood (e.g., loss or separation) so sensitize the receptor sites that relatively minor stressors in later life may suffice to trigger an episode of affective disorder. Chronic stress and deprivation of environmental origin may produce alterations in the catecholaminergic system in response to stimulation from the corticotrophin-releasing hormone–adrenocorticotropic hormone (ACTH) axis.

Primate studies have lent some support to the hypothesis that psychosocial influences have permanent effects on neurophysiological factors. Infant squirrel monkeys who are separated from their mothers experience long-lasting and, in some cases, permanent neurobiological changes. The changes include lasting alterations in the sensitivity of noradrenergic receptors, changes in hypothalamic serotonin secretion, and persistently elevated plasma cortisol concentrations. The sensitivity and number of brain opioid receptors are also significantly affected by repeated separation. Some changes are reversible if the infant monkeys are reunited with their mothers or siblings; other changes are not. Moreover, the separations appear to be more or less damaging during certain developmental periods, possibly because of the correlation with myelinization in the nervous system.

Although depression was once divided into endogenous and reactive subgroups, recent research has rendered these distinctions obsolete. Most depression is triggered by stressors, although at times relatively mild to an outside observer, and all depression involves endogenous biochemical factors. In this regard, every case of depression and mania can be regarded as having both a psychosocial and a biological dimension. Similarly, most depressions, except those of extreme severity, respond equally well to psychotherapy and medication.

One of the most sophisticated efforts to define the relative contributions of psychological vulnerability, genetics, and environmental stressors in major depressive disorder was a prediction study involving female twins. Multiple assessments of 680 female-female twin pairs of known zygosity were made over time, and the findings allowed the investigators to develop an etiological model to predict major depressive episodes. One of the most influential predictors was the presence of recent stressful events. Genetic factors were also important in predicting depression. Two other factors, neuroticism and interpersonal relations, also played substantial etiological roles. Neuroticism seemed to contribute in part by reducing the level of social support for an individual. Interpersonal dimensions of social support, recent difficulties, and parental warmth were all involved in predicting a major depressive episode.

Kenneth Kendler and his colleagues expanded on their twin study by examining information about stressful life events and the onset of major depressive episodes in a population base sample of 1082 female-female twin pairs. They interviewed members of the sample on two different occasions separated by a mean time of 17.3 months. They again found that stressful events significantly predicted onset of major depression in the month of occurrence. The most severe stressors included death of a close relative, serious marital problems, assault, and divorce or breakup. Genetic liability also had a significant impact on the risk of onset of depression. In individuals at lowest genetic risk (i.e., monozygotic twins with an unaffected cotwin), the probability of having a major depression in a particular month was 0.5 percent if unexposed to a severe life event. However, if the individual did experience such an event, the probability went up to 6.2 percent.

An even more dramatic difference was found by looking at individuals at the highest genetic risk (i.e., monozygotic twins with an affected cotwin). In these cases, the probability of having a major depression in a particular month was only 1.1 percent if no stressor was involved. However, if these individuals were exposed to a severe event, their risk of having a depressive episode skyrocketed to 14.6 percent.

The investigators described the best model for the joint effect of genetic liability and stressful life events on the onset of major depression as follows: genetic factors influence the risk of onset of major depression, at least in part, by altering the sensitivity of individuals to the depression-inducing effect of stressful life events. They emphasized that an individual's genetic endowment is not static. It interacts with the environment and reacts to psychosocial stressors.

One factor not captured by research using large samples is the role of the meaning of a particular stressor. What may seem a relatively mild psychological stressor may have conscious or unconscious meanings to an individual patient that greatly magnify its impact. In examining research on the role of stressful events in the course of unipolar and bipolar disorders, Constance Hammen concluded, “The field has reached considerable consensus that it is not the mere occurrence of a negative life event but rather the person's interpretation of the meaning of the event and its significance in the context of its occurrence.” In a longitudinal study of the link between depressive reactions and stressors, Hammen and her colleagues found that stressors whose content matched the patient's area of self-definition were particularly likely to precipitate depressive episodes. In other words, in someone whose sense of self is partly defined by social connectedness, loss of significant interpersonal contacts may precipitate a depressive episode. On the other hand, someone whose self-worth is linked to achievement and efficacy might be more likely to have a depressive episode in response to a perceived failure in the workplace.

Psychodynamic theories of etiology have accumulated over years of clinical experience. These psychodynamic theories must be viewed within a broad dialogue with the neuroscientific data regarding brain changes in affective disorders. The following theories may also be regarded as psychodynamic themes that emerge in the treatment of depressed patients and may assist the clinician in understanding the patient's inner experience.

PSYCHODYNAMIC THEORIES OF DEPRESSION

Anger Turned Inward A common finding in depressed patients is profound self-depreciation. Sigmund Freud, in his classic 1917 paper “Mourning and Melancholia,” attributed that self-reproach to anger turned inward, which he related to object loss, which may or may not be real. A fantasied loss may suffice to trigger a severe depression. Moreover, the patient may actually be unaware of any specific feelings of loss, since the fantasied loss may be entirely unconscious.

Freud drew an analogy between serious melancholic states and normal grief. Both may be time limited, but Freud cited two principal differences. In cases of grief, there is an actual object loss in external reality; in depression the lost object is more likely to be emotional than real. The second difference is that persons with depression experience profound loss of self-esteem, but the self-regard of persons engaged in a mourning process is not diminished.

The observational differences between grief and depression were pivotal in Freud's theory. He reasoned that one way of dealing with the loss of a beloved person is to become like the person. Freud defined that process as *introjection*, a defense mechanism central to the psychodynamics of depression, in which the patient internalizes the lost object so that it becomes an internal presence. Freud later noted that introjection is the only way that the ego can give up a valued and loved object.

Because depressed persons perceive the departed love object as having abandoned them, feelings of hatred and anger are intermingled with feelings of love. Freud

suggested that ambivalence involving the coexistence of love and hate is instrumental in the psychodynamics of depression. As a result of introjecting the lost object, the negative part of the depressed patient's ambivalence—the hatred and anger—is directed inward and results in the pathognomonic picture of self-reproach. In that manner a suicidal act may have the unconscious meaning of murder.

Karl Abraham, one of Freud's early colleagues, shared Freud's view of depression but also extended and elaborated it. Abraham viewed the process of introjection as a defense mechanism that takes two forms. First, he thought that the introjection of the original love object is the basis for building one's ego-ideal, so that the role of the conscience is eventually taken over by the introjected object. In that conceptualization much pathological self-criticism is seen as emanating from the introjected love object. In the second form of introjection, more in keeping with Freud's idea, the content of self-reproach is merciless criticism directed at the object. In other words, Abraham viewed the two processes of introjection as instrumental in creating the superego. Abraham also linked depression to early fixations at the anal and the oral levels of psychosexual development. He viewed oral sadistic tendencies as the primary source of self-punishment in depressed patients, and he inferred that inadequate mothering during the oral stage of development was involved.

The psychodynamic understanding of depression defined by Freud and expanded by Abraham is known as the classical view of depression. That theory involves four key points: (1) disturbances in the infant-mother relationship during the oral phase (the first 12 to 18 months of life) predispose to subsequent vulnerability to depression; (2) depression can be linked to real or imagined object loss; (3) introjection of the departed object is a defense mechanism invoked to deal with the distress connected with the object loss; and (4) because the lost object is regarded with a mixture of love and hate, feelings of anger are directed inward at the self.

Depressive Position Although Melanie Klein understood depression as involving the expression of aggression toward loved ones, much as Freud did, the developmental theory on which her view was based is quite different from Freudian theory. During the first year of life, Klein believed, the infant progresses from the paranoid-schizoid position to the depressive position. In the first few months of life, according to Klein, the infant projects highly destructive fantasies into its mother and then becomes terrified of the mother as a sadistic persecutor. That terrifying “bad” mother is kept separate from the loving, nurturing “good” mother through the defense mechanism of splitting. In that manner the infant's blissful feeding experience remains uncontaminated and undisturbed by persecutory fears of attack by the “bad” mother. In the course of normal development, according to Klein, the positive and negative images of the mother are integrated into a more ambivalent view. In other words, the infant recognizes that the “bad” mother it fears and hates is the same mother as the “good” mother it loves and adores. The recognition that one can hurt loved ones is the essence of the depressive position.

Klein connected clinical depression with an inability to successfully negotiate the depressive position of childhood. She regarded depressed persons as fixated or stuck at a developmental level in which they are extraordinarily concerned that loved good objects have been destroyed by the greed and destructiveness they have directed at them. In the absence of those good objects, depressed persons feel persecuted by the hated bad objects. In short, Klein's view was that depressed patients are longing or pining for the lost love objects while being persecuted by bad objects. In that theoretical framework the feelings of self-depreciation are linked to the fear that one's good parents have been transformed into violent persecutors as a result of one's own destructive tendencies. Also, the bad internal objects are internalized into the superego, which then makes sadistic demands on the patient. Hence, in the kleinian view, the self-reproaches experienced by depressed patients are directed against the self and internal impulses, rather than toward an introjected object, as in Freud's view.

Tension Between Ideals and Reality Whereas most psychodynamic theories of depression incorporate the superego as a significant part of the conceptual understanding, Edward Bibring viewed depression as tension arising from within the ego itself, rather than between the ego and the superego. According to Bibring, the ego has three highly invested narcissistic aspirations—to be good and loving, to be superior or strong, and to be loved and worthy. Those ideals are held up as standards of conduct. Depression sets in when a person becomes aware of the discrepancy between those ideals and reality. Helplessness and powerlessness result from the feeling that one cannot measure up to such high standards. Any blow to the self-esteem or any frustration of the strivings toward those aspirations precipitates depression. Bibring's theory, unlike Freud's and Klein's, does not regard aggression as playing a primary role in depression. The depressed person may ultimately experience anger turned inward, resulting from the awareness of helplessness; however, such expressions of aggression are secondary, rather than primary. The essence of depression, in Bibring's view, is a primary affective state arising within the ego and is based on the tension between what one would like to be and what one is.

Ego as Victim of Superego Edith Jacobson compared the state of depression to a situation in which the ego is a powerless, helpless child, victimized by the superego, which becomes the equivalent of a sadistic and powerful mother who takes delight in torturing the child. Like Freud, Jacobson assumed that depressed persons have identified with ambivalently regarded lost love objects. The self is experienced as identified with the negative aspects of the object, and ultimately the sadistic qualities of the lost love object are transformed into the cruel superego. Hence, depressed persons feel that they are at the mercy of a sadistic internal tormentor that is unrelenting in its victimization. Jacobson also noted that the boundary between self and object may disappear, resulting in a fusion of the bad self with the bad object.

Dominant Other Silvano Arieti studied the psychodynamic underpinnings of depression in severely ill patients who were unresponsive to most somatic treatments. He observed a common psychological theme in those patients that involved living for someone else, rather than for themselves. He referred to the person for whom depressed patients live as the dominant other. In most cases the dominant other is the spouse or a parent, but Arieti also noted that sometimes a principle, an ideal, or an organization serves a similar psychodynamic function. In such cases he referred to the entity as the dominant ideology or the dominant goal.

Depression often sets in when patients realize that the person for whom they have been living is never going to respond in a manner that will meet their expectations. The goal of their lives is regarded as unattainable, and a profound feeling of helplessness sets in. In Arieti's conceptualization of depression, he stressed a marked rigidity in the thinking of depressed persons, so that any alternative to living for the dominant other or the dominant ideology is viewed as unacceptable and even unthinkable. Depressed patients feel locked into an inflexible perspective on how they should live their lives and how gratification or fulfillment can be obtained. Even though they are depressed because living for someone or something other than themselves has been a failure, they nevertheless feel paralyzed and unable to shift their approach to life. If the dominant other will not respond to them in the way they have longed for, they feel that life is worthless, and that rigidity is often involved in a decision that suicide is the only alternative.

A 19-year-old college student consulted a psychiatrist after one semester in school. He told the psychiatrist that he was depressed and discouraged with college and with himself. College was not what he had expected, and he had not performed up to his expectations. He was seriously questioning whether he should return for the second semester, and he had a sense of hopelessness about changing his feelings. Suicidal thoughts had occasionally crossed his mind, although he was not planning to act on them. His sleep was disturbed by awakening in the middle of the night and ruminating about what he should do. He felt a significant diminution in his energy level, and he commented that things he used to find enjoyable no longer gave him pleasure.

The patient attended a prestigious college on the West Coast, but he indicated that he had actually wanted to get into Harvard. His application to Harvard had resulted in his being placed on the waiting list, but he had not been accepted. The psychiatrist he consulted commented that the college he had chosen to attend was certainly highly regarded. The patient responded, “It's not Harvard.” When the psychiatrist asked the patient how he had done academically during the first semester, the patient appeared embarrassed and replied, “I only got a 3.25 grade-point average—one A and three Bs.” The psychiatrist asked him why he seemed embarrassed to reveal such a solid academic record. The patient explained that he had wanted to make the dean's list but that he had fallen short of it, since the list required a 3.5 grade-point average.

The psychiatrist asked the patient if he hoped to be in a different situation after one semester of college. The patient's answer revealed that he had an extraordinarily high internal expectation of himself. He had wanted to be “a star,” a straight-A student at Harvard. He explained that his father had gone to Harvard, and he hoped that, by being a standout there, he would finally achieve the praise and recognition from his father that he had always longed for but had never received. His father seemed disappointed that his son had not been accepted at Harvard, and the patient was convinced that his father was ashamed of him for not making the dean's list.

The above case example illustrates the psychodynamic theories of both Arieti and Bibring. The patient was living his life for a dominant other—his father. He tried to perform beyond his abilities to extract an approving and loving response from his father that was never forthcoming. That longed-for response was rigidly construed as the only thing that mattered in life. Even though he was succeeding at a highly competitive college, his success did not make him feel good about himself. Moreover, the patient's depression can also be linked to his awareness of the disparity between his idealized expectations of himself and the reality of his situation, as described by Bibring. Being a straight-A student at Harvard was his aspiration; the reality was that he was a B + student at a college that did not measure up to Harvard.

The vignette also reflects two other key elements in the psychodynamic etiology of depression. First, in accord with the psychoanalytic notion of multiple causation,

more than one psychodynamic theory may be pertinent in understanding an individual patient's depression. Clearly, both the dominant other and the tension between ideals and realities were significant determinants in causing the patient's depression. Second, the precipitating factors that produce depression do not have to be catastrophic events involving obvious external disasters. To a casual observer the college student had no apparent reason to be depressed, since he was performing successfully at a highly regarded college. Nonetheless, the intrapsychic meaning of his academic performance was such that the patient felt hopeless and despairing as a result. In assessing the psychodynamic factors in depression, clinicians must always attend to idiosyncratic personal meanings of events to understand fully the effects they have on the patient. Otherwise, clinicians run the risk of responding in the same unempathic manner that often characterizes the responses of family members. In the absence of objective evidence of any disastrous events in the depressed person's life, loved ones often react by saying: “You have no reason to be depressed. Everything is going so well in your life.”

Selfobject Failure The ego and the superego do not figure in Heinz Kohut's conceptualization of depression. Kohut's theory, known as self psychology, rests on the assumption that the developing self has specific needs that must be met by parents to give the child a positive sense of self-esteem and self-cohesion and that similar responses are required from others throughout the course of the life cycle. He referred to those needs as mirroring, twinship, and idealization. The *mirroring* responses required by the self are equated with the gleam in the mother's eye when the child exhibitionistically shows off for her. Admiration, validation, and affirmation are responses that are included in the category of mirroring. *Twinship* responses refer to the child's need to be like others. A small boy who is outside playing with his toy lawn mower while his father is mowing the lawn is meeting important psychological needs in asserting his commonality with his father. Finally, the need for *idealization* is an important aspect of the development of the self. Children who grow up with parents they can respect and idealize develop healthy standards of conduct and morality.

Kohut referred to those three needs collectively as selfobject needs. In other words, the responses demanded from others are required by the self, and the needs of the object as a separate person are not taken into account. The other person serves as an object who meets the needs of the self. *Selfobject needs* essentially refer to certain functions that persons in the environment provide rather than to those persons themselves. Kohut felt that selfobject responses continue to be needed throughout life and are as necessary for emotional health as oxygen is for physical health. Within that conceptual framework, depression involves the failure of selfobjects in the environment to provide the self of the depressed person with mirroring, twinship, or idealizing responses necessary for the self to feel whole and sustained. The massive loss of self-esteem seen in depression is regarded by Kohut and the self psychologists as a serious disruption of the self-selfobject connection or bond.

Depression as Affect and Compromise Formation Some contemporary ego psychologists believe that depression is not truly a psychiatric disorder or illness. Instead, depression is regarded as an affect reflecting conflict and compromise formation. Charles Brenner, the principal architect of that view, suggested that concern about such childhood calamities as object loss, loss of love, castration, and punishment are associated with two kinds of unpleasure. One form of unpleasure is anxiety, which involves an anticipated calamity or danger. The other form of unpleasure, depressive affect, involves a calamity that has already happened. That theory of depressive affect differs sharply from the classical views of Freud and Abraham. Brenner pointed out that depression is not always related to object loss or to oral wishes. He also asserted that identification with a lost object is found in some depressed persons but not in all and that anger turned inward is a result of depression, rather than a cause. Depressive affect, in Brenner's view, can be linked to any of the childhood calamities, rather than uniquely to object loss. People can experience depressive affect because they feel unloved, because they feel castrated, or because they feel punished in a variety of ways. Depressive affect is a normal and universal part of the human condition.

A critical feature in Brenner's formulation is the idea of compromise formation, in which a symptom is viewed as simultaneously expressing an unconscious wish or drive and a defense against that wish or drive. A particular compromise formation may be more or less successful in eradicating depressive affect in the same manner as it may succeed to varying degrees in dealing with anxiety. A dog phobia, for example, is a symptomatic compromise formation that succeeds in eliminating anxiety as long as dogs are avoided. Similarly, certain forms of compromise formation may eradicate depressive affect while others do not.

The central point of Brenner's psychodynamic theory is that depressive affect is a universal feature in every pathological conflict, whether it is apparent on the surface or buried in the depths of the compromise formation. Depressive affect is a universal factor in all cases of psychiatric illness. From that standpoint, Brenner believed that classifying certain forms of mental illness as depression simply because depressive affect is part of the conscious symptoms does not make sense. The conscious experience of depression provides information about the efficacy and the nature of a patient's defensive maneuvers and compromise formations, in Brenner's view, but it does not reveal much about the underlying causes of the patient's illness.

Early Trauma and Deprivation Several investigators have noted that consistent, loving, nurturant parental involvement appears to have some value in preventing the development of depression. Conversely, separation from parents early in life or the actual loss of a parent may predispose one to depression. Edith Zetzel observed that adverse experiences in the formative years of childhood, particularly those involving separation and loss, make it difficult for children to tolerate depressive affects without resorting to primitive defensive operations. If caretakers fail to assist children in identifying and tolerating painful feelings that result from an adverse life experience, the child will grow up with inadequate coping mechanisms. That impaired adaptation may contribute to the subsequent development of depression.

Empirical research has provided some corroboration for the view that early deprivation is relevant to the cause of depression. René Spitz demonstrated that infants separated from their mothers during the second 6 months of life have overt signs of depression. In some cases the infants in Spitz's studies wasted away and died in response to the separation. Margaret Mahler and her colleagues, who studied the interactions between normal and abnormal mother-infant pairs, found that children's emotional dependence on their parents is instrumental in the development of their capacity to grieve and mourn. That capacity, in turn, influences children's feelings of self-esteem and helplessness. Although the development of depression may involve genetic and constitutional factors, as well as environmental stressors, most theorists agree that the early relationship between child and parent plays a significant role in causing depression.

One elegantly designed study documented an increased risk for major depressive disorder in those women who had experienced maternal or paternal separation in childhood or adolescence. A prospective study from the United Kingdom found that women with a history of childhood abuse or neglect are twice as likely to have negative relationships and low self-esteem in adulthood. Those abused or neglected women who have these negative relationships and low self-esteem in adulthood are then ten times more likely to experience depression. One of the clinical implications of this finding is that exploration of the impact of childhood trauma or neglect may be crucial in the psychodynamic therapy of depressed patients.

These empirical findings suggest a stress-diathesis model for mood disorders. In other words, a genetic substrate might serve to diminish monoamine levels in synapses or to increase reactivity of the hypothalamic-pituitary-adrenal axis to stress. Corticotropin-releasing factor (CRF), which induces the pituitary to secrete ACTH, is consistently elevated in the cerebrospinal fluid of depressed patients when compared with normal controls. When the brains of laboratory animals are given additional CRF, they have exhibited behavior similar to depression in humans. In keeping with the stress-diathesis model, some investigators have postulated that if there is no serious stress on the individual, the genetically determined threshold is not necessarily sufficient to induce depression. However, experiences of neglect or abuse in childhood may activate the stress response and induce elevated activity in CRF-containing neurons, which are known to be stress responsive and to be excessively active in depressed people. These cells can become supersensitive in certain individuals and then react dramatically to even mild stressors.

Premorbid Personality Factors A comprehensive psychodynamic understanding of depression must include premorbid personality factors in the equation. All persons may become depressed, given sufficient environmental stress, but certain personality types or traits appear to dispose one to depression. For example, the harsh, perfectionistic superego characteristic of persons with obsessive-compulsive personality disorder may lead them to feel that they are always falling short of their own excessive expectations of themselves. That intrapsychic constellation may be critical in the development of a major depressive episode. Similarly, Axis II personality disorders involving dependent yearnings for care—such as dependent, histrionic, and borderline personality disorders—may also be more vulnerable to depression. Personality disorders that use projection and other externalizing defense mechanisms, such as antisocial and paranoid personality disorders, are less likely to decompensate into depression. No particular premorbid personality type has been associated with the development of bipolar disorder.

Evidence is accumulating that an Axis II diagnosis of a personality disorder may complicate the course and treatment of depression. Depressed patients with personality disorders generally have poorer outcomes in the area of social functioning than those without personality disorders. Furthermore, residual depressive symptoms are more likely to present in recovering depressed patients with an Axis II diagnosis. Psychoanalytic clinicians have observed that personality factors frequently serve to maintain a depressed state once it has occurred. In clinical practice the complicating factors of a comorbid personality disorder diagnosis are quite common. One study found that 42 percent of persons with major depressive disorder and 51 percent of patients with dysthymic disorder have an accompanying Axis II diagnosis.

In an 18-year follow-up study of 89 depressed patients, the investigators found that the personality measure of neuroticism led to poorer outcomes in patients with melancholia. Similarly, an examination of the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Project data found that patients with personality disorders who were also depressed had poorer outcomes in social functioning and a much greater likelihood of residual depressive

symptoms than those who did not have personality disorders. In a third study 76 depressed outpatients were treated with interpersonal psychotherapy. The only significant predictor of time to remission was the degree of personality pathology. Even though patients with the most-severe personality disorders, borderline and antisocial, were excluded from the study, those who had other Axis II conditions responded more slowly to psychotherapy or not at all.

Characterological Depression Many patients encountered in clinical practice report feelings of depression even though they lack symptoms of a well-defined Axis I disorder, such as major depressive episode. Many of those patients have a primary diagnosis of a personality disorder on Axis II and experience characterological depression, a feeling of pervasive loneliness or emptiness associated with the perception that others are not meeting one's emotional needs. They can be distinguished from patients with an Axis I diagnosis of major depressive episode by the absence of vegetative symptoms (e.g., psychomotor retardation, loss of libido, diminished appetite, lack of energy, and sleep disturbance) and by the presence of certain qualitative features of their complaint of depression. Loneliness, emptiness, and boredom are often chronic complaints in characterological depression but are much less common in Axis I illnesses. In addition, a conscious sense of rage at not having their needs met may be present. The patients often describe childhood experiences in which they felt deprived of appropriate emotional nurturance from their patients. As a result, they continue to seek parental substitutes in adult life.

Characterological depression is differentiated from Axis II personality disorders by the fact that it is an affective state occurring within the context of certain personality disorders, rather than a constellation of traits forming an overarching personality type.

A 29-year-old woman came to psychotherapy complaining that she was “empty” inside and “needed to be filled up” by a positive experience with a psychotherapist. She said that, while she was growing up, her mother never had time for her and that her mother loved her two sisters more than her. The patient had had a series of romantic relationships with men, but she never felt that she was getting the kind of attention and love that she needed from any of them. The men often ended the relationship because they felt that she was too demanding and that they could not possibly meet all her needs. Her last therapist had “given up” on her because he, too, felt that he was unable to be of help to her. The patient also indicated that she had called her previous therapist almost every night because she would begin to feel lonely and need his reassurance that he still cared. She feared that she had turned off her therapist by being too demanding. She also described several angry outbursts directed at him when he would not talk with her for lengthy periods of time on the phone during the evening. She wondered if her outbursts made him hate her.

The patient had taken four different antidepressive medications with no improvement. She did not meet the diagnostic criteria for an Axis I dysthymic disorder or major depressive episode. However, she did have characteristics in keeping with two different Axis II diagnoses dependent personality disorder and borderline personality disorder.

Other Clinical Entities In addition to the existence of characterological depression and the presence of other Axis II diagnoses, another clinical entity gaining increased acceptance is depressive personality disorder. In Appendix B of the fourth edition of *Diagnostic and Statistical Manuel of Mental Disorders* (DSM-IV) the criteria for the disorder emphasize a constellation of personality traits, while the criteria for dysthymia focus on somatic symptoms. These traits include a mood dominated by unhappiness, dejection, and gloominess; a self-concept centered on worthlessness and low self-esteem; a proneness to feel guilt or remorse; a tendency to blame and criticize oneself; a negativistic and judgmental stance toward others; a pessimistic attitude; and a tendency to worry and to brood. Although this diagnostic entity has a long-standing psychoanalytic tradition, there has been controversy regarding whether or not it is truly distinct from dysthymic disorder. Data are emerging suggesting that the distinction between the two is valid and clinically useful. In a study of 54 patients with early onset long-standing mild depressive features, Katherine Phillips and her collaborators identified 30 subjects with and 24 without depressive personality disorder. Sixty-three percent of the subjects with depressive personality disorder did not have dysthymic disorder, while 60 percent did not have current major depressive disorder. The depressive personality disorder patients were more likely than the comparison group to have another personality disorder, but 40 percent of them had no such disorder. Those who were comorbid for another personality disorder tended to have anxious or Cluster C personality disorder, suggesting that defenses and conflicts at a neurotic level were most prominent, in keeping with the psychoanalytic conceptualization of depressive personality disorder. Finally, the duration of psychotherapy was significantly longer for subjects who had depressive personality disorder when compared to those who did not. Antidepressant drugs may be combined with psychotherapy in the treatment of characterological depression.

Clinicians must remember that depression spans the entire spectrum of pathology and health. In addition to being a discrete psychiatric disorder, depression refers to an emotional state that can occur in normal persons at certain times as well as in persons with characterological or psychotic conditions. Moreover, simply because the patient does not have sufficient symptoms to be given an Axis I diagnosis of a mood disorder does not mean that the depression is benign. In one study, employees with minor forms of depression that did not meet Axis I criteria had 51 percent more disability days than did persons with a diagnosis of major depressive episode.

PSYCHODYNAMIC PSYCHOTHERAPY

Although brief psychodynamic psychotherapy has been shown in meta-analyses to be no better or worse than other forms of psychotherapy, there are few controlled trials of the modality with depressed patients. Even more problematic is the fact that few of the trials that exist have been carried out by advocates of the technique who actively practice it. In most studies brief dynamic therapy was used on a comparison or control group to contrast with the different psychotherapies that the investigators favored. Most studies testing brief dynamic therapy were also done in a group format, even though individual therapy is much more widely practiced by psychodynamic clinicians. A number of the trials involving brief dynamic therapy, in which the investigators favored other therapies, found it significantly less effective than the other interventions.

More recently, well-designed studies show promising results from brief dynamic therapy. In an investigation of depressed caregivers of elderly family members, random assignment was made to one of two treatment cells: brief psychodynamic therapy or cognitive-behavioral therapy. After 20 sessions, 71 percent of the caregivers were no longer clinically depressed. Overall, no differences were found between the two treatment groups. Symptom-oriented measures suggested that those who had been caregivers for more than 3.5 years improved more with cognitive-behavioral therapy, whereas those who had been caregivers for shorter periods showed more improvement from brief psychodynamic therapy.

In the British study known as the second Sheffield Psychotherapy Project, similar findings were obtained in a randomized, controlled trial. Some 120 depressed patients were assigned to either 8 or 16 sessions of cognitive-behavioral therapy or psychodynamic-interpersonal therapy. Both treatments were found equally effective, and their effects equally rapid. Patients with only mild or moderate depression had the same outcome whether they were treated with 8 or 16 weeks of therapy. However, significantly better outcomes were noted in the severely depressed group when 16 weeks of therapy were provided, whether cognitive-behavioral or psychodynamic-interpersonal. At 1-year follow-up no overall differences were found in outcome or maintenance of gains between those receiving the two types of therapy. However, longer periods of therapy appeared to be associated with better long-term outcomes, particularly in the case of psychodynamic-interpersonal therapy.

Systematic controlled studies using randomized assignment are not available for extended psychodynamic psychotherapy and psychoanalysis. Most therapists would agree that these modalities are not indicated as exclusive treatment for patients in the acute phase of major depressive disorder. However, a subgroup of patients who do not respond to medication or brief psychotherapy may benefit from more-extended psychodynamic exploration. Sidney Blatt and his associates reanalyzed the data from the NIMH Treatment of Depression Collaborative Research Project and found that highly perfectionistic and self-critical patients did not respond well to any of the treatment cells, which included 16 weeks of cognitive therapy, 16 weeks of interpersonal therapy, 16 weeks of imipramine (Tofranil) plus clinical management, and 16 weeks of placebo plus clinical management. In two other naturalistic follow-along studies of extended psychotherapy on such patients, self-critical and perfectionistic traits appeared to respond well to intensive psychoanalytic psychotherapy. Hence a subgroup of depressed patients may be particularly well suited to more-extended psychoanalytic therapy. Depressed patients with significant Axis II pathology may also respond poorly to brief therapies or medication and require intensive psychotherapy over a long time. However, no data from controlled trials are available to confirm this clinical impression.

In any depressed patient, certain psychodynamic principles may be useful, no matter which treatment modality is selected. For example, the American Psychiatric Association practice guidelines for depression suggest that psychotherapeutic management should be part of every treatment for depression. Psychodynamic concepts such as therapeutic alliance, transference, resistance, and countertransference apply to all patients, regardless of the type of treatment. The role of the therapeutic alliance may be particularly important in terms of outcome in both psychotherapy and pharmacotherapy. A team of investigators led by Janice Krupnick looked at 225 cases of depressed patients in the NIMH Treatment of Depression Collaborative Research Program. Clinical raters scored videotapes made of the treatments in all four cells. When outcomes were assessed on these cases, the therapeutic alliance was found to significantly affect clinical outcomes of both psychotherapies, placebo treatment, and the imipramine group. In fact, the patient contribution to the therapeutic alliance accounted for 21 percent of the outcome variance on standardized outcome measures, with more of the variance in outcome attributed to alliance than to the treatment method itself. Among the four cells,

none showed significant group differences in terms of the relationship between therapeutic alliance and outcome. Hence attention to the psychotherapeutic relationship is of central importance in all treatment of depression.

Sydney Blatt has suggested that the psychodynamic psychotherapeutic approach must be tailored to the underlying type of depression. His research has delineated two distinct categories of depressed patients: anaclitic and introjective. *Anaclitic depression* is characterized by feelings of helplessness, loneliness, and weakness related to chronic fears of being abandoned and unprotected. These individuals have longings to be nurtured, protected, and loved. They have a vulnerability to having interpersonal relationships disrupted. These types of patients require a therapy that emphasizes the therapeutic relationship rather than interpretation or insight.

By contrast, *introjective depression* is characterized by feelings of unworthiness, failure, guilt, and inferiority. These individuals are also highly self-critical and suffer from a chronic fear of criticism and disapproval from others. They are highly perfectionistic and competitive and are often driven to achieve in work and school. They have a characteristic vulnerability to disruptions of a positive and effective sense of self, and their depression is manifested primarily by dysphoric feelings of guilt, worthlessness, failure, and by a sense that one's autonomy and control have been lost. Interpretive interventions that provide insight seem to be more helpful with this category of depressed patients.

MANIA

Almost all researchers who have studied bipolar disorders have concluded that psychosocial interventions may be crucial in concert with pharmacotherapy for the prevention of relapse. Relapse rates as high as 60 percent over a period of 2 years have been reported, even when maintenance therapy on lithium carbonate (Eskalith) has achieved adequate plasma concentrations. Work status deteriorates in about two-thirds of patients, while social function deteriorates in one-half of patients. In addition, there is a 45 percent separation or divorce rate, compared with 18 percent in controls. In fact, the average 25-year-old female patient can expect to lose nearly 9 years of life, 14 years of effective major activity (work or school), and 12 years of normal health.

A psychodynamic understanding can be extraordinarily helpful in identifying the nature of stressors likely to precipitate episodes. In one 2-year study of 61 outpatients, patients with the highest levels of stress had a 4.53 times greater risk of relapse than patients without stress. In a 10-year follow-up study, two different groups of treatment failures were identified in a cohort of patients with bipolar disorder. One group of patients relapsed because the treating psychiatrist had failed to increase the lithium dosage in response to increased physiological activation before the onset of a manic episode. In the other group of treatment failures, psychological issues that were clearly involved in precipitating manic episodes had not been given appropriate attention by the responsible psychiatrists, and manic episodes had resulted from the stress of those psychosocial factors.

In one study investigators examined both individual differences in stress reactivity and psychological characteristics of a group of 58 patients with bipolar disorders followed for at least 1 year. The stress level predicted relapse, but personality variables, particularly introversion and obsessionality, were also significant predictors. Hence a psychodynamic perspective on the bipolar disorder patient may help both in identifying psychological and environmental stressors and in dealing with characterological issues that may complicate the treatment. There is an increasing recognition that with refined diagnostic methods, a high degree of comorbidity may exist between bipolar disorders and Axis II disorders. In a study of 66 outpatients with a Research Diagnostic Criteria diagnosis of hypomania and a lifetime diagnosis of bipolar disorder, each patient was interviewed during the hypomanic episode and after recovery. Knowledgeable informants were also interviewed at both times. The researchers found that hypomania was associated with higher levels of maladaptive personality traits than the euthymic state, but even when euthymic, 50 percent of the cohort had at least one personality disorder. They also found that the alleged decrease in maladaptive personality traits following recovery was much greater in the reports by patients than in those by informants.

Some of these characterological features of the personality disorder may contribute to compliance problems with medications designed to stabilize mood. They also may alienate family members who then become more critical and more distant, producing an increased risk of relapse. Often these character traits may require individual dynamic psychotherapy in which they can be addressed in the transference-countertransference dimensions of the treatment.

Even in bipolar disorder patients without a personality disorder, compliance is a major problem. At least half of bipolar disorder patients stop taking medication at least once in the course of their illness. Frederick Goodwin and Kay Jamison found that the most important reasons for noncompliance listed by patients were dislike of having medication control their moods, dislike of the idea of having a chronic illness, aversion to feeling depressed, and dislike of adverse effects. The combination of medication and individual psychotherapy is required for comprehensive treatment and optimal compliance.

Patients with bipolar disorder have been studied from the perspective of ongoing psychoanalysis and psychoanalytic psychotherapy, and those clinical investigations have revealed specific psychodynamic factors at work in the onset of manic episodes. In one series of patients, unconscious sexual urges and fantasies seemed to overpower ego defense mechanisms, leading to a clinical picture of hypersexuality and other symptoms of mania. Increasing the lithium dosage decreased the sexual behavior and reinstituted the ego defense mechanisms present before the manic episode. In the course of continued psychotherapeutic or psychoanalytic treatment, the patients became consciously aware of their unconscious sexual desires and of the defenses elicited to deal with those desires. This conscious awareness enabled the patients to identify early warning signals of increased sexual impulses, so that future manic episodes could be avoided by increasing their lithium dosage.

These studies reflect how a psychodynamic understanding of patients with bipolar disorder may be crucial to the effective treatment of the disorder. Most manic patients cannot make use of psychotherapy interventions in the midst of a full-blown manic episode because the essence of mania is a denial of psychological problems. However, after the patient has become euthymic as a result of pharmacological stabilization, psychotherapeutic interventions may have value both in preventing subsequent episodes and in dealing with the feelings of shame and guilt associated with embarrassing behavior that took place during the manic episode.

Psychodynamic Theories of Mania The psychodynamic understanding of mania is usefully applied to clinical instances of hypomania because the differences between the two entities are quantitative, rather than qualitative. Just as mania and depression have been linked from a neurophysiological standpoint, they are similarly connected from a psychodynamic perspective.

Karl Abraham Most theories of mania view manic episodes as defensive against underlying depression. Abraham, for example, believed that manic episodes may reflect an inability to tolerate childhood depression in reaction to a developmental tragedy, such as the loss of a parent. The manic state, in Abraham's view, is understood as a way of removing the shackles of a tyrannical superego by merging the ego and the superego. Self-criticism is then replaced by euphoric self-satisfaction.

Bertram Lewin Lewin regarded the hypomanic patient's ego as a purified pleasure ego. The defense mechanism of denial is appropriated by the ego to disregard unpleasant perceptions and affects as well as distressing psychic realities that may result in self-punishment or self-criticism.

Melanie Klein Klein also viewed mania and hypomania as defensive reactions to depression, but she linked the mechanism to the depressive position, rather than to an overriding of the superego. The essence of the depressive position is intense anxiety that one's own aggression has resulted in the destruction of important love objects, such as parents. In Klein's own words, “Persecution (by ‘bad' objects) and the characteristic defenses against it, on the one hand, and pining for the loved (‘good') object, on the other, constitute the depressive position.” She thought that manic defenses are necessary both to control and to master the dangerous bad objects and to restore and to save the loved good objects.

Manic defenses include omnipotence, denial, idealization, and contempt. Omnipotence serves to deny the need for good objects, to delude oneself into feelings of self-containment and grandiosity, and to help one feel insulated and protected from assault by internal persecutors. Idealization and denial work together so that idealization of self and others serves to deny any destructiveness or aggression in relationships. The euphoric disposition of the manic or hypomanic patient reflects the tendency to gloss over any unpleasant aspects of reality and to treat everything with a sense of humor and a striking disregard for the tragic dimensions of reality, even if the situation is tragic. Idealization, however, may rapidly give way to contempt, which is also linked to denial because it is a way of disregarding the importance of love objects and, therefore, denying the concern that damage has been done to them and reparation is needed. Moreover, the manic patient can then minimize any distressing feelings of sorrow or regret that may arise in connection with concerns about having destroyed love objects.

Klein also observed that a wish to triumph over parents is often an integral part of the manic defensive posture. She noted that a frequent childhood fantasy is to reverse the child-parent relationship and that the fantasy produces feelings of guilt and anxieties of a depressive nature related to the wish to destroy and replace the parents. Feelings of depression may develop after a job promotion or other professional success because the person's unconscious wish to triumph over and to

surpass one's parents has been fulfilled.

The kleinian conceptualization of mania as defensive against feelings of depression is useful in understanding the phenomenon of dysphoria in manic patients when depression breaks through a manic episode, requiring a resurgence of manic denial. This formulation is also useful in understanding the commonly observed phenomenon of elation after the death of a loved one.

> A patient received a phone call that informed him of his mother's death. Rather than feeling grief-stricken or shocked, be noted a sense of expansiveness and power. As he discussed the odd reaction with his psychotherapist, he was able to recognize that the high feeling he experienced was related to a sense that he was finally liberated from feelings of slavish dependence on a tyrannical mother.

Other Theories Other views of mania include Bibring's conceptualization that manic elation is essentially a compensatory reaction secondary to severe depression or an unconscious fulfillment of a person's narcissistic aspirations to be loved, worthy, superior, and virtually flawless. Jacobson understood mania as a transformation of the sadistic superego figure from a punitive tormentor into a loving and forgiving object who is thoroughly idealized. This dramatically altered superego is then projected into persons in the outside world with whom the manic patient establishes idealized relationships that are free from any negative characteristics, such as hatred and anger.

OTHER PSYCHOLOGICAL THEORIES

Adolf Meyer Meyer viewed depression as a person's reaction to a distressing life experience, such as a financial setback, the loss of a job, the death of a loved one, or a serious physical illness. He believed that depression must always be understood in the context of the patient's life history, as an event with psychic causality.

Karen Horney Horney believed that children raised by rejecting and unloving parents are prone to feelings of insecurity and loneliness. In her view, children need to be loved but fear criticism and rejection, which make them susceptible to feelings of depression and helplessness.

Sandor Rado Rado linked depression to a profound feeling of helplessness. He believed that anhedonia, the inability to experience pleasure, is a central phenomenon in depression that develops when persons are not aware of their capacities or cannot provide feelings of emotional self-gratification. Rado connected severe depression with a punitive superego that punishes the patient for unconscious hostility toward a deceased loved one.

John Bowlby Bowlby saw depression from an ethological perspective that emphasized disturbances of the mother-infant attachment bond. He believed that separation of infants from mothers (or other caregivers) early in life leads to feelings of depression and hopelessness that may in some cases continue throughout the life cycle.

Harry Stack Sullivan Although Sullivan concentrated his efforts more on schizophrenia than on mood disorders, his interpersonal perspective applies to both. He thought that adverse interactions between persons and their psychosocial environments were critical to the development of depression.

Cognitive-Behavioral Theory According to the theory developed by Aaron Beck, depression results from specific cognitive distortions present in persons prone to depression. Those distortions, referred to as *depressogenic schemata*, are cognitive templates that perceive both internal and external data in ways that are altered by early experiences. These schemata are associated with four systematic errors in logic: overgeneralization, magnification of negative events with a simultaneous minimization of positive events, arbitrary inference, and selective abstraction.

Learned Helplessness The learned helplessness theory of depression connects depressive phenomena to the experience of uncontrollable events. For example, when dogs in a laboratory were exposed to electrical shocks from which they could not escape, they showed behaviors that differentiated them from dogs who had not been exposed to such uncontrollable events. After exposure to the shocks, they would not cross a barrier to stop the flow of electric shock when put in a new learning situation. According to the learned helplessness theory, the dogs learned that outcomes were independent of responses, so they had both cognitive motivational deficit (i.e., they would not attempt to escape the shock) and emotional deficit (indicating a decreased reactivity to the shock). In the reformulated view of learned helplessness as applied to human depression, internal causal explanations are thought to produce a loss of self-esteem after adverse external events. Behaviorists who subscribe to the theory stress that improvement of depression is contingent on the patient's learning a sense of control and mastery of the environment.

Psychodynamic Treatment Clinicians who combine pharmacotherapy with a mood stabilizer and individual dynamic therapy focus to a large extent on preventing relapse through greater medication compliance, through detailed understanding of problematic stressors, and through careful examination of relationship difficulties, some of which may relate to Axis II psychopathology. Several major psychodynamic themes are often present in bipolar patients, all of which may be relevant at one time or another in the ongoing treatment of the condition. Denial of illness is one of the major defensive postures encountered with these patients. Many argue that their manic or hypomanic symptoms are not part of an illness but rather a reflection of who they really are. Hence when their collaboration is enlisted in a medication regimen or other component of the treatment plan, they frequently deny having a problem that requires treatment. Patients with bipolar illness are notoriously lacking in insight. In one study of 28 manic patients treated on an inpatient unit, insight was measured at admission and discharge. Even when all other symptoms of mania had improved or remitted, insight remained notably absent. The investigators concluded that poor insight is a prominent characteristic of a bipolar disorder regardless of illness phase.

Often related to denial is another psychodynamic theme—splitting or psychic discontinuity. Many bipolar I disorder patients continue to deny the significance of their prior manic episodes when they are euthymic. They may claim that the behavior was simply a result of being exhausted and not taking care of themselves. They also assert that nothing like that will ever happen again. This involves a form of splitting the self-representation so that the self involved in a manic episode is considered entirely disconnected from the self in the euthymic phase. The discontinuity between the different versions of the self may be regarded with bland denial or indifference.

> A 35-year-old dentist was admitted to an inpatient unit with pressured speech, flight of ideas, and extreme hyperactivity. He had been engaging in extensive sexual activity in the days preceding the admission, and he had been spending all of his savings on a variety of outlandish investment schemes. When he finally arrived on the hospital unit, he was singing and dancing around the lounge area to such an extent that he had to be placed in the seclusion room to try to settle him down long enough to conduct a psychiatric interview. The female psychiatric resident who was on call entered the seclusion room with her attending psychiatrist. The patient abruptly stopped his singing and said, “I have a plan.” He then pointed to the female resident and said, “I'll screw her, and you watch.” The patient was told that his plan was not tenable and that he needed to take medication. With considerable coaxing, he finally agreed to take lithium and an antipsychotic agent.
>
> Seven days later his manic episode had subsided. He was calm, polite, and generally socially appropriate. A teaching conference for residents was held in which the patient was interviewed. The attending psychiatrist asked him about his reasons for hospitalization and the nature of his problems. He went on at some length about how he was experiencing “burnout” on his job because people were always telling dentists their problems. He continued to say that he had gotten rather tired and found his job less than fulfilling. He also said that his marriage was troubled because of his long work hours. The impression he created was that of a hardworking professional who was suffering mild stress as a result of occupational difficulties. At no time did he mention anything to the group of residents attending the conference that reflected his bizarre behavior associated with his recent manic episode.
>
> The attending psychiatrist interviewing the patient finally interrupted him and said that he wondered why he was not bringing up any of the behaviors that had brought him into the hospital 7 nights ago, such as singing and dancing, spending vast sums of money on unwise investments, staying up all night, and engaging in extensive sexual activity. The patient stared at the interviewer blandly and said, “Oh, that? Well, that's not a problem now.”
>
> This vignette clearly illustrates how the manic self and the euthymic self are maintained in separate compartments so that no continuity of the self is apparent.

This form of compartmentalization is a major contributor to compliance difficulties. If patients can maintain this kind of discontinuity, they can deny the need for medication or even vulnerability to a subsequent episode. The clinician needs to assemble the self-fragments into a continuous narrative of the patient's life so that maintenance treatment becomes more compelling. Tape recording manic episodes (with the patients' permission) and then playing the recordings back to patients

when they are in a euthymic state may be a useful technique.

Another reason for this form of splitting is a wish to avoid the work of mourning and grief that becomes necessary when the manic self and the euthymic self are connected. There may be mourning of the healthy premorbid self, which seems to be forever gone. Also, following a manic episode, patients may become acutely aware of their own destructiveness and feel intense remorse about the harm they have caused others during the manic phase. Splitting it off and denying its significance spares them the pain of this awareness. If the therapist can gently confront them with this defensive maneuver, they can then acknowledge the harm they have done and attempt to make reparation.

Another relevant psychodynamic theme is ego-syntonic attachment to hypomania. Many patients insist that they are more creative, more full of life, and more engaging when they are in a hypomanic or manic episode. They do not want to comply with their mood-stabilizer medication because they feel it interferes with their creativity and their enjoyment of life. Helping them to see that their behaviors clearly had a negative impact on others may help them deal with the attachment and make it more ego dystonic.

Lithium or other mood stabilizers may take on special meanings to bipolar patients. These agents may come to represent something that is taken away from them. If they comply, they have to resign themselves to no longer experiencing the euphoria of their manic periods, and their denial of their illness is challenged. Maintenance medication may have the meaning of having a severe mental illness that requires lifelong treatment. Integrating this notion into their sense of self also requires the work of mourning. Clinicians must always keep in mind that there is a psychological function of mania that runs in parallel with the biological alteration of the brain—namely, manic denial serves to defend against depression and loss. Medication may also have the meaning of identification with specific family members who have had bipolar illness. To take the medication may unconsciously mean that the course of their illness will be the same as their relative's.

The cornerstone of the psychotherapeutic strategy with bipolar disorder patients is to build a therapeutic alliance. Arguing with the patient about whether or not a bipolar disorder is a correct diagnosis is of little value. Psychotherapeutic exploration, empathy, and education are generally more effective. Creation of a mood chart that helps the patient track highs and lows may also be helpful. Transference tends to shift from idealization to devaluation, often in parallel with shifts in mood. The therapist must also be wary of countertransference acting out in response to anger and frustration with the patient's refusal to cooperate with the treatment plan.

Most clinicians now treating bipolar disorders consider the combination of psychotherapy and pharmacotherapy essential.

Jamison, who wrote a vivid personal account of her struggles with bipolar illness, described it as follows:

Ineffably, psychotherapy heals. It makes some sense of the confusion, reins in the terrifying thoughts and feelings, returns some control and hope and possibility of learning from it all.... No pill can help me deal with the problem of not wanting to take pills; likewise, no amount of psychotherapy alone can prevent my manias and depressions. I need both.

SUGGESTED CROSS-REFERENCES

Further discussion of psychoanalytic theory can be found in Section 6.1. For additional material on characterological depression, see the discussion of borderline personality disorder in Chapter 24.

SECTION REFERENCES

Abraham K: Notes on the psycho-analytical investigation and treatment of manic-depressive insanity and allied conditions. In *Selected Papers of Karl Abraham, M.D.* Basic Books, New York, 1953.

American Psychiatric Association: Practice guidelines for major depressive disorder in adults. Am J Psychiatry *150*(Suppl):126, 1993.

*Barkham M, Shapiro DA, Hardy GE, Rees A: Psychotherapy in two-plus-one sessions: Outcomes of a randomized controlled trial of cognitive-behavioral and psychodynamic-interpersonal therapy for subsyndromal depression. J Consult Clin Psychol *67:*201, 1999.

Bearden C, Lavelle N, Buysse D, Karp JF, Frank E: Personality pathology and time to remission in depressed outpatients treated with interpersonal psychotherapy. J Pers Disord *10:*164, 1996.

*Bibring E: The mechanism of depression. In *Affective Disorders: Psychoanalytic Contributions to Their Study.* P. Greenacre, editor. International Universities Press, New York, 1953.

Bifulco A, Brown GW, Moran P, Ball C, Campbell C: Predicting depression in women: The role of past and present vulnerability. Psychol Med *28:*39, 1998.

*Blatt SJ: Contributions of psychoanalysis to the understanding and treatment of depression. J Am Psychoanal Assoc *46:*723, 1998.

Blatt SJ, Quinlan DM, Pilkonis PA, Shea MT: Impact of perfectionism and need for approval on the brief treatment of depression: the National Institute of Mental Health Treatment of Depression Collaborative Research Program revisited. J Consult Clin Psychol *63:*125, 1995.

Broadhead WE, Blazer DG, George LK, Tse CK: Depression, disability days, and days off from work in a prospective epidemiologic survey. JAMA *264:*2524, 1990.

Coe CL, Wiener SG, Rosenberg LT, Levine S: Endocrine and immune responses to separation and maternal loss in nonhuman primates. In *The Psychobiology of Attachment and Separation*, M Reite, T Field, editors. Academic Press, Orlando, FL, 1985.

Depression Guideline Panel: *Depression in Primary Care*, vol 2, *Treatment of Major Depression.* U.S. Department of Health and Human Services, Rockville, MD, 1993.

Duggan CF, Lee AS, Murray RM: Do different subtypes of hospitalized depressives have different long-term outcomes? Arch Gen Psychiatry *48:*308, 1991.

Ellicott A, Hammen C, Gitlin M, Brown G, Jamison K: Life events in the course of bipolar disorder. Am J Psychiatry *147:*1194, 1990.

Feinstein SC, Wolpert EA: Juvenile manic-depressive illness: Clinical and therapeutic considerations. J Am Acad Child Psychiatry *12:*123, 1973.

Freud S: The ego and the id. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 19. Hogarth Press, London, 1961.

*Freud S: Mourning and melancholia. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 14. Hogarth Press, London, 1963.

Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition.* American Psychiatric Press, Washington, DC, 1994.

Gabbard GO: Psychodynamic psychiatry in the decade of the brain. Am J Psychiatry *149:*991, 1992.

*Gallagher-Thompson D, Steffen AM: Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. J Consult Clin Psychol *62:*543, 1994.

Ghaemi SN, Stoll AL, Pope HG: Lack of insight in bipolar disorder: The acute manic episode. J Nerv Ment Dis *183:*464, 1995.

Gold PW, Goodwin FK, Chrousos GP: Clinical and biochemical manifestations of depression: Relation to the neurobiology of stress, part I. N Engl J Med *319:*348, 1988.

Gold PW, Goodwin FK, Chrousos GP: Clinical and biochemical manifestations of depression: Relation to the neurobiology of stress, part II. N Engl J Med *319:*413, 1988.

*Goldberg JF, Harrow M, editors: *Bipolar Disorders: Clinical Course and Outcome.* American Psychiatric Press, Washington, DC, 1999.

Hammen C, Marks T, Mayol A, DeMayo R: Depressive self-schemas, life stress, and vulnerability to depression. J Abnorm Psychol *94:*308, 1985.

Hammen CL: Stress and the courses of unipolar and bipolar disorders. In *Does Stress Cause Psychiatric Illness?* CM Mazure, editor. American Psychiatric Press, Washington, DC, 1995.

Jacobson E: Psychotic identifications. In *Depression: Comparative Studies of Normal, Neurotic, and Psychotic Conditions.* International Universities Press, New York, 1971.

Jacobson E: Transference problems in depressives. In *Depression: Comparative Studies of Normal, Neurotic, and Psychotic Conditions*. International Universities Press, New York, 1971.

Jamison KR: *An Unquiet Mind: A Memoir of Moods and Madness.* Vintage Books, New York, 1995.

Kendler KS, Kessler RC, Neale MC, Health AC, Eaves LJ: The prediction of major depression in women: Toward an integrated etiological model. Am J Psychiatry *150:*1139, 1993.

Kendler KS, Kessler RC, Walters EE, MacLean C, Neale MC, Heath AC, Eaves LJ: Stressful life events, genetic liability, and onset of an episode of major depression in women. Am J Psychiatry *152:*833, 1995.

Klein M: Mourning and its relation to manic-depressive states. In *Love, Guilt, and Reparation and Other Works 1921–1945.* Free Press, New York, 1975.

Klerman GL, Weissman MM, editors: *New Applications of Interpersonal Therapy.* American Psychiatric Press, Washington, DC, 1993.

Kohut H: *The Analysis of the Self: A Systematic Approach to the Psychoanalytic Approach of Narcissistic Personality Disorders.* International Universities Press, New York, 1971.

Kohut H: *How Does Analysis Cure?* A Goldberg, editor. University of Chicago Press, Chicago, 1984.

Krupnick JL, Sotsky SM, Simmens S, Moyer J, Elkin I, Watkins J, Pilkonis PA: The role of the therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol *64:*532, 1996.

Lewin BD: *The Psychoanalysis of Elation.* Norton, New York, 1950.

Loeb FF, Loeb LR: Psychoanalytic observations on the effect of lithium on manic attacks. J Am Psychoanal Assoc *35:*877, 1987.

Nemeroff CB: The neurobiology of depression. Sci Am *278:*42, 1998.

Phillips KA, Gunderson JG, Hirschfeld RMA, Smith LE: A review of the depressive personality. Am J Psychiatry *147:*830, 1990.

Phillips KA, Gunderson JG, Triebwasser J, Kimble CR, Faedda G, Lyoo IK, Renn J: Reliability and validity of depressive personality disorder. Am J Psychiatry *155:*1044, 1998.

*Salzman C: Integrating pharmacotherapy and psychotherapy in the treatment of a bipolar patient. Am J Psychiatry *155:*686, 1998.

Shapiro DA, Rees A, Barkham M, Hardy G: Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. Society for Psychotherapy Research (1994, York, England). J Consult Clin Psychol *63:*378, 1995.

Shea MT, Pilkonis PA, Beckham E, Collins JF, Elkin I, Sotsky SM, Docherty JP: Personality disorders and treatment outcome in the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry *147:*711, 1990.

Slavney PR: The mind-brain problem, epistemology, and psychiatric education. Acad Psychiatry *17:*59, 1993.

Solomon DA, Keitner GI, Miller IW, Shea MT, Keller MB: Course of illness and maintenance treatments for patients with bipolar disorder. J Clin Psychiatry *56:*5, 1995.

Suomi SJ: The development of affect in rhesus monkeys. In *The Psychobiology of Affective Development*, N Fox, R Davidson, editors. Erlbaum, Hillsdale, NJ, 1984.

Suomi SJ: Early stress and adult emotional reactivity in rhesus monkeys. In *Childhood Environment and Adult Disease: Symposium No. 156*, Ciba Foundation Symposium Staff, editors. Wiley, Chichester, England, 1991.

Swendsen J, Hammen C, Heller T, Gitlin M: Correlates of stress reactivity in patients with bipolar disorder. Am J Psychiatry *152:*795, 1995.

CHAPTER 14. MOOD DISORDERS

14.6 MOOD DISORDERS: CLINICAL FEATURES

HAGOP S. AKISKAL, M.D.

HETEROGENEITY OF MOOD DISORDERS

Terminology Mood disorders are characterized by pervasive dysregulation of mood and psychomotor activity and by related biorhythmic and cognitive disturbances. The rubric of “affective disorder,” which in some European classifications also subsumes morbid anxiety states, is increasingly being replaced by the nosologically more delimited concept of “mood disorder.” Thus *mood disorder* is now the preferred term in both the World Health Organization's (WHO's) 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) and the American Psychiatric Association's (APA's) fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Official mood disorder categories in current use include bipolar disorders (with manic or hypomanic, depressive, or mixed episodes) and major depressive disorders and their respective attenuated variants known as cyclothymic and dysthymic disorders. Conditions that in earlier editions of these manuals were categorized as “endogenous depression,” “involutional melancholia,” and “psychotic depressive reaction” have been incorporated into major depressive disorder, whereas “depressive neurosis” has been largely absorbed by dysthymic disorder. Although the neurotic-endogenous distinction has been officially deleted, the term “melancholic features” is now used to qualify major depressive disorders in which biological concomitants predominate. While both the American and international classifications recognize the common occurrence of mixed anxiety-depressions, whether they should be classified with mood disorders or with anxiety disorders remains unresolved. It is equally uncertain how to classify the classic neurasthenic conditions, which have recently reemerged and overlap to some extent with the so-called chronic fatigue syndrome.

Destigmatization The reshuffling and reclassification of various affective conditions into the mood disorders chapter of the third edition of DSM (DSM-III) and DSM-IV has, on balance, considerably broadened their boundaries. This change reflects, in part, new developments in pharmacotherapy that have resulted in considerable alleviation of suffering for persons whose illnesses fall short of and sometimes beyond the boundaries of classic mood disorders. As a result, many persons with recurrent mood disorders who would have been disabled can now lead productive lives. Such gratifying results have, in turn, helped to destigmatize this group of disorders. Destigmatization has been further facilitated by published self-revelations of famous persons with depressive and bipolar disorders.

Spectrum of Mood Disorders As often happens when new therapeutic interventions prove successful, the past two decades have witnessed an increased readiness to diagnose mood disorders and their variants. These developments should not be dismissed as mere therapeutic fad, however. External validating strategies, such as familial-genetic studies and prospective follow-up, can now be used to buttress the broadened concept of mood disorders. New research comparing monozygotic and dizygotic twins has demonstrated that the genetic propensity to mood disorders embraces entities that extend beyond endogenous depression (melancholia in DSM-IV) to subsume a larger variety of depressions, including some encountered in persons in the community who have never received psychiatric treatment. Although such data might seem counterintuitive to those who would restrict depression to a core primary biological disease, they suggest that the constitutional predisposition for affective dysregulation occurs in as many as one of every three persons. That ratio is similar to the proportion of those who progress to a full depressive syndrome following bereavement, of rhesus monkeys developing depressive-like behavior following a separation paradigm, and of dogs who develop learned helplessness after inescapable shock. The fact that these rates are considerably higher than one observes in clinical populations suggests that many subjects possess protective factors against major depressive episodes; alternatively, the data suggest that other factors determine which person with emotional distress will become a clinical case. A great deal might therefore be revealed about the nature of pathological affective processes through study of self-limiting affective conditions on the border of mood disorders.

The suffering and dysfunction resulting from mood disorders are among the most common reasons for consulting psychiatrists and other physicians. In fully developed cases, all activity stops—including creative powers—and life is grim and in total disarray (as portrayed in Dürer's masterpiece, Fig. 14.6-1).

FIGURE 14.6-1 *Melancholia* (1514) by Albrecht Dürer.

All great physicians of the past, beginning with Hippocrates, have devoted considerable space in their general medical texts to the clinical characterization of melancholic and manic states, as well as alternations in the same patient. Greco-Roman medicine recognized a broad spectrum of affective disturbances, ranging from the relatively mild temperamental variants (represented in the official nosology by dysthymic and cyclothymic disorders) to their severest forms (including what today is considered mood disorder with mood-congruent and mood-incongruent psychotic features). The ancients also recognized the intimate relation of morbid states of fear to melancholia. Furthermore, they noted that melancholia and certain physical diseases shared seasonal incidence and described the common occurrence of alcohol indulgence, especially in those prone to mania.

Boundaries The boundaries between temperament (personality) and mood disorder, grief and melancholia, anxiety and depressive states, depressive and bipolar disorders, mood-congruent and mood-incongruent psychotic features, and other (schizophrenic) psychotic conditions remain unresolved. Mood disorders have long been known to be highly comorbid with alcohol use and somatic disease; these trends continue today, with the addition of substance use disorders.

AFFECTS, MOODS, TEMPERAMENTS, AND MORBID MOOD STATES

Ethological Considerations *Affects* and *moods* refer to different aspects of emotion. Affect is communicated through facial expression, vocal inflection, gestures, and posture and (according to current ethological research) is intended to move human beings and other primates to appraise whether a person is satisfied, distressed, disgusted, or in danger. Thus joy, sadness, anger, and fear are basic affects that serve a communicative function in primates as well as many other mammalian species.

Affects tend to be short-lived expressions, reflecting momentary emotional contingencies. Moods convey sustained emotions; their more-enduring nature means that

they are experienced long enough to be felt inwardly. Moods are also manifested in subtle ways, and their accurate assessment often requires empathic understanding by the interviewer. The words that subjects use to describe their inner emotions may or may not coincide with the technical terms used by researchers or clinicians and often vary from one culture to another. Furthermore, the inward emotion and the prevailing affective tone may be discordant. This conflict could be due to deliberate simulation (i.e., the subject does not wish to reveal his or her inner emotion) or it could result from a pathological lesion or process that has altered the emotions and their neural substrates. Thus, evaluating moods and affective expression requires considerable clinical experience.

Sadness and Joy The normal emotions of sadness and joy are part of everyday life and should be differentiated from major depressive disorder and mania. Sadness, or normal depression, is a universal human response to defeat, disappointment, or other adversities. The response may be adaptive, in an evolutionary sense, by permitting withdrawal to conserve inner resources, or it might signal the need for support from significant others. Transient depressive periods also occur as reactions to certain holidays or anniversaries, as well as during the premenstrual phase and the first week postpartum. Termed, respectively, “holiday blues,” “anniversary reactions,” “premenstrual tension disorder” and “maternity blues,” they are not psychopathological per se, but those predisposed to mood disorder may develop clinical depression during such times.

Premenstrual Dysphoric Mood Changes In view of the higher prevalence of depressive disorders in women, premenstrual affective changes—dysphoria, tension, irritability, hostility, and labile mood—have received both clinical and research attention. The attempt to establish a specific premenstrual dysphoric disorder has neglected the not uncommon occurrence of premenstrual eutonia, increased energy, and sexual drive. The not uncommon occurrence of these positive emotions, along with the labile mixed affective manifestations, tend to point toward a “bipolar” phenomenon. Although women with severe premenstrual complaints appear to have higher rates of lifetime major mood disorders, a recent twin study found that genetic and environmental factors contributing to premenstrual depression and major depressive disorders are largely distinct. Furthermore, events such as migraine, epileptic attacks, and panic states may, in some instances, be associated with the premenstrual phase. The foregoing considerations suggest the hypothesis that premenstrual psychobiological changes exacerbate different neuropsychiatric disorders to which women are otherwise predisposed. Whether the exaggerated premenstrual variability in emotional equilibrium constitutes a variant of mood disorder must await more definitive studies.

Grief Normal bereavement or grief, considered the prototype of reactive depression, occurs in response to significant separations and losses such as death, divorce, romantic disappointment, leaving familiar environments, forced emigration, or civilian catastrophes. DSM-IV tends to limit the concept of normal grief to loss due to death. However, the work of Elie Karam and colleagues showed that losses associated with the civil war in Lebanon served as potent forces in depression formation. In addition to depressed affect appropriate to the loss, bereavement reactions are characterized by the prominence of sympathetic arousal and restlessness, believed to represent (from an evolutionary perspective) physiological and behavioral mechanisms to facilitate the search for the lost object. Like other adversities, bereavement and loss do not generally seem to cause depressive disorder, except in those predisposed to mood disorder.

Elation The positive emotion of elation is popularly linked to success and achievement. However, paradoxical depressions may also follow such positive events, possibly because of the increased responsibilities that often have to be faced alone. Elation is conceptualized psychodynamically as a defense against depression or as a denial of the pain of loss, as exemplified by the so-called maniacal grief, a rare form of bereavement reaction in which elated hyperactivity may replace the expected grief.

Other pseudomanic states include the brief energetic and unusually lucid periods encountered in dying patients or in those who need to take superhuman action in the face of unusual duress, both of which have been conceptualized as “flight into health.” In predisposed persons such reactions might be the prelude to a genuine manic episode. Sleep deprivation, which commonly accompanies major stressors, might represent one of the intermediary mechanisms between stressor and adverse clinical outcome.

Affective Temperaments Another mediating factor between normal and pathological moods is temperament. Most persons have a characteristic pattern of basal affective oscillations that defines their temperament. For instance, some are easily moved to tears by sad or happy circumstances, whereas others tend to remain placid. Normally oscillations in affective tone are relatively minor, tend to resonate with day-to-day events, and do not interfere with functioning. Some exhibit greater variability of emotional responses whereby, with no obvious provocation, the person alternates between normal mood and sadness or elation, or both. Temperaments tend to cluster into basic types, four of which are of the greatest relevance to mood disorders. The depressive temperament, in which the person easily swings into the sad direction, occurs in 3 to 6 percent of the general population; the hyperthymic temperament, in which the person is naturally inclined toward cheerful moods has been reported in 4 to 8 percent and the cyclothymic temperament swinging between cheerful and sad moods characterizes 4 to 6 percent of young adults. All three types have an early insidious onset and tend to persist throughout adult life. An irritable-explosive type occurs in 2 to 3 percent of young subjects and tends to attenuate by middle age.

An examination of the traits associated with these temperaments can provide the rationale for Ernst Kretschmer's hypothesis about the social functions they served. Thus, the person with a depressive temperament is hard working, dependable, and suitable for jobs that require long periods of devotion to meticulous detail (Table 14.6-1). Such persons shoulder the burdens of existence without experiencing its pleasures. A person with the hyperthymic temperament, endowed with high levels of energy, extroversion, and humor (Table 14.6-1), will assume leadership positions in society or excel in the performing arts or entertainment. In talented persons the cyclothymic temperament, which alternates between sadness and elation, could provide the inspiration and the intensity needed for composing music, painting, or writing poetry. One with the irritable temperament, probably a variant of the cyclothymic type, might be best suited for a military career or even revolutionary action. The danger with such temperaments is that they could swing too far in one or the other direction, or in both directions (i.e., major depressive, manic, or mixed episodes). Use of such substances as alcohol, caffeine, and other stimulants might further destabilize affective regulation in persons with those attributes. Some adolescent girls with the irritable temperament might develop the extreme emotional disequilibrium that in contemporary psychiatry is considered borderline personality disorder.

Depressive	Hyperthymic
Gloomy, incapable of fun, complaining	Cheerful and exuberant
Humorless	Articulate and jocular
Pessimistic, and given to brooding	Overoptimistic and carefree
Guilt-prone, low self-esteem, and preoccupied with inadequacy or failure	Overconfident, self-assured, boastful, and grandiose
Introverted with restricted social life	Extroverted and people seeking
Sluggish, living a life out of action	High energy level, full of plans
Few but constant interests	Versatile with broad interests
Passive	Overinvolved and meddlesome
Reliable, dependable, and devoted	Uninhibited and stimulus seeking

Table 14.6-1 Attributes, Assets and Liabilities of Depressive and Hyperthymic Temperaments

Morbid Mood States Mood disorders represent abnormal or extreme variations of mood and associated manifestations and are characterized by the following features.

Pathological Mood Change Pathological moods are distinguished from heir normal counterparts by being out of proportion to any concurrent stressor or situation; being unresponsive to reassurance; being sustained for weeks, months, and sometimes years; and having a pervasive effect on the person, such that judgment is seriously influenced by the mood.

Endoreactive Moods Depression and mania are diagnosed, respectively, when sadness or elation is overly intense and continues beyond the expected impact of a stressful life event. Indeed the morbid mood might arise without apparent or significant life stress. The pathological process in mood disorders is thus partly defined by the ease with which an intense emotional state is released and especially by its tendency to persist autonomously even when the offending stressor is no longer operative. Rather than being endogenous (i.e., occurring in the absence of precipitants), mood disorders are best conceptualized as endoreactive (i.e., once released, they tend to persist autonomously). The homeostatic dyscontrol of mood, which is part of a more pervasive mood dysregulation, resists reversal to the habitual baseline affective tone. DSM-IV, which tends to disparage theory and adhere to a descriptive level of operationalization, gives insufficient weight to this fundamental

characteristic of mood disorders.

Recurrence In a more descriptive vein, what sets mood disorders apart from their normal emotional counterparts is the clustering of signs and symptoms into discrete syndromes that typically recur on an episodic basis or pursue an intermittent, subthreshold course over the span of many years, if not a lifetime. Cyclic course and in some cases regular recurrence, or periodicity, are other signs of mood dysregulation particularly relevant to bipolar disorder.

Impairment Normative reactions to adversity and stress, including biological stress, typically consist of transient admixtures of anxiety and dysphoria that are best captured under the DSM-IV rubric of adjustment disorder with mixed emotional features. That is, the self-limiting reactions are best qualified broadly as normal affective states that produce little, if any, impairment in the main areas of functioning.

Although anxiety, irritability, and anger do occur in various types of mood disorders, pathologically sustained mood states of depression and elation characterize those disorders. Morbid mood states (mood disorders) then consist of protracted emotional reactions that deepen or escalate, respectively, into clinical depression or mania, with a tendency to recur or evolve into unremitting chronicity in 15 to 20 percent of cases. The contribution of temperamental peculiarities to such outcomes should be apparent. The impaired functioning characteristic of mood disorders is thus based on a combination of factors, including severity, autonomy, recurrence, and chronicity of the clinical features.

To recapitulate, dysregulation in mood disorders can take different forms. It could be expressed as a single severe episode that persists autonomously for many months and sometimes years or it might recur with episodes of varying severity, years apart or in rapid succession, with or without interepisodic remission. In general, the earlier the age at onset, the more likely are recurrences, especially those of bipolar nature. Thus, depending on the course of the illness, impairment could be state dependent and occur during an episode or it could extend into the interepisodic period. According to National Institute of Mental Health (NIMH) estimates, on average, a woman with bipolar disorder spends 12 years in florid episodes (often hospitalized), loses 14 years from a productive career and motherhood, and has her life curtailed by 9 years.

Recent observations have also revealed another pattern of impairment. In dysthymic and cyclothymic disorders, which represent an intensification of temperamental instability, impairment is not due to the severity of the mood disturbance per se, but to the cumulative impact of the dysregulation beginning in the juvenile or early adult years and continuing unabated or intermittently over long periods; hence the frequent confusion with character pathology. Here the impairment is more subtle but nonetheless pervasive. Persons with cyclothymic disorder tend to be dilettantes, whereas those with dysthymic disorder often lead morose and colorless lives.

PSYCHOPATHOLOGY

Depressive Syndrome Like other illnesses, depressive disorder clusters into signs and symptoms that constitute what DSM-IV and ICD-10 term major depressive episode (Tables 14.6-2). These criteria attempt to set an operational threshold for depressive disorder based on a specified number of items and their temporal patterns. The diagnosis of clinical depression cannot be accomplished by a checklist: The DSM-IV diagnostic criteria for major depressive disorder provide only a general guide. Only after an in-depth phenomenological approach can a clinician ascertain diagnosis of a depressive disorder. Disturbances in all four spheres (mood, psychomotor activity, cognitive, and vegetative) should be ordinarily present for a definitive diagnosis of major depressive disorder, although that is not specified in DSM-IV.

Table 14.6-2 DSM-IV Criteria for Major Depressive Episode

Mood Disturbances Mood change, usually considered the sine qua non of morbid depression, appears in a variety of disturbances, including (1) painful arousal, (2) hypersensitivity to unpleasant events, (3) insensitivity to pleasant events, (4) insensitivity to unpleasant events, (5) reduced anticipatory pleasure, (6) anhedonia or reduced consummatory pleasure, (7) affective blunting, and (8) apathy. The phenomenology and psychometric properties of this broad range of mood disturbances are under investigation at the Salpêtrière Hospital in Paris. Our focus here is primarily on painfully aroused mood (depression) and diminished capacity for pleasure (anhedonia), two mood disturbances given selective weight in DSM-IV and ICD-10.

DEPRESSED MOOD The term “depressed mood” refers to negative affective arousal, variously described as depressed, anguished, mournful, irritable, or anxious. These terms tend to trivialize a morbidly painful emotion, typically experienced as worse than the severest physical pain. Thus depressed mood has a somatic quality that in the extreme is indescribably painful. Even when not so severe, depressive suffering is qualitatively distinct from its neurotic counterparts, taking the form of groundless apprehensions with severe inner turmoil and torment. This description is particularly apt for middle-aged and elderly persons, who were once considered to be suffering from “involutional melancholia.” The sustained nature of the mood permits no respite, although it tends to lift somewhat in the evening. Suicide may represent an attempt to find deliverance from such unrelenting psychic torment; death can be experienced as comforting (Fig. 14.6-2).

FIGURE 14.6-2 *Death Giving Comfort* by Kaethe Kollwitz (1867–1945).

Patients with a milder form of the malady typically seen in primary care settings might deny experiencing mournful moods and instead complain of physical agony from headache (Fig. 14.6-3), epigastric pain, precordial distress, and so on, in the absence of any evidence of diagnosable physical illness. Such conditions have been described as “depressio sine depressione,” or “masked depression.” In such cases, commonly observed in older patients, the physician should corroborate the presence of mood disturbance by the depressed affect in the patient's facial expression, voice, and overall appearance.

FIGURE 14.6-3 *Headache* by Honoré Daumier (1808–1879).

ANHEDONIA AND LOSS OF INTEREST Paradoxically, the heightened perception of pain in many persons with depressive disorder is accompanied by an inability to experience normal emotions. Patients exhibiting the disturbance may lose the capacity to cry, a deficit that is reversed as the depression is lifting.

In evaluating anhedonia inquiring whether the patient has lost the sense of pleasure is not enough; the clinician must document that the patient has actually given up previously enjoyed pastimes. When mild, anhedonia evidences with decreased interest in life. Later, patients complain that they have lost all interest in things. This is best illustrated by William Shakespeare in Hamlet's disgust: “How weary, stale, flat, and unprofitable seem to me all the uses of the world” (Act I, Scene II). In the extreme, patients lose their feelings for their children or spouses, who once were a source of joy. Thus the hedonic deficit in clinical depression might represent a special instance of a more pervasive inability to experience emotions.

Patients with severe depression may complain of being emotionally cut off from others and experience depersonalization and a world that seems strange to them (derealization). The impact of the loss of emotional resonance can be so pervasive that patients may denounce values and beliefs that had previously given meaning to their lives. For instance, members of the clergy might present with the complaint that they no longer believe in the Church, that they have lost God. The inability of the person with depressive disorder to experience normal emotions (commonly observed among young depressed patients) differs from the schizophrenic patient's flat affect in that the loss of emotions is itself experienced as painful; that is, the patient suffers immensely from the inability to experience emotions.

Psychomotor Disturbances In depression psychomotor changes consist of abnormalities in the motor expression of mental and emotional activity. In severe cases, these changes manifest in specific facial features (Fig. 14.6-4).

FIGURE 14.6-4 The Swiss neuropsychiatrist Otto Veraguth described a peculiar triangle-shaped fold in the nasal corner of the upper eyelid. The fold, often associated with depression, is referred to as Veraguth's fold. The photograph illustrates this physiognomic feature in a 50-year-old man during a major depressive episode. Veraguth's fold may also be seen in persons who are not clinically depressed, usually while they are harboring a mild depressive affect. Distinct changes in the tone of the corrugator and zygomatic facial muscles accompany depression, as shown on electromyograms. (Courtesy of Heinz E. Lehmann, M.D.)

PSYCHOMOTOR AGITATION Although agitation (pressured speech, restlessness, hand wringing, and hair pulling) is the more readily observed abnormality, it appears to be less specific to the illness than retardation (slowing of psychomotor activity). Psychophysiological studies have documented that such slowing often coexists with agitation.

PSYCHOMOTOR RETARDATION Underlying many of the deficits seen in clinical depression, some authorities believe psychomotor retardation to be the core, or primary, pathology in mood disorders. Morbid depression—what patients describe as being “down”—can be understood in terms of moderate-to-extreme psychomotor slowing. The patient experiences inertia, being unable to act physically and mentally. Recent brain imaging research that has revealed subcortical (extrapyramidal system) disturbances in mood disorders tends to support the centrality of psychomotor dysfunction in these disorders.

Long neglected in psychopathological research, psychomotor retardation, can be measured with precision. The Salpêtrière Retardation Scale developed by Daniel Widlöcher and colleagues places special emphasis on the following disturbances: (1) paucity of spontaneous movements; (2) slumped posture with downcast gaze (Fig. 14.6-5); (3) overwhelming fatigue (patients complain that everything is an effort”); (4) reduced flow and amplitude of speech and increased latency of responses, often giving rise to monosyllabic speech; (5) a subjective feeling that time is passing slowly or has stopped; (6) poor concentration and forgetfulness; (7) painful rumination—thinking that dwells on a few (usually unpleasant) topics; and (8) indecisiveness, or an inability to make simple decisions.

FIGURE 14.6-5 A 38-year-old woman during a state of deep retarded depression **(A)** and 2 months later, after recovery **(B).** Note the turned-down corners of her mouth, her stooped posture, her drab clothing, and her hairdo during the depressed episode. (Courtesy of Heinz E. Lehmann, M.D.)

DSM-IV places greater emphasis on the more easily observable objective or physical aspects of retardation. For the patient, however, the subjective sense of slowing is as pervasive and disabling. This more psychological dimension of retardation is most reliably elicited from depressed persons with good verbal skills.

Ms. A, a 34-year-old literature professor, presented to a mood clinic with the following complaint: “I am in a daze, confused, disoriented, staring. My thoughts do not flow, my mind is arrested... I seem to lack any sense of direction, purpose...I have such an inertia, I cannot assert myself. I cannot fight, I have no will.”

Less linguistically sophisticated patients would simply complain of an inability to perform household chores or difficulty in concentrating on their studies. Such psychomotor deficits in turn underlie depressed patients' diminished efficiency or their inability to work.

PSEUDODEMENTIA The slowing of mental functions can be so pronounced in elderly persons that they experience memory difficulties, disorientation, and confusion.

STUPOR Psychomotor slowing in young persons is sometimes so extreme that patients might slide into a stupor, unable to participate even in such basic biological functions as feeding themselves. Such an episode is often the precursor of bipolar disorder, which later declares itself in a manic episode. Today depressive disorder is diagnosed in its earlier stages, and subtle stupor is much more likely to be encountered clinically.

A 20-year-old male college student seen in the emergency room spoke of “being stuck—as if I have fallen into a black hole and can't get out.” Further evaluation revealed that the patient was metaphorically describing his total loss of initiative and drive and was engulfed by the disease process. A clinician without the requisite phenomenological training, might consider such a patient bizarre and perhaps even psychotic. Yet the patient responded dramatically to fluoxetine (Prozac) and in 2 weeks was back in school.

Cognitive Disturbances The cognitive view of depression considers negative evaluations of the self, the world, and the future (the negative triad) central to understanding depressed mood and behavior, but it is equally likely that the depressed mood colors perceptions of the self and others or that disturbed psychomotor activity leads to negative self-evaluations. Therefore, instead of being considered causal, the cognitive triad in depression is best approached empirically as a psychopathological manifestation of depression. Those faulty thinking patterns are clinically expressed as (1) ideas of deprivation and loss; (2) low self-esteem and self-confidence; (3) self-reproach and pathological guilt; (4) helplessness, hopelessness, and pessimism; and (5) recurrent thoughts of death and suicide.

The essential characteristic of depressive thinking is that the sufferer views everything in an extremely negative light. The self-accusations are typically unjustified or are blown out of proportion, as in the case of a middle-aged woman who was tormented by guilt because as a child she had not repaid 5 cents she had borrowed from a classmate. Some of the thoughts may verge on the delusional. For instance, an internationally renowned scientist complained that he was “nothing.” Self-evaluations that indicate an extremely low image of self might nonetheless reflect an accurate perception of one's impairment from psychomotor retardation.

MOOD-CONGRUENT PSYCHOTIC FEATURES In depressive disorder with psychotic features (Table 14.6-3), negative thinking acquires grossly delusional proportions and is maintained with such conviction that the thoughts are not amenable to change by evidence to the contrary. According to Kurt Schneider, delusional thinking in depression derives from humankind's four basic insecurities, those regarding health, financial status, moral worth, and relationship to others. Thus, severely depressed patients may have delusions of worthlessness and sinfulness, reference, and persecution: They believe they are being singled out for their past mistakes and that everyone is aware of their errors. Persecutory ideation in depression is often prosecutory in that it derives from the belief that the person deserves punishment for such transgressions. A severely depressed man may feel so incompetent in all areas of functioning, including the sexual sphere, that he may suspect his wife of having an affair (delusion of infidelity).

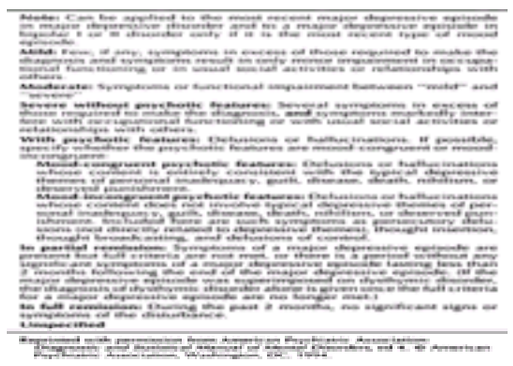

Table 14.6-3 DSM-IV Criteria for Severity/Psychotic/Remission Specifiers for Current (or Most Recent) Major Depressive Episode

Other depressed persons believe that they have mismanaged their finances and their children will starve (delusions of poverty) or that they harbor an occult illness, such as cancer or the acquired immune deficiency syndrome (AIDS) (delusions of ill health) or that parts of their bodies are missing (nihilistic delusions). In more severe illness the patient might feel that the world has changed and that calamity and destruction await everyone. In rare instances a parent with such delusions might kill his or her young children to save them from moral or physical decay and then commit suicide. Finally, a minority of depressed persons have fleeting auditory or visual hallucinations with extremely unpleasant content along the lines of their delusions (e.g., hearing accusatory voices or seeing themselves in coffins or graveyards). All of these psychotic experiences are genuine affective delusions or hallucinations. They are mood congruent in the sense that they are phenomenologically understandable in light of the prevailing pathological mood.

MOOD-INCONGRUENT PSYCHOTIC FEATURES Sometimes so-called first-rank or schneiderian-type symptoms can arise in the setting of a major depressive episode.

A 42-year-old civil servant said she was so paralyzed by depression that she felt that she had no personal initiative and volition left; she believed some malignant force had taken over her actions, and it would comment on every action that she would undertake. The patient recovered fully with thymoleptic medication. There is no reason to believe that in this patient the feelings of somatic passivity and running commentary indicated a schizophrenic process.

Thus, with proper phenomenological probing, certain classes of apparently mood-incongruent psychotic experiences listed in DSM-IV can be understood as arising from the pathological mood and the profound changes in psychomotor activity that accompany them. (In other instances, the clinician must seek a history of alcohol or substance use disorder or withdrawal as a putative explanation for mood incongruence in psychotic depression.) In brief, incidental schneiderian first-rank symptoms should not distract from the diagnosis of an affective disorder if otherwise typical signs and symptoms are present.

HOPELESSNESS AND SUICIDE Given that most, if not all, clinically depressed patients find themselves locked in the private hell of their negative thoughts, it is not surprising that up to 15 percent of untreated or inadequately treated patients give up hope of ever recovering and kill themselves. The suicide attempt is not, however, undertaken in the depth of melancholia. One severely depressed patient asked if she had any suicide plans, replied, “Doctor, I don't exist—I am already dead.”

Thus the risk of suicide is less pronounced during acute severe depression. Emil Kraepelin observed that it is when psychomotor activity is improving, and yet mood and thinking are still dark, that the patient is most likely to muster the requisite energy to commit the suicidal act. Hopelessness on mental status evaluation in a patient recovering from depression should alert the clinician to the possibility of such an outcome.

There is no basis for the common belief that inquiring about suicide provokes such behavior. On the contrary, patients are often relieved that the physician appreciates the magnitude of their suffering. Suicidal ideation is commonly expressed indirectly (e.g., in a wish not to wake up or to die from a malignant disease). Some depressed persons are tormented with suicidal obsessions and are constantly resisting unwanted urges or impulses to destroy themselves. Others might yield to such urges passively (e.g., by careless driving or by walking into high-speed traffic). A third group harbors elaborate plans, carefully preparing a will and taking out insurance. Deliberate planning indicates a very high suicidal risk. The foregoing examples are not exhaustive; they are meant to remind clinicians in charge of depressed patients to be always alert to the possibility of suicide.

Vegetative Disturbances The Greeks considered depression a somatic illness and ascribed it to black bile; hence the term “melancholia.” The mood change in

depressive disorder is accompanied by measurable alterations of biorhythms that implicate midbrain dysfunction. Once the changes occur, they tend to be independent of the environment throughout much of the episode, and as a consequence, they do not respond to interpersonal feedback of a pleasant and upbeat nature. The biological concomitants of melancholia include profound reductions in appetite, sleep, and sexual functioning as well as alterations in other circadian rhythms, especially matinal worsening of mood and psychomotor performances. These disturbances are central to the DSM-IV concept of melancholia (Table 14.6-4), a form of depression in which such biological concomitants predominate. A smaller subgroup of depressed persons exhibits a reversal of the vegetative and circadian functions, with increases in appetite and sleep—and sometimes in sexual functioning—and an evening worsening of mood; in this atypical pattern (Table 14.6-5), patients characteristically exhibit mood reactivity and sensitivity to rejection.

Specify if:
With melancholic features (can be applied to *major depressive* episodes occurring in major depressive disorder, bipolar I disorder or bipolar II disorder only if it is the most recent type of mood episode)

A. Either of the following, occurring during the most severe period of the current episode:
   (1) loss of pleasure in all, or almost all, activities.
   (2) lack of reactivity to usually pleasurable stimuli (does not feel much better, even temporarily, when something good happens)

B. Three (or more) of the following:
   (1) distinct quality of depressed mood (i.e., the depressed mood is perceived as distinctly different from the kind of feeling experienced after the death of a loved one)
   (2) the depression is regularly worse in the morning
   (3) early morning awakening (at least 2 hours before usual time of awakening)
   (4) marked psychomotor retardation or agitation
   (5) significant anorexia or weight loss
   (6) excessive or inappropriate guilt

Table 14.6-4 DSM-IV Criteria for Melancholic Features Specifiers

Specify if:
With atypical features (can be applied when these features predominate during the most recent 2 weeks of a major depressive episode in major depressive disorder or in bipolar I or bipolar II disorder when the major depressive episode is the most recent type of mood episode, or when these features predominate during the most recent 2 years of dysthymic disorder)

A. Mood reactivity (i.e., mood brightens in response to actual or potential positive events).

B. Two (or more) of the following features, present for most of the time, for at least 2 weeks:
   (1) significant weight gain or increase in appetite
   (2) hypersomnia
   (3) leaden paralysis (i.e., heavy, leaden feelings in arms or legs)
   (4) long-standing pattern of interpersonal rejection sensitivity (not limited to episodes of mood disturbance) resulting in significant social or occupational impairment

C. Criteria are not met with melancholic features or with catatonic features during the same episode.

Table 14.6-5 DSM-IV Criteria for Atypical Features Specifier

ANOREXIA AND WEIGHT LOSS The most reliable somatic indicators of depressive disorder include anorexia and weight loss. In addition to the presumed hypothalamic disturbance of depression, anorexia might be secondary to blunted olfactory or taste sensations or a decreased enjoyment of food, or (rarely) it might result from a delusional belief that the food has been poisoned.

If weight loss is severe, especially after the age of 40, the psychiatrist should first use appropriate medical consultation to rule out the likelihood of an occult malignancy. Inanition, especially in elderly persons, can lead to malnutrition and electrolyte disturbances that represent medical emergencies.

WEIGHT GAIN Overeating, decreased activity, or both may result in weight gain. In middle-aged patients it may aggravate preexisting diabetes mellitus, hypertension, or coronary artery disease. In younger patients, especially women, weight problems may conform to a bulimic pattern that is sometimes the expression of the depressive phase of a bipolar disorder with infrequent hypomanic periods (bipolar II disorder).

INSOMNIA Sleep disturbance, a cardinal sign of depression, often is characterized by multiple awakenings, especially in the early hours of the morning, rather than by difficulty falling asleep. The light sleep of a depressed person, in part a reflection of the painful arousal of the disorder, tends to prolong the depressive agony over 24 hours. Thus, deep stages of sleep (3 and 4) are either decreased or deficient. The attempt to overcome the problem by drinking alcohol may initially succeed but ultimately aggravates the sleep patterns and insomnia. This is also true for sedative-hypnotic agents, which are often prescribed by the busy general practitioner who has not spent enough time diagnosing the depressive condition. Although sedatives (including alcohol) effectively reduce the number of awakenings in the short term, they are not effective in the long run because they further diminish stage 3 and stage 4 sleep. They are not antidepressants, and they tend to prolong the depression.

HYPERSOMNIA Young depressed patients, especially those with bipolar tendencies, often exhibit excessive sleep and have difficulty getting up in the morning.

Kevin, a 15-year-old boy, was referred to a sleep center to rule out narcolepsy. His main complaints were fatigue, boredom, and a need to sleep all the time. Although he had always started the day somewhat slowly, he now could not get out of bed to go to school. That alarmed his mother, prompting sleep consultation. Formerly a B student, he had been failing most of his courses in the 6 months before referral. Psychological counseling, predicated on the premise that his family's recent move from another city had led to Kevin's isolation, had not been beneficial. Extensive neurological and general medical workup had also proven negative. He slept 12 to 15 hours a day but denied cataplexy, sleep paralysis, and hypnagogic hallucinations. During psychiatric interview he denied being depressed but admitted that he had lost interest in everything except his dog. He had no drive, participated in no activities, and had gained 30 pounds in 6 months. He believed he was “brain damaged” and wondered whether it was worth living like that. The question of suicide disturbed him as it was contrary to his religious beliefs. These findings led to the prescription of desipramine (Norpramin) in a dosage that was gradually increased to 200 mg a day over 3 weeks. Not only did desipramine reverse the presenting complaints, but it also pushed him to the brink of a manic episode.

The affective nature of the disorder in such patients is often unrecognized, and their behavior is attributed to “laziness.” The vignette also illustrates the emergence of manic behavior during antidepressant treatment. Such shifts in polarity are common in major depressive disorder and necessitate revising the diagnosis to a bipolar disorder (contrary to the admonitions of DSM-IV).

CIRCADIAN DYSREGULATION Many circadian functions, such as temperature regulation and cortisol rhythms, are disrupted in major depressive disorder. Disturbances of sleep rhythms, however, have received the greatest research focus. These include deficits in stage 4 or delta sleep, as well as more intense rapid eye movement (REM) activity in the first third of the night. More specific to depressive disorders—and whether suffering from insomnia or hypersomnia—nearly two-thirds of patients exhibit a marked shortening of REM latency, the period from the onset of sleep to the first REM period. This abnormality is observed throughout the depressive episode and may also be seen during relatively euthymic periods in persons with recurrent depression. The occurrence of short REM latency in the younger “well” relatives of the affectively ill suggests that neurophysiological abnormalities might precede the overt psychopathological manifestations of the illness; upon closer scrutiny, these well relatives will often be found to meet criteria for subthreshold mood conditions such as dysthymic disorder, intermittent depression or labile temperament.

Few data exist on the consistency of sleep electroencephalographic (EEG) abnormalities in patients from episode to episode. However, clinical experience suggests that a patient observed over time (even during the same episode) may exhibit insomnia and morning worsening of mood and activity during one period of the disorder and hypersomnia extending to late morning hours during another period. In either case, persons with depressive disorder are characteristically tired in the morning, which means that even prolonged sleep is not refreshing for them. The propensity to exhibit such divergent patterns of sleep disturbance is more likely in bipolar disorders. Patients with major depressive disorder tend to exhibit insomnia more stereotypically episode after episode; despite extreme fatigue, they rarely oversleep. Such fatigue coexisting with negative affective arousal is even more exhausting.

SEASONALITY Another classic biorhythmic disturbance in mood disorders is seasonal (especially autumn-winter) accentuation or precipitation of depression. Most of those patients experience increased energy and activation, if not frank hypomania, in the spring. In the fall and winter, they complain of fatigue, tend to crave sugars, and overeat and oversleep. The hypersomnia in some of these patients is associated with delayed (rather than short) REM latencies. These data suggest dysregulation of circadian rhythms in depressive disorders rather than mere phase advance. Although autumnal-winter depression has received the greatest attention, there also exist summer depressions; the former appear related to reduction of daylight (photoperiods), and the latter to increased temperature. The DSM-IV criteria for seasonal pattern specifiers are listed in Table 14.6-6.

Specify if:
With seasonal pattern (can be applied to the pattern of major depressive episodes in bipolar I disorder, bipolar II disorder, or major depressive disorder, recurrent)

A. There has been a regular temporal relationship between the onset of major depressive episodes in bipolar I or bipolar II disorder or major depressive disorder, recurrent, and a particular time of the year (e.g., regular appearance of the major depressive episode in the fall or winter)
Note: Do not include cases in which there is an obvious effect of seasonal-related psychosocial stressors (e.g., regularly being unemployed every winter).

B. Full remissions (or a change from depression to mania or hypomania) also occur at a characteristic time of the year (e.g., depression disappears in the spring).

C. In the last 2 years, two major depressive episodes have occurred that demonstrate the temporal seasonal relationships defined in criteria A and B, and no nonseasonal major depressive episodes have occurred during that same period.

D. Seasonal major depressive episodes (as described above) substantially outnumber any nonseasonal major depressive episodes that may have occurred over the individual's lifetime.

Table 14.6-6 DSM-IV Criteria for Seasonal Pattern Specifier

SEXUAL DYSFUNCTION Decreased sexual desire is seen in both depressed men and women. In addition, some women experience temporary interruption of their menses. Depressed women are typically unresponsive to lovemaking or are disinclined to participate in it, a situation that could lead to marital conflict. Psychotherapists might mistakenly ascribe the depression to the marital conflict and devote unnecessarily zealous psychotherapeutic attention to conjugal issues. Decreased or lost libido in men often results in erectile failure, which may prompt endocrinological or urological consultation. Again, depression may be ascribed to the sexual dysfunction rather than the reverse, and definitive treatment may be delayed by the physician's focus on the sexual complaint. Tragically, some men with depressive disorder have been subjected to permanent penile implants before receiving more definitive treatment for their depression. This is less likely to occur in the sildenafil (Viagra) era, but even treatment with such agents would not necessarily resolve the impotence in clinically depressed patients without competent treatment of the mood disorder.

A small subgroup of persons with depressive disorder may exhibit increased sexual drive or activity of a “compulsive” nature. These patients tend to have other atypical features as well; hence the increased sexual drive can be considered the “fifth reverse vegetative sign” (after evening or morning worsening of mood, initial insomnia, hypersomnia, and weight gain). In these depressed persons, increased sexual drive may indicate a mixed episode of bipolar disorder. Further scrutiny in such cases will often reveal a premorbid cyclothymic or hyperthymic temperament.

Manic Syndrome As with clinical depression, the psychopathology of mania (Table 14.6-7) can be conveniently discussed under mood, psychomotor, circadian, and cognitive disturbances. The clinical features of mania are generally the opposite of those of depression. Thus, instead of lowered mood, thinking, activity, and self-esteem, there is elevated mood, a rush of ideas, psychomotor acceleration, and grandiosity. Despite those contrasts, the two disorders share such symptoms as irritability, anger, insomnia, and agitation. Actually, an excess of such symptoms of escalating intensity suggests a mixed phase or mixed episode (Table 14.6-8) of mania and depression occurring simultaneously. Manic and mixed episodes represent the hallmark of what was once termed manic-depressive psychosis and is currently termed bipolar I disorder.

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood, lasting at least 1 week (or any duration if hospitalization is necessary).

B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:
(1) inflated self-esteem or grandiosity
(2) decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
(3) more talkative than usual or pressure to keep talking
(4) flight of ideas or subjective experience that thoughts are racing
(5) distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
(6) increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
(7) excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)

C. The symptoms do not meet criteria for a mixed episode.

D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Note: Manic-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of bipolar I disorder.

Table 14.6-7 DSM-IV Criteria for Manic Episode

A. The criteria are met both for a manic episode and for a major depressive episode (except for duration) nearly every day during at least a 1-week period.

B. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others, or there are psychotic features.

C. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Note: Mixed-like episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of bipolar I disorder.

Table 14.6-8 DSM-IV Criteria for Mixed Episode

Although milder mania (hypomania [Table 14.6-9]) can contribute to success in business, leadership roles, and the arts, recurrences of even mild manic symptomatology are typically disruptive. The elated mood tends to produce overoptimism concerning one's abilities, which coupled with the impulsivity characteristic of mania, often leads to disaster. Thus, accurate and early diagnosis is paramount.

A. A distinct period of persistently elevated, expansive, or irritable mood, lasting throughout 4 days, that is clearly different from the usual nondepressed mood.
B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:
(1) inflated self-esteem or grandiosity
(2) decreased need for sleep (e.g., feels rested after only 3 hours of sleep)
(3) more talkative than usual or pressure to keep talking
(4) flight of ideas or subjective experience that thoughts are racing
(5) distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)
(6) increase in goal-directed activity (either socially, at work or school, or sexually) or psychomotor agitation
(7) excessive involvement in pleasurable activities that have a high potential for painful consequences (e.g., the person engages in unrestrained buying sprees, sexual indiscretions, or foolish business investments)
C. The episode is associated with an unequivocal change in functioning that is uncharacteristic of the person when not symptomatic.
D. The disturbance in mood and the change in functioning are observable by others.
E. The episode is not severe enough to cause marked impairment in social or occupational functioning, or to necessitate hospitalization, and there are no psychotic features.
F. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).
Note: Hypomanic-like episodes that are clearly precipitated by somatic antidepressant treatment (e.g., medication, electroconvulsive therapy, light therapy) should not count toward a diagnosis of bipolar II disorder.

Table 14.6-9 DSM-IV Criteria for Hypomanic Episode

Classic mania as formulated in the DSM-IV operationalization of manic episode (Table 14.6-7) is relatively easy to recognize. Misdiagnosis was once rampant in North American practice as clinicians confused severe mania with schizophrenia, and its milder variants with normality or with narcissistic and sociopathic personality disorders. Like the misdiagnosis of depressive conditions, such errors of clinical judgment are due to a lack of familiarity with the phenomenology of the classic illness. Again, DSM-IV criteria provide only a guideline. The actual diagnosis requires careful history and phenomenologic understanding. The manic patient lifts the observer's mood, makes the examiner smile and even laugh, and can often be irritating. The patient's speech is fast and may even appear “loose,” but it also can often be witty. Finally, the behavior is typically dramatic, expansive, and jesting. For the experienced clinician, the overall gestalt experienced in the presence of such patients is emotionally and qualitatively distinct from that of persons with schizophrenia or frontal lobe diseases; the latter conditions tend to leave the examiner “cold.” These considerations become clearer when the clinical observer systematically examines the psychopathology of mania in the areas of mood, behavior, and thinking.

Mood Disturbance Mood disturbance in mania represents a contrast to that observed in depression, but not entirely.

MOOD ELEVATION The mood in mania is classically one of elation, euphoria, and jubilation, typically associated with laughing, punning, and gesturing.

LABILITY AND IRRITABILITY The prevailing positive mood in mania is not stable, and momentary crying or bursting into tears is common. Also, the high is so excessive that many patients experience it as intense nervousness. When crossed, patients can become extremely irritable and hostile. Thus, lability and irritable hostility are as much features of the manic mood as is elation.

Psychomotor Acceleration Accelerated psychomotor activity, the hallmark of mania, is characterized by overabundant energy and activity and rapid, pressured speech. Subjectively, the patient experiences an unusual sense of physical well-being (eutonia).

FLIGHT OF IDEAS Thinking processes are accelerated, experienced as flight of ideas, and thinking and perception are unusually sharp. The patient may speak with such pressure that associations are difficult to follow; such clang associations are often based on rhyming or chance perceptions and can be lightning fast. The pressure to speak may continue despite development of hoarseness.

IMPULSIVE BEHAVIOR Manic patients are typically impulsive, disinhibited, and meddlesome. They are intrusive in their increased involvement with others, leading to friction with family members, friends, and colleagues. They are distractible and move quickly, not only from one thought to another, but also from one person to another, showing heightened interest in every new activity that strikes their fancy. They are indefatigable and engage in various activities in which they usually display poor social judgment. Examples include preaching or dancing in the street; abuse of long distance calling; buying new cars, hundreds of records, expensive jewelry, or other unnecessary items; paying the bills of total strangers in bars; giving away furniture; impulsive marriages; engaging in risky business ventures; gambling; and sudden trips. Such pursuits can lead to personal and financial ruin.

DELIRIOUS MANIA An extremely severe expression of mania (once known as “Bell's mania”), delirious mania involves frenzied physical activity that continues unabated and leads to a life-threatening medical emergency. This complication, the manic counterpart of stupor, is rare today. (There is no need to invoke here the concept of catatonic features as advocated by DSM-IV (Table 14.6-10). The DSM-IV position is terminologically confusing and phenomenologically imprecise).

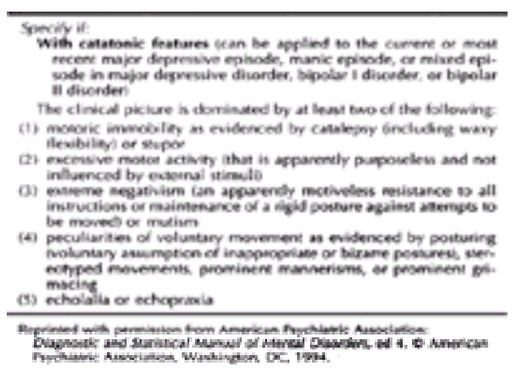

Specify if:
With catatonic features (can be applied to the *current or most recent major depressive episode, manic episode, or mixed episode in major depressive disorder, bipolar I disorder, or bipolar II disorder*)
The clinical picture is dominated by at least two of the following:
(1) motoric immobility as evidenced by catalepsy (including waxy flexibility) or stupor
(2) excessive motor activity (that is apparently purposeless and not influenced by external stimuli)
(3) extreme negativism (an apparently motiveless resistance to all instructions or maintenance of a rigid posture against attempts to be moved) or mutism
(4) peculiarities of voluntary movement as evidenced by posturing (voluntary assumption of inappropriate or bizarre postures), stereotyped movements, prominent mannerisms, or prominent grimacing
(5) echolalia or echopraxia

Table 14.6-10 DSM-IV Criteria for Catatonic Features Specifier

Vegetative Disturbances Vegetative disturbances are more difficult to evaluate in mania than in depression.

HYPOSOMNIA The cardinal sign is decreased need for sleep—the patient sleeps only a few hours but feels energetic on awakening. Some patients may actually go sleepless for several days. This practice could lead to dangerous escalation of manic activity, which might continue despite signs of physical exhaustion.

INATTENTION TO NUTRITION There does not seem to be a clinically significant level of appetite disturbance as such, but weight loss may occur because of increased activity and neglect of nutritional needs.

SEXUAL EXCESSES The sexual appetite is typically increased and may lead to sexual indiscretion. Married women with previously unblemished sexual lives may associate with men below their social status. Men typically overindulge in alcohol, frequent bars, and squander their savings on prostitutes. The sexual misadventures of manic patients result in marital disasters and hence the multiple separations or divorces that are almost pathognomonic of the disorder. Such sexual impulsivity is even more problematic now, in view of the specter of AIDS.

Cognitive Distortions Manic thinking is overly positive, optimistic, and expansive.

GRANDIOSITY, LACK OF INSIGHT, AND DELUSION FORMATION The patient exhibits inflated self-esteem and a grandiose sense of confidence and achievements. Behind that facade, however, may be a vague and painful recognition that the positive self-concepts do not represent reality. However, such insight (if present at all) is transient, and manic patients are notoriously refractory to self-examination and insight. Denial and lack of insight, cardinal psychological derangements of mania, are not listed in the DSM-IV criteria for manic episode or bipolar disorders. This is a serious omission because this lack of insight leads manic patients to engage in activities that harm themselves and their loved ones. It also explains, in part, their noncompliance with medication regimens during the manic phase. Finally, because of their lack of insight, mania nearly always reaches delusional proportions, including delusions of exceptional mental and physical fitness and exceptional talent; delusions of wealth, aristocratic ancestry, or other grandiose identity; delusions of assistance (i.e., well-placed persons or supernatural powers are assisting their endeavors); or delusions of reference and persecution, based on the belief that enemies are observing or following them out of jealousy at

their special abilities. At the height of mania patients may even see visions or hear voices congruent with their euphoric mood and grandiose self-image (e.g., they might see images of heaven or hear cherubs chanting songs to praise them). The denial characteristic of mania—and the frequently psychotic nature of episodes—means that clinicians must routinely obtain diagnostic information about past episodes from significant others. (Lack of insight also unfortunately means that hospitalization must usually be arranged on an involuntary basis).

MOOD-INCONGRUENT PSYCHOSIS Psychosis in the setting of mania and mixed manic episodes is typically mood congruent. The sense of physical well-being and mental alacrity is so extraordinary that it is understandable why manic patients believe that they possess superior powers or perhaps are great scientists or famous reformers. Moreover, their senses are so vivid that reality appears richer and more exotic, and can be easily transformed into a vision. Likewise, their thoughts are so rapid and vibrant that they feel they can hear them. Thus, certain first-rank schneiderian-type symptoms that have been traditionally considered mood incongruent can be understood phenomenologically to arise from the powerful mental experiences of mania.

A 37-year-old engineer, had experienced three manic episodes for which he had been hospitalized; all three episodes were preceded by several weeks of moderate psychomotor retardation. Although he had responded to lithium (Eskalith, Lithobid) each time, once outside the hospital he had been reluctant to take it and eventually refused to do so. Now that he was euthymic, following his third and most disruptive episode during which he had badly beaten his wife, he could more accurately explain how he felt when manic. Mania, he felt, was “like God implanted in him,” so he could serve as “testimony to man's communication with God.” He elaborated as follows: “Ordinary mortals will never, never understand the supreme manic state which I'm privileged to experience every few years. It is so vivid, so intense, so compelling. When I feel that way, there can be no other explanation: To be manic is, ultimately, to be God. God himself must be supermanic: I can feel it, when mania enters through my left brain like laser beams, transforming my sluggish thoughts, recharging them, galvanizing them. My thoughts acquire such momentum, they rush out of my head, to disseminate knowledge about the true nature of mania to psychiatrists and all others concerned. That's why I will never accept lithium again—to do so is to obstruct the divinity in me.” Although he was on the brink of divorce, he would not yield to his wife's plea to go back on lithium.

The vignette illustrates the possibility that even some of the most psychotic manifestations of mania represent explanatory delusions, the patient's attempt to make sense of the experience of mania. The DSM-IV criteria for severity/psychotic specifiers for manic and mixed episode (Table 14.6-11 and Table 14.6-12) are more concerned with operational rigor than with the phenomenological sophistication needed to understand such core manic experiences. (Many manic patients abuse alcohol and stimulants to enhance their mental state; mood incongruence can sometimes be explained on that basis).

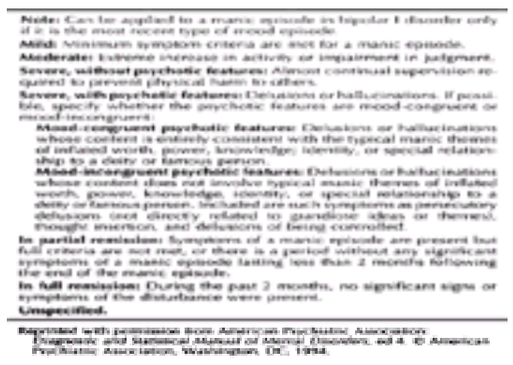

Note: Can be applied to a manic episode in bipolar I disorder only if it is the most recent type of mood episode.

Mild: Minimum symptom criteria are met for a manic episode.

Moderate: Extreme increase in activity or impairment in judgment.

Severe, without psychotic features: Almost continual supervision required to prevent physical harm to others.

Severe, with psychotic features: Delusions or hallucinations. If possible, specify whether the psychotic features are mood-congruent or mood-incongruent:

Mood-congruent psychotic features: Delusions or hallucinations whose content is entirely consistent with the typical manic themes of inflated worth, power, knowledge, identity, or special relationship to a deity or famous person.

Mood-incongruent psychotic features: Delusions or hallucinations whose content does not involve typical manic themes of inflated worth, power, knowledge, identity, or special relationship to a deity or famous person. Included are such symptoms as persecutory delusion (not directly related to grandiose ideas or themes), thought insertion, and delusions of being controlled.

In partial remission: Symptoms of a manic episode are present but full criteria are not met, or there is a period without any significant symptoms of a manic episode lasting less than 2 months following the end of the manic episode.

In full remission: During the past 2 months, no significant signs or symptoms of the disturbance were present.

Unspecified.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 14.6-11 DSM-IV Criteria for Severity/Psychotic/Remission Specifiers for Current (or Most Recent) Manic Episode

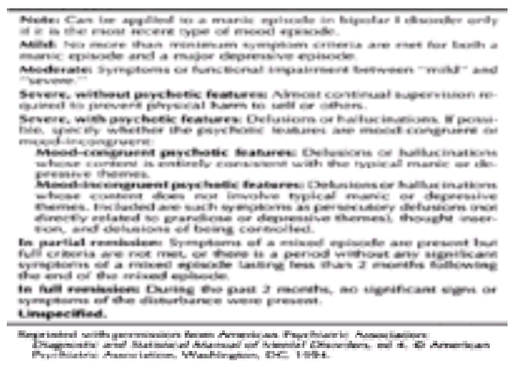

Note: Can be applied to a manic episode in bipolar I disorder only if it is the most recent type of mood episode.

Mild: No more than minimum symptom criteria are met for both a manic episode and a major depressive episode.

Moderate: Symptoms or functional impairment between "mild" and "severe."

Severe, without psychotic features: Almost continual supervision required to prevent physical harm to self or others.

Severe, with psychotic features: Delusions or hallucinations. If possible, specify whether the psychotic features are mood-congruent or mood-incongruent:

Mood-congruent psychotic features: Delusions or hallucinations whose content is entirely consistent with the typical manic or depressive themes.

Mood-incongruent psychotic features: Delusions or hallucinations whose content does not involve typical manic or depressive themes. Included are such symptoms as persecutory delusions (not directly related to grandiose or depressive themes), thought insertion, and delusions of being controlled.

In partial remission: Symptoms of a mixed episode are present but full criteria are not met, or there is a period without any significant symptoms of a mixed episode lasting less than 2 months following the end of the mixed episode.

In full remission: During the past 2 months, no significant signs or symptoms of the disturbance were present.

Unspecified.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 14.6-12 DSM-IV Criteria for Severity/Psychotic/Remission Specifiers for Current (or Most Recent) Mixed Episode

MANIA VERSUS HYPOMANIA Nonpsychotic and nondisruptive variants of mania are much more common and are recognized by DSM-IV as hypomanic episodes. Diagnostically, history of a partial manic syndrome is preferably obtained from significant others who have observed the patient; the experience is often pleasant, and the subject may either be unaware of it or tend to deny it. DSM-IV stipulates a minimum duration of 4 days for hypomania; however, the Memphis and Zurich studies found a modal duration of 2 days. Finally, although DSM-IV states that treatment-emergent hypomania in a depressed patient does not count toward a diagnosis of bipolarity, prospective observations show that nearly all such episodes are followed eventually by spontaneous hypomania (or mania).

DIAGNOSTIC CLASSIFICATION

P>The classification of mood disorders in DSM-IV subsumes a large variety of patients seen in private and public, ambulatory and inpatient settings. The main demarcation in that large clinical terrain is between bipolar and depressive (unipolar) disorders. Thus, bipolar disorders range from the classic manic and depressive episodes of psychotic intensity (bipolar I disorder) through recurrent major depressive episodes, alternating with hypomanic episodes (bipolar II disorder), and cyclothymic mood swings. Likewise, depressive disorders include those with psychotic severity, melancholia, atypical features, and dysthymic variants.

Major and specific attenuated subtypes are distinguished on the basis of severity and duration. In dysthymic and cyclothymic disorders a partial mood syndrome—consisting of such subthreshold features as subdepressive and hypomanic periods—is maintained, intermittently or continuously, for at least 2 years. Subdepressive periods dominate in dysthymia; in cyclothymia, they alternate with hypomania. The onset is typically in adolescence or childhood, and most persons with these diagnoses seen in young adulthood have had low-grade mood symptoms for 5 to 10 years. Major mood disorders, which generally begin much later in life, require the presence of either a full manic episode or a full depressive episode—sustained for at least 1 or 2 weeks, respectively—and an episodic course, typically permitting recovery or remission from episodes. DSM-IV recognizes that a significant minority of persons with major depressive disorders fails to achieve full symptomatic recovery and should thus be qualified as chronic or in partial remission. They are no longer considered dysthymic (the misleading convention in DSM-III).

Dichotomy or Continuum? Although, in the extreme, bipolar and depressive (unipolar) disorders can be discriminated clinically and therapeutically (Table 14.6-13), clinical observations testify to a vast overlap between those extremes. Thus the distinctions between the various affective subtypes are not as hard and fast as DSM-IV attempts to portray. For instance, full-blown bipolar disorder can be superimposed on cyclothymic disorder that tends to persist after the resolution of manic or major depressive episodes. Even more common is major depressive disorder complicating cyclothymic disorder, which should be reclassified as an important course variant of bipolar II disorder. Likewise, recent evidence indicates that dysthymic disorder may precede major depressive disorder in as many as a third of cases. Moreover, one in four persons with major depressive disorder subsequently develops hypomanic or manic episodes and so should be reclassified as having bipolar disorder. Finally, unexpected crossing from dysthymic disorder to hypomanic or manic episodes has also been described, suggesting that some forms of dysthymic disorder are subaffective precursors of bipolar disorder. Such observations are in line with Kraepelin's historic attempt to bring all mood disorders under one rubric. Epidemiological studies in the community have also shown much fluidity between various subthreshold and major mood disorders.

	Bipolar	Unipolar
History of mania or hypomania (definitional)	Yes	No
Temperament/ personality	Cyclothymic/ extroverted	Dysthymic/ introverted
Sex ratio	Equal	More women than men
Age of onset	Teens, 20s, and 30s	30s, 40s, 50s
Postpartum episodes	More common	Less common
Onset of episode	Often abrupt	More insidious
Number of episodes	Numerous	Fewer
Duration of episode	3 to 6 months	3 to 12 months
Psychomotor activity	Retardation > agitation	Agitation > retardation
Sleep	Hypersomnia > insomnia	Insomnia > hypersomnia
Family history		
Bipolar disorder	Yes	∆
Unipolar disorder	Yes	Yes
Alcoholism	∆	Yes
Pharmacological response		
Cyclic antidepressants	Induce hypomania-mania	∆
Lithium carbonate	Acute antidepressant effects	Ineffective

Table 14.6-13 Differentiating Characteristics of Bipolar and Unipolar Depressions

Heterogeneity undoubtedly exists among mood disorders; however, the foregoing observations suggest that much of the unipolar terrain might be “pseudo-unipolar” (i.e., soft bipolar). The clinical significance of these considerations lies in the fact that many DSM-IV subtypes of mood disorders are not pure entities, and considerable overlap and switches in polarity take place. They also provide some rationale, for instance, for why lithium (or lithium augmentation) may be effective in some apparently unipolar depressions; such patients do not experience spontaneous hypomanic episodes, but instead often exhibit a high baseline level of hyperthymic traits. Finally, several studies have shown that bipolar patients with cyclothymic premorbid adjustment and interepisodic adjustment are at considerable risk for antidepressant-induced rapid cycling, defined as a rapid succession of major episodes with few or no intervals of freedom.

Such considerations further testify to the wisdom of supplementing major mood diagnoses with temperamental attributes. DSM-IV only makes subtle or oblique hints concerning this, and instead provides the practitioner with an unwieldly, if not useless, array of episode and course specifiers. The DSM-IV criteria for longitudinal course specifiers are given in Table 14.6-14.

Specify if (can be applied to recurrent major depressive disorder or bipolar I or II disorder):

With full interepisode recovery: if full remission is attained between the two most recent mood episodes

Without full interepisode recovery: if full remission is not attained between the two most recent mood episodes

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 14.6-14 DSM-IV Criteria for Longitudinal Course Specifiers

As Kraepelin illustrated in his monograph, course is best captured graphically. DSM-IV only provides examples of this for depressive disorders (Fig. 14.6-6) and limits itself to four patterns. Kraepelin, after diagramming 18 illustrative patterns for the entire spectrum of manic-depressive illness, declared that the illness pursued an indefinite number of courses.

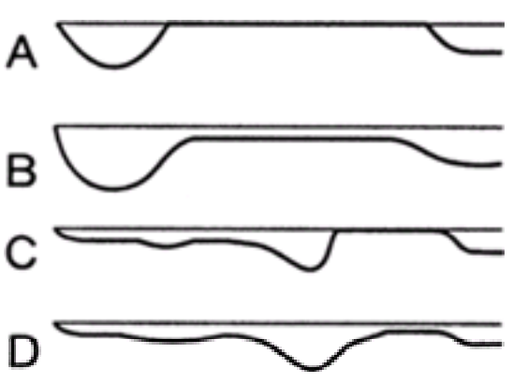

FIGURE 14.6-6 Graphs depicting prototypical courses. **A**, Course of major depressive disorder, recurrent, with no antecedent dysthymic disorder and a period of full remission between the episodes. This pattern predicts the best future prognosis. **B**, Course of major depressive disorder, recurrent, with no antecedent dysthymic disorder but with prominent symptoms persisting between the two most recent episodes (i.e., partial remission is attained). **C**, Rare pattern (present in fewer than 3 percent of persons with major depressive disorder) of major depressive disorder, recurrent with antecedent dysthymic disorder but with full interepisode recovery between the two most recent episodes. **D**, Course of major depressive disorder, recurrent, with antecedent dysthymic disorder and no period of full remission between the two most recent episodes. This pattern, commonly referred to as double depression, is seen in about 20 to 25 percent of persons with major depressive disorder. (Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. Copyright, American Psychiatric Association, Washington, DC, 1994.)

DEPRESSIVE DISORDERS

The broad category of depressive disorders includes major depressive disorder, dysthymic disorder, and depressive disorder not otherwise specified.

Major Depressive Disorder Episodes usually begin over a prodromal period of weeks to months. The DSM-IV diagnosis of major depressive disorder requires (1) dysphoric mood or decreased interest in usual activities and (2) at least four additional classic depressive signs and symptoms, (3) which must be sustained for at least 2 weeks, and (4) cannot be explained by another process known to cause depressive symptoms, such as normal bereavement, certain physical conditions commonly associated with depression, or another mental disorder. It can be single and, more commonly, recurrent. (Table 14.6-15 and Table 14.6-16).

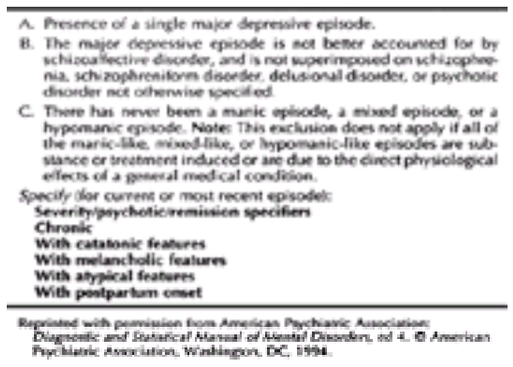

A. Presence of a single major depressive episode.
B. The major depressive episode is not better accounted for by schizoaffective disorder, and is not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.
C. There has never been a manic episode, a mixed episode, or a hypomanic episode. **Note:** This exclusion does not apply if all of the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects of a general medical condition.

Specify (for current or most recent episode):
Severity/psychotic/remission specifiers
Chronic
With catatonic features
With melancholic features
With atypical features
With postpartum onset

Table 14.6-15 DSM-IV Diagnostic Criteria for Major Depressive Disorder, Single Episode

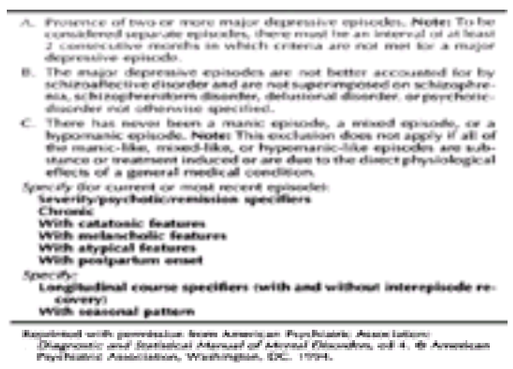

A. Presence of two or more major depressive episodes. **Note:** To be considered separate episodes, there must be an interval of at least 2 consecutive months in which criteria are not met for a major depressive episode.
B. The major depressive episodes are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.
C. There has never been a manic episode, a mixed episode, or a hypomanic episode. **Note:** This exclusion does not apply if all of the manic-like, mixed-like, or hypomanic-like episodes are substance or treatment induced or are due to the direct physiological effects of a general medical condition.

Specify (for current or most recent episode):
Severity/psychotic/remission specifiers
Chronic
With catatonic features
With melancholic features
With atypical features
With postpartum onset

Specify:
Longitudinal course specifiers (with and without interepisode recovery)
With seasonal pattern

Table 14.6-16 DSM-IV Diagnostic Criteria for Major Depressive Disorder, Recurrent

Comorbid Physical Disease Those considerations raise the question whether major depressive disorder should be limited to depressions of unknown etiology (i.e., those without documented physical causes). The DSM-IV approach has basically been that when the cause is known, the condition should be diagnosed as mood disorder due to a general medical condition (Table 14.6-17) which must be specified, or substance-induced mood disorder (Table 14.6-18). The problem with this approach is that many common medical factors historically associated with depression (e.g., use of certain antihypertensive agents) do not seem to be causative in the etiological sense, but rather are triggering agents in otherwise predisposed persons. This is analogous to the situation with life events, which no longer are used in making distinctions between reactive and endogenous subtypes of depression. A more troubling implication is that major depressive disorders without demonstrable physical disease are not medical or otherwise biological. More importantly there appears to be no reliable or valid way for a clinician to decide that a depressive condition is due to a specified medical condition. For this reason it is generally more practical to diagnose the depressive disorder on Axis I and specify the contributing physical condition on Axis III. In brief, the designation “due to a general medical condition” is both cumbersome and redundant. The author considers major depressive disorder to represent the final common pathway of multifactorial interacting factors—both physical and psychological—a syndrome that should be diagnosed irrespective of presumed cause.

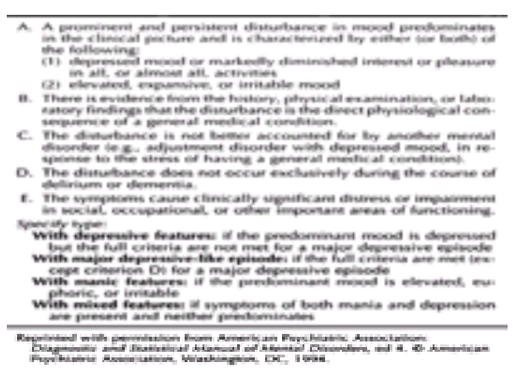

A. A prominent and persistent disturbance in mood predominates in the clinical picture and is characterized by either (or both) of the following:
(1) depressed mood or markedly diminished interest or pleasure in all, or almost all, activities
(2) elevated, expansive, or irritable mood
B. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (e.g., adjustment disorder with depressed mood, in response to the stress of having a general medical condition).
D. The disturbance does not occur exclusively during the course of delirium or dementia.
E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify type:
With depressive features: if the predominant mood is depressed but the full criteria are not met for a major depressive episode
With major depressive–like episode: if the full criteria are met (except criterion D) for a major depressive episode
With manic features: if the predominant mood is elevated, euphoric, or irritable
With mixed features: if symptoms of both mania and depression are present and neither predominates

Table 14.6-17 DSM-IV Diagnostic Criteria for Mood Disorder Due to a General Medical Condition

Table 14.6-18 DSM-IV Diagnostic Criteria for Substance-Induced Mood Disorder

Diagnostic Threshold Another question concerning the DSM-IV definition of major depressive disorders relates to the threshold at which a constellation of depressive features becomes a condition distinct from the ordinary blues. According to the current definition, a person who responds to a setback with lowered spirits and self-doubt, difficulty in sleeping and concentration, and decreased sexual interest for 14 days qualifies for a diagnosis of a major depressive disorder of mild intensity. Many clinicians would consider such a condition a relatively minor departure from normality, probably no more than an adjustment disorder. Obviously, criteria other than signs, symptoms, and duration are necessary to differentiate a depressive disorder from adjustment reactions to life situations. The presence of the following characteristics might assist in such a differentiation.

- By definition, a major depressive disorder should be incapacitating. Previously, much attention was paid to the interpersonal consequences of depression. Recent evidence indicates that measurable deficits in work performance are often early manifestations. Afflicted persons also do not benefit from taking leisure time, and hence prescribing vacations is futile.
- Depressive disorder is usually perceived as a break from a person's usual or premorbid self, which can be so striking that sufferers may feel as though they are losing their minds. The important point is that both the patient and significant others can usually relate the onset of the illness to a given month or quarter of a year, which is not true, for instance, for dysthymic disorder.
- Depressive disorder is often experienced by the sufferer as qualitatively distinct from grief or other understandable reactions to loss or adversity. William James

described it as follows:

There is a pitch of unhappiness so great that the goods of nature may be entirely forgotten, and all sentiment of their existence vanish from the mental field. For this extremity of passion to be reached, something more is needed than adversity; the individual must in his own person become the prey of pathological melancholy. Such sensitiveness and susceptibility of mental pain is a rare occurrence where the nervous constitution is entirely normal: one seldom finds it in a healthy subject even where he is the victim of the most atrocious cruelties of outward fortune; it is an active anguish, a sort of psychical neuraglia wholly unknown to healthy life.

Two additional features, when present, would further validate the diagnosis of major depressive disorder.

- History of past episodes.
- Consecutive-generation family history of mood disorder—especially when a large number of family members are afflicted with depression or mood disorder—is characteristic of clinical depression. For instance, one study that prospectively followed persons with minor or neurotic depression found that such pedigrees predicted the development of future major episodes. DSM-IV makes no provision for considering such familial factors in diagnostic decisions. In clinical practice these factors often strongly influence whether depression is taken seriously.

Single Episode and Recurrent Subtypes About a third of all major depressive episodes do not recur (Table 14.6-15). Such patients tend to be older and less likely to have a positive family history for mood disorders, and have a more protracted (1 to 2 years) course of the disorder. Patients with major depressive disorder, single episode should be distinguished from those experiencing their first episodes of major depressive disorder, recurrent (Table 14.6-16). The latter group tends to be younger, and the disorder is more likely to have been preceded by a depressive temperament or dysthymic disorder.

Research has established that recurrent major depressive disorders are more familial than their single-episode counterparts. The average length of episodes is 6 months, whereas the mean interval between episodes tends to vary (typically years). The mean number of major episodes over a lifetime, according to retrospective and prospective studies, is five to six, in contrast to an average of eight to nine major episodes in bipolar disorder.

Melancholic Features In DSM-III the neurotic-endogenous distinction was deleted. Neurotic depression was largely absorbed by dysthymic disorder and the major depressive disorders that complicate it; endogenous depression became “melancholic features,” a qualifying phrase for major depressive disorders in which anhedonia, guilt, and psychomotor-vegetative disturbances dominate the clinical picture (Table 14.6-4). DSM-IV retains these conventions.

Although the foregoing conventions have received much criticism, they are based on solid data from independent studies in the United States and Germany. Thus neurotic depression, defined as a reactive (i.e., precipitated) nonpsychotic depression of mild to moderate intensity with predominant anxiety and characterologic pathology, does not seem to constitute a distinct nosological entity. Although such a presentation is common in clinical practice, well-conducted studies in the United States and Europe have shown that the prospective follow-up course of those patients is heterogeneous, including melancholic and even psychotic depressions and, in some instances, bipolar transformation. The progression of a precipitated, relatively mild depression (reactive illness) to severe psychotic depression—or one with melancholic autonomy—during prospective observation suggests that so-called endogenous depressions may have their onset in milder depressions, that neurotic and psychotic depressions do not necessarily refer to distinct disorders but to disorders that differ in severity, and that the presence of precipitating stress carries little diagnostic weight in differentiating subtypes of depression (although the absence of such stress might be used to support a melancholic level of major depressive disorder).

At the heart of the concept of morbid depression is its autonomy from stresses that may have precipitated it and its general unresponsiveness to other environmental input. This is embodied in Donald Klein's concept of endogenomorphic depression, which could be precipitated and mild (endoreactive) while exhibiting disturbances of hedonic mechanisms refractory to current interpersonal contexts. Many authorities believe that such features dictate the need to use somatic approaches to reverse the maladaptive autonomy and restore response to interpersonal feedback; that is, psychotherapeutic approaches are deemed largely ineffective until the autonomy is somatically lysed.

Given the somatic connotation of the ancient concept of melancholia, the APA classification has officially adopted it as the preferred nosological term for the revised concept of endogeneity; hence the prominence of the vegetative and biorhythmic features accorded to it in both DSM-III and DSM-IV. However, the APA diagnostic schema risks confusing endogeneity with another classic concept of mood disorder, that of “involutional melancholia.”

Psychotic Features About 15 percent of major depressive disorders, usually from the rank of those with melancholic features, develop into delusional depressions. In young persons they tend to be retarded, even stuporous, and are best considered initial episodes of a bipolar disorder. More typically, psychotic depression that develops for the first time after the age of 50 often presents with severe agitation, delusional guilt, hypochondriacal preoccupations, early-morning awakening, and weight loss. The premorbid adjustment of such patients is classically characterized as “obsessoid.” Their mournful-anxious mood and agitation are autonomous, being refractory to psychological interventions, and they endure great psychic suffering. Except for the fact that generally one to two episodes occur in late-onset (so-called involutional) depressions, they represent a severe variant of DSM-IV melancholia. Kraepelin's postulation of a cerebrovascular basis for such cases makes the ventricular enlargement and white matter opacities reported in psychotic depressions of some interest. Their etiological specificity for persons with late-onset psychotic depression has been controversial, however, since younger (more bipolar) persons with psychotic depression exhibit similar findings. Brain imaging findings tend to be correlated with the neurocognitive deficits observed in psychotic depressions. Those features do not seem to define a distinct depressive subtype, but one of greater severity. Finally, despite attempts to suggest a neurochemical uniqueness based largely on the need for antipsychotic treatment in the acute phase of many of those patients, familial and other external validators have failed to support psychotic depression as a separate entity; hence the decision in DSM-IV to use psychotic features merely as a specifier for major depressive episode (Table 14.6-3). Emerging data, nonetheless, might eventually force a change in this convention. For instance, William Coryell and collaborators in the NIMH collaborative study of depression have shown psychotic depression to be the most consistent unipolar subtype across episodes. Alan Schatzberg's work, originally conducted at Harvard, likewise underscores the uniqueness of psychotic depression based on neuroendocrine and putative neurochemical considerations.

Chronic Depression The DSM-IV criteria for chronic specifier appear in Table 14.6-19. The clinical situation, however, is much more complex than these conventions. For instance, the symptom profile in chronic depressions usually displays low-grade intensity rather than severe syndromal chronicity. Severe depressive disorder in its psychotic forms is so agonizing that the sufferer is at risk of committing suicide before the disorder has a chance to become chronic. More commonly, the psychotic symptoms respond to medication or to electroconvulsive therapy (ECT), but residual depressive symptoms may linger for a long time. In other persons with chronic depressions the chronicity arises from more mundane (nonpsychotic) major depressive episodes, depressive residua following one or several clinical episodes that fail to remit fully. Instead of the customary remission within a year, the patients are ill for years. The level of depression varies, fluctuating between syndromal illness and milder symptoms. The patients often show a sense of resignation, generalized fear of an inability to cope, adherence to rigid routines, and inhibited communication.

Specify if:

Chronic (can be applied to the current or most recent major depressive episode in major depressive disorder and to a major depressive episode in bipolar I or II disorder only if it is the most recent type of mood episode)

Full criteria for a major depressive episode have been met continuously for at least the past 2 years

Table 14.6-19 DSM-IV Diagnostic Criteria for Chronic Specifier

Rather than exhibiting a frankly depressive mood, many persons with chronic depression suffer from deficits in their ability to enjoy leisure and display an attitude of

irritable moroseness. The leisure deficits and irritable humor tend to affect their conjugal lives: their marriages are typically in a state of chronic deadlock, leading neither to divorce nor to reconciliation. In other patients the residual phase is dominated by somatic features, such as sleep and other vegetative or autonomic irregularities. Thus, self-treatment with ethanol or iatrogenic benzodiazepine dependence is common. That these interpersonal, conjugal, and autonomic manifestations represent unresolved depression is shown by persistent sleep EEG (especially REM and delta phase) abnormalities that are indistinguishable from their acute counterparts.

Failure to recover from major depressive disorder is associated with increased familial loading for depression, disabled spouses, deaths of immediate family members, concurrent disabling medical disease, use of depressant pharmacologic agents, and excessive use of alcohol and sedative-hypnotic agents. Social support is often eroded in persons with residual depression, through either the death or illness of significant others. Therefore, a thorough medical evaluation and socially supportive interventions should be essential ingredients of the overall approach to those patients.

Interpersonal disturbances in such patients are usually secondary to the distortions produced by long-standing depression. Therefore, observed pathological characterological changes—clinging or hostile dependence, demandingness, touchiness, pessimism, and low self-esteem—are best considered as “postdepressive personality” changes. A dangerous stereotypical thinking holds that because a patient has not responded adequately to standard treatments (the illness has become chronic), the disorder must have a characterological substrate. The long duration of the disorder often leads the patient to identify with the failing functions of depression, producing the self-image of being a depressed person. This self-image itself represents a malignant cognitive manifestation of the depressive disorder and dictates vigorous treatment targeted at the mood disorder.

Dysthymic Disorder Dysthymic disorder (Table 14.6-20) is distinguished from chronic depressive disorder by the fact that it is not a sequel to well-defined major depressive episodes. Instead, in the most typical cases, patients complain that they have always been depressed. Thus, most cases are of early onset, beginning in childhood or adolescence and certainly by the time patients reach their 20s. A late-onset subtype, much less prevalent and not well characterized clinically, has been identified among middle-aged and geriatric populations, largely through epidemiological studies in the community.

Table 14.6-20 DSM-IV Diagnostic Criteria for Dysthymic Disorder

Although the dysthymic disorder category in DSM-IV can occur as a secondary complication of other psychiatric disorders, the core concept of dysthymic disorder refers to a subaffective disorder with (1) low-grade chronicity for at least 2 years, (2) insidious onset with origin often in childhood or adolescence, and (3) persistent or intermittent course. Although not part of the formal definition of dysthymic disorder, the family history is typically replete with both depressive and bipolar disorders, which is one of the more robust findings supporting its link to primary mood disorder.

Social Adjustment Dysthymic disorder is typically an ambulatory disorder compatible with relatively stable social functioning. However, the stability is precarious; recent data document that many patients invest whatever energy they have in work, leaving none for leisure and family or social activities, which results in marital friction. These empirical findings on the work orientation of persons with dysthymic disorder echo earlier formulations in the German and Japanese literature. For instance, Kraepelin described such persons as follows: “Life with its activity is a burden which they habitually bear with dutiful self-denial without being compensated by the pleasure(s) of existence.”

The dedication of persons with dysthymic disorder to work has been suggested to be an overcompensation and a defense against their battle with depressive disorganization and inertia. Nevertheless, Kretschmer suggested that such persons are the “backbone of society,” dedicating their lives to jobs that require dependability and great attention to detail. Epidemiological studies have demonstrated that some persons with protracted dysthymic complaints, extending over many years, have never experienced clear-cut depressive episodes. Some of them may seek outpatient counseling and psychotherapy for what some clinicians might consider “existential depression,” with feelings of being empty and lacking any joy in life outside their work. Such persons have been described as leading monocategorical existences. Others present clinically because their low-grade dysphoria has intensified into a major depression disorder.

Course An insidious onset of depression dating back to late childhood or the teens, preceding any superimposed major depressive episodes by years or even decades, represents the most typical developmental background of dysthymic disorder. A return to the low-grade depressive pattern is the rule following recovery from superimposed major depressive episodes, if any; hence the designation “double depression” as a prominent course pattern illustrated in DSM-IV for depressive illness (Fig. 14.6-6). This pattern, commonly seen in clinical practice, consists of the baseline dysthymic disorder fluctuating in and out of depressive episodes. The more prototypical patients with dysthymic disorder often complain of having been depressed since birth or of feeling depressed all the time. They seem, in the apt words of Kurt Schneider, to view themselves as belonging to an “aristocracy of suffering.” Such descriptions of chronic gloominess in the absence of more objective signs of depression earn such patients the label of “characterological depression.” The description is further reinforced by the fluctuating depressive picture that merges imperceptibly with the patient's habitual self and thus raises uncertainty as to whether dysthymic disorder belongs in Axis I or Axis II.

Clinical Picture The profile of dysthymic disorder overlaps with that of major depressive disorder but differs from it in that symptoms tend to outnumber signs (more subjective than objective depression). This means that marked disturbances in appetite and libido are uncharacteristic, and psychomotor agitation or retardation is not observed. This all translates into a depression with attenuated symptomatology. However, subtle endogenous features are not uncommonly observed: inertia, lethargy, and anhedonia that are characteristically worse in the morning. Because patients presenting clinically often fluctuate in and out of a major depression, the core DSM-IV criteria for dysthymic disorder tend to emphasize vegetative dysfunction, whereas the alternative criterion B for dysthymic disorder (Table 14.6-21) in a DSM-IV appendix lists cognitive symptoms.

B. Presence, while depressed, of three (or more) of the following:
(1) low self-esteem or self-confidence, or feelings of inadequacy
(2) feelings of pessimism, despair, or hopelessness
(3) generalized loss of interest or pleasure
(4) social withdrawal
(5) chronic fatigue or tiredness
(6) feelings of guilt, brooding about the past
(7) subjective feelings of irritability or excessive anger
(8) decreased activity, effectiveness, or productivity
(9) difficulty in thinking, reflected by poor concentration, poor memory, or indecisiveness

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 14.6-21 DSM-IV Alternative Research Criterion B for Dysthymic Disorder

Although dysthymic disorder represents a more restricted concept than its parent, neurotic depression, it is still quite heterogeneous. Anxiety is not a necessary part of its clinical picture, yet dysthymic disorder is often diagnosed in patients with anxiety and neurotic disorders. That clinical situation is perhaps to be regarded as a secondary or “anxious dysthymia” or, in the framework of Peter Tyrer, as part of a “general neurotic syndrome.” For greater operational clarity it is best to restrict dysthymic disorder to a primary disorder, one that cannot be explained by another psychiatric disorder. The essential features of such primary dysthymic disorder include habitual gloom, brooding, lack of joy in life, and preoccupation with inadequacy. Dysthymic disorder then is best characterized as long-standing fluctuating low-grade depression, experienced as part of the habitual self and representing an accentuation of traits observed in the depressive temperament (Table 14.6-1). Dysthymia then can be viewed as a more symptomatic form of that temperament (introduced in a DSM-IV appendix as a depressive personality disorder). Sleep EEG data indicate that many persons with dysthymic disorder at baseline exhibit the sleep patterns of those with acute major depressive disorder, providing support for the constitutional nature of the disorder. Further evidence for that position comes from studies demonstrating high rates of familial affective disorder in dysthymic disorder, depressive temperament, or both.

The clinical picture of dysthymic disorder that emerges from the foregoing description is quite varied, with some patients proceeding to major depression, while others manifest the pathology largely at the personality level. The foregoing considerations suggest that a clinically satisfactory operationalization of dysthymia must include symptomatic, cognitive, and trait characteristics.

A 27-year-old, male, grade-school teacher presented with the chief complaint that life was a painful duty that had always lacked luster for him. He said he felt enveloped by a sense of gloom that was nearly always with him. Although he was respected by his peers, he felt “like a grotesque failure, a self-concept I have had since childhood.” He stated that he merely performed his responsibilities as a teacher and that he had never derived any pleasure from anything he had done in life. He said he had never had any romantic feelings; sexual activity, in which he had engaged with two different women, had involved pleasureless orgasm. He said he felt empty, going through life without any sense of direction, ambition, or passion, a realization that itself was tormenting. He had bought a pistol to put an end to what he called his “useless existence” but did not carry out suicide, believing that it would hurt his students and the small community in which he lived.

Dysthymic Variants Dysthymia is not uncommon in patients with chronically disabling physical disorders, particularly among elderly adults. Dysthymia-like clinically significant subthreshold depression lasting 6 or more months has also been described in neurological conditions, including stroke. According to a recent WHO conference, this condition aggravates the prognosis of the underlying neurological disease and, therefore, deserves pharmacotherapy. Ongoing studies should provide more explicit clinical recommendations on this topic.

Prospective studies on children have revealed an episodic course of dysthymia with remissions, exacerbations, and eventual complications by major depressive episodes, 15 to 20 percent of which might even progress to hypomanic, manic, or mixed episodes postpuberty. Persons with dysthymic disorder presenting clinically as adults tend to pursue a chronic unipolar course that may or may not be complicated by major depression. They rarely develop spontaneous hypomania or mania. However, when treated with antidepressants, some of them may develop brief hypomanic switches that typically disappear when the antidepressant dose is decreased. Although DSM-IV would not allow the occurrence of such switches in dysthymia, systematic clinical observation has verified their occurrence in as many as a third of dysthymic patients. In this special subgroup of persons with dysthymic disorder, the family histories are often positive for bipolar disorder. Such patients represent a clinical bridge between depressive disorder and bipolar II disorders.

Depressive Disorder Not Otherwise Specified The DSM-IV criteria for depressive disorder not otherwise specified, are presented in Table 14.6-22. What follows are descriptions of conditions that are commonly used in the epidemiological, clinical, or pharmacological literature but do not easily fit into the official nosology of depressive disorders. Some represent complex interweaving of depression with personality constructs. For instance, community studies have revealed a prevalent pattern of intermittent depressive manifestations with brief episodes, below the 2-week duration threshold for major depressive disorder. In so-called minor depressive disorder (Table 14.6-23), observed in primary care settings, the depression is subthreshold, milder than major depression and yet not protracted enough to be considered dysthymic. These varied manifestations of depression argue for a continuum model (Fig. 14.6-7) as originally envisaged by Kraepelin. Lewis Judd and collaborators at the University of California at San Diego have suggested that subthreshold depressive symptoms—without necessarily meeting the criterion for mood change—might actually represent the most common expressions of a depressive diathesis. From such a subsyndromal symptomatic depressive base, individuals predisposed to depressive illness are said to fluctuate in and out of the various DSM-IV and subthreshold subtypes of depressive disorders. This viewpoint is presently most cogent for subsyndromal symptomatic depression that follows major depressive disorder, a strong predictor of subsequent frequent relapse or chronic course. There is an important message for the clinician here: treat subsyndromal symptomatic depression residual to major depressive disorder.

The depressive disorder not otherwise specified category includes disorders with depressive features that do not meet the criteria for major depressive disorder, dysthymic disorder, adjustment disorder with depressed mood, or adjustment disorder with mixed anxiety and depressed mood. Sometimes depressive symptoms can present as part of an anxiety disorder not otherwise specified. Examples of depressive disorder not otherwise specified include:

1. Premenstrual dysphoric disorder: in most menstrual cycles during the past year, symptoms (e.g., markedly depressed mood, marked anxiety, marked affective lability, decreased interest in activities) regularly occurred during the last week of the luteal phase (and remitted within a few days of the onset of menses). These symptoms must be severe enough to markedly interfere with work, school, or usual activities and be entirely absent for at least 1 week postmenses.
2. Minor depressive disorder: episodes of at least 2 weeks of depressive symptoms but with fewer than the five items required for major depressive disorder.
3. Recurrent brief depressive disorder: depressive episodes lasting from 2 days up to 2 weeks, occurring at least once a month for 12 months (not associated with the menstrual cycle).
4. Postpsychotic depressive disorder of schizophrenia: a major depressive episode that occurs during the residual phase of schizophrenia.
5. A major depressive episode superimposed on delusional disorder, psychotic disorder not otherwise specified, or the active phase of schizophrenia.
6. Situations in which the clinician has concluded that a depressive disorder is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 14.6-22 DSM-IV Diagnostic Criteria for Depressive Disorder Not Otherwise Specified

Table 14.6-23 DSM-IV Research Criteria for Minor Depressive Disorder

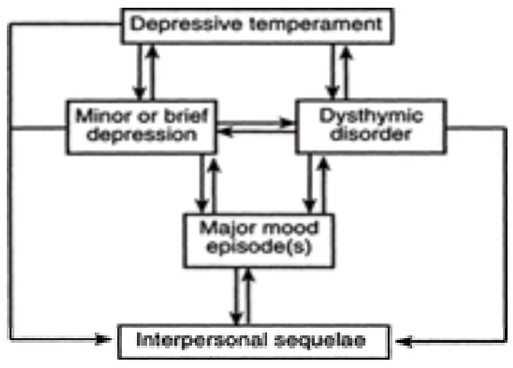

FIGURE 14.6-7 Relation of various depressive conditions supporting a spectrum concept. (Reprinted with permission from Akiskal HS: Dysthymia: Clinical and external validity. Acta Psychiatr Scand *89*(Suppl):19, 1994.)

Recurrent Brief Depressive Disorder Now in a DSM-IV appendix (Table 14.6-24), recurrent brief depressive disorder derives from British work on young adults with frequent suicide attempts and epidemiological studies conducted in a young adult cohort in Zurich. It is described as short-lived depressions that usually recur on a monthly basis but are not menstrually related. They could coexist with major depressive disorder and dysthymic disorder. Such patients are believed to be more prevalent in primary care than in psychiatric settings. Those seen in psychiatric settings are likely to be given Axis II diagnoses such as borderline personality disorder.

A. Criteria, except for duration, are met for a major depressive episode.
B. The depressive periods in criterion A last at least 2 days but less than 2 weeks.
C. The depressive periods occur at least once a month for 12 consecutive months and are not associated with the menstrual cycle.
D. The periods of depressed mood cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hypothyroidism).
F. There has never been a major depressive episode, and criteria are not met for dysthymic disorder.
G. There has never been a manic episode, a mixed episode, or a hypomanic episode, and criteria are not met for cyclothymic disorder. **Note:** This exclusion does not apply if all of the manic-, mixed-, or hypomanic-like episodes are substance or treatment induced.
H. The mood disturbance does not occur exclusively during schizophrenia, schizophreniform disorder, schizoaffective disorder, delusional disorder, or psychotic disorder not otherwise specified.

Table 14.6-24 DSM-IV Research Criteria for Recurrent Brief Depressive Disorder

The current nosological status of those patients is uncertain, but they testify to Kraepelin's observation that many transitional forms link the depressive temperament to affective episodes:

> A permanent gloomy stress in all the experiences of life usually perceptible already in youth, and may persist without essential change throughout the whole of life (or) there is actually an uninterrupted series of transitions to periodic melancholia in which the course is quite indefinite with irregular fluctuations and remissions.

Reactive Depression Classically, reactive depression is defined as resulting from a specific life event. In an ideal case the depression would not have occurred without the event (e.g., love loss) to which it is a reaction. It continues as long as the event is present, and it terminates with the reversal of the event (e.g., return of the lover). Depressions exhibiting all of those features are almost never seen in clinical practice. With interpersonal support most people can face life's reverses, which explains why reactive depression tends to be self-limiting. Hence, adjustment disorder is the more appropriate diagnosis for most cases of reactive depression.

Conceptually, however, one can envision chronically unsatisfactory life situations that might lead to chronic demoralization. However, such a condition, which could warrant the designation of chronic reactive depression, is a contradiction in terms. The question often raised is why a person would continue to stay in the situation. Sometimes psychodynamic authors invoke the concept of masochism to explain why certain persons cannot rid themselves of painful life situations, implying that they somehow contribute to their maintenance. Current thinking is that some of those presumed self-defeating traits are more situation specific than previously believed and might resolve with the elimination of the situation. So-called self-defeating features then are best considered psychodynamic mechanisms rather than indicators of a specific personality. At the present stage of knowledge, they do not deserve to be raised to the level of a nosological entity (hence, their disappearance from DSM-IV). Chronic adjustment disorder might describe the chronic demoralization observed among some individuals stuck in chronically unsatisfactory life situations. Many more might fulfill the criteria for dysthymia.

Neurasthenia A century-old term developed by the American neuropsychiatrist George Beard, neurasthenia refers to a more chronic stage of anxious-depressive symptomatology. The anxiety generated by overstimulation is so excessive that it is replaced by a chronic disposition to irritability, fatigue (especially mental fatigue), lethargy, and exhaustion. It is as if the sufferer's mind refuses to take on new stresses. The clinical picture described by Beard suggests that anxious manifestations were preeminent in his time. They included headache, scalp tenderness, backache, heavy limbs, vague neuralgias, yawning, dyspepsia, palpitations, sweating hands and feet, chills, flushing, sensitivity to weather changes, insomnia, nightmares, pantaphobia, asthenopia, and tinnitus.

Although the diagnosis of neurasthenia is now used more in China than in the rest of the world, the recent worldwide popularity of the concept of chronic fatigue syndrome attests to the clinical acumen of classic physicians. Despite much energy invested in finding a viral or immunological cause, current descriptions tend to suggest an anxiety or mood disorder basis for some (but not all) of those with the syndrome. However, what circumstances would lead anxiety or depression to manifest primarily in fatigue is as elusive as it was 100 years ago. Like many other patients presenting to primary care settings with somatic complaints, those with chronic fatigue tend to denounce psychiatric diagnoses as inadequate explanations for their ills.

Postpsychotic Depressive Disorder of Schizophrenia DSM-IV describes postpsychotic depressive disorder of schizophrenia as follows:

> The essential feature is a Major Depressive Episode that is superimposed on, and occurs only during, the residual phase of Schizophrenia. The residual phase of Schizophrenia follows the active phase (i.e., symptoms meeting Criterion A) of Schizophrenia. It is characterized by the persistence of negative symptoms or of active-phase symptoms that are in an attenuated form (e.g., odd beliefs, unusual perceptual experiences). The superimposed Major Depressive Episode must include depressed mood (i.e., loss of interest or pleasure cannot serve as an alternate for sad or depressed mood). Most typically, the Major Depressive Episode follows immediately after remission of the active-phase symptoms of the psychotic episode. Sometimes it may follow after a short or extended interval during which there are no psychotic symptoms. Mood symptoms due to the direct physiological effects of a drug of abuse, a medication, or a general medical condition are not counted toward postpsychotic depressive disorder of Schizophrenia.

According to DSM-IV, persons whose presentation meets those research criteria (Table 14.6-25) would be diagnosed as having depressive disorder not otherwise specified. As already pointed out, mood or depressive disorder not otherwise specified represents such a hodgepodge of clinical situations that the designation not otherwise specified is at best meaningless and at worst countertherapeutic. In all postpsychotic depressions, one must first exclude a missed bipolar diagnosis. Negative symptoms due to classic antipsychotics—especially depot phenothiazines and those due to the residium of schizophrenia once positive symptoms are brought under control—should be distinguished from the depressive episodes that complicate the course of schizophrenia in young, intelligent patients.

A. Criteria are met for a major depressive episode. **Note:** The major depressive episode must include criterion A1: depressed mood. Do not include symptoms that are better accounted for as medication side effects or negative symptoms of schizophrenia.
B. The major depressive episode is superimposed on and occurs only during the residual phase of schizophrenia.
C. The major depressive episode is not due to the direct physiological effects of a substance or a general medical condition.

Table 14.6-25 DSM-IV Research Criteria for Postpsychotic Depressive Disorder of Schizophrenia

BIPOLAR DISORDERS

Four bipolar disorders are included in DSM-IV: bipolar I disorder, bipolar II disorder, cyclothymic disorder, and bipolar disorder not otherwise specified.

Bipolar I Disorder Typically beginning in the teenage years, the 20s, or the 30s, the first episode could be manic, depressive, or mixed. One common mode of onset is mild retarded depression, or hypersomnia, for a few weeks or months, which then switches into a manic episode. Others begin with a severely psychotic manic episode with schizophreniform features; only when a more classic manic episode occurs is the affective nature of the disorder clarified. In a third group several depressive episodes take place before the first manic episode. A careful history taken from significant others often reveals dysthymic or cyclothymic traits that antedated the frank onset of manic episodes by several years.

According to DSM-IV, bipolar I disorder, single manic episode (Table 14.6-26) describes patients having a first episode of mania (most such patients eventually develop depressive episodes). The remaining subcategorization is used to specify the nature of the current or most recent episode in patients who have had recurrent mood episodes (Table 14.6-27, Table 14.6-28, Table 14.6-29, Table 14.6-30 and Table 14.6-31. For clinicians and researchers alike it is more meaningful to chart a patient's course in color over time—for example using red rectangles for manic, blue for depressive, and violet for mixed episodes, with hypomanic, dysthymic, and cyclothymic periods drawn in the appropriate colors on a smaller scale between the major episodes. Life events, biologic stressors, and treatment can be indicated by arrows on the time axis. This approach, originally championed by Kraepelin, is routinely used in mood clinics. Robert Post at the NIMH has developed this approach into systematic clinical science.

A. Presence of only one manic episode and no past major depressive episodes.
Note: Recurrence is defined as either a change in polarity from depression or an interval of at least 2 months without manic symptoms.
B. The manic episode is not better accounted for by schizoaffective disorder, and is not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.

Specify if:
Mixed: if symptoms meet criteria for a mixed episode

Specify (for current or most recent episode):
Severity/psychotic/remission specifiers
With catatonic features
With postpartum onset

Table 14.6-26 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Single Manic Episode

A. Currently (or most recently) in a manic episode.
B. There has previously been at least one major depressive episode, manic episode, or mixed episode.
C. The mood episodes in criteria A and B are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.

Specify (for current or most recent episode):
Severity/psychotic remission specifiers
With catatonic features
With postpartum onset

Specify:
Longitudinal course specifiers (with and without interepisode recovery)
With seasonal pattern (applies only to the pattern of major depressive episodes)
With rapid cycling

Table 14.6-27 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Most Recent Episode Manic

A. Currently (or most recently) in a hypomanic episode.
B. There has previously been at least one manic episode or mixed episode.
C. The mood symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The mood episodes in criteria A and B are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.

Specify:
Longitudinal course specifiers (with and without interepisode recovery)
With seasonal pattern (applies only to the pattern of major depressive episodes)
With rapid cycling

Table 14.6-28 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Most Recent Episode Hypomanic

A. Currently (or most recently) in a mixed episode.
B. There has previously been at least one major depressive episode, manic episode, or mixed episode.
C. The mood episodes in criteria A and B are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.

Specify (for current or most recent episode):
Severity/psychotic remission specifiers
With catatonic features
With postpartum onset

Specify:
Longitudinal course specifiers (with and without interepisode recovery)
With seasonal pattern (applies only to the pattern of major depressive episodes)
With rapid cycling

Table 14.6-29 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Most Recent Episode Mixed

A. Currently (or most recently) in a major depressive episode.
B. There has previously been at least one manic episode or mixed episode.
C. The mood episodes in criteria A and B are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.

Specify (for current or most recent episode):
Severity/psychotic remission specifiers
Chronic
With catatonic features
With melancholic features

Table 14.6-30 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Most Recent Episode Depressed

A. Criteria, except for duration, are currently (or most recently) met for a manic, a hypomanic, a mixed, or a major depressive episode.
B. There has previously been at least one manic episode or mixed episode.
C. The mood symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The mood symptoms in criteria A and B are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.
E. The mood symptoms in criteria A and B are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication, or other treatment) or a general medical condition (e.g., hyperthyroidism).

Specify:
Longitudinal course specifiers (with and without interepisode recovery)
With seasonal pattern (applies only to the pattern of major depressive episodes)
With rapid cycling

Table 14.6-31 DSM-IV Diagnostic Criteria for Bipolar I Disorder, Most Recent Episode Unspecified

On average, manic episodes predominate in youth, and depressive episodes in later years. Although the overall sex ratio is about one to one, men on average undergo more manic episodes and women experience more mixed and depressive episodes. Bipolar I disorder in children is not as rare as previously thought; however, most reported cases are in boys, and mixed-manic (dysphoric-explosive) presentations are the mode. Childhood-onset depression must also be considered a major risk for ultimate bipolar transformation. This is based on the following characteristics: (1) early age of onset; (2) even sex ratio; (3) prominence of irritability, labile moods, and explosive anger, suggesting mixed episodes; (4) questionable response to antidepressants, hypomanic switches, or both; (5) high recurrence rate after depression; and (6) familial affective loading. Mania can also first appear after age 65, though a diligent search often reveals a past mild, forgotten, or untreated depressive episode in earlier years.

Acute Mania Mania typically escalates over a period of 1 to 2 weeks; more-sudden onsets have also been described. The DSM-IV criteria (Table 14.6-7) stipulate (1) a distinct period that represents a break from premorbid functioning, (2) a duration of at least 1 week, (3) an elevated or irritable mood, (4) at least three to four classic manic signs and symptoms, and (5) the absence of any physical factors that could account for the clinical picture. The irritable mood in mania can deteriorate to cantankerous behavior, especially when the person is rebuffed. Such patients are among the most aggressive seen in the emergency room. Florid grandiose psychosis with paranoid features, a common presentation of mania, further contributes to the aggression. Alcohol use, observed in at least 50 percent of bipolar I patients (often during the manic phase), further disinhibits the patient and might lead to a dangerous frenzy. Such patients may attack loved ones and hurt them physically. So-called crimes of passion have been committed by patients harboring delusions of infidelity on the part of spouses or lovers, usually when under the influence of alcohol.

The genesis of delusional, hallucinatory, even first-rank, psychotic experiences in mania has been described. Recent research has documented that most types of formal thought disorders are common to both schizophrenic and mood psychoses; only poverty of speech content (vagueness) emerges as significantly more common in schizophrenia. Finally, posturing and negativism occur in mania (and, in the author's view, do not warrant the designation of catatonic features as advocated by DSM-IV). Although not specifically mentioned in the DSM-IV definition, confusion, even pseudodemented presentations, can occur in mania. Mania is most commonly expressed as a phase of bipolar type I disorder, which has strong genetic determinants. Available evidence does not permit separating recurrent mania without depressive episodes from that type as a distinct nosological entity.

Secondary Mania Although there is some suggestion that postpartum mania without depression is distinct from familial bipolar I disorder (in which depressive, manic, and especially mixed manic episodes occur in the postpartum period), the evidence for a distinct puerperal mania is not compelling at this time (hence the decision in DSM-IV to include the postpartum-onset specifier [see Table 13.4-3], rather than a separate mood disorder diagnosis). Mania without prior bipolarity can arise in the setting of such somatic illnesses as thyrotoxicosis, systemic lupus erythematosus or its treatment with steroids, rheumatic chorea, multiple sclerosis, Huntington's disease, cerebrovascular disorder, diencephalic and third ventricular tumors, head trauma, complex partial seizures, syphilis, and (most recently) AIDS. The family history is reportedly low in such cases, suggesting a relatively low genetic predisposition and thus a lower risk of recurrence. These patients do not easily fit into the DSM-IV category of mood disorder due to a general medical condition (Table 14.6-17) because most of the conditions appear to be cerebral. Such factors must always be diligently sought in manias of late life.

Less well defined forms of mania are the so-called reactive manias. Personal loss and bereavement are hypothesized to be triggering factors, and the reaction is conceptualized psychodynamically as a denial of loss. Although such explanations may be plausible in individual cases, no systematic data suggest that these patients differ in family history from persons with other manias. The same is generally true for depressed patients who switch to hypomania or mania after abuse of stimulant drugs, treatment with antidepressants, or sleep deprivation. In all of these situations a bipolar diathesis is usually manifest either in a family history of mania or in spontaneous excited episodes during prospective observation. First-onset manic episodes can also occur in persons who abruptly abstain from alcohol after one or more decades of chronic use and then develop classic bipolar I disorder.

Chronic Mania DSM-IV does not specifically address the diagnostic questions posed by the 5 percent of bipolar I patients who have a chronic manic course. These cases commonly represent deterioration of course dominated by recurrent manic episodes grafted on a hyperthymic baseline. Noncompliance with pharmacological treatment is the rule. Recurrent excitement is personally reinforcing, subjective distress is minimal, and insight is seriously impaired. Thus the patient sees no reason to adhere to treatment. Episodic or chronic alcohol abuse, prevalent in such patients, has been suggested as a contributory cause of the chronicity. Some authorities

further consider comorbid cerebral pathology responsible for nonrecovery (and increased mortality) from manic excitements occurring in late life.

Grandiose delusions (e.g., delusions of inventive genius or aristocratic birth) are not uncommon in chronic mania and may lead to the mistaken diagnosis of paranoid schizophrenia. Because of their social deterioration, Kraepelin subsumed such patients under the category “manic dementia.” Organic factors such as head trauma and chronic alcohol abuse may contribute to the deterioration. Nonschizoid premorbid adjustment, a family history of bipolar I disorder, and the absence of flagrant formal thought disorder can be marshaled in establishing the affective basis of these poor-prognosis manic states.

Bipolar Mixed Phase Momentary tearfulness and even depressed mood are commonly observed at the height of mania or during the transition from mania to retarded depression. These transient labile periods, which occur in most bipolar I disorder patients, must be contrasted with mixed episodes proper.

The latter, variously referred to as “mixed mania” or “dysphoric mania,” are characterized by dysphorically excited moods, irritability, anger, panic attacks, pressured speech, agitation, suicidal ideation, severe insomnia, grandiosity, and hypersexuality, as well as persecutory delusions and confusion. Severely psychotic mixed states that involved hallucinations and schneiderian symptoms risk being labeled “schizoaffective.” A correct diagnosis is mandatory because conventional antipsychotic drugs tend to exacerbate the depressive component and failure to use mood stabilizers can prolong the patient's misery.

New research data from mood centers worldwide on mixed mania suggest that dysphoric mania—mania and full-blown depression occurring simultaneously—is relatively uncommon. Two to four depressive symptoms from the list of depressed mood, helplessness, hopelessness, fatigue, anhedonia, guilt, and suicidal ideation, or impulses, or both, in the setting of a manic syndrome, appear to suffice for the diagnosis of mixed manic states, which occurs in 50 percent of patients with bipolar disorder sometime during their lives. Mixed states occur predominantly in females in whom mania is superimposed on a depressive temperament or a dysthymic baseline. These considerations suggest that the DSM-IV concept of mixed episode (Table 14.6-8) as a cross-sectional mixture of mania and depression is simplistic and phenomenologically naive. The emerging conceptualization of mixed mania is a manic state intruding upon long-term depressive traits.

Depressive Phase Psychomotor retardation, with or without hypersomnia, marks the uncomplicated depressive phase of bipolar I disorder. Onset and offset are often abrupt, though onset can also occur gradually over several weeks. Patients may recover into a free interval or switch directly into mania. Switching into an excited phase is particularly likely when antidepressants have been used. However, not all patients develop mania after antidepressant treatment of bipolar depression. Some develop a mixed agitated depression; indeed, patients may be stuck for many months in a severe depressive phase with some manic admixtures such as racing thoughts and sexual arousal. DSM-IV does not specifically recognize a mixed depressive phase with few manic symptoms occurring during full-blown depression. Such recognition is necessary because these patients don't need continued aggressive antidepressant therapy but mood stabilizers, ECT, or both.

Delusional and hallucinatory experiences are less common in the depressive phase of bipolar I disorder than in the manic and mixed manic phases. Stupor is the more common psychotic presentation of bipolar depression, particularly in adolescents and young adults. Pseudodemented organic presentations appear to be the counterpart of stupor in elderly adults.

Cyclothymic Disorder An attenuated bipolar disorder that typically begins insidiously before the age of 21, cyclothymic disorder is characterized in DSM-IV by frequent short cycles of subsyndromal depression and hypomania (Table 14.6-32). The author's research has revealed alternating patterns of moods, activity, and cognition (Table 14.6-33), which are more explicit than the DSM-IV criteria. The course of cyclothymia is continuous or intermittent, with infrequent periods of euthymia. Shifts in mood often lack adequate precipitants (e.g., sudden profound dejection with social withdrawal for a few days switching into cheerful, gregarious behavior). Circadian factors may account for some of the extremes of emotional lability, such as the person's going to sleep in good spirits and waking up early with suicidal urges. The mood changes of cyclothymia are best described as “endoreactive” in the sense that endogenous overreactivity seems to determine the sudden shifts in mood and behavior (e.g., falling in love with a person one has just met and as quickly falling out of love).

A. For at least 2 years, the presence of numerous periods with hypomanic symptoms and numerous periods with depressive symptoms that do not meet criteria for a major depressive episode.
Note: In children and adolescents, the duration must be at least 1 year.
B. During the above 2-year period (1 year in children and adolescents), the person has not been without the symptoms in criterion A for more than 2 months at a time.
C. No major depressive episode, manic episode, or mixed episode has been present during the first 2 years of the disturbance.
Note: After the initial 2 years (1 year in children and adolescents) of cyclothymic disorder, there may be superimposed manic or mixed episodes (in which case both bipolar I disorder and cyclothymic disorder may be diagnosed) or major depressive episodes (in which case both bipolar and cyclothymic disorder may be diagnosed).
D. The symptoms in criterion A are not better accounted for by schizoaffective disorder and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.
E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism).
F. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Table 14.6-32 DSM-IV Diagnostic Criteria for Cyclothymic Disorder

Biphasic dysregulation characterized by abrupt endoreactive shifts from one phase to the other, each phase lasting for a few days at a time, with infrequent euthymia.

Behavioral manifestations:
- Introverted self-absorption versus uninhibited people-seeking
- Taciturn versus talkative
- Unexplained tearfulness versus buoyant jocularity
- Psychomotor inertia versus restless pursuit of activities

Subjective manifestations:
- Lethargy and somatic discomfort versus eutonia
- Dulling of senses versus keen perceptions
- Slow-witted versus sharpened thinking
- Shaky self-esteem alternating between low self-confidence and overconfidence
- Pessimistic brooding versus optimism and carefree attitudes

Updated from Akiskal HS, Khani M, Scott-Strauss A: Cyclothymic temperamental disorders. Psychiatr Clin North Am 2:527, 1979.

Table 14.6-33 Clinical Features of Cyclothymic Disorder

Mood swings in these ambulatory patients are overshadowed by the chaos that the swings produce in their personal lives. Repeated romantic breakups or marital failures are common because of interpersonal friction and episodic promiscuous behavior. Uneven performance at school and work is also common. Persons with cyclothymic disorder are dilettantes; they show great promise in many areas, but rarely bring any of their efforts to fruition. As a result, their lives are often a string of improvident activities. Geographical instability is a characteristic feature; easily attracted to a new locale job, or love partner, they soon lose interest and leave in dissatisfaction. Polysubstance abuse which occurs in as many as 50 percent of such persons, is often an attempt at self-treatment.

Bipolar II Disorder (and the Soft Bipolar Spectrum) Research conducted during the past three decades showed that between the extremes of classic manic-depressive illness defined by at least one acute manic episode (bipolar I disorder) and strictly defined major depressive disorder without any personal or family history of mania (pure unipolar disorder), exists an overlapping group of intermediary forms characterized by recurrent major depressive episodes and hypomania. Table 14.6-34 summarizes the author's observations in defining the clinical subtypes within this intermediary realm best described as “soft bipolarity.” The most accepted of the subtypes is bipolar II disorder (with spontaneous hypomania), elevated to the status of a nosological entity in DSM-IV (Table 14.6-35). Current data worldwide indicate that bipolar II disorder is actually more prevalent than bipolar I disorder. This certainly appears true in the outpatient setting, where 30 to 50 percent of persons with major depressive disorder have been reported to conform to the bipolar II pattern.

	Bipolar I:	At least one manic episode
Soft bipolar	Bipolar II:	Recurrent depressions with hypomania and cyclothymic disorder
Soft bipolar	Bipolar III: (pseudounipolar)	Recurrent depressions without spontaneous hypomania but often with hyperthymic temperament and bipolar family history
Unipolar depressions:		No evidence for hypomania, cyclothymic disorder, hyperthymic disorder, or bipolar family history

Table 14.6-34 Spectrum of Bipolar Disorders Compared With Unipolar Depression

A. Presence (or history) of one or more major depressive episodes.
B. Presence (or history) of at least one hypomanic episode.
C. There has never been a manic episode.
D. The mood symptoms in criteria A and B are not better accounted for by schizoaffective disorder, and are not superimposed on schizophrenia, schizophreniform disorder, delusional disorder, or psychotic disorder not otherwise specified.
E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify current or most recent episode:
Hypomanic: if currently (or most recently) in a hypomanic episode
Depressed: if currently (or most recently) in a major depressive episode

Specify if the current or most recent major depressive episode only if it is the most recent type of mood episode:
Severity/psychotic/remission specifiers
Chronic
With catatonic features
With melancholic features
With atypical features
With postpartum onset

Specify:
Longitudinal course specifiers (with or without interepisode recovery)
With seasonal pattern (applies only to the pattern of major depressive episodes)
With rapid cycling

Table 14.6-35 DSM-IV Diagnostic Criteria for Bipolar II Disorder

The following self-description provided by a 34-year-old poet illustrates the pattern:

> I have known melancholy periods, lasting months at a time, when I would be literally paralyzed: All mental activity comes to a screeching halt, and I cannot even utter one word. I become so dysfunctional that I was once hospitalized. Although the paralysis creeps into me insidiously—often lasting months—it typically reverses within hours. I am suddenly alive and vibrant, I cannot turn off my brain neither during the day nor at night; I usually go on celebrating like this for many weeks, needing no more than few hours of slumber each day.

This vignette is nearly identical to the autobiographical description provided by the British poet William Cowper three centuries earlier:

> I have known many a lifeless and unhallowed hour ... long intervals of darkness interrupted by short returns of peace and joy ... For many succeeding weeks to rejoice day and night was all my employment. Too happy to sleep much, I thought it was lost time that was spent on slumber.

The hypomania at the end of depressive episodes in most bipolar II patients does not persist long; it is usually measured in days. The modal duration of hypomania found in Memphis and Zurich studies was 2 days. Another common form of bipolar II disorder is major depressive disorder superimposed on cyclothymic disorder, in which hypomania precedes and follows major depression, the entire interepisodic period characterized by cyclothymic mood instability. As a result, these are difficult bipolar II patients to manage in clinical practice.

Hypomania in bipolar II disorder can be defined as minimanic episodes occurring spontaneously. Bipolar II disorder—especially when major depressions are superimposed on cyclothymia—is thus best characterized as cyclical or “cyclothymic depression.”

The depressive episodes of patients with bipolar disorder often have admixtures (e.g., flight of ideas, increased drives and impulsivity in sexual and other domains). The phenomenon of lithium augmentation is perhaps best explained by the high prevalence of pseudounipolar depressions with subtle hypomania either during or following a depressive episode, as well as the mixed simultaneous presence of depressive and hypomanic symptoms. The latter are not as severe as dysphoric mixed states, but are refractory to antidepressants nonetheless.

Hypomania The common denominator of the soft spectrum of bipolar disorders is the occurrence of hypomania. Hypomania (Table 14.6-9) refers to a distinct period of at least a few days of mild elevation of mood, sharpened and positive thinking, and increased energy and activity levels, typically without the impairment characteristic of manic episodes. It is not merely a milder form of mania. Hypomania occurring as part of bipolar II disorder rarely progresses to manic psychosis; distractibility is uncommon in hypomania, and insight is relatively preserved. Hypomania is distinguished from mere happiness in that it tends to recur (happiness does not) and can sometimes be mobilized by antidepressants. In cyclothymic disorder it alternates with minidepressions; in hyperthymic temperament it constitutes the person's habitual baseline. These definitions then recognize three patterns of hypomania: brief episodes heralding the termination of a retarded depressive episode (bipolar II disorder), cyclic alternation with minidepressions (cyclothymic disorder), and an elevated baseline of high mood, activity, and cognition (hyperthymic or chronic hypomanic traits).

Because hypomania is experienced either as a rebound relief from depression or as pleasant, short-lived, ego-syntonic mood state, persons with bipolar II disorder rarely report it spontaneously. Skillful questioning is thus required to make the diagnosis of soft bipolar conditions; as in mania, collateral information from family members is crucial. In interviewing the patient the following probes have been found useful to elicit hypomania: “Have you had a distinct sustained high period (1) when your thinking and perceptions were unusually vivid or rapid, (2) your mood was so intense that you felt nervous, and (3) you were endowed with such energy that others could not keep up with you?” The hypomanic manifestations for hypomania in the DSM-IV scheme basically list the signs and symptoms of mania in criterion A and B for mania (Table 14.6-9) but require fewer items and shorter duration. Clinical and epidemiological studies in the United States and Europe have revealed a richer range of manifestations including an increase in cheerfulness and jocularity; gregariousness and people seeking; greater interest in sex; talkativeness, self-confidence, and optimism; and decreased inhibitions and sleep need. The clinician must ascertain that those experiences were not due to stimulant or alcohol withdrawal. Depressive and hypomanic periods are often not easily discerned because chronic caffeinism, stimulant abuse, or both complicate the depression. In such instances, diagnosis should be based on clinical observation for 1 month after detoxification.

When in doubt, direct clinical observation of hypomania—sometimes elicited by antidepressant pharmacotherapy—provides definitive evidence for the bipolar nature of the disorder. Unfortunately DSM-IV denies bipolar status to treatment-emergent hypomanic episodes. Follow-up studies in juvenile and young adults with pharmacological hypomania have demonstrated that nearly all such individuals progress to spontaneous hypomanic (or manic) episodes. Although DSM-IV stipulates a minimum duration of 4 days for hypomania, any recurrent hypomania coupled with major depression should count toward the diagnosis of bipolar II.

Seasonal Patterns Seasonality is observed in many cyclic depressions, often with autumn or winter anergic depression and energetic periods in the spring. This natural propensity explains why phototherapy may provoke mild hypomanic switches. Although not specifically identified by DSM-IV, seasonal depressions conform, in large measure, to the bipolar II or III pattern. Furthermore, preliminary evidence suggests that treatment with classic antidepressants disrupts the baseline seasonality, with the depressive phase appearing in the spring and summer. The changes antidepressants induce in seasonal depressions probably represent a special variant of the rapid-cycling phenomenon.

Temperament and Polarity of Episodes New systematic clinical observations have revealed that bipolar II disorder (characterized predominantly by depressive attacks) arises more often from a hyperthymic or cyclothymic baseline, whereas bipolar I disorder (defined by manic attacks) not uncommonly arises from the substrate of a depressive temperament. When the hyperthymic temperament occurs in bipolar I disorder, it is usually associated with a recurrent mania, which is an

uncommon bipolar course. A prospective 11-year NIMH study of major depressive disorder patients who switched to bipolar II disorder showed that “mood-labile” (cyclothymic) and “energetic-active” (hyperthymic) temperament traits were highly specific and reasonably sensitive predictors of such an outcome.

Bipolarity is conventionally defined by the alternation of manic (or hypomanic) and depressive episodes. The foregoing data on temperaments suggest hat a more fundamental characteristic of bipolarity is the reversal of temperament into its “opposite” episode (in the case of the bipolar II spectrum, from cyclothymia and hyperthymia to major depression). Such findings suggest that the intrusion of cyclothymic and hyperthymic traits into a depressive episode may underlie the instability of the bipolar II subtype and could partly explain why bipolar II depression often has mixed features. These considerations may have important implications for preventing recurrence. For instance, a prospective study of the onset of bipolar disorder in the offspring or sibs of adults with the disorder found that children with depressive onsets as their first episode (and which were usually treated with antidepressants) had significantly higher rates of recurrence than those with manic or mixed onsets (treated with lithium) during a 3-year prospective observation. It appears that temperamental instability in the depressive group might have predisposed them to the cycling effect of antidepressants.

Alcohol, Substance Abuse, and Suicide New evidence supports the high prevalence of alcohol and substance abuse in mood disorder subtypes, especially those with interepisodic cyclothymia and hyperthymia. The relation appears particularly strong in the teenage and early adult years, when the use of such substances often represents self-medication for the mood instability. It is not just self-treatment for selected symptoms associated with the down or up phases (e.g., alcohol to alleviate the insomnia and nervousness characteristic of both phases), it also augments certain desired ends (e.g., stimulants to enhance high-energy performance and sexual behavior associated with hypomania). How many display alcohol and substance abuse secondary to an underlying bipolar diathesis remains to be determined. But in view of findings suggesting a link between polysubstance abuse and suicide in adolescents with bipolar familial backgrounds, the use of mood stabilizers in these adolescents should be strongly considered. Although alcohol and stimulant use continues into adult years in a considerable number of bipolar disorder patients, such use is often unrelated to familial alcoholism, and frequently tends to dwindle during long-term follow-up, which supports the self-medication hypothesis. To complicate matters, in a substantial minority of cases, bipolar mood swings appear for the first time after abrupt cessation of long-term alcohol use; it is not uncommon for such mood swings to escalate into full-blown bipolar syndromes.

Rapid-Cycling Bipolar Disorder Rapid cycling is defined as the occurrence of at least four episodes—both retarded depression and hypomania (or mania)—a year (Table 14.6-36). Thus rapid cyclers are rarely free of affective symptoms and suffer serious vocational and interpersonal incapacitation. Lithium is often only modestly helpful to those patients, as are traditional antipsychotic agents; most antidepressants readily induce excited episodes and thus aggravate the rapid-cycling pattern. A balance among mood stabilizers, antipsychotic drugs, and antidepressants may be difficult to achieve. Many such patients require frequent hospitalization because they develop explosive excitement and precipitous descent into severe psychomotor inhibition. The disorder is a roller coaster nightmare for the patient, significant others, and the treating physician. Treating these patients is an art.

Specify if:

With rapid cycling (can be applied to bipolar I disorder or bipolar II disorder)

At least four episodes of a mood disturbance in the previous 12 months that meet criteria for a major depressive, manic, mixed, or hypomanic episode.

Note: Episodes are demarcated either by partial or full remission for at least 2 months or a switch to an episode of opposite polarity (e.g., major depressive episode to manic episode).

Table 14.6-36 DSM-IV Diagnostic Criteria for Rapid Cycling Specifier

As expected, rapid cycling commonly arises from a cyclothymic substrate, which means that most rapid cyclers have bipolar II disorder. Factors favoring its occurrence include (1) female gender; (2) borderline hypothyroidism; (3) menopause; (4) temporal lobe dysrhythmias; (5) alcohol, minor tranquilizer, stimulant, or caffeine abuse; and (6) long-term, aggressive use of antidepressant medications. Most clinically identified patients are bipolar II women in middle age or upper social classes. Rapid cycling is uncommon from a bipolar I base.

Leadership and Creativity Persons with hyperthymic temperament and soft bipolar conditions in general possess assets that permit them to assume leadership roles in business, the professions, civic life, and politics. Increased energy, sharp thinking, self-confidence and eloquence represent the virtues of an otherwise stormy life.

Creative achievement is relatively uncommon among those with the manic forms of the disorder, which is too severe and disorganizing to permit the necessary concentration and application. Notable artistic achievements are found among those with soft bipolar disorders, especially cyclothymic disorders. Psychosis, including severe bipolar swings, is generally incompatible with creativity. That conclusion, based on recent systematic studies, tends to refute the romantic tendency to idolize insanity as central to the creative process. As talent is the necessary ingredient of creativity, how might soft bipolarity contribute? The simplest hypothesis is that depression might provide insights into the human condition, and the activation associated with hypomania helps in producing the artistic work. A more profound interpretation suggests that the repeated self-doubt that comes with recurrent depression might be an important ingredient of creativity, because original artistic or scientific expression is often initially rejected, and the self-confidence that accompanies repeated bouts of hypomania can help in rehearsing such ideas or expressions until they are perfected. Finally, the tempestuous object relations associated with bipolarity in the parent's or the patient's life often create the unique biographical landmarks that might be immortalized in an artistic medium.

Bipolar Disorder Not Otherwise Specified The criteria for bipolar disorder not otherwise specified are listed in Table 14.6-37.

The bipolar disorder not otherwise specified category includes disorders with bipolar features that do not meet criteria for any specific bipolar disorder. Examples include:

1. Very rapid alternation (over days) between manic symptoms and depressive symptoms that do not meet minimal duration criteria for a manic episode or major depressive episode
2. Recurrent hypomanic episodes without intercurrent depressive symptoms
3. A manic or mixed episode superimposed on delusional disorder, residual schizophrenia, or psychotic disorder not otherwise specified
4. Situations in which the clinician has concluded that a bipolar disorder is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced

Table 14.6-37 DSM-IV Diagnostic Criteria for Bipolar Disorder Not Otherwise Specified

Recurrent hypomanic episodes without intermittent depressions (example 2 in the DSM-IV criteria for bipolar disorder not otherwise specified are almost never observed clinically.

Recurrent Brief Hypomania Recurrent brief depressive disorder as a transitional form between dysthymia and major depression or brief hypomanic episodes often have been missed during evaluations performed by nonclinicians. Some patients who meet the Zurich description might therefore belong in the soft bipolar spectrum. Indeed, subsequent evaluation and analyses have revealed high rates of comorbidity between recurrent brief depression and brief hypomania. Thus, some recurrent brief depressive cases appear to be variants of bipolar disorder. The subtle bipolar nature of recurrent brief depressive disorder is clinically supported by the fact that

the very few such patients the author has encountered in his own practice did poorly with antidepressant monotherapy but benefited from mood stabilizers used alone or combined with antidepressants.

The recurrent hypomanic counterpart of recurrent brief depressive disorder is described under soft bipolar conditions. By DSM-IV criteria, it represents an instance of bipolar disorder not otherwise specified.

Hysteroid Dysphoria The category hysteroid dysphoria combines reverse vegetative signs with the following characteristics: (1) giddy responses to romantic opportunities and an avalanche of dysphoria (angry-depressive, even suicidal responses) upon romantic disappointment; (2) impaired anticipatory pleasure, yet the capability to respond with pleasure when such is provided by others (i.e., preservation of consummatory reward); (3) craving for chocolate and sweets, which contain phenylethylamine compounds and sugars believed to facilitate cellular and neuronal intake of the amino acid L-tryptophan, hypothetically leading to synthesis of endogenous antidepressants in the brain. The use of the epithet “hysteroid” was used to convey that the apparent character pathology was secondary to a biological disturbance in the substrates governing affect, drives, and reward. The intense, giddy, unstable life of the patient with hysteroid dysphoria suggests links to cyclothymic disorder or bipolar II disorder. This suggestion is further supported by the Columbia group's tendency to subsume those patients under atypical depressions (some of which, as indicated, have bipolar affinities). Like patients with bipolar depression, they respond preferentially to monoamine oxidase inhibitors (MAOIs). In brief, hysteroid dysphoria appears to be a variant of bipolar II with cyclothymic-irritable traits. Other variants of bipolar II disorder with hyperthymic-narcissistic traits are described under soft bipolar disorder and represent instances of bipolar disorder not otherwise specified.

Bipolar III Disorder In bipolar III disorder (which is not an official nosological term but can be subsumed under bipolar disorder not otherwise specified) evidence of bipolarity is softer, such as a single brief episode of an antidepressant-mobilized switch. In a related subgroup of cryptic bipolar disorders, strong evidence for familial bipolarity raises the possibility that some phenotypically “unipolar” depressed patients are nonetheless constitutionally bipolar; in such cases, history for hypomania occurring in discrete episodes is not obtained; instead the patient's habitual temperamental baseline is sunny, overenergetic, and overoptimistic (hyperthymic).

Depending on the threshold of traits used in determining the presence of hyperthymia, bipolar III patients may constitute 10 to 20 percent of those with major depressive disorder. Thus, many patients with so-called unipolar depression are actually “pseudounipolar.” The presence of marked narcissistic traits is a helpful clinical clue that a clinically depressed patient might belong to the group of those with hyperthymic depressions.

MOOD DISORDERS NOT OTHERWISE SPECIFIED

After all diagnostic information has been obtained, some depressed and bipolar or otherwise affective patients do not meet the specific criteria for the mood conditions described thus far. Mood disorders not otherwise specified is a statistical concept for filing purposes and not a clinical description. The author prefers to consider such cases as undiagnosed mood disorders rather than using the DSM-IV categories of depression disorder not otherwise specified, bipolar disorder not otherwise specified or mood disorder not otherwise specified.

What follows are descriptions of conditions that commonly appear in the psychiatric literature but do not easily fit into the official nosology of mood disorders. They represent hybrids between mood and anxiety disorders.

Mixed Anxiety-Depressive Disorder The inclusion of anxious depressive states in a DSM-IV appendix acknowledges the simultaneous occurrence of anxious (e.g., the threat loss represents) and depressive (e.g., the despair of loss) cognition in a person confronted with a major aversive life situation. The admixture implies that the psychopathology progresses from anxiety to depression, that the patient's mental state is still in flux, and that the ongoing dynamics partly explains the subacute or chronic nature of the disorder. Anxious depression serves to point to the common presence of anxiety in depressive states, especially its greater visibility when the depression is less prominent. Patients with the latter presentation are reportedly most prevalent in general medical settings. This should not come as a surprise, because depressive symptoms that motivate medical consultation commonly complicate generalized anxiety states with a subthreshold level of symptomatology. Some authorities argue that neurotic depressions arise as maladaptive responses to anxiety and on that basis suggest retaining the “neurotic depression” rubric. Recent preliminary genetic data indirectly support the contention that certain (unipolar) depressive and (generalized) anxiety states are related. However, more research is needed before such an entity can be unequivocally accepted as an official nosological category. The difficulty is that as currently defined, anxious depressions are heterogeneous. In patients refractory to anxiolytic or antidepressant treatment or both, practitioners must entertain the diagnosis of a complex bipolar II disorder with mixed features. Indeed, recent genetic investigations suggest that bipolar II disorder with panic attacks might represent a special form of bipolar disorder.

Atypical Depression Although a delimited version of atypical depression was incorporated into DSM-IV as “atypical features (Table 14.6-5) to qualify the cross-sectional picture of depressive disorders, this construct is much broader in the clinical research literature. Originally developed in England and currently under investigation at Columbia University in New York, atypical depression refers to fatigue superimposed on a history of somatic anxiety and phobias, together with reverse vegetative signs (mood worse in the evening, insomnia, tendency to oversleep and overeat). Sleep is disturbed in the first half of the night in many persons with atypical depressive disorder, so irritability, hypersomnolence, and daytime fatigue would be expected. The temperaments of these patients are characterized by inhibited-sensitive traits. The MAOIs and serotonergic antidepressants seem to show some specificity for such patients, which is the main reason that atypical depression is taken seriously.

Other research suggests that reverse vegetative signs can be classified as either (1) the anxious type just described or (2) a subtle bipolar subtype with protracted hyperphagic-hypersomnic-retarded dysthymic disorder with occasional brief extroverted hypomanic-type behavior, often elicited by antidepressants. Increasing evidence indicates considerable affinity between atypical depression and bipolar II and III disorders. Furthermore, many patients with dysthymic disorder exhibit atypical features at various times. Actually, atypical depression might be an artifact of the DSM-IV definition of hypomania of 4 or more days. Recent Italian research suggests that many patients with atypical depressive meet criteria for brief hypomania or cyclothymic disorder.

The categories of not otherwise specified in the DSM-IV mood disorders schema largely reflect inadequacies of the operational approach to capture patients whose symptomatology falls between or on the boundaries of more classic diagnoses.

DIFFERENTIAL DIAGNOSIS

A missed mood disorder diagnosis means that the disorder does not receive specific treatment, which has serious consequences. Many such persons drop out of school or college, lose their jobs, get divorced, or may commit suicide. Those with unexplained somatic symptoms are frequent users of the general health system. Others are unwell despite interminable psychotherapy. Some, treated with dopamine receptor antagonists develop tardive dyskinesia unnecessarily. As with other medical disorders for which specific treatments are available, accurate diagnosis and early treatment are within the purview of all physicians and mental health professionals. Psychiatrists, in particular, should develop the competence to detect the entire spectrum of mood disorders. Despite massive educational efforts, underdiagnosis and undertreatment of mood disorders remain serious problems worldwide.

Although much enthusiasm was generated a decade ago about the potential use of certain biologic markers (e.g., REM latency, dexamethasone (Decadron) suppression test, and the thyrotropin-releasing-hormone test) to corroborate the differentiation of mood disorder from adjacent disorders, no definitive progress justifies their routine use in clinical practice. Faced with unusual or confusing presentations, a systematic clinical approach is still the best method in differential diagnosis (1) to detail all clinical features of the current episode, (2) to elicit a history of more typical major mood episodes in the past, (3) to assess whether the presenting complaints recur periodically or cyclically, (4) to substantiate adequate social functioning between periods of illness, (5) to obtain a positive family history for classic mood disorder and to construct a family pedigree, and (6) to document a history of unequivocal therapeutic response to thymoleptic medication or ECT in either the patient or the family.

Using the foregoing validating approach, one can examine the affective links of many DSM-IV disorders currently listed under mood disorders not otherwise specified, as well as controversial nosological entities currently categorized as nonmood disorders. The latter include conduct disorders; borderline personality disorder; impulse-control disorders; polysubstance abuse; psychotic disorder not otherwise specified; pain disorder; hypochondriasis; hypoactive sexual desire disorder; circadian rhythm sleep disorder, delayed sleep phase type; bulimia nervosa; and adjustment disorder (with work inhibition). These conditions place special emphasis on selected affective features, such as disinhibited behavior, temperamentality, mood lability, vegetative disturbances, and psychomotor anergia. What follows is a systematic examination of the differential diagnosis of mood disorders with their more classic boundaries.

Alcohol and Substance Use Disorders The high comorbidity of alcohol and substance use disorders with mood disorders cannot be explained as merely the chance

occurrence of two prevalent disorders. Self-medication for mood disorders is insufficiently appreciated by both psychiatrists and other professionals who deal with addiction. Given the clinical dangers of missing an otherwise treatable disorder, mood disorder should be seriously considered as the primary diagnosis if marked affective manifestations persist or escalate after detoxification (e.g. 1 month). This consideration also pertains to cyclothymic disorder and dysthymic disorder, which appear particularly likely to invite self-medication. The clinical validating strategies listed above can further buttress a mood disorder diagnosis.

The DSM-IV category of substance-induced mood disorder (Table 14.6-18) is difficult to validate clinically because in the absence of an affective diathesis, detoxification should, in principle, rapidly clear affective disturbances in persons whose primary problem is that of substance abuse. In the author's view, a dual diagnosis of both a mood disorder and a substance use disorder is a more realistic clinical approach to this group of patients. Bipolarity, particularly bipolar II disorder, should be sought in the interface of mood and substance use disorders.

A 27-year-old married businessman employed in an international family venture owned by his father presented with a court-ordered request for psychiatric treatment. He had been found “bringing” cocaine across the U.S.-Mexican border and was briefly jailed. He had used stimulants since his late teens to enhance his already high level of energy. His family was rich, and he had no difficulty affording cocaine. During the previous year, he had needed more cocaine because of greater moodiness and fleeting suicidal ideation, which he linked to increasing tensions between him and his father: “My father was never satisfied with me and demanded greater and greater performance from me.” His arrest by police was a major embarrassment for him and his family and motivated his compliance with psychiatric hospitalization to detoxify him. He had not had cocaine for 10 days, exhibited marked lability of mood, and gradually sank into a severe hypersomnic-retarded depression of stuporous proportions. He was treated with tranylcypromine (Parnate) 20 mg twice a day, and within 10 days he switched into hypomania, his mind “exploding with creativity and confidence,” marked jocularity and witticisms that entertained other patients, and marked seductiveness toward the nurses. His wife recalled that the patient previously had had several such periods naturally (i.e., “off cocaine”), which had strained their marriage due to “brief sexual liaisons.” Reducing the tranylcypromine dosage by 50 percent did not eliminate the hypomanic behavior and lithium 900 mg a day was added. He has since been maintained on a combination of tranylcypromine and lithium for 4 years; he has not relapsed into cocaine use, and following few psychoeducational sessions involving father and spouse, relationships with family and spouse have been less tempestuous. (Since consultation was sought by the patient's 60-year-old mother, an attractive, sophisticated woman, who confessed that for years she had been engaging in “love relationships” with young artists, with apparently her husband's “silent consent”; since at least her mid-20s, she by history would meet the criteria for bipolar II, only treated “on the couch,” and both her sister and brother had received treatment for “alcohol excesses.”) Patient states that pharmacotherapy—which did require adjustment now and then—has helped in balancing the “rough edges” of his “high-nervous temperament” and his “periodic lapse into paralyzing fatigue states that occurred at stressful times.”

If clinicians had assumed that primarily due to cocaine withdrawal, they would have never treated the patient's bipolar II disorder. DSM-IV conventions in this regard unfortunately bias diagnosis in favor of substance use disorders and, more tragically, against the realistic chance of cure from substances.

There is emerging interest in treating dually diagnosed patients with mood stabilizers, especially anticonvulsants. The intention is to attenuate any withdrawal phenomena from substances of abuse, while treating any underlying or emerging soft bipolar disorders.

Personality Disorders The state dependency of most personality measures is well documented. Accordingly, as exhorted by DSM-IV, clinicians should refrain from using personality disorder labels in describing patients with active affective illness and should focus instead on competent treatment of the mood disorder. Even in those with chronic or subthreshold mood disorders, personality maladjustment is best considered postaffective, arising from the distortions and conflicts that affective disturbances produce in the life of the sufferer. The most problematic of the personality labels used in those with mood disorders is borderline personality disorder, usually applied to teenage and young adult females. The DSM-IV diagnostic criteria for the disorder indicate a liberal mélange of low-grade affective symptoms and behavior. Table 14.6-38 shows that the overlap between borderline personality and mood disorders is extensive, so that giving a “borderline” diagnosis to a person with mood disorder is redundant. Use of personality disorder diagnoses may lead to neglect of the mood disorder or perhaps half-hearted treatment of the mood disorder; failure to respond would then be blamed on the patient's “self-defeating character” or “resistance to getting well,” thus exculpating the clinician.

Familial:	High rates of mood disorder
Phenomenology:	Dysthymic disorder
	Cyclothymic disorder
	Bipolar II disorder
	Mixed state
Pharmacological response:	Worsening on tricyclic antidepressants
	Stabilization on anticonvulsants
Prospective course:	Major mood episodes
	Suicide

Summarized from Akiskal H, Chen S, Davis G, Puzantian V, Kashgarian M, Bolinger M: Borderline: An adjective in search of a noun. Clin Psychiatry 46: 41, 1985.

Table 14.6-38 Overlap of Borderline Personality Disorder and Mood Disorders

Although more-systematic research is needed on the complex interface of personality and mood disorders, clinically they are often inseparable. As with alcohol and substance use disorders, it is generally preferable to diagnose mood disorders at the expense of personality disorders, which should not be difficult to justify in most cases that satisfy the validating strategies outlined above. When features of personality and mood disorders coexist, it is good practice to defer Axis II diagnoses and embark upon competent treatment of the concurrent mood disorder. Although not all personality disturbances recede with the competent treatment of mood disorders, so many experienced clinicians have seen such disturbances melt away with the successful resolution of the mood disorder that erring in favor of mood disorders is justified.

A 19-year-old single woman presented with the chief complaint that “all men are bastards.” Since her early teens, with the onset of her menses, she had complained of extreme variability in her moods on a nearly daily basis; irritability with hostile outbursts was her main affect, though more protracted hypersomnic depressions with multiple overdoses and wrist slashings had led to at least three hospitalizations. She also suffered from migrainous headaches that, according to the mother, had motivated at least one of those overdoses. Despite her tempestuous and suicidal moods that led to these hospitalizations, she complained of “inner emptiness and a bottomless void.” She had used heroin, alcohol, and stimulants to overcome this troubling symptom. She also gave history of ice-cream craving and frequent purging. She was talented in English and wrote much-acclaimed papers on the American confessional poet, Anne Sexton. She said she was mentally disturbed because of a series of stepfathers who had all forced “oral rape” between the ages of 11 and 15. She subsequently gave herself sexually to any man she met in bars, no longer knowing whether she was a “prostitute” or a “nice little girl.” On two occasions she had inflicted cigarette burns on her vagina “to feel something.” She had also engaged in a “brief lesbian relationship” that ultimately left her “emptier” and guilt-ridden; nonetheless, she now believed that she should burn in hell, because she could not get rid of “obsessing” about the excitement of mutual cunnilingus with her much older female partner. The patient's mother, who owned an art gallery, had been married five times and gave history of unmistakable hypomanic episodes; a maternal uncle had died from alcohol-induced cirrhosis. The patient's father, a well-known lawyer known for his “temper and wit,” had committed suicide. The patient was given phenelzine (Nardil), eventually raised to 75 mg a day, at which point the mother described her as “the sweet daughter she was before age 13.” At her next premenstrual phase, patient developed insomnia, ran away from home at night, started “dancing like a go-go girl,” met an “incredibly handsome man” of 45 years (actually, a pornography shop owner) and got married. Lithium augmentation controlled this excited episode. After many dosage adjustments, she is maintained on a combination of lithium (900 mg a day) and divalproex (Depakote) (750 mg a day). The patient now attends college and has completed four semesters in art history. In addition to control of her irritable and suicidal moods, bulimic and migraine attacks have abated considerably. Her marriage has been annulled on the basis that she was not mentally competent at the time of the wedding. She is no longer promiscuous and now expresses fear of intimacy with men she is attracted to. She is receiving individual psychotherapy for this problem.

The author often hears the complaint that even when a mood disorder is diagnosed in a “borderline” patient, response to antidepressants is disappointing. The problem is that affective disorders in these patients usually conform to bipolar II disorder—often complicated by ultrarapid cycling—and many clinicians, including

some with biological orientation, may lack sufficient experience in the art of pharmacologically managing patients who markedly deviate from classic bipolar I disorder.

The interface of mood disorders and behavioral disturbances (conduct and attention-deficit/hyperactivity disturbances) in children is even more problematic than in adult psychiatry. Nonetheless, progress has occurred in clinically recognizing certain behavioral manifestations as possible signs of depression in juvenile subjects, including periodic marked decline in school performance; restlessness and pulling or rubbing hair, skin, or clothing; outbursts of complaining, shouting, or crying; and aggressive or antisocial acts (such as kicking the mother, shoplifting) out of character to the child; as well as other acute personality changes ranging from defiant attitudes to negativism and avoidant behavior. Examined carefully, children and pubescent youth with these characteristics often meet the specific criteria for the diagnosis of major depressive disorder or dysthymic disorder. However, most children do not complain of subjective dysphoria; instead, the clinician can observe the depressed affect in the child's facial expressions or overall demeanor. After much resistance, many child clinicians now accept the existence of childhood mood disorders.

Bipolar disorder in children, even in adolescents, is still grossly underdiagnosed at the expense of so-called externalizing disorders. Table 14.6-39 lists those and related conditions often confused with bipolar disorders in juvenile patients. Many children express bipolar disorder in explosive outbursts of irritable mood and behavior (i.e., as a mixed or dysphoric manic state); another pattern is intermittent hypomania and cyclothymia. Children with bipolar disorder are distinguished from those with so-called externalizing disorders by the fact that they are often, though not always, considered charming and likeable, yet overconfident or delusionally grandiose, and may exhibit age-inappropriate sexual behavior, such as lecherous advances toward adult women (e.g., their elementary school teachers): Moreover, they often get worse on stimulant medication. Correct diagnosis depends on the index of suspicion of a clinician who is convinced that bipolarity exists in juvenile subjects. Depression with first onset before age 18 has an extremely high rate of switching into both bipolar I and bipolar II disorders.

Total (N = 44)	Percent
Adjustment disorder	35
Conduct disorder	15
Attention-deficit/hyperactivity disorder	9
Mental retardation	6
Separation anxiety disorder	9
Overanxious disorder	11
Schizophrenia	15

Adapted from Akiskal HS, Downs J, Watson S, Daugherty D, Pruitt DB: Affective disorders in referred children and younger siblings of manic-depressives. Arch Gen Psychiatry 42:996, 1985.

Table 14.6-39 Misdiagnosis in the Affectively Ill Juvenile Kin of Adults With Bipolar Disorder

Normal Bereavement Bereaved persons exhibit many depressive symptoms during the first 1 to 2 years after their loss, so how can the 5 percent of bereaved persons who have progressed to a depressive disorder be identified?

- Grieving persons and their relatives perceive bereavement as a normal reaction, while those with depressive disorder often view themselves as sick and may actually believe they are losing their minds.
- Unlike the melancholic person, the grieving person reacts to the environment and tends to show a range of positive affects.
- Marked psychomotor retardation is not observed in normal grief.
- Although bereaved persons often feel guilty about not having done certain things that might have saved the life of the deceased loved one (guilt of omission), they typically do not experience guilt of commission.
- Delusions of worthlessness or sin and psychotic experiences in general point toward mood disorder.
- Active suicidal ideation is rare in grief but common in major depressive disorder.
- Mummification (i.e., keeping the belongings of the deceased person exactly as they were before his or her death) indicates serious psychopathology.
- Severe anniversary reactions should alert the clinician to the possibility of psychopathology.

In another form of bereavement depression, the sufferer simply pines away, unable to live without the departed person, usually a spouse. Although not necessarily pathological by the foregoing criteria, such persons do have a serious medical condition. Their immune function is often depressed, and their cardiovascular status is precarious. Death can ensue within a few months of that of a spouse, especially among elderly men. Such considerations (highlighted in the work of Sidney Zisook and his San Diego colleagues at the University of California) suggest that it would be clinically unwise to withhold antidepressants from many persons experiencing an intensely mournful form of grief.

A 75-year-old widow was brought by her daughter because of severe insomnia and total loss of interest in daily routines following her husband's death 1 year before. She had been agitated for the first 2 to 3 months and thereafter “sank into total inactivity—not wanting to get out of bed, not wanting to do anything, not wanting to go out.” According to her daughter, she was married at 21, had four children, and had been a housewife until her husband's death from a heart attack. Past psychiatric history was negative; premorbid adjustment had been characterized by compulsive traits. During the interview she was dressed in black, appeared moderately slowed, and sobbed intermittently, saying “I search everywhere for him ... I don't find him.” When asked about life, she said “everything I see is black.” Although she expressed no interest in food, she did not seem to have lost an appreciable amount of weight. Her dexamethasone suppression test result was 18 μg/dL. The patient declined psychiatric care, stating that she “preferred to join her husband rather than get well.” She was too religious to commit suicide but by refusing treatment she felt she would “pine away ... find relief in death and reunion.”

Anxiety Disorders Anxiety symptoms including panic attacks, morbid fears, and obsessions are common during depressive disorders, and depression is a common complication of anxiety states. Systematic British studies have shown that early-morning awakening, psychomotor retardation, self-reproach, hopelessness, and suicidal ideation are the strongest clinical markers of depression in that differential diagnosis. On follow-up of depressed patients, the manifestations tend to remit, whereas those with anxiety states continue to exhibit marked tension, phobias, panic attacks, vasomotor instability, feelings of unreality, and perceptual distortions as well as hypochondriacal ideas. A predominance of such anxiety features antedating the present disorder suggests the diagnosis of an anxiety disorder. Since anxiety disorders rarely first appear after the age of 40, late appearance of marked anxiety features strongly favors the diagnosis of melancholia. The clinical picture is often one of morbid groundless anxiety with somatization, hypochondriasis, and agitation. The depressive nature of the illness is further supported by a superior response to ECT.

Periodic monosymptomatic phobic and obsessional states exist that can be regarded as affective equivalents on the basis of a family history of mood disorders and their response to thymoleptic agents. Recent data from a large clinical series suggests that 15 percent of patients with obsessive-compulsive disorders develop unmistakable hypomanic symptoms; these patients are best considered to have bipolar II disorder and treated with lithium salts. Social phobias exist that usher in adolescent depression, even a bipolar disorder.

The psychopathological differentiation of anxiety and depressive states has not been entirely resolved. Cognitive factors may differentiate them best (Table 14.6-40). Although recurrent (especially retarded) major depressive disorder is a distinct disorder from anxiety states, at least some forms of depression may share a common diathesis with anxiety disorders, particularly generalized anxiety disorders. Before assigning patients to such a putative mixed anxiety-depressive group (not yet an official nosological entity), the clinician must note that anxiety that arises primarily during depressive episodes is best considered as epiphenomenal to depressive disorder. The same is generally true for anxiety symptoms that occur in a person with depressive disorder who is using alcohol or sedative-hypnotic or stimulant drugs. Finally, anxiety symptoms could be prominent features of mixed bipolar states as well as of complex partial seizures.

Anxiety	Depression
Hypervigilance	Psychomotor retardation
Severe tension and panic	Severe sadness
Perceived danger	Perceived loss
Phobic avoidance	Loss of interest—anhedonia
Doubt and uncertainty	Hopelessness—suicidal
Insecurity	Self-deprecation
Performance anxiety	Loss of libido
	Early-morning awakening
	Weight loss

Reprinted with permission from Akiskal HS: Toward a clinical understanding of the relationship of anxiety and depressive disorders. In *Comorbidity of Mood and Anxiety Disorders*, JF Maser, CR Cloninger, editors. American Psychiatric Press, Washington, DC. 1990.

Table 14.6-40 Unique Cross-Sectional Profiles of Clinical Anxiety and Depression

Physical Disease Somatic complaints are common in depressive disorders. Some, such as vegetative disturbances, represent the hypothalamic pathology that is believed to underlie a depressive disorder. Autonomic arousal, commonly associated with depression, could explain such symptoms as palpitations, sweating, and headache. In some instances the physical symptoms might reflect delusional experiences. The clinician must be vigilant about the likelihood that somatic complaints in depression can also reflect an underlying physical illness. Table 14.6-41 lists the most common medical conditions that have been associated with depression. When depressive symptoms occur in the setting of physical illness, it is not always easy to determine whether they constitute a genuine depressive disorder. Before diagnosing depression, psychiatrists must ensure that they are not dealing with pseudodepression: (1) functional loss due to physical illness; (2) vegetative signs, such as anorexia and weight loss, as manifestations of such an illness; (3) stress and demoralization secondary to the hospitalization; (4) pain and discomfort associated with the physical illness; and (5) medication adverse effects. On the other hand, nonpsychiatric physicians who manage such patients must consider the diagnosis of depression in the presence of persistent anhedonia; observed depressed mood with frequent crying; observed psychomotor retardation or agitation; indecisiveness; convictions of failure, worthlessness, or guilt; and suicidal ideation. The physician should also suspect clinical depression in all patients who refuse to participate in medical care.

Pharmacological	Steroidal contraceptives Reserpine; α-methyldopa Anticholinesterase insecticides Amphetamine or cocaine withdrawal Alcohol or sedative-hypnotic withdrawal Cimetidine; indomethacin Phenothiazine antipsychotic drugs Thallium; mercury Cycloserine Vincristine; vinblastine
Endocrine	Hyperthyroidism and hypothyroidism Hyperparathyroidism Hypopituitarism Addison's disease Cushing's disease Diabetes mellitus
Infectious	General paresis (tertiary syphilis) Toxoplasmosis Influenza; viral pneumonia Viral hepatitis Infectious mononucleosis AIDS
Collagen	Rheumatoid arthritis Lupus erythematosus
Nutritional	Pellagra Pernicious anemia
Neurological	Multiple sclerosis Parkinson's disease Head trauma Complex partial seizures Sleep apnea Cerebral tumors Cerebrovascular disorder
Neoplastic	Abdominal malignancies Disseminated carcinomatosis

Table 14.6-41 Pharmacological Factors and Physical Diseases Associated with Onset of Depression

Diagnosing depression in medically ill elderly patients can be particularly difficult. This task should be undertaken diligently because it was recently reported that (especially in those with cardiovascular disease) mortality is accelerated by depression. Depressed elderly adults often deny being “depressed” but complain of anxiety, fatigue, and worsening memory. Hypochondriacal symptoms and pain are common. Patients may exhibit extreme negativism and querulousness when invited to participate in medical procedures; others develop poor fluid and food intake out of proportion to their physical conditions.

Another important diagnostic problem at the interface of mood disorder and physical disease is the rare development of malignancy in patients with an established mood disorder. Patients who had responded well to a given antidepressant during previous episodes now have an unsatisfactory response to the same medication. Even a small dose may cause such alarming symptoms as agitation, dizziness, depersonalization, and illusions, which might indicate an occult malignancy, perhaps in the abdomen or the brain. The psychiatrist should always be vigilant about the development of life-threatening physical diseases in patients with preestablished depressive disorder.

> A 55-year-old woman had suffered from four previous episodes of severe depression that had responded to ECT or amitriptyline (Elavil) or both in her native city of Cairo, Egypt. She immigrated to the United States at age 43 and encountered several major stresses; including unfamiliarity with English, daughter dating a man of Japanese extraction, and a complicated series of operations for uterine prolapse. Then her husband confessed he had had an affair with a much younger woman. For months the patient had complained of intermittent fatigue and expressed anger toward her husband. Her ensuing fifth depressive episode appeared fully “understandable,” but her physician's prescription of 25 mg of amitriptyline resulted in dizzy spells that culminated in syncope. Paroxetine (Paxil) 10 mg did not fare any better. An extensive medical workup revealed a retroperitoneal lymphoma. She died 6 months later.

Stupor Although less common today, stupor still raises a diagnostic problem in differentiating between a mood disorder and somatic disease as well as other psychiatric disorders. Depressive stupor is relatively easy to distinguish from so-called hysterical mutism; in the latter, behavior is meaningfully directed to significant others in the patient's environment. The rubric of catatonic stupor is best reserved for a phase of schizophrenia; in such patients the schizophrenic origin of the catatonia might be apparent from the patient's history. Otherwise, most acute-onset stupors are probably of affective origin. The main differential diagnosis here is from organic stupor (due to drugs or acute intracranial events); the physical and neurological examination is not always decisive in such cases, and diagnosis depends on a high index of suspicion of possible somatic factors.

Depressive Pseudodementia The geriatric equivalent of semistupor in younger persons with depressive disorder, depressive pseudodementia is distinguished from primary degenerative dementia by its acute onset without prior cognitive disturbance; a personal or family history of past affective episodes; marked psychomotor retardation with reduced social interaction; self reproach; diurnal cognitive dysfunction (worse in the morning); subjective memory dysfunction in excess of objective findings; circumscribed memory deficits that can be reversed with proper coaching; and a tendency to improve with sleep deprivation.

Chronic Fatigue Syndrome Chronic fatigue syndrome is a complex differential diagnostic problem in view of the subtle immunological disturbances presumably associated with it. The following self-report by such a patient illustrates many of the uncertainties marking the present knowledge of the interface between the syndrome and mood disorders.

I am a 39-year-old, never-married woman, trained as a social worker, but currently on disability. I have experienced extreme lethargy and fatigue for many years. I have always felt foggy headed and had trouble thinking and concentrating. My complaint is of fatigue, not of depression. My body feels like lead and aches all over. My brain feels achy and sore. I feel much worse in the morning and I can't get out of bed; I feel better at night. I feel bad every day. I ache all over, as though someone had beaten me up. Exercise has been prescribed to me, but it makes me worse. Also, I am very sensitive to hot and cold. My sexual drive is low. I have a general feeling of anhedonia. As far back as I remember—in junior high school—I was always exhausted. I always complained about fatigue, not depression, because that has been the overwhelming problem. I feel the depression is secondary to the fatigue. In high school I was a compulsive overeater and I was bulimic for a few years, but it was never severe and I was only about 10 pounds overweight. In those days I would sleep 10 or 12 hours a night on the weekend and still feel exhausted; I could not get up for school on Monday. As an adolescent, I felt inferior. I couldn't make decisions, I didn't want to go to camp or leave home for long periods of time—I felt so insecure. Recently I had a sleep study done, which showed a short latency to stage REM sleep (49 minutes). I was diagnosed as having dysthymic disorder and began taking antidepressants. When I took tranylcypromine, it was the first time in my life that I felt like a normal person. I could play sports, I had a sex drive, I had energy, and I was able to think clearly. But the benefits lasted for barely 2 months. My response was equally short-lived to phenelzine, imipramine (Tofranil), selegiline (Eldepryl), and bupropion (Wellbutrin). I have not responded to serotonin-specific reuptake inhibitors (SSRIs) at all. I also wish to point out that I had never experienced high periods before I took antidepressants. My main problem has always been one of exhaustion. When I responded to medications, they worked very quickly (within a few days) and I felt great, but they all stopped working after a short time. The dose would be raised, and again I would feel better. Eventually, when I got to high doses, I either could not tolerate the high dose or the drug would no longer help. I have taken different combinations of drugs for 10 years and I haven't been able to feel well for more than 6 weeks at a time. Recently, I went to an immunologist. He said I have an abnormality in regulating antibody production and recommended gammaglobulin shots. They did not help. When I first started working, I always felt tired and foggy headed, so it was difficult to be sharp while at work. At times I would close the door to my office and put my head down. Working has become increasingly difficult for me. I had two great jobs, which I blew. As of last year I had to go on disability. I am desperate for relief, as my condition has drastically affected my life. Disability has been hard for me. I am single and have no other financial resources. I am very despondent, as I feel that my life is passing by without the hope of my ever really improving.

The foregoing clinical picture is compatible with a pseudounipolar or bipolar III disorder as described by the author. Some virologists and immunologists as well as some psychiatrists believe that abnormal substances circulate in the bloodstream supplying the brains of such patients. Industrial toxins have also been suggested.

While awaiting more definitive research on the etiology of chronic fatigue syndrome, the psychiatrist can cautiously consider certain patients for thymoleptic trials. That decision can be bolstered by the following considerations: the patient wakes up with fatigue and dread of facing the day; fatigue is part of a more generalized psychomotor inertia or lack of initiative; fatigue is associated with anhedonia, including sexual anhedonia; and fatigue coexists with anxious and pessimistic ruminations. Although none of the foregoing alone is pathognomonic for depression, in aggregate they point in that direction. The occurrence of hypomanic-like periods (as in the above vignette) further supports the link between some cases of chronic fatigue and mood disorder. Use of antidepressants without sedative effects, given in gradually increasing doses as tolerated, is a rational strategy; lithium and valproate, though not formally tested in such patients, are rational augmentation choices. Several recent neuroendocrine challenge studies suggest that some chronic fatigue patients might have a strong anxiety substrate and could be managed accordingly. This is not to say that chronic fatigue is largely a matter of missed affective diagnoses; yet it would be a pity to miss potentially treatable diagnoses. A family or past personal history of classic affective illness or episodes should strongly weigh in this direction. Obviously definitive data are lacking on the essential nature of chronic fatigue, and practitioners should be guided by their own clinical experience, while awaiting new research developments.

Schizophrenia Cross-sectionally, young patients with bipolar disorder might seem psychotic and disorganized and thus schizophrenic. Their thought processes are so rapid that they may seem loose, but, unlike those with schizophrenia, they display expansive and elated affect, which is often contagious. By contrast, the severely retarded bipolar depressive person, whose affect may superficially seem flat, almost never exhibits major fragmentation of thought. The clinician, therefore, should place greater emphasis on the pattern of symptoms than on individual symptoms in the differential diagnosis of mood and schizophrenic psychoses. No pathognomonic differentiating signs and symptoms exist. Differential diagnosis should be based on the overall clinical picture, phenomenology, family history, course, and associated features. Because the two groups of disorders entail radically different pharmacological treatments on a long-term basis, the differential diagnosis is of major clinical importance. Table 14.6-42 summarizes the author's clinical experience in the area and lists the most common pitfalls in diagnosis. In the past many bipolar patients, especially those with prominent manic features at onset, were labeled as having “acute schizophrenia” or “schizoaffective schizophrenia.” Such misdiagnoses (which typically led to long-term treatment with antipsychotic agents) has been costly in terms of tardive dyskinesia, vocational and social decline, and even suicide. For instance, some patients with postpsychotic depressive disorder of schizophrenia in the DSM-IV scheme (Table 14.6-25) have postmanic depressions that were treated with neuroleptic monotherapy without the benefit of more definitive thymoleptic agents.

Common pitfalls:
- Reliance on cross-sectional rather than longitudinal picture
- Incomplete interepisodic recovery equated with schizophrenic defect
- Equation of bizarreness with schizophrenic thought disorder
- Ascribing irritable and cantankerous mood to paranoid delusions
- Mistaking depressive anhedonia and depersonalization for schizophrenic emotional blunting
- Flight of ideas perceived as loose associations
- Lack of familiarity with the phenomenological approach in assessing affective delusions and hallucinations
- Heavy weight given to incidental schneiderian symptoms

Adapted from Akiskal HS, Puzantian VR: Psychotic forms of depression and mania. Psychiatr Clin North Am 2:419, 1979.

Table 14.6-42 Misdiagnosis of Mood Disorder as Schizophrenia

Modern treatments, which tend to keep many persons with schizophrenia out of the hospital, do not seem to prevent an overall downhill course. By contrast, the intermorbid periods in bipolar illness are relatively normal or even supernormal, yet over time some social impairment may result from the accumulation of divorces, financial catastrophes, and ruined careers. (Although rapid-cycling disorders, which have sharply risen during the past two decades, cause considerable social impairment, mood symptoms are so prominent that differentiation from schizophrenia is generally not difficult; also such patients usually display more classic bipolar phases before the rapid cycling).

Postpsychotic depressions in persons with established schizophrenia are sometimes due to inadequate control of schizophrenic symptomology. In other patients, especially more intelligent young schizophrenic patients, they reflect the experience of losing one's ego and sanity. It would be more meaningful to give such patients a diagnosis of both schizophrenia and a depressive disorder and treat the patient accordingly.

Schizoaffective Disorder As the above considerations suggest, depression in the setting of a schizophrenic disorder does not necessarily constitute a distinct nosological entity. The concept of schizoaffective (or cycloid) psychosis should be restricted to recurrent psychoses with full affective and schizophrenic symptoms occurring nearly simultaneously during each episode. This diagnosis should not be considered for a mood psychosis in which mood-incongruent psychotic features (e.g., schneiderian and bleulerian symptoms) can be explained on the basis of one of the following: (1) affective psychosis superimposed on mental retardation, giving rise to extremely hyperactive and bizarre manic behavior; (2) affective psychosis complicated by concurrent brain disease, substance abuse, or substance withdrawal, known to give rise to numerous schneiderian symptoms; or (3) mixed episodes of bipolar disorder (which are notorious for signs and symptoms of psychotic disorganization). Official diagnostic systems such as DSM-IV use the category of schizoaffective disorder broadly. Thus patients with clear-cut manic episodes receive a schizoaffective diagnosis if delusions or hallucinations occur in the interepisodic period, in the absence of prominent affective symptoms. Many psychotic symptoms in mood disorders are often explanatory (albeit delusional), whereby the patient tries to make sense of the core experiences of the affective illness. In patients with recurrent episodes, delusional thinking can be carried over into the interepisodic period. Such patients are thus delusional in the absence of prominent mood symptoms and technically (i.e., by research diagnostic or DSM-IV criteria) might be considered schizoaffective.

The author does not concur with that convention. Affective illness is typically a lifelong process, and limiting its features to discrete episodes is artificial. Although neuroleptic agents might be prescribed on an as needed basis to reduce the strong affective charge of those interepisodic delusions, they do not effectively eliminate the affect-laden experiences. Continued thymoleptic treatment (resorting to ECT, if necessary) and an empathic psychotherapeutic approach are more rewarding in

the long run.

A 29-year-old female college graduate, mother of two children and married to a bank president, had experienced several manic and retarded depressive episodes that had responded to lithium carbonate. She was referred to the author, because she had developed the delusion that she had been involved in an international plot. Careful probing revealed that the delusion represented further elaboration, in a rather fantastic fashion, of a grandiose delusion she had experienced during her last postpartum manic episode. She believed she had played an important role in uncovering the plot, thereby becoming a national hero. Nobody knew about it, she contended, as the affair was top secret. She further believed that she had saved her country from the international scheme and suspected that she was singled out for persecution by the perpetrators of the plot. At one point she had even entertained the idea that the plotters sent special radio communications to intercept and interrupt her thoughts. As is typical in such cases, she was on a heavy dosage of a lithium-antipsychotic combination. The consultation was requested because the primary mood symptoms were under control and yet she had not given up her grandiose delusion. She flippantly remarked that one must be “crazy” to believe in her involvement in an international plot, but she could not help but believe in it. Over several months, seen typically in 60-minute sessions weekly, the patient had developed sufficient trust that the author could gently challenge her beliefs.

She was in effect told that her self-professed role in the international scheme was highly implausible and that someone with her superior education and high social standing could not entertain a belief, to use her own words, “as crazy as that.” She eventually broke into tears, saying that everyone in her family was so accomplished and famous that to keep up with them, she had to be involved in something grand; in effect, the international scheme, she said, was her only claim to fame: “Nobody ever gives me credit for raising two kids, and throwing parties for my husband's business colleagues. My mother is a dean, my older brother holds high political office, my sister is a medical researcher with five discoveries to her credit [all true] and who am I? Nothing. Now, do you understand why I need to be a national hero?” As she alternated, over subsequent months, between such momentary flashes of insight and delusional denial, antipsychotic medication was gradually discontinued. Maintained on lithium, she now only makes passing reference to the grand scheme. She was encouraged to pursue her career goal toward a master's degree in library science.

The vignette illustrates how phenomenological understanding, rational pharmacotherapy, and practical psychotherapeutic or vocational guidance can be fruitfully combined in the approach to patients with psychotic mood disorders. At a more fundamental level it suggests that clinical diagnoses in psychiatry cannot be based entirely on operational criteria; one's opinion of patient's illnesses is not infrequently changed by their response to treatment.

In the author's view, DSM-IV represents something good (operationalization of diagnostic criteria) carried to extreme (arbitrary precision often divorced from clinical reality).

ICD-10

The ICD-10 criteria for mood disorders, which are used throughout the world, are listed in Table 14.6-43. Although these criteria derive in part from DSM-III-R, they are more flexible and clinician-friendly: they do not pretend to impose arbitrary precision on the clinical universe of psychiatry.

Table 14.6-43 ICD-10 Diagnostic Criteria for Mood [Affective] Disorders

SUGGESTED CROSS-REFERENCES

Diagnosis and psychiatry are discussed in Chapter 7, the clinical manifestations of psychiatric disorders are covered in Chapter 8, and the classification of mental disorders is presented in Chapter 9. Schizophrenia is the subject of Chapter 12. The somatic treatment of mood disorders is discussed in Section 14.7 and Section 14.8. Psychotherapy is covered in Section 14.9. Mood disorders and suicide in children are the topic of Chapter 45. Anxiety disorders are presented in Chapter 15, and mood disorders in geriatric psychiatry are discussed in Section 51.3d. Somatoform disorder including neurasthenia and chronic fatigue syndrome is covered in Chapter 16.

SECTION REFERENCES

Akiskal HS, Hantouche EG, Bourgeois ML, Azorin JM, Sechter D, Allillaire JP, Lancrenon S, Fraud JP, Châtenet-Duchêne L: Gender, temperament, and the clinical picture in dysphoric mixed mania: Findings from a French National Study (EPIMAN) J Affect Disord *50:*175, 1998.

*Akiskal HS: The prevalent clinical spectrum of bipolar disorders: Beyond DSM-IV. J Clin Psychopharmacol *16*(Suppl):4S, 1996.

Akiskal HS, Cassano GB, editors: *Dysthymia and the Spectrum of Chronic Depressions.* Guilford Press, New York, 1997.

Akiskal HS, Puzantian VR: Psychotic forms of depression and mania. Psychiatr Clin North Am *2:*419, 1979.

Angst J: The emerging epidemiology of hypomania and bipolar-II disorder: The Zurich study. J Affect Disord *50:*143, 1998.

Barefoot JC, Schroll M: Symptoms of depression, acute myocardial infarction and total mortality in a community sample. Circulation *93:*1976, 1996.

*Baschetti R: Psychological factors and chronic fatigue. NZ Med J *112:*58, 1999.

Bauer MS, Calabrese JR, Dunner DL, Post RP, Whybrow PC, Gyulai L, Tay LK: Multi-site data reanalysis: Validity of rapid cycling as a course modifier for bipolar disorder in DSM-IV. Am J Psychiatry *151:*506, 1994.

Beard GM: *A Practical Treatise on Nervous Exhaustion (Neurasthenia): Its Nature, Sequences, and Treatment.* Wood, New York, 1881.

Clayton PJ: The sequelae and nonsequelae of conjugal bereavement. Am J Psychiatry *136:*1530, 1979.

Coryell W, Winokur G, Shea T, Maser J, Endicott J, Akiskal HS: The long-term stability of depressive subtypes. Am J Psychiatry *151:*701, 1994.

Davidson JR, Miller RD, Turnbull CD, Sullivan JL: Atypical depression. Arch Gen Psychiatry *39:*527, 1982.

Dunner DL, Kai Tay L: Diagnostic reliability of the history of hypomania in bipolar II patients with major depression. Compr Psychiatry *34:*303, 1993.

Hantouche E, Akiskal HS, Lancrenon S, Allilaire JF, Sechter D, Azorin JM, Bourgeois M, Fraud JP, Châtenet-Duchêne L: Systematic clinical methodology for validating bipolar II disorder: Data in mid-stream from a French national multi-site study (EPIDEP). J Affect Disord *50:*163, 1998.

James W: *The Varieties of Religious Experience.* Random House, New York, 1902.

*Judd LL, Akiskal HS, Maser JD, Zelle PJ, Endico HJ, Corey W, Parks MP, Kenovac JL, Leon AC, Mueller TJ, Rice JA, Keller MB: A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. Arch Gen Psychiatry *55:*694, 1998.

Karam EG: The nosological status of bereavement-related depressions. Br J Psychiatry *165:*48, 1994.

Kendler KS, Karkowski LM, Corey LA, Neale MC: Longitudinal population-based twin study of retrospectively reported premenstrual symptoms and lifetime major depression. Am J Psychiatry *155:*1234, 1998.

Klein DF: Endogenomorphic depression. A conceptual and terminological revision. Arch Gen Psychiatry *31:*447, 1974.

*Kraepelin E: *Manic-Depressive Insanity and Paranoia*, GM Robertson, editor, RM Barclay, translator. Livingstone, Edinburgh, 1921.

Manning JS, Haykal RF, Connor PD, Akiskal HS: On the nature of depressive and anxious states in a family practice setting: The high prevalence of bipolar II and related disorders in a cohort followed longitudinally. Compr Psychiatry *38:*102, 1997.

McElroy SL, Keck PE, Pope HGJ, Hudson JI, Faedda GL, Swann AC: Clinical and research implications of the diagnosis of dysphoric or mixed mania or hypomania. Am J Psychiatry *149:*1633, 1992.

MacKinron DF, Xu J, McMahon FJ, Simpson SG, Stine OC, McInnis MG, De Paulor JR: Bipolar disorder and panic disorder in families: An analysis of chromosome 18 data. Am J Psychiatry *155:*629, 1998.

*Pagani M, Lucini D: Cronic fatigue syndrome: A hypothesis focusing on the autonomic nervous syndrome. Clin Science *96:*117, 1999.

Perugi G, Akiskal HS, Lattanzi L, Cecconi D, Mastrocinque C, Patronelli A, Vignoli S: The high prevalence of soft bipolar (II) features in atypical depression. Compr Psychiatry *39:*73, 1998.

Reynolds CF, Hoch CC, Kupfer DJ, Buysse DJ, Houck PR, Stack JA, Campbell DW: Bedside differentiation of depressive pseudodementia from dementia. Am J Psychiatry *145:*1099, 1988.

Rosenthal NE: *Winter Blues.* Guilford, New York, 1993.

Roth M, Gurney C, Garside RF, Kerr TA: Studies in the classification of affective disorders: The relationship between anxiety states and depressive illness I. Br J Psychiatry *121:*147, 1972.

Schatzberg AF, Rothschild AJ: Psychotic (delusional) major depression: Should it be included as a distinct syndrome in DSM-IV? Am J Psychiatry *149:*733, 1992.

*Widlocher DJ: Psychomotor retardation: Clinical, theoretical, and psychometric aspects. Psychiatr Clin North Am *6:*27, 1983.

Winokur G, Turvey C, Akiskal HS, Coryell W, Solomon D, Leon A, Mueller T, Endicott J, Maser J, Keller M: Alcoholism and drug abuse in three groups—Bipolar 1, unipolars, and their acquaintances. J Affect Disord *50:*81, 1998.

14.7 MOOD DISORDERS: TREATMENT OF DEPRESSION

A. JOHN RUSH, M.D.

Many available options now exist for treatment of mood disorders. Depressive disorders have enjoyed major therapeutic advances during the past decade, and the plethora of new options confronts clinicians with the problem of tailoring them to individual patients. Which strategies (types of treatment), applied in what order (strategic planning), and delivered by which methods (tactics) (e.g., dosage, duration) produce the best results for most patients in the shortest period of time? Specifying the objectives of each phase of treatment and careful, timely reappraisal of whether they are met, optimize patient outcomes.

Specific objectives of each treatment phase (i.e., acute, continuation, maintenance) provide a strategic map for managing these patients. For mood disorders, initial treatment objectives include (1) symptom remission (acute phase) and restoration of psychosocial functioning (acute and continuation phases), (2) prevention of a relapse (continuation phase), and (3) prevention of recurrences, or new episodes in patients with recurrent depressions (maintenance phase).

STRATEGIES

When initiating acute-phase treatment, practitioners decide where the patient should be treated (e.g., outpatient, day hospital, or inpatient). Treatment location is dictated by factors such as (1) the imminent risk of suicide, (2) the capacity of the patient to recognize and follow instructions or recommendations (adherence, psychosis), (3) the level of psychosocial resources, (4) the level of psychosocial stressors, and (5) the level of functional impairment.

Next, one chooses among the four common acute-phase treatments (medication, psychotherapy, the combination of medication and psychotherapy, or electroconvulsive therapy [ECT]). For some, light therapy alone or in combination with medications may also be an option.

In general, patients who respond to acute-phase medication (alone or combined with psychotherapy) receive continuation-phase medication at the same dosage. Continuation-phase ECT may be indicated for acute-phase ECT responders if continuation-phase medication does not prevent relapse or if prior medications have been ineffective, although the efficacy of this approach rests on open case series rather than randomized controlled trials of continuation-phase ECT.

While no randomized controlled trials of continuation-phase psychotherapy alone are available, a few open studies suggest that patients responding to acute-phase psychotherapy alone may further benefit from continuation-phase psychotherapy at less frequent intervals for the subsequent 6 to 8 months. The comparative efficacy of the combination of medication and formal psychotherapy during the continuation phase versus continuing medication only has not been investigated.

TACTICS

Tactics are devised to ensure an adequate treatment trial (e.g., adequate dosage and duration of treatment). Adequate implementation is required to determine whether any strategic choice was correct. Adherence is the most important tactical issue. Low adherence may be due to adverse effects, the conscious or unconscious meaning of taking medication, or the desire to leave treatment once improved (perhaps because of the shame and stigma that still surround psychiatric disorders).

A second key tactical issue is adequate evaluation of whether the objective (i.e., symptom remission) was met. Symptom severity may be gauged by careful interviewing or by use of a rating scale. For mood disorders, a serious difficulty is accepting a partial response in place of a full remission, a risk for any acute-phase treatment. A full remission carries a better prognosis and minimal residual symptoms.

STRATEGIC CHOICES

Medication The available antidepressants differ in their pharmacology, drug-drug interactions, and short- and long-term adverse effects. They do not differ in overall efficacy, speed of response, or long-term effectiveness. Substantial evidence shows that failure to tolerate or respond to one medication does not imply failure with other medications. In fact, a shift from one medication class to another carries 1 in 2 chance of response to both the initial medication and to the next medication if the first fails to provide a satisfactory response.

Psychotherapy Formal psychotherapy aims at particular objectives. General clinical management, part of any treatment, includes explaining the diagnosis, treatment plan, treatment objectives, anticipated treatment period, counseling, management of both adherence and adverse effects, and a regular assessment of whether or not the treatment objectives are being met. It may involve consulting both the patient and significant others.

The objectives of formal psychotherapy used alone to treat mood disorders are identical to those for medication: (1) symptom remission; (2) psychosocial restoration; and (3) prevention of relapse or recurrence. When used in combination with medication, psychotherapies can achieve such additional objectives as reducing the secondary psychosocial consequences of the disorder (e.g., marital discord, occupational difficulties) or increasing medication adherence.

Clinical management aims to increase adherence, but formal psychotherapy can also be beneficial. Individuals who may need more-formal adherence counseling include those with significant prior or current adherence difficulties and those with relatively fixed negative attitudes toward a clearly indicated treatment. Formal psychotherapy to address the psychosocial consequences of the disorder may include individual, family, couples, or occupational approaches. Evidence suggests that used in combination with medication to control symptoms, such treatments improve the targeted difficulty (e.g., marital counseling improves marriages).

Psychotherapy as monotherapy for symptom remission has shown greater efficacy than wait-listing controls in studies of less severely or chronically ill, nonpsychotic, depressed outpatients. In addition, while some evidence suggests that psychotherapy alone as a maintenance treatment has some benefit in prolonging the well interval, in general, when maintenance treatment is anticipated, medications (alone or combined with psychotherapy) are preferred, given the larger number of medication maintenance studies supporting efficacy.

Choosing Among Psychotherapies No established clinical predictors exist to guide selection of a psychotherapy. Cognitive therapy may be slightly less effective in those with more-dysfunctional attitudes, while interpersonal psychotherapy may be somewhat less effective in those with more-interpersonal problems. However, these predictors lack clinical utility. Time-limited therapies are usually preferred over time-unlimited therapies for symptom reduction because they have established efficacy (time-unlimited therapies do not) and because medication is an effective alternative if psychotherapy alone fails.

Some believe that reconstructive (time-unlimited) psychotherapies are more useful in the treatment of Axis II disorders, while reeducative therapies may be more useful with Axis I conditions. No evidence favors use of psychotherapy alone over medication when a concurrent Axis II disorder exists. On the other hand, psychotherapeutic tactics may benefit the medication management of depressed patients with Axis II conditions by ensuring adherence. Logically, psychotherapy, if

used alone, should be tried for a finite time period and outcome should be evaluated, just as with medication.

Declaration of psychotherapy failures is largely based on lack of efficacy, although a few patients discontinue treatment unilaterally. When to declare psychotherapy a failure is a complex problem. Some patients respond early, while others may take 8 to 10 weeks. The premature discontinuation rate may be higher in actual practice than it is in efficacy trials. Just as with medication, if a symptomatic patient inappropriately discontinues treatment, one should actively attempt to reengage them since the depression has not remitted and, consequently, the prognosis is poor.

What treatment should follow if psychotherapy alone is ineffective? Medication, given its established efficacy, is the next best logical step. The psychotherapy may be continued or discontinued when medication is begun. Whether a different form of psychotherapy would be effective if the initial psychotherapeutic approach has not been tested.

COMBINED TREATMENT

Medication and formal psychotherapy are often combined in practice, yet data from randomized controlled trials suggest that the combination does not predictably add to the symptom-reducing effects of either treatment alone, at least in less complex, chronically ill patients. Conversely, the combination may result in both symptom reduction and psychosocial restoration, which is an additional rationale for using the combined approach. There are basically three ways to develop a combined treatment: (1) initiate the combination as acute-phase treatment, (2) add formal psychotherapy to medication that has elicited a partial response (particularly when there are residual cognitive, psychological, or interpersonal symptoms or difficulties), or (3) add medication after a partial response to psychotherapy alone.

Using the combination of medication and formal psychotherapy at the outset of acute-phase treatment is called for if either (1) formal psychotherapy is used to increase adherence, while medications are used for symptom control or (2) if the targets of each treatment were somewhat distinct and both needed early remediation (e.g., medication for the depressive symptoms and psychotherapy for marital problems). In addition, clinical experience suggests that combination treatment may be preferable to either treatment alone with (1) a coexisting Axis II disorder, (2) a chronic and recurrent pattern with poor interepisode recovery, or (3) a patient who is discouraged and demoralized as well as clinically depressed.

Diagnosis and medication management must allow time for patients with little prior treatment to collaborate in the optimal use of medication. Thus, it is often simpler to initiate medication and clinical management and then determine whether formal psychotherapy is indicated either for complete symptomatic remission or to address psychosocial problems unrelieved by medication. For example, psychotherapy might be added after a partial medication response (e.g., persistence of cognitive and interpersonal difficulties).

When to add psychotherapy to medication is unclear. Evidence suggests that psychosocial and occupational improvements follow response. Thus, routine use of both treatments initially may not be necessary for psychosocial restoration. The need for adjunctive psychotherapy to redress psychosocial difficulties becomes clearer the longer symptom remission obtains and psychosocial problems persist. A history of long-standing psychosocial difficulties, even during remission of chronic depression, may suggest either beginning with combined treatment or adding psychotherapy shortly after symptoms are controlled with medication.

When combined treatment does not produce a full response, a switch of medication classes with continued psychotherapy is logically the next step, since evidence indicates that switching medication classes is effective.

ELECTROCONVULSIVE THERAPY

ECT is effective, even in patients who have failed to respond to one or more medications or combined treatment. It is effective in both psychotic and nonpsychotic forms of depression. Usually, 8 to 12 treatments are needed. Bilateral ECT is somewhat more effective than unilateral ECT, but it appears to have more cognitive adverse effects.

OTHER TREATMENTS

Light therapy has been most clearly evaluated in mood disorder with seasonal pattern, either as monotherapy or in combination with medication. Patients who respond do so within 2 to 4 weeks.

STRATEGIC ISSUES

Role of Diagnosis in Treatment Selection Maintenance medication effectively prevents recurrences of dysthymic disorder, complicated by recurrent major depressive episodes or not. Psychotic depression usually requires both an antidepressant and an antipsychotic agent. Alternatively, ECT is useful in psychotic depression, either as a first-line treatment or after medication has proven ineffective. For those with atypical features, strong evidence indicates that tricyclic drugs are less effective than the monoamine oxidase inhibitors (MAOIs). There is some suggestion of efficacy for the selective serotonin reuptake inhibitors (SSRIs) in atypical depression.

The concurrent presence of another disorder may also affect initial treatment selection. Presence of nonmood Axis I disorder favors use of medications with demonstrated efficacy in both the mood and nonmood disorder. For example, effective treatment of obsessive-compulsive disorder with depressive symptoms usually results in remission of the depression. Similar findings have been reported for anorexia nervosa and bulimia. When panic disorder co-occurs with major depressive disorder, medications with demonstrated efficacy in each condition are preferred (e.g., tricyclic drugs, SSRIs). In general, the nonmood disorder dictates the choice of treatment.

Concurrent substance abuse raises the possibility of a substance-induced mood disorder, which must be evaluated by history or after several weeks of abstinence, since abstinence results in remission of depressive symptoms in substance-induced mood disorders. If significant depressive symptoms continue, even with abstinence, an independent mood disorder is diagnosed and treated.

Axis II disorders frequently accompany mood disorders, but diagnosis of Axis II disorders remains tentative in the presence of a clinical depression. An Axis II disorder should not be mistaken for recurrent major depressive disorder with poor interepisode recovery, since the treatment objectives and strategies differ.

An Axis II disorder does not contraindicate treating the mood disorder, but its presence may prolong the time to acute-phase treatment response, interfere with adherence, or even preclude full symptomatic remission. In general, the presence of Axis II disorders suggests a less optimistic prognosis, because circumstantial evidence suggests that Axis II disorders are risk factors for subsequent relapse or recurrence.

Axis II disorders raise other tactical issues, such as adherence, establishing a therapeutic alliance, or long-term management. In addition, the response to either medication or time-limited psychotherapy is slower, less complete, or both in the presence of an Axis II disorder.

General medical conditions commonly accompany mood disorders and are established risk factors in their development. Recent evidence indicates that a major depressive episode is associated with increased morbidity or mortality of some associated general medical conditions.

Principles that apply to the treatment of depression without a general medical condition generally apply when these conditions are present. However, treatment strategies and tactics are more complex. The initial choice of treatment is influenced by prior response to antidepressant treatments, the relative medical safety of medications, and clinical judgment about whether psychotherapeutic methods might particularly benefit some of these patients. The tactical choice of medications is affected by drug interactions, the pharmacological profile of the compound, the general medical condition, and drug dosing requirements.

Complex, ongoing, stressful life events or social contextual issues (often profoundly disturbing to patients) should not influence whether or not medication is used. Often, patients in major depressive episodes whose symptoms are reduced by medication become less disabled from the mood disorder and are better able to manage these complex life circumstances. On the other hand, chronic, disturbing, life circumstances (e.g., chronic marital discord, spousal abuse) argue for stronger consideration of combined treatment, either initially or sequenced, to obtain both symptom remission and psychosocial restoration. Table 14.7-1 summarizes the

relation between clinical diagnoses and treatment selection.

Diagnosis	Treatment Recommendations
Major depressive disorder (mild- to-moderate severity)	Medication or time-limited, depression-targeted psychotherapies[a] No maintenance-phase treatment
Major depressive disorder, recurrent	Consider maintenance-phase treatment
Major depressive disorder with psychotic features	Antipsychotic plus antidepressant medication Electroconvulsive therapy
Major depressive disorder with melancholic or severe features	Medications essential
Depression with atypical features	Nortricyclic drugs preferred Monoamine oxidase inhibitors have established efficacy
Depression with seasonal pattern	Light therapy or medications
Dysthymic disorder	Medications, time-limited, depression-targeted psychotherapies, or their combination Consider maintenance-phase therapy
Complex[b] or chronic depressions	Medication plus psychotherapy[a]

[a] Interpersonal psychotherapy, cognitive therapy, or behavior therapy.
[b] Complex refers to depression co-occurring with other psychiatric conditions (e.g., anxiety disorders, Axis II conditions).
[c] Psychotherapy may aim at adherence enhancement, symptom reduction, relapse prevention, or psychosocial restoration.

Table 14.7-1 Relation of Diagnosis to Treatment Selection

Selecting Initial Treatment In general, about 45 to 60 percent of all outpatients with nonpsychotic major depressive disorder who begin treatment with medication, psychotherapy, or the combination respond. Consequently, roughly one-half of patients should anticipate a second treatment trial if the initial treatment selected is either intolerable or ineffective. Selection of the initial treatment depends on the chronicity of the condition, the history of recurrences (which predicts the likelihood of future recurrences), family history of illness, symptom severity, associated comorbid general medical or other psychiatric conditions, prior treatment responses to other acute-phase treatments, and patient preference. In general, the less severe, less chronic, and less complex the depression (i.e., less current comorbidity), the greater the role for patient preference, since evidence for selecting between time-limited, depression-targeted psychotherapy and medication is lacking. Furthermore, it is believed that the combination of medication and formal psychotherapy is less likely to be needed for milder, uncomplicated depressions.

Moderate-to-severe mood disorders with prominent chronicity or prior recurrences generally require maintenance treatment. Since medications are the maintenance treatments with established efficacy, medication treatment (alone or combined with psychotherapy) is recommended.

The evidence for the efficacy of medication alone in more-severe depressions is clear; psychotherapy alone is less well studied. Psychotherapy alone appears to be less predictably effective than medication in outpatients with endogenous or melancholic symptom features. Whether psychotherapy alone is effective in depressions with atypical symptom features is under study. However, the MAOIs and SSRIs have established efficacy in this group.

Selecting Second Treatment Options If the first treatment fails (e.g., due to intolerance or lack of efficacy), a strategic decision on the second treatment after the differential diagnosis (including occult general medical condition or substance abuse) has to be reconsidered.

For those receiving medication initially, dose adjustments, extending the trial period, switching to an alternative treatment (either medication or psychotherapy), or adding a second treatment to the initial one are common options. Factors recommending dose escalations are (1) no adverse effects, (2) a prior history consistent with rapid drug metabolism, or (3) low therapeutic blood concentrations. However, blood concentrations of newer-generation medications are related to outcome although they are for desipramine (Norpramin, Pertofrane), imipramine (Tofranil), and nortriptyline (Pamelor). Extending the initial trial is indicated if (1) it has been less than 6 weeks, (2) there is a partial response by 6 weeks, or (3) prior medication trials were unsuccessful and shorter than 6 weeks.

Likewise, partial response to psychotherapy by week 6 argues for extending the trial period. Nonresponse by 8 weeks often predicts an ultimate poor response. Extending a trial of light therapy beyond 3 weeks in nonresponders has not been evaluated. Clinical experience suggests that extending ECT beyond 10 trials with complete nonresponse is unlikely in most cases to elicit a subsequent response, although careful studies are lacking.

The choice to switch from the initial single treatment to a new single treatment (as opposed to adding a second treatment) depends on the philosophy guiding the clinician, the patient's prior treatment history, and other clinical issues. The best-documented augmentation strategies involve inexpensive medicines (e.g., lithium or thyroid hormones) and response, if it occurs, is often within 2 weeks. Conversely, a switching strategy sometimes involves a washout period (e.g., switching from fluoxetine [Prozac] to an MAOI) for safety reasons as well as the need to wait longer than 2 weeks for a full effect. Alternatively, how long to continue augmentation is not clear, and lithium augmentation entails some expense and inconvenience (i.e., blood concentration monitoring).

If the initial trial is the patient's first treatment and other clinical or economic reasons favor monotherapy, switching rather than augmenting is preferred. On the other hand, augmentation strategies, particularly the use of two different medications, seem effective in patients who have failed one or more well-conducted single medication trials. Thus, switching might be preferable for those with only one or two prior treatment attempts, while augmentation is preferable for those who have failed several single-treatment trials. Recent reviews indicate that if the initial medication is ineffective or cannot be tolerated, it is reasonable to switch medication classes. In psychiatric settings, augmentation may be preferable, since more psychiatric patients have failed several adequate prior single treatments.

The value of augmenting medication with psychotherapy is not well evaluated. Many clinicians believe that if the residual symptoms after a partial response to medication are largely cognitive or psychological, either augmentation with psychotherapy or prolonging the initial medication trial are preferred to switching medications or augmenting with another medication, based on the assumption that these symptoms represent residual psychosocial sequelae. On the other hand, if anhedonia persists after an initial medication trial, switching or augmentation with another medication rather than psychotherapy is often preferred since such symptoms suggest ongoing limbic/paralimbic system dysfunction. However, these suggestions are largely based on clinical experience rather than scientific evidence.

TACTICAL ISSUES

The strategic choices of treatment focus on selection of the initial therapy, or for those who fail the initial therapy, the selection of a second treatment option. Implementation of these strategies requires (1) careful attention to adherence, (2) careful evaluation of outcome, (3) proper dosing and duration of the trial, and (4) timely declaration of treatment failure.

Adherence Treatment adherence is increased if patients understand anticipated objectives and common strategies, if fewer daily doses are required (e.g., once-a-day versus three-times-a-day dosing), and no personality disorder is present. Evidence also suggests more-frequent early visits (e.g., weekly versus monthly) improve adherence. Whether other current psychiatric conditions affect adherence is unclear; it is not related to gender, educational level, or socioeconomic status. The best predictor of future adherence is prior adherence.

Thus, general clinical management of medication treatment should include discussions with patients (and, potentially, significant others) about the objectives of treatment, anticipated treatment period, and adherence obstacles. It is best to anticipate and identify obstacles to adherence prior to prescribing medication or initiating psychotherapy and to make adherence checks a routine part of each visit.

Initially, visits should be frequent enough to ensure adherence and permit timely intervention for adverse effects. Several brief telephone contacts during the initial weeks of treatment help adherence by reassuring patients, ensuring that adverse effects are avoided, countering demoralization and pessimism that impairs adherence, and providing information to overcome short-term concentration and recall problems that are part of depressive episodes.

Choosing Among Medications If medication (alone or in combination with psychotherapy) is part of the first step, the practitioner must select from a variety of available compounds. Medications differ in their short- and long-term adverse effects and spectrum of action but not in overall efficacy or speed of response. If maintenance medication is anticipated, long-term adverse effects are more important than short-term effects in selection (e.g., tertiary tricyclic drugs are associated with greater weight gain than SSRIs over the long run).

Table 14.7-2 lists commonly used antidepressant agents presently available in the United States and groups them on the basis of their presumed mechanisms of action (e.g., presynaptic or postsynaptic activity). However, as basic neuroscientific knowledge expands, further actions will likely be discovered. For example, the

number of serotonin receptor types has increased faster than our understanding of their physiological roles. Further, the actions of some (e.g., venlafaxine [Effexor]) are affected by the dosages used or levels attained in the central nervous system (CNS). Venlafaxine exerts proportionally more serotonin than norepinephrine reuptake blockade at lower dosages than at higher dosages.

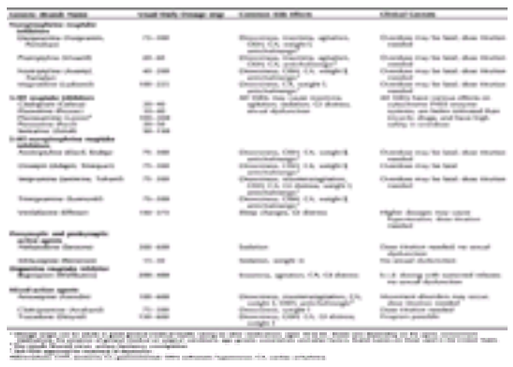

Table 14.7-2 Antidepressant Medications*

Table 14.8-2 lacks two commonly used drug combination treatments (lithium plus antidepressants and liothyronine [Cytomel] augmentation of tricyclic drugs). A Table 14.8-2 provides selected clinical caveats. This list is not exhaustive. Drug-drug interactions, at either the neuronal level (pharmacodynamics) or at the level of absorption, metabolism, and excretion (pharmacokinetics) affect selection of agents, their dosing, and ultimately the risk-benefit equation for individual patients. Patients should be advised what adverse effects to anticipate and encouraged to report them as early as possible. Management of adverse effects may include lowering the dosage, switching medications, or treating the adverse effects with an additional medication.

The MAOIs (tranylcypromine [Parnate] and phenelzine [Nardil]) still have a role in the treatment of depression, the depressed phase of bipolar disorders, and in depressions unresponsive to other treatments. However, because of the necessity of regulating diet (Table 14.7-3) and evaluating concomitant medications (Table 14.7-4), the MAOIs are not used as first-line drugs.

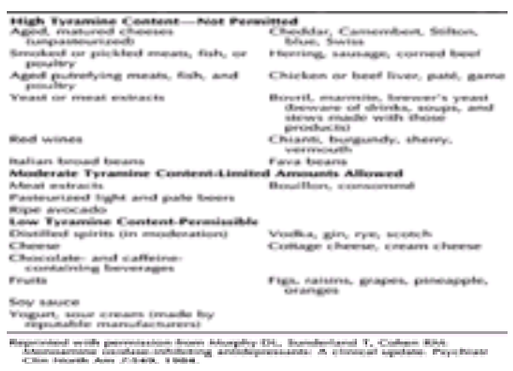

Table 14.7-3 MAOI Dietary Restrictions

Table 14.7-4 MAOI Drug Incompatibilities

Among the tricyclic medications, the secondary amines (desipramine or nortriptyline) have fewer adverse effects than the tertiary amines. Since nortriptyline has a well-established therapeutic window, drug concentration monitoring can ensure that patients who need minimal medication exposure obtain a therapeutic concentration. Conversely, the upper limit of the nortriptyline window may be a disadvantage when deciding to switch or augment treatments, as the blood concentrations may be used to declare a nonresponse.

Choosing more-sedating antidepressants (e.g., amitriptyline [Elavil, Endep]) for more-anxious depressed patients, or more-activating agents (e.g., desipramine) for those more psychomotor retarded is not based on evidence of differential efficacy. However, some clinicians believe that such choices based on adverse-effect profiles increase adherence in the initial weeks of treatment. That is, patients with marked insomnia and anxiety obtain some immediate relief from these associated symptoms before the full antidepressant effect of the drug appears and thus are more likely to comply with acute-phase treatment. These clinical observations are not supported by empirical data, however. In fact, attrition associated with paroxetine (Paxil) or fluoxetine (less-sedating drugs) is lower in acute-phase treatment than that with imipramine or amitriptyline (more-sedating drugs). In addition, the longer-term cost of a possibly beneficial short-term adverse effect advantage must be considered. For example, initially sedating antidepressants often continue to be sedating in the longer run, which may lead patients to prematurely discontinue continuation or maintenance phase treatment, thus increasing the risk of relapse or recurrence.

Some practitioners combine adjunctive medications (e.g., hypnotics or anxiolytics) with antidepressants to provide more immediate symptom relief either routinely from the outset or when the need for such adjunctive medications arises in an individual patient. Adjunctive medications used briefly to cover adverse effects to which most patients ultimately adapt can be useful. Conversely, discontinuing adjunctive medications can result in some return of symptoms or adverse effects to which the patient has not adapted.

Several disadvantages are associated with the routine use of adjunctive medications: (1) the potential risk, inconvenience and expense of unnecessary medications (i.e., many patients may not require them); (2) difficulty identifying the cause of medication intolerance or adverse effects (e.g., an allergic rash) when treatment with an antidepressant and an adjunctive medication is begun at the same time; (3) difficulty in judging response to the antidepressant medication alone if adjunctive medication addresses critical symptoms used to gauge the success of acute-phase treatment, (discontinuing the adjunctive medication to see if the apparent response holds on the antidepressant alone may unnecessarily increase the number of visits or delay a timely revision in the treatment plan); and (4) adjunctive medications may cover adverse effects that if observed would lead to either a dose reduction or to switching treatments. For example, a sedative-hypnotic agent used in conjunction with an SSRI may inappropriately delay a strategic decision to either decrease the dosage or switch to an alternative agent.

In addition to adverse effects, medication choice is affected by prior history of response, cross-sectional symptom features, patient preference, dosing convenience (which affects adherence), drug interactions (if patients are, or will be, taking other medications), current general medical conditions (making one adverse-effect profile preferred over another), and a family history of response. A patient's prior treatment history is important because prior response typically predicts current response. In addition, a documented failure on a properly conducted trial of a particular antidepressant class (e.g., SSRIs, tricyclics, or MAOIs) suggests choosing an agent from an alternative class. Switching classes for those who fail on one class appears to be associated with roughly a 50 percent response rate with the second class of drugs.

History of a first-degree relative responding to a tricyclic drug or an MAOI is associated with a better response to the same class of agents in the patient. Whether family history of response predicts response to the newer antidepressant compounds is not known.

Dosage and Duration Tactical issues surrounding medication use include dosing steps, drug metabolism, pharmacokinetics, drug interactions, and adverse effects. The tricyclic drugs typically are initially given at low dosages and increased to the maximally tolerated dosage or (in the case of nortriptyline) until a therapeutic concentration is obtained. Gradual dose escalations are important to ensure adherence and avoid severe initial adverse effects. Thus, the tricyclic drugs require visits roughly once a week for outpatients as dosages are adjusted. Tricyclic blood concentration monitoring may reduce dosage adjustment time. Dosing is less complicated for the SSRIs than for the tricyclic drugs; fewer dose increments are needed, and the proper dosage is attained earlier because of their better adverse-effect profiles. Some newer agents (e.g., SSRIs, bupropion [Wellbutrin]) need fewer dosage adjustments, but with others (e.g., venlafaxine and nefazodone [Serzone]), raising the dosage increases the likelihood of response, so several adjustments are often helpful.

Safety in overdose is an issue, especially early in treatment. Thus, a 1-week prescription is recommended (without refills) so patients return for frequent medication visits when adverse effects and dosage levels are managed. Tricyclic drugs account for a greater percent of completed suicides than the newer agents, which are far safer in overdose.

Evaluation of Outcome The objective of acute-phase treatment (medication, psychotherapy, their combination, or ECT) is symptom remission, not just symptom reduction. Partial response is associated with a stormier prognosis. Thus, careful interviewing for criterion symptoms at each visit is essential. Self-reported or clinician-rated instruments can facilitate this assessment. Often the patient is slower to recognize the early therapeutic effect of the treatment than the clinician. Thus, a clinician-rated scale may be preferred to a self-reported instrument.

Timely Declaration of Treatment Failures Growing evidence indicates that acute phase medication trials should last 6 (and preferably 8) weeks to determine the full extent of symptom reduction attainable, although most (but not all) patients who ultimately respond fully show at least a partial response by weeks 3 or 4 if the dose is adequate during the initial weeks of treatment. Clinical impression and recent reports suggest that no response by 3 to 4 weeks (e.g., <25 percent reduction in symptoms) indicates that a treatment change is needed (i.e., a few patients respond over the next several weeks), assuming an adequate dosage in the initial 3 to 4 weeks.

Each treatment step should be applied optimally (e.g., dosage and duration) to determine its effectiveness. There is a clinically important tension involved in evaluating the initial treatment—providing sufficient treatment for a long enough time to determine whether it is effective, while at the same time not prolonging (or overdosing) ultimately ineffective treatment.

Medication dosage obviously affects clinical outcome and adverse effect burden. Some patients metabolize certain drugs more rapidly or more slowly than others. Slow metabolizers, especially for the more anticholinergic tricyclic drugs, encounter adverse effects earlier in treatment or at lower dosages. High blood concentrations may cause arrythmias, seizures, or delirium. Fast metabolizers may exhibit virtually no adverse effects or benefits, even with rather large dosages. However, adverse effects, especially for desipramine and nortriptyline, do not predict blood concentrations. Indeed, orthostatic hypotension can occur even with low blood concentrations. Such cases argue for the value of a therapeutic blood concentration to determine dosing strategies, especially when tricyclic drugs are used in medically fragile patients.

Patients may not respond to a medication because (1) they cannot tolerate the adverse effects, even with a good clinical response; (2) an idiosyncratic adverse event occurs; or (3) the clinical response is not adequate. Idiosyncratic or serious adverse effects (e.g. seizures, allergic reactions), while rare, are most likely to be encountered in the first several weeks of treatment and often occur with dosage escalation or as medication concentrations rise to a steady-state level. Some of these adverse effects are dose dependent (e.g., sedation) and can be reduced by decreasing the dosage or slowing the rate of escalation. Moderate adverse effects, when encountered, argue for holding the dosage constant and allowing time for physiological adaptation, which often reduces adverse effects. Some adverse effects are less dose dependent (e.g., orthostatic hypotension), and tolerance to them is less likely. In these cases, gradual dosage escalation is less useful, and a change in treatment is often indicated.

Lack of efficacy is the most common reason for medication failure, but this cannot be fully gauged until patients have had several weeks of treatment at adequate dosages (4 to 6 weeks). Thus, careful evaluation of symptoms during acute-phase treatment (whether formally conducted with a rating scale or informally by assessing each criterion symptom of the mood disorder) is a useful gauge of the adequacy of medication response.

CONTINUATION TREATMENT

Continuation treatment typically lasts 4 to 9 months. In theory, the duration depends on an estimate of when the episode would have remitted spontaneously. Thus, patients with longer prior episodes (e.g., 9 to 15 months) who have had only 2 months of a current depression, for example, would be candidates for 5 to 11 months of continuation treatment, assuming that acute treatment lasted 2 months. Follow-up studies of those with psychotic depressions 1 year after acute-phase treatment indicate a poorer prognosis than for nonpsychotic depression. Thus, continuation-phase treatment for psychotic depressions should be longer.

Continuation-phase medication treatment should end with a gradual taper of medication and careful symptom assessment during, and for several months following, discontinuation. When medication is used alone or in combination with formal psychotherapy for acute-phase treatment, continuation medication treatment is recommended, because early medication discontinuation is associated with a higher relapse rate than later discontinuation. Whenever clinically feasible, continuation medication should be at the dosage used during acute-phase treatment. This recommendation is based on evidence from maintenance trials using lower-dosage tricyclic drugs, which suggested a higher recurrence rate than was obtained with full-dosage treatment.

Psychotherapy may be added to continuation-phase medication if psychosocial residua do not remit with medication alone. Whether to continue psychotherapy following response to acute-phase combined treatment is unclear and is entirely a matter of clinical judgment at this time. Continuation-phase psychotherapy alone (after acute-phase response to psychotherapy alone), has only indirect evidence of efficacy.

MAINTENANCE TREATMENT

Strategic Issues Maintenance treatment aims at preventing new episodes (recurrences). It is appropriate for recurrent (but not for single episode) major depressive disorder. Maintenance medication treatment has been found effective in virtually all studies to date. Strong evidence indicates that those with three or more episodes should receive maintenance-phase treatment, and indeed even at 5 years, maintenance medication has prophylactic efficacy.

Whether those with only two major depressive episodes should receive maintenance treatment is less clear. Information that helps with this decision includes poor recovery between the two episodes, presence of two episodes within the last 3 years, or a positive family history for depression or bipolar disorder; any of these factors increase the likelihood of recurrence. However, clinicians and patients must decide collaboratively whether to initiate maintenance treatment or provide more-diligent monitoring with no treatment until a need is established by the development of a new episode. If a new episode develops when the patient is free of treatment, early intervention shortens the length of the new episode.

Tactical Issues An important issue in both continuation and maintenance treatment is symptom breakthrough, which may be modest and time-limited, requiring only minor shifts in the treatment plan (e.g., dosage adjustment, reassurance). On the other hand, if symptom breakthrough is profound or prolonged or does not respond to dosage adjustment and reassurance, it must be treated. No randomized controlled trials have addressed this issue. Perhaps the simplest approach is to augment the current medication with an additional one (e.g., lithium, thyroid hormone, or another antidepressant). If this strategy is effective, then the augmenting medication may be discontinued after a time, to determine whether it is necessary over the longer term. If the augmenting medication fails, then a switch in treatment to another

medication may be needed.

Symptom breakthrough could also be remediated by psychotherapy, but this option has not been formally studied. Perhaps psychotherapy is indicated if the symptoms were caused by disturbed interpersonal relationships or life events (e.g., divorce or unemployment).

Another tactical problem encountered in both continuation and maintenance treatment is the management of the depression when intercurrent general medical illnesses requiring medication, surgery, or pregnancy occur. For patients who need a “window in time” for surgery or pregnancy, medication discontinuation should be gradual. Pregnancy lasts for a prolonged period and given the evidence for the efficacy of interpersonal therapy alone as a maintenance treatment, psychotherapy without medication may provide an extended drug-free period. The development of other general medical illnesses and the need for nonpsychotropic medications during continuation or maintenance treatment is not uncommon. These circumstances need to be managed with consideration of the pharmacokinetics and drug interactions between the psychotropic and nonpsychotropic agents.

When to discontinue maintenance medication treatment is unclear. As noted above, a recently completed study in patients with highly recurrent depressions (more than three episodes) indicates the efficacy of maintenance treatment continues for at least 5 years. Some patients may require very prolonged (e.g., a decade) or even lifetime maintenance medication treatment. Discontinuation requires careful monitoring, because the first 6 months following discontinuation appear to be a particular risk period for recurrences.

PATIENT PREFERENCE

Patients should become more informed about depressive disorders and their treatment. Even so, some patients are adamantly opposed to medication, while others are equally opposed to psychotherapy. Patient preference can play a greater role when evidence does not strongly support a specific choice. While patients may exercise their first preference initially, a contingency plan should be developed early in the management of the patient in case a second treatment trial is needed. Therefore, it might be wise to plan at the outset for at least two short-term treatment trials, to avoid inappropriate discouragement and consequent premature attrition if the initial treatment fails to provide full remission. Treatment tactics to optimize outcome include attention to adherence, careful titration of medication to attain maximal benefit with minimal adverse effects, and careful symptom evaluation to ensure that remission, not just improvement, has occurred. Establishing explicit goals and following a stepwise plan to attain them can help both practitioners and patients obtain the best outcomes.

SUGGESTED CROSS-REFERENCES

Classification of mental disorders is discussed in Chapter 9, treatment of mood disorders in Chapter 14, psychotherapies in Chapter 30, biological therapies in Chapter 31, mood disorders and suicide in children in Chapter 45, and diagnosis and treatment of psychiatric disorders in late life in Chapter 51.

SECTION REFERENCES

*American Psychiatric Association Task Force: Tricyclic antidepressants—blood level measurements and clinical outcome. Am J Psychiatry *142:*155, 1985.

American Psychiatric Association. Practice guidelines for major depressive disorder in adults. Am J Psychiatry *150*(Suppl): 1, 1993.

Coryell W, Tsuang MT: Primary unipolar depression and the prognostic importance of delusions. Arch Gen Psychiatry *39:*1181, 1982.

*Crismon ML, Trivedi MH, Pigott TA, Rush AJ, Hirschfeld RMA, Kahn DA, DeBattista C, Nelson JC, Nierenberg AA, Sackeim HA, Thase ME: Texas consensus conference panel of medication treatment of major depressive disorder. The Texas Medication Algorithm Project: Report of the Texas Consensus Conference Panel on medication treatment of major depressive disorder. J Clin Psychiatry *60:*142, 1999.

Depression Guideline Panel: *Clinical Practice Guideline. Depression in Primary Care: Volume 2. Treatment of Major Depression.* U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. Rockville, MD, AHCPR Pub no. 93–0551, 1993.

Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ: Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry *47:*1093, 1990.

*Kupfer DJ, Frank E, Perel JM, Cornes C, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ: Five-year outcome for maintenance therapies in recurrent depression. Arch Gen Psychiatry *49:*769, 1992.

Mintz J, Mintz LI, Arruda MJ, Hwang SS. Treatments of depression and the functional capacity to work. Arch Gen Psychiatry *49:*761, 1992.

*Nelson JC, Jatlow PI, Mazure C: Rapid desipramine dose adjustment using 24-hour levels. J Clin Psychopharmacol *7:*72, 1987.

*Robinson DG, Spiker DG: Delusional depression: A one-year followup. J Affect Disord *9:*79, 1985.

Rush AJ: Pharmacotherapy and psychotherapy. In *Clinical Psychopharmacology*, LR Derogatis, editor. Addison-Wesley, Menlo Park, CA, 1986.

*Rush AJ, Crismon ML, Toprac MG, Shon SP, Rago WV, Miller AL, Suppes T, Trivedi MH, Biggs MM, Shores-Wilson K, Kashner TM, Altshuler KZ: Implementing guidelines and systems of care: Experiences with the Texas Medication Algorithm Project (TMAP). J Pract Psychiatry Behav Health *5:*75, 1999.

Rush AJ, Crismon ML, Toprac MG, Trivedi MH, Rago WV, Shon S, Altshuler KZ: Consensus guidelines in the treatment of major depressive disorder. J Clin Psychiatry *59*(Suppl):73, 1998.

Rush AJ, Kupfer DJ: Strategies and tactics in the treatment of depression. In *Treatments of Psychiatric Disorders*, ed 2, vol 1, GO Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

Schatzberg AF, Nemeroff CB, editors: *The American Psychiatric Press Textbook of Psychopharmacology.* American Psychiatric Press, Washington, DC, 1995.

*Thase ME, Rush AJ: Treatment-resistant depression. In *Psychopharmacology: The Fourth Generation of Progress*, F Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Wexler BE, Nelson JC: The treatment of major depressive disorders. Int J Ment Health *22:*7, 1993.

14.8 MOOD DISORDERS: TREATMENT OF BIPOLAR DISORDERS

ROBERT M. POST, M.D.

Treatment of the mood disorders has reached a new level of sophistication based on a variety of advances. The descriptive and diagnostic aspects of bipolar disorders are now explicitly spelled out, recognizing that it is almost always recurrent with potential for severe morbidity and even mortality. There is increasing recognition that bipolar affective disorders have a prominent genetic component interacting with environmental events, and neurobiological alterations have been documented with biochemical assays and functional brain imaging. Thus, different treatment approaches are recognized as a function of type, severity, and course and as in other branches of medicine require the utmost clinical management skills.

Convergent with this increased knowledge about the classification, course, and mechanisms underlying acute episodes and their recurrences is an expanding array of effective psycho- and pharmacotherapeutic modalities and related somatic treatments. Whereas single drugs in one or two classes were available for the treatment of bipolar disorders several decades ago, multiple therapeutic modalities now exist, often including many agents within each class. Thus, the physician must be aware of the nuances in approach to the patient with acute and recurrent mood disorder, so that treatment can be optimized from the outset and the impact of the illness on patients and families minimized.

A growing consensus surrounds a series of new treatment principles. Early recognition and intervention in an acute episode may not only save the patient months of pain and suffering, but also may be life saving. More-careful assessment of the efficacy of an agent at early and regular intervals and early revision of the treatment modality if no improvement is shown are recognized guidelines. This is the case for somatic treatments and also for targeted psychotherapeutic approaches using cognitive, behavioral, and interpersonal therapies.

A substantial body of evidence indicates the efficacy of long-term prophylactic treatments in the management of the recurrent affective disorders. Moreover, earlier institution of long-term prophylaxis is critical to the patient with recurrent mood disorders; it will affect the morbidity of the illness and likely also its subsequent course and treatment responsiveness. Consensus is growing that a patient with a positive family history and a first episode of bipolar disorder is a candidate for continuation therapy after resolution of that episode and for long-term prophylaxis. These factors have dramatically changed the physician's approach to the patient with an acute episode of a mood disorder. The illness should be approached with the same respect accorded the early diagnosis and treatment of a malignancy. Delayed or inadequate treatment may be associated with considerable acute and long-term morbidity from both the illness and its secondary consequences. Thus, it appears appropriate to reconceptualize the recurrent mood disorders not as illusory mental phenomena that can be modified by the patient's will, but as a serious and potentially life-threatening medical illness that has clearly defined mood, cognitive, motor, somatic, and neurobiological concomitants.

Bipolar I disorder occurs in approximately 1 percent of the population, which translates into 2.5 million people in the United States alone. It is estimated that the average woman with onset of bipolar illness at age 25 will lose 14 years of effective lifetime functioning through her illness. Bipolar II disorder and bipolar disorder not otherwise specified may each account for another 1 percent or more. Fifteen to 20 percent of patients with bipolar illness commit suicide. It is against this backdrop of a recurrent, potentially disabling, medical illness that diagnostic and long-term treatment approaches should be conceptualized.

HISTORY

Over the course of this century a revolution has occurred in the treatment of bipolar disorders. In the first half of the century no adequate treatment was available; in the second half, lithium (Eskalith) emerged as a wonder drug for short-term and prophylactic management of the disorder. However, oscillations in the assessment of lithium's safety, efficacy, and utility have persisted. The drug was initially abandoned as unsafe until its concentration could be adequately monitored in blood, which virtually eliminated its most serious and potentially lethal cardiovascular and central nervous system (CNS) toxicities. However, concerns about long-term renal complications have not entirely dissipated after the kidney scare of the 1980s.

There is also greater recognition of lithium's efficacy limitations. As with penicillin, these limitations do not imply that lithium no longer works, but that the spectrum of therapeutic efficacy is narrower. More than 50 percent of patients do not show adequate response to lithium even with adjunctive antidepressant and neuroleptic treatment and using conservative criteria for clinical response (e.g., one episode of illness in a 2-year follow-up).

In the middle of the century electroconvulsive therapy (ECT) emerged as the most effective approach to treating acute episodes. Antipsychotics rapidly became a mainstay of treatment for both mania and psychosis. In the 1960s the first-generation monamine oxidase inhibitors (MAOIs) and tricyclic drugs were introduced and widely used in conjunction with lithium. Now, selective serotonin reuptake inhibitors (SSRIs), bupropion (Wellbutrin), venlafaxine (Effexor), mirtazapine (Remeron), and monoamine oxidase (MAO) type A selective modalities are available. Similarly, the phenothiazine, butyrophenone, and thiothixine antipsychotic classes have given way to the serotonin-dopamine antagonists (atypical antipsychotics), such as clozapine (Clozaril) and risperidone (Risperdal); olanzapine (Zyprexa); quetiapine (Seroquel); and sertindole (Serlect). These new agents include drugs with novel structures and mechanisms of action and more benign adverse effect profiles than the original agents.

This new range of psychopharmacological agents raises a series of important issues for the clinician, particularly when these agents must be chosen on the basis of an inadequate literature on relative efficacy or clinical and biological markers of responsiveness. There is a consensus that with the exception of ECT, no antidepressant modality is more effective or more rapid in onset than another. Thus, the choice of agents is typically based on their adverse effect profile and clinical lore regarding syndromal selectivity of response.

Fortunately, as the limitations of lithium as a mood stabilizer have been increasingly recognized, a variety of other treatment modalities have become available, particularly the anticonvulsants carbamazepine (Tegretol); valproate (Depakene); divalproex (Depakote); as well as the calcium channel inhibitors. Other promising anticonvulsants are being explored as possible third-generation mood stabilizers, including lamotrigine (Lamictal), and possibly gabapentin (Neurontin), and topiramate (Topamax). However, as with targeting therapeutic modalities to specific patients in the depressive disorders, the data are not yet adequate to choose among the accepted mood stabilizers or establish how to use them in combination, which has been increasingly necessary in recurrent bipolar disorders.

Thus, the clinician often has to resort to educated guesses and systematic and sequential clinical trials in individual patients to delineate optimal responsivity (Table 14.8-1). Even with the availability of many new treatments, episodes of illness can often emerge through otherwise partially successful pharmacoprophylaxis and necessitate adjunctive measures. The role of complex combination therapies is well recognized in many branches of medicine and is indispensable in the approach to tuberculosis, acquired immune deficiency syndrome (AIDS), congestive heart failure, or cancer chemotherapy. Systematic research of combination therapies has lagged markedly behind clinical practice, and clinicians are often left to their own devices, without the aid of controlled studies in the literature to guide the optimal

algorithm for approaching the patient who is refractory to standard treatment interventions.

Table 14.8-1 Steps in the Treatment Algorithm of the Bipolar Patient

IMPEDIMENTS TO SHORT- AND LONG-TERM TREATMENT

Although the bipolar disorders are eminently treatable, illness-related variables complicate diagnosis, accessibility to treatment, and the ability of the patient to follow through with treatment. It is estimated that as many as 40 percent of bipolar I disorder patients in community surveys are not in treatment.

Depressed patients often do not recognize that their symptoms are related to a medical illness, and the symptoms themselves (e.g., motor retardation, a sense of inertia, and hopelessness) may preclude the patient's seeking treatment. Thus, the patient's family, acquaintances, and physician may have to actively encourage the patient to initiate treatment.

Treatment must be conducted against the backdrop of the patient's distorted depressive cognitions (e.g., hopelessness, and view of the untreatability of the illness), which must be explained as symptoms of the illness that are not consistent with optimism of treatment response on the basis of the literature and the physician's knowledge base. The therapist's empirical basis for hope of recovery needs to be conveyed to the patient without the promise of immediate results, so that the expected lags in response are not further misinterpreted as a confirmation of the patient's worst fears. Moreover, each phase of the treatment needs continual review in relation to the potential for suicide.

There are major impediments to effective treatment of manic patients. In the early stages of hypomania, the sense of well-being and increased productivity may lead the patient to ignore more severe aspects of the illness, including irritability, argumentativeness, insomnia, poor judgment, and engaging in sexual and other high-risk behaviors without appropriate appreciation of the consequences. These deleterious activities may severely affect the patient's social structure, marriage, and employment. Early recognition that these symptoms and the denial of illness (anosognosia) are components of the illness itself may be crucial to instituting appropriate treatment and preventing escalation to destructive and full-blown manic episodes.

Again, family participation is crucial in both the diagnostic evaluation and the ongoing treatment. The family can assist in overcoming illness denial and thought disorder associated with hypomania and mania, which can be as problematic to receiving adequate treatment as the hopelessness and suicidality of the depression.

Therefore, therapeutic activism, engagement of the family, and early and aggressive treatment of both manic and depressive syndromes are of paramount importance. Individual patients and their families should receive initial and ongoing informational support regarding the medical aspects of the illness, its potential course, and response to treatment, with the long-term goals of increasing compliance, “medicalizing”, and destigmatizing the illness. Destigmatization may become a crucial issue later in therapy when recommendations for long-term prophylaxis may elicit society's negative attitudes toward taking maintenance medications for psychiatric indications (in contrast to most other types of medicine). Conceptualizing the recurrent mood disorders as medical illnesses deserving the same attention, care, and long-term respect as disorders of other organ systems and may help the patient and family accept appropriate long-term treatment options.

A variety of societal, attitudinal, and illness-related variables may interfere with appropriate help-seeking and maintenance behavior in the various treatment phases, including initiation of acute care, continuation treatment, and long-term prophylaxis. During each of these phases patients and their families should be helped to evaluate the medical data and the potential impact of the illness on the patient. Do not introduce all of these variables at the beginning; approach them sequentially in each phase as appropriate. For example, it may be better to discuss the importance of continuation and long-term prophylactic therapy after patients have begun to show an antimanic or antidepressant response, rather than raise this issue with acutely ill patients and possibly frighten them from pursuing further treatment.

Early discussion of long-term prophylaxis—with graphic, statistical, and both written and verbal presentation of the data to the patient and family—may be critical for achieving an optimal outcome. Even in an illness such as juvenile diabetes, where it is unequivocally demonstrated that the patient cannot survive without adequate insulin treatment, many adolescents nevertheless directly or indirectly test the need for insulin and suffer periods of marked hyperglycemia, often requiring hospitalization. In a parallel fashion, patients with bipolar disorder are likely to be tempted to discontinue treatment, especially when the data regarding the morbid or lethal consequences are less well delineated. Nevertheless, the treating clinician must provide the patients and their families with the now overwhelming data showing the high likelihood of a recurrence in a relatively short period of time in patients with one or more prior episodes and the ability of a variety of agents to prevent recurrences of both manic and depressive episodes.

In bipolar disorders the high likelihood of relapse (50 percent in the first 5 months following lithium discontinuation and 80 to 90 percent within the first year and a half) is now also widely recognized and should be explained to the patient. In addition, it has always been assumed that patients who experience a relapse will be readily treatable once their former therapeutic modality has been reinstituted. Most investigators have observed lithium-discontinuation-induced refractoriness in which a small percentage of patients who discontinue successful prophylactic treatment and experience a relapse fail to respond when the treatment is reinstituted. In other instances, patients may not respond as rapidly as they did initially or require increased adjunctive neuroleptic medication.

Many studies report that lithium is less effective in patients who have had more than three or four prior episodes than in those whose prophylaxis is initiated earlier in the illness course. Thus, not only should the potential morbidity and mortality of an episode itself be factored into the decision-making process for long-term prophylaxis, but also the possibility that new episodes could affect the subsequent course of the illness and its pharmacological responsivity.

PSYCHIATRIC HISTORY

A thorough medical history and examination are important, given the many syndromes that mimic both manic and depressive syndromes. The older patient with late-onset illness, in particular, should be approached with the possibility of an associated medical cause, and attention should be paid to obvious or subtle hallmarks of associated pathology. The physician should be alert to symptoms indicating CNS neuropathology, underlying endocrinopathy, or other associated medical illness. Although physicians should aggressively explore these themes with patient and family, they should remember—and even directly tell the patient—that all of the somatic and vegetative symptoms reported are consistent with, and most likely indicate, typical primary affective illness.

The earliest parts of the history can be used to uncover diagnostic clues and to educate the patient about the types of symptoms that are typical of the disorder, are associated with its natural course of spontaneous exacerbation and remission of episodes, and are likely to respond to somatic and pharmacological intervention. The medical history and examination should also seek evidence of cardiac, renal, or thyroid abnormalities that may help guide subsequent treatment choices.

The physician should cover each psychological and somatic symptom category associated with depression while simultaneously educating the patient, providing target symptoms for future assessment of the efficacy of psychological and pharmacological intervention, and constructing the framework for longitudinal monitoring of the patient. The symptoms that are typical for a given patient are likely to be involved in a future episode, and thus they provide an early warning system for illness detection and institution of additional treatment.

A detailed family history of medical and psychiatric illness is also crucial to the initial diagnostic assessment of the patient. Graphic construction of a formal family tree is recommended, with each first-degree relative specifically inquired about for their potential diagnosis, course of illness, and response to therapy (Fig. 14.8-1), since these may help guide the choice of therapies for the patient. A positive family history of bipolar illness may further support the recommendation of long-term prophylaxis after the emergence of the first manic episode.

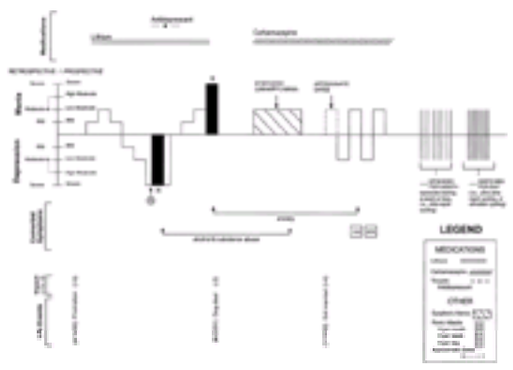

FIGURE 14.8-1 Schema for graphing the course of mood disorders: Retrospective and prospective.

A bipolar disorder family history, especially one with bilineal loading for mood disorder, should markedly raise suspicion of a juvenile- or adolescent-onset bipolar disorder, even if its presentation is less than typical. The clinician should recognize that even in an adult-onset bipolar disorder, the passing of a decade between affective symptom onset meeting diagnostic thresholds and the initiation of treatment is not uncommon. Moreover, in the prepubertal child, a bipolar disorder may present differently from the classic adult picture (Fig. 14.8-2). Instead of showing discrete episodes that easily meet the durational criteria of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), there may be a pattern of tantrums, mood lability, irritability, and marked and rapid fluctuations in mood and behavior. Hypersexuality and grandiosity (if not frank delusions), high-risk behaviors, sleep disturbance, extremes of anger or aggression, or expressions of suicidal ideas may be particular clues that one is dealing with more than attention-deficit/hyperactivity disorder in the hyperactive and inattentive child. Moreover, prepubertal onset of a psychotic depression may herald the beginning of bipolar illness, since a 30 percent switch rate into mania upon treatment with antidepressants has been reported. Similarly, bipolar disorders may present atypically in adolescence, either with extremes of mood lability or with more psychotic and schizophreniform features. In prepubertal children and adolescents, antidepressants may exacerbate the illness, and mood stabilizers, often in combination, are frequently required.

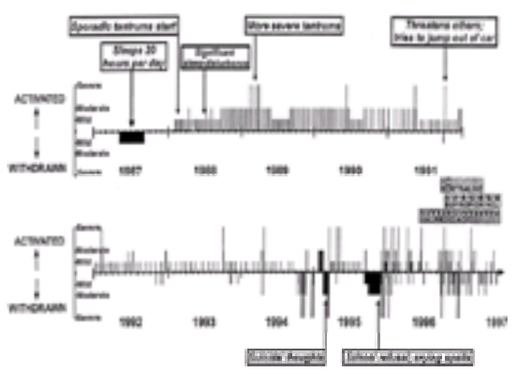

FIGURE 14.8-2 Kiddie Life Chart (K-LCM) of a 10-year-old child with affective dysfunction from the first year of life.

Graphing the Course of Illness The author suggests developing a graphic representation of the patient's prior depressive and manic episodes (Fig. 14.8-3). This graph will form a basis for evaluating the efficacy of previous treatments and assessing current and future prescription. A formal graphic representation of the patient's longitudinal course of illness is useful for several reasons: (1) it provides a clear picture of the earlier illness course, which appears to be the best predictor of the future episode pattern; (2) it clarifies prior and future medication responsiveness; (3) it helps medicalize the patient's history and management, as well as facilitate the recognition of low-level manic symptoms, and (4) it encourages the patient's collaboration and thus may enhance the doctor-patient relationship, making the patient an active partner rather than a passive participant.

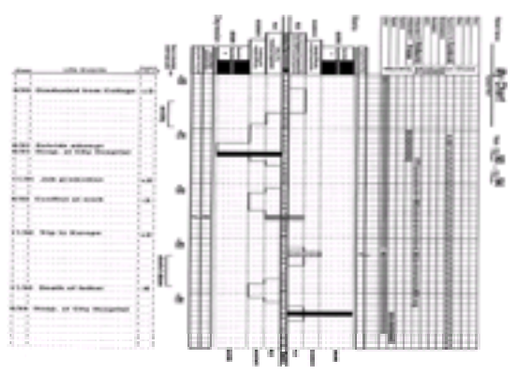

FIGURE 14.8-3 National Institute of Mental Health (NIMH)-life chart method (LCM) self ratings—retrospective.

If a number of past episode recurrences are uncovered in the history, such a graph may also help in the subsequent long-term approach to the illness and in the patient's compliance with prescribed regimens. Moreover, the author has found that this process often uncovers important psychosocial events and possible precipitants of the illness, as well as unique characteristics of the illness, such as cycle characteristics, illness rhythmicity or lack thereof, seasonal variation, relation to anniversaries, and other longitudinal treatment response patterns (such as tolerance or cycle acceleration) not easily uncovered without systematic, graphic representation of the illness. Carefully examining the periods of increased vulnerability to illness provides a template for future treatment intensification or augmentation of therapeutic modalities as appropriate.

With a little practice, the course of an illness can easily be graphically depicted. The author suggests that this be done as part of the initial intake session and be the primary mode of recording a patient's history, even preferable to an extensive written account intended for later conversion to a graph. In this way, both the patient and the physician are immediately and systematically focused on the longitudinal course of the illness and its variation over time, rather than having this focus develop later or possibly be completely sidetracked by attention to acute symptoms. The graphic approach and its associated temporal landmarks can also facilitate recall of important events, dates, and episodes that would otherwise be obscured or forgotten. Typically in this process, supposed first-episode patients will often uncover multiple prior minor or major episodes.

Levels of Severity Physicians can devise their own ways of plotting the course of illness or adopt a system like the one the author and his colleagues have used

successfully over the past decade, that is, graphing three levels (mild, moderate, and severe) of mania or depression on the basis of the degree of associated symptom-driven functional incapacity, which can be most readily assessed retrospectively, on the National Institute of Mental Health NIMH–life chart method (NIMH-LCM) (Fig. 14.8-3).

At the mild level, the patient or family notes a distinct change in the patient's mood, with no notable impairment in the patient's social, educational, or employment roles. This state is readily recognized by depressed patients and may represent the dysthymic baseline from which more severe episodes erupt (i.e., double depression). However, hypomanic patients may deny this mild state, so additional information and input from family members and relatives is important. This observation also reemphasizes the utility of a nonpsychoanalytic approach to the patient's diagnosis and treatment, including the participation and support of family members from the outset. This may be of value both in gaining historical information, and in managing potential suicidality of depression and denial of the adverse consequences of hypomania and mania.

Hypomanic signs and symptoms representing distinct periods of increased energy, productivity, and creativity and decreased need for sleep should be sought directly, and not raised in a negative fashion. These milder periods may be easier to explore after the more severe phases of a patient's illness have been detected and the characteristics of their early presentation agreed upon.

Moderate levels of depression and mania that represent phases with distinct functional impairment are graphed at the next level. Patients can continue their social or employment responsibilities, but only with obvious difficulties, such as absences from work or not being able to perform some routine social tasks. In prospective forms of the NIMH-LCM, moderate dysfunction has been divided into low and high categories, representing some and much dysfunction in usual roles, respectively (Fig. 14.8-4). The manic patient may reveal these levels of dysfunction more easily if asked whether coworkers, friends, or family members are commenting or complaining about the patient's behavior or directing them to seek help.

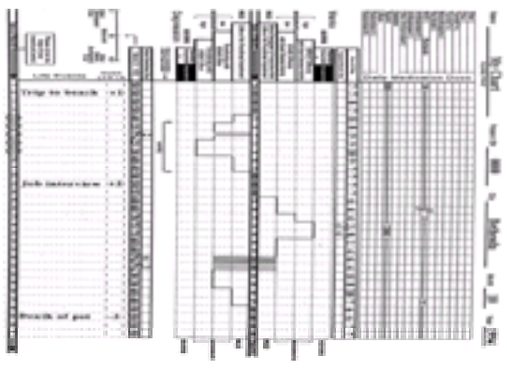

FIGURE 14.8-4 NIMH-LCM self ratings—prospective.

In severe impairment patients are functionally incapacitated and unable to perform their usual roles. Hospitalizations can be coded by shading in the severe manic or depressive episode on the graph. A past episode whose precise timing is unknown or unavailable can be graphed and coded with broken or dotted lines.

For prepubertal onsets a Kiddie-LCM is available that allows graphic depiction of symptom-driven dysfunction whether or not these often highly disturbed and dysfunctional children meet the artificial durational criteria intended for classical adult presentations. Such documentation may be crucial to a child's receiving adequate pharmacological intervention, which is often withheld because of diagnostic controversy and the attendant fears of using medications in children. Professional and societal attitudes and stigma would appear to be highly relevant here, as there is less concern about medications in children with epilepsy, rheumatoid arthritis, infections, malignancies, asthma, and many other medical illnesses with typical or atypical early onsets. Not only is childhood- and adolescent-onset bipolar illness prognostically more disabling than the adult variety in some long-term outcome studies, but the illness also puts the child at increased risk for alcohol and other substance abuse and for other potentially deleterious high-risk behaviors. Moreover, adolescent suicide is one of the fastest growing categories of early mortality.

Psychopharmacological Interventions The history of prior psychopharmacological interventions should be plotted directly above the episodes on this retrospective template of mood fluctuation as illustrated in Figure 14.8-1 of the life-chart schema and in the case example in Figure 14.8-5. When graphed in this fashion, the efficacy of earlier treatments is often more precisely categorized and reclassified. On careful reexamination, a treatment previously deemed ineffective may have partially decreased the frequency or severity of prior episodes. If so, one may now want to consider supplementing this partially effective treatment rather than abandoning it. Previous psychotherapeutic interventions should also be noted, so that their impact on the illness and patient satisfaction can be assessed. As illustrated in Figure 14.8-1 and Figure 14.8-3, important psychosocial events (e.g., significant anniversaries, suicide attempts) and notes about adverse effects (e.g., dosage, reasons for discontinuation of medications) can be entered below the mood graph.

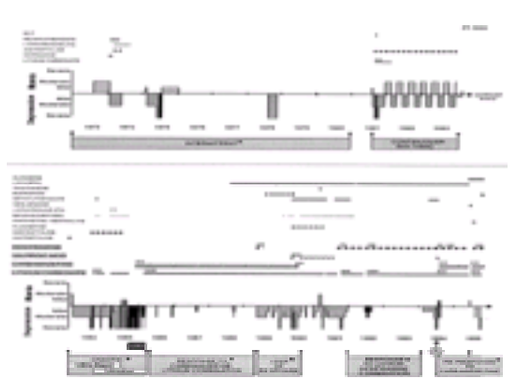

FIGURE 14.8-5 Phases of illness evolution and treatment response in a woman with bipolar II disorder.

Descriptive Symptoms The anamnestic account of specific symptoms and their associated dysfunction also provides the basis for following clinical improvement during an acute episode and during future possible episodes. In particular, one should try to determine the best descriptors and early predictors of an individual patient's episode. For some patients it may be impaired sleep with early morning awakening; for others it may be an inability to concentrate. Still others may experience decreased energy or increased agitation, isolation, anxiety, appetite, or weight gain. The sequential ebbing of symptomatology during a treated episode may also help delineate the need for continued augmentation treatment or reciprocally yield clues to the earliest symptoms of a recurrence. Should these or other residual symptoms break through during prophylaxis or emerge while tapering medication, they can be used to indicate roughening and the need for renewed, more aggressive management of a potential episode.

Similarly, clinicians should contract with their patients in advance about specific symptoms that may herald depressive or manic episodes and require additional monitoring and intervention. Early signs of the emergence of the patients' typical symptoms—such as increased energy or religiosity or decreased sleep—that may be welcomed, may actually presage more consequential manic symptoms. Attention to these early warning signs while insight is still preserved and denial manageable may save patients and their families from a more prolonged or catastrophic episode. A specific contract such as “call if you should have two nights with less than 5 hours of sleep” is often more helpful than a more general suggestion to call if needed.

Prospective Charting For the bipolar patient with several episodes, the author suggests that some elements of this life chart process be continued prospectively, preferably on a daily basis. This can be accomplished quite easily in a number of ways, including having the patient keep a nightly calendar and write a number on a

visual analog scale from 0 (worst ever felt) to 50 (normal or balanced) to 100 (most energized or manic ever). Following discussion with the clinician, this score can be converted to life-charted episodes based on functional incapacity.

In this fashion or with a more complete daily prospective mood chart such as the NIMH-LCM-p patients can track their mood fluctuations in a systematic manner that is unobtrusive and takes only seconds to complete. Pocket life charts and computer graphics packages are now available. Mood rating, like self-assessment of urine glucose concentration by the diabetic patient, may provide an important measure of how well the patient's illness is responding to a given treatment modality, and its dosage adverse-to-effect efficacy ratios and titration. The morbidity and mortality of the mood disorders can be as severe as those of many medical illnesses for which a great deal more attention is paid to the longitudinal and systematic monitoring of fluctuations in symptomatology, biochemistry, and underlying pathology. Patients should be encouraged to help in this life-chart process if they are amenable to it. The completed retrospective and ongoing prospective charts may also help in any future transfers of medical care, orientation of hospital staff, or consultation request should the patient move or suffer future episodes requiring review and revision of treatment.

Subjective and Objective Differences Finally, asking patients to make a calendar, rate their moods (with a specific number on the 100-mm line), and assess illness-driven functional incapacity, may have other secondary benefits. Patients can become better attuned to possible subjective versus objective differences in illness assessment. Some depressed patients can detect mood changes and adverse effects before therapist observation, but many bipolar depressed patients can show remarkable improvement in major aspects of their symptomatology, including sleep, appetite, energy, spontaneity, and sociability, without any subjective sense of mood improvement. If patients do not recognize that their depression is improving despite objective signs, it may lead to further therapeutic pessimism and also may increase the possibility of suicide, as they now may have more energy to carry out such a plan. Moreover, return to previous optimal social and occupational function may lag even further behind objective and subjective appreciation of symptomatic improvement, so that the patient requires additional support, encouragement, and possible retraining during this time frame.

TIME FRAME OF EDUCATION

Although a hopeful perspective about the treatability of a patient's episode should be maintained, a series of drug trials may be needed before the best treatment regimen is found. Evaluation of an acute antidepressant treatment response often requires 3 to 6 weeks, and a given agent's lack of efficacy should be treated as additional information about the patient's illness rather than an indication that the illness is not responsive. Several recent studies in unipolar depression suggest that a lack of some improvement after 4 weeks is a very poor prognostic sign for eventual response with further continuation into 5 and 6 weeks.

Thus, augmentation approaches, in particular, should be triggered at these early times rather than after several months, particularly if high doses or adequate blood concentrations of a given compound have been achieved. In the treatment of acute mania, in which response is typically more rapid (days to weeks), one should consider additional and alternative approaches if improvement is not observed over much shorter periods (days to a week). At the outset of treatment the availability of different effective treatments, with many drugs in each class, should be brought to the patient's attention. This puts possible treatment sequences in their proper perspective and emphasizes that a lack of response to, or an intolerance of, one drug does not portend an ultimate negative therapeutic outcome.

These points should be reemphasized throughout the entire therapeutic process, particularly in light of the different time frames available to the therapist and the patient. The therapist is aware of the many treatment alternatives and the extended treatment course sometimes required to achieve optimal efficacy. The patient (and perhaps even the family) may be overwhelmed by a sense of immediacy and desperation about the current mood-disordered state. Particularly for the depressed patient, this feeling of pain and hopelessness can too easily overtake the realities of the situation and increase the risk of suicide before a positive treatment outcome is achieved.

Reassurance without overpromising immediate therapeutic effects thus appears to be an important part of the treatment. A similar but inverse process may be required for the manic patient, who also may see only the immediate time frame and not the longer-term perspective. The therapist should encourage and help supply the ego for the longer-term view in both of these cases. Thus, supportive, interpersonal, cognitive, and behavioral approaches to the psychopharmacotherapy of the mood disorder may be essential. Patients should also be counseled not to make important long-term decisions on the basis of a distorted view of themselves and their future during an acute depressive or manic episode.

Stressing the time needed to evaluate a given treatment may help maintain the patients' and families' morale and help obtain an adequate informed consent while avoiding malpractice litigation. In this regard, the patient must be told the possible adverse effects of each drug treatment, so that these effects are expected and not seen as dangerously unpredictable, or conversely, they can be recognized as something that merits closer attention, a call to the physician, or an extra visit when indicated.

HOSPITALIZATION

The decision to hospitalize severely depressed or manic patients depends on a variety of clinical and (disappointingly) economic issues. Hospitalization is often indicated for the acutely suicidal patient and may also be considered for the patient with associated medical problems or one who needs close management and monitoring of complicated or novel psychopharmacological regimens. For the knowledgeable patient and supportive family, one or more of a series of psychopharmacological approaches can often be accomplished on an outpatient basis, particularly if there is close coordination between the patient and physician regarding dosage, titration, adverse effects, and indicators of improvement. Despite many of the largely societal criticisms of the modern use of ECT, this modality should be given a higher priority when treating patients with extreme suicidality, associated medical illnesses, difficult adverse reactions to routine psychopharmacological agents, or other medical emergency situations (such as malignant catatonia or hyperthermia) that demand the most rapid treatment response available. For the recurrent manic patient who has refused voluntary hospitalization at the height of an episode or might be likely to do so, obtaining informed consent in advance (during a well interval) for future involuntary hospitalization and pharmacological treatment may avoid many cumbersome practical and medicolegal difficulties should another manic episode requiring hospitalization occur.

PSYCHOTHERAPY AND PHARMACOTHERAPY

Depression is a serious, potentially life-threatening medical illness, and patients and their families deserve much support. Combining psychosocial and pharmacological approaches is important for most because of their potential mutual interaction and support in the context of ongoing pharmacotherapy.

Cognitive-behavioral therapy may enhance medication compliance and provide support for the patient in the interval before psychopharmacological interventions are successful, especially if several agents are required before an effective regimen is found. Moreover, stress and other psychosocial issues may be involved in the onset and recurrence of some depressions and manias and may be indicators for more aggressive pharmacological management.

Frequent meetings with the patient may also help in assessing response to pharmacotherapy and suicide risk. In a depressed patient with severe pain and suffering, frequent meetings may facilitate the fastest maximal application of pharmacological leverage; regimens can be revised as quickly as possible (days to a week in mania, and several weeks in depression) if improvement is not forthcoming. Combined treatment may also be helpful in instances of partial response to monotherapy, a protracted episode, poor interepisode recovery of function, an associated personality disorder, or the presence of acute psychosocial stressors.

NEUROTRANSMITTER THEORIES

Because most of the effective treatments for mood disorders were discovered by serendipity, pharmacological agents have propelled theoretical formulations rather than vice versa. Neurotransmitter theories of the basis of depression have included serotonergic, noradrenergic, cholinergic, dopaminergic, g-aminobutyric acid GABAergic, and, most recently, glutamatergic, each based on presumed mechanisms of effective pharmacotherapeutic interventions. For example, finding that several drugs that acutely potentiated catecholamines and indolamines were antidepressants and that reserpine (which depleted these neurotransmitters) could exacerbate depression and treat mania led to the amine hypotheses of deficiencies in depression and excesses in mania.

Relatively selective manipulations of each of several different neurotransmitter systems (serotonin, noradrenergic, and dopaminergic) now seem to be associated with antidepressant effects (Table 14.8-2 and Table 14.8-3), which raises a critical psychopharmacological question, whether an individual may respond to one type of treatment with a postulated mechanism of action targeting one neurotransmitter system, but not to another (Fig. 14.8-5). In the absence of definitive studies of this question, one is tempted to recommend the sequential use of drugs that act differently within or among classes of agents (e.g., changing from a relatively more

serotonergic to a relatively more noradrenergic tricyclic reuptake inhibitor or from a tricyclic drug to an MAOI to lithium). Few validated clinical or biological markers of response to given treatment agents exist, so one must move through the various treatments or adjuncts for the refractory patient until an effective one is found, largely through trial and error. An MAOI trial should be considered for a patient who has failed to respond to multiple previous antidepressants, as response rates remain 50 to 60 percent with this agent. A similar strategy of using agents with different mechanisms of action sequentially or concurrently in mania may also be warranted (Table 14.8-4 and Table 14.8-5).

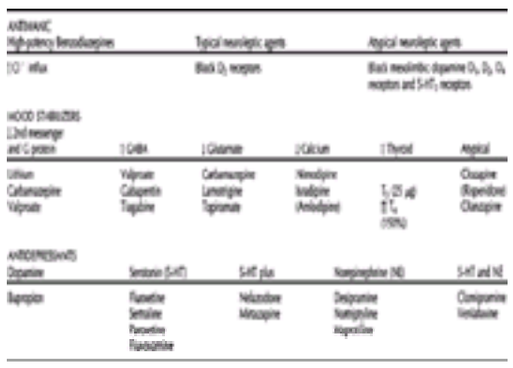

Table 14.8-2 Targets of Action of Antidepressants and Mood Stabilizers

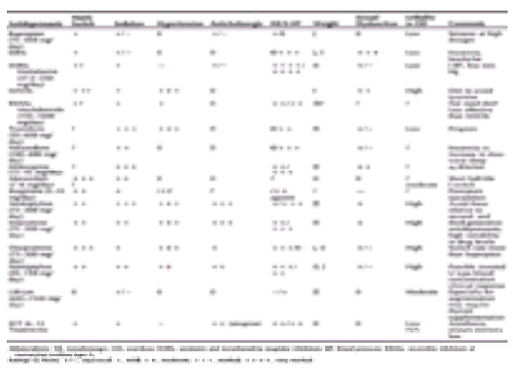

Table 14.8-3 Adverse Effects of Biological Treatments Used to Treat Depression in Bipolar Disorder Patients

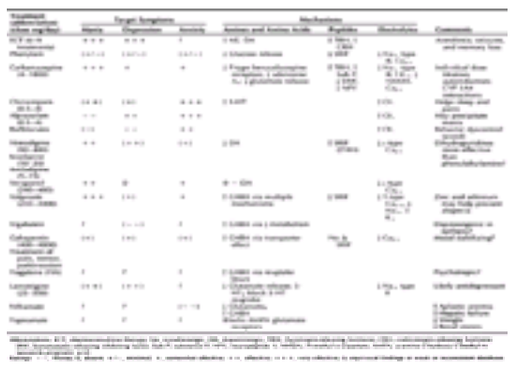

Table 14.8-4 Putative Psychotropic and Mechanistic Effects of the Anticonvulsants

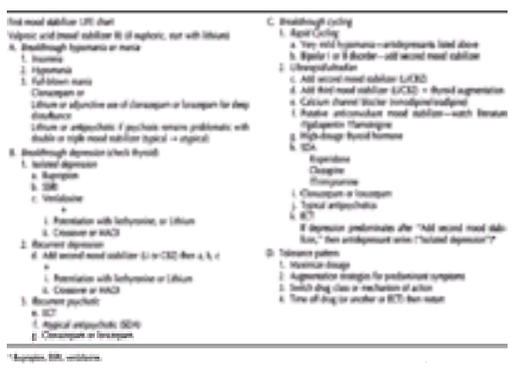

Table 14.8-5 Algorithm for Augmentation and Switching Strategies, Treatment of Bipolar Disorder, Manic Phase, Dysphoric Type

Since 1 month must elapse between treatment with serotonin-active compounds and an MAOI, a transition may be achieved with bupropion, a noradrenergic-selective agent, lithium, or an anticonvulsant. With increasing levels of refractoriness, one should reevaluate the diagnosis (especially with a careful assessment of possible physical, psychological, or drug abuse comorbidity) and consider consultation, psychotherapy revision, as well as more complex combination treatment approaches.

Lithium, valproate, and carbamazepine each alter dopamine and noradrenergic function and upregulate g-aminobutyric acid (GABA) type B ($GABA_B$) receptors in the hippocampus (Fig. 14.8-6). They also target G protein systems, decrease protein kinase C activity, and inhibit calcium influx through the glutamate *N*-methyl-D-aspartate (NMDA) receptor. Lithium and valproate increase myristoylated alanine-rich C kinase substrate (MARCKS) and activator protein–1 (AP-1) binding, as well as messenger ribonucleic acid (mRNA) levels for BCL-2, and brain-derived neurotrophic factor, the latter two of which may be involved in the prevention of apoptosis. GABAergic effects are enhanced by valproate and gabapentin, whereas topiramate, carbamazepine, and lamotrigine decrease excitatory amino acid release or function in part via blockade of sodium ion channels.

FIGURE 14.8-6 Common and different mechanism of mood stabilizers. PI, phosphoinositol; AC, adenylate cyclase; IMPase, inositol monophosphatase; PLC, phospholipase C; cAMP, cyclic adenosine monophosphate; NPY, neuropeptide y; ACh, acetylcholine; DA, dopaminergic; Ne, noradrenergic; PKC, protein kinase C; $A1_R$, adenosine A1 receptors; T_4, thyroxine; CRE, cyclic response element; CBZ, carbamazepine; TRH, thyrotropin-releasing hormone; CREB, cyclic response element binding protein; VPA, valproate; SRIF, somatostatin; t.o., turnover.

TREATMENT OF ACUTE MANIA

Lithium Lithium has been the paradigmatic treatment for acute and prophylactic treatment of mania. In comparative studies with antipsychotic agents, it yields better overall improvement in most aspects of manic symptomatology, including psychomotor activity, grandiosity, manic thought disorder, insomnia, and irritability. However, the onset of antimanic action with lithium can be rather slow (Fig. 14.8-7), even with aggressive dosing; thus acutely deteriorating aggressive or psychotic manic patients may require supplementation of lithium in the early phases of treatment. Until recently, this was traditionally accomplished with the typical neuroleptic drugs, including the phenothiazines, thiothixines, or butyrophenones such as haloperidol (Haldol). Because of the rapidly growing evidence for the parallel acute antimanic efficacy of the mood-stabilizing anticonvulsants carbamazepine (Fig. 14.8-7) and valproic acid, it is suggested that these alternative agents or the more recently available atypical serotonin dopamine antagonists be used as initial supplements rather than in place of the conventional antipsychotics for a variety of reasons related to their adverse effect profiles and the risk of tardive dyskinesia.

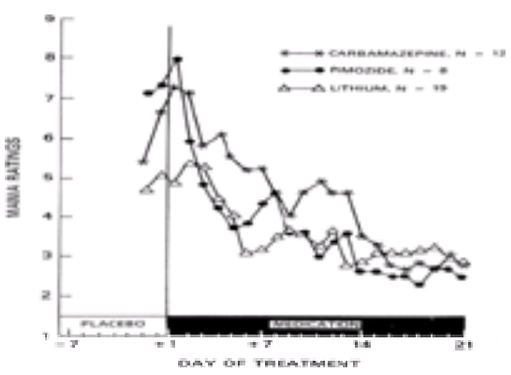

FIGURE 14.8-7 The time course of onset of antimanic response to blind administration of carbamazepine in 12 patients is compared with other similarly diagnosed and rated patients treated with pimozide (*N* = 8) or lithium (*N* = 19).

Double-blind controlled evaluations in many different laboratories have indicated that the onset of antimanic efficacy with carbamazepine is as rapid, or almost as rapid, as it is with traditional antipsychotic drugs, including chlorpromazine (Thorazine), thioridazine (Mellaril), haloperidol, and pimozide (Orap) (Fig. 14.8-7). As of 1998 19 double-blind studies of carbamazepine in acute mania indicated clinical efficacy. Fewer controlled studies have been performed with valproate, but they represent the largest placebo-controlled studies of both lithium and valproate, and they also indicate acute antimanic efficacy. Because initial acute antimanic response may be a guide to subsequent prophylaxis (the major focus of therapeutics of bipolar illness), the author also encourages the investigation of an individual's acute response to these anticonvulsant agents as adjuncts. Antipsychotic agents can always be used later in the sequence if dictated by a lack of clinical response to the mood stabilizers and the adjunctive high-potency benzodiazepines.

Lithium Response The typical clinical profile of the manic patient most responsive to lithium carbonate is one with a classic presentation and euphoric mania, rather than severe or dysphoric mania with paranoid or destructive trends; a pattern of mania followed by a depression and then a well interval rather than a pattern of depression, mania, well interval or continuous cycling; a history of fewer prior episodes and no rapid cycling (i.e., four episodes a year); a positive family history of primary mood disorder in first-degree relatives; and a lack of substance abuse and other comorbidities such as panic disorder. These characteristics can make the difference between a 70 and a 30 percent response rate (Fig. 14.8-8).

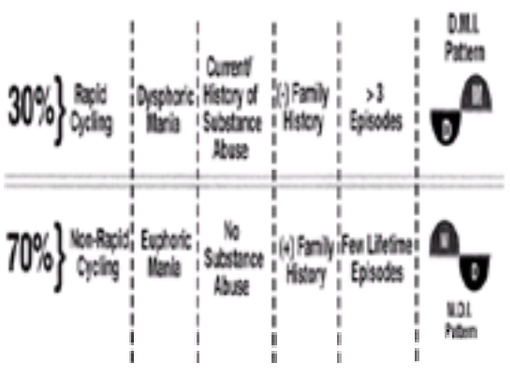

FIGURE 14.8-8 Variable lithium-response rate based on bipolar disorder subtype. D.M.I., depression, mania, well interval; M.D.I., mania, depression, well interval.

Lithium doses should be administered to achieve concentrations in serum between 0.6 and 1.2 mEq per liter. Although a high-dose strategy is advocated by some investigators (concentrations of 1.5 or above), the author has not seen many patients who fail to respond at more typical lithium blood concentrations and respond well when pushed to higher levels that are associated with greater side effects. Dose-limiting adverse effects may include gastrointestinal disturbances (particularly diarrhea) as well as neuropsychiatric syndromes including tremor, subjective sense of cognitive slowing, confusion, and myoclonic twitches. Concentrations of lithium in blood achieved at a given dosage may increase further if the patient switches from mania to depression, leading to greater adverse effects. For the inadequate responder to lithium at levels associated with few side effects, the author recommends potentiation with other agents, rather than discontinuing lithium and requiring a new agent to treat any additional withdrawal emergent symptoms.

The most recent and largest placebo-controlled study of lithium indicated that only 50 percent of patients achieved a 50 percent or greater improvement by 3 weeks. Those with a prior anecdotal history of lithium nonresponse by self-report were particularly at risk for nonresponse in this controlled study. However, even in the responsive group, many symptoms remained at the end of 3 weeks. Thus, even without the increased pressures from managed care for rapid discharge from the hospital, most full-blown manic patients require adjunctive treatment to achieve a timely and adequate antimanic response. A broad range of such options are now

available (Table 14.8-6).

Table 14.8-6 Speculative Differential Approaches to Bipolar Disorder Subtypes

Valproic Acid Typical dosages of valproic acid are 750 to 2000 mg a day, achieving blood concentrations between 50 and 120 μg/mL. Rapid oral loading with divalproex using 15 to 20 mg/kg from day 1 of treatment, has been well tolerated and associated with a rapid onset of response. Blood concentrations above 45 μg/mL have also been associated with earlier response. In several case series, patients with more-typical manic syndromes and fewer schizoaffective symptoms appeared to show a high frequency of response. In contrast to lithium, those with a history of lithium nonresponse, dysphoric mania, or rapid cycling are not less likely to respond to valproate (Table 14.8-6).

Carbamazepine Several preliminary studies have suggested that some of the variables associated with a poor response to or intolerance of lithium may be associated with a good antimanic response to carbamazepine. Thus, the drug may be considered for lithium-nonresponsive manic patients (Fig. 14.8-5).

Typical dosages of carbamazepine to treat mania have ranged between 600 and 1800 mg per day associated with blood concentrations between 4 and 12 μg/mL. However, within this dosage and blood-concentration range, there does not appear to be a clear relation to the degree of clinical response across patients. For an individual patient, however, clinical response and adverse effects are likely dose related. Dosage administration with this anticonvulsant must be individualized, as there is wide variability in the dosage and blood concentration at which adverse effects occur. Increasing the dosage to achieve a clinical effect and titrating the increases against the emergence of adverse effects is the appropriate strategy for such wide individual variability. After several weeks carbamazepine induces hepatic enzymes that lower its levels and may require additional upward dose titration.

Carbamazepine and valproic acid have been used in combination to treat epilepsy, but only preliminary evidence for the efficacy of this combination in acute and prophylactic management of the refractory bipolar patient is available. Valproate may increase concentrations of carbamazepine and its active 10,11-epoxide metabolite, indicating a need for lower carbamazepine dosage when both drugs are used in combination. Typical adverse effects of carbamazepine and other antimanic drugs are listed in Table 14.8-7.

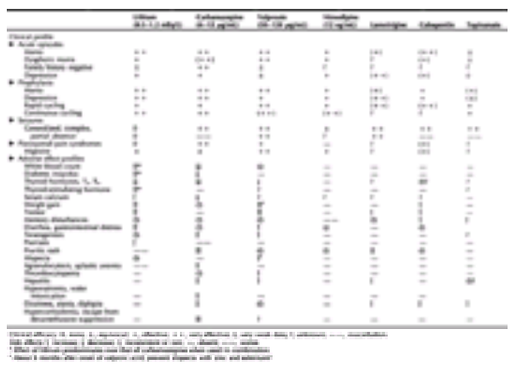

Table 14.8-7 Comparative Clinical and Adverse-Effect Profiles of Lithium, Nimodipine, and the Putative Mood-Stabilizing Anticonvulsants (Preliminary Clinical Impressions)

Clonazepam and Lorazepam The high-potency benzodiazepine anticonvulsants studied in acute mania include clonazepam (Klonopin) and lorazepam (Ativan). Both appear effective and are widely used for adjunctive treatment of acute manic insomnia, agitation, aggression, and dysphoria as well as panic. As noted above, the benign adverse-effect profiles of these agents render them ideal adjuncts to lithium, carbamazepine, or valproate and preferable to the dopamine receptor antagonists (typical antipsychotics) for first-line augmentation. The sedating effects of clonazepam may be problematic in some outpatients, but this property of the drug may be used for bedtime medication for severely insomniac manic patients.

Both of these two high-potency benzodiazepines work rather selectively at the central-type benzodiazepine receptor; in contrast, carbamazepine is not active at this receptor and appears to act at the so-called peripheral-type benzodiazepine receptor (Table 14.8-4). Classic central-type benzodiazepine receptors modulate GABA receptors that facilitate chloride influx and neuronal inhibition (Fig. 14.8-9). In contrast, the peripheral-type benzodiazepine receptor appears to be more closely associated with mitochondrial neurosteroid biosynthesis and calcium channels. These findings are noteworthy in regard to possible differential psychotropic responsiveness between these two classes of anticonvulsants.

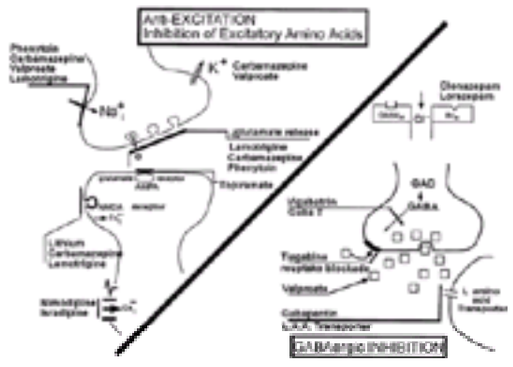

FIGURE 14.8-9 Dual targets of mood stabilizers. Bz_R, benzodiazepine type R receptors, AMPA, a-amino-3-hydroxy-5-methyl-4-isoxazole propionate; GAD, glutamic acid decarboxylase; L.A.A., L amino acids.

Dopamine Receptor Antagonists Studies indicate that short-term use of conventional antipsychotics (Table 14.8-8) in the treatment of mania results in unintended and often unneeded persistence of antipsychotic use 6 months or more after the acute episode. Such intermittent to chronic maintenance treatment with traditional neuroleptic drugs should be avoided, if possible, since bipolar disorder patients are reported to be at high risk for tardive dyskinesia (20 to 40 percent of those

exposed), equal to or greater than that of patients with schizophrenia.

Table 14.8-8 Antipsychotic Drug Use in Mania

Thus, the author recommends exploring potential mood-stabilizing alternatives to lithium (e.g., valproate and carbamazepine), short-term adjuncts (e.g., clonazepine or lorazepam), or serotonin-dopamine antagonists before using dopamine receptor antagonists in bipolar disorder. Intermittent neuroleptic drug use, rather than being a protective factor may actually increase the risk of tardive dyskinesia, according to both clinical and preclinical studies.

The strategy of rapid tranquilization with suprathreshold doses of antipsychotic agents should be avoided. Many double-blind evaluations of this strategy in acutely psychotic schizophrenia patients and one trial in manic patients have shown that the high-dosage strategy (40 to 60 mg a day of haloperidol) is no more efficacious than traditional dosage regimens (10 mg a day) and is associated with greater adverse effects. Particularly for extremely manic patients, the use of heroic antipsychotic dosages to decrease psychomotor activation may not be justifiable considering the added risk of ordinary toxicities, the risk for neuroleptic malignant syndrome, and the sporadic reports of reversible and irreversible organic impairments when used in conjunction with lithium.

Serotonin-Dopamine Antagonists The availability of new atypical antipsychotics with their presumed equal efficacy but superior adverse effect profiles (Table 14.8-9 and Table 14.8-10) may change the algorithm for the treatment of acute mania. A minority of patients appear to require antipsychotics for short- and long-term mood stabilization. Not only can the typical neuroleptic drugs predispose to a variety of acute extrapyramidal effects and long-term tardive dyskinesia, they can also exacerbate the depressive phases of bipolar illness, increasing the frequency or duration of depression or both. These potential liabilities should lead one to use the newer atypical neuroleptic agents if additional antipsychotic treatment is required.

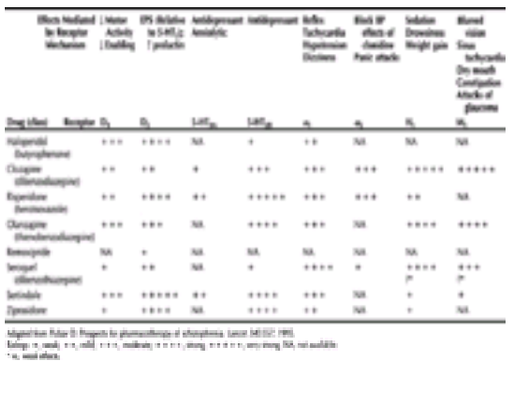

Table 14.8-9 Receptor Activities of Atypical Antipsychotics Relative to Haloperidol

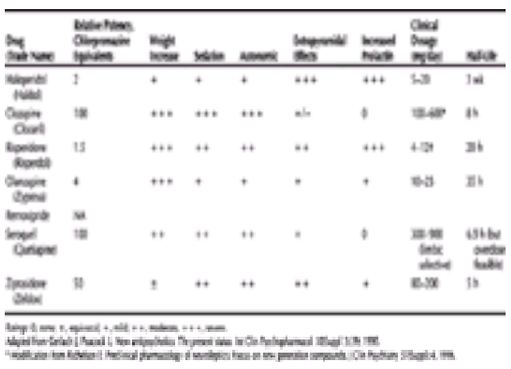

Table 14.8-10 Clinical Profiles of the Atypical Antipsychotics Relative to Haloperidol

Clozapine appears to be particularly efficacious in refractory bipolar I disorder characterized by either dysphoric mania or rapid cycling. Its efficacy in bipolar disorders equals or betters that in schizoaffective disorder and schizophrenia as well. However, it has considerable liabilities: inconvenience, cost, and risk of agranulocytosis, which requires weekly blood monitoring. Olanzapine has a biochemical profile similar to that of clozapine, and preliminary evidence suggests that this newly approved agent will assume a similar role in the treatment of refractory bipolar disorder patients. Initial trials of olanzapine in acute mania yielded positive results, and the drug appears to be generally well tolerated, except for substantial weight gain in some patients.

The use of low dosages of risperidone appears promising, although there are several case reports of depressed patients switching into mania when treated with high dosages of this agent. Low dosages are associated with increased prolactin, and dosages over 8 mg a day are associated with both extrapyramidal effects and weight gain.

Other atypical agents should be watched for their spectrum of efficacy in acute mania and prophylaxis, as they may have beneficial profiles in the treatment of mania such as relative lack of sedation (olanzapine and sertindole); lack of prolactin increases (clozapine, quetiapine, and sertindole); fewer anticholinergic adverse effects (seroquel and sertindole); and less weight gain (ziprasidone). The possible antidepressant aspects of these putative mood stabilizers, perhaps related to their positive effects in the negative symptoms of schizophrenia, also deserve close scrutiny. Table 14.8-9 and Table 14.8-10 summarize preliminary data on the mechanisms of action, adverse effects, and potential efficacies of the serotonin-dopamine antagonist.

A potentially unique approach to the dopamine system might be attained with the tricyclic drug trimipramine (Surmontil). This long-approved antidepressant had an unknown mechanism of action, since it was not a potent reuptake blocker of any of the amine systems. It was recently shown to be a moderately potent antagonist of type 2 dopamine (D_2) and D_4 receptors, somewhat similar to clozapine. It has also been reported in open studies to be effective in monotherapy for acute schizophrenic episodes and psychotic depression. Therefore, in light of its D_2 and D_4 blocking properties as well as its efficacy as an antidepressant, some have ecommended using trimipramine in bipolar patients with refractory rapid cycling depressions.

L-Type Calcium Channel Inhibitors A series of preliminary reports suggest that the calcium channel inhibitor verapamil (Calan, Isoptin) has acute antimanic efficacy.

Whereas a number of controlled studies suggest the antimanic utility of verapamil, several recent studies suggest the superiority of lithium. Moreover, one randomized study in acute depression indicated that verapamil is no more effective than placebo and less effective than routine antidepressant treatment. These data led to a search for a more effective calcium channel blocker that might have better spectrum of antidepressant and prophylactic effects than verapamil.

Several groups chose to study the dihydropyridine L-type calcium channel inhibitor nimodipine (Nimotop) because of its (1) ability to penetrate the CNS, (2) relative lack of tolerance development in the treatment of migraine (in contrast to many other calcium channel inhibitors, (3) better profile in many types of animal seizure models than verapamil, and (4) greater ability to block cocaine-induced hyperactivity and associated dopamine overflow. One study reported that 10 of the first 30 evaluable treatment-refractory patients had a clinically relevant response to nimodipine, including patients with rapid and ultradian cycling frequencies (Fig. 14.8-10 and Fig. 14.8-11). Responsivity was confirmed and reconfirmed in some of these patients in a double-blind, off-on-off-on design. Almost all of these patients needed their regimens further supplemented with another agent such as lithium or carbamazepine for a more complete response, however. Carbamazepine augmentation of nimodipine was effective (moderate or marked response on the Clinical Global Impression Scale) in only 5 of 14 patients treated with the combination.

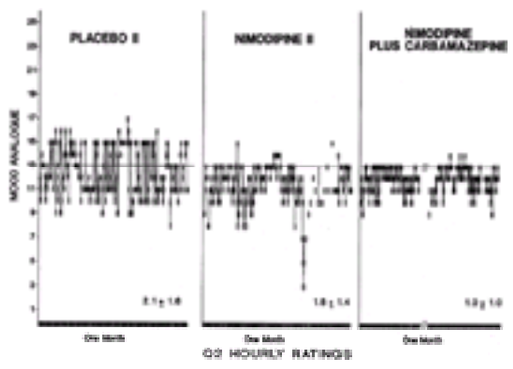

FIGURE 14.8-10 Efficacy of Nimodipine plus carbamazepine combination in a bipolar II female with ultradian cycling.

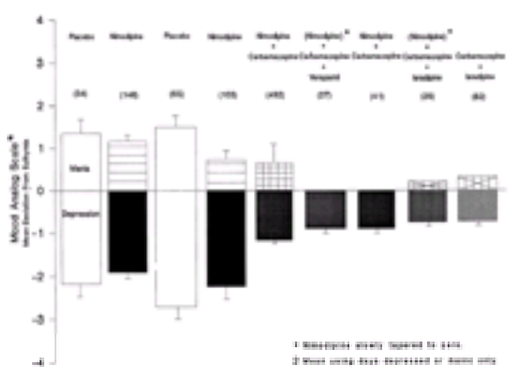

FIGURE 14.8-11 Nurse ratings of the efficacy of dihyropyridine L-type calcium channel inhibitors in a woman with bipolar II disorder.

Of considerable interest were several patients who clearly responded to either monotherapy or combination therapy with nimodipine and transitioned from nimodipine on a double-blind basis to maximally tolerated doses of verapamil, without maintaining response (Fig. 14.8-11). They later either reresponded to nimodipine or to another dihydropyridine L-type calcium channel inhibitor such as isradapine (Dynacirc) further confirming the initial responsivity to this drug class but also suggesting that responsivity might be better conferred by the dihydropyridine subtype of L-type calcium channel inhibitor (with the binding site deep inside the calcium channel) than the phenylalkylamine verapamil (which blocks at the outside of the channel) (Fig. 14.8-10).

Although there is some evidence that patients with extremely rapid cycling fluctuations within a 24-hour period (ultra-ultra rapid cycling) are among those who respond best to the dihydropyridines, the question of whether this subgroup is selectively targeted remains open. Depressed patients with the more classical pattern of global and frontal hypometabolism on positron emission tomography (PET) were among those who responded best to nimodipine. In contrast, those with a hyperactive metabolic pattern were more likely to respond to carbamazepine. Much work remains to define the precise role of the calcium channel blockers in the treatment of bipolar refractory depression (Table 14.8-5 and Table 14.8-6).

Lamotrigine A series of preliminary reports and two controlled trials suggest that lamotrigine, the newly approved anticonvulsant for add-on therapy, has antidepressant and possibly mood-stabilizing properties. In an open study 67 patients were studied usually with the drug as an adjunct to previously ineffective treatment regimens; 27 of 39 (69 percent) who presented in the depressed phase and 19 of 25 (76 percent) in the manic phase showed moderate-to-marked improvement.

A randomized, double-blind study at the NIMH found a significantly greater incidence of good response to blind lamotrigine monotherapy (17 of the first 3 patients [52 percent] than to (9 of 33 gabapentin [27 percent]) or placebo (7 of 32 [22 percent] placebo ($P < .05$). Many patients with refractory depression profiles were among those who showed a good response. Patients with bipolar I disorder depression showed a non-significantly better response than those with bipolar disorder II or a depressive disorder. Another large multicenter study indicated that both 50 and 200 mg a day of lamotrigine were superior to placebo in a 7-week trial in depressed bipolar I disorder patients.

Lamotrigine treatment should be initiated slowly in monotherapy with one 25-mg pill for the first 2 weeks and then 50 mg for 2 weeks, with slow increases thereafter to avoid a moderately high incidence of rash. The rate of increase should be halved if patients are on a regimen including valproate, which can markedly increase lamotrigine blood concentrations and the propensity for rash and more serious dermatologic complications. Conversely, carbamazepine decreases lamotrigine concentrations by approximately 50 percent, and one can start with higher dosages.

The precise anticonvulsant or psychotropic mechanisms of action of lamotrigine remain to be delineated. However, lamotrigine, like valproate, is a broad-spectrum anticonvulsant, effective not only in complex partial and generalized seizures, but also in absence and atonic seizures, in contrast to carbamazepine, which can exacerbate absence seizures. This is important because recent studies have suggested that carbamazepine, lamotrigine, and phenytoin have highly similar properties in the blockade of type 2 sodium channels and consequent inhibition of release of excitatory amino acids such as aspartate and glutamate (Fig. 14.8-9 and Fig. 14.8-12). However, the differential clinical profiles of these drugs in epilepsy and the preliminary evidence that lamotrigine may be effective in some patients who respond inadequately to carbamazepine suggest that lamotrigine has additional mechanisms not shared by carbamazepine. Recent evidence suggests that lamotrigine affects different types of calcium channels (N-type and P-type) and blocks serotonin reuptake and is active at serotonin (5-hydroxytryptamine [5-HT]) type 3 ($5\text{-}HT_3$) receptors, but the high concentrations at which these serotonergic effects are observed suggest that they are not clinically relevant. Preliminary data suggest that depressed patients with the more classical topographic pattern of frontal hypoperfusion on PET studies are likely to respond to lamotrigine.

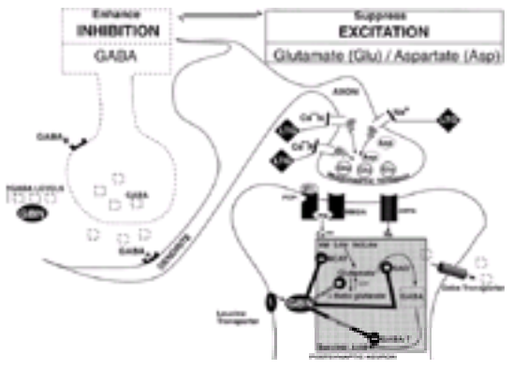

FIGURE 14.8-12 Targets of action of lamotrigine (LTG) and gabapentin (GBN). AMPA, a-amino-3-hydroxy-5-methyl-4-isoxalone propionate; GAD, glutamic acid decarboxylase; GDH, glutamate dehydrogenase; PCP, phencyclidine.

Gabapentin Open adjunctive studies indicate that gabapentin, a newly approved anticonvulsant for adjunctive therapy may also have some mood-stabilizing effects in bipolar patients. The drug appears to have positive effects on sleep and anxiety. However, two double-blind studies of monotherapy, one in acute mania and the other in refractory affectively ill patients showed no benefit over placebo. Whether gabapentin's prominent effects on the L–amino acid transport mechanism (Fig. 14.8-9) and resulting increases in brain GABA concentrations (Fig. 14.8-11) are related to its anticonvulsant or putative psychotropic properties remains to be determined.

Topiramate A recently approved add-on agent for treatment of refractory epilepsy, topiramate, is just beginning to be studied in bipolar disorders (Table 14.8-7). Preliminary uncontrolled data suggest it may have mood-stabilizing properties in rapid-cycling patients, with better antimanic than antidepressant effects. A major asset of topiramate is its positive effect of weight loss in contrast to lithium, valproate, gabapentin, many antidepressants, and most antipsychotic drugs. As a carbonic acid inhibitor it has a 1 percent risk of renal calculi (virtually all occurred in males); the calculi are made of calcium and respond well to emergency lithotripsy. Cognitive slowing and difficulty with word finding, may appear with rapid dosage escalation or in combination therapy. Topiramate is a selective inhibitor of glutamate a-amino-3-hydroxy-5-methyl-4-isoxalone-propionate (AMPA) receptors (Fig. 14.8-9) that also has GABAergic actions and blocks sodium ion channels.

Other Anticonvulsants The clinical utility of other GABA-active anticonvulsants, such as the GABA reuptake inhibitor tiagabine (the transamine inhibitor g-vinyl GABA (vigabatrin [Sabril]) and the agonist progabide (all remain to be further explored, as does the older anticonvulsant phenytoin (Dilantin). Acetazolamide (Diamox) has been reported to be effective for atypical psychoses associated with dreamy, confusional states as well as those occurring premenstrually or in the puerperium. Whether this profile is shared by other carbonic acid inhibitors such as topiramate remains to be determined.

Electroconvulsive Therapy Older clinical observations and recent controlled clinical trials continue to document the efficacy of ECT in acute mania. Bilateral treatments appear to be required; unilateral, nondominant treatments have been reported to be ineffective or to exacerbate manic symptoms in some studies. In light of the many effective pharmacological treatments noted above, and the utility of using the assessment of their acute antimanic efficacy as a surrogate marker for putative efficacy in long-term prophylaxis, ECT may be reserved for the very rare refractory manic patient or one with medical complications, as well as extreme exhaustion (lethal catatonia) or malignant hyperthermia (Table 14.8-6). Antimanic effects of a brief course of repeated transcranial magnetic stimulation (rTMS) of the brain at 20 Hz over the right but not left frontal cortex or 1 Hz rTMS bifrontally have been observed; whether this well-tolerated nonconvulsive strategy will eventually have a role in clinical therapeutics remains open.

OTHER THEORETICAL AND MECHANISTIC CONSIDERATIONS

Antiadrenergic Drugs A series of other nonanticonvulsant compounds with some neurotransmitter selectivity has been reported to be efficacious in treating mania. Clonidine (Catapres) an a_2-adrenergic agonist, is used to treat hypertension. It acutely inhibits the firing of the noradrenergic locus coeruleus and has been reported to show short-term antimanic efficacy in some (but not all) controlled trials. Response in the first few days of treatment may not be associated with an ultimate long-term response, however. Another agent that inhibits noradrenergic function is the b-adrenergic receptor antagonist propranolol (Inderal). Because very high dosages of either the dextrorotatory or levorotatory of this agent isomer form have been effective, it is questionable whether the b-antagonist properties or the membrane-stabilizing effects of this drug account for its short-term antimanic efficacy.

Cholinomimetics One open study reported that high dosages of choline (3 to 8 grams per day) possess antimanic and anticycling effects in refractory bipolar patients. Intravenous administration of the indirect cholinergic agonist physostigmine (Antilirium) has been demonstrated to have an almost immediate antimanic effect. Physostigmine inhibits acetylcholine esterase function, making more acetylcholine available at the synapse. Although this strategy can rapidly decrease manic symptomatology, it also has a rapid half-life and can be associated with rather marked increases in dysphoria and other adverse effects, such that its long-term utility is doubtful. The success of attempts to increase cholinergic function in the long term through other methods (e.g., lecithin, deanol, or direct acetylcholine agonists) has not been adequately delineated.

Overview of Antimanic Agents The ability to achieve rapid antimanic effects with intravenous physostigmine (an acetylcholinesterase inhibitor) suggests that given appropriate pharmacological intervention and pharmacokinetics, there is no theoretical reason why an acute antimanic response cannot be achieved extremely rapidly, even though most antimanic treatments have a moderate delay in onset. Manipulations of a variety of neurotransmitter systems (inhibition of nonadrenergic and dopaminergic and potentiation of the cholinergic, benzodiazepine, GABA, and perhaps serotonergic systems) all can induce antimanic effects. The antipsychotic agents block dopamine receptors; clonidine and propranolol appear to decrease a- and b-noradrenergic function, respectively. Reserpine (Diupres) which depletes catecholamines and indoleamines, has also been reported to have antipsychotic and antimanic effects.

As noted in Table 14.8-2 and Table 14.8-3, awareness of the multiple neurotransmitter approaches to the treatment of mania may be clinically useful in both changing and combining treatments that target different systems in nonresponsive patients. Alterations in endogenous neuropeptide function also have been postulated in mania. Although manipulations of opioids or cholecystokinin have not yielded consistent results in psychotic schizophrenia patients, isolated reports that thyrotropin-releasing hormone (TRH) or calcitonin were successful in treating excited psychotic states, including mania, deserve replication. The potential efficacy of peptide interventions in mania is mentioned because peptides could represent the next generation of antimanic treatments, particularly in light of the evidence of peptide neurotransmitters coexisting in neurons with more classical neurotransmitter substances that have been indirectly linked to the manic syndromes and the early reports of the antidepressant effects of a substance P antagonist.

Preliminary evidence indicates that the L-type calcium channel inhibitors that inhibit calcium influx through voltage-dependent channels may be effective in treating acute mania. New data indicate that lithium, carbamazepine, lamotrigine, and valproate all block calcium cation (Ca_{2+}) influx through the glutamate NMDA receptor (Fig. 14.8-9), which raises the theoretical rationale of using different mechanisms of blockade of Ca_{2+} influx for more effective or complete effects in refractory patients. Multiple studies of platelets and lymphocytes of bipolar disorder patients indicate increased baseline levels of, or serotonin- or thrombin-stimulated rise in, intracellular calcium.

A new approach examining common effects of chronic lithium and valproate has revealed inhibition of protein kinase C as a putative target for antimanic effects. This idea of H. Manji received preliminary support from the observations of rapid-onset acute antimanic effects in six of nine patients treated with the protein kinase C inhibitor tamoxifen (Nolvadex).

Many of the principles of treating unipolar depression are applicable to the treatment of depression in bipolar disorders, but the critical role of concomitant treatment with mood stabilizers and targeting the symptoms most characteristic of bipolar disorder depression—such as its atypicality, and reverse vegetative symptoms of hypersomnia, increased appetite, weight, lethargy, and psychomotor retardation—require emphasis. Antidepressant drugs that are particularly useful in bipolar disorder patients are listed in Table 14.8-2.

MAINTENANCE TREATMENT OF BIPOLAR DISORDERS

Lithium Prophylaxis Lithium originally appeared to be effective in some 70 to 80 percent of bipolar disorder patients, but current estimates suggest that even with adjunctive use of antidepressants and antipsychotics, a response rate of 40 percent or less in many lithium clinics is more accurate (Fig. 14.8-13).

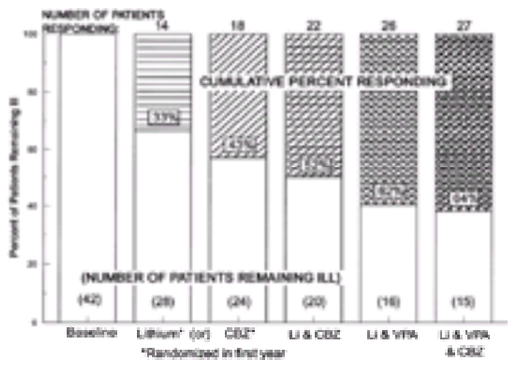

FIGURE 14.8-13 Cumulative response and failure rates on mood-stabilizing regimens. CBZ, carbamazepine; VPA, valproate.

Although initial studies indicated the need for blood concentrations between 0.8 and 1.2 mEq per liter, some series have suggested that concentrations of 0.5 to 0.8 mEq per liter might be effective in maintenance treatment. However, a recent controlled study indicated that the lower levels of adverse effects are achieved at a cost; the relapse rate with a low lithium concentration range (0.4 to 0.6 mg/L) is three times that at higher concentrations (0.8 to 1.0 mg/L). Monitoring of trough levels (performed in the early morning with the AM dose withheld) at 1- to 2-month intervals (or more frequently if the patient's course is unstable) is recommended. A 1-year controlled study reported that the greater the lithium-induced decreases in plasma free thyroxine (T_4) concentration, (but not lithium blood levels) the greater the severity of depression and cycling.

Because of the substantial data on long-term efficacy and prevention of suicide with lithium treatment, preventive treatment should be considered after one or two manic episodes and after a single severe episode of mania, particularly if there is a positive family history of bipolar disorder. The development of a life chart, so that the frequency, severity, and interval between episodes can be accurately assessed, may also assist in arriving at the decision for prophylaxis. If previous episodes were severe—socially incapacitating and requiring hospitalization—or associated with extremely adverse events for the patient and family, one would be more likely to consider prophylaxis earlier rather than later, despite moderately long well intervals between episodes. These factors should be discussed with the patient during a euthymic interval so that the appropriate risk-benefit ratios can be weighed carefully with adequate informed consent. Data from numerous studies indicate that greater numbers of prior episodes (more than three or four) are associated with a poor response to lithium prophylaxis, so delayed prophylaxis may have negative consequences not only for the increased morbidity during these recurrences, but also for ultimate treatment response. Whether greater numbers of prior episodes also predispose to the development of tolerance (Fig. 14.8-14) or lithium discontinuation-induced refractoriness (Fig. 14.8-15) remains to be studied.

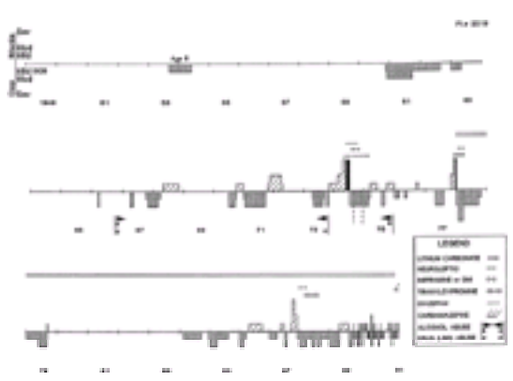

FIGURE 14.8-14 Development of refractoriness to lithium in the course of malignant progression of mood disorder. DMI, desipramine; LSD, lysergic acid diethylamide; DOB, date of birth.

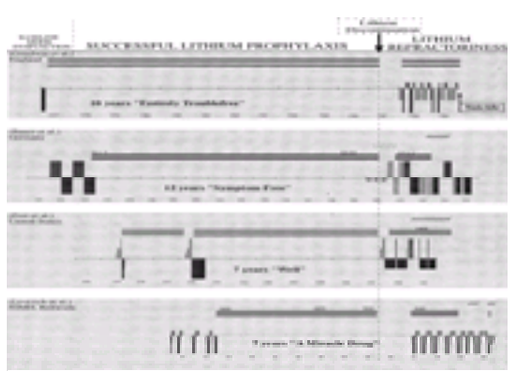

FIGURE 14.8-15 Lithium discontinuation–induced refractory illness.

Carbamazepine and Other Mood Stabilizers One alternative to traditional unimodal antidepressants for depressive breakthroughs is the addition of first-generation anticonvulsant mood stabilizers such as carbamazepine or valproate, or putative second-generation anticonvulsants such as lamotrigine (Table 14.8-11). Although the controlled evidence concerning the efficacy of carbamazepine used as monotherapy for primary depression is inadequate those findings taken with the more substantial emerging literature on the efficacy of carbamazepine prophylaxis for both manic and depressive episodes, raise the priority of using this agent to supplement lithium in depressive breakthroughs, particularly those of the rapid-cycling variety. Although only one-third of refractory depressed patients responded in one study, the responders tended to be the patients with greater initial severity of depression and clearer prior histories of discrete episodes rather than chronic depression.

Table 14.8-11 Drugs in the Prevention of Manic Depressive Episodes and Cycling

When antidepressant response to carbamazepine was observed, it tended to occur with the typical lag observed with other agents, so that only minor improvement was noted in the first and second weeks of treatment, but considerable improvement was observed after the third and fourth weeks. Surprisingly, the degree of antidepressant response was correlated with the degree of decrease in T_4 and free T_4. An abnormal EEG or increased psychosensory symptoms did not predict an acute response to carbamazepine in one series, but did in another when carbamazepine was used for augmentation, and a 53 percent response was observed compared with an even higher rate for lithium augmentation.

More than a dozen controlled studies support the comparative efficacy of carbamazepine and lithium in the prophylaxis of both manic and depressive episodes. One study reported lower antimanic effects of carbamazepine than with lithium as did another study in patients with classic euphoric mania. However, the latter study indicated a better response to carbamazepine in those with atypical presentations (i.e., dysphoric mania, schizoaffective disorder, rapid cycling and comorbidities).

In a small series of patients who responded inadequately to carbamazepine alone, one half showed a rapid onset of antidepressant effect with lithium augmentation. Thus, the combination of carbamazepine and lithium appears to help a substantial subgroup of otherwise refractory patients (Fig. 14.8-5). In one randomized study in bipolar outpatients with a high incidence of rapid cycling, response rates were under 25 percent with either lithium or carbamazepine monotherapy for 1 year, even when adjunctive antidepressants and antipsychotic drugs were allowed, but over 50 percent with both drugs in combination (Fig. 14.8-16). Thus, one might consider combination mood stabilizer treatment from the outset in this rapid-cycling subgroup.

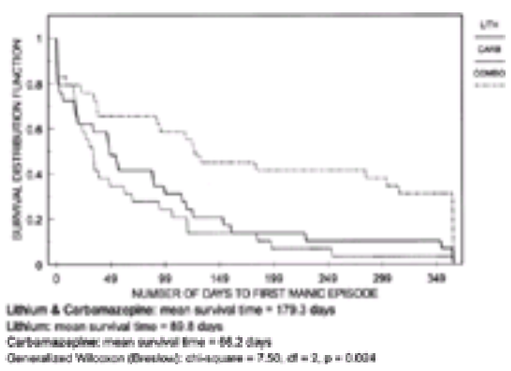

FIGURE 14.8-16 Mean survival time to first manic episode is longer on the combination of lithium and carbamazepine than with either lithium and carbamazepine alone (*N* = 29).

Carbamazepine has been reported effective in some patients failing multiple traditional antidepressant trials, especially in those with a history of head trauma or EEG abnormalities, in one series. As noted for refractory bipolar disorder patients discussed below, the anticonvulsants carbamazepine and valproate may be used in combination with unimodal antidepressants, and the role of these and other combination treatments for the refractory depressed patient deserves much further systematic research to provide adequate statistical and sequence-ordering guidance for the clinician.

Valproic Acid In many open series, valproic acid alone or in addition to lithium has been reported to be successful in the long-term treatment of a substantial subgroup of previously lithium- or carbamazepine-refractory patients. The acute antidepressant efficacy of valproic acid is much less well delineated than its antimanic efficacy, and the utility of this treatment for an acute depressive episode remains to be further elucidated. Nonetheless, the combination of lithium and valproate offers another excellent option in the long-term management of bipolar patients who do not respond to lithium alone. A response to one anticonvulsant may not predict a response to another, and positive long-term effects of valproic acid plus lithium have been noted in patients who did not respond to lithium or carbamazepine prophylaxis (Fig. 14.8-17 and Fig. 14.8-18).

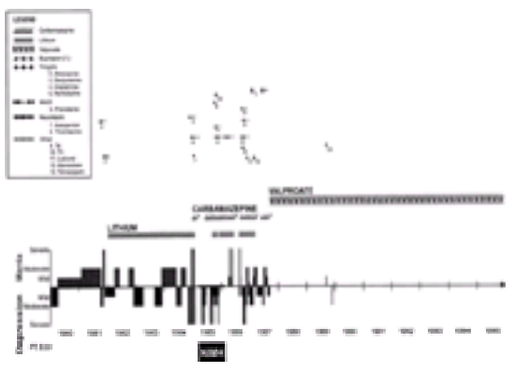

FIGURE 14.8-17 Prophylactic response to valproate in a nonresponder to lithium and carbamazepine.

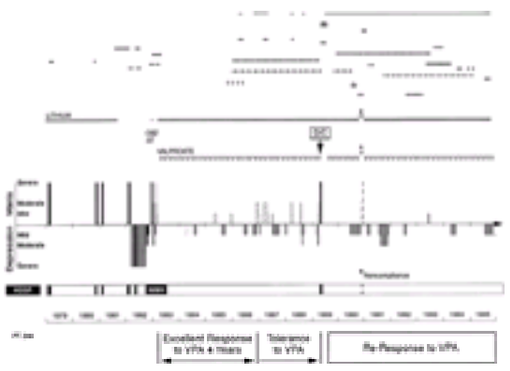

FIGURE 14.8-18 Tolerance and reresponse to the prophylactic effects of valproate (VPA). CBZ, carbamazepine; D/C, discontinuation.

Lamotrigine A series of open studies suggests the possibility of mood stabilizer effects of lamotrigine, in many instances in those not responsive to conventional treatments. One double-blind, controlled study of short-term prophylaxis in ultrarapid-cycling patients revealed significant antidepressant and antimanic effects over those of either gabapentin or placebo. These data, taken with the large controlled trial of lamotrigine in bipolar I disorder depression suggest that the addition of lamotrigine to lithium or another mood stabilizer is an alternative option to the addition of a unimodal antidepressant. Response rates of 60 to 70 percent were observed in open studies when lamotrigine was used adjunctively.

Calcium Channel Inhibitors The few uncontrolled studies of calcium channel inhibitors suggest their promise as prophylactic treatments, but they may be less effective for depressive breakthroughs than for manias. One study reported better prophylactic efficacy of 1 year of the combination of lithium and nimodipine compared with 1 year of either drug alone. Further controlled studies of prophylaxis are sorely needed. Although data are limited to just one double-blind series at the NIMH, in the several patients crossed over blindly from one L-type calcium channel inhibitors to another, agents in the dihydropyridine class (nimodipine and isradapine) had better antidepressant and mood-stabilizing effects than the phenylalkylamine verapamil (Fig. 14.8-10 and Fig. 14.8-11).

Gabapentin Gabapentin may show positive effects on mood, anxiety, and sleep in 30 to 40 percent of refractory bipolar disorder patients, but much higher rates of response are reported in open studies when gabapentin is used as add-on treatment. The mood-stabilizing anticonvulsants (carbamazepine, valproate, lamotrigine, and gabapentin) certainly also deserve more consideration for a patient's profound sleep disturbance, with or without associated posttraumatic stress disorder, especially if the patient has comorbid alcoholism or bipolar disorder depression and benzodiazepines are to be avoided.

Thyroid Although thyroid potentiation similar to that observed in unipolar depression can be attempted, treatment with dosages above suppressive doses should be approached with caution. Several investigators found associated medical toxicities with high-dosage thyroid treatment and inadequate maintenance of long-term prophylaxis unless other routine agents were used concurrently. Thus liothyronine (Cytomel), the levoratatory isomer of triiodothyronine (T_3), is recommended for short-term augmentation strategies because of its short half-life, and levothyroxine (Levoxyl, Synthroid) the levorotatory isomer of T_4 is recommended by some as more appropriate for long-term maintenance during prophylaxis. As noted above, addition of liothyronine to levothyroxine has been reported to benefit nonresponders.

Hypermetabolic Dosages of Levothyroxine Recent data indicate that high-dosage levothyroxine treatment— μg a day targeted toward achieving a free thyroxin index 150 percent of normal—may be helpful as an adjunctive treatment in rapid-cycling patients. Another report further suggests that this high-dosage augmentation strategy may benefit patients with persistent refractory depression. The data indicating improvement in both manic and depressive phases with hypermetabolic levothyroxine augmenting strategies speaks to the potential importance of this modality for bipolar disorders. However, systematic long-term trials must be conducted, and the issue of whether some patients lose responsivity to such thyroid augmentation strategies with the development of tolerance requires further exploration.

Serotonin-Dopamine Antagonists Given their short-term antimanic and longer-term antipsychotic effects against the positive and negative symptoms of schizophrenia, the serotonin-dopamine antagonists are becoming increasingly important in prophylaxis of mood and schizoaffective disorders. Their better profile of acceptability (Table 14.8-8) and safety than the conventional antipsychotics impels their use instead of these conventional antipsychotics even before adequate data on long-term efficacy become available.

POTENTIAL CORRELATES OF RESPONSE TO THE MOOD STABILIZERS

From a clinical and theoretical perspective it would be valuable to know whether the efficacy of carbamazepine, valproate, or the other putative mood-stabilizing anticonvulsants is related to their ability to stabilize neural excitability in temporal lobe and limbic structures (independent of whether or not a seizure disorder underlies the mood disorder) (Fig. 14.8-8 and Fig. 14.8-19). Only a minimum of evidence is available indicating that carbamazepine, for example, acts via its effects on temporal lobe or limbic substrates. However, the approximate rank ordering of carbamazepine, valproate, and phenytoin according to their degree of limbic selectivity is roughly related to their psychotropic efficacy and provides at least indirect support for the limbic hypothesis.

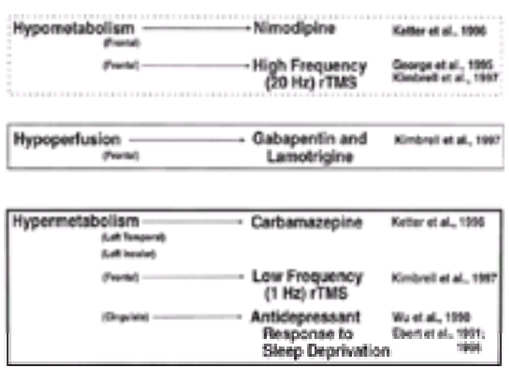

FIGURE 14.8-19 Preliminary predictors of clinical response—functional brain imaging. rTMS, repeated transcraneal magnetic stimulation.

An indirect marker of limbic dysfunction—degree of psychosensory symptoms—is not related to carbamazepine response in primary affective disorder, although others disagree. Recent PET data, however, suggest that patients with initial baseline hypermetabolism, particularly in the left insula, are among those who respond best to carbamazepine, whereas those with a pattern of baseline frontal hypometabolism respond best to nimodipine. These preliminary data are among the first to provide suggestive evidence that increased metabolism in mesiotemporal structures could be associated with the therapeutic response to carbamazepine. Patients with relative hypoperfusion at baseline show normalization upon successful treatment with lamotrigine or gabapentin, whereas nonresponders tend to be closer to normal at baseline and decrease significantly with treatment with these agents. Some preliminary data also link baseline hypometabolism with response to high-frequency rTMS, which normalizes this pattern. Those with relative hypermetabolism at baseline respond to low-frequency (1 Hz) rTMS in association with normalization of this pattern. Although not clinically useful at present, one can only hope that replication and extension of these studies will assist in better matching individual treatments to individual patients.

A number of studies suggest an equally good response to valproate in dysphoric compared to euphoric mania, in contrast to the relatively poor response to lithium in the dysphoric subtype. One small study suggested that carbamazepine responders tended to be more dysphoric at baseline than nonresponders.

P>More than 15 studies have reported a relatively poor response to lithium in rapid-cycling patients; only 2 studies have reported a good response, and 4 have reported no differential response. The relation of carbamazepine response to rapid cycling is less clear. In one acute mania study and several prophylaxis studies, a high response percentage was observed in rapid-cycling patients. However, when rapid-cycling patients were compared with non-rapid-cycling patients in two studies, both found a higher prophylactic response rate to carbamazepine in non-rapid-cycling patients. Data from Japan are revealing in that the 53 percent response rate to carbamazepine in patients with a history of rapid cycling (compared with 76 percent in those without) is still substantially higher than the 30 percent response rate observed for lithium in patients with a history of rapid cycling (compared with 64 percent in those without such a history).

Taken together, these data suggest that rapid cycling is a poor prognostic indicator for both carbamazepine and lithium treatment, although some patients with rapid cycling respond positively to carbamazepine. NIMH data suggest the potential utility of treatment of rapid-cycling patients on the combination of lithium and carbamazepine from the outset (with its 53 percent response rate) in light of the poor response to either monotherapy in these patients (28 percent response to lithium and 19 percent response to carbamazepine).

Rapid cycling and prediction of response to valproate is less well studied, but initial indications from a large double-blind collaborative study are that the short-term antimanic response is robust in these individuals, which agrees with other data from open studies indicating an excellent acute and prophylactic response in several large cohorts of rapid-cycling patients treated with either monotherapy or combination treatment. However, one study reported that an accelerating course of illness was a poor prognosis factor for predicting valproate response. Initiating treatment with valproate (Fig. 14.8-17) or lithium and valproate in rapid-cycling or dysphoric manic patients from the outset may thus be a particularly useful alternative to lithium and carbamazepine for this subgroup.

Some evidence suggests that a negative family history of mood disorders may be associated with a good response to carbamazepine; seven of eight studies reported that a positive family history of affective illness in first-degree relatives is associated with a positive response to lithium. These data, in conjunction with clinical case reports illustrating that patients with evidence of delirium, dementia, and other cognitive disorders show a relatively poor response to lithium, a high potential for toxicity, and a potentially good response to the anticonvulsants carbamazepine or valproate, suggest that the familial genetic subtype of mood disorders may be more responsive to lithium than the subtype mediated through other nonhereditary pathophysiological mechanisms, which may be better targeted with the mood-stabilizing anticonvulsants. These possible mechanisms include neuronal and environmental insult related to birth trauma, infection, secondary mood disorder, and substance

abuse. Further study of this issue is clearly required.

The mood-stabilizing anticonvulsants (carbamazepine, valproate, and possibly lamotrigine and gabapentin) certainly also deserve higher consideration for a patient's profound sleep disturbance (with or without associated posttraumatic stress disorder), especially if the patient has a comorbid alcohol use disorder or a bipolar disorder, and benzodiazepines are to be avoided.

RELATIVE ADVERSE-EFFECT PROFILES OF LITHIUM AND THE MOOD-STABILIZER ANTICONVULSANTS

Since only a modicum of data suggests clinical or biological predictors of response to the mood stabilizers (Fig. 14.8-8 and Fig. 14.8-19), adverse effect profiles and tolerability in long-term prophylaxis (Table 14.8-7) as well as mechanisms of action (Table 14.8-4, Fig. 14.8-6 and Fig. 14.8-9) become potential selection factors.

The general profile of lithium-induced side effects has proven to be relatively benign in the long-term maintenance treatment of bipolar patients. However, several of lithium's more prominent adverse effects deserve comment, as do the relative comparisons with and among the mood-stabilizing anticonvulsants.

Thyroid Function Lithium clearly can impair thyroid function by several different mechanisms; it has even been used to treat hyperthyroidism. Lithium uniformly lowers T_3 and T_4 concentrations in the plasma and, in some patients, increases thyroid-stimulating hormone (TSH). TSH concentrations above normal can be taken to indicate that the hypothalamic-pituitary-axis is working overtime to maintain normal levels of thyroid hormones. Lower free T_4 concentrations during lithium prophylaxis in one study were associated with more-severe depression and more-rapid mood fluctuation. Thus, one might consider thyroid replacement with levothyroxine when levels of TSH are elevated, even when thyroid hormone indexes are still within their normal lower limits. Occasional checks of thyroid function at 6-month to 1-year intervals are wise, as is an earlier check if there is a breakthrough of depressive symptomatology during otherwise adequate lithium maintenance treatment. In these instances, treatment of underlying hypothyroidism can help alleviate a depression that is linked to this hormonal deficit. Whereas levothyroxine is generally used for suppression of TSH and replacement, anecdotal evidence suggests that addition of liothyronine to the levothyroxine replacement may help some patients with refractory depression or cycling.

Carbamazepine tends to decrease T_4, free T_4, and T_3 concentrations (as does lithium), and in combination, the decreases are additive. During carbamazepine treatment there is a negligible incidence of clinical hypothyroidism or above-normal increases in TSH. Consequently, thyroid supplementation of carbamazepine is rarely needed. When the two drugs are used in combination, however, the lithium effect on TSH may override that of carbamazepine, and the patient may then require thyroid supplementation.

Renal Function By the late 1980s the fear regarding the possible high incidence of long-term adverse consequences of lithium on the kidneys had largely dissipated. Original reports of severe nephrotoxicity and pathology with elevated creatinine and low clearance originally attributed to lithium were, in part, related to the absence of an age-matched control group of psychiatric patients not treated with lithium. Thus, although lithium rather consistently impairs vasopressin function at the level of adenylate cyclase and often produces a syndrome of diabetes insipidus, it is less consistently associated with other evidence of renal toxicity, although isolated case reports persist. Preliminary data suggest that less renal toxicity may occur in patients using single nighttime dosing (producing higher peaks, but lower nadirs) than occurs with conventional dosing regimens. Single nighttime dosing may also facilitate compliance.

Current practice suggests that frequent monitoring of renal function during lithium treatment is not generally indicated; however, baseline measures of renal function including creatinine clearance should be obtained before beginning lithium treatment in patients with a history of some renal alterations. Patients must have adequate fluid intake to maintain an appropriate fluid and electrolyte balance because of the induction of diabetes insipidus syndrome related to the blockade of antidiuretic hormone actions. Several patients have been reported in whom high levels of lithium during intoxication were associated with irreversible cerebellar toxicity. Thus lithium levels, fluid and electrolyte status, or both, should be monitored closely during periods of febrile illness, decreased fluid intake, or greater-than-ordinary fluid loss (e.g., during extreme athletic stress or during gastrointestinal illnesses with vomiting or diarrhea).

Amiloride (Midamor) (5 to 10 mg) has been useful in the treatment of lithium-induced diabetes insipidus. If diuretics (furosemide [Lasix] or thiazides) are used, lower dosages of lithium are indicated because these agents will increase lithium concentrations.

Because carbamazepine appears to act as a vasopressin agonist either directly or by potentiating vasopressin effects at the receptor, it will not suffice to reverse lithium-induced diabetes insipidus, which occurs by an action of lithium below the receptor level at the adenylate cyclase second-messenger system. Demeclocycline (Declomycin) and doxycycline (Vibramycin) may counter the hyponatremic effects of carbamazepine, as may lithium. The hyponatremic effects of the ketoderivative oxcarbazepine may be more prominent than those of carbamazepine. To the extent that the minor cognitive impairments of lithium are in part related to its ability to impair vasopressin function in the brain, these data suggest that not only would carbamazepine be less likely to cause this adverse effect, but also during combination treatment, the adverse effects of lithium might override those of carbamazepine.

Carbamazepine tends to induce a benign hypocalcemia that is generally not associated with bone demineralization. In contrast, lithium often produces a transient increase in serum calcium concentration.

Tremor Tremor can be problematic for a small but substantial percentage of patients treated with lithium. The tremor is frequently exacerbated by social stress. When the tremor persists at doses near the lower end of the therapeutic range or at the minimum doses necessary for therapeutic efficacy, attempts can be made to treat it symptomatically. Some investigators find that 10 to 40 mg of the b-blocker propranolol in divided daily doses may reduce lithium tremor. Relief may occur within 30 minutes and may last from 4 to 6 hours. Valproate also has dose-related tremorogenic effects. Gabapentin, in contrast, has been used to treat essential tremor. The dihydropyridine L-type calcium channel inhibitors may provide a nontremorogenic adjunct or alternative to lithium.

Gastrointestinal Effects Gastrointestinal adverse effects (diarrhea and indigestion) can also be problematic for many patients on lithium and valproate but may be attenuated by reducing the dose or giving it at mealtimes (for indigestion). Antidiarrheal agents should be restricted to short-term treatment. The calcium channel blockers (which may be constipating) may substitute for lithium or partially counter its adverse effects when used in combination. Histamine type 2 receptor (H_2) inhibitors may help counter valproate's gastrointestinal adverse effects.

Cognitive Effects Patients may express concern about the effects of lithium on their memory, spontaneity, or creativity. Although some impairment can be objectively delineated in some, but not all, types of detailed neuropsychological testing, most patients either do not experience this effect or do not find it unduly impairing. In fact, productivity and creativity may, overall, be enhanced during lithium treatment, because it prevents unproductive manic and depressive phases. Although no adequate approach to the subjective cognitive effects of lithium has been demonstrated (other than reducing the dose), associated causes for cognitive impairment must be ruled out, including possible hypothyroidism or an inadequately treated coexistent depression. Donepezil (Aricept) has been reported helpful in isolated case reports and deserves further exploration and study.

Many so-called drug-related adverse effects are also evident during placebo treatment phases and thus appear to be more closely associated with illness-related variables than with a particular psychopharmacological treatment. This perspective on lithium maintenance treatment clearly needs to be explored with the patient to avoid premature discontinuation of treatment or noncompliance. Carbamazepine and valproate are noted for their benign cognitive side effects in the epilepsies and may be better tolerated than lithium in some instances, although they, too, can be associated with subjective complaints and word-finding difficulties. Lamotrigine does not appear to have lithium's occasional liability of stabilizing mood slightly below baseline, and some patients may be stabilized at a mood or energy level slightly over baseline (i.e., above 50 millimeters on the mood analogue scale). Topiramate clearly causes cognitive motor slowing, speech, or word-finding difficulties in a small percentage of patients, particularly with high initial doses, rapid upward titration or in combination therapy. Valproate has been associated with a reversible organic brain syndrome with EEG slowing and a dementia-like presentation in isolated patients with epilepsy.

Weight Gain Lithium-induced weight gain can be a vexing problem for a moderate percentage of patients, and in one study was a correlate of better mood-stabilizing response. Thyroid indexes should be rechecked, and the patient reminded not to use calorie-containing beverages when maintaining the necessary increased fluid intake associated with diabetes insipidus.

Weight gain can also be problematic with valproate. Whether this is a correlate or causal link in the reported occurrence of polycystic ovary syndrome in epileptic patients taking valproate remains for further study.

Like most dopamine receptor antagonists (with the possible exception of molindone [Moban]), the serotonin-dopamine antagonists are also associated with weight gain; clozapine, olanzapine, and risperidone are particularly problematic for some patients. One should watch for replication of one anecdotal report of the antidiabetic drug troglitazone (Rezulin) which helped to cause substantial weight loss in a patient who gained considerable weight on olanzapine. Topiramate has a strong tendency to help with weight loss, apparently by both decreasing carbohydrate craving and possibly increasing metabolism as well. Early clinical vignettes suggest it may help overcome lithium or valproate weight increases when used in combination with them. Carbamazepine and gabapentin are less problematic, and, L-type calcium channel inhibitors are relatively weight neutral. Patients lost about 2 pounds in 6 weeks on lamotrigine in one controlled study, compared with a gain of about 2 pounds on gabapentin.

Headache Many mood stabilizers, such as lithium, valproate, and the L-type calcium channel inhibitors are reported to be effective in migraine prophylaxis. Carbamazepine increases substance P concentrations and sensitivity and may treat cluster headaches but can exacerbate migraine. Lamotrigine, with its weak serotonin reuptake effects, may either ameliorate or exacerbate headaches. Lamotrigine and gabapentin have apparent antipain long-term effects in some syndromes.

Rash Lithium may precipitate or exacerbate psoriasis and acne. Lamotrigine treatment must be started extremely slowly to help avoid the otherwise high incidence of rash (10 percent); some estimates suggest that 1 in 500 patients progresses to severe, potentially lethal extremes of Stevens-Johnson or Lyell's syndromes. Risk factors, in addition to rate of titration, include use with valproate (requiring a halving of the lamotrigine dose) and a history of multiple or severe rash on other medications.

Carbamazepine may also produce a common pruritic rash (10 to 15 percent), but severe deterioration to Stevens-Johnson syndrome may be less common than during lamotrigine treatment. Nonetheless, in most instances, carbamazepine treatment should be discontinued with the onset of a rash. However, when patients respond to carbamazepine and other effective treatments are not available, prednisone has reportedly suppressed uncomplicated carbamazepine-induced rashes (i.e., those without evidence of systemic involvement with fever or lymphadenopathy) in a very high percentage of patients. Whether a similar strategy would be effective for lamotrigine remains unstudied.

Hepatitis Valproate has been associated with reports of severe hepatitis in the neurological literature; most of the fatalities have been in children under the age of 2 years, particularly those on polytherapy. Few serious hepatic adverse effects have been reported in adult psychiatric patients on valproic acid, but liver function might be monitored periodically when using this agent, and patients should be warned to report symptoms that might be referable to hepatitis such as fever, right upper quadrant pain, malaise, nausea, anorexia, Coca-Cola–colored urine, and jaundice. Benign elevation of liver function test (LFT) results to two or three times normal are not uncommon in patients taking valproate, carbamazepine, and other anticonvulsants, and LFTs can be followed without drug discontinuation. Zinc and selenium supplements are recommended with valproate, since they are reported anecdotally to decrease the incidence of hepatitis, pancreatitis, and alopecia. Rare cases of carbamazepine-induced hepatitis have been reported, but routine monitoring for this adverse effect does not appear to be indicated. Since lithium and gabapentin are excreted by the kidney, they have no liability in those with evidence of liver pathology or toxicity.

Hematological Effects The side-effect profile of carbamazepine tends to be quite different from that of lithium or valproate (Table 14.8-7). Whenever lithium and carbamazepine act on a common target system, the effects of lithium tend to override those of carbamazepine. In other instances this is a clinical disadvantage, except for the ability of lithium to increase the white count (via increases in colony stimulating factor) which will override the white count-suppressing effects of carbamazepine (via decreasing colony-stimulating factor), and may thus be clinically useful. However, lithium is effective in this regard only against carbamazepine's benign suppression of the white count, and its effects are doubtful if there is evidence of more problematic interference by carbamazepine in hematological function manifest in other cell lines, such as platelets or red cells (RBCs), indicating a possible pancytopenic or aplastic process. The risk of agranulocytosis or aplastic anemia in patients taking carbamazepine has been estimated to be from 1 in 10,000 to 1 in 100,000. If there are normal levels of other blood elements (platelets and RBCs), potentiation with lithium will likely reverse the benign white-count suppression of carbamazepine. Valproate has been associated with thrombocytopenia; the potential impact of lithium on this syndrome has not been reported.

Teratogenic Effects Cardiac (Ebstein's) anomalies of great vessels have been reported to occur with a higher frequency than expected in patients treated with lithium during pregnancy, although recent studies suggest the risk may not be much greater than that in the general population. Thus, the previous prohibition against use of lithium during pregnancy is being reevaluated. In some instances in which the discontinuation of lithium treatment would put the mother at high risk for a severe depression or mania, continuing lithium treatment may be appropriate, especially with increased ability for fetal monitoring.

An increased risk (several percent) of spina bifida has been reported for valproate (which may be dose-related) as well as a slightly smaller risk for carbamazepine, and use of these agents should be avoided in pregnancy if possible. Folate supplements should be used. Persisting biochemical alterations have been found in some animal studies of fetal exposure to typical antipsychotics but have not been assessed systematically in human follow-up studies. Few data are available for lamotrigine and gabapentin, but no specific teratogenic effects have been described. Topiramate causes some bone deformity in animals, but the risk for humans has not been systematically evaluated. ECT may have the lowest risk to the fetus among the somatic treatments, but the effect of maternal seizures has also not been systematically evaluated. The calcium channel blockers have a benign record for fetal abnormalities, and these agents remain among the better candidates for continuation of a putative mood stabilizer during pregnancy. As needed, short-term augmentation with minimal dosages of antipsychotics or high-potency benzodiazepines may be tolerated.

Approaches to Adverse Effects: Dosage Reduction, Adjunctive, Alternative Treatment Dosage reduction may be a first maneuver in treating a variety of lithium-induced problems (tremor, weight gain, thirst, urinary frequency, diarrhea, or psychomotor slowing). If these lower dosages are not adequate for prophylaxis, combination or alternative treatment, especially with carbamazepine or a dihydropyridine calcium channel adverse blocker (which have different side-effect profiles), or other putative mood stabilizers such as valproate may be indicated (Fig. 14.8-5). The renal clearance of lithium appears to decrease with age, so that a lower dose may be necessary and adequate in the older patient on lithium maintenance.

TREATMENT OF BREAKTHROUGH DEPRESSIVE EPISODES DURING LITHIUM AND OTHER MOOD-STABILIZER PROPHYLAXIS

Antidepressant Augmentation Approaches to a depression in an untreated bipolar patient or during an episode emerging during lithium prophylaxis differs from that in a unipolar depressive disorder patient, and few antidepressants have been systematically evaluated in bipolar disorders. The author emphasizes possible choices of agents based on adverse-effect profiles (Table 14.8-3), the differential presenting symptom clusters and comorbid syndromes (Fig. 14.8-20), and mechanisms of action (Table 14.8-2). However, a strong empirical database for these recommendations is lacking. Moreover, given the necessity for protracted clinical trials (of many weeks) to evaluate the clinical efficacy of each individual drug, attempting to potentiate a specific drug treatment once adequate dosages or blood concentration have been reached is recommended before switching treatment modalities. Thus, thyroid or lithium potentiation deserves emphasis in the treatment sequence before multiple trials with single alternative agents. Figure 14.8-21 illustrates a possible treatment algorithm for a depressed patient with a bipolar disorder.

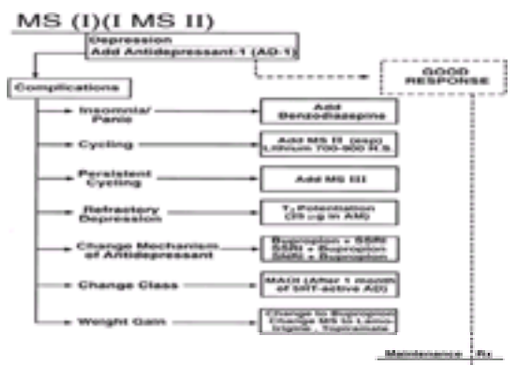

FIGURE 14.8-20 Algorithm for treating bipolar depression and treatment complications. MS I, first mood stabilizer; MS II second mood stabilizer

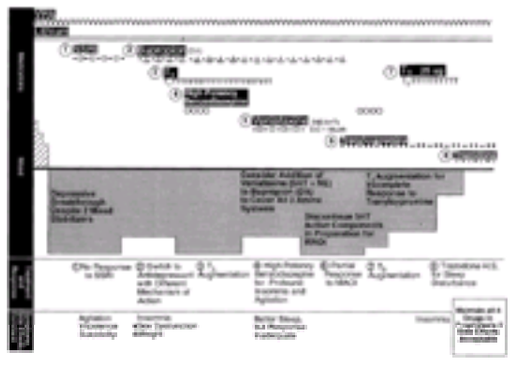

FIGURE 14.8-21 Possible sequential algorithm for the bipolar disorder patient with refractory depression. DA, dopaminergic; NE, noradrenergic.

Potentiation of the Antidepressant Given the long time frame that may be required to evaluate the adequacy of an antidepressant response, antidepressant potentiation should be considered in either the first or second antidepressant trial before switching antidepressants, even to a new category of agents. Thus, when a patient appears to be at either maximally tolerated dosages (or adequate blood concentration of the drug) and has not responded adequately, one might consider adding thyroid hormones or lithium (Table 14.8-5, Fig. 14.8-20).

A sizable but mixed literature exists regarding the efficacy of thyroid potentiation in converting (20 to 50 percent) antidepressant nonresponders to responders. This effect appears to be independent of initial clinical thyroid status or any evidence of hypothyroidism. A response to the addition of liothyronine (25 to 37.5 μg per day in the morning) may occur within days and usually occurs within the first week or two of treatment. Therefore, if there is no response in this time frame, the clinical trial of liothyronine potentiation can be exchanged for other options. Adverse effects are very unusual but could include tachycardia, hypertension, anxiety, or flushing.

A second option is potentiation with lithium. An extensive literature, particularly in unipolar depressions, reveals that addition of lithium carbonate to a variety of antidepressant modalities, including tricyclic drugs, heterocyclic drugs, MAOI, or even carbamazepine, often yields clinical improvement (40 to 60 percent). Response may begin within 24 to 48 hours but may be slower in onset and stretch over 1 to 3 weeks. Dosages of lithium slightly below those conventionally used for monotherapy are generally effective (i.e., 750 to 900 mg in a single dose at bedtime may suffice to reach a target of 0.75 mg/L, which is the concentration reported to be needed for potentiation in unipolar depression). When used in this fashion, the adverse-effect profile of lithium appears quite benign. Lithium potentiation may be effective for all subtypes of depression.

New data suggest that concurrent treatment of acute depression with lithium and antidepressants from the outset also results in more rapid response than with an antidepressant alone. Thus, for the untreated bipolar disorder patient presenting with a depressive episode, the combination of an antidepressant and a mood stabilizer such as lithium is highly recommended. The initial reports of estrogen potentiation of antidepressant response do not appear as promising as those of either thyroid or lithium potentiation but may be considered for postmenopausal women.

Shifting Antidepressant Classes: MAOIs One might consider shifting treatment from one type of antidepressant to another if unacceptable adverse effects appear before adequate blood concentrations or a clinical response has been achieved. If adequate dosages and blood concentrations have been achieved without antidepressant effect, one may switch to a drug with a different biochemical profile within the same class or to a different class, such as an MAOI, but only after a 2- to 4-week period off agents with high potency in blocking serotonin reuptake (Table 14.8-5, Fig. 14.8-20). Problems with orthostatic hypotension may become more prominent in the second and third weeks of MAOI treatment. Salt loading, pressure stockings, and fludrocortisone (Florinef) or the peripherally acting a $_1$ agonist midodrine may prove effective in the treatment of MAOI-induced hypotension. MAOIs can be given in single morning doses or in divided doses. If marked insomnia occurs, nighttime doses of trazadone (Desyrel) have been recommended by some authorities. Bouts of daytime drowsiness and sedation may also become a problem. One might attempt to titrate the dosage against adverse effects, as variations in dosage or timing may be helpful.

The necessity of restricting substances that release tyramine or catecholamines and can produce hypertensive crises during MAOI treatment should be emphasized to the patient. These crises may be clinically manifested as explosive headaches, flushing, palpitations, perspiration, and nausea. Immediate treatment with a slow infusion of phentolamine (Regitine) (5 mg intravenously) in an emergency room is recommended. Most authorities suggest that the patient carry a 10-mg nifedipine (Adalat, Procardia) capsule with them that they could use sublingually or bite and swallow in the event of a presumptive hypertensive crisis.

Although tricyclic and heterocyclic drugs, SSRIs, serotonin-noradrenaline reuptake inhibitors (SNRIs), and MAOIs are central treatments for the unipolar depressed patient, there is reason for caution in their use for the bipolar patient. Some but not all studies have reported an increased incidence of switches into hypomania or mania during tricyclic or MAOI therapy, higher than expected for the patient's natural course of illness, particularly in previous rapid cyclers (Fig. 14.8-22). Whether this increased incidence of switching or cycle acceleration is sufficient to avoid the initial use of antidepressants in favor of mood stabilizers remains controversial. Thus, a shorter depression may occur at the cost of the more rapid onset of the following manic episode, whereas withdrawal of tricyclic drugs and MAOIs has also been shown to slow this cycle acceleration in a small number of patients.

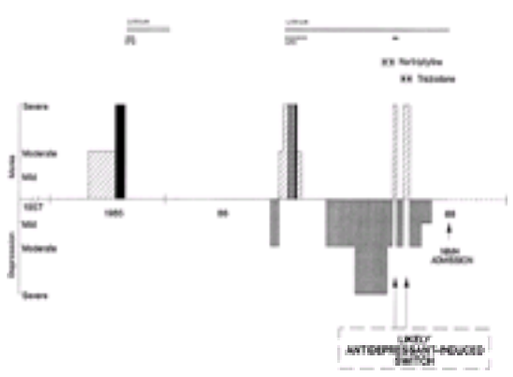

FIGURE 14.8-22 Antidepressant-induced mania in a male patient with bipolar I disorder. CPZ, chlorpromazine.

Uncontrolled observations suggest that tricyclic antidepressants and related compounds may be implicated in the development of continuous cycling phases (i.e., successive episodes without a well interval) (Fig. 14.8-23). This phase of the illness becomes difficult to treat and tends to be relatively lithium refractory. Anecdotal evidence and one double-blind, randomized study indicate that bupropion may not be associated with the same switching tendency as the tricyclic antidepressant desipramine (Norpramine). SSRIs may be less involved in the switch phenomenon and in cycle induction than the tricyclic drugs, but this too requires further investigation, since the commencement of rapid or continuous cycling coincident with the use of SSRIs has been observed anecdotally.

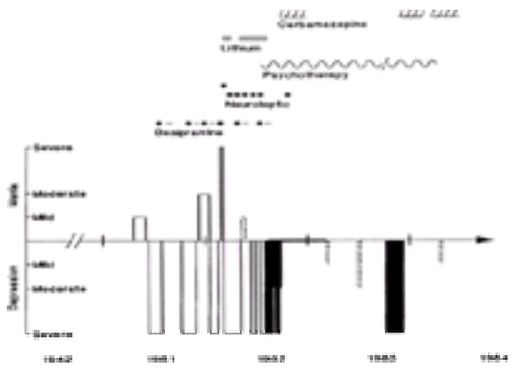

FIGURE 14.8-23 Life chart of a patient rated likely to have had antidepressant-induced cycle acceleration. *Solid black shading* indicates patient was hospitalized.

Once a switch has been observed with an MAOI, reexposure to even a different MAOI may also lead to an earlier onset of a switch, as observed in one controlled study, perhaps reflecting a sensitization phenomenon. Naturalistic data, however, raise questions about whether antidepressant-induced switches occur on each exposure to these drugs. Moreover, it is unclear whether a drug-induced switch appears only in those predestined to have spontaneous switches or whether this occurrence actually predisposes the patient to develop further spontaneous manic episodes. Women and those with rapid cycling may be at higher risk for tricyclic-induced switching or cycle acceleration.

Adding a Second Mood Stabilizer The unimodal antidepressants should be used cautiously and in conjunction with mood stabilizers for bipolar disorder depressive episodes, and particularly if there is a prior history of drug-induced switches, other options should be considered. There is much to recommend adding a second or even a third mood stabilizer (lithium, lamotrigine, carbamazepine, valproate, or a dihydropyridine L-type calcium channel inhibitor) in the rapidly (or ultrarapidly or ultradian) cycling depressed bipolar disorder patient prior to the use of a unimodal antidepressant (Table 14.8-5, Fig. 14.8-24).

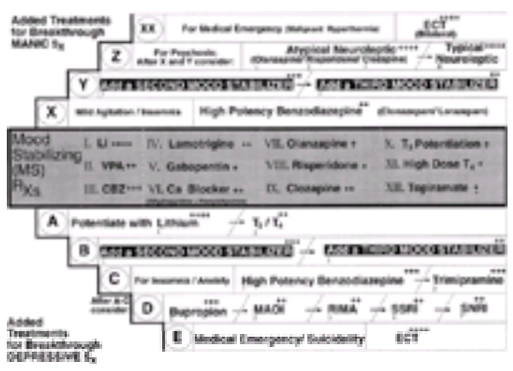

FIGURE 14.8-24 Treatment algorithm for bipolar disorder patients with rapid and ultrarapid cycling. VPA, valproate; CBZ, carbamazepine.

If unimodal antidepressants are used for a bipolar disorder depression, clinical lore suggested that they should be tapered and discontinued as soon as possible, to avoid the potential for drug-induced switches and cycle acceleration, especially since lithium may not be able to prevent these phenomena. However, maintenance therapy with bupropion and lithium has been reported to be effective in rapid-cycling patients, and use of unimodal antidepressants in conjunction with the new putative mood stabilizers deserves study. Several case reports suggest that alprazolam (Xanax) may, like the tricyclic drugs, also induce switches into hypomania and mania (even in unpredisposed patients) and this high-potency benzodiazepine should be avoided in favor of the long half-life compounds clonazepam and lorazepam, which do not appear to share the proclivity of the triazolobenzodiazepine compounds for the induction of mania. These high-potency benzodiazepines may be useful adjuncts to the mood stabilizers; however, the rare patient may experience these classical benzodiazepines as mood destabilizing or even depressogenic.

Some evidence suggests that the MAOIs, in general, may be less prone than tricyclic drugs to induce switches. The MAOIs should be given relatively greater consideration, especially for the reversal of vegetative (anergic, hypersonic, or hyperphagic) states in the bipolar patient. A substantially higher rate of antidepressant response was reported in one controlled series for tranylcypromine (Parnate) (81 percent) compared with imipramine (Tofranil) (48 percent) in bipolar disorder patients. Clorgyline, a selective *MAO type A* inhibitor not yet clinically available has been reported to slow cycling frequency. The reversible inhibitor of monoamine exiduse type A (RIMA) moclobemide (Aurorix) (Table 14.8-3) is widely available in Europe and Canada but not in the United States; this drug is not believed to be as effective an antidepressant as the nonselective MAOIs, such as tranylcypromine and phenelzine (Nardil); however, these MAOIs are unique in potentiating all three amine systems (serotonin, noradrenaline, and dopamine). One could attempt such an equivalent effect by using venlafaxine (for its serotonergic and noradrenergic effects) in combination with bupropion (for its dopaminergic effects).

It is possible that the anticholinergic rather than the noradrenergic effects of the tricyclic drugs makes them prone to cause switches or cycle acceleration. A comparison of SSRIs with the noradrenaline selective agents (desipramine, nortriptyline [Aventyl], and maprotiline [Ludiomil]), or to the SNRI venlafaxine would clarify the putative role of norepinephrine reuptake blockade in inducing these phenomena.

Dopamine-Active Compounds and Other Treatments

Bupropion Bupropion in conjunction with a mood stabilizer has shown promise in short-term and prophylactic management of bipolar disorder patients, including rapid cyclers. Although it may be added to lithium or valproate prophylaxis without pharmacokinetic interactions, when used with carbamazepine its blood concentrations are markedly decreased and those of an active metabolite increased (Fig. 14.8-25).

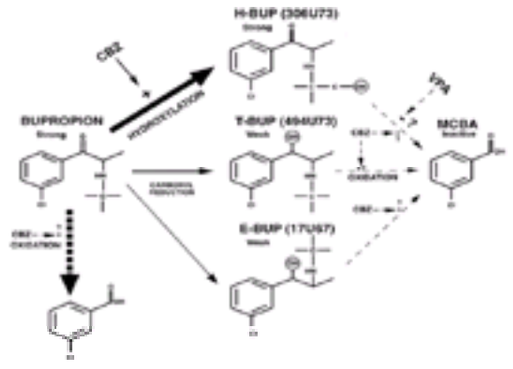

FIGURE 14.8-25 Bupropion metabolism. CBZ, carbamazepine; VPA, valproate; MCBA, metachlorobenzoic acid.

Psychostimulants The role of psychostimulants as short-term augmentation has not been systematically explored, although it is apparently widely used by some experts in the field. However, one investigator has indicated that this is not a useful long-term strategy, since many patients appear to develop tolerance to this

modality. This strategy should perhaps be reserved for temporary augmentation while awaiting more-effective antidepressant response to other modalities.

The same investigator also observed that tolerance does not appear to develop when the psychomotor stimulants are combined with MAOIs. This strategy should be reserved for only the most refractory patients, since the *Physician's Desk Reference* (PDR) lists an absolute contraindication to combining stimulants and MAOIs. Nonetheless, this appears to be effective and tolerated by most patients in many small case series.

Dopamine Agonists Small clinical series have also suggested some antidepressant efficacy of the direct dopamine agonist bromocriptine, which is used to treat parkinsonian patients. One double-blind study indicated that it was as effective as imipramine. A related dopamine agonist, piribedil (Trivastal), has been effective for the occasional refractory depressive patient. Pergolide (Permax) was reported to be an effective augmenting agent in one series on refractory depression but not in another.

Pramipexole, a potent D_3 as well as D_2 agonist recently approved for use in parkinsonism, is reported (at 1 mg a day or higher) to have antidepressant effects equivalent to those of fluoxetine (Prozac). Treatment should be started at low dosages and titrated toward 1 mg a day very slowly to avoid adverse effects such as nausea and orthostatic hypotension.

Dopamine-active drugs has been reported to be more effective in patients with low concentrations of the dopamine metabolite homovanillic acid in their cerebrospinal fluid (CSF). A similar relation to low concentrations of the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA) and a better response to the serotonin active compounds clomipramine (Anafranil) and sertraline (Zoloft) have been reported. The results are inconclusive as to whether urinary concentrations of the noradrenergic metabolite 3-methoxy-4-hydroxyphenylglycol (MHPG) can predict the response to noradrenergically active antidepressants. Consistent biochemical markers of antidepressant response have not yet been identified.

Light in Morning and Melatonin at Night Systematic trials of augmentation with bright light (more than 7500 lux) may also be worth considering in patients with marked disruption of circadian rhythmicity and the typical bipolar disorder hypersomnia. In these instances, high-intensity light might be more useful in the morning, although this issue needs to be reexamined with more-systematic prospective randomized studies.

An additional approach to altering sleep activity cycles (which are common in bipolar disorder patients) might be to use melatonin adjunctively at night, although this, too, requires caution and prospective studies. In addition, isolated reports exist of exacerbation of sleep or mood in some patients when using melatonin supplementation.

Sleep Deprivation as a Short-Term Antidepressant Paradoxically, sleep deprivation may be an adjunctive procedure to hasten antidepressant response (Fig. 14.8-26). A rapid, transient, antidepressant response to 1 night of sleep deprivation has been reported consistently in studies from many different laboratories. Preliminary evidence suggests that sleep deprivation in the last half of the night (from 3 to 7 AM) may be just as effective as total sleep deprivation and may thus be more convenient for clinical use and outpatient treatment.

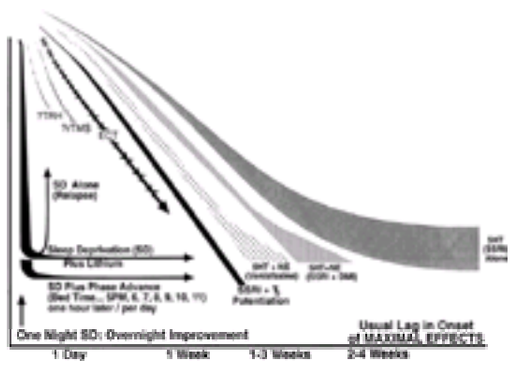

FIGURE 14.8-26 Nine potential clinical approaches to rapid-onset antidepressant effects. rTMS, repeated transcranial magnetic stimulation; NE, noradrenergic; DMI, desipramine.

Although most responsive patients relapse after 1 night's recovery sleep, some modalities (especially lithium) may help sustain the sleep deprivation response. One recent report also indicates that progressive changes in the hours of sleep (from 5 PM on the sleep deprivation day to 6 PM the next, and so forth toward an 11 PM bedtime) also help hold a sleep deprivation response.

The rapid onset (overnight) effects achieved differs from the slower but sustained effects following selective deprivation of rapid eye movement (REM) sleep, that is, a modality that is not readily amenable to clinical use. Response to sleep deprivation may be related to phase or duration of a bipolar disorder depressive episode, with less responsivity early in the episode and a greater responsivity late in the episode. Several, but not all, studies indicate that the degree of antidepressant response to sleep deprivation is correlated with the degree of increase in morning plasma TSH concentration, presumably driven by endogenous increases in TRH. Parenteral TRH administration (500 mg intravenously) has also been reported to have rapid-onset effects, and two case reports suggest more-sustained effects with low dosage (50 μg a day at bedtime).

Inositol Inositol (12 to 16 grams a day) has recently been reported to have antidepressant, antianxiety, and antiobsessive-compulsive effects. This remains to be more systematically explored in bipolar disorder patients (in light of no reports of patients switching in one series reporting a 50 percent response rate in patients who were concomitantly treated with lithium, carbamazepine, or valproate. Theoretically, inositol should not only relieve some lithium-induced adverse effects, but could potentially reverse its therapeutic efficacy to the extent that inositol depletion from reduced phosphoinositide (PI) turnover is related to lithium's mechanism of action. This has not been reported, however. Myoinositol measured by magnetic resonance spectroscopy (MRS) has been reported to be low in brains of bipolar disorder patients (in proportion to the severity of their depression); and lithium lowers it further. One would then predict that inositol augmentation might make carbamazepine even more effective, because carbamazepine has effects opposite to that of lithium, and increases the inositol-1 phosphatase enzyme, which should increase inositol. Lithium, carbamazepine, and valproate are all reported to downregulate the transporter for inositol that has recently been isolated and cloned.

Choline One open study reported that potentiation with choline (3 to 8 grams a day) may be helpful in stabilizing mood in refractory cyclers, and this approach, too, requires further systematic study.

Folate and Ascorbate In two single, randomized, placebocontrolled studies, the vitamins folate (300 to 400 mg a day) and ascorbate (3 grams a day) have each been reported to have some beneficial effects on amelioration of mood in long-term prophylaxis of bipolar disorder patients and, in light of their benign side-effect profiles, might be worthy of consideration in the treatment regimen of the refractory bipolar disorder patient.

Omega-3 Fatty Acids One randomized double-blind study comparing the addition of omega-3 fatty acids (9 grams a day) or an olive oil placebo in bipolar disorder patients poorly controlled on mood stabilizers was terminated early because of the marked superiority of the omega-3 fatty acid to the control. Virtually all of the breakthrough depressive episodes occurred in the placebo control group.

APPROACHES TO SUICIDALITY

Regular psychiatric visits during the prophylactic well phase are recommended on an interval ranging from weekly to biweekly in unstable patients to 1 to 4 months in better stabilized patients, depending on a variety of ancillary circumstances including completeness of response, lack of psychosocial crises, history of compliance, insight into the illness and its treatment, absence of adverse effects, financial constraints, and the wishes of the patient. In addition to periodic assessment of all of these issues, regular treatment visits are recommended to assess the potential risks of suicide independently of the occurrence of discrete episodes. This is particularly important when there is a positive family history of suicide or other risk factors, including male sex, older age, comorbid alcohol abuse, high levels of

anxiety, and prior suicide attempts (particularly if they have been severe).

Suicidal impulses and acts may not always vary directly with either severity of depression or reemergence of a full-blown episode requiring hospitalization and should be part of the careful ongoing clinical assessment of patients in all phases of their illness and treatment. Severe overwhelming psychic anxiety and agitation are predictors of completed suicide, and with the high comorbidity of panic disorder with either phase of bipolar disorder, one should be particularly alert to this and other high-risk factors. Specific contracting for communication with the clinician upon remergence or escalation of suicidal thoughts should be considered in high risk patients (Table 14.8-12).

Dual treatment: Focus acute short-term and prophylaxis
Mania: Treat first, do chemistries later
Load valproate and lithium; slowly start lamotrigine treatment
Use second mood stabilizers over antipsychotics
Benzodiazepines instead of antipsychotics
Combination treatment decreases adverse effects
Chart illness
Augment rather than substitute
Simplify (for adverse effects)
Taper off lithium slowly, if at all
Educate the patient's family
Assess compliance and suicidality
Develop an early warning system
Develop specific contracts
Regular visits; monitor course and adverse effects
Phone contact (PRN) when needed
Develop a "fire drill"
Prevent comorbid alcohol and other substance abuse
Psychotherapy and medicalization of illness
Give statistics: 50% relapse in first 5 months off lithium treatment
Patient as a co-principal investigator
Conservative treatment, if successful
Radical treatment, if inadequate response

Table 14.8-12 Principles in the Treatment of Bipolar Disorders

Electroconvulsive Therapy ECT may benefit bipolar depressed patients who do not respond to lithium or other mood stabilizers and their adjuncts. This is particularly true when intense suicidality presents as a medical emergency. Whether ECT would continue to help abbreviate each bout of recurrent depressive episodes in rapid-cycling patients or whether it would be useful in long-term prophylaxis must be further investigated. The author has observed several instances in which tolerance appeared to develop to the therapeutic effects of repeated series of ECT. Moreover, concern about cognitive adverse effects remains, and the author has recently seen a number of otherwise healthy individuals with rather profound and sustained retrograde memory loss.

TREATMENT OF MANIC BREAKTHROUGHS

A wide range of drugs is available for breakthrough manic episodes occurring during lithium or other prophylactic treatment, including the entire spectrum of drugs indicated for the treatment of acute mania (Table 14.8-5 and Table 14.8-6; Fig. 14.8-24). Ranking high in these treatments are carbamazepine and valproic acid, because of their longer-term prophylactic efficacy. Lamotrigine, nimodipine, and topiramate all require further study. Clonazepam or lorazepam may also be useful acute alternatives to antipsychotic supplementation, even though the benzodiazepines (and antipsychotic agents) appear to have a less primary role in the long-term management of bipolar disorders than the mood stabilizers carbamazepine or valproic acid.

The use of clozapine for bipolar disorder and schizoaffective disorder patients refractory to lithium, carbamazepine, and valproate is now well documented in many case series, especially for those with rapid cycling and dysphoric mania. Its lack of ability to induce tardive dyskinesia makes it particularly attractive, and one looks forward to a possible similar role for other atypical antipsychotics (Table 14.8-8 and Table 14.8-10) without clozapine's liability for agranulocytosis and the attendant need for weekly white blood count monitoring. Given the high liability of tardive dyskinesia even with intermittent use of typical antipsychotics, these drugs should be relatively avoided in favor of the atypical agents and other mood stabilizers whenever possible.

COMPLEX COMBINATION THERAPY

Data from the NIMH also support the notion that many depressed patients with refractory bipolar illness can be treated with a variety of approaches not typically involving the traditional unimodal antidepressant or typical antipsychotic modalities. Analysis of successive 5-year epochs in the tertiary-referral 3-West Clinical Research Unit of the Biological Psychiatry Branch showed that approximately the same high percentage (75 to 80 percent) of patients achieved marked or moderate improvement on the CGI at discharge in the past 25 years. However, patients in the years 1970 to 1974 required only monotherapy on discharge more than 75 percent of the time; this decreased to less than 25 percent in the most recent 5-year period, and the average number of medications on discharge increased from one to three per patient over the same time period (Fig. 14.8-27). Yet, unimodal antidepressants or antipsychotic drugs were used in less than 15 percent of the patients. Although these data based on sequential double-blind trials and augmentation strategies in this discharge phase of the hospitalization, did not involve a systematic randomized approach to therapeutics they nonetheless reveal that the vast majority of patients with refractory depressive and cycling presentations can be managed largely in the absence of the unimodal antidepressants or neuroleptic drugs. Multiple mood stabilizers in combination, often with thyroid augmentation, were used most of the time.

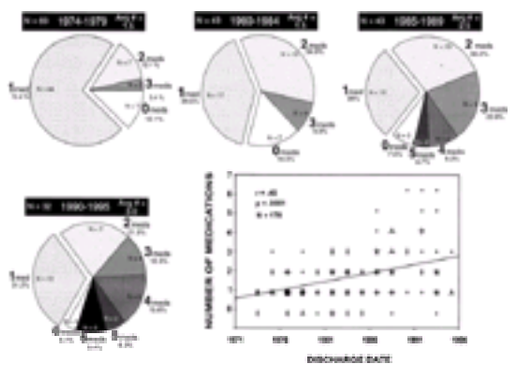

FIGURE 14.8-27 Increasing number of medications at discharge.

As in the late phases of other medical syndromes, complex combination treatment is often required for the treatment-refractory bipolar disorder patient. Although it may at first seem inappropriate to consider a regimen with four or five drugs for the treatment of refractory bipolar illness, this strategy is standard practice in many other areas of medicine such as the treatment of malignancies, tuberculosis, AIDS, or congestive heart failure. In these instances, multiple therapeutic modalities with different symptom targets and different mechanisms of action are often combined for optimal therapeutics.

With the availability of a variety of agents for bipolar disorders, it now befits the field to engage in more systematic clinical trial approaches to delineate the optimal strategies for achieving the most rapid response rates in the highest numbers of patients, so that even the most refractory bipolar disorder patients have an excellent chance at achieving and maintaining clinical remission.

Attempting to supplement the clinical effects of lithium and the mood-stabilizing anticonvulsants such as carbamazepine and valproic acid is often more effective than using lithium or the anticonvulsant alone. By augmenting rather than substituting a mood stabilizer, withdrawal-induced exacerbations will not confound the evaluation of the next agent, and time may be saved in the assessment of the combination in one clinical trial rather than two sequential trials (the anticonvulsant alone and then the combination). For patients who are unable to tolerate lithium carbonate, evidence does suggest that carbamazepine or valproate may be useful long-term maintenance treatments in preventing both manic and depressive episodes. A rather substantial literature on double-blind randomized studies provide evidence of carbamazepine's prophylactic efficacy. Most of the prophylactic data on valproate are based on clinical case series, however, with the exception of one randomized open trial showing efficacy equal to that of lithium and one blind trial showing that valproate has greater efficacy than placebo or lithium for recurrent depression. The

choice of carbamazepine or valproate may depend on the future development of clinical predictors or on the current assessment of their relative side-effects profiles (Table 14.8-13). The data are even more preliminary for the L-type calcium channel inhibitor and the new putative mood-stabilizing agents such as lamotrigine and possibly topiramate and gabapentin.

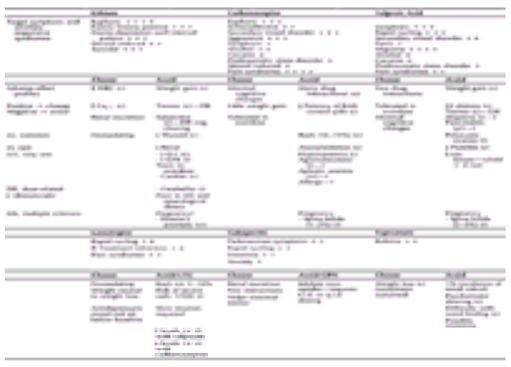

Table 14.8-13 Positive and Negative Selection Factors for Choice of Mood Stabilizer

All of these agents have differential presumptive biochemical mechanisms of action. One can begin to use augmenting strategies (Table 14.8-6), not only empirically (Fig. 14.8-5), but also based on combining agents with different actions (Table 14.8-2, Fig. 14.8-6 and Fig. 14.8-9), symptom targets (Table 14.8-4), side-effect profiles (Table 14.8-7), or benign pharmacokinetic and pharmacodynamic interactions. It is unclear whether using drugs with slightly different actions targeted to a single transmitter system such as GABA (i.e., enhancing GABA by reuptake blockade [tiagabine], transaminase inhibition [g-vinyl GABA], transport effects [gabapentin], or multiple enzymatic actions [valproate]) will be more or less effective than simultaneously targeting two or more entirely different systems (e.g., enhancing GABA and inhibiting glutamate (Fig. 14.8-9).

Using several drugs in combination (Table 14.8-6) may actually help reduce the incidence of adverse effects by keeping each drug below its adverse-effects threshold (Table 14.8-7) rather than pushing one drug to maximum dosages for full therapeutic effect. Preclinical evidence indicates that such a combination strategy may also be less susceptible to the loss of effectiveness via tolerance. Optimally, one would like to use drugs with additive or synergistic therapeutic effects in the relative absence of adverse effects (Table 14.8-13); for this, knowledge of both pharmacokinetic and pharmacodynamic interactions is of considerable importance.

Pharmacokinetic Interactions There do not appear to be major pharmacokinetic interactions among lithium and the other mood stabilizers. There have been occasional reports of neurotoxicity when lithium and carbamazepine are used together. Because both agents can cause these adverse effects at or below clinically accepted dosage ranges, neurotoxicity may occasionally result from the combination treatment as well. In most instances the combination appears to be well tolerated. Many of the adverse effects reported in the literature appear to be caused by starting treatment with relatively large dosages of carbamazepine (rather than using slow increases) in combination with other agents, and assuming that the attendant adverse effects are related to the combination treatment rather than to carbamazepine alone. Carbamazepine enhances its own metabolism and that of other agents metabolized by the hepatic microsomal P450 system (CYP) isoenzyme CYP, 3A4 such as haloperidol or estrogen, and concentrations of these substances (as well as birth control pills) are markedly reduced by carbamazepine (Fig. 14.8-28). Despite this reduction in haloperidol blood concentrations, most studies report improvement with carbamazepine supplementation, suggesting that carbamazepine might potentiate antipsychotic effects because of actions other than dopamine receptor blockade.

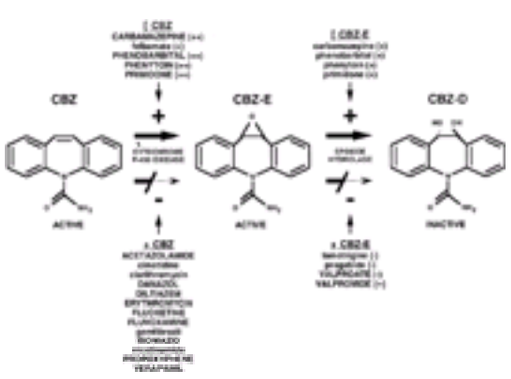

FIGURE 14.8-28 Carbamazepine (CBZ) metabolism.

Agents commonly used in medical practice can markedly increase carbamazepine concentrations and thus produce attendant toxicity (Table 14.8-14, Fig. 14.8-28). The most frequent dose-related toxic manifestations are dizziness, drowsiness, ataxia, diplopia, and confusion. These may occur in someone otherwise tolerating the drug well until other drugs are added. Erythromycin, triacetyloleandomycin (TAO), isoniazid (Nidrazid) (but apparently not other MAOIs), fluoxetine and fluvoxamine (Luvox), and the calcium channel inhibitors verapamil and diltiazem (Cardizem) (but not nifedipine, nimodipine, amlodipine [Norvasc], or isradipine) will increase carbamazepine concentrations. Less marked increases occur during cotreatment with propoxyphene (Darvon) and, transiently, with cimetidine (Tagamet). Carbamazepine will lower the blood concentrations of various agents (especially birth control pills) and interfere with some tests that depend on protein binding, including some pregnancy tests.

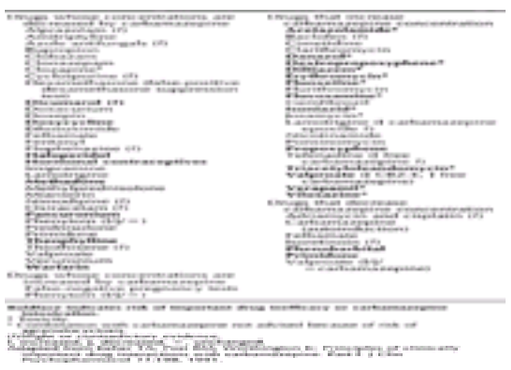

Table 14.8-14 Carbamazepine-Drug Interactions

In contrast to the multiple pharmacokinetic interaction of carbamazepine with other drugs, based largely on its properties as an inducer of hepatic P450 enzymes, valproate has few problematic effects in this area. However, if carbamazepine and valproate are used together, one should consider reducing the carbamazepine dosage (because valproate increases concentrations of the 10,11-epoxide metabolite) as well as increasing the free fraction of carbamazepine (Fig. 14.8-28). Lithium and valproate are generally well tolerated in combination, but effects on tremor, weight gain, and gastrointestinal distress may be additive.

Carbamazepine markedly reduces the concentration of bupropion but increases that of its active hydroxy metabolite (Fig. 14.8-25), an action not shared by valproate.

Valproate approximately doubles blood concentrations of lamotrigine and thus increases the likelihood of a severe lamotrigine rash; starting dosages of lamotrigine should be halved accordingly. If no rash occurs, lamotrigine and valproate have excellent tolerability and efficacy in epilepsy (and perhaps in affective illness) when used in combination. In contrast, carbamazepine decreases lamotrigine concentrations by about one-half, and these two drugs may be less useful in combination because of their similar adverse effects and pharmacodynamics.

Fluoxetine and paroxetine (Paxil) are inhibitors of CYP 2D6, and may increase concentrations of many agents such as desipramine and the secondary tricyclic antidepressants and phenothiazines (Table 14.8-15). Fluvoxamine is a potent inhibitor of CYP 1A2 (potentially producing theophylline toxicity) and CYP 2C19 (increasing warfarin [Coumadin] levels) as well as CYP 3A4. Nefazodone (Serzone) is also potent at CYP 3A4, producing cardiac arrhythmias with astemizole (Hismanal), cisapride (Propulsid), and terfenadine (Seldane), and likely increasing carbamazepine concentrations.

Table 14.8-15 Substrates, Inhibitors, and Inducers of Some Important Cytochrome P450 Isoforms

SENSITIZATION EFFECTS ON THE MOOD DISORDERS AND IMPLICATIONS FOR PROPHYLAXIS

Early clinical observations and more-recent systematic controlled studies suggest that recurrent bipolar affective disorders may undergo a transition from initial episodes that are typically precipitated by psychosocial stressors to later episodes that tend to occur more spontaneously. This transition often occurs in the context of an overall pattern of cycle acceleration with a trend toward shorter well interval between successive episodes (Fig. 14.8-29). Both the effects of psychosocial stressors and recurrent mood episodes themselves may not only cause acute biological perturbations, but may also leave behind residual biological vulnerabilities, or memory traces, based on their ability to alter gene expression (Fig. 14.8-30). Preliminary data from the Stanley Foundation Bipolar Treatment Outcome Network suggest that experiences of early verbal, physical, or sexual abuse are associated with a greater incidence of early-onset bipolar disorder and ultra-ultra-rapid (ultradian) cycling; physical abuse in childhood was associated with increased mania and sexual abuse with increased suicidality.

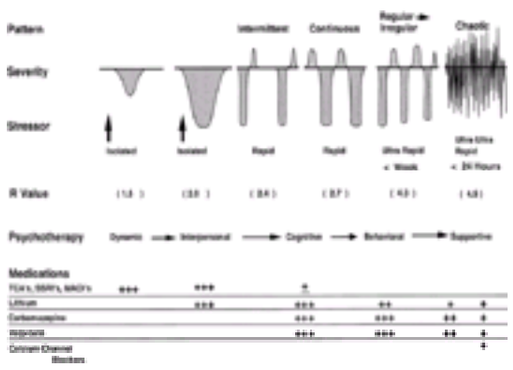

FIGURE 14.8-29 Phases in evolution of mood cycling: potential relation to treatment response.

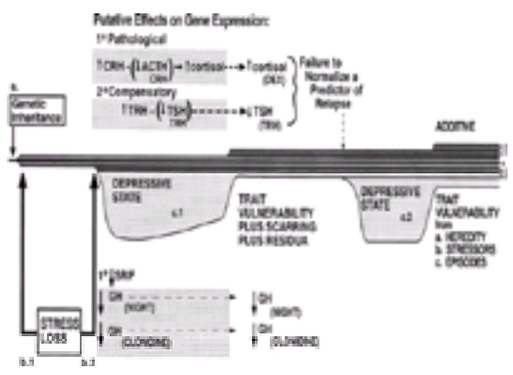

FIGURE 14.8-30 Accumulating experiential genetic vulnerability in recurrent mood disorders. CRH, corticotropin-releasing hormone; ACTH, adrenocorticotropic hormone; DEX, dexamethasone; SRIF, somatostatin release-inhibiting factor.

Following stress and episode-induced changes in neurotransmission, a cascade of neurobiological effects is thought to take place that includes not only short-term adaptations but also longer-lasting alterations initiated by a variety of transcription factors, including the immediate early genes, such as the transcription factors c-*fos* and c-*jun* as well as a variety of neurotrophic factors. These transcription factors can then induce changes in the long-term regulation of late effector genes (i.e., transmitters, receptors, and neuropeptides). This type of episode-related upregulation in transcript could account for the increase in corticotropine releasing hormone (CRH) in brain and CSF of depressed patients and its associated hypercortisolemia as well as the microstructural and synaptic organization of the brain revealed in many models of maternal deprivation and learning and memory.

If this conceptualization proves correct, it suggests the potential dual importance of preventing episodes of affective illness. Not only would episode-associated morbidity and potential mortality be prevented, but the longer-lasting neurobiological vulnerabilities associated with the experience of repeated affective episodes themselves (sensitization) might be attenuated as well. Although the mechanisms and direct causality have not been clarified, greater numbers of prior mood episodes are associated with greater degrees of cognitive dysfunction on a variety of neuropsychological tests in euthymic bipolar disorder outpatients.

Thus, given the increasing evidence that greater numbers of affective episodes are a poor prognostic sign and may be associated with resistance to effective treatment with lithium, both the clinical and theoretical data speak to the importance of early institution and long-term maintenance of prophylaxis, particularly in patients already identified as being at high risk for recurrence.

Treatment efficacy may thus also vary as a function of the stage of illness evolution (Fig. 14.8-29). For example, pharmacotherapies such as lithium and unimodal antidepressants may be more effective in initial and midphases of the illness, but with the emergence of rapid and ultra-rapid cycling, adjunctive treatments with the anticonvulsants may be required, with the antidepressants used more sparingly. In the latest stages of illness, sometimes associated with ultradian cycling,

dihydropyridine *L*-type calcium channel inhibitors augmented by mood-stabilizing anticonvulsants may be required.

Similar transitions may occur in psychotherapeutic approaches wherein psychotherapy may be effective in the early, milder forms of stress-related depressive illness, but with major recurrent episodes (and particularly, melancholic and psychotic syndromes), aggressive acute and maintenance pharmacotherapy may be mandatory. Adjunctive interpersonal, cognitive, and behavioral techniques may become increasingly important in (1) the later, more automatic shifts in mood, (2) helping to maintain active treatment compliance with pharmacotherapies, (3) implementing an early intervention system based on the development of a structured early warning system; and (4) maintaining morale in the face of therapeutic adversity and incomplete response, all of which, in turn, may prevent suicide.

The past several decades have seen important advances in the understanding of the neurobiology of the depressive and bipolar disorders. Clearly these illnesses involve multiple areas of brain dysfunction and affect a variety of organ systems, altering not only mood, but also motor, cognitive, sleep, appetite, hedonic reward, and other somatic systems. Neurobiological alterations are evident at the level of endocrine dysfunction, reflected not only in altered regulation of glucocorticoids and CRH, TRH, and somatostatin release-inhibiting factor (SRIF) but even the size of the pituitary and the adrenals.

Some of these changes may involve episodic and cyclic alterations in gene transcription related to the primary pathology of the illness (such as increases in CRF), but others may be secondary and adaptive (e.g., increases in TRH) (Fig. 14.8-30). The author has postulated that the ratio of a host of pathological factors ("the bad guys") to the compensatory adaptations ("the good guys") determines the proportions of periods of illness versus well intervals (Fig. 14.8-31). To the extent that exogenous medications can change this ratio by either inhibiting the bad guys or enhancing the good guys, sustained remissions can be achieved, if adequate prophylactic treatment is maintained. This conceptual model based on preclinical data also predicts that clinical loss of prophylactic efficacy via development of tolerance would be delayed or prevented by (1) earlier rather than later treatment; (2) sustained full dosages rather than marginally effective ones; (3) use of the most effective agents with a wide therapeutic margin; and (4) combinations of several marginally effective agents (Table 14.8-16, Fig. 14.8-32). These propositions remain to be directly tested, however.

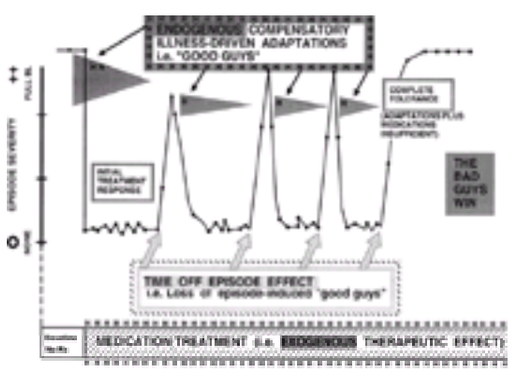

FIGURE 14.8-31 Ratio of "good guys" and "bad guys" drives episode cycling.

Treatment resistance showed by
Higher dosages
Not escalating dosages
More efficacious drugs: valproate carbamazepine
Treatment initiated early in illness
Combination treatment: carbamazepine plus valproate
Reducing illness drive
Response restored by
Period of drug discontinuation then reexposure
Agents with different mechanisms of action: no cross-tolerance
Future studies re predictive validity:
Maximum tolerated dosages
Stable dosing
Differential rate of treatment resistance?
Studies by Gelenberger et al.
O'Connell et al, Sarantidis and Waters, Denicoff et al*
Combination ?= monotherapy?
Treat comorbidities
Randomized study of continuation vs. discontinuation
? Response to gabapentin or lamotrigine

* Denicoff KD, Smith-Jackson EE, Disney ER, Ali SO, Leverich GS, Post RM: Comparative prophylactic efficacy of lithium, carbamazepine, and the combination in bipolar disorder. J Clin Psychiatry 58:470, 1997; Calabrese JR, Kane JM, Kalfler MB, Levori P, Rosenbaum JF, Cole K, Lavelle J: Comparison of standard and low serum levels of lithium for maintenance treatment of bipolar disorder. N Engl J Med 321:1489, 1989; O'Connell RA, Mayo JA, Flatow L, Cuthbertson B, O'Brien BE: Outcome of bipolar disorder on long-term treatment with lithium. Br J Psychiatry 159:123, 1991; Sarantidis D, Waiters B: Predictors of lithium prophylaxis effectiveness. Prog Neuropsychopharmacol 3:507, 1981.

Table 14.8-16 Other Clinical Predictions From the Preclinical Model of Tolerance Development

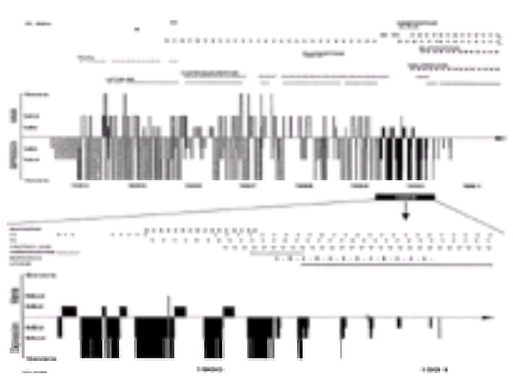

FIGURE 14.8-32 Response to combination therapy in a patient with refractory bipolar I disorder. TCAs, tricyclic antidepressants.

Functional brain imaging has revealed alterations in blood flow and glucose utilization, often reflecting hypofrontality in depression, at times in direct proportion to the severity of the depressive syndrome. Evidence of hyperactivity in the ventral anterior cingulate gyrus and the medial temporal lobes is also present in some patients, and these differential patterns may be linked to differences in therapeutic outcome. Both patient and clinician should be aware that a wealth of clinically relevant neurobiological evidence indicates that the mood disorders are potentially life-threatening medical illnesses not different from those that afflict the other major organ systems of the body and, as such, should be treated with equal care and vigor.

FUTURE DIRECTIONS

During this most exciting era for the development of psychobiological theories and therapies for the mood disorders, one hopes for clearer definitions of the different psychotherapies and pharmacotherapies critical for adequate therapeutic intervention in individual patients whose severities, types, patterns, and stages of illness differ. Since the last half of the century, successive generations of pharmacotherapies have led to critical neurobiological hypotheses of the mechanisms underlying the mood disorders, and one hopes that a continued mutually interactive process of more specific treatments, theories, and therapies will be derived from synergistic clinical and basic research in this area.

Not only is there a wealth of information to be learned through controlled research, but each patient has much to teach the practitioner and the theoretician. Precise life charting of the course of recurrent mood disorders and its response to treatment may be invaluable to the patient and clinician in arriving at optimal therapeutics, particularly when evidence from controlled studies currently provides so little direction for most crucial decisions. Although a host of well-tested, promising treatment alternatives are now available and one can anticipate many novel interventions in the near future, a much wider systematic clinical research base is urgently needed to guide the physician to the best therapeutic regimen for each individual bipolar patient. The author hopes that some of the preliminary data, guidelines, and principles outlined in this chapter will assist in this process and that more rapid progress occurs in the study and treatment of this relatively neglected major

psychiatric illness.

SUGGESTED CROSS-REFERENCES

Biological therapies are discussed in Chapter 31. Section 14.7 provides a thorough discussion of the treatment of depressive (unipolar) disorders. Obsessive-compulsive disorder is covered in Chapter 15. The rest of Chapter 14 can be consulted for other aspects of mood disorders.

SECTION REFERENCES

Altshuler LL, Post RM, Leverich GS, Mikalauskas K, Rosoff A, Ackerman L: Antidepressant-induced mania and cycle acceleration: A controversy revisited. Am J Psychiatry *152:*1130, 1995.

Bauer MS, Whybrow PC: Rapid cycling bipolar affective disorder, II. Treatment of refractory rapid cycling with high-dose levothyroxine: A preliminary study. Arch Gen Psychiatry *47:*435, 1990.

Baumgartner A, Bauer M, Hellweg RTMS: Treatment of intractable non-rapid cycling bipolar affective disorder with high-dose thyroxine: An open clinical trial. Neuropsychopharmacology *10:*183, 1994.

Beydoun A, Uthman BM, Sackellares JC: Gabapentin: Pharmacokinetics, efficacy, and safety. Clin Neuropharmacol *18:*469, 1995.

Bocchetta A, Chillotti C, Severino G, Ardau R, Del Zompo M: Carbamazepine augmentation in lithium-refractory bipolar patients: A prospective study on long-term prophylactic effectiveness. J Clin Psychopharmacol *17:*92, 1997.

Bowden CL, Brugger AM, Swann AC, Calabrese JR, Janicak PG, Petty F, Dilsaver SC, Davis JM, Rush AJ, Small JG, Garza-Trevino ES, Risch SC, Goodnick PJ, Morris DD: Efficacy of divalproex vs lithium and placebo in the treatment of mania. JAMA *271:*918, 1994.

Calabrese JR, Bowden CL, Rhodes LJ, McElroy SL, Cookson J, Anderson J, Woyshville MJ, Keck PE Jr, Kundu K, Ascher JA, Paterson G, Tvarno K, Bolden-Watson C: Lamotrigine in treatment-refractory bipolar disorder (abstract). Presented at the 149th annual meeting of the American Psychiatric Association, May 4–9, 1996. Abstract no. 36, p 15, 1996.

*Calabrese JR, Bowden CL, Sachs GS, Ascher JA, Monaghan E, Rudd GD: A double-blind placebo-controlled study of lamotrigine monotherapy in outpatients with bipolar I depression. Lamictal 602 Study Group. J Clin Psychiatry *60:*79, 1999.

Calabrese JR, Kimmel SE, Woyshville MJ, Rapport DJ, Faust CJ, Thompson PA, Meltzer HY: Clozapine for treatment-refractory mania. Am J Psychiatry *153:*759, 1996.

Calabrese JR, Woyshville MJ, Kimmel SE, Rapport DJ: Predictors of valproate response in bipolar rapid cycling. J Clin Psychopharmacol *13:*280, 1993.

Coppen A, Chaudhry S, Swade C: Folic acid enhances lithium prophylaxis. J Affect Disord *10:*9, 1986.

Denicoff KD, Smith-Jackson EE, Bryan AL, Ali SO, Post RM: Valproate prophylaxis in a prospective clinical trial of refractory bipolar disorder. Am J Psychiatry *154:*1456, 1997.

Denicoff KD, Smith-Jackson EE, Disney ER, Ali SO, Leverich GS, Post RM: Comparative prophylactic efficacy of lithium, carbamazepine, and the combination in the treatment of bipolar disorder. J Clin Psychiatry *58:*470, 1997.

*Dubovsky SL: Calcium channel antagonists as novel agents for manic-depressive disorder. In *Textbook of Psychopharmacology*, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1995.

Eikmeier G, Berger M, Lodemann E, Muszynski K, Kaumeier S, Gastpar M: Trimipramine—an atypical neuroleptic? Int Clin Psychopharmacol *6:*147, 1991.

Faedda GL, Baldessarini RJ, Tohen M, Strakowski SM, Waternaux C: Episode sequence in bipolar disorder and response to lithium treatment. Am J Psychiatry *148:*1237, 1991.

Fawcett J, Kravitz HM, Zajecka JM, Schaff MR: CNS stimulant potentiation of monoamine oxidase inhibitors in treatment-refractory depression. J Clin Psychopharmacol *11:*127, 1993.

Feighner JP, Herbstein J, Damlouji N: Combined MAOI, TCA, and direct stimulant therapy of treatment-resistant depression. J Clin Psychiatry *46:*206, 1985.

Fitton A, Goa KL: Lamotrigine. An update of its pharmacology and therapeutic use in epilepsy. Drugs *50:*691, 1995.

*Frye MA, Ketter TA, Leverich GS, Huggins T, Lantz C, Denicoff KD, Post RM: The increasing use of polypharmacotherapy for refractory mood disorders: Twenty-two years of study. J Clin Psychiatry *60:*152, 1999.

Frye MA, Ketter TA, Kimbrell TA, Cora-Locatelli G, Dunn RT, Post RM: Gabapentin and lamotrigine monotherapy in mood disorder. Presented at the symposium New Anticonvulsants in Mood Disorder, at the 150th meeting of the American Psychiatric Association San Diego, CA, May 1997.

Gelenberg AJ, Kane JM, Keller MB, Lavori P, Rosenbaum JF, Cole K, Lavelle J: Comparison of standard and low serum levels of lithium for maintenance treatment of bipolar disorder. N Engl J Med *321:*1489, 1989.

Gitlin MJ, Swendsen J, Heller TL, Hammen C: Relapse and impairment in bipolar disorder. Am J Psychiatry *152:*1635, 1995.

Goodwin FK, Jamison KR: *Manic-Depressive Illness.* Oxford University Press, New York, 1990.

Gyulai L, Jaggi J, Bauer MS, Younkin S, Rubin L, Attie M, Whybrow PC: Bone mineral density and L-thyroxine treatment in rapidly cycling bipolar disorder. Biol Psychiatry *41:*503, 1997.

Haykal RF, Akiskal HS: Bupropion as a promising approach to rapid cycling bipolar II patients. J Clin Psychiatry *51:*450, 1990.

Himmelhoch JM, Thase ME, Mallinger AG, Houck P: Tranylcypromine versus imipramine in anergic bipolar depression. Am J Psychiatry *148:*910, 1991.

Hoschl C, Kozeny J: Verapamil in affective disorders: A controlled, double-blind study. Biol Psychiatry *25:*128, 1989.

Ketter TA, Post RM: Clinical pharmacology and pharmacokinetics of carbamazepine. In *Anticonvulsants in Mood Disorders*, RT Joffe, JR Calabrese, editors. Marcel Dekker, New York, 1994.

Kores B, Lader MH: Irreversible lithium neurotoxicity: An overview. Clin Neuropharmacol *20:*283, 1997.

*Koukopoulos A, Reginaldi D, Minnai G, Serra G, Pani L, Johnson FN: The long term prophylaxis of affective disorders. Adv Biochem Psychopharmacol *49:*127, 1995.

Pazzaglia PJ, Post RM, Ketter TA, George MS, Marangell LB: Preliminary controlled trial of nimodipine in ultra-rapid cycling affective dysregulation. Psychiatry Res *49:*257, 1993.

*Pazzaglia PJ, Post RM, Ketter TA, Callahan AM, Marangell LB, Frye MA, George MS, Kimbrell TA, Leverich GS, Cora-Locatelli G, Luckenbaugh D: Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. J Clin Psychopharmacol *18:*404, 1998.

Perez V, Gilaberte I, Faries D, Alvarez E, Artigas F: Randomised, double-blind, placebo-controlled trial of pindolol in combination with fluoxetine antidepressant treatment. Lancet *349:*1594, 1997.

Post RM, Denicoff KD, Frye MA, Leverich GS: Algorithms for bipolar mania. In *Mood Disorders, Systematic Medication Management* AJ Rush, editor. Karger, Basel, 1997, p 114.

*Post RM, Ketter TA, Denicoff K, Pazzaglia PJ, Leverich GS, Marangell LB, Callahan AM, George MS, Frye MA: The place of anticonvulsant therapy in bipolar illness. Psychopharmacology *128:*115, 1996.

Post RM, Ketter TA, Pazzaglia PJ, Denicoff K, George MS, Callahan A, Leverich G, Frye M: Rational polypharmacy in the bipolar affective disorders. Epilepsy Res (*11:*153, 1996.

Post RM, Leverich GS, Altshuler L, Mikalauskas K: Lithium discontinuation-induced refractoriness: Preliminary observations. Am J Psychiatry *149:*1727, 1992.

Post RM, Leverich GS, Denicoff KD, Frye MA, Kimbrell TA, Dunn RTMS: Alternative approaches to refractory depression in bipolar illness. Depression Anxiety *5:*175, 1997.

*Post RM, Pazzaglia PJ, Ketter TA, Denicoff K, Weiss SRB, Hough C, Chuang D-M, Stein R, Frye M: Carbamazepine and nimodipine in affective illness: Efficacy, mechanisms of action, and interactions. In *Pharmacotherapy for Mood and Cognition*, SA Montgomery, U Halbreich, editors. American Psychiatric Press, Washington, DC, 1999.

Post RM, Weiss SRB: The neurobiology of treatment-resistant mood disorders. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Post RM, Weiss SRB: A speculative model of affective illness cyclicity based on patterns of drug tolerance observed in amygdala-kindled seizures. Mol Neurobiol *13:*33, 1996.

Post RM, Weiss SRB: Kindling and stress sensitization. In *Bipolar Disorder: Biological Models and Their Clinical Application*, RT Joffe, LT Young, editors. Marcel Dekker, New York, 1997.

*Post RM, Weiss SRB, Clark M, Chuang DM, Hough C, Li H: Lithium, carbamazepine, and valproate in affective illness: Biochemical and neurobiological mechanisms. In *Mechanisms of Action of Bipolar Treatments*, HK Manji, CL Bowden, RH Belmaker, editors. American Psychiatric Press, Washington, DC, 1999.

*Sachs GS, Lafer B, Stoll AL, Banov M, Thibault AB, Tohen M, Rosenbaum JF: A double-blind trial of bupropion versus desipramine for bipolar depression. J Clin Psychiatry *55:*391, 1994.

Sachs GS, Rosenbaum JF, Jones L: Adjunctive clonazepam for maintenance treatment of bipolar affective disorder. J Clin Psychopharmacol *10:*42, 1994.

Stoll AL, Sachs GS, Cohen BM, Lafer B, Christensen JD, Renshaw PF: Choline in the treatment of rapid-cycling bipolar disorder: Clinical and neurochemical findings in lithium-treated patients. Biol Psychiatry *40:*382, 1996.

*Strakowski SM, McElroy SL, Keck PE Jr: Efficacy of valproate in bipolar illness: Comparisons and contrasts with lithium. In *Pharmacotherapy for Mood and Cognition*, SA Montgomery, U Halbreich, editors. American Psychiatric Press, Washington, DC, 1999.

Swann AC, Bowden CL, Morris D, Calabrese JR, Petty F, Small J, Dilsaver SC, Davis JM: Depression during mania: Treatment response to lithium or divalproex. Arch Gen Psychiatry *54:*37, 1997.

Tondo L, Baldessarini RJ, Hennen J, Floris G: Lithium maintenance treatment of depression and mania in bipolar I and bipolar II disorders. Am J Psychiatry *155:*638, 1998.

Vestergaard P: Treatment and prevention of mania: A Scandinavian perspective. Neuropsychopharmacology *7:*249, 1992.

Weiss SRB, Clark M, Rosen JB, Smith MA, Post RM: Contingent tolerance to the anticonvulsant effects of carbamazepine: Relationship to loss of endogenous adaptive mechanisms. Brain Res Rev *20:*305, 1995.

Wolf C, Berky M, Kovacs G: Carbamazepine versus lithium in the prophylaxis of bipolar affective disorders: A randomised, double-blind 1-year study in 168 patients (abstract). Eur Neuropsychopharmacol *7:*S176, 1997.

Young LT, Robb JC, Patelis-Siotis I, MacDonald C, Joffe RT: Acute treatment of bipolar depression with gabapentin. Biol Psychiatry *42:*851, 1997.

14.9 MOOD DISORDERS: PSYCHOTHERAPY

ROBERT M. A. HIRSCHFELD, M.D., AND M. TRACIE SHEA, PH.D.

The interpersonal nature of depression was noted and emphasized in the earliest psychoanalytic writings on depression, as was the centrality of the regulation of self-esteem. In *Mourning and Melancholia*, Sigmund Freud stated that a vulnerability to depression caused by an interpersonal disappointment early in life led to future love and relationships marked by ambivalence. Actual or threatened interpersonal losses in adult life trigger a self-destructive struggle in the ego that is manifested as depression. That theory was significantly refined by later psychoanalysts who described the depression-prone personality as needing constant reassurance, love, and admiration and being dependent on others for narcissistic gratification and maintenance of self-esteem. Frustration of these dependency needs leads to decreased self-esteem and subsequent depression. This notion was later expanded to include any person with a fragile self-esteem system. Another dynamic approach focuses on the cognitive aspects of depression, highlighting the recognition of the disparity between one's actual and idealized situation. That realization leads to a sense of helplessness and powerlessness and ultimately to depression.

Little empirical research has been conducted and published on psychoanalytic and psychodynamic approaches to depression. Psychotherapies have been subjected to clinical trials similar in design to those used to test new antidepressant medications. In addition, modern treatment trends have favored more-focused, short-term therapies at the expense of longer-term therapies, such as most of the psychodynamic therapies.

Although psychoanalytic approaches were the predominant mode of treatment for depression in the early to middle part of this century, now many types of psychotherapy based on a variety of concepts are in use. Psychotherapeutic approaches have been developed specifically for depression that aim to correct specific manifestations including cognition, behavior, and affect. In general, those treatments are short term and seek to alleviate the depressive condition, not to change the character of the patient.

INTERPERSONAL THERAPY (ITP)

Theoretical Concepts Interpersonal therapy was developed by Gerald Klerman and Myrna Weissman as part of their extensive research on the nature and treatment of depression over the past two decades. The theoretical basis of interpersonal therapy includes the work of Adolf Meyer and Harry Stack Sullivan. In contrast with the predominantly intrapsychic orientation of classic psychoanalysis and Emil Kraepelin's biomedical model, Meyer's psychobiological approach emphasizes the interaction between the individual and the psychosocial environment over the patient's entire life course. The patient's current interpersonal experiences and attempts to adapt to environmental change and stress are seen as critical factors in psychiatric illness. Sullivan's interpersonal theory, which views interactions between people as the focus for study and treatment in psychiatry, draws heavily from the social sciences, including anthropology and sociology. A second major influence comes from John Bowlby's studies of attachment. These studies demonstrate the importance of attachment and social bonding to human functioning and the connection between disruption of these bonds and vulnerability to depression.

Interpersonal therapy conceptualizes depression from a medical model: depression is something that happens to the person that requires treatment. The depressed person is allowed to assume the sick role and is not blamed for the affliction any more than someone would be blamed for having cancer, heart disease, or pneumonia. The issue of attribution of blame is important. Many other approaches view depression as something patients have brought on and must end by their own efforts.

The interpersonal therapy approach to depression involves three interacting components: symptom formation, social and interpersonal experiences, and enduring personality patterns. Medication may be recommended or symptom reduction; psychotherapy focuses on improving the patient's interpersonal functioning. Although the causes of depression may vary with regard to a person's biological vulnerability, personality predispositions, or psychosocial precipitants, depression always occurs in a psychosocial and interpersonal context. Depression can predispose a patient to interpersonal problems, or interpersonal problems can precipitate depression. An interpersonal focus in the treatment process is thus presumed essential for recovery.

Goals Interpersonal therapy sets two therapeutic goals: (1) to reduce the patient's depressive symptoms and improve self-esteem and (2) to help the patient develop more-effective strategies for dealing with current social and interpersonal relations. As a short-term psychotherapy, interpersonal therapy does not attempt to restructure the patient's personality. It does, however, recognize the importance of early developmental experiences and assumes the importance of early developmental experiences and that historical conflicts are manifested in current relationships.

General Considerations Interpersonal therapy, a short-term psychotherapy normally consisting of 12 to 16 weekly sessions, was developed specifically to treat nonbipolar, nonpsychotic, ambulatory patients suffering depressive disorders. It is characterized by an active approach on the part of the therapist and by an emphasis on current issues and social functioning in the life of the patient. Intrapsychic phenomena such as defiance mechanisms of internal conflicts are not addressed. Discrete behaviors such as lack of assertiveness, social skills, or distorted thinking may be addressed, but only in the context of their meaning or effect on interpersonal relationships.

Strategies and Techniques

General Strategies For goal 1, reduction of symptoms, an educational approach is used. The patient is told about the clinical syndrome of depression, including its components and course. The therapist reviews the symptoms with the patient, gives a sense of optimism and hope, and emphasizes that depression is a common disorder with a good prognosis. Pharmacotherapy may be considered for symptom reduction if appropriate.

For goal 2, interpersonal therapy defines four major problem areas commonly presented by depressed patients: grief, interpersonal role disputes, role transitions, and interpersonal deficits (Table 14.9-1). Associated therapeutic goals and recommended treatment strategies are outlined for each. The choice of specific strategies and techniques depends on the problem area considered most salient for the patient. The four areas are not mutually exclusive, and patients may have multiple problems in more than one area; however, the focus is on only one or two current interpersonal problems, to allow realistic goals and productive treatment strategies. Abnormal grief may involve delayed or distorted mourning or both. The following example is cited in the interpersonal therapy manual.

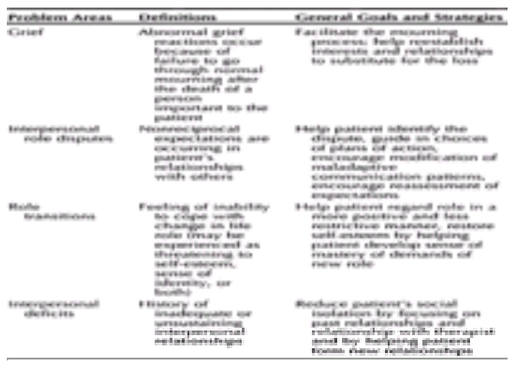

Problem Areas	Definitions	General Goals and Strategies
Grief	Abnormal grief reactions occur because of failure to go through normal mourning after the death of a person important to the patient	Facilitate the mourning process; help reestablish interests and relationships to substitute for the loss
Interpersonal role dispute	Nonreciprocal expectations are occurring in patient's relationships with others	Help patient identify the dispute, guide in choices of plans of action, encourage modification of maladaptive communication patterns, encourage reassessment of expectations
Role transitions	Feeling of inability to cope with change in life role (may be experienced as threatening to self-esteem, sense of identity, or both)	Help patient regard role in a more positive and less restrictive manner, restore self-esteem by helping patient develop sense of mastery of demands of new role
Interpersonal deficits	History of inadequate or unsustaining interpersonal relationships	Reduce patient's social isolation by focusing on past relationships and relationships with therapist and by helping patient form new relationships

Table 14.9-1 Focal Problem Areas of Interpersonal Therapy

A 68-year-old woman became depressed following the death of her husband, who had suffered a long course of physical and mental deterioration that resulted in considerable constraints and isolation on the part of the patient. Her symptoms included pervasive sadness and preoccupation with feelings of guilt and hopelessness. The first aim of treatment was to help the patient successfully mourn the loss, as the mourning process had been blocked by anger. The second aim was to help her to reestablish interests and relationships to substitute for what she had lost.

Interpersonal issues in a troublesome, conflicted marriage may include role disputes or role transitions. The choice between the two problem areas depends on whether the patient believes that the marriage is salvageable and whether the patient wants to stay in the marriage. If the patient decides to leave the marriage and the problem area is defined as role transition, the therapist attempts to help the patient make the transition. That goal may include identifying new sources of emotional support, overcoming irrational fears and regarding the new role more positively, and helping the patient master the demands of the new role. Alternatively, if the problem area is defined as role dispute, the treatment strategies include identifying the dispute and working toward its resolution, improving communication patterns, examining appropriateness of expectations, outlining various options, and deciding on a plan of action.

The interpersonal deficit problem area is appropriate for patients who are socially isolated or who have a sufficient number of relationships but feel unable to enjoy them. Interpersonal deficits may exist in patients who are chronically depressed and experience chronically impaired interpersonal functioning. Problems with social isolation may be longstanding or temporary; for each, treatment strategies aim to reduce social isolation. In the absence of current relationships, discussion of positive and negative features of past relationships may be used as a model for the development of new relationships. Treatment may also focus on the relationship between therapist and patient. The following example of an interpersonal deficit is cited by the interpersonal therapy manual.

A 22-year-old unmarried man became severely depressed 1 month after the breakup of a 3-year relationship with his girlfriend. The patient, a part-time student employed as a cook, lived with his mother, who had stopped working after being hospitalized for physical problems; subsequently, he had become depressed. Discussion of the patient's current relationship revealed that he felt close to no one but his mother.

The patient's history revealed inadequate social relationships and lack of interpersonal skills. Treatment focused on past significant relationships and conflicts over his relationship with his mother. The patient-therapist relationship provided a direct source of information about the patient's style of relating to others, which was used to modify maladaptive interpersonal patterns and improve his ability to form relationships with others.

Specific Techniques The specific techniques used in interpersonal therapy may be applied to any of the four interpersonal problem areas. In the general order of their use in the course of treatment, they are (1) exploratory techniques, (2) encouragement of affect, (3) clarification, (4) communication analysis, (5) use of therapeutic relationship, and (6) behavior change techniques (Table 14.9-2).

Techniques	Definition
Exploratory techniques	Collect (by directive or nondirective methods) information about the patient's symptoms and problems
Encouragement of affect	Help patient recognize and accept painful affects, help patient use and manage affects positively in interpersonal relationships, encourage expression of suppressed affect
Clarification	Restructure and feed back patient's communications
Communication analysis	Identify maladaptive communication patterns, help patient communicate more effectively
Use of therapeutic relationship	Examine patient's feelings and behaviors in therapeutic relationship as model of patient's interactions in other relationships
Behavior change	Use to help patient solve simple life problems, teach patient to consider range of options for solving problems, use role playing to explore and understand patient's relationship with others, and train patient in new ways of interacting with others

Table 14.9-2 Interpersonal Therapy Techniques

Efficacy In depression interpersonal therapy has been tested in two large controlled studies. The first involved four groups (approximately 25 outpatients) treated by interpersonal therapy alone, interpersonal therapy plus amitriptyline (Elavil), amitriptyline alone, and a nonscheduled treatment comparison group. All active treatments, including interpersonal therapy alone, were significantly more effective in reducing depressive symptoms than nonscheduled treatment; the combination of interpersonal therapy and amitriptyline was most effective. In addition, the interpersonal therapy groups had much lower dropout rates than those without interpersonal therapy.

In the second study, the National Institute of Mental Health (NIMH) Treatment of Depression Collaborative Research Program, 250 outpatients with major depressive disorder were randomly assigned to one of four 16-week treatment conditions: interpersonal therapy, cognitive-behavioral therapy, imipramine (Tofranil) with clinical management, and placebo with clinical management. In this study all four treatment conditions significantly reduced depressive symptoms. For severely depressed patients interpersonal therapy was significantly more effective than placebo with clinical management in achieving remission of symptoms at 16 weeks. Imipramine with clinical management tended to have the best outcome, particularly for patients with impaired functioning. Imipramine yielded more-rapid effects, with significantly better outcome than all other conditions at 12 weeks. A reanalysis of the NIMH Treatment of Depression Collaborative Research Program supported the superiority of antidepressant medication for depressions that are more severe on the basis of symptoms or psychosocial status.

Among psychotherapies, depressive severity did not appear to predict outcome to treatment. Interpersonal therapy was compared with supportive therapy in depressed human immunodeficiency virus (HIV)-positive patients. Although both treatments reduced depressive symptoms, interpersonal therapy was superior, and gains were sustained.

Interpersonal therapy has been adapted and tested in several different patient types. Interpersonal therapy late-life maintenance, developed for geriatric patients, incorporates approaches to special problems of elderly adults and allows shortened sessions for those who cannot sustain 50-minute sessions. Interpersonal therapy has been modified for use in adolescents, incorporating developmental and interpersonal issues frequently occurring in this age group.

Interpersonal therapy has also been adapted for use in chronic depression, specifically dysthymia. Pilot results from a 16-week open trial of patients with dysthymia were extremely positive in this group of difficult-to-treat patients. A version for patients with bipolar disorder includes material aimed at normalizing sleep-wake cycles, whose alterations often trigger a manic episode. Interpersonal therapy has been adapted specially for depressed patients with marital disputes. A study comparing patients randomly assigned to regular interpersonal therapy or the adapted therapy found similar reductions in depressive symptoms, but better marital outcomes (e.g., greater marital affection) in those receiving the adapted version. In addition to these modifications for different populations with depression, interpersonal therapy has been modified for use in other psychiatric disorders; among these are interpersonal counseling for distress, for hospitalized elderly patients, for substance abuse, and for bulimia nervosa.

COGNITIVE-BEHAVIORAL THERAPY

Theoretical Concepts Cognitive-behavioral therapy stems from four major previous theories: psychoanalytic theory, phenomenological philosophy, cognitive psychology, and behavioral psychology. Salient common features include recognition of the importance of the subjectiveness of conscious experience (one's perceptual experience of reality rather than the objective reality) and recognition of the emotional consequences of irrational beliefs and thoughts.

Aaron Beck, the originator of cognitive-behavioral therapy, developed a comprehensive, structured theory of depression. According to this theory, depression is associated with negative thought patterns, specific distorted schemas, and cognitive errors or faulty information processing (Table 14.9-3). Such cognitive

dysfunctions form the core of depression, while affective and physical changes and other associated features of depression are its consequences.

Element	Definition
Cognitive triad	Beliefs about oneself, the world, the future
Schemas	Ways of organizing and interpreting experiences
Cognitive distortions	
Arbitrary inference	Drawing a specific conclusion without sufficient evidence
Specific abstraction	Focus on a single detail while ignoring other more important aspects of an experience
Overgeneralization	Forming conclusions based on too little and too narrow experience
Magnification and minimization	Over- or undervaluing the significance of a particular event
Personalization	Tendency to self-reference external events without basis
Absolutist, dichotomous thinking	Tendency to place experience into all-or-none categories

Table 14.9-3 Elements of Cognitive Theory

Cognitive theory conceptualized depression as involving negative cognitions regarding the cognitive triad (ideas of oneself, the world, and one's future). The self is perceived as defective, inadequate, deprived, worthless, and undesirable. The world appears as a negative, demanding, and defeating place, and one expects failure and punishment, continued hardship, deprivation, and failure in the future. Underlying the negative conditions are stable cognitive structures, called schemas, that include core beliefs or assumptions through which one interprets experience. Schemas associated with depression are analogous to viewing the world through dark glasses (e.g., the core belief that one is unlovable). Cognitive errors, or systematic errors in thinking, allow the persistence of negative schemas despite contradictory evidence. A cognitive error frequently associated with depression is dichotomous thinking, the tendency to view one's experiences as black or white without shades of gray or to believe that people are either all bad or all good. Symptoms of depression follow from the cognitive error. For example, apathy and low energy result from the individual's expectation of failure in all areas. Similarly, a paralysis of will stems from the individual's pessimism and feeling of hopelessness.

Goals The goal of cognitive-behavioral therapy is to change the way a person thinks and, subsequently, to alleviate the depressive syndrome and prevent its recurrence. This is accomplished by helping the patient (1) identify the test negative conditions; (2) develop alternative, more flexible schemas, and (3) rehearse both new cognitive and new behavioral responses.

General Considerations Cognitive-behavioral therapy is a short-term, structured therapy that involves active collaboration between the patient and therapist in achieving set goals. It is oriented toward current problems and their resolution. Therapy is usually conducted on an individual basis, although group techniques have been developed and tested. Cognitive-behavioral therapy may be used in conjunction with pharmacotherapy.

Strategies and Techniques As with other psychotherapies, the attributes of the therapist are fundamental to successful cognitive-behavioral therapy. Therapists must be empathetic, able to understand the life experience of each patient, and capable of being genuine and honest with themselves and their patients. Therapists also must be able to relate skillfully to patients in their own experiential world in an interactive way. As a highly structured therapeutic approach, cognitive-behavioral therapy involves setting the agenda at the beginning of each session, assigning homework to be performed between sessions and teaching specific new skills. The active collaboration between the therapist and the patient provides a genuine sense of teamwork.

Cognitive-behavioral therapy has three basic components: didactic aspects, cognitive techniques, and behavioral techniques (Table 14.9-4).

Didactic issues
Learning rationale and strategy of the therapy
Cognitive techniques
Eliciting automatic thoughts
Testing automatic thoughts
Identifying maladaptive underlying assumptions
Analyzing validity of maladaptive assumptions
Behavioral techniques
Scheduling activities
Mastering activities
Graded task assignment
Cognitive rehearsal
Self-reliance training
Role playing
Diversion techniques

Table 14.9-4 Components of Cognitive-Behavioral Therapy

Didactic Aspects The didactic aspects include explaining the cognitive triad, schemas, and faulty logic to the patient. The therapist tells the patient that they will formulate hypotheses together and test them over the course of treatment. The therapist presents a full explanation of the relation between depression and thinking, affect, and behavior as well as the rationale for all aspects of the treatment. This contrasts with the more psychoanalytically oriented therapies, which involve little explanation.

Cognitive Techniques The cognitive approach has four strategies: eliciting automatic thoughts, testing automatic thoughts, identifying maladaptive underlying assumptions, and testing the validity of maladaptive assumptions.

ELICITING AUTOMATIC THOUGHTS Automatic thoughts are cognitions that intervene between external events and the individual's emotional reaction to the event. For example, a man invited to go bowling may think, negatively, “everyone is going to laugh at me when they see how badly I bowl,” before he actually bowls with this group of people. Another example is when a woman thinks “that person doesn't like me” if someone passes her in the hall without saying hello.

TESTING AUTOMATIC THOUGHTS The therapist, acting as teacher, helps the patient test the validity of the automatic thought. The goal is to encourage the patient to formulate alternative possible interpretations and reject inaccurate or exaggerated automatic thoughts, after carefully examining them. For example, patients often set unrealistic expectations for themselves, then blame themselves when they cannot live up to these expectations.

A 32-year-old depressed computer programmer had self-denigrating thoughts about his ability to complete homework assignments.

Patient: I don't know what's been wrong with me this week. I just don't seem to be as interested in doing my homework assignments. I don't know if I'm ever going to get better.

Therapist: Can you think of a specific time this week that you had problems doing homework because of disinterest?

Patient: Yes, on Thursday I tried to do my relaxation exercises, but I eventually gave up.

Therapist: Can you tell me what you were thinking at the time?

Patient: Well, I started doing my breathing, but I couldn't calm my thoughts and stop thinking about other things, like the instructions in the manual said. Then I started thinking about how long I've been working on this and how I should know how to do it by now.

Therapist: And how long have you been working on the breathing technique?

Patient: Uh, one week.

Therapist: Let's review the evidence that supports your statement that you should be performing this exercise with no problem at this time.

In this example, when the patient and therapist carefully reviewed the situation, it became apparent that the patient's expectation that he should be able to perform this exercise perfectly after 1 week of practice was unreasonable. After considering that the ability to breathe and maintain calm thoughts is a skill that normally takes many weeks to perfect, the patient realized that his belief about his inability to learn was distorted and incorrect.

Generating alternative explanations is another technique used to undermine inaccurate and distorted automatic thoughts.

A 29-year-old secretary with a 2-year history of depression reported that she frequently felt sad and hurt at work because of the curt and gruff way her boss interacted with her. She reported an automatic thought following an interaction in which her boss stated “I wish things around here ran smoother.” “He doesn't like me. He doesn't think I'm doing a good job.” The therapist helped her generate a list of other interpretations of her employer's statement and behavior including the possibility that he interacted with all people this way, that he was a generally unhappy person, that he did not like his job and was allowing this unhappiness to influence how he interacted with the patient, and that he was preoccupied with personal problems that made him unhappy at work and inattentive to the way he interacted with his employees.

IDENTIFYING MALADAPTIVE ASSUMPTIONS As the patient and therapist continue to identify automatic thoughts, patterns usually become apparent that represent underlying rules or maladaptive general assumptions that guide the patient's life. Examples of such rules include, “To be happy, I must be perfect” or “If everyone doesn't like me, I'm not lovable.” Such rules inevitably lead to disappointment, failure, and subsequent depression.

ANALYZING MALADAPTIVE ASSUMPTIONS Testing the accuracy of maladaptive assumptions is similar to testing the validity of automatic thoughts. In one particularly effective technique the therapist asks the patient to defend the validity of an assumption.

Patient: I guess I believe that I should always work up to my potential.

Therapist: Why is that?

Patient: Otherwise I would be wasting time.

Therapist: What is the long-range goal in working up to your potential?

Patient: I've never really thought about that. I've just assumed that I should.

Therapist: Are there any positive things you give up by always having to work up to your potential?

Patient: I suppose it makes it hard to relax or take a vacation.

Therapist: What about living up to your potential to enjoy yourself and relax? Is that important at all?

Patient: I've never really thought of it that way.

Therapist: Maybe we can work on giving yourself permission not to work up to your potential at all times.

In this example, the therapist is helping the patient recognize how maladaptive it is to strive to work up to one's potential at all times.

Behavioral Techniques Behavioral techniques are used conjointly with cognitive techniques to test and change maladaptive or inaccurate cognitions, to help patients understand the inaccuracy of their cognitive assumptions and learn new strategies and ways of dealing with issues. A repertoire of behavioral techniques is used in cognitive-behavioral therapy.

1. Among the first things done is scheduling activities on an hourly basis. The patient keeps a record of these activities and reviews it with the therapist.
2. Patients are asked to rate their mastery of these activities and the pleasure derived from them. They are often surprised how much more mastery and pleasure they gain than they had believed.
3. To simplify the situation and allow miniaccomplishments, tasks are often subdivided into subtasks, as in graded task assignments, to show patients they can succeed.
4. In cognitive rehearsal the patient first imagines the various steps involved in meeting and mastering a challenge and then rehearses various aspects of them.
5. Self-reliance training involves encouraging patients to become more self reliant by doing such simple things as making their own beds, doing their own shopping, or preparing their own meals, rather than relying on other people.
6. Role playing, a particularly powerful and useful technique, is used to elicit automatic thoughts and learn new behaviors.
7. Diversion techniques help patients get through particularly difficult times by means of physical activity, social contact, work, play, or visual imagery.

The techniques used are highly structured and goal oriented and require active collaboration between the therapist and patient. Emphasis is on identifying maladaptive, inaccurate cognitions in various forms, seeking alternative explanations, and learning new behaviors to reverse the affective and drive disturbances and other associated features of depression and (it is hoped) help prevent their recurrence.

Efficacy Cognitive therapy has been studied extensively in the treatment of outpatients with major depressive disorder. Of 34 such reports, 9 included a pill placebo, waiting list, or nonspecific treatment as a control. In most studies, cognitive-behavioral therapy was superior to the control condition in reducing depressive symptoms. The one notable exception is the NIMH Treatment of Depression Collaborative Research Program, in which cognitive-behavioral therapy did not differ significantly from the placebo clinical management condition. Cognitive-behavioral therapy was found to be superior to pharmacotherapy alone in two studies conducted in the 1970s. In three more-recent studies, including the Treatment of Depression Collaborative Research Program, there were no differences in efficacy between antidepressant medication and cognitive behavioral therapy. However, in the NIMH study, imipramine (Tofranil) was significantly more effective than cognitive-behavioral therapy for more severely depressed and impaired patients. In six studies that compared cognitive-behavior therapy alone with that therapy plus

pharmacotherapy, five found no differences between the two outcomes and one found the combined treatment superior to cognitive-behavioral therapy alone.

BEHAVIORAL APPROACHES

Theoretical Concepts Although a number of behavioral approaches to depression exist, each with somewhat different theoretical assumptions and specific treatment methods, they have a common source in the work of B. F. Skinner, who incorporated the principles of classical and operant conditioning in an empirical analysis of behavior. Skinner's research provides the basic framework, methodology, and assumptions for current behavioral theories and their clinical applications. Application of that model to complex human behavior led some theorists to expand the framework. For example, social learning theory includes cognitive phenomena, such as emphasizing the role of subjective expectations and value in reinforcement. Although interested in the role of cognition, behavioral theorists assume that cognitions follow the same laws of learning followed by more-observable behavioral events and, while related, do not determine behavior in a causal sense. This assumption distinguishes behavioral approaches from the cognitive-behavioral approach described below. Despite some differences in focus, behavioral therapies are commonly characterized by an emphasis on (1) the links between an observable or operationally definable behavior and the conditions that control or determine it and (2) the role of rewards or reinforcement in determining behavioral change.

The behavioral approach was first applied to depression in 1965 by Charles B. Ferster, who proposed that depression is caused by a person's loss of positive reinforcement (e.g., through separation, death, or sudden environmental change), which results in reduction of the entire behavioral repertoire, depressed behavior, and dysphoric feelings. That concept of depression is central to all behavioral approaches. A change in the rate of reinforcement is believed to be a key factor in generating and maintaining depression (through lack of available reinforcers or when the available reinforcers are not contingent on the person's behavior) and also reversing it. Ferster also proposed that a social skills deficit—characterized by difficulty in obtaining social reinforcement—might increase a person's difficulty in coping with the loss of the usual supply of reinforcement.

Goals The goals of the behavior therapies are to increase the frequency of the patient's positively reinforcing interactions with the environment and decrease the number of negative interactions. Some behavioral treatments aim also at improving social skills. Alteration of personal behavior is believed to be the most effective way to change the associated depressed thoughts and feelings.

At this time behavioral approaches have been incorporated into cognitive approaches to the treatment of depression. An exception to this is the Coping with Depression Course developed by Peter Lewinsohn and colleagues at the University of Oregon, which encourages increasing the number of pleasurable events and improving social skills in a group format.

SELECTION OF TREATMENT APPROACH

The availability of several effective treatment approaches for the treatment of patients with depression means that the clinician must choose a treatment strategy for an individual patient with depression, which raises several questions. When is it appropriate to use these short-term psychotherapies as a sole treatment? Are there patients for whom the use of antidepressants, a less expensive treatment strategy, is equally or more beneficial? When is it warranted or advisable to use a combined-treatment approach including the use of psychotherapy and antidepressants simultaneously or sequentially? In addition, there is the question of whether the different types of psychotherapy might be more or less effective for different kinds of patients

A wealth of studies have examined the association between patient characteristics (including features characterizing the depression) and outcome within and across different psychotherapeutic and antidepressant treatments. Findings have often been inconsistent, but some themes have arisen that have important implications for treatment choice.

Who Should Receive Psychotherapy and Who Should Get Antidepressants? The psychotherapeutic treatment approaches described above were developed for use with outpatients with nonbipolar, nonpsychotic depressions. They were neither developed for nor intended for sole use of severely depressed inpatients or for patients with bipolar disorder depression, although evaluation of their use in combination with pharmacotherapy for such patients has begun.

Nonetheless, within the population of nonbipolar, nonpsychotic outpatients with major depressive disorder, the question of whether or not the use of psychotherapy alone can provide an outcome equal to that with antidepressants and if so, for which outpatients is a controversial subject. There is a long-standing clinical belief that antidepressants should be used for more severely depressed patients or for endogenous depressions. Findings regarding endogenous depression have been mixed, but most show that outpatients with endogenous depression (as defined by the Research Diagnostic Criteria) do not respond better to pharmacotherapy than to psychotherapy. Antidepressants are preferred for patients with melancholia marked by vegetative symptoms and psychomotor disturbances. Outpatients with major depressive disorder are characterized by a range of symptom severity and a range of impaired functioning associated with the depression. The NIMH Treatment of Depression Collaborative Research Program examined the efficacy of cognitive-behavioral therapy, interpersonal therapy, pharmacotherapy (imipramine plus clinical management), and pill placebo plus clinical management for more- and less-severely depressed outpatients with major depressive disorder. Severity was defined in two ways: one based on symptom severity and one based on symptoms and impaired functioning. For the patients with more-severe symptoms, interpersonal therapy was comparable to imipramine treatment, suggesting that interpersonal therapy may be an effective alternative treatment for patients who cannot or do not want to use medications. Imipramine treatment was superior to both psychotherapies for patients with more severely impaired functioning. Patients with less severe depression by both definitions showed no differences in response to the treatments, including placebo plus clinical management. Other studies examining the comparative efficacy of cognitive-behavioral therapy and antidepressants with more severely depressed patients (defined by symptom severity) have not found an advantage for pharmacotherapy. Such studies have not examined whether treatment differences exist for patients with more severely impaired functioning.

Choice of Psychotherapy Providing efficacious psychotherapy the treatment of depression requires trained clinicians in the various modalities. Whether different patients respond preferentially to specific forms of psychotherapy is important for several reasons, including the ability to provide more-efficient, targeted treatments as well as increasing the understanding of the mechanisms by which these treatments effect change. To date, studies examining predictors of treatment response to psychotherapy have been disappointingly inconsistent and inconclusive. Few studies have included more than one form of psychotherapy to allow examination of predictors of differential treatment response within the same study.

The NIMH study found significant interactions between treatment and pretreatment levels of social functioning and dysfunctional attitudes. Patients with relatively less impairment in social functioning responded better to interpersonal therapy than those with more impairment. Patients with lower levels of dysfunctional attitudes did better in cognitive-behavioral therapy than those with higher levels. Other studies have also shown that high scores on measures of dysfunctional attitudes predict a poorer outcome in cognitive-behavioral therapy although one study reported no association. Together these findings suggest that patients may require a minimal level of proficiency in the area of functioning that the treatment targets to benefit from the treatment, at least in the short term.

In general, the more severe the dysfunctional attitudes, the poorer the response to cognitive-behavioral therapy. Since this therapy theoretically and practically focuses on identification and reversal of negative cognitions about oneself, the world, and the future, one would expect that patients with dysfunctional attitudes would profit most from this specific therapy. Most of the studies support just the opposite conclusion. One study did find that a negative life event (e.g., divorce, or death of a loved one) mitigated this association and rendered severity of dysfunctional beliefs nonpredictive.

The beliefs and expectations of the patient regarding depression and treatment should also be considered. Some patients who consider depression to be a psychological disorder that should be amenable to psychotherapeutic approaches are resistant to using medication. Others consider their depression to be a biochemical disturbance that will require medication to be corrected, not psychotherapy. A good therapist may be able to modify such expectations when necessary, but a positive attitude toward treatment on the part of the patient may be significantly important to a successful outcome.

In general, the therapist should be cautious in making attributions about premorbid personality problems during the depressed phase. Many interpersonal and cognitive styles may appear different to the patient and the therapist after the acute phase of the disorder has been alleviated. Nonetheless, several studies have found that the presence of a personality disorder is associated with a slower or generally worse response to treatment. For depression, such patients are likely to need longer periods of treatment.

Role of Biological Measures Evidence is mounting that depressions associated with biological abnormalities respond less well to psychotherapy. In a study of patients with recurrent major depression, 50 patients had normal sleep profiles and 41 had abnormal profiles (in terms of sleep efficiency, rapid eye movement [REM] latency, and REM density). All patients received short-term interpersonal therapy. Those with abnormal sleep profiles responded significantly more poorly than those with normal sleep profiles, and most nonresponders responded to subsequent pharmacotherapy. In a study of unmedicated inpatients with major depressive disorder,

those with abnormally high hypothalamic-pituitary-adrenocortical activity were less responsive to cognitive-behavioral therapy than those with normal hypothalamic-pituitary-adrenocortical activity. Other investigations focused on sleep abnormalities. Reduced REM sleep latency was associated with a poor response to placebo and a favorable response to antidepressants; studies examining reduced REM latency and psychotherapy response have not shown it to predict a poorer response. Michael Thase and colleagues studied a composite measure of sleep abnormality, including REM latency, REM density, and sleep efficiency, and its association with response to cognitive-behavioral therapy or interpersonal therapy. Patients with an abnormal sleep profile have a poorer response to both forms of psychotherapy. These patients were also more likely to have recurrent episodes of depression. The abnormal sleep profiles were not simply markers of depression severity.

Some other patient characteristics have been found to predict response across treatments in general. Longer duration of the current episode and the diagnosis of dysthymic disorder prior to the onset of the major depressive disorder (double depression) predict a poorer response; higher expectations of improvement are associated with better outcomes with different forms of treatment.

Combined Pharmacotherapy and Psychotherapy When should patients be treated with a combination of psychotherapy and pharmacotherapy? Clinical practice commonly provides both forms of treatment; the rationale is that the antidepressant medication is most effective for such vegetative symptoms as sleep disturbance, appetite disturbance, and anhedonia, while psychotherapy improves marital and family relationships, social functioning, and occupational performance. The empirical evidence supporting this belief is minimal.

Nonetheless, is combined treatment more efficacious than either treatment modality alone? Is there a subgroup of patients for whom combined treatment may be more efficacious? Is combined therapy of more benefit when a broader range of outcome measures (e.g., different aspects of functioning, time to remission, prophylactic effects) is considered? Numerous studies have compared combined treatment with single-modality treatment; sometimes pharmacotherapy alone, sometimes psychotherapy alone, and sometimes both single modalities in the same study. The results of such studies have been inconsistent but, more often than not, have shown no clear advantage to the use of combined treatment over a single modality. Results from a meta-analysis showed a modest advantage of combined treatment. More recently, data from several treatment studies of nonbipolar, nonpsychotic, major depressive disorder, using similar assessment and standardized treatment protocols, were combined in a meta-analysis of original data from the same institution, which yielded a much larger sample than previously studied. The single-modality psychotherapy protocols included either cognitive-behavioral therapy or interpersonal therapy; the combined modality was interpersonal therapy plus antidepressant pharmacotherapy. Combined treatment was not significantly more effective in the milder (less severe symptoms) depressions; however, among patients with more-severe and recurrent depressions, combined treatment was of highly significant advantage.

Thus, while no clear, definitive answers exist, evidence is accumulating that patients with milder depressions may be offered psychotherapy first. Patients with more-severe symptoms and more seriously impaired functioning should receive treatment that includes an antidepressant. Biological measures, including indicators of hypothalamic-pituitary-adrenocortical axis abnormality and sleep abnormalities, hold promise as indicators of differential treatment assignment but unfortunately are not feasible for use in most clinical settings currently. The research on combined treatment to date has focused primarily on acute-phase treatment and on simultaneous administration of both forms of treatment. The strategy of combining psychotherapy and pharmacotherapy in sequenced treatments remains to be examined. For more severe depression characterized by impaired ability to think and concentrate, for example, it may make sense to begin treatment with antidepressant treatment, adding psychotherapy when cognitive functioning and energy levels improve.

PHASE OF ILLNESS

Nearly all studies of psychosocial treatments for depression have focused on the acute phase of treatment; that is, they have tested the performance of a specific psychotherapeutic approach in resolving depressive symptoms within 12 to 16 weeks. These studies have generated considerable evidence of the efficacy of interpersonal therapy, cognitive-behavioral therapy, and behavioral therapy in certain groups of patients during this period. However, an episode of depression does not necessarily end when the acute symptoms have abated. In fact, symptoms may relapse if treatment is discontinued too soon after their initial control; presumably because the short-term treatment (especially pharmacotherapy) has not cured the illness but ameliorated or reduced the symptoms temporarily. This situation is analogous to the effect of insulin on diabetes mellitus. Depression is now recognized as recurrent and often chronic; therefore, withdrawal of the treatment may result in return of the illness.

An important consequence of our recognition of the long-term nature of the illness is the need for treatment beyond the acute phase, into the continuation and maintenance phases. Continuation treatment is the ongoing treatment from the point of clinical remission to the point at which spontaneous remission is expected in untreated patients (i.e., to the putative true end of an untreated episode). For depression, the continuation phase in pharmacological treatments generally lasts approximately 6 to 9 months following acute treatment. Maintenance treatment is longer term and is intended to prevent future depressive episodes or decrease their intensity. The model for psychotherapeutic treatments, in contrast, is that the strategies and techniques change maladaptive patterns that are linked to depression and thus should reduce the risk for future episodes or symptoms of depression. Do short-term psychotherapies confer a prophylactic effect in the future? Following a positive response, is it helpful to continue the treatments into the continuation and maintenance phases?

Prophylactic Effect of Short-Term Therapies Follow-up studies of patients responding positively to short-term treatment for depression have attempted to determine whether treatment offers long-term prophylactic effects.

For interpersonal therapy, one study reported no difference in relapse or recurrence at 1-year follow-up among patients in a 16-week clinical trial who were treated with interpersonal therapy, amitriptyline, amitriptyline plus interpersonal therapy, or nonscheduled treatment. However, patients treated with interpersonal therapy had better social functioning at a 1-year reevaluation point. Most follow-up studies have examined relapse rates in patients successfully treated with cognitive therapy or antidepressant medication. These studies have shown a clear pattern of lower relapse rates for patients treated with cognitive-behavioral therapy than for those treated with short-term pharmacotherapy. However, the naturalistic designs of these studies precludes conclusions regarding the reasons for the difference (i.e., did the results represent enduring effects of the cognitive therapy or differences in risk for relapse between patients who respond to drugs and those who respond to psychotherapy). The question of what aspects of cognitive-behavioral therapy (specifically, behavioral activation, automatic thought modification and change of core beliefs) might be responsible for the lower relapse rates that have been observed was examined in a study providing one, two, or all three components and assessing outcome over 2 years.

Despite finding that cognitive behavioral therapy decreases symptoms and possibly lowers relapse rates, the perceived success of both psychotherapy and pharmacological treatments depends on how outcome is defined. When outcome is defined as complete remission of symptoms and maintenance of symptom-free remission for an extended period following treatment, 12 to 16 weeks of treatment (with psychotherapy or pharmacotherapy) is insufficient for most patients with major depression. The NIMH Treatment of Depression Collaborative Research Program reported the proportion of all patients starting treatment who achieved complete remission (at least 8 weeks without symptoms) at the end of treatment and maintained remission for 18 months: 30 percent of patients receiving cognitive-behavioral therapy, 26 percent of those receiving interpersonal therapy and 19 percent of those receiving placebo with clinical management. Another study of cognitive-behavioral therapy with a 2-year follow-up similarly reported that 25 percent of patients recovered and remained in remission. Considering optimal outcome highlights the need for longer treatment for full recovery as well as the need for continuation and maintenance treatments.

Treatment During the Continuation Phase Does continued treatment after successive resolution of symptoms help prevent relapses and recurrences? This clinically important question has received some attention in recent years for cognitive-behavioral therapy. In one study 42 subjects who received short-term therapy were followed for 1 year. At 3 months into the follow-up study, half of those who responded to treatment were given additional treatment (booster sessions) until completion of the study while the other half of the responder group received no additional treatment. The authors found no difference in relapse rates or depressive symptoms between the two groups at 1 year, suggesting that continued treatment with cognitive therapy after successful resolution of symptoms does not improve outcome. However, this is a single study, and further research is needed. There are no studies on the use of interpersonal or behavior therapy in the continuation phase. Another effectiveness study examined whether providing cognitive-behavioral therapy following successful antidepressant treatment of recurrent major depressive disorder reduced relapse rates. Cognitive-behavioral therapy reduced residual symptoms following drug discontinuation and reduced relapse rates over 2 years compared with clinical management.

Treatment During the Maintenance Phase Does therapy continued a year or more after successive treatment help prevent the occurrence of new episodes? In a landmark study by Ellen Frank and colleagues, 128 patients with recurrent major depressive disorder who had responded to a combined short-term and continuation treatment of imipramine and interpersonal therapy were randomly assigned to different maintenance treatment groups. The results strongly support long-term, continued treatment (especially pharmacotherapy) in patients with a history of recurrent episodes of depression. Some evidence indicates that long-term treatment is useful in delaying or preventing recurrences in patients with recurrent depression. However, the value of continuation and maintenance treatment with psychotherapy

awaits further research.

DYSTHYMIC DISORDER AND CHRONICITY

Most of our knowledge on the treatment of depression comes from the study of patients with acute major depression, and far less is known about the treatment of chronic depression. This is unfortunate, given the prevalence of dysthymic disorder. Over 3 percent of adults in the United States suffer from dysthymic disorder during any 6-month period, according to the Epidemiologic Catchment Area (ECA). In addition, approximately one-third of psychiatric outpatients suffer from dysthymic disorder. Nearly one in five patients with a major depressive episode fails to recover and becomes chronically depressed.

The importance and potential usefulness of psychosocial treatments for such patients is demonstrated by (1) the notable morbidity and impaired quality of life associated with dysthymic disorder, which exceeds that associated with most medical illnesses; (2) the fact that a substantial proportion of patients with dysthymic disorder either fail to respond to medication or cannot tolerate the adverse effects; and (3) with or without medication, the long-standing patterns of social withdrawal, lack of assertiveness, impaired family, marital, and occupational functioning, and chronic pessimism and hopelessness associated with dysthymic disorder need to be addressed. When depression is severe, pharmacological treatments are encouraged to alleviate suffering and increase the patient's ability to engage in the therapy. There are no controlled studies of the effectiveness of this treatment approach; however, two naturalistic follow-up studies have suggested that long-term analytic therapy can have long-term benefits.

More-recent developments in the treatment of dysthymic disorder include modification of psychotherapeutic approaches specifically for the treatment of dysthymic disorder as well as preliminary open-trial studies investigating their effectiveness. The chronic interpersonal and social deficits associated with dysthymic disorder provide a strong rationale for the use of interpersonal therapy with dysthymic patients, and a manual was recently developed. Aspects of dysthymic disorder that distinguish it from acute depression and require modification of interpersonal therapy include the lack of an acute precipitant, the characterological features often associated with the presence of a chronic mood disorder (e.g., paucity of interpersonal relationships, lack of self-assertion, poor social skills), and the lack of euthymic memories. Given the typical absence of an acute precipitant in dysthymic disorder, choosing a focus of treatment becomes more difficult. While all four interpersonal therapy problem areas occur in dysthymic patients, their frequency as a primary focus differs from that in acute depression. Grief is rarely the primary focus, interpersonal deficits more frequently are. The frequent absence of interpersonal relationships in the patient's life requires increased focus on the therapeutic relationship, which is used as a model for other interpersonal interactions.

Participation in activities is used to examine social behaviors, expectations, and desires. Relationships that do exist are often unsatisfactory because of the difficulty these individuals have in asserting themselves, expressing anger, or setting limits. Interpersonal therapy for these patients emphasizes exploring what the patient desires from the relationships and what options are available to alter the relationships. The patient is helped to begin to identify personal needs, assert them, and set limits. The expression of anger is encouraged and supported. Preliminary support for the use of interpersonal therapy for dysthymic disorder was provided by a nonrandomized pilot study of 19 patients treated with interpersonal therapy, desipramine (Norpramin), or both.

Cognitive-behavioral therapy was adapted for patients with dysthymic disorder by James McCullough. This modification was piloted in an open study of 10 patients, 9 of whom improved substantially and sustained their remissions for 2 years. The goal of the therapy is to teach the patient to accept total responsibility for the depression and to achieve and maintain mood control by enacting adaptive daily-living strategies. The approach, cognitive-behavior therapy for the chronic depression is being tested in a multicenter random-assignment study of it and nefazodone (Serzone) alone or in combination in a sample of patients with chronic and double depression.

PSYCHOSOCIAL TREATMENT OF BIPOLAR DISORDER

The clinical and research literature on major depressive disorder is replete with both psychotherapeutic and psychopharmacological approaches. In sharp contrast, the literature on treatment of bipolar disorders focuses almost exclusively on psychopharmacology, specifically on the use of lithium (Eskalith) as the treatment of choice. The introduction of lithium has influenced the diagnostic system, clinical practice, and the therapeutic outcome of patients with bipolar disorders for 20 years. It is as close to a wonder drug as has been experienced in psychiatry.

However, lithium is far from a cure for bipolar disorders; about one-third of patients either do not respond or respond only partially. Even for those who respond fully, many serious social, occupational, familial, and marital problems often remain.

Psychological and behavioral problems are frequently associated with bipolar disorders, and alcohol and substance abuse, violence, and suicide can result from inadequate treatment. Psychotherapeutic interventions may be particularly relevant for these problems.

Another major rationale for adjunctive psychotherapy is to improve medication compliance. An estimated 20 to 50 percent of patients with bipolar disorders who are on a prescribed medication regimen either do not comply fully with their doctor's instructions or discontinue treatment altogether. Physical adverse effects and the psychological unwillingness to take pills add to the noncompliance problem. Lithium noncompliance or discontinuation increases relapse. Psychotherapy combined with lithium may increase medication compliance and yield a better clinical outcome.

Adjunctive psychotherapy can also provide important educational benefits. It can help the patient and family members learn to identify early warnings of impending mania so that more-rapid interventions can occur and to identify problems that exacerbate or precipitate episodes.

Little has been written about psychotherapeutic approaches to bipolar disorders since the report of 12 cases by Mabel Blake Cohen in 1954. In recent years, however, several approaches have been developed for psychosocial and psychotherapeutic treatment of bipolar disorders. Unfortunately no empirical research has been published on the efficacy of these approaches, but they are sufficiently important that a description of each is included here. These short-term, outpatient interventions, were inspired by the well-documented success of similar programs used with schizophrenia patients.

Miklowitz and Goldstein The treatment package developed by David J. Miklowitz and Michael J. Goldstein is based on behavioral family-management techniques. Based on the premise that the family attributes considered important in predicting the course of schizophrenia are also associated with the course of bipolar disorders, the program focuses on educating the family about bipolar disorders and aiding in the development of communication and problem-solving skills. This approach (like all psychosocial approaches for bipolar disorders) is not intended to substitute for a traditional medication regimen but to be adjunctive therapy. The program for patients recently discharged after an episode of hypermania includes 21 1-hour sessions conducted in the patient's home over a 9-month period. They include seven sessions dealing with family education, seven on communication-skills training, and seven on problem-solving skills training. In a pilot trial with nine patients, only one patient relapsed over the 9 months during which treatment was implemented. In comparison, a 61 percent relapse rate was reported from a naturalistic outcome study using traditional medication regimens without family management.

Basco and Rush The treatment package developed by Monica R. Basco and A. John Rush is designed around four goals: (1) educating the patient about bipolar disorder; (2) teaching cognitive-behavioral skills for coping with the psychosocial stressors and cognitive and behavioral problems associated with manic and depressive symptoms; (3) facilitating compliance with a prescribed medication regimen; and (4) monitoring the occurrence, severity, and course of manic and depressive symptoms. The protocol is divided into three phases. The first phase, consisting of 1-hour sessions, once a week for 5 weeks, teaches the patient about the causes, systems, and treatment of bipolar disorder. The second phase, which teaches cognitive-behavioral skills, consists of weekly sessions lasting approximately 75 to 90 minutes. The third phase, maintenance, provides an opportunity to monitor the patient's symptoms, reinforce skills, and facilitate medication compliance. This final phase is held in 1-hour sessions no less than once a month and no more than four times a month.

The treatment protocol is highly structured. Each session covers one component of the treatment package and includes (1) a summary of the intention and direction of the session, (2) background information about the intervention technique, (3) goals of the session, (4) a step-by-step description of the intervention procedures, and (5) a homework assignment to reinforce what was learned in the session or prepare for the next session.

Investigators have developed and are currently studying the effectiveness of an adaptation of interpersonal therapy disorder for bipolar patients, called social rhythm interpersonal psychotherapy. Interpersonal interventions are integrated with a structured focus on stabilizing daily activities, including sleep, eating, work, and interpersonal interactions. This treatment is being examined in the acute and maintenance treatment of bipolar disorder patients.

SUGGESTED CROSS-REFERENCES

Information regarding related aspects of mood disorders are discussed further in the other sections of Chapter 14, Chapter 30 on psychotherapies also outlines behavioral and cognitive therapies and other psychosocial treatments. Psychiatric treatments of adolescents are reviewed in Chapter 48, and treatments in the elderly population are included in Section 51.4. Application of psychosocial treatment to schizophrenia may be found in Section 12.9.

SECTION REFERENCES

Basco MR, Rush AJ: Cognitive-behavioral therapy for bipolar disorder. In *Cognitive Behavioral Treatment of Manic Depressive Disorder.* Guilford, New York, 1995.

Beck AT, Rush AJ, Shaw BF, Emery G: *Cognitive Therapy of Depression.* Guilford, New York, 1979.

*Colom F, Vieta E, Martinez A, Jorquera A, Gasto C: What is the role of psychotherapy in the treatment of bipolar disorder? Psychother Psychosom *67:*3, 1998.

Conte HR, Karasu TB: A review of treatment studies of minor depression: 1980–1991. Am J Psychother *46:*58, 1992.

Crits-Christopher P: The efficacy of brief dynamic psychotherapy: A meta-analysis. Am J Psychiatry *149:*151, 1992.

Elkin I, Gibbons RD, Shea MT, Sotsky SM, Watkins JT, Pilkonis PA, Hedeker D: Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol *63:*841, 1995.

Elkin I, Parloff MB, Hadley SW, Autry JH: NIMH Treatment of Depression Collaborative Research Program: Background and research plan. Arch Gen Psychiatry *42:*305, 1985.

*Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB: National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. Arch Gen Psychiatry *46:*971, 1989.

Evans MD, Hollan SD, DeRubeis RJ, Plasecki M, Grove WM, Garvey MJ, Tuason VB: Differential relapse following cognitive therapy and pharmacotherapy for depression. Arch Gen Psychiatry *49:*802, 1992.

Fava GA, Rafanelli C, Grandi S, Conti S, Belluardo P: Prevention of recurrent depression with cognitive behavioral therapy: Preliminary findings. Arch Gen Psychiatry *55:*816, 1998.

Frank E, Hlastala S, Ritenour A, Houck P, Tu XM, Monk TH, Mallinger AG, Kupfer DJ: Inducing lifestyle regularity in recovering bipolar disorder patients: Results form the maintenance therapies in bipolar disorder protocol. Biol Psychiatry *41:*1165, 1997.

Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ: Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry *47:*1093, 1990.

Frank E, Kupfer DJ, Wagner EF, McEachran AB, Cornes C: Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression. Arch Gen Psychiatry *48:*1053, 1991. (Contributing factors [published erratum] Arch Gen Psychiatry *49:*401, 1992.)

Gortner ET, Gollan JK, Dobson KS, Jacobson NS: Cognitive-behavioral treatment for depression: Relapse prevention. J Consult Clin Psychol *66:*377, 1998.

Jarrett RB: Psychosocial aspects of depression and the role of psychotherapy. J Clin Psychiatry *51:*26, 1990.

Kavanagh DJ, Wilson PH: Prediction of outcome with group cognitive therapy for depression. Behav Res Ther *27:*333, 1989.

Keitner GI, Miller IW: Combined psychopharmacological and psychosocial treatment of depression. RI Med J *76:*415, 1993.

*Klerman GL, Weissman MW, Rounsaville BJ, Cheveron ES: *Interpersonal Psychotherapy of Depression.* Basic Books, New York, 1984.

Lewinsohn PM, Antonuccio D, Steinmetz JL, Teri L: *The Coping with Depression Course: A Psychoeducational Intervention for Unipolar Depression.* Castalia, Eugene, OR, 1984.

Markowitz JC, Klerman GL: *Manual for Interpersonal Therapy with Dysthymic Patients.* Payne Whitney Clinic, New York, 1991.

Markowitz JC, Klerman GL, Clougherty KF, Spielman LA, Jacobsberg LB, Fishman B, Frances AJ, Kocsis JH, Perry SW: Individual psychotherapies for depressed HIV-positive patients. Am J Psychiatry *152:*1504, 1995.

McCullough JP: Psychotherapy for dysthymia: A naturalistic study of ten patients. J Nerv Ment Dis *179:*734, 1991.

McCullough JP: *Therapist Manual for Cognitive-Behavior Therapy for the Chronic Depressions.* Virginia Commonwealth University, Richmond, VA, 1995.

Miklowitz D, Goldstein M, Neuchterlein K, Snyder K, Mintz J: Family factors and the course of bipolar affective disorder. Arch Gen Psychiatry *45:*225, 1988.

Moreau D, Mufson L, Weissman MM, Klerman GL: Interpersonal psychotherapy for adolescent depression: Description of modification and preliminary application. J Am Acad Child Adolesc Psychiatry *30:*642, 1991.

*Segal ZV, Gemar M, Williams S: Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. J Abnorm Psychol *108:*3, 1999.

*Shea MT, Elkin I, Hirschfeld RMA: Psychotherapeutic treatment of depression. In *Review of Psychiatry*, vol 7, RE Frances, AJ Hales, editors. American Psychiatric Press, Washington, DC, 1988.

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, Collins JF, Pilkonis PA, Beckham E, Glass DR, Dolan RT, Parloff MB: Course of depressive symptoms over follow-up: Findings from the National Institute of Mental Health treatment of depression collaborative research program. Arch Gen Psychiatry *49:*789, 1992.

Sotsky SM, Glass DR, Shea MT, Pilkonis PA, Collins JF, Elkin I, Watkins JT, Imber SD, Leber WR, Moyer J, Oliveri ME: Patient predictors of response to psychotherapy and pharmacology: Findings in the NIMH Treatment of Depression Collaborative Research Program. Am J Psychiatry *148:*997, 1991.

Spangler DL, Simons AD, Monroe SM, Thase ME: Response to cognitive-behavioral therapy in depression: Effects of pretreatment cognitive dysfunction and life stress. J Consult Clin Psychol *65:*568, 1997.

Thase ME: Long-term treatments of recurrent depressive disorders. J Clin Psychiatry *53*(Suppl):32, 1992.

Thase ME, Dube S, Bowler K, Howland RH, Myers JE, Friedman E, Jarrett DB: Hypothalamic-pituitary-adrenocortical activity and response to cognitive behavior therapy in unmedicated, hospitalized depressed patients. Am J Psychiatry *153:*886, 1996.

Thase MT, Buysse DJ, Frank E, Cherry CR, Cornes CL, Mallinger AG, Kupfer DJ: Which patients will respond to interpersonal psychotherapy? The role of abnormal EEG sleep profiles. Am J Psychiatry *154:*502, 1997.

Watkins JT, Leber WR, Imber SD, Collins JF, Elkin I, Pilkonis PA, Sotsky SM, Shea MT, Slass DR: Temporal course of change of depression. J Consult Clin Psychol *61:*858, 1993.

Weissman MM: The many uses of interpersonal therapy. Harvard Ment Health Lett *14:*4, 1998.

*Weissman MM, Jarrett RB, Rush AJ: Psychotherapy and its relevance to the pharmacotherapy of major depression: A decade later (1976–1985). In *Psychopharmacology: The Third Generation of Progress*, H Meltzer, editor. Raven, New York, 1987.

Weissman MM, Markowitz JC: Interpersonal psychotherapy: Current status. Arch Gen Psychiatry *61:*599, 1994.

CHAPTER 15. ANXIETY DISORDERS

15.1 ANXIETY DISORDERS: INTRODUCTION AND OVERVIEW

JACK M. GORMAN, M.D.

Anxiety disorders are among the most prevalent psychiatric condition in the United States and in most other populations studied. Further, studies have persistently shown that they produce inordinate morbidity, utilization of health care services, and functional impairment. Recent studies also suggest that chronic anxiety disorder may increase the rate of cardiovascular-related mortality. Hence, clinicians in psychiatry and other specialities must make the proper anxiety disorder diagnosis rapidly and initiate treatment.

From a neurobiological perspective, study of anxiety disorders is now seen as compelling. Although validation of animal models of many psychiatric disorders, including depressive disorders, eating disorders, bipolar disorders, and schizophrenia is difficult, constructing animal models of fear and anxiety that appear relevant to human psychopathological states is relatively straightforward. This tool for understanding the neuroanatomy and molecular biology of anxiety promises new insights into etiology and more specific (and thus more effective) treatments in the next decade.

At the same time, the treatments currently available for anxiety disorders are among the most effective in clinical medicine. Pharmacological, cognitive-behavioral, and psychodynamic approaches have all proved useful in combating anxiety disorder. For many conditions (e.g., panic disorder), most patients should expect substantial relief from their symptoms in a relatively brief period.

Another fascinating aspect of anxiety disorders is the exquisite interplay of genetic and experiential factors. While there is little doubt that abnormal genes predispose to pathological anxiety states, evidence clearly indicates that traumatic life events and stress are also etiologically important. Study of the anxiety disorders thus presents a unique opportunity to understand the relation between nature and nurture. This arena should show little regard for purely biological versus psychological explanations about pathogenesis or therapeutics. Rather, anxiety disorder research aims at presenting a view of human function under pathological conditions that integrates multiple sources of theory and information.

The sections that follow review progress in our understanding of panic disorder with and without agoraphobia, social phobia, generalized anxiety disorder, specific phobia, and posttraumatic stress disorder. Each section is intended to provide an up-to-date assessment of the field at the time of its writing.

EPIDEMIOLOGY

Ewald Horwath and Myrna Weissman review epidemiological data and studies on the anxiety disorders. Few published studies have appeared that specifically use the fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV) criteria; hence, most studies of interest involve the older third edition of DSM. Third edition of DSM (DSM-III) and revised third edition (DSM-III-R) criteria sets. This is not expected to make a substantial difference because most changes from DSM-III to DSM-IV have been relatively minor. The data presented demonstrate that anxiety disorders are highly prevalent and that rates of illness are fairly uniform across cultures. In most cases, women are more likely to have anxiety disorder than men, a phenomenon that still begs for adequate explanation. Of particular interest is the finding that social phobia is more common in women than men. Early textbooks gave the opposite view, but that turned out to be a mistake based on overgeneralization of populations seeking treatment rather than community-based studies. While women are about twice as likely to have social phobia as men, men appear more likely to seek treatment. One might speculate that as more women enter the workforce and assume performance roles, they will increasingly find that social phobia symptoms interfere with career success and seek treatment at higher rates.

Horwath and Weissman argue that the data do not substantiate that agoraphobia without panic attacks is nearly as common as an initial glance at available epidemiological studies might suggest. Rather, follow-up studies indicate that panic attacks were missed in such patients or that they actually suffer from other phobic disorders. This remains controversial, and the primacy of the panic attack in leading to agoraphobia remains a subject of investigation (using DSM-IV criteria in the future, one hopes).

CLINICAL FEATURES

The phenomenology of anxiety disorders is reviewed extensively by Daniel Pine. Here, the DSM-IV criteria are used. Some of the differences between DSM-IV and earlier versions are interesting. For example, no longer is a specific number of panic attacks necessary in a specific period of time to meet criteria for panic disorder. Rather, the attacks must be recurrent and at least one attack must be followed by at least 1 month of anticipatory anxiety or phobic avoidance. This recognizes for the first time that although the panic attack is obviously the seminal event for diagnosing panic disorder, the syndrome involves a number of disturbances that go beyond the attack itself. Furthermore, isolated panic attacks without functional disturbance are not uncommon, occurring in approximately 15 percent of the population.

Social phobia has received increasing research and clinical attention in recent years. It is now clear that this condition is not uncommon and that it significantly impairs work and social function. Social phobia is also associated with high rates of substance abuse and, like all anxiety disorders, with depression. One challenge is deciding where to draw the line between social phobia and performance anxiety. Rates of speech phobia may exceed 50 percent in the population, but it is unclear whether such fear and avoidance of public speaking warrants a psychiatric diagnosis and what the appropriate clinical approach should be. On the other hand, generalized social phobia is clearly a serious condition that warrants clinical attention.

Pine notes significant alteration in criteria for posttraumatic stress disorder between DSM-III-R and DSM-IV. The criteria for the earlier version suggest that anyone exposed to particular traumatic events would be expected to develop posttraumatic stress disorder. Furthermore, the qualifying traumas were considered “outside of the range of ordinary human experience.” Two things changed this view. First, clearly fairly ordinary traumatic events can produce posttraumatic stress disorder and there is no adequate definition of “outside the range of ordinary human experience.” Second, only a subset of people exposed to the same traumatic event actually develop posttraumatic stress disorder. The number of previous traumatic events, the severity of the events, and certain precipitating factors such as previous psychiatric illness all seem to increase the likelihood that an individual exposed to severe traumatic stress will actually develop posttraumatic stress disorder. Hence, the DSM-IV criteria make no statement about the commonality of the required traumatic events but stipulate that they must involve a threat of death or physical injury.

What has emerged is a clear picture of separate anxiety syndromes under the rubric of anxiety disorder. These can be reliably diagnosed, but clearly substantial comorbidity exists among them, and each is frequently complicated by depression and substance abuse. Increasingly, investigators are attempting to define common denominators among the mood and anxiety disorders and to discern whether certain childhood conditions specifically predispose to any adult disorders.

GENETICS

Many studies have shown that anxiety disorders tend to run in families. For most medical conditions, this would suffice to conclude that abnormal genes must be etiologically relevant. However, it is not difficult to conceive that growing up with anxious parents or siblings might influence the development of anxiety in any individual. Therefore, family studies are only leads that prompt genetic investigators to attempt to determine whether any anxiety disorders are indeed inherited.

Abby Fyer accomplishes two things in her chapter on genetic aspects of anxiety disorders. First, she instructs the reader about the current techniques used to attempt to disentangle genetic and environmental influences for anxiety disorders. Then she reviews several types of studies that have already been published. What

emerges is a clear picture that genetics has something to do with anxiety disorders but that no anxiety disorder is likely to be the result of a simple Mendelian abnormality. That is, there probably is no specific “gene” for panic disorder or social phobia or generalized anxiety disorder. Fyer's review indicates that what is inheritable is a susceptibility to develop anxiety disorder. It is not even entirely clear at this point whether this susceptibility predisposes to the development of a specific anxiety disorder e.g., panic) or to any one of the anxiety disorders. Further, it is not entirely clear to what extent these genetic influences overlap between susceptibility to anxiety disorder and to depression.

One compelling hypothesis is that individuals inherit a temperament like shyness, hyperactive autonomic nervous system responses, or behavioral inhibition. Then, depending on a variety of life circumstances, these genotypes are expressed as specific phenotypes—one or more of the anxiety disorders themselves. It may also be that more powerful “anxiety genes” require less environmental stress to be expressed.

Modern understanding of molecular genetics makes the above notion entirely plausible. Although the actual genes an individual receives cannot be altered by experience, most genes are only variably expressed. Every cell in the human body contains all of the approximately 100,000 genes in the human genome. A liver cell, for example, has all the same genes as a neuron in the brain's frontal cortex, but the liver cell does not sprout axons and dendrites or generate neural impulses. The genes for these functions are not expressed in the liver cell. Clearly a variety of environmental influences activate latent genes through complex biochemical process, and these operate in the central nervous system. Hence, it is likely that genetic susceptibility to an anxiety disorder becomes an actual anxiety disorder when some set of environmental influences causes anxiety proneness genes to become active.

NEUROBIOLOGY

Attempts are now being made to link the extensive data on the biology of the fear response in animals to observations in humans with anxiety disorder. Two interesting leads are provided from the preclinical literature. First, conditioned fear in animals depends absolutely on normal function of the central nucleus of the amygdala (CNA). This work, pioneered by neurobiologists such as Joseph LeDoux, Trevor Robbins, and Michael Davis, has elegantly shown that afferent projections from the CNA during fear reactions in animals activate autonomic nervous system centers of the brain that reproduce most of the behavioral and physiological manifestations of the acute anxiety response. Second, studies in rodents and nonhuman primates indicate that early separation distress changes function of the hypothalamic-pituitary-adrenal axis that endures throughout the life span of the affected animal.

Gregory Sullivan and Jeremy Coplan review what is now known about the neurobiology of human anxiety. Studies in panic and posttraumatic stress disorders are now extensive. Since the original finding that sodium lactate infusion can induce panic attacks in patients with panic disorder, many substances have shown similar panicogenic properties including the noradrenergic stimulant yohimbine (Yocon), carbon dioxide, the respiratory stimulant doxapram (Dopram), and cholecystokinin. Disordered serotonergic, noradrenergic, and respiratory systems are clearly implicated in panic disorder, and the condition appears to be caused both by a genetic predisposition and traumatic separation distress. More recently, neuroimaging studies revealed that patients with panic disorder have abnormally brisk cerebral vascular responses to stress, showing greater vasoconstriction during hypocapnic respiration than normal controls.

Studies of patients with posttraumatic stress disorder have revealed a downregulated hypothalamic-pituitary-adrenal axis with reduced secretion of cortisol, presumably a result of chronically increased production of corticotropin releasing factor. Studies also suggest that the hippocampus may show neuronal degeneration in patients with posttraumatic stress disorder, probably as a result of increased central glucocorticoid effects. Because the hippocampus is critical for memory function, this could explain any loss of memory for the original traumatic event in these patients.

To date, there is no coherent biology for social phobia. Generalized anxiety disorder has also been inadequately studied, with most interest focused to date on possible abnormalities in the central benzodiazepine receptor and its associated site for binding of the inhibitory neurotransmitter g-aminobutyric acid (GABA). Further research is needed into the biology of these two very common illnesses.

Overall, neurobiological studies, particularly with respect to panic disorder and posttraumatic stress disorder have elucidated a specific set of neurotransmitter and central nervous system abnormalities that can now be addressed with more-specific treatment approaches. For the first time, treatment research can finally be guided to some degree by an understanding of basic pathological mechanisms within the brain.

PSYCHODYNAMIC ASPECTS

Glen Gabbard provides a modern understanding of the psychodynamic aspects of anxiety disorder. One can only admire the insights he derives from Freud's seminal work on the subject, *Inhibitions, Symptoms, and Anxiety*. It is fascinating to recognize that some findings of current neurobiologists substantiate many of Freud's original ideas. For example, the existence of an unconscious memory system for anxiety responses is well substantiated by work demonstrating that the amygdala subserves the fear response without any reference to conscious memory. Gabbard shows the development of Freud's theories through subsequent psychodynamic writers, including Melanie Klein and Heinz Kohut, and demonstrates how psychodynamic concepts of anxiety disorder are indeed fully consonant with observations from neuroscience. For example, he points out that panic attacks may not be entirely “out of the blue” phenomena but are likely to be related to “meaningful stressors” in the patient's life. This view accords with modern ideas about fear conditioning. He also discusses childhood precusors to panic disorder, again finding much in common with preclinical scientists who show that early separation distress in infant animals produces an “anxiety prone” adult animal.

Similarly, in his discussion of phobias, Gabbard draws on the exciting work of Jerome Kagan and colleagues at Harvard University who have identified a state in young children called “behavioral inhibition to the unfamiliar.” This temperamental disposition can apparently be altered by a variety of powerful life experiences. This gives credence to the idea that psychodynamic treatments may be able to alter genetically mediated personality traits.

Although no controlled trials of psychodynamic therapy for anxiety disorders exist, only the rare psychiatrist does not use these techniques even when prescribing medication or using cognitive-behavioral therapy to treat anxiety disorder. When large scale studies are mounted to test the efficacy of psychodynamic treatments for anxiety disorders they are expected to reveal an effective treatment approach.

SOMATIC TREATMENT

Medication to treat anxiety is centuries old if one includes alcohol. As Laszlo Papp details, somatic treatments have gone through many incarnations. Until recently, most antianxiety medications were developed empirically rather than on the basis of known organic pathology. Benzodiazepines remain one of the most often prescribed class of drugs for anxiety, but whether they address any actual brain abnormality or work nonspecifically is not known.

One of the great challenges in prescribing medication to patients with anxiety disorder is their known sensitivity to even relatively minor adverse effects. Because anxious patients generally maintain fear of physical sensations, medication adverse effects are sometimes viewed catastrophically. Hence, as Papp discusses, most treatment should be started at relatively low doses of the selected medication, with slow titration to a fully therapeutic dose. On the other hand, evidence clearly shows that anxiety disorders are, in general, chronic illnesses that often require long-term therapy. Thus the medication selected must be well tolerated and safe even if the patient must continue to take it for a prolonged period to prevent relapse.

Antidepressant medication is increasingly seen as the medication treatment of choice for the anxiety disorders. More specifically, drugs with primary effects on the serotonin neurotransmission system have become first-line recommendations for panic disorder, social phobia, obsessive-compulsive disorder, and posttraumatic stress disorder. Evidence also now exists that such medications are also effective for generalized anxiety disorder. Although they generally take longer to work than benzodiazepines, the selective serotonin reuptake inhibitors such as fluoxetine (Prozac), sertraline (Zoloft), paroxetine (Paxil), fluvoxamine (Luvox), and citalopram (Celexa), as well as venlafaxine (Effexor) and nefazodone (Serzone) are probably more effective than benzodiazepines and easier to discontinue. Increasingly, benzodiazepines are used only for the temporary relief of extreme anxiety as clinician and patient wait for the effects of antidepressants to take hold. Longer-term administration of benzodiazepines is reserved for patients who do not respond to, or cannot tolerate, antidepressants. Monoamine oxidase inhibitors are given only to anxiety disorder patients who do not respond to trials with several medications.

Placebo-controlled trials leave little doubt, as Papp notes, that newer antidepressants are effective for anxiety disorders. Because they work fairly quickly and have fewer adverse effects than tricyclic drugs and monoamine oxidase inhibitors, a low threshold for prescribing them to anxious patients should be maintained. However, most clinicians believe that the best result for anxiety disorder patients comes with combination of medication with one or more types of psychotherapy.

COGNITIVE-BEHAVIORAL APPROACHES

One remarkable recent advance in therapeutics for anxiety disorders is the rigorous testing of cognitive and behavioral psychotherapies. Although many claim that it is difficult to submit psychosocial treatments to experimental investigation, scientists have developed the capacity to apply excellent research designs including randomization and blinded assessment to cognitive behavioral therapy. As Lawrence Welkowitz describes, the result has been documentation that cognitive-behavioral therapy is effective for most anxiety disorders.

A large number of studies have now shown that cognitive-behavioral therapy of various types is effective for panic disorder. The response rate after acute treatment is comparable to that achieved with tricyclic drugs or benzodiazepines. Furthermore, some, but not all, studies indicate that once a patient with panic disorder has completed this therapy, the response may be long-lived. Many different techniques are incorporated into standard cognitive-behavioral packages for panic disorder, including breathing retraining, deconditioning, cognitive restructuring, relaxation, exposure, and psychoeducation. It is not clear which among these is critical for favorable outcome. Studies are also needed to determine if some kind of maintenance, or booster, therapy would help prevent symptomatic relapse.

At least one group has shown that cognitive-behavioral therapy is effective for social phobia, and many studies have documented its benefits for obsessive-compulsive disorder. Studies also report favorable outcome with patients suffering from generalized anxiety disorder or posttraumatic stress disorder. These treatments may be given as first-line approaches to patients who refuse or cannot tolerate medication or in combination with medication. The latter approach may be particularly effective, but empirical justification for it is lacking.

Not only has cognitive-behavioral research provided another effective way to treat anxiety disorders, it has also heartened psychotherapists in general because it proves that one can demonstrate scientifically that psychotherapy works. This should be considered just as exciting a development as anything generated by the neuroscience community. One intriguing challenge for research will be understanding how the psychosocial treatments affect central nervous system processes. Indeed, some imaging studies already suggest that psychotherapy may alter abnormal patterns of brain activation, but much more work is required in this area.

SUGGESTED CROSS-REFERENCES

Chapter 1 covers the neural sciences, and Chapter 2 covers neuropsychiatry and behavioral neurology. The sections provided in this chapter should serve as a guide to research and treatment developments in the anxiety disorder field: Epidemiology of anxiety disorders is discussed in Section 15.2, biochemical aspects of anxiety disorders in Section 15.3, genetics of anxiety disorders in Section 15.4, psychodynamic aspects of anxiety disorders in Section 15.5, diagnosis and clinical features of anxiety disorders in Section 15.6, biological therapies for anxiety disorders in Section 15.7, and psychological therapies for anxiety disorders in Section 15.6. Chapter 46 covers anxiety disorders in children, and Section 51.3c discusses anxiety disorders in the elderly.

SECTION REFERENCES

Barlow DH: *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. Guilford, New York, 1988.

*Brawman-Mintzer O, Lydiard RB: Biological basis of generalized anxiety disorder. J Clin Psychiatry *58*(Suppl):16, 1997.

Fallon BA, Liebowity MR, Compeos R, Schueier FR, Marshall R, Savier S, Goety D, Klein DF: Intravenous clomipramine for obsessive-compulsive disorder refractory to oral clomipramine. Arch Gen Psychiatry *55*:918, 1998.

*Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

*Gorman JG, Shear K: Practice guidelines for the treatment of patients with panic disorder. Am J Psychiatry *155*(Suppl):1, 1998.

*Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: The genetic epidemiology of phobias in women: The interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. Arch Gen Psychiatry *49:*273, 1992.

Smoller JW, Tsuang MT: Panic and phobic anxiety: Defining phenotypes for genetic studies. Am J Psychiatry *155:*1152, 1998.

*Stein MB, Jung KL, Livesley WJ: Heritability of anxiety sensitivity: A twin study. Am J Psychiatry *156:*246, 1999.

*Weissman MM, Markowitz JC, Kleiman GL: *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books, New York, 1999.

Wiborg IM, Dahl AA: Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? Arch Gen Psychiatry *53:*689, 1996.

15.2 ANXIETY DISORDERS: EPIDEMIOLOGY

EWALD HORWATH, M.D., M.SC., AND MYRNA M. WEISSMAN, PH.D.

Anxiety has been recognized as a symptom for centuries. However, only recently, with the incorporation into the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and the revised third edition of DSM (DSM-III-R) of Donald Klein's conceptualization of panic disorder as a separate entity, have anxiety states begun to be subdivided into distinct entities such as panic disorder with and without agoraphobia, social phobia, posttraumatic stress disorder, obsessive-compulsive disorder, and generalized anxiety disorder.

In a review of five population studies conducted in the United States, the United Kingdom, and Sweden prior to the development of specified diagnostic criteria, other workers found that anxiety states were fairly common (about 2.0 to 4.7 per 100, point prevalence) and more prevalent in women than men, particularly women between 16 and 40 years of age. In a separate epidemiological review, the second author identified nine additional community studies of anxiety states that showed rates in the range reported above and also showed that rates were higher in women than in men. This chapter focuses on more recent epidemiological studies in which specific anxiety disorders were diagnosed on the basis of DSM-III or DSM-III-R criteria. Although the criteria for anxiety disorders were modified somewhat in the fourth edition of DSM (DSM-IV), no community studies using DSM-IV criteria are currently available.

The epidemiological approach to the study of anxiety disorders has certain strengths and weaknesses. Epidemiological studies are very informative because they gather data from large numbers of subjects, use powerful statistical techniques, and survey community samples of people who are not in treatment. The study of large numbers of subjects allows comparisons across relevant groups on the basis of differences in sex, race, education, occupation, ethnicity, and other factors. Large numbers also provide the statistical power to use sophisticated analytical strategies such as multivariate regression analysis, which can dissect the effects of complex sociodemographic variables. Community surveys can sample nonclinical populations, which allows investigation of many variables without the confounding factor of treatment seeking, which is strongly influenced by gender, education, and other socioeconomic and cultural factors.

Epidemiological studies also are limited in their capacity to answer certain questions about anxiety disorders. Complex sociodemographic variables such as gender, race, education, ethnicity, language, and religion are measured in epidemiological surveys in an attempt to identify factors that may influence psychiatric symptom formation. But the ability to capture these concepts successfully in fairly simple variables across large community surveys is far from perfect. Further, the diagnostic constructs used in epidemiological surveys may be well established in industrially developed, Western societies, but their reliability and validity may not be tested adequately across other cultures. This review attempts to minimize the problems of diagnosis across different cultures by relying, when possible, on data gathered and analyzed by uniform cross-national procedures.

PANIC DISORDER

Definition The key feature of panic disorder in DSM-III is the occurrence of three or more panic attacks within a 3-week period. These attacks cannot be precipitated only by exposure to a feared situation, cannot be due to a physical disorder, and must be accompanied by at least four of the following symptoms: dyspnea, palpitations, chest pain, smothering or choking, dizziness, feelings of unreality, paresthesias, hot and cold flashes, sweating, faintness, trembling or shaking. In DSM-III-R, the definition was revised to require four attacks in 4 weeks or one or more attacks followed by a persistent fear of having another attack, and the list of potential symptoms was revised to include nausea or abdominal distress and to exclude depersonalization or derealization.

More importantly, DSM-III-R changed the diagnostic hierarchy so that panic disorder could be diagnosed as a primary disorder with or without agoraphobia and dropped the category of agoraphobia with panic attacks. This change placed the emphasis on identifying panic disorder as a discrete entity and reflected the clinical experience that panic attacks tended to occur before development of agoraphobia, which was increasingly viewed as a phobic avoidance response to the frightening experience of spontaneous panic attacks, near-panic experiences, or limited symptom attacks.

DSM-IV criteria require recurrent unexpected panic attacks and persistent concern about having further attacks, worry about the implications of the attacks, or a significant change in behavior because of the attacks. No epidemiological data using these criteria is available.

Prevalence Rates Table 15.2-1 shows prevalence rates for panic disorder from a cross-national collaborative study of 10 countries, using the Diagnostic Interview Schedule (DIS) and DSM-III criteria. These 10 community studies included over 40,000 subjects and were analyzed with appropriate standardization for age and sex differences among subjects from different countries. For comparison purposes, Table 15.2-1 also includes data from the National Comorbidity Study (NCS) of a representative sample of 8098 persons living in the 48 contiguous United States conducted in 1990 to 1992 using the University of Michigan version of the Composite International Diagnostic Interview (UM-CIDI) and DSM-III-R criteria. The annual rate of DSM-III panic disorder ranged from 0.2 per 100 in Taiwan to 2.1 per 100 in Beirut, Lebanon. The NCS reported an annual prevalence of 2.2 per 100 for DSM-III-R panic disorder.

Site	Annual	Lifetime			
		Total	Female	Male	F/M Ratio
United States (ECA)[a]	1.0	1.7	2.3	1.0	2.3
Edmonton	0.9	1.4	1.9	0.9	2.1
Puerto Rico	1.1	1.7	1.8	1.4	1.3
Savigny, France	0.9	2.2	3.0	1.3	2.3
West Germany	1.7	2.6	3.8	1.4	2.7
Florence, Italy	1.3	2.9	3.9	1.2	3.2
Beirut, Lebanon	2.1	2.1	3.1	1.1	2.8
Taiwan	0.2	0.4	0.6	0.2	3.0
Korea	1.5	1.7	2.9	0.5	5.8
New Zealand	1.3	2.1	3.3	0.7	4.7
United States (NCS)[b]	2.2	3.5	5.1	1.9	2.7

[a] Five U.S. sites (New Haven, CT; Baltimore; St Louis, MO; Piedmont County, NC; Los Angeles from the ECA study.
[b] 48 contiguous U.S. states, ages 18–54 from the NCS.
Adapted from Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H-G, Joyce PR, Karam EG, Lee C-K, Lellouch J, Lepine J-P, Newman SC, Oakley-Browne MA, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H-U, Yeh E-K. The cross-national epidemiology of panic disorder. Arch Gen Psychiatry 54:305, 1997.

Table 15.2-1 Prevalence Rates for Panic Disorder per 100 Subjects: Age 18–64 Years

Lifetime rates of DSM-III panic disorder showed excellent agreement, with the prevalence varying from 1.4 per 100 in Edmonton, Canada, to 2.9 per 100 in Florence, Italy. The exception to this narrow range was Taiwan, where DSM-III panic disorder had a lifetime prevalence of 0.4 per 100. The lower rates of panic disorder in Taiwan are consistent with lower Taiwanese rates for most other disorders studied. The reason for these lower rates is not clear. The only study that reported on lifetime DSM-III-R panic disorder was the NCS, which found a rate of 3.5 per 100, somewhat higher than the lifetime rates based upon DSM-III. The higher annual and lifetime rates reported in the NCS may reflect a period effect, with rates increasing between the ECA of the early 1980s and the NCS of the early 1990s. The higher rate in the NCS may also reflect the broader concept of panic disorder in DSM-III-R than in DSM-III or the differences in memory probes used in the NCS interview

(the UM-CIDI) and the DIS used in the cross-national collaboration.

Age at Onset The age at onset of panic disorder is usually in the early to middle 20s, with a later onset in West Germany and Korea (35.5 and 32.1, respectively). The NCS data showed a bimodal distribution of age of onset, with an early mode for panic disorder in the 15- to 24-year age range for both men and women and a later mode in the 45- to 54-year range.

Sex Differences Comparing lifetime prevalence rates, all studies reporting on panic disorder showed higher rates for women than for men. An analysis of the NCS data found uniformly higher rates of panic attacks and panic disorder for women than in men within every age group. Another study analyzed incidence rates from the Epidemiologic Catchment Area (ECA) study and found a twofold higher risk of incident panic disorder in women than in men. This finding is analogous to the higher incidence and prevalence rates for major depression in women than in men and suggests that for both panic disorder and major depression, the higher rates in women reflect a true increase in the risk for new-onset panic disorder and major depression rather than a greater tendency to seek treatment or have longer episodes of illness.

Race and Ethnicity The ECA study found no significant differences in prevalence rates between African-American, Hispanic, and white groups. Similarly, the NCS found no main effects of race or ethnicity but did report an interaction effect of age and race or ethnicity. For young adults in the NCS, differences between racial or ethnic groups were minor, but in the 45 to 54 age range, blacks and Hispanics had substantially lower rates than whites. The cross-national comparisons are more remarkable for the cross-cultural similarities in rates of panic disorder, with the exception of the Taiwan study, which found substantially lower rates of panic. As with major depression, Korean prevalence rates of panic disorder were comparable to those in the West, while Taiwan's were much lower.

Comorbidity At every cross-national site and in the NCS, panic disorder was associated strongly with an increased risk for major depression and agoraphobia. The odds ratios for comorbidity of lifetime panic disorder with agoraphobia ranged from 7.5 in the ECA to 21.4 in Puerto Rico, with the NCS reporting a odds ratio of 10.6 (Table 15.2-2). Based on lifetime rates, the odds ratos for comorbidity of panic disorder with major depression ranged from 3.8 in Savigny, France, to 20.1 in Edmonton, Canada, with an odds ratio of 5.7 reported in the NCS.

	Agoraphobia			Major Depression		
Site	In Persons With Panic (%)	In Persons Without Panic (%)	Odds Ratio (95% CI)	In Persons With Panic (%)	In Persons Without Panic (%)	Odds Ratio (95% CI)
United States (ECA)	33.8	5.5	7.5 (5.3–10.0)	32.8	4.7	8.7 (6.4–11.7)
Edmonton, Alberta	31.8	2.7	15.1 (7.6–29.9)	68.2	9.2	20.1 (10.4–38.8)
Puerto Rico	58.2	6.0	21.4 (9.2–50.2)	38.2	3.7	15.3 (6.4–36.6)
Savigny, France	48.9	6.5	12.7 (6.6–24.8)	48.9	17.3	3.8 (2.1–7.1)
Taiwan	22.5	1.4	17.2 (8.6–37.2)	22.5	1.3	19.6 (9.3–42.1)
Korea	34.7	2.1	17.2 (10.6–27.8)	28.4	2.5	13.2 (8.0–21.6)
New Zealand	29.5	3.4	9.4 (4.2–21.2)	40.8	11.2	4.6 (2.2–9.4)
United States (NCS)	38.8	5.1	10.6 (7.3–15.3)	54.5	16.0	5.7 (4.0–8.1)

Adapted from Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen HU, Yeh EK. The cross-national epidemiology of panic disorder. Arch Gen Psychiatry. 54:305, 1997.

Table 15.2-2 Comorbidity of Panic Disorder With Lifetime Agoraphobia and Major Depression

AGORAPHOBIA

Definition DSM-III agoraphobia is defined as a fear and avoidance of being in places or situations from which escape might be difficult or in which help might not be available in the event of sudden incapacitation. As a result of such fears, the agoraphobic person avoids travel outside the home or requires accompaniment when away from home. Moderate cases may cause some constriction in lifestyle, while severe cases of agoraphobia may result in the person being completely housebound or unable to leave home unaccompanied.

DSM-III-R revised the diagnosis of agoraphobia to a condition accompanying panic disorder (panic disorder with agoraphobia) or paniclike symptoms. Although the diagnosis of agoraphobia without history of panic disorder was retained, this category emphasized the avoidance behavior as a response to the sudden development of anxiety or somatic symptoms. DSM-IV further emphasizes that the agoraphobic avoidance behavior occurs specifically in response to the fear of developing paniclike symptoms.

Prevalence Rates Lifetime prevalence rates of agoraphobia have been reported from a number of studies. If one considers only the studies carried out in primarily English-speaking countries, the lifetime prevalence rates vary over a narrow range, from 2.9 per 100 in Edmonton, Canada, to 6.7 per 100 in the NCS. In spite of the changes in the diagnostic definition between DSM-III and DSM-III-R, the lifetime rates from the ECA and NCS studies showed remarkable consistency (5.6 and 6.7 per 100, respectively). As with other disorders, Taiwan reported a lower rate of agoraphobia.

Risk Factors Lifetime rates of agoraphobia were significantly higher for women than for men in each of the community studies. This is consistent with the sex differences found for panic disorder.

In the ECA study, lifetime prevalence of agoraphobia was higher among African-Americans than among whites or Hispanics. The effects of race or ethnicity and sex or gender combined to produce a considerable range in lifetime prevalence, from 2.9 per 100 white males to 12 per 100 African-American women. In the NCS, current agoraphobia (past month) was associated with higher risk in African-Americans than in whites. Current agoraphobia was inversely related to income and education in a bivariate analysis of the NCS data.

RELATION BETWEEN AGORAPHOBIA AND PANIC ATTACKS

As Table 15.2-2 shows, the Cross-National Collaborative Study and the NCS identified panic disorder as a uniformly powerful risk factor for agoraphobia across a wide array of study sites. Nonetheless, in DSM-III, agoraphobia was considered a separate phobic disorder that might or might not be accompanied by panic attacks. Largely because of the influence of Klein's hypothesis that agoraphobia is a conditioned avoidance response to the aversive stimulus of spontaneous panic attacks, the diagnostic view of agoraphobia changed considerably in DSM-III-R, in which panic disorder is viewed as primary, with or without secondary development of agoraphobia. An important factor in this change was the observation by Klein and others that in clinical settings, agoraphobia rarely occurs without preceding spontaneous panic attacks or limited symptom attacks. However, other investigators have questioned the temporal precedence and causal role of panic attacks in the development of agoraphobia.

Contributing to the controversy are the large differences between clinic and community studies in their estimates of the relative prevalence of agoraphobia with and without panic attacks. Population-based surveys have identified a substantial number of subjects with agoraphobia and no history of panic attacks. For the most part, subjects in these surveys were interviewed by lay persons using the DIS. In contrast, clinic-based studies using less structured interviews administered by clinicians almost invariably found much lower rates of agoraphobia without panic.

Several explanations for this discrepancy have been suggested. One is that samples of treated persons with any illness have higher rates of comorbidity than samples of untreated persons. An alternative explanation is that population studies may have overestimated the rate of agoraphobia without panic disorder.

In a reanalysis of the ECA data on agoraphobia without panic attacks, 22 community cases of agoraphobia without panic attacks were clinically reappraised, and only a single case of probable agoraphobia without panic attacks was found. The diagnostic reappraisal found that 19 (87 percent) of the sample had simple or social phobias rather than agoraphobia or had no DSM-III phobia at all. The reappraisal also identified six cases of panic disorder, panic attacks, or limited symptom attacks that had been missed by the DIS. The authors concluded that community studies using the DIS may have overestimated the prevalence of agoraphobia without panic attacks in the community.

An initial analysis of the NCS data found that only about one third of NCS subjects with agoraphobia reported a history of a panic attack. More-detailed analyses of

the NCS data are under way to determine whether agoraphobia without panic is in fact as common as the initial analyses suggest.

SOCIAL PHOBIA

Definition The central feature of DSM-III social phobia is a persistent, irrational fear accompanied by a compelling desire to avoid situations in which a person might act in a humiliating or embarrassing way while under the scrutiny of others. DSM-III-R allowed the phobic situations to be avoided or endured with intense anxiety and required that the avoidant behavior interfered with occupational or social functioning or that there was marked distress about having the fear. DSM-IV adds that the person recognizes the fear as excessive or unreasonable. Common social phobias involve fears of speaking or eating in public, urinating in public lavatories, writing in front of others, or saying foolish things in social situations.

Prevalence Rates Table 15.2-3 shows the lifetime prevalence of DSM-III social phobia from a cross-national study reporting on the United States, Canada, Puerto Rico, and Korea. For comparison, the lifetime prevalence of DSM-III-R social phobia is also included at the bottom of the table. Lifetime rates of DSM-III social phobia varied somewhat, with a low of 0.5 per 100 in Korea and a high of 2.6 per 100 in the United States. It is not clear whether these contrasting rates reflect true cross-cultural differences or are due to differences in methodology or translation of the DIS.

Site	Males	Females	Total
United States (ECA)	2.1	3.1	2.6
Edmonton, Canada	1.3	2.1	1.7
Puerto Rico	0.8	1.1	1.0
Korea	0.1	1.0	0.5
United States (NCS)	11.1	15.5	13.3

Adapted from Weissman MM, Bland RC, Canino GJ, Greenwald S, Lee C-K, Newman SC, Rubio-Stipec M, Wickramaratne PJ: The cross-national epidemiology of social phobia: A preliminary report. Int Clin Psychopharmacol 11(Suppl):9, 1996.

Table 15.2-3 Lifetime Prevalence Rates of Social Phobia per 100 Subjects

The rate of lifetime DSM-III-R social phobia from the NCS was considerably higher (13.3 per 100) than that in any of the DIS and DSM-III studies. These differences were attributed by some workers to the higher prevalence to differences between the DIS and UM-CIDI. The UM-CIDI uses a stem question based on the broader DSM-III-R criteria allowing either avoidance of a feared situation or endurance with intense anxiety, and it also asks about six specific social phobic fears (versus three in the DIS), including the high-prevalence fears of using a public toilet, writing in front of others, or talking to people and sounding foolish or having nothing to say.

Age at Onset and Comorbidity In the cross-national collaborative data reported in Table 15.2-3, the mean age at onset of first phobia ranged from the midteens to early 20s. The NCS reported a median age at onset for social phobia of 16 years. The vast majority of social phobias occurred in persons who reported at least one other psychiatric disorder. Both the NCS and the cross-national studies reported that persons with social phobia usually suffered a first onset of phobia prior to the onset of other psychiatric disorders.

Risk Factors An analysis of the ECA data from four sites (the New Haven site used a version of the DIS that did not include social phobia items) found that lifetime prevalence rates of social phobia were highest among women and persons who were younger (age 18 to 29 years), less educated, single, and of lower socioeconomic class. In the NCS, higher rates were found in women, those with less education or income, those who never married, students, and those living with their parents. A higher prevalence of lifetime social phobia was also found among women in Korea, Edmonton, and Puerto Rico. The higher rates among women are in contrast to results of several clinical studies, which reported somewhat higher rates of men seeking treatment for social phobia. Men may be more interested in treatment because of impaired work performance. Further study of the gender distribution is required to determine whether the different rates for men and women reflect true differences in prevalence or differences in treatment-seeking patterns.

OBSESSIVE-COMPULSIVE DISORDER

Definition DSM-III obsessive-compulsive disorder required the presence of obsessions or compulsions that are sources of significant distress or impairment and are not due to another mental disorder. DSM-III-R required that the obsessions or compulsions cause marked distress, consume more than 1 hour a day, or significantly interfere with the person's normal routine or occupational or social functioning. DSM-IV added the requirement that the person recognizes that the obsessions or compulsions are excessive or unreasonable. *Obsessions* are defined as recurrent, persistent thoughts, images, or impulses that are experienced as intrusive and inappropriate. *Compulsions* are repetitive behaviors (e.g., checking locked doors or gas jets, handwashing) or mental acts (e.g., counting, repeating words) that the person feels driven to perform in response to an obsession or according to rigid rules.

Prevalence Rates Table 15.2-4 shows lifetime prevalence rates of DSM-III obsessive-compulsive disorder from the Cross National Collaborative Group. Lifetime prevalence of OCD varied from 0.7 per 100 in Taiwan to 2.5 per 100 in Puerto Rico. The studies in English-language sites showed excellent agreement, with lifetime prevalence of 2.2 to 2.3 per 100 in the United States, Canada, and New Zealand. Most remarkably, these rates contradict the previous traditional view of obsessive-compulsive disorder as a rare disorder on the basis of published clinical reports.

Site	Total	Females	Males	F/M Ratio
United States	2.3	2.8	1.7	1.6
Edmonton, Canada	2.3	2.7	2.0	1.3
Puerto Rico	2.5	2.7	2.3	1.2
Munich, Germany	2.1	1.9	2.5	0.8
Taiwan	0.7	0.9	0.5	1.8
Korea	1.9	2.0	1.7	1.2
New Zealand	2.2	3.4	0.9	3.8

Adapted from Weissman MM, Bland RC, Canino GJ, Greenwald S, Hwu H-G, Lee C-K, Newman SC, Oakley-Browne MA, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H-U, Yeh E-K: The cross-national epidemiology of obsessive compulsive disorder. The Cross National Collaborative Group. J Clin Psychiatry 55(Suppl):5, 1994.

Table 15.2-4 Lifetime Prevalence Rates of Obsessive-Compulsive Disorder per 100 Subjects

Age at Onset, Risk Factors, and Comorbidity The mean age at onset of obsessive-compulsive disorder is in the mid-20s to early 30s. The youngest mean age at onset was reported in Edmonton (21.9 years) and the oldest in Puerto Rico (35.5 years).

In the ECA study, prevalence rates of obsessive-compulsive disorder were higher among women than among men. However, when gender comparisons were controlled for marital status, employment status, job status, ethnicity, and age, no differences in prevalence rates for women and men remained. In the Cross National Collaborative Study, rates were generally higher in women than in men, except in Munich, where the rates were higher among men. Persons with obsessive-compulsive disorder had a substantially greater risk of having comorbid major depression or another anxiety disorder than persons without

obsessive-compulsive disorder.

POSTTRAUMATIC STRESS DISORDER

Definition Posttraumatic stress disorder is defined in DSM-III as a constellation of symptoms in response to a stressor, including reexperiencing a traumatic event, numbed responsiveness, and symptoms of an increased level of arousal. In DSM-III-R, the symptoms must persist for at least 1 month, and the criteria are broadened by adding (1) intense psychological distress in response to events that symbolize or resemble an aspect of the trauma and (2) avoidance of stimuli associated with the event. DSM-IV adds the requirement for functional impairment or clinically significant distress.

Prevalence Rates Until recently, no accurate information on the prevalence of posttraumatic stress disorder was available. Earlier studies on the effects of traumatic experiences tended to focus on groups exposed to a common trauma, such as Vietnam war veterans or survivors of natural disasters. In the last decade several surveys have gathered general population data, which have enhanced knowledge of the extent and seriousness of the impact of posttraumatic stress disorder on the community.

Table 15.2-5 shows lifetime prevalence rates of posttraumatic stress disorder from five surveys conducted in the United States using DSM-III or DSM-III-R diagnostic criteria. The DSM-III studies, which were both part of the ECA and used the DIS as the diagnostic instrument, found a low lifetime prevalence rate of 1.0 to 1.3 per 100 subjects. More importantly, these and other studies using DSM-III generated reliable, systematic data on the nature of the response to various traumas, including criminal victimization, sexual assault, exposure to natural disaster, and combat exposure. This empirical information contributed to the revisions of the diagnostic criteria in DSM-III-R. The early studies also led to better understanding of the effects of trauma, and improvements in the assessment of populations for the presence of traumatic life events and the symptoms of posttraumatic stress disorder.

Author (Date)	Females	Males	Total
DSM-III studies			
Helzer (1987)	1.3	0.5	1.0
Davidson (1991)			1.3
DSM-III-R studies			
Kessler (1995)	10.4	5.0	7.8
Resnick (1993)	12.3		
Breslau (1991)	11.3	6.0	9.2

Table 15.2-5 Lifetime Prevalence Rates of Posttraumatic Stress Disorder per 100 Subjects

Later studies using DSM-III-R criteria found a lifetime prevalence of posttraumatic stress disorder ranging from 10.4 to 12.3 per 100 women and 5.0 to 6.0 per 100 men. These later studies seem to confirm that posttraumatic stress disorder is highly prevalent and also provide evidence that the traumatic events that cause it are quite commonly experienced in the community. In the NCS, 61 percent of men and 51 percent of women reported at least one traumatic event. Men were more likely than women to experience physical attacks, combat experience, and being threatened with a weapon, held captive, or kidnapped. Women, on the other hand, were more likely to experience rape, sexual molestation, childhood parental neglect, and childhood physical abuse.

Risk Factors and Comorbidity The epidemiological data suggest that the relation between trauma exposure and development of posttraumatic stress disorder is complex. Men and women differ in the types of traumas to which they are likely to be exposed and in their liability to develop posttraumatic stress disorder once exposed. The lifetime prevalence of posttraumatic stress disorder is significantly higher in women than in men. Women are more likely than men to be exposed to high-impact traumas, or traumas that are associated with a high probability of developing posttraumatic stress disorder. Furthermore, once exposed to traumatic events, a higher proportion of women than men go on to develop the disorder. The vast majority of men and women who met criteria for DSM-III-R posttraumatic stress disorder had a lifetime history of at least one other psychiatric disorder.

GENERALIZED ANXIETY DISORDER

Definition The DSM-III criteria for generalized anxiety disorder require the presence of unrealistic or excessive anxiety and worry, accompanied by symptoms from three of four categories: (1) motor tension, (2) autonomic hyperactivity, (3) vigilance and scanning, and (4) apprehensive expectation. The anxious mood must continue for at least a month, and the diagnosis is not made if phobias, panic disorder, or obsessive-compulsive disorder are present or if the disturbance is due to another physical or mental disorder such as hyperthyroidism, major depression, or schizophrenia. By this definition, generalized anxiety disorder was treated primarily as a residual category after the other major anxiety disorders were excluded. DSM-III-R narrowed the definition further by requiring a minimum of six symptoms and a duration of 6 months. DSM-IV requires only three symptoms from a list of six but adds the requirement that the anxiety causes clinically significant distress or functional impairment.

Prevalence Rates Table 15.2-6 shows the lifetime prevalence rates of generalized anxiety disorder from community studies using DSM-III or DSM-III-R criteria. Lifetime prevalence of generalized anxiety disorder in the ECA study was quite consistent across three study sites, varying from 4.1 per 100 in Los Angeles to 6.6 per 100 in Durham and St. Louis. In spite of differences in diagnostic criteria, the ECA and NCS rates were quite similar. Lifetime prevalence varied considerably more in the Taiwan study, from 3.7 per 100 in Taipei to 10.5 per 100 in small-town areas of Taiwan.

Site	Females	Males	Total
DSM-III studies			
United States (ECA)			
Durham	7.3	5.7	6.6
St. Louis	7.8	5.2	6.6
Los Angeles	5.5	2.6	4.1
Florence, Italy			5.4
Taiwan			
Urban			3.7
Small towns			10.5
Rural			7.8
Korea			3.6
DSM-III-R studies			
United States (NCS)	6.6	3.6	5.1
Florence, Italy			3.9

Table 15.2-6 Lifetime Prevalence Rates of Generalized Anxiety Disorder per 100 Subjects

The Florence study provides an interesting example of the effects of requiring the longer 6-month duration of DSM-III-R. For DSM-III, the lifetime rate was 5.4 per 100, while the narrower DSM-III-R definition resulted in the lower rate of 3.9 per 100.

Risk Factors, Comorbidity, and Impairment The ECA and NCS studies found significantly higher rates of generalized anxiety disorder in women than in men. The NCS also found that being 25 years or older, being previously married, being unemployed, being a homemaker, and living in the Northeast significantly increased the risk of generalized anxiety disorder. The NCS found that 90 percent of subjects with lifetime generalized anxiety disorder had at least one other psychiatric disorder, with affective disorders, panic disorder, and agoraphobia being the most common. The NCS also found substantial evidence for impairment due to generalized anxiety disorders, with 82 percent of those with lifetime generalized anxiety disorders reporting that they sought professional help or took medicine for it or that the

disorder interfered with their life.

ANXIETY DISORDERS AND DISABILITY

Epidemiological studies have shown a strong association between anxiety disorders and functional disability. Recently, the World Health Organization (WHO) Collaborative Study on Psychological Problems in General Health Care examined the relationship between common mental disorders and disability across cultures. This WHO study collected data on psychiatric status and levels of occupational and physical impairment in primary care settings in 15 different sites around the world. Study sites included Ankara, Turkey; Athens, Greece; Bangalore, India; Berlin, Germany9 Groningen, the Netherlands; Ibadan, Nigeria; Mainz, Germany; Manchester, England; Nagasaki, Japan; Paris, France; Rio de Janeiro, Brazil; Santiago, Chile; Seattle, Washington; Shanghai, China; and Verona, Italy. The investigators found panic disorder and generalized anxiety to be strongly associated with functional disability consistently across major cultures around the world. This consistency across cultures is of particular interest given the substantial differences in cultural and socioeconomic environments across included in this study.

FUTURE DIRECTIONS

In spite of methodological limitations, lifetime prevalence rates of panic disorder are remarkably consistent across the community studies and across cultural, racial, and ethnic boundaries, with the exception of the higher rates in the NCS, which may relate to differences in interview method, and much lower rates in Taiwan, where lower rates were reported for several disorders. Cross-nationally, panic disorder is consistently associated with substantial levels of occupational impairment and is more common among women than men. Panic disorder is highly comorbid with agoraphobia and major depression.

In contrast to panic disorder, the epidemiological data on agoraphobia show considerable variation in rates across studies and cross-culturally. A recent clinical reappraisal of the ECA data on agoraphobia without panic suggests that community studies relying upon the DIS and DSM-III may have overestimated the prevalence of agoraphobia without panic. Thus prevalence estimates from such studies should be regarded with caution until their accuracy can be more thoroughly tested.

Analyses of relative risks showed higher rates of agoraphobia for women than for men, just as with panic disorder. Unlike panic disorder, however, agoraphobia was associated with higher rates for African-Americans than whites and higher rates among those with less education or income. The differential effects of race and socioeconomic factors on panic disorder and agoraphobia suggest that the factors that cause panic disorder may differ from those that lead to the subsequent development of agoraphobia.

The ECA and NCS studies found prevalence rates of social phobia to be highest among women, those with less education or income, and those who never married. Posttraumatic stress disorder and generalized anxiety disorder are more prevalent among women than men. Based on community data, obsessive-compulsive disorder turned out to be much more prevalent than previous clinical studies suggested.

In conclusion, community studies have shown anxiety disorders to be highly prevalent and important causes of functional impairment. The data on anxiety disorders are interesting both for their consistency across quite different settings and for some of the questions they raise. Further study is needed to clarify the comorbidity between anxiety disorders, the consistently higher rates of anxiety disorders in women than in men, and the differential effects of socioeconomic and cultural factors on panic disorder and phobias. Further investigation of the complex relation between exposure to traumatic events and the development of posttraumatic stress disorder is needed. Epidemiological data are needed to answer basic questions regarding the etiology and prevention of all anxiety disorders, as well as clinical studies to improve their treatment and prevent the disability they cause.

SUGGESTED CROSS-REFERENCES

Epidemiology is covered in Section 5.1; statistics and experimental design is covered in Section 5.2. Classification of mental disorders is discussed in Chapter 9. Section 15.6 describes the clinical features of anxiety disorders.

SECTION REFERENCES

Angst J, Dobler-Mikola A: The Zurich study: V. Anxiety and phobia in young adults. Eur Arch Psychiatr Neurol Sci *235:*171, 1985.

Anthony JC, Aboraya A: The epidemiology of selected mental disorders in later life. *Handbook of Mental Health and Aging*, ed 2, JE Birren, RB Sloane, GD Cohen, editors. Academic Press, San Diego, 1992.

Argyle N, Roth M: The relationship of panic attacks to anxiety states and depression. In *Abstracts of the World Congress of Biological Psychiatry*, C Shagass, editor. Elsevier Science, New York, 1986.

Aronson TA, Logue CM: On the longitudinal course of panic disorder. Compr Psychiatry *28:*344, 1987.

*Barlow DH: *Anxiety and its Disorders: The Nature and Treatment of Anxiety and Panic*. Guilford, New York, 1988.

Berkson J: Limitations of the application of four fold table analysis to hospital data. Biomet Bull *2:*47, 1946.

P>Bland RC, Newman SC, Orn H: Period prevalence of psychiatric disorders in Edmonton. Acta Psychiatr Scand Suppl *338*:33, 1988.

Blazer DG, Hughes D, George LK, Swartz M, Boyer R: Generalized anxiety disorder. In *Psychiatric Disorders in America. The Epidemiologic Catchment Area Study*, LN Robins, DA Regier, editors. Free Press, New York, 1991.

Breslau N, Davis GC, Andreski P, Peterson E: Traumatic events and post-traumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry *48:*216, 1991.

Canino GJ, Bird HR, Shrout PE, Rubio-Stipec M, Bravo M, Martinez R, Sesman M, Guevara LM: The prevalence of specific psychiatric disorders in Puerto Rico. Arch Gen Psychiatry *44:*727, 1987.

Eaton WW, Kessler RC, Wittchen HU, Magee WJ: Panic and panic disorder in the United States. Am J Psychiatry *151:*413, 1994.

Faravelli C, Degl'Innocenti BG, Giardinelli L: Epidemiology of anxiety disorders in Florence. Acta Psychiatr Scand *79:*308, 1989.

Garvey M, Tuason V: The relationship of panic disorder to agoraphobia. Compr Psychiatry *25:*529, 1984.

*Helzer JE, Robins LN, McEvoy L: Post-traumatic stress disorder in the general population. N Engl J Med *317:*1630, 1987.

Horwath E, Johnson J, Hornig CD: Epidemiology of panic disorder in African-Americans. Am J Psychiatry *150:*465, 1993.

Horwath E, Lish J, Johnson J, Hornig CD, Weissman MM: Agoraphobia without panic: Clinical re-appraisal of an epidemiologic finding. Am J Psychiatry *150:*1496, 1993.

Hwu H-G, Yeh E-K, Chang L-Y: Prevalence of psychiatric disorders in Taiwan defined by the Chinese Diagnostic Interview Schedule. Acta Psychiatr Scand *79:*136, 1989.

Joyce PR, Bushnell JA, Oakley-Brown MA, Wells JE, Hornblow AR: The epidemiology of panic symptomatology and agoraphobic avoidance. Compr Psychiatry *30:*303, 1989.

Karno M, Golding JM, Burnam MA, Hough RL, Escobar JI, Wells KM, Boyer R: Anxiety disorders among Mexican Americans and non-Hispanic whites in Los Angeles. J Nerv Ment Dis *177:*202, 1989.

Karno M, Golding JM, Sorenson SB, Burnam MA: The epidemiology of obsessive-compulsive disorder in 5 U.S. communities. Arch Gen Psychiatry *45:*1094, 1988.

Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Study. Arch Gen Psychiatry *51:*8, 1994.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H-U, Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry *52:*1048, 1995.

*Kessler RC, Zhao S, Katz SJ, Kouzis AC, Frank RG, Edlund M, Leaf P: Past-year use of outpatient services for psychiatric problems in the national comorbidity survey. Am J Psychiatry *156:*115, 1999.

Keyl P, Eaton WW: Risk factors for the onset of panic attacks and panic disorder. Am J Epidemiol *131:*301, 1990.

Lee CK, Kwak YS, Yamamoto J, Rhee H, Kim YS, Han JH, Choi JO, Lee YH: Psychiatric epidemiology in Korea. Part I: Gender and age differences in Seoul. J Nerv Ment Dis *178:*242, 1990.

Lee CK, Kwak YS, Yamamoto J, Rhee H, Kim YS, Han JH, Choi JO, Lee YH: Psychiatric epidemiology in Korea. Part II: Urban and rural differences. J Nerv Ment Dis *178:*247, 1990.

Magee WJ, Eaton WW, Wittchen HU, McGonagle KA, Kessler RC: Agoraphobia, simple phobia and social phobia in the National Comorbidity Survey. Arch Gen Psychiatry *53:*159, 1996.

Marks I, Lader M: Anxiety states (anxiety neurosis): A review. J Nerv Ment Dis *156:*3, 1973.

Ormel J, VonKorff M, Ustun B, Pini S, Korten A, Oldehinkel T: Common mental disorders and disability across cultures. JAMA *272:*1741, 1994.

Pollard CA, Bronson SS, Kenney MR: Prevalence of agoraphobia without panic in clinical settings. Am J Psychiatry *146:*559, 1989.

Resnick HS, Kilpatrick DG, Dansky BS, Saunders BE, Best CL: Prevalence of civilian trauma and post-traumatic stress disorder in a representative national sample of women. J Consult Clin Psychol *6:*984, 1993.

Robins LN, Helzer JE, Croughan J, Ratcliff KS: National Institute of Mental Health Diagnostic Interview Schedule. Arch Gen Psychiatry *38:*381, 1981.

Schneier FR, Johnson J, Hornig CD, Liebowitz MR, Weissman MM: Social phobia: Comorbidity and morbidity in an epidemiological sample. Arch Gen Psychiatry *49:*282, 1992.

Thompson AH, Bland RC, Orn HT: Relationship and chronology of depression, agoraphobia, and panic disorder in the general population. J Nerv Ment Dis *177:*456, 1989.

Thyer BA, Himle J: Temporal relationship between panic attack onset and phobic avoidance in agoraphobia. Behav Res Ther *23:*607, 1985.

Thyer BA, Parrish RT, Curtis GC, Nesse RM, Cameron OG: Ages of onset of DSM-III anxiety disorders. Compr Psychiatry *26:*113, 1985.

Torgersen S: Genetic factors in anxiety disorders. Arch Gen Psychiatry *40:*1085, 1983.

Uhde TW, Boulenger JP, Roy-Byrne PP, Geraci MF, Vittone BJ, Post RM: Longitudinal course of panic disorder. Prog Neuropsychopharmacol Biol Psychiatry *9:*39, 1985.

Weissman MM: The epidemiology of anxiety disorders: Rates, risks, and familial patterns. In *Anxiety and the Anxiety Disorders*, AH Tuma, JD Maser, editors. Erlbaum, Hilldale, NJ, 1985.

*Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu H-G, Joyce PR, Karam EG, Lee C-K, Lellouch J, Lepine J-P, Newman SC, Oakley-Browne MA, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H-U, Yeh E-K: The cross-national epidemiology of panic disorder. Arch Gen Psychiatry *54:*305, 1997.

*Weissman MM, Bland RC, Canino GJ, Greenwald S, Hwu H-G, Lee C-K, Newman SC, Oakley-Browne MA, Rubio-Stipec M, Wells JE, Wickramaratne PJ, Wittchen H-U, Yeh E-K: The cross-national epidemiology of obsessive compulsive disorder. The Cross National Collaborative Group. J Clin Psychiatry *55*(Suppl):5, 1994.

Weissman MM, Bland RC, Canino GJ, Greenwald S, Lee C-K, Newman SC, Rubio-Stipec M, Wickramaratne PJ: The cross-national epidemiology of social phobia: A preliminary report. Int Clin Psychopharmacol *11*(Suppl):9, 1996.

Weissman MM, Leaf PJ, Tischler GL, Blazer DG, Karno M, Bruce ML, Florio LP: Affective disorders in five United States communities. Psychol Med *18:*141, 1988.

*Wittchen H-U, Zhao S, Kessler RC, Eaton WW: DSM-III-R generalized anxiety disorder in the National Comorbidity Survey. Arch Gen Psychiatry *51:*355, 1994.

15.3 ANXIETY DISORDERS: BIOCHEMICAL ASPECTS

GREGORY M. SULLIVAN, M.D. AND JEREMY DAVID COPLAN, M.D.

Anxiety disorders research is undergoing a revolutionary paradigmatic transformation as disparate areas of psychiatric nosology, pharmacologic dissection, and cognitive neuroscience converge toward an integrated neurobiological understanding of the pathologies. The study of fear conditioning in rodents has identified neuroanatomical pathways that (assuming a phylogenetic conservation of function) are putative anatomical substrates for aspects of human fear and anxiety. The limited access to pathophysiological changes in the human brain in anxiety disorders that has historically plagued psychiatry is being removed by the guidance of such preclinical findings combined with the new window offered by rapidly advancing neuroimaging technologies. Highly complex functional neuroanatomical pathways are being proposed for specific aspects of anxiety disorders, and a research boom is aimed at elucidating the true correlates. In keeping with this paradigm shift, the biochemical aspects of anxiety disorders are reviewed in the light of these hypothetical pathways. The anxiety disorders are separated as per the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) although it is assumed that many aspects of the functional-neuroanatomical pathways cross diagnostic boundaries.

AMYGDALA, CORTICOTROPIN-RELEASING FACTOR, AND MONOAMINES

An abundance of preclinical evidence now points to the amygdala as the major mediator of the stress response, fear, and possibly also anxiety. The amygdala receives excitatory glutamatergic thalamic and cortical sensory input that it appears to process in series and in parallel through an organized array of intraamygdaloid circuitries nuclei. Such a pathway allows a stimulus representation to be modulated by different functional systems such as those mediating memories from past experience and information about the hormonal and homeostatic milieu of the organism. The amygdala also has caudal projections to monoaminergic loci such as noradrenergic neurons of the locus ceruleus, dopaminergic neurons of the ventral tegmental area, and serotonergic neurons of the raphe nuclei. The amygdaloid nuclei appear to have the appropriate connections to orchestrate the simultaneous elaboration of individual anxiety disorder correlates such as cognitive misappraisal and fear (cortical areas), escape and freezing behavior (periaqueductal gray), hyperventilation (parabrachial nucleus), sympathetic activation (lateral hypothalamus), endocrine activation (paraventricular nucleus of the hypothalamus), gastrointestinal distress (dorsal motor nucleus of the vagus), increased startle (nucleus caudalis pontis), and motor activation (striatum).

Lesions of the amygdala in animal models lead to attenuated fear and emotional responsiveness as well as blockade of autonomic and neuroendocrine responses to learned stress. Direct electrical stimulation of the amygdala in animal models elicits both fear-like behavior and autonomic arousal. Stereotaxic electrical stimulation of the amygdala in conscious humans has been reported to elicit symptoms of anxiety including fear, anxiety, depersonalization, visceral sensations in the chest and epigastric region, and changes in autonomic function.

Also emerging is the recognition of a wide extrahypothalamic network of corticotropin-releasing factor (CRF)-secreting neurons originating in the central amygdaloid nuclei. These neurons project to regions of neuroendocrine and autonomic function, to monoaminergic nuclei that project back to the forebrain, and to regions that modify or initiate specific components of the response to threat such as the periaqueductal gray. Intracerebroventricular injection of CRF in animals produces physiological changes similar to those produced by stimulation of the amygdala and a range of behaviors analogous to components of human anxiety and affective disorders; these include anorexia, insomnia, decreased libido, hyperactivity, reduced exploratory activity, and increased startle response. Transgenic mice with enhanced CRF expression are described as expressing a constitutive, persistent phenotype of emotionality including prominent anxiety-like behavior. Indeed, CRF is now hypothesized to facilitate both the cognitive and physical symptoms of anxiety and panic by enhancing bidirectional communication through an activated central nucleus of the amygdala. These are just a few examples of the immense amount of preclinical data implicating a fear system coordinated through the amygdaloid nuclei and using CRF as an effector neuropeptide.

One model relevant to anxiety and panic puts forth that the serotonin pathway originating in the dorsal raphe nucleus and running along the medial forebrain bundle innervates the amygdala and frontal cortex, thus facilitating active escape or avoidance behaviors in response to distal threat. This pathway likely involves postsynaptic serotonin (5-hydroxytryptamine [5-HT]) type 2A/2C (5-$HT_{2A/2C}$) and 5-HT_3 receptor activation. It is assumed that such resultant behavior relies on learning and relates to conditioned or anticipatory anxiety. Such a pathway therefore may be relevant to generalized anxiety disorder. A separate pathway originating in the dorsal raphe nucleus involves innervation of the periventricular and periaqueductal gray region such that serotonin neurons inhibit inborn fight or flight reactions in response to proximal danger, acute pain, or asphyxia. This pathway may be relevant to panic attacks and is likely to be mediated by 5-$HT_{2A/2C}$ and 5-HT_{1A} postsynaptic receptors. A further suggestion is that with chronic stress, the serotonin pathway connecting the median raphe nucleus to the hippocampus, likely mediated by postsynaptic 5-HT_{1A} receptors, promotes resistance to such stress by disconnecting the aversive events from processes underlying appetitive and social behaviors. It has been suggested that this pathway may be relevant to avoidance and numbing found in posttraumatic stress disorder.

The locus ceruleus noradrenergic and dopaminergic systems are believed to increase autonomic arousal and vigilance in response to threat. Prefrontal dopamine release is associated with immediate aversive encounter. The bed nucleus of the stria terminalis, known to have dense noradrenergic innervation, has been implicated in conditioned fear to contextual cues and inhibitory avoidance. Noradrenergic neurons of the locus ceruleus appear to be excitatory to serotonin neurons of the dorsal raphe nucleus, while 5-HT neurons of the dorsal raphe nucleus inhibit locus ceruleus firing. The locus ceruleus may also stimulate the periaqueductal gray via an indirect pathway through the amygdala. So again, the amygdala may mediate in a complex orchestration of coping behaviors and adaptive responses to future stressors. Overstimulation of such adaptive mechanisms by repeated or chronic stressors as well as deficits in function of components of these systems could theoretically lead to the pathological responses observed in anxiety disorders.

PANIC DISORDER

Panic attacks, the hallmark of panic disorder, are unexpected and rapidly progressing bursts of anxiety accompanied by an array of cognitive and autonomic symptoms such as immense fear, palpitations, hyperventilation, lightheadedness, and sweating. By definition, the intensity of symptoms should reach a peak within 10 minutes, and most attacks have a relatively short time course of about 5 to 30 minutes. According to DSM-IV, the sequelae of such attacks are necessary for a diagnosis of panic disorder. These can include anticipatory anxiety (“persistent concern about having additional attacks”), worry about the mental and physical implications of the attacks, and phobic avoidance of situations associated with the attacks (or “a significant change in behavior related to the attacks”). To date, no consistent biological differences have been detected between panic disorder patients with and without agoraphobia, therefore these diagnostic categories are discussed together.

Panicogenic and Anxiogenic Challenge Studies Sodium lactate has been used as a panicogenic agent in biological anxiety research since 1967, when it was first infused in patients with “anxiety neurosis.” Sodium lactate provokes acute panic attacks in about two thirds of panic disorder patients and not in healthy controls. Various hypothetical mechanisms of sodium lactate panic induction have been proposed. After the discovery of similar panic induction by inhalation of 5 percent carbon dioxide (CO_2), it was hypothesized that both lactate infusion, via peripheral catabolism to CO_2, and CO_2 inhalation precipitate panic attacks by increasing the partial pressure of CO_2 (pCO_2) in the central nervous system, activating the medullary chemoreceptors. Yet intravenous lactate in nonhuman primates did not change the pCO_2 of the cisternal fluid during or after lactate infusion. Further, the putatively metabolically inactive enantiomer of L-lactate, D-lactate, also is panicogenic at similar rates in panic disorder patients. This raises the question of a recognition site for unmetabolized lactate that is not specific to the metabolically active isomer. A simpler explanation is they both cause an alkalosis which activates the panic pathway. More recently, proton magnetic resonance spectroscopy of humans has shown increased uptake of lactate in the temporoparietal region during intravenous infusion of lactate in normal controls; although signal due to increased lactate in cerebral

blood potentially confounds determination of blood-brain barrier penetration. A recent preclinical study identified the organum vasculosum lamina terminalis, which is not protected by the blood-brain barrier, as a lactate-sensitive site that leads to anxiety-like autonomic activation when g-aminobutyrate (GABA) neurotransmission is inhibited in the dorsomedial hypothalamus. The investigators propose a mechanism for the panicogenesis of lactate that involves direct stimulation of this circumventricular organ by lactate, resulting in activation of a compromised panic-generating site such as the dorsomedial hypothalamus or amygdala.

The body's normal homeostatic response to a metabolic alkalosis is hypoventilation, as a compensatory measure, resulting in lower blood pH. Yet panic disorder patients who panic during lactate infusion paradoxically hyperventilate, suggesting respiratory stimulation that cannot be overridden by the metabolic alkalosis. It has been hypothesized that panic disorder patients have a hypersensitive suffocation alarm system whereby increasing pCO_2 and brain lactate concentrations prematurely activates a physiological asphyxia monitor. The question has also been raised whether hyperventilation leads to local relative vasoconstriction in the brains of panic disorder patients.

A variety of agents have been successfully used to induce paniclike reactions in efforts to elucidate the underlying pathophysiology. A distinction has been proposed between pharmacological challenge agents that induce significant respiratory symptoms and those that have few, if any, respiratory symptoms but significantly activate the hypothalamic-pituitary-adrenal (HPA) axis and result in significant anxiety. The former group of “respiratory panicogens” includes sodium lactate, CO_2, sodium bicarbonate, isoproterenol (Isuprel), and doxapram (Dopram); these agents induce a state more closely resembling spontaneous panic attacks. The latter group of “hypothalamic-pituitary-adrenal”–activating anxiogens” includes yohimbine (Yocon), *m*-chlorophenylpiperazine (mCPP), fenfluramine, and b-carboline. These agents are more notable for inducing a more generalized or anticipatory anxiety. Cholecystokinin (CCK) agonists appear to induce both a prominent respiratory component and robust hypothalamic-pituitary-adrenal activation and thus have qualities of both classes of agents. It has been suggested that over the course of panic disorder, the pathophysiology evolves from one associated with spontaneous panic attacks to one associated with anticipatory anxiety, paralleling the different responses seen with respiratory panicogens versus the hypothalamic-pituitary adrenal–activating anxiogens.

Noradrenergic System A large body of evidence has accumulated implicating noradrenergic system dysregulation in panic disorder. Apparent abnormalities of presynaptic a_2-adrenergic receptors have been identified by pharmacological challenges with the a_2-receptor agonist clonidine (Catapres) and the a_2-receptor antagonist yohimbine. Possibly the most replicable finding from clonidine challenge studies in panic disorder patients is a blunted growth hormone (GH) response compared with controls. It has also been reported, though less consistently, that clonidine administration in panic patients results in less sedation, a greater hypotensive response, and a disproportionate reduction in the principal noradrenergic metabolite, 3-methoxy-4-hydroxyphenylglycol (MHPG). Some have interpreted the blunted GH response to indicate postsynaptic a_2-receptor subsensitivity in panic disorder.

Yohimbine, which stimulates firing of the locus ceruleus, elicits high rates of paniclike anxiety in panic disorder patients accompanied by increased serum MHPG concentration. It has also been proposed that panic disorder patients have increased sensitivity to challenge with isoproterenol because of increased peripheral b-receptor sensitivity. Particular adrenergic agonists and the carotid chemoreceptor stimulator doxapram may in part act via peripheral receptors that relay information via vagal and glossopharyngeal afferents to the nucleus tractus solitarius. The nucleus traclus solitarius is known to have connections with the amygdala, hypothalamus, parabrachial nucleus, and indirect connections with the locus ceruleus. Interestingly, atrial natriuretic factor (ANF) enhances norepinephrine uptake in the circumventricular organs, locus ceruleus, and nucleus tractus solitarius of the rat. Panic disorder patients who have panic attacks in response to lactate infusion have more-pronounced surges in ANF than nonpanicking controls. Aside from the known antagonism of ANF on CRF-stimulated secretion of ACTH, this peptide might possibly attenuate overactivation of panicogenic loci by increasing presynaptic uptake of norepinephrine. Finally, the successful treatment of panic disorder with imipramine (Tofranil) is attributed to noradrenergic modulation.

Serotonergic System An important role for serotonergic dysregulation is also evident in panic disorder, particularly when considering the clinical efficacy of the selective serotonin reuptake inhibitors (SSRIs) and clomipramine (Anafranil), a tricyclic drug with a potent serotonin reuptake component. Challenge studies with the mixed serotonin agonist-antagonist mCPP and the indirect serotonin agonist fenfluramine have demonstrated increased rates of anxiety, though not necessarily overt panic attacks, to these agents, particularly in panic disorder patients. Such exaggerated responses to serotonergic challenges have been hypothesized to be due to postsynaptic serotonin hypersensitivity in panic disorder. Preclinical studies implicate serotonergic neurons originating in the raphe nuclei in modulation of an evolved complex response to threat. Such serotonin pathways appear activating to more evolutionarily sophisticated neuroanatomical centers involved in processing information relevant to threat. They appear restraining to more primitive brainstem centers involved in escape behavior.

GABAergic System High-potency benzodiazepines such as alprazolam (Xanax) and clonazepam (Klonopin) are clearly efficacious in blocking panic attacks. Reduced sensitivity of saccadic eye movements in panic disorder, whose velocity is partially mediated by GABA-sensitive neurons in the midbrain/pons, has been suggested to indicate GABA dysfunction. Preclinical studies indicate that CO_2 alters the tone of GABA receptors, though immediate relevance to panic disorder is lacking. There is also preclinical evidence that attenuation of local inhibitory GABAergic transmission in the basolateral amygdala, midbrain central gray, or dorsomedial hypothalamus can elicit anxiety-like physiological and behavioral responses.

CRF and the Hypothalamic-Pituitary-Adrenal Axis Clonidine induces a greater decrease in cortisol in panic disorder patients than in controls. In a healthy state a close functional relation obtains between the locus ceruleus noradrenergic system and the hypothalamic-pituitary adrenal axis. These two systems are apparently uncoupled in panic disorder, as correlations observed in controls between MHPG and cortisol, at baseline or maximal change to clonidine, are significantly reduced in panic disorder. Analysis of a decade of sodium lactate studies at one institution indicated the strongest predictors of panic during the baseline period before lactate infusion were fear, high cortisol concentration, and low pCO_2. Such predictors are consistent with a hypothetical CRF-mediated amygdaloid system effecting changes in cortical areas (cognitive misappraisal), the paraventricular nucleus of the hypothalamus (hypothalamic-pituitary adrenal axis), and the parabrachial nucleus (hyperventilation), respectively. The development of panic attacks in pharmacological challenge paradigms then may involve an abrupt switch from an anxiogenic pathway to a panicogenic pathway, differing in activation of specific neuroanatomical loci such as the paraventricular nucleus and periaqueductal gray.

GENERALIZED ANXIETY DISORDER

According to DSM-IV, generalized anxiety disorder is characterized by excessive and uncontrollable anxiety or worry persisting for at least 6 months, combined with three of six additional symptoms (“restlessness or feeling keyed up or on edge,” “being easily fatigued,” “difficulty concentrating or mind going blank,” “irritability,” “muscle tension,” and “sleep disturbance”). Recent epidemiological studies indicate that generalized anxiety disorder is one of the most common anxiety disorders; despite such prevalence, a paucity of research on its neurobiology exists.

Panicogenic and Anxiogenic Challenge Studies Both sodium lactate infusion and CO_2 inhalation have been investigated in generalized anxiety disorder. Those studies generally demonstrated that generalized anxiety disorder patients panic at a much lower rate in response to challenge than panic disorder patients but often have higher rates of panic, subjective anxiety, or somatic symptoms than normal controls. On the other hand one small study using the CCK analogue pentagastrin demonstrated panic attacks at a higher rate (5 of 7) in generalized anxiety disorder patients than in controls (1/7). The implications of this are unclear, as CCK agonists have diverse properties such as increasing firing of adrenergic neurons in the locus ceruleus, activation of the hypothalamic-pituitary adrenal axis and cross-regulation with the GABA system, and their anxiogenic-like effects in rodents are blocked by the 5-HT_3 antagonist ondansetron (Zofran).

Autonomic, Noradrenergic, and Neuroendocrine Systems Normal stress reactions are typically accompanied by increased autonomic activity and stress-related hormone production. These include increased heart rate, blood pressure, minute ventilation, muscle tension, skin conductance, and plasma concentrations of epinephrine, norepinephrine, cortisol, GH, and prolactin. Oddly, generalized anxiety disorder patients appear not to differ from controls with any consistency in baseline blood pressure, heart rate, respiration, or skin conductance. More remarkably, particular studies show evidence of hyporeactivity to stress challenge. Women with generalized anxiety disorder have attenuated skin conductance response to stress and slower recovery to baseline, leading some investigators to suggest decreased “autonomic flexibility” in generalized anxiety disorder. Further, after early positive results, it now appears that baseline plasma catecholamine concentrations in patients with generalized anxiety disorder do not differ from those in normals. Challenge with the a_2-adrenergic agonist clonidine did demonstrate a blunted GH response in generalized anxiety disorder (suggesting possible a_2-adrenoceptor subsensitivity) but indicated no difference in blood pressure, heart rate or MHPG concentrations in generalized anxiety disorder patients and controls after clonidine. The a_2-receptor antagonist yohimbine failed to demonstrate differences from controls in subjective anxiety level, blood pressure, heart rate, plasma MHPG, or cortisol. Fewer a_2-adrenergic binding sites exist on platelets of generalized anxiety disorder patients, raising the possibility of a prior hypercatecholaminergic state downregulating postsynaptic a_2-adrenoceptors. Though baseline cortisol levels appear normal in generalized anxiety disorder, challenge with dexamethasone (Decadron), which normally suppresses the hypothalamic-pituitary adrenal axis, results in nonsuppression in up to a third of patients with generalized anxiety disorder. Thus patients with generalized anxiety disorder appear to have autonomic hyporesponsiveness, subtle changes in adrenergic receptor sensitivity, abnormal regulation of GH secretion, and altered glucocorticoid feedback inhibition.

Serotonergic System The anxiolytic effects of 5-HT_{1A} partial agonists (buspirone [BuSpar] and ipsapirone) and 5-HT_{2} antagonists (ritanserin and serazepine) in generalized anxiety disorder patients has suggested serotonin involvement. A challenge study with the mixed serotonin agonist-antagonist mCPP, which has greatest agonist activity for 5-HT_{2C} and 5-HT_{3} receptors and less for 5-$HT_{1A/1D}$, demonstrated greater rates of anxiety and anger in generalized anxiety disorder patients than in controls. These pharmacological dissection results in generalized anxiety disorder patients are consistent with the model of excessive limbic forebrain 5-HT_{2} and 5-HT_{3} activity leading to anxiety by overstimulation and hippocampal 5-HT_{1A} stimulation leading to anxiolysis by adaptation. A contrasting hypothesis suggests that 5-HT_{1A} agonists are efficacious due to stimulation of presynaptic 5-HT_{1A} autoreceptors, thus reducing the elevated serotonin activity responsible for activating the hypothalamus, basal ganglia, and limbic system.

GABAergic System Certainly benzodiazepines have been a mainstay of treatment for generalized anxiety disorder. Further, platelet and lymphocyte benzodiazepine receptors have decreased binding in generalized anxiety disorder patients and increased binding after treatment with benzodiazepines. Scant clinical evidence is available for central GABA receptor complex abnormality, although saccadic eye movement velocity (thought to be partially mediated by GABA-responsive neurons of the midbrain and pons) may be reduced in generalized activity disorder as it is in panic disorder.

SOCIAL PHOBIA

According to DSM-IV, social phobia is characterized by fear of social or performance situations involving exposure “to unfamiliar people or possible scrutiny by others,” combined with fear of acting in a way that “will be humiliating or embarrassing.” The exposure also almost invariably provokes an anxiety reaction, which may have prominent cognitive and autonomic components similar to a situationally bound or predisposed panic attack. The situations are “avoided or else are endured with intense anxiety or distress,” and the fears are recognized as “excessive or unreasonable.”

Challenge Studies Public speaking has been used in a naturalistic challenge model of social phobia. Although no differences in blood pressure, heart rate, or catecholamines between social phobia patients and controls have been found, greater increase in heart rate was observed in patients with public speaking phobia than in those with generalized social phobia. Norepinephrine concentrations increase in normal subjects during the first several minutes of public speaking and decline to levels not significantly different from baseline by 15 minutes. Indirect evidence for abnormal plasma norepinephrine level regulation in social phobia is suggested by orthostatic studies in which higher supine and standing norepinephrine levels were demonstrated than those in panic disorder patients and controls. Increased blood pressure response to Valsalva's maneuver and exaggerated vagal withdrawal in response to isometric exercise has also been found in social phobia. A greater pressor effect of thyrotropin-releasing hormone (TRH [Thyrel TRH]) also differentiated social phobia patients from panic disorder patients and controls. Limited data suggest that yohimbine increases social anxiety in social phobia and is associated with elevated plasma MHPG concentrations. Clonidine challenge has elicited both blunted and normal GH response, depending on route of administration. Investigation of lymphocyte b-receptors showed no difference between social phobia patients and normal controls. Taken together, these preliminary results vaguely suggest abnormality in the noradrenergic system, though more in-depth investigation is necessary.

Results from lactate infusion and CO_2 inhalation studies are limited by small numbers of subjects but indicate low rates of panic compared with patients with panic disorder. Caffeine challenge did differentiate patients with social phobia and panic disorder from normals, but the symptoms reported by subjects with social phobia differed from their naturally occurring symptoms, thus calling into question the specificity of such challenge in social phobia. Pentagastrin, a selective cholecystokinin-B receptor agonist, is another agent that induces panic attacks in social phobia at a high rate, more similar to what is found in panic disorder than in normal controls. Thus, limited data from panicogen studies in social phobia suggest a possible shared neurobiology with panic disorder only with respect to hypothalamic-pituitary adrenal-activating anxiogens not with the pure “respiratory panicogens.”

Investigation of the hypothalamic-pituitary adrenal axis in social phobia by measurement of urinary free cortisol and the dexamethasone suppression test has not revealed differences from controls. The TRH stimulation test also failed to reveal a difference in the change in thyroid-stimulating hormone (TSH) from controls.

Dopaminergic System The favorable clinical effect of the monoamine oxidase inhibitors in the treatment of social phobia and their superiority to tricyclic drugs has led some to suggest a role for dopamine in social phobia. One study of panic disorder patients with social phobia reported significantly lower homovanillic acid concentration than in normal controls, although diagnostic specificity is certainly a question in this group with comorbid disorders. Prolactin and eyeblink response to levodopa (Larodopa) in social phobia patients did not differ from those of controls in one study. Single photon emission computed tomography (SPECT) study with [$_{123}$I]b-CIT, a labeled cocaine analogue, demonstrated decreased striatal dopamine reuptake site density in social phobia patients. Magnetic resonance spectroscopy studies of social phobia patients have suggested decreased cellular energy activity and neuronal activity and abnormal membrane function in regions that include the basal ganglia. Also, magnetic resonance imaging demonstrated greater reduction in the size of the putamen with aging in social phobia. Thus some evidence (albeit limited and indirect) suggests dopaminergic dysfunction in social phobia.

Serotonergic System Fenfluramine challenge in social phobia patients indicated that the prolactin response is not different from that of controls, but social phobia patients did show a significantly greater rise in cortisol concentration. Platelet tritiated paroxetine (Paxil) binding appears the same in social phobia patients and controls. Serotonergic therapies have shown therapeutic success in studies with social phobia patients, but further study of the serotonergic system is needed before any definitive conclusions can be made.

SPECIFIC PHOBIA

According to DSM-IV, specific phobia is characterized by a “marked and persistent fear that is excessive or unreasonable” and is brought on “by the presence or anticipation of a specific object or situation (e.g., flying, heights, animals, receiving an injection, seeing blood).” The response may take the form of a situationally bound or predisposed panic attack, and the phobia causes marked distress or interferes with role functioning. Little biological research has been done on specific phobia, and studies that have been carried out have not found striking differences from controls. A prominent vasovagal response was observed in a subgroup of patients with specific phobia.

POSTTRAUMATIC STRESS DISORDER

DSM-IV defines posttraumatic stress disorder (PTSD) as a disorder in which a person has been exposed to a traumatic event or events that included “actual or threatened death or serious injury, or threat to the physical integrity of self or others,” and “the person's response involved intense fear, helplessness or horror.” The symptomatic sequelae of the event or events include more than 1 month of persistent reexperiencing in thoughts, images, and dreams; behaving or feeling the event is recurring; and intense psychological or physiological reactivity to cues that are reminders of the event. There is also avoidance of stimuli associated with the trauma and numbing of general responsivity. Symptoms of increased arousal are also present and can include sleep disturbance, irritability, poor concentration, and exaggerated startle reflex.

Panicogenic Challenge Though panicogenic challenge studies using “respiratory panicogens” are very limited, one study found sodium lactate–induced flashbacks accompanied by panic attacks in combat veterans with comorbid panic disorder and PTSD. A more recent study indicated that lactate induces flashbacks in patients without comorbid panic disorder. This study also demonstrated lower cortisol and higher epinephrine concentrations at end of infusion than those in panic disorder patients and controls.

Noradrenergic System PTSD has a long and rich history. Soldiers with PTSD-like syndromes were formerly diagnosed with “shell-shock” or “irritable heart.” Pharmacological challenge with epinephrine was carried out in 1919 on traumatized soldiers with “irritable heart.” Unlike soldiers without this condition, these soldiers manifested nervousness with signs of autonomic arousal such as increased blood pressure and heart rate, palpitations, sweating, flushing, and tremors. More recently, stressful trauma-related sound and visual stimuli were used to challenge patients with PTSD. Plasma epinephrine and norepinephrine concentrations rose with temporal relation to subjective psychophysiological arousal elicited by this material. Studies of baseline heart rate and blood pressure indicated elevations in combat veterans, though results were inconsistent. Studies have found increased 24-hour urine epinephrine concentrations in veterans with PTSD, and increased urine catecholamine concentrations in sexually abused girls. Further, platelet a $_{2}$- and lymphocyte b$_{2}$-adrenoceptors are downregulated in PTSD, possibly in response to chronically elevated catecholamine concentrations.

Challenge with agonist yohimbine has elicited panic attacks at rates of 42 to 70 percent, similar to rates observed in panic disorder. These induced panic attacks were

accompanied by increased heart rate, blood pressure, and plasma MHPG. More remarkably, about 30 to 40 percent of PTSD patients report flashbacks after yohimbine administration, and yohimbine can elicit increases in intrusive traumatic thoughts, hypervigalence, startle amplitude, and emotional numbing. Positron emission tomography (fluorodeoxyglucose) measurement of brain metabolism following yohimbine administration demonstrated decreased activity in prefrontal, orbitofrontal, temporal, and parietal regions in PTSD patients who experienced increased anxiety, and increased activity in controls who were without anxiogenic response. Such findings are strong evidence for altered function in the noradrenergic system in PTSD and may indicate neurobiological overlap between PTSD and panic disorder.

Serotonergic System The SSRIs have shown partial efficacy in treating an array of symptoms in at least a subgroup of PTSD patients. As well, platelet paroxetine binding is reduced in PTSD. More recently, challenge with mCPP, in a comparison study with yohimbine administration, demonstrated a panic rate of 31 percent compared with 42 percent with yohimbine. Twenty-seven percent of the patients also had flashbacks with mCPP. An intriguing observation in this study was that 81 percent of the PTSD patients had a panic attack to yohimbine or mCPP but not both, raising the possibility of subtypes of PTSD with either selective-serotonergic or selective-noradrenergic dysfunction.

Opioid System Abnormality in the opioid system is suggested by reduced plasma beta-endorphin concentrations in PTSD. Because hypothalamic b-endorphin is at least partially mediated by CRF, these reduced concentrations may occur via a mechanism similar to that of lowered adrenocorticotropin (ACTH) release in response to CRF challenge. Though plasma concentrations of metenkephalin were similar to those of controls, the in vitro degradation half-life was reduced, suggesting decreased secretion. Although patients with chronic PTSD have a reduced pain threshold, combat veterans with PTSD also demonstrate a naloxone-reversible analgesic response to combat-related stimuli, raising the possibility of opioid system hyperregulation similar to that postulated in the hypothalamic-pituitary adrenal axis. Currently no studies of PTSD are investigating a-melanocyte-stimulating hormone, a hypothalamic peptide known preclinically to cause hyperalgesia and increased arousal and which appears to antagonize many neuroendocrine effects of beta-endorphin.

Dopaminergic System Increased concentrations of dopamine plasma and urine have been reported in PTSD. One study of subjects in addiction treatment who had been exposed to severe combat conditions demonstrated a significantly higher rate of the D2A1 receptor allele in the subjects who met criteria for PTSD. Otherwise, data on the dopamine system in PTSD are scant.

CRF and the Hypothalamic-Pituitary-Adrenal Axis Several lines of evidence point to dysfunction of the hypothalamic-pituitary adrenal axis and possibly extrahypothalamic CRF in PTSD. Studies have demonstrated low plasma and urinary free cortisol concentrations in PTSD. Consistent with low circulating cortisol, PTSD patients have more glucocorticoid receptors on lymphocytes. Challenge with exogenous CRF in PTSD has shown a blunted ACTH response but a normal cortisol response compared with those of controls. Further, suppression of cortisol by challenge with low-dose dexamethasone (Decadron) is enhanced in PTSD. These latter two challenge paradigms indicate hyperregulation of the hypothalamic-pituitary adrenal axis in PTSD. Dexamethasone suppression is also evident in PTSD patients with comorbid depression, which suggests that such hyperregulation overrides the expected nonsuppression usually associated with depression. A negative correlation between urinary cortisol concentration and novelty seeking was found in combat veterans.

Abundant preclinical evidence implicates hippocampal glucocorticoid receptors as a site for negative-feedback regulation of CRF and arginine vasopressin (AVP) synthesis. Also, prolonged high concentrations of glucocorticoids secreted by the adrenals during stress can have neurotoxic effects on hippocampal neurons. Magnetic resonance imaging study of combat veterans with PTSD demonstrated smaller right hippocampal volume than in matched controls.

Cerebrospinal fluid (CSF) concentrations of CRF are higher in PTSD patients than in controls, consistent with a hypothesis of hypersecretion of neuronal CRF in PTSD. Such elevated CSF CRF has also been identified in depression, but in depression cortisol concentrations are high.

Hypothalamic-Pituitary-Thyroid Axis Although little research has investigated the hypothalamic-pituitary-thyroid axis in PTSD, there have been positive findings. Mean total triiodothyronine (T_3) levels are significantly higher in veterans with PTSD than in normal controls. A high rate of blunted TSH response to TRH stimulation was reported in PTSD, and this abnormality did not appear to respond to 4 weeks of desipramine (Norpramin) treatment. Positive correlations have been found between novelty seeking and serum T_3, free T_3, and the T_3 to free thyroxine (FT_4) ratio in Vietnam veterans with PTSD. Positive correlations between free T_3, total T_3, total T_4, and hyperarousal symptoms, measured by the Clinician-Administered PTSD Scale, have also been reported.

OBSESSIVE-COMPULSIVE DISORDER

According to DSM-IV, obsessive-compulsive disorder (OCD) is characterized by repetitive thoughts, impulses, or images that are intrusive and inappropriate and cause anxiety or distress, or repetitive behaviors that the person feels driven to perform in response to an obsession or rigid rules that must be applied. These persons also recognize that the obsessions are a product of their own mind. The obsessions or compulsions are time consuming or interfere with role functioning.

OCD is unique among the anxiety disorders by appearing to be much more dominated by cognitive and related complex behavioral symptomatology, with autonomic dysregulation playing little role. This view is strengthened when considering rare panic and uncommon increased anxiety in OCD patients challenged with typical panicogenic and anxiogenic agents such as lactate, CO_2, yohimbine, and caffeine. Several investigators have questioned the grouping of OCD with the other anxiety disorders; in many ways it has more in common with obsessive-compulsive spectrum disorders such as Tourette's disorder. This spectrum of pathology appears more associated with corticostriatal pathways than with the amygdalofugal pathways that are the current focus of much anxiety disorder research. Neuroimaging in OCD has produced converging data implicating altered function in the neurocircuitry between the orbitofrontal cortex, caudate, and thalamus.

Serotonergic System The neurobiology of OCD is remarkable for consistently robust findings of abnormalities in the serotonergic system. Part of the focus on the serotonergic system is due to the exclusive response of OCD symptoms to agents with predominant serotonin reuptake blockade such as the SSRIs and clomipramine. Also, the high pretreatment CSF concentration of the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA) and correlation of 5-HIAA decrease with clinical improvement during clomipramine treatment focused attention on the serotonergic system. Other studies investigating CSF 5-HIAA concentrations in OCD have produced inconsistent results. Platelets of OCD patients have reduced tritiated-imipramine binding-site density, indicating abnormality in the serotonin transporter, though these findings have not been consistently replicated.

Depletion of the serotonin precursor tryptophan in OCD patients who had responded to SSRIs or clomipramine did not increase obsessions or compulsions, but mean depression ratings significantly increased. Challenge with the serotonin precursor L-tryptophan demonstrated a small yet significant increase in prolactin in OCD patients, compared with controls. Prolactin response to fenfluramine (a nonspecific serotonin agonist) was inconsistent. Challenge with the partial serotonin agonist mCPP, with complex activity involving several serotonin receptor subtypes, was first observed not only to increase anxiety and depression but also to exacerbate acutely OCD symptomatology in about half of patients. Blunted responses of prolactin, cortisol, and ACTH have been found inconsistently during mCPP challenge. Despite inconsistent results in both behavioral and neuroendocrine response to mCPP challenge, the transient exacerbation of OCD symptoms appropriately drew attention to particular aspects of the serotonin system. Prechallenge blockade of 5-HT_{1A}, 5-HT_{1D}, and 5-HT_{2C} receptors with the nonselective serotonin receptor antagonist metergoline appears to attenuate the behavioral and neuroendocrine response to mCPP. Separate challenge study with MK-212, an agonist with 5-HT_{1A} and 5-HT_{2C} affinity, did not alter behavior but did produce a blunted cortisol response. Also, treatment with clomipramine reduced the behavioral response to mCPP, though abnormalities in prolactin and cortisol responses were not significantly different from those of untreated OCD patients. Both blunted prolactin rise to mCPP and blunted cortisol response to fenfluramine were specifically identified in the female subjects of two studies, suggesting a sexual dimorphism in the neuroendocrine component of OCD.

Noradrenergic System Currently less evidence exists for dysfunction in the noradrenergic system in OCD. Still, there are reports of improvement in OCD symptoms with oral clonidine. One small study of intravenous clonidine also demonstrated a dramatic decrease in the severity of symptoms; however, clonidine may have nonspecific sedating effects. Clonidine also may not be a specific probe of the noradrenergic system, as it binds to rodent imidazole receptors and FONT face=symbol>a_2-heteroreceptors located on serotonergic neurons. GH response to clonidine challenge is blunted while remaining intact with GH-releasing hormone challenge, suggesting possible subsensitivity of a_2-adrenergic receptors.

Neuroendocrine Systems Several abnormalities in baseline and challenge response have been found with hypothalamic-pituitary hormones. Increased CSF oxytocin concentrations were found in a subgroup of OCD patients without a personal or family history of tic disorder, and the oxytocin concentration correlated with the current severity of OCD. Concentration of the GH negative-regulator, somatostatin, is higher in the CSF of OCD patients, which may explain the blunted GH response to clonidine. Recently CSF CRF concentrations have come under investigation. Though mostly negative, one study did find that male OCD patients have higher CSF CRF concentrations than panic disorder or generalized anxiety disorder patients and normal controls. Also, cytokine interleukin-6 (IL-6) concentration, known to

increase in the peripheral blood in response to CRF stimulation, correlates with the severity of compulsive (but not obsessive) symptoms in OCD. Baseline elevations in plasma cortisol have also been found in OCD patients as a group, and one small study showed increased circadian secretion of cortisol that was not altered after 8 weeks of symptom-remitting fluoxetine (Prozac) therapy.

Neuroimmunology Interest has developed in a possible link between streptococcal infection and OCD. Group A b-hemolytic streptococcal infection can lead to rheumatic fever about 2 to 3 percent of the time, and approximately 10 to 30 percent of patients with rheumatic fever develop Sydenham's chorea. It has been reported that over 70 percent of patients with Sydenham's chorea manifest obsessive-compulsive symptoms. A monoclonal antibody to B lymphocytes, D8/17, developed using mice immunized with B lymphocytes from a patient with rheumatic carditis, has been used to measure B-cell D8/17 antigen in a small sample of patients with childhood onset of OCD or Tourette's syndrome. All 31 patients, 29 percent of whom had pure OCD, showed an elevated percentage of D8/17 antigen, contrasted with only 1 of 21 controls. Thus this antigen may serve as a marker for vulnerability to some forms of childhood OCD. Further, infectious or immune-mediated factors may play a role in at least a subgroup of OCD patients.

SUGGESTED CROSS-REFERENCES

Neurotransmitters are discussed in Section 1.4 and Section 1.5, and the contributions of the neural sciences in general are the focus of the other sections of Chapter 1. The biological therapies are discussed in Chapter 31.

SECTION REFERENCES

*Biber B, Alkin T: Panic disorder subtypes: Differential response to CO_2 challenge. Am J Psychiatry *156:*739, 1999.

*Brawman-Mintzer O, Lydiard RB: Biological basis of generalized anxiety disorder. J Clin Psychiatry *58*(Suppl):16, 1997.

Charney DS, Grillon C, Bremner JD: The neurobiological basis of anxiety and fear: Circuits, mechanisms, and neurochemical interactions (Part I). Neuroscientist *4:*35, 1998.

Charney DS, Grillon C, Bremner JD: The neurobiological basis of anxiety and fear: Circuits, mechanisms, and neurochemical interactions (Part II). Neuroscientist *4:*122, 1998.

Coplan JD, Klein DF: Pharmacologic probes in panic disorder. In *Advances in the Neurobiology of Anxiety Disorders*, HGM Westenberg, JA Den Boer, DL Murphy, editors. Wiley, New York 1996.

*Coplan JD, Pine DS, Papp LA, Gorman JM: A view on noradrenergic, hypothalamic-pituitary-adrenal axis and extrahypothalamic corticotrophin-releasing factor function in anxiety and affective disorders: The reduced growth hormone response to clonidine. Psychopharmacol Bull *33:*193, 1997.

*Dolberg OT, Iancu I, Sasson Y, Zohar J: The pathogenesis and treatment of obsessive-compulsive disorder. Clin Neuropharmacol *19:*129, 1996.

Goddard AW, Charney DS. Toward an integrated neurobiology of panic disorder. J Clin Psychiatry *58*(Suppl):4, 1997.

Graeff FG, Guimaraes FS, DeAndrade TGCS, Deakin JFW: Role of 5-HT in stress, anxiety, and depression. Pharmacol Biochem Behav *54:*129, 1996.

Gray TS: Amygdaloid CRF pathways: Role in autonomic, neuroendocrine, and behavioral responses to stress. Ann NY Acad Sci *697:*53, 1993.

*Grillon C, Southwick SM, Charney DS: The psychobiological basis of posttraumatic stress disorder. Mol Psychiatry *1:*278, 1996.

Halgren E: Emotional neurophysiology of the amygdala within the context of human cognition. In *The Amygdala: Neurobiological Aspects of Emotion, Memory, and Mental Dysfunction*, JP Aggleton, editor. Wiley-Liss, New York, 1992.

Handley SL: 5-Hydroxytryptamine pathways in anxiety and its treatment. Pharmacol Ther *66:*103, 1995.

Johnson MR, Lydiard RB: The neurobiology of anxiety disorders. Psychiatr Clin North Am *18:*681, 1995.

Katz L, Fleisher W, Kjernisted K, Milanese P: A review of the psychobiology and pharmacotherapy of posttraumatic stress disorder. Can J Psychiatry *41:*233, 1996.

Klein DF: False suffocation alarms, spontaneous panics, and related conditions: An integrative hypothesis. Arch Gen Psychiatry *50:*306, 1993.

Krystal JH, Deutsch DN, Charney DS: The biological basis of panic disorder. J Clin Psychiatry *57*(Suppl):23, 1996.

LeDonx JE: *The Emotional Brain*. Simon & Schuster, New York, 1996.

Miner CM, Davidson JR: Biological characterization of social phobia. Eur Arch Psychiatry Clin Neurosci *244:*304, 1995.

Papp LA, Coplan J, Gorman JM: Anxiety disorders. In *Review of Psychiatry*, vol 13, JM Oldham, MB Riba, editors. American Psychiatric Press, Washington, DC, 1994.

*Potts NL, Book S, Davidson JR: The neurobiology of social phobia. Int Clin Psychopharmacol *11*(Suppl):43, 1996.

Rauch SL, Jenike MA: Neurobiological models of obsessive-compulsive disorder. Psychosomatics *34:*20, 1993.

Shekhar A, Keim SR: The circumventricular organs form a potential neural pathway for lactate sensitivity: Implications for panic disorder. J Neurosci *17:*9726, 1997.

Southwick SM, Krystal JH, Bremner JD, Morgan CA III, Nicolaou AL, Nagy LM, Johnson DR, Heninger GR, Charney DS: Noradrenergic and serotonergic function in posttraumatic stress disorder. Arch Gen Psychiatry *54:*749, 1997.

Tancer ME: Neurobiology of social phobia. J Clin Psychiatry *54*(Suppl):26, 1993.

Trivedi MH: Functional neuroanatomy of obsessive-compulsive disorder. J Clin Psychiatry *57*(Suppl):26, 1996.

Yehuda R: Sensitization of the hypothalamic-pituitary-adrenal axis in posttraumatic stress disorder. Ann NY Acad Sci *821:*57, 1997.

15.4 ANXIETY DISORDERS: GENETICS

ABBY J. FYER, M.D.

What is currently known about the contribution of genetic factors to the development of anxiety disorders is based on data of six types of studies. Twin, adoption, and deoxyribonucleic acid (DNA) marker (linkage or association) studies can provide definite evidence of a genetic contribution. In the case of twin studies genetic influence is demonstrated if monozygotic (MZ) twins (who share 100 percent of their genetic material) are more often alike (concordant) with respect to a trait than are dizygotic (DZ) twins (who, like sibs, share only 50 percent of their DNA). Adoption studies indicate a genetic contribution if adopted offspring are more similar to their biological parents than to the adoptive parents who are raising them. DNA marker studies, not yet successful in psychiatry, offer the possibility of using molecular genetic methods to identify chromosomal loci or actual genes associated with clinical syndromes.

Important corroborative evidence of genetic influence is provided by family and segregation analytic studies as well as animal models. Family studies can indicate intergenerational transmission and are helpful identification of heritable phenotypes; however, they do not distinguish between genetic and environmental influences. Segregation analysis can demonstrate that familial prevalence patterns are consistent with a particular mode of genetic transmission (e.g., major dominant locus, polygenic). Although this does not prove that genes are responsible, identifying the correct genetic model may be critical to the success of linkage studies. Animal models of fear and anxiety may provide more neurophysiologically specific hypotheses concerning the origin of pathology. Many functional gene families are conserved across species, and this in turn may provide clues about genetic mechanisms.

The classification of anxiety disorders has become increasingly specific since the late 1970s. In 1968 the second edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-II) included three anxiety disorders (anxiety, phobic and obsessional neuroses) while the current fourth edition of DSM (DSM-IV) includes six: panic disorder with or without agoraphobia, generalized anxiety disorder, simple and social phobias, obsessive-compulsive disorder, and posttraumatic stress disorder.

Heritability refers to the proportion of the liability for a disorder that is estimated to be due to genetic factors. Disorders (or traits) in which genetic influences dominate (e.g., Huntington's disease, cystic fibrosis) have high heritabilities >.6). Moderate heritability (.3 to .5) suggests that the studied phenotype (1) results from a more even interaction between genetic and nongenetic influences; or (2) contains several genetically different syndromes, some with high and others with low or no heritability. Examples of this latter situation are common in other areas of medicine in which actual genetic mechanisms have been identified (e.g., diabetes, breast cancer, Alzheimer's disease).

PANIC DISORDER AND AGORAPHOBIA

Panic disorder is the best studied anxiety disorder with respect to intergenerational transmission. Accumulated data strongly suggest a genetic contribution to its etiology. However, heritability estimates are variable (from .3 to .6 percent) and the mode of transmission is unknown. Although several DNA marker studies are now under way, no positive results have been reported.

Twin Studies Four twin studies of panic disorder have been reported. Three indicate greater concordance for the diagnosis for MZ twin pairs than for DZ twin pairs, supporting a genetic contribution to the disorder. The fourth study indicates a genetic role in the predisposition to anxiety symptoms but not to any of the specific revised third edition of DSM (DSM-III-R) anxiety disorders.

When 1033 twin pairs from the Virginia Twin Registry were studied, diagnosis of panic disorder was made in 5.8 percent of the 2163 interviewed twins. Assessments were done using a semi-structured clinician-administered interview. Probandwise concordance for DSM-III-R panic disorder was 24 percent in MZ twin pairs and 11 percent in DZ twin pairs. The highest heritability (46 percent) was found using a narrow (strict adherence to the DSM-III-R criteria) definition of panic disorder. The heritability (and by extention the estimate of strength of genetic contribution) was lower when the criteria were broadened to include cases that did not meet full criteria (e.g., recurrent panic attacks with only three rather than four associated symptoms).

Contrasting results were reported in a smaller sample of twins drawn from the Norwegian Twin registry (32 MZ and 53 DZ). Concordance for strict third edition of DSM (DSM-III) panic disorder was 1 of 5 in MZ and 0 of 6 in DZ pairs. However, if the phenotype was broadened to include an anxiety disorder that included recurrent unexpected panic the concordance pattern (4 of 13 versus 0 of 16;31 percent vs 0 percent $P < 0.05$) indicated a strong genetic contribution. A study using a larger but overlapping sample reported similar results; concordance for panic disorder in MZ versus DZ pairs was 42 percent versus 17 percent.

A subset of Australian twin registry was studied using clinical interviews. In 34 twin pairs at least one twin had panic disorder. No difference in concordance for panic disorder was found in MZ versus DZ pairs.

Family Studies Six direct-interview family studies of panic disorder have been reported—all indicate that panic disorder is highly familial. Rates of panic disorder in the relatives of panic disorder patients are 4 to 10 times as high as those in the relatives of never–mentally-ill or community controls. With the exceptions of social phobia and alcohol abuse, excess risk for disorders other than panic was not seen.

Several studies have also used the family study method to investigate possible heterogeneity within the DSM-IV category. This approach is based on the principle that if two syndromes are associated with different patterns of familial aggregation, they may differ with respect to some aspect of their etiology. For example, one investigation divided panic disorder patients into those with early- (at or before 20 years old) versus; later- (after age 20) onset panic disorder. Significantly higher rates of panic disorder were found in relatives of the early-onset patients as compared to the relatives of later-onset patients or controls. A second study using the family history method found that panic disorder patients who reported at least one relative affected with panic disorder also had an earlier onset of panic disorder than did patients who had no affected relatives. These data are of interest in that they suggest that early-onset panic disorder may be a more severe or etiologically different form of the illness; however, further investigations are needed.

Another interesting area of current research concerns response to carbon dioxide. A number of studies indicate that individuals with panic disorder are significantly more likely than well individuals to become anxious or panic during exposure to carbon dioxide (5 to 35 percent). Two recent studies have extended this work by studying never–mentally-ill relatives of panic disorder patients, patients with depression, and never–mentally-ill controls. Mentally healthy relatives of panic patients have significantly higher rates of panic during carbon dioxide exposure than either of the other groups. This finding suggests that carbon dioxide sensitivity may be a familially transmitted marker of an inherited susceptibility to panic disorder. Family and follow-up studies to clarify these findings are now under way.

Some individuals with panic disorder also develop agoraphobia while some do not and the reasons for this variation are not known. It is possible that panic disorder with agoraphobia is a more severe form of panic disorder. Alternatively, the development of agoraphobia may be related to separate inherited or environmental factors; or some mixture of these. Two studies have addressed these questions. Rates of panic disorder with and without agoraphobia in relatives of three groups of individuals patients with panic disorder and no agoraphobia, patients with panic disorder and agoraphobia, and nonanxious controls were compared; results indicated that relatives of the agoraphobia patients had higher rates of both panic disorder (7.0 versus 3.5) and agoraphobia (14.9 versus 3.5) than the relatives of the controls, although the difference reached the .05 level of significance only in the latter group. In contrast, the relatives of the panic disorder patients had significantly higher rates of panic disorder (14.9 versus 3.5, $P < .005$) but not agoraphobia (1.7 versus 3.5, not significant) as compared to relatives of controls. These data may be interpreted as indicating that with respect to intergenerational transmission agoraphobia (with panic attacks) is either a more severe form of panic disorder or possibly

a different but partially overlapping illness. A twin study tested the fit of their data with several genetic models in which panic disorder with avoidance was hypothesized to be a more genetically severe form of panic disorder without avoidance and good fit was found in three out of four cases; heritability estimates ranged from .34 to .40.

Segregation Analyses Several segregation analytic studies of panic disorder have been reported. Results are consistent with genetic transmission but inconclusive as to the genetic model. Support was found for a single major locus dominant model. Another study found equal support for the single major dominant or recessive models in two independent samples; however, in none of these cases could polygenic transmission be excluded.

DNA Marker Linkage and Association Studies of Panic Disorder Although there is significant evidence for a genetic contribution to panic disorder, at this point there is no data indicating association between a specific chromosomal location or mode of transmission and this disorder. Two large genomic screening studies are in progress and both have reported only negative findings. One study has reported exclusion of linkage between the DSM-III-R panic disorder phenotype for over 95 percent and 60 percent of sites in the human genome under the dominant and recessive models of transmission respectively. Other researcher have excluded linkage between panic disorder and several candidate genes including: five adrenergic receptors, a-haptoglobin, and several components of the g-aminobutyric acid (GABA) receptor system.

Accumulated twin and family study data indicate that panic disorder does not follow a Mendelian mode of transmission and is therefore considered a complex disorder in terms of genetic studies. Experience with complex disorders in other areas of medicine indicates that multiple genes may be involved in determination of a single aspect of clinical pathology. Moreover, both pleiotropy (different clinical syndromes associated with the same genetic underpinnings) and genetic heterogeneity (different genes can lead to the same clinical phenotype) are ubiquitous. In this setting it is not surprising that the current effort to identify a gene for panic disorder has not yet been successful. Genetic heterogeneity can greatly impair the power of linkage analyses by introducing inadvertent false positives. Similarly, pleiotropy can lead to inadvertent exclusion of genetically affected individuals and consequent reduction in power to identify linkage. Current efforts to resolve these difficulties are directed at: (1) using epidemiological biological, and family study data to identify phenotypes that may be closer to the gene; (2) enlarging samples to provide sufficient power to identify smaller gene effects; and (3) improvement of molecular genetic technologies.

GENERALIZED ANXIETY DISORDER

In 1980 DSM-III subdivided anxiety neurosis into panic disorder and generalized anxiety disorder depending on the presence or absence of recurrent spontaneous panic attacks. Although several earlier studies suggested a genetic contribution to anxiety neurosis, their relevance to the residual category of generalized anxiety disorder is unclear. This section focuses on work using the DSM-III and subsequent criteria for generalized anxiety disorder. Several twin and family studies indicate a genetic role in the etiology of this syndrome. However, the extent of the relationship between the disorder and depression as well as the symptom and duration boundaries of generalized anxiety disorder itself require further study. Adoption and DNA marker studies of this disorder have not been reported.

Twin Studies Six twin studies have examined heritability of DSM-III or DSM-III-R generalized anxiety disorder. Although results are mixed, taken as a whole these data suggest a moderate genetic contribution to this disorder. Three of the six studies used epidemiological twin samples and bivariate genetic modeling to investigate both the genetic contribution to generalized anxiety disorder and its relation to major depressive disorder. One study evaluated subjects by direct semistructured interview, the other two by questionnaire. Similar results were found across the three samples. Generalized anxiety disorder was found to be moderately heritable (heritability ~.4 to .5). However, according to this model the genetic factors that make an individual vulnerable to this disorder completely overlap with those that predispose an individual to major depressive disorder. Whether an individual with this genetic loading develops generalized anxiety disorder, major depressive disorder, or both depends on individual-specific environmental factors. No significant effect was found for common family environment.

The remaining three twin studies of generalized anxiety disorder used direct interview procedures in smaller samples derived from a combination of population and hospital registries. Genetic contribution was examined by contrasting probandwise concordance for generalized anxiety disorder in MZ versus DZ twin pairs. One study found no difference in rates of concordance for the disorder in monozygotic and dizygotic pairs (0/12 versus 1/20). Major depressive disorder did not aggregate in the cotwins of generalized anxiety disorder probands. The two remaining studies found somewhat higher concordance rates in the MZ pairs. However, these differences did not reach the .05 level of significance. The relationship between generalized anxiety, disorder and major depressive disorder was not systematically examined. One hundred and forty-four twin pairs from a geographical subsample of the Australian Twin Registry in which the index twin met DSM-III criteria for generalized anxiety disorder were interviewed. Concordance for the disorder was 21.5 percent (13/63) in the MZ versus 13 percent (11/81) in the DZ pairs (NS). A sample of 49 twin pairs, which included 12 index twins with generalized anxiety disorder, were found to have MZ versus DZ concordance rates of 3/5 versus 1/7 (NS).

Family Studies Two family studies of generalized anxiety disorder have been reported. In both, psychiatric disorder in relatives of probands with the disorder is compared to that in the relatives of not-ill controls as well as patients with panic disorder or major depressive disorder or both. One hundred and twenty-three relatives of 20 both generalized anxiety disorder probands with similarly evaluated relatives of 20 control, 40 panic disorder, and 40 agoraphobic probands (N = 113, 241, and 236 respectively) were compared. Significantly higher rates of generalized anxiety disorder were found in the relatives of the probands compared to those of the other groups (19.5 percent versus 3.5 percent versus 5.4 percent versus 3.9 percent; $P < .001$) The two-way contrast between the relatives of the subjects with generalized anxiety disorder versus control patients was significant at $P < .001$ (chi squared = 14.37). There were no significant differences between rates of other anxiety and mood disorders (7.3 percent versus 7.1 percent) in the relatives of the anxiety disorder patients versus relatives of controls. The latter argues against the overlap in heritability of generalized anxiety disorder and major depressive disorder reported by the twin studies cited previously. Panic disorder was significantly elevated in relatives of panic patients versus controls. The absence of an excess of panic disorder in the relatives of the generalized anxiety disorder patients versus controls (4.5 versus 3.5 percent) supports the DSM-III separation of these two syndromes.

The second family study compared relatives of four proband groups: generalized anxiety disorder, major depressive disorder, panic disorder, and controls. Rates of the anxiety disorder were higher in relatives of the probands but did not differ significantly across groups (8.9 versus 4.9 versus 4 versus 1.9 percent). There was no evidence of a shared familial diathesis between generalized anxiety disorder and either panic disorder or major depressive disorder.

SOCIAL PHOBIA

Data on genetic contributions to social anxiety are drawn from three sources: (1) twin and family studies of DSM-III-R social phobia; (2) twin studies of irrational social fears in adults; and (3) studies of childhood temperament.

Twin and Family Studies of Social Phobia Of 2163 twins (1033 pairs) drawn from the Virginia Twin Registry, 11.5 percent met criteria for social phobia. Probandwise concordance for social phobia was 24 percent in monozygotic versus 15 percent in dizygotic pairs. The best-fitting model suggested that genetic and environmental factors contribute respectively 30 and 70 percent of the liability for social phobia. Two thirds of the genetic liability is made up of factors specific to social phobia and the remaining one third comes from genetic liability that contributes to all phobias (e.g., phobia proneness).

One direct-interview family study and one family history study of social phobia have been reported. Each found moderate, specific familial aggregation of this disorder. One study compared rates of psychiatric disorders in the directly interviewed relatives of social phobia (N = 105), specific phobia (N = 49), agoraphobia with panic (N = 164), and never-ill controls (N = 231). Rates of social phobia but not of other anxiety disorders were significantly elevated in the relatives of the social phobia probands as compared to the relatives of the other groups. Relatives of social phobia probands had an almost threefold increased risk for social phobia as compared to relatives of controls (16 versus 6 percent, $P < .05$). Similar results were found using the family history method (in this method the data on relatives came from interviewing patients about their relatives). Relatives of social phobia patients were reported to have a higher risk for social phobia (6.6 percent) as compared to both the relatives of panic disorder patients (.4 percent, $P < .001$) and not ill controls (2.2 $P < .1$).

The DSM-III-R and DSM-IV have divided social phobia into two subtypes: generalized (most social situations) and nongeneralized. If these represent two causally different syndromes they would be expected to be associated with different patterns of intergenerational transmission. For example, relatives of patients with generalized social phobia would be expected to have an elevated rate of generalized but not nongeneralized social phobia and vice versa. A 1998 direct interview family study addressed one aspect of this question. Rates of generalized and performance social phobia were compared in relatives (N = 111) of generalized social phobia patients and controls (N = 74). Significantly higher rates of generalized (24 versus 4 percent, odds ratio = 7.6) but not performance (15 versus 14 percent, NS) social phobia were found in the relatives of the generalized social phobia patients. This finding suggests a possible causal distinction between the two subtypes. However, since the study did not include a proband group of performance social phobia patients, it was not possible to determine whether this group would transmit an increased risk for generalized social phobia to their relatives. If so, this would suggest a partial overlap between the subtypes. Further studies in samples that

include larger numbers of nongeneralized social phobia probands are needed to clarify the relationship of these subtypes with respect to heritability.

Twin Studies of Irrational Social Fears A series of twin studies indicate a moderate genetic contribution to irrational social fears and shyness. It is not possible to determine the exact proportion of individuals in these studies who would meet criteria for DSM-IV social phobia. However, the consistency of these data makes a strong argument that genetic factors play an important if partial role in this area of human behavior.

Two reports have been published on a large twin-family study of individuals recruited through university registries in Indiana and the Indiana University Twin panel. A survey questionnaire was used to assess existence and severity of 51 commonly seen irrational fears, several related to social anxiety, in this population. Seven factors accounted for 45 percent of the variance in fear survey responses. Two of these (Negative Social Interactions/Social Criticism and Social Responsibility) are made up of fears that are considered to be part of social phobia. Negative social interactions included fears of making mistakes, looking foolish, or being criticized. Items loading on the social responsibility factor included public speaking, being a leader, and meeting new people.

The 1983 report by these investigators included 222 MZ and 132 DZ twin pairs. The heritability of the Negative Social Interactions and Social Responsibility factors in this sample was .44 and .60, respectively. This is interpreted as indicating that 44 percent of the liability to have irrational fears of negative social interaction and 60 percent of the liability to fears of social responsibility items is determined by genetic factors. The second report included 341 families (250 twin families and 91 sib families); the overlap between this and the 1983 sample is unclear. A multivariate path analysis was used to examine the relative fit of various genetic models. The best fitting model indicated: (1) moderate heritability of the two social fear factors; (2) a single genetic factor that accounted for most of the genetic liability for all 51 fears; (3) smaller genetic contributions that were specific to each of the fear factors; and (4) a moderate contribution from the common twin environment.

A similar study was conducted in 99 same-sex twin pairs (50 MZ and 49 DZ) from the Norwegian National Twin Register. However, in this case fear survey answers were elicited during a face-to-face clinical interview. Five factors accounted for over 50 percent of the variance in fear responses. Heritability of the social fears factor was .50 (fears of eating with strangers; being watched writing, working, or trembling). Using a slightly different approach a study was conducted of 99 MZ and DZ twin pairs using the California Personality Inventory (CPI). Forty-one of the 82 items in the inventory showed differences between MZ and DZ pairs, indicating genetic influence. A factor analysis of the 41 items revealed two factors related to social behavior (conversational poise and social ease). Items in these factors included: fear of finding anything to talk about with people, difficulty talking to or meeting new people, and fear of public speaking.

Childhood Temperaments Related to Social Behavior Two well-studied childhood temperaments (behavioral inhibition and shyness) may also contribute to the understanding of the development of social phobia. Each shares some clinical features with social anxiety and has been shown to have a genetic contribution to its etiology. However, there are also clear differences between each of these syndromes and social phobia. Moreover, as longitudinal studies have not been carried out, there is as yet no prospective evidence that either of these childhood syndromes increases the risk for adult social phobia.

Extensive studies have been conducted on the childhood temperament *behavioral inhibition*, which is found to occur in approximately 10 to 15 percent of white children and is characterized by timid, fearful, and withdrawn behavior on encountering unfamiliar people or situations. Another 10 to 15 percent of children show an opposite temperament (behaviorally uninhibited), which is characterized by outgoing, active, and exploratory behavior in new situations. More recent work indicates that these temperaments are distinguishable at as early as 4 months of age and are stable in 50 to 75 percent of individuals over the first 7 years of life. The suggestion that behavioral inhibition is linked to social phobia comes from both the observation of overlap in certain clinical symptoms and a family study that indicated higher rates of social phobia were found in the parents of behaviorally inhibited versus behaviorally uninhibited children versus controls (18 versus 0 versus 3 percent, $P < .05$).

A recent twin study of behavioral inhibition carried out in connection with a larger study of childhood temperaments reported the heritability at .49 (i.e., about 50 percent of the liability to develop behavioral inhibition comes from genetic factors). This moderate genetic contribution is consistent with those found in other studies of social anxiety. A second piece of evidence on genetic contribution to behavioral inhibition comes from breeding and cross-fostering (i.e., adoption) studies using animal models. There are several animal models of variation in reactivity to the unfamiliar; in each of these cases the more reactive strain shows the timid behavior and increased arousal that characterize Kagan's behaviorally inhibited children. That these behaviors can be bred in strains even if offspring are raised by unaffected individuals suggests a genetic contribution. For example, in a breeding study of rhesus monkeys it was found that the similarity with regard to extent of high reactive (i.e., behaviorally inhibited) behavior was proportional to degree of biological relatedness. High reactivity also persisted in monkeys who were cross-fostered (i.e., adopted) by low reactive mothers, suggesting that pedigree rather than mothering style is the predominant determinant of this trait.

A second set of developmental studies have focused on shyness. In this context shyness is defined as timid, avoidant, and passive behavior when confronted with an unfamiliar person. The Institute for Behavioral Genetics in Colorado has carried out several twin and adoption studies of this temperament in children. The results of the twin studies indicate significantly greater correlation of shyness scale scores between MZ versus DZ twins, supporting a genetic contribution to variation in this behavior. For example, 21 MZ and 25 DZ twin pairs aged 22 months were studied by observing their social response to structured situations involving their mothers and the research rater in their own homes. The types of social situations included: how long it takes the child to engage and interact with the researcher and how the child reacts to playing alone with the researcher or being separated from the mother. Correlation of scores between members of MZ pairs were in the range of .40 to .67 and were higher than those between members of DZ (–.05 to .34) pairs for all but one situation (time to smiling MZ .08 versus DZ .25). Similar results were found for children ranging from 4 to 10 years using both parental rating and direct observation.

In contrast to the twin data, the Colorado Adoption Project data were equivocal with respect to the possibility of a genetic contribution to shyness. This project included 182 adoptive and 164 nonadoptive families. Parents rated infant shyness at 12 and 24 months using a specially developed social ability-and-shyness scale. Typical items included: “child tends to be shy,” child takes a long time to warm up to strangers.” Parental shyness was measured using three self-report adult temperament scales. There were small but significant correlations between infant shyness and parental self-ratings on all three scales in the nonadoptive families (i.e., families in which biological parents raised their biological children) suggesting some degree of familial transmission of these behaviors. A genetic contribution would be indicated by high correlations between biological mothers and their adopted-away children and low correlations between adoptive parents and their adopted children. In actuality, in this sample there were very few strong correlations among any of these groups. Social ease in biological mothers was negatively correlated with shyness in their adopted-away 24-month-old infants (–15, $P < .05$). Correlations of similar strength were seen between this measure in adoptive mothers and fathers and shyness in their adopted (i.e., nonbiological) children.

Twin and family studies of social phobia and social fears indicate a moderate genetic contribution to these behaviors. Two childhood temperaments (shyness and behavioral inhibition), which also have a moderate genetic basis, show symptomatic overlap with social phobia. However, further studies of the adult outcome of these temperaments are needed to clarify their relationship to social phobia. Systematic DNA marker studies have not been reported.

SPECIFIC PHOBIAS

Although specific irrational fears and phobias had often been thought to originate completely through learned conditioning, there is significant evidence that genetic factors also play a definite (though partial) role in their development. However, neither adoption, DNA marker, nor segregation analytic studies have been reported. In addition to the obvious unanswered questions concerning mode and mechanism of transmission, two issues have been the subject of discussion in this field. These are: (1) the extent to which one inherits a susceptibility to a particular phobia versus a general phobia proneness; and (2) whether the genetic factors that influence irrational fears and phobias are the same. Irrational fears are distinguished from phobias by the fact that they do not involve either impairment or distress about the fear and therefore do not meet the DSM-III-R criteria for phobic disorder.

Twin Studies of Specific Fears and Phobias The Virginia Twin Registry study assessed heritability of two specific phobia subtypes: animal and situational (e.g., fears of tunnels, closed places, bridges, airplanes, high places). Each was found to have moderate heritability, although the genetic modeling analyses indicated different mechanisms of transmission. Animal phobias occurred in 11 percent of the sample. Probandwise concordance was 26 percent for MZ and 12 percent for DZ pairs. The best-fitting genetic model suggested that the major contributions to heritability of animal phobias were genetic factors common to all phobias (35 percent) and specific environmental events (59 percent; e.g., getting stuck in a tunnel). A completely different pattern was seen for situational phobias. Here probandwise concordance was similar in MZ versus DZ twin pairs (22 versus 24 percent). The best-fitting model indicated that 53 percent of the liability to these phobias comes from specific environmental events, 20 percent from specific, and 9 percent from common genetic factors.

Several additional twin studies have used the fear survey method to assess specific irrational fears. Subjects were asked to rate the severity of their fear of a variety of different situations on an ordinal scale (e.g., 0 = no fear, 5 = extremely anxious) As fear-related impairment and distress were not assessed, this methodology identified a mixed group some of whom will and some of whom will not meet the DSM-III-R or DSM-IV criteria for specific phobic disorder. These data must be

interpreted in light of this limitation.

Irrational fears were studied in a large twin sample recruited from university students and their families throughout Indiana. The first investigation included 222 MZ and 132 DZ same-sex twin pairs who were assessed for severity of fear of 51 commonly feared situations. Seven factors were found to account for 45 percent of the variance in fear response; two were related to social anxiety. The remaining five factors include irrational fears that are commonly the subject of specific phobic disorders. Each of these five fear factors was found to be influenced by genetic factors; however, the magnitude of heritability varied considerably. Relatively high heritabilities (indicating a large genetic contribution) were found for irrational fears of dangerous places (.58) for example, closed places, crowded places, heights; small organisms (.66) for example, snakes, mice, spiders; and personal death (.72), for example, death, untimely or early. Moderate heritabilities were found for the factors including fears of deep water (.32), including boating, swimming, deep water, and loved one's misfortunes (.28), for example, death, illness or injury of a loved one).

In a second study path analytic methods were applied to genetic analysis of 341 families (250 twin pairs, 91 sib pairs, and their parents). Assessment and data reduction methodology were similar to that used in the previous work. Seven factors accounted for over 50 percent of response variance. Two factors were comprised of social and four of specific irrational fears. The specific fear factors were similar to those identified in the earlier investigation. A seventh factor that is not clearly related to a particular DSM-IV phobia diagnosis included fears of death and illness of loved ones. A single genetic factor accounted for 80 percent of the variance in all fear responses in this sample and contributed to the development of each of the types of fears. In addition, specific genetic factors of varying strength were found for each fear type. Overall heritability was moderate. Heritabilities of each of the irrational fear types were: heights (.22), water (.20), dangerous places and organisms (.39), morbid settings (e.g., cemeteries, dead bodies, needles; .34).

Ninety-nine twin pairs from the Norwegian Twin Registry were studied using a fear survey administered during a clinical interview. Five factors, three of which are relevant to specific phobia, were found to account for over 50 percent of the variation in fear responses. For each of the three specific phobia factors the correlation of fear scale scores was significantly greater between MZ as compared to DZ twins, supporting a genetic contribution. Although the factors largely overlap with those identified in earlier research, the heritabilities were slightly higher. The three factors and their heritabilities were as follows: fears of animals (rats, mice, insects) .47; mutilation (blood, injury, medical procedures) .48; and nature (e.g., bridges, high places, closed places, tunnels, sharp objects, the ocean, fire).

These data indicate a genetic contribution to both irrational fears and phobias but do not address the question of their interrelationship with respect to intergenerational transmission. Two smaller studies provide some preliminary information on this issue. Both suggest that the transmitted component may be a susceptibility to a particular irrational fear whereas the possible consequence of avoidance or impairment is determined by other factors. However, further studies in larger and more well-defined samples are needed to further explore this question.

Forty-nine twin pairs from the Maudesly Twin Registry were studied. Twenty-one index twins had a phobia diagnosis. There was little difference between MZ versus DZ pairs in concordance for treatment of a phobia (13 percent in MZ versus 8 percent in DZ). However, when the diagnostic criteria were broadened to include phobic symptoms with or without treatment, concordance rates for MZ twins were substantially higher although not significantly different from those for DZ twins (88 versus 33 percent).

A subset of the Virginia Twin Registry study participants (541 MZ and 388 DZ twin pairs) were called back a year later to assess presence of phobias and irrational fears of blood, needles, hospitals, and illness. The best-fitting model indicated a genetic contribution to irrational fears of the first three items. However, occurrence of the DSM-IV disorder blood injury phobia (i.e., irrational fear of any of these situations, which interfered with the subject's life) was best explained by environmental factors.

Family Studies Several family and family history studies indicate a greater risk for phobic disorders among the relatives of phobia patients as compared to the relatives of controls. However, these data are somewhat difficult to interpret because most have used mixed proband groups that include patients with both social phobia and agoraphobia. The one direct interview family study of DSM-III-R simple phobia found a threefold increased rate of simple phobia among relatives ($N = 49$) of simple phobia probands as compared to relatives (119) of never–mentally-ill controls (31 versus 11 percent, $P < .005$, relative risk = 3.3). There was no evidence of familial aggregation of subdisorder irrational fears. Although this study is consistent with the twin data it requires replication in a larger sample.

Animal Models One striking feature of phobias is that they are irrational. That is, individuals with phobias are fearful of objects or situations that are not, from most people's point of view, realistically dangerous. This observation has led to the hypothesis that at least some phobias are acquired through conditioning. Conditioning is a process by which the animal acquires fear of a previously neutral stimulus through its repeated association with a noxious stimulus (e.g., shock). In a typical experiment a rat would be trained to fear a soft tone (conditioned stimulus) by repeated exposure to the tone followed by an electric shock (unconditioned stimulus). When conditioning takes place the rat will react to the tone as though it were the shock (i.e., with fear), even if the shock is no longer given. Animal models of conditioned fear are used to study the neurobiology of fear as well as of learning and memory. Experiments with fear conditioning have indicated that: (1) it is possible to breed strains of animals who are more or less fearful of specific stimuli and (2) there is genetic variability between strains in the strength and persistence of conditioned fear. Although this work is not directly applicable to the case of specific phobias in humans, it may provide starting points for development and testing of molecular genetic models.

These data indicate that genes contribute to the development of specific irrational fears and phobias. However, their role is moderate, leaving substantial influence to environmental factors (e.g., interpersonal, chance events). Most studies also indicate complexity in the structure of the genetic contribution. That is, there may be two types of genetic factors involved: those which influence an individual's likelihood to develop any phobia and those that are specific to fears or phobias of a particular situation. This suggestion of genetic heterogeneity within and across the DSM-IV categories is consistent with the diversity seen in studies of onset and course of specific phobias. Recent findings in animal models offer the possibility of dissecting phenotypes that may be more amenable to successful application of molecular genetic techniques.

OBSESSIVE-COMPULSIVE DISORDER

Data on the heritability of obsessive-compulsive disorder are inconclusive. Most but not all modern family studies have been positive, suggesting that there is intergenerational transmission of susceptibility to at least some subset of these cases. However, a genetic contribution to this disorder has not been well documented and there are no large systematic twin studies of this disorder. Recent findings in neuroimaging and treatment studies indicate that some cases of obsessive-compulsive disorder may be associated with specific autoimmune, receptor, or neuroanatomical abnormalities. These data suggest etiologic heterogeneity within the DSM-IV category; they have also increased interest in and provided some clues for understanding the possible genetic underpinnings of obsessive-compulsive disorder.

Twin and Family Studies Only one systematic twin study has assessed obsessive-compulsive symptoms and its results are suggestive but inconclusive concerning a genetic contribution to the disorder. Thirty (15 MZ and 15 DZ) of 49 twin pairs admitted to the Maudesly Twin Register had an index twin with obsessional symptoms. Probandwise concordance for treated obsessional symptoms that met DSM-III criteria was higher but not significantly different in MZ versus DZ twin pairs (33 versus 7 percent) When the diagnostic criteria were broadened to include either obsessive-compulsive disorder or untreated subclinical obsessional features, concordance rates were higher but still not significantly different (MZ versus DZ: 87 versus 47 percent). Diagnostic evaluations were not blind to diagnosis of cotwin.

Thirteen family studies of obsessive-compulsive disorder have been reported; however, only three employed state-of-the-art research methods (i.e., controls, blind assessment of relatives, and operational diagnostic criteria). Of these three studies two indicate familial aggregation in relatives of patients with obsessive-compulsive disorder versus relatives of normal controls, the third did not. One study found the disorder in 10.3 percent of the first-degree relatives of patients with obsessive-compulsive disorder as compared to only 1.9 percent of the relatives of psychiatrically well controls ($P < .005$). In one study the disorder was diagnosed in 7 percent of relatives of patients with the disorder as compared to 2 percent of relatives of controls ($P < .050$). In contrast, the third study found an increased rate of generalized anxiety disorder and any anxiety disorder among the relatives of patients with obsessive-compulsive disorder as compared to relatives of nonpsychiatrically ill controls. However, rates of the disorder did not differ between groups (2.5 percent versus 2.3 percent).

Interestingly, this mixed pattern of results is similar to that seen in earlier, less methodologically rigorous work. For example, of two earlier controlled (but not blind) studies, one reported a significantly higher rate of obsessive-compulsive disorder in relatives of probands with the disorder as compared to relatives of controls while the other study did not. A third study reported on families of early-onset obsessive-compulsive disorder (mean onset 10 years ± 3) using blind diagnostic ratings and DSM-III criteria, but no control group. Considerably higher rates of the disorder (13 percent) and subdisorder of the disorder symptoms (17 percent) among relatives

of childhood-onset probands with obsessive-compulsive disorder than are expected in the general population.

It has been suggested that these mixed findings on familial aggregation of the disorder reflect differences in methodology as well as etiologic heterogeneity within the diagnostic category. That is, some cases of obsessive-compulsive disorder have a heritable contribution whereas others do not. If so, then whether a study shows familial aggregation or not depends on the proportion of familial cases that are included in its sample. To investigate this hypothesis several recent studies have began to subtype patients by such variables as age at onset, presence of neurological symptoms, and symptom patterns of the disorder.

Relation Between Obsessive-Compulsive Disorder and Tourette's Disorder One interesting set of findings concerns the possible relationship between a subset of these cases and certain types of motor tic syndromes (i.e., Tourette's disorder and chronic motor tics). Increased rates of obsessive-compulsive disorder, Tourette's disorder, and chronic motor tics were found in the relatives of Tourette's disorder patients as compared to relatives of controls whether or not the patient had obsessive-compulsive disorder. However, most family studies of probands with obsessive-compulsive disorder have found elevated rates of Tourette's disorder and chronic motor tics only among the relatives of probands with obsessive-compulsive disorder who also have some form of tic disorder. Taken together these data suggest that there is a familial and perhaps genetic relationship between Tourette's disorder and chronic motor tics and some cases of obsessive-compulsive disorder. Cases of the latter in which the individual also manifests tics are the most likely to be related to Tourette's disorder and chronic motor tics. As there is considerable evidence of a genetic contribution to Tourette's disorder this finding also supports a genetic role in a subset of cases of obsessive-compulsive disorders.

POSTTRAUMATIC STRESS DISORDER

Posttraumatic stress disorder (PTSD) is the least studied anxiety disorder with respect to heritable or genetic contributions to etiology; only one twin and a few family studies have been reported. However, the data are intriguing and suggest further investigation in this area will both help clarify the causes of PTSD, and also improve the overall understanding of the multiple ways in which genes and environment can interact in the development of pathology.

The major indication of a genetic contribution to PTSD comes from a large twin study whose subjects were drawn from the Vietnam era veterans registry. Approximately two thirds of the registry twins (N = 4042 male twin pairs) responded to a mailed questionnaire. Independent genetic effects were found to influence both exposure to combat and development of PTSD symptoms. Within-pair correlation for most of the DSM-III-R PTSD symptoms was twice as high for MZ as compared to DZ twins. Although the prevalence of PTSD symptoms increased in proportion to extent of combat exposure, the MZ versus DZ twin pair difference held true even among veterans who did not serve in combat zones. Using a quantitative genetic model, genetic factors were estimated to account for approximately one third of the variance in liability to develop each of the DSM-III-R PTSD clusters in the overall sample (arousal cluster 28 to 32 percent, reexperiencing cluster 30 to 34 percent, and avoidance cluster 13 to 30 percent).

These findings must be considered with caution because the diagnoses are based completely on self-report rather than clinical evaluation and only two thirds of the sample responded. However, the suggestion of genetic influence on both (1) variation in response to stress; and (2) likelihood to expose oneself to situations leading to greater probability of trauma is consistent with several previous findings in epidemiological, personality, and animal model research.

In contrast to these twin study findings, several reported family history studies have not found increased rates of PTSD relatives of PTSD probands as compared to relatives of normal and psychiatrically ill controls. However, the significance of these results is difficult to interpret because the family information or sample sizes were insufficient to examine PTSD rates in relatives who were exposed versus not exposed to trauma. One study yielded family history data indicating an increased rate of anxiety disorders among relatives of PTSD probands compared to relatives of combat-exposed non-PTSD probands. This finding would suggest a heritable relationship between PTSD and other anxiety disorders. However, it was not confirmed in a subsequent direct interview family study and no adoption studies on PTSD have been reported.

Another interesting approach to this problem is reported by researchers who studied the occurrence of PTSD in adolescent Cambodian refugees and their parents. Although subjects were currently living in the United States, all had had significant exposure to war trauma before immigration. Within each family the trauma exposure of the parent and adolescent was similar. Assessments were by direct interview. Significantly higher rates of DSM-III-R PTSD were found among adolescents whose parents had versus those who did not have PTSD (16/60 versus 4/58, 27 versus 7 percent, $P < .05$). This effect was independent of socioeconomic status, amount of reported war trauma, living arrangements, or treatment.

The data summarized above indicate that genetic factors play an important role in the development of anxiety disorders. However, as the heritability of these syndromes is only moderate, nongenetic factors are equally crucial. In addition, as the patterns of intergenerational transmission do not follow Mendelian rules, anxiety disorders fall into the ever-enlarging category of genetically complex disorders. Future research will have to take into account the high probability that there is widespread pleiotropy (i.e., different clinical syndromes associated with the same genetic underpinnings) and genetic heterogeneity (i.e., different genes can lead to the same clinical phenotype). It will also need to focus not only on genetic mechanisms that underlie susceptibility to pathological anxiety states, but also on nongenetic contributions and how (and if) there is interaction between several different genes as well as between genetic and nongenetic causes.

The classification of anxiety disorders has become increasingly specific. Overall the family study and genetic data are supportive of these new distinctions. For several anxiety disorders there is evidence of specific genetic contributions that increase susceptibility to that illness but not to the other anxiety syndromes. However, there are also data that indicate the existence of genetic factors that increase an individual's likelihood for any anxiety disorder; at least one investigation finds that both types of genetic effects (general and disorder specific) are important. These data are completely consistent with the current knowledge of the etiological heterogeneity seen in genetically complex disorders. However, taken together they suggest that the actual genetics of anxiety disorders may be much more complicated than suggested by previous etiological theories.

SUGGESTED CROSS-REFERENCES

Population genetics in psychiatry is discussed in Section 1.17 and genetic linkage analysis of psychiatric disorders is in Section 1.18. Epidemiology is covered in Section 5.1.

SECTION REFERENCES

Black DW, Noyes R Jr, Goldstein RB, Blum N: A family study of obsessive-compulsive disorder. Arch Gen Psychiatry *49:*362, 1992.

Brown FW: Heredity in the psychoneuroses. Proc R Soc Med. *35:*785, 1942.

Carey G, Gottsman II: Defining cases by genetic criteria. In *What is a Case?* JK Wing, P Bebbington, LN Robins, editors. Grant McIntyre, London, 1981.

*Carey G, Gottesman II: Twin and family studies of anxiety, phobic, and obsessive disorders. In *Anxiety New Research and Changing Concepts*, DF Klein, JG Rabkin, editors. Raven, New York, 1981.

Cohen ME, Badal DW, Kilpatrick A, Reed EW, White PD: The high familial prevalence of neurocirculatory asthenia (anxiety neurosis, effort syndrome). Am J Hum Genet *3:*126, 1951.

Coryell W: Hypersensitivity to carbon dioxide as a disease-specific trait marker. Biol Psychiatry *41:*259, 1997.

*Crowe RR, Noye R, Pauls DL, Slymen DJ: A family study of panic disorder. Arch Gen Psychiatry *40:*1065, 1983.

Fyer AJ, Mannuzza S, Chapman T, Liebowitz MR, Klein DF: A direct interview family study of social phobia. Arch Gen Psychiatry *50:*286, 1993.

Fyer AJ, Mannuzza S, Chapman TF, Martin LY, Klein DF: Specificity in familial aggregation of phobic disorders. Arch Gen Psychiatry *52:*564, 1995.

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King MC: Linkage of early-onset familial breast cancer to chromosome 17q21. Science *250:*1684, 1990.

Henikoff S, Greene EA, Pietrokousk S, Bork P, Attwood TK, Hood L: Gene families the taxonomy of protein paralogs and chimeras. Science *278:*609, 1997.

Kagan J, Snidman N: Temperamental factors in human development. Am Psychol *46:*856, 1991.

*Katsuragi S, Kunugi H, Sano A, Tsutsumi T, Isogawa K, Nanko S, Akiyoshi J: Association between serotonin transporter gene polymorphism and anxiety-related traits. Biol Psychiatry *45:*368, 1999.

Kendler KS, Heath AC, Martin NG, Eaves LJ: Symptoms of anxiety and symptoms of depression: Same genes, different environments? Arch Gen Psychiatry *44:*451, 1987.

*Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: The genetic epidemiology of phobias in women: The interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. Arch Gen Psychiatry *49:*273, 1992.

Kendler KS, Neale MC, Kessler RC, Heath AC, Eaves LJ: Panic disorder in women: A population-based twin study. Psychol Med *23:*397, 1993.

Knowles JA, Fyer AJ, Vieland VJ, Weissman MM, Hodge SE, Heiman GA, Haghighi F, De Jesus GM, Rassnick H, Preud'homme-Rivelli X, Autsin T, Cunjak J, Mick S, Fine JD, Woodley KA, Das K, Maier W, Adams PB, Freimer NB, Klein DF, Gilliam TC: Results of a genome-wide genetic screen for panic disorder. Am J Med Genetic *81:*139, 1998.

*Lander ES, Schork NJ: Genetic dissection of complex traits. Science *265:*2037, 1994.

Lyons MJ, Goldberg J, Eisen SA, True W, Tsuaung MT, Meyer JM, Henderson WG: Do genes influence exposure to trauma: A twin study of combat. Am J Med Genet *48:*22, 1993.

Noyes R Jr, Clarkson C, Crowe RR, Yates WR, McChesney CM: A family study of generalized anxiety disorder. Am J Psychiatry *144:*1019, 1987.

Ohman A: Face the beast and fear the face: Animal and social fears as prototypes for evolutionary analyses of emotion. Psychophysiology *23:*123, 1986.

Pauls DL, Alsobrook JP, Goodman W, Rasmussen S, Leckman JF: A family study of obsessive-compulsive disorder. Am J Psychiatry *152:*76, 1995.

Perna G, Cocchi S, Bertani A, Arancio C, Bellodi L: Sensitivity to 35% CO_2 in healthy first-degree relatives of patients with panic disorder. Am J Psychiatry *152:*623, 1995.

Phillips K, Fulker DW, Rose RJ: Path analysis of seven fear factors in adult twin and sibling pairs and their parents. Genet Epidemiol *4:*345, 1987.

Plomin R, Daniels D: Genetics and shyness. In *Shyness*, WH Jones, M Cheek, SR Briggs, editors. Plenum, New York, 1986.

Rose RJ, Miller JZ, Pogue-Gelle MF, Cardwell GF: Twin-family studies of common fears and phobias. In *Twin Research 3: Intelligence, Personality, and Development*. Liss, New York, 1981.

P>Rosenbaum JF, Biederman J, Hirshfeld DR, Bolduc EA, Faraone SV, Kagan J, Snidman N, Reznick JS: Further evidence of an association between behavioral inhibition and anxiety disorders: Results from a family study of children from a nonclinical sample. J Psychiatr Res *25:*49, 1991.

Sack WH, Clarke GN, Seeley J: Posttraumatic stress disorder across two generations of Cambodian refugees. J Am Acad Child Adolesc Psychiatry *34(9)*:1160, 1995.

*Stein MB, Chartier MJ, Hazen AL, Kozak MA, Tancer ME, Lander S, Furer P, Chubaty D, Walker JR: A direct-interview family study of generalized social phobia. Am J Psychiatry *155:*1, 1998.

Suomi SJ: Genetic, maternal, and environmental influences on social development in rhesus monkeys. In *Primate Behavior and Social Biology*, AB Chiarelli, RS Corruccini, editors. Springer, New York, 1987.

Suomi SJ, Kraemer GW, Baysinger CM, DeLizio RD: Inherited and experiential factors associated with individual differences in anxious behavior displayed by rhesus monkeys. In *Anxiety, New Research and Changing Concepts*, Klein DR, Rabkin J, editors. Raven, New York, 1981.

Swedo SE, Leonard HL, Mittleman BB, Allen AJ, Rapoport JL, Dow SP, Kanter ME, Chapman F, Zabriskie J: Identification of children with pediatric autoimmune neuropsychiatric disorders associated with streptococcal infections by a marker associated with rheumatic fever. Am J Psychiatry *154:*110, 1997.

Tatusov RL, Koonin EV, Lipman DJ: A genomic perspective on protein families. Science *278:*631, 1997.

Torgersen S: Genetic factors in anxiety disorders. Arch Gen Psychiatry *40:*1085, 1983.

Torgersen S: The nature and origin of common phobic fears. Br J Psychiatry *134:*1979.

True WR, Rice J, Eisen SA, Heath AC, Goldberg J, Lyons MJ, Nowak J: A twin study of genetic and environmental contributions to liability for posttraumatic stress symptoms. Arch Gen Psychiatry *50:*257, 1993.

Vieland VJ, Hodge SE, Lish JD, Adams P, Weissman MM: Segregation analysis of panic disorder. Psychiatr Genet *3:*63, 1993.

Weissman MM, Warner V, Wickramaratne P, Moreau D, Olfson M: Offspring of depressed parents. 10 years later. Arch Gen Psychiatry *54:*932, 1997.

Weissman MM, Wickramaratne P, Adams P, Lish JD, Horwath E, Charney D, Woods SW, Leeman E, Frosch E: The relationship between panic disorder and major depression: A new family study. Arch Gen Psychiatry *50:*767, 1993.

15.5 ANXIETY DISORDERS: PSYCHODYNAMIC ASPECTS

GLEN O. GABBARD, M.D

FREUD'S THEORY OF ANXIETY

Sigmund Freud placed anxiety at the core of neurosis; hence the history of the development of psychoanalytic theory can be linked to the evolution of Freud's understanding of anxiety. In 1895 he put forth his original theory of anxiety, which postulated a biologically based cause. Specifically Freud believed that the accumulation of undischarged libido was transformed into anxiety. In his view the "dammed-up" libido could either result from external obstacles to its appropriate discharge or from internal inhibitions related to unconscious conflicts about sexual gratification. Freud did not fully explain how the transformation of accumulated libido into symptomatic anxiety occurred—he merely suggested that such a transformation must occur.

Freud referred to this form of anxiety as *actual neurosis*, and this designation clearly referred to a pathological type of overwhelming fear. It was directly linked to drive energy and associated with a number of physiological symptoms, including manifestations of autonomic discharge, such as profuse sweating, diarrhea, and increased respiratory and heart rates. He also connected it with an overwhelming sense of panic and terror.

This rather nonpsychological formulation of anxiety was retained until 1926, when Freud wrote his classic monograph *Inhibitions, Symptoms, and Anxiety*. This publication appeared 3 years after *The Ego and the Id*, which suggested that the tripartite structural model of *id, ego, and superego*, was the key to understanding mental phenomena. His 1926 monograph was essentially a new theory of anxiety based on his understanding of interagency conflict in the structural model. Although he had now abandoned the original idea that undischarged libido is transformed into anxiety, he nevertheless continued to regard anxiety as having a biological or genetic basis. He recognized that there was clearly a Darwinian survival value in reacting to external threats with anxiety; hence he believed that every human organism must be congenitally endowed with this capacity to respond to danger situations both physiologically and psychologically. However, Freud's theory of anxiety was not designed to identify the origin of anxiety as a human phenomenon; instead, its emphasis was on how to understand the role of anxiety in the individual's mental life.

Traumatic situations were regarded as central to the development of anxiety according to Freud's new theory. The prototype of the traumatic situation was the experience of birth, when the infant is overwhelmed by a flood of stimuli arising from external sources as well as from internal sensations. Freud believed that this prototypical model could be used to understand anxiety in other situations in the course of development where individuals are overwhelmed by an influx of stimuli that they are incapable of mastering. The emphasis in his theory was primarily on stimuli arising from the id, specifically aggressive and sexual wishes that had been repressed and are striving to make their presence known.

A second aspect of Freud's new theory on anxiety involved "danger" situations. Freud suggested that as children develop they learn to anticipate a traumatic situation before it happens and to react to it with anxiety. Freud termed this reaction *signal anxiety* because it was produced by a situation in which the individual anticipated danger. In response to this signal of danger, the ego mobilized defense mechanisms to prevent unacceptable thoughts and feelings from emerging into conscious awareness.

Freud used the situation of an infant being left alone by its mother as a way of illustrating a typical danger situation. Infants learn that a relationship exists between the absence of the mother and the experiencing of unpleasurable states of internal turmoil. Based on this experience, the ego learns that separation from the mother is a danger situation and responds accordingly with signal anxiety. If the signal anxiety fails to activate the ego's defensive resources adequately, more persistent or even neurotic symptoms will result. In this regard, Freud conceptualized anxiety as both a symptomatic manifestation of neurotic conflict and an adaptive signal to avoid awareness of neurotic conflict.

In this second aspect of Freud's new theory of anxiety, the pleasure principle was intimately related to the function of the anxiety. The signal anxiety that the ego learns to produce in response to the anticipation of danger was viewed as relatively less unpleasurable than the anxiety that would develop if there were no signal and the traumatic situation of abandonment developed fully. Signal anxiety is a way of attenuating a more profound and terrifying anxiety.

In this model the phenomenon of anxiety is an ego affect. The signal of danger produced by the ego offers opposition to the emergence of repressed id impulses into conscious awareness. The ego's defense mechanisms are marshaled in response to the signal anxiety. One such defense mechanism is *repression*, which keeps undesirable id impulses, whether desires, wish-fulfilling fantasies, memories, or affects, from entering the consciousness. The ego serves to censor both the impulse itself and the corresponding intrapsychic representation of that impulse. Nevertheless, the impulse or drive-related wish that has been repressed may still find a way to express itself in the form of a symptom. By the time it manifests itself as a symptom, however, it is likely to be disguised and displaced. For example, the symptom of hysterical paralysis might reflect an underlying wish to strike a parent with murderous intent.

The underlying wish related to drive pressures may also be defended against through *reaction formation*. This defense mechanism involves an overemphasis on one side of an ambivalent attitude. If love and hate are both directed towards a parent, the defense of reaction formation maintains the hate as an unconscious part of the affective state while emphasizing love. The manifestation of this reaction formation might be excessive devotion to the parent and professions of love without any conscious awareness of negative feelings. Hence an exaggeration of a feeling or attitude led Freud to wonder about the other side of the equation buried in the unconscious. This awareness of reaction formation had been observed by writers years before Freud's formulation of anxiety. Ralph Waldo Emerson, for example, once noted, "The louder he spoke of his honor, the faster we counted our spoons."

KLEIN'S VIEWS OF ANXIETY

Melanie Klein expanded on Freud's view by developing a theory of internal object relations linked to drives. She regarded fear of annihilation as the most fundamental anxiety and related it to Freud's death instinct. In her view, the ego engaged in a splitting process to deal with that fear of annihilation. All the derivatives of the death instinct, such as sadism, hatred, aggression, and any form of "badness," were evacuated from the infant and projected into the mother. The infant then began to suffer *persecutory anxiety* from the death instinct derivatives projected into the mother. This form of anxiety, which Klein linked to the paranoid-schizoid position, involved a fear that the "bad" mother created by the infant's projections would invade the infant and destroy all the "good" aspects of the infant.

As the infant moved from the paranoid-schizoid position to the *depressive position*, splitting of the ego was overcome by an integrative effort that was designed to link good and bad aspects of the self and of objects. Instead of viewing the mother as "all bad," the child is now able to see that mother has both good and bad qualities. Similarly, infants also become aware of the presence of good and bad within themselves. When the child becomes aware that the loving and nurturing mother is basically the same person as the hateful, rejecting mother, an important developmental moment has been achieved. Now the infant becomes concerned that it may have harmed or destroyed its mother through its hostile and sadistic fantasies directed toward her. This developmental achievement leads to *depressive anxiety*. While the infant was mainly concerned about being attacked by the persecutory object during the paranoid-schizoid position, in the depressive position the infant's primary anxiety is that it may hurt love objects, particularly the mother. The child then worries that the good object may be lost through the child's own aggression and sadism. Hence depressive anxiety can be summarized as a concern about the loss of the love object through one's own destructiveness. The child learns to deal with

these guilt feelings through a process of *reparation*, in which the child attempts to repair the perceived damage through loving behavior toward the ambivalently regarded object.

SELF PSYCHOLOGICAL VIEWS OF ANXIETY

Heinz Kohut developed self psychology in the late 1960s and continued his work until his death in 1981. He deemphasized drives and conflicts and emphasized a deficit model in which the self was regarded as immature and lacking in maternal supplies. The emphasis in self psychology is on infantile needs rather than repressed wishes or drives. Kohut felt that the need to maintain self-esteem and well-being was just as powerful as sexuality and aggression in molding the human organism. In his view the child has powerful needs to idealize a parent; to receive affirmation, validation, and empathy from that parent; and to maintain a sense of wholeness of the self.

Within this model, *disintegration anxiety* is the most fundamental concern of the individual. This anxiety is generated by the child's concern that the failure of adequate selfobject responses from persons in the environment will lead to fragmentation of the self. If empathy, affirmation, and validation are not forthcoming from significant individuals, such as parents, the child may then resort to pathological behavior to restore harmony to the self. In adult life these behaviors may include such things as gambling, perverse sexual behavior, disordered eating behaviors, substance abuse, and a myriad of other so-called “acting-out” behaviors. Kohut believed that the anxieties based on the internal drive pressures, such as sexuality and aggression, were breakdown products of preOedipal failures in the provision of selfobject needs.

HIERARCHY OF ANXIETY

It is possible to construct a developmental hierarchy of anxiety based on Freud's contributions to the subsequent elaborations by Klein and Kohut (Table 15.5-1).

- Superego anxiety
- Castration anxiety
- Fear of loss of love
- Separation anxiety (fear of the loss of the object—Kleinian depressive anxiety)
- Persecutory anxiety (Klein)
- Disintegration anxiety (Kohut)

Adapted from Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

Table 15.5-1 A Developmental Hierarchy of Anxiety

The psychodynamic clinician can use this hierarchy to guide the understanding of the unconscious origins of anxiety in a specific patient. These may be regarded as “danger situations” in terms of Freud's 1926 model. The highest level of developmental danger is that of guilt. The presence of superego anxiety reflects that moral standards of conduct have been internalized in the form of parental introjects. The superego provides love and approval or punishment and disapproval based on how closely the individual's behavior (or fantasy life) adheres to these standards of conduct.

Moving down the developmental hierarchy, the next typical danger situation is associated with the oedipal phase of development and focuses on the potential damage to or loss of the genitals at the hands of a retaliatory parental figure. This fear is often expressed metaphorically as a loss of a different body part or any other form of physical injury. The next type of anxiety involves the loss of parental love. In this situation, even though the parent (or parent substitute) is present physically, the child may still fear the loss of that object's love. In the adult patient this anxiety may manifest itself as a preoccupation with behaving in a manner that ensures a significant object's ongoing love and attention to fend off the anxiety that the object may become disappointed with the patient.

Although the foregoing forms of anxiety are typical of neurotically organized patients, the last three points in the hierarchy are often found as the predominant anxiety in more severely disturbed patients. For example, separation anxiety or the fear of abandonment by the love object is typical of patients with borderline personality disorder. Many of these patients lack object constancy, so they cannot internalize a soothing internal presence in the absence of the parental object. They often develop intense anxiety that a significant other, such as a therapist, a parent, a spouse, or a close friend, will suddenly disappear and no longer be there as a source of support. Margaret Mahler linked this to the rapprochement subphase of separation-individuation.

The most primitive forms of anxiety are often found in patients with borderline personality disorder or narcissistic personality disorder and those with psychotic disorders. Those with persecutory anxiety often use projection as a defense and have a paranoid mode of organizing experience, believing that there are outside malevolent forces that will invade and destroy them from within unless vigilance is maintained. Disintegration anxiety can manifest itself either through the fear of fragmentation (the loss of one's sense of a cohesive self) or through merger or fusion with an object. This form of anxiety is particularly prominent in patients who have poorly developed ego boundaries.

The hierarchical nature of these developmental anxieties may lead the clinician to assume that the more primitive levels of anxiety are “outgrown” as the child develops. Clinical wisdom, however, has taught us that to some degree the most primitive levels of anxiety persist in everyone. In a well-integrated and high-functioning individual superego anxiety may be the most prominent or leading anxiety, the more fundamental forms of anxiety may be activated in situations of great stress or trauma. Similarly, large group situations are often highly conducive to triggering persecutory anxiety, in which “outsiders” or those with different political views are seen as monstrous attackers. The history of Western civilization is characterized by the development of warring factions based on relatively minor differences in culture, religion, or politics.

A corollary of this view that all levels of anxiety persist to some extent in every individual is that symptoms may be multiply determined by anxiety stemming from different developmental levels. Consider the following example:

A 26-year-old man came to psychoanalytic treatment because of anxieties triggered by sexual intercourse with his fiancée. Although he professed to love this woman and considered their relationship to be mutually satisfying, he found himself much less interested in sexual intercourse than she was. During the act of intercourse he found himself wishing his fiancée would reach orgasm more quickly so he could retract his penis from her vagina. At the same time he felt reluctant to ejaculate in her, and he often stopped intercourse before reaching orgasm.

When his analyst asked him about his fear of ejaculating during intercourse, he said that he wanted to avoid an out-of-wedlock pregnancy. The analyst noted that the patient had previously conveyed that his fiancée was taking oral contraceptives. The patient confirmed that indeed she was, but he recognized that there was still a slight risk of pregnancy associated with birth control pills. The analyst asked the patient if he had any other fears associated with ejaculation.

The patient initially said that he did not, but he then associated to a dream in which he was having intercourse with a former high school girlfriend, and after the sexual act was consummated, he noticed that his voice had gone up two octaves and he “talked like a girl.” He looked in the mirror and he had become a woman who looked exactly like his fiancée. As the analyst and the patient worked together to analyze the dream through understanding the patient's associations, it became clear that two separate but related anxieties were operating together to produce the symptom. At one level the patient was worried about losing his penis in his fiancée's vagina. This form of castration anxiety was acting in concert with a more primitive concern—that at the moment of orgasm he would fuse and become one with his fiancée and lose his own identity. The patient stressed that orgasm seemed to be the one moment when he tended to lose a sense of his own boundaries and felt like he was merging with the other person.

This clinical vignette illustrates that a patient often is not aware of the source of anxiety because it is unconscious. Although in this case a vivid dream and its accompanying associations made the meaning of the anxiety much clearer, in many other cases much more psychotherapeutic or psychoanalytic exploration is necessary before the unconscious sources are accessible. The developmental hierarchy provides a set of guidelines for clinicians to think about when they are trying to understand a patient's anxiety. However, there are infinite individual variations on these different themes of loss of the object, persecutory anxiety, loss of the object's love, and so forth. The psychodynamic approach to anxiety is one that acknowledges the uniqueness of each individual's intrapsychic world. In each case the clinician must allow for the possibility that the origins of anxiety may not fit these guidelines precisely and may involve a blend of other concerns that may expand psychodynamic theory.

CONTEMPORARY VIEWS OF ANXIETY

The publication of the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) marked the beginning of an effort to make the official nomenclature of the American Psychiatric Association atheoretical. An unfortunate consequence of this shift has been the decline of interest in the classical neurotic entities along with the psychodynamic model of symptom formation inherent in them. Three separate categories have replaced the neuroses: dissociative disorders, somatoform disorders, and anxiety disorders. The category of anxiety disorders has been further subdivided in the fourth edition of DSM (DSM-IV) into the following entities: panic disorder, phobias, obsessive-compulsive disorder, posttraumatic stress disorder, acute stress disorder, and generalized anxiety disorder.

This classification reflects growing empirical evidence linking biological mechanisms to the generation of anxiety. Panic attacks, for example, can be experimentally induced by lactate infusion. Moreover, genetic studies suggest a higher concordance rate for certain anxiety disorders in monozygotic twins than in dizygotic twins. Acute panic attacks appear to arise in the brainstem and represent spontaneous hyperactivity of noradrenergic nuclei in conjunction with lowered firing thresholds in medullary respiratory chemoreceptors. By contrast, anticipatory anxiety seems to represent the effect of a kindling in the limbic lobe. Phobic avoidance appears to be cortical in origin.

Although these advances in the field of neuroscience are impressive and of considerable heuristic value in understanding anxiety, they by no means preclude a psychodynamic understanding or the use of psychotherapeutic techniques. Of greater concern is that the rise of the atheoretical categorizations in the official nomenclature and the associated research on the biological underpinnings of anxiety may encourage clinicians to think about anxiety as only an illness rather than an overdetermined symptom of unconscious conflict. This view of anxiety may be linked to a specific form of treatment designed to eradicate or at least to reduce the anxiety. This perspective fails to recognize the value of anxiety as a signal of internal distress and conflict that may require some reflection and introspection. In the Menninger Foundation's Psychotherapy Research Project, 18 of 35 patients showed increased anxiety at the termination of psychoanalysis or psychotherapy even though 13 of these 18 cases were judged by independent raters to have undergone substantial improvement. These results were regarded by the investigators as a sign that primary anxiety, which is disorganizing to the patient, must be differentiated from anxiety used as a signal. They stressed that an increase in *anxiety tolerance*, defined as the capacity to experience anxiety without having to discharge it, may be one positive outcome of dynamic psychotherapy or psychoanalysis and may reflect an expansion of the ego's capacity to deal with anxiety. The presence or absence of anxiety after treatment was insufficient to understand or evaluate the change, because many of the patients with good outcomes showed dramatic improvement in their capacity to use ideational activity more efficiently in the service of understanding the anxiety.

Anxiety is likely to surface at various developmental crises throughout one's life. It is a regular accompaniment of existential uncertainty inherent in life. The expansion of ego mastery over anxiety as a result of psychoanalytic therapy appears to facilitate the patient's capacity to use and understand anxiety in a variety of circumstances in which it arises. The anxiety can be normal, adaptive, or maladaptive, and the assumption that all anxiety should be eradicated is certainly unwarranted.

Any neurophysiological mechanism involved in the generation of anxiety may produce an adaptive form of signal anxiety as well as the more pathological forms of chronic symptomatic anxiety. It would be erroneous to assume that the biological mechanisms are only associated with highly pathological or symptomatic anxieties, or that psychotherapeutic treatments should deal only with signal aspects of anxiety. Psychoanalytic therapy obviously must work by altering brain functioning.

Recent genetic studies have demonstrated that individuals with a somewhat shorter gene involved in serotonin transport may have greater anxiety than those with the longer version. Although the investigators have speculated that many other genes as well as numerous environmental factors may be involved in the generation of anxiety, the findings are encouraging because the shorter gene also is functionally different from the longer gene. It is relatively weaker in terms of its capacity to transport serotonin compared to the longer gene. Of particular interest to this discussion is that nearly 70 percent of people have the shorter and less vigorous version of the gene, which results in greater anxiety. This skewed distribution may well reflect natural selection in that individuals with greater anxiety might be better equipped to survive the dangers in the environment than those who experience less worry or concern.

Other problems also have resulted from the differentiation of Axis I anxiety disorders in DSM-IV according to specific symptom profiles and temporal criteria. Many researchers as well as a good many more clinicians have argued that the anxiety syndromes have reified distinctions that are more illusory than real. In actual clinical practice there is a growing recognition of comorbidity among anxiety disorders, largely because many of the disorders are not as separate from one another as the criteria imply. Nevertheless, in this chapter the psychodynamic considerations of each of the major anxiety disorders will retain the DSM-IV classification to promote uniformity across the chapters and to illustrate that a psychodynamic perspective can be applied separately or be integrated with the descriptive and biological approaches to these entities.

PANIC DISORDER

Patients with panic disorder generally experience a good deal of distress during the few minutes that the typical panic attack lasts. They may find themselves gasping for breath, shaking uncontrollably, feeling lightheaded, sweating profusely, and feeling certain that death is imminent. The majority of such patients also struggle with agoraphobia, in that they are intensely afraid of being stuck in a situation from which they cannot escape. Many patients with panic disorder develop anticipatory anxiety secondarily because they are worried most of the time about when and where the next panic attack will strike them. In some cases patients with panic disorder accompanied by agoraphobia severely restrict their travel so they will not have to encounter a situation in which a panic attack occurs in a place from which they cannot make a ready escape.

Many patients suffering from panic disorder experience the attacks as if they “came out of the blue.” They cannot necessarily identify a precipitating factor linked to the onset. Such patients, and often clinicians as well, may thus regard panic attacks as psychologically “contentless.” Yet careful psychodynamic interviewing of such patients often reveals overlooked factors that may be involved in triggering the onset of panic.

Experienced psychoanalysts interviewed nine consecutive patients with panic disorder on videotape. A research psychiatrist with psychoanalytic training then reviewed the videotapes and was able to identify psychologically meaningful stressors preceding the onset of panic in each patient evaluated in this manner. These stressors were connected with an alteration of or decrease in the level of expectations placed on the patient. The expectation changes were often related to job situations in which patients had to take on increased responsibilities. Loss events generally involved emotional or physical separations from key figures in the patients' lives. These events were typically associated with childhood experiences where an attachment to an important person in the patients' lives was threatened. Sexual conflicts did not appear to be prominent in triggering the onset of panic, but seven of the nine patients had considerable difficulty dealing with anger. Another common denominator was the perception of parents as frightening, controlling, demanding, temperamental, and critical.

Further analysis of the interviews suggested a pattern of childhood difficulties. Anxiety about socializing with others, parental relationships that were perceived as unsupportive, a sense of frustration or resentment, and a feeling of being trapped were experienced by the patients. Aggression was difficult to handle, and most of the patients described feelings of self-reproach and inadequacy.

Some of these observations have been borne out by empirical studies. Compared to normal control subjects, patients with panic disorder report higher rates of dysfunctional parenting and more intense separation anxiety in childhood. In a large study of 1018 female twin pairs, the relationship between parental loss prior to the age of 17 and adult psychopathology was investigated. Panic disorder was strongly and significantly associated with both parental separation and death. Early maternal separation in particular was linked to panic disorder. Other studies have suggested higher rates of stressful life events, especially loss, when compared with controls in the months preceding the onset of panic disorder. There is also preliminary evidence of attachment disturbance in adulthood as well as a pattern of insecurity and dependency in relationships.

There may be a predisposing neurophysiological vulnerability that may interact with certain kinds of environmental stressors to produce the end result of a panic attack. Researchers have identified an inborn temperamental characteristic in a number of children that has been termed behavioral inhibition to the unfamiliar. These children have a tendency to be easily frightened by anything unfamiliar in the environment. To deal with their fear, they tend to rely on their parents to protect them. As they grow they often are afraid that their parents will not be there to protect and comfort them. They may externalize their own inadequacies on their parents, whom they regard as unpredictable and unreliable. These children then become angry at their parents' inconsistent availability, but the anger creates new problems in that they worry that their angry fantasies will be destructive and drive their parents away, leaving them with the loss of a parent who they depend on to provide safety. A vicious cycle results where the child's anger threatens the connection with the parent and thus increases the child's hostile and fearful dependence.

The alternative scenario is that an individual without the inborn temperamental predisposition to fear the unfamiliar is sufficiently traumatized by a series of stressors in development such that the same type of cycle results. These individuals may experience rejection or extremely controlling behavior by parents or parental substitutes that contribute to the same pattern of fearful dependence.

Some form of attachment difficulty appears to be involved in the pathogenesis of panic disorder. A small preliminary study of attachment style in 18 women suffering from anxiety disorders suggested that all had problematic attachment styles. Fourteen of the 18 were diagnosed with panic disorder. These patients had higher rates of preoccupied attachment. Other studies have shown patients with panic disorder to have greater separation anxiety. Fearful individuals tend to view separation and attachment as mutually exclusive. They have difficulty developing the normal oscillation between separation and attachment because they have a heightened sensitivity to both loss of safety or protection and loss of freedom. Hence they end up operating in an extremely narrow range of behavior that attempts to avoid attachment that is too intense or separation that is too frightening. The end result is an overcontrolling style of interaction.

To use Freud's terminology, the signal anxiety function is insufficient to deal with anxiety, leading to an overwhelming and uncontrollable level of panic. Threats to attachment in particular appear to trigger this kind of overwhelming panic. Patients who develop panic disorder are prone to feelings of self-fragmentation and may need a companion (or therapist) to help them feel that they have a firm sense of identity. Some investigators have suggested that the presence of ego defects involving self-other confusion may be related to difficulties in using anxiety as a signal. The intensity of panic attacks may appear to be so biologically driven that medication appears to be the only viable treatment approach. However, when the underlying fear of separation and abandonment is seen to be connected with the panic, psychotherapeutic interventions may be equally effective.

A patient with borderline personality disorder and poorly established object constancy could not deal with her psychotherapist's absence over extended weekends. While sitting in her apartment she would be unable to visualize her therapist's face or imagine his voice as a way of soothing herself. She began to hyperventilate, experience tachycardia, feel that her death was imminent, and would become terrified that her therapist would not be there for her next session. When the panic became overwhelming, she would call her therapist on the weekend. The call would generally last less than 30 seconds. The patient would report that she was afraid something terrible had happened to her therapist and that she would never see him again. The therapist would reassuringly state that everything was fine and he would see her at their Tuesday appointment. The patient's panic symptoms including tachycardia and hyperventilation, immediately dissipated on hearing the therapist's voice, and she was able to get off the telephone and return to her weekend activities. In the psychotherapy sessions the patient corroborated that hearing her therapist's voice restored a sense of attachment to him. Such situations confirm the fact that the physiological symptoms such as tachycardia and hyperventilation may be reduced just as dramatically with psychotherapeutic as with pharmacotherapeutic interventions.

Another etiological factor in women patients appears to be childhood physical and sexual abuse. Both forms of abuse have been found to be significantly higher in persons with anxiety disorder when compared to a control group. Childhood sexual abuse in particular has been found to occur at a rate of 45.1 percent in women with anxiety disorders compared to a rate of 15.4 percent in a comparison group. The link with panic disorder is even more striking—60 percent of women with panic disorder have a history of childhood sexual abuse compared to 31 percent of women with other anxiety disorders. Since childhood trauma interferes with the child's attachment to the parents, childhood sexual abuse could account for some of the difficulties that panic disorder patients have in feeling safe and secure with significant objects in their lives. These patients may have internal representations of abusive relationships that interfere with the development of trust.

The psychodynamics of panic disorder are summarized in Table 15.5-2.

1. Difficulty tolerating anger.
2. Physical or emotional separations from significant person both in childhood and in adult life.
3. May be triggered by situations of increased work responsibilities.
4. Perception of parents as controlling, frightening, critical, and demanding.
5. Internal representations of relationships involving sexual or physical abuse.
6. A chronic sense of feeling trapped.
7. Vicious cycle of anger at parental-rejecting behavior followed by anxiety that the fantasy will destroy the tie to parents.
8. Failure of signal anxiety function in ego related to self fragmentation and self-other boundary confusion.
9. Typical defense mechanisms: reaction formation, undoing, somatization, and externalization.

Table 15.5-2 Psychodynamic Themes in Panic Disorder

Psychotherapeutic Techniques In the course of psychodynamic therapy the patient's difficulties in relationships often become centered in the transference to the therapist. Conflicts around separation, independence, and anger are particularly prominent. Active exploration of the patient's fears of becoming overly dependent on the therapist are a crucial part of the psychotherapeutic work. Conversely, anxieties over the loss of the therapist may be equally meaningful to the patient.

The patient's difficulty with anger may take the form of avoidance. Patients who feel particularly threatened by anger towards the therapist may simply forget the session or find an excuse not to appear. The fears of anger can often be clarified by an examination of childhood situations in which anger was expressed in the family. Anger may be regarded as uncontrollable because of the patient's experiences with parental anger. When the patient became angry with the parent as a child, the parent may have emotionally abandoned the child. Because patients with panic disorder often avoid expressing anger, the therapist's initial task may be to actively explore the resistances to acknowledging anger.

As the transference fantasies about the therapist develop, there is generally a great deal of useful information contained in these fantasies. For example, patients may be convinced that they must operate within a narrow range of behavior or risk losing the therapist's interest or investment. These fantasies can often be traced back to childhood fears of being alone or independent.

An integral part of the psychodynamic approach to psychotherapy is the examination of characteristic defense mechanisms. Patients with panic disorder typically use any combination of the following defenses: reaction formation, undoing, somatization, and externalization. Both reaction formation and undoing serve to reassure the patient that negative affects are either absent or completely controllable. The psychotherapist's strategy with such defenses is to make the patient aware of his or her anxiety about the expression of anger and the associated need to disavow it. Hence a patient who repeatedly assures the therapist that he loves his father and that his father can do no wrong may be actively disavowing his anger. The therapist might intervene by noting the defense, as in the following observation: “I often feel that you're very invested in letting me know how much you love your father. It makes me wonder what other feelings you have about him.”

The defenses of somatization and externalization are both designed to avoid reflection and introspection. Somatization places the patient's focus on what is happening physiologically. It is an attempt to make the panic attack a phenomenon that is devoid of psychological causes or meaning. Externalization places the origin of the problem in external persons who are viewed as mistreating the patient in some way. In some cases the focus on somatization may unconsciously create a specific form of object relationship in which the patient enlists others as healers who are supposed to fix something in the body. This object relationship may be activated in the transference, and the therapist may clarify how the preoccupation with bodily symptoms establishes a dependent relationship with the therapist. Anger at the therapist may then be explored by examining the patient's disappointment that the therapist is unable to “fix” the patient's problems. A long-range

psychotherapeutic strategy is to help the patient view the panic as an internally generated phenomenon rather than to blame others for it.

There is a good deal of outcome data suggesting that cognitive therapy and pharmacotherapy are effective in treating panic disorder. Psychodynamic therapy has not been studied as extensively as the other two modalities, but there are numerous case reports in the literature suggesting that dynamic therapy is useful. Moreover, at least one study used a randomized, controlled design that demonstrates the effectiveness of dynamic therapy at reducing relapse in patients with panic disorder. In this study patients were randomized to treatment with either clomipramine (Anafranil) for 9 months or clomipramine for 9 months combined with 15 weekly sessions of brief dynamic therapy. Whereas all patients in both groups were free of panic attacks within 26 weeks of the start of treatment, the relapse rate was significantly higher in the group that received only clomipramine when the pharmacotherapy was terminated. The investigators concluded that brief dynamic therapy may reduce the psychosocial vulnerability that is associated with panic disorder.

Integrations of cognitive techniques with dynamic psychotherapy are also useful. In virtually every case of panic disorder, distortions of thinking need to be addressed as part of the psychotherapeutic strategy.

Although both cognitive therapy and pharmacotherapy are effective in the short run, long-term follow-up studies suggest that some patients continue to experience symptoms when the medication is withdrawn or when the brief cognitive therapy has terminated; most patients will require additional treatment. Studies are needed to demonstrate which patients may be particularly suited to extended psychoanalytic therapy either alone or in combination with medication and other techniques.

PHOBIAS

Phobias are the most common of all anxiety disorders. In DSM-IV they are subdivided into three categories: (1) agoraphobia without history of panic disorder, (2) specific phobia, and (3) social phobia. Social phobia, which is the most common of all phobias, is further subdivided into two subtypes. The *nongeneralized type* is characterized by a fear of public situations such as public speaking or performing on stage. As many as 20 percent of the populations surveyed have acknowledged this fear. The second variety is a *generalized type* in which almost all social interactions are feared. This variant may be difficult to differentiate from avoidant personality disorder.

Freud regarded the phobias as neurotic entities and therefore understood the mechanism of symptom formation to be consistent with his 1926 theory of anxiety. When unacceptable aggressive or sexual thoughts that could lead to a fantasy of punishment begin to emerge from the unconscious, signal anxiety is produced, which then leads to the deployment of specific ego defense mechanisms. The first of these is *displacement*, which involves the redirection of anxiety associated with an unconscious source to a conscious substitute that is often intrinsically harmless. *Projection* is the second specific defense mechanism used by phobics to get the source outside of themselves and into the external world. The third defense is *avoidance*, which is simply a systematic process of not coming into contact with the displaced and projected item that the anxiety is associated with. If that item is elevators, the individual always takes the stairs. If the item is dogs, the individual avoids dogs. The end result is that the three combined defenses may eliminate the anxiety because the unacceptable or forbidden thought is re-repressed. However, the anxiety generated by the thought is controlled at the cost of creating a phobic neurosis with all its associated inconvenience.

Behavioral therapies using relaxation and in vivo exposure procedures have been highly effective in treating phobias. Cognitive therapies, which usually include an exposure element, have also been effective. There are few controlled studies of psychotherapies other than those of the cognitive or behavioral orientation. Nonetheless, psychodynamic understanding may enhance a broad-based overall treatment plan.

Phobias are best understood as reflecting a genetic-constitutional diathesis that is acted upon by environmental stressors. A study of 2163 female twins concluded that the best model for the disorder is the inheritance of a vulnerability to phobias that require environmental factors specific to the individual to produce a diagnosable phobic syndrome. In this study population, parental death before the age of 17 was clearly associated with an increased risk for phobia.

Work on the temperamental construct of behavior inhibition to the unfamiliar appears to be applicable to social phobia in much the same way it is relative to panic disorder. The expectation that one will be embarrassed and humiliated by the critical scrutiny of others appears to have some origin in temperamental disposition. Children who are behaviorally inhibited appear to be born with a much lower threshold for limbic-hypothalamic arousal in response to unexpected or novel events in the environment that are difficult to assimilate. To result in shy, withdrawn, and timid behavior at 2 years of age, however, chronic environmental stressors must act on the basic temperamental disposition that was present at birth. It has been suggested that some of these stressors might include criticism and humiliation from an older sibling, parental fighting, and separation from a parent through death or divorce. It was also possible to differentiate children who develop shyness purely in response to environmental factors alone without the underlying temperament.

A clinical cohort study examined the families of these children and found that the parents were at greater risk for most anxiety disorders, and particularly social phobia. Compared with parents in control groups, they had higher rates of two or more anxiety disorders. The investigators speculate that such children with behavioral inhibition may go on to develop anxiety disorders in part because of genetic factors and in part because their parents may have greater anxiety and thus convey messages to their children that the world is unsafe.

Social phobia is an example of an anxiety disorder that may have been artificially distinguished from other related disorders. The lifetime comorbidity approaches 70 percent with persons with social phobia. This entity is rarely treated by mental health professionals unless it is accompanied by other disorders that bring the person to attention. An underlying genetic constitutional diathesis may actually predispose an individual to several anxiety disorders.

Psychotherapeutic Techniques Dynamic psychotherapy with individuals who have social phobia attempts to get to the heart of the patient's fears. Embarrassment and shame are central affects, and the patient's affect states are a good starting point. As the patient begins to feel embarrassed, the therapist encourages exploration of the patient's fantasies of how others will react to him or her. Within these perceptions or fears of other people's reactions, the therapist will find a characteristic pattern of internal object relationships within the patient that is then externalized in any social situation. Patients with social phobia often have internalized representations of others, usually parents, siblings, or caretakers, that involve shaming, ridiculing, humiliating, criticizing, and abandoning the patient. In some cases these representations will be distortions of the way childhood interactions actually occurred; in others they will be closely related to the actual behavior by family members. In either case the therapist's task is to help the patient see that the fantasies of how others react are largely based on adverse childhood experiences rather than on the reality of the interactions in the patient's current environment. Often therapeutic work within the transference is extremely helpful in this regard, as in the following example:

A 27-year-old woman who could not finish her PhD in history came to psychotherapy because a selective serotonin reuptake inhibitor had not significantly helped her with her social phobia. She was too anxious to meet with her dissertation committee so her academic progress had been seriously delayed. After missing one session of her therapy, she came to the next session and apologized for her absence. The therapist asked, “What made you stay away last time?”

The patient responded, “I was afraid you were going to criticize me for not going to my meeting with my dissertation committee.”

The therapist asked, “Have you experienced me as critical of you in the time we've worked together?

The patient replied, “No, not really. You're actually the opposite of critical. You're patient with me, supportive, and always trying to help in any way you can.”

The therapist then said, “It may be then that the perception of me being critical arises from within yourself rather than being based on my actual behavior with you.”

The patient reflected for a moment and said, “There must be something to that. I'm always imagining that others are going to criticize or embarrass me, even when I don't have any hard evidence that such a response is likely.”

The therapist made the following observation, “Your dissertation committee has not actually been critical of you, so it may be that the same concerns you have here with me are applicable with your committee. In other words, your fear of negative responses from them may largely be an internal one rather than a reality-based fear.”

Discussion In this case, the therapist is drawing a link between the anxiety in the transference and the anxiety in a present-day situation outside of the therapy. As the therapy progressed, he also made a linkage to the patient's childhood when the patient felt repeatedly humiliated and criticized by her parents.

The interpersonal dimensions of phobias are often crucially important in the treatment of phobias. In the case of individuals for example, one often finds that the spouse of such a patient may be instrumental in maintaining the symptomatic behavior. For example, the husband of a woman with agoraphobia may feel much more secure with his wife in the house because he fears that she will find another man if she leaves the home. The husband's jealousy is controlled by keeping his wife indoors. Any treatment effort that tries to change the nature of this system may be doomed because the husband will not cooperate with the treatment plan. The patient herself may also resist the treatment because she fears her husband's jealous rage. Hence in many cases of resistance to behavioral or cognitive-behavioral interventions, a psychodynamic approach may be needed either in a marital or an individual context.

An overview of psychodynamic aspects of phobias is summarized in Table 15.5-3.

- Principle defense mechanisms include: displacement, projection, and avoidance.
- Environmental stressors, including humiliation and criticism from an older sibling, parental fights, or loss and separation from parents, interact with a genetic-constitutional diathesis.
- A characteristic pattern of internal object relations is externalized in social situations in the case of social phobia.
- Anticipation of humiliation, criticism, and ridicule is projected onto individuals in the environment.
- Shame and embarrassment are the principal affect states.
- Family members may encourage phobic behavior and serve as obstacles to any treatment plan.
- Self-exposure to the feared situation is a basic principle of all treatment

Table 15.5-3 Psychodynamic Themes in Phobias

OBSESSIVE-COMPULSIVE DISORDER

Within the psychoanalytic literature there has been a frequent confusion between obsessive-compulsive personality disorder and obsessive-compulsive disorder. While they were previously thought to reside on the same continuum of symptomatology, the evidence is mounting that the two are somewhat distinct entities. Patients with obsessive-compulsive disorder view their symptoms ego-dystonically, that is, they experience distress connected with their symptoms. By contrast, the traits of patients with obsessive-compulsive personality disorder are often ego-syntonic and may disturb others more than the patient. Moreover, the vast majority of patients with obsessive-compulsive disorder do not have symptoms of the other condition. Finally, although obsessive-compulsive personality disorder generally responds very well to psychoanalytic therapy, obsessive-compulsive disorder is notoriously refractory to psychodynamic treatment.

Much of the confusion began with Freud's classic 1909 paper describing his psychoanalysis of the “Rat Man.” He referred to this condition as obsessive-compulsive neurosis and developed a compelling psychodynamic formulation of the symptoms. According to this formulation the anxiety produced by the fear of retaliation associated with the oedipal situation leads the individual with obsessive-compulsive neurosis to regress to a more primitive psychosexual developmental stage, specifically to the anal phase. Freud suggested that certain characteristic defense mechanisms were associated with the anal phase, including reaction formation, doing and undoing, and isolation of affect. He believed that the regression was more likely because of the long standing presence of anal fixations related to disturbances during toilet training.

Freud's sophisticated psychodynamic formulation has not been able to eradicate the rituals and obsessional thoughts of most patients with obsessive-compulsive disorder. The resistance of these symptoms to interpretation may relate to the biological underpinnings of the disorder. In comparison to normal controls, patients with obsessive-compulsive disorder manifest increased metabolism in the orbitofrontal cortex, the anterior cingulate cortex, and the caudate nuclei. Some studies using magnetic resonance imaging (MRI) have found significantly less total white matter associated with significantly greater total cortex and opercular volumes in these patients as compared to normal controls. There are also higher rates of concordance for obsessive-compulsive disorder in monozygotic than in dizygotic twins. Studies of computed tomography have found a smaller caudate nucleus volume in these patients as well. Animal models have suggested serotonergic abnormalities in obsessive-compulsive disorder and the fact that symptomatic improvement occurs with serotonin reuptake inhibitors suggests that this hypothesis has some merit.

Behavior therapy has been shown in numerous studies to be successful in reducing the symptoms of obsessive-compulsive disorder. Many clinicians now regard the combination of a selective serotonin reuptake inhibitor (SSRI) and behavior therapy as the treatment of choice. Despite the lack of data that psychoanalytic or psychodynamic therapy reduces the symptoms of obsessive-compulsive disorder, a psychodynamic approach may still be of considerable value. Psychodynamic clinicians are always thinking in terms of resistance, transference, countertransference, and unconscious meanings, even if they are not involved in the provision of formal psychoanalytic psychotherapy. Hence when treating patients who have obsessive-compulsive disorder, even if the treatment involves prescribing an SSRI, a clinician nevertheless may think about the patient's situation and the overall treatment within a psychodynamic framework. Part of this framework is the notion that symptoms, no matter how biologically based, nevertheless have meanings to the patient. Psychodynamic conflicts frequently attach themselves to neurophysiologically driven symptoms and use them as a vehicle for their expression. Moreover, the symptoms often result in secondary gains that serve as powerful resistances to treatment. Often a psychodynamic understanding of the resistance may lead a patient who is reluctant to seek help to consider behavior therapy and an SSRI.

Another area in which psychodynamic theory has relevance to obsessive-compulsive disorder is in understanding the impact of stressors on the condition. Research suggests that while genetic processes are clearly involved in the development of the disorder, environmental stressors also are important. Issues involving child care and pregnancy appear to be particularly influential. In one study 69 percent of patients with obsessive-compulsive disorder could link exacerbation or onset of their symptoms to childbirth, pregnancy, or child care issues. Another study found that pregnancy was associated with the onset of symptoms of obsessive-compulsive

disorder far more than any other life event. In clinical work with young mothers or young women who are pregnant, unconscious or nearly conscious feelings of hostility toward the child often increase anxiety, which in turn increases a reliance on the rituals of obsessive-compulsive disorder to bind the anxiety.

A 24-year-old woman who was learning to take care of her 2-month-old infant was deeply distressed by thoughts that she might accidentally kill the baby by dropping her. She found herself constantly checking the floor to see if there was a soft area in case the baby fell. She also found herself washing her hands up to 60 times a day because of her concern that she might transmit an infection of some sort to the baby, who would then die from the transmitted agent. When she was evaluated for treatment, she adamantly denied any hostile thoughts toward the child. However, when the clinician explained to her that wishes to harm the child and hostile feelings toward the child are part of the normal mothering experience, the patient's defensiveness decreased and she was able to acknowledge that she harbored such thoughts. She immediately condemned herself for having them and was deeply worried that the interviewer would be just as harsh on her as she was on herself.

Psychodynamically Informed Treatment Psychodynamic therapy is not usually indicated as the exclusive treatment for obsessive-compulsive disorder, but in patients who tenaciously hang on to their symptoms because of their special meanings or because of the interpersonal control they exercise over others a psychodynamic approach may be extremely useful. Similarly, problems of compliance with medication and difficulties in interpersonal relationships may all be productively addressed through psychodynamically informed strategies. The mean reduction of symptoms of obsessive-compulsive disorder when treated with pharmacotherapy is in the range of 38 to 40 percent, so a variety of problems persist despite the partial efficacy of medications. Many of these difficulties lie in the area of personal relationships. A high risk of separation or divorce has been linked to obsessive-compulsive disorder. The illness itself is extraordinarily stressful to those in close relationships with the patient, and psychodynamic understanding may be useful in this regard.

A study of 34 parents or spouses of obsessive-compulsive disorder patients examined the extent to which these relatives would accommodate the patient through active participation in rituals or significant modifications of their daily routines. Of those relatives, 88.2 percent reported accommodating the patient in some way. This form of family accommodation was clearly correlated with stress in the family, rejecting attitudes toward the patient, and poor family functioning. These accommodations often involve an effort to reduce the patient's anxiety or to control expressions of anger by the patient. Many relatives describe feeling bullied by the patient into doing things to accommodate the patient's obsessions or compulsions.

A 26-year-old single man was admitted to a psychiatric hospital with obsessive-compulsive disorder because he had refused to cooperate with pharmacotherapy and behavior therapy approaches. He was obsessed with the concern that his father and mother might bring the “AIDS virus” into the home, so he asked that everything be sprayed with disinfectants. He also insisted that his parents wear gloves around the house and specifically when handling his personal items or touching the doorknob of his room. His parents had reluctantly complied with these instructions because they feared his wrath when they did not go along with him.

When he arrived in the hospital, the young man became concerned that the previous occupant of his room might have masturbated and left semen stains in his room that could transmit HIV infection to him. Even though the hospital staff tried to engage him in discussions of other issues, he repeatedly wanted to get everyone to discuss the possibility of human immunodeficiency virus (HIV) contamination in the room. One nurse who was assigned to him was told each day not to touch anything in his room unless he was wearing gloves. Eventually, the nurse capitulated and began wearing gloves when entering the patient's room. The nurse felt embarrassed about this capitulation and tried to keep it secret from the other hospital staff, but he was eventually discovered wearing gloves when he emerged from the patient's room. The behavior led to a useful discussion of countertransference in treating such patients. It also led to a productive exploration of the patient's entitlement. He was indignant when people did not comply with his expectations. In this regard his hospital psychiatrist confronted him with the observation that he tended to deny the autonomy and subjectivity of anyone else in his life. All family members and all treaters were supposed to respond primarily to his needs rather than their own needs. These confrontations eventually helped the patient to recognize his omnipotent control of others, and he began to withdraw his demandingness.

Discussion This case example reflects how patterns of relationships become internalized because of the specific needs of the patient with obsessive-compulsive disorder. His needs to avoid contamination became paramount over every other concern in his environment. This pattern of omnipotent control was internalized as his characteristic mode of object relatedness. His admission into the hospital and the recapitulation of that pattern of relatedness is an example of the time-honored principle of psychodynamic hospital treatment—namely, that the patient attempts to re-create his family situation in the milieu of the hospital with various staff members.

This case also illustrates the extent to which significant characterological issues may operate in concert with the symptoms of obsessive-compulsive disorder. To some degree the patient had everyone in his family waiting on him hand and foot and doing exactly what he wanted them to do. He was thus reluctant to enter into any treatment program that might change his behaviors and loosen the hold he had on his family. Studies have suggested that patients with comorbid personality disorders in association with obsessive-compulsive disorder are much more likely to terminate behavior therapy and be classified as treatment resistant.

Psychodynamically informed treatment has a continued role to play with patients who have obsessive-compulsive disorder. These symptoms, however biologically driven, are nevertheless rich in unconscious meaning. Psychodynamic understanding of those meanings may reveal the key to addressing poor patient compliance with pharmacotherapy or other forms of resistance. A psychodynamic approach is also helpful in understanding the meaning of psychosocial triggers that exacerbate these symptoms, which almost always have interpersonal meanings that must be addressed. Psychodynamic conflicts often use biologically determined symptoms to serve as a vehicle for their expression. Family members and others in significant relationships with the patient often accommodate to the patient's demands as a way of placating the patient. This characteristic pattern of relating to others may also surface in the transference-countertransference dimensions of the treatment. Characterological features are often significant in obsessive-compulsive disorder patients and may influence the response to treatment rather dramatically. Addressing these characterological features must be an overall part of the treatment plan in such cases. The psychodynamic themes of the disorder are summarized in Table 15.5-4.

- Classical psychoanalytic formulation was linked to an anal-phase regression with a specific constellation of defenses: reaction formation, doing and undoing, and isolation of affect.
- The psychodynamic meaning of psychosocial stressors, such as pregnancy or childbirth, may help to explain the onset or exacerbation of symptoms of obsessive-compulsive disorder.
- Psychodynamic conflicts frequently appropriate biologically driven symptoms and use them as a vehicle for the expression of those conflicts.
- Treatment resistance, such as poor compliance with medication, often involves the tendency to hang on to the symptoms because of special meanings to the patient or because of characterological resistances to receiving help.
- Symptoms of obsessive-compulsive disorder almost always have interpersonal meanings, including omnipotent control of family members and others in significant interpersonal relationships.

Table 15.5-4 Psychodynamic Themes in Obsessive-Compulsive Disorder

POSTTRAUMATIC STRESS DISORDER

Posttraumatic stress disorder first entered the official diagnostic nomenclature of the American Psychiatric Association in 1980 when DSM-III was published. The identification of trauma-related psychopathology in returning Vietnam War veterans appeared to have a major influence on the acceptance of the diagnosis in American psychiatry. Nevertheless, there has been a long tradition of psychoanalytic exploration of trauma, beginning with Freud's observations that a splitting of consciousness appeared to occur in hysterical patients who reported a history of childhood sexual trauma. Although his emphasis shifted to infantile sexuality and fantasy in the pathogenesis of neurosis, he never abandoned his belief that childhood sexual abuse was at the core of some neuroses. Freud also noted the existence of traumatic war neuroses and suggested that the external trauma reawakened infantile trauma in these conditions. He postulated that the phenomenon of the repetition compulsion in the form of repeating the traumatic event was a response to overwhelming assault on the individual's stimulus barrier, which activated a

regression.

The seminal work of Abraham Kardiner and Herbert Spiegel with traumatized soldiers in World War II suggested that the external factors related to the trauma were most decisive but that in some instances a preexisting conflict might be symbolically reawakened by the traumatic event. However, their view was that the conflict did not cause the event to be traumatic. In more contemporary studies of the impact of trauma, both the traumatic event itself and its meaning to the individual based on that individual's earlier life experiences are regarded as important in determining whether an event becomes traumatic. Empirical studies support this view in that as the intensity of a violent event increases, so does the number of individuals traumatized by that event. Also, certain types of events appear to be more likely to create traumatic reactions in victims than others. In one study of burn patients, for example, the severity of the posttraumatic symptoms was not necessarily proportional to the severity of the burn. Posttraumatic stress disorder was predicted by smaller burns, by greater emotional distress, and by the perception of inadequate emotional support. Approximately one fourth of people witnessing a violent event may develop posttraumatic stress disorder. Those who are more likely to develop the disorder have a history of early separation from parents; a family history of anxiety and neuroticism; and preexisting depression or anxiety. Hence most researchers today believe that a personal predisposition is necessary to develop posttraumatic stress disorder in the face of trauma.

As the DSM-IV symptoms of this disorder suggest, the basic processes that result from psychic trauma revolve around three different symptoms clusters: (1) intrusive and repetitive reexperiencing of the trauma; (2) avoidance of this reexperiencing; (3) overactivation of the autonomic nervous system that is, sleep difficulties, hypervigilance, startle reactions, and difficulty concentrating. Many other patients adapt to psychic trauma in childhood by shutting down any affective development. As a result, they cannot use affects as signals. Powerful feelings are viewed as a threat that the original traumatic situation will recur, so these patients often use somatization as a defense but do so at a cost. They cannot soothe themselves or relax and are often deficient in self-care. These patients may be incapable of speaking the language of emotions and hence have been referred to as having alexithymia.

Dissociation is another response to trauma. One of the best predictors of the ultimate development of posttraumatic stress disorder is dissociation at the moment of the trauma. Through dissociation trauma victims symbolically remove themselves from the trauma by depersonalizing or perceiving the incident as though it is happening to someone else rather than to them. Dissociation is also connected with affect dysregulation, and this difficulty modulating various affects states, including anxiety, has led the disorder to be grouped among the anxiety disorders.

This multifaceted response to trauma explains why the occurrence of posttraumatic disorder in pure form is unusual. Most people with its symptoms have other diagnoses as well, such as dissociative or somatization disorders. Many borderline personality disorder patients also have significant histories of trauma.

Trauma shatters the individual's view of the world as a place that is safe, predictable, and controllable and forces a confrontation with one's own vulnerability. Consequently, a variety of defense mechanisms are activated to deal with it. Many of these defenses marshaled by the ego in the face of trauma are those commonly thought to be primitive or immature. Hence there is frequently a regression to developmentally earlier modes of dealing with helplessness, vulnerability, fear, and anger.

Many patients use rather extraordinary degrees of denial to avoid facing the severity of the trauma or the extent of distress they feel as a result of it. A related defense is that of minimization, often seen in the wake of trauma when victimized individuals repeatedly tell friends and family that the trauma was really not that bad and that they are doing fine.

The sense of rage at being victimized may be disavowed and projected into others who they view as aggressive and angry. They may become hypervigilant in an effort to protect themselves from the aggression they perceive in others. This defensive projective disavowal may be accompanied by splitting, in which they view themselves as completely free of angry or aggressive impulses. In other patients anger itself may be used as a defense against more disturbing feelings of vulnerability. Some patients pursue litigation in an angry manner that suggests a powerful revenge motive and that also reflects a wish to master the feelings of helplessness by gaining some form of compensation for their suffering.

Another common defense in the wake of trauma is guilt. Many rape victims, for example, feel responsible for what happened to them. Often beneath this feeling of guilt and responsibility is a more disturbing feeling that they are completely helpless in a malevolent universe where violence is random. In more chronic cases of posttraumatic stress disorder a particular relational mode is also observable in the clinical setting as it manifests itself in the transference and countertransference aspects of the therapy. This way of relating to others reflects a constellation of internal self and object representations that are best described as characters in a psychological drama. They include a victim, an omnipotent rescuer, and an abuser. These roles are projected onto the therapist, reintrojected into the patient, and reprojected, creating an oscillating pattern in which at times the therapist is viewed as an abuser and at other times the patient may regard the therapist as an idealized and omnipotent rescuer. Sometimes therapists find themselves feeling victimized by such patients because the patients are demanding and feel entitled to be compensated for their injuries. A common countertransference pattern is to try to avoid the role of the abuser by resorting to a *disidentification with the aggressor* in which the therapist strives to be the most compassionate and loving parent imaginable to make up for the damage that the patient feels was caused in childhood by parental abuse. Psychodynamic themes in posttraumatic stress disorder are summarized in Table 15.5-5.

- The subjective meaning of a stressor may determine its traumatogenicity.
- Traumatic events may resonate with childhood traumas.
- Inability to regulate affect may result from trauma.
- Somatization and alexythymia may be among the aftereffects of trauma.
- Common defenses used include denial, minimization, splitting, projective disavowal, dissociation, and guilt (as a defense against underlying helplessness).
- Mode of object relatedness involves projection and introjection of the following roles: omnipotent rescuer, abuser, and victim.

Table 15.5-5 Psychodynamic Themes in Posttraumatic Stress Disorder

Psychotherapeutic Approaches The psychotherapist must always keep in mind that these roles or “characters” are not unidimensional but involve complexities that vary with the uniqueness of each individual patient and that involve specific meanings and affects that complicate the perception of who is doing what to whom at any given moment.

The research on psychotherapy with posttraumatic stress disorder patients is limited in its ability to answer what kind of treatment works best with which patients. One well-controlled study of brief psychodynamic psychotherapy with victims of fire demonstrated significant improvement in the patients who participated. Nineteen of the 30 patients met DSM-III criteria for posttraumatic stress disorder alone or with comorbid depression. The therapy consisted of 6 to 12 sessions following a manualized protocol that encouraged exposure to feared situations. Another study compared psychodynamic therapy with systemic desensitization, hypnotherapy, and a waiting list control. All three of the treatments resulted in improvements; dynamic therapy seemed to be most helpful in reducing avoidant symptoms and systemic desensitization and hypnotherapy seemed to have a greater impact on intrusive thoughts. However, only 23 of the 112 patients in this sample had actually experienced trauma; the remaining patients had suffered bereavement.

The treatment goals of dynamic psychotherapy with patients who have posttraumatic stress disorder include reducing anxiety, integrating and accepting the trauma as part of the self, regaining a sense of mastery, re-establishing a sense of personal integrity, and moving away from a sense of being haunted by the past toward feeling fully engaged in the present. The earlier that a clinician intervenes in the course of the posttraumatic reaction, the more successful the intervention is likely to be.

Much of the art of the therapy involves striking a balance between forcing the patient to reconstruct a complete picture of the trauma and assuming an observing, detached posture that allows much information to be left out. The therapist must tactfully strive to integrate the memory of the trauma with the patient's continuous sense of self while also being aware that the patients need to proceed at their own pace because the traumatic memories can be overwhelming. Building a therapeutic alliance of safety and security is essential. Without a solid alliance, the patient may not trust the therapist sufficiently to be able to discuss the trauma. The therapeutic

alliance may be particularly problematic because many traumatized patients profoundly distrust everyone, including the therapist.

Educating the patient about the common reactions to trauma may be useful in facilitating an alliance. Similarly, empathically validating patients' rights to feel the way that they do may also create the conditions necessary to explore the trauma. In the case of childhood trauma, such as sexual or physical abuse, the therapist must help the patient see the difference between having no responsibility in childhood for what a parent did and taking full responsibility as an adult for choosing abusive partners that repeat the trauma from childhood.

Differentiating reality from fantasy is a basic technique used to facilitate the integration of the traumatic information. Much of this differentiation involves transference work to help the patient see the distinction between the reality of the therapist's attitude and the representation of the therapist in the patient's mind. Four types of transference have been identified: (1) the transfer of specific disavowed memories of the traumatic event into the treatment situation, (2) the transfer of figures involved in the traumatic event onto the therapist, (3) the transfer onto the therapist of intrapsychic functions in the patient that have been distorted as a result of the trauma (with the hope that healthier function will be restored), and (4) the transfer onto the therapist of an omnipotent and wise role in which the therapist can help the patient make sense out of what happened and restore a sense of personal meaning.

The psychotherapist in turn may feel helpless, overwhelmed, angry for the patient's tenacious hanging on the trauma, omnipotent in fantasies of saving or rescuing the patient from the trauma, indifferent, and hopeless. Some of these transference and countertransference paradigms are reflected in the following case example:

> A 35-year-old woman who had witnessed her son's death in an auto accident felt she could never get over what happened. Much of her psychotherapy focused on her repeated reexperiencing of the event, including vivid nightmares that caused her to wake up crying. The therapist repeatedly tried to get the patient to mourn the loss of her child, but the patient steadfastly refused, saying she did not know how to grieve and she did not think she could ever get over it. The therapist interpreted her wish to hang on the images of the son as a way of avoiding the mourning. The interpretations fell on deaf ears, and the patient continued to talk about the accident and her feelings of guilt in connection with it. After months of this repetitive interaction in the therapy, the therapist began to feel angry and irritated at the patient for thwarting all his attempts. He became more and more forceful in his efforts to provide her with insight about how she was holding on to the past. Finally, the patient experienced him as intrusive and unsympathetic, and she shouted at him: “Would you stop trying to make me do what you want me to do and just leave me alone?!” As the therapist reflected on this outburst, he recognized that he had felt so helpless and inadequate that he had had re-created an abusive situation in the therapeutic relationship. The patient was again feeling victimized, not only by the cruel hand of fate in the death of her son and by the drunken driver who killed him but also by the therapist himself. As this fragment of a psychotherapy process reveals, many patients with posttraumatic stress disorder present formidable challenges to the psychotherapist:

Some cases appear to respond much better than others. Although precise data are not available, a certain percentage of patients do not respond to psychotherapeutic efforts and their symptoms may become chronic. Many of these patients feel that time has frozen for them at the moment of the trauma, and they are unable to move forward in their lives. In the national comorbidity survey, survival analysis showed that more than one third of people with an index episode of posttraumatic stress disorder failed to recover even after many years.

ACUTE STRESS DISORDER

Acute stress disorder is a relatively new diagnostic entity. Although it shares many of the symptoms of posttraumatic stress disorder, this disorder occurs within 4 weeks of the traumatic event; in contrast, posttraumatic stress disorder has its onset at least 1 month after the trauma. Many of the same principles involved in the psychodynamic treatment of the latter disorder apply to acute stress disorder as well. Therapy should begin as soon as possible after the trauma has occurred because better outcomes have been associated with more immediate crisis intervention. The individual who has experienced the trauma must be helped to integrate and process it so that some mastery is gained over the experience. Containment of intense feelings in a safe environment, support, and active listening are all important therapeutic interventions. Normalizing the experience through education and understanding is also helpful. Specific work on grief and loss may be extremely beneficial to patients who have experienced an overwhelming trauma.

GENERALIZED ANXIETY DISORDER

Generalized anxiety disorder was recognized as a disorder in 1987 with the publication of the revised third edition of DSM (DSM-III-R); however, it continues to be a controversial diagnostic category. When classified as the principal diagnosis in a patient, it is associated with the highest rate of comorbidity of any of the other anxiety disorders. Some critics prefer to conceptualize this disorder as a vulnerability located on Axis II. Nevertheless, with the addition of specific somatic symptoms, this disorder has a fairly high degree of construct validity. Moreover, psychodynamic clinicians commonly encounter patients who are chronic worriers. The lifetime prevalence of generalized anxiety disorder is thought to range from 4.1 to 6.6 percent. The condition is also one that tends to be chronic and to lead to a significant amount of disability and impairment in the quality of life. In one survey 27 percent of respondents were receiving disability payments, and only about 50 percent were able to work full time.

Most of the systematic studies on generalized anxiety disorder have involved behavioral or cognitive treatments, which have been shown to be helpful, or pharmacotherapy, which also is efficacious in reducing anxiety. However, there are a variety of limitations inherent in using benzodiazepines, including the fact that as many as 30 percent of patients do not respond to them. Other problems include the fact that benzodiazepines often do not affect the core symptom of worry; cause a variety of adverse effects; have a potential for abuse, physical dependence, and withdrawal; and have fairly high relapse rates when the medication is withdrawn.

Psychotherapeutic Techniques Whereas generalized anxiety disorder was formerly thought to be something of a “wastebasket” category into which all other forms of anxiety were placed, more recent research has begun to identify specific interpersonal issues and traumatic events connected with the diagnosis. In a survey study of over 1000 subjects, the individuals meeting criteria for generalized anxiety disorder had more frequent traumatic events in their history than those who did not meet the criteria. Subjects with this disorder reported trauma involving injury, illness, or death at a rate that was one and a half times greater than the reports from normal controls. Also, traumas related to assault, emotional events, and miscellaneous other traumas were four to six times more prevalent in subjects with generalized anxiety disorder than in those without this condition. Affected patients also tend to avoid thinking about the past events they consider traumatic; worrying appears to distract these patients with superficial matters that prevent them from worrying about more disturbing underlying concerns. In addition to this characteristic defensive pattern of avoidance, generalized anxiety disorder has also been linked to an insecure on conflicted attachment in childhood.

Because the object of the worry is often misleading, the psychodynamic clinician must embark on a collaborative search with the patient to discover the underlying conflicts and the true source of anxiety. As noted previously, the developmental hierarchy may help guide such a search, but the clinician must remain open to surprises rather than forcing the data to fit within a favorite theoretical formulation. Often the underlying conflicts surface in maladaptive and repetitive relationship patterns that are ultimately self-defeating. Table 15.5-6 illustrates the common psychodynamic themes in generalized anxiety disorder.

- Worrying serves a defensive function to avoid thinking about more disturbing issues.
- Increased prevalence of past trauma is highly characteristic.
- Link with an insecure/conflicted attachment in childhood.
- The underlying conflict that creates the anxiety can be related to any number of developmental themes.
- The unconscious conflict continues to be “alive” in self-defeating patterns in relationships.
- Resistance is common in moving below the level of symptoms to underlying sources of conflict.

Table 15.5-6 Psychodynamic Themes in Generalized Anxiety Disorder

When patients with generalized anxiety disorder begin psychotherapy they are likely to focus a great deal on symptoms, especially those related to the body. After listening to this catalog of symptoms, the therapist needs to move the patient further and inquire about work situation, family relationships, and any interpersonal difficulties encountered. Although the patient's goals may be geared to relief of somatic symptoms and anxiety, the therapist should encourage the patient to think in broader terms and to include broader patterns of adaptation to work and to love relationships. The therapist then tries to formulate the major conflicts that arise in interpersonal relationships, looking for this pattern as it recurs in the transference, in discussions of past relationships, and in reports of current relationships. As far as possible the therapist helps the patient to see how patterns of pathological relatedness invariably evoke more anxiety and worry.

Often these patterns can be linked to early traumas and difficulties attaching to others. The therapist demonstrates how anxiety is generated from these recurrent conflicts as well as how they can be mastered through the understanding of the patient's expectations of failure and the experience of trying new ways of relating to others. When attention is drawn away from the somatic focus of worrying, patients often feel that they have renewed energy to devote to the true underlying conflicts.

SUGGESTED CROSS-REFERENCES

Section 6.1 on Psychoanalysis provides a more detailed discussion of the psychoanalytic theory of anxiety and Section 30.1 on psychoanalysis and psychoanalytic psychotherapy provides more information on technique.

SECTION REFERENCES

Brom D, Kleber RJ, Defares PB: Brief psychotherapy for posttraumatic stress disorder. J Consult Clin Psychol *57:*607, 1989.

Bush FN, Shear MK, Cooper AM, Shapiro T, Leon AC: An empirical study of defense mechanisms in panic disorder. J Nerv Ment Dis *183:*299, 1995.

Calvocoressi L, Lewis B, Harris M, Trufan SJ, Goodman WK, McDougle CJ, Price LH: Family accomodation in obsessive-compulsive disorder. Am J Psychiatry *152:*441, 1995.

Compton A: An investigation of anxious thought in patients with DSM-IV agoraphobial panic disorder: rationale and design. J Am Psychoanal Assoc *46:*691, 1998.

Crits-Christoph P, Crits-Christoph K, Wolf-Palacio D, Fichter M, Rudick D: Brief supportive-expressive psychodynamic therapy for generalized anxiety disorder. In *Dynamic Therapies for Psychiatric Disorders (Axis I)*, JP Barber, P Crits-Christoph, editors. Basic Books, New York, 1995.

*Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

Gabbard GO: Psychoanalytically informed approaches to the treatment of obsessive-compulsive disorder. Psychoanal Inquiry (in press).

Gaston L: Dynamic therapy for post-traumatic stress disorder. In *Dynamic Therapies for Psychiatric Disorders (Axis I)*, JP Barber, P Crits-Christoph, editors. Basic Books, New York, 1995.

*Kagan J, Reznick JS, Snidman N: Biological bases of childhood shyness. Science *240:*167, 1988.

Lindy JD, Green BL, Grace M, Tichener J: Survivors of the Beverly Hills Supper Club fire. Am J Psychother *4:*593, 1983.

Marshall JR: The course and impact of panic disorder. J Clin Psychiatry *58:*36, 1997.

Milrod B: Unconscious pregnancy fantasies as an underlying dynamism in panic disorder. Journal of the American Psychoanalytic Association *46:*673, 1998.

*Milrod B, Busch F, Cooper A, Shapiro T: *Manual of Panic-Focused Psychodynamic Psychotherapy*. American Psychiatric Press, Washington, DC, 1997.

Neziroglu F, Anemone R, Yaryura-Tobias JA: Onset of obsessive-compulsive disorder in pregnancy. Am J Psychiatry *149:*947, 1992.

Rosenbaum JF, Biederman J, Bolduc EA, Hirshfeld DR, Faraone SV, Kagan J: Comorbidity of parental anxiety disorders as risk for childhood-onset anxiety in inhibited children. Am J Psychiatry *149:*475, 1992.

Schneier FR, Marshall RD, Street L, Heimberg RG, Juster HR: Social phobia and specific phobias. In *Treatments of Psychiatric Disorders*, ed 2, vol 1, GO Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

*Shear MK: Factors in the etiology and pathogenesis of panic disorder: Revisiting the attachment-separation paradigm. Am J Psychiatry *153*(Suppl):125, 1996.

Stein MB, Walker JR, Anderson G, Hazen AL, Ross CA, Eldredge G, Forde DR: Childhood physical and sexual abuse in patients with anxiety disorders and in a community sample. Am J Psychiatry *153:*275, 1996.

*van der Kolk BA, McFarlane AC, Weisaeth L, editors: *Traumatic Stress: The Effects of Overwhelming Experience on Mind, Body, and Society* . Guilford, New York, 1996.

van der Kolk BA, Pelcovitz D, Roth S, Mandel FS, McFarlane A, Herman JL: Dissociation, somatization, and affect dysregulation: The complexity of adaptation to trauma. Am J Psychiatry *153*(Suppl):83, 1996.

Wiborg IM, Dahl AA: Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? Arch Gen Psychiatry *53:*689, 1996.

Yehuda R, McFarlane AC, editors: *Psychobiology of Posttraumatic Stress Disorder*. New York Academy of Sciences, New York, 1997.

15.6 ANXIETY DISORDERS: CLINICAL FEATURES

DANIEL S. PINE, M.D.

The last 30 years have seen a dramatic increase in clinical research on a group of mental conditions labeled “anxiety disorders” in the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) nomenclature. While the term “anxiety” has been applied to diverse phenomena in the psychoanalytic, learning-based, and neurobiological literature, the term “anxiety” in the clinical psychopathological literature refers to the presence of fear or apprehension that is out of proportion to the context of the life situation. Hence, extreme fear or apprehension can be considered “clinical anxiety” if it is developmentally inappropriate (e.g., fear of separation in a 10-year-old child) or if it is inappropriate to an individual's life circumstances (e.g., worries about unemployment in a successful business executive). The last 30 years of clinical research has led to progressive refinement of the nosology for clinical anxiety disorders. While these disorders were broadly conceptualized in the early twentieth century, narrower definitions have arisen, partially stimulated by Donald Klein's observations on pharmacological distinctions between panic and nonpanic anxiety.

Consensus has emerged on the view of anxiety disorders as a family of related but distinct mental disorders. This consensus is reflected in the relatively minor changes to the broad categorization of anxiety disorders over the last 15 years, between the third edition of DSM (DSM-III) and the fourth edition (DSM-IV). Both DSM-IV and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) recognize similar groups of syndromes as discrete diagnostic entities. There is some disagreement, however, on whether all of these syndromes are most properly considered anxiety disorders. While DSM-IV considers a group of nine disorders to be the primary anxiety disorders, ICD-10 adopts a broader category of neurotic, stress-related, and somatoform disorders that includes each of the conditions in DSM-IV, as well as a number of disorders not considered anxiety disorders in DSM-IV. Prior to DSM-III, which brought a relatively major revision to the nosology of mental disorders in the United States, anxiety disorders were classified in a group of conditions that included many of the disorders currently listed in DSM-IV along with a set of disorders that have been reclassified. These reclassified disorders included affective disorders (formerly classified as “depressive neurosis”), somatoform and dissociative disorders (formerly classified as “hysterical neurosis”), and neurasthenia, a disorder that was eliminated with the writing of DSM-III.

DSM-III, with its emphasis on empiricism and the validity of nosological categories, reclassified anxiety disorders into categories that were quite similar to the disorders included in the current anxiety section of DSM-IV, which include panic disorder with and without agoraphobia, agoraphobia with and without panic disorder, specific phobia, social phobia, obsessive-compulsive disorder (OCD), posttraumatic stress disorder (PTSD), acute stress disorder, and generalized anxiety disorder. This chapter reviews the clinical features of these nine conditions, as conceptualized in DSM-IV. This includes the primary symptomatology, history, epidemiology, differential diagnosis, and course of each disorder, along with a clinical vignette designed to capture the essential features of each disorder as it typically presents in the clinic.

PANIC DISORDER AND AGORAPHOBIA

The panic attack is the key feature of panic disorder, which is characterized by a pattern of recurrent panic attacks. The *panic attack* is defined as an episode of abrupt intense fear that is accompanied by at least four autonomic or cognitive symptoms listed in Table 15.6-1. Such episodes of abrupt fear occur in many situations. For example, a healthy person might experience a panic attack when confronted with sudden extreme danger, and an individual with a phobia of heights might experience a panic attack when confronted with the feared situation.

Note: A panic attack is not a codable disorder. Code the specific diagnosis in which the panic attack occurs (e.g., panic disorder with agoraphobia).

A discrete period of intense fear or discomfort, in which four (or more) of the following symptoms developed abruptly and reached a peak within 10 minutes:

(1) palpitations, pounding heart, or accelerated heart rate
(2) sweating
(3) trembling or shaking
(4) sensations of shortness of breath or smothering
(5) feeling of choking
(6) chest pain or discomfort
(7) nausea or abdominal distress
(8) feeling dizzy, unsteady, lightheaded, or faint
(9) derealization (feelings of unreality) or depersonalization (being detached from oneself)
(10) fear of losing control or going crazy
(11) fear of dying
(12) paresthesias (numbness or tingling sensations)
(13) chills or hot flushes

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 15.6-1 DSM-IV Criteria for Panic Attack

DSM-IV recognizes three types of panic attacks. The unexpected or spontaneous panic attack occurs without cue or warning; the situationally bound panic attack occurs upon exposure to, or in anticipation of, exposure to a feared stimulus; and the situationally predisposed panic attack is more likely to occur, but does not necessarily have to occur, on exposure to a situational trigger. In panic disorder, panic attacks occur spontaneously, arising without any trigger or environmental cue. As shown in Table 15.6-2, panic disorder requires at least two spontaneous panic attacks, at least one of which is associated with either concern with additional attacks, worry about attacks, or changes in behavior.

Note: Agoraphobia is not a codable disorder. Code the specific disorder in which the agoraphobia occurs (e.g., panic disorder with agoraphobia or agoraphobia without history of panic disorder).

A. Anxiety about being in places or situations from which escape might be difficult (or embarrassing) or in which help may not be available in the event of having an unexpected or situationally predisposed panic attack or panic-like symptoms. Agoraphobic fears typically involve characteristic clusters of situations that include being outside the home alone; being in a crowd or standing in a line; being on a bridge; and traveling in a bus, train, or automobile.

Note: Consider the diagnosis of specific phobia if the avoidance is limited to one or only a few specific situations, or social phobia if the avoidance is limited to social situations.

B. The situations are avoided (e.g., travel is restricted) or else are endured with marked distress or with anxiety about having a panic attack or panic-like symptoms, or require the presence of a companion.

C. The anxiety or phobic avoidance is not better accounted for by another mental disorder, such as social phobia (e.g., avoidance limited to social situations because of fear of embarrassment), specific phobia (e.g., avoidance limited to a single situation like elevators), obsessive-compulsive disorder (e.g., avoidance of dirt in someone with an obsession about contamination), posttraumatic stress disorder (e.g., avoidance of stimuli associated with a severe stressor), or separation anxiety disorder (e.g., avoidance of leaving home or relatives).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 15.6-2 DSM-IV Criteria for Agoraphobia

Patients with panic disorder present with a number of comorbid conditions, but there has been considerable interest in the relationship between panic disorder and agoraphobia, which refers to fear or anxiety of places from where escape might be difficult (Table 15.6-2). There has in fact been some debate about whether agoraphobia is best conceptualized as a complication of panic disorder or as a separate condition. This controversy centers on the frequency with which patients develop agoraphobia in the absence of panic disorder or panic attacks (Table 15.6-3) DSM-IV suggests that such patients do exist, noting the existence of panic disorder with agoraphobia (Table 15.6-4) and without agoraphobia (Table 15.6-5). However, even in agoraphobia without history of panic disorder, agoraphobia is considered related to the fear of developing paniclike symptoms.

A. The presence of agoraphobia related to fear of developing panic-like symptoms (e.g., dizziness or diarrhea).
B. Criteria have never been met for panic disorder.
C. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.
D. If an associated general medical condition is present, the fear described in criterion A is clearly in excess of that usually associated with the condition.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 15.6-3 DSM-IV Diagnostic Criteria for Agoraphobia Without History of Panic Disorder

A. Both (1) and (2):
   (1) recurrent unexpected panic attacks
   (2) at least one of the attacks has been followed by 1 month (or more) of the following:
      (a) persistent concern about having additional attacks
      (b) worry about the implications of the attack or its consequences (e.g., losing control, having a heart attack, "going crazy")
      (c) a significant change in behavior related to the attacks
B. The presence of agoraphobia.
C. The panic attacks are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism).
D. The panic attacks are not better accounted for by another mental disorder, such as social phobia (e.g., occurring on exposure to feared social situations), specific phobia (e.g., on exposure to a specific phobic situation), obsessive-compulsive disorder (e.g., on exposure to dirt in someone with an obsession about contamination), posttraumatic stress disorder (e.g., in response to stimuli associated with a severe stressor), or separation anxiety disorder (e.g., in response to being away from home or close relatives).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 15.6-4 DSM-IV Diagnostic Criteria for Panic Disorder With Agoraphobia

A. Both (1) and (2):
   (1) recurrent unexpected panic attacks
   (2) at least one of the attacks has been followed by at least 1 month (or more) of the following:
      (a) persistent concern about having additional attacks
      (b) worry about the implications of the attack or its consequences (e.g., losing control, having a heart attack, "going crazy")
      (c) a significant change in behavior related to the attacks
B. Absence of agoraphobia.
C. The panic attacks are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism).
D. The panic attacks are not better accounted for by another mental disorder, such as social phobia (e.g., occurring on exposure to feared social situations), specific phobia (e.g., on exposure to a specific phobic situation), obsessive-compulsive disorder (e.g., on exposure to dirt in someone with an obsession about contamination), posttraumatic stress disorder (e.g., in response to stimuli associated with a severe stressor), or separation anxiety disorder (e.g., in response to being away from home or close relatives).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington DC, 1994.

Table 15.6-5 DSM-IV Diagnostic Criteria for Panic Disorder Without Agoraphobia

As with most anxiety disorders, panic disorder often co-occurs with a number of mental conditions beyond agoraphobia, particularly anxiety and depressive disorders. These include specific and social phobia, generalized anxiety disorder, and major depressive disorder. Some data also suggest associations with substance use disorders, bipolar disorder, and suicidal behavior. While the high comorbidity seen in the clinic partially reflects referral bias, considerable comorbidity with these anxiety and depressive disorders is also found in epidemiological studies, which suggests that panic disorder in the community is frequently compounded by comorbid mental conditions.

History and Comparative Nosology While the term “panic disorder” was first coined in DSM-III, a syndrome characterized by recurrent episodes of spontaneous fear has been recognized for more than 100 years. This syndrome has been given various labels throughout history, including DaCosta's syndrome in the late nineteenth century, and effort syndrome or neurocirculatory asthenia in the earlier part of the twentieth century. Even Freud's descriptions of “anxiety neurosis” invoked many of the key features of panic disorder.

A major change to the DSM conceptualization of anxiety occurred in 1980, with SM-III, where panic disorder was first recognized as a distinct entity. Between 1980 and 1994 one significant change to the conceptualization of the disorder involved refining the view of panic disorder and agoraphobia as tightly linked constructs. As conceptualized in DSM-IV, agoraphobia invariably involves at least some form of spontaneous crescendo anxiety, even if such episodes do not meet formal criteria for panic attacks (Table 15.6-3 and Table 15.6-4). In the earlier versions of the DSM and in ICD-10, agoraphobia is considered less closely linked to panic disorder. Indeed, ICD-10 classifies agoraphobia as one of many phobic disorders and does not emphasize the relationship with panic disorder to the same degree as DSM-IV. Table 15.6-6 presents the ICD-10 diagnostic criteria for phobic anxiety disorders, including agoraphobia. Table 15.6-7 presents the ICD-10 criteria for other anxiety disorders, including panic disorder.

Table 15.6-6 ICD-10 Diagnostic Criteria for Phobic Anxiety Disorders

Table 15.6-7 ICD-10 Diagnostic Criteria for Other Anxiety Disorders

Differential Diagnosis Panic disorder with or without agoraphobia must be differentiated from a number of medical conditions that produce similar symptomatology. Panic attacks are associated with a variety of endocrinological disorders, including both hypo- and hyperthyroid states, hyperparathyroidism, and pheochromocytomas. Episodic hypoglycemia associated with insulinomas can also produce paniclike states, as can primary neuropathological processes. These include seizure disorders, vestibular dysfunction, neoplasms, or the effects of both prescribed and illicit substances on the central nervous system. Finally, disorders of the cardiac and pulmonary systems, including arrhythmias, chronic obstructive pulmonary disease, and asthma, can produce autonomic symptoms and accompanying crescendo anxiety that can be difficult to distinguish from panic disorder. Clues of an underlying medical cause for paniclike symptoms include atypical features during panic attacks, such as ataxia, alterations in consciousness, or bladder dyscontrol; onset of panic disorder relatively late in life; or physical signs or symptoms indicating a medical disorder.

Panic disorder also must be differentiated from a number of psychiatric disorders, particularly other anxiety disorders. Since panic attacks occur in many anxiety disorders, including social and specific phobia, posttraumatic stress disorders, or even obsessive-compulsive disorder, the key to correctly diagnosing panic disorder involves documenting recurrent spontaneous panic attacks at some point in the illness. Differentiation with generalized anxiety disorder can also be difficult. Classically, panic attacks are characterized by their rapid onset, within minutes, and short duration, usually less than 10 to 15 minutes, in contrast to the anxiety associated with generalized anxiety disorder, which emerges and dissipates more slowly. This distinction can be difficult, however, as the anxiety surrounding panic attacks can be more diffuse and dissipate more slowly. Since anxiety is a frequent concomitant of many other psychiatric disorders, including the psychoses and affective disorders, distinctions between panic disorder and a multitude of disorders can also be difficult.

Course Panic disorder typically has its onset in late adolescence or early adulthood, although cases of childhood-onset and late adulthood–onset disorder have been described. Only tentative data exist on the natural course of panic disorder. The best evidence on the course of any disorder, including panic disorder, derives from prospective epidemiological research, since both retrospective and clinically based studies are vulnerable to biases that preclude firm conclusions on course. Unfortunately, few such studies exist. Retrospective or clinical studies suggest that panic disorder tends to exhibit a fluctuating course, with varying levels of persistence over the life span. Approximately one third to one half of patients are psychiatrically healthy at follow-up, with most living relatively normal lives, despite either fluctuating or recurrent symptoms. Typically, patients with chronic disorders exhibit a pattern of exacerbation and remissions rather than chronic disability.

Ms. S. was a 25-year-old student who was referred for a psychiatric evaluation from the medical emergency room at a larger university-based medical center. Ms. S. had been evaluated three times over the preceding 3 weeks in this emergency room. Her first visit was prompted by a paroxysm of extreme dyspnea and terror that occurred while she was working on a term paper. The dyspnea was accompanied by palpitations, choking sensations, sweating, shakiness, and a strong urge to flee. Ms. S. thought that she was having a heart attack, and she immediately went to the emergency room. She received a full medical evaluation, including an electrocardiogram (ECG) and routine blood work, which revealed no sign of cardiovascular, pulmonary, or other illness. Although Ms. S. was given the number of a local psychiatrist, she did not make a follow-up appointment, since she did not think that her episode would recur. She developed two other similar episodes, one while she was on her way to visit a friend and a second that woke her up from sleep. She immediately went to the emergency room after experiencing both paroxysms, receiving full medical workups that showed no sign of illness.

SPECIFIC AND SOCIAL PHOBIAS

The term “phobia” refers to an excessive fear of a specific object, circumstance, or situation. Phobias are classified on the basis of the feared object or situation, and DSM-IV recognizes three distinct classes of phobia: agoraphobia (discussed above as it is considered closely related to panic disorder), specific phobia, and social phobia. Criteria for specific and social phobia are shown in Table 15.6-8 and Table 15.6-9. Both specific and social phobia require the development of intense anxiety, to the point of even situationally bound panic, upon exposure to the feared object or situation. Both conditions also require that fear either interferes with functioning or causes marked distress. Finally, both conditions require that an individual recognizes the fear as excessive or irrational and that the feared object or situation is either avoided or endured with great difficulty.

Table 15.6-8 DSM-IV Diagnostic Criteria for Specific Phobia

Table 15.6-9 DSM-IV Diagnostic Criteria for Social Phobia

Specific phobia is divided into four subtypes (animal type, natural environment type, blood-injection-injury type, situational type, and other type), with a residual category for phobias that do not clearly fall into any of these four categories. The key feature in each type of phobia is that the fear is circumscribed to a specific object, both temporally and with respect to other objects. Hence, an individual with specific phobia becomes immediately frightened when presented with a feared object. This fear may relate to concern about harm from a feared object, concern about embarrassment, or fear of consequences related to exposure to the feared object. For example, individuals with blood-injury phobia may be afraid of fainting on exposure to blood, and individuals with fear of heights may be afraid of becoming dizzy.

Specific phobia often involves fears of more than one object, particularly within a specific subcategory of phobia. For example, it is common for an individual with a phobia of thunderstorms to also have a phobia of water, both phobias being classified as natural environment type phobias. Further, in the clinical setting, specific phobia often occurs with other anxiety or mood disorders. Since it is rare for patients to seek treatment for an isolated phobia, some of the comorbidity seen in the clinic reflects referral bias. Community-based studies also suggest that specific phobia is associated with other anxiety disorders, although at lower rates than seen in the clinic. Quantifying the impairment associated with a specific phobia is sometimes difficult, since the comorbid disorders typically tend to cause more impairment than specific phobia and since individuals with isolated specific phobia are rarely seen in the clinic. Impairment associated with specific phobia typically restricts the social or professional activities of the individual.

Social phobia involves fear of social situations, including situations that involve scrutiny or contact with strangers. Individuals with social phobia typically fear embarrassing themselves in social situations (e.g., social gatherings, oral presentations, or meeting new people). This can involve specific fears about performing certain activities, such as writing, eating, or speaking in front of others. It can also involve a vague, nonspecific fear of embarrassing oneself. DSM-IV provides a specifier for the diagnosis of social phobia. Individuals with social phobia who fear most situations are considered to suffer from generalized social phobia. Such individuals are fearful of initiating conversations in many situations, about dating or participating in most group activities or social gatherings, and about speaking with authority figures.

The clinician should recognize that many patients exhibit at least some social anxiety or self-consciousness. In fact, community studies suggest that roughly a third of all people consider themselves to be far more anxious than other people in social situations. Such anxiety only becomes social phobia when the anxiety either prevents an individual from participating in desired activities or causes marked distress in such activities. Individuals with the more specific form of social phobia possess fear of specific, circumscribed social situations. For example, extreme anxiety about public speaking that interferes with an individual's job performance is a common type of specific social phobia; it would not be considered generalized social phobia unless it was associated with fears related to many other social situations besides public speaking.

As with other anxiety disorders, social phobia frequently co-occurs with other mood and anxiety disorders. The association of social phobia with both panic disorder and major depression has received considerable attention in recent literature. Associations with substance use disorders and childhood conduct problems have also been documented.

History and Comparative Nosology Phobias have been recognized as incapacitating mental disorders for more than 100 years. The prominent place of phobia in the history of modern mental health science is indicated by the major role case histories of phobic patients played in the development of both psychoanalytic and cognitive therapies. The category of phobia has undergone progressive refinement over the past 20 years, as research has focused on each of the specific classes of phobia described above. Much of this refinement crystallized in DSM-III, which was based on emerging evidence that phobias represent a group of related but distinct conditions, rather than one heterogenous disorder. Such evidence included Isaac Mark's work on differentiating social and specific phobias. The refinement in DSM-III categorized agoraphobia as a condition closely related to panic disorder and distinguished social and simple phobia, which was relabeled specific phobia in DSM-IV.

The view of phobias has changed since the writing of DSM-III. While discussion of agoraphobia has emphasized the role of panic since DSM-III, DSM-IV also contains descriptions of paniclike phenomena in both the specific and social phobia sections, as well as in the discussion of agoraphobia. Beyond this change, the most significant other change for specific phobia between DSM-III and DSM-IV involved inclusion of the above subcategories of phobia, based on research noting distinct physiology and demographics of the subtypes. For social phobia, the most significant other change occurred with the revised third edition (DSM-III-R), which distinguished between generalized and more specific forms of social phobia. This change was based on descriptive phenomenology, epidemiology, and pharmacology studies that validated the two variants of the condition.

The approach to categorization of phobias in the ICD system is quite similar to that in DSM-IV. ICD-10 recognizes specific phobia as a distinct category, including the subtypes in DSM-IV. Social phobia is also classified in ICD-10, although without the qualifier in DSM-IV. Perhaps the major difference between DSM-IV and ICD-10 in the consideration of phobia relates to agoraphobia. While DSM-IV emphasizes the relation between panic disorder and agoraphobia, in ICD-10 the term “panic disorder” is restricted to cases without phobia, and the term “agoraphobia” is applied to all cases that meet criteria, regardless of the presence of absence of panic attacks (Table 15.6-6 and Table 15.6-7).

Differential Diagnosis Specific phobia is usually quite easily distinguished from anxiety stemming from primary medical problems by the focused nature of the anxiety, which is not typical of anxiety disorders related to medical problems. The most difficult diagnostic issues involve differentiating specific phobia from other anxiety disorder. As DSM-IV emphasizes the presence of paniclike symptoms with specific phobia, including situationally bound panic attacks, specific phobia must be differentiated from panic disorder, in which panic attacks occur without a cue. Specific phobia can occasionally be confused with PTSD, as both conditions can involve focused fears of specific objects or situations. The two disorders are most easily differentiated by the marked other features of PTSD, such as reexperiencing the trauma, avoidance, and enhanced startle, which are absent in specific phobia. Similarly, specific phobia can occasionally be confused with generalized anxiety disorder, as both conditions may involve worry about exposure to specific situations. The two disorders are differentiated on the basis of the focused nature of the fear, both over time and with respect to objects, in specific phobia.

Like specific phobia, social phobia is rarely confused with anxiety that is the primary result of medical disorders. However, the number of psychiatric disorders that are associated with social withdrawal make it difficult to diagnose social phobia correctly. Perhaps the most difficult distinction involves differentiating social phobia and agoraphobia, since both conditions involve fears of situations where people typically gather. The key distinction between the disorders centers on the nature of the feared object. Patients with social phobia are specifically afraid of encountering people; individuals with agoraphobia are afraid of situations from which escape would be difficult but do not specifically fear people. Hence, an individual with agoraphobia might be reassured in the presence of other people, provided the physical properties of the location are suitable, while an individual with social phobia flees other people. The clinician might also encounter difficulty in distinguishing social phobia from the social isolation that accompanies a number of psychiatric disorders, including major depression and the early stages of psychosis. Two factors are essential in making this distinction. First, the individual with social phobia must experience anxiety or fear in social situations; individuals who are isolated due to depression or indolent psychosis often isolate themselves for other reasons. Second, with social phobia, symptomatology is restricted to the fears of social situations; with other disorders social isolation is accompanied by a broad array of symptoms that are not found in social phobia.

Course Specific phobia exhibits a bimodal age of onset, with a childhood peak for animal phobia, natural environment phobia, and blood-injury phobia and an early adulthood peak for other phobias, such as situational phobia. As with other anxiety disorders, limited prospective epidemiological data exist on the natural course of specific phobia. Because patients with isolated specific phobias rarely present for treatment, research on the course of the disorder in the clinic is limited. The information that is available suggests that most specific phobias that begin in childhood and persist into adulthood continue to persist for many years. The severity of the condition is thought to remain relatively constant, without the waxing and waning course seen with other anxiety disorders.

> Mr. A. was a successful businessman who presented for treatment following a change in his business schedule. While he had formerly worked largely from an office near his home, a promotion led to a schedule of frequent out-of-town meetings, requiring weekly flights. Mr. A. reported being “deathly afraid” of flying. Even the thought of getting on an airplane led to thoughts of impending doom as he envisioned his airplane crashing to the ground. These thoughts were associated with intense fear, palpitations, sweating, clammy feelings, and stomach upset. While the thought of flying was terrifying enough, Mr. A. became nearly incapacitated when he went to the airport. Immediately before boarding, Mr. A. often had to turn back from the plane and run to the bathroom to vomit.

Social phobia tends to have its onset in late childhood or early adolescence. Social phobia tends to be a chronic disorder, although as with the other anxiety disorders, prospective epidemiological data are limited. Both retrospective epidemiological studies and prospective clinical studies suggest that the disorder can profoundly disrupt the life of an individual over many years. This can include disruption in school or academic achievement, interference with job performance, and

social development.

> Ms. M. was a successful secretary working in a law firm. While she reported a long history of feeling uncomfortable in social situations, Ms. M. came for treatment when she began to feel that her uneasiness was interfering with her social life and job performance. Ms. M. reported that she noticed herself feeling increasingly nervous whenever she met a new person. For example, upon meeting a new member of the law firm, she described feeling suddenly tense and sweaty, noticing that her heart was beating very fast. She had the sudden thought that she would say something foolish in these situations or commit a terrible social gaffe that would make people laugh at her. At social gatherings she described similar feelings that led her to either leave the gathering early or decline invitations to attend.

OBSESSIVE-COMPULSIVE DISORDER

Obsessions and compulsions are the essential features of OCD. As shown in Table 15.6-10, an individual must exhibit either obsessions or compulsions to meet DSM-IV criteria. DSM-IV recognizes *obsessions* as "persistent ideas, thoughts, impulses, or images that are experienced as intrusive and inappropriate," causing distress. Obsessions are anxiety provoking, accounting for the categorization of OCD as an anxiety disorder; they must be differentiated from excessive worries about real life problems; and they must be associated with either efforts to ignore or suppress the obsessions. Typical obsessions associated with OCD include thoughts about contamination ("my hands are dirty") or doubts ("I forgot to turn off the stove").

Table 15.6-10 DSM-IV Diagnostic Criteria for Obsessive-Compulsive Disorder

Compulsions are defined as repetitive acts, behaviors, or thoughts that are designed to counteract the anxiety associated with an obsession. A compulsion reduces the anxiety associated with the obsession. While many compulsions are acts associated with specific obsessions, such as hand-washing or checking, other compulsions can be thoughts. For example, patients may have the obsession that they have committed a sin and might relieve the resultant anxiety by repetitively saying a silent prayer to themselves.

Obsessions and compulsions must cause marked distress, consume at least 1 hour a day, or interfere with functioning to be considered above the diagnostic threshold. At least during some point in the illness, symptoms of OCD must be recognized as unreasonable for adults, though this varies greatly both across individuals and in a given individual over time. For example, early in the course of the disorder patients may recognize their hand washing is excessive or irrational but, over a number of years, lose this recognition. DSM-IV recognizes a poor insight subtype of OCD when an individual fails to recognize the irrational or unreasonable nature of the obsessions. This subtype of OCD has been labeled the psychotic subtype in some of the clinical literature, prompting trials of antipsychotic pharmacotherapy. The criterion related to insight does not apply to children, who may either not possess the insight to recognize the unreasonableness of their condition or may be too embarrassed to discuss the condition as unreasonable.

OCD frequently co-occurs with other disorders. The association with major depression is particularly prominent, although comorbidity with panic disorder, phobias, and eating disorders is also not uncommon. Finally, OCD exhibits a particularly interesting association with Tourette's disorder. Approximately half of all patients with Tourette's disorder meet criteria for OCD, although less than 10 percent of patients with OCD meet criteria for Tourette's disorder. There is also evidence of cotransmission of Tourette's disorder, OCD, and chronic motor tics within families.

History and Comparative Nosology Descriptions of patients with a syndrome of recurrent obsessions and compulsions are found in the nineteenth century, when these conditions were viewed as a form of "depressive state." Descriptions of OCD also played a prominent role in Freud's writings, as captured in the case history of the Rat Man, and in early learning-based theories that attempted to apply treatments developed among patients with phobias to patients with OCD. Research on OCD changed with the ECA study in the early 1980s. Prior to this study, OCD was recognized as a discrete but rare entity, stimulating a modest degree of research. The Epidemiologic Catchment Area (ECA) Study noted that OCD had a prevalence in excess of 1 percent in the population and was associated with marked impairment. This stimulated extensive research on all aspects of OCD, including its phenomenology.

The major change in OCD from DSM-III to DSM-IV involved the conceptualization of compulsions. While DSM-III-R viewed compulsions as behaviors, DSM-IV recognized compulsions as either behaviors or mental acts designed to reduce the anxiety-provoking nature of the obsession. The conceptualization of OCD in the ICD and DSM systems is generally similar with a few exceptions in the emphasis on specific features of the condition. For example, ICD-10 emphasizes that a compulsive act must not be pleasurable. ICD-10 also stipulates that obsessions or compulsions must be present most days for 2 weeks, a requirement not included in DSM-IV, and ICD-10 does not stipulate the amount of time a patient must spend on compulsions. Perhaps the major difference between the DSM-IV and ICD-10 view of the disorder relates to the categorization of the disorder with respect to other anxiety disorders. DSM-IV recognizes OCD as one of the nine anxiety disorders discussed in the current chapter. There has been some debate in the United States and Europe about whether OCD is more properly classified in a distinct category. ICD-10 has adopted such a scheme, using OCD to designate a group of syndromes considered distinct from anxiety disorders (Table 15.6-11).

Table 15.6-11 ICD-10 Diagnostic Criteria for Obsessive-Compulsive Disorder

Differential Diagnosis A number of primary medical disorders can produce syndromes bearing a striking resemblance to OCD. In fact, the current conceptualization of OCD as a disorder of the basal ganglia derives from the phenomenological similarity between idiopathic OCD and OCD-like disorders that are associated with basal ganglia diseases, such as Sydenham's chorea and Huntington's disease. Hence, neurological signs of such basal ganglia pathology must be assessed when considering the diagnosis of OCD in a patient presenting for psychiatric treatment. OCD frequently develops before age 30, and new-onset OCD in an older individual should raise questions about potential neurological contributions to the disorder. Finally, some evidence suggests an association between an immune reaction to streptococcal infections and either initial manifestations or dramatic exacerbation of OCD in children. This syndrome appears to emerge relatively acutely, in contrast

to the more indolent onset of other cases of childhood OCD. Hence, in children with such presentations, the role of such an infectious process should be considered.

Obsessive-compulsive behavior is also found in a host of other psychiatric disorders, and the clinician must also rule out these conditions when diagnosing OCD. OCD exhibits a superficial resemblance to obsessive-compulsive personality disorder, which is associated with an obsessive concern for details, perfectionism, and similar personality traits. The conditions are easily distinguished by the fact that only OCD is associated with a true syndrome of obsessions and compulsions, as described above.

Psychotic symptoms often lead to obsessive thoughts and compulsive behavior that can be difficult to distinguish from OCD with poor insight, in which obsessions border on psychosis. The key to distinguishing OCD from psychosis involves the fact that patients with OCD can almost always acknowledge the unreasonable nature of their symptoms and the fact that psychotic illnesses are typically associated with a host of other features that are not characteristic of OCD. Similarly, OCD can be difficult to differentiate from depression since it is often associated with comorbid major depression and since major depression is often associated with obsessive thoughts that can at times border on true obsessions, as conceived in OCD. The two conditions are best differentiated by their course. Obsessive symptoms associated with depression are only found in a depressive episode; true OCD persists despite remission of depression.

Finally, OCD is closely related to Tourette's disorder; the two conditions frequently co-occur, both in an individual over time and within families. In its classic form, Tourette's disorder is associated with a pattern of recurrent vocal and motor tics that bears only a slight resemblance to OCD. However, the premonitory urges that precede tics often bear a striking resemblance to obsessions, and many of the more complicated motor tics can bear a close resemblance to compulsions.

Course OCD typically begins in late adolescence, though onset in childhood is not uncommon. The disorder tends to exhibit a waxing and waning course over the life span, with periods of relative good functioning and limited symptoms punctuated by periods of symptomatic exacerbation. A small minority of subjects exhibit either complete remission of their disorder or a progressive, deteriorating course.

Ms. B. presented for psychiatric admission after being transferred from a medical floor where she had been treated for malnutrition. Ms. B. had been found unconscious in her apartment by a neighbor. When brought to the emergency room by ambulance, she was found to be hypotensive and hypokalemic. At psychiatric admission, Ms. B. described a long history of recurrent obsessions about cleanliness, particularly related to food items. She reported that it was difficult for her to eat any food unless it had been washed by her three to four times, since she often thought that a food item was dirty. She reported that washing her food decreased the anxiety she felt about the dirtiness of food. While Ms. B. reported that she occasionally tried to eat food that she did not wash (e.g., in a restaurant), she became so worried about contracting an illness from eating such food that she could no longer dine in restaurants. Ms. B. reported that her obsessions about the cleanliness of food had become so extreme over the past 3 months that she could eat very few foods, even if she washed them excessively. She recognized the irrational nature of these obsessive concerns, but either could not bring herself to eat or became extremely nervous and nauseous after eating.

POSTTRAUMATIC STRESS AND ACUTE STRESS DISORDERS

Symptomatology Both PTSD and acute stress disorder are characterized by the onset of psychiatric symptoms immediately following exposure to a traumatic event. As noted in Table 15.6-12, DSM-IV specifies that the traumatic event involves either witnessing or experiencing threatened death or injury or witnessing or experiencing threat to physical integrity. Further, the response to the traumatic event must involve intense fear or horror. Such traumatic experiences might include being involved in or witnessing a violent accident or crime, military combat, assault, being kidnapped, being involved in natural disasters, being diagnosed with a life-threatening illness, or experiencing systematic physical or sexual abuse. Both PTSD and acute stress disorder also require characteristic symptoms following such trauma. There is evidence of a relation between the degree of trauma and the likelihood of symptoms. The proximity to, and intensity of, the trauma relate to the probability of developing symptomatology.

Table 15.6-12 DSM-IV Diagnostic Criteria for Posttraumatic Stress Disorder

In PTSD, the individual develops symptoms in three domains: reexperiencing the trauma, avoiding stimuli associated with the trauma, and experiencing symptoms of increased autonomic arousal, such as an enhanced startle. Flashbacks represent the classic form of reexperiencing: the individual may act and feel as if the trauma were recurring. Other forms of reexperiencing include distressing recollections or dreams and either physiological or psychological stress reactions upon exposure to stimuli linked to the trauma. An individual must exhibit at least one symptom of reexperiencing to meet criteria for PTSD. Symptoms of avoidance associated with PTSD include efforts to avoid thoughts or activities related to the trauma, anhedonia, reduced capacity to remember events related to the trauma, blunted affect, feelings of detachment or derealization, and a sense of a foreshortened future. An individual must exhibit at least three such symptoms. Symptoms of increased arousal include insomnia, irritability, hypervigilance, and exaggerated startle. An individual must exhibit at least two such symptoms. Finally, the diagnosis of PTSD is only made when these symptoms persist for at least 1 month; the diagnosis of acute stress disorder is made in the interim. DSM-IV acknowledges three subtypes of PTSD that differentiate syndromes with varying time courses. *Acute PTSD* refers to an episode that lasts less than 3 months. *Chronic PTSD* refers to an episode lasting 3 months or longer. *PTSD with delayed onset* refers to an episode that develops 6 months or more after exposure to the traumatic event

The diagnosis of acute stress disorder is applied to syndromes that resemble PTSD but last less than 1 month after a trauma. Acute stress disorder is characterized by reexperiencing, avoidance, and increased arousal, much like PTSD. Acute stress disorder is also associated with at least three of the dissociative symptoms listed in Table 15.6-13.

Table 15.6-13 DSM-IV Diagnostic Criteria for Acute Stress Disorder

Because individuals often exhibit complex biological and behavioral responses to extreme trauma, the clinician must recognize other medical and psychiatric conditions in the traumatized patient. Particularly after traumatic events that involve physical injury, the clinician must always consider neurological causes of symptoms that develop after trauma. Traumatized patients also can develop mood disorders, including dysthymia and major depression, as well as other anxiety disorders, such as generalized anxiety disorder or panic disorder, and substance use disorders. Finally, recent research suggests that some psychiatric features of posttraumatic syndromes can relate to the state of a patient before the trauma. For example, patients with premorbid anxiety or affective syndromes may be more likely to develop posttraumatic symptoms than individuals who are free of mental illness before the trauma. Thus, the clinician should consider the premorbid mental state of the traumatized patient to enhance understanding of symptoms that develop following a traumatic event.

History and Comparative Nosology Astute clinicians have recognized the juxtaposition of acute mental syndromes to traumatic events for more than 200 years. Observations of trauma-related syndromes were documented following the Civil War, and early psychoanalytic writers, including Freud, noted the relation between neurosis and trauma. Considerable interest in posttraumatic mental disorders was stimulated by observations of “battle fatigue,” “shell shock,” and “soldier's heart” in World War I and World War II. Moreover, increasing documentation of mental reactions to the Holocaust, to a series of natural disasters, and to assault contributed to the growing recognition of a close relationship between trauma and psychopathology.

The syndrome of PTSD was first recognized in the DSM nosology in DSM-III in 1980; acute stress disorder was first recognized in DSM-IV in 1994. The recognition of acute stress disorder followed observations suggesting that many individuals exhibit mental syndromes immediately following trauma; such individuals might face an elevated risk for PTSD. The original DSM-III definition of PTSD required only one symptom of reexperiencing, two symptoms of “psychic numbing,” and one symptom from a list of miscellaneous items, with no duration criteria. DSM-III-R added a number of symptoms to the DSM-III definition and removed the DSM-III symptom of guilt. DSM-III-R also adopted the symptom groupings found in DSM-IV, where symptoms are classified as manifestations of either reexperiencing, avoidance, or hyperarousal. The major change to the definition in DSM-IV involved the definition of trauma. While DSM-III-R emphasized trauma as an event that was “outside of normal experience,” a number of field studies suggested that the typical traumatic precipitants of PTSD were relatively common events. As a result, DSM-IV emphasizes the threat and fear-provoking nature of a trauma, without reference to “normal experience.”

Some variation exists in the definitions of PTSD and acute stress disorder in DSM-IV and ICD-10. While ICD-10 acknowledges the same core group of symptoms as DSM-IV for PTSD, including exposure to a trauma, reexperiencing, avoidance, and increased arousal, ICD-10 provides considerably less detail than DSM-IV for each of the criteria. For example, unlike DSM-IV, ICD-10 provides only brief examples of reexperiencing or avoidance symptoms. The broader view of PTSD and acute stress disorder also differs in the DSM and ICD systems. As with OCD, ICD-10 groups PTSD and acute stress reaction in a distinct category rather than including them with other anxiety disorders (Table 15.6-14).

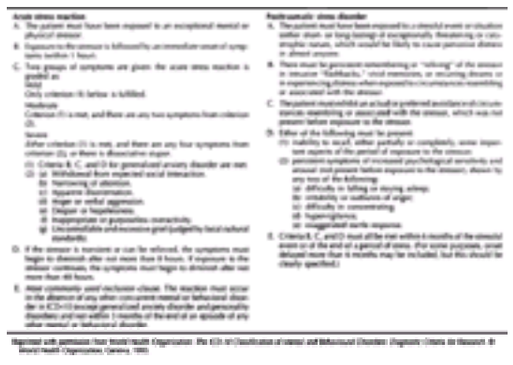

Table 15.6-14 ICD-10 Diagnostic Criteria for Reactions to Severe Stress

Differential Diagnosis Because patients often exhibit complex reactions to trauma, the clinician must be careful to exclude syndromes other than PTSD and acute stress disorder when evaluating patients presenting in the wake of trauma. Recognizing potentially treatable medical contributors to posttraumatic symptomatology is particularly important. For example, neurological injury following head trauma can contribute to the clinical picture, as can psychoactive substance use disorders or withdrawal syndromes, either in the period immediately surrounding the trauma or many weeks after the trauma. Medical contributors can usually be detected through a careful history and physical examination, if the clinician remembers to consider such factors.

Symptoms of PTSD can be difficult to distinguish from those of either panic disorder or generalized anxiety disorder, as all three syndromes are associated with prominent anxiety and autonomic arousal. Keys to correctly diagnosing PTSD involve a careful review of the time course relating the symptoms to a traumatic event. Further, PTSD is associated with reexperiencing and avoidance of a trauma, features typically not present in panic or generalized anxiety disorder. Major depression is a frequent concomitant of PTSD. While the two syndromes are generally readily distinguishable phenomenologically, comorbid depression should be noted because it may affect treatment of PTSD. Finally, PTSD must be differentiated from a series of related disorders that can exhibit phenomenological similarities, including borderline personality disorder, dissociative disorders, and factitious disorders.

Course Much recent research on the course of psychological reactions to trauma focused on the time course of symptoms immediately following a trauma. The likelihood of developing symptoms, the severity of such symptoms, and the duration of the symptoms are each proportional to the proximity, duration, and intensity of the trauma. Many individuals develop acute stress reactions when faced with close, persistent, intense trauma. Moreover, many individuals who develop PTSD exhibit features of the acute stress syndrome prior to developing PTSD, although many individuals with acute stress syndromes do not develop PTSD. Finally, the full syndrome of PTSD also exhibits a variable course, with some evidence that this also relates to the nature of the trauma. A large minority of patients experience complete remissions, while another large group exhibits only mild symptoms. Approximately 10 percent of patients with PTSD exhibit a persistent or chronic course to their disorder.

Mr. F. sought treatment for symptoms that he developed in the wake of an automobile accident that had occurred about 6 weeks prior to his psychiatric evaluation. While driving to work on a mid-January morning, Mr. F. lost control of his car on an icy road. His car swerved out of control into oncoming traffic in another lane, collided with another car, and then hit a nearby pedestrian. Mr. F. was trapped in his car for 3 hours while rescue workers cut the door of his car. Upon referral, Mr. F. reported frequent intrusive thoughts about the accident, including nightmares of the event and recurrent intrusive visions of his car slamming into the pedestrian. He reported that he had altered his driving route to work to avoid the scene of the accident, and he found himself switching the television channel whenever a commercial for snow tires appeared. Mr. F. described frequent difficulty falling asleep, poor concentration, and an increased focus on his environment, particularly when he was driving.

GENERALIZED ANXIETY DISORDER

Generalized anxiety disorder is characterized by a pattern of frequent, persistent worry and anxiety that is out of proportion to the impact of the event or circumstance that is the focus of the worry (Table 15.6-15). For example, while college students often worry about examinations, a student who persistently worries about failure despite consistently outstanding grades displays the pattern of worry typical of generalized anxiety disorder. Patients with generalized anxiety disorder may not acknowledge the excessive nature of their worry, but they must be bothered by their degree of worry. This pattern must occur “more days than not” for at least 6 months. The patients must find it difficult to control this worry and must report at least three of six somatic or cognitive symptoms, including feelings of restlessness, fatigue, muscle tension, or insomnia. Worry is a ubiquitous feature of many anxiety disorders: patients with panic disorder often worry about panic attacks, patients with social phobia worry about social encounters, and patients with OCD worry about their obsessions. The worries in generalized anxiety disorder must be beyond the those that characterize these other anxiety disorders. Children exhibiting characteristic symptoms are also considered to meet criteria for generalized anxiety disorder, but they need only meet one of the six somatic or cognitive symptom criteria rather than three.

A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least 6 months, about a number of events or activities (such as work or school performance).
B. The person finds it difficult to control the worry.
C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms present for more days than not for the past 6 months). **Note:** Only one item is required in children.
(1) restlessness or feeling keyed up or on edge
(2) being easily fatigued
(3) difficulty concentrating or mind going blank
(4) irritability
(5) muscle tension
(6) sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)
D. The focus of the anxiety and worry is not confined to features of an Axis I disorder, e.g., the anxiety or worry is not about having a panic attack (as in panic disorder), being embarrassed in public (as in social phobia), being contaminated (as in obsessive-compulsive disorder), being away from home or close relatives (as in separation anxiety disorder), gaining weight (as in anorexia nervosa), having multiple physical complaints (as in somatization disorder), or having a serious illness (as in hypochondriasis), and the anxiety and worry do not occur exclusively during post-traumatic stress disorder.
E. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
F. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., hyperthyroidism), and does not occur exclusively during a mood disorder, psychotic disorder, or a pervasive developmental disorder.

Table 15.6-15 DSM-IV Diagnostic Criteria for Generalized Anxiety Disorder

History and Comparative Nosology Like those of other anxiety disorders symptoms of generalized anxiety disorder have been noted for more than 100 years. Many syndromes considered to be related to panic disorder (e.g., DaCosta's syndrome or neurocirculatory asthenia) also closely resemble generalized anxiety disorder. In fact, prior to DSM-III, panic disorder and generalized anxiety disorder were both subsumed under the broader category of anxiety neurosis.

The conceptualization of generalized anxiety disorder has changed gradually from DSM-III to DSM-IV. The disorder was originally considered a residual category in DSM-III for anxiety disorders that did not fulfill criteria for another disorder. DSM-III required only 1-month duration of symptoms, and concerns arose about the low reliability of the diagnosis. DSM-III-R increased the duration criterion to 6 months, placed more emphasis on the symptom of worry, and added a list of 18 symptoms, from which patients had to exhibit at least 6. DSM-III-R also removed some of the hierarchical rules that had limited the diagnosis to individuals who were free of specific disorders in DSM-III. Finally, DSM-IV further revised the diagnosis. The list of associated symptoms was narrowed from 18 to 6, from which patients had to exhibit at least 3, and more emphasis was placed on the pervasiveness of the worry. DSM-IV attempted to integrate the approach to worry across development. While DSM-III-R possessed the diagnosis of overanxious disorder for use among children, DSM-IV integrated this DSM-III-R diagnosis into generalized anxiety disorder, with some minor threshold differences in making the diagnosis in children.

ICD-10 includes the diagnosis of generalized anxiety disorder and emphasizes the distinction between generalized anxiety and panic disorders (Table 15.6-7). While ICD-10 places an emphasis on worry that is similar to the emphasis in DSM-IV, ICD-10 adopts an approach to the other symptoms of generalized anxiety disorder that is closer to the approach in DSM-III-R than that in DSM-IV. For example, in ICD-10 the diagnosis requires 4 associated symptoms from a list of 22.

Differential Diagnosis Like other anxiety disorders, particularly panic disorder, generalized anxiety disorder must be differentiated from both medical and psychiatric disorders. Neurological, endocrinological, metabolic, and medication-related disorders similar to those considered in the differential diagnosis of panic disorder must be considered in the differential diagnosis of generalized anxiety disorder. Common co-occurring anxiety disorders must also be considered, including panic disorder, phobias, OCD, and PTSD. A patient must exhibit the full syndrome of generalized anxiety disorder that cannot be explained by the presence of such a comorbid anxiety disorder. Diagnosing generalized anxiety disorder in the presence of such other anxiety disorders requires documenting anxiety or worry related to circumstances or topics that are either unrelated or only minimally related to other disorders. Hence, proper diagnosis involves both definitively establishing the diagnosis and properly diagnosing other anxiety disorders. Patients with generalized anxiety disorder also frequently develop major depressive disorder; thus this condition must also be recognized and distinguished. Again, the key feature to making a correct diagnosis is documenting anxiety or worry that is unrelated to the depressive disorder.

Course The lack of prospective epidemiological studies precludes firm conclusions about the course of generalized anxiety disorder. Prospective research, even among clinical samples, is also insufficient. The most complete data on the course of the disorder derive from retrospective epidemiologically based studies. These studies suggest that generalized anxiety disorder is a chronic disorder, as most patients report symptoms for many years prior to assessment. Given the possible biases in such studies, no definitive statement on the course of the disorder can be made.

Ms. X. was a successful, married, 30-year-old attorney who presented for a psychiatric evaluation to treat growing symptoms of worry and anxiety. For the preceding 8 months, Ms. X. had noted increased worry about her job performance. For example, while she had always been a superb litigator, she increasingly found herself worrying about her ability to win each new case she was presented. Similarly, while she had always been in outstanding physical condition, she increasingly worried that her health had begun to deteriorate. Ms. X. noted frequent somatic symptoms that accompanied her worries. For example, she often felt restless while she worked and while she commuted to her office, thinking about the upcoming challenges of the day. She reported feeling increasingly fatigued, irritable, and tense. She noted that she had increasing difficulty falling asleep at night as she worried about her job performance and impending trials.

SUBSTANCE-INDUCED ANXIETY AND ANXIETY DUE TO A GENERAL MEDICAL CONDITION

These conditions are characterized by prominent anxiety that arises as the direct result of some underlying physiological perturbation. Hence, for patients with substance-induced anxiety disorder (Table 15.6-16), clinically significant symptoms of panic, worry, phobia, or obsessions emerge in the context of prescribed or illicit substance use. For example, both prescribed and illicit sympathomimetic substances can often produce relatively marked degrees of anxiety. Similarly, for anxiety due to a general medical condition (Table 15.6-17), symptoms develop in the context of an identifiable medical syndrome. For example, panic attacks have been tied to various medical conditions, including endocrinologic, cardiac, and respiratory illnesses.

Specify if:
With generalized anxiety: if excessive anxiety or worry about a number of events or activities predominates in the clinical presentation
With panic attacks: if panic attacks predominate in the clinical presentation
With obsessive-compulsive symptoms: if obsessions or compulsions predominate in the clinical presentation
Specify if:
With onset during intoxication: if the criteria are met for substance intoxication and the symptoms develop during the intoxication
With onset during withdrawal: if criteria are met for withdrawal from a substance and the symptoms develop during or shortly after withdrawal

Table 15.6-16 DSM-IV Diagnostic Criteria for Substance-Induced Anxiety Disorder

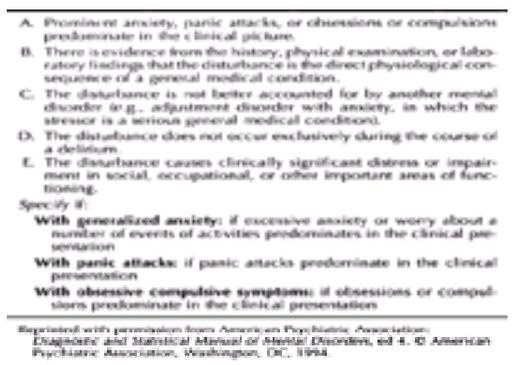

A. Prominent anxiety, panic attacks, or obsessions or compulsions predominate in the clinical picture.
B. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (e.g., adjustment disorder with anxiety, in which the stressor is a serious general medical condition).
D. The disturbance does not occur exclusively during the course of a delirium.
E. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify if:
With generalized anxiety: if excessive anxiety or worry about a number of events or activities predominates in the clinical presentation
With panic attacks: if panic attacks predominate in the clinical presentation
With obsessive compulsive symptoms: if obsessions or compulsions predominate in the clinical presentation

Table 15.6-17 DSM-IV Diagnostic Criteria for Anxiety Disorder Due to a General Medical Condition

The first step in identifying anxiety disorder due to a general medical condition or substance-induced anxiety disorder is to confirm the presence of one or the other complicating factor. Clearly, practitioners should routinely document the medical and substance use status of all patients. However, the clinician should be particularly wary when encountering a patient with an unusual symptomatic presentation. For example, changes in consciousness or neurological function almost never occur in acute anxiety states unless there is also an underlying medical component to the syndrome. In patients where there is a suspicion of such complicating factors, the presence of substance use or medical problems first must be definitively confirmed by obtaining the necessary medical history or evaluative procedures. Next, the clinician must determine that this underlying problem is the cause of the ongoing anxiety symptoms. While there is no definitive test to establish such a causal relationship, several factors can help confirm the diagnosis. These include the timing of the symptoms, the existing literature pertaining to the strength of the association between anxiety and the potential complicating factor, and signs or symptoms (e.g., changes in consciousness) that are atypical for an anxiety disorder. Finally, even more suggestive evidence can be provided if alleviation of the complicating medical factor produces an amelioration of the anxiety symptoms.

ANXIETY DISORDER NOT OTHERWISE SPECIFIED

Anxiety represents one of the most common psychiatric symptoms encountered in various settings, including primary care settings, and it is relatively common to encounter patients who exhibit impairment from anxiety but who do not meet criteria for one of the specific anxiety disorders. These patients are appropriately classified as suffering from anxiety disorder not otherwise specified (Table 15.6-18).

This category includes disorders with prominent anxiety or phobic avoidance that do not meet criteria for any specific anxiety, disorder, adjustment disorder with anxiety, or adjustment disorder with mixed anxiety and depressed mood. Examples include

1. Mixed anxiety-depressive disorder: clinically significant symptoms of anxiety and depression, but the criteria are not met for either a specific mood disorder or a specific anxiety disorder.
2. Clinically significant social phobic symptoms that are related to the social impact of having a general medical condition or mental disorder (e.g., Parkinson's disease, dermatological conditions, stuttering, anorexia nervosa, body dysmorphic disorder).
3. Situations in which the clinician has concluded that an anxiety disorder is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced.

Table 15.6-18 DSM-IV Diagnostic Criteria for Anxiety Disorder Not Otherwise Specified

Two clinical features of this disorder must be recognized to properly identify the condition. First, the anxiety described by the patients must be distressing and interfere with some aspect of functioning. Second, the anxiety must not be attributable to another psychiatric condition. For example, patients with generalized anxiety disorder may not initially report sufficient associated symptoms to meet criteria for this condition. However, on further probing such symptoms may be identified. Particularly in patients with long standing anxiety, it is important to establish that another anxiety disorder does not account for the complaints. Anxiety concerning an embarrassing medical problem or scenario is another frequently encountered form of anxiety disorder not otherwise specified. For example, patients who exhibit excessive concern regarding a dermatological condition might exhibit symptoms of this syndrome.

Perhaps the most consistent research on this condition examines patients with mixed anxiety-depressive disorder, a condition described in an Appendix of DSM-IV (Table 15.6-19). Patients with mixed anxiety-depressive disorder exhibit symptoms both of depression and anxiety that do not meet criteria for another mood or anxiety disorder. Such patients must show signs of consistent low mood for at least a month, accompanied by additional symptoms that include prominent worries. Longitudinal studies find a relatively high risk for later mood or anxiety disorders with this condition, particularly major depressive disorder. Due to the paucity of data on treatment for the condition, clinicians often use approaches that are effective in other mood or anxiety disorders.

A. Persistent or recurrent dysphoric mood lasting at least 1 month.
B. The dysphoric mood is accompanied by at least 1 month of four (or more) of the following symptoms:
   (1) difficulty concentrating or mind going blank
   (2) sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)
   (3) fatigue or low energy
   (4) irritability
   (5) worry
   (6) being easily moved to tears
   (7) hypervigilance
   (8) anticipating the worst
   (9) hopelessness (pervasive pessimism about the future)
   (10) low self-esteem or feelings of worthlessness
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.
E. All of the following:
   (1) criteria have never been met for major depressive disorder, dysthymic disorder, panic disorder or generalized anxiety disorder
   (2) criteria are not currently met for any other anxiety or mood disorder (including an anxiety or mood disorder, in partial remission)
   (3) the symptoms are not better accounted for by any other mental disorder

Table 15.6-19 DSM-IV Research Criteria for Mixed Anxiety-Depressive Disorder

SUGGESTED CROSS-REFERENCES

Other aspects of these disorders are discussed elsewhere, such as theories on etiology (Section 15.3, Section 15.4, and Section 15.5) and treatment (Section 15.7 and Section 15.8). Similarly, substance-induced anxiety disorder and anxiety disorder due to a general medical condition as distinct anxiety disorders are also discussed in Chapter 11 on substance-use disorders and Chapter 10 on the psychiatric complications of medical conditions.

SECTION REFERENCES

Angst J, Vollrath M: The natural history of anxiety disorders. Acta Psychiatr Scand *84:*446, 1991.

Ballenger JC, Fyer AJ: DSM-IV in progress: Examining criteria for panic disorder. Hosp Community Psychiatry *44:*226, 1993.

*Barsky AJ, Delamater BA, Orav JE: Panic disorder patients and their medical care. Psychosomatics *40:*50, 1999.

Barlow DH: *Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic*. Guilford, New York, 1988.

Bresalau N, Davis GC, Andreski P, Peterson E: Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry *48:*216, 1991.

*Calamari JE, Wiegarts PS, Janeck AS: Obsessive-compulsive disorder subgroups: A symptom-based clustering approach. Behav Res Ther *37:*113, 1999.

Carden E, Speigel D: Dissociative reaction to the San Francisco Bay Area earthquake of 1989. Am J Psychiatry *150:*474, 1993.

Davidson JRT, Foa EB: Diagnostic issues in posttraumatic stress disorder: Considerations for the DSM-IV. J Abnorm Psychol *1900:*346, 1991.

Davidson JRT, Hughes DL, George LK, Blazer DG: The epidemiology of social phobia: Findings from the Epidemiologic Catchment Area study. Psychol Med *23:*709, 1993.

Freud S: Inhibitions, symptoms and anxiety. In standard edition of *The Complete Psychological Works of Sigmund Freud*, vol 20. Hogarth Press, London, 1966, p 77.

Freud S: Obsessions and phobias. In standard edition of *The Complete Psychological Works of Sigmund Freud*, vol 3. Hogarth Press, London 1966.

Fyer AJ, Mannuzza S, Chapman T, Liebowitz MR, Klein DF: A direct interview family study of social phobia. Arch Gen Psychiatry *50:*286, 1994.

Gater R, Tansella M, Korten A, Tiemens BG, Maurbas VG, Olatawura MO: Sex differences in the prevalence and detection of depressive and anxiety disorders in general care settings: Report from the WHO collaborative study on psychological problems in general health care. Arch Gen Psychiatry *55:*405, 1998.

Gorman JM, Papp LA: Chronic anxiety: Deciding the length of treatment. J Clin Psychiatry *5:*11, 1990.

Gorman JM, Papp LA, editors: Anxiety disorders. In *Annual Review of Psychiatry*, vol 11, A Tasman, editor. American Psychiatric Association Press, Washington, DC, 1992.

Horwath E, Wolk SI, Goldstein RB, Wickramaratne P, Sobin C, Adams P, Lish JD, Weissman MM: Is the comorbidity between social phobia and panic disorder due to familial contransmission or other factors? Arch Gen Psychiatry *52:*574, 1995.

Jarrell MP, Ballenger JC: Psychiatric comorbidity in patients with generalized anxiety disorder. Am J Psychiatry *150:*1216, 1993.

Jenike MA, Baer L, Minichiello WE: *Obsessive-Compulsive Disorders: Theory and Management*, ed 2. Year Book Publishing, Chicago, 1990.

Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the national comorbidity survey. Arch Gen Psychiatry *52:*1048, 1995.

Klein DF: Delineation of two drug-responsive anxiety syndromes. Psychopharmacology *5:*397, 1964.

Klein DF: False suffocation alarms, spontaneous panics, and related conditions: An integrative hypothesis. Arch Gen Psychiatry *50:*306, 1993.

Klein DF, Rabkin JG, editors: *Anxiety: New Research and Changing Concepts*. Raven, New York, 1981.

Klerman GL, Weissman MM, Ouellette R, Johnson J, Greenwald S: Panic attacks in the community: Social morbidity and health care utilization. JAMA *265:*742, 1991.

Liebowitz MR, Gorman JM, Fyer AF, Klein DF: Social phobia: Review of a neglected anxiety disorder. Arch Gen Psychiatry *42:*729, 1985.

Magee WJ, Eaton WW, Wittchen HU, McGonagle KA, Kessler RC: Agoraphobia, simple phobia, and social phobia in the National Comorbidity Survey. Arch Gen Psychiatry *53:*159, 1996.

Marks IM: *Fears, Phobias, and Rituals: Panic, Anxiety, and Their Disorders*. Oxford University Press, New York, 1988.

McFarlane AC: The phenomenology of post-traumatic stress disorders following a natural disaster. J Nerv Ment Dis *176:*22, 1988.

Pine DS, Cohen P, Gurley D, Brook J, Ma Y: The risk for early-adult anxiety and depressive disorders in adolescents with anxiety and depressive disorders. Arch Gen Psychiatry *56:*775, 1998.

*Rapee RM, Barlow DH, editors: *Chronic Anxiety*. Guilford, New York, 1991.

Rapoport JL: *The Boy Who Couldn't Stop Washing*. Dutton, New York, 1989.

Shalev AY, Sakar T, Freedman S, Peri T, Glick N, Brandes D, Orr SP, Pitman RK: A prospective study of heart rate response following trauma and the subsequent development of PTSD. Arch Gen Psychiatry *56:*553, 1998.

Stein MB, Walker JR, Forde DR: Public-speaking fears in a community sample: Prevalence, impact on functioning, and diagnostic classification. Arch Gen Psychiatry *53:*169, 1995.

Smoller JW, Tsuang MT: Panic and phobic anxiety: Defining phenotypes for genetic studies. Am J Psychiatry *155:*1152, 1998.

Weissman MM: Panic and generalized anxiety: Are they separate disorders? J Psychiatr Res *24*(Suppl):157, 1990.

Wittchen HV, Reed V, Kessler RC: The relationship of agorophobia and panic in a community sample of adolescents and young adults. Arch Gen Psychiatry *56:*1017, 1998.

15.7 ANXIETY DISORDERS: SOMATIC TREATMENT

LASZLO A. PAPP, M.D.

After decades of neglect and controversy, anxiety disorders are increasingly recognized as legitimate medical conditions requiring specific treatment. The unproductive debate over the primacy of biological or psychological factors in the pathophysiology of anxiety is gradually being replaced by a pragmatic approach based on research on the relative contributions of both. A parallel, unbiased approach in treatment research began to examine the merits of combined somatic and psychological treatments in anxiety. Recognition of the role of psychological factors in most medical illnesses is breaking down the artificial barrier between psychiatric and medical conditions. While the stigma of psychiatric illness remains a strong deterrent, the availability of effective treatments, frequently a combination of somatic and psychological, is a powerful incentive for patients to come forward.

To date panic disorder remains the best-researched anxiety disorder. Pharmacological treatment trials in panic disorder outnumber those conducted in other anxiety disorders, and most pharmacological challenge studies were first initiated in panic disorder.

HISTORY

The oldest antianxiety drug is alcohol, and it remains the most frequently used and most easily accessible tranquilizer. Modern medical anxiolysis began with the introduction of paraldehyde and bromides around the turn of the century, followed by the first medical use of barbiturates in 1903. While the availability of a number of safer and more effective anxiolytics today makes them mostly obsolete, paraldehyde remains an old-fashioned treatment for alcohol withdrawal, and certain barbiturates retain limited use in anesthesiology and addiction medicine. The development of the so-called nonbarbiturate nonbenzodiazepine hypnotic drugs in the 1930s did not address any of the deficiencies of the barbiturates, and in many cases these drugs proved more problematic. Meprobamate (Equanil, Miltown) and tybamate possess very low therapeutic index, methaqualone and methyprylone (Noludar) are highly addicting and can be fatal in overdose, and glutethimide (Doriden) overdose can result in convulsion and fluctuating coma. The synthesis in 1957 of the first benzodiazepine, chlordiazepoxide (Librium), heralded a new era of safe and effective medical management of anxiety. Because of their safety, efficacy, and high therapeutic index, benzodiazepines have for the most part replaced barbiturates and the nonbarbiturate, nonbenzodiazepine type drugs.

The demonstration in the early 1960s that imipramine (Tofranil) controls panic attacks was the first evidence that antidepressant drugs may alleviate anxiety and that this effect may be independent of their antidepressant property. The historic observation that panic attacks were specifically responsive to antidepressants also marked the beginning of a new diagnostic system that differentiates the subtypes of anxiety neuroses on the basis of medication response. Parallel to the work with imipramine in the United States, British investigators found that another class of antidepressants, the monoamine oxidase inhibitors (MAOIs), specifically benefit hysterical patients with phobic symptoms. Since these patients show many features of panic disorder and agoraphobia, tricyclic and tetracyclic drugs and MAOIs quickly became the first-line treatment choice in panic disorder. The pharmacological dissection of the formally homogeneous anxiety neuroses category into distinct anxiety disorders not only changed diagnostic thinking and the treatment of anxiety but also revolutionized psychopharmacology in general. Symptom-specific rather than diagnosis-specific drug development helped refine and create diagnostic categories for all psychiatric conditions.

Of the many subsequently introduced antidepressants with anxiolytic properties, fluoxetine (Prozac) was the next milestone in the pharmacology of anxiety. This first drug in a series of serotonergic agents became the best-selling antidepressant by 1990. The efficacy of serotonergic drugs in the treatment of panic disorder and obsessive-compulsive disorder significantly advanced the treatment of these anxiety disorders and gave rise to new theories implicating the serotonergic system in the neurobiology of anxiety.

In addition to benzodiazepines and several classes of antidepressants currently available anxiolytic agents include b-adrenergic receptor antagonists and the azapyrone buspirone (BuSpar). New drug development targets neurotransmitter systems identified primarily by pharmacological challenges as pertinent to the neurobiology of anxiety. Candidates include partial benzodiazepine agonists and reverse benzodiazepine antagonists, neurosteroids, neuropeptide agonists and antagonists such as cholecystokinin B (CCK-B) antagonists, corticotropin-releasing factor (CRF) antagonists, neuropeptide Y agonists and serotonergic agents acting on specific serotonin receptor subtypes (e.g., serotonin [5-hydroxytryptamine (5-HT)] subtype 1A [5-HT_{1A}] agonists, and 5-HT_2 and 5-HT_3 antagonists). The accelerated drug development process promises highly effective anxiolytic agents with minimal adverse effects in the near future.

GENERAL PRINCIPLES

The general premise of medication use in anxiety disorders is that an underlying neurochemical imbalance is responsible for the disorder, which can be corrected by administration of drugs affecting specific neurotransmitter systems. Another view of medication use in anxiety disorders posits that anxiety disorders result from underlying cognitive, behavioral, or psychodynamic abnormalities and that the role of medications is to allow the patient to participate in appropriate psychotherapeutic work. The evidence is simply not available at present to adjudicate the primacy of one approach over the other. The knowledge base is rapidly evolving, but it will be many years before definitive statements can be made and causal relationships among neurophysiology, psychopathology, drugs, specific psychotherapeutic interventions, and behavioral outcome can be established. However, medication management of most anxiety disorders with or without psychotherapy is one of the most successful treatments in medicine today. Treatment response rates can reach 80 percent in panic disorder, with steadily improving rates for many other anxiety disorders. Nevertheless, a nonpharmacological treatment will almost always be preferable if comparable efficacy can be established. However benign and tolerable they may be, medication use usually entails some adverse effects. Medications may also be contraindicated because of underlying medical conditions or pregnancy.

EVALUATION

Evaluation of the anxious patient for somatic treatment begins with the psychiatric interview. Particular attention is paid to onset, course, symptomatology and comorbidity. While the anxiety disorders represent distinct diagnostic categories, they also frequently present as comorbid conditions in a given patient. Other comorbid conditions include depression, substance use disorders, and personality disorders. Comorbidity usually complicates the management of anxiety disorders and often predicts poorer outcome. The difficulty in making a differential diagnosis in the presence of comorbid depression and anxiety prompted the suggestion to create a separate diagnostic category, mixed anxiety-depressive disorder. Some believe that this new category will discourage rigorous diagnostic thinking; others consider the issue moot from a pharmacological point of view, since most antianxiety medications have antidepressant properties. In sum, establishing the correct diagnosis remains one of the cornerstones of any therapeutic endeavor.

The initial interview should elicit detailed information about prior treatments. In addition to doses, length of treatment period, response, and adverse effects for each agent, the interviewer should ask about prejudices, attitudes, and reservations about taking medications in general. Family history may help in diagnosis, and treatment history of family members can provide important clues about medication response, which also runs in families.

The initial evaluation should include (when appropriate) a thorough physical examination blood tests for chemistry, hematology, and thyroid function; urinalysis; and an electrocardiogram (ECG). If neurological abnormalities are suspected, a neurological examination may need to be supplemented with appropriate brain imaging

studies. Prominent nocturnal pathology should be investigated in a sleep laboratory.

Choosing the Medication When the decision to use medication is made, the psychiatrist frequently faces a dilemma. Scientifically valid recommendations about anxiety management should be based on rigorous double-blind, placebo-controlled trials. However, even when the data are available, the results of controlled trials do not necessarily reflect clinical reality or guide practice. Patients participating in (usually short-term) treatment trials are preselected to meet entry criteria. Natural comorbidity is limited and severity of symptoms is calibrated to ensure sample homogeneity. Elderly, pediatric, and medically ill patients are specifically excluded from most therapeutic trials. Therefore, the bulk of what psychiatrists do in clinical practice remains empirical, and many recommendations are based on expert clinical consensus.

While the choice of medication for a particular patient involves many considerations, treatment algorithms have recently been developed to guide the practitioner. These algorithms consist of decision trees based on numerous, but necessarily limited factors generally applicable to typical patients. In the absence of conclusive evidence for differential efficacy, the choice is usually determined by adverse-effect profiles. These algorithms frequently include combination treatments. Since most of these recommendations are not supported by systematic data, they should be viewed as preliminary, and periodic revisions are imperative. Whenever available, the most salient features of these algorithms are presented for the specific anxiety disorder.

Duration of Treatment Most information on the pharmacological management of anxiety is limited to short-term treatment trials. Long-term trials and follow-ups are rare; sequential treatment research is practically nonexistent. While some studies suggest that longer maintenance treatments lower relapse rates, the optimal length of medication maintenance treatment is still unknown. Since most anxiety disorders are chronic, frequently life-long conditions, treatment should probably be continued indefinitely but perhaps intermittently. In general, the lowest dosage of medication that controls the patient's symptoms should be prescribed. Slow, gradual taper of medication dosage may be attempted after 1 year in symptom-free patients; recurrence should be retreated promptly. The benefit of freedom from adverse effects should be weighed against developing drug resistance. Retreatment with a previously efficacious agent may not produce the same benefits. Frequently, a former drug responder requires several medication trials before an efficacious drug is again found.

Another, possibly related, problem associated with long-term pharmacotherapy of anxiety is drug burnout. *Burnout* refers to patients who, after a few months of response, develop drug resistance while on medication. In these patients, adjusting the dosage up or down may help, but usually one must augment or replace the original medication. Burnout is unrelated to drug tolerance, which is rare in patients suffering from pure anxiety disorders. Even when treated with benzodiazepines, the drug group claimed to be most addicting, patients with anxiety disorders usually continue to respond to stable or decreasing doses.

Treatment Nonresponse Treatment nonresponse should prompt careful reassessment of the target symptoms, review of adverse effects, discussion of patient expectations, reevaluation of the diagnosis, checking compliance with treatment, identification of comorbid conditions (including personality disorders, substance use, and medical conditions), and potential interactions with concomitant medications (including consideration of pharmacokinetics and pharmacodynamics). The most frequent reason for medication nonresponse remains undermedication (i.e., inadequate dosage or duration). Dosing and length of treatment should be based on valid clinical trials or (in the absence of data) collective clinical experience.

Managing Adverse Effects One of the main reasons for treatment resistance is noncompliance, frequently related to anticipated or actual adverse effects. With the availability of many agents of comparable efficacy, side-effect profiles are increasingly considered in choosing a drug. A less efficacious drug may even be selected because its has a more favorable side-effect profile.

Patients should be well informed about adverse reactions and their management; however, the education and reassurance should match the needs of the patient. An overinclusive presentation may frighten a suggestible patient with anxiety disorder. As an introduction, patients should always be told that most side effects are benign and do not represent clinically significant limitations.

Lower-than-therapeutic starting dosages are generally better tolerated. Even if the initial agitation reaction of panic patients is not expected, a gradual, slow increase gives the anxious patient the best chance to accommodate to adverse reactions. Patients with dry mouth should increase fluid intake, use sugarless gum or candy, and may also try bethanecol (Urecholine). Constipation usually responds to high-bulk diet with plenty of fluid, stool softeners, or milk of magnesia. Avoiding sudden postural changes and using constrictive support hose usually minimizes orthostatic hypotension. If it persists in spite of increased fluid and salt intake, salt tablets, mineralocorticoids, amphetamines, or yohimbine (Yocon) may be tried unless specifically contraindicated. Urinary hesitancy should prompt dosage reduction, and bethanecol may also be tried. In general, anticholinergic side effects are best handled by switching to a less anticholinergic drug. Anticholinergic drugs should not be used in patients with prostatic hypertrophy or narrow-angle glaucoma. Sexual dysfunction is one of the most problematic and resistant adverse effects of drugs with significant serotonergic activity. Dosage reduction is only occasionally helpful. Anorgasmia may respond to bethanecol, cyproheptadine (Periactin), or yohimbine; diminished libido may improve if buspirone (Dulper) or bupropion (Wellbutrin) are added. Sildenfanil (Viagra) is under investigation for use in reversing the adverse sexual effects of SSRIs.

Medication-induced insomnia may improve over time. Switching the dose to the morning should be tried before hypnotic drugs such as a benzodiazepine, zolpidem (Ambien), or trazodone (Desyrel) are added. Sedating drugs should be given at bedtime or in split doses. Patients should be warned that most antidepressants cause photosensitivity. They should also be told that tricyclic and tetracyclic antidepressants may reduce parasympathetic innervation to the heart resulting in decreased beat to beat heart rate variability and potentially increased risks of coronary heart disease and arrhythmia. Antidepressant-induced hypomania may respond to dosage reduction, but if the patient has a history of bipolar disorder, mood stabilizers may be indicated.

Weight gain is one of the most troublesome and limiting adverse effects of antidepressant antianxiety drugs. Tricyclic and tetracyclic drugs with strong anticholinergic profiles and MAOIs are probably the worst offenders, followed by the serotonergic class. Atypical antidepressants such as bupropion, trazodone, and nefazadone (Serzone) are less likely to cause weight gain. If diet and exercise fail, dosage reduction or a switch to a different agent are likely to be more successful.

Antipsychotics As a general rule, antipsychotic medications should not be used for the management of anxiety disorders. However, the new generation of safer, atypical antipsychotics, (i.e., serotonin-dopamine antagonists) appears to have dramatically reduced risks of extrapyramidal side effects and tardive dyskinesia. Antipsychotic drugs may be considered in the treatment of agitation in elderly patients, managing anxiety in patients with organic brain disease, and potentiating antidepressants in patients with refractory obsessive-compulsive disorder with tic disorder or schizotypal personality disorder. Periodic, short-term antipsychotic drug may also be used in patients with histories of multiple substance dependence and those for whom all other alternatives have failed.

Medication Monitoring Duration and frequency of contact between the patient and the medicating doctor vary, depending on many factors (logistics, severity of illness, the patient's response to the medication, and the availability and use of ancillary mental health and medical services), but general guidelines apply. The initial evaluation should last for at least 1 hour. If the patient begins to take medication, a follow-up visit lasting from 20 to 30 minutes should occur within a week or two after the first visit. Another follow-up visit is usually indicated approximately 1 month after the initial evaluation, and if the condition of the patient is stable, monthly visits usually suffice to monitor progress. Face-to-face contact could be less frequent in certain situations, but a visit every 3 months is usually the outside limit. The prescribing physician should always make arrangements for 24-hour coverage and should be available to answer questions on the phone.

PANIC DISORDER WITH OR WITHOUT AGORAPHOBIA

The symptomatic triad of panic disorder consists of panic attacks, usually complicated by anticipatory anxiety and phobic avoidance, or agoraphobia. For a long time the pharmacological approach to panic disorder assumed that blocking panic attacks with medications lead to gradual improvement in both anticipatory anxiety and phobic avoidance. Therefore, panic blockade was the main focus of pharmacotherapy. Response rates to panic blockade alone range from 50 to 80 percent, depending on the definition of response. Since some of these panic-free patients remain avoidant and apprehensive, clinical practice must consider the full spectrum of panic disorder, including anticipatory anxiety, phobia, and comorbid conditions such as mood disorders.

The natural course of panic disorder varies. A significant proportion of these patients have very favorable outcomes, others suffer from long-term disability, and some have an alternating course. To address these problems, the traditional, relatively straightforward, antipanic drug regimens are gradually replaced or modified in clinical practice by complex decision trees, treatment algorithms, drug combinations, and augmentation strategies. The recommendation for refractory cases frequently includes addition of specific psychotherapeutic techniques as well.

At the time of this writing only five drugs have been approved by the Food and Drug Administration (FDA) to treat panic disorder, but the number of efficacious, off-label antipanic agents are in the dozens. They include high- and low-potency benzodiazepines, selective serotonin reuptake inhibitors (SSRIs), tricyclic and

tetracyclic, MAOIs, reversible MAOIs (RIMAs), and novel antidepressants. In explaining the pharmacological options, the clinician should tell the patient about the risks and benefits of these medications. Specifically, the patients should know that most of these medications are equally efficacious and that the choice is usually made on the basis of side effects.

Current consensus is to start treating a patient with uncomplicated panic disorder with low dosages of an SSRI (Table 15.7-1). Even at these low starting dosages many panic patients may experience initial agitation and more-frequent panic attacks. This so-called supersensitivity syndrome is usually limited to the first week or two of treatment. Further dosage reduction, switching to a different compound in the same family, or addition of a high-potency benzodiazepine, such as clonazepam (Klonopin) or alprazolam (Xanax), usually gets the patient through this relatively short period. Dosages can subsequently be increased over several weeks until they reach therapeutic range. For full panic blockade, most patients require dosages equivalent to those used in depression. A therapeutic trial with these agents should last for at least 5 weeks. Panic blockade often leads to improvement in both anticipatory anxiety and phobic avoidance. Preliminary evidence suggests that improvement in phobic avoidance may require higher dosages than panic blockade.

	Starting (mg)	Maintenance (mg)
SSRI		
Paroxetine	5–10	20–60
Fluoxetine	2–5	20–60
Sertraline	12.5–25	50–200
Fluvoxamine	12.5	100–150
Citalopram	10	20–40
Tricyclic antidepressants		
Clomipramine	5–12.5	50–125
Imipramine	10–25	150–300
Desipramine	10–25	150–200
Benzodiazepines		
Alprazolam	0.25–0.5 t.i.d.	0.5–2 t.i.d.
Clonazepam	0.25–0.5 b.i.d.	0.5–2 b.i.d.
Diazepam	2–5 b.i.d.	5–30 b.i.d.
Lorazepam	0.25–0.5 b.i.d.	0.5–2 b.i.d.
MAOIs		
Phenelzine	15 b.i.d.	15–45 b.i.d.
Tranylcypromine	10 b.i.d.	10–30 b.i.d.
RIMAs		
Moclobemide	50	300–600
Brofaromine	50	150–200
Atypical antidepressants		
Venlafaxine	6.25–25	50–150
Nefazodone	50 b.i.d.	100–300 b.i.d.
Other agents		
Valproic acid	125 b.i.d.	500–750 b.i.d.
Inositol	6000 b.i.d.	6000 b.i.d.

Table 15.7-1 Recommended Dosages for Antipanic Drugs (Daily Unless Indicated Otherwise)

Approximately 60 percent of patients respond to this approach. Partial responders can benefit from the addition of a high-potency benzodiazepine or buspirone. Partial responders with tachycardia may try a combination of an SSRI and b-adrenergic receptor antagonists. b-adrenergic receptor antagonists may also alleviate cardiac discomfort associated with mitral valve prolapse. The dosage of b-adrenergic receptor antagonists such as propranolol (Inderal) or atenolol (Tenormin) should be titrated until the heart rate is reduced by 5 to 10 beats per minute. Contraindications to using b-adrenergic receptor antagonists include bradycardia, heart block and asthma. Drug-drug interactions between some SSRIs and some b-adrenergic receptor antagonists should be kept in mind.

Nonresponders should be tapered off the SSRI and offered one of the following options: a second SSRI, venlafaxine (Effexor), a high-potency benzodiazepine, a tricyclic or tetracyclic, or an MAOI. Persistent adverse effects such as weight gain, hypomania, or sexual dysfunction, may also necessitate a switch from an SSRI in otherwise fully responding panic patients, although first an attempt should be made to address these adverse effects.

Originally, one of the most important advantages of SSRI treatment was considered the absence of withdrawal reaction upon discontinuation. However, withdrawal reactions ranging from mild, transient anxiety and insomnia to severe headache, nausea, dizziness, and “electric jolts” lasting for several months have since been reported in up to 86 percent of patients who abruptly discontinue taking SSRIs. Therefore, gradual taper over several weeks is strongly recommended. Anecdotal reports suggest that the addition of benzodiazepines, or trazodone for sleep may alleviate SSRI withdrawal.

MAOIs MAOIs such as phenelzine (Nardil) or tranylcypromine (Parnate) are the most likely to be efficacious in panic disorder, but they present the dilemma of dietary restrictions and the danger of hypertensive reaction, weight gain, orthostatic hypotension, insomnia, and a series of anticholinergic adverse effects. While somewhat less effective, a RIMA, such as moclobemide (Aurorix), is an alternative. After 2 weeks of washout (6 weeks for fluoxetine), the MAOI can be given and the dosage gradually raised to usual antidepressant levels. Comorbid major depressive in disorder in panic disorder patients may respond best to MAOIs. Because of severe withdrawal reactions such as disinhibition, irritability, agitation, insomnia, myoclonic jerks, and occasionally, delirium, thought disorder, cognitive impairment, and mania, gradual taper is strongly recommended before MAOIs are discontinued. The dietary restriction should continue for at least 2 weeks beyond discontinuation.

High-Potency Benzodiazepine High-potency benzodiazepines should be considered if the adverse effects of all other alternatives are unacceptable to the patient, or they may be first-line choices if the patient is unwilling or unable to wait out the 4- to 5-week delay in response associated with most antidepressants. The advantage of switching to a benzodiazepine also includes no wash-out time for the SSRI. In fact, a benzodiazepine can be added to the SSRI during taper. A typical switch would entail adding 0.5 mg of clonazepam twice a day to 40 mg of paroxetine (Paxil). A few days later the paroxetine dose can be lowered to 20 mg, and depending on its sedative effects, clonazepam dosage can be raised to 1 mg twice a day. Every 3 to 4 days the dosage of paroxetine can be halved while clonazepam dosage is adjusted as tolerated.

High-potency benzodiapines and perhaps low-potency benzodiazepines as well are powerful antipanic drugs. Their sedative and habituating potentials are easily managed in most patients and should not significantly diminish their clinical utility. A major advantage of benzodiazepines is their quick onset of action. Most responders benefit within 1 week. Panic patients with prior history of substance use disorder should not be treated with benzodiazepines. Older patients should take significantly lower doses of benzodiazepines because they are more sensitive to sedation and the potential cognitive side effects of benzodiazepines. A history of organic brain disease or significant character pathology may predispose patients to experience disinhibition while taking benzodiazepines.

The serious and clinically limiting problem of withdrawal symptoms from benzodiazepines prompted many studies attempting to alleviate the withdrawal reaction, also referred to as discontinuance syndrome. Studies with clonidine (Catapres), propranolol (Inderal), and carbamazepine (Tegretol) yielded inconsistent results, and substituting other medications such as antidepressants or azapirones that are easier to taper has not been studied sufficiently. Cognitive-behavioral therapy may best counter benzodiazepine withdrawal. The severity of withdrawal is related to the dosage and length of use and to the rate of taper. The rule of lowest efficacious dose for the shortest time needed for improvement should always be observed. Tapering should never exceed a 10 percent dosage reduction every 3 days. Several months of tapering after years of high-dose benzodiazepine use is frequently unavoidable.

Follow-up studies suggest that benzodiazepine-responsive panic patients maintain their gains over several years and do not develop tolerance to the drugs' antianxiety effects. In fact, maintenance dosages of benzodiazepines are frequently lower than dosages during acute treatment. Residual symptoms may require additional treatment. Since use of short-half-life benzodiazepines such as alprazolam may be associated with between-dose rebound anxiety, some clinicians prefer benzodiazepines with longer half-lives, which may also be easier to taper.

Tricyclics and Tetracyclics Switching to a cyclic antidepressant should begin with tapering the SSRI dosage down to the starting dosage. Some clinicians consider it safe to add the tricyclic at low starting dosages while others recommend a drug-free period lasting from 24 hours to 1 week before administering the second drug. The reason for caution is the potential development of the serotonin syndrome, a rare but severe reaction that may occur when two serotonergic drugs are combined. Also, plasma concentrations of tricyclic and tetracyclics may increase due to inhibition of their metabolism by the SSRI.

The first drug ever shown to possess antipanic efficacy is the tricyclic drug imipramine, and many clinicians still consider it the gold standard in panic disorder. Tricyclic and tetracyclic are generally well tolerated. Their clinical utility is clearly limited, however, by anticholinergic and cardiovascular adverse effects and toxicity in overdose. One of the most effective tricyclic drug for panic is clomipramine (Anafranil), which seems to be effective at lower than full antidepressant dosage. A linear relationship may exist between total plasma tricyclic drug concentrations and response rate. In general, agoraphobia may respond to higher dosages than panic attacks alone. Because of the variable nature of the withdrawal reaction upon discontinuation of tricyclic or tetracyclic drugs (somatic distress, sleep disturbance, behavioral activation, cardiac arrhythmias), taper should be tailored to the patient's needs. Halving the original dosage will give an early indication of withdrawal severity.

Secondary Alternatives While controlled data are unavailable, clinicians increasingly turn to venlafaxine (Effexor) as an alternative to SSRIs and tricyclic drugs in the treatment of panic disorder. If such initial adverse effects as nausea and agitation are overcome, venlafaxine is a powerful antipanic agent at a dosage

significantly below its antidepressant dosage. Since the safety of combining venlafaxine and an SSRI has not been systematically assessed, the SSRI should be tapered and discontinued before switching to venlafaxine. A potential limitation of venlafaxine is an unusually severe reaction to taper. Many successfully treated patients complain of nausea, headache, fatigue, dizziness, and gastrointestinal disturbance when trying to reduce the venlafaxine dosage. Anecdotal evidence suggests that the temporary addition of fluoxetine (10 mg a day) may reduce the withdrawal reaction and allow discontinuation of venlafaxine.

Switching to a second SSRI may be justified if the adverse effects of all alternative possibilities are unacceptable to the patient or there is a relative contraindication to using anything but an SSRI. Again, in the absence of controlled data, opinions differ on how to switch between two SSRIs. Since the potential for the serotonin syndrome is even higher than in the case of combining an SSRI and a tricyclic antidepressant, caution is warranted, and most clinicians recommend a drug-free period between two SSRIs.

Buspirone was promoted as a less-sedating alternative to benzodiazepines in the treatment of panic disorder. Buspirone has lower potential for abuse and dependence than benzodiazepines and produces relatively few adverse effects and no withdrawal syndrome. Buspirone does not alter cognitive or psychomotor function, does not interact with alcohol, and is not a muscle relaxant or an anticonvulsant. However, the efficacy of buspirone in panic disorder is disappointing, and with its further handicap of delayed onset of action and the need for multiple dosing, its use is limited to potentiating the efficacy of other antidepressants and countering the adverse sexual effects of SSRIs. Also, buspirone seems even less effective in patients previously exposed to benzodiazepines.

Bupropion, maprotiline (Ludiomil), and trazodone have not been found efficacious for panic disorder in controlled studies, while the anticonvulsants divalproex (Depakote) and gabapentin (Neurontin), the polyol second-messenger precursor inositol, nefazodone, and the calcium channel inhibitor verapamil (Calan, Isoptin) have shown promise as antipanic agents.

While the short-term efficacy of antipanic medications has been established, the question of how long to treat a panic patient who responds to treatment remains open. The results of follow-up studies are mixed. Several reports indicate that most panic patients relapse within 2 months to 2 years after the medication is discontinued. A recent review concludes that following medication discontinuation, only about 30 to 45 percent of the patients remain well, and even remitted patients may display a variety of symptoms. Others find that while occasional panic attacks are quite common, otherwise responding patients rarely revert back to significant phobic avoidance or serious vocational or social disability. Improvement may continue for years following a single course of medication treatment. This favorable outcome may be explained by the heterogeneity of panic disorder, spontaneous learning experience of patients in clinical trials, and concomitant self-monitoring. Given the uncertainty about the optimal duration of treatment, the current recommendation is to continue full-dosage medication for panic-free patients for at least 1 year. Medication taper should be slow, with careful monitoring of symptoms. Distinction should be made among return symptoms, withdrawal, and rebound anxiety.

Since longer duration of illness at baseline predicts poor long-term outcome, all efforts should be made to identify and treat panic patients as early as possible. More-severe phobic avoidance and comorbid depression and social phobia at baseline also predict poor long-term outcome. Higher depression scores coincide with greater severity of avoidance and disability. The poorer overall outcome in panic disorder patients with comorbid recurrent depression is more likely due to the simultaneous presence of the two conditions. Comorbid depression usually improves in parallel with panic symptoms.

Atypical responses to medications have been reported in panic patients whose first panic attacks were precipitated by cocaine use. These patients respond preferentially to benzodiazepines and anticonvulsants, while tricyclic drugs seem to worsen their anxiety symptoms. This pattern of medication response suggests that cocaine-induced panic attacks may be related to a kindling-like phenomenon.

In a small series of patients with comorbid depression and panic disorder electroconvulsive therapy (ECT) seemed highly effective. None of these patients experienced panic attacks after their fourth ECT treatment until discharge.

GENERALIZED ANXIETY DISORDER

Patients with generalized anxiety disorder suffer from excessive and uncontrollable anxiety and worry for at least 6 months and experience a series of somatic symptoms such as restlessness, irritability, insomnia, and muscle tension. The illness is chronic, with periodic exacerbations and relative quiescence. The relative sparsity of biological data and pharmacotherapy research is due to a number of factors. First, because of their multiple somatic complaints, patients with generalized anxiety disorder are usually seen by generalists and medical specialists other than psychiatrists; generalized anxiety disorder is more likely to be diagnosed as a comorbid condition in psychiatric practices. Second, pharmacotherapy is considered less effective in generalized anxiety disorder than in some other anxiety disorders. Third, the diagnostic features are not clear-cut, and comorbid conditions make the diagnosis difficult.

The efficacy of benzodiazepines in the pharmacological treatment of generalized anxiety disorder gave rise to theories implicating the benzodiazepine–g-aminobutyric acid (GABA) receptor system in the pathophysiology of generalized anxiety disorder, but evidence exists for the involvement of the serotonergic and noradrenergic systems as well. Benzodiazepines remain the traditional medication choice for patients with generalized anxiety disorder. Data do not support the advantage of any one benzodiazepine over others, and no correlation has been established between clinical response and dosage or plasma concentration. A daily equivalent of 15 to 25 mg of diazepam usually suffices to relieve most symptoms in up to 70 percent of generalized anxiety disorder patients. Both somatic and psychic anxiety symptoms respond within the first week of treatment. Tolerance to the sedative effects of benzodiazepines develops quickly, but the antianxiety effect of a given dosage is well maintained over time in generalized anxiety disorder. However, the relapse rate upon discontinuation of benzodiazepines is high.

Buspirone, the only currently available azapirone, is a potential alternative to benzodiazepine treatment in generalized anxiety disorder. Response rates between 60 and 80 percent have been reported at dosages ranging from 30 to 60 mg a day in three divided doses. While response rates seem comparable, more patients drop out of buspirone trials than benzodiazepine trials. The relative merits of buspirone and benzodiazepines are further detailed under panic disorder. One notable exception is that generalized anxiety disorder patients exposed to benzodiazepines may still be responsive to buspirone unlike panic patients.

Antidepressants are also effective in generalized anxiety disorder patients. Dosages and response patterns are similar to those observed in panic disorder (Table 15.7-1). Increased initial physiological symptoms and anxiety may be related to adverse effects such as dry mouth, constipation, sedation, and positional hypotension rather than to the hypersensitivity syndrome described in panic disorder.

Isolated (but prominent) symptoms such as palpitation, tremor, and sweating may respond to b-blockers within 1 week of treatment, but the full generalized anxiety disorder picture usually requires the use of benzodiazepines, antidepressants, or buspirone. Controlled studies are unavailable, but clinical experience suggests the benefits of combination treatments for generalized anxiety disorder. For instance, the combination of benzodiazepines or b-adrenergic receptor antagonists with antidepressants could yield a rapid response, and when the antidepressant becomes effective, the benzodiazepine or b-adrenergic receptor antagonists can be tapered off. Several new drugs, currently unavailable in the United States, have also been found promising in recent clinical trials. The only double-blind, placebo-controlled study with the partial benzodiazepine agonist abecarnil showed significant improvement on many measures in patients with generalized anxiety disorder. Tropisetrone, a 5-HT_3 antagonist, has good antianxiety efficacy with few, mild adverse effects, and serazepine, a 5-HT_2 antagonist, also worked well. Gepirone and ipsapirone, two compounds related to buspirone, have demonstrated efficacy comparable to that of buspirone.

Maintenance pharmacotherapy of patients with generalized anxiety disorder should follow the principles given above for long-term treatment of anxiety disorders. Given the fluctuating course of this disorder, periodic discontinuation may be attempted.

OBSESSIVE-COMPULSIVE DISORDER

Patients with OCD suffer from recurrent ruminations, the need for repeated performance of useless stereotyped rituals, or both. The initial assessment of patients with OCD should include a detailed description of type, severity, and onset and identification of target symptoms. Comorbid conditions frequently complicating the treatment of OCD include depression, other anxiety disorders, substance use disorders, schizophrenia, bipolar disorders, and personality disorders. OCD is probably the most difficult anxiety disorder to treat and has the highest rate of nonresponse.

Finding that medications that increase serotonergic transmission in the central nervous system (CNS) are efficacious in OCD revolutionized treatment and suggested that the pathophysiology of OCD is related to changes in serotonin function. Subsequently, a series of these medications such as clomipramine, fluvoxamine (Luvox), fluoxetine, paroxetine, and sertraline (Zoloft) have been shown in double-blind controlled trials to alleviate the symptoms of OCD. These drugs are used at relatively high daily dosages (clomipramine, 250 mg; fluvoxamine, 300 mg; fluoxetine, 80 mg; paroxetine 60 mg; sertraline, 200 mg a day or more), although recently the high-dosage strategy is being challenged. Some investigators believe that longer duration of treatment rather than higher dosage is required for response. Ten weeks

of treatment trial is currently the minimum recommendation, but it is not unusual to have treatment continue for several months before peak efficacy is reached. The average response rate to these agents ranges from 40 to 60 percent, but nonresponders to one drug can respond to another. Therefore, several anti–obsessive-compulsive OCD drugs should be tried sequentially before making conclusions about drug resistance. Recent studies suggest that intravenous pulse loading or gradual infusion of clomipramine may convert oral clomipramine–refractory patients into responders.

Based on meta-analysis of effect sizes in placebo-controlled trials, clomipramine is considered more effective than the SSRIs. However, recent direct comparisons failed to confirm this observation. Also, the potential advantage of clomipramine should be weighed against its more troublesome (primarily anticholinergic) adverse-effect profile.

A significant proportion of OCD patients remain refractory to SSRI monotherapy, which suggests diagnostic heterogeneity and also that the psychophysiology of OCD may involve abnormalities other than serotonin dysregulation. Patients with comorbid chronic tic disorders such as Tourette's disorder may represent a subgroup of OCD and may require the combination of SSRI and an antipsychotic drug such as pimozide (Orap) or haloperidol (Haldol). Addition of a low-dosage antipsychotic to an SSRI may also be justified for comorbid schizotypal personality disorder. In general, however, addition of an antipsychotic should only be considered if all other approaches including trials of several SSRIs and cognitive-behavioral therapy have all failed. Since most of these patients require long-term pharmacotherapy, the potential risk of tardive dyskinesia should be weighed against the benefits. The availability of serotonin-dopamine antagonists, with more-favorable adverse-effect profiles, may change this conservative view.

The presence of neurological abnormalities or abnormal electroencephalogram (EEG) may indicate a need for a trial concomitant use of anticonvulsants such as valproate (Depakene) or carbamazepine, while prominent impulsivity may respond to sympathomimetics (psychostimulants) alone or in combination with other antiobsessional drugs. Earlier case reports of the antiobsessional effects of phenelzine were confirmed in one double-blind trial but not in a subsequent controlled study. The efficacy of phenelzine may be limited to patients with symmetry obsessions. Augmentation strategies to enhance the anti–obsessive-compulsive properties of SSRIs have used lithium (Eskalith), buspirone, pindolol (Visken), fenfluramine, 5-hydroxytryptophan, and L-tryptophan. In the absence of controlled studies confirming their efficacy, the addition of most of these agents cannot be recommended as augmenting strategies with scientific certainty. For example, lithium and buspirone augmentation failed in double-blind controlled trials. Case reports suggest that pindolol (2.5 mg three times a day) may help with comorbid depression but has no additive antiobsessional effect.

Relapse rates are high following medication discontinuation. In one study over 90 percent of the patients relapsed after abrupt discontinuation of clomipramine. Medication maintenance treatment seems to be associated with maintained, and even improved, overall functioning. Dosage titration during maintenance treatment could further improve outcome with few late-emergent adverse events.

The most important predictor of poor outcome in OCD is early onset. The presence of schizotypal, borderline, and avoidant personality disorders, and greater number of personality disorders also predict poorer outcome. Severity and duration of illness, sex, age, and type of symptoms have no predictive value.

Psychosurgery remains the last-resort treatment for intractable OCD. The most frequently performed operation, stereotactic subcaudate tractotomies or capsulotomy, intersects the connections between the frontal lobe and the thalamus and exerts its efficacy on the proposed functional imbalance in OCD between the frontal lobe and other parts of the brain. Stereotactic leukotomy seems to decrease glucose metabolism in the orbital part of the frontal lobe. Cingulotomy resulted in significant improvement in 25 to 30 percent of patients previously unresponsive to medication and behavioral treatments. This rate of improvement seems independent of the changes in anxiety and depression scores. Personality changes following psychosurgery are infrequent, but most patients display lower initiative, drive, and energy levels at follow-up. Capsulotomy can lead to perservative behavioral responses. Striatal lesions may lead to the development of substance use disorders following surgery. Phenytoin (Dilantin) usually controls the symptoms in the few cases when seizure develops following surgery.

The few ECT case studies in intractable OCD show considerable benefits up to a 1-year follow-up and suggest that these benefits are independent of changes in measures of depression. These reports must be viewed with caution until confirmed in controlled trials. Preliminary evidence suggests that plasmapheresis, antibiotics, or both may be indicated in a subgroup of children whose obsessive-compulsive symptoms are associated with a specific autoimmune response in the brain.

SOCIAL PHOBIA

Social phobia is an exaggerated fear of negative evaluations, ranging from specific to generalized types. Some consider avoidant personality disorder to be the most severe form of social phobia. Social phobia is a chronic, disabling condition whose prevalence is being increasingly recognized.

Double-blind, placebo-controlled studies confirmed that the most effective medication for the treatment of generalized social phobia is the MAOI phenelzine. Therapeutic dosages of phenelzine range from 45 to 90 mg a day, with response rates ranging from 50 to 70 percent. Approximately 5 to 6 weeks are needed to assess the efficacy of phenelzine. Responders should continue treatment for at least a year. Dietary restriction and numerous troublesome adverse effects put MAOIs at a significant disadvantage compared with less-efficacious but more acceptable medication choices. For instance, the SSRIs have recently emerged as viable alternatives to phenelzine. Although not subjected to the same rigorous evaluations, the results of controlled trials (fluvoxamine, paroxetine, sertraline), a controlled discontinuation study (paroxetine), and open-label case reports (citalopram [Celexa]) with SSRIs suggest that their ease of administration and tolerability more than make up for their slightly lower efficacy.

Most clinicians consider SSRIs the first-line treatment choice for patients with generalized social phobia. Since patients with social phobia do not exhibit the supersensitivity syndrome described in panic disorder, SSRI administration can be initiated at usual antidepressant starting dosages (e.g., fluoxetine, 20 mg; paroxetine, 20 mg) and titrated on the basis of clinical response. The benzodiazepines alprazolam and clonazepam also seem efficacious in both generalized and specific social phobias. Their favorable adverse-effect profile and quick onset of action are countered by lower response rates and earlier relapse upon discontinuation than with phenelzine. As in panic disorder, tolerance does not develop to the therapeutic effects of benzodiazepines. Starting and therapeutic dosages for benzodiazepines are the same as those for panic disorder (Table 15.7-1).

Specific social phobias such as fear of public speaking respond moderately well to b-adrenergic receptor antagonists, although the data are mostly anecdotal. Propranolol (20 to 60 mg per dose) counters several of the physiological symptoms of excessive autonomic arousal. Propranolol's short half-life makes it necessary to repeat the effective dose every 2 to 4 hours. If propranolol works but the social phobic situations are unpredictable during the day, a b-adrenergic receptor antagonists with a longer half-life can be tried. Atenolol (Tenormin), 50 to 100 mg before sleep, should control the symptoms for 24 hours. The adverse effects of b-blockers may include sedation, fatigue, dry mouth, gastric cramping, and occasionally confusion and memory problems. Long-term use can lead to depression. Because of the adverse effects and unpredictability of response, the patient is strongly encouraged to test these drugs before the actual event or performance. The utility of b-adrenergic receptor antagonists in generalized social phobia is much more limited. Preliminary evidence supports the use of buspirone as well. Modest response has been noted in patients with social phobia at dosage above 45 mg a day.

Hyperhidrosis (excessive sweating) may be controlled by topical application of aluminum chloride in alcohol (Drysol) while a generalized form of the symptom usually responds to clonidine or terazosin (Hytrin). Ondansetron (Zofran) is the preferred medication for patients with social phobia if the most disabling symptom is nausea or the fear of vomiting. Nausea may also respond to cisapride (Propulsid), but caution should be exercised when prescribed concomitantly with drugs that inhibit cytochrome P450 (CYP) A4 (CYP 3A4). Paruresis a frequent symptom of social phobia, manifesting as the inability to void in public restrooms, usually requires in vivo exposure and relaxation training. Pharmacological approaches have been largely disappointing but a trial of furosemide and bethanechol may be warranted during behavioral treatment.

The RIMA brofaromine (150 mg a day) was found efficacious in a controlled trial. The finding that most patients maintained their gains at a 9-month follow-up assessment suggests that brofaromine could play an important role in the management of social phobia if it becomes available in the United States.

SPECIFIC PHOBIAS

Specific phobias are quite common and only require treatment if they interfere significantly with the functioning. Treatment is usually behavioral exposure. Medications are used occasionally to alleviate the anticipatory anxiety associated with beginning exposure treatment. Low-dose benzodiazepines and b-adrenergic receptor antagonists can be used for this purpose on an as-needed basis.

POSTTRAUMATIC STRESS DISORDER

Patients with posttraumatic stress disorder (PTSD) by definition have experienced a life-threatening trauma. These patients continue to relieve the experience and complain of flashbacks and vivid dreams of the original trauma. Blunted emotions coupled with a wide range of anxiety, mood, and sleep disorders are frequent manifestations of PTSD.

Prior to being assigned to a separate diagnostic entity, trauma survivors were treated according to the most prominent symptom clusters (e.g., anxiety or depression). Consequently, antidepressants and benzodiazepines were the mainstay of pharmacotherapy, with occasional use of antipsychotics for flashbacks and behavioral discontrol. Interpretation of medication trials in PTSD populations is complicated by a number of factors. First, most early medication trials were conducted in war veterans with chronic PTSD and a number of comorbid conditions such as substance use and severe personality disorders. They tended to have undergone several unsuccessful treatment trials and were considered refractory to psychiatric interventions. It has been suggested recently that the results of these earlier trials may not be applicable to well-diagnosed, previously untreated civilian populations. Civilian trauma survivors seem more responsive to pharmacotherapy than veterans. Second, some studies excluded PTSD patients with comorbid depression. Since this strategy eliminated over 50 percent of the otherwise eligible patients, the clinical relevance of these studies may be quite limited. Third, there is no consensus on how to rate improvement in PTSD. Some studies focus on intrusive thoughts; others on avoidance and hyperarousal. The field is in the process of teasing out medication-responsive symptom clusters and perhaps identifying subgroups of PTSD patients. As biological abnormalities are increasingly recognized in the pathophysiology of PTSD, recent pharmacological trials have begun to target neurotransmitter systems thought to be relevant in PTSD.

Currently, the medication treatment of PTSD is mostly based on clinical experience, case studies, open trials, and retrospective chart reviews; comparatively few controlled trials exist. Almost all antidepressants and antianxiety medications have been tried. It is rather unusual to have an agent show no therapeutic benefit in PTSD, but no medication has covered the full spectrum of PTSD symptomatology. Substantial heterogeneity with regard to type and severity of trauma, developmental phase of the patient at the time of the trauma, comorbid conditions, and psychiatric history may explain relative treatment resistance.

Intrusive thoughts, insomnia, recurring dreams and memories may respond to phenelzine at antidepressant dosages (45 to 90 mg a day). Brofaromine, while not yet available in the United States, may be a good alternative to the MAOIs in PTSD; over half of PTSD patients responded to brofaromine in a double-blind trial. Tricyclic drugs such as amytriptyline (Elavil) or imipramine may improve depressive and anxiety symptoms but have modest effect on intrusive thought. Global improvement is evident even in the absence of prominent depression, but it may not be observable before 8 weeks of treatment. Benzodiazepines, such as alprazolam, have only modest effect on anxiety, and the risk of dependence and abuse potential contraindicate their use in PTSD patients with substance use disorders. Numbing and hyperarousal in civilians but not in veterans, and impulse control, labile affect, and disturbed interpersonal relationships in both populations may respond to fluoxetine in dosages of 60 mg a day and above. Antipanic, low starting dosages of fluoxetine usually circumvent the initial agitation in patients with prominent anxiety; paroxetine may be less activating than fluoxetine in these patients. Fluvoxamine and sertraline reduce hyperarousal, intrusion, and explosiveness. Since SSRIs may be the only medications targeting the full symptom spectrum of PTSD, they are usually considered the first-line pharmacological treatment.

While not considered first-line approaches, clonidine (Catapres) buspirone, and high-dosage propranolol may be used alone or as adjuncts to an SSRI or a tricyclic drug for hyperarousal, nightmares, and flashbacks. Refractory impulsivity and aggressive outbursts in combat veterans may be treated with lithium, carbamazepine, or valproate, alone or in combination with an SSRI. Mood stabilizers have not been tested in civilian PTSD populations.

In the absence of data about optimal duration of pharmacotherapy and relapse upon discontinuation, the recommendations should follow those given for other chronic anxiety and affective disorders. Medication treatment should continue for at least 1 year. Given the considerable uncertainty surrounding the illness, flexibility in tailoring medication combinations according to individual patients is recommended, and psychotherapy should always be part of the management of PTSD.

SUGGESTED CROSS-REFERENCES

Generic discussion of medications used in anxiety disorders can be found in Chapter 31. Side effects, interactions, pharmacodynamics, and pharmacokinetics are described in Section 31.1. The neurobiology of anxiety disorders and the hypothesized basis of the efficacy of antianxiety medications is given in Section 15.3. The serotonin syndrome, encountered when several serotonergic agents are combined, is described in Section 31.25. Detailed diagnostic description of various anxiety disorders is given in Section 15.6. The epidemiology of anxiety disorders is discussed in Section 15.2.

SECTION REFERENCES

*American Psychiatric Association: Practice guideline for the treatment of patients with panic disorder. Am J Psychiatry *155*(Suppl):1, 1998.

Coplan JD, Papp LA, Pine D, Martinez JM, Cooper T, Rosenblum LA, Klein DF, Gorman JM: Clinical improvement with fluoxetine therapy and noradrenergic function in patients with panic disorder. Arch Gen Psychiatry *54:*643, 1997.

Expert consensus guideline for the treatment of obsessive-compulsive disorder. J Clin Psychiatry *58*(Suppl):2, 1997.

Fallon BA, Liebowitz, MR, Campea R, Schneier FR, Marshall R, Javier S, Goetz D, Klein DF: Intravenous clomipramine for obsessive-compulsive disorder refractory to oral clomipramine. Arch Gen Psychiatry *55:*918, 1998.

*Greist JH, Jefferson JW, Kobak KA, Wenzel KW, Bailey TM: Efficacy and tolerability of serotonin transport inhibitors in obsessive-compulsive disorder: A meta-analysis. Arch Gen Psychiatry *52:*53, 1995.

Irle E, Exher C, Thieleu K, Weniger G, Rüther E: Obsessive-compulsive disorder and ventromedial frontal lesions: Clinical and neuropsychological findings. Am J Psychiatry *155:*255, 1998.

Jenike MA, Baer L, Minichiello WE, Ranch SL, Buttolph ML: Placebo-controlled trial of fluoxetine and phenelzine for obsessive-compulsive disorder. Am J Psychiatry *154:*1261, 1997.

*Katschnig H, Amering M, Stolk JM, Ballenger JC, Briggs A, Buller R, Cassano G, Garvey M, Roth M: Long-term follow-up after a drug trial for panic disorder. Br J Psychiatry *167:*487, 1995.

Liebowitz MR, Schneier F, Campeas R, Hollander E, Hatterer J, Fyer A, Gorman JM, Papp LA, Davies S, Gully R: Phenelzine vs atenolol in social phobia: A placebo controlled comparison. Arch Gen Psychiatry *49:*290, 1992.

*Marshall RD, Klein DF: Pharmacotherapy in the treatment of posttraumatic stress disorder. Psychiatr Ann *25:*588, 1995.

Papp LA, Schneier FR, Fyer AJ, Liebowitz MR, Gorman JM, Coplan JD, Saoud J, Campeas R, Fallon B, Jusino C, Klein DF: Clomipramine treatment of panic disorder: Pros and cons. J Clin Psychiatry *154:*1557, 1997.

Papp LA, Sinha SS, Martinez JM, Coplan JD, Amchin J, Gorman JM: Low dose venlafaxine treatment in panic disorder. Psychopharmacol Bull *34:*207, 1998.

Pigott TA, Pato MT, Bernstein SE, Grover GN, Hill JL, Telliver TJ, Murphy DL: Controlled comparisons of clomipramine and fluoxetine I in the treatment of obsessive-compulsive disorder: Behavioral and biological results. Arch Gen Psychiatry *47:*926, 1990.

*Pigott TA, Seay SM: A review of the efficacy of selective serotonin reuptake inhibitors in obsessive-compulsive disorder. J Clin Psychiatry *60:*101, 1999.

Pohl RB, Wolkow RM, Clary CM: Sertraline in the treatment of panic disorder: A double-blind multicenter trial. Am J Psychiatry *155:*1189, 1998.

Price JS, Waller PC, Wood SM, MacKay AV: A comparison of the postmarketing safety of four selective serotonin re-uptake inhibitors including the investigation of symptoms occurring on withdrawal. Br J Clin Pharmacol *42:*757, 1996.

*Rickels K, Downing R, Schweizer E, Hassman H: Antidepressants for the treatment of generalized anxiety disorder. Arch Gen Psychiatry *50:*884, 1993.

Schweizer E, Rickels K, Lucki I: Resistance to the antianxiety effect of buspirone in patients with a history of benzodiazepine use. N Engl J Med *313:*719, 1986.

15.8 ANXIETY DISORDERS: PSYCHOLOGICAL TREATMENTS

LAWRENCE A. WELKOWITZ, PH.D.

Since the late 1980s there has been tremendous progress in the nonpharmacological treatment of anxiety disorders. Cognitive-behavioral therapies, which reflect a recent integration of the cognitive theories and methods associated with Aaron Beck, Albert Ellis, and David Clark, and the behavioral theories and methods of B.F. Skinner and Ivan Pavlov, currently have the greatest degree of empirical validation. Panic disorder, which has been said to be the most disabling of the anxiety disorders in terms of social and occupational functioning is a case in point in that recently developed treatments contain elements based on classical or Pavlovian conditioning, behavioral techniques of exposure, and cognitive restructuring of irrational beliefs and overvalued ideas. Other anxiety disorders for which cognitive-behavioral approaches have been particularly effective include social phobia (both generalized and nongeneralized type), obsessive-compulsive disorder, specific phobia, social phobia, and generalized anxiety disorder. Further validation is needed for two recently developed treatments for posttraumatic stress disorder, one based on exposure methods, and the other based on classical conditioning *(eye movement desensitization* and *reprocessing* [EMD/R]). Psychosocial (e.g., supportive and psychodynamic) therapies have also been used in the treatment of anxiety disorders.

COGNITIVE-BEHAVIORAL THERAPY

Panic Disorder Cognitive-behavioral treatment of panic disorder or *panic control therapy* produces 80 to 90 percent panic-free status, at least within 6 months of treatment. Two-year follow-up indicates that more than 50 percent of those patients who originally responded to panic control therapy have occasional panic attacks, and more than a quarter will seek additional treatment. Nonetheless, these treatment responders do tend to have a significant decline in panic-related symptoms and most maintain many of their treatment gains. A large multi-center study comparing medication and panic control therapy treatment being carried out at the University of Pittsburgh, Yale University, State University of New York at Albany, and Columbia University should help to resolve many questions about long-term gains in panic control therapy and pharmacological treatment.

Panic control therapy is an amalgam of techniques drawn from a variety of cognitive and behavioral methods and based largely on Peter Lang's three-system model of cognitive psychophysiology, which focuses on the interactions of three systems affecting the human experience of panic and anxiety: physiological (e.g., palpitation, sweating, dizziness, nausea), cognitive (fears of losing control or going crazy), and behavioral (avoidance, pacing). Panic control therapy techniques are thus directed towards each of these systems: breathing control techniques are designed to control the physiological effects of hyperventilation and progressive muscle relaxation helps reduce the overall negative effects of muscle tension on anxiety. Cognitive restructuring techniques, which focus on catastrophic thinking errors (e.g., “I will have a heart attack and die”) as well as on misinterpretation of harmless bodily associations, are directed towards problems in the cognitive system. Education about the nature of anxiety and panic, which also counters the myths of panic attacks, (i.e., that they are associated with schizophrenia or heart disease) also serves as a form of stress inoculation in the cognitive domain. Exposure to feared situations and events, such as crowded public areas, helps to reduce avoidance symptoms of the behavioral domain. A novel technique called *interoceptive exposure* involves repeated exposure to the physical sensations associated with panic. This method, which is based on a classical or Pavlovian conditioning model of panic, consists of a series of activities, such as deliberately overbreathing (to produce physical effects of hyperventilation), breathing through a straw while holding one's nose (feelings of breathlessness), spinning in a chair (dizziness), running in place (increased heart rate), staring at one's hand (feelings of unreality), and so on. These activities are rehearsed until an habituation of the anxiety response has been achieved.

In clinical practice the degree to which various techniques are emphasized may depend on particular features of the panic disorder. For example, a patient with a high degree of agoraphobic avoidance may require additional exposure to fearful situations. The house-bound person with agoraphobia may require therapist-assisted in vivo exposure therapy as opposed to less labor intensive and more instruction-oriented exposure therapy for the patient with minimal agoraphobia and panic disorder. These panic control therapy techniques may also be carried out successfully by therapists whose primary mode of treatment is pharmacological. Finally, further research is needed to identify effective components of treatment and to determine whether certain subtypes of patients may require certain components of panic control therapy over others. In clinical practice, for example, it appears that a substantial number of patients with panic disorder respond well to education and breathing control alone whereas others seem to require more extensive treatment.

> José was a 27-year-old laboratory technician who began having full-blown panic attacks 8 months prior to seeking help at our research clinic. While he was unable to identify specific situations that elicited attacks, he was particularly concerned about the possibility of them occurring while he was engaged in laboratory procedures with patients. His attacks typically involved a sudden explosion of autonomic arousal and included palpitations, sweating, dizziness, feelings of unreality, and tingling in his arms and legs. He dreaded the idea that the attacks might reoccur. In the beginning of his cognitive-behavioral program, he found an educational handout that described the myths of panic attacks (e.g., that they will lead to heart attacks, losing control, or going crazy) particularly reassuring. He began practicing diaphragmatic breathing each evening, and after several weeks, became effective in challenging his negative way of thinking about the consequences of panic attacks. In the latter few weeks of his 12-week program, he practiced exposing himself to physical sensations of panic by doing a variety of interoceptive exercises at home, including hyperventilating for 1 or 2 minutes at a time (designed to help José acclimate to the physical sensations associated with overbreathing), and spinning in a chair repeatedly (designed to help acclimate him to symptoms of dizziness and feelings of unreality). At the conclusion of the treatment program José's panic attacks had disappeared, and at 6-month follow-up he had maintained his treatment gains by attending “booster sessions” with his therapist once every 2 months.

In contrast to cognitive-behavioral formulations, it has been reported that patients who develop panic symptoms in their late 20s have histories replete with relationship problems as well as other neurotic symptoms stemming from parent-child disruptions. These patterns, in turn, are said to lead to disturbances in empathy and understanding, which may ultimately elicit panic symptoms. Recently, detailed treatment protocols have been developed that are based on psychodynamic formulations of anxiety, consisting primarily of supportive-expressive psychodynamic psychotherapy. Preliminary outcome research indicates that these psychodynamically oriented protocols are useful in modifying anxiety, depression, and related interpersonal problems. Studies on hypnotherapy and hypnoanalysis, in which early childhood conflicts are identified and ultimately tied to the origins of panic and anxiety, provide some support for a more traditional analytic approach.

The importance of cultural issues in the treatment of anxiety disorders should not be overlooked. African-Americans with panic disorder appear to respond less with panic control therapy compared with whites. This may be related to the high prevalance of untreated hypertension among African-Americans, which may, in turn, exacerbate anxiety symptoms or possibly lead to higher comorbidity of emotional problems. Conversely, it has been observed that Asian Americans and Native-Americans, whose philosophical systems emphasize the control of mind over body and who have also utilized breathing control methods for centuries, are particularly receptive to the breathing and cognitive control aspects of panic control therapy. Interestingly, data from the Epidemiologic Catchment Area (ECA) study did not detect any differences in prevalence of panic disorder between African-American, Hispanic, and white groups.

Social Phobia As with panic disorder, considerable progress in the psychological treatment of social anxiety or social phobia is linked to the application of cognitive-behavioral methods. Unlike more traditional psychotherapies, cognitive-behavioral approaches do not address the origins of social anxiety, but instead focus on the use of coping strategies that can be implemented in current fearful situations. The most thoroughly studied form of cognitive-behavioral therapy for social phobia is a group therapy consisting of several discrete entities including (1) presentation of a three-system (cognitive-behavioral-physiological) model of social anxiety; (2) training in identification and restructuring of irrational beliefs regarding social performance; (3) in-session exposure to feared social situations via group role-playing scenarios; and (4) homework assignments directing patients to utilize cognitive and exposure techniques in vivo). Groups are particularly amenable to the treatment of social phobia in that they provide natural opportunities for patients to practice feared behaviors in a supportive and informative context.

Outcome research is somewhat limited but one study showed that cognitive-behavioral group therapy was nearly twice as effective as standard educational-supportive group psychotherapy. Responders to cognitive-behavioral group therapy were also shown to maintain treatment gains to a considerable extent

at 5-year follow-up. Questions remain as to the effective treatment component in these therapies that blend cognitive and exposure-based methods. For example, it is unclear whether exposure to feared situations alone, without cognitive therapy, would be just as effective as the combined treatment. Also, it is not known whether group therapies other than educational or supportive group therapy, such as interpersonal group therapy, may be effective in treating social phobia.

In terms of the differential effects of cognitive-behavioral group therapy on patients of varying subtypes (generalized versus nongeneralized social phobia), it appears that patients with more discrete or nongeneralized fears, such as of public speaking, respond best to these type of treatments. Pharmacological research has also shown differential effects based on subtype, with beta-blockers being more effective for nongeneralized fears, and antidepressant agents such as the monoamine oxidase inhibitor (MAOI) phenelzine (Nardil), being more effective for generalized social phobia. For generalized as well as more severe forms of social phobia, a combination of psychological and pharmacological therapy may be in order. Alternatively, it may be that longer-term cognitive-behavioral therapy, with or without medication, is warranted.

Despite the lack of empirical validation via controlled studies on large samples, social phobia is commonly treated with other behavioral methods. Systematic desensitization, first described in 1958, is based on a classical conditioning model of fears. It involves a pairing of previously conditioned stimuli (e.g., hierarchical presentation of images of feared social situations) with a relaxed state (promoted by means of progressive muscle relaxation). This new pairing is said to create a new set of associations that lead the patient to no longer fear previously feared stimuli. Regardless of the specific behavioral technique utilized, the key mechanism of effective treatment is related to the development of self-efficacy, or a sense of confidence that a patient experiences when confronting challenging social events or situations.

Jane was a 31-year-old marketing executive at a rapidly expanding technology company. She contacted our research clinic for help with her fear of public speaking, which had become so severe in recent months that she was at risk for losing her job, which required a considerable amount of public presentation and leading training programs. Jane was assigned to a treatment program that entailed a combination of medication (b-adrenergic receptor antagonist the beta-blocker atenolol [Tenormin]) and a 12-week group-based cognitive-behavioral treatment program for nongeneralized social phobias, such as fear of public speaking. Before beginning the group program, Jane met with her behavior therapist who helped her develop a hierarchy of the public-speaking situations she feared. Difficulty levels ranged from speaking up at meetings involving five or fewer colleagues, which caused a low level of anxiety, to leading a training program for 100 or more employees and colleagues at annual corporate gatherings, which caused a very high level of anxiety. In her group treatment, Jane began by identifying and challenging negative thoughts associated with each situation (e.g., “They won't understand what I am saying” became “I have always explained myself clearly in the past”). The other group members then assisted Jane in designing and carrying out role-plays that allowed her to practice employing her new coping responses and helped her to develop new public speaking skills. By the end of treatment, Jane still experienced occasional anxiety when making a public presentation to large numbers people, but she no longer avoided such situations.

Finally, it is important to consider cultural factors when treating social phobia. In Japan, a condition called *Taijin Kyofusho* refers to a type of social anxiety in which the sufferers become preoccupied with the idea that they are causing others to be embarrassed or humiliated. Other *Taijin* fears include fear of emitting a body odor and fear of one's facial expression becoming too stiff. In utilizing cognitive therapy with Japanese and other Asian patients, it is important to consider general cultural characteristics, such as the importance of displaying hierarchical deference (i.e., respect for those in authority) as well as an emphasis on communal versus individual expression. The emphasis on rigid moral codes and religious ritual and custom in Arab culture is said to place individuals at higher risk of social phobia, reflecting an underlying fear of betraying valued ideas and formalities. It has been reported that 13 percent of all outpatient psychiatric visits (excluding for schizophrenia) in Saudi Arabia are for treatment of social phobia.

Posttraumatic Stress Disorder Treatment of posttraumatic stress disorder is based largely on the aversive effects of a traumatic event or series of traumatic events. Specifically, fears related to this disorder are developed via classical or Pavlovian conditioning in which the unconditioned stimulus (e.g., a horrible accident or sexual assault) that automatically produces a fear response or unconditioned response is paired with other stimuli (e.g., the sounds, smells, and sights associated with the traumatic event). These conditioned stimuli then come to elicit a fear response independent of the original unconditioned stimuli. According to O. Hobart Mowrer's two-factor theory these new classically conditioned fears are subsequently maintained via operant conditioning in that escape or avoidance behaviors are reinforced by their anxiety-relieving effects. Through a process of *higher-order conditioning*, increasing numbers and types of stimuli may become elicitors of an anxiety or a posttraumatic stress disorder response. For example, if particular smells present during the original trauma are later present in more innocuous situations, such as when walking down the street or visiting a friend or relative, the patient may eventually evidence a fear response.

It has been argued that Mowrer's two-factor theory only partially explains the cognitive, affective, and physiological response unique to posttraumatic stress disorder. Symptoms such as trauma-related nightmares and psychological numbing are believed to be best understood by cognitive network theories, which posit the development of a “fear network” of memories. These cognitive networks, which include cognitive, behavioral, and physiological information about trauma responses, are formed following a traumatic event and are activated by trauma-related stimuli. According to this model, posttraumatic stress disorder responses can change across time as the *interpretation* of original trauma events produces changes in its meaning. Thus, a delayed response in this disorder can be explained by a subsequent change in the personal meaning of the original trauma, such as when an originally asymptomatic woman who has been sexually assaulted later discovers that her assailant has been released from prison and has been stalking previous victims.

Regardless of the emphasis on cognitive or conditioning processes, empirically validated treatments of posttraumatic stress disorder employ similar components, including (1) cognitive monitoring and restructuring, (2) exposure in imagination to the original traumatic event, and (3) exposure in vivo to trauma-related stimuli. The overall goal is to reduce or eliminate the patient's anxiety reaction to trauma-related cognitive and environmental stimuli. Cognitive restructuring is based on the notion that fear networks or structures can be directly modified by identifying and modifying negative or nonproductive thinking patterns related to concerns about threat or danger. Cognitive therapy may proceed in traditional fashion, based largely on therapist-patient dialogue, or may utilize homework-based written exercises in which the patient records irrational beliefs and composes more adaptive, coping-oriented responses that are based on new interpretation of the meaning of these trauma-related events. At least one study has found this cognitive processing therapy to be more effective than a no-treatment control group in reducing the symptoms of posttraumatic stress disorder.

In vivo and imaginal exposure techniques, which involve graduated exposure either to real life or cognitive trauma-related stimuli, are clearly useful in reducing overall anxiety as well as trauma-related fears. Both imaginal and in vivo techniques may involve return to the original place where the traumatic event occurred, or may be limited to situations that produce anxiety in the present (e.g., approaching certain people or places that evoke trauma-related memories). These methods have been shown to be effective for combat veterans suffering from the disorder and for women who have been sexually assaulted. More recently, a 4-week program has been developed to prevent the symptoms of this disorder for women who have been recently raped or assaulted. This program consists of (1) education about common reactions to trauma, (2) relaxation skills, (3) exposure in imagination to the original trauma, and (4) cognitive restructuring. A preliminary study comparing 10 women participants and 10 nonparticipants showed that 70 percent of the untreated women and only 10 percent of the treated women suffered symptoms of posttraumatic stress disorder at 2 months following the original trauma.

Jennifer was a 21-year-old college student who sought treatment in the anxiety clinic after being raped repeatedly at a fraternity party. She had difficulty immediately remembering details of the rape because she had ingested an unknown recreational drug that had been slipped into the punch bowl at the party. In the days following the rape, she slowly began to remember details of the traumatic event, often in the form of flashbacks and nightmares. She also experienced increased generalized anxiety, feelings of unreality, and an intense fear reaction to stimuli associated with the attacks, such as the sight of college men wearing sweat shirts with fraternity names and logos, as well as certain odors present during the party. Her treatment program was administered one-on-one by a psychologist who specialized in posttraumatic stress disorder. In the first few sessions, Jennifer was educated about the disorder and the nature of anxiety and was provided a rationale for an exposure-based treatment program. Ensuing sessions involved extensive discussion about the details of the traumatic event, as well as in-session exposures in-imagination to the original trauma. Exposure-based homework assignments complemented these in-session activities and were designed to treat Jennifer's new tendency to avoid trauma-related situations, including campus parties and other social gatherings. Slowly, Jennifer experienced decreased reactivity to trauma-related stimuli and began to enter situations she had previously avoided. After approximately twenty-five sessions Jennifer's symptoms were almost gone and she had nearly returned to normal in terms of participating in campus-related social events and parties. Despite her gains, Jennifer remains, in her words, “more careful” about certain high-risk social events (e.g., she will rarely enter a fraternity house), and finds that her trust in men is markedly diminished. A previous supporter of fraternities on campus, Jennifer is now an ardent advocate of removing fraternities from all colleges.

Unfortunately, little has been written about cross-cultural aspects of treatment of posttraumatic stress disorder. This is surprising in light of its particular relevance for minority groups in the United States, such as African-Americans from socially disadvantaged groups who witness a great deal of violence in their communities. For example, anxious black children, compared with their non-anxious peers, were found to experience a significantly greater degree of trauma. Observers of

Caribbean-American immigrants have noted an association between acculturation and symptomatology of posttraumatic stress disorder. These researchers support the use of group supportive treatments of the disorder for this ethnic group, but warn against the practice of mixing patients with those whose disorder is of differing causes (e.g., sexual assault versus acculturation).

Generalized Anxiety Disorder Recent data from the National Anxiety Disorders Screening Day (NADSD) revealed a surprisingly high 1-month prevalence (11.5 percent) of generalized anxiety disorder for those participants in an anxiety screening program who met criteria for one anxiety disorder only as compared with 5.4 percent for panic disorder, 3 percent for posttraumatic stress disorder, 3.4 percent for social phobia, and 0.7 percent for obsessive-compulsive disorder. Generalized anxiety disorder is particularly prevalent among those with multiple anxiety problems; For example, in the NADSD sample 10.2 percent met the criteria for panic disorder and generalized anxiety disorder, and 7.9 percent met criteria for panic, social phobia, and generalized anxiety disorder. The latter may therefore be considered a hidden problem in that other comorbid anxiety problems may supercede it in terms of focus of treatment.

Cognitive-behavioral formulations and treatment programs dominate this area, despite the lack of data on etiology or psychotherapy outcome. Debate continues as to whether generalized anxiety disorder should be considered a discrete anxiety disorder at all, with some arguing that high comorbidity with other anxiety and depressive disorders suggests that it is merely a basic anxiety trait that serves as a platform for the development of other related problems. Citing the inability of the Center for Epidemiologic Studies Depression Scale (CES-D) to distinguish between depression and generalized anxiety disorder, at least in a sample of mothers of handicapped children, symptoms were noted as being, in some circumstances, a natural component of more severe depression. Data from the ECA study revealed that generalized anxiety disorder was more common among people of lower socioeconomic status as well as in African-Americans, thus suggesting a sociocultural factor in etiology.

Psychological theories and treatment programs alike have tended to focus on the issue of chronic and excessive worrying among individuals who have generalized anxiety disorder. Unlike those suffering panic attacks who experience catastrophic-type cognition (e.g., “I'm going to have a heart attack and die”), those with generalized anxiety disorder exhibit less emotionally charged cognitions on an ongoing basis (e.g., “I'll never make it through the day,” or “I'm sure something will go wrong”). Excessive worriers actually experience less physiological arousal (e.g., decreased heart rate) in response to a fearful stimuli as compared with nonworriers, and it has been speculated that generalized anxiety disorder may actually serve as a type of defense against response to more fearful events.

Treatment programs to date contain one or both of the following components: (1) *cognitive restructuring* designed to identify worry-related thoughts and replace them with more positive coping responses and (2) *relaxation training* designed to reduce excessive physiological arousal. While controlled studies have shown that these methods are superior to either nondirective therapies or no-treatment controls, identification of the most effective treatment components remains unclear as does the superiority of these methods to other types of treatments, such as interpersonal-oriented therapies or other more alternative treatments, such as biofeedback or meditation.

A large-scale study comparing cognitive restructuring, relaxation training, a combined cognitive plus relaxation program, and a no-treatment group revealed improvements in all three active treatments relative to the control condition. While responders were remarkably successful in maintaining gains at 2-year follow-up, no differences were found among the three active treatments. Additionally, study dropouts and nonresponders remained symptomatic at follow-up. In similar fashion, a study comparing relaxation training, a combined cognitive therapy with relaxation training, and a nondirective psychotherapy demonstrated the superiority of the cognitive-behavioral programs compared to a treatment consisting of Rogerian-style reflecting about worrisome thoughts. A longer-term follow-up indicated the slight superiority of the combined cognitive-behavioral treatment over the relaxation-alone condition.

Melinda's friends and family refer to her as a “worrywart.” She constantly worries about something, whether it be that the chicken she is cooking may have salmonella, that her children may not be safe while driving to school, or that her family will not have enough money to pay for expenses. As Melinda put it during her initial evaluation, “I never seem to stop worrying.” In addition to her chronic worrying, Melinda experiences ongoing physical discomfort, including headaches, nausea, sweating, and occasional shortness of breath. She does not, however, experience full-blown panic attacks. Her treatment consisted of a combination of relaxation (specifically, progressive muscle relaxation practiced on a twice-daily basis) and cognitive therapy (identifying and challenging worry-related thoughts and replacing them with more productive coping responses). For example, she applied a type of “probability analysis” to her constant fear that her children would be harmed in a car accident and produced a coping response along the lines of “It is unlikely that they will be in an accident” as well as “My son is an excellent driver.” Following completion of treatment, Melinda reported that her family saw a tremendous improvement in terms of her chronic worrying. While worry-type thoughts would still occasionally enter her consciousness, she now felt that she had effective tools for dealing with them.

PSYCHOSOCIAL THERAPY

Supportive and psychodynamic therapies have been widely used in the treatment of generalized anxiety disorder, panic disorder, and posttraumatic stress disorder, although further research is needed to establish their efficacy.

According to the 1998 American Psychiatric Association's Practical Guidelines for the Treatment of Patients with Panic Disorder, psychosocial therapies may be of use. They have some application to the other anxiety disorders discussed in this section. A summary of the relevant areas of the report follows.

Psychodynamic Psychotherapy Psychodynamic psychotherapy is based on the concept that symptoms result from mental processes that may be outside of the patient's conscious awareness and that elucidating these processes can lead to remission of symptoms. Moreover, in order to lessen the patient's vulnerability to panic, the psychodynamic therapist considers it necessary to identify and alter core conflicts. The goals of psychodynamic psychotherapy may be more ambitious and require more time to achieve than those of a more symptom-focused treatment approach. There are some case reports of brief dynamic psychotherapies that took no longer than cognitive-behavioral therapy to achieve reasonable treatment goals for patients with panic disorder.

In psychodynamic psychotherapy, the successful emotional and cognitive understanding of the various elements of psychic conflict (impulses, conscience, internal standards that are often excessively harsh, psychological defense patterns, and realistic concerns) and reintegration of these elements in a more realistic and adaptive way may result in symptom resolution and fewer relapses. To achieve this insight and acceptance, the therapist places the symptoms in the context of the patient's life history and current realities and extensively uses the therapeutic relationship to focus on unconscious symptom determinants.

Combined Treatments Investigators have examined use of the combination of medication and cognitive behavior therapy for patients with panic disorder and agoraphobia. Several short-term treatment studies have shown that the combination of the tricylic medication imipramine (Imipramine) with one component of cognitive behavior therapy, behavioral exposure, may be superior to either treatment alone. Another study showed that selective serotonin reuptake inhibitors, such as paroxetine (Paxil), plus cognitive therapy worked significantly better for patients with panic disorder than cognitive therapy plus placebo. There has been one study of the combination of psychodynamic psychotherapy with medication. This study suggested that psychodynamic psychotherapy may improve the long-term outcome of medication-treated patients.

Group Therapy Reports on group therapy in the treatment of panic disorder have consisted primarily of cognitive behavioral approaches; the improvements with group cognitive behavior therapy were comparable to those in studies of individually administered cognitive behavior therapy and pharmacological treatment.

Patient Support Groups Patient support groups are very helpful for some patients suffering from panic disorder. Patients have the opportunity to learn that they are not unique in experiencing panic attacks and to share ways of coping with the illness. Support groups may also have a positive effect in encouraging patients to confront phobic situations. Finally, family members of patients with panic disorder may benefit from the educational aspects of patient support groups. In deciding to refer a patient to a support group, however, it is imperative that the psychiatrist obtain information about the nature of the group and the credentials of its leader. Support groups are not a substitute for effective treatment but are complementary to it.

FUTURE DIRECTIONS

There has been significant progress in recent years in the psychological treatment of anxiety disorders. Most of this progress can be linked to research on cognitive behavioral therapies, which are symptom-directed and relatively short-term. A common feature of all these treatments is active exposure to distressing stimuli, either in vivo or in imagination. These exposures vary depending on the type of anxiety disorder (e.g., the patient with social phobia is exposed to distressing social stimuli, such as parties or public-speaking events). Psychological treatments also include various forms of cognitive therapy that target irrational ideas emblematic of a particular disorder (e.g., the catastrophic fears of a patient with panic disorder). In all cases the goal is to assist the patient to endure discomfort by regularly practicing

stress-reducing techniques.

SUGGESTED CROSS-REFERENCES

The other sections of Chapter 15 describe the various aspects and areas of study of anxiety disorders. Cognitive therapy is presented in Section 30.6 and behavioral therapy is covered in Section 30.2.

SECTION REFERENCES

*Amir M, Kaplan Z, Efroni R, Kotler M: Suicide risk and coping styles in posttraumatic stress disorder patients. Psychother Psychosom *68:*76, 1999.

Bandura A: *Social Foundations of Thought and Acquisition: A Social Cognitive Theory*. Prentice Hall, Englewood Cliffs, NJ, 1986.

Barlow DH, Craske MG, Cerny JA, Klosko JS: Behavioral treatment of panic disorder. Behav Ther *20:*261, 1989.

*Barlow DH, Rapee RM, Brown TA: Behavioral treatment of generalized anxiety disorder. Behav Ther *23:*551, 1992.

Blazer DG, Hughes D, George LK, Swartz M, Boyer R: Generalized anxiety disorder. In *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*, LN Robins, DA Regier, editors. Maxwell MacMillan International, New York, 1991.

Borkovec TD, Abel JL, Newman H: Effects of psychotherapy on comorbid conditions in generalized anxiety disorder. J Consult Clin Psychol *63:*479, 1995.

Borkovec TD, Hu S: The effect of worry on cardiovascular response to phobic imagery. Behav Res Ther *28:*69, 1990.

Boudewyns PA, Hyer L: Physiological response to combat memories and 5 preliminary treatment outcomes in Vietnam veteran PTSD patients treated with direct therapeutic exposure. Behav Ther *21:*63, 1990.

Breslau N: Depressive symptoms, major depression and generalized anxiety. A comparison of self-reports on the CES-D and results from diagnostic interviews. Psychiatry Res *15:*219, 1985.

Brown TA, Barlow DH: Long-term outcome in cognitive-behavioral treatment of panic disorder: Clinical predictors and alternative strategies for assessment. J Consult Clin Psychol *63:*754, 1995.

Calhoun KS, Resnick PA: Post-traumatic stress disorder. In *Clinical Handbook of Psychological Disorders*, ed 2, Barlow DH, editor. Guilford, New York, 1994.

Chaleby K: Social phobia in Saudis. Soc Psychiatry *22:*167, 1987.

Chambless DL, Williams KE: A preliminary study of African Americans with agoraphobia: Symptom severity and outcome of treatment with in vivo exposure. Behav Ther *26:*501, 1995.

Craske M, Barlow DH: Panic disorder and agoraphobia. In *Clinical Handbook of Psychological Disorders: A Step-By-Step Treatment Manual*, ed 2, DH Barlow, editor. Guilford, New York, 1993.

Creamer M, Burgess P, Pattison P: Reactions to trauma: A cognitive processing model. J Abnorm Psychol *101:*452, 1992.

Crits-Christoph P, Connolly MB, Azarian K, Crits-Christoph K: An open trial of brief supportive-expressive psychotherapy in the treatment of generalized anxiety disorder. Psychotherapy *33:*418, 1996.

DelMonte MM: The use of hypnotic regression with panic disorder: A case report. Aust J Clin Hypnother Hypn *17:*1, 1996.

Foa EB, Steketee G, Rothbaum BO: Behavioral/cognitive conceptualizations of post-traumatic stress disorder. Behav Ther *20:*155, 1989.

Foa EB, Rothbaum BO, Riggs DS, Murdock TB: Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive-behavioral procedures and counseling. J Consult Clin Psychol *59:*715, 1991.

Foa EB, Zinbarg R, Olasov-Rothbaum B: Uncontrollability and unpredictability in post-traumatic stress disorder: An animal model. Psychol Bull *112:*218, 1992.

Foa EB, Hearst-Ikeda D, Perry KJ: Evaluation of a brief cognitive-behavioral program for the prevention of chronic PTSD in recent assault victims. J Consult Clin Psychol *63:*948, 1995.

Friedman S, Paradis C: African-American patients with panic disorder and agoraphobia. J Anxiety Disord *5:*35, 1991.

Gopaul-McNicol S, Brice-Baker J: Caribbean Americans. In *Cultural Issues in the Treatment of Anxiety*, S Friedman, editor. Guilford, New York, 1997.

*Gorman JG, Shear K: Practice guidelines for the treatment of patients with panic disorder. Am J Psychiatry *155*(Suppl):1, 1998.

Heimberg RG, Dodge CS, Hope DA, Kennedy CR, Zollo L, Becker RE: Cognitive behavioral group treatment of social phobia: Comparison with a credible placebo control. Cogn Ther Res *14:*1, 1990.

Heimberg RG, Salzman D, Holt CS, Blendell K: Cognitive behavioral group treatment of social phobia: Effectiveness at 5-year follow-up. Cogn Ther Res *17:*325, 1993.

Heimberg RG, Juster HR: Treatment of social phobia in cognitive behavioral groups. J Clin Psychiatry *55*(Suppl):38, 1994.

Horwath E, Johnson J, Hornig CD: Epidemiology of panic disorder in African-Americans. Am J Psychiatry *150:*465, 1993.

Lang PJ: The cognitive psychophysiology of emotion: Fear and anxiety. In *Anxiety and the Anxiety Disorders*. AH Tuma, JD Maser, editors. Erlbaum, Hillsdale, NJ, 1985.

*Liebowitz MR, Schneier F, Campeas R, Hollander E, Hatterer J, Fyer A, Gorman J, Papp L, Davies S, Gully R, Klein DF: Phenelzine vs atenolol in social phobia: A placebo-controlled comparison. Arch Gen Psychiatry *48:*938, 1992.

Margraf J, Barlow D, Clark D, Telch M: Psychological treatment of panic: Work in progress on outcome, active ingredients, and follow-up. Behav Res Ther *31:*1, 1993.

Markowitz J, Weissman MM, Ouellette R, Lieb R: Quality of life in panic disorder. Arch Gen Psychiatry *46:*984, 1989.

Massion A, Warshaw M, Keller M: Quality of life and psychiatric morbidity in panic disorder and generalized anxiety disorder. Am J Psychiatry *150:*600, 1993.

Mowrer OH: On the dual nature of learning: A reinterpretation of “conditioning” and “problem solving.” Harv Educ Rev *17:*102, 1947.

Neal AM, Ward-Brown BJ: Fears and anxiety disorders in African American children. In *Anxiety Disorders in African Americans*, S Friedman, editor. Springer, New York, 1994.

Resick P, Schnicke M: Cognitive processing therapy for sexual assault victims. J Consult Clin Psychol *60:*748, 1992.

Roth M: The panic-agoraphobic syndrome: A paradigm of the anxiety group of disorders and its implications for psychiatric practice and theory. Am J Psychiatry *153* (Suppl):111, 1996.

Struening E, Pittman J, Welkowitz LA, Guardino M: Results of a national screening for anxiety disorders. Paper presented at the American Psychiatric Association, New York, 1996.

Takahashi T: Social phobia syndrome in Japan. Compr Psychiatry *30:*45, 1989.

Viswanathan R, Shah MR, Ahad A: Asian-Indian Americans. In *Cultural Issues in the Treatment of Anxiety*, S Friedman, editor. Guilford, New York, 1997.

*Welkowitz LA, Papp LA, Cloitre M, Liebowitz MR, Martin LY, Gorman JM: Cognitive-behavior therapy for panic disorder delivered by psychopharmacologically oriented clinicians. J Nerv Ment Dis *179:*473, 1991.

Wolpe J: *Psychotherapy by Reciprocal Inhibition*. Stanford University Press, Stanford, CA, 1958.

CHAPTER 16. SOMATOFORM DISORDERS

FREDERICK G. GUGGENHEIM, M.D.

Disorders of the mind are often linked to physical symptoms and signs of psychiatric disorders: for example, the trembling, palpitations, and hyperpnea of panic disorder and the constipation, dry mouth, furrowed brow (omega sign), and weight loss of depressive disorders. The connection between somatic symptoms and depression or panic disorder, for example, is almost always readily accepted by patient and primary care physician; the task is to chart a course of action to treat symptoms with medication and psychological intervention. However, the primary care physician often misses the diagnosis of somatoform disorders for weeks to years. When finally informed of the diagnosis by the primary care physician, the patient is typically not prone to accept it. The patient is convinced of the physical origin of the undiagnosed distress and is equally convinced that the psyche plays no part in the disease process. Thus, when a patient has a somatoform disorder, a disconnect is initially established between patient and physician in the diagnostic process.

As the term *somatoform* implies, this heterogeneous group of disorders are at the mind-body interface: bodily symptoms as psychiatric disorder. Somatoform disorders thus exemplify mind-brain interactions: the brain, in ways still not well understood, sends various signals that impinge on the patient's awareness, indicating a serious problem in the body.

From a nosological perspective somatoform disorders were only grouped together for the first time in 1980 in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) as those disorders in which bodily sensations or functions, as the patient's predominant focus, are influenced by a disorder of the mind. This clustering was not based on theoretical construct or laboratory findings.

There are now five specific somatoform disorders: (1) somatization disorder, (2) conversion disorder, (3) pain disorder, (4) hypochondriasis, and (5) body dysmorphic disorder. There are also two residual categories that do not meet the fullblown criteria of the other five: (1) undifferentiated somatoform disorder and (2) somatoform disorder not otherwise specified.

In routine medical practice, classic cases of somatoform disorders are less common than less well differentiated syndromes.

Chronic fatigue syndrome is a recently prevalent disease without identified pathogen. The Centers for Disease Control (CDC) has recently set up criteria for this disorder. Several features of this syndrome seem to overlap with some of the somatoform disorders and might be considered in the differential diagnosis of somatization disorder. From a symptom perspective, chronic fatigue syndrome has much in common with certain other ill-defined maladies such as multiple chemical sensitivity syndrome, the syndrome of clinical hypoglycemia, the effort syndrome, and neurasthenia. There is no doubt that those afflicted with these disorders of unknown cause suffer. Those disorders do have some syndromic overlap with clearly defined psychiatric disorders such as the somatoform disorders.

CORE FEATURES

Characteristic of somatoform disorders are three enduring clinical features: (1) somatic complaints that suggest major medical maladies yet have no associated serious, demonstrable, peripheral organ disorder; (2) psychological factors and conflicts that seem important in initiating, exacerbating, and maintaining the disturbance; and (3) symptoms or magnified health concerns that are not under the patient's conscious control.

Because of their intense bodily perceptions, restricted level of physical functioning, and morbid beliefs, these patients have become convinced they harbor serious physical problems. Moreover, their symptoms are not willfully controlled. Whatever their faults and problems, these patients are not malingerers. Yet their physicians' physical and laboratory examinations persistently fail to evince significant substantiating data about physical infirmity other than the patients' vigorous and sincere complaints. Patients with somatoform disorders are convinced that their suffering comes from some type of presumably undetected and untreated bodily derangement.

DIAGNOSTIC ISSUES

When a somatoform disorder occurs without another comorbid psychiatric condition, the primary care physician and the patient usually do not initially consider a psychiatric condition. The patient's morbid preoccupation with bodily concerns, not emotional feelings or disordered interpersonal relationships, is paramount. Often this preoccupation is so severe that it interferes with the patient's capacity for living, loving, or working; it usually has sent the patient on a ceaseless odyssey from physician to physician looking for effective symptomatic relief.

Because of the patient's focus on bodily issues, the psychiatrist's questions (should such a referral ever occur) about stress and family matters might seem off the mark to the patient. However, research literature as well as the clinical experience of seasoned psychiatric consultants has demonstrated the utility of psychiatric input into managing somatoform patients from a psychosocial perspective with attention to physical health status, mental health status, global outcome, and cost.

The 10th edition of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), which includes most of the somatoform disorders in the fourth edition of DSM (DSM-IV) (except for conversion disorders, which are included with the dissociative disorders), states that “the main feature of [the somatoform] disorders is the repeated presentation of physical symptoms together with persistent requests for medical investigations, in spite of repeated negative findings and reassurances by doctors that the symptoms have no physical basis. If any physical disorders are present, they do not explain the nature and the extent of the symptoms or the distress and preoccupation of the patient.” The ICD-10 criteria for somatoform disorders are presented in Table 16-1.

Table 16-1 ICD-10 Diagnostic Criteria for Somatoform Disorders

Primary Physician's Diagnostic Process Physicians have been trained to record faithfully the patient's medical history, to perform a physical examination, and to make use of laboratory tests. Often during the primary care physician's evaluation of the patient for tiredness, poor appetite, weakness, or tense nerves, an underlying psychiatric condition is spotted by the astute clinician who senses considerable emotional turmoil. The clinician then elicits specific findings on the mental status examination. In the routine practice of medicine, as long as the patient can render a concise chronicle of events and present a straightforward account of bodily perceptions underlying pathophysiological events, the primary care physician has a remarkably good chance of making an appropriate diagnosis and instituting corrective treatment. However, it becomes understandably difficult for the physician to develop an accurate diagnosis when the somatoform patient forgets (represses) or refuses (suppresses) to share with the physician certain medically relevant, critical events or when the somatoform patient continually attributes potentially fatal diseases to benign internal stimuli.

Moreover, when the patient's expectations of a clear-cut diagnosis of physical illness are not substantiated by the results of the physical examination and laboratory findings, the clinician may have questions about the somatoform patient's competency, motivation, or even integrity. Is this account of suffering actually deliberate distortion? Is it simple forgetfulness when the patient fails to tell the physician that there was a similar episode of such symptoms 10 years ago (with benign outcome)? To the probing clinician it might appear at first that somatoform patients are guilty of blatant, intentional manipulation, or at the least of considerable distortion. Most physicians are not primed to consider that the lack of concordance between subjective suffering and objective physical findings can stem from a set of distorted perceptions and attributions or unusual beliefs.

In an emergency room or office practice setting, the application of specific diagnostic criteria for the somatoform disorders can be helpful. Even patients with somatoform disorders do get ill and eventually die of something. In the past physicians have been far too ready to attribute atypical presentations of certain protean medical diseases (e.g., multiple sclerosis, lupus, or myasthenia gravis) to psychiatric causes—coupling such reductionistic thinking with the conclusion that a putative psychiatric disorder does not merit further investigation or even referral.

Since the 1970s clinical research has allowed the inclusion criteria for some of the somatoform disorders to be considerably simplified, which has helped to make some of these diagnoses easier to use. More concise criteria have further supported the maxim that it is inappropriate to make a psychiatric diagnosis just because another medical diagnosis cannot be determined (i.e., the process of psychiatric diagnosis by exclusion). Table 16-2 lists a few of the disorders commonly confused with somatoform disorders, especially early in their course.

Multiple sclerosis	Acute intermittent porphyria
Central nervous system syphilis	Lupus erythematosus
Brain tumor	Hyperthyroidism
Hyperparathyroidism	Myasthenia gravis

Table 16-2 Conditions Commonly Confused With Somatoform Disorder

BIOMEDICAL THEORIES

Of the few research findings on somatoform disorders that are relevant to underlying physiological or structural brain abnormalities, few have been widely disseminated. This makes the internist's task of telling patients that they have a psychiatric disorder even more complicated.

To comprehend the underlying basis of the somatoform disorders, the brain should be viewed both as a transducer of experience and as a practiced, highly trained organizer of perceptions from the *milieu interieur* and the *milieu exterieur.* The brain filters, amplifies, or dampens afferent and efferent stimuli from all parts of the body and from the brain itself. It then produces messages that explain the matrix of the patient's experiential world; that is, the brain edits and interprets its material (signals) and then sends out “sound bites” and an executive summary about bodily function in light of past incidents (including material seen on television or in the press). Somatoform disorders may involve a considerable variety of neuronal pathways—from brain-brain signals to pain pathways and perceptual pathways, in addition to efferent signals to motor apparatus and blood vessels.

CONVERSION DISORDER

Definition A conversion disorder is a disturbance of bodily functioning that does not conform to current concepts of the anatomy and physiology of the central or the peripheral nervous system. It typically occurs in a setting of stress and produces considerable dysfunction.

Many conversion disorders simulate acute neurological pathology (e.g., strokes and disturbances of speech, hearing, or vision). However, conversion disorders are not associated with the usual pathological neurodiagnostic signs or the underlying somatic pathology. Conversion symptoms (e.g., anesthesias and paresthesias produced by a conversion disorder) do not conform to usual dermatomal distribution of the underlying peripheral nerves; rather, the signs and symptoms of a conversion disorder typically conform to the patient's concept of the medical condition.

Conversion disorders seem to change or convert the psychic energy of the turmoil of acute conflict into a personally meaningful metaphor of bodily dysfunction. Turbulence of the mind is transformed into a somatic statement, condensing and focusing concepts, role models, and communicative meanings into one or several physical signs or symptoms of dysfunction. These somatic representations often simulate an acute medical calamity; initiate urgent, sometimes expensive medical investigation; and produce disability. In primitive settings, however, certain conversion symptoms have been taken as tokens of religious faith and even as expressions of witchcraft.

Although most conversion reactions are transient (hours to days), some can persist. Chronic conversion disorders can actually produce permanent conversion complications, such as disuse contractures of a “paralyzed” limb that remains long after the psychic strife that prompted the conversion has been resolved. In many cases a chronic conversion disorder serves to help stabilize an otherwise dysfunctional family. In addition to sensorimotor symptoms, marked autonomic disturbances such as protracted (psychogenic) vomiting, hyperemesis gravidarum, urinary retention, and pseudocyesis are also seen, although less commonly. Conversion

disorders challenge the diagnostic competence of internists, neurologists, otolaryngologists, ophthalmologists, and psychiatrists.

Like the other somatoform disorders, conversion disorders are not volitional. Rather, ego defense mechanisms of repression and dissociation act outside of the patient's awareness. Many patients with conversion disorders experience *la belle indifference*, an emotional unconcern or even flatness in a setting of catastrophic illness; but some patients do experience considerable anguish over their new symptoms.

A conversion disorder can be considered when a patient manifests a loss or alteration in physical functioning suggesting a medical or neurological disorder and the condition cannot be explained by any other known medical disorder or pathophysiological process. A conversion disorder cannot be diagnosed just because a medical disorder cannot be ruled in. Failure to prove a physical illness is a necessary but not sufficient condition for making the diagnosis of conversion disorder.

HISTORY

Until the middle of the nineteenth century, somatization disorder and conversion disorder (which often travel together) were considered to be one condition called *hysteria.* The term *hysteria* was derived from the Greek word *hystera*, meaning uterus. Descriptions of conversion disorders appeared as far back as 1900 BC when multiple symptoms were attributed by Egyptian physicians to a wandering of the uterus within the body.

In the middle of the century, Pierre Briquet originated the modern concept of conversion disorder. He considered the disorder to result from a dysfunction of the central nervous system (CNS). He proposed that conversion symptoms occurred in those with a constitutional predisposition when a receptive part of the brain was impacted by extreme stress. Later, Russel Reynolds described clinical cases in which the loss of function or the persistence of severe pain could be attributed to an idea that the patient had about the body.

Jean-Martin Charcot then expanded on the biological concepts of Briquet and the psychological constructs of Reynolds, adding heredity to factors that influence predisposition. Moreover, Charcot suggested that a traumatic event gave rise to the idea, which then led to the brain's dynamic dysfunction; Charcot also suggested that the idea could be produced in the brain by hypnosis.

The term *conversion* was first used by Sigmund Freud and his associate Josef Breuer. It was used to describe the clinical case of Anna O., whose undischarged psychic energy was bound in a somatic symptom. This symptom represented the unconscious conflict. That is, a repressed thought was converted to a somatic symptom. Freud then worked out his concept of talking therapy as a catharsis through which unconsciously repressed material might become conscious. With catharsis in psychotherapy and with hypnotic suggestion, somatic conversion symptoms were shown to diminish and even disappear.

In 1929, following from Charcot, Pierre Janet observed that conversion disorders were preceded by a lowering of conscious threshold and were associated with dissociation. He recognized that a constitutional weakness in an individual might be accentuated by shock or fatigue, resulting in aspects of consciousness being split off. His concept of what is now considered to be conversion disorder did not, however, include the concept of repression; thus, he was not concerned with the significance of the dynamic unconscious as Freud was.

COMPARATIVE NOSOLOGY

In 1952 the American Psychiatric Association's first edition of DSM (DSM-I) used the diagnostic term “conversion reaction,” stressing (1) the reactive part of the disorder, (2) issues of symbolism—symbols or ciphers from the individual's unconscious that relate to significant life experiences, and (3) secondary gain—tangible benefits that accrue to the individual upon assumption of the sick role. In 1967 the second edition of DSM (DSM-II) changed the diagnostic term to “hysterical psychoneurosis, conversion type.” Important in this formulation was *la belle indifférence.* Follow-up studies of specific phenomena associated with conversion disorder were constructed by assigning relative weights to certain features associated with conversion studies. These studies indicated, however, that there is no pathognomonic validity (using outcome as a gold standard) to symbolism of the symptom, secondary gain, hysterical personality, and *la belle indifférence.*

In 1980 DSM-III revised the diagnostic term again to *conversion disorder.* Removed from the cluster of symptoms was pain disorder, a symptom-based pain disorder with criteria otherwise comparable to that of conversion disorder. DSM-III and the revised third edition of DSM (DSM-III-R) required conversion disorder to be judged by the clinician to be etiologically related to the conversion symptom because of a temporal relationship between it and a significant psychosocial stressor or else a demonstrated coupling of conflict and psychological need and the initiation or the exacerbation of a preexisting symptom. That is, it was one of the few diagnoses where the judgment of the clinician was explicitly sought in the making of a diagnosis based on psychodynamic mechanisms. A subjective rather than an objective component was written into the diagnostic criterion; this raised the issue of inter-rater reliability in the diagnostic process.

One of the major changes from DSM-III and DSM-III-R to DSM-IV has been the further removal of the etiological inference of unconscious mechanisms and psychodynamics involved in the productions of symptoms, with the statement that “psychological factors are judged to be associated with the symptom or deficit because it was preceded by conflicts or other stressors.” That is, since diagnosticians are not mind readers, the role of unconscious motivation must be inferred.

A second difference in DSM-IV from DSM-III-R is in the exclusion category associated with the concept of “not fully explained by a known physical disorder.” This was broadened to include culturally sanctioned behavior or experience, general medical condition, and the direct use of a substance.

A third change came in the distress-disability category, in which the concept has now been broadened to include important areas of functioning to the individual other than just social and occupational and the phrase “warrants further medical attention” has been added (e.g., distress to family or physician because of potential medical implications associated with the symptom or deficit).

A fourth change focused on the wording used to eliminate factitious disorder and malingering. The words “not conscious of intentionally producing the symptom” have been removed in favor of the simple concept “not intentionally produced or feigned.”

In 1978 the ninth revision of *International Statistical Classification of Diseases, Injuries, and Causes of Death* (ICD-9) used the term “hysteria” to include conversion disorder and dissociative phenomena. Both conditions were defined as mental disorders in which a mechanism beyond the patient's awareness produced either a restriction of the field of awareness or a disturbance of motor or sensory function. Hysteria was associated with psychological advantage or symbolic value. Conversion symptoms involved the body's function, whereas dissociative symptoms involved the mind's function.

When ICD-10 considered the somatoform disorders, it listed somatization disorder and its subthreshold companion diagnosis, undifferentiated somatoform disorder, plus hypochondriacal disorder, persistent somatoform pain disorder, autonomic (psychogenic aerophagia) and nonautonomic (psychogenic pruritus) somatoform disorders as part of the group of the somatoform disorders. However, ICD-10 assigned conversion disorder to the dissociative disorders. Presumably this ICD-10 change was made because conversion disorders, like dissociative disorders, typically have a sudden onset and a short duration whereas the other somatoform disorders have a gradual onset and a much more chronic course. Thus, the term *conversion hysteria*, coupled loosely and interchangeably over centuries now disappears whereas *dissociative (conversion) disorders* appear in ICD-10 with separate codings for dissociative disorders of movement and sensation, dissociative motor disorders, dissociative convulsions, and dissociative anesthesia and sensory loss.

EPIDEMIOLOGY

Conversion disorders are the most frequently occurring of the somatoform disorders. Affected persons can range in age from early childhood into old age. The annual incidence of conversion disorders seen by psychiatrists in a New York county has been estimated to be 22 cases per 100,000 population. In a general hospital setting 5 to 16 percent of all psychiatric consultation patients manifest some conversion symptoms. In a study of a rural Veterans Administration general hospital, 25 to 30 percent of all male patients had a conversion symptom at some time during their admission. By contrast, in a psychiatric emergency room or psychiatric clinic, the incidence of conversion disorder is far lower (1 percent of all psychiatric admissions), as different selection factors supervene. Lifetime figures for ever having any conversion symptom, even if only on a transient basis, are far higher, with some studies reporting a 33 percent prevalence rate. Conversion disorder occurs mainly in women, with a ratio of 2 to 1 up to 5 to 1 in some studies. However, there does not seem to be an overrepresentation of conversion disorders in female children.

The prevalence of the disorder is highest in rural areas and among the undereducated and the lower socioeconomic classes. It is more prevalent in military populations, especially in those exposed to combat. It is also more common in underprivileged persons, in those of subnormal intelligence, and in industrial settings

where compensation neurosis may become an issue. There may be a tendency for familial aggregation and for the patient to be the youngest sibling in the family. The incidence of the disorder may be on the decline.

Clinicians frequently involved with these patients in general hospital settings contend that conversion disorders are more likely to be seen in individuals who are psychologically naive, who are not particularly introspective, and who are disinclined to believe that psychic factors can affect physical processes than in individuals who are psychologically sophisticated.

Comorbidity Medical and especially neurological disorders occur frequently among patients with conversion disorders. Indeed in some series the majority of conversion patients have a well-documented neurological condition. What is typically seen in these comorbid neurological or medical conditions is an elaboration of symptoms stemming from the original organic lesion. Whether this tells the clinician something about the nature of the sick role or about the nature of the brain with compromised functioning and an altered state of consciousness remains an open question.

Preexisting or emerging psychopathology also seems to predispose an individual to the development of a conversion. Among the Axis I psychiatric conditions, depressive disorders, anxiety disorders, and somatization disorders are especially noted for their association with conversion disorder. Conversion in schizophrenia is reported but is very uncommon. Studies of patients admitted to a psychiatric hospital for conversion disorder reveal that, on further study, one quarter to one half have a clinically significant mood disorder or schizophrenia.

Axis II personality disorders also frequently accompany a conversion disorder, especially the histrionic type (in 5 to 21 percent of cases); the passive-dependent type (9 to 40 percent of cases), and the passive-aggressive type of personality disorder. However, conversion disorders can occur in persons with no predisposing medical, neurological, or psychiatric disorder.

ETIOLOGY

Biological Factors Conversion disorders represent pathology in the mind as well as dysfunction within the brain. Recent etiological research on conversion disorder has involved event-related potentials, structural and functional brain imaging, and neuropsychological testing to investigate aspects of corticofugal inhibition of afferent stimuli.

Imaging Pierre Flor-Henry's important work on the etiopathology of conversion disorder has used bipolar derivations of the electroencephalograph (EEG) with measures of coherence to reflect synchrony, not between individual electrode sites but between regions of the brain. His findings emphasize (1) hypofunction of the dominant hemisphere systems, (2) a consequent dysfunctional overactivity of the nondominant hemisphere, and (3) abnormal interhemispheric relations. Defects in processing endogenous somatic signals and integrating sensorimotor signals appear to be the consequence of altered dominant hemispheric systems. Particularly in women there seems to be a secondary disorganization of the contralateral hemisphere that in turn is capable of producing the characteristic somatic symptoms seen in a conversion reaction. An EEG study of a woman with a left-sided paralysis as a conversion disorder has proved most illuminating. The attempt to move the paralyzed leg failed to activate the right primary motor cortex. Instead, the right orbito-frontal and right anterior cingulate cortex were significantly activated, suggesting these two areas inhibit the prefrontal (willed) effects on the right primary motor cortex when the patient tries to move her left leg.

NEUROPSYCHOLOGICAL TESTS

Impaired vigilance-attention and short-term memory have been demonstrated. Localizing studies using the Halstead-Reitan Battery and other neuropsychological tests have manifested dysfunction of both nondominant right and especially dominant left hemispheres. Increased field dependency and heightened suggestibility is also present.

Unilateral Symptoms and Localization Taken together, clinical and research findings could suggest that patients with conversion disorder under extraordinary circumstances experience impaired intercortical communication and blockade of ordinary channels of verbal associations. The preponderance of left-sided, unilateral symptoms seen in conversion disorder plus the strong association of conversion disorders with depressive disorders could be used as evidence of a nondominant right-hemispheric vulnerability. Additional complementary evidence for localization comes from the fact that the left hemisphere is phylogenetically associated with inhibitory influences. Thus, the motor and sensory symptoms of conversion suggest defects in processing and in analysis of sensorimotor signals, which leads to a failure in the integration of endogenous somatic signals. The proposed defect in understanding the signals in a conversion disorder is in some ways analogous to the failure of comprehension in a stroke (i.e., with receptive and expressive aphasia when acoustic-motor coordination of auditory signals involving language fails to occur).

Sex and Brain Localization The fact that conversion disorders occur mainly in women is another piece of evidence used to support a theory of brain localization. Studies from a variety of sources indicate that women have greater instability of right-hemispheric organization. Thus, it has been proposed that a primary defect in the left hemisphere interferes with the normal transcallosal inhibitory stabilizing functions of the unstable contralateral right hemisphere. These circumstances could account for the symptoms of conversion disorder and for its almost exclusive restriction to the female pattern of cerebral organization.

Much more clinical and experimental work remains to be done if these heuristic hypotheses are to have widespread clinical relevance. The phenomena of conversion disorder and hypnosis have both been considered to result from blockade of corticofugal impulses induced by emotional rapport or intense emotional experience. Both conditions can lead to selective diminution of awareness of a bodily function. Interestingly, hypnosis can bring about temporary remission of conversion symptoms and can also produce a mimicry of conversion symptoms in those not afflicted with the condition.

These biomedical theories account for the *how*, but not the *what* or the *why* of conversion disorder. Obviously a multifactorial explanation is needed to render an understanding of the patient's plight and to serve as a framework for testing the most effective and efficient methods of treatment. Conversion symptoms represent a common pathway for the expression of a complex biopsychosocial event. A patient with conversion disorder, having a specific diathesis, experiences and creates (outside of his or her level of awareness) an illness in a setting of stress that is shaped to some extent on his or her model of disease.

PSYCHOSOCIAL FACTORS

Long before neuropsychological, neurophysiological, and imaging evidence was available to contribute to an understanding of symptoms of conversion, astute nineteenth-century clinicians focused their attention on the psychological aspects of their patients' internal and external worlds. Sigmund Freud and Josef Breuer carefully studied the ongoing emotional struggles and the relevant life circumstances immediately antecedent to conversion symptoms in patients with neuroses. In such a fashion the context of the acute psychic trauma and other aspects of the patient's life story often could be coupled meaningfully to the development and maintenance of the patient's conversion symptoms. Moreover, psychoanalytic and hypnotic techniques sometimes produced dramatic and sometimes permanent remission of the patient's symptoms.

Psychodynamic Factors According to psychoanalytic theory a conversion reaction results when the anxiety of unconscious intrapsychic conflict is converted into somatic symptoms. When aggressive or sexual impulses emerge in a field of strong inhibition of their expression, the resultant intrapsychic conflict overwhelms the person's ordinary ego defense mechanisms. In such a setting unconscious mechanisms facilitate a compromise as conversion symptoms emerge. The settlement allows a partial expression of the primitive impulse but disguises it so that the individual is unaware of the unconscious wish and the unacceptable desire. However, the symptom formation may impose a considerable price. Suffering and disability then serve as atonement for having had the unacceptable wish or impulse.

The decrease in anxiety and psychological distress after formation of the conversion symptom is the primary gain. Benefits that also accrue to the individual after the sick role is assumed are the secondary gain. Both primary and secondary gain are typically part of the syndrome associated with conversion disorder. Its permanence and severity can be reinforced by patients being enmeshed in irresolvable conflict for which they feel no responsibility.

In psychodynamic terms conversion symptoms represent a solution to an unconscious conflict between instinctual drives and superego prohibitions to their expression. The conversion symptom emerges when a latent conflict is activated by an event unconsciously perceived as related to the conflict. The specific meaning of the precipitating event resonates with the personal vulnerability of the patient and the patient's life experiences. Thus the formation of the conversion symptom embodies a symbolic aspect of the intrapsychic conflict. The conversion symptoms may derive from identification with a significant individual, often someone whom the patient associates with loss and who has also experienced such a symptom (a conversion model). Sometimes the conversion model is the patient's own somatic

ailment, with symptoms now called forth in an effort to resolve some intrapersonal or interpersonal crisis.

Sociocultural Theory Viewed alone or as complementary to psychoanalytic theory, conversion symptoms can be understood in sociocultural terms as a form of communication concerning an emotionally charged feeling or idea blocked from expression by personal or cultural restraints. Conversion symptoms can express the forbidden, using mimicry or pantomime instead of words. Moreover, the symptoms of a conversion disorder allow the individual to enter into the sick role, avoiding certain responsibilities or noxious situations. As such, the patient can control or otherwise manipulate the behavior of others. However, it is most likely that clinicians who view conversion disorder only as a nonspecific organic response to stress and merely as a manifestation of illness behavior have: (1) had difficulty obtaining clinical data from or about the patient; (2) have failed to engage the patient from a psychodynamic perspective; or (3) have not used an amobarbital (Amytal) or other drug-assisted interview.

Learning Theory In terms of conditioned learning theory a conversion symptom can be seen as a piece of classically conditioned learned behavior; symptoms of illness, learned in childhood, are called forth as a means of coping with an otherwise impossible situation.

DIAGNOSIS AND CLINICAL FEATURES

A frequent but not invariable common denominator of conversion disorder is the pseudoneurological nature of the symptom. Common types of conversion symptoms are listed in Table 16-3.

Motor Symptoms	Sensory Deficits
Involuntary movements	Anesthesia, especially of extremities
Tics	Midline anesthesia
Blepharospasm	Blindness
Torticollis	Tunnel vision
Opisthotonos	Deafness
Seizures	**Visceral Symptoms**
Abnormal gait	Psychogenic vomiting
Falling	Pseudocyesis
Astasia-abasia	*Globus hystericus*
Paralysis	Swooning or syncope
Weakness	Urinary retention
Aphonia	Diarrhea

Table 16-3 Common Symptoms of Conversion Disorder

Motor symptoms—Abnormal gait, weakness, and paralysis may occur. There can be involuntary movements, rhythmical tremors, episodic jerks, tics, seizures, and falling. Yet when patients with conversion symptoms fall, they rarely are severely hurt. Blepharospasm, torticollis, and opisthotonos may also occur. All such symptoms tend to grow more intense when observed. Many neurological symptoms of patients with a known pathophysiological basis, such as Parkinson's disease, also tend to intensify when the individual experiences increased anxiety. The DSM-IV criteria for conversion disorder are listed in Table 16-4.

A. One or more symptoms or deficits affecting voluntary motor or sensory function that suggest a neurological or other general medical condition.
B. Psychological factors are judged to be associated with the symptom or deficit because the initiation or exacerbation of the symptom or deficit is preceded by conflicts or other stressors.
C. The symptom or deficit is not intentionally produced or feigned (as in factitious disorder or malingering).
D. The symptoms or deficit cannot, after appropriate investigation, be fully explained by a general medical condition, or by the direct effects of a substance, or as a culturally sanctioned behavior or experience.
E. The symptom or deficit causes clinically significant distress or impairment in social, occupational, or other important areas of functioning or warrants medical evaluation.
F. The symptom or deficit is not limited to pain or sexual dysfunction, does not occur exclusively during the course of somatization disorder, and is not better accounted for by another mental disorder.
Specify type of symptom or deficit: **With motor symptom or deficit** **With sensory symptom or deficit** **With seizures or convulsions** **With mixed presentation**
Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. th American Psychiatric Association, Washington, DC, 1994.

Table 16-4 DSM-IV Diagnostic Criteria for Conversion Disorder

Astasia-Abasia A staggering, ataxic gait with gross jerks and thrashing or wild waving of the upper extremities, often with an inability to stand without support, is called *astasia-abasia.* Surprisingly, such patients sometimes can dance to music after the clinician suggests that the ability to successfully perform a dance such as the fox trot or the Texas two-step is not impacted by the patient's inability to stand or walk.

Pseudoseizures *Pseudoseizures* are paroxysmal episodes of altered behavior resembling epileptic attacks but devoid of the characteristic clinical epileptic and electrographic features. Convulsive behaviors identified as a conversion disorder often take place when the clinician walks into the patient's room or when the family visits. Variously termed *psychogenic* or *hysterical seizures* in the past, these clinical episodes terminate without the patient having a period of sluggishness, sleepiness, or confusion (as might be seen following a true convulsion). Pseudoseizures, but not complex partial seizures, tend not to manifest extreme stereotypy in the overt motor sequence and lack a neurological indicator. These patients do not evidence an elevated serum prolactin level immediately following the clinical episode, and the patient's interictal cortical EEG is normal or ambiguous. The EEG of the pseudoseizure patient during the clinical episode does not show any of the correlates of an epileptic seizure, such as increasing frequency of spike discharges, sudden onset of focal or diffuse rhythmic activity, or postictal slow waves.

Patients with pseudoseizures tend to have a history of psychiatric treatment, suicide attempts, and borderline personality disorder. However, the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) in one study demonstrated an overall accuracy rate of 67 percent (correctly identifying 71 percent of the group with pseudoseizures and 65 percent of the patients with epileptic seizures) using the video-monitored EEG as the gold standard. Families of pseudoseizure patients had considerably more health problems, distress, and criticism than families of patients with epilepsy. There have now been a number of reports linking an increased incidence of childhood sexual abuse with the initiation, precipitation, or exacerbation of symptoms in patients with pseudoseizures. A diagnosis of pseudoseizures is also more likely if there is a financial incentive. Pseudoseizures can also be seen as a form of chronic maladaptive behavior utilizing the benefits of the sick role.

The induction of a seizure by suggestion, formerly considered a hallmark of hysterical seizures, is also seen in complex partial seizures when there is elevated electrical lability. Semi-purposeful movements, thrashing and pelvic thrusting, often considered to be the hallmarks of pseudoseizures, can be seen with direct stimulation of the cingulate region and are common correlates of complete partial seizures with frontotemporal foci. Moreover, more than 70 percent of patients with pseudoseizures without documented bona fide seizures show interictal EEG abnormalities, neuropsychological impairment, or abnormalities on magnetic resonance imaging (MRI) and computed tomography (CT) scans.

Although tongue-biting, urinary incontinence, injury during falls, and seeming loss of consciousness do not usually occur with a pseudoseizure, all of these can occur. However, the preservation of corneal, pupillary, and gag reflexes, plus the absence of extensor plantar responses and the preservation of normal color during the attack all suggest a pseudoseizure. The eyes of a patient in a pseudoseizure reportedly deviate toward the ground when the patient is placed on his or her side.

The proportion of pseudoseizures in a given study typically reflects the nature of the referral source and their relationship with the evaluation. From a clinician's perspective about one third of patients evaluated for a pseudoseizure do not have a convulsive disorder or pseudoseizures but rather have some other neurological condition, another third of patients have true convulsions as well as pseudoseizures, and another third have just pseudoseizures. From a treatment perspective it is imperative to know the correct diagnosis so as to avoid polypharmacy and toxicity, hazardous diagnostic interventions, and neglect of underlying psychological distress. Although patients with pseudoseizures typically are psychologically naive, their very lack of sophistication may make them more respectful of the power of

the physician's authority. They may be more suggestible and easier to manipulate, which may be effectively utilized in the service of their treatment.

Other Common Motor Symptoms Other symptoms include paralysis or paresis, more frequently on the patient's nondominant side (i.e., on the left side if the patient is right-handed). Paralyses can occur in one, several, or all limbs. In contrast to a patient with neuropathy, the reflexes in a patient with conversion disorder with anesthesia or paralysis remain normal. There is no fasciculation, no other electromyographic abnormality, and no atrophy. However, in cases of longstanding pseudoparalysis, disuse atrophy and even contractures can occur as a complication of conversion.

Sensory Symptoms Anesthesias, hypesthesias, and paresthesias are common conversion symptoms, especially in the extremities (Table 16-3). The pattern of distribution of the anesthesia does not conform to the underlying central or peripheral nerve distribution. Typical are glove or stocking and strict midline anesthesias. Despite a claimed loss of total sensation in the legs and feet, for example, conversion patients can walk in the dark without stumbling (unlike those with tabes dorsalis who lack position sense). Conversion patients told to answer “yes” if they can feel anything when pricked with a pin in the anesthetic zone often will respond “no” when pricked, even when they are not looking at the area being tested.

Hysterical Blindness Other sensory modalities can also be affected. Patients with hysterical blindness typically do not hurt themselves seriously when they bump into stationary objects. Despite their “lack” of vision, their pupils react to light and their visual evoked potentials on the EEG are consistent with those of normal vision. Some of the other findings in monocular and binocular blindness associated with conversion are shown in Table 16-3.

Hallucinations Patients with conversion disorder can have positive or negative hallucinations. Patients who have positive hallucinations perceive an image or hear a sound that is not there. Hallucinations in conversion disorder are usually associated with intact insight, and the hallucinations are often visual, auditory, and tactile. They tend to be described by the person as part of an interesting story. Conversely, with a negative hallucination the patient with an intact nervous system does not see an object that others can see or does not hear a sound that others can hear. When generalized to more than a specific object that is not seen, the individual seems to be blind even though the visual apparatus is still working.

Visceral Symptoms Psychogenic vomiting can occur as a conversion disorder (Table 16-3). Typically such a patient will not suffer significant weight loss when observed for a week on an inpatient medical service; gastrointestinal workup will show no significant disease or disorder. Another visceral conversion disorder is urinary retention. On urological workup, conversion patients show normal intracystometric dynamics. Conversion as pseudocyesis manifests as a cessation of menses, a protuberant abdomen, and an elevation of serum hormones seen in early pregnancy. Other visceral conversions include globus hystericus, syncope, and diarrhea.

Important Caveats As with all disorders with a low specificity and no confirmatory laboratory test, substantiation of the diagnosis is facilitated by the passage of time with no other countervening diagnosis evolving. Also helping to confirm the diagnosis of conversion disorder is the development of another bout of conversion symptoms. About 25 percent of patients with conversion disorder will develop another episode during the following 1 to 6 years.

The diagnosis of a conversion disorder is made more secure if details of a prior set of conversion symptoms can be elicited. However, complicating the collecting of the past medical history in many patients with somatoform disorders is their frequent use of repression. Such patients thus may not recall important pieces of historical data. In adding collateral information a review of the patient's old clinical chart can be very helpful.

Another important caveat is that the diagnosis cannot be based solely on inexplicable neurological findings accompanied by relevant psychological factors. Patients with lesion-based neurological disorders also wrestle with the sick role. Moreover, although suggestibility is often seen in patients with conversion, some patients with organic disorders can also respond to suggestion, briefly altering their symptoms. It is not uncommon to find neurological disease coexisting with a conversion disorder.

Critical review of the literature finds little or no empirical support for the necessity of a number of previously widely accepted classic accompaniments of conversion disorder including: *la belle indifférence*, hysterical personality, the presence of secondary gain, the symptom as symbolism, sibling position, disturbed sexuality, and conversion V pattern on the MMPI-2 (elevated hysteria and hypochondriasis scales, even higher than the depression scale).

Suggestibility, with a patient seizing on command (or when induced in a particular situation), used to be considered a sign that a convulsion was a pseudoseizure. However, recent clinical and EEG evidence indicates that the precipitation of overt, nonstereotyped seizures by instruction may not be a reliable diagnostic method when an insidious process is slowly elevating temporal and limbic lobe lability. When taken by itself, no single associated finding is pathognomonic for the diagnosis of conversion disorder. However, a number of features taken together help the clinician to determine the likelihood (i.e., possible, probable, definite) of a conversion disorder.

Setting Marked psychological stress is almost always present. Precipitants typically may have been acute rage, truncated grief, sexual abuse, or physical abuse. Somatic symptoms may develop abruptly following a dramatic psychological blow, mechanical trauma, or a life-threatening experience. However, studies on patients with conversion disorder demonstrate that their life experiences are not more extreme than those with other types of psychiatric disorders.

The initial mental status examination of the patient with conversion disorder may prove to be quite unremarkable: a calm person who may or may not be troubled by a new somatic symptom may have no insight into the symptom's underlying dynamics. At first glance the patient's family may appear to be happy and integrated, with the family predicament (if present) being covert. Family difficulties are common in patients with conversion disorders, but not more so than for families of patients who attend a psychiatric clinic for other psychiatric disorders.

> Mrs. A. was a 22-year-old right-handed fundamentalist farmer's wife, homemaker, and mother of three from a sparsely settled Western state. Her past medical history was benign except for a motor vehicle accident 2 years previously that produced a sharp blow to the right temporal area, resulting in several hours' loss of consciousness. She had an unremarkable behavioral history without substance abuse, prolonged depressions, or unexplained somatic symptoms. Her demeanor had always been placid and unassuming. There was no family history of antisocial behavior or substance abuse.
>
> On Thanksgiving Day, while taking her usual solitary afternoon walk along the creek behind the kitchen, she came upon the floating, lifeless bodies of two of her children. She shrieked, swooned, and fell to the ground. Relatives in the house rushed out to assist but were unable to revive the children. When she was helped up, she asked that her husband guide her back to her room. Later that afternoon she seemed calm, even detached, as others scurried about making arrangements. She admitted to a visitor that she seemed to have lost the gift of sight.
>
> That evening the family physician was called to examine the newly sightless woman. He noted that her pupils were round, equal, and constricted briskly with a bright light; she was unable to touch the tips of her index fingers together in front of her; she failed to look at her own hands when instructed to do so; and she had no other neurological abnormalities, asymmetries, or complaints. The physician explained to the gathered family and patient that the woman was suffering from nervous shock, needed kind and quiet support, and should refrain from routine household chores for the moment. The physician also suggested that her eyesight would gradually return over the next week or so, perhaps following the funerals of her children. The patient's vision did slowly return over the next days and she gradually resumed her usual level of care for the home, her surviving child, and other members of the family.

PSYCHOPATHOLOGY AND LABORATORY EXAMINATION

Diagnostic workup is a multistep process that begins with a very thorough history and physical examination. There are a number of simple but specialized examinations that can be worked into a seemingly routine physical examination. Although some diagnoses can be made from the foot of the bed, chair, or stretcher with substantiating data collected in just a few minutes, it is not infrequent for an initial workup of a patient with a conversion disorder to take many hours even when conducted by a skilled clinician. Collateral sources need to be included to build a case based on circumstantial evidence of what was going on in the patient's life at the time of the acute shock. Table 16-5 lists examples of important tests that are relevant to conversion disorder symptoms.

Condition	Test	Conversion Findings
Anesthesia	Map dermatomes	Sensory loss does not conform to recognized pattern of distribution
Hemianesthesia	Check midline	Strict hall-body split
Astasia-abasia	Walking, dancing	With suggestion, those who cannot walk may still be able to dance; alteration of sensory and motor findings with suggestion
Paralysis, paresis	Drop paralyzed hand onto face	Hand falls next to face, not on it
	Hoover test	Pressure noted in examiner's hand while paralyzed leg when attempting straight leg raising
	Check motor strength	Give-away weakness
Coma	Examiner attempts to open eyes	Resists opening, gaze preference is away from doctor
	Ocular cephalic maneuver	Eyes stare straight ahead, do not move from side to side
Aphonia	Request a cough	Essentially normal coughing sound indicates cords are closing
Intractable sneezing	Observe	Short nasal grunts with little or no sneezing on inspiratory phase; little or no aerosolization of secretions; minimal facial expression; eyes open; stops when asleep, abates when alone
Syncope	Head-up tilt test	Magnitude of changes in vital signs and venous pooling do not explain continuing symptoms
Tunnel vision	Visual fields	Changing pattern on multiple examinations
Profound monocular blindness	Swinging flash light sign (Marcus Gunn)	Absence of relative afferent pupillary defect
	Binocular visual fields	Sufficient vision or "bad eye" precludes plotting normal physiological blind spot in good eye
Severe bilateral blindness	"Wiggle your fingers, I'm just testing coordination"	Patient may begin to mimic new movements before realizing the slip
	Sudden flash of bright light	Patient flinches
	"Look at your hand"	Patient does not look at them
	"Touch your index fingers"	Even blind patients can do this by proprioception

Table 16-5 Distinctive Physical Examination Findings in Conversion Disorder

There are also a few specialized laboratory-based procedures that can assist in the diagnostic workup of some patients with conversion disorder. Helpful in specific instances are simultaneous EEG and videotaping of behavior along with lack of elevation of serum prolactin levels in pseudoseizures; optokinetic drum test in conversion blindness; cortical evoked potentials for auditory and visual deficits; and electromyogram for fasciculation in lower motor neuron paralysis.

DIFFERENTIAL DIAGNOSIS

Almost any neurological symptom can have a conversion basis. Many serious neurological diseases can be mistaken for the conversion disorders (Table 16-6), and a number of psychiatric conditions need to be considered in the differential diagnosis of conversion disorder (Table 16-7). Until the diagnosis is clarified, ongoing evaluation by a psychiatrist and a neurologist may be useful; premature closure obviously aborts further diagnostic consideration.

Myasthenia gravis
Periodic paralysis
Brain tumor
Multiple sclerosis
Optic neuritis
Partial vocal cord paralysis
Guillain-Barré syndrome
On-off syndrome of Parkinson's disease
Degenerative diseases of basal ganglia and peripheral nerves
Acquired myopathies including polymyositis
Subdural hematoma
Acquired and hereditary dystonias
Drug-induced dystonia
Creutzfeldt-Jacob disease
Early manifestations of acquired immune deficiency syndrome (AIDS)

Table 16-6 Neurological Conditions in the Differential Diagnosis of Conversion Disorder

Major depressive episodes	Histrionic personality disorder
Catatonic schizophrenia	Adjustment disorder
Pain disorder	Posttraumatic stress disorder
Somatization disorder	Malingering

Table 16-7 Psychiatric Conditions in the Differential Diagnosis of Conversion Disorder

COURSE AND PROGNOSIS

Course Most patients diagnosed with conversion disorder experience a quick symptomatic recovery. Rapid improvement is especially seen in cases where symptoms are of recent onset, suggestive of voluntary control, inconsistent or variable, and of obvious immediate benefit.

Follow-up studies in the literature from the 1960s demonstrated multiple diagnostic errors in series of conversion patients followed for 2 to 20 years because clinicians, using subjective criteria, proved to be too liberal in their use of the diagnosis. Some long-term follow-up studies have shown that 25 percent of conversion disorder patients on psychiatric wards who subsequently died had inaccurate initial psychiatric diagnoses of conversion disorder. Long-term follow-up studies using selection criteria heavily weighted to psychodynamic issues (symbolism, secondary gain, *la belle indifférence*) from DSM-I, DSM-II, ICD-9, and even DSM-III have shown an appalling lack of diagnostic consistency. A frequently quoted British inpatient study reported that over half of the patients had an organic disease 7 to 11 years later that accounted for the symptoms of the supposed conversion disorder. Of note, however, two thirds of the patients in that series were not diagnosed as having conversion disorder by a psychiatrist before they were accepted into the series. Many cases from that inpatient study who did not develop neurological disease instead developed or were discovered to have some other type of disabling psychiatric disorder.

An American study found that neurological disease explained the original conversion symptoms in only one fifth of the cases. The emergence of neurological disease was more frequent in conversion patients who did not have somatization disorder as a comorbid condition. Another study with a 20-year follow-up noted that fully one third of patients later experienced a psychotic illness, often paranoid schizophrenia. In that sample central nervous system disease, especially epilepsy, was often mistaken for hysteria. A third major study with a 10-year follow-up noted that a quarter of patients developed organic disorders over time that accounted for the presenting conversion symptoms: mainly degenerative diseases of the spinal cord, peripheral nerves, bones, muscle, and connective tissue.

Thus, follow-up studies from tertiary medical centers show that there is a distinct possibility of the emergence of other medical, neurological, or disabling psychiatric disorders to account for the original conversion symptoms. Thus, one can conclude from the older literature that either the inclusion criteria used for the initial diagnosis of conversion disorder were far too inclusive, or else that the diagnosis of conversion disorder must be made more cautiously.

Conversion disorder has been imprecisely and improperly diagnosed when clinicians used it as a diagnosis of exclusion in the face of the neurological examination not being consistent with full-blown neurological disease. Clinicians needs to be wary that they do not erroneously think the patient has a psychiatric illness such as conversion disorder because of inexplicable neurological signs and the coincidence of emotional conflict.

Prognosis A favorable prognosis of conversion disorder is associated with sudden onset; readily identifiable stressful events; good premorbid health with no comorbid psychiatric, medical, or neurological disease; and no ongoing compensation litigation.

In a retrospective American study of patients with conversion disorder without comorbid neurological or medical disorders, 90 of 100 patients recovered by the time of

psychiatric hospital discharge; on follow-up 5 years later, 75 percent remained well. Another general hospital study demonstrated that half of patients on a medical surgical unit found to have a conversion disorder during a psychiatric consult experienced remission of their conversion symptoms by the time of medical discharge. Another study of conversion patients successfully treated with a 1-year follow-up, only one fifth had relapsed; symptom substitution was minimal.

Patients with chronic conversion disorders, however, do not have a good prognosis. In a recent 10-year British follow-up study of 56 patients hospitalized and well studied for their pseudoneurological symptoms, 30 had no relief from their original symptoms. Additionally, 11 patients had neurological findings, tentatively considered at symptom onset, which in retrospect accounted for their conversion symptoms.

TREATMENT

Most conversion symptoms remit spontaneously or after behavioral treatment, suggestion, and a supportive environment. Thus, for symptoms of very recent onset a variety of other therapies have also been utilized successfully. In practice clinicians tend to choose therapies that reflect their training. Irrespective of the technique used, most approaches seem to work when symptoms are not reinforced, when an authoritarian approach is used, and when the patient's psychosocial plight is the focus of attention.

Common Denominators of Success What seems least likely to be effective is trying to get the newly afflicted patient to accept the therapist's opinion that the somatic symptom is a direct manifestation of a psychosocial problem (e.g., that the physical disability is the representation of a psychiatric problem). Rather, the common denominator of successful treatment (irrespective of the clinician's theoretical framework) is the building of a caring, authoritative relationship. It is important to provide the patient with a safe environment to facilitate the gradual decrease in symptoms. The clinician can then deal indirectly with interpretations of the conversion symptom while trying to minimize or eliminate the symptom. It is not helpful to argue with the patient about the cause of the conversion disorder.

A multiplicity of types of therapies have their adherents (Table 16-8). When successful, some common elements in many of these treatments are the following: (1) a nonconfrontational approach, (2) discouraging retention of symptoms, and (3) manipulating the environment.

Faradic stimulation	Hypnosis and other suggestive techniques
Physical therapy	Behavioral approach to symptom elimination
Electrosleep	Family therapy
Inexact interpretations	Long-term insight-oriented psychotherapy

Table 16-8 Therapies Used in Conversion Disorder

Psychodynamic techniques and insight-oriented intensive psychotherapy may need considerable modification with psychologically naive patients. At times placing patients in a double bind by telling them that full recovery constitutes proof of an organic cause and failure to recover constitutes proof of nonorganic or psychiatric etiology, facilitates rapid recovery. For example, when the patient is failing to progress, three possible explanations can be made for the lack of progress: (1) the patient has psychological problems that will necessitate long-term self-examination in treatment; (2) the patient is not trying, which constitutes grounds for dismissal; or (3) the symptoms are caused by excessive overstimulation and fatigue, which would necessitate periods of deep rest without stimulation of any kind. Patients usually choose this latter option as the most likely explanation. Therapy would then begin with behavioral goals augmented daily and deep rest if the patient fails to meet those goals. Deep rest precludes reading, watching television, and socializing with staff members or other patients until the next therapy session. This precaution should be presented as an expression of concern about the patient's level of fatigue and as a means of preventing overstimulation. This behavioral paradigm must be seen not as punitive but as very supportive. In this way, patients rapidly lose symptoms and gain discharge; however, each behavioral intervention needs to be tailored to the specifics of the case.

If combined with a supportive and problem-solving approach, hypnosis may be helpful. On some occasions parenteral injections of amobarbital or lorazepam (Ativan) have also been seen as helpful. For example, 50-mg doses of amobarbital from an intravenous infusion in 5 percent dextrose could be administered until the patient feels sleepy. An interview might require a total of 200 to 700 mg of amobarbital. A nurse should always be present as witness and a crash cart must be available lest there be respiratory failure or cardiac arrest. Videotaping the interview can be a powerful tool to allow later feedback to the patient under the supervision of the therapist. This gives patients the opportunity to see how their unconscious works; however, it is the physician's responsibility to help the patient integrate this information.

Oral anxiolytic medications also may reduce anxiety and allow the patient to engage in a psychotherapeutic process that might otherwise be too overwhelming to handle. Decreasing the need for secondary gain by opening up other channels of communication seems to help. Eventually the more emotionally healthy patients seem to be able to gain insight into the meaning of their symptoms. Using a variety of techniques to work at resolving the patient's problems as well as eliminating the conversion symptom seems to be the best approach. Symptom substitution may occur if the patient still needs the conversion symptom, but most authors believe that substitution is usually not seen if one works patiently and with tact, allowing patients to save face and to accomplish some of their covert goals.

Poor Outcomes Not all symptoms remit in hours or days; some linger tenaciously despite skilled inpatient treatment. In such instances other psychopathology and immutably malignant social pathology are often present. The presence of a conversion disorder in one member of a family, the putative patient, may well serve as a beneficial cohesive factor in an otherwise chaotic dysfunctional family. Symptom removal and clinical remission seems to be least successful if compensation is at issue.

SOMATIZATION DISORDER

Definition The essential feature of somatization disorder is recurrent, multiple somatic complaints requiring medical attention but not associated with any physical disorder. Somatization disorder is the expression of personal and social distress in a deeply ingrained idiom of bodily complaints linked with medical help-seeking behavior. The diagnosis requires a history of many physical complaints of several years' duration and a lifetime history, beginning before age 30, which result in medical treatment or alteration in life-style. The symptoms must not be fully explained by a known nonpsychiatric medical condition, or the resulting complaints or impairment must be excessive. Typically these symptoms, coupled to persistent complaints plus demands placed on caregivers and medical services, result in gain to the patient through relief from responsibilities, through caring responses and attention from others, or both.

The diagnostic syndrome of somatization disorder has been evolving since the 1950s with various refinements and simplifications in criteria of a disorder that has the same stable, chronic course. Like a personality disorder, somatization disorder begins at an early age, is chronic and unremitting and appears to be a deeply ingrained adaptation to life. Currently, eight symptoms are required, which must meet a specific pattern: four different sites of pain; two different gastrointestinal symptoms; one sexual or reproductive system symptom other than pain; and one neurological symptom. The subjective severity of the symptoms must be sufficient to lead the patient to consult a physician, take medicine, or make life-style changes.

History The history of somatization disorder is complex. Essentially, over the centuries two complementary syndromes have been described: one monosymptomatic and the other polysymptomatic. The monosymptomatic syndrome is currently recognized as conversion disorder whereas the polysymptomatic syndrome has become known as somatization disorder. Historically, the two disorders have often been interrelated and commingled.

Hysteria Somatization disorder has had many names and many antecedents; one such predecessor has been referred to in older texts as *hysteria*, first recognized by the ancient Egyptians who believed that hysteria was caused by upward dislocation of the uterus and displacement of other organs. Migration of the uterus throughout

the body thus provided the basis for the multiple symptoms. Doubts about the uterine origin of hysteria began in the seventeenth century. Thomas Syndenham dissociated hysteria from the uterus and linked it with a psychological disturbance known at that time as “antecedent sorrows,” therein recognizing the emotional origin of the disorder. Further, Sydenham was also the first to recognize the disorder in men.

In 1859 Briquet emphasized the multisymptomatic aspects of the disease and its protracted course. His report of the 430 cases observed at the Hospital de la Charité in Paris focused on polysymptomatic aspects of the disorder. Briquet also recognized hysteria in men and attributed the disorder to emotional causes.

Modern Era An important series of papers published between 1951 and 1953 presented the first modern conceptualization of the multisymptomatic concept of hysteria. The Washington University group in St. Louis concluded that hysteria is a definable syndrome with a characteristic clinical picture that begins before the age of 35. Using objectifiable criteria, they defined a prevalence of the multisymptomatic disorder in the general hospital of 2.2 percent of all admissions. While noting the similarities of their work to Briquet's, they initially deviated by suggesting that men did not have the disorder.

In the early 1960s two studies confirmed the original findings of a definable clinical syndrome, demonstrating diagnostic stability of the multisymptomatic concept of hysteria. In 1970 the eponym *Briquet's syndrome* was proposed to denote multisymptomatic hysteria. The disorder, characterized by at least 25 symptoms from 10 symptom groups, was known as *Briquet's syndrome* until the publication of the DSM-III. Ironically, after the decision was made to incorporate Briquet's syndrome as part of the new diagnostic nomenclature, an unrelated decision was made to drop all eponyms. Hence a new name—*somatization disorder*—had to be created.

DSM-III streamlined the criteria to 14 lifetime symptoms in women (12 in men) from a list of 37 symptoms; moreover, a requirement for symptom grouping by organ systems was dropped. With the advent of DSM-III-R and further time for detailed follow-up studies, the number of symptoms required for men and women were both changed to 13.

Comparative Nosology There have been multiple predecessors to the current diagnosis of somatization disorder, the best validated of which has been Briquet's syndrome. Later studies have demonstrated only a moderate degree of diagnostic concordance between Briquet's syndrome and somatization disorder. In spite of these limitations, in most situations it appears reasonable to apply the findings from the literature on Briquet's syndrome and somatization disorder. Somatization disorder did not appear in DSM-I or DSM-II; it first appears in DSM-III and later in DSM-III-R. For DSM-IV the diagnostic criteria were simplified to require one or more symptoms from each of four symptom groups. The ICD-10 criteria for somatization disorder require (1) at least 2 years of multiple and variable physical symptoms with no adequate physical explanation; (2) persistent refusal to accept advice; and (3) some degree of impairment of functioning. There must be a total of 6 or more from a grouping of 14 symptoms in 2 of the 4 different areas (gastrointestinal, cardiovascular, genitourinary, plus skin and pain). There are no pseudoneurological symptoms in ICD-10 and DSM-IV has different gynecological and sexual symptoms from those found in ICD-10. Moreover, ICD-10 requires at least 6 symptoms from 4 groupings whereas DSM-IV requires 8 symptoms from its 4 groupings. Finally, ICD-10 is specific about duration of the disorder—2 years—whereas DSM-IV requires that some of the physical complaints have their beginnings before age 30.

Epidemiology Several studies based on large populations have estimated the lifetime prevalence of somatization disorder as 0.13 percent of the general population. Other community-based surveys estimate the lifetime prevalence to be 0.2 to 2 percent in women and less in men. Because patients with somatization disorder believe themselves to be medically ill, they can be assumed to frequent physicians' offices. Recent work indicates that as many as 5 percent of patients seen in family practice settings meet criteria for the disorder whereas 9 percent in a general hospital population and 12 percent in a group of patients with chronic pain met criteria for somatization disorder. In a group of patients with irritable bowel syndrome in a private outpatient clinic, 17 percent of patients met the criteria. In one random selection of 49 women who underwent hysterectomy for noncancerous reasons, 27 percent were diagnosed with somatization disorder.

Persons that meet the full criteria for somatization disorder are typically unmarried, nonwhite, poorly educated, and from rural areas. In families with somatization disorder, children have 11.7 times as many emergency room visits as families less afflicted with somatizing. Moreover, children's somatization is predicted by parental somatization, substance abuse, and antisocial symptoms.

Although there may not be a large number of these patients in the general population that meet full-blown criteria, individually they are expensive to care for. Recent studies estimate that somatization disorder patients incur 6 times the national average of per capital expenditures for hospitals and 14 times the national average of per capital expenditures for physician services. Moreover, there may be thirty times as many patients with subthreshold somatization, that is, those that meet partial criteria for somatization disorder (i.e., undifferentiated somatoform disorder).

Sex Early reports found somatization disorder exclusively in women. However, it is now recognized that somatization disorder afflicts men, but less commonly than women. Men comprise 5 to 20 percent of those with somatization disorder.

Comorbidity As the severity, chronicity, and invalidism of somatization increase, so does the likelihood of a major psychiatric disorder. Frequent concomitants of somatization disorder are major depressive disorder, anxiety disorders, and personality disorders. Over half of somatization disorder patients from a primary care sample were found to have a lifetime history of a major depressive episode in addition to their somatization disorder. Several recent studies using structured diagnostic interviews in patients from primary care settings found that 61 to 72 percent of patients with somatization disorder also have co-occurring personality disorders. This rate would seem to be 2.5 to 11.6 times more common in somatization disorder patients than in general medical patients.

Somatization disorder does not seem to be associated with any one or any combination of currently defined personality disorders. In one study, the most frequent types in order were avoidant, paranoid, self-defeating, obsessive-compulsive, schizotypal, and histrionic personality disorders. Earlier studies from patients attending psychiatric clinics had only reported an association with histrionic personality disorder and antisocial personality disorder. Many patients with somatization disorder also have conversion symptoms.

Anxiety disorders, especially phobias, panic disorder, and generalized anxiety disorder, are also present in patients with somatization disorder. Suicide threats are common in patients with somatization disorder, as are suicide gestures; however, suicide attempts rarely are lethal or near-lethal. Doctor shopping is frequent, with patients moving from one doctor to another. Typically, somatization patients have very chaotic social lives with frequent divorces, separations, and remarriages; similarly, they have trouble maintaining jobs and often become too disabled to hold gainful employment.

Etiology Numerous theories have been advanced to explain the psychosocial mechanisms involved in the process of somatization. By contrast few theories have been yet proposed to account for the biological basis of somatization disorder. The act of somatization can be understood as social and emotional communication. It can also be explained as the result of an intrapsychic dynamic, with somatization carried to an extreme.

Social Communication Somatization as social communication includes the use of bodily symptoms to manipulate or control relationships (e.g., an adolescent girl's developing unexplained abdominal pain to prevent her parents from going away for the weekend). Somatizing also can serve as emotional communication. Patients may be unable to verbally express their emotions; therefore, they may use somatic symptoms and somatic complaints to express their emotional state. Symptoms may be used to symbolically communicate emotions, as they are in conversion disorder. Some patients also use medical complaints as a coping device to deal with stress. Finally, physical symptoms may be used as a solution to an intrapsychic conflict, again as in conversion symptoms. Studies of psychological tests in somatization disorder have reported that compared to sex- and age-matched controls these patients have significantly more scale elevations on the MMPI-2.

Psychodynamic Factors Classical psychoanalytic theory has held that hysteria represents a substitution of somatic symptoms for repressed instinctual impulses. Freud postulated that the conflict was a phallic Oedipal one. However, more recent articles in the analytic literature emphasize a pregenital conflict as well.

Biological Factors Interesting preliminary data are now finally available concerning the biomedical underpinnings of somatization disorder. Neuropsychological testing demonstrates equal bifrontal impairment of the cerebral hemispheres and nondominant hemispheric dysfunction in patients with somatization disorder. However, some reports indicate that symptoms referable to the left side of the body may indicate that the right hemisphere of the brain is more involved than the left side. Preliminary evidence indicates that patients with somatization disorder may have an abnormality in cortical functioning, as evidenced by abnormal auditory-evoked potentials. In contrast to controls, patients with somatization disorder responded similarly to relevant and irrelevant stimuli, suggesting an impairment in selective attention. The data require much more extensive follow-through, without the confound of comorbid personality disorders influencing findings.

Christopher Bass at Oxford University is leading an effort to define the pathophysiology of symptoms experienced by patients with somatization disorder because not all patients with unexplained medical complaints have emotional disorders. The awareness, reporting, and seeking of medical help is complex and involves biopsychosocial processes. From a perceptual perspective, awareness of symptoms varies with the attention paid to them. Symptom reporting is elevated among

individuals who live alone or work in unstimulating settings. The awareness of symptoms and of disease is increased by reading about or observing medical disorders. Depressed and anxious moods are associated with greater reporting of bodily symptoms. Personality traits, neuroticism, and negative affectivity with introspective characteristics all tend to lower the threshold for noticing and reporting bodily sensations. Muscular tension, hyperventilation, sleep, and inactivity are all conditions that could contribute to the symptoms experienced by patients who have somatization disorder.

Considerable evidence now also points to familial and genetic associations in somatization disorder. Some data support the findings that groups of patients with somatization disorder have a higher-than-expected prevalence of antisocial personality disorder or manifest several traits of antisocial personality disorder. Other data, however, do not support such a strong association. One theory holds that antisocial personality disorder and somatization disorder may have a common genetic background. Some scholars consider somatization disorder to be the female expression of a genetic tendency, with antisocial personality disorder being its male counterpart.

Recent data is now emerging from a number of studies on families of patients with somatization disorder. In comparison to control families, those of patients with somatization disorder have had more physical and sexual abuse, and more disabling medical illnesses in one or both parents—all severe socio-environmental challenges. Thus somatizing behavior in many patients with somatization disorder seems often to emerge against a backdrop of emotionally disturbed or impoverished family life lacking in ordinary parental care in which physical symptoms had established meanings. Indeed Christopher Bass has argued that somatization disorder should be classified as a developmental personality disorder because of its patients' characteristic lifestyles and modes of relating to self and others. Relevant to this argument is evidence of persistent impairment associated with an illness-shaping background in an individual's formative years.

Diagnosis and Clinical Features Table 16-9 presents the DSM-IV diagnostic criteria of somatization disorder. It is important to note that specific symptoms do not need to be considered legitimate by the clinician. Rather, patient's reports that they have the symptom is sufficient as long as the symptom meets the severity criteria. A patient can have somatization disorder even if a current or presenting symptom did not begin before the age of 30 years. A careful review for an early onset of any of the unexplained symptoms for which the patient has had problems is necessary to make the diagnosis and at least one of these symptoms must begin before age 30.

Table 16-9 DSM-IV Diagnostic Criteria for Somatization Disorder

Patients with somatization disorder consider themselves to be severely ill. They report their health is worse than those with chronic, lesion-based medical conditions. In contrast, the mortality rate of somatization patients is similar to that of the general population and is substantially less than patients with major depressive disorder.

Ms. D. is a 52-year-old white woman who was referred to a general internist in the city for evaluation of persistent back pain and multiple other complaints. At hospitalization it was noted that the patient was disabled from her job as a machine operator at a shoe factory. Ms. D. gave a history of 10 operations: removal of a tumor from her right wrist, a dilation and curettage, a hysterectomy, three abdominal gastric operations, three breast biopsies, and leg surgery. She had received care from five different hospitals and seven different physicians in the past 2 years.

On physical examination, Ms. D. was an obese, chronically ill-appearing woman who came to the hospital wearing her transcutaneous electrical nerve stimulation unit. She was cooperative and showed her various scars with a certain amount of enthusiasm. The remainder of her physical examination was within normal limits except for a decreased range of motion in the area of her lumbar spine and local muscle guarding, with some tenderness. Spinal radiographs revealed some degeneration of vertebral bodies L2 to L5. On mental status examination she was cooperative and pleasant, and her behavior was somewhat seductive. There was no pressure or eccentricities in her speech. She showed little hesitation in discussing intimate details of her life. Her mood was euthymic; her affect was appropriate to mood but possibly a little shallow. The remainder of her mental status examination was within normal limits.

Disallowing all back-related symptoms, Ms. D. was positive for eight pain symptoms: four gastrointestinal symptoms, two sexual symptoms, and two pseudoneurological symptoms with an age of onset of 26 years. During the previous 12 months, Ms. D. reported that she had been in bed 21 days, had made seven office visits to four physicians, and had been hospitalized for a total of 52 days.

Ms. D.'s case illustrates that the diagnosis of somatization disorder can and should be made in the presence of comorbid medical conditions. Patients with somatization disorder do become ill, and their problems need to be appropriately diagnosed and treated. However, the management of somatization disorder should continue unchanged.

Pathology and Laboratory Examination There are no known neuropathological or routine laboratory findings specific for somatization disorder.

Differential Diagnosis An important aspect of the differential diagnosis is distinguishing a somatic symptom secondary to another psychiatric disorder from a symptom of somatization disorder. Table 16-10 explains how somatic complaints in other disorders differ from those found in somatization disorder.

Anxiety or depressive disorders	Usually one or two somatic symptoms of acute onset and short duration
Panic disorder	Somatic symptoms experienced only during panic episode
Hypochondriasis	Patient's focus is on fear of disease, not focus on a symptom
Conversion disorder	Only one or two complaints
Pain disorder	One or two unexplained pain complaints, not a lifetime history of multiple complaints

Table 16-10 Nonsomatization Somatic Symptoms

Hypochondriasis can sometimes be difficult to distinguish from somatization disorder. Hypochondriasis is more likely to begin at a later age than somatization disorder whereas the latter is more likely to occur in women than in men. Hypochondriacal patients often misinterpret normal bodily sensations as an indication of a pathological process. They believe that they harbor a specific lethal disease whereas somatization patients focus on complaints about multisystem symptoms that

lead them to have an inordinately high number of surgical procedures and consume an excessive amount of health care.

Course and Prognosis By definition, somatization disorder is a chronic relapsing condition with no known cure. It usually begins in middle to late adolescence, but may start as late as the third decade of life.

Typically, patients develop a new symptom, or symptoms, during times of emotional distress. No research data are yet available as to how long a modal episode of illness lasts. Clinical wisdom indicates that a typical episode lasts 6 to 9 months, with quiescent periods lasting from 9 months to a year. It is unlikely that patients with somatization disorder can go for more than a year without developing a new symptom and seeking some type of health care. Periods of psychosocial distress seem to coincide either with the onset of new symptoms or with increased health—care-seeking behavior associated with some preexisting symptom. Although no research data exist on whether stress precipitates the relapse, anecdotally there does seem to be an association. This association is especially problematic for patients because somatization disorder considerably disrupts social aspects of living.

Poor Long-Term Health Somatization disorder patients typically consider their health to be poor. When standard measures for health status are applied, the patients report that all aspects of their health—physical, social, and mental—as well as their general health perceptions are severely impaired. Patients with somatization disorder report worse health than do those with chronic medical conditions. Further, because patients who have somatization disorder perceive themselves to be sicker than the sick, it is not at all incongruous that they usually deem themselves disabled from work.

Treatment Because the cause of somatization disorder is unknown and no curative or ameliorative treatment has been found, the clinician needs to focus on management rather than treatment, on coping rather than curing (care rather than cure).

General Management Strategies While taking the medical history, physicians should take note of and store away for future use psychosocial material that it brought up spontaneously by the patient: references to personal difficulties and life events that cause distress. These matters can be explored at a later time after the relationship with the patient is more developed. It is important initially to avoid confrontations or to alienate the patient in other ways. The physical examination is an opportunity for discussing symptoms and providing clear information about clinical findings.

Information to the patient must be clear and unambiguous. During regularly scheduled repeat visits, as the patient endlessly drones on about physical symptoms, the clinician should gradually change the agenda to inquire about psychosocial issues in an empathic manner. In actual practice, this type of inquiry does not consume inappropriate amounts to time, contrary to clinicians' fears. Limits also need to be placed on the numbers and types of investigations that the patient requests. Crises in the patient's life, provoked by psychosocial stresses, are often accompanied by an increase in somatic symptoms. At such times the physician may be more able to draw the patient's attention to the relationship between emotional and physical suffering.

Best practice would indicate that the patient's primary care physician follows the patient closely and treats, or refers for treatment, any complicating comorbid psychiatric condition, such as a mood disorder. When ordering laboratory studies, the physician should order for the patient the same set of tests as if the patient were not a somatizer. If called into the case after all medical evaluations have been completed, the psychiatrist is advised to inquire about symptoms and the patient's medical odyssey, avoiding the temptation to express any opinions prematurely and maintaining a position of empathic curiosity during the initial intake in an effort to understand the patient's experiences.

Several types of treatment techniques have been advocated, including behavioral (avoid prolonged bed rest, which only increases weakness and encourages the patient to focus on symptoms); cognitive (avoid catastrophizing); and interpersonal (elucidate and deal with family conflicts). Progress should be aimed at improving functional capacity and the patient's progress will typically be very gradual.

Management of somatization disorder has only been tested empirically in several studies, findings from which revealed that when certain specific management strategies were undertaken by the primary care physician, patients with somatization disorder improved their physical functioning; their health care utilization decreased at the same time too.

Cost estimates of the savings of a simple consultation letter to these patients' primary care physicians about specific principles was associated with a 12 percent decrease in per-patient annual health care cost. Net annual cost savings were $295. Improvement to the patient is best reflected in the ability to perform the activities of daily living. Patients whose physicians received a simple consultation letter from a psychiatrist were significantly clinically improved in this domain compared to patients whose primary care physician did not receive such a consultation letter. Table 16-11 presents helpful management strategies for somatization disorder.

Establish primary care physician as patient's main and (if possible) only physician
Set up regularly scheduled visits every four to six weeks
Keep outpatient visits brief
Perform at least a partial physical examination during each visit directed at the organ system of complaint
Understand symptoms as emotional communication rather than the harbinger of new disease
Look for signs of disease rather than being symptom focused
Avoid diagnostic tests, laboratory evaluations, and operative procedures unless clearly indicated
Set a goal of getting at least selected somatization disorder patients referral-ready for mental health care

Table 16-11 Helpful Management Strategies for Somatization Disorder

There is little data on the treatment of depression in patients with somatization disorder, but what evidence there is suggests that depression can be successfully treated; when it is treated successfully, the somatization itself may also improve.

Doctor-Patient Relationship The cornerstone for successful management of the patient with somatization disorder is establishing a trusting relationship between the patient and one community physician. The doctor-hopping that frequently occurs in these patients is both frustrating and countertherapeutic.

Recently a form of group treatment was shown to be effective in somatization patients. The intervention tested was a time-limited, behaviorally oriented group with a structured protocol. The overall goals of the group were to be a source of peer support, to share methods of coping, to increase the ability to perceive an expressed emotion, and to allow the patient to enjoy the group experience. The study demonstrated that in the year following treatment the experimental group of patients with somatization disorder demonstrated better physical and emotional health and evidenced decreased health care charges than untreated controls with the same disorder. Moreover, the more group sessions the subjects attended, the greater their sense of well-being and the lower their net health care costs. Other studies using short-term cognitive-behavioral therapy, individually or in groups, for somatizing or somatization disorder patients have demonstrated decreased somatic preoccupation, anxiety, depression, and medication usage.

UNDIFFERENTIATED SOMATOFORM DISORDER

Definition Undifferentiated somatoform disorder is characterized by one or more unexplained physical complaints of at least 6 months' duration. These symptoms impair the patient in some domain and are temporally associated with a stressor. Psychological factors are assumed to be associated with the symptoms or complaints because of a contemporaneous relationship between the initiation or exacerbation of the symptoms and stressors, conflicts, or needs. The complaint must be unattributable to any other known psychiatric condition or pathophysiological mechanism or, when it is related to a nonpsychiatric condition, the physical complaints or resulting social and occupational impairments must be grossly in excess of what would ordinarily be expected from the findings.

History and Comparative Nosology Many who work in the general medical setting find the diagnosis of undifferentiated somatoform disorder helpful. Research indicates the validity or distinguishing undifferentiated somatoform disorder from somatization disorder. There appears to be a dimensional or quantitative difference

between undifferentiated somatoform disorder and somatization disorder rather than a qualitative difference between the two. However, the natural history of both disorders seems to be similar. The disorder was not included in DSM-I, DSM-II, and DSM-III; it was first introduced in DSM-III-R and remains unchanged in DSM-IV. Undifferentiated somatoform disorder was introduced to the psychiatric lexicon in DSM-III-R because somatization disorder was considered to be too restrictive by primary care providers to provide adequate coverage for many patients with significant somatoform complaints. Thus the subsyndromal grouping was formalized to facilitate learning about the natural course of a large cluster of patients. The criteria for undifferentiated somatoform disorder in ICD-10 are similar to the diagnostic criteria in DSM-IV.

Somatizing Syndrome This disorder, characterized by a lifetime history of four unexplained somatic complaints for men and six for women, has been called various names. It forms a subset of patients with undifferentiated somatoform disorder. All research data so far have used the term *somatization syndrome* rather than the broader term *undifferentiated somatoform disorder.* The remainder of this section refers to data from research on somatizing syndrome.

Epidemiology The importance of undifferentiated somatoform disorder comes from the fact that it may be 30 to 100 times more prevalent than full-blown somatization disorder. Undifferentiated somatoform disorder has an estimated lifetime prevalence in the general population of between 4 and 11 percent with an estimated 1 percent 6-month prevalence. Javier Escobar has described a group of patients with this disorder who have six unexplained symptoms for women and four for men; these patients have many of the same socioeconomic and clinical manifestations of somatization disorder.

Undifferentiated somatoform disorder typically affects women. It has also been shown in some reports to be associated with lower socioeconomic status, older age, and Hispanic or African-American origin. One study found a slight association with antisocial personality disorder; however, other studies have not confirmed this. One recent study reported these patients had higher warning salivary cortisol concentrations and higher heart rates under the stress of a mental task and less habituation than controls. Presumably, those with undifferentiated somatoform disorder are comorbid with at least one other lifetime psychiatric disorder, but this has not yet been firmly established.

Approximately 50 percent of patients with the disorder have comorbid psychiatric conditions (e.g., depression, anxiety, personality disorders) compared with 7 percent in the general population.

Etiology There are multiple theories about the process of somatization; none of which is specific to undifferentiated somatoform disorder.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria are listed in Table 16-12. Patients with undifferentiated somatoform disorder have more comorbid psychiatric diseases than do general medical patients without the disorder. These comorbid conditions are primarily depressive disorders, anxiety disorders, and personality disorders.

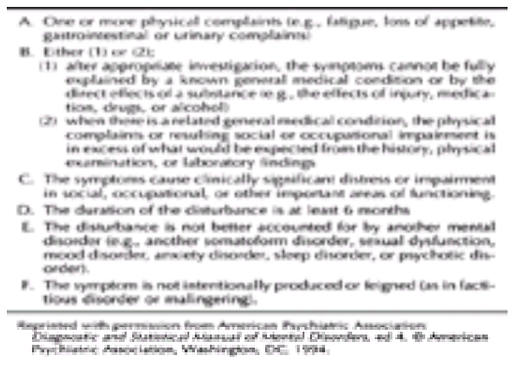

A. One or more physical complaints (e.g., fatigue, loss of appetite, gastrointestinal or urinary complaints)
B. Either (1) or (2):
   (1) after appropriate investigation, the symptoms cannot be fully explained by a known general medical condition or by the direct effects of a substance (e.g., the effects of injury, medication, drugs, or alcohol)
   (2) when there is a related general medical condition, the physical complaints or resulting social or occupational impairment is in excess of what would be expected from the history, physical examination, or laboratory findings
C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The duration of the disturbance is at least 6 months
E. The disturbance is not better accounted for by another mental disorder (e.g., another somatoform disorder, sexual dysfunction, mood disorder, anxiety disorder, sleep disorder, or psychotic disorder).
F. The symptom is not intentionally produced or feigned (as in factitious disorder or malingering).

Reprinted with permission from American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 16-12 DSM-IV Diagnostic Criteria for Undifferentiated Somatoform Disorder

Many patients with this disorder have illnesses that wax and wane but continue to play a persistent role in their lives. In a study of 46 patients with chest pain, those without coronary artery disease had a higher rate of anxiety disorders than those with coronary artery disease. Eleven years later, those with chest pain but without coronary artery disease were found to have led lives of disability and chronic distress. They had had repeated interruptions in their lives associated with functional symptoms. Those in the study with chest pain without coronary disease also had higher research scale ratings for depression and anxiety and reported more physical symptoms.

Pathology and Laboratory Examination There are no specific pathological or routine laboratory features for undifferentiated somatoform disorder.

Differential Diagnosis In the differential diagnosis of undifferentiated somatoform disorder care must be taken to differentiate somatic symptoms that could be a part of many other psychiatric disorders. For example, somatic symptoms frequently occur in major depressive disorder. Similarly, the somatic delusion of schizophrenia and major depressive disorder also needs to be differentiated from undifferentiated somatoform disorder.

Several other psychiatric disorders need to be differentiated from undifferentiated somatoform disorder. Somatization disorder is a more severe disorder characterized by a chronic history of multiple unexplained somatic complaints that fit a specific pattern and begin before the age of 30. Adjustment disorder with physical symptoms may also evidence unexplained somatic complaints; however, the duration of this disorder is by definition less than 6 months. Finally, the diagnosis of psychological factors affecting medical condition may at first appear similar; however, the disorder affects a known Axis III disorder rather than mimicking an Axis I disorder.

Course and Prognosis The course of undifferentiated somatoform disorder is generally chronic and relapsing; however, little systematic research on the disorder has been accomplished to date. It is likely that some cases of the disorder can resolve after a single episode. There is substantial disability, work impairment, and excessive health care utilization in undifferentiated somatoform disorder. The dysfunction is not seen to the same extent as in somatization disorder. Extensive comorbidity has been shown for somatization disorder with depressive, anxiety, and personality disorders. With the less restrictive criteria for undifferentiated somatoform disorder, there is also considerable comorbidity, although the rates are not quite as high.

Treatment Recent studies have indicated that patients with undifferentiated somatoform disorder respond to the same treatment or management approach as patients with somatization disorder.

One recent study indicates that somatizing syndrome patients of physicians who received a simple psychiatric consultation letter explaining management principles of intervention were able to lower the annual medical care charges of these patients by $289 (in constant 1990 dollars), which equals a 32.9 percent reduction in the annual median cost of their medical care. In addition, these atients experienced clinically significantly improved physical functioning after the intervention.

HYPOCHONDRIASIS

Definition Hypochondriasis is a disorder characterized by preoccupation with the fear of developing a serious disease or the belief that one has a serious disease. The fear is based on the patient's interpretation of physical signs or sensations as evidence of disease even though the physician's physical examination does not support the diagnosis of any physical disorder. The unwarranted fear of or belief in a diseased state persists in spite of medical reassurance. However, the belief does not have the certainty of delusional intensity.

History The concept of hypochondriasis has been a part of medical lore since ancient times. Prior to the early nineteenth century the area of the body below the rib cage, the abdomen, was called the hypochondrium. Thus, hypochondriasis referred to somatic complaints occurring in the abdomen. In the late 1920s, R.D. Gillespie provided the first modern description of the disorder.

Comparative Nosology Hypochondriacal symptoms can be a part of another disorder such as major depressive disorder, dysthymic disorders, generalized anxiety

disorder, or adjustment disorder. However, primary hypochondriasis or hypochondriacal disorder is a chronic and somewhat disabling disorder with hypochondriacal symptoms, not merely a part of another psychiatric condition.

Hypochondriasis was included as a diagnostic entity in DSM-I. The diagnostic criteria continued to be revised in DSM-II, DSM-III, and DSM-III-R; however, the changes have been primarily linguistic, not substantive. The only change between DSM-III-R and DSM-IV is the addition of a specifier to note that the patient has poor insight during the current episode. The ICD-10 criteria for hypochondriasis are essentially the same as those of DSM-IV.

Epidemiology Hypochondriasis is rather common in primary care settings. In various settings, the prevalence has varied from 3 to 14 percent. Recent work indicates that in a 6-month period of observation, 4 to 6 percent of the general medical population meet the specific criteria for this disorder. The prevalence in either sex is comparable to that within the general medical population. There are no specific tendencies for overrepresentation based on social position, education, marital status, or other sociodemographic descriptors. There is a wide range of ages at onset. Although the disorder can begin at any age, onset is thought to be most common between 20 and 30 years of age. A preliminary family study of 19 cases and their 72 first-degree relatives demonstrated no increase in the rate of hypochondriasis among their relatives compared with a control group.

Comorbidity with other psychiatric disorders is common with hypochondriasis, and must be treated accordingly. As an example, when case-matched controls and 42 hypochondriasis patients from a medical clinic were evaluated psychiatrically, the hypochondriasis patients had twice as many lifetime Axis I disorders and three times the number of personality disorders. Of the hypochondriasis patients in the study, 88 percent had one or more additional Axis I disorders, the overlap being greatest with depressive and anxiety disorders. Depression only accounts for a minor part of the total picture in hypochondriasis, so it is a mistake to think that all hypochondriasis is the result of some other Axis I disorder.

Etiology There are four major etiological theories concerning hypochondriasis: (1) amplification of normal bodily sensations; (2) psychodynamic formulations; (3) social learning concepts; and (4) syndromic variant of some other psychiatric condition, such as a mood or an anxiety disorder.

Amplification The amplification hypothesis posits that hypochondriasis results from the augmentation of normal bodily sensations. Of the four hypotheses it has the most research interest: hypochondriacal patients might amplify their normal somatic sensations and then attribute them to some sort of pathological condition. For instance, a change in a patient's perception of peristalsis might be interpreted as abnormal, hence representing disease. However, studies of hypochondriacal patients' awareness of their own normal heart-beat now shows that they are no more accurate than are a comparison group of nonhypochondriacal patients. Moreover, in a test of tactile sensitivity, patients with hypochondriasis did not have a greater ability to discriminate on two-point touch. Yet hypochondriacal patients consider themselves to be more sensitive to benign bodily sensations and certainly report more functional somatic symptoms. When hypochondriacal patients develop clinical arrhythmias, they are not more sensitive to or more accurately aware of subtle changes in cardiac activity than nonhypochondriacal patients. A hypochondriacal patient's behavior would seem to indicate a response bias toward reporting somatic and physiological distress with an attributional tendency to invoke somatic pathology.

Psychodynamic Factors A variety of hypotheses purport that intrapsychological factors are responsible for hypochondriasis. These factors run the gamut from Freud's early theories about disturbed object relations and intense preoccupation with the self to the concept of hypochondriasis as ego defense mechanisms against guilt.

Other theorists advocate that in hypochondriasis aggressive and hostile wishes toward others are transferred into physical complaints via repression or displacement. The frustration or anger that these patients often express can be theorized to be caused by past losses, rejections, or disappointments. Sometimes these patients express anger by first soliciting and then rejecting the help and concern of others. Alternatively, hypochondriasis may be viewed as a defense against guilt, a result of low self-esteem, or a sign of excessive self-concern. Pain and somatic suffering then symbolically become a means of atonement or can be experienced as deserved punishment for past real or imagined wrongdoing.

Learning Theory Learning theory postulates that psychosocial learning has a strong etiological component in hypochondriacal disorder. The concept contends that a patient learns the sick role and that role is sufficiently reinforced through either social contact or some need gratification. Interestingly, when compared to case-matched controls, patients with hypochondriasis believe that to be in good health means to be relatively symptom free, and they consider the presence of symptoms as indicative of sickness. Hence, personal theories of attribution, when coupled to a catastrophizing cognitive style, may be relevant to the focus on illness. The sick role then becomes a means of receiving caretaking from others.

Variants The variant theory holds that hypochondriasis is a modification of some other psychiatric disorder, such as depressive disorders, anxiety disorders, and certain personality disorders, such as obsessive-compulsive personality disorder. Although some scholars have maintained that hypochondriasis is a variant of a depressive condition, current research has not supported this hypothesis. Some theorists believe that hypochondriasis is a variant of an anxiety disorder, such as obsessive-compulsive disorder or a panic disorder, but the research data on this is not yet compelling.

Diagnosis and Clinical Features Table 16-13 lists DSM-IV criteria for the diagnosis of hypochondriasis. The developmental background of hypochondriacal patients is of interest in that significantly more of these patients than matched controls report traumatic sexual contacts, physical violence, and major parental upheaval before the age of 17. Significantly more hypochondriacal patients also report being sick as children and missing school time for health reasons; thus, they recall more childhood trauma and illnesses even though they are not currently more medically sick.

A. Preoccupation with fears of having, or the idea that one has, a serious disease based on the person's misinterpretation of bodily symptoms.
B. The preoccupation persists despite appropriate medical evaluation and reassurance.
C. The belief in criterion A is not of delusional intensity (as in delusional disorder, somatic type) and is not restricted to a circumscribed concern about appearance (as in body dysmorphic disorder).
D. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
E. The duration of the disturbance is at least 6 months.
F. The preoccupation is not better accounted for by generalized anxiety disorder, obsessive-compulsive disorder, panic disorder, a major depressive episode, separation anxiety, or another somatoform disorder.

Specify if:
With poor insight: if, for most of the time during the current episode, the person does not recognize that the concern about having a serious illness is excessive or unreasonable

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 16-13 DSM-IV Diagnostic Criteria for Hypochondriasis

Patients typically present for the first time with the complaint of a symptom, pain, or sensation. With some gentle questioning they quickly move from concern about the symptom to fear of a disease. Almost uniformly these patients are not concerned transiently about a minor disease, but rather are persistently worried about a serious disease. The specific disease may change from over time. The duration of an episode of a feared disease may run from months to years; alternatively, the feared disease may not change at all throughout the course of the disorder.

Pathology and Laboratory Examination At present there are no known pathological somatic features of the disorder nor are there known routine laboratory or electrophysiological tests that help to elucidate this disorder.

Differential Diagnoses Essential to the process of differential diagnoses is ruling out underlying organic disease. In the vast majority of cases this can be completed by the primary care physician without referral to a specialist. The workup is usually a straightforward process and typically focuses around the disease that the patient is concerned about.

Somatization disorder is the main somatoform disorder that needs to be differentiated from hypochondriasis. Although there are numerous differences between somatization disorder and hypochondriasis, the major distinction is that patients with somatization disorder are concerned about their symptoms, even dramatizing them, and are relatively indifferent to concerns about underlying disease. Patients with hypochondriasis have an exactly opposite preoccupation; they fear an

underlying disease and concern about symptoms quickly fades. Typically, hypochondriacal patients do not have the plethora of symptoms present in somatization disorder patients. Hypochondriasis needs to be distinguished from factitious disorder with predominantly physical signs and symptoms and from malingering. In hypochondriasis, patients actually experience the symptoms they report rather than simulating them.

Patients with psychotic disorders, particularly major depressive disorder and schizophrenia, may have somatic delusions or concerns about the presence of a disease. Hypochondriacal concerns secondary to major psychiatric disorders are categorized with the more serious disorders. Candidate conditions include major depressive disorder, dysthymic disorder, generalized anxiety disorder, obsessive-compulsive disorder, and panic disorder.

Course and Prognosis Hypochondriasis is a chronic, relapsing condition with waxing and waning symptoms. The disorder is usually longstanding, frequently over several years. Episodes typically last months and even a few years. There are often quiescent periods between episodes; however, frequently there are recurrences during times of psychosocial distress.

Symptom severity of the disease is such that some degree of psychosocial impairment occurs. Typically, familial and spousal relationships become strained by living with a patient who persistently fears disease. The individuals ability to work may or may not be affected.

Historically hypochondriasis has been given a pessimistic prognosis; however, that reputation may be exaggerated. Some authors now state that approximately 50 percent of patients show improvement; the remainder show a chronic, fluctuating course. A recent case-matched study to examine diagnostic stability and outcome of 50 cases meeting DSM-III-R criteria indicated that after 1 year two thirds of the subjects continued to meet criteria for hypochondriasis and the remaining one third had persisting hypochondriacal symptoms. The hypochondriacal subjects were improved on most measures. However, severe symptoms, longer duration of illness, and coexisting psychiatric illness were predictive of a worse outcome. In a 2-year follow-up study of 24 patients with transient hypochondriacal symptoms that did not meet DSM-III-R criteria for hypochondriasis, these patients continued to have their hypochondriacal symptoms with only one patient going on to meet full DSM-III-R criteria of the full-blown disorder. These patients continued to magnify the importance of their physical symptoms, had more functional disability, and utilized more medical care than case-matched controls. Thus, recent research has validated the clinical view that the diagnosis of hypochondriasis appears to be stable over time, with symptoms waxing and waning. Hypochondriasis carries a very substantial, long-term burden of morbidity, functional impairment, and personal distress.

The following characteristics, which bode well for a patient's general health status, bode well for patients with hypochondriasis as well: a high socioeconomic status; the presence of other treatable conditions, such as an anxiety or depressive disorder; a sudden onset; an absence of a personality disorder; and the absence of comorbid medical disease.

Treatment To date there are no controlled clinical trials on which to base rational treatment, therefore inferences have to be made from the treatment of other disorders and from clinical lore used in treating hypochondriasis. The simplest treatment approach is to look for and treat any comorbid psychiatric conditions, such as obsessive-compulsive disorder, panic disorder, and depressive disorder. When these conditions are treated pharmacologically and with appropriate therapy, the hypochondriasis often will improve.

A recent open trial of high dosages of fluoxetine (Prozac) on hypochondriasis patients not meeting criteria for comorbid depressive disorders showed much promise, with 10 of 16 patients much improved at the end of a 12-week trial. Several other trials with selective serotonin reuptake inhibitors (SSRIs) with primary hypochondriasis have also shown some positive results. Previously this group of patients had been regarded as refractory to all psychopharmacological treatment.

The management of hypochondriasis has typically been in the domain of the primary care physician. At least initially, these patients strongly resist psychiatric referral. Their mental model of the problem is that they have a covert physical illness and so they do not believe they need to see a psychiatrist. Secondly, psychiatry's track record using traditional psychotherapies for these patients has been rather dismal. Psychoanalytic treatment and sometimes even psychodynamically oriented therapy of patients with full-blown primary hypochondriasis have been unsuccessful in removing symptoms and improving the function of these patients. One pragmatic approach that has heuristic merit is the use of paradoxical intention, telling the patient that you recognize the suffering but that you are not certain that you can help very much. When little is expected, sometimes small gains are appreciated and there is some improvement.

Another approach that may possess merit is group treatment. A cognitive-educational group treatment has recently been proposed that at face value appears to have validity, especially given the success demonstrated by a similar approach in patients with somatization disorder. Currently several trials of cognitive-behavioral treatment, both in groups and individually, are ongoing.

Hypochondriasis patients usually accept referral for treatment of a comorbid psychiatric condition, and such conditions are not rare for them. Finally, management suggestions that have been shown to be effective in somatization disorder should be tried in patients with hypochondriasis.

PAIN DISORDER

Definition With their expertise in the use of psychoactive medication plus their interest in the personal and family dynamics of patients, psychiatrists have the capacity to be involved in the treatment of patients with chronic pain. Astute clinicians have long known that some patients use their pain as a way of seeking human relationships; the secondary gain they receive from their infirmity then can be reinforced, maintaining the pain.

Co-occurring psychiatric disorders are not rare when pain is severe. It is always gratifying when a pain syndrome dissipates as an accompanying psychiatric disorder is treated, both receding together. However, the clinical question remains: was pain the cause of a coexisting psychiatric disorder or was pain the consequence of another psychiatric syndrome? One study found that in 89 percent of one clinic's patients who had both depression and pain, there was no antecedent history of depressions. This data would be consistent with the hypothesis that pain often causes depression rather than vice versa.

Pain is a highly subjective matter and is extremely difficult to quantify. For example, it has always been difficult to specify just how much chronic pain ought to be associated with a given lesion. Assessing the degree to which emotional factors are intensifying the patients' complaints is also complicated. Moreover, patients with chronic pain are not always immediately motivated to give an intricate personal and family history because they fail to see its relevance to their pain.

There is no accepted objective way to measure pain, and a variety of different measures of pain intensity and severity are currently used in the field, which makes it difficult to compare different clinical studies. Finally, many studies in the pain literature have demonstrated considerable differences in study populations, with different personalities and different cultures experiencing and expressing themselves in divergent fashions.

Wilder Penfield's clinical wisdom about patients with chronic pain has now become rather widely accepted in the field: if the patient is not a malingerer, the complaints about the extent of the pain are to be believed. The issues that need to be explored clinically are: (1) how disabled by the pain is the patient and (2) to what extent are there complicating emotional factors and comorbid psychiatric conditions. The physician attendant to the suffering of a patient with chronic pain needs to render a diagnosis that communicates important underlying aspects of the case and leads to a treatment plan that encompasses the patient's emotional needs.

Comparative Nosology There have been a variety of informal and formal diagnostic terms associated with pain. For example, the euphemistic term *atypical* has been used, as in *atypical facial pain*, or *atypical pelvic pain*, to indicate that psychiatric issues needed to be considered in the cause or the maintenance of the severe pain syndrome. ICD-9 used the term *psychalgia* and DSM-III incorporated the term *psychogenic pain.*

Previous terms such as *chronic pain syndrome, psychogenic pain*, and DSM-III-R's *somatoform pain disorder* were used as psychiatric diagnoses for patients in pain wherein psychological factors have played a pronounced role. DSM-III-R required preoccupation with pain for at least 6 months and either no organic pathology or no pathophysiological mechanism found after appropriate evaluation; if organic pathology is present, the extent of the complaints about the pain or the degree of social-occupational disability should be greater than could be expected by the examiner based on the physical findings.

Applying the diagnostic terminology of DSM-III and DSM-III-R, however, has proved difficult. Even when diagnostic criteria required only no organic pathology or impairment in excess of what would be expected, precise application of criteria was problematic. Judging whether the amount of a patient's subjective pain is excessive became a daunting task and lacked validity and reliability studies. Furthermore, these diagnoses failed to adequately describe persons who have a physical cause for pain but react to their pain in a dysfunctional manner.

DSM-IV provides a more focused approach, not relying on the extent of the pain as a criterion but asking the clinician to make a judgment as to whether psychological factors have a major role in the onset, severity, exacerbation, or maintenance of the pain, and as to whether or not there are factors from the general medical condition that also make a major contribution to the symptoms.

DSM-IV divides pain disorder into *acute* and *chronic* subsets. Another new feature of DSM-IV is that it splits pain disorders into three subtypes: pain disorder associated with psychological factors, pain disorder associated with both psychological factors and a general medical condition, and pain disorder associated with a general medical condition (which is not considered to be a mental disorder but is included to help in the differential diagnosis). Patients with chronic pain do not readily fall into dichotomous roups, those either with physical or with mental illness. According to one student of chronic pain, “Many patients have both; a few have neither.”

ICD-10 labels pain disorder as “persistent somatoform pain disorder.” ICD-10 does not subdivide pain into subtypes in which psychological factors play a major role along with a general medical condition and in which psychological factors are the determinant of pain. ICD-10 excludes backache not otherwise specified, tension headache, muscle tension, and migraine. Additionally, ICD-10 retains more of a DSM-III-R approach regarding emotional factors involved in the cause of the pain, indicating that “persistent somatoform pain occurs in association with emotional conflict or psychosocial problems that are sufficient to allow the conclusion that they are the main causative influences.” ICD-10 is inconsistent in that it does not specify, as DSM-IV does, that emotional factors can impinge on certain aspects of the pain condition (onset, severity, exacerbation, or maintenance). Finally, ICD-10 goes beyond DSM-IV's statement that “pain... is the predominant focus of clinical attention and is of sufficient severity to warrant clinical attention” when it describes the outcome of persistent somatoform pain disorder in this way: “Usually [there is] a marked increase in support and attention, either personal or medical.”

Epidemiology Pain of one sort or another is among the most frequent reasons that patients consult their physician. One study found that 13 percent of patients in a private internal medicine practice suffered from chronic pain. In one sample of health maintenance organizations, 8 percent of patients had severe persistent pain and 2.7 percent had severe persistent pain with at least 7 or more days of pain-related activity restriction in the past 6 months. Most pain is remediable, but chronic pain carries with it a high price: more than $10 billion was spent in 1980 in disability payments for chronic pain problems. Recent data from numerous local and regional surveys leads to a yearly estimate of $60 billion as the annual cost to the nation from health care costs and loss of work from patients of chronic pain. Low back pain alone currently affects more than 7 million Americans. Another cost of pain disorders is suicide, with one report stating that the suicide rate among pain sufferers is 9 times that of the general population.

Because diagnostic criteria for the disorder have been changing (from DSM-III to DSM-III-R to DSM-IV), detailed investigations that pull together the databases are still lacking. Studies seem to indicate that there are twice as many women as men with pain disorder, with peak incidence in the fourth and fifth decades, especially in those in blue-collar jobs. There is also some tendency for familial aggregation, along with more depression and more chemical dependency use disorders in those families.

Comorbidity There is a high frequency of comorbid psychiatric diagnoses in most groups of chronic pain patients. One study reported that 56 percent of their sample of 283 chronic pain patients suffered from one of the depressive disorders and that alcohol or drug dependence occurred in 15 percent. Finally, 58 percent met criteria for a personality disorder (most frequently dependent, passive-aggressive, or histrionic). Since these studies of personality disorder are retrospective, it remains unclear whether the pain itself has caused the patient's personality traits to become a personality disorder, or whether having a personality disorder predisposes one to developing a pain disorder.

Etiology Pain comes from many sources, but the peripheral afferent nerves, central nervous system processing, and central nervous system interpretation of the localization and the severity of the pain are almost always involved. In some cases of pain disorder of the psychological or the combined type, a purely neurosurgical approach (eliminating a painful lesion and its connectivity) can successively chop away at parts of a person's peripheral and central nervous system until the cortex is no longer in contact with the relevant nerves or with the involved spinal cord from an afflicted area. In some cases the pain can still continue.

No adult registers an experience on a tabula rasa; rather, persons all encounter life happenings in a context of (1) personal training based on critical developmental incidents; (2) family-centered culture transmitted early in life; and (3) societal values about appropriate behaviors, including the expression of feelings. These past experiences all interact with a person's constitutional hardwiring, including preprogrammed response modes.

Chronic pain syndromes are among the most frustrating conditions in all of medicine for patients and physicians. To even have a chance at being successful in dealing with patients in pain clinicians need to be able to appreciate the patient in a biopsychosocial context. Successful academic pain clinics have discovered the necessity of a multidisciplinary approach. Indeed, physicians caring for such a patient need to do more than just take a detailed history; they also need to understand the patient's life from the patient's as well as the family's perspective.

For reasons they may or may not be aware of, some individuals magnify a given pain; others pay little attention to it, diminishing their response to a lesion that might otherwise cause great discomfort. Pain thresholds, which can be clinically measured, vary from person to person. Pain tolerance, likewise, has considerable latitude. Moreover, there is a broad range of ways in which individuals express their discomfort or even their agony. Some with a broken ankle will be stoic; others with the same fracture will scream out in agony, facies pinched, arms flailing.

Although there are conditions in which persons do not have pain fibers (familial dysautonomia), this is so rare that they are not a factor in everyday practice. It is easier to divide the discussion of issues in pain into four categories: intrapersonal factors, interpersonal factors, ethnic factors, and biological factors.

Intrapersonal Factors Intrapersonal factors can impact how pain is experienced. Psychodynamically formative events in the early years can carry over later into how one processes illness and other types of distress. One psychodynamically oriented theory is that having pain and expressing distress attracts others to deliver needed attention. In later years the experience of pain can be associated with obtaining love from an otherwise absent significant caregiver.

Another psychodynamically oriented theory holds that painful experiences, such as repeated severe corporal punishment from an abusive parent, can set the stage many years later for the person's response to a wrongful piece of behavior with self-imposed retribution. In this way a pain disorder may be used to pay for some self-perceived sin or transgression. However, simple interpretation (illuminating the repressed, suppressed, or even grudgingly embellished facts) rarely serves to remove the pain dramatically.

Recent research data suggests that pain patients have a background of past sexual abuse or trauma and the resultant posttraumatic stress disorder. Some patients can be vaguely or poignantly aware of the associations between past life events and being pain prone or even being overresponsive to a painful experience. For other pain patients, the associated early-life experiences of pain are deeply buried and emerge only through a lengthy process of free association, dream work, or psychological testing.

Not every one with severe pain and unusual pain behavior has had dysfunctional early-life experiences that caused a psychopathological organization of the character structure. However, astute clinicians are always alert to the possibility that there may be important masochistic, dependent, or narcissistic features to the personality, or traumatic life experiences that impact how a given patient handles pain.

No one psychological test can accurately delineate the extent of the psychological component in a pain syndrome. Any painful chronic disease highlights and even changes certain components of personality. Psychological testing and patient interviews can yield clear evidence that some patients with considerable emotional overlay to their pain are at least initially resistant to suggestions that their pain symptoms are caused by anything other than physical factors.

Finally, it is likely that intense pain affects attitudes and emotions, and that attitudes and emotions influence the reported perceptions of pain intensity.

Interpersonal Factors Interpersonal factors may also impact on the pain experience and behaviors can be reinforced or inhibited. For example, reinforcements given by others important to the pain-afflicted person can play a major role in day-to-day and even second-to-second experiences. Lavishing attention on an individual when chronic pain intensifies creates a different inner and outer world for the patient than when the pain is either ignored or treated with additional painful experiences.

Some persons use their pain to manipulate others. Clinicians often observe this when there is a chaotic and otherwise inattentive family or in the case of a dysfunctional marriage that becomes reorganized around the continued crisis of chronic pain. Surprisingly, in some persons the persistence of pain is far less

important than what it is traded for: a relationship that is far more to the patient's liking.

The consequences of pain are often gain. Secondary gains, such as monetary rewards or avoidance of distasteful activities, are not an unusual consequence of having prolonged pain. It is only when interpersonal and other rewards are stripped away from the pain experience and its attendant disability that the individual with pain can develop a more normal life.

Like a somatization disorder, a pain disorder can become a way of life that alters the behavior of other people. The caring physician must always be aware of interpersonal factors. Family, financial, and legal dynamics of a situation can have a dramatic influence on the patient's pain experience and pain behavior.

In one study, male patients with chronic pain who received financial compensation following injury at work showed a particularly high rate of personality disorder (83 percent) which supports a contention in the literature that patients with “compensation neurosis” were frequently passive-aggressive, hostile, and inadequate.

Ethnic or Cultural Factors Quantitative and qualitative research projects of chronic pain sufferers in New England and Puerto Rico by the same investigators reinforce what astute clinicians have known all along—that similar lesions in patients from different backgrounds can produce profound variations in pain experience and pain intensity. The style of the typical taciturn “old Mainer” is quite different from the emotionally labile and highly expressive native Puerto Rican. Despite higher pain intensity and more emotional responses among Puerto Ricans, there was no significant difference between the two groups regarding interference in daily activities. Attitudes, beliefs, and emotional and psychological states can differ according to ethnic group affiliation. Moreover, the individual's psychological coping style and internal or external locus-of-control style, which are important determinants of how an individual handles the pain, are to a considerable extent culturally determined.

Biological Factors Biological factors also play a large role in the onset and perpetuation of chronic pain. Cortical and subcortical centers process and filter afferent impulses. Sensory and limbic structural abnormalities can determine the severity of a pain experience. Neural pathways serving pain are not “hard wired,” but demonstrate considerable plasticity in response to injury. After an injury, changes occur in the sensitivity of the pain receptors themselves and also in the excitability of neurons in the dorsal horn of the spinal cord. Thus symptoms of chronic pain following injury may have a physiological basis in changes in the brain, spinal cord, and peripheral nerves.

Neurological changes resulting from recalibration of pain thresholds may be the result not of gross external injuries but rather of nonobvious injury resulting from mechanical stresses such as increased muscle tension (associated with anxiety or depression), and mechanical stresses of the lumbar spine. Additionally, there could be an associated muscular ischemia with the release of pain-producing substances such as substance P and histamine that alter the pain threshold.

The gate control theory purports that a gating mechanism in the dorsal horn of the spinal cord handles afferent pain signals; competing signals as well as neurotransmitters then can open or close the gate on painful perceptions. Endorphin deficiency seems to correlate with the augmentation of afferent stimuli. Serotonin, presumably relatively diminished in some forms of depression, has now been implicated as the main neurotransmitter in descending inhibitory pathways. Such mechanistic studies still remain preliminary but offer interesting explanations into a patient's plight when pain is burdensome but organic lesions do not appear major.

Thus, the underlying psychodynamic and neurobiological framework can impact on the complex ways a given person experiences and re-experiences pain. Pathophysiology and psychopathology may well be reciprocally linked in self-perpetuating cycles.

Diagnosis and Clinical Features Pain disorder as a psychiatric condition is diagnosed when a patient's preoccupation with pain is consuming and to some extent disabling. That is, pain becomes the predominant focus of the clinical presentation and the pain itself causes clinically significant distress or impairment and the patient's life becomes organized around the pain; also, psychological factors are judged to play a role in this disorder. Table 16-14 lists the diagnostic criteria for pain disorder.

Table 16-14 DSM-IV Diagnostic Criteria for Pain Disorder

Pain disorder patients make repeated visits to physicians, often successively or even concomitantly and doctor “shopping” is very common. There is a risk of excessive use of narcotic or sedative hypnotic agents because the patient is enveloped by pain and often has associated complaints of anxiety, depression, and insomnia. Thus, any related comorbid psychiatric conditions must also be treated to alleviate the patient's pain.

The major feature of pain disorder is an all-encompassing focus on the pain such that any other concern takes on less than its ordinary significance. The pain is believed by the patient to be the source for all the patient's misery.

If the patient is depressed, anxious, or has insomnia, those symptoms are almost always considered by the patient to be secondary to the pain. Other frequent ancillary symptoms include changes in appetite, loss of energy, decreased interest in social activities, decline in sexual interest, diminished physical exercise, and even diminished exertion. Also often noted are a change in normal recreational pursuits, a breakdown in family relationships, an increased amount of time spent in bed or lying down, an increased amount of self-absorption (including hypochondriasis), and multiple drug use or abuse. Frequent visits to physicians and requests for medical or even surgical approaches to obtain pain relief are commonplace.

One of the possible predisposing factors for handling pain poorly is childhood sexual abuse, for example, in patients with chronic pelvic pain. Another predisposing factor is pending litigation, as in the case of whiplash injury. Here the personal meaning of a symptom can mitigate or increase the pain's intensity and the severity of the incapacitation.

By contrast, sometimes refractory pain with a presumed strong psychogenic component turns out to be a treatable neurological or orthopedic condition that has been missed by the referring clinician. In one report from a multidisciplinary pain clinic keeping tally on these matters for 120 patients with chronic pain, the overall rate of inaccurate or incomplete diagnosis at referral was 40 percent. Commonly missed diagnoses were myofascial disease, facet disease, peripheral nerve entrapment, temperomandibular joint disease, thoracic outlet syndrome, and herniated discs. After comprehensive workup, an organic origin for the pain was found in 98 percent of these patients.

Sites of Pain Typical sites and types of pain involved in pain disorder are headache, atypical facial pain, low back pain, and chronic pelvic pain; however, any site is a potential target. Pain can emanate from almost any point source in the body or it can be a conversion symptom involving just pain, using a previous pain pattern from organic disease as a conversion pain model. If the pain is used as a metaphor for psychic turmoil converted into a somatic form, it may bring to the afflicted person primary and secondary gains that reinforce the initiation, maintenance, and even exacerbation of pain symptoms for psychological purposes. When pain is a conversion, its distribution may not conform to known anatomical patterns. From a diagnostic perspective, pain cannot be diagnosed as a conversion disorder according to DSM-IV.

Mr. L., a 72-year-old married, Ukrainian-born, pious, wealthy retailer and father of a large family from an east coast city was admitted to the orthopedic service of a general hospital for evaluation of unbearable pain in the arches of his feet. He had fled his native country following a pogrom when he was 9. During the year of flight he had endured enormous physical hardships, starvation, and beatings until the surviving family members finally were able to emigrate to the United States. With incessant hard work, he had prospered economically; married a patient, supportive wife; and witnessed his six children develop promising careers. He became the major contributor to his temple and gave unstintingly to local charities for the needy and unfortunate. He had little time for personal enjoyment. Over the years each time he and his wife had time alone together and she had been affectionate with him, he would develop some excruciating bodily pain shortly thereafter: blinding headache, severe back spasm, abdominal pain, facial pain, or pelvic pain. These pains usually receded several days after the weekend was over or the trip was completed. Some pains occurred more frequently than others. He sought medical attention rarely except for these pains, which occurred every few months. His mood varied from glum to gloomy but he denied that he was depressed. He often claimed to have been blessed with good fortune. He led a temperate life, drank little, and had relatively good health between these episodes of pain.

Over the four decades that his physicians cared for him, they had become frustrated with this unassuming and humble man; his ardent complaints of pain were always so nonspecific and fluctuating in nature that they could not describe adequately for themselves any pathophysiological mechanisms to account for his pain. Their diagnostic tests were not revealing and Mr. L. usually refused their offers of narcotic or other analgesic relief. Laboratory workup for pains in the arches was noncontributory, and he was discharged when his symptoms cleared in 3 days. Four months later he was readmitted to the surgical service of the general hospital with severe, unrelenting, left-side upper abdominal pain. This time Mr. L. described this new pain in meticulous detail. A brief workup revealed very advanced carcinoma of the tail of the pancreas. He took the news from his physician stoically and asked to be discharged home that day.

Differential Diagnosis All pain is subjective and only obvious if the patient tells the physician about it. Acute pain tends to distort the facies in a grimace, causing muscle tightening and elevation of autonomic measures, such as pulse and blood pressure. Chronic pain, however, produces none of these signs. Clinicians trying to explore the extent of psychiatric involvement have used a number of clues, the validity of which still remains to be proved (Table 16-15). Those are the underlying dynamics and the presence or absence of conflicts and stressors are open to multiple interpretations; only careful long-term follow-up studies of large series of patients will help increase validity.

Antecedent history of poor premorbid adjustment: alcohol or drug abuse, sexual difficulties, multiple marriage, inability to hold a job
Temporal relationship between an environmental stimulus, relevant intrapsychic conflict, and pain
Utility of pain in either gaining compensation or avoiding situation deemed noxious by the patient
Lack of variation in the amount of pain from distraction, suggestion, or fear

Table 16-15 Clues to Underlying Psychopathology in Pain Disorder

Patients deal with their chronic illnesses in ways characteristic of their dynamics and their own psychopathology. Those with a histrionic nature exaggerate; those who use a history of ill luck to manipulate their environment make the most of their pain for their own ends; those with a need for punishment for real or imagined sins may fulfill their needs for atonement.

Both acute and chronic pain tend to be very field dependent. For example, soldiers who have survived fierce fighting on the battlefield typically need far less morphine for relief than their civilian counterparts with the same degree of tissue injury. Absolute pronouncements about no somatic pathology or no emotional components to the clinical picture must be viewed with skepticism. A positive response to a trial of placebo medication does not indicate anything about the organic nature of the patient's pain or the extent of the underlying psychopathology. The placebo response is positive in about one third of whatever population is tested. Pain associated with both melancholia and metastases can respond to placebo. Patients and their clinical syndromes need to be treated without too much loss of time deliberating about whether the pain caused the depression or the depression caused the pain. Individuals with personality disorders and chronic pain tend to have a more difficult time coping with any type of difficulty, both in the hospital and at home, and they are a challenge to their physicians. Fortunately, intensive inpatient treatment of refractory chronic pain has had some success.

Depressive Disorder Of the psychiatric diseases, depression presents most often as a pain disorder. At times the individual will have had previous depressive episodes with an altogether different clinical presentation and with no associated pain. Sometimes the best clue to the presence of a mood disorder is just the family history of severe depression, with or without pain being part of the clinical picture. There is no specific type of pain picture to alert the clinician to the presence of depression. Moreover, the sadness accompanying the pain disorder can almost always be ascribed to the pain.

Supporting the diagnosis of a mood disorder are the vegetative signs of depressive disease, especially sleep and libidinal disturbances. Cognitive signs of an accompanying depression include apathy, decreased interest in work, suicidal ideation, and a preoccupation with death. Collaborative sources of information are often needed to fully understand the patient's level of functioning.

Sometimes the best way to make the diagnosis of depressive disorder in a patient presenting with pain is simply to initiate a treatment trial of an antidepressant medication in addition to nonspecific modalities. Dramatic clinical relief of both the pain and the vegetative signs of depression provides the answer to the clinical question. Suicide in chronic pain patients must always be kept in mind, especially during the first 10 days of treatment with an antidepressant when the patient begins to look better and has more energy; such a state often occurs before the feeling of despair has lifted.

Conversion Disorder When a conversion mechanism is involved in the pain disorder, the clinician determines that (1) details of the pain picture have a symbolic or other specific meaning for the patient; (2) there is a conversion model, often a parent or other close family member; and (3) the pain follows no known anatomical pathway. By convention, pain involving a conversion mechanism is diagnosed as a pain disorder, not as a conversion disorder.

Treatment for this is similar to the psychotherapy of a patient with a conversion disorder. In addition, the clinician will use the same nonspecific treatment meted out to many or all pain disorder patients: physiotherapy, water pool massage, infrared heat, nonsteroidal anti-inflammatory agents, and other such modalities.

Psychosis When a psychosis associated with schizophrenia, medical disorder, or a dementing illness presents with pain, the pain itself is delusional and the pain pattern is atypical in quality and in distribution. Treatment must relieve the underlying psychiatric disorder. Delusional beliefs, however, may prevent the patient from cooperating with the treatment plan.

Malingering When someone is purposely lying about pain, the pain is hard to detect. Such individuals are not interested in the same outcome as the physician, which is relief of pain. Sometimes old records betray that the patient has tried to malinger in other settings and in other ways. Unlike the patient with Munchausen syndrome (facticious disorder with predominantly physical signs and symptoms), who is interested in staying in the sick role, the malingerer is consciously interested in some other benefit. Some clinicians have found that Munchausen syndrome patients will willingly agree to painful procedures whereas the malingerer will not.

Emotional Overlay Determining the extent of the emotional component of chronic pain is helpful, and care plans could then be effected, along with appropriate expectations and appropriate treatment. The well-known MADISON scale (Table 16-16) has not been rigorously validated but has been of considerable help to those working in the field. Another instrument that has proven to be clinically valuable in assessing whether a patient has an organic lesion that potentially amenable to a corrective surgical approach is the Mensana Pain Test, a 10-minute structured medical interview.

M =	**Multiplicity:** Pain is either in more than one place or of more than one variety; when treated, may recur elsewhere.
A =	**Authenticity:** More interested in clinician's acceptance of pain as genuine than in a cure.
D =	**Denial:** Especially exaggerated marital or family harmony; when admitting depression or anxiety, no impact on pain is admitted.
I =	**Interpersonal relationship:** Although the connection to the presence of any particular person's company as worsening the pain may be denied, observation of the patient's nonverbal and interactive behavior indicates otherwise.
S =	**Singularity:** When the pain is described as unlike that of anyone else, ever.
O =	**"Only you":** When the patient immediately idealizes the physician as savior, despite numerous failures by other competent experts.
N =	**Nothing helps, or no change:** When there is no relief whatsoever from any type of intervention, although all are tried (including narcotics) and there is no hour-to-hour or day-to-day fluctuation under a variety of circumstances.

Tabby adapted from Hackett TP, Cassem NH: *Massachusetts General Handbook of General Hospital Psychiatry*., St. Louis, 1978.

Table 16-16 MADISON Scale for Markers of Considerable Emotional Overlay

Course and Prognosis Once diagnosis is complete, treatment on an outpatient basis can be carried out by concerned physicians who will see the patient on a regular basis, be interested in the patient's complaints, and assure the patient that treatment will continue if there is some improvement. Pain disorder patients with pending litigation often attain no, or only minimal, relief until after the legal proceedings are finished.

Pain-Prone Individuals Persons who are painprone often experience pain in response to losses or other stresses, presumably as a way of dealing with unresolved guilt. Outcomes from treating the pain-prone are typically described as transient or poor. Techniques that are most promising are supportive therapy, a modicum of insight, and, when indicated, antidepressant medication.

Treatment Many patients, especially those with depression, can be considerably helped. The key to dealing with the pain disorder patient is accurate diagnosis and focus on the patient's level of functioning. The role of the family in the pain patient's life must be determined and efforts must be made to intervene in the dysfunctional components of family life. The patient's pain and predicament are always acknowledged at each visit, but the major focus is in moving on and regaining function. Pain clinics and inpatient pain treatment units both have had their share of success.

Treatment of the patient with any of the pain disorder subtypes needs to be multidisciplinary and multidimensional. Controlled trials seem to indicate that group therapy with antidepressant medication seems to have an advantage over individual therapy with antidepressant agents.

Treatment programs need to be culturally sensitive in terms of intake procedures and treatment planning. Treatment teams need to be aware of the potential effect of ethnic background on the communication, concerns, and coping styles related to the chronic pain experience of these patients.

Treatment team planning is important from the onset, coordinated as needed on a periodic basis. Multiple regimens often help in reducing pain intensity. Tailored to the patient, options include nonsteroidal anti-inflammatory agents, tricyclic medications, selective serotonin reuptake inhibitors, nerve blocks, localized electrical stimulation, visual imaging, relaxation, physical therapy, hypnosis, counseling, cognitive behavioral therapy, and supportive psychotherapy. Currently the choice of which form of psychotherapy to use is based on empiric grounds and personal preference because no case-matched studies have yet been performed. Research data is clear, however, that success in treatment result in a decrease in the patient's catastrophizing and an increase in the patient's perceived control over the pain.

If opioids have been used for chronic refractory pain, they should be tapered and discontinued because obviously they were not helping and may have been interfering with the patient's progress. Abuse of anti-anxiety medications may also be present in patients with chronic pain and must be dealt with. Antidepressant medications often play an important role in decreasing pain, even in the absence of a frank depressive syndrome. The possibility of a chronic sleep disorder must be investigated and appropriately treated, if present.

Price of Chronic Pain The clinician needs to be aware that any type of chronic pain can alter the patient's personality and family dynamics. Pain can also alter the way the patient relates to the environment and perhaps also the patient's internal locus of control style. At the onset of treatment the clinician needs to set guidelines for continuing and discontinuing treatment. No improvement in any way after 6 months is grounds for termination of treatment. If improvement is shown, the patient will need continued support. Often the patient's objective improvement far exceeds the subjective improvement.

BODY DYSMORPHIC DISORDER

Definition Body dysmorphic disorder focuses on the patient's feelings of dislike or even loathing for some aspect of the body's appearance. Despite the starkness of the complaint, few empirical data about the condition have been gathered although clinical investigations into this disorder are greatly increasing.

Patients with body dysmorphic disorder have a pervasive subjective feeling of ugliness of some aspect of their appearance despite a normal or nearly normal appearance. The core of the disorder is the person's strong belief or fear that he or she is unattractive or even repulsive. This fear is rarely assuaged by reassurance or compliments even though the typical patient with the disorder is quite normal in appearance. Although a minority of patients with the disorder do have a minor defect, the patient's concern is disproportionate to the degree of defect. So great is the individual's preoccupation that there is significant impairment in social or occupational functioning, or there is marked personal distress. All other problems are illogically attributed to the perceived cosmetic defect, accompanied by the unrealistic expectation that surgery will correct the defect and the patient's deficient life. Finally, the preoccupation is not better accounted for by any other mental disorder.

History Body dysmorphic disorder has been described in the European, Japanese, and Russian psychiatric literature for almost a century with various names. Emil Kraepelin considered it to be a compulsive neurosis. Pierre Janet called it an obsession of shame of the body, and Freud's famous Wolf-Man case was obsessively concerned about the size of his nose. In the United States this disorder was initially classified as an atypical somatoform disorder, labeled first as *dysmorphophobia*; however, until the late 1980s it has been little used and little studied in the United States.

Comparative Nosology In the United States body dysmorphic disorder first appeared in DSM-III with the term “dysmorphophobia” in the residual category of atypical somatoform disorder. Because dysmorphophobia inaccurately implied the presence of a behavioral pattern of prominent avoidance of the body, the term “body dysmorphic disorder” was introduced in DSM-III-R as a nondelusional somatoform disorder of undue preoccupation with imagined defects of appearance. Clinicians have wrestled over whether a patient's strong conviction about a feature of appearance was or was not of delusional intensity (i.e., whether the patient had a somatoform disorder or a delusional disorder with somatoform features). DSM-IV resolves that issue by not requiring that the conviction be defined in intensity as a criteria for inclusion. Delusional body dysmorphic disorder can be double coded as both a delusional disorder and as body dysmorphic disorder—a compromise that underscores the uncertainty about whether they are the same or different disorders. The delusional form of the illness may be a more severe type of the disorder. It was first incorporated into ICD-9 with the somatoform disorders and is grouped with them in ICD-10.

Epidemiology Accurate prevalence data for this disorder in the general population are not available. From the clinician's perspective, the full-blown syndrome is uncommon. Concern over appearance is culturally determined and dissatisfaction of a transient nature is particularly high in adolescents. However, to justify the diagnosis the individual must have clinically significant distress or impairment, which is not often found in teenagers.

The largest series to date, with 188 cases, had an equal balance of men and women. The average age of patients first diagnosed with body dysmorphic disorder is 30 years, and a high percentage have never married and are unemployed. Patients with the disorder first develop symptoms in adolescence or young adulthood. Typically they come from middle-class families and constitute a small, underrecognized and distinct group of patients in the general population. Only about 2 percent of those attending a university hospital plastic surgery clinic meet criteria for the disorder.

Body dysmorphic disorder is not uncommon as a comorbid condition in patients with major depressive disorder (current rate of 60 percent; lifetime rate of 80 percent), obsessive-compulsive disorder, and social phobia. Indeed in one study of 30 patients, all met DSM-III-R criteria for at least one other psychiatric diagnosis at some point in their lives, and usually concurrently.

Etiology The cause of body dysmorphic disorder is unknown. The high comorbidity with depressive disorders, a higher-than-expected family history of mood disorders and obsessive-compulsive disorder, and the reported responsiveness to SSRIs indicates that, in at least some patients, the pathophysiology of the disorder may involve serotonin and may be related to other disorders. There may be significant cultural or social effects on patients with body dysmorphic disorder because of the emphasis on stereotyped concepts of beauty that may be emphasized in certain families and within the culture at large. In psychodynamic models body dysmorphic disorder is seen as reflecting the displacement of a sexual or an emotional conflict onto a nonrelated body part; such a putative association occurs through the defense mechanisms of repression, dissociation, distortion, symbolization, and projection.

Diagnosis and Clinical Features There are several likely places to find patients with this uncommon disorder: in a mood disorders clinic, in a plastic surgery clinic, and in a dermatology clinic. A patient with a body dysmorphic disorder might request rhinoplasty; removal of facial sags, jowls, wrinkles, or puffiness; or breast reduction or augmentation. In a clinic for refractory and recurrent depression a patient with the disorder might request relief from a depression caused by an imagined but burdensome bodily defect, but because of embarrassment or shame might not reveal the self-loathing of a body part. In a dermatology clinic, the concern could be acne, scarring, or blemishes. It is a secret disorder, often not admitted to, yet it is often a time-consuming one that engages the patient for many hours during the day with activities such as excessive checking of the defect in mirrors or other reflective surfaces, skin picking, hair combing, or frequently asking others for reassurance. Frequent accompanying symptoms, presumably related to suffering from discomfort about appearance in social situations, are insomnia, depression, and anxiety. Social isolation, dissatisfaction with relationships, shame, and low self-esteem are also common.

Facial flaws are the most common defect in body dysmorphic disorder. Other body parts that sometimes become a focus include hair, breasts, and genitalia. Commonly associated with the distorted belief about appearance is an unrealistic concept of how much one's image can be improved by surgical intervention. Some clinicians report that these patients, for all their preoccupation with their defect, still are vague or inconsistent in the description of it.

Family histories of substance abuse and mood disorder are common in reported cases. Also predisposing to the disorder may be certain types of personality characteristics, especially a mixture of obsessional and avoidant traits, but no single personality pattern predominates. Reportedly the patients are shy, self-absorbed, and overly sensitive to their imagined defect as a focus of notice or criticism. As adolescents, some reports hold that they had few friends and seldom dated. There may be some ideas of reference related to the body dysmorphic disorder, but otherwise there is no formal thought disorder. The major abnormality on mental status examination is the lack of insight into the nature of the problem. Table 16-17 lists the DSM-IV criteria for body dysmorphic disorder.

A. Preoccupation with an imagined defect in appearance. If a slight physical anomaly is present, the person's concern is markedly excessive.
B. The preoccupation causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The preoccupation is not better accounted for by another mental disorder (e.g., dissatisfaction with body shape and size in anorexia nervosa).

Table 16-17 DSM-IV Diagnostic Criteria for Body Dysmorphic Disorder

Differential Diagnosis Body dysmorphic disorder patients have an overvalued idea about their defective appearance. A number of other psychiatric disorders are also accompanied by odd or unusual ideas about the body. In general such patients are best accounted for diagnostically by considering other major psychiatric disorders as primary, and adding that body dysmorphic symptoms are also present. For example, especially in the early phases of schizophrenia some patients may have somatic delusions for which they seek correction by cosmetic surgery. In such cases there is an absolute conviction of ugliness as well as other bizarre delusions and perhaps even hallucinations.

Other disorders with body dysmorphic symptoms include mood disorder, narcissistic personality disorder, and anorexia nervosa; in such instances, both diagnoses should be made. Melancholically depressed patients often have additional delusions about some aspect of their body's inadequacy or defect. The presence of a mood disorder, if suspected, can be further elucidated through the taking of the appropriate medical history. Patients with narcissistic personality disorder have a continual interest in their appearance and a long history of interpersonal difficulties that overshadows the body image problem. Constant preoccupation with feared obesity and inadequacy in an obviously malnourished person may be best classified as anorexia nervosa.

A certain proportion of patients with obsessive-compulsive disorder also have body dysmorphic disorder, in which case both disorders should be diagnosed. If the patient has obsessions about appearance and associated compulsive behaviors (such as mirror-checking), the diagnosis of obsessive-compulsive disorder is not made if the obsessions only concern aspects of appearance.

Isolated somatic delusions unrelated to appearance and accompanied by absolute certainty without other evidence of thought disorder are best classified with the delusional disorders. The belief that one's gender is not rightly assigned (according to the appearance of the external genitalia at birth) is better accounted for by a diagnosis of gender identity disorder. Opinions still vary as to whether monosymptomatic hypochondriacal psychosis (a delusion about a disease or disfigurement in a body part) and body dysmorphic disorder are one or two disorders. DSM-IV allows for double coding if it seems relevant to the clinician.

Course and Prognosis Onset of the disorder may be gradual, during childhood, adolescence, or in the 20s. Discontentment may build for several years before the person considers some type of definitive surgical correction; however, surgical treatment produces no relief. Several long-term follow-up series of patients who have undergone cosmetic surgery have indicated the frequent later emergence of even more severe psychopathology.

Treatment Neither surgical nor psychotherapeutic intervention has had any significant long-term impact on decreasing the preoccupation with defective bodily appearance in most patients with body dysmorphic disorder. There are case reports that psychotherapy alone has helped some patients with this disorder: presumably those helped most by this treatment modality are more emotionally intact, but current thinking in the field is that psychotherapy and medication are both indicated.

Case reports note positive effects from tricyclic drugs, monoamine oxidase inhibitors (MAOIs), and pimozide (Orap), a dopamine-receptor antagonist. Recent preliminary data now support the efficacy of SSRIs and cognitive-behavioral therapy. Moreover, the efficacy of SSRIs appears to be greater than that of tricyclic drugs, which in turn are more effective than the MAOIs. Augmentation of SSRIs with buspirone (BuSpar) and antipsychotic agents, as well as the combination of SSRIs also appears to be promising. However, to date there have been no controlled pharmacotherapy trials.

Some patients refractory to the tricyclics and MAOIs have recently been shown in open trials to respond to antidepressant and antiobsessional agents with potent serotonin receptor blockade, such as fluoxetine (Prozac) and clomipramine (Anafranil).

After maximal treatment response, some patients with body dysmorphic disorder may still retain remnants of their bodily preoccupation, but symptom intensity is often diminished enough to allow patients to resume full social and personal lives. However, relapse when patients are taken off medication is common.

SOMATOFORM DISORDER NOT OTHERWISE SPECIFIED

This is the second of the residual categories for the somatoform disorders. Like the first residual category, undifferentiated somatoform disorder, this category was created in order to classify certain somatoform disorder patients whose symptoms and associated disability do not fit the full criteria for other somatoform disorders. The diagnostic criteria for somatoform disorder not otherwise specified are presented in Table 16-18. Patients with this disorder have clinically significant distress or

disabilities in social, occupational, or other important areas of functioning.

This category includes disorders with somatoform symptoms that do not meet the criteria for any specific somatoform disorder. Examples include:

1. pseudocyesis: a false belief of being pregnant that is associated with objective signs of pregnancy, which may include abdominal enlargement (although the umbilicus does not become everted), reduced menstrual flow, amenorrhea, subjective sensation of fetal movement, nausea, breast engorgement and secretions, and labor pains at the expected date of delivery. Endocrine changes may be present but the syndrome cannot be explained by a general medical condition that causes endocrine changes (e.g., hormone-secreting tumor).
2. a disorder involving nonpsychotic hypochondriacal symptoms of less than 6 months' duration.
3. a disorder involving unexplained physical complains (e.g., fatigue or body weakness) of less than 6 months' duration that are not due to another mental disorder.

Table 16-18 DSM-IV Diagnostic Criteria for Somatoform Disorder Not Otherwise Specified

NEURASTHENIA

Neurasthenia, literally a lack of nerve energy, was described by George M. Beard at the end of the nineteenth century to account for the physical and mental exhaustion arising from the depletion of nervous resources. This finding has given rise to numerous theories but little controlled research. The diagnosis is not included in DSM-IV, nor was it included in DSM-III or DSM-III-R. In DSM-I the condition was called *psychologic nervous system reaction* and in DSM-II it was called *neurasthenic neurosis.*

Neurasthenia is still diagnosed with some frequency in Asia and Russia but not the United States. Particularly in Asia, researchers have found neurasthenia to be a culturally sanctioned idiom of distress. In ICD-10, two main types of neurasthenia occur, with considerable overlap. One type has a predominance of increased fatigue after mental effort, thus associated with a decrease in job performance or coping with activities of daily living. Difficulty in concentration, unpleasant distracting associations or recollections, and generally inefficient thinking have all been reported. The other type has physical weakness and exhaustion after only a minimal physical effort, accompanied by muscular aches and pains and an inability to relax. Both types also have a variety of unpleasant physical feelings including dizziness, tension headaches, and feelings of general instability. Patients also worry about decreased mental or physical well-being and have irritability and anhedonia. Varying degrees of minor anxiety and depression are also common. Sleep may be disturbed with initial or middle-phase insomnia or hypersomnia.

According to ICD-10, the condition must be present for more than 3 months and have either (1) persistent and distressing complaints of feelings of exhaustion after minor mental effort or (2) persistent and distressing complaints of feelings of fatigue and bodily weakness after minor physical effort (Table 16-19). Moreover, the patient is unable to recover from the symptoms of exhaustion or fatigue by rest, relaxation, or entertainment.

A. Either of the following must be present:
   (1) persistent and distressing complaints of feelings of exhaustion after a minor mental effort (such as performing or attempting to perform everyday tasks that do not require unusual mental effort);
   (2) persistent and distressing complaints of feelings of fatigue and bodily weakness after minor physical effort;
B. At least one of the following symptoms must be present:
   (1) feelings of muscular aches and pains
   (2) dizziness;
   (3) tension headaches;
   (4) sleep disturbances;
   (5) inability to relax;
   (6) irritability.
C. The patient is unable to recover from the symptoms in criterion A (1) or (2) by means of rest, relaxation, or entertainment.
D. The duration of the disorder is at least 3 months.
E. *Most commonly used exclusion clause.* The disorder does not occur in the presence of organic emotionally labile disorder, postencephalitic syndrome, postconcussional syndrome, mood disorders, panic disorder, or generalized anxiety disorder.

Table 16-19 ICD-10 Diagnostic Criteria for Neurasthenia

Prevalence studies indicate that, using ICD-10 criteria, about 1 percent of the population have had this condition for more than 3 months and 10 percent have it for more than a month during the past 10 years. An epidemiological study of Chinese-Americans reported that 6.4 percent of the subjects had neurasthenia, and only 3.6 percent of those had no other current or lifetime psychiatric diagnoses.

The cause of neurasthenia seems to be pleomorphic, with proponents implicating mood disorders, anxiety disorders, or conversion disorders. Others propose viral sequelae while still other investigators search for trace mineral deficiencies. Obscure neuromuscular conditions, cryptic immune dysfunction, and chronic fatigue syndrome have also been implicated in the cause of neurasthenia. Despite the broad range of unproven theories, individuals can be considerably incapacitated with symptoms of fatigue and fatigability, with or without accompanying anxiety. Diagnostic workup of the fatigue syndrome by definition is unrewarding because a medical condition causing fatigue would by definition exclude the diagnosis of neurasthenia.

There is no best treatment for neurasthenia. Randomized controlled trials have not been completed. In clinical practice, trials on antidepressant agents such as the activating selective serotonin reuptake inhibitors and psychostimulants such as methylphenidate (Ritalin) or amphetamines have been tried. It may be that sleep, tincture of time, and empathic support are the best treatment modalities, but this has not been proven.

Mrs. B. is a 45-year-old married accountant who had an especially promising career with a large national firm for 5 years until, 10 years ago she developed the gradual onset of neurasthenic symptoms of fatigue, diffuse muscle aches, and fitful sleeping following a series of major psychological stressors including a secret extramarital affair. Her primary care physician prescribed nonspecific treatments including nonsteroidal anti-inflammatory agents and rest, all to little avail.

Two years after the onset of her neurasthenia, she developed an acute depressive episode with multiple somatic complaints. She was referred to a psychiatrist who also noted that her mother had multiple sclerosis, requiring considerable physical care from all her children as they grew up.

During treatment for this major depressive episode, Mrs. B.'s depression and neurasthenia improved gradually, but she experienced multiple side effects from fluoxetine, cyclobenzaporine (Flexeril), and even the low-impact swimming program. In addition to medication, she was treated with behavioral therapy that focused on diminishing her catastrophizing about her bodily sensations. In supportive psychotherapy she benefitted from working through both some of her insoluble family conflicts and her grief about letting go of her high career aspirations.

Her energy gradually increased toward normal. Her sleep pattern of multiple interruptions and early-morning awakening improved considerably. Her overall outlook on a life of diminished expectations ameliorated markedly.

Now 10 years since the onset of her neurasthenia, she still has mini-relapses of persistent and vexing exhaustion; when others see her faring better, they make demands on her for emotional or physical support that she can not directly refuse. She has not returned to gainful employment.

CHRONIC FATIGUE SYNDROME

Fatigue is one of the most common symptoms in all of medical practice. The nature of chronic fatigue syndrome, however, remains very controversial. This syndrome has been defined by the Centers for Disease Control (CDC) in 1988 as a disabling disorder with a combination of a certain number of nonspecific symptoms such as fluctuating levels of fatigue, various combinations of neuromuscular and neuropsychological symptoms, chronic pain, malaise, mild fevers, and anxiety. According to the latest CDC criteria, it is a condition that has been clinically evaluated and still remains as an unexplained, persistent, or relapsing chronic fatigue that is of new or definite onset in a previously healthy person; is not the result of exertion; is not substantially alleviated by rest; and results in substantial reduction in previous levels of occupational, educational, social, or personal activities (Table 16-20).

A. Severe unexplained fatigue for over 6 months that is:
(1) of a new or definite onset
(2) not due to continuing exertion
(3) not resolved by rest
(4) functionally impairing
B. The presence of four or more of the following new symptoms:
(1) impaired memory or concentration
(2) sore throat
(3) tender lymph nodes
(4) muscle pain
(5) pain in several joints
(6) new pattern of headaches
(7) unrefreshing sleep
(8) postexertional malaise lasting more than 24 hours

Table 16-20 1994 CDC Criteria for Chronic Fatigue Syndrome

Since the time of the original CDC definition there have been a few minor revisions in criteria. The debate continues on whether chronic fatigue syndrome is a psychiatric illness or a somatic one with emotional components.

The prevalence of chronic fatigue syndrome is yet to be determined. In one recent primary care clinic study, of 686 patients specifically examined for fatigue, 77 patients were identified as having chronic fatigue and 17 cases met criteria for chronic fatigue syndrome. To date, there has been no community-based epidemiological investigation of this condition, only studies from primary-care practices. Hence there is no data of this condition that is not biased by help-seeking and access to health care.

In the absence of a clear cause, chronic fatigue syndrome has been disparagingly called “the yuppie flu,” neurasthenia, and masked depression. Viral or postviral etiologies have also been considered, with diagnoses such as myalgic encephalomyelitis and chronic Epstein-Barr virus disorder being in vogue for a while. Studies of these viral etiologies have not seemed to be relevant to a large segment of those with chronic fatigue syndrome, but recent studies note a persistent enterovirus or herpes virus 6 in some cases. An overlap with fibrositis or fibromyalgia has also been considered, but this is not particularly illuminating because it only links a poorly understood rheumatic condition with another even less clearly defined, fatigue-defined condition of unclear etiology.

Medical anthropologists have seen chronic fatigue syndrome as a vehicle for negotiation of change in interpersonal worlds. A physiologist recently suggested that there might be a nasal fatigue reflex, akin to the atavistic but well-documented diving reflex. According to this hypothesis, the nasal fatigue reflex could produce the debilitating fatigue that would in turn give the afflicted individual the time to heal before having to face a hostile environment. Brain magnetic resonance imaging (MRI) studies of 43 patients with chronic fatigue syndrome, compared to controls, demonstrated that no MRI pattern of white matter abnormalities is specific.

Psychiatric factors have been strongly associated in the etiology of chronic fatigue syndrome. For example, certain clinical samples have reported that almost half of their cases have had antecedent psychiatric disorders such as depression, phobias, or other anxiety disorders. In a recent matched study of 214 subjects with chronic fatigue syndrome from a nonspecialist, nonreferral setting, most of the index subjects were at considerably greater risk of current psychiatric disorder than were control subjects. The likelihood of psychiatric disorder was six times greater in these chronic fatigue syndrome patients than in the matched controls when evaluated either by interview or by questionnaire. Other studies of subjects who have undergone neuropsychological testing indicate that at least a subset of patients with chronic fatigue syndrome experience significant impairments in learning and memory. However, a primary psychiatric etiology for chronic fatigue syndrome is typically stoutly denied by patients with this syndrome, especially by those who are members of the chronic fatigue syndrome national peer support groups.

Chronic fatigue syndrome is most likely to be a heterogenous condition, with fatigue being only a final common pathway. At this phase of research on the condition, it is not possible to speak with certainty about the etiology of this complicated condition.

Whether chronic fatigue syndrome should be considered as a special class of mood disorder with somatic symptoms (specifically fatigue), or a somatoform disorder not otherwise specified, or a combination of a psychiatric disorder with an unidentified infectious agent, or even some combination of these conditions, remains to be clarified. Using strict DSM-III-R criteria, a recent study demonstrates that very few chronic fatigue syndrome patients have somatization disorder.

Treatments for a presumably heterogenous condition with an unknown cause typically should involve a multidisciplinary approach involving psychological, physiological, and social factors. Possible concomitant psychiatric disorders could most likely benefit from a psychopharmacological trial. Many other types of treatments are being used for this debilitating illness, with even electric plum blossom needle therapy having its adherents in the literature. Two randomized controlled trials of cognitive-behavioral therapy (compared with relaxation therapy or routine practitioner care) indicate that cognitive behavioral therapy has substantial impact on the disability and symptoms of patients with this disorder.

Miss J. was a 35-year-old single white librarian with a benign medical past and no psychiatric symptoms prior to developing a flu-like illness. After 10 days the acute episode passed, but she continued to feel lethargic and readily fatigued. Two weeks after the onset of this illness, she returned to work but was unable to complete her usual 8-hour days because of increasing exhaustion and a newly developed set of symptoms: gradually evolving, diffuse muscle and joint pain.

Her primary care physician suggested naproxen (Naprosyn) and gave her encouragement while counseling patience. The physician noted nothing unusual about her mood, and prescribed hypnotic agents to improve her sleep. There was no improvement, however, on 10 mg of zolpidem (Ambien). She then started having squeezing bitemporal headaches. After 3 months, she was referred to a rheumatologist who tried to start her on amitriptyline (Elavil) 50 mg at night. The patient protested vehemently, denying that she was depressed, just in pain.

Previously she had been a conscientious employee and had rarely taken leave or missed work because of illness. After 3 months of this illness, however, she was forced to take a leave of absence, returning to live with her mother since she no longer had any income. She continued to “hurt all over,” was lethargic and irritable, and slept poorly because of pain. When she slept, she reported that she no longer awoke refreshed.

Six months after the onset of her original symptoms, she self-referred to an academic health center's rheumatology clinic where she presented as an afebrile and otherwise healthy woman who was angry about her protracted illness and her living situation. She admitted to difficulty with concentration. Joint examination revealed full range of motion with no red, hot, or swollen joints; tender points were present at all 18 sites.

Her rheumatologist prescribed amitriptyline 25 mg at night for 4 days and then told to increase the dose by one tablet until she achieved better sleep or reached a dosage of 150 mg. Still protesting that she was not depressed, she took the antidepressant medication because she was desperate for relief. A month later she returned to the rheumatology clinic, still hostile and impatient, with little change, and she was then prescribed 20 mg of fluoxetine in the morning in addition to the amitriptyline at night.

Within a month of this regimen, she was somewhat improved in her mood, sleep, and joint symptoms. However, she still continues to have episodes of fatigue, usually related to stressful life events. She had not yet returned to the work force.

SUGGESTED CROSS-REFERENCES

Related disorders are discussed in Chapters 17 on factitious disorders and Chapter 18 on dissociative disorders. Section 25.12 on consultation-liaison psychiatry, Section 27.1 on noncompliance with treatment, and Section 27.2 on malingering also have information relevant to somatoform disorders. Chronic pain is presented in Section 28.6.

CHAPTER REFERENCES

Adler RH, Zlot S, Hurny C, Minder C: Engel's psychogenic pain and the pain-prone patient: A retrospective, controlled clinical study. Psychosom Med *51:*87, 1989.

Barsky AJ, Fama JM, Bailey ED, Ahern DK: A prospective 4 to 5 year study of DSM-III-R hypochondriasis. Arch Gen Psychiatry *55:*737, 1998.

Bass C, Murphy M: Somatoform and personality disorders: Syndromal comorbidity and overlapping developmental pathways. J Psychosom Res *39:*403, 1995.

*Bates MS, Rankin-Hill L, Sanchez-Ayendez M, Mendez-Bryan R: A cross-cultural comparison of adaptation to chronic pain among Anglo-Americans and native Puerto Ricans. Med Anthrop *16:*141, 1995.

Bendefeldt F, Miller LL, Ludwig AM: Cognitive performance in conversion hysteria. Arch Gen Psychiatry *33:*1250, 1976.

Buckwald D: Fibromyalgia and chronic fatigue syndrome: Similarities and differences. Rheum Dis Clin North Am *22:*219, 1996.

Buckwald D, Pearlman T, Kith P, Katon W, Schmaling K: Screening for psychiatric disorders in chronic fatigue and chronic fatigue syndrome. J Psychosom Res *42:*87, 1997.

Corbishley MA, Hendrickson R, Beutler LE, Engle D: Behavior, affect, and cognition among psychogenic pain patients in group expressive psychotherapy. J Pain Symptom Manage *5:*241, 1990.

Deale A, Chalder T, Marks I, Wessely S: Cognitive behavior therapy for chronic fatigue syndrome: A randomized controlled trial. Am J Psychiatry *154:*408, 1997.

Deluca J, Johnson SK, Ellis SP, Natelson BH: Sudden vs gradual onset of chronic fatigue syndrome differentiates individuals on cognitive and psychiatric measures. J Psychiatr Res *31:*83, 1997.

Drake ME Jr: Conversion hysteria and dominant hemisphere lesions. Psychosomatics *34:*524, 1993.

Escobar JI: Developing practical indexes of somatization for use in primary care. J Psychosom Res *42:*323, 1997.

Flor-Henry P, Fromm-Auch D, Tapper M, Schopflocher D: A neuropsychological study of the stable syndrome of hysteria. Biol Psychiatry *16:*601, 1981.

Flor-Henry P, Toomer R, Kumpula I, Koles Z, Yeudall I: Neurophysiological and neuropsychological study of two cases of multiple personality syndrome and comparison with chronic hysteria. Int J Psychophysiol *10:*151, 1990.

Ford CV: The somatizing disorders. Psychosomatics *27:*327, 1986.

Ford CV: Dimensions of somatization and hypochondriasis. Neurol Clin *13:*241, 1995.

Ford CV, Folks DG: Conversion disorders: An overview. Psychosomatics *26:*371, 1985.

Hackett TP: The pain patient: Evaluation and treatment. In *The Massachusetts General Hospital Handbook of General Hospital Psychiatry*, TP Hackett, NH Cassem, editors. Mosby, St Louis, 1978.

Harden CL: Pseudoseizures and dissociative disorders: A common mechanism involving traumatic experiences. Seizure *6:*151, 1997.

Hendler N, Bergson C, Morrison C: Overlooked physical diagnoses in chronic pain patients involved in litigation, part 2. The addition of MRI, nerve blocks, 3-D CT and quantitative flow meter. Psychosomatics *37:*509, 1996.

Hollander E, Neville D, Frenkel M, Josephson S, Liebowitz MR: Body dysmorphic disorder: Diagnostic issues and related disorders. Psychosomatics *33:*156, 1992.

James L, Gordon E, Kraiuhin C, Howson A, Meares R: Augmentation of auditory evoked potentials in somatization disorder. J Psychiatr Res *24:*155, 1990.

Jenkins PLG: Psychogenic abdominal pain. Gen Hosp Psychiatry *13:*27, 1991.

Kellner R: Hypochondriasis and somatization. JAMA *258:*2718, 1987.

Kirmayer LJ, Robbins JM, editors: Current concepts of somatization: Research and clinical perspectives. American Psychiatric Press, Washington, DC, 1991.

Kroenke K, Spitzer RL, deGruy FV, Swindle R: A symptom checklist to screen for somatoform disorders in primary care. Psychosomatics *39:*263, 1998.

*Lipowski ZJ: Somatization: The concept and its clinical application. Am J Psychiatry *145:*1358, 1988.

Livingston R, Witt A, Smith GR: Families who somatize. Dev Behav Pediatr *16:*42, 1995.

Mace CJ, Trimble MR: Ten-year prognosis of conversion disorder. Br J Psychiatry *169:*282, 1996.

Marshall JC, Halligan PW, Fink GR, Wade DT, Frackowiak RS: The functional anatomy of a hysterical paralysis. Cognition *64:*B1, 1977.

Min SK, Lee BO: Laterality in somatization. Psychosom Med *59:*236, 1997.

Noyles R Jr, Holt CS, Happel RL, Kathol RG, Yagla SJ: A family study of hypochondriasis. J Nerv Ment Dis *185:*223, 1997.

Newman NJ: Neuro-ophthalmology and psychiatry. Gen Hosp Psychiatry *15:*102, 1993.

*Noyes R: The relationship of hypochondriasis to anxiety disorders. Gen Hosp Psychiatry *21:*8, 1999.

*Phillips KA: Pharmacologic treatment of dysmorphic disorder. Psychopharmacol Bull *32:*597, 1996.

*Pilowsky I, Barrow CG: A controlled study of psychotherapy and amitriptyline used individually and in combination in the treatment of chronic intractable, 'psychogenic' pain. Pain *40:*3, 1990.

Rief W, Shaw R, Fichter MM: Elevated levels of psychophysiological arousal and cortisol in patients with somatization syndrome. Psychosom Med *60:*198, 1998.

Rost K, Kashner TM: Effectiveness of psychiatric intervention with somatization disorder patients: Improved outcomes at reduced costs. Gen Hosp Psychiatry *16:*381, 1994.

Sackeim HA: Lateral asymmetry in bodily response to hypnotic suggestions. Biol Psychiatry *17:*437, 1982.

Sharpe M, Bass C: Pathophysiological mechanisms in somatization. Intl Rev Psychiatry *4:*81, 1992.

Sharpe M, Chalder T, Palmer I, Wessely S: Chronic fatigue syndrome. A practical guide to assessment and management. Gen Hosp Psychiatry *19:*185, 1997.

Shorter E: The borderline between neurology and history: Conversion reactions. Neurol Clin *13:*229, 1995.

*Simon GE, Gureje O: Stability of somatization disorder and somatization symptoms among primary care patients. Arch Gen Psychiatry *56:*90, 1999.

Simon GE, Von Korff M: Somatization and psychiatric disorder in the NIMH epidemiologic catchment area study. Am J Psychiatry *148:*1494, 1991.

*Smith GR: *Somatization Disorder in the Medical Setting.* American Psychiatric Press, Washington, DC, 1991.

Smith GR, Monson RA, Ray DC: Psychiatric consultation in somatization disorder. New Engl J Med *314:*1407, 1986.

*Smith GR, Rost K, Kashner M: A trial of the effect of standardized psychiatric consultation on health outcomes and costs in somatizing patients. Arch Gen Psychiatry *52:*238, 1995.

*Stuart S, Noyes R: Attachment and interpersonal communication in somatization. Psychosomatics *40:*34, 1999.

Valdes M, Garcia L, Treserra J, Pablo J, Flores T: Psychogenic pain and depressive disorders: An empirical study. J Affect Disord *16:*21, 1989.

Von Korff M, Resche LL: First onset of common pain symptoms: A prospective study of depression as a risk factor. Pain *55:*251, 1993.

Wessely S, Chalder T, Hirsch S, Wallace P, Wright D: Psychological symptoms, somatic symptoms, and psychiatric disorder in chronic fatigue and chronic fatigue syndrome: A prospective study in a primary care setting. Am J Psychiatry *153:*1050, 1996.

Wexler BE: Cerebral laterality and psychiatry: A review of the literature. Am J Psychiatry *137:*279, 1980.

Wood BL, McDaniel S, Burchfiel K, Erba G: Factors distinguishing families of patients with psychogenic seizures from families of patients with epilepsy. Epilepsia *39:*432, 1998.

Wynick S, Hobson RP, Jones RB: Psychogenic disorders of vision in childhood ("visual conversion reactions"): Perspectives from adolescence: A research note. J Child Psychol Psychiatry *38:*375, 1997.

CHAPTER 17. FACTITIOUS DISORDERS

MARC D. FELDMAN, M.D., AND CHARLES V. FORD, M.D.

Factitious disorders are among the most challenging phenomena in clinical practice. They involve efforts to garner gratifications intrinsic to the sick role through the simulation, exaggeration, aggravation, or induction of physical or psychological signs and symptoms. Factitious disorders can be conceptualized to fall within a continuum of "abnormal illness behavior," in which individuals create or amplify medical signs or symptoms. Somatoform disorders, factitious disorders, and malingering represent various points on the continuum, with the respective goals traversing the spectrum from fully unconscious and nonvolitional to exclusively conscious and willful. Commenting upon the "purloined sick role" in *The Spectrum of Factitious Disorders*, Don R. Lipsitt writes,

> In all likelihood, factitious disorder is both disease and deception, presenting one of the most challenging (and potentially vexing) variants of psychopathology in medical experience. If we impute a conscious motive to the wish to be sick, we regard this as strange behavior indeed; only in the case of the true malingerer, whose behavior is quite transparently designed for some personal gain, does logic prevail. If we can detect no reason for such a desire, our preference is to assume that there must be strong propulsion from deeply unconscious and complex roots.
>
> Such behavior lies outside the parameters of the 'sick role' as defined by Talcott Parsons in 1951. In this role, sick individuals are excused from social responsibility, are expected to perceive their condition as undesirable, lack voluntary control over it and therefore are not considered at fault, and are expected to seek competent help in ameliorating or curing the condition . . . . The concept of 'abnormal illness behavior' . . . can be extended emphatically to the voluntary induction of illness, whether for primary or secondary gain.

Not surprisingly, clinicians confronted by factitious disorders often find themselves frustrated and bewildered by individuals who ostensibly present for thorough medical care but simultaneously conceal the cause of the malady, thus making definitive care impossible.

DEFINITION

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) contains a single set of criteria for factitious disorder, but delineates three subtypes: factitious disorder with predominantly psychological signs and symptoms, factitious disorder with predominantly physical signs and symptoms, and factitious disorder with combined psychological and physical signs and symptoms. The diagnostic criteria for factitious disorder and explanations of the subtypes appear in Table 17-1. The category of factitious disorder not otherwise specified includes disorders with factitious symptoms that do not meet the criteria for the other factitious disorders. The DSM-IV diagnostic criteria for factitious disorder not otherwise specified are outlined in Table 17-2.

A. Intentional production or feigning of physical or psychological signs or symptoms.
B. The motivation for the behavior is to assume the sick role.
C. External incentives for the behavior (such as economic gain, avoiding legal responsibility, or improving physical well-being, as in malingering) are absent.

Code based on type:

With predominantly psychological signs and symptoms: If psychological signs and symptoms predominate in the clinical presentation

With predominantly physical signs and symptoms: If physical signs and symptoms predominate in the clinical presentation

With combined psychological and physical signs and symptoms: If both psychological and physical signs and symptoms are present but neither predominates in the clinical presentation

Table 17-1 DSM-IV Diagnostic Criteria for Factitious Disorder

This category includes disorders with factitious symptoms that do not meet the criteria for factitious disorder. An example is factitious disorder by proxy: the intentional production or feigning of physical or psychological signs or symptoms in another person who is under the individual's care for the purpose of indirectly assuming the sick role.

Table 17-2 DSM-IV Diagnostic Criteria for Factitious Disorder Not Otherwise Specified

DSM-IV illustrates factitious disorder not otherwise specified with the example of factitious disorder by proxy. However, the DSM-IV Task Force determined that there was insufficient information about factitious disorder by proxy to warrant its inclusion as an official category, and so research criteria for this disorder are consigned to Appendix B, "Criteria Sets and Axes Provided for Further Study." Factitious disorder by proxy includes situations in which one person produces symptoms in another for the purpose of assuming the sick role indirectly (i.e., by proxy). In representative cases, a mother fabricates illness in her child, then presents the child for medical treatment while disclaiming knowledge of the origin of the problem. Risks to the health of these children stem from diagnostic maneuvers and treatments as well as actual illness induced by the parent. The research criteria for factitious disorder by proxy are presented in Table 17.3.

A. Intentional production or feigning of physical or psychological signs or symptoms in another person who is under the individual's care.
B. The motivation for the perpetrator's behavior is to assume the sick role by proxy.
C. External incentives for the behavior (such as economic gain) are absent.
D. The behavior is not better accounted for by another mental disorder.

Reprinted with permission from DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 17-3 DSM-IV Research Criteria for Factitious Disorder by Proxy

Patients with factitious disorders consciously perpetrate their deceptive behaviors although their underlying motivations may be unconscious. By convention, individuals who readily acknowledge that they have produced their own medical signs (e.g., patients who self-mutilate) are not diagnosed with factitious disorders.

HISTORY

Although reports in the professional and lay literature have flourished during the past several decades, the phenomenon of falsified physical or psychological ailments is hardly unique to modern medicine. The Bible (1 Samuel 21:11–15) describes King David's feigning of madness before Achis, the King of Geth. Facing Achis's servants,

> [David] changed his countenance before them, and slipt down between their hands. And he stumbled against the doors of the gate; and his spittle ran down upon his beard.
>
> And Achis said to his servants: You saw the man was mad. Why have you brought him to me?
>
> Have we need of madmen, that you have brought in this fellow to play the madman in my presence? Shall this fellow come into my house?

In the second century AD, Galen wrote a treatise on the subject, listing medical signs and symptoms that people had self-induced or simulated, including vomiting and rectal bleeding. In the first half of the nineteenth century, several treatises on feigned insanity appeared; late in this period, authors such as Isaac Ray disputed earlier writers who had maintained that falsified insanity was difficult to detect. In 1838 the British physician Hector Gavin published an essay, “On Feigned and Factitious Diseases,” on the array of medical and psychological ruses plied by soldiers and seamen. Gavin recognized that while most of these errant conscripts sought to procure discharge or evade unpleasant duties, others seemed driven by a less concrete goal: they wished simply “to excite compassion or interest,” a motivation that is often the basis for contemporary cases of factitious disorder. In his book *Hysteria* (1853), Dr. Robert Carter coined the term “tertiary hysteria” to refer to those who would bask in the “fuss and parade of illness” to elicit sympathy. Speculation about the psychodynamic underpinnings of factitious behavior was advanced in 1934 when Karl Menninger described “polysurgical addiction.” He postulated an intrapsychic struggle in these patients characterized by intense aggression both toward themselves and their physicians, the latter representing the “perceived sadistic parent.”

Perhaps the most influential article to date on the subject appeared in *Lancet* in 1951. In this work of whimsical medical literature, Richard Asher coined the term “Münchausen syndrome” to refer to a pattern of factitious disorder “which most doctors have seen, but about which little has been written.” Recounting several brief case reports, Asher distinguished Munchausen syndrome by the chronic use of fabricated symptoms to gain hospital admission, embellishment of the personal history, and migration from hospital to hospital. Asher named the syndrome after Baron Hieronymus Karl Friedrich Freiherr von Münchausen (1720–1791), a retired German cavalry officer and raconteur who had been ruthlessly parodied by R. E. Raspe in an immensely popular book first published in 1785 (Fig. 17.1). Asher advised physicians to be alert to the possibility of Munchausen syndrome if a patient (1) exhibited numerous surgical scars, usually in the abdominal area; (2) displayed a truculent and evasive manner; (3) recounted a medical history fraught with harrowing episodes of dubious veracity; and (4) attempted to conceal documents such as hospital discharge forms or insurance claim filings. In 1962 J. C. Barker argued that the fascinating but false stories of personal triumph and tragedy these patients told were not simply an intriguing feature, but instead a cardinal feature of the diagnosis. Barker referred to these tales as “pseudologia fantastica,” a term included in DSM-IV as a common feature of factitious disorder but one that is not essential for the diagnosis. *Pseudologia fantastica* is typified by stories that are not entirely improbable and that often contain a matrix of truth and falsehood. The stories are enduring and have a self-aggrandizing component.

FIGURE 17-1 A commemorative stamp issued by Germany in 1970 to mark the two hundred and fiftieth anniversary of Baron von Münchausen's birth. The Baron is depicted riding a severed horse into battle.

Asher suggested that Munchausen syndrome patients “desire to be the centre of interest and attention,” but still found himself unable to explain their often-startling behavior: “The most remarkable feature of the syndrome is the apparent senselessness of it. Unlike the malingerer, who may gain a definite end, these patients often seem to gain nothing except the discomfiture of unnecessary investigations or operations.” The resolution to this apparent paradox—one that tends to trouble observers but not the patients themselves—has remained a source of speculation. For example, in 1968 H.R. Spiro expanded upon Menninger's psychodynamic formulation and invoked disturbed early parent-child relationships as a critical etiological factor. He suggested that factitious behaviors were the consequence of early deprivation, an incomplete self-identity, superego defects, and abortive attempts at mastery of early trauma. He further noted that hospitals are a natural setting in which such unresolved issues could be reenacted because, like parents, medical professionals bestow caretaking even as they exert control over their charges. Spiro also chronicled one of the legacies of Asher's often-whimsical article, the ever-evolving and vaguely pejorative lingo used to describe factitious disorders. Terms such as “hospital hoboes” often seemed to reflect the authors' feelings of both bemusement and betrayal. Refuting the implication that these patients were mere “tricksters, liars, and swindlers,” Spiro advocated a more “impersonal nomenclature.” He suggested that, rather than “epithets,” these patients receive primary psychiatric diagnoses qualified with the phrase “with chronic factitious symptomatology.” Despite the 1980 introduction of the term “chronic factitial disorder” into the third edition of DSM (DSM-III), its retention in the revised third edition of DSM (DSM-III-R) and DSM-IV as “factitious disorder,” the appeal of the colorful *Munchausen* eponym has remained. “Munchausen syndrome” is mentioned only parenthetically in DSM-IV as a severe subtype involving repeated hospitalization, but the terms “factitious disorder” and “Munchausen syndrome” are often used interchangeably, especially by nonpsychiatrists. As a result, health care professionals often experience a reflexive therapeutic nihilism whenever a patient is determined to have feigned or induced an ailment, no matter the severity or chronicity.

The term *Munchausen syndrome by proxy* was coined in 1976 by John Money and June Werlwas to refer to psychosocial dwarfism, but came into broad use only as a result of its having been applied by Roy Meadow in a 1977 article. Consistent with a move in the official nomenclature away from eponyms, the syndrome entered

DSM-IV as “factitious disorder by proxy.”

COMPARATIVE NOSOLOGY

Arising from the Munchausen archetype, DSM-III emphasized chronic factitious disorder with physical symptoms. DSM-III-R offered an expanded nomenclature that reflected emerging evidence that factitious disorders may occur briefly or episodically, and that patients sometimes manifest mixed or exclusively psychological symptoms. It outlined distinct criteria sets for factitious disorder with physical symptoms and factitious disorder with psychological symptoms, and required the use of the not otherwise specified category for persons who presented with mixed symptom pictures. DSM-IV has merged the criteria sets into a single one, labeled *factitious disorder*, but has delineated the three subtypes described. In contrast, the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) lists factitious disorder within the category of disorders of adult personality and behavior under the name of “intentional production or feigning of symptoms or disabilities, either physical or psychological.” The diagnostic criteria in ICD-10 are listed in Table 17-4. Similar to DSM-IV, ICD-10 criteria imply that the demarcation between “internal” motivation (factitious disorder) and “external motivation” (malingering) is clearcut. In practice, a given patient's dissimulation may stem from a combination of incentives, both internal and external, which vary over time and even coexist; for example, a patient may feign severe pain both to acquire opioids and to obtain nurturance from a solicitous nursing staff. Despite this shared limitation, the DSM-IV entry appears more clinically useful for a number of reasons. First, ICD-10 goes further than DSM-IV in proposing that even the motivation of accessing “more medical care” is “external” and therefore suggests malingering; the task of differentiating the pursuit of the sick role from the pursuit of more medical care is taxing and speculative at best. Second, the ICD-10 criteria do not differentiate self-mutilation (as in borderline personality disorder) from the self-harm that can occur in factitious disorder; however, the latter but not the former typically involves deception about the cause of the injuries. Third, in the preamble to the criteria, the authors of ICD-10 select for inclusion two of the pejorative terms formerly applied to factitious disorder patients (“hospital hopper syndrome” and “peregrinating patient”); they also choose to list, without definition, the term “Münchhausen's syndrome” although this phrase is accurately applied only to a small subset of patients with factitious disorder. Other differences are that ICD-10 arbitrarily excludes factitial dermatitis and categorizes “Münchhausen by proxy” as a form of child abuse but not of factitious disorder.

Elaboration of physical symptoms for psychological reasons

A. Physical symptoms originally due to a confirmed physical disorder, disease, or disability become exaggerated or prolonged in excess of what can be explained by the physical disorder itself.

B. There is evidence for a psychological causation for the excess symptoms (such as evidence fear of disability or death, possible financial compensation, disappointment at the standard of care experienced).

Intentional production or feigning of symptoms or disabilities, either physical or psychological [factitious disorder]

A. The individual exhibits a persistent pattern of intentional production or feigning of symptoms and/or self-infliction of wounds in order to produce symptoms.

B. No evidence can be found for an external motivation such as financial compensation, escape from danger, or more medical care. (If such evidence can be found, the category malingering, should be used.)

C. Most commonly used exclusion clause. There is no confirmed physical or mental disorder that could explain the symptoms.

Other specified disorders of adult personality and behavior. This category should be used for coding any specified disorder of adult personality and behavior that cannot be classified under any one of the preceding headings.

Table 17-4 ICD-10 Diagnostic Criteria for Other Disorders of Adult Personality and Behavior

EPIDEMIOLOGY

Because factitious disorders always involve willful deception and sometimes involve itinerancy as well, standard epidemiological techniques are not applicable. Information must be inferred from individual case studies, a few reported series, and referral patterns, typically within university medical centers. For example, at a large teaching hospital in Toronto, factitious disorders were diagnosed in 10 out of 1288 (0.8 percent) patients referred consecutively to the psychiatric consultation-liaison service. At a New York City medical center, a comparable proportion of psychiatric consultation patients—1 percent—were diagnosed with factitious disorders. In one general hospital setting, 3.5 percent of fevers of unknown origin proved to be factitious, a rate higher than that of drug-related fevers. Since primary treating physicians often do not detect psychiatric illness in patients with physical complaints, these identified cases probably represented only a fraction of the actual prevalence of factitious disorders at these sites. Factitious disorders were identified in 32 of 343 patients (9.3 percent) referred for evaluation of fever of unknown origin in a report from the National Institute for Allergy and Infectious Disease in which nonpsychiatric physicians explicitly considered the diagnosis. Five of 1538 patients (0.3 percent) who were treated for ostensible neurological disorders at a Berlin hospital had an ultimate diagnosis of factitious disorder. Finally, 3 percent of the “kidney stones” submitted by patients in one report were determined to have been nonphysiological and to have represented probable attempts to deceive physicians. Experts concur that factitious psychological disorders that receive professional attention are much less frequent than factitious physical disorders. The prevalence has been estimated at 0.1 to 0.5 percent of inpatient psychiatric admissions.

In the Toronto study, factitious disorder patients had a median age of 26 years. The average age at onset of the dissimulations was 21 years, and other reports have also suggested onset in early adulthood. Overall, demographic analyses in the literature have distinguished two general patterns. Factitious disorder patients with Munchausen syndrome are typically middle-aged men who are unmarried, unemployed, and estranged from their families; the remaining patients are generally women aged 20 to 40 years. A number of reports suggest that those in the second group are commonly employed in or intimately familiar with health care occupations such as nursing and physical therapy. In a 10-year retrospective study of hospitalized patients, 28 of 41 patients with factitious disorder identified worked in medically related fields, 15 as nurses.

No genetic pattern has been established, although a large number of persons diagnosed with factitious disorders have comorbid psychiatric diagnoses (e.g., mood disorders, personality disorders, or substance-related disorders) that are genetically influenced.

Factitious disorder by proxy currently accounts for fewer than 1000 of the almost 3 million cases of child abuse reported each year in the United States, but this number may rise as mass media and professional attention increase recognition of these cases. Authors have attempted to elucidate the prevalence of factitious disorder by proxy within particular populations, such as children presenting with apnea (0.27 percent), allergy (5 percent), asthma (1 percent), apparent life-threatening episodes (1.5 percent), and life-threatening episodes treated with cardiopulmonary resuscitation (over 9 percent among children in whom final diagnoses were established). The average length of time to establish a diagnosis of factitious disorder by proxy after the initial presentation is 15 months, and often a sibling has died of undiagnosed causes before the disorder is recognized.

ETIOLOGY

Biological Factors Over the past decade, several researchers have proposed that brain dysfunction may play a role in factitious disorders, particularly the Munchausen variant. In single case reports or small series, some patients with Munchausen syndrome have displayed abnormalities on brain imaging, such as frontal temporal cortical atrophy, or on neuropsychological testing, particularly in the areas of conceptual organization, management of complex information, and judgment. It has been speculated that impaired information processing contributes to the aberrant behavior and pseudologia fantastica of Munchausen patients. Electroencephalography (EEG) studies have been small in scale, and abnormalities observed in a minority of the patients have been nonspecific.

Psychosocial Factors A number of psychodynamic causes for Munchausen syndrome have been proposed, and they appear applicable to persons with non-Munchausen factitious disorders as well. For many patients, the production of illness simultaneously meets multiple needs ranging from conscious to fully unconscious. Dependency needs are typically among the salient ones. These patients often have had emotionally deprived backgrounds with parents who were rejecting, and illness becomes a way of inducing others to gratify these needs and to satisfy the longing for nurturance. For other patients, particularly those prone to a decline in reality testing, factitious behavior organizes them, establishing a role to be played. Through this role, a helpless individual becomes a clever manipulator able to control health care providers, who may also represent surrogates of the parents. Similarly, factitious behavior also meets the need for identity. Patients with factitious disorders frequently have a poor sense of self and have failed to define their value systems and goals. By simulating or inducing disease, they assume the well-defined role of sick persons; through the pseudologia fantastica in Munchausen syndrome, they even become important and interesting people with exotic backgrounds. Another need gratified through dissimulation is the need for mastery. Individuals with poor coping skills, a limited social support system, and poor vocational skills may feel helpless. Successful appropriation of the status of “patient” stifles this weak and vulnerable feeling. In particular, the illusion of glamorous roles (e.g., professional athlete or war hero) in pseudologia fantastica provides them the vicarious pleasure of being someone who is respected and powerful. A classical psychoanalytic interpretation of the patient's eager acceptance of the sick role is underlying masochism. This theory suggests that inadequate, sometimes

abusive parenting has resulted in a distorted self-image, and that anger directed toward the child has resulted in a devalued sense of the self as one who deserves abuse. The Munchausen syndrome patient may therefore seek out situations that mix nurturing with pain, as in surgical procedures. Incorporated into this formulation is the concept of compromise formation: that a part of the body may need to be sacrificed to save the whole—that is, the sense of self.

FACTITIOUS DISORDER BY PROXY

By definition, factitious disorder by proxy requires that any external gains for the victim's fabricated or induced illnesses, such as disability payments or respite from child-rearing responsibilities during hospitalization, are incidental to the pursuit of the vicarious sick role. Several analyses have referred to a disorder of empathy among perpetrating mothers fueled by depression and isolation. With their spouses typically unavailable or uninvolved, they vitiate these painful feelings by mobilizing attention and nurturance through this disorder. A related hypothesis involves projective identification. Through this defense mechanism, the mother projects onto her child her own unconscious longings for nurturance, then ensures—through her own indefatigable attention as well as that of health care providers and others—that the child receives the attention she herself so desperately craves. Others have referred to factitious disorder by proxy as an epiphenomenon of the parent's narcissism and sociopathy, with glee arising from the capacity to dupe highly educated professionals. The perpetrator may also be displacing sadistic impulses toward herself or others onto the child. A recent theory suggests that an unsatisfactory relationship with a desired but unavailable father contributes to a perverse relationship with physicians and hospital staff members, surrogate parents whom the abusing mother simultaneously pursues and punishes. In this context the mother perceives her child as an object to be used to manipulate an intensely ambivalent relationship with the medical establishment. Concrete evidence for this last observation has been provided through covert videotapes. In contrast to the devoted, even symbiotic parenting style they reveal in public, these mothers are commonly unresponsive to their infants when their behavior is unwitnessed. Despite the perversity of their behavior, they rarely suffer from psychotic or dissociative disorders, although they often have personal histories of factitious or somatoform disorders. Although they may have been neglected or undervalued, most perpetrators did not suffer direct abuse in childhood.

DIAGNOSIS AND CLINICAL FEATURES

Factitious Disorder With Predominantly Psychological Signs and Symptoms The essential feature is a clinical presentation characterized by the intentional production or feigning of psychological signs and symptoms that suggest a mental disorder. The individual's goal is to assume the sick role, not to accrue external gain. The patient's symptomatology typically appears inconsistent with the known features of the mental disorder, responds poorly to treatments expected to be effective if the malady were authentic, and worsens when the patient is aware of being observed.

The preponderance of published cases of factitious disorder involve only physical signs and symptoms. Factitious psychological symptoms are most frequently seen in conjunction with physical complaints. In the accounts in which the factitious disorder consists exclusively of psychological symptoms, the individuals often present as the typical Munchausen syndrome patients, exhibiting peregrination, antisocial behavior, and a lack of intimate or sustained relationships. In addition, since these patients rely on verbal exchange to capture the attention of health professionals, pseudologia fantastica may be more prominent in those with factitious psychological rather than physical symptoms.

As Table 17-5 indicates, patients have displayed a formidable array of psychological manifestations. Presentations involving depression or suicidal ideation have often been linked to false claims of bereavement. In a study of 20 patients who had embellished or falsely reported the deaths of loved ones in order to procure the sick role, 15 had also feigned physical symptoms in the past. Of 12 patients in another report, 9 had claimed that they were grieving for more than one deceased family member and 5 had been hospitalized previously with factitious physical symptoms. Features commonly noted in factitious bereavement include a lack of witnesses to corroborate the patient's story and deaths of loved ones that are purported to have been especially dramatic or gruesome. Such dramatic but unverifiable events may also be reported in factitious victimization, in which an individual contrives baseless claims of rape, assault, or other crimes.

Bereavement	Eating disorder
Depression	Amnesia
Posttraumatic stress disorder	Substance-related disorder
Pain disorder	Paraphilias
Psychosis	Hypersomnia
Bipolar I disorder	Transsexualism
Dissociative identity disorder	

Adapted with permission from Feldman MD, Eisendrath SJ: *The Spectrum of Factitious Disorders*. © American Psychiatric Press, Washington, DC, 1996.

Table 17-5 Presentations in Factitious Disorder With Predominantly Psychological Signs and Symptoms

Although evidence exists in isolated cases for feigned psychosis, some authors contend that such cases need to be examined longitudinally before a diagnosis of factitious disorder is proffered. Such patients tend to have family histories of mental illness and, over time, often develop authentic psychotic disorders, such as schizophrenia. Indeed, feigning psychosis may be a defense against the emergence of a genuine psychotic disorder. Based on these considerations, at least one researcher has questioned the diagnostic legitimacy of the entire category of factitious psychological disorders. The debate about this category is likely to grow more vigorous; the distinction between psychological and physical is becoming increasingly artificial in the face of neurophysiological advances pointing to the interplay between abnormal psychology and neurobiology. Further, almost all patients with factitious psychological disorders have comorbid personality disorders.

The Minnesota Multiphasic Personality Inventory-2 (MMPI-2) may help to confirm the diagnosis of factitious disorder with predominantly psychological signs and symptoms. An invalid test profile and elevations of all clinical scales indicate an attempt to appear more disturbed than is the case (“fake bad”). Also, patients with factitious disorders, whether physical or psychological, most often demonstrate an average or above-average intelligence quotient (I.Q.), absence of a formal thought disorder, a poor sense of identity, poor sexual adjustment, poor frustration tolerance, strong dependency needs, and narcissism.

Ms. A. was a 35-year-old unemployed woman who presented to the emergency room complaining of auditory hallucinations and both suicidal and homicidal ideation. She acknowledged 10 prior psychiatric hospitalizations and diagnoses that had included major depressive disorder, bipolar I disorder, and dissociative identity disorder. She reported that “another personality,” whom she called Krystal, would assume control at times and that Krystal was responsible for her frequent episodes of rage. In addition, the voice instructing her to harm others, which she felt increasingly unable to resist, was that of Krystal. Ms. A. added that at times she would go 8 full days without sleep because of her continual shifts in identity states.

Ms. A. indicated that she had been married and divorced four times. After receiving a Bachelor's degree in psychology, she said, she had worked as a substance abuse counselor but had recently had to quit her job. The limited records available showed that she had been detoxified from benzodiazepines at age 18 and cocaine at age 28 and had attempted suicide on multiple occasions.

Following admission, Ms. A. abruptly dropped her complaints of hallucinations and disclaimed further suicidal or homicidal ideation. She was noted by the staff to socialize freely with other patients, but abruptly to assume a demeanor of perplexity or intense preoccupation when she was aware of being observed. Shortly after the attending psychiatrist began discussions with her about discharge planning, Ms. A. began to speak unintelligibly and wander seemingly aimlessly into other patient's rooms. However, her behavior normalized almost immediately after she was transferred to a more secure, notably less plush unit, and she insisted upon a return to the original ward. She later blamed Krystal for the uncontrolled behavior. Psychological testing indicated severe borderline personality disorder. Telephone calls to a family member and a former psychiatrist revealed that Ms. A. had never exhibited an alternate identity state. It was also learned that she had neither received a degree in psychology nor had she worked as a substance abuse counselor.

Ms. A. was discharged to the care of a local psychiatrist. Two weeks later, she was readmitted for less than 48 hours with complaints of suicidal intent. Her suicidal ideation promptly resolved when she learned that transfer to the secure unit was again being arranged, and she returned to the local physician.

Discussion Ms. A.'s diagnoses were factitious disorder with predominantly psychological signs and symptoms, and borderline personality disorder. Her apparent dissociation in the hospital appeared contrived and resolved almost instantly when she was transferred to a less desirable ward; this degree of control is absent in authentic dissociative identity disorder. Similarly, she abandoned the claim of suicidal and homicidal ideation as soon as admission was assured. No history of additional identity states could be culled from the records of a former treating psychiatrist, and a relative stated that any claim of dissociative identity disorder was spurious. Ms. A. did not solicit medications, disability payments, or other external gains with her fallacious dissociative history, but instead primarily seemed to be seeking the sick role. Ms. A.'s willful creation of Krystal may also have served as a ready explanation for the rage and impulsivity that were in reality attributable to her personality disorder. She gratuitiously enhanced her educational and work histories, and her relaxed sociability when she was not aware of being observed supported the impression of factitious disorder. She tailored her presentation to the settings and people around her and, in so doing, diverted attention from and impeded potential assistance with her actual emotional problems.

Factitious Disorder With Predominantly Physical Signs and Symptoms The essential feature is the intentional production or feigning of physical signs and symptoms, motivated solely by a psychological need to assume the sick role. The presentation of factitious physical symptoms spans a considerable range. Individuals may engage in: (1) total fabrications (e.g., asserting falsely that one is infected with the human immunodeficiency virus [HIV]), (2) simulations (e.g., mimicking a grand mal seizure); (3) illness aggravations (e.g., manually manipulating a preexisting wound so it will not heal); or (4) illness inductions (e.g., injecting oneself with fecal bacteria to cause sepsis). The specific examples in the literature are as varied and intricate as the human imagination, and are constrained only by the limitation of the particular individual's creativity and motivation. The maladies may be relatively common (e.g., urinary tract infections) or so esoteric that most physicians would have only a passing familiarity with them (e.g., intermittent Mediterranean fever; necrotizing fasciitis). The potential sophistication and multiplicity of the deceptions is well illustrated by five patients in a German report who feigned neurological syndromes: they had enacted meningoencephalitis, grand mal seizures with postictal agitated states, sudden loss of consciousness, progressive brainstem syndrome, status epilepticus, dysarthria, spastic paraparesis, hemiparesis, and paresis of one arm. The most common factitious physical signs, symptoms, and diseases reported in the literature include iron deficiency anemia, rash, chronic diarrhea, seizures, fever of unknown origin, renal stones, hematuria, hypoglycemia, cancer, and intestinal bleeding. Less frequent reports have involved every medical or surgical specialty, and have included such disparate conditions as anaplastic anemia, septic arthritis, lead poisoning, pheochromocytoma, Goodpasture's syndrome (hemoptysis and glomerulonephiritis), ectopic pregnancy, ventricular tachycardia, and torsion dystonia.

Ms. B., a 24-year-old woman, presented to a medical clinic for treatment of chronic vomiting. She described herself as having C5 quadriplegia and used a motorized wheelchair. Her limbs were markedly atrophic but she retained sufficient movement to operate her wheelchair and perform most activities of daily living. Ms. B. attributed her quadriplegia to a motor vehicle accident she had suffered at the age of 13, and she had a scar on her neck at the site of subsequent spinal fusion. Due to severe spasticity, an in-dwelling intrathecal baclofen pump had been inserted and, due to apparent bladder atony, she had had an ileal loop procedure for urine drainage. Based upon the patient's report of unremitting fecal incontinence, a colostomy had been performed 1 week prior to her presentation at the clinic.

Ms. B.'s vomiting failed to respond to dietary manipulations, and increasing dehydration and malnutrition necessitated hospitalization. She was ultimately discharged on total parenteral nutrition. A series of readmissions ensued, however, for line sepsis caused by unusual microorganisms such as *Clostridia* and fungi. It was during one of these hospitalizations that a nurse recognized Ms. B. and recalled having taken care of her after the car accident years earlier. She remembered clearly that the patient had had no lingering physical sequelae at discharge. For the first time, the diagnosis of factitious disorder was entertained.

Ms. B. was not confronted with this suspicion. Instead, she was referred for extensive neurological testing that confirmed that her spinal cord was intact and that her peripheral nerve function was normal. Although Ms. B. had consistently denied permission for the staff to speak with her parents, she finally acceded when her primary physician threatened to resign from her case. A family meeting was scheduled.

During the family session, Ms. B.'s parents stated that they had always felt badly that Ms. B. discouraged their involvement, going so far as to insist that they not visit her in the hospital. They confirmed that she had been physically normal following the accident. It was not until her senior year of high school as she was preparing for college that she developed neurological dysfunction that defied diagnosis. Her incapacity progressed until she became wheelchair-bound. Confronted in the context of the family meeting, Ms. B. acknowledged hesitantly that perhaps she did not have objective physical pathology and proposed that she had “acted quadriplegic” because her health care providers had told her she was quadriplegic.

After the meeting, a multidisciplinary treatment team was formed. The team met regularly and made it clear to Ms. B. that the goals for her were nothing short of full recovery. Within this firm and expectant context, she received physical therapy, occupational therapy, and psychological counseling supplemented by home visits by a psychiatric nurse. Regularly scheduled medical follow-up was provided, and her baclofen pump was eventually removed. She eventually became able to walk with leg braces and a walker. She regained the lost weight and has required no additional hospitalizations since the family session occurred.

Discussion Consistent with the diagnosis of factitious disorder, Ms. B. sought to assume the sick role, particularly to mobilize the attention and concern of skilled medical professionals and to gain relief from responsibilities and expectations such as establishing autonomy. Here, as in many other cases, repeated diagnostic studies, invasive procedures, and major surgical operations were performed in an effort to manage what ultimately proved to be feigned or self-induced ailments.

Ms. B.'s behavior was characteristic in that she demonstrated essentially no anxiety even when confronted with painful and disfiguring surgery. In addition, she strenuously resisted efforts to confirm the details of her stories and almost never had visitors. In this case the patient's deceptions proved so effective that the diagnosis was considered only when a nurse coincidentally recognized her from a much earlier hospitalization. Psychological testing confirmed that Ms. B. had a severe personality disorder with avoidant and dependent features. (Adapted from Feldman MD, Duval NJ: Factitious quadriplegia: A rare new case and literature review. Psychosomatics *38*:76, 1997.)

Factitious Disorder With Combined Psychological and Physical Signs and Symptoms If a patient presents with factitious disorder with both physical and psychological features and neither type of disorder predominates clinically, the appropriate diagnosis is factitious disorder with combined psychological and physical signs and symptoms. In one representative report, a patient alternated among feigned dementia, bereavement, rape, and seizures. Another patient, confronted about his factitious posttraumatic stress disorder, began to offer complaints of incapacitating pain associated with metastatic cancer that proved factitious as well (Table 17-1).

Factitious Disorder Not Otherwise Specified This category is reserved for patients who do not meet the criteria for the other factitious disorder classifications. The example provided in DSM-IV is factitious disorder by proxy. According to the research criteria for this disorder the essential feature is the intentional production or feigning of physical or psychological signs and symptoms in another person who is under the individual's care. The motivation for the perpetrator's behavior is to assume the sick role by proxy, and external incentives for the behavior (such as economic gain) are absent. In 95 to 98 percent of reported cases, the abusive individual is the mother and the victim is her own child; fewer than 20 cases have been published in which the father is the perpetrator. Typically the child is an infant or toddler, but the ages reported have varied from neonates to 14-year-olds. In extremely rare cases, an adult will produce an illness in another adult. The manufactured signs and symptoms are almost always physical, and the perpetrator creates the dissimulation through behaviors such as spurious or exaggerated symptom reports, manipulation of laboratory samples or data, or actual induction of illness through maneuvers such as surreptitious administration of drugs. Table 17-6 summarizes the most common presentations of factitious disorder by proxy. Such descriptions have been reaffirmed by hundreds of case reports in the medical literature from around the world.

Poisoning (includes Munchausen syndrome by proxy and intentional poisonings)
Seizures or vomiting
Apnea
Diarrhea
Unconsciousness
Fevers
Lethargy
Dehydration or hematemesis
Ataxia or hematuria

Adapted with permission from Schreier HA, Libow JA: *Hurting for Love: Münchausen by Proxy Syndrome.* © Guilford Press, New York, 1993; and Rosenberg DA: Web of deceit: A literature review of Münchausen syndrome by proxy. *Child Abuse Negl* 11:553, 1987.

Table 17-6 Ranking of the Most Common Bibliographic References to Signs and Symptoms of Factitious Disorder by Proxy

In a typical case of factitious disorder by proxy, the child is brought to the clinic or hospital with dramatic symptoms. The mother characteristically assists the staff and readily consents to any invasive diagnostic procedures proposed for the child, often seeming more concerned about obtaining additional tests than about the child's emotional status. Discovery of her role in the production of the child's symptoms may occur accidentally, such as finding her smothering the child with a pillow to induce hypoxic seizures. At other times, mothers have been found administering ipecac to cause vomiting or introducing feces into intravenous tubing to create sepsis. Barring such direct findings, detection can be supported by interventions such as finding toxic amounts of sodium in food items that the mother has brought for the child; drug screening tests unexpectedly indicating surreptitious administration of drugs; or anomalous, inexplicably variable, or flatly impossible laboratory results. Older siblings may have had similar findings that were medically unexplained at the time. Gross evidence of abuse—such as burns, bruises, or fractures—is uncommon in factitious disorder by proxy. Instead, clinicians should consider the diagnosis based upon the indicators that have emerged in the literature (Table 17-7).

The symptoms and pattern of illness are extremely unusual, or inexplicable physiologically.
Repeated hospitalizations and workups by numerous caregivers fail to reveal a conclusive diagnosis or cause.
Physiological parameters are consistent with induced illness; e.g., apnea monitor tracings disclose massive muscle artifact prior to respiratory arrest, suggesting that the child has been struggling against an obstruction to the airways.
The patient fails to respond to appropriate treatments.
The vitality of the patient is inconsistent with the laboratory findings.
The signs and symptoms abate when the mother has not had access to the child.
The mother is the only witness to the onset of signs and symptoms.
Unexplained illnesses have occurred in the mother or her other children.
The mother has had medical or nursing education, or exposure to models of the illnesses afflicting the child (e.g., a parent with sleep apnea).
The mother welcomes even invasive and painful tests.
The mother grows anxious if the child improves.
Maternal lying is proved.
Medical observations yield information that is inconsistent with parental reports.

Adapted with permission from Feldman MD, Eisendrath SJ: *The Spectrum of Factitious Disorders.* © American Psychiatric Press, Washington, DC, 1996.

Table 17-7 Clinical Indicators That May Suggest Factitious Disorder by Proxy

By 9 weeks of age, Carter C. had spent the majority of his life in the hospital for recurrent idiopathic apnea. Evaluations by multiple subspecialty teams had been unrevealing. Similarly, medical tests—which had been expanded to include lumbar puncture, EEG, continuous electrocardiogram (ECG) monitoring, upper gastrointestinal tract series, pH probe, overnight sleep study, herpesvirus culture, mandible films, chest X-ray, and numerous blood studies—were notable only for minor findings of doubtful significance. An empiric trial of metoclopramide (Reglan) was ineffective. Mrs. C. had been the only witness to the onset of the apneic spells, and her reports sometimes struck the staff as unconvincing. For example, she had claimed that the apneic episode precipitating the most recent admission had lasted less than 1 minute, but Carter's blood pH had been 6.9, suggesting that the apnea had been sustained.

These facts, combined with Mrs. C.'s eagerness to have Carter undergo invasive tests, led to a request for psychiatric consultation. Review of the records by the psychiatrist revealed that Mrs. C.'s first child, Denise, had died at 18 months of age of aspiration pneumonia. Denise had had 3 hospitalizations and 9 emergency room presentations for seizures or apnea observed only by Mrs. C. Phenobarbital had been prescribed empirically, but there were unaccountable variations in the blood concentration. During the interview, Mrs. C. disclosed that she had been unable to realize her goal of becoming a paramedic, but often accompanied her mother, a nurse, to continuing-education courses. She used medical terms freely and expressed pride in having resuscitated Carter on numerous occasions. Now divorced from Mr. C., her third husband, she lived with her sister, who was disabled.

Based upon the accumulated evidence of factitious disorder by proxy, and despite Mrs. C.'s emphatic denial, the Department of Human Resources placed Carter with a foster mother. The child remained well during this period. At the Shelter Care and Dependency trial, the medical examiner voiced concerns about Denise's illnesses and death, and noted disturbing parallels with Carter's course to date. Following the testimony of the state's psychiatric expert, and before a judgment was rendered, Mrs. C. agreed unilaterally to an award of temporary custody to Mr. C. Six months later, Carter was thriving without medical intervention, and Mr. C. was awarded permanent custody. While continuing to deny having engaged in factitious disorder by proxy Mrs. C. stated that she would “acquiesce” to this arrangement. The State elected, without explanation, not to pursue an inquiry into Denise's death.

Discussion In DSM-IV, factitious disorder by proxy is a diagnosis that accrues to the abusive caretaker. The victim is not given a psychiatric diagnosis, but is identified with a V code, “physical abuse of child” (or in rare cases involving an adult victim, “physical abuse of adult”).

This case was typical in that the conclusion that the mother actually produced the symptoms in the child was based on an accumulation of circumstantial evidence. In cases of factitious disorder by proxy the mother almost never acknowledges her role in the dissimulation. This denial complicates detection and management, although intervention, as in this example, must focus on ensuring the safety of the child regardless of the caregiver's response. Mrs. C. appeared to be motivated by the desire to experience the sick role indirectly through her child and to realize her unfulfilled goal of becoming part of the health care system.

Like most cases of factitious disorder by proxy, Mrs. C. appeared superficially quite normal; in fact, this diagnosis is not an “attribute” of an individual that can be elicited during an interview. Psychiatric evaluations and psychological testing do not enable an examiner to confirm or disprove the diagnosis, although one report claims that many of these parents display nonspecific features such as limited stores of information (except in the medical arena); poor abstract conceptual abilities; superficial social skills; outgoing behavior; rigid, defensive styles that mask rebelliousness; emotional immaturity; self-centeredness; a lack of social conformity; and intense passive resentment.

DIFFERENTIAL DIAGNOSIS

Factitious Disorder With Predominantly Physical Signs and Symptoms

True Physical Illness An important item in the differential diagnosis is an authentic physical illness. A true disease may account for the clinical presentation, as in the report of a patient who died of encephalitis after having been spurned by the staff as having feigned her presentation. The complexity of the challenge to the diagnostician is heightened by the fact that patients may self-induce ailments such as sepsis, conditions that are life-threatening and require vigorous medical intervention despite being manifestations of factitious disorder. As another example, the history of numerous abdominal surgical procedures in many cases of Munchausen syndrome predisposes the patient to actual intestinal obstruction, necessitating even further surgery.

Somatoform Disorders In somatoform disorders, such as conversion disorder, somatization disorder, and hypochondriasis, the patient reports or manifests physical symptoms that suggest a general medical condition but are not fully explained by a medical ailment, the direct effects of a substance, or another mental disorder. Unlike patients with factitious disorders, those with somatoform disorders do not intentionally produce their somatic symptoms. However, the presence of factitious or malingered symptoms mixed with other nonintentional symptoms is not uncommon, particularly in somatization disorder.

Personality Disorders Patients with Munchausen syndrome frequently meet criteria for antisocial personality disorder as well, but this comorbidity is not invariable. Like the patient with Munchausen syndrome, individuals with antisocial personality disorder alone may exhibit pseudologia fantastica, shoddy work records, substance-related disorders, and few sustained or intimate relationships, but they do not center their lives around hospitalization. Borderline and other personality disorders are also extremely common in Munchausen syndrome and factitious disorder as a whole. The diagnosis of these personality disorders is based on explicit criteria—in the case of borderline personality disorder, for example, the diagnosis is based on the person's identity disturbance, affective instability, suicidal or self-mutilating behavior, impulsivity, and chronic feelings of emptiness. If the patient meets criteria for both personality disorder and factitious disorder, the Axis II and Axis I diagnoses may be made concurrently.

Malingering In DSM-IV malingering is listed as an additional condition that may be a focus of clinical attention. In factitious disorders, the motivation is to assume the sick role, whereas in malingering, external incentives such as financial compensation, avoidance of military duty, or evasion of criminal prosecution, are apparent. However, when perceived incentives are used as a diagnostic criterion, imprecision in diagnosis is bound to occur. In addition, patients may feign illness to achieve different goals—internal or external—at different times.

Factitious Disorder With Predominantly Psychological Signs and Symptoms The patient's simulation of a mental disorder may actually represent the prodrome to an authentic mental disorder with a serious outcome. In particular, clinicians should be cautious in diagnosing factitious psychosis because in two small series a majority of these patients eventually manifested clear-cut psychotic disorders such as schizophrenia. In other cases, an ostensibly feigned condition such as depression has responded to psychotropic medications, validating at least some element of the presentation. Since virtually all patients with this type of factitious disorder have a personality disorder (usually borderline, histrionic, or antisocial), caregivers must also recognize that the simulated mental disorder coexists with an authentic one. Comorbidity with substance-related disorders and somatoform disorders has been reported as well, and dissociative disorders may result in alterations of memory and the patient's providing inconsistent factual information. In the past, Ganser's syndrome had been considered by a number of authors to fall within the realm of factitious psychological disorders, but it is classified in DSM-IV as an example of dissociative disorder not otherwise specified. The hallmark of Ganser's syndrome is approximate answers, or *vorbeireden*, to questions (e.g., Examiner: “How many legs does have a spider have?”; Patient: “Seven”). Although the syndrome has been shown to be present in some patients with impaired cognitive abilities and previous brain injury, there is clearly some conscious component to the responses provided by the patient because random answers would not be approximate, and it has been described in one patient who also had clearcut factitious disorder with combined physical and psychological signs and symptoms.

The clinician's level of suspicion that a factitious psychological disorder is present should be raised by a number of factors. At times, the symptoms are patent caricatures, reflecting a layperson's view of how those with mental disorders would behave. Cases involving feigned depression, however, may be modeled from the individual's personal experiences and thus more difficult to detect. Recognition of inconsistencies in the patient's account, or between the professed history and the patient's objective presentation, facilitates detection. For example, individuals falsely reporting posttraumatic stress disorder as a result of combat may be unmasked as a result of inherently flawed narratives or the staff's acquisition of military records or other collateral information; indeed, these patients may be discovered to have never served in the military at all. Other clues arise when the patient's symptoms are only present when observation by others is evident; when new symptoms arise that are incongruous with the condition already diagnosed; when new symptoms emerge that are similar to those exhibited by other patients on the ward; and when there is a consistent and chronic lack of response to standard treatments. In some cases, psychological testing (e.g., the Malingering/Dissimulation Scale of the MMPI-2, Structured Interview of Reported Symptoms, M Test, projective tests, or the Bender Visual Motor Gestalt Test) may aid in arriving at the correct diagnosis.

Factitious Disorder by Proxy Several categories of illness presentation can be confused with factitious disorder by proxy. First, overanxious parents may continually bring their children for medical evaluation and treatment, therefore putting them at risk for iatrogenic complications. However, this behavior is founded in a genuine belief that their offspring are ill, and they are at least transiently relieved when illness is ruled out. The signs and symptoms reported in such cases are usually not exotic and they respond predictably to medical intervention. Second, parents of children who are recovering from a documented health crisis such as prematurity may appear to overreact to mild symptoms of illness. The physician can identify these families by documenting the veracity of the previous diagnoses. Once recovery is well established, these parents will begin to attenuate their reactions to illness. Third, some parents discontinue prescribed medications because of concerns about adverse effects, not because they wish to mislead medical caregivers or worsen the child's symptoms. After they receive additional information, these parents comply or at least become more forthright in expressing their reservations about the medication regimen. Fourth, a parent's nominal exaggeration of her child's symptoms may represent an attempt to call attention to symptoms that concern her but seem not to impress the pediatrician. In contrast to cases of factitious disorder by proxy, these parents welcome recovery when it occurs and exaggeration of symptoms does not represent a pattern of behavior. Fifth, a number of illnesses, such as asthma, present with such variability that it is possible for the same child to appear remarkably different within a short period of time. An innocent parent may appear to be an illness exaggerator if she is describing a child whose symptoms no longer fit her description; independent corroboration will provide the necessary clarification. Sixth, an older child may malinger for secondary gain. The innocent parent witnesses and presents the child's symptoms to the physician, and appears to have manufactured the symptom reports when illness is disconfirmed by examination. Finally, a legitimate medical disorder occasionally presents with symptoms that can be confused with various manifestations of factitious disorder by proxy. For example, although sudden infant death syndrome is a well-described phenomenon, the absence of a clearcut marker renders parents vulnerable to charges of abuse. Guidelines are in the process of development that will help to differentiate this syndrome from induced apnea. For instance, repeated episodes of apnea culminating in the death in a child over the age of 6 months point to factitious disorder by proxy much more commonly than to sudden infant death syndrome. Covert video surveillance of mother and child in the hospital room has been used to confirm the diagnosis of factitious disorder by proxy although this method is controversial from a legal and ethical standpoint. The decision to undertake covert video surveillance should ideally involve a team comprised not only of medical personnel, but of legal, child protection, and law enforcement representatives as well. Prevailing case law in the United States suggests that the serious, potentially life-threatening nature of the abuse in factitious disorder by proxy provides a strong countervailing interest when balanced against a parent's or caretaker's right of privacy.

COURSE AND PROGNOSIS

Information concerning the course and prognosis of factitious disorder with predominantly physical signs and symptoms is limited to anecdotal case reports and small series. Patients who meet the criteria for Munchausen syndrome appear generally to have an unremitting course. Death may occur from miscalculations in disease-producing behavior (e.g., an overdose of chemotherapeutic agents that were being taken surreptitiously) or as a result of iatrogenic complications. Patients with Munchausen syndrome rarely accept psychiatric treatment; when they do, such intervention serves to ameliorate factitious behavior only transiently; the patients almost inevitably move on and resume their previous lifestyle. The course and prognosis for patients with simple (nonperegrinating) factitious disorder is more variable. When confronted with their behavior, most of these patients will deny the factitious nature of their illnesses; some will then quietly discontinue the dissimulation, some will change doctors and resume their illness-producing behavior, and a few will accept a referral for psychiatric treatment. If patients persist, they—like Munchausen patients—are at risk for a panoply of complications that include unnecessary amputations, removal of internal organs, and death. Non-Munchausen forms of the disorder appear usually to remit spontaneously after about age 40. Perhaps this remission is related to the observation that features of borderline personality disorder also tend to abate at midlife.

The literature on factitious disorder with predominantly psychological signs and symptoms is notable for the magnitude of the psychological dysfunction present in the patients. Almost all have serious personality disorders, often associated with substance abuse. Several authors have reported that there is a high rate of suicide in

this population and that factitious psychological disorders have a worse prognosis than most other Axis I disorders.

A comprehensive 1987 literature review of factitious disorder by proxy showed a 9 percent mortality rate. Although this rate was probably inflated from its having been derived only from published reports, a later study highlighted other risks. In this article, British researchers followed up 41 victims of factitious disorder by proxy an average of 5.6 years after the condition was identified. They showed that, of the moderately severe cases in which the children had been returned home, the fabrications were repeated in 33 percent. In addition to the caretaker's abuse of the child, diagnostic and treatment maneuvers in pursuit of the elusive diagnosis may also engender their own short- and long-term morbidity. Behaviorally, a range of conduct, school, and emotional disorders was observed in around half of the British children whether they had been returned home or placed with others. In a more recent study of 10 adults who identified themselves as having experienced factitious disorder by proxy in childhood, persistent insecurity, reality-testing deficits, and symptoms of posttraumatic stress disorder were uncovered. In several other reports, individuals who had been victimized by factitious disorder by proxy have had an increased likelihood of developing factitious disorder later in life although some individuals, consistent with the profile of posttraumatic stress disorder, have avoided even essential testing and procedures that remind them of the original trauma.

TREATMENT

Factitious Disorder The primary goals in treatment of factitious disorders are to reduce the patient's disease dissimulation and the overutilization of medical and surgical services. Psychiatric consultation should be initiated when the diagnosis is first seriously considered. Consultation, perhaps using the format of a task force or management team, should be pursued in conjunction with nursing, social work, and legal professionals, and should include an ethicist when available. This multidisciplinary approach can be invaluable because patients with factitious disorder often mobilize interpersonal conflicts among their caretakers, including splitting, and raise complex management issues—for example, is it legal, ethical, and medically warranted to perform a surreptitious room search based on the suspicion that a patient is self-injecting with insulin?

Although there are isolated anecdotal reports of successful treatment of factitious disorder, there is no specific treatment that has been proved to be the gold standard. Nonetheless, authorities generally advocate that, following confirmation of the diagnosis, the patient should be confronted in a supportive manner that redefines the patient's illness from that of physical disease to that of psychological distress. Rather than angrily rejecting the patient as a fraud, the psychiatrist should extend an offer of help. If the patient accedes, a careful diagnostic evaluation is then needed to determine the presence or absence of comorbidity. Although concurrent Axis I disorders are relatively uncommon, they should be vigorously treated if present. Essentially all patients with factitious disorder are comorbid for a personality disorder, and an assessment of ego strengths is mandatory because many of these patients are at risk for acting-out behaviors, including psychosis and suicide.

Pharmacotherapy of factitious disorders is largely limited to treatment of any comorbid Axis I disorders. Medications should be carefully monitored because of patients' propensity to act in self-destructive ways. The selective serotonin reuptake inhibitors (SSRIs) may prove to be of some use in reducing impulsivity. While not treating the underlying causes for an individual's factitious behavior, other somatic interventions, such as the provision of physical therapy, may provide both needed attention and a face-saving mechanism for the patient to curtail or abandon factitious behavior.

The psychotherapy of patients with factitious disorders is directed to the underlying personality disorder. Thus, the indications, techniques, and complications are similar to the those that apply to the treatment of any person with a Cluster B personality disorder. Persons with severe personality disorders, if seen in reconstructive-type therapies, are prone to intense transference responses and acting-out behaviors as a result of elicited anxiety; in this respect, the patient with factitious disorder may engage once again in disease forgery or flee treatment to resume previous behavior in a new locale. Supportive psychotherapies, although not curative, may be more effective in the containment of symptoms. Supportive techniques include the establishment of an ongoing relationship in which there is minimal confrontation of the patient's factitious behaviors or pseudologia fantastica even though the therapist does not express agreement with them. Instead, increased attention to material involving day-to-day conflicts provides a helpful reinforcement of the patient's attempt to struggle with real problems while serving also to reduce the patient's preoccupation with fantasy. Approximate or general interpretations may be particularly useful at times—for example, “You find it difficult when your mother leaves town on a business trip,” rather than “You feel abandoned and helpless when your mother leaves town and resort to factitious behavior to create a sense of power.”

The psychotherapist who treats these patients must have modest expectations, anticipating relapse and accepting periods of symptomatic relief rather than “cure.” It is the relationship itself that at least transiently serves to provide the nurturance that these patients crave.

Family intervention, especially for the younger patient, offers promise as an effective adjunctive approach (as in the case of Ms. B.). Such therapy can decrease family secrets and help the family to provide more attention that is not contingent on perceived illness, thereby diminishing the reinforcement of sick-role behaviors.

Behavior therapy has been suggested for some factitious disorder patients. The goals of such strategies are to allow face-saving ways for the patient to relinquish symptoms and learn socially appropriate techniques to have needs met.

Overall, several clinical indicators of enhanced treatment responsiveness have been elucidated. They include: (1) underlying psychiatric syndromes, such as mood, anxiety, substance-related, or conversion disorders; (2) personality traits without borderline or antisocial elements; (3) psychosocial supports, such as ongoing relationships with significant others, employment or employability, or religious affiliations; and (4) ability to establish a therapeutic alliance as characterized by the capacity to establish and maintain rapport, accept confrontation, and comply with treatment recommendations.

As suggested, clinicians should be aware of their countertransference toward patients with factitious disorders. The patient's overt behaviors, such as actual bodily damage, as well as their underlying emotional conflicts can mobilize particularly intense reactions in health professionals. These reactions may include therapeutic nihilism, anger and aversion, nonemergent breaches of confidentiality, overidentification, or feelings of personal responsibility for the patient's ongoing behavior. Identification of these countertransferential feelings will enhance the clinician's understanding of the patient.

Factitious Disorder by Proxy Authors differ in the intensity with which they have confronted the abusive parent and in the amount of supporting evidence they have produced. Categorical denial is extremely common whether a gentle suggestion is made, previously concealed medications or medical equipment are displayed, or definitive evidence from video surveillance is provided. Unless legal mechanisms to prevent it are in place, escalation of the behavior or immediate discharge of the child against medical advice is common. The management plan must therefore involve the immediate protection of the child. By law in all 50 states, health professionals who suspect or diagnose factitious disorder by proxy must alert child abuse authorities. In many situations the child's safety can be assured only by removing him or her from the home, at least temporarily. Reunification of families should be undertaken only if a plan, typically one mandated by the court system, is in place to help ensure the child's continued safety. Criminal prosecution of the mother may be indicated as well.

For the victimized child, additional medical care to address residual physical problems may be necessary; the child should also receive a psychiatric evaluation and therapy. Despite the abuse, children remain attached to their parents and suffer from the separation. If possible, perpetrating mothers and other family members should be engaged in long-term therapy. Ideally, the therapy will focus on teaching the parent adaptive ways to get her needs met, including expressing painful affects in words rather than with abusive actions. However, no author has yet described a consistently effective or specific treatment program for the factitious disorder by proxy maltreater, especially if an acknowledgment of culpability is not forthcoming. Without definitive intervention, the risk remains that the parent will relocate to an area where she is unknown and where the dissimulations can be successfully repeated. A listing of guidelines endorsed by a number of researchers and clinicians is provided in Table 17-8.

The child should generally be removed from the care of the mother, and separation should continue at least until the mother's evaluation is completed and the child's health status is clarified.
A pediatrician or family practitioner familiar with the case should become the long-term, court-ordered gatekeeper for the child's health care needs.
The gatekeeper should perform a periodic audit of the child's health and school records to help detect school absences or developmental and behavioral consequences of the disorder.
Audits of the health status of siblings should be performed for an extended period.
Individual psychotherapy is indicated for the mother and, if old enough, for the child.
Family therapy should be provided to the parents, siblings, and affected child.
If the family is reunited, supervision should be continued and arrangements made with the courts to have monitoring continue following any relocations.

Adapted with permission from Feldman MD, Eisendrath SJ: *The Spectrum of Factitious Disorders*. © American Psychiatric Press, Washington, DC, 1996.

Table 17-8 Management Recommendations Following Recognition of Factitious Disorder by Proxy

SUGGESTED CROSS-REFERENCES

Chapter 16 discusses somatoform disorders; Section 27.2 focuses on malingering. Personality disorders are the subject of Chapter 24; information on behavior therapy is presented in Section 30.2; information on insight-oriented psychotherapy is discussed in Section 30.1; psychological testing is discussed in Section 7.4 and Section 7.5.

CHAPTER REFERENCES

*Adshead G, Brooke D, editors: *Münchausen's Syndrome by Proxy: Current Issues in Assessment, Treatment and Research*. Imperial College Press, London, 1999.

*Asher R: Munchausen's syndrome. Lancet *1:*339, 1951.

Barker JC: The syndrome of hospital addiction (Munchausen syndrome). A report on the investigation of seven cases. J Ment Sci *108:*167, 1962.

Bauer M, Boegner F: Neurological syndromes in factitious disorder. J Nerv Ment Dis *184:*281, 1996.

Bools C, Neale B, Meadow R: Co-morbidity associated with fabricated illness (Munchausen syndrome by proxy). Arch Dis Child *67:*77, 1992.

Bools CN, Neale BA, Meadow SR: Follow up of victims of fabricated illness (Munchausen syndrome by proxy). Arch Dis Child *69:*625, 1993.

Byard RW, Burnell RH: Covert video surveillance in Munchausen syndrome by proxy. Med J Australia *160:*352, 1994.

Carter RB: *On the Pathology and Treatment of Hysteria*. John Churchill, London, 1853.

Davis P, McClure RJ, Rolfe K, Chessman N, Pearson S, Sibert JR, Meadow R: Procedures, placement, and risks of further abuse after Munchausen syndrome by proxy, non-accidental poisoning, and non-accidental suffocation. Arch Dis Child *78:*217, 1998.

Earle JR, Folks DG: Factitious disorder and coexisting depression: A report of successful psychiatric consultation and case management. Gen Hosp Psychiatry *8:*448, 1986.

Edi-Osagie EC, Hopkins RE, Edi-Osagie NE: Munchausen's syndrome in obstetrics and gynecology: A review. Obstet Gynecol Surv *53:*45, 1998.

Eisendrath SJ: Factitious physical disorders: Treatment without confrontation. Psychosomatics *30:*383, 1989.

Feldman MD: Denial in Münchausen syndrome by proxy: The consulting psychiatrist's dilemma. Int J Psychiatry Med *24:*121, 1994.

Feldman MD, Bibby M, Crites SD: “Virtual” factitious disorders and Munchausen by proxy. West J Med *168:*537, 1998.

Feldman MD, Duval NJ: Factitious quadriplegia: A rare new case and literature review. Psychosomatics *38:*76, 1997.

*Feldman MD, Eisendrath SJ, editors: *The Spectrum of Factitious Disorders*. American Psychiatric Press, Washington, DC, 1996.

Feldman MD, Feldman JM: Tangled in the web: Countertransference in the therapy of factitious disorders. Int J Psychiatry Med *25:*389, 1995.

Feldman MD, Ford CV: *Patient or Pretender: Inside the Strange World of Factitious Disorders*. John Wiley & Sons, New York, 1994.

Feldman MD, Rosenquist PB, Bond JP: Concurrent factitious disorder and factitious disorder by proxy. Gen Hosp Psychiatry *19:*24, 1997.

Folks DG: Munchausen's syndrome and other factitious disorders. Neurol Clin *13:*267, 1995.

Ford CV: *Lies! Lies!! Lies!!! The Psychology of Deceit*. American Psychiatric Press, Washington, DC, 1996.

Ford CV: The Munchausen syndrome: A report of four new cases and a review of psychodynamic considerations. Psychiatr Med *4:*31, 1973.

Ford CV, Abernethy V: Factitious illness: A multidisciplinary consideration of ethical issues. Gen Hosp Psychiatry *3:*329, 1981.

Gavin H: *On the Feigned and Factitious Diseases of Soliders and Seamen, on the Means Used to Simulate or Produce Them, and on the Best Modes of Discovering Impostors*. University Press, Edinburgh, 1838.

*Goldfarb J, Lawry KW, Steffen R, et al: Infectious disease presentations of Munchausen syndrome by proxy. Clin Pediatr (Phila) *37:*179, 1998.

King BH, Ford CV: Pseudologia fantastica. Acta Psychiatr Scand *77:*1, 1988.

Libow JA: Münchausen by proxy victims in adulthood: A first look. Child Abuse Negl *19:*1131, 1995.

Meadow R: Munchausen syndrome by proxy abuse perpetrated by men. Arch Dis Child *78:*210, 1998.

*Meadow R: Munchausen syndrome by proxy: The hinterland of child abuse. Lancet *2:*343, 1977.

*Meadow R: Unnatural sudden infant death syndrome. Arch Dis Child *80:*7, 1999.

Meropol NJ, Ford CV, Zaner RM: Factitious illness: An exploration in ethics. Perspect Biol Med *28:*269, 1985.

Parsons T: *The Social System*. Free Press, Glencoe, 1951.

Pilowsky I: A general classification of abnormal illness behaviours. Br J Med Psychol *51:*131, 1978.

Reich P, Gottfried LA: Factitious disorders in a teaching hospital. Ann Intern Med *99:*240, 1983.

Rosenberg DA: Web of deceit: A literature review of Munchausen syndrome by proxy. Child Abuse Negl *11:*547, 1987.

*Schreier HA, Libow JA: *Hurting for Love: Munchausen by Proxy Syndrome*. Guilford, New York, 1993.

Sigal MD, Altmark D, Carmel I: Munchausen syndrome by adult proxy: A perpetrator abusing two adults. J Nerv Ment Dis *174:*696, 1986.

Solyom C, Solyom L: A treatment program for functional paraplegia/Munchausen syndrome. J Behav Ther Exp Psychiatry *21:*225, 1990.

Spiro HR: Chronic factitious illness: Munchausen's syndrome. Arch Gen Psychiatry *18:*569, 1968.

Sutherland AJ, Rodin GM: Factitious disorders in a general hospital setting: Clinical features and a review of the literature. Psychosomatics *31:*532, 1990.

CHAPTER 18. DISSOCIATIVE DISORDERS

18.1 DISSOCIATIVE AMNESIA

MARLENE STEINBERG, M.D.

Amnesia is the black hole of dissociation: neither the dead star nor the symptom can be directly observed; their existence and nature must be inferred from only the surrounding effects. Patients usually become aware of their amnesia only after it begins to disrupt their social or work environments. However, recent advancements in psychiatry—including new diagnostic tools—enable clinicians and researchers to reliably assess the severity of amnesia symptoms and to diagnose the syndrome of dissociative amnesia.

DEFINITION

The fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) defines *dissociative amnesia* as “an inability to recall important personal information, usually of a traumatic or stressful nature, that is too extensive to be explained by normal forgetfulness ... and is not due to the direct physiological effects of a substance or a neurological or other general medical condition.” DSM-IV further distinguishes between five types of memory disturbances included under the heading of dissociative amnesia: (1) *localized amnesia*, in which “the individual fails to recall events that occurred during a circumscribed period of time, usually the first few hours following a profoundly disturbing event”; (2) *selective amnesia*, in which “the person can recall some, but not all, of the events during a circumscribed period of time”; (3) *generalized amnesia*, when “failure of recall encompasses the person's entire life”; (4) *continuous amnesia*, “the inability to recall events subsequent to a specific time up to and including the present; (5) *systematized amnesia*, the “loss of memory for certain categories of information, such as all memories relating to one's family or to a particular person.” The latter three types of dissociative amnesia are more rare. When dissociative amnesia occurs recurrently, clinicians should rule out the diagnosis of more complex dissociative disorder, such as dissociative identity disorder.

HISTORY

Paul Briquet and Jean-Martin Charcot were two of the first to record and comment on the frequency of amnesia in their classic studies of hysteria. Pierre Janet, however, was the true pioneer in the field of dissociation. Demonstrating a surprising similarity to our modern conceptions, Janet's discussion of amnesia is fundamental and remains one of the most extensive in the literature. Josef Breuer and Sigmund Freud also describe dissociative amnesia in their studies of hysterical patients; many were described as experiencing blackouts and extended amnesic gaps in memory.

Although there have been a number of studies and case reports of amnesia during the twentieth century—which have been extremely influential in the formation of modern notions and texts—many of these suffer from methodological weakness. Nonetheless, most support the classic concept which notes that most patients with dissociative amnesia come from massively stressful psychosocial environments characterized by feelings of shame, guilt, desperation, and conflict. During the 1940s, a group of clinicians became interested in dissociation and published several case reports of amnesia including a series of case studies on amnesia and other disturbances resulting from combat experiences in the World War II. Modern studies have begun to address the range of amnestic experiences and the relationship between amnesia and trauma in a more systematic fashion. Research based on the analysis of 141 clinical interviews utilizing Steinberg's *Structured Clinical Interview for DSM-IV Dissociative Disorders–Revised (SCID-D–R)* indicates that amnesia occurs across a spectrum, ranging from the occasional forgetfulness of normal controls, to occasional memory difficulties of patients with a variety of psychiatric disorders, to recurrent or persistent episodes of amnesia (lasting days or longer) in patients suffering from dissociative disorders.

COMPARATIVE NOSOLOGY

Little data has accumulated suggesting that the description or diagnostic criteria of dissociative amnesia be revised since the publication of the third edition of DSM (DSM-III) in 1980. However, one noteworthy change is the disorder's name, which was changed from *psychogenic amnesia* to *dissociative amnesia* in DSM-IV. Most other changes merely reflect attempts at clarification, emphasizing the sudden nature of the condition and its relationship to trauma, and instructing the clinician to rule out *dissociative identity disorder (multiple personality disorder)* before making a diagnosis of dissociative amnesia.

The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) defines dissociative amnesia in basically the same terms: “The main feature is loss of memory, usually of important recent events, that is not due in organic mental disorder and is too great to be explained by ordinary forgetfulness or fatigue (see Table 18.5-3). The amnesia is usually centered on traumatic events” although it does not differentiate between subsets of memory disturbance, as DSM-IV does (Table 18.1-1).

A. The predominant disturbance is one or more episodes of inability to recall important personal information, usually of a traumatic or stressful nature, that is too extensive to be explained by ordinary forgetfulness.

B. The disturbance does not occur exclusively as a symptom of dissociative identity disorder, dissociative fugue, post-traumatic stress disorder, acute stress disorder, or somatization disorder and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a neurological or other general medical condition (e.g., amnestic disorder due to head trauma).

C. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Table 18.1-1 DSM-IV Diagnostic Criteria for Dissociative Amnesia

EPIDEMIOLOGY

The epidemiology of dissociative amnesia has not been adequately examined but the few studies that have been conducted provide some data. The majority of studies, excluding those of veterans, indicated a relatively equal sex incidence. Similarly, although a wide range of ages have been reported, most studies report a peak incidence in the third and fourth decades. Although the frequency of occurrence awaits systematic research, researchers believe that dissociative amnesia may be more common than previously recognized. In modern clinical practice, dissociative amnesia most frequently presents in hospital emergency rooms, where amnesic patients are typically brought after being found on the street.

ETIOLOGY

According to the DSM-IV criteria B for dissociative amnesia, the disturbance is not due to a medical or neurological condition or to the effects of substance abuse. Thus the etiology of dissociative amnesia excludes all organic causes. Janet was one of the earliest researchers to emphasize the role of trauma in cognitive impairments. Although systematic research specific to the etiology of dissociative amnesia is limited, numerous subsequent investigators have noted a causative link between emotional trauma and dissociative amnesia. The amnestic episode is thought to be an intrapsychic defense, excluding painful memories from conscious awareness, and can result from a single overwhelming traumatic event or from a series of smaller precipitants. Commonly reported circumstances include childhood abuse, kidnapping, incest, rape, past wartime combat experience (“shell shock”), and other threats of death or physical violence, and even being a witness to violence. Intensity, duration, and age of exposure to the traumatic event seem to be critical factors in the development of dissociative amnesia; generally the stronger, longer, and earlier the exposure, the worse the amnesia. Research by the author reports that recurrent dissociative amnestic episodes often occur in individuals suffering from a range of other dissociative symptoms, and frequently occur in the most severe dissociative disorder, *dissociative identity disorder*.

The DSM-IV diagnostic criteria for dissociative amnesia are listed in Table 18.1-1.

DIAGNOSIS AND CLINICAL FEATURES

In order to systematically assess dissociative amnesia, clinicians may wish to use specialized diagnostic tools. Currently the only tool that assesses both the degree of amnesia and diagnoses dissociative amnesia is SCID-D-R. This semistructured interview evaluates the severity of five dissociative symptoms (amnesia, depersonalization, derealization, identity confusion, and identity alteration) and diagnoses the five dissociative disorders (dissociative amnesia, dissociative fugue, depersonalization disorder, dissociative identity disorder, and dissociative disorder not otherwise specified). The interview contains a total of 276 items (36 items relate to amnesia), with follow-up items if the examiner requires additional information. Particularly important in the evaluation of substance abusers, *SCID-D-R* enables the interviewer to rule out amnesia secondary to organic etiology and drugs or alcohol prior to diagnosing dissociative amnesia. Good to excellent reliability and discriminant validity have been reported for SCID-D-R in the United States and in cross-national reliability studies.

Dissociative amnesia is more likely to involve interruption of the episodic-autobiographical memory than the implicit-semantic memory. The memories unavailable for recall tend toward historical factual information (i.e., where was I; whom was I with; and what did I do, think, and feel during the unaccountable period of time?) rather than interruption of general cognitive functioning or language capacity. This period of amnesia usually centers around a traumatic event or series of events and is usually localized, occurring during a specific period of time lasting anywhere from a few hours to several years. Occasionally the amnesia may be selective or systematized, whereby it is restricted to certain memories such as those involving a particular individual. Other forms of amnesia also occur (*generalized amnesia*, when the amnesia extends over the patient's entire life, and *continuous amnesia*, when the amnesia extends from a specific time up to the present), but are much more rare and are often associated with more severe dissociative disorders.

Dissociative amnesia is one of the most difficult disorders to assess because it cannot be observed directly except in cases of global amnesia; patients rarely complain about amnesia itself. Clinically, the patient may present symptoms of anxiety, depression, confusion, difficulty concentrating, and a history of blank spells or gaps in memory. Even once amnesia is confirmed, clinicians typically find it difficult to obtain from patients reliable estimates of the frequency and extent of their amnestic episodes.

Clinicians should pay particular attention to patients' verbal responses and nonverbal behaviors because these may exhibit patterns that suggest amnesia. For example, research based on the administration of Steinberg's *SCID-D-R* notes that vagueness and inconsistency of response and difficulty in narration of past events may indicate amnesia. Other clues that suggest amnesia include a history of unexplained displacement, possession of unfamiliar objects, disavowal of confirmed actions, and any instance of identity loss. The following clinical descriptions, accompanied by actual patient quotations, should help to clarify some of the most prominent individual manifestations of dissociative amnesia.

Confabulation and Self-Monitoring Because amnesia can have a disastrous effect on a patient's day-to-day life, many people with chronic amnesia develop adaptive strategies. One such strategy is *confabulation*, the invention of false information to cover up a gap in memory. One woman with dissociative identity disorder spoke of the marital tension caused by her amnesia:

> “When Charlie and the kids were infants ... I would have to wonder if I fed them—wonder what I did with them—wonder where we went. Often, as I think of it, I couldn't figure out what we had done. When my husband would come home from work and he'd say, “Gee, what did he have for lunch? How did he eat? How many times did he eat?” I wouldn't be able to tell him ... After a while I began to lie.”

Other patients will resort to various forms of self-monitoring to protect themselves from memory loss, such as note-taking or the cessation of regular activities.

Distortions in Time Perception Episodes of amnesia are often characterized by the subject's misperception of elapsed time. The patient may have a distorted sense of the rapid passage of time from the last moment of fully remembered consciousness to the present. The following example is typical:

> “It's like I have specific things that flash on the screen, specific, I'm only talking about minutes, hours at the most. It's like my whole, from the time I was not quite 5 till not quite 13, it seems like my memory is in an hour's time. Then from the time I am like 14 through high school seems pretty ... I can remember all that. In between like from the time I'm through high school till I'm 21 that seems pretty clear. And from the time I am 21 forward there sometimes seems to be entire years like I arrived, I feel like I arrived at this age.”

Amnesia and Memories of Abuse Patients are often amnestic for traumatic experiences. Research investigations found that 59 percent and 64 percent of the study sample were amnestic for having been sexually abused prior to the age of 18. Independent corroboration was reported of 74 percent of subjects' reports of abuse. However, a person may also experience intrusive traumatic memories during stressful periods or following exposure to a sensory trigger. Sometimes a consciously unretrievable memory can cause anxiety without the patient's recollection of any historical connection, as in the following example:

> “I'm paranoid of handkerchiefs. I was scared to death ... I was afraid of a handkerchief. When I finally traced my mind and forced myself to remember and everything else, I remember why now. At, I don't know, I guess 4 years old, a babysitter used to gag me with a handkerchief, shove it down my throat, and rape me... . It's like after a while you learn to remember it. I don't want to remember it, but in order to help myself I knew I had to. God, I was afraid of the things I was learning, because it's like totally new. Your mind blanks it out. I didn't remember it and now I know why ... too painful.”

DIFFERENTIAL DIAGNOSIS

When evaluating a patient presenting for amnesia, clinicians must determine whether the amnesia is a result of a dissociative disorder, another psychiatric disorder, or any number of organic causes.

Amnestic Disorders Amnestic disorders are caused by a variety of organic conditions. Examples of organic causes of amnesia are epileptic seizures, to head trauma, alcoholic blackouts, Korsakoff's syndrome, stroke, postoperative amnesia, postinfectious amnesia, post-ECT, surgery, infection, electroconvulsive therapy (ECT), and transient global amnesia. Less common causes are cerebrovascular disease, metabolic abnormalities, and toxic states.

Amnesia that occurs in the context of an epileptic seizure is called *ictal amnesia* and is characterized by short duration and epileptiform electroencephalograph (EEG) results. In cases of ictal amnesia, there is often anterograde amnesia during the seizure and retrograde amnesia following the seizure. The individual retains a sense of identity and appears otherwise normal. A systematic investigation of amnesia and other dissociative symptoms performed in patients with epilepsy and pseudoseizures found that patients with epilepsy experienced significantly less amnesia, depersonalization, identity confusion, and alteration than patients with pseudoseizures when assessed by SCID-D-R. Patients with a history of nonictal amnesia should be evaluated further to rule out a concomitant dissociative disorder.

Patients who have had a head injury usually experience a brief period of retrograde amnesia and a longer period of anterograde amnesia. Blackouts during alcohol intoxication involve partial to complete amnesia for events occurring during the intoxicated state. In Korsakoff's syndrome, there is significant anterograde amnesia

and variable retrograde amnesia; other cognitive functions are usually left intact. In dementia, the memory loss occurs in combination with multiple cognitive deficits, and many different kinds of memory, including remote memory, are disturbed.

Organic amnesias have several distinguishing features: they do not normally involve recurrent identity alteration, the amnesia is not selectively limited to personal information, the memories do not focus on or result from an emotionally traumatic event, and the amnesia is more often anterograde than retrograde. In cases of amnesia of organic etiology (excluding substance abuse, transient global amnesia, or metabolic abnormalities), the amnesia is usually permanent and does not lend itself to therapeutic techniques. Whereas dissociative amnesia represents a displacement of the memory from awareness, organic amnesias represent the erasure or destruction of that memory through disturbance of the neuropsychological process. If the patient presents with amnesia and symptoms suggestive of an underlying medical disorder, a complete history of the patient's behavior during the amnesia should be obtained as well as a physical and neurological examination, including all necessary laboratory tests and an EEG. Once an organic cause has been ruled out the clinician can evaluate the possible causes of psychogenic amnesia. Use of diagnostic tools such as the SCID-D-R can allow assessment of the severity of the amnesia as well as other core dissociative symptoms, thus enabling the clinician to rule out the presence of dissociative disorder.

Amnesia in Psychiatric Disorders

Amnesia in Depression Amnesia occurs commonly in patients with moderate to severe depression, as well as in patients suffering from the dissociative disorders. Individuals suffering from major depressive disorder frequently experience cognitive impairments that may superficially mimic the organic dementias. Thus, individuals with severe depression have been referred to as having *pseudodementia*. However, amnesia associated with depression is characterized by “I don't knows” rather than near-misses or confabulations, and it is less severe than that in senescent dementia. Individuals suffering from dementia of depressive origin are also more likely to complain about their amnesia than those suffering from dementia caused by medical conditions.

Amnesia in Substance Use Disorders Amnestic experiences are common to patients with histories of alcohol abuse, which is not surprising given the frequency of alcohol-related blackouts. Investigation of amnesia in patients with substance abuse histories using the SCID-D-R found that amnesia by itself is not a diagnostically discriminating symptom in subjects with histories of alcohol abuse; clinicians must evaluate patients for all of the five core dissociative symptoms (amnesia, depersonalization, derealization, identity confusion, identity alteration) as assessed by the SCID-D-R. Substance abuse patients without coexisting dissociative disorders reported amnesia only in the context of substance use (i.e., alcoholic blackouts precipitated by drinking). The diagnosis of dissociative amnesia is excluded when substance abuse patients report that their amnesia occurs only in the context of alcohol consumption. In contrast, amnesia associated with dissociative disorders in substance abusers was characterized by recurrent amnestic episodes as well as other dissociative symptoms such as depersonalization or identity confusion, occuring without the use of drugs or alcohol.”

Amnesia in the Dissociative Disorders as Assessed by SCID-D-R In dissociative amnesia and in the other dissociative disorders amnesia is functional—it results from psychogenic rather than organic factors. In contrast to patients with substance abuse disorders, patients diagnosed with dissociative disorders reported memory gaps in the absence of alcohol intake when assessed with the SCID-D-R. Severe amnesia in patients with underlying dissociative disorders is often characterized by their inability to recall their age, name, address, or other essential personal information. If the memory gap is the predominant disturbance and results in dysfunction or distress, then dissociative amnesia should be ruled out. If the memory gaps occur in conjunction with other dissociative symptoms, the clinician should rule out one of the other dissociative disorders. SCID-D-R-based research indicates that recurrent memory gaps in conjunction with depersonalization, derealization, identity confusion, and identity alteration characterize the symptomatic profile of patients with dissociative identity disorder. Table 18.1-1 presents a summary of the spectrum of amnesia on the SCID-D-R; Figure 18.1-1 is a visual summary of the differential diagnosis of amnesia.

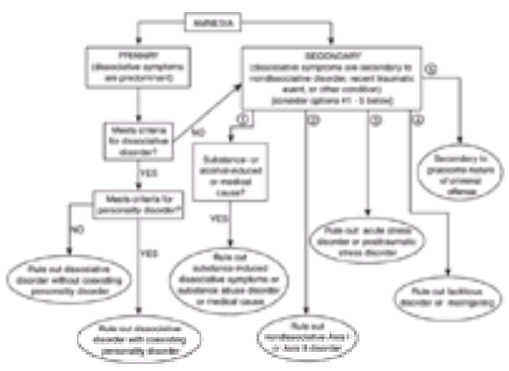

FIGURE 18.1-1 Differential diagnosis decision tree of amnesia. (Reprinted with permission from Steinberg M: *Handbook for the Assessment of Dissociation: A Clinical Guide*. American Psychiatric Press, Washington, DC, 1995.)

Dissociative identity disorder and Dissociative disorder not otherwise specified	Nondissociative and Personality Disorders	No Psychiatric Disorder
Gaps for hours or years	Rare memory gaps; forgetfulness of a minor nature	Minor forgetfulness associated with stress
Associated with fugues	No fugues	No fugues
Recurrent-persistent episodes	Isolated episodes	None—isolated episodes

Reprinted with permission from Steinberg M: *Interviewer's Guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised* (SCID-D-R). American Psychiatric Press, Washington, DC, 1994.

Table 18.1-2 The Spectrum of Amnesias on the SCID-D-R

Proper treatment of amnesia depends on accurate diagnosis, as in the following case study of a patient presenting with global amnesia. Diagnosis based on the administration of SCID-D-R further clarifies the constellation of symptoms associated with the amnesia, and thus allows for accurate diagnostic assessment. This case is presented in the form of a report suitable for inclusion in the patient's records, in order to demonstrate the clinical applications of the SCID-D-R for the documentation of patient's symptoms.

SCID-D-R Psychological Report Mr. Smith

Referral Source: Patient's psychiatrist

Brief summary: Mr. Smith was referred for a diagnostic evaluation due to global amnesia that developed immediately following an episode of syncope at his home in July of 1996. He is a 41-year-old owner of a small business in Chicago and lives with his wife and three children in Lake Forest, Illinois. Immediately following his fainting episode he was taken to a local emergency room where he received a comprehensive medical evaluation. Physical and neurological examinations were unremarkable, and all laboratory tests as well as the computed tomography (CT) scan were within normal limits. He was referred for an EEG with nasopharyngeal leads that was subsequently performed on an outpatient basis; this also was without abnormal findings. The patient had no history of substance abuse, is not a military veteran, and had no known medical problems other than an allergy to ragweed pollen.

Past Psychiatric History: The patient had no history of psychiatric disturbance, and is currently in outpatient treatment for the amnesia with Dr. Jones, a local psychiatrist. The patient is distressed about his inability to recover his memory, particularly because there are apparently no immediate stressors severe enough to account for either the fainting or the amnesia. He also acknowledged recent attacks of agoraphobia, including fear of driving farther than five miles from his house. His wife accompanied him to the interview because of his fear of driving and fear of the dark.

Mr. Smith's only psychiatric hospitalization was following his fainting episode and subsequent global amnesia. He was hospitalized for four weeks and treated with psychotherapy without medication. His discharge diagnosis was dissociative amnesia. Due to his global amnesia Mr. Smith was unable to provide information regarding his childhood or parents, stating that he could not remember the details of his childhood relationships. He was able to report that he is the oldest of six children, and that he felt that he had been punished by physical beatings by his father at unclear intervals of frequency. He completed 3 years of college, and has been married to his present wife since 1973. They have three daughters, aged 18, 15, and 11, respectively. He reports that the marriage has been stable with no separations. The oldest daughter attends college in Michigan, and the younger two reside at home. Mr. Smith indicated that his wife and all three children have been supportive throughout his hospitalization and his present inability to work.

SCID-D-R Evaluation Summary: The findings of Mr. Smith's SCID-D-R interview indicated a modification of his first diagnosis. His endorsement of severe retrograde amnesia is not inconsistent with a tentative diagnosis of dissociative amnesia; Mr. Smith reported a nearly total failure of recall for his life prior to his fainting, with the exception of a few major events such as his wedding. He did not suffer from anterograde amnesia, indicating at a number of points during the interview that he had no difficulty learning new material. However, Mr. Smith also reported the presence of other dissociative symptoms. Had the initial diagnosis of dissociative amnesia been correct, Mr. Smith's amnesia would have been the predominant symptom. However, he mentioned recurrent experiences of depersonalization that included feelings of sitting in two different chairs simultaneously, and out-of-body episodes in which he floated on the ceiling while watching his body on the floor. In terms of derealization, Mr. Smith remarked that his surroundings sometimes seemed “fuzzy” or unreal to him, and that he had frequent “trances” that sometimes lasted for an hour or two. In addition, he was unable to recognize several close relatives at a family funeral. The patient also described experiences suggestive of identity confusion and identity alteration; he reported that he saw “dark beings” resembling humans in his trance states, and that these beings had communicated with his wife and present therapist. On one occasion he walked two of these beings around his bedroom. He mentioned that he experienced an ongoing struggle between two different “sides” of himself, one of which wanted to retrieve his past and one that opposed attempts at recollection. It appeared that one of these sides represented a childlike part of Mr. Smith, in that he would curl up in a fetal position from time to time, act childlike, and state that he was afraid of the dark. The follow-up section of SCID-D-R on “internal dialogues” was administered to explore the extent of Mr. Smith's identity disturbance.

Mr. Smith's responses throughout the administration of the SCID-D-R indicated that his reality testing was excellent, in that he was clearly aware of the distinctions between the “unreal” or “strange” quality of his dissociative episodes and “ordinary” reality. His comment that the “dark beings” in his trance communicated “all kinds of crazy things” to his therapist was typical of his reality testing. During the mental status examination Mr. Smith spoke clearly and intelligently and answered all questions with relevant replies.

Discussion: Based on the SCID-D-R evaluation, Mr. Smith's suffers from severe amnesia, as well as severe depersonalization, derealization, identity confusion and alteration. Although Mr. Smith's different sides did not represent full-fledged alternate personalities, his symptoms of identity confusion and alteration were sufficiently severe to meet the criteria for a subcategory of dissociative disorder not otherwise specified, which includes “clinical presentations similar to Dissociative Identity Disorder that fail to meet full criteria for this disorder. Examples include presentation in which there are not two or more distinct personality states.”

Mr. Smith's case illustrates the importance of a comprehensive evaluation for a range of dissociative symptoms and symptom re-evaluation in patients presenting with dissociative amnesia. Later progress reports indicated that Mr. Smith's symptoms were responding satisfactorily to treatment aimed at reducing the frequency of his trance episodes, and at increasing the cooperation between the different sides of his personality.

COURSE AND PROGNOSIS

Individuals with dissociative amnesia typically experience sudden onset of amnesia, usually following severe psychosocial stressors. When amnesia is localized or selective, recovery is usually complete and termination may be rapid. In cases of generalized amnesia, recovery of memory is usually gradual. Functional impairment varies from mild to severe, depending on the extent of the amnesia. When the dissociative amnesia is recurrent, the clinician should rule out the presence of another more severe dissociative disorder, specifically dissociative identity disorder.

TREATMENT

The more acute and the more recent the instance of dissociative amnesia, the more likely and the more quickly it is to be resolved. However, the clinician should be careful because intrusive attempts to retrieve dissociated memories can result in retraumatization if the patient is not properly prepared. This risk is especially great for longstanding or childhood-onset amnesias. The clinician should control the pace of suggested recollection, usually within the framework of a broader psychotherapy gauged to resolve the complex of events producing the amnesia. In particularly extreme cases, hospitalization may be necessary. Once the patient has been stabilized and a comfortable therapeutic relationship has been established, the clinician may proceed with other measures.

Group Psychotherapy Group psychotherapy has been successful in helping combat veterans and survivors of childhood abuse to regain function. Group interventions may lead to integration of dissociated memories.

Hypnosis Hypnosis can help to contain and modulate the symptoms, to facilitate recall of dissociated memories, and to reintegrate the dissociated material.

Pharmacotherapy To date there have been no double-blind studies of medication for the treatment of dissociative amnesia. Double-blind trials are necessary to evaluate the relative efficacy of pharmacotherapeutic agents in the treatment of dissociative amnesia. The recent development of diagnostic tools such as the *SCID-D-R*, which allows for the assessment of the presence and severity of dissociative amnesia based on operational definitions of symptom severity, can allow for systematic research for the treatment of dissociative amnesia.

Drug-Assisted Interviews Although there have been no reliable studies assessing their efficacy, case reports of drug-facilitated interviews have been reported to have positive results in some patients with especially intransigent dissociations. Agents used include sodium amytal and oral benzodiazepines. Such drug-facilitated interviews should be administered only by mental health professionals experienced in the treatment of individuals with dissociative disorders.

Amnesia occurs in a variety of psychiatric, medical, and substance-related disorders. Comprehensive assessment of the individual's amnesia and its nature, severity, and context can distinguish amnesia in cases of dissociative disorder from amnesia in other nondissociative disorders. Research based on SCID-D-R indicates that amnesia occurs across a spectrum, ranging from the occasional forgetfulness of normal controls to the occasional memory difficulties of patients with a variety of psychiatric disorders, to recurrent or persistent episodes of amnesia (lasting days or longer) in patients suffering from dissociative disorders. Future research is necessary to evaluate the relative efficacy of psychotherapy and pharmacotherapeutic agents in the treatment of dissociative amnesia.

SUGGESTED CROSS-REFERENCES

Dissociative identity disorder is discussed in Section 18.3, and dissociative disorder not otherwise specified and the ICD-10 classification for dissociative disorders are covered in Section 18.5. Alcohol-related disorders are described in Section 11.2. Amnestic disorders, other cognitive disorders, and mental disorders due to a general medical condition are discussed in Chapter 10. Neuropsychiatric aspects of head trauma is presented in Section 2.5.

SECTION REFERENCES

Bowman E: The differential diagnosis of epilepsy, pseudoseizures, dissociative identity disorder and dissociative disorder not otherwise specified. Psychiatr Ann (in press)

Briere J, Conte J: Self-reported amnesia for abuse in adults molested as children. J Trauma Stress *6:*21, 1993.

*Coons PM, Milstein V: Psychogenic amnesia: A clinical investigation of 25 cases. Dissociation *5:*73, 1992.

Croft PB, Heathfield KWG, Swash M: Differential diagnosis of transient amnesia. Br Med J *4:*593, 1973.

Herman J, Schatzow E: Recovery and verification of memories of childhood sexual trauma. Psychoanal Q *4:*1, 1987.

Herman JL: *Trauma and Recovery*. Basic Books, New York, 1992.

Kopelman MD: Amnesia: Organic and psychogenic. Br J Psychiatry *150:*428, 1987.

Kasniak AW, Nussbaum PD, Berren MR, Santiago J: Amnesia as a consequence of male rape: A case report. J Abnorm Psychol *97:*100, 1988.

Kluft RP: The dissociative disorders. In *The American Psychiatric Press Textbook of Psychiatry*, J Talbott, R Hales, S Yudofsky, editors. American Psychiatric Press, Washington, DC, 1988.

La Rue A: Memory loss and aging: Distinguishing dementia from benign senescent forgetfulness and depressive pseudomentia. Psychiatr Clin North Am *5:*89, 1982.

Lowenstein RJ: Psychogenic amnesia and psychogenic fugue: A comprehensive review. In *American Psychiatric Press Annual Review of Psychiatry*, vol 10, A Tasman, S Goldfinger, editors. American Psychiatric Press, Washington, DC, 1991.

*Pope HG Jr, Oliva PS, Hudson JI, Bodkin JA, Gruber AJ: Attitudes toward DSM-IV dissociative disorders diagnoses among board-certified American psychiatrists. Am J Psychiatry *156:*321, 1999.

Ruedrich SL, Chu C-C, Wadle CV: The Amytal interview in the treatment of psychogenic amnesia. Hosp Community Psychiatry *36:*1045, 1985.

Schacter DL, Kihlstrom JF: Functional amnesia. In *Handbook of Neuropsychology*, F Boller, J Grafman editors. Elsevier Science, New York, 1989.

Spiegel D: Dissociation and hypnosis in posttraumatic stress disorders. J Trauma Stress *1:*17, 1988.

Steinberg M, Rounsaville B, Cicchetti D: The structured clinical interview for DSM-III-R dissociative disorders: Preliminary report on a new diagnostic instrument. Am J Psychiatry *147:*76, 1990.

*Steinberg M: *Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised (SCID-D-R)*. American Psychiatric Press, Washington, DC, 1994.

*Steinberg M: *Interviewer's Guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised (SCID-D-R)*. American Psychiatric Press, Washington, DC, 1994.

Steinberg M: Systematizing dissociation: Symptomatology and diagnostic assessment. In *Dissociation: Culture, Mind and Body*, D Spiegel, editor. American Psychiatric Press, Washington, DC, 1994.

*Steinberg M: *Handbook for the Assessment of Dissociation: A Clinical Guide*. American Psychiatric Press, Washington, DC, 1995.

Steinberg M: *A Clinician's Guide to Diagnosing Dissociative Symptoms and Disorders: The Structured Clinical Interview for DSM-IV Dissociative Disorders (SCID-D)*. Health Systems, Toronto, 1996.

*Steinberg M: *Tips and Techniques for Assessing and Planning Treatment with Dissociative Disorder Patients: A Practical Guide to the SCID-D*. Multi-Health Systems, Toronto, 1996.

Steinberg M, Hall P: The SCID-D diagnostic interview and treatment planning in dissociative disorders. Bull Menninger Clin *61:*1, 1996.

van der Hart O, Friedman B: A reader's guide to Pierre Janet on dissociation: A neglected intellectual heritage. Dissociation *2:*3, 1989.

CHAPTER 18. DISSOCIATIVE DISORDERS

18.2 DISSOCIATIVE FUGUE

PHILIP M. COONS, M.D.

Dissociative fugue (formerly called psychogenic fugue) is an unusual but dramatic type of dissociative disorder. Owing to its rarity, dissociative fugue has been the least studied of the dissociative disorders in recent years.

DEFINITION

In general, a *fugue* is characterized by amnesia for one's identity, which is coupled with sudden, unexpected travel. Both psychological and organic forms of fugue exist.

HISTORY

Although fugues have undoubtedly occurred in individuals for a long time, fugues were not described in detail until the nineteenth century when Freud and his contemporaries, Jean-Martin Charcot and Pierre Janet, began to take an interest in hypnosis and various hysterical phenomena.

Perhaps the most famous case of fugue was the Reverend Ansel Bourne. Bourne was an itinerant preacher, who in January 1887 withdrew over $500 from his bank account in Providence, Rhode Island, paid some bills, and then disappeared. Two months later Rev. Bourne “awoke” in Norristown, Pennsylvania. He apparently had forgotten his original identity and had assumed a new identity, that of A.J. Brown, a shopkeeper. After returning to Providence, Bourne was completely amnesic for his 2-month absence. During his fugue state he was reported to have behaved completely normally by the Norristown residents. When examined under hypnosis 13 years later, it was possible to talk to A.J. Brown and recover the memories of the fugue state, but it was not possible to integrate Brown's memory or identity into that of Reverend Bourne. So far as is known, the identities of Bourne and Brown did not alternate as in dissociative identity disorder.

COMPARATIVE NOSOLOGY

Dissociative fugue was formerly called *psychogenic fugue* in the revised third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R). Diagnostic criteria at the time required the assumption of a new identity, either partial or complete. When during the literature search phase for the fourth edition of (DSM-IV) it became obvious that only a minority of historical cases were associated with new identity formation, this criterion was relaxed to allow either confusion about identity or the assumption of a new identity. The 10th revision of *International Statistical Classification of Disease and Related Health Problem* (ICD-10) has not yet adopted this relaxed standard and provides no place to classify psychological fugue states that occur without new identity formation (Table 18.5-3).

EPIDEMIOLOGY

Because of its rarity, very little is known about the epidemiology of dissociative fugue. In a study done on psychiatric inpatient admissions at the Dayton, Ohio Wright Patterson Air Force Medical Center in 1973, dissociative fugues were found in only 0.3 percent of active duty servicemen and female dependents. In a more recent study done in Winnipeg, Canada, no cases of dissociative fugue were found in 502 members of the general population.

Fugues are known to increase in prevalence during times of extremely stressful events, such as wars or natural disasters. The few studies, which have been done previously and which contained relatively large numbers, consisted of collections of cases occurring over a number of years such as during World War II. Unfortunately, these studies were confounded by the different diagnostic conventions of the time and were in many instances gender biased in that they contained mostly men. When dissociative fugues occur, they can be quite dramatic and are frequently the subject of exaggerated accounts in the popular media.

Although most dissociative fugues occur in adults, primarily in the second through fourth decades of life, dissociative fugues have been described in adolescents and in children as young as age 7. The rarity of dissociative fugue in the younger age group is probably related to the inability of most children and adolescents to drive or arrange for independent transportation. It has been speculated that some running-away behavior in children or adolescents may occur during dissociative fugue states but this has never been rigorously studied.

Although clinical lore teaches that dissociative fugue is more frequent in men, this has never been proven. Dissociative fugues may occur in either gender and in one study were approximately equal in gender distribution. Other studies, done in a military setting, have shown a gender bias towards subjects.

ETIOLOGY

Dissociative fugue is thought to be related either to traumatic or overwhelmingly stressful life events and thus may begin after exposure to natural disasters or military combat. Dissociative fugue may also be related to overwhelming stress such as a bankruptcy or an impending divorce. In many cases dissociative fugue is related to an avoidance of responsibility concerning legal or financial matters, sexual indiscretions, or fear of combat. Many individuals who suffer from dissociative fugue have histories of childhood abuse or neglect, although this finding has not been rigorously studied. In some cases the dissociative fugue can be understood as an amnesic state during which forbidden wishes can be symbolically expressed. In other dissociative fugue states, the amnesia may protect against a forbidden wish, such as suicide. Frequently, an underlying dysphoria or depression is present with accompanying shame or guilt. In most cases the dissociative fugue appears to be a symbolic running away from stressful situations. As in all of the other dissociative disorders, the process of dissociation appears to play a central role and results in the symptoms of amnesia and identity change.

DIAGNOSIS AND CLINICAL FEATURES

A baseline evaluation of a fugue state should include a thorough clinical psychiatric history. The clinician should pay particular attention to obtaining history about previous childhood and adult trauma, especially child abuse, financial and marital difficulties, depression, substance abuse, legal difficulties, and sexual indiscretions. A thorough mental status examination should pay particular attention to cognition, memory, and orientation. Because fugues may have either a psychological or organic basis, a thorough physical and neurological examination is indicated in all but the most obvious cases.

Increasingly, the initial evaluation of dissociative disorders involves the use of screening instruments or structured clinical interviews. An excellent instrument for screening for dissociative symptoms is the Dissociative Experiences Scale (DES), a 28-item patient questionnaire that assesses symptoms of amnesia, absorption, depersonalization, derealization, and identity alteration. Patients with dissociative fugue would be expected to elevate only amnesic items and possibly identity alteration items on this easily administered test. The Structured Clinical Interview for DSM-IV Dissociative Disorders (SCID-D) is a clinician-administered semi-structured clinical interview that assesses the dissociative symptoms of amnesia, fugue, depersonalization, derealization, identity confusion, and identity alteration. The SCID-D takes longer to administer, 60 to 90 minutes compared to 10 minutes for the DES. Individuals with dissociative fugue typically endorse only

items on amnesia and some times identity alteration and identity confusion.

The DSM-IV diagnostic criteria for the psychological type of fugue, *dissociative fugue*, are listed in Table 18.2-1.

A. The predominant disturbance is sudden, unexpected travel away from home or one's customary place of work, with inability to recall one's past.
B. Confusion about personal identity or assumption of a new identity (partial or complete).
C. The disturbance does not occur exclusively during the course of dissociative identity disorder and is not due to the direct physiological effects of a substance (e.g., a drug of abuse a medication) or a general medical condition (e.g., temporal lobe epilepsy).
D. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 18.2-1 DSM-IV Diagnostic Criteria for Dissociative Fugue

Patients with dissociative fugue suddenly travel away from their home or usual place of work. During such travel they are amnesic for their identities (e.g., autobiographical memory), but their procedural memories are preserved. For example, they know how to operate a motor vehicle or feed themselves. During short dissociative fugues the individual may appear somewhat confused. However, if the dissociative fugue is prolonged, the individual will usually take on a new identity, either partial or complete. Unlike dissociative identity disorder, however, there is no alternation of personality states.

Dissociative fugues may last from less than an hour to a year or more, but generally last in the range of days to weeks. During such time the individual may travel great distances, even across continents or oceans, but generally the distance traveled is anywhere from 10 miles to several hundred miles, which may take the person across state lines.

During most dissociative fugues individuals do not appear to be confused or bizarre and do not attract attention. However, in some cases when the dissociative fugue ceases, they may still be amnesic, are lost, and present themselves to the police or local emergency rooms for help. Such cases often draw dramatic media attention in order to help discover the identities of the individuals involved. Usually, however, when dissociative fugues cease, individuals resume their old identity and are amnesic for events that occurred during the fugue. Even in these situations significant others may be concerned about their loved one's absence and amnesia and will bring the individual for medical attention.

The patient was a 45-year-old man, an attorney, who was in the midst of bankruptcy proceedings and a deteriorating marriage. One day he suddenly disappeared from his office, much to the consternation of his law partners. Fearing foul play, his wife notified the police and an investigation was begun in an attempt to trace his whereabouts. A day later his car was found abandoned and out of gas at an interstate highway rest stop. Nothing was heard from the gentleman for almost a month when he presented himself to a large inner city hospital emergency room several states away from his home. He remained amnesic for his identity and was admitted to the psychiatric unit. His picture and accompanying story was picked up by the wire services and his wife recognized him on television. She was eventually summoned to the hospital where her husband was staying and in a collateral interview provided a full history of her husband to the attending psychiatrist. The patient was confronted with this information but remained amnesic until hypnosis was used to return him to his original personality state. Upon resuming this state, he was transported home and was hospitalized locally by his family physician. He continued to be amnesic for his fugue state. In consultation with a neurologist, a full physical and neurological examination was performed, the results of which were normal. An electroencephalogram and magnetic resonance imaging were performed, neither of which was abnormal. Referral was then made to a psychologist who began psychodynamic psychotherapy. The precipitants, including the patient's failing marriage and poor financial status, were explored, and marital therapy was began as an adjunct. Eventually his memory was restored. However, his financial problems remained unresolved and he left on a shorter 5-day fugue but found his own way home. His treatment continued as the problems in his professional life, including gambling and misappropriation of funds from his legal clients, were explored.

PATHOLOGY AND LABORATORY EXAMINATIONS

Baseline laboratory screening should include an electrocardiogram, blood chemistry screening examination, drug screening, and a blood alcohol level. Further testing may be required and might include an electroencephalogram, toxicology screening test, computed tomography or magnetic resonance imaging of the brain, or neuropsychological testing.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of fugue states is vast. Genuine dissociative fugue must be excluded as a diagnosis if dissociative identity disorder is present or if the fugue is caused by the direct physiological effects of drugs, medications, or alcohol; it cannot be diagnosed if it is due to a general medical condition such as epilepsy.

Probably the most common organic fugue is secondary to epilepsy, especially complex partial seizure disorder. During such seizures or postictally the individual may exhibit wandering behaviors. The clinical differentiation is usually fairly easily accomplished by a good clinical history regarding epileptic symptoms and electroencephalographic studies.

Several medical conditions besides epilepsy can cause organic fugue. These conditions include brain tumor, head trauma, migraine, cerebrovascular accidents, hypertensive neuropathy, limbic system dysfunction, hypoglycemia, uremia, dementia, and malaria.

Organic fugue states may be caused by a wide variety of medications including hallucinogenic drugs, steroids, barbiturates, phenothiazines, triazolam (Halcion), and L-asparaginase. The alcohol blackout can be easily confused with dissociative fugue, but this can be easily differentiated through a good clinical history and alcohol concentrations, if drawn during acute intoxication. The clinician should remember, however, that dissociative fugue and alcohol blackouts may coexist in the same individual.

Dissociative fugue states have been reported to exist cormorbidly with a number of other nondissociative psychiatric diagnoses. These conditions include bipolar I disorder, schizophrenia, various depressive disorders, and alcohol use disorders.

Perhaps the most difficult differential is between dissociative fugue and dissociative identity disorder. At least 50 percent of all patients with dissociative identity disorder experience symptoms of fugue. These fugues are generally but not always brief in duration and involve only short distances of travel. Clinical history and use of the DES and SCID-D should differentiate between the two conditions because patients with dissociative identity disorder experience all types of dissociative symptoms including amnesia, fugue, depersonalization, derealization, identity confusion, and identity alteration.

There are a number of culture-bound psychiatric syndromes in which fugue is a prominent feature. These syndromes include the “running” syndromes, which include *latah* and *amok* that occur in several nations along the western Pacific rim; *grisi siknis* occurring among the Miskito of Nicaragua and Honduras, and *piblokto* (Arctic hysteria) occurring among the Eskimos of northern Greenland. These syndromes are characterized by a high level of agitation, running about, trance-like states, and amnesia for the episode.

Finally, malingering should always be high on the clinician's index of suspicion, especially in a forensic context of presentation. Malingering may be very difficult to differentiate from genuine dissociative fugue, however, because it may be associated with the same stressors as dissociative fugue (e.g., combat, financial problems,

or marital indiscretions).

Hypnosis or amobarbital (Amytal) interviews may be useful in helping the clinician make the differential diagnosis between dissociative fugue and these other conditions. In many cases of genuine dissociative fugue, memories for the events that occurred during the fugue can be elicited; however, the same is true of dissociative identity disorder. The patient with schizophrenia would be expected to exhibit even more trouble thinking and with looser associations while under the effects of amobarbital. It may be impossible to recover memory in fugues caused by physical conditions, but this may be true of dissociative fugue as well. In one study of patients who had an organic etiology of fugue, 41 percent had memory return with sodium amobarbital. Sometimes malingerers vehemently object to either technique. The clinician must keep in mind, however, that the malingerer may consciously deceive the examiner during the use of either hypnosis or sodium amobarbital. The clinician is also cautioned to use a nonsuggestive interviewing style to avoid the possibility of inducing pseudomemories.

COURSE AND PROGNOSIS

Most dissociative fugues are typically short, lasting from a few days to a few weeks. In many of these cases, recovery is spontaneous. In a only a few cases are symptoms refractory to treatment. Dissociative fugues may be recurrent. In one study, 65 percent of those experiencing dissociative fugue had more than one fugue.

TREATMENT

The mainstay of treatment of dissociative fugue is psychodynamic psychotherapy. A gently exploratory and expressive form is preferred in most cases, although a largely supportive form will usually suffice for those of low ego strength. The clinician should begin with a thorough clinical history and pay close attention to possible precipitating events. In many cases, encouraging persons with dissociative fugue to talk about what they already remember will bring the return of other memories; in some cases free association has proven helpful.

In situations where acute traumatic events have precipitated the dissociative fugue, a gentle abreaction of the trauma is indicated. However, the clinician should be very careful to not proceed with abreactive work until a stable therapeutic alliance has been established. In addition, abreactive work should be suspended, at least temporarily, if the patient's condition worsens (e.g., if the patient becomes depressed or suicidal).

If the individual with dissociative fugue continues to be densely amnesic for identity and autobiographical memory, the use of hypnosis or sodium amobarbital interviewing may be tried cautiously, keeping in mind that the dissociative fugue serves a defensive purpose, and that if the amnesia is suddenly lifted the individual may become depressed or even suicidal. Also, after the hypnotic session is completed, the amnesia may recur. Informed consent should be obtained whenever hypnosis or sodium amobarbital is used.

After the amnesia has been lifted, continued psychotherapy is indicated to help the individual cope with the underlying psychological conflicts that initially caused the dissociative fugue. Ideally, the patient should be helped to integrate the memories of the dissociative fugue state into a cohesive self and memory.

SUGGESTED CROSS-REFERENCES

Dissociative mechanisms are also discussed in Chapter 6 on psychoanalytic theory. Dissociative amnesia is discussed in Section 18.1. Psychotherapies are covered in Chapter 30. Dissociative disorders in children and adolescents are discussed in Section 49.8.

SECTION REFERENCES

Abeles M, Schilder P: Psychogenic loss of personal identity: Amnesia. Arch Neurol Psychiatry *34:*587, 1935.

*Akhtar S, Brenner I: Differential diagnosis of fugue-like states. J Clin Psychiatry *40:*381, 1979.

Bernstein E, Putnam FW: Development, reliability, and validity of a dissociation scale. J Nerv Ment Dis *174:*727, 1986.

Berrington WP, Liddell DW, Foulds GA: A reevaluation of the fugue. J Ment Sci *102:*280, 1956.

Brown D, Scheflin AW, Hammond DC: *Memory, Trauma Treatment and the Law*. WW Norton, New York, 1998.

Coons PM: The dissociative disorders: Rarely considered and underdiagnosed. Psychiatr Clin North Am *21:*637, 1998.

*Coons PM: Psychogenic or dissociative fugue: A clinical investigation of five cases. Psychol Reports *84:*881, 1999.

Ellenberger HF: *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*. Basic Books, New York, 1970.

Fisher C: Amnesic states in war neuroses: The psychogenesis of fugues. Psychoanal Q *14:*437, 1945.

Fisher C: The psychogenesis of fugue states. Am J Psychother *1:*211, 1947.

Fisher C, Joseph ED: Fugue with loss of personal identity. Psychoanal Q *18:*480, 1949.

*Ford CV: Psychogenic fugue. In *Treatments of Psychiatric Disorders*, vol 3, TB Karasu, editor. American Psychiatric Press, Washington, DC, 1989.

Gifford S, Murawski B, Kline NS, Sachar EJ: An unusual adverse reaction to self-medication with prednisone: An irrational crime during a fugue state. Int J Psychiatry Med *7:*97, 1977.

Good MI: Substance-induced dissociative disorders and psychiatric nosology. J Clin Psychopharmacol *9:*88, 1989.

Kanzer M: Amnesia: A statistical study. Am J Psychiatry *96:*711, 1939.

Kapur N: Amnesia in relation to fugue states—Distinguishing a neurological from a psychogenic basis. Br J Psychiatry *159:*872, 1991.

Kenny MG: *The Passion of Ansel Bourne: Multiple Personality in American Culture*. Smithsonian Institution Press, Washington, DC, 1986.

Kopelman MD, Panayiotopoulos CP, Lewis P: Transient epileptic amnesia differentiated from psychogenic “fugue”: Neuropsychological, EEG, and PET findings. J Neurol Neurosurg Psychiatry *57:*1002, 1994.

*Loewenstein RJ: Psychogenic amnesia and psychogenic fugue: A comprehensive review. In *American Psychiatric Press Review of Psychiatry*, vol 10, A Tasman, SM Goldfinger, editors. American Psychiatric Press, Washington, DC, 1991.

MacHovec FJ: Hypnosis to facilitate recall in psychogenic amnesia and fugue states: Treatment variables. Am J Clin Hypn *24:*7, 1981.

Mohan KJ, Salo WW, Nagaswami S: A case of limbic system dysfunction with hypersexuality and fugue state. Dis Nerv Syst *36:*621, 1975.

*Pope KS, Brown L: *Recovered Memories and Abuse: Assessment, Therapy, and Forensics*. American Psychological Association, Washington, DC, 1996.

Ross CA: Epidemiology of multiple personality and dissociation. Psychiatr Clin North Am *14:*503, 1991.

Rowan AJ, Rosenbaum DH: Ictal amnesia and fugue states. Adv Neurol *55:*357, 1991.

Rubinsky EW, Brandt J: Amnesia and criminal law: A review. Behav Sci Law *4:*27, 1986.

Ruedrich SL, Chu C, Wadle CV: The amytal interview in the treatment of psychogenic amnesia. Hosp Community Psychiatry *36:*1045, 1985.

*Steinberg M: *Structured Clinical Interview for DSM-IV Dissociative Disorders*. American Psychiatric Press, Washington, DC, 1993.

Stengel E: On the aetiology of fugue states. J Ment Sci *87:*572, 1941.

Stengel E: Further studies on pathological wandering (fugues with the impulse to wander). J Ment Sci *89:*224, 1943.

Takahashi Y: Aokigahara—jukai: Suicide and amnesia in Mt. Fuji's Black Forest. Suicide Life Threat Behav *18:*164, 1988.

Van der Hart O: Metaphoric and symbolic imagery in the hypnotic treatment of an urge to wander: A case report. Aust J Clin Exp Hypn *13:*83, 1985.

Venn J: Family etiology and remission in a case of psychogenic fugue. Fam Proc *23:*429, 1984.

*Zlotlow M: Temporal lobe “spike focus” associated with confusion, complete amnesia, and fugues in a paranoid schizophrenic. Psychiatr Q *42:*138, 1988.

18.3 DISSOCIATIVE IDENTITY DISORDER

FRANK W. PUTNAM, M.D., AND RICHARD J. LOEWENSTEIN, M.D.

Interest in dissociative identity disorder (previously called *multiple personality disorder*) has increased in recent years as a result of improved screening and diagnostic tools and a greater awareness of the dissociative sequelae of trauma, especially childhood maltreatment. The cardinal feature of dissociative identity disorder is the existence of two or more distinct personality states that exchange executive control of the individual's behavior. In addition, almost all such patients report dissociative amnesia, a loss of memory for personal history that cannot be explained by ordinary forgetfulness or organic memory impairment.

Debate about the existence of dissociative identity disorder has waxed and waned for more than a century. Critics usually raise the possibility of an iatrogenic cause, although no empirical studies in clinical populations support this notion. On the other hand, sufficient data on pathological dissociation in general and on this disorder in particular have accrued to judge dissociative disorders by the same standards that are applied to the validity of other psychiatric diagnoses.

The most widely accepted standards are based on a set of criteria first articulated by Eli Robins and Samuel Guze and subsequently refined by others. In essence, a psychiatric diagnosis is considered valid if it satisfies three basic requirements: content validity, criterion-related validity, and construct validity. Content validity requires a detailed clinical description of the disorder that is repeatedly independently replicated. More than a dozen clinical phenomenological studies of dissociative identity disorder including those from North America, South America, Europe, Turkey, and Asia, document the presence of a core dissociative psychopathology in these patients that fulfills this requirement. Criterion-related validity requires that laboratory tests or reliable psychological tests are consistent with the defined clinical picture. This stipulation is met by the reliable and valid diagnostic interviews and scales that have increasingly been employed in dissociation research as well as psychological testing protocols that discriminate these patients from normal controls, and subjects with schizophrenia and major depressive disorder, among others.

Construct or discriminant validity requires that the disorder be differentiable from other disorders. This has been empirically demonstrated for dissociative identity disorder with respect to disorders such schizophrenia, borderline and other personality disorders, and mood disorders. Also, patients with dissociative identity disorder can be discriminated from normal individuals and from other psychiatric patient groups, including those with personality disorders, posttraumatic stress disorder, and dissociative disorder not otherwise specified by structured interviews such as the Structured Clinical Interview for DSM-IV Dissociative Disorders. Psychological test batteries and experimental cognitive and psychophysiological studies have discriminated patients with dissociative identity disorder from other subjects including simulators. Thus, this disorder satisfies widely accepted standards for psychiatric validity and should be regarded as a legitimate disorder requiring an informed diagnostic and treatment approach.

HISTORY

Individuals with a multiplicity of selves have been described in Greco-Roman mythology, the Bible, and other writings prior to the modern medical era. Writing in 1646, Paracelsus is credited with the first medical report of an individual with alternating selves. In 1791 Eberhardt Gmelin described a German woman who alternatingly “exchanged” her peasant personality for that of an aristocratic French lady, each amnesic for the other's existence. By the early nineteenth century, such cases, with diagnoses of dual, double, or duplex consciousness, were being regularly reported on both sides of the Atlantic. In 1811 Benjamin Rush, father of American psychiatry, included a classic example in his medical school lectures. Rush proposed that dual personality reflected a functional disconnection between the two cerebral hemispheres. In the same year the case of Mary Reynolds, who became the American archetype of dissociative identity disorder for the remainder of the century, was first published. European cases such as Despine's “Estelle,” a child with this disorder, and Azam's “Felida” drew widespread public as well as professional attention as the nineteenth century advanced, just as these cases continue to do.

By the beginning of the twentieth century, Pierre Janet in France, Morton Prince and William James in the United States, and others were engaged in a lively transatlantic discussion about possible psychological and neurological mechanisms underlying cases of multiple personality disorder. Medical models of the time invoked phenomena such as sleep and dreams, hypnosis and somnambulism, epilepsy, and disconnections between the cerebral hemispheres to explain this condition. Pierre Janet's research and clinical theory, in particular his emphasis on the role of traumatic antecedents of dissociation, is widely regarded as the foundation for modern views of dissociation. Shortly thereafter, however, interest in dissociation waned in the face of the ascendancy of psychoanalytic theory as well as a variety of other historical forces in Western psychiatry and psychology. Clinical interest in dissociative phenomena has recurred in every war since the turn of the century, however, with the observation of amnesia, fugues, and conversion symptoms in traumatized soldiers. Subsequently, these observations tended to be forgotten by the psychiatric community as the wars themselves faded from memory.

A number of cases of dissociative identity disorder, including the famous “Three Faces of Eve,” continued to be published over the next decades. However, the modern era in the study of multiple personality disorder begins with the work of Arnold Ludwig and colleagues at the University of Kentucky during the 1970s. Cornelia Wilbur, widely identified with the case of ‘Sybil,' was influential in describing the clinical features of multiple personality disorder and identifying the role of childhood trauma as a factor in the etiology of the disorder. Wilbur, Richard Kluft, and others began the articulation of a modern treatment approach.

Since the 1980s the study of multiple personality and pathological dissociation has moved beyond single case reports to systematic clinical description of large groups of patients and empirical studies using standardized measures. The increased attention has led to more cases being diagnosed, which in turn has resurrected old controversies about the nature of this disorder. These include the possibility of iatrogenesis in its creation, the actual prevalence of dissociative disorders, the relationship of traumatic experiences and dissociation, and the criminal responsibility of the individual with the disorder.

EPIDEMIOLOGY

Few systematic epidemiological data exist for dissociative identity disorder. One study yielded a prevalence rate of 3.1 percent for a stratified sample ($N = 1055$) of the general population of Winnipeg, Canada, although a more conservative analysis of these data suggests a prevalence of about 1.3 percent. Independent analysis of the Dissociative Experiences Scale (DES) data collected on the same sample found a prevalence rate of 3.3 percent for pathological dissociation. Several studies examine the prevalence rate of dissociative identity disorder in general psychiatric patient samples. Results from the United States, Canada, Turkey, and several western European countries suggest that between 6 to 20 percent of adult inpatients may have the disorder. Available epidemiological data are insufficient and a large-scale, population-based study is necessary to resolve controversies about its prevalence.

Clinicians have long noted gender differences in the frequency of dissociative identity disorder. Clinical studies report between 5:1 to 9:1 female to male ratios for diagnosed cases. Research with dissociation measures such as the DES, however, find no evidence of gender differences in the propensity or capacity to dissociate. Developmental studies indicate that the ratio of female to male dissociative identity disorder cases steadily increases from 1:1 in early childhood to about 8:1 by late adolescence. Reasons proposed for the increased numbers of female patients relative to males include sex-related differences in the types, age of onset, and duration of maltreatment experienced by males and females; differences in clinical presentations such that male cases are more likely to be missed; and the possibility that more male cases end up in the criminal justice system or alcohol and other substance treatment programs rather than the general mental health system.

ETIOLOGY

Pathological levels of dissociation are strongly associated with histories of antecedent trauma. This has proven true across many types of trauma (e.g., rape, combat, natural disasters, and child maltreatment) in clinical and nonclinical samples. In some, but not all, studies indices of the severity of the trauma (e.g., age of onset and number of perpetrators for child abuse or intensity and duration of combat) are significantly related to the degree of dissociation. Studies of *peritraumatic dissociation*, dissociation manifest in the immediate context of trauma, consistently find that it is the best predictor of subsequent posttraumatic stress disorder 6 or more months later. Multiple lines of research strongly associate increased levels of dissociation with antecedent trauma for many different kinds of trauma.

Dissociative identity disorder is strongly linked to severe experiences of early childhood trauma, usually maltreatment, in all studies in both Western and non-Western cultures that have systematically examined this question. The rates of reported severe childhood trauma for both child and adult patients ranges from 85 to 97 percent of cases. Physical and sexual abuse, usually in combination, are the most frequently reported sources of childhood trauma in clinical research studies. Critics have raised questions about the validity of patients' self-reports of childhood trauma. Recent studies that now include rigorous independent corroboration of the patients' reports of maltreatment continue to strongly support a developmental linkage between childhood trauma and dissociative identity disorder.

Early life experiences resulting in disturbances in attachment relationship with the primary caregiver and other abnormal family processes have been implicated in the genesis of pathological levels of dissociation and the development of dissociative identity disorder. Recent research indicates that high levels of dissociation in mothers is associated with disturbed, often dissociative-like, attachment behavior in their children. In another study early presence of these attachment disturbances prospectively predicted higher levels of dissociation in late adolescence. The contribution of genetic factors is only now being systematically assessed, but preliminary studies have not found evidence of a significant genetic contribution.

Memory Disturbances Functional (i.e., nonorganic) amnesias are a hallmark of dissociative psychopathology. Researchers have sought to examine these for the past century. The total number of patients intensively studied is small, but a number of common findings are emerging. Research designs typically investigate differential encoding and retrieval of learned information across the alter personality states of patients with dissociative identity disorder. A variety of memory tests have been used, generally measuring explicit or implicit memory performance. Explicit memory refers to memory for learned information that involves facts or events and is associated with an awareness of the source of that information. Implicit memory refers to memory for information or experience whose source is largely out of the individual's awareness, but nonetheless may influence behavior. In several studies the information to be learned was presented (implicitly or explicitly) to one alter personality state and retrieval performance was measured in the same and different alter personality states.

Memory studies have most consistently found dissociative effects on implicit memory functions. Alter personality state-dependent compartmentalization of explicit information has also been demonstrated in some cases. The degree of ambiguity of the information appears to be a key factor in determining whether learned information is available for recall across different alter personality states. Highly ambiguous information is the most likely to become compartmentalized within the specific alter personality state in which it was acquired and relatively unavailable in other personality states. Patients with this disorder also show alter personality state-dependent disturbances in the retrieval of autobiographical information such that specific memories, typically those containing negative experiences and affects, are most available while in certain personality states. Experimental studies indicate that alter personality state-dependent compartmentalization of learned information is generally relative and rarely absolute, so that some of the information can usually be retrieved across personality states. Finally, comparison of individuals with this disorder with normals in standardized autobiographical memory studies show that the former subjects recall significantly fewer memories for childhood.

Differential Psychophysiology Psychophysiological studies of these patients have largely been directed towards the implicit question of “Are the alter personalities real?” Beginning with Morton Prince's 1908 galvanometer study of electrodermal responses to emotional cue words, investigators have sought to identify differential responses to the same stimulus by alter personality states in patients with dissociative identity disorder. The result is a number of single case or small sample studies reporting differential psychophysiological responses that are not duplicated by simulating control subjects. These include power spectral differences in the electroencephalogram (EEG), differences in the amplitude and latency of visual and auditory evoked potentials, ophthalmological differences, and differences in hemispheric and regional cerebral blood flow as measured by several techniques. Personality state-dependent differences have also been reported for autonomic nervous system variables such as heart rate and blood pressure. None of these studies is definitive. In aggregate, they support long-standing clinical observations of differential alter personality physiological reactivity to medications, allergens, and other stimuli.

Epileptic-and-Limbic System Model Modern scientific models echo the explanations proposed by Pierre Janet, Morton Prince, William James and others at the end of the last century. The epileptic- and-limbic system dysfunction model arises from clinical observations of a higher-than-expected incidence in patients with dissociative identity disorder of apparent seizures and symptoms such as depersonalization, déjà vu, and fugue behaviors that also occur in patients with complex partial seizures. These observations led to a model suggesting that epileptiform discharges were a major factor in the etiology of this disorder. Also, unusually high rates of left temporal lobe EEG abnormalities have been reported in children and adults with histories of childhood maltreatment. However, video EEG studies have not correlated epileptoid activity with dissociative epileptiform symptoms. Also, studies using standardized diagnostic measures have compared epilepsy patients, including those with refractory temporal lobe pathology, to patients with dissociative identity disorder; little clinical similarity was found between seizure patients and dissociative patients in these studies.

Several single case reports document changes in hemispheric lateralization of motor functions across alter personality states, for example, changes in dominant handedness; larger controlled studies have not yet replicated these reports. Overall, when all the available psychophysiological data are considered, patients with dissociative identity disorder generally show significantly greater variability compared to controls when measurements are made comparing different personality states on a specific measure.

Autohypnotic Model The autohypnotic model is widely subscribed to by clinicians working with dissociative patients. It postulates that pathological dissociation is an extreme form of self-hypnosis or autohypnosis. Autohypnosis is postulated to be adaptive in the immediate context of trauma or abuse, but subsequently becomes maladaptively elaborated into dissociative alter personality states. Proponents point to similarities between the phenomenology of deep trance states and some of the clinical phenomenology seen in dissociative identity disorder. Also, adjunctive hypnotherapeutic interventions can be quite helpful in the treatment of many such patients. In addition, studies of hypnotizability using standardized scales have shown that dissociative patients have the highest hypnotizability compared to patients with other diagnoses such as mood disorders, panic disorder, personality disorders, and schizophrenia as well as normal controls.

On the other hand, more than a dozen studies find only low correlations between measures of hypnotizability and dissociation in clinical and nonclinical subjects. A history of trauma is not necessarily associated with increased hypnotizability, although a subgroup of traumatized individuals will show high levels of hypnotizability both clinically and on standardized measures. These findings indicate that hypnotizability and clinical dissociation, as defined by standard measures, are different processes. Although high hypnotizability may be a correlate of traumatization in some patients, the autohypnotic model of the etiology of dissociative identity disorder does not by itself appear to account for the disorder.

Social Role Model The social role model is frequently invoked by skeptics, who regard this disorder as an iatrogenic condition. These critics argue that it is a socially induced behavior that arises through the effects of implicit or explicit suggestion by an authority figure such as a therapist on a susceptible patient. Experiments with undergraduate students demonstrate that with sufficient cueing some individuals will say that they have a second personality that is responsible for alleged misbehaviors. These staged simulations, however, are psychologically transparent and lack any clinical features of dissociative identity disorder. Psychophysiological studies consistently find differences between patients with the disorder and simulating controls, which further indicates that the simulation or malingering social role model is not sufficient to explain the disorder.

Behavioral State Model The behavioral state model conceptualizes the disorder as a developmental failure by a traumatized child to consolidate a core sense of identity. Drawing on research demonstrating the key role of discrete behavioral states in the patterning and organization of normal early childhood behavior and affect regulation, the behavioral state model postulates that trauma disrupts unification of identity in a least two key ways. The first is through the creation of discrete behavioral states associated with the mitigation of and restitution from repetitive traumatic experiences such as incest. These dissociative behavioral states psychologically encapsulate intolerable memories and affects through cognitive mechanisms such as state-dependent learning and memory retrieval. Secondly, repetitive traumatic experiences, together with disturbed caretaker-child attachment and parenting, disrupt the development of normal metacognitive processes involved in the elaboration and consolidation of a unified sense of self. These metacognitive processes, which flower between ages 1 to 6 years, enable the child to integrate the different experiences of self that normally occur across different contexts, for example, with parents, peers, and others. A corollary of this notion is the idea that the failure of integration of self may preserve aspects of parent-child attachment necessary for development because the child may continue to perceive the caretaker as good, despite mistreatment or neglect. Encapsulation of traumatic experiences may also permit more normal maturation in other developmental dimensions such as educational and intellectual tasks, interpersonal relations, and artistic endeavors.

Overall, however, the long-term outcome of these developmental deficits and deformations operating over childhood and adolescence is an individual with multiple,

relatively concretized, quasi-independent senses of self, which are often in psychological conflict with each other. The secondary structuring of these self-states due to a variety of developmental pressures and intrapsychic needs results in the concrete elaboration of the alter identities with names, personal descriptors, and variable ways of presenting themselves to others. These secondary elaborations are not the core aspect of the disorder. However, they may be highly invested in by some patients and thus may be quite resistant to change. At the other extreme, some patients may show significant liability to influence and suggestion in outward presentational features of the alter identities.

Nosology Early clinical conceptualizations considered multiple personality disorder as a manifestation of hysteria. In the first edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I), published in 1952, multiple personality disorder was included under the category of dissociative reaction. In DSM-II, published in 1968, multiple personality disorder was included as hysterical neurosis, dissociative type. In DSM-III, a new diagnostic category of dissociative disorders was created and has been retained in the revised third edition of DSM (DSM-III-R) and the fourth edition of DSM (DSM-IV). The diagnostic criteria have remained essentially unchanged since DSM-III-R, other than the addition of the requirement of dissociative amnesia in DSM-IV.

DIAGNOSIS AND CLINICAL FEATURES

Table 18.3-1 lists the DSM-IV criteria for dissociative identity disorder. The comparable ICD-10 disorder, multiple personality disorder, falls under the other dissociative (conversion) disorder category (see Table 18.5-3). The DSM and ICD criteria are virtually identical; both require that organic disorders, (e.g., general medical conditions, substance abuse) be ruled out. DSM-IV adds that in children the symptoms cannot be attributable to imaginary playmates or other fantasy play.

A. The presence of two or more distinct identities or personality states (each with its own relatively enduring pattern of perceiving, relating to, and thinking about the environment and self).
B. At least two of these identities or personality states recurrently take control of the person's behavior.
C. Inability to recall important personal information that is too extensive to be explained by ordinary forgetfulness.
D. The disturbance is not due to the direct physiological effects of a substance (e.g., blackouts or chaotic behavior during alcohol intoxication) or a general medical condition (e.g., complex partial seizures). **Note:** In children, the symptoms are not attributable to imaginary playmates or other fantasy play.

Table 18.3-1 DSM-IV Diagnostic Criteria for Dissociative Identity Disorder

Dimensions of Trauma Traditional characterizations of dissociative symptoms largely derive from nineteenth-century formulations and do not incorporate recent understanding of the psychiatric effects of trauma. Comparative study of psychiatric symptoms associated with different types of trauma suggests that a number of common dimensions underlie traumatic sequelae. Affect modulation is frequently disturbed, giving rise to mood swings, depression, suicidality, and generalized irritability. Impulse control is often impaired, leading to risk taking, substance abuse, and inappropriate or self-destructive behaviors; high levels of anxiety and panic are common. A variety of disturbances in sense of self, from the identity diffusion seen in patients with borderline personality disorder to the alter identities of patients with dissociative identity disorder, reflect disruptions in the psychological integration of traumatic and nontraumatic aspects of self. Eating disorders are common in a subgroup of trauma patients and may also relate to disorders of body image and identity. Frequent somatization, conversion, and psychophysiological disorders may represent disruptions in the integration of psychic and somatic representations of overwhelming recollections, intolerable affects, posttraumatic cognitive schema, and intrapsychic conflicts. In addition, studies suggest that childhood sexual abuse survivors with psychophysiological disorders when compared with controls are more likely to have a lower threshold for experiencing physiological phenomena as noxious or painful.

Initial clinical presentations in trauma victims of all kinds may encompass or mimic a variety of psychiatric conditions including mood and anxiety disorders, somatoform disorders, personality disorders, and psychosis. Some of these disorders may also be comorbidly associated with dissociative identity disorder, especially posttraumatic stress disorder, mood disorders, somatoform disorders, substance use disorders, and a mixed personality disorder most commonly with some combination of avoidant, obsessive-compulsive, dependent, and borderline traits. All systematic studies of the clinical phenomenology of dissociative identity disorder emphasize the polysymptomatic presentations of these patients. Therefore, the detection and diagnosis of the disorder involve looking beyond a confusing plethora of symptoms for core dissociative symptoms of functional amnesias, depersonalization and derealization, passive influence experiences, and identity alterations.

Memory Symptoms Dissociative disturbances of memory are manifest in several basic ways and are frequently observable in clinical settings. As part of the general mental status examination, clinicians should routinely inquire about experiences of losing time, blackout spells, and major gaps in the continuity of recall for personal information. Patients rarely report these experiences spontaneously and require active inquiry by the interviewer to uncover amnesia. Positive responses should be documented with specific examples provided by the patient. In some instances, patients report coming to or waking up in the midst of some activity with little or no recall of how they came to be involved in that activity. In other instances patients find evidence of having done or acquired things for which they have no recall. Friends and family members may tell them about significant things that they have said or done that they cannot remember. Patients may find that they have unknowingly traveled some distance (a fugue episode) or that days or even weeks have passed unaccounted for. Dissociative time loss experiences are too extensive to be explained by normal forgetting and typically have sharply demarcated onsets and offsets. It is important to establish that such time-loss experiences occur in the absence of intoxication or substance abuse, although high rates of drug and alcohol abuse in dissociative patients may complicate this determination.

Patients with severe dissociative memory disturbances also report perplexing fluctuations in skills, habits, or well-learned abilities such as fluency in a foreign language or athletic abilities. Patients report drawing a complete blank for skills or knowledge at times, while at other times they easily and reliably access the information in question. This perplexing forgetfulness is believed to be related to the dissociative state-dependent disturbances of implicit memory functions that have been documented in laboratory settings.

Dissociative patients often report significant gaps in autobiographical memory, especially for childhood events. Dissociative gaps in autobiographical recall are usually sharply demarcated and do not fit the normal decline in autobiographical recall for younger ages. For example, a patient may complain of being unable to recall anything between the ages 8 to 12 years, while reporting readily available memories before and afterwards. Another patient may report having no memories available for the first 10 years of life. Available autobiographical memories may have a depersonalized quality such that recalled events seem to be memories of a dream or as if the patient had seen them happen to someone else.

Process Symptoms Dissociative process symptoms include depersonalization and derealization, dissociative hallucinations, passive influence or interference experiences, and dissociative cognition. Some authorities include dissociative alterations in identity in this category. Symptoms of depersonalization and derealization are commonly reported by patients and may include profound out-of-body experiences. Patients frequently report feeling “spaced out” or disconnected from themselves and others. The world is perceived as distant or unreal with a hazy or foggy quality. Patients may report feeling as if they exist in a waking dream state. Out-of-body experiences commonly take the form of watching oneself from a distance (inner as well as outer) as if observing another person, with little or no ability to affect one's actions.

Dissociative auditory hallucinations commonly take the form of voices heard as originating from within the person (pseudohallucinations) as opposed to coming from outside. Individual hallucinated voices typically have distinctive age and gender attributes. They may comment negatively about the patient, argue with each other, command the patient to perform certain acts, discuss neutral topics, and sometimes provide useful information or comfort. Patients generally recognize that the voices are hallucinations and may be reluctant to reveal their existence for fear of being considered psychotic. Many patients report some ability to ignore or disregard hallucinations unless they are stressed. Hallucinated voices often come to be identified with specific alter personality states. Visual hallucinations typically take the form of detailed images with traumatic or frightening content. Other visual hallucinations may be understood as depicting the alter identities or may even have a complex artistic quality. Tactile, gustatory, and olfactory hallucinations may also occur, leading to misdiagnoses of seizure disorder or other organic mental disorders. Intrusive posttraumatic flashbacks and images may also be experienced as complex multimodal hallucinations.

Negative hallucinations, in which external percepts and stimuli are not consciously registered, are not uncommon. Patients also report the volitional ability to block out

various perceptions or sensations, including pain. Most hallucinations, pseudohallucinations, and negative hallucinations in patients with dissociative identity disorder are likely to be homologous to phenomena that can be experienced in deep trance states among highly hypnotizable individuals, and are not a manifestation of a process psychosis.

Passive influence or interference symptoms include many first-rank Schneiderian symptoms such as audible thoughts; voices arguing with each other; influences playing on the body; thought withdrawal and insertion; and ‘made' feelings, impulses, and actions. These symptoms were once considered to be pathognomonic of schizophrenia, but can also be found in patients with mood, organic, and dissociative disorders. Passive influence symptoms now have been demonstrated to be more common in patients who have dissociative identity disorder than in patients with psychotic mental disorders. However, the agents of the passive-influence symptoms are usually experienced by dissociative patients as internal, not external, as in psychotic disorders. In addition, patients may report strong affects or impulses that they experience without a sense of personal ownership, but with a peremptory sense of intrusion and control.

The recognition that dissociative patients frequently manifest subtle but often clinically significant cognitive impairments emerges from clinical research with psychological and cognitive test batteries. Research using projective testing finds distinctive cognitive process markers, including evidence of confusing and contradictory responses to the same stimulus. Distinctive responses to standardized projective testing can often be helpful in distinguishing dissociative patients from other diagnostic groups such as patients with mood disorders, nondissociative forms of posttraumatic stress disorder personality disorders, psychotic disorders, and factitious disorders.

Dissociative Alterations of Identity Clinically, dissociative alterations in identity may first be manifest by odd first person plural or third person singular or plural self-references. In addition, patients may refer to themselves by their own first names or make depersonalized self-references such as “the body” when describing themselves and others, for example, “The father hurt the body so she was upset. We tried to protect her, but it didn't work.” Patients often describe a profound sense of concretized internal division or personified internal conflicts between “parts” of themselves. In some instances these parts may have proper names or be designated by their predominant affect or function (e.g., the angry one or the wife). Patients may suddenly change the way in which they refer to others (e.g., “the son” instead of “my son”).

A set of behaviors, collectively referred to as “switching behaviors,” may be manifest during evaluation or therapy sessions. Switching behaviors include intrainterview amnesias, in which the patient does not seem to recall or is very confused about the process and content of the session. These microdissociative episodes may be manifest by abrupt shifts in the train of thought or sudden inexplicable changes in affect or rapport. A variety of physical signs may occur in conjunction with microdissociative episodes including pronounced upward eye rolls or bursts of rapid blinking and eyelid fluttering. The patient's tone of voice and manner of speaking, posture, and demeanor may show marked alteration. When clinicians encounter evidence of possible microdissociative episodes they should seek to clarify what the patient is experiencing and can recall with nondirective, open-ended questions.

These cognitive, behavioral, and physical shifts are manifestations of alter personality switching or overlap and interference between alter states. The alter personalities of patients with dissociative identity disorder are best conceptualized as discrete behavioral states each organized around a prevailing affect, sense of self (often including a distinct body image), a set of state-dependent autobiographical memories, and a limited behavioral repertoire. Authorities have long cautioned that alter personalities should not be regard as separate people. Rather, the alter personalities are conceptualized as relatively stable and enduring patterns of behavior that are largely unintegrated with each other and often in direct conflict.

The set of alter personality states, usually referred to as the personality system, constitutes the personality of the individual. Most psychotherapeutic work is directed towards this larger personality system and thus towards the individual as a whole. Much has been made of the psychological and physiological differences among the alter personality states of these individuals and popular accounts emphasize the presentation differences among alters. Laboratory studies support clinical accounts of significant differences; however, much general information and many functions and abilities are shared across alter personality states and indicate the fundamental unity of the mental processes of the dissociative individual. This provides the foundation for therapeutic efforts directed at the development of a more consciously integrated sense of self in the patient with dissociative identity disorder.

Apparent differences in the organization and dynamics of alter personality systems have been used to classify dissociative patients into various categories for more than a century. The validity of these classifications remains to be proven, but as a group these patients show considerable variability in the complexity and therapeutic tractability of their alter personality systems. Several alter personality types are commonly reported, including child alter personalities, internalized persecutory alters who inflict pain and may attempt to kill the individual, and depleted and depressed “host” personality states, who function as the primary identity with respect to the world at large. Alter personality states often reflect painful psychological issues for the individual and frequently take the form of polarized pairs representing antithetical positions, although alters representing more neutral and conflict-free processes also commonly occur. In some individuals, virtually every aspect of mental life is structuralized and personified in this form.

Child and Adolescent Presentations A growing clinical research literature documents the diagnosis and treatment of child and adolescent dissociative disorders including dissociative identity disorder. In most respects children and adolescents manifest the same core dissociative symptoms and secondary clinical phenomena as adults. Age-related differences in autonomy and life-style, however, may significantly influence the clinical expression of dissociative symptoms in youth. For example, dissociative amnesias and perplexing forgetfulness are more apparent in school situations rather than work or family life. Younger children in particular have a less linear and less continuous sense of time and are often not able to self-identify dissociative discontinuities in their behavior. Fortunately, there are often additional informants, such teachers and relatives, available to help document dissociative behaviors. A number of normal childhood phenomena, such as imaginary companionship and elaborated daydreams, must be carefully differentiated from pathological dissociation in younger children.

Screening Instruments and Diagnostic Interviews The development and increasing utilization of screening instruments and standardized diagnostic interviews has contributed to the increase in numbers of identified cases of dissociative identity disorder seen since the 1980s. A variety of self-report questionnaires are available for clinical screening. The most widely used, the DES, takes about 10 minutes to complete and has good reliability, sensitivity, and specificity. Two well-validated diagnostic interviews are available. SCID-D is a 277-item, clinician-administered, semi-structured interview that addresses all five of the DSM-IV dissociative disorders. It is currently considered the gold standard. Many clinicians rely on the shorter Dissociative Disorders Interview Schedule (DDIS) for its ease and speed of administration. Child and adolescent dissociation measures are just becoming available and are not as well validated. The Child Dissociative Checklist (CDC) and Adolescent Dissociative Experiences Scale (A-DES) are two commonly used screening measures. Dissociative modules have been drafted for the standard child diagnostic interviews, but these sections are as yet untested. Several stand-alone child or adolescent diagnostic interviews have also recently been introduced.

Ms. A. is a 33-year-old married woman, employed as a librarian in a school for disturbed children. She presented to psychiatric attention after discovering her 5-year-old daughter “playing doctor” with several neighborhood children. Although this event was of little consequence, the patient began to become fearful that her daughter would be molested. Ms. A. became panicked and increasingly obsessed with this idea, much to the bafflement of her husband. The patient was seen by her internist, treated with antianxiety agents and antidepressants, but with little improvement. Ms. A. became increasingly anxious, phobic, depressed, and preoccupied. She sought psychiatric consultation from several clinicians, but repeated good trials of antidepressants, antianxiety agents, and supportive psychotherapy resulted in limited improvement. After the death of her father from complications of alcoholism, the patient became more symptomatic. He had been estranged from the family since the patient was about 12 years old because of his drinking and associated antisocial behavior.

Ms. A. developed a variety of somatic complaints including headaches, abdominal pain, menstrual and gastrointestinal problems, back pain, and sleep difficulties. Repeated medical workup was unrevealing leading to diagnoses such as fibromyalgia, irritable bowel syndrome, and premenstrual tension. Family and marital difficulties increased as the patient withdrew from her husband and was increasingly dysfunctional in taking care of her children; work function also deteriorated. Psychiatric hospitalization was precipitated by the patient's arrest for disorderly conduct in a nearby city. She was found in a hotel, dressed in revealing clothing, engaged in an altercation with a man. She denied knowledge of how she had come to this hotel, although the man insisted she had come there under a different name for a voluntary sexual encounter.

On psychiatric examination the patient described dense amnesia for the first 12 years of her life with the feeling that her “life started at 12 years old.” She reported that for as long as she could remember she had had an imaginary companion, an elderly black woman, who advised her and kept her company. She reported hearing other voices in her head: several women and children, as well as her father's voice repeatedly speaking to her in a derogatory way. She reported that much of her life since age 12 was also punctuated by episodes of amnesia: for work, for her marriage, for the birth of her children, and for her sex life with her husband. She reported perplexing changes in skills; for example, she was often told she played the piano well, but had no conscious awareness that she could do so. Her husband reported that she had always been “forgetful” of conversations and family activities. He also noted that at times she would speak like a child, at times adopt a southern accent, and at other times be angry and provocative. She frequently had little recall of these episodes.

Questioned more closely about her early life, the patient appeared to enter a trance and stated; “I just don't want to be locked in the closet” in a childlike voice. Inquiry about this produced rapid shifts in state between alter identities who differed in manifest age, facial expression, voice tone, and knowledge of the patient's history. One identified itself by a diminutive of the patient's name and appeared childlike. Another spoke angrily, used expletives, and appeared irritable and preoccupied with sex. She discussed the episode with the man in the hotel and stated that it was she who had arranged it. A third alter personality identified itself as a protective entity, experiencing itself as an elderly African-American woman who commented sadly and philosophically about “this whole situation.” Gradually the alter identities described a history of family chaos, brutality, and neglect during the first 12 years of the patient's life, until her mother, also alcoholic, achieved sobriety and left her husband, taking her children with her. In the alter identities, the patient described episodes of physical abuse, sexual abuse, and emotional torment by her father, her siblings, and her mother.

Family sessions with the mother and siblings confirmed many of these reports, with the family recalling episodes of maltreatment that the patient did not recollect. The patient's mother had bid the family never to speak of the earlier difficulties, hoping that everyone would “just forget the whole thing.” After additional assessment of family members, the patient's mother also met diagnostic criteria for dissociative identity disorder, as did her older sister who also had been molested. A brother met diagnostic criteria for posttraumatic stress disorder, major depressive disorder, and alcohol dependence.

The patient improved significantly with psychotherapy directed at stabilization of her dissociative identity disorder and posttraumatic stress disorder. She responded well to clomipramine (Anafranil) with a marked reduction in obsessive-compulsive and depressive symptoms Family therapy was helpful in stabilizing the patient's marriage and helping her husband with the aftermath of the hospitalization and its precipitants. The husband also reported a family history of abuse, although primarily directed at his mother and siblings; he had always seen himself as the family protector. The patient's mother and siblings were already in treatment, but were helped by the opening up of the family history and clarification of their diagnoses. At 3-year follow-up the patient reported fusion of most alter identities, marked diminution of dissociative, somatoform and posttraumatic stress disorder symptoms, but she still required clomipramine (Anafranil) for stabilization of mood and for the symptoms of obsessive-compulsive disorder.

Differential Diagnosis On average, more than 6 years pass between first psychiatric contact and the diagnosis of dissociative identity disorder. Although often portrayed as flamboyant hysterics, only a small minority of patients present in this way. These patients are typically reticent about revealing their dissociative symptoms, especially hallucinations, amnesia, and identity divisions. They most commonly present as relatively inhibited and obsessional, with affective and somatic complaints, and typically acquire three or more psychiatric diagnoses before the disorder is recognized. A subgroup of patients show interpersonal dynamics reminiscent of borderline personality disorder, which some patients with dissociative identity disorder qualify for as a secondary diagnosis once posttraumatic stress disorder and dissociative symptoms are stabilized. The presence of auditory hallucinations, disturbed thinking and behavior, confusion due to amnesic gaps, and Schneiderian first-rank symptoms contributes to the misdiagnosis of schizophrenia in about half of dissociative patients at some point in their psychiatric histories. Rapid changes in mood associated with alter personality switching may suggest a rapid-cycling mood disorder or schizoaffective disorder. Table 18.3-2 lists the most common disorders that must be differentiated from dissociative identity disorder.

Other dissociative disorder (dissociative amnesia, dissociative fugue, dissociative disorder not otherwise specified)
Schizophrenia and other psychotic disorders
Rapid-cycling mood disorder
Borderline personality disorder
Malingering and factitious disorder
Partial complex seizure disorder
Posttraumatic stress disorder

Table 18.3-2 Differential Diagnosis of Dissociative Identity Disorder

COURSE AND PROGNOSIS

Little is known about the natural history of untreated dissociative identity disorder. A few case studies of partially treated patients followed up many years later suggest that the disorder becomes less overt over time, with a decrease in florid dissociative symptoms and less intrapsychic conflict among the alter personality states. These cases, however, are too few and too selected to generalize. Some untreated individuals are thought to continue to be involved involvement in abusive relationships or violent subcultures that may result in the traumatization of their children, with the potential for additional family transmission of the disorder. Many authorities believe that some percentage of undiagnosed or untreated patients die by suicide or as a result of their risk-taking behaviors. Experience with a large number of dissociative identity disorder cases suggests that there are several subgroups of these individuals, ranging from those who function at quite high levels for long periods of time to others who have severely impaired and dysfunctional life trajectories often beginning early in development.

Patient presentations and prognosis vary somewhat across the life span. Children with dissociative identity disorder show many dissociative symptoms and behaviors, but typically have fewer and less crystallized alter personality states that are less invested in their individuality. If diagnosed early, children often have an excellent prognosis and many seem to have relatively spontaneous resolutions when removed from abusive and neglectful environments. In adolescence, alter personality states become more distinct and more invested in their autonomy. Additional alter personalities associated with life stresses such as academic, athletic, social, or sexual challenges may appear and the personality system dynamics become more complicated and polarized. In general, adolescents have a poorer prognosis than children or adults, in part because they are often not invested in their treatment. Better outcome with dissociative adolescents has been reported when the patients' families were successfully engaged in treatment. Young adults typically present in crisis and have a layering of affective, somatic, posttraumatic, and personality

disorder symptoms in addition to their core dissociative pathology.

First presentations in middle-aged or older patients frequently involve a life event such as the loss of a job, death of a parent, or a revictimzation such as a rape that reactivates earlier conflicts and destabilizes the alter personality system. In adult patients, there appear to be both treatment-responsive and treatment-refractory subgroups of patients with dissociative identity disorder. Time course of improvement also may vary, with some patients improving relatively quickly and others requiring intensive treatment over long periods of time.

Prognosis is poorer in patients with comorbid cognitive disorders, psychotic disorders (not dissociative pseudopsychosis), and severe medical illnesses. Refractory substance abuse and eating disorders also suggest a poorer prognosis. Other factors that usually indicate a poorer outcome include significant antisocial personality features, current criminal activity, ongoing perpetration of abuse, and current victimization with refusal to leave abusive relationships. Repeated traumas with recurrent episodes of acute stress disorder may severely complicate the clinical course. Severe personality disorders with overinvestment in multiplicity as a way of life, engagement in psychotherapy primarily to seek gratification, or a refusal to take responsibility for behavior change and symptom management are also generally associated with a poorer prognosis. However, the number of alter personality states has only a moderate effect on treatment course.

TREATMENT

Stages and Goals Current treatment approaches to dissociative identity disorder have evolved considerably with the conceptualization of the disorder as a complex developmental trauma disorder and as the spectrum of patients has been better appreciated. Appropriate treatment of the dissociative patient follows a three-stage model: (1) a phase of symptom stabilization; (2) an optional phase of focused in-depth attention to traumatic material; and (3) a phase of integration or reintegration in which the patient moves more completely away from a life adaptation based on long-term traumatization and victimization. Obviously, these phases are relatively heuristic and aspects of each may be part of the others.

Stabilization of the patient is vital to permit more successful negotiation of all aspects of treatment. Stabilization focuses on safety, stability, and management of core and comorbid symptoms. The vast majority of patients engage in some form of self-destructive behavior, including suicide attempts, self-mutilation, eating disorders, substance abuse, promiscuity, risk-taking activities, and involvement in abusive, violence-based relationships. Many male and some female patients with dissociative identity disorder have difficulty with aggression, violence, and homicidality, including perpetration of child abuse. It is incumbent on the clinician to make these issues the basic focus of treatment. In general, cognitive-behavioral therapy approaches are used, framing these behaviors as part of a set of quasiaddictive trauma-related homeostatic mechanisms. Experienced clinicians find that many dissociative patients can bring self-destructive and high-risk behaviors under control. Cognitive-behavioral therapy methods are used to develop therapeutic agreements (*safety contracts*) so that patients have a repertoire of techniques to manage their difficulties instead of by self-destruction. Eating disorders, severe substance abuse, enmeshment in abusive relationships, and perpetration of violence may be more refractory to these methods and may require concurrent specialty treatment interventions and involvement of police, social service, and community agencies.

Stabilization of severe symptoms generally involves work with posttraumatic stress and dissociative symptoms. Patients with dissociative identity disorder frequently present with highly disturbing symptoms of posttraumatic stress disorder such as intrusive thoughts, imagery, somatic sensations, hyperarousal, and flashbacks. The latter may present as acute behavioral reexperiencing episodes with loss of reality orientation or more subtle and pervasive reliving experiences. Patients may be overwhelmed by dissociative hallucinations, amnesia and fugue episodes, passive influence experiences, or profound identity confusion and disorganization. Sorting out genuinely comorbid disorders from the plethora of posttraumatic, anxiety, mood, and somatic symptoms commonly manifest by dissociative patients in crisis is important because some disorders (e.g., mood disorders) require additional treatment interventions. Assessment of available family and community supports is also important.

Psychotherapy A survey of over 300 clinicians treating patients with dissociative identity disorder found that the vast majority considered psychotherapy to be their primary and most efficacious treatment modality. Successful psychotherapy for the such a patient requires the clinician to be comfortable with a range of psychotherapeutic interventions and willing to actively work to structure the treatment. These modalities include psychoanalytic psychotherapy, cognitive-behavioral therapy, hypnotherapy, and a familiarity with the psychotherapy of the traumatized patient.

Effective stabilization in many patients requires psychotherapeutic work with individual alter personality states. Many patients will not be able to stabilize symptoms in the long term if alter identities who control these symptoms are not therapeutically engaged. Clinicians new to dissociative identity disorder are frequently uncomfortable or perplexed by the need to work with individual alter states and with how to do this without producing a chaotic regression. Certain basic principles are important to understand. No alter is any more or less real than any other or better or worse than another. All are aspects of a single human being and have psychological importance that needs to be heard and respected. All alter states will be held accountable and responsible for the behavior of any part, even if experienced with amnesia or lack of subjective ownership. In the context of psychotherapy, alter personalities can be understood as developmentally concretized, trauma-based metaphors and forms of mental symbolization. These unusual metaphors have many problematic cognitions and affects as well as adaptive attributes such as the alter personality state-dependent memory data. As such, the fullest participation of all mental symbols and sources of knowledge, skills, and adaptations is likely to produce the best outcome in psychotherapy.

Concerns about iatrogenesis are often raised in this context. Understanding the alter personalities as forms of intrapsychic symbolization reduces this concern to a more routine psychotherapeutic task of understanding the meaning of mental contents, rather than a suppression of some mental states as less authentic than others. Most experienced clinicians avoid regression in dealing with dissociative patients' alter identities by insisting that the patient has resources to counter regression. Tendency for regression can be managed by: (1) firm limits and boundaries on the length of sessions; (2) payment of fees; (3) allowing extra-session contacts, and acceptable behavior during sessions; (4) limiting regressive interactions with child alter personalities and insisting that “caretaking” of these entities is the job of the patient, not the clinician; (5) actively structuring sessions to focus on goals of stabilization; (6) using cognitive-behavioral containment and imagery techniques to attenuate the impact of posttraumatic stress disorder intrusions, rather than opening up this material; (7) helping the patient develop new skills to manage symptoms such as dysfunctional switching; (8) using therapeutic contracting to help the patient take active responsibility for symptom management and behavioral control outside of sessions; and (9) actively countering the manifold cognitive distortions presented by the patient.

Typical cognitive distortions include the insistence that alter personalities inhabit separate bodies and are unaffected by the actions of one another (delusional separateness); that patients are unable to control themselves and require the clinician to manage all difficulties; that the clinician is completely untrustworthy and must be not be allowed any access to the patient's mind; that the patient is “bad” and deserved to abused; that anger and violence are the same; that love and sex are the same; that self-injury is safety; and that since trauma and abuse are inevitable it is best to invite them so that at least their timing and intensity can be better controlled.

The continuity, stability, and respectful impartiality of the therapist towards the different alter personalities provides an important therapeutic experience that helps the patient to experience, examine, and integrate these dissociated aspects of the self. The processing of negative life events, often from the multiple perspectives of different alter personality states, is an important part of the narrative reorganization of the patient's fragmented identity into a more coherent whole.

A subgroup of patients with dissociative identity disorder will not progress beyond a long-term supportive treatment entirely directed towards stabilization of their multiple multiaxial difficulties. To the extent that they can be engaged in treatment at all, these patients will require a long-term treatment focused on symptom containment and management of their overall life dysfunction, as would be the case with any other severely and persistently ill psychiatric patient population.

For patients who can stabilize and form a reasonable working alliance in treatment, longer-term treatment goals involve the psychotherapeutic processing of life experiences, especially traumatic experiences, and the transformation of the meaning of these experiences for the individual. Authorities emphasize that in most cases intensive, detailed psychotherapeutic work with traumatic memories should only be initiated after the patient is well stabilized and a solid therapeutic relationship has been established. Experienced clinicians attempt to carefully structure affectively intense sessions focused on traumatic material. In addition, many sessions may be needed to fully explicate the cognitive and emotional meaning of traumatic events so that they may become part of the patient's repertoire of nondissociated, ordinary memories for life experience.

Traumatic Memories Media attention and legal cases have led to concerns over the authenticity of traumatic recollections of patients with dissociative identity disorder. Recent rigorous case series have supported the experience of clinicians who have found corollary information to corroborate patients' recollections or even unearthed data about traumas that the patient does not recall. Based on corollary information, however, some patients can be shown to misinterpret and misrepresent contemporary information or confabulate aspects of their past history. In addition, dissociative patients are hardly immune to ordinary human emotions such as greed, envy, wishes for revenge, wishes to evade consequences for misbehavior, and desires to placate significant others. All these may complicate the patient's veracity,

especially in situations of potential financial gain, media attention, evasion of legal consequences, and possible loss of family contacts. Conversely, collateral informants, such as family members, may be subject to the same sources of unreliability as the index patient and their input should be weighed accordingly. The American Psychiatric Association's statement on recollections of childhood abuse is a helpful guide to these complex matters.

Patients with dissociative identity disorder typically oscillate from regarding their recollections as all true to all false. Specific alter identities may take opposing positions. The clinician is best served by maintaining a stance of neutrality towards the patient's recollections. In this regard it is generally most helpful for the clinician to identify for the patient internal conflicts over the veracity of recollections and invite an open airing of all points of view. Education about the complex factors that may affect autobiographical recall may be helpful, but also may be too abstract for some patients to absorb, especially early in treatment.

The therapist should respond respectfully and thoughtfully to all clinical material brought by the patient. Patients can be helped to not come to premature closure about their views of events. Some patients come to understand life events quite differently as they become less dissociative over the course of treatment. Consequently, clinicians must avoid validating memories for the patient, or dismissing them out of hand, in the absence of reliable collateral information. Therapy is most successful for the patient if the clinician maintains the role of therapist, not personal advocate, detective, or derisive skeptic.

Transference and Countertransference Patients with dissociative identity disorder often manifest a complex multilayered transference as a whole, as a system of personality states, and as individual alter personality states. Commonly, transference is dominated by themes of trauma and abuse with the therapist most commonly experienced as potentially exploitative and abusive, uninvolved or uncaring about the patient's difficulties, or a helpless victim like the patient faced with an implacable other. Countertransference responses may vary accordingly with overinvolvement, detached hostile skepticism, or a sense of being overwhelmed and unskilled being quite common. Burnout and “secondary” posttraumatic stress disorder has been reported for therapists and treatment teams; adequate attention must be paid to limit-setting, boundaries, and assistance to clinicians who experience reports of extremely abusive trauma adversely.

Therapeutic Resolution Over the course of treatment significant unification of dissociated mental processes may be observed. Alter personalities lose distinctness and decrease compartmentalization of thoughts, memories, and affects and the patient develops a more unified sense of self. Transference is modified consistent with these changes. Amnesia and switching become less apparent. Fusion of alter personalities results in psychological merging of two or more entities at a point in time with a subjective experience of loss of all separateness. Some patients proceed to what appears to be a complete fusion of all alter identities, with a shift in self-representation from that of a dissociated individual to one with a consistent and continuous sense of self across all behavioral states. Many patients never attain a full integration of their alter personalities but leave treatment when they have achieved relative stability, adequate function, and some measure of internal harmony.

As this integrative process occurs, symptoms of posttraumatic stress disorder usually improve significantly. Patients often experience a freeing of energy towards everyday life and away from trauma-focused ways of living. At the same time, nondissociative and integrative coping strategies must be identified and substituted for dissociative responses to life stressors. Losses must be mourned and the patient must be helped to connect and to cope with the larger world in a more functional manner.

Hypnosis Despite the controversy about its use, hypnosis was endorsed by approximately two thirds of respondents as a psychotherapeutic adjunct. Hypnotherapeutic interventions can often alleviate self-destructive impulses or reduce symptoms such as flashbacks, dissociative hallucinations, and passive-influence experiences. Teaching the patient self-hypnosis may help with crises that occur outside of sessions. Hypnosis may be useful for accessing specific alter personality states and their sequestered affects and memories. Hypnosis is also used to create relaxed mental states in which negative life events can be examined without overwhelming anxiety. Clinicians using hypnosis should be trained in its use in general and in trauma populations. Clinicians should be aware of current controversies over the impact of hypnosis on accurate reporting of recollections and obtain appropriate informed consent for its use.

Pharmacotherapy Pharmacotherapy was the third most commonly endorsed treatment modality. Although double-blind, controlled clinical trials have not been conducted, a variety of medications are considered clinically effective with many patients with dissociative identity disorder. Guidelines for the use of medications with dissociative patients emphasize the need to identify specific treatment-responsive symptoms rather than attempting to treat the dissociation per se. Medications may be helpful in attenuating symptoms to assist the patient in stabilizing in treatment. Patients should be advised that medication response is likely to be partial. In general, success is more likely if medication target symptoms are present across a range of alter personality states rather than confined to one or a few personality states.

Among the target symptoms considered most responsive to medication are affective symptoms. In many instances, these are secondary and show a more heterogeneous and less robust response than primary mood disorders. Nonetheless, antidepressant medications are often important in the reduction of depression, and stabilization of mood. A variety of symptoms of posttraumatic stress disorder, especially intrusive and hyperarousal symptoms, are partially responsive to medication. Guided by clinical experience and research on patients with posttraumatic stress disorder, clinicians report some success with antidepressant medications, b-adrenergic receptor antagonists, clonidine (Catapres), anticonvulsants, and benzodiazepines in reducing intrusive symptoms, hyperarousal, and anxiety in these patients. Sleep disturbances and traumatic nightmares may also be improved by these medications, although patients should be cautioned that sleep disturbances may be particularly refractory to medications. Case reports suggest that aggression may respond to carbamazepine (Tegretol) in some individuals if EEG abnormalities are present. Many patients show significant obsessive-compulsive symptoms, and they may preferentially respond to antidepressant medications that have antiobsessive efficacy. Open-label studies suggest that naltrexone (ReVia) may be helpful for amelioration of recurrent self-injurious behaviors in a subset of traumatized patients.

Questions are often raised about the efficacy of antipsychotic medications, particularly for symptoms such as hallucinations. Although these drugs are only minimally effective for quasipsychotic symptoms such as hallucinations in many patients with dissociative identity disorder, low doses may be useful in some cases for severe anxiety and for the subtle cognitive slippage sometimes seen. The newer drugs, such as risperidone (Risperdal) and olanzapine (Zyprexa), may be more effective and better tolerated for overwhelming anxiety and intrusive posttraumatic stress disorder symptoms. Dissociative memory and process symptoms have proven refractory to medication in general. Dissociative-like symptoms and behaviors, however, can be induced in some patients with posttraumatic stress disorder and in normal individuals with drugs (e.g., phencyclidine, cannabinoids), suggesting the presence of pharmacologically sensitive neurobiological mechanisms that may lead to future medications specifically targeting dissociative symptoms per se.

Inpatient Treatment Inpatient treatment is most commonly used for stabilization of patients in acute crisis and, less often, for providing a safe environment in which to do intensive work with painful moods and memories. For a number of reasons, patients with overt dissociative identity disorder can have disruptive effects on the milieu of general psychiatric units and often stimulate splitting and conflict among staff about the best way in which to relate to the patient and the alter personalities. Inpatient programs specializing in the treatment of dissociative and posttraumatic disorder patients may be more successful because they provide a unified treatment approach in which rigorous boundaries and firm limit-setting are combined with specific supportive treatment interventions to assist with symptoms. Inpatient treatment on general hospital units is most successful if “special” treatment is avoided for the dissociative patient with a focus on a unified approach to managing specific target symptoms.

Partial hospital treatment can be a very effective modality for the dissociative patient if clinical interventions target trauma-based issues and adaptations. Specialized trauma-partial hospital programs with groups emphasizing symptom containment, cognitive-behavioral interventions, and psychoeducation can be very effective in helping to stabilize the more severely and persistently ill dissociative patients as well as in providing an intensive stabilization experience for higher-functioning dissociative patients.

Adjunctive Treatments Authorities agree that in therapy groups that include general psychiatric patients, the emergence of alter personalities can be disruptive to the group process by eliciting excess fascination or by frightening other patients. Therapy groups composed only of dissociative patients are reported to be more successful, although the groups must be carefully structured, set firm limits, and should generally focus only on here-and-now issues of coping and adaptation. Family or couples therapy is often important for long-term stabilization and to address pathological family and marital processes that are common in these patients and their family members. Family therapy with the family of origin of the dissociative patient may be helpful in clarifying on resolving the conflictual emotions these patients frequently experience towards family members. Care must be exercised in setting up such meetings since some interactions with family members may lead to a decompensation in the patient. Confrontation of family members about past traumas in an accusatory manner almost invariably has a disastrous outcome both for the dissociative patient and family members. Accordingly, there is usually little place for such interventions in treatment of dissociative identity disorder.

Expressive and occupational therapies, such as art and movement therapy, have proven particularly helpful in treatment of dissociative patients. Art therapy may permit patients to safely express thoughts, feelings, mental images, and conflicts that they have difficulty verbalizing. Movement therapy may facilitate normalization of

body image for these severely traumatized patients.

Outcome Studies Single case descriptions of successful treatments for dissociative identity disorder date back more than a century. Systematic outcome studies, however, have only appeared since the 1980s. The first such study followed up 20 patients an average of 3 years after intake. The majority were being treated by therapists who were unfamiliar with this disorder. Nonetheless, two thirds of the clinicians reported moderate to great improvement in their patients. A history of severe retraumatization during the course of treatment was associated with poorer outcomes. In the Netherlands a chart review study of 101 dissociative disorder patients in outpatient treatment for an average of 6 years found that clinical improvement was related to the intensity of the treatment, with more comprehensive therapies having better outcomes. A study using the DES to track treatment progress of 21 dissociative inpatients found a significant drop in overall scores over a 4-week hospitalization.

The largest and most systematic treatment outcome study re-evaluated 54 inpatients 2 years after discharge to outpatient treatment. As a group, there were significant overall decreases in psychopathology including number of Axis I and Axis II disorders, decreased DES scores, decreased depression on the Beck Depression Inventory and Hamilton Rating Scales for Depression and decreased dissociative symptoms on all of the DDIS subscales. Patients who were reported integrated according to the criteria defined by Richard Kluft in 1984 were the most improved. Two studies investigating cost efficacy of treatment for dissociative identity disorder have concordant findings suggesting that outcome depends on clinical subgroup. The more treatment-responsive group of patients showed significant remission of symptoms within 3 to 5 years of beginning appropriate treatment. A second group with more alter identities and more personality disorder features showed good outcome but required hospitalizations in addition to outpatient treatment. A third group, characterized by the longest period of treatment before diagnosis, largest number of alter identities, and most personality disorder problems, had a much longer, more expensive, and more difficult course. Overall, however, treatment approaches specifically targeting this disorder showed reductions in overall psychiatric treatment cost after the first year compared with prior treatment for these patients. Some health maintenance organizations report that more intensive treatment benefits for dissociative patients has not only reduced overall psychiatric costs but has also reduced costs for medical utilization for somatoform symptoms.

These preliminary studies have notable limitations including the diverse and nonstandardized nature of the therapy and lack of comparison groups. Nonetheless, in aggregate they indicate that many patients with dissociative identity disorder improve with treatments focused on their dissociative symptoms and that overall treatment costs may be saved in the long term with better treatment for these patients.

FORENSIC ISSUES

The intersection between the diagnosis dissociative identity disorder and the legal system has proven exceedingly controversial. The contentious dispute over its existence, the often sensationalized media attention, and the need to rule out simulation or malingering have made the forensic evaluation of the disorder difficult to perform and to defend. Issues of competency to stand trial and degree of responsibility for the behavior of different alter personality states have received contradictory judicial opinions. The most common defenses are: (1) dissociative defendants do not have control over or are not conscious of their alter personalities and therefore cannot be held responsible for their actions; (2) these defendants cannot recall the actions of their alter personalities and therefore cannot participate in their own defense; and (3) a diagnosis of dissociative identity disorder makes it impossible for a defendant to conform to the law or to know right from wrong. Evidentiary questions, such as the admissibility of hypnotic or amobarbital (Amytal) interviews and the independence of testimony by different alter personalities have proven problematic.

In assessing possible dissociative identity disorder in criminal defendants, a number of guidelines may be helpful: (1) Follow basic principles of forensic assessment; (2) undertake a comprehensive assessment of all available documentary materials concerning the defendant and the case; (3) review all available past psychiatric, social service, and related materials; (4) perform a standard diagnostic assessment on the defendant before introducing specialized testing; where possible, videotape the interview following forensic guidelines for videotaping; (5) conduct a longitudinal life-history interview, which may be useful in cases where malingered amnesia or dissociative identity disorder are suspected: ask for the defendant's entire life history beginning with the first memory and proceeding to the present time, following up information as necessary in a neutral manner; (6) after completion of the basic clinical assessment, psychological testing, specialized dissociative interviews (e.g., the SCID-D), or neuropsychiatric testing may be done; (7) corollary sources (e.g., friends and family) may be necessary to corroborate the history in addition to written records; and (8) use forensic guidelines for hypnotically facilitated or drug-facilitated interviews; avoid performing these unless you have been trained in forensic hypnosis.

There is no pathognomonic presentation of dissociative identity disorder in the forensic setting. However, authorities suggest that genuine defendants are more likely to show a complex alter system, not just a good-or-bad alter dichotomy. Individuals with bonafide dissociative identity disorder may have bizarre explanations for why and how they (in alter identities) committed crimes, not just that “the bad one did it.” Authorities also state that a number of these defendants minimize their psychiatric symptoms and attempt to avoid being labeled psychiatrically ill, although a few will exaggerate their symptoms.

Clinicians may find themselves treating patients who face criminal charges for minor crimes such as shoplifting, check forgery, credit card fraud, or driving while intoxicated. It may be impossible for the patient to hire an independent forensic expert. In these cases the clinician may have to provide a written report or appear for court testimony at the request of the patient or the patient's attorney. The clinician should clarify with the patient on the differing roles of the forensic expert and the therapist. It may be difficult for the therapist to separate the roles of the advocate for a disturbed patient from that of a forensic examiner. However, the clinician should take seriously the importance of protecting the public from the patient. Clinically, the forensic situation may be useful to underscore for the patient the responsibility of the whole person for behavior committed in dissociated states of consciousness and to gain better adherence to responsible behavior from acting-out alter identities.

In general, most courts have not found dissociation sufficient grounds for a claim of legal incompetence. Also, the clinical standard of holding the whole human being responsible for the behavior of any part should be the foundation for any legal consideration of diminished responsibility for criminal conduct in the dissociative patient. An affirmative case should be made that the defendant fails to meet the standard for legal sanity or has diminished capacity, just as would be done for any psychiatric defense.

SUGGESTED CROSS-REFERENCES

Further descriptions of the nature and the function of dissociative mechanisms are found in Chapter 6, dealing with the theories of personality and psychopathology, and in Chapter 9, dealing with the classification of mental disorders. The diagnostic distinction between dissociative disorders and other mental disorders is clarified in Chapter 12 on Schizophrenia and in Chapter 15 on anxiety disorders. Chapter 16 on the Somatoform disorders provides a detailed description of the somatic symptoms that in this chapter are viewed as manifestations of dissociation. Expanded descriptions of the various psychotherapeutic approaches appears in Chapter 30.

SECTION REFERENCES

American Society for Clinical Hypnosis: *Guidelines for Clinicians Working with Hypnosis and Memory and for the Conduct of Forensic Hypnosis Interviews. American Society for Clinical Hypnosis*, Des Plaines, IL, 1994.

Armstrong JG: Reflections on multiple personality disorders as a developmentally complex adaptation. Psychoanal Study Child *50:*349, 1995.

Armstrong JG, Loewenstein RJ: Characteristics of patients with multiple personality and dissociative disorders on psychological testing. J Nerv Ment Dis *178:*448, 1990.

Behnke SH: Confusion in the courtroom: How judges have assessed the criminal responsibility of individuals with multiple personality disorder. Int J Law Psychiatry *20:*293, 1997.

Bremner JD, Marmar CR: Trauma, memory and dissociation. American Psychiatric Press, Washington, DC, 1998.

Coons PM: Iatrogenesis and malingering of multiple personality disorder in the forensic evaluation of homicide defendants. Psychiatr Clin North Am *14:*757, 1991.

Coons PM: Confirmation of childhood abuse in child and adolescent cases of multiple personality disorder and dissociative disorder not otherwise specified. J Nerv Ment Dis *182:*461, 1994.

*Eich E, Macaulay D, Loewenstein RJ, Dihle PH: Memory, amnesia and dissociative identity disorder. Psychol Sci *8:*417, 1997.

*Ellason JW, Ross CA: Two-year follow-up of inpatients with dissociative identity disorder. Am J Psychiatry *154:*832, 1997.

Fujii Y, Suzuki K, Sato T, Murakami Y, Takahashi T: Multiple personality disorder in Japan. Psychiatry Clin Neurosci *52:*299, 1998.

Griffin MG, Resick PA, Mechanic MB: Objective assessment of peritraumatic dissociation: Psychophysiological indicators. Am J Psychiatry *154:*1081, 1997.

Groenndijk I, van der Hart O: Treatment of DID and DDNOS patients in regional institute for ambulatory mental care in the Netherlands: A survey. Dissociation *8:*73, 1995.

Hornstein N, Putnam FW: Clinical phenomenology of child and adolescent dissociative disorders. J Am Acad Child Adolesc Psychiatry *31:*1077, 1992.

Kluft RP, Fine CG: *Clinical Perspectives on Multiple Personality Disorder*. American Psychiatric Press, Washington, DC, 1993.

*Lewis DO, Putnam FW: Dissociative identity disorder/multiple personality disorder. Child Adolesc Psychiatr Clin North Am *5:*263, 1996.

Lewis DO, Yeager C, Swica Y, Pincus J, Lewis M: Objective documentation of child abuse and dissociation in 12 murders with dissociative identity disorder. Am J Psychiatry *154:*1703, 1997.

Loewenstein RJ: An office mental status examination for complex chronic dissociative symptoms and multiple personality disorder. Psychiatr Clin North Am *14:*567, 1991.

Loewenstein RJ: Rational psychopharmacology in the treatment of multiple personality disorder. Psychiatr Clin North Am *14:*721, 1991.

Ogawa J, Sroufe L, Weinfield N, Carlson E, Egeland B: Development and the fragmented self: Longitudinal study of dissociative symptomatology in a nonclinical sample. Dev Psychopathol *9:*855, 1997.

*Pope HG Jr, Oliva PS, Hudson JI, Bodkin JA, Gruber AJ: Attitudes toward DSM-IV dissociative disorders diagnoses among board-certified American psychiatrists. Am J Psychiatry *156:*321, 1999.

Putnam FW: *Dissociation in Children and Adolescents: A Developmental Perspective*. Guilford, New York, 1997.

Putnam FW, Loewenstein RJ: Treatment of multiple personality disorder: A survey of current practices. Am J Psychiatry *150:*1048, 1993.

Robins L, Barrett J, editors: *The Validity of Psychiatric Diagnoses*. Raven, New York, 1989.

Ross CA: *Dissociative Identity Disorder*. Wiley, New York, 1997.

Scroppo JC, Drob SL, Weinberger, Eagle P: Identifying dissociative identity disorder: A self-report and projective study. J Abnorm Psychol *107:*272, 1998.

*van Ijzendoorn M, Schuengel C: The measurement of dissociation in normal and clinical populations: Meta-analytic validation of the Dissociative Experiences Scale (DES). Clin Psychol Rev *16:*365, 1996.

*Waller N, Ross C: The prevalence and biometric structure of pathological dissociation in the general population: Taxometric and behavior genetic findings. J Abnorm Psychol *106:*499, 1997.

18.4 DEPERSONALIZATION DISORDER

MARLENE STEINBERG, M.D.

Although listed as a dissociative disorder in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders*, (DSM-IV), coverage of depersonalization in the scientific literature has until recent years been limited. However, interest in the dissociative disorders has steadily increased since the 1970s, and many advances have recently been made in the study of depersonalization and the other dissociative disorders, including the development of reliable diagnostic instruments.

DEFINITION

DSM-IV defines depersonalization as “a feeling of detachment or estrangement from one's self,” in which “The individual may feel like an automaton or as if he or she is living in a dream or movie.” This may be accompanied with “a sensation of being an outside observer of one's mental processes, one's body, or parts of one's body.” Additional examples of depersonalization include emotional and mental numbness, not recognizing oneself in the mirror, the feeling of lack of control over emotions, and feeling as if body parts are detached or unreal.

HISTORY

Although first described by M. Krishnaber in 1872, depersonalization was not named as such until 1898, when L. Dugas coined the term to contrast “the feeling of loss of ego” with a “real loss.” While various other nineteenth- and early–twentieth-century scholars in psychology—such as Eugén Bleuler, Sigmund Freud, Josef Breuer, and Pierke Janet—referred to the phenomenon in passing or listed it as a symptom in their celebrated case studies, relatively little attention was paid to it until 1980, when *depersonalization disorder* was officially listed as a psychiatric diagnosis in the third edition of DSM (DSM-III). However, a few notable studies on depersonalization occurred during this time. W. Mayer-Gross, in his classical paper of 1935, began the modern study of depersonalization and was followed by H. Shorvonn in 1946 and J. Saperstein in 1949. The next major contributor to the study of depersonalization was C. Ackner, who in his 1954 paper noted that the phenomenon lacked clear boundaries and therefore described four salient features that continue to be widely accepted: (1) the feeling of unreality or strangeness apropos the self; (2) the retention of insight and lack of delusional elaboration; (3) the lack of affective response (“numbness”) except for the discomfort regarding depersonalization; and (4) the unpleasant property that may vary in intensity inversely with the subject's familiarity with the phenomenon.

In recent years interest in depersonalization has been increasing. An extensive and systematic investigation of depersonalization has been performed based on the administration of the *Structured Clinical Interview for DSM-IV Dissociative Disorders–Revised* (SCID-D-R), a diagnostic tool for the assessment of specific dissociative symptoms and disorders. SCID-D-R research indicates that the range and severity of depersonalization distinguishes dissociative identity disorder from the occasional occurrence of depersonalization, which may be seen in patients with other psychiatric disorders. In addition, depersonalization has been included as one of the five core symptoms of dissociation, the others being amnesia, derealization, identity confusion, and identity alteration. In the evaluation of patients experiencing depersonalization, it is essential to assess the presence of the other core symptoms of dissociation in order to arrive at an accurate diagnosis.

COMPARATIVE NOSOLOGY

With each edition of DSM, the thrust has been toward a more specific definition of depersonalization disorder and a greater emphasis on subjective stress. The most recent (fourth) edition of DSM (DSM-IV), is the most useful in that its wording better clarifies the criteria and presents terminology and examples that are easily understood.

While depersonalization disorder was classified as a dissociative disorder in DSM-III and all subsequent editions, the International Classification of Diseases (ICD) does not list it as such. In its most recent incarnation, the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), depersonalization-derealization syndrome is listed under other neurotic disorders. This distinction is fairly arbitrary, but some scholars support this exclusion because memory disturbance, which is found in the other dissociative disorders, is absent (Table 18.4-1).

A. Either of the following must be present:
   (1) *Depersonalization*. The patient complains of a feeling of being distant or "not really here." For example, individuals may complain that their emotions, feelings or experience of the inner self are detached, strange, not their own, or unpleasantly lost, or that they feel as if acting in a play.
   (2) *Derealization*. The patient complains of a feeling of unreality. For example, there may be complaints that the surroundings or specific objects look strange, distorted, flat, colorless, lifeless, dreary, uninteresting, or like a stage upon which everyone is acting.
B. There is retention of insight, in that the patient realizes that the change is not imposed from outside by other persons or forces.

Table 18.4-1 ICD-10 Diagnostic Criteria for Depersonalization-Derealization disorder

EPIDEMIOLOGY

Transient depersonalization has been reported to be the third most common complaint among psychiatric patients, following depression and anxiety; in fact, it is experienced to some degree by 80 percent of the psychiatric population. However, the symptom of depersonalization often occurs in association with other disorders and may not be the predominant syndrome. Recurrent depersonalization is a common occurrence in the context of dissociative identity disorder and dissociative disorder not otherwise specified. The sex ratio, while inadequately documented and unclear, seems to favor females. It has been reported widely in adults and adolescents, but few cases have been reported in children. There is no known pattern of inheritance for depersonalization disorder.

ETIOLOGY

The etiology for depersonalization is diverse and symptoms are probably produced by an array of biological and psychological factors. Several biological and psychodynamic theories have been proposed. First, depersonalization may result from temporal lobe dysfunction and various metabolic and toxic states. This theory has associated depersonalization with epilepsy and other diseases of the central nervous system, as well as the ingestion of psychotomimetic drugs like mescaline and lysergic acid diethylamide (LSD). Second, depersonalization may result from a predetermined functional brain response adaptive to overwhelming trauma, as

evidenced by its occurrence in a variety of other psychiatric disorders and in nonpsychiatric populations. Third, depersonalization may be a defense against painful conflictual affects such as guilt, phobic anxiety, anger, rage, paranoia, conflictual ego identifications, primitive fusion fantasies, and exhibitionism. Systematic research into these theories is limited, and further research is necessary to clarify and provide conclusive support for any current theories of etiology.

Research has documented the psychological triggering of depersonalization. Depersonalization frequently appears as a response to life-threatening danger such as accidents, serious illnesses, cardiac arrests, anaphylactic reactions, and complications of surgery, as well as a response to emotional stress from a variety of situations such as overwhelming anxiety, anger, or severe conflict. Depersonalization appears to occur commonly in conjunction with posttraumatic stress disorder, dissociative identity disorder and hallucinogen persisting perception disorder (flashbacks) and is frequently reported by survivors of severe physical, emotional, or sexual abuse; political imprisonment; torture; and cult indoctrination.

DIAGNOSIS AND CLINICAL FEATURES

Depersonalization, in the sense of detachment from the self, occurs as a transient symptom of a variety of Axis I and Axis II disorders. However, it is important to distinguish between this mild and episodic depersonalization and the more severe, prolonged, and recurrent syndrome of depersonalization disorder. The DSM-IV diagnostic criteria of depersonalization disorder are listed in Table 18.4-2.

A. Persistent or recurrent experiences of feeling detached from, and as if one is an outside observer of, one's mental processes or body (e.g., feeling like one is in a dream).
B. During the depersonalizing experience, reality testing remains intact.
C. The depersonalization causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The depersonalization experience does not occur exclusively during the course of another mental disorder, such as schizophrenia, panic disorder, acute stress disorder, or another dissociative disorder, and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., temporal lobe epilepsy).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 18.4-2 DSM-IV Diagnostic Criteria for Depersonalization Disorder

Although the central characteristic of depersonalization is the quality of unreality and estrangement from the self—wherein both internal mental processes and external worldly events remain unchanged while their relation and meaning to the patient alter or even appear to lose all coherence—this disorder can manifest itself in a variety of ways. Common manifestations of depersonalization include a feeling of being unreal, of being dead, of parts of the body being disconnected, or of observing a movie of the self. The depersonalization experience is evident in the following excerpt from a patient interview:

> It's really weird. It's sort of like I'm here, but I'm really not here and that I kind of stepped out of myself, like a ghost... . I feel really light, you know. I feel kind of empty and light, like I'm going to float away... . Sometimes I really look at myself that way... . It's kind of a cold, eerie feeling. I'm just totally numbed by it.

Many people who suffer from chronic depersonalization do not report these experiences spontaneously. Patients often are reluctant to relate these experiences for fear that the clinician will think that they are crazy; some even adapt to and learn to accept their chronic condition, and may actually find it comforting. This makes detection and diagnosis particularly difficult, since depersonalization disorder is marked only by an altered internal state of consciousness and perception, not by any apparent change in external social behavior. Even those who do report feelings of depersonalization often find that their experiences are difficult to put into words. For example, they may describe their feelings as follows: “[It's like] being high on drugs,” “I would have the ability to fly, to escape from my body for a short period of time,” “I can see myself [from the outside] and what I must look like [to others] or what I'm doing or how I'm walking,” or “It's like feeling like you're floating in space and watching [yourself].” Due to the elusive nature of depersonalization, and to the fact that it is rarely the patient's presenting complaint, it is essential that clinicians assess for the varied manifestations of depersonalization in mental status examinations.

Bodily Perceptions Parts of one's body or one's entire body may feel insubstantial, unreal, or foreign.

> ... it just doesn't seem real. Everything to me doesn't seem real, my body included... . I was looking in the mirror and all of a sudden it just felt as if the image in the mirror was looking back out upon myself.

Emotional, Mental and Behavioral Perceptions One's internal mental and emotional operations may seem altered or foreign, sometimes resulting in behavioral transformations or limitations.

> You just feel like you're never being yourself, uh, you're not thinking as you normally do. You just feel strange... . I need to concentrate on things a lot more... . I feel as if most people have their brain on automatic. I feel like I have to crank it, like I'm always working at it and always consciously trying to remember things, trying to concentrate... . It's almost as if I can't visualize things in my mind the way you normally would, when you actually see things ... I can't do that very well.

External Perceptions Feelings of unreality, strangeness, and discontinuity may pervade one's perceptions of objects and people in the outside world. Thus, individuals experiencing depersonalization may experience it in conjunction with derealization.

> [Everything] just feels different, just like it's in a dream, in a daze, things are dulled... . I remember when this first started and I was in college and I went home and the thing that kept going over in my mind was I want to go home, I want to go home and I was home, but that just kept coming into my mind... . everything always seems unreal... . Other people, myself, just everything... .

Anxiety Although many patients report a sense of numbness or emotional detachment, this lack often does not extend to fear of depersonalization itself; in fact, many patients become upset by the sense of unreality. They may have attacks of panic as a result of either hampered interaction with the outside world or fear of encroaching insanity.

> Um well, yeah, I guess you'd call it a panic attack... . if someone's talking to me [and] I can't understand what they're saying or I'm just not thinking clearly, then it's kind of a panic that they're going to notice that, or if I'm driving. Sometimes I'd have problems with driving for a while.

Reality Testing Although patients may report feeling as if they were unreal or automated, reality testing—that is, the ability to distinguish between the real and the unreal—always remains intact; depersonalization is typified by the “as if” quality of the experience.

> I know what is really real... . [things] just don't seem real... . I'm used to it now. I'm used to compensating... . I know pretty much what it's going to do, how bad it is at its worst or how good it is at its best... . everything you're saying feels like it's not real, but yet you know it's real and you know what you really just said.

Screening Tools and Diagnostic Interviews Until recently, the diagnosis of depersonalization disorder and of the rest of the dissociative disorders was difficult to make because of the multiple causes for transient depersonalization and the lack of tools for systematic assessment. In recent years, however, the development of structured interviews for diagnosis have improved the accuracy of diagnoses. Screening instruments for dissociative symptoms include the Questionnaire of Experiences of Dissociation, which consists of 26 true-false items, the Dissociative Experiences Scale, which consists of 28 questions, and the Perceptual Alteration

Scale, which includes items from the Minnesota Multiphasic Personality Inventory relative to dissociation. The Dissociative Disorders Interview Schedule is a structured interview for the assessment of dissociative disorders. Poor inter-examiner reliability has been reported for depersonalization on this tool.

Because screening tools cannot diagnose depersonalization disorder, individuals who have dissociative symptoms should be evaluated further with a thorough clinical evaluation and a more comprehensive, structured interview, such as SCID-D-R in order to rule out the presence of a dissociative disorder. The SCID-D-R is a semistructured interview that evaluates the severity of five dissociative symptoms (amnesia, depersonalization, derealization, identity confusion, and identity alteration) and diagnoses dissociative disorders, including depersonalization disorder. The average duration of the interview is 45 to 90 minutes, and contains 40 items specific to depersonalization, with follow-up items if the examiner wishes to obtain additional information. Good to excellent reliability and discriminant validity have been reported for the SCID-D-R, with the symptom of depersonalization receiving the highest rating for inter-examiner reliability. The *Interviewer's Guide to the SCID-D-R* provides guidelines for the administration, scoring and interpretation of SCID-D-R.

Proper treatment of depersonalization depends upon accurate diagnosis of its presence and severity, as well as the symptom constellation surrounding it, as the following a case study of a patient diagnosed with depersonalization disorder on the SCID-D-R demonstrates. This case is presented in the form of a report suitable for inclusion in the patient's records, in order to demonstrate the utility of the SCID-D-R in patient education and treatment planning as well as diagnostic assessment. Figure 18.4-1 is a visual outline of the process of differential diagnosis.

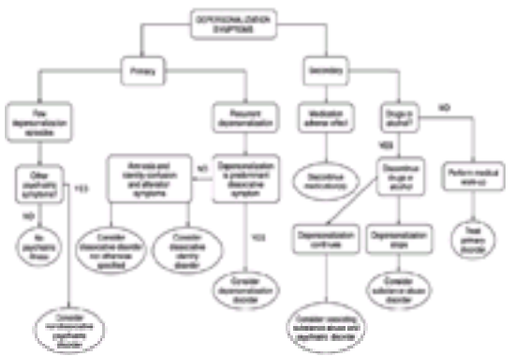

FIGURE 18.4-1 Differential diagnosis decision tree of depersonalization. (Reprinted with permission from Steinberg M: *Interviewer's Guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised [SCID-D-R].* American Psychiatric Press, Washington, DC, 1994.)

SCID-D-R Psychological Report Ms. Walker

Dates of Evaluation: 8/11/94 and 8/18/94

Referral Source: present therapist

Reason for Referral: diagnostic evaluation

Information obtained from: Ms. Walker; present therapist

Brief Summary: Ms. Walker is a 31-year-old unmarried Caucasian woman who presented with the complaint of “feeling detached” from herself. Ms. Walker currently commutes to her job as an administrative assistant at a nearby community college. Her present therapist, from whom she has been receiving outpatient treatment for the past 3 years, referred her.

Past Psychiatric History: The patient has no history of hospitalization for psychiatric disturbance. In 1995 she began receiving treatment with her present therapist when periods of depression began to interfere with both her job and her social relationships. Although she had experimented with marijuana during college, she has never required treatment for a substance abuse disorder. While she indicates that the severity of her depression has diminished somewhat, she still describes her feelings as those of being “detached.”

Family History: Ms. Walker is the older of two siblings. Her family is intact but is in general emotionally unsupportive. She claimed that there is no history of alcohol or drug abuse in her family. While she also stated that neither of her parents had been treated for a psychiatric disorder, she indicated that her mother and father experienced episodes of emotional instability. She said that both her parents were prone to mood swings and spontaneous temper outbursts.

Mental Status Exam: Dressed neatly in a sweater and skirt, Ms. Walker was calm and cooperative. She spoke fluently and answered questions with cogent replies. Though she appeared mildly depressed, her affect seemed to be full range. Ms. Walker denied the presence of auditory or visual hallucinations and did not display any symptoms of psychotic thinking. Furthermore, she claimed not to have experienced any suicidal or homicidal ideation. For a summary of her dissociative symptoms, see the SCID-R Evaluation below.

SCID-D-R Evaluation: In addition to a routine mental status examination, the SCID-D-R was administered in order to systematically evaluate the patient's dissociative symptoms (chronic feelings of detachment from self) and the presence of dissociative disorders. The SCID-D-R interview was then scored according to the guidelines described in the Interviewer's Guide to the SCID-D-R. A review of the significant findings from the SCID-D-R interview includes the following:

Ms. Walker denied experiencing amnestic episodes. However, she endorsed experiencing recurrent episodes of detachment from herself. These feelings of depersonalization cause her considerable distress and inhibit her ability to function effectively in either an occupational or a social environment. This ongoing feeling of depersonalization began at age 18 and has continued since. The patient describes her depersonalization as occurring “all the time”; however, she noted that these feelings vary directly with her overall level of stress. Her experience of depersonalization entail feelings of detachment from life and psychic numbing. Her depersonalization does not manifest itself in disturbances in body image or a sense of a split between participating and observing parts of the self. She endorsed only one out-of-body experience, which she considers an isolated occurrence. With respect to the symptom of derealization, Ms. Walker endorsed experiences of derealization, which vary in intensity together with her depersonalization. However, she reported that her derealization was secondary to her depersonalization, and she never experiences the former without the latter. Her feelings of depersonalization trigger episodes of anxiety and panic, which in turn affect her ability to drive. Ms. Walker entered therapy in 1995, when the combination of depersonalization and panic attacks overwhelmed her, causing feelings of depression. Ms. Walker stated that the depersonalization has negatively affected her ability to function at her job and in life in general. With respect to the symptoms of identity confusion or alteration, Ms. Walker denied experiencing puzzlement regarding her identity, severe mood changes, on alterations in her sense of identity. In addition, Ms. Walker denied experiencing internal dialogues, feelings of being possessed, or the acquisition of unexplained possessions or skills.

Assessment: Based on the SCID-D-R evaluation, Ms. Walker's symptoms are consistent with a primary diagnosis of a dissociative disorder. More specifically, her recurrent episodes of depersonalization with associated derealization, in the absence of amnesia, identity confusion, or alteration is characteristic of SCID-D-R symptom profiles for patients with depersonalization disorder. Her chronic feelings of detachment and unreality towards herself, associated with dysfunction, in the absence of substance abuse disorder or temporal lobe epilepsy, are consistent with DSM-IV criteria for depersonalization disorder. In addition, her descriptions of internal struggle are focused on feelings of unreality, which are characteristic of patients with depersonalization disorder, rather than on conflicts between different personalities, as is described by patients with dissociative identity disorder. Ms. Walker's recurrent depersonalization has in turn resulted in episodes of secondary panic, which interfere with driving, as well as recurrent feelings of depression.

Recommendations: (1) a follow-up interview for review of the findings of the SCID-D-R evaluation, and for educating the patient regarding the nature of her symptoms; (2) continuation of individual psychotherapy. Initial treatment goals should include identification of specific triggers of depersonalization episodes and review of specific techniques for alleviation of the symptom. Longer-range goals should include decrease in the frequency and severity of symptom

disorder, rather than on conflicts between different personalities, as is described by patients with dissociative identity disorder. Ms. Walker's recurrent depersonalization has in turn resulted in episodes of secondary panic, which interfere with driving, as well as recurrent feelings of depression.

Recommendations: (1) a follow-up interview for review of the findings of the SCID-D-R evaluation, and for educating the patient regarding the nature of her symptoms; (2) continuation of individual psychotherapy. Initial treatment goals should include identification of specific triggers of depersonalization episodes and review of specific techniques for alleviation of the symptom. Longer-range goals should include decrease in the frequency and severity of symptom episodes and resultant dysfunction. Trial administration of an antidepressant or antianxiety medication may alleviate this patient's coexisting symptoms of depression and anxiety.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of patients who experience depersonalization should take into account both psychological and organic causes of depersonalization. Patients who present with recurrent depersonalization should receive a comprehensive psychiatric evaluation, including a specialized diagnostic interview in order to rule out the presence of a dissociative disorder.

Medical Etiology If the depersonalization is isolated and coexists with other medical or neurological symptoms, a medical work-up including physical and neurological examination is essential. Blood screening chemistries—such as thyroid function tests, chemical profiles, and a complete blood count—may be necessary. An index of suspicion for temporal lobe epilepsy requires that an electroencephalogram (EEG) be performed. Drug toxicology tests may be helpful in ruling out substance-induced depersonalization, including prescribed medication that may produce depersonalization as an adverse effect.

Individuals with the following disorders may experience depersonalization: (1) seizure disorders such as temporal lobe epilepsy, postconcussional syndrome, structural pathology (i.e., brain tumor), cerebral arteriopathy, and meningioma; (2) infectious causes (i.e., encephalitis); (3) metabolic abnormalities (i.e., hyperglycemia); (4) other conditions, including migraine headaches, vestibular dysfunction, Meniere's disease, and Korsakoff's syndrome.

In addition, depersonalization may occur as an adverse effect of both drug abuse and prescribed medications (including the antipsychotic agent haloperidol [Haldol], the anti-inflammatory agent indomethacin [Indocin], methyldopa [Aldomet], and the amphetamine-like agent fenfluramine). Drug-induced depersonalization is typically transient and disappears on discontinuation of the medication. Since case reports of drug-induced depersonalization have not evaluated the baseline level of depersonalization prior to pharmacotherapy, it is unclear whether the drugs caused the depersonalization, worsened pre-existing depersonalization, or whether the association is coincidental.

It is important to note that organically induced depersonalization typically has a repetitive content, in contrast to the elaborate descriptions and internal dialogues associated with depersonalization in individuals who suffer from dissociative disorders.

Spectrum of Depersonalization Research using SCID-D-R indicates that the severity of depersonalization ranges from isolated episodes experienced by normal subjects under stress or sleep deprivation to recurrent and persistent depersonalization in individuals with dissociative disorders and posttraumatic stress disorder. Patients with a variety of nondissociative psychiatric disorders, including the anxiety disorders (particularly panic disorders, obsessive-compulsive disorder, agoraphobia), mood disorders (depressive disorders, bipolar I disorder), psychotic disorders (schizophrenia, psychotic disorder not otherwise specified), the personality disorders (particularly borderline personality disorder), and the substance-related disorders (particularly those involving marijuana and hallucinogens such as LSD and mescaline, although depersonalization may also occur with alcohol, cocaine, phencyclidine, opioids, sedatives, and stimulants), may also experience occasional depersonalization episodes. On the other hand, patients with dissociative disorders exhibit the most severe depersonalization, including individuals with depersonalization disorder and dissociative identity disorder.

Investigations based on SCID-D-R have found that depersonalization that occurs as a part of depersonalization disorder will occur in the relative absence of other dissociative symptoms. However, in individuals with dissociative identity disorder, depersonalization can occur along with the other four dissociative symptoms—amnesia, derealization, identity confusion, and identity alteration. In addition, research has found that depersonalization that exists within the context of dissociative identity disorder is often distinguished by the occurrence of ongoing, interactive dialogues that occur between the observing and the participating self. It is also important to note that depersonalization occurring in an individual with a dissociative disorder typically has an “as if” quality and that reality testing is intact. In patients with schizophrenia, the depersonalization typically occurs in a psychotic context and reality testing is impaired. In order for a diagnosis of depersonalization disorder to be made, depersonalization must be chronic and of long duration. Once an organic cause has been ruled out, systematic evaluation of depersonalization, including its severity and nature, can distinguish dissociative disorders from other psychiatric disorders.

Table 18.4-3 presents a summary of the diagnostic distinctions between normal and pathological depersonalization, Table 18.4-4 is a summary of the spectrum of depersonalization, and Figure 18.4-1 is a visual summary of the differential diagnosis of depersonalization.

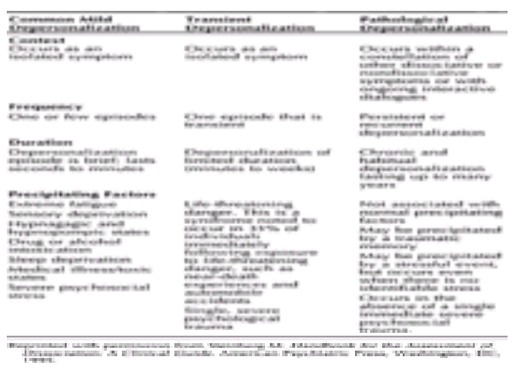

Common Mild Depersonalization	Transient Depersonalization	Pathological Depersonalization
Context		
Occurs as an isolated symptom	Occurs as an isolated symptom	Occurs within a constellation of other dissociative or nondissociative symptoms or with ongoing interactive dialogues
Frequency		
One or few episodes	One episode that is transient	Persistent or recurrent depersonalization
Duration		
Depersonalization episode is brief; lasts seconds to minutes	Depersonalization of limited duration; minutes to weeks)	Chronic and flattened depersonalization lasting up to many years
Precipitating Factors		
Extreme fatigue Sensory deprivation Hypnagogic and hypnopompic states Drug or alcohol intoxication Sleep deprivation Medical illness/toxic states Severe psychosocial stress	Life-threatening danger. This is a syndrome noted to occur in 35% of individuals immediately following exposure to life-threatening danger, such as near-death experiences and automobile accidents Single, severe psychological trauma	Not associated with normal precipitating factors May be precipitated by a traumatic memory May be precipitated by a stressful event, but occurs even when there is no identifiable stress Occurs in the absence of a single immediate severe psychosocial trauma.

Table 18.4-3 Distinguishing Between Normal and Pathological Depersonalization

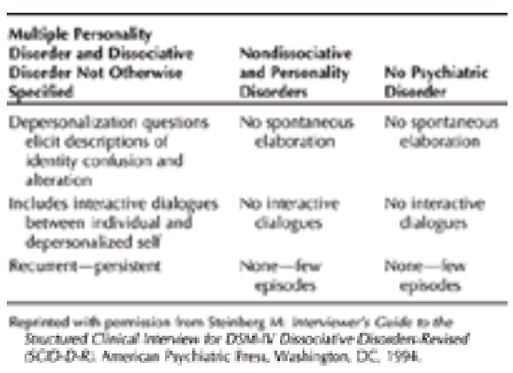

Multiple Personality Disorder and Dissociative Disorder Not Otherwise Specified	Nondissociative and Personality Disorders	No Psychiatric Disorder
Depersonalization questions elicit descriptions of identity confusion and alteration	No spontaneous elaboration	No spontaneous elaboration
Includes interactive dialogues between individual and depersonalized self	No interactive dialogues	No interactive dialogues
Recurrent—persistent	None—few episodes	None—few episodes

Reprinted with permission from Steinberg M. *Interviewer's Guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised (SCID-D-R)*. American Psychiatric Press, Washington, DC. 1994.

Table 18.4-4 The Spectrum of Depersonalization on the SCID-D-R

COURSE AND PROGNOSIS

Depersonalization usually begins between age 15 and age 30 although there have been reports of its onset in childhood. The symptoms typically begin abruptly, after which point they tend to continue on a steady course without significant changes in intensity. However, in some cases the symptoms occur as a series of repeated attacks, with symptom-free intervals in between. In individuals who are exposed to life-threatening trauma or near-death experiences, depersonalization may develop instantaneously upon the onset of trauma and usually remits after the traumatic experience has ended. The prognosis of individuals with depersonalization disorder has not been investigated systematically.

TREATMENT

Therapy for depersonalization is usually indicated if the depersonalization is recurrent or if it results in distress or functional impairment; if depersonalization is secondary to an underlying primary disorder, then that primary disorder should be treated. If depersonalization is the primary disorder, therapy can be directed at alleviating the depersonalization. A variety of treatments have been reported, including supportive psychodynamic psychotherapy, pharmacological treatment, behavioral and directive techniques, and hypnosis.

Psychodynamic Psychotherapy Traditional psychotherapy that teaches the patient to identify the underlying dynamics of and to achieve control over the symptoms has been reported to be effective. For this treatment to be successful, it is important for the patient to learn to accept and tolerate the depersonalization; hence patients should be encouraged to accept the experience rather than to attempt to change it directly.

Cognitive Therapies In patients with depersonalization and a trauma history, cognitive restructuring approaches may be particularly useful. As depersonalization may be triggered by feelings of loss of control or victimization, cognitive therapies that counteract anxiety can help to alleviate anxiety as well as the depersonalization.

Patient Education and Specific Techniques The patient should be educated regarding depersonalization, its onset, development, triggers, and treatment. A review of the defensive nature of this symptom and its triggers can be the first step in reducing anxiety and minimizing future episodes of depersonalization. In addition, techniques are used to reduce recurrent depersonalization.

Review of techniques specific to symptom alleviation.

Grounding (reorienting) is a method in which the patient can make contact with objects in the environment (e.g., touching a chair where they are sitting) during a depersonalization episode and state his or her name, location, and a statement of self-support (e.g., “I'm John, I'm in the library in my house; it is safe here”).

In creative visualization the patient is taught to visualize a safe place as an attempt to manage their depersonalization episodes.

Distraction techniques may be used to focus the patient's attention to an external activity that they enjoy (e.g., drawing, exercising, talking to a friend) in an attempt to divert their attention from the episodes of depersonalization.

In controlled dissociation patients are taught to rate their episodes of depersonalization on a scale of 1 to 10. They are then asked to increase or decrease the intensity of an episode, so that they can regain a sense of control over their symptom.

Hypnosis Since dissociation is a feature of the hypnotic trance, this treatment may help to demonstrate to patients how to control their own dissociation. When combined with well-designed psychotherapy, hypnosis can lead to a reduction in depersonalization symptoms. In addition, hypnosis is an effective tool for gaining access to and reintegrating dissociated traumatic memories. Because dissociative symptoms can occur secondary to hypnosis, it has been recommended that clinicians should document their patients' baseline level of dissociative symptoms prior to induction using the SCID; preferably, the interview should be audiotaped.

Behavior Therapy Reports have indicated successful response mostly to positive reinforcement, although a few report good results with negative reinforcement. Flooding treatment through fantasy and paradoxical intention has also been reported successfully.

Pharmacotherapy In addition to psychological and social factors, biological factors also seem to contribute to dissociation. Some patients may have a physiological predisposition toward depersonalization. While preliminary data on pharmacological treatments is encouraging, systematic research is limited and current treatment of patients with depersonalization should include a flexible individualized approach.

Their have been no double-blind trials of medications for the treatment of depersonalization. Case reports have found antidepressant and antianxiety medications helpful in alleviation of depersonalization, including the tricyclic desipramine (Norpramin) the selective serotonin reuptake inhibitor fluoxetine (Prozac), and the benzodiazepine clonazepam (Klonopin). One study found fluoxetine to be helpful in decreasing depersonalization in patients with coexisting panic or obsessive-compulsive symptoms. In addition, a variety of biological therapies used to treat depersonalization disorder, because of their significant adverse effects, were previously reported to be effective in studies on patients suffering from severe depersonalization. These medications include pentylenetetrazole, methamphetamine (Desoxyn), electroconvulsive therapy (ECT), dextroamphetamine (Dexed), amobarbitol (Amytal), and thiopental (Pentathol). ECT has also been reported to worsen symptoms of depersonalization, as has caffeine. Mixed results have been reported for antipsychotic agents.

Further research is necessary in the form of controlled double-blind studies utilizing reliable tools comparing the efficacy of pharmacotherapeutic agents. Pharmacotherapy trials can now evaluate baseline and postmedication severity levels of depersonalization using recently developed diagnostic tools that allow for the assessment of the severity of depersonalization based on operationalized criteria.

SUGGESTED CROSS-REFERENCES

Dissociative identity disorder is discussed in Section 18.3, dissociative disorder not otherwise specified in Section 18.5, and posttraumatic stress disorder in Chapter 15. Hypnosis is covered in Section 30.3, behavior therapy in Section 30.2, and biological therapies in Chapter 31.

SECTION REFERENCES

Ackner B: Depersonalization, I: Aetiology and phenomenology. J Ment Sci *100:*838, 1954.

Bernstein E, Putnam F: Development, reliability, and validity of a dissociation scale. J Nerv Ment Dis *174:*727, 1986.

Blue F: Use of directive therapy in the treatment of depersonalization neurosis. Psychol Rep *45:*904, 1979.

Cattell JP, Cattell JS: Depersonalization: Psychological and social perspectives. In *American Handbook of Psychiatry*, S Arieti, editor. Basic Books, New York, 1974.

Dugas L: Un cas de depersonalization (A case of depersonalization). Rev Philos *45:*500, 1898.

Frances A, Sacks M, Aronoff M: Depersonalization: A self relations perspective. Int J Psychoanal *58:*325, 1977.

*Hammond C: *Hypnosis in the Treatment of Dissociative Identity Disorder in Multiple Personality Disorder: Continuum of Care*, B Cohen, J Turkus, editors. Bruner Mazel, New York, 1999.

Hollander E, Fairbanks J, Decaria C, Liebowitz M: Pharmacological dissection of panic and depersonalization (letter to the editor). Am J Psychiatry *146:*402, 1989.

*Hollander E, Liebowitz M, Decaria C, Fairbanks J, Fallon B, Klein D: Treatment of depersonalization with serotonin reuptake blockers. J Clin Psychopharmacol *10:*200, 1990.

Jacobson E: Depersonalization. J Am Psychoanal Assoc *7:*581, 1959.

Mayer-Gross W: On depersonalization. Br J Med Psychol *15:*103, 1936.

Noyes R Jr, Kletti R: Depersonalization in response to life-threatening danger. Compr Psychiatry *18:*375, 1977.

Riley K: Measurement of dissociation. J Nerv Ment Dis *176:*449, 1988.

Roberts W: Normal and abnormal depersonalization. J Med Sci *106:*478, 1960.

Shimizu M, Sakamoto S: Depersonalization in early adolescence. Jap J Psychiatry *40:*603, 1986.

Shorvon H: The depersonalization syndrome. Proc R Soc Med *39:*779, 1946.

Sookman D, Solyom L: Severe depersonalization treated by behavior therapy. Am J Psychiatry *135:*1543, 1978.

Spiegal D: Dissociation and hypnosis in post-traumatic stress disorders. J Traum Stress *1:*17, 1988.

Steinberg M: The spectrum of depersonalization: Assessment and treatment. In *American Psychiatric Press Review of Psychiatry*, vol 10, A Tasman, S Goldfinger, editors. American Psychiatric Press, Washington, DC, 1991.

*Steinberg M: *Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised (SCID-D-R)*. American Psychiatric Press, Washington, DC, 1994.

*Steinberg M: *Interviewer's Guide to the Structured Clinical Interview for DSM-IV Dissociative Disorders-Revised (SCID-D-R)*. American Psychiatric Press, Washington, DC, 1994.

*Steinberg M: *Handbook for the Assessment of Dissociation: A Clinical Guide*. American Psychiatric Press, Washington, DC, 1995.

Steinberg M: *A Clinician's Guide to Diagnosing Dissociative Symptoms and Disorders: The Structured Clinical Interview for DSM-IV Dissociative Disorders (SCID-D)*. Multi-Health Systems, Toronto, 1996.

Steinberg M: *Tips and Techniques for Assessing and Planning Treatment with Dissociative Disorder Patients: A Practical Guide to the SCID-D*. Multi-Health Systems, Toronto, 1996.

Steinberg M, Hall P: *The SCID-D Diagnostic Interview and Treatment Planning in Dissociative Disorders*. Bull Menninger Clin *61:*1, 1996.

18.5 DISSOCIATIVE DISORDERS NOT OTHERWISE SPECIFIED

DAPHNE SIMEON, M.D., AND ERIC HOLLANDER, M.D.

The category of *dissociative disorders not otherwise specified* includes all the dissociative conditions whose manifestations do not meet diagnostic criteria for the better delineated, specific dissociative disorders from the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I) that is, dissociative amnesia, dissociative fugue, dissociative identity disorder, and depersonalization disorder. The DSM-IV criteria for dissociative disorder not otherwise specified are presented in Table 18.5-1. Given the surge of interest and the return of dissociation to the mainstream of psychiatry in recent years after many decades of relative neglect, it is hoped that this residual category of dissociative pathology will encourage further detection and research into dissociative syndromes that are less well understood. Importantly represented in this group are nonwestern dissociative ailments that until recently were considered more the subject of folklore and medical anthropology. It is widely known that dissociation is common in the third world, probably more so overall than in industrialized societies, and comprises an important ingredient of various culturally defined ritual and healing practices. Dissociation is culturally more accepted as a idiom of distress in these countries and therefore is often the medium through which significant dysfunction and distress may be expressed, frequently qualifying as psychiatric nosology by the criteria of mental illness. These non-Western dissociative disorders, which fall under the rubric of *dissociative trance disorder*, are to a great extent unfamiliar to Western clinicians and researchers, limiting cross-cultural understanding as well as the universal applicability of the nosological system. With the inclusion of dissociative trance disorder in the appendix of the latest edition of DSM-IV, the elucidation of these widespread, variable, and fascinating conditions appears closer.

This category is included for disorders in which the predominant feature is a dissociative symptom (i.e., a disruption in the usually integrated functions of consciousness, memory, identity, or perception of the environment) that does not meet the criteria for any specific dissociative disorder. Examples include:

1. Clinical presentations similar to dissociative identity disorder that fail to meet full criteria for this disorder. Examples include presentations in which (a) there are not two or more distinct personality states, or (b) amnesia for important personal information does not occur.
2. Depersonalization unaccompanied by derealization in adults.
3. States of dissociation that occur in individuals who have been subjected to periods of prolonged and intense coercive persuasion (e.g., brainwashing, thought reform, or indoctrination while captive).
4. Dissociative trance disorder: single or episodic disturbances in the state of consciousness, identity, or memory that are indigenous to particular locations and cultures. Dissociative trance involves narrowing of awareness of immediate surroundings or stereotyped behaviors or movements that are experienced as being beyond one's control. Possession trance involves replacement of the customary sense of personal identity by a new identity, attributed to the influence of a spirit, power, deity, or other person, and associated with stereotyped "involuntary" movements or amnesia. Examples include *latah* (Indonesia), *bebainan* (Indonesia), *latah* (Malaysia), *pibloktoq* (Arctic), *ataque de nervios* (Latin America), and possession (India). The dissociative or trance disorder is not a normal part of a broadly accepted collective cultural or religious practice.
5. Loss of consciousness, stupor, or coma not attributable to a general medical condition.
6. Ganser's syndrome: the giving of approximate answers to questions (e.g., "2 plus 2 equals 5") when not associated with dissociative amnesia or dissociative fugue.

Reprinted with permission from American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, ed 4. Washington, DC, 1994, American Psychiatric Association.

Table 18.5-1 DSM-IV Diagnostic Criteria for Dissociative Disorder Not Otherwise Specified

DISSOCIATIVE TRANCE DISORDER

Definition The diagnostic criteria for dissociative trance disorder are presented in Table 18.5-2. It encompasses a variety of dissociative phenomena that share an alteration of normal consciousness, or trance state, which is distressing, impairing, and beyond the nonpathological trance manifestations that may be indigenous to particular cultures. Dissociative trance involves a narrowing of awareness of the immediate surroundings with stereotypical behaviors or movements that are experienced as involuntary and for which there may be partial or total amnesia. The most widely written about dissociative trance of this sort is the *ataque de nervios* encountered in South America. Alternatively, *possession trance* is characterized by a transient alteration in identity whereby one's normal identity is temporarily replaced (possessed) by a spirit, ghost, deity, or other person. Descriptions of possession trance are to be found in a variety of cultures and have been most studied in India.

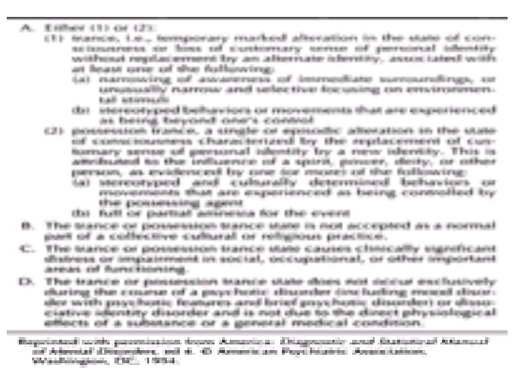

A. Either (1) or (2):
   (1) trance, i.e., temporary marked alteration in the state of consciousness or loss of customary sense of personal identity without replacement by an alternate identity, associated with at least one of the following:
      (a) narrowing of awareness of immediate surroundings, or unusually narrow and selective focusing on environmental stimuli
      (b) stereotyped behaviors or movements that are experienced as being beyond one's control
   (2) possession trance, a single or episodic alteration in the state of consciousness characterized by the replacement of customary sense of personal identity by a new identity. This is attributed to the influence of a spirit, power, deity, or other person, as evidenced by one (or more) of the following:
      (a) stereotyped and culturally determined behaviors or movements that are experienced as being controlled by the possessing agent
      (b) full or partial amnesia for the event
B. The trance or possession trance state is not accepted as a normal part of a collective cultural or religious practice.
C. The trance or possession trance state causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The trance or possession trance state does not occur exclusively during the course of a psychotic disorder (including mood disorder with psychotic features and brief psychotic disorder) or dissociative identity disorder and is not due to the direct physiological effects of a substance or a general medical condition.

Reprinted with permission from American Diagnostic and Statistical Manual of Mental Disorders, ed 4. American Psychiatric Association, Washington, DC, 1994.

Table 18.5-2 DSM-IV Research Criteria for Dissociative Trance Disorder

ATAQUE DE NERVIOS

Diagnosis and Clinical Features *Ataque de nervios* is characterized by a number of somatic symptoms including trembling, heart palpitations, a sense of heat rising to the chest and head, faintness, paresthesias, and darkened vision. There can be difficulty moving or seizure-like convulsive movements, falling to the ground, or lying still as if dead. Individuals also typically cry, shout, moan, or swear uncontrollably; become hysterical; and attempt to harm themselves or strike out at others or throw objects. The episode ends suddenly or gradually, sometimes with the intervention of others who attempt to calm down the victim, pray, or perform alcohol rubs *(alcoholado)*. During the episode there is typically a narrowing of attention and general lack of awareness of the surroundings, and after the episode there can be partial or total amnesia for what transpired. After the episode there is a return to normal consciousness or a prolonged period of exhaustion that can last for hours. Acute family conflict appears to comprise the hallmark trigger of this condition. The episode typically commences with a sense of brooding, feeling overwhelmed, and of shock, followed by the acute shift in consciousness and the narrowing of attentional focus. Overall affective and somatic alterations are pronounced in the context of a dissociated state. *Ataque de nervios* may occur as isolated incidents, lasting minutes to hours, or can become long-term and recurring, sometimes lasting for years and causing significant disability.

The inclusion of *ataque de nervios* in the dissociative disorders, with their characteristic alterations in sensorimotor control, broadens the traditional Western concept of dissociative disorders, which focuses primarily on alterations in perception, memory, and identity. It begs the broader question of whether conversion syndromes are more closely related to the dissociative disorders than their current DSM classification under the somatoform disorders would suggest. In addition, this broadened spectrum of dissociation more closely approximates the 10th revision of *International Statistical Classification of Diseases and Related Health Problem* (ICD-10), in which dissociative (conversion) disorders are grouped together. Trance and possession disorders are classified in this section of the ICD-10, and the specific criteria

are given in Table 18.5-3.

Table 18.5-3 ICD-10 Diagnostic Criteria for Dissociative (Conversion) Disorders

Etiology *Ataque de nervios* is precipitated by acute traumatic stressors, typically loss, grief, danger, fear, or vulnerability, which may vary from relatively mild to extreme. Typical triggers are arguments with a spouse, death of a close relative, drunkenness and threats from a spouse, physical or sexual abuse, an accident, or even natural disasters.

Epidemiology *Ataque de nervios* has best been studied in Puerto Rico, and a recent epidemiological study estimated their lifetime prevalence at 13.8 percent of the population. The demographic characteristics of the sufferers have been consistent in the various reports and include female sex, age over 45, low socioeconomic status (education less than high school and unemployment) and divorced, separated, or widowed marital status.

Differential Diagnosis Although there is some overlap between the symptoms of *ataque de nervios* and those of panic attacks, *ataque de nervous* is additionly characterized by alterations in consciousness and sensorimotor dissociation that render them clearly distinct. There is significant comorbidity between *ataque de nervios* and other Axis I disorders, in approximately 60 percent of sufferers, including depressive disorders, generalized anxiety disorder, panic disorder, agoraphobia, and posttraumatic stress disorder. *Ataque de nervios* largely accounts for the high rates of somatization found in these cultures, and the expression of distress in such a somatic form appears to be more culturally sanctioned and accepted than the stigma of mental illness. Nevertheless, individuals who suffer from *ataque de nervios* commonly experience high levels of distress and dysfunction and the high rates of associated psychopathology contradict the notion that these are normal cultural expressions of suffering.

Treatment Typically, others perceive the victim of *ataque de nervios* as being in significant distress and requiring some form of assistance. If the trigger was overwhelming, such as death or marital abuse, home-based remedies may be attempted, such as emotional support, concrete assistance, or interventions to resolve the conflict. However, if the stressor is not deemed to be as significant, it is more likely that professional help will be sought, either in nonpsychiatric or in psychiatric settings, and *ataque de nervios* patients may frequently be prescribed antidepressant or antianxiety medications or may receive some type of spiritual healing or counseling.

POSSESSION TRANCE

Diagnosis and Clinical Features Possession trance is a category of dissociative disorders that encompasses a variety of different cultural presentations throughout the world, best characterized in African and South Asian cultures. It is defined by the presence of alterations in consciousness, behavior, memory, and, in particular, identity that go beyond the culturally accepted norms and bring dysfunction and distress to the sufferer.

During the trance state a variety of behaviors similar to those seen in *ataque de nervios* can occur: uncontrolled bodily movements such as shaking, flailing, or falling to the ground; moaning and shrieking; violent behaviors directed at the self or others; and derogatory or menacing verbalizations. Paresthesias, dizziness, and other somatic symptoms may also occur, although they may be less pronounced and intense than in *ataque de nervios*. A personality emerges that is distinct from that characteristic of the individual, and these personalities are typically regarded as arising outside of the person. This is in contrast to dissociative identity disorder, in which the identities are experienced and viewed as distinct aspects of some sense of self. The possessing personalities may not disclose their identity immediately, until tricked into doing so by family members or healing practitioners, and they typically involve deities, demons, spirits, ghosts, or deceased family members, neighbors, or friends who were known to the particular individual and underwent similar stress. The individual then takes on their particular attributes. Demands are often made by this possessing personality for specific changes in the individual's circumstances and for better treatment, and the possessing entity may agree to recede if these demands are met. These verbalizations tend to be more purposeful, organized, and coherent than the affective storm quality that characterizes *ataque de nervios*. After the episode the individual almost always collapses, loses consciousness, is disoriented, and has partial or total amnesia for the events that occurred and may remain dazed, exhausted, and confused for a number of hours. This aftermath tends to be more pronounced than in *ataque de nervios*.

In contrast to the individualistic character and the stricter personal boundaries of western cultures, it has been proposed that the more porous and fluid characteristics of Indian culture, which is more open to external human and spiritual influences, contribute to this particular pathological manifestation. Rurality and lack of formal education also probably contribute to the phenomenon and permit the often helpless and powerless individual to express, in a dissociated form, feelings of fear, anger, vulnerability, and grief that may be more effectively heard and attended to than otherwise. Indeed, the very low standing of women in the low strata of these cultures, whose worth is often at best equated with a dowry, often renders it very unlikely that their distress would merit attention in other shapes or forms. Typically the episodes recur at variable durations and intervals, may involve different consecutive possessing personalities, and may continue for days or weeks until the underlying stressors are addressed, such as by a visit to the young bride's hometown or by more respectful treatment by her in-laws. If not attended to in some fashion the possession attacks can in some cases become severely disabling.

Etiology Possession trance in India typically occurs in young adult women of underpriviledged socioeconomic status and low caste in the context of prolonged subacute stress usually surrounding the transition from the family of origin to wedlock in the absence of a solid support network. The transition typically demands a multitude of difficult adaptations, including leaving one's home town, family, and prior status and adjusting to the demands made upon a new bride by her in-laws, forced marriages, consummation of the marriage, as well as marital conflict, abuse, or neglect. Other circumstances that have been described in relation to possession include a variety of traumas, such as death of the husband or children; discovery of incest, sexual abuse, murder, or suicide of close female friends; or natural disasters. Although much has been written about earlier antecedents to these states, including an underlying primitive hysterical personality, oedipal dynamics, or the possibility of reactivated early childhood abuse, none of these is well substantiated. The most obvious determinant is the largely culturally determined prolonged and subacute trauma that precedes the condition.

Epidemiology Possession trance disorders have been described in India, Bangladesh, Nepal, Taiwan, Japan, Hong Kong, Malaysia, Niger, and Brazil; they have been most extensively described in the standard psychiatric literature in India. A review of 2651 case records of all patients seen in the adult outpatient clinic of a large Indian hospital found that 62 patients definitely suffered from dissociative disorders. Of those, the vast majority (90 percent) were atypical dissociative disorders that would not fit criteria for the established Western category of major dissociative disorders, and at least 10 percent were cases of possession trance. The prevalence of possession syndromes has been estimated at 1 to 4 percent of the population in community surveys, although the methodologies may be flawed.

Differential Diagnosis In terms of associated psychopathology and not studied in a systematic fashion, hysterical or paranoid states are more commonly diagnosed in Indian possession in contrast to the high prevalence of anxiety and depression in Puerto Rican *ataque de nervios*; in both, some associated psychopathology is often present.

Possession trance has been compared to Western dissociative identity disorder, and it appears that although the disorders have similarities they comprise distinct entities. Possession trance is a more time-limited condition that relates to more current stressors, and the basic organization of the underlying personality is more intact. Indeed, cases of dissociative identity disorder have been found to be much rarer in India than in the United States, with the first case reported in 1956 and a recently estimated prevalence of 0.15 of 1000 psychiatric patients. However, the degree to which these two disorders are distinct remains not fully understood. It may

be due to a true difference in childhood abuse, as the neglect of children is probably more common in India. Alternatively, it may be that the presence of possession in a patient from the West is more likely to be diagnosed as a psychotic state whereas Indian patients with dissociation are unlikely to manifest different personal identities because of cultural and religious beliefs that emphasize external influences.

Treatment Treatment is commonly obtained within the family setting, as family members attempt to provide support and appease the possessing spirits by meeting some of their demands. Professional intervention is more likely be sought from traditional healers who perform different types of exorcism that are often successful in ridding the individual of the possession. Sometimes professional psychiatric help may be sought.

BRAINWASHING

Definition The DSM-IV describes this dissociative disorder not otherwise specified as "states of dissociation that occur in individuals who have been subjected to periods of prolonged and intense coercive persuasion (e.g., brainwashing, thought reform, or indoctrination while captive)" (Table 18.5-1). The 10th edition of *Merriam Webster's Collegiate Dictionary* defines brainwashing as "a forcible indoctrination to induce someone to give up basic political, social or religious beliefs and attitudes and to accept contrasting regimented ideas." The term *brainwashing* was first coined by an American journalist in the 1950s from Chinese ideographs that would more accurately have been translated as "thought reform." Indeed, the terms *brainwashing, thought reform, coercive persuasion*, and *mind control* are for all practical purposes used synonymously and interchangeably. Depending on differing perspectives, the degrees and methods of what constitutes brainwashing vary. In general, psychiatrists have largely stayed outside of such issues, which involve larger ideological, societal, political, and religious implications. The average psychiatrist is typically not better informed in this area than an educated lay person, and with few exceptions, psychiatrists have for the most part been uncomfortable with this issue.

Diagnosis and Clinical Features Brainwashing occurs largely in the settings of political reform, as has been lengthily described with the cultural revolution in communist China, war imprisonment, torture of political dissidents, terrorist hostages, and, more familiarly in Western culture: totalitarian cult indoctrination. It implies that under conditions of adequate stress and duress, individuals can be made to comply with the demands of those in power, thereby undergoing major changes in their personality, beliefs, and behaviors. People submitted to such conditions can undergo considerable harm, including loss of health and life, and typically manifest a variety of posttraumatic and dissociative symptoms. The first stage in coercive processes has been likened to the artificial creation of an identity crisis, with the emergence of a new pseudoidentity that manifests characteristics of a dissociative state. Under circumstances of extreme and malignant dependency, overwhelming vulnerability, and danger to one's existence, individuals develop a state characterized by extreme idealization of their captors, with ensuing identification with the aggressor and externalization of their superego, regressive adaptation known as *traumatic infantilism*, paralysis of will, and a state of frozen fright. The coercive techniques that are typically employed to induce such a state in the victim have been amply described and include isolation of the subject, degradation, control over all communications and basic daily functions, induction of fear and confusion, peer pressure, assignment of repetitive and monotonous routines, unpredictibility of environmental supplies, renunciation of past relationships and values, and various deprivations. Even though physical or sexual abuse, torture, and extreme sensory deprivation and physical neglect can be parts of this process, they need not necessarily occur to define a coercive process. As a result, victims who have been extensively studied manifest extensive posttraumatic and dissociative symptomatology, including a drastic alteration of their identity, values, and beliefs; a reduction of cognitive flexibility with a regression to simplistic perceptions of good-evil and dominance-submission; a numbing of experience and blunting of affect; trance-like states and diminished environmental responsiveness; and in some cases more severe dissociative symptoms such as amnesia, depersonalization, and shifts in identity.

It has been suggested that an individual subjected to severe inescapable coercion can have three distinctive response patterns: to sham compliance, to submit and identify, or to defend against persuasion. Individuals capable of the latter option appear more able to spontaneously use autohypnotic techniques, to become involved in restitutive hypnotic fantasy, and to dissociate from their suffering as if it were happening to someone else. Thus, in a paradoxical fashion the use of such dissociative techniques might be protective against the development later of more severe psychopathology and deterioration and has been proposed as a preventive training technique for those who are at risk of becoming victims of coercion (e.g., military personnel).

Proponents of religious freedom and the right to choose membership in cults and new religious movements have argued that in the absence of physical coercion and in the presence of a real option to depart, the term *brainwashing* is misapplied. They further argue that any set of beliefs, religious or otherwise, involves an element of arbitrary faith and a background of indoctrination that may be sanctioned by the general culture, thus blurring the boundaries of coercion. As members of such groups, some individuals may find a sense of stability and direction that society may otherwise not be able to offer them. Opponents of this view argue that coercion is primarily a mental process that need not involve extreme physical means, and that adoption of the former view approaches blaming the victim. Individuals who have greater predisposing psychopathology; identity issues; family conflicts; difficulties negotiating their aggressive, sexual, or individuating needs; or are in transitional stressful phases of life are then in effect held wholly responsible for submitting to the coercion. No one disputes that predisposing psychopathology is one of the factors that can affect the outcome of the coercive process, but the process nonetheless remains the same.

Treatment The treatment of the victims of coercion can vary considerably depending on their particular background, the circumstances involved, and the setting in which help is sought. Although there are no systematic studies in this domain, basic principles involve validation of the traumatic experience and coercive techniques used, cognitive reframing of the events that transpired, exploration of preexisting psychopathology and vulnerabilities where this is applicable, and general techniques used in treating posttraumatic and dissociative states. In addition, family interventions and therapy may be required, at least in cases of cult indoctrination, as significant family duress and disruption is a common occurrence.

GANSER'S SYNDROME

Definition *Ganser's syndrome* is a fascinating, poorly studied, and uncommon dissociative condition whose exact etiology, classification, and symptom constellation are still debated in the literature. Based on the gradual accumulation of relatively limited data and a clearer conceptualization of the disorder, Ganser's syndrome was moved from its status as a factitious disorder in the third edition of DSM (DSM-III) to that of dissociative disorders not otherwise specified in the revised third edition (DSM-III-R), and has remained so in DSM-IV (Table 18.5-1).

Diagnosis and Clinical Features The hallmark symptom of Ganser's syndrome is paralogia, the giving of approximate answers to questions, as originally described by Sigbert Ganser in 1897 in a paper entitled "Concerning an Unusual Hysterical Confusional State." Although afflicted individuals typically know their names, they often only approximate other key personal information such as age, address, and occupation. They also give approximate answers to questions like the number of days in a week or months in a year, to simple calculations (e.g., 3 + 1 = 5), and perform approximate naming of objects (e.g., pencil for pen, picture for stamp). It is generally agreed that for the diagnosis of the syndrome additional dissociative symptoms must be present, including disorientation, dazing or clouding of consciousness, and amnesia for past events or typically, once the episode clears, for prior responses. Various case reports and literature reviews tend to concur that the presence of these additional symptoms may be variable so that the Ganser phenomenon may occur on a continuum, at times possibly warranting the diagnosis of *Ganser's symptoms* instead of the full-blown Ganser's syndrome. The current edition of DSM does not specify criteria for the severity, duration, number, and types of symptoms that must be present to definitively make the diagnosis of Ganser's syndrome, probably because conclusive data are lacking.

Epidemiology According to all reports, Ganser's syndrome is much more common in men than in women. One study reported the mean age of subjects as 35 years old and the mean duration of the episode as 1 month. In addition to Western cultures, it has been described in the Far East and in India. Its incidence in the West may have declined over the past century, paralleling the generally accepted decline in hysteria.

Etiology Precipitating stressors can often be identified, including an organic or mental illness, personal conflicts, and, not uncommonly, litigation, imprisonment, and disability claims. The latter have given rise to considerable debate about whether the syndrome may involve conscious faking or secondary gains, such as in factitious disorders or malingering. However, extensive review of this topic has fairly consistently excluded this as a significant component of the syndrome in the large majority of cases.

Differential Diagnosis Part of the difficulty in making the discrete diagnosis is the very common overlap with major organic or psychiatric disorders, most commonly psychosis, organic brain syndromes such as head trauma, seizures, or other of the central nervous system diseases major depressive disorder, use disorders. A careful review of all convincing cases of Ganser's syndrome in the literature revealed that the majority of patients suffer such comorbidity. At other times, Ganser's syndrome has been described as the prelude to an acute and severe psychotic or depressive decompensation. It does appear, however, that these underlying disorders are not adequate in accounting for the manifestations of Ganser's syndrome, as these patients do not manifest them before or after the resolution of the episode, are in a dissociated state at the time, and have normal enough intelligence to intellectually know the correct answers. Rather, it has been conceptualized that the underlying condition may serve as an acute stressor that in vulnerable individuals precipitates a hysterical overlaid reaction.

In considering the differential diagnosis of Ganser's syndrome, Ganser's paralogia must be differentiated primarily from simple hysterical or depressive pseudodementia, from the organic dysphasias or dementias, from the thought disorders of psychosis, and from the confabulation of Korsakoff's syndrome. In addition to the general consideration of the clinical picture, this differential is based on the presence of the other accompanying acute dissociative symptoms that are typically absent in other conditions. A comprehensive review of the literature has found that amnesia is present in 69 percent of subjects, disorientation or confusion in 62 percent, conversion symptoms in 54 percent, fugue in 29 percent, and loss of personal identity in 23 percent. In addition, hallucinations occur in about half the subjects, leading some to speculate that the syndrome might be a type of brief reactive psychosis to stress although hallucinations are well known to also occur in severe dissociative conditions. The types of amnesia can be variable, including amnesia for past events prior to the episode, amnesia during the episode, anterograde amnesia, and most typically amnesia for the episode as the person recovers and the sensorium clears.

Treatment It has been speculated that the paralogia of Ganser's syndrome, consistent with the variety of other dissociative symptoms present, is based on trance logic, that is, a processing of information that is reminiscent of primary process, that is unaffected by inherent contradictions, and that allows the subject to simultaneously know and not know troubling realities. Consequently, the indicated treatment parallels that of other dissociative states. An insistent confrontation of the presenting contradictions is probably not fruitful, and can lead either to an intensification of the paralogia or to a compliant suppression of the paralogia, which may further confuse the presentation. A structured, protective, and supportive environment; treatment of the comorbid conditions; and exploration of the precipitating stressors have been advocated and may lead to a rapid resolution of the syndrome.

SUGGESTED CROSS-REFERENCES

Further descriptions of the nature and the function of dissociative mechanisms are found in Chapter 6 dealing with the theories of personality and psychopathology, and in Chapter 9, dealing with the classification of mental disorders. The diagnostic distinction between dissociative disorders and other mental disorders is clarified in Chapter 12 on schizophrenia and in Chapter 15 on posttraumatic stress disorder and anxiety disorders. Chapter 16 on the somatoform disorders, provides a detailed description of the somatic symptoms that in this chapter are viewed as manifestations of dissociation. Expanded descriptions of the various psychotherapeutic approaches appear in Chapter 30 which is devoted to psychotherapies.

SECTION REFERENCES

Adityanjee, Raju GSP, Khandelwal SK: Current status of multiple personality disorder in India. Am J Psychiatry *146:*1607, 1989.

Aktar S: Four culture-bound psychiatric syndromes in India. Int J Soc Psychiatry *34:*70, 1988.

Anthony D, Robbins T: Law, social science and the “brainwashing” exception to the First Amendment. Behav Sci Law *10:*5, 1992.

Balson PM, Dempster CR, Brooks FR: Auto-hypnosis as a defense against coercive persuasion. Am J Clin Hypn *26:*252, 1984.

*Cardena E: Trance and possession as dissociative disorders. Transcult Psychiatr Res Rev *29:*287, 1992.

Cardena E, Lewis-Fernandez R, Bear D, Pakianathan I, Spiegel D: Dissociative disorders. In *DSM-IV Sourcebook*, vol 1, AJ Frances, editor. American Psychiatric Press, Washington, DC, 1994.

Castillo RJ: Spirit possession in South Asia, dissociation or hysteria? Part I: Theoretical background. Cult Med Psychiatry *18:*1, 1994.

Castillo RJ: Spirit possession in South Asia, dissociation or hysteria? Part II: Case histories. Cult Med Psychiatry *18:*141, 1994.

Chowdhury AN, Nath AK, Chakraborty J: An atypical hysteria epidemic in Tripura, India. Transcult Psychiatr Res Rev *30:*143, 1993.

Cocores JA, Santa WG, Patel MD: The Ganser syndrome: Evidence suggesting its classification as a dissociative disorder. Int J Psychiatry Med *14:*47, 1984.

Coons PM: The differential diagnosis of possession states. Dissociation *6:*213, 1993.

Eguchi S: Between folk concepts of illness and psychiatric diagnosis: Kitsune-tsuki (fox possession) in a mountain village of Western Japan. Cult Med Psychiatry *15:*421, 1991.

*Epstein RS: Ganser syndrome, trance logic and the question of malingering. Psychiatr Ann *21:*238, 1991.

Freed RS, Freed SA: Ghost illness in a North Indian village. Soc Sci Med *30:*617, 1990.

*Guarnaccia PJ, Canino G, Rubio-Stipec M, Bravo M: The prevalence of *ataques de nervios* in the Puerto Rico disaster study: The role of culture in psychiatric epidemiology. J Nerv Ment Dis *181:*157, 1993.

Guarnaccia PJ, Good BJ, Kleinman A: A critical review of epidemiological studies of Puerto Rican mental health. Am J Psychiatry *147:*1449, 1990.

Guarnaccia PJ, Rubio-Stipec M, Canino G: *Ataques de nervios* in the Puerto Rican diagnostic interview schedule: The impact of cultural categories on psychiatric epidemiology. Cult Med Psychiatry *13:*275, 1989.

Kirmayer LJ: Pacing the void: Social and cultural dimensions of dissociation. In *Dissociation: Body, Mind and Culture*, D Spiegel, editor. American Psychiatric Press, Washington, DC, 1994.

Lewis-Fernandez R: The proposed DSM-IV trance and possession disorder category: Potential benefits and risks. Transcult Psychiatr Res Rev *29:*301, 1992.

*Lewis-Fernandez R: Culture and dissociation: A comparison of *ataque de nervios* among Puerto Ricans and possession syndrome in India. In *Dissociation: Culture, Mind and Body*, D Spiegel, editor. American Psychiatric Press, Washington, DC, 1994.

Maleson FG: Dilemmas in the evaluation and management of religious cultists. Am J Psychiatry *138:*925, 1981.

*Pope HG Jr, Oliva PS, Hudson JI, Bodkin JA, Gruber AJ: Attitudes toward DSM-IV dissociative disorders diagnoses among board-certified American psychiatrists. Am J Psychiatry *156:*321, 1999.

Post SG: Psychiatry and ethics: The problematics of respect for religious meanings. Cult Med Psychiatry *17:*363, 1993.

Ross CA: The dissociated executive self and the cultural dissociation barrier. Dissociation *4:*55, 1991.

Saxena S, Prasad KVSR: DSM-III subclassification of dissociative disorders applied to psychiatric outpatients in India. Am J Psychiatry *146:*261, 1989.

Spiegel D, Cardena E: Disintegrated experience: The dissociative disorders revisited. J Abnorm Psychol *100:*366, 1991.

Steinberg M: Transcultural issues in psychiatry: The *ataque* and multiple personality disorder. Dissociation *3:*287, 1990.

Tsoi WF: The Ganser syndrome in Singapore: A report of ten cases. Br J Psychiatry *123:*567, 1973.

Varma VK, Bouri M, Wig NN: Multiple personality in India: Comparison with hysterical possession state. Am J Psychother *35:*113, 1981.

*West LJ: A psychiatric overview of cult-related phenomena. J Am Acad Psychoanal *21:*1, 1993.

Whitlock FA: The Ganser syndrome. Br J Psychiatry *113:*19, 1967.

CHAPTER 19. NORMAL HUMAN SEXUALITY AND SEXUAL AND GENDER IDENTITY DISORDERS

19.1 NORMAL HUMAN SEXUALITY

19.1A NORMAL HUMAN SEXUALITY AND SEXUAL DYSFUNCTIONS

VIRGINIA A. SADOCK, M.D.

Sexuality is determined by anatomy, physiology, psychology, the culture in which one lives, one's relationships with others, and developmental experiences throughout the life cycle. It includes the perception of being male or female and all those thoughts, feelings and behaviors connected with sexual gratification and reproduction including the attraction of one person to another.

The 1990s have seen advances in the field of sexuality in the areas of pharmacology and psychology and in the study of the interaction of sex and the social milieu. Significant new developments are the availability of medications that enable men to gain and maintain erections later in their lives and hormonal therapies that allow women to have pleasurable coitus postmenopausally. These medications have helped breach the taboo against sex in elderly adults. Theories on the psychology of sex have examined compulsive sexual behavior, which is not currently an official diagnosis in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Finally, a 1994 study of sex in the United States (discussed below) yielded the most authoritative survey of sexual practices to date and placed the sexual behavior of Americans in a social context.

HISTORY

Cultural mores regarding sexual behavior have varied throughout the history of Western civilization. Attitudes have oscillated between the liberal and the puritanical, between the acceptance and the repression of human sexuality. During the past several decades, the prevalent attitudes toward sex in the United States have been markedly liberal. However, recent studies indicate a trend toward accepting more-conservative values. That shift is attributed largely to the fear of acquired immune deficiency syndrome (AIDS). One poll reported that 40 percent of Americans are concerned about contracting AIDS and are altering their sexual behavior because of that fear. The greatest concern was expressed by young adults, who are now more likely to use condoms as a precaution and to choose their sexual partners with greater care. In 1997 the rate of teenage pregnancy declined for the first time in 40 years, and in 1998 the number of teenagers who had sexual intercourse fell below 50 percent for the first time in a decade. Conservative segments of society emphasize abstinence before marriage as the answer to the fear of AIDS. The recurrence of conservative attitudes in response to the threat of illness has parallels in history. The sexual liberality of the Renaissance ended when syphilis swept the European continent and became a major argument for chastity among proponents of the Reformation. Other factors that predispose to more-restrictive mores are periods of economic recession that tend to bring people to more puritanical positions and limited gratification when sexual freedom is used as a substitute for intimate relationships. Few of these issues have been resolved definitively in the form of new social mores, however, and the legacies of the sexual revolution of the 1960s and 1970s exert a strong effect on current sexual behavior.

That particular sexual revolution derived from social and scientific sources. The advent of effective birth control methods and legalized abortion clearly differentiated the pleasure of sexual activity from its procreative function. The feminist movement attacked the double standard for acceptable sexual behavior for men and women, encouraged women to accept sexual responsibility for the gratification of their needs, and challenged society to reevaluate stereotyped male and female roles. The women's movement also focused attention on rape and incest. Gerontologists and elderly people alike have drawn attention to the sexual needs of the aged. Middle-class adolescents became sexually active, and gay rights groups urged acceptance of their sexual orientation, and succeeded in 1980 in having homosexuality dropped as a diagnostic category in the revised third edition of DSM (DSM-III-R).

Concurrent with the cultural changes of the sexual revolution there was growth in scientific research into sexual physiology and sexual dysfunctions. William Masters and Virginia Johnson published their pioneering work on the physiology of sexual response in 1966 and reported on their program for treating sexual complaints in 1970. Most medical centers now have programs specifically geared to the treatment of sexual dysfunctions.

Historically, problems of sexual conflict and sexual dysfunction have always been the province of psychiatry. Such pioneers as Havelock Ellis (Fig. 19.1a-1), Richard Krafft-Ebing (Fig. 19.1a-2), and Sigmund Freud focused broadly on human sexuality. Later workers focused more intensively on sexual physiology and dysfunctions. Dysfunction problems are particularly distressing to patients and have often been resistant to treatment. The current approach to sexual dysfunctions reflects the cultural and scientific developments of recent years, the development of specific techniques for the treatment of these problems, the historical interest of psychiatry in this area, and the recognition of its importance in psychiatric practice.

FIGURE 19.1a-1 Havelock Ellis, 1859–1939. In his book *Studies in the Psychology of Sex* (1896), Ellis recorded examples of normal and abnormal sexuality. It remains a classic in the field of sexology. (Courtesy of New York Academy of Medicine.)

FIGURE 19.1a-2 Richard von Krafft-Ebing (1840–1903), a psychiatrist who published a classic text, *Psychopathia Sexualis* (1898), in which he documented every variation in sexuality, including zoophilia, necrophilia, urolagnia, and lust murder among others. Case reports were so lurid and detailed that early editions were published in Latin. (Courtesy of New York Academy of Medicine.)

NORMAL SEXUALITY

Anatomical and Physiological Bases Knowledge about the organs of sexuality and the normal physiological sequence of male and female response is necessary for an informed understanding of the sexual dysfunctions. In fact, over the past decade greater emphasis has been placed on the genetic, neuroanatomical, and neurochemical model of human sexuality than on psychological and social factors.

Anatomy of the Male The external genitalia of the normal adult male include the penis, scrotum, testes, epididymis, and parts of the vas deferens. Internal components include the vas deferens, ejaculatory ducts, and prostate gland.

Freud referred to the penis as the executive organ of sexuality. Since antiquity, culture has represented the penis in a variety of art forms. In ancient Greece, the cults of Dionysus, Priapus, and the satyrs used the phallus as a recurrent symbol of fertility and rejuvenation. The word “penis” has been traced from the Latin, meaning variously “tail” or “to hang,” and refers to the pendant position of the organ in its resting or flaccid state. The size of the penis varies within a fairly constant range, but sex researchers over the years have disagreed on the dimensions of the range. All agree, however, that concern over the size of the penis is practically universal among men. Masters and Johnson report a range of 7 to 11 cm in the flaccid state, and 14 to 18 cm in the erect state. Of particular interest was their observation that the flaccid dimension bears little relation to the erect dimension: the smaller penis erects proportionally more than the larger one.

Circumcision, in which the prepuce is surgically removed, has been practiced for centuries as a religious rite by Jews and Moslems and is a common medical procedure in the United States today. The circumcised penis with its exposed glans was once believed to be less sensitive because of cornification of the epithelium. In laboratory studies, however, researchers have found no difference in tactile threshold between the circumcised penis and the uncircumcised penis. Intravaginally the prepuce of the uncircumcised penis remains retracted behind the glans during penile thrusting, dispelling the myth that premature ejaculation may be more common in uncircumcised men because of increased stimulation caused by preputial movements. In 1999, the American Academy of Pediatrics recommended that male circumcision not be performed as a routine procedure except for religious reasons.

Ejaculation is the forceful propulsion of semen and seminal fluid from the epididymis, vas deferens, seminal vesicles, and prostate into the urethra. The dilation of the prostatic urethra and the passage of fluid into the penile urethra provide the man with a sensation of impending climax, the emission phase of the ejaculatory process. Indeed, once the prostate contracts, ejaculation is inevitable. The ejaculate is then propelled through the penile urethra by contractions of the striated pelvic and perineal muscles. This phase of ejaculation is essentially under somatic efferent control. The ejaculate consists of about 1 teaspoon (2.5 mL) of fluid and contains about 120 million sperm cells. It is believed that the larger the ejaculate, the more pleasurable the orgasm—but this belief is highly subjective. *Orgasm*, the sense of pleasure usually associated with ejaculation, is thought to be a cortical experience.

Anatomy of the Female The external genitalia of the normal female, also called the *vulva*, include the mons pubis, major and minor lips, clitoris, glans, vestibule of the vagina, and vaginal orifice. The internal system includes the ovaries, fallopian tubes, uterus, and vagina.

The word *vagina* comes from the Latin word meaning “sheath.” The vagina is usually collapsed, a potential rather than an actual space. About 8 cm long, the vagina extends from the cervix of the uterus above to the vestibule of the vagina or vaginal opening below. In most virgins a membranous fold, the hymen, separates the vestibule and opening from the rest of the vaginal canal. The mucous membrane lining the vaginal walls rests in numerous transverse folds. To accommodate the penis during sexual intercourse, the vagina expands in both length and width. After menopause, because circulating estrogen concentrations decrease, the vagina loses much of its elasticity.

Hippocrates first described the clitoris in the medical literature, referring to it as the site of sexual excitation. Masters and Johnson described the clitoris as the primary female sexual organ, because orgasm depends physiologically on adequate clitoral stimulation. Anatomically, the clitoris has a nerve net that is proportionally three times as large as that of the penis.

Alfred Kinsey found that when women masturbate, most prefer clitoral stimulation. That finding was refined further by Masters and Johnson, who reported that women prefer the shaft of the clitoris to the glans, because the glans is hypersensitive if stimulated excessively.

The clitoral prepuce is contiguous with the labia minora and during coitus, the penis does not stimulate the clitoris directly. Rather, penile thrusting exerts traction on the minor lips, which in turn stimulate the clitoris sufficiently for orgasm. During heightened excitement, just before orgasm, the clitoris retracts under the clitoral hood because of contraction of the ischiocavernosus muscles. Retracting, the clitoris moves away from the vaginal barrel, which makes clitoral-penile contact impossible. The size of the clitoris varies considerably and is unrelated to the sexual responsiveness of a particular female.

G SPOT In 1950 Ernst Graefenberg described an area surrounding the female urethra in the anterior wall of the vagina that has come to be called the G spot. About 0.5 to 1 cm in size, it becomes engorged during sexual stimulation. Many women report that stimulation of the area is highly pleasurable and in some, can induce orgasm. Graefenberg believed that the tissue here was analogous to the prostate and might account for the spurt of fluid during orgasm reported by some women, similar to male ejaculation.

Innervation of Sex Organs Innervation of the sexual organs is mediated primarily through the autonomic nervous system (ANS). Penile tumescence occurs through the synergistic activity of two neurophysiologic pathways. A parasympathetic (cholinergic) component mediates reflexogenic erections via impulses that pass through the pelvic splanchnic nerves (S2, S3, and S4). A thoracolumbar, mainly sympathetic pathway transmits psychologically induced impulses. Both parasympathetic and sympathetic mechanisms are thought to play a part in relaxing the smooth muscles of the penile corpora cavernosa, which allow the penile arteries to dilate and cause the inflow of blood that results in penile erection. Relaxation of cavernosal smooth muscles is aided by the release of nitric oxide, an endothelium-derived relaxing factor. Clitoral engorgement and vaginal lubrication also result from parasympathetic stimulation that increases blood flow to genital tissue. Adrian Zorgniotti has compared the erection phenomenon that results from increased penile blood inflow and decreased blood outflow with the inflation of an automobile tire, which requires an inflow of air and an intact casing.

Evidence indicates that the sympathetic (adrenergic) system is responsible for ejaculation. Through its hypogastric plexus the adrenergic impulses enervate the urethral crest, the muscles of the epididymis, and the muscles of the vas deferens, seminal vesicles, and prostate. Stimulation of the plexus causes emission. In women the sympathetic system facilitates the smooth muscle contraction of the vagina, urethra, and uterus that occurs during orgasm.

The ANS functions outside of voluntary control and is influenced by external events (e.g., stress, drugs) and internal events (hypothalamic, limbic, and cortical stimuli). It is not surprising, therefore, that erection and orgasm are so vulnerable to dysfunction.

Endocrinology From the time of conception, hormones play a major role in human sexual development (Fig. 19.1a-3). Unlike the fetal gonads, which are under chromosomal influence, the fetal external genitalia are very susceptible to hormones. Recently a gene was found on the X chromosome that is believed capable of disrupting the normal development of male genitals. The finding requires replication.

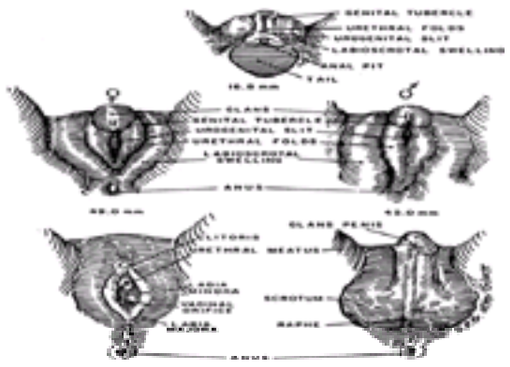

FIGURE 19.1a-3 Differentiation of male and female external genitalia from indifferent primordia. Male differentiation occurs only in the presence of androgenic stimulation during the first 12 weeks of fetal life. (From Van Wyk and Grumbach, 1968. Reprinted with permission from Brobeck JR, editor: *Best and Taylor's Physiological Basis of Medical Practice*, ed 9. Williams & Wilkins, Baltimore, 1973.)

Exogenous hormonal administration can cause external genital development inconsistent with the fetal sex gland development. For instance, if the pregnant mother receives sufficient exogenous androgen, a female fetus possessing an ovary can develop external genitalia resembling those of a male. Fetal, maternal, or exogenous hormones administered to a pregnant woman may all affect development of the external genitalia of the fetus. Deprived of male and female gonads and the respective hormones testosterone and estrogen, the human adult does not develop normal secondary sexual characteristics, is incapable of reproduction, and, in the case of the female, does not develop a menstrual cycle.

Testosterone is the hormone believed to be connected with libido in both men and women. In men stress is inversely correlated with testosterone blood concentration. Other factors, such as sleep, mood, and lifestyle, influence circulating levels of the hormone.

The release of testosterone in men is under the control of the hypothalamic-gonadal-pituitary axis. The hormone is secreted in a pulsatile manner and in a diurnal rhythm, with the highest levels occurring in the morning and the lowest levels in the evening. Normal concentrations range from 270 to 1100 ng/dL. Decreased testosterone concentrations are apparent by age 50 and proceed at the rate of about 100 ng/dL per decade. However, many healthy, aging men never become hypogonadal. It has also been suggested that the sensitivity of androgen receptor sites decreases in the aging male.

Administration of androgens to patients complaining of sexual dysfunction is usually futile if normal hormonal function is present. Androgen administered to men complaining of loss of potency and loss of libido is usually unsuccessful unless testosterone concentrations are below normal, and administration to women may precipitate disturbing virilization. Many clinicians correct the hormone deficiency of the postmenopausal period with estrogen replacement therapy. Testosterone has been used in combination with estrogen in women who do not respond to estrogen alone. The combination is especially useful in treating headache, depression, and reduced libido. Oxytocin, secreted by the hypothalamus, stimulates lactation and uterine contractions. It may enhance sexual activity. Plasma oxytocin concentrations increase in men and women during orgasm.

Diethylstilbestrol Diethylstilbestrol (DES), an androgenic steroid, was prescribed in the 1950s and 1960s for pregnant women with threatened abortion. However, the drug had untoward effects on the children (especially female children) born to these mothers. Reports of cervical and uterine abnormalities were reported in women and reproductive tract abnormalities in men. The children (called DES daughters and sons) organized a National DES Education Program sponsored by the National Cancer Institute to provide information on the potential medical problems confronting those born to DES mothers.

Brain and Sexual Behavior Experimentation with animals has demonstrated that the limbic system is directly involved with elements of sexual functioning. In all mammals the limbic system is involved in behavior required for self-preservation and the preservation of the species.

Chemical or electrical stimulation of various sites of the limbic system—the lower part of the septum and the contiguous medial preoptic area, the fimbria of the hippocampus, the mammillary bodies, and the anterior thalamic nuclei—have all elicited penile erection. The hippocampus is believed to influence genital tumescence and affect the regulation of the release of gonadotropins. Stimulation of the amygdala in primates initiates first oral (chewing, lip smacking) and then genital (penile erection) behavior. Researchers have stated that the closeness of those functions may derive from the evolutionary fact that the olfactory sense was strongly involved in both feeding and mating. They speculate that the evolution of the third subdivision of the limbic system may reflect a shift in importance from olfactory contact to visual communication in sociosexual behavior.

Brain Neurotransmitters A vast array of neurotransmitters are produced by the brain. They include dopamine, epinephrine, norepinephrine, and serotonin. All affect sexual function. For example, an increase in dopamine is presumed to increase libido. Serotonin produced in the upper pons and midbrain is presumed to have an inhibitory effect on sexual function. Basic science and clinical research on brain neurotransmitters and their effects on behavior (including sex) is one of the most rapidly expanding fields. Some of those findings are summarized in Table 19.1a-1.

	Adrenergic Impulses		Cholinergic Impulses
Effector Organs	Receptor Type	Responses	Responses
Urinary bladder			
Detrusor	β_2	Relaxation (usually) +	Contraction +++
Trigone and sphincter	α_1	Contraction ++	Relaxation ++
Ureter			
Motility and tone	α_1	Increase	Increase (?)
Uterus	α_1, β_2	Pregnant: contraction (α_1), relaxation (β_2); nonpregnant: relaxation (β_2)	Variable
Sex organs, male	α_1	Ejaculation +++	Erection +++
Skin			
Pilomotor muscles	α_1	Contraction ++	
Sweat glands	α_1	Localized secretion +	Generalized secretion +++

Adapted with permission from Goodman and Gilman's The Pharmacological Basis of Therapeutics, ed 8, A Goodman Gilman, TW Rall, AS Nies, P Taylor, editors. Pergamon, New York, 1990.

Table 19.1a-1 Responses of Sex Organs to Autonomic Nerve Impulses

PSYCHOSEXUALITY

Sexuality and total personality are so entwined that to speak of sexuality as a separate entity is virtually impossible. The term “psychosexual” is therefore used to describe personality development and functioning as these are affected by sexuality. *Psychosexual* applies to more than sexual feelings and behavior, and it is not synonymous with libido in the broad Freudian sense.

Freud's generalization that all pleasurable impulses and activities are originally sexual has given lay people a somewhat distorted view of sexual concepts and has given psychiatrists a confused picture of motivation. For example, some oral activities are directed toward obtaining food, and others are directed toward achieving sexual gratification. Both activities are pleasure seeking and use the same organs, but they are not, as Freud contended, both necessarily sexual. Labeling all pleasure-seeking behaviors *sexual* obviates specifying precise motivation. People may also use sexual activities to gratify nonsexual needs, such as dependency, aggression, power, and status. Although sexual and nonsexual impulses may jointly motivate behavior, the analysis of behavior depends on understanding the underlying individual motivation and their interactions. Sexuality is something more than physical sex, coital or noncoital, and something less than all behaviors directed toward gaining pleasure.

Psychosexual Factors Sexuality depends on four interrelated psychosexual factors: sexual identity, gender identity, sexual orientation, and sexual behavior. These

factors affect personality, development, and functioning.

Sexual identity is the pattern of a person's biological sexual characteristics: chromosomes, external and internal genitalia, hormonal composition, gonads and secondary sex characteristics. In normal development, these characteristics form a cohesive pattern that leaves persons in no doubt about their sex.

Gender identity is a person's sense of maleness or femaleness. By the age of 2 or 3 years, almost everyone has a firm conviction that “I am male” or “I am female.” Gender identity results from an almost infinite series of clues derived from experiences with family members, peers and teachers and from cultural phenomena. For instance, male infants tend to be handled more vigorously and female infants to be cuddled more. Fathers spend more time with their infant sons than with their daughters, and they also tend to be more aware of their sons' adolescent concerns than of their daughters' anxieties. Boys are more likely than girls to be physically disciplined. A child's sex affects parental tolerance for aggression and reinforcement or extinction of activity and of intellectual, aesthetic, and athletic interests. Physical characteristics derived from a person's biological sex (e.g., physique, body shape and physical dimensions) interrelate with an intricate system of stimuli, including rewards, punishment, and parental gender labels, to establish gender goals.

Sexual orientation describes the object of a person's sexual impulses: heterosexual (opposite sex), homosexual (same sex), or bisexual (both sexes). The overwhelming majority of people have a heterosexual orientation. In the United States, 2.8 percent of men and 1.4 percent of women identify themselves as homosexual. These numbers are compatible with figures from Western European countries as well. However, a higher percentage of persons have had at least one same-sex experience in their lives. Additionally, homosexuals congregate in urban areas, so the incidence of homosexuality in some large cities is as high as 8 or 9 percent.

Sexual behavior includes desire, fantasies, pursuit of partners, autoeroticism, and all the activities engaged in to express and gratify sexual needs. It is an amalgam of psychological and physiological responses to internal and external stimuli.

SEXUAL LEARNING AND SEXUAL BEHAVIOR

The sex drive is innate and varies in intensity in different people, but much sexual behavior is learned. Early sexual experiences, particularly those during puberty and early adolescence, can have an imprinting effect. If they are strongly associated with pleasure and release of tension they are likely to be repeated and the person is conditioned to a particular form of sexual expression. In the normal person sexual learning and experimentation continue throughout the life cycle, the repertoire of sexual behavior expands, and the behaviors are compatible with cultural norms.

Childhood Sexual learning begins in childhood. In a broad sense, that learning occurs through parent-child interaction, including the meeting of the infant's needs, cuddling, and the reinforcement or discouragement of gender-associated activities. Cuddling and physical touching engender emotional security and positive feelings in infants toward their bodies. That physicality lays the groundwork for a healthy body image that is a component of sexual self-esteem. A good body image also derives from mastery of early physical activity and a positive parental approach to tasks such as toilet training.

Genital self-stimulation is a normal activity of babies. It is particularly pronounced between the ages of 15 and 19 months, and it is part of the general interest of children in their bodies. The activity is reinforced by the pleasurable sensations it produces. As youngsters acquire playmates, curiosity about their own and others' genitalia motivates episodes of exhibitionism or genital exploration. Unless the child is unduly shamed, such experiences contribute to continued pleasure from sexual stimulation.

Children also learn by watching the interaction of their parents. They observe demonstrations of physical affection (although usually not the sex act itself), and they are sensitive to the sexual undertone of flirtation, bantering, or seductive interchange.

Adolescence With the approach of puberty, the upsurge of sex hormones, and the development of secondary sex characteristics, sexual curiosity is intensified. Adolescents are physically capable of coitus and orgasm but are usually inhibited by social restraints. The dual, often conflicting pressures of establishing their sexual identities and controlling their sexual impulses produce a strong physiological sexual tension in teenagers that demands release, and masturbation is a normal way to reduce sexual tension. In general, males learn to masturbate to orgasm earlier than females and masturbate more frequently. Consequently, many males integrate their sexuality as an autonomous characteristic earlier than some females. An important emotional difference between the adolescent and the younger child is the presence of coital fantasies during masturbation in the adolescent. These fantasies are an important adjunct to the development of sexual identity; in the comparative safety of the imagination, the adolescent learns to perform the adult sex role. Fantasies that accompany masturbation vary and reflect individual psychodynamics; however, in general, fantasies differ for the sexes. Males respond to visual stimuli of nude or barely dressed women and images of explicit physical acts. Females report responding to romantic stories in which a man demonstrates intense passion for, and commitment to, a woman. Their fantasies focus more on touching, emotions, and the partner's response than on visualizing an explicit sexual act.

In addition to masturbation, adolescents learn sexually through caressing and kissing with partners. In early adolescence, sex play may involve a partner of the same sex for a short period of time for heterosexuals as well as for homosexuals. Adolescence is also when one's body image becomes more definitive and a sense of sexual adequacy and sexual desirability begins to develop. Peer acceptance by the same sex and the opposite sex is of paramount importance. Engaging in sexual talk and jokes, kissing, touching genitalia, experimenting with degrees of nudity, experimenting with different partners or with one partner are part of the process of learning about sexuality. These experiences reinforce the adolescent's sense of being a sexual male or female.

First Coitus The first coitus is a rite of passage for both sexes. The modal age for first coitus is 16 for males and 17 for females. Currently, both considerable peer pressure and individual sex drive impel adolescents or young adults to their first coital experience. In the past, virginity conferred status on the young person. Today, many young people feel embarrassed or inadequate about being virgins. A small backlash is present in some groups in which adolescents sign premarital chastity pledges in a public forum. Also, a survey by the Centers for Disease Control reported that 49 percent of male teenagers had intercourse in 1998, down from 57 percent in 1991. The corresponding number for girls was 48 percent, down from 51 percent.

For the male, anxiety involved in his first coitus relates to performance. Will he be able to get an erection, to penetrate the vagina, to last for some period of time before he ejaculates? He is vulnerable in his masculine pride and may fear being judged inadequate by his female partner and his male peers.

First coitus for a female has been surrounded by cultural ambivalence and concern about the meaning of her loss of virginity and her assumption of the risk of pregnancy and responsibility for the next generation. That risk has been eased by the availability of contraceptive methods. Development of the birth control pill in the 1960s changed societal attitudes toward premarital sex for women and created a more permissive climate. The cohort of females who came of age sexually after the late 1960s report a much higher frequency of premarital intercourse than women before that time. However, only 30 percent of young women use contraception with their first sexual mate and many use it inconsistently. In effect, many young women deny that they are planning to have sex by not being prepared in terms of contraception, reflecting cultural ambivalence about their sexual activity.

Studies have shown that first coitus is most likely to be a positive experience for girls with strong feelings for their partners. In general, women report a greater need to experience sex in the context of an affectionate relationship.

Adulthood By their late 20s, 70 percent of men and 85 percent of women have formed a union, either exclusive cohabitation or marriage. The sexual impulse is catalytic in forming and maintaining adult love relationships. Ideally, a mature sexual relationship encompasses the capacity for intimacy and love for one's partner.

As sexual access ceases to be problematic, more attention is focused on the activity itself. Interference with sexual activity arises from the time and energy required by the pursuit of careers, child rearing, and other family and community obligations. Nonetheless, studies reveal a much higher frequency of sexual interaction among married persons than among single persons. The modal frequency for married persons is three times a month with many couples relating sexually twice a week. The highest frequency of sexual interaction, 4 or more times a week, exists among cohabiting couples.

Kissing, intercourse, and masturbation are frequent sexual activities. Even after a permanent sexual relationship has been established, masturbation remains a healthy practice during the illness or absence of a partner or when intercourse is unsatisfactory. Only when it is a compulsive activity or when it is preferred to partner interaction should it be considered maladaptive. Many adults have some experience with oral sex. Studies show that although it is not a regular activity for most couples, many people are likely to engage in it occasionally throughout their lifetimes. The incidence of the practice of cunnilingus and fellatio are the same; the lifetime prevalence is 75 percent of sexually active people. Anal sex is not part of the repertoire of regular sexual activities for most men and women in the United

States. Although a sizeable minority (25 percent of men and 20 percent of women) report one experience of anal sex, very few repeat the practice. In general, a graph of sexual practices does not form a bell curve. Most curves of sexual behavior are strongly skewed, with many persons indulging in a particular behavior or, conversely, very few people doing so.

Middle Age During middle age the frequency of marital intercourse may decline. The rates of interaction depend more on male interest, and the middle-aged man may devote much of his energy to his career at this point. The decline in sexual interaction often reflects deeroticization of the woman because of her wife-mother role. Familiarity also contributes to a decreased passion, and some men have difficulty connecting marital sexual activity with the erotic scripts they elaborated during their adolescence. Conversely, a couple's experience of each other makes them comfortable and augments their skill at mutual arousal. For women, interest in sexual competence typically increases at this time. A woman's erotic pleasure and sexual commitment are strongly connected to feelings of attachment toward her partner and the security of being in a loving relationship.

With late middle age the biological drive decreases in intensity. It takes the man longer to reach orgasm, he has a longer refractory period, and he requires more stimulation to achieve an erection. The woman must adjust to the hormonal fluctuations of her perimenopausal years, and she too requires more stimulation to become aroused. Although medications are available to address these specific physiological changes, they do not eliminate all need to adjust to the aging process. In terms of family dynamics, children often leave the house for work or further education at this time. That fact plus freedom from concern over unwanted pregnancy may enhance sexual activity for persons who have more time to direct attention toward each other and renew their life as a couple.

However, middle age is the period of rising extramarital activity, although the frequencies of such occurrences are low; 75 percent of men and 85 percent of women remain faithful throughout the lifetime of their marriages. Daniel Offer has explained that the patterning of extramarital sexual activity for both sexes continues to express earlier patterns of psychosexual development:

> For men, it predominantly has the capacity for detachment that in adolescence was directly related to the pursuit of sexual fantasies and the homosocial validation of masculinity. For women, on the other hand, it resembles a quest for circumstances that justify and confirm a romantic self-image rather than a quest for orgasms.

In the context of a loving, communicative, and committed relationship, a decreased frequency of sexual interaction does not herald the onset of extramarital affairs or threaten the stability of the marriage.

SEX IN AMERICA

A 1994 study conducted by the University of Chicago is the latest sex survey and the most authoritative. Based on a representative United States population between the ages of 18 and 59, it found the following:

1. Eighty-five percent of married women and 75 percent of married men are faithful to their spouses.
2. Forty-one percent of married couples have sex twice a week or more, compared with 23 percent of single persons.
3. Cohabiting single persons have the most sex of all, twice a week or more.
4. The median number of sexual partners over a lifetime for men is six; for women, two.
5. A homosexual orientation was reported by 2.8 percent of men and 1.4 percent of women, with 9 percent of men and 5 percent of women reporting that they had at least one homosexual experience after puberty.
6. Vaginal intercourse was considered the most appealing type of sexual experience by 83 percent of men and 78 percent of women.
7. Among married partners, 93 percent are of the same race, 82 percent are of similar educational level, 78 percent are within 5 years of each other's age, and 72 percent are of the same religion.
8. Both men and women who as children had been sexually abused by an adult were more likely as adults to have had more than 10 sex partners, to engage in group sex, to report a homosexual or bisexual identification, and to be unhappy.
9. Less than 8 percent of the participants reported having sex more than four times a week; about two thirds said they had sex a few times a month or less, and about 3 in 10 have sex a few times a year or less.
10. About 1 man in 4 and 1 woman in 10 masturbates at least once a week, and masturbation is less common among those 18 to 24 years of age than among those 24 to 34 years old.
11. Three quarters of the married women said they usually or always had an orgasm during sexual intercourse, compared with 62 percent of the single women. Among men, married or single, 95 percent said they usually or always had an orgasm.
12. More than half of the men said that they thought about sex every day or several times a day, compared with only 19 percent of the women.
13. More than four in five Americans had only one sexual partner or no partner in the past year. Generally, African-Americans reported the most sexual partners and Asian-Americans the fewest.

PHYSIOLOGICAL RESPONSES

Normal men and women experience a sequence of physiological responses to sexual stimulation. In the first detailed description of these responses, Masters and Johnson observed that the physiological process involves increasing vasocongestion and myotonia tumescence and subsequent release of the vascular activity and muscle tone as a result of orgasm detumescence. Table 19.1a-2 and Table 19.1a-3 describe the female and male sexual response cycles, respectively. DSM-IV defines a four-phase response cycle: phase I, desire; phase II, excitement; phase III, orgasm; phase IV, resolution.

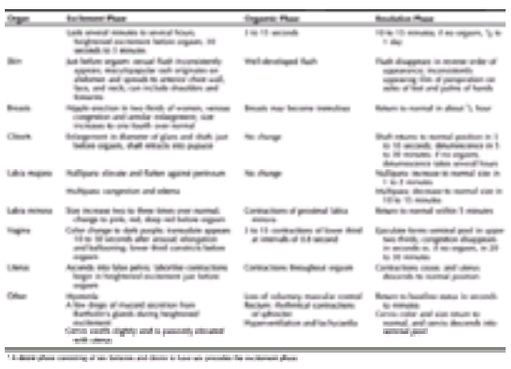

Organ	Excitement Phase	Orgasmic Phase	Resolution Phase
	Lasts several minutes to several hours; heightened excitement before orgasm; 10 seconds to 3 minutes	3 to 15 seconds	10 to 15 minutes; if no orgasm, ½ to 1 day
Skin	Just before orgasm: sexual flush inconsistently appears; maculopapular rash originates on abdomen and spreads to anterior chest wall, face, and neck; can include shoulders and forearms	Well-developed flush	Flush disappears in reverse order of appearance, inconsistently appearing; film of perspiration on soles of feet and palms of hands
Breasts	Nipple erection in two thirds of women; venous congestion and areolar enlargement; size increases by one fourth over normal	Breasts may become meteoritous	Return to normal in about ½ hour
Clitoris	Enlargement in diameter of glans and shaft just before orgasm; shaft retracts into prepuce	No change	Shaft returns to normal position in 1 to 10 seconds; detumescence in 5 to 30 minutes; if no orgasm, detumescence takes several hours
Labia majora	Nullipara: elevate and flatten against perineum Multipara: congestion and edema	No change	Nullipara: decrease to normal size in 1 to 5 minutes Multipara: decrease to normal size in 10 to 15 minutes
Labia minora	Size increase two to three times over normal; change to pink, red, deep red before orgasm	Contractions of proximal labia minora	Return to normal within 5 minutes
Vagina	Color change to dark purple; transudate appears 10 to 30 seconds after arousal; elongation and flattening; lower third constrict before orgasm	3 to 15 contractions of lower third at intervals of 0.8 second	Ejaculate forms seminal pool in upper two thirds; congestion disappears in seconds or, if no orgasm, in 20 to 30 minutes
Uterus	Ascends into false pelvis; ballooning contractions begin in heightened excitement just before orgasm	Contractions throughout orgasm	Contractions cease, and uterus descends to normal position
Other	Myotonia A few drops of mucoid secretion from Bartholin's glands during heightened excitement	Loss of voluntary muscular control Reclam: rhythmical contractions of sphincter Hyperventilation and facial grimacing	Return to baseline status in seconds to minutes Cervix color and size return to normal, and cervix descends into

* A dotted phase consisting of sex hormone and desire to have sex precedes the excitement phase.

Table 19.1a-2 Female Sexual Response Cycle*

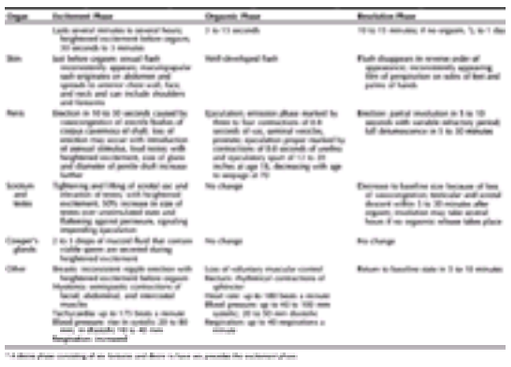

Organ	Excitement Phase	Orgasmic Phase	Resolution Phase
	Lasts several minutes to several hours; heightened excitement before orgasm; 30 seconds to 3 minutes	3 to 15 seconds	10 to 15 minutes; if no orgasm, ½ to 1 day
Skin	Just before orgasm: sexual flush inconsistently appears; maculopapular rash originates on abdomen and spreads to anterior chest wall, face, and neck; can include shoulders and forearms	Well-developed flush	Flush disappears in reverse order of appearance, inconsistently appearing; film of perspiration on soles of feet and palms of hands
Penis	Erection in 10 to 30 seconds caused by vasocongestion of erectile bodies of corpus cavernosa of shaft; loss of erection may occur with introduction of sexual stimulus, loud noise, with heightened excitement, size of glans and diameter of penile shaft increase further	Ejaculation, emission phase marked by three to four contractions of 0.8 seconds of vas, seminal vesicles, prostate; ejaculation proper marked by contractions at 0.8 seconds of urethra and ejaculatory spurt of 12 to 30 inches at age 18, decreasing with age to seepage at 70	Erection: partial resolution in 5 to 10 seconds with variable refractory period; full detumescence in 5 to 30 minutes
Scrotum and testes	Tightening and lifting of scrotal sac and retraction of testes; with heightened excitement, 50% increase in size of testes over unstimulated state and flattening against perineum, signaling impending ejaculation	No change	Decrease to baseline size because of loss of vasocongestion; testicular and scrotal descent within 5 to 30 minutes after orgasm; resolution may take several hours if no orgasmic release takes place
Cowper's glands	2 to 3 drops of mucoid fluid that contain visible sperm are secreted during heightened excitement	No change	No change
Other	Erratic, inconsistent nipple erection with heightened excitement before orgasm Myotonia: semispastic contractions of facial, abdominal, and extremities muscles Tachycardia: up to 175 beats a minute Blood pressure: rise in systolic 20 to 80 mm; in diastolic 10 to 40 mm Respiration: increased	Loss of voluntary muscular control Reclam: rhythmical contractions of sphincter Heart rate: up to 180 beats a minute Blood pressure: up to 40 to 100 mm systolic; 20 to 50 mm diastolic Respiration: up to 40 respirations a minute	Return to baseline state in 5 to 10 minutes

* A dotted phase consisting of sex hormone and desire to have sex precedes the excitement phase.

Table 19.1a-3 Male Sexual Response Cycle.*

Phase I: Desire Phase I is distinct from any identified solely through physiology and reflects the psychiatrist's fundamental concern with motivations, drives, and personality. It is characterized by sexual fantasies and the desire to have sexual activity.

Phase II: Excitement Phase II is brought on by psychological stimulation (fantasy or the presence of a love object), physiological stimulation (stroking or kissing), or a combination of the two. It consists of a subjective sense of pleasure. The excitement phase is characterized by penile tumescence leading to erection in the man and vaginal lubrication in the woman. The nipples of both sexes become erect, although nipple erection is more common in women than in men. The woman's clitoris becomes hard and turgid, and her labia minora become thicker as a result of venous engorgement. Initial excitement may last several minutes to several hours. With continued stimulation, the man's testes increase in size 50 percent and elevate. The woman's vaginal barrel shows a characteristic constriction along the outer third known as the orgasmic platform. The clitoris elevates and retracts behind the symphysis pubis; hence it is not easily accessible. However, stimulation of the area causes traction on the labia minora and the prepuce, and there is intrapreputial movement of the clitoral shaft. Breast size in the woman increases 25 percent. Continued engorgement of the penis and vagina produces specific color changes, particularly in the labia minora, which become bright or deep red. Voluntary contractions of large muscle groups occur, rate of heartbeat and respiration increases, and blood pressure rises. Heightened excitement lasts 30 seconds to several minutes.

Phase III: Orgasm Phase III consists of peaking sexual pleasure, with release of sexual tension and rhythmic contraction of the perineal muscles and pelvic reproductive organs. A subjective sense of ejaculatory inevitability triggers the man's orgasm. Forceful emission of semen follows. The male orgasm is also associated with four to five rhythmic spasms of the prostate, seminal vesicles, vas, and urethra. In the woman orgasm is characterized by 3 to 15 involuntary contractions of the lower third of the vagina and by strong, sustained contractions of the uterus, flowing from the fundus downward to the cervix. Both men and women have involuntary contractions of the internal and external sphincter. Those and the other contractions during orgasm occur at 0.8-second intervals. Other manifestations include voluntary and involuntary movements of the large muscle groups, including facial grimacing and carpopedal spasm. Blood pressure rises 20 to 40 mm (both systolic and diastolic), and the heart rate increases up to 160 beats a minute. Orgasm lasts from 3 to 25 seconds and is associated with a slight clouding of consciousness.

Phase IV: Resolution Resolution consists of the disgorgement of blood from the genitalia (detumescence), which brings the body back to its resting state. If orgasm occurs, resolution is rapid; if it does not occur, resolution may take 2 to 6 hours and be associated with irritability and discomfort. Resolution through orgasm is characterized by a subjective sense of well-being, general relaxation, and muscular relaxation. After orgasm men have a refractory period that may last from several minutes to many hours; in that period they cannot be stimulated to further orgasm. The refractory period does not exist in women, who are capable of multiple and successive orgasms.

Sexual response is a true psychophysiological experience. Arousal is triggered by both psychological and physical stimuli, levels of tension are experienced both physiologically and emotionally, and with orgasm, there is normally a subjective perception of a peak of physical reaction and release. Psychosexual development, psychological attitude toward sexuality, and attitudes toward one's sexual partner are directly involved with, and affect, the physiology of human sexual response.

ABNORMAL SEXUALITY AND SEXUAL DYSFUNCTIONS

Seven major categories of sexual dysfunction are listed in DSM-IV: (1) sexual desire disorders, (2) sexual arousal disorders, (3) orgasm disorders, (4) sexual pain disorders, (5) sexual dysfunction due to a general medical condition, (6) substance-induced sexual dysfunction, and (7) sexual dysfunction not otherwise specified.

Definition In DSM-IV sexual dysfunctions are categorized as Axis I disorders. The syndromes listed are correlated with the sexual physiological response, which is divided into the four phases discussed above. The essential feature of the sexual dysfunctions is inhibition in one or more of the phases, including disturbance in the subjective sense of pleasure or desire or disturbance in the objective performance (Table 19.1a-4 and Table 19.1a-5). Either type of disturbance can occur alone or in combination. Sexual dysfunctions are diagnosed only when they are a major part of the clinical picture. They can be lifelong or acquired, generalized or situational, and due to psychological factors, physiological factors or combined factors. If they are attributable entirely to a general medical condition, substance use, or adverse effects of medication, then sexual dysfunction due to a general medical condition or substance-induced sexual dysfunction is diagnosed.

Phase	Characteristics	Dysfunction
1. Desire	Distinct from any identified solely through physiology and reflects the patient's motivations, drives, and personality; characterized by sexual fantasies and the desire to have sex	Hypoactive sexual desire disorder, sexual aversion disorder, hypoactive sexual desire disorder due to a general medical condition (male or female); substance-induced sexual dysfunction with impaired desire
2. Excitement	Subjective sense of sexual pleasure and accompanying physiological changes; all physiological responses noted in Masters and Johnson's excitement and plateau phases are combined in this phase	Female sexual arousal disorder, male erectile disorder (may also occur in stages 3 and 4); male erectile disorder due to a general medical condition; dyspareunia due to a general medical condition (male or female); substance-induced sexual dysfunction with impaired arousal
3. Orgasm	Peaking of sexual pleasure, with release of sexual tension and rhythmic contraction of the perineal muscles and pelvic reproductive organs	Female orgasmic disorder, male orgasmic disorder, premature ejaculation; other sexual dysfunction due to a general medical condition (male or female); substance-induced sexual dysfunction with impaired orgasm
4. Resolution	A sense of general relaxation, well-being, and muscle relaxation; men are refractory to orgasm for a period of time that increases with age, whereas women can have multiple orgasms without a refractory period	Postcoital dysphoria, postcoital headache

* DSM-IV-TR consolidates the Masters and Johnson excitement and plateau phases into a single excitement phase, which is preceded by the desire (appetitive) phase. The orgasm and resolution phases remain the same as originally described by Masters and Johnson.

Table 19.1a-4 DSM-IV Phases of the Sexual Response Cycle and Associated Sexual Dysfunctions*

Category	Dysfunctions
Sexual pain disorders	Vaginismus (female) Dyspareunia (female and male)
Other	Sexual dysfunctions not otherwise specified. Examples: 1. No erotic sensation despite normal physiological response to sexual stimulation (e.g., orgasmic anhedonia) 2. Female analogue of premature ejaculation 3. Genital pain occurring during masturbation

Table 19.1a-5 Sexual Dysfunction Not Correlated With Phases of the Sexual Response Cycle

According to the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), sexual dysfunction refers to a person's inability to "participate in a sexual relationship as he or she would wish." This dysfunction is expressed in various ways: as a lack of desire or of pleasure or as a physiological inability to begin, maintain, or complete sexual interaction. Because sexual response is psychosomatic, it may be difficult to determine "the relative importance of psychological and/or organic factors."

Sexual dysfunction such as lack of desire can occur in both men and women, but women more often complain of the "subjective quality" of the experience than of the "failure of a specific response." ICD-10 advises looking "beyond the presenting complaint to find the most appropriate diagnostic category." Table 19.1a-6 presents the ICD-10 diagnostic criteria.

Table 19.1a-6 ICD-10 Diagnostic Criteria for Sexual Dysfunction, Not Caused by Organic Disorder or Disease

With the possible exception of premature ejaculation, sexual dysfunctions are rarely found separate from other psychiatric syndromes. Sexual disorders may lead to or result from relational problems, and patients invariably develop an increasing fear of failure and self-consciousness about their sexual performance. Sexual dysfunctions are frequently associated with other mental disorders, such as depressive disorders, anxiety disorders, personality disorders, and schizophrenia. In many instances, a sexual dysfunction may be diagnosed in conjunction with another psychiatric disorder; in other cases, however, it is only one of many signs or symptoms of the psychiatric disorder.

A sexual disorder can be symptomatic of biological problems, intrapsychic conflicts, interpersonal difficulties, or a combination of these factors. The sexual function can be affected by stress of any kind, by emotional disorders, and by a lack of sexual knowledge.

Taking a Sexual History As with all psychiatric interviews, sexual history taking is not only an information-gathering time, it also permits the development of a positive doctor-patient relationship. The development of confidence and rapport requires an accepting atmosphere and a nonjudgmental attitude on the part of the therapist toward the patients' sexual values, ideas, and practices.

The sexual history is more structured than the rest of the psychiatric interview, although patients are encouraged to take their own lead in areas of great personal significance. In general, the therapist structures the interview so that both recent and early sexual histories are covered. The therapist must ascertain the specific current sexual complaint, the person's sexual practices and pattern of interaction with partners, the person's sexual goal and fantasies, the masturbatory history, the presence or extent of extramarital relationships, and the degree of commitment to the marriage or the partner. Patients describe their view of the problem and when it began. If married, the courtship, honeymoon, and reproductive history are examined in detail. Premarital expectations, mutual physical attraction, periods of separation, the type of contraception used, and the effect of children on the couple's sexual life are covered. The satisfying aspects of the marriage must also be discussed. The patient is particularly asked to evaluate the partner's contribution to the present distress.

Early sexual development and education are also thoroughly discussed. The interviewer asks for the patient's view of the parents' marriage as seen in retrospect and as perceived in childhood. Relationships to peers, siblings, and important familial figures other than parents are also explored. Particular attention is paid to ways in which affection was expressed in the family and the degree of physical contact between family members. The sexual climate in which the patient grew up is seen through reported parental attitudes, memories of sexual games played as a child, the way in which the patient learned sexual facts, the specifics of religious training, reactions to masturbation and nocturnal emissions or the menarche, dating patterns, an adolescent rebellious phase, and any significant premarital involvements. Ethnic background and the socioeconomic level of the patient's primary family are also taken into account. As the interview progresses, the patient's self-image emerges. The interviewer must be sensitive to any event that was exceptional in the patient's sexual life in either a destructive or a highly pleasant manner and should take particular note of the people who contributed to the patient's sexual education, identity, and mores.

The interviewer must also ask specific questions to elicit information that may be outside the patient's view of the socially acceptable, such as premarital and extramarital affairs, group sex, homosexual involvements, and abortions. The sexual orientation of the person being interviewed should be ascertained, and questions related to same-sex interactions explored. All interviews should review high-risk sexual behavior regardless of sexual orientation, since transmission of the human immunodeficiency virus (HIV) is increasing rapidly in all groups. Additionally, the issue of sexual abuse must be explored, particularly since a history of abuse predisposes to the development of sexual dysfunction.

Similar disorders exist among both homosexual and heterosexual partners, with variations imposed by anatomical differences. For example, although penile-vaginal dysfunction cannot be described among homosexuals, penile-anal dysfunction may occur. In spite of sexual orientation, each phase of the sex cycle applies equally to same-sex and heterosexual partners, and the methods and principles for treatment are essentially similar.

Rating Scales In addition to the interview, several questionnaires are available to assess sexual function. They were developed primarily to evaluate illness or medication-related effects on sexual functioning. The most commonly used scales are the Arizona Sexual Experience Scale, the Brief Sexual Function Questionnaire, the Changes in Sexual Functioning Questionnaire, the Derogatis Sexual Function Inventory, and the Rush Sexual Inventory. The scales vary in length, reliability, validity, and method of administration. Some are rated by the patients themselves, others by the therapist. The formats include structured and semistructured approaches. The scales differ in number of symptoms assessed, in the dysfunctions they target, and in the time frame they cover.

Sexual Desire Disorders DSM-IV divides sexual desire disorders into two classes: hypoactive sexual desire disorder, characterized by a deficiency or lack of sexual fantasies and desire for sexual activity, and sexual aversion disorder, characterized by an aversion to, and avoidance of, genital sexual contact with a sexual partner. The former condition is more common than the latter.

Hypoactive Sexual Desire Disorder Hypoactive sexual desire disorder (Table 19.1a-7) is experienced by both men and women; however, they may not be hampered by any dysfunction once they are involved in the sex act. Conversely, hypoactive desire may be used to mask another sexual dysfunction. Lack of desire may be expressed by decreased frequency of coitus, perception of the partner as unattractive, or overt complaints of lack of desire. Upon questioning, the patient is found to have few or no sexual thoughts or fantasies, a lack of awareness of sexual cues, and little interest in initiating sexual experiences.

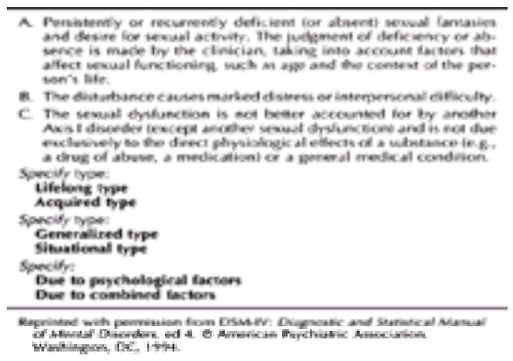

A. Persistently or recurrently deficient (or absent) sexual fantasies and desire for sexual activity. The judgment of deficiency or absence is made by the clinician, taking into account factors that affect sexual functioning, such as age and the context of the person's life.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The sexual dysfunction is not better accounted for by another Axis I disorder (except another sexual dysfunction) and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Reprinted with permission from DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-7 Diagnostic Criteria for Hypoactive Sexual Desire Disorder

Sometimes biochemical correlates are associated with hypoactive desire. A recent study found markedly low serum testosterone concentrations in men complaining of this dysfunction when they were compared with normal controls in a sleep-laboratory situation. Also, a central dopamine blockage is known to decrease desire (Table 19.1a-7). The need for sexual contact and satisfaction varies among different persons, as well as in the same person over time. In a group of 100 couples with stable

marriages, 8 percent reported having intercourse less than once a month. In another group of couples, one third reported lack of sexual relations for periods averaging 8 weeks. In a survey of a general medical practice in England, 25 percent of that sample reported no sexual activity most of the time. It has been estimated that 20 percent of the total population have hypoactive sexual desire disorder. The complaint is more common among women.

Patients with desire problems often have good ego strengths and use inhibition of desire defensively to protect against unconscious fears about sex. Lack of desire can also result from chronic stress, anxiety, or depression. Abstinence from sex for a prolonged period sometimes suppresses the sexual impulse. Desire problems may also be an expression of hostility or signal a deteriorating relationship.

The presence of desire depends on several factors: biological drive, adequate self-esteem, previous good experiences with sex, the availability of an appropriate partner, and a good relationship in nonsexual areas with one's partner. Damage to any of those factors may result in diminished desire.

Hypoactive sexual desire disorders often become manifest during puberty and may remain a lifelong condition. A general medical workup is always indicated to rule out a medical cause, which would be diagnosed as male or female sexual desire disorder due to a general medical condition according to DSM-IV.

Mr. and Mrs. L. were a professional couple in their mid-40s who presented with a chief complaint of lack of desire on the wife's part. Although Mrs. L. stated she never found her husband “a terrific turn-on”; she had felt sufficiently attracted to him for the couple to have sustained a satisfactory sex life up to the year before they entered therapy. She agreed to seek help because of his increasing complaints about their lack of sexual interaction.

They were both Ph.D.s and taught at different colleges. To juggle her obligations at home and work, Mrs. L. had been forced to reduce her teaching hours and resign from academic committees, and she had been unable to complete several scholarly articles she had outlined for publication. Mr. L. took time for conferences, networking, and writing, He felt he could not put his career on hold. Also, he objected to his wife's wish to have him as available to her as he had been during their time as graduate students.

Mrs. L. was diagnosed as having hypoactive sexual desire disorder, acquired type. She had experienced desire in the past and was a sexually responsive woman. Her current lack of desire was associated with anger at her husband whom she perceived as no longer willing to support her career and needing “to have everything go his way.”

Sexual Aversion Disorder Sexual aversion disorder (Table 19.1a-8) is defined in DSM-IV as a “persistent or recurrent and extreme aversion to, and avoidance of, all or almost all, genital sexual contact with a sexual partner.” Some researchers consider the line between hypoactive desire disorder and sexual aversion disorder blurred, and in some cases both diagnoses are appropriate. Low frequency of sexual interaction is a symptom common to both disorders. The clinician should think of the words “repugnance” and “phobia” in relation to the patient with sexual aversion disorder. Freud conceptualized sexual aversion as the result of inhibition during the phallic psychosexual phase and unresolved oedipal conflicts. Some men, fixated at the phallic stage of development, fear the vagina and believe that they will be castrated if they approach it (a concept Freud called *vagina dentata*), because they believe unconsciously that the vagina has teeth. Hence they avoid contact with the female genitalia entirely.

A. Persistent or recurrent extreme aversion to, and avoidance of, all (or almost all) genital sexual contact with a sexual partner.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The sexual dysfunction is not better accounted for by another Axis I disorder (except another sexual dysfunction).

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Table 19.1a-8 Diagnostic Criteria for Sexual Aversion Disorder

The disorder may result from a traumatic sexual assault, such as rape or childhood abuse, from repeated painful experiences with coitus, or from early developmental conflicts that have left the patient with unconscious connections between the sexual impulse and overwhelming feelings of shame and guilt. The disorder may also be a reaction to a perceived psychological assault by the partner and to relationship difficulties.

Mr. and Mrs. J. were 38 and 36 years old, respectively, when they presented for treatment stating they had not had coitus for 3 years. Three years previously Mrs. J. revealed that she had been having an extramarital affair of 2 months duration, which she then ended. At that time, she told her husband that she had been unhappy with their lovemaking.

Upon hearing of her affair, the husband was angry and refused to approach his wife sexually. When he finally did so, she would allow caressing of her body but not her genitalia although she was willing to stimulate him sexually. At this point the husband also had an affair but stopped when he realized it would not solve his problems with his wife, and she then agreed to enter sex therapy. When they presented for sex therapy their sexual interaction consisted of mutual kissing and her manual stimulation of his penis until he reached orgasm.

In the individual interviews that were part of the evaluation of the case, Mrs. J. reiterated that she found her husband's lovemaking unsatisfactory; but that she did not have a problem allowing other men to caress her genitalia and that she could have coitus with other men. Mr. J. had been traumatized by his wife's rejection and sexual betrayal, and he in turn was averse to touching her genitalia or attempting intromission.

This was a case in which both partners suffered from sexual aversion disorder. The husband feared touching his wife in an explicitly sexual way; the wife was averse to his touching her genitalia, and both spouses avoided coitus. For both partners the dysfunction was acquired and situational.

Sexual Arousal Disorders DSM-IV divides the sexual arousal disorders into (1) female sexual arousal disorder, characterized by the persistent or recurrent partial or complete failure to attain or maintain the lubrication-swelling response of sexual excitement until the completion of the sexual act, and (2) male erectile disorder, characterized by the recurrent and persistent partial or complete failure to attain or maintain an erection until the completion of the sex act. The diagnosis takes into account the focus, the intensity, and the duration of the sexual activity in which the patient engages. If sexual stimulation is inadequate in focus, intensity, or duration, the diagnosis should not be made.

Female Sexual Arousal Disorder Women who have excitement-phase dysfunction often have orgastic problems as well. In one series of relatively happily married couples, 33 percent of the women described difficulty in maintaining sexual excitement. Numerous psychological factors are associated with female sexual inhibition. Those conflicts may be expressed through inhibition of excitement or orgasm and are discussed under orgasmic phase dysfunctions. In some women arousal disorders are associated with dyspareunia or lack of desire.

Less research has been done on physiological components of dysfunction in women than in men, and there have been conflicting results. Masters and Johnson found normally responsive women to particularly desire sex premenstrually. A recent study found that dysfunctional women tended to be more responsive immediately following their periods. A third group of dysfunctional women felt the greatest sexual excitement at the time of ovulation. Some evidence indicates that dysfunctional women are less aware of such physiological responses of their bodies as vasocongestion, during arousal.

There are some organic causes for female sexual arousal disorder. Alterations in testosterone, estrogen, prolactin, and thyroxine concentrations have been implicated. Medications with antihistaminic or anticholinergic properties lessen vaginal lubrication and interfere with arousal. Also, postmenopausal women require

longer stimulation for lubrication to occur, and there is generally less vaginal transudate after menopause. An artificial lubricant is frequently useful in that situation.

Table 19.1a-9 presents the diagnostic criteria for female sexual arousal disorder.

A. Persistent or recurrent inability to attain, or to maintain until completion of the sexual activity, an adequate lubrication-swelling response of sexual excitement.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The sexual dysfunction is not better accounted for by another Axis I disorder (except another sexual dysfunction) and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Table 19.1a-9 Diagnostic Criteria for Female Sexual Arousal Disorder

Male Erectile Disorder Male erectile disorder (Table 19.1a-10) is also called *erectile dysfunction* and *impotence*. A man with lifelong male erectile disorder has never obtained an erection sufficient for vaginal insertion. In acquired male erectile disorder, the man successfully achieved vaginal penetration at some time in his sexual life but later cannot do so. In situational male erectile disorder, the man can have coitus in certain circumstances but not in others; for example, a man may function effectively with a prostitute but not with his wife. Kinsey estimated that a few men (2 to 4 percent) are dysfunctional at age 35 but 77 percent are at age 80. Ten percent of the men in the University of Chicago study reported an experience with erectile dysfunction in the past year, and between 15 and 20 percent experienced anxiety about performing. More recently it was estimated that the incidence of erectile dysfunction in young men is about 8 percent. However, that sexual dysfunction may first appear later in life. Masters and Johnson reported a fear of impotence in all men over 40, which the researchers believed reflects the masculine fear of loss of virility with advancing age. (As it happens, however, erectile dysfunction is not a regularly occurring phenomenon in aged men; good health and an available sexual partner are more closely related to continuing potency than is age per se.) From 35 to 50 percent of all men treated for sexual disorders have erectile dysfunction as the chief complaint.

A. Persistent or recurrent inability to attain, or to maintain until completion of the sexual activity, an adequate erection.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The erectile dysfunction is not better accounted for by another Axis I disorder (other than a sexual dysfunction) and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Table 19.1a-10 Diagnostic Criteria for Male Erectile Disorder

Many recent studies have focused on the relative incidences of psychological and organic erectile dysfunction. Physiologically, erectile dysfunction may be due to a variety of medical causes (Table 19.1a-11). In the United States it is estimated that two million men cannot gain erections because they suffer from diabetes mellitus; an additional 300,000 are dysfunctional because of other endocrine diseases; 1.5 million are dysfunctional as a result of vascular disease; 180,000 because of multiple sclerosis; 400,000 because of traumas and fractures leading to pelvic fractures or spinal cord injuries; and another 650,000 as a result of radical surgery, including prostatectomies, colostomies, and cystectomies. In addition, the clinician should be aware of the possible pharmacological effects of medication on sexual functioning. The increased incidence of organic causes for erectile dysfunction in the past 15 years may partly reflect the increased use of psychotropic and antihypertensive medications. Statistics indicate that 50 to 80 percent of men with erectile dysfunction have a medical basis for their problem. The following case is an example of male erectile dysfunction.

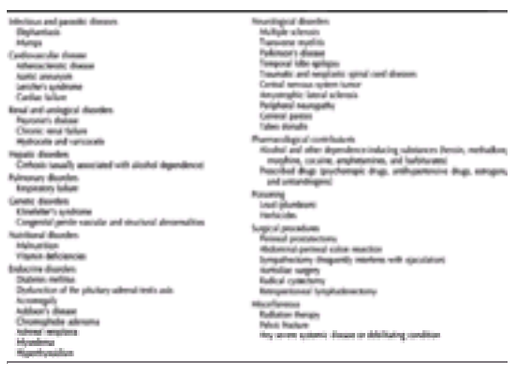

Table 19.1a-11 Diseases and Other Medical Conditions Implicated in Erectile Dysfunction

Mr. R., a 45-year-old doctor, presented with a chief complaint of difficulty gaining and maintaining erections. He had been suffering from erectile dysfunction for the past 4 years, with a few exceptions, ever since he divorced his wife. At the age of 29 he had married a woman to whom he was very attracted. He never experienced any dysfunction with her, and they had frequent intercourse. However, he described her as very passive during their sex play and stated that she rarely caressed him or directly stimulated his genitals. After the birth of their child his wife began complaining about the amount of time he spent working. At the same time she began rebuffing his sexual overtures. Their sex life dwindled and then ceased. The patient stated the reason he divorced his wife was his dissatisfaction with her social skills—he described her as withdrawn and shy—and his unhappiness with her demands and coldness.

Mr. R.'s sexual experience prior to his marriage consisted of masturbation and four or five visits to prostitutes. As a youngster and adolescent he felt unattractive to girls, short and homely, and he was convinced his penis was smaller than average. Although he was the spouse who terminated the marriage, his wife's sexual rejection of him revived his sense of sexual inadequacy. All his earlier sexual anxiety resurfaced, and he became unable to function.

He was diagnosed as having male erectile disorder, acquired type.

Orgasmic Disorders Female orgasmic disorder (also known as *inhibited female orgasm*, or anorgasmia) is defined as the recurrent and persistent inhibition of the female orgasm, as becomes manifest by the absence or delay of orgasm after a normal sexual excitement phase that the clinician judges to be adequate in focus, intensity, and duration. Women who can achieve orgasm with noncoital clitoral stimulation but cannot experience it during coitus in the absence of manual clitoral

stimulation are not necessarily categorized as anorgasmic.

Physiological research on the female sexual response has demonstrated that orgasms caused by clitoral stimulation and those caused by vaginal stimulation are physiologically identical. Freud's theory that women must give up clitoral sensitivity for vaginal sensitivity to achieve sexual maturity is now considered misleading, although some women say that they gain a special sense of satisfaction from an orgasm precipitated by coitus. Some workers attribute that to the psychological feeling of closeness engendered by the act of coitus, but others maintain that the coital orgasm is a physiologically different experience. Many women achieve orgasm during coitus by a combination of manual clitoral stimulation and penile vaginal stimulation.

Lifelong female orgasmic disorder exists when the woman has never experienced orgasm by any kind of stimulation. Acquired orgasmic dysfunction exists if the woman has previously experienced at least one orgasm regardless of the circumstances or means of stimulation, whether by masturbation or during sleep while dreaming. Kinsey found that the proportion of married women over 35 years of age who had never achieved orgasm by any means was only 5 percent. The incidence of orgasm increased with age. According to Kinsey, the first orgasm occurs in late adolescence in about 50 percent of women. The rest usually experience orgasm by some means as they get older. Lifelong female orgasmic disorder is more common among unmarried women than among married women; 39 percent of the unmarried women over age 35 in Kinsey's study had never experienced orgasm. Increased orgasmic potential in women over 35 has been explained on the basis of less psychological inhibition, greater sexual experience, or both. Also, orgasmic consistency has been correlated with marital happiness, although cause and effect have not been determined. In the University of Chicago study, three quarters of the female respondents usually or always experienced orgasm during sex, compared with two thirds of the single women. One woman in 10 complained of difficulty in achieving orgasm.

Mr. and Mrs. Z. were a childless couple, both in their late 20s. She was a college instructor, and he was a free-lance writer. The couple had been married for 4 years and came to therapy with a mutual complaint of steadily lessening sexual frequency. During the year before they were seen, they had had intercourse six times. In the initial interview, Mrs. Z. stated that she had never been able to have an orgasm with her husband and had never experienced an orgasm during intercourse. Her frustration made her increasingly reluctant to have coitus and rejecting of her husband sexually. Mrs. Z. could masturbate to orgasm while indulging in masochistic fantasies, but she did so very infrequently.

She was the oldest of four children, raised by rigid and intellectual, but undemonstrative, parents. She stated that she had been a very docile child and adolescent but openly rebelled in marrying her husband. Her parents disapproved of him, since he came from a different religious background. Mrs. Z.'s father particularly objected, and relations between Mr. Z. and his father-in-law were strained. Mrs. Z. stated that she often felt caught between them.

Mr. Z. was the middle child and the only boy in a family of three sisters. There was a 20-year age difference between his parents, and the father was not actively involved with the children. He felt he had always been his mother's favorite.

Mr. and Mrs. Z. met in college, and although neither was sexually experienced, they were strongly attracted to one another and indulged in enjoyable sexual play, short of intercourse. Mrs. Z. perceived her husband as a very assertive man and was impressed by the way he stood up to her family. When he completed school, a year ahead of Mrs. Z., he broke off the relationship with her. During that year Mrs. Z. became involved with another man and enjoyed sexual play with him. She did not have intercourse with him but sometimes came to orgasm through manual manipulation. At the time of her graduation, she terminated this relationship and managed to regain Mr. Z.'s attention. The couple married 6 months after her graduation. Intercourse was a disappointment to Mrs. Z. from the beginning. She felt unfulfilled, and her husband felt rejected and inadequate.

In therapy, the couple's premarital expectations were discussed. In effect, each had expected more assertiveness from the other, and in reality, each had some problem with passivity. Several sessions were also spent dealing with relations with in-laws. Mrs. Z. was encouraged to give priority to her relationship with her husband, and Mr. Z. was encouraged to face and restrain his competitiveness with her father. They responded to the behavioral exercises that were prescribed. At the time of their discharge from therapy, the couple was having intercourse 1 to 2 times a week. Mrs. Z. could not reach orgasm during intercourse but could frequently achieve climax after coitus through manual stimulation.

Acquired female orgasmic disorder is a common complaint in the clinical population. One clinical treatment facility described nonorgasmic women as about four times more common in its practice than patients with all other sexual disorders. In another study 46 percent of the women complained of difficulty in reaching orgasm, and 15 percent described an inability to have orgasm. The overall prevalence of inhibited orgasm in women is estimated to be 30 percent.

Numerous psychological factors are associated with female sexual inhibition. They include fears of impregnation, rejection by the sexual partner, or damage to the vagina; hostility toward men; and feelings of guilt regarding sexual impulses. Some women equate orgasm with loss of control or with aggressive, destructive, or violent behavior. Fear of those impulses may be expressed through inhibition of excitement or orgasm. The expression of orgasmic inhibition varies. Some women feel unentitled to gratify themselves and cannot masturbate to climax. Others enjoy self-stimulation but cannot reach orgasm with a partner present. Cultural expectations and societal restrictions on women are also relevant. Nonorgasmic women may be otherwise symptom free or may experience frustration in a variety of ways, including such pelvic complaints as lower abdominal pain, itching, and vaginal discharge, as well as increased tension, irritability, and fatigue. The diagnostic criteria for female orgasmic disorder are presented in Table 19.1a-12.

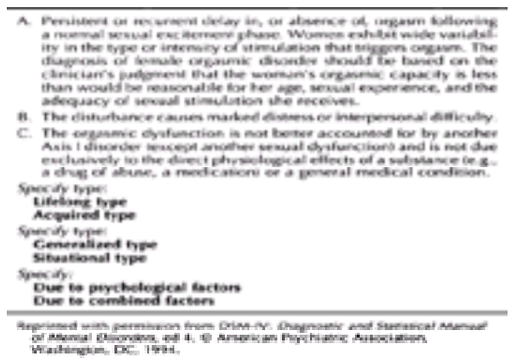

A. Persistent or recurrent delay in, or absence of, orgasm following a normal sexual excitement phase. Women exhibit wide variability in the type or intensity of stimulation that triggers orgasm. The diagnosis of female orgasmic disorder should be based on the clinician's judgment that the woman's orgasmic capacity is less than would be reasonable for her age, sexual experience, and the adequacy of sexual stimulation she receives.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The orgasmic dysfunction is not better accounted for by another Axis I disorder (except another sexual dysfunction) and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, edition 4, © American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-12 Diagnostic Criteria for Female Orgasmic Disorder

Male Orgasmic Disorder In male orgasmic disorder (previously inhibited male orgasm and called *retarded ejaculation*), the male achieves climax during coitus with great difficulty, if at all. A man suffers from lifelong orgasmic disorder if he has never ejaculated during coitus. The disorder is diagnosed as acquired if it develops after previous normal functioning.

Some workers suggest that orgasm and ejaculation should be differentiated. Certainly, inhibited orgasm must be differentiated from retrograde ejaculation, in which ejaculation occurs but the seminal fluid passes backward into the bladder. This condition always has an organic cause. Retrograde ejaculation can develop after genitourinary surgery and is also associated with medications that have anticholinergic adverse effects, such as the phenothiazines, particularly thioridazine (Mellaril).

The incidence of male orgasmic disorder is much lower than those of premature ejaculation and erectile dysfunction. Masters and Johnson reported only 3.8 percent in one group of 447 sex-dysfunction cases. This problem is more common among men with obsessive-compulsive disorders than among others. Male orgasmic disorder may have physiological causes and can occur after surgery of the genitourinary tract, such as prostatectomy. Male orgasmic dysfunction may also be associated with Parkinson's disease and other neurological disorders involving the lumbar or sacral sections of the spinal cord. The antihypertensive drugs guanethidine monosulfate (Esmelin) and methyldopa (Aldomet) have been implicated in retarded ejaculation. Phenothiazines have also been associated with the disorder, as have almost all the antidepressants. Transient retarded ejaculation may occur with excessive alcohol intake or with hyperglycemia, whether caused by drugs or by pituitary adenoma. Strictly organic cases and problems that are symptomatic of other Axis I psychiatric syndromes are not to be included in the diagnosis.

Primary male orgasmic disorder indicates more-severe psychopathology. The man often comes from a rigid, puritanical background; he perceives sex as sinful and the genitals as dirty, and he may have conscious or unconscious incest wishes and guilt. Usually difficulties with closeness exist that extend beyond the area of

sexual relations.

In an ongoing relationship, secondary ejaculatory inhibition frequently reflects interpersonal difficulties. The disorder may be the man's way of coping with real or fantasized changes in the relationship. Those changes may include plans for a pregnancy about which the man is ambivalent, the loss of sexual attraction to the partner, or demands by the partner for greater commitment as expressed by sexual performance. In some men the inability to ejaculate reflects unexpressed hostility toward women.

In a version of the dysfunction, some men experience partial inhibition of ejaculation. Those men experience a slow dribbling of ejaculation (not related to age) rather than an ejaculatory spurt. They usually do not experience the pleasurable sensations of orgasm.

The following case is an example of ejaculatory inhibition.

> Mr. and Mrs. G. presented with the chief complaint of the man being unable to achieve climax and ejaculate. The problem had begun some 10 weeks prior to the couple's entering treatment. The man denied having masturbated during that time, but after being instructed to do so in the privacy of his home as part of the therapeutic evaluation, he reported having no difficulty reaching orgasm and ejaculating with masturbation.
>
> The couple (male aged 46 and female aged 39) had been married for 2 years. Mr. G. had a child from a previous marriage. This was Mrs. G.'s first marriage and she had no children. Although Mr. G. did not want more children, he felt it would be unfair to deprive his wife of them, and he had agreed to have a new family prior to the marriage. Ten months after they were married, Mrs. G. suffered a miscarriage. The couple decided to wait before trying to have her conceive again but had resumed trying 3 months before entering therapy.
>
> It was clear following the initial interviews that Mr. G. felt he had made “a good faith effort” with his wife's first conception. His genuine distress for her at the time of the miscarriage had increased his ambivalence about having her conceive again. His current problem of retarded ejaculation was a dysfunctional attempt to resolve the conflict of fulfilling his promise to his wife with his desire to have no more children. He was diagnosed with male orgasmic disorder, acquired.

The diagnostic criteria for male orgasmic disorder are presented in Table 19.1a-13.

A. Persistent or recurrent delay in, or absence of, orgasm following a normal sexual excitement phase during sexual activity that the clinician, taking into account the person's age, judges to be adequate in focus, intensity, and duration.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The orgasmic dysfunction is not better accounted for by another Axis I disorder (except another sexual dysfunction) and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Table 19.1a-13 DSM-IV Diagnostic Criteria for Male Orgasmic Disorder

Premature Ejaculation In *premature ejaculation* the man recurrently achieves orgasm and ejaculates before he wishes to do so. There is no definite time frame within which to define the dysfunction. The diagnosis is made when the man regularly ejaculates before or immediately after entering the vagina or following minimal sexual stimulation. The clinician should consider actors that affect duration of the excitement phase, such as age, novelty of the sexual partner, and the frequency and duration of coitus. Masters and Johnson conceptualized the disorder in terms of the couple and consider a man a premature ejaculator if he cannot control ejaculation long enough during intravaginal containment to satisfy his partner in at least half of their episodes of coitus. That definition assumes that the female partner is capable of an orgasmic response. As with other dysfunctions, the disturbance is diagnosed only if it is not caused exclusively by medical factors or is not symptomatic of any other Axis I syndrome.

Premature ejaculation is more common today among college-educated men than among men with less education and is thought to be related to their concern for partner satisfaction. It is estimated that 30 percent of the male population have the dysfunction, and about 40 percent of men treated for sexual disorders have premature ejaculation as the chief complaint.

Difficulty in ejaculatory control may be associated with anxiety regarding the sex act. Both anxiety and ejaculation are mediated by the sympathetic nervous system. Other psychological factors that have been noted include sexual guilt, a history of parent-child conflict, interpersonal hypersensitivity, and perfectionism or unrealistic expectations about sexual performance.

Current research also suggests that a subgroup of premature ejaculators (particularly those with a lifelong history of premature ejaculation) may be biologically predisposed to this dysfunction. Some researchers believe that certain males are constitutionally more vulnerable to sympathetic stimulation, hence they ejaculate rapidly. Other workers have found a shorter bulbocavernosus reflex nerve latency time in men with lifelong premature ejaculation than in men who had acquired the dysfunction.

Premature ejaculation also may result from negative cultural conditioning. The man who has most of his early sexual contacts with prostitutes who demand that the sex act proceed quickly or in situations in which discovery would be embarrassing, such as in an apartment shared with roommates or in the parental home, may become conditioned to achieving orgasm rapidly. In ongoing relationships the partner has been found to have great influence on the premature ejaculator. A stressful marriage exacerbates the disorder. Table 19.1a-14 gives the diagnostic criteria for premature ejaculation.

A. Persistent or recurrent ejaculation with minimal sexual stimulation before, on, or shortly after penetration and before the person wishes it. The clinician must take into account factors that affect duration of the excitement phase, such as age, novelty of the sexual partner or situation, and recent frequency of sexual activity.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The premature ejaculation is not due exclusively to the direct effects of a substance (e.g., withdrawal from opioids).

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Table 19.1a-14 Diagnostic Criteria for Premature Ejaculation

Sexual Pain Disorders

Dyspareunia *Dyspareunia* refers to recurrent and persistent pain during intercourse in either a man or a woman. In women the dysfunction is related to and often coincides with vaginismus. Repeated episodes of vaginismus may lead to dyspareunia and vice versa, but in either case somatic causes must be ruled out. Dyspareunia should not be diagnosed as such when a medical basis for the pain is found or when (in a woman) it is associated with vaginismus or with lack of lubrication.

The true incidence of dyspareunia is unknown, but it has been estimated that 30 percent of surgical procedures on the female genital area result in temporary dyspareunia. Additionally, among women seen in sex therapy clinics, the complaint is more common in women with a history of rape or childhood sexual abuse.

Dynamic factors are usually considered causative, although situational factors probably account for more secondary dysfunction. Painful coitus may result from tense vaginal muscles. The pain is real and makes intercourse unbearable or unpleasant. Anticipation of further pain may cause the woman to avoid coitus altogether. If the partner proceeds with intercourse regardless of the woman's state of readiness, the condition is aggravated.

Dyspareunia can also occur in men, but it is uncommon and usually associated with a medical condition such as Peyronie's disease, prostatitis, or gonorrheal or herpetic infections. Vasocongestion during sexual activity without orgasmic release also may lead to discomfort. Rarely, some men experience pain upon ejaculation (postejaculatory pain disorder). This pain is caused by an involuntary spasm of the perineal muscles that may be due to physiological conflicts about the sex act or be an adverse effect of some antidepressant medications. Table 19.1a-15 lists the diagnostic criteria for dyspareunia. In DSM-IV, dyspareunia due to a general medical condition is used when the medical condition is the sole or major causal factor.

A. Recurrent or persistent genital pain associated with sexual intercourse in either a male or a female.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The disturbance is not caused exclusively by vaginismus or lack of lubrication, is not better accounted for by another Axis I disorder (except another sexual dysfunction), and is not due exclusively to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Reprinted with permission from DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. ©, American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-15 DSM-IV Diagnostic Criteria for Dyspareunia

Vaginismus *Vaginismus* is an involuntary and persistent constriction of the outer one third of the vagina that prevents penile insertion and intercourse. The response may be demonstrated during a gynecological examination when involuntary vaginal constriction prevents introduction of the speculum into the vagina, although some women only have vaginismus during coitus. The diagnosis is not made if the dysfunction is caused exclusively by medical or surgical factors or if it is symptomatic of another Axis I psychiatric syndrome (Table 19.1a-16). Vaginismus is less prevalent than anorgasmia. It most often afflicts highly educated women and those in the higher socioeconomic groups. A milder form of the dysfunction in which some vaginal tightness makes penile entry difficult is experienced by more women on an intermittent or chronic basis.

A. Recurrent or persistent involuntary spasm of the musculature of the outer third of the vagina that interferes with sexual intercourse.
B. The disturbance causes marked distress or interpersonal difficulty.
C. The disturbance is not better accounted for by another Axis I disorder (e.g., somatization disorder) and is not due exclusively to the direct physiological effects of a general medical condition.

Specify type:
Lifelong type
Acquired type

Specify type:
Generalized type
Situational type

Specify:
Due to psychological factors
Due to combined factors

Reprinted with permission from DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. Copyright American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-16 DSM-IV Diagnostic Criteria for Vaginismus

A woman suffering from vaginismus may consciously wish to have coitus but unconsciously prevent penile entrance into her body. A sexual trauma, such as rape, may result in vaginismus. Women who have experienced pain with nonsexual bodily traumas, through accidents or because of illness or surgery, may become sensitized to the idea of penetration. Women with psychosexual conflicts may perceive the penis as a dangerous weapon. Pain or the anticipation of pain at the first coital experience causes vaginismus in some women. A strict religious upbringing that associates sex with sin is frequently noted in such cases. Others have problems in the dyadic relationship; a woman who feels emotionally abused by her partner may protest in this nonverbal fashion.

Miss M. was a 31-year-old accountant who presented for therapy stating that she was afraid to have intercourse. She was increasingly distressed by this problem as she was becoming more interested in forming a serious, loving relationship with a man. Miss M. was a virgin who had never had a serious boyfriend and who had little dating experience. However, she had experienced some petting sessions with men in which she allowed them to stimulate her genitally over her underwear, and she reported becoming excited. She also had sexual fantasies and masturbated to climax.

As a girl, the patient had been squeamish about menstruation, and she was unable to use tampons. Prior to therapy the she had never gone for a gynecological examination. A gynecological consultation was included in her workup. The gynecologist reported no abnormal findings, with the exception of extreme difficulty performing the examination even using the smallest available speculum. The patient reported anxiety and pain associated with the examination and felt light headed for some time after its completion.

She was given the diagnosis of vaginismus, lifelong type.

Sexual Dysfunction Due to a General Medical Condition The category sexual dysfunction due to a general medical condition covers sexual dysfunction that results in marked distress and interpersonal difficulty when there is evidence from the history, the physical examination, or the laboratory findings of a general medical condition judged to be causally related to the sexual dysfunction (Table 19.1a-17).

A. Clinically significant sexual dysfunction that results in marked distress or interpersonal difficulty predominates in the clinical picture.
B. There is evidence from the history, physical examination, or laboratory findings that the sexual dysfunction is fully explained by the direct physiological effects of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (e.g., major depressive disorder).

Select code and term based on the predominant sexual dysfunction:

Female hypoactive sexual desire disorder due to a general medical condition: if deficient or absent sexual desire is the predominant feature.

Male hypoactive sexual desire disorder due to a general medical condition: if deficient or absent sexual desire is the predominant feature.

Male erectile disorder due to a general medical condition: if male erectile dysfunction is the predominant feature.

Female dyspareunia due to a general medical condition: if pain associated with intercourse is the predominant feature.

Male dyspareunia due to a general medical condition: if pain associated with intercourse is the predominant feature.

Other female sexual dysfunction due to a general medical condition: if some other feature is predominant (e.g., orgasmic disorder) or no feature predominates.

Other male sexual dysfunction due to a general medical condition: if some other feature is predominant (e.g., orgasmic disorder) or no feature predominates.

Coding note: Include the name of the general medical condition on Axis I (e.g., male erectile disorder due to diabetes mellitus; also code the general medical condition on Axis III.

Table 19.1a-17 DSM-IV Diagnostic Criteria for Sexual Dysfunction Due to a General Medical Condition

Male Erectile Disorder Due to a General Medical Condition Many studies have focused on the relative incidences of psychological and organic male erectile disorder. Statistics indicate that 50 to 80 percent of men with erectile disorder have an organic basis for the disorder. The medical causes of male erectile dysfunction are listed in Table 19.1a-11. Adverse effects of medication may impair male sexual functioning in a variety of ways. Castration (removal of the testes) does not always lead to sexual dysfunction, depending on the person. Erection may still occur after castration, via a reflex arc fired when the inner thigh is stimulated that passes through the sacral cord erectile center.

PHYSIOLOGICAL TESTS A number of procedures, benign and invasive, are used to help differentiate psychogenic erectile dysfunction. The procedures include monitoring nocturnal penile tumescence (erection that occur during sleep) normally associated with rapid eye movement; monitoring tumescence with a strain gauge; measuring blood pressure in the penis with a penile plethysmograph or an ultrasound (Doppler) flow meter, both of which assess blood flow in the internal pudendal artery; and measuring pudendal nerve latency time. Neurological impairment of penile function may be indicated by decreased vibratory perception in the penis. Other diagnostic tests that delineate organic bases for erectile disorder include glucose tolerance test, plasma hormone assays, liver and thyroid function tests, prolactin and follicle-stimulating hormone (FSH) determinations, and cystometric examinations. Invasive diagnostic studies include penile arteriography, infusion cavernosography, and radioactive xenon penography. Invasive procedures require expert interpretation and are used only for patients who are candidates for vascular reconstructive procedures.

MEDICAL VERSUS PSYCHOGENIC CAUSES A good history is crucial to determining the cause of the male erectile disorder. If a man reports having spontaneous erections at times when he does not plan to have intercourse, having morning erections or only sporadic erectile dysfunction, or having good erections with masturbation or with partners other than his usual one, then organic causes for his disorder can be considered negligible, and costly diagnostic procedures can be avoided. When a medical basis for erectile dysfunction is found, psychological factors often contribute to the dysfunction, and psychiatric treatment may be helpful. Some diabetics, for instance, may experience psychogenic erectile dysfunction.

In general, the psychological conflicts that cause erectile dysfunction are related to an inability to express the sexual impulse because of fear, anxiety, anger, or moral prohibition. Lifelong dysfunction is a more serious, but less common, condition than acquired erectile disorder, and is less amenable to treatment.

Many developmental factors have been cited as contributing to erectile disorder. Any experience that hinders the ability to be intimate, that leads to a feeling of inadequacy or distrust, or that develops a sense of being unloving or unlovable may result in this problem. Erectile dysfunction in an ongoing relationship may reflect difficulties between the partners, particularly if one person cannot communicate his or her needs or angry feelings in a direct and constructive way. Successive episodes of impotence are reinforcing, with the man becoming increasingly anxious about his next sexual encounter. Regardless of the original cause of the dysfunction, his anticipatory anxiety about achieving and maintaining an erection interferes with his pleasure in sexual contact and his ability to respond to stimulation, thus perpetuating the problem.

Dyspareunia Due to a General Medical Condition An estimated 30 percent of all surgical procedures on the female genital area result in temporary dyspareunia. In addition, 30 to 40 percent of women with the complaint who are seen in sex therapy clinics have pelvic pathology. Organic abnormalities leading to dyspareunia and vaginismus include irritated or infected hymenal remnants, episiotomy scars, Bartholin's gland infection, various forms of vaginitis and cervicitis, and endometriosis. Postcoital pain reported by women with myomata and endometriosis is attributed to the uterine contractions during orgasm. Postmenopausal women may have dyspareunia because of thinning of the vaginal mucosa and reduced lubrication.

Dyspareunia can also occur in men, but it is uncommon and is usually associated with an organic condition such as Peyronie's disease, which consists of sclerotic plaques on the penis that cause penile curvature.

Hypoactive Sexual Desire Disorder Due to a General Medical Condition Desire commonly decreases after major illness or surgery, particularly when the body image is affected after such procedures as mastectomy, ileostomy, hysterectomy, and prostatectomy. Illnesses that deplete a person's energy, chronic conditions that require physical and psychological adaptation, and serious illnesses that may cause the person to become depressed can all result in a marked lessening of sexual desire in both men and women. In some cases, biochemical correlates are associated with hypoactive sexual desire disorder (Table 19.1a-18). A recent study found markedly lower serum testosterone concentrations in men complaining of low desire than in normal controls in a sleep-laboratory situation. Drugs that depress the CNS or decrease testosterone production can decrease desire.

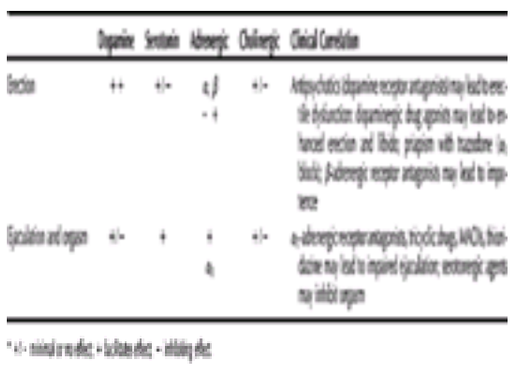

	Dopamine	Serotonin	Adrenergic	Cholinergic	Clinical Correlation
Erection	++	+/-	α, β - +	+/-	Antipsychotics (dopamine receptor antagonists) may lead to erectile dysfunction; dopaminergic drug agonists may lead to enhanced erection and libido; yohimbine (α2 blocks), β-adrenergic receptor antagonists may lead to impotence
Ejaculation and orgasm	+/-	+	+ α1	+/-	α-adrenergic receptor antagonists, tricyclic drugs, MAOIs, thioridazine may lead to impaired ejaculation; serotonergic agents

* +/- = minimal or no effect; + = facilitates effect; - = inhibits effect.

Table 19.1a-18 Neurotransmitter Effects on Sex Function*

Other Male Sexual Dysfunction Due to a General Medical Condition The category other male sexual dysfunction due to a general medical condition is used when some other dysfunctional feature predominates (e.g., orgasmic disorder) or no feature predominates. Male orgasmic dysfunction may have physiological causes and can occur after surgery on the genitourinary tract, such as prostatectomy. It may also be associated with Parkinson's disease and other neurological disorders involving the lumbar or sacral sections of the spinal cord. The antihypertensive drug guanethidine monosulfate, methyldopa, the phenothiazines, the tricyclic drugs, and serotonin reuptake inhibitors, among others, have been implicated in retarded ejaculation. Male orgasmic disorder must also be differentiated from retrograde ejaculation, in which ejaculation occurs but the seminal fluid passes backward into the bladder. Retrograde ejaculation always has an organic cause. It can develop after genitourinary surgery and is also associated with medications with anticholinergic adverse effects, such as the phenothiazines.

Other Female Sexual Dysfunction Due to a General Medical Condition The category other female sexual dysfunction due to a general medical disorder is used when some other feature (e.g., orgasmic disorder) predominates or no feature predominates.

Some medical conditions—specifically such endocrine diseases as hypothyroidism, diabetes mellitus, and primary hyperprolactinemia—can affect a woman's ability to have orgasms. Also, a number of drugs affect some women's ability to have orgasms. Antihypertensive medications, CNS stimulants, tricyclic drugs, serotonin reuptake inhibitors, and frequently monoamine oxidase inhibitors (MAOIs) have interfered with female orgasmic capacity (Table 19.1a-19). However, one study of women taking MAOIs found that after 16 to 18 weeks of pharmacotherapy, that adverse effect of the medication disappeared, and the women could reexperience orgasms, although they continued taking an undiminished dosage of the drug.

Tricyclic antidepressants
Imipramine (Tofranil)
Clomipramine (Anafranil)
Nortriptyline (Aventyl)
Monoamine oxidase inhibitors
Tranylcypromine (Parnate)
Phenelzine (Nardil)
Isocarboxazid (Marplan)
Dopamine receptor antagonists
Thioridazine (Mellaril)
Trifluoperazine (Stelazine)
Selective serotonergic receptor inhibitors
Fluoxetine (Prozac)
Paroxetine (Paxil)
Sertraline (Zoloft)
Fluvoxamine (Luvox)
Citalopram (Celexa)

* The interrelation between female sexual dysfunction and pharmacological agents has been less extensively evaluated than have male reactions. Oral contraceptives are reported to decrease libido in some women, and some drugs with anticholinergic side effects may impair arousal as well as orgasm. Benzodiazepines have been reported to decrease libido, but in some patients the diminution of anxiety caused by these drugs enhances sexual function. Both increase and decrease in libido have been reported with psychoactive agents. It is difficult to separate these effects from the underlying condition or from improvement of the condition. Sexual dysfunction associated with the use of a drug disappears when the drug is discontinued.

Table 19.1a-19 Some Psychiatric Drugs Implemented in Inhibited Female Orgasm*

Substance-Induced Sexual Dysfunction The diagnosis substance-induced sexual dysfunction is used when evidence from the history, physical examination, or laboratory findings indicates substance intoxication or withdrawal when dysfunction follows the use of prescribed medication. Distressing sexual dysfunction occurs within a month of significant substance intoxication or withdrawal (Table 19.1a-20). Specified substances include alcohol; amphetamines or related substances; cocaine; opioids; sedatives, hypnotics, or anxiolytics; and other or unknown substances.

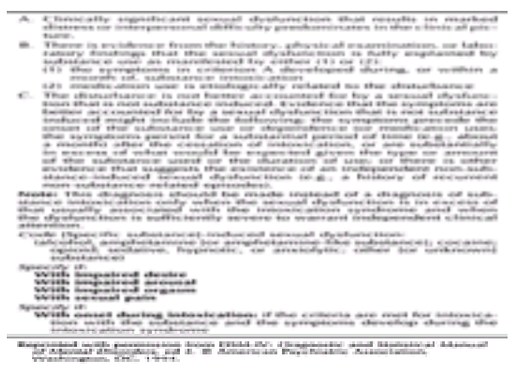

A. Clinically significant sexual dysfunction that results in marked distress or interpersonal difficulty predominates in the clinical picture.

B. There is evidence from the history, physical examination, or laboratory findings that the sexual dysfunction is fully explained by substance use as manifested by either (1) or (2):
(1) the symptoms in criterion A developed during, or within a month of, substance intoxication
(2) medication use is etiologically related to the disturbance

C. The disturbance is not better accounted for by a sexual dysfunction that is not substance induced. Evidence that the symptoms are better accounted for by a sexual dysfunction that is not substance induced might include the following: the symptoms precede the onset of the substance use or dependence or medication use; the symptoms persist for a substantial period of time (e.g., about a month) after the cessation of intoxication or are substantially in excess of what would be expected given the type or amount of the substance used or the duration of use; or there is other evidence that suggests the existence of an independent non-substance-induced sexual dysfunction (e.g., a history of recurrent non-substance-related episodes).

Note: This diagnosis should be made instead of a diagnosis of substance intoxication only when the sexual dysfunction is in excess of that usually associated with the intoxication syndrome and when the dysfunction is sufficiently severe to warrant independent clinical attention.

Code (Specify substance-induced sexual dysfunction):
(alcohol; amphetamine (or amphetamine-like substance); cocaine; opioid; sedative, hypnotic, or anxiolytic; other (or unknown) substance)

Specify (if):
With impaired desire
With impaired arousal
With impaired orgasm
With sexual pain

Specify (if):
With onset during intoxication: if the criteria are met for intoxication with the substance and the symptoms develop during the intoxication syndrome

Reprinted with permission from DSM-IV, Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-20 DSM-IV Diagnostic Criteria for Substance-Induced Sexual Dysfunction

Abused recreational substances affect sexual function in various ways. In small doses many substances enhance sexual performance by decreasing inhibition or anxiety or by temporarily elevating mood. However, continuous use impairs erectile, orgasmic, and ejaculatory capacities. The abuse of sedatives, anxiolytics, and particularly opiates and opioids nearly always depresses desire. Alcohol may foster the initiation of sexual activity by removing inhibitions, but it impairs performance. Cocaine and amphetamines produce similar effects. Although no direct evidence indicates that sexual drive is enhanced, the user initially has a feeling of increased energy and may become sexually active. Ultimately, dysfunction occurs. Men usually go through two stages: prolonged erection without ejaculation and then a gradual loss of erectile capacity.

Recovering substance-dependant patients may need therapy to regain sexual function. In part, this is one piece of psychological readjustment to a nondependent state. Many substance abusers have always had difficulty with intimate interactions. Others have missed the experience that would have enabled them to learn social and sexual skills because they spent their crucial developmental years under the influence of some substance.

Pharmacological Agents Implicated in Sex Dysfunction Many pharmacological agents, particularly those used in psychiatry, have been associated with an effect on sexuality. In men those effects include decreased sex drive, erectile failure (impotence), decreased volume of ejaculate, and delayed or retrograde ejaculation. In women decreased sex drive, decreased vaginal lubrication, inhibited or delayed orgasm, and decreased or absent vaginal contractions may occur. Drugs may also enhance the sexual response and increase the sex drive, but that effect is less common than are adverse effects.

Antipsychotic Drugs Most antipsychotic drugs are dopamine receptor antagonists that also block adrenergic and cholinergic receptors, which accounts for their adverse sex effects. Chlorpromazine (Thorazine), thioridazine, trifluoperazine (Stelazine), and haloperidol (Haldol) are potent anticholinergic agents that impair erection and ejaculation in men and inhibit vaginal lubrication and orgasm in women. Chlorpromazine and thioridazine also block adrenergic receptors. Thioridazine has a particular adverse effect of causing retrograde ejaculation in which the seminal fluid backs up into the bladder rather than being propelled through the penile urethra. Patients still have a pleasurable sensation of orgasm, but it is dry. When urinating after orgasm, the urine may be milky white since it contains the ejaculate. The condition is startling but harmless and may occur in up to 50 percent of patients taking the drug. Paradoxically, rare cases of priapism have been reported with antipsychotics.

Antidepressant Drugs The antidepressant drugs may be subdivided into heterocyclic drugs, such as imipramine (Tofranil); the SSRIs, such as fluoxetine; and the MAOIs, such as phenelzine (Nardil).

Sexual dysfunction has been reported with almost all antidepressants, including changes in libido, impaired erectile capacity, and such ejaculatory/orgasmic dysfunction as associated discomfort and delayed or inhibited ejaculation or orgasm. The sexual adverse effects of these drugs are of particular concern since over 65 million prescriptions a year for antidepressants have been dispensed since 1997.

The most commonly prescribed group of antidepressants are the selective serotonin reuptake inhibitors (SSRIs). Adverse sexual effects may occur with this group of drugs because of increased serotonin concentration. A lowering of the sex drive and difficulty reaching orgasm occur in both sexes. Of the SSRIs, the most frequent sexual adverse effects are seen with paroxetine (Paxil), next with fluoxetine (Prozac) and the least with sertraline (Zoloft). Similar symptoms have been associated with fluvoxamine (Luvox) and citaloprom (Celexa). Reversal of negative effects has been achieved with cyproheptadine (Periactin), an antihistamine with antiserotonergic effects; amantadine (Symmetrel), a dopamine agonist; yohimbine, (Yocon) a central a_2 adrenergic receptor antagonist used in treating male erectile disorder, and methylphenidate (Ritalin), or dextroamphetomine which are dopaminergic and have adrenergic effects. There have been anecdotal reports of sildenafil (Viagra), a nitric oxide enhancer used to treat erectile dysfunction, overcoming orgasmic problems associated with the SSRIs. However, this remedy has not been systematically studied.

The tricyclic and tetracyclic antidepressants have anticholinergic effects that interfere with erection and delay ejaculation. Since the anticholinergic effects vary among the cyclic antidepressants, those with the fewest side effects (e.g., desipramine [Norpramin]) produce the fewest sexual side effects. The effects of the heterocyclic drugs in women have not been studied sufficiently; however, few women seem to complain of any effects. Selegiline (Eldepryl) is a selective monoamine oxidase type B (MAO_B) inhibitor reported to increase sex drive, possibly by dopaminergic activity and increased production of norepinephrine.

Some men reported a pleasurable increased sensitivity of the glans that does not interfere with erection, although it delays ejaculation. In some cases, however, the tricyclic drug causes a painful ejaculation, perhaps as the result of interference with seminal propulsion caused by interference with urethral, prostatic, vas, and epididymal smooth muscle contractions. Clomipramine (Anafranil) has been reported to increase sex drive in some persons. Buspirone (Buspar) helps some patients overcome adverse sexual effect of SSRI's, possibly because it is an 5-HT_A agonist or because it suppresses SSRI induced elevation of prolactin.

The MAOIs affect biogenic amines broadly. Accordingly, they produce impaired erection, delayed or retrograde ejaculation, vaginal dryness, and inhibited orgasm. Tranylcypromine (Parnate) has a paradoxical sexually stimulating effect in some persons, possibly as a result of its amphetamine-like properties.

GENERAL EFFECTS Since depression is associated with a decreased libido, varying levels of sexual dysfunction and anhedonia are part of the disease process. That phenomenon makes assessment of dysfunction as a result of sexual side effects in patients taking the drugs difficult. Some patients report improved sexual function as their depression improves with antidepressant medication. Sometimes sexual side effects disappear with time, perhaps because a biogenic amine homeostatic mechanism comes into play. In many cases, antidepressants without associated side effects of sexual dysfunction are substituted, such as bupropion (Wellbutrin), nefazadone (Serzone), or mirtazapine (Remeron). There have been rare, individual reports of orgasmic dysfunction with the latter two drugs.

Lithium Lithium regulates mood and in the manic state may reduce hypersexuality, possibly by a dopamine antagonist activity. Others have reported impaired erection.

Psychostimulants Psychostimulants are sometimes used in the treatment of depression and include amphetamine, methylphenidate, and pemoline (Cylert), which raise plasma concentrations of norepinephrine and dopamine. Libido is increased; however, with prolonged use, men may experience a loss of desire and erections. One study found that ephedrine sulfate facilitated arousal in functional women.

a-Adrenergic and b-Adrenergic Receptor Antagonists Adrenergic receptor antagonists are used to treat hypertension, angina, and certain cardiac arrhythmias. They diminish tonic sympathetic nerve outflow from vasomotor centers in the brain. As a result, they can cause impotence, decrease the volume of ejaculate, and produce retrograde ejaculation. Changes in libido have been reported in both sexes. Suggestions have been made to use the side effects of drugs therapeutically. Thus a drug that delays or interferes with ejaculation (such as fluoxetine) might be used to treat premature ejaculation.

Anticholinergics The anticholinergics block cholinergic receptors and include such drugs as amantadine and benztropine (Cogentin). They can produce dryness of the mucous membranes (including those of the vagina) and erectile dysfunction.

Antihistamines Drugs such as diphenhydramine (Benadryl) have anticholinergic activity and are mildly hypnotic. They may inhibit sexual function as a result. Cyproheptadine, although an antihistamine, also has potent activity as a serotonin antagonist. It is used to block the serotonergic sexual adverse effects produced by SSRIs, such as delayed orgasm and erectile dysfunction.

Antianxiety Agents The major class of antianxiety drugs is the benzodiazepines (e.g., diazepam [Valium]). They act on the g-aminobutyric (GABA) receptors, which are believed to be involved in cognition, memory, and motor control. Because they decrease plasma epinephrine concentration, they diminish anxiety; thus they improve sexual function in persons inhibited by anxiety.

Alcohol Alcohol suppresses CNS activity generally and hence can produce erectile disorders in men. Alcohol has a direct gonadal effect that decreases testosterone concentrations in men; paradoxically, it can produce a slight increase in testosterone concentrations in women. This may account for increased libido in women after drinking small amounts of alcohol. Long-term use of alcohol reduces the ability of the liver to metabolize estrogenic compounds; in men this produces signs of feminization (e.g., gynecomastia as a result of testicular atrophy).

Opioids Opioids such as heroin have such adverse sexual effects as erectile failure and decreased libido. Altered consciousness may enhance the sexual experience in occasional users.

Hallucinogens The hallucinogens include lysergic acid diethylamide (LSD), phencyclidine (PCP), psilocybin (from some mushrooms), and mescaline (from peyote cactus). In addition to inducing hallucinations, the drugs cause loss of contact with reality and an expanding and heightening of consciousness. Some users report that the sexual experience is similarly enhanced; others experience anxiety, delirium, or psychosis, which clearly interferes with sex function.

Cannabis The altered state of consciousness produced by cannabis may enhance sexual pleasure for some persons. Its prolonged use depresses testosterone concentrations.

Barbiturates and Similarly Acting Drugs Barbiturates are sedative-hypnotics that may enhance sexual responsiveness in persons who are sexually unresponsive because of anxiety. They have no direct effect on the sex organs; however, they do alter consciousness, which some persons find pleasurable. They are subject to abuse and may be fatal when combined with alcohol or other CNS depressants. Methaqualone (Quaalude) acquired a reputation as a sexual enhancer that had no biological basis in fact. It is no longer marketed in the United States.

Sexual Dysfunction and Sexual Disorder Not Otherwise Specified DSM-IV uses two categories-sexual dysfunction not otherwise specified and sexual disorder not otherwise specified. The diagnostic criteria are listed in Table 19.1a-21 and Table 19.1a-22, respectively. The distinction between the two categories is unclear, however, and there is overlap between them. Many sexual disorders are not classifiable as sexual dysfunctions or as paraphilias. These unclassified disorders are rare, poorly documented, not easily classified, or not specifically described in DSM-IV. ICD-10 has a similar residual category for problems related to sexual development or preference (Table 19.1a-23).

This category includes sexual dysfunctions that do not meet criteria for any specific sexual dysfunction: Examples include.

1. No (or substantially diminished) subjective erotic feelings despite otherwise normal arousal and orgasm
2. Situations in which the clinician has concluded that a sexual dysfunction is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced

Table 19.1a-21 DSM-IV Diagnostic Criteria for Sexual Dysfunction Not Otherwise Specified

This category is included for coding a sexual disturbance that does not meet the criteria for any specific sexual disorder and is neither a sexual dysfunction nor a paraphilia. Examples include

1. Marked feelings of inadequacy concerning sexual performance or other traits related to self-imposed standards of masculinity or femininity
2. Distress about a pattern of repeated sexual relationships involving a succession of lovers who are experienced by the individual only as things to be used
3. Persistent and marked distress about sexual orientation

Table 19.1a-22 DSM-IV Diagnostic Criteria for Sexual Disorder Not Otherwise Specified

This section is intended to cover those types of problems that derive from variations of sexual development or orientation, when the sexual preference per se is not necessarily problematic or abnormal.

Sexual maturation disorder
The patient suffers from uncertainty about his or her gender identity or sexual orientation, which causes anxiety or depression.

Ego-dystonic sexual orientation
The gender identity or sexual preference is not in doubt, but the individual wishes it were different.

Sexual relationship disorder
The abnormality of gender identity or sexual preference is responsible for difficulties in forming or maintaining a relationship with a sexual partner.

Other psychosexual development disorders

Psychosexual development disorder, unspecified

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.1a-23 ICD-10 Diagnostic Criteria for Psychological and Behavioral Disorders Associated With Sexual Development and Orientation

Examples include persons who experience the physiological components of sexual excitement and orgasm but report no erotic sensation or even anesthesia and the male experience of orgasm with a flaccid penis. The orgasmic female who desires, but has not experienced, multiple orgasms can be classified under this heading as well. Also, disorders of excessive rather than inhibited function, such as compulsive masturbation, might be diagnosed under atypical dysfunction. Other sexual practices exist that are not listed in DSM-IV, for example, behaviors that attempt to enhance sexual arousal by oxygen deprivation (hypoxyphilia) or other deviant methods. Atypical dysfunction also might be used to cover complaints engendered by couple, rather than individual, dysfunction; for example, a couple in which one partner prefers morning sex and the other functions more readily at night or a couple with unequal frequencies of desire.

Compulsive Sexual Behavior The concept of compulsive sexual behavior, or sex addiction, developed over the past two decades to describe persons who compulsively seek out sexual experiences and whose behavior becomes impaired if they cannot gratify their sexual impulses. The concept of sex addiction derived from the model of addiction to such drugs as heroin, or addiction to behavioral patterns, such as gambling. Addiction implies psychological dependence, physical dependence, and a withdrawal symptom if the substance (e.g., the drug) is unavailable or the behavior (e.g., gambling) is frustrated.

DSM-IV does not use the terms *sex addiction* or *compulsive sexual behavior*, nor is this disorder universally recognized or accepted. Nevertheless, the person whose entire life revolves around sex-seeking behavior and activities, who spends an excessive amount of time in such behavior, and who often tries to stop such behavior but cannot do so is well known to clinicians. Such persons show repeated and increasingly frequent attempts to have a sexual experience, and deprivation evokes symptoms of distress. In the author's view sex addiction is a useful concept heuristically because it can alert the clinician to seek an underlying cause for the manifest behavior.

DIAGNOSIS Sex addicts cannot control their sexual impulses, which can involve the entire spectrum of sexual fantasy or behavior. The need for sexual activity increases, and the person's behavior is motivated largely by the persistent desire to experience the sex act. The history usually reveals a long-standing pattern of such behavior, which the person repeatedly tried to stop without success. Although feelings of guilt and remorse may exist after the act, they do not prevent its recurrence. The patient may report that the need to act out is most severe during stressful periods or when angry, depressed, anxious, or otherwise dysphoric. Most acts culminate in a sexual orgasm, although a sense of excitement (a high) usually accompanies the sex-seeking behavior even in the absence of orgasm. Eventually, the sexual activity interferes with the person's social, vocational, or marital life, which begins to deteriorate.

The signs of sexual addiction are listed in Table 19.1a-24.

1. Out-of-control behavior
2. Severe adverse consequences (medical, legal, interpersonal) due to sexual behavior
3. Persistent pursuit of self-destructive or high-risk sexual behavior
4. Repeated attempts to limit or stop sexual behavior
5. Sexual obsession and fantasy as a primary coping mechanism
6. Need for increasing amounts of sexual activity
7. Severe mood changes related to sexual activity (e.g., depression, euphoria)
8. Inordinate amount of time spent in obtaining sex, being sexual, or recovering from sexual experience
9. Interference of sexual behavior in social, occupational, or recreational activities

Data from Carnes P: *Don't Call It Love*. Bantam Books, New York, 1991.

Table 19.1a-24 Signs of Sexual Addiction or Compulsive Sexual Behavior

TYPES OF BEHAVIORAL PATTERN Compulsive sexual behavior is seen in males more than females in a ratio of 4 to 1. The paraphilias constitute the behavioral patterns frequently found in the sex addict. As defined in DSM-IV, the essential features of a paraphilia are recurrent intense sexual urges or behaviors, including exhibitionism, fetishism, frotteurism, sadomasochism, cross-dressing, voyeurism, and pedophilia. Paraphilias are associated with clinically significant distress and almost invariably interfere with interpersonal relationships; they often lead to legal complications. In addition to the paraphilias, however, sex addiction can also include behavior considered normal, such as coitus and masturbation, except that it is promiscuous and uncontrolled. Paradoxically, the sex addict may suffer from a sexual disorder such as erectile dysfunction or orgasmic inhibition.

In the nineteenth century Krafft-Ebing reported several cases of abnormally increased sexual desire. One involved a 36-year-old married teacher, the father of seven children, who masturbated repeatedly while sitting at his desk in front of his pupils, after which he was “penitent and filled with shame.” He indulged in coitus three or four times a day in addition to his repeated masturbatory acts. In another case, a young woman masturbated almost incessantly and was unable to control her impulses. She had frequent coitus with many men, but neither coitus or masturbation sufficed, and she eventually was placed in an institution. Krafft-Ebing referred to the condition as *sexual hyperaesthesia*, which he believed could occur in otherwise normal persons.

In many cases sex addiction is the final common pathway of a variety of other disorders. In addition to the paraphilias that are frequently present, there may be an associated major mental disorder, such as anxiety, depressive disorder, bipolar disorder, or schizophrenia. Antisocial personality disorder and borderline personality disorder are common.

COMORBIDITY Comorbidity (dual diagnosis) refers to the presence of an addiction that coexists with another psychiatric disorder. For example, about 50 percent of patients with substance-use disorder also have an associated psychiatric disorder. Dual diagnosis implies that the psychiatric illness and the addiction are separate disorders; one does not cause the other. The diagnosis of comorbidity is often difficult to make because addictive behavior (of all types) can produce extreme anxiety and severe disturbances in mood and affect, especially while the addictive behavior is treated. If symptoms of a psychiatric disorder remain after a period of abstinence, the comorbid condition is more easily recognized and diagnosed. Finally, sex addiction and substance-use disorders are highly correlated (up to 80 percent in some studies), which complicates both diagnosis and treatment.

MANAGEMENT Self-help groups based on the 12-step concept used in Alcoholics Anonymous (AA) have been employed successfully with many sex addicts. They

include such groups as Sexaholics Anonymous (SA), Sex and Love Addicts Anonymous (SLAA), and Sex Addicts Anonymous (SAA). The groups differ in that some are for men or women or for married persons or couples. All advocate some abstinence from either the addictive behavior or sex in general. If a substance-use disorder is also present, the patient often requires referral to AA or Narcotics Anonymous (NA) as well. Patients may enter an impatient treatment unit when they lack sufficient motivation to control their behavior on an outpatient basis or are a danger to self or others. Additionally, medical or psychiatric symptoms may require careful supervision and treatment best carried out in a hospital.

Mr. P., a 44-year-old man, was married for the second time. He presented with feelings of depression. His history revealed a pattern of chronic dissatisfaction in his intimate relationships, a persistent problem of orgasmic inhibition in his sexual interactions, and a need to masturbate four or five times a day. He had no problem ejaculating with masturbation. He would get up at 4 o'clock in the morning to watch pornography without his wife knowing and masturbate, He would interrupt work several times a day to go to the men's room and masturbate, historically with pornographic magazines, and after computers were installed in his office, following pornographic interchange with unknown women on the computer. He constantly fantasized about women he would see at work or on the street and occasionally followed them, although he did not accost them. The patient believed that one of the women that appealed to him sexually would turn out to be the love of his life. However, he had felt that way initially about both wives and about many women with whom he had had affairs. All his interactions ended with his feeling dissatisfied and disappointed.

Mr. P. was one of four children in a middle-class family. His father became disabled when the patient was 6 years old, and he remembers his father either lying in bed or sitting alone in the basement. He died when the patient was 10. The mother was described as caring but stressed by the responsibility of raising the family herself and not emotionally available. Mr. P. masturbated “incessantly” throughout his adolescence and married the first woman with whom he had coitus, during his senior year in college. She was 2 years old than he. He interrupted his habit of masturbation only to allow his wife to conceive when she pressured him to have a baby. However, after the couple had a child he had a series of affairs, one involving the woman who became his second wife.

Mr. P., an intelligent and successful man, continually put his relationships and career in jeopardy with his sexual compulsivity. He would not stop masturbation even though he knew he might be caught at work; he had experienced sex as an anodyne for his depression and emotional needs during adolescence and could not give it up. Therapy for Mr. P. entailed a multimodal approach including insight-orientated psychotherapy, group work with other sex addicts, couple therapy, and the use of a specific serotonin reuptake inhibitor, both to treat his depression and for its side effect of decreasing libido.

PSYCHOTHERAPY Insight-oriented psychotherapy may help patients understand the dynamics of their behavioral patterns. Supportive psychotherapy can help repair the interpersonal, social, or occupational damage that occurs. Cognitive-behavioral therapy helps the patient recognize dysphoric states that precipitate sexual acting out. Marital therapy or couples therapy can help the couple establish or reestablish communication, which is usually severely impaired by the time a treatment program is begun. Finally, psychotherapy may help in the treatment of any associated psychiatric disorder.

PHARMACOTHERAPY Most specialists in general addiction avoid the use of pharmacological agents, especially in the early stages of treatment. Substance-dependent persons tend to abuse these agents, especially those with a high abuse potential, such as benzodiazepines. Pharmacotherapy is of use in the treatment of such associated psychiatric disorders as major depressive disorder and schizophrenia.

Certain medications may help the sex addict, however, because of their specific effects on reducing the sex drive. Fluoxetine and other SSRIs reduce libido in some persons, a side effect that is used therapeutically. Compulsive masturbation is an example of a behavioral pattern that may benefit from such medication. Medroxyprogesterone acetate (Depo-Provera) diminishes libido in men and thus helps them control sexually addictive behavior. However, this medication is rarely prescribed by therapists.

The use of antiandrogens to control hypersexuality in women has not been sufficiently tested, but since androgenic compounds contribute to the sex drive in women, antiandrogens could be of benefit. Antiandrogenic agents are not available in the United States but are used in Europe with varying success.

Postcoital Dysphoria Postcoital dysphoria is not listed in DSM-IV. It occurs during the resolution phase of sexual activity, when persons normally experience a sense of general well-being and muscular and psychological relaxation. Some persons, however, experience postcoital dysphoria; after an otherwise satisfactory sexual experience, they become depressed, tense, anxious, and irritable and show psychomotor agitation. They often want to get away from the partner and may become verbally or even physically abusive. The incidence of the disorder is unknown, but it is more common in men than in women. Its several causes relate to the person's attitude toward sex in general and the partner in particular. It may occur in adulterous sex and with prostitutes or when individuals cannot experience sex without consequential strong feelings of guilt. The fear of AIDS causes some persons to experience postcoital dysphoria. Treatment requires insight-oriented psychotherapy to help patients understand the unconscious antecedents to their behavior and attitudes.

Unconsummated Marriage A couple involved in an unconsummated marriage have never had coitus and are typically uninformed and inhibited about sexuality. Their feelings of guilt, shame, or inadequacy are increased by their problems, and they experience conflict between their need to seek help and their need to conceal their difficulty. Couples present with the problem after having been married several months or several years. Masters and Johnson reported an unconsummated marriage of 17 years duration.

Frequently, the couple does not seek help directly, but the woman may reveal the problem to her gynecologist on a visit ostensibly concerned with vague vaginal or other somatic complaints. On examining her, the gynecologist may find an intact hymen. In some cases though, the wife may have undergone a hymenectomy to resolve the problem. That surgical procedure is another stress and sometimes increases the feelings of inadequacy in the couple. The wife may feel put upon, abused, or mutilated, and the husband's concern about his manliness may increase. Inquiry by a physician who is comfortable in dealing with sexual problems may be the first opening to a frank discussion of the couple's distress. Often, the pretext of the medical visit is a discussion of contraceptive methods or—even more ironically—a request for an infertility workup. Once presented, the complaint can often be successfully treated. The duration of the problem does not significantly affect the prognosis or the outcome of the case.

The causes of unconsummated marriage are varied: lack of sex education, sexual prohibitions overly stressed by parents or society, problems of an oedipal nature, immaturity in both partners, overdependence on primary families, and problems in sexual identification. Religious orthodoxy, with severe control of sexual and social development or the equation of sexuality with sin or uncleanliness has also been cited as a dominant cause. Many women involved in an unconsummated marriage have distorted concepts about their vaginas. They may fear that the vagina is too small or too soft, or they may confuse the vagina with the rectum, leading to feelings of being unclean. The man may share in those distortions about the vagina and, in addition, perceive it as dangerous to himself. Similarly, both partners may have distortions about the man's penis, perceiving it as a weapon, as too large, or as too small. Many patients can be helped by simple education about genital anatomy and physiology, by suggestions for self-exploration, and by correct information from a physician. The problem of the unconsummated marriage is best treated by seeing both members of the couple. Dual-sex therapy has been markedly effective. However, other forms of conjoint therapy, marital counseling, traditional psychotherapy on a one-to-one basis, and counseling from a sensitive family physician, gynecologist, or urologist are all helpful.

Body Image Problems Some persons are ashamed of their bodies and experience feelings of inadequacy related to self-imposed standards of masculinity or femininity. They may insist on sex only during total darkness, not allow certain body parts to be seen or touched, or seek unnecessary operative procedures to deal with their imagined inadequacies. Body dysmorphic disorder should be ruled out.

Don Juanism Some men who appear to be hypersexual, as shown by their need to have many sexual encounters or conquests, use their sexual activities to mask deep feelings of inferiority. Some have unconscious homosexual impulses, which they deny by compulsive sexual contacts with women. After having sex, most Don Juans are no longer interested in the woman. The condition is also referred to as satyriasis or a form of sex addiction.

Nymphomania Nymphomania signifies excessive or pathological desire for coitus in a woman. There have been few scientific studies of the condition. Those patients who have been studied usually have had one or more sexual disorders, usually including female orgasmic disorder. The woman often has an intense fear of loss of love. She attempts to satisfy her dependency needs, rather than to gratify her sexual impulses through her actions. It is sometimes classified as a form of sex addiction.

Fantasies Other atypical disorders are found in persons who have one or more sexual fantasies about which they obsess, feel guilty, or are otherwise dysphoric. As indicated in Table 19.1a-25 however, the range of common sexual fantasies is broad.

Men	Women
Heterosexual	
Replacement of established partner	Replacement of established partner
Forced sexual encounters with woman	Forced sexual encounters with man
Observing sexual activity	Observing sexual activity
Sexual encounters with man	Idyllic encounters with unknown man
Group sex	Sexual encounters with woman
Homosexual	
Images of male anatomy	Forced sexual encounters with women
Forced sexual encounters with men	Idyllic encounters with established partner
Sexual encounters with women	Sexual encounters with man
Idyllic encounters with unknown men	Memories of past sexual experiences
Group sex	Sadistic imagery

* Listed in order of occurrence. A 1994 study found that one in five persons experienced a same-sex sexual fantasy at some time in their lives.
Adapted with permission from Masters WF, Johnson VE: The Masters and Johnson treatment program for dissatisfied homosexual men. Am J Psychiatry 145:173, 1988.

Table 19.1a-25 Common Sexual Fantasies*

Persistent and Marked Distress About Sexual Orientation Distress about sexual orientation is characterized by a dissatisfaction with homosexual arousal patterns, a desire to increase heterosexual arousal, and strong negative feelings about being homosexual. Occasional statements to the effect that life would be easier if the person were not homosexual do not constitute persistent and marked distress about sexual orientation.

Treatment of sexual orientation distress is controversial. One study reported that with a minimum of 350 hours of psychoanalytic therapy, about a third of about 100 bisexual and homosexual men achieved a heterosexual reorientation at 5-year follow-up, but that study was challenged and was never replicated. Behavior therapy and avoidance conditioning techniques have also been used, but a basic problem with behavioral technique is that the behavior may be changed in the laboratory setting but not outside the laboratory. Prognostic factors weighing in favor of heterosexual reorientation for men include being under 35 years of age, having some experience of heterosexual arousal, and having a high motivation for reorientation.

Another style of intervention is directed at enabling the person with persistent and marked distress about sexual orientation to live comfortably as a homosexual without shame, guilt, anxiety, or depression. Gay counseling centers are engaged with patients in such treatment programs. At present, outcome studies of such centers have not been reported in detail.

Few data are available about the treatment of women with persistent and marked distress about sexual orientation, and those are primarily single-case studies with variable outcomes.

Postcoital Headache Postcoital headache is a headache immediately after coitus and may last for several hours. It is usually described as throbbing, and it is localized in the occipital or frontal area. The cause is unknown but may be vascular, muscle contraction (tension), or psychogenic. Coitus may precipitate migraine or cluster headaches in predisposed persons.

Orgasmic Anhedonia Orgasmic anhedonia is a condition in which the person had no physical sensation of orgasm, even though the physiological component (e.g., ejaculation) remains intact. Medical causes, such as sacral and cephalic lesions that interfere with afferent pathways from the genitalia to the cortex, must be ruled out. Psychic causes usually relate to extreme guilt about experiencing sexual pleasure. Those feelings produce a type of dissociative response that isolates the affective component of the orgasmic experience form consciousness.

Female Premature Orgasm Data on female premature orgasm are lacking; no separate category of premature orgasm for women is included in DSM-IV. However, in the University of Chicago study, 10 percent of women felt they reached orgasm too quickly.

Masturbatory Pain Some persons may experience pain during masturbation. Organic causes should always be ruled out. A small vaginal tear or early Peyronie's disease may produce a painful sensation. The condition should be differentiated from compulsive masturbation. People may masturbate to the extent that they do physical damage to their genitals and eventually experience pain during subsequent masturbatory acts. Such cases constitute a separate sexual disorder and should be so classified.

Autoerotic Asphyxiation Certain masturbatory practices have resulted in what has been called autoerotic asphyxiation. The practices may involve masturbating while hanging by the neck to heighten erotic sensations and the intensity of the orgasm through the mechanism of mild hypoxia. Although they intend to release themselves from the noose after orgasm, an estimated 500 to 1000 persons a year accidentally kill themselves by hanging. Most who indulge in the practice are men; transvestism is often associated with the habit, and most deaths occur among adolescents. Such masochistic practices are usually associated with severe mental disorders such as schizophrenia and major mood disorders.

TREATMENT

The treatment of sexual disorders has evolved significantly since the 1970s when Masters and Johnson refocused the attention of the psychiatric community on sexual disorders. Innovations in treatment reflect the results of research and changes in the patient population. For example, in the late 1960s most cases of erectile dysfunction were considered psychological in origin, with about 20 percent of erectile problems having an organic cause. Currently, the numbers are reversed, with the majority of cases considered to result from a physiological problem. Actually, many cases are of mixed origin. Similarly, early patients in sex therapy were assumed to suffer in part from lack of sexual information and culturally reinforced negative attitudes toward sex. In the past 30 years the public has received a great deal of accurate information about sex through the media, and cultural attitudes regarding sexual behavior have become markedly more liberal. Patients now have more knowledge and greater sexual sophistication. Their dysfunctions have a complex etiology frequently involving psychodynamic and relational issues. Problems of desire are seen with increasing frequency and are among the most challenging cases for therapists. Couple issues involve problems of trust, intimacy, lack of sexual attraction, and struggles for dominance. Finally, along with the rest of psychiatry, sex therapy has experienced medicalization. Biological treatment approaches have developed rapidly. An eclectic approach that allows the use of several techniques sequentially or in combination may be necessary.

Before entering therapy, the patient should have a thorough medical evaluation, including a medical history, physical examination, and appropriate laboratory studies when necessary. If a medical cause for the disorder is found, treatment should be directed toward ameliorating that cause.

Prior to 1970 the most common treatment of sexual dysfunction was individual psychotherapy. Classic psychodynamic theory considers sexual inadequacy to have its roots in early developmental conflicts, and the sexual disorder is treated as part of a more pervasive emotional disturbance. Treatment focuses on the exploration of unconscious conflicts, motivation, fantasy, and various interpersonal difficulties. Therapy assumes that removal of the conflicts will allow the sexual impulse to become structurally acceptable to the patient's ego and thereby find appropriate means of satisfaction in the environment. Unfortunately, the symptom of sexual dysfunction frequently becomes secondarily autonomous and persists after resolution of the other problems evolving from the patient's pathology. The addition of behavioral techniques are often necessary to cure the sexual problem.

Dual-Sex Therapy The theoretical basis of the dual-sex therapy approach is the concept of the marital unit or dyad as the object of therapy. The method of dual-sex therapy was originated and developed by Masters and Johnson. Dual-sex therapy does not accept the idea of a sick half of a patient couple. Both are involved in a relationship in which there is sexual distress, and thus both must participate in the therapy program.

The sexual problem often reflects other areas of disharmony or misunderstanding in the marriage. The marital relationship as a whole is treated, with emphasis on sexual functioning as a part of that relationship. Improved communication in sexual and nonsexual areas is a specific goal of treatment. Psychological and physiological aspects of sexual functioning are discussed with an educational attitude. Suggestions are followed in the privacy of the couple's home.

Initial histories are taken to determine suitability for this type of treatment. Evidence of major underlying psychopathology suggests further psychiatric evaluation, and participation in the program may be deferred until the patient seems better able to benefit from it. Concurrent psychotherapy with a psychiatrist while participating in

dual-sex therapy is sometimes recommended.

Each patient is interviewed individually early in the course of treatment. A complete sexual history is obtained, which is later reflected back to the couple, to help them understand their present problem. The individual sessions also help the therapist understand the patients' lifestyle and make suggestions that fit into that lifestyle.

Behavioral Exercises Treatment is short term and behaviorally oriented. Specific exercises are prescribed to help the couple with their particular problem. Sexual dysfunction often involves a fear of inadequate performance; thus couples are specifically prohibited from any sexual play other than that prescribed by the therapist. Initially, intercourse is interdicted, and couples learn to give and receive bodily pleasure without the pressure of performance. Beginning exercises usually focus on heightening sensory awareness to touch, sight, sound, and smell.

During those exercises, called *sensate focus exercises*, the couple is given much reinforcement to lessen anxiety. They are urged to use fantasies to distract them from obsessive concerns about performance, which is termed *spectatoring*. The needs of both the dysfunctional partner and the nondysfunctional partner are considered. If either partner becomes sexually excited by the exercises, the other is encouraged to bring him or her to orgasm by manual or oral means. That procedure is important to keep the nondysfunctional partner from sabotaging the treatment. Open communication between the partners is urged, and the expression of mutual needs is encouraged. Resistances, such as claims of fatigue or not enough time to complete the exercises, are common and must be dealt with by the therapist. Genital stimulation is eventually added to general body stimulation. The couple is sequentially taught to try various positions for intercourse without necessarily completing the act, and to use varieties of stimulating techniques before they are permitted to proceed with intercourse.

The specific exercises vary with differing presenting complaints, and special techniques are used to treat the various dysfunctions. In cases of vaginismus, for instance, the woman is advised to dilate her vaginal opening with her fingers or with size-graduated vaginal dilators as part of the therapy. In cases of premature ejaculation an exercise known as the *squeeze technique* is used to raise the threshold of penile excitability. In that exercise the man or the woman stimulates the erect penis until the earliest sensations of impending orgasm and ejaculation are felt. Penile stimulation is then stopped abruptly, and the coronal ridge of the penis is forcibly squeezed for several seconds. The technique is repeated several times. A variation is the *stop-start technique* in which stimulation is interrupted for several seconds but no squeeze is applied. Masturbation to the point of imminent orgasm raises the threshold of excitability to a more tolerant stimulation level. The man is encouraged to focus on sensations of excitement rather than distract himself from them. This makes him more familiar with his excitement pattern and lets him feel in control rather than overwhelmed by sensations of arousal. Communication between the partners is improved, because the man must let his partner know his level of sexual excitement so that she can squeeze the penis before the ejaculatory process has started. Sex therapy has been successful with some premature ejaculators; however, a subgroup of dysfunctional men may need pharmacotherapy as well.

A man with sexual desire disorder or erectile disorder is sometimes told to masturbate to demonstrate that full erection and ejaculation are possible. A woman with lifelong female orgasmic disorder is directed to masturbate, sometimes using a vibrator. Kegel's exercises may be introduced to strengthen the pubococcygeal muscles; that is, the woman is encouraged to contract her abdominal and perineal muscles during masturbation and coitus. When a man has erectile disorder, the woman may be instructed to stimulate or tease his penis. The same technique is used with men who suffer from retarded ejaculation, with stimulation sometimes involving a vibrator. Retarded ejaculation is managed by extravaginal ejaculation initially and gradual vaginal entry after stimulation to the point of near ejaculation.

Treatment Goals The overall goal of treatment is to initiate an educational process, to diminish the fears of performance felt by both sexes, and to facilitate communication in sexual and nonsexual areas. Therapy sessions follow each new exercise period, and problems and satisfactions (both sexual and nonsexual) are discussed. Specific instructions and new exercises geared to the individual couple's progress are reviewed in each session. Gradually, the couple gains confidence and learns (or relearns) to communicate verbally and sexually. Dual-sex therapy is most effective when the sexual dysfunction exists apart from other psychopathology.

Hypnotherapy Hyponterapists focus specifically on the anxiety-producing symptom, that is, the particular sexual dysfunction. Successful use of hypnosis helps the patient gain control over the symptom that has been lowering self-esteem and disrupting psychological homeostasis. Patient cooperation is first obtained and encouraged during a series of nonhypnotic sessions with the therapist, designed to develop a secure doctor-patient relationship and a sense of physical and psychological comfort on the part of the patient and establish mutually desired treatment goals. During that time, the therapist assesses the patient's capacity for the trance experience. The nonhypnotic sessions also permit the clinician to take a careful psychiatric history and do a mental status examination before beginning hypnotherapy. Treatment focuses on symptom removal and attitude alteration. In a trance state, patients can entertain ideas incongruent with their usual (nonhypnotized) perceptions of reality. Patients are instructed in developing alternative means of dealing with the anxiety-provoking situation (i.e., the sexual encounter).

For example, a woman suffering from vaginismus is given the posthypnotic suggestion that she will feel no pain during intercourse and will be able to relax the muscles surrounding her vagina. If compliance with the suggestion is successful, she can deal with the anxiety produced by the sex act. She is also taught new attitudes, such as being entitled to sexual pleasure. Under hypnosis, her fear or anger at sexual contact can be examined, and she learns how her emotions are expressed by involuntary vaginal spasms.

Some patients respond particularly well to the use of self-hypnosis and indirect suggestion. These techniques allow them to retain a greater sense of control over their situation. Typically, patients are instructed to conjure up images and develop ideas antithetical to their dysfunctional responses. For example, a woman with an arousal disorder may first agree to concentrate on imagery that causes her to salivate. She is then told that just as she has made her mouth water by focusing on stimulating images, she can effect the lubricating response of her vagina by focusing on images she finds erotic or romantic. At the same time, the therapist helps her deal with her anxieties about a positive sexual response. Patients are also taught relaxing techniques to use before sexual relations. With these methods to alleviate anxiety, the physiological responses to sexual stimulation can more readily result in pleasurable excitation and discharge. Hypnosis may be added to a basic individual psychotherapy program to accelerate the impact of psychotherapeutic intervention.

Behavior Therapy Behavior therapists assume that sexual dysfunction is learned, maladaptive behavior. Behavioral approaches were initially designed to treat phobias. In cases of sexual dysfunction the therapist sees the patient as phobic of sexual interaction. Using traditional techniques, the therapist sets up a hierarchy of anxiety-provoking situations for the patient, ranging from the least threatening to the most threatening. Mild anxiety may be experienced at the thought of kissing, and massive anxiety may be felt when imagining penile penetration. The behavior therapist helps the patient master the anxiety through a standard program of systematic desensitization. The program is designed to inhibit the learned anxious response by encouraging behaviors antithetical to anxiety. The patient first deals with the least anxiety–producing situation in fantasy and progresses by steps to the most anxiety–producing situation. Medication, hypnosis, or special training in deep muscle relaxation is sometimes used to help with the initial mastery of anxiety.

Assertiveness training helps teach patients to express their sexual needs openly and without fear. Exercises in assertiveness are given in conjunction with sex therapy, and patients are encouraged both to make sexual requests and to refuse to comply with requests perceived as unreasonable. Sexual exercises may be prescribed for patients to perform at home, and a hierarchy may be established, starting with activities that proved most pleasurable and successful in the past.

One treatment variation involves the participation of the patient's sexual partner in the desensitization program. The partner, rather than the therapist, presents the hierarchical items to the patient. In such situations a cooperative partner is necessary to help the patient carry gains made during treatment sessions to sexual activity at home.

Behavior therapy techniques have been particularly effective in treating women with severe inhibition of excitement and orgasm when such feelings were accompanied by strong feelings of anxiety, anger, or disgust.

Group Therapy Methods of group therapy have been used to examine both intrapsychic and interpersonal problems in patients with sexual disorders. The therapy group provides a strong support system for patients who feel ashamed, anxious, or guilty about a particular sexual problem. It is a useful forum in which to counteract sexual myths, correct misconceptions, and provide accurate information regarding sexual anatomy, physiology, and varieties of behavior.

Groups for the treatment of sexual disorders can be organized in several ways. Members may all share the same problem, such as premature ejaculation; members may all be of the same sex with different sexual problems; or groups may be composed of both men and women who are experiencing different sexual problems. Group therapy may be an adjunct to other forms of therapy or the prime mode of treatment. Groups organized to cure a particular dysfunction usually have a behavioral approach. For example, patients suffering from anorgasmia may participate with others suffering from the same problem in a short-term, intensive group experience. Sexual histories, feelings of inadequacy, and concerns about body image are shared. Specific physiological information, sometimes with the aid of

audiovisual materials, is presented to the group members. Members are given homework assignments (e.g., they may be instructed to masturbate). A combination of group support and group pressure helps some of the participants complete assignments they might otherwise avoid. As the short-term group process nears termination, members are encouraged to talk about their experiences with their partners.

Groups have also been effective when composed of sexually dysfunctional married couples. The group provides an opportunity to gather accurate information, provides consensual validation of individual preferences, and enhances self-esteem and self-acceptance. Such techniques as role playing and psychodrama may be used in treatment. These groups are not indicated for couples in which one partner is uncooperative, a patient is suffering from a severe depression or psychosis, or a strong repugnance for explicit sexual audiovisual materials or a strong fear of groups exists.

Integrated Sex Therapy One of the most effective treatment modalities is the use of sex therapy integrated with supportive, psychodynamic, or insight-orientated psychotherapy. Adding psychodynamic conceptualizations to the behavioral techniques used to treat sexual dysfunctions allows treatment of patients with sex disorders associated with other psychopathology. Also, the therapy is appropriate for patients suffering from hypoactive desire disorders. Insight-oriented therapy helps them deal with problems in their interpersonal relationships or intrapsychic conflicts that frequently are at the root of the problem. The themes and dynamics that emerge in patients in analytically oriented sex therapy are the same as those seen in psychoanalytic therapy, such as relevant dreams, fear of punishment, aggressive feelings, difficulty with trusting the partner, fear of intimacy, oedipal feelings, and fear of genital mutilation.

A 34-year-old widow presented for therapy with a chief complaint of vaginismus. Her marriage of 3 years, which had been unconsummated, ended when her husband was killed in a car accident. Approximately 1 year after she lost her husband, the patient became involved with a married man. She was very attracted to him and became highly aroused during their sexual encounters. Although she could reach orgasm through manual or oral stimulation, she could not tolerate penetration. Although she never considered therapy when she was married, in spite of her husband's requests to do so, she was motivated to seek help for her problem because she felt sure her lover would leave his wife for her if they could share a more complete sexual experience.

The patient's vaginismus resulted partly from unresolved developmental conflicts. Her parents had been loving but rigid people who came from different socioeconomic backgrounds. Their values often conflicted, and they frequently fought over their daughter as she entered adolescence. The mother insisted that she take an academic course in high school to prepare for college, whereas the father pushed a more practical business program. The patient sided with her mother and felt that her father, whom she had always perceived as cold, became more distant than before. Some of her difficulties were due to unresolved oedipal problems; both her husband and her lover were more than 20 years older than she was, and her lover (reflecting her parental situation) was married to a woman who was more successful than he was. In addition, she had identified with some of her mother's negative feelings about men. The mother had once told the patient she hoped that she would be spared marriage. Vaginismus protected the patient from the closeness with men that she consciously wanted but unconsciously perceived as hurtful and dangerous.

A 56-year-old man came for treatment because of an erectile disorder. In general, he functioned better in extramarital affairs than in his marriage. Although he loved his wife and considered her an attractive woman, he believed that she was not interested in sex. He could rarely achieve an erection with her and he gradually stopped approaching her sexually. His wife felt deprived by their lack of sexual relations and frequently masturbated.

The patient had been a sickly child, with a mother he described as devoted but smothering. He remembered her cuddling him in bed until he was 8 years old, and he felt that she was inappropriately affectionate in general (“she embarrassed me”). At the same time, he remembered his father as an earthy man and had a childhood recollection of hearing his mother ask his father, “How could you, how could you?” The patient believed it had been his mother's response to a sexual overture or act. In part, his disorder derived from his unconscious oedipal associations to his wife, which made her taboo for him as a sex partner. The women to whom he responded had to be blatantly sexual and signal their acceptance of him before he would risk an advance. Therapy involved both individual sessions with the patient and joint sessions with him and his wife. Communication, which had been strained partly because of the sexual distance between them, was encouraged, and a behavioral approach was used to reestablish some physical interaction. Individual work focused on his deeper psychological problems.

The dynamics and the emotional difficulties evident in the vignettes are seen every day by the psychiatrist. Psychiatrists can readily absorb the techniques of sex therapy into their treatment armamentarium, just as they have modified and absorbed any number of specialized techniques, from classic analytic dynamic formulations to the use of pharmacotherapy, group therapy techniques, and behavioral and other directive modalities. The combined approach of individual and sex therapy is used by the general psychiatrist, who carefully judges the optimal timing of sex therapy and the ability of patients to tolerate the directive approach that focuses on their sexual difficulties.

Biological Treatment Methods Biological treatments, including pharmacotherapy and surgery, have application in specific cases of sexual disorder. Advances in biological treatment methods, particularly pharmacological methods of treating sexual dysfunction have significantly augmented the therapist's catalogue of therapeutic approaches. Most of the recent advances involve male sexual dysfunctions. Current studies are under way to test pharmacological treatment of sexual dysfunctions in women.

Pharmacotherapy A variety of drugs have been explored in the treatment of sexual dysfunction. The major new medications are sildenafil; oral prostaglandin (Vasomax); alprostadil (Caverject), an injectable phentolamine; and a transurethral alprostadil (MUSE), all used in the treatment of erectile disorder.

Sildenafil Sildenafil is a nitric oxide enhancer that facilitates the inflow of blood to the penis necessary for an erection. The physiological mechanism of penile erection involves release of nitric oxide in the corpus cavernosum during sexual stimulation. Nitric oxide activates the enzyme guanylate cyclase, which increases cyclic guanosine monophosphate concentration and produces smooth muscle relaxation in the corpus cavernosum that allows the penile vessels to dilate and admit blood. Sildenafil enhances the effect of nitric oxide by inhibiting the enzyme that degrades cyclic guanosine monophosphate. In short, sidenafil augments the natural process involved in gaining and maintaining an erection during sexual stimulation. The drug takes effect about 1 hour after ingestion, and its effect can last up to 4 hours. Sildenafil has no effect in the absence of sexual stimulation.

The most common adverse events associated with sidenafil are headaches, flushing, and dyspepsia. Some sildenafil users see things in a blue tint for several hours after taking the medication; because of this, airline pilots have been prohibited from taking the drug, so that this visual artifact does not interfere with safe landings. The use of sildenafil is contraindicated for people taking organic nitrates. The concomitant action of the two drugs can result in large, sudden, and sometimes fatal drops in systemic blood pressure. The Food and Drug Administration (FDA) has posted 130 deaths in which sidenafil was listed as an associated medication and advises caution in prescribing sidenafil to men with a recent (6 month) history of myocardial infarction, stroke, life threatening arrhythmia, significant hypotension or hypertension, cardiac failure, angina or retinitis pigmentosa. However, the FDA has reiterated that it is a safe drug. Sildenafil is not effective in all cases of erectile dysfunction. It fails to produce an erection rigid enough for penetration in about 50 percent of men who have had radical prostate surgery or in those with long-standing insulin-dependent diabetes. It is also ineffective in certain cases of nerve damage.

Oral phentolamine has proved effective as a potency enhancer in men with minimal erectile dysfunction. It may prove useful for men with cardiac problems, since sildenafil is contraindicated for men using organic nitrates, but it is not currently approved by the FDA. Apomorphine is also being tested as an oral remedy for erectile dysfunction.

Alprostadil In contrast to the oral medications, injectable and transurethral alprostadil act locally on the penis and can produce erections in the absence of sexual stimulation. Alprostadil contains a naturally occurring form of prostaglandin E. Prostaglandins are composed of complex hydroxy fatty acids, and they have wide biological influences. Although some prostaglandins are vasoconstrictive, prostaglandin E_1 found in alprostadil is a powerful vasodilating agent, especially in local vascular areas such as the corpus cavernosum of the penis. The drug causes direct smooth muscle relaxation of penile vessels and erectile tissue; this reaction lowers the vascular resistance of the corpus and significantly increases blood flow to the penis. The firm erection produced with 2 to 3 minutes by increased blood flow may last as long as 1 hour. Treatment consists of the patient's self-injection of alprostadil into the corpus before coitus. This technique is easily taught and relatively painless. Infrequent adverse effects include penile bruising and changes in liver function test results, which are readily reversible when a man stops the injections. However, possible hazardous sequelae exist, including priapism and sclerosis of the small veins of the penis. Another substance being tried is vasoactive intestinal polypeptide (VIP). Intracavernous injection of VIP causes erection and has a parasympathomimetic effect. Some researchers speculate that this substance, which has been found in the hypothalamus and the female genital organs, is the essential factor in male and female arousal. In Europe phenoxybenzamine (Dibenzyline) is used to produce erections by injection into the penis. Serious adverse effects include priapism and pain accompanying the injection, and the drug is not allowed as a therapy in the United States.

Alprostadil can also be delivered via the urethra, eliminating the need for self-injection. Some men prefer the local, nonsystemic effects of alprostadil to oral sildenafil, others prefer sildenafil because it seems more like a nonpharmacologically aided response to them and their partners. A small trial found a topical cream effective in alleviating erectile dysfunction. The cream consists of three vasoactive substances that are a mixture of ergot alkaloids.

The pharmacological treatments described above are useful in treatment of erectile dysfunction of various causes: psychogenic, neurogenic, arterial insufficiency, venous leakage, and mixed. The following case demonstrates the use of pharmacotherapy to treat erectile dysfunction of psychogenic origin.

> Mr. B. and Ms. C. (his fiancee) presented for sex therapy with Mr. B. having experienced erectile dysfunction for 7 months' as well as having a lifelong history of premature ejaculation. Mr B. was a 42-year-old professional, and Ms. C. was a 38-year-old corporate executive. She had never been married; it was to be his second marriage.
>
> Mr. B.'s sexual history was remarkable for a late first coitus (age 25) and a general sexual insecurity that contrasted with his professional confidence and social poise. Ms. C. was sexually responsive in spite of a history of anxiety attacks for which she had been treated in the past. Their early sexual activity together had been frequent and satisfying to both in spite of his prematurity. After a few months, however, she began to complain about both the premature ejaculation and the secrecy of their relationship. Although Mr. B. was legally separated when they became involved, he was worried about divorce negotiations if his spouse learned about his new relationship. Treatment revealed that he found the secrecy in their relationship exciting.
>
> The two were treated with integrated sex therapy, a combination of behavioral and insight-oriented technique, and made substantial progress. They successfully controlled the premature ejaculation by practicing the squeeze technique (i.e., forcibly squeezing the coronal ridge of the penis before ejaculation, to increase the threshold of penile sensitivity). In genital-caressing sessions Mr. B.'s erections returned, and he could maintain a good erection to climax, with manual and oral stimulation. However, he continued to lose his erection when he attempted vaginal penetration. It was decided to supplement psychotherapy with intracavernosal injections of alprostadil (sildenafil was not yet available). Ms. C. was present when he was instructed in the injection technique and supported the process. Mr. B. was delighted with the results, and after 1 month of pharmacologically assisted coitus, he successfully achieved penetration without assistance. Currently, he has coitus once a week, with occasional use of sildenafil rather than intracavernosal injection when he is feeling stressed. The availability of a medication to treat his erectile problem significantly relieves his performance anxiety.

Erectile dysfunction of psychologic or mixed origin should not be treated by medication alone, even though the dysfunction can be corrected pharmacologically. The drugs should be used in conjunction with, not as a replacement for, sex therapy. In some cases, erectile dysfunction serves as a defense against unconscious conflicts that must be faced when the sexual symptom is removed. Therapeutic intervention is essential in such cases to prepare the patient to deal with his conflicts or to judge that he would be unable to do so and then to postpone pharmacological treatment. Sometimes the patient deals with his conflicts by simply not using the medications prescribed for his erectile problem. Also, patients in long-standing or marital relationships require the cooperation of their partners for this treatment to be effective. Although most women are very accommodating to their partner's desire for treatment, several concerns occur with some frequency: the element of romance is an important part of sexual interaction for many women, and pharmacological assistance of erections may eliminate that sense of romance; the woman, or the couple, may bemoan the lack of spontaneity when pharmacotherapy is part of the sex act; and some women feel deprived of feedback about their desirability. Joint sessions help the couple cope with these issues.

Sildenafil Use in Women The use of sildenafil in women is being researched, but no completed studies have been published. Anecdotal reports describe individual women who have experienced intensified excitement when they have tried sildenafil (presumably their partner's).

The physical sign of sexual excitement in the female is vaginal lubrication. That lubrication is a transudate believed to result from increased vasocongestion of the extensive capillary net in the vaginal walls. The same physiological mechanism, vasocongestion, results in erection in the male. Researchers believe that sildenafil may facilitate blood flow in the female just as it does in the male and thus give women with inhibited excitement of psychological or physiological origin a pharmacological remedy for their dysfunction. However, as in men, pharmacological treatment may need to be combined with psychological modalities to be effective.

Recognition of sexual excitement may be more complex for women than it is for men. For example, in studies of response to pornography, men and women underwent physiological measurements of excitement following visual stimuli. In these studies, men were more accurate in correlating their subjective sense of arousal with the objectively measured physiological response. Also, 50 percent of women experience postpubertal sexual stimulation and orgasm for the first time in the course of sex play with a partner, while most men first experience these sensations with masturbation. Many women face a greater challenge in identifying themselves as autonomously sexual than do men and consequently may not respond to pharmacotherapy alone.

Other Pharmacological Agents Numerous other pharmacological agents have been used to cure the various sexual disorders. Intravenous methohexital sodium (Brevital) has been used in desensitization therapy. Antianxiety agents may have some application in tense patients, although these drugs can also interfere with the sexual response. The side effects of antidepressants, in particular the SSRIs and tricyclic drugs, have been used to prolong the sexual response in patients with premature ejaculation. This approach is particularly useful in patients refractory to behavioral techniques, who may fall in the category now being investigated of physiologically determined premature ejaculation. The use of antidepressants has also been advocated in the treatment of patients who are phobic of sex and in those with a posttraumatic stress disorder following rape. The risks of taking such medications must be carefully weighed against their possible benefits. Bromocriptine is used in the treatment of hyperprolactinemia which is frequently associated with hypogonadism. Such cases are first worked up to rule out pituitary tumors. Bromocriptine (Parlodel), a dopamine agonist, may improve sexual function impaired by hyperprolactinemia. Sometimes the problem requires androgen therapy.

A number of substances have popular standing as aphrodisiacs, for example, ginseng root and yohimbine. However, studies have not confirmed any aphrodisiac properties. Yohimbine, an a-adrenergic receptor antagonist, may cause dilation of the penile artery by that mechanism. Many recreational drugs, including cocaine, amphetamines, alcohol, and cannabis, are considered enhancers of sexual performance. Although they may provide the user with an initial benefit because of their tranquilizing, disinhibiting, or mood-elevating effects, consistent or prolonged use of any of these substances impairs sexual functioning.

Ginseng has been reported to have androgenic effects. One report described the case of a mother who ingested large amounts of ginseng during her pregnancy, resulting in androgenization of the neonate, who was born with pubic hair and enlarged testes.

Dopaminergic agents have been reported to increase libido and improve sex function. Those drugs include L-dopa, a dopamine precursor, and bromocriptine, a dopamine agonist. The antidepressant bupropion has dopaminergic effects and has increased sex drive in some patients. Selegiline, an MAOI, is selective for MAO_B and is dopaminergic. It improves sexual functioning in older persons.

Hormone Therapy Androgens increase the sex drive in women and in men with low testosterone concentrations. Women may experience virilizing effects, some of which are irreversible (e.g., deepening of the voice). In men, prolonged use of androgens produces hypertension and prostatic enlargement. Testosterone is most effective when given parenterally; however, effective oral and transdermal preparations are available.

Gonadotrophin-releasing hormone (GnRH), also known as luteinizing hormone–releasing hormone (LHRH), stimulates the release of luteinizing hormone (LH), which increases testosterone secretion in both sexes. GnRH is used as an inhalant in Europe. It stimulates desire and increases potency. Since GnRH is released normally in a pulsatile fashion, portable infusion pumps have been developed that simulate pulsatile delivery. An excess of GnRH suppresses estrogen and testosterone; thus the therapeutic use of GnRH is limited by a narrow therapeutic window.

Women who use estrogens for replacement therapy or for contraception may report decreased libido; in such cases a combined preparation of estrogen and testosterone has been used effectively. Estrogen, itself, prevents thinning of the vaginal mucous membrane and facilitates lubrication.

Pheromones are sexual scents that are found in animals and may be present in humans. They produce dramatic sex-seeking behavioral patterns in animals (e.g., male deer following female deer in estrus and mounting behavior in primates). Human pheromones are believed to be short-acting fatty acids present in vaginal secretions and male sweat. In one study women were consistently more attracted to items impregnated with a chemical derived from male sweat (a-androstenol) than to control items. In another study, the sweat of women was preserved in underarm pads they changed daily, and the date each pad was worn was correlated with the women's menstrual cycle. A second group of women was asked to smell these pads as they were rubbed above their upper lips, without knowing what they were and recognizing no scent but the alcohol preservative used on the pads. Depending on whether they were exposed to pads from the early or late part of the first group's

(wearers) menstrual cycle, the second group (sniffers) saw their own menstrual cycles shortened or lengthened. This area is still being researched.

Antiandrogens and Antiestrogens Estrogen and progesterone are antiandrogens that have been used to treat compulsive sexual behavior in men, usually in sex offenders. Clomiphene (Clomid) and tamoxifen (Nolvadex) are both antiestrogenic and both stimulate GnRH secretion and increase testosterone concentration, thereby increasing libido. Women being treated for breast cancer with tamoxifen report an increased libido.

Cyprosterone acetate is a strong antiandrogen used in male sex offenders. At doses of 100 to 200 mg a day, the sex drive disappears within 2 weeks. It is available in Europe but is in the investigational stage in the United States.

Mechanical Treatment Approaches

Steal Syndrome In male patients with arteriosclerosis (especially of the distal aorta, known as Leriche's syndrome), the erection may be lost during active pelvic thrusting. The need for increased blood in the gluteal muscles and others served by the ilial or hypogastric arteries takes blood away (steals) from the pudendal artery and thus interferes with penile blood flow. Relief may be obtained by decreasing pelvic thrusting, which is also aided by the woman's superior coital position.

Vacuum Pump Vacuum pumps are mechanical devices that patients without vascular disease can use to obtain erections. The blood drawn into the penis following the creation of the vacuum is kept there by a ring placed around the base of the penis. This device has no adverse effects, but it is cumbersome, and partners must be willing to accept its use. Some women complain that the penis is redder and cooler than when erection is produced by natural circumstances, and they find the process and the result objectionable.

Surgical Treatment

Male Prostheses Surgical treatment is infrequently advocated, but improved penile prosthetic devices are available for men with inadequate erectile response who are resistant to other treatment methods or who have medically caused deficiencies. There are two main types of prosthesis: a semirigid rod prosthesis that produces a permanent erection that can be positioned close to the body for concealment and an inflatable type that is implanted with its own reservoir and pump for inflation and deflation. The latter type is designed to mimic normal physiological functioning. Placing a penile prosthesis in a man who has lost the ability to ejaculate or to have an orgasm as a result of medical causes will not restore those functions. Men with prosthetic devices have generally reported satisfaction with their subsequent sexual functioning, but their wives report much less satisfaction. Presurgical counseling is strongly recommended so that the couple has a realistic expectation of what the prosthesis can do for their sex lives. Postsurgical counseling may also be necessary to help the couple adapt to their rediscovered ability to have intercourse. They may experience a high level of anxiety if their sex life had been inactive for a prolonged period before surgery. Prosthetic devices have been associated with severe adverse effects including perforation, infection, urinary retention, and persistent pain.

Some surgeons are attempting revascularization of the penis as a direct approach to treating erectile dysfunction resulting from vascular disorders. Such surgical procedures are indicated in patients with corporal shunts that allow normally entrapped blood to leak from the corporal spaces, leading to inadequate erections (steal phenomenon). Limited reports exist of prolonged success with the technique. Endarterectomy can be of benefit if aortoiliac occlusive disease is responsible for the erectile dysfunction.

Another medical treatment being studied for erectile disorders is electrostimulation at the base of the penis. Initial reports indicate minimal physical discomfort in patients receiving this therapy. However, response to treatment is inconsistent, and a problem exists in terms of maintaining erections. At the present time, the treatment seems to have no benefits.

Female Procedures Surgical approaches to female dysfunctions include hymenectomy in the case of dyspareunia in an unconsummated marriage, vaginoplasty in multiparous women complaining of lessened vaginal sensations, or freeing clitoral adhesions in women with inhibited excitement. Such surgical treatments have not been carefully studied and should be considered cautiously.

Outcome The results of different treatment methods have varied considerably since Masters and Johnson first reported positive results for their treatment approach in 1970. Masters and Johnson studied the failure rates of their patients (defined as the failure to initiate reversal of the basic symptom of the presenting dysfunction). They compared initial failure rates with 5-year follow-up findings for the same couples. Although some have criticized their definition of the percentage of presumed successes, other studies have confirmed the effectiveness of their approach. Demonstrating the effectiveness of traditional outpatient psychotherapy is just as difficult when therapy is oriented to sexual problems as it is in general. In some cases the patient improves in all areas except the sexual area. The more severe the psychopathology associated with a problem of long duration, the more adverse the outcome is likely to be.

The more difficult treatment cases involve couples with severe marital discord. Cases involving problems of fear of intimacy, excessive dependency, or excessive hostility are also complex. Other challenges are posed by patients with impulse disorders, unresolved homosexual conflicts, or fetishistic defenses. Patients phobic of sex also present treatment difficulties, as do patients diagnosed with lifelong dysfunctions. Desire disorders are particularly difficult to treat. They require longer, more intensive therapy than some other disorders, and their outcomes are very variable.

When behavioral approaches are used, empirical criteria that are supposed to predict outcome are more easily isolated. Using these criteria, for instance, couples who regularly practice assigned exercises appear to have a much greater likelihood of successful outcome than do more-resistant couples or those whose interaction involves sadomasochistic or depressive features or mechanisms of blame and projection. Flexibility of attitude is also a positive prognostic factor. Overall, younger couples tend to complete sex therapy more often than older couples. Couples whose interactional difficulties center on their sex problems, such as inhibition, frustration, or fear of performance failure, are also likely to respond well to therapy.

In general, methods that have proved effective singly or in combination include training in behavioral-sexual skills, systematic desensitization, directive marital counseling, traditional psychodynamic approaches, group therapy, and pharmacotherapy. Although most workers prefer to treat a couple for sexual dysfunctions, treatment of individual persons has also been successful.

The frequency of sessions is not a significant factor in treatment success. Thus, whether patients have intensive daily therapy over a period of 2 weeks, weekly therapy, or biweekly therapy appears to have little effect on the outcome of treatment. Also, the use of one therapist to treat a couple instead of a dual-sex cotherapy team is nearly as effective and certainly more practical.

Patients seen today are frequently older when they present for therapy, more informed about sex, and more likely to have dysfunctions of mixed etiology than those seen 30 years ago. A multimodal treatment regimen is often necessary. Future studies will evaluate results of an eclectic therapeutic approach to sexual disorders.

SUGGESTED CROSS-REFERENCES

Homosexuality is discussed in Section 19.1b , paraphilias in Section 19.2, and gender identity disorders in Section 19.3. The neuropsychological and neuropsychiatric aspects of HIV infection are covered in Section 2.8. Couples therapy is discussed in Section 30.5 and yohimbine and other pharmacological therapies are discussed in Section 31.33. The physical and sexual abuse of children, including incest, is the subject of Section 49.4.

SECTION REFERENCES

Araoz DL: Uses of hypnosis in the treatment of psychogenic sexual dysfunctions. Psychiatr Ann *16:* 102, 1986.

Assalian P, Margolese H: Treatment of antidepressant induced side effects. J Sex Marital Ther *22:* 3, 1996.

Brady JP: Behavior therapy and sex therapy. Am J Psychiatry *133:* 896, 1976.

Chessick RD: Thirty unresolved psycho dynamic questions pertaining to feminine psychology. Am J Psychother *42:* 86, 1988.

*Delgardo PL, McGauey CA, Moreno FA, Laukes C, Gelenberg AJ: Treatment strategies for depression and sexual dysfunction. J Clin Psychiatry *17:* 22, 1999.

Ellis A: *Studies in the Psychology of Sex*. Preston House, New York, 1936.

Goldstein I, Lue T, Padma-Nathatan H, Rosen R, Steers WD, Wicker PA: The Sildenafil study group, oral sildenafil in the treatment of erectile dysfunction. N Engl J Med *338:* 1397, 1996.

Herman J, LoPiccolo J: Clinical outcome of sex therapy. Arch Gen Psychiatry *40:* 443, 1983.

Kegeles SM, Adler NE, Irwin CE: Sexually active adolescents and condoms: Changes over one year in knowledge, attitudes and use. Am J Public Health *78:* 460, 1988.

Koppelman M, Parry BL, Hamilton JA, Alogna SW, Loreaux PL: Effect of bromocriptine on affect and libido in hyperprolactinemia. Am J Psychiatry *144:* 1037, 1987.

Koren G: Maternal ginseng use associated with neonatal androgenization. JAMA *264:* 2866, 1990.

Krafft-Ebing R: *Psychopathia Sexual*. Matthes & Seitz, Munich, 1984.

*Laughman E, Gagnon J, Michael R, Michaels S: *Sex in America*. University of Chicago Press, Chicago, 1994.

Leitenberg H, Detzer M, Srebnik D: Gender differences in masturbation and the relation of masturbation experience in preadolescence and/or early adolescence to sexual behavior and sexual adjustment in young adulthood. Arch Sex Behav *22:* 87, 1993.

Linet OI, Ogrinc FG (for the Alprostadil Study Group): Efficacy and safety of intracavernosal alprostadil in men with erectile dysfunction. N Engl J Med *334:* 875, 1996.

Loosen PT, Purdon SE, Pavlou SN: Effects on behavior of modulation of gonadal function in men with gonadotropin-releasing hormone antagonists. Am J Psychiatry *151:* 271, 1994.

*Masters WH, Johnson VE: *Human Sexual Response*. Little, Brown, Boston, 1970.

*Masters WH, Johnson VE: *Human Sexual Inadequacy*. Little, Brown, Boston, 1970.

Padma-Nathan H, Hellstrom WJG, Kaiser FE, Labasky RF, Lue TF, Nolten WE, Norwood PC, Peterson CA, Shabsigh R, Tam PY, Place VA, Gesundheit N (for the Medicated Urethral System for Erection [MUSE] Study Group: Treatment of men with erectile dysfunction with transurethral alprostadil. N Engl J Med *336:* 1, 1997.

*Purnine DM, Carey MP, Jorgensen RS: Gender differences regarding preferences for specific heterosexual practices. J Sex Marital Ther *20:* 271, 1994.

*Riley AJ: Life-long absence of sexual drive in a woman associated with 5-dihydrotestosterone deficiency. J Sex Marital Ther *25:* 13, 1999.

*Rosen RC, Lane RM, Menza M: Effects of SSRIs on sexual function: A critical review. J Clin Psychopharmacol *19:* 67, 1999.

*Sadock VA: The treatment of psychosexual dysfunctions: An overview. In *Psychiatry 1982. The American Psychiatric Association Annual Review*, L Grinspoon, editor. American Psychiatric Press, Washington, DC, 1982.

Sadock VA: Group psychotherapy of psychosexual dysfunctions. In *Comprehensive Group Psychotherapy*, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1983, p 286.

Seagraves RT, Seagraves KB: Human sexuality and aging. J Sex Educ Ther *21:* 88, 1995.

Semans JH: Premature ejaculation: A new approach. South Med J *49:* 353, 1956.

Shrainer-Engel P, Schiavi R: Lifetime psychopathology in individuals with low sexual desire. J Nerv Ment Dis *174:* 646, 1986.

Stein DJ, Hollander E, Anthony DT, Schneier FR: Serotonergic medications for sexual addictions, and paraphilias. J Clin Psychiatry *53:* 267, 1992.

Sternbach H: Age associated testosterone decline in men: Clinical issues for psychiatry. Am J Psychiatry *155:* 10, 1998.

Thase M, Reynolds C, Glanz L, Jennings JR, Sweitz DE, Kupper DJ, Frank E: Nocturnal penile tumescence in depressed men. Am J Psychiatry *144:* 89, 1987.

Waldinger MD, Hengeveld WH, Zwinderman A, Olivier B: Effect of SSRI antidepressants on ejaculation: A double-blind, randomized, placebo-controlled study with fluoxetine, fluvoxamine, paroxetine and sertaline. J Clin Psychopharmacol *189:* 274, 1998.

Wiley D, Borts WM: Sexuality and aging—usual and successful. J Gerontol *51:* 22, 1996.

CHAPTER 19. NORMAL HUMAN SEXUALITY AND SEXUAL AND GENDER IDENTITY DISORDERS

19.1 NORMAL HUMAN SEXUALITY

19.1B HOMOSEXUALITY AND HOMOSEXUAL BEHAVIOR

TERRY S. STEIN, M.D.

Perspectives on homosexuality—concerning its origins, social and personal meanings, diagnostic implications, and clinical relevance—have persistently represented one of the most contentious issues within psychiatry during the latter part of the twentieth century. In spite of a remarkable expansion in the understanding of homosexuality and bisexuality since the work of Alfred C. Kinsey, many psychiatrists continue to view sexual orientation as either homosexual, heterosexual, or bisexual and have failed to understand its complex developmental, biological, behavioral, and psychosocial dimensions. This reductionistic viewpoint ignores the richness and diversity of development, experience, and identity associated with sexual orientation in contemporary society. This chapter examines both the experience of homosexuality as it is expressed in individuals and the variety of theories and clinical and research findings available to explain sexual orientation.

DEFINITIONS

Increased understanding of sexuality in general has led to greater specificity in the use of terms related to sex, gender, sexuality, and sexual orientation. Terms such as *gender identity* and *gender* or *sex role* tended to be confused with sexual orientation in early writing about homosexuality; however, it is now recognized that these are related but discrete characteristics that may interact with but are not synonymous with homosexuality and have distinct developmental pathways. It is important to understand the precise meanings of these terms.

Sexual orientation refers to a person's erotic response tendency—homosexual, bisexual, or heterosexual—toward other persons of the same or other sex as reflected in such indicators as the proportion of dreams and fantasies directed to one or the other sex, the sex of one's sexual partners, and the extent of physiological response to erotic stimuli associated with one or both genders. Sexual orientation consists of three components: desire, behavior, and identity, which may or may not be congruent in an individual. The Kinsey heterosexual-homosexual scale—a 7-point continuous scale with 0 representing exclusive heterosexuality, 6 representing exclusive homosexuality, and 3 representing equal amounts of both, or bisexuality—is the scale most widely used to describe sexual orientation. The Kinsey scale has been criticized for its unidimensional and bipolar conceptualization of sexual orientation that suggests a diminished attraction toward one sex in proportion to an increased attraction to the other sex, as opposed to independently describing the degree of erotic attraction in an individual to each sex.

The terms *sexual orientation* and *sexual preference* are often used interchangeably; however, the former term has had a wider currency in contemporary professional and popular usage because it is generally accepted to refer more narrowly to a person's erotic response tendency. In contrast, sexual preference is viewed as implicitly suggesting that an individual prefers and therefore also chooses an erotic attraction toward one sex or the other. Such an association between desire and choice is often thought, particularly by those biological determinists who believe that sexual attractions are irrevocably set by inborn factors, to be inappropriate and potentially politically dangerous because it is seen as open to attempts to forcibly change a sexual attraction that is viewed negatively by society. Neither term necessarily implies anything about the degree of volition persons have with regard to their sexual desire; historically medical, legal, and moral justifications have been used to try to change both personal preferences and biologically determined characteristics. In this section *sexual orientation* is used to refer to erotic response tendency, not because it is semantically a better term, but only because it is more commonly used in the psychiatric literature.

Homosexuality was first used as a term in the second half of the nineteenth century to refer to erotic desire for persons of the same sex. *Heterosexuality* was developed later as a term to describe erotic desire for persons of the opposite sex. A third category of sexual orientation, *bisexuality*, was used by Sigmund Freud and others to describe an attraction to members of both sexes. These terms have not been applied consistently in theoretical, research, or popular discourse; they have been used to refer to a wide range of constructs, including categories of sexual desire, gender role attributes, forms of sexual behavior, personal and social identities, types of personalities and persons, degrees of normality and abnormality, and the presence or absence of mental illness. This conceptual confusion, along with an underlying assumption that sexual orientation as a concept refers to some unitary quality within different individuals, has complicated efforts to understand the complex personal and social meanings of sexual desire and relationships.

In this section *homosexuality* is used solely to refer to an erotic desire for someone of the same biological sex, and its presence in an individual does not imply the concurrent existence of any other characteristics in that individual. When used as an adjective, *homosexual* is intended to refer to sexual ideation or activity involving members of the same sex. Thus, for example, a woman may engage in homosexual behavior, demonstrate typical feminine gender role characteristics, be married to a man, and experience a heterosexual identity; or a man may have homosexual desire, have sex only with women, and show gender role nonconformity. These different characteristics have distinct developmental and expressive pathways in each individual; however, as the result of socialization and because of the possible existence of predisposing biological factors, their development will be interactive and reinforcing in some individuals.

Bisexuality refers to an erotic propensity, an individual identity, and sexual behavior. It may occur both sequentially and simultaneously as an expression of a sexual orientation directed to the same and the other sex. Although bisexuality is often devalued or ignored and has been portrayed as a transitional state in development to a heterosexual or gay or lesbian identity, it describes the sexual behavior and the sexual orientation of a large number of men and women at some point in their lives and is a persistent sexual identity for an increasingly visible group of individuals. There are frequent parallels between the experiences of gay men, lesbian women, and bisexual men and women.

In the latter part of the twentieth century, the terms *gay, gay man, lesbian*, and *bisexual* have been used to refer to men and women who have developed a sexual orientation identity that is homosexual or bisexual. In much of the early psychological and psychoanalytic literature about homosexuality, *homosexual* was used as a noun to designate a person with homosexual desire or behavior, however, at this point the term is used in this manner less often in the scientific and popular literature. When used as a noun the term may have a negative connotation because of its origins in medicine, its historical association with psychopathology, and its reification of the idea of a category of person, generally presented as a negative stereotype, designated solely by sexual orientation. Although *gay* is sometimes used as an inclusive term to refer to both men and women, reference to two distinct groups, gay men and lesbians, is more appropriate because of the significant differences in the development and experiences of these two groups.

Homosexual, or lesbian and gay, and bisexual identities are the subjective experiences of being homosexual or bisexual—being significantly or exclusively attracted to someone of the same sex—with the identities being both self- and other-ascribed. The acquisition of lesbian, gay, and bisexual identities is generally conceptualized as a developmental process that occurs over time. Models of lesbian, gay, and bisexual identity development usually portray a series of linear progressive stages involving tasks such as *coming out* (the process of recognizing one's homosexual or bisexual attraction and acknowledging it to oneself and to others); involvement in lesbian, gay, and bisexual communities; establishing same-sex relationships; and integrating one's sexual identity into other aspects of the self. It is important to understand that not all persons who experience homoerotic desire or participate in homosexual behavior develop a lesbian, gay, or bisexual identity. Thus, the clinician must be careful when evaluating a patient's sexuality to inquire about all of the components of sexual orientation: desire, behavior, and identity. Congruence between these components should not be assumed based upon information about only one or two of them.

Although a *heterosexual identity* refers in theory to the experience of those men and women who primarily are aware of attraction to persons of the other sex, this construct is not as well developed as that of lesbian, gay, and bisexual identities. Heterosexual identity is often not consciously experienced by individuals in our society because it is associated with the dominant sexual orientation. A heterosexual developmental outcome has been described as compulsory and is usually

assumed to be a person's identity unless the individual does something to disavow it or to proclaim an alternative identity. Thus, acquisition of a heterosexual identity cannot by definition involve coming out because it is the default identity in contemporary Western society.

In contrast to sexual orientation, *gender identity* is generally understood to refer to a persistent sense of oneself as being male, female, or ambivalent and is contrasted with *sex*, which refers to the biological attributes of being male or female. *Gender identity* describes an inner experience of being male or female, *gender role* and *social-sex role* are defined as the degree to which an individual's outer behavior or appearance can be described as masculine, feminine, or androgynous. The large majority of persons, regardless of their sexual orientation, have a gender identity and gender role consistent with their biological sex, although sometimes gay and lesbian persons may demonstrate a greater range of gender role attributes.

Sometimes people with variant gender identities and gender roles are diagnosed as having a type of mental disorder. *Transsexualism* consists of a strong and persistent cross-gender identification, discomfort with one's biological sex, and a wish to acquire the characteristics of the other sex. When these characteristics are present in a person, some researchers believe that they should be diagnosed as a *gender identity disorder. Transgendered* is a more recent term that is used to describe a person who identifies with and adopts the gender identity of a member of the other biological sex. *Transvestism* consists of sexual urges and fantasies involving cross-dressing which, when distressing to a patient, may be diagnosed as a *transvestic fetishism*, a type of *paraphilia*. Historically these experiences have often been confused with homosexuality, but they are separate and distinct phenomena.

THEORIES

Prior to the time homosexuality was first studied as a psychiatric condition in the second half of the nineteenth century, the meaning of homosexual behavior was very different than it is today in Western culture. Homosexual practices played an important role for men in ancient Greece and other cultures and homosexual acts have been both ritualized and prohibited in various societies, but the idea of identifying a type of person based on regular participation in homosexual behavior is generally believed to be a product of the medicalization of homosexuality during the nineteenth century. The study of the forms and meanings of homosexual behavior in different cultures is a rich field that has helped to provide information about the range of sexual practices and types of gender-based relationships transculturally and transhistorically. However, the forms of socialization and development of identity associated with homosexual behavior in contemporary society are generally recognized to be unique to this period of time in history and, until recently, to Euro-American culture. The history of the development of these contemporary forms of homosexual expression is the focus of the following review of theoretical perspectives about homosexuality.

History The modern view of homosexuality had its origins in the second half of the nineteenth century when a number of European sexologists, including several prominent psychiatrists, began to study homosexuality from a scientific perspective. Medical and psychiatric views of homosexuality passed through two major periods in the ensuing 100 years: the first from the second half of the nineteenth century through the first half of the twentieth century, marked by the gradual predominance of biological and psychological, largely psychoanalytic, theories of homosexuality over the earlier religious and moral conceptualizations; and the second, roughly from the time of the publication of the two Kinsey volumes on male and female sexuality in 1948 and 1953, respectively, until the removal of homosexuality from the list of mental disorders by the American Psychiatric Association (APA) in 1973, marked by a growing attempt to normalize homosexuality and to study the phenomena of sexual orientation from a wide range of social and scientific perspectives.

Medicalization of Homosexuality: From Morality to Pathology

LATE–NINETEENTH- AND EARLY–TWENTIETH-CENTURY BIOLOGICAL VIEWS During the second half of the nineteenth century researchers in Europe began to study homosexuality scientifically in an effort that reflected a shift away from the previous religious and moral view that homosexual acts were an expression of degeneracy. *Homosexual* was first coined by the Hungarian physician Károli Mária Kertbeny in 1869, and was later adopted by the influential German psychiatrist, Richard von Krafft-Ebing. The conceptual basis for homosexuality arose from the earlier ideological construction of two sexes, male and female, with dichotomous characteristics, and the attempt to find explanations for homosexuality derivative from these male and female categories. Thus, the homosexual was conceived of as a person of one biological sex with the sexual longing of the other sex, as portrayed, for example, in the German jurist and homosexual rights advocate Karl Heinrich Ulrich's description of the male homosexual as a female soul in a male body. Ulrich's belief that because homosexuality was biologically determined it was natural, and his reliance on the idea of separate male and female “love drives” in the brain that developed in *Urninge* (male homosexuals) and *Urninden* (female homosexuals) in opposite directions to their genital sex, continue to be reflected in contemporary efforts to locate an anatomical basis for homosexuality in brain sexual dimorphism.

The psychiatrists Carl Westphal and von Krafft-Ebing adopted many of Ulrich's ideas, but they considered the traits he described to be unnatural. The German physician and sexual emancipator, Magnus Hirschfeld, also believed that there was a genetic basis for homosexuality. Although he understood homosexuality to be a malformation in development and ascribed it to a number of different biological etiologies, he also asserted that it was natural. Similarly, the English sex reformer Havelock Ellis based his efforts to alter attitudes about homosexuality on the belief that homosexuality was an inborn trait. The sexual reforms advocated by Hirschfeld and others in Germany came to an end when the Nazis came to power in 1933. Other biological bases for homosexuality began to be examined during this period, including a variety of constitutional factors, such as different anatomical structures in the brains and nerve structures in the rectums of homosexual men, and endocrinological determinants, including atrophied and abnormal testicles in homosexual men, which led to experiments involving the removal of the testicles of homosexual men and replacement with what were believed to be the normal testicles of heterosexual men.

These early biological theories about the origin of homosexuality, with their emphasis on heredity, endocrinological influences, and anatomical or structural differences between homosexual and heterosexual persons, presage many of the current efforts to find a biological basis for homosexuality. They also served as the justification for subsequent eugenic efforts in Germany to eradicate homosexual persons. Furthermore, the almost exclusive emphasis on the study of homosexuality in men parallels later twentieth-century biological research that has largely ignored homosexuality in women. The historical belief in a biological basis for homosexuality, whether it was viewed as natural or unnatural, has had a major influence on scientific approaches to homosexuality, but from the beginning this belief has been challenged by those who assert that homosexuality results from environmental, psychological, and developmental causes. The biological arguments helped to create the notion of a homosexual person by emphasizing the existence of fundamental constitutional differences between those who are attracted to persons of the same sex and those who are attracted to persons of the other sex. This conflation of behavior and personhood can be seen as the progenitor for the later development of gay and lesbian identities.

PSYCHOANALYTIC APPROACHES Like the biological theorists of his time, Freud believed that heterosexuality was the natural outcome of normal development. However, unlike many of his historic contemporaries and some of his later followers, he did not view homosexuality as a sign of degeneracy and asserted that homosexuality could occur in persons who had no other signs of deviation and no impairment in functioning. For Freud homosexuality represented an arrest in development from an instinctual bisexuality to a mature heterosexuality that could arise as the result of a variety of factors; he presented both biological and environmental explanations for its development. Freud's position on homosexuality, neither completely based in biology or in environment, further influenced the movement away from the moral degeneracy and religious views that predominated prior to the nineteenth century.

Later psychoanalytic theorists, including Sandor Rado, who refuted Freud's model of innate bisexuality and developed his own theory of homosexuality as a type of phobia, heavily emphasized the beliefs that homosexuality is the result of problems in psychosexual development and that its occurrence is a sign of severe psychopathology. However regardless of the degree of pathology homosexuality was believed to represent, psychoanalysis largely derived its theories of homosexuality from a biological and medical model that relied on dichotomous categories of male or female, masculine or feminine, and heterosexual or homosexual. Various etiological theories were hypothesized, but they all were based on the same assumptions of gender, sexual, and sexual orientation polarities that shaped the research in the biological sciences during this period. Although some discussion of the etiology of female homosexuality was presented in the psychoanalytic literature, the primary emphasis in psychoanalysis, like in the biological sciences, was on male homosexuality; often the causes and types of homosexuality in women were simply treated as the opposite of those for male homosexuality.

During the latter part of the nineteenth and the first half of the twentieth century, psychoanalysis, with its emphasis on intrapsychic and interpersonal development, represented an alternative approach to the biological sciences to studying and understanding homosexuality; however, its fundamental assumptions were derived from the same materialistic beliefs upon which the biological sciences rested. Thus, the outcomes of the efflorescence of interest in studying homosexuality within the medical and the psychoanalytic fields during this period were similar: to establish a focus of interest about homosexuality outside of the realms of morality and religion within science and medicine; to reify a category of person, the homosexual, associated with homosexual behavior; and to confirm the abnormality and the pathology inherent in homosexuality. These developments set the stage for the movement toward the normalization of homosexuality that began in the middle part of the twentieth century.

Normalization of Homosexuality: From Diagnosis to Normality

INFLUENCE OF NEWER RESEARCH FINDINGS FROM SEXOLOGY, CROSS-CULTURAL, ANIMAL, AND PSYCHOLOGICAL STUDIES The publication of the two volumes on male and female sexuality by Kinsey and his colleagues in 1948 and 1953, respectively, marked the beginning of the move away from the view of homosexuality as pathology and toward an understanding of homosexuality as a normal variant of sexual desire and behavior. Although Kinsey's studies have subsequently been criticized for their methodology, at the time of their publication their demonstration of the widespread presence of homosexual feelings and behavior among samples of several thousand American men and women suggested that homosexuality was not an isolated and aberrant phenomenon. Kinsey's conceptualization of sexual orientation as a continuum also introduced the idea that homosexuality and heterosexuality were not dichotomous categories but rather occurred in varying degrees in individuals. Reflecting his nominalist view of the world, Kinsey stated that the world cannot "be divided into sheep and goats.... It is a fundamental of taxonomy that nature rarely deals with discrete categories. Only the human mind invents categories and tries to force facts into separated pigeon-holes. The living world is a continuum in each and every one of its aspects."

Cleland Ford and Frank Beach's work, published in 1951, demonstrated that homosexuality was common across cultures and occurred in almost all nonhuman primate species. These findings were interpreted to support the notion that homosexuality was both natural and widespread. The psychologist Evelyn Hooker's publication in 1957 of the results of her groundbreaking study comparing the projective test results from 30 homosexual men and 30 heterosexual men showed that experienced judges could not distinguish between the two groups based on their test results and thereby established for the first time the idea that homosexuality in men was not associated with mental illness. The effect of this significant finding was to require that future views of homosexuality as an illness could only be based on an a priori assumption that homosexuality was itself pathological. By the 1970s influential psychiatrists like Judd Marmor, a psychoanalyst and president of the APA, began to publish works that moved the psychiatric profession away from the perspective of homosexuality as a type of mental illness and toward a consideration of the perspectives offered in these newer research findings.

During this period of transition and in spite of the availability of contradictory findings from empirical studies, other psychologists, particularly behaviorists who attempted to extinguish homosexual expression through aversive techniques, and some psychiatrists and psychoanalysts, espoused a pathological basis of homosexuality and published works in the 1960s that presented homosexuality as the outcome of developmental disturbance and as a sign of severe psychopathology. Such work, often criticized for its methodology, was particularly influential in portraying a pathogenic family type, consisting of a detached and rejecting father and a close-binding and domineering mother, that led to homosexuality in adult men. Ignoring the fact that such family configurations occur for heterosexual men as well and relying on questionable methods for analyzing results, those workers strengthened the belief within psychoanalysis and psychiatry that homosexuality was a sign of severe psychopathology and intensified efforts to attempt to "cure" persons who were homosexual through reconstructive treatment to make them heterosexual. The rate of cure for conversion of men from homosexuality to heterosexuality was in fact quite small and subjects were never independently evaluated to determine if reported changes in behavior were maintained.

Pathologizing views of homosexuality within psychoanalysis and psychiatry reflected the negative attitudes about homosexuality in the larger society, which allowed an assumption of pathology in homosexual persons whether or not it had been demonstrated to exist. The impact of these works was to reinforce the societal association of homosexuality with mental illness and stigma and to increase the experience of guilt and self-hatred in homosexual persons and their families. The clash within psychiatry between these traditional views of homosexuality and the perspectives offered in the newer research and articulated by spokespersons for an increasingly vocal gay rights' movement led to the next phase of development in theory about homosexuality.

DECLASSIFICATION BY THE AMERICAN PSYCHIATRIC ASSOCIATION Homosexuality was classified as a type of mental disorder since the publication of the first edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I) in 1952, in which it was designated, along with the other conditions labeled as sexual deviations, as a type of sociopathic personality disturbance, defined by the existence of certain behaviors and not by the presence of distress or dysfunction. This classification of homosexuality as a form of mental illness was not controversial at the time DSM-I was published, reflecting as it did the negative societal attitudes about homosexuality and accepting the prevailing medical views of homosexuality as an illness as previously defined in the International Classification of Disease (ICD). In the second edition of DSM (DSM-II), published in 1968, homosexuality continued to be listed as a sexual deviation, but the sexual deviations were no longer categorized as types of sociopathic personality disturbance.

During the next several years the gay rights' movement became more vocal in its opposition to classifying homosexuality as a type of mental disorder, and increasing numbers of psychiatrists were responsive to the research findings that demonstrated its occurrence in large numbers of people, in persons who demonstrated normal adjustment, and in a variety of cultures. The psychiatric view of homosexuality as a disease was now seen as part of a larger historical movement to replace moral and religious judgments of homosexuality with supposedly scientific and objective but equally negative portrayals of homosexuality as a sickness.

Beginning in 1968 gay activists began to confront the APA about its position, and following a series of dramatic encounters between gay rights' advocates and psychiatrists at several annual meetings of the APA, the organization began officially to reevaluate its position of diagnosing homosexuality as a mental disorder. Following considerable political activity and significant scientific study of the issue, a proposal from the Nomenclature Committee of the APA to eliminate homosexuality from DSM was approved by the APA's Council on Research and Development, Reference Committee, and Assembly of District Branches before being accepted by the APA's Board of Trustees in December 1973. As a result, for the first time in modern history homosexuality per se was not classified as a type of illness. This action is viewed as having enormous social significance in shifting the predominant and official view of homosexuality as a sickness to viewing it as acceptable behavior. Psychiatrists who continued to adhere to the traditional views of homosexuality criticized the action taken by the Board of Trustees and challenged it in a referendum of the APA membership, but the decision was upheld by 58 percent majority of voting members.

Other major mental health professional organizations, including the American Psychological Association and the National Association of Social Workers, moved to endorse the action taken by the APA. The decision to declassify homosexuality was accompanied by passage of a position statement by the APA that supported the protection of the civil rights of homosexual persons:

> Whereas homosexuality in and of itself implies no impairment in judgment, stability, reliability, or vocational capabilities, therefore, be it resolved that the American Psychiatric Association deplores all public and private discrimination against homosexuals in such areas as employment, housing, public accomodations, and licensing, and declares that no burden of proof of such judgment, capacity, or reliability shall be placed on homosexuals greater than that imposed on any other persons. Further, the APA supports and urges the enactment of civil rights legislation at local, state, and federal levels that would insure homosexual citizens the same protections now guaranteed to others. Further, the APA supports and urges the repeal of all legislation making criminal offenses of sexual acts performed by consenting adults in private.

The rationale for removing homosexuality from DSM was based in part on the clarification, as formulated by Robert Spitzer, of the definition of a mental disorder as meeting the criteria of either experience of subjective distress or generalized impairment in social effectiveness or functioning as a result of the condition. Neither of these criteria applied to homosexual persons who were satisfied with their sexual orientation and who did not demonstrate impaired functioning. However, in what appeared to be a compromise between opposing views, a new diagnostic category, sexual orientation disturbance, was introduced to apply to those homosexual persons who "are either disturbed by, in conflict with, or wish to change their sexual orientation." When third edition of DSM (DSM-III) was published in 1980, sexual orientation disturbance was replaced with a new diagnostic category, "ego-dystonic homosexuality," a type of general psychosexual disorder that labeled homosexuality as a mental disorder only when it was accompanied by persistent distress. Finally, after further review of the appropriateness and utility of this diagnosis, all reference to homosexuality was removed from the revised third edition of DSM (DSM-III-R), published in 1986, representing the final acknowledgment by the American psychiatric profession that homosexuality by itself is not a type of mental illness. This change in the official view of homosexuality signified not only the acceptance of findings from newer research about homosexuality, but also a dramatic conceptual shift in the cultural meaning and significance of homosexual behavior within American society. This shift set the stage for the publication during the last three decades of the twentieth century of a profusion of new theory and research findings about homosexuality and gay and lesbian persons based on more complex and multidimensional conceptualizations of homosexuality as a normal variant of sexual desire and expression.

Postmodern Theories: Essentialist and Constructivist Arguments The declassification of homosexuality per se as a type of mental illness in 1973 and the elimination of any reference to homosexuality in DSM-III-R paralleled a significant cultural shift in approaches to epistemology and knowledge during the latter part of the twentieth century away from realist and materialist beliefs and toward a way of viewing the world that has been called postmodernism. *Postmodernism* is characterized by an attitude that challenges fixed beliefs and categories of knowledge, uses constructivist methods to unravel the particular historical and social influences that determine language and understanding, and seeks to identify the links between knowledge and social and political influences. Thus, at the same time that a specific paradigm shift in the conceptualization of homosexuality was occurring, signaled by its declassification by the APA as a type of mental illness, a more

fundamental alteration in philosophical and analytic approaches to understanding was taking place.

Postmodernism has had its strongest influence within the humanities and the social sciences whereas the biomedical field—focused as it is on biology and the body—has continued to rely on more traditional materialist approaches to studying and viewing the world. During the latter part of the twentieth century a bridge between traditional and postmodern methods for the study of homosexuality arose from the debate that took place between essentialist and constructivist (also known as social constructionist) theorists. This debate originated within the fields of philosophy, feminism, and identity politics and built upon the critiques of sex, gender, and racial categories developed by feminist and civil rights' thinkers. *Constructivist ideas*, first defined in the 1970s by the French philosopher Michel Foucault and others, maintained with respect to sexuality that there is no inner sexual drive, but rather that the human potential for thinking and acting is shaped by social forces of regulation and categorization into various types of sexual desire at different times in history and in different societies. Constructivist theory critiques classical theories of sexuality that posit inherent and largely immutable qualities that are associated with sex, gender, and sexual orientation. Thus, constructivist methods provide one way of questioning and examining contemporary notions of sexuality, including homosexuality.

In contrast, *essentialism* refers to the view that sexual desires and identities are fixed personal characteristics that are inherent, objective, transcultural, and transhistorical. Essentialism treats sexuality as a biological force that underlies genuine sexual differences, which in turn serve as the basis for sexual categories and identities. With respect to sexual orientation, essentialist theorists believe that fundamental differences in sexual desire are the basis for creating categories of persons known as homosexual, heterosexual, and bisexual.

The debate between essentialists and constructivists reflects the earlier debate within psychiatry regarding the importance of nature and nurture in creating mental illness and in determining the etiology of homosexuality. However, the scope of this postmodern debate encompasses much more than origins or etiology alone and entails challenges to other dimensions of understanding about homosexuality as well. Although these dimensions are often described in terms of oppositional positions representing essentialist or constructivist arguments, a continuum of beliefs about them is contained within both schools of thoughts.

Homosexuality: Innate Versus Constructed One polarity in the debate consists of the contrasting beliefs that homosexuality is innate versus that homosexuality is constructed. These two positions represent a fundamental distinction between essentialist and constructivist views respectively. The essentialist view that homosexuality is innate is represented most directly in attempts to locate a biological basis for homosexuality. Many of these efforts reflect a conceptualization identical to what was in fashion during the nineteenth century when researchers sought to identify a source for homosexuality in the brain and other parts of the body and during the early part of the twentieth century when endocrinologists first began to search for a possible hormonal basis for homosexuality. Corollary essentialist assumptions include the notion that types of persons exist who can be defined by their homoerotic desire and that these persons have shared fundamental characteristics throughout history and in different cultures.

In contrast, constructivist theorists argue that different forms of sexual desire and expression arise as the result of social forces that vary across time and cultures. Thus, while the individual may experience same-sex desire and even some sense of being lesbian or gay, these experiences are viewed as the product of larger social forces—like the market place, urbanization, and government regulation—that influence sexual desire and behavior. The implications of this belief include the idea that homosexual desire and gay and lesbian identity are variable and fluid in their origins and expression.

Homosexuality: Fixed Versus Mutable Another dimension of the debate between essentialists and constructivists involves beliefs about the extent to which homosexuality is fixed or mutable. There is considerable variability in the positions represented by essentialists and constructivists on this matter and any distinction between the two schools cannot be collapsed into disagreement about the possibility of changing sexual orientation. For example, persons believing that homosexuality should be changed through psychotherapy or other means are often essentialists who believe in the essential nature of heterosexuality; those who strongly endorse a constructivist conceptualization of sexuality may at the same time recognize that most individuals are unable to alter their sexual attraction. Nonetheless, a political position has arisen that argues that if sexual orientation is an essential characteristic of a person, it cannot be changed, and this position has been contrasted with constructivist ideas that are believed to imply the possibility for changing sexual orientation. However, neither of these theoretical beliefs necessarily suggests that sexual orientation can either be changed easily or be changed through psychotherapy or other means such as religious conversion. In this debate the fixed-mutable dimension of sexual orientation has often been conflated with two other issues: the political concern about attempts to coercively change a homosexual or bisexual orientation to a heterosexual orientation and the question of the degree of individual freedom in relation to choosing a particular sexual orientation. It is important to understand that discussion of the degree to which underlying sexual orientation is fixed or mutable can be separated from the notions of how change can be accomplished and to what extent individuals have choice in relation to their sexual desire, behavior, and identity.

The belief that homosexuality is a stable phenomenon that is relatively resistant to change derives from several sources, including the experiences of many people who are lesbian, gay, and bisexual concerning the difficulty and undesirability of changing their sexual attractions; the empirical finding that homosexuality rarely changes as the result of psychotherapy or other external attempts to alter it; the idea that sexual orientation is the outer representation of some constitutionally determined inner characteristic; the political stance that external agents should not attempt to alter sexual orientation; and the deterministic assumption that individuals have no choice about the direction of their sexual desire. In contrast, the idea that sexual orientation can change for some persons derives from a different set of observations about sexual orientation: the experiences of some persons who report fluctuations in sexual attractions across the course of their lives; the belief that sexuality is the product of social forces that vary over time and in different cultures; adherence to a belief in the need to protect individual freedom in relation to choosing sexual partners regardless of their sex; and the idea that the acceptance of a potential for change in an individual's sexual orientation can be separated from the concern about the use of coercive and harmful means for attempting such change. Thus, the apparent conundrum present in trying to determine the degree of change in sexual orientation that is possible can only be understood by appreciating the variable origins and values associated with such arguments and defining which of these beliefs is actually the topic of discourse.

Towards an Interactionist Model An important distinction between essentialist and constructivist approaches to homosexuality is the manner in which they conceptually organize the phenomena they address. Strict essentialists often assume that homosexuality is a unitary phenomenon comprised of similar experiences and behaviors that fundamentally represent the same characteristic. As a result, observations that are at variance with a unitary notion of homosexuality may not be viewed by strict essentialists as part of it. For example, people who experience fluctuations in their sexual desire for one or the other sex may be considered to be bisexual, and only persons who go through certain developmental stages or who demonstrate certain biological markers may be described as “true homosexuals.” As a result of demarcating such a conceptual boundary, other essentialist assumptions that homosexuality is innate, fixed, and uniform are reinforced. By excluding variations and focusing on means, a more discrete descriptive category can be created.

In contrast, constructivists tend to focus on variability and to see multiple phenomena within the construct of “homosexuality.” The constructivist methods of analyzing and deconstructing fixed categories and ideas result in the recognition of multiple layers and types of experiences associated with same-sex desire, behavior, and identity in different cultures and periods and which vary for some individuals across the lifespan. The result of this perspective may be that the constructivist sees more clearly the dehumanizing effects of narrow categorization and labeling but may also be speaking about something that is very different from what the essentialist is addressing.

If the differences between the two theoretical stances are understood, the focusing lens of the essentialist and the dispersing crystal of the constructivist can both be used to help view different aspects of what is called homosexuality in contemporary society and together can serve to drive, from different vantage points, new thinking and research that will deepen the understanding of human sexuality. The apparent contradictions between the essentialist and constructivist positions, like those in the earlier arguments about nature and nurture, can then be seen to reflect sometimes complementary aspects of the complex phenomena that are associated with sexual orientation. From this perspective, the biological, genetic, and structural matrix of the human body becomes shaped by volition and external influences into a person's sexual orientation.

FREQUENCY

Studies that purport to demonstrate the extent of homosexuality in a population have generally failed to distinguish between the various components of sexual orientation: desire, behavior, and identity. If an interview or survey question simply asks if a person is gay or lesbian or homosexual, respondents who only have sexual relations with other persons of the same sex but do not identify themselves as gay or lesbian may deny that they are gay or lesbian, and the stigma associated with being a member of a sexual minority group often interferes with full disclosure. Similarly, if persons are not open about their sexual orientation, they may also fail to respond fully or accurately to questions about their sexual behavior or identity. Even when questions about sexual orientation are constructed in a manner that takes into account such problems of definition and disclosure, it may be difficult to obtain a representative sampling of individuals with different sexual orientations. Studies have shown that self-reports of sexual orientation vary significantly across age groups, for example, from early to late teenage years, and in different geographical locations. These variations are compounded when frequency rates are compared cross-culturally because of the variable meanings of sexual behaviors

and practices in different cultures.

Given these limitations in the existing studies of the frequency of homosexuality, it is impossible to determine definitively the number of persons who have same-sex desire, behavior, or identity in the United States today. The 1948 Kinsey study of male sexuality, based on a sample of men between the ages of 16 and 55 years, found that 4 percent of the men were exclusively homosexual, 50 percent were exclusively heterosexual, and 46 percent were somewhere in between. The Kinsey study has subsequently been criticized for its sampling errors, which are believed to have led to inflated estimates of the extent of homosexual and bisexual behavior. Later studies of the frequency of same-sex sexual contacts between men, utilizing results from national surveys conducted in 1970 and 1988, determined that 20 percent of men had sexual contact to orgasm with another man at some time in their lives, and that 5 to 7 percent of men had such behavior in adulthood, but that only one quarter to one half of these men also reported having such contact in the preceding 12 months. Because of problems in reporting and with missing data, these figures were considered to be the lower bounds of actual frequency of same-sex behavior among men. Although reports about the frequency of homosexuality among women in the United States have been almost absent since the publication of Kinsey's 1953 volume on sexuality in women, the number of women reporting same-sex sexual contact or identifying themselves as lesbian is generally considered to be considerably lower than for men, often one half or less than the number of men.

A large scale study of sexuality in the United States published in 1994 by Edward Laumann and his colleagues provides the most reliable current estimates of the prevalence of homosexuality in the United States. This study utilized more precise definitions of homosexuality and heterosexuality, inquiring about individual desire, behavior, and identity, and obtained a carefully stratified representative sampling of the general population. However, because the total number of persons reporting at least one of the components of homosexuality was only 293 in a total sample of 3159, 150 of 1749 women and 143 of 1410 men, the number of persons reporting same-sex desire, behavior, and identity in specific age groups, in different geographical locations, and in other demographic categories was rather small. This study showed that, of the total sample 1.3 percent of the women and 2.7 percent of the men participated in same-sex sexual behavior during the preceding year, and 4.1 percent of women and 4.9 percent of men had done so since age 18; 7.5 percent of women and 7.7 percent of men reported the presence of sexual desire for someone of the same sex; and 1.4 percent of the women and 2.8 percent of the men reported a homosexual or bisexual identity. These figures varied considerably across groups based on age, marital status, education, religion, race, and place of residence.

The frequency figures from these studies of the extent of homosexuality in the United States during the second half of the twentieth century demonstrate that the figure of 1 in 10, or 10 percent often used to describe the number of persons who are gay and lesbian seems to be an overestimate of the actual number of persons who report same-sex desire or behavior or who acknowledge a gay or lesbian identity. However, it is also clear from the results of these studies that a larger number of persons report the existence of same-sex desire and that in certain population groups, for example, individuals living in the twelve largest U.S. cities, the number of gay and lesbian persons does approach the 10 percent figure. The clinician must understand the variable prevalence depending on which component of homosexuality and which social or demographic groups are being considered. Studies of prevalence of homosexuality in other Western countries show similar trends but estimates of the extent of homosexuality in a given country must take into consideration such factors as the degree of repression of homosexuality and the cultural meaning of same-sex desire, behavior, and identity.

ORIGINS OF HOMOSEXUALITY

Prior to the declassification of homosexuality as a form of mental disorder scientific studies into its origins consisted primarily of attempts to understand the nature of biological, psychological, or environmental causation of a pathological condition. Since the declassification of homosexuality in 1973, studies of its origin have been free to examine broader aspects of homosexuality and bisexuality and have largely been conducted with the assumption that homosexuality is a normal variation of human desire and behavior. Four areas in which the origins of homosexuality have been focused are biology, cross-cultural studies, psychological development, and acquisition of gay and lesbian identities.

Biological Studies Studies of the biological origins of homosexuality conducted since the 1960s have been based on the essentialist assumption that the cause of homosexuality, or at least of some of its forms, can be discovered in genetic, endocrinological, or anatomical differences between homosexual and heterosexual persons. Findings from studies into the biology of homosexuality have claimed the attention of the biomedical sciences and much of the general public and have continued to focus on the determination of a single cause or basis for homosexuality.

Genetic studies of homosexuality have been reported within the field of molecular biology through genetic linkage studies that attempt to locate a “gay gene” and within family studies of the twins and siblings of gay and lesbian individuals. One genetic study reported evidence for genetic linkage between alleles on the X chromosome in pairs of gay brothers. A number of studies have reported increased concordance for homosexuality among sets of monozygotic or identical twins and increased presence of homosexuality in the siblings of gay and lesbian subjects. The findings from such family studies vary in the degree of concordance they demonstrate and in the extent to which they reflect environmental or genetic influences. Explanations for discordant phenotypes in genotypically identical individuals have not been accounted for in these studies but suggest the influence of nongenetic factors in determining sexual orientation.

Early studies of a hormonal basis for homosexuality hypothesized a decreased level of testosterone in homosexual men and an increased level in homosexual women. However, there is currently no evidence for hormonal differences between adult homosexual and heterosexual men and women. More recent studies have focused on prenatal hormonal influences. This research has relied on studies in animals involving diminished prenatal exposure to androgens, the so-called male hormones, and comparison of the degree of sex-typed behaviors, primarily mounting and lordosis, in demasculinized rats with normally developing rats. These studies have not reported consistent findings and have been criticized for assuming an animal analog for human sexual orientation that compares demasculinized rodents with homosexual men who have normal testicles and hormones and that equates sex-typed behaviors in animals with sexual orientation in humans.

The final area of current biological studies of homosexuality is of brain function and structure. Based on the arguable assumption that the human brain is sexually dimorphic, these studies attempt to demonstrate that the brains of lesbians and gay men are in some fashion intersex, or intermediary in function or structure, between heterosexual men and women. This research has attempted to find evidence for differences between the brains of homosexuals and heterosexuals based on neuroendocrine function, cognition and lateralization, and brain structure, including in the corpus callosum and in the third interstitial nucleus and the suprachiasmatic nucleus of the hypothalamus. The results of these studies have not been replicated, and none of the criticisms of the studies has been adequately addressed through the design of new studies that correct the theoretical and technical problems that have been identified.

Recent efforts to locate a biological basis for homosexuality have narrowed their focus from the expansive arena of investigation in earlier generations of such research, which hypothesized a wide range of constitutional differences between homosexuals and heterosexuals. Current biological research often continues to rely on nineteenth-century assumptions that conflate sex and gender differences with differences in sexual orientation, that utilize the study of animal models and behavior to explain more complex human phenomena, and that visualize a structural-functional equation between anatomy and behavior. Biological studies have also have been plagued by ongoing problems with definition and methodology, including lack of precise definitions in describing subjects, use of small sample sizes, failure to replicate study results, and generalizations from findings about specific samples to the general population. Virtually all the biological studies have focused on men and results from the few studies of lesbians demonstrate even fewer differences from heterosexual women than for men. Thus far the results of this research at best suggest that some aspects of homosexuality in some individuals may be linked to genetic and familial origins. Carefully designed future research may further clarify the nature of biological contributions to the determination of sexual orientation.

Cross-Cultural Studies Anthropological studies of human sexuality have demonstrated the enormous variation in gender and sexual organization and behavior in different cultures, thereby providing further evidence for the fluidity and complexity of sexual behavior and experience at individual and social levels. The findings from ethnographic studies of gender and sexuality in different cultural groups have helped to dislocate the cultural bias underlying many approaches to examining these phenomena that assumes developmental continuity in individuals and cross-cultural similarities in the definition and expression of sexuality. Building upon the pioneering work of Margaret Mead, Ruth Benedict, and other anthropologists during the first half of the twentieth century in identifying variations in sexual patterns in different societies, more recent researchers have been able to classify structural types of sexual development and forms of homosexual organization in different cultures. Gilbert Herdt in particular has explicated these cross-cultural models based on his study of ritualized homosexuality among the Sambia people in New Guinea and on his review of findings from other anthropological studies of sexual practices. Herdt has incorporated the idea of “discontinuity” in sexual development to explain variation in the psychological experience and symbolic meaning of sexuality, in contrast to the continuity in sexual development across the life span that is assumed in most traditional approaches to understanding sexual development and that underlies important aspects of biological development. He has also explored the cultural aspects of fluidity in experience and expression of sexual orientation that has defined much of the theoretical discourse on bisexuality.

Herdt has identified three types of sexual development—linear, sequential, and emergent—that occur cross-culturally and help to explain variations in patterns of sexual orientation. *Linear development* refers to sexual behavior that occurs without significant change in behavior or orientation across a lifetime. *Sequential development* describes a developmental pattern that incorporates important change at different life stages, for example, in Victorian England, and to some extent modern America, where children were assumed to be sexually naïve but adults were expected to have a fully developed range of sexual practices, and among the

Sambia, who practice ritualized homosexual behavior during childhood and universal heterosexuality as adults. *Emergent development* occurs in societies that allow some degree of ambiguity and uncertainty in adult sexual outcomes based on childhood socialization. It can be identified in societies that have undergone massive social change that has affected sexual norms and values and in segments of modern American society that allow a range of sexual options for adults. Each of these patterns of sexual development can be shown to significantly influence the forms and expression of sexual orientation that exist in different cultures.

Herdt has also described a typology of homosexuality based on the organization of same-sex sexual behavior cross-culturally. The first type is *age-structured homosexuality*, usually involving older and younger males and generally including a sequential pattern of childhood same-sex practices that yield to primarily heterosexual behaviors in adults. Another form of homosexuality is *gender-reversed homosexuality*, which consists of a reversal in the normative sex-role dress and behavior. One example of such gender-reversed homosexuality is the American Indian *berdache*. The third type identified by Herdt is *role-specialized homosexuality*, in which same-sex sexual activity is restricted to certain social roles and positions. This type occurs among some American Indian tribes that identify their shamans, who then cross-dress and engage in homosexual behavior, following dreams or visions that reveal their special status, and among certain working-class women in nineteenth-century China. Finally, Herdt describes the *modern gay movement*, consisting of egalitarian same-sex practices, as the fourth type of homosexual organization. While acknowledging that hierarchical relationships still exist in parallel with traditional forms of heterosexual relationships, he states that homosexuality in contemporary America has arisen as a new form of homosexual practice that comprises a sexual orientation, a social identity, and a political movement.

Findings from the cross-cultural study of sexuality and homosexuality highlight the limitations of any narrow explanation for the development and experience of homosexuality in a particular society. Variations in the organization and meaning of same-sex practices are enormous in their range across cultures, and changes within a given society over time have been amply demonstrated. The universality of the existence of same-sex sexual expression has been shown to be matched by the constancy of change and variation in its organization and meaning. Future cross-cultural studies can help to further explicate the social factors, such as industrialization, urbanization, religion, and class stratification, that influence the structural forms into which individual sexual desire and behavior are shaped within cultures. By challenging the culture-bound assumptions that are brought to scientific conceptualizations of homosexuality, cross-cultural approaches encourage a broader understanding of sexual development and behavior.

Psychosocial Studies of Development Classical theories of psychological development are based on linear pathways of development that progress sequentially from childhood to adulthood and posit the origins of adult personality and identity in childhood experience. Only in more recent years have the distinct processes of adult development been studied. An understanding of the psychological development of sexual orientation must be based on an examination of psychological and interpersonal events throughout the life cycle and must incorporate assumptions of both continuous and discontinuous patterns of development. Furthermore, when applied to sexuality and sexual orientation, models of development must be sensitive to the physical, cognitive, and emotional maturation that occurs across the lifespan and must not assume that childhood behaviors, awareness, and feelings are isomorphic with adult experiences. In addition, until recent years psychological studies of homosexuality have been based on the examination of very specific and unrepresentative populations, such as psychiatric patients, incarcerated individuals, and persons in the military. Most psychosocial studies continue to be plagued by problems of sampling and, although they may no longer be limited to these special populations, have tended to focus more on white, urban, openly identified, and community-affiliated gay men and lesbians, whose development and experience may not adequately reflect other groups.

The findings from psychosocial studies may also vary depending on what research methodology is followed. For example, findings from retrospective studies may demonstrate greater consistency with popular models of development as a result of social learning about those models that leads to their internalization as part of the life scripts of individuals. Studies of purported childhood precursors of adult homosexuality may focus too narrowly on selected characteristics, such as effeminacy in boys, that are associated with stereotyped notions of adult homosexuality, but ignore the enormous individual variability in patterns of development toward and characteristics associated with adult sexual identity.

Only recently has psychological research begun to examine potential childhood and adolescent antecedents of adult homosexuality. Even now, empirical studies of homoeroticism in American society before puberty are nonexistent, and only a few studies on same-sex desire during adolescence have been published. Inhibitions on acknowledging sexual attractions and behaviors, particularly same-sex desire and behavior, during childhood have fettered scientific inquiry and led to a profusion of models of development derived from the retrospective reconstruction of childhood experience. Virtually the only perspective that has been available to examine the psychosocial origins of homosexuality has been that of the clinician, whose vision is framed by a focus on dysfunction, pathology, and abnormality. As a result, children who are different in their gendered behavior and adults who are disturbed have continued to serve until recently as the primary source for the depiction of the psychological development of homosexuality. The potential psychosocial understanding that might derive from ethnographic and other social science approaches to studying sexuality in children has been applied only in other cultures and has thus not been utilized as a tool for the study of the development of sexual orientation in American society.

Current knowledge about the psychosocial origins of homosexuality has been acquired through studies of children who demonstrate early gender-role nonconformity, reconstruction of childhood histories within psychotherapy, and a small number of empirical reports about sexuality and homosexuality in adolescence. The origins of adult gay, lesbian, and bisexual identities in Western society have been described in theoretical models of identity acquisition and in the findings from a limited number of empirical studies of the specific behaviors and events associated with establishing these identities.

Origins of Homosexuality in Childhood and Adolescence Early attempts to describe the childhood psychosocial antecedents of adult homosexuality consisted primarily of psychoanalytic formulations derived from the histories of psychiatric patients that depicted trauma and disturbed family relationships during childhood as causal mechanisms. The conceptual and empirical inadequacies of these portrayals for explaining the multiple forms of same-sex desire and behavior in both psychologically healthy and disturbed individuals led gradually to the investigation of alternative explanations for the development of a variety of forms of normal sexual expression and identity. Current explorations into homosexuality in childhood continue to be guided primarily by essentialist views of sexual orientation, which assume that a particular personality characteristic or specific behavior is a reliable predictor of adult sexual orientation. It is unlikely, however, that a single set of characteristics or a unitary pathway leading to adult homosexuality will be identified, given the varied and fluid experiences associated with same-sex desire, behavior, and identity in American society.

The most common characteristics examined in studies of childhood factors associated with adult homosexuality are early gender nonconformity and the recollection of an awareness of being different. The appearance of early gender nonconformity—femininity in boys and masculinity in girls—appears to be correlated with the development of homosexuality in adulthood for some individuals. Similarly, many adult gay men and lesbians report a recollection of an awareness of “feeling different” during childhood. These characteristics seem to be more common among children who will become homosexual than in those who will become heterosexual.

Findings from large-scale surveys of sexual attraction during adolescence show that younger adolescents tend to be less aware of same-sex feelings and to identify themselves less often as lesbian, gay, or bisexual than older adolescents. Reports of same-sex attraction in teenage girls range from 7.5 to 18 percent and in boys from 7.7 to 21 percent, with an estimated 15 percent of youth overall experiencing some same-sex attraction. Fewer youth report same-sex sexual activity, and even fewer identify themselves as gay, lesbian, or bisexual, although this latter figure increases with age. Although exact figures are unavailable for comparison, more youth currently identify themselves as bisexual or uncertain about their sexual orientation than in previous generations as a reflection of the greater awareness of flexibility in sexual expression and of the diminishing social stigma associated with bisexual identity in contrast to lesbian and gay identities.

The study of homoeroticism and of the development of homosexual or bisexual identities in adolescence continues to be difficult because of the existence of strong peer pressures to conform to heterosexual norms, ongoing and significant societal stigmatization of same-sex desire and behavior, and institutional and familial repression and denial of the existence of same-sex sexual feelings and behaviors in youth. Recent studies have focused on the adjustment, health, and mental health of young people who self-identify as gay, lesbian, or bisexual and who present to community centers that provide services for these individuals. As a result of access to and concentration on this segment of the youth population, more may be known at this point about the problems experienced by youth with an emerging or solidified gay, lesbian, or bisexual identity than about the normative events associated with development.

Acquisition of Gay and Lesbian Identities Much of the contemporary discussion about the psychosocial development of homosexuality has taken place within the field of identity development, where it has flourished as a primary theme in the broader discourse of social constructionism and its analysis of the political and social meaning of identity in late–twentieth-century postindustrial Western society. Thus, it is important to understand not only the theoretical description of gay, lesbian, and bisexual identity development in an individual, but also to appreciate the increased importance of personal identities at a cultural level. From this perspective, the definition of lesbian, gay, and bisexual identities and the detailed description of their emergence and multifaceted expression within both society and the individual can be seen as an outcome both of personal growth and development for the individual and of the cultural evolution of the concept of the “homosexual person.” The appearance of fully developed gay, lesbian, and bisexual identities in individual men and women can thus be seen to be as much a reflection of the creation and acceptance of the late–nineteenth-century biomedical belief that sexual desire and behavior are core characteristics that define selfhood as an endpoint or process of

individual growth and maturation.

Viewing gay, lesbian, and bisexual identities as both personal narratives and cultural constructs should not detract from the recognition of their centrality in organizing the subjective experiences of individual women and men. Rather, the psychiatrist who works with persons who talk about their experiences of becoming gay or lesbian should view these personal stories as framed within the contemporary language that most accurately describes their experience. The language used to encapsulate these subjective experiences is helpful when understood and applied within the clinical setting, but models of development and categories of sexual orientation must not be used in a manner that objectifies or simplifies the enormous range of individual experiences associated with sexuality.

Although theoretical descriptions of lesbian, gay, and bisexual identity acquisition are inconsistent and sometimes fail to distinguish between such concepts as identity, self, and self-concept, the experience of these identities by lesbians, gay men, and bisexuals is widespread and subjectively meaningful. For this reason it is important for the psychiatrist to be familiar with the principles of identity development relevant for these men and women. Models of lesbian, gay, and bisexual identity acquisition describe an intrapsychic and interpersonal process whereby a subjective experience of oneself as lesbian, gay, or bisexual becomes increasingly congruent with the perception by others that one is lesbian, gay, or bisexual. This process is generally conceived of as occurring over time, in sequential stages, and in a manner that leads to an increasing integration of one's experience of being lesbian, gay, or bisexual into other aspects of one's sense of self. Gay, lesbian, and bisexual identity development has generally been conceived as consisting of a series of stages involving a period of growing awareness and confusion about the presence and meaning of same-sex erotic feelings; followed by a time during which these feelings are assumed to have relevance to the person's life; and ending with an extended phase of growing acceptance of and pride in the new identity.

The process of identity development leads to different outcomes for different individuals, is influenced by interpersonal effects and consequences and feedback from others, and may be interrupted or terminated at any point. Most models of gay, lesbian, and bisexual identity development incorporate the completion of several events or tasks: coming out, or becoming aware of same-sex attraction and disclosing it at some point to others; forming relationships with other gay and lesbian persons; engaging in same-sex sexual behaviors; becoming involved in gay and lesbian activities and communities; establishing primary relationships with persons of the same sex: and learning how to integrate one's sexual identity into other aspects of the self.

Models of gay and lesbian identity acquisition present an image of progressive elaboration and synthesis of sexual orientation with other aspects of the self, but the lives of individual men and women vary considerably in the degree to which they follow such models. Development may be truncated for some persons in response to a number of factors, including a particularly punitive or harsh internal or environmental response to aspects of their new identity and a redirection of focus on other developmental tasks, such as schooling or career development. Because of a greater potential fluidity in their sexual orientation, women in particular may not follow any ideal of lesbian identity development, tending rather to emphasize the nature of the relationship they are in regardless of the sex of their partner. Similarly, because models of gay, lesbian, or bisexual identities have been developed based on the experiences of urban white persons, individuals who live in situations that do not provide reinforcement for adopting these identities (e.g., individuals in rural areas and people from some communities of color) may participate in same-sex sexual activities without ever acquiring an identity associated with these behaviors.

Individual variations in the extent of development of gay, lesbian, and bisexual identity may result both from repressive factors that interfere with optimal personal growth and from normative psychosocial influences reflecting diverse individual experiences and different adaptational needs. The psychiatrist needs to appreciate the relevance of such identity development for each patient, to understand their particular psychological and social resources and needs in this area, and to appreciate the costs and benefits of different pathways of development. Ultimately it is the individual's responsibility to develop an identity; the clinician can only help to optimize the understanding of the meaning of various options and to establish opportunities for maximizing freedom in making choices about often-conflicting outcomes.

LIFE-CYCLE AND DEVELOPMENTAL ISSUES

In most respects the lives of lesbians and gay men are very similar to the lives of heterosexual persons; however, because of the experience of difference and the negative effects of antihomosexual attitudes, lesbians, bisexual persons, and gay men are presented with unique developmental tasks throughout the life cycle. For example, during childhood and adolescence, youth who feel they are different and who become aware of significant attractions to someone of the same sex will need to understand their difference and often begin to disclose their homosexuality to others through the process of coming out. In adulthood these men and women will confront the unique challenges associated with establishing close relationships and creating families in a society that fails to provide rituals for celebrating and laws for protecting their families. All men and women are faced at different ages with important tasks such as establishing a sense of trust in themselves and others, forming healthy personal identities, creating and maintaining intimate relationships, learning how to lead productive lives, and sustaining a sense of generativity and personal integrity in old age. Success in completing these tasks is dependent on many factors, including biological and environmental resources, the impact of larger social and historical events, personal opportunities and traumas, and individual personality characteristics that interact at different developmental stages to shape a unique life for each person. In navigating the life cycle, gay men and lesbians and bisexual men and women will also need to contend with the negative factors resulting from the stigmatization of their sexuality as well as the potentially positive aspects of being part of a minority community.

Antihomosexual Attitudes and Heterosexism The term *homophobia* was coined by George Weinberg in 1972 to refer to fear or hatred of homosexuality and of gay and lesbian persons. This term has had widespread popular currency and has been applied to a variety of antihomosexual attitudes, but it has also been criticized for implying that the occurrence of these attitudes in an individual suggests the presence of a formal psychiatric diagnosis. Although extensive research on homophobic attitudes has demonstrated that they are correlated with certain demographic and personality characteristics, including sex, area of residence, religion, and degree of authoritarianism, the presence of these attitudes does not represent a type of mental illness; in fact, they are considered normative among some groups and in many social and institutional settings, including many schools, the military, and some religious organizations. Currently, the broader phrase, *antigay and antilesbian attitudes*, is often used to refer to the wide range of negative beliefs and feelings about homosexuality and gay men and lesbians.

Heterosexism is a related term that is defined as “an ideological system that denies, denigrates, and stigmatizes any nonheterosexual form of behavior, identity, relationship, or community.” Heterosexism is similar to other forms of ideological oppression, like racism and sexism, and occurs at both individual or psychological and societal or institutional levels. Antihomosexual attitudes may be expressed by individuals through verbal abuse, prohibitions, and physical violence directed toward gay, lesbian, and bisexual persons. Institutions promulgate such attitudes by excluding gay men and lesbians and by maintaining discriminatory policies and practices. Antihomosexual attitudes are an expression of heterosexist beliefs and can serve to enforce the maintenance of a “compulsory heterosexuality” in society. Heterosexism and antihomosexual attitudes may not be viewed as negative in a society or among groups that devalue and wish to eliminate erotic feelings and behaviors between persons of the same sex. Indeed, these systems serve to restrict awareness and expression of such feelings. For example, many religious institutions, military groups, and school systems actively prohibit homosexuality and promote negative attitudes toward nonheterosexual forms of sexual and affectional relationships. Outdated psychiatric and psychological views of homosexuality as a form of mental illness and religious teachings that homosexuality and homosexual behavior are immoral may be used to reinforce such negative attitudes.

All lesbian, gay, and bisexual persons will to some extent contend with what has been called *internalized homophobia*, which occurs when the developing gay or lesbian person psychologically incorporates negative societal views of homosexuality and then experiences these feelings and beliefs in the form of a negative self-evaluation. Although the effects of internalized homophobia on the development of sexual orientation differ enormously at an individual level, these negative feelings can lead to the suppression of awareness of same-sex attraction, an interference with the acceptance of one's homosexuality, and the truncating of a healthy integration of sexual orientation into other aspects of identity. Many gay men and lesbians experience internalized homophobia as an oppressive sense of devaluation and limitation in relation to their homosexuality, and for some it can produce a variety of psychiatric and behavioral symptoms, including depression, anxiety, denial, and even suicide. Internalized homophobia interacts with individual life histories to produce a unique expression in each gay and lesbian person, but its common roots lie in experiences such as criticism and shaming because of gender-nonconforming behavior in childhood in adolescence, being systematically excluded from institutions like the military, and being discriminated against by prohibition of same-sex marriage and denial of equal health and other benefits to same-sex partners.

Antigay and antilesbian attitudes, heterosexism, and internalized homophobia serve to define important aspects of development and daily experience in the lives of lesbians and gay men. For some persons these negative forces impinge upon the successful completion of normative developmental tasks and may delay or prevent successful adaptation and achievement of a healthy sense of self. The psychiatrist can be helpful by recognizing the influence of these factors on the personalities and experiences of gay, lesbian, and bisexual persons and carefully differentiating the intrapsychic and interpersonal wounds that result from stigmatization, discrimination, and victimization associated with being lesbian, gay, or bisexual from other aspects of the self. The psychiatrist's tendency to perceive problems that arise because of a need to deny being gay or in response to negative societal reactions to homosexuality as problems with sexual orientation itself has historically put gay men and lesbians at risk for further injury when seeking psychiatric help.

Coming Out and Meaning of Disclosure *Coming out*, defined previously as the process of recognizing one's homosexuality and disclosing it to oneself and others,

is a concept that is essential for understanding the lives of lesbians, gay men, and bisexual persons. It is particularly useful because of its experiential relevance for these persons and because of its widespread application in the behavioral and social science as well as in the popular literature. Coming out is not synonymous with identity development. One way of understanding the difference between these two concepts is to view coming out as one part of the process whereby a lesbian or gay identity is formed as the outcome. Behaviors associated with coming out reflect underlying cognitive and affective transformations that occur as part of identity development. Thus, coming out to oneself—becoming aware that one is attracted to persons of the same sex—facilitates disclosure of these feelings to others, which in turn can lead to social reinforcement and support. Some gay men and lesbians may come out in certain spheres of their lives, for example, within friendship groups, but remain closeted in other areas, such as in the work setting or within families.

The extensive literature on coming out generally describes it at the most basic level as the state of being clearly and openly gay or lesbian, in contrast to remaining *in the closet* (hinding one's homosexuality from oneself or from others). Early gay liberation publications portrayed the political distinction between coming out and being closeted as a struggle between liberation and oppression. Subsequently, coming out has also been defined as a psychological and social process that can help the gay person to overcome self-hatred and other negative effects of internalized homophobia.

Coming out is an ongoing process because gay and lesbian persons must repeatedly inform others of their identity. In spite of the belief that stereotyped personalities and appearances are associated with homosexuality, gay men and lesbians cannot be identified without some intentional or unintentional disclosure of their identity. Persons are generally assumed to be heterosexual unless they declare themselves to be otherwise. Adoption of certain styles of dress or behavior, residing in areas where other gay men and lesbians live, participation in gay and lesbian communities, and being involved in relationships with persons of the same sex may serve to confirm the identity of the gay or lesbian person, but these acts are themselves signifiers of coming out that must be maintained in order to inform others that one is gay or lesbian. Thus, disclosure of one's homosexuality involves either telling other persons that one is gay or lesbian or showing them through the display of some commonly accepted signal of sexual orientation identity. Gender ambiguity resulting from deliberately playing with gender and sexual portrayals by many young people in a postmodern culture may also lead to confusion about or inaccurate assessment of the sexual orientation of these individuals on the part of persons who rely on traditional signifiers of gender.

Most empirical studies of coming out have described the ages at which gay men have reported events associated with coming out; fewer studies have reported on the process for lesbians. However, studies show that women tend in general to have their first sexual experiences and to come out later than men. Younger lesbians may experience stronger prohibitions on coming out because of intense expectations of heterosexual dating for women. More recent studies have also suggested that coming out, much like the age of the first sexual experience, is occurring at progressively earlier ages in adolescents and that there is significant variation in the process based on a variety of factors other than sex, such as socioeconomic status, education, and ethnicity. The study of various cultures also demonstrates variation in the coming-out process based on the degree of sexual restrictiveness of a society.

Coming out is a complicated process that involves both internal and external spheres, occurs over a lifetime, and may vary in different domains of a person's life. The clinician must be aware that even persons who appear to be gay or lesbian may have never verbally disclosed their sexual orientation, and that others who provide no stereotypical indications of being gay or lesbian may have widely disclosed their identity to family, friends, and co-workers. Some persons who are fully aware of their homosexuality may never disclose this fact to others, and the range of time that elapses between awareness of same-sex attraction and self-labeling as gay or lesbian may be from less than a year to decades. Each individual's coming-out process must be understood and appreciated as unique and multidimensional.

Effects of Diversity Homosexuality and its various forms of expression develop in relation to other aspects of the self, which are influenced by a variety of individual and group characteristics. Lesbians and gay men are members of the same subgroups of diverse individuals that exist in the general population, and their experiences are significantly shaped by membership in these groups, which are defined by such characteristics as sex, racial and ethnic background, age, and class. Thus, understanding the development and experiences of lesbians, gay men, and bisexual persons often requires appreciation for their multiple affiliations and identifications, including but not limited to those defined by sexual orientation. Membership in many of these groups may involve further experience of discrimination resulting both from the prejudice directed toward people because of their race, sex, religion, age, or other characteristics, and from particularly intense antihomosexual attitudes that exist within some of these populations, such as certain conservative religious and ethnic groups. For some lesbians and gay men, membership in such groups may be a far more important determinant of their identity than their sexual orientation, and it is helpful for the psychiatrist to understand the variations in meaning of sexual orientation within these groups. For example, an older woman of color who has had a long-term relationship with another woman generally has a completely different construction of identity than that of a young gay man.

Sex and Gender Undoubtedly the most important group characteristic affecting the expression of sexual orientation is biological sex; men and women have very different experiences in relation to homosexuality because sex and gender are profound determinants of their identity. *Sexism* the domination and exploitation of women by men, can operate in conjunction with homosexuality in parallel to its function in defining certain aspects of heterosexuality. Definitions of gender are used to categorize people based on their sex, whether they are gay or lesbian, bisexual, or heterosexual. Therefore, the belief that lesbians and gay men necessarily share common qualities and interests should be replaced with a recognition that lesbian psychology and development may relate far more fundamentally to the psychology of women than to the psychology of gay men. In addition, the existing body of research and theory about homosexuality has largely focused on homosexuality in men and may have much less relevance for lesbian women.

Growing up as a boy or girl influences the experience of sexuality and sexual orientation in two general ways: through the development of different gender characteristics associated with men and women and through the imposition of different expectations for men and women, which often take the form of limitation and even discrimination based on sex. Gender socialization of boys and girls is one of the most profound forces influencing the quality of social interactions, and as a result, men and women generally behave in different ways. Although a person's biological sex is ultimately not the determining factor in the development of gender characteristics, sex and gender are inextricably linked in our society at the experiential level. Consequently, a variety of features of same-sex relationships will be influenced by the gender characteristics of the partners. For example, men often are more sexually active and competitive, and resolve conflict differently than women, and women as a group may be more emotionally expressive and relationally oriented than men. Gay men and lesbians may also be less likely than heterosexual persons to conform to traditional gender stereotypes and therefore may be more variant in their manifestations of gender-related characteristics.

Associations of sex and gender with homosexuality have historically led to inaccurate views of gay men and lesbians that portray them as persons who have the gender characteristics of the other biological sex, or who form relationships based on the assumption of complementary gender roles by two men or two women. In fact, the range of gender attributes of gay men and lesbians is as large and their relationships are as varied as those of heterosexual men and women. The psychiatrist working with gay men, lesbians, and bisexuals should set aside any preconceived notions about their gender-role characteristics; appreciate the significant role of gender socialization in shaping the personalities of all men and women and, as a result, the effect of this socialization on gay men and lesbians and on the dynamics of same-sex relationships; and assess the specific gender-related issues that might be present for a particular individual or couple. Such an approach recognizes the important influence of sex and gender in the lives of lesbians, gay men, and bisexuals, as well as heterosexuals, and avoids introducing inaccurate and denigrating stereotypes into the clinical setting.

Age and Generation The age at which a gay man, lesbian, or bisexual came out, the era during which they grew up, and the length of time since they came out will significantly affect their experiences of being gay, lesbian, or bisexual. Comparing a 50-year-old man or woman who came out at age 20 with someone of the same age who came out at age 45 or with a young person coming out today will generally demonstrate vastly different perceptions of reactions by others to their homosexuality. Generally, older lesbians and gay men were confronted with a much more negative set of social responses to their homosexuality than young people are today, and people who remain closeted during their adolescence and much of their adulthood will generally feel a greater sense of deprivation and loss in relation to their homosexuality than people who come out and establish a gay or lesbian identity earlier.

The differences between people of similar ages who have come out relatively earlier and later is a function not only of the discrepancy in social reactions to homosexuality during different historical periods but also of the interaction between general developmental maturation and ages of coming out and acquiring a gay or lesbian identity. For example, the impact of coming out on someone who establishes an openly gay or lesbian identity before being involved in an intimate primary relationship will be very different than for someone who comes out after being involved with another person in a heterosexual relationship or who has avoided any expression of intimacy because of fear of being identified as gay or lesbian. In addition to such developmental factors, each decade has brought a very different set of responses to homosexuality in American society and has seen the creation of increasing numbers of social structures and organizations available to gay and lesbian persons. Access to these structures and organizations will vary considerably as a function of the historical period and the chronological age of the individual. Many settings and groups are age-specific and may not welcome persons who are older or struggling with their sexual identity.

The psychiatrist should carefully evaluate the influences of age and generation on the individual gay man, lesbian, or bisexual and appreciate the enormous potential effects of these factors on their experiences of their homosexuality. Most young people growing up gay or lesbian today have been exposed to a vastly different set of social depictions of homosexuality than older persons who grew up decades ago. The portrayals available today are not always more positive and may arise from

sources that are very negative toward homosexuality and gay and lesbian people, but there is greater discussion of the topic and more visibility of a wider range of possible role models for young people. The older person who is struggling with coming out today will often have more deeply ingrained internalized homophobia with which to contend than someone of the same age who came out earlier. The greater number of gay and lesbian venues and the increased visibility of gay and lesbian persons can, for some persons who come out at an older age, lead to increased frustration and lowered self-esteem as a result of problems in accessing these potential supports and self-recrimination about the inability to come out earlier. These variable responses must be carefully assessed in clinical work with lesbians and gay men.

Race and Ethnicity With respect to the race or ethnic background of the gay, lesbian, or bisexual individual, consideration must be given to two further issues: the attitudes within the specific group toward sexuality and homosexuality and the additional prejudice directed toward some racial and ethnic groups within society. These two factors may compound the experiences of stigma and discrimination with which gay men and lesbians of color and from some cultural groups will have to contend. One example occurs within communities of African Americans, which may have varying attitudes and beliefs about sexuality and homosexuality resulting from such factors as family structure, religion, and class. The interaction between the effects of racism in the gay community and homophobia in the African-American community may lead to inadequate coping techniques and poor self-concepts among African-American gay men. Finding validation within either the African-American or gay and lesbian communities may be difficult, and integration of potentially conflicting identities based on race and sexual orientation may be an important need for persons from these groups. Similar issues may arise for persons from other racial and ethnic groups. Hispanic individuals are often influenced by antigay and antilesbian religious attitudes and traditional gender role and family expectations. Native-American concepts of gender, sexuality, and sexual orientation often diverge considerably from non-Native notions of these constructs and may incorporate beliefs about spirituality as well. Because Asian-American gay men and lesbians come from a large number of different language and cultural groups that have diverse attitudes toward homosexuality and gay, lesbian, and bisexual persons, it is impossible to generalize about their experiences. Thus, although membership in a particular group may assist in predicting family, religious, and social reactions to homosexuality, the diversity of backgrounds within many legally designated minority groups requires careful assessment of each individual's experience, level of cultural identity, and degree of acculturation.

The expression of sexual attraction and identity may be strongly influenced by racial and ethnic identity. These identities may be experienced as conflictual, resulting in such strategies as maintaining complete secrecy about homoerotic feelings, adopting a bisexual instead of a gay or lesbian identity, or even feeling the need to choose between cultural and sexual orientation identities. For some persons the conflict between identities may lead to profound confusion and alienation from both groups. The helpful merging or compartmentalization of potentially oppositional personal identities into an integrated sense of self may be extremely difficult to achieve, and this struggle must be appreciated and understood by psychiatrists who work with multicultural gay and lesbian populations.

Religion The impact of religion on the lives of gay men and lesbians derives from the importance of religion both in individual history and in society. Individuals come from different religious backgrounds that may involve exposure to a wide range of attitudes toward homosexuality, and they may or may not maintain ties to the religions of their families of origin. Socially, religion influences gay men and lesbians who belong to particular religious groups that may present supportive, tolerant, or negative views of homosexuality as well as all gay and lesbian persons through shaping the larger cultural and political climate toward homosexuality. Many gay men and lesbians attempt to avoid negative religious teachings about homosexuality by eschewing any contact with religion; others work within their religious groups to try to change negative attitudes about homosexuality; and still others stay attached to, while feeling powerless to change, religions that reinforce their internalized homophobia. Many gay men and lesbians have actively sought a reconciliation between their sexual identity and their spirituality within existing religious groups, and others have moved to create new churches and spiritual settings that will affirm their sexual orientation and their relationships.

Increasingly, religious organizations within American society have been forced to deal with homosexuality both as a personal issue for their members and as a source of theological contention, and a growing number of religious groups have moved to positions that either tolerate or actively embrace gay men and lesbians. At the same time, in response to the rising social discourse about homosexuality some conservative religious organizations have reaffirmed their opposition to homosexuality and become more outspoken about the immorality of individuals who adopt gay, lesbian, and bisexual identities. Regardless of their personal relationship to religion or their involvement with a specific religious group, all gay men and lesbians have been profoundly affected by religious teachings about homosexuality.

Negative religious views about homosexuality often become internalized in lesbians and gay men and may seriously interfere with healthy psychosexual development, particularly involving sexual expression and relationships. Spiritual development itself may be impaired because of an inability to reconcile the conflicting internal and external messages concerning the erotic and spiritual aspects of the self. Until the gay or lesbian person is able to overcome such negative religiously based self-evaluations, it may be impossible for them to achieve a healthy identity. Some individuals may not be able or wish to disengage from their religions even when their homosexuality is judged harshly, and many others struggle to establish a connection to some other, more affirming spiritual community. The psychiatrist must understand the important need felt by some gay and lesbian persons to resolve these difficult situations without totally cutting themselves off from their religious groups or families. However, the pull to retain connections to such groups is sometimes very strong and it is essential to provide appropriate support for individuals who are unable to extricate themselves from this situation. Even when the separation is completely successful, however, serious psychological wounds may remain that continue to distort or truncate those dimensions of experience associated with spirituality and sexuality. These concerns may become an important focus in clinical work with lesbians and gay men.

Interaction of Sexual Orientation With Stage of Development Experiences associated with sexual orientation lead to a variety of tasks, challenges, and opportunities for gay men and lesbians at different points in their lives in relation both to levels of biological and psychosocial maturation and to the extent of coming out and of acquiring a gay or lesbian identity. In general the psychosocial and physical issues confronting gay men and lesbians during different developmental stages are similar to those for heterosexual persons, but there are also specific issues that occur as a result of being gay and lesbian during childhood, adolescence, young adulthood, midlife, and old age. Although the developmental tasks that occur across the life cycle do not occur at the same age for all persons, even when there are similar experiences related to awareness and expression of sexual orientation, most persons will have to confront certain tasks during each stage of development.

Childhood During childhood the issues for children who will grow up to be gay or lesbian derive primarily from the experiences on the part of some of them of being different and the related gender-role variations that some of these children demonstrate. Feeling different can lead to a sense of alienation that may result in reactions such as social isolation and denial. Boys who appear less masculine may have deficits in self-esteem, which may be reinforced by being shunned, humiliated, and derided by their peers in school and other social settings. Such children may also be criticized or otherwise devalued within the family, and some parents and mental health professionals may try to shape or enforce the formation of more traditional gender-role behaviors. Other parents may ignore or shun their children when they fail to conform to expectations of heterosexual development. These experiences usually lead to an internalization of the negative reactions of others and serve to pressure some children to reject the devalued aspects of self they associate with their sexuality and gender. If these effects are severe or are accompanied by incidents of overt discrimination or violence during childhood, the personality may be further damaged and lead to later impairment in the capacity for psychological, social, and work adaptation.

Adolescence Adolescence is a particularly vulnerable time for most youth, who need to begin during this period to establish a coherent identity, separate from their families of origin, and practice patterns of relating and working that will be further developed in early adulthood. Many gay and lesbian youth first become aware of their same-sex attraction during puberty when they begin to mature physically. Gay and lesbian adults often recall the development of a cognitive awareness of their sexual orientation during adolescence as well and many of them may also move through the early stages of establishing a gay or lesbian identity, begin to have affectional and sexual relationships with persons of the same sex, and develop connections with gay and lesbian communities. Successful accomplishment of these tasks can be inhibited or facilitated by the reactions of families and others to the emerging gay or lesbian identities and social interactions. Most gay and lesbian youth are also involved in heterosexual relationships during adolescence.

After telling his parents that he was gay, a 16-year-old high school student sought therapy from an openly gay psychiatrist. The parents insisted that the adolescent male enter therapy for his problem but the youth insisted that he would only see a lesbian or gay psychotherapist. During the evaluation the patient reported depression of several years standing, difficulty in his school work, and social isolation. The parents also were interviewed and stated their unhappiness with their son's declaration to them and their wish for him to become heterosexual.

The patient entered individual psychotherapy accompanied by some joint sessions with his parents. During the course of therapy he described his awareness of his sexual orientation prior to adolescence, his fear of being labeled gay in school, and his anger at his parents' refusal to accept him and his assertion about his sexual orientation. Over the course of two years in psychotherapy his long-standing symptoms of depression (including lethargy, withdrawal, lack of interest in school, and feelings of self-hatred) diminished and his mood improved considerably. He established better communication with his parents who were less openly critical of his sexual orientation. His grades improved and he developed a small network of close friends in high school who were aware that he was gay. Regular therapy ended when the patient moved to a large city to attend college. During a number of follow-up appointments over a period of several years when he visited his parents, the patient reported good adjustment at school and success in establishing satisfying relationships with other men. His parents were now supportive of him and seemed to accept his sexual orientation.

Most gay and lesbian adolescents either consciously or unconsciously delay some aspect of the development of their sexual orientation identity until they can obtain a greater degree of safety and support, which might be accomplished by going away to school, moving to an urban environment that allows more anonymity and acceptance, or creating a network of friends who will be accepting of the gay or lesbian identity. Some adolescents may suppress any awareness of same-sex attractions, others may defensively attempt to establish heterosexual relationships and identity, and others may consciously avoid any outward display of their sexual orientation. Less often but increasingly frequently in some urban settings, gay and lesbian youth may clearly identify their sexual orientation at an early age, disclose their feelings to others, and complete significant aspects of gay and lesbian identity formation during adolescence. Such youth may be able to avoid some of the problematic developmental decisions made by previous generations of gay men and lesbians, such as entering into unsatisfactory heterosexual relationships, believing that having and raising children could only occur within traditional heterosexual relationships, and forgoing any intimate relationships with persons of either sex because of the fear of disclosure of a gay or lesbian sexual orientation.

Because of the unavailability of adequate role models and resources for support, the ongoing stigma associated with being gay or lesbian, the frequent victimization of visible gay and lesbian youth in schools, and the absence of acknowledgment and affirmation from parents and other significant figures in their lives, most pregay and prelesbian adolescents undoubtedly continue to experience significant difficulty in forming and integrating a healthy gay or lesbian identity during adolescence. The psychiatrist should be aware of these negative influences on development for youth who may be gay or lesbian and be sensitive to the potential effects of these factors on their mental health.

Young Adulthood The period of transition from adolescence to adulthood is significant in the lives of many lesbians and gay men who often have their first opportunity to come out and to establish gay and lesbian identities independent of their families of origin during this time. Gay men and lesbians in their 20s and 30s confront the same developmental tasks as heterosexual men and women of the same age, including developing a career, establishing a social identity, achieving the capacity for intimate relationships, and perhaps childrearing, but these tasks may be accomplished in different ways and at different times because of the influence of being gay and lesbian. For example, as a result of negative external forces and internalized homophobia, some gay men and lesbians may experience a delay in the consolidation of their sexual identity and may compensate by investing relatively more energy in their careers during early adulthood, and because many young gay and lesbian adults have not been able to date others of the same sex during adolescence, they may engage in sexual and peer-group relationships that do not follow traditional heterosexual patterns of relating. The psychiatrist should not view these relational patterns as inadequate or abnormal but rather as adaptive for persons who have had to struggle to overcome societal disapproval of their identities and behaviors and to learn to create satisfying sexual and affectional bonds in the absence of visible social structures or role models for such relationships.

During early adulthood many heterosexual persons marry and have children. Gay men and lesbians are less likely to form long-term relationships at this age, but most of them will enter into primary relationships at some point. In previous generations a significant proportion of men and women with attractions to persons of the same sex adopted a heterosexual persona and entered into heterosexual marriages during early adulthood; many of these individuals also had children and raised families. Although currently there are a greater variety of options available to gay and lesbian people during this period, such as living in gay communities with supportive friendship networks, forming primary same-sex relationships, attempting in a self-aware manner to be in heterosexual relationships, and choosing to have or to adopt children, many gay and lesbian young adults continue to feel limitations in their choices as a result of discrimination and the effects of internalized homophobia. The threat of human immunodeficiency virus (HIV) infection and acquired immune deficiency syndrome (AIDS) has also had enormous impact on the sexuality and relationships of gay men by creating new barriers to establishing a healthy sense of well-being and to forming intimate relationships.

In the area of creativity and work, gay men and lesbians may in some instances choose careers during early adulthood that tend to ensure greater acceptance of their sexual identity. Although gay men and lesbians are represented in every type of occupation, they may tend to migrate in greater numbers to those fields that are not stereotyped as traditionally masculine or feminine. Some reports suggest that because gay men and lesbians have had to contend with and overcome negative reactions to them based on stereotypes of gender-nonconforming behaviors, they are better able to transcend gender limitations and to integrate masculine and feminine attributes more fully into their personalities. As a result, in the area of work some lesbians may be more able than many heterosexual women to achieve successful careers, and some gay men are able to be employed in fields historically associated with women, such as nursing and the fashion industry. Nonetheless, discrimination in hiring and promotion continues to be widespread for gay men and lesbians and the lack of legal protection creates considerable work-related stress for many of them.

Midlife The midlife decades of the 40s and 50s are marked for gay men and lesbians, as they are for heterosexual men and women, by consolidation of career identity, concerns about generativity, and emerging problems with health and aging. Gay men and lesbians currently in midlife represent the first generation to have openly identified themselves as gay and lesbian in significant numbers and as a result are forging new developmental patterns in relation to their sexual orientation. Many midlife lesbians and gay men appear to have been able to achieve a greater balance between their career and relationship than other groups, particularly when compared with heterosexual men.

Midlife lesbians are considerably more likely to be employed in professional or technical and managerial positions when compared with other women, although they do not always receive incomes commensurate with their level of employment. Lesbians and gay men may have to struggle with integrating their identities into the workplace and many are confronted with limitations on their advancement at work because of discrimination. Individuals who have invested more heavily in their careers in order to compensate for a lack of fulfillment in their relationships or in other areas of their lives resulting from unresolved issues about being gay or lesbian may be particularly vulnerable to such restrictions and rejections at work.

Many lesbians and gay men may have children to raise during midlife, as the result of previous marriage, adoption, or decisions to have and raise children outside of traditional heterosexual relationships. However, many of these parents do not receive the same support from their extended families as heterosexual parents do and thus may rely much more heavily on friendship networks for support of their families. During midlife and earlier, some gay and lesbian parents may be involved in extensive legal battles over custody and support of their children. Many gay men and lesbians who do not have children will seek outlets for their need to be generative in midlife in their friendship circles, through service as mentors or foster parents, or in other types of family and community involvement.

Gay men and lesbians have the same concerns about health during midlife as heterosexual men and women; however, in addition, they may suffer from a lack of access to adequate health care as a result of fear of disclosure of their sexual orientation and inadequate knowledge on the part of health care providers about their special needs. As a result of high rates of HIV infection, a large number of gay men are also confronted during midlife with issues of death and dying and increasingly with the need to live with long-term disease and uncertainty about health status.

Old Age Persons of any sexual orientation in their 60s and older have a wide range of experiences in relation to aging; some will develop serious illnesses and become incapacitated while others will live healthy lives for two or more decades. All older persons need to deal with the approach of death at some point, and older gay men and lesbians have additional factors to face. Stereotypes of older gay and lesbian persons are usually negative, portraying them as lonely and unhappy. However, they have been shown to be just as well adjusted as their heterosexual counterparts, and they may be better able than younger gay people to contend with the negative reactions to their sexual orientation because they may no longer be concerned about disclosure at work and they may have learned to accept their sexuality.

The best predictors of psychological adjustment in older gay men appear to be their commitment to their sexual identity and their integration into the gay community. Older gay men generally prefer to associate with their age peers, and many of them report that they are in relationships. A large majority of older gay men and older

lesbians report that they are happy, deny significant concerns about loneliness or fear of death, and describe good integration into social networks.

Problems for older gay men and lesbians exist because of the lack of adequate resources for older gay people, the widespread reluctance to disclose a nonheterosexual sexual orientation to health care and other providers, and the almost total absence of acknowledgment of gay and lesbian persons and their relationships and needs within institutions and agencies that provide care for older persons. The failure to acknowledge older gay and lesbian persons is a function both of the denial of sexuality in older people in general and of an unwillingness to recognize relationships that are not legally sanctioned. The resulting invisibility of gay men and lesbians in most health care and custodial settings that provide services to older persons perpetuates the stigmatization and discrimination that many of these men and women have experienced in relation to their sexual orientation throughout their lives and may reawaken old psychological wounds. The psychiatrist working with older persons should make every effort to acknowledge and affirm the sexual orientation and the intimate relationships and friendships of these patients. Same-sex partners and gay and lesbian friends may often be excluded within the health care setting, which may significantly compromise the quality of life and care for the older gay or lesbian patient.

Families and Relationships Gay, lesbian, and bisexual persons are born into the same families that produce heterosexual children; however, growing up in their families provides few of these children with the opportunity to have their emerging sexual orientation acknowledged and affirmed or to be exposed to positive adult gay and lesbian role models. The vicissitudes of being raised in families that often fail to support their sexual feelings and identity lead many gay men and lesbians to create what have been called *families of choice*, friendship networks that provide the interpersonal bonds and resources that were lacking in their families of origin or that generally occur outside of these families, such as in adult friendships. The majority of gay men and lesbians form their own nuclear family structures as well, in the form of steady romantic same-sex relationships, long-term cohabiting relationships, and single or co-parent households with children. The findings from a growing body of research have begun to describe these couples and families.

Families of Origin Versus Families of Choice Many gay men and lesbians have grown up in families that have either ignored or denigrated their sexual feelings and emerging identities; sometimes they report significant physical and emotional abuse from their families related to their sexual orientation. The reactions of parents and other family members to the disclosure by a young person of feelings of sexual attraction to persons of the same sex or of a gay or lesbian identity can significantly influence the self-esteem and self-acceptance of gay and lesbian youth and help shape expectations of acceptance or rejection by others in the world. Even in optimal situations, however, gay and lesbian youth may expect no more than tolerance of their sexual orientation within their families; rarely will they receive active acceptance or empathic support for the experiences and concerns that occur in relation to their emerging sexuality.

A 25-year-old woman sought treatment because of problems in her graduate degree program. Initially, she did not reveal her attraction to women but instead talked about the emotional abuse she felt from her conservative family who had never accepted her personal or educational choices. She gradually provided a history of unsatisfactory attempts to establish romantic relationships with men and her early recognition of her primary sexual and affectional preference for women.

The focus of the treatment was initially on her conflicts with her family and within her educational program. She described a longstanding pattern of disagreement with and ridicule by her parents and siblings because of her refusal to follow their conventional gender role expectations for her in terms of her physical appearance, dress, behavior, and career aspirations. Parallel conflicts had developed in her educational program, as evidenced particularly in her relationships with women instructors. Over the course of three years in weekly psychotherapy, she gradually began to discuss her reluctance to come out as a lesbian because of fear of her family's reaction to her and her own guilt about religious prohibitions. She tentatively began to explore the lesbian community in her town and eventually formed a relationship with a woman. Her conflicts within her graduate school program resolved as she developed a more secure lesbian identity and established satisfying relationships within the gay and lesbian community. She also came out to her parents and confronted what she experienced as their abusive treatment of her.

At the conclusion of therapy the patient moved in with the woman she had been dating. In subsequent follow-up appointments she reported improved relationships with her family and successful completion of her graduate degree program. She and her partner celebrated a religious commitment ceremony after living together for 1 year, and the patient appeared satisfied and quite happy.

Once gay and lesbian young people achieve appropriate separation from their families of origin, they often create networks of friends and acquaintances who provide them with the type of support and mirroring that was previously lacking. These families of choice may be extremely important in the lives of lesbians and gay men, serving the same functions as an extended family for many by helping them to pass through developmental crises, being present at celebrations and rituals, and providing them with a source of comfort and intimacy. In some cases these friendships may be more enduring than romantic relationships and more important than a lover or partner in the daily life of the gay man or lesbian. Families of choice often coexist with biological families, with each providing different types of nurturance. Holidays and social gatherings may involve combinations of individuals from biological and friendship relationships. The clinician who works with these individuals must understand and accept the importance of families of choice in the lives of lesbians and gay men.

Couples The majority of gay men and lesbians report being in a committed romantic relationship, with surveys indicating that 45 to 80 percent of lesbians and 40 to 60 percent of gay men are currently in such relationships. From 8 to 14 percent of lesbian couples and from 18 to 25 percent of gay male couples report that they have lived together for more than 10 years. In contrast to stereotypes of gay men and lesbians, they clearly form and maintain intimate same-sex relationships. While same-sex marriage is still legally prohibited, many same-sex couples live in relationships that are as enduring and emotionally significant as any heterosexual relationship, in spite of discrimination against these relationships, the absence of rituals and laws that support them, and the denial of equal protection and benefits to these couples.

Findings from the limited but growing research on gay and lesbian couples reveal that lesbian couples tend more frequently to be sexually exclusive than male couples. In addition, gay male and, particularly, lesbian couples, in comparison with heterosexual couples, generally have more equality in their relationships and do not differentiate household functions according to gender-based categories of work. Gay men and lesbians report the same degree of global satisfaction in their relationships as heterosexual men and women. However, undoubtedly as a result of gender socialization, women tend to appraise the value of and rewards from their relationships more highly than men, but gay and heterosexual men and lesbian and heterosexual women describe similar levels of valuing and rewards.

Empirical research by David McWhirter and Andrew Mattison led to the description of six stages in the development of male-male relationships: blending, nesting, maintaining, building, releasing, and renewing. Each stage is marked by specific characteristics and a time span although the stages may overlap and occur in a different sequence for some couples and the characteristics may appear in more than one stage. The relevance of this stage model for lesbians has not been determined, but some variation from the gay male experience should be expected. For example, lesbians may spend longer in the tasks associated with blending and the reappearance of the individual during the maintaining stage may not develop until later in lesbian relationships. Understanding the stages of gay and lesbian relationships can be helpful in identifying discrepancies between partners in progressing through the stages and in assessing stage-related problems in same-sex relationships.

Gay and Lesbian Parents and Their Children Many lesbians and gay men have had children, historically most often within heterosexual marriages and more recently through other avenues as well, including adoption, alternative insemination, and co-parenting arrangements outside of marriage. Numerous studies since the late 1970s have demonstrated the characteristics of gay and lesbian families and the impact of having a gay or lesbian parent on children. A significant change in recent years has been the greater visibility of families with gay and lesbian parents and the increased viability for gay men and lesbians of the choice to be involved in childrearing.

Research on gay and lesbian families was initially undertaken to disprove psychological and judicial assumptions about the harm that might be done to children raised by gay and lesbian parents. The findings resulting from such research have confirmed that lesbian and gay parents neither impact on their children in a negative way psychologically nor produce significantly different outcomes in the gender identity, gender role, or sexual orientation of their children when compared with the children of heterosexual parents. Children with gay and lesbian parents do have to struggle with their difference from heterosexual families and may have difficulty overcoming the stigma associated with this situation. The first era of research into the effects of gay and lesbian parents on their children has now yielded to a period of greater emphasis on understanding the dynamics of and determining the potential differences between families with heterosexual, gay, and lesbian parents. Future studies may involve greater methodological precision, focusing more on family process than on structure alone and allowing for improved understanding of differences between gay and lesbian families, derived from characteristics such as religion, age, class, race, and other characteristics, as well as sex and sexual orientation.

The emergence of families with openly gay, lesbian, and bisexual parents and children since the 1970s represents one of the most significant shifts in the structure of the family during the latter part of the twentieth century. The psychiatrist must recognize and validate the importance of this phenomenon in the lives of millions of

men, women, and children and not confound the clinical evaluation and approach to treatment of patients from these families with negative portrayals derived from political, religious, or psychological sources.

CLINICAL APPROACHES

For the most part the psychiatric concerns of lesbians, gay men, and bisexuals will be the same as those of the general population, and clinical work with them should involve approaches to assessment and treatment that are identical to those used with other patients. In addition, the psychiatrist must be aware of certain unique aspects of clinical work with this population.

General Considerations In spite of the fact that homosexuality and bisexuality are not considered to be mental disorders, many health care providers, including psychiatrists, continue to believe that homosexuality is associated with mental illness and dysfunction, report feelings of discomfort with gay and lesbian individuals, and admit to varying degrees of bias toward these men and women. Some even engage in attempts to change a person's sexual orientation when it is not heterosexual. As a result of the real risks of encountering disapproval, discrimination, insensitivity to their concerns, or abusive treatment within clinical settings, many gay, lesbian, and bisexual persons are reluctant to disclose their sexual orientation without a clear demonstration of a nonjudgmental and accepting attitude on the part of the clinician. Because of psychiatry's historical role in defining homosexuality as a type of mental illness and the persistence of negative views toward homosexuality on the part of some psychiatrists, clinicians must be particularly sensitive to the fear of mistreatment within psychiatric settings on the part of lesbians, gay men, and bisexuals. Optimal evaluation and treatment will therefore need to include a respectful and thorough assessment of all aspects of sexual orientation relevant to the care of a patient and careful monitoring of the reactions of patients to such inquiry and of the demeanor of the psychiatrist when discussing a potentially volatile and painful topic.

Inclusive and Special Clinical Settings Gay men and lesbians are for the most part treated in the same inpatient, outpatient, and residential clinical settings as other patients. In recent years a small number of special units have been developed for working with gay, lesbian, and bisexual persons—either community health centers intended to serve the needs of the local lesbian and gay communities or units geared to particular problems like HIV infection and substance abuse. Whether services are provided in general or specialized settings, staff need to be trained in how to deliver sensitive and unbiased care to gay men and lesbians and to be knowledgeable about the special concerns of this population. The clinical setting and staff with whom the patient comes into contact are important determinants of the climate of inclusiveness perceived by gay, lesbian, and bisexual patients. The explicit communication in these settings of awareness of and respect for diverse sexual feelings and identities is an essential component of successful psychiatric treatment.

The language and behavior of psychiatrists and other staff members should convey an awareness that not all individuals and families are heterosexual. For example, questions about relationships should be neutral with respect to the sex of partners. Clinicians should not assume that all married persons are heterosexual or sexually exclusive with their marriage partners. Questions about sexuality are most helpful when they inquire about desire, behavior, and identity as distinct entities. The availability of publications and forms indicating a recognition of diverse sexual orientations and family arrangements in the clinical setting and waiting area can help to reassure the gay, lesbian, or bisexual person as well as persons who are uncertain about their sexual orientation that they will be accepted in this setting and can talk about their concerns. Openly acknowledging that people have different types of sexual and emotional attractions and relationships can be a powerful tool for facilitating the honest presentation of a wide range of feelings and concerns about sexuality.

Another important strategy in clinical work with gay and lesbian patients is to routinely include appropriate references to significant friends and members of nontraditional families during evaluation and treatment. Same-sex partners and children in these families can be involved in the care of gay and lesbian patients to the same extent heterosexual partners and their children would be involved. For example, it may be important to obtain collaborating information from partners, to identify the potential source of a patient's problems in the context of the couple or family, or to include consideration of a partner's behavior in maintaining compliance. The same restrictions on confidentiality and decision making that apply to other families will also be relevant when working with gay and lesbian families.

Sexual Orientation of the Psychiatrist The effects of the psychiatrist's sexual orientation and of disclosing this orientation to the patient are complex. In general, the psychiatrist's sexual orientation will not be the most significant factor in determining clinical outcomes. Any therapist working with gay men and lesbians needs to maintain a conscious awareness of the potential for the development of antigay attitudes and heterosexist assumptions; possess a clear knowledge base about lesbian, gay, and bisexual issues; consistently apply gay-sensitive and gay-affirmative approaches to psychotherapy; and obtain adequate supervision when necessary.

When the lesbian or gay patient works with a psychiatrist who is openly gay or lesbian, there are several possible benefits and risks. Benefits of working with a psychiatrist of the same sexual orientation may include an easier rapport, greater efficiency resulting from shared knowledge about experiences associated with being lesbian or gay, and positive role-modeling effects of therapists who have established a healthy gay or lesbian identity. Possible risks for such pairs include the occurrence of unwarranted assumptions about shared experiences or personal characteristics based on sexual orientation alone, overidentification of the therapist or patient with each other, and a tendency to avoid discussion of certain events or affects associated with being gay or lesbian because of common knowledge about the pain associated with them.

It is also important to understand how the psychiatrist's sexual orientation is disclosed. For example, sometimes the patient has requested to work with an openly gay or lesbian therapist and has this knowledge before beginning treatment; in other cases this information becomes known after treatment has begun, either as the result of self-disclosure by the psychiatrist or from some other source. If self-disclosure is involved, the timing of the disclosure may have a significant impact on facilitating or inhibiting the work in therapy. It is important for the psychiatrist to be able to identify those gay and lesbian individuals and those situations for which working with an openly gay or lesbian therapist might be most helpful.

Sexual orientation may be far less significant than other personal characteristics in determining the effectiveness of a psychiatrist in working with gay, lesbian, and bisexual patients. A gay or lesbian therapist who has not fully come out or established a comfortable identity may have trouble working with patients struggling with similar concerns, and a therapist of any sexual orientation who is uncomfortable talking about sexual problems will have more trouble being helpful to gay and lesbian patients who have sexual concerns. Ultimately, the degree of empathy experienced and expressed by the therapist for the patient will in most cases be the most important factor in determining the effectiveness of treatment. Psychiatrists need to be vigilant about the dynamics of the ongoing interaction between their real or perceived personal characteristics, including but not limited to their sexual orientation, and the work of their patients.

Assessment of Sexual Orientation A careful and thorough assessment and description of sexual orientation is rarely completed in the clinical setting. When undertaken at all, an assessment usually does not go beyond designating a person as “homosexual,” or even less often, as “heterosexual” (heterosexuality usually being assumed). Sexual orientation is multifaceted and there are a variety of scales that can be useful in its assessment. These scales assess different components of sexual orientation—including desire, behavior, and identity—and evaluate the relative degree of sexual orientation for each component. Additional variables may also be included, such as changes over time in sexual orientation, social and emotional preference for one sex or the other, life-style choices, and gender-role identity. Sex research may involve psychophysiological quantification of sexual attraction through techniques like penile and vaginal plethysmography.

A 63-year-old man sought psychotherapy with complaints of depression and alcoholism, exacerbated since his father's death 2 years earlier. He did not present with concerns about his sexual orientation, and he described himself as “homosexual,” rather than as “gay.” He was employed in a very public setting, where he had always avoided questions about his private life. He reported that he did not feel comfortable with other homosexual people or in gay social settings. He had maintained a relationship for many years with another man who lived in a nearby city, spending one or two nights a week at this man's home and all holidays with him. When asked about various aspects of his sexuality, he stated that he had always only been attracted to men, had infrequent sexual interactions with other men in the past, but he did not consider himself to be gay. He described gay people as “flamboyant” and used derogatory terms to describe a “gay lifestyle.” Physically he was a large and attractive man who dressed impeccably and formally in a traditionally masculine manner.

During the course of psychotherapy, this man discussed his lifelong effort to hide his sexual orientation from others but he did not seek to change his attitudes about his homosexuality. He clearly had exclusively homosexual desire and behavior, but he denied any experience of a gay identity. His work in psychotherapy focused on his problematic relationships with his parents, his reactions to their deaths, his longstanding alcohol abuse, and his wish to retire and seek a new career. Although he discussed his views of his homosexuality extensively and attended an Alcoholics Anonymous (AA) retreat for gay men and lesbians while in therapy, he did not seek wider connections with other gay people and never saw himself as having a gay identity. At the end of psychotherapy, his depression resolved, he had no relapses of his alcoholism, and he successfully navigated his retirement. He was also less critical in his attitudes toward homosexuality in general, but his sexual orientation was not fully integrated into his self-identity.

The utility for the individual psychiatrist of the various approaches to assessment will depend on the extent to which information about sexual orientation is relevant to a particular patient's concerns. However, knowledge about the various scales and techniques informs the clinician about the complex and multidimensional nature of sexual orientation and helps prevent the tendency to attribute stereotyped global characteristics to individuals based on knowledge of one aspect of their sexuality. Assessment of sexual orientation on several dimensions can help to better understand the meaning and purpose of sexuality in an individual's life.

In evaluating sexual orientation the psychiatrist must balance the need to obtain all relevant information with the need to be sensitive when inquiring into an area about which a patient may feel very vulnerable. In addition, sexual orientation must not be approached solely as a sexual matter, as it also involves relational patterns, family structures, friendship and support networks, and community and cultural activities for many lesbians and gay men. Consideration of each of these psychosocial dimensions is important in developing approaches to treatment in work with lesbians and gay men.

Special Concerns Certain concerns are unique to lesbians, gay men, and bisexuals. The presentation of other concerns that do not directly relate to sexual orientation may also be shaped by the fact that a person is gay, lesbian, or bisexual, and for some persons being a member of a sexual minority group may interfere with and even prevent interactions with the health care system.

Societal and Domestic Violence The problem of hate-based violence directed against lesbians and gay men is widespread and well documented; it ranges from verbal abuse to violent homicide and is perpetrated by individuals, within families, by gangs, and by authorities such as the police. In addition, although studies have failed to show any causal connection between childhood sexual abuse and developing a gay, lesbian, or bisexual sexual orientation as an adult, sexual abuse of gay men and lesbians during childhood appears to occur at about the same rate as in the general population. Recent studies demonstrating the existence of domestic violence and abuse in lesbian and gay male relationships demonstrate that physical abuse occurs at about the same rate as in heterosexual relationships.

The effects of these various types of violence on gay men and lesbians are the same as they are on other victims of violence, including posttraumatic stress disorder, depressive disorder and a range of other psychological and emotional problems, but may also involve specific symptoms associated with being members of a stigmatized sexual minority group. The internalized homophobia of the victim of antigay or antilesbian violence may be triggered by the assault and lead to the belief that being gay or lesbian is the cause of the assault, and to feelings of inadequacy, embarrassment, or shame, or—depending on the severity of the attack and on the vulnerability of the victim—even to denial of one's sexuality and identity and avoidance of other gay and lesbian individuals. Gay and lesbian survivors of incidents of hate-based violence and of sexual or domestic abuse may encounter more difficulty reporting their experiences and seeking help than heterosexual persons because of fear of disclosure of their identity and of further harassment. The psychiatrist should be prepared to respond to the immediate crisis that may follow such an incident of abuse or violence, to the potential long-term effects of these events on the individual, and to the ongoing concerns about the various forms of violence directed against them on the part of all gay men and lesbians. Treatment and prevention can best occur in the clinical setting and in the community following the development of programs that will ensure adequate protection and safety for gay men and lesbians.

Increased Suicidality Determining the rate of suicide attempts and completed suicide among lesbian, gay, and bisexual persons is difficult because of such factors as the absence of information identifying sexual orientation on death certificates, the reluctance of many individuals who are too distressed about their emerging sexuality to discuss their feelings with mental health professionals, the general failure to assess sexual orientation adequately in most health care settings, and the inability to obtain representative samples. In spite of these problems in research, results from recent studies demonstrate that among youth being gay, lesbian, or bisexual increases the risk for attempting suicide. Among all youth, young women are more at risk for attempting suicide and young men are at greater risk for completing suicide. Suicide rates among gay, lesbian, and bisexual youth are also likely to vary based on the presence of other risk characteristics, such as mental illness, substance abuse, rejection by family, and degree of gender role nonconformity. In addition, increased risk of suicide may accompany struggles with internalized homophobia and problems with coming out and establishing a healthy gay, lesbian, or bisexual identity at any age.

Determining the exact numbers of persons who attempt or complete suicide who are gay, lesbian, bisexual, or struggling with their identity is ultimately less important than understanding the characteristics of these groups that contribute to risk for suicide. For gay youth, the invisibility of sexual-minority young people and the lack of adequate resources to address their health and mental health needs often prevent the necessary assessment of their concerns and the development of specialized programs for prevention and treatment. The politicization of funding for research to study suicidality among gay, lesbian, and bisexual young people has resulted from the equation of efforts to identify mental health concerns with attempts to promote minority sexual orientations, as though the need to obtain information about a medical problem is itself the source of the problem.

All youth in distress must be carefully evaluated for the presence of concerns about sexuality and sexual identity, and the psychiatrist can utilize approaches to assessment and treatment that do not place these persons at further risk following disclosure of their minority sexual orientation. Premature coming out can increase the potential for victimization of various kinds, including rejection by peers and family. The risk of suicide for persons of any age who are struggling with coming out and establishing a gay, lesbian, or bisexual identity must be recognized and clinical services appropriate for managing suicidality and for clarifying needs and providing support to these persons should be available when relevant.

Increased Rates of Alcohol and Substance Abuse Although findings from research into the problem vary considerably and controlled studies comparing comparable groups of heterosexuals with lesbians and gay men are absent, alcohol and other substance use and abuse appear to be increased among gay men and lesbians, with an incidence perhaps as much as two to three times higher than in the general population. Although lesbians and gay men consume more alcohol and are more likely to abuse substances at different points in their lives than the general population, they do not appear to have greater rates of drug or alcohol dependency. The causes of the problems with alcohol and other substance use are generally attributed to the psychosocial stressors associated with being lesbian, gay, and bisexual and to the tendency for gay and lesbian people to socialize in bars and clubs that encourage alcohol and drug use. Rates of substance abuse appear to be equally high among lesbians and gay men.

For some persons, links between substance use and abuse and problems with internalized homophobia are established when coming out and perpetuated during the stages of establishing a gay, lesbian, or bisexual identity, creating a profound and disturbing connection between sexual orientation and addiction. In addition, substances are often used to disinhibit prohibitions on sexual expression, and as a result many lesbians and gay men develop their sexual behaviors and patterns in relation to use of these substances. Identification of the possible comorbidity between substance abuse and problems with development of sexual orientation can be essential in defining adequate approaches to treatment of substance abuse among gay men, lesbians, and bisexuals. Specialized programs that simultaneously address problems with substance use and with establishing a healthy identity have been shown to be helpful for treating substance abuse in many gay, lesbian, and bisexual persons. Some 12-step and other self-help programs provide special settings in which gay men and lesbians can openly discuss their sexuality and identity as well as their use of substances.

The psychiatrist should be prepared to identify the overlap between problems with acceptance of one's sexual orientation and substance abuse. The signs of disturbance or delay in acquiring a healthy lesbian, gay, or bisexual identity and of addiction may be similar, including traits such as denial, fear, anger, guilt, helplessness, hopelessness, dishonesty, low self-esteem, self-hatred, and social isolation and alienation. Assessment of these problems and awareness of how each reinforces the other is essential if adequate treatment is to be accomplished. It is unlikely that recovery from substance abuse in gay, lesbian, and bisexual persons can occur without also establishing some degree of acceptance of sexual orientation; similarly, problems with sexual identity are very difficult to overcome while abusing substances that numb and interfere with self-awareness. As a result, programs designed for treatment of substance abuse must be both gay-sensitive and gay-affirming if they are to treat substance abuse disorders among gay men and lesbians successfully. Prevention of substance abuse among this population must involve efforts to address social and cultural factors that may predispose to use of substances, including the development of settings for socializing outside of bars.

Problems With Sexuality and Sexual Dysfunction The majority of problems concerning sexual functioning for gay men and lesbians are similar to those for the heterosexual population. Often, however, psychiatrists and other health professionals are uncomfortable talking about sexuality in general; dealing with the specific sexual concerns of gay men and lesbians may be particularly difficult because of unfamiliarity with or disapproval of same-sex sexual behaviors. The psychiatrist needs to be able to discuss the sexual concerns of gay and lesbian patients objectively and comprehensively and to be aware of special diagnostic issues and approaches to sex therapy with these men and women. Descriptions of existing diagnostic categories for sexual dysfunction and disorders have a heterosexist bias, which may lead to an avoidance of reference to concerns of same-sex couples, such as problems with anal intercourse and problems secondary to concerns about HIV infection for gay men and affectional and orgasmic difficulties for lesbians.

A variety of special problems may develop in the area of sexuality and sexual function for gay, lesbian, and bisexual persons. Gay men, for example, may struggle more with issues of sexual compulsivity, and bisexual men and women may also struggle with mastering safer-sex practices. Patterns of same-sex sexuality have been shown to be different from those in heterosexual couples, with female couples having decreased sexual interaction and male couples initially having increased amounts of sexual interaction in comparison with heterosexual couples. Typologies of male couples have sometimes described them in terms of their degree of sexual

exclusivity and openness, and issues related to sexual exclusivity may concern male couples. In contrast, female couples present more often with problems related to decreased sexual desire, which may be associated with a tendency for greater emotional and interpersonal intimacy and dependence among women.

The psychiatrist working with sexual problems in gay and lesbian individuals and couples will need to consider organic, developmental, interpersonal, and social causes. Assessment and treatment should adapt traditional heterosexually oriented paradigms of sexuality to include consideration of issues related to gay and lesbian development and to the dynamics of same-sex relationships. Special attention should be paid to cultural factors that may contribute to sexual problems among gay men and lesbians, including a failure to obtain accurate information about same-sex sexual behaviors resulting from prohibition on education about the topic and internalization of negative societal attitudes about same-sex sexuality.

Physical Problems Lesbians and gay men present with the same medical problems as the general population but also have particular problems in relation to their health and medical care that are associated with their sexual orientation; these include dealing with negative attitudes of health care providers; accessing providers who are sensitive and knowledgeable about their concerns; and dealing with an increased risk for some diseases. Knowledge of a person's current gay or lesbian sexual orientation and behaviors should not exclude careful inquiry about current or past heterosexual relationships and experiences that could also put them at risk for certain diseases.

There are no gynecological problems that are unique or occur more frequently among lesbians than among heterosexual women, although they may be at greater risk for undetected disease as the result of receiving fewer pelvic examinations and Pap smears. Sexually transmitted diseases are rare among women who are sexually active only with other women, and there are few cases of HIV infection in these women. The risk for some types of cancer in women may be affected by factors such as history of sexual intercourse with men, number of pregnancies, and breastfeeding. Historically lesbians have had a lower number of pregnancies than heterosexual women, but an increasing number of lesbians are choosing to have children through alternative insemination. It is important to consider the special needs of these women, a higher proportion of whom may seek care from nontraditional providers such as midwives.

In general, gay men have increased rates of sexually transmitted diseases, HIV infection, and use of certain substances such as inhaled nitrites (poppers). Since the 1980s the influence of HIV infection and AIDS on the individual lives and the communities of gay men has been enormous and has affected almost every aspect of the personal and public experience of being gay. As a result, HIV infection must be considered as a potential problem in the life of every gay man who enters the health care system, in relation to such issues as testing for HIV infection, ensuring early treatment intervention in infected individuals, preventing further HIV infection, responding to fears about contracting the disease, dealing with the chronic stress resulting from pervasive loss and illness, and taking care of gay men who are already infected. For many gay men HIV infection is negatively associated with aspects of their sexual identity and exacerbates pre-existing problems resulting from internalized homophobia. Recent advances in the treatment of HIV infection may diminish the importance of this disease in the lives of gay men, but current generations of these men have been profoundly affected by it. The psychiatrist can play an important role in differentiating functional and organic symptoms that occur in relation to HIV infection.

Physical illness of any type in gay men and lesbians may interact with experiences associated with their sexual identity. Life-style factors, family structures, and social networks can determine the extent and type of support available; the stress of being sick may evoke psychological reactions involving internalized homophobia; and the capacities for intimacy, affection, and sexuality, whether with the same or the other sex, may be significantly altered as the result of being sick. The psychiatrist should be sensitive to the unique issues that may arise for the gay man or lesbian when confronted with physical problems, including the lack of support from families who have rejected many of them because of their sexual orientation.

Psychotherapy The topic of psychotherapy with gay men and lesbians has been explored in a large number of publications that identify the special problems of these men and women and describe effective approaches to treatment. Treatment modalities include individual, couples, family, and group therapies. Since the 1970s most of these volumes have been written from the perspective of *gay and lesbian–affirmative psychotherapy*, described by Alan Malyon as a “theoretical disposition [that] regards homosexuality as a non-pathological human potential. The goals of gay-affirmative psychotherapy are similar to those of most traditional approaches to psychological treatment and include both conflict resolution and self-actualization.” Gay and lesbian–affirmative psychotherapy recognizes the centrality of sexual attraction, behaviors, and identity in a person's life but also appreciates that sexual orientation is only one aspect of personality and, depending on the patient's needs and goals, may or may not be a central focus of therapeutic work.

Antigay, Antilesbian, and Heterosexist Bias Within Psychotherapy In spite of the availability of helpful gay and lesbian affirmative approaches to psychotherapy, gay men, lesbians, and bisexuals are still at risk for encountering negative reactions to their sexuality and identities within psychotherapy. One study has identified 25 different ways in which bias can affect psychotherapy with gay men and lesbians in association with assessment, intervention, identity, intimate relationships, and therapist expertise and education. The psychiatrist should be informed about the detrimental effects of inappropriate and negative assumptions about patients based on their sexual orientation. Often these assumptions are a reflection of inaccurate information and outdated knowledge, but they may also result from personal prejudice. Regardless of their source, they may reinforce negative feelings in vulnerable persons seeking psychotherapy.

A limited number of authors continue to present homosexuality as a form of mental illness, to portray gay and lesbian identities as pathological adaptations, and to describe approaches to curing homosexuality. These views are inconsistent with current knowledge about sexual orientation and fail to follow contemporary diagnostic standards. These approaches generally seek to eliminate sexual attraction to persons of the same sex and to encourage the formation of heterosexual relationships. Careful review of reports of these so-called reparative therapies have failed to demonstrate their effectiveness in changing sexual orientation and have clarified the extent to which they confound an understanding of sexual orientation with gender identity and gender role. Behavioral approaches to eliminating same-sex desire and sexual behaviors developed from the 1950s through the 1970s were abandoned because of recognition of their lack of efficacy and of the harm they inflicted on persons struggling with their sexual feelings and identity. Currently these negative approaches are described almost exclusively by psychoanalytic and religiously affiliated writers.

Generally, persons who knowingly seek such treatment are experiencing reactions to internalized homophobia that are normative developmental events for gay and lesbian persons growing up in a heterosexist society with widespread antigay and antilesbian attitudes. The response to these reactions by a therapist who subtly or directly denigrates same-sex sexual feelings and expressions may interrupt or totally prevent development of a healthy gay or lesbian adaptation and can lead to acting out and other self-destructive behaviors. Temporary changes in behavior or relationships may result from a wish to please the therapist, particularly in persons who have already had heterosexual relationships, but long-term changes in sexual desire or expression resulting from psychotherapy have rarely been described or verified. Clinical reports of subsequent work in psychotherapy with patients who have been the subjects of such attempts to change their sexual orientation have demonstrated their negative effects in encouraging acting out, maintaining lack of self-acceptance, and prolonging struggles with identity acquisition.

A 52-year-old man entered psychotherapy with symptoms of depression and anxiety. He had been maintained on an antidepressant medication prescribed by his physician for the past year and reported a good response; however, recently he had experienced some return of his symptoms of tiredness, distraction, and self-doubt as well as concerns about his physical health. The patient reported a history of having been arrested three years previously after having sex with another man in a public setting. As a result of this incident, he lost his job and his marriage of twenty-five years had ended. After his arrest he was in psychotherapy with a psychiatrist for one year. The patient also reported that prior to his arrest he had seen two therapists, both of whom tried to help him stop his homosexual behavior and to restrict his sexual activity to his marriage. The patient reported that the new psychiatrist attempted to help him achieve a heterosexual adaptation and to suppress his homosexual feelings. At the end of this period the patient, who was also alcoholic, attempted suicide and was admitted to a residential substance abuse treatment program. While in this program, he stopped seeing the psychiatrist, entered an AA program, and began to work on accepting his homosexuality. The patient then became involved in local gay community activities and begun to date a man who was his own age. The were talking about living together.

During the course of once a week psychotherapy, this man talked about his lifelong sexual attraction to men and of his efforts to achieve heterosexual satisfaction with his wife to whom he had been deeply committed. He believed that he would have killed himself if he had not ended the therapy with the psychiatrist who wanted him to suppress his homosexual feeling and not begun a treatment program for his alcohol abuse. His symptoms of depression and anxiety decreased after examination of his fears about forming an intimate relationship with another man. He eventually purchased a home with the other man and reported satisfaction with his involvement in the gay community, becoming a leader in a number of political and social organizations. In follow-up appointments over a period of five years the patient reported episodic exacerbation of his symptoms of depression but no longer felt any conflict about his sexual orientation. He maintained sobriety, was satisfied in his gay relationship, and was successful in a new career.

Role of Gay and Lesbian Affirmative Psychotherapy Acceptance, acknowledgment, and affirmation of gay, lesbian, and bisexual persons are essential elements in work in psychotherapy with these men and women. These therapeutic stances reflect the therapist's regard and empathy for the patient's sexual feelings and for their

same-sex relationships, in contrast to the denial and derogation with which these aspects of the patient have often been met. Ultimately, as with any patient, the psychiatrist working with gay and lesbian patients should maintain a position of neutrality that will both help to contain the struggles of the patient and at the same time serve to convey a recognition of the ongoing negative impact of societal prejudice against gay and lesbian persons. Lack of awareness about and failure to explicitly acknowledge the effects of the negative interpersonal and social forces working against the expression of same-sex desire may suggest to the patient that the psychiatrist agrees with these attitudes, thereby representing a deviation from the appearance of neutrality. Explicit agreement with these negative attitudes—even when these views may be shared by some patients, as in the case of the therapist who utilizes so-called reparative approaches—constitutes therapeutic bias.

Being lesbian, gay, or bisexual represents a fundamental experience of feeling different and frequently alienated from mainstream culture. Often the sexual, romantic, and affectional expressions of gay and lesbian patients have been ridiculed and rejected. For many gay men and lesbians, speaking with a therapist may represent the first chance they have had to discuss their stigmatized sexuality with another person. The fundamental importance in therapy of having these feelings listened to without judgment, either positive or negative, having them acknowledged as real by a person in authority, and having them affirmed as commonly felt and normal cannot be emphasized enough. The tolerance and validation of such feelings by the therapist will allow the patient to express negative feelings as well, such as a fear of the reactions of others, negative feelings about being gay or lesbian, and regret about losses associated with being gay or lesbian. At these times the therapist should not reassure patients prematurely but rather should allow them to express and to come to terms with these experiences—which are often manifestations of internalized homophobia—at their own pace.

Special Issues Gay men and lesbians present to psychotherapy with the same concerns as other patients, seeking help with the full range of mental disorders and adjustment to interpersonal, work, and social situations. In many cases the sexual feelings, behaviors, or identity of the individual may be incidental to treatment; in other cases these issues may be central to the treatment. For example, some gay and lesbian patients who may have no problems with their sexual feelings or identity will present with symptoms of sexual dysfunction whereas others will seek help directly because of confusion about their attraction to someone of the same sex or of struggles with coming out or identity formation. As long as they are lesbian and gay affirming, practitioners of all schools of psychotherapy can be effective in working with lesbians and gay men; the same tools of evaluation, planning, and treatment used for other patients should be employed with these men and women.

It is helpful for the psychiatrist to be aware of several special treatment issues relevant to psychotherapy with lesbians and gay men. These issues can be clustered into general topics that have been discussed in other sections of this chapter: developmental experiences, including problems of youth, coming out, and acquiring a gay or lesbian identity; effects of heterosexism and antigay and antilesbian attitudes, including violence and internalized homophobia; families and relationships, including families of origin and families of choice, same-sex relationships, and having children; and physical and mental problems, such as alcohol and substance abuse, sexual dysfunction, and HIV infection. The experience of all these issues will be significantly influenced by the patient's biological sex, race, and other personal and social characteristics.

ETHICAL AND LEGAL ISSUES

Psychiatrists have an ethical responsibility to their patients to avoid harm and to attempt to ensure their well-being. In addition to this general responsibility several specific ethical and legal issues are relevant to psychiatric work with lesbians, gay men, and bisexuals. These involve confidentiality, medical records, the potential for abuse in military and other discriminatory settings, and maintenance of standards of care, including the avoidance of heterosexist and antigay and antilesbian bias. Confidentiality is important in relation to sexual orientation because of the risk of discrimination against lesbians, gay men, and bisexuals when their sexual orientation is known within the health care setting and by health insurers. Furthermore, because sexual relations between persons of the same sex are still illegal in more than 20 states, a medical record that states that someone has had such relationships could put the person at risk for being charged with a criminal act if this information were revealed. Although knowledge of the sexual identity of a patient may be essential in order to provide quality care, it is inappropriate to include this information in a medical record or to transmit it to others without the informed consent of the patient. In all instances the psychiatrist should discuss with the patient the intent to record this information before it is entered into the chart. When the gay or lesbian patient does not wish this information to be recorded and when the psychiatrist cannot agree to withhold this information, referral to another provider should be offered. Information about sexual feelings, behaviors, identity, and relationships should only be entered into the medical record when it is relevant to the evaluation and treatment of the patient and must always be treated in a confidential manner.

Psychiatrists working in the military, prisons, certain religious organizations, or other settings where knowledge about a person's sexual orientation could be used to harm the individual are faced with a difficult ethical dilemma. If the psychiatrist's knowledge that patients are gay or lesbian must be shared with authorities who will then use this knowledge to discriminate against them, for example, by discharging them from the military, then the psychiatrist may be torn between taking part in a discriminatory act that is prohibited by the code of professional ethics and violating the operating rules of the organization. Disclosure of the risk of harm that may come to patients who reveal their sexual orientation to providers who are then obligated to report this information to discriminatory organizations, like the military, is the minimal action required by psychiatrists in such circumstances. In the author's opinion, refusal to participate in the discriminatory process is the only behavior fully consistent with the professional code of ethics for psychiatrists.

Maintenance of appropriate standards of care in relation to sexual minority status requires the avoidance of heterosexist and antigay and antilesbian bias in providing assessment and treatment to gay men, lesbians, and bisexuals. Patients have suffered historically from psychiatric diagnosis of homosexuality as a type of mental illness; currently, a number of psychiatrists and other mental health practitioners continue to treat homosexuality as a form of mental illness even though it is no longer included in DSM and no empirical justification for this view exists. Diagnosis of sexual minority status as a mental illness or developmental problem and offering treatment for it represent overt expressions of bias on the part of the psychiatrist. Such pathologizing allies the authority of the mental health field with prejudicial attitudes and actions. These views not only place individual patients at risk but are also used to support the maintenance of institutionalized acts of legal and legislative discrimination, such as the denial of custody to gay and lesbian parents.

Other expressions of bias may be more subtle, including, for example, inappropriate recommendations and acting-out behavior on the part of gay and lesbian psychiatrists who are struggling with coming out themselves and dealing with feelings of internalized self-hatred, or the failure of psychiatrists to achieve competency in treating gay and lesbian patients by acquiring adequate knowledge about this group and recognizing the effects of discrimination on development. Gay and lesbian psychiatrists struggling with their own acceptance may need to refrain for a period of time from working with patients who have similar problems. Psychiatrists should undertake training to keep them informed about the concerns of gay and lesbian patients as well as of those from other minority groups. Problems in working with gay men and lesbians resulting from bias are often expressed indirectly and may be manifested as a tendency on the part of the therapist to marginalize or diminish the extent of problems such people experience. Psychiatrists should assume that they have learned heterosexist and antigay and antilesbian attitudes, and be alert to both direct and indirect forms of expressing such bias.

TRAINING AND RESEARCH NEEDS

Concepts of sexuality and sexual orientation have significantly shifted in the second half of the twentieth century. Psychiatric residents need to receive systematic training about the new body of knowledge concerning sexuality and the dimensions of sexual orientation and to develop attitudes and skills that will allow them to interact sensitively and competently with persons of all sexual orientations. Teaching about gay men, lesbians, and bisexual persons cannot be isolated to classes on sexuality but needs to be integrated into the general curriculum and should provide clinical examples of gay, lesbian, and bisexual persons with different forms of mental illness who are well adjusted to their same-sex relationships and identities. Many lesbian, gay, and bisexual residents have needs for special support during their training, and all residents can benefit from exposure to respected role models who are openly gay, lesbian, and bisexual. In the author's opinion, medical schools and professional organizations must increase the number of opportunities available for training about the clinical concerns of sexual minority individuals and encourage gay and lesbian faculty and other leaders to be visible and active in a wide range of professional activities.

Although much has been learned during the past several decades about sexual orientation and its development, a great deal more remains to be understood. Similarly, although attitudes toward gay, lesbian, and bisexual individuals have improved considerably during this period, antigay and antilesbian bias continues to be widespread and to have harmful effects on the development of gay and lesbian people. Since the 1800s the biomedical and social science research agenda in relation to sexual orientation has focused largely on finding a cause for homosexuality, whether in family dynamics, anatomical structures, or hormonal and gene variations. Cultural and psychological factors contributing to the development of sexual orientation have generally been of less interest. Until the 1970s virtually all this research on homosexuality attempted to locate a cause for a condition that was considered pathological. Since the removal of homosexuality from the list of mental disorders in 1973, researchers have been able to ask much more complex questions about the phenomena based on conceptualizations of sexual orientation that recognize its diverse forms and expressions over the lifespan.

As a result of the deepened understanding of sexual orientation, preliminary research has been undertaken to describe variations in sexual desire, behavior, and

identity associated with a range of demographic and personal characteristics, such as biological sex, age, race, class, geographical location, and religion. Although many studies continue to focus on white, urban, middle-class men, the need for further examination of variations of sexual orientation in other groups, such as women and persons of color, is now broadly recognized.

Research on the topic also needs to utilize longitudinal studies that will describe the development and experience of sexual orientation over time and across the life span. Retrospective methods alone are inadequate for describing the experiences of persons because of the inevitable subjective reconstruction involved when recalling history. Longitudinal studies will also expand understanding about those characteristics in childhood that contribute to the development of adult sexual orientation, as well as factors that shape individual expression of desire and behavior in adults.

Future research into sexual orientation will inevitably continue to evolve further beyond attempts to locate the cause for a single phenomenon representing some unitary factor or cluster of factors in all persons and developing consistently in a linear fashion across the life span. The development of a postmodern perspective has led to an appreciation for the wide range of cultural, familial, interpersonal, and individual elements that contribute to the construction of our sexual and relational selves and to an awareness of the profound and changing influence historical and social forces have on language and meaning, including the meaning of sexual orientation. Future research about sexual orientation and attitudes toward it must expand beyond the use of traditional biological and psychosocial approaches to utilize constructivist methodologies that also account for the role of fluctuating cultural and other historical forces in shaping individual lives and experience.

SUGGESTED CROSS-REFERENCES

Normal child development is discussed in Section 32.2, and normal adolescent development in Section 32.3. Normal human sexuality, other than homosexuality and homosexual behavior, and sexual dysfunctions are discussed in Section 19.1a. Gender identity disorders are discussed in Section 19.3 and the psychotherapies are discussed in Chapter 30.

SECTION REFERENCES

*Bayer R: *Homosexuality and American Psychiatry*. Princeton University Press, Princeton, NJ, 1987.

Bell AP, Weinberg MS: *Homosexualities: A Study of Diversity Among Men and Women*. Simon & Schuster, New York, 1978.

Bell AP, Weinberg MS, Hammersmith FK: *Sexual Preference: Its Development in Men and Women*. Indiana University Press, Bloomington, IN, 1981.

Bieber I, Dain HJ, Dince PR, Drellich MG, Grand HG, Grundlach RH, Kremer MW, Rifkin AH, Wilbur CB, Bieber TB: *Homosexuality*. Basic Books, New York, 1962.

*Cabaj RP, Stein TS, editors: *Textbook of Homosexuality and Mental Health*. American Psychiatric Press, Washington, DC, 1996.

*D'Augelli AR, Patterson CJ: *Lesbian, Gay, and Bisexual Identities Over the Lifespan: Psychological Perspectives*. Oxford University Press, New York, 1995.

DeCecco JP, Elia JP, editors: *If You Seduce a Straight Person, Can You Make Them Gay? Issues in Biological Essentialism Versus Social Constructionism in Gay and Lesbian Identities*. Haworth Press, New York, 1993.

DeCecco JP, Parker DA, editors: *Sex, Cells, and Same-Sex Desire: The Biology of Sexual Preference*. Haworth Press, New York, 1995.

Drescher J: *Psychoanalytic Therapy & the Gay Man*. Analytic Press, Hillsdale, NJ, 1998.

Ford CS, Beach FA: *Patterns of Sexual Behavior*. Harper & Row, New York, 1951.

Foucault M: *The History of Sexuality, Volume I: An Introduction*, R Hurley, translator. Pantheon, New York, 1978.

Friedman RC: *Male Homosexuality: A Contemporary Psychoanalytic Perspective*. Yale University Press, New Haven, CT, 1988.

Garnets LD, Kimmel DC, editors: *Psychological Perspectives on Lesbian and Gay Male Experiences*. Columbia University Press, New York, 1993.

Glassgold JM, Iasenza S, editors: *Lesbians and Psychoanalysis*. Free Press, New York, 1995.

Gonsiorek JC, Weinrich JD, editors: *Homosexuality: Research Implications for Public Policy*. Sage Publications, Newbury Park, CA, 1991.

Greenberg DF: *The Construction of Homosexuality*. University of Chicago Press, Chicago, 1988.

Hamer D, Copeland P: *The Science of Desire: The Search for the Gay Gene and the Biology of Behavior*. Simon & Schuster, New York, 1994.

Herdt G: *Guardians of the Flutes: Idioms of Masculinity*. McGraw-Hill, New York, 1981.

Herdt GM, editor: *Gay and Lesbian Youth*. Harrington Park Press, New York, 1989.

Herek GM, editor: *Stigma and Sexual Orientation: Understanding Prejudice Against Lesbians, Gay Men, and Bisexuals*. Sage Publications, Thousand Oaks, CA, 1998.

Herek GM, Berrill KT, editors: *Hate Crimes*. Sage Publications, Newbury Park, CA, 1992.

*Hooker EA: The adjustment of the male overt homosexual. J Project Techn *21:* 17, 1957.

Isay RA: *Being Homosexual: Gay Men and Their Development*. Farrar, Straus, Giroux, New York, 1989.

Kinsey AC, Pomeroy WB, Martin CE: *Sexual Behavior in the Human Male*. WB Saunders, Philadelphia, 1948.

Kinsey AC, Pomeroy WB, Martin CE: *Sexual Behavior in the Human Female*. WB Saunders, Philadelphia, 1953.

Laumann EO, Gagnon JH, Michael RT, Michaels S: *The Social Organization of Sexuality: Sexual Practices in the United States*. University of Chicago Press, Chicago, 1994.

Magee M, Miller DC: *Lesbian Lives: Psychoanalytic Narratives Old and New*. Analytic Press, Hillsdale, NJ, 1997.

*Malyon A: Psychotherapeutic implications of internalized homophobia in gay men. J Homosex *1:* 59, 1982.

McWhirter DP, Mattison AM: *The Male Couple*. Prentice-Hall, Englewood Cliffs, NJ, 1984.

McWhirter DP, Sanders SA, Reinisch JM, editors: *Homosexuality/Heterosexuality: Concepts of Sexual Orientation*. Oxford University Press, New York, 1990.

Murray SO: *American Gay*. University of Chicago Press, Chicago, 1996.

Patterson CJ, D'Augelli AR, editors: *Lesbian, Gay, and Bisexual Identities in Families: Psychological Perspectives*. Oxford University Press, New York, 1998.

Savin-Williams RC, Cohen KM, editors: *The Lives of Lesbians, Gays, and Bisexuals: Children to Adults*. Harcourt Brace College Publishers, Fort Worth, TX, 1996.

Stein TS, Cohen CJ, editors: *Contemporary Perspectives on Psychotherapy with Lesbians and Gay Men*. Plenum, New York, 1986.

Weinberg MS, Williams CJ, Pryor DW: *Dual Attraction: Understanding Bisexuality*. Oxford University Press, New York, 1994.

19.2 PARAPHILIAS

STEPHEN B. LEVINE, M.D.

Sexual behaviors that deviate from intuitive notions of normality create uneasiness in large segments of the culture. Because psychiatry defines sexual abnormality and takes some responsibility for alleviating the suffering that abnormality generates, the profession's roles are complicated. A wide variety of sexual identity disorders and sexual dysfunctions exist. The paraphilias have almost forty different pathological manifestations. Also, it is easier to conceptually define an abnormality than to agree about where normality ceases and pathology begins. The paraphilias range from dangerous to relatively benign, from the bizarre to the almost ordinary, from the compulsively pursued to the incidentally noted. Even the dangerous, bizarre, and compulsive paraphilias blend uncertainly into ordinary sexual expression. Finally, the classic psychiatric approaches to any problem—definition, etiology, and treatment—are complicated for the paraphilias by the inherent tension between the conflicting social needs of sexual freedom and sexual regulation. Despite these formidable intellectual difficulties, the paraphilias bring mental health professionals to an awareness of the paradoxes, dilemmas, and enigmas of human sexual life—particularly male sexual life.

Society worries about the paraphilias because they can be detrimental to mental and physical well-being. Some of them, like rape and pedophilia, victimize others. Others can be ultimately self-destructive, as in autoerotic asphyxia, or chillingly self-destructive as when people pay to be injured. Still other paraphilias, such as shoe fetishism, seem harmless, unless the perspective of the partner is taken into account. Regardless of their patterns, persons who have a paraphilia cannot be easily identified objectively—some are socially unconventional and many others live rather ordinary nonsexual lives. Society identifies offenders only after victims have been created. Although the diagnosis of paraphilia is not well known, sexual victimization is. Some victims are created by individuals who are on a compulsive quest for sexual gratification through conventional sexual activities such as seduction. In seeking frequent relief, sexually compulsive persons victimize themselves and others. Psychiatry has not yet decided to classify these patterns as a variant form of paraphilia.

Three forces seem to keep sexual behaviors within acceptable social boundaries. The first is the presence of biological normality. This fortunate state of bodily health creates the substrate for personality development that is predisposed towards civilized sexual behavior. Continuing physical and mental health enables an adult to continue to uphold the requirements of nonvictimizing sexual expression. The second force consists of the socialization processes of family, educational, and religious life. These inculcate values about what is moral and immoral within relationships. The third is the legal system, which defines sexual crimes. Legal systems criminalize the egregious victimization of children, adolescents, and adults and reinforce moral standards through punishment. These general forces—health, education, and the law—are civilizing influences on every individual's development.

Psychiatry is now highly interested in the biological forces that may account for the paraphilias. For most of the twentieth century, however, individual psychological development was used to explain these disorders. Development meant the subtle private, conscious and unconscious, cognitive, emotional, and motivational processes that integrated all that occurred to and within a growing person. Whatever their ultimate causes, paraphiliac symptoms seem to be the result of (1) an abnormal sexual preoccupation—for example, “I am excited by the image of humiliating a person,”; (2) a deficiency of control over sexual attractions—for example, “I touched her on the escalator because she had nice thighs”; or (3) both—for example “the urge to exhibit my penis overcomes me and I cannot stop myself.” From these descriptions alone, it is clear that society expects a civilized person to be excited only by certain behaviors and to choose the context for sexual expression carefully.

A disorder such as a paraphilia is a pathological manifestation of something that more often exists in a different form. Paraphilia is an abnormal intention. *Intention* refers to what people want to do with their partners during sexual behavior and what they want their partners to do with them. Intention is the seldom-recognized third component of sexual identity. Like the widely recognized components of gender identity and orientation, intention has separate private and behavioral aspects. *Erotic intentions* are the mental aspects—they consist of attractions, fantasies, ideas, or scripts that generate intense arousals. *Sexual intentions* refer to the behavioral aspects—they consist of what one actually does with others or to oneself to become excited.

A person's sexual identity is largely developed in a private psychological manner. It is consciously derived from the individual's evolving self-observations on: (1) the erotic and behavioral aspects of his or her gender identity, orientation, and intention; (2) the labels the person eventually assigns to these components; and (3) the meanings that the components have to the person. Normal intentions, which are rarely even conceptualized by conventional people, involve desires to cooperate with a partner to create pleasure in a mutually acceptable, often mutually orgasmic fashion. Abnormal intentions, which are typically thought about a great deal by the paraphilic person, involve aggression, victimization, and extreme one-sidedness; that is, behaviors that are difficult to negotiate with others because of their bizarre or idiosyncratic nature.

The paraphilias are specific abnormalities of intention. Problems of gender identity or orientation phenomena are not paraphilias. Paraphilic behaviors occur among individuals with conventional and unconventional gender identities and among those of all orientations. A female who lives as a male and desires to be beaten during arousal is paraphiliac only on the basis of masochism, not because of her gender adaptation. A rapist is paraphiliac because his mental and behavioral intentions are aggressive; his masculinity and heterosexuality are irrelevant to the diagnosis.

DEFINITION

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) recognizes the paraphilias as consisting of recurrent, intensely sexually arousing fantasies, sexual urges, or sexual behaviors that involve either nonhuman objects, the suffering of the self or partner, children, or nonconsenting persons. To qualify as a DSM-IV diagnosis, however, these patterns must have existed for at least 6 months and they have to cause clinically significant distress or impairment in social, occupational, or some other important area of function, such as sexual function.

The DSM-IV specifies nine paraphiliac diagnoses: *exhibitionism* or genital exposure; *voyeurism* or clandestine observation of another person's undressing, toileting, or sexual behavior; *sadism* or causing suffering during sexual behavior; *masochism* or being humiliated during sexual behavior; *pedophilia* or sexual behavior with prepubescent or peripubertal children; *fetishism* or use of nonliving objects for sexual behavior; *frotteurism* or rubbing against or touching a nonconsenting person; *transvestic fetishism* or use of clothing of the opposite sex for arousal; and *paraphilia not otherwise specified* for the other 30 or so observed atypical sexual patterns such as dressing in diapers, requiring a partner who has an amputated limb, and others.

Other Diagnostic Distinctions

Criminalized Sexual Behavior Includes Paraphilia and Nonparaphilia While horrifying acts—rape, murder after or before sexual behavior, and penetration of children—have long been crimes, international awareness about the extent of sexual victimization has been growing since the 1970s. Now other sexual cruelties and childhood victimizations are routinely prosecuted. Date rape, incest, sexual behavior with children and young adolescents, sexual stalking, sexual harassment, professional sexual misconduct, and sexual violence against homosexual persons occupy the judicial system. Nonviolent offenses involving exhibitionism, voyeurism, frotteurism, and obscene phone calls are criminal acts as well, but prosecutors tend to view them as relatively harmless nuisance crimes even though they may not be. Nonviolent perpetrators are apt to be referred to mental health professionals as a condition of probation; more dangerous and destructive sexual cruelties are punished with incarceration.

The terms “sex offender” and “sexual perpetrator” refer to a person who has committed a sexual crime. However, all sexual criminals do not meet the criteria for a paraphilia—for instance, many incest offenders and rapists do not, and homosexual men arrested for public indecency during solicitation in public parks are usually not diagnosable as having exhibitionism. Sexual crimes are sometimes referred to as hands-on versus hands-off offenses. Typical hands-off crimes are exhibitionism, voyeurism, sexual harassment, and stealing underwear.

Parameters of Paraphiliac Severity

PERSONAL-RELATEDNESS Clinicians are often helped in their treatment planning by assessing the paraphilia patient's ability to relate to another person sexually. Four categories of relatedness can be observed among most of the paraphilia diagnoses.

Solitary Paraphiliac Expressions All paraphilia behaviors are rehearsed repeatedly in fantasy; often these unusual fantasies have been present since childhood or puberty. Some persons with paraphilias never experience any sexual behavior with partners. The shame of the paraphilia interest and the fear of negative consequences may contribute to a lifelong avoidance of intimate contact. Sexual behaviors consist entirely of masturbation, often with magazine or Internet pictures and bulletin boards, videotapes, reading material, clothing, or mechanical props. People in this category are the most psychologically unprepared for sexual expression within an ordinary relationship.

Paraphiliac Behaviors with Hired or Unmet Partners Female and male prostitutes have a long collective experience in dealing with paraphiliac men. Men pay to be urinated upon, spanked, observed masturbating, humiliated, beaten, cross-dressed, or told a paraphilial story. Many businesses sell the opportunity to watch others undress and dance in seminudity; they may provide private booths where men can masturbate in front of the dancer for an additional fee. Prior to the appearance of acquired immune deficiency syndrome (AIDS), some people used to belong to clubs for group sex activities. Men now pay for strangers to talk to them about sex over the telephone while they masturbate. For some computer-literate men, nude magazine collections have given way to the Internet's opportunities to download pictures and videos or to “talk” to others with similar interests in chat rooms. In each of these arenas, the man obtains sexual gratification with an illusion of an interpersonal relationship. Although many nonparaphiliac men partake in these activities, paraphilic customers are able to gratify their masochistic, exhibitionistic, and voyeuristic desires through public masturbation or with a prostitute. For some, this illusion of connection or relatedness is as far as their capacities have taken them.

Secret Paraphiliac Fantasies With Long-Term Partners Some paraphilia persons are able to have a more complex relationship with a partner that includes conventional sexual activity. Their paraphiliac fantasies occur during conventional sexual behavior with their partners. However, their partners are not told of the scenes that are occurring in their minds. Spanking fantasies, for instance, may occupy the thoughts of men and women during conventional sexual activities, but their partners may never know of the accompanying mental masochistic imagery.

Negotiated Paraphiliac Behaviors with Long-Term Partners Couples seen in clinical practice usually consist of a paraphilia patient and a nonparaphiliac partner. In order to sustain a paraphiliac script, the partners must negotiate about behaviors. Success at this process means that the partner understands the specific requirements of the paraphiliac lover; appreciates and accepts that people often have quirky means of attaining sexual gratification; and consents to cooperate. The partner with typical intentions has to learn to accommodate to her partner—for instance, to his wearing women's underwear or to his desire that she wear a shoe to bed. Consent is given because the partner realizes that cooperation is the best way to ensure a consistent opportunity for sexual intercourse; sometimes the partner consents in an attempt to preserve monogamy. Once given, consent may be withdrawn because partners cannot endure the narcissistic injury of not being desired for themselves, or because even with cooperation the paraphiliac partner is unable to have satisfactory sexual activities. Theoretically, the complementary paraphilias of sadism and masochism should sustain each other's script, particularly if each is willing and able to explore both sides of sadomasochism.

REQUIREMENT FOR AROUSAL While the usual functional significance of any paraphilia is that the person needs to use the paraphiliac fantasy or behavior in order to desire a partner, obtain arousal, and reach orgasm, the degree to which this occurs varies within most of the categories of paraphilia.

Absolutely Required In the most severe forms of paraphilia, the specific paraphiliac imagery or activity is absolutely necessary for any sexual function. Without it, the person is always sexually apathetic—there is no desire for sexual behavior with the partner, no sustainable arousal, or no orgasmic attainment. When embarrassment prevents the paraphiliac from sharing the secret, the relationship may never be consummated and the partner is either baffled or settles on a wrong explanation. If the paraphiliac is able to integrate the paraphiliac requirements into the lovemaking, either by engaging in the behavior or imagining it, sexual functioning may even be normal.

Usually Required in Some Manner In less severe forms, ordinary sexual behavior is possible. However, in many of these circumstances, sex only occurs if initiated by the partner, if a fetish is used, if the paraphiliac fantasy is sometimes spoken of, or if it is almost always privately imagined. Personal and relationship dynamics appear to affect the need for paraphiliac assistance, which sometimes is not present.

Not Required but Imagery Occasionally Appears In the least severe form, sexual behavior with a partner is normal but during times of stress, unhappiness, loneliness, or marital discord a longstanding paraphiliac intention appears in masturbation fantasy, during a daydream, or during sexual behavior with a partner. The paraphiliac imagery may be a source of shame, which only creates further distress. These erotic circumstances are sometimes referred to as deviant arousal coexisting with normal arousal.

Enhancements Heterosexual and homosexual couples may on occasion tie one another up, cross-dress, talk dirty, peek at other couples, or watch pornography to create novelty and alternate forms of excitement in their long-term relationship. These behaviors may either arouse, amuse, bore, or disgust the couple, but they do not become a requirement for sexual arousal; they are only experimental efforts to temporarily enhance excitement. Although this is not to be confused with paraphilia, it may be that the use of these paraphilia-like behaviors are only enhancements for one partner while the other partner is highly excited by them but unable to share their actual private significance.

COMPULSIVITY OR IMPULSIVITY The final parameter of severity is the degree of drivenness to masturbate or act out the fantasy with a partner.

Out of Control The most severe form of compulsivity is the loss of autonomy. The loss of autonomy has three characteristics: (1) the need for sexual behavior consumes so much money, time, concentration, and energy that the patient describes himself as out of control; (2) intrusive unwanted paraphiliac thoughts prevent concentration on other life demands and are the source of anxiety; and (3) orgasm does not produce satiety in the way it typically does for age mates. Thus, a man may masturbate six times a day, visit a prostitute twice in a day, or spend $200 on phone sex. Although such drivenness may exist for long periods of time and is presented to the clinician as uncontrollable, a careful history often reveals that both the thoughts and the behaviors can be interrupted by other life demands.

Autonomy Preserved The same paraphiliac desires—for instance, sadomasochism—may be present, distressing, and yet be contained to ordinary frequencies of masturbation and only occasional paraphiliac sexual behavior with another person. The imagery is present but the pressure to act upon it, although distressing, is manageable.

Impulsivity The lack of control of some individuals with paraphilia does not appear to be the result of a constant preoccupation with sexual thoughts; rather, it is because they suddenly perceive an opportunity to behave in a manner they have long fantasized about. For example, an exhibitionist sees two young girls are alone in the street and seizes the moment to expose himself. Often such impulsive behavior is explained as “something came over me.” This often-repeated pattern is not planned for; it occurs without much internal struggle or consideration of the consequences. The trigger for the behavior is the absence of external social control. For example, a pedophiliac grandfather who had not touched a child inappropriately since age 22 and “pretty much forgot about it,” began periodically licking the genitals of his first infant granddaughter because he had the urge and his recently too-busy wife was preoccupied upstairs. Paraphiliac impulsivity often is seen in those with brain pathologies, but organicity is certainly not required.

Internal Compulsivity and External Impulsivity Although there are paraphiliac behaviors that clinicians might agree are either compulsive or impulsive, the distinction between the two concepts often quickly blurs. Clinicians need to be mindful of two circumstances. The first is when the clinician cannot decide whether the person is compulsive or impulsive; he seems to be both. Compulsivity and impulsivity seem to trigger one another. Intrusive sexually arousing thoughts culminate in one episode of sudden acting out or one impulsive act leads to intense preoccupation, masturbatory activity, and a binge of planned paraphiliac acts. The second circumstance is patient deceit. Most pedophiles and incest offenders, for example, scheme to arrange for sexual behavior with a child, and they groom a victim. The perpetrator falsely claims impulsivity. The opportunity that he had been visualizing and organizing his social behavior toward for years finally presents itself, but he hides behind the excuse of being overcome by an irresistible impulse. Grooming is not usually admitted unless clinicians carefully question the patient about the

antecedents to the impulsive act.

Nonparaphiliac Compulsivity and Addiction Clinicians have a diagnostic problem when they evaluate patients with major social and vocational impairments due to sexual compulsivity if the urges are for ordinary sexual behaviors with adults. This is a common circumstance. Devotion to seduction of strangers, frequent uses of prostitutes, relentless requests for intercourse or fellatio within an established relationship, or masturbatory preoccupation with pictures of naked women can erode the autonomy of heterosexual men. Similarly, cruising for sex can occupy days at a time to the detriment of homosexual men's jobs, relationships, safety, and physical health. Women often describe themselves as compulsively seeking love through sexual behaviors with strangers. These patients often say, “I am driven.” Are such cases to be classified as paraphilia not otherwise specified or as sexual preference disorder unspecified?

The answer seems to depend upon the clinician's ideological framework. Psychopharmacologists tend to view these cases as equivalent to paraphilias because of similar consequences, comorbidities, and responses to medication; they are now labeled paraphilia-related disorders. Psychoanalysts use the term near-perversions and assume that the paraphiliac fantasies are unconscious and will emerge in intensive treatment. Other terms used are *sexual addictions* because the patients' destructive loss of autonomy directly parallels consequences of substance abuse or simply *compulsive sexual behavior* because that is what it seems to be. Seven or more orgasms per week in an adult is generally considered a reasonable definition of *sexual compulsivity*. The point of agreement in all of this is that people with “normal” desires can be quite destructive to themselves and their partners.

The term *sexual addiction* arose from Alcoholics Anonymous treatment programs to describe, explain, and treat sexual compulsivity. The term does not distinguish paraphiliac from nonparaphiliac compulsivity and does not have a specific definition. It does convey, however, the person's sense of being unable to control the incessant desire for some sexual behavior. Men and women of all orientations describe themselves as sexually addicted. Many use this term to denote the relentless use of anything sexual: masturbation, pornography, prostitutes, strip bars, cruising, telephone sex, school-age children, or a partner. Some professionals use *addiction* only when the conscious intentions are conventional because paraphilia adequately describes bizarre intentions.

HISTORY

Inversion, Perversion, and Deviance Throughout the late-nineteenth and twentieth century, these three terms were used by mental health professionals to denote the paraphilias. They are officially out of favor now because they no longer connote acceptable standards of objectivity and they have been generically applied to any unconventional aspect of sexual identity—cross-dressing, homosexual orientation, or sexual sadism.

Character Pathology and Sociopathy Early editions of DSM assumed that the paraphilias were manifestations of a severe defect in character formation due to various preoedipal developmental difficulties. As nosologies embraced a more observational basis for classification, the paraphilias came to be recognized as worthy of a distinct category under the sexual diagnoses. Character problems are now carefully defined on Axis II and causes are recognized to be hypotheses.

COMPARATIVE NOSOLOGY

ICD-10 The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) describes the paraphilias as nine categories of disorders of sexual preference (Table 19.2-1). Each of these disorders must meet three criteria: (1) The individual experiences recurrent intense sexual urges and fantasies involving unusual sexual objects or activities; (2) the individual either acts on the urges or is markedly distressed by them; and (3) the preference has been present for at least 6 months. The term *fetishism* is applied to those who rely on use of nonliving objects—for example, footwear, leather, rubber—as a stimulus for sexual arousal and gratification *Fetishistic transvestism* is applied to those who wear articles of clothing of the opposite sex in order to create the appearance and feeling of being a member of that sex. Cross-dressing is associated with sexual arousal and once orgasm occurs and sexual arousal declines, there is a strong desire to remove the clothing. *Exhibitionism* is used for those who have a recurrent or persistent tendency to expose their genitals to unsuspecting strangers, usually of the opposite sex. The act is associated with sexual arousal and masturbation but there is no intention to have sexual intercourse with the witness of the exhibition. *Voyeurism* is the term applied to those who have a recurrent or persistent tendency to look at people engaging in sexual or intimate behavior such as undressing. The act is associated with sexual arousal and masturbation but there is no intention to be sexually involved with the person observed. *Pedophilia* is the diagnosis of those who have a persistent or predominant preference for sexual activity with a prepubescent or early pubertal child or children. The individual must be at least 16 years of age and at least 5 years older than the child victim. *Sadomasochism* is a preference for sexual activity as a recipient or provider or both of pain, humiliation, or bondage. The activity is the most important source of stimulation or is necessary for sexual gratification. *Multiple disorders of sexual preference* is used for those whose sexual preferences include more than one of the above and none has a first-rank position; the most common combination is fetishism, transvestism, and sadomasochism. Other disorders of sexual preference is used for those with any one of a variety of patterns including obscene phone calling, rubbing up against people in crowded places for sexual stimulation, sexual activity with animals, use of strangulation for intensifying sexual excitement, sex with a dead person, preference for partners with a specific anatomical abnormality such as amputated limbs, and other such preferences. *Disorder of sexual preference unspecified* is applied to those whose patterns do not fit any of the preceding categories.

Table 19.2-1 ICD-10 Diagnostic Criteria for Disorders of Sexual Preference

While there are some differences in the labels used for the nine categories of the DSM-IV and the ICD-10, the basic concepts are identical. Both these classifications separate issues of gender identity and orientation from those of intention or preference.

EPIDEMIOLOGY

The sophisticated epidemiological studies of the prevalence of mental disorders published since the late 1980s did not investigate the sexual psychopathologies. Had this been undertaken for the paraphilias and sexual compulsivities, researchers would have faced formidable methodological obstacles such as asking the right questions to delineate cases of the major paraphiliac forms in a relaxed and nonjudgmental fashion; overcoming the political problem of gaining funding for the project from the government or a foundation source; and ascertaining the extent of subject lying about illegal behaviors. There are no large-scale sophisticated community-based studies on the topic. However, studies on sexual criminals, many of whom are paraphiliac, do exist.

Sex Crimes Most adult perpetrators of sex crimes experience multiple paraphilias over time. In a study of 561 nonincarcerated sex criminal volunteers, the average number of paraphilias per person was 2.02. When paraphilias evolve, the behaviors tend to get more destructive to victims over time. Within any particular paraphiliac category, the number of offenses committed by the perpetrators ranged widely. Rape had consistently lower offense rates (7.2 times) than others, such as exhibitionism (500 times). Compared with pedophiles, exhibitionists, frotteurs, and voyeurs were high-frequency offenders. Nonincestuous pedophiles who targeted boys had far more victims than those who targeted girls. Incest offenders had fewer sexual victims than nonincestuous pedophiles, but they had far greater numbers of sexual encounters with their victims than nonincestuous offenders.

Data from 1616 perpetrators between 5 and 21 years of age, most of whom were between 10 and 18 years old, demonstrate that the backgrounds of youthful sexual criminals typically involve emotional neglect and physical and sexual abuse. Nonsexual psychopathic features are common. Perpetrators often recognize their paraphiliac interests by puberty and begin offending early in their lives. These data estimated that 15 percent of forcible rapes are committed by boys under age 18

and 20 percent of all sex offenses are committed by adolescents. At intake 42 percent reported a history of physical abuse, 39 percent of sexual abuse, 26 percent of neglect, and 63 percent of witnessing severe domestic violence. The majority of youngsters had a history of engaging in other antisocial activities. Ninety-one percent of their victims were between the ages of 3 and 16 years; 63 percent were less than 9 years of age. The majority of the sexual assaults involved orifice penetration; most of the perpetrators showed no empathy for the victims; only 19 percent accepted full responsibility for the assaults. Another study estimated that between 30 and 50 percent of child molestations are committed by juvenile offenders.

Reporting of pedophiliac offenses increased at least 300 percent in the United States during the 1980s and is still thought to underestimate the sexual exploitation of children and young adolescents. The number of treatment centers for adolescent sex offenders has increased more than thirty fold. Estimates of the percentage of American children who are sexually abused vary considerably. A conservative estimate is that 15 percent of girls and 8 percent of boys are victimized. Sexual abuse of children is a rampant public health problem.

Paraphiliac Eroticism In a questionnaire study of patients seeking care from family practitioners, 4 percent of 230 patients acknowledged paraphiliac erotic imagery; none admitted to acting upon their fantasies.

Comorbidity Significant psychiatric and physical comorbidities are common among men and women diagnosed with paraphilias, compulsive sexual behavior syndromes, and sex-offending behaviors. Substance abuse, anxiety disorders, and dysthymic and other mood disorders, conduct disorders, or character pathology, for example, are to be expected in the majority of patients who are seen. In a study of volunteers who claimed to be sexually compulsive, for example, the lifetime histories of substance abuse was 64 percent, phobic disorder was 42 percent, mood disorder was 39 percent, and personality disorder was 44 percent. Among 60 consecutive paraphiliac and nonparaphiliac sexually impulsive men seeking treatment, the lifetime presence of comorbid Axis I disorders were: mood—77 percent; dysthymic disorder—53 percent; substance abuse—47 percent; anxiety disorders—47 percent; and social phobias—32 percent. In general psychiatric settings these comorbidities are what are first perceived.

ETIOLOGY

Like most behavioral disorders, paraphilia is multifactorial and eludes a simple explanation. Three broad approximations of causality can be seen among paraphiliac persons: an abnormal brain, gross psychological adversity, and idiosyncratic development. Each of these categories contains many variations and the categories are not mutually exclusive. Paraphiliac phenomena are usually broader than any clinician's experience; what the physician encounters depends to a large extent on the setting. A forensic clinic, adolescent residential treatment center, adult alcohol treatment center, and a private outpatient practice generate different impressions about the forces that commonly shape abnormal intentions.

Brain Abnormalities The association of paraphilia with abnormalities of brain structure or function may be explained as follows: (1) the sexual disorder is a direct result of the abnormal neural organization; (2) the brain abnormality places the person at a social disadvantage, which leads to abnormal intentions; (3) the brain abnormality impairs the person's social judgment; and (4) the brain abnormality decreases the person's control over antisocial impulses. Seizure disorders, mental retardation, attention-deficit/hyperactivity disorder, cerebral palsy, dementia of the Alzheimer's type, multiple cortical infarcts, and postconcussional disorder are occasionally important Axis III diagnoses.

A mentally retarded 22-year-old ruined his first community job as a grocery clerk when on his second work day he spotted a 7-year-old girl wearing MaryJane shoes. He stopped what he was doing, followed the girl for a few minutes and asked her excitedly if he could caress her shoes. The girl's mother almost immediately stopped him and complained to the manager. He could not explain why he has been interested in little girl's shoes since he was young, but he acknowledged masturbating to images of stroking little girl's shoes throughout adolescence. MaryJanes were his favorite since he first saw them in a magazine.

Egregious Developmental Adversity Parental abandonment; recurrent physical, psychological, and sexual abuse; exposure to parental violence; prostitution; or chronic intoxication are often part of the backgrounds of persons with paraphilia and sex offenders. To the extent that a dependable, stable loving relationship with a parent is required for normal sexual development, the opportunities of some individuals with paraphilia have been severely limited from early childhood.

A 15-year-old was brought for assessment to a residential setting by his third foster mother who could no longer tolerate his stealing of her underwear and his sexually provocative behavior with the two other foster children in the home. In the residential treatment center he constantly masturbated and repeatedly initiated sexual behaviors with staff and boy and girl residents. His developmental history was rife with early physical abuse, substance-abusing parents, paternal then maternal abandonment, and cross-dressing since age 6.

A 29-year-old bank loan officer sought help for his addiction to prostitutes, hostile sexual preoccupations, and lack of sexual interest in his wife. His father is a practicing sadomasochist. As his father was leaving his wife and children, he gave the 11-year-old patient his collection of pornographic magazines, the use of which triggered the beginning of more than daily masturbation. The father arranged that his new wife and her live-in best friend separately have intercourse with his son, who was then 16 years old. After this, the patient's incessant seeking of sexual experiences with classmates and prostitutes began. Throughout his adolescence the patient was frequently berated by his mother for being “just like his father—a bum.” The patient always calls the prostitutes some demeaning name before intercourse and thinks of this as getting even with his mother.

Less Extreme Developmental Adversity Often the cause of paraphiliac problems seems to consist of sexual trauma at various stages of life, less than ideal relationships with parents, or idiosyncratic responses to seemingly ordinary events.

A 28-year-old single man on a golf outing thought that he had successfully seduced a woman he met in a bar. As they were having intercourse, the woman's husband appeared. The patient's panic quickly subsided when the couple explained that the husband was a friendly voyeur who just wanted to watch and masturbate. Their negotiations led to 3 days of drunken sexual experimentation for the young man, which culminated in him being dressed in women's clothes by the couple and then having intercourse in front of the voyeur who also was cross-dressed. Never having cross-dressed or ever having indulged in any unconventional sexual behavior, the man sought help 2 years later because of his inability to stop thinking about the weekend with excitement and his continuing experimentation with wearing women's clothing to masturbate.

A substance-abusing professional was finally able to attain sobriety at age 33 years. With this accomplishment, he met a woman and got married, began to work steadily for the first time in his life, and was able to impregnate his new wife. His preferred sexual activity had been masturbation in semi-public places. The patient had a strong sense that his mother had always thought him to be inadequate, did not like to spend time with him, and constantly made negative comparisons between him and his “all-boy” younger brother. He recalled several times when his father had tried to explain his mother's antipathy: “It is just one of those things, son; your mother does not seem to like you.” Without substance abuse, he gave up his exhibitionism, but he quickly developed sexual incapacity with his wife and became “addicted” to phone sex.

A 38-year-old thin, stylish mother of three with lifelong inability to attain an orgasm with a partner sought help for dysthymic disorder with recurrent suicidal ideation, sedative abuse, and aversion to sexual experience with her highly valued husband. She had masturbated almost every day of her life to orgasm to thoughts of being painfully penetrated with medical instruments, burned, or given painful enemas. The medical instrumentation fantasies began at age 19 when she worked in a urologist's office. She thought of her fantasies as “sick, uncontrollable, but necessary to relieve my tensions.” A painfully shy child with school and social phobias until adult life, she was unable to recall any childhood medical procedure, burn experience, or physical or sexual abuse that might explain why her erotic life has been dominated by masochism. She was extremely close to her grossly obese mother, rarely saw her hard-working idealized father, and was reluctant to discuss the possibility that she had experienced any pain and suffering in her home.

Psychodynamic Features Many new ideas have been offered to explain the paraphilias. Some simply ignore psychodynamic considerations. They pass over the fact that intentions are part of sexual identity development and bear important relationships to masculinity and femininity, attachment to others, self-esteem, the intrapsychic balance of warmth and aggressiveness, and the ability to cope with difficult realities. This results in the paraphilias being referred to as variants of other nosological categories: obsessive-compulsive disorder, posttraumatic stress disorder, mood disorders, impulse-control disorders, or addiction, with which they are frequently comorbid. The evidence given for paraphilias being variants is that paraphilia patients seeking help may respond to selective serotonin reuptake inhibitors (SSRI) tricyclic drugs, anxiolytics, or 12-step programs. Comorbidity and response to therapy are not infallible evidence of pathogenesis, however. Even the lives of

patients with classic psychiatric entities have been shaped by psychodynamic forces.

Earlier in the twentieth century psychodynamic explanations were focused almost entirely on intrapsychic developmental considerations. Explanations such as preoedipal helplessness with a frightening mother, castration anxiety, passive-into-active erotization of hated persons or of humiliating events placed little importance on developmental processes that follow the preoedipal and oedipal stages. All psychodynamic hypotheses are moored by two anchors: (1) that the vital intrapsychic developmental steps that are required for the development of normal intentions are not known and (2) that mental life and its erotic aspects continuously develop, recombine, and dynamically fluctuate. Trauma, deprivation, abuse, neglect, and idiosyncratic developmental explanations focus on the original organizing forces for the abnormal intention but they do not necessarily explain their persistence.

Once established, paraphiliac intentions are impressive in their ability to provide an escape from various unpleasant emotional states and life situations. Thus, paraphiliac preoccupations and behavior function as a defense against dysphoria or a frightening life challenge. “Rather than looking for a job and risking being told ‘no' once again, my husband just spends all day looking at pornography, drinking, and masturbating. Life has become too much for him.” Psychodynamic hypotheses that emphasize what paraphilias have in common include denial of the differences between the sexes, fear of castration, fear of the woman's genitalia, and a symbolic act that preserves the gender identity of the individual with a paraphilia.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of a paraphilia or compulsive sexual behavior is easy to make once the patterns of private and behavioral arousal are ascertained and a sexual developmental history is obtained. However, before this, several conditions must be met. The patient has to perceive a good reason to discuss this subject with the clinician. The patient needs to know whether confidentiality exists. The clinician has to know enough about this subject to ask relevant questions about it in calm, nonjudgmental manner that indicates an understanding of the subject. When there is suspicion of paraphilia and patient denial, it is important to obtain descriptions of sexual crimes from police reports or victim statements to help confront the denial.

Transvestic Fetishism This pattern usually begins with private childhood or pubertal use of female undergarments. In the peripubertal period the clothing becomes a stimulus to masturbatory arousal. Years of experimentation with lingerie may give way to the use of outer garments and the imagining of the self as a woman. Cross-dressing and its attendant imagery captures erotic life, retards heterosexual development, and generates shame about this habit. As the young man learns that his pattern is called transvestism and that others like him exist, he may begin to experiment with dressing as a woman in public. “Passing” typically begins with walks around the block at night, drives in a car, or brief appearances at a mall. Eventually, sexual arousal to lingerie weakens. Although masturbation may still occur after cross-dressing, his mental life becomes captured by his dual roles: masculine-in-the-world and feminine-in-solitude.

In adulthood, fetishistic transvestism may take several subsequent paths: private cross-dressing with occasional passing as a woman; part-time living as a woman (these two paths are sometimes referred to as transgenderism); introducing the spouse to his secret and integrating the cross-dressing into sex play (adult form of transvestitic fetishism); and the total abandonment of the male gender role (*transsexualism*).

Most fetishistic transvestites are perceived by others as inconspicuously masculine-enough during grade school children and adolescence. As adults, however, their memories of these times involve feeling uncomfortable as a boy or not being like other boys. Female clothing must have powerful symbolic meanings to these boys because the texture of the underwear or the underwear itself generates sexual arousal. Some young adolescents who are easily excited by female undergarments use them as a prop to their imagining being with a girl who is wearing the garments whereas these boys use the garments to imagine themselves as girls. This experience may contain three aspects: (1) the boy's inadequate sense of his masculinity; (2) his envy of girls for their nurturant abilities; and (3) his hopelessness in becoming an aggressive and popular boy. The cross-dressing adolescent or adult often cannot explain dynamics because the cross-dressing enables him to be excited without thinking or feeling these issues. Such adolescents can, however, recognize them when pointed out. They tend to think of themselves as literally having separate male and female selves, with the female self being far more interesting. They fail to integrate male and female elements into a singular self and often spend their lives in unhappy masculine compartments where they may marry and father children. Married transvestites' sexual desire for their partners is enhanced if they are allowed to wear female garments, but this often impairs their wives' capacities to sustain personal sexual interest in them. Most women are uneasy with or insulted by their husband's cross-dressing; many try to eroticize it but fail. Mere accommodation is the more typical outcome.

Not all fetishistic transvestites are heterosexuals. Some transvestitic males are bisexual, uncertain about their orientation, autogynephilic—that is, only attracted to the image of themselves as females, or homoerotic homosexuals. Homosexual cross-dressers tend not to be as strongly fetishistic at puberty. Gay communities provide far greater acceptance for cross-dressing than does the dominant culture for heterosexual transvestites.

Several psychological hypotheses exist for understanding cross-dressing. The first is that the child, feeling deprived of adequate maternal nurturance and embarrassed about his continued longings for closeness to an unavailable person, invents a feminine self to provide for himself through his imagination and play activities. The second is that the child simply has subthreshold form of gender identity disorder manifested by subjective unmasculinity rather than behaviorly expressed wishes to be girl. The third is that he has concluded that girls have a better chance of being loved and he envies their opportunities to grow up to be helpful like mothers rather than frightening like fathers. The fourth is that the child has identified with his transvestitic father.

Socially, dual-role transvestites, transgenderists, and transsexuals come together in organizations that are formed to provide environments that enable these men to appear as women. These cross-dressers network, build relationships, and learn more about their life possibilities at regional and national meetings. Participants have occasionally been the object of questionnaire studies and are the source of information from nonpatient transvestites. These samples raise the possibility that there are some transvestites who are not paraphiliac because they sense no occupational, relationship, or other social impairment. Such studies also point to the fact that the gratifying aspects of networking and displaying the self as a female represent a good adaptation to this proclivity. It contrasts to those socially isolated ashamed men who cross-dress in private and who remain mortified at the possibility of anyone knowing about their situation. Table 19.2-2 lists the DSM-IV diagnostic criteria for tranvestic fetishism.

A. Over a period of at least 6 months, in a heterosexual male, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving cross-dressing.
B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify if:
With gender dysphoria: if the person has persistent discomfort with gender role or identity.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.2-2 DSM-IV Diagnostic Criteria for Transvestic Fetishism

Fetishism Unlike transvestic fetishism where the man imagines himself as a woman, simpler forms of fetishism focus on an item of clothing (e.g., shoes, gloves, stockings). The fetish is linked to someone closely involved with a patient during childhood and has a quality associated with this loved, needed, or even traumatizing person. Usually, the disorder begins by adolescence, although the fetish may have been established in childhood.

Sexual activity may be directed toward the fetish itself (for example, masturbation with or into a shoe), or the fetish may be incorporated into sexual intercourse (e.g., the demand that high-heeled shoes be worn).

Table 19.2-3 presents the DSM-IV diagnostic criteria for fetishism.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving the use of nonliving objects (e.g., female undergarments).
B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The fetish objects are not articles of female clothing used in cross-dressing (as in transvestic fetishism) or devices designed for the purpose of tactile genital stimulation (e.g., vibrator).

Table 19.2-3 DSM-IV Diagnostic Criteria for Fetishism

The questionnaire responses of 262 of 1000 gay and bisexual men who belonged to a national organization, the Foot Fraternity, provides a nonclinical view of this pattern of arousal to clean feet, boots, shoes, and smelly socks. About half of the respondents were also aroused by clothes. About 75 percent were particular about the type of footwear—wingtips versus penny loafers, for example—but their preferences changed over time. Over half wanted the footwear to have been worn by someone so that they could fantasize about the man whose feet had been in the shoes. One fifth of the respondents had no idea why feet or shoes aroused them; 50 percent pointed to the sensual qualities of feet—smell, taste, sight, and touch and fondly remembered being tickled as children or teenagers on the foot; 30 percent emphasized that the shoe symbolized the essence of masculinity—power, strength, success, and youth. Partner play was much less frequent than masturbation. Seventy-seven percent had foot play with a partner in the past year and 66 percent had sex with a partner without foot play. For 75 percent, foot play with a partner was an important part of sex with partners. About 66 percent of the foot play activities were part of a larger interest in sadomasochistic relationships—dominance and submission games—but for the rest, massaging, caressing, kissing, and licking without any role playing enhanced arousal.

Seventy percent of the respondents were not socially isolated from friends within the gay subculture. The Foot Fraternity helped fetishists to meet one another socially and for foot play, to learn more about how to enjoy feet, and to increase self-acceptance. The authors concluded that most of the respondents would not meet DSM-IV criteria for paraphilia.

Sexual Masochism Table 19.2-4 lists the DSM-IV diagnostic criteria for sexual masochism. Masochism occurs in both sexes, but has been thought to be more common among men. In questionnaire studies of masochistic women, about one fourth had been prostitutes; the majority of others had had large numbers of sexual partners. Respondents of both sexes can be subgrouped into those who simply prefer submissiveness, those who prefer the dominant role in sadomasochistic behavior, and those who enjoy playing both roles. Masochism's mild behaviors, such as the desire to be spanked, and its more dangerous forms, such as whipping or urethral insertions, are known to exist in individuals of every orientation. The persistence of masochistic behaviors throughout the life cycle, although possibly frequent, should not be presumed. Masochism tends to coexist with various forms of fetishism in both sexes and is part of the evolution of paraphiliac forms. Some masochists advertise, join clubs, frequent particular bars, or purchase sexual services to find the right partner. Whereas the sadist is actually dominating, humiliating, or causing physical pain, the masochist typically thinks that he or she is actually controlling the abusive episode. During sexual games, some sadomasochists are aware that while dominating they are identifying with someone who hurt them as a child. “I sometimes know that I am beating her like my mother beat me. After hurting her for a while, I identify with her. I feel love for her because she knows how I felt as a child. I then asked to be beaten. In the delicious experience of pain, I feel loved and loving... Most of the time I don't feel that good about myself or her.” Although studies of nonpatient devotees of masochism suggest that many do not consider that their sexual patterns impair their nonsexual lives, clinicians may see those who have been impaired.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving the act (real, not simulated) of being humiliated, beaten, bound, or otherwise made to suffer.
B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Table 19.2-4 DSM-IV Diagnostic Criteria for Sexual Masochism

A 53-year-old, house-bound man with agoraphobia who had been married four times sought help for a recent loss of control over his usual need for foot fetishism and fantasies of humiliation and domination. He claimed that his masochistic hunger dominated all his waking moments. He recalls being sexually aroused by women's feet since age 4. When his mother sat in a chair and placed her sole on his face, he developed an erection. He spent many of his childhood hours orchestrating games in which babysitters and neighbors placed their feet on his face. As an adult, he says he likes to be degraded, humiliated, and owned by a woman who leaves him with no decision-making capacities. He is excited by leather and the smell of a foot—particularly by being forced to endure the disgusting odor from a foot. His new wife's insistence that he seek help came after he sought total domination from her and told her that “I want to be broken, with no way out, then be used as a slave for sex and to cook and clean. I even want to watch you have sex with another man.” She threatened to leave with their 4-year-old daughter and insisted that he seek professional help. When his first wife refused such behavior, he had an affair with a woman who whipped him. He and his second wife were “into bondage” until he became so obsessed that she became uncomfortable. His third marriage at age 50 was to a very aggressive thief who sexually dominated him but did not care about him outside their bedroom dramas. In order to conduct intercourse during the last decade, he has had to fantasize about being dominated. He has masturbated at least daily since age 17 and continues to attempt to regulate his frequent sexual tensions 3 to 4 times per day via masturbation or intercourse with his wife.

A 36-year old man with a lifelong gender identity disorder was sexually abused for several years by his stepfather beginning at age 12. Ever since he has been erotically excited by being ordered to do the sexual bidding of men. Such passivity increased his sense of his femininity. For several years in his late 20s, he and his wife ran a sex business that used a photoadvertisement that depicted him as a boy who dresses as a woman who is interested in submission. His preferred sexual activity with clients was intracrural intercourse because that is what his stepfather did before he hugged him, held him, and told him that he loved him. His masochism, however, did not seem to be restricted to his sexual life. Now in prison for armed robbery and being a fugitive from the criminal justice system, his life story seemed to contain repeated episodes of dramatically bad judgment. He might someday be described as having a self-defeating personality disorder.

Autoerotic Asphyxia This form of masochism is characterized by the use of self-strangulation up to the point of loss of consciousness in order to enhance sexual arousal. Oxygen deprivation is brought about by the use of a noose, ligature, a plastic bag, mask, a volatile nitrite, or chest compression. Death may occur because of equipment failures, planning errors, or suicide. Most of the 250 to 1000 deaths per year in the United States are thought to be accidental. Experiments that have shown that as little as 7 pounds of pressure on the carotid arteries can cause unconsciousness in 7 seconds. Males die of asphyxia 25 times more often than females. Deceased men are often found cross-dressed, with dildos in the anus or objects that cause pain to the nipples or genitals, or surrounded by pornography. Adolescent victims are thought to be more often accidental; older persons are more often thought to be suicides. Some couples practice asphyxia (hypoxyphilia) in pairs. Sexual asphyxia has been known for centuries in both the eastern and western worlds. A study of five adolescent practitioners of this elaborate masturbatory ritual found that each had an early history of being choked by parents and other episodes of physical or sexual abuse. Their developmental histories were abject.

Sexual Sadism Table 19.2-5 lists the DSM-IV diagnostic criteria for sexual sadism. Sexual sadists are most commonly encountered by forensic mental health professionals after the person's sexual cruelties have become known. Rape or sexual assault of a nonconsenting adult is the usual issue. Women are by far the most

common victims. Coerced vaginal, oral, or anal penetration may be brought about by various degrees of verbal, physical, and psychological threats and humiliation. The rapist may be a stranger, an acquaintance, a separated or exspouse, or a spouse batterer. The mental health professional may be called upon to distinguish between paraphilic and power-aggression rape. This judgment is typically made by considering three questions: (1) Does the man fulfill the 6-months duration criterion for paraphilia? (2) Has the man had other paraphiliac manifestations such as pedophilia, exhibitionism, voyeurism, or frotteurism? (3) If physiological testing is done, does he show a maximum arousal to violent sexual stimuli? Rapists are a heterogeneous group. Estimates of the frequency of paraphilia among rapists vary widely, beginning at only 5 percent.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving acts (real, not simulated) in which the psychological or physical suffering (including humiliation) of the victim is sexually exciting to the person.

B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.2-5 DSM-IV Diagnostic Criteria for Sexual Sadism

Sexual sadism is not restricted to rape, however. A large minority of college-aged men in numerous studies admit to being aroused by and fantasizing about coercive sex with strangers, including rape. The fantasies often end with the woman's enjoyment. A smaller but sizable minority of women admit to being aroused by fantasies of being coerced into sex until they enjoy the sexual activity.

Occasionally sexual sadism other than rape becomes obvious in a general psychiatric setting.

A massively obese 28-year-old rarely employed man with a high school education sought help for a chronic depressive disorder and revealed that he spent most of his days masturbating to the image of hurting prostitutes. He claimed to reach as many as 12 orgasms per day as he thought about torturing, cutting, and killing these women. Also having a stocking fetish, the sight of women's legs often mesmerized him. While he claimed that his childhood was deeply unhappy, he generally refused to discuss it preferring never to remember his critical family members.

A controlling, narcissistic physician, raised alone by his widowed mother since age 2, has been preoccupied with spanking's erotic charge for him since age 6. Socially awkward during adolescence and his 20s, he married the first woman he dated and gradually introduced her to his secret arousal pattern of imagining himself spanking women. Although horrified, she episodically agreed to indulge him on an infrequent schedule to supplement their frequent ordinary sexual behavior. He ejaculated only when imagining spanking. Following her sixth episode of anxious, sullen, depression in 20 years of marriage, her psychologist instructed her to tell him, “No more.” He fell into despair, was diagnosed with a major depressive disorder, and wrote a long letter to her about why he was entitled to spank her. He claimed to have had little idea that her participation in this humiliation was negatively affecting her mental health (“She even had orgasm sometimes after I spanked her!”). He became suicidal as a solution to the dilemma of choosing between his or her happiness and that what he was asking was evil. He was shocked to discover that she had long considered suicide as a solution to her marital trap of loving an otherwise good husband and father who had an unexplained sick sexual need.

Pedophilia Of the terms used to describe those who have genital contacts with young adolescents and children who are at least 5 years their junior— *child molester, incest offender, sex offender, child rapist, statutory rapist, sexual abuser, sexual deviant, serial child molester, ephebophile*, and *pedophile*—only the latter is a formal diagnosis. Its criteria, however, do not cover all those who are arrested for child molestation. The term “pedophile” has two meanings: (1) a person who fulfills the DSM-IV criteria (Table 19.2-6) for the diagnosis and (2) a person with a lifelong erotic preference for children.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving sexual activity with a prepubescent child or children (generally age 13 years or younger).

B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The person is at least 16 years and at least 5 years older than the child or children in criterion A.
Note: Do not include an individual in late adolescence involved in an ongoing sexual relationship with a 12- or 13-year-old.

Specify if:
Sexually attracted to males
Sexually attracted to females
Sexually attracted to both

Specify if:
Limited to incest

Specify type:
Exclusive type (attracted only to children)
Nonexclusive type

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.2-6 DSM-IV Diagnostic Criteria for Pedophilia

Child molesters are subdivided in four ways. First, by age of the perpetrators: (1) A large minority of molestations are committed by adolescents seemingly in their attempt to learn about sex with naive and accessible children; (2) young and middle-aged adults who molest their own children, stepchildren, or friends' children; and (3) elderly men who may be organically impaired, isolated, or lonely. Second, by sex of the offender: Females are found to molest children far less frequently than males. When they are discovered, they are often deteriorated by substance abuse, major mental illness, major organic illness, or thought to be coerced into the behavior by a dominant man. However, many adolescent sex offenders claim to have been abused by women in their prepubertal period. Third, by orientation: Men who have been sexually aroused by children all their lives have the largest numbers of child victims. The most destructive of these individuals find work or recreational opportunities to be in the company of children and young adolescents. They are synonymously labeled as having an anomalous orientation or a preference for children or as fixated pedophiles. Pedophiliac orientation is often seen in pure form. When it coexists with an adult heterosexual or homosexual orientation, the person may be called a regressed pedophile who under extenuating circumstances has sex with children. Fourth, by victim characteristics: (1) boys versus girls (androphilic versus gynephilic); (2) infants and toddlers, preschoolers, grade schoolers, or adolescents; and (3) indiscriminant. Although these four categories are not mutually exclusive, they suggest that children are molested under such a wide array of circumstances that clinicians should not seek a single explanation for pedophiliac offenses.

Within each of these categories, however, child molesters display a startling set of rationalizations for their behavior. These include: (1) “I was educating the child about sex to learn the right way”; (2) the child initiated the behavior; (3) the child readily consented to it; (4) it was merely a caring act to provide pleasure; (5) it caused no harm whatsoever; and (6) it was less problematic than having an affair with an adult. These notions are referred to as cognitive distortions that demonstrate the molesters' lack of empathy for their victims. While children of many ages are subjectively attractive to many adults who never behave sexually with children, the compelling question about child molesters is what accounts for their behavior. Since the answer seems to vary significantly from case to case, the clinician should formulate a sophisticated answer with each perpetrator.

A 67-year-old man with a doctorate degree whose aspiration for entrepreneurial success remained undaunted despite his lifetime of failures touched his 9-year-old granddaughter's genitals under the guise of teaching her what others should never do to her. The visibly disturbed child told her mother who quickly called a family meeting to expose a previous secret she and her father had maintained: between the ages of 13 and 16 her father had sexual intimacies with her and told her it would not be good for her depressed mother to know this. Her younger sister too had been approached when she was pubescent, but she had strongly rebuffed the father. The man described that he and his beloved granddaughter were having their usual warm affectionate time reading a story together when for “no good reason” he began educating her. He had not been erotically preoccupied with her; he was just in his ordinary state of sexual and financial frustration. Although potent, he had no interest in sex with his unemployable dysthymic wife who could not stop complaining about her husband's vocational failures and their serious financial problems.

A 62-year-old married janitor had worked as a fourth-grade school teacher for 26 years before he transferred school districts and finally several years later mysteriously lost his second job. He was referred for help after his family discovered that he had repeatedly fondled the genitals of his 4- and 6-year-old granddaughters. A father of five who had not had sex with his wife for 30 years after strenuously objecting to her cigarette smoking, he was generous, helpful, and cooperative with his children and grandchildren. Intellectually slow, he preferred comic books and had a charming manner of playing with young children “like he was one himself.” By his estimate he had touched the buttocks and genitals of at least 300 girl students, thinking only of how they did not know what he was doing because he was being affectionate and they were too young to realize what was happening. He loved the anticipation and excitement of this behavior. His teaching career ended when parents complained to a principal. The principal discovered that the new teacher had been transferred, not fired, from his longstanding teaching job for the same reason. The patient had tried to touch his 12-year-old daughter who angrily warned him to stay away from her, but he had also managed to touch her girlfriends and his best friend's daughters as they neared puberty.

An ebullient 60-year-old rural clergyman was arrested after parental complaints of his inappropriate behaviors within a youth program. He recently had confessed his love for a 14-year-old boy in his parish. Limited by an osteoarthritic hip, widowed for 10 years from a woman with multiple sclerosis whom he married when he was 35 years old he learned to deal with his social isolation and loneliness by creating programs for seventh-grade boys. In the last decade he made friends with adult homosexual men and occasionally had sexual relationships with them. These typically ended when he changed churches. Between these more satisfying adult relationships, he provided fellatio to at least 30 teenagers in three different church settings. He offered them these services “when your girlfriend was not being friendly to you.” He admired the boys' masculinity, which contrasted with his own sense of effeminacy. “They were normal guys and I was just a momma's boy who loved to perform musical theater.” “But I did not count on falling in love with him. I guess I did because he was just like I would like to have been.”

Studies have suggested a correspondence between the victimizers' personality characteristics and the age of their victims. Pedophiles are described as immature with poor social skills, low self-esteem, insufficient self-assertiveness, and more signs of psychotic thinking whereas those who offend against adolescents are intermediate between the immaturity of pedophiles and the psychopathology of those who rape adults. Such findings, however, may be encompassed by other observations such as many pedophiles are males of low social status.

Exhibitionism Table 19.2-7 lists the DSM-IV diagnostic criteria for exhibitionism. The desire to display one's penis, erection, or naked body to a girl or woman who is a stranger is usually motivated by the hope of shocking or exciting the witness. This idea is so exciting that the behavior culminates in masturbation at the site of the exhibition or soon after fleeing to safety. When men exhibit themselves exclusively to girls, clinicians wonder whether this pattern is a form of pedophilia. Although exhibitionists account for approximately one third of arrests for all sexual crimes, the vast majority of indecent exposures are not reported to the police. Exhibitionists often seem socially unskilled, shy, and unable to negotiate a psychologically intimate relationship with an adult woman that includes regular sexual intercourse. They are commonly regarded as harmless hands-off offenders, but some are involved in aggressive sexual and nonsexual assaults. A large minority of exhibitionists report cross-dressing; a smaller percentage also are voyeurs.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving the exposure of one's genitals to an unsuspecting stranger.
B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.2-7 DSM-IV Diagnostic Criteria for Exhibitionism

A divorced 62-year-old schizoid, chronically depressed radiologist who had overcome an opiate addiction was asked to retire after being arrested in his community for walking naked in a public place at night. Since early adolescence, he has had a fantasy of walking naked in public. The fantasy only occasionally was acted out except that he did often walk naked on his property after dark and masturbated in the woods to the image of being seen exhibiting himself. Considered to be competent in his field, he never could explain to himself what “comes over me and causes me to want to do that,” but “it is just the way I am.”

Voyeurism Table 19.2-8 lists the DSM-IV diagnostic criteria for voyeurism. Voyeurism is a pathological manifestation of the widespread arousal responses that men and women have at the sight of other people having sex. Voyeurs' erotic lives are dominated by surreptitious watching, their sexual behavior is masturbatory, and is not used to enhance a sexual relationship with a socially appropriate partner. Secretly watching another undressing, showering, masturbating, or having sex tends to consume the person, undermine other sexual behaviors, and eventually get the person into marital, familial, or legal difficulty. Most incidents of voyeurism are undetected because victims do not realize that they have been spied upon. When voyeurs are detected, however, they are only rarely found to be purely voyeuristic; about three quarters of voyeurs exhibit themselves and a minority cross-dress.

A. Over a period of at least 6 months, recurrent, intense sexually arousing fantasies, sexual urges, or behaviors involving the act of observing an unsuspecting person who is naked, in the process of disrobing, or engaging in sexual activity.
B. The fantasies, sexual urges, or behaviors cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.2-8 DSM-IV Diagnostic Criteria for Voyeurism

A 52-year-old married sociopathic chronic alcoholic nearly illiterate contractor had been hospitalized 12 times for binge drinking before he revealed his pattern of exhibitionism. A more careful history revealed a 30-year history of 2 to 5 exposures per week, repeated searches for windows into which to peek, frequent date rape and coercive sex with women he met in bars, and episodes of cross-dressing. Most recently he felt he was losing control of his temptations to touch his daughters' girl friends. Arrested only once for a sexual crime, his voyeuristic preoccupations and his genital exposures to young teenage girls make him fear that he will jeopardize his marriage, family, and business because of greater risk taking in recent years. “I feel like I am a boy in a man's body.”

Paraphilia Not Otherwise Specified The classification of paraphilia not otherwise specified includes varied paraphilias that do not meet the criteria for any of the aforementioned categories (Table 19.2-9).

This category is included for coding paraphilias that do not meet the criteria for any of the specific categories. Examples include, but are not limited to, telephone scatologia (obscene phone calls), necrophilia (corpses), partialism (exclusive focus on part of body), zoophilia (animals), coprophilia (feces), klismaphilia (enemas), and urophilia (urine).

Table 19.2-9 DSM-IV Diagnostic Criteria for Paraphilia not Otherwise Specified

Telephone and Computer Scatologia Telephone scatologia is characterized by obscene phone calling and involves an unsuspecting partner. Tension and arousal begin in anticipation of phoning; the recipient of the call listens while the telephoner (usually male) verbally exposes his preoccupations or induces her to talk about her sexual activity. The conversation is accompanied by masturbation, which is often completed after the contact is interrupted.

People also use interactive computer networks, sometimes compulsively, to send obscene messages by electronic mail and to transmit sexually explicit messages and video images. Because of the anonymity of the users in chat rooms who use aliases, on-line sex (cybersex) allows some people to play the role of the opposite sex (“genderbending”), which represents an alternative method of expressing transvestite or transsexual fantasies. A danger of on-line cybersex is that pedophiles often make contact with children or adolescents who are hired into meeting and then molested. Many on-line contacts develop into off-line liaisons. While some people report that the off-line encounters develop into meaningful relationships, most such meetings usually are characterized by disappointment and disillusionment. In other situations, when adults meet, rape or even homicide may occur.

Necrophilia Necrophilia is an obsession with obtaining sexual gratification from cadavers. Most people with this disorder find corpses in morgues, but some have been known to rob graves or even to murder to satisfy their sexual urges. In the few cases studied, those with necrophilia believed that they were inflicting the greatest conceivable humiliation on their lifeless victims.

Partialism People with this paraphilia concentrate their sexual activity on one part of the body to the exclusion of all others. Mouth-genital contact—such as cunnilingus, fellatio, and anilingus—is normally associated with foreplay. But when a person uses these activities as the sole source of sexual gratification and cannot have or refuses to have coitus, a paraphilia exists.

Zoophilia In zoophilia, animals—which may be trained to participate—are preferentially incorporated into arousal fantasies or sexual activities, including intercourse, masturbation, and oral-genital contact. Zoophilia as an organized paraphilia is rare. For many people, animals are the major source of relatedness, so it is not surprising that a broad variety of domestic animals are sensually or sexually used.

Sexual relations with animals may occasionally be an outgrowth of availability or convenience, especially in parts of the world where rigid convention precludes premarital sexuality and in situations of enforced isolation. Because masturbation is also available in such situations, however, a predilection for animal contact is probably present in opportunistic zoophilia.

Coprophilia and Klismaphilia Coprophilia is attraction to sexual pleasure associated with the desire to defecate on a partner, to be defecated on, or to eat feces (coprophagia). A variant is the compulsive utterance of obscene words (coprolalia). These paraphilias are associated with fixation at the anal stage of psychosexual development. Similarly, klismaphilia, the use of enemas as part of sexual stimulation, is related to anal fixation.

Urophilia Urophilia, is interest in sexual pleasure associated with the desire to urinate on a partner or to be urinated on. In both men and women, the disorder may be associated with masturbatory techniques involving the insertion of foreign objects into the urethra for sexual stimulation.

PATHOLOGY AND LABORATORY EXAMINATION

Newer diagnostic probes of brain function have not yet been applied to paraphiliac problems. However, using the clinical history and neurological examination, several investigators have suggested a higher-than-expected frequency of soft neurological signs and learning disabilities among sex offenders. Technology for assessing paraphilia has largely focused on pedophilic offenders, rapists, and juvenile offenders. Penile plethysmography measures penile circumference while subjects are exposed to audiotapes, slides, or videotapes of paraphiliac and nonparaphiliac stimuli. Penile enlargement data provides the clinician with objective evidence of arousal to a specific stimuli—for instance, children—to counter the person's denial that he is not sexually interested in children. Plethysmography is limited by the ability of some perpetrators to enhance their arousal to nonparaphiliac stimuli, to suppress their arousal to children, and by the fact that arousal to images of children does not mean that the person behaves sexually with children. Some men demonstrate limited or no arousal in the laboratory setting. A proprietary system with scientifically definable and acceptable limits, Abel Screening, can detect pedophiliac interest without measuring penile circumference changes. Some centers use the polygraph for initial and subsequent assessments. None of these approaches, however helpful at times, can alone resolve the issue of a person's guilt or innocence. Pedophiliac arousal in the laboratory can be highly useful in confronting a person who denies any interest in children and may trigger a more truthful process between the clinician and the accused. Laboratory assessments are generally not considered admissible evidence in courts.

DIFFERENTIAL DIAGNOSIS

Commonly made mistakes in the evaluation of the paraphilias stem from failure to take a comprehensive sexual developmental and psychiatric history. This is prevented by taking enough time for evaluation, tracing the person's sexual behavior and fantasies from childhood to the present, asking about unmentioned categories, and thoroughly investigating comorbid conditions. A large percentage of patients will not be entirely truthful about their sexual behaviors at first. Since paraphilias are frequently comorbid with a wide variety of psychopathologies, hypomanic episodes, substance intoxication, and dementia or organic brain disease should be given serious consideration. However, it is safer to think about them as releasing factors than to ignore the Axis I diagnosis of paraphilia. Some brain diseases may release behaviors that at first glance seem paraphiliac, for instance, postictal undressing, but are not part of a longstanding pattern of arousal.

COURSE AND PROGNOSIS

When paraphiliac behaviors appear to be released during new pathology, for example, tumor, epilepsy, intoxication, or social or psychological deterioration due to schizophrenia or chronic substance abuse, it is useful to recall that abnormal intentions are developmentally constructed, usually manifested in childhood or shortly after puberty and managed by internal mental and social control mechanisms. What is so dangerous about the chronic problem of paraphilias, particularly paraphilias that victimize others, is that the mental homeostasis that keeps the impulses unconscious or behaviorally suppressed may fail under conditions that are idiosyncratic

to the patient or that are generally understood as stressful. Thus, clinicians need to be wary whenever they find themselves thinking that the “problem is resolved”—unless they add, “for now.”

While the diagnosis of a paraphilia carries a pessimistic prognosis when the person remains untreated, a more realistic view of its natural history may be approximated by returning to the parameters of severity and the etiological concepts. Many paraphiliacs are highly damaged persons, others seem to have more psychological capacities to use what therapists have to offer. Thus, each case must be comprehensively assessed beyond what is necessary to make a formal diagnosis.

TREATMENT

Five types of psychiatric interventions are used to assist the paraphilia patient to rebalance internal control mechanisms, cease victimization of others, and enhance the capacities to relate to others: external control, reduction of sexual drives, treatment of comorbid conditions, cognitive behavioral therapy, and dynamic psychotherapy. The art of therapy is to select and modify these various elements for the individual patient.

Provision of External Controls When sexual victimization of others has occurred, new external controls should be instituted. All relevant persons in the environment need to know what the person has done and is capable of doing again under opportune conditions. For intrafamilial abuse of children, for instance, the adults and other children in the family are informed of the abuse. The children in the family are not permitted to be alone with the offender again as long as they are unable to adequately protect themselves. With such controls implemented, those who victimize girls in their own home have recidivism rates of less than 5 percent. The more socially, economically, and psychologically disruptive alternative is to remove the offender from the home. When professionals who have sexually offended return to their work settings, external controls take the form of restricting access to potential victims and monitoring by administration, co-workers, and sometimes by patients with the use of exit questionnaires. Perpetrators and their professional colleagues can be educated about boundary crossings. When a therapist agrees to keep the knowledge of the patient's victimizing sexual patterns from family members, administrators, or colleagues on the basis of confidentiality, the single most effective psychiatric intervention is removed from the treatment plan. The therapist will probably be the last to know of the resumption of victimizing behaviors.

Prison is an external control mechanism for sexual crimes, but it is not part of rehabilitation; it is designed for public safety and punishment. Many states are considering instituting laws “Megan's Law” that would inform relevant persons in the community when a person who has served time for extrafamilial child abuse was about to be released from prison. The tension between the constitutional rights of privacy for the sex offender and rights of the community to protect itself has yet to be resolved in a legally satisfactory manner.

Reduction of Sexual Drives Several medications can be used to reduce sexual drive for those assessed to have lost their self-control or autonomy. Each medication is relatively safe in physically healthy men and is effective and reversible. The administration of medroxyprogesterone (Provera) to males in dosages ranging from 20 to 120 mg by mouth daily or 400 to 600 mg intramuscularly a week (intramuscular treatment has considerably more adverse effects) may within 96 hours begin to reduce the drive to behave sexually. The reduction is reflected in lower rates of masturbation and partner sex, less intense erotic preoccupations, a greater ability to concentrate on nonsexual matters, and an increase in calmness. Medroxyprogesterone typically reduces serum testosterone to subnormal concentrations of 100 to 250 µg/dL, which is usually sufficient to bring about the desired effect. Medroxyprogesterone decreases sperm production, which may require 10 months for total recovery. Outside the United States cyproterone is used for this purpose. The pretreatment serum testosterone concentrations of male paraphilia patients are normal not supranormal. These compounds empirically assist men who feel out of control sexually to gain more autonomy. They may exert their beneficial effects by decreasing testosterone synthesis, increasing testosterone reductase activity, displacing testosterone from its plasma binding protein, and blocking cellular membrane receptors and intracellular metabolism. The SSRIs are more commonly used to reduce the craving for sexual expression and in one open-label trial helped 70 percent of men with paraphilia or paraphilia-related conditions. No large-scale double-blind conditions studies of the SSRIs have been done, however, and no comparative studies of testosterone-lowering drugs and the SSRIs have been published. Other more expensive drugs, such as a once-monthly dosage of leuprolide (Lupron), may achieve the same testosterone-lowering effect, but generally are unnecessary.

Treatment of Comorbid Conditions Psychiatrists need to consider the role of inadequately treated comorbid states when planning to treat sexual compulsivity or impulsivity. Alcohol and substance abuse, major depressive disorder, grief, psychotic disorder, attention-deficit/hyperactivity disorder, bipolar II disorder, and others each may be the co-factor that enables a compensated sexual pattern to deteriorate and come to clinical attention.

Cognitive-Behavioral Therapy It is frequently observed that sex offenders lack the social skills necessary to live effectively and create nonproblematic sexual relationships. Correcting some of these deficits is a goal of most cognitive-behavioral treatment programs for sexually offending paraphiliac men. Each intervention is an aspect of a therapy approach that assumes that a paraphiliac lifestyle is learned and can be significantly modified. The specific techniques can be done in individual or group settings. They include: role playing how to initiate and maintain conversations *(social skills training)*; role playing for acceptable methods of self-assertion *(assertiveness training); sex education*; confrontation about the rationalizations used to support victimization of others *(cognitive restructuring)*; learning about the feelings and adverse consequences experienced by victims *(victim empathy)*. Other more complex overlapping techniques deal directly with deviant arousal: *imaginal desensitization* teaches the patient to use progressive muscular relaxation when feeling anxious from resisting the paraphiliac impulse; *relapse prevention* by identifying the triggers to paraphiliac urges and teaching new social and psychological coping strategies; *modified aversive behavioral rehearsal*, the most ambitious technique, in which the offender demonstrates his problematic sexual pattern—cross-dressing, exhibitionism, or pedophilia with a mannequin—while being videotaped and in the presence of a therapist and other offenders. The audience asks questions about the perpetrator's feelings, thoughts, and motives and repeatedly confronts the cognitive distortions and lack of victim empathy. Work on victim empathy is predicated on the assumption that a large minority of perpetrators have been victims of childhood sexual trauma. The cognitive-behavioral approach is an integrated treatment approach that can be tailored for the individual patient; it readily incorporates 12-step programs for alcoholism or sex addictions into the treatment plan.

Dynamic Psychotherapy The nonviolent paraphilias and the paraphilia-related disorders are often treated with traditional individual or group therapies using a combination of supportive, growth-promoting tactics. Such therapies aim at creating an evolving hypothesis about the unique developmental origin of the patient's eroticism. The defensive function of the impulse to act out (the anxiety reduction function) is defined so that the person can deal directly with the unpleasant feelings that trigger the impulses. Hypotheses get more complex as more developmental details emerge. The vast majority of patients need many therapeutic opportunities to work through the paraphiliac defense so that they can simply feel and cope with their particular life issue. Dynamic psychotherapy approaches emphasize the importance of the trusting relationship to the therapist to enable the work to occur. Therapists often face a delicate situation because most paraphilia embodies the capacity to transform the pain of childhood misery—for instance, “Mother didn't love me” into the pleasure of sexual arousal—“I'll exhibit myself to a teenager.” Although the patient may no longer want to act out sexually, his desire to not re-experience his painful past may be greater. When the therapist decides to assist the patient in working through some of his painful memories, a great deal of support may be needed. Generally speaking the more globally impaired the paraphiliac is, the less likely it is that he or she can be treated in a dynamic psychotherapy.

Negative Countertransference Therapist hostility, disgust, and fear are important obstacles to providing care for paraphilia patients. Their presence is often manifested by resistance to initially seeing a patient, treating the comorbid symptoms and ignoring the paraphilia in the hope that disorder of intention will spontaneously remit, or quickly referring the patient. Many therapists tell their patients that they do not know anything about this subject. The patient's attempts to get help are further frustrated by the fact that many insurance contracts will not cover a paraphilia diagnosis. Negative countertransference reactions should be defined, discussed, and thoughtfully considered so that the therapist can be objective and discover the patient's developmental issues and attempt to ameliorate them. The problem is that many therapists are so offended by the victimization of others that they refuse to learn about the patient's life experiences and avoid abandoning their assumption that perpetrators are “simply evil through and through.”

Ideological Synthesis As experience with paraphiliac persons has accumulated since the 1970s the behavioral, psychodynamic, and nosological camps have come together in agreement about the profound deficits observed, the formidable therapeutic tasks, and the social problems that these disorders create.

Treatment Results The results of various interventions have profound methodological limitations that rest upon the many distinctions developed in this chapter. For sex offenders, the parameter of success that is of greatest interest is the lack of recurrence of violence towards others, not the disappearance of paraphiliac eroticism in itself. Behavior therapist define outcomes and constantly evolve new treatment approaches for offenders. With selected groups of rapists treated with cognitive restructuring and procedures to enhance victim empathy, for example, the recividism rates in some hands are less than 10 percent. Despite periodic reports of therapy success, however, the impression exists that good results are most often attained when a single paraphilia is present, a successful adult attachment has been made that supports nonparaphiliac behaviors, an intelligence quotient (I.Q.) above 70; nonsexual antisocial personality traits are absent; alcohol and drug problems are absent; rationalizations of the crimes are not tenacious; and the degree of paraphiliac arousal on physiological testing is not great. When incest offenders are

considered, the prognosis is better if intercourse with the child has not occurred and there is no evidence of an exclusive attraction to children.

Centers that provide comprehensive evaluation tend to use multiple modalities of therapy. Psychodynamically oriented clinicians do not generate outcome data, but are anecdotally known to assist people in various ways: (1) to act out less frequently and to deal with life stresses more competently; (2) to enhance their capacity to relate nonsexually and sexually to a life partner; and (3) to accept their eroticism with more comfort and seek the company of others with similar erotic interests to enhance general adjustment—for example, joining a support group for cross-dressers. Regardless of the therapy methods used, the data available from intense research efforts, or the ideological preferences of the therapist, paraphiliac problems are recognized to be serious chronic maladaptive attempts to relate to others.

SUGGESTED CROSS-REFERENCES

Gender identity disorder is discussed in Section 19.3, personality disorders in Chapter 24, and sexual dysfunctions in Section 19.1b. The discussions of pervasive developmental disorders in Chapter 38 and separation anxiety disorders in children in Section 46.3 are also relevant. Child abuse is covered in Section 49.4.

SECTION REFERENCES

*Abel GG, Becker JV, Cunningham-Rathner J, Rouleau JL, Murphy WD: Self-reported sex crimes of nonincarcerated paraphiliacs. J Interpers Violence *2:* 3, 1987.

Abel GG, Osborn CA: Behavioral therapy treatment for sex offenders. In *Sexual Deviation*, ed 3, I Rosen, editor. Oxford University Press, London, 1996.

Abel GG, Rouleau JL: Sexual abuses in clinical sexuality. Psychiatr Clin North Am *18:* 139, 1995.

Briere J, Malamuth NM: Self-reported likelihood of sexually aggressive behavior: Attitudinal versus sexual explanations. J Res Pers *17:* 315, 1983.

Black DW, Kehrberg LLD, Flumerfelt DL, Schlosser SS: Characteristics of 36 subjects reporting compulsive sexual behavior. Am J Psychiatry *154:* 243, 1997.

Brown GR, Wise TN, Costa PT, Herbst JH, Fagan PJ, Schmidt CW: Personality characteristics and sexual functioning of 188 cross-dressing men. J Nerv Ment Dis *184:* 265, 1996.

Chivers M, Blanchard R: Prostitution advertisements suggest association of transvertism and masochism. J Sex Marital Ther *22:* 97, 1998.

Coleman E: Is your patient suffering from compulsive sexual behavior? Psychiatr Ann *22:* 320, 1992.

Epstein RS, Simon RI: The Exploitation Index: An early warning indicator of boundary violations in psychotherapy. Bull Menninger Clin *54:* 450, 1990.

Feierman JR: Pedophilia: Paraphilic attraction to children. In *The Handbook of Forensic Sexology*, JJ Krivacska, J Money, editors. Prometheus Books, Amherst, NY, 1994.

Finkelhor D, Araji S: Explanations of pedophilia: A four-factor model. J Sex Res *22:* 145, 1987.

Friedrich WN, Gerber PN: Autoerotic asphyxia: The development of a paraphilia. J Am Acad Child Adolesc Psychiatry *33:* 970, 1974.

*Gijs L, Gooren L: Hormonal and psychopharmacological interventions in the treatment of paraphilias: An update. J Sex Res *33:* 273, 1996.

Greenberg DM: Sexual recidivism in sex offenders. Can J Psychiatry *43:* 459, 1998.

Haywood T, Cavanaugh JL: Sexual deviancy. Curr Opin Psychiatry *9:* 384, 1996.

Johnstone J, Hews R: Autoerotic asphyxia: A case report. J Sex Marital Ther *23:* 326, 1997.

Kalichman SC: Psychopathology and personality characteristics of criminal sex offenders as a function of victim age. Arch Sex Behav *20:* 187, 1991.

*Kafka MP, Prentky RA: Preliminary observations of DSM-II-R Axis I comorbidity in men with paraphilias and paraphilia-related disorders. J Clin Psychiatry *55:* 481, 1994.

Kafka MP, Prentky RA. Attention-deficit/hypersensitivity disorder in males with paraphilia & paraphilia related disorders: A co-morbidity study. J Clin Psychiatry *59:* 388, 1998.

Kafka MP: Sertraline pharmacotherapy for paraphilias and paraphilia-related disorders: An open trial. Ann Clin Psychiatry *6:* 189, 1994.

Kafka MP: Hypersexual desire in males: An operational definition and clinical implications for males with paraphilias and paraphilia-related disorders. Arch Sex Behav *26:* 505, 1997.

Kravitz HM, Haywood TW, Kelly J, Wahlstrom C, Liles S, Cavanaugh JL: Medroxyprogesterone treatment for paraphiliacs. Bull Am Acad Psychiatry Law *23:* 19, 1995.

*Levine SB: *Sexual Life: A Clinician's Guide*. Plenum, New York, 1992.

Levitt EE, Moser C, Jamison KV: The prevalence and some attributes of females in the sadomasochistic subculture: A second report. Arch Sex Behav *23:* 465, 1994.

*Shaw JA, editor: *Sexual Aggression*. American Psychiatric Press, Washington, DC, 1999.

Weinberg MS, Williams CJ, Calhan C: Homosexual foot fetishism. Arch Sex Behav *23:* 611, 1994.

19.3 GENDER IDENTITY DISORDERS

RICHARD GREEN, M.D., J.D., AND RAY BLANCHARD, PH.D.

Cross-gender identity (gender identity contradicted by anatomical sex characteristics) in adulthood virtually always causes distress to the individual, and cross-gender identity in childhood represents a distinct risk of cross-gender identity in adulthood. Cross-gender identity at any age, therefore, is appropriately regarded as a disorder and a possible reason for clinical intervention. Many persons with gender identity problems—particularly those functioning well in other areas of life—dislike the classification of such problems as psychiatric disorders because they regard all psychiatric diagnoses as stigmatizing. For this reason, activists in the transsexual community have proposed alternative conceptualizations. These include the argument, intended to eliminate the perceived stigma of psychiatric disorder while preserving the possibility of third-party payment for sex-reassignment surgery, that transsexualism is a somatic rather than a psychiatric condition. This position cannot at present be justified on medical or scientific grounds. Therefore, the best solution to the problem of psychiatric nosology is to retain the classification of severe gender identity problems as psychiatric disorders while emphasizing, in clinical teaching, clinical care, and research writings, that gender identity disorders can be quite circumscribed in their effects, leaving the individual free to function in other areas of life at a satisfactory or even outstanding level.

DEFINITIONS

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) defines gender identity disorders as a heterogeneous group of disorders whose common feature is a strong and persistent preference for the status and role of the opposite sex. These disorders may be manifested verbally, in assertions that one properly belongs to the opposite sex, or nonverbally, in cross-sex behavior. The affective component of gender identity disorders is commonly referred to as *gender dysphoria*, which may be defined as discontent with one's biological sex, the desire to possess the body of the opposite sex, and the wish to be regarded as a member of the opposite sex. The extreme forms of gender identity disorders, collectively referred to as *transsexualism* in the third edition of DSM (DSM-III) and revised third edition (DSM-III-R), commonly involve attempts to pass as a member of the opposite sex in society and to obtain hormonal and surgical treatment to simulate the phenotype of the opposite biological sex.

HISTORY

Transsexualism became popularly known with the sex change of George Jorgensen into Christine Jorgensen in 1952. The 1966 professional book by Harry Benjamin, the pioneer who evaluated or treated hundreds of patients, and the introduction of sex-reassignment surgery at the Johns Hopkins Hospital in 1966 legitimized its treatment.

In 1960 a paper described behaviors in children that were consistent with the later-described gender identity disorder of childhood, and in 1974 a text was published that described several dozen boys with sexual identity conflict. Drawing on this and related work, the psychosexual disorders advisory committee for DSM-III introduced the diagnostic entity gender identity disorder of childhood in 1980.

COMPARATIVE NOSOLOGY

DSM Gender identity disorders first entered the American Psychiatric Association's official nomenclature in DSM-III, reflecting the growth of that area of human sexuality in psychiatry since the second edition of DSM (DSM-II).

Nosological Position of Gender Identity Disorders In DSM-III gender identity disorders were included in the category of psychosexual disorders along with the paraphilias and sexual dysfunctions. In DSM-III-R the gender identity disorders were placed in the section on disorders usually first evident in infancy, childhood, or adolescence.

In the fourth edition of DSM (DSM-IV) gender identity disorders was placed in a separate section called sexual and gender identity disorders. The heading has the same status as mood disorder, anxiety disorders, and other major classifications.

Number of Major Categories In DSM-III two specific categories of gender identity disorders were coded, each with its own diagnostic criteria: transsexualism and gender identity disorder of childhood. DSM-III-R added a third category, gender identity disorder of adolescence or adulthood, nontranssexual *type*, which appears to apply to persons with mild or fluctuating gender dysphoria.

DSM-IV reversed the trend toward greater differentiation by reducing the number of major diagnostic categories to just one, gender identity disorders. The main objective of the change was to unify the diagnostic criteria for children, adolescents, and adults. However, inspection of the DSM-IV diagnostic criteria shows that that goal was only partially realized; the diagnostic criteria for children do not fully parallel those for adults.

Subtypes Following a long tradition of classifying cross-gender syndromes according to the patient's sexual orientation, DSM-III and DSM-III-R listed three subtypes of gender identity disorder. The heterosexual subtype is attracted to members of the opposite genetic sex, the homosexual subtype to members of their own genetic sex, and the asexual subtype to neither. DSM-IV continues the tradition while avoiding the customary labels for those orientations. It also added a fourth subtype, bisexual, based on evidence that among gender dysphoric men the bisexual subtype is at least as common as the heterosexual or asexual subtypes. DSM-IV subtypes are (1) sexually attracted to males, (2) sexually attracted to females, (3) sexually attracted to both, and (4) sexually attracted to neither.

The current diagnostic criteria for both children and adults are organized under two main headings, cross-gender identification and *discomfort* with the assigned sex. A potential problem is that the arrangement promotes a questionable distinction between the two categories of symptoms. Criterion A for children, for example, includes the intense desire to participate in the games and activities of the other sex whereas criterion B includes rejection of sex-conventional toys.

DSM-IV requires the diagnosis of gender identity disorder not to be concurrent with a physical intersex condition (e.g., hermaphroditism). This criterion assumes, not always correctly, that when the symptoms of gender identity disorder are manifested by a child with some intersex status, the behaviors are due to that status. It also assumes that children without obvious intersex status have no physiological basis for their disorder, which also may ultimately be proved false.

In the ninth revision of *International Statistical Classification of Diseases* (ICD-9), gender identity disorders, without diagnostic criteria, were placed in the section on sexual deviations and disorders. As in DSM-III, ICD-9 had two major categories of gender identity disorders: transsexualism and disorders of psychosexual identity. The latter corresponded to DSM-III's gender identity disorder of childhood.

In ICD-10, gender identity disorders are placed in the section on disorders of adult personality and behavior. Three main categories are included: transsexualism, dual role transvestism, and—strangely, given the section heading—gender identity disorder of childhood.

The diagnosis of dual-role transvestism the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) corresponds

roughly to the DSM-III-R diagnosis of gender identity disorder of adolescence or adulthood, nontranssexual type, a diagnosis eliminated in DSM-IV. The diagnosis of dual-role transvestism applies to individuals who cross-dress episodically without any sexual excitement or those who desire surgical sex reassignment. It could be applied equally well to homosexual “drag queens” or to transvestic fetishists in their later years, who often continue to cross-dress even though this activity is no longer accompanied by penile erection or masturbation.

EPIDEMIOLOGY

Prevalence in Children The prevalence of the gender identity disorder of childhood can only be estimated because no epidemiological studies have been published. A rough estimate can be obtained from two items on Thomas Achenbach's Child Behavior Checklist that are consistent with components of the diagnosis: behaves like opposite sex and wishes to be of opposite sex. Among a sample of 4- to 5-year-old boys referred for a range of clinical problems, the reported desire to be of the opposite sex was 15 percent. Among 4- to 5-year-old boys not referred for behavioral problems, it was only 1 percent. For ages 6 to 7, the rates were 2.7 and 0 percent; for ages 8 to 9, 5.1 and 0 percent; and for ages 10 to 11, 1.1 and 2.3 percent. For clinically referred girls, there was more uniformity across the ages, with the highest being 8 percent at age 9, and the low 4 percent. For nonreferred girls, the highest rate was 5 percent at ages 4 to 5, and then less than 3 percent for other ages.

Parents reported cross-gender behavior for 16 percent of the clinically referred boys at ages 4 to 5, and for about 10 percent with the other age groups. Among the nonreferred boys the rates were about 5 percent. With clinically referred girls, nearly 19 percent reportedly showed cross-gender behaviors at ages 4 to 5, as did between 9 and 14 percent of girls in the other age groups. The rate was about 11 percent for all ages of nonreferred girls. These data offer only an approximation for gender identity disorder in children because they do not assess the longitudinal persistence of the reported behavior and do not elucidate what constitutes “behaves like opposite sex.”

Another way to estimate the prevalence of gender identity disorders in children is to use the percentage of adults with a predominant or exclusive homosexual orientation. The adult rates can then be compared with the percentage of homosexual men and women in various studies who report childhood cross-gender behavior. Predominant or exclusive homosexuality is estimated at 3 to 4 percent for men and 1½ to 2 percent for women. Estimates for homosexual men recalling childhood cross-gender behavior are between 50 and 65 percent, and for homosexual women perhaps 50 percent. A methodological problem is that the behaviors in those retrospective reports are contaminated by time and are different from study to study. Further, they rarely address the key item of wanting to be the opposite sex as a young child. Based on these imprecise assumptions, a high estimate of about 3 percent is reached for boys and of 1 percent for girls.

Sex Ratio in Children As many as five boys are referred for each girl referred, for which several explanations are possible. First, there is greater parental concern with sissiness than with tomboyishness, and greater peer group stigma attaches to substantial cross-gender behavior in boys. Thus, there may be an equal prevalence of gender identity disorder in boys and girls but a differential referral rate. Another possibility is that a genuine disparity results from the male's more perilous developmental course. The fundamental mammalian state is female. No sex hormones are required for prenatal female anatomical development (XO children with gonadal dysgenesis [Turner's syndrome] appear female at birth). Sex hormones are required at critical developmental times for male anatomical differentiation. If the mechanisms of behavioral development track anatomical development, the masculine behavioral system requires adequate levels of hormone at the appropriate time for normative expression. Finally, the psychodynamic developmental model explaining the disparate referral rates views both boys and girls as initially identifying with the female parent, with only boys needing to make the developmental shift for later normative identification.

Age of Onset in Children Most children with a gender identity disorder are referred for clinical evaluation in early grade school. Parents typically report that cross-gender behaviors were apparent before age 3.

Prevalence in Adults There is no basis for estimating the proportion of adults who would qualify for a DSM-IV diagnosis of gender identity disorder. The only relevant data are for transsexuals, who comprise only a subgroup of gender-dysphoric adults, and even those figures may be underestimates. The available data (from the United Kingdom, the Netherlands, Sweden, and Australia) place the prevalence rate of transsexualism at about 1 case per 10,000 males and 1 in 30,000 females.

Sex Ratio in Adults Recent data suggest that the sex ratio of adults with sufficient gender dysphoria to present at a gender identity center is about four males for each female. Many factors undoubtedly influence sex ratios at inception of treatment, including the greater publicity given to male-to-female transsexuals and the greater cosmetic and functional success of vaginoplasty compared with phalloplasty. Another factor is that gender-dysphoric females nearly all belong to the homosexual subtype whereas gender-dysphoric males include an equal number of homosexuals and a large contingent, which has no comparable counterpart among the females, of nonhomosexual (heterosexual, bisexual, and asexual) individuals.

Age of Onset in Adults Most gender-dysphoric adults of the homosexual subtype would have qualified for a DSM-IV diagnosis of gender identity disorder in childhood, with their adult behavior simply being a continuation of their childhood disorder. Both males and females of that subtype tend to present for clinical attention for the first time in their mid-20s.

Nonhomosexual gender-dysphoric adults (virtually all of whom are males) typically seek help for their disorder in their mid-30s—a striking difference from the homosexual subtype. If the age of onset is considered to be the point at which they first qualify for a DSM-IV diagnosis of gender identity disorder, they generally have an adolescent or adult onset. However, in most cases prodromes of the disorder were present before puberty.

ETIOLOGY

Much empirical evidence suggests that the three main types of nonhomosexual gender identity disorder in males (heterosexual, bisexual, and asexual) are superficially variant forms of the same condition; that nonhomosexual and homosexual gender identity disorder are etiologically different conditions; that nonhomosexual gender identity disorder is etiologically related to transvestic fetishism; and that homosexual gender identity disorder is etiologically related to typical homosexuality.

The conclusion that heterosexual, bisexual, and asexual gender identity disorders are superficially variant forms of the same condition is based on a wide variety of evidence. Similar majorities of men with heterosexual, bisexual, and asexual gender identity disorder acknowledge some history of transvestic fetishism; such self-reports are rare in men with homosexual gender identity disorder. Men with heterosexual, bisexual, and asexual gender identity disorder are also similar to each other, and dissimilar to men with homosexual gender identity disorder with regard to their degree of recalled childhood femininity, age at clinical presentation, extent of interpersonal heterosexual experience, and a history of erotic arousal in association with thoughts of being a woman.

It is possible that the common denominator linking transvestic fetishism and heterosexual, bisexual, and asexual gender identity disorder is *autogynephilia*, a male's tendency to be sexually aroused by the thought or image of himself as a woman. Autogynephilia is highly variable in its manifestations. It may be expressed in fantasies of dressing as a woman (transvestic fetishism); in (masturbatory) fantasies of engaging in stereotypically feminine behavior like knitting; in fantasies of gestating, lactating, or menstruating; in fantasies of being treated by other people as a woman; or in fantasies of possessing a woman's body. When an autogynephilic man's favorite sexual fantasy is that of possessing a vagina, he is very likely to develop cross-gender wishes that persist even when he is not sexually aroused, along with a desire for surgical sex reassignment.

Autogynephilia may be conceived as a modified form of heterosexuality, in which a man's sexual approaches are directed not at external women but at a feminized version of himself. It seems to involve some developmental anomaly in the learning of sexual behavior, because the man's principal erotic object in many cases—for example, the thought or image of himself wearing pantyhose, applying make-up, or knitting—cannot be innate but must have been assembled from experiences. It remains to be discovered whether some men are relatively prone to such developmental anomalies for neurological reasons.

The conclusion that homosexual gender identity disorder and typical homsexuality (i.e., homosexuality without gender identity disorder) have etiological commonalities is based on two lines of evidence. The first is that the early manifestations of homosexuality and homosexual gender identity disorder appear rather similar. Research has consistently shown that at least 50 percent of adult homosexual men with no gender identity problems nonetheless recall significant amounts of cross-gender behavior in childhood. Similar although somewhat less striking findings obtain for homosexual women. These observations suggest that the difference between ordinary homosexuality and homosexual gender identity disorder may begin as a difference in degree, which develops during adolescence into a difference in kind, when the less severely affected children shed their cross-gender traits and the more severely affected children elaborate them into a full-blown cross-gender identity.

The second line of evidence is epidemiological in nature and pertains only to males. Research on homosexual men without gender identity disorders has established that homosexual men are on average born later in the sibling order than comparable heterosexual men. Recent studies have established that this difference in birth order is caused by homosexual males having a greater number of older brothers; they do not have a greater number of older sisters, once their number of older brothers has been taken into account. Studies of Dutch, Canadian, and British male patients with gender identity disorder have produced the similar finding that homosexual patients are on average born later than nonhomosexual patients. These observations suggest that whatever etiological factor is reflected by high birth order contributes to the development both of homosexuality and of homosexual gender identity disorder.

The foregoing discussion illustrates that theories developed to explain homosexuality or transvestic fetishism may also apply to homosexual or nonhomosexual gender identity disorder, respectively. Furthermore, theories developed to explain gender identity disorder without further qualification may apply to only one of the two main types.

Biological Factors Theories of homosexual development, notably in males, have taken on an increasingly biological basis as opposed to an experiential one.

Genetic Factors The Franz Kallmann twin study of the 1950s found a 100 percent concordance for homosexuality between presumably monozygotic male twins. Further research identified discordant pairs, and methodological critiques of the Kallmann study resulted in a general decline of interest in the genetic basis. However, in recent years twin studies and other family studies of sexual orientation have promoted new interest.

A 1991 study of 56 male monozygotic pairs of twins raised together found a 52 percent concordance for homosexuality compared with 22 percent for 54 dizygotic pairs. A 1992 study found that of 71 female monozygotic twin pairs, 48 percent were concordant for homosexuality or bisexuality compared with 16 percent for 37 dizygotic pairs.

Monozygotic twins separated at birth, although rare, provide a better model for testing the relative influences of environment and genetics than do twins reared together, where the two factors are confounded. A report of two pairs of males separated at birth argues for an inherited influence on homosexual orientation. In one pair, both men were homosexually oriented. In the second pair, one twin was homosexual, and the other, while heterosexually married, had had a 3-year homosexual relationship in adolescence. By contrast, in four pairs of separated female-female twins where one twin in each pair was lesbian, none of the cotwins was lesbian.

Family studies of nontwin siblings of homosexual men and women also lend support to a genetic basis, although the confound of a similar environment is considerable. Two studies found higher rates of homosexuality in brothers than is expected in the general male population. No corresponding increase in the number of lesbian siblings was reported.

Two gene linkage studies add further weight to a genetic basis of male homosexuality. When families are selected for having male homosexuals on the mother's side of the family tree, and two of the mother's sons are homosexual, there is an increased probability of a marker for a shared gene on the sons' X chromosome (contributed by the mother). The marker is less often shared between a homosexual and heterosexual brother.

Hormonal Factors Evidence for a hormonal influence on gender identity disorder derives from several research sources. One possible source is congenital virilizing adrenal hyperplasia. Girls with this condition overproduce adrenal androgen from before birth. They are more rough-and-tumble, less interested in doll play, and more likely to be considered tomboys than girls without the condition. Conversely, there is limited evidence that prenatal exposure of males to estrogenic or progestational agents may reduce the expression of conventional boy-type behaviors. Atypical levels of sex-typed hormones before birth and the attendant effects on specific sex-typed behaviors can substantially modify the child's early social experiences. Boys who are disinclined to rough-and-tumble play or who play with dolls have different father-son and mother-son relationships and a different peer group experience from more conventionally masculine boys. Similarly, girls who prefer rough-and-tumble activity and sports to doll play have a different early socialization experience with parents and peers from girls who are conventionally feminine. Thus, hormonal influences may act through a pathway of affecting sex-typed behaviours that interact with socialization experiences.

Reported neuroendocrine and neuroanatomical differences also suggest an inborn contribution to sexual orientation, particularly in men. One phenomenon tested is the feedback response on luteinizing hormone (LH) after an intravenous pulse of estradiol. In women there is a marked rebound after an initial drop (the hormonal basis of ovulation). The original research found an attenuated female-like response in homosexual men, which theoretically reflected a deficiency in prenatal androgenization of the central nervous system (CNS). In another study using the same methodology, more than half of a sample of homosexual men showed a response more like that of the heterosexual women than of the heterosexual men in the study. However, a subsequent study, which used a different approach to elicit the luteinizing hormone feedback phenomenon, found no significant group difference, and another study with a methodology similar to that used in the original research also failed to confirm a difference.

A related phenomenon that suggests that a deficiency in male hormone in utero leads to a homosexual orientation in men derives from the prenatal stress theory.

Stressing pregnant rodents results in feminized behavior in male offspring, owing either to the competition between adrenal stress steroids and testicular androgens or to the mistiming of testicular androgen secretion as a result of stress. In one study a higher-than-average rate of homosexuality was found in men who were born in Germany between 1941 and 1946, the stressful years of World War II. However, an environmental explanation is also possible, because fathers were more likely to be away from their sons during the war. A second study, based on retrospective reports by homosexual, bisexual, and heterosexual men describing stress in their mothers, found more stress during the pregnancies of the mothers of homosexual men.

Other research has been less supportive of an association between stress and homosexuality. Some research found no connection. One American study found a marginally significant relationship, based on the reports of mothers of college students. Another found a low correlation between reported pregnancy stress and lesbianism, but not with male homosexuality. No prospective studies are reported.

Although medical histories given by parents of children with gender identity disorder do not provide a basis for grossly abnormal hormone levels before birth, a neuroendocrine base may still be posited at a more subtle level. If the range of prenatal androgen levels is as wide as that in adult life, the fetus may also be exposed to a wide range of androgen. Another factor could be the androgen surge that occurs in boys between about 3 weeks and 3 months of age.

Immunological Theories Two immunological theories have been advanced to explain the finding that homosexual men (including those with gender dysphoria of the homosexual subtype) have a greater average number of older brothers than do heterosexual men. Both theories propose that male homosexuality may result from a maternal immune reaction, which is provoked only by male fetuses and which becomes stronger after each pregnancy with a male fetus. The earlier theory proposed that antibodies to testosterone, produced by a woman pregnant with a male fetus and passed through the placenta from the mother to the fetus, could reduce the hormone's biological activity and thus compromise the sexual differentiation of the fetal brain. This seems unlikely because steroid hormones are not ordinarily antigenic. An alternative theory is that the relevant fetal antigen might be one of the male-specific, Y-linked, minor histocompatibility antigens, often referred to collectively as H-Y antigen. Although there is no direct evidence for this theory, it is consistent with a variety of observations, including the finding that male mice whose mothers were immunized to H-Y prior to pregnancy are much less likely to mate successfully with receptive females.

Brain and CNS Involvement A difference in a nucleus of the anterior hypothalamus may represent a CNS difference related to sexual orientation. The area known as interstitial nucleus of the anterior hypothalamus-3 (INAH-3) was compared in autopsy between homosexual men, heterosexual men, and heterosexual women. Although there was some overlap between the size of the nucleus between the groups, it was smaller on average in the homosexual men and women compared with the heterosexual men. All the homosexual men and some of the heterosexual men and women had died of acquired immune deficiency syndrome (AIDS), but death from AIDS was not a factor. No homosexual women were studied to determine whether the size of their nucleus was similar to that of the heterosexual men. INAH-3 is embedded in the hypothalamic area that appears to be related to some aspects of sexual behavior in male nonhuman primates. This study has neither been confirmed nor refuted by subsequent research. Another finding, of a larger suprachiasmatic nucleus in a sample of homosexual men, may be less relevant because that area is not known to be associated with sexual behavior. However, it may be related to endocrine function. A more recent finding points to a difference in the brain of male-to-female transsexuals. In a postmortem sample of 6, the bed nucleus corresponded in size to that of typical females rather than to that of typical males; it was not relevant whether the male transsexual was heterosexual or homosexual.

Psychosocial Theories Psychodynamic and behavioral influences may lead to extensive cross-gender identification. In an early study boys with an excessive mother-son symbiosis in the early years, replete with extensive mother-son skin-to-skin contact, appeared later to manifest significant feminine behavior. This is attributed to the inability to differentiate psychologically from the mother. Male-identified females have been reported to have mothers who were removed in affect

from their children, frequently by depression, and fathers who did not support their daughters' femininity. The girl becomes a substitute husband to treat the mother. Other reports describe traumatic psychological losses to boys and girls in the earliest years that appear related to the onset of cross-gender behavior.

Research on a sample of 66 boys with gender identity disorder found a positive correlation between the extent to which parents supported cross-gender behaviors in their sons and the extent of that cross-gender behavior. In most of the families at least initially there was no discouragement of cross-gender behaviors. In more limited work with girls with gender identity disorder, initial parental reactions were similar. A study of cross-gendered boys found the extent of father-son involvement in the early years to be related to later sexual orientation. The association emerged not only between the two groups of boys studied (gender identity disorder and control) but within the subgroup of boys with gender identity disorder. Less father-son involvement was associated with a more homosexual orientation.

Social Learning Theories Social learning theories typically focus on the differential reinforcement by parents of sex-typed behaviors, starting shortly after birth. This reinforcement shapes conduct into conventional masculinity or femininity. Cause and effect are hard to distinguish here. On the one hand, sex differences are reported early in life, probably before any major differential impact of parental reinforcement; on the other hand, mothers and fathers apparently treat male and female newborns differently.

In Baby X experiments adults are told, sometimes incorrectly, the sex of a clothed child and asked to describe the child's attributes or to provide it with toys. Perceived boys are encouraged more to physical action and are given more whole-body stimulation than perceived girls.

When 6-month-old children were similarly clothed, toy choice by adults was related to perceived gender of the child. Boys were presented with footballs, girls with dolls. Strong bald babies were seen as male, soft fragile ones as female.

At 1 year, boys may be more exploratory and active and toy preferences may differ. Girls were found to prefer soft toys and dolls whereas boys preferred transportation toys and robots. A preference for same-sex playmates emerges early. When 3½- to 4½-year-olds were shown photographs of boys and girls and asked to select those with whom they would like to play, boys preferred boys and girls preferred girls. By age 2 to 3 years, boys appear to be more aggressive toward peers and to show more rough-and-tumble play. Fathers are as likely to give a 1-year-old daughter a truck as a doll but more likely to give the son a truck. However, when children are given dolls, boys play with them less than girls. Fathers more than mothers give negative responses to boys playing with dolls. Boys receive more positive responses for playing with blocks and girls receive more positive responses for playing with dolls.

Imitative and vicarious learning pervade general theories of social learning of sex typing. In imitative learning, behaviors are adopted that simulate those of a significant other person, the model. In vicarious learning, if something happens to a model the viewer's behavior is modified to resemble the model because the child perceives the model as possessing desirable attributes or obtaining desirable goals. The cognitive developmental theory, by contrast, sees the child first labeling itself as male or female and then finding the behaviors associated with that label rewarding.

Nature Versus Nurture The classic research on intersexed or hermaphroditic children pointed to the early-life emergence of gender identity as being influenced primarily by environment and as irreversible. In the studies by John Money, Joan Hampson, and John Hampson, a range of anatomical features discordant with the gender of rearing were found to be less relevant to the adoption of a male or female gender identity than the gender of rearing.

Studies of matched pairs, for example, demonstrate that with the syndrome of congenital virilizing adrenal hyperplasia, the newborn female, if considered to be male and designated male, matures with a male identity in spite of having the XX female chromosomal pattern, ovaries, and a uterus. If considered female the child matures with a female identity. However, questions have been raised about the generalizability of those findings to nonintersexed children because of the atypical prenatal endocrine environment and other atypical genetic or anatomical influences of intersexed children.

Studies of children born with normal sex characteristics who undergo gender reassignment early in life may be a more relevant test of nature versus nurture. The tragedy of penile amputation, usually through negligent circumcision, has provided such a model. In one celebrated case a reassigned male monozygotic twin who was reportedly being raised successfully as a girl failed to incorporate a female identity and now lives as a heterosexual male. Reassignment in that case was at 22 months, which may have been after core identity as a male was in place. A somewhat more complicated outcome was revealed in the very recent psychosexual follow-up of another biologically normal male whose penis was accidentally ablated during circumcision at the age of 2 months. The decision to reassign the infant as a female occurred sometime between 2 and 7 months, at which point surgical castration occurred. At age 26 years, clinical interviews and self-report questionnaires were used to obtain information on the patient's gender identity, gender role, sexual orientation, and sexual identity. In adulthood the patient lived socially as a woman and her gender identity was unequivocally female, with no evidence of gender dysphoria. However, the patient's childhood gender role behavior had been predominantly masculine, and her current occupation was male-dominated. Moreover, she fantasized sexually about women more often than about men. On the other hand, her objective sexual history included roughly equal amounts of sexual experience with women and men. At the last follow-up, the patient was living with a new female partner, in a lesbian relationship.

Psychoanalytic Theories As in other areas of psychopathology, psychoanalytic theories about gender identity disorders constitute a tradition distinct from biological and other nonbiological approaches. One influential theory is that of Ethel Person and Lionel Ovesey, who advanced the hypothesis that transsexualism in males originates from unresolved separation anxiety during the separation-individuation phase of infantile development. To cope with this anxiety, the child resorts to a reparative fantasy of symbiotic fusion with his mother. Adult transsexualism may be understood as an attempt to master that anxiety through sex reassignment surgery, through which the transsexual acts out his unconscious fantasy and symbolically becomes his mother.

According to this hypothesis, male transsexuals vary in the directness with which they proceed to the transsexual resolution. Some individuals never develop any other psychosexual phenomena as defenses against separation anxiety, and they proceed to the transsexual outcome in a straightforward manner. Others develop transvestism or effeminate homosexuality as initial defenses. When those defenses fail in the face of various stressors, the individual regresses to the primitive fantasy of symbiotic fusion with his mother and begins to experience transsexual impulses.

The other major psychoanalytic theory was developed by Robert Stoller to explain the etiology of transsexualism in a specific group of biological males, who would fall within the DSM-IV category of gender identity disorder, sexually attracted to males. Stoller called those males true transsexuals.

The theory begins with the grandmother of the future transsexual who treats her daughter coldly and neither encourages nor models femininity for her. The grandfather has a closer relationship with the daughter, but he encourages masculinity in her. In consequence the mother of the future transsexual develops a mild gender identity disorder of her own. In adolescence, however, she abandons her conscious transsexual wishes of someday being male and adopts a heterosexual façade. At the unconscious level she nevertheless retains a strong penis envy.

The transsexual's mother eventually enters an empty marriage with a passive and withdrawn husband who is psychologically if not physically absent from the household. The final pathogenic process becomes operative when the mother gives birth to an infant son whom she perceives as particularly beautiful and graceful. The boy, who represents her feminized phallus, fulfills her lifelong wish for a penis. The mother-son interaction, described by Stoller as a blissful symbiosis, includes excessively close and prolonged body contact, sometimes with the infant's nude body cradled against the mother's nude body. The mother's behavior expresses her need to treat her son as an extension of her own body.

The transsexual's early experiences, especially the continuous skin-to-skin contact, produce an overidentification with his mother, a blurring of ego boundaries, and eventually a feminine gender identity. The transsexual boy never develops a “heterosexual” relationship with his mother and therefore never develops an oedipal conflict. His femininity is produced nonconflictually and remains a nonconflictual, autonomous form of behavior. This theory does not account for “secondary” transsexuals, notably those who evolve through a transvestite, heterosexual pattern.

DIAGNOSIS AND CLINICAL FEATURES

The DSM-IV diagnostic criteria for gender identity disorder are listed in Table 19.3-1. The ICD-10 diagnostic criteria are presented in Table 19.3-2.

Table 19.3-1 DSM-IV Diagnostic Criteria for Gender Identity Disorder

Table 19.3-2 ICD-10 Diagnostic Criteria for Gender Identity Disorders

Children A longitudinal study of 66 boys provides a picture of gender identity disorder in children. The age range at the initial evaluation was 4 to 12 years. One third of the boys frequently stated their wish to be girls and three fifths did so occasionally; three quarters cross-dressed frequently. The age of onset of cross-dressing was before the fifth birthday with 90 percent. A female-type doll, such as Barbie, was the favorite toy for one fifth of the boys and a frequently played with toy for another two fifths. Three-fifths of the boys took a female role when playing house; over four-fifths had a primarily female peer group; and three fifths of the boys were rejected by other children or were voluntary loners.

The full prepubertal age range for evaluating gender identity disorders in children yields a variety of presenting behaviors. Younger children aged 3 to 5 may believe that they are of the other sex or that they can easily become the other sex. Between the ages of 6 to 9 years the children's gestures and mannerisms may be cross-sex stereotypes. Older children have often been subjected to increased peer teasing and so may have gone underground with their cross-gender behaviors, particularly cross-dressing.

A 5-year-old boy, referred by his general practitioner, was brought for evaluation by both parents. He had been saying either he was a girl or had wanted to be one since age 3. His preferred clothing was that of his sister. He wanted to wear makeup and his mother's jewelry. His favorite toys were dress-up dolls, especially those with long hair. He refused to stand to urinate. In make-believe games he was usually the mommy or a female character. Most of his friends were girls. He showed no interest in sports. All the pictures he draws are of women. The cross-gender behavior was recalled by both parents since the boy was age 2. Neither attempted to interrupt it or redirect his interests until recently, when their efforts were met by the boy with resistance. They had been advised by a preschool teacher and a regular babysitter to ignore the behavior as it would go away. Neither parent expressed concern that their son would become homosexual but they are concerned that he may be transsexual.

A 7-year-old girl referred through the school had been insisting to other children that she was a boy. Since the age of 2 years she has said that she was a boy and that she did not want to be a girl. Her clothing style preference is that of boys and attempts to make her wear a dress are met with refusal. She watches her father shave and ignores her mother applying makeup. Her friends are boys and her favorite activity is sport. She has no interest in doll play but loves action and soldier toys and guns. She frequently attempts to urinate while standing.

Associated Features Boys with gender identity disorder have been shown in some reports to evidence greater general psychopathology than nonclinical control boys. Using the Child Behavioral Checklist, a study found that boys with gender identity disorder had indices of psychopathology similar to a clinic-referred group used in the instrument's standardization. However, another report by the same investigator did not find more behavioral problems on the checklist than among concurrently assessed, demographically comparable clinical controls.

Some clinical experience argues that much of the behavioral problem seen in gender-atypical boys is secondary to discomfort over gender and the consequent social ostracism and teasing. That social ostracism is the basis of psychopathology in cross-gendered boys is supported by the finding that checklist symptom scores increase with the age of the child (when stigma increases).

Some clinicians have found separation anxiety disorder in boys with gender identity disorder. They claim that separation anxiety disorder precedes feminine behavior, with cross-gender behavior emerging to restore the emotional tie with a mother who is perceived as unavailable. A recent study using liberal criteria for diagnosing separation anxiety disorder found a correlation between that disorder and gender identity disorder. In the absence of convincing data demonstrating that separation anxiety disorder precedes gender identity disorder and because most children with separation anxiety disorder do not have a gender identity problem, the connection is still unclear.

Psychological Tests No psychological test is diagnostic of gender identity disorder in children. However, two tests, the It-Scale for Children and the Draw-A-Person Test, discriminate boys with gender identity disorder from gender-typical boys.

The It-Scale presents a child with a neuter stick figure (It). “It” then selects from a series of cards depicting gender-typed accessories and activities. Cross-gendered boys more often select feminine or girl-type cards. The Draw-A-Person Test in its basic format requires a child to draw a person. Most gender-typical children draw a person of their sex. By contrast, the majority of cross-gendered boys draw a female first. Conversely, with a sample of nonclinical tomboys, the majority drew a male first, in contrast to a matched sample of nontomboys. These tests can be used as ancillary evaluation procedures. The Draw-A-Person Test, in particular, is a good technique for making a child comfortable during the initial interview.

Adults Virtually all adults who present complaining of gender dysphoria are self-referred. In most cases the relevant diagnostic question is not whether the patient is gender dysphoric but how severe the condition is.

Central features of gender identity disorders in adults are the persistent sense that one was born into the wrong sex, the belief that one would have been happier as the opposite sex, and the conviction that one has the typical feelings and reactions of the opposite sex. The discontent extends to the social, interpersonal, and somatic realms of the person's existence although the relative emphases differ from case to case.

Social Realm At the social level there is a consistent indifference to or distaste for roles and activities traditionally associated with the original sex, together with recurrent desires to live in society as a part-time or full-time member of the opposite sex. In response to these longings, the gender-dysphoric person begins cross-dressing in isolation, in the company of friends, or attempts to pass as a member of the opposite sex in public. These behaviors may stabilize as intermittent

activities or expand to permanent adoption of the cross-gender role, depending in part on the intensity of the gender dysphoria and the person's success in simulating the appearance and manner of the opposite sex.

Interpersonal Factors The enjoyment of relationships, especially romantic ones, is significantly diminished by the conflict between the person's subjective feelings of masculinity or femininity and the discrepant perceptions of others. A female-to-male transsexual, for example, may reject an otherwise acceptable female partner if that partner clearly regarded the transsexual as essentially female and the relationship as a lesbian one. Efforts to find sexual or romantic partners who perceive the gender-dysphoric person as a member of the opposite sex may lead the person into dangerous circumstances. For example, adolescent gender-dysphoric females sometimes pass themselves off as males when first dating or courting potential girlfriends. Male-to-female transexuals sometimes complete entire sexual encounters with unsuspecting men, discouraging interest in coitus with the excuse of menstruation and offering fellatio instead.

Somatic Factors The gender-dysphoric person often fails to value the primary and secondary sexual characteristics of his or her body, in many cases developing an aversion to those characteristics. A male-to-female transsexual, for example, may describe his penis as an “ugly growth” and avoid touching it as much as possible. Somatic dysphoria is usually accompanied by the desire for some or all of the physical characteristics of the opposite sex. Gender-dysphoric persons use various means to simulate the desired phenotype. Women, for example, commonly flatten their breasts with elastic binding to produce a masculine chest contour, men wear prosthetic breasts. Transsexuals of both sexes seek hormonal and surgical treatment to complete the transformation.

A 37-year-old man presented with a request for sex reassignment surgery. He was dressed as a man, and he was masculine in his physical appearance and mannerisms.

The patient had not been observably effeminate in boyhood. He had, however, been attracted to and fascinated by feminine activities from an early age. He particularly recalled having liked to watch women apply makeup. He began cross-dressing around puberty. During adolescence he sometimes stole women's underwear from clotheslines. At that stage he preferred women's clothes that had been worn because he felt that they were somehow infused with femininity. Cross-dressing was sexually arousing from puberty until his early 30s, when sexual excitement yielded to feelings of comfort and naturalness.

The patient married at 25. He initially hoped that marriage would cure his cross-dressing but that hope soon proved false. The couple's sex life was poor and the marriage dissolved after 8 years. The patient began another heterosexual relationship a few years after his divorce, but that relationship also failed in the face of his growing gender dysphoria.

Following clinical assessment, the patient began making systematic plans to move into the female role and commenced living and going to work as a woman full-time at 39. The attempt to live as a woman proved successful, and the patient underwent surgical sex reassignment at 41.

After moving into the female role, the patient went on a few dates with men but found that she could not really develop an interest in them. After surgery she began moving in lesbian social circles and having sexual relationships with women. She was content with her life in general and had no regrets about her decision to undergo sex reassignment.

A case of classic transsexualism, corresponding to the popular notion of a “woman trapped in a man's body,” is illustrated by this homosexual male gender-dysphoric patient, who presented at a university-affiliated gender identity clinic at the age of 22, requesting sex reassignment surgery. The patient presented in the female role, wearing a dress; well-applied make-up; clean, long, curly hair with bangs; and pink nail polish. The patient's voice and mannerisms were convincingly feminine, and the overall impression was that of an attractive young woman.

The patient was the second of three children, having one older brother and one younger sister. The patient reported that from earliest childhood he preferred girls' games, dolls, and female playmates. His obviously feminine behavior continued after he entered school, and he was teased and ridiculed by his classmates, sometimes being asked if he was a boy or a girl. The patient's mother, who was also interviewed as part of his assessment, reported that she had noted the teasing by his classmates when he was in school but claimed that she had not attributed this to his feminine behavior and had not known what was causing it. The mother stated that she never tried to feminize the patient in any way, but she did allow him to dress up as a pretty young witch on Halloween.

The patient began cross-dressing secretly in Grade 8, in his mother's nylons, brassieres, and dresses. However, he wore his own underwear because he did not want to soil his mother's clothing. He avoided detection by being very careful about how he returned his mother's clothes to her drawers. These sessions were never accompanied by sexual arousal.

The patient reacted negatively to his physical changes at puberty, especially the appearance of facial hair. At age 15, he began to hide this with makeup, which he wore outside the home and even to school.

He never made any real friends at school and he did not participate in any school activities. He was very lonely and unhappy during these years, and he made a suicidal gesture in Grade 10 with pills belonging to his mother. The only effect of this was that he awoke the next morning with a bad headache, and he did not seek medical or psychiatric attention. Despite his psychological and social problems, he maintained average grades and graduated high school at the normal age.

After graduating high school, the patient remained at home. His father died in a construction accident around this same time. The patient was still living at home with his mother and younger sister at the time he presented at the gender identity clinic. He was unemployed and had never held a job.

He informed his mother and sister about his cross-gender feelings shortly after his father's death. The mother's initial reaction was: “I hate to see it this way. I pray every night it will change, but I know it won't.” During this period the patient began cross-dressing at home, but he usually waited for his mother and sister to go out, for fear of making them uncomfortable. The patient began going out in public as a female in the year prior to his presentation at the gender identity clinic.

At the time of clinical intake, the patient's sexual history was virtually nil. He reported that he had masturbated only once, around the age of 18. He found the experience “horrible and disgusting.” He had had no sexual experience with either men or women. He reported that he was romantically interested in men but that he would not feel comfortable attempting sexual intercourse at the present time. He explained his reluctance as follows: “I guess I'm just an old-fashioned girl—I don't believe in sex before surgery.”

Following the intake assessment, the patient (now “she”) adopted the female role full-time. She changed most of her identifying documents so that they showed her as a female, and she obtained a job as a secretary. She also obtained her driver's license—something she had thought she could never achieve. Her morale improved substantially, and with it, the degree of acceptance from her family. The patient's mother commented that “living with John was a living hell,” but that Joanna was a completely different person. She cited, as an example, the fact that the patient had a conversation with her brother for the first time that the mother could remember.

The patient eventually acquired a boyfriend who had no idea that she was biologically male. She avoided sexual intercourse and his proposal of marriage with the excuse that she had “intersex problems” and required surgery to make her a more complete woman. She underwent surgery at the age of 28 and married her boyfriend a year later. The couple were still happily married at our last contact with the patient, 4 years later. The patient was at this time managing her husband's trucking business. He still did not know the nature of her “intersex” problems and was aware only that she could not have children.

A 21-year-old woman complained that she felt uncomfortable with her body, that she should have been a male, and that she wanted to be a male. She presented in the female role, but all of her clothes (jeans, shoes, shirt, vest) were men's style. Her mannerisms and her voice were convincingly masculine.

The patient recounted a childhood history typical of gender identity disorder in girls. Her dissatisfaction with her sex intensified at puberty. She hated menses and her developing breasts, which she began hiding with jackets and sweatshirts. The last time she wore a dress was at her eighth-grade graduation.

The patient had no sexual experience with men. Her first homosexual relationship occurred in high school and lasted about 2 years. Her second was with her current partner, a divorcee 10 years older than she. She was cohabiting with her partner and her two young children. Their sexual relationship was reported by both partners to be satisfactory although it was one-sided; the patient brought her partner to orgasm but would not allow her own breasts or vulva to be touched because that reminded her of her anatomical sex.

The patient regarded herself as gay in high school but eventually came to realize that she was transsexual. Her goals at clinical presentation were to undergo sex reassignment, to marry her partner, and to be a father to her partner's children.

Associated Features Gender identity disorders may be associated with character pathology. Clinical authors sometimes assert that gender identity disorder patients tend to have narcissistic or borderline personality disorders, perhaps because a significant minority of patients seem self-absorbed, demanding, unempathic, selfish, inconsiderate, or interpersonally shallow. It is likely that only a fraction of these, however, would meet DSM-IV criteria for a formal diagnosis of narcissistic personality disorder or borderline personality disorder. It is possible that, in some cases, the labels “narcissistic” and “borderline” are simply used as the closest available description for personality patterns that in fact may be specifically associated with gender identity disorders.

Associated features vary markedly for different subtypes. Gender-dysphoric women exhibit the least associated psychopathology, and what they do exhibit appears to fall into no particular category. The sexually attracted to male subtype of males includes a proportion of persons with histories of drug abuse, property offenses, and prostitution. Prostitution is usually carried out in the female role, or, more precisely, in women's attire; the customers are frequently men specifically seeking transsexual prostitutes in preference to biologically female prostitutes.

The majority of men with nonhomosexual subtypes of gender identity disorder report past histories of erotic arousal in association with cross-dressing, and some qualify for a concurrent diagnosis of transvestic fetishism. Careful questioning usually reveals that they are more aroused by the thought or image of themselves as women (autogynephilia) than by items of clothing per se. Therefore, the label “fetishism” applied to their sexual behavior, may be somewhat misleading.

One feature that is occasionally found among nonhomosexual gender-dysphoric men is sexual interest in (other) feminized males. It has been previously argued that persistent and recurrent sexual interest in such males represents a distinct paraphilia, *gynandromorphophilia*. The feminized males of interest to gynandromorphophiles include men wearing women's attire and men—known in the vernacular as “she-males”—who have surgically or hormonally feminized bodily contours but intact male genitalia. In some cases the key feature that makes a feminized male attractive to a gynandromorphophile is simply the thought or knowledge that this individual had been transformed from male to female; thus, postoperative male-to-female transsexuals are of interest to some gynandromorphophiles, even though the reassigned transsexual lacks male genitals.

The clinical impression that gynandromorphophilia is sometimes associated with gender dysphoria, transvestism, or both was supported by a statistical study of male subscribers to a voice mail system devoted to personal advertisements for sexual or romantic partners. About 40 percent of the men who sought to meet cross-dressers, transvestites, transsexuals, or she-males for sexual or romantic purposes were cross-dressers themselves.

This 39-year-old lawyer was referred to a gender identity clinic because of various related complaints involving gender identity. He was unremarkably masculine in manner and demeanor, wearing a full beard and appropriate male attire. The patient reported that he had been thinking about sex reassignment; however, he appeared to have no real desire for women's genitals and actually said that he might prefer to live as a woman with a penis.

The patient's earliest recollection of anything that presaged his later gender identity problems was a vague memory involving his mother's petticoat. His second memory, at about the age of 6, was that of encountering a woman's shoe with a stiletto heel lying in the gutter. He remembered this because it made such an impression on him. He cross-dressed for the first time at about the age of 8, in a slip that he found outside. He began cross-dressing in earnest at about the age of 12 or 13, using both his mother's and his sister's clothes. Cross-dressing was usually accompanied by masturbation. In his late 20s he lived by himself during one period, during which time he cross-dressed very frequently. He had never gone out in public in women's attire, however, and he had ceased cross-dressing at the time he presented clinically.

The patient had his first heterosexual experience at about age 20. This was a casual encounter with a woman he met at a party or in some similar situation; he could not fully recall the details of this first intercourse. He had no long-term sexual relationships until he met his future wife, at age 29. They were married about 2 years later; the marriage lasted 5 years. There were no children, because neither partner desired them. Although his wife knew about his cross-dressing early in their relationship, she became progressively less accepting over time. The patient and his wife were divorced 4 years before he presented at the gender identity clinic.

When he was having heterosexual intercourse, he had difficulty ejaculating unless he engaged in private fantasies. He mentioned three fantasies that could excite him enough to ejaculate: the first was that he was cross-dressed, the second was that some other male was cross-dressed, and the third was that his partner was really a postoperative male-to-female transsexual (rather than a biological female). The patient had no sexual interest at all in (untransformed) men. He had had only one homosexual experience in his life, at age 12 or 13, with another boy about the same age. He did not really understand what was happening at the time, and this was a meaningless experience for him.

The patient's erotic fascination with feminized males appeared to center on the thought of a male being (or having been) transformed into a female. The erotic potential of this thought did not depend on concrete evidence of maleness, such as intact male genitals. On the contrary, the presence of male genitals on an otherwise attractive feminized male would kill any possible sexual interest on his part. The only type of feminized male with whom he could contemplate actual intercourse was a postoperative male-to-female transsexual who simulated a biological female perfectly, and whom he simply knew previously to have been male. On the one occasion when he attempted sexual intercourse with a real postoperative transsexual, the experience was not successful because he was turned off by the transsexual's residual male anatomical features, such as beard stubble.

The patient's first recollection of being stimulated by the idea of another feminized male was at about age 15 or 16, when he first heard of or saw female impersonators. He was excited by this idea as soon as he was exposed to it. At the time he presented, he was occasionally going to bars where she-males or transsexuals congregated. He did not attempt to pick any up because he was afraid of contracting HIV. He did, however, talk to them and appeared to derive some erotic pleasure simply from being in that milieu. Although his gynandromorphophilic interests were longstanding, recurrent, and well differentiated, they do not appear to have been his strongest sexual interests. He reported that his main masturbatory fantasy was that of being cross-dressed in the presence of a woman.

Other gender-dysphoric patients with gynandromorphophilia illustrate similarities to the patient in the foregoing vignette as well as differences. In the next vignette, a striking similarity is the patient's self-report—accurate or otherwise—that his very first exposure to the image of another cross-dressed male produced an immediate and powerful emotional reaction. A striking difference is the patient's absolute requirement that the male genitals be present rather than absent for a feminized male to be sexually arousing.

A heterosexual male gender-dysphoric patient recalled that when he was 4 or 5 years old he watched a Bugs Bunny cartoon in which Elmer Fudd, for reasons relevant to the plot, cross-dresses as a female. The patient wrote that he could still remember the impact of seeing Elmer Fudd dressed as a female: “The moment my young mind saw Fudd changed into a woman something hit me like a fist in the chest, releasing in me a desire to become a woman myself—a desire that has pursued me my entire life.” When he reached puberty, his favorite masturbatory fantasy was that some imagined male—not himself—was being transformed into a female. In adulthood, he spent considerable sums of money on she-male pornography. It was essential to him that the male genitals be exposed in the photograph or videotape. He reported that the sight of the genitals confirmed the person's biological sex and that this was what produced sexual excitement. The sight of the penis and scrotum was not arousing in and of itself.

DIFFERENTIAL DIAGNOSIS

Children Gender atypical children without a gender identity disorder must be distinguished from those with a diagnosable disorder. Tomboys without gender identity disorder prefer functional and casual clothing or gender-neutral clothing. By contrast, gender identity disorder girls refuse to wear girl's clothes and usually reject gender-neutral clothes. Many girls prefer shirts and pants to dresses, enjoy rough-and-tumble play or sports, and show little interest in doll play. They may say it is better to be a boy because of perceived social advantages, but such girls do not necessarily have a gender identity disorder. What distinguishes girls who do is their repeated statements of being or wanting to be a boy and wanting to grow up to be a man, along with repeated cross-sex fantasy play.

With boys the differential diagnosis must distinguish those who do not conform to traditional masculine sex-typed expectations but who do not show extensive cross-gender identification, and who are not discontented with their anatomical sex. It is not uncommon for boys to reject rough-and-tumble play or sports, to prefer sedentary or aesthetic activities, or occasionally to role-play as a girl, to play with a doll, or to dress up in a girl's or woman's costume. Such boys do not necessarily have a gender identity disorder. This fact must be stressed to parents, especially to fathers who may have vigorous athletic expectations for their sons. What distinguishes boys who do have a gender identity disorder is the stated preference for being a girl and for growing up to become a woman, along with repeated cross-sex fantasy play, a strong preference for traditionally female-type activities, including cross-dressing, and a female peer group.

The clinician may sometimes see boys with repetitive cross-dressing but no other observable evidence of cross-gender identification. Those boys are often brought by parents when they discover their son cross-dressed, usually in his mother's underclothes. Such children would not satisfy the diagnostic criteria for gender identity disorder in children, although it is possible that some of them will warrant a diagnosis of gender identity disorder by the time they are grown. Monosymptomatic cross-dressing of this nature is sometimes recalled by adults with transvestic fetishism, but no studies are available predicting the sexual outcome of early isolated cross-dressing.

Because the diagnosis of gender identity disorder excludes children with anatomical intersex, a careful family medical history needs to be taken, with a focus on any suggestion of hermaphroditism in the child, such as an unusual appearance of the genitalia. When there is doubt, referral to a pediatric endocrinologist is indicated for appropriate hormonal and cytological examinations.

Adults Psychotic persons rarely develop the delusion that their sex is changing. A male patient, for example, may claim that he can feel his breasts growing or his penis shrinking. Such beliefs are never expressed by adult patients with gender identity disorder, who understand that sex reversal does not occur spontaneously. In any event there are other signs (e.g., hallucinations, ideas of reference) to aid in the differential diagnosis.

A more difficult diagnostic problem is posed by some late-adolescent patients who report that they were extreme sissies (or tomboys) as children, that they have always felt like members of the opposite sex, and that they are uncomfortable with their bodies, but who go on to say that they feel guilty about being attracted to members of their own sex, that they believe homosexuality to be worse than transsexualism, and that they want sex reassignment so that they can lead normal heterosexual lives. It can be difficult at initial assessment to determine how much of the patient's behavior is driven by true gender dysphoria and how much by internalized homophobia. Fortunately, the diagnosis becomes clearer when such patients are followed into young adulthood.

COURSE AND PROGNOSIS IN CHILDREN

In a 15-year prospective study, 44 of 66 cross-gendered boys, most of whom would be diagnosed today with gender identity disorder, were followed to adolescence or young adulthood. Sexual orientation was determined by interviews tapping erotic fantasies and erotic behaviors. Fantasies were determined from questioning about masturbation content, erotic nocturnal dreams, and experiences of arousal when seeing pornography or attractive persons. Behaviors were assessed by reports of interpersonal genital sexuality.

On the dimension of erotic fantasy, 33 of 44 (75 percent) previously gender-atypical boys were bisexual to homosexual (rated 2 to 6 on the Kinsey 7-point scale of sexual orientation, where 0 is exclusive heterosexuality and 6 is exclusive homosexuality). With behavior, 24 of 30 (80 percent) were bisexual to homosexual. One boy at 18 years was transsexual. None of the boys was sexually aroused by cross-dressing.

Behavioral changes were evident in the boys during the years prior to the determination of sexual orientation. There was a general lessening of specific cross-gender behaviors and a denial by all but one of the continued desire to be female.

There are several possible reasons for the considerable behavioral change from the time of initial evaluation to adolescence, with or without formal treatment. When deciding to seek professional counsel, parents have usually concluded, or may conclude after consultation, that their child's extensive cross-gender behaviors should be limited or eliminated. The child is also receiving negative reactions from age-mates for obviously cross-gendered activities. The developmental course of activity preferences in typical children also dictates change. Because even typical girls play less with dress-up dolls as they get older, cross-gendered boys can also be expected to show less interest.

There have been no large studies of atypical females evolving into adolescence. A community-based sample of 50 tomboys found that the girls shared some (but not all) features with girls with gender identity disorder. Because federal research funding was not forthcoming, no follow-up data are available for this unique sample. Confusion and controversy have resulted from the inclusion of gender identity disorder as a childhood mental disorder and the finding from longitudinal research that most such children (especially boys) become homosexual or bisexual adults and on that basis do not meet the criteria for mental disorder. Notice must be given to the cardinal complaint of gender identity disorder, which is gender dysphoria or discontent with one's birth sex. When such children grow out of their dysphoria they no longer meet the criteria for gender identity disorder because erotic attraction to same-sex partners is not a disorder. However, if dysphoria endures, it can be diagnosed as gender identity disorder or transsexualism.

COURSE AND PROGNOSIS IN ADULTS

Biological Males Sexually Attracted to Males and Biological Females Sexually Attracted to Females Biological males sexually attracted to males and biological females sexually attracted to females are similar in course and prognosis. In many regards the characteristics of the one group are mirror images of the characteristics of the other.

Course Adult homosexual gender-dysphoric persons primarily represent that fraction of gender identity disorder children who did not normalize in gender identity by the end of adolescence. The course of the disorder is continuous, although certain manifestations of it may be driven underground in late childhood or the early teens. A feminine boy, for example, may cross-dress frequently in connection with fantasy play; cease cross-dressing entirely from junior high school to early adulthood as he attempts to fit into society or at least to minimize conflicts with family and peers; and then resume cross-dressing in his 20s with the intention of pursuing sex-assignment surgery.

Homoerotic feelings begin in puberty. Some adolescents label themselves initially as homosexual and seek companionship in a gay crowd. They find they do not really fit in there either, although some adolescent male transsexuals continue to socialize in the homosexual drag circuit simply because it offers the only available supportive environment. Eventually, homosexual gender-dysphoric persons of both sexes begin to label their erotic attractions in terms of their subjective gender identity. Thus, for example, a female-to-male transsexual will assert that her romantic interest in another woman is heterosexual not lesbian because inside she is really a man.

In young adulthood there is often an increase in cross-dressing and in attempts to pass as the opposite sex. Some patients have already moved into the cross-gender role full-time by the time they present for clinical attention. These developments reflect freedom from parental controls and increased opportunities for self-expression more than intensification of the gender dysphoria. The homosexual type (biological male sexually attracted to males and biological female sexually attracted to females) of gender identity disorder does not have the character of a progressive disorder. In young males, adoption of the female role is sometimes accompanied by withdrawal from society at large into a subculture consisting mostly of other male-to-female transsexuals.

At initial presentation female patients more often than males are involved in love relationships with same-sex partners. The female partner typically concurs with the patient's self-evaluation that she is really a man and reports that she perceives her lover as a man without a penis. Such partners are often anxious to be part of the

transsexual's evaluation.

Prognosis Young adults who are still living in the original gender role when they first present sometimes relinquish the desire for sex reassignment and make a satisfactory adjustment to a homosexual lifestyle. In most cases, however, the disorder is chronic.

Impairment and Complications Peer ostracism often makes school attendance so difficult for biological males with gender identity disorder, sexually attracted to males, that they drop out of high school without graduating. Lack of education and job skills contribute to chronic unemployment and further marginalization of the individual in society. Societal rejection and its sequelae may lead to prostitution or to drug and alcohol abuse. Gender identity disorder has fewer such outcomes for women, possibly because of society's greater tolerance of cross-gender behavior in girls and women.

Biological Males Sexually Attracted to Females, Both Sexes, or Neither Sex There are marked and probably qualitative differences between the type of gender identity disorder that may occur in homosexual males and females (biological males sexually attracted to males and biological females sexually attracted to females) and the types that occur in nonhomosexual males (biological males sexually attracted to females, sexually attracted to both, and sexually attracted to neither). The differences in their natural histories suggest that homosexual and nonhomosexual gender identity disorders are etiologically different conditions.

Course In terms of observable symptoms, nonhomosexual gender identity disorders may be characterized as progressive disorders with an insidious onset. The course is fairly continuous in some cases; in others, the intensity of symptoms fluctuates to the point that the course might be called episodic. The subjective experience is of a lifelong struggle with feminine longings that change their focus from time to time and may temporarily recede in the face of conflicting desires, but which have always been present in one form or another from the early years.

In most cases the first outward manifestation of the disorder is surreptitious cross-dressing in childhood (e.g., in mother's or sister's clothing). Many men also report that they first began wishing they were female during that period. The boyhood behavior does not, however, exhibit the pervasive pattern of effeminacy required for a childhood diagnosis of gender identity disorder. At puberty, cross-dressing begins to elicit penile erection, and for the next few years or decades, the individual may qualify for a diagnosis of transvestic fetishism. In his 20s or 30s, the person's penile response to cross-dressing begins to wane while at the same time his desire to have a woman's body grows stronger and more insistent. In general, the cross-gender wishes attain the highest intensity by the mid-30s and remain about the same thereafter. It must be emphasized, however, that the course is highly variable.

Most men with gender identity disorder, sexually attracted to females, sexually attracted to both, or sexually attracted to neither establish relationships with women at some point in their lives and many marry and father children. Those who fall in love with a woman often report that, during the early days of the romance, they lose their interest in cross-dressing or surgical sex reassignment. When the relationship becomes routine and the initial excitement subsides, however, the desire to dress or live as a woman reasserts itself.

The course of nonhomosexual gender identity disorders may also be punctuated by periods of increased symptomatology. Men of that type occasionally present during an episode of intensified gender dysphoria, with anguished longings to be female accompanied by frustration and despair at their male state.

Prognosis The disorder tends to be chronic. Its tendency towards cyclical variation in many cases may mislead the patient or his therapist into thinking that the patient has been cured when he would be better regarded as in remission. As a group, nonhomosexual gender-dysphoric patients are less likely than homosexual gender-dysphoric patients to pursue surgical sex reassignment to completion. Many simply learn to live with their feelings. Acute episodes of intensified gender dysphoria usually resolve to the preepisode level after a few months, and patients should be discouraged from taking steps toward establishing themselves as females during such periods.

Impairment and Complications Nonhomosexual gender identity disorders tend to interfere with heterosexual relationships. Feelings of sexual interest aroused by the sight of a beautiful woman may turn, in the next instant, into feelings of envy. Heterosexual intercourse may require the person to fantasize that he is the woman and his partner is the man or that he and his partner are two women having lesbian relations; even with the help of such autogynephilic fantasies, erectile problems are common.

Gender identity disorders commonly result in marital breakdown, either because the husband wishes to be free to pursue sex reassignment or because the wife can no longer tolerate her husband's cross-dressing or other cross-gender behaviors. Men contemplating or undergoing divorce often feel considerable guilt about the effect of their behavior on their children and anxiety about their prospects for continued access to them. These concerns are often part of the clinical picture in patients of the nonhomosexual type first presenting for assessment.

TREATMENT

Children Evaluating the effects of intervention is difficult. Whereas short-term behavioral changes toward more conventional gender-typed activities are described, few data are available about adult sexuality. In a prospective study, a subsample of boys with gender identity disorder was treated by several therapists with a variety of approaches, including psychoanalysis, family therapy, individual psychotherapy, and behavior modification. With each intervention the boys showed lessening of cross-gender behaviors. At follow-up none expressed the desire to be female. However, the rates of homosexual or bisexual orientation did not differ from those boys who had no formal treatment, who also showed a lessening of cross-gender behaviors.

At present there is no convincing evidence from any controlled research that psychiatric or psychological intervention with children with gender identity disorder affects the direction of subsequent sexual orientation. Transsexualism, however, may be affected. Transsexuals are unable to cope socially as persons of their anatomical sex. The treatment of gender identity disorder in children is directed largely at developing social skills and comfort in the sex role dictated at birth. To the extent that treatment is successful, transsexual development is interrupted. Clinicians experienced in the treatment of gender dysphoria in adulthood can attest to the extraordinary difficulties faced by such patients. It is the rare transsexual who does not acknowledge that life would have been better if the overwhelming discontent over birth sex could have been overcome in childhood. The low prevalence of transsexualism in the general population, however, even in the special population of cross-gendered children, thwarts the testing of this.

The most systematic attack on specific cross-gender behaviors is the behaviorist approach. Specific cross-gender behaviors in boys have been substantially reduced by the use of a token economy or differential social reinforcement. To the extent that such changes reduce stigma and enhance the boy's self-image, the results could be considered positive. However, one goal of George Rekers, the principal proponent of that type of intervention, appears not to be realized. Rekers wished to prevent the boy from yielding to “homosexual temptation” with homosexuality having “been sold to the unwary public as a right between consenting adults.” At least two graduates of that program are bisexual. Systematic follow-up of adult sexual orientation on the others in that series has not been reported, although it is over two decades since many were treated. One current treatment approach is eclectic in that it addresses the child's interaction with each parent, the child's perception of sex roles, and the child's relations with peers.

In a typical case it has been a year or more during which parents have observed behaviors in their child that constitute gender identity disorder. During that time, one parent may have taken a firmer stand against some of the behaviors, but for the most part the children have not been interrupted in these activities. Parents have typically been uncertain or ambivalent about its meaning, and until recently, if at all, the child has not been aware that his or her parents object to the atypical behavior. A child may interpret a parent's neutral stance to atypical behaviors as positive. Some parents may have begun to be concerned, but they have been advised by preschool teachers that the behaviors are normal and should not be discouraged. A parent who is concerned about the excessive nature of the child's cross-gender behaviors may have difficulty convincing the school of that child's special needs, notwithstanding any ideal of androgyny. The family doctor may dismiss the behavior as a “passing phase.” A grandparent or housekeeper may be supportive of the child's cross-gender activities or at least may not discourage them, again undermining efforts at behavioral change at least toward gender-neutral activities. Initial limit-setting by parents usually meets with resistance and testing by the child. It may result in behaviors continuing in secret. Some children will relabel the cross-gender behaviors as gender-appropriate. For example, a long, flowing robe that was previously a princess gown is now a Superman cape.

One strategy of therapy looks to the child's level of cognitive development. Young children paint the world in black and white. In the area of gender, there are no grays. A child who does not like the activities he or she associates exclusively with that child's sex concludes that being the other sex is the only solution. A girl will say that only boys play sports or a boy will say boys play too rough. Grays must be introduced. Boys need to know that they can participate in sedentary play with other children, both boys and girls. Girls needs to know that girls play sports and can be as good as or better than many boys. Parents should find children who

demonstrate those behaviors to play with their own child.

Goals with parents should be set early. Many parents are motivated to the initial evaluation by fears that their children will become transsexual or homosexual. Parents should be redirected from hypothetical concerns of decades ahead to the immediacy of the child's life. At present, the child is unhappy being a boy or a girl, and the focus should be on making the child content with who he or she is. In the immediate term, the child is experiencing social stigma. The child should be integrated more effectively into the peer group.

It has been demonstrated that by ages 4 to 5 boys and girls differ in their manners of walking, running, throwing a ball, and narrating a story. Cross-gender gestures and mannerisms should be pointed out to the child. The reason given to the child and parents for intervention is that otherwise the child will be teased. Children are often unaware of cross-gender mannerisms, but alerting them can help bring them into consciousness and under control.

A boy with gender identity disorder typically has a strained relationship with his father. Identifying cause and effect in the distant father-son relationship is difficult. Some fathers were not available to their sons in the earliest years, and the boy gravitated toward his mother's activities. Then the father found that the boy did not respond to his belated attempts to engage him in sports or other activities. Alternatively, a father may have been available from the outset, but the child was temperamentally attuned to his mother's activities. The father became discouraged and his attention focused elsewhere, perhaps to another sibling with whom he shared interests. The need for a positive father-son experience must be emphasized to fathers. Nonathletic activities can be mutually enjoyable. Taking the son to work provides a better image of who father is. Board games, video games, and a shared father-son activity, such as Indian Guides, a program sponsored by the YMCA that emphasizes crafts and camping, can be helpful. The father's busy work schedule must be compromised lest the best years of their relationship be sacrificed. Whatever the outcome of the boy's sexuality, a more positive father-son relationship is to be sought; it is good for both father and son.

The child may believe that the parents wanted a child of the other sex. Sometimes parents did and conveyed the wish to the child. Parents must convey the message to the child that this is not so (if it ever was) and that they are happy with the sex of their child.

Young children should be taught that sex is irreversible. Not yet having achieved gender constancy at ages 4 to 6, they may think that by cross-dressing or changing hair length they change their sex. They should know the anatomical differences between the sexes and that superficial changes will not achieve their goal. Older children who are aware of genital differences and are cognitively more advanced may also be sophisticated about sex-change surgery. The parade of transsexuals on television has educated older children so that the clinician's statement of irreversibility of sex can be challenged.

The same treatment strategies used for boys with gender identity disorder are applicable to girls. Parental responses that have been supportive of behaviors causing the girl to be stigmatized can be interrupted. Same-sex peer group experiences are to be encouraged. Where a swagger has evolved and teasing results, it should be modified as are the extensive feminine mannerisms in a boy.

No hormonal or psychopharmacological treatments for gender identity disorders in childhood have been identified.

In recent years controversy over treatment of gender identity disorders has increased. The Human Rights Commission in the City and County of San Francisco in September 1996 issued a resolution condemning the use of a gender identity disorder diagnosis against children and young people. It contends that the diagnosis of gender identity disorder in children and young people was established to identify “pre-homosexual” and “pre-transsexual” children to prevent them from growing up to be gay or transsexual. The resolution also asserted that gender identity disorder of childhood and adolescence is used to stigmatize children and young people who have been discriminated against because of gender nonconformity. The diagnosis of gender identity disorder of childhood and adolescence was seen as reinforcing bias against gender-atypical young people.

From this perspective, the Commission called on the American Psychiatric Association and Psychological Association to oppose treatment designed to “manipulate a young person's sexual or gender identity” and for a ban on funding for treatment. On the other hand, a transgendered child psychiatrist sees this perspective as myopic and as ignoring the extreme distress continuing for decades in the untreated or unchanged gender-dysphoric child.

Although critics of treatment for childhood gender identity disorder currently point to the high association between this disorder and later homosexual orientation (itself not a disorder), the “corollary” that gender identity disorder was introduced as a substitute for homosexuality in DSM-III is factually wrong. The psychiatrists who were the strongest advocates for omitting homosexuality from DSM and who were very concerned with the stigma of pathologizing homosexual behavior also agreed with the new diagnosis of gender identity disorder. The reason that gender identity disorder was included is that it was consistent with the new DSM criteria for mental disorder, namely subjective distress and social limitations. Children with gender identity disorder are unhappy with their gender and are socially stigmatized.

Whether or not someone else agrees, parents have the legal right to bring a child for therapy to modify behavior they disapprove of and with the goal of preventing a later behavior of which they disapprove. Under American law a parent, for example, may withdraw a child from a state mainstream school in the child's early teens to further education only in Amish ways. A parent may require a child to observe religious dietary laws or attend religious services on a regular basis and not to participate in a sex education program. Therefore, parents may also engage a psychotherapist to treat what is diagnosed as gender identity disorder or attempt to change their child's behavior without professional intervention.

Adolescents Young persons whose previous gender identity disorder has normalized (with or without treatment) by the end of puberty may experience new conflicts when homosexual feelings emerge in adolescence. Although homosexuality is not a psychiatric disorder, it may be a source of anxiety to the adolescent and cause intrafamily conflict. Teenagers should be reassured about the prevalence and the nonpathological aspects of a same-sex partner preference. They should understand that sexual experimentation in adolescence need not reflect their final orientation. Parents must also be informed of the nonpathological nature of a same-sex orientation. They must be dissuaded from obsessing over (or even asking the question) “Where did I go wrong?” The goal of family intervention is to keep the family stable and to provide a supportive environment for the adolescent.

Adolescents whose gender identity disorders have persisted beyond puberty present different treatment problems. One problem in treating adolescent transsexuals is how to manage the rapid emergence of unwanted secondary sex characteristics. Legal issues may confront the clinician who would prescribe hormone-inhibiting substances or later contrasex hormones to thwart that development, even with parental approval. Social cross-gender living is a possible alternative. Some schools have permitted teenagers to enroll as opposite-sex students.

Adults Adult patients with gender identity disorders present with various agendas. Some are seeking help in suppressing their cross-gender feelings, some are gathering information about gender identity disorders and treatment options, and still others come with straightforward requests for surgical sex reassignment. For all the patients seeking a cure or information, and many of those seeking surgery, the first consideration should be to help the patient reconcile to the original gender role or at least learn to function reasonably well in it.

Psychotherapy At present, only psychological therapies are available to help patients accept their biological sex. No drug treatment has been shown to be effective in reducing cross-gender desires per se.

There are no general guidelines for individual psychotherapy with gender-dysphoric patients. Some authorities recommend that when treating patients with unrealistic hopes of sex reassignment surgery, it is best not to confront the patient on that point directly. Patients may be more open to critical examination of their goals and alternatives if they perceive the therapist as open-minded about the surgical option. There are exceptions: Some patients are relieved to be told frankly by an experienced professional that they are not suitable—or ready—for sex reassignment surgery, and they drop the idea without resistance.

Psychotherapy may be appropriate either when the patient requests it or when the patient willingly accepts a recommendation for psychotherapy from the attending clinician. Some gender identity clinics go further than this and require all patients to undergo psychotherapy as a condition for approving them for sex-reassignment surgery. There is, however, variability on this point and other reputable clinics do not have such a requirement.

Gender identity disorder patients are often amenable to group therapy, especially if the group is homogeneous. Mixing patients who are attempting to become reconciled to their original gender roles with patients who are pursuing sex reassignment may make it more difficult to promote group cohesion.

Medical Treatment When the patient's gender dysphoria is severe and intractable, sex reassignment may be the best solution. The first medical intervention in this process is hormone therapy. Dosage should be set with endocrinological consultation. Biological males are usually treated with daily doses of oral estrogens. These

produce breast enlargement, which continues for about 2 years at any given dose. The final amount of gynecomastia varies greatly among patients. The other major effects of estrogen treatment are testicular atrophy, decreased libido, and diminished erectile capacity. There may be a slight decrease in the growth of facial and body hair together with arrest of male-pattern baldness.

Biological females are treated with biweekly or monthly intramuscular injections of testosterone. Because the effects of exogenous testosterone are more profound than those of estrogen, clinicians should be more conservative about commencing females on hormone treatment. The pitch of the voice drops permanently into the male range as the vocal cords thicken. The clitoris enlarges to about three times its pretreatment length, with a growth plateau of 1 year; that is often accompanied by increased libido. A full, thick beard often develops. Male pattern baldness may develop, and acne may appear. Ovarian function is suppressed, with the menses ceasing within 3 months of the start of treatment. When treatment of adolescents is contemplated a pediatric endocrinologist should be consulted.

The second major stage in the medical treatment of transsexualism is sex reassignment surgery. All major gender identity clinics in North America and western Europe require their patients to live full-time in the cross-gender role for some time—usually 1 to 2 years—prior to surgery. Most gender identity clinics further require that patients either maintain gainful employment, attend school, or perform bona fide charity work in the cross-gender role during the probationary period. It is a good idea to require patients to provide documentation verifying such activities; these documents may take various forms, including pay receipts, school transcripts, or letters on known institutional letterheads.

The requirement of work, school, or charity in the cross-gender role is intended to demonstrate that the patient is capable of interacting successfully with members of the general public in the cross-gender role. Without this requirement, it would be possible for patients who claim to be cross-living to be in fact leading completely reclusive lives, restricting their activities to the privacy of their home, to the company of a few sympathetic friends or relatives, or to contacts with members of the helping professions. A reclusive life-style does not demonstrate to the clinician that the patient has the ability to function normally in society as a member of the other sex. It also does not ensure that the patient has the genuine experience of what it is like to live in society as the other sex or test the patient's expectations and fantasies against the reality of everyday living. Because of the importance of documented evidence that the patient can function successfully in society at large as a member of the other sex—which includes functioning successfully in public and among strangers—it is not sufficient for male-to-female patients to be full-time housewives during the presurgical probationary period or for female-to-male patients to be full-time househusbands.

Sex reassignment surgery for biological males consists principally of vaginoplasty. Operative techniques vary, mainly in the method of obtaining material to line the neovagina. One technique uses a section of the rectosigmoid colon. The standard methods involve lining the neovagina with penile skin flaps, scrotal skin flaps, and free grafts from the thigh, in various combinations. Occasional postoperative complications include urethral strictures, caused by retraction of the urethra, and rectovaginal fistulas. Vaginal stenosis sometimes occurs; this is usually caused by inadequate postoperative care of the neovagina by the patient. Inadequate vaginal width or depth may make penile penetration difficult.

Surgical reassignment procedures for females are more variable. Construction of a male chest contour is virtually always performed. Panhysterectomy and bilateral salpingo-oophorectomy are usually carried out, although some patients, satisfied with the menstrual suppression produced by testosterone therapy, do not request those procedures. Phalloplasty is the least commonly performed, partly because it is relatively expensive and partly because it is not as successful, cosmetically or functionally, as vaginoplasty. Surgical techniques for phalloplasty have recently improved, however, so that picture may be changing.

An alternative to phalloplasty is a procedure most commonly known as metoidioplasty. The administration of testosterone medication to female-to-male transsexual stimulates growth of the clitoris. In some patients the degree of enlargement is sufficient for the hypertrophied clitoris to be used as a penis. Metoidioplasty capitalizes on this phenomenon to construct a neophallus from the clitoris, using techniques similar to those used for correction of severe hypospadias in biological males with transposition of the scrotum and penis. In this procedure the clitoris is moved anteriorly to the penis position, the labia majora are moved posteriorly and fused together into a scrotal configuration, and the chordee is released and excised according to the principles of hypospadias surgery. The urethra is formed from a vaginal flap and the inner part of the labia minora. Testicular prostheses are inserted into the neoscrotum, and the feminine mons veneris is reduced with liposuction or direct excision of fat.

The best technique for constructing a male chest contour varies according to the size of the patient's breasts. For small-breasted females, the surgeon may perform a subcutaneous mastectomy through a keyhole incision without disturbing the nipple-areola complex. Large-breasted patients require breast amputation and replacement of the nipple as a free graft. Postoperative complications of the latter procedure include obvious chest scars and widening of the nipple-areola grafts.

Additional plastic surgery is commonly sought by biological males, rarely or never by biological females. Male patients most often undergo thyroid cartilage shave (Adam's apple reduction), rhinoplasty (to create a more feminine nose), and breast implants. They require facial hair electrolysis and may need vocal pitch surgery.

Clinical management of transsexuals incarcerated in the prison system—in particular, prescribing hormones for the inmate or recommending sex reassignment surgery—poses special problems. Many jurisdictions follow a formal or informal policy of “freeze-framing” the inmate at the hormonal and surgical status at the time of incarceration. The rationale would appear to be that the prison environment is not representative of society at large, so there can be no adequate evaluation of an individual's ability and motivation to function successfully in the cross-gender role as normally assessed during the presurgical probationary period.

Treatment Outcome Numerous studies have investigated the postoperative adjustment of reassigned transsexuals, primarily among patients assessed and approved for surgery by established gender identity clinics. The findings are generalizable only to properly screened patients. The most reliable conclusion is that the overwhelming majority of postoperative transsexuals are content with their decision to undergo sex reassignment.

Outcome studies as a whole suggest that surgical sex reassignment produces additional improvements in psychosocial adjustment. The main areas of benefit are interpersonal relationships and psychological symptomatology, especially morale and mood. It should be noted that only one outcome study has included random assignment to surgically treated and (waiting list) control groups. Its findings bolster the conclusion that postoperative improvements in social integration, sexual adjustment, and psychological symptomatology can be attributed to the surgical intervention.

The available evidence for socioeconomic improvement is weaker, perhaps because most studies fail to look at outcome variables for males and females separately. Studies that have examined socioeconomic change separately for male and female transsexuals have found that the socioeconomic status of biological females improves when they move into the male role whereas the socioeconomic status of males worsens when they adopt the female role.

It has been shown that in male-to-female transsexuals the cosmetic and functional adequacy of surgical interventions affects the self-image of transsexuals and the degree to which they and their partners are reminded of the sex-reassigned status. Thus, the better the surgical result—other things being equal—the better the postoperative adjustment.

Finally, that sex reassignment surgery has little or no effect on major mental illness or serious character pathology. Patients with preoperative histories of psychiatric hospitalization for depression or multiple suicide attempts are liable to be at risk for further episodes after surgery.

Intersex Disorders Patients who present with uncertainty or confusion about their correct gender or are discontented with the gender in which they have been living may have an intersexed physiological status. Differences between gender identity patients with and without physical intersexuality suggest distinct causes of gender identity complaints in the two groups. There are, on the other hand, occasional patients (e.g., adult men with Klinefelter's syndrome) who appear similar in their clinical presentations to otherwise comparable men with normal karyotypes. For this reason, there is continuing discussion among specialists about whether the presence of intersexuality should preclude the diagnosis of gender identity disorder. For research purposes, intersexed patients should be excluded from gender-dysphoric samples. According to DSM-IV, intersex conditions are an example of a gender identity disorder not otherwise specified (Table 19.3-3).

This category is included for coding disorders in gender identity that are not classifiable as a specific gender identity disorder. Examples include 1. Intersex conditions (e.g., androgen insensitivity syndrome or congenital adrenal hyperplasia) and accompanying gender dysphoria. 2. Transient, stress-related cross-dressing behavior. 3. Persistent preoccupation with castration or penectomy without a desire to acquire the sex characteristics of the other sex.

Reprinted with permission from DSM-IV: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 19.3-3 DSM-IV Diagnostic Criteria for Gender Identity Disorder Not Otherwise Specified

Congenital Virilizing Adrenal Hyperplasia Formerly called *adrenogenital syndrome*, this disorder is characterized by an enzymatic defect in the production of cortisol by the adrenal gland. As a result, excessive androgenic adrenal hormone production begins prenatally. At birth, affected females show varying degrees of genital virilization; after birth, excessive androgen production can be largely controlled by cortisone.

With diagnosis of congenital virilizing adrenal hyperplasia neonatally, male or female children appear to develop a gender identity consistent with their chromosomal and gonadal sex and sex of rearing. Girls may show more than typical boy-type play behaviors. In earlier cases before the diagnosis was made neonatally and when the diagnosis for a woman was not made in childhood and socialization was as a man, a male identity typically evolved. There may be higher rates of homosexual behavior in affected women, whether or not treated with cortisone in childhood.

Turner's Syndrome One sex chromosome is missing in Turner's syndrome, so that the sexual karyotype is simply X (designated XO). Affected children have poorly developed ovaries and possibly other anomalies, including an unusual appearance of the chest and neck. Their genitalia are female at birth. They develop into short women who require replacement estrogen for female secondary sex characteristics. Identity is usually female and gender identity conflict is rare, but they may experience psychosexual conflict over infertility.

Klinefelter's Syndrome An additional X chromosome is present in this condition along with the typical XY male pattern. At birth affected infants appear to be normal males. There may be excessive gynecomastia in adolescence, and the testes are small, usually without spermatogenesis. Testosterone levels are low and the body habitus is eunichoid. A higher rate of gender identity disorder has been suggested in affected persons, but unrepresentative sampling clouds the issue.

5-a-Reductase Deficiency An enzymatic defect here prevents the conversion of testosterone to dihydrotestosterone, which is required for prenatal virilization of the genitalia. Thus, the XY-affected individual is born with female-appearing genitalia, although there is some evidence of abnormality. At puberty extensive virilization occurs from testosterone and some dihydrotestosterone; there is phallic growth and a male chest pattern.

Gender identity development is the subject of controversy. Many of the earliest described cases were raised as girls who with virilization during adolescence evolved a male identity. They were later heterosexually active as men. This transition was variously interpreted from neuroendocrine and social learning perspectives. It was posited that testosterone organized the prenatal central nervous system to mediate a later male identity, although that hormone was inadequate to virilize the genitalia in utero. Alternatively, it was posited that social pressures on these male-appearing adolescents to act like men and avoid the stigma of homosexuality forced them into their masculine heterosexual status. Subsequent generations born with the disorder are recognized as those who will virilize at puberty and so are raised with that hermaphroditic identity. When raised as men, perhaps with androgen supplementation, they function as relatively normal men, although their phallus is not entirely normal. When castrated in infancy and raised as girls, they may be markedly tomboyish in childhood. Insufficient numbers of cases are available for predicting their adult psychosexual status, but some appear to be feminine heterosexual women, without menses.

Androgen Insensitivity Syndrome Formerly called testicular feminization, androgen insensitivity syndrome is a disorder of metabolism in the XY individual where tissue cells are unable to use testosterone or other androgens; consequently, the person appears to be a normal female at birth. Testes are undescended and no internal female reproductive structures are present. Puberty brings female-type breast development (from the conversion of testosterone to estradiol). A vagina needs to be constructed. Gender identity usually evolves as female, with a sexual interest in males. There may be psychosexual conflict over the absence of menses and infertility.

SUGGESTED CROSS-REFERENCES

Related discussions include Section 19.1a on normal human sexuality and sexual dysfunctions, Section 19.1b on homosexuality, Section 32.2 on normal child development, and Section 32.3 on normal adolescent development. Transvestic fetishism is discussed in Section 19.2 on paraphilias, and intersex disorders are discussed in Section 25.6 on endocrine and metabolic disorders.

SECTION REFERENCES

Bailey JM, Pillard R: A genetic study of male sexual orientation. Arch Gen Psychiatry *48:* 1089, 1991.

Bailey JM, Zucker KJ: Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. Dev Psychol *31:* 43, 1995.

Benjamin H: *The Transsexual Phenomenon*. Julian Press, New York, 1966.

Blanchard R: Cross-gender fetishism scale. In *Handbook of Sexuality Related Measures*, CM Davis, WL Yarber, R Bauserman, G Schreer, S Davis, editors. Sage, Thousand Oaks, CA 1998.

Blanchard R: Partial versus complete autogynephilia and gender dysphoria. J Sex Marital Ther *19:* 301, 1993.

Blanchard R: The she-male phenomenon and the concept of partial autogynephilia. J Sex Marital Ther *19:* 69, 1993.

Blanchard R: Varieties of autogynephilia and their relationship to gender dysphoria. Arch Sex Behav *22:* 241, 1993.

Blanchard R, Clemmensen LH, Steiner BW: Heterosexual and homosexual gender dysphoria. Arch Sex Behav *16:* 139, 1987.

Blanchard R, Collins PI: Men with sexual interest in transvestites, transsexuals, and she-males. J Nerv Ment Dis *181:* 570, 1993.

Blanchard R, Klassen P: H-Y antigen and homosexuality in men. J Theor Biol *185:* 373, 1997.

*Blanchard R, Steiner BW, editors: *Clinical Management of Gender Identity Disorders in Children and Adults*. American Psychiatric Press, Washington, DC, 1990.

Blanchard R, Zucker KJ, Cohen-Kettenis P, Gooren LJG, Bailey M: Birth order and sibling sex ratio in two samples of Dutch gender-dysphoric homosexual males. Arch Sex Behav *25:* 495, 1996.

Coates S, Person E: Extreme boyhood femininity: Isolated behavior or pervasive disorder? J Am Acad Child Psychiatry *24:* 702, 1985.

*Daskalos C: Changes in the sexual orientation of six heterosexual male-to-female transsexuals. Arch Sex Behav *27:* 605, 1999.

Diamond M, Sigmundson HK: Sex reassignment at birth: Long-term review and clinical implications. Arch Pediatr Adolesc Med *151:* 298, 1997.

Dörner G, Rhode W, Stall F: A neuroendocrine predisposition for homosexuality in men. Arch Sex Behav *4:* 1, 1975.

Dörner G, Geiser T, Ahrens L: Prenatal stress and possible aetiogenic factor for homosexuals in human males. Endocrinologie *75:* 365, 1980.

Futterweit W: Endocrine therapy of transsexualism and potential complications of long-term treatment. Arch Sex Behav *27:* 209, 1998.

Green R: Atypical psychosexual development. In *Child and Adolescent Psychiatry*, ed 3, M Rutter, L Hersov, E Taylor, editors. Blackwell Scientific, London, 1994.

*Green R: *Sexual Identity Conflict in Children and Adults*. Basic Books, New York, 1974.

*Green R: *The "Sissy Boy Syndrome" and the Development of Homosexuality*. Yale University Press, New Haven, CT, 1987.

Green R, Fleming DT: Transsexual surgery follow-up: Status in the 1990s. Annu Rev Sex Res *1:* 63, 1990.

Green R, Money J: Incongruous gender role: Nongenital manifestations in prepubertal boys. J Nerv Ment Dis *131:* 160, 1960.

Green R, Money J, editors: *Transsexualism and Sex Reassignment*. Johns Hopkins Press, Baltimore, 1969.

Hamer D, Hu S, Magnuson V, Hu N, Pattatucci A: A linkage between DNA markers on the X chromosome and male sexual orientation. Science *261:* 321, 1993.

Isay R: Remove gender identity disorder in DSM. Psychiatr News *13* (9):32, 1997.

LeVay S: A difference in hypothalamic structure between heterosexual and homosexual men. Science *253:* 1034, 1991.

*Mate-Kole C, Freschi M, Robin A: A controlled study of psychological and social change after surgical gender reassignment in selected male transsexuals. Br J Psychiatry *157:* 261, 1990.

Money J, Hampson J, Hampson J: An examination of some basic sexual concepts. Bull Johns Hopkins Hosp *97:* 301, 1955.

Person E, Ovesey L: The transsexual syndrome in males. I. Primary transsexualism. Am J Psychother *28:* 4, 1974.

Person E, Ovesey L: The transsexual syndrome in males. II. Secondary transsexualism. Am J Psychother *28:* 174, 1974.

Rekers G: *Shaping Your Child's Sexual Identity*. Baker Book House, Grand Rapids, MI, 1982.

Schlatterer K, Yassourdis A, von Werder K, Poland D, Kemper J, Stalla G: A follow-up study for estimating the effectiveness of a cross-gender hormone substitution therapy on transsexual patients. Arch Sex Behav *27* :475–492, 1998.

Stoller R: *Sex and Gender*. Science House, New York, 1968.

Stoller R: *Sex and Gender*, vol 2, *The Transsexual Experiment*. Aronson, New York, 1975.

Ward I: The prenatal stress syndrome. Psychoneuroendocrinol *9:* 3, 1984.

Williams K, Green R, Goodman M: Patterns of sexual identity development: A preliminary report on the "tomboy." In *Research in Community and Mental Health*, R Simmons, editor. JAI Press, Greenwich, CT, 1979.

Zhou J, Hoffman M, Gooren L, Swaab D: A sex difference in the human brain and its relation to transsexuality. Nature *378:* 68–70, 1995.

*Zucker K: Gender identity disorder in the DSM-IV (Letter to Editor). J Sex Marital Ther *25:* 5, 1999.

*Zucker K, Green R: Gender identity disorders in children and adolescents. In *Child and Adolescent Psychiatry*, M Lewis, editor. Williams & Wilkins, Baltimore, 1991.

Zucker KJ, Green R: Psychological and familial aspects of gender identity disorder. In *Child and Adolescent Psychiatric Clinics of North America. Sexual and Gender Identity Disorders*, A Yates, editor. WB Saunders, Philadelphia, 1993.

Zucker K, Green R, Garofano C, Bradley S, Williams K, Rebach M, Sullivan C: Prenatal gender preference of mothers of feminine and masculine boys: Relation to sibling sex compositions and birth order. J Abnorm Child Psychol *22:* 1, 1994.

CHAPTER 20. EATING DISORDERS

KATHERINE A. HALMI, M.D.

The eating disorders anorexia nervosa, bulimia nervosa, and variants thereof have been identified in the history of western civilization for the past 15 centuries. The eating disorders are syndromes and are classified on the basis of the clusters of symptoms with which they present. The prevalence of anorexia nervosa and bulimia nervosa has increased since the late 1960s to the extent that they are now commonly encountered in clinical practice. Over a long-term course such as 30 years, 15 to 20 percent of anorexia nervosa patients will die as a result of the disorder. Another 25 percent remain chronically ill, either maintaining a low body weight or developing bulimia with wide fluctuations in weight. About 40 percent will eventually recover and the others will function well with mild eating disorder symptomatology. There is a high prevalence of mood, anxiety, and substance use disorders with the eating disorders. These comorbidities further impair the quality of life of the patient. The diverse nature of the eating disorders has stimulated the development of a variety of treatments.

DEFINITIONS

Anorexia nervosa is characterized by willful and purposeful behavior directed towards losing weight, weight loss, preoccupation with body weight and food, peculiar patterns of handling food, intense fear of gaining weight, disturbance of body image, and amenorrhea. About half of these persons will lose weight by drastically reducing their total food intake and some will also develop rigorous exercising programs. The other half of these patients will also rigorously diet but will lose control and regularly engage in binge eating followed by purging behaviors. Some patients routinely purge after eating small amounts of food.

Bulimia is merely a term that means binge eating, which is defined as eating more food than most people in similar circumstances and in a similar period of time, accompanied by a strong sense of losing control. When binge eating occurs in relatively normal weight or overweight persons who are also excessively concerned with their body shape and weight, and who regularly engage in behaviors to counteract the calorie gain in binges, the binge eating is in the context of the disorder known as *bulimia nervosa.*

HISTORY

Willful dieting leading to irreversible self-starvation was carefully documented in the fasting female saints of the middle ages. Princess Margaret of Hungary, who lived from 1242 through 1271, could be an example of anorexia nervosa—restricting type. She excelled in all her studies, in the chores of the monastery, and in the austerities of fasting and was never idle. She died with a clear alert mind and a poor wasted body at the age of 28. Catherine of Sienna could well be diagnosed with anorexia nervosa—binge eating purging type. She inserted and then withdrew a specific reed from her throat to self-induce vomiting. Whether dieting for sainthood or for thinness produces the same disorder remains a matter of controversy. The psychobiological vulnerability factors that induce the development of the anorectic symptom cluster are probably similar across the ages.

In the second half of the seventeenth century case descriptions of self-starvation appeared. One of these was of an 18-year old English girl named Martha Taylor who had lost her menses, stopped taking all solid food, and became emaciated. She was examined by John Reynolds who wrote, “most of these damsels fall into this abstinence between the age of 14 and 20 years.” Twenty-two years later Richard Morton described two cases of typical anorexia nervosa symptomatology and distinguished them from consumption.

In the nineteenth century Dr. Louis-Victor Marce of Paris described several cases of “young girls, who at the period of puberty, and after a precocious development, became subject to inappetance carried to the utmost limits.” He conceptualized a gastric nervous disorder becoming cerebro-nervous. In 1873 Sir William Gull of London and Ernest Charles Laseque almost simultaneously published papers on the description and treatment of “hysterical anorexia.” Gull recommended a treatment that to some extent is still followed: “the patient should be fed at regular intervals, and surrounded by persons who would have moral control over them; relations and friends being generally the worse attendance.” During this period very little attention was paid to these cases of fasting women in the United States. However, two neurologists, William Alexander Hammond and Silas Weir Mitchell described cases of great self-starvation. The term *anorexia nervosa* was commonly used after Gull's publications.

In the early twentieth century there was confusion between pituitary insufficiency and the disorder of anorexia nervosa. This was definitively clarified in a book by E.L. Bliss and C.H. Branch in which the current endocrine studies as well as the history and psychological descriptions of anorexia nervosa were summarized. Later in the 1950s Hilda Bruch described a core psychological facet of anorexia nervosa; “a paralyzing sense of ineffectiveness which pervades all thinking and activities.” She also laid the foundation for cognitive therapy by describing methods for treating the disturbed cognitions of such patients.

Bulimia nervosa was first described as a distinct syndrome in 1979 by Gerald Russell. Subsequent extensive studies of this disorder have continued to define it more precisely.

COMPARATIVE NOSOLOGY

The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) criteria for anorexia nervosa (Table 20-1) are very similar to those of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) in that they acknowledge the weight loss, the fear of becoming fat, and the amenorrhea. The criterion concerning body image disturbance is limited to the perception of being too fat. The ICD-10 criteria for bulimia nervosa do not clearly separate bulimia nervosa from anorexia nervosa—binge eating/purging type. There is also a lack of clear distinction between bulimia nervosa as defined in ICD-10 and another condition in this classification system labeled *hyperorexia nervosa.* The ambiguous and overlapping definitions of various bulimia subtypes listed in ICD-10 pose problems for research studies in this area; reliability and validity studies are needed to justify these classifications.

Table 20-1 ICD-10 Diagnostic Criteria for Eating Disorders

EPIDEMIOLOGY

Incidence rates are commonly expressed with eating disorders as the rate per 100,000 of a population per year. One of the major problems in studying the epidemiology of eating disorders over time is the change in criteria for these syndromes since the late 1960s. Despite this problem, studies suggest that the overall incidence of both anorexia nervosa and bulimia nervosa have increased in the past 50 years. There was a consistent increase in the registered incidence of anorexia nervosa over a period from 1931 to 1986 in cases presenting to the health care system in several industrialized countries. For example, a study in northeastern Scotland showed that between 1965 and 1991 there was almost a sixfold increase in the incidence of anorexia nervosa—from 3/100,000 to 17/100,000. A study conducted in southwest London between the period of July 1991 and June 1992 showed an incidence of anorexia nervosa to be 2.7 cases per 100,000 total population. In females aged 15 to 29 years the incidence was 19.2 cases per 100,000. In Rochester, Minnesota a recent study found an overall adjusted incidence for females of 14.6 per 100,000 per year; for men the corresponding figure was 1.8. When estimates are based on the population at large, the incidence rate of anorexia nervosa in industrialized countries is estimated at 8.1 per 100,000 population per year. A recent prevalence study of anorexia nervosa was found to be 20.2 cases per 100,000 population (0.02 percent total population) in the United Kingdom. In a recent study conducted in the United States, lifetime prevalence rates for anorexia nervosa were found to be 0.51 percent (narrowly defined) and 3.7 percent (broadly defined). A summary of numerous prevalence studies across European countries reported an average prevalence of anorexia nervosa using strict diagnostic criteria of 0.28 percent of young females.

There are very few studies of the incidence of bulimia nervosa. A large representative sample of the Dutch population in Holland showed an incidence of bulimia nervosa at 11.4 per 100,000 population per year during the period from 1985 through 1989.

A review of over 50 prevalence studies of bulimia nervosa conducted from 1981 through 1989 in Europe and the United States showed a remarkable consistency in finding that bulimia nervosa had a prevalence rate of about 1 percent in adolescent and young adult women. Partial eating disorders occur at a far greater frequency in the general community—in about 5 to 10 percent of young women. In a recent Canadian study the lifetime prevalence rate for bulimia nervosa was 1.1 percent for females and 0.1 percent for males. In industrialized countries the prevalence is about 1 percent for bulimia nervosa.

All studies show that eating disorders are rare among males. In clinical samples the male-to-female ratio for eating disorders lies consistently between 1:10 and 1:20. Studies conducted in the early 1970s and 1980s found a higher prevalence of eating disorders among the middle and upper classes; however, eight studies from the 1980s to the present failed to demonstrate this relationship. This is true of both anorexia nervosa and bulimia nervosa.

Transcultural studies show that anorexia nervosa is rare in non-Western developing countries. When non-Westerners are exposed to Western ideals of thinness, they are significantly affected. For example, a study of the prevalence of eating disorders among Greek and Turkish girls who remained in their homeland and those who were living in Germany showed twice the prevalence in the latter compared to the former group. A recent review of eating disturbances among American minority groups found that compared to Caucasian females, eating disturbances were equally common among Hispanic females, more frequent among Native Americans, and less frequent among black and Asian females. These studies also indicated that minority females who are younger, better educated, and more identified with white, middle-class values are at greater risk for developing an eating disorder.

ETIOLOGY

A specific cause and pathogenesis for the eating disorders anorexia nervosa and bulimia nervosa is unknown. The current concepts of the development of these disorders are best represented by a multidimensional model that emphasizes that dieting behavior is the instigating and common stimulus that leads to the development of the serious eating disorders. Behaviors and influences antecedent to the dieting experience can be categorized into problems of biological vulnerability, psychological predispositions, and societal influences. Disturbances in these categories propel the dieting behavior so that starvation effects, weight loss and nutrition effects, as well as psychological changes occur. Psychological and physiological reinforcement of the maladaptive eating behavior continues a sustaining cycle of core dysfunctional eating disorder behaviors.

Sociocultural Factors Slenderness as an expression of attractiveness is prominently emphasized in Western culture. Much willful dieting occurs for the purpose of being more attractive. Other willful dieting occurs for the purpose of being more professionally competent; for example, ballet dancers, gymnasts, jockeys, and wrestlers—all of whom have a high risk for developing an eating disorder. It has been hypothesized but not proven by systematic studies that some women diet strenuously in response to the conflict of professional and social demands and nurturing expectations.

Psychological Factors Earlier psychological theories on the etiology of anorexia nervosa centered mostly on phobic mechanisms and psychodynamic formulations. It has been postulated that anorexia nervosa constitutes a phobic avoidance response to food, resulting from the sexual and social tension generated by the physical changes associated with puberty. Psychodynamic theories have focused on fantasies of oral impregnation and dependent, seductive relationships with warm, passive fathers and guilt over aggression toward ambivalently regarded mothers. A cognitive and perceptual developmental defect has been theorized to be the cause of anorexia nervosa, with the disturbances of body image (denial of emaciation), disturbances in perception (lack of recognition or denial of fatigue, weakness, hunger), and a sense of ineffectiveness being caused by untoward learning experiences.

Comprehensive studies assessing personality traits before the onset of anorexia or bulimia nervosa are extremely difficult to conduct, thus, most studies assess these patients after recovery as well as during the illness. Recovered patients often continue to have obessional and inflexible thinking, social introversion, greater interpersonal insecurity and minimization of affect, excessive conformance, and more regimentation of behavior. An obsessional, perfectionistic personality is considered a risk factor for the restricting type of anorexia nervosa. These patients have often had a lack of childhood experiences that foster personal independence.

Although personality development can be strongly influenced by family interactions, studies that have sound methodology and specificity are difficult to conduct. Studies of family interactions during the acute stages of these disorders can give some idea of what factors may be contributing to the maintenance of the eating disorder. One carefully controlled study showed that families of bulimia patients had greater hostility, chaos, feelings of isolation, and substantial deficits in nurturing and empathy compared with control families. This same study showed that parents of daughters who had anorexia nervosa communicated mixed messages with antithetical meanings. These parents were often very affectionate as they controlled their daughter and suffocated her separateness.

The expression of personality is also influenced by comorbid psychiatric illnesses. Anorexia nervosa patients had a high rate of major depression (68 percent), anxiety disorders (65 percent), obsessive-compulsive disorder (26 percent), and social phobia (34 percent) in one large study. About one fourth of patients with the restricting type of anorexia nervosa have a Cluster C anxious personality disorders. About 40 percent of patients with the binge eating/purging type of anorexia nervosa have a cluster B, the impulsive cluster personality disorder diagnosis. They also have a high prevalence (32 percent of the anxious cluster of personality disorders. Bulimia nervosa patients have considerable comorbidity with major depressive disorder varying from 36 to 70 percent, substance abuse from 18 percent to 32 percent, and Axis II personality disorders from 28 percent to 77 percent, with a predominance of Cluster B (impulsive) personality disorders.

Biological Factors

Familial Genetic Factors Family studies of anorexia nervosa have shown a tendency to familial occurrence of this disorder. Female siblings of anorexia nervosa patients in one study had 6.6 percent risk of developing anorexia nervosa. Another study showed that the concordance for restricting anorexia nervosa was markedly higher for monozygotic twins (66 percent) than for dyzygotic twins (0 percent). In a large population-based twin registry study, the concordance for bulimia nervosa was significantly higher in monozygotic than in dyzygotic twin pairs. Several studies have found a familial aggregation of anorexia nervosa, bulimia nervosa, and variants of these disorders in the first-degree relatives of both anorexia nervosa and bulimia nervosa probands. Some authors have proposed that anorexia nervosa has a genetic predisposition that could become manifest under adverse conditions, such as inappropriate dieting or emotional stress. This genetic vulnerability might implicate a particular personality type, a general susceptibility to psychiatric instability (in particular, mood disorder), or may directly involve a hypothalamic dysfunction. Family studies of anorexia nervosa and bulimia nervosa have shown an increased frequency of mood disorder in the first-degree relatives of the eating disorder probands compared with the first-degree relatives of normal control subjects. Another study showed an increased prevalence of obsessive-compulsive disorder in the mothers of anorexia nervosa patients compared with the mothers of control persons. Thus, it is likely that a family history of mood disorders and perhaps of anxiety disorders or more specifically of obsessive-compulsive disorder places a person at risk for developing an eating disorder.

Hypothalamic-Neuromediator Factors A primary hypothalamic dysfunction in anorexia nervosa has received some support from the fact that many persons develop amenorrhea before weight loss has occurred, and the return of the normal menstrual cycles often lags behind the return to a normal body weight. The resumption of

menses in anorexia nervosa is associated with marked psychological improvement after weight gain. In the area of the paraventricular nucleus in the hypothalmus, neurotransmitters have complex interactions in regulating eating behavior. Norepinephrine can have an appetite-stimulating effect; serotonin increases satiation, decreases food intake, and modulates aggressive, impulsive, and obsessive-compulsive behaviors. Underactivation or overactivation of dopamine systems can produce eating deficits. Drugs and hormones are in vivo environmental stresses that interfere with dopamine transmission and can cause disinhibition of eating. Current studies have shown there is evidence for dysfunction in all three neurotransmitters in anorexia nervosa. Because of the prohibitive cost of investigating large numbers of people before they develop anorexia or bulimia nervosa, the only possible strategy is to compare the acutely ill patient with the recovered patient and controls. Recently more focus has been placed on serotonin because of its widespread influences on behaviors that are disturbed in anorexia nervosa. Long-term weight-restored anorexia nervosa patients in one study had an elevated cerebral spinal fluid (CSF) 5-hydroxyindoleacetic acid (5-HIAA) concentration compared with controls. This finding is compatible with the hypothesis that obsessional traits in anorexia patients are associated with a hyperfunctioning of this serotonergic neurotransmitter system. Corticotrophin-releasing factor (CRF) is a potent inhibitor of feeding. Emaciated patients with anorexia nervosa have increased concentrations of CSF CRF, which return to normal range with weight gain. The increased CRF may be one of the physiological mechanisms that maintains the anorexia and abnormal eating behavior in patients with anorexia nervosa. Neuropeptide Y, which stimulates feeding behavior, was found to be elevated in the CSF of underweight and recently weight-restored anorexia nervosa patients. However, in long-term weight-restored anorexia nervosa patients these concentrations return to normal. A reduction of food intake may produce a homeostatic increase in neuropeptide secretion that should serve to stimulate feeding, but this mechanism seems to be ineffective in the anorectic patient. Decreased secretion of gonadotrophin-releasing hormone (GnRH) is the cause of amenorrhea in anorexia nervosa. The secretion of GnRH is highly influenced by the neurotransmitters norepinephrine and serotonin, which help to regulate eating behavior and to influence mood. The relation of these neurotransmitters to GnRH, neuropeptide Y, and CRF secretion in anorexia nervosa is an area of ongoing research and may help to reveal biological risk factors. Starvation-induced changes such as reduced norepinephrine turnover, hyperactivity of the hypothalamic-pituitary-adrenal axis, diminished thyroid-stimulating hormone (TSH) response to thyrotrophin-releasing hormone (TRH), low plasma triiodothyronine (T_3) concentrations, and increased basal growth hormone concentrations revert to within the normal range with nutritional rehabilitation.

Because of the role of serotonin in producing satiety or fullness and in regulating impulsive behaviors, a dysfunction of the serotonergic neurotransmitter system was suspected in bulimia. Normal-weight bulimia patients have decreased prolactin response to the pharmacological challenge test of fenfluramine and methchlorophenylpiperazine (mCPP) compared with controls. High binge frequency in bulimia patients is significantly related to lower CSF5 HIAA than is present in controls. This is some evidence for a cerebral-CNS serotonergic deficiency in bulimia nervosa patients. Table 20-2 presents neuroendocrine changes caused by anorexia nervosa.

Table 20-2 Neuroendocrine Changes in Anorexia Nervosa and Experimental Starvation

DIAGNOSIS

Anorexia Nervosa The three criteria for anorexia nervosa proposed by G.F.M. Russell over 20 years ago are still embedded in the latest DSM-IV criteria (Table 20-3). Russell's criteria are: (1) behavior that is designed to produce marked weight loss, (2) a morbid fear of becoming fat, and (3) evidence of an endocrine disorder that in the postpubertal girl causes the cessation of menstruation. Because there is no specific amount of weight loss associated with other symptoms of anorexia nervosa, the current DSM-IV first criterion simply states, “refusal to maintain body at/or above a minimally normal weight for age and height.” The second criterion, “intense fear of gaining weight,” has remained unchanged through the decades. Patients with anorexia nervosa fear gaining weight even when they are emaciated.

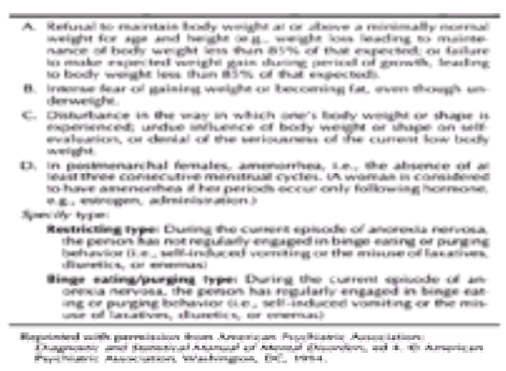

A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected).
B. Intense fear of gaining weight or becoming fat, even though underweight.
C. Disturbance in the way in which one's body weight or shape is experienced; undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight.
D. In postmenarchal females, amenorrhea, i.e., the absence of at least three consecutive menstrual cycles. (A woman is considered to have amenorrhea if her periods occur only following hormone, e.g., estrogen, administration.)

Specify type:

Restricting type: During the current episode of anorexia nervosa, the person has not regularly engaged in binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Binge eating/purging type: During the current episode of anorexia nervosa, the person has regularly engaged in binge eating or purging behavior (i.e., self-induced vomiting or the misuse of laxatives, diuretics, or enemas).

Table 20-3 DSM-IV Diagnostic Criteria for Anorexia Nervosa

The third criterion in DSM-IV for anorexia nervosa pertains to body image disturbance and has evolved since the late 1970s into a more complex concept. In the third edition of DSM (DSM-III) body image disturbance was narrowly defined as body image related to visual self-perception. Because studies have shown that many anorexia patients do not visually overestimate their sizes, and that overestimation is not unique to those with anorexia nervosa, the criterion was then reworded in the revised third edition of DSM-III-R to focus on attitudinal and affective dimensions of body image. This newly worded criterion of “overconcern with body size and shape” has significant overlap between anorexia nervosa patients and the general female population. In order to articulate the core cognitive distortions of the anorexia nervosa patient more precisely, this criterion was again revised in DSM-IV and has three components. “Disturbance in the way in which one's body weight or shape is experienced” reflects the person's claiming to feel fat or believing that one area of the body such as thighs are too fat even when obviously underweight. “Undue influence of body weight or shape on self-evaluation” reflects the dominating importance of the latter to the person. Losing weight and dieting are skills these persons can perform competently and from which they receive strong positive reinforcements. The security of being thin and of dieting compensates for vast insecurities in other areas of life such as interpersonal relationships and adult role expectations. “Denial of the seriousness of the current low body weight” reflects the person's unwillingness to acknowledge that her emaciated body is not healthy or attractive. DSM-IV divides anorexia nervosa into two subtypes: the restricting type and the binge-eating/purging type. There are significant differences between anorexia nervosa patients who engage in bulimic or purging behaviors compared with those who consistently restrict their dietary intake. Impulsive behaviors such as stealing, drug abuse, suicide attempts, and self-mutilations are significantly more prevalent in the binge-eating/purging type compared with the restricting type. The bulimic group of anorexia nervosa patients also has a higher prevalence of premorbid obesity, familial obesity, mood ability, and debilitating personality characteristics; the latter are often in the context of a Cluster B or the impulsive personality disorder cluster. The binge-eating/purging type of patients are actually similar to persons with bulimia nervosa in all of these behaviors and characteristics. They are not classified as having bulimia nervosa because they have lost large amounts of weight and meet the criteria for anorexia nervosa. The specific mechanisms by which the binge-eating/purging type patients are able to obtain and maintain large weight losses are still unknown and are being investigated.

The fourth criterion for the diagnosis of anorexia nervosa in the DSM-IV is amenorrhea, which can occur before noticeable weight loss has occurred and may be caused by a primary disturbance of hypothalamic function. A common hypothesis is that psychological stress can bring about the full expression of this vulnerable risk factor, which then expresses itself in the form of amenorrhea. Malnutrition can perpetuate the amenorrhea but is not primarily responsible for the endocrine disorder.

Specific mechanisms of hypothalamic dysfunction in anorexia nervosa are still under investigation.

Bulimia Nervosa The DSM-III-R criteria for bulimia nervosa are more arbitrary and more ambiguous compared with the criteria for anorexia nervosa. When G.F.M. Russell defined bulimia nervosa in 1979 he proposed three criteria: a powerful and intractable urge to overeat, resulting in episodes of overeating; avoidance of the "fattening" effects of food by inducing vomiting or abusing purgatives or both; and a morbid fear of becoming fat. These criteria do not distinguish bulimia nervosa patients from anorexia nervosa patients of the binge-purge subtype. Because there are physiological differences between bulimic people who lose large amounts of weight and have anorexia nervosa and those who never lose weight, it became necessary to define bulimia nervosa more precisely. Although binge eating is a necessary diagnostic criterion for bulimia nervosa, there is no consensus on what constitutes a binge and what is acceptable as the minimum binge frequency for this disorder. In the first criterion for bulimia nervosa (Table 20-4) binge eating is defined as eating more food than most people would in similar circumstances in a similar period of time. The sense of losing control is a significant subjective aspect that occurs in binge eating and is present as a second part of this criterion. The emphasis on a discrete time period is to differentiate binge eating from continual snacking throughout the day.

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   (1) eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat in a similar period of time under similar circumstances
   (2) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating)
B. Recurrent inappropriate compensatory behavior in order to prevent weight gain, such as self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise.
C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for 3 months.
D. Self-evaluation is unduly influenced by body shape and weight.
E. The disturbance does not occur exclusively during episodes of anorexia nervosa.

Specify type:

Purging type: during the current episode of bulimia nervosa, the person has regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas

Nonpurging type: during the current episode of bulimia nervosa, the person has used other inappropriate compensatory behaviors, such as fasting or excessive exercise, but has not regularly engaged in self-induced vomiting or the misuse of laxatives, diuretics, or enemas

Table 20-4 DSM-IV Diagnostic Criteria for Bulimia Nervosa

The second criterion in the diagnosis of bulimia nervosa is the recurrent use of inappropriate compensatory behaviors to avoid weight gain. These behaviors can be purging, such as self-induced vomiting or laxative abuse, or very restrictive dieting. There is evidence that bulimic persons who purge differ from binge eaters who do not purge in that the latter tend to have less body-image disturbance and less anxiety concerning eating. Bulimia nervosa patients who do not purge tend to be obese. There are also distinct physiological differences between bulimia patients who purge and those who do not. Because of all these differences the diagnosis of bulimia nervosa is subtyped into a purging type, for those who regularly engage in self-induced vomiting or the use of laxatives or diuretics, and a nonpurging type for those who use strict dieting, fasting, or vigorous exercise but do not regularly engage in purging.

The third chronicity and frequency criterion is arbitrary because there was no specific research to support that a minimum average of two episodes of binge eating and two inappropriate compensatory behaviors a week for at least 3 months was associated with impaired behavior and other symptoms of bulimia nervosa. However, when earlier definitions of bulimia nervosa excluded the chronicity and frequency criterion, the prevalence of this disorder seemed unusually high and not always associated with an impairment of function. Ongoing research should eventually provide data to define this criterion more specifically.

The fourth criterion states that self-evaluation is unduly influenced by body shape and weight and more accurately describes the preoccupation of bulimia nervosa patients with their physical appearance than Russell's previous "morbid fear of becoming fat." Further data are needed to define this criterion more specifically.

The fifth and final criterion for bulimia nervosa is that the disturbance does not occur exclusively during episodes of anorexia nervosa. This clearly separates bulimia nervosa from the binge-eating/purging type of anorexia nervosa.

Eating Disorder Not Otherwise Specified This category is for patients who display most of the core clinical features of anorexia nervosa and bulimia nervosa but do not meet criteria for the full syndrome (Table 20-5).

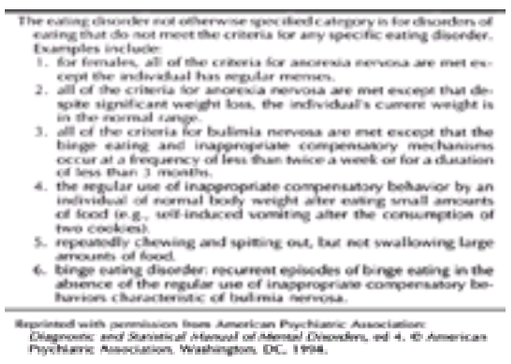

The eating disorder not otherwise specified category is for disorders of eating that do not meet the criteria for any specific eating disorder. Examples include:
1. for females, all of the criteria for anorexia nervosa are met except the individual has regular menses.
2. all of the criteria for anorexia nervosa are met except that despite significant weight loss, the individual's current weight is in the normal range.
3. all of the criteria for bulimia nervosa are met except that the binge eating and inappropriate compensatory mechanisms occur at a frequency of less than twice a week or for a duration of less than 3 months.
4. the regular use of inappropriate compensatory behavior by an individual of normal body weight after eating small amounts of food (e.g., self-induced vomiting after the consumption of two cookies).
5. repeatedly chewing and spitting out, but not swallowing large amounts of food.
6. binge eating disorder: recurrent episodes of binge eating in the absence of the regular use of inappropriate compensatory behaviors characteristic of bulimia nervosa.

Table 20-5 DSM-IV Diagnostic Criteria for Eating Disorder Not Otherwise Specified

Based on clinical work and a survey of the literature researchers came to the conclusion that a significant number of people had serious problems with binge eating but did not meet criteria for bulimia nervosa. These people lacked the compensatory weight-control behaviors and the overconcern with body weight and shape. This proposed disorder was labeled "binge eating disorder" (Table 20-6) and is classified as an eating disorder not otherwise specified in DSM-IV. It is distinguished from bulimia nervosa, nonpurging type by the lack of any compensatory behavior to avoid weight gain and the necessity of binge eating twice a week for a 6-month period (Table 20-4); additional field trials are needed to provide sufficient evidence that this is indeed a specific diagnostic entity.

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by both of the following:
   (1) eating, in a discrete period of time (e.g., within any 2-hour period), an amount of food that is definitely larger than most people would eat in a similar period of time under similar circumstances
   (2) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating)
B. The binge eating episodes are associated with three (or more) of the following:
   (1) eating much more rapidly than normal
   (2) eating until feeling uncomfortably full
   (3) eating large amounts of food when not feeling physically hungry
   (4) eating alone because of being embarrassed by how much one is eating
   (5) feeling disgusted with oneself, depressed, or very guilty after overeating
C. Marked distress regarding binge eating is present.
D. The binge eating occurs, on average, at least 2 days a week for 6 months.
   Note: The method of determining frequency differs from that used for bulimia nervosa; future research should address whether the preferred method of setting a frequency threshold is counting the number of days on which binges occur or counting the number of episodes of binge eating.
E. The binge eating is not associated with the regular use of inappropriate compensatory behaviors (e.g., purging, fasting, excessive exercise) and does not occur exclusively during the course of anorexia nervosa or bulimia nervosa.

Table 20-6 DSM-IV Research Criteria for Binge Eating Disorder

CLINICAL FEATURES

Anorexia Nervosa Anorexia nervosa patients (Fig. 20-1) continually engage in behaviors to maintain or increase their marked weight loss. They try to get rid of food surreptitiously to avoid eating it. They drastically reduce their total food intake and disproportionately decrease the intake of high carbohydrate and fatty foods. Some anorexia patients will follow rigorous exercising programs and others will simply be as active as possible at all times. The binge-eating/purging type of patients will self-induce vomiting and use laxatives and diuretics to lose weight. Many anorexia nervosa patients are preoccupied with food, often collecting recipes and preparing elaborate meals for their families. They may hide carbohydrate-rich foods and horde large quantities of candies, carrying them in their pockets and purses.

FIGURE 20-1 Patient with anorexia nervosa. Note the normal amount of pubic hair. (From Novak ER, Jones GS, Jones HW: *Novak's Textbook of Gynecology*, ed 8. Williams & Wilkins, Baltimore, 1970.)

Anorexia nervosa patients typically express an intense fear of gaining weight; they stare in the mirror frequently to reassure themselves that they are not gaining weight.

The significance of body weight and shape are greatly distorted in anorexia nervosa patients. Some feel globally overweight and others realize they are thin but think certain parts of their body, particularly the abdomen, buttock, and thighs, are too fat. They fail to recognize that they are dangerously emaciated. They often feel very ineffective and for them losing weight is an impressive achievement that boosts their self-esteem and sense of control. The denial of the seriousness of the medical implications of their malnourished state allows them to believe that thinness can solely define their self-worth. Obsessive-compulsive behavior often develops after the onset of anorexia nervosa. An obsession with cleanliness is often apparent as an increase in housecleaning activities occurs. In students compulsive studying may develop, although difficulty concentrating often forces students to spend many more hours at their studies.

Amenorrhea is due to diminished secretion of GnRH, which in turn diminishes the pituitary secretion of follicle-stimulating hormone (FSH) and luteinizing hormone (LH); this in turn is responsible for abnormally low levels of estrogen.

The obsessions concerning food and weight are egosyntonic in these patients. Patients have characteristic cognitive distortions such as dichotomous thinking (all-or-none reasoning). Persons with anorexia nervosa have little or no motivation for treatment because the illness serves a strong positive function in their life by allowing them to escape uncomfortable interpersonal problems or developmental issues. Poor sexual adjustment is frequent in anorectic patients. Many anorectic adolescents have delayed psychosocial sexual development, and adults often have a markedly decreased interest in sex with the onset of anorexia nervosa.

Medical Complications Most of the physiological and metabolic changes in anorexia nervosa are secondary to the starvation state or purging behaviors (Table 20-7 and Table 20-8). These changes revert to normal with nutritional rehabilitation or cessation of the purging behavior. Abnormalities in hematopoiesis include leukopenia and a relative lymphocytosis. Patients who engage in self-induced vomiting (Fig. 20-2) or abuse laxatives and diuretics may develop hypokalemic alkalosis. This is diagnosed by elevated serum bicarbonate, hypochloremia, and hypokalemia. Patients with electrolyte disturbances may have physical symptoms of weakness, lethargy, and cardiac arrhythmia. The latter condition may result in a sudden cardiac arrest, which is a cause of death in patients who purge. Elevation of serum enzymes reflects fatty degeneration of the liver and is present both in the emaciated anorectic phase and during refeeding. Elevated serum cholesterol levels tend to occur more frequently in younger patients. Carotenemia is often observed in malnourished anorectic patients.

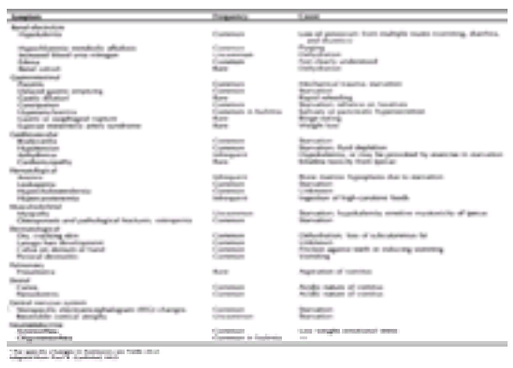

Table 20-7 Biological Complications of Anorexia Nervosa and Bulimia Nervosa*

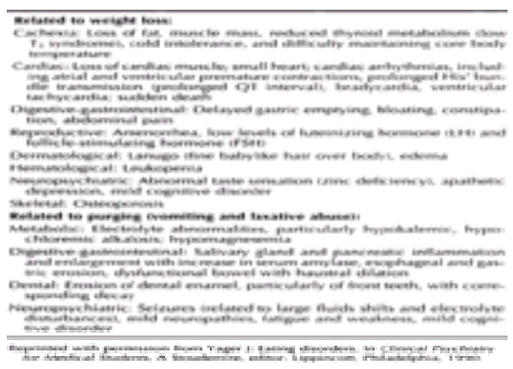

Table 20-8 Medical Complications of Eating Disorders

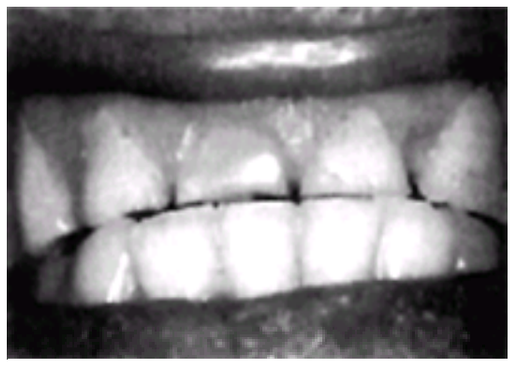

FIGURE 20-2 Dental erosion of upper front teeth of a patient with longstanding bulimia nervosa, binge eating/purging type. (Reprinted with permission from Walsh BT: Eating disorders. In *Psychiatry*, vol 2, A Tasman, J Kay, JA Lieberman, editors. Saunders, Philadelphia, 1997.)

Sue, a 19-year-old single white female, has just completed 6 months in an eating disorder specialty unit for the treatment of anorexia nervosa and obsessive-compulsive disorder. She arrived weighing 30 kg (66 lbs) and was 162 cm (5'1") tall.

Sue recalls that as early as age 4 she had to do everything “perfectly” and had rituals for putting things away and cleaning things thoroughly. These behaviors continued throughout grade school. In school she was extremely competitive and always obtained straight As. She began gymnastics when she was 8 years old and continued with this until the onset of her anorexia nervosa at age 12, when she began eating foods slowly and gradually eating less food. She denies any reason for doing this or any stress at the time. The following summer she was hospitalized because of weight loss at a time when she should have been growing and gaining weight. It was then she was diagnosed as having anorexia nervosa. Her obsessions and compulsions became more severe as her anorexia worsened. She developed rituals around cleaning, straightening and arranging objects, counting, and exercising. She felt and continues to feel that she must be constantly industrious, moving and accomplishing things because she fears becoming “lazy.” She was often unable to leave her home in the morning for hours because she had to touch objects and perform a number of rituals, which caused her to frequently be late for school. The following year her weight dropped to 50 lbs and she was hospitalized again. This was the beginning of a series of repeated hospitalizations and weight losses up to the time she graduated from high school as valedictorian despite her almost continued absence from school. After graduation she worked as a cashier in a grocery store for 1 year while living with her parents and was able to stay out of the hospital for 18 months. She then had a series of several hospitalizations and short-term jobs working as a copy center associate and as an aide in a nursing home.

Sue denies any binge-eating; purging; illegal drug use, or use of alcohol, cigarettes, or caffeine. She only restricts her intake and vigorously exercises. She is very distressed about her obsessions and compulsions but not about her eating behavior. She has never shown the signs and symptoms of depression. She has primary amenorrhea.

Sue stated she always had girlfriends throughout high school.

She has never dated or had a boyfriend and has never been sexually active. She is extremely angry with her father and has an openly antagonistic relationship with him. She states she has always wanted attention and approval from him but has never received it. She is the youngest of three children; her two older sisters have done well academically and professionally. She has a close and enmeshed relationship with her mother.

On admission to the specialized eating disorder program, Sue was placed on a liquid diet. She had enormous difficulty adhering to unit rules and procedures. She was perpetually in motion. She was very upset and angry when confronted by staff and was extremely sensitive to criticism. She was exceedingly harsh, critical, and self-punishing. After 6 months of hospitalization, she was able to successfully maintain her weight within a normal range. Her confidence was greatly improved in her ability to eat three meals and maintain her weight. She stated she actually enjoyed eating because she was allowed to be a vegetarian. She was much more sociable, interacting with peers and vocal in community meetings and groups. Her rituals had gradually decreased and she was actually able to get through several days with no rituals. She found it more difficult to abstain from rituals in the evening and she still felt very concerned about “being lazy.” She was also more comfortable and open in talking with her father.

During her fifth month of hospitalization when she had reached her target weight, Sue was placed on 225 mg a day of clomipramine (Anafranil), which was helpful in reducing her obsessions and compulsions. Throughout her hospitalization Sue was treated with a cognitive-behavioral therapy program and family counseling. She was discharged to a specialized eating disorder outpatient program with cognitive psychotherapy, family counseling, and group therapy.

Discussion Sue had perfectionistic traits, a feature of obsessive-compulsive personality disorder, as a young child. There is growing evidence that this type of personality trait puts a young woman at risk for developing anorexia nervosa—restricting type, which is Sue's diagnosis. After several years in addition to having an obsession-compulsive personality disorder, Sue actually met diagnostic criteria for Axis I obsessive-compulsive disorder. These behaviors were not severe until she developed anorexia nervosa. There is a distinct difference between obsessions and compulsions such as counting and checking compared with eating disorder rituals and preoccupations. Sue was distressed by the former, which were egodystonic, but was not bothered by the latter eating disorder rituals, which were egosyntonic. Because she failed to respond to fluoxetine (Prozac) in her previous hospitalizations she was placed on clomipramine and responded well to this drug, as evidenced by a decrease in obsessions and compulsions.

Bulimia Nervosa Bulimia nervosa usually begins from a period of a few weeks to a year or longer after dieting. The dieting may or may not be successful in achieving weight loss but the weight loss is never enough to qualify for a diagnosis of anorexia nervosa. Restrained eating gives way in these persons to binge-eating episodes, which are usually terminated by abdominal pain or discomfort, self-induced vomiting, or social interruption. The bulimic episode is often followed by feelings of guilt, depression, or self-disparagement. Some bulimia nervosa patients will use laxatives for weight control and have a pattern of alternating binges and fasting. A minority of bulimic patients use diuretics for weight control. Patients with bulimia nervosa have a fear of not being able to stop eating voluntarily and thus develop a pattern of severe dieting, interrupted by binge-eating episodes. The food consumed during a binge usually has a high dense calorie content and a texture that facilitates rapid eating. Frequent weight fluctuations can occur in bulimia nervosa patients.

The average length of a bingeing episode is about 1 hour. Some patients with the disorder have abrasions and scars on the backs of their hands from sticking their fingers down their throat to induce vomiting. Bulimia nervosa patients will abuse amphetamines to reduce their appetite and to lose weight; they will also use ipecac to induce vomiting. Most bulimia nervosa patients do not eat regular meals and have difficulty feeling satiety at the end of a normal meal. They usually prefer to eat alone and in their homes. Although many have been at the upper end of a normal weight range or slightly overweight they will often choose a weight at the lower end of a normal weight range as their ideal body weight. About 10 percent of bulimia nervosa patients have been markedly overweight. They can be very concerned about their body shape and image but are not driven to lose weight like anorexia nervosa patients are. For many bulimia nervosa patients the actual binge-eating episode often gives them a relief of tension and anxiety and thus can become a rewarding behavior.

Patients with bulimia nervosa have a wide range of psychological and behavior problems. Their lives are often chaotic with troublesome interpersonal relationships, impulsive behaviors, and high levels of anxiety and compulsivity. They have poor self-concepts and a high prevalence of mood disorders. Substance abuse is not unusual in patients with this disorder, with alcohol abuse being the most common. About one fourth of bulimic patients have problems of impulsively stealing, with food, clothing, and jewelry being most commonly stolen.

Medical Complications Bulimia nervosa patients who engage in self-induced vomiting and abuse purgative or diuretic medications are susceptible to the same complications as anorexia nervosa patients involved in this behavior. Exposure to gastric juices through vomiting can cause severe erosion of the teeth, pathological pulp exposures, diminished masticatory ability, and an unaesthetic appearance. Parotid gland enlargement is associated with elevated serum amylase concentrations and is commonly observed in patients who binge and vomit. Acute dilatation of the stomach is a rare emergency condition for patients who binge eat, and esophageal tears can occur in the process of self-induced vomiting. Severe abdominal pain in the bulimia nervosa patient should alert the physician to a diagnosis of gastric dilatation and the need for nasal gastric suction, X-rays, and surgical consultation. Cardiac failure caused by cardiomyopathy from ipecac toxication is a medical emergency that usually results in death. Symptoms of pericardial pain, dyspnea, and generalized muscle weakness associated with hypotension, tachycardia, and

electrocardiogram (EEG) abnormalities should alert medical personnel to possible ipecac intoxication.

Alice was 21 years old and had been binge-eating and vomiting for 8 years before she was referred to an eating disorder treatment program. She had achieved her greatest weight of 59 kg when she was a sophomore in high school and 15 years old. She had several stresses during that year—she did not make the cheerleading team; her grandfather, with whom she was very close, died; and some close friends moved away. She developed a general feeling of unhappiness, loneliness, and loss. During that year, her pets had been given away by her parents because they were too much trouble. It was during this time that Alice decided she wanted to lose weight and soon developed the habit of self-induced vomiting. She gradually increased the amount of vomiting until her senior year of high school, when she was vomiting four times a week. When Alice began to lose weight, she resorted to a series of fad diets. She refused to eat meals with her family and after about 6 months of dieting she began to get up at night for binge-eating episodes followed by self-induced vomiting, which she had perfected for the past several months.

After graduating from high school, Alice went to college for 1 year and majored in animal science. She did not like this curriculum and dropped out after a year. However, during that time her symptomatology increased and she actually wallpapered one of her walls with M&M wrappers. She then went back to live at home and had a series of jobs in food stores, such as a donut shop and a deli, as well as a clothing store. Finally, she was unable to hold a job for more than a few months because of her binge-eating and vomiting behavior. Her most stable relationship was with her boyfriend, with whom she was living a year before hospitalization. During that time, at her boyfriend's insistence she saw a social worker weekly for several months but then stopped treatment abruptly. Her binge-eating and vomiting got out of control, so her boyfriend took her to a local hospital. She was there for 3 months and then transferred directly to another program in a different hospital. During her initial hospitalization she was allowed to come and go as she pleased and she continued her binge-eating and vomiting behavior. At the time of admission to the special eating disorder treatment program, Alice admitted to feeling depressed and having frequent crying spells. She also had difficulty falling asleep at night. She was totally preoccupied with thoughts of binge-eating and of food. The significant findings on her physical examination were a slight scoliosis and marked deterioration of her teeth.

Alice began her menses at age 14. At about the age of 16 she began to have irregular menses; however, she continued to menstruate. Her lowest weight was 50.9 kg and that was just a few months before her admission to an eating disorder specialty unit.

She was placed on a multifaceted treatment program with twice-weekly individual psychotherapy sessions, daily group therapies, and biweekly family sessions. Initially she expressed considerable denial and anger and was later able to describe feelings of emptiness, loneliness, and a sense of inadequacy. Response prevention techniques were used to stop the vomiting and binge-eating behavior and she responded very well in a structured setting. She also received 60 mg of fluoxetine daily. Alice gradually gained control over the eating behavior and developed considerable insight into her problems and insecurities. During her 4 months in a special eating disorder treatment program, she responded well to the gradual introduction of responsibility for her eating behavior and other aspects of her life. After discharge from the hospital, she continued her treatment in an outpatient setting and received individual and family therapy. She returned to live with her parents and brother, where she stayed for 6 months until she moved into a supportive apartment. Several months later she began attending a community college. Her binge-eating and vomiting had stopped entirely and she was eating normally and maintaining her weight within a normal range. Her prognosis, however, is guarded because of her long history of dysfunctional eating behavior.

Discussion Alice's dieting began when she had a series of stresses that challenged her feelings of competency and effectiveness. She was mildly overweight when she began dieting and was not very effective at it until she started self-induced vomiting. Even with this behavior, as is typical in bulimia nervosa patients, she was not able to become emaciated. She was able to maintain a sexual relationship with her boyfriend, which is in contrast to the more typical asexual nature of the anorexia nervosa patient. The rest of her life, however, was chaotic in that she could not keep a job for longer than a few months and was unable to have stable relationships with most people. She suffered from depression, which responded to fluoxetine. Over the course of a year of cognitive psychotherapy, she ceased bingeing and vomiting. Although most bulimia nervosa patients have an immediate reduction in their bingeing and vomiting as a response to cognitive therapy, complete cessation of this behavior often occurs 1 or 2 years after the intensive therapy has occurred.

PATHOLOGY AND LABORATORY EXAMINATION

Anorexia Nervosa There are no laboratory tests that can provide a diagnosis of anorexia nervosa. The medical phenomena present in this disorder result from the starvation or purging behaviors. There are several relevant laboratory tests that should be obtained in these patients. A complete blood count will often reveal a leukopenia with a relative lymphocytosis in emaciated anorexia nervosa patients. If binge-eating and purging are present, serum electrolytes will reveal a hypokalemic alkalosis. Fasting serum glucose concentrations are often low during the emaciated phase, and serum salivary amylase concentrations are often elevated if the patient is vomiting. An electrocardiogram may show S-T segment and T-wave changes, which are usually secondary to electrolyte disturbances; emaciated patients will have hypotension and bradycardia. Young girls may have an elevated serum cholesterol level. All these values revert to normal with nutritional rehabilitation and cessation of purging behaviors. Endocrine changes that occur, such as amenorrhea, mild hypothyroidism, and hypersecretion of corticotrophin-releasing hormone are due to the underweight condition and revert to normal with weight gain.

Bulimia Nervosa The predominant laboratory abnormalities in bulimia nervosa patients occur in those who purge and are seen as changes in serum electrolyte values. The purging behavior can produce hypokalemia and a metabolic alkalosis. Patients who vomit frequently may have an elevated salivary serum amylase. On rare occasions the bulimia nervosa patient will have to be hospitalized because of severe electrolyte disturbances or depression with suicidal ideation.

DIFFERENTIAL DIAGNOSIS

Anorexia Nervosa Because anorexia nervosa patients are usually in the physician's office at the insistence of family, friends, or employers and have a strong denial of their illness, obtaining an accurate history is often difficult. It is important to ascertain that the patient has no medical illness that can account for the weight loss. In rare circumstances a patient can have both anorexia nervosa and a medical illness that contributes to the weight loss. In such a situation the diagnosis of anorexia nervosa is made by the positive criteria for anorexia nervosa and both the underlying medical condition and the anorexia nervosa have to be treated.

Weight loss, peculiar eating behavior, and vomiting can occur in several psychiatric disorders. Weight loss frequently occurs in depressive disorders, which have several features in common with anorexia nervosa; for example, depressed moods, crying spells, sleep disturbances, obsessive ruminations, and occasional suicidal thoughts. There are, however, several distinguishing features between the disorders. Generally, a patient with a depressive disorder has a decreased appetite whereas an anorexia nervosa patient denies the existence of an appetite. It is only in the severe stages of anorexia nervosa that the patient actually has a decreased appetite. The hyperactivity seen in anorexia nervosa patients differs from the agitative activity seen in patients with depressive disorders in that much of the activity of the patient with anorexia nervosa is planned and ritualistic—exercising, jogging, and cycling programs. The preoccupation with the calorie content of food, collecting recipes, and preparing gourmet feasts, which is typical of the anorectic patient, is not present in the patient with a depressive disorder. Patients with depression do not fear becoming fat nor do they have a disturbance of their body image, as anorexia nervosa patients do.

Weight fluctuations, vomiting, and peculiar food handling may occur in somatization disorder. The weight loss in somatization disorder is usually not as severe as in anorexia nervosa, nor does the patient in the former disorder express a morbid fear of becoming overweight. Also, amenorrhea that lasts for 3 months or longer is unusual in a patient with somatization disorder.

Delusions about food in schizophrenia are seldom concerned with the calorie content of food. A patient with schizophrenia is not preoccupied with a fear of becoming obese and does not have the hyperactivity seen in an anorectic patient. Chronic medical illnesses frequently associated with weight loss are Crohn's disease, hyperthyroidism, Addison's disease, and diabetes mellitus.

Bulimia Nervosa If bingeing and purging behaviors are occurring in a person who meets diagnostic criteria for anorexia nervosa, bulimia nervosa cannot be the diagnosis. Severe weight loss and amenorrhea are the two features distinguishing these eating disorders.

On rare occasions a central nervous system tumor may be associated with bulimic behaviors. Overeating episodes also occur in the Kluver-Bucy syndrome, which consists of visual agnosia, compulsive licking and biting, inability to ignore any stimulus, and hypersexuality. Another uncommon syndrome associated with hyperphagia is the Kleine-Levin syndrome, which is characterized by periodic hypersomnia lasting for several weeks.

COURSE AND PROGNOSIS

The only long-term studies of eating disorders are based on patient samples. There are no data available on people who have eating disorders but do not get treatment. Large–sample-size long-term studies have occurred predominantly in anorexia nervosa patients because bulimia nervosa was defined as a separate entity and has been studied as such only since in the 1980s.

Anorexia Nervosa Anorexia nervosa has a bimodal peak of onset either between the ages of 14 and 15 or at age 18. Although infrequent, cases are reported in girls as young as 9 years and some first episodes have occurred in postmenopausal women. The course of anorexia nervosa can vary from one episode with full recovery, commonly occurring partial recovery with relapses, to a chronic severe course that results in death.

Ten-year outcome studies in the United States have shown that about one fourth of the patients recover completely and another one half are markedly improved and functioning fairly well. The other one fourth includes an overall 7 percent mortality rate and those who are functioning poorly with a chronic underweight condition. Swedish and English studies over a 20- and 30-year period have a mortality rate of 18 percent. Good prognostic indicators across studies include an earlier age of onset (under age 18), no previous hospitalization for the illness, and no purging behaviors. Some factors such as parental conflict, degree of denial, immaturity, and self-esteem have been related to outcome in some studies but not others. About half of anorexia nervosa patients will eventually have the symptoms of bulimia, usually within the first year after the onset of anorexia nervosa.

Bulimia Nervosa Course of illness and prognostic studies of bulimia nervosa are now appearing in the literature. Of these, the majority have a relatively short-term follow-up of between 6 months and 2 years posttreatment. A recent review of 88 articles on the naturalistic and treatment follow-up of bulimia nervosa with periods ranging from 6 months to 10 years reached the following conclusions: the typical age of onset for bulimia nervosa is between 18 and 19 years and mortality rates can be estimated between 0 and 3 percent. After between 5 and 10 years about 50 percent of bulimic patients will be fully recovered, 20 percent will continue to meet diagnostic criteria for bulimia nervosa. About one third of recovered bulimic patients will relapse within 4 years of presentation. Patients with personality disorders marked by problems with impulse control generally had a worse prognosis, compared with bulimia nervosa patients with no personality disorder problems.

TREATMENT

Anorexia Nervosa The severity of illness will determine the intensity of treatment required for the anorexia nervosa patient. Treatment levels can range from an inpatient specialized eating disorder unit to a partial ospitalization or day program to outpatient care depending on the weight, medical status, and other psychiatric comorbidity of the patient. Decisions about the particular treatment modalities and strategies must be based on the needs of the patient as well as the capabilities of the treatment setting. Understandably it has been extremely difficult to subject severely medically ill patients with anorexia nervosa to controlled treatment studies in the emaciated state of the illness. Since patients with anorexia nervosa are resistant to and disinterested in treatment there are very few outpatient controlled treatment studies. Open studies have indicated that a multifaceted treatment approach is the most effective, which includes medical management, psychoeducation, and individual therapy utilizing cognitive and behavior therapy principles. Controlled studies have shown that children under the age of 18 do better if they have family therapy. Nutritional counseling and pharmacological intervention can also be useful components to the treatment plan. Guidelines for admission to a hospitalized or day treatment program for anorexia nervosa are presented in Table 20-9. To obtain the patient's cooperation for treatment it is helpful to emphasize the benefits of treatment and reassure the patient that treatment can bring about relief of distressing symptoms related to anorexia such as insomnia, depression, and obsessive thoughts about food and body weight that interfere with the ability to concentrate on other matters, and can result in increased physical well-being, energy, and improved peer relationships.

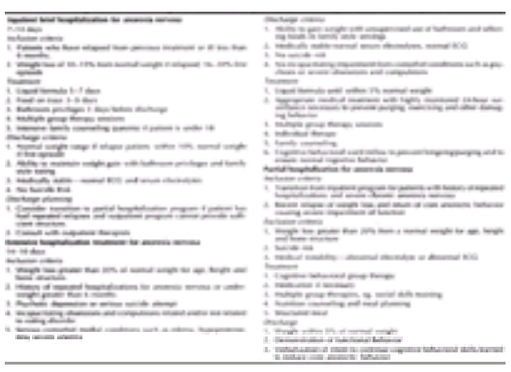

Table 20-9 Guidelines for Hospitalizations and Day Treatment for Anorexia Nervosa

The severely ill anorexia nervosa patient can present an extremely difficult medical-behavioral management challenge. These patients should be hospitalized in a specialized inpatient eating disorder treatment setting that has a team of individuals highly skilled in the multidisciplinary treatment of anorexia nervosa. Outpatient therapy has the best chance for success in anorectic patients who: (1) have had the illness for less than 6 months, (2) are not bingeing and vomiting, and (3) have parents who are likely to cooperate and participate effectively in family therapy.

Medical Management Medical management requires weight restoration, nutritional rehabilitation, rehydration, and correction of serum electrolytes. Inpatient hospitalization should include daily monitoring of weight, food, calorie intake, and urine output. Patients should be monitored closely for attempts to vomit. Outpatients should be weighed weekly in the physician's office with periodic physical examinations and measurement of serum electrolytes if the patient is purging.

Cognitive-Behavioral Therapy Cognitive and behavior therapy principles can be applied in both inpatient and outpatient settings. Behavior therapy has been found to be effective for inducing weight gain. There are no large–sample-size controlled studies of cognitive therapy with behavior therapy in anorexia nervosa patients. Monitoring is an essential component of cognitive-behavioral therapy. Patients are taught to monitor their food intake, their feelings and emotions, their bingeing and purging behaviors, and their problems in interpersonal relationships. Cognitive restructuring is a method that patients are taught in order to identify autonomic thoughts and to challenge their core beliefs. Problem solving is a specific method whereby patients learn how to think through and devise strategies to cope with their food-related and interpersonal problems. Patients' vulnerability to rely on anorexic behaviors as a means of coping can be reduced if they can learn to use these techniques effectively.

Family Therapy A family analysis should be done on all anorexia nervosa patients who are living with their families. On the basis of this analysis a clinical judgment can be made as to what type of family therapy or counseling is advisable. In some cases family therapy is not possible. However, in those cases issues of family relationships can be addressed in individual therapy. In some cases brief counseling sessions with immediate family members may be the extent of family therapy required. In one controlled family therapy study in London, anorectic patients under the age of 18 benefited from family therapy whereas patients over the age of 18 did worse in family therapy compared with the control therapy. There are no controlled studies of the combination of individual and family therapy. However, in actual practice most clinicians provide individual therapy and some form of family counseling in managing anorexia nervosa patients.

Pharmacotherapy Medications can be useful adjuncts in the treatment of anorexia nervosa. The first drug used in treating anorectic patients was chlorpromazine. This medication is particularly helpful for severely ill patients who are overwhelmed with constant thoughts of losing weight and behavioral rituals for losing weight. Surprisingly there are no double-blind controlled studies to definitively prove this drug's effectiveness for calming these patients and inducing weight gain. It may be necessary to start at a low dosage of the liquid form such as 10 mg 3 times a day and to gradually increase the dosage. Cyproheptadine (Periactin) in high dosages (up to 28 mg a day) can facilitate weight gain in anorectic restrictors and also has an antidepressant effect. Some recent small–sample-size open studies indicate that fluoxetine may be effective in preventing relapse in patients with anorexia nervosa.

Bulimia Nervosa In contrast to the relatively few outpatient treatment studies of anorexia nervosa, treatment outcome studies of bulimia nervosa have proliferated since the late 1980s.

Cognitive-Behavioral Therapy Cognitive-behavioral therapy should be considered the benchmark, first-line treatment for bulimia nervosa. It has been found to be the most effective treatment in over 35 controlled psychosocial studies. About 40 to 50 percent of patients are abstinent from both binge eating and purging at the end of treatment (16 to 20 weeks). Altogether, improvement by a reduction in binge-eating and purging occurred in a range from 70 to 95 percent of patients. Additionally, another 30 percent of those who did not show improvement immediately posttreatment showed improvement to full recovery 1 year after treatment. In patients with a depressive disorder and bulimia nervosa, cognitive-behavioral therapy was also found to decrease depression.

The data supporting the efficacy of cognitive-behavioral therapy is based upon strict adherence to rigorously implemented, highly detailed, manual-guided treatments that include about 18 to 20 sessions over 5 to 6 months. It should be noted that all controlled studies have been conducted by therapists specifically trained in these manual-guided treatments written for bulimia nervosa. Cognitive-behavioral therapy implements a number of cognitive and behavioral procedures to: (1) interrupt this self-maintaining behavioral cycle of bingeing and dieting, and (2) alter the individual's dysfunctional cognitions and beliefs about food, weight, body image, and overall self-concept.

Family Therapy Family therapy is not widely used in the treatment of bulimia nervosa as it is for anorexia nervosa because most patients with bulimia nervosa are in their 20s and live away from their family of origin. There is a consensus that families of younger patients should be involved with their treatment; however, controlled studies are needed to prove this.

Interpersonal Therapy Interpersonal therapy has been applied as a treatment for bulimia nervosa because interpersonal problems frequently accompany this disorder. This therapy, which focuses on interpersonal functioning and does not target eating behaviors, has shown promising results. In one controlled study interpersonal therapy was not as effective as cognitive behavioral therapy at the end of treatment but at 1 and 6 years following treatment it had produced equally significant changes. These findings suggest that interpersonal therapy has a delayed but powerful effect.

Pharmacotherapy Since the late 1980s over a dozen double-blind, placebo-controlled trials of antidepressant medications, such as desipramine (Norpramin), imipramine (Tofranil), amitriptyline (Elavil), nortriptyline (Pamelor), phenelzine (Nardil), and fluoxetine, have been conducted in normal-weight outpatients with bulimia nervosa. The dosage of antidepressant medication used was similar to that used for the treatment of depression. In all trials the drugs were significantly more effective than placebo in reducing binge eating. In addition, the medication improved mood and reduced eating disorder symptoms such as preoccupation with shape and weight. However, the average abstinence rate from binge eating and purging in these studies was 22 percent, indicating that the majority of patients were still symptomatic at the end of treatment. The few long-term studies to evaluate maintenance of change showed disappointing results, with over 80 percent of patients relapsing. There is some indication that if medications are used the treatment should be continued for 6 months to prevent relapse. However, studies are needed to assess the adequate length of time for pharmacotherapy. One large multicenter collaborative study showed fluoxetine in doses of 60 mg a day to be effective in reducing episodes of binge eating and purging. Controlled double-blind placebo trials are needed for other selective serotonin reuptake inhibitors (SSRI).

Three studies have examined the combined effects of cognitive behavioral therapy and antidepressant medication in treating bulimia nervosa. One study found that group cognitive-behavioral therapy was superior to imipramine in decreasing binge eating and purging, but the combined treatment demonstrated no additive effects over those of the cognitive-behavioral therapy group alone. Another study compared individual cognitive-behavioral therapy, desipramine, or both and found that all groups were similar at 16 weeks. At 32 weeks, only the combined treatment given for 24 weeks was superior to medication given for 16 weeks. In the third study cognitive-behavioral therapy plus medication (desipramine, followed by fluoxetine in nonresponders) was superior to medication alone, but supportive psychotherapy plus medication was not. Cognitive-behavioral therapy plus medication was superior to cognitive-behavioral therapy alone. Cognitive-behavioral group therapy is especially cost-effective for patients as well as for treatment centers. Group therapy is also appealing because patients with bulimia nervosa often keep their eating disorder a secret and feel isolated, ashamed, and embarassed about their symptoms. They have difficulty asking for help and frequently have interpersonal problems. Although widely practiced using cognitive-behavioral therapy techniques, group therapy for bulimia has not been studied in a controlled assessment that compares it with individual cognitive-behavioral therapy.

Guided Self-Help Guided self-help is a new approach to the treatment of bulimia nervosa. Two self-help programs based on the cognitive behavioral therapy approach have been published in which the patient follows a self-help program with some support and guidance from a therapist; the therapist's involvement is minimal. Methodologically sound controlled studies are needed to establish evidence for the effectiveness of these self-help groups.

SUGGESTED CROSS-REFERENCES

Some of the specific syndromes that can be associated with eating disorders are found in Chapter 14 on mood disorders, in Chapter 15 on anxiety disorders, in Chapter 16 on somatoform disorders, in Chapter 11 on substance-related disorders, Chapter 22 on impulse-control disorders not elsewhere classified, in Section 25.6 on endocrine and metabolic disorders, and in Section 7.3 on typical signs and symptoms of psychiatric illness. Personality disorders are discussed in Chapter 24, and the relationship between eating disorders and feeding in childhood, rumination, and pica are discussed in Chapter 41. Other areas that relate to this chapter include consultation-liaison psychiatry (Section 25.12), relational problems (Chapter 26), psychodynamic therapy (Section 30.1), and psychiatric treatment of infant, child, and adolescent (Chapter 48). Sections on genetics and psychiatry (Section 1.18), psychopharmacology (Chapter 31), and family therapy (Section 30.5) provide more detailed background to understanding eating disorders.

CHAPTER REFERENCES

Agras WS, Rossiter EM, Arnow B: Pharmacologic and cognitive-behavioral treatment for bulimia nervosa: A controlled comparison. Am J Psychiatry *149:*82, 1992.

Bell RM: *Holy Anorexia.* University of Chicago Press, Chicago, 1985.

Braun DL, Sunday SR, Halmi KA: Psychiatric comorbidity in patients with eating disorders. Psychol Med *24:*859, 1994.

*Bruch H: *Eating Disorders: Obesity, Anorexia Nervosa, and the Person Within.* Basic Books, New York, 1973.

Casper RC: Personality features of women with good outcome from restricting anorexia nervosa. Psychosom Med *52:*156, 1990.

Cooper PJ: *Bulimia Nervosa: A Guide to Recovery.* Robinson, London, 1993.

Crisp AH, Palmer RL, Kaluchy RS: How common is anorexia nervosa? A prevalent study. Br J Psychiatry *128:*549, 1976.

Devlin MJ, Walsh BT, Spitzer RL: Is there another binge eating disorder? A review of the literature on overeating in the absence of bulimia nervosa. Int J Eating Disord *11:*341, 1992.

Eckert ED, Halmi KA, Marchi EP: Ten-year follow-up of anorexia nervosa: Clinical course and outcome. Psychol Med *25:*143, 1995.

Fairburn CG, Welch S, Doll HA: Risk factors for bulimia nervosa. Arch Gen Psychiatry *54:*509, 1997.

Fairburn CG, Jones R, Pelevar RC: Three psychological treatments for bulimia nervosa: A comparative trial. Arch Gen Psychiatry *48:*463, 1992.

Halmi KA: Clinical crossroads—a 24-year-old woman with anorexia nervosa. JAMA *279:*1992, 1998.

*Halmi KA, Eckert E, Ladu TJ: Anorexia nervosa: Treatment efficacy of cyproheptadine and amitriptyline. Arch Gen Psychiatry *48:*177, 1986.

Halmi KA, Eckert E, Marchi EP, Sampugnaro R, Apple R, Cohen J: Comorbidity of psychiatric diagnosis in anorexia nervosa. Arch Gen Psychiatry *48:*712, 1991.

Halmi KA: Basic biological overview of eating disorders. In *Psychopharmacology: The Fourth Generation of Progress*, Bloom FE, Kupfer DJ, editors. Raven, New York, 1995.

Halmi KA: Comorbidity of the eating disorders. In *Bailliere's Clinical Psychiatry Series*, D Jimmerson, W Kaye, editors. Bailliere Tindall, London, 1997.

*Hoek HW: Review of the epidemiological studies of eating disorders. Int Rev Psychiatry *5:*61, 1993.

Kaye WH, George DT, Gwirtsman HE: Altered serotonin activity in anorexia nervosa after long-term weight restoration. Arch Gen Psychiatry *48:*556, 1991.

Keel PK, Mitchell JE: Outcome and bulimia nervosa. Am J Psychiatry *154:*313, 1997.

*Keel PK, Mitchell J, Miller K, Davis T, Crow S: Long term outcome of bulimia nervosa. Arch Gen Psychiatry *55:*63, 1999.

Mitchell JE, Raymond NC: Cognitive-behavioral therapy in treatment of bulimia nervosa. In *The Psychobiology and Treatment of Anorexia Nervosa and Bulimia Nervosa*, K Halmi, editor. American Psychiatric Press, Washington, DC, 1992.

Mitchell JE, Pyle RL, Eckert ED: A comparison study of antidepressants and structured group psychotherapy in the treatment of bulimia nervosa. Arch Gen Psychiatry *47:*149, 1990.

Russell GFM, Beardwood CJ: Amenorrhea and the eating disorders: Anorexia nervosa and obesity. Psychother Psychosom *18:*358, 1970.

Russell GFM: Bulimia nervosa: An ominous variant of anorexia nervosa. Psychol Med *9:*492, 1979.

Russell GFM, Shmuckler GI, Dare C: An evaluation of family therapy in anorexia nervosa and bulimia nervosa. Arch Gen Psychiatry *44:*1047, 1987.

*Silverman JA: Historical development. In *Psychobiology and Treatment of Anorexia Nervosa and Bulimia Nervosa*, K Halmi, editor. American Psychiatric Press, Washington, DC, 1992.

Strober M: Personality and symptomalogical features in young, non-chronic anorexia nervosa patients. J Psychosom Res. *24:*353, 1980.

Theander S: Outcome and prognosis in anorexia nervosa and bulimia: Some results of previous investigations compared with those of a Swedish long-term study. J Psychiatry Res *19:*493, 1985.

*Walsh BT, Hadigan CM, Devlin MJ: Long-term outcome of antidepressant treatment for bulimia nervosa. Am J Psychiatry *148:*1206, 1991.

Walsh BT, Wilson GT, Loeb KL, Devlin MJ: Medication and psychotherapy in the treatment of bulimia nervosa. Am J Psychiatry *154:*523, 1997.

Wilson GT, Fairburn CG: Cognitive treatments for eating disorders. J Consult Clin Psychol *61:*261, 1993.

CHAPTER 21. SLEEP DISORDERS

CONSTANCE A. MOORE, M.D., ROBERT L. WILLIAMS, M.D. AND MAX HIRSHKOWITZ, PH.D.

Sleep is a process the brain requires for proper functioning. The seriousness of sleep disorders is poorly recognized by the general public and the clinical community. Although individuals afflicted with sleep problems desperately seek help, they are often met with indifference on the part of health care providers. Whether this comes from lack of knowledge or a jaded attitude resulting from the sleepless-overworked "rite of passage" during internship, residency, or graduate study is not known; however, the old saying "no one ever died from a sleep problem" is incorrect. Many sleep disorders are life threatening, either directly (e.g., fatal familial insomnia and obstructive sleep apnea) or indirectly as a result of sleep-related accidents. Investigations link many major industrial catastrophes to sleepiness. The indirect costs of sleep-related accidents total approximately 100 billion dollars per year in the United States. Thus, sleepiness is a serious, potentially life-threatening condition that affects not only the individual but also the family, coworkers, and society in general.

The 1959–1960 American Cancer Society survey of more than one million Americans included several questions related to sleep habits. Results showed increased mortality among individuals who habitually slept significantly more or less than normal. Specifically, men 30 years old or older sleeping less than 4 hours per night were 2.8 times more likely to die within 6 years than those who slept 7 to 8 hours per night. Adult men sleeping more than 10 hours per night had higher death rates (1.8) than normal sleepers. Follow-up studies, controlling for comorbidity factors, confirmed these results.

Sleep disorders medicine is a rapidly evolving subspecialty. Interest in sleep disorders was initially dominated by psychiatry, psychology, and neurology. The past two decades have witnessed discoveries that now make sleep medicine truly multidisciplinary. Recognizing the consequences of sleep-disordered breathing has attracted many pulmonary and internal medicine specialists. Sleep-wake cycle–related endocrinologic and circadian rhythm research has migrated from the laboratory to the clinical arena. Finally, maturation of the Clinical Sleep Society into the American Sleep Disorders Association (ASDA), provides an intellectual home for practitioners interested in sleep disorders.

The ASDA sponsored the development of the International Classification of Sleep Disorders (ICSD). ICSD is the most comprehensive classification of sleep disorders and is referred to frequently (Table 21-1). The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders*, (DSM-IV) includes many ICSD diagnostic classifications. However, the organizational scheme differs, and DSM-IV frequently lumps several ICSD nosological entities into a single diagnostic classification. Nonetheless, because this is a textbook of psychiatry, this chapter organizationally follows the DSM-IV classification (Table 21-2). Overall, DSM-IV differentiates between primary sleep disorders, sleep disorders related to another mental disorder, sleep disorder due to a general medical condition, and substance-induced sleep disorder. In addition to DSM-IV and ICSD, the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) has its own classification of sleep disorders (Table 21-3).

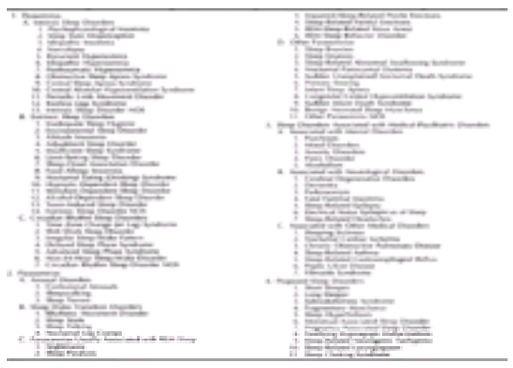

Table 21-1 International Classification of Sleep Disorders

Table 21-2 DSM-IV Classification of Sleep Disorders

Table 21-3 ICD-10 Diagnostic Criteria for Nonorganic Sleep Disorders

PRIMARY SLEEP DISORDERS

Primary sleep disorders are dichotomized into dyssomnias and parasomnias. Dyssomnias include insomnias and hypersomnias. DSM-IV defines insomnia as difficulty initiating sleep or maintaining sleep. Many sleep specialists also would include nonrestful or poor-quality sleep. Hypersomnia is excessive sleepiness. Dyssomnias can be characterized by disturbance in amount, quality, or timing of sleep. By contrast, parasomnias are characterized by abnormal behaviors or physiological events associated with sleep. These events can occur during specific sleep states, during any stage of sleep, or at the transition between sleep and wakefulness.

Dyssomnias are thought of as disturbances involving sleep-wake generating or timing mechanisms. Parasomnias, however, involve inappropriately timed activation of (or failure to suppress) behavioral or physiological systems or both during sleep and sleep-wake transitions.

Dyssomnias DSM-IV lists five specified classifications for primary dyssomnias: primary insomnia, primary hypersomnia, narcolepsy, breathing-related sleep disorders, and circadian rhythm sleep disorder as well as a “not otherwise specified” category. Table 21-4 cross-references DSM-IV primary dyssomnias to the ICSD classification. The table shows ICSD classification differentiates diagnostic categories more fully than DSM-IV. Because in most cases this affects the treatment plan, it is of more than passing academic interest.

Table 21-4 Cross-References Between DSM-IV Primary Sleep Disorders and ICSD Classifications

For example, a patient with psychophysiological insomnia may have a prolonged latency to sleep as a classically conditioned arousal response to getting in bed; treatment would focus on deconditioning. By contrast, if a clinical interview reveals that the patient is engaging in behaviors counterproductive to being able to sleep, inadequate sleep hygiene may be diagnosed. In this case, intervention would focus on improving sleep habits by decreasing or eliminating the sleep-defeating behaviors. However, both cases would be diagnosed as primary insomnia.

Primary Insomnia Primary insomnia is diagnosed only after a patient complains of sleeplessness or poor quality sleep for 1 month or more. The insomnia, or consequent sleepiness, must produce significant distress or impair activities of daily life. Recognizing that insomnia occurs as a symptom of many different underlying conditions, DSM-IV lists various exclusionary criteria for diagnosing insomnia as primary. The clinician must not be able to account for the insomnia as a feature of another primary sleep disorder or secondary to another mental disorder, medical disorder, or substance use. DSM-IV diagnostic criteria are shown on Table 21-5. ICSD defines four distinct nosological entities that would be included in, and further differentiate, primary insomnia: (1) inadequate sleep hygiene, (2) psychophysiological insomnia, (3) sleep state misperception, and (4) idiopathic insomnia.

A. The predominant complaint is difficulty initiating or maintaining sleep, or nonrestorative sleep, for at least one month.

B. The sleep disturbance (or associated daytime fatigue) causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The sleep disturbance does not occur exclusively during the course of narcolepsy, breathing-related sleep disorder, circadian rhythm sleep disorder, or a parasomnia.

D. The disturbance does not occur exclusively during the course of another mental disorder (e.g., major depressive disorder, generalized anxiety disorder, a delirium).

E. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Table 21-5 DSM-IV Diagnostic Criteria for Primary Insomnia

Ms. W. was a 41-year-old, divorced, white female who presented with a 2 ½ year complaint of sleeplessness. She had some difficulty falling asleep (30- to 45-minute sleep-onset latency) and awakened every hour or two after sleep onset. These awakenings might last 15 minutes to several hours, and she estimated having approximately 4.5 hours of sleep on an average night. She rarely takes daytime naps notwithstanding feeling tired and edgy. The patient described her sleep problem with the following words. “It seems like I never get into a deep sleep. I have never been a heavy sleeper, but now the slightest noise wakes me up. Sometimes I have a hard time getting my mind to shut down.” She viewed the bedroom as an unpleasant place of sleeplessness and stated: “I tried staying at a friend's house where it is quiet, but then I couldn't sleep because of the silence.”

At times, Ms. W. was unsure whether she was asleep or awake. She had a history of clock watching (to time her wakefulness) but stopped doing this when she realized it was contributing to the problem. Reportedly the insomnia is unrelated to seasonal changes, menstrual cycle, or time-zone translocation. Her basic sleep hygiene was good. Appetite and libido were unchanged. She denied mood disturbance, except for being quite frustrated and concerned about sleeplessness and its effect on her work. Her work involved sitting at a microscope 6 hours of a 9-hour working day and meticulously documenting her findings. Her final output hadn't suffered, but she had to “double check” for accuracy.

She described herself as a worrier and a type-A personality. The patient did not know how to relax. For example, on vacation she continually worried about things that could go wrong and would not even begin to unwind until she had arrived at the destination, checked in, and unpacked. Even then, she was unable to relax.

Medical history was unremarkable except for tonsillectomy (age 16 years), migraine headaches (current), and diet-controlled hypercholesterolemia. She took naproxen (Naprosyn) as needed for headache. She did not currently drink caffeinated beverages, smoke tobacco, or drink alcoholic beverages. She did not use recreational drugs.

The problem with insomnia began after relocation to a new city and place of employment. She attributed her insomnia to the noisy neighborhood in which she lived. She first sought treatment 18 months earlier. Her family-practice physician diagnosed depression and fluoxetine therapy was started. This made her “climb the walls.” Antihistamines were tried next with similar results. She was then switched to low-dose trazodone (Desyrel) (for sleep) and developed nausea. After these medical interventions, she sought medical care elsewhere. Zolpidem (Ambien), 5 mg was prescribed, but it made her feel drugged, and upon discontinuation she had withdrawal effects. Another family-practice physician diagnosed “nonspecific anxiety disorder” and began treating her with buspirone (BuSpar); an experience she described as “having an alien try to climb out of my skin.” Buspirone treatment was discontinued. Paroxetine (Paxil) was tried for 8 weeks with no effect. Finally, a psychiatrist was consulted, who diagnosed adult attention-deficit disorder (without hyperactivity) and suggested treatment with methylphenidate. At this point, the patient was convinced that a stimulant would not help her insomnia and demanded referral to a sleep disorders center.

Discussion Ms W.'s symptoms fell into the broad category of an insomnia, and the symptoms began after having moved from one city to another. Environmental sleep disorder (noise) and adjustment sleep disorder (new job, city, apartment) were likely initial diagnoses. However, a more-chronic, endogenous problem became operative. What was it? Ms. W. was a “worrier” and meticulous but did not meet diagnostic criteria for personality or anxiety disorders. Dyssomnia associated with mood disorder should be considered in any patient with sleep maintenance problems and early morning awakening insomnia. However, this patient did not have other significant signs of depression. Unfortunately, many patients are misdiagnosed with depression or “masked depression” on the sole basis of an insomnia complaint and are unsuccessfully treated with antidepressant medication. Ms. W.'s job demands long hours with focused concentration. Her job performance had been superior for many years notwithstanding insomnia. Thus, a diagnosis of attention-deficit disorder is unlikely. Idiopathic insomnia implies a childhood complaint, which Ms. W. denied.

The likely working diagnosis was psychophysiological insomnia. She may have some sleep state misperception (sometimes unclear whether she was awake or asleep), but this cannot adequately account for the constellation of symptoms. An initial treatment plan should include further documentation of the sleep pattern using a sleep log. Behavioral treatments would likely benefit this patient. Medications with sedative effects are sometimes useful during initial treatment of psychophysiological insomnia. However, thus far in this patient they have done more harm than good. She is likely to be a challenging patient to treat.

INADEQUATE SLEEP HYGIENE DSM-IV indicates that inadequate sleep hygiene sometimes falls within the primary insomnia classification, depending on the specific sleep hygiene factor involved. ICSD defines inadequate sleep hygiene as “a sleep disorder due to the performance of daily living activities that are inconsistent with the maintenance of good quality sleep and full daytime alertness.” Many behaviors can interfere with sleep and may do so by increasing nervous system arousal proximal to bedtime or by altering circadian rhythms. Treatment should focus on only two or three problem areas at a time. Overwhelming the patient with too many lifestyle changes or a complex regimen seldom succeeds. Some general dos and don'ts are instructive.

PSYCHOPHYSIOLOGICAL INSOMNIA The patient with psychophysiological insomnia has developed a conditioned arousal associated with attempts to sleep. Objects associated with sleep (e.g., the bed, the bedroom) likewise become conditioned stimuli that evoke insomnia. Thus, psychophysiological insomnia is sometimes called conditioned insomnia. Psychophysiological insomnia often occurs in combination with other causes of insomnia, including episodes of stress and anxiety disorders, delayed sleep phase syndrome, and hypnotic drug use and withdrawal. In contrast to the insomnia in patients with psychiatric disorders, daytime adaptation is generally good. Work and relationships are satisfying; however, extreme tiredness can exist. Other features include (1) excessive worry about not being able to sleep; (2) trying too hard to sleep; (3) rumination, inability to clear one's mind while trying to sleep; (4) increased muscle tension when attempting to sleep; (5) other somatic manifestations of anxiety; (6) being able to sleep better away from one's own bedroom; and (7) being able to fall asleep when not trying (e.g., watching TV). The sleep complaint becomes fixed over time. Interestingly, many patients with psychophysiological insomnia sleep well in the laboratory.

Treatment can be difficult. Sleeping pills should be used only sparingly and at the lowest effective dose. Sleeplessness during withdrawal from long-term sleeping pill use typically exacerbates the problem. Stimulus control therapy is recommended to break the conditioning and improve the association between going to bed and being able to fall asleep. Because many patients with psychophysiological insomnia have developed poor sleep habits, improving sleep hygiene is usually beneficial. If muscle tension and rumination at bedtime are prominent features, relaxation therapy is a useful ancillary treatment.

SLEEP STATE MISPERCEPTION For most persons, loss of awareness is a cognitive marker for sleep-onset; however, under some circumstances mental and biological processes become discoordinated. Sleep state misperception is diagnosed when a patient complains of difficulty initiating or maintaining sleep and no objective evidence of sleep disruption is found. For example, a patient sleeping in the laboratory reports taking more than an hour to fall asleep, awakening more than 30 times, and sleeping less than 2 hours the entire night. By contrast, the polysomnogram shows sleep onset occurring within 15 minutes, few awakenings, a 90 percent sleep efficiency, and total sleep time exceeding 7 hours. Sleep state misperception can occur in individuals who are apparently free from psychopathology or represent a somatic delusion or hypochondriasis. Some patients with sleep state misperception have obsessional features concerning somatic functions. Short-term sleep state misperception can occur during periods of stress, and some clinicians believe it can result from latent or ineffectively treated anxiety or depressive disorders. Cognitive relabeling, diffusing the worry about being unable to sleep, or both can help. Interestingly, anxiolytics can profoundly reduce the perception of sleeplessness without remarkably changing sleep physiologically.

IDIOPATHIC INSOMNIA Idiopathic insomnia typically starts early in life, sometimes at birth, and continues throughout life. As the name implies, its cause is unknown; suspected causes include neurochemical imbalance in the brainstem reticular formation, impaired regulation of brainstem sleep generators (e.g., raphe nuclei, locus coeruleus), or basal forebrain dysfunction. Treatment is difficult, but improved sleep hygiene, relaxation therapy, and judicious use of hypnotic medicines are reportedly helpful.

PHARMACOLOGICAL TREATMENTS FOR INSOMNIA For many years, benzodiazepines were far and away the most commonly prescribed sedative-hypnotic medications for treating insomnia. Chloral hydrate (Noctec) is sometimes useful because adverse effects limit its abuse potential. The imidazopyridine-derivative zolbidem (Ambien) has rapidly replaced benzodiazepines. Also, a variety of over-the-counter sleep aids are available. Nonprescription formulas include sedating antihistamines, protein precursors, and other substances. L-Tryptophan was popular and readily available at health food stores until an outbreak of eosinophilia several years ago. Melatonin is now the leader in self-administered food additives believed by some to alleviate sleeplessness.

Regardless of their pharmacology, long-term use of sleeping pills is not generally recommended. Since prescription medicines are rigorously tested in clinical trials, they hold an advantage over the virtually untested over-the-counter remedies. Nonetheless, most premarket study of prescription hypnotics involves short-, not long-term, use. Issues surrounding safety and efficacy of self-administered sedating drugs should not be underestimated. Most sleep-promoting substances work best with time-limited use to overcome short-term insomnia. When properly used, sedative-hypnotics can provide immediate and adequate relief from sleeplessness. One

should prescribe the lowest possible dose for the shortest period of time. As a rule the authors recommend that patients take the medication for two or three nights consecutively and repeat this sequence no more than two or three times. If this does not suffice, a review of the diagnostic possibilities is in order, and other treatments should be considered. Continued long-term dosing can perpetuate the sleep problem. The clinician should realize that these medicines can mask underlying disorders that need to be addressed.

NONPHARMACOLOGICAL TREATMENTS FOR INSOMNIA Some dyssomnias result directly from extrinsic factors. Often problems arise from environmental conditions or maladaptive habits. *Sleep hygiene* refers to basic rules designed to provide circumstances and conditions conducive to sleep. They include a list of things to incorporate into a good sleep ritual and things to avoid.

Universal Sleep Hygiene Sleep-enhancing directives include maintaining a regular sleep-wake schedule; keeping a steady program of daily exercise; insulating the bedroom against excessive noise, light, cold, and heat; eating a light snack before retiring if hungry; and setting time aside to relax before getting into bed. The final item on this list is especially important. It attempts to keep a person from developing conditioned arousal associated with the bed and bedroom. Developing a pattern of emotional arousal upon getting into bed can lay the foundation for psychophysiological insomnia.

Sleep hygiene don'ts are designed to prevent behaviors or relieve conditions incompatible with restful sleep. They include avoiding strenuous exercise immediately before bedtime; abstaining from alcohol, tobacco, and caffeinated beverages in the evening (or early in the day if one is especially sensitive); not watching television in bed; and not chronically taking sleeping pills. Finally, the bed must not become an emotional or sexual battlefield on which couples work through or act out their anger.

Often a few simple alterations in a patient's habits or sleep environment can be effective. The clinician, however, needs to spend time during the interview to review both the patient's routine and its irregularity. In some respects, the essence of insomnia is its variability. The day-to-day changes in behavior and the changing severity of sleeplessness can obscure the factors responsible for the problem. A carefully explained program of sleep hygiene, with follow-up, represents a fairly inexpensive but effective intervention. Furthermore, improving sleep habits can enhance sleep even when the major cause of insomnia is physical.

Stimulus Control Therapy Stimulus control therapy aims to break the cycle of problems commonly associated with difficulty initiating sleep. By attempting to undo conditioning that undermines sleep, stimulus control therapy helps reduce both primary and reactive factors involved in insomnia. The rules attempt to enhance stimulus cues for sleeping and diminish associations with sleeplessness. The instructions are simple, but they must be followed consistently.

The first rule is to go to bed only when sleepy, to maximize success. Second, use the bed only for sleeping. While in bed do not watch television, do not read, do not eat, and do not talk on the telephone. Rule three instructs the patient not to lie in bed and become frustrated if unable to sleep. After a few minutes (do not watch the clock), get up, go to another room, and do something nonarousing until sleepiness returns. The goal is to associate the bed with rapid sleep onset. Rule three should be repeated as often as needed. The last two instructions, four and five, attempt to enhance the mechanisms underlying the circadian and sleep-wake cycles. Rule four is to awaken at the same time every morning, regardless of bedtime, total sleep time, or day of week. The final rule is to totally avoid napping. Stimulus control therapy does work; results may not appear during the first few weeks or month. If rules are continually practiced, the bouts of insomnia lessen in both frequency and severity.

Sleep Restriction Therapy Restricting time in bed can help consolidate sleep for patients who find themselves lying awake in bed unable to sleep. If the patient reports sleeping only 5 hours of a scheduled 8 hours in bed, reduce the time in bed. It is advised, however, not to reduce bedtime to less than 4 hours per night and to warn the patient about the hazards of daytime sleepiness. Sleep at other times during the day must be avoided, except that elderly adults may take a 30-minute nap. The clinician then monitors sleep efficiency (time asleep as a percentage of the time in bed). When patient-estimated sleep efficiency reaches 85 percent (averaged over five nights), time in bed is increased by 15 minutes. Sleep restriction therapy produces a gradual and steady decline in nocturnal wakefulness.

Relaxation Therapy and Biofeedback The most important aspect of relaxation therapy is that it be performed properly and overlearned. Self-hypnosis, progressive relaxation, and deep breathing exercises are all effective if they produce relaxation. Patients require careful and adequate training, simply giving them an instruction tape is not especially helpful. Biofeedback, to provide stimulus cues for physiological markers of relaxation can increase self-awareness. Techniques should be mastered during the day for several weeks before application to the sleep problem. Relaxation techniques readily lend themselves to being combined with sleep hygiene and stimulus control therapies.

Primary Hypersomnia Primary hypersomnia is diagnosed when a patient complains of excessive sleepiness that cannot be accounted for by another primary sleep disorder, mental disorder, medical disorder, or substance use. Unless the hypersomnia is a recurrent type, symptoms must persist for at least 1 month. The most common causes of debilitating sleepiness to exclude are sleep-disordered breathing, insufficient sleep, depression-related sleepiness, and narcolepsy. DSM-IV diagnostic criteria are shown in Table 21-6. ICSD diagnostic entities are well matched to DSM-IV primary hypersomnia, with minor differences in nomenclature; that is, primary hypersomnia is called idiopathic hypersomnia and recurrent type primary hypersomnia is called recurrent hypersomnia. ICSD divides these into distinct nosological classifications because presentation, afflicted group, and presumed etiology differ considerably.

A. The predominant complaint is excessive sleepiness for at least one month (or less if recurrent) as evidenced by either prolonged sleep episodes or daytime sleep episodes that occur almost daily.
B. The excessive sleepiness causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The excessive sleepiness is not better accounted for by insomnia and does not occur exclusively during the course of another sleep disorder (e.g., narcolepsy, breathing-related sleep disorder, circadian rhythm sleep disorder, or a parasomnia) and cannot be accounted for by an inadequate amount of sleep.
D. The disturbance does not occur exclusively during the course of another mental disorder.
E. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify if:
Recurrent: if there are periods of excessive sleepiness that last at least three days occurring several times a year for at least two years.

Table 21-6 DSM-IV Diagnostic Criteria for Primary Hypersomnia

Mr. J. was a 28-year-old, single, African-American male with an approximately 10-year history of fatigue and sleepiness in the daytime. He began to recognize the daytime sleepiness as a problem in his freshman year of college, when he would fall asleep in class or in the dormitory. He admitted that his sleep-wake schedule was disrupted during college due to taking long naps and then having to stay up until 1:00 or 2:00 AM to complete his studies. His grades and social life suffered and he described himself as depressed, isolated, and hopeless about his planned future as a certified public accountant (CPA).

As a child, Mr. J. said he slept “normally.” In high school he felt best with 10 hours of sleep per night and was able to function well in the daytime. Mr. J. denied abuse of alcohol or drugs. He did not use tobacco and drank about 8 to 10 cups of coffee per day. Family history was negative for known sleep or psychiatric disorders. Physical examination findings were noncontributory except for a body mass index of 0.29. Routine laboratory test results were normal, including thyroid-stimulating hormone (TSH).

Mr. J.'s excessive sleepiness continued, notwithstanding some improved sleep hygiene. Improvements included more-consistent bedtime, trying not to nap, and a torturous month-long trial without caffeine. He remained dysphoric and discouraged about his future, blaming his chronic sleepiness as the continuing impediment to his life plans. “I'm just tired of being tired,” he said.

When last seen his bedtime was between 10:00 and 10:30 PM; his wake-up alarm was set for 6:30 AM. He oversleeps at least once a week on work days and sleeps from 10:30 PM until 10:00 AM on weekends in an attempt to “catch up.” He has difficulty awakening and feels unrefreshed or mildly refreshed. By drinking six to eight cups of coffee in the morning, he can usually avoid dozing during the morning. Luckily, he is an independent CPA and can schedule client appointments during this relatively alert time. After lunch he routinely falls asleep at the computer while working. He sleeps for 20 to 60 minutes and is usually awakened by his secretary. He then drinks another two cups of coffee and continues with his work. Unexpected napping can also occur later in the afternoon or evening, and he has “nodded off” while driving. He sleeps alone; however, he has been told that he snores loudly. He does not awaken gasping or choking. He denied hypnagogic hallucinations and sleep paralysis but thought he might feel weak after the rare occasions when he participated in a heated argument.

Discussion Mr. J. had one of the hypersomnias. Most consistent with his history are obstructive sleep apnea syndrome, idiopathic hypersomnia, sleep deprivation in a long sleeper, dyssomnia associated with mood disorder, and narcolepsy. The ancillary symptoms of narcolepsy are absent, with the possible exception of cataplexy. When cataplexy occurs clearly, the diagnosis of narcolepsy is strongly indicated. However, Mr. J.'s possible infrequent weakness during heated arguments is equivocal for cataplexy. His persistent desire for a 10-hour sleep period would be unusual for a patient with narcolepsy.

A long sleeper (ICSD proposed sleep disorder) or an individual with idiopathic hypersomnia requires prolonged sleep periods and may awaken groggy as does Mr. J. The main differentiating feature is that whenever consistently given a chance to have a full nightly sleep period (usually 10 to 12 hours), the long sleeper does not experience excessive daytime sleepiness. Furthermore, patients with idiopathic hypersomnia may display associated autonomic nervous system dysfunction or polysomnographic evidence of elevated slow wave sleep percentage.

Sleepiness associated with a low-grade chronic depression may be difficult to differentiate from other causes of hypersomnia. Polysomnography, psychiatric interview, and psychometric testing can be helpful. Mr. J. related his dysphoria to sleepiness and not vice versa; nonetheless, dyssomnia associated with mood disorder should be considered.

Obstructive sleep apnea syndrome is a strong possibly. Mr. J. is overweight (BMI = 0.29) and snores loudly. Many patients are unaware of gasping or choking for breath. Often, family members witness cessation of breathing during sleep and urge patients to seek treatment. However, Mr. J. lives and sleeps alone.

Polysomnography is recommended for patients suspected of obstructive sleep apnea syndrome, narcolepsy, or idiopathic hypersomnia. These disorders usually require lifelong treatment and have significant morbidity and mortality if untreated.

IDIOPATHIC HYPERSOMNIA Patients with idiopathic hypersomnia have prolonged periods of well-consolidated but nonrestorative sleep. They may sleep more than 10 to 14 hours and awaken unrefreshed. Long naps (1 to 2 hours) are common and likewise not refreshing. Patients may awaken disoriented and groggy (sleep drunkeness). Electroencephalographic studies confirm long sleep periods and sometimes reveal increased slow wave sleep. Otherwise, the sleep pattern is essentially normal (Fig. 21-1). Notwithstanding prolonged nocturnal sleep, patients with idiopathic hypersomnia usually fall asleep rapidly if allowed to nap. Typical age of onset is 15 to 30 years, and once established, it is a lifelong problem.

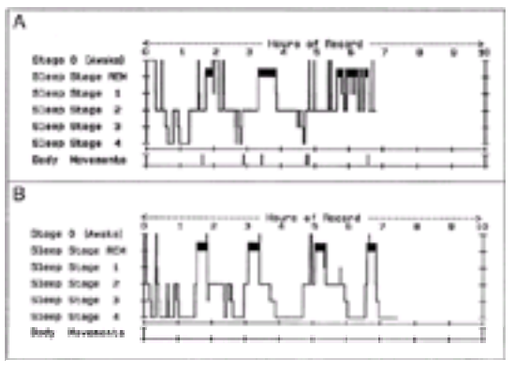

FIGURE 21-1 Sleep stage histograms comparing a normal sleep profile with that of a patient with idiopathic hypersomnia. **A** illustrates the sleep pattern recorded in a healthy young adult. His sleep was interrupted several times by awakening associated with other concurrent laboratory procedures; however, all parameters were within normal limits. **B** depicts the sleep pattern of a patient with severe daytime sleepiness. The patient did not have narcolepsy or apnea. In the laboratory, sleep onset was rapid, sleep continued virtually without interruption, and he was extremely difficult to arouse. This recording was terminated after 7.5 hours to begin daytime testing; however, the patient's sleep revealed little indication of progressing toward a lighter state. In contrast to the normal sleep pattern, slow wave sleep duration was elevated and it continued to be alternated with REM sleep throughout the night.

Although unproved, this dyssomnia is presumed to be of central nervous system origin. Three subgroupings have been proposed. Subgroup 1 includes individuals who have other affected family members, are HLA Cw2 positive, and often have autonomic nervous system dysfunctions. Subgroup 2 is composed of patients with a history of viral infection such as Guillain-Barré syndrome, mononucleosis, and atypical viral pneumonia. The final subgroup are those in whom the dyssomnia is neither postviral or familial.

Response to treatment is usually poor. Psychostimulants are used palliatively to alleviate sleepiness. Pemoline (Cylert), mazindol (Sanonex), methylphenidate (Ritalin), and dextroamphetamine (Dexedrine) are used. Cardiovascular complications must be monitored when using stimulants. Behavioral approaches are recommended but usually provide limited benefit.

RECURRENT HYPERSOMNIA Recurrent hypersomnia, sometimes called periodic hypersomnia, involves episodic bouts of extreme sleepiness that can occur weeks or months apart. Underlying hypothalamic or limbic dysregulation is the suspected cause. Kleine-Levin syndrome, the best-recognized recurrent hypersomnia, predominantly afflicts males in early adolescence, but can occur later in life and in females. In its classic form, the recurrent episodes are associated with extreme sleepiness (18-to-20–hour sleep periods), voracious eating, hypersexuality, and disinhibition (e.g., aggression). A monosymptomatic hypersomnolent form can occur.

Narcolepsy

DIAGNOSIS Narcolepsy is a syndrome of unknown origin characterized by irresistible urges to sleep. These sleep attacks typically occur two to six times a day and last 10 to 20 minutes. They may occur at inappropriate times (e.g., while eating, talking, or driving, and during sex). Narcolepsy may occur in as many as 10 to 60

individuals per 10,000, making it nearly as common as multiple sclerosis. Symptoms typically appear in the second decade of life, but narcolepsy may remain undiagnosed for many years. It is usually associated with one or more of the following symptoms: cataplexy, sleep paralysis, and hypnagogic hallucinations. DSM-IV and ICSD agree on terminology and classification of narcolepsy. Table 21-7 lists DSM-IV diagnostic criteria.

A. Irresistible attacks of refreshing sleep that occur daily over at least three months.

B. The presence of one or both of the following:
   (1) cataplexy (i.e., brief episodes of sudden bilateral loss of muscle tone, most often in association with intense emotion)
   (2) recurrent intrusions of elements of rapid eye movement (REM) sleep into the transition between sleep and wakefulness, as manifested by either hypnopompic or hypnagogic hallucinations or sleep paralysis at the beginning or end of sleep episodes

C. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or another general medical condition.

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 21-7 DSM-IV Diagnostic Criteria for Narcolepsy

Patients with narcolepsy frequently have sleep-onset rapid eye movement (REM) sleep both at night and during daytime naps (Fig. 21-2). The multiple sleep latency test (MSLT) takes advantage of this phenomenon and is used diagnostically. MSLT provides a patient four to six nap opportunities at 2-hour intervals throughout the day. The latency to sleep on each nap opportunity is used to quantify sleepiness and the presence of sleep-onset REM sleep on two or more naps helps confirm narcolepsy.

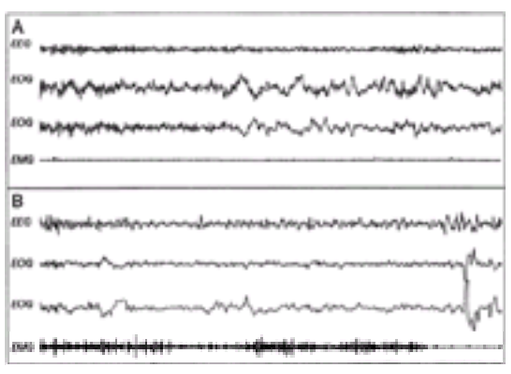

FIGURE 21-2 Polygraphic tracing comparing normal sleep onset with that of a patient with narcolepsy. Each panel illustrates approximately 30 seconds of polysomnographic recording beginning with relaxed wakefulness. **A** (normal sleep progression) shows reduced EEG alpha activity and development of slow rolling eye movements. **B** shows the normally expected abatement of EEG alpha activity associated with increased theta activity and the appearance of a few slow eye movements. However, within 25 seconds (*far right of figure*) a swift loss of muscle tone occurs accompanied by rapid eye movements. This appearance of sleep-onset REM sleep characterizes narcolepsy and is part of the diagnostic criteria.

Cataplexy ranges from transient weakness in the knees to total loss of skeletal tone during full consciousness. It is triggered by emotion (e.g., anger, laughter) and usually lasts for several seconds to several minutes. The patient may fall to the floor and be unable to speak. *Sleep paralysis* is similar to the muscle atonia that normally occurs during REM sleep; however, the patient is awake and unable to move. It also lasts seconds to minutes and can be extremely frightening to the patient. Reportedly, attacks may be terminated by patients vigorously moving their eyes or by someone touching or speaking to them. *Hypnagogic hallucinations* are vivid dreams that occur while the patient is still conscious; they are associated with sleep-onset REM sleep periods.

TREATMENT There is no cure for narcolepsy, but symptom management is possible. Psychostimulants are commonly used to manage sleepiness. REM sleep–suppressing drugs (e.g., many antidepressants) are used to treat cataplexy. Although drug therapy is the treatment of choice, the overall therapeutic approach should include scheduled naps, lifestyle adjustment, psychological counseling, drug holidays to reduce tolerance, and careful monitoring of drug refills, general health, and cardiac status.

Stimulants are used to treat hypersomnolence associated with some sleep disorders. Stimulants are appropriate for certain intrinsic dyssomnias (e.g., narcolepsy, idiopathic hypersomnia, and recurrent hypersomnia) and for some dyssomnias associated with sleep-related medical conditions (e.g., sleep-related asthma). However, they should not be used to treat the sleepiness associated with other intrinsic dyssomnias (e.g., obstructive sleep apnea) or most of the extrinsic dyssomnias. Taking stimulants to overcome daytime sleepiness caused by inadequate sleep hygiene or hypnotic-dependent sleep disorder constitutes drug abuse.

Sleepiness associated with narcolepsy can usually be managed with pemoline or methylphenidate. Sometimes however, amphetamines are needed to combat sleepiness. There is also a newly approved drug called modafinil (Provigil) that is somnolytic without some of the adverse side effects of traditional psychostimulants. Nonetheless, the clinician must monitor usage and be sensitive to developing tolerance. Drug-free holidays may help forestall loss of efficacy. Scheduled naps, as stimulant dose replacements, are also thought to help.

Sleep specialists often prescribe tricyclic drugs or selective serotonin reuptake inhibitors (SSRIs) to reduce cataplexy. This approach capitalizes on the REM sleep–suppressant properties of these drugs. Since cataplexy is presumably an intrusion of REM sleep phenomena into the awake state, the rationale is clear. Many reports indicate that imipramine (Tofranil), protriptyline (Vivactil), and flouxetine (Prozac) are quite effective in reducing or eliminating cataplexy.

Breathing-Related Sleep Disorders

DIAGNOSIS The DSM-IV classification breathing-related sleep disorders includes three ICSD nosological entities: obstructive sleep apnea, (OSA), central sleep apnea, and central alveolar hypoventilation. Sleep apnea is the cessation of breathing during sleep for 10 seconds or more. Episodes of apneas can be associated with oxygen desaturation and cardiac arrhythmias. An apnea episode may be obstructive, central, or mixed. In *obstructive apnea*, breathing stops because of airway blockage; however, respiratory effort continues (Fig. 21-3). *Central apnea* is an absence of breathing due to lack of respiratory effort. In *mixed apnea*, the episode usually starts as a central apnea and then progresses to an obstructive apnea. OSA is most common. In the past few years, upper airway resistance syndrome has gained wide recognition as another form of sleep-disordered breathing. This syndrome is marked by snoring, hypersomnia, and other symptoms associated with OSA; however, cessations of breathing are minimal. A patient with loud disruptive snoring, no apneic events, and neither daytime sleepiness nor insomnia is classified as having the parasomnia primary snoring (Fig. 21-4). Table 21-8 lists DSM-IV criteria for breathing-related sleep disorders.

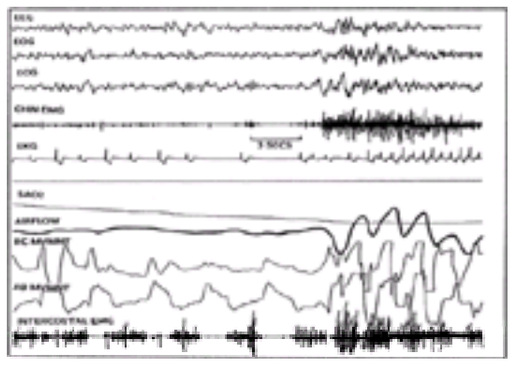

FIGURE 21-3 Obstructive sleep apnea. The airflow tracing shows cessation of breathing while the ribcage (*RC*) and abdominal (*AB*) movement (*MVMNT*) tracings clearly reveal respiratory effort. *Intercostal EMG* also confirms attempt to breathe. As blood oxygen saturation (*SAO2*) plummets, electrocardiographic abnormalities appear. Sinus pauses of more than 3 second's duration occurred. When the patient finally aroused (note *EEG-EOG–Chin EMG*), breathing resumed and was associated with a rebound tachycardia. This patient, whose chief complaint was excessive daytime sleepiness, had more than 200 similar events during a single night of sleep.

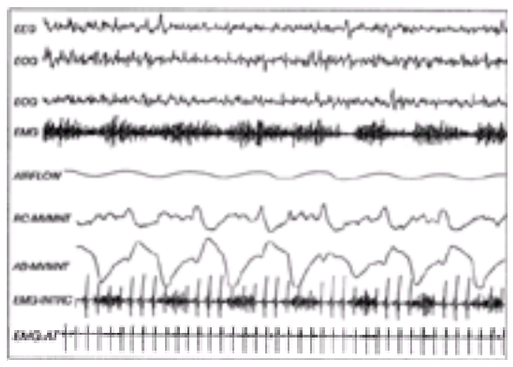

FIGURE 21-4 Polygraphic tracing of snoring. For comparison with obstructive sleep apnea, this tracing shows continual reduction in, but no cessation of, airflow. The movement (*MVMNT*) pattern of the ribcage (*RC*) and abdomen (*AB*), as well as the firing pattern of intercostal muscles (*EMG-INTRC*) indicates considerable respiratory effort, presumably caused by increased airway resistance. Vibration artifact of submental *EMG (fourth tracing from top)* in cadence with breathing reflects the patient's loud, disruptive snoring.

A. Sleep disruption, leading to excessive sleepiness or insomnia, that is judged to be due to a sleep-related breathing condition (e.g., obstructive or central sleep apnea syndrome or central alveolar hypoventilation syndrome).

B. The disturbance is not better accounted for by another mental disorder and is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or another general medical condition (other than a breathing-related disorder).

Coding note: Also code sleep-related breathing disorder on Axis II.

Table 21-8 DSM-IV Diagnostic Criteria for Breathing-Related Sleep Disorder

Predisposing factors for OSA include being male, reaching middle-age, obesity, micrognathia, retrognathia, nasopharyngeal abnormalities, hypothyroidism, and acromegaly. Clinical features associated with obstructive sleep apnea include profound hypersomnolence, snoring, obesity, restless sleep, nocturnal awakenings with choking or gasping for breath, morning dry mouth, morning headaches, and heavy nocturnal sweating. Patients may also have hypertension, erectile dysfunction, depression, nocturia, polycythemia, and memory impairment as a result of OSA.

TREATMENT A wide assortment of treatments are available for OSA. Weight loss helps many patients. However, because losing weight and keeping it off is difficult and unreliably achieved, the prudent clinician should recommend weight loss but rely on other therapies. Because OSA has pathophysiological and potentially life-threatening consequences, aggressive treatments evolved soon after its recognition. The earliest surgical intervention was designed to create a patent airway; thus, in the late 1970s tracheostomies were performed on individuals with severe apnea. There is little doubt that tracheostomy succeeds in creating an airway. Although no longer the preferred treatment, it remains a standard against which newer, more refined therapies are judged.

Second-generation surgical approaches to treating OSA attempt to correct airway obstructions and malformations. Early studies of uvuolopalatopharyngoplasty (UPPP) suggested that soft palate modification effectively relieved most sleep apnea. More-recent follow-up results are less impressive. Approximately 30 to 50 percent of patients with sleep apnea benefit from UPPP. These patients are those with oropharyngeal obstruction; thus, careful attention to selection criteria undoubtedly improves outcome. However, if obstruction occurs in the posterior airway space, maxillomandibular surgery may be appropriate. In retrognathic patients or in patients with cephalometrics revealing compromised posterior airway space moving the jaw forward can achieve impressive normalization of breathing during sleep.

Currently, nasal continuous positive airway pressure is the preferred treatment for sleep-disordered breathing. The continuous positive airway pressure apparatus consists of a fan-driven blower, a nasal mask, and tubing connecting the two. It provides constant pressure through the nasal mask to offset oropharyngeal collapse produced by inspiratory negative thoracic pressure. In this manner it acts as a pneumatic splint, thereby maintaining the airway. When the pressure is properly titrated, even the most severe sleep apnea can be alleviated (Fig. 21-5). Results are usually dramatic. Continuous positive airway pressure therapy is the standard treatment for patients with sleep apnea who can tolerate sleeping with the machine. Its success has made surgical intervention a secondary option, unless patient acceptance issues arise. Sometimes a patient has difficulty exhaling against the constant column of air. In these cases, a variant that allows differential setting of inspiratory and expiratory pressures can be used.

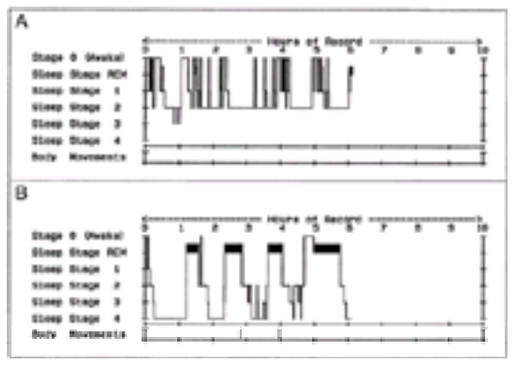

FIGURE 21-5 Sleep stage histogram illustrating the immediate, dramatic improvement in sleep architecture produced by treating obstructive sleep apnea with continuous positive airway pressure (CPAP) therapy. **A** illustrates the abnormal sleep pattern on a night when the patient had more than 200 episodes of obstructive sleep apnea. Sleep is disturbed by frequent awakenings while REM and slow wave (stages 3 and 4) sleep are nearly absent. **B** shows data from the same patient being treated with CPAP on the next night. Normalization of sleep continuity with a massive rebound in REM and slow wave sleep is evident.

A variety of oral appliances have also been used to treat snoring and, in some cases, sleep apnea. The general approach is to either manipulate the position of the mandible or retain the tongue or both. In some cases, these devices may improve airway patency and reduce soft palate vibration. Improved sleep-related breathing does not always reach satisfactory levels; therefore, follow-up sleep studies are needed.

In some patients, sleep-disordered breathing occurs only in the supine position. In such situations, preventing patients from sleeping on their backs may produce beneficial results. Tennis balls sewn onto or placed into pockets on the back of the nightshirt may accomplish this goal.

Finally, several drug therapies have been tried for sleep-disordered breathing. Methoxyprogesterone acetate was originally thought to be helpful but is seldom used now. Likewise, tricyclic drugs sometimes decrease apnea severity by reducing REM sleep, the stage of sleep in which obstructive apnea is usually more frequent. Recently, theophylline (Theo-Dur) was reported to reduce apnea, but further study is needed.

Circadian Rhythm Sleep Disorder Circadian rhythm sleep disorders share the same basic underlying cause, that is, a desynchrony between an individual's internal circadian biological clock and the desired sleep-wake cycle. The circadian [*circa* + *di*(*em*), approximately 1 day] pacemaker usually cycles once every 24 to 26 hours. It is located in the suprachiasmatic nucleus, whose firing patterns oscillate with a pronounced daily rhythm that correlates with the core body temperature cycle. Suprachiasmatic nucleus firing patterns persists even in isole preparations. Mismatched circadian clock and desired schedules can arise from improper phase relationships between the two, travel across time zones, or dysfunctions in the basic biological rhythm. DSM-IV diagnostic criteria (Table 21-9) provide coding-specific subtype coding for three of six ICSD circadian rhythm sleep disorders. The additional three specific ICSD entities can be coded as circadian rhythm sleep disorder unspecified type.

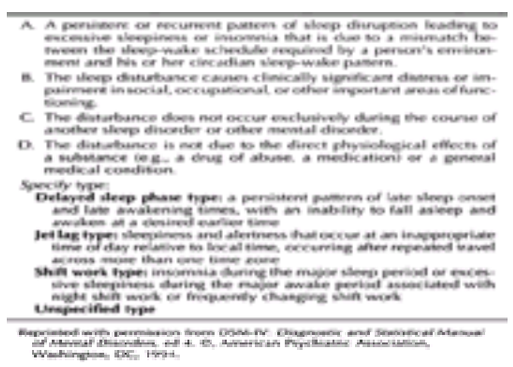

A. A persistent or recurrent pattern of sleep disruption leading to excessive sleepiness or insomnia that is due to a mismatch between the sleep-wake schedule required by a person's environment and his or her circadian sleep-wake pattern.
B. The sleep disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The disturbance does not occur exclusively during the course of another sleep disorder or other mental disorder.
D. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Specify type:

Delayed sleep phase type: a persistent pattern of late sleep onset and late awakening times, with an inability to fall asleep and awaken at a desired earlier time

Jet lag type: sleepiness and alertness that occur at an inappropriate time of day relative to local time, occurring after repeated travel across more than one time zone

Shift work type: insomnia during the major sleep period or excessive sleepiness during the major awake period associated with night shift work or frequently changing shift work

Unspecified type

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4, ©, American Psychiatric Association, Washington, DC, 1994.

Table 21-9 DSM-IV Diagnostic Criteria for Circadian Rhythm Sleep Disorder

DELAYED SLEEP PHASE SYNDROME In an optimal schedule, hours in bed must coincide with the sleepy phase of the circadian cycle. When the circadian sleep-wake cycle lags behind the desired schedule, mismatch is called *phase delay*. Individuals with a delayed sleep phase are more alert in the evening and early nighttime, stay up later, and are more tired in the morning. These individuals are characterized as “owls.” When this desynchrony is severe enough to interfere with daily living, delayed sleep phase syndrome is diagnosed.

ADVANCE SLEEP PHASE SYNDROME In contrast to delayed sleep phase, individuals with *advanced sleep phase* are drowsy in the evening, want to retire to bed earlier, awaken earlier, and are more alert in the early morning. Individuals with this pattern of advanced sleep phase are sometimes called “larks.” When severity impairs routine daily function, the diagnosis is made.

TIME ZONE CHANGE (JET LAG) SYNDROME With the advent of high-speed air travel, an induced desynchrony between circadian and environmental clocks became possible. Thus, the term *jet lag* came into use. Individuals who rapidly cross many time zones induce either a circadian phase advance or a phase delay, depending upon the direction of travel. Typically, a one or two time zone translocation does not produce a sustained problem; however, overseas travel can be marked by great difficulty in adjusting one's sleep-wake routine. Individuals who frequently travel for business can find themselves quite impaired at the time they need to make important decisions. Owls experience greater difficulty adjusting to eastward travel because resynchronization requires phase advance. Similarly, larks theoretically have more difficulty with westward travel. The number of time zones crossed is a critical factor. Normally, healthy individuals can easily adapt to one to two time zone changes per day; therefore, natural adjustment to an 8-hour translocation may take 4 or more days.

SHIFT WORK SLEEP DISORDER Many service industries require 24-hour operation (e.g., transportation, health care). Similarly, as our cultures became more capital intensive, mining and manufacturing became around-the-clock enterprises. The number of individuals doing shift work has been increasing steadily for decades. Shiftworkers commonly suffer from insomnia, excessive sleepiness, or both.

Some individuals require only a short time to adjust to a shift change while others have great difficulty. Frequent shift rotation adds to the problems. Also, to meet social demands, shiftworkers often adopt a nonshifted sleep-wake schedule on weekends and holidays. The result can be severe sleep deprivation and a circadian rhythm poorly entrained to the sleep-wake schedule. The natural low point in the normal sleep-wake rhythm occurs at approximately 3:00 to 5:00 AM, precisely the time frame during which transportation and industrial accidents commonly occur as a direct consequence of sleepiness.

NON-24-HOUR SLEEP-WAKE DISORDER As the name implies, the non-24-hour sleep-wake disorder occurs when the circadian sleep-wake pacemaker has a cycle length longer or shorter than 24 hours and does not reset each morning. The normal cycle length is 24 to 26 hours; however, resynchronization occurs daily. Problems occur incrementally when internal and environmental clocks become more and more out of phase. If the circadian clock's period exceeds 24 hours and does not reset each day, the patient experiences progressively worsening sleep-onset insomnia and daytime sleepiness. Sleep problems peak when circadian and environmental clocks are 12 hours out of phase and then begin to lessen, emulating progressively resolving advanced sleep phase. Eventually the clocks correlate and sleep-wake cycle is normal for a few days, after which the insomnia-hypersomnia cycle begins again. For this reason, non-24-hour sleep-wake disorder has been called periodic insomnia and periodic excessive sleepiness.

IRREGULAR SLEEP-WAKE PATTERN In contrast to the non-24-hour sleep-wake disorder, the irregular sleep-wake pattern occurs when the circadian sleep-wake rhythm is absent or pathologically diminished. The sleep-wake pattern is temporally disorganized, and the timing of sleep and wakefulness is unpredictable. Individuals with this condition have a normal amount of sleep during a 24-hour period; however, it is fragmented into three or more episodes that occur irregularly. Long daytime naps and inappropriate nocturnal wakefulness occur. Except in unusual circumstances, activities of daily life are significantly impaired.

TREATING CIRCADIAN RHYTHM SLEEP DISORDERS Under normal circumstances the internal circadian pacemaker is reset each day by bright light, social cues, stimulants, and activity. When these factors fail to reentrain the circadian rhythm, circadian sleep disorders occur.

Chronotherapy is one technique used to reset the biological clock. It involves progressive phase delays until the circadian oscillator is synchronized with the desired sleep-wake schedule. When individuals are deprived of environmental time cues and told to sleep when they feel sleepy, the typical “day” lasts 25 to 26 hours. This “free running” rhythm suggests that most young and middle-aged adults naturally phase delay. Thus, phase delaying each night by 2 or 3 hours is thought to be easier than phase advancing because it capitalizes on a natural tendency. Halting the phase delay at the appropriate moment and maintaining the desired synchrony can be

a challenge. The patient also has to cope with an odd sleep-wake schedule for the better part of a week during therapy (which can interfere with school or work). For these reasons, light therapy has overshadowed chronotherapy in the past few years.

Recent advances in sleep disorders research indicate that exposing an individual to bright lights (more than 10,000 lux) can alter the endogenous biological rhythm. With precise timing of bright light exposure, the biological clock can be reset. Exposure to light modifies the setpoint of the biological clock. Using core body temperature as a physiological marker, bright lights presented before the temperature nadir produce sleep phase delay. By contrast, light exposure after temperature nadir evokes phase advance. The closer one presents light to the point of inflection (temperature nadir), the more robust the response in altering the cycle. Thus, early morning bright light therapy can be used to phase advance individuals with delayed sleep phase syndrome. Similarly, exposure to bright light in the evening can help patients with advanced sleep phase syndrome.

Light therapy is now being used to reset the circadian rhythm of shift workers, astronauts, and individuals experiencing jet lag. Results are encouraging. However, additional outcome studies are needed to assess the efficacy of this mode of therapy properly. Using bright light exposure to overcome circadian desynchrony represents an exciting frontier in sleep research.

Recent experimental use of melatonin to treat circadian rhythm disorders in the blind has proved successful. Researchers posit that melatonin secretion acts as the biological substrate for the internal circadian oscillator; thus its use may generalize beyond treatment of the blind. Its efficacy in managing sighted patients with disturbed sleep-wake cycles is controversial and currently under investigation.

Dyssomnia Not Otherwise Specified Nine ICSD diagnostic classifications fall within DSM-IV's dyssomia not otherwise specified (Table 21-10) grouping: insufficient sleep syndrome, restless legs syndrome, periodic limb movement disorder, altitude insomnia, environmental sleep disorder, adjustment sleep disorder, limit-setting sleep disorder, sleep-onset association disorder, and nocturnal eating (or drinking) syndrome.

The dyssomnia not otherwise specified category is for insomnias, hypersomnias, or circadian rhythm disturbances that do not meet criteria for any specific dyssomnia. Examples include:

1. Complaints of clinically significant insomnia or hypersomnia that are attributable to environmental factors (e.g., noise, light, frequent interruptions).
2. Excessive sleepiness that is attributable to ongoing sleep deprivation.
3. Idiopathic "restless legs syndrome": uncomfortable sensations (e.g., discomfort, crawling sensations, or restlessness) that lead to an intense urge to move the legs. Typically, the sensations begin in the evening before sleep onset and are temporarily relieved by moving the legs or walking, only to begin again when the legs are immobile. The sensations can delay sleep onset or awaken the individual from sleep.
4. Idiopathic periodic limb movements ("nocturnal myoclonus"): repeated low-amplitude brief limb jerks, particularly in the lower extremities. These movements begin near sleep onset and decrease during stage 3 or 4 non-rapid eye movement (NREM) and rapid eye movement (REM) sleep. Movements usually occur rhythmically every 20–60 seconds, leading to repeated, brief arousals. Individuals are typically unaware of the actual movements, but may complain of insomnia, frequent awakenings, or daytime sleepiness if the number of movements is very large.
5. Situations in which the clinician has concluded that a dyssomnia is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4, Washington, DC, 1994, American Psychiatric Association.

Table 21-10 DSM-IV Diagnostic Criteria for Dyssomnia Not Otherwise Specified

INSUFFICIENT SLEEP SYNDROME Insufficient sleep syndrome results from an individual's disregard for the sleep-wake schedule. It is usually subclinical and occurs in much of the population. Medical help is generally not sought because individuals are aware of the cause of their sleepiness. Insufficient sleep, however, is an insidious killer and is related to many vehicular and industrial accidents (Fig. 21-6). When an individual becomes progressively more and more sleep deprived, eventual payment for the sleep debt will be exacted. Excessive sleepiness associated with insufficient sleep can be unmasked by a heavy meal, low-dose alcohol ingestion, a warm room, or sedentary activity. Although caffeinated beverages are commonly self-administered, sleep extension is the appropriate intervention for insufficient sleep syndrome.

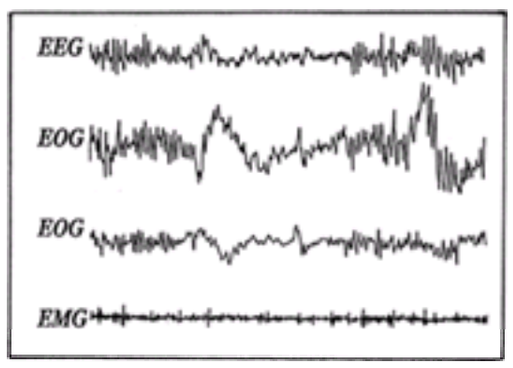

FIGURE 21-6 Microsleep. Figure illustrating one of the most dangerous consequences of excessive daytime sleepiness, regardless of its origin. This patient lapses into a 3-second episode of microsleep (*center one third of figure*) during performance of a vigilance test. Note the transient disappearance of EEG alpha activity and the appearance of slow rolling eye movements. Awakening occurs spontaneously, thereby aborting the microsleep's progression into frank sleep. At normal highway speed, an individual lapsing into microsleep of this duration, would travel almost the length of a football field.

RESTLESS LEGS SYNDROME Restless legs syndrome, once known as Ekbom's syndrome, is characterized by the irresistible urge to move one's legs while trying to fall asleep. Patients often report crawling feelings in their lower legs. A patient may describe the sensation as feeling as if worms or ants are crawling inside the legs (Fig. 21-7). The legs may feel tired or heavy. Moving the legs or walking around are often the only ways to alleviate the discomfort. Patients may feel compelled to pace, swing their legs from the side of the bed, or march in place. Restless legs can be caused by pregnancy, anemia, renal failure and other metabolic disorders. Various drug treatments have been used in restless legs syndrome including benzodiazepines, opiates, and dopaminergic medications (levodopa [Larodopa], bromocryptine [Parlodel], pergolide [Permax], pramipexole [Mirapex], and ropinirole).

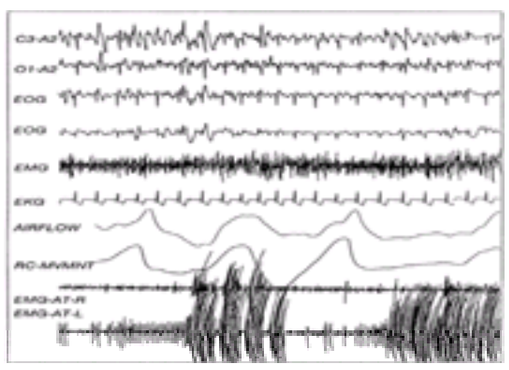

FIGURE 21-7 Restless legs syndrome. This patient presented with complaints of uncomfortable, crawling sensations in the legs when trying to fall asleep. Patients commonly report an urge to move the leg to dispel the sensation. This figure shows a bilateral pattern of leg EMG activity; however, the discharge is more pronounced in the left anterior tibialis (*EMG-AT-L*) than the right (*EMG-AT-R*). This pattern continued for more than an hour as the patient attempted to fall asleep; note that the sharp activity in central and occipital EEG (*C3-A2* and *O1-A2*, respectively) and *EOG* is an EKG artifact and not an EEG abnormality.

PERIODIC LIMB MOVEMENT DISORDER Periodic limb movement disorder (formerly called nocturnal myoclonus) involves brief, stereotypic, repetitive, nonepileptiform movements of the limbs, usually the legs. It occurs primarily in non-REM (NREM) sleep and involves an extension of the big toe. Partial flexion of the ankle, knee, and hip may also occur. These movements range from 0.5 to 5 seconds in duration and occur every 20 to 40 seconds (Fig. 21-8). The leg movements are frequently associated with brief arousals from sleep and as a result can (but do not always) disturb sleep architecture. The prevalence of periodic limb movement disorder increases with aging. It can occur in association with folate deficiency, renal disease, anemia, and the use of antidepressant drugs. Patients with periodic limb movement disorder often complain of nonrefreshing sleep.

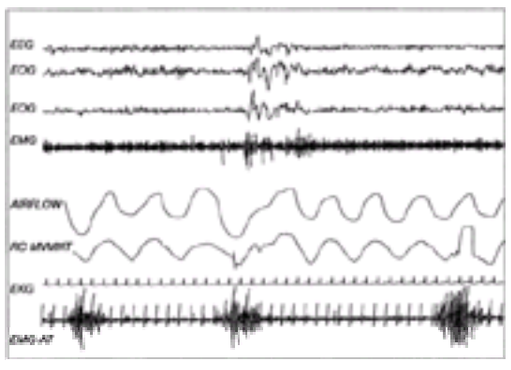

FIGURE 21-8 Periodic leg movements during sleep. This figure of approximately 33 seconds of recording clearly shows a pattern of muscle activity in anterior tibialis (*EMG-AT*). The activity pattern repeats at regular intervals and sometimes produces transient arousals (note *EEG-EOG-EMG*) after movement at center of figure. This patient, whose chief complaint was nonrefreshing sleep with sleep maintenance insomnia, had more than 250 such leg movements during each night he slept in the laboratory.

Benzodiazepines are commonly used to treat periodic limb movement disorder. Clonazepam (Klonopin) and other benzodiazepine hypnotics may not eliminate the leg movements; however, they blunt arousal. Thus, the patient sleeps through the muscle activity, and sleep does not become fragmented by repeated, brief arousals. However, the development of tolerance to the medication remains problematic. Many sleep medicine practitioners now prescribe a low dose L-dopa for periodic limb movement disorder. Unlike sedative-hypnotics, L-dopa suppresses the leg or limb muscle activity rather than merely preventing arousals and awakenings. While the drug is sometimes effective, some clinicians are hesitant to use L-dopa. Potential hazards include unbalancing the dopamine system (risking emergence of Parkinson-like movement disorders) and the chance of dopamine-induced psychosis. However, 5-year follow-up data indicate that these problems are minimal.

ALTITUDE INSOMNIA Altitude insomnia, also known as acute mountain sickness, results from breathing problems associated with sleep-onset ventilatory drive setpoint change. Because oxygen concentration declines as altitude increases, insomnia worsens with increasing elevation. The patient may awaken realizing the apneic event. A person who experiences symptoms usually does so after having been at an altitude of 4000 meters for 72 hours or more. Acetazolamide (Diamox) can increase ventilatory drive via medullary pH receptors and thereby decrease hypoxemia.

ENVIRONMENTAL SLEEP DISORDER Noise, heat, cold, light, bedpartner activity, or perceived danger can induce an environmental sleep disorder. The insomnia or hypersomnia is directly caused by the disturbing environmental factor. Onset, course, and termination of the problem are correlated with introduction, presence, and removal of the specific factor or factors. Thus, treatment involves identifying and removing the environmental irritant.

ADJUSTMENT SLEEP DISORDER Adjustment sleep disorder is a transient insomnia caused by acute stress, conflict, or change. Course is usually brief, lasting a week to several months. The problem quickly resolves when the stressor is removed. The first day of school, an impending examination, a divorce or separation, sudden financial crisis, a change in employment, or the death of a loved one may briefly disturb the normal sleep-wake pattern.

LIMIT-SETTING SLEEP DISORDER Limit-setting sleep disorder occurs when a child takes advantage of an inadequately enforced bedtime schedule. *Childhood insomnia*, as it is sometimes called, involves a child refusing to, or delaying, going to bed when the caregiver stipulates. If unresolved when children reach adolescence and can set their own schedule, inadequate sleep hygiene, insufficient sleep syndrome, or irregular sleep-wake pattern may become the problem. Delay strategies used by children include getting up to use the bathroom, requests for another drink, wanting another bedtime story, or complaints of hunger. Sometimes children claim that they are afraid of the dark, monsters under the bed, or some other factitious creation. The well-intentioned parent may be manipulated out of emotional concern. The problem often begins when the child can climb out of the crib or has been moved to a bed. The caregivers' knowledge and ability to properly set limits is critical; therefore, psychoeducational therapy leads to effective treatment.

SLEEP-ONSET ASSOCIATION DISORDER The sleep-onset association disorder is another dyssomnia characteristic of childhood; however, adult cases have been reported. The classic example involves insomnia when children do not have their pillow, blanket, or teddy bear. Children may become conditioned to fall asleep only while being nursed, rocked, or held. When the sleep-onset associated objects, conditions, or both are not met, sleep is disturbed. It is a problem for the caregiver who is the conditioned sleep-onset queue. In adults, sleep may be disturbed if the television, radio, or night light is not on. Between ages 6 months and 3 years, an estimated 15 to 20 percent of children have this disorder; however, prevalence declines precipitously thereafter.

NOCTURNAL EATING SYNDROME Nocturnal eating (or drinking) syndrome involves an inability to get back to sleep after awakening unless the individual has something to eat or drink. After eating or drinking, return to sleep is normal. Nocturnal eating (drinking) syndrome predominantly affects infants and children; however, adult cases have been reported. The problem is believed to be mainly associated with breast nursing, bottle feeding, or both. An infant will drink 4 to 8 ounces or more at each awakening. Wetting is also excessive. Infants should be able to sleep through the night without feeding after age 6 months; however, afflicted individuals cannot. The syndrome invariably leads to the caregiver becoming sleep deprived. In adults, eating may become obsessional, and patients may be unaware of their behavior. As with sleepwalking, complex behaviors are possible; however, they are not always executed properly. Weight gain may become a problem.

Parasomnias The parasomnias are behavioral or physiological phenomena that occur during, or are potentiated by, sleep. The frequency of occurrence of the different parasomnias vary from nightly to a few times per year. This can make it difficult to document the disorder in a sleep laboratory. When a patient with a parasomnia complaint (e.g., nightmares) is studied polysomnographically and no evidence is found, one should not necessarily conclude that the patient is not afflicted by the suspected sleep disorder.

Ms. R. was a 20-year-old white woman who was referred with symptoms of talking, mumbling, and crying out during sleep. At least twice per week she screamed in her sleep. She was bothered by excessive sleepiness and falling asleep inappropriately (e.g., during a conversation). When inactive, she was tired and sleepy, even after a full 8-hour night of sleep. However, she had energy when motivated and led a vigorous life. Once, she awakened outside of her apartment and her roommate had to let her back in because she had locked herself out. She did not recall the sleepwalking episode or other nocturnal wanderings but remembered yelling sometimes. From the history, crying seemed to occur in light sleep but she rarely recalled any sleep-related thoughts or dreams. She had a history of occasional nightmares and bruxism. The patient used an oral appliance to protect her teeth. Leg kicking and mild snoring without gasping or choking were noted. The patient also complained of leg kicking during sleep. Her sleep-wake schedule was irregular and she averaged between 5 and 7 hours of sleep per night. She occasionally awakened with a headache in the morning.

Previous health history included a hospitalization for febrile convulsions during infancy, ophthamological surgery for strabismus during childhood, and tonsillectomy as a teenager. Health was otherwise excellent. The patient did not smoke tobacco or drink alcohol.

Discussion By history Ms. R. had one or more of the parasomnias. Sleep talking alone does not require a sleep study, but this patient had nocturnal wanderings. Polysomnography, with clinical EEG were indicated to rule out unrecognized nocturnal seizure disorder or other organic factors inducing sleep walking. Sleepwalking is common and not necessarily considered abnormal in young children; however, in the adult it is rare and merits careful evaluation. Ms. R.'s excessive daytime sleepiness was likely due to insufficient sleep (5 to 7 hours per night) and possibly parasomnia-related disruption. Interestingly, many parasomnias are exacerbated by sleep deprivation as is nocturnal seizure disorder.

Sleep studies were performed using comprehensive, attended, laboratory polysomnography. Prior to the overnight study, a clinical EEG was performed, which did not reveal any significant abnormal EEG activity during baseline, photic stimulation, or hyperventilation. An extended EEG montage was used during the sleep study. Overall sleep quality was within the normal range. Sleep efficiency was 96 percent, and latency to sleep was 1 minute. REM sleep percentage was elevated (31 percent), and latency to REM sleep was below normal (57 minutes). Slow wave sleep was normal in percentage, but EEG delta activity was of very high amplitude. The overall macroarchitectural sleep pattern suggested rebound from sleep deprivation.

By contrast, sleep microarchitecture contained many abnormal features. We observed high amplitude paroxysmal EEG bursts. Excessively prolonged sleep spindles were noted and rhythmic K complexes were observed. There was one arousal out of slow wave sleep with rhythmic EEG discharges alternating with sharp waves. Sharps and spikes occurred several times; however the focus was difficult to localize (possibly right temporal lobe). She exhibited frequent body movements and full body jerks, most of which occurred during NREM sleep. Episodes of moaning during slow wave sleep and laughing during stage 2 sleep were followed by high-amplitude theta bursts and REM sleep. Frequent movements and arousals from REM sleep were observed but no REM-related spikes or sharp waves. Seizurelike EEG activity was noted during the night and occurred predominantly during slow wave sleep. However, the patient did not attempt to sleepwalk. Sharp wave and spike activity increased during the final 45 minutes of the sleep study.

The patient did not have any sleep-related breathing impairment and SaO_2 nadir was 90 percent. She had no periodic limb movements during sleep and polygraphic features associated with restless legs syndrome were absent.

Nightmare Disorder Nightmares are anxiety-provoking dreams that occur during REM sleep. Nightmares are typically complex dreams that become increasingly frightening toward the end, culminating in an awakening. Because they are REM sleep–related dreams, terminated with arousal the individual remembers the content. There is seldom any talking, screaming, walking, or striking out associated with nightmares (in contrast to sleep terrors). Table 21-11 lists criteria used to diagnose nightmare disorder.

A. Repeated awakenings from the major sleep period or naps with detailed recall of extended and extremely frightening dreams, usually involving threats to survival, security, or self-esteem. The awakenings generally occur during the second half of the sleep period.
B. On awakening from the frightening dreams, the person rapidly becomes oriented and alert (in contrast to the confusion and disorientation seen in sleep terror disorder and some forms of epilepsy).
C. The dream experience, or the sleep disturbance resulting from the awakening, causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The nightmares do not occur exclusively during the course of another mental disorder (e.g., a delirium, posttraumatic stress disorder) and are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Table 21-11 DSM-IV Diagnostic Criteria for Nightmare Disorder

Ten to 50 percent of children between ages 3 and 6 years may experience occasional nightmares. Approximately 1 percent of adults have nightmares weekly or more often. There may be nightmares in children younger than age 3 years, but unless the child is linguistically fluent, the reported frightening dream may be misunderstood by parents. A subgroup of patients continues to have nightmares after childhood into adolescence and adulthood. In general, however, the prevalence of nightmares decreases with advancing age. Patients with schizotypal personality, borderline personality disorder, and schizophrenia are predisposed to nightmares. However, 50 percent of patients suffering from nightmares have no diagnosable psychiatric illness. One theory is that nightmares are more common in individuals with "thin boundaries" who are open, trusting, and who often have creative or artistic inclinations. Levoclapa b-adrenergic receptor antagonists antagonists), and withdrawal from REM-suppressant medications can induce or potentiate nightmares. Severity ranges from occasional (less than 1 episode per week) nightmares with no evidence of psychosocial impairment to nightly episodes that severely impair daily functioning.

Sleep Terror Disorder ICSD defines the essential feature of sleep terrors as "characterized by a sudden arousal from slow wave sleep with a piercing scream or cry, accompanied by autonomic and behavioral manifestations of intense fear" (Fig. 21-9). Table 21-12 lists DSM-IV criteria for diagnosing sleep terror disorder.

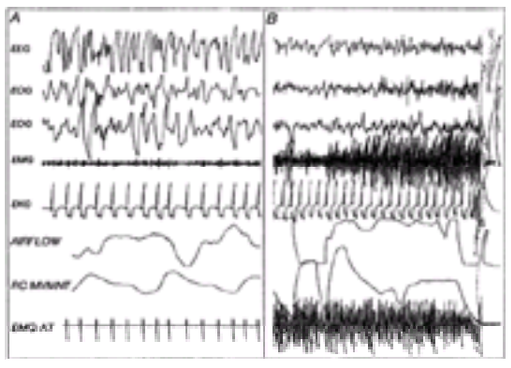

FIGURE 21-9 Polysomnogram of a sleep terror. **A** shows approximately 14 seconds of tracing occurring immediately before the sleep terror. Prominent EEG slow wave activity and other characteristics of stage 4 sleep are seen. **B** shows the awakening, accompanied by tachycardia and movement. EEG activity is ambiguous, and the patient eventually disconnected his electrodes as he thrashed about in bed (visible at *far right of figure*). Although the patient was screaming and greatly agitated, no dreaming was reported. In the morning, he had little recollection of anything having occurred during the night.

A. Recurrent episodes of abrupt awakening from sleep, usually occurring during the first third of the major sleep episode and beginning with a panicky scream.
B. Intense fear and signs of autonomic arousal, such as tachycardia, rapid breathing, and sweating, during each episode.
C. Relative unresponsiveness to efforts of others to comfort the person during the episode.
D. No detailed dream is recalled and there is amnesia for the episode.
E. The episodes cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.
F. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 21-12 DSM-IV Diagnostic Criteria for Sleep Terror Disorder

Sleep terrors occurs in 3 percent of children and less than 1 percent of adults. Children have higher percentages of slow wave sleep than adults. Furthermore, during childhood, delta electroencephalographic (EEG) activity has greater amplitude and is more hypersynchronous. Sleep terrors are not dreamlike, and usually there is no memory of what provoked the fright; however, fragments of brief, vivid images may be reported. After awakening from a sleep terror, a patient is usually unresponsive to stimuli, confused, or disoriented. Vocalizations are usually incoherent. Fever, sleep deprivation, and central nervous system (CNS) depressants may potentiate sleep terror episodes. A familiar pattern has been reported. In children, sleep terrors are not associated with psychopathology; however, adults often have a positive psychiatric history. Severity ranges from less than once per month to almost nightly occurrence (with injury to patient or others).

Sleepwalking Disorder Sleepwalking is another parasomnia associated with slow wave sleep. In its most extreme form, it consists of ambulating during sleep (somnambulism). However, many types of complex behaviors arising from slow wave sleep are typically regarded as sleepwalking. Because it arises from slow wave sleep, the patient is difficult to awaken, confused, and amnestic. Sleep talking may also occur. Table 21-13 lists DSM-IV criteria for diagnosing sleepwalking disorder.

A. Repeated episodes of rising from bed during sleep and walking about, usually occurring during the first third of the major sleep episode.
B. While sleepwalking, the person has a blank, staring face, is relatively unresponsive to the efforts of others to communicate with him or her, and can be awakened only with great difficulty.
C. On awakening (either from the sleepwalking episode or the next morning), the person has amnesia for the episode.
D. Within several minutes after awakening from the sleepwalking episode, there is no impairment of mental activity or behavior (although there may initially be a short period of confusion or disorientation).
E. The sleepwalking causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
F. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 21-13 DSM-IV Diagnostic Criteria for Sleepwalking Disorder

There have been rare reports of homicide and attempted suicide associated with sleepwalking. The person attempting to awaken the sleepwalker may be violently attacked. Common in children, sleepwalking peaks between ages 4 and 8 years and usually disappears after adolescence. Severity ranges from less than once per month without injury to nightly episodes associated with physical injury to patient and others.

Parasomnia Not Otherwise Specified Of the 23 specific parasomnias described in ICSD, DSM-IV lists 3. In DSM-IV, all parasomnias other than nightmares, sleep terrors, and sleepwalking should be coded as parasomnia not otherwise specified (Table 21-14).

The parasomnia not otherwise specified category is for disturbances that are characterized by abnormal behavioral or physiological events during sleep or sleep-wake transitions, but that do not meet criteria for a more specific parasomnia. Examples include:
1. REM sleep behavior disorder: motor activity, often of a violent nature, that arises during rapid eye movement (REM) sleep. Unlike sleepwalking, these episodes tend to occur later in the night and are associated with vivid dream recall.
2. Sleep paralysis: an inability to perform voluntary movement during the transition between wakefulness and sleep. The episodes may occur at sleep onset (hypnagogic) or with awakening (hypnopompic). The episodes are usually associated with extreme anxiety and in some cases, fear of impending death. Sleep paralysis occurs commonly as an ancillary symptom of narcolepsy and, in such cases, should not be coded separately.
3. Situations in which the clinician has concluded that a parasomnia is present but is unable to determine whether it is primary, due to a general medical condition, or substance induced.

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed. 4, © American Psychiatric Association, Washington, DC, 1994.

Table 21-14 DSM-IV Diagnostic Criteria for Parasomnia Not Otherwise Specified

CONFUSIONAL AROUSALS Confusional arousals are the mildest form of the three slow wave sleep–related parasomnias and are common in young children. Typically the child partially awakens from slow wave sleep and sits up. The episodes are marked by confusion, but usually the child lies back down and resumes sleep. Confusional arousals, sleepwalking, and sleep terrors are thought to lie on a continuum.

RHYTHMIC MOVEMENT DISORDER Rhythmic movement disorder has also been called jactatio capitis nocturna, headbanging, headrolling, bodyrocking, and rhythmie du sommeil. It is marked by repetitive, rhythmic movements usually involving the head and neck. It usually occurs at the transition from wakefulness to sleep and may continue during light sleep. Most infants bodyrock. Some think bodyrocking develops from the soothing effect of vestibular stimulation. If the rhythmic movement persists into childhood and involves headbanging, the risk of injury increases. Male to female ratio is 4 to 1. Severity ranges from less than one episode weekly to nightly episodes that produce injury.

SLEEP STARTS Sleep starts are sudden, brief muscle contractions that occur at the transition between wakefulness and sleep in 60 to 70 percent of adults. The contractions commonly involve the legs; however, sometimes there is movement in the arms and head. This “hypnic jerk,” as it is sometimes called, is usually benign. The sleep start, however, can interfere with the ability to fall asleep and may be accompanied by sensations of falling, a hallucinated flash of light, or a loud crackling sound. In severe cases the sleep start produces profound sleep-onset insomnia.

SLEEP TALKING As the name implies, sleep talking in classic form involves unconscious speech during sleep. It is seldom recognized in an individual unless it annoys the bedpartner. It can be induced by fever, stress, or conversing with the sleeper. Somniloquy may accompany sleep terror, sleepwalking, confusional arousals, OSA, and REM sleep behavior disorder.

NOCTURNAL LEG CRAMPS Nocturnal leg cramps are much like wakeful leg cramps. They usually affect the calf and are painful muscle contractions. The pain awakens the sleeper and thereby disrupts sleep. Metabolic disorders, diabetes, and pregnancy are known precipitators. Why some individuals have repeated leg cramps during sleep and not during the day is not known.

SLEEP PARALYSIS Usually occurring at sleep onset or upon awakening, sleep paralysis is an inability to move volitionally. Although it is usually very frightening, it is actually a normal REM sleep feature briefly intruding into wakefulness. Eye movements and respiratory movements are unaffected. In predisposed individuals, sleep deprivation and stress may exacerbate sleep paralysis. The paralysis may last from 1 to several minutes, and voluntary rapid eye movements or being touched by someone often terminates the episode.

IMPAIRED SLEEP-RELATED PENILE ERECTIONS Under normal circumstances, sleep-related erections occur several times nightly in association with REM sleep in all healthy, potent men (Fig. 21-10). Impaired sleep-related erections in the presence of normal REM sleep usually result from organic impotence. The impairment can range from reduction to absence of sleep erections. Sleep-related erections are of particular interest because they are one of the few reliable, objective techniques for differentially diagnosing impotence as psychogenic or organic.

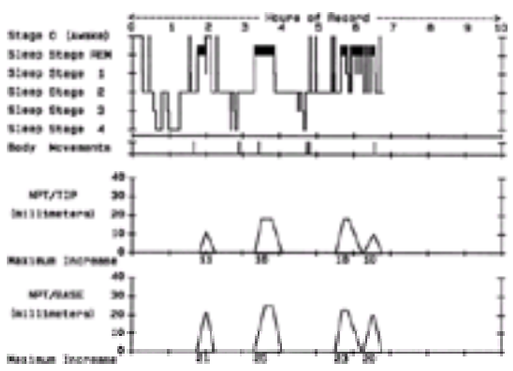

FIGURE 21-10 Nocturnal penile tumescence (NPT). Figure illustrating the pattern of sleep-related penile erections that occur in all normal potent men. Recordings of penile circumference increase from the flaccid stage are made at the penile base and coronal sulcus (tip). Numbers under each episode indicate the maximum circumference increase in millimeters. Under normal circumstances, tumescence episodes are closely linked to REM sleep. Sleep erection monitoring provides useful diagnostic information for parceling out organic and psychogenic components of impotence in men reporting erectile failure.

SLEEP-RELATED PAINFUL ERECTIONS Although sleep-related painful erections are rare, afflicted men have severely disturbed sleep. In its severe form, a patient may awaken with erectile pain from every REM sleep episode, thereby producing profound REM sleep deprivation, hypersomnia, and impaired daytime function. Daytime erections are normal and not painful. Why the normal, naturally occurring REM sleep–related erections are associated with pain is unknown, and to the authors' knowledge no reliable treatment has been described.

REM SLEEP–RELATED SINUS ARREST The rare condition, REM sleep–related sinus arrest is usually discovered incidentally during nocturnal cardiac monitoring for another purpose. REM sleep–related sinus arrest was originally described in otherwise healthy volunteer subjects participating in a sleep research project. There may be vague complaints of chest pain, tightness, or intermittent palpitations during the daytime and lightheadedness, faintness, and blurred vision upon awakening from sleep. Sinus pauses can range from 3 to 9 seconds and may occur in clusters. When clinically warranted, REM sleep–related sinus arrest is managed by implantation of a pacemaker.

REM SLEEP BEHAVIOR DISORDER REM sleep behavior disorder involves dream enactment permitted by a failure of mechanisms that normally produce REM sleep atonia. A patient may run, leap, kick, or punch as part of fictive dream action. These behaviors frequently injure the sleeper, the bed partner, or both. In some cases injuries are serious, involving lacerations, ecchymoses, and fractures. In animals, presumed REM sleep behavior disorder (RBD) can be produced with bilateral peri–locus coeruleus lesions. In humans, there is a suggestion that this parasomnia may result from diffuse hemispheric lesions, bilateral thalamic abnormalities, or brainstem lesions. It may also be prodromal for Parkinson's disease. Clonazepam has been used successfully to treat REM sleep behavior disorder.

SLEEP BRUXISM Sleep-related teeth grinding can produce periodontal damage and the sound can disturb the bed partner (Fig. 21-11). Facial pain and headaches may also result from teeth grinding or clenching. An estimated 85 to 90 percent of the population brux at some time during their lives, but it is clinically significant in only about 5 percent of people. Sleep bruxism may occur during stage 2 or REM sleep or both. Some evidence indicates that REM sleep–related bruxism may be more disruptive and destructive. Oral appliances are commonly used to treat destructive sleep bruxism. Teeth guards, while they do not prevent bruxism, can help protect the teeth from the consequences of destructive grinding. However, such devices usually do not alleviate mandibular pain produced by clenching or grinding the teeth.

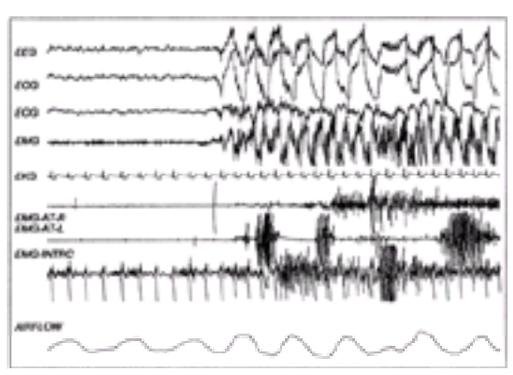

FIGURE 21-11 Sleep-related bruxism. Figure showing approximately 25 seconds of tracing was obtained from a patient during an episode of bruxism. Bruxism can occur during any stage of sleep or wakefulness. The interference pattern of EEG-EOG-EMG channels is typical and reflects the rhythmic jaw movement and grinding of the teeth. This patient had many such episodes, and some caused awakening. Readily observable tooth damage and jaw pain were noted.

SLEEP ENURESIS Nocturnal bed wetting is subclassified as primary (since infancy) or secondary (relapse after toilet training was complete). Prior to toilet training, sleep enuresis is normal and resolves spontaneously before age 6 years. Prevalence progressively declines (30, 10, 5, 3, and 1 to 3 percent at ages 4, 6, 10, 12, and 18 years, respectively). A single recessive gene is suspected in cases of primary enuresis. Embarrassment is usually the most serious direct problem; however, it may lead to psychosocial impairment without appropriate intervention. Properly administered behavioral treatments have good success. Fluid restriction and avoiding caffeine in the evening is helpful. Drug therapy with vasopressin (DDAVP) is also used, especially in older children. Severity can range from less than weekly episodes to nightly wettings associated with severe shame and guilt.

SLEEP-RELATED ABNORMAL SWALLOWING SYNDROME Individuals with sleep-related abnormal swallowing syndrome have inadequate swallowing of saliva during sleep, which produces aspiration. The resulting coughing and choking lead to arousals from sleep and should not be confused with sleep apnea. Sedating medications exacerbate sleep-related abnormal swallowing syndrome.

NOCTURNAL PAROXYSMAL DYSTONIA ICSD describes nocturnal paroxysmal dystonia as characterized by “repeated dystonic or dyskinetic (ballistic, choreo-athetoid) episodes that are stereotyped and occur during NREM sleep.” Additionally, the patient may vocalize. Episodes can last from 15 to 60 seconds to an hour. After its description in ICSD as a parasomnia, nocturnal paroxysmal dystonia was found to result from seizure activity. Carbamazepine (Tegretol) at low dosage usually produces therapeutic benefit.

SUDDEN UNEXPLAINED NOCTURNAL DEATH SYNDROME Sudden unexplained nocturnal death syndrome is sudden death during sleep in apparently healthy

young adults. Laotian, Kampuchean, and Vietnamese male refugees are particularly affected. Asian languages contain descriptors for this syndrome (e.g., *non-laita* [sleep death], *gangungut* [to arise and moan], *pokkuri* [sudden death]). No anatomical pathology has been found to account for the deaths.

PRIMARY SNORING Primary snoring consists of loud snoring in the absence of recurrent apnea or hypopnea episodes. The sound may cause the bedpartner to sleep in another room. To be classified as primary snoring, the individual must not be suffering from excessive sleepiness. Snoring may become louder when the individual sleeps supine or during REM sleep. A variety of oral appliances have been developed to decrease snoring.

INFANT SLEEP APNEA Infant sleep apnea is sleep-disordered breathing that occurs in infants. Breathing events may be central or obstructive sleep apnea episodes. Some data suggest that these "apparently life-threatening events" represent "near-miss" sudden infant death syndrome (SIDS). The episodes may be associated with color change (pallor or cyanosis), noisy breathing during leep, significant oxygen desaturations, bradycardia, or sustained hypoventilation. Complications include need for cardiopulmonary resuscitation, right heart failure, chronic hypercapnia, and severe hypoxia.

CONGENITAL CENTRAL HYPOVENTILATION SYNDROME Sometimes called "Ondine's curse," congenital central hypoventilation syndrome cannot be explained by primary pulmonary disease or ventilatory muscle weakness. The sleep-related hypoventilation results from a failure in automatic control of breathing. Although present at birth, congenital central hypoventilation syndrome may initially be unrecognized. In severe forms, treatment requires continual ventilatory support.

SUDDEN INFANT DEATH SYNDROME Often referred to as "crib death," 80 percent of SIDS occurs when the infants were presumed to be sleeping. No postmortem results explain pathophysiology. There is debate whether the SIDS originates with cardiac or respiratory failure. Ninety percent of deaths occur before age 6 months, with a peak rate at 10 to 12 weeks. Incidence is estimated as 1 to 2 per 1000 live births. Factors increasing SIDS risk include teenage pregnancy, short interpregnancy interval, smoking during pregnancy, preterm birth, lower socioeconomic status, and a familial pattern. The psychological effect of SIDS on the surviving family is usually devastating. Divorce, social isolation, relocation, and depression are common without professional counseling or support group help.

BENIGN NEONATAL SLEEP MYOCLONUS Benign neonatal sleep myoclonus is characterized by asynchronous jerking of limbs and trunk during quiet sleep in neonates. This benign, apparently rare, parasomnia usually begins within the first week of life and may last for a few days or several months. No treatment is recommended.

SLEEP DISORDERS RELATED TO ANOTHER MENTAL DISORDER

Among patients seen at sleep disorders centers with a chief complaint of insomnia, an underlying psychiatric disorder is the most common cause (35 percent). Half (50 percent) of these patients have major depressive disorders. Table 21-15 provides DSM-IV diagnostic criteria for insomnia related to another mental disorder.

A. The predominant complaint is difficulty initiating or maintaining sleep, or nonrestorative sleep, for at least one month that is associated with daytime fatigue or impaired daytime functioning.
B. The sleep disturbance (or daytime sequelae) causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The insomnia is judged to be related to another Axis I or Axis II disorder (e.g., major depressive disorder, generalized anxiety disorder, adjustment disorder with anxiety), but is sufficiently severe to warrant independent clinical attention.
D. The disturbance is not better accounted for by another sleep disorder (e.g., narcolepsy, breathing-related sleep disorder, a parasomnia).
E. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Table 21-15 DSM-IV Diagnostic Criteria for Insomnia Related to Another Mental Disorder

Ninety percent of patients with major depressive disorders have sleep disturbances by EEG criteria. Sleep alterations associated with depression include both general and architectural sleep disturbances. General sleep disturbance consists of increased latency to sleep, increased awakenings, and early morning awakenings with difficulty returning to sleep. Macroarchitectural anomalies include shifting of the center of mass of slow wave sleep from the first NREM-REM cycle to later in the sleep period. Latency to REM sleep is also shortened (Fig. 21-12). Microarchitectural analysis reveals increased eye movement rate (density), especially early in the sleep period.

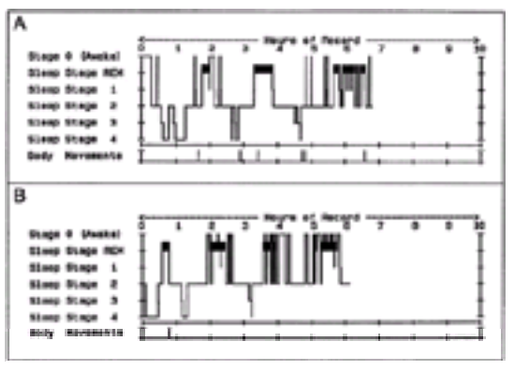

FIGURE 21-12 Sleep stage histograms comparing normal sleep **(A)** with that found in a patient with major depressive disorders **(B)**. Difficulty maintaining sleep and early morning awakenings are common complaints in patients with depression. **B** illustrates the electrophysiological correlates of these complaints beginning after approximately 2 hours of sleep. Sleep continuity becomes disrupted as morning approaches. Also present is a markedly reduced latency to REM sleep, a feature characteristic of this patient population and thought by some to reflect cholinergic-aminergic imbalance.

Selective REM sleep deprivation can improve mood. The effect may persists for days in some patients. Interestingly, traditional medications used to treat depression are powerful REM sleep suppressors (tricyclic drugs and monoamine oxidase inhibitors). SSRIs also suppress REM sleep. Arecoline infusion induces REM sleep in depressed patients, and scopolamine withdrawal mimics depression in normal subjects (including EEG sleep changes). These sleep phenomena support the cholinergic-aminergic imbalance theory of depression.

Patients with mania and hypomania seldom complain of sleep problems, even though they sleep only a short time (2 to 4 hours per night), have very prolonged latency to sleep, and sometimes have reduced slow wave sleep. Patients with schizophrenia have no consistent EEG sleep changes except that they do not have REM sleep rebound in response to REM sleep deprivation. The only other consistent finding is that the patients with schizophrenia frequently deny having slept, even though polysomnographic sleep appears normal. Other findings include reduced slow wave sleep and sometimes reduced REM sleep latency. Patients with anxiety and personality disorders often have sleep-onset, and sleep maintenance insomnia.

Since most depressive disorders are associated with disturbed sleep, a sedating antidepressant, taken at bedtime, can be doubly helpful. The general paradigm is to treat the depression and monitor for improvement in sleep. Interestingly, some antidepressants, especially tricyclics drugs and fluoxetine can produce or exacerbate sleep-related movement dyssomnias (i.e., periodic limb movement disorder, restless legs syndrome, and fragmentary myoclonus).

Sleep disturbance accompanies many psychiatric conditions. Insomnia is included as a diagnostic feature in DSM descriptions of major depressive, generalized anxiety, bipolar, posttraumatic stress, and obsessive-compulsive personality disorders. When anxiety represents a critical component of the sleep problem, anxiolytic

medications may be helpful. Typically taken at bedtime, these antianxiety agents can relieve tension that prevents sleep initiation. Sleep symptoms should be carefully monitored to ensure that a sleep-onset problem is not reemerging as sleep maintenance insomnia. Anxiolytics may also provide relief for patients afflicted by sleep-related panic attacks.

Psychiatric consultation, psychotherapy, or both can facilitate management when applied in conjunction with pharmacological or behavioral therapy. Many patients with insomnia deny emotional problems. Insomnia may be one of several psychosomatic manifestations in “repressors.” Unresolved grief can also underlie insomnia. In some cases, excessive sleepiness can be a symptom of an Axis I or Axis II mental disorder. Table 21-16 shows DSM-IV diagnostic criteria for such hypersomnia related to another mental disorder.

A. The predominant complaint is excessive sleepiness for at least 1 month as evidenced by either prolonged sleep episodes or daytime sleep episodes that occur almost daily.
B. The excessive sleepiness causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.
C. The hypersomnia is judged to be related to another Axis I or Axis II disorder (e.g., major depressive disorder, dysthymic disorder), but is sufficiently severe to warrant independent clinical attention.
D. The disturbance is not better accounted for by another sleep disorder (e.g., narcolepsy, breathing-related sleep disorder, a parasomnia) or by an inadequate amount of sleep.
E. The disturbance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

Table 21-16 DSM-IV Diagnostic Criteria for Hypersomnia Related to Another Mental Disorder

Although not specifically mentioned, psychiatric conditions may also underlie some parasomnias. Sleep terror and nightmares in older children and adults should alert the clinician to a possible history of trauma or abuse. Psychotherapy plays an important role in treating survivors of sexual assault or other forms of posttraumatic stress disorder. Some consider nightmares to be the cardinal feature of posttraumatic stress disorder. The association between sleep and terrifying dreams produces a fear of sleep in some patients. A hypervigilant state ensues, and sleep integrity declines.

The suggestion of psychotherapeutic treatment meets strong opposition from most patients with insomnia. In one review, psychotherapy was recommended to a carefully selected subgroup of patients. Only one third of the group was willing to try it. However, of those patients who began psychotherapy, 75 percent rated it as beneficial.

SLEEP DISORDER DUE TO GENERAL MEDICAL CONDITION

Insomnia, excessive sleepiness, or a parasomnia may be related to a physical condition. Constant pain, obstructive lung disease, neurological diseases, and stroke can disturb the integrity and architecture of sleep, rendering it restless and unrefreshing (see Table 21-17 for DSM-IV diagnostic criteria). Sleep can sometimes unmask pathophysiologies that are difficult to detect with awake-state diagnostic procedures (Fig. 21-13).

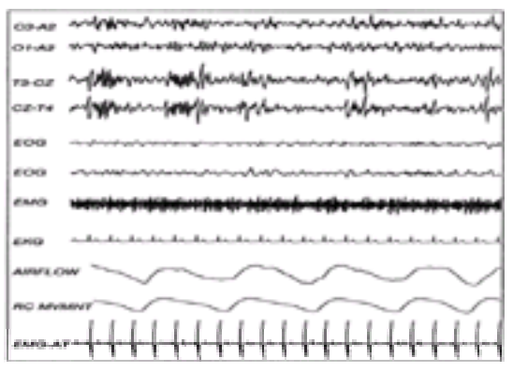

FIGURE 21-13 Polygraphic tracing from a patient with complex partial seizures. EEG sleep spindles of excessive magnitude and duration mixed with sharp activity are visible, especially in temporal lobe derivations. The slow paper speed used in sleep studies makes it difficult to classify this activity precisely; however, the abnormality is apparent. Such a finding should be confirmed with standard clinical EEG techniques during both wakefulness and sleep.

A. A prominent disturbance in sleep that is sufficiently severe to warrant independent clinical attention.
B. There is evidence from the history, physical examination, or laboratory findings that the sleep disturbance is the direct physiological consequence of a general medical condition.
C. The disturbance is not better accounted for by another mental disorder (e.g., an adjustment disorder in which the stressor is a serious medical illness).
D. The disturbance does not occur exclusively during the course of a delirium.
E. The disturbance does not meet the criteria for breathing-related sleep disorder or narcolepsy.
F. The sleep disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify type:
Insomnia type: if the predominant sleep disturbance is insomnia
Hypersomnia type: if the predominant sleep disturbance is hypersomnia
Parasomnia type: if the predominant sleep disturbance is a parasomnia
Mixed type: if more than one sleep disturbance is present and none predominates

Table 21-17 DSM-IV Diagnostic Criteria for Sleep Disorder Due to a General Medical Condition

SUBSTANCE-INDUCED SLEEP DISORDER

Alcohol, anxiolytics, opioids, and sedative-hypnotics all promote sleep by sedation. However, the resulting sleep, while apparently of greater quantity, is of poorer quality. Many benzodiazepine hypnotic medications alter the basic architecture of sleep. Most reduce slow wave sleep, and some reduce REM sleep. Abnormally increased EEG beta and sleep spindle activity result from ingesting some hypnotic drugs (Fig. 21-14). Alcohol may relax a tense person and thereby decrease latency to sleep; however, sleep later in the night is fragmented by arousals. As tolerance develops to chronic drug and alcohol use, increased dosage is needed to sustain effects; lower dosage produces an abstinence syndrome, and sleep regresses to its initial abnormal pattern. Furthermore, during withdrawal or after tolerance has developed, the sleep disturbance can rebound to a more severe level than the initial problem. DSM-IV diagnostic criteria are shown in Table 21-18.

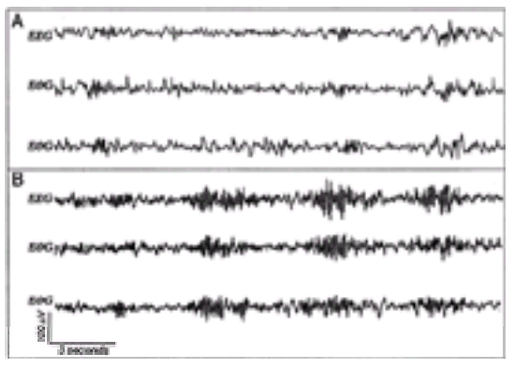

FIGURE 21-14 Tracings comparing normal sleep spindle activity during stage 2 sleep **(A)** with that of a patient chronically using a benzodiazepine **(B)**. This patient had been treated with benzodiazepines for more than a decade (and was currently taking an extremely high dose) before being seen at the sleep disorders center. Sleep was grossly abnormal. The most obvious aberration was the tremendous increase in the frequency, magnitude, and duration of EEG spindle activity **(B)**. Additionally, slow wave sleep was absent, stage 2 was grossly elevated, and spindles even intruded into REM sleep.

Table 21-18 DSM-IV Diagnostic Criteria for Substance-Induced Sleep Disorder

By contrast, psychostimulant use poses a different problem. Cocaine, amphetamine and related stimulants, caffeine, and theobromine all produce CNS arousal that may persist into the sleep period and produce insomnia. Especially in cases of stimulant abuse, an individual usually becomes severely sleep deprived. Over time a massive sleep debt accumulates, and upon substance discontinuation, profound hypersomnia results. This compensatory sleep, or sleep rebound, continues for an extended time (several weeks or more in some instances).

FUTURE DIRECTIONS

Sleep disorders significantly affect an individual's cognitive function, mood, and overall quality of life. Sleep disorders may be primary. However, many are secondary to other mental or medical conditions. Finally, disturbances in sleep or wakefulness may result from substance use or abuse. For psychiatrists, leep disorders are particularly important because insomnia and hypersomnia commonly occur in a wide range of psychiatric conditions. With appropriate treatment, most sleep disorders can be effectively managed. The extent of improvement in a patient's general well being after appropriate treatment is usually dramatic. A good night of sleep is as important to general health as nutrition and exercise.

SUGGESTED CROSS-REFERENCES

The basic science of sleep is discussed in Section 1.19. Mood disorders are covered in Chapter 14 and anxiety disorders in Chapter 15. Light therapy, sleep deprivation, and sleep delay are discussed in Section 31.33 and sleep disorders among the elderly in Section 51.3b.

CHAPTER REFERENCES

*American Sleep Disorders Association: *The International Classification of Sleep Disorders, Revised: Diagnostic and Coding Manual*. American Sleep Disorders Association, Rochester, MN, 1997.

American Sleep Disorders Association: Practice parameters for the indications for polysomnography and related procedures. Sleep *20:*406, 1997.

Chesson AL, Ferber RA, Fry JM, Grigg-Damberger M, Hartse KM, Hurwitz TD, Johnson S, Kader GA, Littner M, Rosen G, Sangal RB, Schmidt-Nowara W, Sher A: The indications for polysomnography and related procedures: An American Sleep Disorders Association review. Sleep *20:*423, 1997.

*Chokroverty S: *Sleep Disorders Medicine: Basic Science, Technical Considerations, and Clinical Aspects*. Butterworth-Honeymoon, Boston, 1994.

*Kryger MH, Roth T, Dement WC, editors: *Principles and Practice of Sleep Medicine*. Saunders, Philadelphia, 1994.

*Loube DI, Gay PC, Strohl KP, Pack AI, White DP, Collop NA: Indications for positive airway pressure treatment of adult obstructive sleep apnea patients: A consensus statement. Chest *115:*863, 1999.

Maquet P: Positron emission tomography studies of sleep and sleep disorders. J Neurol *244*(4 Suppl):S23, 1997.

*Morin CM, Colecchi C, Stone J, Sood R, Brink D: Behavioral and pharmacological therapies for late-life insomnia: A randomized controlled trial. *JAMA 281:*991, 1999.

Nishino S, Mignot E: Pharmacological aspects of human and canine narcolepsy. Prog Neurobiol *52:*27, 1997.

*Pressman MR, Orr WC, editors: *Understanding Sleep: The Evaluation and Treatment of Sleep Disorders*. American Psychological Association, Washington, DC, 1997.

Sack RL, Hughes RJ, Edgar DM, Lewy AJ: Sleep-promoting effects of melatonin: At what dose, in whom, under what conditions, and by what mechanisms? Sleep *20:*908, 1997.

*Thompson DF, Pierce DR: Drug-induced nightmares. *Ann Pharmacother 33:*93, 1999.

*Thorpy MH, editor: *Handbook of Sleep Disorders*. Marcel Dekker, New York, 1990.

Williams RL, Karacan I, Hursch CJ, editors: *Electroencephalography (EEG) of Human Sleep: Clinical Applications*. Wiley, New York, 1974.

CHAPTER 22. IMPULSE-CONTROL DISORDERS NOT ELSEWHERE CLASSIFIED

VIVIEN K. BURT, M.D., PH.D., AND JEFFREY WILLIAM KATZMAN, M.D.

Impulse-control disorders not elsewhere classified include impulse disorders not classified in other sections of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). All of the disorders in this grouping are characterized by the failure to resist an impulse, drive, or temptation to perform some act that is harmful to the patient or others. In most cases the person senses increasing tension or arousal prior to the act and experiences pleasure, gratification, or relief during or following the act. Impulse-control disorders not elsewhere classified include six disorders: kleptomania, pathological gambling, trichotillomania, pyromania, intermittent explosive disorder, and the residual category of impulse-control disorder not otherwise specified, which includes clinical entities such as self-mutilation and compulsive buying.

The 10th edition of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) classifies these entities as habit and impulse disorders. It states that these disorders are “characterized by repeated acts that have no clear rational motivation, cannot be controlled and generally harm the patient's own interests and those of other people. The patient reports that the behavior is associated with impulses to action.” The ICD-10 further states that these disorders “are grouped together because of broad descriptive similarities, not because they are known to share any other important features.” This category excludes habitual use of alcohol or psychoactive substances as well as habitual behavior involving sexual activity.

Disorders of impulse control have long been recognized. In 1838 Jean Etienne Esquirol proposed the term *monomanies instinctives* to describe behaviors characterized by irresistible urges and without apparent motive. The impulse-control disorders were first categorized in the third edition of DSM (DSM-III). The revised third edition of DSM (DSM-III-R) added the diagnosis of trichotillomania, eliminated the diagnosis of isolated explosive disorder, and modified the criteria for the remaining categories. DSM-IV retains the class of impulse-control disorders not elsewhere classified and (with some modification) the diagnoses of pathological gambling, kleptomania, pyromania, trichotillomania, intermittent explosive disorder, and impulse-control disorder not otherwise specified.

Controversy continues about whether these disorders should be placed in a distinct category or should be listed as variations of other Axis I or Axis II disorders. Those who argue that impulse-control disorders are subgroups of other psychiatric disorders point to similarities with obsessive-compulsive disorder, mood disorders, substance dependence, paraphilias, and mental disorders due to a medical condition. Others suggest that impulse-control disorders share features with certain personality disorders such as antisocial personality disorder or borderline personality disorder. DSM-IV, however, emphasizes that the category of impulse-control disorders is a residual group separate from all others.

Recently, investigators have proposed that the impulse-control disorders are more aptly included in an entity that has been called obsessive-compulsive pectrum disorders. Included within this category are the disorders involving preoccupation with bodily appearance or sensations such as the somatoform disorders (i.e., body dysmorphic disorder and hypochondriasis), eating disorders, dissociative disorders, and neurological disorders (Tourette's disorder, Sydenham's chorea, parkinsonism, and autistic disorder). Features proposed to be common to these disorders include epidemiological data (age of onset, family history, and clinical course), clinical presentation (repetitive thoughts or behaviors), proposed causes (serotonin alterations and modified frontal lobe activity), and response to common treatments such as selective serotonin reuptake inhibitors (SSRIs) and behavior therapy.

Some attempt has been made to delineate these disorders further as a subcategory of obsessive-compulsive spectrum disorder, based on the presence of risk-seeking (impulsive) endpoints versus risk-aversion (compulsive) endpoints (Fig. 22-1).

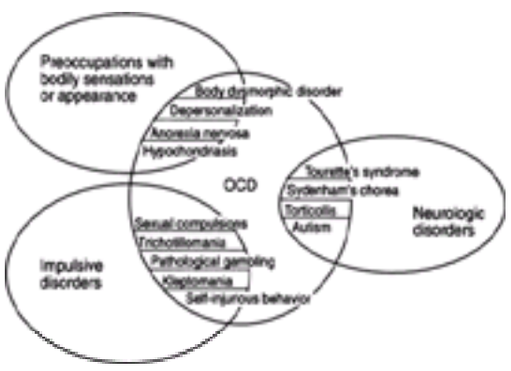

FIGURE 22-1 Obsessive-compulsive-related disorders. (Reprinted with permission from Hollander E, Kwon JH, Stein DJ, Broatch J, Rowland CT, Himelein CA: Obsessive-compulsive and spectrum disorders: Overview and quality of life issues. J Clin Psychiatry *57*[Suppl]:3, 1996.)

Researchers have also pointed to the similarities between impulse-control disorders and mood disorders, suggesting that in fact they may belong to a spectrum of affective disorders. This idea derives primarily from the fact that many patients with these disorders have a comorbid major depressive disorder at some point in their lives, and both disorders respond to selective serotonin reuptake inhibitors. Similarities between impulse-control disorders and bipolar disorders include impulsive and often harmful, dangerous, or pleasurable behaviors, affective lability, high comorbidity with one another and similar comorbidity with other psychiatric disorders, and similar age of onset, clinical course, and response to mood stabilizers and SSRIs.

Thus, while similarities exist between the impulse-control disorders and other major psychiatric disorders, the nature of these relationships awaits further delineation. Similarly, despite theoretical constructs, the primary etiological factor in impulse-control disorders remains unknown. Much remains to be learned about the phenomenology, epidemiology, etiology, and treatment of impulse-control disorders.

KLEPTOMANIA

Definition and History

References to kleptomania, a disorder of nonsensical pilfering, date to the early nineteenth century. The term *kleptomaniac* was coined in 1838 by Jean Etienne Esquirol and Charles-Chretien-Henri Marc, who used it to describe the behavior of a number of kings who stole worthless objects. Kleptomania was also historically considered part of a hysterical disorder in women, often felt to be associated with diseases of the uterus. Kleptomania is included as a disease of impulse control because of the characteristic irresistible urge to steal that is relieved by the act of stealing itself. Persons with kleptomania do not steal items for personal use or monetary gain, often can afford the objects, and may give away, hide, or return the stolen goods. The stealing behavior is usually inconsistent with the general character of the individual. Since the original reference to the disorder more than 150 years ago, there have been no formal, rigorous studies of kleptomania.

Comparative Nosology Kleptomania has been recognized as a discrete disease entity by DSM-III, DSM-III-R, and DSM-IV. The disorder was excluded from the second edition of DSM (DSM-II) and was mentioned in the first edition of DSM (DSM-I) as an accessory term only. Features include a recurrent impulse to steal objects that are not needed for personal use or monetary value. DSM-IV diagnostic criteria stipulate that the stealing is not an expression of anger or revenge, is not

associated with a delusion or hallucination, and is not due to a conduct disorder, manic episode, or antisocial personality disorder.

ICD-10 characterizes pathological stealing as a repeated failure to resist impulses to steal objects that are not acquired for personal use or monetary gain. The objects may be discarded, given away, or hoarded. It goes on to describe a characteristic sense of tension prior to the act of stealing and a sense of gratification during and immediately after the act. The ICD-10 criteria for kleptomania are listed in Table 22-1.

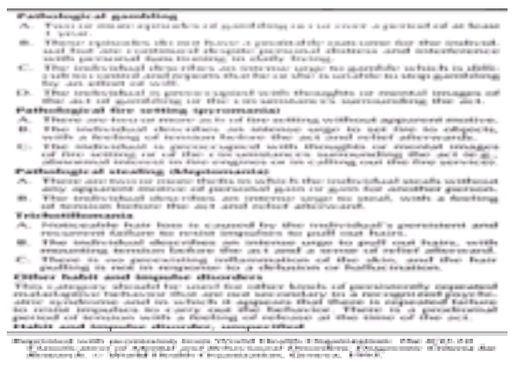

Table 22-1 ICD-10 Diagnostic Criteria for Habit and Impulse Disorders

Epidemiology Because of the absence of studies, little is known about kleptomania. An estimated 6 out of 1000 people in the general population suffer from kleptomania. According to DSM-IV fewer than 5 percent of identified shoplifters actually have kleptomania. However, many factors that influence these data result in an underestimated rate of kleptomania. Persons with kleptomania are ashamed and rarely report their behavior. Since kleptomania is a recurrent disorder, individuals apprehended multiple times are often viewed as recidivist criminals and thus not referred for psychiatric evaluation. Furthermore, some apprehended shoplifters are misdiagnosed as having an antisocial personality.

Information from case reports points to a preponderance of kleptomania in women. Recent studies show that approximately three quarters of reported cases meeting DSM-III-R criteria for kleptomania were in women. However, women are more likely to present for psychiatric evaluation than men. Similarly, they are more likely to be referred by the courts for psychiatric evaluations, whereas male shoplifters are more likely to be sent to prison. Men tend to be older at first presentation, approximately 50 years old, compared with 35 years old for women. The time between onset of symptoms and presentation may be years to several decades.

Etiology No definitive cause for kleptomania is known. It appears that rates of shoplifting in general increase when the availability of goods increases. Psychodynamic theories center on the concept of pathological stealing as a defense against forbidden unconscious impulses, wishes, conflicts, or needs. These impulses or wishes may reflect sexual or masochistic themes, and the act of stealing may represent a mechanism whereby a narcissistically vulnerable individual prevents a fragmentation of the self by carrying out the impulse. Most studies suggest that persons with kleptomania have typically had a stormy and dysfunctional childhood; this history is used to support the analytic premise that impulsive stealing is an attempt to restore early childhood losses.

Phenomenological theories suggest that kleptomania may in fact be part of a larger spectrum disorder. Studies have demonstrated a high comorbidity with mood disorders, both major depressive disorder and bipolar disorders. A higher rate of mood disorders is reported in the families of these patients as well, which suggests that pathological stealing may in fact be part of affective spectrum disorders. However, many individuals with kleptomania demonstrate obsessive and compulsive symptoms including compulsive hand washing, cleaning, checking, hoarding, collecting, and buying. In one study of 20 patients with kleptomania, 7 percent of first-degree relatives met criteria for obsessive-compulsive disorder. This supports the idea of pathological stealing as part of obsessive-compulsive spectrum disorders. A high rate of pathological stealing also appears among patients with eating disorders (particularly bulimia nervosa). Given that kleptomania is at least partially responsive to serotonergic agents, these conditions may in fact have a common pathophysiological link. However, further definitive studies are necessary to determine the viability of this hypothesis.

Diagnosis and Clinical Features The DSM-IV criteria for kleptomania are listed in Table 22-2. The individual with kleptomania usually comes to clinical attention through court referral or through a disclosure in the course of treatment for a comorbid diagnosis. Most patients describe their impulse to steal as intrusive and unpleasurable. However, during the act of stealing, many patients report feeling a sense of pleasure or thrill. Some individuals sexualize the act in some fashion. The act is often done impulsively without any premeditation. Most patients describe their activity as neutralizing a sense of discomfort. Once the impulse has arisen, they report mounting anxiety if they do not carry out the act of stealing. Afterward, individuals frequently feel shame or remorse about the act. Most patients feel that what they do is wrong. Some go to great lengths to protect others and themselves from their impulse to steal and to relieve their sense of guilt about their activity. Individuals may warn shops that are identified as a potential target, return to stores to pay for stolen items, or donate stolen goods to charity. They may avoid places where stealing is a temptation or avoid shopping completely. These individuals rarely reveal their problem to family or friends. Many individuals report traumatic childhoods, but few pinpoint one traumatic event as an immediate precipitant for stealing. Patients with kleptomania have an increased lifetime rate of major mood disorders, anxiety disorders, and eating disorders. They frequently have a history of sexual dysfunction. Persons with kleptomania do not meet the criteria for antisocial personality disorder. Those with a psychiatric disorder in addition to kleptomania generally state that of all their difficulties, stealing causes them the greatest grief.

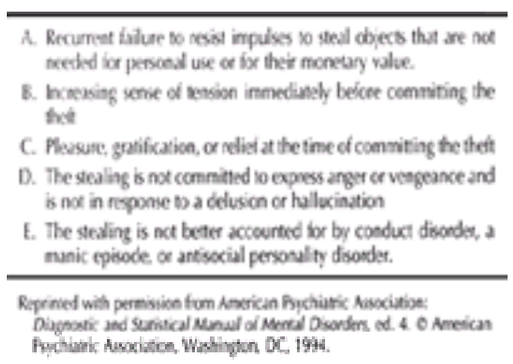

A. Recurrent failure to resist impulses to steal objects that are not needed for personal use or for their monetary value.
B. Increasing sense of tension immediately before committing the theft
C. Pleasure, gratification, or relief at the time of committing the theft
D. The stealing is not committed to express anger or vengeance and is not in response to a delusion or hallucination
E. The stealing is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed. 4. © American Psychiatric Association, Washington, DC, 1994.

Table 22-2 DSM-IV Diagnostic Criteria for Kleptomania

Differential Diagnosis Kleptomania may be distinguished from ordinary stealing in that the objects stolen are not taken for personal use or monetary gain. Occasionally, a shoplifter may feign kleptomania to avoid prosecution by the law; malingering should therefore be considered in the differential diagnosis. Other disorders to be considered include conduct disorder, antisocial personality disorder, and manic episodes that exist as part of bipolar I disorder. In each of those cases the act of stealing is clearly related to the primary disorder, and the patient exhibits a spectrum of associated unacceptable behaviors, of which stealing is only one example. On occasion, a schizophrenia patient may steal in response to a command hallucination. Finally, in certain mental disorders due to a general medical condition, ostensible stealing results from a memory lapse involving failure to remember to pay for merchandise.

Course and Prognosis Kleptomania is a chronic illness, generally beginning in late adolescence and continuing over many years. The spontaneous remission rate and long-term prognosis are unknown.

Treatment The literature describes the use of psychodynamic psychotherapy, behavioral techniques, and somatic interventions in the treatment of kleptomania, with variable outcomes. However, there are no controlled studies in the treatment of kleptomania. The use of long-term insight-oriented psychotherapy is of questionable

value to this patient population. Behavioral techniques have met with some success. Successful pharmacological interventions have been described in case reports, including the use of fluvoxamine (Luvox), amitriptyline (Elavil), imipramine (Tofranil), nortriptyline (Pamelor), trazodone (Desyrel), fluoxetine (Prozac), lithium (Eskalitte), and valproate (Depakote). The use of electroconvulsion therapy (ECT) has also met with some success in case reports.

PATHOLOGICAL GAMBLING

Definition and History Pathological gambling is characterized by the failure to resist the impulse to gamble despite severe and devastating personal, family, or vocational consequences. Pathological gambling should be distinguished from heavy social gambling, professional gambling, or gambling problems that do not meet criteria for pathological gambling.

Comparative Nosology Pathological gambling was first officially recognized in DSM-III. In DSM-III-R the criteria were modified to reflect similarities between pathological gambling and psychoactive substance use and dependence. Included in the DSM-III-R criteria were items that emphasized physiological symptoms such as tolerance and withdrawal. Nevertheless, a number of criticisms were leveled at the DSM-III-R criteria. Some appeared too vague, while others overlapped one another. A survey of treating professionals revealed dissatisfaction with the DSM-III-R criteria and some preference for a return to the DSM-III criteria that emphasized effects on legal, social, vocational, and interpersonal issues in the life of the pathological gambler. A questionnaire combining criteria from DSM-III and DSM-III-R as well as other modifications was sent to several hundred pathological gamblers and substance-abusing control subjects. As a result, a new set of criteria that combined elements of DSM-III and DSM-III-R was incorporated into the revised criteria published in DSM-IV. Importantly, DSM-IV has reinstated the essence of DSM-III descriptors that detail the multilevel consequences of pathological gambling.

ICD-10 characterizes pathological gambling as “a disorder of frequent, repeated episodes of gambling that dominate the patient's life to the detriment of social, occupational, material and family values and commitments.” Diagnoses of exclusion include “excessive gambling by manic patients, gambling and betting not otherwise specified, and gambling in dissocial personality disorder.” ICD-10 criteria are listed in Table 22-1.

Epidemiology Up to 3 percent of adults in the general population may be classified with probable pathological gambling. Based on treatment samples, the typical pathological gambler is an upper-middle-class or middle-class white man between the ages of 40 and 50. However, pathological gamblers in treatment may differ significantly from those in the general population. Surveys demonstrate that rates of pathological gambling are higher among the poor and minorities and that these individuals are underserved by current treatment resources. Although male pathological gamblers outnumber women, the previous ratio of 2 to 1 may be high. Individuals under the age of 30 are probably underrepresented in treatment centers, and data suggest that the prevalence of pathological gambling among adolescents is increasing. Some surveys have shown higher rates of pathological gambling among high school students than in the general population. Pathological gamblers tend to have had an alcohol- or other substance-abusing parent, and approximately 25 percent had a parent who was probably a pathological gambler. Surveys also demonstrate that rates of pathological gambling are considerably higher in locations where gambling is legal.

Pathological gamblers have multiple addictions to alcohol and other substances. The prevalence of pathological gambling among substance abuse patients in treatment programs is also quite high. Substance abusers who are also pathological gamblers appear to have greater social dysfunction than those who are exclusively substance dependent. Dually addicted individuals report higher rates of substance use, greater unemployment, and higher rates of interpersonal conflict. The presence of multiple addictions places pathological gamblers at increased risk for incarceration. One study demonstrated that when substance abuse coexisted with pathological gambling, there was often a history of childhood exposure to gambling and a large family size. In another recent study, early adolescents who revealed impulse-control deficits by self-report were more likely to exhibit gambling problems by late adolescence.

Etiology Numerous theories have been advanced to explain the cause of compulsive gambling despite ongoing losses. Psychoanalytic theory has focused on a number of core character difficulties. One suggestion is that the gamblers are narcissists whose grandiose and omnipotent fantasies lead them to believe they can control events and even predict their outcome. More-classical analytic theories extending back to Sigmund Freud have suggested that compulsive gamblers have an unconscious desire to lose and gamble to relieve unconscious psychic guilt. Behavior theorists view uncontrolled gambling as a learned maladaptive behavior, while cognitive scientists hypothesize that gamblers suffer from a number of erroneous perceptions regarding locus of control.

Several studies have suggested that gamblers' risk-taking behavior may have an underlying neurobiological cause. These theories have centered around both serotonergic and noradrenergic receptor systems. Male pathological gamblers may have subnormal 3-methoxy-4-hydroxyphenylglycol (MHPG) concentrations in plasma, increased cerebrospinal fluid (CSF) MHPG concentrations, and increased urinary output of norepinephrine. Evidence also implicates serotonergic regulatory dysfunction in the pathological gambler. Chronic gamblers have low platelet monoamine oxidase (MAO) activity, a marker of serotonin activity also linked to difficulties with inhibition. However, a recent study of 27 male pathological gamblers found no correlation between platelet MAO activity and psychological ratings. Another study evaluating CSF 5-hydroxyindoleacetic acid (5-HIAA), a serotonin metabolite, found no difference in concentrations between pathological gamblers and controls. Neuroendocrine studies evaluating prolactin response to challenge agents have also yielded mixed results; some demonstrate a serotonin receptor hyposensitivity, and others show a hypersensitivity. The conflicting data call for further studies into the neurobiology of pathological gambling.

Diagnosis and Clinical Features The DSM-IV criteria for pathological gambling are outlined in Table 22-3. Features essential to the diagnosis of pathological gambling include a progressive course, inability to control the behavior despite heavy losses, and disregard for the consequences of the behavior. Like substance abusers, pathological gamblers are driven by the need to achieve an aroused or euphoric state. Pathological gamblers also exhibit tolerance and withdrawal. Thus, they tend to increase the size of their bets or the odds against them to achieve adequate arousal, and when forced to decrease the extent or frequency of gambling, they exhibit characteristic withdrawal symptoms, including irritability, restlessness, depressed mood, and poor concentration.

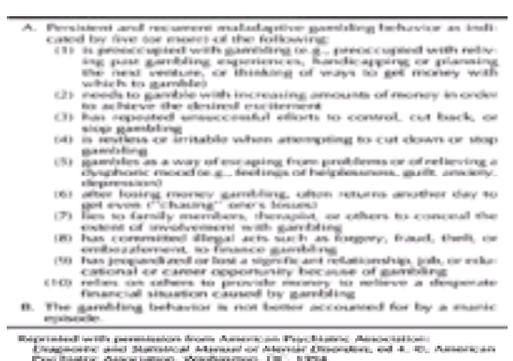
A. Persistent and recurrent maladaptive gambling behavior as indicated by five (or more) of the following:
(1) is preoccupied with gambling (e.g., preoccupied with reliving past gambling experiences, handicapping or planning the next venture, or thinking of ways to get money with which to gamble)
(2) needs to gamble with increasing amounts of money in order to achieve the desired excitement
(3) has repeated unsuccessful efforts to control, cut back, or stop gambling
(4) is restless or irritable when attempting to cut down or stop gambling
(5) gambles as a way of escaping from problems or of relieving a dysphoric mood (e.g., feelings of helplessness, guilt, anxiety, depression)
(6) after losing money gambling, often returns another day to get even ("chasing" one's losses)
(7) lies to family members, therapist, or others to conceal the extent of involvement with gambling
(8) has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
(9) has jeopardized or lost a significant relationship, job, or educational or career opportunity because of gambling
(10) relies on others to provide money to relieve a desperate financial situation caused by gambling
B. The gambling behavior is not better accounted for by a manic episode.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. Ed, American Psychiatric Association, Washington, DC, 1994.

Table 22-3 DSM-IV Diagnostic Criteria for Pathological Gambling

Differential Diagnosis Pathological gambling should be distinguished from social gambling, which does not seriously compromise relationships or vocational performance. Patients whose judgment is impaired by a manic episode or by acute substance abuse do not exhibit the insidious, protracted course that characterizes pathological gambling.

A number of diagnoses are frequently comorbid with pathological gambling. Coexisting and preexisting disorders include multiple substances and alcohol use disorders, major depressive disorder and bipolar disorders, anxiety disorders, attention-deficit/hyperactivity disorder, Tourette's disorder, and personality disorders, particularly narcissistic and antisocial personality disorders.

Course and Prognosis The course of pathological gambling is insidious, and conversion to pathological gambling probably is precipitated either by increased exposure to gambling or by the occurrence of a psychological stressor or significant loss. In males, the onset of pathological gambling begins in adolescence; in females the onset occurs later in life.

The natural history of the illness has been divided into four phases. In the first (winning) phase, a big win stimulates feelings of omnipotence. Women do not generally experience a big win initially. They may see gambling as a means of escaping overwhelming problems in their environment or in their past. Thus, there are apparently

two possible motivators for ongoing gambling activity: action seeking (characterized by the big win) or escape seeking. In the second (losing) phase, the person either has a string of bad luck or begins to find losing intolerable. Gamblers then alter their strategy in an attempt to win back everything at once (chasing). Debts accrue, and there is a sense of urgency and an attempt to cover up both the behavior and the losses by lies. Relationships suffer as the gambler becomes irritable and secretive. In the third (desperation) phase gamblers engage in uncharacteristic, often illegal behaviors. Bad checks are written, funds are embezzled, and they desperately seek ways to obtain money to continue gambling, both to recoup losses and to regain the feeling of arousal characteristic of the initial phase. Relationships deteriorate further. Symptoms of depression appear, including neurovegetative signs, suicidal ideation, and suicide attempts. The fourth and final phase (hopelessness) involves an acceptance that losses can never be made good. Nevertheless, gambling continues, with the main motivator being the attainment of arousal or excitement.

Although a few gamblers seek help while in the winning phase, most seek help much later, generally because their relationships are threatened or they have committed illegal acts.

The course of the disorder is accelerated by the use of alcohol or drugs, the death or loss (possibly through divorce) of a significant other, the birth of a child, physical illness, a job or career disappointment, or increasing interpersonal difficulties. Job promotion or success may also hasten the course of the disorder.

Treatment Some pathological gamblers find participation in Gamblers Anonymous extremely helpful and require nothing more. However, the rate of dropout from this organization is quite high and rate of continued abstinence is low. Psychodynamic psychotherapy attempts to confront the sense of omnipotence and self-deceptions and should address the maladaptive nature of the various defenses. Family therapy is often valuable. Comorbid disorders such as major depressive disorder or substance abuse should be addressed and treated. Inpatient hospitalization may be considered, particularly if the patient is severely depressed and suicidal. Behavioral approaches such as imaginal desensitization, in which relaxation is paired with visualization of avoidance of gambling, have had some success. Little is known about the efficacy of psychopharmacology with pathological gambling. Case reports indicate some benefit from lithium and clomipramine (Anafranil).

Mr. T. was a 47-year-old white male who presented to a community mental health center for evaluation of depression. In the course of the initial interview with the psychiatrist and social worker, he disclosed that he had previously been a quite successful manager of a restaurant in a Las Vegas hotel. He had been married, had two children, and supported a nice lifestyle. He became well known in several casinos and was given free nightly stays by a number of casinos. His gambling accelerated to the point at which he spent the entire family savings. He became unavailable at work as he spent his time around gambling activities. The business collapsed, and his wife later divorced him. He then moved to Los Angeles where he lived in an apartment while managing the other units in the building. He was responsible for collecting the rent from the tenants and turning it over to the building owner. However, on three occasions, he gambled a good portion of this money at the race track. The third episode involved a loss of $8000. He was fired from this job and over the past year was unable to find work as he had no references. Prior to his presentation to the mental health clinic, he was homeless for 6 months and was increasingly hopeless and suicidal.

An initial treatment plan involved mandatory daily attendance at Gamblers Anonymous and work with a sponsor in the 12-step model. He was started on a course of fluoxetine. He met weekly with his case manager in the clinic. His short-term goals involved acquiring housing and employment with continued abstinence from gambling. His long-term goals included reconnecting with his children whom he had not seen in many years.

TRICHOTILLOMANIA

Definition and History Trichotillomania is a chronic disorder characterized by the irresistible urge to pull out one's hair.

Trichotillomania was first characterized by the French dermatologist Francois Hallopeau in 1889. Though the description of the disorder dates back over a century, understanding of the condition remains elusive. Patients with this diagnosis wait years to seek treatment and may never be seen by a physician at all. Patients often disguise the condition with makeup or wigs. This appears to be related to shame about the condition and lack of knowledge about possible treatment strategies.

Comparative Nosology Trichotillomania was first formally recognized as a distinct entity in DSM-III-R, which classified it as an impulse-control disorder largely because like other impulse-control disorders, it is characterized by heightened tension prior to the behavior, inability to resist the impulse, and a feeling of gratification immediately after the act. DSM-IV added to the exclusion criteria both a primary medical etiology and another mental disorder (such as delusional disorder). Additionally, DSM-IV includes severe distress or impairment as a diagnostic criterion.

ICD-10 includes trichotillomania as a disorder of habit and impulse in which “noticeable hair loss is caused by the individual's persistent and recurrent failure to resist impulses to pull out hairs with mounting tension before its act and a sense of relief afterward.” Hair pulling in response to preexisting inflammation of the skin, command hallucinations, and stereotyped movements is not considered trichotillomania. ICD-10 criteria for trichotillomania are in Table 22-1.

Researchers have questioned whether trichotillomania is in fact more similar to the other impulse-control disorders or whether the ritualistic, repetitive nature of the disorder is more properly related to obsessive-compulsive disorder. Some have proposed that hair-pullers are heterogeneous in their presentation; some patients describe class symptoms of impulse-control disorder including a rising tension preceding the action and relief following, while others pull or pluck their hair without these tension and gratification experiences. Many describe pulling their hair in an automatic ritualistic fashion during sedentary activity such as watching television. Still others describe a more highly ritualized behavior including searching for particular kinds of hair, pulling the hair out in a particular manner, ingestion of all or part of the hair, or having to pull out hair to retain body symmetry.

Epidemiology Studies applying strict DSM-IV criteria have found the prevalence to be below 1 percent. However, in one large-scale study of 2579 college freshmen, when the criteria of rising tension and gratification were relaxed to include all hair pullers, the prevalence rose to over 3 percent in women and 1.5 percent of men. In large-scale sample of Israeli 17-year-olds, none of the identified hair pullers met the tension and gratification criteria, and the prevalence of obsessive-compulsive symptoms was significantly elevated. Previous studies that suggested female to male ratios up to 9 to 1 have involved much smaller groups of subjects. These clinical population studies may also have been confounded by such factors as the relative reluctance of males to seek psychiatric treatment. Males with trichotillomania tend to pull from the beard and moustache area and can shave this area to camouflage the problem; this too may confound determination of the female to male ratio.

The mean age of onset of trichotillomania is in the early teens, most frequently before age 17. Most patients do not present for treatment at this time and may wait for decades before detection. The literature describes both early- and late-onset forms of the disorder, though this has not been demonstrated systematically. The early-onset form, beginning before age 6, appears to have a more evenly divided sex distribution and tends to remit more readily, responding to simple interventions such as suggestion, support, and simple behavioral techniques. The late-onset form, typically occurring after age 13, tends to become a chronic disorder with a less hopeful prognosis.

Etiology The cause of trichotillomania remains unclear and both psychological and biological mechanisms have been proposed. Psychodynamic explanations suggest that the disorder is a response to loss or separation in childhood. Mothers of patients are often characterized as critical, and fathers are frequently passive or emotionally weak. Alternatively, behaviorists point out that the disorder is common in children and adolescents, and when it persists, it takes on the characteristics of a habit that remits with behavioral techniques. Biochemical explanations stem from the finding that patients have an increased prevalence of both mood and anxiety disorders and some symptoms of obsessive-compulsive disorder. The family history of patients with trichotillomania often includes tics, habits, and obsessive-compulsive symptoms. This disorder responds preferentially to serotonergic agents, which lends credence to a biochemical relation to either mood or anxiety disorders. However, since serotonergic agents are not always effective, perhaps the disease has more than one biochemical cause. Another biochemical theory points to the potential role of the opiate system, largely on the basis of the response of dogs with canine acral lick dermatitis to opioid antagonists. It was proposed that patients with trichotillomania have a general hypoalgesia that allows them to continue plucking hair without the perception of pain, but a recent study found no difference in pain perception thresholds.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for trichotillomania are presented in Table 22-4. Trichotillomania can be quite difficult to diagnose for a number of reasons. Although this disorder disrupts most patients' lives, they tend to deny the illness and frequently disguise it successfully for decades. Most often, people with this disorder pull hair from the scalp, most commonly the vertex, though people also pull frequently from the temperoparietal, occipital, and frontal regions. Generally, hair plucking involves multiple sites on the scalp or other body regions such as the eyebrows, eyelashes, facial hair, and the pubic area. The average dermatologist sees three to seven cases of trichotillomania a year. The affected site usually shows a mixture of short and long hairs in a linear or circular pattern. A biopsy is often necessary to confirm the diagnosis and distinguish it from alopecia areata or tinea capitis. (Fig. 22-2, Fig. 22-3 and Fig. 22-4).

A. Recurrent pulling out of one's hair resulting in noticeable hair loss.
B. An increasing sense of tension immediately before pulling out the hair or when attempting to resist the behavior.
C. Pleasure, gratification, or relief when pulling out the hair.
D. The disturbance is not better accounted for by another mental disorder and not due to a general medical condition (e.g., a dermatologic condition).
E. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 22-4 DSM-IV Diagnostic Criteria for Trichotillomania

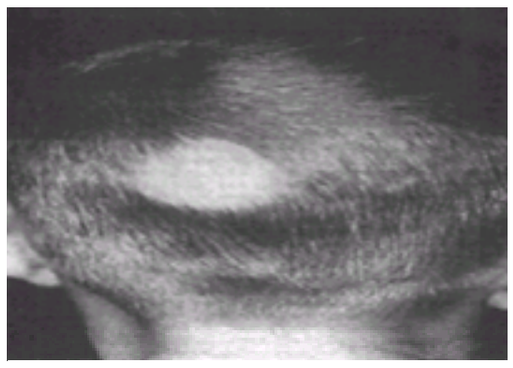

FIGURE 22-2 Tinea capitus in a young male due to *Microsporum audouinii*. (Courtesy of Victor Newcomer, M.D.)

FIGURE 22-3 Alopecia areata. (Courtesy of Victor Newcomer, M.D.)

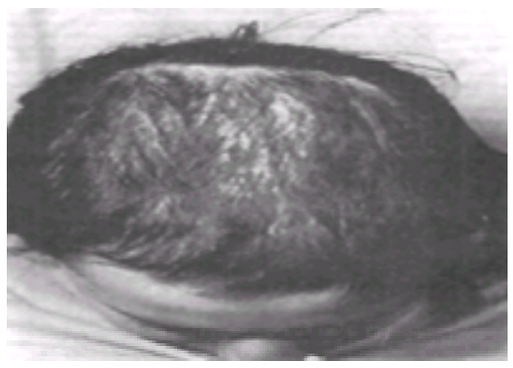

FIGURE 22-4 Trichotillomania with a mixture of short and long hairs with hair loss in a linear fashion. (Courtesy of Victor Newcomer, M.D.)

A 48-year-old woman presented to her internist for a routine physical examination. She became somewhat tearful during the interview, explaining that she was under tremendous pressure at home taking care of her three children and that she recently separated from her husband. During the physical examination, the patient pointed to her head saying that her hair had been falling out during this recent period of stress and that she was somewhat concerned about this. The physician noted that the patient had two distinct areas of hair thinning, one at the vertex of the scalp and one in the frontal region. He also noted that the eyebrow regions were covered with pencil. Otherwise results of the physical examination and routine blood work were normal. He suggested the possibility of a condition other than spontaneous stress-induced hair loss and referred her to a dermatologist. Quite reluctantly, she agreed to the referral.

The patient followed up with the dermatologist who noted a mixture of long and short hairs in the thinning regions of the scalp. Pathological specimens were negative. When the dermatologist asked the woman whether she might be pulling out her hair, she hesitatingly disclosed that she was. Tearfully, she requested complete confidentiality and asked the physician if he had ever encountered this before. When he reassured her that he had numerous patients with her condition and that treatment was available, she was visibly relieved.

Differential Diagnosis Trichotillomania may be mistaken for obsessive-compulsive disorder because both involve patterned, ritualistic behavior. The types of behavior in which patients with obsessive-compulsive disorder participate involve a conscious urge to achieve a precise goal, whereas trichotillomania occurs in response to an irresistible urge without a goal as the primary motive. Recently, however, discussion has focused on features common to trichotillomania and obsessive-compulsive disorder. Furthermore, since a large proportion of patients with trichotillomania also meet criteria for obsessive-compulsive disorder, the idea that these disorders are in fact related in an obsessive-compulsive spectrum has become more popular.

In a recent small, preliminary study of 61 patients with Tourette's disorder, obsessive-compulsive disorder, or both, it appeared that having a tic disorder increased the likelihood of trichotillomania. Although obsessive-compulsive disorder was not correlated with hair pulling, some form of trichotillomania may be more likely to coexist with the disorder than with obsessive-compulsive disorder.

Since patients with trichotillomania often avoid medical attention for years despite desiring relief from their symptoms, trichotillomania may be distinguished from factitious disorder with physical features, a disorder in which people mutilate themselves to get medical attention. Patients with stereotypic movement disorder have stereotypic, rhythmic movements and, unlike those with trichotillomania, are depressed about their symptoms. Because of high rates of comorbidity with multiple other psychiatric disorders including major depressive disorder, generalized anxiety disorder, alcohol and other substance abuse or dependence, social phobia, and various

personality disorders, trichotillomania is often not recognized as a discrete disorder. Trichotillomania can also be difficult to distinguish from hair loss from organic causes.

Course and Prognosis The clinical course is varied. Some evidence suggests that early-onset trichotillomania may be more self-limited and more easily treated, whereas the later-onset version is more chronic and tends to be refractory to treatment. The reluctance of the patient to seek treatment and the inconsistency of treatment strategies to bring about long-term relief of symptoms are both responsible for the chronicity of late-onset trichotillomania.

Trichotillomania can lead to serious long-term complications including infection at the hair-pulling site, change in texture of the hair, and carpal tunnel syndrome from repetitive pulling. More serious complications may arise from the avoidance of pelvic examinations because of shame over the exposure of pulled pubic hair. Finally, hair ingestion can lead to development of trichobezoars and resultant bowel obstruction and peritonitis.

Treatment Treatment strategies include a wide range of psychotherapeutic and pharmacological modalities. Regardless of the method used, the treating physician must tell the patient what is known about the condition and underscore that the patient is not alone with the problem. The physician must remember that patients often wait decades before seeking treatment and frequently feel both humiliated and devalued when their condition is finally revealed.

Psychological treatment approaches have been varied. For young children, kind, supportive acknowledgment of the problem with common-sense instructions about ways to forestall future hair pulling may be all that is necessary to treat the illness. For more chronic forms of the disorder, treatment approaches have included insight-oriented dynamic psychotherapy, behavioral techniques, and hypnotherapy. For children, family therapy should be included to understand and attempt to reduce the various family stressors that may have partially precipitated the illness. Psychoanalytic psychotherapy has underscored losses that may have precipitated the hair-pulling behavior and has been intermittently successful. Behavioral modalities are directed toward maximizing the assets of the patient and may involve simple self-monitoring strategies or more-complex programs such as covert desensitization or habit reversal, the latter seeming to have the greatest efficacy.

Psychopharmacological treatment has incorporated a wide range of drug classes. In recent years, trials have investigated serotonergic agents, primarily because of the initial finding that canine acral lick dermatitis responds to serotonergic agents and the growing awareness that certain features are common to trichotillomania and obsessive-compulsive disorder. Though multiple case reports and open studies have demonstrated the efficacy of these agents for trichotillomania, double-blind, placebo-controlled studies have yielded mixed results. Efficacy was demonstrated for the serotonergic agent clomipramine, though the sustainability of symptom improvement remains controversial. Two placebo-controlled studies conducted with fluoxetine showed no advantage of drug over placebo, but a study comparing fluoxetine and clomipramine found them equally efficacious in promoting symptom reduction. The efficacy of fluoxetine therefore remains unclear. Case reports have also shown efficacy for sertraline (Zoloft), paroxetine (Paxil), and trazodone, though no controlled studies have been done. The addition of pimozide (Orap) to fluoxetine or clomipramine also appears to be beneficial.

Case reports have also shown efficacy of some of the tricyclic drugs, lithium, the MAO inhibitor isocarboxazide (Marplan), the anxiolytic agents buspirone (BuSpar) and clonazepam (Klonopin), the progestogen levonorgestrel (Ovnette), and the anorectic agent fenfluramine. One placebo-controlled study of the opioid antagonist naltrexone (ReVia) showed reduction in severity of symptoms.

PYROMANIA

Definition and History Pyromania is the recurrent, deliberate, purposeful setting of fires. The disorder has been identified as a discrete entity for more than a century. With the designation of specific criteria to define it as a characterizable disorder in the DSM-III, most fire-setting behavior does not appear to qualify as pyromania. Nevertheless, despite the addition of pyromania in DSM-III-R and again in DSM-IV as a disorder under the category of impulse-control disorders not elsewhere classified, there are no rigorous studies of the disorder and no clear data on epidemiology, course, prognosis, etiology, or treatment. Pyromania as a valid diagnosis, either individually or as a subcategory under impulse-control disorder, has been questioned.

References in the literature to fire setting extend back to the nineteenth century, when conflicting schools of thought were represented by Marc, who believed that pyromania was an authentic disorder, and Wilhelm Griesinger, who viewed pyromania as symptomatic of other psychological processes. Prominent references to pyromania as a discrete disorder surfaced in the 1930s and in a 1951 monograph on the subject by Nolan Lewis and Helen Yarnell.

Comparative Nosology Pyromania was first mentioned in DSM-I as an accessory term but was excluded in DSM-II. The disorder was officially recognized in DSM-III, where it was classified under the larger category of impulse-control disorders not elsewhere classified, along with pathological gambling, kleptomania, intermittent explosive disorder, and isolated explosive disorder. Formal reference to pyromania was retained in DSM-III-R, with some modification of the criteria to objectify the characterization. DSM-IV retains pyromania with no significant changes to the defining criteria.

To qualify as pyromania, fire-setting behavior must, like other impulse-control disorders, be a response to an irresistible urge and a rise in tension and be followed by relief, pleasure, or gratification on commission of the act or behavior. Fire setting for purposes of monetary gain (arson), expression of political ideas, concealment of criminal activity, or demonstration of anger or vengeance may not be classified as pyromania.

DSM-III-R criteria incorporated the impulsive nature of the act of setting fires operationally, removing the DSM-III criterion that characterized pyromania as an inability to resist impulses to set fires. Additionally, DSM-III-R added criterion C, which describes the disorder in terms of preoccupation or fascination with fires or associated effects. Criterion E expanded the exclusionary characteristics, stipulating that fire setting cannot occur in response to a delusion or hallucination. In DSM-IV criterion F stipulates that the fire setting may not occur as part of a manic episode and cannot be accounted for by conduct disorder or antisocial personality disorder.

ICD-10 criteria for pyromania are similar (Table 22-1). Diagnostic criteria include two or more acts of fire setting, lack of motive, feelings of tension before the act with subsequent relief, and preoccupation with fire and mental images of fire setting.

Epidemiology Retrospective studies of the literature on arsonists indicate that true pyromania is rare. Fire setters tend not to meet DSM-IV criteria for pyromania because of the exclusionary criteria. Most arsonists set fires in search of monetary gain, in response to psychotic experiences, or in conjunction with a conduct or antisocial personality disorder. Because few studies of pyromania using DSM-IV criteria exist, little is really known about the epidemiology of pyromania. Nevertheless, studies of fire setters have revealed a preponderance of males with a history of fascination or involvement with fire dating back to childhood or adolescence. Recent studies have shown that a high percentage of fire setters are under the influence of alcohol at the time of the arson and that many have alcohol dependence. The correlation of fire setting and alcohol appears stronger in adults than in adolescents. Recent studies have also shown a high degree of psychosis among arsonists. Finally, a substantial number of fires are set by homeless, mentally ill persons.

Etiology The cause of pyromania is not known. Both psychodynamic and biological hypotheses have been proposed. Sigmund Freud suggested that fire setting represents a masturbatory equivalent with homosexual underpinnings. Some researchers proposed revenge as an underlying motive for repeated acts of fire setting. Some have suggested that fire setting represents a primitive attempt by persons with limited social skills to communicate.

Attempts have been made to identify biological markers common to fire setters. Lowered 5-HIAA and CSF MHPG concentrations suggest possible serotonergic or adrenergic involvement, which suggests that pyromania may be part of an affective spectrum disorder. The possibility of reactive hypoglycemia has been raised by reported blood glucose concentrations on glucose tolerance testing. Further studies using controlled subjects and operational criteria are needed to expand these preliminary findings.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for pyromania are listed in Table 22-5. Fire setting has been suggested to be common to many different diagnoses, which raises questions about the entity of pyromania as an impulse-control disorder. Fire setters, however, seem to share some common clinical features. Many have a history of involvement with fire or fascination with fire early in life. DSM-IV suggests that there may be advanced preparation for the act itself. Most fire setters deny the act when questioned or apprehended. Many are often seen lingering in the vicinity of the arson after the fact or are involved in setting off false fire alarms.

A. Deliberate and purposeful fire setting on more than one occasion.
B. Tension or affective arousal before the act.
C. Fascination with, interest in, curiosity about, or attraction to fire and its situational contexts (e.g., paraphernalia, uses, consequences).
D. Pleasure, gratification, or relief when setting fires, or when witnessing or participating in their aftermath.
E. The fire setting is not done for monetary gain, as an expression of sociopolitical ideology, to conceal criminal activity, to express anger or vengeance, to improve one's living circumstances, in response to a delusion or hallucination, or as a result of impaired judgment (e.g., in dementia, mental retardation, substance intoxication).
F. The fire setting is not better accounted for by conduct disorder, a manic episode, or antisocial personality disorder.

Table 22-5 DSM-IV Diagnostic Criteria for Pyromania

Differential Diagnosis Before a diagnosis of pyromania can be made, a number of other conditions must be excluded. Fires set for monetary gain, for political statement, or for revenge do not qualify as pyromania. Patients with schizophrenia and mania can have psychotic experiences leading them to set fires. In certain cases, patients with dementia may set fires because of their inability to anticipate the consequences of an act. Children with conduct disorder may set fires as one of a number of inappropriate and antisocial activities. Individuals who fulfill criteria for antisocial personality disorder are excluded from the diagnosis of pyromania.

Treatment There is little to suggest a predictably effective treatment for pyromania. Psychodynamic psychotherapy is limited by patients' denial, lack of insight, and the frequent coexistence of alcohol abuse. Behavioral techniques include aversive therapy and positive reinforcement, though the usefulness of this approach is questionable. Conclusive information regarding the efficacy of pharmacotherapy is also lacking. Until substantive studies suggest the benefits of a single modality of treatment, an appropriate approach would be to invoke a number of associated treatments, including behavioral approaches and, in the case of children, adolescents, or young adults, family therapy.

INTERMITTENT EXPLOSIVE DISORDER

Definition The notion that explosive violence may be linked to a discrete diagnosable condition is controversial. In DSM-IV intermittent explosive disorder is characterized by aggressive impulses out of proportion to any precipitating psychosocial stressor. In the intervals between episodes there is no sign of impulsiveness or aggressiveness.

Comparative Nosology Intermittent explosive disorder was not recognized as a mental disorder in DSM-I. The closest diagnoses at the time were emotionally unstable personality and passive-aggressive personality. DSM-II described explosive personality, specifying character pathology associated with outbursts of aggression, perhaps linked to ictal episodes. The first formal recognition of intermittent explosive disorder came with DSM-III, when, with isolated explosive disorder, it was one of two recognized aggressive disorders of impulse control. DSM-III-R deleted isolated explosive disorder but retained intermittent explosive disorder. The category of intermittent explosive disorder is retained in DSM-IV with criteria that exclude the diagnosis if symptoms occur during the course of another mental disorder, as part of antisocial or borderline personality disorder, or as the direct physiological effect of substance use or a general medical condition, such as head trauma or Alzheimer's disease.

Intermittent explosive disorder is not recognized as a unique subcategory of habit and impulse disorders in ICD-10. It is mentioned as an example of other habit and impulse disorders with no diagnostic criteria provided specifically for intermittent explosive disorder (Table 22-1).

The existence of intermittent explosive disorder as a unique entity remains controversial. Many have difficulty with the idea of a normal baseline with superimposed periods of aggressive episodes. In addition, anger outbursts are a part of many other disease entities.

Epidemiology Intermittent explosive disorder is thought to be rare and occurs more frequently in males. High rates of fire-setting behavior in persons with the disorder have been reported. Recent studies suggest higher than normal rates of intermittent explosive disorder in families of patients with the diagnosis. First-degree relatives of patients with the disorder appear to have higher than expected rates of depressive disorders and alcohol and substance abuse.

Etiology Hypotheses about the cause of impulsive aggression derive from both psychological and biological perspectives. Psychoanalytic reports suggest that such outbursts occur in response to narcissistic injurious events. Rage outbursts serve as defenses that regulate interpersonal distance and protect against further narcissistic wounding. Typical patients appear to be large men with dependent personality features who respond to feelings of uselessness or impotence with violent outbursts.

Researchers supporting biological causes of aggression cite a number of important facts. In most case series of individuals with aggression, an extremely high number appear to display organic causes, a medical condition resulting in rage outbursts or neurological soft signs of neuropathology. One small case study of patients with episodic rage showed that they had lower levels of platelet serotonin reuptake than control subjects. Another sample demonstrated an association between a low CSF 5-HIAA concentration and impulsivity, whereas a high CSF testotesterone concentration correlated with aggressiveness and interpersonal violence. Restoring serotonergic activity by administration of the serotonin precursor L-tryptophan or drugs that increase synaptic serotonergic levels appears to restore control of episodic violent tendencies. Further, biological relatives of patients meeting criteria for categories A and B of DSM-IV criteria for this disorder were found more likely to have histories of temper outbursts than were adopted relatives of these patients.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for intermittent explosive disorder are presented in Table 22-6. Intermittent explosive disorder should not be diagnosed on the basis of one discrete episode of violence. A complete developmental history is needed to make the diagnosis, since the condition is characterized by recurrent episodic aggressive outbursts. Patients with this disorder also characteristically have a developmental history that includes alcohol dependence, violence, and emotional instability. Unstable interpersonal relationships, repeated job losses, and illegal behavior are also typical.

A. Several discrete episodes of failure to resist aggressive impulses that result in serious assaultive acts or destruction of property
B. The degree of aggressiveness expressed during the episodes is grossly out of proportion to any precipitating psychosocial stressors.
C. The aggressive episodes are not better accounted for by another mental disorder (e.g., antisocial personality disorder, borderline personality disorder, a psychotic disorder, a manic episode, conduct disorder, or attention-deficit/hyperactivity disorder) and are not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., head trauma, Alzheimer's disease).

Table 22-6 DSM-IV Diagnostic Criteria for Intermittent Explosive Disorder

Intermittent explosive disorder is mainly a diagnosis of exclusion. The clinician must first rule out medical conditions that can account for recurrent aggression. Assessment includes a comprehensive neurological examination and blood chemistry (with fasting blood glucose, liver function tests, electrolytes, and thyroid function tests), syphilis serology, urinalysis, and a urine toxicology screen. Further evaluation would include neuropsychological assessment, a computed tomographic (CT) scan of the head, and an electroencephalogram. If an organic condition is detected, a diagnosis of personality change due to a general medical condition, disinhibited type, should be made. If the diagnosis of a medical condition is equivocal, then other psychiatric disorders that might account for the rage episodes must be ruled out

(e.g., borderline and antisocial personality disorders or mood disorders). If the condition is not clearly due to organic causes or other psychiatric conditions, the diagnosis of intermittent explosive disorder is made.

Differential Diagnosis Impulsive acts of aggression that happen in the context of an organic condition must be ruled out. These conditions are better understood as a personality change due to a general medical condition, aggressive or disinhibited type. Impulsive aggression that occurs during a delirium or dementia also rules out intermittent explosive disorder as a diagnosis. Other psychiatric conditions must also be ruled out. Individuals with borderline personality disorder, antisocial personality disorder, and narcissistic personality disorder can certainly display explosive aggression. The difference in general is that the aggression seen in those with intermittent explosive disorder is more impulsive and continuous. Irritability can be seen in individuals with major depressive illness or in hypomanic or manic states. However, patients with intermittent explosive disorder do not display other components of a mood disorder. Although individuals with posttraumatic stress disorder can also display highly charged anger and irritability as a component of a hyperaroused state, other symptoms including the reexperiencing of a traumatic event and avoidance of reminders of this event are also present. Psychotic individuals can also display impulsive aggression. Careful assessment often reveals such features as command hallucinations and delusions of persecution. Finally, aggressive outbursts secondary to substance intoxication or substance withdrawal rule out a primary diagnosis of intermittent explosive disorder.

Course and Prognosis The course of intermittent explosive disorder is episodic and chronic. The unpredictable aggressive outbursts result in impaired interpersonal relationships and social isolation.

Treatment Treatment methods include both pharmacological and psychosocial approaches. Anticonvulsants and b-adrenergic receptor antagonists have been used for episodic aggressive outbursts. Some reports suggest that carbamazapine (Tegretol) may be preferable for intermittent explosive disorder and b-adrenergic receptor antagonists for temper proneness of organic etiology. Antianxiety agents have been used to decrease the anxiety that may lead to temper outbursts, though at times these agents in fact disinhibit certain patients, making control of their aggression more difficult. Multiple studies have evaluated the use of SSRIs to control anger across multiple diagnoses. Patients with a personality disorder and impulsive aggression appear to show significantly reduced anger and irritability after treatment with these agents. Additional studies need to assess the efficacy of these drugs for individuals specifically diagnosed with intermittent explosive disorder. An important adjunct to treatment with pharmacological agents is an ongoing and supportive therapeutic alliance between clinician and patient. Such a relationship increases the likelihood that a patient will immediately seek help in time of stress and thus avoid a dangerous outburst.

A 36-year-old real estate agent sought assistance for difficulty with his anger. He was quite competent at his job, though he frequently lost clients when he became enraged over their indecisiveness. On a number of occasions he became verbally abusive, leading clients to find ways out of escrow closings. The impulsive aggression also led to termination of multiple relationships because sudden angry outbursts contained demeaning accusations toward his girlfriends. This occurred frequently in the absence of any clear conflict. On multiple occasions, the patient became so uncontrollably enraged that he threw things across the room including books, his desk, and the contents of the refrigerator. Between episodes he was a kind and likable individual with many friends. He enjoyed drinking on the weekends and had a history of two arrests for driving while intoxicated. On one of these occasions he became involved in a verbal altercation with a police officer. He had a history of drug experimentation in college that included cocaine and marijuana.

Mental status examination revealed a generally cooperative patient. However, he became quite defensive when questioned about his anger and easily felt accused and blamed by the interviewer for his past behaviors. He had no significant medical history and no signs of neurological problems. He had never been in psychiatric treatment prior to this evaluation. He was on no medications. He denied any symptoms of a mood disorder or any other antisocial activity.

Treatment included the use of carbamazepine and a combination of supportive and cognitive-behavioral psychotherapy. The patient's angry outbursts improved as he became aware of early signs that he was about to lose control. He learned techniques to avoid confrontation when he was faced with these warning signs.

IMPULSE-CONTROL DISORDERS NOT OTHERWISE SPECIFIED

Impulse-control disorder not otherwise specified is a diagnosis reserved for disorders that involve the inability to resist an impulse but are not described by any of the other five categories (Table 22-7). Examples include compulsive sexual behavior, compulsive face picking, and self-mutilation. Recently it was suggested that compulsive buying also falls in this category.

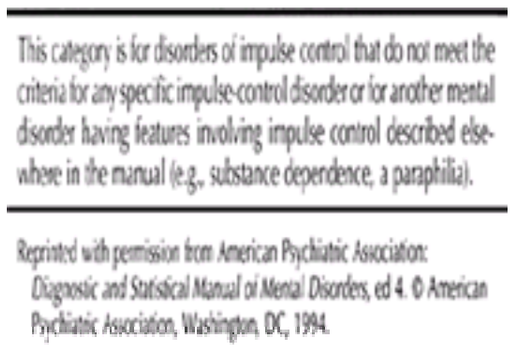

This category is for disorders of impulse control that do not meet the criteria for any specific impulse-control disorder or for another mental disorder having features involving impulse control described elsewhere in the manual (e.g., substance dependence, a paraphilia).

Table 22-7 DSM-IV Diagnostic Criteria for Impulse-Control Disorder Not Otherwise Specified

ICD-10 provides two categories similar to the category of impulse-control disorder not otherwise specified in DSM-IV (Table 22-1). The category of other habit and impulse disorders includes kinds of “persistently repeated maladaptive behavior that are not secondary to a recognized psychiatric syndrome and in which it appears that the patient is repeatedly failing to resist impulses to carry out the behavior. There is a prodromal period of tension with a feeling of release at the time of the act.” Intermittent explosive disorder is included as one such potential entity. ICD-10 also provides a category of habit and impulse disorder, unspecified, with no description of criteria for this category.

Compulsive Buying

Definition and History Although little attention has been paid until recently to compulsive buying, the entity was recognized by both Emil Kraeplin and Eugen Bleuler. It was originally referred to as oniomania and categorized as one of the “reactive impulses” or “impulsive insanities.” Although not recognized by DSM-IV or ICD-10 as a unique subcategory of the impulse-control disorders, some attempt has been made to develop a formal definition and diagnostic criteria of compulsive buying for both research and clinical purposes, based on the phenomenology of cases in the literature to date.

Epidemiology Current estimates of the prevalence of compulsive buying range from 1.1 to 5.9 percent of the general population. General population surveys of people meeting criteria for compulsive buying have shown that 80 to 92 percent are women. The onset of the disorder appears to be approximately 18 years of age, though frequently a decade passes before the buying pattern is recognized as a problem.

Etiology The cause of compulsive buying is not known. The psychoanalytic literature has suggested that a variety of central issues are responsible for this behavior. These patients have diminished self-esteem, and buying temporarily allows them to feel better about themselves. This helps explain why the most frequently purchased items include clothing and jewelry, items that often attract recognition from the external world. Others consider castration anxiety etiological, stating that women attempt to resolve the loss of the penis through the purchase of desired items. Another theory suggests that the act of purchasing makes a statement that there is a future, thus transiently reducing a central death anxiety. Other theorists suggest that compulsive buying is a form of psychological dependence like that with drugs and alcohol and that it shares qualities addictions such as tolerance, in which the shopper must make larger and larger purchases to feel relief.

Biological theorists point to similarities between compulsive buying and mood disorders, obsessive-compulsive disorder, and the other impulse-control disorders.

Compulsive buying shares symptoms with these entities, and some comorbidity seems to exist with these disorders, particularly mood disorders.

Diagnosis and Clinical Features Compulsive buyers frequently describe feelings of tension, power, or excitement before and while shopping, with relief or pleasure immediately following the experience. Nevertheless, at times distant from shopping, the experience is ego-dystonic; compulsive buyers appreciate the negative impact that the behavior has on their lives. Buying urges are episodic and tend to last about an hour. Urges may be as frequent as every hour but can occur as infrequently as once a month. Urges most commonly arise at home but can arise anywhere and anytime throughout the day. Most people at some point attempt to resist the urges but are often unsuccessful. Compulsive shoppers generally buy for themselves, although sometimes they shop for others. They tend to purchase items that they do not need and often give the purchases away as gifts. They generally shop in stores but also use catalogs and the home shopping network. Most purchases are made on credit, and these individuals tend to have numerous credit cards.

Buying urges occur throughout the year and are not centered specifically around holidays and birthdays. Buyers frequently shop alone and tend to buy a large number of inexpensive things rather than a few expensive items. Frequently purchased items are those that are worn (e.g., clothing and perfume), though men tend to focus somewhat more on electronic equipment, automobile equipment, and hardware. Proposed diagnostic criteria for compulsive buying are listed in Table 22-8.

A. Maladaptive preoccupation with buying or shopping, or maladaptive buying or shopping impulses or behavior, as indicated by at least one of the following:
   1. Frequent preoccupation with buying or impulses to buy that are experienced as irresistible, intrusive, and/or senseless.
   2. Frequent buying of more than can be afforded, frequent buying of items that are not needed, or shopping for longer periods of time than intended.
B. The buying preoccupations, impulses, or behaviors cause marked distress, are time consuming, significantly interfere with social or occupational functioning, or result in financial problems (e.g., indebtedness or bankruptcy).
C. The excessive buying or shopping behavior does not occur exclusively during periods of hypomania or mania.

Reprinted with permission from McElroy SL, Keck PE Jr, Pope HG Jr, Smith JM, Strakowski SM: Compulsive buying: A report of 20 cases. J Clin Psychiatry 55: 242, 1994.

Table 22-8 Diagnostic Criteria for Compulsive Buying

Differential Diagnosis Compulsive buying must be distinguished from spending that occurs exclusively during a hypomanic or manic period. Compulsive shoppers demonstrate a more enduring pattern of behavior that is not limited to episodes in which other symptoms of mania can be observed.

These individuals show a high comorbidity with other Axis I conditions. In one study of 20 patients, all met criteria for two or more Axis I conditions, and 13 met criteria for four or more conditions. All of these patients met criteria for some form of mood disorder, most commonly bipolar I or bipolar II disorder. Symptoms seemed to increase when individuals felt more dysphoric and to decrease when patients were hypomanic. Many of these individuals stated that buying relieved their depressive symptoms.

Other disorders comorbid with compulsive buying include anxiety disorders such as obsessive-compulsive disorder, panic disorder, and phobias. Substance abuse and dependence, eating disorders, and other impulse-control disorders are frequently seen in these patients.

Course and Prognosis Compulsive buying is a chronic condition that can have devastating financial, marital, and vocational consequences. Though individuals frequently attempt to stop the behavior on their own, they are usually unsuccessful. Limiting access to shopping including credit cards, home catalogs, the Internet, and the home shopping network has met with some success for this disorder.

Most patients run into debt problems, with the average debt in one study being $23,000. Some are forced to declare bankruptcy. This financial burden results in discord within relationships and, not infrequently, divorce. Jobs are compromised because of the increasing amount of time devoted to compulsive shopping.

Treatment Information regarding treatment is based on case reports; no formal, rigorously controlled, treatment studies exist for compulsive shopping. Treatment of compulsive buying has included both psychological and pharmacological modalities, though information is based upon case reports rather than formal clinical trials. Some patients receiving supportive or insight-oriented therapy report gaining some control over their buying compulsions. Others have been helped by supportive self-help groups like Debtors Anonymous. Pharmacological data are limited, with mixed results. In one series, 9 of 20 patients receiving pharmacological interventions had complete or partial remission of symptoms. Medications that were at least partially effective included anitidepressants, mood-stabilizing agents, anxiolytics, and antipsychotics used singly or in conjunction with other agents. Some individuals who did not improve had discontinued pharmacotherapy early in treatment because of adverse effects or induction of hypomania.

A 46-year-old married woman presented to a psychiatrist for help with grieving the loss of her father. She met with the psychiatrist twice before disclosing that she was experiencing great discord in her marriage. This centered primarily on a financial debt that was now building as a result of her continued purchases. Her husband had been a highly successful attorney. However, he had had a stroke and could no longer continue his practice, which greatly limited the family income. Although the patient's spending had not been a problem in the past, she could no longer control it despite increasingly limited funds. She continued to spend over $10,000 a month on credit card purchases, generally items of clothing and jewelry for herself. She was extremely attracted to the lights and smells of a particular clothing designer, stating that she felt “forever young” when she shopped there.

Despite numerous attempts to control her spending, she was unable to do so. When she limited her visits to shopping malls, she turned to the home shopping network and continued to build a large debt. Despite the painful knowledge that her family could no longer afford this and that her action might threaten her son's private school tuition, she was unable to resist the impulse to shop. She described feelings of tension before shopping and power while she was shopping. Shortly afterward, however, she felt quite remorseful with the realization that she could not control herself and that this behavior was becoming a great problem for her family.

Treatment consisted of a combination of insight-oriented psychotherapy and psychopharmacology. Dynamic therapy focused on long-term issues of vulnerability that were stimulated by the loss of her father and the illness of her husband. She was placed on an SSRI to treat her continued dysphoric feelings. Over the next 2 years, she was able to curtail her uncontrolled shopping greatly. Ultimately the family's financial crisis was resolved.

SUGGESTED CROSS-REFERENCES

Obsessive-compulsive disorder, discussed in Chapter 15, has both significant similarities to and differences from the various impulse-control disorders. Chapter 14 on mood disorders and Chapter 20 on eating disorders contain details on features common to a number of possibly related syndromes, including impulse-control disorders not elsewhere specified. Section 11.1, describing substance dependence and abuse, and Section 11.2 on alcohol-related disorders describe features such as dependence and withdrawal that parallel those of impulse-control disorders (especially pathological gambling). Information in Chapter 10 on personality change due to a medical condition is important in distinguishing organic from presumably functional causes of explosive episodes. Compulsive sexual behavior and sexual addiction are discussed in Section 19.1a.

CHAPTER REFERENCES

*Black DW: Compulsive buying: A review. J Clin Psychiatry *57*(Suppl):50, 1996.

*Black DW, Monahan P, Gabel J: Fluvoxamine in the treatment of compulsive buying. J Clin Psychiatry *58:*159, 1997.

*Campbell F, Lester D: The impact of gambling opportunities on compulsive gambling. J Soc Psychol *139:*126, 1999.

*Chong SA, Low BL: Treatment of kleptomania with fluvoxamine. Acta Psychiatr Scand *93:*314, 1996.

*Christenson GA, Crow SJ: The characterization and treatment of trichotillomania. J Clin Psychiatry *57*(Suppl):42, 1996.

*Christenson GA, Faber RJ, de Zwaan M: Compulsive buying: Descriptive characteristics and psychiatric comorbidity. J Clin Psychiatry *55:*5, 1994.

*Christenson GA, Ghernoff-Clementz E, Clementz BA: Personality and clinical characteristics in patients with trichotillomania. J Clin Psychiatry *53:*407, 1992.

*Christenson GA, Pyle RL, Mitchell JE: Estimated lifetime prevalence of trichotillomania in college students. J Clin Psychiatry *52:*415, 1991.

*Cohen LJ, Stein DJ, Simeon D, Spadaccini E, Rosen J, Aronowitz B, Hollander E: Clinical profile, comorbidity, and treatment history in 123 hair pullers: A survey study. J Clin Psychiatry *56:*319, 1995.

*Daghestani AN, Elenz E, Crayton J: Pathological gambling in hospitalized substance abusing veterans. J Clin Psychiatry *57:*360, 1996.

*DeCaria CM, Hollander E, Grossman R, Wong CM, Mosovich SA, Cherkasky S: Diagnosis, neurobiology, and treatment of pathological gambling. J Clin Psychiatry *57*(Suppl):80, 1996.

*Goldman MJ: Kleptomania: Making sense of the nonsensical. Am J Psychiatry. *148:*986, 1991.

*Hollander E, Kwon JH, Stein DJ, Broatch J, Rowland CT, Himelein CA: Obsessive-compulsive and spectrum disorders: Overview and quality of life issues. J Clin Psychiatry *57*(Suppl):3, 1996.

*Hudson JI, Pope HG: Affective spectrum disorder: Does antidepressant response identify a family of disorders with a common pathophysiology? Am J Psychiatry *147:*552, 1990.

*Kavoussi R, Armstead P, Coccaroa E: The neurobiology of impulsive aggression. Psychiatr Clin N Am *20:*395, 1997.

*Kavoussi R, Liu J, Coccaro E: An open trial of sertraline in personality disordered patients with impulsive aggression. J Clin Psychiatry *55:*137, 1994.

*Keuthen NJ, O'Sullivan RL, Goodchild P, Rodriguez D, Jenike MA, Baer L: Retrospective review of treatment outcome for 63 patients with trichotillomania. Am J Psychiatry *155:*560, 1998.

*Lawrence L: The psychodynamics of the female shopper. Am J Psychoanal *50:*67, 1990.

*Lejoyeux M, Tassin V, Solomon J, Ades J: Study of compulsive buying in depressed patients. J Clin Psychiatry *58:*169, 1997.

*Leong GB: A psychiatric study of persons charged with arson. J Forensic Sci *37:*1319, 1995.

*McElroy SL, Hudson JI, Pope HG, Keck PE, Aizley HG: The DSM-III-R impulse control disorders not elsewhere classified: Clinical characteristics and relationship to other psychiatric disorders. Am J Psychiatry *149:*318, 1992.

*McElroy SL, Keck PE, Phillips KA: Kleptomania, compulsive buying, and binge eating disorder. J Clin Psychiatry *56*(Suppl):14, 1995.

*McElroy SL, Keck PE Jr, Pope HG Jr, Smith JM, Strakowski SM: Compulsive buying: A report of 20 cases. J Clin Psychiatry *55:*242, 1994.

*McElroy SL, Pope HG Jr, Keck PE Jr, Hudson JI, Philips KA, Strakowski SM: Are impulse-control disorders related to bipolar disorder? Compr Psychiatry *37:*229, 1996.

*O'Sullivan R, Miguel EC, Coffey B, Rauch SL, Savage C, Keuthen NJ, Baer L, Jenike MA: Trichotillomania, obsessive-compulsive disorder, and Tourette syndrome: Comorbid relationships and risks for expression. CNS Spectrums *3:*49, 1998.

*Rapoport JL, Ryland DH, Kriete M: Drug treatment of canine acral lick: An animal model of obsessive-compulsive disorder. Arch Gen Psychiatry *49:*517, 1992.

*Rosenthal R: Pathological gambling. Psychiatr Ann *22:*72, 1992.

*Roy A, DeJong J, Ferraro T, Adinoff B, Gold P, Rubinow D, Linnoila M: CSF, GABA and neuropeptides in pathological gamblers and normal controls. Psychiatr Rev *30:*137, 1989.

*Roy A, DeJong J, Linnoila M: Extroversion in pathological gamblers: Correlates with indexes of noradrenergic function. Arch Gen Psychiatry *46:*679, 1989.

*Schlosser S, Black DW, Repertinger S, Freet D: Compulsive buying: Demography, phenomenology, and comorbidity in 46 subjects. Gen Hosp Psychiatry *16:*205, 1994.

*Simeon D, Stein DJ, Gross S, Islam N, Schmeidler JS, Hollander E: A double-blind trial of fluoxetine in pathologic skin picking. J Clin Psychiatry *58:*341, 1997.

*Soriano JL, O'Sulliven RL, Baer L, Phillips KA, McNally RJ, Jenike MA: Trichotillomania and self esteem: A survey of 62 female hair pullers. J Clin Psychiatry *57:*77, 1996.

*Stein DJ, Hollander E: Low dose pimozide augmentation of serotonin reuptake blockers in the treatment of trichotillomania. J Clin Psychiatry *53:*123, 1992.

*Stein DJ, Simeon D, Cohen LJ, Hollander E: Trichotillomania and obsessive-compulsive disorder. J Clin Psychiatry *56*(Suppl):28, 1995.

*Virkkunen M, Rawlings R, Tokola R, Poland R, Guidotti A, Nemeroff C, Bissette G, Kalogeras K, Karonen SL, Linnoila M: CSF biochemistries, glucose metabolism, and diurnal activity rhythms in alcoholic, violent offenders, fire setters, and healthy volunteers. Arch Gen Psychiatry *51:*20, 1994.

*Vitaro F, Arseneault L, Tremblay RE: Dispositional predictors of problem gambling in male adolescents. Am J Psychiatry *154:*1769, 1997.

*Volberg RA, Abbott MW: Lifetime prevalence estimates of pathological gambling in New Zealand. Int J Epidemiol *23:*976, 1994.

*Winchell RM: Trichotillomania: Presentation and treatment. Psychiatr Ann *22:*84, 1992.

CHAPTER 23. ADJUSTMENT DISORDERS

JEFFREY H. NEWCORN, M.D., JAMES J. STRAIN, M.D., AND JUAN E. MEZZICH, M.D., IH.D.

The adjustment disorders are diagnoses that are seldom the subject of research, but are nonetheless widely used in clinical practice. The appeal of this diagnostic category may be attributable to several factors: (1) In a diagnostic system that is principally atheoretical, adjustment disorders remain one of the few conditions that are linked to an etiological event; (2) the notion that adjustment problems follow from stressful events has been a mainstay of psychodynamic thinking, and often forms the basis of psychotherapeutic treatment; and (3) because the development of transient psychiatric symptoms in the context of stress is virtually a universal experience, an adjustment disorders is considered by many to be a nonstigmatizing diagnosis to assign when making a patient's psychiatric status public.

Despite these considerable strengths, adjustment disorders is one of the most problematic diagnostic categories in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). The fact that the relationship between stress and psychiatric disorder is both complex and uncertain has caused many to question the theoretical basis of adjustment disorders. In addition, the absence of operationalized, symptom-based criteria and a threshold level of symptomatology required for diagnosis has resulted in the use of this category for patients who might otherwise fulfill criteria for another, more specific mental disorder. Since incorrect diagnosis may result in inadequate treatment, inappropriately conceptualizing a patient's problem as constituting an adjustment disorder may result in delays or inaccuracies in treatment planning. Misdiagnosis of other, more specific disorders when an adjustment disorder should appropriately be diagnosed is also a problem. For example, in one study, medical residents frequently diagnosed major depressive disorder when adjustment disorder with depressed mood was considered the correct diagnosis by an attending psychiatrist.

Problems with the use of the adjustment disorders category among child and adolescent psychiatrists were highlighted in a survey that was conducted as a means of informing the DSM-IV work groups about the use of psychiatric disorders in clinical practice. Of those who responded to the survey, 55 percent indicated that they used adjustment disorders to avoid stigmatization of patients. Many of those who favored the use of this category were not formally trained in the revised third edition of DSM (DSM-III-R) and were not inclined to use not otherwise specified categories. Over half of these psychiatrists did not consider the temporal onset criterion for adjustment disorders or the relevant exclusionary criteria in applying this diagnosis.

The survey results indicate that the adjustment disorders diagnoses may be used excessively and incorrectly by some clinicians. Nevertheless, there are individual patients for whom use of this category represents the correct formulation. Examples include high-functioning adults who experience mood or anxiety symptoms in response to change of occupational status or loss of a loved one, adolescents who become distraught in the face of parental separation or academic failure, and individuals who develop transient psychiatric symptoms in the context of a newly diagnosed or ongoing medical condition.

DEFINITION

Adjustment disorders are characterized in DSM-IV by the development of emotional or behavioral symptoms in the context of one or more identified psychosocial stressors. The resultant symptomatology is deemed to be clinically significant by virtue of either impairment in social, occupational, or educational function, or the subjective experience of distress in excess of what would normally be expected for the given stressors. The nature and severity of the stressors are not specified. However, the stressors are more often everyday events that are ubiquitous (e.g., loss of a loved one, change of employment or financial situation) rather than rare, catastrophic events (e.g., natural disasters, violent crimes). The symptomatology must, by definition, occur within 3 months of the occurrence of the stressor, and must remit within 6 months following the cessation of the stressor. Finally, the disturbance must not fulfill the criteria for another major psychiatric disorder or bereavement (not considered a mental disorder, although it may be a focus of clinical attention). A variety of symptomatic presentation subtypes of adjustment disorders are identified. The scope of symptomatology covered by these subtypes indicates that virtually any subthreshold condition deemed to be associated with a psychosocial stressor could potentially meet the criteria for adjustment disorders.

HISTORY

The notion that psychiatric disorders frequently occur in the context of stressful life events has a time-honored tradition. In the second edition of DSM (DSM-II) the category transient situational disorder was used to describe the phenomenon now labeled adjustment disorders. The disorder was understood in a developmental context, and the subtypes were defined accordingly. Although the condition was described as transient and situational, the clinical presentation was linked to severe stress. Dissatisfaction with this definition followed from the observations that the disorders were not always transient and did not necessarily require an unusually severe precipitating event.

The diagnosis of transient situational disorder was reworked considerably in the third edition of DSM (DSM-III) and was renamed adjustment disorders. Severe and unusual stress was no longer required, and the subtypes were redefined according to their symptomatic presentation. The duration of adjustment disorders was not specified; however, the symptoms were presumed to remit with the cessation of the stressor. There were no major theoretical changes from the third edition of DSM (DSM-III) to DSM-III-R. However, the duration of the disorder was restricted to 6 months, with the expectation of return to baseline function or establishment of a new steady state characterized by a more specific diagnosis; in addition, the subtype adjustment disorders with physical complaints was added.

The most important change from DSM-III-R to DSM-IV was that the duration of the disorder was allowed to extend beyond the 6-month limit imposed in DSM-III-R. The current criterion states that the symptoms may not persist for greater than 6 months following the cessation of the stressor. The duration of the disorder is coded as *acute* (less than 6 months) or *chronic* (greater than 6 months). This change better accounts for the persistence of adjustment disorders symptoms in the context of chronic stress, while still maintaining some limitations to prevent overuse once the stressor remits or a new steady state has been reestablished. Other changes involved rewording of criteria to reflect current thinking regarding this condition. The symptoms are no longer described as a reaction to a stressor (implying causality), nor are they described as maladaptive, but rather as clinically significant. Finally, the exclusionary criterion for other mental disorders is more clearly articulated to permit the diagnosis of comorbid conditions, provided that the stress-related symptoms are not better accounted for by the coexisting disorder.

COMPARATIVE NOSOLOGY

DSM Criticism of the adjustment disorders diagnosis has focused on uncertainty regarding the nature and meaning of the predisposing psychosocial stressors, the extent to which inference and value judgement may be required in order to make the diagnosis, and problems in reliability and validity. Equally problematic but not as clearly articulated in the published literature are difficulties in distinguishing adjustment disorders from subthreshold pathology that is not stress-related (e.g., not otherwise specified categories) and other posttraumatic disorders. Both these latter categories were expanded in DSM-IV, making their distinction from adjustment disorders more challenging.

Stressor Criterion The linear model of stress-disease interaction, which has served as the model for adjustment disorders in DSM, has been criticized by several authors. This model presupposes that a direct and clearly identifiable pathological reaction follows a stressful event. However, there are many possible ways in which stress may be related to psychiatric illness which are not adequately taken into account by this model. For example, there may be multiple stressors or insidious or chronic circumstances, as opposed to discrete events. Recent studies suggest that pathology is more often associated with an accumulation of stressors over time, rather than the occurrence of any one stressor. Stressors that result in dependent function are particularly noxious. Classification of individuals with chronic disturbance, chronic stress, or both may be difficult using the DSM model of adjustment disorders. The development of psychopathology in individuals exposed to chronic stress is not always direct, and relatively minor precipitating events may generate symptomatology in individuals who have previously been sensitized to

stress.

The stressor criterion in adjustment disorders has also been criticized for its lack of specificity. Several authors note that there is no mechanism to measure the stressors in adjustment disorders and, as a result, their clinical implications are often uncertain. These authors question whether patients with adjustment disorders are unusually sensitive to psychosocial events that are not likely to cause disturbance in others, or whether these individuals have been exposed to high levels of stress, the severity or accumulation of which would likely produce negative consequences in most people. The current DSM definition clearly favors the former explanation because unusually high levels of stress are more characteristic of posttraumatic stress disorder and acute stress disorders than adjustment disorders. Another criticism of the stressor criterion in adjustment disorders is that it is not specific enough to adequately facilitate discrimination of adjustment disorders from other subthreshold disorders that are not linked to stress. The failure to define what represents a normal and expectable reaction to stress has also been a source of criticism. Although cultural factors often play an important role in determining what types of events are most stressful and define how an individual is likely to respond to stress, they are not factored into the diagnostic algorithm.

Implicit in many of these criticisms is the importance of individual differences relative to the development of psychopathology in the context of stressful events, and the failure of current classification schemes to adequately account for these factors. In addition to individual or group sensitivity to stress, factors that buffer the individual from stress or those that mitigate the development of psychopathology need to be included in the assessment matrix. Recent publications that have focused on protective features relative to the development of psychopathology in children and adolescents demonstrate the importance of a variety of individual (e.g., positive temperament, absence of preexisting mood or conduct problems) and family (e.g., warm and nurturant relationship with primary caregivers) factors in mitigating pathological response to stress. How to create an interactive model that takes into account both stress and resiliency factors, which is responsive to individual and cultural differences, remains a challenge for future editions of DSM.

Reliability The few extant studies of reliability in adjustment disorders have produced unimpressive results. One study determined the inter-rater agreement (k) for adjustment disorders to be 0.05 (*P* = not significant) in a survey of psychiatrists and psychologists using 27 case histories of child and adolescent cases. The k for the DSM-II category transient situational disorder was somewhat higher in this study (k = 0.28; $P < 0.05$). The results of the U.K.–World Health Organization (UK-WHO) study of reliability of the ninth revision of *International Statistical Classification of Disease* (ICD-9) categories in children and adolescents were consistent with these findings. The k for adjustment disorders was 0.23, which was considerably lower than for many other categories. Reclassification using a glossary improved reliability to 0.33, suggesting that structured assessment can partially ameliorate the limited reliability of adjustment disorders.

There are many potential sources of poor reliability, although the lack of an operationalized symptom checklist and a threshold level of symptoms that demarcate entrance into the diagnosis seem paramount. Other sources of poor reliability include (1) difficulties in determining when subjective distress or observable symptomatology exceeds what would normally be expected for a given stressor in an “average” individual; (2) the absence of any impairment criterion in the diagnostic algorithm that defines maladaptation to stress; and (3) the variability produced by cultural expectations regarding the reaction to and management of stressful events. The low reliability of adjustment disorders is consistent with the fact that measurement of psychosocial stress on Axis IV has repeatedly been found to be questionable. Nevertheless, recent efforts to create a more rigorous adjustment disorders diagnosis offer promise for improved reliability in the future. One such study utilized an extensive symptom checklist and required a minimum of four symptoms for diagnosis, and was able to accurately classify the majority of cases of adjustment disorder.

Temporal Criteria Two studies have examined the onset of adjustment disorders following stress. One was conducted in 92 school-age children with a recent diagnosis of new-onset insulin-dependent diabetes; 33 of these children developed adjustment disorders. The highest risk occurred in the first month after diagnosis (73 percent), and 31 of the 33 had onset of symptoms by 3 months. Another study, conducted in adults who underwent cardiac surgery, observed that over half developed psychiatric symptoms before surgery, probably as a consequence of the impairment resulting from the illness or the anticipation of surgery. The other half experienced symptom formation 1 month afterwards, more likely as a consequence of the recuperation from surgery or the resultant change in lifestyle. Both studies demonstrate the close temporal association of psychiatric symptoms with the stressor, and thereby support the DSM onset criterion of 3 months. However, the latter study also illustrates that it may be difficult to ascertain what the actual stressor is in any given situation, thereby complicating the determination of onset of symptoms.

Several published studies have provided data regarding the duration of adjustment disorders. A chart review study at the University of Iowa reported that 66 percent of adolescents and 35 percent of adults with adjustment disorders were symptomatic for greater than 6 months, while 47 percent of adolescents and 23 percent of adults were ill for over 1 year. Another study, which examined the correlates of depressive disorders in children, including adjustment disorders with depressed mood, found that the mean duration of an episode was 25 weeks. The peak interval-specific probability of recovery was 6 to 9 months. Similar findings were reported in the study that examined the psychological adjustment of children following the onset of insulin-dependent diabetes. The mean episode length was 3 months, with the majority recovering within 6 months. However, 15 percent of the sample was symptomatic for greater than 6 months. In the study of adjustment disorders following cardiac surgery, 30 percent of patients had not recovered within 6 months. These findings all support the decisions in DSM-IV to extend the duration of symptoms in adjustment disorder beyond 6 months and to establish acute and chronic designations.

Results of studies examining the onset of psychopathology following stress, not linked to any specific disorder, also provide useful data with which to evaluate the temporal criteria in adjustment disorders. One such study assessed the timing and number of life events occurring in the year prior to the onset of emotional or behavioral problems or both in school-aged children compared with community controls. The onset of symptoms occurred within 16 weeks of the event in only 49 percent of cases, with a subgroup of children developing symptoms more than 6 months following the event. This finding of a relatively long latency of onset of symptoms following stress stands in contrast to the results in samples with a circumscribed diagnosis of adjustment disorders. It may be that symptomatology in response to stress is quicker in adjustment disorders than other disorders, or it may be that latency is shorter following serious medical illness as opposed to other stressors.

ICD There is an analogous category of adjustment disorders in the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), which is placed in a cluster of stress-related disorders. Although the construct of “adjustment disorders” is quite similar in DSM-IV and ICD-10, certain defining characteristics differ. In ICD-10 the disorder has its onset within 1 month of the occurrence of the stressor, rather than the 3 months required in DSM-IV. Also, there are differences in the subtypes identified in the two systems. ICD-10 identifies a brief depressive reaction and a prolonged depressive reaction; the latter subtype may last up to 2 years. Similar to DSM-IV, ICD-10 identifies subtypes for mixed anxiety and depression, disturbance of conduct, mixed disturbance of emotions and conduct, and other specific predominant symptoms. However, ICD also highlights particular features of maladjustment to stress in children, best exemplified by regressive behaviors such as bedwetting, babyish speech, or thumb-sucking. These behaviors, when present, are coded under the subtype adjustment disorders with predominant disturbance of other emotions.

Unresolved Questions Several important questions regarding adjustment disorders remain unresolved and should receive attention in future classification efforts. What is the relationship between adjustment disorder and other stress-related disorders, such as posttraumatic stress disorder and acute stress disorder, and should these conditions be grouped together, as in ICD-10, or separately, as in DSM-IV? Are the subtypes of adjustment disorders more appropriately conceptualized as minor disorders of mood, anxiety, and behavior? As increased attention is focused on these various subthreshold disorders, and criteria for their use are developed, it may be more difficult to distinguish adjustment disorders from subthreshold categories that are not stress-related. Finally, should adjustment disorders be placed with the problem-level diagnoses rather than the psychiatric disorders? Against the latter option are the findings from a large outpatient sample that indicate that, although patients with adjustment disorders present with less severe symptomatology than those with specific psychiatric disorders, they more closely resemble the patients with specific disorders in terms of impairment and level of function than they do the individuals who do not qualify for a psychiatric diagnosis.

EPIDEMIOLOGY

There is only one published epidemiological study that included adjustment disorders, and that was conducted in children and adolescents in Puerto Rico. The prevalence rate of adjustment disorders was determined to be 7.6 percent using a cutoff of 70 for “caseness” on the Children's Global Assessment Scale (CGAS). However, if children with CGAS scores between 61 to 70 were excluded, the prevalence of adjustment disorders dropped to 4.2 percent. While this suggests a low level of impairment in about 40 percent of children with adjustment disorders, the 4.2 percent figure indicates that a large number of individuals have the disorder and that it is often associated with impairment.

Similar epidemiological data are not available in adults because the structured interviews used as a source of diagnosis in the major epidemiological studies in this age group did not include algorithms for adjustment disorders. However, reanalyses of data from the Epidemiologic Catchment Area study have identified a subthreshold mood disorder, subsyndromal symptomatic depression, whose prevalence exceeded that of all proposed DSM-III mood disorders combined. Approximately half of the individuals who met criteria for this disorder presented with depressed mood. Another large group presented with two or more related symptoms, such as vegetative symptoms, trouble concentrating, decreased energy or suicidal thoughts, for over 2 weeks, but did not report depressed mood. It is

likely that many of these individuals would have qualified for diagnoses of minor depressive disorder, another subthreshold diagnosis (e.g., anxiety disorder not otherwise specified) or adjustment disorders. However, it is not known what percentage of cases were related to psychosocial stressors.

The prevalence of adjustment disorders has more often been evaluated in clinical samples. The largest and most systematically assessed database was obtained by evaluating all patients on entry to the Western Psychiatric Institute clinical services using a semistructured assessment form and DSM-III criteria sheets. Ten percent of a sample of over 11,000 patients (all ages) were found to have adjustment disorders, making it the second largest diagnostic category. In children and adolescents under 18 years old, over 16 percent had adjustment disorders. In adults, females predominated over males by approximately 2 to 1. The sex ratio was more equal in children and adolescents, although there was still a slight excess of female cases.

Prevalence estimates of adjustment disorders in other clinical populations have been characterized by considerable variability, though reported rates are generally high. In a study of over 1300 psychiatric outpatients in Switzerland, 26 percent qualified for a diagnosis of adjustment disorders as determined by two evaluating psychiatrists. A survey of emergency room visits found that 13 percent of adults and 42 percent of adolescents were diagnosed with adjustment disorders. Adjustment disorders was twice as common as any other diagnosis among adolescents in that study, including substance abuse. Adjustment disorders is also highly prevalent in medical settings. In a multisite consortium of teaching hospitals, adjustment disorder was diagnosed in 125 out of over 1000 patients (12 percent). In another study, 9 percent of 1072 patients evaluated on an adult medical-surgical consultation-liaison service were diagnosed with this disorder. Approximately 50 percent of 71 patients who received cardiac surgery met criteria for adjustment disorders, almost all had adjustment disorders with depressed mood. Among pediatric patients sampled from four different clinics, 25 to 65 percent of those who presented with psychiatric disturbance were given a diagnosis of adjustment disorders. In evaluating these 11 prevalence figures, it is important to remember that structured methods are generally not used in clinical settings, and that clinician preferences regarding diagnosis or sample bias may have contributed to the high prevalence estimates of adjustment disorders.

ETIOLOGY

By definition, adjustment disorders is associated with psychosocial stress; it is presumed that without the occurrence of the stressor, the condition would not arise. However, many important questions regarding etiology remain unanswered. For example, why is there so much individual variation in the development of psychopathology in individuals who experience apparently similar life events? Are all individuals likely to develop symptoms if stress levels are increased sufficiently? When a pathological reaction to stress does occur, why is the nature of the pathology so variable—ranging from no disorder, to subthreshold pathology or adjustment disorders, to major disorders? Does an initial episode of adjustment disorders predict the subsequent development of more specific and severe pathology? What modifiers of stress attenuate or accentuate its effects?

Few studies have evaluated the types and duration of stressors in adjustment disorders. One such study highlighted differences between adolescents and adults in the nature and duration of stress. Stressors were more likely to be chronic (>1 year) in adolescents than adults (59 percent versus 36 percent). School problems were the most frequently identified precipitant for symptomatology among adolescents (>60 percent), although a variety of family, boyfriend-girlfriend, and substance use problems were also common. In adults, marital problems, including separation and divorce, were the most common stressors. A similar profile of stressors was found in a recent study of Swiss outpatients with adjustment disorders.

The results of studies that have examined the relationship of stress to the development of psychopathology provide additional information to be considered in evaluating the model of stress-disease interaction in adjustment disorders. Several studies have found that individuals with and without preexisting symptoms respond differently to the presence of stressful events. This has been observed both in samples with adjustment disorders as well as other conditions. Preexisting mood symptomatology was the only factor that predicted a prolonged course of adjustment disorders following cardiac surgery. In Israeli children who suffered the loss of their father in war, preexisting conduct symptoms were highly correlated with poor adjustment (although not necessarily adjustment disorders).

Other studies have focused on those factors that protect individuals from developing stress-related symptoms. For example, preexisting adaptive skills decrease the likelihood of symptoms developing in the face of stress. In child and adolescent populations, a warm and supportive relationship with the primary caregiver, easy and adaptable child temperament, and healthier adjustment of the family to stress all predict a more positive response to stress within the child. Finally, the nature of the life events has important implications for the level of adjustment. Control over life events has generally been associated with improved adjustment even though it may increase stress. Similarly, pleasurable events are generally linked to psychological well-being, even though they may be stressful. In contrast, life events of a more dependent nature, which are undesirable and outside the control of the affected individual, are more likely to produce psychiatric symptoms.

The relationship of family, environmental, and genetic factors to adverse life events must also be considered. Findings from a study of over 2000 twin pairs indicate that life events are modestly correlated in twin pairs, with monozygotes showing greater concordance than dizygotes. Family-environmental and genetic factors each accounted for approximately 20 percent of the variance in that study. Another twin study that examined genetic contributions to the development of symptoms of posttraumatic stress disorder (not necessarily at the level of full disorder, and therefore relevant to adjustment disorders) similarly concluded that the likelihood of developing symptoms in the context of traumatic life events is partially under genetic control. The findings of these studies suggest that the occurrence of adverse life events and their consequences are not necessarily random. Certain individuals appear to be at increased risk both for the occurrence of these events and for the development of pathology once they occur.

In summary, although the occurrence of psychosocial stress is an essential component of adjustment disorders, data indicate that stress is just one of many factors that determines the development, nature, and extent of psychopathology. Specific types of stressful events and individual patterns of stress response appear to be preferentially related to the development of psychiatric symptoms in vulnerable individuals. Protective factors play a role in mitigating the stress response. Moreover, development of psychiatric symptoms is most often associated with an accumulation of stressors rather than any single event. How to best integrate these findings into a more robust conceptual framework for adjustment disorders represents a challenge for the future.

DIAGNOSIS AND CLINICAL FEATURES

Table 23-1 lists the DSM-IV diagnostic criteria for adjustment disorders, and Table 23-2 presents the ICD-10 diagnostic criteria for adjustment disorders.

A. The development of emotional or behavioral symptoms in response to an identifiable stressor(s) occurring within 3 months of the onset of the stressor(s).
B. These symptoms or behaviors are clinically significant as evidenced by either of the following:
(1) marked distress that is in excess of what would be expected from exposure to the stressor
(2) significant impairment in social or occupational (academic) functioning
C. The stress-related disturbance does not meet the criteria for another specific Axis I disorder and is not merely an exacerbation of a preexisting Axis I or Axis II disorder.
D. The symptoms do not represent bereavement.
E. Once the stressor (or its consequences) has terminated, the symptoms do not persist for more than an additional 6 months.

Specify if:
Acute: if the disturbance lasts for less than 6 months
Chronic: if the disturbance lasts for 6 months or longer

Adjustment disorders are coded based on the subtype, which is selected according to the predominant symptoms. The specific stressor(s) can be specified on Axis IV

With depressed mood
With anxiety
With mixed anxiety and depressed mood
With disturbance of conduct
With mixed disturbance of emotions and conduct
Unspecified

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 23-1 DSM-IV Diagnostic Criteria for Adjustment Disorders

Table 23-2 ICD-10 Diagnostic Criteria for Adjustment Disorders

Several studies have described the nature of the symptomatic presentation of adjustment disorders in clinical samples. In most samples depressive symptoms have been found to predominate, leading some authors to assert that depressive symptoms represent a core feature of adjustment disorders. In the Western Psychiatric Institute study half of a mixed child and adult population with adjustment disorders had depressed mood. An earlier publication from the same data set reported that 76 percent of adjustment disorders patients had depressed mood. Additional symptoms prominently represented in the adjustment disorders group were insomnia (53 percent), other vegetative symptoms, social withdrawal (29 percent), and suicidal indicators (29 percent).

The development of depressive symptoms is characteristic of children and adolescents with adjustment disorders as it is in adults, although behavioral symptoms and mixed clinical presentations are also frequently encountered. These findings are consistent with data from studies of mood disorders in children and adolescents, which indicate that comorbidity with behavior disorders is common. In the University of Iowa study, a large percentage of both adults and adolescents with adjustment disorders had depressive symptoms (87 percent of adults and 63 percent of adolescents). However, 77 percent of adolescents but only 25 percent of adults had behavioral symptoms. The prominence of mixed syndromes in youth was highlighted in two other studies of children and adolescents with adjustment disorders, in which the majority of subjects were found to have mixed emotional or mixed emotional and behavioral features.

Several recent investigations have begun to delineate symptom profiles for adjustment disorders. Youth who developed adjustment disorders following the onset of diabetes mellitus presented with an average of five symptoms, including feeling sad, suicidal ideation, pessimism, anhedonia, irritability or anger, and fatigue. The most common symptoms of adjustment disorders in the Western Psychiatric Institute data set reanalyzed for DSM-IV included depressed mood, low self-esteem, suicidal behavior, increased motor activity, impulsive behavior, substance use, hypervigilance, and hostility. Many of these symptoms have also been described in subthreshold mood and anxiety disorders that are not stress related. One proposed category of subsyndromal symptomatic depression included symptoms of trouble falling asleep or early waking, fatigue, thoughts of death, sadness, and trouble concentrating. Similarly, the field trial for mixed anxiety-depression conducted for DSM-IV indicated that difficulty concentrating, hypervigilance, hopelessness, feelings of worthlessness, tearfulness, fatigue, worry, anticipating the worst, irritability, and sleep disturbance are frequently present. The collective findings from these studies indicate that it is possible to identify core symptoms for subthreshold mood and anxiety disorders. However, similarity of symptom profiles among the various subthreshold conditions, as well as with related, above-threshold disorders, suggests that symptomatology alone may not be the most important distinguishing feature of the subthreshold disorders. In addition, it is unknown whether symptoms of subthreshold disorders are in any way modified when they present in the context of stress.

Adjustment disorders has been described as a transitional diagnostic category, because the level of symptomatology and impairment in adjustment disorders was found to be intermediate between that observed in comparison groups who only had a DSM-III problem-level diagnosis, and patients with specific, above-threshold diagnoses. The symptomatic profile and level of impairment in patients with adjustment disorders with depressed mood was quite similar to that found in dysthymic disorder and atypical (minor) depression, although it was distinct from major depressive disorder and bipolar I disorder, suggesting poor discriminant validity of adjustment disorders with depressed mood with respect to minor depressive disorders. Adjustment disorders has also been described as an admission diagnosis or initial diagnosis for many adolescent and adult psychiatric inpatients. In one study, a large number (40 percent) of patients given this preliminary diagnosis were assigned a different diagnosis at the time of discharge. An even smaller number (18 percent) of those who were subsequently readmitted to the hospital were again diagnosed as having adjustment disorders.

Studies of adjustment disorders using structured diagnostic instruments have reported a fairly high level of comorbidity. In a mixed group of children, adolescents, and adults, approximately 70 percent of patients with adjustment disorders had at least one additional Axis I diagnosis. In the study of correlates of depressive disorder in children, 45 percent of those with adjustment disorders with depressed mood had another diagnosis. However, comorbidity of adjustment disorders with other Axis I disorders was less than in dysthymic disorder or major depressive disorder, suggesting a more pure or encapsulated disturbance in adjustment disorders. The extent to which comorbidity exists may be an important determinant in the level of impairment and risk for poor outcome in samples with adjustment disorders. One study, which compared the risk for continued pathology in individuals with adjustment disorders and psychiatric controls matched for comorbidity, found that adjustment disorders added no risk above that accounted for by comorbidity.

Studies have consistently reported a significant association of adjustment disorders with suicidal behavior, particularly in adolescents and young adults. In an urban hospital setting, 56 percent of all admissions for suicidal behavior were classified as having transient situational disorders using DSM-II criteria. Similarly, a retrospective review of 325 consecutive hospital admissions for deliberate self-poisoning found that 58 percent of cases met criteria for adjustment disorders with depressed mood. The majority were aged 15 to 24, with females predominating over males. Psychological autopsy data from several studies also indicate a high representation of adjustment disorders among suicide completers. In a Scandinavian sample of 58 consecutive suicide victims aged 15 to 29, 14 percent were classified as having adjustment disorders with depressed mood. In a U.S. population, 9 percent of suicide victims aged 10 to 29 were reported to have adjustment disorders. When compared with a group of patients with major depression who made suicide attempts, suicidal patients with adjustment disorders were more likely to make their attempt in the context of substance abuse. Moreover, suicidal behavior among patients with adjustment disorders was less likely to be planned. These findings underscore the seriousness of adjustment disorders in a subset of individuals and suggest that although the diagnosis may be subthreshold its morbidity may be highly significant, including risk of serious self-harm and death.

A 21-year-old college senior was in her usual state of health, vacationing with her aunt and uncle in Arizona during spring break. While she was away, a letter containing a copy of her full college transcript arrived at her parents' home. On reading the letter, her parents noticed that the grades on the transcript did not match those which they had previously seen at the end of each semester since freshman year. The mother called her daughter and confronted her with this information. During the conversation it became apparent that the daughter had been changing the grades on the transcripts that were sent home so that the grades her parents saw were better than was actually the case. An argument ensued and harsh words were exchanged. Following these events, the daughter became extremely upset and experienced feelings of anxiety and depressed mood. She felt ashamed about the situation and saw no way out. She spent most of the next 2 days in her room and did not participate in pleasurable activities; there were periods of crying as well as decreased eating and sleeping. Her aunt and uncle were only somewhat able to comfort her. Two days later, without warning, the patient ingested ten tablets of an over-the-counter sleep preparation. Later that day, the aunt found the patient to be fairly lethargic, and upon questioning, determined that she had made a suicide attempt. The patient was brought to a local emergency room, where she was admitted and retained in the psychiatric inpatient unit for several days. The patient's mother flew to Arizona to meet with her daughter and the hospital staff. Following crisis intervention, the patient was released with her mother to return home, where they consulted with another psychiatrist. By the time of this second evaluation, there were no longer signs of depressed or anxious mood, nor was there any suicidal ideation. The treatment plan consisted of crisis intervention, continued individual therapy for the patient on return to school, counseling for the parents, and conjoint family sessions to be held monthly during scheduled home visits. Despite continued individual and family upset over the precipitating event and the subsequent suicide attempt, the patient was able to return to school the following week to complete her spring semester.

Subtypes The diagnostic criteria for adjustment disorders define the contextual and temporal characteristics of a subthreshold response to a psychosocial stressor; the specific nature of the resultant symptom presentation is used as a means of subtyping. Six subtypes of adjustment disorders are described. Three of the subtypes define discrete symptomatic presentations (e.g., depressed mood, anxious mood, disturbance of conduct); two describe mixed clinical presentations (e.g., mixed emotional features; mixed disturbance of emotions and conduct); and one is a residual category for presentations not accurately described by any of the other subtypes (e.g., adjustment disorders unspecified). Previous editions of DSM included additional subtypes of adjustment disorders (e.g., adjustment disorders with physical complaints, work or academic inhibition, and withdrawal), but these were eliminated in DSM-IV because they were found to be infrequently used, even in

medical settings with physically ill patients. One additional subtype, adjustment disorders with suicidal behavior, was proposed for inclusion in DSM-IV. However, it was ultimately rejected because of concern that inclusion of this subtype might discourage systematic assessment and differential diagnosis of suicidal behavior, possibly resulting in the overidentification of adjustment disorders instead of a specific above-threshold diagnosis when suicidal behavior was present.

DSM-IV does not identify specific symptom profiles for the individual subtypes of adjustment disorders. However, an attempt was made to more accurately describe these subtypes and determine whether they are unique from above-threshold disorders. Symptoms found to be particularly salient for individual adjustment disorders subtypes were (1) depressed mood: depression, hyposomnia, low self-esteem and suicidal indicators; (2) anxiety: generalized anxiety, increased motor activity and situational anxiety; (3) disturbance conduct: impulsivity, lack of insight and violent behavior; and (4) mixed disturbance of emotions and conduct: excessive alcohol ingestion, depressant drugs that do not act on the central nervous system (CNS), suspiciousness, hostility, defrauding behavior, and homicidal ideation. Adjustment disorder subtypes did not differ from each other in the overall stressor severity or the highest level of functioning in the past year. However, impairment in current function was greatest in the depressive and mixed emotions and conduct subtypes.

DIFFERENTIAL DIAGNOSIS

The combination of subsyndromal symptomatology and the presence of an identified psychosocial stressor distinguishes adjustment disorders from all other Axis I and Axis II disorders in DSM-IV. Other conditions are either stress related but have an established diagnostic threshold, or subthreshold but not linked to stress. Posttraumatic stress disorder is characterized by a more narrowly defined concept of severe psychosocial stress, and a specific constellation of affective and autonomic symptoms lasting a minimum of 3 months. Acute stress disorder describes a similar, although possibly less severe, clinical presentation of less than 3 months' duration. In contrast, adjustment disorders can be triggered by a stressor of any severity, may present with a wide range of possible symptoms, and has no minimum or maximum duration. The individual subtypes of adjustment disorders must also be distinguished from atypical or subthreshold presentations of many of the major mental disorders (e.g., anxiety mood, conduct). However, the latter conditions need not be stress related and are therefore classified under the appropriate symptom domain using the designation not otherwise specified. Differential diagnosis should also include conditions that may be associated with impairment but are not deemed to be of sufficient severity to warrant a psychiatric diagnosis (e.g., bereavement).

Differences between adjustment disorders subtypes with respect to each other and to relevant above-threshold disorders were investigated in a reanalysis of the Western Psychiatric Institute data set in anticipation of DSM-IV. Each of the adjustment disorders subtypes was characterized by a set of 5 to 8 relatively discrete symptoms. In adults, adjustment disorders with depressed mood could be distinguished from above-threshold mood disorders both in terms of symptomatic presentation, severity and duration of impairment, and stressor severity. Similar trends were observed between adjustment disorders with anxious mood and above-threshold anxiety disorders. However, stressor severity was not found to be significantly different in adjustment disorders with anxious mood and posttraumatic stress disorder due to the low number of cases. In youth, symptomatic and functional differences were also detected between adjustment disorders with depressed mood and above-threshold depressive disorders. However, stressor severity was not found to differ between adjustment disorders and a variety of other disorders, quite possibly because stressor severity was high for all disorders in this age group. The fact that adjustment disorders does not appear to be more highly associated with stress in youth than a variety of other disorders raises questions regarding discriminant validity of the construct in this age group.

Similarities and differences between adjustment disorders and other subthreshold disorders represents an area of increasing interest. DSM-IV conducted field trials to investigate the proposed category of mixed anxiety-depression, in which threshold levels of symptomatology for each disorder alone were not achieved, but the total number of symptoms of both disorders combined exceeded a pre-determined threshold. Symptom profiles and descriptive features of mixed anxiety-depression were found to be relatively distinct from those of above-threshold mood and anxiety disorders. However, this study did not adequately examine the relationship of subthreshold symptomatology to stress, and ruled out overlap with adjustment disorders only by the extended duration of symptomatology in the proposed anxiety-depression category (i.e., distinguishing it from adjustment disorders as defined in DSM-III-R but not as modified in DSM-IV. A similar symptoms profile was also determined to be present in individuals who fulfilled criteria for the proposed diagnosis of subthreshold symptomatic depression. Interestingly, the symptom profiles for mixed anxiety-depression and subthreshold symptomatic depression were strikingly similar to those most frequently encountered among individuals with adjustment disorders in the Western Psychiatric Institute data. Taken together, these findings indicate that it is possible to more accurately describe symptomatic profiles of individuals with adjustment disorders and other subthreshold disorders. However, adequate discrimination among these various conditions remains elusive.

COURSE AND PROGNOSIS

Historically, adjustment disorders has not been considered to be an enduring diagnosis; individuals with this disorder have generally been thought to either experience a resolution of symptoms or progress to a more serious illness. However, this is not always the case, and DSM-IV now recognizes chronic as well as acute presentations of adjustment disorder. Nevertheless, the question of whether adjustment disorder predicts future psychiatric morbidity has considerable relevance for understanding whether the symptoms in adjustment disorder constitute a normal or pathological reaction to stress. It is also important to understand whether an episode of adjustment disorder might predict subsequent occurrences of symptomatically related, above-threshold disorders.

Studies of adults with adjustment disorder have generally indicated a benign clinical course. A follow-up study of military recruits using DSM-II criteria found that transient situational disorder was the least severe and disabling condition of those studied. Approximately 60 percent of this group was well at follow-up. Many of the 40 percent who demonstrated continued impairment were diagnosed as having a personality disorder. Follow-up of the University of Iowa sample indicated a high rate of recovery in adults with adjustment disorder, but a lower likelihood of recovery in adolescents (71 percent versus 44 percent). An additional 13 percent of the adolescent sample were well at follow-up but had intervening difficulties. In total, nearly 70 percent of the original sample of adolescents with adjustment disorder was ill at one time during the follow-up period. Adolescents were more likely than adults to have severe diagnoses at follow-up; some developed major mental disorders such as schizophrenia, bipolar I disorder, major depressive disorder, antisocial personality disorder, and drug abuse. Chronicity of behavioral problems was the best predictor of poor outcome. However, the original adjustment disorder subtype did not predict the follow-up diagnosis.

Other studies of adjustment disorder among children and adolescents have also reported poor outcome. In a survey of all adolescents receiving mental health treatment in a single geographical region, 52 percent of those originally diagnosed with transient situational disorder required treatment on 10-year follow-up. Another study, which examined the outcome of children with speech and language disorders, also indicated poor outcome in children with adjustment disorder. Of the 19 patients with this diagnosis at the time of initial assessment, only 26 percent were recovered after 4 years. The most frequent follow-up diagnoses were attention-deficit disorder, overanxious disorder, oppositional disorder, and avoidant disorder.

The longitudinal course of adjustment disorder with depressed mood was not characterized by increased risk for major depressive disorder in a study at the University of Pittsburgh that examined the course and prognosis of depressive disorders in children and adolescents. It is not known whether these children went on to develop different disorders outside the spectrum of mood disorders (and therefore may have still experienced poor outcome). However, a subsequent publication using a sample that included those and other children with adjustment disorder (not all of whom had adjustment disorder with depressed mood) found that adjustment disorder added no additional risk for poor outcome when compared with a psychopathological control group matched for comorbidity who did not have adjustment disorder. Those findings raise questions as to whether some of the risk for poor outcome reported in children and adolescents with adjustment disorder might be attributable to comorbidity with other disorders.

In contrast to this latter finding, high school students with adjustment disorder with depressed mood were at increased risk for major depressive disorder in young adulthood. Also, the University of Pittsburg group reported that adjustment disorder with depressed mood was commonly observed following the acute onset of diabetes mellitus, and predicted later psychiatric maladjustment. Of 92 children studied, 38 (41 percent) developed a new psychiatric disorder; 30 of these children met criteria for adjustment disorder. The disorder had a relatively rapid onset after diagnosis and was resolved in most cases within 2 to 3 months. Nevertheless, the cumulative probability of developing a psychiatric disorder in the subsequent 5 years was over three times as great in the children with adjustment disorder as compared to those without adjustment difficulties. Thus, development of adjustment disorder following the new onset of a medical illness predicted subsequent psychiatric dysfunction in a previously healthy population whereas it did not appear to affect later adjustment in a population with a high degree of psychiatric comorbidity. Taken together, these findings suggest that while the relationship between adjustment disorder and subsequent mental health function may be less dramatic than it is for other disorders, it is nonetheless real.

TREATMENT

Adjustment disorder may present with subthreshold pathology in one or several symptom domains. A variety of treatment interventions may therefore be useful in treating the heterogeneous group of individuals with this disorder. There are few, if any, treatment studies in individuals with adjustment disorder; indeed, very little is known or written about the treatment of individuals with subthreshold pathology of any kind. However, since adjustment disorder and other subclinical conditions

represent common psychiatric diagnoses in private practice, consultation-liaison situations, and primary care settings, systematic clinical trials with these conditions are warranted.

The treatment of individuals with adjustment disorder entails careful assessment of the nature and severity of the disturbance, taking into account the presence of risk factors that are known to be associated with poor outcome—for example, continuation of stressors, poor premorbid functioning. Interventions should be designed to minimize the impact of these stressors on day-to-day function. It is essential to understand the meaning of the stressor to the patient and why it seems to have been associated with the development of psychiatric symptoms at this time. Moreover, the clinician should appreciate the patient's level of vulnerability and capacity for adaptation. Is this a previously healthy, high-functioning individual who has become symptomatic in the context of a serious life stress? Or is this an individual with chronic, ill-defined coping problems who is having difficulty dealing with a variety of longstanding circumstances?

It is also important to understand and facilitate those factors that may mitigate the pathological response to stress. Examples include helping to develop a warm and supportive relationship with family members and helping the patient increase the breadth of social support. Attempting to minimize factors that maintain symptoms represents another important strategy. Examples might include working with parents of a child with adjustment disorder to help them better cope with the stressful events in their family because positive parental adjustment minimizes the likelihood of pathological adjustment in children.

In addition to these general recommendations, a range of specific treatment interventions may be considered in individuals with adjustment disorder, including psychodynamic, cognitive-behavioral, and supportive psychological interventions, counseling regarding life circumstance problems, and concrete assistance in the resolution of problematic situations. Short-term treatment may be sufficient for many patients with adjustment disorder, consistent with the conceptualization of the disorder as time limited. However, referral for more extended treatment following the reestablishment of baseline functioning may be desirable if there was pre-existing symptomatology or if there are individual characteristics that predispose the patient to stress intolerance. Crisis intervention and case management may be required on a short-term basis, often involving direct environmental manipulation in order to bring about a lessening of stress and enhanced external support. Individual psychotherapy may be either supportive or exploratory, although the latter strategy may be more suitable following resolution of the acute situation. Individual psychotherapy offers the patient an opportunity to better understand the meaning of the stressor and why it resulted in an impaired ability to function. Group therapies may be particularly useful for individuals who have experienced similar stresses—for example, in individuals who have recovered from a specific medical illness or surgical procedure.

There are no studies assessing the efficacy of pharmacological interventions in individuals with subsyndromal pathology; however, it may be reasonable to consider the judicious use of medication to treat specific symptoms associated with adjustment disorder. One report presented a rationale for the use of benzodiazepines in the treatment of adjustment disorder with anxious mood. Selective serotonin reuptake inhibitors have been found to be useful in treating symptoms of traumatic grief, and may be useful for other groups with subthreshold depressive symptoms. Indeed, a recent survey of prescribing practices among outpatient physicians indicated a robust increase in antidepressant use in individuals with adjustment disorder. It should be stressed, however, that pharmacological intervention in this population is most often used to augment psychosocial strategies rather than serving as the primary modality. Controlled clinical trials are needed to more accurately specify the role of pharmacotherapy in adjustment disorder and other minor disorders.

SUGGESTED CROSS-REFERENCES

Anxiety disorders, including posttraumatic stress disorder, are discussed in Chapter 15 and mood disorders in Chapter 14. Suicide is presented in Section 29.1 and Chapter 45. Separation and divorce are covered in Section 30.5 on couples therapy, Section 25.9 on stress, and Chapter 26 on relational problems. Bereavement is discussed in Section 28.5, and physical and sexual abuse is discussed in Section 28.7.

CHAPTER REFERENCES

Andreasen NC, Hoenk PR: The predictive value of adjustment disorders: A follow-up study. Am J Psychiatry *139:*589, 1982.

*Andreasen NC, Wasek P: Adjustment disorders in adolescents and adults. Arch Gen Psychiatry *37:*1166, 1980.

Bird HR, Canino G, Rubio-Stipec M, Gould MD, Ribera J, Sesman M, Woodbury M, Huertas-Goldman S, Pagan H, Sanchez-Lacay A, Moscoso M: Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico: The use of combined measures. Arch Gen Psychiatry *45:*1120, 1988.

Cantwell DP, Baker L: Stability and natural history of DSM-III childhood diagnoses. J Am Acad Child Adolesc Psychiatry *28:*691, 1989.

Despland JN, Monod L, Ferrero F: Clinical relevance of adjustment disorder in DSM-III-R and DSM-IV. Compr Psychiatry *36:*454, 1995.

Fabrega H Jr, Mezzich J: Adjustment disorder and psychiatric practice: Cultural and historical aspects. Psychiatry *50:*31, 1987.

*Fabrega H Jr, Mezzich J: Adjustment disorder as a marginal or transitional illness category in DSM-III. Arch Gen Psychiatry *44:*567, 1987.

Fowler RC, Rich CL, Young D: San Diego Suicide Study II. Substance abuse in young cases. Arch Gen Psychiatry *43:*962, 1986.

Goodyear JM, Kolvin I, Gatzanis S: The impact of recent undesirable life events on psychiatric disorders in childhood and adolescence. Br J Psychiatry *151:*179, 1987.

Gould M, Rutter M, Shaffer D, Sturge C: UK/WHO study of ICD-9. In *Assessment and Diagnosis in Child Psychopathology*, M Rutter, AH Tuma, IS Lann, editors. Guilford Press, New York, 1988.

Greenberg WM, Rosenfeld DN, Ortega EA: Adjustment disorder as an admission diagnosis. Am J Psychiatry *152:*459, 1995.

Judd LJ, Rapaport MH, Paulus MP, Brown JL: Subsyndromal symptomatic depression: A new mood disorder? J Clin Psychiatry *55*(Suppl):18, 1994.

Kendler KS, Neale M, Kessler R, Heath A, Eaves L: A twin study of recent life events and difficulties. Arch Gen Psychiatry *50:*789, 1993.

*Kovacs M, Gatsonis C, Pollock M, Parrone PL: A controlled prospective study of DSM-III adjustment disorder in childhood. Arch Gen Psychiatry *51:*535, 1994.

*Kovacs M, Ho V, Pollock MH: Criterion and predictive validity of the diagnosis of adjustment disorder: A prospective study of youths with new-onset insulin-dependent diabetes mellitus. Am J Psychiatry *152:*523, 1995.

Kovacs M, Feinberg T, Paulauskas S, Finkelstein R, Pollock M: Initial coping responses and psychosocial characteristics of children with insulin-dependent diabetes mellitus. J Pediatr *106:*827, 1985.

Kovacs M, Feinberg TL, Crouse-Novak MA, Paulauskas SL, Finkelstein R: Depressive disorders in childhood. I. A longitudinal prospective study of characteristics and recovery. Arch Gen Psychiatry *41:*229, 1984.

*Lewinsohn PM, Rhode P, Klein DN, Seeley JR: Natural course of adolescent major depressive disorder: I. Continuity into young adulthood. J Am Acad Child Adolesc Psychiatry *38:*56, 1999.

Margolis RL: Nonpsychiatric house staff frequently misdiagnose psychiatric disorders in general hospital inpatients. Psychosomatics *35:*485, 1994.

Marttunen MJ, Aro HM, Henriksson MM, Lonnqvist JK: Psychosocial stressors more common in adolescent suicides with alcohol abuse compared with depressive adolescent suicides. J Am Acad Child Adolesc Psychiatry *33:*490, 1994.

Mezzich AC, Mezzich JE, Coffman GA: Reliability of DSM-III and DSM-II in child psychopathology. J Am Acad Child Psychiatry *24:*273, 1985.

Mezzich JE, Fabrega H Jr, Coffman GA, Haley R: DSM-III disorders in a large sample of psychiatric patients: Frequency and specificity of diagnoses. Am J Psychiatry *146:*212, 1989.

Newcorn JH, Strain J: Adjustment disorder in children and adolescents. J Am Acad Child Adolesc Psychiatry *31:*318, 1992.

Olfson M, Marcus SC, Pincus HA, Zito JM, Thompson JW, Zarin DA: Antidepresant prescribing practices of outpatient psychiatrists. Arch Gen Psychiatry *55:*310, 1998.

Oxman TE, Barrett JE, Freeman DH, Manheimer E: Frequency and correlates of adjustment disorder related to cardiac surgery in older patients. Psychosomatics *35:*557, 1994.

Polyakova I, Knobler HY, Ambrumova A, Lerner V: Characteristics of suicidal attempts in major depression versus adjustment reactions. J Affect Disord *47:*159, 1998.

Runeson B: Mental disorder in youth suicide: DSM-III-R Axes I and II. Acta Psychiatr Scand *79:*490, 1989.

Setterberg SR, Ernst M, Rao U, Campbell M, Carlson GA, Shaffer D, Staghezza BM: Child psychiatrists' views of DSM-III-R: Survey of usage and opinions. J Am Acad Child Adolesc Psychiatry *30:*652, 1991.

*Strain JJ, Newcorn J, Mezzich J, Kirisci L: Adjustment disorder: The McArthur reanalysis. In *DSM-IV Sourcebook*, vol 4, TA Widiger, AJ Frances, HA Pincus, R Ross, MB First, W Davis, editors. American Psychiatric Press, Washington, DC, 1998.

Strain JJ, Newcorn J, Wolf D, Fulop G: Adjustment disorder. In *American Psychiatric Press Textbook of Psychiatry*, RE Hales, SC Yudofsky, JA Talbott, editors. American Psychiatric Association Press, Washington, DC, 1994.

Strain JJ, Smith GC, Hammer JS, McKenzie DP, Blumenfield M, Muskin P, Newstadt G, Wallack J, Wilmer A, Schleifer SS: Adjustment disorder: A multisite study of its utilization and interventions in the consultation-liaison psychiatry setting. Gen Hosp Psychiatry *20:*139, 1998.

True WR, Rice RJ, Eisen SA, Heath AC, Goldberg J, Lyons MJ, Nowak J: A twin study of genetic and environmental contribution to liability for posttraumatic stress symptoms. Arch Gen Psychiatry *50:*257, 1993.

Woolston JL: Theoretical considerations of the adjustment disorders. J Am Acad Child Adolesc Psychiatry *27:*280, 1988.

Zinbarg RE, Barlow DH, Liebowitz MR, Street L, Broad E, Katon W, Roy-Byrne P, Lepine JP, Telerani M, Richards Brantley PJ, Kraemer H: The DSM-IV field trial for mixed anxiety-depression. In *DSM-IV Sourcebook*, vol 4, TA Widiger, AJ Frances, HA Pincus, R Ross, MB First, W Davis, editors. American Psychiatric Press, Washington, DC, 1990.

Zygmont R, Prigerson HG, Houck PR, Miller MD, Sheer M, Jacobs S, Reynolds CF III: A posthoc comparison of paroxetine and nortriptyline for symptoms of traumatic grief. J Clin Psychiatry *59:*241, 1998.

CHAPTER 24. PERSONALITY DISORDERS

C. ROBERT CLONINGER, M.D., AND DRAGAN M. SVRAKIC, M.D., ḥH.D.

Personality disorders have become a major social, medical, and scientific problem. Serious concerns about these disorders have been raised by both experts and lay persons, including mental health professionals, sociologists, public health workers, politicians, charity activists, community leaders, and news analysts. No demographic group is immune to personality disorders. Indeed, estimates of the prevalence for these disorders in the general population range from 11 to 23 percent, depending on the severity of impairment required for diagnosis. This translates into an alarming 1 in every 4 to 10 individuals in our neighborhoods having personality disorder, regardless of location or socioeconomic status. These individuals have chronic impairments in their ability to work and love and tend to be less educated, single, drug addicts, and sex offenders, and to have marital difficulties and be unemployed. In addition, many perpetrators of violent and nonviolent crime as well as most prison inmates have a personality disorder.

In psychiatry, about one half of all patients have personality disorder, frequently comorbid with Axis I conditions. Personality factors interfere with the response to treatment of all Axis I syndromes and increase personal incapacitation, morbidity, and mortality of these patients. Moreover, personality disorders are a predisposing factor for many other psychiatric disorders, including substance use disorders, suicide, mood disorders, impulse-control disorders, eating disorders, and anxiety disorders. Patients with personality disorders are often perceived as “aggravating” and “demanding,” and may be rejected by therapists who expect them to have a poor prognosis. Alternatively, they may be seductive and dependent and elicit inappropriate responses from their therapist, such as sexual interest or the urge to rescue. This blurring of professional boundaries can lead to loss of objectivity, when patient's verbal and nonverbal communication takes on personal, rather than diagnostic and therapeutic, significance. Indeed, these patients test the limits of any physician's theoretical knowledge, practical skills, and maturity of character. The rare physician with low narcissism, high energy, and high tolerance is optimal for treating patients with personality disorder.

A somewhat lower, but still substantial, proportion of patients in somatic medicine have personality disorders comorbid with their physical illnesses. Personality factors have been associated with increased risk for coronary artery disease, angina pectoris, human immunodeficiency virus (HIV) infection, psoriasis, ulcerative colitis, and many other so-called psychosomatic diseases. Yet, patients with personality problems who seek medical help frequently have a negative workup and no medical explanation for their complaints. Clearly, people with personality disorder consume a large portion of community services, public health resources, charity, and social programs.

Since the 1980s there has been a dramatic increase in interest in personality disorders. For example, in 1975, the MEDLINE search service of the National Library of Medicine listed 69 published articles about personality disorder, in 1985 the number was 262, and in the years 1996 and 1997 the number escalated to 3079. In addition, numerous monographs dealing with personality disorders have been published over the last two decades. Reflecting this trend, the International Society for the Study of Personality Disorders and its official journal, the *Journal of Personality Disorders,* were founded during the past decade.

Several phenomena help explain the dramatic shift in professional sensitivity to personality disorders and a growing interest in their systematic study. First, personality disorders are diagnosed far more frequently nowadays than in the past. This reflects the introduction of operationalized diagnostic criteria and the multiaxial diagnostic system in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) in 1980. By specifying precise descriptive criteria based on typical behaviors and feelings for each of the classified personality syndromes, DSM-III enabled clinicians to diagnose personality disorders more systematically. These criteria were revised and standardized in the revised third edition of DSM (DSM-III-R) in 1987 and the fourth edition of DSM (DSM-IV) in 1994. The multiaxial diagnostic system means that personality disorders and mental retardation are classified on Axis II, whereas other clinical syndromes are classified on Axis I. The availability of these two diagnostic axes facilitates the diagnosis of personality disorders because it ensures that personality disorders are not overlooked, even when they co-occur with a clinically more prominent Axis I syndrome.

Second, the increased clinical recognition of personality disorders probably reflects their increased prevalence in the general population. At the turn of the twentieth century, when Freud was pioneering his understanding of our human nature, neuroses and neurotic problems were the dominant “psychopathology of every day life.” At the time, the central psychological task people faced was finding socially acceptable ways to express their “asocial” impulses, predominantly aggression and sexuality. The full complexity of this struggle with personal conflicts and demands for self control is illustrated in Karen Horney's masterpiece *The Neurotic Personality of Our Time* in the 1950s. Obviously, behavior codes of our present ethically more liberal society differ substantially from the ethical conservativism at the turn of the century. Both aggression and any kind of sexuality are much easier to express. Personal descriptors such as “dominant,” “powerful,” or “competitive” have gained wide social approval. Even modern psychology has begun to distinguish between “angry” (unacceptable) and “instrumental” (competitive, acceptable) aggression, the former having as reward “discomfort of the victim” and the latter “dominance in the society.” Instead of searching for ways to express their impulses, people are currently faced with questions of identity, meaning, and choice. Clearly, the objective of the psychological puzzle has shifted from morality to identity and, therefore, from the society to the person. The modern day “psychopathology of everyday life” has changed accordingly. Nowadays, one more frequently encounters patients who need help but cannot precisely describe their problems, who are more angry than anxious, who struggle with questions of purpose rather than with guilt, who are ambivalent rather than inhibited, who feel empty rather than sad, and who manifest a peculiar inability to learn from experience and tend to repeat maladaptive behaviors over and over again. Such people are likely to be diagnosed as having personality disorder. Due to the obviously increased number of such patients, and motivated by the desire to help them, psychiatrists have started to study them more thoroughly and more frequently.

Third, personality disorder is the clinical arena in which psychodynamic psychiatry and biological psychiatry is most likely to be reconciled. In many ways, this disorder is somewhere in between minor and major psychiatric problems (e.g., adjustment disorder and schizophrenia), and it is thus interesting from both sociodynamic and biological standpoints.

Fourth, personality disorders are fairly common and underlie susceptibility to many other psychiatric and medical problems. Accordingly, the diagnosis and treatment of any patient with psychosocial problems, as well as prevention planning in many medical syndromes, are inadequate without a systematic approach to assessing and classifying personality.

Fifth, the development of psychiatry as a whole has generated increased interest in personality disorders. Every science in its initial stages has a tendency to define its subject as clearly as possible. Once sufficiently developed, it naturally begins to recognize various mixed and transitional forms. Clearly personality disorders represent a deviance from normality that psychiatry, because of its development, has begun to recognize. Consequently, numerous questions about personality disorder remain to be answered definitively, but recent progress in understanding their underlying psychobiological mechanisms is giving psychiatry a stronger foundation for understanding both normal and abnormal behavior.

BASIC DEFINITIONS AND TERMINOLOGY

Personality, temperament, motivation, and *character* are often used interchangeably. Unfortunately, this ambiguity is both misleading and unjustified. In what follows, each of these concepts is reviewed and clearly distinguished.

Personality Personality is both complex and unique in that individuals differ greatly from one another in multiple components of behavior, and each expresses only one of many potential lifestyles. All existing definitions of personality are functional (i.e., they focus on questions related to motivation and mental adaptation of the organism). Specifically, Gordon Allport defined personality as the “dynamic organization within the individual of those psychophysical systems that determine his or her unique adjustment to the environment.”

Allport elaborated on this definition by explaining that “dynamic organization” emphasizes that personality is an organized system (“unitas multiplex”) that is constantly evolving and changing. The phrase “within the individual” means that personality is what lies behind a specific individual's acts. The term “psychophysical” reminds one that personality is neither exclusively mental nor exclusively neural but a combination of the two. The verb “determine” indicates that systems that constitute personality guide expressive and adaptive behaviors. The expression “adjustment to the environment” has both functional and evolutionary significance pointing to personality as a mode of survival and, more generally, adaptation.

It is widely accepted that personality develops through the interaction of hereditary dispositions and environmental influences. The notion of genetic canalization (or “epigenetic landscape”) has been revised to include the reciprocal necessity of genetic endowment and environmental stimulation in the development of behaviors. Genetic differences account for about half of the variance in most normally distributed temperament traits. Of the remaining 50 percent of the variance, 25 to 30 percent is explained by nonshared environmental effects (i.e., experiences unique to the individual), and 15 to 20 percent by measurement error. Contrary to common belief, environmental factors shared by siblings reared together have little or no effect on basic temperament traits but do influence other aspects of personality.

From the structural standpoint, most authors agree that personality consists of temperament, character, and intelligence. Roughly speaking, temperament reflects biological contributions, and character reflects social and cultural contributions, to personality. Intelligence impregnates both constitutional and social traits and modifies overall personality functions. Basic functions of personality are to feel, think, and perceive and to incorporate these into purposeful behaviors.

Temperament Pioneering work by Alexander Thomas and Stella Chess conceptualized temperament as the stylistic component (how) of behavior, as differentiated from motivation (why) and content (what) of behavior. Modern concepts of temperament, however, emphasize its emotional, motivational, and adaptive aspects. Specifically, four major temperament traits have been identified: harm avoidance, novelty seeking, reward dependence, and persistence. Remarkably, this four-factor model of temperament can, in retrospect, be seen as a modern interpretation of the ancient four temperaments: individuals differ in the degree to which they are melancholic (harm avoidant), choleric (novelty seeking), sanguine (reward dependent), and phlegmatic (persistent). However, the four temperaments are now considered genetically independent dimensions that occur in all possible factorial combinations rather than mutually exclusive categories.

Distinction Between Temperament and Character The four dimensions of human temperament correspond closely to those observed in other mammals, such as rodents or dogs. Multiple levels in the phylogeny of learning abilities in animals from invertebrates to man indicate that learning has multiple component processes that are hierarchically organized and interact extensively throughout development (Fig. 24-1).

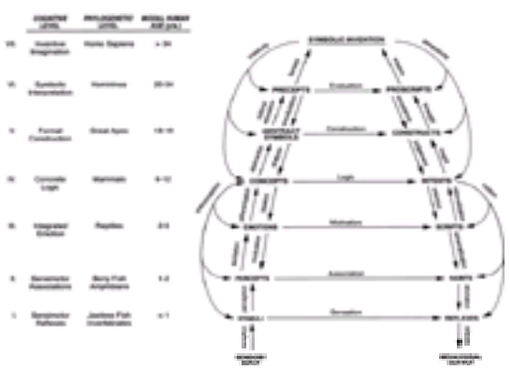

FIGURE 24-1 A hierarchical model of learning.

In Figure 24-1, temperament corresponds to the processes of sensation, association, and motivation that underlie the integration of skills and habits based on emotion. Character corresponds to the processes of symbolization and abstraction that are based on conceptual learning. Figure 24-1 is useful to specify the hierarchical phylogenetic and ontogenetic organization of learning and its relevance to the concepts of temperament and character not to compare learning in humans at a particular age with learning in lower animals. Specifically, temperament and character are conceptualized on the basis of two types of memory and learning, propositional and procedural, which have been described in humans and primates. Temperament (or the “emotional core” of personality) involves procedural memory regulated by the corticostriatolimbic system, primarily the sensory cortical areas, amygdala, and the caudate and putamen. Procedural memory underlies associative learning and involves presemantic perceptual processing of visuospatial information and affective valence that can operate independently of abstract conceptual or volitional processes or both. Propositional learning involves higher cognitive functions of abstraction and symbolization. These two major memory and learning systems can be dissociated functionally. For example, individuals with Parkinson's disease, characterized by striatal lesions, exhibit deficits in procedural learning but not in propositional learning. In contrast, individuals with an amnestic syndrome, characterized by lesions in the medial temporal lobe, have deficits in propositional learning but not in procedural learning.

Temperament traits of harm avoidance, novelty seeking, reward dependence, and persistence are defined as heritable differences underlying one's automatic response to danger, novelty, and various types of reward, respectively. These four temperament traits are closely associated with the four basic emotions of fear (harm avoidance), anger (novelty seeking), attachment (reward dependence), and mastery (persistence). Individual differences in temperament and basic emotions modify the processing of sensory information and critically shape early learning characteristics, especially associative conditioning of unconscious behavior responses. In other words, temperament is conceptualized as heritable biases in emotionality and learning that underlie the acquisition of emotion-based, automatic behavior traits and habits observable early in life and relatively stable over one's life span.

Each of the four major dimensions is a normally distributed quantitative trait. The four dimensions were shown to be genetically homogeneous and independently inherited from one another in large, independent twin studies in the United States and Australia. Temperamental differences, which are not very stable initially, tend to stabilize during the second and third years of life. Accordingly, ratings of these four temperament traits at age 10 years predicted personality traits at ages 15, 18, and 27 years in a large sample of Swedish children. The four dimensions have been repeatedly shown to be universal across different cultures, ethnic groups, and political systems on five continents. In summary, these aspects of personality are called “temperament” because they are heritable, observable early in childhood, relatively stable in time, moderately predictive of adolescent and adult behavior, and consistent in different cultures. Table 24-1 summarizes contrasting sets of behaviors that distinguish extreme scorers on the four dimensions of temperament. Note that each extreme of these dimensions has specific adaptive advantages and disadvantages (i.e., neither high nor low scores inherently means better adaptation).

Temperament Dimension	Descriptors of Extreme Variants	
	High	Low
Harm avoidance	Pessimistic	Optimistic
	Fearful	Daring
	Shy	Outgoing
	Fatigable	Energetic
Novelty seeking	Exploratory	Reserved
	Impulsive	Deliberate
	Extravagant	Thrifty
	Irritable	Stoical
Reward dependence	Sentimental	Detached
	Open	Aloof
	Warm	Cold
	Affectionate	Independent
Persistence	Industrious	Lazy
	Determined	Spoiled
	Enthusiastic	Underachieving
	Perfectionist	Pragmatic

Table 24-1 Descriptors of Individuals Who Score High and Low on the Four Temperament Dimensions

The component traits (“facets”) for each of the four temperament dimension are more strongly correlated with one another than with any other component; they have

distinct learning characteristics and independent genetic antecedents. In other words, they share a common source of covariation that is strong and invariant regardless of environmental change and past experience.

Neurobiological studies of animals, most often rodents, have been highly informative about the functional organization of brain systems underlying procedural learning and temperament. Explicit animal models have been described on the basis of extensive work in rodents and other nonprimates, providing hypotheses being tested in humans with modern techniques of brain imaging, neurochemistry, and neurogenetics. The most comprehensive neurobiological model of learning in animals that has been systematically related to the structure of human temperament is summarized in Table 24-2. This model distinguishes four dissociable brain systems for behavioral activation (novelty seeking), behavioral inhibition (harm avoidance), social attachment (reward dependence), and partial reinforcement (persistence).

Brain System–Related Personality Dimension	Principal Neuromodulators	Relevant Stimuli	Behavioral Response
Behavioral activation–novelty seeking	Dopamine	Novelty CS of reward CS or LCS of relief of monotony or punishment	Exploratory pursuit Appetitive approach Active avoidance Escape
Behavioral inhibition–harm avoidance	GABA Serotonin (dorsal raphe)	Aversive conditioning (pairing CS+LCS) Conditioned signals for punishment, novelty, frustrative nonreward	Formation of aversive CS Passive avoidance, extinction
Social attachment–reward dependence	Norepinephrine Serotonin (median raphe)	Reward conditioning (pairing CS+UCS)	Formation of appetitive CS
Partial reinforcement–persistence	Glutamate Serotonin (dorsal raphe)	Intermittent reinforcement	Resistance to extinction

GABA, γ-aminobutyric acid; CS, conditioned signals; UCS, unconditioned signals.

Table 24-2 Four Dissociable Brain Systems Influencing Stimulus-Response Patterns Underlying Temperament

HARM AVOIDANCE Harm avoidance involves a heritable bias in the inhibition of behavior in response to signals of punishment and frustrative nonreward. High harm avoidance is observed as fear of uncertainty, social inhibition, shyness with strangers, rapid fatigability, and pessimistic worry in anticipation of problems, even in situations that do not worry other people. Adaptive advantages of high harm avoidance are cautiousness and careful planning when hazard is likely. The disadvantages occur when hazard is unlikely but still anticipated, which leads to maladaptive inhibition and anxiety. Persons low in harm avoidance are carefree, courageous, energetic, outgoing, and optimistic, even in situations that worry most people. The advantage of low harm avoidance is confidence in the face of danger and uncertainty, leading to optimistic and energetic efforts with little or no distress. The disadvantages are related to unresponsiveness to danger or unreasonable optimism with potentially severe consequences when hazard is likely.

The psychobiology of harm avoidance is complex. In animal studies, ascending serotonergic projections from the dorsal raphe nuclei to the substantia nigra inhibit nigrostriatal dopaminergic neurons and are essential for conditioned inhibition of activity by signals of punishment and frustrative nonreward. Benzodiazepines disinhibit passive avoidance conditioning by g-aminobutyric acid (GABA)-ergic inhibition of serotonergic neurons originating in the dorsal raphe nuclei. The anterior serotonergic cells in the dorsal raphe nucleus intermingle with the dopaminergic cells of the ventral tegmental area, and both groups innervate the same structures (e.g., basal ganglia, accumbens, amygdala), providing opposing dopaminergic-serotonergic influences in the modulation of approach and avoidance behavior. The anterior serotonergic projections from the dorsal raphe to striatum, accumbens, amygdala, and frontal cortex are usually associated with serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptors. Individuals high in both harm avoidance and novelty seeking are expected to have frequent approach-avoidance conflicts, as seen in cycles of binging and purging in patients with bulimia. More generally, excessive behavioral inhibition (i.e., high harm avoidance) predisposes individuals to anxiety, depression, and low self-esteem. Effective antidepressant treatment lowers harm avoidance scores, but higher harm avoidance scores predict poorer responses to antidepressants, including tricyclic drugs and selective serotonin reuptake inhibitors.

Recent neuropsychological studies confirm that harm avoidance is associated with individual differences in classical aversive conditioning, whereas other dimensions of personality are uncorrelated (Table 24-3). Harm avoidance (and not other dimensions of temperament) is also replicably associated with potentiation of startle responses. Neuropsychological tests also confirm that harm avoidance is associated with other aspects of behavioral inhibition. For example, harm avoidance is associated with the Posner validity effect. The Posner task uses a detection reaction time paradigm with discrete presentation of simple visual stimuli in the periphery. Cues that correctly direct attention to the spatial location where the target stimulus will appear are called *valid cues,* whereas those that direct attention to incorrect locations are called *invalid cues.* In three large samples of healthy college students, those higher in harm avoidance had greater slowing of their reactions after invalid cues, or less benefit from valid cues, than others.

Variable	Effect
Neuroanatomy (PET)	
Medial prefrontal (L)	Increased activity (behavioral inhibition)
Anterior paralimbic (R)	Increased activity (sensitivity to threat)
Neuropsychology	
Aversive conditioning	Greater associative pairing with punishment ($r = .4$)
Eyeblink startle reflex	Potentiation of response to aversive stimulus (effect size, 1.9)
Posner validity effect	Greater slowing of responses after invalid cues ($r = .3$)
Spatial delayed response	Better ability to delay responses after ingesting amphetamines ($r = .5$)
Neurochemistry	
Platelet 5-HT_2 receptor	Fewer receptors ($r = -.6$)
Plasma GABA	Lower concentration ($r = -.5$)
Neurogenetics	
5-HT transporter promoter	Greater reuptake

PET, positron emission tomography.

Table 24-3 Reported Psychobiological Correlates of Harm Avoidance

Positron emission tomography (PET) at the National Institute of Mental Health (NIMH) with ^{18}F-deoxyglucose (FDG) in 31 healthy adult volunteers during a simple, continuous, performance task showed that harm avoidance was associated with increased activity in the anterior paralimbic circuit, specifically the right amygdala and insula, the right orbitofrontal cortex, and the left medial prefrontal cortex. This activation pattern corresponds well to the 5-HT_2 terminal projections of the dorsal raphe. However, 5-HT_2 receptor numbers have been correlated with harm avoidance only in studies of platelets (Table 24-3).

Higher plasma GABA concentrations have also been correlated with low harm avoidance. Plasma GABA concentration has also been correlated with other measures of anxiety proneness, and it correlates highly with brain GABA concentration. Finally, a gene on chromosome 17q12 that regulates the expression of the serotonin transporter accounts for 4 to 9 percent of the total variance in harm avoidance in four of six tests of this relation. These findings support a role for both GABA and serotonergic projections from the dorsal raphe underlying individual differences in behavioral inhibition as measured by harm avoidance.

NOVELTY SEEKING Novelty seeking reflects a heritable bias in the initiation or activation of appetitive approach in response to novelty, approach to signals of reward, active avoidance of conditioned signals of punishment, and skilled escape from unconditioned punishment. All four of these behaviors are hypothesized to covary as part of one heritable system of learning. They are observed as exploratory activity in response to novelty, impulsiveness, extravagance in approach to cues of reward, and active avoidance of frustration. Individuals high in novelty seeking are quick-tempered, curious, easily bored, impulsive, extravagant, and disorderly. Adaptive advantages of high novelty seeking are enthusiastic explorations of new and unfamiliar stimuli, potentially leading to originality, discoveries, and reward. The disadvantages are frequent and easy boredom, excessive impulsivity and angry outbursts, potential fickleness in relationships, and impressionism in efforts.

Persons low in novelty seeking are slow tempered, uninquiring, stoical, reflective, frugal, reserved, tolerant of monotony, and orderly. Their reflectiveness, stoical resilience, systematic efforts, and meticulous approach are clearly advantageous when these features are adaptively needed. The disadvantages reflect an uninquiring attitude, lack of enthusiasm, and tolerance of monotony, potentially leading to prosaic routinization of activities.

Mesolimbic and mesofrontal dopaminergic projections have a crucial role in incentive activation of each aspect of novelty seeking in animals. For example, dopamine-depleting lesions in the nucleus accumbens or the ventral tegmentum lead to neglect of novel environmental stimuli and reduce both spontaneous activity and investigative behavior. Behavioral activation by dopaminergic agonists depends on integrity of the nucleus accumbens but not the caudate nucleus. In human studies, individuals at risk for Parkinson's disease have low premorbid scores in novelty seeking but not other dimensions of personality, supporting the importance of dopamine in incentive activation of pleasurable behavior. The initiation and frequency of hyperactivity, binge eating, sexual hedonism, drinking, smoking, and other substance abuse (especially stimulants) are each associated with high scores in novelty seeking.

The psychobiological correlates reported for novelty seeking are summarized in Table 24-4. High scores correlate with increased metabolic activity during continuous performance tasks on PET in cingulate cortex and left caudate. In addition, high novelty seeking is associated with decreased activity in the left prefrontal cortex, including the medial prefrontal region associated with increased activity in individuals scoring high in harm avoidance. This suggests that medial prefrontal cortex may be an important site in the processing of approach-avoidance conflicts.

Variable	Effect
Neuroanatomy (PET)	
Medial prefrontal (L)	Decreased activity (behavioral disinhibition)
Dorsolateral prefrontal	Decreased activity (behavioral disinhibition)
Cingulate	Increased activity (behavioral activation)
Caudate (L)	Increased activity (behavioral activation)
Neuropsychology	
Reaction time	Slower to respond if not reinforced (neutral stimuli, $r = -.4$)
Stimulus intensity (N1/P2 ERP)*	Augmentation of intensity of cortical responses to novel auditory stimuli ($r = .5$)
Sedation threshold	More easily sedated by diazepam (Valium) (lower threshold, $r = -.3$)
Key word list memory	Deterioration of verbal memory when excited (after ingesting amphetamine, $r = .6$)
Neurochemistry	
Dopamine transporter	Higher density observed in striatum
Platelet MAO_B	Lower activity (associated with cigarette smoking)
Neurogenetics	
Dopamine receptor D_4	Association with long alleles of exon III variant
Dopamine transporter	Greater reuptake

Table 24-4 Reported Psychobiological Correlates of Novelty Seeking

Evoked potential studies of stimulus-intensity dependence confirm that novelty seeking is associated with augmentation of stimulus intensity, particularly with novel stimuli. Novelty seekers are also unusually sensitive to both sedatives and stimulants; they are easily oversedated by benzodiazepines and overstimulated by amphetamines, leading to deterioration in their information processing. Their reaction times are slow when they are not reinforced.

The association of increased striatal activity with high novelty seeking is more specifically associated with higher density of the dopamine transporter, suggesting that novelty seeking involves increased reuptake of dopamine at presynaptic terminals, thereby requiring frequent stimulation to maintain optimal levels of postsynaptic dopaminergic stimulation. Novelty seeking leads to various pleasure-seeking behaviors, including cigarette smoking, which may explain the frequent observation of low platelet monoamine oxidase type B MAO_B activity because cigarette smoking has the effect of inhibiting MAO_B activity in platelets and brain.

Studies of candidate genes involved in dopamine neurotransmission, such as the dopamine transporter gene (*DAT*1) and the type 4 dopamine receptor gene (*DRD*4) have provided evidence of association with novelty seeking and no other dimension of temperament. The dopamine transporter, responsible for presynaptic reuptake of dopamine and a major site of action of drugs including stimulants such as methylphenidate (Ritalin) and the antidepressant bupropion (Wellbutrin) is encoded by locus SLC6A3 (alias *DAT*1) on chromosome 5p. Polymorphisms at this locus are associated with attention-deficit disorder and other disorders related to variation in novelty seeking. Likewise polymorphisms at the *DRD*4 locus have been associated with attention-deficit disorders, opioid dependence, and other traits related to novelty seeking.

REWARD DEPENDENCE Reward dependence reflects a heritable bias in the maintenance of behavior in response to cues of social reward. It is observed as sentimentality, social sensitivity, attachment, and dependence on approval by others. Individuals high in reward dependence are tender hearted, sensitive, socially dependent, and sociable. One of the major adaptive advantages of high reward dependence is the sensitivity to social cues that facilitates affectionate social relations and genuine care for others. The disadvantage is related to suggestibility and loss of objectivity frequently encountered with people who are excessively socially dependent.

Individuals low in reward dependence are practical, tough minded, cold, socially insensitive, irresolute, and indifferent if alone. The advantages of low reward dependence are personal independence and objectivity not biased by efforts to please others. Its adaptive disadvantage is related to social withdrawal, detachment, and coldness in social attitudes.

Noradrenergic projections from the locus ceruleus and serotonergic projections from the median raphe are thought to influence such reward conditioning (Table 24-2). In animals, stimulation of the noradrenergic locus ceruleus or its dorsal bundle or direct application of norepinephrine decreases the firing rate of terminal neurons and increases their sensitivity to other afferents, so that targeted stimuli can stand out from nontargeted stimuli. In humans, short-term reduction of norepinephrine release by short-term infusion of the a_2 presynaptic agonist clonidine (Catapres) selectively impairs paired-associate learning, particularly the acquisition of novel associations. The noradrenergic locus ceruleus is located at the same posterior level of the brainstem as the serotonergic median raphe, and both of these posterior monoamine cells innervate structures important to formation of paired associations, such as the thalamus, neocortex, and hippocampus. Neurophysiological studies show that the anterior temporal lobe decodes social signals, such as facial images and social gestures. Consequently, individuals high in reward dependence are expected to be particularly sensitive in their social communication, whereas those low in reward dependence are expected to be socially aloof.

The psychobiological correlates reported about reward dependence are summarized in Table 24-5. As predicted, reward dependence is associated with individual differences in formation of conditioned signals of reward. This is also supported by its association with individual differences in paired-associate learning.

Variable	Effect
Neuroanatomy (PET)	
Thalamus	Increased activity (facilitates sensory processing)
Neuropsychology	
Reward conditioning	Increased associative pairing with rewards ($r = .3$)
Paired associates	Better learning of novel associations ($r = .5$)
Posner validity effect	Faster responses after valid cues ($r = -.4$)
Neurochemistry	
Urinary MHPG	Less excretion of norepinephrine metabolite ($r = -.4$)
Plasma cortisol	Higher morning cortisol concentration when depressed ($r = .3$)
Urinary harman	Greater excretion of indoleamine product in alcoholics high in reward dependence ($r = .7$)
Neurogenetics	
5-HT_{1d} receptor	Allelic association (effect size, 2.0)

Table 24-5 Reported Psychobiological Correlates of Reward Dependence

High reward dependence is associated with increased activity in the thalamus, which is consistent with proposals about the importance of serotonergic projections to the thalamus from the median raphe in modulation of social communication. This is further supported by the finding of low 3-methoxy-4-hydroxyphenylglycol (MHPG) concentration with high reward dependence. High reward dependence is also associated with hypercortisolemia in patients with melancholia but not in individuals who are not depressed.

PERSISTENCE Persistence reflects a heritable bias in the maintenance of behavior despite frustration, fatigue, and intermittent reinforcement. It is observed as

industriousness, determination, ambitiousness, and perfectionism. Highly persistent people are hard-working, perseverant, and ambitious overachievers who tend to intensify their effort in response to anticipated reward and perceive frustration and fatigue as a personal challenge. High persistence is an adaptive behavioral strategy when rewards are intermittent but contingencies remain stable. When the contingencies change rapidly, perseveration becomes maladaptive. Individuals low in persistence are indolent, inactive, unstable, and erratic; they tend to give up easily when faced with frustration, rarely strive for higher accomplishments, and manifest little perseverance even in response to intermittent reward. Accordingly, low persistence is an adaptive strategy when reward contingencies change rapidly and may be maladaptive when rewards are infrequent but occur in the long run.

Persistence can be objectively measured by the partial reinforcement extinction effect, in which persistent individuals are more resistant to the extinction of previously intermittently rewarded behavior than individuals who have been continuously reinforced. Recent work in rodents showed that the integrity of the partial reinforcement extinction effect depends on projections from the hippocampal subiculum to the nucleus accumbens. This glutaminergic projection may be considered a short circuit from the behavioral inhibition system to the behavioral activation system, thereby converting a conditioned signal of punishment into a conditioned signal of anticipated reward. This connection is probably disrupted in humans by lesions of the orbitomedial cortex that may have a specific antipersistence effect of therapeutic benefit to some severely obsessive-compulsive patients. Bilateral cingulotomy, which reduces harm avoidance only, is less effective in reducing persistent compulsive behavior than cingulotomy combined with orbitomedial lesions.

The reported psychobiological correlates of persistence are summarized in Table 24-6. Less work is available about this than the other three temperament dimensions, because it has been distinguished as an independently inherited dimension only since 1993. Neuropsychological studies do confirm that people high in persistence make more effort to learn without reinforcement. In laboratory gambling tasks, individuals high in persistence are more perseverative; that is, they exhibit the “gambler's fallacy” by maintaining the same bet size despite repeated losses. There is one report that a polymorphism at the 5-HT $_{2C}$ receptor gene locus is associated with both reward dependence and persistence. These effects were further facilitated by interactions with *DRD*4 and *DRD*3.

Variable	Effect
Neuroanatomy	
Orbitomedial cortex	Disconnection reduces perseveration
Neuropsychology	
Gambling style	Perseveration in bet size despite continuing losses
Key word list learning	Learning without reinforcement (r = .5)
Neurochemistry	
Glutaminergic connection Subiculum to nucleus accumbens	Essential for PREE* in rodents
Neurogenetics	
5-HT$_{2C}$ receptor	Allelic association (effect size, 2.0)

* Partial reinforcement extinction effect (PREE) is persistence or increased resistance to extinction following intermittent reinforcement.

Table 24-6 Reported Psychobiological Correlates of Persistence

Motivation Survival and reproduction, the basic drives for all animal species, are expressed in humans primarily through their experiential derivatives, emotions. In contrast to the limited motivational spectrum of the basic drives, emotions have an independent motivational power that makes them not only the primary motivational system for human beings, but also the personality processes that give meaning to human existence.

Temperament traits of harm avoidance, novelty seeking, reward dependence, and persistence, with their respective primary emotions of fear, anger, attachment, and ambition, are observable early in development. Research in children has shown that during the first several months of life, fear and anger differentiate from the disposition to *distress* (a tendency to become upset and aroused autonomically easily and intensely). In this period, characterized by active regressive changes in the organization of neuronal circuits and density of synapses, especially in frontal, temporal, and limbic cortical areas, many other complex and social functions in human behavior emerge. These new functions are, however, organized around an early set of enduring behavioral and emotional dispositions referred to in this work as *temperament traits*.

Emotional states associated with the four temperaments depend on whether a person has a high or low disposition for that particular temperament trait and whether the occurring experience is positive (gratifying) or negative (frustrating) (Table 24-7).

Temperament Dimension	High Scorers		Low Scorers	
	+	–	+	–
Harm avoidance	Anxious (agitated)	Depressed (retarded)	Cheerful	Fearless
Novelty seeking	Euphoric	Angry	Placid	Stoical
Reward dependence	Sympathetic	Disgusted	Aloof	Indifferent
Persistence	Enthusiastic	Steadfast	Unstable	Discouraged

Table 24-7 Effects of Positive (+) and Negative (–) Reinforcement on Emotional State of Four Temperaments

As can be inferred from Table 24-7, the same external stimulus is likely to elicit responses via activation of multiple temperament dimensions. For example, novel or unfamiliar stimuli elicit interest in approach in proportion to novelty seeking, as well as inhibition of approach in proportion to harm avoidance. Each temperament trait clearly involves an integration of multiple emotional drives that may be conflicting (competitive) or facilitatory (synergistic). Such shared environmental effects mean that the genetic and phenotypic structure of personality cannot be assumed to be the same.

Temperament involves a relatively small set of emotions associated with one's basic needs (e.g., safety), so-called primary motives. Excessive fear and anger, associated with temperament, are motivationally monopolistic and take over the personality by altering perception, learning, and behavior in a biased way. However, under normal circumstances, after survival needs are met, the goals of normally developing personality change to include not only the integrity of the physical self, but also that of the mental self (e.g., self-esteem) as well as numerous social goals and emotions (e.g., shame, pride, empathy). These are called “secondary,” “social,” or “growth” motives. These secondary emotions are closely related etiologically to the development of character. Specifically, basic emotions of fear, anger, disgust, and excitement are transformed to more-complex, predominantly positive secondary emotions, such as love, empathy, compassion, and resourcefulness. This transformation occurs through the interaction of internalized concepts associated with character with basic emotions associated with temperament. Even though some basic character components develop early in life, such as trust and confidence, the completion of self-object differentiation (“me” vs. “not-me”) between 18 months and 3 years of age sets the stage for the full development of character traits and secondary emotions.

The secondary emotions take over as primary motivators of further character development and maturation. They are not as monopolistic motivationally, and they initiate development of more-flexible and adaptable personality traits. As shown in Table 24-8, each of the three character traits is associated with a characteristic pattern of secondary emotions.

Character Dimension	High Scorers		Low Scorers	
	+	-	+	-
Self-directed	Hopeful	Resourceful	Vain	Shameful
Cooperative	Loving	Forgiving	Scornful	Revengeful
Transcendent	Joyful	Peaceful	Greedy	Miserable

Table 24-8 Effects of Positive (+) and Negative (–) Reinforcement on Emotional State of Three Characters

In conclusion, abnormal (deviant, immature) motivation derives from two or three excessive, monopolistic, elementary emotional needs associated with threats to survival and physical integrity. In contrast, mature motivation develops after basic needs are met and the person is freed to experience numerous secondary growth motives. This explains the motivational inflexibility and poverty of deviant personality and accounts for all the rich motivational diversity and flexibility of those with a mature personality.

Character In contrast to temperament, which is mostly inherited, character is less heritable and is moderately influenced by social learning, culture, and random life events unique to the individual. The seven-factor psychobiological model of personality is the only model that has addressed the issue of character and its structure in a systematic way. The psychodynamic concepts of character have created many intriguing clinical formulations on the subject.

Psychobiology of Character Character (or the conceptual core of personality) involves higher cognitive functions regulated by the hippocampus and neocortex (e.g., abstraction and symbolic interpretation, analytical and nductive logic). These functions (also called *propositional memory*) are critical for cognitive processing of sensory percepts and affects regulated by temperament, leading to the development of concepts about the self and the external world. Three major character traits have been distinguished: self-directedness, cooperativeness, and self-transcendence. When fully developed, these traits define mature personality. These character traits are adaptive, but their low ends are less advantageous because of a limited spectrum of circumstances in which immaturity, especially low self-directedness and cooperativeness, means better adaptation than maturity.

Self-directedness quantifies the extent to which an individual is responsible, reliable, resourceful, goal-oriented, and self-confident. The most advantageous summary feature of self-directed individuals is that they are realistic and effective; that is, they are able to adapt their behavior in accord with individually chosen, voluntary goals based on a realistic assessment of facts. Individuals low in self-directedness are blaming, helpless, irresponsible, unreliable, reactive, and unable to define, set, and pursue meaningful internal goals, which is often disadvantageous to the individual.

Cooperativeness quantifies the extent to which individuals consider themselves integral parts of human society. Highly cooperative persons are described as empathetic, tolerant, compassionate, supportive, and principled. These features are advantageous in teamwork and social groups but not for individuals who prefer to live in a solitary manner. Low cooperativeness involves self-absorbed, intolerant, critical, unhelpful, revengeful, and opportunistic behavior in persons who primarily look out for themselves and tend to be inconsiderate of others' rights or feelings.

Self-transcendence quantifies the extent to which individuals consider themselves integral parts of the universe as a whole. Self-transcendent individuals are spiritual, unpretentious, humble, and fulfilled. These traits are adaptively advantageous when people are confronted with suffering, illness, or death, which is inevitable with advancing age. They are disadvantageous in most modern societies where idealism, modesty, and meditative search for meaning might interfere with the acquisition of wealth and power. People who are low in self-transcendence are described as practical, self-conscious, materialistic, and controlling. Such individuals are expected to be well adapted in most Western societies because of their rational objectivity and materialistic success. However, they consistently have difficulty accepting suffering, loss of control, personal and material losses, and death, which lead to adjustment problems, particularly with advancing age. Contrasting sets of descriptors that distinguish the three dimensions of character are summarized in Table 24-9.

Character Dimension	Descriptors of Extreme Variants	
	High	Low
Self-directedness	Responsible	Blaming
	Purposeful	Goal-less
	Resourceful	Passive
	Self-accepting	Wishful
	Disciplined	Undisciplined
Cooperativeness	Tender hearted	Intolerant
	Empathic	Insensitive
	Helpful	Selfish
	Compassionate	Revengeful
	Principled	Opportunistic
Self-transcendent	Imaginative	Conventional
	Intuitive	Logical
	Acquiescent	Doubtful
	Spiritual	Materialistic
	Idealistic	Relativistic

Table 24-9 Descriptors of Individuals Who Score High and Low on the Three Character Dimensions

As shown in Table 24-9, high and low scorers in each character dimension are distinguished by behavior traits that arise from differences in concepts that are each internally consistent but not logically falsifiable. For example, people low in self-transcendence live in the material world, skeptical of whatever they cannot prove objectively and use practically. In contrast for individuals who are highly self-transcendent, the meaning of life goes beyond material things and includes issues that cannot always be known objectively. Each set of beliefs is internally consistent and may be satisfying to different people, but spiritual people cannot prove a materialist is wrong or vice versa.

Character matures in a stepwise manner in incremental shifts from infancy through late adulthood. The timing and rate of transitions between levels of maturity are nonlinear functions of antecedent temperament configurations, systematic cultural biases, and random life events. The developing character traits (i.e., newly internalized concepts about one's self and the external world) optimize adaptation of temperament to the environment by reducing discrepancies between one's emotional needs and norm-favoring social pressures.

In Figure 24-1, character corresponds to the processes of logic, construction, evaluation, and invention of abstract symbols that are based on conceptual representation of information and are well developed only in some mature humans. Recent work shows that psychophysiological markers of neocortical processing, such as the P300 event-related potential and contingent negative variation, are correlated with measures of character but not temperament. More specifically, self-directedness, but not other temperament or character traits, correlates moderately with the evoked potential P300 in parietal leads. Likewise, cooperativeness correlates moderately with contingent negative variation. Individuals with Parkinson's disease differ from others in temperament (they are lower in novelty seeking premorbidly) but not in character. These clinical and empirical observations associate character, but not temperament, to higher cortical functions in the central nervous system (CNS).

Basic differences between temperament and character are summarized in Table 24-10.

Variable Properties	Temperament	Character
Awareness level	Automatic	Intentional
Memory form	Percepts Procedures	Concepts Propositions
Learning principles	Associative Conditioning	Conceptual Insight
Role of subject in mental activity	Passive Reproductive	Active Constructive
Key brain system	Limbic system Striatum	Temporal cortex Hippocampus
Form of mental representation	Stimulus-response sequences varying additively in strength	Interactive networks (conceptual schema) varying qualitatively in configuration
Etiological components		
Genetic heritability	40–60%	15–40%
Family environment	0%	10–15%
Random environment	40–60%	25–35%

Table 24-10 Key Differences between Temperament (Associative or Procedural Learning) and Character (Conceptual or Propositional Learning)

Psychodynamic Concept of Character Psychodynamic thinking about character is based on the concept of *defense mechanisms*, defined as automatic, unconscious psychological processes (both cognitive and emotional) that protect against anxiety generated by intrapsychic conflicts. All defense mechanisms are grouped into three levels: mature, neurotic, immature. Psychodynamic concepts developed by Anna Freud and by Otto Fenichel define *character traits* as residues of previous mature defense mechanisms (e.g., sublimation, anticipation, altruism, and humor) that have become incorporated into everyday behavior patterns. In other words, normal character reflects one's capacity to postpone immediate gratification of internal needs through sublimation, anticipation, altruism, and humor. Behaviors reflecting these particular defense mechanisms correspond to the above description of cooperativeness (especially altruism) and self-transcendence (especially anticipation). Psychodynamic theories have also recognized that the ego not only protects from, but shifts, organizes, and reacts to, internal impulses and external stimuli. The pattern in which one utilizes and combines these integrative functions to satisfy multiple external and internal pressures also constitutes character. Again, ego strength corresponds closely to the description of self-directedness.

Psychodynamic theories describe two forms of character disorders: neurotic character and character neurosis. Neurotic character traits are postulated to derive from neurotic defenses (e.g., reaction formation, projection, repression, undoing) that have dissociated from their original conflict and have become inflexible, pervasive, and ego-syntonic patterns of everyday behavior (“the armor plating” of character). As extensions of previous defensive solutions to conflicts between one's emotional needs and opposing social pressures, neurotic character traits protect a person from being involved in such situations again. For example, hypermorality is a “reaction formation” against aggression that protects obsessive persons from consciously experiencing aggression and thus from intrapsychic conflicts. Character neurosis is observed when an inflexible neurotic character trait interferes with healthy parts of personality. For example, when obsessive hypermorality interferes with the need for social contacts generated in healthy parts of obsessive personality and, instead of being perceived as natural, is perceived as frustrating, one observes character neurosis. Character neurosis has been included in the 10th edition of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) but not in DSM IV.

Individual differences in defense mechanisms (i.e., normal, neurotic, and immature) appear to covary with the severity of psychopathology (i.e., normality, anxiety disorders, personality disorder, respectively). Interestingly, patients with personality disorder tend to “favor” certain immature defense mechanisms, which in turn shape their behavioral style and clinical presentation. These “favorite” defenses are, for example, projection for those with paranoid personality disorder, acting out for those with antisocial personality disorder, and fantasy for patients with schizoid personality disorder. Note, however, that in addition to these favorite immature defenses, persons with personality disorders also use other immature defenses that contribute to their irresponsibility, tendency to blame others, and so forth. Reflecting these and similar observations about the importance of defense mechanisms for the overall understanding of deviant personality, the DSM-IV classification now enables clinicians to classify the use of a particular, dominant defense mechanism on Axis II.

Descriptively, psychodynamic and psychobiological perceptions of character are obviously similar. Conceptually, both approaches underscore the adaptive function of character traits in the overall adaptation of emotional needs to the environment. However, the two concepts differ in other etiological and classificatory aspects of character. The psychobiological model explains individual differences in character as quantitative variation on the three character traits. In contrast to this dimensional approach, the psychodynamic concept distinguishes categorically between character subtypes (e.g., neurotic character is further divided into anal, oral, phallic) and discrete character disorders (i.e., neurotic characters and character neurosis). Some dimensionality is preserved, since most people use all defense mechanisms, and it is the relative predominance of immature versus mature defenses that classifies them into personality disorders (i.e., the transition between character types is gradual, not abrupt). The psychobiological model underlines the preconscious or conscious nature of character, whereas the psychodynamic concept understands character traits, (with the possible exception of anticipation) as predominantly unconscious processes. Moreover, the psychodynamic concept does not provide guidelines for further etiological studies of character. However, this concept is very important in treatment, especially for the revision of emotional and conceptual biases inherent in immature defenses and deviant character traits.

Integrated View of Personality Recent reports about complex and fundamental processes underlying neurophysiology, development, and phenomenology of personality provide powerful guidelines for the formulation of a comprehensive psychobiological model of personality. The model integrates earlier phenotypic observations about behavior with contemporary concepts based on neuroimaging, biology of behavior, genetics, psychometric advances, and nonlinear modeling of normal and deviant personality development. Personality is conceptualized as a complex adaptive system involving a bidirectional interaction between temperament and character once self-concepts emerge with the maturation of the conceptual or propositional learning system (Fig. 24-2).

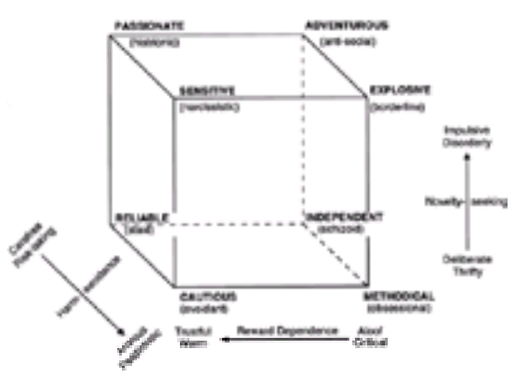

FIGURE 24-2 Temperament configurations.

Through the interaction of temperament and character, different aspects of internalized concepts of the self and the external world modify the significance and the salience of sensory percepts and affects regulated by temperament and vice versa. In other words, temperament regulates what one notices, and in turn, character modifies its meaning, so that the salience and significance of all experience depends on both one's temperament and one's character.

CONCEPTUAL AND PSYCHOMETRIC ISSUES

Personality Traits: Person Versus Situation Debate Personality traits are neuropsychic structures with the capacity to render many stimuli functionally equivalent and to initiate and guide equivalent (consistent) forms of adaptive and expressive behavior. Accordingly, the DSM defines *personality traits* as “enduring patterns of perceiving, relating to, and thinking about oneself and the environment” (i.e., other people and the world as a whole). The major value of personality traits, therefore, lies in their usefulness in identifying predictable regularities in an individual's behavior.

The stability of personality traits in time and across situations has been the central issue in personality theory for decades, as it is directly relevant to whether internal dispositions or external situations determine behavior. The “person versus situation” debate was initiated in the early 1900s by two groups of scientists (“situationists”

versus “personologists”). The former viewed behavior as highly situationally specific, the latter as centrally organized and purposive. The dilemma remained unresolved for decades because it was not realized that single items of behavior have limited reliability and generality. They must be aggregated over situations and occasions to reveal stable traits that allow predictions of behavioral tendencies but, less accurately, prediction of single behavioral acts.

On the other hand, global personality traits also do not explain behavior completely because people do not exhibit unmodulated consistencies in behavior across time and situations. Complete invariance in behavior is associated more with psychopathology than with normality. The current understanding of traits differs from the conservative one of 20 years ago. The emerging consensus is an interactionist position allowing room for both situational (external) and dispositional (internal) determinants of behavior in a variety of complex combinations.

Phenotypic and Developmental Trait Personality Models The major difficulty in relating observed variation in personality to its underlying biological processes is that the observed phenotypic structure differs from the underlying biogenetic structure because learning and environmental factors also influence phenotypic variation. Phenotypic traits do not explain behavior but themselves require etiological explanation. Factor analysis can determine only the minimal number of distinguishing behavior traits, not their underlying causal structure. Hence, the phenotypic trait personality models, which are usually factor analytically derived (e.g., Hans Eysenck's tridimensional model, Paul Costa and Robert McCrae's five factor model), account for the much of the observed phenotypic variance but may not correspond to the underlying biogenetic processes. For example, sociability and impulsivity, phenotypic behavioral traits described by Heinz Eysenck, have independent genetic antecedents; still, Eysenck regards them as components of a single phenotypic trait—extraversion. Likewise, the five factor model, as measured by the Neuroticism Openness Extraversion (NEO) personality inventory, defines neuroticism, extraversion, openness, agreeableness, and conscientiousness as basic personality traits. Extensive heterogeneity within and among these traits has been suggested (e.g., benzodiazepines increase extraversion and decrease neuroticism, suggesting common biological background for these two phenotypic traits). In addition an exonic polymorphism of *DRD*4 that accounts for a portion of the genetic variance in novelty seeking is correlated with 3 of 6 facets of extraversion and 1 of 6 facets of conscientiousness in the five factor model. This indicates that extraversion and neuroticism are heterogeneous composites of the underlying genetic factors. The dopamine transporter is associated with novelty seeking but with none of the five factors of the NEO model.

In contrast, models of personality as developmental traits take into account both the underlying biological dispositions to observable behaviors and individual differences in responses to experience during personality development. The most recognized dimensional models are those of Henrik Sjobring, Marvin Zuckermann, and Robert Cloninger. Despite some limitations, Sjobring's developmental model in the late 1940s was a heuristic attempt to take personality research beyond the level of phenotypic observations and factor analysis. The model defines three dimensions of latent variation predisposing to personality: “validity” (the degree of energy available in adaptation to experience), “solidity” (the consistency of adaptation), and “stability” (referring to the maximum potential of a person to develop skillful habits in adapting to routine). These three dispositions were postulated to be causally independent from one another and from the factor of general intelligence (referred to as “capacity”) and also to define one's susceptibility to behavior disorders.

Personality has often been described in terms of a linear sequence of qualitatively discrete developmental stages and structural types. The frequent occurrence of heterotypic individuals with intermediate or mixed features calls into question the assumption that personality structures are limited to a finite number of discrete and homogeneous types. Likewise, there is not only variation in the rate and direction of personality development as a whole, but also asynchrony and inconsistency among its major components that are reorganized through time. The following section outlines a nonlinear quantitative theory of personality development that does not make restrictive assumptions about a finite number of qualitatively discrete structural types.

PERSONALITY DEVELOPMENT: FUNDAMENTALS OF A SELF-ORGANIZING PSYCHOBIOLOGICAL COMPLEX

Nenad Svrakic and coworkers in 1996 formulated a quantitative model of normal and deviant personality development as a complex dynamic system that is self-organizing and partly molded by familial and sociocultural influences. The model allows for nonlinear interactions among etiologically distinct components of personality. It also accounts for the frequent, but not invariant, development of stagelike periods of moderately stable personality configurations (similar to personality types) punctuated by abrupt transitions with structural reorganization and emergence of qualitatively new features. The model is based on a sophisticated mathematical framework that can be implemented and tested with readily accessible data.

The central ideas behind this model are the following. First, Personality, as a dynamic multidimensional system, comprises more-elementary operating components (traits) organized in an interdependent way that is critical to carrying out a particular function. Such a system is characterized by particular rules of operation (e.g., the principles of associative learning within the temperament function or propositional learning within the character function). Satisfying such multiple constraints results in nonlinear dynamics, which is characteristic of all systems involving growth and development in biology, neuroscience, psychology, and sociology. Such multidimensional dynamic systems are usually called *complex adaptive systems*. Second, the correlations observed among multiple personality traits can be used to explain the spontaneous organization of stable multidimensional configurations (i.e., personality types); these types develop in a stagelike fashion, with successive periods of prolonged stability punctuated by rapid transitions. Third, the model accounts for functional interactions among multiple types of influence, including genetic factors, family environment, sociocultural norms, and random events unique to each individual. Based on the specified basic differences between temperament and character (Table 24-10), this model takes into account all aspects of information processing from associative conditioning of habits and skills to enculturation about goals and values.

Personality development is presented as a walk on an adaptive (or fitness) landscape with two or more hills (representing high adaptive values) separated by valleys (representing low adaptive values). *Fitness* is defined as the ability to produce change in personality. In general, a complex system subjected to constraints responds to these constraints by optimizing fitness, that is, by adaptive changes in personality. As personality change is motivated by optimization of fitness, individuals most likely and most rapidly move toward the greatest increase in adaptive value (i.e., the nearest hill). Once they have reached the peak of the hill, they stay there for a relatively long time because they would first have to decrease fitness (i.e., descend into the valley) to find a hill with a higher peak (i.e., better adaptation). Such adaptive development is called *U-shaped*.

Temperament and character traits are etiologically distinct but functionally related. Such an interactive system is self-organizing because of the collective dynamics among its multiple components. In other words, the organism is spontaneously driven to find patterns of behavior that satisfy all internal and external constraints. For example, harm avoidance interacts with both novelty seeking and reward dependence, inhibiting approach to novel stimuli and inhibiting social attachments by increasing fear of the unfamiliar and sensitivity to social criticism. Accordingly, the authors have found weak, but consistent, negative correlations of harm avoidance with novelty seeking and reward dependence. In contrast, novelty seeking and reward dependence are positively synergistic, facilitating sociability and seeking of social approval. Such interactions influence the stability of each of the eight possible multidimensional configurations or profiles of temperament traits (Fig. 24-3) and make some more stable than others.

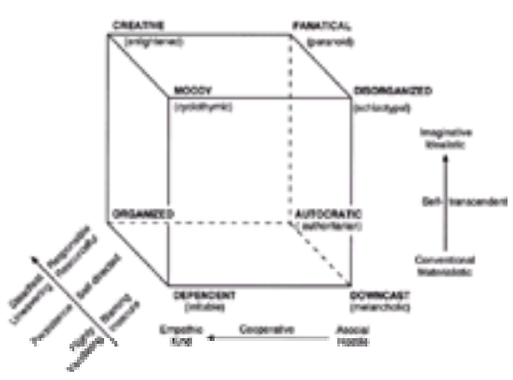

FIGURE 24-3 Character configurations.

In the fitness landscape, the valleys (or adaptive minima) are unstable because even such small perturbations as random events will drive the system away from these points of low fitness. In contrast, hills (i.e., adaptive maxima) are stable because they act as “attractors” for all lower points in the neighborhood. Consequently, personality interactions tend to self-organize into an attractor state and to remain in that configuration in the absence of external pressure or maturational processes.

Understanding both the chronicity and the treatment resistance of personality disorders requires realizing that a particular personality configuration may be highly stable, even though it may not be the most adaptive behavior possible for that individual.

The authors are mainly concerned with patterns of change of character traits as a function of relatively stable temperament traits and various other influences discussed below. The three character traits interact to produce eight possible character configurations or profiles (Fig. 24-4).

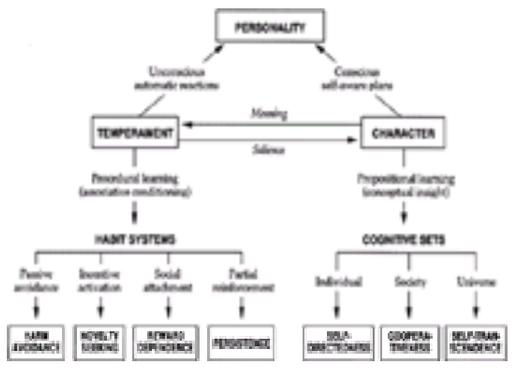

FIGURE 24-4 Personality: An integrated psychobiological schema.

Each of these character configurations can be predicted from the initial temperament configuration, taking into account temperament-character interactions and the effects of sociocultural norms, random events unique to the individual, social learning within the family, and genetic factors. All these sources of individual differences influence the likelihood for different stable character configurations in the outcome. The sociocultural norms can be additionally accounted for as the systematic bias in development associated with demographic variables such as subculture, gender, race, age, and occupation. These norms and social education in the family create environments that support or oppress specific character development. With all this taken into account, a single initial temperament configuration may lead to several different stable character configurations, an aspect of development referred to as *multifinality*. These results are graphically represented in Figure 24-5, Figure 24-6 and Figure 24-7, with four different initial temperament configuration and their possible character outcomes.

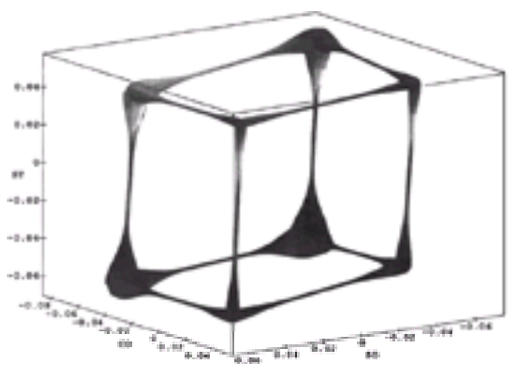

FIGURE 24-5 Most-probable character outcomes for the adventurous antecedent temperament traits.

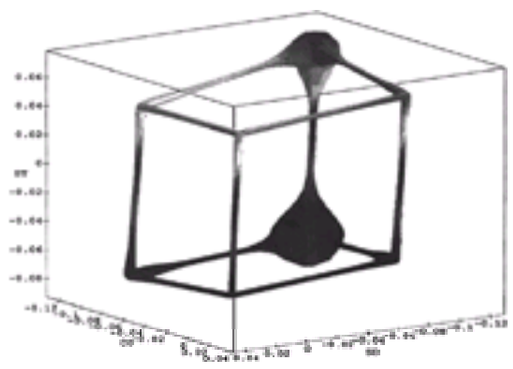

FIGURE 24-6 Most-probable character outcomes for the explosive (borderline) antecedent temperament traits.

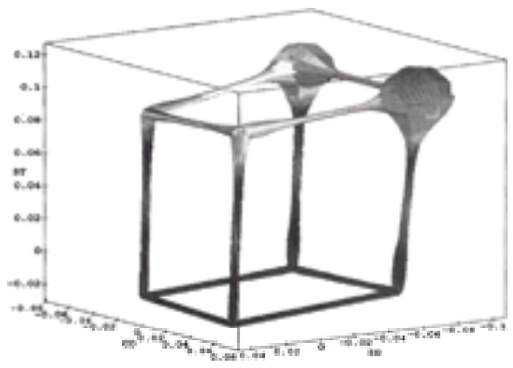

FIGURE 24-7 Most-probable character outcomes for the sensitive antecedent temperament traits.

Three dimensions of character (self-directedness [SD], cooperativeness [CO], and self-transcendence [ST]) are drawn along three orthogonal axes, and each character configuration is represented by a point in this abstract space. Eight possible combinations of high versus low scores on the three character dimensions are represented by the corners of the cube. The size of the bulging at a particular character configuration is proportional to the probability of that outcome. These figures show that the same temperament type leads to several possible stable character outcomes, each with certain probability, a feature characteristic of nonlinear systems. Figures 20-5 through 20-7 illustrate such multifinality for the adventurous (high NS, low HA, low RD), explosive (high NS, high HA, low RD), and sensitive (high NS, high HA, high RD) temperament types. Similar results can be obtained for the remaining five temperament configurations.

Individuals are born with one temperament profile, which has several possible character outcomes. This multifinality reflects the fact that the fitness landscape in personality development has both objective and subjective components. Feelings and internalized concepts unique to the individual do influence one's evaluation of the “objectively” worthwhile goals and values. The observed discrepancies between the natural directions influenced by antecedent temperament traits and actually

achieved character configurations point to the importance of external events such as social learning and random fluctuations in the final character outcome.

Because points of high and low adaptive values exist in the fitness landscape, personality development is characterized by periods of relative stability alternating with more-rapid transitions to new adaptive levels. The subscale structure of the Temperament and Character Inventory (TCI) was formulated to specify character in terms of 15 component steps of its development, as shown in Table 24-11.

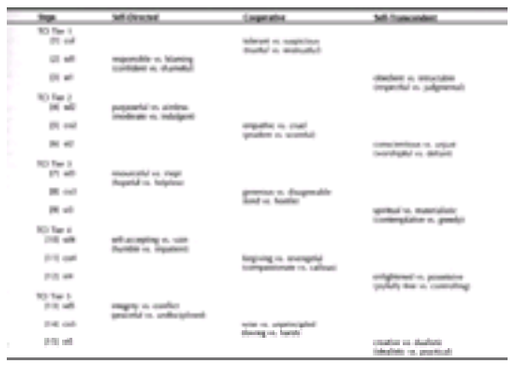

Table 24-11 15-Step Personality Development

The 15 steps in character development are a theoretical ideal, corresponding to the modal pathway that leads to full character development with high scores on all three character dimensions, which is not optimal for everybody. Observe that the overall character developmental line describes a spiral, in which each revolution around the spiral (presented as each of the five TCI tiers in Table 24-11) introduces a new set of developmental issues associated with the component facets of the major character traits. For example, during the first revolution around the spiral, a person encounters problems of trust versus mistrust (cooperativeness, facet 1: social acceptance), confidence versus shame (self-directedness, facet 1: responsibility), and obedience versus rebelliousness (self-transcendence, facet 1: self-forgetfulness). As can be inferred from Table 24-5, as individuals move along the spiral, they successively face new developmental tasks associated with the component facets in self-directedness, cooperativeness, and self-transcendence. This spiral pattern of character development provides an opportunity to correct developmental errors that were made at any earlier step within the same trait. For example, as character develops around the spiral to reach a new step of self-directedness (e.g., resourcefulness), one gets “in line” historically with issues of purposefulness and responsibility (steps of self-directedness encountered during earlier revolutions) (Table 24-11). This alignment facilitates retrospective revisions of errors at earlier developmental levels. The possibility of retrospective healing of old character errors has implications for psychotherapy planning, specifically, simultaneous intervention at different levels along the same trait. However, such retrospective healing does not always occur, so character development cannot be fully described by any simple linear schema. This 15-step developmental sequence is consistent with prior qualitative descriptions of developmental stages by Jean Piaget, Sigmund Freud, and Erik Erikson but allows more complete quantitative specification (Table 24-12).

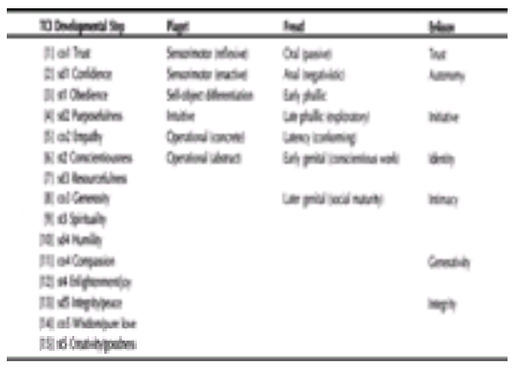

Table 24-12 Different Descriptions of Personality Development

Computer-simulated self-organized development of character, beginning with average initial temperament traits and taking into account all of the above contributing factors, demonstrated that children first increase in self-transcendence (i.e., become imaginative and enjoy fantasy), then increase in cooperativeness (i.e., become conforming and rule based), and only later increase in self-directedness (i.e., behavior becomes increasingly self-reliant and autonomous). This prediction is consistent with the description of the stages of ego-development by Jane Loevinger and her colleagues. This early sequence does not exclude subsequent further development of other character dimensions in response to demands of changing social roles with age. This subsequent development in response to external social pressures may explain adult self-actualization and moral development as described by Erikson, Lawrence Kohlberg, and others.

PERSONALITY DISORDER

Normal Versus Deviant Personality Normal personality is usually defined (1) directly, using criteria of health ideals; (2) indirectly, as the opposite to deviant personality; or most frequently (3) statistically, by behaviors most common in the given environment. However, the distinction between “normal” and “abnormal” personality is inherently relative as it relies on a largely arbitrary cutoff point on the continuum between two extremes of a behavior trait. This distinction is also context dependent, as the same behavior manifested in different situations could be viewed as normal or maladaptive (e.g., invariant fearfulness when danger is unlikely versus when it is likely). However, attempts to define *abnormal personality* based solely on the social context are not satisfactory because personality disorders involve many noninterpersonal traits as well (e.g., narcissistic persons satisfy many aspects of their grandiosity internally, in fantasy). Furthermore, personal deviance alone does not reliably distinguish between normal behavior and personality disorders (e.g., some individuals are very socially withdrawn without impairment in professional functioning or signs of personal suffering and distress). In other words, both personal and social aspects are needed to account fully for the symptoms of personality disorder.

Classification of Personality Disorders The leading classifications of personality disorders are the International Classification of Diseases (ICD) of the World Health Organization, and DSM of the American Psychiatric Association. The ICD classification of personality disorders, ninth revision of *International Statistical Classification of Diseases* (ICD-9) relied extensively on Kurt Schneider's book *Psychopathic Personalities* and his description of 10 discrete, socially deviant personality types. ICD-10, published in 1987, corresponds more closely to DSM, however. Anglo-American concepts of personality disorder originate in James Pritchard's description of moral insanity, which was later termed “sociopathy” and finally became antisocial personality disorder in the second edition of DSM (DSM-II) in 1968. However, as John Gunderson and Katharine Phillips showed in the previous edition of this textbook, other classified personality disorders in the DSM system can be traced to Kurt Schneider's work. This chapter uses the DSM-IV classification of personality disorders, which is descriptively similar to that of ICD-10. A noteworthy exception to this is that ICD-10 classifies schizotypal personality disorder on Axis I, among schizophrenic disorders, whereas DSM-IV keeps this disorder on Axis II among personality disorders.

DSM-IV Classification According to DSM-IV, the crucial criterion for distinguishing deviant personality traits is the presence (evidence) of long-term maladaptation and inflexibility, manifested as subjective distress, or socioprofessional functional impairment, or both. DSM-IV defines personality disorders as:

> an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture. The pattern is manifested in two (or more) of the following areas:
>
> 1. cognition (i.e., ways of perceiving and interpreting self, other people, and events)

2. affectivity (i.e., the range, intensity, lability, and appropriateness of emotional response)
3. interpersonal functioning
4. impulse control

The pattern is stable and of long duration and its onset can be traced back at least to adolescence or early adulthood. It is inflexible and pervasive across a broad range of personal and social situations and leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Personality disorder subtypes classified in DSM-IV are schizotypal, schizoid, paranoid, narcissistic, borderline, antisocial, histrionic, obsessive-compulsive, dependent, and avoidant. In addition to these 10 standard disorders, DSM-IV classifies two disorders, passive-aggressive and depressive among “Criteria Sets and Axes Provided for Further Study” in Appendix B.

DSM-IV arranges categorical personality disorders into three clusters, each sharing some clinical features: Cluster A includes three disorders with odd, aloof, and eccentric features (paranoid, schizoid, and schizotypal); Cluster B includes four disorders with dramatic, impulsive, and erratic features (borderline, antisocial, narcissistic, and histrionic); and Cluster C includes three disorders sharing anxious and fearful features (avoidant, dependent, and obsessive-compulsive). Several studies have supported the construct validity of these clusters except that the symptoms for compulsive disorder sometimes tend to form a fourth cluster. Note that the three dimensions underlying Clusters A, B, and C (i.e., detachment, impulsivity, and fearfulness) correspond closely to normal temperament traits (reward dependence, novelty seeking, and harm avoidance, respectively), suggesting that variation in these temperament traits might be significant in distinguishing among the three clusters of disorders.

ICD-10 In ICD-10, personality disorders are described as severe disturbance of personality and behavior that are pronounced deviations, from normal cultural patterns. ICD-10's diagnostic guidelines include disturbances of long-standing duration in several areas of functioning; pervasive and maladaptive behavior; onset in childhood or adolescence; continuation into adulthood; considerable personality distress (although sometimes apparent only late in the disorder's course); and usually, but not always, significant problems in work and in social behavior. ICD-10 also allows for the possibility of criteria developed to describe personality disorders in different cultures. The diagnostic criteria for specific personality disorders appear in Table 24-13.

Table 24-13 ICD-10 Diagnostic Criteria for Specific Personality Disorders

Other related disorders in ICD-10 are mixed and other personality disorders (Table 24-14); enduring personality changes not attributed to brain damage and disease (Table 24-15); other disorders of adult personality and behavior (Table 24-16); and a residual category, unspecified disorder of adult personality and behavior.

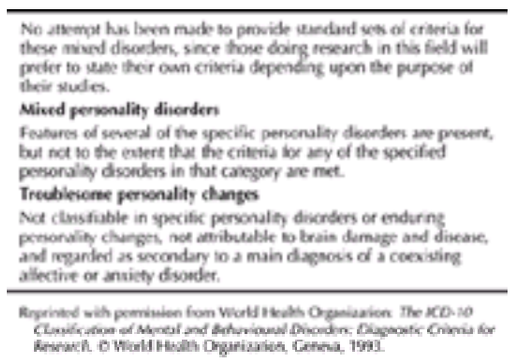

No attempt has been made to provide standard sets of criteria for these mixed disorders, since those doing research in this field will prefer to state their own criteria depending upon the purpose of their studies.

Mixed personality disorders

Features of several of the specific personality disorders are present, but not to the extent that the criteria for any of the specified personality disorders in that category are met.

Troublesome personality changes

Not classifiable in specific personality disorders or enduring personality changes, not attributable to brain damage and disease, and regarded as secondary to a main diagnosis of a coexisting affective or anxiety disorder.

Reprinted with permission from World Health Organization. *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 24-14 ICD-10 Diagnostic Criteria for Mixed and Other Personality Disorders

Table 24-15 ICD-10 Diagnostic Criteria for Enduring Personality Changes, Not Attributable to Brain Damage and Disease

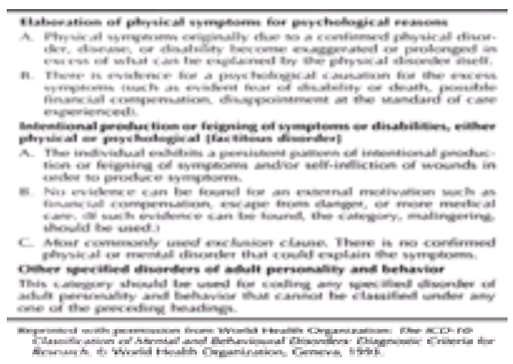

Elaboration of physical symptoms for psychological reasons

A. Physical symptoms originally due to a confirmed physical disorder, disease, or disability become exaggerated or prolonged in excess of what can be explained by the physical disorder itself.

B. There is evidence for a psychological causation for the excess symptoms (such as evident fear of disability or death, possible financial compensation, disappointment at the standard of care experienced).

Intentional production or feigning of symptoms or disabilities, either physical or psychological (factitious disorder)

A. The individual exhibits a *persistent pattern* of intentional production or feigning of symptoms and/or self-infliction of wounds in order to produce symptoms.

B. No evidence can be found for an external motivation such as financial compensation, escape from danger, or more medical care. (If such evidence can be found, the category, malingering, should be used.)

C. *Most commonly used exclusion clause.* There is no confirmed physical or mental disorder that could explain the symptoms.

Other specified disorders of adult personality and behavior

This category should be used for coding *any* specified disorder of adult personality and behavior that cannot be classified under any one of the preceding headings.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 24-16 ICD-10 Diagnostic Criteria for Other Disorders of Adult Personality and Behavior

DSM-IV PERSONALITY DISORDERS

Qualitative features considered diagnostic of a personality disorder in DSM-IV are summarized in Table 24-17.

Cluster	Subtype	Discriminating Features
Odd/eccentric	Schizoid	Socially indifferent
	Paranoid	Suspicious
	Schizotypal	Eccentric
Erratic/impulsive	Antisocial	Disagree
	Borderline	Unstable
	Histrionic	Attention seeking
	Narcissistic	Self-centered
Anxious/fearful	Avoidant	Inhibited
	Dependent	Submissive
	Obsessive	Perfectionistic
Not otherwise specified	Passive-aggressive	Negativistic
	Depressive	Pessimistic

Data from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 24-17 DSM-IV Qualitative Clusters and Subtypes of Personality Disorders

General Diagnostic Guidelines DSM-IV introduces the following general guidelines for the diagnosis of personality disorder (Table 24-18). Many of the features of various personality disorders may also be seen during an episode of another mental disorder. The diagnosis of personality disorder is made only when the features are typical of long-term functioning and are not limited to a discrete episode of another mental disorder. Likewise, when maladaptive behavior features are due to the direct psychological effects of a substance (e.g., various psychoactive substances including medication), general medical condition, or catastrophic experience, diagnosis of personality disorder is not warranted.

A. An enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's culture. This pattern is manifested in two (or more) of the following areas:
   (1) cognition (i.e., ways of perceiving and interpreting self, other people, and events)
   (2) affectivity (i.e., the range, intensity, lability, and appropriateness of emotional response)
   (3) interpersonal functioning
   (4) impulse control
B. The enduring pattern is inflexible and pervasive across a broad range of personal and social situations.
C. The enduring pattern leads to clinically significant distress or impairment in social, occupational, or other important areas of functioning.
D. The pattern is stable and of long duration and its onset can be traced back at least to adolescence or early adulthood.
E. The enduring pattern is not better accounted for as a manifestation or consequence of another mental disorder.
F. The enduring pattern is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition (e.g., head trauma).

Reprinted with permission from *DSM-IV: Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 24-18 DSM-IV General Diagnostic Criteria for a Personality Disorder

Judgments about personality functioning must take into account an individual's ethical, social, and cultural background. Personality disorder should not be confused with acculturation following immigration or with the expression of customs and religious and political values characteristic of a person's culture of origin.

Diagnosis of specific personality disorder may be made in children or adolescents when observed maladaptive personality traits are pervasive, persistent, and unlikely to be limited to a particular developmental stage or an episode of an Axis I disorder. Diagnosis of a personality disorder in an individual under 18 years of age requires that the features be present for more than 1 year. The only exception to this is antisocial personality disorder, which cannot be diagnosed in individuals under 18 years of age.

Clinical experience points to a potential sex bias in diagnosing personality disorders. Certain personality disorders are diagnosed more frequently in men (e.g., antisocial and schizoid), whereas some disorders are more frequently diagnosed in women (e.g., borderline, histrionic, and dependent). Even though real gender differences likely exist in the prevalence of these disorders, clinicians are cautioned not to diagnose certain personality disorders in males and females because of social stereotypes about typical gender roles and behaviors.

Observed maladaptive behavior traits must be pervasive, that is, manifest in a wide range of personal and social contexts (e.g., at home, at work, with family and friends), not isolated aspects of the person's life. Data collection from collateral informants is thus considered critical to ensure high-quality personality assessment, diagnostic reliability, and validity.

In addition to designated diagnostic criteria, DSM-IV specifies a group of “associated features” that are given as loose descriptions of optional, but fairly frequent, behaviors, intended to help clinicians when diagnosis is not certain.

Cluster A Personality Disorders

Paranoid Personality Disorder The hallmarks of paranoid personality disorder are excessive suspiciousness and distrust of others expressed as a pervasive tendency to interpret actions of others as deliberately demeaning, malevolent, threatening, exploiting, or deceiving (Table 24-19). Frequently impairment is mild, but the disorder typically includes occupational and social difficulties.

A. A pervasive distrust and suspiciousness of others such that their motives are interpreted as malevolent, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following:
   (1) suspects, without sufficient basis, that others are exploiting, harming, or deceiving him or her
   (2) is preoccupied with unjustified doubts about the loyalty or trustworthiness of friends or associates
   (3) is reluctant to confide in others because of unwarranted fear that the information will be used maliciously against him or her
   (4) reads hidden, demeaning, or threatening meanings into benign remarks or events
   (5) persistently bears grudges, i.e., is unforgiving of insults, injuries, or slights
   (6) perceives attacks on his or her character or reputation that are not apparent to others and is quick to react angrily or to counterattack
   (7) has recurrent suspicions, without justification, regarding fidelity of spouse or sexual partner
B. Does not occur exclusively during the course of schizophrenia, a mood disorder with psychotic features, or another psychotic disorder and is not due to the direct physiological effects of a general medical condition

Note: *If criteria are met prior to the onset of schizophrenia add "premorbid," e.g., "paranoid personality disorder (premorbid)."*

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 24-19 DSM-IV Diagnostic Criteria for Paranoid Personality Disorder

COMORBIDITY These patients are at increased risk for major depressive disorder, obsessive-compulsive disorder, agoraphobia, and substance abuse or dependence. The most common co-occurring personality disorders are schizotypal, schizoid, narcissistic, avoidant, and borderline personality disorders. The disorder may be complicated by brief psychotic disorder, particularly in response to stress. Paranoid personality disorder has been postulated to be a prepsychotic antecedent of delusional disorder, paranoid type.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 0.5 to 2.5 percent in the general population, 10 to 30 percent for psychiatric inpatients, and 2 to 10 percent for psychiatric outpatients. According to DSM-IV, this disorder is more commonly diagnosed in males.

ETIOLOGY Some studies have demonstrated increased prevalence of this personality disorder among relatives of probands with chronic schizophrenia and delusional disorder, paranoid type.

DIFFERENTIAL DIAGNOSIS Paranoid personality disorder is distinguished from schizophrenia (especially paranoid type), delusional disorder, paranoid type, and mood disorder with psychotic features on the basis of periods with positive psychotic symptoms such as delusions and hallucinations in the latter. When a brief reactive psychosis with delusions complicates the clinical picture of paranoid personality disorder, this distinction is far more difficult. The duration of the latter and its frequent association with stress usually suffices for differential diagnosis.

Paranoid personality disorder is sometimes difficult to distinguish from the following personality disorders: (1) schizotypal personality disorder (which includes magical thinking, unusual perceptual experiences, oddities in speech, appearance, and thought processes), (2) obsessive-compulsive, schizoid, borderline, and histrionic personality disorders (all with no prominent paranoid ideation), (3) avoidant personality disorder (which includes fear of embarrassment), (4) antisocial personality disorder (which includes personal gains from antisocial behavior), and (5) narcissistic personality disorder (which includes fear of having their “hidden” imperfections and flaws revealed).

Schizoid Personality Disorder The hallmarks of schizoid personality disorder are a pervasive pattern of social detachment and a restricted range of expressed emotions in interpersonal settings (Table 24-20). Frequently these individuals exhibit severe problems in social relations and occupational problems when interpersonal involvement is required. Social isolation sometimes favorably affects overall performance.

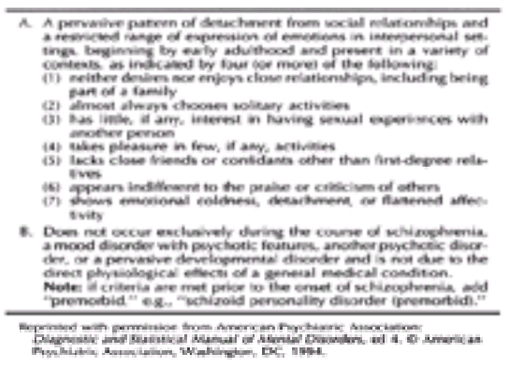

A. A pervasive pattern of detachment from social relationships and a restricted range of expression of emotions in interpersonal settings, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following:
(1) neither desires nor enjoys close relationships, including being part of a family
(2) almost always chooses solitary activities
(3) has little, if any, interest in having sexual experiences with another person
(4) takes pleasure in few, if any, activities
(5) lacks close friends or confidants other than first-degree relatives
(6) appears indifferent to the praise or criticism of others
(7) shows emotional coldness, detachment, or flattened affectivity

B. Does not occur exclusively during the course of schizophrenia, a mood disorder with psychotic features, another psychotic disorder, or a pervasive developmental disorder and is not due to the direct physiological effects of a general medical condition.
Note: If criteria are met prior to the onset of schizophrenia, add "premorbid," e.g., "schizoid personality disorder (premorbid)."

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 24-20 DSM-IV Diagnostic Criteria for Schizoid Personality Disorder

COMORBIDITY This personality disorder sometimes appears as the prepsychotic antecedent of delusional disorder, schizophrenia, or (rarely) major depressive disorder. These individuals may exhibit brief psychotic disorder, particularly in response to stress. The most common co-occurring personality disorders are paranoid, schizotypal, and avoidant personality disorders.

EPIDEMIOLOGY Reported prevalence rates vary from “uncommon” (DSM-IV) to 7.5 percent in the general population. According to DSM-IV, this disorder is more commonly diagnosed in males and may cause more impairment in them.

ETIOLOGY An increased prevalence among the relatives of probands with schizophrenia or schizotypal personality disorder has been reported.

DIFFERENTIAL DIAGNOSIS Schizoid personality disorder is distinguished from schizophrenia, delusional disorder, and mood disorder with psychotic features on the basis of periods with positive psychotic symptoms, such as delusions and hallucinations, in the latter. When brief psychotic disorder complicates the clinical picture of schizoid personality disorder, this distinction is far more difficult. The duration of the latter and its frequent association with stress usually suffices for differential diagnosis.

Schizoid personality disorder is distinguished from autistic disorder and Asperger's disorder by more severely impaired social interactions and stereotypic behaviors and interests in the latter two disorders. Schizoid disorder is distinguished from the following personality disorders: schizotypal personality disorders (which include magical thinking, unusual perceptual experiences, oddities in speech, appearance, and thought processes), avoidant personality disorder (adequate emotionality in the latter, also social isolation because of fear of embarrassment, not indifference); obsessive-compulsive personality disorder (adequate capacity for intimacy, despite sometimes excessive isolation due to perfectionism and workaholic attitudes); and paranoid personality disorder (which includes suspiciousness, ideas of reference, and guarded facade).

Schizotypal Personality Disorder The hallmarks of schizoid personality disorder are pervasive discomfort with and reduced capacity for close relationships, as well as cognitive and perceptual distortions and eccentric behavior (not severe enough to meet criteria for schizophrenia) (Table 24-21). Patients with this disorder typically experience occupational and social difficulties.

A. A pervasive pattern of social and interpersonal deficits marked by acute discomfort with, and reduced capacity for, close relationships as well as by cognitive or perceptual distortions and eccentricities of behavior, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:
(1) ideas of reference (excluding delusions of reference)
(2) odd beliefs or magical thinking that influences behavior and are inconsistent with subcultural norms (e.g., superstitiousness, belief in clairvoyance, telepathy, or "sixth sense"; in children and adolescents, bizarre fantasies or preoccupations)
(3) unusual perceptual experiences, including bodily illusions
(4) odd thinking and speech (e.g., vague, circumstantial, metaphorical, overelaborate, or stereotyped)
(5) suspiciousness or paranoid ideation
(6) inappropriate or constricted affect
(7) behavior or appearance that is odd, eccentric, or peculiar
(8) lack of close friends or confidants other than first-degree relatives
(9) excessive social anxiety that does not diminish with familiarity and tends to be associated with paranoid fears rather than negative judgments about self

B. Does not occur exclusively during the course of schizophrenia, a mood disorder with psychotic features, another psychotic disorder, or a pervasive developmental disorder.
Note: If criteria are met prior to the onset of schizophrenia, add "premorbid," e.g., "schizotypal personality disorder (premorbid)."

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 24-21 DSM-IV Diagnostic Criteria for Schizotypal Personality Disorder

COMORBIDITY Transient psychotic episodes can complicate this disorder, particularly in response to stress. Symptoms sometimes become so significant that subjects may meet criteria for schizophreniform disorder, delusional disorder, or brief psychotic disorder. More than a half of these patients have had at least one episode of major depression, and 30 to 50 percent have major depression concurrent with this personality disorder. The most common co-occurring personality disorders are

schizoid, paranoid, avoidant, and borderline.

EPIDEMIOLOGY DSM-IV reports a prevalence rate of 3 percent in the general population. Earlier reports suggested a range between 2 and 6 percent. The sex ratio is unknown. The disorder is frequently diagnosed in females with fragile X syndrome.

ETIOLOGY Empirical findings about this disorder are difficult to interpret because it appears to have several clinical subtypes with potentially differential relation to the schizophrenic spectrum. Prevalence of this personality disorder is increased among the first-degree relatives of probands with schizophrenia, and the prevalence of schizophrenia and other psychoses is increased in the relatives of probands with schizotypal personality disorder. Adoption, family, and twin studies demonstrated an increased prevalence of schizotypal features in the families of schizophrenia patients, especially when schizotypal features were not associated with comorbid affective symptoms. Small anticipatory saccades that disrupt smooth pursuit eye movement (hypothesized to be a marker of genetic vulnerability to schizophrenia), have also been detected in schizotypal and introverted personalities. The relation of this personality disorder and schizophrenia is still controversial; some genetic studies find no relation, and some studies find a gradation in multifactorial liability from schizotypal personality (mild), to broad schizophrenia (moderate) to narrow schizophrenia (severe).

DIFFERENTIAL DIAGNOSIS Schizotypal personality disorder is distinguished from schizophrenia, delusional disorder, and mood disorder with psychotic features on the basis of periods with positive psychotic symptoms such as delusions and hallucinations in the latter. When brief psychotic disorder with delusions complicates the clinical picture of schizotypal personality disorder this distinction is far more difficult.

Schizotypal personality disorder is difficult to distinguish from the heterogeneous group of disorders in solitary, odd children whose behavior is characterized by social isolation, eccentricity, and peculiarities in language (autistic disorder, Asperger's disorder, expressive language disorder, and mixed receptive-expressive language disorder). Communication disorders might be distinguished by the primacy and the severity of the disorder in language accompanied by compensatory efforts of the child to communicate by other means and also by specialized language assessment. Autistic and Asperger's disorders are distinguished on the basis of the more severely impaired social interactions and stereotypic behaviors and interests in these two disorders. Schizotypal personality disorder is distinguished from the following personality disorders: schizoid and paranoid personality disorders (which rarely have magical thinking, unusual perceptual experiences, or oddities in speech, appearance, and thought processes), narcissistic personality disorders (with predominant sense of grandiosity, fragile self-esteem, and fear of having “hidden” imperfections or flaws revealed), avoidant personality disorders (which rarely have oddities in appearance and behavior, and in which fear of embarrassment, not disinterest and detachment, causes social avoidance and isolation); and borderline personality disorders (characterized by affective instability and stormy relationships as well as impulsive and manipulative behavior).

Cluster B Personality Disorders Narcissistic, antisocial, and histrionic personality disorder qualify for the so-called spectrum disorders. The three disorders have been shown to aggregate in the same family and co-occur in the same person. Phenotypically distinguishable disorders may be referred to as spectrum disorders if they meet the above two conditions. Spectrum disorders sometimes reflect differential expression of the same liability. Antisocial, narcissistic, and histrionic personality disorders exemplify the spectrum in which a tendency to aggressivity (associated with high novelty seeking) interferes with character development and maturity. Empirical data show that symptoms of the three disorders tend to group around impulsivity, aggression, and dramatic affects. A spectrum disorder may also reflect a less deviant form of the other on an underlying liability scale. In that regard, the underlying antisocial behaviors decrease in severity as one proceeds from antisocial (violent and property crime), via narcissistic (exploitation), to histrionic personality disorder (manipulation).

Antisocial Personality Disorder Antisocial personality disorder is characterized by pervasive disregard for, and violation of, rights of others since the age of 15. A person must be at least 18 years old and have displayed evidence of conduct disorder before the age of 15 (Table 24-22). Antisocial personality disorder may be complicated by dysphoria, tension, low tolerance for boredom, depressed mood, and premature violent death. Impairment is extremely variable, but typically includes social difficulties.

A. There is a pervasive pattern of disregard for and violation of the rights of others occurring since age 15 years, as indicated by three (or more) of the following:
   (1) failure to conform to social norms with respect to lawful behaviors as indicated by repeatedly performing acts that are grounds for arrest
   (2) deceitfulness, as indicated by repeated lying, use of aliases, or conning others for personal profit or pleasure
   (3) impulsivity or failure to plan ahead
   (4) irritability and aggressiveness, as indicated by repeated physical fights or assaults
   (5) reckless disregard for safety of self or others
   (6) consistent irresponsibility, as indicated by repeated failure to sustain consistent work behavior or honor financial obligations
   (7) lack of remorse, as indicated by being indifferent to or rationalizing having hurt, mistreated, or stolen from another
B. The individual is at least age 18 years.
C. There is evidence of conduct disorder with onset before age 15 years.
D. The occurrence of antisocial behavior is not exclusively during the course of schizophrenia or a manic episode.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 24-22 DSM-IV Diagnostic Criteria for Antisocial Personality Disorder

COMORBIDITY These patients are at increased risk for impulse control disorders, major depression, substance abuse or dependence, pathological gambling, anxiety disorders, and somatization disorder. The most common co-occurring personality disorders are narcissistic, borderline, and histrionic.

EPIDEMIOLOGY Prevalence rates of 3 percent for males and 1 percent for females in the general population and 3 to 30 percent in clinical settings, with even higher rates for forensic samples and substance abusers, have been reported. DSM-IV notes concern that this disorder may be underdiagnosed in females, given the emphasis on aggressive items in diagnosing conduct disorder. A high frequency of antisocial personality disorder is associated with low socioeconomic status and urban settings.

ETIOLOGY Antisocial personality disorder is more frequent among the first-degree biological relatives of probands with this disorder. Biological relatives of women with antisocial personality disorder are at greater risk for the same disorder than biological relatives of men with antisocial personality disorder. Genetic studies have suggested familial transmission of antisocial personality disorder, substance use, and somatization disorder, with the former two being characteristic of males and the latter of females in the same family. Adoption studies show that both genetic and environmental factors contribute to the risk for this disorder since both adopted and biological children of parents with antisocial personality disorder are at increased risk for this disorder.

Conduct disorder (before the age of 10) and accompanying attention-deficit/hyperactivity disorder increase the likelihood of developing antisocial personality disorder in adult life. Conduct disorder is more likely to develop into antisocial personality disorder when there is erratic parenting, neglect, or inconsistent parental discipline.

DIFFERENTIAL DIAGNOSIS Antisocial personality disorder is distinguished from manic episodes of bipolar disorder I on the basis of episodic course and euphoric mood in the latter. Antisocial disorder is distinguished from the following personality disorders: narcissistic personality disorder (which rarely manifests serious criminality, aggression, and deceit, and is characterized by excessive need for admiration and envy of others), histrionic personality disorder (which includes seductiveness, attention seeking, superficiality, and rarely serious criminality and aggressiveness), paranoid personality disorder (which includes suspiciousness, guarded attitude, rarely serious antisocial behaviors), and borderline personality disorder (which includes manipulativeness to gain nurturance, and affective instability). Contrary to common belief, persons with explosive temperaments (which underlie borderline symptoms and are characterized by high harm avoidance with high novelty seeking and low reward dependence) frequently manifest antisocial behaviors (this corresponds to the old term “secondary psychopathy” in the literature). Secondary psychopathy is distinguished from antisocial personality proper (or “primary psychopathy”) by high novelty seeking with low harm avoidance and low reward dependence, which corresponds to the adventurous temperament profile. Antisocial personality disorder must also be distinguished from adult antisocial behavior (with no personality pathology in the background).

Course After the age of 30, both the most flagrant antisocial behaviors (promiscuity, crime) and less severe behaviors and substance use tend to decrease.

Narcissistic Personality Disorder The hallmarks of narcissistic personality disorder are a pervasive sense of grandiosity (in fantasy or in behavior), need for

admiration, lack of empathy, and chronic intense envy (Table 24-23).

A pervasive pattern of grandiosity (in fantasy or behavior), need for admiration, and lack of empathy, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

(1) has a grandiose sense of self-importance (e.g., exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements)
(2) is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love
(3) believes that he or she is "special" and unique and can only be understood by, or should associate with, other special or high-status people (or institutions)
(4) requires excessive admiration
(5) has a sense of entitlement, i.e., unreasonable expectations of especially favorable treatment or automatic compliance with his or her expectations
(6) is interpersonally exploitative, i.e., takes advantage of others to achieve his or her own ends
(7) lacks empathy: is unwilling to recognize or identify with the feelings and needs of others
(8) is often envious of others or believes that others are envious of him or her
(9) shows arrogant, haughty behaviors or attitudes

Table 24-23 DSM-IV Diagnostic Criteria for Narcissistic Personality Disorder

This disorder may be complicated by social withdrawal or depressed mood. Frequently severe, impairment typically includes marital problems and poor interpersonal relationships in general.

COMORBIDITY These patients are at increased risk for major depression and substance abuse or dependence (especially cocaine use). Dysthymic disorder or major depressive disorder may develop in reaction to criticism or failure. The most common co-occurring personality disorders are borderline, antisocial, histrionic, and paranoid personality disorders.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 2 to 16 percent in clinical populations and less than 1 percent in the general population. According to DSM-IV, this disorder is more commonly diagnosed in males.

ETIOLOGY A higher risk for this personality disorder may exist in the offspring of narcissistic parents who give their children an unrealistic sense of grandiosity. In addition, most narcissistic persons are realistically talented, beautiful, or highly intelligent as these features serve as the nucleus around which the sense of specialness is further organized.

DIFFERENTIAL DIAGNOSIS Narcissistic personality disorder is distinguished from a manic or hypomanic episode by the episodic course, euphoria, and functional impairments in the latter two. Narcissistic personality disorder is distinguished from the following personality disorders: Antisocial personality disorder (narcissistic exploitativeness is more driven by the wish to establish one's dominance than by material gains, history of conduct disorder, and no excessive need for admiration); borderline personality disorder (which includes unstable self-concept, chaotic behaviors, self-destructive gestures, chronic anxiety, and rarely high achievements); histrionic personality disorder (which includes capacity for empathy, emotional display, and rarely unscrupulousness and exploitation of others); obsessive-compulsive personality disorder (which includes inflexibility, orientation to detail, and social isolation in addition to perfectionism and belief that others cannot do things as well); and paranoid and schizotypal personality disorders (which include suspiciousness and social withdrawal).

COURSE The course is chronic; however, narcissistic symptoms tend to diminish after the age of 40, when pessimism usually develops.

Histrionic Personality Disorder The hallmarks of histrionic personality disorder are pervasive and excessive self-dramatization, excessive emotionality, and attention seeking (Table 24-24).

A pervasive pattern of excessive emotionality and attention seeking, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

(1) is uncomfortable in situations in which he or she is not the center of attention
(2) interaction with others is often characterized by inappropriate sexually seductive or provocative behavior
(3) displays rapidly shifting and shallow expression of emotions
(4) consistently uses physical appearance to draw attention to self
(5) has a style of speech that is excessively impressionistic and lacking in detail
(6) shows self-dramatization, theatricality, and exaggerated expression of emotion
(7) is suggestible, i.e., easily influenced by others or circumstances
(8) considers relationships to be more intimate than they actually are

Table 24-24 DSM-IV Diagnostic Criteria for Histrionic Personality Disorder

Frequently impairment is mild. The patients may exhibit frequent suicidal gestures and threats to coerce better caregiving. Interpersonal relations are unstable, shallow, and generally ungratifying. Frequent marital problems result from the tendency to neglect long-term relationships for the excitement of new relationships.

COMORBIDITY These patients are at increased risk for major depression, somatization disorder, and conversion disorder. The most common co-occurring disorders are narcissistic, borderline, antisocial, and dependent.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 2 to 3 percent in the general population and 10 to 15 percent for psychiatric inpatients and outpatients. This disorder is generally agreed to occur far more frequently among women. According to DSM-IV, it might be equally frequent among men and women.

ETIOLOGY This disorder tends to run in families. A genetic link between histrionic and antisocial personality disorders and alcohol use disorders has been suggested, but many of those previously described as having histrionic personality disorder would now be classified as having borderline personality disorder.

DIFFERENTIAL DIAGNOSIS Histrionic personality disorder is distinguished from the following personality disorders: antisocial personality disorder (which includes antisocial behaviors and crime to gain profit, power, or some other material gratification; history of conduct disorder; no excessive self-dramatization; no exaggerated emotional expression), borderline personality disorder (which includes bunstable self-concept, chaotic behaviors, self-destructive gestures, chronic anxiety, and identity disturbance, and narcissistic personality disorder (which includes fear of having “hidden” imperfections and flaws revealed, sense of grandiosity, and sense of specialness).

Borderline Personality Disorder The hallmarks of borderline personality disorder are pervasive and excessive instability of affects, self-image, and interpersonal relationships as well as marked impulsivity (Table 24-25).

A pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:
(1) frantic efforts to avoid real or imagined abandonment. **Note:** do not include suicidal or self-mutilating behavior covered in criterion 5.
(2) a pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation
(3) identity disturbance: markedly and persistently unstable self-image or sense of self
(4) impulsivity in at least two areas that are potentially self-damaging (e.g., spending, sex, substance abuse, reckless driving, binge eating). **Note:** do not include suicidal or self-mutilating behavior covered in criterion 5.
(5) recurrent suicidal behavior, gestures, or threats, or self-mutilating behavior
(6) affective instability due to a marked reactivity of mood (e.g., intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days)
(7) chronic feelings of emptiness
(8) inappropriate, intense anger or difficulty controlling anger (e.g., frequent displays of temper, constant anger, recurrent physical fights)
(9) transient, stress-related paranoid ideation or severe dissociative symptoms

Table 24-25 DSM-IV Diagnostic Criteria for Borderline Personality Disorder

The disorder may be complicated by psychotic-like symptoms (hallucinations, body image distortions, hypnagogic phenomena, ideas of reference) in response to stress and premature death or physical handicap from suicide and suicidal gestures, failed suicide, and self-injurious behavior. Frequent and severe impairment may involve frequent job losses, interrupted education, and broken marriages.

Comorbidity These patients are at increased risk for major depression, substance abuse or dependence, eating disorder (notably bulimia), posttraumatic stress disorder, and attention-deficit/hyperactivity disorder. Borderline personality disorder co-occurs with most other personality disorders.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 2 percent in the general population, 10 percent for psychiatric outpatients, 20 percent for psychiatric inpatients, and 30 to 60 percent among patients with personality disorders. According to DSM IV, this disorder is more commonly iagnosed in females.

ETIOLOGY Numerous studies have pointed to early traumatic experiences as a cause of this personality disorder. Recently, a tripartite etiological model, including childhood trauma, vulnerable temperament, and a series of triggering events was formulated. Dynamic and biological psychiatry agree that a combination of early traumatic events and certain biological vulnerabilities (mostly in the emotional domain) represent primary etiological factors for this disorder. Physical and sexual abuse, neglect, hostile conflict, and early parental loss or separation are more common in childhood histories of patients with this disorder. Familial aggregation of borderline personality disorder has been repeatedly demonstrated. Borderline personality disorder is five times more common among relatives of probands with this disorder than in the general population. This disorder is also associated with increased familial risk for antisocial personality disorder, substance abuse, and mood disorders.

DIFFERENTIAL DIAGNOSIS Borderline personality disorder is distinguished from mood disorder, dysthymic disorder, and cyclothymia (with depression and mood swings mimicking borderline affective problems) on the basis of efforts to avoid abandonment, unstable relationships with alternation between idealization and devaluation, identity disturbance, impulsivity in potentially self-damaging areas, chronic feelings of emptiness, and inappropriately intensive anger or difficulty controlling anger, which are rarely observed in mood disorder, dysthymia, or cyclothymia).

Borderline disorder shares many features and is difficult to distinguish from all other personality disorders. Frequently it is an exclusion diagnosis based on typical clinical symptoms for other personality disorders. Borderline disorder is distinguished from identity problems, which is limited to a developmental stage.

COURSE The variable course most commonly follows a pattern of chronic instability in early adulthood with episodes of serious affective and impulsive dyscontrol. Impairment and the risk of suicide are greatest in the young adult years and gradually wane with advancing age. In the fourth and fifth decade these individuals tend to attain greater stability in their relationships and functioning.

Cluster C Personality Disorders

Avoidant Personality Disorder Clinical Criteria Avoidant personality disorder is characterized by pervasive and excessive hypersensitivity to negative evaluation, social inhibition, and feelings of inadequacy (Table 24-26). Impairment can be severe and typically includes occupational and social difficulties.

A pervasive pattern of social inhibition, feelings of inadequacy, and hypersensitivity to negative evaluation, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following:
(1) avoids occupational activities that involve significant interpersonal contact, because of fears of criticism, disapproval, or rejection
(2) is unwilling to get involved with people unless certain of being liked
(3) shows restraint within intimate relationships because of the fear of being shamed or ridiculed
(4) is preoccupied with being criticized or rejected in social situations
(5) is inhibited in new interpersonal situations because of feelings of inadequacy
(6) views self as socially inept, personally unappealing, or inferior to others
(7) is unusually reluctant to take personal risks or to engage in any new activities because they may prove embarrassing

Table 24-26 DSM-IV Diagnostic Criteria for Avoidant Personality Disorder

COMORBIDITY These patients are at increased risk for mood and anxiety disorders (especially social phobia, generalized type). The most common co-occurring personality disorders are schizotypal, schizoid, paranoid, dependent, and borderline personality disorders.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 0.5 to 1 percent in the general population and 10 percent for psychiatric outpatients. According to DSM-IV, this disorder is equally frequent in males and females.

DIFFERENTIAL DIAGNOSIS Avoidant personality disorder is very difficult to distinguish from social phobia (many authors believe these are alternative labels for the same or similar condition). In social phobia, specific situations rather than interpersonal contact are avoided. Panic disorder with agoraphobia also manifests avoidance but usually after the onset of panic attacks.

Avoidant disorder is distinguished from schizotypal and schizoid personality disorders (social isolation of avoidant personalities is accompanied by the desire for social relations, which is not observed in schizoid and schizotypal personality disorder), paranoid personality disorder (which includes guarded attitude, preoccupation with hidden meanings and conspiratorial explanations of events), and dependent personality disorder (which focuses on being taken care of rather than on the fear of negative evaluation).

COURSE Frequently beginning in childhood with shyness and fear of strangers and new situations, disfiguring illness and shyness in childhood predispose children to avoidant personality disorder.

DEPENDENT PERSONALITY DISORDER The hallmark of dependent personality disorder is a pervasive, excessive need to be taken care of, leading to clinging behavior, submissiveness, fear of separation, and interpersonal dependency (Table 24-27). Complications include low socioeconomic status and poor family and marital functioning.

A pervasive and excessive need to be taken care of that leads to submissive and clinging behavior and fears of separation, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

(1) has difficulty making everyday decisions without an excessive amount of advice and reassurance from others
(2) needs others to assume responsibility for most major areas of his or her life
(3) has difficulty expressing disagreement with others because of fear of loss of support or approval. **Note:** do not include realistic fears of retribution.
(4) has difficulty initiating projects or doing things on his or her own (because of a lack of self-confidence in judgment or abilities rather than a lack of motivation or energy)
(5) goes to excessive lengths to obtain nurturance and support from others, to the point of volunteering to do things that are unpleasant
(6) feels uncomfortable or helpless when alone because of exaggerated fears of being unable to care for himself or herself
(7) urgently seeks another relationship as a source of care and support when a close relationship ends
(8) is unrealistically preoccupied with fears of being left to take care of himself or herself

Table 24-27 DSM-IV Diagnostic Criteria for Dependent Personality Disorder

Impairment is frequently only mild and typically includes poor interpersonal relationships and occupational functioning if independence is required.

COMORBIDITY These patients are at increased risk for major depressive and other mood disorders, anxiety disorders, adjustment disorder, and social phobia. The most common co-occurring personality disorders are histrionic, avoidant, and borderline personality disorders.

EPIDEMIOLOGY DSM-IV reports it to be the most frequent personality disorder. The disorder is equally frequent in males and females.

ETIOLOGY Chronic physical illness or separation anxiety disorder may predispose for dependent personality disorder, whose familial pattern and genetics are unknown.

DIFFERENTIAL DIAGNOSIS Dependent personality disorder is distinguished from dependency seen in mood disorders, panic disorder, agoraphobia, and as a result of general medical condition.

Dependent disorder is distinguished from borderline personality disorder (which includes unstable, stormy relationships and reaction to abandonment with rage, emptiness, and demands as opposed to increasing appeasement and submissiveness seen with dependent personalities), histrionic personality disorder (which includes gregarious flamboyance with active demands for attention), and avoidant personality disorder (which includes social isolation because of the fear of negative evaluation, rather than the clinging and submissive behavior of dependent personalities).

Obsessive-Compulsive Personality Disorder The hallmark of obsessive-compulsive personality disorder is pervasive preoccupation with orderliness, perfectionism, and mental and interpersonal control, at the expense of flexibility, openness, and efficiency (Table 24-28). Complications include distress and difficulties when confronted with new situations that require flexibility and compromise and myocardial infarction (secondary to features typical of type A personalities, such as time urgency, hostility, competitiveness). Frequently severe, impairment typically includes occupational and social difficulties.

A pervasive pattern of preoccupation with orderliness, perfectionism, and mental and interpersonal control, at the expense of flexibility, openness, and efficiency, beginning by early adulthood and present in a variety of contexts, as indicated by four (or more) of the following:

(1) is preoccupied with details, rules, lists, order, organization, or schedules to the extent that the major point of the activity is lost
(2) shows perfectionism that interferes with task completion (e.g., is unable to complete a project because his or her own overly strict standards are not met)
(3) is excessively devoted to work and productivity to the exclusion of leisure activities and friendships (not accounted for by obvious economic necessity)
(4) is overconscientious, scrupulous, and inflexible about matters of morality, ethics, or values (not accounted for by cultural or religious identification)
(5) is unable to discard worn-out or worthless objects even when they have no sentimental value
(6) is reluctant to delegate tasks or to work with others unless they submit to exactly his or her way of doing things
(7) adopts a miserly spending style toward both self and others; money is viewed as something to be hoarded for future catastrophes
(8) shows rigidity and stubbornness

Table 24-28 DSM-IV Diagnostic Criteria for Obsessive-Compulsive Personality Disorder

COMORBIDITY These patients are at increased risk for major depression, anxiety disorder, and obsessive-compulsive disorder.

EPIDEMIOLOGY DSM-IV reports prevalence rates of 1 percent in the general population and 3 to 10 percent for psychiatric outpatients. According to DSM-IV, this disorder is twice as common in males as in females.

ETIOLOGY Obsessions and compulsions have been repeatedly linked to high central serotonergic function. However, the latter is associated with anxiety in general, which supports the hypothesis that obsessions and compulsions represent psychological and behavioral mechanisms against anxiety. Some studies have demonstrated familial aggregation of this disorder.

Differential Diagnosis Obsessive-compulsive personality disorder is distinguished from obsessive-compulsive disorder on the basis of true obsessions and compulsions in the latter.

This personality disorder is distinguished from schizoid personality disorder (which includes lack of capacity for intimacy and social isolation secondary to emotional detachment, as opposed to devotion to work and discomfort with emotions), antisocial personality disorder (which includes material goals in antisocial behavior and criminality as opposed to hypermorality of obsessive personalities), and narcissistic personality disorder (which includes sense of grandiosity, self-aggrandizement, exhibitionism, and fear of having “hidden” imperfections and flaws revealed).

Disadvantages of DSM-IV Classification In addition to general problems shared by all categorical models of personality, the DSM classification has its own specific limitations. First, DSM-IV is an etiologically atheoretical classification of personality disorder based solely on descriptive criteria derived from observations of prototypical cases. This atheoretical approach was expected to stimulate work on the etiology of deviant behaviors classified as personality disorders. Unfortunately, this has not happened. The introduction of descriptive criteria in DSM-III inspired extensive research on some basic psychometric aspects of the personality disorder diagnosis but rarely studies of etiology, underlying dynamics, or pathogenesis. While this emphasis on elementary psychometrics obviously reflects an attempt of this new field to delineate its subject, it also reflects the purely descriptive approach that provided no testable etiopathogenetic models or hypotheses. Consequently, after almost two decades of research based on DSM criteria, our conceptual understanding of personality disorders as specified by DSM is still at a basic descriptive level similar to that introduced in 1980.

Second, DSM-IV describes as pathognomonic for personality disorders maladaptive response patterns that are enduring, that is, appearing before adulthood and typical of long-term functioning. In practice, however, it can be difficult to distinguish long-term maladaptation indicating personality disorder from chronic personality changes caused by other factors, such as other mental disorders (e.g., chronic depression), long-term situational factors (e.g., job-related behavior requirements), and other medical disorders (e.g., irritability associated with hyperthyroidism).

Third, the DSM definition of personality disorders is inherently imprecise because it requires that maladaptive behaviors cause clinically significant subjective distress or clinically significant impairment in social and occupational function or both. The quantifier “significant” is an arbitrary diagnostic element that can hardly be

objectivized. In addition, subjective distress usually means low self-esteem, which often leads to anxiety, guilt, depression, and hypochondriasis, which are frequently diagnosed and treated as independent Axis I disorders while the background personality disorder is overlooked.

CLINICAL AND PSYCHOMETRIC ISSUES

Ego-Syntonic and Alloplastic Nature of Personality Disorder Patients with personality disorder typically blame other people or unfavorable circumstances for their own problems. This externalizing of responsibility results from two characteristics of these disorders. First, most of these patients perceive their own deviant behaviors as adequate; in other words, their symptoms are ego-syntonic. The exception are patients with cluster C dependent and avoidant personality disorders. Both of these disorders are characterized by prominent anxiety associated with maladaptive behaviors, which frequently causes patients to perceive their symptoms as disturbing (i.e., ego-dystonic). Second, patients with personality disorder try to change others, instead of changing themselves; in other words, their attitude is described as alloplastic. The exception are patients with avoidant and schizoid personality disorders. Avoidant persons usually try to improve their own performance to avoid negative evaluation by others. Those with schizoid personalities are so socially detached that they are usually indifferent to what others might do, think, or feel.

Personality Disorder: Social or Clinical Diagnosis? One of the general DSM-IV requirements for personality disorder is that maladaptive behavior “deviates markedly from the expectations of the individual's culture.” This requirement introduces a significant social connotation to a clinical classification of behaviors. Indeed, clinical diagnosis of a personality disorder is often based solely on the extent to which certain behavior is adjusted to the local society. To what extent should mental health professionals serve as designated gate keepers for behavior traits considered unusual in a particular social setting? Most temperament profiles with high reward dependence have a low incidence of diagnosed personality disorder, whereas those with high harm avoidance, high novelty seeking, or both increase this risk. In general, sociocultural pressures promote phenotypes that fall within the range of accepted behavior norms. In Western societies, the normative phenotype is not the one with average personality traits (as one would expect given the adaptive flexibility of such configuration) but the one with high reward dependence. High reward dependence leads to much easier conditioning of socially accepted behaviors than do novelty seeking and harm avoidance. Similarly child psychiatry classifies three broad groups of mental disorders characterized by high novelty seeking (e.g., attention-deficit/hyperactivity disorder, conduct disorder), high harm avoidance (e.g., depressive and anxiety disorders), and low reward dependence (e.g., schizophrenia, autistic disorder). Most children with these extreme temperaments are diagnosed and treated when young, whereas children with high reward dependence (sociable, attached, dependent) are selectively encouraged to mature without psychiatric intervention. This raises a general question about the role of our society in shaping psychological profiles of its members that is beyond the scope of this chapter. Suffice it to say, anatomical adaptation appears to have become less critical for our survival than behavioral adaptation. Natural selection may be targeting not our morphological features but our personality features and intelligence. Through its suppression of certain temperament traits and its modeling effects on character development, modern human society obviously has an important role in this process.

Categorical Versus Dimensional Approach to Personality Disorder Medical diagnosis has historically been categorical, and most clinicians tend to think categorically for two reasons. First, one of the most important functions of diagnosis is to prescribe appropriate corrective action (i.e., treatment). Deciding whether a person is affected or unaffected is functionally equivalent to deciding whether or not that person needs treatment (hence, categorical diagnosis facilitates treatment decisions in practical work). Second, categorical systems simplify professional communication because they describe prototypical and severe cases that meet all or most of the specified criteria. Just one categorical label (e.g., antisocial personality disorder) can communicate a great deal of vivid clinical information about the patient. However, categorical descriptions convey useful information about prototypical and severe cases, but little or no information about atypical, mixed, or mild cases.

Categorical Approach The categorical approach adopts the medical diagnostic model that a personality disorder is either present or absent. Categorical models are optimal for discontinuous variables (e.g., pregnancy, skin cancer). However, when the distinguishing characteristics of a disorder are quantitative variants of normally occuring phenomena, such as personality traits, any categorical decision about the presence or absence of a disorder is arbitrary. In other words, one of the major disadvantages of categorical models of personality disorders is that they establish arbitrary cutoff points or thresholds for continuous behavior traits. This becomes clear especially in those with mild cases close to the cutoff point who cannot be well classified as affected or unaffected. In an attempt to reduce this problem, DSM-IV now allows clinicians to classify maladaptive personality traits that do not meet the threshold for a personality disorder on Axis II.

Next, a categorical approach is not optimal for patients who do not perfectly fit their pigeonhole in the classification, and an additional wastebasket category for “atypical” or “mixed” cases must be established. Finally, categorical approach to personality disorder does not establish a prescriptive relation between diagnosis and treatment. Reflecting a significant overlap of diagnostic criteria for personality disorder subtypes, categorical systems usually yield multiple personality diagnoses for individual patients. In such cases, treatment priorities are easily and frequently confused.

Dimensional Approach Dimensional models define a number of graded and continuous behavior dimensions and specify individual differences as quantitative variations along these dimensions. These models are optimal for features that vary quantitatively and have different adaptive significance in different situations (such as personality traits). These models recognize that everyone has multiple personality traits that are more or less prominent and adaptive rather than simply present or absent. This eliminates the need for multiple, overlapping categorical diagnoses. By quantifying their position on a continuous personality trait, dimensional models conserve more information about individual patients than categorical models. Personality dimensions are as easily applied to common atypical and mild cases as they are to the rare prototypical, severe cases. Finally, a dimensional system facilitates mathematical manipulation of complex data with many variables.

Research shows that normal personality traits tend to generalize to personality disorders, indicating that deviant traits may be conceptualized as extreme variants of normal, adaptive behaviors. Note, however, that dimensional models do explain what makes some traits disorders. Statistical deviance alone does not suffice to define dysfunction. In addition, extreme variants can be labeled normal or maladaptive in different contexts. Finally, natural breaks or points of rarity on the continuum of normal and maladaptive personality traits have not been detected. Hence, dimensional models must introduce cutoff points for continuous traits (which are usually arbitrary, unless there are natural steps as described for character in Table 24-12) or use probability estimates for categorical personality disorders on the basis of the severity of measured personality traits.

Measurement of Personality Disorders

Self-report or Interview? No convincing evidence favors self-report or interview, still, interviews are considered more reliable and valid diagnostic tools than self-reports for personality disorders. Many of these patients tend to be biased in reporting their ego-syntonic deviant behaviors. The problem is reduced when the patient is interviewed because unclear, inconsistent, or defensive responses can be clarified, and the patient's demeanor and appearance observed. On the negative side, interviews tend to be affected by systematic biases and rating idiosyncrasies of interviewers. In selfreports, various response sets (e.g., careless inconsistency, defensiveness, and exaggeration) and validity items can aid interpretation. Recently, a set of performance scales, independent of the item content, based on the observed regularities in the technical pattern of responding to questions was defined in some tests, such as the TCI. These performance scales are used to test the validity of self-reported personality traits. Specifically, self-report to questions, regardless of the content of the questions, tends to follow specific patterns that correlate with personality traits. For example, low reward dependence is indicated by endorsement of a large number of items that are rare in the general population. These performance scales are used to predict scores on each of the temperament and character dimensions, which are then compared with self-reported, content-dependent scores on these same traits. Substantial score differences between subjective and performance-based scores on personality traits indicate either significant cognitive distortion about own behaviors or intentional mispresentation.

Note that both self-reports and interviews depend on the patient's accuracy, honesty, and insightfulness. Hence, collection of data from collateral informants and expert ratings is usually considered critical to ensuring high-quality personality diagnosis.

State-Trait Effect Most Axis I disorders have less-severe variants, or “representatives,” on Axis II (Table 24-29). The state-trait distinction between symptoms reflecting long-term personality traits and those reflecting transient Axis I states can be difficult. Indeed, transient variation in one's emotional state can influence both self-report and clinical evaluation of long-term personality characteristics. The susceptibility to such state effects has been demonstrated for some self-report categorical measures of personality disorders. In contrast, only those personality dimensions that underlie susceptibility to depression and anxiety (e.g., neuroticism and harm avoidance) tend to covary with mood and anxiety syndromes. Also, increased anxiety or depression leads to mild and transient decreases in self-directedness and cooperativeness. Consequently, a person with no enduring personality disorder may act immaturely when depressed or under stress. Self-reports of temperament traits unrelated to this susceptibility (e.g., novelty seeking and reward dependence) are largely independent of comorbid mood and anxiety. This remarkable finding suggests that most aspects of personality are unaffected by current anxiety and mood states. The state effect can be reduced (but not eliminated) when subjects are repeatedly reminded to describe their usual, not their current, behavior, feelings, and attitudes.

Axis I Disorders "Major Variant"	Axis II Disorders "Minor Variant"	Shared Symptoms
Schizophrenia spectrum	Schizoid PD* Schizotypal PD	Negative and positive
Delusional disorder	Paranoid PD	Suspiciousness
Major depression	Depressive PD Borderline PD	Proneness to depressed mood; self-destruction;
Cyclothymia/bipolar disorder/mania	Narcissistic PD Histrionic PD Antisocial PD	Mood swings; impulsiveness;
Obsessive-compulsive disorder	Obsessive-compulsive PD	Hypochondriasis; inflexibility
Social phobia	Avoidant PD	Shyness; avoidant behavior;
Panic disorder with agoraphobia	Dependent PD	Dependency

* PD, personality disorder.

Table 24-29 Phenotypic Similarity Between Axis I and Axis II Disorders

Categorical Tests for Personality Disorder Most categorical tests for personality disorders are based on the DSM criteria and share all the advantages and disadvantages of that polythetic categorical approach. In the DSM-III field trials, the interrater reliability was .61 and the test-retest reliability was .54 for any personality disorder. The reliability is usually lower for individual subtypes than for any personality disorder because of the difficulty of distinguishing among the subtypes.

Among the categorical self-reports, the two most frequently used are the Personality Diagnostic Questionnaire (PDQ-R), and the Minnesota Multiphase Personality Inventory (MMPI) scales for personality disorders (MMPI-PD). The PDQ-R is a self-report version of the DSM-III-R criteria for personality disorders, along with the impairment/distress scale to quantify the significance of impairment or subjective distress required by DSM criteria. In practice, this instrument overdiagnoses personality disorders compared with expert clinical diagnosis. To the authors' knowledge, no PDQ-IV version matching the DSM-IV criteria for personality disorders has been published.

The MMPI-PD scales were derived from the MMPI item pool to evaluate symptoms for DSM-III personality disorders. The MMPI-PD scales are incorporated into MMPI so that, in addition to obtaining a reliable evaluation of personality disorders and other clinical syndromes, clinicians and researchers can also use other MMPI scales (e.g., Lie, Validity, Defensiveness) for validity and response-set analyses.

Among the interviews that categorically diagnose personality disorders, the most commonly used is the Structured Interview for DSM-IV Personality Disorders (SIDP-IV). The original SIDP was developed for personality disorders described in DSM-III and it has been regularly updated with each new DSM revision and edition (SIDP-R for DSM-III-R, SIDP-IV for DSM-IV). To increase its diagnostic validity, the SIDP-IV requires that answers for selected questions be obtained from a collateral informant. This test can be combined with a 10-minute SIDP-IV screen for personality disorder, which can considerably reduce testing time in opulations with many individuals who do not have personality disorder.

Other popular interviews for categorical classification of personality disorders include the Structured Clinical Interview for DSM-III-R Personality Disorders Revised (SCID-II), the Diagnostic Interview for Personality Disorders (DIPD), and the Personality Disorder Examination (PDE). The PDE, available in an international version from the World Health Organization in several languages, makes diagnoses according to ICD in addition to DSM.

Dimensional Assessment of Personality Disorders Historically, personality has been evaluated dimensionally and personality disorders categorically. Reflecting this unfortunate separation, conceptual advances in the field of normal behavior and personality have had little or no effect on research in the field of personality disorder, and vice versa. For example, if the features associated with the extremes of Eysenck's personality traits (i.e., neuroticism, extraversion, and tough mindedness) are combined, the resulting combinations do not correspond closely to traditional categories of personality disorders.

The popular five factor model of normal personality was recently advocated as a potential model for personality disorders. However, the homogeneity of postulated traits and their specificity for personality disorders is questionable. The ability of the five factor model to account for the structure of abnormal personality was tested primarily in nonclinical samples and normal individuals. It provides no measures of response bias or validity as needed in clinical samples. The few studies that have tested the NEO Personality Inventory (NEO-PI) in clinical samples showed that neuroticism, extraversion, and low agreeableness predict personality disorder symptoms. However, these studies did not control for mood and anxiety disorders, leaving open to question the generalizability of the findings to clinical subjects with comorbid Axis I syndromes. In fact, after controlling for age and depression (which is considered standard for high-quality personality diagnosis), only high neuroticism and low agreeableness remained significant predictors of personality disorder symptoms. Obviously, these two dimensions define personality disorder nonspecifically, as a general predisposition to psychopathology (high neuroticism) accompanied by an antagonistic facade (low agreeableness). In addition, neuroticism appears to be confounded with nonspecific factors (e.g., depression or anxiety) that reduce its ability to distinguish those with personality disorder, those with other psychopathology (e.g., mood and anxiety disorders), and well-adjusted individuals with high neuroticism.

Dimensional tests of personality disorders are formatted as self-reports (except for Tyrer's Personality Assessment Schedule). The most frequently used are the Millon Clinical Multiaxial Inventory (MCMI), the NEO PI (which evaluates the five factor personality model), the TCI, and the Dimensional Assessment of Personality Pathology Disorder (DAPP) by John Livesley. These tests have been originally designed to evaluate normal personality (NEO-PI), psychiatric patients (MCMI), personality disorders (the DAPP), and both normal and deviant personality (TCI).

The Millon Clinical Multiaxial Inventory-II (MCMI-II), designed to match DSM-III-R personality disorders, evaluates long-term behavior traits systematized as 10 basic personality patterns (dependent, avoidant, schizoid, passive-aggressive, narcissistic, antisocial, hysterical, compulsive, aggressive, and self-defeating) and 3 pathological personality disorders (borderline, schizotypal, and paranoid). When present, the latter indicate the severity of the 10 basic personality patterns. The test also evaluates nine Axis I clinical syndromes. Some reports suggest that MCMI-II tends to overdiagnose personality disorders in comparison to expert clinicians. A study comparing the MCMI-II and the TCI pointed to a considerable overlap of Millon's measures of Axis I syndromes and Axis II personality disorders.

Several lines of evidence suggest that some common, core features, distinct from other forms of psychopathology that causes personal dysfunction, characterize personality disorder subtypes classified in DSM-IV as discrete categories. In the literature, the number of proposed dimensions for deviant personality ranges from 2 to 24, but, in most cases, can be reduced to 4 basic traits: fearful, agressive or impulsive, detached, and compulsive. Most of the proposed dimensions are phenotypic, factor analytically derived. More importantly, no consensus has been reached about neither the optimal number nor the content of dimensions that purport to describe personality and its disorders most efficiently. These commonly described traits are quite similar to the DSM-IV Clusters A, B, and C (plus compulsivity) and to the four temperament traits described in the TCI (high harm avoidance, high novelty seeking, low reward dependence, and persistence). Livesley and colleagues recently reported that four phenotypic and genetic factors account for most of the variance in both normal personality and personality disorder. The factors are: emotional dysregulation, dissocial behavior, inhibition, and compulsivity. These traits fit nicely the description of personality disorder, but emotional dysregulation, dyssocial behavior, inhibition, and compulsivity do not seem optimal to describe adaptive behavior traits observed in normal personality. Moreover, these four traits do not answer the question of what makes individuals with personality disorder maladaptive. In brief, these four traits most closely correspond to extreme variants of normal temperament traits described in the TCI. The structure of character needs to be determined in order to evaluate maturity versus immaturity and maladaptive versus adaptive behaviors. In other words, the distinction between temperament and character is essential to improve both the etiopathogenetic understanding and the diagnosis of personality disorder. Fortunately, structural equation analysis of the causes of individual differences in the resemblance in twin pairs varying in degree of genetic relationship can specify a unique model of the genetic structure of normal and deviant personality.

TCI The TCI evaluates four major temperament dimensions (harm avoidance, novelty seeking, reward dependence, and persistence) and three major character dimensions (self-directedness, cooperativeness, and self-transcendence). Each of these higher-order temperament and character dimensions is composed of component facets (or subscales) to evaluate response patterns associated with specific stimuli (e.g., harm avoidance can be manifested as worry and pessimism, fear of uncertainty, shyness, or fatigability depending on the stimulus). Recently, the TCI home page has been developed on the World Wide Web. The seven major dimensions and their component traits are presented Table 24-30.

Table 24-30 Temperament and Character Inventory (TCI) Scales and Subscales

The TCI is a family of tests with several specialized versions designed for varying types of informants (self-report, peer ratings, interviewers), varying age groups (for 7 to 14 years, the so-called Junior TCI; for 15 years or above, Adult TCI), scope of information (temperament, character, both), and level of clinical detail (144 items for major dimensions only, or 295 items for multiple facets of each dimension). The test measuring temperament dimensions only was originally called the Tridimensional Personality Questionnaire. This test measured novelty seeking, harm avoidance, reward dependence, and persistence, but persistence was originally scored as a component of reward dependence. The name of the test was changed to TCI when the character scales were added and persistence was recognized as a fourth, separately inherited temperament dimension in twin studies in Australia and the United States. The self-report version of the TCI is a 295-item, true/false test. Its psychometric properties are presented in the *TCI Manual*.

The ability of the TCI to predict categorical symptoms and diagnoses of personality disorders was clinically tested in samples of psychiatric inpatients and outpatients with and without personality disorders and varying mood and anxiety states. Low scores on character dimensions, especially self-directedness and cooperativeness, were associated with high symptom counts for personality disorders in general and for each of the DSM clusters of personality disorders. Note that each DSM cluster is primarily predicted by one of the TCI temperament dimensions (Table 24-31).

		Symptoms		
TCI Scales	Total	Cluster A	Cluster B	Cluster C
Novelty seeking	.22‡	.02	.44*	−.06
Harm avoidance	.31†	.23†	.08	.43*
Reward dependence	−.14	−.37*	−.08	−.04
Persistence	.00	−.07	.04	−.01
Self-directedness	−.56*	−.35*	−.43*	−.50*
Cooperativeness	−.44*	−.44*	−.40*	−.28†
Self-transcendence	.02	−.08	.03	.04

*$p < .0001$; †$p < .001$; ‡$p < .01$.

Table 24-31 Correlations Between TCI Scales and the Total Number of Symptoms for Inpatients With Cluster A, Cluster B, and Cluster C Personality Disorders (*N* = 136)

Low scores on character dimensions have been repeatedly shown to correlate with symptoms for each personality disorder subtype. Clearly, low scores on character traits represent a core feature extending across clusters and subtypes of personality disorder. In addition, four TCI temperament traits correlate with each subtype of personality disorder in a unique way (Fig. 24-8). Based on this, personality disorder subtypes are distinguished on the basis of temperament scores with no overlap.

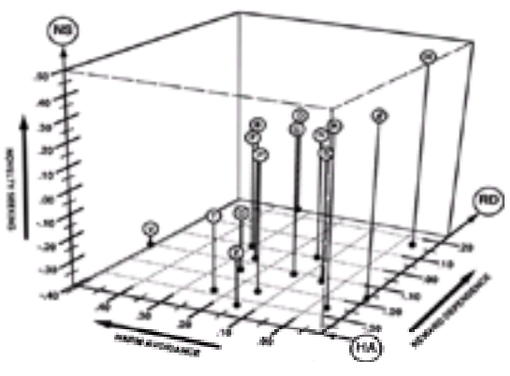

FIGURE 24-8 Correlations between individual personality disorder subtypes and temperament traits of harm avoidance, novelty seeking, and reward dependence.

Regression analyses confirmed that the TCI character scales of self-directedness, cooperativeness, and self-transcendence predict the number of personality disorder symptoms, after controlling for age, anxiety, and depression (Table 24-32).

		Partial R^2	*F*
Inpatients ($N = 136$)			
Total number of interview PD symptoms predicted by TCI (controlled for age and depression)			
Age and depression		.35	.0001
Novelty seeking		—	ns
Harm avoidance		—	ns
Reward dependence		—	ns
Persistence		—	ns
Self-directedness		.13	.0001
Cooperativeness		.03	.0076
Self-transcendence		.02	.0376
Cumulative	R^2	.18	.0001
Model	R^2	.53	.0001
Outpatients ($N = 109$)			
Total number of self-report PD symptoms predicted by TCI (controlled for age, depression, and anxiety)			
Age, anxiety and depression		.45	.0001
Novelty seeking		—	ns
Harm avoidance		—	ns
Reward dependence		—	ns
Persistence		—	ns
Self-directedness		.20	.0001
Cooperativeness		.05	.0001
Self-transcendence		.05	.0001
Cumulative	R^2	.30	.0001
Model	R^2	.75	.0001

Table 24-32 Regression Analyses of Personality Disorder Symptoms Predicted by TCI Scales and Observed

These results have critical diagnostic and treatment ramifications for personality disorders.

DIAGNOSIS OF PERSONALITY DISORDERS Character and temperament traits delineate both the core features and the distinguishing features of personality disorders

(Table 24-30, Fig. 24-8). Poorly developed character traits, especially self-directedness and cooperativeness, are a common denominator extending across all subtypes of personality disorder. Together with self-transcendence, character traits predict categorical personality disorder symptoms (Table 24-21). Clinically, this explains chronic difficulties in accepting responsibility, setting long-term goals, having fragile self-esteem, and other features associated with low self-directedness that are so characteristic of individuals with personality disorders. Usually (but not always) these patients are also uncooperative as well (i.e., revengeful, opportunistic, self-centered, socially intolerant, unhelpful to others, and lacking in empathy, compassion, or principles). Self-transcendence correlates with schizotypal and paranoid symptoms (depicting primary process thinking and fanaticism, respectively) and with personality subtypes characterized by dissociation and mood swings (e.g., borderline, histrionic), much as predicted in Figure 24-4. In addition, self-transcendence appears to be a critical component in defining the underlying susceptibility to psychosis (when the other two character traits are low) or to high creativity (when the other two character traits are high).

Note that high self-directedness does not always protect against personality disorder. Some narcissistic and antisocial persons can be resourceful, self-disciplined, and successful, but their low cooperativeness (i.e., intolerance, revengefulness, low empathy, etc.) makes them maladapted.

Once the probability for the presence of personality disorder is established on the basis of character, temperament traits are used for differential diagnosis. Membership in cluster A (aloof, eccentric personalities) is most clearly determined by low reward dependence, which is characterized by social detachment. Membership in cluster B (dramatic, erratic personalities) is determined by high novelty seeking, which is characterized by impulsivity. Membership in cluster C (anxious personalities) is determined by high harm avoidance, which is characterized by fearfulness (Table 24-20). Recent data indicate that persistence predicts obsessive-compulsive traits. In summary, individual subtypes are distinguished on the basis of their unique pattern of correlations with temperament traits without any overlap (Fig. 24-8). Based on these data and those cited above about the structure of deviant personality, it seems safe to claim that the clinical presentation of personality disorder varies along four underlying dimensions, corresponding to the four temperament traits of harm avoidance, novelty seeking, reward dependence, and persistence. Quantifiable dimensional features of personality disorders are summarized in Table 24-33.

Consistent features
Low self-directedness
Irresponsible, blaming
No mature goals
Resourceless, helpless
Poor self-esteem
Undisciplined
Low cooperativeness
Intolerant of others
Lack of empathy
Unhelpful
Revengeful
Unprincipled
Variable features
High persistence (obsessive-compulsive symptoms only)
Low reward dependence (odd cluster only)
High novelty seeking (erratic cluster only)
High harm avoidance (anxious cluster only)

Table 24-33 Quantifiable (Dimensional) Features of Personality Disorder

This model does not require arbitrary cutoffs for the presence or absence of personality disorder. Personality disorder can be quantified in terms of severity or the probability of diagnosis can be estimated. As shown in Table 24-34, the probability of the presence or absence of personality disorder based on the self-directedness scores alone ranges from very low (about 10 percent) to very high (about 95 percent) in clinical samples.

Self-Directedness Scores	Predicted Percentage (%) with PD	Observed with PD f/N	%
44–39	13	4/18	22 ± 0.10
38–34	23	5/17	29 ± 0.11
33–30	33	4/16	25 ± 0.11
29–27	44	8/20	40 ± 0.11
26–23	55	8/19	42 ± 0.11
22–20	66	6/12	50 ± 0.14
19–16	70	14/16	88 ± 0.08
<16	85	17/18	94 ± 0.05

* Predicted % = 100 e^y/(1 + e^y), where y = −3.37 + .13 (SD score).

Table 24-34 Predictions of Personality Disorders (PD) From Self-Directedness (SD) Scores*

This model integrates categorical and dimensional aspects with personality disorders. It specifies temperament and character dimensions to account for individual differences in behavior. However, as shown in Figure 24-3 and Figure 24-4, unique combinations of temperament and character dimensions are categorically classified as eight clinical types, profiles, or configurations. In other words, this model is dimensional but retains the notion of categorical clinical diagnosis because each temperament and character profile corresponds to one multidimensional clinical category (which is useful in practice for description and prediction of clinical course). On the other hand, quantified dimensions constituting a particular temperament and character profile are used for individualized treatment planning and studies of etiology and pathophysiology.

ETIOPATHOGENESIS OF PERSONALITY DISORDER

Only a few etiopathogenetic models of personality disorder are available. Of those, the psychodynamic model elaborates the effects of early traumatic experiences, whereas the spectrum disorder hypothesis postulates a common cause of personality syndromes with Axis I disorders. A growing body of evidence points to a multifactorial, nonlinear etiopathogenesis (including genetic, neurophysiological, learning, maturational, and cultural factors, which are fully incorporated into the psychobiological model of personality disorder).

Spectrum Disorder Hypothesis Echoing Ernest Kretchmer's work from many decades ago, some experts still consider subtypes of personality disorders minor variants of major Axis I syndromes. Indeed, most Axis I disorders have their less severe variants, or representatives, on Axis II (e.g., schizophrenia is represented by schizotypal and schizoid personality disorders; delusional disorder, paranoid type, by paranoid personality disorder; social phobia by avoidant personality disorder; cyclothymic disorder by the group of emotionally unstable cluster B personality disorders).

These Axis I and Axis II disorders are symptomatologically similar and are frequently differentiated only by a severity criterion. The etiopathogenetic nature of this Axis I–Axis II relationship has not been established, aside from the obvious symptomatic similarity. Plausible alternatives include Axis II disorders as predisposing factors, premorbid features, subclinical expression, or *forme frustes* of Axis I disorders. The possibility of a common etiopathogenesis thus remains open. So far, a close etiological relation has been solidly supported for schizotypal personality disorder and the spectrum of schizophrenia.

Psychodynamic Model of Personality Disorder The central idea behind this model is that personality disorder subtypes classified as separate nosologic units reflect different behavioral expressions of the same core deficit in personality. The so-called borderline level of personality organization has been postulated to represent the nucleus shared by most subtypes of personality disorder.

As defined by Otto Kernberg in 1975, the borderline level of personality organization is characterized by nonspecific manifestations of ego weakness (e.g., lack of impulse control, lack of anxiety tolerance, and lack of sublimatory potentials), specific ego defects (manifested as partially blurred self-object boundaries, mild-to-moderate primary process thinking, periodically distorted reality testing), partial object relations (manifested as alternations between “all good” and “all bad” perception of the same object), primitive defense mechanisms (centered around splitting), and identity disturbance. In addition, an inadequately developed superego

is either sadistic (e.g., obsessive personalities) or rigid in some, but completely absent in other areas, permitting conflict-free expression of impulses and behaviors.

Immature defenses and a fragmentary self-concept, very similar to the above-described borderline personality organization, are a normal early phase in mental development. In the case of personality disorder, however, this early developmental phase pathologically persists in later periods. Etiologically, this pathological persistence is postulated to reflect either constitutional factors (e.g., poor anxiety tolerance, high aggressivity, genetic vulnerability to certain affects) or environmental traumatic factors (e.g., early separation, loss of a parent, physical or mental abuse, neglect). According to the theory, because of such traumatic etiological factors, aggressively charged, negative representations of the self and external objects are predominantly incorporated into the internal world. This, in turn, interferes with the crystallization of the early ego, which normally occurs around positive, libidinally cathected representations of the self and external objects. In other words, normal early motivation and growth are centered around positive primary emotions, particularly interest and joy. It is the relative predominance of strong negative emotions, particularly anger and fear, that interferes with normal development. As a consequence, primitive defenses (splitting and the related defenses), which normally predominate in this early phase of life, pathologically persist in the inner world and interfere with normal development. More-mature defenses, such as repression, require more energy for their operation than splitting (e.g., repression requires at least as much energy for its operation as is contained in unacceptable impulses). In contrast, splitting simply keeps opposite good and bad self- and object-representations apart. As Kernberg pointed out, splitting and early ego mutually perpetuate each other. By keeping predominantly negative self- and object-representations apart from a few positive representations, primitive defenses ensure at least some development around positive experiences but interfere with neutralization of aggression and fear. These primary emotions and their related pregenital impulses (or a chaotic combination of pregenital and genital ones) persist as dominant motivators of behavior and significantly interfere with normal personality development.

Borderline Core of Personality Disorders Clinically, the borderline personality organization is observed as (1) chronic, free-floating anxiety, (2) polysymptomatic neurosis (e.g., multiple phobias related to one's body or appearance, bizarre conversion symptoms, dissociative reactions, hypochondriasis), (3) polymorphous perverse sexual trends (e.g., coexistence of genital and pregenital elements, bizarre forms of perversions involving aggression), (4) impulse control problems and addictions, (5) shift toward primary process thinking (magical thinking, primitive fantasies), and (6) partially impaired reality testing. These borderline symptoms and signs are shared by other subtypes of personality disorder as well. James Reich and Allen Frances concluded in 1984 that the diagnosis of borderline personality organization is “essentially equivalent” to the diagnosis of any personality disorder. Likewise, a hierarchical cluster analysis of identical and fraternal twins, their siblings, and parents showed that 12 DSM personality disorder subtypes can be grouped into eight heritable syndromes, which can be further aggregated into one borderline condition.

Christopher Perry and George Vaillant wrote: “Just as fever, pus, and callus formation are the body's physiological reactions to insults of disease, personality disorder reflects persons' efforts to heal themselves.” Indeed, a personality disorder reflects a person's attempt to overcome underlying fragility through maladaptive, but purposeful, behaviors. These maladaptive behaviors are referred to as “purposeful” because they compensate for the fragility and proneness to decompensation associated with the borderline nucleus. Persons with personality disorder become less fragile (because the personality deviation becomes the principal motivator of specific behaviors) but more abnormal (because an already immature personality is being complicated by additional distortions of self- and object-representations). These compensatory behaviors can have different types of symptomatic styles (e.g., avoidant, paranoid, antisocial, schizoid, dependent). Note that compensatory behavior styles help us give a syndrome a descriptive label (e.g., avoidant, antisocial) but tell us little about the disorder underlying that description. Clinical presentation (i.e., distinguishing behavior style) depends on internal factors (e.g., fearfulness causes avoidant behaviors), external factors (e.g., social class shapes antisocial behaviors), or both. In other words, depending on certain constitutional and environmental factors, persons with a borderline level of personality organization create the clinical picture of their personality disorder.

A subgroup of borderline patients manifests the core symptoms (e.g. storrmy affects, vague identity, unstable rrelationship) in a rather stable way (stable instability). These patients are fragile and prone to fragmentation but do not develop any defensive maladaptive behavior facade to compensate for this fragility. Some of them can periodically mimic behaviors of any personality disorder subtype (avoidant, antisocial, obsessive) as a temporary, chameleon-like solution to their chronic internal problems. These core borderline patients correspond closely to the DSM-IV category of borderline personality disorder.

The described compensatory process generates a two-level personality structure with the dominant compensatory self-concept and the split-off real self-concept. Consequently, clinical expression of personality disorders is dominated by overt, prototypical behaviors, but a set of subtle, covert symptoms and signs associated with the underlying borderline nucleus are regularly observed clinically. These difficult-to-detect symptoms and signs are usually overlooked as diagnostic criteria of personality disorders. For example, the clinical picture of narcissistic personality disorder is dominated by the grandiose self, a deviant self-concept that generates an unrealistic sense of one's specialness, entitlement, arrogance, and other prototypical narcissistic behaviors or symptoms. The hidden real self of a narcissistic person is recognizable through chronic feelings of inferiority, hypochondriasis, envy, and pessimism that regularly accompany their clinical picture. Similar overt and covert clinical features have been established for other personality disorders as well.

Further Conceptual and Methodological Issues The borderline level of personality organization can be reliably distinguished from the schizophrenic spectrum disorders but less reliably from affective syndromes. This is related to the fact that the same immature personality processes that increase susceptibility to deviant behaviors may also increase susceptibility to mood disorders. Indeed, extreme temperament variants increase the risk for personality disorder and are associated with excessive basic emotions such as fear and anger. In many cases, mood disorders and personality disorders are comorbid and can be so interwoven that no meaningful distinction between them can be made (as implied in concepts of depressive personality, hysteroid dysphoria, or characterological depression).

An increasing number of psychometric tests for psychodynamic constructs are available (e.g., the Defense Style Questionnaire, Kernberg's Structural Interview). Satisfactory diagnostic stability and discriminant validity in patients given the diagnosis borderline personality organization has been demonstrated.

PSYCHOBIOLOGICAL SUMMARY OF PERSONALITY DISORDER

The psychodynamic and psychobiological model of personality disorder share many conceptual views. Both models define a common denominator (the core feature) that extends across discrete subtypes of personality syndromes. This core feature is the borderline personality organization in the psychodynamic model and poorly developed character traits, mostly self-directedness, in the psychobiological model. Descriptively, low self-directedness and borderline personality organization are for all practical purposes identical (e.g., both describe blaming, helpless, irresponsible, unreliable individuals with no internal direction and with difficulties in defining and pursuing stable goals; rather, they tend to experience numerous short-term, frequently mutually exclusive motives, none of which can develop to the point of real personal significance). Low self-directedness correlates highly with the use of immature defense mechanisms that are central in the psychodynamic concept of borderline personality. Etiologically both concepts postulate that excessive negative emotionality, such as fear and anger (or extreme harm avoidance and novelty seeking), interferes with developmental processes and increases the risk of personality disorder. Clearly, the psychodynamic understanding of personality disorder was a useful first step in understanding human behaviors. However, the psychodynamic model neither accounts for nor provides, directions for expanding our understanding of the full etiopathogenetic complexity of observed behaviors (e.g., neurochemical, genetic, learning factors). As an etiopathogenetic theory, the psychodynamic model addresses only a portion of the iceberg of deviant behaviors. However, as a treatment tool, its unique understanding of emotional and cognitive processes behind deviant behaviors is used effectively in the therapy of personality disorder, especially in reorganizing internalized concepts about the self and the external world. In summary, psychodynamic strategies are an integral part of a comprehensive psychobiological understanding of and approach to the treatment of personality disorder.

In contrast, modern psychobiology of behaviors provides an integrative, multifactorial, and developmental etiopathogenetic model of personality and its disorders. This approach is based on four fundamental ethiopathogenetic perspectives: genetic, neurophysiological, learning, and phenomenological. These factors interact in a nonlinear way to produce phenotypic differences in cognitive styles and behavior traits. Psychobiology of temperament and character traits is discussed throughout this chapter (see Table 24-2, Table 24-3, Table 24-4, Table 24-5 and Table 24-6 for review).

Genetic and environmental factors interact in complex ways to influence the risk of personality disorder. Available genetic observations about twins, adoptees, and families are explained by the hypothesis that quantitative inheritance of underlying personality dimensions influences the risk of personality disorder, rather than positing separate inheritance of personality disorder subtypes. More than one half of the variance in the four major temperament traits is inherited. These temperament traits have their dissociable genetic, neurobiological, and phenomenological correlates (Table 24-2, Table 24-3, Table 24-4, Table 24-5 and Table 24-6). Temperament traits determine one's susceptibility to specific neurochemical processes leading to individual differences in basic emotions and biased learning. These antecedent temperament factors, along with systematic cultural biases and random life events, critically influence character development, represented as internalized concepts about the self and the external world. As shown in Table 24-18, various temperament types differentially affect one's risk of immature character and personality disorder. Some configurations (mostly those with high reward dependence) protect against personality disorders, whereas some increase this risk (e.g., the explosive or borderline profile with low reward dependence, high novelty seeking, and high harm avoidance). Contrary to common belief, average scores on the temperament dimensions do not protect against maladaptation and immaturity. People with average temperament traits have an average (not a decreased) risk of

personality disorder (Table 24-35).

Temperament Type	Configuration*	N	% Immature
High risk			
Borderline	NHr	39	72
Obsessional	nHr	44	59
Antisocial	Nhr	25	48
Passive-aggressive	NHR	30	40
Average risk			
Average profile	—	15	33
Low risk			
Avoidant	nHR	30	17
Schizoid	nhr	31	16
Histrionic	NhR	50	12
Reliable	nhR	16	6
Total		300	33

* Lower-case letters (h, n, r) and upper-case letters (H, N, R) indicate low and high values for temperament traits, respectively.

Table 24-35 Relative Risk of Personality Disorder as a Function of Temperament Type in a General Community Sample

Extreme temperament variants do not necessarily indicate personality pathology. They are expected to be associated with long-term personal, social, or occupational impairments that warrant the personality disorder diagnosis only when accompanied by low character traits. In other words, poorly developed character is what makes some behavior traits maladaptive and increases the risk of personality disorders. An individual high in novelty seeking and low in harm avoidance may have an impulsive personality disorder if they are low in self-directedness and cooperativeness or may be an energetic businessman or inquisitive scientist without personality disorder if they are self-directed and cooperative. Mature character traits (i.e., mature concepts about one's self and the external world) optimize adaptation of temperament (i.e., basic emotionality) to the environment by reducing discrepancies between one's emotional needs and norm-favoring social pressures. In personality disorder, immature character traits and extreme temperament configurations mutually perpetuate each other.

The concepts of temperament and character are essential to divide the symptoms of personality disorder into common features shared by all personality disorders and distinguishing features that permit discrimination among personality disorder clusters and individual subtypes. This is not possible in models that confound temperament and character. This also indicates that personality disorder reflects deviations in both temperament and character, that is, in both emotional and conceptual personality processes with critical treatment implications.

Extreme temperament configurations tend to be stable. The stability of nonoptimal configurations (i.e., they are maladaptive relative to the global adaptive optimum possible for that individual) occurs because each step in personality development must increase the adaptive fitness so that it is more successful in balancing multiple internal and external constraints. In other words, search for higher adaptive maxima is discouraged because it initially necessitates decreasing overall fitness (the U-shaped pattern). This necessity to decrease current adaptation to achieve a better one accounts for both the chronicity and the treatment resistance of personality disorders.

Behavioral geneticists have demonstrated that the effect of sociocultural factors on personality is less specific than that of genetic factors and influences success in adaptation rather than its form or personality style. This is consistent with recent findings about the importance of family and local culture in character development. Family environment does not influence temperament, but explains about 35 percent of variability of character traits. Hence, psychosocial disorganization in the rearing environment of a child substantially influences the risk of personality disorders. This is essential for preventive strategies as even temperament configurations with high risk of personality disorder might be overcome in homes and communities that provide security and limit behavior in a warm, compassionate manner, as well as encouraging self-directed choice and respect for other people.

Personality disorder is associated with younger age, which indicates maturation (i.e., remission of deviant behaviors) with increasing age. In general, three dimensions of personality change substantially with age. Novelty seeking decreases with age by approximately 18 percent; thus older individuals are less impulsive (more reflective), less rule breaking (more orderly), and less quick-tempered (more stoical). Cooperativeness increases markedly in most children during school age and then increases by 12 percent on average after age 18. Self-directedness increases markedly in most people during adolescence and young adulthood, increasing on average by 9 percent after age 18.

The decreasing prevalence of personality disorder with age is attributable to increased development of both self-directedness and cooperativeness with age. The additional tendency for novelty seeking to decrease with age explains why patients with impulsive personality disorders show more improvement than those with anxious or eccentric personality disorders. The best-documented finding about change in deviant behaviors is the remission of criminal behavior in individuals with antisocial personality disorder. These individuals nearly always remained impulsive (high novelty seeking), risk taking (low harm avoidance), and aloof (low reward dependence) but became mature enough to maintain work and family life in a stable manner.

TREATMENT

Most individuals with personality disorder perceive their lifestyles as normal and seldom seek treatment. Typically, they seek help when their maladaptive behaviors cultminate in unbearable marital, family, and career circumstances with secondary anxiety, depression, substance abuse, or eating disorders.

Temperament traits are primarily treated by pharmacological intervention. However, given the importance of associative learning in the development of these traits, some psychotherapeutic correction is theoretically possible, especially with behavioral techniques. Alternatively, commonly used psychotropic drugs rarely induce changes in internalized self-concepts, that is, character, which is more amenable to psychotherapeutic intervention. Character optimizes adaptation of temperament to the environment because it modulates the salience of percepts and primary emotions, thereby reducing the maladaptive impact of extreme temperament traits. Hence, extreme temperament and immature character traits are optimally treated simultaneously with combined psychotherapy and pharmacotherapy.

Psychotherapy It is hard to find a psychotherapeutic method that has not been tried for personality disorders. This primarily reflects ambiguities in the conceptual understanding of these disorders. The psychobiological evaluation of temperament and character traits and their component facets enables clinicians to tailor psychotherapy and pharmacotherapy to the specific needs of each individual patient.

General Issues Each school of psychotherapy provides a specific understanding of behavior and a particular method of intervention. In practice, different schools are not mutually exclusive, but overlap and complement one another. Dynamic psychotherapy addresses the internal world of the patient's emotions and needs and treats symptoms as external manifestations of internal conflicting motivations. Behavior therapy focuses on external manifestations (or symptoms) and enables patients to change behavior or better control their behaviors. Cognitive therapy helps patients to correct their distorted cognitive appraisal of the significance of environmental cues and their underlying core beliefs that lead to maladaptive behaviors. Humanistic approaches, by increasing both self-directedness and cooperativeness, help patients achieve personal and social maturity in a form of altruistic individualism. No one of the above goals is invariably more correct than the others, and each orientation is expected to make a specific contribution to the overall efficacy of treatment. This seems especially relevant to personality disorders, where the reduction of symptoms and the improvement of social functioning are equally important. For example, a combination of dynamic therapy (insight oriented) and cognitive-behavioral therapy (action oriented) efficiently helps patients transform their insights into actual behavioral change. Recently, dialectical behavioral therapy based on a biosocial theory that borderline symptoms primarily reflect dysfunction of the emotion-regulation system has shown superiority in both reducing core symptoms and increasing social adjustment of borderline patients. A growing number of therapists are beginning to ignore ideological barriers dividing different schools and attempt both technical synthesis (eclecticism) and theoretical synthesis (integration) of various orientations (called *integrativeeclectic psychotherapy*).

As a rule of thumb, combinations of various orientations and formats and emphasis on teamwork are optimal in the psychotherapy of personality disorders. Some basic rules and values are strictly observed, however. Probably most crucial is having a stable therapeutic framework with consistent and reliable care. Next, behavior and feelings are used as the principal mode of communication. The therapist is active and uses high-energy confrontation and care (so-called therapeutic pressing). The central message is always “doing something with the patient, not something to the patient.” This way patients feel somewhat in control, which might keep them in treatment. Reflecting their splitting mechanism, these patients alternatively feel inferior and omnipotent, angry at others and self-destructive, sensitive to rejection but usually provoking it. Flexibility in approach, but firmness in basic values, with creativity and readiness to step away from the rules to get out of frustrating “no way out”

situations, is essential. Many of these patients cannot tolerate feeling better, as this means that the therapist is successful. These and similar frustrative situations cause countertransference problems, with a potential loss of professional objectivity; constant supervision and a support network are therefore necessary. Remember, these patients are almost never as good as they look when they are doing well and almost never as bad as they look when they are not doing well.

Pure supportive psychotherapy is rarely used for personality disorder. Pure support, aimed at strengthening existing coping styles (which are by definition maladaptive in personality disorder), often reinforces the problems of these patients. Kernberg has formulated a modification of supportive therapy that can be useful in treating patients with personality disorders. Also, modified supportive therapy can be used as the initial phase of the treatment, during the contract and trust-building phase.

The psychobiological approach incorporates these psychodynamic strategies into a comprehensive treatment plan aimed at stimulating character development, primarily self-directedness and cooperativeness. The primary focus is to change internalized conceptual representations of the self and the external objects (i.e., concepts about self, society, and the world as a whole). This is attempted either with cognitive methods (aimed at revising these concepts), with dynamic methods (aimed at stimulating maturation of internalized object-relations), or most frequently with a combination of the two. Cognitive methods use emotions as the royal road to cognition. Also, object relations are a special kind of internalized concept about self and others associated with either positive or negative affects, depending on the nature of the relation. In other words, both cognitive and psychodynamic methods address emotional and conceptual aspects of deviant behaviors. Dynamic and cognitive methods are complemented by behavioral and experiential techniques, which are efficient in transforming concepts and insights into everyday life.

During therapy, as character matures and new concepts and their associated secondary emotions develop, they neutralize extreme temperament traits and their related basic emotions of fear and anger. Behaviors change accordingly, from being primarily reactive (i.e., steered by basic emotions and automatic responses regulated by temperament) to being primarily proactive (i.e., steered predominantly by secondary emotions and active symbolic constructs regulated by character traits).

Practical Issues Psychotherapy of personality disorders is a specific strategy rather than a strictly defined method because the therapy takes place constantly (during any contact with the patient) not only during the psychotherapy sessions proper. In fact, what is happening between sessions may be critical for the outcome of the treatment as a whole. With these patients, psychotherapy essentially means reparenting, which, although sometimes demanding, leaves space for various types of interventions, such as education, help with real-life problems, and encouragement. Frequent sessions (at least once a week) are needed to develop reasonably complex interactions in the relationship for diagnostic and (especially) treatment purposes.

Patients with personality disorder are required to clarify their goals and objectives in treatment. These may be simple and concrete (e.g., “to reduce alcohol use”), more complex (e.g., “to become independent”), or very ambitious (e.g., “to be able to love”). The therapist evaluates each treatment goal and determines the likelihood of successful outcome. Most authors agree that the therapist and the patient must be in agreement regarding these goals in at least one area. Patients with personality disorder cannot all be helped to achieve their goals. In general, patients who have never had a meaningful relationship with at least one person are less likely to benefit from psychotherapy. Additional prognostic factors are the patient's intelligence quotient (I.Q.) and psychological mindedness, the therapist's training, and the compatibility of desired treatment goals with baseline personality characteristics. Efficient treatment is expected to stimulate development of character traits; this, in turn, sets the stage for a better adaptation of congenital temperament traits to the environment. As shown in Figure 24-5, Figure 24-6, and Figure 24-7, one temperament configuration may lead to several, but not all possible, character outcomes. Obviously, some treatment goals chosen by individuals with incompatible temperament traits simply might not be achievable. In other cases, where desired treatment goals are not incompatible with the baseline temperament configuration desired goals are achieved through character changes that range from ideal to optimal to compromises. Ideal outcomes correspond to preset concepts of an ideally mature personality with full development of all three character traits. These outcomes are not always possible, especially for patients with personality disorders, who usually have temperament configurations incompatible with high scores on all character dimensions (Fig. 24-5, Fig. 24-6, and Fig. 24-7). However, whenever possible, ideal outcomes are pursued as the ultimate standard of maturity. More frequently, treatment is planned to enable patients to achieve the best possible adaptation for their given temperament traits. This is called an “optimal character outcome” If for any reason optimal outcome cannot be achieved, the alternative is the so-called compromise outcome, which improves the patient's adaptation over baseline but does not achieve ideal or optimal adaptation possible for that patient.

For patients who are unlikely to benefit from psychotherapy, symptom control might be achieved through pharmacological intervention. In addition, many areas of everyday life (e.g., friendships, romantic relationships, media, education, and such valued life opportunities as stable marriage, work-related progress, and religious conversion) increase their chances for maturation. Often the emergent changes are sudden and associated with a new perspective on life and new goals and values, which cannot be achieved by logic, medication, or advice alone.

At least initially, the therapy is supplemented by as much structuring of the patient's life as needed. This may range from directing behaviors, to day-hospital programs, to hospitalization. Complications often require additional structuring such as phone calls or extra sessions. Structuring, advice, and logic are not expected to generate personality change, only to temporarily improve behavior control. As Kernberg noted, psychotherapy begins where common sense ends. Prolonged structuring robs patients of the opportunity to become more self-reliant or to learn from their failure to do so.

These patients usually want psychotherapy their way, with many conditions. In general, genuine personal motivation or strong social pressure are preconditions for a real change in personality. These patients rarely manifest personal motivation for change (especially after depression or anxiety is reduced). In fact, they frequently minimize or deny the existence of even serious problems, such as a suicide attempt. These blind spots for their problems reflect the fact that these patients, since an early age, have developed efficient mechanisms to avoid disturbing feelings and insights. Hence, they generally feel much less distress with their symptoms than their environment. Confrontations (not explanations!) are used to increase the patients' discomfort with their symptoms, which then improves their recognition of these symptoms and their motivation to change them. Most of these patients tolerate confrontations well if they trust their therapist. They need an extraordinary amount of stimulation (sometimes achieved through confrontations) in any relationship, including their therapy. This hunger for stimuli is related to either lack of stimulation (neglect) or extreme stimulation (abuse) by their caregivers.

Resistance to change is strong in patients with personality disorders, whose symptoms reflect their attempt to overcome their own fragility through compensatory deviant behaviors. The achieved compensatory facade (e.g., avoidant, schizoid, antisocial) is maladaptive but stable. In psychotherapy, these patients are expected to give up their purposeful, if maladaptive, behavior traits. In the process, they temporarily become similar to the core borderline patients without any behavioral compensation for their fragility. In other words, most patients with personality disorders first must regress briefly to a more primitive (but less deviant) level of personality organization and then develop more mature object relations and better adapted personality traits. This exposure of their fragility and proneness to fragmentation underlies their strong resistance to change. The progressive regression, or back-to-the-future, strategy derives directly from the U-shaped pattern, in which better adaptation is achieved only after initially decreased adaptation. Each personality disorder subtype requires specific modifications and careful timing of this strategy in addressing their distinguishing symptomatic facade and specific social deviations.

Group therapy is generally considered useful for persons with personality disorders as it both exposes and treats their social deviance. It is usually done in conjunction with individual therapy, which often elaborates experiences from group sessions.

Pharmacotherapy Increasing evidence shows pharmacotherapy to be at least as important as psychotherapy in the overall treatment of personality disorders. Pharmacotherapy is either (1) causal, aimed at correcting neurobiological dispositions underlying deviant traits, or (2) symptomatic, aimed at correcting target behaviors and symptoms of personality disorders. The third approach, based on the expectation that the treatment of diagnosable comorbid Axis I disorders might indirectly improve personality symptoms, receives least support in the literature.

Causal Pharmacotherapy of Personality Disorder The central idea behind causal pharmacotherapy is that enduring personality changes may result from pharmacological manipulation of the underlying neurobiological dispositions to deviant traits (the so-called trait vulnerability). When directed toward the deepest level of personality (i.e., the neurochemical systems mediating dispositions to mood and learning), pharmacotherapy may modify these dispositions. This, in turn, is expected to facilitate changes in affective and learning processes, leading ultimately to changes in observable behaviors and better overall adaptation.

In contrast to factor-analytically derived models, psychobiological models provide testable guidelines for pharmacological manipulation of the underlying neurochemical trait-vulnerability. Harm avoidance, novelty seeking, and reward dependence have been postulated to reflect individual differences in serotonergic, dopaminergic, and noradrenergic neurotransmitter systems. Pharmacological trials based on these and similar postulates might lead to pharmacological dissection of the underlying biological vulnerabilities.

How different patients with major depression will respond to antidepressants is better predicted by their temperament than by the number, type, severity, or course of their depressive symptoms. Patients who are dysphoric and highly sensitive to social approval (i.e., who are high in reward dependence and harm avoidance) are

most likely to improve on selective serotonin reuptake inhibitors (SSRIs). In contrast, those who are highly fearful but not socially dependent are most likely to improve on noradrenergic uptake inhibitors such as desipramine (Norpramine). Children with attention-deficit/hyperactivity disorder (who are high in novelty seeking) are efficiently treated with drugs that increase dopamine release and inhibit its reuptake, such as methylphenidate. Likewise, central serotonergic mediation is important in those with obsessive-compulsive behaviors, characterized by high harm avoidance. These behaviors are modified by drugs that change serotonergic system activity, such as SSRIs. Cerebrospinal fluid (CSF) shows high basal serotonergic activity in obsessive patients who respond to serotonin reuptake blockers and normalization of serotonin activity following long-term treatment (the treatment suppresses serotonin activity, presumably by reducing postsynaptic serotonin sensitivity).

Most of the presented treatment strategies remain tentative and experimental and need further testing. However, they are based on modern biology of behavior and provide guidelines that may potentially lead to a better understanding of biological vulnerabilities underlying deviant personality traits, to a more prescriptive relation between diagnosis and treatment, and to a more-specific therapy of deviant personality and mental disorder in general.

Symptomatic Pharmacotherapy of Personality Disorders Pharmacotherapy cannot be usefully organized around discrete subtypes of personality disorder. First, the efficacy of drug treatment is best evaluated on a symptom level, not the syndrome level. Second, the target symptoms likely to respond to particular drugs are not unique to any subtype, but are shared by various personality disorders. Third, the state-trait effect tends to interfere with effective evaluation of pharmacotherapy of personality disorders, which are frequently comorbid with mood and anxiety states. Fourth, some of the classified personality disorder subtypes are heterogenous composites that can be further subtyped into one or more subcategories, each potentially requiring specific pharmacotherapy.

Phenotypic similarity between some Axis I and Axis II disorders (Table 24-18) together with the lack of understanding of the underlying mechanisms of deviant personality traits explains the attempts to treat patients with personality disorders with drugs proved effective for the corresponding Axis I disorders. However, no common pathogenesis for the Axis I syndromes and their corresponding Axis II counterparts has been established, leaving open to question the validity of these treatments by analogy. Such treatments are neither theory driven nor hypothesis testing, and none has proved superior or ptimal for any individual personality disorder.

Nonetheless, symptomatic treatments are the standard of care for patients with personality disorder at present. In general, symptom improvement of 20 to 30 percent over placebo is expected. Pharmacological trials usually focus on short-term symptoms (e.g., suicidality, paranoia), but an increasing number of authors advocate treatment of long-term pathology (e.g., impulsiveness, anxiety, affective dysregulation) as well. In that regard, most authors agree that four symptom domains underlie chronic pathology of personality disorders: (1) aggression and behavioral dyscontrol, (2) affective symptoms and mood dysregulation, (3) anxiety, and (4) cognitive-perceptual distortions including psychotic symptoms.

These four symptom domains are clinically related to the underlying temperament dimensions that distinguish DSM-IV clusters of personality disorders: high novelty seeking and cluster B disorders correspond to the target symptom domain of impulsiveness and aggression; high harm avoidance and cluster C disorders correspond to the target symptom domain of anxiety and depression symptoms; and low reward dependence and cluster A disorders correspond to the target symptom domain of affective dysregulation, detachment, and cognitive disturbances. In other words, identification of these four target symptom domains has narrowed the gap between causal and symptomatic pharmacotherapy of personality disorders.

The interaction between biological and psychological factors in deviant behaviors is complex. For example, high aggressivity may cause splitting, which in turn prevents neutralization of aggressivity. Hence, the feed-back pattern between the unfavorable biology and unfavorable psychology can only be interrupted when drug treatment of the biological vulnerability is combined with psychotherapy of associated psychological mechanisms.

Pharmacological treatment of the four target symptom domains is nonspecific, as the selected drug regularly affects both its target symptom domain and other symptom domains. This nonspecificity may reflect many factors, such as diagnostic heterogeneity of subjects, or the equifinality phenomenon in which similar behaviors derive from different underlying antecedents, or the nonspecific action profile of the drug, affecting several neurophysiological systems at the same time. More-sophisticated, receptor-specific drugs might help answer some of these questions.

AGGRESSION It is useful, though sometimes difficult, to distinguish different types of aggression. The most common form of aggression occurs when a quick-tempered person is provoked by frustration or threats (affective aggression), which is frequent in impulsive-aggressive individuals. Biological correlates of impulsive aggression and poor behavior inhibition include low concentrations of CSF 5-hydroxyindoleacetic acid (5-HIAA), and altered serotonin neurotransmission. Unprovoked aggression can occur in patients with cerebral instability documented by an abnormal electroencephalogram (EEG) (ictal aggression), regardless of any associated personality traits. Predatory aggression (cruelty) involves hostile revengefulness and taking pleasure in victimizing others, often with intact impulse control; predatory aggression is most frequent in individuals with low cooperativeness, which is most likely in antisocial and borderline personalities. Organic-like impulsivity and aggression, often accompanied by poor social judgment, is distinguished from other impulsive-aggressive syndromes by prominent distractibility, inattention, emotional lability, and high somatic anxiety with panic and cardiorespiratory symptoms, often seen in patients with frontal lobe lesions.

Multiple double-blind trials have shown that lithium (Eskalith) reduces affective display and aggression in both normal subjects and impulsive-aggressive individuals. Lithium may help reduce cruelty and hostility, but this may be an indirect result of reducing impulsivity. Anticonvulsant mood stabilizers such as carbamazepine (Tegretol) and valproate (Depakene) reduce both the intensity and the frequency of unprovoked angry outbursts, improve behavior dyscontrol, and reduce anxiety and suicidality in some patients regardless of the normality of their EEGs. Carbamazepine is nonspecific in targeting both impulsivity and other chronic symptom domains and acute manifestations. Ictal aggression can also be treated with benzodiazepines, but anticonvulsants are recommended, because of frequent tolerance to the anticonvulsive effects of benzodiazepines.

Double-blind trials have shown that psychostimulants such as methylphenidate benefit inattentive and hyperactive adults who are impulsive and aggressive, especially when their symptoms began in early childhood.

As noted by Paul Soloff, impulsivity is most dramatically expressed as self-destructive or assaultive acts, but can also manifest in a wide spectrum of symptomatic behaviors (e.g., binge eating, spending, sex, drugs), or as cognitive impulsivity (e.g., rush to judgment). Consistent with the postulated serotonergic correlates for impulsive aggression, antidepressants (mostly SSRIs) benefit those with some chronically impulsive personality disorders (e.g., borderline personality disorders). In many cases, SSRIs, in addition to improving depressed mood and impulsivity, also nonspecifically improve affective lability, rejection sensitivity, impulsiveness, self-mutilation, and psychoticism. Monoamine, oxidase inhibitors (MAOIs) such as tranylcypromine (Parnate) benefit some patients with hysteroid dysphoria with somatic anxiety, hostility, and destructive impulsivity.

Patients with organic aggression subtype may respond to imipramine, psychostimulants (e.g., methylphenidate) and some of the novel cholinergic agonists (donepezil [Aricept]).

b-Adrenergic receptor antagonists (beta-blockers) reduce aggression and violence in patients with dementia, brain injury, schizophrenia, mental retardation, and organic brain syndrome.

Low-dose neuroleptics may be useful in modifying affective or predatory aggression in some patients. A variety of neuroleptics, from thioridazine (Mellaril) to haloperidol (Haldol), have been used at dosages much lower than those for psychoses (full antipsychotic dosages have generally not proved useful). The usefulness of new atypical neuroleptics for aggressive behaviors has yet to be established. In general, neuroleptics are used cautiously and short term to avoid potentially irreversible adverse movement effects. Also, dosage adjustment is crucial to maintain compliance because patients with personality disorders often have little tolerance for adverse effects.

Some relative pharmacological contraindications exist in treating impulsivity and aggression. For example, lithium should not be given to antisocial persons without aggression and impulsivity; it does not diminish nonaggressive antisocial behaviors (such as lying or stealing). It is also poorly tolerated by anxious schizoid individuals. Likewise, benzodiazepines and alcohol have disinhibiting effects on violence, reduce conditioned avoidance behavior (loosen inhibitions), and further impair passive avoidance learning in impulsive antisocial persons. Benzodiazepine use seems appropriate only in those with nonaggressive, dyssocial behaviors (e.g., schizoid personalities). In some borderline patients, carbamazepine was behaviorally toxic as it precipitated melancholic depression.

MOOD DYSREGULATION Target symptoms in the domain of mood dysregulation include emotional instability, emotional detachment, depression, and dysphoria. Emotional instability and mood swings respond to lithium, carbamazepine, or valproate. Low-dose neuroleptics such as haloperidol also have a mood-stabilizing effect. Tricyclic drugs such as amitriptyline (Elavil) increase impulsivity, suicidality, and assaultiveness in emotionally unstable patients with borderline features and

depression. These paradoxical effects are unrelated to anticholinergic or sedating effects of tricyclic drugs and are postulated to reflect cognitive and behavioral disorganization under catecholamine stress (i.e., increased catecholamine levels with tricyclic drug treatment).

Emotional detachment, cold and aloof emotions, and disinterest in social relations (chronic asociality) is typical of schizoid, schizotypal, and some antisocial and paranoid personalities. When social withdrawal reflects an underlying depression, antidepressants (SSRIs or MAOIs) may help. One should use tricyclic drugs cautiously in those with schizotypal personality disorder, for they may worsen psychosis. Emotional detachment often responds to serotonin-dopamine antagonists (atypical antipsychotics) such as risperidone (Risperdal), olanzapine (Zyprexa), or quetiapine (Seroquel), which may reduce social withdrawal and other features of eccentric personality disorders and carry less risk of extrapyramidal symptoms than the dopamine receptor antagonists (typical antipsychotics). These pharmacological responses suggest that the psychobiology of affective dysregulation in personality disorder subtypes may involve multiple neurotransmitter systems (e.g., dopaminergic, serotonergic) that regulate anger and inhibition.

Atypical depression and dysphoria, frequent in personality disorders, rarely respond to tricyclic drugs but rather to SSRIs or MAOIs. A lack of specificity of effect is also seen with MAOIs and SSRIs in borderline patients; several target symptom domains respond to these drugs: impulsivity, anger, hostility, mood reactivity, hypersomnia, hyperphagia, rejection sensitivity. In fact, at least half of patients with personality disorder and atypical depression worsen when given tricyclic drugs. In contrast, typical depressive episodes, which may complicate any personality disorder, are treated with antidepressants, including heterocyclic drugs, in dosages suggested for Axis I major depression. Paradoxically, neuroleptics are often effective in treating the affective symptom domain. Low-dose trifluoperazine (Stelazine) significantly reduces anxiety, suicidality, depression, and rejection sensitivity. Likewise, depoflupenthixol improves recurrent parasuicidal behavior in Cluster B patients (the same could not be demonstrated for antidepressants). This nonspecific effect has been attributed to reduced depersonalization, obsessive rumination, sense of helplessness, and paranoid ideation, not to improvement in core symptoms of depression. Recently, an open label trial of venlafaxine (Effexor), which is believed to block both serotonergic and noradrenergic reuptake, demonstrated that patients with borderline personality disorder improve globally on all Symptom Checklist-90 (SCL-90) scales (after 12 weeks of treatment with dosages as high as 400 mg a day), including reduction in somatic anxiety and complaints and self-injurious behavior. Atypical antipsychotics may reduce cognitive deficits and increase self-directedness.

ANXIETY Cluster C personality disorders (dependent, avoidant, and obsessive-compulsive personality disorders) are characterized by high harm avoidance and are thus expected to manifest chronic anxiety. With other personality disorder subtypes a state effect increase in their harm avoidance scores up to 25 percent with episodes of depression has been repeatedly observed. Empirical studies have demonstrated the background role of serotonin in harm avoidant behaviors and anxiety. This is clinically supported by the antianxiety efficacy of the SSRIs, 5-HT $_{1A}$ receptor agonists such as buspirone (BuSpar) and mixed serotonin reuptake inhibition and 5-HT$_{1C}$ and 5-HT$_2$ antagonism such as nefazodone (Serzone). The postulated role of the GABA system is supported by the efficacy of benzodiazepines. Patients with personality disorder usually manifest chronic anxiety, which makes the state-trait distinction very difficult; that is, it is not always clear whether a chronic state symptom (anxiety) becomes a stable personality characteristic, or vice versa. In addition, they often exhibit both cognitive anxiety (i.e., anticipatory worrying) and somatic anxiety (i.e., concern about bodily pains and psychophysiological reactions). Cognitive anxiety is most responsive to benzodiazepines, whereas somatic anxiety is more responsive to MAOIs and SSRIs. Again, benzodiazepines have a nonspecific effect as they also tend to reduce hostility, suspiciousness, cognitive disturbance, and sleep problems in borderline patients. Remember, they can also cause severe behavior toxicity and disinhibition. Patients with severe behavior inhibition (avoidance, in particular) may improve with phenelzine (Nardil) or fluoxetine (analogous to social phobia). Buspirone has a potentially important role in the treatment of anxiety associated with personality disorder given its low potential for dependence. Some components of somatic anxiety (e.g., sweating, palpitations, diarrhea, and tremor) can be treated with b-adrenergic receptor antagonists. Severe, psychotic-like anxiety responds to low-dosage antipsychotics, including the atypical antipsychotics.

COGNITIVE-PERCEPTUAL DISTURBANCES Cognitive disturbances, as defined by most authors, refer to magical thinking, odd beliefs, illusions, and long-term, low-grade psychotic symptoms observable primarily in schizotypal personality disorder. Every personality disorder subtype manifests cognitive disturbances and biased perception of reality because of the underlying specific learning and emotional biases. These disturbances are alternatively called a *nonpsychotic thought disorder* or *partial loss of reality testing*. Brief psychotic disorder may complicate most personality disorders. These are treated symptomatically, according to accepted pharmacological practices. In general, psychotic patients with personality disorder are likely to respond to, and comply with, low doses of neuroleptics. Acute psychotic symptoms requiring medication may subside when environmental stressors are brought under control; then one should lower the dosage or discontinue the medication.

Some patients with certain personality disorders (particularly borderline and schizotypal personality disorders) manifest chronic, low-grade cognitive symptoms such as nonpsychotic thought disorder (ideas of reference, magical thinking, odd fantasies, suspiciousness), unusual perceptual experiences (illusions), depersonalization, derealization, and eccentric behaviors. These chronic, low-level, psychotic-like symptoms respond to low-dose neuroleptics, typical and atypical. Sometimes, chronic cognitive disturbances, such as mild ideas of reference or suspiciousness, may subside when the background emotional tension is reduced by anxiolytics.

Neuroleptics also act nonspecifically as they improve several target symptom domains in borderline and schizotypal personalities. Symptoms that respond most clearly to low-dose neuroleptics are anger, hostility, suspiciousness, illusions, ideas of reference, anxiety, and obsessive-compulsive symptoms. The effect is most impressive when symptoms are severe, which leads some authors to speculate that these drugs might be a nonspecific treatment for symptom severity.

Table 24-36 summarizes drug choices for various target symptoms of personality disorders.

Table 24-36 Pharmacotherapy of Target Symptom Domains of Personality Disorders

Psychobiological Integration of Treatment The presented psychobiological model distinguishes components of personality that differ in etiology, pattern of development, and responses to psychotherapy and pharmacotherapy. This provides a foundation for integrating diagnosis and treatment planning in a manner that is generalizable but sensitive to differences among individual patients. Based on the established structural and clinical, and postulated neurochemical, characteristics of temperament and character, pharmacotherapy and psychotherapy can be systematically matched to the personality structure and stage of character development of each individual, clearly a unique advantage over other available approaches. Personality development can be described in terms of a sequence of 15 character steps in which attitudes, values, and emotions are transformed as a result of complex interactions among heritable predispositions, social learning, and individual experiences (Table 24-11). Each step in development results from a complex set of nonlinear facilitating and inhibitory influences.

A hierarchical model of the different types of underlying problems occurring at each step in personality development is shown in Table 24-37.

Character Deficits	Associated Mental Health Features
Level I "Walking together"	(Disorganized disorders)
[1] co1 - restless	Self-injurious behavior
[2] st1 - doubt and shame	Sadomasochistic sexuality
Level II "Working together"	(Severe personality disorders)
[3] st1 - poor impulse control	Bipolar affective lability
[4] st2 - aimlessness	Unemployment, criminality
[5] co2 - lack of empathy	Polysubstance abuse
Level III "Feeling together"	
[6] st2 - no conscience	Mild personality disorder
[7] st3 - low self-esteem	Social dysfunction
[8] co3 - selfishness	Frequent negative emotions
Level IV "Reflecting together"	
[9] st3 - materialism	Personal dissatisfaction
[10] st4 - arrogance	Lack of generativity
[11] co4 - revengefulness	Occasional negative emotions
Level V "Minds thinking as one"	
[12] st4 - possessiveness	Mature but lack of joy and lack of serenity
[13] st5 - integrated	
Level VI "Listening in silence"	
[14] co5 - unloving at times	Mature but lack of intuitive wisdom
Level VII "Experiencing in unity"	
[15] st5 - dualistic	Mature but judgmental and not fully harmonious

Relationships are designated by tail (self-directedness), co (cooperativeness), or st# (transcendence), and the number of the subscale shown in Table 24-37.

Table 24-37 A Hierarchical Model of Mental Order and Disorder Based on Level of Personality Development

Problems at the first level, involving basic trust (step 1) and confidence (step 2), are characteristic of individuals with a history of sexual or physical abuse beginning in early infancy and are associated with highly disorganized disorders. Following T. Byram Karasu, such patients are considered to have severe borderline and narcissistic disorders with dyadic deficits if they are arrested at step 1 (which leads to an impaired sense of self and impaired mother-child relations) or dyadic conflicts if they are arrested at step 2 (which leads to severely impaired object relations, such as difficulty of the child separating from the mother). Such fundamental impairments predispose to vulnerability to psychosis.

Problems at the second level involve severe personality disorders with negativism, disobedience, and lack of purposefulness, and empathy. They are described as mild borderline and narcissistic disorders characterized by problematic three-person (i.e., mother-child-father) relationships. Following Karasu, this group includes patients with triadic deficits if they are arrested at step 3 (which leads to poor impulse control and severe Oedipal problems, such as a lack of capacity for intimacy and social commitment) or triadic conflicts (arrested at steps 4 or 5, leading to milder Oedipal conflicts, such as inhibited sexuality or impaired internalization of group values). Such individuals have severe problems in working, socialization, and impulse-control.

Problems at the third level, with low conscientiousness, low self-confidence and resourcefulness, and little generosity, typify individuals with mild personality disorders or type 1 alcoholism and many with anxiety disorder or major depression, which are associated with mild problems in self-esteem, social intimacy, or group identification. Progress through level three reflects ego strength, which can be measured as the sum of TCI scores on self-directedness and cooperativeness, consistent with Freud's notions of mature personality organization.

The fourth and higher levels involve progressive steps in cognitive and spiritual development among socially mature individuals seeking fulfillment with better health and happiness. Such overall personality integration can be measured by the sum of scores on all three character dimensions, consistent with Carl Jung's notion of the self-transcendent leader. These higher levels of personality integration may be associated with psychopathology at times of existential crisis.

Practical Strategies Character development depends on both temperament configuration and prior development. Taking such information into account requires reliable and systematic assessments. This information provides guidelines for comprehensive treatment planning tailored for each patient. Early in treatment, nonlinear modeling can be used to evaluate possible character outcomes to predict alternative outcomes for given baseline conditions and interventions. These developmental observations indicate which techniques are likely to benefit the patient depends on the stage of personality integration. A brief outline of this integrative psychobiological approach is given in Table 24-38.

Table 24-38 A 15-Step Program for Integrative Psychobiological Treatment of Personality and Psychopathology

Treatment is organized into seven levels corresponding to the seven underlying problem levels described above. Levels 1, 2, 3, and (rarely) 4 are of particular interest to personality disorders.

The first treatment level involves supportive and reality-based techniques, including basic trust-building, encouragement, and teaching basic living skills (“walking together”). The goal is to ensure the safety of patients with severe disorganization and destructive impulses. Trust building and encouragement within safe limits, basic components of all therapy, are essential at this level because patients are filled with all the negative emotions, such as most patients whose temperaments are explosive (borderline) or whose characters are severely immature (melancholic or schizotypal). However, with psychotic or prepsychotic patients, therapy may be supportive, with such somatic therapies as electroconvulsive therapy or high doses of antipsychotic medication. The holding, or reparenting, aspects of the psychotherapeutic environment are crucial at this level, including a therapist who is dependable, nonretaliatory, and compassionate despite frequent crises, who has a more optimistic understanding of the patient's needs and opportunities than the patient. This allows the patient to build trust and self-confidence.

The second treatment level involves direction and cognitive-behavioral techniques that emphasize rational cognitive analysis and repetitive behavioral drills to improve discipline and impulse control (“working together”). This level involves initial treatment of individuals with moderately severe personality and mood disorders with poor impulse control and emotional lability (e.g., most antisocial, severe obsessional, or mild borderline patients). Pharmacotherapy for labile affect and impulsivity (e.g., mood stabilizers) and for hostility (e.g., low-dose neuroleptics) may be beneficial. In addition, firm structure and limit setting, direction, and advice are of critical importance at this level. Based on their improved understanding of their temperament structure and basic emotional needs, patients are encouraged to explore more mature ways of satisfying those needs. For example, hatha yoga or martial arts training are appealing ways to teach self-control at this level, leading to enhanced self-esteem and impulse control. Such structured approaches minimize conflict about setting limits but require much repetition of practical problem-solving skills and encouragement by discussion of attractive role models. This makes constructive use of such a patient's craving for pleasure and power.

The third treatment level involves nondirective, dynamic, humanistic, and interpersonal techniques to foster increased conscientiousness, resourcefulness, and generosity (“feeling together”). Direction is counterproductive at this stage because patients need to internalize group values and to develop confidence in their self-willed action. Depression and anxiety are common and may be treated with antidepressants as needed in combination with the nondirective therapy. The standard six exercises of autogenic training are useful and provide relaxation and preparation for meditative exercises at more advanced levels.

The fourth treatment level involves experiential and existential therapy techniques such as meditation and spiritual identification exercises (“reflecting together”). This involves conscious expansion of the self-concept to include a transpersonal (i.e., spiritual component in addition to the mind and body). Antidepressant or antianxiety medication may be useful during existential crises but are used sparingly and transiently because strong motivation is needed to promote the leaps of faith needed to transform consciousness and self-concepts.

Levels five through seven involve advanced guidance in meditation and other techniques needed to transcend ordinary self-concepts. Patients with such mature

personality development are often little impaired by ordinary standards, but they seek superior character integration, emotional fulfillment, and healthy longevity.

SUGGESTED CROSS-REFERENCES

Adult antisocial behavior is discussed in Section 27.3; personality disorder in elderly adults is discussed in Section 51.3g; theories of personality and psychopathology are discussed in Chapter 6; and impulse-control disorders are discussed in Chapter 22. Psychotherapy is discussed in Chapter 30, and Pharmacotherapy is discussed in Chapter 31.

CHAPTER REFERENCES

Allport G: *Personality: A Psychological Interpretation*. Constable, London, 1937.

Bayon C, Hill K, Svrakic DM, Przybeck TR, Cloninger CR: Dimensional assessment of personality in an outpatient sample: Relations of the systems of Millon and Cloninger. J Psychiatr Res *30:*341, 1996.

*Cloninger CR: A systematic method for clinical description and classification of personality variables. Arch Gen Psychiatry *44:*573, 1987.

Cloninger CR: The genetics and psychobiology of the seven factor model of personality. In *Annual Review of Psychiatry*, vol 17, K Silk, editor. American Psychiatric Association, Washington, DC, 1988.

*Cloninger CR, Przybeck TR, Svrakic DM, Wetzel R: *The Temperament and Character Inventory: A Guide to Its Development and Use*. Center for Psychobiology of Personality, Washington University, St. Louis, 1994.

Cloninger CR, Svrakic DM: Integrative psychobiological approach to psychiatric assessment and treatment. Psychiatry *60:*120, 1997.

*Cloninger CR, Svrakic DM, Przybeck TR: A psychobiological model of temperament and character. Arch Gen Psychiatry *50:*975, 1993.

Costa PT, McCrae RR: *The NEO Personality Inventory Manual*. Psychological Assessment Resources, Odessa, FL, 1985

Eysenck HJ: *The Biological Basis of Personality*. Thomas, Springfield, IL, 1967.

Gabbard G: Psychotherapy of personality disorders. J Pract Psychiatry Behav Health *3:*327, 1997.

Gunderson J, Philips C: Personality disorders. In *Comprehensive Textbook of Psychiatry,* ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Joffe RT, Bagby RM, Levitt AJ, Regan JJ, Parker JD: The Tridimensional Personality Questionnaire in major depression. Am J Psychiatry *150:*959, 1993.

*Johnson DL, Wiebe JS, Gold SM, Andreasen NC, Hichwa G, Watkins L, Boles Ponto LL: Cerebral blood flow and personality: A positron emission tomography study. Am J Psychiatry *156:*252, 1999.

Joyce PR, Mulder RT, Cloninger CR: Temperament predicts clomipramine and desipramine response in major depression. J Affect Disord *30:*35, 1994.

Kernberg O: *Borderline Conditions and Pathological Narcissism*. Aronson, New York, 1975.

Livesley WJ, Jang KL, Vernon PA: Phenotypic and genetic structure of traits delineating personality disorder. Arch Gen Psychiatry *55:*941, 1998.

Millon T: *Manual for the MCMI II,* ed 2. National Computer Systems, Minneapolis, MN, 1987.

Mulder R, Joyce P: Temperament and structure of personality disorder symptoms. Psychol Med *27:*99, 1997.

Perry C, Vaillant G: Personality disorders. In *Comprehensive Textbook of Psychiatry,* ed 5, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1989.

Reich J, Frances A: The structural interview method for diagnosing borderline disorders. Psychiatr Q *56:*229, 1984.

Schneider K: *Psychopathic Personalities*. Casell, London, 1958.

Soloff, P: Symptom-oriented psychopharmacology for personality disorders. J Pract Psychiatry Behav Health *1:*3, 1998.

Sjobring H: Personality structure and development: A model and its application. Acta Psychiatr Scand Suppl *244:*1, 1973.

*Svrakic DM, Whitehead C, Przybeck TR, Cloninger CR: Dimensional diagnosis of personality disorders by the seven factor model of temperament and character. Arch Gen Psychiatry *50:*991, 1993.

*Svrakic NM, Svrakic DM, Cloninger CR: A general quantitative theory of personality development: Fundamentals of a self-organizing psychobiological complex. Dev Psychopathol *8:*247, 1996.

Tome MB, Cloninger CR, Watson JP, Isaac MT: Serotonergic autoreceptor blockade in the reduction of antidepressant latency: Personality variables and response to paroxetine and pindolol. J Affect Disord *44:*101, 1997.

Zuckerman M: The psychobiological model for impulsive unsocialized sensation seeking: A comparative approach. Neuropsychobiology *34:*125, 1996.

CHAPTER 25. PSYCHOLOGICAL FACTORS AFFECTING MEDICAL CONDITIONS

25.1 HISTORY, CLASSIFICATION, AND CURRENT TRENDS IN PSYCHOSOMATIC MEDICINE

ALAN STOUDEMIRE, M.D., AND JOHN STEPHEN McDANIEL, M.D.

Although use of the word “psychosomatic” dates to the early 1800s, concepts of psychosomatic orientation in health beliefs have been traced to primitive societies in which spiritual powers and religion dominated the practice of medicine. Regular use of the term “psychosomatic” can be found in the nineteenth-century writings of the German psychiatrist Johann Christian Heinroth and the British psychiatrist John Charles Bucknil. Although the word “psychophysiologic” was in common use by the late 1800s and was used in association with the scientific investigations of Ivan Petrovich Pavlov and Walter Cannon, the term “psychosomatic” was rarely used prior to the 1930s. The phrase “psychosomatic medicine” was introduced by Felix Deutsch in 1922. Helen Flanders Dunbar firmly established the term and designated it a legitimate area of scientific investigation.

Psychosomatic medicine is an area of scientific investigation concerned with the relation between psychological factors and physiological phenomena in general and disease pathogenesis in particular. In a broader context, the term is applied to a philosophical approach to the care of patients that emphasizes the psychosocial aspects of medical care. These two meanings have predominated during most of the twentieth century. Conceptual problems with the word “psychosomatic” have been acknowledged (primarily based on the split in the mind-body relation implied by use of the compound word), but no other satisfactory term has gained complete acceptance, although many terms have been proposed, such as “integrative,” “holistic,” and “biopsychosocial.” Acceptance of the field by the medical community and the general public has been complicated by the pejorative use *psychosomatic*, referring to illness generated by persons who imagine or otherwise psychologically generate physical symptoms.

COMPARATIVE NOSOLOGY

In the American Psychiatric Association's diagnostic manuals, the term “psychophysiologic” formerly designated conditions in which psychological and emotional factors result in somatic symptom formation. In the second edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-II), published in 1968, such conditions were designated psychophysiologic disorders.

The third edition of DSM (DSM-III), published in 1980, saw a substantial change in terminology with adoption of the category psychological factors affecting physical conditions. The argument proposed by Robert Spitzer for abandoning the term “psychophysiologic” was based on eight major points: (1) the DSM-II term “psychophysiologic” was rarely used as a diagnosis; (2) the decision about whether a patient's condition is psychophysiologic or organic was arbitrary; (3) the term was believed to decrease collaboration between specialists; (4) the term perpetuated simplistic ideas about causes of diseases; (5) the term was often used as a last resort when previous efforts at medical diagnosis and treatment have failed; (6) the term referred only to causation and did not address the issue of how psychosocial factors may perpetuate or exacerbate illness; (7) no clear operational criteria were defined; and (8) the DSM-II classification of psychophysiologic disorders was inadequate for research purposes.

The DSM-III category was believed to offer advantages over the DSM-II terminology because it integrated psychological contributions to medical illness into a multiaxial diagnostic system. However, psychological factors affecting physical conditions was considered a category, not a specific diagnosis per se. By enlarging the scope of DSM-II psychophysiologic disorders, the framers of DSM-III hoped the category and multiaxial system would broaden the range of psychological factors considered to contribute to the onset or exacerbation of physical illnesses.

The DSM-III category was reevaluated in the process of developing the fourth edition of DSM (DSM-IV). To make the category more clinically useful, DSM-IV contains a subcategorization format that allows clinicians to specify the type of psychological or behavioral factor that affects the patient's medical condition; the factors so designated include the broad range of psychological and behavioral phenomena that appear to affect physical health (Table 25.1-1).

A. A general medical condition (coded on Axis III) is present.
B. Psychological factors adversely affect the general medical condition in one of the following ways:
(1) the factors have influenced the course of the general medical condition as shown by a close temporal association between the psychological factors and the development or exacerbation of, or delayed recovery from, the general medical condition.
(2) the factors interfere with the treatment of the general medical condition.
(3) the factors constitute additional health risks for the individual.
(4) stress-related physiological responses precipitate or exacerbate symptoms of a general medical condition.
Choose name based on the nature of the psychological factors (if more than one factor is present, indicate the most prominent):
Mental disorder affecting medical condition (e.g., an Axis I disorder such as major depressive disorder delaying recovery from a myocardial infarction)
Psychological symptoms affecting medical condition (e.g., depressive symptoms delaying recovery from surgery; anxiety exacerbating asthma)
Personality traits or coping style affecting medical condition (e.g., pathologic denial of the need for surgery in a patient with cancer; hostile, pressured behavior contributing to cardiovascular disease)
Maladaptive health behaviors affecting medical condition (e.g., lack of exercise, smoke ses, overeating)
Stress-related physiological response affecting general medical condition (e.g., stress-related exacerbations of ulcer, hypertension, arrhythmia, or tension headache)
Other unspecified psychological factors affecting medical condition (e.g., interpersonal, cultural, or religious factors)

Reprinted with permission from the American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 25.1-1 DSM-IV Diagnostic Criteria for Psychological Factors Affecting Medical Condition

The DSM-IV strategy continues to allow the designation of comorbidity between Axes I and II and Axis III conditions and specific subtyping of multiple factors that can affect the onset, exacerbation, and outcome of physical disorders. The revised category attempts to embrace the basic concepts involved in such terms as “psychosomatic,” “psychophysiological,” “psychobiological,” “integrative,” “holistic,” and “biopsychosocial.”

The 10th revision of *International Statistical Classification of Diseases and Related Health Problems*, published in 1992, classifies psychosomatic disorders as “psychological and behavioural factors associated with disorders or diseases classified elsewhere,” referring to factors that influence the manifestation or alter the course of physical illnesses. Additional codes are used in this system to identify specific disorders, including asthma, dermatitis, gastric ulcer, mucous colitis, ulcerative colitis, and urticaria. This system, which complements the DSM-IV classification of psychosomatic disorders, may become increasingly important in the future if is adopted worldwide (Table 25.1-2).

This category should be used to record the presence of psychological or behavioral factors thought to have influenced the manifestation, or affected the course, of physical disorders that can be classified using other chapters of ICD-10. Any resulting mental disturbances are usually mild and often prolonged (such as worry, emotional conflict, apprehension) and do not of themselves justify the use of any of the categories described in the rest of the book. An additional code should be used to identify the physical disorder. (In rare instances in which an overt psychiatric disorder is thought to have caused a physical disorder, a second additional code should be used to record the psychiatric disorder.)

Reprinted with permission from *The ICD-10 Classification of Mental and Behavioral Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 25.1-2 ICD-10 Diagnostic Criteria for Research (DCR-10) for Psychological and Behavioral Factors Associated With Disorders or Diseases Classified Elsewhere

HISTORY

Early Theories Mind-body concepts date to ancient beliefs that the body can be affected by external forces. In prehistoric times, disease was understood to be caused by evil spirits or forces, and the concept of disease as a result of divine intervention to test or punish the guilty or ungodly is well entrenched in Judeo-Christian religion as reflected in the Book of Job and the exorcisms practiced by Jesus in the *New Testament*. What are now understood to be mental disorders were believed to result from possession by evil spirits or demons that could be cast out by exorcistic rites or destroyed by killing the possessed persons. Alternatively, one could be cured or cleaned of illness by supplication, sacrifice, prayer, or ritualistic appeasement of the gods. Belief in the power of prayer and other religious and spiritual rituals to invoke divine intervention to alter the forces of disease remain a component of most religious belief systems.

The roots of intellectual approaches to psychosomatic theory date to the ancient Greeks. Mind-body relations were noted by Hippocrates, Aristotle, Plato, and Socrates. Theodore Brown noted that the doctrine of the passions was a basic component of both classical and neoclassical medical theory; it held that relations exist between emotional states (anger, fear, joy, love) and physical states. Hippocrates established ancient psychosomatic concepts based on the humoral theory, and Galen subsequently validated emotions as important in the pathogenesis of disease.

Although psychosomatic scholars have incriminated René Descartes for creating the mind-body dualism that prevented the development of integrated models of psychosomatic relations in the evolution of modern medicine, that conclusion may be a misunderstanding of Cartesian theory. Careful analysis of Descartes' ideas reveals that he did not rigidly split the mind from the body, but regarded mental forces as interacting with the physical. Brown proposed that rather than denying theoretical mind-body interactions, Descartes actually facilitated them. He specified a location, the pineal gland, where mind and body regularly and readily interact, and he provided a logical basis for the extensive interaction of affective states with bodily conditions by explaining that most aspects of affective states are primarily somatic. Descartes' views provided the impetus for the replacement of the passion-based neoclassical understanding of mind-body relations by more scientific views of illness (Table 25.1-3).

Table 25.1-3 History of Psychosomatic Medicine

Psychoanalytic Theories Many psychosomatic concepts in modern medicine and psychiatry can be traced to Sigmund Freud, whose psychoanalytic formulations and theories seemed to confirm the role of psychological factors in causing and determining the manifestations of such conversion reactions as paralysis and blindness (Table 25.1-4). Neo-Freudians elaborated and extrapolated his basic theory, perhaps prototypically represented by Sandor Ferenczi, who applied the notion of conversion phenomenon to the autonomic nervous system and used the theory to explain such diseases as ulcerative colitis.

Table 25.1-4 Major Conceptual Trends in 20th-Century Psychosomatic Medicine

The psychoanalytic tradition in psychosomatic theory culminated in the work of Franz Alexander. He conceptualized hysterical conversion as caused by unconscious conflicts between aggressive or dependent wishes and opposing effects of the ego and superego, leading to chronic emotional distress. Specific types of psychodynamic conflicts were purported to cause specific types of diseases from stress-induced autonomic and physiological distress. Alexander applied the theory to seven illnesses: essential hypertension, rheumatoid arthritis, thyrotoxicosis, peptic ulcer, ulcerative colitis, bronchial asthma, and neurodermatitis.

Although Alexander's psychosomatic specificity theory has been largely discarded, his specificity theory was but one component of his overall approach. Careful analysis of his work reveals that Alexander believed that psychosomatic illness always occurred in the context of multiple interacting variables, such as biogenetic vulnerabilities and social environmental influences. Alexander developed a biopsychosocial model that was overshadowed by psychoanalytic emphasis on the “specificity” aspect of his model.

Recent Theoretical Approaches A parallel theory was developed by Adolph Meyer, whose work continues to influence mainstream psychiatry and psychosomatic theory. Meyer's psychobiology was an inherently multifactorial approach to the cause and the treatment of mental illness. His psychobiological approach attempted to integrate mind and body into a psychobiological unit; he proposed studying psychological and biological processes as dynamic interacting systems. Meyer's approach was eclectic and encouraged evaluating each aspect of the patient's life—developmental, psychological, social, environmental, and physiological.

One of Meyer's followers, Dunbar, in 1935 coalesced the literature and compiled reports of psychosomatic reactions observed by clinicians over a 23-year period. In 1943 she postulated the first formulation of what is now known as the type A personality theory and theorized other personality profiles associated with psychosomatic illnesses. Dunbar established the American Psychosomatic Society and is acknowledged as the founder of modern psychosomatic medicine. The continued efforts of many researchers to identify specific personality attributes that cause or predispose to physical illness directly relates to her original contributions.

The psychophysiological approach in psychosomatic theory dates to Pavlovian studies of conditioned physiological responses in animals and has continued through current biofeedback and behavioral physiological research. Cannon was among the first scientists to demonstrate the physiological sequelae of emotional states and to study the influence of the autonomic nervous system in modulating such responses. The classic study by William Beaumont examining the variations in gastric fluid

induced by emotional states in a patient with a gastric fistula, formed the basis for such later observations as those by the internist and psychoanalyst George Engel, who coined the term “biopsychosocial.”

Stress Theory The concept of stress has been central in the development of psychosomatic theory. Cannon conducted the first systematic study involving stress paradigms relevant to psychosomatic medicine. His model of stress was derived from physics: under certain circumstances, physical or emotional stimuli can strain an animal beyond its ability to adapt successfully.

The research of Harold Wolff and Stewart Wolf continues to serve as a model for scientific investigations of stress. One of Wolff's fundamental premises was that disease is a failure or inability to adapt to life stress. Wolff's theory heralded the concept that the way in which a person copes with a stressful event is critical in determining the magnitude of subsequent physiological effects. Events are deemed stressful only if the person perceives the stress as threatening to life, well-being, or emotional security.

Wolff and Wolf also observed that the physiological states of the gastrointestinal tract appear to correlate with specific emotional states (hyperfunction with hostility and hypofunction with sadness). Nevertheless, they regarded such reactions as relatively nonspecific, believing that the patient's reaction is determined by the general life situation and perceptual appraisal of the stressful event. Wolff also emphasized that the capacity to adapt to a threatening event (e.g., familial discord, emotional deprivation, goal frustration, object loss, separation, and unemployment) determines the nature and severity of psychophysiological response patterns.

Hans Selye's model of stress was based on experiments in which toxic substances were injected into animals and the subsequent reactions observed. Selye outlined the general adaptation syndrome as consisting of three phases: (1) the alarm reaction; (2) the stage of resistance, in which adaptation is ideally achieved; and (3) the stage of exhaustion, in which acquired adaptation or resistance may be lost. Selye's concept of stress was originally used to describe the effects of a force acting against an organism's innate resistance. He considered stress a nonspecific bodily response to any demand caused by either pleasant or unpleasant conditions. Selye's basic theory that physical and emotional stressful stimuli can produce relatively predictable responses formed a fundamental model for much of the behaviorally oriented stress experimentation in psychosomatic research that followed.

Selye's concept of a single nonspecific physiological response to stressors was superseded by experiments demonstrating complex adaptational reactions to stress. For example, John Mason performed elegant research in psychoendocrinology demonstrating that in experimental stress situations, physiological reactions are largely influenced by the emotional response to the stimuli—that is, not determined just by the nature of the stimulus itself. Mason focused on intervening psychological variables that determine both acute hormonal reactions to stressful stimuli and the effects of psychological variables on long-term adaptational responses.

Richard Lazarus elaborated the concept of individualized stress responses, which are determined by cognitive factors. He proposed that responses are determined by the manner in which a person cognitively appraises and copes with stressful events. Hence, persons' reactions to stress depend on their appraisal of the event and their belief in their ability to cope with the stress; their attitude regarding the significance of the outcome of the event is also considered important. The reality of the stress is of less importance than an individual's subjective cognitive assessment of it in determining the subsequent emotional and physiological reactions.

MEASURES OF STRESS A major area of psychosomatic research involving social stress is based on the work of Thomas Holmes and Richard Rahe, using the Schedule of Recent Experience (SRE). Recent life events (e.g., the death of a close relative, a job change, or a divorce) are assigned life change units. An expanded and revised version of the SRE, the Recent Life Change Questionnaire, also asks subjects to score recent life changes on the degree of perceived adjustment. Both instruments are designed to measure recent life stress and to correlate the degree of life stress with subsequent illness. The basic hypothesis is that stressful life occurrences are risk factors for the development of physical illness. Numerous studies demonstrate a probability relation between stressful life events and one's chances of having a physical illness. Epidemiological research involving other types of stressful life events (e.g., natural disasters, social disruption, social changes, poor social support, and work-related stress) has shown similar trends toward an increased probability of physical illness and relatively poor medical outcome. Current models of psychosomatic research integrate the interaction of psychological variables, stressful social situations, and biological vulnerability with latent physical disease.

X-Y-Z Model. The National Academy of Sciences' Institute of Medicine set forth a definition of stress in an attempt to present a scientifically integrated model to aid in mind-body research. The three major elements of environmental stress are identified as the environmental activator, the reaction to the activator (stress reaction), and the resulting consequences. The series is called the x-y-z sequence.

An *activator* is defined as any internal or external event or condition that can alter a person's physical or psychological state. Activators can be quantified and qualified on the basis of intensity, duration, frequency, and the part of the body they act on (i.e., globally or molecularly). If an activator becomes a stressor, its effect is considered sufficient to cause a significant physical or psychological reaction. A *reaction* to the activator (stress reaction) can also be quantified and described physiologically and psychologically. Reactions can also be characterized as transient responses to specific activators. *Consequences* are the prolonged effects of the reactions. Consequences may be biological, psychological, and sociological. *Mediators* are qualities and characteristics of the person that account for the variability in the psychological and physiological reaction-consequences sequence. Mediators may be qualities (physical or psychological) of the person, the environment, or the social milieu that provide buffering (protective effects) or that impart vulnerability.

An essential aspect of the x-y-z model that has proved helpful in conceptualizing stress paradigms is the concept of *dynamic interaction* between the various components of the stress-response system. Each part of the system interacts with and modifies the other parts. For example, mediators may facilitate each aspect of the x-y-z sequence, but change and modification may result because of past responses and the success or failure of efforts to adapt.

The main value of the x-y-z model derives from the possibility of identifying *risk factors*, characteristics that appear to predispose a person to an illness. Risk-factor research is in many respects basic probability assessment. In general, because of difficulties in the definition and measurement of stress and because of individual variability in responsiveness, the association between stress-related variables and physical illness has remained statistically weak.

DEVELOPMENTAL PSYCHOBIOLOGY

Although not traditionally considered part of mainstream psychosomatic medicine, basic science research examining the biological effects of attachment and traumatic separation on developing animals and humans forms a strong scientific underpinning for the validity of many traditional psychosomatic concepts. For example, a long tradition in psychosomatic medicine has associated the onset of physical illness with the stressful effects of bereavement and loss. From his anecdotal clinical experience, Engel popularized the notion of the “giving up–given up syndrome.” Psychosomatic theory has long held that losses that are poorly coped with or that otherwise overwhelm a person's adaptive capabilities may cause a breakdown in resistance to illness and precipitate the onset or the exacerbation of illnesses ranging from diabetes to cancer. Epidemiological data also indicate increased rates of mortality among the bereaved.

As far back as Richard Burton's *Anatomy of Melancholy*, published in 1621, traumatic separations and losses first described by Reneé Spitz have been deemed the cause of psychiatric illness. Anaclitic depression in infants deprived of maternal nurturing leads to failure to thrive and death. In a similar phenomenon described in infant primates, a period of frantic searching for the mother is followed by a depressive response characterized by sad facial expressions, slouched posture, cessation of play, withdrawal, and other social changes. These observations, including the classic primate-separation experiments by Harry Harlow, have led to research examining the behavioral, social, and biological effects of separation and loss.

Physical Correlates. Through studying traumatic separation in primates, investigators have found that agitated states of loss are associated with elevations in heart rate and body temperature. The agitation phase of the separation response is marked predominantly by activation of the sympathetic adrenergic system and adrenocortical response with increased serum cortisol concentration. Almost opposite effects are observed in the bereavement period, which is associated with diminished central b-adrenergic tone and a net decrease in central b-adrenergic activity. The depressed state is typified by fragmented circadian rhythms and sleep, increased cardiac arrhythmias, and decreased immune responsiveness. The reaction has been described as a compensatory homeostatic response to adrenergic activation in the agitated-protest phase of a separation experience.

The involvement of biogenic amine systems in such reactions is further suggested by the observation that behavioral responses to separation can be modified by treatment with pharmaceutical agents. For example, in separation experiments infant rhesus monkeys given imipramine (Tofranil) display less passivity and more interest in the environment than do placebo-treated animals. Similar studies have been performed in rodent models.

Data from a variety of psychobiological investigations confirm the effects that early experience can have on the neurophysiology of the developing organism. Experiments have shown that postnatal handling changes subsequent behavioral and neuroendocrine responses to stress. Research showing that rats previously subjected to stressful handling experiences show attenuation of fear reactions in strange or new environments and less-pronounced increases in the secretion of adrenal glucocorticoids has led to research examining how early handling experiences may influence the development of certain brain regions that regulate glucocorticoid negative-feedback inhibition of hypothalamic-pituitary-adrenal axis activity. Handling increases glucocorticoid receptor density in the hippocampus and the frontal cortex, increasing the sensitivity of these neuronal structures to the negative-feedback effects of higher glucocorticoid concentrations, thereby increasing the efficiency of neuronal inhibition of adrenocorticotropic hormone (ACTH) secretion. It is well established that children who are physically or sexually abused develop enduring dysfunction of their neuroendocrine and immune systems.

Current Advances in Psychosomatic Medicine Historically and in current research, the essential challenge in psychosomatic-psychobiological research is to delineate the mechanisms by which experiences cause certain types of physiological reactions that result in disease states. In addition, the mechanisms of neurobiological modulating systems must be understood in molecular and genetic terms. A host of research advances in psychosomatic medicine have delineated the role of stress in psychological states, the role of behavior in immunity, and the complex interactions of psychological conditions and various disease conditions.

Psychocardiology Psychocardiology encompasses the spectrum of interactions of psychiatric disorders, cardiac symptoms, and cardiac disease, including associated complicating health behaviors. For the past several decades attention to the psychosocial and behavioral factors in cardiovascular disease has increased significantly. The research has taken two primary pathways: one has examined hypertension and the other has looked generally at coronary artery disease, including myocardial infarction and sudden cardiac death.

Hypertension Hypertension, one of the originally hypothesized psychosomatic illnesses, is a major risk factor for coronary artery disease and cerebral vascular disease. Psychological factors have been closely studied as part of the pathogenesis of the condition; these factors have been categorized as pressure reactivity and personality and behavioral factors. Physiological hyperreactivity to environmental stimuli (pressure reactivity) has been studied many times, and although results are contradictory, some specific commonalties appear. Relatively strong evidence indicates that some persons have greater blood pressure reactivity than do others to a variety of stressors, ranging from experimental stress induced in the laboratory to such social and societal stressful conditions as racism. However, the evidence linking reactivity in normotensive persons with hypertension is equivocal. Perhaps most important, pressure reactivity in hypertensive individuals may exacerbate and even accelerate the disease process.

Other research examining psychological aspects of hypertension has focused on personality traits or coping styles. Traits such as submissiveness and distorted expression of anger have emerged as correlates of hypertension. The most consistent correlates have involved anger-coping styles, both inhibited anger expression and excessive anger expression. Epidemiological researchers have noted that persons using an active coping style under environmental conditions that are not conducive to success (e.g., low education and socioeconomic status) may be predisposed to hypertension.

As evidence has clarified how psychological factors affect hypertension, investigators have focused on treatment interventions. Various behavioral procedures including biofeedback, relaxation training, and psychotherapy have been used as interventions. Some investigators have reported clinically significant success in controlled studies. For example, studies that used 24-hour monitoring of blood pressure to examine the effects of combined relaxation therapy and medication found the combination to be more effective in controlling blood pressure than medication alone.

Coronary Artery Disease

STRESS Stress causes a sympathicoadrenal medullary alarm reaction characterized by excess catecholamine secretion. Specifically, excess epinephrine is secreted under what the body interprets as stressful conditions. The outpouring of epinephrine raises blood pressure and heart and respiratory rates, enhances neuromuscular transmission, elevates the concentration of blood sugar by glycogenolysis, mobilizes fat, redirects hemodynamic patterns to suit muscular activity, and while increasing blood oxygenation, increases oxygen consumption. More-specific b-adrenergically mediated cardiac effects include increased heart rate, contractility, and conduction velocity and a short arteriovenous refractory period. These catecholamine-mediated cardiac effects are thought to be pathogenically related to adverse cardiac events.

Studies of stress-induced cardiac changes and studies examining stress, arrhythmias, and sudden cardiac death suggest a significant relation in the pathophysiology of coronary artery disease. Early researchers examined temporally related stressful life experiences including stressful states described among patients who had experienced sudden death attributed to arrhythmias. One study found that the cardiac surgery patients at greatest risk for complications, including arrhythmias and sudden death, were depressed, anxious and in denial of their anxiety, or both.

Recent research has examined the direct cardiac effect of controlled stress. For example, in a study of the effects of psychological stress on patients with ventricular arrhythmias, the stress of mental arithmetic and recalling past traumatic events increased the ventricular premature beat frequency in most patients. Other studies have documented diminished cardiac perfusion during mental stress via positron emission tomography and radionucleotide ventriculography in patients with coronary artery disease. One group of investigators who also used 48-hour Holter monitoring examined mental stress–induced myocardial changes in patients with coronary artery disease and found that ischemia was associated with significantly higher rates of subsequent fatal and nonfatal cardiac events independent of age, baseline left ventricular ejection fraction, and previous myocardial infarction.

TYPE A BEHAVIOR Most of the studies examining the influence of psychosocial and behavioral risk factors in the etiology of coronary artery disease have focused on type A behavior (Fig. 25.1-1). Individuals with type A behavior exhibit enhanced aggressiveness, ambitiousness, competitive drive, impatience, and a chronic sense of time urgency. Associated speech and motor characteristics are rapid body movements, tense facial and body musculature, explosive conversational speech, and hand or teeth clenching. Three major prospective studies have found the type A behavior pattern to be a risk factor for clinical coronary artery disease; however, once the disease is present, type A behavior does not appear to increase the risk of subsequent cardiac morbidity.

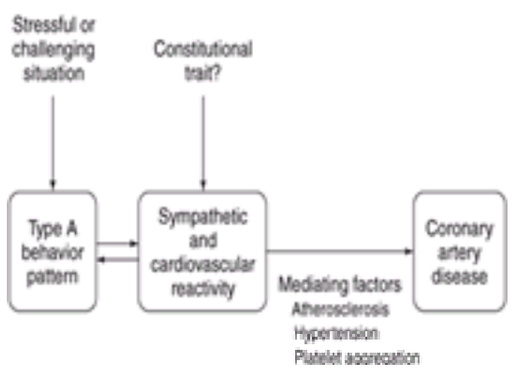

FIGURE 25.1-1 Conceptual model of type A behavior and the development of coronary artery disease. (Adapted from Goldstein MG, Niaura R: Cardiovascular disease, Part I: Coronary artery disease and sudden death. In *Psychological Factors Affecting Medical Conditions,* A Stoudemire, editor. American Psychiatric Press, Washington, DC, 1995.)

The initial enthusiasm for the global type A concept waned in the middle 1980s, as hostility was found to be the more toxic element of the syndrome. Whereas global type A behavior does not always predict risk of coronary artery disease, hostility is consistently linked to coronary artery diseases and appears to be pathophysiologically related to the disease by numerous mechanisms. Expressive hostility and antagonistic interactions appear to be the subcomponents of type A behavior that are most strongly related to the risk of coronary artery disease, especially among female patients and middle-aged male patients.

Physiological correlates of type A behavior have also been studied with regard to cardiac morbidity. Type A behavior is believed to be part of a stress paradigm. Numerous studies report that persons with type A behavior patterns display large, episodic increases in blood pressure, heart rate, and catecholamine concentrations

when confronted by stressful tasks. Interestingly, evidence from primate studies links atherosclerosis in coronary disease and sympathetic nervous system activation. Findings suggest a link between psychological states, physiological activity, and subsequent cardiovascular disease.

A meta-analysis of 18 controlled studies examining effects of psychological treatment on type A behavior concluded that psychological treatment aimed at reducing the behavior had a positive outcome. When type A behavior decreased, treatment had a significant improved effect on coronary events and mortality at 1-year follow-up.

A recent study by Dean Ornish revealed that lifestyle modification (e.g., aerobic exercise, stress management, group therapy) actually promoted revascularization, and improved the prognosis of patients with cardiac disease.

MOOD STATES A number of mood states have been linked with cardiovascular disease, and an increasing body of literature links major depression to increased morbidity and mortality in individuals with coronary artery disease. Recent investigations using standardized instruments and current diagnostic categories have found the point prevalence of major depression in patients with coronary artery disease to be between 17 and 22 percent—approximately twice that in primary care patients and three to four times the 30-day point prevalence in the general population. One study showed that a major depressive episode was the best predictor of major cardiac events during the 12 months after cardiac catheterization; the events were independent of such variables as severity of cardiac disease, left ventricular function, and smoking. In a recent prospective study of patients hospitalized following a myocardial infarction, major depressive disorder was found to be an independent risk factor for mortality at 6 months. The risk was at least equivalent to that of left ventricular dysfunction and previous myocardial infarction. When these patients were studied 18 months after their hospitalization for myocardial infarction, depression was a significant predictor of cardiac mortality, particularly among patients with 10 or more premature ventricular contractions per hour.

Increasing evidence of an association between depressive disorders and coronary artery disease makes it important to ascertain the strength of this association because of the important practical health implications. Therefore, researchers have assessed both the persistence of this relation over time and its magnitude. In most patients with coronary artery disease, depressive disorders do not appear to be a transient adjustment to a major illness, but rather a clinically significant comorbid disorder that is often chronic. Moreover, recent studies have shown that among individuals with coronary artery disease, women appear to have a higher prevalence of depressive disorders than men. The impact of depressive disorders on poor outcome of coronary artery disease appears to equal or exceed that of other well-known cardiovascular risk factors. Current explanations for this relation include biological correlates related to sympathetic activity and autonomic dysregulation, neuroendocrine-immunological interactions producing cytokine-mediated coronary artery occlusion including serotonin-mediated platelet aggregation, and behavioral factors related to the effect of depression on cardiac treatment adherence. The importance of aggressive treatment of depression in coronary artery disease is clear, and research examining treatment benefits and prevention strategies is under way.

A convincing body of literature focuses on the mood states related to acute situational disturbances. A number of studies have demonstrated a connection between sudden death and acutely disturbing life events. One study found that the onset of malignant ventricular arrhythmia was associated with identifiable emotional triggers in 21 percent of patients referred for antiarrhythmic treatment. Finally, such sociological factors as work overload and life stress in the face of a lack of social support enhance general coronary risk (Table 25.1-5). Research is increasing into the possibility that central or autonomic nervous system events or both mediate the increase in cardiovascular morbidity and mortality.

Risk Factor	Cardiac Effects
Stressful life experiences	Sudden death, myocardial ischemia, arrhythmias
Type A behavior pattern	
Expressive hostility	
Antagonistic interactions	Risk of coronary artery disease; episodic increases in blood pressure, heart rate, and catecholamine levels (linked to atherosclerosis in primates)
Cynicism	
Hostile affect	
Aggressive responding	Significant relationships to cardiac mortality
Psychiatric diagnosis	
Major depression	Significant predictor of morbidity after cardiac catheterization; significant predictor of cardiac mortality up to 18 months following myocardial infarction may increase platelet aggregation and thrombus formation
Anxiety	Enhanced risk of cardiac morbidity
Sociocultural factors	
Work overload and life stress	
Lack of social support	Enhanced risk of cardiac morbidity

Table 25.1-5 Psychological Factors Related to Coronary Artery Disease

Gastrointestinal Conditions Historically, the interactions of emotions and the gastrointestinal system have been predominant in psychosomatic theory and research. Alexander identified emotional factors believed to be affect disturbances of appetite and eating, swallowing, digestive functions, and eliminative disorders. Although most studies examining psychological factors and the gastrointestinal tract have been methodologically flawed, a few have important historical significance. For example, in 1957, a double-blind prospective study of military inductees measured serum pepsinogen concentrations and psychological profiles to examine the role of psychological factors in causing peptic ulcer disease. A profile characterized by unresolved conflicts about dependency and oral gratification and relatively high pepsinogen secretion reliably predicted a risk for subsequent development of peptic ulcers; only high secretors developed ulcers, and combined measures were the best predictors. Persons with a genetic predisposition to secrete high levels of pepsinogen were considered at risk for peptic ulcer if they were psychologically vulnerable and were placed in situations that activated the conflict.

Peptic ulcer disease, a condition now etiologically linked to heterogeneous factors affecting mucosal integrity, including *Helicobacter pylori* infection, continues to be a primary focus of psychosomatic investigation. The corrosive action of gastric acid was once widely accepted as a complicating factor in peptic ulcer disease, as affecting illness course and outcome. Recent studies have shown that personality features of hostility, irritability, hypersensitivity, and impaired coping ability correlate significantly with serum pepsinogen concentration in peptic ulcer disease patients. The incidence of cigarette smoking, alcohol intake, and aspirin ingestion is also higher in peptic ulcer disease patients but is not related to coexistent psychopathology. The fact that antibiotic treatment of *H. pylori* is effective in gastric and peptic ulcer disease brings into serious question previous theories and studies linking psychological factors, acid secretion, and ulcerative stomach disease. However, there remains evidence for other mediators, such as stress, given the fact that more than 80 percent of *H. pylori*-infected people never develop an ulcer.

Although the relation between psychological factors and inflammatory bowel disease (ulcerative colitis and regional enteritis) remains unclear, irritable bowel syndrome, first described in 1820, is the most frequently documented gastroenterological syndrome that has been related to psychiatric influences. Irritable bowel syndrome is defined by a cluster of bowel syndromes, which make classification of the disorder difficult; predominant symptoms are abdominal pain and change in bowel habit. If irritable bowel syndrome is conceptualized as a condition in which the gut is hypersensitive and hyperactive to mechanical and chemical stimuli, it may be comparable with asthma in the respiratory system. The psychophysiological disturbances found with irritable bowel syndrome suggest an underlying biological predisposition toward abnormal electrical rhythms in the smooth muscle of the rectum and rectosigmoid colon. Irritable bowel syndrome is presumed then to be associated with disturbances in motility and be influenced by psychological factors in individuals whose symptoms tend to recur in a chronic, relapsing, and remitting course. Individuals with irritable bowel syndrome have significantly greater reactivity to stress than matched controls with abdominal pain without irritable bowel syndrome. Moreover, environmental stress, life events, and psychological factors are now known to be associated with colonic physiology.

A strong, consistent association between functional gastrointestinal disorders and psychological factors suggests the need for well-controlled psychological interventions other than prescribing drugs or diets. To date, intervention studies have shown benefit from supportive psychotherapy, behavioral interventions (e.g., hypnotherapy, stress management) and education. However, the long-term efficacy of these interventions needs further study. Low dosages of tricyclic antidepressants have also resulted in symptom improvement. Nonetheless, the interrelatedness of behavioral factors and gastrointestinal conditions continues to offer insight into the complex interactions of the mind and body.

Psychoneuroimmunology In 1964 the term “psychoimmunology,” later amended to *psychoneuroimmunology*, was coined to identify the field of investigation directed at understanding how behavior and mental states affect immune function. Robert Ader's landmark studies of classical conditioning laid much of the foundation for this area of specialty. His research paired specific hemagglutinating antibody titers with nausea induced by an immunosuppressant agent. By carefully controlling all possible alternative explanations, Ader demonstrated a brain–immune system influence, an effect now replicated in humans.

In recent years major advances have been made in establishing direct evidence for a bidirectional brain–immune system relation. Evidence for such a communication system ranges from anatomical confirmation of central nervous system (CNS) innervation of immune organs to reports documenting behavioral effects on immune response and tumor acquisition in experimental animals. A number of studies involving experimental animals and human autopsy specimens have shown direct sympathetic nervous system innervation of the spleen, thymus, and lymph nodes. Cholinergic innervation of the thymus gland has been documented, and investigators have described noradrenergic innervation of lymphoid tissue in a variety of mammalian species, including humans. Lending further evidence is the presence of receptors for various neurotransmitters, neurohormones, and neuropeptides on cells of the immune system. Recent evidence indicates that lymphocytes may synthesize some neurohormones de novo.

Lymphokines Studies in both humans and animals indicate that activation of the hypothalamic-pituitary-adrenal axis is associated with immune activation. Just as cells of the peripheral immune system are influenced by circulating factors of CNS origin, a reciprocal relation is demonstrated by the brain's susceptibility to lymphokines (substances produced by lymphocytes). Interleukin-1, a widely studied lymphokine, plays a major role in the initiation of events leading to T-cell responses. Interleukin-1 (IL-1) directly stimulates the hypothalamus to produce corticotropin-releasing factor (CRF). Stimulation of CRF is a physiologically important part of an inhibitory feedback loop that regulates the immune response. CRF increases ACTH secretion, which in turn increases cortisol secretion by the adrenal glands, which then inhibits immune function. Glucocorticoids have well-known immunosuppressive effects, including the reduction of circulating T-cell subsets. The inhibitory feedback loop is of particular interest to researchers studying mood disorders who have documented increased hypothalamic CRF production among some populations of depressed patients. Aldesleuken, a human IL-2 derivative, has proven to be an important component of immunotherapy for certain cancers such as malignant melanoma. Hence, augmentation of immunity shows promise as an effective strategy for cancer treatment.

Behavioral States Considerable psychoneuroimmunology research has examined behavioral states and associated in vitro immune suppression; however, these studies are correlational and do not imply direction of causality. Investigators have examined various forms of life stress to ascertain associated changes in immunity, and although most have examined depressed states, researchers have examined a variety of stressful conditions, including laboratory stress, examination stress, loneliness, divorce, caregiver stress, and bereavement. All of these behavioral states have been associated with suppression of markers of immune function, usually measured by lymphocyte responses to antigens and natural killer (NK) cell activity (Table 25.1-6).

Disturbed sleep function
Examination stress
Loneliness
Unemployment
Marital discord
Divorce
Alzheimer's disease–caregivers' stress
Bereaved spouses (including anticipatory bereavement)
Clinical anxiety
Major depressive disorder

Table 25.1-6 Behavioral States Associated With In Vitro Immune Suppression

A growing number of reports have provided data on depression and the immune system, a topic of considerable public interest in recent years. However, the studies have led to disagreement and confusion regarding conceptualizations, methods, experimental designs, and results. Alterations in the immune system in major depressive disorders do not appear to be specific biological correlates of depression but occur in association with other variables that characterize depressed patients, including age and symptom severity. A recent study found significant age-related differences between depressed patients and controls. Specifically, the depressed patients did not show increases in lymphocyte function or in the number of T-helper lymphocytes with advancing age, as did the controls, suggesting diminished immune functioning.

Psycho-Oncology The growing interface between the social sciences and oncology between the 1930s and 1950s was the foundation for emergence of the subspecialty called psychosocial oncology or psycho-oncology. This area, which has become a major contender in psychosomatic medicine research, seeks to study both the impact of cancer on psychological functioning and the role that psychological and behavioral variables may play in cancer risk and survival. A hallmark of psycho-oncology research has been intervention studies that attempt to influence the course of illness in patients with cancer. While most psychosocial interventions are aimed at providing psychological support rather than direct treatment of comorbid conditions such as major depressive disorder, important clinical observations have emerged that focus on such variables as cancer outcome and psychoneuroimmunology. For example, one landmark study demonstrated that women with metastatic breast cancer who received weekly group psychotherapy survived an average of 18 months longer than control patients randomly assigned to routine care. In their study of patients with malignant melanoma, one group of investigators found that control patients who did not receive a structured group intervention had a statistically significant recurrence of cancer and a greater mortality rate than patients who did receive such therapy. The malignant melanoma patients who received the group intervention also exhibited a significant increase in the number of large granular lymphocytes and NK cells as well as indications of increased NK cell activity. Another group of investigators used a group behavioral intervention (relaxation, guided imagery, and biofeedback training) in patients with breast cancer to demonstrate increased NK cell activity and lymphocyte mitogen responses in patients receiving treatment compared with controls.

Increasingly, intervention studies in psycho-oncology are focusing on complex variables of disease outcome including neuroendocrine and immune parameters. Investigators have found that psychosocial variables such as lack of social support and depressive symptoms may be linked to reduced NK cell activity in women with breast cancer and that more metastatic nodes and decreased NK cell activity are associated with depressive symptoms, emphasizing the need for more research in the area. Addressing psychoneuroimmunological aspects of depression in patients with cancer may have important treatment implications, particularly regarding hypothalamic-pituitary-adrenal axis hyperactivity associated with depression. In fact, hypothalamic-pituitary-adrenal hyperactivity in depressed cancer patients may have important prognostic implications, particularly since hypothalamic-pituitary-adrenal hyperactivity induced by exposure of rats to stress is associated with increased tumor growth, especially in older rats. Psychosocial treatment of patients with malignant melanoma has been shown by Fawzy Fawzy to improve prognoses and enhance certain immune parameters.

Psychoneuroendocrinology Selye first described the adaptation syndrome in rats exposed to chronic stress, a response characterized by adrenal hypertrophy, gastric ulcers, and thymus and lymph node involution. Since then, psychoneuroendocrinology, the subspecialty that addresses the relation between hormones and behavior, has continued as an area of study. While earlier researchers in psychosomatic medicine were more interested in endocrine disturbances that produce psychiatric syndromes (Cushing's syndrome, hyperthyroidism, hypothyroidism), exogenous hormone–induced syndromes (glucocorticoid administration), and psychiatric aspects of other endocrine abnormalities (diabetes mellitus), more-recent research has focused on the multifaceted role of neuroendocrine modulation and its relation to behavioral disturbances and psychiatric symptomatology.

For example, exposure of rats to stress reliably increases plasma concentrations of ACTH and corticosterone and decreases secretion of growth hormone and gonadotropins. Increased secretion of CRF from the hypothalamic-hypophysial portal system mediates a complex cascade. Within the CNS, CRF activates the sympathetic nervous system, raising plasma concentrations of epinephrine and norepinephrine, thereby increasing heart rate, blood pressure, and plasma glucose concentrations. Stress-induced alterations in immune function may also be mediated by increased CRF secretion. The CRF concentration in the locus ceruleus increases markedly after acute stress.

When administered directly into the CNS of laboratory animals, CRF produces a number of physiological and behavioral changes similar to the physiological changes observed in stressed animals, which resemble signs and symptoms of depression and anxiety disorders. Centrally administered CRF increases mean arterial pressure, heart rate, oxygen consumption, and plasma glucose and catecholamine concentrations. It also alters locomotor activity, decreases sexual activity, diminishes food consumption, increases emotionality, and induces sleep disturbances.

Depression Much recent psychoendocrinology research has focused on neuroendocrine alterations in psychiatric disorders. Some of the most reproducible findings

were observed in depressed patients. Disorders of both the hypothalamic-pituitary-thyroid axis and the hypothalamic-pituitary-adrenal axis have been documented.

One of the most reproducible findings in biological psychiatry is the hyperactivity of the hypothalamic-pituitary-adrenal axis in some patients with major depressive disorder. Early studies measured glucocorticosteroid concentrations in plasma, urine, or cerebrospinal fluid (CSF) of depressed patients and matched controls. Those early findings led to development of the dexamethasone (Decadron)-suppression test (DST), which received considerable attention. The sensitivity and specificity of the test have prevented it from performing as a reliable diagnostic probe.

Considerable research has explored the locus of the hypothalamic-pituitary-adrenal axis abnormality in depression. Some research has focused on the role of ACTH and the possibility of adrenal supersensitivity to ACTH. Results are supported by postmortem and computerized tomographic studies that have provided evidence for adrenal gland enlargement in some depressed patients. Further, studies using the steroid hormonal receptor antagonist RU486, which blocks cortisol receptors and stimulates pituitary secretion of ACTH and beta-endorphin, have provided evidence for ACTH hypersecretion in a percentage of depressed patients.

Much recent research pertaining to hypercortisolism has focused on the role of CRF in hypothalamic-pituitary-adrenal–axis hyperactivity. For example, markedly higher CRF concentrations were found in CSF samples from suicide victims than in similar samples from sudden-death controls. In a study of CSF-CRF concentrations in drug-free depressed patients before and after a series of electroconvulsive shock treatments, the values at baseline were significantly elevated, but the values normalized after the patients had undergone electroconvulsive therapy and showed improvement in their condition. Increased CSF-CRF concentrations have now been found in other settings, including patients with anorexia nervosa and comorbid depression, as well as depressed patients with Huntington's disease. These findings illuminate earlier animal research examining the role of intracerebral administration of CRF in the production of a peripheral behavioral response resembling neurovegetative symptoms of depression. A confounding factor in these studies is that the stress of lumbar puncture itself may elevate CSF-CRH concentrations.

FUTURE DIRECTIONS

When psychological factors influence the onset, course, exacerbation, or outcome of illness, they do so in a complex interaction that depends on preexisting biogenetic factors predisposing the person to disease. Current psychosomatic theory emphasizes psychological factors as interacting dynamically with other biological and social factors in influencing disease processes. In fact, Leon Eisenberg has reviewed the spectrum of psychiatric disorders to emphasize how psychopathology arises at the interface between the brain and social experience. He has clarified the concept that contemporary psychiatric research has demonstrated that the brain responds to biological and social vectors and is jointly constructed by both. Such reasoning further delineates current views of the interrelatedness of psychological and somatic systems.

Emphasis in psychosomatic research will continue to focus on basic biological mediating mechanisms—derived from neuroimmunological, neurophysiological, and psychoneuroimmunological investigations—to identify more clearly the processes by which behavioral factors and psychological responses affect physiological activity and disease vulnerability. Simultaneously, this research must focus on the importance of psychosocial stressors, psychological influences of well-being, and modifiers of outcome in therapeutic interventions. Important research is continuing on the psychobiological effects of developmental trauma, such as that by Bessal van der Kolk, who has related such trauma to the development of posttraumatic stress disorder and borderline personality disorder.

SUGGESTED CROSS-REFERENCES

A general history of psychiatry appears in Section 55.1. Psychoneuroendocrinology is discussed in depth in Section 1.11, and endocrine and metabolic disorders in Section 25.6. Neurotransmitters are discussed in Section 1.4 and Section 1.5, cardiovascular disorders and psychosomatic medicine are reviewed in Section 25.4, stress in Section 25.9, behavior and immunity in Section 25.10, gastrointestinal conditions in 25.2, psycho-oncology in Section 25.11, and consultation-liaison psychiatry in Section 25.12.

SECTION REFERENCES

Alexander FG, Selesnick ST: *The History of Psychiatry: An Evaluation of Psychiatric Thought and Practice from Prehistoric Times to the Present*. Harper & Row, New York, 1966.

Brown TM: Cartesian dualism and psychosomatics. Psychosomatics *30:*322, 1989.

Cannon WB: *Bodily Changes in Pain, Hunger, Fear, and Rage*. Appleton, New York, 1915.

Cunningham AJ, Edmonds CVI, Jenkins GP, Pollack H, Lockwood GA, Warr D: A randomized controlled trial of the effects of group psychological therapy on survival in women with metastatic breast cancer. Psychooncology *7:*508, 1998.

Dunbar HF: *Emotions and Bodily Changes: A Survey of Literature on Psychosomatic Interrelationships 1910–1933*. Columbia University Press, New York, 1935.

Dwight MM, Stoudemire A: Effects of depressive disorders on coronary artery disease: A review. Harvard Rev Psychiatry *5:*115, 1997.

Eisenberg L: The social construction of the human brain. Am J Psychiatry *152:*1563, 1995.

Elliott G: Stress and illness. In *Psychosomatic Medicine: Theory, Physiology, and Practice*, vol 1, S Cheren, editor. International Universities Press, Madison, CT, 1989.

Engel GL: Sudden and rapid death during psychological stress. Ann Intern Med *74:*771, 1971.

*Evans DL, Staab JP, Petitto JM, Morrison MF, Szuba MP, Ward HE, Wingate B, et al: Depression in the medical setting: Biopsychosocial interactions and treatment considerations. J Clin Psychiatry *60*(Suppl):40, 1999.

Fawzy FI, Fawzy NW, Hyun CS, Elashoff R, Guthrie D, Fahey JL, Morton DL: Malignant melanoma: Effects of an early structured psychiatric intervention, coping, and affective state on recurrence and survival six years later. Arch Gen Psychiatry *50:*681, 1993.

Frasure-Smith N, Lesperance F, Talajic M: Depression and 18-month prognosis after myocardial infarction. Circulation *91:*999, 1995.

Goodkin K, Appels A: Behavioral-neuroendocrine-immunologic interactions in myocardial infarction. Med Hypotheses *48:*209, 1997.

Jiang W, Babyak M, Krantz DS, Waugh RA, Coleman RE, Hanson MM, Frid DJ, McNulty S, Morris JJ, O'Conner CM, Blumenthal JA: Mental stress-induced myocardial ischemia and cardiac events. JAMA *275:*1651, 1996.

Leonard BE, Miller K: *Stress, the Immune System and Psychiatry*. Wiley, Chichester, England, 1995.

*Levenstein S, Ackerman S, Kiecolt-Glaser JK, Dubois A: Stress and peptic ulcer disease. JAMA *281:*10, 1999.

*Lipowski ZJ: Psychosomatic medicine: Past and present. Part I: Historical background. Can J Psychiatry *31:*2, 1986.

*Lipowski ZJ: Psychosomatic medicine: Past and present. Part II: Current state. Can J Psychiatry *31:*8, 1986.

*Lipowski ZJ: Psychosomatic medicine: Past and present. Part III: Current research. Can J Psychiatry *31:*14, 1986.

Meyer A: *Psychobiology: A Science of Man*. Thomas, Springfield, IL, 1958.

*McDaniel JS, Moran M, Levenson J, Stoudemire A: Psychological factors affecting medical conditions. In *American Psychiatric Press Textbook of Psychiatry*, ed 3, RE Hales, SC Yudofsky, JA Talbott, editors. American Psychiatric Press, Washington, DC, 1999.

Miller TW: *Theory and Assessment of Stressful Life Events*. International Universities Press, Madison, WI, 1996.

Musselman DL, Tomer A, Manatunga AK, Knight BT, Porter MR, Kassey S, Marzec U, Harker LA, Nemeroff CB: Exaggerated platelet reactivity in major depression. Am J Psychiatry *153:*1313, 1996.

Neunes EV, Frank KA, Kornfield DS: Psychologic treatment for type A behavior pattern and for CAD: A meta-analysis of the literature. Psychosom Med *48:*159, 1987.

Ornish D, Scherwitz LW, Billings J, Gould L, Merritt T, Sparler S, Armstrong W, Ports T, Kireede R, Hogeboom C, Brand R: Intensive lifestyle changes for reversal of coronary heart disease. JAMA *280:*2001, 1998.

Prakash C, Lustman PJ, Freedland KE, Clouse RE: Tricyclic antidepressants for functional nausea and vomiting: Clinical outcome in 37 patients. Dig Dis Sci *43:*1951, 1998.

*Reiser MF: Changing theoretical concepts in psychosomatic medicine. In *American Handbook of Psychiatry*, ed 2, vol 4, *Organic Disorders and Psychosomatic Medicine*, MF Reiser, editor. Basic Books, New York, 1975.

Reite M, Capitanio JP: On the nature of social separation and social attachment. In *The Psychobiology of Attachment and Separation*, M Reite, T Field, editors. Academic Press, Orlando, 1985.

Rundell JR, Wise MG: *Textbook of Consultation-Liaison Psychiatry*. American Psychiatric Press, Washington, DC, 1996.

Schatzberg AF, Nemeroff CB: *Textbook of Psychopharmacology*, ed 2. American Psychiatric Press, Washington, DC, 1998.

Selye H: *The Stress of Life*. McGraw-Hill, New York, 1978.

Spiegel D, Bloom JA, Kraemer HC, Goettheil E: Effect of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet *2:*881, 1989.

*Stoudemire A, Fogel BS, Greenberg D, editors: *Psychiatric Care of the Medical Patient*. Oxford University Press, New York, 1999.

Stoudemire A: *Psychological Factors Affecting Medical Conditions*. American Psychiatric Press, Washington, DC, 1995.

van der Kolk B: The psychology and psychobiology of developmental trauma. In *Human Behavior: An Introduction for Medical Students*, ed 3, A Stoudemire, editor. Lippincott-Raven, Philadelphia, 1998.

*Weiner H: *Psychobiology and Human Disease*. Elsevier, New York, 1977.

Wolf S, Wolff HG: *Human Gastric Function*, ed 2. Oxford University Press, London, 1947.

25.2 GASTROINTESTINAL DISORDERS

WILLIAM R. YATES, M.D.

The psychiatric and psychological connections to gastrointestinal diseases have a rich history that continues to evolve as knowledge accumulates about both medical and psychiatric contributions to disease. Peptic ulcer disease and ulcerative colitis emerged as two of the earliest proposed examples of psychosomatic illnesses. Psychological factors such as personality characteristics were felt to be of primary importance in the development of both peptic ulcer and ulcerative colitis. Some early psychosomatic theories have not been confirmed, as medical knowledge of the causes of gastrointestinal diseases advanced. Recent understanding of the role of *Helicobacter pylori* in peptic ulcer disease now supports a primary medical mechanism for this gastrointestinal disease. However, this does not eliminate the need to examine biopsychosocial factors in the disease management of peptic ulcer. Gastrointestinal disorders underscore the importance of a biopsychosocial model of illness management.

The term *psychosomatic* indicates simply a mind-body relationship. At times, *psychosomatic* has inaccurately been used to mean that a physical disorder was caused by a mental disorder or process. More accurately, psychosomatic syndromes or disorders are physical disorders in which emotional processes play a role in some individuals.

Understanding psychiatric contributions and their importance to gastrointestinal disorders is growing. This is particularly true for the importance of psychological factors in the functional gastrointestinal disorders. Psychiatric illnesses frequently produce somatic symptoms including gastrointestinal symptoms. Patients frequently consult medical physicians for the gastrointestinal symptoms produced by psychiatric disorders. Medications used for gastrointestinal disorders can induce psychiatric symptoms, and medications used for psychiatric disorder can induce gastrointestinal symptoms. Understanding the interaction between gastrointestinal disorders and psychiatric disorders is important for primary care physicians, gastroenterologists, general psychiatrists, and consultation-liaison psychiatrists.

HISTORY

In the mid–nineteenth century, William Beaumont conducted longitudinal observational studies of the stomach of a man who had accidently shot himself in the abdomen. This accident resulted in the opportunity to directly observe the gastric mucosa while modifying external factors such as sight, taste, smell, and emotional arousal. Beaumont noted that emotional factors directly influence the appearance and function of the stomach.

Ivan Pavlov used a direct observational model in studying the principles of behavior. Pavlov studied gastrointestinal function, including digestion, in the dog by developing a surgical technique that allowed observation and sampling of the stomach. He found that the sight of food elicited the salivation response. Other cues could also be linked to the salivation response by pairing them with the sight of food. For example, if a tone accompanied the sight of food or feeding, the tone itself would eventually elicit a salivation response when presented without food. This finding led to the development of the concept of a conditioned stimulus and conditioned response. Pavlov developed a model for behavioral therapy using the results of his studies on the link between the gastrointestinal tract and the brain.

In the twentieth century, George Engel and others continued the study of the gastrointestinal tract and emotions. Engel was able to study the role of developmental factors on gastrointestinal function in a girl with a gastric fistula whom he observed from infancy to adulthood. Engel noted that developmental factors, interpersonal events, and emotional state all affect gastrointestinal function. Changes in gastrointestinal secretion, motility, and color were linked to specific emotions of anxiety, depression, and anger. Disruption in interpersonal function adversely affected gastrointestinal function.

Today, the study of the link between gastrointestinal function and emotional state involves analysis of the interaction between the cortex, hypothalamic axis, neurotransmitter function, and the gastrointestinal tract. Many gastrointestinal hormones and transmitters are also found in the brain and appear to mediate some of the complex interactions between the brain and gut. Emotional factors continue to appear to influence the function of the gastrointestinal tract and hence the gastrointestinal symptoms humans experience.

DEFINITION AND COMPARATIVE NOSOLOGY

It is important to understand the distinction between the terms *gastrointestinal disease* and *functional gastrointestinal disorders*. Gastrointestinal disease indicates a medical condition that can be documented by changes in pathophysiology, for example, peptic ulcer disease. Peptic ulcer disease can be diagnosed by direct examination of the anatomical features by endoscopy or radiography. Additionally, blood and breath tests confirm the presence of *H. pylori*, a bacterial agent that frequently causes peptic ulcer. The term *functional gastrointestinal disorders* indicates clinically significant distress producing a symptom related to the gastrointestinal tract (e.g., heartburn, dyspepsia, diarrhea) that does not demonstrate a pathophysiological mechanism. The term *functional* implies a disturbance of function without a disturbance of structure.

Functional Gastrointestinal Disorders Functional gastrointestinal disorders are common syndromes associated with significant subjective distress, abnormalities of bowel function, without evidence of structural abnormalities. Functional gastrointestinal disorders frequently have high rates of psychiatric comorbidity. Psychological factors can contribute significantly to the level of subjective gastrointestinal distress, use of medical resources, and the course and outcome of the functional gastrointestinal distress. Successful treatment of comorbid psychiatric conditions appears to reduce the severity of many functional gastrointestinal symptoms. Psychotropic drugs are commonly used by primary care physicians and subspecialists who treat functional bowel syndromes.

Table 25.2-1 outlines the spectrum of functional gastrointestinal disorders. Functional gastrointestinal disorders can include symptoms identified throughout the gastrointestinal tract. Individual functional gastrointestinal disorders with the highest prevalence rates and clinical importance are reviewed below.

Table 25.2-1 Functional Gastrointestinal Disorders

Globus and Fear of Choking *Globus*, the Latin word for lump, indicates a sensation of having a lump in the throat. Alternate terms for globus include *globus pharyngeus, globus hystericus*, and *globus syndrome*. Globus must be distinguished from *dysphagia*, or difficulty swallowing. Patients complaining of globus may also report dysphagia and fear of choking, but most patients with globus endorse neither of these other symptoms. Historically, globus constituted one of the symptoms of hysteria and was described by Hippocrates as a symptom related to the wandering uterus putting pressure on the neck. Approximately half of patients complaining of globus have no medical diagnosis after extensive testing. Acid reflux can produce globus and is the most common medical cause of the sensation. Other pharyngeal disorders including pharyngeal cancer can also produce globus.

Functional Heartburn Functional heartburn is the most common functional gastrointestinal disorder. Functional heartburn needs to be distinguished from gastroesophageal reflux disease. In functional heartburn, symptoms of acid reflux are present (i.e., heartburn, regurgitation of food), but there is no evidence of anatomical abnormality or esophagitis on endoscopy or radiography.

Functional Abdominal Bloating and Aerophagia Emotional distress and anxiety disorders can result in abnormal respiration and air swallowing or aerophagia. Swallowed air produces distension of the stomach and feelings of abdominal fullness or bloating along with belching.

Functional Chest Pain Chest pain frequently prompts attention for potential cardiac causes. However, functional esophageal motility disorders can produce chest pain and need to be considered in patients with chest pain and no evidence of cardiac abnormalities. Pain related to esophageal motility disturbances can be described as angina-like in location and character. High-amplitude esophageal contractions in the distal esophagus, “nutcracker esophagus,” can produce significant pain.

Referral to gastroenterology for evaluation of gastrointestinal causes of chest pain frequently occurs after cardiac causes have been ruled out. It is not unusual for several specialists to see the chest pain patient prior to appropriate psychiatric referral. However, at times psychiatrists may be involved before a comprehensive medical evaluation has been completed. Psychiatrists need to be aware of the possible gastrointestinal causes for chest pain, particularly when there is limited evidence for a psychiatric disorder in an unexplained chest pain presentation.

Irritable Bowel Syndrome Irritable bowel syndrome is the prototypical functional gastrointestinal disorder characterized by abdominal pain and diarrhea or constipation. The International Congress of Gastroenterology has developed a standardized set of criteria for irritable bowel syndrome.

1. Abdominal pain relieved by defecation or associated with change in frequency or consistency of stool.
2. Disturbed defecation involving two or more of the following:
   altered stool frequency
   altered stool form (hard or loose and watery)
   altered stool passage (straining or urgency, feeling of incomplete evacuation)
   passage of mucus

Irritable bowel syndrome can often be categorized into diarrhea-predominant, constipation-predominant, and mixed subtypes. Medical treatment often targets the predominant symptom. Some studies suggest that irritable bowel syndrome accounts for up to 50 percent of all outpatient evaluations done by gastroenterologists. Comorbid psychiatric disorders appear to increase the likelihood of health-care-seeking behavior for people with symptoms of irritable bowel syndrome.

Some patients with irritable bowel syndrome may demonstrate physiological abnormalities including abnormal intestinal myoelectric activity, gastrointestinal hormonal abnormalities, or allergic responses to some foods. Most clinicians agree that both physiological and psychological factors contribute to the clinical picture of irritable bowel syndrome in most patients.

Dyschezia and Functional Anorectal Pain Dyschezia (difficulty with evacuation of rectal contents) and functional anorectal pain appear commonly in the general population. Two types of functional anorectal pain have been described: one is associated with constant rectal pain (levator syndrome) and the other is characterized by intermittent sharp pain that often disappears completely (proctalgia fugax).

Classification of Functional Gastrointestinal Disturbances The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) classification system shares and differs in some ways from the diagnoses and categories of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). ICD-10 recognizes a category of somatoform disorders with somatization disorders and hypochondriacal disorders similar to their DSM-IV counterparts. However, ICD-10 also includes a diagnosis called somatoform autonomic dysfunction. Two subtypes are noted in this category relevant to gastrointestinal disorders: upper gastrointestinal tract and lower gastrointestinal tract. Such diagnoses may be appropriate in patients with functional gastrointestinal disorders where the ICD-10 system is used.

Functional Gastrointestinal Disorders and Somatothymia Somatothymia has been proposed by Alan Stoudemire as a way to understand the mechanism of unexplained somatic symptoms in some patients. Somatothymia is the tendency to communicate psychological distress (i.e., affective distress) to others through somatic symptoms. Patients who tend to somatize may be using their own culturally determined communication strategies to acknowledge psychological distress. Some functional gastrointestinal disorders may be elicited through this mechanism. Patients from lower socioeconomic classes and third-world countries may be more likely to have somatothymia.

Somatization Somatization indicates the presentation of somatic symptoms in the absence of disease or tissue damage. Often somatization increases during periods of stress or during conflict. Functional somatic symptoms including functional gastrointestinal disorders are more likely to be endorsed by patients with somatization. Somatization occurs in a variety of psychiatric disorders including major depressive disorder, anxiety disorders, and somatoform disorders. The term *functional* does not imply that the symptoms serve some function for the patient, only that a medical disease cannot be confirmed as the direct cause of the patient's symptoms.

Psychological Factors Affecting Medical Conditions Psychiatric disorders can influence gastrointestinal diseases and functional gastrointestinal disorders in a variety of ways. DSM-IV defines psychological factors affecting physical conditions to address the spectrum of ways that psychiatric disorders and psychological mechanisms influence medical conditions. Psychological factors affecting medical conditions include the following subtypes: (1) mental disorder affecting a medical condition, (2) psychological symptoms affecting a medical condition, (3) personality traits or coping style affecting a medical condition, (4) maladaptive health behaviors affecting a medical condition, and (5) stress-related physiologic response affecting a medical condition. The corresponding classification category in ICD-10 is F54 psychological and behavioral factors associated with disorders or diseases classified elsewhere. All of the psychological factors affecting medical condition categories can be found with gastrointestinal diseases and disorders.

EPIDEMIOLOGY

Gastrointestinal diseases occur frequently in the general population. Primary care physicians commonly encounter and manage many of the common gastrointestinal disorders. Gastroenterologists provide specialized care for more serious or complicated gastrointestinal disorders.

Table 25.2-2 lists the common gastrointestinal diseases by anatomical region. Diseases of the esophagus, stomach, intestines, anorectum, pancreas, liver, and gallbladder constitute the spectrum of gastroenterology. Consultation psychiatrists are likely to be involved in assessment and psychiatric treatment for patients who have one of these gastrointestinal diseases.

Disorders of the esophagus
Reflux esophagitis (GERD)
Infectious esophagitis
Esophageal motility disorders
Disorders of the stomach and intestines
Peptic ulcer disease
Gastroparesis
Malabsorption and maldigestion
Inflammatory bowel disease
Crohn's disease
Ulcerative colitis
Diverticular disease
Diseases of the anorectum
Hemorrhoids
Anal fissures
Perirectal abscess
Diseases of the pancreas
Acute pancreatitis
Chronic pancreatitis
Diseases of the liver and gallbladder
Infectious hepatitis
Toxic and drug-induced hepatitis
Primary biliary cirrhosis
Primary sclerosing cholangitis
Metabolic liver disease
Gallstones and cholecystitis
Cancer of the gastrointestinal tract
Colon and rectal cancer
Pancreatic cancer
Esophageal cancer
Stomach cancer
Liver cancer

Table 25.2-2 Common Gastrointestinal Diseases

Cancer of the gastrointestinal tract represents a major cause of cancer mortality in the United States and throughout the world. Colon and rectal cancer is estimated to be the third leading cause of death in men (following lung and prostate cancer) and the second leading cause of death in women (following breast cancer). Pancreas, stomach, and esophageal cancer rank in the top ten leading sites of cancer in men. Pancreas and stomach cancer rank in the top ten leading sites of cancer in women. The reader is directed to the chapter on oncology for further discussion of psychiatric issues related to cancer.

Prevalence Rates and Correlates Drossman and his colleagues examined the prevalence and correlates of functional gastrointestinal disorders in a community survey of U.S. households. Sixty-nine percent of this sample reported at least one of 20 functional gastrointestinal symptoms in the 3 months prior to the survey. The prevalence rates for various functional gastrointestinal syndromes are noted in Table 25.2-3. Most individuals who experience functional gastrointestinal symptoms do not seek medical attention. Severity of symptoms and psychiatric comorbidity increase the likelihood of seeking medical attention.

Disorder	Prevalence (%)
Functional heartburn	32.6
Functional abdominal bloating	30.8
Aerophagia	23.4
Dyschezia	13.8
Functional chest pain	13.6
Globus	12.5
Functional anorectal pain	11.6
Irritable bowel syndrome	11.6
Rumination syndrome	10.9
Functional incontinence	7.8
Functional dysphagia	7.5
Functional constipation	3.6
Functional dyspepsia	2.9
Chronic abdominal pain	2.2
Functional diarrhea	1.8
Sphincter of Oddi dyskinesia	1.5

Adapted from Drossman DA, Li Z, Andruzzi E, Temple RD, Talley NJ, Thompson WG, Whitehead WE, Janssens J, Funch-Jensen P, Corazziari E: U.S. householder survey of functional gastrointestinal disorders. Dig Dis Sci 38:1569, 1993.

Table 25.2-3 Prevalence of Functional Gastrointestinal Disorders in the Community

Functional heartburn ranked as the most frequent functional gastrointestinal disorder, experienced by nearly one-third of the sample in this community survey. Functional abdominal bloating and aerophagia were reported by 31 percent and 23 percent of the sample, respectively. A variety of other functional gastrointestinal disorders of the esophagus, stomach, intestine, and anorectal areas were endorsed by 10 percent or more of the population.

Risk Factors Women are more likely to report globus, functional dysphagia, irritable bowel syndrome, functional constipation, functional abdominal pain, functional biliary pain, and dyschezia. Men are more likely to report aerophagia and functional bloating. Functional gastrointestinal syndromes increase the rate of absenteeism at work or school. Occupational impairment appears to be specifically associated with chronic abdominal pain, functional biliary pain, functional dyspepsia, and irritable bowel syndrome.

Psychiatric disorders and the behaviors associated with them can increase the risk of gastrointestinal disease. A good example of this is alcohol abuse or dependence. Chronic alcohol consumption is associated with a variety of gastrointestinal disorders including esophageal reflux, gastritis, pancreatitis, and several of the gastrointestinal cancers. Intravenous drug abuse or dependence increases the risk for hepatitis B and hepatitis C. Sexual promiscuity associated with antisocial personality disorder may also increase the risk for viral hepatitis.

Nicotine and caffeine produce direct physiological changes in gastrointestinal function. Both compounds can increase gastric acid secretion and exacerbate gastritis, esophagitis, or peptic ulcer disease. Patients with psychiatric disorders have higher rates of nicotine dependence and caffeine dependence than the general population. Modification of smoking behaviors may be more difficult in patients with depressive and other concurrent psychiatric disorders than in those with nicotine dependence alone.

PSYCHIATRIC COMORBIDITY IN FUNCTIONAL GASTROINTESTINAL DISORDER

A large part of the literature reports the link between stress, anxiety, and physiologic responsivity of the gastrointestinal system. Anxiety can produce disturbances in gastrointestinal function through a central control mechanism or via humoral effects such as the release of catecholamines. Electrical stimulation studies suggest that sympathetic autonomic responses can be generated in the lateral hypothalamus, a region with neural interactions within the limbic forebrain. Parasympathetic autonomic responses also influence gastrointestinal function. Parasympathetic impulses originate in the periventricular and lateral hypothalamus and travel to the dorsal motor nucleus of the vagus, the main parasympathetic output pathway. The vagus is also modulated by the limbic brain, producing an emotional-gut pathway of response.

Acute stress can induce physiological responses in several gastrointestinal target organs. In the esophagus, acute stress increases resting tone of the upper esophageal sphincter and increases contraction amplitude in the distal esophagus. Such physiological responses may result in symptoms consistent with globus or esophageal spasm syndrome. In the stomach acute stress induces decreased antral motor activity, potentially producing functional nausea and vomiting. Reduced migrating motor function can occur in the small intestine while in the large intestine there can be increased myoelectrical and motility activity under acute stress. These effects in the small and large intestine may be responsible for bowel symptoms associated with irritable bowel syndrome.

Patients with contraction abnormalities and functional esophageal syndromes demonstrate high rates of psychiatric comorbidity. Functional esophageal symptoms include globus, dysphagia, chest pain, and regurgitation. Such symptoms can occur in conjunction with esophageal smooth muscle contraction abnormalities. Not all patients with functional esophageal symptoms display contraction abnormalities. Anxiety disorders ranked highest in a study of psychiatric comorbidity in functional esophageal spasm, being present in 67 percent of subjects referred to a gastrointestinal motility laboratory for testing. Generalized anxiety disorder topped the list of anxiety disorder diagnoses in this series. Many patients in this study had anxiety disorder symptoms prior to the onset of esophageal symptoms. This suggests that anxiety disorder may induce physiological changes in the esophagus that can produce functional esophageal symptoms.

One of the earliest studies of psychiatric comorbidity that used structured psychiatric interviews focused on patients with esophageal motility disorders. A nonspecific esophageal motility group (the functional group) was compared with a group of patients with organic esophageal disorders (i.e., achalasia) and a group of healthy controls. The functional esophageal motility group ranked highest in rate of psychiatric comorbidity, with 84 percent of the group receiving at least one psychiatric diagnosis. The organic esophageal motility group and healthy control groups had psychiatric comorbidity rates of 33 and 31 percent, respectively. The psychiatric disorders noted in the functional motility group included depressive disorders, anxiety disorders, somatization disorder, and phobias.

Gastroesophageal Reflux Disease Gastrointestinal reflux disease is the most common disorder of the esophagus and accounts for most over-the-counter antacid

consumption. The predominant symptom is heartburn, which may be accompanied by regurgitation and pain with swallowing. Multiple factors appear to be important in the generation of reflux and heartburn: (1) presence of a hiatal hernia, (2) effectiveness of the lower esophageal sphincter in blocking the reflux of stomach acid, (3) effectiveness of the esophagus in clearing and neutralizing reflux, (4) ability of the esophagus to protect itself against acid and pepsin, and (5) delayed gastric emptying and acid hypersecretion. Up to 80 percent of patients with gastrointestinal reflux disease have a hiatal hernia. However, 50 percent of patients with a hiatal hernia do not have gastrointestinal reflux disease.

Many patients with gastrointestinal reflux disease along with many clinical gastroenterologists believe that psychological distress increases symptom severity in those prone to this disease. In a survey of gastrointestinal reflux disease sufferers, excessive stress, too much excitement, family arguments, and temporary depression were felt to trigger heartburn.

Irritable Bowel Syndrome Psychiatric disorders complicate the diagnosis and management of many patients with irritable bowel syndrome. Because many studies of psychiatric comorbidity have focused on clinical and hospitalized patients, a potential sample bias effect could explain this association. However, recent studies of general population samples found evidence of increased psychiatric comorbidity in community subjects with unexplained gastrointestinal symptoms. In the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area (ECA) Study, over 18,000 adults in the United States received a direct interview for psychiatric illness. Six gastrointestinal symptoms from the somatization disorder section were reviewed with the participants (abdominal pain, diarrhea, gas and bloating, constipation, and nausea and vomiting). Subjects who experienced multiple gastrointestinal symptoms were compared with those without gastrointestinal symptoms.

One gastrointestinal symptom increased the rate of lifetime major depressive disorder (7.5 versus 2.9 percent), panic disorder (2.5 versus 0.7 percent), and agoraphobia (10.0 versus 3.6 percent). Community subjects with two gastrointestinal symptoms had even higher rates of major depressive disorder (13.4 percent), panic disorder (5.2 percent), and agoraphobia (17.8 percent). These rates of psychiatric disorders are lower than those in referred tertiary-care populations with functional bowel syndromes. However, the increased rates compared with controls in this community population suggests a true link between psychiatric illness and functional gastrointestinal disorders and cannot be simply explained by sample selection bias.

Other studies of psychiatric comorbidity in irritable bowel syndrome estimate comorbidity rates of 42 to 64 percent of all irritable bowel syndrome patients. Control comorbidity rates in these populations generally range from 15 to 20 percent. The exact mechanism for high rates of psychiatric comorbidity in irritable bowel syndrome is unknown.

Four models have been proposed to explain the relation between irritable bowel syndrome and high rates of psychiatric comorbidity. The first model is the somatization disorder hypothesis, which proposes that some people display anxious and depressed mood with multiple nonspecific somatic symptoms. This model classifies irritable bowel syndrome as one of a group of diagnoses that can be made from a primary somatization disorder or other somatoform disorder. Other similar functional medical disorders that could occur in these patients include fibromyalgia. Although somatization disorder and other somatoform disorders occur with irritable bowel syndrome, they do not occur in all patients.

A second model to explain the role of psychiatric comorbidity in irritable bowel syndrome is the somatopsychic hypothesis. This hypothesis states that psychological symptoms are the result of chronic gastrointestinal distress and unsatisfactory interaction with health care providers who do not accurately diagnose and treat irritable bowel syndrome. One argument against the somatopsychic hypothesis is the relatively lower rates of psychiatric disorder in more severe and chronic gastrointestinal disorders such as inflammatory bowel disease and lactose intolerance. These disorders would be expected to have higher rates of psychiatric comorbidity than irritable bowel syndrome if the somatopsychic hypothesis explained the psychological manifestation of symptoms in irritable bowel syndrome.

A third explanatory model, the psychogenic hypothesis, states that specific psychiatric disorders cause irritable bowel syndrome in a significant proportion of patients. Panic disorder in particular is proposed to cause secondary irritable bowel syndrome. Although a significant percentage of irritable bowel syndrome patients have panic disorder, most do not, so this third model is unlikely to be a sole explanation for the psychological manifestation of irritable bowel syndrome.

The final model is the self-selection model. This model proposes that psychiatric comorbidity increases the rate of treatment seeking by patients who have irritable bowel syndrome. Under this model, irritable bowel syndrome without psychiatric comorbidity would be accompanied by the lowest rate of health care use, while irritable bowel syndrome with psychiatric comorbidity would be associated with the highest rate. There is some support for this model in the medical clinic setting; psychiatric comorbidity does seem to be more prevalent in clinic samples than in community samples. However, population studies (including the NIMH ECA study) suggest that this model cannot fully explain the relation. A significant number of community subjects with irritable bowel syndrome who do not seek medical attention suffer from significant psychiatric comorbidity. Possibly several of the proposed models contribute to the overlap between irritable bowel syndrome and psychiatric illness.

One method of sorting out the relation between irritable bowel syndrome and psychiatric illness is the family study method. If psychiatric disorders simply follow the distress of irritable bowel syndrome, one would not expect a higher rate of psychiatric illness in the relatives of patients with the syndrome. Twenty patients with irritable bowel syndrome and 20 patients undergoing laparoscopic cholecystectomy were compared. Family rates of depressive disorders and anxiety disorder were higher in family members of the irritable bowel syndrome probands than in control probands. This study supports the psychogenic hypothesis for irritable bowel syndrome, although the association could be due to a treatment selection bias since probands in this study were identified in a tertiary-care center.

Physiological abnormalities also appear to contribute to the symptom profile of irritable bowel syndrome. Autonomic nervous system abnormalities in irritable bowel syndrome may vary by the predominant symptom noted by individual patients; one study found separate physiological abnormalities in diarrhea- and constipation-predominant patients. Diarrhea-predominant patients exhibited abnormalities in adrenergic function as measured by blood flow in the hand in various postures. Constipation-predominant patients exhibited vagal cholinergic function abnormalities as measured by R-R interval changes during deep breathing. Constipation-predominant patients exhibited higher levels of psychological distress (i.e., anxiety and depression) than the diarrhea-predominant group.

Peptic Ulcer Disease Peptic ulcer refers to mucosal ulceration involving the distal stomach or proximal duodenum. Symptoms of peptic ulcer disease include a gnawing or burning epigastric pain that occurs 1 to 3 hours after meals and is relieved by food or antacids. Accompanying symptoms can include nausea, vomiting, dyspepsia, or signs of gastrointestinal bleeding such as hematemesis or melena. Lesions are generally small, one centimeter or less in diameter.

Early theories identified excess gastric acid secretion as the most important etiologic factor, but the importance of infection with *H. pylori* is becoming more acknowledged. *H. pylori* is associated with 95 to 99 percent of duodenal ulcers and 70 to 90 percent of gastric ulcers. Antibiotic therapy that targets *H. pylori* results in much higher healing and cure rates than antacid and histamine inhibitor therapy. Standard regimens for the treatment of *H. pylori* infection often include combinations of two or three antibiotic agents. Commonly used antibiotic agents with efficacy against *H. pylori* include amoxacillin (Amoxil), metronidazole (Flagyl), tetracycline (Achromycin), and clarithromycin (Biaxin).

Early studies of peptic ulcer disease suggested a role of psychological factors in the production of ulcer vulnerability. This effect was felt to be mediated through the increased gastric acid excretion associated with psychological stress. Studies of prisoners of war during World War II documented rates of peptic ulcer formation twice as high as controls. Recent evidence for a primary role of *H. pylori* in peptic ulcer initiation, suggests that psychosocial factors are primarily involved in the clinical expression of symptoms. Stressful life events may also reduce immune responses, resulting in higher vulnerability to *H. pylori* infection. There is no consensus on whether specific psychiatric disorders are related to peptic ulcer disease.

Ulcerative Colitis Ulcerative colitis is an inflammatory bowel disease affecting primarily the large intestine. The cause of ulcerative colitis is unknown, but it must be differentiated from forms of infectious colitis such as campylobacteriosis, amebiasis, and shigellosis and pseudomembranous enterocolitis and ischemic colitis. The yearly incidence of ulcerative colitis is about 10 per 100,000. Risk factors for ulcerative colitis include Jewish ethnicity. The predominant symptom of ulcerative colitis is bloody diarrhea. Extracolonic manifestations can include uveitis, iritis, skin diseases, and primary sclerosing cholangitis. Diagnosis is made mainly by colonoscopy or protoscopy. Surgical resection of portions of the large bowel or entire bowel can result in cure for some patients.

Over 100 studies of ulcerative colitis have examined associated psychiatric factors. A review of these studies found serious flaws in research design, including lack of control subjects, unspecific data collection methodology, and failure to use standardized diagnostic criteria. Studies without adequate controls were more likely to suggest an association of life events, abnormal personality, and psychiatric disorders. When the seven studies with the best methodological design were examined, no association between ulcerative colitis and psychiatric factors was found. This review emphasizes the need for caution in interpreting psychiatric factors associated with organic gastrointestinal disease. Undoubtedly, for individual patients, psychiatric factors may play a key role in the manifestations and complexity of disorders

such as ulcerative colitis. However, generalizations about psychological mechanisms for ulcerative colitis must be considered unproved.

Crohn's Disease Crohn's disease is an inflammatory bowel disease affecting primarily the small intestine and colon. Common symptoms in Crohn's disease include diarrhea, abdominal pain, and weight loss. The yearly incidence of Crohn's disease is about half that of ulcerative colitis, about 5 per 100,000 population. The course is chronic, often with periods of remission followed by periods of acute symptoms. Treatment consists of the use of antibiotic agents such as metronidazole, sulfasalazine (Azulfidine) or mesalamine (Asacol). Immunosuppressive drugs also are commonly used to control flare-ups. Prednisone is the most frequently used corticosteroid, and azathioprine (Imuran) a commonly used immunosuppressive agent.

Because Crohn's disease is a chronic illness, most studies of psychiatric comorbidity focus on psychiatric disorders occurring after the onset of the disorder. A study of psychiatric symptoms in Crohn's disease prior to the onset of physical symptoms found high rates (23 percent) of preexisting panic disorder compared with control subjects and subjects with ulcerative colitis. No statistically significant preexisting psychiatric comorbidity in ulcerative colitis occurred in this study. Longitudinal studies and careful retrospective studies in gastrointestinal chronic gastrointestinal disorders can be helpful in sorting out psychiatric disorder as a risk factor, consequence, or chance association with specific gastrointestinal disorder.

A young woman from a small Midwestern town developed Crohn's disease while in high school. Her symptoms included intermittent abdominal pain and diarrhea. Prior to the onset of her inflammatory bowel disease, she had been obese, with a maximum weight of 225 lbs. She had no history of psychiatric illness prior to the development of Crohn's disease but had been a somewhat shy individual. Her family became quite distressed by the development of her medical illness. They developed a protective style of interacting with her, and she continued to live at home following completion of high school.

Her Crohn's disease became chronic, with intermittent periods of flare-up and remission. Over time her gastroenterologist noted that her complaints of abdominal pain did not always correspond with periods of inflammatory disease activity. She developed a fear of eating and drinking and associated oral intake with abdominal pain. She became preoccupied with her weight and began vigorous dieting and restricting her oral fluid intake. By age 24 her weight began to drop, and over 6 months, she lost 100 pounds, down to a weight of 125. Her gastroenterologist suspected she might be developing an eating disorder, and she was referred for a psychiatric consultation.

The psychiatric consultant evaluated the patient and noted chronic low mood consistent with a diagnosis of dysthymia. She was preoccupied with concerns about her weight and had rigid rules about eating. Dieting complicated the medical management of her Crohn's disease. A diagnosis of eating disorder not otherwise specified along with dysthymic disorder was made, therapy with a selective serotonin reuptake inhibitor (SSRI) was started, and she was referred to an eating disorder partial hospitalization program. In the partial program her compliance was poor, and one day she collapsed in group therapy. Medical evaluation revealed orthostatic hypotension that resolved with hospitalization and intravenous fluid and electrolyte replacement. She was transferred to an inpatient eating disorder program. However, her family checked her out of the program against medical advice. The patient and family complained that a psychiatric disorder was not her problem and that the psychiatric treatment team did not understand her Crohn's disease and proper medical management.

This case illustrates the effect that chronic gastrointestinal disorders can have on individuals vulnerable to psychiatric illness. Chronic gastrointestinal illness can worsen abnormal family dynamics and stimulate development of psychiatric illnesses such as mood and eating disorders. When both a gastrointestinal illness and a psychiatric illness are present, the outcome and prognosis for each individual condition becomes complicated. Coordination of medical and psychiatric services is necessary to maximize the likelihood of a successful outcome.

Hepatitis and Liver Failure Hepatitis can be caused by viral agents, drugs, or poisons that cause hepatotoxicity. Viral agents associated with hepatitis include hepatitis A, hepatitis B, hepatitis C, hepatitis D, hepatitis E viruses, cytomegalovirus, and Epstein-Barr virus. Intravenous drug use increases risk for hepatitis B and C. Hepatitis C is most commonly due to transfusions that occurred prior to blood bank screening for this agent. Hepatitis C can become a chronic infection taking decades to progress to the symptomatic phase. Hepatitis C is becoming increasingly prevalent because chronic transfusion-associated infections are now moving into symptomatic and cirrhotic phases.

Alcohol is a hepatotoxic agent that causes alcoholic hepatitis and cirrhosis. Typically, long periods of heavy drinking are required for development of cirrhosis—estimates of 12 standard drinks a day for 20 years have been made for the alcohol burden necessary to cause cirrhosis. Concurrent hepatitis C infection may increase the risk of alcohol-related liver disease. Women appear to be more susceptible to alcoholic cirrhosis than men.

The symptoms of hepatitis include malaise, nausea, anorexia, fatigue, vomiting, diarrhea, and weight loss. Treatment for hepatitis differs by specific agent. Antiviral therapies continue to evolve and become more widely used. Interferon therapy for hepatitis C carries neuropsychiatric risks, including an increased risk of depression and suicide.

Psychiatric comorbidity is common with infectious and toxic hepatitis. Acetaminophen overdose can produce acute toxic hepatitis. This drug is commonly ingested in overdose by patients with psychiatric disorders such as mood disorders, personality disorders, alcohol and other substance abuse, and adjustment disorders. The management of alcoholic hepatitis requires attention to the alcohol abuse and dependence issues common in this patient population. Liver transplantation for alcohol-related end-stage liver disease commonly requires psychiatric consultation. Consultation-liaison services increase the likelihood of successful transplant outcomes.

PSYCHIATRIC CONSULTATION IN GASTROINTESTINAL DISEASE

Patients with gastrointestinal disorders frequently require psychiatric consultation. In a series of 1250 patients referred for psychiatric consultation over a 1-year period, 103 (8 percent) had gastrointestinal disorder as their primary medical diagnosis (Table 25.2-4). This series of psychiatric consultations occurred in a tertiary-care university hospital setting. Consultation requests originated from medical inpatients (55 percent) and medical outpatients (45 percent) referred from primary care internists and gastroenterologists. Most patients were female (56 percent). All patients were diagnosed by a faculty psychiatrist using DSM-IV criteria. A database of primary medical and psychiatric disorders coded for the ninth revision of *International Statistical Classification of Diseases* (ICD-9) and DSM-IV diagnoses for these consultation patients provides some insight into the epidemiology of psychiatric illness in the gastrointestinal disorders.

Medical Disorder	Frequency (%)	Psychiatric Diagnosis	Frequency (%)
Hepatitis	25	Major depressive disorder or dysthymic disorder	30
Abdominal pain	12	Alcohol or other substance dependence	17
Pancreatitis	12	Somatoform disorder	10
Crohn's disease or ulcerative colitis	10	Anxiety disorder	7
Peptic ulcer or gastritis	8	Personality disorder	7
Cirrhosis	7	Adjustment disorder	6
Gastroesophageal reflux	6	Bipolar I disorder	5
Diarrhea or constipation	6	Psychotic disorder	5
Esophageal disorder	5	Factitious disorder	4
Gastroparesis	5	No psychiatric diagnosis	4

* Rankings and frequencies in series of 103 patients referred from internal medicine for psychiatric consultation. Patients had a primary diagnosis of a gastrointestinal disorder in a university hospital setting.

Table 25.2-4 Frequency of Medical and Psychiatric Disorders in Psychiatric Consultation for Gastrointestinal Disease*

Specific gastrointestinal disorders encountered in this recent series are ranked in Table 25.2-4. Hepatitis ranked first in this series. Most hepatitis patients presented with hepatitis C and were seen with mood disorders or referred for consultation prior to interferon therapy. Undiagnosed gastrointestinal symptoms of abdominal pain, diarrhea, and constipation were common, with many representing functional gastrointestinal disorders. Pancreatitis presented primarily in the inpatient setting, with alcohol dependence as the primary psychiatric comorbid condition. Inflammatory bowel disease, peptic ulcer, gastritis, gastroesophageal reflux disease, esophageal disorders, and gastroparesis all represented at least 5 percent of the series total. The type of gastrointestinal disorders seen in consultation can vary by the site of

practice, and tertiary-care-center distributions may not reflect the patient profiles seen in other practice settings.

Major depressive disorder and dysthymic disorder ranked first in the psychiatric diagnosis category for this series. This is consistent with the most prevalent diagnoses seen overall in internal medicine referrals. Substance dependence and somatoform disorders both occurred in 10 percent or more of consultations. Anxiety disorders, personality disorders, adjustment disorders, bipolar I disorder, and psychotic disorders occurred in at least 5 percent of the diagnoses. Factitious disorder, a very rare psychiatric disorder, also made the top 10 ranking for this series of patients. Many patients in this series also met criteria for psychological factors affecting a medical condition, a diagnosis that would have been made in addition to the primary psychiatric disorder but not coded in the series database.

DIAGNOSIS AND CLINICAL FEATURES

Gastrointestinal Symptoms in Psychiatric Disorders Gastrointestinal symptoms occur commonly in a variety of psychiatric disorders. This association is so well known that many of the diagnostic criteria for some mental disorders include the presence of specific gastrointestinal symptoms and behaviors. Table 25.2-5 shows some relations between specific psychiatric disorders and gastrointestinal symptoms. Somatization disorder and the associated somatoform disorders commonly present with gastrointestinal complaints. In DSM-IV abdominal and rectal pain are two possible symptoms for meeting the criteria for medically unexplained pain in four separate sites. In addition, somatization disorder requires two additional gastrointestinal symptoms. Examples given to meet these criteria include nausea, bloating, vomiting other than during pregnancy, diarrhea, or intolerance to several varieties of food.

Symptom	Somatization Disorder	Major Depressive Disorder	Panic Disorder	Social Phobia	Eating Disorders	Substance Withdrawal	Substance Intoxication
Nausea	•		•			†	†
Vomiting	•				•	†	†
Abdominal pain	•		•				†
Diarrhea	•			†	†	†	†
Appetite disturbance		•			•	†	†
Food intolerance	•						

• = "A diagnostic criteria," † = disorder description.

Table 25.2-5 Gastrointestinal Symptoms Noted in DSM-IV

Changes in appetite are a hallmark feature of major depressive disorders. Nausea frequently accompanies a panic attack, as does abdominal distress. Diarrhea or fear of uncontrollable diarrhea in a social situation can be present in social phobia. Eating disorders usually include a disturbance of appetite, and self-induced vomiting is the most common cause of purging behavior in bulimia. A variety of substances produce gastrointestinal disturbances during the intoxication and withdrawal phases. Physicians need to keep these psychiatric disorders in mind when considering the differential diagnosis of patients presenting with gastrointestinal symptoms.

Psychotropic Drug Adverse Effects on Gastrointestinal Function Another potential link between psychiatric illness and gastrointestinal symptoms is the potential for psychotropic drugs to influence gastrointestinal function adversely. These drug-induced effects can produce several clinical challenges. First, patients may elect to discontinue necessary treatment because of the gastrointestinal adverse effects. Second, prescribers may need to consider the possibility of serious gastrointestinal disease or exacerbation of functional gastrointestinal disturbances when drug-induced symptoms develop. Clinicians may need to carefully consider the side-effect profile of specific psychotropic drugs when treating patients with gastrointestinal disorders.

The SSRIs can produce significant gastrointestinal symptoms. These gastrointestinal adverse effects tend to be noted at the initiation of therapy and to be dose related, with higher dosages producing higher rates of adverse effects. Nausea is a significant adverse effect in the profile of the SSRIs. In placebo-controlled trials of SSRIs for depression, rates of nausea attributed to SSRIs were 11 to 26 percent above the rates of nausea related to placebo. Fluvoxamine (Luvox) appears to have higher rates of nausea. Fluoxetine (Prozac) and Sertraline (Zoloft) studies indicate relatively lower rates of nausea. Additional gastrointestinal adverse effects noted in the placebo-controlled studies of SSRI include diarrhea (0 to 8 percent), anorexia (1 to 9 percent), and dyspepsia (1 to 3 percent). All these rates are adjusted for placebo rates and are statistically significant.

It is not suprising that drugs affecting serotonin induce changes in gastrointestinal function. Serotonin-rich areas of the brain are located in the brainstem and spinal cord. Efferent tracts from these areas innervate multiple areas throughout the brain that regulate many bodily functions including sleep, appetite, gastrointestinal motility, and reproductive function.

Standard tricyclic antidepressant drugs can also produce gastrointestinal effects, specifically dry mouth and constipation. These effects appear primarily related to the anticholinergic effect of tricyclic compounds. Tricyclic drugs with lower gastrointestinal adverse-effect profiles include desipramine (Norpramin) and nortriptyline (Pamelor), compounds with lower anticholinergic potencies. Higher rates of gastrointestinal adverse effects are noted in amitriptyline (Elavil) and imipramine (Tofranil). With imipramine nearly 50 percent experience dry mouth, and constipation is noted in nearly 20 percent. Tricyclic drugs can reduce gastric motility and need careful consideration for patients with gastrointestinal reflux or gastric hypomotility syndromes such as gastroparesis and for geriatric patients prone to problems related to constipation. Low-potency antipsychotic agents such as chlorpromazine (Thorazine) and thioridazine (Mellanil) share anticholinergic properties with the tricyclic drugs.

Use of nortriptyline rather than amitriptyline as the standard tricyclic drug has several advantages. Because of the lower anticholinergic effect, nortriptyline is better tolerated by most patients and is likely to induce fewer gastrointestinal adverse effects. Amitriptyline has a much higher discontinuation rate than nortriptyline for gastrointestinal effects. Patients who are prescribed amitriptyline as a first-line agent may discontinue the drug and psychiatric treatment because of adverse effects. If a tricyclic drug is used to treat depression, a less anticholinergic agent such as nortriptyline is likely to be tolerated well enough for an adequate therapeutic trial.

Psychiatric Adverse Effects Associated With Commonly Prescribed Drugs for Gastrointestinal Conditions Gastrointestinal prescription drugs represent a significant percentage of all prescriptions. Several of these agents can produce psychiatric symptoms. Additionally, several of the gastrointestinal drugs can interact with common psychiatric prescription drugs. Table 25.2-6 summarizes some of the important potential adverse psychiatric effects and drug interactions with common prescription drugs.

Generic	Brand	Indications	Adverse Psychiatric Effects and Drug Interactions
Histamine receptor antagonists		Peptic ulcer disease	
Nizatidine	Axid		Cimetidine can cause delirium especially in severely ill and hospitalized patients; cimetidine can increase blood levels of tricyclic drugs, carbamazepine, valproate, SSRI compounds; fewer CNS effects and metabolic disturbances are noted with other histamine receptor agonists
Famotidine	Pepcid		
Cimetidine	Tagamet		
Ranitidine	Zantac		
Proton pump inhibitor		Peptic ulcer disease	
Omeprazole	Prilosec		Can increase carbamazepine concentrations
Gastrointestinal stimulants		Cosmogenics	
Cisapride	Propulsid		
Metoclopramide	Reglan		Can cause depression and less frequently dystonic reactions and parkinsonism
Antiemetics		Nausea	
Prochlorperazine	Compazine		Prochlorperazine is an phenothiazine and like others can increase tricyclic drug levels
Dronabinol	Marinol		A cannabinoid found in marijuana; can cause drowsiness, dizziness, euphoria
Ondansetron	Zofran		A selective serotonin (5-hydroxytryptamine [5-HT]) type 3 receptor antagonist; interactions with SSRIs unclear
Antidiarrheal Agents			
Interferon alpha	Intron A	Hepatitis C	Depression in up to 10%, also insomnia, anxiety, constipation
Metronidazole	Flagyl	Giardiasis, Helicobacter	Potential nephrotoxic effect when used with lithium, interacts with disulfiram

Table 25.2-6 Common Drugs Used for Gastrointestinal Disorders—Psychiatric Effects and Interactions With Psychiatric Drugs

Cimetidine (Tagamet) has been available for over 20 years as a histamine receptor antagonist blocking gastric acid secretion. Cimetidine appears to have some anticholinergic effect and can produce delirium, especially in hospitalized patients with severe medical or surgical illness. The newer histamine receptor antagonists (ranitidine [Zantac], nizatidine [Axid], and famotidine [Pepcid]) have fewer adverse central nervous system (CNS) effects and drug interaction problems with psychotropic drugs. Cimetidine use has increased in many hospitals to reduce hospital costs and is a preferred drug in some managed care settings. This change has resulted in a risk for more adverse CNS and drug interaction problems. Omeprazole (Prilosec) is a blocker of acid secretion that works by inhibiting the proton pump in gastric cells. A possible drug interaction with carbamazepine (Tegretol) that increases carbamazepine concentrations has been reported.

Gastrointestinal stimulants target delayed gastric emptying associated with diabetes and advanced age. Two drugs in this class include cisapride (Propulsid) and metoclopramide. Metoclopramide (Reglan) appears to induce a depressive state in some individuals and can also occasionally cause neuroleptic-induced adverse effects. Parkinsonian symptoms noted with metoclopraminde include dystonic reactions, akathisia, and muscular rigidity. Dystonia associated with metoclopramide has been noted in up to 25 percent of children and young adults treated for chemotherapy-associated nausea. Prophylactic diphenhydramine (Benadryl) is used commonly to reduce the incidence of metoclopramide-associated dystonia. Metoclopramide doses may need to be reduced with renal compromise to avoid adverse effects.

Akathisia related to metoclopramide appears commonly in elderly patients receiving the drug. This may be related to reduced renal function and reduced clearance of standard doses in the geriatric age group. Renal impairment may result in high blood concentrations of metoclopramide and higher rates of adverse effects in this age group.

Antipsychotic drugs reduce nausea and vomiting. Some common antipsychotic drugs like haloperidol (Haldol) and chlorpromazine are used in the medical setting for their antiemetic effect. The antipsychotic prochlorperzine (Compazine) can produce the adverse effects of other phenothiazines and potentiate the sedative effects of some psychotropics. Like other phenothiazines, concurrent tricyclic drug administration results in higher blood concentrations of the tricyclic agents. As interferon therapy for hepatitis C carries a risk of depression with reports of suicide attempts and completed suicide in several cases, some centers recommend psychiatric consultation prior to the institution of interferon therapy.

Metronidazole is commonly prescribed for giardiasis and peptic ulcer due to *Helicobacter* infection. This agent may increase risk for nephrogenic toxicity associated with lithium (Eskalith) and also can produce a disulfiram (Antabuse)-like reaction when alcohol is consumed.

Psychiatrists and other physicians prescribing psychotropic agents need to consider these potential interactions when treating patients with gastrointestinal diseases. Collaboration between psychiatrists and physicians treating gastrointestinal disorders is frequently necessary for patients with combined illnesses. This collaboration is often also necessary for separating out gastrointestinal and psychiatric causes in the differential diagnosis of complex cases.

A 30-year-old woman developed fatigue and consulted her family physician. During her teenage years, she had a period of experimentation with alcohol and drug use. For a brief period, she used intravenous methamphetamine. After several years of intermittent drug use she stopped all drug use at her boyfriend's insistence. They later married, and she had two children without obstetrical complications.

Her family physician completed a medical workup. Blood tests revealed elevated liver transaminase levels. She was referred to a gastroenterologist for further diagnostic workup and treatment. Serum tests and a liver biopsy confirmed a diagnosis of hepatitis C. Interferon therapy was recommended, and she was referred for psychiatric consultation prior to initiating interferon to rule out psychiatric contraindications to interferon therapy. Psychiatric consultation identified her previous drug use but showed no evidence for a history of depression or active depression. No contraindications to interferon therapy were found.

Interferon therapy began, and she did well for several weeks. However, she began to notice mood lability, irritability, and sleep disturbance. Her symptoms distressed her a great deal, and she began to have difficulty with occupational and social functioning. She denied suicidal ideation. Psychiatric reevaluation resulted in a diagnosis of mood disorder due to interferon. Her initial liver response to interferon was excellent, with liver transaminase levels reduced to normal. Since her response to interferon had been very positive, her gastroenterologist and psychiatrist elected to continue the interferon therapy while treatment with an SSRI was initiated.

Following 3 weeks of antidepressant therapy, her mood symptoms improved, including a return to normal sleep patterns. She continued antidepressant therapy during the year of interferon therapy. Her mood remained normal during interferon therapy. Following completion of interferon therapy, her antidepressant medication was tapered without recurrence of depression.

Discussion This case illustrates the potential psychiatric effects of drugs used for gastrointestinal disorders. Interferon can produce a significant mood disorder in some individuals. Since suicide has been reported during interferon therapy, mood symptoms should not be ignored. The *Physician's Desk Reference* (PDR) includes risk for suicide and suicidal ideation in the warning section for interferon with a note that depression may occur in up to 15 percent of patients treated with interferon alpha. Additionally, it notes that if depression occurs during the course of interferon, discontinuation of the drug does not guarantee clinical remission of the depressive disorder. The PDR goes on to note that suicide has occurred following the discontinuation of interferon. Early consultation with a psychiatrist can facilitate coordinated care. Active psychiatric treatment can allow successful completion of medical therapy.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of gastrointestinal complaints requires careful attention to the medical and psychiatric history as well as the pattern of the patient's symptoms. A psychiatric cause is more likely when the symptoms are chronic, occur in the context of multiple somatic symptoms, and occur in a patient with a psychiatric comorbid Axis I or Axis II diagnosis or multiple psychosocial stressors. A medical cause is more likely when these factors are absent and the patient has a physical examination (e.g., fever or abnormalities on laboratory testing).

A 45-year-old woman presented to her primary care physician with complaints of chronic intermittent abdominal pain and diarrhea. Medical workup did not reveal a cause for her symptoms, and a diagnosis of irritable bowel syndrome was made. The patient received treatment with fiber supplementation and bulking agents without relief. Various antidiarrheal medications did not improve her complaints of frequent loose stools. In an attempt to provide relief, her primary care physician prescribed acetaminophen with codeine. The patient reported some improvement in her symptoms with this medication.

Over time the patient's symptoms worsened, despite the acetaminophen with codeine. She requested larger doses than she was given and refilled prescriptions before they were scheduled to expire. She started to make desperate claims to her physician that she didn't know if she could go on if there was no improvement in her pain. Her physician referred her for psychiatric consultation to assess for somatization, depression, and possible prescription drug abuse.

On psychiatric evaluation, the patient denied any psychiatric problems, claiming she had only a physical problem. She had a history of recurrent pain complaints dating back to age 22. These pain complaints involved multiple sites and had often lacked evidence of a medical cause. On further questioning, it became apparent she was receiving narcotic pain prescriptions from several physicians. Her physicians were unaware that she had other sources of prescription drug supply. Her family history was positive for alcohol dependence in her father and alcohol and drug dependence in her brother. A diagnosis of somatization disorder and prescription drug abuse was made.

Discussion The case demonstrates the relation of somatization disorder to a functional gastrointestinal disorder. Patients with functional gastrointestinal disease who do not unrespond to treatment may have psychiatric comorbidity that affects the medical long-term prognosis. Patients with multiple functional complaints need careful coordination of care. Prescription drug abuse and misuse by these individuals is common and requires careful management.

COURSE AND PROGNOSIS

The course and prognosis of many of the functional somatic syndromes are often favorable, given time and reassurance. One group of patients presenting to a

physician with globus followed and interviewed after 15 years showed only 10 percent with continued symptoms, 43 percent had no symptoms in the interval and a similar number had minor recurrences during the follow-up period.

Similarly, studies of irritable bowel patients suggest an overall favorable course for most patients. Studies 2 and 5 years after diagnosis report that 70 to 80 percent of patients were improved. Compliance with a high-fiber diet and bulking agents appears to improve the prognosis. The longest outcome study of irritable bowel syndrome covered an average of 29 years following the index evaluation. An organic gastrointestinal disorder was detected in the follow-up in only 10 of 112 patients. Survival in irritable bowel syndrome was identical to expected survival, and most patients demonstrated improvement with time.

Some evidence indicates that men with irritable bowel syndrome may respond better to treatment than women and that those with constipation-predominant symptoms do better than those with diarrhea-predominant symptoms. Those with a shorter duration of illness have a more favorable prognosis. Irritable bowel syndrome patients with high anxiety measures or other psychiatric comorbidity have a poorer prognosis than those without anxiety. More-recent treatment studies suggest a better prognosis with completion of psychotherapy.

Follow-up studies of functional dyspepsia also appear to support an overall favorable prognosis. In a series of 102 patients followed 6 years, 76 percent reported no symptoms of dyspepsia, and only 3 percent developed a peptic ulcer. Not all studies of dyspepsia support a favorable prognosis. Case selection factors may influence the likelihood of improvement.

Successful attention to psychological factors and the development of a strong patient relationship can contribute to an improved prognosis for functional gastrointestinal disorders. In a series of patients with irritable bowel syndrome, a strong patient-physician relationship appeared to decrease the frequency of return visits needed for the management of the syndrome. Some physician behaviors increase the likelihood of successful management of irritable bowel syndrome patients. Primary care physicians who attend to important factors in the psychosocial history and to symptom precipitants appear to increase the rate of successful outcome. Providing reassurance and discussing psychologic issues in irritable bowel syndrome also appear to favorably influence outcome.

TREATMENT

Psychotropic Treatment Psychotropic drug use is common in the treatment of a variety of gastrointestinal disorders and is complicated by disturbances in gastric motility, absorption, and metabolism related to the underlying gastrointestinal disorder. Many gastrointestinal effects of psychotropic drugs can be used therapeutically in functional gastrointestinal disorders. For example, tricyclic antidepressant agent can be used to reduce gastric motility in patients with irritable bowel syndrome with diarrhea. However, pscychotropic gastrointestinal side effects can exacerbate a gastrointestinal disorder (e.g., prescribing a tricyclic drug to treat a depressed patient with gastroesophageal reflux).

Psychopharmacological agents provide one step in a stepped-care approach to the management of irritable bowel syndrome. For milder cases of irritable bowel syndrome, education, reassurance, and avoidance of foods that increase symptoms forms the base of treatment. Moderate cases of irritable bowel frequently respond to nonpsychopharmacological drug treatment targeted at the patient's primary symptom. Dicyclomine (Bentyl) can be helpful for abdominal pain, and loperamide (Imodium) frequently provides relief for diarrhea associated with irritable bowel syndrome.

Tricyclic antidepressants can significantly reduce irritable bowel symptoms for many more-severe cases of irritable bowel syndrome and usually are reserved for more severe cases. The exact mechanism for tricyclic drug efficacy is unknown but may include controlling pain through central analgesia, reduced colonic motility because of a direct anticholinergic action on the gut, or reduction in comorbid depression. Retrospective studies of outpatients with irritable bowel syndrome found that 89 percent of patients had improvement in bowel symptoms with use of tricyclics or anxiolytics. Sixty-one percent of patients reported complete remission of symptoms. The most frequently used tricyclic drugs in this series were amitriptyline (about 50 mg a day) and doxepin (Adapin) (about 25 to 50 mg daily). These response rates to relatively low doses of tricyclic drugs suggest that the mechanism is not likely to involve remission of a comorbid depressive disorder.

Antidepressant treatment can also reduce symptoms in other functional disorders. In a placebo-controlled trial, trazodone (Desyrel) (100 to 150 mg daily) significantly reduced esophageal symptoms in a series of patients. Improvement in this series of patients did not correlate with reduction in esophageal contraction abnormalities, presence of psychiatric illness at entry, or changes in measures of anxiety or depression. Reduction of functional esophageal symptoms by antidepressants does not appear to depend on psychiatric comorbidity, suggesting that all patients with the gastrointestinal symptoms should be considered for treatment. Tricyclic drugs appear effective for the management of functional nausea and vomiting. An open-label study documented moderate reduction in symptoms of functional nausea and vomiting in 84 percent of patients taking an average dosage of 50 mg a day.

Gastrointestinal diseases can complicate the administration of psychotropic agents. Patients unable to take oral agents because of acute or chronic gastrointestinal diseases may need to use available parenteral forms of psychotropic agents. Lorazepam (Ativan) can be administered by the sublingual, intravenous, or intramuscular route. Many antipsychotic agents have parenteral forms including haloperidol, fluphenazine (Prolixin), trifluoperazine (Stelazine), thiothixene (Navane), perphenazine (Tritofon), and loxapine (Loxitane). Haloperidol has been commonly used by the intravenous route although this route is not approved by the Food and Drug Administration (FDA). Intensive care unit management of agitation associated with delirium often includes intravenous haloperidol as one treatment option, particularly for patients without evidence of rhythm abnormalities or other cardiac conditions.

Psychotropic drug treatment is complicated by acute and chronic liver disease. Most psychotropic agents are metabolized by the liver, and many of them can be associated with hepatoxicity. When acute changes in liver function tests occur with tricylic drugs, carbamazepine, or antipsychotic agents, it may be necessary to discontinue the drugs. During periods of discontinuation, lorazepam or lithium can be used, because they are excreted by the kidney. Electroconvulsive therapy (ECT) can also be used in the patient with liver disease, although the anesthesiologist must carefully select anesthetic agents that carry a minimal risk for hepatotoxicity.

Antidepressant treatment for patients on total parenteral nutrition can be a significant clinical challenge. Common antidepressant agents do not have parenteral forms. Fluoxetine is available in a liquid form that may be used for some patients who can tolerate and absorb some oral intake. ECT can be accomplished for patients without the option for oral antidepressant agents. For some patients with gastrointestinal diseases and disorders, aggressive use of psychotherapy may be the most appropriate choice in providing psychiatric treatment.

Psychotherapy Psychotherapy can be a key component in the stepped-care approach to the treatment of irritable bowel syndrome and other functional gastrointestinal disorders. Multiple different models of psychotherapy have been applied to clinical samples of patients with irritable bowel syndrome, including short-term, dynamically oriented individual psychotherapy, supportive psychotherapy, hypnotherapy, relaxation techniques, and cognitive therapy. There appears to be a significant placebo effect for psychotherapy in this population for both pharmacological and psychological treatment approaches. At least one study found the response to psychological treatment alone for irritable bowel syndrome to be comparable to that of pharmacological treatment. Patients with significant Axis I psychopathology appear to be less responsive to psychological intervention than those without. This may be related to the increased severity of irritable bowel syndrome symptoms associated with psychiatric comorbidity.

Most studies of psychotherapeutic approaches to the treatment of irritable bowel syndrome have been case series, not controlled trials. Some well-controlled studies are beginning to be published that use various models of psychotherapy. A cognitive therapy model for the treatment of irritable bowel syndrome has been developed that considers irritable bowel syndrome to be autonomic nervous system response to stress, with three components: cognitions, behaviors, and physiological response. Cognitions form the key element in this therapy. Therapy focuses on identifying life stressors and the thoughts associated with these stressors. Subjects are taught to identify threatening stimuli and modify their appraisals and interpretations of such stimuli. Subjects record automatic thoughts in their daily lives, and the therapist focuses on identifying central themes. Therapeutic work centers around three processes: rational self-analysis, decentering, and experimental disconfirmation.

Using this model, a study of 20 patients compared gastrointestinal symptom response of 10 subjects receiving cognitive therapy with that of 10 subjects who simply monitored their gastrointestinal symptoms; 80 percent of the cognitive therapy group showed clinically significant improvement compared with only 10 percent of the control. These small, early studies suggest that specific psychotherapy approaches may become a key intervention in irritable bowel syndrome and other functional gastrointestinal syndromes.

Because of the lack of controlled studies, it is not possible to recommend a specific model of psychotherapy for irritable bowel syndrome. Controlled comparison studies with rigorous selection criteria are needed before conclusions about the relative efficacy of different psychotherapeutic models can be made. Despite this lack

of information, psychological treatments appear promising and can be a legitimate part of a comprehensive treatment plan for irritable bowel syndrome.

SUGGESTED CROSS-REFERENCES

Sections related to the gastrointestinal disorders include Section 25.2, Section 1.11 on psychoneuroendocrinology, Section 1.4 and Section 1.5 on neurotransmitters, Section 25.1 on psychosomatic medicine, and Section 25.11 on psycho-oncology.

SECTION REFERENCES

Aggarwal A, Cutts T, Abell T, Cardoso S, Familoni B, Bremer J, Karas J: Predominant symptoms in irritable bowel syndrome correlated with specific autonomic nervous system abnormalities. Gastroenterology *106:*945, 1994.

Blanchard E, Schwarz S, Suls J, Gerardi MA, Scharff L, Greene B, Taylor AE, Berreman C, Malamood HS: Two controlled evaluations of multicomponent psychological treatment of irritable bowel syndrome. Behav Res Ther *30:*175, 1992.

Clouse R, Lustman P, Geisman R, Alpers DH: Antidepressant therapy in 138 patients with irritable bowel syndrome: A five-year clinical experience. Aliment Pharmacol Ther *8:*409, 1994.

Clouse RE: Antidepressants for functional gastrointestinal symptoms. Dig Dis Sci *39:*2352, 1994.

Drossman DA: Diagnosing and treating patients with refractory functional gastrointestinal disorders. Ann Intern Med *123:*688, 1995.

Drossman DA: Presidential address: Gastrointestinal illness and the biopsychosocial model. Psychosom Med *60:*258, 1998.

*Drossman D, Li Z, Andruzzi E, Temple RD, Talley NJ, Thompson WG, Whitehead WE, Janssens J, Funch-Jensen P, Corazziari E: U.S. householder survey of functional gastrointestinal disorders: Prevalence, sociodemography, and health impact. Dig Dis Sci *38:*1569, 1993.

*Drossman DA, Thompson WG: The irritable bowel syndrome: Review and a graduated multicomponent treatment approach. Ann Intern Med *116:*1009, 1992.

Drossman DA, Thompson WG, Talley NJ, Funch-Jensen P, Janssens J, Whitehead WE: Identification of sub-groups of functional gastrointestinal disorders. Gastroenterol Int *3:*159, 1990.

Epstein SA: Psychotropic medications in gastrointestinal and hepatic disease. Adv Psychosom Med *21:*49, 1994.

Folks D, Kinney F: The role of psychological factors in gastrointestinal conditions: A review pertinent to DSM-IV. Psychosomatics *33:*257, 1992.

Greene B, Blanchard E: Cognitive therapy for irritable bowel syndrome. J Consult Clin Psychol *62:*576, 1994.

Kay L, Jorgensen T, Jensen KH: The epidemiology of irritable bowel syndrome in a random population: Prevalence, incidence, natural history, and risk factors. J Intern Med *236:*23, 1994.

*Kellner R: *Psychosomatic Syndromes and Somatic Symptoms*. American Psychiatric Press, Washington, DC, 1991.

Lydiard RB, Greenwald S, Weissman MM, Johnson J, Drossman DA, Ballenger JC: Panic disorder and gastrointestinal symptoms: Findings from the NIMH Epidemiologic Catchment Area Project. Am J Psychiatry *151:*64, 1994.

Maunder RG: Panic disorder associated with gastrointestinal disease: review and hypotheses. J Psychosom Res *44:*91, 1998.

*Mayou R, Bass C, Sharpe M, editors: *Treatment of Functional Somatic Symptoms*. Oxford University Press, Oxford, 1995.

*North CS, Alpers DH, Thompson SJ, Spitznagel EL: Gastrointestinal symptoms and psychiatric disorders in the general population. Findings from the NIMH Epidemiologic Catchment Area Project. Dig Dis Sci *41:*633, 1996.

Noyes R, Cook B, Garvey M, Summers R: Reduction of gastrointestinal symptoms following treatment for panic disorder. Psychosomatics *31:*75, 1990.

Owens D, Nelson D, Talley N: The irritable bowel syndrome: Long-term prognosis and the physician-patient interaction. Ann Intern Med *122:*107, 1995.

Peterson WG: *Helicobacter pylori* and peptic ulcer disease. N Engl J Med *324:*1043, 1991.

**Physicians Desk Reference*, ed 53. Medical Economics Company, Montvale, NJ, 1999.

Prakash C, Lustman PJ, Freedland KE, Clouse RE: Tricyclic antidepressants for functional nausea and vomiting: clinical outcome in 37 patients. Dig Dis Sci *43:*1951, 1998.

Richards H, Prendergast M, Booth I: Psychiatric presentation of Crohn's disease: Diagnostic delay and increased morbidity. Br J Psychiatry *164:*256, 1994.

Smith R, Greenblaum D, Vancouver J: Psychosocial factors are associated with health care seeking rather than diagnosis in irritable bowel syndrome. Gastroenterology *98:*293, 1990.

Stoudemire A: Somatothymia. Psychosomatics *32:*365, 1991.

Walker EA, Katon WJ, Jemelka RP, Roy-Byrne PP: Comorbidity of gastrointestinal complaints, depression, and anxiety in the Epidemiologic Catchment Area (ECA) Study. Am J Med *92:*26S, 1992.

Whitehead WE: Psychosocial aspects of functional gastrointestinal disorders. Gastroenterol Clin North Am *25:*21, 1996.

Woodman C, Noyes R: The relationship between panic disorder and irritable bowel syndrome: A review. Ann Clin Psychiatry *4:*175, 1992.

Yates WR, Gleason O: Hepatitis C and depression. Depress Anxiety *7:*188, 1998.

25.3 OBESITY

KELLY D. BROWNELL, PH.D., AND THOMAS A. WADDEN, PH.D.

Obesity is a problem as mysterious as it is pervasive. At first glance the cause is apparent—an individual consumes more energy than the body expends. The more complex question is why this occurs. Society's view that overweight individuals are responsible for their problem and that an overweight body reflects a troubled personality is at odds with a robust literature yet persists to the point that severe bias and discrimination against overweight persons are widespread. Health professionals are not immune from this bias. Overweight persons often feel blamed, scolded, and overtly criticized by professionals to the point they are much less likely to seek medical help for a given disease than are persons of normal weight.

What has been missing but is becoming more common is a social and scientific climate where the causes of obesity can be evaluated in the absence of social bias, the disease gets the needed attention given its toll on the public health and the lives of those affected, prevention can become a priority, the resources necessary to develop safe and effective treatments will be forthcoming, health care professionals skilled in the management of obesity and its complications are more abundant, and compassionate treatment is the norm.

DEFINITION

Obesity refers to an excess of body fat. In healthy individuals body fat accounts for approximately 25 percent of body weight in women and 18 percent in men. Estimating body fat accurately is expensive, although advances in the assessment of body composition may change this picture. *Overweight* refers to elevated weight above some reference norm, typically standards derived from actuarial or epidemiological data. In most cases increasing weight reflects increasing obesity, but not always. Muscular individuals might be overweight (weight might be elevated given height) but not be obese, and a person might have normal weight but have elevated body fat.

Indexes have been developed using height and weight to estimate level of obesity. The most common of these is the body mass index (BMI). BMI is calculated by dividing weight in kilograms by height squared in meters. Although there is debate about the ideal BMI, it is generally thought that a BMI of 20 to 25 kg/m ^{2} represents healthy weight, a BMI of 25 to 27 kg/m^2 is associated with somewhat elevated risk, a BMI of above 27 kg/m^2 is where the increase in risk is clear, and a BMI above 30 kg/m^2 is where there is greatly increased risk. Figure 25.3-1 presents a nomogram for calculating BMI from height and weight.

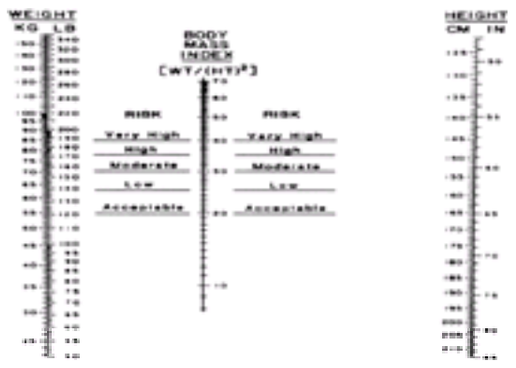

FIGURE 25.3-1 Nomogram for calculating body mass index (BMI) from height and weight. To determine BMI, place a ruler or other straight edge between the body weight column on the left and the height column on the right and read the BMI from the point where it crosses the center. (Reprinted with permission from GA Bray: Definition, measurement, and classification of the syndromes of obesity. Int J Obesity *2:*99, 1978.)

There has been a move in the obesity field to embrace BMI as an index of obesity. This is positive because BMI better estimates level of obesity than does weight per se, but there is the illusion that BMI avoids the problem of people focusing on “ideal” weight. However, ideal BMI levels do exist; it is only a matter of time before these are in wide use and the problems of overemphasis on ideal weights are still an issue, only using a different metric.

EPIDEMIOLOGY

The prevalence of obesity varies greatly by nation, by social groups within nations, and in some cases, within a given nation over time. In the United States approximately 35 percent of women and 31 percent of men are significantly overweight (BMI 27 or above—about 20 percent overweight). If one defines obesity as BMI over 25, there are now more obese than nonobese Americans. Using BMI over 31 (approximately 40 percent overweight), 11 percent of women and 8 percent of men are severely overweight. The prevalence of obesity in America has tripled since the early 1900s. Through the 1960s, 1970s, and early 1980s there was a steady rise in prevalence, but the dramatic changes occurred in subsequent years; there has been a 25 percent increase since the 1980s alone.

The prevalence of obesity is highest in minority populations, particularly among women. Fully 60 percent of African-American women ages 45 years and older are overweight, as defined by a BMI of 27 or greater. The high prevalence of obesity among minorities appears attributable primarily to lower income and educational attainment rather than race. Obesity is six times more prevalent in women of low socioeconomic status as compared with women of high socioeconomic status.

Weight gain is most pronounced in both sexes between the ages of 25 and 44. During this time, men gain an average of 4 kg and women 7 kg. Pregnancy probably contributes to the greater increase in women, who on average, begin each successive pregnancy approximately 2.5 kg heavier than at the last. After age 50, weights of men stabilize, and even decline slightly between ages 60 and 74. Women, in contrast, continue to increase in weight until age 60, at which time weight begins to decline.

Obesity is a massive public health problem by any standard. The prevalence is extreme, is increasing rather than decreasing, and in what speaks to a bleak future, is especially high and growing in children. Obesity rivals the most serious illnesses in the public health toll it takes.

COURSE AND PROGNOSIS

Recent developments in treatment offer new hope for overweight individuals, but obesity remains one of the most refractory conditions confronted by health professionals. There is little evidence of spontaneous remission, and the natural history of obese individuals left untreated is weight gain over time. As people age, the likelihood of chronic diseases such as diabetes, hypertension, cardiovascular disease, and cancer increases, due in part to the contribution of weight and the

behaviors that affect weight (poor diet and sedentary life-style).

Medical Complications There is a clear association of obesity with serious diseases and with mortality. Figure 25.3-2, from a study of 750,000 men and women from the general population by the American Cancer Society (ACS), shows a curve that has been replicated a number of times. An orderly relationship is seen between increasing weight and the risk of premature death. The major contribution to mortality in overweight persons in the ACS study was from coronary heart disease. At the low end of the weight curve, there was an increase in mortality attributable to low weight and excess mortality occurring in smokers. When smokers are excluded from such analyses, the relationship between weight and mortality falls nearly on a straight line. The minimum mortality ratio in the ACS study occurred at a BMI of approximately 22, but there was no substantial increase in mortality until BMI exceeded 27.

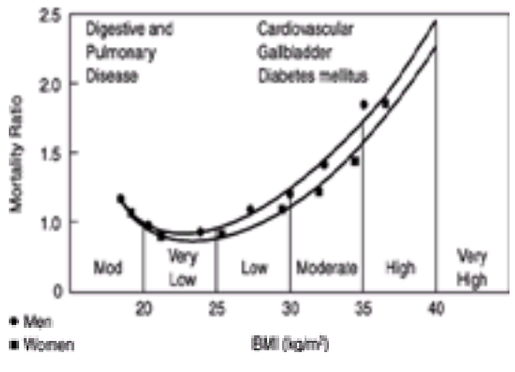

FIGURE 25.3-2 Obesity and mortality risk in the American Cancer Society Study of 750,000 men and women from the general population. Mortality rates were calculated as a ratio of observed deaths to expected deaths. Increased mortality in those at the lowest weights is attributable primarily to both low weights and elevated mortality occurring in smokers. (Data abstracted from Lew EA, Garfinkel L: Variations in mortality by weight among 750,000 men and women. J Chron Dis *32:*563, 1979.)

Table 25.3-1 lists the medical conditions most often associated with obesity. Nearly every organ system and part of the body can be affected, although this varies from person to person and depends in part on the extent of the obesity. Some weight-sensitive people will have comorbid problems such as hypertension and diabetes with only a small degree of overweight whereas others may gain large amounts of weight and have little or no evidence of medical problems.

Cardiovascular disease	Degenerative joint disease
Type II diabetes	Some cancers
Hypertension	Gallstones
Dyslipidemia	Gynecological irregularities
Sleep apnea	

Table 25.3-1 Major Health Consequences of Obesity

At one point there was lively debate in the obesity field about whether excess weight is a risk factor for morbidity and mortality independent of its influence on other disease processes (e.g., hypertension, diabetes, dyslipidemia). This question has been put to rest, not only because obesity does cause these comorbid problems in many people, and because the poor diet and inactive life-style that lead to obesity are themselves dangerous, but because studies have shown an independent contribution of obesity beyond its effect on risk factors. As shown in Table 25.3-2, the risk for a number of serious diseases increases with rising weight.

	20–25 kg/m²	25–30 kg/m²	30–35 kg/m²	>35 kg/m²
Diabetes	3.7	7.4	12.2	15.3
Hypertension	17.7	29.4	36.7	39.4
Elevated cholesterol	12.6	18.6	19.5	14.0
Coronary heart disease	9.9	14.8	16.8	16.9
Depression	12.8	13.5	16.4	22.4
Musculoskeletal pain	21.2	24.1	28.3	30.6

Data abstracted from Queseberry CF Jr, Caan B, Jacobson A: Obesity, health services use, and health care costs among members of a health maintenance organization. *Arch Int Med* 158:466, 1998.

Table 25.3-2 Percentage of Patients With Medical Conditions by Increasing Levels of Body Mass Index (BMI)

The most common diseases associated with obesity are coronary heart disease, diabetes, hypertension, gallbladder disease, sleep apnea, and some cancers. For many people, joint problems, difficulty breathing, and the psychosocial consequences of obesity are of greatest personal concern because of their effect on the quality of life.

There has been strident debate about the exact weight at which risk begins to increase, with some estimates being as low as 5 percent overweight and others being six times or more that figure. In addition, some have argued that tables of ideal weights need to vary by age and that some weight gain with age is healthy. Prevailing sentiment is that risk increases when BMI exceeds 27, which is approximately 20 percent overweight.

Body Fat Distribution Body fat distribution is a better predictor of cardiovascular morbidity and mortality than is weight, total body fat, or BMI. Persons whose fat is confined primarily to the upper body (more common in males) are at greater risk of type II diabetes and cardiovascular morbidity and mortality than are persons whose adiposity is carried primarily in the lower body. Within the upper body, increased visceral fat is more damaging to health than is subcutaneous fat. The mechanisms by which visceral fat leads to increased metabolic abnormalities are not well understood but appear to be related to an increase in portal free fatty acid concentrations, which are associated with elevated hepatic gluconeogenesis and very–low-density lipoprotein secretion as well as decreased hepatic insulin clearance.

Findings for fat distribution have important implications. Such data suggest possible factors that would argue in favor of targeting the most aggressive treatments at those at greatest risk. For example, women are much more likely than men to diet and to enroll in weight loss programs, yet men may have a more compelling reason to do so, given their greater likelihood of having upper body fat distribution.

Social and Psychological Consequences It is a clear social disadvantage to be obese, and in some cases, even to be mildly overweight, particularly for women. In

a culture where beauty ideals are pathologically thin and highly unrealistic, overweight people are blamed for their condition and are the subject of teasing, bias, discrimination, and torment. Income and earning power are lower in overweight people, and untoward social conditions such as absence of romantic relationships are more common.

The bias against overweight people begins at early ages, and both the general public and health care professionals are involved. Children as young as age 6 describe obese children shown in silhouettes as “lazy,” “dirty,” “stupid,” “cheats,” and “lies.” When shown drawings of an obese child and children with various handicaps, including facial disfigurement, both children and adults pick the obese child as the one they would least like to play with.

Health care providers appear to share this prejudice. Some research finds physicians describing their obese patients as “weak-willed, ugly, and awkward.” Patients appear to be conscious of such attitudes, as indicated by the results of a recent study of individuals who underwent surgery for their obesity. Approximately 80 percent of respondents indicated that they had “been treated disrespectfully by the medical profession” because of their weight. The finding from a study in the family practice literature that overweight people are less likely to seek medical help for a given condition than are people without weight problems reflects embarrassment, shame, and fear of being criticized.

The literature on the psychological consequences of obesity has been mixed. Several studies report no differences in psychological function (using general measures of psychopathology such as depression and anxiety) between overweight and nonoverweight individuals. This seems counterintuitive given the strong social bias and the recurrent clinical observation that many overweight individuals internalize the negative messages and make very negative attributions for their problem. More recent papers have noted the methodological shortcomings of earlier work and propose that obesity does increase risk for psychological problems, particularly in the presence of possible risk factors such as binge eating.

Health Care Costs Its massive prevalence and serious effects on health make obesity a major contributor to health care costs. In the early 1990s the health care costs of obesity were estimated at $40 billion annually. More sophisticated and recent data show even more disturbing figures. The estimated annual cost of obesity, including direct costs (costs of preventive, diagnostic, and treatment services related to disease) and indirect costs (value of lost wages, value of future earnings lost due to premature mortality) is $51.64 billion annually, which is 5.7 percent of the national health care expenditure.

On a more local level, a study of 17,118 patients in the Kaiser Permanente Medical Care Program in northern California examined health care costs across different levels of weight. Compared to patients at normal weight, patients with a BMI between 30 to 35 kg/m^2 had 25 percent higher annual costs. The costs were 44 percent higher for those with BMI greater than 35 kb/m^2. These incremental costs attributable to obesity accounted for 6 percent of the total healthcare costs in the Kaiser health plan.

Quality of Life The toll of obesity to individuals can be seen in discrete medical conditions, psychological disturbances, and social disadvantages. Recent attempts have been made to quantify the overall impact on quality of life. Table 25.3-3 shows a comparison of individuals in different weight groups in various quality-of-life domains. These figures show clearly increased problems in living with elevated weight. The problems are in areas of disease and general health, and also in mental health and social and emotional functioning. Excess weight can have a profound effect on the quality of an individual's life.

SF-36 Item	Mildly Obese	Moderately Obese	Morbidly Obese
Physical functioning	85.6	79.4	51.9
Role—physical	85.0	77.8	46.3
Bodily pain	66.4	54.5	43.2
General health	72.0	65.7	54.3
Vitality	49.5	48.1	38.0
Social functioning	84.9	79.3	67.9
Role—emotional	85.4	75.3	69.3
Mental health	74.3	68.6	65.6

*In 278 mildly, moderately, and morbidly obese individuals using the SF-36. Figures represent mean SF-36 scores. Lower scores represent lower satisfaction with quality of life.
SF-36 = Medical Outcomes Study Short Form-36 Health Survey Data abstracted from Fontaine KR, Cheskin LJ, Barofsky I: Health-related quality of life in obese persons seeking treatment. *J Fam Pract* 40:265, 1996.

Table 25.3-3 Comparison of Quality-of-Life Scores*

ETIOLOGY

Obesity is a heterogeneous disorder with causes in the environment and in the individual's psychology and biology. The field has progressed from ideas that a single cause might exist and that obesity is a singular entity, to recognition that multiple causes are necessary to explain obesity across individuals, within the same individual at a given time, and even within a given individual over time.

Fat storage remains the end product of consuming more energy than the body requires, but a variety of factors may contribute to positive energy balance. The etiology of obesity can be examined in terms of risks to the population as a whole, and of genetic and behavioral factors that place individuals in the population at even greater risk. Illnesses that give rise to obesity are also discussed.

Risks to the Population Merely by residing in the United States a person has a greatly increased risk of obesity. With the highest prevalence in the world and climbing, it is important to examine the American environment in the search for causes; epidemiological data from other countries can also be informative in this context.

There are data from country after country showing increased prevalence of obesity as cultures become more like that of the United States. Prevalence is increasing in nearly every such country as development and modernization occur. The rate of increase in the United Kingdom, for example, is much the same as that in the United States, but at a lower starting point. In peoples as diverse as inhabitants of Mauritius (an island in the Indian Ocean), Alaskan Eskimos, natives of Samoa, Pima Indians living in the United States and Mexico, and immigrants from countries like India and Japan moving to countries like the United Kingdom and the United States, the risk of obesity increases dramatically as individuals are exposed to a changed environment.

Energy balance in the U.S. population is affected in negative ways by the food supply and diminishing need to be physically active. Both have changed in dramatic fashion in recent years.

High-Fat, High-Calorie, Palatable Diet From 1910 to 1985 the percentage of calories in the U.S. diet derived from fat increased from 32 to 43 percent, whereas that from carbohydrate declined from 57 to 46 percent. Protein remained relatively constant at about 11 percent. An extensive body of literature shows that both animals and humans gain weight when fat in the diet is high. Weight gain results from increased caloric intake, attributable to the high energy value of fat (9 kcal/g for fat versus 4 kcal for protein and carbohydrate), and to fat's palatable flavor and texture. In addition, dietary fat is converted to body fat with approximately 25 percent greater efficiency than is carbohydrate, such that a person in positive energy balance will gain more weight when consuming fat than when consuming the same number of calories from carbohydrate. Premium ice creams and a host of sweet and salty foods derive 50 percent or more of their calories from fat.

High-calorie, high-fat, and highly palatable foods are ubiquitous. Fast-food restaurants have exploded in number, most now with drive-through windows, breakfast, packaged meals, and massive serving sizes. Most gas stations have been remodeled to have convenience food markets. Malls have food courts, airports and train stations have fast-food franchises, and there are even cases of fast-food restaurants in the lobbies of children's hospitals and in schools. In no country on earth is food more available, promoted, and inexpensive. Some experts have labeled this a “toxic environment” and claim that the outcome is entirely predictable—climbing rates of obesity.

The American lifestyle also contributes. Americans are eating out in record numbers, with the greatest expenditures occuring at fast-food restaurants. Fully 7 percent of the U.S. population eats at McDonald's on any given day. More processed foods are available. In the late 1990s the majority of the best-selling new food products were snack foods and beer. With single-parent families and a stressful lifestyle, less and less cooking is done at home, and good nutrition becomes more difficult to

achieve.

Declining Physical Activity Decreased physical activity is also a major threat to the nation's weight control. Daily energy expenditure decreased substantially as this nation changed from an agricultural to an industrial economy and, more recently, from a service to an information economy. From 1965 to 1977 alone, energy expenditure was thought to decrease by approximately 200 kcal per day. The increasing reliance upon automobiles, elevators, escalators, extension telephones, computers, and remote control devices all contribute to increasing obesity.

GENETIC RISKS

Although all Americans, as a result of national life-style changes, are at risk of weight gain, persons at greatest risk are those with a genetic predisposition to obesity. Studies of twins and adoptees have shown that body weight is under substantial genetic control. Current estimates are that 35 to 40 percent of the variance in BMI is attributable to genes and the remainder to environmental influences or gene-environment interactions.

Recent studies have identified five separate genes that cause obesity in laboratory animals. The best-known animal model is the *ob*/*ob* mouse. This animal grows markedly obese because of a deficiency in leptin, a protein that is synthesized in adipose tissue and acts on central neural networks that control ingestive behavior and energy expenditure. Under normal conditions leptin inhibits food intake and increases energy expenditure. As an animal's adipose mass increases, leptin secretion increases proportionately, thus reducing food intake and eventually body fat. The *ob*/*ob* mouse, because of a mutation in the ob gene, produces no leptin. As a result the animal is hyperphagic and extremely sedentary, producing its marked obesity. Peripheral or central administration of recombinant leptin in the *ob*/*ob* mouse results in rapid normalization of food intake, activity, and body weight.

With the discovery of the *ob* gene, investigators expected to find low leptin concentrations in obese humans. To date, only a handful of individuals have been identified who produce little or no leptin (comparable to the *ob*/*ob* mouse). Instead, the overwhelming majority of obese individuals have high circulating leptin levels; the higher the body fat, the higher the leptin excretion. Attention has now shifted to whether some obese individuals are leptin insensitive, in the same manner that many overweight individuals develop insulin insensitivity. Clinical trials are currently investigating whether leptin administration in humans reduces body weight.

Energy Expenditure With a few exceptions, such as Prader-Willi syndrome, researchers have not identified the specific gene(s) responsible for human obesity or the mechanisms of transmission. However, inherited differences have been shown not only in body fat but also in body fat distribution (i.e., upper- versus lower-body obesity), level of spontaneous physical activity, resting metabolic rate, and the thermic effect of food (i.e., energy expended in the digestion and absorption of food). Of these, resting metabolic rate (RMR), which accounts for approximately 60 to 75 percent of daily energy expenditure in sedentary individuals, appears the most significant. Longitudinal investigations of infants and adults have shown that those with low metabolic rates are more likely to become obese than are those with normal or elevated rates. Recent studies have also suggested that African-American women have lower resting energy requirements than Caucasian women, a factor that could contribute to the higher rates of obesity in black women.

Discovery of the leptin protein has greatly advanced the understanding of the complex neurohumoral regulation of body weight. Using the techniques of molecular biology, future research is likely to reveal that obese individuals differ from lean individuals in the mechanisms that regulate food intake (such as neuropeptide Y) or energy expenditure (such as uncoupling proteins). The identification of these mechanisms should provide specific targets for therapeutic intervention, whether by pharmacotherapy or gene therapy.

Behavioral Risks Behavioral and psychological factors also play a major role in the etiology of obesity. People makes choices about their eating and physical activity. A person's psychological well-being can influence weight and vice versa.

Energy Intake and Food Choices Investigators sought for years to determine whether obese individuals eat more than their average-weight peers; almost two dozen studies that relied upon self-reports of caloric intake concluded that they did not.

Recent studies have used doubly labeled water to assess 24-hour energy expenditure, and find that obese individuals typically underestimate their caloric intake by 30 to 40 percent, as compared with only 5 to 20 percent for average-weight controls. Hence, obese individuals generally do have higher caloric intakes than the nonobese, although the excess energy consumed is generally limited to that required to support the extra fat-free mass associated with obesity (approximately 25 to 30 percent of excess weight in significantly obese individuals consists of fat-free mass). These findings suggest that although practitioners should take seriously a patient's report of low caloric requirements, they must stress to the patient the need to keep very precise records of food intake.

Patients also make choices about the type of food they eat. This is influenced by family upbringing and by the general culture the individual is exposed to. Risk for obesity is increased in individuals who have preferences for high-fat, high-calorie food and in those who use food as a source of emotional comfort.

Binge Eating Approximately 20 to 30 percent of individuals who seek professional treatment of their obesity suffer from binge eating—episodes of uncontrolled eating in which they consume large quantities of food. Persons diagnosed with binge-eating disorder do not purge following overeating episodes, which distinguishes them from individuals with bulimia nervosa.

The absence of purging may contribute to the development of obesity, although there are few data on this topic. Little also is known about the etiology of binge eating. Early theorists proposed that binge eating resulted from severe caloric restriction (i.e., dieting), but more recent reports suggest that binge eating can precede dieting or may occur in the absence of dieting. This suggests the presence of serious psychological disturbance, created perhaps by interpersonal distress, leading the person to use food as a means of escaping from unpleasant life circumstances. It is common for binge eaters to report being “numb” during these episodes and using food as a means for blocking out the world.

Psychological Status With the possible exception of binge eating, there is little evidence to support popular notions that obesity results from overeating in response to feelings of anxiety or depression. Longitudinal investigations have not found psychological status to predict the development of obesity in children, and in cases in which psychopathology is observed in overweight adults, it is probably a consequence rather than cause of the obesity. A minority of obese individuals will present with significant psychological distress, but these must be addressed in treatment. However, practitioners should not expect that resolution of such distress will automatically result in weight loss.

A consistent finding in binge eaters is that of increased depression and psychopathology. As many as 50 percent of obese binge eaters present with clinically significant depression, as compared with only 5 percent of obese nonbingers, whose scores on depression scales typically fall within the normal range. It is not clear whether depression results from binge eating or causes it. Alternatively, the two disorders may be linked to a common neurotransmitter deficiency, given the favorable response of each to serotonergic agents such as fluoxetine (Prozac).

The psychological meaning of food to individuals who struggle with weight is an important topic that has been studied far too little. From clinical experience it is clear that some individuals use food to escape and food may become their best friend in many ways. They look forward to times when they can be alone with food and use food in the way others may use close personal relationships. This is generated by intense personal needs, but no research has occurred on the personality or situational factors that predispose an individual to this phenomenon, how exactly this contributes to weight gain, and how such individuals might be treated.

Illness Risks In only a small number of cases is obesity the consequence of identifiable illness. Such cases include a variety of rare genetic disorders, as well as the neuroendocrine abnormalities shown in Table 25.3-4. Hypothalamic obesity results from damage to the ventromedial region of the hypothalamus, which has been studied extensively in laboratory animals and is a known center of appetite and weight regulation. In humans, damage to the ventromedial region of the hypothalamus may result from trauma, surgery, malignancy, or inflammatory disease. If the patient has had a recent physical examination and laboratory tests, with satisfactory results, the disorders in Table 25.3-4 generally can be ruled out.

Genetic (dysmorphic) obesities
Autosomal recessive
X-linked
Chromosomal
Neuroendocrine obesities
Hypothalamic syndrome
Cushing's syndrome
Hypothyroidism
Polycystic ovary (Stein-Leventhal) syndrome
Pseudohypoparathyroidism
Hypogonadism
Growth hormone deficiency
Insulinoma and hyperinsulinism
Iatrogenic obesities
Drugs (psychiatric)
Hypothalamic surgery (neuroendocrine)

Adapted from Bray GA: An approach to the classification and evaluation of obesity. In Obesity, P Björntorp, BN Brodoff, editors. Lippincott, Philadelphia, 1992.

Table 25.3-4 Illnesses That Can Explain Some Cases of Obesity

Some forms of depression, particularly mood disorder with seasonal pattern, are associated with weight gain. Most persons who live in seasonal climates report increases in appetite and weight during the fall and winter months, with decreases in the spring and summer. These variations, however, are usually more marked in persons diagnosed with mood disorder with seasonal pattern (seasonal affective disorder). One study found that 67 percent of patients gained 4 kg or more during the winter and that many had trouble losing the weight over the summer; only 33 percent of normals gained a comparable amount. It has been reported that 38 percent of patients diagnosed with major depressive disorder gained a significant amount of weight, as compared with 51 percent who lost weight. Thus, weight change may prove to be a marker for subtypes of depressive disorder.

PSYCHIATRIC MEDICATIONS

Long-term use of steroid medications is associated with significant weight gain, as is the use of several psychiatric agents. Patients treated for major depressive disorder, psychotic disturbances, and bipolar I disorder typically gain 3 to 10 kg, with even larger gains with long-term care. Table 25.3-5 shows the antidepressant and antipsychotic medications that are most and least likely to produce significant weight gain.

Tendency to Increase Appetite and Body Weight		
Greatest	**Intermediate**	**Least**
Antidepressant Drugs		
Amitriptyline (Elavil)	Imipramine (Tofranil) Trimipramine (Surmontil) Nortriptyline (Pamelor) Doxepin (Adapin, Sinequan) Phenelzine (Nardil)	Amoxapine (Asendin) Desipramine (Norpramin) Trazodone (Desyrel) Tranylcypromine (Parnate) Fluoxetine (Prozac)* Bupropion (Wellbutrin)*
Antipsychotic Drugs		
Chlorpromazine (Thorazine) Thioridazine (Mellaril) Mesoridazine (Serentil)	Trifluoperazine (Stellazine) Perphenazine (Trilafon) Thiothixene (Navane)	Haloperidol (Haldol) Loxapine (Loxitane) Molindone (Moban)*

*May decrease appetite and facilitate weight loss.
Adapted from Bernstein IL: Management of psychotropic drug-induced obesity. In Obesity, P Björntorp, BN Brodoff, editors. Lippincott, Philadelphia, 1992.

Table 25.3-5 Psychiatric Medications and Changes in Body Weight

The tricyclic and heterocyclic antidepressant drugs are associated with increased appetite and cravings for carbohydrates and sweets. In addition, anticholinergic effects produce dry mouth and increased thirst, which may lead to the consumption of high-caloric beverages and candy to freshen the mouth. By contrast, fluoxetine, sertraline (Zoloft), and other selective serotonin reuptake inhibitors are usually associated with reduced appetite and food cravings and may initially induce small weight losses (0.5 kg a week) in depressed and obese individuals. Similarly, bupropion (Wellbutrin), a weak reuptake inhibitor of dopamine reuptake, appears to be associated with modest weight loss rather than gain.

Of the antipsychotic agents, chlorpromazine (Thorazine) is generally associated with the greatest weight gain and molindone (Moban) with the least (and sometimes even with weight loss). Weight gain on antipsychotic agents appears to result from their antiserotonergic and anticholinergic effects; patients should be encouraged to drink water rather than high-calorie beverages when trying to reduce thirst and dry mouth.

In many cases patients with bipolar I disorder can be maintained atisfactorily on carbamazepine (Tegretol) or valproic acid (Depakene) without the significant weight gain typically associated with lithium (Eskalith). Like those treated by antidepressant and antipsychotic agents, these individuals should be provided dietary counseling to ensure that they select foods and beverages that are low in fat and sugar.

DIAGNOSIS

The diagnosis of obesity, if done in a sophisticated way, would involve assessment of body fat. This is rarely practical, so use of height and weight to calculate body mass index is recommended (Fig. 25.3-2).

Binge eating disorder may accompany and complicate obesity. Binge eaters are more likely than obese non–binge-eaters to suffer significant psychiatric complications that may require pharmacological or other treatment to regulate mood and appetite. Binge eating disorder, in turn, must be differentiated from bulimia nervosa. The management of binge eating disorder is discussed below.

In most cases of obesity it is not possible to identify its precise etiology, given the multitude of possible causes and their interactions. Instances of secondary obesity (described in Table 25.3-4) are rare but should not be overlooked. In Cushing's disease, for example, a distinct weight pattern is seen (Fig. 25.3-3).

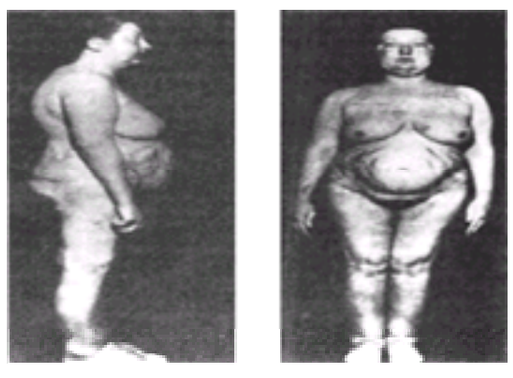

FIGURE 25.3-3 Cushing's syndrome. Obesity and round face are evident. Hirsutism and abdominal striae are noted. (Reprinted with permission from Spillane JD, Spillane JA: *An Atlas of Clinical Neurology*, ed 3. Oxford University Press, New York, 1982.)

TREATMENT

Many treatments are available for obesity and different approaches are likely to be effective for different people. The most thoroughly evaluated approaches, behavior therapy, pharmacotherapy, very–low-calorie diets, and surgery, represent only a fraction of the means people use to lose weight. Self-help books, commercial and self-help programs, input from a variety of health professionals, and just doing it entirely on one's own have not been evaluated to the extent that firm conclusions can be drawn.

It is generally accepted that life-style change is necessary no matter what approach to weight control is used. Behavior therapy is the means for altering life-style and hence could be considered sufficient by itself for some individuals or a necessary adjunct for people needing additional assistance.

Preparing for Treatment More thought has been given in recent years to whether treatment is more important or more likely to be successful in some obese individuals than others. Having upper body fat distribution confers greater risk than does lower body obesity, but this is not widely used as a criteria for aggressive intervention. The field has not yet accepted a scheme by which treatment might be targeted to certain individuals. This is an important area of inquiry because some individuals will have more medical and psychological complications from their weight and some will have a better prognosis for treatment.

Matching Individuals to Treatments Some experts in the obesity field believe that a search for a single best treatment will not be fruitful and that finding the best fit between individuals and different treatments is a more productive way to proceed. This is based on the fact that most treatments are effective for at least some people, and obesity has multiple etiologies.

There has been some speculation of how patients might be matched, but there is little in the way of validated criteria to undertake a matching process. An example of this speculation is that Overeaters Anonymous might be especially helpful for binge eaters, given its focus on “compulsive overeating.”

In the absence of criteria for matching, several steps are still possible. First, speaking with patients about weight loss resources in the community is useful. It can be helpful to develop a list of local weight control programs (commercial, self-help, professionally directed), relevant facilities such as health clubs, and professionals who may be valuable (dietitians, exercise specialists, etc.). Patients can be told that a variety of approaches are available and that they should think about what might be most effective for them as individuals. Second, patients can be encouraged to combine or sequence programs in creative ways. Seeing a professional once each week and attending a self-help group on alternate days might be a helpful combination for some individuals. Finally, patients can think of various resources as means of maintaining weight loss. Re-entering a program if weight begins to increase might be used as crisis intervention. Periodic attendance at a program might be viewed as a booster to strengthen defenses.

Treatment Provider It is important for practitioners to clarify the role that they will play in the treatment of a patient's obesity. Psychiatrists and other mental health professionals certainly can assume primary responsibility for the patient's weight loss efforts and will probably achieve the most successful weight loss outcome by using a structured treatment manual, in some cases accompanied by pharmacotherapy.

Psychiatrists, however, may choose not to direct the patient's weight loss effort when obesity is not the primary presenting problem. Treatment time devoted to weight loss reduces time available for other problems and may confuse the therapeutic agenda. Specifically, the treatment of obesity is usually goal-directed and includes homework assignments that the patient must complete and turn in to the therapist. This approach contrasts with more open-ended models of traditional psychotherapy. In addition, patients may attempt to avoid the discussion of difficult topics in psychotherapy by indicating that they want to “work on their weight.” Alternatively, when it is time to discuss weight-related tasks (such as keeping a diet diary) that were not completed, patients often object that they want to discuss the “real things that are troubling me.”

Obesity is not a principal complaint in the majority of cases referred to psychiatrists or other mental health professionals, perhaps because obesity is not recognized as a psychiatric disorder except when it is accompanied by an eating disorder. In cases in which weight loss is raised as a secondary goal in ongoing therapy, the most favorable outcome is usually obtained by referring the patient to a weight loss program or specialist. This allows the mental health professional to discuss the significance of weight-related issues and to support the patient's efforts to lose weight without losing sight of the original presenting problem.

Setting Weight Loss Goals Individuals begin with an image of what they will accomplish. It may be a picture of how the new body will look, or may be a goal weight, and is shaped by multiple factors, including media portrayals of beauty, messages from the general culture about how people can and should look, family, health ideals (e.g., height-weight charts), and personal experience. What is common is the highly unrealistic nature of these ideas.

One study examined goals in 60 women beginning a weight loss program at an average weight of 218 pounds. The women were asked to list weights at which they would be successful, including a dream weight, happy weight (less than the dream but still satisfying), acceptable weight (not satisfying, but reasonable), and disappoint weight (better than nothing, but clearly disappointing). The weight losses needed to attain these weights, expressed as percent of initial body weight, are shown in Figure 25.3-4.

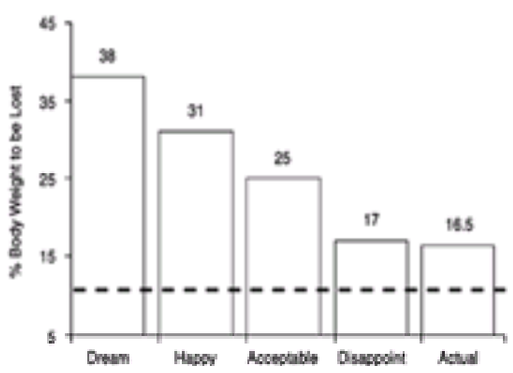

FIGURE 25.3-4 Amount of weight patients said they needed to lose to attain a dream weight, weight at which they would be happy, acceptable weight, and weight at which they would feel disappointed, and finally, their actual weight loss. The dashed line represents a weight loss of 10 percent of initial body weight. This is the approximate magnitude of weight loss produced by most approaches to obesity except for larger losses produced by very aggressive diets or surgery. The data show that actual weight loss in a program with quite good results is far below what patients feel they need to lose to feel at all satisfied with their effort. (Data from Foster GD, Wadden TA, Vogt RA, Brewer G: What is a reasonable weight loss? Patients' expectations and evaluations of obesity treatment outcomes. J Consult Clin Psychol *65:*79, 1997.)

The average weight loss needed to attain the dream weight was 38 percent of initial body weight. The weight at which the subjects felt they would be disappointed was a 17 percent loss, far above the 10 percent loss in body weight that most programs produce. In this study the average loss was 16.5 percent, far above average for most studies but still below the disappoint weight. It is clear from these data that unrealistic goals and expectations are the norm.

Unrealistic goals create a discrepancy between expected and actual performance. This can lead to mood changes, cognitive distortions, and a series of mental events that create disappointment, frustration, self-blame, and poor self-esteem. People feel that failure is inevitable and dismiss as inconsequential weight losses that their physicians might feel are beneficial.

Helping people develop reasonable expectations for their rate of weight loss, the ease with which weight loss will occur, and the ultimate magnitude of weight loss is crucial. However, this requires more than a single admonition to accept modest losses. A 200-pound woman making a 10 percent loss in body weight will weigh 180 pounds. It will not be easy for her to accept this weight, and in order to do so she must uncouple weight from self-esteem and challenge a number of fundamental body image assumptions. Current thinking is that a 10 percent loss in initial body weight is likely to be beneficial from a health perspective; some experts feel the number may be as low as 5 percent. Certainly greater weight loss is desirable, but not always possible, so thinking of moving a person toward a healthier weight

rather than an ideal weight offers a greater chance of having goals attained. This brings goals more in line with what modern treatments can deliver.

Behavior Therapy Behavior therapy became a presence in the field in the late 1960s and 1970s and has become the cornerstone of a great many treatment approaches for obesity. This approach, as the foundation of many commercial programs such as Weight Watchers and Jenny Craig, is used in some form in most obesity centers, and is the primary feature of lifestyle change that is thought to be essential even when treatments such as pharmacotherapy are used. It is based in theory, provides a conceptual scheme for encouraging behavior change, offers a multitude of techniques for initiating change, is the most thoroughly evaluated approach, and is designed to promote long-term changes in important aspects of lifestyle (diet, physical activity, and attitudes).

Behavioral treatment is designed to help patients make gradual habit changes to produce slow but steady weight loss. This approach has several distinguishing features. Its hallmark is in being goal-oriented. The objectives of treatment are clearly specified in weekly homework assignments that patients complete and discuss with the therapist the following week. Patients might be asked to limit eating to two rooms of the home, to walk after dinner three times weekly, or to record self-critical statements following overeating episodes.

Treatment is designed to change behavior and to help the individual establish new skills. It differs from traditional psychodynamic therapy, which would focus on unconscious motives thought to control overeating. Treatment is also process-oriented; it offers patients a method of analyzing eating and exercise habits to determine the behaviors to be modified and the best strategies for changing them. When the patient fails to acquire new behaviors, treatment examines emotional conflicts in a manner similar to more traditional psychotherapy. This step is taken, however, only after more direct methods of behavior change have proven ineffective.

The key elements to behavioral approaches are structured lifestyle change, methods to increase physical activity, cognitive techniques for attitude change, strategies for increasing social support, and methods for dietary change.

Record Keeping Patients keep extensive records of food intake throughout treatment. In the initial weeks they record the types and amounts of foods eaten and their caloric value. They are asked to reduce consumption of high-fat and high-calorie foods as a means of decreasing caloric intake by approximately 500 kcal a day. Record keeping is increased over time to include information about times, places, and feelings associated with eating, along with exercise habits. In the latter stages of treatment the records help identify high-risk situations that are associated with dietary lapses and the thoughts and feelings that occur in response to these situations.

Physical Activity Patients increase their life-style activity by walking more, using stairs rather than escalators, and reducing their dependence on energy-saving devices such as extension phones and remote control devices. Most patients also adopt a structured exercise program, such as walking or swimming, but not until they understand that the activity need not be heroic or punishing. Because there are psychological advantages to exercise (effects on body image, self-esteem, and mood), patients are encouraged to be active as often as possible, in ways that can be sustained, and to be certain that every activity they engage in, no matter how small, “counts” as exercise in their minds.

Cognitive Change People often have attitudes that are significant barriers to weight control. Body image disparagement is an example, but others include issues with self-concept, the tendency for small slips to become grounds for relapse, rigid rules about what constitutes adherence to or breaking a diet, and highly unrealistic goals about how rapidly weight loss will occur and how much weight will ultimately be lost. Behavioral approaches, now more often called *cognitive-behavioral treatments*, provide a systematic way for identifying problem attitudes and for replacing them with thoughts that are more constructive.

Relapse Prevention Based in large part on the pioneering work on relapse by G. Alan Marlatt and Judith Gordon, behavioral treatments have a strong focus on the maintenance of weight loss. Relapse-prevention strategies involve identifying high-risk situations (e.g., negative moods, interpersonal difficulties), planning to prevent initial mistakes (lapses), and building cognitive and behavioral methods for preventing lapses from becoming relapse.

Treatment Mechanics Treatment appears to be most cost effective when delivered in small groups (of 10 to 12 persons) in which participants discuss weekly homework assignments and related issues. Individual treatment may also result in adequate weight loss although usually not the emotional support associated with group care. The most favorable outcome is likely to be facilitated in either case by the use of a structured treatment manual, such as *The LEARN Program for Weight Control*.

Patients treated by group behavioral therapy lose an average of 8 kg in 15 to 20 weeks, approximating a weekly loss of 0.5 kg (1 lb) a week. Longer therapy is associated with greater losses but the average rarely exceeds 15 kg, even in treatment of 40 to 52 weeks. Patients regain approximately one-third of their weight loss in the year following therapy, although better maintenance has been reported in some studies, with increasing regain over time.

Maintenance of Weight Loss Weight regain can be minimized and even prevented. Persons in one research study, who following weight reduction attended every-other week meetings for 1 year maintained their full weight loss during this time whereas those without the benefit of such care regained almost half the weight they had lost. Regular therapist-patient contact by telephone or mail also helps to keep the weight off.

Physical activity is a key issue in considering maintenance. Correlational studies as well as controlled trials show consistently that persons who engage in regular physical activity are the ones most likely to maintain a reduced weight. The exercise need not be punishing as revealed by favorable findings for lifestyle changes such as walking and using stairs. The mechanisms by which exercise controls weight are not well understood. Physical activity clearly burns calories but may also have favorable effects on mood, metabolic rate, and appetite.

That the issue of maintenance is complicated is an understatement. Most overweight people have lost weight, but many have regained it. There are a number of methods for producing weight loss, but long-term maintenance of the loss is far too rare. The future looks hopeful, however, because fundamental changes are occurring in the conceptualization of obesity and its treatment. The view of obesity as a chronic disease that requires long-term treatment should lead to new approaches to therapy (e.g., sustained pharmacotherapy), may open the door to better coverage of the costs of obesity treatment, and should enhance attention among researchers to the issues that govern maintenance and methods for improving it.

Very-Low-Calorie Diets Very-low-calorie diets, which provide 400 to 800 kcal daily, are not as popular today as they were in the 1980s, but remain an important treatment option for patients with a BMI >30 kg/$m^{2.}$ The popularity of liquid diets, in particular, lay in their ease of preparation and to their induction of large, rapid weight losses—as great as 20 kg in 12 weeks. Findings, however, that patients regained 35 to 50 percent of their weight loss in the year following treatment decreased the attractiveness of this approach, as did concerns that the diets were associated with an increased risk of gallstones and binge eating.

The most appropriate candidates for very-low-calorie diets are individuals who need to lose weight quickly in order to undergo surgery, as well as those who are unable to lose weight on a conventional diet. Even with these individuals, severe caloric restriction is not necessary. Recent studies have shown excellent weight losses (i.e., 15 kg) with diets that provide 900 to 1000 kcal a day and that include a daily meal of conventional foods, usually in the form of a frozen-food entree. These dietary changes reduce patients' anxiety and problem eating during the refeeding period, during which patients terminate the liquid diet and resume consumption of conventional foods. As compared with traditional very-low-calorie diets low-calorie mixed diets (i.e., liquid diet plus conventional foods) also require less intensive medical monitoring, which can reduce the cost of treatment substantially.

The best long-term results are likely to be obtained with very-low-calorie diets (and low-calorie diets) when they are combined with a comprehensive program of life-style modification that focuses, in particular, on increasing physical activity. Promising results have also been reported by using anorectic agents, such as sibutramine (Meridia), to facilitate maintenance of weight loss following discontinuation of the very-low-calorie diet. Patients interested in this approach should be referred to multidisciplinary programs that specialize in such care.

Pharmacotherapy Weight loss medications will play an increasingly important role in the treatment of obesity in years to come, based on the discovery of new agents growing out of research on the regulation of body weight. The identification of safe and effective medications will allow practitioners to use long-term pharmacotherapy to manage obesity (and its health complications) in the same manner that medications are used to control long-term conditions including depressive disorder, diabetes, and hypertension.

The pharmacological treatment of obesity suffered a setback in September, 1997 when fenfluramine and dexfenfluramine, both serotonin releasers and reuptake inhibitors, were withdrawn from the market because of concerns that they were associated with valvular heart disease. The Food and Drug Administration (FDA)

received reports that approximately 30 percent of individuals who used these medications for 6 months or more experienced aortic regurgitation of mild or greater severity and mitral regurgitation of moderate or greater severity. These findings abruptly halted use of the popular fenfluramine-phentermine combination ("Fen-Phen") that had been tried by millions of Americans, following demonstration that it reduced weight by approximately 15 percent in only 34 weeks.

Current weight-loss medications can be divided into those that (1) increase energy expenditure, (2) reduce food intake, or (3) affect nutrient absorption. Historically, drugs in the second class have been prescribed most widely and consist primarily of noradrenergic agents (including diethylpropion [Tenuate], mazindol [Sanorex], and phentermine [Ionamin, Eastin]), which were approved by the FDA for only short-term use (i.e., £3 months). This restriction severely limits their usefulness; weight is likely to be regained rapidly when medication is terminated.

Sibutramine is a B-phenylethylamine that inhibits the reuptake of serotonin and norepinephrine (and dopamine to a very limited extent). It was approved by the FDA in 1997 for weight loss and the maintenance of weight loss (i.e., long-term use) in patients with a BMI £30 kg/m ^{2} (or a BMI £27 kg/m^2 in the presence of comorbidities). Clinical trials showed that sibutramine (10 to 15 mg a day) in combination with a 1200 to 1500 kcal-a-day diet, reduced initial weight by approximately to 9 percent in the first 6 months and maintained this loss at 12 months while patients remained on medication. In contrast to the fenfluramine and dexfenfluramine, sibutramine has not been linked to valvular heart disease. Its apparent safety resembles that of other agents that are serotonin or norepinephrine reuptake inhibitors but not releasers (e.g., selective serotonin reuptake inhibitors). No cases of primary pulmonary hypertension, another complication linked to the fenfluramines, have been reported in the more than 6000 patients treated by sibutramine. More patients, however, must be studied because only 1 case of primary pulmonary hypertension is expected in every 25,000 to 50,000 treated patients.

Sibutramine is contraindicated with the use of monoamine oxidase inhibitors (MAOIs), as well as with centrally acting appetite-suppressant medications. It also should not be used in combination with SSRIs and migraine medications (i.e., sumatriptin [Imitrex] and dihydroergotamine [D.H.E]). Such use could precipitate a *serotonin syndrome*, characterized by excitability, agitation, and even psychosis. Sibutramine also substantially increases blood pressure in some patients and should not be used in those with uncontrolled hypertension.

The other weight-loss medication approved by the FDA (in 1999) for long-term use is orlistat (Xenical), which is a selective gastric and pancreatic lipase inhibitor that reduces the absorption of dietary fat (which is then excreted in stool). In clinical trials orlistat (120 mg, three times a day), in combination with a low-calorie diet, induced losses of approximately to 10 percent of initial weight in the first 6 months, which were generally well maintained for periods up to 24 months. Because of its peripheral mechanism of action, orlistat is generally free of the central nervous system effects (i.e., increased pulse, dry mouth, insomnia, etc.) that are associated with most weight-loss medications. Thus, orlistat may be preferable to sibutramine in managing obese patients with hypertension or other cardiovascular complications, as well as patients who are treated by serotonergic medications for depression or migraine headache. The principal adverse effects of orlistat are gastrointestinal; patients must consume 30 percent or fewer calories from fat in order to prevent adverse events that include oily stool, flatulence with discharge, and fecal urgency.

The practitioner should emphasize three points in treating patients with weight-loss medications. First, the best results are likely to be obtained if medication is combined with the patient's own efforts to change eating and activity habits. Patients should not assume that they can eat whatever they wish simply because they are taking medication. Second, patients should be prepared for weight loss to slow and eventually stop after the first 4 to 6 months. Most will lose about 10 percent of their initial weight, which will leave them far short of ideal weight. Patients will need support in accepting that a 10 percent loss is a therapeutic success. Third, serious consideration should be given to using weight-loss medications for the long term (i.e., indefinitely). Pharmacotherapy effectively maintains weight loss and improvements in health complications. These benefits are likely to be lost if medication is discontinued, just as blood pressure is likely to increase if antihypertensive medications are withdrawn.

Surgery Surgery is an option for persons 100 percent or more overweight (i.e., BMI >40 kg/m ^{2}) who have failed to lose weight using more conservative measures. There are two principal procedures. With *vertical-banded gastroplasty*, a staple line is used just below the gastroesophogeal junction to create a small vertical pouch (15–30 mL) with a narrow opening (10 mm) to the remaining stomach. The intestines are not altered in any fashion. The pouch drastically reduces the amount of food that can be eaten at a given meal, thus inducing weight loss. Patients vomit frequently during the first few months until they learn to reduce their portion sizes. Weight losses of 50 kg are common in the first year, although patients can circumvent the operation by consuming high-calorie liquids or easily dissolvable foods such as chocolate or potato chips.

In the *gastric bypass*, a staple line is used just below the gastroesophogeal junction to create a horizontal pouch, which is attached to a loop of the small bowel. Thus, nutrients bypass the remainder of the stomach and the duodenum. The procedure is frequently associated with diarrhea (which can be managed by dietary modification and medication) and with *dumping*—a syndrome characterized by palpitations, weakness, and possible diaphoresis resulting from excessive intake of carbohydrate.

Patients typically reduce their pretreatment weight by 25 to 35 percent in the first 1 to 2 years following surgery. Gastric bypass is associated with greater weight loss than is gastroplasty, although patients treated by both procedures frequently remain as much as 50 percent over their desired weight because of their cell hyperplasia. Both interventions are associated with generally good maintenance of weight loss 5 years postsurgery. Moreover, there is a marked reduction in both physical and psychological complications of obesity. The most striking indication of success is the improved life expectancy of operated patients, which approaches actuarial standards of the general population.

It is imperative that patients who are candidates for surgery be referred to centers that specialize in the procedures and perform a large series annually. Under such conditions, operative mortality is as low as 0.5 percent. It may increase, however, to 1 to 5 percent at centers that perform only an occasional gastric procedure.

Role of the Psychiatrist Patients with chronic psychotic illness are inappropriate for surgery, as may be those with a history of alcohol or other substance abuse. Psychiatrists can play an important role in screening such individuals and in identifying potential psychosocial complications of marked weight loss. Although surgery is usually accompanied by improvements in mood, self-esteem, and body image, patients should anticipate changes in their personal relationships as family and friends adjust to the patient's reduced weight and more active life-style. Many patients, for example, leave unsatisfactory marriages following surgery—an action that is potentially positive but also stressful. Finally, some reports suggest that persons with personality disorders (particularly borderline personality) decompensate following surgery, although studies have failed to find that psychopathology consistently predicts weight loss or psychological response to surgery.

Management of Binge Eating Binge eating was treated initially by researchers working in the eating disorders field. Based on a classic study by Christopher Fairburn with bulimia nervosa showing equivalent results using interpersonal psychotherapy and cognitive-behavioral therapy, these two treatments were applied in randomized research to binge eating and once again showing equivalent results. This led some in the field to believe that binge eating was a disorder distinct from obesity and that eating disorders treatment for the binge eating was necessary before the issue of weight loss could be addressed.

Some recent preliminary research suggests that binge eating responds as well to traditional cognitive behavioral treatment for obesity as it does to eating disorders treatment. Furthermore, patients receiving weight loss treatment lose weight whereas obese binge eaters receiving eating disorders treatment do not—this is a robust finding. This would suggest that treating binge eating as an eating disorder may help to remedy the disorder, but will not influence weight. Hence, priorities of treatment (weight or the binge eating) must be defined before intervention should proceed.

Comprehensive, Compassionate Approach A truly comprehensive approach for managing obesity would require facilities (e.g., metabolic measurement rooms) and personnel (e.g., dietitians and exercise physiologists) that are rarely available. High-quality help can still be provided by use of readily available resources (e.g., treatment manuals) and the judicious use of referral. Deciding on the need for psychotherapy or weight management alone is an important task, and then deciding with the patient which combination of resources for weight control would be most suitable is the next logical step.

Overweight individuals too seldom receive compassionate treatment. The suffering experienced by such people can be considerable. The blame assessed by society is often internalized as self-blame, and it often comes to pass that people hate their bodies and hate themselves. To help reverse this, health professionals must be sensitive to the suffering, resist the cultural bias about assigning blame, recognize that the patient faces a long-term illness, and be persistent and understanding in providing help. When this occurs, there can be striking effects on the health and well-being of people who struggle with their weight.

SUGGESTED CROSS-REFERENCES

Eating disorders, including binge eating disorder, are discussed in Chapter 20. Learning theory is discussed in Section 3.3, behavior therapy in Section 30.2, and

cognitive therapy in Section 30.6. Group psychotherapy is discussed in Section 30.4.

SECTION REFERENCES

Allison DB, editor: *Handbook of Assessment Methods for Eating Disorders and Weight-Related Problems: Measures, Theory, and Research*. Sage, Thousand Oaks, CA, 1995.

*Andersen RE, Wadden TA, Bartlett SJ, Zemel B, Verde TJ, Franckowiak SC: Effects of lifestype activity versus structured aerobic exercise in obese women: A randomized trial. JAMA *281:*335, 1999.

Bar-Or O, Foreyt J, Bouchard C, Brownell KD, Dietz WH, Ravussin E, Salbe AD, Schwenger S, St Jeor S, Torun B: Physical activity, genetic, and nutritional considerations in childhood weight management. Med Sci Sports Exerc *30:*2, 1998.

Bjorntorp P, Brodoff BN, editors: *Obesity*. Lippincott, Philadelphia, 1992.

Borecki IB, Blangero J, Rice T, Perusse L, Bouchard C, Rao DC: Evidence for at least two major loci influencing human fatness. Am J Hum Genet *63:*831, 1998.

Bray GA: Obesity: A time bomb to be defused. Lancet *352:*160, 1998.

Brownell KD: *The LEARN Program for Weight Control, ed 7*. American Health Publishing Company, Dallas, 1997.

*Brownell KD, Fairburn CG, editors: *Eating Disorders and Obesity: A Comprehensive Handbook*. Guilford, New York, 1995.

Campfield LA, Smith FJ, Burn P: The ob protein (leptin) pathway—A link between adipose tissue mass and central neural networks. Horm Metab Res *28:*619, 1996.

Cash TF: *What Do You See When You Look In The Mirror? Helping Yourself to a Positive Body Image*. Bantam, New York, 1995.

*Ditschuneit HH, Flechtner-Mors M, Johnson TD, Adler G: Metabolic and weight-loss effects of a long-term dietary intervention in obese patients. Am J Clin Nutr *69:*198, 1999.

*Dunn AL, Marcus BH, Kampert JB, Garcia ME, Kohl HW III, Blair SN: Comparison of lifestyle and structured interventions to increase physical activity and cardiorespiratory fitness: A randomized trial. JAMA *281:*327, 1999.

Foster GD, Wadden TA, Vogt RA, Brewer G: What is a reasonable weight loss? Patients' expectations and evaluations of obesity treatment outcomes. J Consult Clin Psychol *65:*79, 1997.

Friedman MA, Brownell KD: Psychological correlates of obesity: Moving to the next research generation. Psychol Bulletin *117:*3, 1995.

*Institute of Medicine: *Weighing the Options: Criteria for Evaluating Weight Management Programs*. U.S. Government Printing Office, Washington, DC, 1995.

Klem ML, Wing RR, McGuire MT, Seagle HM, Hill JO: A descriptive study of individuals successful at long-term maintenance of substantial weight loss. Am J Clin Nutr *66:*239, 1997.

*Mantzorus CS: The role of leptin in human obesity and disease. Ann Intern Med *130:*671, 1999.

Perri MG, Nezu AM, Viegener BJ: *Improving the Long-term Management of Obesity*. Wiley, New York, 1992.

*Pi-Sunyer FX. Energy balance: Role of genetics and activity. Ann NY Acad Sci *819:*29, 1997.

Rosmond R, Dallman MF, Bjorntorp P: Stress-related cortisol secretion in men: Relationships with abdominal obesity and endocrine, metabolic and hemodynamic abnormalities. J Clin Endocrinol Metab *83:*1853, 1998.

Stunkard AJ, Wadden TA, editors: *Obesity: Theory and Therapy*. Lippincott-Raven, Philadelphia, 1993.

Wadden TA, Berkowitz RI, Vogt RA, Steen SA, Stunkard AJ, Foster GD: Lifestyle modification in the pharmacologic treatment of obesity: A pilot investigation of a potential primary care approach. Obesity Res *5:*218, 1997.

Wadden TA, Vogt RA, Andersen RE, Bartlett SJ, Foster GD, Kuehnel RH, Wilk J, Weinstock R, Buckenmeyer P, Berkowitz RI, Steen SN: Exercise in the treatment of obesity: Effects of four interventions on body composition, resting energy expenditure, appetite, and mood. J Consult Clin Psychol *65:*269, 1997.

Wadden TA, Vogt RA, Foster GD, Anderson DA: Exercise and the maintenance of weight loss: 1-year follow-up of a controlled clinical trial. J Consult Clin Psychol *66:*429, 1998.

25.4 CARDIOVASCULAR DISORDERS

PETER A. SHAPIRO, M.D.

Cardiovascular disease is the leading cause of death in the United States and in most of the industrialized world. About one-third of all adults over age 35 will ultimately die of cardiovascular disease, most often of complications of atherosclerotic coronary artery disease. In the United States, the incidence of coronary artery disease is over 5,000,000 cases per year, and 500,000 persons per year have a myocardial infarction; 20 percent of survivors of the acute infarction will die within a year. Some 400,000 persons develop congestive heart failure each year, and heart failure is cited as the principal cause in 40,000 deaths and as a contributing factor in 250,000 deaths annually. Although rapid advances in cardiovascular therapeutics have reduced mortality for patients admitted to the hospital with acute myocardial infarction, about one-third of cardiac deaths occur within the first hour of the patient's first heart attack, before any opportunity for medical intervention exists. This fact underscores the importance of primary prevention, efforts to reduce the risk of coronary artery disease before its clinical presentation. Behavioral risk factors for coronary heart disease, such as smoking, overeating, physical inactivity, and poor compliance with management of diabetes, hypercholesterolemia, and hypertension, are the main targets for primary prevention interventions. Achieving behavioral change with respect to these risk factors poses a considerable challenge for medicine and psychiatry. In addition, certain psychiatric disorders, acute mental stress, and psychological traits are risk factors for development and clinical progression of cardiovascular disease.

Psychiatric disorders frequently occur as complications or comorbid conditions in individuals with cardiovascular disease. Depression, anxiety, delirium, and cognitive disorders are especially prevalent problems. Surveys of ambulatory cardiology patients with documented heart disease indicate a point prevalence of 5 to 10 percent with anxiety disorders (predominantly panic attacks and phobias) and 10 to 15 percent with mood disorders (predominantly depressive episodes and minor depression or dysthymia). Major depressive disorder occurs in 15 to 20 percent of patients following myocardial infarction.

DEPRESSION, ANXIETY, AND OTHER RISK FACTORS

Depression In overwhelming support of long-held popular views, numerous recent investigations strongly support the hypothesis that depression increases the risk of development and progression of coronary artery disease. Over the past 60 years, several studies of institutionalized or treated mentally ill patients suggested an excess of cardiovascular mortality in those with depression compared with the general population, but these studies were potentially confounded by effects of the setting or treatment that defined the population. More recently, numerous large-scale, prospective epidemiological studies of community-dwelling subjects who were not psychiatric patients have yielded converging estimates of increased relative risks of incident myocardial infarction and cardiac related mortality of about 1.6-2.2 to 1 in association with depression. This finding holds even after controlling for smoking, a potent risk factor for cardiovascular morbidity and mortality, which is far more prevalent in those with depression than in the population at large, and the effect of depression persists, even in long-term follow-up over 10 to 20 years. Furthermore, studies of patients with preexisting coronary artery disease also demonstrate a near doubling of risk for adverse coronary disease–related outcomes, including myocardial infarction, revascularization procedures for unstable angina, and death, in association with depression. Symptoms of depression and the diagnosis of major depressive disorder carry a 3.5 to 6.6-fold increased adjusted relative risk of death in 6- and 18-month follow-up of myocardial infarction patients. In these patients the predominant mode of death is sudden cardiac death. Cooccurrence of depression and frequent premature ventricular contractions after myocardial infarction appears to increase mortality risk substantially, suggesting arrhythmia as the mechanism of death. Whether this risk can be reduced by treatment of depression is currently under investigation. One recent study of psychosocial intervention for postmyocardial infarction patients with depression or social isolation used monthly telephone contact, followed by a nurse's home visit to patients who expressed distress, and variable subsequent contacts. Although a preliminary study in men had suggested a beneficial effect of this program on recurrent cardiac event and mortality rates, replication in a cohort of both men and women failed to demonstrate an overall benefit, and women receiving the intervention had a poorer outcome than a control group receiving usual care. Further psychosocial intervention and pharmacotherapy trials are ongoing.

Possible Mechanisms Mechanisms by which depression may increase coronary disease risk are uncertain. Autonomic dysregulation with diminished cardiac vagal modulation occurs in depression and may provide a substrate for increased arrhythmic activity and sudden death. Disordered platelet aggregation leading to increased thrombus formation may also play a role in increasing risk of coronary events in depression. Measures of in vivo platelet activation and aggregation after overnight bed rest and following orthostatic challenge in medication-free, otherwise healthy depressed patients and normal control subjects show that depressed patients exhibit greater procoagulant activity at baseline and greater platelet activation on orthostatic challenge. The findings suggest that increased concentrations of plasma neuroregulators that can induce platelet activation (e.g., epinephrine or serotonin), intrinsic platelet factors, intraplatelet catecholamine or monoamine concentration changes, or a combination of these increase platelet reactivity in depressed patients. Patients with ischemic heart disease and depression have significantly elevated concentrations of circulating platelet factor 4 and b-thromboglobulin, factors associated with platelet activation; patients with ischemic heart disease alone also have elevated levels of these factors, but to a much lesser extent than patients with depression. Treatment with the serotonergic antidepressant paroxetine, but not with the tricyclic antidepressant nortriptyline, was associated with reduced levels of these indexes of platelet activation in one small study.

Anxiety, Depression, and Arrhythmias Malignant ventricular arrhythmias may occur as an “electrical accident” in individuals with no structural heart disease but are more frequent in those with underlying heart disease. Because conduction of myocardial depolarization may be delayed in regions of myocardial tissue damage, regions distal to the affected area may be stimulated to depolarization while adjacent tissue is undergoing repolarization, setting off a premature depolarization of the adjacent tissue that may precipitate ventricular fibrillation. The rate of local repolarization and readiness to initiate the next cardiac cycle is modulated by local innervation of the heart by branches of both the sympathetic and parasympathetic nervous system. In general, sympathetic fibers predominate in the left ventricular myocardium and have the effect of increasing irritability and stimulating depolarizations. Parasympathetic fibers predominate in the innervation of the specialized cardiac conduction system and have the effect of slowing heart rate and reducing automaticity.

Because autonomic cardiac modulation is profoundly sensitive to acute emotional stress such as intense anger, fear, or sadness, it is not surprising that acute emotions can stimulate arrhythmias. Indeed, instances of sudden cardiac death related to sudden emotional distress have been noted throughout history in all cultures. Two studies have demonstrated that in addition to depression, a high level of anxiety symptoms raises the risk of further coronary events in patients following myocardial infarction by two to five times that for nonanxious comparison patients. This effect is independent of the effect of depression. Risk is increased in both the immediate postinfarct period and over 18-month follow-up. Recently completed analyses of prospectively recorded epidemiological data in community samples demonstrate that phobic anxiety and mixed anxiety symptoms pose an increased risk of sudden cardiac death but not of nonfatal myocardial infarction, even after adjusting for other cardiovascular risk factors. High levels of anxiety symptoms are associated with a tripling of risk of sudden cardiac death.

Hostility and Type A Behavior Pattern The relation between a behavior pattern characterized by easily aroused anger, impatience, aggression, competitive striving, and time urgency (Type A) and coronary heart disease dominated studies in psychosomatic cardiology in the 1970s and 1980s. Several large prospective epidemiological studies found the Type A pattern to be associated with a nearly twofold increased risk of incident myocardial infarction and coronary disease–related mortality. In the 1980s, however, studies using a variety of measures of the Type A pattern and a variety of clinical endpoints, in a variety of patient populations, failed to lend convergent validity to the Type A construct as a risk factor for adverse cardiovascular outcomes across a range of clinical populations. For example, Type A pattern is associated with reduced (not increased) mortality in myocardial infarction patients with frequent premature ventricular contractions. Nevertheless, the observation of increased risk of coronary disease incidence and mortality associated with Type A pattern in nonclinical populations still appears valid. Moreover, group therapy to modify Type A behavior pattern was associated with reduced reinfarction and mortality in a 4.5-year study of patients with prior myocardial infarction. Type A behavior modification therapy has also been demonstrated to reduce episodes of silent ischemia on ambulatory electrocardiographic monitoring. A recent meta-analysis of 23 randomized controlled trials evaluated the additional impact of psychosocial treatment on rehabilitation from documented coronary artery disease. Relaxation training, stress management, and group social support were the predominant modalities of psychosocial intervention. Anxiety, depression, biological risk factors, mortality, and recurrent cardiac events were the clinical endpoints studied. These studies included a total of 2024 patients in intervention groups and 1156 control subjects. Psychosocial treatment patients had greater reductions in emotional distress, systolic blood pressure, heart rate, and blood cholesterol level than

comparison subjects. Patients who did not receive psychosocial intervention had 70 percent greater mortality and 84 percent higher cardiac recurrent event rates during the first 2 years of follow-up.

Hostility as a core component of the original Type A concept has received considerable empirical support as a predictor of coronary heart disease outcomes. Low hostility is associated with low coronary disease risk in studies of workplace populations. In addition, hostility is associated with several physiological processes that in turn are associated with coronary disease, such as reduced parasympathetic modulation of heart rate, increased levels of circulating catecholamines, and increased lipid concentrations during interpersonal conflict. Conversely, submissiveness has been found to be protective against coronary disease risk in women. Adrenergic receptor function is downregulated in hostile men, presumably an adaptive response to heightened sympathetic-adrenergic drive and chronic overproduction of catecholamines resulting from chronic and frequent anger. Because anger and hostility frequently coexist with depression, anxiety, and low social support, recent investigations of psychosocial factors associated with coronary heart disease have increasingly invoked the more global descriptors of low social support and poverty.

Vital Exhaustion Vital exhaustion, a state of fatigue associated with decreased interest in work, pessimism, and burnout, has been identified as a risk factor for incident myocardial infarction in European (predominantly Dutch) studies. Because the vital exhaustion concept overlaps with depression, its independent contribution remains uncertain.

Acute Mental Stress Acute mental stress has a significant effect on coronary artery blood flow, which may be of significance in patients with preexisting coronary disease. States of fear, excitement, and especially acute anger reduce blood flow through atherosclerotic coronary segments, provoke coronary spasm, and are associated with abnormal left ventricular wall motion and electrocardiographic evidence of myocardial ischemia. Mental stress–induced ischemia is associated with increased risk of subsequent cardiac events in patients with known coronary disease and may occur even in patients who do not demonstrate evidence of ischemia during exercise stress testing. Recent studies, however, indicate that relaxation training can alter autonomic activation during mental stress, implying a potential therapeutic role for such training in stress-induced ischemia.

Sudden cardiac death triggered by mental stress was demonstrated in a study of deaths in the aftermath of a major California earthquake. Following a surge of sudden deaths on the day of the earthquake, there was a several-day period of reduced incidence of sudden cardiac death, suggesting that only those predisposed to sudden cardiac death by underlying disease were affected by the monumental, acute stressor. Deaths were primarily associated with emotional stress rather than physical exertion. Sudden death without antecedent angina, suggesting cardiac arrhythmia, as well as deaths preceded by chest pain, suggesting acute coronary occlusions, were observed. Other studies have led to estimates that between 20 and 40 percent of sudden cardiac deaths are precipitated by acute emotional stressors.

Valvular Heart Disease and Anxiety Disorders The relationships between valvular heart disease and psychiatric disorder have been a matter of considerable interest over the past several decades. In panic disorder, mitral valve prolapse is detected in 10 to 25 percent of patients studied with echocardiography. However, prolapse also occurs in a substantial portion of the population without panic disorder, and the nature of the relationship remains uncertain. The subjective experience of valve prolapse (fluttering, chest pressure, etc.) may trigger panic sensations; alternatively, the association may be purely coincidental. Obsessive-compulsive disorders, tic disorders, and Tourette's disorder are associated with poststreptococcal immune system–mediated inflammatory responses similar to those leading to glomerulonephritis and rheumatic heart disease, a finding stimulated by consideration of the ticlike quality of movement disorders occurring in rheumatic fever. This association would lead to the expectation of an excess of rheumatic valvular heart disease in patients with disorders in the obsessive-compulsive spectrum.

Smoking The relationship between cigarette smoking and mental illness, particularly depression, has been well documented. Individuals with a lifetime history of depression are far more likely to have smoked than those without a depression history. Moreover, individuals with a history of depression, who have smoked, are far less likely to quit smoking successfully. As the prevalence of smoking trends downward in the United States, this association is strengthened. Inasmuch as smoking is such a powerful risk factor for ischemic heart disease, smoking cessation intervention is a matter of considerable interest for the psychiatrist, cardiologist, and primary physician. Buproprion (Wellbutrin) has been demonstrated to improve “quit rate” in smokers as a supplement to nicotine replacement and behavior therapy. The duration of pharmacotherapy is typically 2 to 3 months. Nortriptyline has also been shown to improve smoking cessation rates. Success in smoking cessation appears to be augmented by longer-term behavioral treatment.

PSYCHIATRIC PROBLEMS IN PATIENTS WITH CARDIAC DISEASE

Arrhythmias Ventricular arrhythmias may be asymptomatic or may cause palpitations, lightheadedness, dizziness, syncope, or sudden cardiac death. Patients who experience life-threatening rhythm disturbances are prone to secondary adjustment, mood, and anxiety disorders. The cardiovascular symptoms may lead to profound disruption of social roles and capacity for autonomous functioning. Even patients without symptoms may be counseled to avoid activities such as driving, which may be hazardous in the event of an arrhythmic event. Psychodynamically, because of the recurring, unpredictable, and sudden quality of the course of illness, issues of dependence on others and loss of control are especially prominent, as well as anxiety about death itself.

Treatment of cardiac rhythm disorders can engender psychiatric complications as well. Many antiarrhythmic agents have psychiatric side effects, including delirium, hallucinations, paranoid ideation, and mood disturbance. Lidocaine (Xylocaine) and procainamide (Pronestyl) are especially likely to have these effects. Implanted automatic pacemaker-defibrillators have supplanted chronic antiarrhythmic medical therapy in a substantial portion of patients who have experienced ventricular tachycardia–ventricular fibrillation. Regrettably, cardioversion by one of these devices may occur in a conscious patient, who experiences the sensation of one or more powerful shocks or blows to the chest. Although recent surveys have shown that most patients with automatic pacemaker-defibrilators seem satisfied with their devices, approximately 10 percent develop clinically significant anxious or depressed mood symptoms, including symptoms of posttraumatic stress, associated with defibillator discharges. Avoidance behaviors and anxiety in the setting of discharge-associated stimuli reduce quality of life for patients who have experienced shocks during such ordinary activities as sexual relations, showering, and low-intensity exercise. Risk factors for psychiatric complications in patients with automatic defibrillator-pacemakers include discharges early after implantation, repetitive discharges, and a high cumulative number of discharges.

Valve Replacement Valve replacement was among the first of the open heart surgeries to be developed, and pioneering work in psychiatric aspects of cardiovascular surgery demonstrated a high prevalence of delirium in early postcardiotomy patients. This delirium can be attributed to toxic or metabolic processes in many cases. Prolonged exposure to the intensive care unit (ICU) environment with sleep deprivation, sensory stimulation, and simultaneous monotony led to the phenomenon of delirium following a lucid interval (so-called ICU psychosis), but shortened ICU stays and better attention to maintenance of day-night cycles, noise reduction, and the provision of privacy have greatly reduced the extent of this phenomenon.

Delirium Delirium is a common problem in severely ill cardiac patients. Three main categories of patients are at risk: patients with severe congestive heart failure, patients receiving antiarrhythmic agents for tachyarrhythmias early after myocardial infarction or cardiac surgery, and patients following cardiac surgery. In congestive heart failure patients, delirium commonly results from hypoxia, hyponatremia, hyperammonemia, or azotemia, as pulmonary congestion and poor end-organ perfusion progress. Patients receiving lidocaine and procainamide may appear psychotic or delirious, even at nominally therapeutic blood levels. The clinical picture in postoperative patients may include all of these elements; in addition, cerebrovascular incidents during surgery, infection, sedatives, and narcotics may contribute to delirium. Management relies on correcting the underlying abnormality while treating psychosis or agitation with antipsychotic agents. Other sedatives should generally be avoided, although the agitated patient may benefit from concurrent administration of lorazepam (Ativan) with haloperidol (Haldol). Intravenous haloperidol can be administered frequently, especially if hemodynamic monitoring is in place. Very high dosages, however, may be associated with arrhythmias, including torsade de pointes. Prolonged cardiac conduction can occur with thioridazine (Mellaril) and chlorpromazine (Thorazine), especially in patients taking type 1A antiarrhythmic agents. b-Adrenergic receptor antagonists (beta-blockers) may cause elevated concentrations of antipsychotic drugs.

Coronary Artery Bypass Graft Surgery Coronary artery bypass graft surgery does not have an advantage over medical management of coronary artery disease with respect to survival, except in cases of left main coronary artery or left anterior descending artery obstruction, but surgery may be more effective than medication in ameliorating angina symptoms and improving exercise tolerance, functional capacity, and overall quality of life. Coronary artery bypass graft surgery is one of the most frequently performed surgeries in the United States. Psychiatric complications after this surgery have been noted since its inception and persist despite improvements in cardiopulmonary bypass and anesthetic and surgical technique. Persistent, subtle, memory and cognitive impairment may occur after this surgery. In a recent study of over 2100 elective coronary artery bypass graft surgical patients at 24 centers, clinical observation yielded an adverse neurological event rate of 6.1 percent. Death, focal neurological impairment (stroke), or stupor occurred in 3.1 percent; these events were predicted by older age, proximal aortic atherosclerosis, and prior history of neurological disease. Nonspecific impairment in intellectual function occurred in 2.6 percent, and seizures in 0.4 percent. Risk factors for these events were older age, systolic hypertension on admission, pulmonary disease, and excessive alcohol use. Identified cases were limited to those whose clinical record recorded the neurological impairment; it is possible that unrecorded impairment occurred with a substantially higher incidence. Adverse cerebral outcomes

substantially increase hospitalization duration, in-hospital mortality, and the need for discharge to institutional supportive care.

Mild-to-moderate depression occurs in approximately one-third of patients following coronary bypass surgery but may remit within weeks to months. Depressive symptoms are present in almost 40 percent of coronary artery bypass graft patients 6 months after surgery, but the presence of depression in the early postoperative period is not correlated with depression at 6-month follow-up. Depression at this later point is associated with increased mortality at 3-year follow-up. Subtle changes in cognition are frequently associated with depression after coronary bypass surgery. The linkage of cognitive and affective disturbances after this surgery may be attributable to small vessel cerebrovascular changes seen in many elderly persons with depression. Such cerebrovascular pathology may be exacerbated by cardiopulmonary bypass. Trials randomizing patients to coronary artery bypass graft surgery or angioplasty have shown near equivalence in quality-of-life measures over the first 3 years after treatment but greater improvement in daily role function after the former.

Congestive Heart Failure The syndrome of congestive heart failure results from numerous causes of heart disease but represents the most common final common pathway toward disability and mortality. Treatment of heart failure patients usually includes digoxin (Lanoxin), diuretics, and angiotensin-converting enzyme inhibitors or angiotensin antagonists. Heart transplantation and left ventricular assist devices are options in refractory cases. Apart from their evaluation as heart transplant candidates, there has been relatively little study of the psychiatric problems of heart failure patients. Clinically, it appears that adjustment disorders, depression, and anxiety are relatively common problems in this population. Treatment of depression in heart failure patients is discussed below. Disorders of alertness and cognition occur in patients with hypotension, hepatic failure, or renal failure secondary to heart failure. Diagnosis of psychiatric disorders in patients with congestive heart failure requires clinical judgment, as symptoms of heart failure may confound the psychiatric presentation. Dyspnea and orthopnea, for example, may resemble panic, and orthopnea in particular may interfere with sleep. Hopelessness, decreased self-esteem, suicidal ideation, and other cognitive aspects of affective illness are not normally present in heart failure patients and may help to make the distinction.

Heart Transplantation and Left Ventricular Assist Devices Heart transplantation is available for about 2500 patients annually in the United States. It provides approximately 75 percent 5-year survival for patients with severe heart failure, who would otherwise have less than 50 percent 2-year survival. Candidates for heart transplantation typically experience a series of adaptive challenges as they proceed through evaluation, waiting, perioperative management, postoperative recuperation, and long-term adaptation to life with a transplant. These stages of adaptation typically elicit anxiety, depression, elation, and working through of grief. Although quality of life improves in most respects for most patients who undergo heart transplantation, myriad difficulties can occur. Common medical problems of transplant survivors include infections, graft rejection, graft coronary artery disease, lymphoproliferative disorders and other malignancies, cerebrovascular accidents, renal insufficiency, hypertension, and adverse effects of immunosuppressive drugs. Return to work is problematic, and other social role functioning may remain limited even after physical recovery from heart transplant surgery. Family adjustments to shifting patterns of chronic illness and recovery may be chaotic. Sexual dysfunction affects a third or more of male heart transplant recipients. Psychiatric difficulties in heart transplant patients include primary and secondary mood disorders, delirium, addiction to analgesic and anxiolytic agents, and anxiety. Serotonergic antidepressants and nortriptyline (Aventyl) are well tolerated in heart transplant recipients with depression.

LEFT VENTRICULAR ASSIST DEVICES Left ventricular assist devices have been approved as a bridge to heart transplantation and are undergoing clinical trials for long-term treatment of congestive heart failure. These devices consist of a conduit that drains the left ventricle of the heart, connected to a mechanical pump implanted in the abdomen, which in turn drains into a conduit connected to the ascending aorta. The pump is powered by an external power source, such as a battery that can be worn on a vest, permitting the patient to ambulate freely. Patients with left ventricular assist devices may recover from the clinical complications of left ventricular functional impairment, with improved renal, hepatic, cerebral, and peripheral circulation leading to improved exercise tolerance, functional capacity, and general well-being. They are subject, however, to complications such as bleeding, infection, anorexia (presumably related to gastroparesis or reduced gastric filling capacity), and stroke. Depression and organic mental syndromes occur frequently. Prior cerebrovascular disease appears to be a risk factor for psychiatric and neuro-psychiatric complications in patients with left ventricular assist devices. Caregiver burden is substantial for family members of assist device patients when they are discharged from the hospital.

PSYCHOSOCIAL EVALUATION IN TRANSPLANTATION An emerging body of evidence has demonstrated that preoperatively assessed psychosocial variables predict not only psychiatric but also medical outcomes after heart transplantation. Personality disorders, substance abuse disorders, dimensional measures of coping skills and social support, and clinically assessed compliance history have been linked to increased risk of poor postoperative compliance, rejection episodes, and increased mortality. Psychiatric assessment of transplant candidates may identify opportunities for intervention prior to transplantation, with subsequent improvement in outcome. Ethical quandaries are inherent in the role of transplant service psychiatrist, as conflicting ethical agendas present themselves in the course of work with potential transplant patients. Only by systematically explicating the issues can the psychiatrist protect the dialogue with the patient that is the basis of further therapeutics.

CARDIOVASCULAR FACTORS IN PSYCHIATRIC TREATMENTS

Tricyclic agents cause orthostatic hypotension and cardiac conduction disturbances. In combination with diuretics, vasodilators, or benzodiazepines, their effect on blood pressure is exaggerated. In toxic levels they may precipitate ventricular arrhythmias, but at therapeutic doses they exhibit type 1A antiarrhythmic properties, similar to those of quinidine (Duraquin). Studies of type 1 antiarrhythmic therapy for patients with myocardial infarction and ventricular ectopy demonstrated that antiarrhythmic therapy was associated with increased mortality. Consequently, tricyclic agents are no longer recommended as first-line drugs in treatment of depression in patients with ischemic heart disease. Cardiovascular effects of serotonergic antidepressants in patients with impaired left ventricular function, including patients with coronary artery disease and prior myocardial infarction, are lowering of heart rate by two to four beats per minute and increased left ventricular ejection fraction. They have no effect on cardiac conduction, blood pressure, or arrhythmias. Compared with nortriptyline treatment, side effects occur less frequently and result in lower rates of treatment discontinuation in depressed heart disease patients treated with selective serotonin reuptake inhibitors. Buproprion in high doses may be associated with increased blood pressure and ventricular ectopic activity. Trazodone (Desyrel) and nefazodone (Serzone) may be associated with ventricular ectopy and orthostatic hypotension. Venlafaxine (Effexor) therapy in higher dosages is associated with sustained elevations of blood pressure (Table 25.4-1).

Agent	Effects
Tricyclic antidepressants	Orthostatic hypotension Delayed cardiac conduction, heart block Type 1A antiarrhythmic effect Ventricular arrhythmias in overdose Tachycardia
Selective serotonin reuptake inhibitors	Bradycardia (rare) Increased ejection fraction
Monoamine oxidase inhibitors	Hypotension Peripheral edema
Bupropion	Hypertension
Nefazodone, trazodone	Hypotension Premature ventricular contractions
Venlafaxine	Hypertension, especially at high dosage

Table 25.4-1 Cardiovascular Effects of Antidepressants in Patients With Heart Disease

In an open-label study, patients treated with sertraline (Zoloft) for depression following myocardial infarction demonstrated improved mood ratings, a low rate of cardiovascular side effects, and a low rate of withdrawal from treatment because of adverse events. There were no significant hemodynamic effects or effects on cardiac conduction or arrhythmias. A small study of interpersonal psychotherapy for depressed patients after myocardial infarction also demonstrated improvement in mood.

Drug Interactions and the Cytochrome System Drug-drug interactions mediated by the hepatic cytochrome P450 system are of theoretical importance, although few of these interactions have been proven to be clinically significant. Some potential interactions are listed on Table 25.4-2. Of clinical note, serotonergic antidepressants may prolong bleeding in occasional patients via an effect on platelets; they do not have a systematic effect on coagulation as measured by the Protime test even in patients treated with warfarin (Coumadin). Although fluoxetine (Prozac) inhibits the cytochrome P450 (CYP) isoenzyme CYP 3A4, which metabolizes cyclosporine, studies have indicated no significant effect on cyclosporine concentrations in heart and renal transplant patients. One study suggests that cyclosporine concentrations may increase during nortriptyline treatment, although nortriptyline is not regarded as an inhibitor of this enzyme system. Clinically significant bradycardia has been reported very rarely in patients treated with serotonergic agents along with b-adrenergic receptor antagonism. Lithium levels may be

raised by the use of diuretics affecting the proximal renal tubule (thiazides) and by angiotensin-converting enzyme inhibitors.

P450 Isoenzyme	Metabolizes	Inhibited by
CYP 2D6	Propranolol Timolol Metoprolol Encainide Mexiletine Flecainide Fluoxetine Paroxetine Tricyclic antidepressants Venlafaxine	Fluoxetine Paroxetine Sertraline (weak)
CYP 1A2	Amitriptyline Imipramine Propranolol	Fluvoxamine
CYP 3A4	Lidocaine Quinidine Nefazodone Sertraline Venlafaxine Alprazolam Diltiazem Nifedipine Verapamil Cyclosporine	Fluoxetine Fluvoxamine Nefazodone Sertraline
CYP 2C19	Citalopram Clomipramine Diazepam Imipramine Propranolol	Fluoxetine Sertraline

Table 25.4-2 Selected P450 System-Related Potential Psychiatric and Cardiovascular Drug Interactions

Electroconvulsive Therapy in Patients With Cardiovascular Disease Electroconvulsive therapy (ECT) is a highly effective treatment for severe depressive and manic episodes. During ECT, heart rate and blood pressure increase substantially, followed by a vagally mediated bradycardia. Patients with ischemic heart disease or ventricular arrhythmias are at increased risk of myocardial infarction or ventricular tachyarrhythmias. Atropine is commonly used to reduce airway secretions during ECT and may reduce the extent of bradycardia, but it may exacerbate the tachyarrhythmia associated with the ECT stimulus. b-Adrenergic receptor antagonists, conversely, may precipitate heart block during ECT. Few studies have systematically controlled concurrent treatment and cardiac diagnoses in assessing effects of ECT in cardiac patients, but available information suggests that patients with cardiac disease are more likely to experience cardiovascular complications during ECT than comparison patients without heart disease. Most complications are transient, however, and do not necessarily preclude use of this modality. In patients with preexisting heart disease ECT-related arrhythmias can be reduced and almost eliminated by pretreatment with b-adrenergic receptor antagonists such as labetalol (Normadyne).

CARDIOVASCULAR PRESENTATIONS OF PSYCHIATRIC DISORDERS: CHEST PAIN, ARRHYTHMIA, PALPITATIONS

Somatization disorder, panic disorder, anxiety, and depression can all present with somatic complaints and represent a substantial issue in ambulatory and emergency cardiology practice. In studies of patients presenting with the chief complaint of palpitations, these diagnoses account for about 30 percent of cases. In this population, psychiatric disorder is associated with more frequent recurrent symptoms, emergency room visits, hypochondriacal concerns, and impairment in activities of daily living.

SUGGESTED CROSS-REFERENCES

Section 14.6, Section 14.7, and Section 14.9 contain further information on depression and its treatment, and Chapter 15 and Chapter 16 provide reviews of panic disorder and somatoform disorders. Chapter 10 provides useful information on neurophysiology, diagnosis, and management of delirium and cognitive disorders. Section 25.3 on obesity and Section 25.9 on stress provide information related to primary and secondary prevention efforts in cardiac disease. The clinical approach to the patient and the general hospital setting in which many of these contacts occur is discussed in Section 25.12. General issues pertaining to the experience of surgery and of serious general medical illness are discussed in Sections 28.1 on surgery and Section 28.5 on death, dying, and bereavement.

SECTION REFERENCES

Barsky AJ, Delamater BA, Clancy SA, Antman EM, Ahern DK: Somatized psychiatric disorder presenting as palpitations. Arch Intern Med *156:*1102, 1996.

Blumenthal JA, Jiang W, Waugh RA, Frid DJ, Morris JJ, Coleman RE, Hanson M, Babyak M, Thyrum ET, Krantz DS, O'Connor C: Mental stress-induced ischemia in the laboratory and ambulatory ischemia during daily life. Association and hemodynamic features. Circulation *92:*2102, 1995.

*Booth-Kewley S, Friedman HS: Psychological predictors of heart disease: A quantitative review. Psychol Bull *101:*343, 1987.

Carney RM, Rich MW, Freedland KE, Saini J, TeVelde A, Simeone C, Clark K: Major depressive disorder predicts cardiac events in patients with coronary artery disease. Psychosom Med *50:*627, 1988.

Chacko RC, Harper RG, Gotto J, Young J: Psychiatric interview and psychometric predictors of cardiac transplant survival. Am J Psychiatry *153:*1607, 1996.

Crow SJ, Collins J, Justic M, Goetz R, Adler S: Psychopathology following cardioverter defibrillator implantation. Psychosomatics *39:*305, 1998.

Dew MA, Kormos RL, Nastala C, Pristas JM, Winowich S, Canning RD, Griffith BP: Psychiatric and psychosocial issues and intervention among ventricular assist device patients. In *Psychosomatic Aspects and Quality of Life in Mechanical Circulation and Heart Transplantation,* W Albert, A Bittner, R Hetzer, editors. Verlag, Steinkopff, Darmstadt, Germany, 1995.

Ford DE, Mead LA, Chang PP, Cooper-Patrick L, Wang N, Klag MJ: Depression is a risk factor for coronary artery disease in men. Arch Intern Med *158:*1422, 1998.

Frasure-Smith N, Lesperance F, Prince RH, Verrier P, Garber RA, Juneau M, Wolfson C, Bourassa MG: Randomised trial of home-based psychosocial nursing intervention for patients recovering from myocardial infarction. Lancet *350:*473, 1997.

Frasure-Smith N, Lesperance F, Talajic M: Depression following myocardial infarction. Impact on 6-month survival. JAMA *270:*1819, 1993.

*Frasure-Smith N, Lesperance F, Talajic M: Depression and 18-month prognosis following myocardial infarction. Circulation *91:*999, 1995.

*Fricchione GL, Vlay LC, Vlay SC: Cardiac psychiatry and the management of malignant ventricular arrhythmias with the internal cardioverter-defibrillator. Am Heart J *128:*1050, 1994.

Friedman M, Thoresen CE, Gill JJ, Ulmer D, Powell LH, Price VA, Brown B, Thompson L, Rabin DD, Breall WS, Bourg E, Levy R, Dixon T: Alteration of type A behavior and its effect on cardiac recurrences in postmyocardial infarction patients: Summary results of the recurrent coronary prevention project. Am Heart J *112:*653, 1986.

Glassman AH, Roose SP, Bigger JT Jr: The safety of tricyclic antidepressants in cardiac patients. Risk-benefit reconsidered. JAMA *269:*2673, 1993.

*Glassman AH, Shapiro PA: Depression and the course of coronary artery disease. Am J Psychiatry *155:*4, 1998.

Hurt RD, Sachs DPL, Glover ED, Offord KP, Johnston JA, Dale LC, Khayrallah MA, Schroeder DR, Glover PN, Sullivan CR, Croghan IT, Sullivan PM: A comparison of sustained-release bupropion and placebo for smoking cessation. N Engl J Med *337:*1195, 1997.

Jiang W, Babyak M, Krantz DS, Waugh RA, Coleman RE, Hanson MM, Frid DJ, McNulty S, Morris JJ, O'Connor CM, Blumenthal JA: Mental stress—induced myocardial ischemia and cardiac events. JAMA *275:*1651, 1996.

Katon W: Panic disorder: Relationship to high medical utilization, unexplained physical symptoms, and medical costs. J Clin Psychiatry *57*(Suppl):11, 1996.

P>Kawachi I, Colditz GA, Ascherio A, Rimm E, Giovannucci E, Stampfer M, Willet WC: Prospective study of phobic anxiety and risk of coronary heart disease in men. Circulation *89:*1992, 1994.

Kawachi I, Sparrow D, Vokonas PS, Weiss ST: Symptoms of anxiety and risk of coronary heart disease. The Normative Aging Study. Circulation *90:*2225, 1994.

Leor WJ, Poole WK, Kloner RA: Sudden cardiac death triggered by an earthquake. N Engl J Med *334:*413, 1996.

Levenson J, Dwight M: Cardiovascular disease. In *Psychiatric Care of the Medical Patient*, A Stoudemire, BF Fogel, PT Greenberg, editors. Oxford University Press, New York 1998.

*Linden W, Stossel C, Maurice J: Psychosocial interventions for patients with coronary artery disease. Arch Intern Med *156:*745, 1996.

Matthews K: CHD and type A behaviors: Update on and alternative to the Booth-Kewley and Friedman quantitative review. Psychol Bull *104:*373, 1988.

Musselman DL, Tomer A, Manatunga AK, Knight BT, Porter MR, Kasey S, Marzec U, Harker LA, Nemeroff CB: Exaggerated platelet reactivity in major depression. Am J Psychiatry *153:*1313, 1996.

Musselman D, Evans DL, Nemeroff CB: The relationship of depression to cardiovascular disease. Arch Gen Psychiatry *55:*580, 1998.

Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry *153:*311, 1996.

Roach GW, Kanchuger M, Mangano CM, Newman M, Nussmeier N, Wolman R, Aggarwal A, Marschall K, Graham SH, Ley C, Ozanne G, Mangano DT: Multicenter study of Perioperative Ischemia Research Group, Ischemia Research and Education Foundation Investigators: Adverse cerebral outcomes after coronary bypass surgery. N Engl J Med *335:*1857, 1996.

Roose SP, Laghrissi-Thode F, Kennedy JS, Nelson JC, Bigger JT Jr, Pollock BG, Gaffney A, Narayan M, Finkel MS, McCafferty J, Gergel I: Comparison of paroxetine and nortriptyline in depressed patients with ischemic heart disease. JAMA *279:*287, 1998.

Roose SP, Glassman AH, Ahia E, Woodring S, Gardina EGV, Bigger JT: Cardiovascular effects of fluoxetine in depressed patients with heart disease. Am J Psychiatry *155:*660, 1998.

Shapiro PA, Levin HR, Oz MC: Left ventricular assist devices: Psychosocial burden and implications for heart transplant programs. Gen Hosp Psychiatry *18:*30S, 1996.

*Shapiro PA, Lespérance F, Frasure-Smith N, O'Connor CM, Baker B, Jiang JW, Dorian P, Harrison W, Glassman AH: An open-label preliminary trial of sertraline for treatment of major depression after acute myocardial infarction (The SADHAT trial). Am Heart J *137:*22, 1999.

*Sloan RP, Shapiro PA, Bagiella E, Myers MM, Gorman JM: Cardiac autonomic control buffers blood pressure variability responses to challenge: A psychophysiological model of coronary artery disease. Psychosom Med *61:*58, 1999.

Stoudemire A, Knos G, Gladson M, Markwalter H, Sung YF, Morris R, Cooper R: Labetalol in the control of cardiovascular responses to electroconvulsive therapy in high-risk depressed medical patients. J Clin Psychiatry *51:*508, 1990.

Stoudemire A, Moran MG: Psychopharmacology in the medically ill patient. In *Textbook of Psychopharmacology*, ed 2, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1998.

*Wulsin LR, Vaillant GE, Wells VE: A systematic review of the mortality of depression. Psychosom Med *61:*6, 1999.

25.5 RESPIRATORY DISORDERS

MICHAEL G. MORAN, M.D.

Although usually unconscious and autonomous, breathing is intimately embedded in the psychic experience of human beings. Psychological distress commonly expresses itself in alterations of breathing, such as tachypnea during anxiety or panic. Conversely, disordered or impaired breathing quickly disrupts a person's sense of calm and even survival (e.g., the distress of any pulmonary patient who is acutely hypoxemic). Respiratory disease is likely to become a more prevalent problem for physicians caring for any patient population as the smoking habits of an older population begin to take their toll in the form of chronic obstructive pulmonary disease. In addition, the prevalence of marijuana smoking among younger persons may adversely interact with other pulmonary conditions, such as asthma.

The interaction of psychic and somatic influences becomes more complex when the etiology and maintenance of symptoms of pulmonary disorders is taken into consideration. For example, self-destructive or depressive feelings may generate or give momentum to a smoking habit, which could eventually result in chronic obstructive pulmonary disease.

The psychiatrist may be asked to evaluate pulmonary patients as to (1) the possible psychiatric origins of pulmonary symptoms, (2) the issues impeding compliance with a medical regimen, (3) the emotional distress arising from loss of function due to the pulmonary illness, and (4) psychiatric issues, such as delirium, caused by the treatment of the pulmonary condition.

ASTHMA

Diagnosis and Pathophysiology Asthma is a chronic and episodic illness, characterized by widespread narrowing of the tracheobronchial tree. Symptoms include coughing, wheezing, chest tightness, and dyspnea. Nocturnal symptoms and exacerbations are common. During an episode, nonspecific irritants such as perfume and smoke may trigger a worsening of the symptoms. Some patients may also have sinusitis and aspirin sensitivity. Diagnosis requires confirmation of reversible airway obstruction, usually established by spirometry. Pulmonary function tests demonstrate reduced maximal forced expiratory flows, with elevated residual volumes or functional residual capacity, or both. Irreversible changes in the lung are unlikely to occur.

The pathophysiology of airway inflammation in asthma has not been fully specified. Cells that may participate in the process include mast cells, macrophages, eosinophils, and activated T helper cells. These and other cells interact in a complex manner that results in the production of numerous chemical mediators. Neural influences may amplify the resulting intense local inflammatory process. Vascular permeability, smooth muscle contraction, mucosal edema, and mucus production result.

Prevalence, Morbidity, and Mortality Asthma is a common illness, affecting up to 5 percent of Americans. Onset may occur in infancy or early childhood and remissions during the teenage years are common. The mortality rate from asthma is reported to have increased, but at least one carefully constructed study suggests that morbidity, not mortality, is on the rise. Coexisting chronic obstructive pulmonary disease, usually caused by smoking, significantly reduces survival among patients with asthma. Factors contributing to the increase in morbidity include inappropriate use of b_2 adrenergic receptor agonists, underestimation of illness severity by the patient or physician, socioeconomic impediments to medical care access, and psychiatric factors.

Asthma as a Psychosomatic Illness The definition of *psychosomatic illness* has evolved considerably over the past 50 years. In the 1940s, Franz Alexander and others worked with patients who had medical illnesses they considered psychic in origin. Asthma was one of a group of the “Holy Seven” illnesses (i.e., peptic ulcer, asthma, ulcerative colitis, rheumatoid arthritis, hypertension, neurodermatitis, and hyperthyroidism) thought to be psychosomatic. The wheezing and bronchial secretions of the asthma patient were interpreted as a suppressed cry for the mother. Although Alexander took a predominantly psychoanalytic approach, he understood the intricacies of what is now called a biopsychosocial matrix. His long-term research projects on asthma and six other illnesses opened the door to the study of psychosomatic disorders and helped to solidify the significance of psychiatry's relationship to the rest of medicine. Alexander's work led to certain developments by some of his followers, and to what may now be considered a radicalization of treatment for these conditions, sometimes involving only psychotherapy or psychoanalysis. Poor outcomes and an expanded understanding of the physiology of asthma and the other six illnesses subsequently produced a backlash: a radical dismissal and underestimation of psychiatric contributions to symptoms and morbidity. Current formulations of the pathogenesis and maintenance of asthma embrace subtle interactions between organic and psychiatric elements.

Psychiatric Aspects of Asthma Research suggests that psychiatric forces may affect the clinical expression of asthma in several ways: altered awareness of airway resistance, suggestibility to airway constriction, and comorbidity with panic disorder and depression.

Suggestibility and Airway Awareness The assumption that suggestibility to airway constriction confirms the presence of psychopathology or psychosomatic vulnerability is weak but pervasive. Suggestible patients are more likely to be anxious and to feel physically vulnerable than nonsuggestible patients, but their actual comparative risk for more asthma episodes or symptoms as a result of stressful life events is unknown.

Perception of changes in airway resistance appears to be an important event in asthma self-care. The patient's timely reports of dyspnea can also help to initiate an appropriate medical response. From patient to patient there is considerable variability in reported breathlessness for a given degree of experimentally induced airway obstruction. Up to 15 percent are unable to detect even marked changes in airway resistance. However, in an individual patient, increasing breathlessness usually correlates closely with decreased forced expiratory volume (FEV) in 1 second (FEV_1). Patients with labile airways (over 50 percent variability in FEV_1 in a day) have an especially limited perception of increases in airway resistance, as do the elderly. Some research suggests that poor ability to perceive airway changes predicts a worse course for the illness; other reports show no difference in this ability between severely ill and less ill patients.

The respiratory response to hypoxemia is mediated by the carotid body. Some asthma patients have an impaired response that is probably an inherited defect of that structure; these individuals perceive dyspnea poorly and are at increased risk for fatal asthma.

Anxiety and Panic Disorder Coexisting anxiety or panic disorder probably worsens the course of asthma, and the prevalence of panic disorder and agoraphobia among asthma patients is higher than in the general population. Up to 30 percent of persons with asthma meet the criteria for panic disorder or agoraphobia. Panic disorder appears to be underrecognized in asthma patients, and its symptoms may be misunderstood by both the patient and physician as arising from an exacerbation of asthma. Hyperventilation, a common symptom of both anxiety and panic disorder, may trigger an exacerbation of a pulmonary illness. The mechanism may be via brainstem overreactivity to low arterial carbon dioxide (CO_2) levels. Lactate infusion tests, breathing elevated levels of carbon dioxide, and voluntary hyperventilation have been shown experimentally to induce panic attacks in individuals with panic disorder. Voluntary hyperventilation may be the most specific test and lactate infusion the most sensitive.

Medullary brainstem centers are connected to the hippocampus, which receives data from a number of sensory modes, and the parahippocampus, which may be abnormal in panic disorder patients. Incoming body sensory data, including chest wall and intratracheal air movement, together with abnormal brainstem sensitivity to elevated partial pressure of carbon dioxide (Pco_2) or elevated serum lactate, may provoke the panic response and hyperventilation. The final pathway appears to be through the locus coeruleus, which fires in response to hypercapnia, and may be a controlling site for the intense autonomic outflow of panic attacks.

Repeated experience with severe dyspnea or near-suffocation provides little in the way of cognitive or emotional protection for future episodes. Such events often seem only to traumatize patients, and may lead to the development of anxiety disorders or panic disorders. Cohorts of panic disorder patients are more likely than

patients with other anxiety disorders to have an antecedent history of lung disease.

Patients with both asthma and panic disorder seem to have greater anticipatory anxiety about both dyspnea and anxiety itself than patients who are not comorbid. The fear of dyspnea may directly trigger asthma attacks. An extremely high level of anxiety predicts increased rates of hospitalization and asthma-associated mortality. Certain personality traits in asthma patients appear to be associated with greater use of corticosteroids and bronchodilators, as well as longer hospitalizations than would be predicted from pulmonary function alone. Such traits include intense fear, emotional lability, sensitivity to rejection, and lack of persistence in difficult situations.

If a patient attributes symptoms of panic to asthma, the use of antianxiety medications such as corticosteroids, b_2 adrenergic receptor agonists, and theophylline (Theo-Dur) may worsen the panic disorder. Careful use of peak flow meters during periods of dyspnea and a diary correlating peak flow rates with experiences of breathlessness can help patients to determine the origin of their symptoms.

Depression and the High-Risk Asthma Patient Many researchers suggest that depression has a meaningful and negative effect on the course of asthma. Shame and low self-esteem, common symptoms of depression, are seen in many chronically ill patients and are risk factors for a severe course. Patients who feel they are worth little tend to manage their asthma poorly, and compliance with a medication regimen may suffer. The sleep disturbances of depression could adversely affect self-care, as could the concentration and attention deficits of the mood disorder. Sleep deprivation may reduce a patient's proprioceptive capacity to detect changes in airway resistance and can erode daytime cognitive performance. In a severely depressed patient unconscious suicidal wishes may find expression in self-neglect. In addition to these behavioral manifestations, the hypothetical parasympathetic dominance of depression could contribute to airway reactivity and constriction.

Although there is no asthma-prone personality, psychopathology of the level of personality disorders raises the risk for near-fatal episodes. High levels of aggression, denial of or disregard for asthma symptoms, and psychotic signs are associated with poor outcome or fatal episodes.

Asthma deaths occur by one of two courses: gradual deterioration and abrupt onset, with the gradual course being more common. Typically, the patient and the physician underestimate the severity of the episode and are not aggressive enough in their approach. b_2 adrenergic receptor agonist inhalants may actually provide an illusion of reduced bronchoconstriction, which gives the patient false assurance; progressive airflow obstruction in these instances is detectable only by peak flow measurements. The abrupt decline is less common and the mechanism less understood. The immunohistology differs from that in patients with a gradual course, and eosinophils are less plentiful in the airways.

Certain demographic characteristics have a clear association with greater risk for death from asthma. Minority group status, once thought to be a risk factor, is now believed to be a surrogate marker for poverty. With low socioeconomic position comes poor access to high-quality health care, indigent housing with high levels of dust-mite and cockroach antigens, low levels of social support, and low educational status.

Family members of patients with severe asthma tend to have higher than predicted prevalence rates of mood disorders, posttraumatic stress disorder, substance use, and antisocial personality disorder. How these conditions contribute to the genesis or maintenance of asthma in an individual patient is unknown. The familial and current social environment may interact with a genetic predisposition for asthma to influence the timing and severity of the clinical picture. The interaction between family psychopathology and the clinical course of asthma may be especially insidious in adolescents. In turn, the adolescent's developmental need for and fear of emotional separation from the family often becomes entangled in battles over medication adherence as well as other modes of diligent self-care.

Vocal Cord Dysfunction This syndrome of abnormal vocal cord movement may imitate asthma and lead to misdiagnosis. Patients may present a history of years of refractory asthma with intubations, corticosteroid treatment, even immunomodulator use or tracheostomy. Diagnosis is very difficult on physical examination because the abnormal vocal cord movement can produces laryngeal sounds that resemble wheezing (over the chest) or stridor (over the larynx). Laryngoscopy is usually necessary to make the diagnosis. Adduction of the anterior two-thirds of the cords is the classic finding.

The psychiatrist may be asked to see the patient with vocal cord dysfunction for several reasons. The change in diagnosis to vocal cord dysfunction from asthma may be disorienting, and regret and anger about past unnecessary treatment can ensue. Some patients with vocal cord dysfunction may have a comorbid psychiatric diagnosis and traumatic childhood history that contributes to the appearance or maintenance of the symptoms, and interferes with their care.

Adverse Effects of Drugs Many of the oral and inhaled medications for treating asthma produce varying degrees of alertness, stimulation, tremor of the extremities, agitation, and sleep disruption. Both physician and patient may misinterpret these symptoms and signs as an anxiety disorder that requires separate treatment, often with a benzodiazepine. The physician must try to separate drug effects from the signs and symptoms of a new anxiety disorder. Adjustment of the dose and timing of administration of the offending pulmonary medication can often address the patient's complaints adequately.

Oral corticosteroids produce varying degrees of memory disturbance, sleep disturbance, emotional lability, depression, and hypomania or mania. Most patients experience some kind of dysphoria, a few feel energized. These effects tend to be dose-dependent, and usually present few problems at doses lower than the equivalent of 30 mg of prednisone (Deltasone) per day; however, even at these dosages, patients report problems with sleep continuity and irritability.

Treatment The psychiatrist should address (1) psychiatric factors contributing to the severity of asthma, (2) family system problems, and (3) drug adverse effects and interactions. There are other important considerations, such as countertransference problems among staff and physicians treating chronically ill patients.

If an accurate diagnosis of panic disorder, other anxiety disorder, or depressive disorder is made, medication may be indicated. Benzodiazepines are usually not needed for anxiety disorder or panic disorder but may be indicated for short-term use while beginning an antidepressant agent. These medications are generally acceptable for most asthma patients, as Pco_2 levels tend to be normal unless the patient is in overt respiratory failure. Despite mild stimulating effects, many patients tolerate conventional doses of selective serotonin reuptake inhibitors. Nortriptyline (Pamelor) is a good alternative drug. Monoamine oxidase inhibitors should be avoided because of potentially catastrophic interactions with other asthma drugs that have direct and indirect sympathomimetic activity. Cognitive-behavioral psychotherapy or stress reduction practices such as meditation and yoga are useful for many patients. Psychodynamic psychotherapy is recommended as a component for complete therapy of depression. Family system interventions, such as education and multifamily groups, can help to identify members who need special support and treatment.

Although offending medications such as corticosteroids can sometimes be stopped or reduced, the psychiatrist may need to plan the treatment of any syndrome resulting from the adverse effects of medication without making such adjustments. Mood disorders should be treated as in other situations. The chief difference in treating such syndromes in the medically ill patient resides in the need for vigilance for delirium or other cognitive disorders, and for drug-drug interactions. Patients with asthma and other medical illnesses have an increased vulnerability to infection and metabolic disturbances and are usually on several other medications. As a result, there may be a lowered threshold for adverse events, such as delirium from anticholinergic effects from an added tricyclic antidepressant medication. The psychiatrist should start at lower doses and proceed with close follow-up for dosage adjustment.

A 32-year-old respiratory therapist with a history of asthma presented to the emergency room with severe dyspnea of several hours' duration. She had been working “doubles” in her hospital for the preceding 3 weeks and was fatigued and sleep deprived. She noticed a sudden increase in her wheezing earlier in the day, and had been unable to control symptoms with her usual regimen of inhaled bronchodilators. On examination, she was found to be anxious and depressed: she had recently broken up with a boyfriend. After the patient achieved stabilization with nebulized bronchodilators, a psychiatrist was asked to see her before she left the emergency room. During that interview conscious feelings of rage and sadness were noted and interpreted in connection with the breakup. The psychiatrist also noted upper airway stridor when the patient cried, and recommended that she have a laryngoscopic examination. This was accomplished, and a coincident diagnosis of vocal cord dysfunction was made. For her asthma the patient was placed on a regimen of inhaled steroids, and was asked to keep careful daily records of peak expiratory flows. She was referred to a speech therapist for the vocal cord dysfunction. The psychiatrist recommended outpatient psychotherapy and nortriptyline for her mood symptoms.

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

Diagnosis and Pathophysiology Chronic obstructive pulmonary disease refers to a spectrum of disorders that are characterized by three pathophysiological aspects. The bronchial component is manifested as chronic bronchitis with chronic cough and sputum production. The second component is the acinar enlargement of emphysema, and is associated with smoking or $alpha_1$-antitrypsin deficiency. The third component is inflammation, which produces fibrosis and narrowing of the

airways, thought to be important in the production of airway resistance and airflow obstruction.

Clinicians usually recognize two forms of the disease: chronic bronchitis and emphysema. The former classically presents with CO_2 retention, hypoxemia, bronchial symptoms, dyspnea on exertion, and a history of recurrent bouts of acute bronchitis (“the blue bloater”). The latter form typically manifests severe hypoxemia, very little CO_2 retention, dyspnea on exertion, asthenic habitus, and severe air trapping with large residual volumes (“the pink puffer”).

Smoking is recognized as the single greatest risk factor and most important cause of chronic obstructive pulmonary disease. Other environmental elements may serve a modifying role, including air pollutants, childhood respiratory infections, the level of bronchial hyperresponsiveness, and constitutional and genetic factors other than $alpha_1$-antitrypsin deficiency.

Prevalence, Morbidity, and Mortality Chronic obstructive pulmonary disease affects more than 16 million Americans. The rate of cigarette use is in the decline among the population at large, but is rising among teenagers and predicts a large cohort of future patients with chronic obstructive pulmonary disease. Morbidity of the disease appears in the forms of lost time at work, decreased capacity for joyful exercise and exertion, early medical retirement, and recurrent episodes of respiratory deterioration leading to hospitalization. Mortality rates are increasing, especially among women. Chronic obstructive pulmonary disease is now the fifth leading cause of death in this country.

Measurement of FEV_1 is the standard for assessment of clinical course and response to therapeutic interventions, and provides data for predicting prognosis and mortality.

Psychiatric Aspects Psychiatric issues affect many facets of the course of chronic obstructive pulmonary disease from etiology to ongoing symptoms and the appearance of related respiratory syndromes, as well as psychiatric comorbidity and its effects on rehabilitation.

Smoking Emotional and social factors probably influence many persons to begin or continue cigarette smoking, which is a major element in the etiology of the disease. Education is notoriously weak as an isolated antismoking intervention, but probably serves as a positive influence on vacillating ambivalent smokers. Several researchers have shown that depression among smokers is associated with low rates of successful cessation, suggesting the possible utility of a concerted effort to diagnose and treat depression in this group of patients.

The older patient with chronic obstructive pulmonary disease commonly resists counseling to stop smoking by countering that there will be no benefit: “It doesn't matter now.” Some of these patients will respond to education about the significance of ongoing irritant effects of smoke and the effects of carbon monoxide on oxygen-carrying capacity and thus on exercise tolerance. In making the case for cessation, attempts should be made to appeal to aspects of the self or health that are important to that individual: personal attractiveness smoking increases facial wrinkling, affects the smell of the breath and clothes); exercise capacity (smoking reduces endurance by several mechanisms); independence (continued disease progress will erode autonomy through loss of function); longevity and ability to enjoy children and grandchildren (increased morbidity and mortality).

Panic Disorder, Anxiety Disorders, and Drug Effects As for asthma, prevalence rates for both panic disorder and anxiety disorders are increased among patients with chronic obstructive pulmonary disease. Anxiety disorders occur at rates of 16 to 34 percent, above the rate for the general population of 15 percent. Panic disorder prevalence rates among chronic pulmonary obstructive disease patients range from 8 to 24 percent, higher than the general prevalence of 1.5 percent.

Patients with chronic obstructive pulmonary disease can benefit from the use of inhaled sympathomimetic agents, but two points deserve emphasis. First, use of high doses can produce hypokalemia. Second, inadequate bronchodilatation with inhaled agents may lead to the use of oral b_2 adrenergic receptor agonists, which have a high incidence of adverse effects, including tremor, anxiety, and interference with sleep.

Anticholinergic drugs such as atropine and ipratropium (Atrovent) reduce the volume of sputum without changing viscosity, and induce beneficial bronchodilatation. There is limited absorption of these inhaled drugs, thus there is little concern for central nervous system effects.

The role of theophylline and related medications in chronic obstructive pulmonary disease is controversial. Since its use continues, psychiatrists must be aware of the stimulating adverse effects and the resulting tremor, nervousness, and sleep disruptions. Long-acting preparations and careful timing of the dose can help reduce effects on sleep. Any of the oral bronchodilators can cause psychiatric symptoms, but extended-release albuterol (Ventolin) is probably the biggest offender in terms of sleep disruption, agitation, and anxiety. Avoidance of this medication may be particularly important among elderly patients who are prone to delirium.

Corticosteroid use is also controversial in patients who have chronic obstructive pulmonary disease. Some small fraction of patients who have coincident asthma or a significant bronchospastic component to their disease exacerbations may benefit from either short-term or long-term courses of corticosteroid treatment. For some of these patients, alternate-day oral administration and inhaled steroid use may be useful. Both these measures reduce blood and brain concentrations of the steroid, and thus the likelihood of psychiatric adverse effects, including mood alteration, cognitive disturbances, delirium, and sleep disorders.

Nighttime symptoms in the patient with chronic obstructive pulmonary disease deserve special attention. Reports of choking or erratic breathing, heavy snoring, or abrupt awakenings due to dyspnea should suggest the diagnosis of obstructive sleep apnea. The patient may report none of these symptoms, as the awakenings that occur are usually not enough to register in memory. Daytime somnolence, depressed mood, and cognitive disturbances are the more likely complaints, but the patient may not mention these and the psychiatrist must solicit them. Nocturnal polysomnography is usually necessary to make the diagnosis because it permits adequate documentation of the number and severity of apneic episodes.

Rehabilitation The care of many patients with chronic obstructive pulmonary disease remains suboptimal, largely because of reversible factors. Many physicians do not intervene early enough in the disease, consider it only irreversible, avoid attending to comorbid depression and anxiety, and believe that rehabilitation is only an in-hospital treatment that most patients do not need.

Exercise conditioning is an essential aspect of rehabilitation, and should include upper-extremity and respiratory muscle training. Malnutrition may affect between 25 and 50 percent of patients with chronic obstructive pulmonary disease. These and other measures improve mood and sense of well-being even among patients whose objective measures of pulmonary function do not change in an extended rehabilitation program.

Mood appears to influence exercise tolerance and activity level. Patients with depressed mood or poor self-esteem have less exercise tolerance for a given level of pulmonary function. The effect of pulmonary function on mood is less clear. For one group of patients not yet on supplemental oxygen, level of pulmonary symptoms directly influenced mood and level of function. In another study of a group with severe airflow impairment, functions such as self-care, eating, and ambulating were affected but mood was not. These data suggest a resilience of mood over time to the effects of functional loss.

Treatment The psychiatrist should (1) diagnose and treat comorbid psychiatric disorders, (2) be alert to drug adverse effects and give advice regarding their dosing and timing, and (3) maintain a high index of suspicion for treatable coexisting sleep disorders.

Treatment of comorbid psychiatric disorders can help with day-to-day functioning as well as promote the best results from rehabilitation programs and ongoing rehabilitation efforts. Antidepressant medications are the best choice for anxiety and panic. Benzodiazepines are problematic, chiefly for chronic bronchitic retainers of CO_2. Levels of CO_2 as low as 44 mg Hg have been associated with ventilatory suppression by these medications. Amelioration of depression with psychotherapy and an antidepressant agent can improve quality of life, interpersonal relationships (and thus family support), and overall physical comfort.

In a patient who is chronically ill with chronic obstructive pulmonary disease daytime symptoms of fatigue, depression, and even somnolence will sometimes be dismissed as to be expected and result in no further investigation. Patients will usually not present symptoms of sleep disturbances as such in obstructive sleep apnea, so the psychiatrist must be vigilant about these symptoms in this context and must help make appropriate referrals to a sleep laboratory.

A 61-year-old man, a former cigarette smoker with chronic obstructive pulmonary disease with chronic elevation of Pco_2 presented to the pulmonary clinic with chronic fatigue and an acute exacerbation of his dyspnea. He was on 24-hour oxygen supplementation by nasal cannula. Evaluation revealed an overweight man with normal vital signs, no change in his FEV_1, and a chest X-ray that was unchanged from previous studies. His pulmonologist referred him for psychiatric evaluation because no organic cause could be found for the dyspnea. The psychiatrist found the man to be slightly confused, with some mild changes in short-term memory; the patient had a history of occasional use of alprazolam (Xanax) for anxiety. The patient said that when anxious, he would also sometimes increase his oxygen flow rate above what was recommended by his pulmonologist. Arterial blood gases evaluations were ordered. The psychiatrist obtained a history of severe snoring, occasional awakenings with panic, and chronic daytime somnolence. To rule out obstructive sleep apnea, the psychiatrist recommended a sleep study and recommended cessation of the benzodiazepine, to which the patient may have been very sensitive because of his chronic Pco_2 elevation. The benzodiazepine may also have worsened the sleep apnea and contributed to a mild daytime delirium. The blood gases obtained at that visit revealed an acute deterioration with greater than usual elevations of Pco_2 and slightly worse hypoxemia. The psychiatrist recommended careful adherence to prescribed flow rates of oxygen to avoid “oxygen toxicity” and suppression of hypercapneic ventilatory drive, another possible source of the delirium. For the panic and anxiety the psychiatrist prescribed paroxetine (Paxil).

SUGGESTED CROSS-REFERENCES

Many psychiatric diagnoses have a respiratory component. Section 11.4 discusses caffeine-related disorders, Section 11.9 nicotine-related disorders, and Section 11.5 cannabis-related disorders. Chapter 15 discusses anxiety disorders, and Chapter 21 discusses sleep disorders (including sleep apnea). Somatoform disorders are discussed in Chapter 16, and stress and psychiatry is covered in Section 25.9.

SECTION REFERENCES

Borson B, McDonald GJ, Gayle T, Deffenbach M, Lakshminarayan S, Van Tuinen C: Improvement in mood, physical symptoms, and function with nortriptyline for depression in patients with chronic obstructive pulmonary disease. Psychosomatics *33:*190, 1992.

Breslau N, Kibley NM, Andreski P: Vulnerability to psychopathology in nicotine-dependent smokers: An epidemiologic study of young adults. Am J Psychiatry *150:*941, 1993.

Buist AS, Vollmer WM: Preventing deaths from asthma (editorial). N Engl J Med *331:*1584, 1994.

Covino NA, Dirks JF, Kinsman RA, Seidel JV: Patterns of depression in chronic illness. Psychother Psychosom *37:*144, 1982.

Ferguson GT, Cherniak RM: Management of chronic obstructive pulmonary disease. N Engl J Med *328:*1017, 1993.

Karajgi B, Rifkin A, Doddi S, Kolli R: The prevalence of anxiety disorders in patients with chronic obstructive pulmonary disease. Am J Psychiatry *147:*200, 1990.

Kikuchi Y, Okabe S, Tamura G, Hida W, Homma M, Shirato K, Takishima T: Chemosensitivity and perception of dyspnea in patients with a history of near-fatal asthma. N Engl J Med *330:*1329, 1994.

Kinsman RA, Luparello T, O'Banion K, Spector S: Multidimensional analysis of the subjective symptomatology of asthma. Psychsom Med *35:*250, 1973.

136 Ley R: Pulmonary function and dyspnea suffocation theory of panic. J Behav Ther Exp Psychiatry *29:*1, 1998.

McFadden ER Jr, Gilbert IA: Asthma. N Engl J Med *327:*1928, 1992.

Miller BD, Wood BL: Psychophysiologic reactivity in asthmatic children: A cholinergically mediated confluence of pathways. J Am Acad Child Adolesc Psychiatry *33:*1236, 1994.

Miller BD IW: Influence of specific emotional states on autonomic reactivity and pulmonary function in asthmatic children. J Am Acad Child Adolesc Psychiatry *36:*669, 1997.

*Moran MG: Psychiatric aspects of asthma. Semin Resp Crit Care Med *15:*168, 1994.

Nelms BC: Emotional behaviors in chronically ill children. J Abnorm Child Psychol *17:*657, 1989.

Newman KB, Dubester SN: Vocal cord dysfunction: Masquerader of asthma. Semin Resp Crit Care Med *15:*161, 1994.

Philibert RA, Richards L, Lynch CF, Winokur G: The effect of gender and age at onset of depression on mortality. J Clin Psychiatry, *58:*355, 1997.

Pine DS, Weese-Mayer DE, Silvestri JM, Davies M, Whitaker AH, Klein DF: Anxiety and congenital central hypoventilation syndrome. Am J Psychiatry *151:*864, 1994.

Ringsberg KC, Segesten K, Akerlind I: Walking around in circles—the life situation of patients with asthma-like symptoms but negative asthma tests. Scand J Caring Sci *11:*103, 1997.

Rocco PL, Barboni E, Balestrieri M: Psychiatric symptoms and psychological profile of patients with near fatal asthma: Absence of positive feelings. Psychother Psychosom *67:*105, 1998.

*Scharloo M, Kaptein AA, Weinmann J, Hazes JM, Willems LN, Bergman W, Rooijmans HG: Illness perceptions, coping and functioning in patients with rheumatoid arthritis, chronic obstructive pulmonary disease and psoriasis. J Psychosom Res *44:*573, 1998.

Silverstein MD, Reed CE, O'Connell EJ, Melton LJ, O'Fallon WM, Yunginger JW: Long-term survival of a cohort of community residents with asthma. N Engl J Med *331:*1537, 1994.

*Smoller JW, Pollack MH, Otto M, Rosenbaum JF, Kradin RL: Panic anxiety, dyspnea, and respiratory disease. Theoretical and clinical considerations. Am J Resp Crit Care Med *154:*6, 1996.

Smoller JW, Pollack MH, Systrom D, Kradin RL. Sertraline effects on dyspnea in patients with obstructive airways disease. Psychosomatics *39:*24, 1998.

Spinhoven P, van Peski-Oosterbaan AS, Van der Does AJ, Willems LN, Sterk PJ: Association of anxiety with perception of histamine-induced bronchoconstriction in patients with asthma. Thorax *52:*149, 1997.

Stout C, Kotses H, Creer TL: Improving perception of air flow obstruction in asthma patients. Psychosom Med *59:*201, 1997.

Strunk RC, Mrazek DA, Fuhrmann GS, LaBrecque JF: Physiologic and psychological characteristics associated with deaths due to asthma in childhood: A case-controlled study. JAMA *254:*1193, 1985.

Wamboldt MZ, Weintraub P, Krafchick D, Wamboldt FS: Psychiatric family history in adolescents with severe asthma. J Am Acad Child Adolesc Psychiatry *35:*1042, 1996.

*Weinberger SE: Recent advances in pulmonary medicine. N Engl J Med *328:*1389, 1993.

*Weinstein AG, Chenkin C, Faust D: Caring for the severely asthmatic child and family. I. The rationale for family systems integrated medical/psychological treatment. J Asthma *34:*345, 1997.

Wingate BJ, Hansen-Flaschen J: Anxiety and depression in advanced lung disease. Clin Chest Med *18:*495, 1997.

Yellowlees PM, Kalucy RS: Psychobiological aspects of asthma and the consequent research implications. Chest *97:*628, 1990.

25.6 ENDOCRINE AND METABOLIC DISORDERS

VICTORIA C. HENDRICK, M.D., AND THOMAS R. GARRICK, M.D.

Endocrine disorders may produce symptoms that are indistinguishable from psychiatric illnesses. While physical manifestations of endocrine disease provide clues to the diagnosis, they are not always present. The effect of endocrinopathies on psychiatric symptomatology has been studied particularly for disorders of the thyroid and adrenal glands. Less is known about psychiatric sequelae of other endocrine disorders such as acromegaly, diabetes mellitus, prolactin-secreting tumors, and hyperparathyroidism. While several hormones directly influence neurotransmitter function (e.g., thyroid hormones and corticosteroids affect b-adrenergic receptor binding and density), biogenic amines have a reciprocal role, influencing hormone synthesis and release. However, little research has been done on the effect of psychiatric or psychological factors on endocrine disease. Stressful life events have been postulated to play a role in the onset of Graves' and Cushing's diseases, but studies have been limited by methodological deficiencies, including retrospective designs and absence of control groups.

GENERAL PRINCIPLES

Endocrinology is the study of hormones, or chemical mediators, that travel from their site of production to influence the function of other tissues. They exist as peptides, steroids, or amines. Endocrinopathies result from excessive or deficient amounts of a hormone or from resistance to a hormone's action (e.g., when target tissues do not respond to normal levels of the hormone. These latter cases are primarily hereditary and are characterized by normal or elevated concentrations of a hormone despite an absence or deficiency of the hormone's action. Increased concentrations result from the lack of normal negative feedback on the hormone's production, a consequence of the failure or deficiency of the hormone's action.

Hormone production is regulated largely by the concentrations and activity of the hormones themselves and additionally by hormones from other sites. For example, hypothalamic releasing factors regulate the production of pituitary hormones. The pituitary gland, situated in the sella turcica at the base of the skull, is divided into the anterior and posterior lobes. The anterior pituitary produces six hormones: thyroid-stimulating hormone (TSH), which regulates the production of thyroid hormones; growth hormone (GH), which promotes growth and development; prolactin, which stimulates lactation; luteinizing hormone (LH) and follicle-stimulating hormone (FSH), which are necessary for gonadal function in men and women; and adrenocorticotropic hormone (ACTH), which controls adrenal cortex production of glucocorticoids. The posterior pituitary stores two hormones: antidiuretic hormone (arginine vasopressin [AVP]), which regulates fluid balance, and oxytocin, which is necessary for milk letdown during breast feeding. Both are produced by the hypothalamus. The hypothalamus additionally produces releasing factors that influence pituitary hormone production: growth hormone–releasing hormone (GHRH) and growth hormone–release-inhibiting factor (i.e., somatostatin), which influence production of GH; gonadotropin-releasing hormone (GnRH), which influences production of LH and FSH; prolactin-release-inhibiting factor (i.e., dopamine), and prolactin-releasing factor (PRF), influencing prolactin production; and thyrotropin-releasing hormone (TRH) and corticotropin-releasing hormone (CRH), which influence TSH and ACTH release, respectively.

In addition to the hypothalamus and the pituitary, the major endocrine organs include the adrenal, thyroid, and parathyroid glands and the ovaries and testes. Endocrinopathies may produce a variety of psychiatric symptoms, as hormones affect a variety of organ systems. Patients are more likely to present primarily with psychiatric symptoms early in the course of endocrine disease and to experience a worsening of the psychiatric condition and emergence of physical symptoms with the progression of the disorder.

Endocrine disorders are common medical conditions, particularly in elderly adults. Thus, especially in this population, an evaluation for endocrine function is an essential part of the initial diagnostic workup. To identify potential endocrine etiologies, a laboratory evaluation should include serum electrolytes, fasting blood sugar, TSH, and calcium determinations. Psychiatric symptoms will not resolve until the underlying endocrine disturbance is corrected and may worsen with psychotropic agents. Psychiatric disorders that appear to result from an endocrine condition are categorized in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) under mental disorders due to a general medical condition, with the medical condition stated on Axis III.

THYROID DISORDERS

The thyroid hormones, triiodothyronine (T_3) and thyroxine (T_4), are produced from the thyroid gland and are necessary for regulation of metabolic activity and protein synthesis. They are formed from a common precursor, thyroglobulin, stored in the thyroid follicles. In extrathyroidal tissues, T_4 is converted to T_3, the metabolically active hormone. TSH produced by the anterior pituitary, promotes the synthesis and release of these hormones and is in turn regulated by TRH, produced in the hypothalamus. Low serum concentrations of thyroid hormones stimulate release of TRH and TSH. The central nervous systems (CNS) neurotransmitter dopamine decreases TSH secretion, as do glucocorticoids.

T_3 and T_4 are tightly protein bound to thyroxine-binding globulin (TBG), so that only 0.3 percent of T_3 and 0.02 percent of T_4 are free. Total serum T_4 concentrations vary with TBG concentrations; for example, exposure to estrogens (pregnancy and oral contraceptives) raises TBG and therefore serum T_4 concentration, while androgens have the opposite effect. Free thyroxine concentration (FT_4I), which measures the percentage of free T_4, produces a more accurate determination of thyroid function. FT_4I is elevated in hyperthyroid patients and low in hypothyroid patients and is thus an essential test in the assessment of thyroid status. Measures of TSH concentrations are also necessary in the workup of thyroid disorders. Concentrations are elevated in thyroidal hypothyroidism and usually low in pituitary-hypothalamic hypothyroidism. The normal concentration range of TSH is 0.3 to 4.7 mIU/L. FT_4I and free T_3 index (FT_3I) range from 5 to 11 and 75 to 175, respectively.

The TRH stimulation test (TRH infusion test) is sometimes used in the workup of thyroid function. After the patient's baseline TSH concentration is obtained, TRH is administered intravenously and TSH concentrations are then measured at regular intervals. A normal response is a rise in TSH of 5 µU/mL or more above baseline. In hypothyroid patients the response is exaggerated; in hyperthyroid patients it is blunted. With the development in the past decade of highly sensitive TSH testing, the TRH stimulation test is seldom used clinically. Its use as a diagnostic tool in the psychiatric assessment of depression has similarly waned, as its specificity and sensitivity are poor. For example, while some patients with depression display an increased TSH concentration following TRH infusion, others have a blunted response. Tests of radioactive iodine uptake are used to determine the activity of the thyroid gland. A patient receives a small amount of oral radioactive iodine and a scan then determines the amount taken up by the thyroid gland.

Disorders of the thyroid gland fall into two general categories: hyperthyroidism and hypothyroidism. Of all the endocrine disorders, disorders of the thyroid are the most common.

Hyperthyroidism

Definition and Comparative Nosology Hyperthyroidism, or thyrotoxicosis, results from overproduction of thyroid hormone by the thyroid gland. The most common cause is Graves' disease. Toxic nodular goiter causes another 10 percent of cases among middle-aged and elderly patients. Other causes include hyperfunctioning solitary thyroid adenomas, thyroiditis, use of exogenous thyroid hormone, TSH-producing pituitary adenoma, pituitary resistance of suppression of TSH secretion by thyroid hormone, thyroid carcinoma, choriocarcinoma, hydatiform moles, and struma ovarii.

Subclinical hyperthyroidism describes a condition in which circulating thyroid hormone concentrations are normal but the TSH response to TRH is blunted or absent. Patients may experience symptoms of hyperthyroidism such as nervousness, irritability, fatigue, and tachycardia. Subclinical hyperthyroidism may progress to overt hyperthyroidism.

Thyroid storm (thyrotoxic crisis), a life-threatening syndrome, is usually precipitated by illness or injury but may also develop following withdrawal of antithyroid drugs or following therapy with radioactive iodine. Symptoms include marked tachycardia, weakness, fever, and altered mental status.

Epidemiology In the general population, the prevalence of hyperthyroidism is about 0.5 percent. Graves' disease is the most common cause, accounting for 80 percent of cases. Graves' disease occurs four times more frequently in women than men and occurs primarily in the third and fourth decades. The prevalence of thyroiditis is estimated to be 5 to 9 percent in women in the postpartum period, usually resolving spontaneously by 1 year following delivery.

Etiology Patients with Graves' disease produce thyroid-stimulating antibodies and frequently have a family history of other autoimmune disorders such as systemic lupus erythematosus and myasthenia gravis. Hyperthyroidism from toxic nodular goiter results from the lack of response of thyroid nodules to normal feedback regulation. In the case of thyroiditis, hyperthyroidism results from inflammation-induced damage to the thyroid gland, which then spills the stored colloidal thyroid hormones.

Diagnosis and Clinical Features Signs and symptoms of hyperthyroidism include increased pulse, arrhythmias, elevated blood pressure, fine tremor, heat intolerance, excessive sweating, increased appetite, weight loss, palpitations, tachycardia, frequent bowel movements, menstrual irregularities, muscle weakness from catabolism of muscle protein, exophthalmos, lid lag, infrequent blinking, and hyperactive deep tendon reflexes. The patient with thyroiditis may experience pain and tenderness in the thyroid region.

Psychiatric features include nervousness, fatigue, insomnia, mood lability, and dysphoria. Speech may be pressured, and patients may demonstrate a heightened activity level. Cognitive symptoms include a short attention span, impaired recent memory, and an exaggerated startle response. Patients with severe cases may exhibit visual hallucinations, paranoid ideation, and delirium. While some symptoms of hyperthyroidism resemble those of a manic episode, an association between hyperthyroidism and mania has rarely been observed. A patient with Grave's disease (also called exophthalmic goiter) is illustrated in Figure 25.6-1.

FIGURE 25.6-1 Exophthalmic goiter. Note lid retraction and enlarged thyroid. (Reprinted with permission from Douthwaite AH, editor: *French's Index of Differential Diagnosis*, ed 7. Williams & Wilkins, Baltimore, 1954.)

Pathology and Laboratory Examination Laboratory examination will reveal increased concentrations of total and free T_4 (FT4). Concentrations of total and free T_3 are usually elevated, but this is not necessary to make the diagnosis of hyperthyroidism. The TSH concentration is abnormally low, as its production is suppressed by the elevated thyroid hormone concentrations, and the TSH response to TRH is blunted. Uptake of radioactive iodine by thyroid hormone is increased, except in cases of hyperthyroidism due to thyroiditis or ingestion of thyroid hormone.

Differential Diagnosis Panic disorder, generalized anxiety disorder, and social and specific phobias present with symptoms that resemble hyperthyroidism, including tachycardia, diaphoresis, and tremulousness. Other conditions that produce symptoms of autonomic hyperarousal include hypoglycemia, diabetes, dehydration, and pheochromocytoma. Chronic obstructive pulmonary disease may produce anxiety, diaphoresis, dyspnea, and tachycardia. In some cases, hyperthyroidism may produce severe agitation resembling a manic state. Certain medications (e.g., appetite suppressants, stimulants, and street drugs such as cocaine and amphetamines) may also produce these symptoms. Laboratory testing will distinguish between hyperthyroidism and other medical conditions.

Course and Prognosis Rates of remission of Graves' disease depend on the type and length of treatment: patients treated with antithyroid drugs for 6 months have a rate of remission of approximately 30 percent, compared with 80 percent in patients treated for 2 years. Patients treated with radioiodine rarely experience a recurrence of hyperthyroidism. However, permanent hypothyroidism is a common complication of radioactive iodine treatment. Patients who undergo surgical ablation have a relapse rate of approximately 10 percent and are also at risk for permanent hypothyroidism. Hyperthyroidism due to thyroid nodular disease is permanent but remits with radioactive iodine therapy; in contrast, thyroiditis is a short-lived condition, lasting only a few weeks and resolving spontaneously.

Treatment Treatments for Graves' disease are (1) antithyroid drugs, (2) radioactive iodine (RAI), and (3) surgical thyroidectomy. The antithyroid drugs, propylthiouracil (PTU) and methimazole (Tapazole), inhibit synthesis of thyroid hormone. Propylthiouracil additionally inhibits the monodeiodination of T_4 to T_3. Methimazole has the advantage of single-day dosing. Initial doses of methimazole range from 10 to 40 mg, and of propylthioruracil from 75 to 100 mg, three times a day. The doses should be adjusted as the patient returns to a euthyroid state. Usual maintenance doses of methimazole are 5 to 10 mg a day and of propylthiouracil, 50 to 100 mg a day. A rare and serious adverse effect of these medications is agranulocytosis. Other adverse effects include pruritus, rash, and fever. Length of treatment varies from 6 months to 2 years. These medications may not produce a therapeutic effect for several weeks, because of the long half-life of T_4. Some clinicians combine the antithyroid drugs with thyroxine replacement to prevent iatrogenic hypothyroidism.

While some patients with Graves' disease respond to antithyroid drugs, most require RAI. An RAI uptake test done prior to administration of ^{131}I allows determination of the appropriate dose. Hypothyroidism may complicate RAI therapy, and patients may subsequently require thyroid hormone replacement.

Subtotal thyroidectomy is an option for patients who have an allergy to antithyroid drugs or are unwilling to undergo RAI therapy. Patients should receive antithyroid drugs, potassium iodide, or b-adrenergic receptor antagonists so that they are euthyroid prior to surgery, to reduce the risk of postoperative thyrotoxic crisis. Subtotal thyroidectomy has several potential complications, such as hypoparathyroidism and injury to the recurrent laryngeal nerve. Patients may have recurrent hyperthyroidism after surgery if insufficient thyroid tissue is removed.

b-Adrenergic receptor antagonists such as propanolol (Inderal), metoprolol (Lopressor), atenolol (Tenormin), or nadolol (Congard) can provide symptomatic relief while treatment is in progress. Atenolol (50 to 100 mg a day) and nadolol (80 mg a day) have the advantage of once-daily dosing. These medications should be used cautiously in patients with asthma or heart failure.

Iodine (given as a solution of potassium iodide or as Lugol's solution) inhibits thyroid hormone. The brief duration of its antithyroid action (a few weeks) limits its use to pretreatment prior to surgery, management of thyroid storm, and speeding the reduction in thyroid hormone concentration following RAI therapy.

Treatment of thyroid nodular goiter consists of b-adrenergic antagonists and RAI. Treatment of thyroiditis consists of a brief course (a few weeks) of b-adrenergic antagonists, as this condition is short-lived.

For patients with psychotic symptoms, medium-potency antipsychotics are preferable to low-potency drugs, as the latter can worsen tachycardia. Similarly, tricyclic

drugs should be used with caution, if at all, in these patients. In general, the psychiatric symptoms resolve with successful treatment of the hyperthyroidism.

Hypothyroidism

Definition and Comparative Nosology Hypothyroidism results from inadequate synthesis of thyroid hormone and is categorized as either overt or subclinical. In overt hypothyroidism, thyroid hormone concentrations are abnormally low, TSH is elevated, and patients are symptomatic; in subclinical hypothyroidism, patients are by definition asymptomatic, with normal thyroid hormone concentrations but elevated TSH. Patients with normal levels of circulating T_3 and T_4 but an elevated TSH or exaggerated TSH response to TRH infusion are described as having subclinical hypothyroidism.

Hashimoto's thyroiditis (chronic lymphocytic thyroiditis) is the most common cause of hypothyroidism in adults. Most patients have a mild-to-moderate goiter but are asymptomatic and euthyroid; approximately 20 percent are hypothyroid. Occasionally, especially as the illness first develops, patients may be hyperthyroid.

Other causes of hypothyroidism include idiopathic atrophy, endemic hypothyroidism from a deficiency of dietary iodine, hypopituitarism (e.g., from postpartum pituitary necrosis), hypothalamic disease resulting in deficient production of TRH, and iatrogenic hypothyroidism (e.g., caused by drugs such as lithium [Eskalith] or antithyroid drugs or from surgical or chemical thyroidectomy). The risk of developing overt (grade I) hypothyroidism following long-term lithium use is estimated at 20 percent. The risk of developing subclinical hypothyroidism is another 20 percent and is higher in women and in patients with thyroid autoantibodies. Lithium's antithyroid effects are related to the length of treatment and appear to be more persistent in the presence of antithyroid antibodies. Patients with rapid-cycling bipolar disorder are especially vulnerable to lithium's antithyroid effects.

Hypothyroidism must be distinguished from euthyroid sick syndrome, in which any severe nonthyroidal illness may decrease conversion of T_4 to T_3, resulting in abnormally low T_3 concentrations.

Myxedema is a syndrome in which patients with hypothyroidism have infiltration of the skin by a mucopolysaccharide that produces puffy subcutaneous swelling, especially in the eyes, face, and extremities.

Epidemiology The prevalence of overt, or clinical, hypothyroidism is approximately 2 percent in women and less than 0.1 percent in men. Subclinical hypothyroidism also predominates in women, occuring in approximately 7.5 percent of women and 3 percent in men. Women appear to have a higher rate of hypothyroidism because of their greater rates of autoimmune disease. As the prevalence of autoimmune disease rises with age, the prevalence of both overt and subclinical hypothyroidism also rises, particularly in women.

Etiology Antibodies against the thyroid are the most prevalent cause of hypothyroidism. Because autoimmune disease occurs more often in women—the result of female reproductive hormones on immunological response—hypothyroidism predominates in women. The symptoms of hypothyroidism may result from decreased postsynaptic b-receptor number and function associated with low concentrations of thyroid hormones. Hypothyroidism additionally is associated with reduced central serotonin activity. Brain synthesis of serotonin is reduced in hypothyroidism, and the prolactin and cortisol responses to the serotonin agonist fenfluramine are blunted in hypothyroid patients (compared with euthyroid controls), suggesting reduced central serotonin activity. Also, serum serotonin concentrations correlate positively with serum T_3 concentrations.

The increased T_4 activity observed in psychiatric patients may reflect a compensatory mechanism to maintain normal levels of thyroid function in the brain. The increase in T_4, the prohormone, rather than T_3, the active hormone, makes more T_4 available for uptake into the brain without increasing peripheral thyroid activity, which would place a metabolic strain on the body. Alternatively, the increased T_4 concentrations may reflect a lower rate of peripheral conversion of T_4 to T_3.

Diagnosis and Clinical Features Symptoms of hypothyroidism tend to develop with TSH concentrations of 8 to 12 mEg/L or above and correspond to the degree of hormone deficiency. Signs and symptoms are varied and include cold intolerance, constipation, muscle cramps, paresthesias, menstrual disturbances (amenorrhea or menorrhagia), dyspnea, dizziness, syncope, reduced hearing, poor appetite, weight gain, brittle and thinning hair, husky voice, deep tendon reflexes, bradycardia, cardiomegaly, low-voltage complexes on electrocardiography (ECG), increased cholesterol and triglyceride concentrations, and normochromic, normocytic anemia from a reduced rate of red blood cell production.

Psychiatric symptoms include depressed mood, apathy, impaired memory and concentration, and a long response latency. Also, hypothyroidism may contribute to treatment-refractory depression. A psychotic syndrome of auditory hallucinations and paranoia, named “myxedema madness,” has been described in some patients. Patients with severe cases may exhibit diminished cerebral blood flow, with subsequent coma or death.

Subclinical hypothyroidism may produce depressive symptoms and cognitive deficits, although these tend to be less severe than those produced by overt hypothyroidism. Further, subclinical hypothyroidism more than doubles the lifetime risk of developing major depressive disorder. Subclinical hypothyroidism has been reported to reduce the efficacy of antidepressant treatment. The lifetime prevalence of depression in patients with subclinical hypothyroidism is approximately double that of the general population. These patients display a lower response rate to antidepressants and a greater likelihood of responding to T_3 augmentation than euthyroid patients with depression. After controlling for age and gender, the presence of antithyroid antibodies appears to be weakly associated with episodes of bipolar disorder but not with depressive (unipolar) disorders.

> A 45-year-old male with bipolar I disorder had been well on lithium monotherapy for 5 years, but over the course of several months he became increasingly withdrawn and reported fatigue and poor concentration. The patient's family noted he was forgetful and uncharacteristically listless and apathetic. No stressors were identified, and the patient's lithium concentration was therapeutic. His TSH µg concentration was elevated at 14 mEq/L. The patient was treated with levothyroxine 0.05 µg and within 6 weeks experienced an improvement in energy and a subjective return to baseline cognitive functioning. His TSH determined 8 weeks following initiation of thyroid replacement therapy was normal at 3.3 mEq/L.

Pathology and Laboratory Examination Of the thyroid function assays, TSH testing is the most sensitive assay for hypothyroidism. TSH levels are low in those with hypothyroidism produced by hypothalamic or pituitary disease and high if caused by thyroid gland dysfunction. Patients with thyroid autoimmune disease exhibit antithyroid antibody titers (of antimicrosomal or antithyroglobulin antibodies). T_4 concentrations are low, and T_3 concentrations may be either low or normal. Among patients with depression, a relative increase of total or free T_4 or both and a relative decrease in T_3 concentration is common. This condition, termed “euthyroid hyperthyroxinemia,” has also been observed among acutely admitted psychiatric and medical patients and thus is not specific to depression.

Both blunted and exaggerated TSH responses to TRH have been found in patients with refractory depression to TRH; the exaggerated TSH response (indicating subclinical hypothyroidism) is more consistently associated with treatment refractoriness. Although data are inconclusive, some studies have reported a greater prevalence of antithyroid antibody titers in patients with bipolar disorder and depression (estimated at 9 to 25 percent) than the general population rate of approximately 5 percent.

Differential Diagnosis Depression may present with symptoms similar to hypothyroidism, including lethargy, poor concentration, impaired memory, and apathy. Bipolar patients with depression may have lithium-induced hypothyroidism and thus thyroid testing is essential for diagnostics. Certain medications, including barbiturates, carbamazepine (Tegretol), and phenytoin (Dilantin), may lower total and free thyroid hormone concentrations by increasing their hepatic metabolism. Pernicious anemia presents with macrocytic anemia, lethargy, and decreased appetite. The incidence of pernicious anemia is increased in patients with autoimmune illnesses such as thyroiditis. The changes accompanying normal aging, like hypothyroidism, are associated with slight weight gain, thinning hair, dry skin, and slowed response latency.

Course and Prognosis The onset of hypothyroidism is usually gradual, except in cases of surgical or chemical thyroid ablation. Hypothyroidism in children (cretinism) may produce permanent physical and mental retardation. The physical and mental symptoms in adults are usually reversed by thyroid replacement. However, long-standing hypothyroidism may produce permanent cognitive deficits, possibly from neuronal death resulting from thyroid deficiency in the brain.

Depression in hypothyroid patients is unlikely to resolve until thyroid levels return to normal. Thyroid hormone alone may suffice to induce remission of the depressive changes, although standard antidepressant treatment may also be necessary for those with more severe cases.

Subclinical hypothyroidism has been associated with rapid-cycling bipolar disorders and with refractoriness to antidepressant treatment. Additionally, subclinical hypothyroidism has a high likelihood of progressing to overt hypothyroidism, particularly in women.

Treatment Current preparations of exogenous thyroid hormone include levothyroxine (Levothroid, Lexoxyl), synthetic T_4, liothyronine (Cytomel), synthetic T_3; liotrix (Thyrolar), mixed synthetic T_4 and T_3, and desiccated thyroid (Armour). Levothyroxine is preferred as it provides the most uniform thyroid replacement. Doses are typically between 50 and 200 mg a day. Replacement of thyroid hormone should be gradual, especially in patients over age 50 and those with a history of cardiac disease, to avoid precipitating angina or cardiac arrhythmias. The starting dose of levothyroxine for these patients should not exceed 0.05 mg daily. TSH concentrations should be measured 6 weeks after initiation of thyroid hormone treatment. Clinical improvement may not occur for several weeks.

For patients presenting with severe psychiatric symptoms (e.g., psychosis or suicidal depression), urgent psychiatric treatment is necessary. Psychotropic agents should be given at low doses initially, as the reduced metabolic rate in hypothyroid patients may produce reduced breakdown and therefore higher blood concentrations of medications.

The treatment of subclinical hypothyroidism is controversial. Since patients are by definition asymptomatic, there is no clear medical reason to treat. On the other hand, some patients experience increased energy and decreased constipation and skin dryness following treatment with thyroid supplementation. Thyroid testing should be followed at least yearly. Thyroid supplementation, however, should be considered for the following patients with subclinical hypothyroidism: (1) patients who exhibit thyroid antibodies, as they are at high risk of progression to overt hypothyroidism; (2) patients who are depressed, as subclinical hypothyroidism may produce treatment refractoriness (these patients should receive a 6-week trial of thyroid replacement before receiving antidepressant medications); (3) those who have a goiter; (4) those who have hypercholesterolemia (as subclinical and overt hypothyroidism increase serum lipid concentrations and hence the risk of coronary heart disease); and (5) patients taking lithium. Further, a trial of thyroid augmentation should be initiated for patients who (1) have a serum TSH concentration above 10 mU/L, (2) exhibit deficits in memory or cognitive functioning, or (3) have a rapid-cycling bipolar disorder, as some (but not all) studies have found subclinical hypothyroidism to be a risk factor in rapid cycling.

Thyroid supplementation can be administered in the form of levothyroxine 0.05 to 0.2 mg a day or liothyronine 0.025 to 0.1 mg a day. Thyroid supplementation can similarly be used for patients who become hypothyroid while taking lithium. Valproate (Depakene) and carbamazepine are reasonable alternatives to lithium for patients with a history of thyroid dysfunction or with lithium-induced hypothyroidism, as it has no effects on thyroid functioning.

The response to treatment should be assessed by frequent clinical and laboratory evaluations for thyroid hormone and TSH concentrations. TSH will diminish as T_3 and T_4 return to normal ranges. Care should be taken to not overtreat and place patients in a state of hyperthyroidism. A patient showing improvement after treatment for myxedema is illustrated in Figure 25.6-2 and Figure 25.6-3.

FIGURE 25.6-2 Patient with myxedema on admission from hospital. (Reprinted with permission from Haughton CS: Psychosis with myxoedema. Med J Aust *2:*766, 1959.)

FIGURE 25.6-3 Patient with myxedema following treatment (2 months later). (Reprinted with permission from Haughton CS: Psychosis with myxoedema. Med J Aust *2:*766, 1959.)

ADRENAL DISORDERS

The adrenal gland produces three major steroid hormones, glucocorticoids, mineralocorticoids, and sex steroids. Of these, the main steroids produced by the adrenal gland are cortisol, aldosterone, testosterone, and estrone/estradiol. Glucocorticoids promote gluconeogenesis, lipolysis, and glycogenesis, increase catecholamine activity during stress, and inhibit inflammatory reactions. Aldosterone's effect is primarily to increase renal sodium reabsorption and potassium excretion. Only a small proportion of circulating sex steroids are produced by the adrenal glands; their major site of production is the gonads.

Cortisol production is under the control of ACTH from the anterior pituitary, which in turn is regulated by CRH from the hypothalamus. Negative cortisol feedback occurs at pituitary, hypothalamic, and suprahypothalamic levels. CRH secretion follows a circadian pattern, subsequently producing circadian patterns of ACTH and cortisol concentrations, with their concentrations peaking in early morning and reaching a nadir in the evening. CRH release is also influenced by emotional and physical stress.

Hypercortisolism

Definition and Comparative Nosology Spontaneous Cushing's syndrome results from adrenocortical hyperfunction and can develop from either excessive secretion of ACTH (which stimulates the adrenal gland to produce cortisol) or from adrenal pathology (e.g., a cortisol-producing adrenal tumor). Cushing's disease, the most common form of spontaneous Cushing's syndrome, results from excessive pituitary secretion of ACTH, usually from a pituitary adenoma. Cushing's syndrome is ACTH dependent if the syndrome results from the effect of ACTH on the adrenal gland, and non-ACTH-dependent, in cases of excessive cortisol regardless of ACTH levels (e.g., as a result of an adrenal adenoma).

The hypothalamic-pituitary-adrenal axis responsivity to stressors may be influenced by stress in infancy and even during in utero development. Studies in primates and rodents show that prenatal stress may induce long-lasting effects on hypothalamic-pituitary-adrenal axis regulation; offspring whose mothers were stressed during

pregnancy show greater hypothalamic-pituitary-adrenal axis response to stressors than control offspring.

Epidemiology and Etiology The prevalence of Cushing's disease is approximately 40 per million, and it occurs far more frequently in women, particularly in the third and fourth decades of life. The most common cause of spontaneous Cushing's syndrome is an ACTH-producing pituitary adenoma. Other causes are autonomous ACTH-producing tumors such as oat cell carcinoma of the lung, islet cell pancreatic tumors, or thymomas. An adrenal adenoma or adrenocortical carcinoma and nodular adrenal hyperplasia are ACTH-independent causes of increased cortisol concentrations.

The relation of stressful life events to the onset of Cushing's syndrome is controversial; some studies find a greater frequency of adverse life events prior to the development of the disorder, while others find no such association. Similarly, the relation between depressive symptoms and the onset of Cushing's syndrome remains inconclusive; some studies report a high frequency of depression preceding the physical manifestations of the illness, while others find no greater rate of depressive symptoms than noted in control groups.

Cognitive impairment in hypercortisolemic patients is believed to result from the effect of corticosteroids on hippocampal formation. Among other tasks, the hippocampus is involved in selective attention, learning, and memory. Hippocampal damage can result from prolonged exposure to high corticosteroid concentrations, and higher plasma cortisol concentrations are associated with smaller hippocampal volume. Reduced hippocampal volume is in turn associated with lower scores on memory tests. In electrophysiological studies, the response to relevant (but not to irrelevant) stimuli is reduced following short-term cortisol administration, which suggests that cortisol impairs selective attention and discrimination of important information. Several researchers have suggested that the cognitive impairments in patients with clinical depression (many of whom have hypothalamic-pituitary-adrenal axis dysregulation) may result in part from cortisol effects on the hippocampus.

The exogenous use of steroids may also produce psychiatric sequelae. The likelihood of psychiatric adverse effects increases in direct proportion to prednisone dosage. Dosages under 40 mg a day are associated with an incidence of psychiatric symptoms below 2 percent, but this rate rises to almost 20 percent when the dosage exceeds 80 mg a day.

Diagnosis and Clinical Features The clinical features of Cushing's disease include a characteristic “moonfacies” or rounded face from accumulation of adipose tissue around the zygomatic arch (Fig. 25.6-4). Truncal obesity, a “buffalo hump” appearance, results from cervicodorsal adipose tissue deposition. The catabolic effects of cortisol on protein produce muscle wasting, slow wound healing, easy bruising, and thinning of the skin leading to abdominal striae. Bones become osteoporotic, sometimes resulting in pathological fractures and loss of height (Fig. 25.6-5). Diabetes may occur in some patients as a result of the gluconeogenesis stimulated by the high glucocorticoids concentrations. Chronic ACTH stimulation of the adrenal glands may lead to excess androgen production, with secondary hirsutism, acne, and amenorrhea. Hyperpigmentation may also occur, as ACTH is a precursor of a-melanocyte-stimulating hormone, which produces a skin-darkening effect. Although cortisol's mineralocorticoid effects are mild, at high concentrations it produces sodium retention, potassium loss, metabolic alkalosis, and edema. Hypertension occurs in most patients.

FIGURE 25.6-4 Cushing's syndrome. Plethoric “moonface.” (Reprinted with permission from Douthwaite AH, editor: *French's Index of Differential Diagnosis*, ed 7. Williams & Wilkins, Baltimore, 1954.)

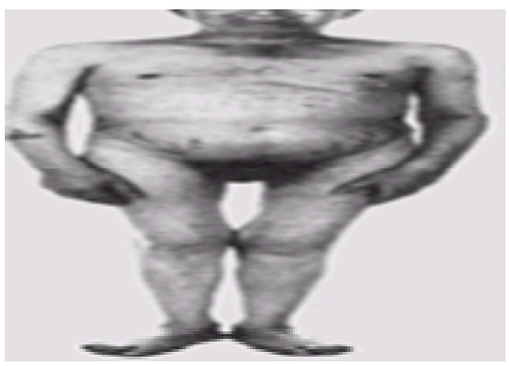

FIGURE 25.6-5 Cushing's syndrome. Legs thin owing to atrophy of thigh muscles. Some abdominal obesity with marked striae. (Reprinted with permission from Douthwaite AH, editor: *French's Index of Differential Diagnosis*, ed 7. Williams & Wilkins, Baltimore, 1954.)

Psychiatric symptoms are myriad. Most patients experience fatigue, and approximately 75 percent report depressed mood. Of these, approximately 60 percent experience moderate or severe depression. The severity of depression does not appear to be influenced by the underlying cause of the Cushing's syndrome. Depressive symptoms occur more commonly in female patients than in male patients with Cushing's syndrome.

Emotional lability, irritability, decreased libido, anxiety, and hypersensitivity to stimuli are common. Somatic symptoms and elevated euroticism scores on the Eysenck Personality Inventory have also been reported, with significant improvement following normalization of cortisol concentrations. Patients may develop social withdrawal as a result of shame regarding their physical appearance. Paranoia, hallucinations, and depersonalization are estimated to occur in 5 to 15 percent of cases. Cognitive changes are common, and approximately 83 percent of patients experience deficits in concentration and memory. The severity of these deficits correlates with plasma cortisol and ACTH concentrations.

Manic and psychotic symptoms occur much less frequently than depression, in approximately 3 to 8 percent of patients, but reach 40 percent in patients with adrenal carcinomas. In iatrogenic hypercortisolism and adrenal carcinomas, however, mania and psychosis may predominate. The psychiatric disturbances in prednisone-treated patients tend to appear within the first 2 weeks of treatment and occur more commonly in women than men.

> T. S. was a 40-year-old white woman who was diagnosed with membranoproliferative glomerulonephritis and began being treated with prednisone 20 mg a day. Within 10 days of beginning the steroid treatment, her mood became elevated, her speech was pressured, her sleep diminished from 8 to 6 hours a night, and her activity level was heightened. She reported that her house “has never been so clean!” Within 2 weeks of discontinuing the prednisone, her mental state returned to baseline.

Steroids withdrawal can also produce psychiatric disturbances, particularly depression, weakness, anorexia, and arthralgia. Other steroid-induced withdrawal symptoms include emotional lability, memory impairment, and delirium. Withdrawal symptoms can persist for up to 8 weeks following corticosteroid withdrawal.

Patients presenting with mood lability or depression in association with muscle weakness, obesity, diabetes, easy bruising, cutaneous striae, acne, hypertension, and

(in women) hirsutism and oligomenorrhea or amenorrhea benefit from an endocrinological evaluation.

Pathology and Laboratory Examination In Cushing's syndrome, plasma and urine concentration of cortisol increase, and normal circadian rhythms of cortisol secretion are blunted or absent. Administration of metyrapone, which blocks cortisol synthesis, is normally followed by increased cortisol production. No such increase occurs in cases of adrenal or ectopic origin; in cases of pituitary origin, postmetyrapone response is augmented. Another test that distinguishes pituitary from adrenal cause of Cushing's syndrome is the dexamethasone (Decadron) suppression test (DST). Administration of 1 mg of the synthetic steroid dexamethasone does not suppress cortisol in patients with Cushing's syndrome of adrenal origin or other nonpituitary origin but does suppress cortisol in pituitary Cushing's syndrome.

The overnight DST is considered the procedure of choice for screening for Cushing's syndrome. It involves administering 1 mg dexamethasone at midnight and determining the plasma cortisol concentration the following morning. Plasma cortisol concentrations above 200 nmol/L indicate a high likelihood of Cushing's syndrome. The overnight 1-mg DST is simple to administer and has greater specificity and sensitivity than other screening procedures such as measurements of urinary free cortisol. Because plasma cortisol concentrations vary considerably during the day, an isolated measurement of plasma cortisol is not adequate to screen for Cushing's syndrome.

In the early 1980s the DST was a popular neuroendocrine challenge test in psychiatry, because a positive result (i.e., nonsuppression of cortisol following dexamethasone administration) was believed to be a useful marker for melancholic depression. However, the DST is now used infrequently because its clinical use is limited by low specificity and sensitivity. Additionally, medical illnesses, weight loss, stress, and concomitant medications confound the DST results.

Differential Diagnosis Patients with Cushing's syndrome commonly have mood changes very similar to those of patients with a major depressive episode. Both groups of patients have a shortened rapid eye movement (REM) latency, increased cortisol concentration, and abnormal findings on DSTs. The physical manifestations of Cushing's syndrome help distinguish between the two diagnoses, but they may be absent early in the course of the illness. Other distinguishing symptoms of Cushing's syndrome are greater irritability and mood lability; greater tendency to feel best in the morning (in contrast to patients with major depression, who tend to feel worse then); and less guilt and hopelessness than patients with major depression.

Treatment and Course The treatment of pituitary ACTH-producing tumors involves surgical resection or pituitary irradiation. Adrenal adenomas and carcinomas are removed surgically, and in the case of carcinomas, chemotherapy is instituted subsequently. Medications that antagonize cortisol production (e.g., metyrapone or mitotane [tysobren]) or suppress ACTH (e.g., serotonin antagonists such as cyproheptadine [Periactin]) are sometimes used but have met with limited success.

Following treatment, hypertension and electrolyte imbalances correct quickly. The physical stigmata and subtle cognitive deficits may take months or years to resolve. Following successful treatment of Cushing's syndrome, most patients experience significant improvement in depressive and anxiety symptoms.

For patients undergoing prednisone treatment, lithium and neuroleptic medications during the course of treatment help prevent the development of manic or psychotic symptoms. Steroid psychosis resolves following discontinuation of the medication. Prednisone should be tapered rather than discontinued abruptly, to allow the pituitary gland, which has been suppressed, to resume its production of ACTH.

Adrenocortical Insufficiency

Definition and Comparative Nosology Adrenocortical insufficiency results from inadequate production of three major steroid hormones, glucocorticoids, mineralocorticoids, and sex steroids, by the adrenal gland. In *primary adrenal insufficiency,* adrenal hypofunction results from adrenal gland disease (e.g., autoimmune disease, infection, idiopathic atrophy, metastatic tumor). ACTH and CRH concentrations increase in response to low concentrations of adrenal steroids. Because ACTH is metabolized to a-melanocyte-stimulating hormone, which promotes melanocyte activity, the skin may become hyperpigmented as ACTH concentrations increase.

Secondary adrenal insufficiency results from deficient ACTH secretion caused by pituitary disease, and *tertiary adrenal insufficiency* refers to deficient hypothalamic secretion of CRH. Adrenal aldosterone production tends to be less affected by secondary or tertiary forms of adrenal insufficiency.

Epidemiology The prevalence of Addison's disease is estimated at 110 per million population, with an incidence of 5.6 per million per year.

Etiology Autoimmune destruction of the adrenal cortex produces over 90 percent of cases of Addison's disease. Secondary and tertiary adrenal insufficiency result primarily from withdrawal of exogenous corticosteroids or ACTH, as their administration suppresses the hypothalamic-pituitary-adrenal axis. Less common causes include pituitary tumors, trauma, infectious disease (e.g., human immunodeficiency virus [HIV]), infarction, and surgical ablation.

Diagnosis and Clinical Features Clinical symptoms usually do not become apparent until more than 90 percent of the adrenal cortex is damaged. Adrenal insufficiency can be life threatening, particularly in cases of acute adrenal insufficiency (e.g., following adrenal hemorrhage from infectious processes or use of anticoagulants). Symptoms include weakness, hypoglycemia, hyponatremia, hyperkalemia, nausea, diarrhea, fever, psychotic symptoms, and shock.

Chronic adrenal insufficiency produces more-subtle symptoms, including fatigability, salt craving, weight loss, vitiligo, nausea, hyperpigmentation, loss of body hair, muscle cramps, apathy, irritability, and mild-to-severe depression. Symptoms develop insidiously over months or years. A diagnosis of adrenocortical insufficiency is made primarily on the basis of plasma cortisol concentrations during the ACTH stimulation test. Following intravenous administration of 250 µg of cosyntropin (Cortrosyn), cortisol concentrations normally peak in 30 to 90 minutes. A normal test result is defined as a baseline cortisol concentration of at least 5 µg per 100 mL, a rise between baseline and stimulated cortisol concentrations of more than 7 µg per 100 mL, and a stimulated cortisol concentration above 18 µg per 100 mL. Low cortisol responses to this test indicate adrenal insufficiency.

Pathology and Laboratory Examination Common laboratory findings include low serum concentrations of sodium, high concentrations of potassium, and low or normal plasma cortisol concentrations. The diagnosis of adrenal insufficiency is particularly likely if plasma cortisol concentrations in the early morning (when they should be peaking) are low.

In cases of primary adrenal insufficiency, ACTH concentration rises and cortisol concentration does not rise following ACTH stimulation. In cases of adrenal insufficiency resulting from pituitary disease, ACTH concentrations are low or normal, and the response to CRH is blunted. If the adrenal insufficiency results from hypothalamic disease, the ACTH response to CRH administration is adequate but cortisol concentrations may fail to rise because of adrenal atrophy following prolonged ACTH deficiency.

Differential Diagnosis The symptoms of chronic adrenal insufficiency overlap those of depression. Easy fatigability, muscle cramps, anorexia, and lethargy may also lead to a diagnosis of chronic fatigue syndrome. In cases of acute adrenal insufficiency (e.g., following abrupt discontinuation of steroid administration), patients may develop a psychotic or delirious syndrome.

Course and Prognosis Physical and mental symptoms respond quickly to steroid replacement. A patient's prognosis depends largely on the underlying cause of the adrenal dysfunction. In cases of chronic adrenal insufficiency, patients have an unremitting illness requiring lifelong steroid treatment.

Treatment Acute adrenal insufficiency requires immediate treatment with intravenous hydrocortisone in addition to fluid replacement with saline solution and potassium supplementation. In cases of primary adrenal insufficiency, a mineralocorticoid (e.g., fludrocortisone [Florinef]) is also required. In cases of chronic adrenal insufficiency, prednisone or hydrocortisone is administered orally as maintenance treatment. Patients should be encouraged to carry a card or bracelet listing their diagnosis, medications, dosages, and physician.

HYPERPROLACTINEMIA

Definition and Comparative Nosology Prolactin, produced by the anterior pituitary, stimulates milk production from the breast and modulates maternal behavior. Its production is inhibited by dopamine (also known as prolactin-inhibiting factor) produced by the tuberoinfundibular neurons of the arcuate nucleus of the hypothalamus. Normal concentrations (5 to 25 ng/mL in women and 5 to 15 ng/mL in men) fluctuate during the day, peaking during sleep. Exercise and emotional

stress can increase prolactin concentration.

Epidemiology and Etiology Hyperprolactinemia most frequently results from prolactin-secreting pituitary adenomas (prolactinomas) and idiopathic hyperprolactinemia, two conditions that occur more commonly in women. Medications that block dopamine action (e.g., antipsychotics, methyldopa [Aldomet], reserpine [Diupres]) raise prolactin concentrations up to 20 times. Neuroleptic medications raise prolactin concentrations up to 20-fold in a dose-dependent fashion, with prolactin concentrations reaching plateau. All antipsychotics appear equally likely to raise prolactin concentrations, with the exception of clozapine (Clozaril) and olanzapine (Zyprexa). Other medications that may increase prolactin concentrations include oral contraceptives, estrogens, tricyclic drugs, serotonergic antidepressants, and propranolol. Hypothyroidism raises prolactin concentration because TRH stimulates prolactin release. Acromegaly is also associated with hyperprolactinemia, as a result of the lactogenic properties of GH. Hypothalamic disease (e.g., craniopharyngiomas), may also result in an elevation of prolactin levels, as the hypothalamus exerts an inhibitory influence on prolactin secretion. Physiological hyperprolactinemia occurs in pregnant and breast-feeding women; nipple stimulation also increases prolactin concentrations.

Traumatic childhood experiences such as separation from parents or living with an alcoholic father have been reported to produce predisposition to hyperprolactinemia. Stressful life events are also associated with galactorrhea, even in the absence of increased prolactin concentrations.

Diagnosis, Clinical Features, and Laboratory Examination Patients with high prolactin concentrations may be asymptomatic, although symptoms usually develop when prolactin concentrations reach 60 to 100 ng/mL. Women may experience amenorrhea because prolactin suppresses normal menstrual cycling, and men may develop gynecomastia. Both women and men may experience low libido, decreased fertility, and galactorrhea (inappropriate lactation), although the latter occurs primarily in women. Headache and visual impairment may result in cases of pituitary adenomas.

Women with pathological hyperprolactinemia have been reported to experience higher levels of depression, anxiety, and hostility than controls, which lessen with normalization of prolactin concentration.

Treatment A laboratory workup, the initial step in treatment, should include thyroid function tests, a pregnancy test, and a chemistry panel. A brain imaging study, either magnetic resonance imaging (MRI) with gadolinium or computed tomography (CT) with contrast, should be done.

Bromocriptine (Parlodel), a dopamine agonist, reduces the synthesis and secretion of prolactin. In approximately 80 percent of cases, bromocriptine at doses of 2.5 to 5 mg twice or thrice daily is effective treatment. However, many patients cannot tolerate the adverse effects of bromocriptine, which can include severe nausea and dizziness. Alternative medications include quinagolide, pergolide (Permax), and cabergoline (Dostinex). Medication use must continue indefinitely, as medication discontinuation usually leads to a relapse of hyperprolactinemia.

Radiotherapy is another treatment strategy for patients who are refractory to medication treatment. Surgical removal through transphenoidal surgery is used far less often now than 10 years ago, because of the high rate of recurrence of prolactinomas.

ACROMEGALY AND GIGANTISM

GH, produced by the anterior pituitary, stimulates protein synthesis, lypolysis, and growth of skeletal cartilage. Its production is stimulated by GH-releasing factor and inhibited by somatostatin, both of hypothalamic origin. If GH concentrations are excessively high in childhood, gigantism may develop. Acromegaly results from an excess of GH in adulthood, after fusion of the epiphyses of the long bones, so that height is not affected. The prevalence of acromegaly is approximately 4 in 100,000. Acromegaly most frequently results from a pituitary tumor (Fig. 25.6-6). Other causes include ectopic production or excessive hypothalamic production of growth hormone releasing factor. The longer the illness, the more extreme the symptoms. The nose, jaw, tongue, and soft tissues of the hands and feet become enlarged, as do the heart, liver, and kidneys. Organ enlargement can lead to congestive heart failure. No specific psychiatric symptoms have been consistently associated with acromegaly or gigantism, or with elevated GH levels. Adjustment disorder may occur from changes in physical appearance and from living with a chronic illness. Laboratory studies show high GH concentrations. The most reliable test for acromegaly is measurement of GH following an oral glucose tolerance test. In acromegalic patients, GH secretion does not show the normal suppression by glucose. Brain imaging studies may show a pituitary tumor. Treatment involves pituitary ablation through surgery or radiation. Dopamine agonists such as bromocriptine and apomorphine, which normally increase GH levels, have the opposite effect in patients with acromegaly. The serotonin antagonist cyproheptadine may also reduce GH levels. A historical case of gigantism is illustrated in Figure 25.6-7.

FIGURE 25.6-6 A. Before onset of acromegaly. **B**. Acromegaly: enlargement of the mandible, nose, and lips is obvious. (Reprinted with permission from Spillane JD, Spillane JA: *An Atlas of Clinical Neurology,* ed 3. Oxford University Press, New York, 1982.)

FIGURE 25.6-7 A case of simple (primary) gigantism. The Austrian giant, Winkelmeyer, 7 ft. 6 in. tall. (Reprinted with permission from Douthwaite AH, editor: *French's Index of Differential Diagnosis*, ed 7. Williams & Wilkins, Baltimore, 1954.)

POLYCYSTIC OVARY SYNDROME

Polycystic ovary syndrome is one of the most widespread endocrine disorders in women. Clinical features include hirsutism, obesity, male-pattern alopecia, acne, and irregular menstrual cycles. Ultrasonography shows enlarged ovaries with multiple cystic follicles. Normal-appearing ovaries, however, do not rule out the condition. LH concentrations are usually high while FSH concentrations are low or normal. Other laboratory findings include elevated concentrations of androstenedione, testosterone, estrone, and prolactin. Because of the absence of ovulation, estrogen is unopposed by progesterone, placing patients at risk for endometrial hyperplasia and carcinoma. Another long-term risk is type II diabetes, which occurs seven times as frequently as in the reference population, secondary to insulin resistance associated with hyperandrogenism. An association between depression and hyperandrogenism in women has been noted in several reports. Initiation of oral

contraceptive treatment has been reported to improve mood, possibly by the reduction in androgen concentration that results from suppression of ovarian activity.

Prevalence rates vary depending on the criteria used to make the diagnosis but are estimated to be approximately 8 percent of women of reproductive age. Rates reach 30 to 75 percent when they are based on women presenting for treatment of menstrual irregularities or hirsutism.

The pathogenesis of polycystic ovary syndrome is not clear. Dysregulation of the rate-limiting enzyme involved in androgen biosynthesis in the ovaries has been implicated. Polycystic ovary syndrome appears to follow an autosomal dominant mode of inheritance. Emotional stress in puberty has been reported as a risk factor for the development of polycystic ovary syndrome.

The differential diagnosis includes pituitary and adrenal disorders such as hyperprolactinemia, acromegaly, and congenital adrenal hyperplasia, which can produce menstrual irregularities and hirsutism. These can be distinguished from polycystic ovary syndrome by laboratory and clinical assessments.

Treatments include antiestrogens such as clomiphene (Clomid), gonadotropins, laparoscopic surgery, and low-dose oral contraceptives. Weight reduction is also recommended. Hirsutism may respond to spironolactone, an androgen receptor antagonist, but it may be difficult to tolerate as it commonly produces mood changes, decreased libido, and fatigue.

ANDROGEN INSENSITIVITY SYNDROME

Androgen insensitivity syndrome, or testicular feminization, is a rare X-linked recessive disorder that produces a variable appearance of external genitalia depending on the degree of androgen receptor defectiveness. In cases of complete androgen insensitivity, patients with a 46 XY karyotype are phenotypically female. Body fat and hair distribution are typically female, and breasts develop after puberty. The vagina, however, is short and blind or may be absent altogether (Fig. 25.6-8). An inguinal hernia is a common clinical presentation in prepubertal children, and the diagnosis may not be made until surgical repair of the hernia reveals a testis. Postpubertally, primary amenorrhea is a common presentation. With the exception of the undescended testes, internal genitalia are absent. Surgical removal of the testes is recommended, as the development of gonadal tumors is a concern. The psychosexual development is feminine, and it is therefore recommended that these individuals be raised as females. In patients with incomplete testicular feminization, the appearance of external genitalia is more virilized and ambiguous. The degree of virilization varies, depending on the extent of androgen insensitivity. Patients who have been raised as males and who have undergone genital reconstructive surgery have generally experienced poor anatomical and functional results. In most cases, it is recommended that patients be raised as females, and that gonadectomy be performed before puberty to avert further virilization. Careful evaluation of infants and children with ambiguous genitalia is important for early diagnosis and gonadectomy.

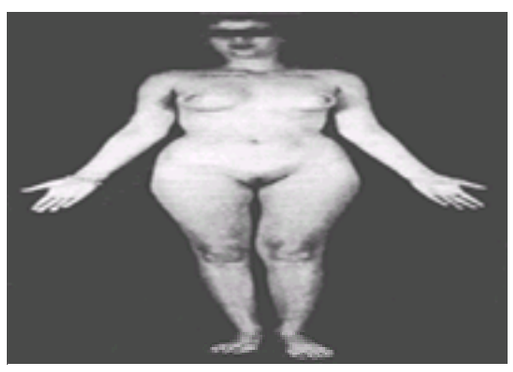

FIGURE 25.6-8 A phenotypic female with abdominal testes and an XY chromosomal karyotype. Note the excellent breast development and the absence of pubic hair. A normal blind vagina was present without clitoral enlargement. (Courtesy of R.B. Greenblatt, M.D., and V.P. McNamara, M.D.)

SUGGESTED CROSS-REFERENCES

Related basic information about the biology of brain function can be found in Section 1.2 on functional neuroanatomy, in Section 1.6 on neuropeptides, and in Section 1.11 on psychoneuroendocrinology. Reproductive medicine is discussed in Section 28.2, postpartum psychiatric syndromes in Section 13.4, and premenstrual dysphoric disorder in Section 28.3. The classification of mental disorders is presented in Chapter 9. Delirium, dementia, and amnestic and other cognitive disorders, as well as mental disorders due to a general medical condition, are presented in Chapter 10. Schizophrenia is discussed in Chapter 12, other psychotic disorders in Chapter 13, mood disorders in Chapter 14, and anxiety disorders in Chapter 15. Intersexual conditions are discussed in Section 19.3 on gender identity disorders.

SECTION REFERENCES

Assies J, Vingerhoets JJM, Poppelaars K: Psychosocial aspects of hyperprolactinemia. Psychoneuroendocrinology *17:*673, 1992.

Backer JR: Autoimmune endocrine disease. JAMA *278:*1931, 1997.

Bauer MS, Whybrow PC: The effect of changing thyroid function on cyclic affective illness in a human subject. Am J Psychiatry *143:*633, 1986.

Bocchetta A, Bernardi F, Burrai C, Pedditzi M, Loviselli A, Velluzi F, Martino E, Del Zompo M: The course of thyroid abnormalities during lithium treatment: A two-year follow-up study. Acta Psychiatr Scand *86:*38, 1992.

Brosnan CM, Gowing NF: Addison's disease. Br Med J *312:*1085, 1996.

*Brown ES, Suppes T: Mood symptoms during corticosteroid therapy: A review. Harvard Rev Psychiatry *5:*239, 1998.

Brown TR: Human androgen insensitivity syndrome. J Androl *16:*299, 1995.

Cleare AJ, McGregor A, O'Kean V: Neuroendocrine evidence for an association between hypothyroidism, reduced central 5HT activity, and depression. Clin Endocrinol *43:*713, 1995.

Franklyn JA: The management of hyperthyroidism. N Engl J Med *330:*1731, 1994.

Franks S: Polycystic ovary syndrome. N Engl J Med *333:*853, 1995.

Haggerty JJ, Prange AJ: Borderline hypothyroidism and depression. Annu Rev Med *46:*37, 1995.

Haggerty JJ Jr, Silva SG, Marquardt M, Mason GA, Chang HY, Evans DL, Golder RN, Pedersen C: Prevalence of antithyroid antibodies in mood disorders. Depress Anxiety *5:*91, 1997.

*Hauser ST, Jacobson AM, Benes KA, Anderson BJ: Psychosocial aspects of diabetes mellitus in children and adolescents: Implications and interventions. In *Handbook of Child and Adolescent Psychiatry,* JD Noshpitz, editor. Wiley, New York, 1997.

Henry C, Kabbaj M, Simon H, Le Moal M, Maccari S: Prenatal stress increases the hypothalamo-pituitary-adrenal axis response in young and adult rats. J Neuroendocrinol *6:*341, 1994.

*Kelly WF, Kelly MJ, Faragher B: A prospective study of psychiatric and psychological aspects of Cushing's syndrome. Clin Endocrinol *45:*715, 1996.

Kong MF, Jeffcoate W: Eighty-six cases of Addison's disease. Clin Endocrinol *41:*757, 1994.

Jacobson AM: The psychological care of patients with insulin-dependent diabetes mellitus. N Engl J Med *334:*1249, 1996.

Joffe RR, Levitt A: Major depression and subclinical (grade 2) hypothyroidism. Psychoneuroendocrinology *17:*215, 1992.

Langer M, Fiegl J, Riegel V, Prohaska R, Kubista E, Ringler M: Psychosomatic aspects of hyperprolactinemia. Arch Gynecol Obstet *248:*167, 1991.

*Lasser RA, Baldessarini RJ: Thyroid hormones in depressive disorders: A reappraisal of clinical utility. Harvard Rev Psychiatry *4:*291, 1997.

Levin TR, Terrell TR, Stoudemire A: Organic mood disorder associated with the HAIR-AN syndrome. J Neuropsychiatry Clin Neurosci *4:*51, 1992.

Patterson MN, McPhaul MJ, Hughes IA: Androgen insensitivity syndrome. Baillieres Clin Endocrinol Metab *8:*379, 1994.

Serri O: Progress in the management of hyperprolactinemia. N Engl J Med *331:*942, 1994.

Shah R, Woolley MM, Costin G: Testicular feminization: The androgen insensitivity syndrome. J Pediatr Surg *27:*757, 1992.

*Singer PA, Cooper DS, Levy EG, Ladenson PW, Braverman LE, Daniels F, Greenspan F, McDougall R, Nikolai T: Treatment guidelines for patients with hyperthyroidism and hypothyroidism. JAMA *273:*808, 1995.

Sobrinho LG: The psychogenic effects of prolactin. Acta Endocrinol Suppl *1:*38, 1993.

Starkman MN: The HPA axis and psychopathology: Cushing's syndrome. Psychiatr Ann *23:*691, 1993.

Surks MI, Ocampo E: Subclinical thyroid disease. Am J Med *100:*217, 1996.

Surks MI, Sievert R: Drugs and thyroid function. N Engl J Med *333:*1688, 1994.

Whybrow P: Sex differences in thyroid axis function: Relevance to affective disorder and its treatment. Depression *3:*33, 1995.

Wiersinga WM: Subclinical hypothyroidism and hyperthyroidism. I. Prevalence and clinical relevance. Neth J Med *46:*197, 1995.

*Wolkowitz OM: Prospective controlled studies of the behavioral and biological effects of exogenous corticosteroids. Psychoneuroendocrinology *3:*233, 1994.

25.7 PSYCHOCUTANEOUS DISORDERS

LESLEY M. ARNOLD, M.D.

Psychocutaneous disorders include dermatological diseases affected by the presence of psychiatric symptoms or stress and psychiatric illnesses in which the skin is the target of disordered thinking, behavior, or perception. Evidence of the nervous system's involvement in skin pathophysiology provides possible links between stress or psychiatric disturbances and many dermatological diseases.

CLASSIFICATION

The psychocutaneous disorders discussed in this section are classified in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) (Table 25.7-1). Psychological factors affecting medical condition include dermatological diseases in which comorbid psychiatric disorders or symptoms or stress affect their course. The dermatological disorders listed are most commonly affected by psychiatric factors, but many other dermatological diseases (e.g., genital herpes) may also be influenced by these factors. Somatoform disorders contain dermatological problems that are not fully explained by a known dermatological disease. Delusional disorder, somatic type, is the most likely diagnosis for dermatological patients with delusions of parasitosis. The differential diagnosis of delusions of parasitosis includes major depressive disorder, bipolar I disorder, schizophrenia, psychotic disorder due to a general medical condition, and substance-induced psychotic disorder. Delusional patients must also be differentiated from patients with nondelusional obsessive concerns about infestation (obsessive-compulsive disorder). A preoccupation with a defect in appearance (body dysmorphic disorder) has a delusional variant, classified as a delusional disorder, somatic type. Trichotillomania is classified under the impulse-control disorders. Psychogenic excoriations and onychophagia would probably be included under the category of impulse control disorder not otherwise specified, but like trichotillomania, these disorders frequently have symptoms that overlap with obsessive-compulsive disorder. Many propose to include all of the impulse-control disorders under the broad category of obsessive-compulsive spectrum disorders, although this is not part of the DSM-IV classification. Finally, factitious disorders include factitious dermatitis (dermatitis artefacta) and psychogenic purpura. Some cases of psychogenic purpura are thought to result from a conversion disorder but further confirmation is needed.

I. Psychological factors affecting medical condition
Atopic dermatitis
Psoriasis
Alopecia areata
Urticaria and angioedema
Acne vulgaris
Rosacea
Telogen defluvium
Primary hyperhidrosis
Seborrheic dermatitis
Prurigo nodularis
II. Undifferentiated somatoform disorder
Chronic idiopathic pruritis
Idiopathic pruritis ani, vulvae, and scroti
III. Pain disorder
Idiopathic glossodynia
Dysesthetic (essential) vulvodynia
IV. Other somatoform disorder
Body dysmorphic disorder
V. Delusional disorder, somatic type
Delusions of parasitosis
Delusions of a defect in appearance
Delusions of a foul body odor
VI. Impulse-control disorder (obsessive-compulsive spectrum disorders)
Trichotillomania
Psychogenic excoriation
Onychophagia
VII. Factitious disorder
Factitious dermatitis
Psychogenic purpura

Table 25.7-1 DSM-IV Classification of Psychocutaneous Disorders

PSYCHOLOGICAL FACTORS AFFECTING MEDICAL CONDITION

Atopic Dermatitis Atopic dermatitis (also called atopic eczema or neurodermatitis) is a chronic skin disorder characterized by pruritus and inflammation (eczema), which often begins as an erythematous, pruritic, maculopapular eruption (Fig. 25.7-1). Scratching in response to the pruritus can lead to lichenification, excoriations, and infections. Atopic dermatitis arises most commonly during early infancy, childhood, or adolescence and is frequently associated with a personal or family history of atopic dermatitis, allergic rhinitis, or asthma. A national census of dermatological disease in the United States in 1977 found atopic dermatitis to be a common disorder, affecting 7 per 1000 individuals and 24 per 1000 children, with a female to male ratio of 1.2 to 1.0. The prevalence of atopic dermatitis has increased to more than 10 percent in the past decade, possibly because of greater exposure to provocative factors such as outdoor pollution, reagents in highly insulated buildings, house mites, food additives, and increased parental and physician awareness of the disorder. Atopic dermatitis can resolve spontaneously, especially in mild cases, but most patients experience persistent or relapsing symptoms. A 20-year follow-up study of 47 patients with a diagnosis of atopic dermatitis by the age of 2 found that 72 percent continued to have signs of the disorder. In another study, symptoms persisted or recurred in 77 to 91 percent of adult patients who had moderate or severe symptoms as teens. In general, comorbid asthma or hayfever, late onset, severe dermatitis and a family history of atopic dermatitis correlate positively with disease chronicity.

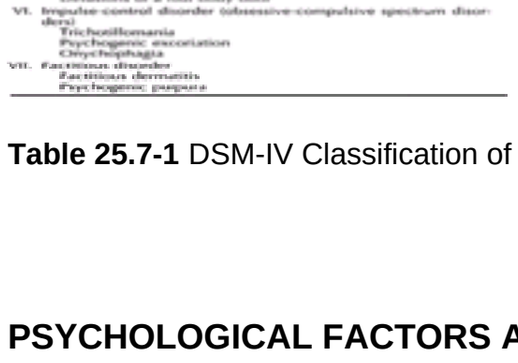

FIGURE 25.7-1 Atopic dermatitis. Note the maculopapular eruption behind the knees. (Courtesy of D.F. Mutasim.)

The cause of atopic dermatitis is unknown, although recent studies suggest a genetic influence. Concordance rates were higher in monozygotic (0.75) than dizygotic (0.25) twins, and the prevalence of atopic dermatitis in children with two affected parents was 81 percent versus 56 percent for those with a single affected parent. Heritable traits identified so far appear to affect the immune system, which exhibits an imbalance of immunoregulatory T cells that may explain the defective cell-mediated immunity and increased immunoglobulin E (IgE) production seen in atopic dermatitis. Another important feature of atopic dermatitis is the abnormal mediator release by mast cells and basophils. Increased release of histamine contributes to the characteristically severe pruritis. Growing evidence for the role of neuropeptides such as substance P, a neuropeptide released by cutaneous nerves that can cause histamine secretion from mast cells, provides a possible link

between the central nervous system (CNS) and atopic dermatitis.

Although genetic factors probably underlie the development of atopic dermatitis, environmental factors frequently trigger or exacerbate the disease. Affected patients seem to have a lower response threshold and more prolonged reaction to pruritic stimuli than controls. Environmental triggers include food allergy or intolerance, contact irritants and allergens, aeroallergens (house dust mites, pollen, molds, human and animal dander), microbes, hormones, climate, sweating, and stress. A vicious circle of itching, scratching, and lesion aggravation frequently develops and contributes to symptom chronicity. Stressful life events often precede the onset and exacerbation of atopic dermatitis. How stress affects this disorder is unclear, but it may involve the relationship between stress, neuropeptides, and immune function.

Methodological problems have limited many studies examining the role of psychiatric factors in atopic dermatitis that infer causal relationships when only an association or correlation was found in cross-section. However, well-controlled studies have found adult atopic dermatitis patients to be more anxious and depressed than clinical and disease-free control groups. Anxiety or depression exacerbates atopic dermatitis by eliciting scratching behavior. In another study of pruritus associated with atopic dermatitis and other dermatological conditions, depressive symptoms appeared to amplify the itch perception. Recent studies of children with atopic dermatitis found most to be well adjusted psychiatrically with secure maternal attachments, compared with normal controls. About a third of the children with severe atopic dermatitis symptoms had significantly higher morbidity levels on behavioral screening questionnaires. The emotional state in children with atopic dermatitis was closely related to the severity of the illness. Certain dimensions of family environment such as independence and organization correlated with less severe symptoms of atopic dermatitis, while parental responses of attention or physical contact reinforced scratching.

Different modalities of psychiatric treatment for atopic dermatitis exist. Some strive to reduce stress and interrupt the vicious circle of itching and scratching. Controlled studies have established that relaxation training, habit reversal training, cognitive-behavioral techniques, and stress management training led to significant and stable adjunctive treatment responses beyond those of standard medical care as well as reducing anxiety and depression. Controlled trials of psychotropic drug treatment found that topical doxepin (Sinequan) cream was effective in reducing pruritus in atopic dermatitis patients, probably related to doxepin's potent histamine antagonism. Trimipramine (Surmontil), another antidepressant with histamine receptor antagonism, decreased the fragmentation of sleep and reduced the time spent in stage 1 sleep, which secondarily reduced the amount of scratching during the night in atopic dermatitis patients.

Psoriasis Psoriasis is a chronic, relapsing disease of the skin with variable clinical features. Characteristic lesions involve both the vasculature and the epidermis and have clear-cut borders and noncoherent silvery scales with a glossy, homogeneous erythema under the scales (Fig. 25.7-2). Some patients also develop nail dystrophy and arthritis. It affects 1 to 2 percent of the United States general population and is equally common in women and men, with most developing initial lesions in the third decade of life. Early onset (before age 15) predicts greater disease severity with a higher percentage of body surface involvement and a worse treatment response. A family history of psoriasis is more common in early-onset patients. For most patients psoriasis persists throughout life, with unpredictable exacerbations. Spontaneous remissions have been reported and have lasted between 1 and 54 years.

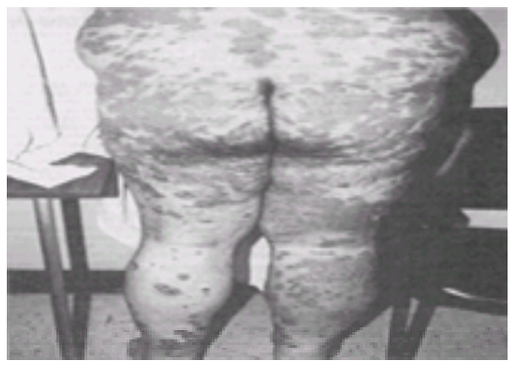

FIGURE 25.7-2 Psoriasis. The characteristic lesions have clear-cut borders and silvery scales. (Courtesy of D.F. Mutasim.)

The pathogenesis of psoriasis is not completely understood, but it appears that several systems involved in skin repair, inflammatory defense mechanisms, and immunity may contribute to its development. A genetic predisposition to the disease is indicated by a higher incidence of psoriasis in relatives of patients and higher rates of monozygotic (0.65) than dizygotic (0.30) concordance as well as correlations with the human major histocompatibility gene complex (HLA) and non-HLA loci. However, external factors frequently affect the course of psoriasis. Common triggers of psoriasis include cold weather, physical trauma, acute bacterial and viral infections, and drug-related effects associated with corticosteroid withdrawal and with the use of b-adrenergic receptor antagonists and lithium. Lithium-induced psoriasis typically occurs within the first few years of treatment, is resistant to treatment, and resolves after discontinuation of lithium treatment.

The adverse effect of psoriasis on the quality of life can lead to stress that can in turn trigger more psoriasis. In a recent survey of psoriatic patients, 46 percent reported daily problems secondary to psoriasis. The 40 to 80 percent of patients who reported that stress triggered psoriasis often described disease-related stress, resulting mainly from the cosmetic disfigurement and social stigma of psoriasis, rather than stressful major life events. Psoriasis-related stress may have more to do with psychosocial difficulties inherent in the interpersonal relationships of patients with psoriasis than with the severity or chronicity of psoriasis activity. The mechanism of stress-induced exacerbations is unknown but may involve the nervous, endocrine, and immune systems in such a way that descending autonomic information from the CNS is transmitted to sensory nerves in the skin, resulting in release of neuropeptides such as substance P into the skin. These neuropeptides help initiate and maintain the inflammatory response in psoriatic lesions.

Controlled studies have found psoriatic patients to have high levels of anxiety and depression and significant comorbidity with a wide array of personality disorders from the revised third edition of DSM (DSM-III-R), including schizoid, avoidant, passive-aggressive, and obsessive-compulsive personality disorders. Patients' self-report of psoriasis severity correlated directly with depression and suicidal ideation, and comorbid depression reduced the threshold for pruritus in psoriatic patients. Heavy alcohol drinking (>80 grams of ethanol daily) by male psoriatic patients may predict a poor treatment outcome.

These possible links between the CNS and psoriasis have led to the development of psychosocial interventions in its treatment. Controlled studies have shown meditation, hypnosis, relaxation training, cognitive-behavioral stress management, and symptom control imagery training to be effective in reducing psoriasis activity.

Alopecia Areata Alopecia areata is a nonscarring hair loss in patches of typically well-demarcated smooth skin with breakage of the hair shaft resulting in characteristic “exclamation-mark hairs.” Hair loss often involves the scalp but can also affect the brows, lashes, beard, and body hair and varies from a single patch to multiple patches or total hair loss (Fig. 25.7-3). Alopecia areata accounts for about 2 percent of new dermatologic outpatient visits in the United States, with an estimated 1 percent of the U.S. population having at least one episode by the age of 50. The incidence is equal in men and women and peaks during the third to fifth decade. The prognosis is variable; 20 to 30 percent never recover from the first episode, and another 30 percent recover completely. Factors associated with poor prognosis include rapid evolution, prepubertal onset, ophiasis (loss of hair along the scalp margins), multiple episodes, associated atopy, severe nail changes, and loss of eyebrows and eyelashes.

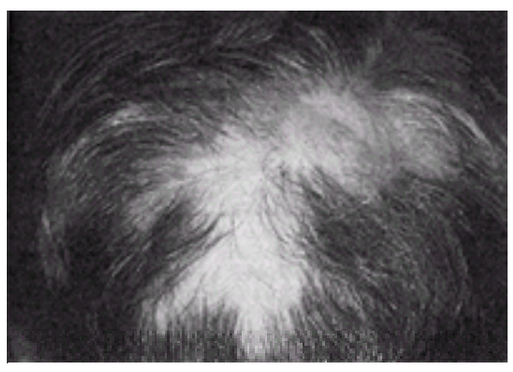

FIGURE 25.7-3 Alopecia areata. Note the typical patch of hair loss on the scalp. (Courtesy of D.F. Mutasim.)

The pathogenesis of alopecia areata is not completely understood but probably involves immunological and genetic factors. About 30 percent of patients report a positive family history. Some preliminary evidence implies that alopecia areata is an autoimmune disorder with T suppressor cell dysfunction, association with other autoimmune conditions, and frequently positive responses to immunosuppressive treatment such as corticosteroids or to induced, delayed-hypersensitivity reactions to a variety of sensitizers.

The role of psychiatric factors in the development of alopecia areata has been debated. The few controlled studies have yielded contradictory findings regarding the rates of psychopathology and the role of psychosocial events or stress in the onset of alopecia areata. In the 6 months preceding the onset of symptoms, events with a negative impact, exits from social fields (e.g., death, divorce), uncontrolled events, and socially desirable and undesirable events were significantly more frequent in alopecia areata patients than in controls. However the actual impact of these events on the development of alopecia areata is unclear. Recent, uncontrolled studies of lifetime comorbid psychiatric disorders in adults with alopecia areata found that many had lifetime psychiatric diagnoses, particularly major depression, generalized anxiety disorder, and paranoid disorder, suggesting that adult patients with alopecia areata may be at risk for comorbid psychiatric disorders. In another study, symptoms of depression and anxiety were more common in children with alopecia areata than in controls. Patients with highly stress reactive alopecia areata report more depressive symptoms than patients who do not note a relationship between stress and alopecia areata. A double-blind, placebo-controlled trial of imipramine (Tofranil) 75 mg daily found that the patients on imipramine had significantly more hair regrowth than controls, and this effect was independent of a reduction in anxiety or depression.

Urticaria and Angioedema Urticaria (also known as hives) is characterized by circumscribed, raised, erythematous, usually pruritic areas of edema that involve the superficial dermis (Fig. 25.7-4). When the edema extends into the deep dermis, subcutaneous or submucosal layers, it is called angioedema. About 15 to 20 percent of the general population develop urticaria at some point in their lives, with a peak incidence between ages 20 and 40. The male to female ratio is equal in children, but more women than men develop urticaria in adulthood. Most patients with acute urticaria (duration less than 6 weeks) readily respond to treatment of the underlying cause, usually infection or intolerance to specific drugs or food. However, in about 70 percent of patients with chronic urticaria or angioedema (duration greater than 6 weeks), the cause is unknown. Chronic idiopathic urticaria often responds poorly to usual dermatological treatment, although the prognosis improves with mild severity, short duration of the episode, and positive response to antihistamines. About 20 percent of patients with chronic idiopathic urticaria still have symptoms after 10 years. The manifestations of urticaria result from the release of vasoactive mediators in the skin, primarily histamine from mast cells or basophils. The release of these mediators can result from either immunological (IgE dependent or complement mediated) or nonimmunological mechanisms (agents that directly release mediators from mast cells and basophils).

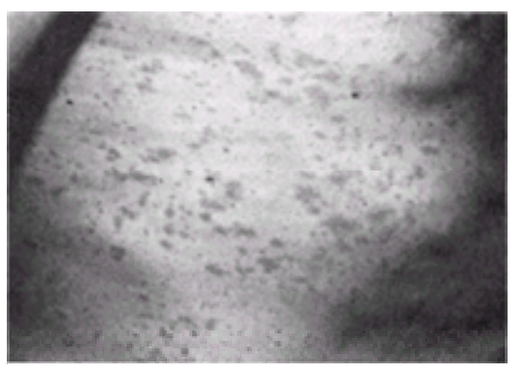

FIGURE 25.7-4 Urticaria. Note the circumscribed, raised areas of edema on the chest and abdomen. (Courtesy of D.F. Mutasim.)

Psychiatric factors have been implicated in the development of some types of urticaria. The typical “halo-hives” (papules surrounded by a white halo) of adrenergic urticaria that develops after acute emotional stress are associated with increased plasma noradrenaline and adrenaline concentrations, are reproduced with intradermal injection of noradrenaline, and resolve after treatment with the b-adrenergic receptor antagonist propranolol (Inderal). Most psychiatric studies, however, have focused on chronic idiopathic urticaria. Early psychodynamic theories about urticaria have been abandoned because no association between a specific personality conflict and urticaria could be proved, but recent controlled studies using well-documented psychometric instruments have found that patients with chronic idiopathic urticaria were frequently depressed and anxious and that women were more likely to experience significant psychiatric symptoms. However, whether the psychiatric symptoms resulted from urticaria or were a contributing causal factor in its development or exacerbation is unclear. Stress may also lead to the secretion of such neuropeptides as vasoactive intestinal peptide and substance P, which can cause vasodilation and contribute to the development of urticarial wheals.

While only one well-controlled study found that psychological treatment (hypnosis with relaxation) reduced pruritus (but not number of hives), controlled studies of psychotropic drugs demonstrated that antidepressant medications are effective in the management of chronic idiopathic urticaria. Doxepin at low dosages (10 mg three times a day) was more effective than diphenhydramine (Benadryl) (25 mg three times a day) for the control of chronic idiopathic urticaria, and 25 mg of nortriptyline (Pamelor) three times a day was significantly better than placebo in the treatment of pruritus and wheals. Whether the efficacy of antidepressants was solely a product of antihistaminic activity or involved a central effect as well is not known. An open trial found that both comorbid panic disorder and urticaria responded to treatment with selective serotonin reuptake inhibitors (SSRIs), suggesting a possible role for serotonergic mechanisms as well.

Acne Vulgaris Acne vulgaris is a common sebaceous gland disease typified by a variety of lesions including comedones, papules, pustules, or nodules and possible development of pitted or hypertrophic scars as a result. Occasionally present at birth, most cases develop in patients between the middle to late teens. The course is usually self-limited with spontaneous remission after several years, although acne may persist through the third decade and later, particularly in women. Acne is more common and tends to be more severe in men.

The etiology of acne is multifactorial. Patients frequently report stress-induced aggravation of acne, although the mechanism is unknown. One theory is that via the hypothalamic-pituitary-adrenal axis, stress leads to higher glucocorticoid and androgen concentrations, which can both exacerbate acne. Acne interferes significantly with social and occupational functioning, with a positive association between the severity of the acne, anxiety, and poor self-image. Patients with even mild-to-moderate acne experience significant psychological distress and body image concerns. Interestingly, successful treatment of acne with isotretinoin (Accutane) reduced both anxiety and depressive symptoms. A controlled study of adjunctive psychological treatment of acne patients receiving medical dermatological treatment demonstrated that biofeedback-assisted relaxation and cognitive imagery treatment significantly reduced acne severity compared with that in medical control groups.

Other Dermatologic Conditions

Rosacea Rosacea, a disorder of the cutaneous vasculature that occurs in adults between the ages of 30 and 60, is characterized by facial flushing, erythema, inflammatory lesions, lymphedema, and rhinophyma (hypertrophy of vascular and connective tissue of the nose). Rosacea is more likely to present on the cheeks and chin of women, while men tend to get rosacea on the nose. Development is probably multifactorial, with genetically susceptible individuals experiencing recurrent facial flushing (an important first stage in the pathogenesis of rosacea) when exposed to a variety of provocative environmental factors including emotional stimuli such as anxiety, excitement, or embarrassment. In past studies, patients frequently reported a connection between stress and the development of rosacea and were inclined to be more anxious than patients with alopecia areata or lichen planus.

Telogen Defluvium Telogen defluvium (also called telogen effluvium) is the increased loss of normal club hairs because of a perturbation of the hair cycle by drugs, high fever, childbirth, and crash dieting. More women than men seek clinical attention for telogen defluvium. Reports suggest that emotional stress can induce telogen defluvium, but the limited documentation in these older reports makes it difficult to conclude firmly that the patients had telogen defluvium and not other diagnoses

such as alopecia areata.

Primary Hyperhidrosis Primary hyperhidrosis is excessive sweating in response to mental stimuli and typically affects the palms, soles, and axillae. Patients usually present before the age of 20 and often have a positive family history. Although the pathophysiology is not completely understood, it involves the sympathetic nervous system and integration at the level of the hypothalamus. The disorder can cause considerable embarrassment and avoidance, but open studies have successfully used hypnosis, assertiveness training, systematic desensitization, and instrumental conditioning in the treatment of hyperhidrosis.

Seborrheic Dermatitis Seborrheic dermatitis, a common disorder affecting about 2 to 5 percent of the population, occurs slightly more often in male patients and at two age peaks: within the first 3 months and in the fourth to seventh decade of life. The cause of the increased sebum production (seborrhea) of the scalp and sebaceous follicle-rich areas of the face and trunk is unknown, but reports suggest that emotional stress may aggravate the symptoms.

Prurigo Nodularis Prurigo nodularis, a chronic disorder distinguished by intense localized pruritus and discrete hard nodules, occurs mainly in middle-aged women. The cause is unknown, but stress-induced exacerbations have been reported.

UNDIFFERENTIATED SOMATOFORM DISORDER

Chronic Idiopathic Pruritus Pruritus, or itching, is the most common symptom of dermatological disorders and of several systemic diseases including chronic renal disease, hepatic disease, hematopoietic disorders, endocrine disorders, malignant neoplasms, drug toxicity, and neurological syndromes (e.g., multiple sclerosis). Other problems such as advanced age, infections with internal parasites, and viremia can also be associated with pruritus.

Although pruritus is a common and distressing symptom, its pathophysiology is not completely understood. In histamine-induced pruritus, psychic trauma lowers itch threshold, aggravates itch intensity and prolongs itch duration, and its duration was more pronounced in subjects reporting “moodiness” during stressor exposure. Recent stressful life events have also been correlated with an increased ability to detect itch, and a study of pruritus in psoriasis, atopic dermatitis, and chronic idiopathic urticaria drew a direct correlation between pruritus severity and the degree of depressive symptoms, possibly due to a reduced itch threshold. Psychiatric factors may influence the perception of itch by several mechanisms. The hypothalamus and other components of the limbic system may modulate sensory perception through the sensory cortex, and elevation of corticotropin-releasing factor in some depressed patients may increase CNS opiate concentrations, enhancing the perception of pruritus. Neuropeptides such as substance P may also be released in response to stress and produce itching. Chronic idiopathic pruritus and idiopathic pruritus ani (itching in the anal area), vulvae (itching in the vaginal area), and scroti (itching in the scrotum) frequently have been called psychogenic, but more study is needed to determine how psychiatric and other CNS disorders contribute to the development of pruritus. Antidepressant medications, particularly the tricyclic drugs, can relieve pruritus of many origins. Behavioral treatment aimed at interrupting the itch-scratch cycle may be important in the management of pruritus and the prevention of complications such as *lichen simplex chronicus*, a condition of prominent skin markings and thickening of the tissue that develops secondary to long-term scratching. Techniques include habit reversal training and cognitive-behavioral therapy.

PAIN DISORDER

Glossodynia Glossodynia (also called burning mouth syndrome) is an unexplained, prolonged sensation of pain, burning, or both inside the oral cavity, most frequently at the tip and lateral borders of the tongue and often accompanied by other symptoms such as dryness, paresthesia, and changes in taste and smell. It affects up to 5 percent of the general dental population and occurs mostly in women with a mean age of 55 to 60. Symptoms are long-lived, lasting from months to up to 18 years. Several possible causes for glossodynia have been proposed including candida glossitis, vitamin B deficiencies, diabetes mellitus, estrogen deficiency at menopause, inadequate dental fillings, denture wear, parafunctional oral habits, salivary disturbances, and galvanism. The condition is often considered idiopathic.

Psychiatric factors are thought to play an important role in many patients with glossodynia. Controlled studies found higher rates of depression and anxiety in glossodynia patients, but causal relationships are uncertain. Currently no evidence supports a correlation between stressful life events and glossodynia. In open trials antidepressants (monoamine oxidase inhibitors, tricyclic drugs, and SSRIs) and benzodiazepines have shown beneficial effects in patients with glossodynia.

Vulvodynia Vulvodynia is chronic vulvar and perineal discomfort of variable severity with burning, stinging, irritation, or rawness. It is a clinical problem of importance to dermatologists, gynecologists, and primary care physicians. In general gynecological outpatient practices, the prevalence can be as high as 15 percent. The age distribution is variable, ranging from the mid-20s to the late 60s and older. Vulvar pain frequently has an acute onset but can become chronic, lasting for several months or years. Many patients seek help from several physicians and try multiple therapies, frequently with unsatisfactory results.

Vulvodynia without an identifiable medical cause was once thought to be psychological, but more-recent diagnostic techniques and discoveries have permitted identification of the medical cause of many cases, including vulvar dermatoses, cyclic candidal vulvitis, vulvar papillomatosis, and vulvar vestibulitis. However, many cases of “essential” or dysesthetic vulvodynia remain, in which the cause is unknown. Dysesthetic vulvodynia is postulated to result from a problem with cutaneous perception, either centrally or at the nerve root. It occurs more frequently in perimenopausal or postmenopausal women and is characterized by constant vulvar or perineal discomfort with frequent urethral and rectal discomfort as well. These patients have less dyspareunia and less point tenderness on examination than patients with vulvar vestibulitis.

In a controlled study, patients with vulvodynia of any cause were significantly more likely than patients with other vulvar pathology to somatize and to be anxious, were more bodily aware, and were more bothered by many different symptoms in addition to vulvar complaints. Vulvodynia patients also consulted more doctors and experienced more sexual dysfunction related to their symptoms. Patients with dysesthetic vulvodynia were even more distressed than other vulvodynia patients with more anxiety and suggestibility. The role of comorbid psychiatric symptoms needs further study, but they may influence the perception of pain through central mechanisms. In a retrospective study of psychotropic treatment of vulvodynia, the patients with dysesthetic vulvodynia responded to the tricyclic drug amitriptyline (Elavil). Other types of vulvodynia were less responsive to antidepressant treatment. The amitriptyline dosage ranged from 10 mg to 100 mg per day, with an average daily dosage of 60 mg and an average treatment duration of 7 months. Other tricyclic drugs such as nortriptyline and desipramine (Norpramine) may also be effective and possibly better tolerated than amitriptyline, but more study is needed.

Body Dysmorphic Disorder Body dysmorphic disorder is a preoccupation with an imagined defect in appearance, and patients with this disorder who are concerned about the appearance of their skin or hair frequently consult dermatologists. Some patients with body dysmorphic disorder pick the skin in an attempt to improve their appearance, as discussed below. Although body dysmorphic disorder is classified as a somatoform disorder, it has phenomenological similarities to obsessive-compulsive disorder.

DELUSIONAL DISORDER, SOMATIC TYPE

Delusions of Parasitosis Delusions of parasitosis is the fixed belief that one is infested with living organisms despite a lack of medical evidence of such infestation. Patients also frequently experience tactile hallucinations described in many cases as crawling, biting, or stinging or other perceptual abnormalities such as auditory hallucinations of buzzing or other sounds. Delusions are often of a cutaneous invasion, so patients frequently present to dermatologists. Delusions of parasitosis appears to be uncommon, with an equal sex distribution in patients under the age of 50 and a female to male ratio of 3 to 1 in patients 50 and older.

Delusions of parasitosis has been associated with many medical disorders including tuberculosis, syphilis, chronic lymphocytic leukemia, polycythemia vera, malignant lymphoma, congestive heart failure, arteriosclerosis, diabetes mellitus, vitamin B_{12} deficiency, pellagra, renal failure, hypothyroidism, and hepatic disease and the dermatological disorders stasis dermatitis and vitiligo. Neurological conditions associated with delusions of parasitosis include dementia, Parkinson's disease, Huntington's disease, cerebral infarction, and CNS tumors. Some of the disorders appear to cause delusions of parasitosis, while others seem to at least contribute to its development. Delusions of parasitosis can also occur after long-term use of amphetamines, methylphenidate (Ritalin), cocaine, or alcohol or use of phenelzine (Nardil), pargyline, and corticosteroids. The disorder usually remits with discontinuation of the drug.

Several psychiatric disorders may also exhibit delusions of parasitosis as part of their clinical presentation, including major depressive disorder, bipolar I disorder, and schizophrenia. Delusions of parasitosis also occurs as a single somatic delusion with no other impairment of thought or thought process, called monosymptomatic hypochondriacal psychosis and classified under delusional disorder, somatic type, in DSM-IV. Because their mental status is otherwise intact, these patients may be more likely than those with other psychiatric disorders to seek care from a dermatologist. Other patients develop tactile hallucinations (formication) but do not have

delusions about the cause of the symptoms and are less likely to have a history of psychiatric disorders than those with delusions of parasitosis.

Characteristic features of delusions of parasitosis include a specific precipitant, a history of potential or actual exposure to contagious organisms or actual infestation, and a history of multiple consultations with physicians and other professionals such as entomologists and exterminators. However, many patients, particularly those with an isolated delusion of infestation, are reluctant to accept psychiatric evaluation because of the implication that the infestation is not real. Patients make multiple attempts at treatment and often use nontraditional therapies to eradicate the "bugs," such as repetitive washing, checking, and cleaning; excoriation of the skin with needles, knives, or fingernails; and discarding or destroying possessions, including the family pet. The delusion may be shared by others *(folie à deux)*, and the fear of contaminating others is often present. The patients usually bring in specimens for examination by the physician but all that is found is inanimate material, nonharmful insects, or bits of hair and skin. Descriptions of the offending parasite vary from the imprecise "things" or "bugs" to detailed explanations of the organism's (usually insects, spiders, or worms) appearance, behavior, and life cycle. Patients frequently report social isolation and loss of employment as a result of the delusions, and patients have rarely been reported to commit suicide. Dermatologists have noted that the mistaken worries about infestation span a spectrum from nondelusional to delusional thinking. Sometimes the preoccupation with infestation resembles obsessions in that it is distressing, anxiety provoking, persistent, recurrent, and difficult to resist or control. Furthermore the repetitive behaviors that often accompany the preoccupation are compulsive. Many patients have both obsessional and delusional thinking, which may have important treatment implications.

One must seek underlying medical and psychiatric conditions that may be contributing to the delusions of parasitosis. Patients with delusions of parasitosis as an isolated delusion (delusional disorder, somatic type) respond to treatment with the potent neuroleptic pimozide (Orap). Only one controlled study of the treatment of delusions of parasitosis with pimozide exists, and 10 of 11 patients improved after 6 weeks of 1 to 5 mg per day. A follow-up study indicated that many patients treated with pimozide could discontinue the medication after a mean treatment duration of 5 months without recurrence of the delusions. In other anecdotal reports pimozide doses up to 12 mg per day have been used. There are uncontrolled reports of efficacy with other antipsychotic medications including chlorpromazine (Thorazine), trifluoperazine (Stelazine), haloperidol (Haldol), risperidone (Risperdal), and fluphenazine (Prolixin). Although pimozide is most commonly advocated for the treatment of delusions of parasitosis (particularly delusional disorder, somatic type), there is no conclusive evidence that pimozide is superior to other antipsychotics. A case report of successful treatment with naloxone (Narcan) and naltrexone (ReVia) suggests a disorder in the endogenous opiate system, and interestingly, pimozide is the only neuroleptic used in delusions of parasitosis that is a potent opiate antagonist. Pimozide may be antipruritic as a result of its opiate antagonism, and the relief of pruritus or paresthesias could contribute to improvement in delusions of parasitosis. There are also case reports of successful treatment of delusions of parasitosis with tricyclic drugs and electroconvulsive therapy. Because of the obsessive characteristics of many patients with delusions of parasitosis, the symptoms may possibly respond to the SSRIs, but more study is needed. Psychotherapy has minimal efficacy in the treatment of this disorder. However, behavior therapy may have a role in treatment, especially for the obsessive-compulsive characteristics of many patients with delusions of parasitosis, but more study is needed. Dermatologists have advocated an approach to patients with delusions of parasitosis that allows development of a therapeutic relationship with them. This is accomplished by listening to patients' stories, asking how the condition affects their lives, being alert to any area in which they will accept help, and trying to reduce their sense of isolation.

Other Delusions In the delusional variant of body dysmorphic disorder, the preoccupations with a defect in appearance are of delusional intensity, although the distinction between an obsessive concern and delusions is not always clear. Delusional body dysmorphic disorder may respond preferentially to SSRIs, and if a patient does not completely respond to an SSRI, the addition of an antipsychotic such as pimozide may help.

The delusion of a foul body odor is another encapsulated somatic delusion that a dermatologist may encounter. Treatment data are limited, but pimozide and other antipsychotic medications may be effective.

OBSESSIVE-COMPULSIVE SPECTRUM DISORDERS

Psychogenic Excoriation Psychogenic excoriations (also called *neurotic excoriations*) are lesions caused by scratching or picking in response to an itch or other skin sensation or because of an urge to remove an irregularity on the skin from preexisting dermatoses such as acne *(acné excoriée)*. Lesions are typically found in areas that the patient can easily reach (e.g., the face, upper back, and the upper and lower extremities) and are a few millimeters in diameter and weeping, crusted, or scarred, with occasional postinflammatory hypopigmentation or hyperpigmentation (Fig. 25.7-5). The patients usually acknowledge the self-inflicted nature of the lesions. There is an incidence of 2 percent among dermatology clinic patients and a predominance of women. Most studies report a mean age of onset between 30 and 45 years of age, although *acné excoriée* can begin in adolescence and persist into adulthood. The mean duration of symptoms is 5 years, with a better prognosis for patients who have had the symptoms for less than 1 year.

FIGURE 25.7-5 Psychogenic excoriation. The self-induced nature of the condition is suggested by the relative sparing of the lateral upper back, where the patient cannot easily reach.

The behavior in psychogenic excoriation sometimes resembles obsessive-compulsive disorder in that it is repetitive, ritualistic, and tension reducing, and patients attempt (often unsuccessfully) to resist excoriating, a behavior they find ego-dystonic. Patients also sometimes describe obsessions about an irregularity on the skin or preoccupations with having smooth skin and may excoriate in response to the thoughts. The preoccupation with appearance can be severe enough to meet criteria for body dysmorphic disorder, a disorder also thought to be related to obsessive-compulsive disorder. The excoriative behavior can also have features characteristic of impulse-control disorders in that the patients often act automatically and sometimes experience increased tension before the behavior, with transient pleasure or relief immediately afterward. Thus, behaviors can span a compulsivity-impulsivity continuum from purely obsessive-compulsive to purely impulsive with mixed symptoms in between. The case below illustrates the mixed compulsivity-impulsivity of psychogenic excoriation.

> A 22-year-old woman with no past medical history presented with a 3-year history of excoriation of acne lesions. She described herself as "obsessed" with having a "flawless" complexion. When blemish eruptions occurred, she became depressed and worried constantly that people would look at her skin and think she was unclean. She became reluctant to leave her home and repeatedly examined and washed her face. She picked facial lesions until "something came out" and could "work at" her face for 30 minutes at a time, becoming completely focused on the behavior. At other times, she found herself picking her face automatically, while doing homework or watching television. She tried to resist the urge to excoriate the lesions but felt she had little control over the behavior. She reported feeling tension build with the urge to pick her face and some relief of tension and anxiety and frequently pleasure upon acting. However, the behavior was often followed by self-reproach, as she was aware that she was making her skin worse.

The repetitive and tension-reducing nature of the behavior and the resistance to performing the behavior are consistent with compulsions. Obsessive features include the preoccupation with a flawless complexion and cleanliness. The automatic enactment of the excoriation and the sense of tension before the behavior and pleasure upon acting are characteristic of impulsivity.

Studies examining the psychiatric comorbidity of patients with psychogenic excoriation found depressive and anxiety disorders to be common. A recent report noted that patients with impulsive excoriative behavior had a high rate of comorbid bipolar disorders, consistent with studies of other impulse-control disorders in which bipolar disorders are frequently comorbid.

Case reports and open trials demonstrate the responsiveness of compulsive skin picking or scratching to treatment with the serotonin reuptake inhibitors fluoxetine

(Prozac), sertraline (Zoloft), clomipramine (Anafranil), and fluvoxamine (Luvox). One controlled trial of the treatment of psychogenic excoriation found that fluoxetine may be beneficial. There are also case reports of successful treatment of psychogenic excoriation and acné excoriée with behavioral therapy.

Trichotillomania Trichotillomania is a disorder of chronic hair pulling currently classified as an impulse-control disorder in DSM-IV, but like psychogenic excoriations, the hair pulling has characteristics of both impulsivity and compulsivity. It causes significant distress or impairment in functioning and leads to notable alopecia, most commonly of the scalp hair, but lashes, brows, pubic hair, and other body hair may be involved (Fig. 25.7-6). The extracted hair may also be chewed or swallowed, leading to trichobezoars. Patients may also develop infections at the site of hair pulling, a change in texture or color of the hair or carpal tunnel syndrome from pulling. In a survey of college freshmen, the lifetime prevalence of trichotillomania (by strict criteria) was 0.6 percent for both men and women but rose to 3.4 percent for women and 1.5 percent for men if all hair pulling with noticeable hair loss was included. The condition typically persists for years, with a mean age of onset of 13 years. The prognosis appears to be better with hair pulling of 6 months' duration or less, but most do not seek treatment.

FIGURE 25.7-6 Trichotillomania. Note the typical findings of an area of incomplete alopecia involving the frontal and vertex scalp.

Researchers have identified two types of hair pulling. In the focused style, patients focus solely on the pulling without attention to other thoughts and activities. Symptoms resemble obsessive-compulsive disorder in that there is resistance to the pulling and relief with pulling. However, most patients engage in an automatic style of pulling that occurs during situations described as “sedentary or contemplative” such as reading or watching TV. Many patients have a combination of these two styles.

In clinical samples of trichotillomania, Axis I comorbidity included anxiety, mood, obsessive-compulsive, substance use, and eating disorders. The most frequent comorbid Axis II disorders included Cluster B (histrionic, borderline) personality disorders and Cluster C (obsessive-compulsive, passive-aggressive) personality disorders. However, whether one particular personality disorder is characteristic of patients with trichotillomania is unclear. Patients with trichotillomania frequently have first-degree relatives with obsessive-compulsive disorder and trichotillomania, and family histories of mood disorders, substance use disorders, anxiety disorders, and schizophrenia.

In a double-blind crossover study, trichotillomania responded preferentially to the antiobsessional agent clomipramine over desipramine, and in another report, response to clomipramine negatively correlated with anterior cingulate and orbital frontal metabolism on position emission tomography (PET) examination, a finding also characteristic of obsessive-compulsive disorder. Open studies of the antiobsessional SSRI fluoxetine reported positive short- and long-term results, but controlled studies have yielded mixed results. Two placebo-controlled studies found that fluoxetine was not superior to placebo in the treatment of trichotillomania, but in a double-blind crossover study of clomipramine and fluoxetine, both medications had positive treatment effects on trichotillomania. Case reports and open trials also relate successful treatment with other SSRIs including fluvoxamine, paroxetine (Paxil), and sertraline, and other antidepressants, lithium, and buspirone (Buspar). Pimozide augmentation of clomipramine or fluoxetine and olanzapine augmentation of fluoxetine have also proved useful in case studies. The behavioral treatment of habit reversal was reported effective in trichotillomania and involves increasing awareness of situations or stressors associated with hair pulling, relaxation training, and competing response training. Hypnosis and other behavioral treatments have also been useful in case reports.

Onychophagia Onychophagia (repetitive nail biting) is a common behavior that can begin as early as age 4, with a peak between the ages of 10 and 18 years of age and appears to be familial. Severe onychophagia can lead to significant medical and dental problems such as hand infection and craniomandibular disorders. Onychotillomania, the picking or tearing of the nail, may be a variant of the behavior. Like trichotillomania and psychogenic excoriation, onychophagia possesses the phenomenological features of repetition, resistance, and relief and may therefore be one of the obsessive-compulsive spectrum disorders. Like trichotillomania, a double-blind crossover study demonstrated that onychophagia responded preferentially to clomipramine over desipramine. Many forms of behavioral therapy, including habit reversal, have been efficacious.

FACTITIOUS DISORDERS

Factitious Dermatitis Factitious dermatitis (also called dermatitis artefacta) is a disorder in which the skin is the target of self-inflicted injury. As in other factitious disorders, patients intentionally produce lesions to assume the sick role and typically deny the self-inflicted nature of the disorder. Factitious dermatitis can present with a wide variety of lesions depending on the methods used by the patient. Sometimes patients simulate actual medical dermatoses or aggravate a preexisting dermatological condition. Lesions consist of excoriations, blisters, purpura, ulcers, erythema, edema, sinuses, and nodules. Methods used by patients include rubbing, scratching, picking, cutting, puncturing, sucking, biting, application of suction cups, occlusion, applications of dye or heat or caustic substances onto the skin, or the injection of caustics, infected material, blood, feces, or other substances into the skin. Skin damage can be extensive with full-thickness skin loss and severe scarring requiring plastic surgery or even amputation. Despite the variety of factitious lesions, they have some common features. Affected sites tend to be easily accessible to the patient and are more prominent on one side of the body, depending on the handedness of the patient. Excoriated lesions frequently have sharp geometric borders with normal surrounding skin or lines at the margin if a caustic is used and some liquid dripped out of the site. There is no evidence of the natural progression of the lesions through the different stages of development. The history is often vague and hollow, and patients are often very suggestible as to the site of the next lesion. The use of an occlusive dressing such as an Unna Boot (a dressing that hardens to form a protective covering) can help make the diagnosis because the lesions heal when the patients are unable to reach them.

Factitious dermatitis occurs in about 0.3 percent of dermatology patients, with a reported female to male ratio between 3 to 1 and 8 to 1 and greatest frequency in adolescents and young adults. It reportedly begins after severe psychosocial stress, usually involving loss, threatened loss, or isolation. As in other factitious disorders, the patient often has experience in medicine either from prior exposure to illness or employment in a health care field. The prognosis varies; some cases resolve after a brief episode while others become lifelong problems. In a classic long-term study, 30 percent of patients continued to produce lesions more than 12 years after the onset of their symptoms.

Physicians must distinguish factitious dermatitis from other forms of self-inflicted dermatoses described in the sections above. Although borderline personality disorder is commonly comorbid with factitious dermatitis, some patients with borderline personality disorder readily acknowledge that they damage their skin in an impulsive act that resembles obsessive-compulsive spectrum disorders. Self-mutilation by impulsive patients may result from low serotonergic function and may respond to the SSRI fluoxetine. Other self-mutilating patients with psychosis or mental retardation do not have the other characteristics of a factitious disorder. Physicians must also eliminate the possibility of malingering.

The management of factitious dermatitis involves developing an empathic, therapeutic relationship between the dermatologist and the patient. Since the patients deny the self-inflicted nature of the lesions, they usually do not immediately accept a referral to a psychiatrist, and direct confrontation may disrupt the doctor-patient relationship. However, once a therapeutic relationship has been established, some patients will accept a psychotherapeutic approach.

Psychogenic Purpura The spontaneous appearance of recurrent bruising (purpura) is a rare dermatological disorder often following an injury or surgical procedure and typically initiated by pain, burning, or stinging succeeded by warmth, erythema, swelling, sometimes pruritus, and, after hours or days, ecchymosis. Blood coagulation and hemostatic test results, however, remain normal. The incidence is unknown, but women are affected more frequently than men, with a female to male ratio of 20 to 1. The age of onset varies from 9 to 53 years, with most between the ages of 14 and 40. Three different proposed mechanisms currently exist:

autoerythrocyte sensitization, conversion reaction, and factitious disorder.

Autoerythrocyte sensitization (Gardner-Diamond syndrome) attempts to explain the appearance of lesions after trauma and postulates that during the injury patients became sensitized to the stroma of extravasated erythrocytes. Many patients' bruises are reproduced by intracutaneous injection of washed red blood cells or erythrocyte stroma, but some patients do not have a positive response to the injections. Furthermore, tests for circulating antibodies to red blood cell components have been negative, no abnormalities in immunological function have been evident, and immunosuppressive treatment and desensitization with autologous blood products have been unsuccessful.

The limited evidence for autoerythrocyte sensitization has led some authors to advocate a psychogenic cause for the purpura. Purpura often begins after a significant psychosocial stressor, and most patients present with several other unexplained somatic symptoms and conversion symptoms and have significant comorbid psychiatric symptoms, including depression, anxiety, and personality disorders, particularly borderline and histrionic personality disorders. However, the differentiation between conversion and factitious purpura has been less clear because patients thought to have conversion symptoms and those found to have factitious disorder share many common features. In addition, physicians can identify no mechanism for a conversion reaction or convincingly disprove factitious causation in many cases. The underlying psychiatric cause is therefore unresolved and awaits further information about possible psychophysiological mechanisms.

PSYCHIATRIC CONSULTATION-LIAISON WITH DERMATOLOGISTS

Disorders at the interface of psychiatry and dermatology are varied. Developments in both fields have improved understanding of psychocutaneous disorders. Continued collaboration between dermatologists and psychiatrists is needed to further the understanding of these disorders.

Psychocutaneous disorders are best managed when the psychiatrist and dermatologist have a close working relationship. Because many patients with psychocutaneous disorders are initially reluctant to accept a psychiatric referral, dermatologists must educate the patients about the role of stress and psychiatric problems in their skin disorder and sometimes initiate treatment with psychotropic medications. In some medical centers, psychodermatology clinics in which a psychiatrist is present in the dermatology setting have been established to improve the management of psychocutaneous disorders. Because the psychiatrist is readily available at these clinics and part of the medical team, patients may be more willing to accept psychiatric evaluation and treatment.

SUGGESTED CROSS-REFERENCES

Somatoform disorders are discussed in Chapter 16, delusional disorders in Section 13.2, impulse-control disorders in Chapter 22, obsessive-compulsive disorders in Chapter 15, and factitious disorders in Chapter 17.

SECTION REFERENCES

Arnold LM, McElroy SL, Mutasim DF, Dwight MM, Lamerson CL, Morris EM: Characteristics of 34 adults with psychogenic excoriation. J Clin Psychiatry *59:*509, 1998.

*Arnold LM, Mutasim DF, Dwight MM, Lamerson CL, Morris EM, McElroy SL: Open clinical trial of fluvoxamine treatment of psychogenic excoriation. J Clin Psychopharmacol *19:*15, 1999.

Badoux A, Levy DA: Psychologic symptoms in asthma and chronic urticaria. Ann Allergy *72:*229, 1994.

Christenson GA, Crow SJ: The characterization and treatment of trichotillomania. J Clin Psychiatry *57:*42, 1996.

Ehlers A, Stangier U, Gieler U: Treatment of atopic dermatitis: A comparison of psychological and dermatological approaches to relapse prevention. J Consult Clin Psychol *63:*624, 1995.

Fitzpatrick TB, Eisen AZ, Wolff K, Freedberg IM, Austen KF, editors: *Dermatology in General Medicine,* ed 4. McGraw-Hill, New York, 1993.

Fortune DG, Main CJ, O'Sullivan TM, Griffiths CEM: Assessing illness-related stress in psoriasis: The psychometric properties of the psoriasis life stress inventory. J Psychosom Res *42:*467, 1997.

Fried RG: Evaluation and treatment of “psychogenic” pruritus and self-excoriation. J Am Acad Dermatol *30:*993, 1994.

Gil KM, Sampson HA: Psychological and social factors of atopic dermatitis. Allergy *44:*84, 1989.

*Gupta MA, Gupta AK: Psychodermatology: An update. J Am Acad Dermatol *34:*1030, 1996.

Gupta MA, Gupta AK, Ellis CN, Voorhees JJ: Some psychosomatic aspects of psoriasis. Adv Dermatol *5:*21, 1990.

*Gupta MA, Gupta AK, Haberman HF: The self-inflicted dermatoses: A critical review. Gen Hosp Psychiatry *9:*45, 1987.

Gupta MA, Gupta AK, Schork NJ, Ellis CN: Depression modulates pruritus perception: A study of pruritus in psoriasis, atopic dermatitis, and chronic idiopathic urticaria. Psychosom Med *56:*36, 1994.

Gupta MA, Gupta AK, Watteel GN: Stress and alopecia areata: A psychodermatologic study. Acta Derm Venereol *77:*296, 1997.

Hashiro M, Okumura M: Anxiety, depression and psychosomatic symptoms in patients with atopic dermatitis: Comparison with normal controls and among groups of different degrees of severity. J Dermatol Sci *14:*63, 1997.

Headington JT: Telogen effluvium. Arch Dermatol *129:*356, 1993.

Johnson GC, Anton RF: Delusions of parasitosis: Differential diagnosis and treatment. South Med J *78:*914, 1985.

Keuthen NJ, O'Sullivan RL, Sprich-Buckminster S: Trichotillomania: Current issues in conceptualization and treatment. Psychother Psychosom *67:*202, 1998.

Koblenzer CS: *Psychocutaneous Disease*. Grune & Stratton, Orlando, 1987.

Koo JY, Shellow WV, Hallman CP, Edwards JE: Alopecia areata and increased prevalence of psychiatric disorders. Int J Dermatol *33:*849, 1994.

Koo JY, Smith LL: Psychologic aspects of acne. Pediatr Dermatol *8:*185, 1991.

Krahn LE, Goldberg RL: Psychotropic medications and the skin. In *Psychotropic Drug Use in the Medically Ill,* vol 21, PA Silver, editor. S Karger AG, Basel, Switzerland, 1994.

Leonard HL, Lenane MC, Swedo SE, Rettew DC, Rapoport JL: A double-blind comparison of clomipramine and desipramine treatment of severe onychophagia (nail biting). Arch Gen Psychiatry *48:*821, 1991.

Lerer B: Hyperhidrosis: A review of its psychological aspects. Psychosomatics *18:*28, 1977.

Liakopoulou M, Alifieraki T, Katideniou A, Kakourou T, Tselalidou E, Tsiantis J, Stratigos J: Children with alopecia areata: Psychiatric symptomatology and life events. J Am Acad Child Adolesc Psychiatry *36:*678, 1997.

*McElroy SL, Phillips KA, Keck PE: Obsessive-compulsive spectrum disorder. J Clin Psychiatry *55:*33, 1994.

McKay M: Vulvodynia. Dermatol Clin *10:*423, 1992.

*Morris M: Delusional infestation. Br J Psychiatry *159:*83, 1991.

Park BS, Youn JI: Factors influencing psoriasis: An analysis based upon the extent of involvement and clinical type. J Dermatol *25:*97, 1998.

Phillips KA, Taub SL: Skin picking as a symptom of body dysmorphic disorder. Psychopharmacol Bull *31:*279, 1995.

Potenza MN, Wasylink S, Epperson CN, McDougle CJ: Olanzapine augmentation of fluoxetine in the treatment of trichotillomania. Am J Psychiatry *155:*1299, 1998.

Ratnoff OD: Psychogenic purpura (autoerythrocyte sensitization): An unsolved dilemma. Am J Med *87:*3, 1989.

Rojo L, Silvestre FJ, Bagan JV, DeVicente T: Prevalence of psychopathology in burning mouth syndrome. Oral Surg Oral Med Oral Pathol *78:*312, 1994.

Rothe MJ, Grant-Kels JM: Atopic dermatitis: An update. J Am Acad Dermatol *35:*1, 1996.

Simeon D, Stein DJ, Gross S, Islam N, Schmeidler J, Hollander E: A double-blind trial of fluoxetine in pathologic skin picking. J Clin Psychiatry *58:*341, 1997.

*Stein DJ, Hollander E: Dermatology and conditions related to obsessive-compulsive disorder. J Am Acad Dermatol *26:*237, 1992.

Stewart DE, Reicher AE, Gerulath AH, Boydell KM: Vulvodynia and psychological distress. Obstet Gynecol *84:*587, 1994.

Stocker WW, McIntyre OR, Clendenning WE: Psychogenic purpura. Arch Dermatol *113:*606, 1977.

Thiboutot DM: Acne rosacea. Am Fam Physician *50:*1691, 1994.

Zachariae R, Oster H, Bjerring P, Kragballe K: Effects of psychologic intervention on psoriasis: A preliminary report. J Am Acad Dermatol *34:*1008, 1996.

Zomer SF, Dewitt RFE, Van Bronswijk JEHM, Nabarro G, Van Vloten WA: Delusions of parasitosis. A psychiatric disorder to be treated by dermatologists? An analysis of 33 patients. Br J Dermatol *138:*1030, 1998.

25.8 MUSCULOSKELETAL DISORDERS

TERESA A. RUMMANS, M.D., KEMUEL L. PHILBRICK, M.D., AND M. KEVIN O'CONNOR, M.D.

Numerous diseases come under the umbrella of musculoskeletal disorders. The three most common musculoskeletal disorders encountered by psychiatrists are rheumatoid arthritis, systemic lupus erythematosus, and fibromyalgia. Although there is no single unifying theory for the development and pathogenesis of these conditions, multidimensional explanations currently aid in integrating them into one category. They involve primarily biological but also psychological and social components. This section focuses on the etiology, presentation, diagnosis, and treatment of these diverse but similar musculoskeletal disorders.

RHEUMATOID ARTHRITIS

The name rheumatoid arthritis was coined in 1859, only a generation after the first distinct clinical description in 1800 by Augustin Landre-Beauvais. However, ample evidence indicates that rheumatoid arthritis afflicted diverse individuals prior to surfacing in the medical literature. Thomas Sydenham offered a detailed description of a chronic joint disease with distinguishing features suggesting rheumatoid arthritis in 1676. Disiderius Erasmus, who died in 1536, was portrayed by the Flemish painter Quinten Metsys with swelling of three metacarpophalangeal joints clearly visible on his right hand. Constantine IX, the Byzantine emperor who ruled a half-millenium earlier, was afflicted with a chronic, polyarticular, deforming joint disorder; the description offered by a courtier inclines historians to suspect it was rheumatoid arthritis. Reaching even further back in time, 13 of 213 well-preserved skeletons of Native Americans dating back at least 3000 years were subjected to both gross and radiological examination that revealed extensive joint erosions consistent with rheumatoid arthritis.

The specific nosology of rheumatic diseases continues to pitch and roll with continental preferences and episodic surging advances in the understanding of pathophysiology, but rheumatoid arthritis remains under the rubric of disorders involving the immune system, connective tissue, and joints. The first two editions of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I, DSM-II) regarded rheumatoid arthritis as psychosomatic. By contrast, the disease is not directly mentioned in the fourth edition of DSM (DSM-IV), although rheumatological disorders are cited as one example of illness in which psychological factors may affect a medical condition.

Epidemiology Rheumatoid arthritis does not respect geography or race; it has been identified in all areas of the world and every ethnic group that has been studied. Prevalence rates, however, do show ethnic and geographic partiality. Asian populations enjoy a lower prevalence rate, while North American and many European populations have a lifetime prevalence of definite rheumatoid arthritis among white adults that hovers around 1 percent. Peak prevalence rates of 5.3 percent were discovered in select Native American populations. Rates in African populations vary; African-Americans have a rate roughly parallel to that of white Americans, whereas in Africa, this prevalence holds for urban populations but is significantly lower in many rural populations. Some attribute this disparity to excess mortality from infections associated with rheumatoid arthritis among rural Africans with limited access to medical care. Other epidemiological studies have concluded that this disparity in disease incidence despite similar genetic background reflects the supplemental role of environmental factors in the genesis of rheumatoid arthritis.

The disease afflicts women two to three times more often than men, regardless of age. The peak age of onset (usually placed in the early to mid-40s) is difficult to determine, owing to the varied diagnostic criteria used, but men are more commonly affected several years later than women. Studies on both sides of the Atlantic have demonstrated a decrease in age-adjusted annual incidence rates in women but not men; this has prompted the suggestion that oral contraceptives may influence the onset of rheumatoid arthritis.

Family and twin studies support a definite genetic predisposition. An association between rheumatoid arthritis and various genes and genomic subtypes in the human lymphocyte antigen (HLA)-D complex is established. Different ethnic groups demonstrate selective associations (e.g., North American whites possess an association with HLA-DR4, Israeli Jews with HLA-DR1, and Yakima Indians with HLA-Dw16). These associations do not offer complete accountability for the susceptibility to rheumatoid arthritis, however, suggesting that as yet unidentified genes outside the HLA complex also participate. These genetic factors appear to extend to the propensity for particular toxic reactions to medications encountered by patients with rheumatoid arthritis. For example, adverse effects of gold therapy such as proteinuria and thrombocytopenia appear genetically linked to the presence of the HLA-DR3 allele. Thus far, however, HLA typing has not been able to predict response to a particular treatment regimen.

Etiology Despite impressive advances in knowledge regarding the genetic and immunological contributions to rheumatoid arthritis, it remains a condition of unknown cause whose progression is well described but without identification of the specific factors that fuel disease maintenance and progression. Considerable current research attention is focused on the possibility that it may result from infection by an unknown but ubiquitous organism occurring in a genetically susceptible host.

Clinical tradition, buttressed by anecdotal lore and flawed research, long attributed causal strength to a psychological style characterized by restricted emotional expression (with particular emphasis upon the inability to express anger), obsessionality, dependency, and depression. However, appreciation grew for the weaknesses inherent in retrospective research (conducted without appropriate control groups and with inadequate attention to the concurrent medical status of participants) and the breadth of clinical severity of rheumatoid arthritis and subsequent studies have not validated any etiological contribution of the “arthritic personality.”

Studies attempting to clarify whether acute stress might play a precipitant role in the onset of rheumatoid arthritis have come to varying conclusions, leading some to suggest that subtypes of patients may exhibit differential sensitivity to the potential triggering effects of stress. Other reviewers attribute the diversity of results to the equal diversity of methodological inadequacies of these studies. Distressed interpersonal relationships are especially prominent in the literature of stressors experienced by patients with rheumatoid arthritis, despite the lack of evidence for etiological significance.

The capacity to measure suppression of cellular and humoral immune responsiveness following stress has appropriately fueled speculation that there may be a psychoneuroimmunologic circuit whereby psychological issues wield biochemical central nervous system effects that trigger immunological responses, forming the foundation for rheumatoid arthritis. Much of this research remains limited to animal models, with no significant contribution as yet to the understanding of a potential role of psychological stress in rheumatoid arthritis. Nonetheless, suggestive results in animal studies continues to invite translation research involving rheumatoid arthritis patients.

Diagnosis and Clinical Features Rheumatoid arthritis is a systemic, chronic, inflammatory, and progressive disease. Patients usually experience a prodromal syndrome that may include vague constitutional symptoms, fleeting musculoskeletal pain, and variable morning stiffness that may last for weeks or months without yielding to diagnosis. Although a few patients experience abrupt onset, most note an insidious beginning followed by gradual involvement of multiple joints in a symmetrical pattern. A few patients have an asymmetrical pattern.

Although no joint is necessarily spared, symmetrical polyarthritis of the peripheral joints is most common. Morning stiffness is almost pathognomonic and can serve as a rough gauge of disease activity. Constitutional symptoms as well as local inflammation in affected joints continue. Over years, these joints suffer articular destruction, consequent instability, and subsequent deformity with collateral pain and functional impairment (Fig. 25.8-1). Neurological, ocular, pulmonary, cardiac, vascular, hematopoietic, gastrointestinal, hepatic, renal, muscular, and dermatological sequelae are well described with persistent active rheumatoid arthritis. Neurological corollaries of the disease are generally limited to the peripheral nervous system, relieving the central nervous system of primary consequences, although rare instances of delirium have occurred, perhaps related to central rheumatoid vasculitis.

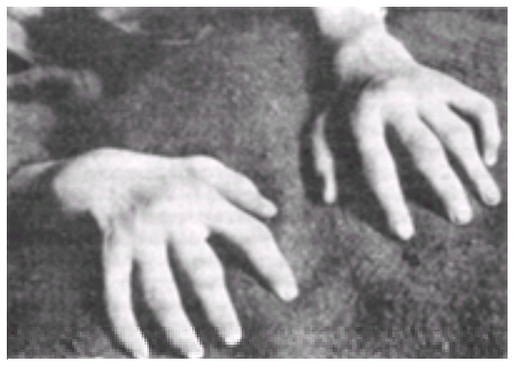

FIGURE 25.8-1 Acute rheumatoid arthritis: showing the spindle-shaped swelling of the joints between the first and second phalanges, and the swelling in connection with the wrist and metacarpophalangeal joints. (Reprinted with permission from Douthwaite AH, editor: *French's Index of Differential Diagnosis*, ed 7. Williams & Wilkins, Baltimore, 1954.)

The American College of Rheumatology has simplified diagnostic criteria for rheumatoid arthritis, presently requiring four of the following seven clinical features: morning stiffness, arthritis of three or more joint areas, arthritis of the hand joints, symmetrical arthritis, rheumatoid nodules, serum rheumatoid factor, and typical radiographic changes. The first four of these must be present for at least 6 weeks. These criteria provide a sensitivity in excess of 90 percent and a specificity of 89 percent when used with control patients who have rheumatic diseases other than rheumatoid arthritis.

Pathology and Laboratory Examination Synovial hyperplasia and hypertrophy associated with vascular changes, inflammatory edema, and mononuclear cellular infiltration form the pathological hallmarks of early rheumatoid activation. Continuing immunological activation, including T and B lymphocyte infiltration with subsequent B cell transformation to antibody-producing plasma cells, marks the ongoing inflammation of the rheumatoid synovium. Cartilagenous and bony destruction is attributed to the production of various degradative and proteolytic enzymes as well as cytokinetic activation of demineralizing osteoclasts. The systemic sequelae of rheumatoid arthritis, including constitutional symptoms, are ascribed to inflammatory effector molecules produced by the synovium.

No single test offers explicit confirmation of rheumatoid arthritis. However, rheumatoid factors (largely immunoglobulin-M [IgM] autoantibodies) are found in nearly 70 percent of afflicted adults and only 5 percent of healthy adults. With advancing age, the latter percentage increases, and up to 20 percent of adults over 65 may test positive for a rheumatoid factor without clinical evidence of disease. Adding to the potential for misapprehensions, numerous other rheumatic and infectious diseases are associated with the presence of rheumatoid factor. Other laboratory findings that commonly accompany active RA include rheumatoid arthritis elevated acute phase reactants such as C-reactive protein and the erythrocyte sedimentation rate; normochromic, normocytic anemia; thrombocytosis; and turbid synovial fluid with increased protein content and white cell count.

Course and Prognosis Although the course of rheumatoid arthritis is variable, defying prediction for any given patient, more than 80 percent of patients experience persistent disease activity with some joint deformity 10 years after diagnosis. Radiographic evidence of joint damage can precede overt evidence by 6 to 7 years. Although remissions are most apt to occur in the first year, the rate of progression of joint destruction is greatest during the first 6 years, slowing considerably thereafter. Infection, drug therapy, and secondary gastrointestinal bleeding contribute to increased mortality, particularly in patients with especially severe articular disease. Overall, the median life expectancy of individuals with rheumatoid arthritis is estimated to be reduced by 3 to 7 years. Prognosis is somewhat better for patients who experience a more abrupt onset of illness, respond to first-line therapies, and have neither rheumatoid factor nor rheumatoid nodules.

Hormonal influences appear important in women with rheumatoid arthritis. Most women experience amelioration of symptoms during pregnancy, with early recurrence after parturition. Women with prior pregnancies and women taking oral contraceptives share a reduced incidence of the disease. Oddly, although women are clearly more prone to rheumatoid arthritis, female hormones may attenuate disease activity.

Most of the attention given to ascertaining the comorbidity of psychiatric disorders and rheumatoid arthritis has focused upon depression. There are suggestions in the literature that the incidence of anxiety spectrum disorders is mildly increased in patients with rheumatoid arthritis and that the incidence of psychotic disorders is decreased; however, both of these observations are controversial, and study results vary.

A recurring challenge in the determination of the comorbidity of depression with rheumatoid arthritis is the considerable overlap between somatic neurovegetative signs and symptoms suggesting depression and the physical manifestations of rheumatoid arthritis itself. Attempts to excise the diagnostic influence of such symptoms have supported the view that depression is comorbid with rheumatoid arthritis in about 20 percent of individuals, a rate approximating that in other chronic illness populations but about three times the general community rate. Studies attempting to delineate concomitant factors serve to expose associations but do not reliably imply causality, as these efforts have largely been cross-sectional rather than prospective. One possible exception is the relationship between pain and depression; increasingly, it appears that when pain and depression surface in concert, pain antedates development of mood disruption.

A linear relation between raw measures of disease activity and depression does not exist. Collateral factors that do influence the experience of depression include pain (both as reflected in self-report instruments and as documented with observed pain behaviors), perception of unpredictability regarding the clinical course of rheumatoid arthritis, degree of functional impairment (with particular emphasis on the loss of valued activities rather than sterile measures of strength or mobility), individual coping style, and the quality of environmental support, especially from significant relationships. Among individuals with rheumatoid arthritis, depression is more frequent in those who are not married, whose duration of disease is greater, and who have a higher occurrence of medical comorbidities.

Patients' preconceptions and expectations regarding their likely clinical course, their perception of the extent to which they can exercise some control over this course, and their personal cognitive style are important modifiers of their experience of the illness and their potential for depressive complications. Patients who tend to be passive, avoidant, and inordinately prone to fantasy, emotional expressivity, and catastrophization regarding their illness fare less well (with regard to experience of depression) than those who demonstrate active involvement and a problem-solving focus toward their disease.

The presence of depression in patients with rheumatoid arthritis deserves vigorous clinical attention as it may significantly affect clinical course. Individuals with rheumatoid arthritis and depression (compared with patients with similar objective measures of arthritic activity without depression) commonly demonstrate poorer functional status, an increased number of reported painful joints, more pronounced experience of pain, increased health use, increased bed days, and increased inability to work.

Treatment Pharmacotherapy for rheumatoid arthritis is stratified in four broad camps: anti-inflammatory agents, disease-modifying drugs, immunosuppressive agents, and experimental therapies. Aspirin, nonsteroidal anti-inflammatory drugs (NSAIDs), and glucocorticoids effectively decrease acute signs and symptoms but they offer little benefit in retarding disease progression. In contrast, although they offer little anti-inflammatory or analgesic benefit, disease modifiers such as gold compounds, antimalarial agents, penicillamine (Cuprimine), and sulfasalazine (Azulfidine) offer both clinical and serological evidence of benefit. In patients who cannot tolerate these agents or whose disease has progressed despite these medications, a third strategy uses immunosuppressive agents such as azathioprine (Imuran), cyclophosphamide (Cytoxan), methotrexate (Folex), or even cyclosporine (Sandimmune). Experimental approaches include total lymphoid irradiation, lymphoplasmapheresis, or administration of monoclonal antibodies to T cells. The conventional approaches are all associated with potential toxicity and they cannot ensure sufficient disease amelioration or arrest to satisfy each individual with rheumatoid arthritis. This has prompted diverse suggestions, many of which are expensive and avidly promoted, from the alternative medicine community, but none of these have weathered close scrutiny.

Potential psychiatric side effects of commonly used medications for rheumatoid arthritis are numerous. Cognitive dysfunction and delirium may accompany toxic levels of salicylates. The NSAIDs cross the blood-brain barrier and contribute to blockade of prostaglandin production, which in turn affects the function of various neurotransmitters. NSAIDs have been associated with cognitive impairment, frank delirium, depression, mania, and psychotic symptoms. Geriatric patients are especially susceptible to these adverse effects. Mood lability, sleep disruption, delirium, and psychosis are well-recognized potential effects of the glucocorticoids and are often dose sensitive. Patients who require corticosteroids for effective management but tolerate them poorly may benefit from concomitant prophylaxis with lithium (Eskalith, Lithobid). A variety of NSAIDs increase serum lithium concentrations and require cautious monitoring to avoid lithium intoxication. Antimalaria agents are associated with delirium, depression, depersonalization, and psychotic symptoms, but the gold compounds enjoy a benign history with respect to psychiatric adverse

effects. Immunosuppressive agents are typically used in low doses for patients with rheumatoid arthritis, enabling most patients to tolerate them without significant psychiatric sequelae.

In addition to pharmacotherapy, these patients commonly benefit from physical therapy guidance to balance appropriate elements of rest and exercise, as well as possible use of occasional splinting or reliance upon orthotic devices. With more advanced disease, surgery can offer significant relief of pain, correction of deformity, and consequent improvement in functional capacity.

The psychiatrist caring for patients with rheumatoid arthritis will do well to ascertain the patient's opportunities for education about the disease. Maximizing a patient's understanding of rheumatoid arthritis addresses the potent desire for control and thereby alleviates some measure of the vulnerability these patients often experience because of the unpredictable course of the disease. Frank discussion of compliance with medical pharmacotherapy is indicated, as multiple studies have shown wide variations in adherence to medical recommendations, particularly among patients with disease of longer duration and greater severity. Generally, NSAIDs are those most frequently ignored. Cognitive-behavioral therapy has consistently demonstrated effectiveness in reducing both pain behaviors and patients' self-report of pain. Studies also underscore the value of social support in ameliorating the experience of pain and functional impairment.

Psychotropic agents selected for use in patients with rheumatoid arthritis should minimize the side effects that prove especially problematic for individuals contending with this disease. Sleep, which is often disrupted by pain, can be assisted by the combination of an NSAID and trazodone (Desyrel), with appropriate cautionary advice regarding orthostasis to patients who face decreased mobility. There are no studies demonstrating differential efficacy of one antidepressant agent versus another. However, the tricyclic drugs exert mild anti-inflammatory effects independent of their mood-altering benefit. Unfortunately, anticholinergic effects (prominent among the tricyclic drugs and also present with the serotonergic agent paroxetine [Paxil]) can aggravate the significant difficulties with dry oral and ocular membrances in patients whose rheumatoid arthritis is associated with Sjögren's syndrome. Patients with rheumatoid arthritis who also require mood stabilization may experience fewer fluctuations in serum concentrations with use of carbamazepine (Tegretol) or valproic acid (Depakote), as the NSAIDs can effectively increase lithium levels and toxicity by increasing renal tubular reabsorption of the lithium ion. Gold compounds, as well as amyloid or vasculitic involvement of the kidneys, can impair renal function and alter serum lithium concentrations. On the other hand, carbamazepine can induce hepatic metabolism of cyclosporine and is therefore not ideal in that combination. The choice of a lopamine receptor antagonist for a patient with rheumatoid arthritis who develops delirum can be challenging; the lower-potency agents present heightened anticholinergic toxicities, whereas the high-potency agents may provoke extrapyramidal effects that can be both difficult to recognize in a relatively immobile rheumatoid patient and especially threatening for a patient who can ill-afford further compromise of flexibility and mobility. The serotonin-dopamine antagonists, such as risperidone (Risperdal) or olanzapine (Zyprexa), may offer appropriate efficacy with less risk to such patients.

SYSTEMIC LUPUS ERYTHEMATOSUS

Systemic lupus erythematosus is a connective tissue disease of unclear etiology, characterized by recurrent episodes of destructive inflammation of several organs including the skin, joints, kidneys, blood vessels, and central nervous system (CNS). *Lupus* is the Latin word for wolf, and the choice of this term to designate the group of patients suffering from combination of skin, joint, nervous system, and vascular manifestations of this disease seems unfortunate in retrospect, for only rarely today do the facial edema and skin changes that occur in this disease cause its victims to take on a lupine visage. They are more likely to take on the appearance of Cushing's disease because of the high-dose corticosteroids used in treatment.

Epidemiology In the United States, systemic lupus erythematosus is estimated to appear in between 1.8 and 4.6 persons per 100,000 population per year and to be present in 0.1 percent of women at any one time. Its peak incidence occurs in women between the ages of 15 and 25 years.

Etiology Systemic lupus erythematosus is thought to result from dysregulation of the immune system, which results in production of antibodies against the self. The deposition of immune complexes along with complement leads to inflammation. Why this occurs is likely to become apparent only as more is learned about the cellular biology of the immune system.

Diagnosis and Clinical Features The diagnosis of systemic lupus erythematosus is not a simple matter, particularly early in the course of the disease before the characteristic immunological markers become clearly abnormal. In a manner similar to what is contained in the American Psychiatric Association's DSM, the American Rheumatism Association formulated and in 1982 revised criteria for the diagnosis of systemic lupus. A person is said to have the disease if any four of the following are or have been present: malar rash (Fig. 25.8-2), discoid rash, photosensitivity, oral ulcers, arthritis, serositis (pleuritis or pericarditis), renal disorder (proteinuria or cellular casts), neurological disorder (seizures or psychosis), hematological disorder (anemia, leukopenia, thrombocytopenia), immunological disorder, abnormal titer of antinuclear antibody. Fatigue commonly associated with active systemic lupus erythematosus has been found to correlate with other conditions found occasionally without systemic lupus erythematosus such as fibromyalgia.

FIGURE 25.8-2 Woman with lupus erythematosus malar rash. (Courtesy of M. Kevin O'Connor, M.D.) (See Color Plate 8.)

Course and Prognosis The course is highly variable, and prognosis for the most part rests on whether there is involvement of the kidney or other vital organs as well as responsivity to immunosuppressant treatments.

Treatment Biological treatment of systemic lupus erythematosus is determined by the severity of the disease and whether or not there is involvement of a vital organ such as the kidney. NSAIDs, salicylates (particularly nonacetylated salicylates), and antimalarial agents are used in the management of the arthritis and dermatitis seen in lupus. Glucocorticoids and the immunosuppressant agents cyclophosphamide (Cytoxan) and azothioprine are reserved for the more severe life-threatening forms of the disease.

This disease is highly unpredictable, often incapacitating, potentially disfiguring, and not infrequently death dealing, and its treatment requires administration of potentially toxic drugs. It is difficult to conjure up a more demanding challenge to the adaptational capacities of systemic lupus erythematosus patients, many of whom are adolescents or young adults. Appropriate application of the biopsychosocial model that emphasizes attending to the interplay between the biological and psychosocial elements of a disease such as lupus seems to hold out the greatest opportunity for positive therapeutic intervention.

The psychiatrist involved in a systemic lupus erythematosus or rheumatological treatment program would be interested in promoting positive interactions between patients and the program staff and ensuring a tolerant attitude on the part of these staff members. A high-quality psychoeducational program to help patients acquire the knowledge and maturity necessary to deal as effectively as possible with their disease is essential. Installation of hope, avoidance of demoralization, and finding of meaning in the experience of such an illness all seem necessary to having the patients live as well as possible with this often chronic and disabling illness. No controlled studies exist indicating the best programmatic approaches for this patient population. The needs of lupus patients are highly individualized, and it is difficult to imagine that one monolithic approach would answer all of the difficulties encountered by them. Working through the affects associated with incapacity and loss is important. How one goes about this involves considerable skill in brief psychotherapy.

The consulting psychiatrist is asked to see such patients for three broad reasons. One is an adverse psychological response to the disease, characterized by poor compliance with medical regimens and other self-defeating behaviors. Adjustment level symptoms of anxiety and minor depression are commonly encountered. The

need for close medical scrutiny, limitations on personal freedom, mandated avoidance of exposure to the sun, and daily use of sometimes toxic medications all require a level of personal maturity and acceptance that is difficult to achieve under the best of circumstances, particularly by individuals struggling with other major developmental issues and concerns in their lives. Treatment consists of the known effective interventions for adjustment disorders, primarily education, support, and processing emotional responses associated with loss, as well as facilitating a degree of mastery and deriving meaning in the midst of difficult life circumstances. The judicious use of psychoactive medications for symptomatic relief has a place here as well.

The psychiatrist may also be consulted because of the neuropsychiatric complications of direct involvement of the CNS. Lupus causes small vessel vasculitis in the brain. Microhemorrhages and microinfarcts are common pathological findings at autopsy. Cognitive impairment, psychotic symptoms, and florid delirium with marked behavioral abnormalities are all encountered in lupus cerebritis patients. No reliable or sufficiently sensitive test exists to confirm or exclude CNS involvement. The diagnosis is often made by history. The appearance of new symptoms or intensification of preexisting symptoms in sites outside the CNS that coincides with a major behavioral change is one of the more reliable signs suggesting CNS disease. Following the exclusion of CNS infections or other acute processes that could conceivably explain the behavioral and neurological alterations seen in lupus cerebritis, a trial of high-dosage immunosuppressant drug therapy is often administered to determine whether a therapeutic response can be achieved.

Psychiatric treatment is dictated by the predominant clinical picture as it relates to similar primary conditions known to respond to such treatments. No well-designed control trials exist indicating superiority of one treatment over another. Judicious use of antipsychotic agents, mood-stabilizing agents, or both in patients with hallucinations and delusions is reasonable and reportedly effective. However, like most psychiatric disorders thought to be due to general medical conditions, the response of these syndromes to treatment and comparison with similar or primary psychiatric illness has not been determined in longitudinal comparison trials.

The third reason for a psychiatric consultation is psychiatric complications of the treatment of lupus. The immunosuppressant agents used to modify autoantibody production that seems integral to lupus disease activity can in and of themselves cause psychiatric illness including depression, hypomania, delirium, and various psychotic states. NSAID have been reported to cause delirium as well as mild forms of cognitive dysfunction. Adrenal corticosteroid hormones are notorious for causing psychopathology.

OTHER CONNECTIVE TISSUE DISEASES

A number of connective tissue disorders (including Sjögren's syndrome, periarteritis nodosa, and rheumatoid arthritis) can be complicated by inflammation of the blood vessels supplying the brain. These are associated with a variety of neuropsychiatric syndromes, particularly in patients with the more severe form of vasculitis. When the CNS is involved, a number of psychiatric symptoms can appear. Treatment is directed as for a primary psychiatric illness. Symptomatic treatment with known psychoactive drugs, protection from harm and injury, and support in either the hospital or the outpatient setting are the most useful interventions. However, virtually all of the connective tissue diseases create limitation, uncertainty, and occasional psychological dysfunction. Close and careful medical assessment and follow-up is needed. These diseases can create considerable turmoil in the lives of their victims.

FIBROMYALGIA

Fibromyalgia is a nonspecific disorder characterized by many diffuse complaints including pain, stiffness, tender muscles and joints, overwhelming fatigue, distress, and sleep disturbances. Diffuse musculoskeletal pain has been described throughout the history of mankind, but the symptoms were clustered into a named syndrome within the past 100 years or so. In 1904, William Gowers hypothesized that the chronic back pain resulted from an inflammatory change in the fibrous tissue of the muscles and the back producing “muscular rheumatism.” In 1976, the term “fibromyalgia” was introduced by P. Kahler Hench.

The syndrome of fibromyalgia has been called many names over the years, including fibrositis, fibromyositis, myofibrositis, and myofascial pain syndrome. As a result, in 1990 the American College of Rheumatology (ACR) established criteria defining fibromyalgia (Table 25.8-1), standardization that allows the syndrome to be classified and studied.

I. Widespread Pain
Pain must be present for 3 months, widespread and not localized to one area. Involvement includes left and right side of the body, above and below the waist, and axial-skeletal pain.

II. Presence of 11 of 18 tender-point sites
Digital palpation must elicit pain in at least 11 of possible 18 tender point sites. These bilateral sites include occiput, lower cervical, trapezius, supraspinatus, second rib, lateral epicondyle, gluteal, greater trochanter, and knees.

Table 25.8-1 The 1990 American College of Rheumatology Criteria for the Classification of Fibromyalgia

Epidemiology The consensus document on fibromyalgia from the second world conference on myofascial pain and fibromyalgia, Copenhagen, 1992, found fibromyalgia to be the most common cause of widespread musculoskeletal pain, with prevalence rates varying between 0.7 and 3.2 percent. Using the ACR criteria for fibromyalgia, Frederick Wolfe and coworkers estimated the point prevalence of fibromyalgia in a randomized, population-based sample of individuals living in Wichita, Kansas, to be 2 percent overall (3.4 percent for women and 0.5 percent for men). Progressively higher rates have generally been noted in medical clinics and rheumatology clinics. The prevalence of fibromyalgia increases with age up to age 80 in both sexes.

Additional findings from this study included a stronger than expected association with failure to complete high school, reduced household income, divorced status, and increased visits to physicians. Psychological associations were found with elevated T scores on somatization, depression, and anxiety subscales of Derogates' Symptom Checklist-90-Revised (SCL-90-R). Depression and anxiety subscales on the Arthritis Impact Measurement Scales (AIMS) were also elevated.

Etiology The etiology and pathogenesis of fibromyalgia are unknown. Extensive histochemical and biochemical evaluations of musculoskeletal and neurological systems in patients with fibromyalgia have demonstrated only findings consistent with inactivity and disuse. However, these findings may only represent the final outcome of a more complicated neuropsychiatric process.

Possible clues pointing to a more complex etiological picture are found by exploring the associated symptoms of fibromyalgia, especially sleep disturbances, fatigue, and pain. Sleep deprivation can produce transient symptoms of fibromyalgia such as fatigue. Inactivity and deconditioning may render the patient more susceptible to muscle pain. This creates a vicious cycle and a downward spiral producing significant disability.

Indeed, sleep studies reveal that patients with fibromyalgia syndromes have alpha rhythm electroencephalogram (EEG) abnormalities during non–rapid eye movement sleep, increase in stage 1, a reduction in delta sleep, and an increased number of arousal. Other conditions such as periodic limb movement, obstructive sleep apnea, inflammatory joint diseases, some psychiatric illnesses, and noxious environmental stimuli can produce similar sleep abnormalities that are associated with fatigue, pain, and emotional distress. Harvey Moldofsky has postulated that serotonin metabolism, which is associated with sleep and wake cycle and pain, may be connected to sleep disturbances and affective disturbances found in patients with fibromyalgia.

Given the numerous psychological and psychiatric features that have been associated with fibromyalgia, many psychosomatic etiologies have been investigated. Don Goldenberg reviewed the literature on the relationship between fibromyalgia to psychiatric disorders. He found that studies using the Minnesota Multiphase Personality Inventory-2 (MMPI-2) demonstrated elevations in hypochondriacal and hysteria (less so depression) subscales similar to those of other chronic pain patients. Additional evaluation using a number of psychological instruments including the SCL-90-R, the Spielberger State Trait Anxiety Inventory, and the Beck Depression Inventory have found that fibromyalgia patients do not differ significantly from medically ill patients, especially those with pain. However, patients with fibromyalgia have a higher lifetime prevalence of major depression, panic disorder, obsessive-compulsive disorder, and somatization disorder than the general

population.

Diagnosis and Clinical Features Diagnosis of fibromyalgia can be difficult. Fibromyalgia is really a syndrome rather than an entity with clear signs and symptoms, laboratory findings, and specific treatments. The signs and symptoms include widespread pain with tenderness at multiple specified anatomical sites. Additional features may include sleep disturbance, fatigue, complaints of weakness, headaches, cold sensitivity, paresthesia or dysesthesia, swelling, Raynaud's phenomena, restless legs, exercise intolerance, and irritable bowel and bladder. Psychological abnormalities, especially depression, anxiety, and chronic pain with functional disability, often develop and aggravate the condition. No one sign or symptom is diagnostic, and no laboratory data including serological studies, electromyography (EMG), or muscle or nerve biopsies exist to confirm the diagnosis of fibromyalgia.

Because of the diagnostic difficulties, fibromyalgia is often confused with a number of other conditions (Table 25.8-2). The evaluation should focus on ruling out disorders with similar features. Evaluations should include a complete history and examination; laboratory assessment including electrolytes, calcium and magnesium concentrations, serum creatinine, liver function tests, complete blood count, sedimentation rate, antinuclear antibody, thyroid-stimulating hormone test and (when appropriate) MMPI-2 or a similar psychological screening test. Additionally, more extensive but specific serological markers for specific diseases such as cortisol concentrations or viral titers, X-rays, EMG, and even muscle or nerve biopsies should be considered when the signs or symptoms point in a specific direction.

Musculoskeletal disorders
Rheumatoid arthritis
Osteoarthritis
Polymyalgia rheumatica
Giant cell arteritis
Polymyositis/dermatomyositis
Systemic lupus erythematosis
Metabolic-endocrine myopathies
Hypothyroidism, hyperthyroidism
Hypoparathyroidism, hyperparathyroidism
Hypocorticalism, hypercorticalism
Drug induced (alcohol, steroids)
Paraneoplastic syndrome associated with cancer
Infections especially viral infections
Psychiatric disturbances
Dysthymic disorder
Generalized anxiety disorder
Somatization
Chronic pain syndrome

Table 25.8-2 Differential Diagnosis of the Fibromyalgia Patient

Course and Prognosis The natural history of the syndrome is not well delineated. Therefore, the course and prognosis are quite variable. Reports range from complete remission to total disability, with no identifiable factors differentiating those who will improve from those who will have a chronic disabling condition.

Treatment No single medical or psychiatric intervention has been shown to be uniformly effective for fibromyalgia. The current approach to the patient with fibromyalgia combines supportive counseling, behavioral modification, education, physical conditioning, and limited pharmacological interventions. Regular evaluations to identify and address any concomitant aggravating conditions such as sleep disturbance and depression should be performed routinely.

Control trials of current pharmacological and nonpharmacological interventions have provided hope for alleviating some of the suffering associated with this condition. Trials have shown beneficial effects with amitriptyline (Elavil), clomipramine (Anafranil), dothiepin, alprazolam (Xanax), and cyclobenzaprine (Flexeril). Optimal dosages and duration of treatment have not been fully established. Nonpharmacological interventions have been studied in controlled trials. Those noted to improve overall status include cardiovascular fitness, education and physical training, EMG biofeedback, and hypnotherapy.

OTHER ISSUES

Bone Mineral Density and Depression Major depressive disorder has been associated with neuroendocrine abnormalities, especially hypothalamic dysfunction. The changes include hypercortisolism, diminished secretion of growth hormone, hypothalamic hypogonadism, and anorexia. These changes have also been associated with changes in the bone mineral density in women.

David Michelson and colleagues studied the relationship between past or current history of major depression and bone mineral density in 48 women (24 with history of depression and 24 without). They found that the women with history of depression had lower bone mineral density. Additionally, these women had higher urinary cortisol excretion, lower serum osteocalcin concentration, and lower urinary excretion of dexoypyridinoline. This study highlights the musculoskeletal problems—specifically lower bone mineral density—that people with psychiatric problems can develop. Along with depression, lower bone mineral density has been reported for schizophrenia, anorexia nervosa, and other psychiatric conditions. The underlying pathophysiology is not clear, but the implications may be quite significant. Many with psychiatric conditions are young. Studies such as this one suggest a higher lifetime risk of complications such as bone fractures from the decreased in bone mineral density associated with depression.

Steroid-Induced Psychiatric Problems In addition to the problems of adaptation and the neuropsychiatric comorbidity encountered in patients with musculoskeletal diseases that involve the CNS, considerable psychiatric morbidity is associated with the treatment of these illnesses, particularly by administration of corticosteroid hormones. The adrenal corticosteroids are used widely in the treatment of most connective tissue diseases to modulate the production of autoantibodies. Corticosteroids have been known to cause significant psychopathology for nearly 50 years. Given the profound effect of corticosteroids on cellular metabolism including that of the neuron, it is perhaps surprising that their rate of psychiatric complication is not higher than the reported 17 percent.

Clinically, the observed psychopathology tends to defy precise criteria-based classification such as that in DSM-IV. A mixture of affective, psychotic, and neurocognitive deficit symptoms are often encountered. More rarely, pure forms of major depressive disorder or bipolar disorder can be seen. One of the more interesting paradoxes noted in the literature is that excessive endogenous production of corticosteroids seen in such disorders as Cushing's syndrome is more likely to cause depression, whereas exogenous administration of pharmacological doses of corticosteroids often produces euphoria resembling that in manic depressive illness. The risk of inducing severe psychopathology associated with the administration of corticosteroids is dose related but is highly variable and unpredictable. Patients known to have had serious psychiatric complications from the use of corticosteroids in the past may not be at increased risk when retreated with the same agent. A past history of psychiatric disorders does not seem to increase the risk of an adverse reaction.

SUGGESTED CROSS-REFERENCES

Some aspects related to musculoskeletal disorders and rheumatoid arthritis are discussed in different contexts in Section 1.11 on psychoneuroendocrinology, Section 1.12 on the immune system and central nervous system interactions, Chapter 16 on somatoform disorders, Section 25.1 on the history, classification, and current trends in psychosomatic medicine, Section 25.9 on stress and psychiatry, and Section 25.10 on behavior and immunity.

SECTION REFERENCES

Abdel-Nasser AM, Abd El-Azim S, Taal E, El-Badawy SA, Rasker JJ, Valkenburg HA: Depression and depressive symptoms in rheumatoid arthritis patients: an analysis of their occurrence and determinants. Br J Rheumatol *37:*391, 1998.

Anderson KO, Bradley LA, Young LD, McDaniel LK, Wise CM: Rheumatoid arthritis: Review of psychological factors related to etiology, effects, and treatment. Psychol Bull *98:*358, 1985.

Beckham JC, D'Amico CJ, Rice JR, Jordan JS, Divine GW, Brook WB: Depression and level of functioning in patients with rheumatoid arthritis. Can J Psychiatry *37:*539, 1992.

Buchi S, Sensky T, Allard S, Stoll T, Schnyder U, Klaghofer R, Buddeberg C: Sense of coherence—protective factor for depression in rheumatoid arthritis. J Rheum *25:*869, 1998.

Creed F, Murphy S, Jayson MV: Measurement of psychiatric disorder in rheumatoid arthritis. J Psychosom Res *34:*79, 1990.

Crofford LJ: Neuroendocrine abnormalities in fibromyalgia and related disorders (Review). Am J Med Sci *315:*359, 1998.

*Fessler BJ, Boumpas DT: Severe major organ involvement in systemic lupus erythematosus. Diagnosis and management. Rheum Dis Clin North Am *21:*81, 1995.

Firestein GS: Rheumatoid arthritis: rheumatoid synovitis and pannus. In *Rheumatology,* JH Klippel, PA Dieppe, editors. Mosby, London, 1994.

Goldenberg D: Psychiatric and psychologic aspects of fibromyalgia syndrome. Rheum Dis Clin NA *15:*105, 1989.

Gordon DA, Hastings DE: Rheumatoid arthritis: clinical features. In *Rheumatology,* JH Klippel, PA Dieppe, editors. Mosby, London, 1994.

Gowers WR: Lumbago: Its lesions and analogues. Br Med J *1:*117, 1904.

Grant I, Heaton RK, Marcotte TD: Evaluating the neurocognitive complications of SLE. Lessons from HIV disease. Ann NY Acad Sci *823:*18, 1997.

Halberg P: Rheumatoid arthritis: History. In *Rheumatology*, JH Klippel, PA Dieppe, editors. Mosby, London, 1994.

Hall RCW, Popkin MK, Stickney SK, Gardner Emergency Room: Presentation of the steroid psychoses. J Nerv Ment Dis *167:*229, 1979.

Harding SM: Sleep in fibromyalgia patients: Subjective and objective findings (Review). Am J Med Sci *315:*367, 1998.

Hawley DJ, Wolfe F: Depression is not more common in rheumatoid arthritis: A 10-year longitudinal study of 6,153 patients with rheumatic disease. J Rheumatol *20:*2025, 1993.

Hench PK: Nonarticular rheumatism. Arthritis Rheum *19:*1088, 1976.

*Iverson GL, Anderson KW, McCracken LM: Research methods for investigating casual relations between SLE disease variables and psychiatric symptomatology. Lupus *4:*249, 1995.

Katz PP, Yelin EH: The development of depressive symptoms among women with rheumatoid arthritis. Arthritis Rheum *38:*49, 1995.

Katz PP, Yelin EH: Prevalence and correlates of depressive symptoms among persons with rheumatoid arthritis. J Rheumatol *20:*790, 1993.

Kelly WN, Harris ED, Ruddy S, Smedge BC: *Textbook of Rheumatology, ed 5*. Saunders, Philadelphia, 1997.

Lorenzen I: Fibromyalgia: A clinical challenge. J Intern Med *235:*199, 1994.

Maini RN, Feldmann M: Immunopathogenesis of rheumatoid arthritis. In *Oxford Textbook of Rheumatology,* ed 2, PJ Maddison, editor. Oxford University Press, New York, 1998.

Mason LW, Gool Kasian P, McCain GA: Evaluation of multimodal treatment program for fibromyalgia. J Behav Med *21:*163, 1998.

Michelson D, Stratakis C, Hill L, Reynolds J, Gallivon E, Chrousos G, Gold P: Bone mineral density in women with depression. N Engl J Med *335:*1176, 1996.

Moldofsky H: Fibromyalgia, sleep disorder, and chronic fatigue syndrome (review). Ciba Found Symp *173:*262, 1993.

Rogers MP: Rheumatoid arthritis: Psychiatric aspects and use of psychotropics. Psychosomatics *26:*915, 1985.

Rome HP, Braceland FJ: Psychological response to corticotropin, cortisone, and related steroid substances. JAMA *148:*27, 1952.

Simms RW: Fibromyalgia is not a muscle disorder Am J Med Sci *315:*346, 1998.

Smedstad LM, Vaglum P, Kvien TK, Moum T: The relationship between self-reported pain and sociodemographic variables, anxiety, and depressive symptoms in rheumatoid arthritis. J Rheumatol *22:*514, 1995.

Smedstad LM, Vaglum P, Moum T, Kvien TK: The relationship between psychological distress and traditional clinical variables: A 2-year prospective study of 216 patients with early rheumatoid arthritis. Br J Rheumatol *36:*1304, 1997.

Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, Schaller JG, Talal N, Winchester RJ: Special article: The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum *25:*1271, 1982.

Wang B, Gladmon DD, Urowitz MB: Fatigue in lupus is not correlated with disease activity. J Rheumatol *25:*892, 1998.

*West SG: Lupus and the central nervous system. Curr Opin Rheumatol *8:*408, 1996.

White K, Harth M: An analytical review of 24 controlled clinical trials for fibromyalgia syndrome. Pain *64:*211, 1996.

Wolfe F, Ross K, Anderson J, Russel I, Hebert L: The prevalence and characteristics of fibromyalgia in the general population. Arthritis Rheum *38:*19, 1995.

*Wolfe F, Smythe HA, Yunus MB, et al: The American College of Rheumatology 1990 Criteria for Classification of Fribromyalgia: Report of the Multicenter Criteria Committee. Arthritis Rheum *33:*160, 1990.

*Wollheim FA: Rheumatoid arthritis-the clinical picture. In *Oxford Textbook of Rheumatology,* ed 2, PJ Maddison, editor. Oxford University Press, New York, 1998.

Wright GE, Parker JC, Smarr KL, Johnson JC, Hewett JE, Walker SE: Age, depressive symptoms, and rheumatoid arthritis. Arthritis Rheum *41:*298, 1998.

25.9 STRESS AND PSYCHIATRY

JOEL E. DIMSDALE, M.D., FRANCIS J. KEEFE, ĦH.D., AND MURRAY B. STEIN, M.D.

Stress is one of the old war-horses of psychiatry. It has been invoked as a cause of major psychopathology, a precipitator or trigger of psychiatric illness, and a contributor to considerable mental anguish. Similar roles have been cast for stress vis-a-vis illness throughout medicine. Many of the suggested links are intuitively appealing because somatic distress is commonly experienced in the face of unpleasant events and thoughts. However, the removal of stress is frequently associated with the cessation of diverse physical symptoms. At Appomatox, Ulysses S. Grant had an incapacitating migraine that disappeared instantaneously on the morning when General Lee announced his intention to surrender. Such instantaneous associations between stress and various symptoms are obvious, but whether stress causes an illness is much more difficult to demonstrate.

The slipperiness of the term “stress” is both an advantage and a disadvantage. It is an advantageous term because it is all-embracing and allows a synthesis of many aspects of life's experience; it is disadvantageous because its referents are diverse. *Stress* refers both to a subject and a predicate, an event and the consequences of that event. For clarity, it is probably wise to eschew use of the word “stress” altogether and to use different terms for its various meanings. Thus, it is much clearer to refer to stressful events or stressors and the resulting strain or stress responses incurred.

HISTORY

Stress was coined by Hans Selye (1907–1982) who observed that many highly diverse ways of perturbing the organism resulted in common physiological responses. The concept was unusual at the time because medical research focused on specificity—that specific pathogens elicited unique pathological responses. Selye found that highly diverse stimuli seemed to result in similar changes. From animal studies he reported that there is a general adaptation system to perturbation. He called it general because the response was similar across diverse stimuli; “adaptation” because he believed that the physiological changes represented the organism's efforts to respond to the stressor; a “system” because these responses were well orchestrated, involved diverse organ systems, and followed certain phases. Selye invoked the adrenocortical system as the crucial responder to stressful stimulation. He observed that any novelty or perturbation of the system was associated with an elevation of adrenocortical activity, at least transiently, and that if stressors were unremitting, diverse pathological changes would be evident. Like all groundbreakers, Selye was guilty of oversimplification, a weakness that has become abundantly clear with the accumulation of additional empirical data. Not all organisms respond to all perturbations with the general adaptation system. However, his insights were crucial in shaping the field. As a result of his prolific scholarship, the term “stress,” has become so popular that it is used constantly, even when the researchers would prefer more precision.

The other groundbreaker in the field of stress research was Walter Cannon (1871–1945). Although he made major contributions in gastroenterology and roentgenology, his long-abiding interest was in the physiology of emotion and the autonomic nervous system. Cannon methodically investigated the other great pathway of stressful responses, the sympathetic nervous system. Cannon is noted for his studies of gastric erosion and of cardiovascular response; he also focused on more immediate or short-term responses to stressors.

There are earlier antecedents to the concept of stress. For centuries physicians and patients have made the association between adverse life events and illness. In law, for instance, there is a long history of tort cases alleging that emotional arousal secondary to a stressful event causes miscarriage. The links between emotion and sudden cardiac death have been repeatedly noted in sources as diverse as the Bible, anthropological texts, and clinical experience. Such links were repeatedly observed by crucial figures in the history of medicine. William Harvey in the seventeenth century and William Osler in the nineteenth century frequently allude to the relationship between adverse life events and illness onset. Many contemporary cultures regard illness as the outcome of being out of balance with the environment and its demands, a manifestation of inadequate coping with diverse stressors.

The evidence for this relation is hardly unanimous. Physicians who see their patients briefly and from a highly specialized perspective are rarely familiar with the nuances of their patient's lives, and thus it is not surprising that many specialists deny the existence of such connections. On the other hand, researchers criticize the oversimplification of most studies that examine stress and illness.

A major conceptual problem that has hindered progress in this field is the confusion of *cause* and *trigger* of illness. The allegation that a stressor “caused” an illness is difficult to demonstrate in most human studies, which impose a short-term stressor and examine the resultant physiological changes. Illnesses such as myocardial infarction usually become evident only after a long accumulation of coronary artery disease. As a result, how can one say that “stress caused a myocardial infarction”? There is, however, an alternative hypothesis that provides very strong evidence for a link between stressors and illness. This argument emphasizes that stressors trigger illness in individuals whose underlying illness is latent or subclinical. Such a case is demonstrated routinely in psychophysiology and in psychosomatic medicine studies. For instance, studies of the physiological response to stressors in normal individuals demonstrate large-amplitude changes in physiological levels of blood pressure, gastric motility, and hormonal levels. When such psychophysiological studies are extended to patient groups, incontrovertible evidence is seen for the role of stressors as trigger events in eliciting pathophysiological phenomena. For instance, even discussing emotional conflicts can elicit life-threatening arrhythmias in vulnerable individuals (Fig. 25.9-1).

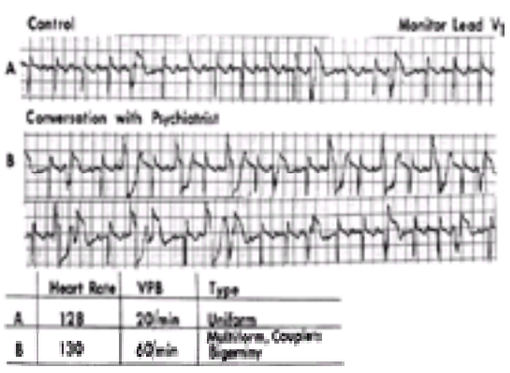

FIGURE 25.9-1 Impact of stress on electrocardiogram. (Reprinted with permission from Lown B: Sudden cardiac death: Biobehavioral perspective. Circulation *76* [Suppl]:I186, 1987.)

Another key conceptual problem centers on the issue of vulnerability versus resilience or stress versus coping. Following a long medical tradition, researchers initially focused on the virulent stressor, treating it like a virulent pathogen that would infect all individuals despite their varied life experiences. Thus, the initial stress research examined the physiological consequences of stressor imposition. The variability in the data motivated investigators to study the subtleties of how people interpreted the stressful challenges. Initially these interpretations were framed in terms of defensive styles, that is, how individuals guard themselves against painful emotional stressors. Subsequently, the inquiry broadened and investigators examined individuals' defensive behaviors, information-seeking behaviors, affiliative behaviors, and overall problem-solving behaviors. This heterogeneous combination is referred to as coping behavior. One of the clearest explications of this complex

process of stressor stimulation and coping mediation was suggested by Richard Rahe in 1974 (Fig. 25.9-2).

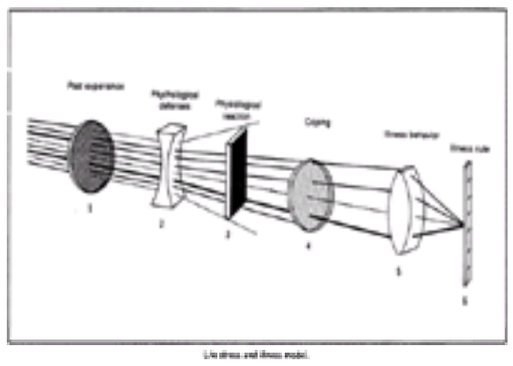

FIGURE 25.9-2 Step 1 in the stress and coping model represents a person's perception of his or her recent life events, which are indicated by light rays of differing intensities. (Thicknesses of the lines indicate the intensities of life-change magnitudes assigned to those events.) Much like a polarizing filter that can reduce or augment selected light rays, a person's perception of the events can alter their magnitudes. Such alterations depend, in part, on a person's prior experience with an event, on current social supports, and on biographical assets, such as income and education. Steps 2 and 3 in the stress and coping model indicate the effects of psychological defenses that can diffract away signals resulting from perceived recent life events that typically stimulate a variety of psychophysiological responses. A negative lens is used in step 2 to represent the potential function of psychological defenses. A box drawn for step 3 indicates the arena of psychophysiological responsivity. Once perceived recent life events can stimulate a number of psychophysiological responses (step 3). If those responses are interpreted by the person and are potentially dangerous to the person's health, they are called symptoms. Step 4 in the model is indicated by a color filter that absorbs light rays of certain frequencies. In much the same way, certain response-reduction (coping) techniques can absorb selected symptoms. The coping activities listed successfully absorbed two of the four symptoms. Two persisting symptoms not absorbed by the coping activities outlined may lead a person to adopt various illness behaviors. Those behaviors entail symptom recognition and concern, beliefs that the medical profession can help, and a willingness to conform to various sick-role expectations. A positive lens in step 5 indicates a patient's focusing attention on persistent symptoms in the presence of a physician, who then arrives at a diagnosis. The data are then entered into a medical record, step 6. (Reprinted with permission from Rahe RH, Floistad I, Bergan T, Ringal R, Gerhardt R, Gunderson E, Arthur R: A model for life changes and illness research: Cross-cultural data from Norwegian Navy. Arch Gen Psychiatry *31:*172, 1974.)

MEASURING STRESS PSYCHOLOGICALLY

Doctors have always asked about the stressors in their patients' lives because obtaining such information is at the core of history taking. “What sorts of things aggravate your symptoms? Are you facing any unusual burdens or challenges?” With certain symptoms such as angina, the link between stressor and illness is immediately obvious. However, it appears that many illnesses not traditionally thought of as stress-sensitive are aggravated by stressful conditions and the mechanisms for this link are complex.

Approximately 70 years ago Adolf Meyer suggested that psychiatrists employ a “life chart” in evaluating their patients. The idea is elegant in its simplicity. A patient's life is plotted on a line, beginning with birth and ending in the present. Various important life events are penciled in, as are the onset or worsening of health problems. The life chart is a very effective way of eliciting a clinical history. Sometimes, recurrent patterns of association between life events and illness episodes are suggested (Fig. 25.9-3).

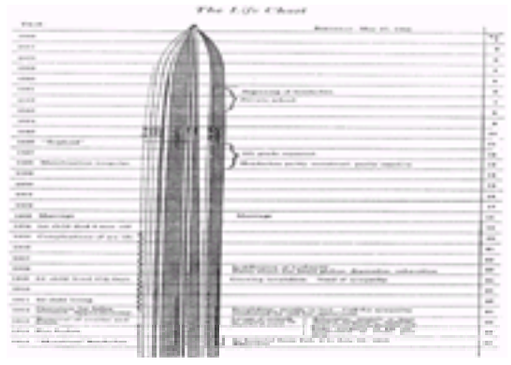

FIGURE 25.9-3 Meyerian life chart. (Reprinted with permission from *The Collected Papers of Adolf Meyer,* vol 3. The John Hopkins Press, Baltimore, 1951.)

The next step in the evolving field of life events was developed by the fruitful collaboration of Thomas Holmes and Richard Rahe who believed that some kind of list of commonly encountered stressors would be more useful than the relatively free-form process of taking an unstructured history. After considerable research they developed a list of 43 life events ranging in severity from death of a spouse to receipt of a traffic ticket (see Table 25.9-1). The checklist or Schedule of Recent Events can be self-administered and requires very little time to complete. It is noteworthy that both positive and negative life events are tallied by the form. The underlying assumption is that all life events, be they positive or negative, impose a demand for adaptation and that such demands are stressful. While at first blush this may seem counterintuitive, it is indubitably true that positive life events like getting married bring with them a host of stressors. Another characteristic of the Holmes-Rahe approach is that the items are weighted; they are not treated as equivalent in adaptational demand. After considerable research, the investigators observed that most people rank stressors similarly and thus it is possible to assign a specific “number” of life change units to the occurrence of each specific stressor. This technique has been used in hundreds of studies and has demonstrated that numerous illnesses follow on the heels of increases in life stress. It is not that the high-scoring individual is at risk, but rather that the individual with an increase in life stress units is at risk for any number of health consequences.

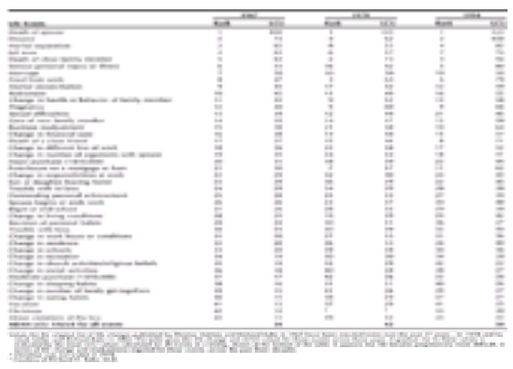

Table 25.9-1 Life-Change Scaling Results Across Time†

There are many criticisms of this approach, and alternative ways of measuring stressful life events have been proposed. The criticisms take one of six forms.

1. Some investigators feel that negative life events carry with them far greater anguish and adaptational pressure, and that the positive life events may be ignored.
2. Some investigators question the utility of adding all the life change units together and coming up with one number to represent the total stressful exposure of the individual. Whereas the Holmes-Rahe approach may be extremely useful for epidemiological research, for clinical use it may be more helpful to have subscales that list the stressors associated with job pressure, home pressure, and others, in separate categories.
3. Some theorists believe that inventories of relatively rarely occurring stressful events are of limited utility, and that more is to be gained by measuring the accumulation of the more common microstressors that individuals encounter frequently.
4. Some investigators doubt the validity of a universal weighting to stressful events and favor allowing individuals to apply their own idiosyncratic assessment of the stressfulness or emotional travail elicited by events.
5. Some investigators believe that the nuances of stressful life events are so complex that they can be identified only by skilled interviewers and that self-administered scales are of questionable validity.
6. Finally, common to all scales that measure stressful events is the problem of retrospective distortion. Patients forget, disavow, deny, and repress. As a result, when patients are asked to retrospectively report life events, they tend to forget events that are shrouded in the past. On the other hand, patients' beliefs sometimes cause them to ‘look for' events harder if they have sustained an illness that they feel is elicited by stress.

SELECTED PHYSIOLOGICAL RESPONSES TO STRESSORS

The body's reaction to stress—in this sense defined as anything (real, symbolic, or imagined) that threatens an individual's survival—is to put into motion a set of responses that seeks to diminish the impact of the stressor. Another way of viewing this is that a stressor disrupts an organism's equilibrium, and the stress response consists of the initiation of physiological systems that seek to react to the stressor, bring about an adaptive response, and restore homeostasis. Much is known about the physiological response to acute stress, but considerably less is known about the response to chronic stress. This represents a major gap in the ability to relate the effects of stress to the occurrence of psychiatric illness in humans. Many stressors occur over a prolonged period of time or have long-lasting repercussions (e.g., loss of a spouse is followed by months or years of loneliness; a violent sexual assault is followed by years of apprehension and worry about walking outside at night or about being alone at home). Animal models that can capture this reverberation or echo of stress are lacking and, as a result, what is known about the effects of real-life stressors on humans is limited. However, what is known about the neuroendocrine and immune response to laboratory stressors is growing apace and will ultimately inform the understanding of why and how stress has a deleterious effect on humans.

Neurotransmitter Responses to Stress Stressors of many kinds activate noradrenergic systems in the brain (most notably in the locus coeruleus) and cause the release of catecholamines from the autonomic nervous system. Prior exposure to chronic stress results in enhanced synthesis of brain norepinephrine, apparently through the induction of tyrosine hydroxylase, the rate-limiting step in catecholamine synthesis. Although an animal exposed chronically to a given stressor may develop tolerance (both behaviorally and in terms of noradrenergic activation) to it over time, this same animal may exhibit an enhanced response to a novel stressor. In other words, chronic stress may dampen the autonomic nervous system response to that particular stress but leave the animal more sensitive to the effects of ensuing stressors. Stressors also activate serotonergic systems in the brain as evidenced by increased serotonin turnover. Recent evidence suggests that although glucocorticoids tend to enhance overall 5HT functioning, there may be differences in glucocorticoid regulation of serotonin-receptor subtypes that may have implications for serotonergic functioning in depression and related illnesses. For example, glucocorticoids may increase serotonin (5-hydroxytryptamine [5-HT] type 2 (5-HT_2)-mediated actions of 5-HT, possibly contributing to the intensification of actions of the central nervous system of these receptor types that have been implicated in the pathophysiology of major depressive disorder. Stress also has the effect of increasing dopaminergic neurotransmission in mesoprefrontal pathway.

Since the late 1980s it has become clear that amino acid and peptidergic neurotransmitters are also intricately involved in the stress response. Studies have shown that corticotropin-releasing factor (CRF) (as a neurotransmitter, not just as hormonal regulator of hypothalamic-pituitary-adrenal axis functioning), glutamate—through *N*-methyl-D-aspartate (NMDA) receptors—and g-aminobutyric acid (GABA)—all play important roles in the generation of the stress response or in modulation of other stress-responsive systems such as dopaminergic and noradrenergic brain circuitry. As these interactions come to be better understood, it is likely that they will enhance the knowledge base of how stress responses are coordinated and individualized for specific stressors.

Endocrine Responses to Stress In response to stress, CRF is secreted from the hypothalamus into the hypophysial-pituitary-portal system. CRF acts at the anterior pituitary to trigger release of adrenocorticotrophic hormone (ACTH). Once ACTH is released, it acts at the adrenal cortex to stimulate the synthesis and release of glucocorticoids. Glucocorticoids themselves have myriad effects within the body, but their actions can be summarized in the very short term as promoting energy utilization, increasing cardiovascular activity (in the service of the “flight-or-fight” response), and inhibiting functions such as growth, reproduction, and immunity.

This hypothalamo-pituitary-adrenal axis is subject to tight negative feedback control by its own endproducts (i.e., ACTH and cortisol) at multiple levels including the anterior pituitary, the hypothalamus, and suprahypothalamic brain regions such as the hippocampus. It is now known that the hypothalamo-pituitary-adrenal axis is also influenced by multiple extrinsic systems, which should not be surprising given the coordinated nature of the stress response. For example, in addition to CRF, numerous secretagogues (i.e., substances that elicit ACTH release) exist that can bypass CRF release and act directly to initiate the glucocorticoid cascade. Examples of such secretagogues include catecholamines, vasopressin, and oxytocin. Interestingly, different stressors (e.g., cold stress versus hypotension) trigger different patterns of secretagogue release, again demonstrating that the notion of a uniform stress response to a generic stressor is an oversimplification.

A relevant feature of the hypothalamo-pituitary-adrenal axis response to stress is that it varies depending on whether the organism (most of these data come from rats) is exposed to the stressor on a single occasion, more than once, or over a prolonged period of time. Interestingly, it has not been possible yet to rely on a model that can predict whether sensitization (an enhanced response to the same stress over time) or *desensitization* (a reduced response to the same stress over time) will occur for a given stressor. Nonetheless, the fact that the stress response can change dramatically depending on the frequency and duration of exposure highlights the necessity of specifying these elements when trying to understand the role of stress in clinical psychiatric illness.

Immune Response to Stress Part of the stress response consists of the inhibition of immune functioning by glucocorticoids. However, this inhibition may reflect a compensatory action of the hypothalamo-pituitary-adrenal axis to mitigate other physiological effects of stress. Stress can cause immune activation through a variety of pathways. CRF itself can stimulate norepinephrine release via CRF receptors located on the locus coeruleus, which activates the sympathetic nervous system, both centrally and peripherally and increases epinephrine release from the adrenal medulla. In addition, there are direct links of norepinephrine neurons that synapse upon immune target cells. Thus, in the face of stressors there is also a profound immune activation, including the release of humoral immune factors (*cytokines*) such as interleukin-1 (IL-1) and IL-6. These cytokines can themselves cause further release of CRF, which in theory serves to increase glucocorticoid effects and thereby self-limit the immune activation.

Life stresses such as divorce, school examinations, and caring for a sick relative have been shown in numerous studies to have a deleterious effect on antiviral immunity (e.g., as manifest in increased herpes simplex virus recurrence rates or as reduced cellular immunity) in humans. In one well-controlled study volunteer subjects who had been exposed to high levels of stress in the preceding year were at greater risk than low-stressed subjects for developing clinical colds following the intranasal administration of respiratory viruses by droplet. Acute stress has been shown in most studies to increase the activity of natural killer cells, which are involved in the early, nonspecific cellularly mediated response to infection. In contrast, the effects of chronic stress on natural killer cell activity are less clearly defined. The tentative consensus about the effects of stress on immune function is that whereas small amounts of stress may enhance immune function, excessive stress seems to impair it. The functional health significance of these stress-related alterations in immunity remains to be determined.

STRESS AND PSYCHIATRIC ILLNESS

Historical Conceptualization Mental disorders have been ascribed to an imbalance in bodily humors, to the influence of external spiritual or other supernatural forces, and to moral or somatic deficiency. In the early to mid-1800s a school of thought emerged, led by Philippe Pinel (1745–1826) of the Salpetriere in Paris and Amariah Brigham (1798–1849) in America, that the expression of mental illness was affected by life circumstances and, more broadly, by societal factors. These various influences converge in the biopsychosocial model proposed by Meyer. This orientation held that an individual's response to stress is modified by a number of intrinsic factors (e.g., genetic vulnerability, premorbid personality) and extrinsic factors (e.g., social support). The model incorporated individual temperamental and experiential characteristics such as potential vulnerability (or resiliency) factors, stressful life events as initiating or exacerbating factors, and a variety of support networks as potentially modifying factors for the occurrence of mental illness. For the most part, this multideterminate model of mental illness dominates modern psychiatric thinking.

The current diagnostic nomenclature of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) does not emphasize this biopsychosocial model of the nature and genesis of psychiatric disorders because of the decision to have the DSM expound a descriptive, phenomenological typology, one that would

be least governed by theoretical or other subjective interpretations. Thus, in DSM-IV, major depressive disorder is major depressive disorder whether it appears in the context of job loss, divorce, or for no apparent reason. There are interesting inconsistencies in this approach. For example, major depressive disorder may not be termed major depressive disorder in DSM-IV when it appears in the context of bereavement. There is, in this instance, an acquiescence to the fact that bereavement so frequently results in a depressive syndrome (albeit temporary and perhaps slightly different in character than non–bereavement-related depression), that it should be differentiated from major depressive disorder.

What if individuals develop major depressive disorder following the loss of their home in an earthquake? According to DSM-IV, if such a person develops a depressive syndrome, it might be considered an adjustment disorder (if short-lived) but, if it is persistent, then a diagnosis of major depressive disorder applies. Why, one might ask, is this scenario treated differently from a depressive syndrome following bereavement? Following a traumatic stressor (e.g., earthquake), one might better diagnose posttraumatic stress disorder. Not necessarily, because although posttraumatic stress disorder is another exception to the DSM-IV rule that the nomenclature should be atheoretical (it is one of several disorders for which the occurrence of a stressful life event is pivotal for diagnosis) it has its own characteristic symptoms that are not equivalent to the depressive syndrome. Thus, like many psychiatrists who struggle to understand the influence of life events on the emotional functioning of patients, DSM-IV continues to struggle with the best way to classify and incorporate stressful life events into diagnosis. In the face of this ongoing debate, however, much research has been conducted that serves to enhance the knowledge base of the relationship between stress and psychiatric illness.

Stress and Psychotic Disorders There is little reason to suspect that stress plays a role in the pathogenesis of schizophrenia. Most current etiological theories of schizophrenia focus on genetic and prenatal environmental factors or a combination thereof. There is, however, ample evidence that adverse life events and a stressful social or familial milieu play an important role in determining the course of illness in general and episodes of relapse in particular.

Chronic interpersonal stress, usually studied in the context of the family, has been shown to be an important risk factor for relapse in schizophrenia. Studies of *expressed* emotion, a proxy for familial tumult, have demonstrated that persons with schizophrenia residing in homes with high expressed emotion tend to relapse at twice the rate of those who live in families with low expressed emotion. Other stressors that plague the lives of many patients with schizophrenia—poverty, homelessness, and criminal victimization—also powerfully affect the clinical course of such patients. Combined with evidence that schizophrenia renders the individual more susceptible and sensitive to the negative effects of even minor stressors, it is obvious that stress plays a major role in the course of schizophrenia. Considerable research has been conducted examining the effects on relapse rates of psychosocial interventions that seek to enhance the ability of schizophrenia patients to cope with stressful events.

Researchers have been interested in determining the neurobiological mechanisms through which stress exacerbates schizophrenia. Schizophrenia has been hypothesized to involve dysfunction within brain neuronal circuits involving dopamine and other excitatory neurotransmitters such as glutamate. Deficits in the cognitive functions of a patient with schizophrenia mediated by the dorsolateral prefrontal cortex have strongly implicated this region of the brain in the pathophysiology of schizophrenia. Abnormal dopaminergic functioning within the mesocortical system (which provides innervation to the prefrontal cortex and cingulate cortex) might underlie some of these deficits in cognitive functioning. Some animal studies suggest that the mesocortical projections to the prefrontal cortex are particularly sensitive to stress-induced activation.

Alan Breier et al. used 2-deoxyglucose (2DG) infusions to induce glucose deprivation and monitored the ensuing physiological stress responses in patients with schizophrenia and healthy control subjects. They found that intravenous 2DG injection resulted in significantly larger increases in plasma levels of the dopamine metabolite homovanillic acid in patients with schizophrenia than in controls. Moreover, these investigators found that patients with the smallest prefrontal cortices were the most susceptible to the 2DG-induced stress, at least in terms of their plasma homovanillic acid response. These data lend support to the hypothesis that dysfunction within prefrontal-subcortical connections in schizophrenia may mediate the susceptibility to stress-induced exacerbation of symptoms seen in this disorder. The extent to which this metabolic stressor represents a suitable analog for psychosocial stressors remains to be determined.

Mood Disorders Recent advances in psychiatric epidemiology suggest that a simple relationship may not exist between stressful life events and mood disorders. Most studies of stress and depression suffer from a number of methodological shortcomings. There is a tendency for depressed patients to selectively recall negative life events and to associate these events with their current difficulties. Although these effects can be minimized by ascertaining life events when patients are not in episodes of depression, there is additional evidence that depressed patients really do experience more negative life events than nondepressed patients. This difference may be accounted for predominantly by an excess of *dependent events*—that is, events that may conceivably have been caused by the individual's actions while depressed (e.g., losing a job due to poor performance, failure in school) as opposed to being independent of the depression (e.g., property loss, family illness). This confound complicates the interpretation of most early studies of life events and depression.

Another way of thinking about the relationship between stress and depression is to focus not on specific life events but rather on the effects of chronic stress. Studies in this genre are few (e.g., examining rates of depression in mothers of chronically ill children compared to mothers of healthy children), but those that do exist tend to support a relationship between chronic stress and the propensity for major depressive disorder. To the extent that most animal models of depression (e.g., chronic restraint, forced swim) are subchronic or chronic stress paradigms, the study of the effects of chronic stress on depression in humans offers the possibility for convergence and cross-fertilization between the preclinical and epidemiological knowledge bases.

There is a fundamental similarity between the neuroendocrine effects of stress on humans and the neuroendocrine abnormalities in major depressive disorder. For over twenty years major depressive disorder has been recognized as associated with overactivity of the hypothalamo-pituitary-adrenal axis, as inferred by increased plasma cortisol and ACTH concentrations, increased urinary free cortisol, increased cerebrospinal fluid cortisol, and cortisol resistance to dexamethasone suppression. A critical question is where in the hypothalamo-pituitary-adrenal axis the physiological abnormality in depression might lie. Challenge studies with corticotropin-releasing hormone (CRH), which have demonstrated reduced ACTH responses to CRH in depressed patients compared to control subjects, suggest that chronic CRH secretion (with compensatory downregulation of CRF receptors at the pituitary level) may be an important feature of major depressive disorder. Current thinking tends to position hypothalamic CRH hypersecretion as the driving force behind the hypothalamo-pituitary-adrenal abnormalities found in major depressive disorder but the neurochemical basis for CRH hypersecretion remains a topic of ongoing research.

Might persons with a genetic diathesis for depression be more susceptible to the depression-inducing effects of stressful life events? Information has been collected about stressful life events and the onset of major depressive episodes in a population-based sample of female-female twins in Virginia. Several life events (e.g., assault, serious marital problems, divorce or breakup) were found to powerfully predict onset of major depressive disorder in the month of occurrence. However, genetic liability was also found to have a significant impact on risk of major depressive disorder. The researchers concluded that the onset of major depressive disorder was best explained by a model that posited genetic vulnerability to the depressogenic effects of stressful life events. Thus, this study provides rare empirical evidence of what has been viewed as a truism in contemporary psychiatric teachings: genes and stressful life events operate in concert to elicit depressive episodes.

Anxiety Disorders In comparison to the voluminous literature on stress and depression, there is less empirical research on the relationship between stress and anxiety disorders. These studies are marred by similar methodological shortcomings found in the literature on depression. Despite these limitations it seems reasonable to conclude that panic disorder frequently has its onset or recrudescence in the context of stressful life events. In particular, there is some evidence to suggest that either interpersonal conflict or serious illness (in a significant other) may trigger the onset of panic disorder in susceptible individuals. Given the fact that patients with panic disorder are often somatically preoccupied and likely to interpret physical sensations as life-threatening, it is not too surprising that the occurrence of illness in someone close to the patient might precipitate the onset or an exacerbation of panic in the patient.

The effects of chronic stress on the course of anxiety disorders has not been empirically studied, although it seems intuitive to expect that a relationship should exist. In terms of the effects of early life stressors, there is growing evidence that certain adverse early life events such as sexual or physical abuse may be risk factors for the later development of panic disorder, particularly in women.

In light of the known effects of stress on hypothalamic-pituitary-adrenal axis functioning, an association between panic attacks—themselves a stressful experience—and hypothalamic-pituitary-adrenal axis activation would be expected. Surprisingly, no significant increase in plasma cortisol levels has been found in association with spontaneous panic attacks or with lactate-induced panic attacks. Moreover, 24-hour blood sampling in patients with panic disorder reveals considerable heterogeneity in plasma cortisol across patients. Taken together, these observations underscore the notion that a typical stress response, from a neuroendocrine perspective, is not seen consistently across psychiatric disorders. This conclusion is strengthened by the findings with regard to hypothalamic-pituitary adrenal axis functioning in another anxiety disorder, posttraumatic stress disorder.

Posttraumatic Stress Disorder DSM-IV recognizes the existence of a group of disorders that are, by definition, stress-related. Included here are posttraumatic stress disorder and, new to DSM-IV, acute stress disorder. The latter has its onset following particularly traumatic (often life-threatening) events such as violent assault or serious accidents, and is denoted by the presence of prominent dissociative symptoms (e.g., derealization, numbing). At present, it is not known whether acute stress

disorder is of prognostic significance in terms of heralding the onset of more chronic psychiatric symptoms, nor is it known what risk factors exist for the development of acute stress disorder itself. In contrast, much is known about the prevalence of posttraumatic stress disorder and risk factors for its development.

Whereas posttraumatic stress disorder was not long ago considered to be a disorder only of war veterans (15 percent of Vietnam veterans may suffer from this disorder), it is now recognized as a highly prevalent disorder affecting a much larger segment of the population. Among the traumatic stressors that can elicit this disorder are relatively rare events such as natural disasters and frighteningly common events such as criminal victimization. In prior editions of the DSM, it was noted that in order to qualify for this diagnosis, the traumatic event needed to be “beyond the range of normal human experience.” Epidemiological research has shown that horrible things happen to many people and must be considered part of the human experience. Consequently, although debate continues, it is now generally accepted that any life-threatening or perhaps, life-altering event can be considered sufficiently traumatic that it is capable of eliciting posttraumatic stress disorder. With this definition in place, recent surveys place the lifetime rate of the disorder in the community at around 10 percent.

The occurrence of posttraumatic stress disorder following severe psychological trauma provides the most persuasive evidence in favor of certain kinds of stress causing psychiatric illness. It is important to note, however, that the neuroendocrine profile of patients with posttraumatic stress disorder is not what one might expect to see following chronic stress, nor is it akin to what is observed in major depressive disorder. Whereas the latter two conditions would be expected to be associated with hypercortisolism, most recent studies of patients with posttraumatic stress disorder have demonstrated rather subtle hypocortisolism. Although the neurobiological implications of this finding remain to be elucidated, the message for stress researchers is clear: there is no longer any hope that a uniform neuroendocrine stress response explains the heterogeneity of psychiatric disorders encountered in the wake of psychological stress.

Most severe psychological stressors do not result in the development of posttraumatic stress disorder or other forms of psychiatric illness. Certain characteristics of the stressor seem to increase the propensity to develop posttraumatic stress disorder following exposure; these include the magnitude of the stressor, the occurrence of physical injury, and the occurrence of peritraumatic dissociation. However, individual differences also contribute to the variance in posttraumatic stress disorder expression following trauma exposure; these include female gender, genetic vulnerability, and a history of exposure to early childhood trauma.

The latter finding has been both particularly intriguing and exceptionally controversial. The controversy has been spurred in part by the methodological difficulties inherent in assessing the nature and severity of childhood trauma. Concerns range from the possibility that some individuals might recall being sexually victimized in childhood when this has not actually occurred (i.e., the so-called false memory syndrome) to the possibility that some traumatized individuals might have suppressed the memories of childhood victimization and therefore may be unable to report that they were indeed traumatized. With all these potential sources of “noise” coming into play, it is all the more impressive that a consistent “signal” emerges from a variety of studies, all suggesting that early experiences of victimization predispose an individual to developing posttraumatic stress disorder when later exposed to psychologic trauma.

Much more speculative but on the cutting edge of traumatic stress research are recent studies that suggest that psychological trauma may result in quantifiable abnormalities in neurocognition (e.g., attention) and even in neuroanatomy (e.g., hippocampal volume). If replicated, these studies promise to dramatically alter the understanding of the effects of stress on psychiatric disorders and ultimately on normal and pathological brain development.

STRESS AND NONPSYCHIATRIC ILLNESS

Stressors are relevant to all illnesses, not just primary psychiatric conditions. There is an enormous literature linking stressors to the entire panoply of illness, a literature generally referred to as psychosomatic medicine. The term “psychosomatic” has been dropped from current nomenclature because of a misperception that it refers primarily to psychoanalytic theories of behavior and illness or to a simplistic assumption that all illness has its origin in struggles with adversity and with emotional conflict. Stress and the psychosomatic link pertain to some of the most prevalent and deadly diseases. The links are multiple—ranging from physiological perturbations elicited by stressors to various illness-promoting behaviors that are resorted to under conditions of emotional travail.

TREATMENT

Pharmacotherapy Many different psychiatric conditions are worsened by stressors. When major psychiatric conditions like depression or psychosis are present, the most appropriate medications are those for the Axis I diagnosis. However, more typically the picture is primarily one of anxiety or insomnia, perhaps not of as great a magnitude as that found typically in anxiety disorder, but still very distressing to the patient. In such settings pharmacological therapy with anxiolytics or short-term hypnotic agents can be extremely helpful. The field has lurched from one extreme to another in terms of prescription of such medication. Clearly, the unsupervised long-term prescription of this type of medication brings with it risks for drug dependence. On the other hand, depriving patients of such medication when they are in crisis is unconscionable. Short-term use of low-dose benzodiazepines is of considerable utility in helping patients struggle with unusual stressors. This intervention in association with supportive psychotherapy is extremely effective in crisis intervention.

Cognitive-Behavioral Therapy Cognitive-behavioral therapy methods are increasingly used to help individuals better manage their responses to stressful life events. These treatment methods are based on the notion that cognitive appraisals of stressful events and the coping efforts related to these appraisals play a major role in determining the response to stress.

A model of stress and coping has been developed that serves as the conceptual foundation for most cognitive-behavioral stress-management protocols. This model maintains that there are two types of cognitive appraisal that are especially important in mediating reactions to stress. The first, *primary appraisal,* refers to the way in which an individual evaluates the significance or meaning of a given event. When events (e.g., being diagnosed with a chronic illness such as rheumatoid arthritis) are appraised as harmful and threatening (e.g., “Because of my rheumatoid arthritis, I will never be able to do anything I really care about”), the individual is more likely to become anxious, depressed, and withdrawn. However, if the same event is viewed as challenging (e.g., “Although my arthritis prevents me from doing some physically demanding activities, there are many other important and meaningful activities that I can do and enjoy”), then more positive outcomes are more likely to occur. *Secondary appraisal* refers to the process of evaluating what can be done about the stressful event. Several factors can influence such secondary appraisals including the breadth of individuals' repertoire of coping skills, their mastery of specific coping skills, and their expectation that their skills will be effective.

Cognitive-behavior therapy approaches to stress management have three major aims. The first aim is to help individuals become more aware of their own cognitive appraisals of stressful events. The second aim is to educate individuals about how their appraisals of stressful events can influence negative emotional and behavioral responses and to help them reconceptualize their abilities to alter these appraisals. The third aim is to teach individuals how to develop and maintain the use of a variety of effective cognitive and behavioral stress management skills.

Efficacy Stress Management A number of controlled studies have documented the efficacy of cognitive-behavioral therapy stress management protocols. Two recent controlled studies examining this therapy in the treatment of individuals suffering chronic diseases are good examples of such research.

Individuals with rheumatoid arthritis not only must cope with frequent pain episodes, but also with major disruptions in their work and family life. One hundred and forty-one rheumatoid arthritis patients were randomly assigned to one of three groups: (1) a comprehensive cognitive-behavioral therapy stress management program, (2) an attention control group, or (3) a standard care control condition. Patients in the stress management and attention control conditions met weekly for 10 weeks and then monthly for an additional 15-month maintenance phase. At the end of the 10-week treatment phase of the study, patients in the cognitive-behavioral therapy group had significant improvements in pain, helplessness, self-efficacy, coping, and health status. When the outcomes for the three treatment conditions were compared at 15-month follow-up it was found that subjects in the cognitive-behavioral therapy group were much more likely to report maintenance of their treatment gains. These results suggest that cognitive-behavioral therapy may provide an effective means of reducing pain and psychological distress in patients with rheumatoid arthritis.

The efficacy of stress management has also been demonstrated in individuals suffering from inflammatory bowel disease. Eighty such subjects were randomly assigned to either a stress management intervention or a control group. Subjects in the stress management condition had six group sessions focusing on training with relaxation techniques, time management skills, and communication skills. There were significant improvements in measures of physical and psychological functioning in patients undergoing stress management whereas there were no changes in these measures in patients in the control condition. These benefits were not due to changes in subjects' medication status since no changes in medications were noted over the course of the study.

Stress-Management Training When cognitive-behavioral therapy is used for stress management, training is provided in a wide range of stress management skills. Although space limitations preclude a discussion of all these skills, it is important to highlight five skills that form the core of almost all such stress management

programs: self-observation, cognitive restructuring, relaxation training, time management, and problem solving.

Self-Observation One of the most effective ways of helping individuals become more aware of how they respond to problem situations is to have them keep a daily record of their behavior. A daily diary format (Table 25.9-2) is often used with patients being asked to keep a record of how they responded to challenging or stressful events that occurred each day. Entries are made in three columns: antecedents, behaviors, and consequences. In the "antecedents" column, patients record a specific environmental event that they perceived as stressful. The event might be an interpersonal stressor such as having an argument with a spouse or a confrontation with a coworker, or an intrapersonal stressor such as a flare in pain symptoms, or a major natural disaster. To assess appraisals of the stressor, patients are asked to rate how stressful they found this event on a 0 to 100 scale (e.g., 0 = not at all stressful; 100 = the most stressful event I have ever experienced). In the column marked behavior, patients are instructed to record their cognitive and behavioral reactions to the stressful event. The term "behavior" is broadly defined so that it not only encompasses overt behavioral responses, but also more covert cognitive, affective, and physiological responses. Thus, in the behavioral response column individuals might indicate specific coping behaviors they engaged in (e.g., avoidance), cognitive responses to the stressor (e.g., "I am overwhelmed and can't deal with this right now"), negative emotions (e.g., anxiety, depression), and physiological responses (e.g., increased sweating, heart rate, or nausea). In the column marked "consequences," patients record the outcomes resulting from their behavior in this situation.

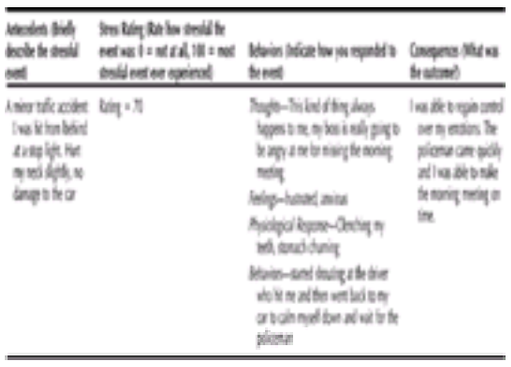

Table 25.9-2 Daily Diary Format for Recording Stress-Related Behaviors

Self-observation is effective for several reasons. First, it makes individuals aware of behaviors that they usually fail to notice. An individual whose style of dealing with social anxiety is to use avoidance, for example, may avoid potentially stressful situations in a more or less automatic fashion (e.g., routinely refuse invitations), use very subtle forms of avoidance (e.g., arriving late or leaving a party early), or arrange their lifestyle (e.g., work alone) so that they have very few chances to confront their fears. By keeping a record of their own behavior, individuals begin to recognize these behavioral patterns. A second important feature of self-observation is that it helps individuals to see how their cognitive-behavioral responses are linked to situational antecedents and consequences. Participants who keep daily diaries are often struck by the variations in their own reactions across different events. For example, while they may cope effectively with one stressful event (e.g., meeting an important work deadline), they find themselves unable to deal with another stressful event (e.g., an argument with their teenage son or daughter). Self-observation highlights the role that coping efforts, cognitive appraisals, and physiological responses play in explaining these very different responses to stressful situations. A third reason that self-observation is effective is that it makes individuals aware of how they might change their behavior to more effectively manage stress. Interestingly, individuals who keep daily diary records of stress-related behaviors often make changes in their own behavior, even before other stress management methods are introduced.

Cognitive Restructuring A hallmark of cognitive-behavioral therapy is its insistence that cognition plays a central role in the stress and coping process. In cognitive-behavioral therapy cognitive appraisals about stressful events are considered to be the key factor in determining stress-related responding. Given this emphasis on cognition, it is not surprising that a major thrust of cognitive-behavioral therapy approaches to stress management is on helping participants to become aware of and to change their maladaptive thoughts, beliefs, and expectations.

Cognitive therapists have developed protocols for helping distressed individuals restructure dysfunctional or maladaptive thinking patterns. Such cognitive restructuring begins by educating patients about the cognitive model and that the underlying cognition is a major determinant of emotional responding. According to the cognitive model, emotional reactions that are excessive or prolonged are often the result of cognitions that are distorted or dysfunctional.

The second step in cognitive restructuring is monitoring and analyzing dysfunctional thoughts. Table 25.9-3 lists a number of stress-related thoughts that might be identified through daily diary records. As can be seen, each of these thoughts can be linked to a negative emotion. A careful analysis of each thought also reveals an inherent error in logic. Certain types of errors or cognitive distortions occur quite frequently and individuals can thus be taught to recognize them.

Stressor	Cognition	Emotion
Painful medical procedure	I know I won't be able to tolerate the pain	Fear
Argument with a friend	She will never speak to me again	Discouragement
Co-worker injured on the job	It is my fault. I should have warned him about the danger	Guilt
Late for a deadline	It's hopeless, I'm always messing things up	Depression
Spouse criticizes you	He just doesn't get it. He will never understand what I am trying to do.	Anger

Table 25.9-3 Stress-Related Negative Thoughts

The third step in cognitive restructuring is to challenge and change cognitive distortions. Participants are often provided with a list of key questions (Table 25.9-4) that can be applied to their own negative thoughts. They are asked to select one or two questions that are useful in identifying the underlying logical problems in their thinking. Finally, participants develop a rational response that represents a more accurate and helpful cognitive response to the situation.

Is there an alternative explanation?
What is the evidence that this thought is true?
What will be the effect of continuing to think this way?
What is the best outcome, worst outcome, and most realistic outcome?
What is the likelihood that this will happen?

Adapted from Beck JS: *Cognitive Therapy: Basics and Beyond*. Guilford Press, New York, 1995.

Table 25.9-4 Key Questions for Challenging Negative Automatic Thoughts

Table 25.9-5 illustrates how one might change various overly negative stress-related thoughts to more adaptive rational responses. If the rational response is to be effective, it must be used frequently, ideally each time the negative thought occurs. To ensure that this occurs, patients are given home practice assignments in which they are asked to expose themselves to stressful events, monitor their negative thoughts, and respond with a more rational thought pattern. They are also asked to keep records of the resulting consequences, both in terms of changes in emotions and behaviors.

Negative Thought	Rational Response
I have no control over this situation.	It's true that I can't change what happened, but I can control how I respond.
I can't return to my old job, I am worthless.	Even though I may not be able to do the same kind of work, there a number of other jobs that I may be able to do.
It's not fair that I am being treated this way.	Life is not always fair, sometimes things go my way, other times they don't.
I am a burden on my family.	My family has had to give me more support lately, but I do a lot of things for them as well.
No one really cares about me.	There are people in my life (e.g., spouse or child) that I can turn to for support and love.

Table 25.9-5 Negative Thoughts and Rational Responses

Relaxation Training Relaxation skills can be very helpful in managing stress. When individuals learn to relax, their overall muscle tension is reduced, as is their overall level of autonomic arousal. Individuals who are able to relax are also more likely to be able to think more rationally and to be able to restructure negative cognitions when faced with stressful events. Finally, relaxation skills may be helpful in reducing maladpative behavior patterns.

In the 1920s Edmund Jacobson developed progressive relaxation training as a way to help individuals to control excessive muscle tension. Jacobson approached relaxation as a motor skill and emphasized that, like any skill, repeated practice was necessary for skill mastery. In traditional Jacobsonian relaxation training the goal is to heighten an individual's awareness of very low levels of muscular tension. The patient is instructed to focus on sensations that occur when tensing a single large muscle group (e.g., the forearm flexors—tensed by bending the left hand back at the wrist). The individual is then instructed to release the tension in that muscle group and study the resulting physical sensations. With repeated practice, trainees often become quite adept at recognizing “control signals” indicative of tension and letting the tension go. Traditional Jacobson relaxation training is effective; however, it does require a substantial time commitment on the part of the participant.

In the 1950s Jacobsonian relaxation training was modified by behavior therapists so that it could be more easily incorporated into treatment programs. The modified training program was similar to Jacobson's in emphasizing that relaxation could be achieved by using a series of exercises involving tensing and relaxing muscle groups, but differed in that it was much briefer. Numerous modified versions of Jacobson's original method have been developed and refined over the years.

One of the most important tasks in relaxation training is helping patients learn how to generalize their skills in relaxation from home practice sessions to stressful and demanding daily life events. Several techniques are used to enhance generalization. The first is a method developed by Jacobson called differential relaxation. In differential relaxation, the trainee is instructed to engage in a daily task and only use the muscles that are necessary for carrying out the task. For example, while writing one's name there needs to be activity in muscles of the hands and arms, but the muscles in the face, lower trunk, legs, and feet can be deeply relaxed. Differential relaxation is initially practiced in relatively easy situations (e.g., sitting or standing) where the demands are minimal. Over time, however, patients learn how to use differential relaxation in situations that are challenging and stressful. A second generalization method is termed the mini-practice. This consists of a brief relaxation procedure in which the trainee takes a deep breath, and then while slowly exhaling focuses on sensations of relaxation flowing downward from the muscles of the face to those of the neck, shoulders, trunk, and legs. The mini-practice takes approximately 30 seconds and thus can be done repeatedly over the course of a day. Individuals are encouraged to develop innovative reminders to cue themselves to perform frequent mini-practices every day. As individuals increase this frequency, they often report that relaxation becomes more or less an automatic and habitual response.

Time Management A stressful event, (e.g., caring for a sick child), can place inordinate demands on time. When exposure to such an event is prolonged or is combined with exposure to other stressors, the demands on time and energy are multiplied. In such circumstances individuals often report feeling that they have lost control of their daily schedule and, that as a result, they have little or no time in which to attend to their own needs.

Time management methods are designed to help individuals restore a sense of balance to their lives. The first step in training in time management skills is designed to enhance awareness of current patterns of time use. To accomplish this goal, individuals might be asked to keep a record of how they spend their time each day, noting the amount of time spent in important categories such as work, family, exercise, or leisure activities. Alternatively, they may be asked to list the important areas in their lives and then asked to provide two time estimates: (1) the amount of time they currently spend engaging in these activities, and (2) the amount of time they would like to spend engaging in these activities. Frequently, there is a substantial difference in the time individuals would like to spend on important activities and the amount of time they actually spend on them. With awareness of this difference comes increased motivation to make changes.

The second step in time management is designed to help individuals set their priorities. In setting priorities an important distinction can be drawn between urgency and importance. To explore this distinction it is helpful to envision a table that has the dimension of importance along one axis and the dimension of urgency along the other axis. Patients are then encouraged to enter into each quadrant of the table examples of activities from their own lives and to reflect on the patterns they observe. Individuals may note that much of their time is spent doing activities that are in the Urgent-Not Important quadrant or the Not Urgent-Not Important quadrant. Alternatively, they may report that they spend little or no time engaging in activities in the Not Urgent-But Important quadrant. Patients are encouraged to discuss the benefits that can come from increasing the priority they give to such important but less urgent activities as exercising, resting, relaxing, or doing volunteer work. They are also encouraged to try to match the time spent engaging in such meaningful activities with the priority they place on the activity.

Once participants have developed a better sense of their priorities, they are ready to move on to the third step in time management: goal setting. Goal setting focuses on specific behavior change targets that fit with an individual's long-term priorities. First, the individual is asked to identify a goal to be accomplished in the next week. The goal, for example, might be to lose 2 pounds or to sign up for an exercise program. Second, patients are asked to critically analyze the goal to ensure that it is reasonable, specific, and personally meaningful. Many of us set unrealistic goals (e.g., losing 10 pounds in a week) when trying to change our own behavior and as a result fail, feel frustrated, and give up. By helping a participant to set goals that are more reasonable, one can enhance the likelihood of success and increase compliance with stress management efforts. In the third step of goal setting, individuals are asked to reconfirm a time frame (e.g., a week) in which they will meet the goal. This step is important in reducing procrastination, a common obstacle to effective time management. Finally, the individual's efforts in reaching the goal are reviewed. Successful completion of a goal that has been long neglected can be very rewarding. Even if the goal was not achieved, important information about goal setting can be gleaned by reviewing performance. Perhaps the goal was set too high or the individual lacks the skills necessary to achieve the goal.

Problem Solving Problem solving is a skill that is introduced in the later stages of stress management training. As trainees attempt to apply what they have learned about stress management, they may find that some problem situations are particularly challenging and difficult to manage using only one or two stress management techniques.

Problem solving involves several basic steps. The first is problem identification during which the key problematic aspects of a stressful event such as problematic behaviors, thoughts, feelings, and physiological responses are identified. For example, dealing with the aftermath of a hurricane may be perceived as stressful because it is associated with particular (1) behavioral problems—having to clean up while simultaneously caring for one's children, (2) overly negative thoughts—“We will never recover from this event,” (3) negative feelings—fear, anger, depression, and (4) physiological responses—muscle tension, fatigue, and sleep problems. Helping an individual understand the various facets of their problematic responses is often helpful because it suggests what skills might be particularly useful. For example, if the problem is mainly a somatic one (increased tension), then relaxation-based strategies may be the most effective.

The second step in problem solving is generating alternatives. Brainstorming is used to generate a wide range of alternative problem solutions. The rules of brainstorming include: (1) the individual should offer as many solutions as they can, even if the solutions do not immediately seem reasonable; (2) no criticism is allowed; and (3) there should be an attempt to mix and match solutions in creative ways. One reason that brainstorming is effective in problem solving is that because of its rules it reduces the tendency to prematurely judge and reject possible coping options. Using this approach a participant can often generate a list of 5 to 10 potential solutions.

The third step in problem solving involves evaluating the alternatives and selecting the best solution. To begin this process a participant can be asked to rate the likelihood that each solution will have a positive outcome. It is often helpful to discuss some of the factors that influence how the participant makes these ratings (e.g., cost, availability of resources, effect of the solution on others). The result is a full review of the pros and cons of each solution. Typically, one solution stands out as the best and most practical.

The final step of problem solving involves implementing the solution. Participants basically try to apply the best solution in the problem situation and then review their progress with the therapist.

There is growing recognition that developing problem-solving skills is crucial to the maintenance of the effects of cognitive-behavioral therapy. Participants who learn how to approach a problem situation and formulate and implement effective coping options benefit in several important ways. First, they are often able to anticipate problems and make plans to avoid them. Second, they cope with setbacks more effectively and are able to get their coping efforts back on track after a setback so as to prevent a major relapse. Finally, they are better prepared for termination because they have learned to engage in a form of self therapy in which they can pinpoint and solve their own problems.

A 55-year-old married man was seen in psychiatric consultation because of symptoms of profound anxiety and depression following the destruction of his home in an earthquake. While he and his family survived intact, his home was a total loss. He developed chest pains and a myocardial infarction while trying to argue with insurance adjusters regarding the loss. In the coronary care unit he was apprehensive, tremulous, and tearful.

In the coronary care unit he was treated with low dose benzodiazepines to alleviate some of his anxiety symptoms. He made an uneventful recovery. During the next 6 months there were frequent aftershocks, and he became fearful, complained of difficulty sleeping, and felt guilty for not having taken better earthquake precautions to preserve his former house. Symptoms progressed to frank depression, and his chest pain symptoms worsened. The depression responded to treatment with a selective serotonin reuptake inhibitor; however, he continued to be preoccupied about his losses and the unpredictability of the future.

In cognitive behavioral therapy he began to chart the occurrence of his worries about future earthquakes and noted that the worries were more evident when he was paying bills for house renovation. He began to attend to some of the negative thoughts associated with the bills (“I'll never receive enough insurance reimbursement; my credit rating will be destroyed unless I pay the bill immediately”). In addition, he began a program of cardiac rehabilitation with exercise training that reassured him that he was recovering well. He also began regular meditation twice daily which he found helpful for his anxiety symptoms. After 10 weeks of treatment, he reported considerable improvement in his symptoms.

This case history demonstrates the confluence of adverse life events, physical illness, and emotional distress and the pharmacologic and psychotherapeutic treatment of stress-related psychopathology.

SUGGESTED CROSS-REFERENCES

Section 1.17 discusses population genetics in psychiatry. Acute and transient psychotic disorders and culture are discussed in Section 13.3, and depressive disorders are included in Section 14.6. Behavior and immunity is discussed in Section 25.10. Death, dying, and bereavement are discussed in Section 25.6.

SECTION REFERENCES

Abelson JL, Curtis GC: Hypothalamic-pituitary-adrenal axis activity in panic disorder: 24-hour secretion of corticotropin and cortisol. Arch Gen Psychiatry *53:*323, 1996.

*Beck JS: *Cognitive Therapy: Basics and Beyond*. Guilford Press, New York, 1995.

Bernstein DA, Borkovec TD: *Progressive Relaxation Training: A Manual for the Helping Professions*. Research Press, Champaign, IL, 1973.

Boscarino JA: Diseases among men 20 years after exposure to severe stress: Implications for clinical research and medical care. Psychosom Med *59:*605, 1997.

Breier A, Davis OR, Buchanan RW, Moricle LA, Munson RC: Effects of metabolic perturbation on plasma homovanillic acid in schizophrenia: Relationship to prefrontal cortex volume. Arch Gen Psychiatry *50:*541, 1993.

Cohen S, Tyrrell DA, Smith AP: Psychological stress and susceptibility to the common cold. N Engl J Med *325:*606, 1991.

Covey S, Merrill S, Merrill R: *First Things First*. Fireside Books, New York, 1995.

Dimsdale J: Emotional causes of sudden death. Am J Psychiatry *134:*1361, 1977.

Engel B: Stress is a noun! No, a verb! No, an adjective! In *Stress and Coping,* TM Field, PM McCabe, N Schneiderman, editors. Erlbaum, Hillsdale, NJ, 1985.

*Friedman MJ, Charney DS, Deutch AY, editors. *Neurobiological and Clinical Consequences of Stress: From Normal Adaptation to Post-Traumatic Stress Disorder*. Lippincott-Raven, Hagerstown, MD, 1995.

Gil KM, Ross SL, Keefe FJ: Behavioral treatment of chronic pain: Four pain management protocols. In *Chronic Pain*, RD France, KRR Krishnan, editors. American Psychiatric Press, Washington, DC, 1988.

Goldfried MR, Davidson GC: *Clinical Behavior Therapy*. Holt, Rinehart, & Winston, New York, 1976.

Jacobson E: *Progressive Relaxation*. University of Chicago Press, Chicago, 1938.

*Kessler Ronald C: The effects of stressful life events on depression. Annu Rev Psychol *48:*191, 1997.

Kiecolt-Glaser JK, Page GG, Maruch PT, MacCallum RC, Gaser R: Psychological influences on surgical recovery. Am Psychol *53:*1209, 1998.

*Lazarus RS, Folkman S: *Stress, Appraisal, and Coping*. Springer, New York, 1984.

Lesserman J, Li Z, Hu YJB, Drossman DA: How multiple types of stressors impact on health. Psychosom Med *60:*175, 1998.

Marlatt GA, Gordon JR: *Relapse Prevention*. Guilford, New York, 1985.

Mazure CM, editor: *Does Stress Cause Psychiatric Illness?* American Psychiatric Press, Washington, DC, 1995.

Milne B, Joachim G, Niedhardt J: A stress management programme for inflammatory bowel disease patients. J Adv Nurs *11:*561, 1986.

Parker JC, Smarr KL, Buckelew SP, Stucky-Ropp RC, Hewett JE, Johnson JC, Wright GE, Irvin WS, Walker SE: Effects of stress management on clinical outcomes in rheumatoid arthritis. Arthritis Rheum *38:*1807, 1995.

Post R: Transduction of psychosocial stress into the neurobiology of recurrent affective disorder. Am J Psychiatry *149:*999, 1992.

*Rahe R, Floistad I, Bergan T, Ringdal R, Gerhardt R, Gunderson E, Arthur R: A model for life changes and illness research: Cross-cultural data from the Norwegian navy. Arch Gen Psychiatry

*31:*172, 1974.

*Rice PL: *Stress and Health.* Brooks/Cole, Pacific Grove, CA, 1999.

Romas JA, Sharma M: *Practical Stress Management: A Comprehensive Workbook for Managing Change and Promoting Health*. Allyn and Bacon, Boston, 1995.

Rubin RT: Neuroendocrine aspects of stress in major depression. In *Stress in Psychiatric Disorders,* J Yager, RP Liberman, editors. Springer, New York, 1994.

Sapolsky RM: Stress, the aging brain, and the mechanisms of neuron death. The MIT Press, Cambridge, MA, 1992.

Schnall PL, Schwartz JE, Landsbergis PA, Warren K, Pickering TG: A longitudinal study of job strain and ambulatory blood pressure: results from a three-year follow-up. Psychosom Med *60:*697, 1998.

Stein MB, Walker JR, Anderson G, Hazen AL, Ross CA, Eldridge G, Forde DR: Childhood physical and sexual abuse in patients with anxiety disorders and in a community sample. Am J Psychiatry *153:*275, 1996.

Wolpe J: *Psychotherapy by Reciprocal Inhibition*. Stanford University Press, Stanford, 1958.

25.10 BEHAVIOR AND IMMUNITY

JOHN M. PETITTO, M.D., AND DWIGHT L. EVANS, M.D.

Researchers in psychiatry were instrumental in founding the discipline of psychoneuroimmunology, research that is attempting to dissect the complex interactions between the brain and immune systems as well as identify common signaling pathways shared by these systems. In the 1960s two psychiatrists performed landmark studies demonstrating that the brain could modulate parameters of the immune response. George Solomon and his colleagues demonstrated that early experience could produce long-term changes in immune function in animals. In a series of studies, Marvin Stein and coworkers showed that hypothalamic lesions could modify immune processes and anaphylactic responses across species of laboratory animals. They speculated that the hypothalamus transmits signals to the immune system via its control of neuroendocrine and autonomic activity that could modulate immune physiology. Subsequent research, in particular the pioneering collaborative studies of psychologist Robert Ader and immunologist Nicholas Cohen, provided the most compelling and powerful illustrations of the interrelationships between the central nervous system (CNS) and the immune system and heightened interest in psychoneuroimmunological research. Ader and Cohen behaviorally conditioned mice by pairing saccharin (a conditioned stimulus that has no intrinsic immunomodulatory actions) with an immunosuppressant, cyclophosphamide (an unconditioned stimulus). They showed that when behaviorally conditioned mice were exposed to saccharin alone at a later time, saccharin itself resulted in inhibition of antibody production following immunization with sheep red blood cells. Because immunologists had long held that the immune system was a self-regulated autonomous system, Ader and Cohen initially faced considerable skepticism. To demonstrate further that the physiological changes in immune activity elicited by behavioral conditioning were clinically meaningful and could alter the health status of the organism, Ader and Cohen subsequently showed that exposure to saccharin alone could modify the course of murine systemic lupus erythematosus.

This area of inquiry has considerable potential to provide new insights into understanding processes that may affect the development and pathogenesis of neuropsychiatric disorders and could lead to novel strategies to treat and prevent some forms of mental illness.

NEUROBIOLOGY OF BRAIN-IMMUNE INTERACTIONS

Regulation of systemic homeostatic processes such as modulation of blood pressure and gastrointestinal physiology by the CNS has been recognized for a number of years. Thus, it is not surprising that the brain might also play a role in immunoregulation as well as receive sensory input from the immune system, which in turn can help inform the brain to engage in adaptive behaviors that can affect the well-being of the organism. As the field has progressed, the need for multidisciplinary research strategies to disentangle the complex interactive mechanisms between systems (e.g., CNS, endocrine system, and immune system) and within systems has become increasingly apparent. Figure 25.10-1 is a schematic summarizing the various levels of neuroimmune interactions identified by basic research studies to date.

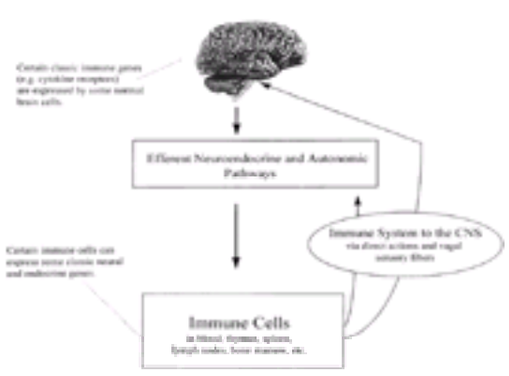

FIGURE 25.10-1 Schematic diagram of central nervous and immune system interactions. Neural and endocrine pathways can modulate parameters of immune system physiology, and the peripheral immune system can transmit signals to the brain. The central nervous and immune systems are now known to share particular endogenous signaling molecules previously thought to be unique to either the CNS or peripheral immune cells.

More than three decades ago psychiatric researchers began to explore the hypothesis that stressful emotional states and their accompanying changes in the brain (e.g., hypothalamus) could modify immune homeostasis by neuroendocrine and sympathetic outflow pathways from the CNS. It is now known that various hypothalamic-pituitary signals possess modulatory properties on some immune cell subsets. Much attention has been given to the actions of the hypothalamic-pituitary-adrenal axis on components of immune system functioning, in part because alterations in hypothalamic-pituitary-adrenal axis physiology have been well documented in the field of biological psychiatry (e.g., major depressive disorder). Although preclinical studies have documented varied immunomodulatory effects of different neuropeptides (e.g., adrenocorticotropic hormone [ACTH], corticotropin-releasing factor [CRF], endorphins, thyrotropin-releasing hormone [TSH]), most emphasis has been given to the role of adrenal steroids. The known immunosuppressive effects of high doses of synthetic steroids led to the commonly held view that behavioral state–induced (e.g., stress, bereavement, depression) elevations in adrenal corticosteroids would suppress immune activity. It is now appreciated, however, that under physiological conditions, endogenous adrenal steroids have bidirectional and complex immunomodulatory actions on immune cell circulation patterns between blood and immune organs (e.g., spleen) and on immune cell function. Attesting to the complexity of brain-immune interactions in vivo, it is apparent that even the neuroimmunomodulatory actions of adrenal cortisol in association with stressful events and psychiatric disorders such as major depressive disorder are not yet clear. Immune organs such as the thymus and spleen are innervated by sympathetic efferent fibers. David Felten and colleagues and others have found that noradrenergic nerve terminals are in close proximity to lymphocytes forming functional neuroeffector contacts.

A working hypothesis of research in the field has been that neurobiological changes in the brain associated with stress and mental illness may result in alterations in the activity of the hypothalamic-pituitary-adrenal axis, and descending sympathetic autonomic pathways that innervate immune organs, or both, which in turn can modify immune status and susceptibility to disease in vulnerable individuals. Preclinical research indicates that the neural modulation of immune activity involves limbic forebrain and cortical regions, brain areas involved in the expression of emotional, cognitive, and thought processes. Relevant to understanding how stress and depression might alter immune physiology in humans, Michael Irwin and colleagues, for example, have shown that intracerebroventricular CRF administration produces increased emotional reactivity and central activation of descending sympathetic pathways, which leads to reductions in the function of natural killer cells residing in the spleen.

To begin to understand how complex phenomena such as behavioral conditioning can affect immunity, several research laboratories have used multidisciplinary approaches to begin to dissect the neural, endocrine, and immune mechanisms involved. Donald Lysle and colleagues, for example, have extended the work of Ader and Cohen by examining how behavioral conditioning effects are translated to changes in immunity. They have shown that conditioned environmental stimuli (inherently nonaversive stimuli) previously paired with an aversive event (footshock stress) can themselves induce pronounced changes in immune status and modify disease course in animals. In this model, they found that opioid systems are involved centrally, that sympathetic pathways transduce the effects to cells of the immune system peripherally (e.g., subsets of splenic lymphocytes), and that release of nitric oxide by macrophages is a key immunoregulatory step in modulating other components of the immune response (e.g., T cell responsiveness).

Analogous to the discovery years ago that digestive system peptides such as cholecystokinin (CCK) and vasoactive intestinal polypeptide (VIP) were present in the CNS, cytokines and their receptors (e.g., interleukin-1 [IL-1] receptors) are now being identified in brain, and their role as neuromodulators, growth factors, and

mediators of immunelike responses in the CNS is being explored. Immune cell–derived signals such as cytokines communicate with endocrine and brain cells, and it is now known that cells of the immune, endocrine and central nervous systems share common signals and receptors. The function of immunomodulators (e.g., IL-1) on brain cells and neuromodulators such as norepinepherine on immune cells is the focus of ongoing research.

Recently, several areas of inquiry have begun to shed light on the potential clinical significance of immune cell–derived signals on the brain and behavior. Cytokines derived by immune cells, in particular proinflammatory cytokines (e.g., IL-1, IL-6, tumor necrosis factor-alpha), can cross the blood-brain barrier. Because cytokines are large molecules, however, most do not readily cross the blood-brain barrier. New evidence from the laboratories of Steve Maier, and colleagues shows that the afferent sensory fibers of the vagus nerve (once viewed as an exclusively efferent cranial nerve) can carry signals initiated by peripheral cytokines to brainstem areas including the nucleus tractus solitarii. Increased concentrations of certain peripheral cytokines can elicit a sickness syndrome in animals that includes mild leukocytosis, fever, increased levels of slow-wave sleep, increased production of acute phase proteins by the liver, decreased activity and exploration, reduced social interaction, increased sensitivity to pain and depressive-like symptoms. It has been theorized that this signal by the immune system provides a message to the brain to aid in initiation of an adaptive, recuperative process by modifying the organism's physiology (e.g., fever can act to reduce severity of infection by a subset of microorganisms) and behavior (e.g., decreased exploration of the environment to preserve energy). Maier and Linda Watkins have also posed the hypothesis that both psychological and physical stressors activate much of the same neurobiological circuitry that is activated by infectious agents or immune cell–derived products that elicit the sickness syndrome. This intriguing avenue of investigation may provide new insights into understanding some of the enigmatic immune and acute phase–like changes described in some individuals experiencing major depressive disorder or other psychiatric syndromes.

PSYCHOSOCIAL FACTORS, DEPRESSION, AND IMMUNE STATUS

The effects of various forms of stress on measures of immune status in humans are well documented. Several notable studies relating psychosocial and emotional factors to immune-based diseases are discussed in this section. Although the role of emotional factors in the onset and course of immune-based diseases has not been firmly established, considerable evidence indicates that psychosocial factors can influence the course of general medical illnesses. Depression has been associated, for example, with increased mortality in elderly patients in nursing homes and with lower 6- and 18-month survival rates following myocardial infarction. Amanda Ramirez and colleagues showed that severe life stress increased the likelihood of relapse in women with breast cancer by nearly sixfold, and research studies by our group on the effects of life stress and depression in men infected with the human immunodeficiency virus (HIV) indicate that high levels of severe life stresses (but not stresses common to daily living) were associated with greater risk of early HIV disease progression and reductions in natural killer (NK) cells and $CD8^+$ T cells over time. Although the studies cited support the hypothesis that stressful life events may affect immune status and disease course, generally research studies in the field from different groups have yielded mixed results. Also, in the particular studies cited above, stress and depression were found to influence individuals who were vulnerable; individuals with existing impairments or reduced immunological plasticity, rendering them more susceptible to the influence of adverse events on health outcome.

In one of the most promising studies to date David Spiegel and coworkers found that a psychosocial group intervention was associated with improved survival in women with breast cancer. They examined the effects of weekly supportive group therapy on survival time in a randomized, long-term follow-up study of patients with metastatic breast cancer. They predicted that the psychotherapeutic intervention would improve quality of life; however, it also influenced the course of the disease. Breast cancer patients who took part in weekly supportive group therapy lived twice as long as matched breast cancer patients who received the same cancer treatment interventions but did not participate in group therapy. Fawzy and colleagues also found that a group intervention increased NK cell function and survival in men with malignant melanoma. Studies are currently under way to replicate and extend these findings.

Perhaps the strongest evidence relating psychosocial events to susceptibility to clinically meaningful endpoints comes from experimental laboratory studies in humans. Sheldon Cohen and colleagues inoculated subject volunteers with cold virus (e.g., rhinovirus) and found that psychosocial stress preceding inoculation was associated with an increased rate of acute respiratory infection. A. A. Stone and coworkers showed that psychosocial factors were correlated with increased cold symptoms in subjects with confirmed rhinovirus infection following experimental inoculation. Moreover, because as many as one-third of subjects with known rhinovirus infection do not exhibit cold symptoms, they also determined whether recent life events, mood, and perceived stress were related to development of cold symptoms following confirmed infection with rhinovirus virus. They found that significantly more major life events occurred during the year preceding rhinovirus infection in the subjects who developed cold symptoms. Current mood and perceived stress were not associated with an increased likelihood of developing cold symptoms.

After many years of research in the field, Stein and his colleagues concluded that the available literature on immune correlates of depression were “findings in search of meaning.” Although studies spanning more than 20 years have examined the relation between clinical depression and immunity, results are inconclusive and data conflicting. The same immune parameters have been reported as increased, decreased, or unchanged by different clinical studies. These discrepancies in the literature may stem in part from differences across studies, including duration and severity of symptoms, temporal parameters (e.g., when samples are drawn for assays), and subject variables such as sex, age, and trait variable (e.g., heritable personality characteristics) among others. Understanding how these variables may interact to produce changes in immune responsiveness may be particularly important for designing clinical psychiatric investigations. In a meta-analytic assessment of this literature several years ago Tracy Hebert and Shelda Cohen found that the most consistent immune changes associated with major depression were reductions in NK activity, mitogen-induced T cell activity, and numbers of certain lymphocyte subsets. Reductions in these in vitro laboratory immune measures should be viewed cautiously. Substantive data indicating that alterations in these in vitro laboratory immune measures correlate with reduced immunocompetence in vivo are lacking. More-recent data may provide another viable interpretation. The work of Michael Maes and colleagues and others now suggests that at least in subgroups of depressed patients, reduced NK activity and altered T cell proliferative responses may present together with estimates of immune activation and inflammatory-like responses (and possibly autoimmune-like laboratory correlates). Many of these findings, such as elevations in acute phase reactants (e.g., haptoglobin concentrations) and increased concentrations of inflammatory cytokines, together with some of the behavioral characteristics (e.g., reduced activity, reduced interest in typically pleasurable activities, reduced social interaction) seem to fit within the theoretical framework of the sickness syndrome described above. Reductions in select cellular immune parameters such as NK activity with concomitant measures of immune activation (e.g., nonspecific antinuclear and antiphospholipid antibodies, acute phase reactants, and increased production of certain proinflammatory cytokines) are often exhibited by individuals diagnosed with certain systemic autoimmune diseases.

NEUROIMMUNE INTERACTIONS ASSOCIATED WITH MAJOR NEUROPSYCHIATRIC DISORDERS

Recent advances have led to new insights into potentially effective treatments that could slow the course of neurodegeneration in patients with Alzheimer's disease. Knowledge of basic processes in the CNS such as the fundamental neuroimmunology of microglial cells, astrocytes, and their secretory products is a vital area of research that has recently led to discovery of immune-based drug therapies (e.g., new cyclooxygenase inhibitors) with the potential to treat Alzheimer's disease and other neurodegenerative diseases. This sets an important precedent for future research seeking to determine the clinical potential for understanding the role of neuroimmune interactions in mental illness.

The field is advancing to the stage at which testable hypotheses examining the role of neuroimmune processes in the pathogenesis of these major neuropsychiatric disorders can be examined. As noted above, certain cytokines and their receptors are now known to be constituents of the normal brain. Understanding their neurobiological function could provide important new clues for unraveling the etiology and pathogenesis of major neuropsychiatric disorders such as schizophrenia and dementia of the Alzheimer's type. For example, Petitto and colleagues and others are determining the molecular neurobiology of the prototypical cytokine, IL-2. Clinical investigations have detected increased levels of IL-2 in the concentrations cerebrospinal fluid (CSF) of schizophrenic patients manifesting symptoms of psychosis, and increased concentrations of IL-2 have been measured in the hippocampi of Alzheimer's disease patients postmortem. The authors have cloned and sequenced the subunits of the IL-2 receptor from the normal brain and have found that IL-2 can modulate the release of endogenous dopamine from striatal slices in vitro and modify behaviors in animals in vivo that are known to be mediated by forebrain dopamine pathways. This prototypical cytokine has also been shown to have trophic effects on both fetal glial cells and neurons and to possess a variety of other actions involved in brain development (e.g., effects on myelin basic protein expression).

Epidemiological findings prompted reconsideration of whether prenatal viral infection or the maternal immune response to such a putative agent may serve as an environmental trigger for the expression of schizophrenia in individuals with genetic loading for the disorder. The hypothesis that early neurodevelopmental alterations may account for key pathophysiological abnormalities seen in the mature brain of individuals with schizophrenia is being examined extensively by neuroscientists. For example, during early development an event such as viral infection or birth trauma in a genetically susceptible individual could possibly disrupt the normal timing for IL-2 and other neurotrophic cytokine stimulation of neuronal growth and migration. If this occurred in a region such as the hippocampus in which IL-2 receptors are enriched, it could contribute to the alterations in this region (e.g., abnormal orientation of subsets of hippocampal neurons) that are seen in postmortem brains of individuals with schizophrenia.

Although there has been ongoing interest in whether autoimmune-like processes may be operative in neuropsychiatric disorders such as schizophrenia, clinical research support for this theory comes largely from nonspecific, correlative data. Often the group mean values for various laboratory measures of immune status (e.g., cytokines, elevated $CD5^+$ lymphocytes) differ significantly from those of normal controls; however, for the most part the actual values themselves fall within the range of normal limits. Some of the issues that researchers are struggling with in this field are similar to those being dealt with by researchers studying systemic autoimmune disorders. People with a chronic schizophrenia (including never-medicated subjects) have a significantly higher frequency of certain autoantibodies than control subjects. This also occurs in people with certain systemic autoimmune disorders; like in groups of individuals with schizophrenia, it is often unclear what role they play in the pathogenesis of autoimmunity against the tissue targeted in the disease (e.g., inflammation and destruction of the pancreas in autoimmune diabetes). Many such autoantibodies are frequently found to be nonspecific. Therefore the various autoimmune-like or proinflammatory-like findings suggesting autoimmune involvement in many neuropsychiatric disorders may be equally likely to reflect epiphenomena. Since the brain is a more immune-privileged organ than others, classic end-organ autoimmune signs (e.g., infiltration of inflammatory cells) are not common to the brain unless the integrity of the blood-brain barrier is compromised. The brain's internal immune cells such as microglia could play a significant role, and microglial responses are an important focus of current research in Alzheimer's disease, for example.

FUTURE DIRECTIONS

Research on brain-immune interactions could provide vital new clues to advance our understanding of psychiatric disorders of the brain such as schizophrenia, depressive disorders, and dementia of the Alzheimer's type. Fundamental understanding of the molecular basis of neuroimmune interactions operative in the pathogenesis of these psychiatric disorders could lead to development of novel therapeutic agents and intervention strategies. Learning more about how behavioral conditioning may exert potent effects on immune physiology, and identifying the combination of psychosocial factors that may affect the course of immune-based diseases such as HIV disease, cancer, and lupus could also lead to useful behavioral therapies to complement conventional medical treatment approaches. These are a few of the many promising directions that research into neuroimmune interactions may go.

SUGGESTED CROSS-REFERENCES

Psychoneuroendocrinology in discussed in Section 25.9. The immune system and central nervous system are discussed in Section 1.12. Psycho-oncology is covered in Section 25.11. See Section 25.9 for a discussion of stress and psychiatry.

SECTION REFERENCES

Ader R, Cohen N: Behaviorally conditioned immunosuppression and murine systemic lupus erythematosus. Science *215:*1534, 1982.

*Ader R, Felten DL, Cohen N, editors: *Psychoneuroimmunology*. Academic Press, New York, 1991.

Blalock JE: The syntax of immune-neuroendocrine communications. Immunol Today *15:*504, 1994.

Cohen S, Tyrrell DAJ, Smith AP: Psychological stress and susceptibility to the common cold. N Engl J Med *325:*606, 1991.

*Dunn AJ: Interactions between the nervous system and the immune system: Implications for psychopharmacology. In *Psychopharmacology: The Fourth Gerneration of Progress,* FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Evans DL, Folds JD, Petitto JM, Golden RN, Pedersen CA, Corrigan M, Gilmore JH, Silva SG, Quade D, Ozer H: Circulating natural killer cell phenotypes in males and females with major depression: Relation to cytotoxic activity and severity of depression. Arch Gen Psychiatry *495:*388, 1992.

Evans DL, Leserman J, Perkins DO, Stern RA, Murphy C, Zheng B, Longmate JA, Silva SG, van der Horst CM, Hall CD, Folds JD, Golden RN, Petitto JM: Severe life stress: Association with HIV disease progression. Am J Psychiatry *154:*630, 1997.

Fawzy FI, Fawzy NW, Hyun CS, Eashoff R, Guthrie D, Fahey JL, Morton DL: Malignant melanoma effects of an early structures psychiatric intervention, coping, and affective state on recurrence and survival 6 years later. Arch Gen Psychiatry *50:*681, 1993.

Felten DL, Felten SY: Immune interactions with specific neural structures. Brain Behav Immun *1:*287, 1987.

Frasure-Smith N, Lesperance F, Talajic M: Depression following myocardial infarction: Impact on 6-month survival. JAMA *270:*1819, 1993.

Ganguli R, Brar JS, Chengappa KN, Yang ZW, Nimgaonkar VL, Rabin BS: Autoimmunity in schizophrenia: A review of recent findings. Ann Med *25:*489, 1993.

*Goehler LE: Interleukin-1 beta in immune cells of the abdominal vagus nerve: A link between the immune and nervous systems? J Neurosci *19:*2799, 1999.

Herbert TB, Cohen S: Depression and immunity: A meta-analytic review. Psychol Bull *1133:*472, 1993.

*Irwin M: Psychoneuroimmunology of depression. In *Psychopharmacology: The Fourth Generation of Progress,* FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

*Kiecolt Glaser JK, Glaser R: Psychoneuroimmunology and health consequences: Data and shared mechanisms. Psychosom Med *57:*269, 1995.

Leserman J, Petitto JM, Perkins DO, Folds JD, Golden RN, Evans DL: Severe stress, depressive symptoms, and changes in lymphocyte subsets in human immunodeficiency virus-infected men. Arch Gen Psychiatry *54:*279, 1997.

Licinio J, Seibyl JP, Altemus M, Charney DS, Krystal JH: Elevated CSF levels of interleukin-2 in neuroleptic-free schizophrenic patients. Am J Psychiatry *150:*1408, 1993.

Lysle DT, Coussons-Read ME: Mechanisms of conditioned immunomodulation. Int J Immunopharmacol *17:*641, 1995.

*Lysle DT: Endogenous opioids regulate the expression of inducible nitric oxide synthase by splenocytes. J Pharmacol Exp Ther *288:*502, 1999.

Maes M: Evidence for an immune response in major depression: A review and hypothesis. Prog Neuropsychopharmacol Biol Psychiatry *19:*11, 1995.

*Maier SF, Watkins LR: Cytokines for psychologists: Implications of bidirectional immune-to-brain communication for understanding behavior, mood and cognition. Psychol Rev *105:*83, 1998.

McEwen BS: Protective and damaging effects of stress mediators. N Engl J Med *338:*171, 1998.

McGeer PL, McGeer EG: The inflammatory response system of brain: Implications for therapy of Alzheimer's and other neurodegenerative diseases. Brain Res Brain Res Rev *21:*195, 1995.

Miller AH, Spencer RL, McEwen BS, Stein M: Depression, adrenal steroids, and the immune system. Ann Med *25:*481, 1993.

Petitto JM, Huang Z, Raizada M, Rinker CM, McCarthy DB: Molecular cloning of the cDNA coding sequence of IL-2 receptor-g (g_c) from human and murine forebrain: Expression in the hippocampus in situ and by brain cells in vitro. Mol Brain Res *53:*152, 1998.

Petitto JM, McCarthy DB, Rinker CM, Zhi Huang H, Getty T: Modulation of behavioral and neurochemical measures of forebrain dopamine function in mice by species-specific interleukin-2. J Neuroimmunol *73:*183, 1997.

Ramirez AJ, Craig KJT, Watson JP, Fentiman IS, North WRS, Rubens RD: Stress and relapse of breast cancer. Br Med J *298:*291, 1989.

Reichlin S: Mechanisms of disease: Neuroendocrine-immune interactions. N Engl J Med *329:*1246, 1993.

Solomon GF, Levine S, Kraft JK: Early experience and immunity. Nature *220:*821, 1968.

Spiegel D, Bloom J, Kraemer HC, Gottheil E: The beneficial effect of psychosocial treatment on survival of metastatic breast cancer patients: A randomized prospective outcome study. Lancet *14:*888, 1989.

Stein M, Miller AH, Trestman RL: Depression, the immune system and health and illness. Arch Gen Psychiatry *48:*171, 1991.

Stone AA, Bovbjerg DH, Neale JM, Napoli A, Valdimarsdottir H, Cox D, Hayden FG, Gwaltney JM Jr: Development of common cold symptoms following experimental rhinovirus infection is related to prior stressful life events. Behav Med *18:*115, 1992.

25.11 PSYCHO-ONCOLOGY

MARGUERITE S. LEDERBERG, M.D., AND JIMMIE C. HOLLAND, M.D.

Psycho-oncology, or the study of the psychological aspects of cancer, continues its exponential growth, enriched by the new perspective of Eastern European researchers and other international contributions. After describing the phenomenology, prevalence, and treatment of psychiatric syndromes in cancer patients, another generation of researchers focused on subsyndromal stress responses through quality-of-life research. The needs of a large and growing population of long-term survivors continue to be elucidated. The third generation is studying the families of cancer patients, with results that are already of practical importance. The ethical issues of the postmodern era are sharper than ever, as exemplified by the physician-assisted suicide debate, which reached the United States Supreme Court. Deoxyribonucleic acid (DNA) technology continues to spawn ever larger populations of healthy individuals made aware of their genetic risk for cancer, creating psychological and ethical problems we must yet learn to manage. Finally, the role of psychological factors in cancer risk and survival continues to evoke a mix of provocative scientific findings and public overinterpretations.

ADAPTIVE RESPONSE

The diagnosis of cancer creates a crisis that requires patients to adapt quickly to catastrophic news. They must try to control the level of emotional distress while making crucial treatment decisions. Major concerns are fear of death, dependency, disfigurement, disability, and abandonment as well as disruptions in relationships, role functioning, and financial status. Patient responses are modulated by medical, psychological, and interpersonal factors. Medical factors include site of disease, symptoms, and predicted course; psychological factors include preexisting character style, coping ability, ego strength, developmental stage of life, and the impact and meaning of the cancer at that stage. Interpersonal factors refer to family and social support and include the input of the health care team. Affectively, patients may experience reactions ranging from extreme anxiety, sadness, fear, and anger to numbness and lack of reactivity. Guilt and attribution mechanisms play a major role. Cognitively, patients may become highly focused and seek information aggressively or become confused, paralyzed, and unable to concentrate. Somatic complaints may increase, and daily activity, appetite, and sleep are often disrupted. This acute stress response may be severe, but it is usually variable and transient, as patients respond to their rapidly changing reality. Generally, the stress is not so long-lasting as to interfere with the significant areas of functioning outlined in Table 25.11-1. When impairment lasts more than 2 weeks or symptoms are incapacitating, the patient should be evaluated for a psychiatric disorder.

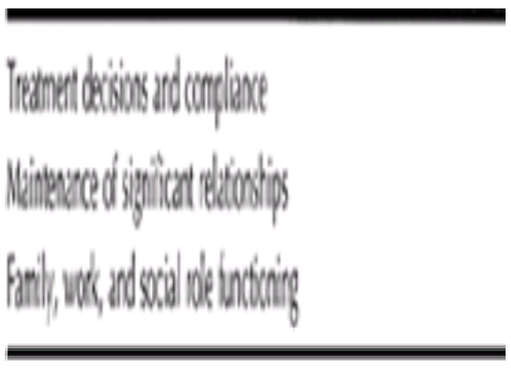

Table 25.11-1 Significant Areas of Functioning for the Cancer Patient

SUBSYNDROMAL STRESS RESPONSES AND QUALITY OF LIFE

In absolute terms, subsyndromal responses to chronic stress account for much human suffering that is amenable to treatment, particularly in cancer patients, whose significant psychological problems are primarily reactive. Two categories in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) recognize the importance of stress in medical conditions: (1) psychological factors affecting medical condition and (2) acute stress disorder.

Well-validated measures can assess quality-of-life at any point in the illness. They help predict response to future treatment and are factored into the outcome measures of therapeutic trials for noncurative treatments, addressing the crucial question: is the toxicity worth the gain?

PSYCHIATRIC SYNDROMES IN CANCER PATIENTS

Epidemiology

Symptoms Table 25.11-2 lists the most common symptoms found in a group of cancer patients with advanced disease. Five of the top seven are in the psychological realm and are present in more than 50 percent of patients. Three of the next ten are present in 12 to 45 percent of patients.

Symptom	% Positive
Fatigue	74
Worry	71
Sadness	66
Pain	63
Drowsiness	61
Dry mouth	56
Insomnia	54
Poor appetite	45
Nausea	44
Bloating	39
Difficulty in concentration	38
Constipation	33
Change in taste	36
Cough	30
Sexual dysfunction	24
Incontinence	12
Nightmares	11

Table 25.11-2 Symptoms in Patients With Cancer

Syndromes The incidence of psychiatric disorders in mixed inpatient and outpatient populations is about 50 percent, two-thirds of which represent adjustment disorders and psychological problems directly related to the cancer. Thus most of the observed pathology can be assumed to reflect the burden of the illness. Later studies have shown higher incidence of these disorders in inpatients, between 20 and 45 percent depression, depending upon the diagnostic criteria, and an incidence of delirium rising 15 percent to 75 percent with worsening disease. These figures are similar to those found in other comparably ill patients, contrary to persistent myths about how much worse cancer patients must feel. This distortion still affects members of the medical profession, leading to therapeutic nihilism, with negative consequences for patients. It raises the threshold of psychological distress that doctors consider “abnormal” enough to require treatment. Studies of psychiatric consultation data reveal that treatable syndromes such as major depression and delirium continue to be underdiagnosed and undertreated, even in hospice populations, despite known effectiveness in cancer patients.

Treatment

General Principles The hallmark of psycho-oncological treatment is therapeutic activism, with simultaneous use of several modalities in an aggressive attempt at rapid relief of symptoms. Our growing understanding of cancer and cancer treatments as traumatic events makes this not only humane, but also essential to improving subsequent adjustment. The physician must also address the family and the family-staff interface and must remain active throughout the terminal stages of the illness. Psychiatrists must be prepared for the existential plight of the cancer patient, which intensifies the expectable countertransference reactions. They must resist the urge to distance themselves from a painful situation, while guarding against overinvolvement. They must be aware of the patient's physical condition, yet not allow acute physical or emotional distress, legitimate despair, and bars to communication such as tracheostomies, to prevent a thorough evaluation. This requires overcoming fear and guilt and engaging the patient as deeply as one's interviewing skills allow.

Areas of Assessment Assessment of cancer patients is outlined in Table 25.11-3. It begins with the standard psychiatric history, which should also include an exploration of the patient's understanding of the illness and its meaning. Of particular importance is the prognosis: What is it? What are the patient and family being told? What are they absorbing? Major discrepancies in this area can exist among patients, family members, and staff, which lead to unnecessary stress and disruption. The next area is the impact of medical factors on mental state, and the last is the role of the environment, especially the patient's family, and their interface with the health care system. The recurrent demands of cancer treatments promote intense relationships between the patient or family and their medical caregivers, which, if dysfunctional, can lead to psychiatric symptoms and impair care.

Psychiatric
Past history
Current mental state
Understanding of the illness
Meaning of the illness
Medical
Cancer site and stage
Cancer treatment
Associated medical conditions and treatments
Environmental
Interface with the family
Interface with the medical team
Other social supports
Financial issues

Table 25.11-3 Areas of Assessment of Cancer Patients

Treatment Modalities

MEDICAL TREATMENT The psycho-oncologist must recommend medical treatments such as correcting metabolic imbalances and removing offending drugs. Optimum symptom relief must be sought, especially for pain and insomnia. Insomnia is readily acknowledged but is often mistreated with hypnotics or antihistamines that cause habituation or daytime sedation and may worsen symptoms of depression and delirium. Depressed patients benefit from a sedating antidepressant, such as amitriptyline (Elavil) or trazodone (Desyrel): mildly confused or anxious ones benefit from a low dose of a neuroleptic at bedtime. Short- to medium-acting benzodiazepines or zolpidem (Ambien) are the best hypnotics in uncomplicated situations. The psychiatrist should also work with the attending oncologist on such symptoms as nausea, vomiting, diarrhea, and dry mouth, following classically described principles of liaison psychiatry.

PSYCHOTHERAPEUTIC INTERVENTIONS A review of more than 40 psychotherapeutic interventions, which include crisis intervention, cognitive problem solving, supportive therapy with both an educational and a psychodynamic focus, showed that all had a beneficial effect. Working with individual patients or groups, the psychotherapies lowered distress, increased self-esteem and self-image, decreased anxiety, and gave patients a better sense of control. Current innovations focus on populations with special needs. Telephone counseling is an effective form of intervention in rural areas. Videophones have an added advantage for isolated homebound patients, patients in isolation, or children who need to maintain a sense of connection to a hospitalized parent. Patient reports indicate that instructional videotapes are well received.

The usual case begins in a crisis intervention mode, but if the patient remains physically well, may shift into formal therapy. Many cancer patients are happy to leave psychotherapy when they feel better, and return only episodically, when in relapse or crisis. This is not treatment resistance, it is inherent in the original treatment contract. Two major issues for patients are the reality of a shortened life span and the coming death. Despite their close connection, these are often dealt with separately, the first as a negotiable issue on a continuum with other life problems, the second as the mysterious and frightening absolute that it is. Psychiatrists must overcome their own reflexes to stay with the first and avoid the second issue. Later in their course, many patients come to an ambivalent acceptance of, even a yearning for, death. It is important to recognize both sides of the ambivalence and help families and staff tolerate and accept it as well. Families may be feeling ambivalent at this time also and need support. If the patient dies, previously known caregivers are in a unique position to offer bereavement counseling.

Groups are helpful to cancer patients because of mutual support, validation, ventilation, and learning. Formats include education groups, often led by oncology staff with minimal emphasis on expression of feelings; support groups, often led by a mental health professional with emphasis on affective expression; hybrid groups, which include both structured and unstructured portions; and self-help groups, which increase patients' sense of empowerment. Patients who received time-limited adjunct psychological therapies have shown persistence of the positive effects on coping, self-esteem, and quality of life measures, long after the intervention is terminated.

Given psychological prescreening and support, veteran patient volunteers can be invaluable to new patients for orientation, counseling, and rehabilitation. Educational and support programs can be built into the medical routine, as in presurgery education groups, postsurgery rehabilitation groups, and waiting room groups. The growing number of survivors has prompted development of innovative posttreatment programs that include educational and psychotherapeutic modalities, self-help groups, and social activism networks.

BEHAVIORAL TECHNIQUES Behavioral techniques such as hypnosis, self-hypnosis, active and passive relaxation, desensitization, and distraction are useful in a growing range of situations in oncology, including relief of acute and chronic anxiety related to surgery and painful procedures; relief or prevention of conditioned symptoms, such as anticipatory nausea and vomiting associated with chemotherapy; and improved management of almost all pediatric problems.

SOCIAL SERVICES Social workers who are available in large centers or community institutions are a first line of psychological support. They perform a triage function, alerting staff to patients with serious psychopathology. Their classic role is to help patients and families find their way to the large array of community resources. If social workers are not available, psychiatrists and other psychotherapists must fulfill this function.

PSYCHOPHARMACOLOGY Medications play an important role which is discussed below under each psychiatric syndrome. There is an important general caveat: given that cancer patients are usually on many drugs, it is imperative to know of any possible psychotropic drug interactions, especially with respect to the cytochrome P450 system.

Specific Syndromes: Diagnosis and Treatment

Psychological Factors Affecting Medical Conditions The category psychological factors affecting medical conditions should be reserved for conditions and behaviors that affect the course of the existing illness, such as nicotine addiction causing continued smoking in lung cancer patients or compulsive high-risk behaviors in immunosuppressed patients.

Adjustment Disorders Adjustment disorders with depressed mood, anxious mood, and mixed emotional features represent the largest group of Axis I diagnoses found in cancer patients. They are on a continuum, with acute stress responses at one end and mood and anxiety disorders at the other. Commonly used interventions include (1) brief individual psychotherapy focused on helping patients clarify and normalize their intense conflicted feelings and on setting and achieving short-term goals, especially those related to the illness; (2) couple or family interventions to shore up coping and reinforce communication and bonding; (3) educational and referral interventions to maximize the patient's and family's mobilization and integration into a network of support; (4) behavioral techniques to dampen physiological arousal and increase the patient's sense of control; and (5) pharmacological interventions, usually benzodiazepines, (using lower doses in very ill or older patients).

Anxiety Disorders Anxiety in response to a major illness is universal, and until recent changes in DSM classifications, it was difficult to describe the full range of patient reactions. Anxiety may be part of a normal stress response, or an acute stress disorder, or an anxiety disorder due to a general medical condition. The latter two are comparatively new diagnoses whose presentation and course in a cancer population have not been authoritatively described in a research context, but they are a welcome addition to the diagnostic lexicon. The importance of dissociative phenomena for the former must be emphasized, while the latter requires the symptoms to be a direct physiological consequence of the illness and not a subsyndromal adjustment reaction. Anxiety may be a component of delirium and major depression. It is rarely a full-blown anxiety disorder, in which it usually represents an exacerbation or a recurrence of preexisting disorder.

ETIOLOGY Drugs commonly found to cause anxiety in cancer patients include bronchodilators, frequently used in lung cancer patients; steroids, the mainstay of many protocols; interferon; and antipsychotics (including metoclopramide [Reglan] commonly used as an antiemetic) that cause akathisia, which is often mistaken for anxiety by both patient and staff. All the drugs that can cause encephalopathy and delirium can also give a picture of anxiety.

DIAGNOSIS Anxiety may be a panic disorder, with or without agoraphobia, which responds to the usual treatments plus supportive psychotherapy in relation to the cancer illness. Posttraumatic stress disorder has aroused much recent interest, and several studies have shown a high incidence of posttraumatic symptoms in patients and their parents. The full syndrome is rare, and the diagnosis may be overused, but it is useful to conceptualize many of the reactions seen in cancer patients as a reactivation of earlier traumatic responses by the cancer experience, and to see cancer recurrences as reactivations of the earlier trauma. When a diagnosis of cancer precipitates the emergence, or relapse, of delayed posttraumatic stress disorder due to previous stressors (e.g., the Vietnam war or the Holocaust), supportive psychotherapy must attempt to defuse the unconscious resonance between the cancer and the original trauma. In patients with generalized anxiety disorder and anxiety disorders not otherwise specified, one must be careful to rule out other causes of anxiety. Common ones include agitated depression, delirium (especially in a patient taking several anticholinergic drugs), and other drug adverse effects.

TREATMENT Treatment of anxiety disorders includes removal of the cause when possible, behavioral techniques, environmental interventions, and supportive psychotherapy. Pharmacological interventions follow the usual lines, but antipsychotics are often used empirically to good effect, especially when early delirium is a possibility. In more-chronic conditions, the lack of psychomotor, respiratory, and cognitive adverse effects make the buspirone (Buspar) a useful alternative. Pharmacological intervention in posttraumatic stress disorder is under intense investigation; clonidine (Catapres) and several types of antidepressants have been found useful in selected cases.

Major Depression Depressive symptoms in cancer patients may be coincidental, a symptom of a medical disorder, or a functional reaction. Several authors have attempted to discriminate the core symptoms of major depression in medically ill patients, since vegetative signs are often due to the illness. The *inclusive* approach counts all symptoms regardless of etiology. The *etiological* one counts them only if they are not related to physical illness. The *substitutive* one replaces somatic criteria with cognitive ones. The *exclusive* one uses the strictest nonsomatic research criteria. This explains the 4 to 10-fold differences in reported incidence. Self-report symptoms inventories such as the Hamilton Rating Scale for Depression (HAM-D) or the Beck Depression Inventory (BDI) are useful for screening purposes but are not very discriminating. Replacing the somatic criteria with other psychological ones improves the results, but there is no substitute for a clinical interview.

CLINICAL FEATURES Feelings of worthlessness and excessive guilt are powerful discriminants between normal sadness and major depression. Although common in cancer patients, they are mild and transient, and patients usually maintain intact self-esteem. Depressed patients, in contrast, experience intense self-loathing. Similarly, recurrent thoughts of death are common in cancer patients, but they do not have the morbid intensity and desire characteristic of depressed patients. Cognitive impairment is also a useful discriminant, although one must consider anxiety states and delirium when it is prominent. Other criteria—appetite changes, sleep changes, and loss of energy—can still be used to diagnose the cancer patient who is physically well.

DIFFERENTIAL DIAGNOSIS The differential diagnosis includes adjustment reactions with depressive mood and with mixed emotional features, dysthymia, mood disorder due to a general medical condition, and substance-induced mood disorder. A large number of medical disorders and drugs may cause an organic mood syndrome (Table 25.11-4).

Drugs
Chemotherapeutic agents: prednisone, dexamethasone, procarbazine, vincristine, vinblastine, L-asparaginase, tamoxifen, interferon
Addictive effect of narcotics and many other drugs known to cause depression, such as antihypertensives, benzodiazepines, antiparkinson agents, and β-adrenergic receptor antagonists
Tumor effects
Hormone-secreting tumors
Central nervous system tumors
Associated medical conditions
Uremia
Viral encephalopathies
Electrolyte imbalances

Table 25.11-4 Causes of Mood Disorders in Cancer Patients

TREATMENT Depressed cancer patients respond well to pharmacological treatment. Selective serotonin reuptake inhibitors have become first-line drugs, especially sertraline (Zoloft) and paroxetine (Paxil). Fluoxetine (Prozac) requires a long washout period, especially before using procarbazine (Matulane), a drug with incidental monoamine oxidase inhibitor (MAOI) features. It in turn requires a washout period before antidepressants can be used. Antidepressant selection is guided by the adverse-effect profile, gearing the choice to the patient's medical condition. Maprotiline (Ludiomil) has dopamine-blocking activity and hence may cause an increased incidence of dyskinesias in conjunction with commonly used antiemetics such as metoclopramide. Cardiac adverse effects are not usually a problem in patients with a normal electrocardiogram (ECG). Postural hypotension usually does not disturb bedridden patients. Anticholinergic adverse effects should be avoided in elderly patients at risk for anticholinergic delirium or patients at risk for gastrointestinal obstruction. Sedation is frequently a desired effect, and the use of sedating antidepressants such as nefazodone (Serzone) or trazodone obviates the need for hypnotics. Several tricyclic drugs improve the effectiveness of opiates. An average starting dose of amitriptyline is 10 to 25 mg orally at bedtime raised in 10- to 25-mg increments up to 100 mg or more. If needed, routes of administration may be varied: the oral and intramuscular routes are well known: the per rectum route has been described for amitriptyline but its pharmacokinetics are not known. Intravenous antidepressants have been rarely used outside the research setting.

MAOIs are seldom used in cancer patients because of the complex dietary restrictions and the numerous drug interactions involved with their administration. Low-dose lithium combined with an antidepressant may be useful for treatment-resistant medically ill patients. It can usually be continued uneventfully, although attention must be paid to hydration and sodium balance. Acute lithium toxicity has been reported with combined fluoxetine and lithium use. Renal insufficiency is a

contraindication, but end-stage renal disease can be managed with a 300-mg dose after each dialysis. Medically ill patients must be monitored for lithium toxicity in the presence of vomiting and dehydration (e.g., during acute cytotoxic therapy), hypercalcemia, renal impairment (e.g., with cisplatin [Platinol] treatment), possible hypothyroidism (e.g., radiation to the neck or treatment with lymphokines), or when there is an increased cardiac arrhythmia (e.g., preexisting heart disease, or use of digoxin [Lanoxin], bronchodilators, or cardiotoxic chemotherapy agents such as doxorubicin [Adriamycin]). Radiation therapy is not a contraindication, except possibly for cranial irradiation whose adverse effects mimic lithium toxicity, making monitoring difficult. Lithium's modest ability to stimulate the bone marrow is an asset during the recovery phase of chemotherapy, although colony-stimulating factor (G-CSF) has made that obsolete. It is also used to prevent and treat steroid-induced psychotic disorders, although haloperidol (Haldol), and other antipsychotic agents as well as valproate (Depakote), and clonazepam (Klonopin) are more frequent alternatives.

Psychostimulants have been used in patients with advanced cancer. Dextroamphetamine (Dexedrine) and methylphenidate (Ritalin) are given in dosages from 2.5 to 10 or 15 mg per day in divided doses, as early in the day as needed to prevent insomnia. Pemoline (Cylert) has caused rare but serious liver impairment. Increased appetite, energy, and well-being are seen within 24 hours; hence these drugs are used in terminal patients or as diagnostic probes before a course of antidepressant treatment. The dosage may be increased if tolerance develops. The main adverse effects are cardiac but are very rare, even in patients with predisposing factors.

SUICIDE Suicidal thoughts and wishes are common in cancer patients, but studies show that the incidence of actual suicide is no more than approximately twice that of the general population. The true incidence may be higher due to underreporting. Table 25.11-5 outlines suicide vulnerability factors in cancer patients. In responding to suicide threats, remember that many patients harbor suicidal feelings sometimes but learn to hold them back because of family and staff disapproval. Patients feel guilty, burdened, and angry. It is therapeutic to acknowledge the validity of their suicidal feelings and encourage ventilation without encouraging action. Treatment must be instituted for the first eight factors listed in Table 25.11-5. Inadequate symptom control is a frequent cause, and good symptom relief often stops the suicidal threats. Rarely, a competent patient presents a difficult problem by continuing to threaten or ask for help in dying. The hospitalized patient must be given intense ongoing support including open discussions about treatment options. Ambulatory patients are free to exercise all their options, although they rarely do so. Medical staff and family may need to be reminded of the competent patient's right to refuse all treatments, even lifesaving ones. This also extends to the right to stop treatment.

Depression and hopelessness
Poorly controlled pain
Mild delirium (disinhibition)
Feeling of loss of control
Exhaustion
Anxiety
Preexisting psychopathology (substance abuse, character pathology, major psychiatric disorder)
Acute family problems
Threats and history of prior attempts of suicide
Family history of suicide
Other usually described risk factors in psychiatric patients

Adapted from Breitbart W: Suicide in cancer patients. Oncology 1:49, 1987.

Table 25.11-5 Suicide Vulnerability Factors in Cancer Patients

Delirium, Dementia, and Other Cognitive Disorders

ETIOLOGY Delirium can be caused by the primary disease, by associated medical diseases, and by treatment (Table 25.11-6). Drug toxicity is probably the most common cause, most often (1) narcotic analgesics, especially drugs with a long plasma half-life, such as methadone (Dolophine) or morphine; (2) anticholinergic drugs, which are often multiple and cumulative in effect and include antiemetics (the centrally active drugs most commonly prescribed in cancer patients), antidepressants, antipsychotics, antiparkinsonian drugs, and anesthesia preparations; (3) sedatives and hypnotics, especially long-acting ones that are cumulative or those taken after ketamine (Ketalar) and halothane (Fluothane) anesthesia; and (4) drugs known to have psychoactive adverse effects, such as digitalis glycosides, b-adrenergic receptor antagonists, (beta-blockers), bronchodilators, antihypertensives, and antibiotics, such as amphotericin (Fungizone). In addition, drug withdrawal may be a common cause of delirium, especially in postoperative patients whose drug treatment was interrupted for surgery and not promptly restarted.

Medications (worsened by polypharmacy and by hypoalbuminemia, which increases the amount of circulating unbound drug)
Opioids
Benzodiazepines
Anticholinergic drugs
Sedative-hypnotics
Steroids
Cardiac glycosides, β-adrenergic receptor antagonists, antihypertensives
Anticonvulsants
Antihistamines
Bronchodilators
Antibiotics
Drug withdrawal
Metabolic and nutritional causes
Oxygen deprivation
Fluid, electrolyte, and osmolarity imbalance (sodium, calcium, phosphorus, magnesium; hyperglycemia and hypoglycemia, hypercarbia)
Vitamin deficiencies (thiamine, niacin, B_{12}, folate)
Infections, especially in immunosuppressed hosts
Vital organ failure (liver, kidney)
Vascular disorders
Endocrine and hormonal disorders
Tumors of the pituitary, thyroid, parathyroid, adrenal, and pancreas
Multiple endocrine neoplasia syndromes
Carcinoid tumors (mainly ovary, lung, stomach, and intestinal tract

Table 25.11-6 Medical Causes of Delirium in Cancer Patients

Other risk factors include increased age, preexisting central nervous system (CNS) problems, low serum albumin which raises the serum concentration of unbound drugs, other metabolic abnormalities, a history of alcohol or substance abuse, and temperature dysregulation. With an increase from 0 to 3 risk factors, the incidence rises from 2 to 50 percent, bringing longer hospital stays and poorer outcomes. Elderly persons whose delirium may unmask an underlying dementia may never return to premorbid levels of function.

DIAGNOSIS Psychiatric consultation is frequently requested by the time a patient manifests the blatant clouding of consciousness and the cognitive, perceptual, and behavioral abnormalities of delirium. In the earlier stages, however, the symptoms of delirium are subtle and routinely missed. Even as the patient becomes overtly withdrawn, suspicious, and erratic in mood and behavior, family and staff may interpret this as stress, anxiety, poor coping, or anger. They may collude in finding excuses for the patient's behavior, deny the impairment, and make inappropriate demands or scapegoat patients and blame them for their bad behavior. Benzodiazepines are often given to treat anxiety, but they worsen the symptoms of delirium because of disinhibition; more rarely, antidepressants are prescribed. Diagnosis requires a high index of suspicion. Delirium may be a hyperactive syndrome with hallucinations, delusions, and hyperarousal, or it may be hypoactive with partial obtundation. Its cause is often multifactorial, and treatment must be instituted even while the causes are being investigated.

TREATMENT Management includes the usual environmental measures to decrease confusion and guarantee patient safety (e.g., use of bed rails, night lights, repeated orientation) and an antipsychotic such as haloperidol, chosen because of its lack of sedation and overall safety. The beginning dose ranges from 0.5 mg in a cachectic patient to 2 mg or even 5 to 10 mg in a stronger one, using the intramuscular or intravenous route in agitated or disruptive patients. The dose is halved for intravenous administration; it takes effect within 15 minutes and may be repeated hourly two or three times, after which it is preferable to supplement any further use with intravenous lorazepam (Ativan) in doses of 0.5 to 2 mg every 6 hours as needed, to achieve adequate sedation. Raising the dose of haloperidol much higher increases the incidence of seizures and other adverse effects without adding much therapeutic effectiveness. However, there are reports of treating resistant delirium with doses of haloperidol up to 1 to 1.5 grams in a 24-hour period without ill effect. Thioridazine (Mellaril) or chlorpromazine (Thorazine) may be used when sedation is desired and hypotension is not an acute problem. Molindone (Moban) carries the lowest risk of seizures. Risperidine (Risperdal) and olanzepine (Zyprexa) are being increasingly used, but their use is limited in the very ill by the lack of parenteral forms. These new preparations are under development and will soon be

available.

The main adverse effects of antipsychotics relevant to cancer patients are the dystonias, including rare ones that involve muscles of respiration and deglutition, which can be mistaken for underlying disease and go untreated; akathisia, which is mistaken for agitation, delirium, anxiety, and agitated depression; and neuroleptic malignant syndrome.

Follow-up evaluations are more reliable with the sequential use of a standardized instrument, such as the Delirium Rating Scale (DRS) and the Memorial Delirium Assessment Scale (MDAS). The well-known Mini-Mental State Examination is focused on cognition alone, hence is more useful for dementia. The discussion of delirium can also be applied to other mental syndromes—namely amnestic disorder; psychotic disorder due to an organic delusional syndrome, general medical condition, with delusions or with hallucinations, and mood disorder due to a general medical condition. They are rare in pure form, and their cause should be established by the appropriate workup.

Personality Disorders Personality disorders are an infrequent cause of psychiatric consultation. Since the patients' symptoms are chronic and egosyntonic, it is often the family and staff who feel stressed by a worsening of the patient's usual dysfunctional modes of adaptation and who ask for help because of increased conflict and disrupted treatment. Nevertheless, personality-disordered patients do have acute stress syndromes, which may be overlooked because of the chronic issues. Evaluation should search for both Axis I and Axis II disorders, and any disorders should be actively treated. Environmental measures should be implemented to decrease conflict and improve patient care, including both family and staff in the intervention.

The dynamics of difficult patients and ways of dealing with them have been well described in the psychiatric literature. The feelings of anger and guilt that patients with personality disorders routinely induce in mental health staff also occur in oncology staff. The latter may find them harder to deal with because they are not trained to recognize these feelings as part of the patient's pathology and are disarmed by the genuine pathos of the patient's condition. Explanations, support, and guidance are helpful.

Schizophrenia and Other Psychotic Disorders Contrary to earlier beliefs, cancer is no rarer in patients with schizophrenia than in the general population. Schizophrenia patients may appear less often in doctors' offices because their disease interferes with early and consistent seeking of care. They give idiosyncratic meanings to their symptoms and may tolerate them without complaint, often because of a high pain threshold. The patients' unreliability continues throughout the treatment course and may complicate postoperative and long-term medical management. When treating hospitalized schizophrenia patients it is important to ensure continuity of care. Ongoing medications can usually be continued, except for a brief operative period. The medical staff needs help in handling bizarre behaviors and reassurance about intimidating mannerisms. They need a precise protocol and written orders for the management of aggressive or disruptive episodes, as well as clear understanding of the legal and ethical considerations governing the patient's right to refuse medical treatment and leave the hospital.

Bipolar Disorders Patients with bipolar disorders in remission can be maintained on lithium with few problems. Acutely manic patients must be brought under control as rapidly as possible, using intravenous haloperidol if necessary. Otherwise, they may be treated as usual.

Substance Use Disorders Alcohol withdrawal on the second or third day in the hospital may elicit a psychiatric consultation for anxiety, poor coping, or postoperative delirium. Thoracic, head, and neck cancer patients are at high risk for this, as well as for nicotine withdrawal since they stop smoking upon diagnosis or admission. The symptoms are those of anxiety and can be severe enough to include depersonalization and marked cognitive distortions. Fear is eased by giving explanations and symptoms are relieved by benzodiazepines. Transdermal nicotine patches (Nicotrol, Habitrol, Nicoderm CQ), are increasingly used to ease the distress of patients forced to stop smoking suddenly because of a diagnosis of cancer. Other substance withdrawal syndromes rarely occur. Surreptitious drug use is more likely, manifesting itself as a fluctuating organic mental state with erratic difficulties in the delivery of care.

Somatoform Disorders The development of cancer in patients with a somatoform disorder can create difficult diagnostic problems because patients incorporate the cancer into their psychiatric disorder and elaborate on physical symptoms in their usual manner. That is, the somatization disorder patients continue to plead for surgery for a multiplicity of symptoms; the conversion reaction patients become paralyzed at the first hint of real leg weakness; and the psychogenic pain patients maintain the same abnormal relationship to pain they have always had. The resultant syndromes are complex, involving both organic and psychogenic components. Physicians often feel manipulated and frustrated by these compulsive behaviors and in trying to correct them either underestimate them and delay treatment or give up and overtreat psychosomatic symptoms. A psychiatric evaluation and ongoing psychotherapeutic relationship will let the psychiatrist work with the oncologist and staff to set limits and treat the patient appropriately.

Factitious Disorders Patients may enter a hospital claiming to have cancer and turn out to have a chronic factitious disorder with physical symptoms. Their presentation and course fit the usual description, and on being confronted, they often leave angrily to seek other care.

Cancerophobia The negative reaction to cancer in our society often makes it the object of excessive fear, comparable to syphilis or germs in the past and acquired immune deficiency syndrome (AIDS) today. Transient cancerophobia may be a response to a specific event (e.g., a cancer death in the family) and often occurs when new personnel start work on a cancer ward. Fixed cancerophobia ranges from a neurotic symptom to a hypochondriacal fixation, an obsession, or a somatic delusion. It is associated with a broad spectrum of diagnoses, and management depends on making the appropriate diagnosis. The symptoms are often quite resistant.

Psychiatric Syndromes and Tumor Site

Primary Nervous System Cancers Tumors that arise in the CNS (brain or spinal cord) account for 3 percent of deaths in all cancer patients. They account for 20 percent of all childhood tumors and are the second most common tumor in children under age 15. Psychiatric symptoms are present in almost all patients with supratentorial tumors and are the presenting symptom in 25 percent of these patients. In one series, they represented the presenting complaint in 12 percent of patients with subtentorial tumors. There are also reported cases of spinal cord tumors initially diagnosed as conversion reactions. Presenting symptoms vary enormously depending on the location of the tumor and whether it is invasive or encapsulated. They commonly include personality change, anxiety disorders, and mood disorders.

Metastatic Neoplasms Brain metastases are the most common complication of systemic cancer, occurring in 25 to 30 percent of patients. Their frequency is increasing as cancer treatment improves survival, and the central nervous system provides a sanctuary for malignant cells hiding behind the blood-brain barrier. Lung tumors account for 35 to 40 percent of CNS metastases; breast, kidney, and colon tumors account for approximately 15 percent each; and other tumors, malignant melanoma, carcinomas of the bladder, liver, thyroid, testicle, uterus, and ovary account for the rest. Leukemia spread to the brain was more common before CNS prophylaxis by radiation and intrathecal therapy. In a series of 199 patients, headache (49 percent) and mental disturbances (32 percent) were the most common presenting symptoms, and cognitive impairment (58 percent) the most common presenting sign. It was far more frequent than papilledema.

SPINAL METASTASES Spinal metastases occur in 5 to 10 percent of patients with cancer, usually in late stages of the disease, but they may be the presenting sign in as many as 8 percent of patients. The diagnosis is often delayed, even in advanced disease, if patients' pain complaints are not taken seriously. In one series, epidural spinal cord compression was the presenting symptom for 47 percent of primary care patients.

LEPTOMENINGEAL METASTASES Also known as carcinomatous meningitis, leptomeningeal metastases are the diffuse, multifocal seeding of the leptomeninges by cancer. They are found in 5 to 8 percent of patients with solid tumors, 5 to 29 percent of patients with non-Hodgkin's lymphoma, and 11 to 70 percent of patients with leukemia. Psychiatric symptoms are often present (e.g., anxiety, mood instability, personality change), sometimes without focal signs. When neurological symptoms are present, they relate to widely separated parts of the neuraxis and may lead to diagnostic skepticism and an early psychiatric diagnosis.

Nonmetastatic Neurological Complications of Cancer Many local and systemic factors contribute to neurological complications in cancer patients. These are particularly relevant to psychiatry because a mild or subacute encephalopathic presentation is common in debilitated cancer patients and can be initially attributed to psychological factors.

CENTRAL NERVOUS SYSTEM INFECTIONS Because of severe immunosuppression caused by chemotherapy, cancer patients are susceptible to a wide range of infectious agents. Mild headache and low-grade fever may be the only systemic factors accompanying mental state changes. Fungal agents have become as common as bacterial ones, also accounting for the less acute presentation. Opportunistic viral infections can cause leukoencephalopathy.

CEREBROVASCULAR DISEASE Cerebrovascular diseases cause symptoms in only 7 percent of cancer patients, but they present as encephalopathy more

commonly than with acute or localizing signs. They are associated with the bleeding tendencies secondary to bone marrow depression, with the coagulopathies commonly found in cancer patients, and with locally induced tissue damage and treatment toxicity.

PARANEOPLASTIC SYNDROMES The *paraneoplastic* refers to an uncommon group of disorders, defined as remote, nonmetastatic effects of cancer on the nervous system. In 50 percent of cases, the symptoms precede the discovery of the cancer and run an independent course. They can affect all parts of the nervous system. Dementia, particularly with loss of memory, is a common feature in several variants, which are grouped under the name “carcinomatous encephalomyelitis,” of which several cases have turned out to be leukoencephalopathies. Mild-to-severe cerebellar dysfunction and abnormalities of peripheral nerve muscle function are commonly found as well. Small cell lung cancer, Hodgkin's lymphoma and ovarian cancer are the most frequent associated tumor types. Suggested causes include toxins secreted by the tumor, competition for essential substrates, opportunistic infections, and most likely autoimmune reactions mounted against the underlying tumor.

Endocrine and Hormonal Effects of Tumors Tumors of endocrine glands may be physiologically active and produce anxiety and mood disorders, confusional states, and occasionally psychosis. Carcinoid tumors can cause anxiety syndrome. Carcinoma of the pancreas reportedly has been associated with an increased incidence of depression, sometimes as the presenting symptom.

General Clinical Considerations No specific picture is associated with this wide range of CNS disorders, which require a high index of suspicion for accurate diagnosis. Signs of cognitive impairment in older patients are too easily dismissed without a full medical workup. Signs of affective ability are labeled mood disorder. Symptoms of pain are too easily thought to be functional if the first workup does not yield an organic cause. Symptoms of a chronic nature may be too readily labeled character pathology, and associated physical complaints dismissed as conversion reactions. This is complicated by the fact that true conversion reactions in cancer patients frequently present as an elaboration of an organic symptom. Suspicion should be aroused by an abrupt change from the premorbid personality, an unusual age of onset for symptoms, a lack of family history, associated physical symptoms, and a poor response to treatment.

Psychiatric Problems Due to Cancer Treatment Radiation and most of the drugs used in cancer treatment have significant adverse effects, and are even carcinogenic in tumoricidal doses. Patients may become transiently more ill from the treatment than from the disease, developing adverse effects such as hair loss that have great emotional impact. Steady compliance requires more courage and endurance than medical staff realizes. Patients may become upset about the adverse effects, convinced they represent disease progression, depressed about the paradoxical nature of treatment, and generally so exhausted they develop clinical depression and anxiety. If untreated, they delay or stop treatment, moving to alternative treatments and jeopardizing their survival. Later adaptation is also impaired. Many patients, however, come to feel dependent on continued treatment as a means of holding the cancer at bay. Anxiety levels rise at the end of treatment, associated with this fear and with the difficulty of making a transition to the survivor status. This can be assuaged by acknowledging it, exploring its cause, and making concrete plans for continued medical supervision and psychological support.

Radiation Therapy Radiation-induced encephalopathy is the most common neuropsychiatric complication of radiation treatment. It takes several forms. The first is an uncommon, immediate (minutes to hours) syndrome with symptoms of increased intracranial pressure. The second is an early-delayed (6 to 16 weeks) syndrome with headaches, somnolence, and sometimes no focal signs. Both respond to steroids and are usually self-limited. The third is an irreversible late-delayed (months to years) form associated with atrophy or necrosis. It presents with headaches and personality changes and only later progresses to focal signs and seizures. Irreversible focal radiation necrosis has also been described.

Other forms of radiation-induced cognitive dysfunctions have come to the fore in a growing population of long-term survivors (mainly from Hodgkin's disease and childhood leukemia) who have undergone whole brain radiation. A short-term memory deficit has been observed, as well as a long-term decrease in intelligence quotient (I.Q.) of as much as 10 to 25 points. Even in the presence of a normal I.Q., neuropsychological problems and learning disorders have been documented. Although the studies in this area have many shortcomings, it is generally believed that the threshold of the brain's tolerance to radiation must be revised downward. Elderly and very young persons are the most affected. A recent controlled study in postradiation brain tumor patients showed that methylphenidate (Ritalin) improved cognition, mood, and functioning.

Radiation therapy can be associated with a high level of psychological distress. Nausea, vomiting, and mild-to-moderate fatigue can lead to fear, anxiety, depression, and occasionally major depressive episodes. Sexual impairment is common following radiation to the pelvic area—impotence in men and dyspareunia in women. The latter, related to vaginal fibrosis, is improved by dilatation. Acute transient gastrointestinal problems (e.g., diarrhea and food intolerance) are common; severe irreversible ones are rare.

Chemotherapy The immediate effects of chemotherapy include nausea and vomiting, which occur within 2 to 6 hours and last up to 24 hours. Table 25.11-7 lists the drugs most likely to cause these symptoms. The most emetogenic drugs cause symptoms in virtually all patients; the less emetogenic ones are variable. The reasons for the marked interpatient variability are not understood. Both chemotherapeutic drugs and radiation cause vomiting by central mechanisms.

Severe and moderate emesis (requires prophylactic combination antiemetics)	Mild emesis (requires occasional antiemetics)
Cisplatin	Etoposide
Dacarbazine (DTIC)	Mitomycin
Dactinomycin	Plicamycin
Mechlorethamine (nitrogen mustard)	Ifosfamide
Carboplatin*	Methotrexate
Cyclophosphamide*	5-Fluorouracil
Doxorubicin	Hydroxyurea
Daunorubicin	Bleomycin
Idarubicin	Vinblastine
Cytarabine	Vincristine
Procarbazine	Chlorambucil
Carmustine (BCNU)	
Lomustine (CCNU)	

* Peak emesis: 8 to 12 hours after administration.

Table 25.11-7 Emetogenic Potential of Commonly Used Anticancer

Classical conditioning is an important mechanism in the causation of several reactions to chemotherapy: anticipatory nausea and vomiting, which occurred in 25 to 75 percent of patients before the use of better antiemetics; anticipatory immune suppression, whose clinical significance has not been established; and learned food aversion, which can be minimized by avoidance of staples and favorites. Patients develop a conditioned response to sights and smells reminiscent of the chemotherapy experience. After termination of treatment, these symptoms lessen, but one study found that 6 to 140 months after treatment, 63 percent of patients still reported nausea, 5 percent reported vomiting, and 80 percent still reported anxiety with specific stimuli.

All these reactions are presumed to represent conditioned responses, and treatment approaches have been formulated accordingly. Prophylactic antiemetic and antianxiety treatments are used to prevent or decrease the initial episode or conditioned response (Table 25.11-8). With the advent of highly specific, centrally acting antiemetic agents such as the serotonin (5-hydroxytryptamine [5-HT]) type 3 (5-HT_3) blockers ondansetron (Zofran) and granisetron (Kytril), the success of antiemetic regimens has greatly increased. Other cheaper agents such as metoclopromide, lorazepam, and dexamethasone (Decadron) continue to be commonly used as well. The usual regimens still include lorazepam, steroids, metoclopramide, or prochlorperazine (Compazine). The latter two are neuroleptics with the usual side effects, but the briefness of administration makes tardive dyskinesia rare. Akathisia occurs in 10 to 20 percent of patients on metoclopramide, but coincident use of lorazepam lowers the incidence to below 3 percent. Mild sedation is acceptable or even welcome. Hypotension may limit the use of prochloperazine and chlorpromazine. Dronabinol (Marinol) a tetrahydrocannabinol, is no more effective than other drugs and has a similar adverse effect profile plus a significant incidence of dysphoric reactions in older patients. However, the inhalation route is effective in experienced users. Its legalization is a controversial issue in several states, where its use may constitute a felony. Dexamethasone (Decadron) in one intravenous bolus is generally well tolerated.

Neurotransmitter blocking agent	Metoclopramide* 3 mg/kg IV, 30 min before therapy and 1 1/2 hours after therapy or Ondansetron 0.15 mg/kg, IV, 30 min before therapy and 1 1/2 and 3 hours after therapy
plus	
Steroid	Dexamethasone* 20 mg IV, 20 min before therapy
plus	
Benzodiazepine	Lorazepam 1.5 mg/m² (max, 3 mg) IVPB before thetapy
or	or
Antihistamine	Diphenhydramine* 50 mg—oral, IV, or IM—every 4 hours as needed for restlessness or acute dystonic reaction

* Should also be used in oral form for delayed nausea and vomiting, starting 24 hours after cisplatin therapy.

Table 25.11-8 Current Chemotherapy Antiemetic Regimens

Behavioral methods are important, effective ways to lessen the reactions to chemotherapy, and are most effective when instituted before the conditioned response is established. They include hypnosis and progressive muscular relaxation with guided imagery or relaxation with cognitive distraction, to block both the conditioned stimulus and the conditioned response; systemic desensitization, which extinguishes the conditioned response; and cognitive distraction, which blocks perception of the conditioned stimulus. Patients are also helped by reassurance about the frequency and normalcy of their responses.

CNS Toxicity Chemotherapeutic agents may cause neuropsychiatric problems indirectly by causing metabolic abnormalities or other organ failure, or by direct CNS toxicity. Although all chemotherapeutic agents are selected because they interfere with basic cell processes, they are remarkably well tolerated and have surprisingly little CNS toxicity. However, few problems are found because few are looked for, and mild impairment may be the rule rather than the exception. This view is of pragmatic importance because underdiagnosed organic mental disorders lead to further disqualification of patient complaints and predispose to further psychological decompensation and management problems. Table 25.11-9 lists the agents most likely to be associated with neuropsychiatric complications. Some drugs, such as vincristine (Oncovin, Vincasar), cause neurotoxicity in all dosage schedules; others, such as cisplatin cause neurotoxicity only at higher doses. Carmustine (BiCNU) is neurotoxic mainly when given by the intracarotid route or in conjunction with radiation therapy. Some drugs have mild, transient toxicity; others have delayed or irreversible toxicity, or both. It is usually (but not always) dose related and cumulative. Some drugs, like methotrexate (Folex, Mexate), when given intrathecally, may cause several syndromes: a reversible encephalopathy, a delayed irreversible one, and a variable dementia or developmental delay when coadministered with radiation therapy. Initially, mood changes, personality changes, and subtle cognitive changes may predominate, leading to psychiatric consultation. Only later does the encephalopathic nature of the condition become manifest.

Table 25.11-9 Neurological Complications of Chemotherapeutic Drugs

Combination of radiation with chemotherapy and the use of the intrathecal route or of very high doses leads to a higher incidence of problems. Recognition may be complicated because the changes may occur months to years after treatment.

Table 25.11-10 lists agents for which mood or psychotic symptoms have figured prominently. All of these also cause cognitive disorders. (Anxiety has not been included because it is ubiquitous.) These syndromes are often reversible and remit spontaneously if drug administration is stopped. This may take only a few days or up to several weeks. If the offending drug cannot be discontinued, symptomatic treatment should be instituted.

Dacarbazine: depression and suicide reported, especially when used with hexamethylmelamine
Vinblastine: frequent reversible depression
Vincristine: 5 percent incidence of hallucinations: depression noted
L-Asparaginase: reversible depression noted
Procarbazine: MAOIs, concurrent tricyclic drugs, and selective serotonin reuptake inhibitors are contraindicated: associated with mania and depression: potentiates alcohol, barbiturates, phenothiazines
Hydroxyurea: hallucinations reported
Interferon: anxiety, depression with suicidal ideation common at high doses
Steroids: frequent alterations of mental state, ranging from emotional lability through mania or severe, suicidal depression to frank psychosis
Tamoxifen: depression: rare delusional syndrome

Table 25.11-10 Chemotherapy Agents Associated With Mood and Psychotic Symptoms

Disorders of special senses include ototoxicity, associated with cisplatin and several commonly used antibiotics; visual disturbances, associated with mitotane (Lysodren) and tamoxifen (Nolvodex); and gustatory and olfactory hallucinations described with vinblastine (Velban). These disorders may occur slowly or suddenly, reversibly or permanently, and often result in acute psychological distress requiring psychiatric consultation. Long-term effects of ototoxicity in children add considerably to the educational and neuropsychiatric impairment of these patients. The diagnosis is often missed in younger children, the very ones in whom early intervention is essential to minimize later developmental impairment.

Antineoplastic Hormones and Biological Modifiers The hormonal agents listed in Table 25.11-11 are used mainly in prostate and breast cancer. They are associated with depressive symptoms and many adjustment disorders related to their sexual adverse effects such as feminization, masculinization, and disorders of sexual functioning. Of the biological modifiers listed in Table 25.11-11, the cytokines cause the widest range of confusional and affective states. The interferons, interleukin-2, tumor necrosis factor, and colony stimulating factors are naturally present in the body in low concentrations. Their normal elevation during naturally occurring infections is responsible for what patients describe as “flu” symptoms. Recombinant DNA technology has made them available for use against cancer. Interferon causes an acute reversible encephalopathy with confusion, lethargy, and hallucinations. One case of delayed progressive encephalopathy was reported. A reversible depression is common, with severe fatigue and loss of concentration. Aldesleukin (Proleukin), a human interleukin-2 derivative, causes headache, agitation, disorientation, and confusion, which are dose-related and reversible. A rare, severe encephalopathy was also reported. These drugs have become frightening to patients who hear of the multiple adverse effects. The severe depression may respond to antidepressant medication.

Types of antineoplastic hormones
Androgens (danazol)
Antiandrogens (flutamide for prostate cancer)
Estrogens (diethylstilbestrol [DES])
Antiestrogen (tamoxifen for breast and ovarian cancer)
Progestins (megestrol for anorexia)
Adrenocorticoids (dexamethasone)
Antiadrenals (aminogluthamide)
Antiprogestins (mifepristone)
Aromatase inhibitors (aminoglutethimide)
Types of antineoplastic biological
Antibodies (epidermal growth factor receptor)
Immune cells (lymphokine-activated killer cells, tumor-infiltrating lymphocytes)
Vaccines (melanoma cells)
Immune enhancers (levamisole)
Differentiation agents (retinoic acid)
Cytokines (interleukin-2, interferon)
Immunosuppressants (immunoglobulins)

Table 25.11-11 Antineoplastic Hormones and Biological Modifiers

Tailor-made vaccines against various parts of the malignant cell are being developed in increasingly sophisticated ways. They remain difficult to make, the results are partial, they are limited to research-based programs, but they are of great scientific interest and have much unexplored potential. They also give rise to near magical hopes in many sophisticated patients who must deal with the currently inevitable disappointments.

Surgery Many studies on the psychological effects of mastectomy, hysterectomy, colostomy, and prostatectomy report the high cost paid in quality of life and recommend that more emphasis be placed on this dimension. Laryngectomy patients and those who undergo extensive head and neck procedures suffer considerable psychiatric morbidity; the loss of voice is catastrophic for many, and the blow to self-image severe. Patients who have had curative limb amputations can have a good psychological outcome. There is little systematic information about patients who undergo unusual and extreme procedures, such as pelvic exenteration, hemipelvectomies, and forequarter amputations.

Bone Marrow Transplantation Bone marrow transplantation is being used for an increasing number of childhood and adult cancers. It is no longer an innovative treatment, but is expensive enough that insurance companies sometimes still call it “experimental” in some cancer patients, to avoid paying for it. It is commonly recommended for leukemias and lymphomas whose prognosis is poor with other treatments, and it is increasingly used to treat advanced solid tumors such as breast cancer. It is a potentially curative treatment for otherwise lethal disorders. It succeeds best when given to patients in remission, with as little previous treatment as possible. Yet it has a baseline mortality of 5 to 15 percent, an uncomfortable course, and a significant incidence of graft-versus-host disease. Undertaking such a course of action, especially when the patient is doing relatively well, can be very stressful for patients, families, and caregivers.

New Supportive Technologies The use of life supports and nutritional supports in the critically ill patient has created a spate of psychological, legal, and ethical issues that are widely discussed in the lay press. Less well known are developments in the nutritional support of cancer patients who can no longer ingest food because of physical obstruction (e.g., in head and neck cancers) or because of physiological failure (e.g., radiation enteritis). In the former, a percutaneous endoscopic gastrostomy or jejunostomy, and in the latter, home parenteral nutrition allow a significant improvement in quality of life. It also relieves the family from the anguish of watching the patient “starve to death,” which engenders feelings of guilt, which may turn to anger toward the medical staff. Of these home-based technologies, home parenteral nutrition creates the most major changes in family structuring routines such as the preparing and sharing of meals and are time and labor intensive, causing adaptational strains that may bring them to psychiatric attention.

Cancer Pain Management

Incidence Cancer pain is prevalent and varies with stage and site of disease. An overall incidence of 51 percent rises to 74 percent in advanced disease, of which 40 to 50 percent is said to be moderate to severe and 20 to 30 percent excruciating. Treatment failures range from 10 to 30 percent, depending on the skill of the treating physician. The psychiatrist has a special role to play, especially as pain specialists are rare.

Pain and Psychiatric Symptoms Patients with pain have a significantly higher incidence of depression and anxiety. It has been suggested that chronic pain may be a depressive equivalent, that pain may cause psychiatric syndromes, and that pain may coexist with psychopathology in vulnerable subjects. In cancer patients, the evidence suggests that these emotional reactions both result from, and contribute to, the experience of pain and that treatment of one improves the other. A review of the impact of biobehavioral factors on adult cancer pain concluded that the role of personality factors was inconsistent, the relationship to affective states minimal, the environmental influences weak, and the role of cognitive factors unexplored. Hence, the psychiatric consultant is wise to avoid diagnostic inferences that minimize the patient's complaints. Patients with significant psychopathology are indeed more difficult to evaluate, hence the consultant must help the staff make the same aggressive efforts at symptom relief for these patients as for other patients. The incidence of psychiatric complications is particularly high when pain is underestimated and undermedicated by caretakers, an event that recurs with a regularity that cries out for explanations.

Undertreatment The undertreatment of pain in cancer patients is a resistant problem that has frustrated specialists in the field for many years. The most concrete causes are lack of available resources and knowledge gaps. However, even knowledgeable, compassionate staff are afraid of “addicting” patients with metastatic cancer. And if not that, they may continue to minimize patient complaints and measure them against some unspoken internal standard of how much pain is expectable in a given situation. Their need to underestimate pain is intimately related to their deepest motivations and defenses in doing this difficult work. They must protect themselves against too much empathic identification as well as sadistic arousal and guilt. An occasional lapse into blaming the victim is an all-but-irresistible human response to such a difficult situation. Practitioners become less attuned to their patients' pain as the pain becomes more severe. They also often do not differentiate between depression, anxiety, pain, or other inadequate symptom control.

Psychiatric Interventions Psychiatrists and other outside consultants, such as pain specialists, can reorient staff responses in a supportive, nonjudgmental way. They can emphasize the need to respect the patient's statements and model the use of a concrete instrument such as the Memorial Pain Assessment Scale. Daily charting makes it harder for staff to ignore persistent patient complaints. Psychiatrists can also give staff a sympathetic and respectful insight into the patient's psychological state and undermine the staff's covert pain standard by emphasizing the extent of the patient's illness and the progressiveness of the disease; they can sympathize with the staff's own difficulties and uncertainties, ensure adequate pain management along the principles outlined below, and educate (and reeducate) about the low addiction risk. Psychiatrists' role in bringing about optimal pain relief is a powerful intervention in itself, but it is most complete when integrated into supportive psychotherapy oriented to helping the patient to differentiate between pain and suffering and develop appropriate responses to both (Table 25.11-12).

Behavioral approaches have been effectively used by patients who claim an increased sense of control, decreased anxiety, and decreased pain. In one study, doctors appeared to underestimate the value of these methods, compared with the patient's own estimate of them. Hence, it is useful for psychiatrists to make a point of recommending and administering them. They include relaxation techniques, guided imagery, distraction techniques, and biofeedback (Table 25.11-13).

Goals
Support
Providing continuity
Knowledge
Providing information
Skills
Relaxation
Cognitive coping
Use of analgesics
Communication
Self-observation
Documentation
Forms of therapy
Individual
Supportive, crisis intervention
Family
Patient and family are the unit of concern
Group
Sharing experience
Identifying successful coping strategies

Reprinted with permission from Breitbart W: Psychiatric management of cancer pain. Cancer 63:2316, 1989.

Table 25.11-12 Psychotherapy for Cancer Pain

Cognitive therapy	Behavior therapy
Preparatory information	Self-monitoring
Cognitive restructuring	Systematic desensitization
Focusing	Graded task management
Controlled mental imagery	Contingency management
Distraction	Modeling
Controlled attention	Behavioral rehearsal
Mental, behavioral	Relaxation
Music therapy	Passive, progressive
Hypnosis	Meditation
Biofeedback	Music therapy
	Hypnosis
	Biofeedback

Table 25.11-13 Cognitive-Behavior Therapy Techniques for Cancer Patients

Types of Patients For treatment purposes, cancer patients with pain fall into 5 categories.

1. *Acute pain cancer patients* respond well to treatment but develop tolerance to all known strong analgesics and require more medication if the pain lasts beyond a few days or weeks. This is often inappropriately viewed as “addiction” or drug-seeking behavior by staff. Studies show that these patients wean themselves easily and voluntarily when their pain eases.
2. *Chronic pain cancer patients;* the treating physician must not lose sight of the progression of disease in these patients, even if it is very slow. Unlike nonmalignant chronic pain patients, whose symptoms may not change over many years, cancer patients generally do have worsening pain and the doctor must respond.
3. *Chronic pain patients who develop cancer* may have addictive problems and are at high risk for escalating psychological problems. Nevertheless, their cancer pain creates an additive requirement for medication.
4. *Cancer patients with a drug abuse problem*. If actively abusing or taking methadone, these patients are already tolerant and must be given a correspondingly larger amount of medication to achieve adequate pain control. This must not be viewed as abusing behavior. If not currently abusing, these patients are at risk for relapse, but they need sympathy and support, not a suspicious, withholding approach.
5. *Dying patients* may be adequately medicated, but one needs to address all physical symptoms, emotional suffering, and existential issues as well. A poignant dilemma is that of pain relief versus alertness. If the need for opioids is so great, moments of wakefulness can be increased by using psychostimulants.

Types of Pain Cancer patients are subject to several different kinds of pain, summarized in Table 25.11-14. The distinctions are important because they call for different treatment strategies; somatic and visceral pain is more responsive to opiates, while neuropathic and sympathetically maintained pain responds better to adjuvant medications. Most advanced cancer patients have more than one kind of pain and require complex treatment regimens. Table 25.11-15 and Table 25.11-16 outline the neurophysiology of pain. Table 25.11-17 shows pain syndromes in cancer patients.

Nociceptive pain	
Somatic pain	Usually but not always constant, aching, gnawing, and well localized; e.g., bone metastases
Visceral pain	Usually but not always constant, deep, squeezing, poorly localized, with possible cutaneous referral; e.g., pleural effusion leading to (1) deep chest pain, (2) diaphragmatic irritation referred to shoulder
Neuropathic pain	Burning dysesthetic pain with shocklike paroxysms associated with direct damage to peripheral receptors, afferent fibers, or CNS, leading to loss of central inhibitory modulation and spontaneous firing; e.g., phantom limb pain; can involve sympathetic somatic afferents
Psychogenic pain	Variable characteristics, secondary to psychological factors in the absence of medical factors; vanishingly rare as a pure phenomenon in cancer patients but often an additional factor in the presence of organic pain

Table 25.11-14 Types of Pain

Table 25.11-15 Neurophysiology of Nociceptive Pain

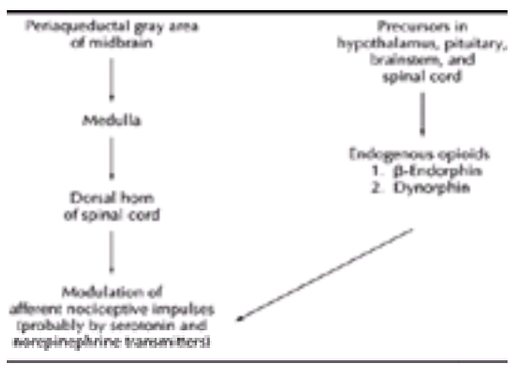

Table 25.11-16 Endogenous Pain Suppression Pathways

Table 25.11-17 Pain Syndromes in Cancer Patients

Diagnosis A careful evaluation may require an etiologic workup, even in a patient receiving only supportive care, because certain pains can be eradicated even in very ill patients. One should weigh the benefits of better pain relief against the intrusion of diagnostic and therapeutic procedures in each individual case.

Treatment Aggressive multimodal treatment should be the norm. Drugs are the mainstay but should always be part of a broader approach. Other modalities include palliative radiation and anesthetic and neurosurgical techniques (Table 25.11-18), behavioral interventions, and psychological support.

Anesthetic approaches
Myofascial trigger-point injections
Peripheral, autonomic, and intrathecal nerve blocks
Nitrous oxide
Surgical approaches
Tumor resection and debulking
Shunting and decompression
Neurosurgical anesthetic approaches
Neuroablative procedures; e.g., neurectomy, tractectomy, cordotomy
Neurostimulatory procedures; e.g., transcutaneous electrical nerve stimulation (TENS)
Neuropharmacological procedures; e.g., epidural and intrathecal opioids and local anesthetics

Table 25.11-18 Anesthetic and Surgical Approaches to Cancer Pain Management

The World Health Organization (WHO) three-step ladder for cancer pain control (Fig. 25.11-1) is a drug-treatment method with documented effectiveness in many settings. The use of drugs begins with nonnarcotic analgesics, moves to moderately strong opiates, then to strong opiates, culminating in a continuous morphine drip if necessary. Nonnarcotic analgesics may be used throughout because they potentiate stronger ones: aspirin and indomethacin (Indocin) inhibit prostaglandin E_2, which is known to be active in the development of bone metastases. Along with biphosphanates, they play a special role in treating bone pain. Adjuvant drugs should be used simultaneously throughout. Table 25.11-19 and Table 25.11-20 give the drug characteristics and Table 25.11-21 gives the treatment guidelines.

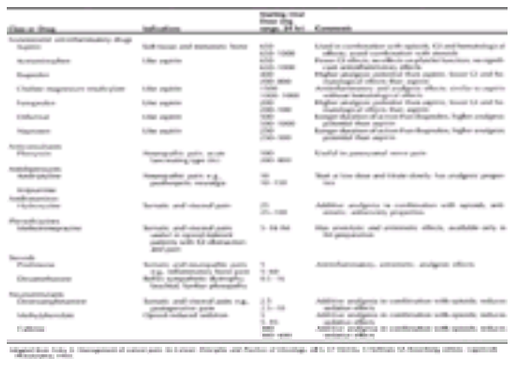

Table 25.11-19 Nonopioid and Adjuvant Analgesic Drugs in the Management of Cancer Pain

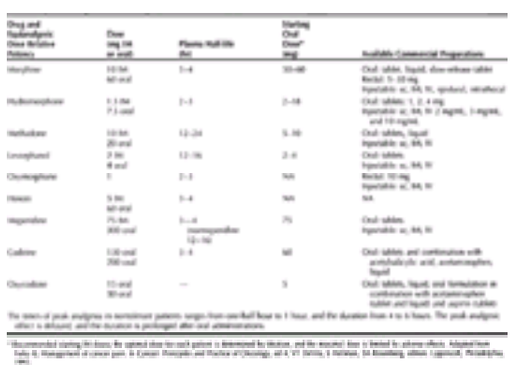

Table 25.11-20 Opioid Analgesics for Management of Cancer Pain

Table 25.11-21 Guidelines for the Rational Use of Analgesics in the Management of Cancer Pain

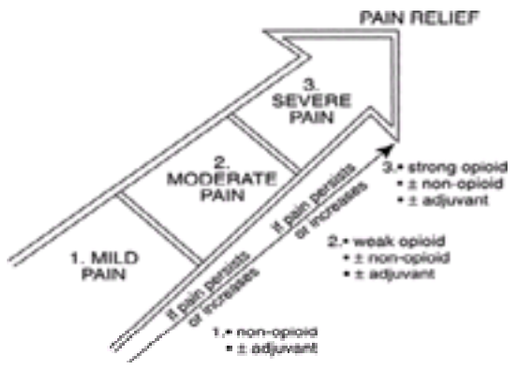

FIGURE 25.11-1 A simple paradigm for the appropriate drug treatment of increasing levels of pain. The World Health Organization's three-step ladder for pain control was developed to rationalize the pharmacological treatment of differential levels of pain. The combinations of drugs that are most appropriated for mild, moderate, and severe pain are indicated. The drugs recommended for a higher level should not be used until the drugs for lower levels of pain have exhausted their effectiveness. However, drugs of increasing strength should be readily used as needed to keep pain under control. (Adapted from Foley KM, Inturrisi EC: Analgesic drug therapy in cancer pain: Principles and practice. In *Medical Clinics of North America: Cancer Pain*, R Payne, KM Foley, editors. Raven, New York, 1987.)

A good pain regimen may require several drugs or the same drug used in different ways and administered via different routes. For example, intravenous morphine may be supplemented by self-administered oral rescue doses, or a continuous epidural drip may be supplemented by bolus intravenous doses. Transdermal patches may provide baseline concentrations in patients for whom intravenous or oral intake may be difficult. Patient-controlled analgesia systems for intravenous opiate administration result in better pain relief with lower total dose and patient self-tapering if pain improves.

Opioids are a common cause of delirium and hallucinosis. A frequent mechanism of psychotoxicity is the accumulation of drug or metabolites whose duration of analgesia is shorter than their plasma half-life (morphine, levorphanol [Levo-Dromoran] and methadone). Using drugs such as hydromorphone (Dilaudid) with half-lives closer to their analgesic duration can relieve the problem without loss of pain control. Cross tolerance is incomplete between opiates, hence several should be tried in any patient, with the dosage lowered when switching drugs.

The importance of maintenance as opposed to as-needed administration cannot be overemphasized. It improves pain control, increases drug efficiency, and relieves patient anxiety; as-needed orders do not provide the immediate response of patient-controlled analgesia and allow pain to increase while waiting for the rug to be given. It perversely sets up the patient for staff complaints about drug-seeking behavior. Even on maintenance treatment, extra doses should be available for breakthrough pain, and their repeated use should signal the need for raising the maintenance dose.

Knowing equianalgesic doses of different drugs and different routes of administration is important to avoid accidental undermedication. For example, when changing a patient from intramuscular to oral morphine, one should multiply the intramuscular dose by 6 to avoid causing the patient pain and provoking drug-seeking behavior.

Many adjuvant drugs are psychotropics with which the psychiatrist is very familiar, but in some cases, their analgesic effect is separate from their primary psychotropic effect. Commonly used adjuvants include antidepressants, phenothiazines, butyrophenones, antihistamines, amphetamines, and steroids (Table 25.11-19). They are particularly important in neuropathic and sympathetically maintained pain, in which they can be the mainstay of treatment.

Other developments in pain control include more-intrusive procedures such as nerve blocks or the use of continuous epidural infusions starting before surgery to prevent phantom pain and other chronic postoperative pain syndromes.

Palliative Care Despite increasingly visible efforts by dedicated specialists, palliative care is still not the rule. If it were, the right-to-die movement would arouse far less public support. Lack of training is a major cause, but the transition from active treatment to palliative care can also be emotionally difficult for physicians, patients, and families. It is sometimes avoided altogether, and active or experimental treatment is continued to the end; this may be satisfactory if it truly meets the patient's and family's emotional needs. Sometimes, a pretense of active treatment is continued. This may work but is likely to cause psychological problems if it is being done because of medical avoidance in the face of patient and family ambivalence. A well-negotiated transition to palliative care often decreases anxiety after the patient and family go through an appropriate grief reaction. A positive outcome is much more likely if the physician and staff project a conviction that palliative care will be an active, involved process, with no hint of withdrawal or abandonment. When this fails or when the family cannot tolerate the transition, the ensuing stress frequently results in psychiatric consultation.

The psychiatrist must correct any therapeutic defeatism present, especially in the areas of relief of psychiatric symptoms, pain, insomnia, and vomiting, and support the oncologist if disengagement is an issue. This is best done by role-modeling careful attention to all aspects of patient care. Strengthening caregivers' intellectual defenses and providing concrete opportunities for intervention diminishes withdrawal behaviors and benefits patients directly. Other symptoms that benefit from psychopharmacological interventions include fatigue, cachexia, anorexia, dyspnea, and sedation as well as opioid and steroid adverse effects.

Another major transition involves deciding whether to have the patient die at home or in a hospice—a sensitive question with many implications for survivors. Hospices often have built-in psychological supports that are not available for home care patients and must be provided as needed on an individual basis.

One should engage terminally ill patients in dynamic issues to the extent that they are physically able to participate. Often, the most precious gift one can give them is to reawaken their interest in the richness of their own experiences. This may be their one remaining possession, as the others slip away. Existential despair and spiritual distress often evoke feelings of helplessness and ignorance in the therapist, leading to their being ignored or shunted to a religious advisor. Indeed, religious and spiritual beliefs are an important, but not exclusive, vehicle for these experiences, which are intimately embedded in the patient's psychological coping style. All of them should be sympathetically and respectfully included in therapy as they are often the key to a sense of emotional healing. Table 25.11-22 suggests ways of helping patients recast their experiences.

Futility → discover a sense of meaning/purpose
Disappointment → find sources of satisfaction in past and current events
Remorse → carry out remediation, taking care of unfinished business

Adapted from N. Cherny.

Table 25.11-22 Common Dimensions of Spiritual Suffering

Ethical Issues

Psychiatry-Ethics Interface In today's contentious social climate, ethical and legal dilemmas multiply with increasing stress, often resulting in psychiatric consultation. Often, reactive psychopathology has developed and must be assessed and treated. If inappropriate psychiatric labeling has occurred, the psychiatrist must remove the label, redefine the situation, bring out the underlying ethical issues, and set a decision-making process in motion. To be effective in these situations, psychiatrists must be clear about the difference between psychiatric issues on the one hand and ethical and legal ones on the other. They must diagnose and act decisively in the former, be restrained and humble in the latter, and educate others about the distinction. If they understand the ethical issues, have a precise knowledge of the legal constraints operating in the case, and call for appropriate consultation, their interviewing skills and experience in conflict reduction can make them valuable participants. A few major ethical issues in the care of cancer patients are discussed below, with emphasis on their psychological correlates.

Truth-Telling At the extremes, cultures that put a high value on autonomy consider the need for truth telling as self-evident, while cultures in which the patient is tightly embedded in a wider family unit consider truth-telling an exercise in maleficence. The spreading breakdown in close kinship ties and the steady improvement of modern cancer treatments are combining to increase truth-telling in growing segments of the world. However, local cultural deviance continues to arouse strong objections to it.

Informed Consent In the United States, informed consent is legally required for both conventional and experimental treatment. Given the unique toxicity and noncurative nature of much cancer treatment, this does not seem unreasonable, and it is spreading to other Western countries. However, it comes at some psychological cost. Anxiety reactions and occasional psychiatric decompensations are seen when patients feel overburdened by demands to make decisions. These often occur in predisposed individuals, whose premorbid adjustment was nevertheless good. Patients respond to active interventions, including support, medication, behavioral techniques, and active environmental manipulation of both home and treatment setting. These interventions usually suffice to help patients make decisions on their own behalf and undergo treatment, but posttreatment relapses may occur and are resistant to therapy.

Terminal Care Decisions Modern society is unequipped to cope with the life-and-death decisions spawned by technology. During the buoyant 1950s and 1960s, public enthusiasm about cardiopulmonary resuscitation was supported by the medical profession. It was quickly endowed with exaggerated, quasi-magical power and became a ritualized right rather than an optional medical treatment. That practice played into the paternalistic therapeutic activism characteristic of many physicians. The 1970s and early 1980s saw a countermovement. First the right to refuse treatment was established, thanks in large part to synergy between the consumer movement and the bioethics movement with its emphasis on patient autonomy. Next, the legality of do-not-resuscitate orders and the moral equivalence of stopping and not starting treatment were established. The medical profession is less enthusiastic than the public, perhaps because practitioners are familiar with the emotional ambiguities that inevitably surround death and must repeatedly experience them. Mechanisms are now in place which encourage competent patients to state their wishes concerning the terminal care period through the use of health proxies and advanced directives. Such mechanisms culminated in the 1991 Federal Patient Self-Determination Act (PSDA), but their use continues to be limited to about 20 percent of surveyed populations, unless very pointed, labor-intensive routines are put in place to encourage patients to prepare them. Such programs are under intense investigation, along with attempts to understand patients and surrogate decision-making processes.

At the bedside patients and their families vacillate between a suspicious, self-protective stance and a less conscious but strongly motivated, needy, and dependent one. These behaviors may appear erratic and paradoxical to medical staff who do not understand the emotional agendas being expressed. Caregivers are not immune to defensiveness and confusion as they try to rise above the doctor-bashing which has become commonplace. Federal mandates and regulations cannot resolve the emotional ambiguities of this difficult time.

With incompetent patients who have left no directive or who were never competent, active legal and ethical debates are continuing. Legally sanctioned behaviors vary with the jurisdiction. Psychiatric consultants must be very clear about these issues in their area. The local branch of one's professional society usually has an ethics committee, which can be consulted.

Euthanasia and Physician-Assisted Suicide Cancer patients have been central to the euthanasia and assisted-suicide debate. The legal rules governing physician behavior are unclear and constantly changing under the impact of precedent-setting court decisions and new legislation. Physician-assisted suicide had been legal in Australia's Northern Territory until it was overruled by the Federal Parliament. It has been accepted in the Netherlands, where it is regulated without being formally legalized, an ambiguous state of affairs that is unworkable in larger heterogeneous societies. Furthermore, longitudinal studies have pointed out a rising incidence of euthanasia as opposed to assisted suicide, in the face of a constant total number of assisted deaths, leading some to ask whether the Dutch are on a slippery slope.

In the United States, euthanasia remains unacceptable, but physician-assisted suicide is garnering increasing support. Two federal circuit courts of appeals have supported it. One decriminalized it on the basis of equal protection under the law, equating it to the available right to discontinue life-preserving treatment. The other legalized it on the basis of liberty and privacy rights. Both decisions were halted by injunctions and were reviewed by the United States Supreme Court. In June 1997 the Court denied a constitutional right to assisted suicide and opened the door to a state-by-state process of continuing experimentation. Oregon voters promptly voted in a law permitting physician-assisted suicide under carefully defined conditions. At the end of one year, the right had been exercised by eight patients. For some it was eight too many, for others it was few enough to prove that the law would not be abused. Meanwhile, the practical and emotional dilemmas will only multiply as regulations, laws, and initiatives proliferate. Some commentators pointed out that the Court had implied a right to active, effective pain management and palliative care. It has not yet been tested in the courts, but doubtless will be in the future.

Since 1982 a presidential commission, supported by major Catholic and Protestant authorities has affirmed the moral acceptability of withholding extraordinary measures and of stopping them later. They also support “the double effect,” namely, the unintentional possible hastening of death in the process of improving comfort. Intent is the important variable, meaning that one cannot deliver a large bolus of morphine that renders an alert patient deeply comatose or dead, but it is appropriate and humane to titrate morphine doses upward at a rate that maintains good pain relief regardless of potential respiratory impairment. In the habituated patient without primary respiratory failure, pain relief from opioids is usually achieved before respiratory depression occurs, making terminal struggle over opioids an unnecessary torment.

Despite this, the reasoning that surrounds the double effect remains controversial, with one faction using it to justify the legalization of assisted suicide and another calling it euthanasia or even murder. The moral situation is even less clear when non–terminally ill patients ask for help in dying. The physician who actually performs the act is practicing active euthanasia, which is against the law and against the medical code of ethics. But the role of the physicians who provide information on how to die or knowingly allow patients to hoard prescriptions is also under intense debate. Derek Humphrys' suicide handbook *Final Exit*, was on the best seller list for several months, underlining public hunger for knowledge in this area and making moot the issue of the medical role in providing information to many.

Proponents of legalization emphasize dignity and autonomy, pointing to Dutch studies that name this as the most frequent reason for help-in-dying requests. Opponents point to the lack of good pain management and other palliative care, citing American and Canadian studies that found depression and poor symptom control to be the major factors underlying patient suicidality. Given severe retrenchment of health care expenditures, with politicians, insurance companies, and managed care corporations calling the tune, there is cause for concern even among individuals who have no philosophical or moral objection to assisted suicide. African-Americans are far more opposed to legalizing assisted suicide than whites.

This controversy is emblematic of the moral and emotional malaise of the postmodern era and surfaces in the practice of psychiatrists, whether or not they are directly involved with dying patients. They must explore their own feelings to prevent their confusion from warping the treatment. Whatever personal conclusion they reach and whatever the outcome of the debate, psychiatrists should always keep the following fact uppermost in their mind and remind their medical colleagues, that a dying human being whose physical, emotional, and spiritual needs are being effectively attended to, seldom exercises the option to commit suicide.

Investigational Treatment Since the Nuremberg trials, and the Helsinki declaration, the moral requirement for informed consent from research subjects has been clearly spelled out and buttressed by regulations and institutions in the United States. Critics feel it is honored in letter more than in spirit, but competent patients systematically suppress information given to them, to create the illusion of an old-fashioned paternalistic relationship; hence the situation remains ambiguous despite the efforts of lawyers and bioethicists to rationalize it.

Critics of clinical trials claim that randomization is incompatible with the doctor's basic responsibility to the patient. Some patients find randomization intolerable and refuse to participate or later demand to break the code. Most patients find trials with a placebo-versus-treatment design emotionally more difficult to understand and tolerate than one with two treatment arms. Overall, patient's reactions are defined less by the protocol than by their relationship to the physician. Many patients seek experimental treatment with desperate intensity, peddling influence to enter limited protocols or pleading to use untested agents that have not shown enough efficacy to warrant clinical trials. Other patients participate in clinical trials out of altruistic motives. Alternatives to randomized clinical trials have been suggested, but they are

unwieldy and less powerful than the trials. Any solution requires well-designed trials (Table 25.11-23).

A sound clinical trial must
-Ask a useful question
-Have the power to answer it
-Have an objective review of the methodological and ethical issues
-Have a condition of equipoise (i.e., no treatment arm is known to be inferior to any of the others.)
-Have a mechanism for premature closure

Table 25.11-23 Characteristics of an Ethically Sound Clinical Trial

The role of the commercialization of scientific knowledge and of the media in promoting premature, exaggerated claims contribute heavily to consumer anguish as hopes rise with the first headlines, only to founder in subsequent months as more thoughtful voices make themselves heard or results fail to be reproduced. Savvy patients often outwit themselves as they learn about numerous protocols in different academic and commercial institutions and agonize about whether they are missing the very one that would save them.

Surrogate Decision Making Current legal doctrine specifies that surrogates must use substituted judgment (i.e., do what the patient would have wanted). Family members are presumed best equipped to carry out this charge, by virtue of their close association with the patient. However, experienced clinicians know that family members often respond to their own needs, once patients can no longer express their own, but they usually do not confront this issue unless surrogates are incompetent or clearly acting against the patient's interest. Psychiatry often plays a major role in furthering negotiations, developing compromises, and arranging treatment for the surrogate. However, these situations are often insoluble because existing laws make no provisions for incompetence in surrogates except via adversarial proceedings, which are painful and often destructive. The psychiatrist can still play a useful role in helping caregivers cope with their frustration.

Consent in Pediatrics All pediatric consents are third-party consents that carry the possibility of conflicts between parent and child. The parent may make a decision that is not in the interests of the child, as in the refusal of lifesaving treatment. The courts have tended to overrule parental authority in such cases. Parent and child may disagree, especially adolescents, although disagreements occur with school-age children as well, presenting staff with painful problems. Psychiatric intervention may help resolve the conflict, if the consultant does not impose personal beliefs, but the pain of watching a child caught between his need to please his parents and his right to the best available treatment does not always have a satisfactory solution. Legal recourse is used only after all attempts at conflict resolution have been unsuccessful.

Resource Allocation and Managed Care A basic tenet of the medical profession is that the physician attends solely to the patient's welfare, excluding other considerations. That tenet has not formally changed, but in fact, physician autonomy has been eroded by economic and political interests, and medical decision making is now seriously constrained by third parties with a profit-making agenda. Forced early discharges, limitations on expensive treatments, and the notorious gag rule are the result. The well-publicized refusal of managed care companies to pay for bone marrow transplants may represent only the tip of the iceberg.

Sexuality Cancer patients hide their sexual concerns because they are embarrassed and recognize that their doctor may be embarrassed; but they are usually eager for information, which is seldom provided. Formal observations on sexual dysfunction related to cancer type are sparse for women (breast cancer excepted) and men and inconsistent for several reasons. First, early studies did not stratify well by diagnosis, stage, or treatment. Except for obvious sexual ablations, dysfunction relates to type of treatment as much as to site. For example, testicular cancer with lymph node dissection has a high incidence of dysfunction. Absent that dissection, patients should have normal erection and orgasm—despite loss of ejaculation—but in fact, about 20 percent of men experience sexual performance problems on a psychosexual basis. The sexual impairments associated with prostate, bladder, and rectal cancers matched the amount of postsurgical sympathetic nerve damage rather than gross anatomical change. Addition of hormonal therapy or surgical castration to prostate cancer regimens results in near total cessation of activity, regardless of anatomy. Any course of radiation or chemotherapy can result in some sexual dysfunction, with radiation causing a high rate of dyspareunia. Patients with nondisfiguring cancers like leukemias and lymphomas have an unexpectedly large and persistent decrease of psychosexual functioning.

Second, evidence continues to accumulate showing that the incidence and severity of sexual dysfunction increase over time and that spontaneous remissions or stabilization does not occur, even in the presence of stable or improved physical states. Early studies had variable and usually short follow-ups, accounting for inconsistencies. Dysfunction is greater in younger patients across all diagnoses and is particularly high for cancers of breast, head, and neck and other physically disfiguring cancers. Significant psychiatric morbidity ensues and is only beginning to be studied.

Finally, one problem may be artifactual, because many texts are written by nonpsychiatric physicians who focus on the anatomically and physiologically possible, rather than on psychological factors. Patients report two- to threefold drops in level of sexual activity after treatment, for no clear physical reason.

Current research design is likely to define dysfunction more carefully, select the study population by both cancer diagnosis and treatment regimen, include partner assessment, extend follow-up much longer, and include interventions in the study design.

In an ideal world sexual concerns should be routinely addressed before treatment, with a baseline assessment that models open communication from the start. Patients need detailed information about what problems to expect and their reversibility. The information should be combined with reassurance that some form of sexual expression is always possible. If there is a spouse or sexual partner, early joint involvement is beneficial.

After the initial treatment period, patients should again be queried about sexual concerns. Follow-up clinic appointments are a good time to advertise workshops, support groups, and information sessions. A flexible format, with a mix of information, emotional ventilation, and focused problem-solving has been well received by several types of cancer patients. Emotional concerns include lowered self-esteem, negative body image, fear of rejection, anxiety, and depression and frequently create self-perpetuating cycles. Severe reactions and antecedent problems call for treatment. Recent studies reinforce the observation that psychological and relational factors play a large (often the largest) role in determining the degree of posttreatment recovery of sexual functioning. Given the neglect of sexual issues characteristic of much medical treatment, psychiatrists and other mental health professionals will continue to be the first to hear about them. They need to know the basic anatomy and physiological potential of a patient's condition and should refer the patient for further workup and treatment if needed (it often is). Psychological treatment includes couple psychotherapy and sex therapy. Sexual rehabilitation involves an adaptation of sex therapy principles and can be very helpful. It is based on the desire-arousal-orgasm model of sexual response and requires a phase-specific diagnosis. Treatment of late-phase dysfunction is quite effective if there is no anatomical interdiction. Early-phase dysfunction is more difficult to treat and benefits from a specialized referral. Table 25.11-24 summarizes the main treatment issues.

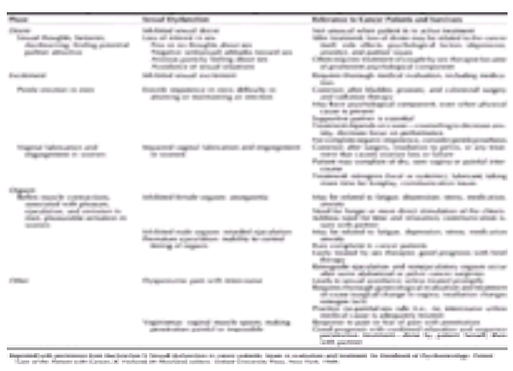

Table 25.11-24 Application of the Triphasic Model to Assessment of Sexual Dysfunction in Cancer Patients

Cancer Survivors At least 8 million Americans are cancer survivors. Many of them can be presumed cured, but they will never be the same. Their existence is a tribute to the human intellect; their resiliency, a tribute to the human spirit. But these partial victories should not blind one to subjective distress and residual problems in this traumatized population.

Most survivors are adolescents and young adults cured of childhood cancers and leukemia or lymphomas. Breast cancer survivors are another large group, but colon, cervix, head and neck, testicular, and thyroid cancers also figure among those in which early diagnosis and good treatment can result in cure.

In early studies, survivors were believed to be free of psychological residue because many of them used denial and emphasized positive beliefs such as having learned deeper values and being grateful for life. Recent studies of leukemia/lymphoma survivors and breast cancer patients, using the Brief Psychiatric Symptom Inventory, have shown a level of psychiatric symptomatology one standard deviation above the norm, or halfway between normal controls and psychiatric patients. Psychosexual functioning has been found to be notably impaired. Many survivors have posttraumatic stress symptoms that overlap with, but are not identical to, patients with persistent conditioned nausea and vomiting.

Becoming a survivor as opposed to being a cancer patient is both a real and symbolic transition of great complexity. Some of the coping techniques that enabled the survivor to be a good patient (e.g., the ability to regress, to be more passive, and to be dependent) must now be jettisoned, and a new mantle of action must be assumed. The survivor's environment may be out of step. Families may continue to require an untoward degree of dependency or may demand an instant transition to a preillness level of behavior.

With more-visible disfigurement and greater disability comes a greater impact on intrapsychic and interpersonal adjustment. When the face, voice, breasts, or sexual organs are affected or when the patient must use an appliance or prosthesis, a substantial psychological burden overlies the physical impairment. But even with no obvious deformity, survivors feel damaged, vulnerable, and sensitive to any abnormal signs or symptoms in their bodies. Their awareness of death seems to last indefinitely and is comparable to the altered self and world view seen in trauma victims. Thus, survivors' increased somatic complaints have both psychological and physiological determinants.

Delayed Consequences of Treatment Four categories have major psychological impact.

1. Infertility problems, premature ovarian failure, and interference with sexual response are particularly distressing if patients have not been educated and prepared to face them. But evidence indicates persistent patient denial even in the face of preparation, such as sperm banking. Intervention should include endocrinological, gynecological or urological, and medical evaluation and treatment. Psychotherapy, couple therapy, group therapy, veteran patient programs, and self-help programs can all be used to good effect.
2. Pituitary, thyroid, and other endocrine failures occur as a consequence of radiation therapy. Patients with these problems often present with psychiatric symptoms. Patients treated before or during puberty suffer from short stature due to lack of growth hormone. Chronic fatigue is a common symptom that calls for patient and family education, since it is often given an unjustified psychological label.
3. Survivors have increased vulnerability to late organ failure and secondary malignancies and to the chronic anxiety that this vulnerability generates. If such a complication comes to pass, it finds the survivor less able to cope and more prone to a posttraumatic response.
4. In a frightened, superstitious world, survivors face social rejection, employment discrimination, and serious insurance problems. Survivors well enough to return to work plateau in their careers because they are passed over for advancement by employers who fear a recurrence or because they are averse to moving due to having become uninsurable.

Predictors of poor coping include previous psychiatric problems, lack of social supports, and stage of disease at diagnosis. In treating cancer survivors psychotherapeutically, one should strive for a detailed understanding of the patient's experience. When the content is anxiety provoking to the therapist, it is sometimes overlooked, leading to poorer outcomes even in the hands of seasoned therapists. Full knowledge is required to understand when the trauma response is primary and when it is a shield for deeper psychodynamic issues. Lingering too sympathetically on the trauma may deprive the patient of the opportunity to grow beyond it. Yet, failing to be sufficiently empathic in understanding its impact is obviously unproductive. Many survivor programs offer workshops, lectures, and groups for patients, couples, and families. The National Coalition of Cancer Survivors (NCCS) is emblematic of the possibilities for self-awareness and psychological and political empowerment.

PSYCHOLOGICAL PROBLEMS OF FAMILIES

Research on the family of the cancer patient is an active and exciting field, especially because of the increasing integration of family therapy concepts. Some findings are disease specific, but many can serve as a paradigm for understanding the impact of any major illness on the family.

Psychological Reciprocity Awareness of the powerful reciprocity of emotional states and behaviors between cancer patient and family members is well established from numerous studies of parents, spouses, and siblings that document levels of distress and functional impairment that can rival or even exceed that of the patient. More disturbingly, primary caregivers have been noted to worsen over time in the face of stabilization or improvement in the patient. Finally, studies of families after bereavement or cancer survival show a significant incidence of impaired functioning over extended periods of time and, again, a worsening over time in the distressed families. These findings have some urgency, since most families will experience cancer at some time and it will probably be a drawn-out, chronic ordeal. High-risk families can be identified early and significantly helped by professional intervention.

Family Evaluation Families can be evaluated from several points of view. The most traditional one is to examine how well the family is taking care of the patient. Medical staff are quite sensitive to this and will often call for help when families fail in this area. Another approach is to scan for psychopathology in family members. If the family member is acutely symptomatic and is visible to the treatment team, a consultation will follow, but absent members will go untreated.

Categories of family support include

1. Providing emotional support and containment. Family members are expected, and expect themselves, to contain their own reactions and support the patient.
2. Sharing responsibility for decision making. Family members often carry the lion's share in researching and selecting options, with or without patient approval. This goes hand in hand with an enormous and universally reported need for information.
3. Concrete caretaking. In today's budget-conscious climate and vanishingly short hospital stays, an ever-increasing amount of care is being thrown onto the family. This is not acknowledged, hence families receive little help or training.
4. Meeting financial costs. Even under socialized medicine, families absorb massive expenses and suffer significant downward mobility.
5. Absorbing social costs. In one study, 25 to 30 percent of primary caregivers lost their jobs, and more than half of the remaining moved to lower-paying jobs to accommodate their need for flexibility.
6. Maintaining stability. The patient's role and contribution must be replaced, a difficult task in the case of young mothers or breadwinners. The new demands must be folded into ongoing functioning. Some members develop increased emotional needs that must be met. And all the standard family functions, emotional nurturing, feeding, clothing, sheltering, educating, and socializing, must continue unabated.
7. Adapting to change. It has been said change may be the only constant during a cancer illness. But it is useful to divide events into acute and chronic phases. An acute event is met with effective mobilization of family members, but psychologically, they may be reacting quite asynchronously, with the denial of some and the pessimism of others causing significant friction. Members may protect themselves, under the guise of protecting each other, by colluding in a conspiracy of silence that can rigidify and be hard to change later. There is a universal need for information, to decrease uncertainty, to help the family problem solve effectively, and to ensure better communication, cohesion, and emotional containment. In the chronic phases the family must return to developmental and maintenance tasks, which may expose conflicts between the patient's continuing needs and the needs of others. Initial mobilization is replaced with a paradoxical increase in psychological symptoms. Families isolate themselves, even when outside support is available, and may put major decisions on hold.

Family-Centered Approaches A new and powerful method of evaluation uses a family-centered approach to examine how the family sustains itself and copes with

the illness, while still meeting the needs of all its members. Four concepts are important for understanding the impact of the illness and devising interventions that will not be instantly neutralized for the sake of family homeostasis.

1. Developmental stages: Families go through developmental stages with specific tasks at each stage. Cancer disrupts the tasks at any point, sometimes permanently if no help is offered. Any intervention must be compatible with helping the family carry out its phase-specific tasks.
2. Family structure: One must understand how a family is organized, to understand the caregiving dynamics and intervene in ways that will reinforce the family rather than threaten and destabilize it.
3. Family history, beliefs, and myths: Every family is permeated by stories that dictate unconscious responses to illness, dependency, loss, and other fundamental life experiences. They underlie behaviors that seem idiosyncratic or bizarre to untrained observers and must be factored into treatment. Explaining them to the treatment team minimizes scapegoating and polarization.
4. Family style of interaction: Good communication among members is the best predictor of good outcomes, with family cohesion a close second. Theoretically too much cohesion should be deleterious to individual development, but most studies have found that during a serious illness, high cohesion remains a positive factor.

Nontraditional Families Divorced or blended families, single-parent families, homosexual families, and families that are ethnically, culturally, or religiously discrepant from the ambient medical culture experience many of the problems described above but also have specific characteristics that are being more systematically recognized.

Staff-Family Interface During a life-threatening illness such as cancer, the medical staff come to play an enormous role in the family dynamics, invading the family system to create a new ad hoc hybrid system. That hybrid is the milieu in which consultation requests are generated. It must be understood and used for any intervention to be effective.

Guidelines for Intervention While most families show generosity, courage, and resiliency, as many as 25 percent might need intervention, and more might well benefit. Of those, a much smaller number require formal family therapy, which must be done with a realistic understanding of the medical facts and the medical milieu. Most families manage with short-term crisis interventions designed with an accurate understanding of the family factors described above and a commitment to the family as a whole, not just selected members. The staff-family conference deserves special mention. It can be a powerful mutative event that resolves what had seemed very knotty problems.

PSYCHOLOGICAL ISSUES FOR STAFF

Anatomy of Stress The stress on caregivers has been documented in studies across many settings and many countries. Some stress factors are comparable to those found in other high-technology, high-morbidity areas of medical care outlined in Table 25.11-25. Of recent note are the massive downsizing and cost-ratcheting sweeping medical institutions as they are being turned over to commercial entities, creating severe individual hardships and serious ethical problems in the delivery of care. Communication problems between doctors and patients lead not only to patient dissatisfaction, but also to lower job satisfaction and self-esteem. The following stressors are more particular to the oncology setting: the negative aura that still surrounds it, even in the AIDS era; the toxic nature of treatment and its inevitable adverse effects; the occupational hazards surrounding chemotherapeutic and radioactive agents, and the resistant infections to which immunocompromised patients are prone; the ambiguity of many treatment decisions, which often breeds intrastaff conflict and hostility; the unsettled social and ethical climate that surrounds palliative care and terminal care decisions today; and the extreme emotional reactions engendered by cancer and projected onto staff in the form of overvaluation, ingratiation, and covert bargaining, followed by disillusionment and anger.

High morbidity, high mortality
Complex technology used under high pressure
High frequency of life-and-death decisions
Terminal-care issues
Third-party conflicts
Interstaff conflicts
Limelight medicine (medical care administered under the impact of unusual publicity or public interest)
Response to severe debilitation and disfigurement
Response to difficult patients (excessive dependence, anger, uncooperativeness)
Response to suicidal ideation
Issue of inflicting pain as part of treatment

Table 25.11-25 Staff Stresses Common to High-Stress Medical Settings

One must add the staff's own reactions to cancer. At a conscious level, caregivers depend on their intellectual grasp of the situation and make a point of rejecting myth and superstition. They are proud of their ability to function under pressure with reliable emotional controls and to offer something in all situations, even hopeless ones. Yet, at an unconscious level, covert adaptations exist that are idiosyncratic and irrational, often determined by historical precedent, and resistant to change. They include superstitious beliefs, stubborn fears, and feelings. Their eruption into awareness makes the worker feel embarrassed and inadequate. Hence the feelings are kept secret, and the caregivers believe they alone have them. Thus, staff workers isolate themselves from one another. Outside the workplace, staff isolation may be reinforced by family, friends, and a society that avoids reminders of a frightening topic. Caregivers are not singled out personally but are discouraged from talking about their work and risk being shunned if they do not accept this taboo.

Clinical Spectrum of Staff Responses

Normal Adaptation and Coping Staff undergo a developmental process when they begin to work intensively with cancer patients. A very high initial level of dysphoria, anxiety, sadness, and numbing recedes over the first few months, displaced by the need to demonstrate competence and ability to survive. Over the next few months, existential issues are allowed to surface, and staff members shape a deeper adaptation that engages their whole personality. Formal support is not necessary, but informal support and peer-group interactions are important for successful resolution.

Adjustment Reaction Adjustment reactions may occur during the formative stages of providing cancer care, but they also overtake staff members at any point in their careers when a personal loss resonates with their work (cancer in self or loved ones in particular). These reactions are acute but time limited and respond well to support or a brief change of venue. Rarely is psychiatric intervention needed. Sometimes adjustment does not improve with time, exposing an underlying psychiatric disorder, often preexisting and precipitated by work stresses. Psychiatric evaluation is indicated and may result in transfer to a less emotionally stressful situation.

Supportive Interventions for Cancer Staff Staff support interventions generally fall into three categories: organizational, individual, and group. Organizational issues include leaders who model realistic expectations and monitor overinvolvement in themselves as well as others. A team approach encourages sharing difficult tasks, prevents a sense of being indispensable, promotes staff cohesion, and eases time off and vacations. Open communications and good orientation procedures are important. Individual responses include peer support, supervisor support, access to an employee assistance program, and other confidential psychological support or psychiatric evaluation. Positive reports about the benefits of staff support groups in many forms and many settings continue to appear despite the paucity of rigorous outcomes studies. The group must not be perceived as undermining unit norms or leadership. Structure must be provided as needed to alleviate staff anxiety, while still allowing unscheduled topics and feelings to be aired. Therapy issues must be avoided, and discussions should clearly flow from work-based topics. An active area of new and productive intervention is the development of group-based, workshop formats for improving communication. Many have been described and found helpful; more are under development.

CANCER IN THE PSYCHOTHERAPIST

Unlike degenerative diseases with their linear course and associated impairments, cancer may have an episodic course with no loss of cognitive functioning. It inevitably forces a blurring of psychological, and later overt, boundaries in therapy. An all-but-intolerable burden is placed on doctor and patient when therapists with

cancer continue treatment too long because of their own need to deny their fate. This suggests that difficult as it is, the therapist's most ethical course of action is to set a fixed termination date that allows patients to transfer to someone else while the therapist is still alive and well enough to negotiate the termination. This may be asking too much of an individual in whom denial is both the main coping mechanism and the only impairment. The help of colleagues is invaluable at this time.

CANCER IN CHILDREN

Magnitude of the Problem Cancer is second only to accidents as a cause of death in children. The overall survival rate is about 60 percent, but it may be higher for Wilms' tumors, Hodgkin's disease, and certain leukemias. These outcomes are bought at the cost of long, intensive treatments that put patient and family under severe stress.

Child Patient Sick children are benefitting from social and legal trends toward empowering them. They are being informed more honestly and their cooperation is being thoughtfully enlisted. Like adults, children know or intuit a great deal without being told and are relieved by open communication. Its provision is complicated by the need to be age appropriate and is more painful because of adult vulnerability to suffering in the young. As a result, adult denial often prevails, and children respond by increased behavioral symptoms or withdrawal.

Discussion About Diagnosis and Prognosis A child can hardly remain oblivious, with endless clinic visits, repeated chemotherapies, flagrant symptoms such as hair loss, repeated exposure to other patients, and statements of people around them. Studies note that perception outweighs reality, hence intrusive procedures can loom larger than prognosis, especially for younger children. The developmental stages of death awareness in healthy children are well described but are probably not applicable to children with cancer nor to other children who have been exposed to death in an untimely and personally threatening way that results in earlier and keener awareness. Thus, discussions are clearly indicated, but they must follow the child's lead, probing and responding to the child's sources of anxiety and remaining reassuring about care and comfort if they cannot be so about outcome. Children who received truthful information have been found to have less subsequent anxiety than those who did not.

Psychiatric Diagnosis and Treatment A syndrome-oriented approach to diagnosis is not useful with children. It is more important to know the child's previous developmental profile and to assess current symptomatology, arrests, or regressions. But in a few situations diagnosis is important. Preexisting disorders are aggravated by cancer and must be monitored and treated throughout. The stress of cancer can make a mild disorder manifest for the first time. Anxiety, depression, and delirium occur routinely in children and respond well to treatment. Anxiety is ubiquitous. Behavioral interventions are the first line of treatment, but some situations may require the use of anxiolytics. Benzodiazepines are used but can cause a disinhibition syndrome that may be mistaken for more anxiety. The use of sublingual alprazolam (Xanax) allows a rapid effect in children who cannot swallow. Mild depressive symptoms are common and rarely progress to severe, retarded depression with vegetative symptoms that respond to low doses of antidepressant medication. Desipramine is contraindicated because of several reports of sudden death. Delirium is common in very ill children and is often mistaken for psychosis, which it can resemble. Antipsychotics are helpful. Pediatric psycho-oncologists should be quicker to use drugs than child psychiatrists, but they should start with small doses, watch closely for adverse effects, and try to taper as quickly as possible. Further studies will be welcome.

Pain Management Children are even more undermedicated than adults, and some caregivers still doubt that neonates and infants feel pain. Therefore, as-needed dosing in children always results in serious undermedication. Very young children cannot make their complaints known clearly. Older children often make complaints in the context of a normal regression, which is then used to disqualify the complaints. Exhausted, critically ill children, given relief from suffering, may enjoy opioids in a way that upsets some parents and caretakers.

Management of Stressful and Painful Procedures Recent work on posttraumatic stress disorder in children reinforces the principle that aggressive prophylactic relief during stressful procedures is important not only for current comfort, but also for long-term adaptation. Poorly managed painful procedures not only brutalize the child, but also result in anticipatory fear of pain, loss of trust and cooperativeness, phobias, and posttraumatic symptoms that do not improve over time. Multimodal behavioral methods for use by both parents and staff are widely and effectively used. Painful medical and or surgical procedures should always be eased by the use of pharmacotherapy up to and including general anesthesia.

Family Involvement The families of pediatric patients are even more vulnerable than those of adults because of the child's total dependency and the biological, emotional, and social factors that make families uniquely responsible. No intervention has any hope of success if it does not address the family as much as the child. In one study, mothers had a greater rate of posttraumatic symptoms than their children.

Staff Problems The human species survives in part because the helpless young elicit nurturing responses from adults. This principle is strongly in evidence on pediatric oncology units. The staff's reward is a deep sense of engagement and meaning; the price they pay is relentless pain held at bay by a variety of mechanisms. The maintenance of staff morale, the prevention of dysfunctional norms, the support of faltering members, and the defusing of unconscious competitive feelings toward parents are important ways of contributing to the well-being of all participants.

Team Approach The multidisciplinary approach to patient care is most widely implemented in pediatric cancer patients than in adults. Child psychiatrists can play a useful role and, where the model is not in place, are in good position to help organize it. Local teachers and clergy should be involved, and community resources should be used to help families reintegrate when they return home.

PATIENTS AT GENETIC RISK FOR CANCER

The advent of DNA technology and sophisticated epidemiological studies have permitted identification of many healthy individuals at high genetic risk for developing cancer. This work is at the forefront of basic research and is paradigmatic of the ambiguous interplay between basic research, commercial interests, clinical practice, and social policy. More and more cancers are being found to have a genetic component (Table 25.11-26). Among the best studied are breast cancers, of which about 5 percent are hereditary and associated with an increased incidence of ovarian cancer, 25 percent are familial, and the rest are assumed to be sporadic. Hereditary forms have also been identified for colon cancer, retinoblastoma, Wilms' tumors, and melanoma. The Li-Fraumeni syndrome identifies families in which a mutation of the p53 gene on chromosome 17 is associated with the incidence of several early-onset tumors, particularly breast and soft tissue sarcomas.

Bilateral acoustic neuroma
Bladder cancer
Breast cancer
Colorectal cancer
Familial renal cell carcinoma
Gastric carcinoma
Multiple endocrine neoplasia (type 2A)
Neuroblastoma
Ovarian cancer
Small cell lung cancer
Uveal melanoma
Wilms' tumor

Adapted from Bodmer WF: Cancer genetics and the human genome. Hosp Pract 26(7), 1991.

Table 25.11-26 Selected Human Cancers Associated With Recessive Chromosomal Abnormalities

High-risk individuals have chronic anxiety, lowered self-esteem, and problems concerning marriage and child-bearing. Their self-estimates of risk are usually much higher than their real risk, even for those in the highest risk bracket. Some show no increased psychological distress when found to be carriers, perhaps because, like human immunodeficiency virus (HIV)-infected individuals, their anticipation amounts to near certainty. Previously unsuspecting carriers do show increased distress, while those who decline testing remain anxious throughout. High risk and high anxiety often go hand in hand with denial and poor health maintenance behaviors such as avoiding breast self-examination and surveillance testing. A group therapy intervention with educational and expressive components has been effective in reducing psychological distress, normalizing risk estimation, and improving health behaviors. Individual educational and supportive interventions have shown promise. Some form of psychological intervention should be a regular part of the surveillance programs that have been developed in many centers. Similarly, psychological support is

crucial to deciding whether to enter surveillance trials, have prophylactic mastectomy, and above all whether or not to have testing. One study has shown that while only 71 percent of high-risk women would undergo testing for breast cancer, as many as 42 percent of average-risk women wished to undergo it as well. On the other hand, intention is not always translated into practice for the high-risk group. Testing is rapidly moving from an experimental procedure offered in a research context to a commercial enterprise. In response to this, the field of genetic counseling is undergoing exponential growth. But it does not include long-term support, other than occasional booster sessions; hence, mental health professionals will continue to provide the help needed to cope with later psychological repercussions.

The consequences of testing go far beyond the psychological distress of the individual. Cooperation of unwilling family members may be required, and unwelcome knowledge forced upon them. Marital, sibling, and parent-child relations may suffer, and whole families may be exposed to insurance and job discrimination in today's world of thoroughly eroded privacy. The individuals involved must not be given false reassurances by eager researchers or naive mental health professionals, but well-informed counselors can provide essential information and problem-solving techniques. The broader social issues are in other hands, in recognition of which, 5 percent of the budget for the human genome project is going to the study of the ethical, legal, and social implications of genetic knowledge, through the so-called ELSI office.

CANCER RISK AND SURVIVAL

The role of psychosocial factors in altering cancer risk and survival fascinates the public, but the discussion is often blighted by wishful thinking and genuine confusion. Cancer risk and survival are often conflated, and great emphasis is placed on the direct role of the individual's psychoneuroimmune mechanisms in preventing or curing cancer, while ignoring many psychologically mediated effects that are far stronger and better established. The common denominator behind all these distortions is the deeply grounded wish to believe that all human beings have within themselves the power to prevent or cure cancer. A well-designed study in which cancer mortality in a group of Chinese women showed a significant drop during a holiday honoring older women, followed by a compensatory peak after the holiday, does lend some encouragement to these views, but there is enough distortion and overinterpretation abroad that psychiatrists must know the current state-of-the-art to clarify their own thinking and better understand their patients.

Cancer Biology The natural history of a given cancer begins when changes in a cell's DNA establish the potential for this cell to become malignant. This cancer initiation precedes by many years (anywhere from 3 to 30) the second event, known as cancer promotion, in which growth suppression is lost and malignant multiplication begins. This again precedes by months to years the appearance of clinical symptoms. Finally, there is a greatly variable period between the appearance of any clinical symptoms and the seeking of care. Cancer risk can relate to the risk of initiation or the risk of promotion. Cancer survival is measured from time of diagnosis, but this ignores the important factor of delayed diagnosis for which initial disease staging is a partial but incomplete solution. With new knowledge have come opportunities to explore the role of psychosocial factors that propel a given cancer through all stages of biological development.

Psychosocial Factors One attempt at clarifying the multifaceted impact of psychosocial factors on cancer is to think in terms of the external and internal loops. The *external loop* refers to ways in which psychological, behavioral, and socioenvironmental factors affect the person's behavior and lifestyle and hence cancer risk. Even there, the actual behaviors must be teased out from their psychological determinants. The *internal loop* refers to the ways in which the person's psychological status affects cancer risk and survival via changes in the internal milieu. This effect can be mediated in many ways, such as neurological, immune, and endocrinologic mechanisms individually and in varying relations to each other.

Psychologically Mediated Environmental Factors By far the largest role in the development of cancer is played by environmental insults such as radiation, viruses, and chemicals acting cumulatively over time on each individual's predetermined complement of potentially cancer-causing genes. Exposure to injury in turn results from social and psychological factors over which the individual has variable control. By far the greatest contributors of psyche to cancer incidence are personal lifestyle and habits. Smoking cessation, abstinence from alcohol, more exercise, and better eating habits would probably decrease the incidence of cancer in the United States by 50 percent overall. The greatest contributor to decreased survival is poor use of screening, since early diagnosis is still the best predictor of prolonged survival. These facts are unequivocal, yet they leave the public, the media, and many policymakers remarkably passive in comparison to the excitement generated by other, far less significant and less proven psychological factors. Recent developments in which individuals and state governments may succeed in recovering medical costs from tobacco companies seemed to herald a sea change, although the settlements negotiated by the government and the tobacco companies have been seen by many as inadequate.

Methodological Considerations The ways of studying the role of psychosocial variables all involve large populations over long periods of time. The first is the case-control study, in which a group of patients is matched to a control group and differences between the two groups are identified. The index population is usually unrepresentative of broader populations, and the matching is incomplete; hence it is unlikely that the findings can be generalized to larger unselected populations unless the effects are large enough to be inescapable and easily verified. More reliable is the retrospective method, in which a cohort is opportunistically identified at a late point in time, and factors related to cancer are sought retrospectively. The index population is equally unrepresentative, but these are essential pilot studies pointing the way to later larger ones. The latter can be designed prospectively, with sophisticated methods of controlling for known confounding variables, but the task remains methodologically daunting. In semiprospective studies, patients are assessed in the prebiopsy stage and at various later times, to compare positive and negative biopsy groups. But, breast cancer patients correctly predicted their biopsy results more frequently than lung cancer patients, and all of them had greater than chance accuracy, casting doubt on the significance of pre- and postbiopsy self-reports. Even crisper facts such as negative life events can be of uncertain significance because individual reports vary with the subject's current mood and chronic disposition.

Current Findings All this notwithstanding, over 50 psychosocial factors have been claimed to play a role in cancer initiation or promotion. Some of them are surviving critical scrutiny, and the field continues to be deservedly active and exciting. Stress was initially thought to be significant, but later studies disproved this in most situations. However, stress plus passive coping styles may have a negative impact on the progression of virus-induced tumors such as cancer of the cervix, an effect that may be immunologically mediated. These account for about 5 percent of human cancers. No impact for bereavement has been demonstrated except for two studies that showed a small increase in diagnoses that washed out after 1 year, too short a time to support a role for bereavement in the development of cancer.

On the other hand, several recent studies have shown that social support is a significant positive survival factor, a fact that has important implications for mental health professionals. Social class has a variable effect on cancer risk, with some cancers labeled “cancers of poverty” and others “cancers of affluence.” But class has an unequivocal negative effect on cancer survival, probably because lower-class individuals are less knowledgeable about, and less able to avoid, toxic exposures and unhealthy lifestyles and lack access to good medical care. A crucial, hard-to-measure factor is the mix of mistrust, exhaustion, fatalism, and despair that leads the poor and the discriminated-against to underutilize even what little is available to them.

Race is also a strong predictor of decreased survival, but the effect decreases after correction for social class and shows great variability within racial groups. The effectiveness of interventions aimed at improving lifestyle and health behaviors has often been disappointing, usually because the complexity of the problem is underestimated, unrealistically short-term results are being demanded, or both. A few encouraging studies have been described: one that used an understanding of the role of peer pressure to affect adolescent smoking, and another that combined the provision of health care with the provision of food to the very poor. These interventions are cost-effective in the long run, but the up-front cost is high enough that they often lack reliable funding. Managed care institutions have claimed to put more emphasis on preventive care as a counterweight to their drastic curtailing of available medical care, but the population turnover is too great, their lack of open accountability is thus far notorious, and many are going out of business.

Internally Mediated Risk and Survival Factors Recent developments in psychoneuroendocrine-immunology have confirmed the feedback loop between feelings and perceptions on the one hand and neurological, endocrine, and immune changes on the other. In small part these update classic psychosomatic concepts. But more urgently, in this area of research recent developments lend theoretical support to current popular beliefs about “the power of the mind over the body.” A rereading of William James' 1902 book *The Varieties of Religious Experience* should disabuse any reader from thinking that these convictions represent anything new, either in their mood-enhancing effects or their potential for abuse and backlash. What is new is contemporary science's ability to explore the physiological underpinnings of these age-old and cherished human beliefs.

Personality Traits, Feeling States, and Coping Styles There are conflicting results on the role of personality in cancer prevalence and survival, beyond what is mediated through risk-avoidant behavior. Studies are handicapped by the absence of comprehensive prognostic indicator profiles. This becomes more urgent as more tumor markers such as prostate-specific antigen (PSA) come into use and as more prognostic tumor markers such as estrogen receptors in breast cancer are defined. These indicators may someday define subgroups of patients in whom the role of psychological factors may be very clear. A few centers have refined their research design over many years to allow a few conclusions. No strong relationship has been established between personality style and disease outcome. This may be due to the inadequacy of personality typologies as well as the difficulty of differentiating between state and trait when observing an individual in crisis. The well-known type A personality (driven and quick to anger) was studied mainly in reference to cardiovascular disease, although one study found a 1.5 times greater risk of cancer, likely related to smoking. Evidence for the role of its opposite, type C personality, is much flimsier than the enthusiasm with which it was received. Long-standing psychiatric

symptoms such as chronic neuroticism have not been proved to be related to disease outcome. Feeling states such as level of distress or anxiety have not been found to have any predictive value. An association between passive coping styles and poor disease outcomes was demonstrated in many studies using measures of helplessness and hopelessness, “fighting spirit,” stoic acceptance, and anxious preoccupation. The role of denial remains open and points again to the complexity of what is being measured. Affective denial in the face of an effective coping style may have the opposite impact from that of effective denial associated with a passive coping style.

Immune Mechanisms Now that measurements of immune function are available, numerous studies are showing small, but statistically significant, differences in bereavement, disease states, and many kinds of stress. In cancer, a conditioned negative immune response has been found with chemotherapy. An improved immune response was seen following a group intervention in melanoma patients, and differences have been found in the families of cancer patients depending on their coping effectiveness. More work needs to be done to establish the clinical significance of these preliminary findings observed in very small samples.

Endocrine Mechanisms The discovery that breast cancers with estrogen-sensitive receptors are associated with a better outcome is of clinical use when receptor typing is done on biopsy specimens to guide subsequent management. Since the receptors respond to emotional mediation through the hypothalamic-pituitary axis, a relationship to emotional states and psychological adjustment can be postulated, although studies have shown conflicting results.

Effect of Psychological Interventions The direct effect of psychological interventions on cancer risk through internally mediated mechanisms has not been studied and may never be fully disentangled from the large, well-known, externally mediated effects on lifestyle and health behaviors. However, studies of the effect on cancer survival have produced significant results. Over 20 intervention studies with patients in traditional treatment and 3 studies of alternative therapies have shown clear improvement in quality of life and no effect on mortality. One well-designed, widely known study was thought to show a significantly increased survival time (>18 months) in a small sample of women with metastatic cancer after a group intervention designed to address pain control, interpersonal relationships, and general illness-coping behaviors. The study is small, but large replications are under way. However, a subsequent paper studied the outcome figures for a large sample of women in the same geographic area and in the whole state and found that the control group did more poorly, while the intervention group reflected the larger norms. Several replication studies are under way. Another widely discussed intervention study demonstrated improved quality of life and improved immune parameters in early melanoma patients who had received a time-limited group therapy intervention with an educational and problem-solving emphasis. In a follow-up study, the intervention group had a longer relapse-free interval than the control group. Publication of a longer follow-up period is expected. It is not accidental that both of these studies involved a group intervention that provided a close social group and promoted more-active coping behavior, two of the factors already observed to have a positive effect on survival.

Discussion Given the unequivocal improvement in quality of life shown by psychosocial intervention studies, their continued use and further development is an obvious imperative. If they also turn out to affect disease outcome, the health care delivery system should be reexamined and adapted to maximize the effect. The extent to which their effect on quality of life has been undervalued by a medical establishment that prizes small improvements in survival so highly, speaks to fundamental differences between the medical and the mental health establishments and between doctors and the public.

On the other hand, premature acceptance of a major role for internally mediated psychological factors can have negative implications for patients. Patients feel anguish over their inability to have a positive attitude, and families conspire to worsen their guilt by demanding an upbeat stance from patients far too ill to do much except endure. The reaction of medical personnel is ambiguous. They reject a strong role for psychological causation, yet they are privately intrigued. Some clarification was suggested by a paper that differentiated a belief in control over the cause from control over the course of the cancer. The former was associated with anxious preoccupations about cancer, while the latter was associated with the beneficial fighting spirit. Another clinically useful distinction was made between weak and strong believers in psychological causation. Strong believers give psychological factors primary importance in determining their prognosis, equaling or superior to the role of conventional treatment. Weak believers view them as modifiers and do not experience as much anguish over their own inability to maintain a relentlessly positive attitude. Unfortunately, even seasoned psychiatrists find it difficult to move self-punitive patients from the strong to the weak category. Given the current state of the art, it would seem best to work actively to improve patients' quality of life, while at the same time encouraging them to enjoy these improvements for their own sake, independent of any effect on survival time. Group interventions are particularly desirable because they tap into the well-demonstrated protective effect of social supports, even as they address their other stated goals. If survival is improved, so much the better. If it is not, patients will be spared unnecessary guilt.

ALTERNATIVE TREATMENTS

The use of unproved methods of treatment is deeply entrenched in human history. The scientific developments described above have blurred the boundary between them and traditional treatment, and many have moved into the mainstream. The role of diet in cancer risk is now well respected. Programs that use meditation and lifestyle changes are being reimbursed by third-party payers who view them as effective, safe, and inexpensive adjuncts to traditional treatment. So-called integrative or complementary medicine clinics are being opened in a rapidly growing number of tradition medical centers. They often occur in the setting of an intense charismatic relationship, with a great emphasis on faith, support, and encouragement. Patients regard this experience alone as valuable.

The motives that drive providers range from mercenary ones to idealistic beliefs about the power of mind over body. The motives that drive patients include universal concerns, such as the often justified fear of failure of traditional treatment, the hope for a miraculous cure, and the desire to achieve mastery and control. The stigma of cancer propels people irresistibly into magical thinking. The toxicity of traditional treatments also makes patients so anxious that other options become compelling. The out-of-pocket expense of nontraditional treatment is in the billions. In one study, a large percentage of patients at a tertiary medical center was seeking alternative treatment, now called complementary medicine, while still in traditional treatment. They did so early in their course, took both simultaneously, dropping the traditional treatment before they dropped the alternative one. Educated and wealthy patients were best represented. The main treatments were physical ones to cleanse and strengthen the body; special diets to purify, fortify, and avoid harmful substances; immune therapies, faith healing; and psychological methods to improve mental attitudes and reinforce the body's defenses.

Many such treatments give patients a sense of purpose and control. Faith healing harnesses the great power of religious belief. The last category, improving attitudes, is most troubling to the psychiatrist, because it leads so readily into blaming the victim. Patients are told that they can influence their cancer by mental efforts. The unspoken corollary is that further disease will be due to inadequate efforts and, worse yet, that they caused the cancer in the first place. This plays directly on attribution mechanisms to which cancer patients are already universally prone. One can also see desperate family members grasping at these explanations and trapping the patient ever deeper in a web of guilt and responsibility as a means of coping with their own grief and frustration, or an angry patient blaming a family member for causing the cancer.

The psychiatrist must remember that many patients are secretly involved in alternative care, which makes them embarrassed and distant if they cannot share it. As time passes, if their condition worsens, the burden of failure and guilt makes a greater contribution to psychic distress. Hence clinicians must encourage disclosure and help patients extract something positive from the experience. This is difficult to do if patients and families retreat into anger at the establishment and reduce cognitive dissonance by clinging desperately to their alternate beliefs. However, this picture is in a place of rapid change.

SPIRITUAL BELIEFS AND HEALTH BEHAVIORS

The inclusion of a section on religious or spiritual problems in DSM-IV is but one sign of increasing awareness of the importance of this area to patients, families, and many staff as well. Several studies have shown that religious beliefs are often associated with mature and active coping methods, and the field of psychological and spiritual interfaces in the experiences of acutely ill patients is spawning a whole new area of psychological research within the traditional medical establishment. The psychiatric consultant should inquire about faith, its meaning, associated religious practices, and impact on the coping response. It can be a source of strength or guilt at all stages of the disease, ranging from the earliest “What did I do to cause this?” through “Will God give me only what I can carry?” to the poignant life review of the late stage. It is often a primary factor in the reaction to suicidality and in attitudes toward terminal care decisions. Mental health professionals should be prepared to deal with these areas in an unselfconscious, uncondescending manner, enabling patients to fully integrate this aspect of their personality into the current crisis. They should also work in harmony with the patient's spiritual guide, if one is available. Sometimes an experienced, effective chaplain working with the appropriate patient can achieve positive results more directly than any psychotherapy. Also, sometimes chaplains should call for psychiatric help at the proper moment.

The study of the psychological aspects of cancer continues to lead to many fundamental human issues, starting with bedrock existential confrontations with death and the sense of self. The recent work on families illuminates our interdependence upon one another. The issues surrounding genetic counseling, assisted-suicide, and other terminal-care dilemmas underline the grinding moral inadequacies of our social institutions. Ironically, half of the problem could be eliminated by altering a few personal habits, yet we nearly ignore this, while clinging to wisps of evidence about the “power of the mind to cure cancer.” This highlights the need for continuing

efforts to understand and alter the human response to the trauma of severe illness.

SUGGESTED CROSS-REFERENCES

Psychoneuroendocrinology is discussed in Section 1.1, the neuropsychiatric aspects of tumors in Section 2.2, cognitive-impairment disorders in Chapter 10, schizophrenia in Chapter 12, mood disorders in Chapter 14, anxiety disorders in Chapter 15, somatoform disorders (including pain management) in Chapter 16, and sexuality in Chapter 19. Death, dying, and bereavement are discussed in Section 28.5. Personality disorders are discussed in Chapter 24, noncompliance with treatment in Section 27.1, and suicide in Section 29.1. Psychotherapies are discussed in Chapter 30 and biological therapies in Chapter 31. Disruptive behavior in children is discussed in Chapter 40, and children's reactions to illness, hospitalization, and surgery are discussed in Section 49.5. Psychiatric problems in the medically ill elderly are discussed in Section 51.3a and the psychiatric aspects of long-term care in Section 51.6a. Legal issues are discussed in Section 54.1 and ethics in Section 54.2.

SECTION REFERENCES

*American Pain Society Quality of Care Committee: Quality improvement guidelines for the treatment of acute pain and cancer pain. JAMA *274:*1874, 1995.

Astin JA: Why patients use alternative medicine: Results of a national survey. JAMA *279:*1548, 1998.

Breitbart WS, Passik SD: Psychiatric aspects of palliative care. In *Oxford Texbook of Palliative Medicine,* D Doyle, GWC Hanks, I Macdonald, editors. Oxford University Press, New York, 1993.

*Edmonds CVI, Lockwood GA, Cunnigham AJ: Psychologocial response to long-term group therapy: A randomized trial with metastatic breast cancer patients. Psychooncology *8:*74, 1999.

Ernst E: Harmless herbs? A review of the recent literature. Am J Med *104:*170, 1998.

Geller G, Botkin JR, Green MJ, Press N, Biesecker BB, Wilfond B, Grana G, Daly MB, Schneider K, Kahn MJE: Genetic testing for susceptibility to adult-onset cancer. JAMA *277:*18, 1997.

Goodkin K, Antoni MH, Sevin B, Fox BH: A partially testable predictive model of psychosocial factors in the etiology of cervical cancer. I. Biological, psychological and social aspects. Psychooncology *2:*79, 1993.

Grassi L, Rosti G: Psychological morbidity and adjustment to illness among long-term cancer survivors. Psychosomatics *37:*523, 1996.

Green WH: *Child and Adolescent Clinical Psychopharmacology,* ed 2. Williams & Wilkins, Baltimore, 1995.

Holland JC, editor: *Psycho-Oncology*. Oxford University Press, New York, 1998.

Holland JC, Kash KM, Passik S, Gronert MK, Sison A, Lederberg MS, Russak S, Baider L, Fox B: A brief spiritual belief inventory for use in quality of life research in life-threatening illness. Psychooncology *7:*460, 1998.

Hubbard R, Lewontin RC: Pitfalls of genetic testing. N Engl J Med *334:*1192, 1996.

Kathol RG, Mutgi A, Williams J, Clamon G, Noyes R Jr: Diagnosis of major depression in cancer patients according to four sets of criteria. Am J Psychiatry *147:*1021, 1990.

Kissane DW, Bloch S, Burns I, McKenzies D, Posterino M: Psychological morbidity in the families of patients with cancer. Psychooncology *3:*47, 1994.

Kissane DW, Bloch S, McKenzie AC, McDowell AC, Nitzan R: Family grief therapy: A preliminary account of a new model to promote healthy family functioning during palliative care and bereavement. Psychooncology *7:*14, 1998.

Kreitler S, Chaitchik S, Kreitler H: Repressiveness: Cause or result of cancer? Psychooncology *2:*43, 1993.

Mackillop WJ, Zhang-Salomons J, Groome PA, Paszat L, Holowaty E: Socioeconomic status and cancer survival in Ontario. J Clin Oncol *15:*1680, 1997.

Manne SL, Bakeman R, Jacobsen PB, Redd WH: Childrens' coping during invasive medical procedures. Behav Ther *24:*143, 1993.

*Massie MJ, Gagnon P, Holland JC: Depression and suicide in cancer patients. J Pain Sympt Manage *9:*234, 1994.

Meier DE, Emmons CA, Wallenstein S, Quill T, Morrison RS, Cassel CK: A national survey of physician-assisted suicide and euthanasia in the United States. N Engl J Med *338:*1193, 1998.

Meyers CA, Weitzner MA, Valentine AD, Levin VA: Methylphenidate therapy improved cognition, mood, and function of brain tumor patients. J Clin Oncol *16:*2522, 1998.

Moorey S, Greer S, Bliss, Law M: A comparison of adjuvant psychological therapy and supportive counselling in patients with cancer. Psychooncology *7:*218, 1998.

Parle M, Jones B, Maguire P: Maladaptive coping and affective disorders among cancer patients. Psychol Med *26:*735, 1996.

*Posner JP: Neurologic complications of cancer. Davis, Philadelphia, 1996.

*Roth AJ, Breitbart W: Psychiatric emergencies in terminally ill cancer patients. Pain Palliative Care *10:*235, 1996.

Sirkia K, Hovi L, Pouttu J, Saarinen-Pihkala UM: Pain medication during terminal care of children with cancer. J Pain Symptom Manage *15:*220, 1998.

Sourkes B: *The Deepening Shade: Psychological Aspects of Life-Threatening Illness*. University of Pittsburgh Press, Pittsburgh, 1982.

Syrjala KL, Donaldson GW, Davis MW, Kippes ME, Carr JE: Relaxation and imagery and cognitive-behavioral training reduce pain during cancer treatment: A controlled clinical trial. Pain *2829:*189, 1994.

25.12 CONSULTATION-LIAISON PSYCHIATRY

JAMES J. STRAIN, M.D.

Consultation-liaison (C-L) psychiatry is the study, practice, and teaching of medical (Axis III) and psychiatric (Axis I and II) comorbidity. The dimension of this subspecialty of psychiatry ranges from screening, a spectrum of problem-level diagnosis to disorders, interventions, outcome, random control trials, costs, and ethical implications. A flow sheet for these crucial steps in the C-L process was developed to be a template to assist the consultant to include the steps in the process of the consultation, for example, screening, history taking, assessment scales, treatment approaches, and when drugs are used to have access to random control-trial results, drug-drug interactions, critical laboratory studies, technical reports with new findings, compliance measures, and documented patient outcome results (Fig. 25.12-1). Such a carefully orchestrated flow sheet would also assist in documenting that the consultation met quality assurance standards and legal requirements and assure that appropriate consents were obtained after capacity, and competence were established. Reliable and valid instruments that may help to document that the critical steps have been executed (e.g., diagnosis and outcome have been verified by appropriate measures) are also outlined (Fig. 25.12-1). As the steps of the consultation are completed, alerts occur to remind the psychiatrist that certain key issues must be addressed (e.g., noncompliance of the patient or nonadherence of the consultee to the suggested recommendations). This method ensures that crucial steps in the consultation process are completed and documented, and data are available for clinical and research purposes.

FIGURE 25.12-1 Computer-assisted physician management system. (Reprinted with permission from Strain JJ, Hammer JS, Caliendo G, Mayou R, Smith GC, Malt U, Lyons J, Kurosawa H: Further evolution of a literature data-base: Content and Software. Gen Hosp Psychiatry *18:*294 , 1996.)

Since C-L psychiatrists may be asked to consult on any patient in the outpatient or inpatient medical or surgical setting they may encounter any medical disorder or Current Procedural Terminology (CPT) intervention in the ninth revision of *International Statistical Classification of Diseases* (ICD-9) and any of the thousands of pharmaceutical agents that this broad spectrum of patients may be prescribed. The C-L psychiatrist must be conversant with interactions between psychotropic and medical and surgical drugs. Consequently, practitioners of this subspecialty must possess or have access to a monumental database if they are to manage their patients in a safe, therapeutically effective manner together with their medical and surgical colleagues. Possession of this knowledge base—psychological, psychiatric, medical, surgical—is essential for one of the missions of this subspecialty: teaching physicians, nurses, social workers, and patients.

The pedagogic functions of the C-L psychiatrist cannot be overstated. C-L psychiatry is the linchpin between psychiatry and the other medical specialties. The C-L role as teacher and purveyor of psychiatric skills in the medical setting maintains psychiatry as the *discipline of record* to assist with psychological, psychiatric, and psychophysiological morbidity in the medical setting. The C-L psychiatrist in many ways is the ambassador for psychiatry in the acute care and teaching hospital for tomorrow's doctors. Whether tomorrow's doctors turn to psychiatry or other mental health professionals for assistance (e.g., social work, psychology, clergy) depends upon their experiences with psychiatry in their formative years of training as medical students and house staff. Thus the C-L psychiatrist needs both the necessary knowledge base and the teaching skills to impart this knowledge and skill effectively to nonpsychiatric physicians.

The knowledge base in medicine is doubling every 5 1/2 years. In fact, 6000 to 7000 publications occur daily in science, and many of these pertain to medicine. Clearly, C-L psychiatrists need help in accessing and processing this outpouring of new information that concerns the optimal approach to the patient. Medical informatics, artificial intelligence, archives of clinically relevant information, and reference databases of seminal investigations are needed for clinical decision making. Whereas most subspecialties in psychiatry by their distinct focus limit the information with which they must be conversant, C-L psychiatrists must maintain an enormous range of information because they may be asked to consult on a patient with any physical disorder, who may be taking any medical drug, and who may be in any medical or surgical setting—a situation resembling that of primary care physicians.

COMORBIDITY IN THE MEDICAL SETTING: RATIONALE FOR C-L PSYCHIATRY

Psychosocial disability and psychiatric disorders are prevalent in the population and more often encountered in medical inpatient and outpatient settings than in psychiatric settings. One-month and point-prevalence studies found that 16 percent of community cohorts exhibited psychiatric morbidity. It is estimated that 21 to 26 percent of medical outpatients have psychiatric disorders. The age- and sex-adjusted prevalence of mental disorders was 25 percent in patients with a chronic medical condition and 17.5 percent in those without. Lifetime prevalence of a mental disorder in chronically physically ill patients reached 42 percent (most often substance abuse and mood or anxiety disorders) compared to 33.0 percent in those who did not have long-term physical disability. Finally, 30 to 60 percent of those in short term general medical or surgical inpatient settings have significant psychosocial or psychiatric morbidity.

Medical patients have a much higher morbidity for specific mental disorders than those in the general population: delirium in 15 to 30 percent of those hospitalized, depression (2 to 3 times higher), panic and somatization (10 to 20 times higher), and substance abuse (3 to 5 times higher). Thus the medical setting—where patients have already put themselves within a medical facility—is an excellent venue for screening, triage, and confronting psychiatric morbidity. In fact, the medical setting is an important de facto mental health setting from which to launch an effort to ameliorate mental suffering.

In addition, psychiatric morbidity is associated with high utilization of general medical services and an increase in several parameters of health services utilization. Investigators found that 50 percent of high utilizers had psychiatric morbidity: major depressive disorder or dysthymic disorder (40 percent), anxiety disorders (21.8 percent), somatization disorder (20 percent), panic disorder (12 percent), alcohol or other substance abuse (5 percent). Depressed patients in the medical setting use three times more health resources, encounter twice the costs, and make seven times more visits to the emergency facilities than those not depressed. Patients with panic disorder visit the emergency room 10 times more often than those without (70 percent of those with panic disorders encounter 10 or more doctors before a correct diagnosis is achieved). The need for short-term hospitalization can triple when asthma patients experience comorbid anxiety disorders.

Patients with alcohol abuse or dependence have twice the total health care costs of those without, and, one-fourth to one-half of patients with alcohol disorders are not identified as such when admitted to an acute care medical and surgical hospital. Patients with diagnosed depressive disorders used more health care resources than those not depressed ($4246 vs. $2371). Similar differences were found in those with anxiety disorders and those without ($2390 vs. $1397). When patients with specific illnesses are investigated the economic consequences of psychiatric morbidity in the medical setting become even more pronounced: Mean rehospitalization costs in patients with cardiac disease and psychological distress versus those without were $9504 and $2146, respectively. In a review of 26 studies, a significant association was found between length of hospital stay and psychiatric morbidity in 89 percent of studies with sample sizes above 110 patients. This increased health services use involved the emergency room, hospitalization, and rehospitalization, and it continued for at least 4 years after the first hospital discharge.

Psychiatric morbidity compromises patients' functional status and quality of life. The Medical Outcomes Study states that depressive morbidity carries profound physical impairment equivalent to that observed in several other chronic medical conditions (e.g., diabetes, hypertension, arthritis, coronary artery disease), and the social impairment is even greater. (Studies such as that of David Spiegel and coworkers observed improved quality of life after a C-L psychiatric intervention. Improved quality of life is an important research agenda.)

Since the majority of medical, surgical, obstetrical, and gynecological patients receive most of their medical care (including attention to their psychosocial needs and psychiatric morbidity) in the primary care setting, methods are needed to transfer knowledge and skills to this nonpsychiatric setting on a continuing basis. In fact, 44 percent of patients with mental disorders receive their only mental health care in the general health setting, and detection rates in primary care for substance abuse, alcohol abuse or dependence, anxiety disorders, and depressive disorders are only about one-third to one-half. In the general hospital, less than half of the inpatients with depression, delirium, dementia or alcohol abuse or dependence are identified (and detection without appropriate treatment and follow-up does no good). Rand studies demonstrate insufficient mental health care in the general medical setting: minimal psychopharmacological intervention for depression with subtherapeutic dosage; inadequate information to the patient, prompting too-early drug discontinuation; inadequate number of visits and lack of attention to long-term monitoring resulting in half of the patients with depression treated by primary physicians remaining symptomatic months later.

By and large, the venue of primary care does not lend itself to encouraging enough attention to psychiatric issues. The physician must direct a significant effort to ensuring that there are no physiological abnormalities, and payment for services is geared to cover only modest amounts of the physician's time for mental health screening and treatment. Patients may not readily expose their psychiatric problems, but rather emphasize somatic complaints that encourage the doctor to perform repetitive physiological examinations. Through the process of *somatization* (expressing mental feelings and conflicts via the body), neither the patient nor the primary care physician may be aware that the physical complaints emanate from psychological mechanisms. Mind-body dualism prevails, with patients and their doctors locked into endless examination of the soma in order to not miss something that could be life threatening. But they may miss another mechanism for the somatic disturbances—that is, psychiatric morbidity. They are practicing the biomedical model rather than the biopsychosocial interactional model espoused by George Engel in 1977.

SCREENING

Medical students and physicians need practical training and education in clinical diagnosis and treatment of mental disorders, particularly in the medical and surgical environment. Appropriately trained physicians are the best screening tools for mental pathology in the medically ill; however, as an interim measure, easily administered screening devices have been developed. Such devices become more necessary when mental health care is placed within the context of primary care. Most patients with comorbidity first appear in the general medical setting, and these physicians must identify the patient in need of a psychiatric consultant if they cannot treat the patient themselves. Reliable and valid screening instruments are needed by nonpsychiatric physicians and their assistants to facilitate detection of mental disorders, but adequate diagnosis and appropriate treatment must follow. Screening is only the beginning of a process.

David Goldberg developed one of the first mental health screening devices for the medical and surgical setting, the General Health Questionnaire, which has had numerous iterations and forms since its introduction in 1974. However, it is a measure of morbidity and is neither syndrome nor medical disorder specific. It identifies subthreshold states, disturbance of mood, and decreased well-being. Such broad inclusiveness will identify too many individuals beyond those with psychiatric morbidity. It has been superseded by contemporary instruments, some of which can even be administered over the telephone from home before arriving at the doctor's office or hospital, for example, Symptom Driven Diagnostic Scale (SDDS) developed by Eugene Broadhead, A.C. Leon, Myrna Weissman, and coworkers. The SDDS has a computer-based format—the CATI—that provides a telephone or admission screening result with a scored statement of the potential risk for a specific or multiple psychiatric morbidity; depression, panic, obsessive-compulsiveness, anxiety, and suicidality are automatically forwarded to the physician as a mental health alert. Presenting this alert to physicians before they evaluate the patient should focus history taking to confirmation the suspicion of a psychiatric disorder. This instrument can also be used for diagnostic and follow-up evaluation of psychiatric morbidity during and after a course of treatment.

However, it is only a screen and as such requires confirmation of the presence of the essential attributes of a diagnostic algorithm from the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) namely, the vegetative signs and symptoms should not be secondary to physical causes. Such an alert would mandate further examination by history, physical examination, and mental status assessment for the presence of core psychiatric disorders and save the physician initial triage for every patient. Only those at risk for the disorders are further screened by the history and physical investigations. Thus, much like the hematocrit, which alerts one to an anemia that requires further exploration, the screen points to the need for a mental health workup, treatment, and follow-up of the patient. No patient should leave the medical setting with a known hematocrit of 30 or less without an anemia workup, and no individual should leave a medical setting with undetected depression, anxiety, panic, suicidality, or substance abuse and lacking expert management. The screening assessment does not require the physician's time or presence for its completion, thus accommodating the significant demands and time constraints of contemporary medical and surgical practice.

The PRIME-MD, developed by Robert Spitzer, is another commonly used screening device. Such instruments allow modestly skilled persons to develop a profile of patients at risk for mental morbidity and immediately bring it to the attention of those who can perform a more rigorous clinical evaluation and provide the necessary mental health treatment.

Screening devices must be used with caution. The assessment of cognitive dysfunction by two commonly used instruments—the Mini Mental Screening Examination and the Cognitive Capacity Screening Device—are confounded by the frequent occurrence of false-positive and false-negative results, up to 25 percent in some studies. In the former case the patient may be exposed to an unnecessary workup, and in the latter (the false-negative result), critical assessment to determine the underlying mechanism for cognitive dysfunction (of which even the patient may be unaware) is overlooked, at times a life-threatening omission. The first challenge is developing a mechanism to screen every patient entering the health care system for psychosocial and psychiatric morbidity, with valid tools to alert the caregiver to the potential presence of mental states that must be addressed, which is of cost and risk benefit. The second step is confirming or ruling out this suspicion. The third requirement is formulating a diagnostic and treatment plan to ensure that the mental morbidity is properly identified and treated.

Concomitant medical illness confounds screening for mental illness. Reliability and validity are challenged by the presence of physical symptoms that can emanate from either the physical or mental sphere. However, the source is not always known and for illnesses such as depression, anxiety, and somatization it is essential that physical symptoms that emanate from physical causes are not counted in the diagnostic algorithm. The instruments commonly used to assess for the presence of mental disorders (e.g., Beck Anxiety Inventory, Beck Depression Inventory, Hamilton Rating Scale for Depression or Anxiety Scales) do not sufficiently take into account the contamination of their scores by physical symptoms that result from physical illness (e.g., anorexia, insomnia, decreased energy, diminished libido) or ideational signs that commonly accompany physical illness (e.g., guilt over smoking and helpless or hopeless feelings in cancer), all of which are expected human responses to serious medical morbidity and its treatment.

PROBLEM-LEVEL DIAGNOSIS

In the C-L psychiatric setting interaction with the patient usually begins with a problem-level diagnosis (the chief complaint). The patient may be referred for noncompliance, poor coping, strange or bizarre behavior, confusion, or staff conflicts. It may be difficult for the consultee to be more specific, especially in the face of acute or life-threatening physical illness. Primary care medicine has long known and respected the problem-level of diagnosis, knowing that many patients will not meet the criteria for a specific diagnosis; as many as 50 percent of those seen in the ambulatory setting do not receive a physiological, psychological, social, or combined diagnosis but remain at the reason-for-visit–level or problem-level of diagnosis.

Diagnosis proceeds along a continuum: (1) understanding the stated reason for the patient's visit or referral; (2) assessing the reason for contact; (3) formulating (if possible) a diagnosis or multiple diagnoses with a known cause and prognosis. The patient can be managed (treated) at any point on this continuum. The patient with acute chest pains or a bizarre psychotic reaction is managed (e.g., hospitalized) and medicated even before the underlying problem has been definitively diagnosed,

if it ever is. The patient could be signed out with chest pain of unknown origin and no more specific diagnosis. Noncompliance or phase of life problem may not be differentiated into a depressive, anxiety, or psychotic disorder but be left at the V-Code level of diagnosis (i.e., conditions that may be the focus of clinical attention).

On the other hand, use of computerized clinical database assessment techniques showed that many of the consultee's problem-level diagnoses (e.g., noncompliance) could be further reduced to specific psychiatric categories such as depressive disorders, anxiety disorders, or cognitive disorders. Such differentiation was not possible by the primary care physician. Less than 50 and 22 percent, respectively, of alcohol and delirium, dementia and amnestic and other cognitive disorders were properly identified by the consultee in the inpatient medical or surgical setting. Several reasons may explain the difficulty the primary physician has with identification. The clinical state may be at a subthreshold level and has not reached criteria for an apparent psychiatric diagnosis. Primary care physicians may not switch hierarchical levels from the physical to the psychological mechanism of causality; it may be difficult for them to discern the interrelation between somatic and mental phenomena and incorporate such mechanisms into their diagnostic formulations (e.g., delirium due to metabolic causes). Even the identification of biological, psychological, social dysfunction (triaxial diagnostic systems) does not mean the physician is aware of the interactions among the systems to arrive at a higher-order, more complex, or multiple interactional diagnosis, such as a depressive disorder due to cancer of the pancreas, psychotic disorder due to multiple sclerosis. Therefore, some cases remain diagnostic problems for the primary care physician, while others can be further specified when the medical morbidity improves and allows the psychiatric morbidity to be more apparent or disappear.

DISORDER-LEVEL DIAGNOSES

To ensure appropriate treatment within the medical model, the clinician strives for diagnostic certainty. The major dilemma is the attribution of physical symptoms, since the decision-making rules of DSM-IV state that physical symptoms secondary to a medical illness should not be counted in the algorithm for mental disorders. However, it is difficult to ascertain whether a symptom emanates from the body or has a psychological origin or both.

Three pain disorder categories are now available to the diagnostician: (1) pain associated with psychological factors, (2) pain disorder associated with both psychological factors and a general medical condition, and (3) pain disorder associated with a general medical condition (in which case the pain diagnosis is placed on Axis III and not Axis I). This classification takes into account that the most common reason for psychiatric consultation is combined psychological and physical pain, that it was an appropriate reason for referral, and that it is a knowledge and skill area in which the psychiatrist should be trained (e.g., a part of the residency curriculum). It also incorporates the reality of diagnostic uncertainty in physical symptoms that may emanate from a mental or physical sphere. The traditional somatoform-conversion pain disorder was relatively rare compared with the psychiatrist's need to assist with combined pain. The old classification had no Axis I for coding or billing purposes for appropriate psychiatric consultative effort. A similar conundrum obtains for the somatoform disorders, psychiatric disorders that present with physical symptoms, which the patient is convinced have a somatic origin. Somatization disorder (Briquet's disorder) patients besiege the primary care physician to find the physical source of their plight. They have numerous examinations and frequent emergency room visits, become high medical-resource utilizers, and continue to have negative physical findings in one workup after another. They drive themselves and their doctors to distraction and are outraged with each other.

> A 32-year-old black unmarried woman came to the physician for recurrent chest pain, after seeing several clinic and emergency room physicians for assistance in relieving the pain. Despite countless examinations and two hospitalizations, the patient insisted that something was physiologically wrong and that she must have a serious illness that the doctors could not diagnose. Finally, she was sent to a prestigious medical facility in a neighboring state, where a different set of physicians performed a complete, extensive workup and found no discernible pathology. The patient wanted one more computerized tomographic (CT) or magnetic resonance imagery (MRI) scan, and even wanted to have a surgeon look into her chest to find the disease. With increasing despair and mounting depression, the patient threatened to kill herself and blamed her family and physicians for her plight. Despite being a highly educated professional, she could not function or leave her apartment except to see the next physician, and she believed that her life was over. Observation and treatment on a psychiatric unit considerably diminished her physical complaints as she underwent cognitive-behavioral treatment, physical rehabilitation, and psychotropic medication, and was reassured that nothing had been missed (reframing her understanding of her significant pain that would in time be reduced), that gradually she would gain greater control over her body, and that she did not have a life-threatening illness. The psychiatrist knew that at times she had unbearable physical symptoms, that she would learn to manage them better and better, and that in time she would resume her previously productive and rewarding life.

These patients are certain that they have physical illness and often eschew any reference or referral to psychiatry. Close collaboration of the primary care physician and the psychiatrist is needed to shift the focus to psychological processes and conflicts without discounting the bodily symptoms and the patient's concern.

Factitious disorder patients are aware that they are causing the symptoms but do not know why they are compelled to do so.

> A 26-year-old registered nurse was admitted to the medical ward with a fever of unknown origin. Extensive workup revealed no physiological mechanism. On her fourth hospital day her primary care nurse found her injecting herself in the thigh. When questioned, the patient was embarrassed and tried to hide the syringe. The nurse took the syringe, and when its contents were examined, she learned that the patient had been injecting herself with her own feces to incite a fever and restart the process of a fever workup, which had confounded the staff for the last 2 years. When questioned about her need to do this, the patient had no explanation. Despite recurrent psychiatric consultations, she remained oblivious to the mechanism underlying her behavior and demanded to leave the hospital against medical advice. Having been exposed, she wanted to leave that medical setting and had no interest in learning more about her need to hurt herself or what was behind her self-destructive behavior.

Factitious disorders can occur with physical symptoms, mental symptoms, or by proxy where the patient has a perpetrator who claims the patient (the victim) has physical symptoms (e.g., mother of a child, or an adult child of an elderly parent, in which the perpetrator wants the victim (to be ill).

Finally, *malingering* (i.e., the feigning physical symptoms for secondary gain, in which the patient displays self-simulated symptoms) is a problem-level (V-code) diagnosis demanding attention by the psychiatrist, usually within the medical setting. This diagnosis requires proof of simulation and indentification of the secondary gain.

Unfortunately, patients who present with physical symptoms with no detectable physiological origin are too often labeled malingerers. The differentiation of the somatoform disorders, factitious disorders, and malingering is problematic not only for the psychiatrist but for the primary care physician in particular. In fact, this triad challenges the most experienced clinician and highlights the diagnostic dilemma for the psychiatrist in the medical setting when unexplained physical symptoms are under deliberation. Unfortunately, malingering is too often the default diagnosis for unexplained physical symptoms. In addition, once real malingerers are found to be simulating for secondary gain (and exposed), they usually leave that medical environment and simply move on to the next physician, with their behavior unchanged.

Physical symptoms are common in the entire spectrum of mental disorders. It is medically unsafe to categorize a symptom as psychological simply because no organic attribution can be established. If the disorder does not reach the threshold for a mental disorder, then the symptoms must remain at the problem level of diagnosis, with their cause regarded as uncertain. The adjustment disorders themselves are problematic because they have no symptom checklist but rather rely on the presence of a maladaptive response to a stressor (which may be medical illness), with the descriptor taken from the prevailing mood. Since there are no valid or reliable criteria to reach threshold for these diagnoses (so commonly used in the medical setting) and since maladaptation may include problems in mental functioning or bodily competence, this ubiquitous diagnosis in medically ill patients remains problematic. Because the standard instruments of assessment have not included calibrations for the physical symptoms they assess (e.g., 13 of 20 on the Hamilton Rating Scale for Depression) the likelihood of scoring depression in the medically ill is extremely high. Similarly, such instruments for anxiety disorder, which itself has many physical symptoms, and those for somatization disorder confront the diagnostician with, at times, insurmountable validity questions.

Nor are dementia and other cognitive disorders, mental disorders due to a general medical condition (e.g., stroke, multiple sclerosis, systemic lupus erythematosus) and substance–related disorders immune from the confounds of diagnosing psychiatric morbidity in the medical setting.

> A 72-year-old patient with alcohol abuse seen in the intensive care unit after surgery for the extravasation of a subdural hematoma was delirious postoperatively, had experienced blood loss, anesthetic agents, analgesic drugs, and exposure to a strange and confusing environment. Classification of the patient's delirium is problematic, since it could be secondary to substance abuse, medical drugs, blood loss, detoxification, or the foreign environment, and the acute state observed could cover an underlying chronic state such as dementia.

Therefore, diagnosis in the medical or surgical setting remains a process requiring initial impressions and follow-up observations to document the mental status as

medications change, the medical condition evolves (including the diminution of pain), the environment becomes more familiar and less frightening, and underlying personality traits and chronic coping styles and strategies emerge. Even in retrospect, diagnostic certainty may be elusive. Yet, this does not preclude interventions to enhance behavioral control and compliance and diminish anxiety in the medical care setting.

TREATMENT

Treatment follows the best diagnosis possible given the potential confounding factors. In recent years, treatment guidelines have been developed on the basis of the best evidence available.

According to the Institute of Medicine's *Guidelines for Clinical Practice: From Development to Use*, developers of guidelines should incorporate eight attributes: (1) validity, including strength of the evidence and estimated outcomes; (2) reliability and reproducibility; (3) clinical applicability; (4) clinical flexibility; (5) clarity; (6) multidisciplinary process; (7) scheduled review; and (8) documentation. These attributes were incorporated into the American Psychiatric Association's guidelines for the treatment of depression in primary care. The United States Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research (AHCPR) has sponsored over 19 guidelines for treatment, many of particular importance to the C-L psychiatrist. For example, guidelines on management of cancer pain, acute pain management in adults: operative procedures, managing early HIV infection, acute low back pain, post stroke rehabilitation, cardiac rehabilitation, smoking cessation, Alzheimer's disease, anxiety and panic disorder, and chronic headache pain were constructed by incorporating these same attributes, among which one of the most important is scheduled review and constant updating and reconsideration of the strength of the evidence.

It is anticipated that guidelines can become the textbook of tomorrow, incorporating new findings and being published in a few weeks to months for rapid updating, almost the way medical journals are now distributed but even more efficiently. Amendments can be distributed overnight to update diagnostic and treatment approaches, using electronic alerts to translate new findings into therapeutic strategy and tactics for the clinician. The C-L psychiatrist must be aware of these guidelines and provide feedback to developers and promoters to enhance their utility. (Many guidelines have few psychiatric or behavioral statements or recommendations). Therefore, guidelines not specifically geared to psychiatry (e.g., Diagnosing and Managing Unstable Angina, Urinary Incontinence in Adults, Benign Prostatic Hyperplasia: Diagnosis and Treatment) may benefit from the input of C-L psychiatrists to augment the psychosocial or psychiatric morbidity component. Specialty organizations such as the American Psychiatric Association have also developed important guidelines to be used for comorbidity in the medical setting such as Practice Guideline for Major Depressive Disorder in Adults. These guidelines developed with the psychiatric patient must be amended to incorporate conditions and modifications for individuals who have psychiatric and medical comorbidity. They, like the assessment instruments described above, too often lack the dimensionality considerations that concurrent medical illness superimposes on psychiatric morbidity. Tomorrow's guidelines for the treatment of psychiatric morbidity must include algorithms and parameters to account for medical illness, which may add risks and injunctions to treatment in the medical setting.

The strength of the evidence for treatment has been codified and enhanced by the Cochrane Collaboration, Oxford University, funded by the United Kingdom's National Health Service, which has established criteria for the assessment of validity of random control trials or treatment trials. This is the highest level of assessment, promulgated in the Cochrane Database of Systematic Reviews which includes “Stroke Unit Trialists' Collaboration: A Systematic Review of Specialist Multidisciplinary Team (Stroke Unit) Care for Stroke Inpatients,” “Elective versus Caesarean Delivery of the Small Baby,” “The Effects of Family Intervention for Those with Schizophrenia,” and “Support from Caregivers for Socially Disadvantaged Mothers.” These meta-analytical studies by clinical experts in a given discipline offer the C-L psychiatrists a cross-comparison of studies that heightens the strength of the evidence and thus should have an important impact on treatment and clinical decision making. New reviews are continually being developed, and in time such efforts will help move the world's repository of random control trials to a more refined definition of reality than exists in most medical treatment areas to date. This effort goes well beyond one research team or one finding. These reviews are labor intensive; they require hand examination of journals.

PSYCHOTROPIC-MEDICAL DRUG INTERACTIONS

One of the most critical areas for the C-L psychiatrist involves medical drug–psychotropic drug interactions. In addition, the C-L psychiatrist must understand pharmacokinetics as it relates to end-organ dysfunction (kidney, liver, lung, heart) and the aging process. Since, by definition, most patients in the C-L setting present with physical illness, prescription of psychotropic drugs entails knowing the pharmacokinetics and interactions of the medical drugs their consultation patients are taking. Thus authors, Gina Caliendo and Carol Himelein have developed an approach to (1) review the literature, (2) codify and describe the most important psychotropic drug–medical drug interactions, and (3) access and update new findings from a variety of sources not customarily used by the psychiatrist. Many psychotropic medications have not been tested against medical drugs, especially newer ones such as protease inhibitors for human immunodeficiency virus (HIV) infection. Thus, the C-L psychiatrist must have access to technical reports and a way to update their knowledge of potential interactions. There is a lag time from publication to inclusion in the MEDLINE system, and as stated above, many published studies do not appear in MEDLINES. Updates on medical drug–psychotropic drug interactions are often in the form of case reports and letters to the editor. Although C-L psychiatrists may not have ready access in their offices to the following materials, they must know about them and where to access them (usually the pharmacy department of the hospitals in which they practice).

One important source for updates in this crucial area is the *F-D-and-C Pink Sheets*, a weekly publication that highlights regulatory issues for new drugs and those soon to be approved and important labeling changes. Computerized methods of accessing important pharmacological literature exist; most hospital pharmacy services have access to these instruments and will assist the C-L psychiatrist with queries regarding psychotropic drug–medical drug interactions. MICROMEDEX is a computerized database compiled from multiple drug information systems. DRUG-REAX is a subsection of MICROMEDEX that searches multiple drugs simultaneously for drug-drug, drug-food, and drug-disease interaction data. DRUGDEX is a software system that contains drug monographs with drug interactions obtained from case reports and drug interaction references. In addition, many journals and publications include information on medical drug-psychotropic drug interactions: *Pharmacotherapy, Annals of Pharmacotherapy, Clinical Pharmacology and Therapeutics, Drug Safety, and Drugs* . Monitoring services include *Clin-Alert; Reactions, Clinical Abstracts/Current Therapeutic Findings; Current Contents/Adverse Reactions and Excerpta Medical/Adverse Reactions*; and *American Journal of Health-System Pharmacy, Current Literature* (a section of the *Journal of the American Society of Health-System Pharmacists)*. Most of these resources are not known or used by psychiatrists but are the information systems of pharmacists and consultants. Most large institutional settings (e.g., university teaching hospitals) have established drug information centers within their pharmacy departments. Trained pharmacists with access to the information sources above are more adept at determining potential medical drug–psychotropic drug interactions or can provide the best advice given the state of current knowledge. Pharmaceutical companies also have information services and can often assist with potential or reported interactions with their products. These in-house information services cull data from the published literature, clinical trials, foreign databases, and unpublished spontaneous case reports from medical professionals. An example of one hospital's attempt to systematize psychotropic drug–medical drug interactions is given below and described in Table 25.12-1 and Table 25.12-2.

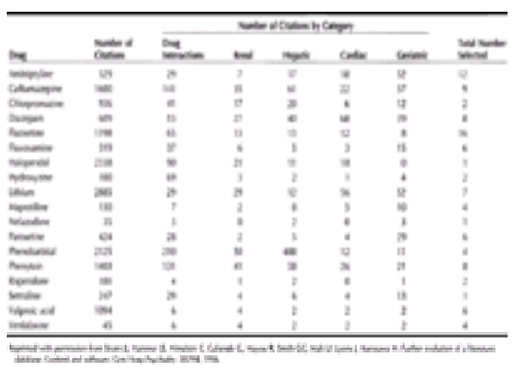

Table 25.12-1 Sample of Search Results Psychotropic Drug Medical Drug Interactions

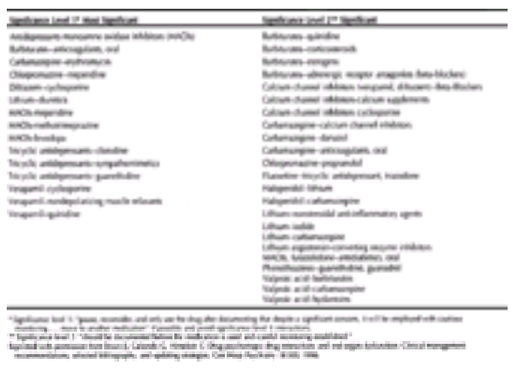

Table 25.12-2 Psychotropic-Medical Drug Interactions and Significance Level

Interaction: Carbamazepine–Erythromycin Significance Level 1: Most Significant

Advice to Practitioner: Erythromycin may inhibit the hepatic metabolism of carbamazepine and toxicity may result. Interactions of this type present problems when the drug that alters hepatic metabolism is added after a patient is stabilized on the affected drug. A problem may also arise if a patient is stabilized on the affected drug while receiving an enzyme-altering drug and that enzyme-altering drug is discontinued. Interactions of this type may take a few days to become evident and resolve because effects of the enzyme altering agent may take a few days to become evident or dissipate. If a patient is stabilized on the combination, there is little need to alter therapy. If a patient is on erythromycin and carbamazepine is added, the dosage of carbamazepine will be determined by serum concentrations and clinical efficacy, and no additional intervention is required. If a patient is on carbamazepine and erythromycin is added, serum concentrations of carbamazepine may increase during combination therapy, and may result in toxicity. An alternative antibiotic should be used if possible.

Action: Monitor carbamazepine serum concentrations when erythromycin is added to carbamazepine, and an alternative antibiotic is not required.

Alan Stoudemire and T.M. Brown have an important publication that addresses the issue of psychiatric adverse effects of prescription and over-the-counter medications, which may also be useful to the C-L psychiatrist.

COMPLIANCE

Not only research but clinical care is undermined by noncompliance in the patient, the consultee, or the consultant; the patient may fail to comply with the prescribed regimen; the consultee may fail to follow the consultant's recommendation (*nonconcordance* with recommendations); while the consultant may not comply with the suggested treatment protocol, algorithm, or guideline *(nonadherence)*. Any of these three behaviors leads to error measurement in the outcome results of research. For the patient, noncompliance may mean less than optimal behavioral changes that at times can be life saving (e.g., tearing out intravenous lines in the intensive care unit because of a delirious agitated state).

One study found that consultees used the C-L psychiatrists' recommendations for psychopharmacology, consultations, and diagnostic labeling on 71, 43, and 39 percent of the prescribing occasions, respectively. It is not surprising that consultees may be reluctant to prescribe medications that they are unfamiliar with or for a frail or especially vulnerable patient, those with the recent acute surgery, elderly patients, or those with nascent trauma. The C-L psychiatrist needs to tell the consultee about expected adverse effects of psychotropic drugs as well as central nervous system (CNS) manifestations that may accompany medical drugs; for example, many of the antiviral, anti-infection, and antineoplastic chemotherapeutic agents used to combat (HIV) (e.g., zidovudine (Retrovir) affect mood and cognition (Table 25.12-3). The consultant must also encourage the use of indicated psychotropics when there is reluctance on the part of the consultee because of inadequate information, misguided beliefs, or inadequate understanding.

Table 25.12-3 Psychiatric and Neurological Side Effects of Medications Used in Treating AIDS

Just as diagnostic impressions must be confirmed with reliable and valid rating scales and treatment strategies must be titrated against specialty-developed guidelines, compliance must be assessed as a part of the consultation process by consulting with the patient, the caregivers, and the medical chart. Until it is confirmed that the recommendations were invoked, assessment and the meaning of outcome are in peril of error measurement. Since noncompliance in all its forms is ubiquitous in medicine and since it may have such a pernicious effect on preferred outcome, it becomes an essential attribute of the consultation process, and the fate of recommendations must be documented. Currently, compliance is seldom commented upon, and its documentation in the consultant's note is even rarer.

PATIENT OUTCOMES

Patient outcome studies from C-L psychiatric interventions in the inpatient or outpatient medical setting have shown important contributions to the well-being of patients and often decreased costs of medical treatment.

Inpatients Steven Saravay and Mark Lavin examined 26 outcome studies (17 U.S.) on the effects of psychiatric morbidity and hospital length of stay. Eighty percent of the studies selected and 89 percent of those with a sample size larger than 110 reported a significant association between increased length of hospital stay and psychiatric morbidity, and that depressed mood, impaired cognition, and other personality variables were associated with increased length of hospital stay and greater utilization of rehospitalization and general medical services for up to 4 years after discharge from an initial hospitalization.

C-L psychiatric intervention studies for acutely hospitalized medical or surgical patients demonstrated improved patient status and decreased costs. When psychiatrists became part of the orthopaedic team and examined every elderly patient admitted during 1 year for the surgical repair of an acute hip fracture and treated those requiring psychiatric intervention, patients (1) were found to have significant diagnosable DSM-III-R disorders (most commonly, organic mental and mood disorders, 56 percent), (2) were discharged with less psychiatric morbidity, (3) left the hospital on average 2 days sooner than those in comparison groups, (4) had fewer rehospitalizations, and (5) needed fewer rehabilitation treatment days. There was no evidence of cost shifting to ambulatory settings. A $20,000.00 C-L psychiatric intervention cost saved $167,000.00 hospital-stay dollars on one illness in 1 year. Other studies demonstrated that early discharge planning and timely, focused interventions for psychosocial factors that affect length of hospital stay produced significant financial savings. Since inpatient hospital care remains the most expensive delivery component of medical service in our health care system, the Health Care Financing Administration has been issuing guidelines since 1979 in an

attempt to limit length of hospital stay to only what is medically necessary.

With both the high incidence of psychiatric morbidity in the general, short-term care, inpatient, medical and surgical hospital setting and the demonstration that a C-L psychiatric intervention increases recognition of morbidity, improves patients' status upon discharge, and decreases costs, provision of C-L psychiatry in these settings appears to offer a significant cost-to-benefit ratio and should be a mandatory service for high-risk patients (e.g., elderly persons with hip fracture, persons with HIV infection, and those with acute myocardial infarction with depression).

Outpatients In a review, G. Richard Smith found the prevalence of somatization disorder in the primary care outpatient setting to be 9 percent, three times higher than that in the general population. Somatization disorder patients also had more significant functional impairment (e.g., sick days, days in bed, physical restrictions, poor psychological health, impaired parenting) than patients with medical illness alone, and somatization dysfunction is additive to what may be experienced from physical morbidity per se, which results in seriously impaired and dysfunctional patients.

PRIMARY CARE PHYSICIANS

Wayne Katon and his coworkers used double-blind, controlled intervention studies to determine what training algorithms are required to promote better patient outcome for depression treated by primary care physicians. Screening for depression, informing the physician of its presence, and a brief concurrent intervention by the psychiatrist with the patient and the primary care physician had little impact on physician behavior or patient improvement. A more comprehensive intervention by the psychiatrist with the primary care physician, which included joint interviews, educational efforts for the physician and patient, and repeat visits with physicians, did improve psychiatric morbidity. This conjoined evaluation, treatment, and educational paradigm was effective when lesser interventions did not ameliorate the depression.

By analyzing massive databases of three large health studies (the Health Insurance Experiment, the Medical Outcomes Study, and the Prospective Payment System Quality of Care Study), Kenneth Wells and his collaborators found that both outcomes and quality of care of depressed patients in primary care treatment settings are inadequate. For example, 40 percent of depressed patients were treated with subtherapeutic dosages of medication; prepaid care had lower rates of detection and referral for psychosocial treatments for depressed patients than fee-for-service programs. Significant numbers of depressed patients cared for in primary care experienced inadequate interventions and exhibited marginal improvement.

Spiegel demonstrated that women with metastatic breast neoplasia (stage IV) who were treated in an ongoing group therapy protocol lived 18 months longer than controls and had a better quality of life. This important study is now under replication at several national and international sites. Fawzy Fawzy observed that patients with metastatic melanoma who were entered into group and educational therapy protocols survived longer and had better quality of life than matched controls. These studies suggest that a psychosocial or psychiatric intervention not only improves life, but also prolongs it.

The current literature indicates that psychiatric morbidity in both outpatients and inpatients in the medical setting is generally undetected, and that these patients use more health resources, increase the cost of health care, and have more-impaired functional status than patients with medical illness alone. Lower-intensity interventions (e.g., notifying the primary care physician of psychiatric morbidity) are of minimal consequence in effecting change. Structural change in the pattern of health care provided by primary care is needed to ensure that adequate C-L psychiatric involvement occurs at every phase of treatment: assessment, treatment, and follow-up. The liaison skills—case-finding, education, and training of medical caregivers—of the C-L psychiatrist need to be incorporated into the structure of primary care so that appropriate screening, triage, treatment, and follow-up are accomplished early in the patient's encounter with primary care, whether in the inpatient or outpatient setting.

RANDOM CONTROL TRIALS: FUTURE RESEARCH AGENDA

Random control trials are confounded by psychiatric and medical comorbidity. Two recent random control trials in chronic fatigue syndrome and in patients with medically unexplained physical illness found significant lessening of symptoms of those receiving cognitive-behavioral treatment compared with placebo controls. The Spiegel and Fawzy studies of psychiatric interventions for stage IV breast cancer patients and melanoma patients, respectively, used random control trial methodology, but questions of severity of medical illness and comparability of medical treatment remain. Controlling for severity of psychiatric illness and medical illness is difficult. Measures of both spheres have many questions about reliability and validity at this time. Most measures of physical illness have not been standardized to take into account psychiatric morbidity, and psychiatric rating scales have not been modified to correct for potential challenges to validity by concurrent medical illness. How one random-control trial compares with another examining the same issue is the key to legitimizing the value of even this highest form of rigor in scientific endeavor. Since it is difficult to ascertain the reliability and validity of some psychiatric diagnoses in the medical setting and whether a psychosocial or psychiatric intervention has been fully implemented is often uncertain, errors in measurement often occur. If the random control trial is to be a system intervention (e.g., a structural consultation–liaison psychiatric change on one inpatient ward versus another, a new technique on one ward in contrast to a traditional technique on another unit), the study may be compromised by nonrandom admissions on one unit or the other. Hospitals cannot afford to hold a bed open for random assignment of a study patient. Random control trials may be approached more easily in the outpatient setting, but standardization of medical treatment and complete delivery of the psychiatric intervention enhance uncertainty and raise questions about the integrity of the therapeutic arm of the study.

An important new venue of care for comorbidity that needs to be studied with random control trial methodology is the psychiatry-medical unit (if under the Department of Psychiatry), or the medical-psychiatry unit (if it is a unit of the Department of Medicine). Roger Kathol has developed this genre of care facility at the University of Iowa to one of the highest levels in the United States. This inpatient care facility attends to medically ill patients too psychiatrically ill to be on a medical ward and too medically ill to be on a psychiatric inpatient unit. It studies chronic pain patients, somatization disorder, chronic fatigue syndrome, and those with profound somatic presentation of psychiatric symptomotology. These patients with diagnostic dilemmas and the need for short-term care for both their psyche and their soma require special physicians, nurses, and a ward structure that can address diagnoses and therapeutics in both dimensions.

Empirically, it makes sense to have a setting that can optimize care in both spheres of the human being. However, these units are costly to maintain, require staff competent in both domains (not easy to locate), and must deal with patients that abuse the medical system with their constant complaints and dissatisfaction with medical care. The effects of these units on their unique patient cohort and the cost-benefit ratio in the use of health care resources have yet to be subjected to random control trials. Until then, they are experimental, and short-term care hospitals should not universally set out to establish psychiatry-medicine units, as they have currently for coronary care, intensive care, and respiratory intensive care without ascertaining the effect size on patient outcome and comparing costs with those of traditional medical or psychiatric care units. Furthermore, existing psychiatric units need to expand their capacity to care for more seriously medically ill, neurologically complex, and elderly patients, in essence becoming de facto psychiatry-medicine units.

Transplantation services exist in many university teaching hospitals. Their medical effectiveness has been sufficiently proved to warrant payments by the Health Care Financing Administration and many third-party payers. What remains to be determined is the place of psychiatric consultation and treatment for this special cohort of patients. Although many transplant units have psychiatrists attached and have psychiatric assessment built into the selection procedure for organ transplantation, the contribution of the psychiatrist to selection, treatment, and patient outcome remains in question. No study has compared outcomes with and without psychiatric assessment in a cohort of patients receiving liver, heart, lung, pancreas, kidney, bone marrow or intestine transplants.

However, the psychiatric assessment is so essential to selection in some academic teaching hospitals that an indication of inability to comply or insufficient capacity to maintain a scarce resource (the organ) can eliminate a candidate for transplantation, and this usually means the patient will die. In such settings the psychiatrist holds as much power as the surgeon in deciding whether or not a patient will receive the life-saving liver. Random control trials of the psychiatric contribution to clinical decision making and patient outcome are essential in transplantation medicine, in which scarce resources mandate using vital organs in persons who can maintain them with the greatest likelihood of their and the organs' survival. The behavioral capacity of the organ recipient to comply with complex and demanding treatment protocols and the presence of a support system to assist in this compliance may be as important to the successful outcome of the transplantation as the optimal donor-recipient match itself. The additional cost of psychiatric assessment and treatment is a minimal part of the cost of a liver transplant (estimated to be $250,000 in New York City in 1997) if it demonstrably benefits patient outcome.

The ability to substitute one mental health specialist for another, *product substitution,* has not been studied under rigorous conditions in the medical setting or with patients with comorbidity. Random control trials comparing the ability of psychologists, social workers, clinical nurse specialists, primary care physicians, and gerontologists to diagnose and effect biological and psychological intervention strategies for biopsychosocial problems in patients with comorbidity, which may include psychotropic drug–medical drug interactions, end-organ dysfunction, and somatic and pharmacological effects on the CNS (including cortical-cognitive-affective functioning) have yet to be undertaken. As managed care replaces mental health specialists on the basis of least cost, it is compromised by a paucity of outcome

studies showing which mental health professionals optimally perform what functions. The C-L psychiatrist who manages comorbidity in the intensive care unit, with patients emerging from surgery and anesthesia who are on multiple drugs in strange and spectacular environments, and who may have impaired cognition, including orientation, from the admixture of all the above etiological agents and mechanisms, and so demand psychopharmacological intervention, needs the knowledge and skills of a highly trained physician. Without studies demonstrating that the C-L psychiatrist makes a significant difference in patient outcome, health maintenance organizations will use the least costly and lesser-skilled professional in their efforts to contain costs.

Finally, it will be important to conduct a random control trial comparing consultation (awaiting referral) with C-L psychiatry, in which the psychiatrist becomes a member of the medical-surgical team and makes rounds on all patients. Although there has been considerable debate about the virtue of these two forms of psychiatry, no study has compared their cost-benefit ratios or their effects on screening, diagnosing, treating, and outcome on a medical or surgical service. Studies such as the one on hip fracture in elderly persons and their length of hospital stay demonstrate clearly the superiority of liaison over pure consultation psychiatry, but random control trials comparing the two structures are needed not only to foster optimum patient care in the medical setting, but also to convince third-party payers of their gain from structural changes and intervention expenditures. Such studies would enhance the funding potential for C-L psychiatry as part of the team or hospital care centers, following the path of social work services who are paid from the hospital budget to be on every ward, clinic, and emergency service. This would ameliorate the current attrition of this subspecialty of psychiatry from lack of support from departments of psychiatry. It would promote autonomy, and allow the C-L psychiatrist to work at a more advanced level rather than waiting to be called by the consultee and being funded (often marginally) by third-party payers. It would also allow access to the entire hospital population to screen and triage psychiatric morbidity that cannot be referred to the consultant because it is not identified and therefore not treated by the primary care physician or a mental health specialist.

ETHICS AND ETHICS COMMITTEES

An increasing function of C-L psychiatrists is their participation in, and often chairing of, ethics committees in the general hospital setting. To this end the psychiatrist needs to be aware of basic ethical principles and theories and their application in the medical setting. The rules of autonomy, beneficence, distributive justice, and malfeasance are essential to ethical decision making and are recurrent themes encountered in the medically ill. Capacity to make decisions (e.g., respect for autonomy versus paternalistic medicine, truth telling), do no harm, prima nocere (e.g., not offering procedures that carry risk and offer little to no benefit), and the optimum use of scarce resources (e.g., what to do when two persons need the same liver to preserve their lives) are the common issues for of the C-L psychiatrist. The understanding of these dilemmas is coupled with the need in the medical setting to translate manifest content into latent meaning so that the physician can know what the patient really feels and wishes to communicate. Manifest statements by patients that they want to kill themselves may really be conscious expressions of a wish to not be abandoned, not be left in chronic pain, and not be sent to a hospice but rather be permitted to die at home. It is necessary to try to interpret patients' refusal for a procedure, chemotherapy, hospitalization, or surgical intervention.

> A painter having a one-man show on Madison Avenue heard command hallucinations that he must cut off his painting hand with an ax on the bumper of his car in a mall parking lot for sins he had committed on the astral plain some 50,000 years ago. The patient followed the admonitions of the internal voices, placed his belt around his severed wrist, put the hand in a paper bag, and brought it to the emergency room. He yelled at the doctors that if they touched him or tried to sew the hand back on or repair his arm in any way he would kill them, sever the hand again, injure it by smashing it against the wall, and the doctors would never operate again. Understandably, the surgeons were initially reluctant to begin the surgery. This was all manifest behavior and easily observed. However, the patient's presence in the emergency room with his severed hand was a latent statement to disregard his manifest verbal comment and not let him harm himself. When the patient was no longer psychotic he was most grateful that the doctors had not listened to him and had disregarded his manifest pleas to be left alone.

C-L psychiatrists need to add ethical decision-making skills to their psychological armamentarium to assist in clinical decision making that extends to moral and ethical reasoning. The psychiatrist does not judge behavior, but attempts to explain it both in psychological terms and within an ethical dimension that considers the rights and duties of both patients and their doctors. Ethical issues, such as whether to treat or not to treat; use of scarce resources; denying treatment on the basis of mental capacity (in contrast to denying autonomy on the basis of the legally derived decision of lack of competency); confidentiality; surrogate decision making; and evaluating the patient with regard to wanting to die, will only gain in importance as technological advances enhance the array of possible interventions and their risks and benefits. Western society values self-determination highly, and these conflicts—to tell, not to tell, to favor autonomy or bow to beneficence—are more current, and patients are expected to make their own choices. Critical reasoning may be mastered by working with medical ethicists, enrolling in courses, and reading cases that highlight the central problems in critical thinking. The synthesis of psychological and ethical decision-making knowledge and skills positions the C-L psychiatrist to assist patients and their caregivers with the ethical dilemmas their illness and its treatment engenders.

SUGGESTED CROSS-REFERENCES

The neural sciences are discussed in Chapter 1 and neuropsychiatry and behavioral neurology in Chapter 2. The role of epidemiology in psychiatry is detailed in Section 5.1. Information regarding the evaluation and workup of the patient is examined in Chapter 7. Specific attention is directed to Section 7.7 on medical assessment and laboratory testing in psychiatry. Cognitive disorders and mental disorders due to a general medical condition are discussed in Chapter 10, adjustment disorders in Chapter 23, anxiety disorders in Chapter 15, somatoform disorders in Chapter 16, mood disorders in Chapter 14, factitious disorders in Chapter 17, and psychological factors affecting medical condition (psychosomatic disorders) in Chapter 25. Biological therapies are discussed in Chapter 31 and psychotherapies in Chapter 30; of special interest is cognitive therapy, Section 30.6. Psychiatric problems in the medically ill elderly patient are discussed in Section 54.4k. The recurrent themes of the patient's competence and participation in decision making are discussed in Section 51.6b, forensic issues, and Section 54.1, legal issues in psychiatry.

SECTION REFERENCES

American Psychiatric Association: Practice guidelines for major affective disorder in adults. Am J Psychiatry *150*(Suppl):1, 1993.

Broadhead WE, Leon AC, Weissman MM, Barrett JE, Blacklow RS, Gilbert TT, Keller MB, Olfson M, Higgines ES: Development and validation of the SDDS-PC screen for multiple mental disorders in primary care. Arch Fam Med *4:*211, 1995.

Bronheim HE, Fulop G, Kunkel EJ, Muskin PR, Schindler BA, Yates WR, Shaw R, Steiner H, Stern TA, Stoudemire A: The Academy of Psychosomatic Medicine Practice Guidelines for Psychiatric Consultation in the General Medical Setting. Psychosomatics *39:*S8, 1998.

Chalmers I, Dickersin K, Chalmers TC: Getting to grips with Archie Cochrane's agenda. Br Med J *305:*786, 1992.

Cochrane Collaboration: *The Cochrane Database of Systematic Reviews*. British Medical Journal Publishing Group, London, 1995.

Cohen-Cole SA, Kaufman K: Major depression in physical illness: Diagnosis, prevalence and antidepressant treatment (A ten year review: 1982–1992). Depression *1:*181, 1993.

*Druss BG, Rohrbaugh RM, Rosenheck RA: Depressive symptoms and health costs in older medical patients. Am J Psychiatry *156:*477, 1999.

Field MJ, Lohr KN: *Guidelines for Clinical Practice: From development to use*. National Academy Press, Washington, DC, 1992.

*Fink P, Ewald H, Jensen J, Sorensen L, Engberg M, Holm M, Munk-Jorgensen P: Screening for somatization and hypochonriasis in primary care and neurological inpatients: A seven-item scale for hypochondriasis and somatization. J Psychosom Res *46:*261, 1999.

Frasure-Smith N, Lesperance F, Talajic M: Depression following myocardial infarction: Impact on 6 months survival. JAMA *270:*1819, 1993.

Fuller MG: More is less: Increasing access as a strategy for managing health care costs. Psychiatr Serv *46:*1015, 1995.

Hall CW, Wise MG: The clinical and financial burden of mood disorder: Cost and outcome. Psychosomatics *36*(Suppl):11, 1995.

Hammer JS, Strain JJ, Fulop G, Friedberg A: Operationalizing a bedside Pen Entry notebook clinical database system in consultation-liaison psychiatry. Gen Hosp Psychiatry *17:*165, 1995.

Hasan S, Buckley P: Novel antipsychotics and the neuroleptic malignant syndrome: A review and critique. Am J Psychiatry *155:*1113, 1998.

Hays RD, Wells KB, Sherbourne CD, Rogers W, Spritzer K: Functioning and well-being outcomes of patients with depression compared with chronic general medical illness. Arch Gen Psychiatry

*52:*11, 1995.

Johnson J, Weissman M, Klerman GL: Service utilization and social morbidity associated with depressive symptoms in the community. JAMA *267:*1478, 1992.

Katon W, Gonzales J: A review of randomized trials of psychiatric consultation-liaison studies in primary care. Psychosomatics *35:*268, 1994.

Katon W, VonKorff M, Lin E, Walker E, Simon G, Bush T, Robinson P, Russo J: Collaborative management to achieve treatment guidelines: Impact on depression in primary care. JAMA *273:*1026, 1995.

Katon W, VonKorff M, Lin E, Walker E, Simon G, Bush T, Robinson P, Russo J: Distressed high utilizers of medical care: DSM-III-R diagnosis and treatment needs. Gen Hosp Psychiatry *12:*355, 1990.

Kornfeld DS, Youngner SJ, Steinberg MD, Powell T, Lederberg MS: Clinical ethics series. Psychosomatics *38:*307, 1997.

Levenson JL: Psychosocial intervention in chronic medical illness: An overview of outcome research. Gen Hosp Psychiatry *14:*43S, 1992.

Lin E, VonKorff M, Katon W, Bush T, Simon GE, Walker E, Robinson P: Primary care physician behavior and patient's adherence to antidepressant therapy. Med Care *33:*67, 1995.

Moros D, Rhodes R, Baumrin S, Strain JJ: Thinking critically in medicine and its ethics: Relating applied science and applied ethics. J Appl Philos *4:*229, 1987.

Morris PLF, Robinson RG, Andrzejewski P, Samuels J, Price TR: Association of depression with 10-year poststroke mortality. Am J Psychiatry *150:*124, 1993.

Piemme TE: Computer-assisted learning and evaluation in medicine. JAMA *260:*367, 1988.

Regier DA, Narrow WE, Rae DS, Manderscherd RW, Locke BZ, Goodwin FK: The de facto U.S. mental and addictive disorders service systems: Epidemiologic Catchment Area prospective 1-year prevalence rates of disorders and services. Arch Gen Psychiatry *50:*85, 1993.

Rost K, Kashner TM, Smith GR: Effectiveness of psychiatric intervention with somatization disorder patients: Improved outcomes at reduced costs. Gen Hosp Psychiatry *16:*381, 1994.

Rush AJ, Golden WE, Hall GW: *Depression in Primary Care: Clinical Practice Guidelines: Agency for Health Care Policy and Research*. AHCPR publication no. 93-0550. U.S. Department of Health and Human Services, Rockville, MD, 1993.

Saravay SM, Lavin M: Psychiatric comorbidity and length of stay in the general hospital: A critical review of outcome studies. Psychosomatics *35:*233, 1994.

Shortliffe EH, Perreault LE, Fagan LM, Wiederholds G, editors: *Medical Informatics Computer-Applications in Health Care*. Addison-Wesley, Reading, MA, 1990.

Smith GC, Strain JJ, Hammer JS, Wallack JL, Bialer PA, Schleifer SS, McKenzie DP: Organic mental disorders in the consultation-liaison psychiatry setting: A multi-site study. Psychosomatics *38:*363, 1997.

Smith GR Jr: The course of somatization and its effects on utilization of health care resources. Psychosomatics *35:*263, 1994.

Smith GR Jr, Monson RA, Ray DC: Psychiatric consultation in somatization disorder: A randomized controlled study. N Engl J Med *314:*1407, 1986.

Smith GR Jr, Rost K, Kashner TM: A trial of the effect of a standardized consultation on health outcome and costs in somatizing patients. Arch Gen Psychiatry *52:*238, 1995.

Spiegel D, Bloom JR, Kraemer HC, Goltheil E: Effects of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet *2:*888, 1989.

Spitzer RL, Williams JBW, Kroenke K, Linzer M, de Gruy FV III, Hahn SR, Brody D, Johnson JG: Utility of a new procedure for diagnosing mental disorders in primary care: The PRIME-MD 1000 study. JAMA *272:*1749, 1994.

Stoudemire A, Brown TM: *Psychiatric Side Effects of Prescription and Over the Counter Medications*. American Psychiatric Press, Washington, DC, 1998.

*Strain JJ, Caliendo G, Himelein C: Drug-psychotropic drug interactions and end organ dysfunction: Clinical management recommendations, selected bibliography, and updating strategies. Gen Hosp Psychiatry *18:*300, 1996.

*Strain JJ, Hammer JS, Fulop G: APM task force on psychosocial interventions in the general hospital inpatient setting. Psychosomatics *35:*253, 1994.

Strain JJ, Hammer JS, Himelein C, Caliendo G, Mayou R, Smith GC, Malt U, Lyons J, Kurosawa H: Further evolution of a literature database: Content and software. Gen Hosp Psychiatry *18:*294, 1996.

*Strain JJ, Pincus HA, Gise LH, Houpt J: Models of mental health training in primary care. Psychosom Med *47:*95, 1985.

Strain JJ, Lyons JS, Hammer JS, Fahs M, Lebovits A, Poddison PL, Snyder S, Strauss E, Burton R, Nuber G, Abernathy T, Sacks H, Noralic J, Sacks C: Cost offset from a psychiatric consultation-liaison intervention with elderly hip fracture patients. Am J Psychiatry *148:*1044, 1991.

Thompson D, Hylan TR, McMullen W, Romeis ME, Buesching D, Oster G: Predictors of a medical-offset effect among patients receiving antidepressant therapy. Am J Psychiatry *155:*824, 1998.

Way BB, Allen MH, Mumpower JL, Stewart TR, Banks SM: Interrater agreement among psychiatrists in psychiatric emergency assessments. Am J Psychiatry *155:*1423, 1998.

Wells KB: Cost containment and mental health outcomes: Experiences from U.S. studies. Br J Psychiatry *27:*43, 1995.

CHAPTER 26. RELATIONAL PROBLEMS

JOHAN M. F. VERHULST, M.D.

A person's psychological health and sense of well-being depends to a significant degree upon the quality of one's relationships, that is, upon patterns of interaction with one's partner and children, parents and siblings, friends and colleagues. Close relationships are relied upon for personal growth, need fulfillment, and support. Problems in the interaction between significant others may lead to clinical symptoms and impaired functioning among one or more members of the relational unit. Relational problems can also exacerbate the course or complicate the management of a mental or general medical condition. Conversely, an individual's medical or psychiatric condition may provoke dysfunctional patterns of interaction with significant others. Relational problems are often chronic and progressive, with increasing impairment in the functioning of the couple or family unit itself.

Relationships have been the object of a growing number of research studies. Certain aspects of relational functioning have been extensively investigated. The relative frequency and patterning of the interactions between intimate partners have been analyzed. Approach-avoidance, problem solving, communication deviance, behavioral control, and coercive processes have been studied. There is an extensive literature on sexual interactions and on gender issues in relationships. Cognitive elements such as interpersonal attributions and schemas have been studied, as well as the expression and regulation of emotional tension between partners and in the family. A remarkable continuity of attachment patterns in adult relationships has been demonstrated. Researchers have explored the development of partner relationships, differences between distressed and nondistressed couples, and predictors of divorce. Attempts have been made to classify relationships on the basis of resources exchanged, rules established, roles played by the participants, and styles used in conflict resolution. Unfortunately, few investigators have developed a culturally oriented understanding of observed interaction patterns, and the research findings are generally not applicable across cultures. Furthermore, the field lacks a coherent unifying theoretical framework within which the various lines of investigation can be integrated. Many important theoretical contributions developed in the clinical practice of couples and family therapy have not yet been empirically tested.

Relational problems may be a focus of clinical attention (1) when a relational unit is distressed and dysfunctional or threatened with dissolution and (2) when the relational problems precede, accompany, or follow other mental or general medical disorders. Indeed, medical or psychiatric symptoms are influenced by the relational context of the patient. Conversely, the functioning of a relational unit is affected by a member's medical or psychiatric illness. Relational disorders require a different clinical stance than other disorders. Instead of focusing on the link between symptoms and the workings of the individual mind, the clinician has to focus on interactions between people and how these interactions are related to the medical or psychological symptoms in a meaningful way.

DEFINITION

According to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), relational problems are patterns of interaction between members of a relational unit that are associated with symptoms or significant impairment in functioning in one or more individual members, or with significant impairment in the functioning of the relational unit itself. DSM-IV distinguishes five categories of relational problems. The first category, relational problem related to a mental or general medical condition, deals with the association between relationships and health. The other categories focus on problems in specific relational units: parent-child relational problem, partner relational problem, sibling relational problem, and relational problem not otherwise specified. In addition, DSM-IV features in appendix B the Global Assessment of Relational Functioning (GARF) Scale (see Table 7.8-4), which allows the clinician to rate the degree to which a relational unit meets the needs of its members.

ETIOLOGY

People communicate not only about personal issues and external events but also about their relationship. Interpersonal problems are generally related to interactions that deal with the various components and functions of the relationship itself. The topics of these interactions (i.e., what they are about) are classified in broad categories. It is hypothesized that all close relationships must at least deal with issues of (1) attachment, (2) ranking order, status, and dominance, (3) autonomy and territorial control, (4) sexuality and gender, (5) boundaries and loyalties, and (6) cognitive interpretations of the world. The list is somewhat arbitrary and not exhaustive. Interestingly, except for cognitive interpretations, all other issues can also be observed in interactions among nonhuman primates.

Attachment A core issue in all close personal relationships is the establishment and regulation of the affiliative connection between the participants. In a typical attachment interaction, one person seeks more proximity and affection and the other either reciprocates, rejects, or disqualifies the request. A pattern is shaped through repeated exchanges. Attachment behavior between an infant and its primary caregiver will lead to the development of an *attachment style*, which is a relatively stable communication pattern exhibited in close relationships. Distinct attachment styles have been observed in children and adults. Adults with an *anxious-ambivalent* attachment style tend to be obsessed with romantic partners, suffer from extreme jealousy, and have a high divorce rate. People with an *avoidant* attachment style are relatively uninvested in close relationships, though they often feel lonely. They seem afraid of intimacy and tend to withdraw when there is stress or conflict in the relationship. Break-up rates are high. People with a *secure* attachment style are highly invested in relationships and tend to behave without much possessiveness or fear of rejection.

Ranking Order Interactions that center on the acquisition of status and dominance (i.e., on decision-making power and control over the partner) can be called ranking-order interactions. Though the bases of interpersonal influence are diverse and difficult to specify, power and dominance are always an issue in close relationships. People tend to resent strategies of influencing to which they are particularly susceptible. Some partners are sensitive to threats of withdrawal of affection, others to guilt induction, and so on. Functional relationships require a balance of power that is satisfactory to the participants and facilitates conflict resolution rather than leading to interactional gridlock. Acute power struggles usually involve a (perceived) challenge to an existing dominance structure. An example of an especially dysfunctional exertion of power is the verbal or physical assault of a child with supposedly noncompliant behavior or of a spouse perceived as being insufficiently obsequious. Violence is less likely to occur in egalitarian relationships.

Territory People develop a sense of ownership over their own bodies, their personal space and possessions, their areas of responsibility, and the privacy of their minds. Territorial interactions deal with personal autonomy and control over resources and center on the acquisition, management, and defense of ownership rights. Effective parenting requires a gradual transfer of territorial control from the parents to the increasingly autonomous child. Especially important is the progressive validation of children's sense of ownership and control over their bodies (who can touch? who controls food intake?). Issues of personal autonomy often come to a head in adolescence. Many conflicts in couples are about territorial issues such as privacy, control over one's time, responsibility for tasks and chores, and financial control. The person who performs the quality control over an area is considered the owner of the territory, which is why comments, advice, and criticism are often resented. In healthy relationships each partner is able to relinquish autonomy and territorial control at times of stress and illness. Problems arise when such dependency is not tolerated, but also when it is systematically encouraged and maintained. Contentious territorial interactions are also common in the work environment when a person's area of responsibility is expanded to an unmanageable level or when a competing colleague threatens it.

Sexuality and Gender Communications that express gender role expectations or signal the presence or the absence of sexual attraction are a part of most close relationships. Sexual interactions center on the expression of sexual interest and on the management of sexual tension between potential partners. Cultural norms and individual maturity dictate the relationships in which sexual interactions are allowed and those that must be kept desexualized. Clearly, parents and older adults carry the responsibility of neutralizing the sexual tension that can arise in the relationship with children. Many sexual partner relationships are challenged by differences in desire (with women's sexual drive being less insistent than men's) and by the task of synchronizing individual sexual rhythms and maintaining sexual tension in the relationship.

Boundaries and Loyalties Relationships are constantly influenced by other people, by external pressures, and by each partner's outside activities and experiences.

A marriage can be undermined by an adulterous relationship as well as by excessive jealousy about a partner's emotional involvement with an outside person or activity. Loyalty conflicts involving one's spouse and one's family of origin are common in the beginning of a marriage. Work is often experienced as competing with the relationship. Even a medical condition can be seen as a rival when the patient seems to prefer the emotional involvement with symptoms and treatments to involvement with the partner. Every close relationship has to define its boundaries (e.g., the acceptable degree of outside interference as well as the acceptable degree of each partner's involvement with outside elements).

Cognitive Interpretations Humans live in a symbolic world, the world of language, ideas, values, and goals. A major socialization task of parents is transmitting this symbolic world to their children. One's sense of belonging depends upon the social validation of one's values and interpretations of the world. One's sense of identity is, to a large degree, a summary description of how one has been defined by others. Differences in worldview, values, and beliefs tend to cause interpersonal conflicts. Differences in how a conflict is interpreted by the participants tend to maintain it. A psychological mechanism that has consistently been found to create problems is that people tend to interpret their own behavior as *state-dependent*, (a normal response to the circumstances) and their partner's behavior as *trait-dependent* (determined by stable character traits). Another consistent finding is that of gender differences in thinking about relationships. Women think about relationships more and with more complexity than men do. They are more distressed about problems and more likely to take action to resolve them. Conceivably, women carry more than an equal share of the responsibility for relationships because they are defined by society as the relationships experts.

These interactions can take place without the participants paying conscious attention to them. They often consist of an exchange of nonlinguistic signals while the overt communication is focusing on something completely different and relationship neutral. For instance, while discussing vacation plans, a couple may use tone of voice and nonverbal messages to also exchange information about who is going to be in charge. These subtle exchanges may confirm or challenge the existing power structure. If enough tension develops, the issue may be brought to the forefront and discussed openly.

It is useful to distinguish between *conflicts of interest* and *conflicts of meaning.* In conflicts of interest the partners agree on what the interaction is about, but each one pursues a different outcome. In conflicts of meaning, the partners see the interaction from a different perspective and give it a different meaning. For instance, one partner may have a need for intimate conversation; the other partner has had a rough day and wants to be left alone. The first partner may experience the other's reluctance to talk as a refusal to engage in intimacy, while that partner may experience the insistence to talk as an insensitive invasion of privacy. Such misunderstanding can escalate and lead to distorted global interpretations of the partner and the relationship.

RELATIONAL PROBLEM RELATED TO A MENTAL DISORDER OR GENERAL MEDICAL CONDITION

According to DSM-IV:

> This category should be used when the focus of clinical attention is a pattern of impaired interaction associated with a mental disorder or a general medical condition in a family member.

Studies indicate that satisfying relationships may have a health-protective influence and that relationship distress tends to be associated with an increased incidence of illness. The influence of relational systems on health has been explained through psychophysiological mechanisms that link the intense emotions generated in human attachment systems to vascular reactivity and immune processes. Stress-related psychological or physical symptoms can be an expression of family dysfunction.

A person who feels ill will first turn to the members of the family. Their opinions will influence whether a person sees the symptoms as trivial or as a cause for concern, as requiring home remedies or professional attention. The family also participates, directly or indirectly, in the ongoing discourse between patient and physician, which defines the type of medical problem the patient is suffering from, how it should be dealt with, and what the family's role should be in managing the illness and supporting the patient. To a large extent, the psychological meaning of the illness and the patient's cooperation and compliance with treatment depend on the outcome of these interactions. Family dynamics influence an individual's course of illness. Conversely, the illness of a family member influences the family's interactional dynamics.

If the disease process is chronic, an adjustment of family members' roles and responsibilities is likely to be needed. The patient's identification with the sick role and the concomitant regression and disability are strongly influenced by family dynamics. In some cases the family joins with the patient in a struggle against physicians and others who appear to refuse to acknowledge the patient's disability and deny financial benefits. Sometimes medical symptoms are recruited to play a meaningful role in the interactional dynamics of the relational unit (e.g., stabilize the patient's family, regulate closeness and distance among family members, or communicate information that cannot be openly expressed).

The role of marital and family interactions in the onset and course of psychiatric disorders has been extensively studied, in particular for mood and anxiety disorders, schizophrenia, substance-related disorders, eating disorders, and personality disorders. Attempts were made to link distinct family dynamics to specific disorders. On the other hand, the following outline provides a somewhat simplified and generalized summary of the observed interactions.

At first, the psychiatric symptoms displayed by a family member elicit concern and care. The roles within the family are adjusted, often with other family members taking charge and assuming some of the patient's duties and responsibilities. Family members define the patient's problem, give advice, propose solutions, and encourage the patient to overcome the difficulties, seek help, or both. When the patient fails to improve, however, the family members may feel increasingly frustrated. Indeed, psychiatric symptoms are often hard to live with, and they can undermine the family's energy and morale. The mentally ill family member often behaves in a way that discourages social interaction and increases interpersonal conflict. The extra efforts and tasks required of the other family members become more burdensome over time. In addition, they may hold psychiatric patients more responsible for their symptoms than they would medical patients. As frustration mounts, the family members have limited options for dealing with their feelings. Direct expression of anger, disapproval, and criticism is likely to aggravate the psychiatric condition. On the other hand, if family members withdraw and exclude the patient from activities, the patient may feel neglected and abandoned. Finally, family members may feel guilty about their feelings of frustration and try to compensate for this with self-sacrifice, overprotection, or overidentification with the patient. Such overcompensation indirectly expresses the underlying hostility.

These understandable family adaptations represent behaviors that are rated high on scales for hostility, criticism, and overinvolvement, a constellation that has come to be known as “expressed emotion.” Expressed emotion is a significant and robust predictor of relapse in several psychiatric conditions, including schizophrenia, mood disorders, and eating disorders. High expressed emotion might also be associated with a worsening of the course of Alzheimer's disease and may even play a role in medical conditions such as diabetes. The interactional mechanisms that account for the different expressed emotion response styles of family members are unknown.

The relationship between family interaction patterns and psychiatric illness is likely to be bidirectional. There is some suggestion that when a vulnerable family member feels excluded or rejected and treated with criticism and overinvolvement, psychiatric symptoms may develop. For instance, the risk of developing depression has been found to be 10 times greater for individuals experiencing relationship distress than the risk for individuals in nondistressed relationships.

The mechanisms described above do not support the notion that families cause mental illness or relapse. Indeed, neither the research data nor the complex and recursive nature of family interaction patterns warrants such a conclusion. And yet mental health professions have in the past either neglected the families of patients in the name of confidentiality and therapeutic alliance or blamed them for causing the illness. The clinician must appreciate the dedication of most families, their need to be involved in treatment decisions that affect them, and the fact that they are entitled to assistance with the considerable burden of family caregiving. Indeed, families are the primary caregivers of the sick and the disabled, and it is estimated that half to two-thirds of the persistently mentally ill population live with families. Caregiving is stressful, and the burden of caring for ill family members falls disproportionately on the women in a family. Family members who live with mental illness report high levels of stress, anxiety, depression, and resentment. They cite increased marital strain, disrupted social life, and hardship for siblings. In recent years, research into the stress of caregiving and how to alleviate it has been initiated, partly in response to the advocacy of support organizations such as the National Alliance for the Mentally Ill.

PARENT-CHILD RELATIONAL PROBLEM

According to DSM-IV:

> This category should be used when the focus of clinical attention is a pattern of interaction between parent and child (e.g., impaired communication,

> overprotection, inadequate discipline) that is associated with clinically significant impairment in individual or family functioning or the development of clinically significant symptoms in parent or child.

Parents differ widely in sensing the needs of their infants. Some quickly note their child's moods and needs, others are slow to respond. Parental responsiveness in interaction with the infant's temperament affects the quality of the attachment between infant and parent. A considerable body of research links an infant's attachment style with social and emotional adjustment in childhood. *Securely* attached children showed enhanced resilience in a high-stress environment. *Insecure* children are more likely to be anxious, aggressive, and low on social competence. Furthermore, attachment-related factors such as maternal deprivation, major separations, and loss of attachment figures and the development of disorganized-insecure attachment in response to maltreatment have been found to be risk factors in the development of psychiatric disorders.

Research on parenting skills has isolated two major dimensions: (1) a permissive-restrictive dimension and (2) a warm-and-accepting versus cold-and-hostile dimension. A typology that separates parents on these dimensions distinguishes between *authoritarian* (restrictive and cold), *permissive* (minimally restrictive and accepting), and *authoritative* (restrictive as needed but also warm and accepting) parenting styles. The children of authoritarian parents tend to be withdrawn or conflicted; those of permissive parents are likely to be more aggressive, impulsive, and low achievers; the children of authoritative parents seem to be functioning at the highest level, both socially and cognitively. Switching from an authoritarian to a permissive mode may create a negative reinforcement pattern. This interactional pattern starts with increasingly aversive child behavior, such as whining or aggressive hyperactivity, in response to insistent parental demands to “stop it.” Then, the parent backs down, which rewards the child for behaving aversively. At this point the child deescalates, thus reinforcing the parent's permissive attitude. Such coercive processes have been associated with a child's lack of self-regulation, with rejection by peers, later academic failure, depression, delinquency, and antisocial behavior.

Substantial evidence indicates that marital discord leads to problems in children, from depression and withdrawal to conduct disorder and poor performance at school. This negative effect may be partly mediated through *triangulation* of the parent-child relationships. Triangulation refers to the process in which conflicted parents attempt to win the sympathy and support of their child, who is recruited as an ally in the struggle with the partner. The term has also been used for the process by which parents focus on a behavioral or general health problem of the child as a strategy for defusing marital conflicts. Marital discord can also affect the parent-child relationship in a direct way, as negative emotions begin to contaminate all family relationships. For instance, angry interactions between spouses have been linked to negative affect in the relationship between father and infant.

The parent-child relationship is greatly influenced by the marital and family environment. Divorce and remarriage stress the parent-child relationship and may create painful loyalty conflicts. Stepparents often find it difficult to assume a parental role and may resent the special relationship that exists between the new marital partner and the children from that partner's previous marriage. Single-parent families usually consist of a mother and children, and their relationship is often affected by financial and emotional problems. Women who work and pursue careers often feel guilty about not having enough time with their children, yet they are likely to spend much more time in child-rearing activities than fathers.

Normal developmental crises can also be related to parent-child problems. For instance, adolescence is a time of frequent conflict, as the adolescent resists rules and demands increasing autonomy, while at the same time eliciting protective control by displaying immature and dangerous behavior.

PARTNER RELATIONAL PROBLEM

According to DSM-IV:

> This category should be used when the focus of clinical attention is a pattern of interaction between spouses or partners characterized by negative communication (e.g. criticisms), distorted communication (e.g. unrealistic expectations), or noncommunication (e.g. withdrawal) that is associated with clinically significant impairment in individual or family functioning or the development of symptoms in one or both partners.

A partner relationship involves two people who have made a commitment to maintain their relationship. Usually the partners live together and have sexual relations with one another. Many partner relationships are socially sanctioned as marriage. Separation and divorce rates have increased rapidly in Western societies. An unexpected announcement by one spouse of a decision to divorce often elicits an acute relational crisis in which the partners desperately seek professional help. Sociologists have found that increases in the divorce rate correlate with measures of women's economic and psychological independence. A suggested explanation is that women are less willing to stay in unhappy marriages to the degree that they are no longer forced to do so for economic reasons. Nevertheless, for most women, divorce still has severe economic consequences.

Using observational methods, John Gottman has identified criticism, contempt, defensiveness, and stonewalling (e.g., withdrawal from interaction) as interactive processes that longitudinally predict which couples will separate and divorce. The best single predictor is contempt, especially the wife's contempt. There are gender differences; women tend to show more criticism than men do, and men show more stonewalling. A cascade has been proposed in which complaining and criticizing leads to contempt, which leads to defensiveness, which leads to withdrawal from the interaction. These same patterns probably occur also in gay and lesbian relationships, although comparable data have not been reported.

Relatively minor problems can escalate to dysfunctional levels as a result of natural interactional processes such as the *amplification of differences* and the *polarization of ambivalence.* Small differences between partners, such as a subtle difference in sexual need, can amplify spontaneously. Gradually, the partner with the greater need may become the only sexual initiator, overly focusing on the other's sexual availability, while the latter is likely to adopt a style of avoiding sex and focusing on the first one's sexual aggressiveness. This preoccupation with one another's sexual desire detracts from the awareness of a person's own real needs, as one person seems to always be “in the mood” and the other never. In a spontaneous process of polarization, each partner expresses an opposite position with regard to an issue about which they both feel ambivalent. An internal dilemma is thus externalized in the relationship. The polarization of a couple's ambivalence about commitment or about the decision to have children can lead to chronic emotional conflict.

Many partners expect the other to regularly provide proof of commitment. This may result from insecurity or concerns about the power balance in the relationship. Indeed, the person who is more committed to and dependent on the relationship may feel powerless in the face of threats of divorce. Violent behavior can be an attempt to deny dependency and exert power. Being rejected by an intimate partner can elicit traumatic memories of past attachment experiences and lead to impulsive suicide attempts or to violence toward the partner.

An issue with special importance for partner relational problems is *passionate love*, defined as a special state of mind characterized by emotional dependence on the person one is in love with and by symptoms such as an intrusive preoccupation with that person, an intense desire for reciprocity of feelings, idealization, and an uncanny ability to see hope even when there is none. “Falling in love” can occur at any age and has been described in all cultures. Unrequited love can lead to depression and suicidality. Distress and alienation in one's primary relationship puts one more at risk for falling in love with someone else.

Violence can be associated with relational problems. A pattern of wife beating, however, indicates individual impulse-control problems, impaired socialization, and usually alcohol and substance abuse. Wife beating is found in every class and at every income level. Pregnancy is an extremely high risk time for domestic violence toward the woman. Being subjected to violence often results in depression and learned helplessness. The patient may feel inhibited about divulging the secret of domestic violence; thus clinicians must develop an appropriate sensitivity to the possibility of spousal abuse.

SIBLING RELATIONAL PROBLEM

According to DSM-IV:

> This category should be used when the focus of clinical attention is a pattern of interaction between siblings associated with clinically significant impairment of family functioning or symptoms in one or more of the siblings.

Sibling relationships tend to be characterized by competition, comparison and cooperation. Intense sibling rivalry can occur with the birth of a child and may persist as the children grow up, compete for parental approval, and measure their accomplishments against one another. Alliances between siblings are equally common. Siblings may learn to protect one another against parental control or aggression. In households with three children, one pair tends to become closely involved with

one another, leaving the extra child in the position of outsider.

Parents tend to be more empathetic with the child who has the same sibling position that they experienced. Relational problems can arise when siblings are not treated equally; for instance, when one child is being idealized while another is cast in the role of the family scapegoat. Differences in gender roles and expectations expressed by the parents can underlie sibling rivalry. The notion that a child's resentment directed at a parental figure or a child's own disavowed dark emotions can be projected onto a sibling and fuel an intense hate relationship is an interesting psychodynamic hypothesis. As people grow older they often experience a desire to reconnect with brothers and sisters, and they may take great effort to resolve long-standing sibling conflicts.

A child's medical or psychiatric condition always stresses the sibling relationships. Parental concern and attention can elicit envy in the siblings. Chronic disability can leave the sick child feeling devalued and rejected by siblings, who may develop a sense of superiority and feel embarrassed about having a disabled brother or sister.

RELATIONAL PROBLEM NOT OTHERWISE SPECIFIED

According to DSM-IV:

> This category should be used when the focus of clinical attention is on relational problems not classifiable by any of the specific problems above (e.g. difficulties with coworkers).

People become emotionally involved in peer relationships at school or at work, in relationships with teachers, superiors, students, and employees. They develop intensive friendships and engage in complex interactions around avocational interests and goals. In such relationships, conflicts are common and can bring about stress-related symptoms.

Many relational problems of children occur in the school setting and involve peers. Impaired peer relationships can be the chief complaint in attention-deficit or conduct disorders, as well as in depressive and other psychiatric disorders of childhood and adolescence.

Racial, ethnic, and religious prejudices cause problems in interpersonal relationships. In the workplace, sexual harassment is often a combination of inappropriate sexual interactions, displays of power and dominance, and expressions of negative gender stereotypes.

SUGGESTED CROSS-REFERENCES

Child psychiatry is the subject of Chapter 32. Couples and family therapy are discussed in Section 30.5. Child physical and sexual abuse and neglect is dealt with in Section 49.4. Physical and sexual abuse of adults is discussed in Section 28.7. Chapter 19.1a deals with normal human sexuality and sexual disorders.

CHAPTER REFERENCES

Bakker C, Bakker-Rabdau MK: *No Trespassing: Explorations in Human Territoriality.* Chandler and Sharp, San Francisco, 1973.

Bank SP, Kahn MD: *The Sibling Bond.* Basic Books, New York, 1982.

Baumrind D, Black AE: Socialization practices associated with dimensions of competence in preschool boys and girls. Child Dev *38:*291, 1967.

*Borge L, Martinsen E, Ruud T, Watne O, Friis S: Quality of life, loneliness, and lucid contact among long-term psychiatric patients. Psychiatr Serv *50:*81, 1999.

Burman B, Margolin G: Analysis of the association between marital relationships and health problems: An interactional perspective. Psychol Bull *112:*39, 1992.

*Butzlaff RA, Hooley JM: Expressed emotion and psychiatric relapse. Arch Gen Psychiatry *55:*547, 1998.

Dunn J, Kendrick C: *Siblings.* Harvard University Press, Cambridge, MA, 1982.

Fletcher GJO, Fitness J, editors: *Knowledge Structures in Close Relationships: A Social Psychological Approach.* Erlbaum, Mahwah, NJ, 1996.

*Gottman JM: *The Marriage Clinic: A Scientifically Based Marital Therapy.* Norton, New York, 1999.

*Gottman JM: *What Predicts Divorce? The Relationship Between Marital Processes and Marital Outcomes.* Erlbaum, Hillsdale, NJ, 1994.

Gottman JM, Katz LF, Hooven C: *Meta-Emotion: How Families Communicate Emotionally.* Erlbaum, Mahwah, NJ, 1997.

Hazan C, Shaver PR: Attachment as an organizational framework for research on close relationships. Psychol Inquiry *5:*1, 1994.

Hinde RA: *Relationships: A Dialectical Perspective.* Psychology Press, Hove, UK, 1997.

*Kaslow FW, editor: *Handbook of Relational Diagnosis and Dysfunctional Family Patterns.* Wiley, New York, 1996.

Lefley HP, Johnson DL, editors: *Families as Allies in Treatment of the Mentally Ill: New Directions for Mental Health Professionals.* American Psychiatric Press, Washington, DC, 1990.

*Lewis JM: For better or worse: Interpersonal relationships and individual outcome. Am J Psychiatry *115:*582, 1998.

Main M: Introduction to the special section on attachment and psychopathology: Overview of the field of attachment. J Consult Clin Psychol *64:*237, 1996.

O'Leary KD, Christian JL, Mendell NR: A closer look at the link between marital discord and depressive symptomatology. J Soc Clin Psychol *13:*33, 1994.

Patterson GR: *Coerceive Family Process.* Castalia, Eugene, OR, 1982.

Patterson GR, Stoolmiller M: Replication of a dual failure model for boys' depressed mood. J Consult Clin Psychol *59:*49, 1991.

Perlmutter RA: *A Family Approach to Psychiatric Disorders.* American Psychiatric Press, Washington, DC, 1996.

Tennov D: *Love and Limerence: The Experience of Being in Love.* Stein and Day, New York, 1979.

Tseung W, Hsu J: *Culture and Family: Problems and Therapy.* Haworth Press, New York, 1991.

Verhulst J: Limerence: Notes on the nature and function of passionate love. Psychoanal Contemp Thought *7:*115, 1984.

Verhulst J, Heiman J: A systems perspective on sexual desire. In *Sexual Desire Disorder*s, SR Leiblum, RC Rosen, editors. Guilford, New York, 1988.

Vitaliano PP, Young H, Russo J, Romano J, Magana-Amato A: Does expressed emotion in spouses predict subsequent problems among care recipients with Alzheimer's disease? J Gerontol *48:*202, 1993.

CHAPTER 27. ADDITIONAL CONDITIONS THAT MAY BE A FOCUS OF CLINICAL ATTENTION

27.1 TREATMENT COMPLIANCE

BARRY BLACKWELL, M.D., F.R.C.PSYCH.

Although noncompliance with treatment is relegated in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), as a condition that may be the focus of clinical attention, no problem is more ubiquitous in practice or better illustrates the scope of psychiatry in understanding the biopsychosocial influences and interventions that determine treatment outcome. DSM-IV also indicates that our professional involvement may be related to the treatment of mental health and medical disorders directly as providers or indirectly through consultation. For this reason and because those who suffer from mental illness are people first, this review includes the entire medical domain as a background to understanding the specific problems of compliance in psychiatric disorders.

HISTORY

Interest in compliance has fluctuated through history with concerns about physician authority and patient autonomy. During the first half of the twentieth century, biomedical advances and technology rendered patients subservient to physicians. Moving toward the millennium, the social, political, and scientific zeitgeist shifted back in the direction of patient autonomy. Distrust of authoritarianism after World War II ushered in the civil, women's, and patient rights movements. The physician's unique prerogative to prescribe was eroded by other professions. Acute diseases were controlled by modern drugs, and chronic conditions predominated, raising concerns about risk reduction, disease prevention, and lifestyle change. This resulted in complex, long-term treatment regimens that created quality-of-life concerns for consumers.

In 1975 the term “patient dropout” was replaced by the word “compliance.” This coincided with an upsurge of research on the topic, and international conferences on compliance were held at McMaster University in 1974 and 1977. In the last quarter-century, over 12,000 publications have appeared, reaching a plateau in 1989 of about 800 articles a year. Almost half contain original data; the remainder are reviews.

Interest in compliance is likely to continue and increase for many reasons. Aging populations afflicted with chronic conditions are seeking longevity and improved quality of life. As more and more people (including those receiving Medicaid and Medicare) are placed in managed-care organizations, compliance will become a compelling social policy issue when the economics and quality of care are closely scrutinized.

New developments will sustain the field into the next century, including microchip circuitry and computer technology. A recent report in the *New England Journal of Medicine* argued that the Internet may become a valuable tool for transforming the delivery of medicine:

> Responsibility for decisions could be shared by the patient and physician. The patient could tap into authoritative medical data bases including textbooks and newsletters formulated expressly for lay audiences. If done right, the online computer system would function as a ‘virtual physician' in a new kind of house call.

DEFINITION

The most resilient definition of compliance is that adopted at the First International Conference, which states: “Compliance is the extent to which a person's behavior (in terms of taking medications, following diets or executing lifestyle change) coincides with medical or health advice.” This definition, which places the patient in a passive role, does not meet contemporary demands for a less subservient consumer. Soren Holm, an ethicist, noted: “It is not patients who should comply with their doctors' demands but doctors who should comply with their patients' informed and considered desires.”

The term *noncompliance* subsumes manifold behaviors, including reluctance to seek help, rejection of screening procedures, and failure to attend initial appointments. Later problems may include irregular attendance, premature termination, or failure to follow advice. These may involve diet, lifestyle, or drugs. Faulty compliance with medication may include errors of purpose, timing or dosage, partial or total omission, or use of inadvertent combinations.

Noncompliance with treatment is included in DSM-IV as a condition that may be a focus of clinical attention. DSM-IV states:

> This category can be used when the focus of clinical attention is noncompliance with an important aspect of the treatment for a mental disorder or a general medical condition. The reasons for noncompliance may include discomfort resulting from treatment (e.g., medication side effects), expense of treatment, decisions based on personal value judgments or religious or cultural beliefs about the advantages and disadvantages of the proposed treatment, maladaptive personality traits or coping styles (e.g., denial of illness), or the presence of a mental disorder (e.g., Schizophrenia, Avoidant Personality Disorder). This category should be used only when the problem is sufficiently severe to warrant independent clinical attention.

MEASUREMENT METHODS AND PROBLEMS

In the compliance field, simple measures are not accurate and accurate measures are not simple. Both patients and providers minimize the scope and understate the severity of the problem. Pill counts are unreliable, and biological markers are often only qualitative. New and more precise electronic methods are now available that use time-recording microcircuitry installed in pill containers or eye droppers. Only this last method circumvents a problem that pervades all compliance research—noncompliance disappears under scrutiny because people invariably modify their behavior to conform with the observer's expectations.

Compliance is often described in vague terms as “good, fair or bad,” but the most accurate descriptions use index measures that take into account receiving the most appropriate treatment in adequate amounts to achieve a defined outcome. Deviation from this ideal may be consistent, intermittent, or sporadic.

How much compliance is required also depends on the nature of the illness, the individual, and the treatment. For example, control of hypertension is often achieved with 80 percent compliance, but this level of adherence to a weight loss regimen usually results in failure. Even if a person is 100 percent compliant, a regimen may fail if the person takes an ineffective drug, metabolizes it slowly, excretes it rapidly, or is exposed to conditions that aggravate the disorder (e.g., stress).

One property of a medication that influences the outcome of noncompliance is “forgiveness.” This concept is illustrated by contrasting two psychotropic medications, alprazolam (Xanax) and fluoxetine (Prozac). Alprazolam is metabolized rapidly and induces its effect by immediate receptor binding. Its action is dramatic, and the effects of cessation are almost immediate. By contrast, fluoxetine is metabolized slowly and acts by delayed receptor regulation. Its effects are insidious, and the results of cessation are delayed and prolonged. A missed dose of alprazolam is noticed immediately, while an occasional missed dose of fluoxetine is immaterial.

OTHER TERMINOLOGY

Some cultures in which authority is seldom questioned lack words to convey the concept of compliance; more egalitarian societies consider the term coercive. One researcher noted the disfavor attached to the word compliance: “Because it conjures up images of patient or client sin and serfdom.” The alternative word “adherence” has been used but appears in less than 1 percent of articles published. Alvan Feinstein made the tongue-in-cheek observation that “adherence seems too sticky,

fidelity has too many connotations, and maintenance suggests a repair crew."

Irrespective of terminology, the concept of compliance has taken firm hold. Holm, notes: "New terminology will not be enough, a change of underlying attitudes is absolutely necessary." This change has begun; the slogan adopted by the National Council on Patient Information and Education is "communicate before you medicate." Further along still is the suggestion by B. A. Merrill that rehabilitation therapists like "successful salespeople know their product, learn the patient's dominant buying motive, combine facts with benefits and present arguments that fit the person's belief system and norms."

COST OF NONCOMPLIANCE

The costs of noncompliance can be computed in both social and individual terms. From a social perspective, there is always a balance between the cost incurred by the iatrogenic consequences of overuse of procedures or medications (an estimated 20 percent of health expenditures) and the economic effects of poor compliance, which have been estimated to exceed 100 billion dollars annually in lost work productivity, unnecessary hospitalizations, and premature deaths. A society that shuns preventive methods to check the spread of the human immunodeficiency virus (HIV) infection embraces genocide.

The individual costs of noncompliance were insignificant when most treatments were panaceas, but as therapy has become more effective, personal costs can be considerable. An individual denied a scarce kidney or heart transplant because of suspected compliance problems receives a death sentence. Recent research in a number of disorders has shown that patients who are compliant with placebo have better outcomes than noncompliant patients treated with an active drug. This is presumably due to the fact that people who are compliant with pill taking are more committed to all aspects of the treatment regimen (exercise, diet, and lifestyle). In another possible manifestation of this phenomenon, each of three different interventions was equally effective in the treatment of alcoholism if the type of treatment was matched to the needs and expectations of the patient; in other words, steps taken to ensure compliance with *any* intervention may be a more powerful determinant of outcome than differences between interventions.

At the same time, attempts to improve compliance are not without potential adverse consequences. Intervention at the work site to improve compliance with antihypertensive treatment increases absenteeism and decreases productivity, presumably because it increases the individual's sense of vulnerability and sick role behavior.

RESEARCH

Types Over 50 different diseases or conditions have been studied, and although no single disorder accounts for more than 10 percent of the output, the two most commonly studied areas throughout the past quarter-century have been pediatrics and psychiatry. Hypertension, smoking, asthma, diabetes, and screening programs are the other areas of common interest. Organ transplantation and HIV infection are of increasing contemporary concern while interest in seat belt use has declined.

All aspects of therapy have been studied including disease factors, referrals and appointments, treatment regimen, patient-provider interaction, and patient characteristics. What has been easiest to study has been most studied—disease and sociodemographic characteristics. Far more attention has been paid to patient characteristics and behaviors than to those of providers.

Attempts to study and understand compliance have employed five major models. The oldest and most basic is the biomedical model, which focuses on technical and mechanistic issues such as the treatment regimen. The second, the operant behavioral model, emphasizes structuring environmental rewards and teaching people specific skills. A third model is educational and aims to enhance information exchange between patient and provider. The fourth, particularly popular, is the health beliefs model based on an appraisal of the benefits and barriers to complying with treatment. The fifth and final model is the self-regulatory systems approach, which considers the patient's own models of illness and treatment and seeks to obtain congruence with providers in respect to action and coping behaviors.

Problems Research design in compliance has often lacked sophistication, and less than half of all studies have yielded positive findings. Psychosocial variables have tended to do so much more than demographic or disease characteristics. Even positive findings often lack generalizability. For example, factors that influence clinic attendance differ from those that determine medication taking. Even within a specific compliance behavior, no variable studied (e.g., marital status) has a universal or consistent impact on outcome.

Another problem is that most research is cross-sectional, while life is longitudinal. An individual's compliance may alter radically across the life cycle or between episodes of illness. A juvenile diabetic who is passively compliant with parental wishes during latency may rebel by rejecting treatment in adolescence. In early adulthood the rewards of a healthy lifestyle may be compelling, but in old age the burdens of treatment may exceed the benefits of longevity.

UNDERSTANDING COMPLIANCE

The determinants of compliance are kaleidoscopic and everchanging. There is no stereotypical noncompliant person or situation. Nonetheless, there are areas of consensus.

Noncompliance is more common and roughly double among outpatients than among inpatients. Medication compliance is more readily achieved than modification in lifestyle.

Compliance is a particular problem in disorders that have no symptoms, are persistent, and have no method for self-monitoring such as hypertension, diabetes, and glaucoma. Glaucoma presents the most difficult problem because currently there is no home tonometer to measure intraocular pressure corresponding to the sphygmometer and glucometer that measure blood pressure and blood sugar concentrations, respectively.

Compliance decays across time in disorders that are chronic or long lasting. Compliance generally (but not invariably) improves when patients' expectations are met, when they are satisfied, and when they are supervised. It helps if patients view their disease as a serious one to which they are susceptible and if they have developed compliance strategies of their own. Supportive family members or friends are important. Factors that generally (but not invariably) impede compliance are complicated regimens, troublesome adverse effects, social stress, isolation, and alcohol dependence.

Technological devices have been introduced recently to increase compliance such as special pillboxes, electronic alarms, and beepers that flash messages to take the prescribed medication. Another device is a pill-bottle cap fitted with a computer chip that records when the bottle is opened and closed. These are expensive and have not been tested adequately.

INFLUENCING COMPLIANCE

Two principles apply to choice of interventions. First, parsimony is important; some changes are more simple to make. Examples include modifying the treatment regimen, reducing waiting times, and combining written and verbal instructions. Other factors are more difficult to influence such as memory impairment or a person's attitudes and beliefs about illness, treatment, or quality of life.

Secondly, although single interventions may contribute to outcome, sustained success often requires a multimodal approach. Table 27.1-1 illustrates the types of intervention, intermediary goals, and final outcomes that are part of a truly comprehensive approach.

Intervention		Intermediary Goals	Final Outcome(s)
Education	Didactic Interactive	Knowledge Congruence	Composite compliance (to criterion) and Autonomy (at ideal)
Supervision	Professional Social	Appointment keeping	
Convenience	Regimen Access to Care	Medication taking	
Skills training	Tailoring Risk Reduction	Risk reduction	
Patient participation	Monitoring Self-Management	Active patient	

Table 27.1-1 Categorization of Interventions and Outcomes

The interventions can be conceptualized in three stages. An initial phase of information exchange (education) is followed by a negotiated alliance (supervised compliance) that consists of supervision, convenience, and skills training, leading to an ideal of autonomous self-care (active patient participation).

Education is necessary but seldom sufficient to sustain compliance. Excessive use of fear-provoking messages may also be counterproductive, because they cause anxiety and encourage avoidance. What works best is a negotiated alliance in which information is shared and congruence is reached between the provider's and the patient's beliefs about illness, treatment, and recovery. Some form of support or supervision is often essential from providers, family members, peer groups, or significant others. Although autonomy is an ideal, not everyone is eager to assume complete independence, and not all physicians are comfortable relinquishing control.

COMPLIANCE IN PSYCHIATRIC DISORDERS

The published rates of noncompliance are very similar for medical and psychiatric disorders, partly because it is such a multiply determined behavior that similarities and differences tend to average out in overall outcome. Nevertheless, these similarities and differences are worth considering in both understanding and influencing compliance in individual patients who suffer from a psychiatric disorder. Factors in the disorder, the patient, and the medications, influence compliance.

Disorders Most psychiatric disorders have two features that facilitate compliance: they are symptomatic in ways that diminish the patient's quality of life and they are episodic or intermittent. Unlike asymptomatic and persistent medical disorders (e.g., hypertension, diabetes, and glaucoma) these two features create an opportunity for the patient to learn ways in which treatment can decrease symptoms and improve overall quality of life. In addition to these features of illness, most psychiatric disorders have their peak prevalence early in life when there is less medical comorbidity, cognition is relatively intact, and social connections are preserved.

Each major mental disorder may also impose unique burdens on the individual's adaptation to treatment. Impaired concentration, poor memory, thought disorder, or anxiety all interfere with effective information exchange. Delusional ideas (particularly paranoid or grandiose types) can erode the trust necessary for a working alliance. Depression, apathy, and the negative symptoms of schizophrenia are barriers to the motivation and persistence necessary for long-term treatment. Lack of awareness or acceptance of illness (insight) is a particular problem in some patients with bipolar disorder or schizophrenia. Insight is a complex construct, difficult to define or measure. It is also misused at times to describe differences of opinion or underlying attitudes and beliefs that are part of the patient's personality or value system, not their illness.

Of great practical significance are the difficulties encountered by patients with extreme forms of symptom sensitivity (somatization disorder) or bodily preoccupation (conversion disorder). Some of these individuals suffer from panic disorders with episodic autonomic arousal that create fears about being allergic to medications. As a consequence, they are often attracted to homeopathic or alternative therapies.

Added to these problems is comorbidity with substance abuse. This is a risk factor for all forms of noncompliance but is a special problem when the effects of the substance may interfere with the therapeutic action of medication or may be used by the patient to stifle psychotic symptoms or alleviate depression.

Special problems are posed by people who have developmental disorders, since a crucial aspect of their treatment is encouraging independence and autonomy. This can conflict with simplistic notions about compliance.

Patient All long-term illness, medical or psychiatric, imposes generic burdens on those who suffer. These challenge the core human needs to feel in control, autonomous, intact, and connected to others. Severe mental illness arouses feelings of ambiguity, loss of integrity, dependence, and stigma or isolation. Irrespective of the specific disorder, these needs must be addressed before treatment recommendations of any kind can prove acceptable.

Like physicians, patients possess unique beliefs, attitudes, and experiences relating to illness, which may include its cause, treatment, and outcome. People with strong psychological or spiritual beliefs may resist medication. Patients in homeless clinics often report bruising experiences with uncaring or coercive providers that make them reluctant to accept help. Many people share a belief that the body sometimes deserves a rest from medication and as a result they take drug holidays.

Medication Psychoactive drugs possess several properties that may appear counterintuitive to the patient and so invite noncompliance. Because the brain is well protected, large amounts are required to pass the liver and penetrate the blood-brain barrier. The receptors with which drugs interact are also present elsewhere in the body so that adverse effects are virtually inevitable, with disturbances in cardiovascular, sexual, or gastrointestinal function that compromise quality of life. The newer, more selective compounds are less prone to this drawback, but the cost-benefit balance is always an issue.

Secondly, changes in brain chemistry often occur slowly because of up- or down-regulation of receptors. People are used to the immediate mind-altering effects of alcohol, nicotine, caffeine, and cocaine. The fact that therapeutic effects of psychotropic drugs are delayed (while many side effects are immediate) is an incentive to poor compliance.

Finally, the natural history of psychiatric disorders and their response to treatment may conflict with commonsense lay experience. In most acute illnesses people take medicine until they feel better and then stop. In major psychiatric disorders, treatment must continue beyond the restoration of normal function, often for many months. Because repeated episodes generate the likelihood of recurrence (kindling), long-term maintenance or prophylaxis is often desirable, but it may be resisted by those who fear dependency on medication.

Clinicians To facilitate compliance the treating psychiatrist must accomplish three objectives: developing an alliance, enabling a support system, and tailoring the regimen to fit the convenience and lifestyle of the patient.

Alliance A physician's need for control and instantaneous cure may confront a patient's desire for autonomy and independence. Although psychotic illness can sometimes diminish a person's decision-making capacity or ability to make informed choices, attempts to evoke fear or use coercion are seldom successful and usually elicit resistance or regression. Even when safety is an issue, the short-term protective benefits of coercion quickly dissipate over the long-term management of an illness. Coercion is not limited to obvious forms of involuntary confinement or outpatient commitment procedures; it can be a subtle and pervasive influence over many of the interactions between a provider and patient.

To negotiate a productive alliance the physician must explore the patient's knowledge, attitudes, and beliefs about the disorder and its treatment. These can often be accommodated without sacrificing the essential ingredients of effective treatment. Persons who believe in megavitamin therapy experience no harm if they also take traditional medicine.

Not every patient seeks or is willing at first to accept complete responsibility for coping with the burden of illness. The physician must explore resistance or regression, avoid an overly parental approach that may provoke these defenses, and discourage passive adoption of the sick role. The patient's ability to assume control may

also depend on the phase and severity of the illness. In acute episodes, psychotic symptoms, cognitive impairment, and poor impulse control are impediments. Later, apathy or negative symptoms may interfere. Moves toward autonomy should be gradual; an individual who initially receives monthly parenteral injections is slowly weaned toward daily or weekly attendance for oral medication. For the physician to relinquish control and encourage autonomy may involve calculated risks including the possibility of relapse or recurrence of dangerousness to self and others.

Patients must learn for themselves that medication stifles voices or impulses and must calculate the personal benefit of these improvements with the discomfort of side effects. The physician's advice and predictions are just that—inexact but (one hopes) well-informed guesses. The more often they are accurate, the stronger the patient's trust and the better the likelihood of later acceptance of suggestions in the lifelong course of a relapsing illness. As patients make their own decisions, they develop a stronger sense of self and personal responsibility.

Support System A physician's tolerance of ambiguity and risk is bolstered when the patient has a strong attachment to others. The presence of concerned persons is among the most potent facilitators of participation in treatment. Support may be provided by relatives, friends, other health care providers, a case manager, or a payee. Critical, coercive, or intrusive involvement may be counterproductive and can provoke relapse in vulnerable individuals. At most health care clinics and shelters for homeless persons, staff strive to create a nurturant and optimistic environment in which encouragement is gentle but persistent. Failed appointments are followed by telephone calls, and people who live on the streets or in shelters can attend daily or are given weekly supplies of medication in pill containers. Social, medical, and psychiatric needs are cared for simultaneously, and people receive hot drinks, snacks, and clothing as well as pills. Communication with case managers and payees is especially productive since they encounter clients on a regular basis and can monitor medications and be alert for use of other drugs or alcohol that interfere with medication taking or its efficacy.

Whenever possible, involvement with family members is crucial to improving compliance and monitoring medication taking or side effects. The dynamics of family interactions should be respected, but lack of involvement is sometimes due to the spurious use of confidentiality as an excuse to avoid the time and trouble required to communicate effectively.

Treatment Regimen Treatment must be tailored to the convenience and lifestyle of the patient. Taking medications can be cued to repetitive daily events. Although pharmaceutical companies have gone to considerable trouble to develop once-daily medications, compliance is not usually compromised unless the patient is taking multiple medications or medications are prescribed more than three times daily, because most people use meal time as their reminder. The benefits of medication are often unclear unless the physician explains them in plain language. This can include restful sleep, increased energy, clear thinking, and absence of voices or alien ideas. The physician should also try to explain and minimize adverse effects—for example, by timing sedative drugs to facilitate sleep.

The linkage between relief of symptoms and taking medication can sometimes be demonstrated by keeping a daily checklist of symptoms, adverse effects, and predicted benefits. This may be particularly helpful with symptom-sensitive people, since record keeping capitalizes on their compulsive skills and need for control.

Randolph was a 30-year-old former intravenous drug abuser with a diagnosis of schizoaffective disorder who was recently informed that he had tested positive for HIV infection. During a recent hospital admission for medical treatment, the nursing staff felt he was paranoid and depressed. Randolph had previously rejected offers of medication or had grudgingly accepted but never taken it. On this admission his urine tested positive for cocaine. When questioned at the bedside, he freely admitted using this to counter feelings of “disgust.” Randolph also agreed that after each brief euphoric interlude he was back again in the real world, feeling even more suspicious and disgusted.

The psychiatrist began with a detailed analysis of why Randolph used cocaine, what it did for him, and how he thought it worked. Based on his experience, Randolph agreed that chemicals could alter mood and behavior but at some cost. It was suggested that he might consider the alternative of prescribed medications. The point was made that unlike cocaine, these worked slowly with gradual benefit but that also unlike cocaine, the results persisted. If he was willing to attempt an experiment, chances were he would feel less suspicious and disgusted. In addition, he might no longer crave cocaine. Randolph agreed it was worth a try. After discharge, he came to the homeless clinic and was given an antidepressant and an antipsychotic. Two weeks later he was pleased and surprised to report that his voices were slowly fading away and he was sleeping soundly and feeling much less disgusted with life. His urine tested negative for cocaine and other substances of abuse.

Agnes was a 42-year-old single female well known throughout the city for frequent admissions to several hospitals over many years. She usually arrived accompanied by police, in an manic episode, combative, provocative and totally lacking any capacity to view herself as ill or in need of treatment. Although highly intelligent, Agnes was unable to hold a job and received Social Security with Medicaid health benefits. She was unpopular with the inpatient staff who viewed her behavior as “borderline.”

Her first several months with the psychiatrist were spent getting acquainted, and during this time Agnes was admitted twice to hospital. While Agnes was willing to take medication, she detested the way mood stabilizers stifled her creativity. As a result, she experimented continuously with various forms of meditation and marginal religious cults. When doing so, she gave herself frequent drug holidays. In her initial dialogue with the psychiatrist each attempted to convert the other. Agnes was sensitive to control issues; her father had been a benevolent tyrant, and her mother was critical and unloving. Agnes agreed that she found her experiences with the police and in hospital humiliating. The psychiatrist encouraged her ongoing search for spiritual meaning and nondrug methods of thought control but also maintained that these could be compatible with medication if she were willing to explore the tradeoff between a benefit she valued (staying out of hospital) and side effects she didn't (slowing of thoughts). Together they agreed on a target—to persist with medication in a sustained way until Agnes had been out of hospital longer than in any previous interlude (about 11 months).

It is now 2 years since Agnes last suffered the humiliation of hospitalization; she has found a source of spiritual solace, and she continues to complain bitterly about side effects. Together she and the psychiatrist are looking forward to a new generation of more specific psychotropic compounds that will stabilize moods without stifling her creative thought.

A paradigm shift has occurred away from a simple interest in treatment compliance to a more sophisticated understanding of the therapeutic alliance. This was brought about by the evolving systems of community outreach and support that have developed to supplement the community mental health system and to complete the process of deinstitutionalization.

Psychiatrists are now challenged to meet the needs of people striving to lead normal lives in the community despite sustained mental illness, homelessness, and substance abuse.

Prescribing is often necessary but never sufficient. Coercion is ineffective, and the current concept of compliance is inadequate. Medications are now part of an alliance that embraces each person's beliefs, experience, and lifestyle—not simply that individual's diagnosis, chemistry, or genes.

SUGGESTED CROSS-REFERENCES

Consultation-liaison psychiatry is discussed in Section 25.12, and primary care psychiatry is discussed in Section 28.1. Biological therapies are covered in Chapter 31.

SECTION REFERENCES

Blackwell B: From compliance to alliance: A quarter century of research. Neth J Med *48:*140, 1996.

*Blackwell B, editor: *Patient Compliance and the Treatment Alliance*. Harwood Academic, Sidney, Australia, 1997.

*Bleyer AJ, Hylander B, Sudo H, Nomoto Y, de la Torre E, Chen RA, Burkart JM: An international study of patient compliance with hemodialysis. JAMA *281:*1211, 1999.

Breen R, Thornhil JT: Noncompliance with medication for psychiatric disorders: Reasons and remedies. CNS Drugs *9:*457, 1998.

*Carling PJ: *Return to Community: Building Support Systems for People with Psychiatric Disabilities*. Guildford, New York, 1995.

Carrion PG, Swann A, Kellert-Cecil H, Barber M: Compliance with clinic attendance by outpatients with schizophrenia. Hosp Community Psychiatry *44:*764, 1993.

Chang P: Effects of interviewer type on compliance: An analogue study. J Counsel Psychol *41:*74, 1994.

Curtis LC, Hodge M: Ethics and boundaries in community support services: New challenges. In *Maturing Mental Health Systems: New Challenges and Opportunities*, LI Stein, EJ Hollingsworth, editors, *New Directions for Mental Health Services*, no 66. Jossey-Bass, San Francisco, 1995.

David A, Van Os J, Jones P, Fahy T, Harvey I: Insight and course of psychotic illness: Cross-sectional and longitudinal associations. Br J Psychiatry *167:*621, 1995.

De Araujo LA, Ito LM, Marks IM: Early compliance and other factors predicting outcome of exposure for obsessive-compulsive disorder. Br J Psychiatry *169:*747, 1996.

Diamond RJ: Coercion and tenacious treatment in the community: Applications to the real world. In *Maturing Mental Health Systems: New Challenges and Opportunities*, LI Stein, EJ Hollingsworth, editors, *New Directions for Mental Health Services*, no 66. Jossey-Bass, San Francisco, 1995.

*DiMatteo MR, DiNicola DD: *Achieving Patient Compliance: The Psychology of the Medical Practitioner's Role*. Pergamon, New York, 1982.

Draine J, Solomon P: Explaining attitudes toward medication compliance among a seriously mentally ill population. J Nerv Ment Dis *182:*50, 1994.

Ellenor GL, Dishman BR: Pharmaceutical care role model in psychiatry—Pharmacist prescribing. Hosp Pharm *30:*371, 1995.

Fisher D: Hope, humanity and voice in recovery from psychiatric disability. J Calif Alliance Mentally Ill *5:*13, 1994.

*Haynes RB, Taylor DW, Sackett DL, editors: *Compliance with Therapeutic and Preventive Regimens*. Johns Hopkins University Press, Baltimore, 1979.

Hayward P, Chan N, Kemp R, Youle S, David A: Illness self-management in psychotic patients. J Ment Health *4:*513, 1995.

Healey A, Knapp M, Astin J, Beecham J, Kemp R, David A, Kirov G: Cost-effectiveness evaluation of compliance therapy for people with psychosis. Br J Psychiatry *172:*420, 1988.

Holm S: What is wrong with compliance? J Med Ethics *19:*108, 1993.

*Horne R: Adherence to medication: A review of existing research. In *Adherence to Treatment in Medical Conditions*, LB Myers, K Midence Kenny, editors. Harwood Academic, Amsterdam, 1998.

Kassirer JP: The next transformation in the delivery of health care (editorial). N Engl J Med *332:*52, 1995.

Kemp R, Hayward P, Applewhaite G, Everitt B, David A: Compliance therapy in psychotic patients: A randomised controlled trial. Br Med J *312:*345, 1996.

Kuhlman TL: *Psychology on the Streets: Mental Health Practice with Homeless Persons*. Wiley, New York, 1994.

Leventhal H, Diefenbach MA, Leventhal EA: Illness cognition: Using commonsense to understand treatment adherence and affect cognition interactions. Cogn Ther Res *16:*143, 1992.

Levy RM: Involuntary treatment: Walking the tightrope between freedom and paternalism. In *Choice and Responsibility: Legal and Ethical Dilemmas in Services for Persons with Mental Disabilities*, CJ Sundaram, editor. New York State Commission on Quality of Care for the Mentally Disabled, Albany, 1994.

Lucksted A, Coursey RD: Consumer perceptions of pressure and force in psychiatric treatments. Psychiatr Serv *46:*146, 1995.

McCabe S, Unzicker RE: Changing roles of consumer/survivors in mature mental health systems. In *Maturing Mental Health Systems: New Challenges and Opportunities*, LI Stein, EJ Hollingsworth, editors, *New Directions for Mental Health Services*, no 66. Jossey-Bass, San Francisco, 1995.

McDonnell AP: Ethical considerations in teaching compliance to individuals with mental retardation. Educ Train Ment Retard *28:*3, 1993.

Melnikow J, Kiefe C: Patient compliance and medical research: Issues in methodology. J Gen Intern Med *9:*96, 1994.

Merrill BA: A global look at compliance in health/safety and rehabilitation. J Orthop Sports Phys Ther *19:*242, 1994.

Sellwood W, Tarrier N: Demographic factors associated with extreme non-compliance in schizophrenia. Soc Psychiatry Psychiatr Epidemiol *29:*172, 1994.

Solomon P, Draine J, Delaney MA: The working alliance and consumer case management. J Ment Health Admin *22:*126, 1995.

Urquhart J: Role of patient compliance in clinical pharmacokinetics: A review of recent research. Clin Pharmacokinet *27:*202, 1994.

Weiden P, Rapkin B, Zygmunt A, Mott T, Goldman D, Frances A: Postdischarge medication compliance of inpatients converted from an oral to a depot neuroleptic regimen. Psychiatr Serv *46:*1049, 1995.

27.2 MALINGERING

MARK S. LIPIAN, M.D., FH.D. AND MARK J. MILLS, J.D., M.D.

The medical concept of malingering has its origins in the philosophical dichotomy between reality and unreality, the ultimate expression of which is the question of the truth and falsity of life in general. Jesus called Satan “the father of the lie” (John); that lie has been interpreted by Christian theologians as man's denial of God who referred to the statement in John: “Who is a liar but he that denieth that Jesus is the Christ?”.

Central to philosophical moral thought has been debate of the rightfulness or wrongfulness of a person's deliberate employment of the capacity to deceive. At one conceptual extreme is Friedrich Nietzche, writing in *The Will to Power:*

> There is only one world, and that world is false, cruel, contradictory, misleading, senseless.... We need lies to vanquish this reality, this “truth,” we need lies in order to live... That lying is a necessity of life is itself a part of the terrifying and problematic character of existence.

At the opposite philosophical pole is Nicolai Hartmann, who argues in his *Ethics* that any lie “injures the deceived person in his life; it leads him astray.”

A young engineering assistant, a recent emigré from Lebanon, began to complain of burning pain in his feet. He attributed this malady to “having to walk so much at work,” where he was required to journey three or four times a day to and fro between his office and a common work space. The pain increased, and the man left the workplace on a workers' compensation disability claim.

Shortly after leaving work, and despite his “disabling” foot pain, the man embarked upon a three-week holiday to Lebanon where he visited family; did some sightseeing; and met, courted, proposed to, and married a young woman. He returned to the United States with his new wife and began the workers' compensation treatment process.

A major effort, including two podiatric surgeries and months of chiropractic manipulation, failed to improve the man's burning foot pain, and he was referred to psychiatry for “depression in the face of intractable and apparently nonphysiologic pain.”

The psychiatrist diagnosed a psychotic disorder, and began to write weekly session-based reports to the workers' compensation insurance carrier describing increasingly ominous paranoia. Over the course of several years, these reports documented escalating threats against the original employer and the insurance carrier, whom the man perceived as having deliberately produced his disabling pain, and as now toying with him by refusing to settle his case for the award of one million dollars that he felt he deserved. The psychiatrist energetically advocated giving the man his million dollars so that he might “return to Lebanon in peace.” Even as he proposed this “treatment,” the psychiatrist told of his patient bringing to therapy guns and bullets earmarked for insurance executives; appearing at sessions bloodied by deep wounds sustained “while sharpening my machete so I can cut off their feet,” and snapping out of typically docile, melancholy reveries into sudden rageful tirades directed at his alleged “tormentors”—the tirades replete with threats, vulgarities, and gory, vengeful vows.

With the help of an outside forensic consultant engaged some 3 years after psychiatric treatment began, the insurance company moved toward aggressive intervention. The man was hospitalized involuntarily at a major university center. He lost several civil commitment appeals. Faced with the probability of a court-ordered conservatorship and involuntary electroconvulsive therapy, the man miraculously ceased his threats and his bellicose behavior. His psychosis melted away, and he developed remarkable insight into not only his prior “paranoia,” but also into the psychogenic origins of his “chronic foot pain.” He even began to exercise on a treadmill. Shortly after discharge, the man's workers' compensation case was settled for a figure eminently acceptable to all. The man is not known to have interacted with the mental health community again.

HISTORY

In medicine, the tension between truth and lies has found expression in the domain of malingering and its detection. Mental illness, because it is difficult to verify objectively and is dramatic and emotional subjectively, is an age-old favorite of the malingerer. An early description of feigned madness appears in the Bible:

> And David arose, and fled that day for fear of Saul and went to Achish the king of Gath. And the servants of Achish said unto him, Is not this David the king of the land? Did they not sing one to another of him in dances, saying, Saul hath slain his thousands, and David his ten thousands? And David laid up these words in his heart, and was sore afraid of Achish the king of Gath. And he changed his behavior before them, and feigned mad in their hands, and scrabbled on the doors of the gate, and let his spittle fall down upon his beard. Then said Achish unto his servants, Lo, ye see the man is mad: wherefore then have ye brought him to me? (Samuel)

That individuals might feign or produce illness or disability for gain of some kind or to avoid duty has been known since antiquity. The Greeks considered malingering in military service to be analogous to forgery, and both offenses were punishable by death; the penalty was later mitigated to forced public exposure for three days, wearing floridly female regalia. Ulysses is said to have feigned insanity to avoid duty in the Trojan War, employing such tactics as yoking a bull and a horse together, plowing the seashore, and sowing salt instead of grain. In an early attempt at lie detection, the infant son of the King of Ithaca was placed in the furrow, directly in the line of Ulysses' oncoming plow. If the man was mad, it was reasoned, he would take no heed of the obstacle. Ulysses swerved to avoid his son, and that action was considered proof that his madness was a ruse.

In the second century AD Galen wrote a treatise, *On Feigned Diseases and the Detection of Them*, in which he described Roman conscripts who cut off thumbs or fingers to ensure unfitness for duty. In the sixteenth century Paolo Zacchias, considered by many to be the father of forensic medicine, wrote of madness, “There is no disease more easily feigned, or more difficult to detect.”

Truth versus imposture was discussed in Theodric Beck's 1823 *Elements of Medical Jurisprudence*, perhaps the first notable American text on the subject:

> In almost every age, impostors have sprung up who affect various maladies and operate on the superstition or the curiosity of the vulgar. And even the higher ranks of society, from motives as unworthy, have occasionally, like the courtiers of Dionysus and Louis XIV, given a sanction to such practices.

DEFINITION AND COMPARATIVE NOSOLOGY

Contemporary theorists share with their historical antecedents the construct that the fundamental characteristic of malingering is intentional falsity, with an underlying incentive of gain of some kind. According to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV):

> The essential feature of Malingering is the intentional production of false or grossly exaggerated physical or psychological symptoms, motivated by

external incentives such as avoiding military duty, avoiding work, obtaining financial compensation, evading criminal prosecution, or obtaining drugs. Under some circumstances, malingering may represent adaptive behavior—for example, feigning illness while a captive of the enemy during wartime.

Malingering should be strongly suspected if any combination of the following is noted: (1) Medicolegal context of presentation (e.g., the person is referred by an attorney to the clinician for examination); (2) marked discrepancy between the person's claimed stress or disability and the objective findings; (3) lack of cooperation during the diagnostic evaluation and in complying with the prescribed treatment regimen; and (4) the presence of Antisocial Personality isorder.

Malingering differs from factitious disorder in that the motivation for the symptom production in malingering is an external incentive, whereas in factitious disorder external incentives are absent. Evidence of an intrapsychic need to maintain the sick role suggests factitious disorder. Malingering is differentiated from conversion disorder and other somatoform disorders by the intentional production of symptoms and by the obvious, external incentives associated with it. In malingering (in contrast to conversion disorder), symptom relief is not often obtained by suggestion or hypnosis.

Thus, malingering is distinguished from factitious disorder in its motivation: whereas malingering is prompted by a conscious desire to obtain tangible external rewards or environmental outcomes, factitious disorder is not. In the latter, a matrix of intrapsychic needs manifesting themselves as a nearly irresistible desire to assume the sick role is thought to motivate the intentional deception of malady. The problems in diagnostic differentiation are obvious. A person might malinger to obtain tangible rewards (such as disability payments or chicken soup) associated with the sick role but might still enjoy the nurturing and care that such a role provides. A person with factitious disorder might be highly resistant to returning these financially and gustatorily satisfying aspects of being sick, while still clinging to and needing the emotional gratification that a person with this disorder is supposed to crave.

Differentiating malingering from the somatoform disorders (e.g., conversion disorder) is easier in that the somatoform disorders lack the volitional component of malingering. In the somatoform disorders an underlying emotional conflict is thought to be unconsciously transformed into a physical manifestation of some kind. No external or environmental outcome or reward is consciously sought. Rather, the defensive makeup of the person with a somatoform disorder is believed to be more tolerant of manifest turmoil in observable, extrapsychic form than of unobservable, emotional upset.

Whereas the various nosological schemes employed historically and internationally agree on the centrality of volitional deception to malingering, there is more controversy over whether malingering should be considered a mental disorder at all. That is of some importance because in most civil litigation, only mental disorders are compensable. DSM-IV classifies malingering with additional conditions that may be the focus of clinical attention, and thus, if malingering occurs in association with such mental disorders as antisocial personality disorder, factititious disorder, or a somatoform disorder, the diagnostician is enjoined to consider those diagnoses primary. In fact, a diagnosis of factitious disorder excludes a diagnosis of malingering.

Definitions of Subtypes of Malingering It has been suggested that malingering be considered not as a dichotomous variable (a condition that is either present or absent) but as falling along a continuum in terms of (1) degree of intentionality, (2) degree of symptom exaggeration involved, and (3) degree of actual impairment (if any). In keeping with the concept of a continuum, the following definitions have been proposed:

Pure Malingering Feigning a disease or disability when it does not exist to any extent.

Partial Malingering Consciously exaggerating symptoms that really exist.

False Imputation Ascribing actual symptoms to a cause consciously understood to have no relation to the symptoms.

Misattribution Ascribing actual symptoms to a cause erroneously believed to have given rise to them. Such misperception is often the product of unconscious processes that have interfered with reality testing and in its pure form, it does not constitute malingering. To the extent that the misattribution is consciously augmented (false imputation), malingering is at play.

A female police officer filed a lawsuit against her former department, alleging that sexual harassment at the workplace and wrongful termination for "whistleblowing" had resulted in psychiatric damages. Psychiatric evaluation of the case suggested that while the officer was suffering from a bout of major depressive disorder, "harassment" had not caused it. Instead, the officer was found to suffer from profound narcissistic and borderline personality disorders, and a series of inappropriate interactions and behaviors born of these disorders had given rise to the tensions and improprieties ultimately resulting in her professional demise. While the officer's narcissism did, to a great extent, cause her to distort and truly misperceive the sequence of events leading to termination (misattribution to her gender rather than to her professional failures), the officer also clearly and blatantly "invented" episodes of harassment, pointing to these fantastic events as the specific origins of her (exaggerated) symptoms (false imputation and malingering).

In addition to the various degrees of malingering, several forms of malingering have been identified and defined:

Simulation Feigning symptoms that do not exist, or the gross, conscious exaggeration of preexisting symptoms. Simulation has sometimes been referred to as *faking bad* and *positive malingering*.

Dissimulation Concealing or minimizing existing symptoms. Dissimulation has also been called *faking good, negative malingering*, and *defensiveness*. The term is somewhat confusing because it has sometimes been used to refer to medical faking in general—that is, as a synonym for malingering.

Staged Events Carefully planning, orchestrating, and executing events, with the desired result being either an actual injury or a credible explanation for a disability that will later be feigned.

Data Tampering Altering diagnostic data or records to simulate a disorder. Such alteration might take the form of self-mutilation (to influence the outcome of a physical examination), physical addition to or removal of substances from laboratory specimens (to influence the results of analyses performed on the specimens), and defacing or adjusting laboratory reports, diagnostic instruments, and medicohistoricolegal documents.

Opportunistic Malingering Exploiting a naturally occurring event or preexisting medical condition for gain. Opportunistic malingering is distinguished from partial malingering, which involves the exaggeration of specific preexisting symptoms.

Symptom Invention Falsely and consciously complaining of symptoms that are unrelated to any current or preexisting disorder or injury.

An air traffic controller was hit by a car while riding his bicycle, and thrown by the impact into a nearby ravine. Despite minimal trauma, no loss of consciousness, perfect Glasgow Coma Scale scores, and negative neurological workup, the man was diagnosed as suffering from closed head trauma and postconcussional disorder. He filed suit against the offending driver, claiming to have cognitive impairments including left-right confusions (mistaking left for right and vice versa); analogous up-down confusions, and similar object confusions (mistaking a car for a bus, a plane for a bird). His doctors judged him to be permanently disabled from work as an air traffic controller.

Although neuropsychological testing and psychiatric evaluation revealed none of the claimed impairments and, in fact, suggested malingering, the man was charismatic and appealing, and a generous settlement package was offered and accepted. Undercover investigation discovered the man 8 months later living in Iceland under a new name, once again working happily and successfully as an air traffic controller.

EPIDEMIOLOGY

In the United States, malingering is not considered as a medical or psychiatric diagnosis. Therefore, such traditional epidemiological considerations as the reliability of diagnostic criteria and the validity of the diagnostic construct are largely irrelevant. Further, estimations as to the typical age of onset, sex ratio, prevalence, or age-specific features are hampered by systematic underreporting of malingering by health care practitioners. Several concerns underlie the reluctance of professionals to label patients as malingerers and the tendency to refer instead to a vague lack of objective evidence in support of the patient's particular subjective

complaints.

Most clinicians were drawn to the mental health field out of some desire to help others; they find it distasteful and a violation of the doctor-patient relationship to call a patient a liar. Concern over legal liability also inhibits the widespread labeling of malingering: although expert testimony about malingering, given in court and in good faith, is protected by immunity, a clinical label applied in the outpatient office or emergency room is not. Finally, physicians are concerned about the anger and possibly physical outbursts that might result if a person's conscious attempts at deception are foiled and his or her conscious scheme to manipulate the system for reward is dismantled.

Nevertheless, according to some epidemiological data it is estimated that the average person in the average day lies twice, although the same person may be completely truthful at other times. A 1 percent prevalence of malingering has been estimated among mental health patients in civilian clinical practice, with the estimate rising to 5 percent in the military. In a litigious context, during interviews of criminal defendants, the estimated prevalence of malingering is much higher—between 10 and 20 percent.

Most criminals are skillful liars, and the capacity to lie develops early; about 50 percent of children presenting with conduct disorders are described as bringing to the clinician serious lying-related issues. It can reasonably be assumed that in civil lawsuits, particularly those involving child custody or evaluations pertaining to civil commitment, a similar elevation of the prevalence of dissimulation (faking good) would become apparent if the appropriate data were amassed.

Thus, although no familial or genetic patterns have been reported and no clear sex bias or age at onset has been delineated, malingering does appear to be highly prevalent in certain military, prison, and litigious populations, and, in Western society, in men from youth through middle age. Associated disorders include conduct disorder and anxiety disorders in children and antisocial, borderline, and narcissistic personality disorders in adults.

ETIOLOGY

Although no biological factors have been found to be causally related to malingering, its frequent association with antisocial personality disorder raises the possibility that hypoarousability may be an underlying metabolic factor. No predisposing genetic, neurophysiological, neurochemical, or neuroendocrinological forces are known.

During the height of psychoanalytic influence, malingering was considered to be a form of mental disease. Kurt Eissler wrote, “It can be rightly claimed that malingering is always a sign of a disease often more severe than a neurotic disorder because it concerns an arrest of development at an early phase.” Although most analysts today accept the nondisease concept of malingering, the malingerer is still thought to be as enslaved to his or her past as the neurotic is, except that the malingerer is unable to produce an actual neurotic symptom as a result of primitive ego and superego development. The malingerer acts out consciously that which the neurotic is able to convert unconsciously; the psychodynamic connection between malingering and the somatoform disorders is a close one.

At the other end of the etiological spectrum lie the theories of Thomas Szasz who has argued that malingering can be understood only in the context of the sociopsychology of games, systems, and interactions. Malingering would be meaningless in a strictly psychiatric situation limited to physician and patient; in fact, if it were to emerge only in such a context, it would have to be considered factitious disorder. According to Szasz, malingering is a relevant concept only when the physician and patient are part of a larger social structure and each are players in a social game. The patient is out to score; so is society. The physician's job is to officiate the proceedings, to prevent cheating, and to minimize any unfair advantage. As such, the prudent psychiatrist approaches every evaluation with the assumption that the patient being evaluated may stand ready to use mental infirmity to achieve questionable ends.

Social and cultural forces must be considered as etiologically relevant in any analysis of malingering behavior. Although studies showing marked differences in readiness to complain of illness among different ethnic groups have been criticized on methodological grounds, it is generally accepted that the culture of work, ethical responsibility, and duty to country in the form of military service (with the potential for self-sacrifice) varies from society to society, subculture to subculture, and epoch to epoch. How acceptable, condemnable, or laudable malingering may be within any sociocultural context will surely influence its prevalence and manifestations.

DIAGNOSIS AND CLINICAL FEATURES

In his 1823 *Medical Jurisprudence*, Beck described the three contexts that have most stimulated malingering behavior throughout history.

> Diseases are generally feigned from one of three causes—fear, shame, or the hope of gain. Thus the individual ordered on service will pretend being afflicted with various maladies to escape the performance of military duty—the mendicant, to avoid labour, and to impose on public or private beneficence—and the criminal, to prevent the infliction of punishment. The spirit of revenge, and the hope of receiving exorbitant damages, have also induced some to magnify slight ailments into serious and alarming illness.

Contemporary authors have elaborated on those core motivations and added a few more. It is thought that the twentieth-century malingerer is hoping to achieve one of the following seven ends:

Avoidance of Criminal Responsibility, Trial, and Punishment Criminals may pretend to be incompetent to avoid standing trial; they may feign insanity at the time of perpetration of the crime; they may malinger symptoms in order to receive a less harsh penalty; or they may attempt to act too incapacitated (incompetent) to be executed. The following transcript from an east-coast newspaper during the 1940s was quoted by Henry Davidson in the chapter on malingering in his *Forensic Psychiatry:*

> X and Y who acted like mad dogs in a murderous payroll robbery and like apes when brought to trial yesterday, were locked up out of the public gaze today to await their next appearance in court.... The defense lawyer contended that they were insane and they gave no outward indications to the contrary. They walked into the courtroom with an apelike gait, arms hanging loosely, heads wabbling [sic] from side to side with their chins held against their chests for pivots.... Attendants had to prod them along like animals. X poked the fingers of his hand into his mouth and gnawed them. Y ate pieces of paper from the table. X then rolled up a paper napkin and ate it. Y lowered his brow to the edge of the table and rubbed it like a horse scratching his head on a post. X took a pair of soiled underpants from his pocket and wrapped it around his neck. Then he licked it and wrapped it around his head. At recess, both prisoners had to be lifted from their chairs like sacks. They sloughed out of the room like apes....

Avoidance of Military Service or of Particularly Hazardous Duties Persons may malinger to avoid conscription into the armed forces, and once conscripted, they may feign illness to escape from particularly onerous or hazardous duty.

Financial Gain Modern malingerers may seek financial gain, in the form of undeserved disability insurance, veterans' benefits, workers' compensation, or tort damages for alleged psychological injury. In the chapter on malingering and associated syndromes in the second edition of *American Handbook of Psychiatry*, David Davis and James Weiss paraphrased Punton's description of a malingerer, circa 1903:

> A man named Moffett described his “cane and screw” racket, in which he used a specially prepared cane to loosen the floor screws on streetcars and railroad cars, over which he would then pretend to stumble and then institute a claim for injuries sustained. He stated that he made it a universal rule to employ the very best doctors, as he found, by experience, that they were the most easily fooled, while the companies he fleeced were better satisfied with their opinions. In this connection, he noted that he paid the doctors' bills promptly and willingly, even though at times he thought they were exorbitant, but this was done, he said, in order to impress them with the honesty of his actions.

In a more modern example at the cutting edge of “political correctness,” a male high school junior claimed his female drama teacher (widely regarded by students as personifying female desirability) had seduced him into a brief sexual relationship, which the boy finally ended “because it wasn't right.” The boy and his mother then sued the teacher and the school department for “ruining his life, his self-concept, his sexual identity, and his future.” At the time of the alleged affair, the student was sexually involved with a female peer; shortly after allegedly “spurning” his drama teacher, he began a new sexual relationship with a different female classmate. The young man went on to graduate from high school, enter the University of California, earn a 3.5 grade point average, and achieve considerable social success on campus. Nevertheless, he, his mother, and his therapist steadfastly reported chronic depression, social awkwardness, sexual inhibition, and irreversible identity disturbance, all the result of the alleged lust of the insatiable drama teacher.

Avoidance of Work, Social Responsibility, and Social Consequences Individuals may malinger so as to escape from unpleasant vocational or social circumstances, or to avoid the social and litigation-related consequences of vocational or social improprieties.

> A lawyer had mishandled her assets to the point of bankruptcy; she similarly mishandled her legal practice to the point of being served with several professional negligence suits as well as formal charges of misconduct by the American Bar Association. She proceeded to implement the following plan:
>
> (a) She shifted bankruptcy status so as to protect her existing assets against all her disgruntled former clients.
>
> (b) She convinced a psychiatrist that she suffered from disabling “anxiety and depression,” thereby activating her generous disability insurance policy and achieving financial independence while escaping from what had been a most unpleasant marriage to her legal practice. (A variety of data gathered by the insurance company and its psychiatric consultant confirmed the frank falsity of the attorney's claimed psychiatric difficulties.)
>
> (c) Through her psychiatrist, the malingerer convinced attorneys representing the American Bar Association that she was “incompetent” to assist in the defense of the various professional transgressions of which she was accused, thereby achieving an indefinite “stay” from facing the professional and social consequences of her alleged legal misbehavior.

Facilitation of Transfer From Prison to Hospital Prisoners may malinger (fake bad) with the goal of obtaining a transfer to a psychiatric hospital, from which they may hope to escape or in which they expect to do easier time. However, the prison context may also give rise to dissimulation or faking good: the prospect of an indeterminate number of days on a mental health ward may prompt an inmate with true psychiatric symptoms to make every effort to conceal them.

Admission to a Hospital In this era of deinstitutionalization and homelessness, persons may malinger in an effort to gain admission to a psychiatric hospital. Such institutions may be seen as providing free room and board, safe haven from the police, or refuge from rival gang members or disgruntled drug cronies who have made street life even more unbearable and hazardous than it usually is.

> A young, healthy, but disheveled-looking man presented to a psychiatric emergency room in the early morning hours. He declared that “voices told me to come down to the VA and see the doc.” Pressed to further elucidate, the patient confided, “The voices told me I should be coming on in to the hospital for a few days.” When he was confronted with the atypical nature of this hallucinatory experience, the young man became increasingly agitated, demanding, and confrontational. Ultimately, he admitted there in fact had been no voices at all; a drug deal gone sour had created a need for a safe haven and the hope that the Veterans Administration would comply.

Drug-Seeking Malingerers may feign illness in an effort to obtain favored medications, either for personal use or, in a prison setting, as currency to barter for cigarettes, protection, or other inmate-provided favors.

> A woman was rear-ended while parked at a stoplight, giving rise to a herniated cervical disc and chronic pain. A first attempted surgical intervention was botched when the wrong disc was removed and the wrong section of spine fused. A second was botched when the anesthesiologist managed to pharmacologically paralyze his patient, but failed to put her to sleep. Thus, the woman was conscious and alert throughout the entire discectomy and fusion procedure, but unable to speak or gesture to alert anyone to her plight. Several malpractice lawsuits were filed, and posttraumatic stress disorder was alleged.
>
> Frequent, impulsive wanderings across the country were held by the plaintiff's attorney to be the most disabling feature of this unfortunate woman's posttraumatic stress disorder. However, psychiatric evaluation revealed that the woman had no significant symptoms of posttraumatic stress disorder. Rather, she displayed a classic picture of substance-induced depressive disorder and manipulative drug-seeking behavior.
>
> It turned out that the plaintiff's cross-country wanderings centered around visits to random emergency departments, with desperate complaints of pain and demands for morphine. Further investigation revealed a pre-accident history of chronic polysubstance dependence, and similar cross-country drug-seeking meanderings beginning in the early teens.

DETECTION OF MALINGERING

In his article on the detection of malingered mental illness in *Behavioral Sciences and the Law*, Philip Resnick noted that clues to aid in the detection of deception have been sought for at least 3000 years. He quoted the following description of a liar, taken from a papyrus drafted around 900 BC: “He does not answer questions, or gives evasive answers; he speaks nonsense, rubs the great toe along the ground, and shivers; he rubs the roots of the hair with his fingers.”

The clinician must rely primarily on interviewing skills to detect malingering. Untrained observers seem to do little better than chance in lie detection, and some studies have found police detectives to do hardly better than undergraduates in judging guilt versus innocence. Despite conceptual and methodological flaws, David L. Rosenhan's study of malingerers faking insanity as inpatients on a psychiatric ward casts doubt on the ability of trained clinicians to differentiate real from feigned mental illness in certain contexts.

Nevertheless, several clues exist, and if properly applied, they are useful for the clinician who needs to detect malingering. The psychiatric conditions most likely to be malingered are mental retardation, organic impairment, amnesia, psychosis, and posttraumatic residua, including depression and posttraumatic stress disorder.

Detecting Deception Perpetrators' urges to confess seem greatest shortly after they have transgressed; during this period their guilt is apparently at its most burdensome and is as yet unrelieved by mental defense mechanisms. Memory has yet to be distorted, either consciously or unconsciously. Thus, every effort should be made to interview the criminal shortly after the crime.

Intentional malingering is harder to maintain as the evaluative interview becomes increasingly lengthy because of basic fatigue and a pull toward reality. Therefore, when malingering is suspected, the clinical interview should be as long, detailed, and grueling as possible.

Research on lying and malingering has shown that liars often speak in high-pitched voices, make errors of grammar, and make slips of the tongue. Students instructed to lie, hesitate or pause while lying and tend to make irrelevant, rambling, and negative comments; the negativity is thought to be related to guilt. The passive voice is more common than the active, discrepancies between verbal and nonverbal expression (blinking, dilated pupils, rubbing or stroking of the self) are frequent, and answers may appear rehearsed, overly facile, and rote. The clinician's suspicion should be aroused if an interviewee issues spontaneous assurances of veracity: “Doc, would I tell you a lie?” or “To be perfectly honest....”

Planning a lie before the fact allows the liar to rehearse more carefully. Thus, during the lie he or she requires fewer pauses for words, phrases, or ideas; has more control over tone of voice; and is able to project more confidently. Planned lies are consequently less easy to detect than unplanned ones. Among the most successful liars are those who exaggerate their feigned sentiments histrionically (*hamming*); those who have had some acting experience; those who are intelligent, creative, resourceful, and have good memories; and those who are smooth and practiced verbally.

Contrary to intuition and popular myth, facial expression and eye contact are poor indicators of truthfulness and may actually interfere with more sensitive lie-detection modalities. One study measured the ability to detect deception under three different conditions: (1) while watching and hearing a videotaped interview, (2) while listening to an audio recording of the interview but seeing no visuals, and (3) while reading a transcript of the interview. Listeners and readers were far more successful at identifying deception than were watchers, which suggests that visual cues during the interview serve primarily as distractions. At a fairly early age children learn how to control their facial expressions so as to conceal their emotions, refining those skills through game playing and social interaction; the face is particularly adept at deception.

Liars do not necessarily make less eye contact, have shifty eyes, smile or gaze less, or adjust posture more than nonliars. However, Sigmund Freud's assertion that a liar's unconscious guilt “oozes out of every pore” may at least be true below the neckline. A marked incongruity between calmness of the face and tension, fidgeting,

and active movement of the less controllable arms, hands, legs, and feet is a good reason to suspect deception.

Detection of Specific Malingered Conditions

Malingered Mental Deficiency or Mental Retardation Mental deficiency or mental retardation is difficult to feign in the United States, primarily because the educational system is so energetic in its passion to evaluate and classify intelligence. In general, the school system is paced at 1 school year for each year of mental age on the Stanford-Binet Intelligence Scale, with the first grade adjusted to a mental age of 6. Thus, a student who has completed the ninth grade should have a mental age of approximately 14 years. A sudden adult score more than 2 or 3 years out of line with that formula, in the context of prior school scores consistently adhering to the predicted pattern, should arouse suspicion. In addition, malingerers may miss easy questions on psychometric tests while correctly answering more difficult ones.

It is useful to check vocational and military records if mental deficiency is claimed. These sources may provide further corroborative or contradictory instances of formal intelligence testing, and a marked discrepancy between claimed capacity and historical accomplishment would be further cause to suspect falsity. Table 27.2-1 summarizes some clues to the detection of malingered mental deficiency or mental retardation.

1. Striking discrepancy between level of educational achievement and level of tested or reported intelligence (>3 years' difference)
2. Striking discrepancy between military and employment testing records and presenting behavior and test performance
3. Striking discrepancy between adult test performance and prior pattern of test performance
4. Failure on easy items in the face of success on difficult items during evaluative testing
5. Incongruity of vocational and social performance with presentation capabilities and behaviors

Table 27.2-1 Clues to the Detection of Malingered Mental Deficiency or Mental Retardation

Malingered Cognitive Disorders The most noteworthy discrepancy reported between those who have dementia and those who are attempting to fake a cognitive disorder has been the presence of marked perseveration in the former and its absence in the latter. If the suspected malingering follows a traumatic injury (e.g., head trauma or toxin exposure), a careful study of the known effects of that insult is necessary before the physician can compare the claimed sequelae with the plausible ones. Most persons attempting to malinger feign fairly blatant and dramatic symptoms—paranoia, morbid depression, suicidal behavior, gross forgetfulness—that may not be reasonable outcomes of a specific class of physical injury.

The patient's performance before and after the alleged incident should be thoroughly analyzed as should the patient's behavior in all spheres of function since the onset of the alleged deterioration. A person who performs well socially and when at leisure while claiming inability to work should be suspected of malingering. Table 27.2-2 summarizes some guidelines on the detection of malingered cognitive disorders.

1. Lack of marked perseveration
2. Implausible symptom profile given reported injury
3. Psychotic symptoms confused with cognitive impairments
4. Unimpaired function in social and recreational realms in the face of gross disability in domain of vocation and responsibility

Table 27.2-2 Clues to the Detection of Malingered Cognitive Disorders

Malingered Amnesia Amnesia, probably the most common clinical presentation of malingering, is claimed by 30 to 55 percent of perpetrators of homicide. It is easy to feign and is particularly difficult to detect.

At least six possible causes have been suggested for amnesia: (1) conversion disorder, (2) psychosis, (3) alcoholism, (4) head injury, (5) epilepsy, and (6) malingering. Before malingering is ascribed, the clinician should review and eliminate the other five potential causes. A good diagnostic battery would include negative results on skull X-ray, head computed tomography or magnetic resonance imaging, and electroencephalography; normal findings on a neurological examination; a life history inconsistent with either conversion disorder or alcoholism (or other causes of intoxication); and a clinical examination and history inconsistent with either alcoholic amnesia (alcohol-induced persisting amnestic disorder) or psychosis (alcohol-induced psychotic disorder).

If those tests are negative, the clinician faces the difficult task of amassing evidence, albeit inferential, of malingering. Motivation is a key indicator. Previous amnestic episodes without apparent motivational precursors lower the likelihood that the patient is malingering. Similarly, a patient with histrionic personality traits is more likely to be experiencing true dissociative amnesia than one with primarily antisocial traits.

The timing of onset and recovery, and correlation of the alleged amnestic episode with convenience, are other clues to the presence of malingering. Global amnesia is somewhat more convincing than spotty, patchy, self-serving amnesia. There have been several reported epidemics of copy-cat amnesias following famous or highly publicized cases; an eye to recent sensational litigation is prudent. Table 27.2-3 summarizes some guidelines on detecting malingered amnesia.

1. No history of amnestic episodes
2. Antisocial personality traits more prominent than histrionic personality traits
3. Spotty, episode-specific amnesia rather than global amnesia
4. Self-serving timing of onset and recovery
5. Recent, widely publicized, suspiciously familiar cases involving amnesia

Table 27.2-3 Clues to the Detection of Malingered Amnesia

In the context of forensic psychiatry and exculpation from criminal blame, a claim of amnesia following a crime is of little help to a defendant. Loss of memory following severe upset (such as committing a murder) is a fairly common phenomenon. To be useful in an insanity defense, the amnesia claimed would have to have existed for an appreciable period prior to the criminal act and would have to have continued unabated through it; confusion on that issue has precipitated the downfall of many a malingerer.

Malingered Psychosis In assessing an apparently psychotic patient, the clinician should obtain as much historical and collateral information as possible, especially if malingering is suspected. Motivation, availability of coaching, and adherence of the symptom picture to known disorder profiles are of use in assessing the plausibility of an evaluee's psychotic presentation.

Beyond these initial assessments, a number of indicators have been identified that, if present, raise the probability of malingering. Malingerers tend to overact their part, often mistakenly believing that the more bizarre they appear, the more convincing they are. The earlier example of the two ape-men is a case in point.

Schizophrenic patients tend to be reluctant to discuss their symptoms, particularly if they are in the throes of a persecutory delusional psychosis. On the other hand, malingerers may be anxious to call attention to their illnesses and to whatever outlandish belief systems they are attempting to feign. The form of schizophrenic thinking is far more difficult for malingerers to imitate than its content. Thus, it is much less likely that malingerers will mimic loose associations or tangential or circumstantial reasoning than that they will describe or act out odd beliefs.

Unsophisticated malingerers often confuse madness with dumbness, supposing that silly or childlike responses go hand in hand with bizarre, delusional experiences. Such faulty juxtaposition may be evident during clinical interviews and can be highlighted on psychological testing.

Malingerers may claim the sudden onset of a delusion. Realistically, delusional symptoms take weeks, months, or years to coalesce. Truly psychotic persons experiencing an acute episode are likely to act in direct accordance with their delusional system; malingerers claim to have done so only when such concordance is opportunistic. The discrepancy is less striking once the psychosis has become more longstanding; the burned-out schizophrenic person may not be as flamboyant in acting on delusional beliefs as a younger, more acutely psychotic person.

Malingerers are likely to contradict themselves in their accounts of the illness, and those contradictions may multiply as the interview progresses. They may be led astray by leading questions that the examiner knows are psychologically absurd, and they may make mistakes when questioned rapidly or directly confronted. Their symptoms may fit no diagnostic entity, and when caught in such a discrepancy they may sulk or laugh from embarrassment.

Malingerers sometimes attempt to take control of the interview, behaving in an intimidating and blustery manner. They are likely to repeat questions or answer questions slowly, working to give themselves more time to fabricate convincing responses. They present themselves as blameless within their feigned illness, yet they are likely to have nonpsychotic alternative motives for their behavior, such as killing out of revenge or out of paraphiliac desire.

As is true of the form of schizophrenic thinking, malingerers are unlikely to successfully imitate the subtle signs of residual schizophrenia. Blunted affect, concreteness, and odd, schizoid relatedness are presentations familiar only to the most talented malingerers. Table 27.2-4 summarizes some features of malingered delusions.

1. Abrupt onset and termination rather than gradual development and hesitant abandonment
2. Eagerness to call attention to delusions and symptoms rather than reluctance to acknowledge them
3. Behavior inconsistent with delusional content rather than reflective of delusional content
4. Thought content grossly disturbed in the face of quite conventional and goal-directed thought process

Table 27.2-4 Features of Malingered Delusions

Criminal defendants attempting to malinger psychosis most frequently fake auditory hallucinations; thus, detailed knowledge of the characteristics of genuine hallucinations and detailed questioning into the nature of the claimed experience are mandatory if the clinician suspects malingering.

HALLUCINATIONS True hallucinations tend to be associated with delusions and are often a psychotic expression of some intrapsychic need. Accusatory voices may represent unacceptable (or psychotic) guilt, horrifying creatures or threatening animals may represent unacceptable aspects of the personality, and the belief that one is giving off a foul odor may represent severe depression and self-deprecatory delusion.

Persons with schizophrenia frequently report that voices speak directly to them or pass judgment on them. Schizophrenic hallucinations tend to be intermittent rather than continuous, and over 50 percent of schizophrenic persons eventually acknowledge that they may have imagined their hallucinations. Some studies reveal that 66 percent of schizophrenic patients have auditory hallucinations and 33 percent have visual hallucinations; the visual hallucinations almost always accompany the auditory hallucinations. Visual hallucinations tend to be in color and of normal-sized people. Olfactory and gustatory hallucinations have been reported in acute episodes of schizophrenia; olfactory hallucinations are of unpleasant odors.

Auditory hallucinations are reported to come from outside the head by 88 percent of schizophrenic persons, and 75 percent report that they hear both male and female voices. The message is usually clear; it is vague only 7 percent of the time, and is accusatory about one-third of the time. About 30 percent of schizophrenic persons answer the voices that they hear.

Command hallucinations, although clinically noteworthy, are ignored by true psychotic persons up to 60 percent of the time. They are more likely to be obeyed if the voice is familiar to the hallucinator or if they are one element in a developed delusional system.

Most truly psychotic individuals have developed strategies for coping with hallucinatory episodes. Commonly helpful are changes of posture, distraction through activity or interpersonal contact, or ingestion of antipsychotic medication. Although the auditory hallucinations of truly psychotic persons tend to be general and basic, those of malingerers are often stilted, specific, and self-serving.

Alcohol-induced psychotic disorder with hallucinations, which follows the abrupt cessation of previously high alcohol intake, ordinarily produces vivid hallucinations. Although voices are extremely common, experience of noise, music, unintelligible voices, or tactile hallucinations (insects) are more frequent than in schizophrenia. Substance-induced psychotic disorders generally involve unformed, indistinct hallucinatory noises. Table 27.2-5 summarizes some features of malingered auditory hallucinations.

1. Continuous rather than intermittent
2. Vague, inaudible, or unintelligible rather than distinct
3. Free-standing rather than associated with delusions
4. Stilted in language and specific in tone rather than basic and general
5. Reported in the first person rather than in the third person
6. Uncontrollable rather than susceptible to strategies for containment
7. Irresistible rather than susceptible to indifference

Table 27.2-5 Features of Malingered Auditory Hallucinations

Malingered Posttraumatic Residua In today's litigious society, the psychiatrist may be asked to determine the veracity of psychiatric or behavioral symptoms that a person claims resulted from trauma of some sort, whether psychological or physical in nature. The actual incidence of malingered psychological symptoms following physical injury is unknown, but one study found that 48 out of 50 cases of postaccident neurosis had cleared up 2 years after the legal claim had been settled, and a United States General Accounting Office follow-up study reported in 1980 that approximately 40 percent of persons considered to be totally disabled had no disability whatsoever 1 year after that declaration.

Malingerers in personal injury cases seldom feign psychosis. Much more common are claims of depression and, now that it has been highly publicized, of posttraumatic stress disorder. Such ambulatory illnesses allow malingerers to avoid responsibility and to escape from work while sparing them the unpleasantness and the 24-hour scrutiny of hospitalization in a psychiatric setting.

In attempting to distinguish malingered from actual posttraumatic depression, it is helpful to consider subtle signs and symptoms of depression not commonly known to the lay individual. The absence of diurnal variation, of specifically early-morning awakening (as opposed to vague, generalized insomnia), of angry irritability, and of diminished sexual interest suggest uninformed imitation. Similarly, the lack of a family history of the disorder or of prior incidents is suspicious. Possibly most revealing of all is a lack of depressive withdrawal from enjoyable domains (such as social and recreational activities) in the face of apparent total incapacity in aversive ones (such as work).

Posttraumatic stress disorder has become a particularly productive field for fraud, especially since the Veterans Affairs (VA) widely publicized the posttraumatic stress disorder criteria in the early 1980s and announced that any veteran meeting the criteria was eligible for governmental compensation. In posttraumatic stress disorder claims as in posttraumatic depression claims, amassing a detailed picture of pretrauma functioning is crucial. The reasonableness of the relation between the degree of the stressor and the alleged symptom cluster, the temporal pattern, and any prior, concurrent, or subsequent symptoms and stressors figure centrally in the evaluation.

Some specific clues to the malingering of posttraumatic stress disorder exist. The patient with true posttraumatic stress disorder frequently focuses on the negative features of the disorder—the emotional numbness, indifference, and social withdrawal—whereas the malingerer may be more impressed by (and therefore hopes to impress more with) the expressionistic, flamboyant nightmares and flashbacks. In true posttraumatic stress disorder victims the nightmares tend to vary in content while hewing to the constant themes of terror and helplessness; malingerers are more likely to report reexperiencing exactly the same dream, often described as a videotape-like reliving of the alleged trauma itself. Many malingerers are misinformed about the nature of the true flashback experience, confusing it with—and describing it as—a benign pictorial memory. Particularly in posttraumatic stress disorder with its clearly delineated criteria and entirely descriptive diagnosis, a textbook-perfect presentation should raise greater suspicion than a presentation characterized by a vague, more approximate symptom cluster.

Table 27.2-6 lists a number of items suggestive of malingering of psychological distress following a traumatic incident.

1. Assertion of inability to work in the face of unimpaired capacity for pleasurable activity (recreation, social interaction)
2. Subscription to more obvious symptoms of widely publicized disorders (depression, posttraumatic stress disorder) in the face of denial of more subtle features of the disorders
3. Spotty, questionable vocational history; tendency to drift; fringe member of society
4. Evasiveness during interview; unwillingness to concretely address a return to work, responsibility, and social expectation
5. General presentation of sullenness, suspicious guardedness, uncooperativeness, or resentment
6. Refusal to comply with recommended diagnostic or treatment procedures; avoidance of direct examination
7. History of disabling injuries and unusually frequent absences from work
8. Traits common to antisocial, narcissistic, borderline, or histrionic personality disorders
9. Energetic and concerted pursuit of legal claim in the face of alleged debility caused by depression and posttraumatic stress disorder
10. Refusal of employment suggested as plausible despite alleged disability
11. Self-depiction in excessively favorable and capable terms prior to alleged trauma and behavioral collapse

Table 27.2-6 Factors Suggesting Malingering of Psychological Distress Following Trauma

Objective Testing For centuries objective measures, such as physiological reactions, have been used to differentiate the conniving from the sincere. For example, realizing that stress may cause a reduction in saliva production, the Bedouins required conflicting witnesses to lick a hot iron; the person whose tongue stuck to the iron was adjudged to be the liar. Similarly, in ancient China individuals in conflict were each required to chew on a ball of rice and then to spit it out; the person whose ejected rice was drier was considered to be the charlatan.

The more contemporary technique for measuring psychophysiological stress is the polygraph. However, the polygraph is not foolproof; at best it is 80 to 90 percent reliable, which has caused many jurisdictions to forbid its use in forensic evaluations. The polygraph seems best at detecting lying in anxious but honest individuals; it is therefore prone to false-positive results. Particularly disheartening is the boast, empirically verified, of many practiced sociopaths that they can beat the lie detector through exercises such as meditation, holding the body rigid, or the pretest ingestion of contraband benzodiazepines or other sedatives.

Amobarbital (Amytal), sometimes mislabeled truth serum, and hypnotic techniques have also proved disappointing candidates in the search for objective arbitration. About 50 percent of tested persons are able to maintain a lie under either of these relaxation techniques, and in fact, some lies can actually become solidified under botched hypnotic circumstances. There is much debate among modern practitioners of hypnosis about whether a person can be forced to reveal, experience, or accomplish anything at all under hypnotic trance that the same person is disinclined to reveal, experience, or accomplish when fully alert.

Psychological testing has proved somewhat more successful than the foregoing techniques in the detection of malingering. By far the most widely employed psychological test for the purpose, and the only one with an empirically validated indicator, is the Minnesota Multiphasic Personality Inventory-2 (MMPI-2). The useful gauge is known as the F – K scale, obtained by subtracting the K raw score from the F raw score; the higher the F – K index, the greater the likelihood that the subject is malingering. With an F – K index of +10, one would be correct about 97.5 percent of the time to assume that the entire MMPI-2 profile was malingered.

The F scale from MMPI-2 alone has also been proposed as an indicator of malingering, particularly when the score is above 100. However, an elevated F score alone could also signal simple uncooperativeness, misunderstanding of the items, a plea for help from an adversarial system perceived as adversarial, or frank psychosis. Thus, it has been suggested that the MMPI-2 may be most fruitfully consulted if both F and F – K scores are computed and the subject is then reinterviewed concerning the endorsed critical items. Such an analysis can shed further light on the reasoning that prompted the observed response pattern.

A further use of the MMPI-2 in the detection of malingering has involved an analysis of the ratio of subtle versus obvious indicators of pathology endorsed. In such an

analysis, as in the setting of a clinical interview, malingerers are more likely to subscribe to blatant, popular indicators of pathology than to more subtle ones.

How useful a Rorschach profile is to the detection of malingering remains controversial. According to conventional wisdom, the Rorschach is easily fooled and therefore is not very useful. However, John Exner, Jr., employing a highly developed scoring system emphasizing interrelations among the underlying structural indicators (texture, color, form, movement, and shading), has claimed that he is much less likely to be tricked. Exner teaches that a highly dramatized and personalized content identifies simulators, while a very constricted profile suggests dissimulation.

Neuropsychological testing has been advanced as an aid in the evaluation of head-injured patients. In one study, the Halstead-Reitan Battery, the Wechsler Adult Intelligence Scale (WAIS), and the first edition of Minnesota Mutiphasic Personality Inventory (MMPI) were administered to 16 volunteer malingerers and 16 nonlitigating head-injured patients, and the test results were sent to 10 neuropsychologists for blind ratings. Although the overall levels of impairment were equivalent in both groups, the patterns of impairment were not. The malingerers scored most poorly on the obvious items, whereas the injured individuals failed more subtle ones. Considering the results of all three tests together, the neuropsychologists correctly classified 100 percent of the malingerers and 94 percent of the nonlitigating injured patients.

Thus, although psychological and neuropsychological testing is helpful, it is not infallible. As for all clinical tests, a given test result should be considered in the context of a battery of converging objective indices, and that battery should be considered as only one entity in the overall clinical evaluative matrix.

DIFFERENTIAL DIAGNOSIS

Malingering must be differentiated from the actual physical or psychiatric illness suspected of being feigned. Furthermore, the possibility of partial malingering, which is an exaggeration of existing symptoms, must be entertained. There also exists the possibility of unintentional, dynamically driven misattribution of genuine symptoms (e.g., of depression) to an incorrect environmental cause (e.g., to sexual harassment rather than to narcissistic injury).

> A 51-year-old manager of a large department store made a critical error in his twenty-third year of service. When two department supervisors, one retiring and the other incoming, disputed who was the rightful recipient of a seasonal bonus check, the manager did not appeal to upper management for a solution. An individual with clear-cut narcissistic personality disorder, the manager took matters into his own hands. He divided the check three ways, one-third going to each supervisor, and one-third going to himself. The manager was put on suspension for inappropriate dispensation of company funds.
>
> This episode was only the most serious of an escalating series of poor decisions, and the company suggested that the manager allow himself to be psychiatrically evaluated at company expense. Outraged at the suspension “after all these years of devoted service” and the suggestion that “I would do anything wrong,” the manager instead quit his position, and sued the company for race discrimination (he was Hispanic) and age discrimination. Psychiatric examination revealed a man suffering from depression, clearly the product of a narcissistic wound, but who appeared utterly convinced of the truth of the misattribution of his misfortune, which had been dynamically driven outward. A diagnosis of malingering would have been as inappropriate in this case as were the plaintiff's allegations of race- and age-motivated discrimination.

It should also be remembered that a real psychiatric disorder and malingering are not mutually exclusive.

> A patient was admitted to a VA hospital with complaints consistent with posttraumatic stress disorder. Although he denied auditory hallucinations, he was troubled by recurrent nightmares and frightening flashback experiences, both involving a terrifying event he had witnessed while serving with the infantry in Vietnam.
>
> Time, and growing trust in the therapist, resulted in an admission by the patient that he experienced persecutory auditory hallucinations and had for years; in fact, his history became increasingly clearly that of a schizophrenic person, and he was assigned that diagnosis. Antipsychotic agents greatly relieved his symptoms.
>
> Ultimately, it was discovered that the patient had never served in Vietnam, had never witnessed any life-threatening trauma, and had never experienced flashbacks or recurrent nightmares. He did, however, continue to benefit greatly from treatment with antipsychotics and was eventually discharged with the diagnosis of chronic paranoid schizophrenia.

Potentially confused with malingering are factitious disorder, Ganser's syndrome, and somatoform disorders (especially conversion disorder). Factitious disorder is distinguished from malingering by motivation (sick role versus tangible pain), whereas the somatoform disorders involve no conscious volition. In conversion disorder as in malingering, objective signs cannot account for subjective experience, and differentiation between the two disorders can be quite difficult; Table 27.2-7 lists some variables that may aid in distinguishing between these two conditions.

1. Malingerers more likely to be suspicious, uncooperative, aloof, and unfriendly; patients with conversion disorder likely to be friendly, cooperative, appealing, dependent, and clinging
2. Malingerers may try to avoid diagnostic evaluations and refuse recommended treatment; patients with conversion disorder likely to welcome evaluation and treatment, “searching for an answer”
3. Malingerers likely to refuse employment opportunities designed to circumvent their disability; patients with conversion disorder likely to accept such opportunities
4. Malingerers more likely to provide extremely detailed and exacting descriptions of events precipitating their “illness”; patients with conversion disorder more likely to report historical gaps, inaccuracies, and vagaries

Table 27.2-7 Factors Aiding in the Differentiation Between Malingering and Conversion Disorder

Ganser's syndrome, or the syndrome of approximate answers, is rare and described primarily in prison populations. It involves the production of answers to questions that are relevant but not quite correct, such as stating that the product of 7 times 4 is 29. It is classified in DSM-IV as a dissociative disorder not otherwise specified, but it has variously been regarded as a factitious disorder, an hysterical phenomenon, a psychotic disorder, and frank malingering. Several studies of undergraduates instructed to feign madness resulted in the malingerers producing approximate answers strikingly reminiscent of Ganser's syndrome. That finding lends credence to Frederick Wertham's 1949 assertion that “A Ganser reaction is a hypothetical pseudostupidity which occurs almost exclusively in jails and in old-fashioned German textbooks. It is now known to be almost always due more to conscious malingering than to unconscious stupefaction.”

COURSE AND PROGNOSIS

Malingering persists as long as the malingerer thinks it to be productive of the desired rewards. In the absence of concurrent diagnoses, once the rewards have been attained, the feigned symptoms disappear. In some structured settings, such as the military or prison units, ignoring the malingered behavior may result in its disappearance, particularly if an expectation of continued productive performance despite complaints is made clear. In children, malingering is most likely associated with a predisposing anxiety or conduct disorder; proper attention to that developing problem may alleviate the child's propensity to malinger.

Malingerers are unlikely to comply with disorder-specific treatments that are offered. If they are confronted with their malingering directly, they are likely to seek out other doctors until they can find a physician who unwittingly complies with their manipulations. Symptoms are likely to abate only when the desired outcome has been achieved or when it becomes clear that the malingering is futile. At that point, malingerers are likely to discontinue treatment abruptly. However, as would be expected, the more the malingering has been reinforced, the more likely it is to recur. Successful malingerers are apt to malinger repeatedly throughout their lives, on

just as many occasions as society willingly rewards them for doing so.

TREATMENT

In his 1940 *Behind the Scenes of Murder*, Joseph Catton describes how he added whiskey to the food of an individual arrested for committing a murder while intoxicated, a defendant who had remained persistently mute and was now consequently a patient. “The insanity defense,” wrote Catton, “was dissolved in five ounces of the same liquid that precipitated the killing.”

Malingerers do not wish to be treated; the last thing they desire is to have their condition diagnosed. They are consciously gaming the system, manipulating it in the hopes of achieving personal gain of some kind.

The appropriate stance for the physician is initial clinical neutrality. If malingering is suspected, a careful differential investigation should ensue. If, at the conclusion of the diagnostic evaluation, malingering seems most likely, the patient should be tactfully but firmly confronted with that outcome. However, he or she must not be abruptly shunned; the reasons underlying the ruse should be elicited and possible alternative pathways to the desired outcome explored. Coexisting psychiatric disorders should be thoroughly addressed. Misattribution patterns and their characterological underpinnings should be gradually exposed. Only if the patient is utterly unwilling to interact with the physician under any terms other than manipulation should the possibility of an ultimate, positive therapeutic outcome be abandoned.

SUGGESTED CROSS-REFERENCES

The somatoform disorders, including conversion disorder, are discussed in Chapter 16; further differentiation of those disorders from malingering may be found in that chapter. Factitious disorders are discussed in Chapter 17. Disturbances of perception, including hallucinations, and of experience, including delusions, are discussed in Chapter 8 on clinical manifestations of psychiatric disorders. Schizophrenia is discussed in Chapter 12, dementia is discussed in Chapter 10, and mental retardation is discussed in Chapter 34. Dissociative amnesia is addressed in Chapter 18. Depressive disorders are reviewed in Chapter 14. Various personality disorders and traits, including borderline, narcissistic, histrionic, and antisocial personality disorder, are elucidated in Chapter 24. Psychological and neuropsychological testing are addressed in Section 7.4, Section 7.5, and Section 7.6.

SECTION REFERENCES

Bagby RM, Nicholson R, Buis T: Utility of the deceptive-subtle items in the detection of malingering. J Pers Assess *70:*405, 1998.

Bash IY, Alpert M: The determination of malingering. Ann NY Acad Sci *347:*86, 1980.

Beck TR: *Elements of Medical Jurisprudence*, vol 1. Websters & Skinners, Albany, NY, 1823.

Ben-Avi I, Rabin S, Melamed S, Kreiner H, Ribak J: Malingering assessment in behavioral toxicology: What, why, and how. Am J Indus Med *34:*325, 1998.

Bok S: *Lying: Moral Choice in Public and Private Life*. Vantage Books, New York, 1978.

*Butcher JN, Miller KB: Personality assessment in personal injury litigation. In *The Handbook of Forensic Psychology*, ed 2, AK Hess, IB Weiner, editors. Wiley, New York, NY, 1999.

Catton J: *Behind the Scenes of Murder*. Norton, New York, 1940.

Chouinard MJ, Rouleau I: The 48-Pictures Test: a two alternative forced-choice recognition test for the detection of malingering. J Int Neuropsychol Soc *3:*545, 1997.

Davidson HA: *Forensic Psychiatry*, ed 2. Ronald Press, New York, 1965.

Davis D, Weiss JMA: Malingering and associated syndromes. In *American Handbook of Psychiatry*, ed 2, vol 3, S Arieti, EB Brody, editors. Basic Books, New York, 1974.

DePaulo BM, Rosenthal R: Telling lies. J Pers Soc Psychol *37:*1713, 1979.

Eissler KR: Malingering. In *Psychoanalysis and Culture*, GB Wilbur, W Muensterberger, editors. International Universities Press, New York, 1951, p 218.

*Ekman P: *Telling Lies*. Norton, New York, 1985.

Exner JE: *The Rorschach: A Comprehensive System*, vol 1: *Basic Foundation*, ed 3. Wiley, New York, 1993.

Ford CV: *Lies! Lies! Lies!!! The Psychology of Deceit*. American Psychiatric Press, Washington, DC, 1996.

Goodwin DW, Alderson P, Rosenthal R: Clinical significance of hallucinations in psychiatric disorders: A study of 116 hallucinatory patients. Arch Gen Psychiatry *24:*76, 1971.

Griffin GA, Glassmire DM, Henderson EA, McCann C: Rey II: Redesigning the Rey screening test of malingering. J Clin Psychol *53:*757, 1997.

*Hadjistavropoulos T: Chronic pain on trial: The influence of litigation and compensation on chronic pain syndromes. In *Handbook of Pain Syndromes: Biopsychosocial Perspectives*, AR Block, EF Kremer, editors. Erlbaum, Mahwah, NJ, 1999.

Hall HV, Pritchard DA: *Detecting Malingering and Deception*. St. Lucie Press, Delray Beach, FL, 1996.

Harris PL, Lipian MS: Understanding emotion and experiencing emotion. In *Children's Understanding of Emotion*, PL Harris, C Saarni, editors. Cambridge University Press, New York, 1989.

Hawk GL, Cornell DG: MMPI profiles of malingerers diagnosed in pretrial forensic evaluations. J Clin Psychol *45:*673, 1989.

Lipian MS: Somatoform disorders. In *The Insurer's Handbook of Psychological Claims*, DR Price, PR Lees-Haley, editors. I.W. Publications, Seattle, WA, 1995, p 83.

Lykken DT: Trial by polygraph. Behav Sci Law *2:*75, 1984.

Marcus EH: The dilemma of the malingering patient-litigant. Am J Forensic Psychiatry *7:*3, 1987.

McKinzey RK, Russell EW: A partial cross-validation of a Halstead-Reitan Battery malingering formula. J Clin Exp Neuropsychol *19:*484, 1997.

Merydith SP, Humphreys JK, Ebener DJ: Motivational distortion of the 16PF by welfare recipients. J Pers Assess *69:*376, 1997.

Osimani A, Alon A, Berger A, Abarbanel JM: Use of the Stroop phenomenon as a diagnostic tool for malingering. J Neurol Neurosurg Psychiatry *62:*617, 1997.

Pollock PH, Quigley B, Worley KO, Bashford C: Feigned mental disorder in prisoners referred to forensic mental health services. J Psychiatr Ment Health Nurs *4:*9, 1997.

Price JR: Malingering and symptom exaggeration. In *The Insurer's Handbook of Psychological Claims*, DR Price, PR Lees-Haley, editors. I.W. Publications, Seattle, WA, 1995.

*Resnick PJ: Malingering. J Forensic Psychiatry *5:*1, 1994.

Resnick PJ: The detection of malingered mental illness. Behav Sci Law *2:*21, 1984.

Rogers R, editor: *Clinical Assessment of Malingering and Assessment*. Guilford, New York, 1988.

*Rogers R, Cruise KR: Assessment of malingering with simulation designs: Threats to external validity. Law Hum Behav *22:*273, 1998.

Rogers R, Harrell EH, Liff CD: Feigning neuropsychological impairment: A critical review of methodological and clinical considerations. Clin Psychol Rev *13:*255, 1993.

*Rogers R, Salekin RT, Sewell KW, Goldstein A, Leonard K: A comparison of forensic and nonforensic malingerers: A prototypical analysis of explanatory models. Law Hum Behav 22:353, 1998.

Rosenhan D: On being sane in insane places. Science *179:*250, 1973.

Schmand B, Lindeboom J, Schagen S, Heijt R, Koene T, Hamburger HL: Cognitive complaints in patients after whiplash injury: The impact of malingering. J Neurol Neurosurg Psychiatry *64:*339, 1998.

Schretlen D: The use of psychological tests to identify malingered symptoms of mental disorders. Clin Psychol Review *8:*451, 1988.

Siegler M, Osmond H: The “sick role” revisited. Hastings Center Studies *1:*41, 1973.

Smith GP, Burger GK: Detection of malingering: Validation of the Structured Inventory of Malingered Symptomatology (SIMS). J Am Acad Psychiatry Law *25:*183, 1997.

Szasz TS: Malingering: Diagnosis of social condemnation. Arch Neurol Psychiatry *76:*432, 1956.

*Trueblood W, Schmidt M: Malingering and other validity considerations in neuropsychological evaluation of mild head injury. J Clin Exper Neuropsychol *15:*578, 1993.

Wang EW, Rogers R, Giles CL, Diamond PM, Herrington-Wang LE, Taylor ER: A pilot study of the Personality Assessment Inventory (PAI) in corrections: Assessment of malingering, suicide risk, and aggression in male inmates. Behav Sci Law *15:*469, 1997.

Wertham F: *The Show of Violence*. Doubleday, New York, 1949.

*Williams RW, Carlin M: Malingering on the WAIS–R among disability claimants and applicants for vocational assistance. Am J Foren Psychology *17:*35, 1999.

Ziskin J: Malingering of psychological disorders. Behav Sci Law *2:*39, 1984.

27.3 ADULT ANTISOCIAL BEHAVIOR AND CRIMINALITY

KENNETH TARDIFF, M.D., M.P.H.

Antisocial behavior refers to a number of behaviors that are immoral, illegal, or both and do harm to other persons and society. Antisocial behavior often involves aggression and violence. In the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) the following statement about adult antisocial behavior appears:

> This category can be used when the focus of clinical attention is adult antisocial behavior that is not due to a mental disorder (e.g., Conduct Disorder, Antisocial Personality Disorder, or an Impulse Control Disorder). Examples include the behavior of some professional thieves, racketeers, or dealers in illegal substances.

The term *antisocial behavior* is sometimes confusing because it refers to behavior both by persons whose behavior is not due to a mental disorder and by persons who have never received an adequate neuropsychiatric workup to determine the presence or absence of a mental disorder. The violence of antisocial behavior must be differentiated from violence due to psychiatric disorders defined by DSM-IV. Physical violence is associated with a number of psychiatric disorders and varies in severity, frequency, and nature, depending on the psychopathology of the specific disorder and the environmental context of the specific incident. For some psychiatric disorders violent behavior is part of the diagnostic criteria listed in DSM-IV, such as intermittent explosive disorder, antisocial personality disorder, borderline personality disorder, and conduct disorder. In other psychiatric disorders violence may not be among the diagnostic criteria but may emerge as part of the expression of the disorder. Among these disorders are substance-related disorders, cognitive disorders, schizophrenia, psychotic disorders, and bipolar disorders.

HISTORY

Although antisocial behaviors have been part of society since history has been recorded, the medical profession's involvement with such behaviors is fairly recent. In the nineteenth century Benjamin Rush described disorders of the will, and J. C. Prichard formulated the concept of moral insanity to describe antisocial behaviors. In the twentieth century the person who committed crimes was thought to have a hereditary defect and was called a psychopath. Later, as a larger role was imputed to social factors in causing crime, such persons were called sociopaths. In 1950 Sheldon Glueck compared delinquents and nondelinquents on hundreds of traits and determined that both heredity and socioeconomic factors were related to antisocial behaviors. August Aickhorn formulated deviant behavior in psychoanalytical terms. He emphasized the faulty development of the superego and noted that psychopaths lacked anxiety or guilt feelings about their behaviors. As researchers studied criminal populations, it became apparent that criminals had a broad range of psychiatric disorders. In addition to psychopaths, persons with schizophrenia, mental retardation, psychoneurosis, and alcoholism were in prison populations.

EPIDEMIOLOGY

In 1933 the rate of homicide in the United States was 9.7 homicides per 100,000 persons. The rate of homicide declined in the 1950s but began to climb in the late 1960s and 1970s and peaked at 10.7 per 100,000 persons in 1980. From the high in 1980 homicide rate declined to a low of 7.9 per 100,000 persons in 1985. This dip was short-lived, as the homicide rate climbed steadily to a high of 9.5 per 100,000 in 1993. In contrast, rates of nonfatal assaults, monitored by the National Victimization Survey since 1973, did not fluctuate as much as the homicide rate between the early 1970s and early 1990s.

In other words, violence per se did not increase, only lethal violence increased from the mid-1980s to the early 1990s. This was probably due to a number of factors. First, there was greater access to firearms, particularly the more lethal automatic and semiautomatic weapons. Access to these weapons and carrying weapons increased especially among young males. Not surprisingly, homicide rates increased most among young males. For example, between 1965 and 1985, homicide rates for 15- to 19-year-old males were one-third to one-half the rates for men 20 to 34 years of age. Between 1985 and 1992, annual rates of homicide for 15- to 19-year-old males rose 154 percent. Victims of these young perpetrators were more likely to be strangers or acquaintances than family members. The rise of crack cocaine from the mid-1980s was associated with chaotic drug dealing, with violence related to territorial disputes and robberies and death of drug dealers, customers, and innocent bystanders. African-Americans and Latinos are overrepresented as homicide victims and perpetrators, probably because of socioeconomic disparities, not race per se. These include poverty, low education, unemployment, and living in crime-ridden neighborhoods.

Men are more likely than women to be victims of violent crime, except rape and sexual assault. In 1994 the National Victimization Survey found the rate of nonfatal assault to be 5100 per 100,000 men and 3500 per 100,000 women 12 years and older. Rates of homicide were 18 per 100,000 men and 4 per 100,000 women. Rates of robbery were 800 per 100,000 men and 400 per 100,000 women. Rates of rape and sexual assault were 400 per 100,000 women and 20 per 100,000 men. Rates of women being victims of violent crime have increased over the past 20 years. Women are more likely than men to be attacked by intimates, relatives, or acquaintances.

There has been good news in regard to violent crime in the United States over the past few years: it is decreasing. From 1993 to 1996, overall violent crime dropped 16 percent; nonfatal assaults decreased 27 percent, rape decreased 44 percent, and other sexual assaults decreased 37 percent. According to the 1998 Federal Bureau of Investigation report, there was a downward trend in murders of about 8 percent from 1997 to 1998. Speculation about the reason for the decrease in violent crime involved more-aggressive police work with enforcement of “quality-of-life” crimes stabilizing neighborhoods and decreasing drug-dealing, particularly the violence-prone selling of crack cocaine.

Although accounting for few homicides, mass murders and serial murders receive disproportionate attention in the media. In fact, some speculate that media attention has caused an increase in these homicides over the past two decades. The first mass murder that received widespread coverage by the media in the United States was the sniper attack by Charles Whitman from a tower at the University of Texas. Since then, mass murders have been media events and the number of dead bodies are compared with that in earlier mass murders. Mass murderers who followed Whitman in the United States have acquired more-lethal weapons that are easier to carry and conceal. While Whitman lugged a case of rifles and ammunition to the top of the tower, more-recent mass murderers use small automatic pistols or automatic rifles. Consider the Long Island Railroad mass murder in which Colin Ferguson carried an automatic pistol and clips of ammunition in a small knapsack. Originally, a mass murderer was most likely to be a white man preoccupied with guns and militaristic themes. The mass murderer was often a disgruntled employee or some person who felt discriminated against and paranoid about society. Most recently, mass murderers have been children who feel ridiculed or rejected by classmates and teachers in schools and who have easy access to guns. The Littleton, Colorado, massacre in 1999 by two high school students who killed 13 people and then killed themselves may fall into this category.

Serial murderers usually kill one victim at a time over a period of time rather than a group of people at one time. Serial murderers have existed in England and Europe since the nineteenth century (e.g., “Jack the Ripper”). In the United States the first infamous serial killer in this century was Albert DeSalvo, the “Boston Strangler.” Serial murderers often kill for sexual reasons, and the victims often resemble each other. Ted Bundy raped and killed many young women who resembled a former fiancé who called off their engagement. John Wayne Gacy and Jeffrey Dahmer primarily molested and killed young men. Along with sexual themes, there may be cult or satanistic themes as in the cases of Charles Manson and David Berkowitz, the “Son of Sam” killer. Another type of serial murder involves killing elderly persons in rooming houses for pension or Social Security checks, as in the case of Dorothy Perente in California.

A third type of homicide that constitutes a small number of homicides but receives wide media coverage is the murder-suicide. As with mass murders and serial murders, those who murder someone and then commit suicide are usually men. There are several types of murder-suicide. The most common type, representing one-half to three-fourths of murder-suicides in the United States, involves amorous jealousy. Typically a man develops suspicion or knowledge of his spouse's or girlfriend's infidelity. He kills her and perhaps her lover and commits suicide with a gun. Another type of murder-suicide involves an elderly man who is in poor health or has an ailing wife. A new variant of this type is a person with acquired immune deficiency syndrome (AIDS) who kills a lover and then himself. A third type involves the murder of a spouse, children, or other family members and the suicide of a depressed male head of a household or a mother of a child. Last, a paranoid disgruntled mass murderer may kill himself or put himself in a position so that the police kill him.

ETIOLOGY

Socioeconomic Factors Criminal violence in deprived inner cities has been explained in terms of recourse to fighting as a means of achievement. The breakup of families, alienation, discrimination, and frustration are believed to be compounding factors. Some have hypothesized that blacks live in a violent subculture; others have found there is no difference between blacks and whites in domestic violence if socioeconomic status is controlled for in the analysis.

In a study that compared rates of violent crimes in large standard metropolitan statistical areas, racial inequality and stress were associated with higher rates of homicides. Economic inequality denoted not merely poverty, but also relative income differences among individuals. Violence may occur because of hostility in persons who perceive that they are economically disadvantaged relative to other persons. Some studies have found that economic inequality is not related to violence but that absolute poverty is related to violence and other criminal behavior.

These contradictory results are probably due to the use of large metropolitan areas as units of analysis. A study by the author used the smaller, more naturalistic units of neighborhoods to test the hypothesis that economic inequality is related to homicide. The study found that economic inequality and race were not related to homicide and that the prime determinants of homicide were absolute poverty and marital disruption.

The social determinants of violent crimes are interconnected. Poverty, inability to acquire the basic necessities of life, marital disruption, single- or no-parent families, substance abuse, unemployment, and lack of education lead to social disintegration, decreased social control, and violence.

Physical Factors Physical crowding is probably related to violent crime in terms of the number of buildings in a geographic area or the number of inhabitants in a building. Some believe that with increased contact in high-density areas, there is less defensible space and an increase in violence; whereas others argue that an increased density affords increased social control and decreases violent crime. The latter situation may be obtained if bystanders serve informally or formally in a neighborhood crime-watch program.

Metropolitan areas with central cities have the highest rates of homicide, followed by nonmetropolitan counties, and then metropolitan areas without central cities. It appears that the difference between urban and rural areas in primary homicides is less than the difference in secondary homicide rates, or homicides committed secondary to crime, which are higher in urban areas.

Biological Factors Genetic inheritance may be related to criminal acts in general. Studies have found more criminal behavior in monozygotic twins than in dizygotic twins. Insofar as twins share the same environmental and genetic backgrounds, further research has turned to adoption studies. In two studies reported from Scandinavia there was no support for a genetic basis to homicide and some support for an association with other crimes. There appears to be no specific chromosomal abnormality that accounts for violent behavior, and studies of inheritance support only a genetic relation for economic, property crimes, not for violence per se. There is evidence that low serotonin and corticotropin concentrations in the brain and high free testosterone concentration in cerebrospinal fluid may be associated with episodic violence.

Developmental Factors Several retrospective studies and one prospective study have shown that being abused as a child is related to becoming a physically abusive adult. Further, witnessing intrafamilial violence (e.g., spouse abuse) is related to increased problems with violence among children. Abuse as a child and growing up with domestic violence are not the sole causes of violence in adults. Other factors affect whether there is violence in a family, such as the presence of psychiatric disorders, alcohol and drug abuse, or other pathology in the perpetrator, or socioeconomic factors such as poverty, unemployment, education, or culture. The role of the media, such as violent television programs, movies, and music videos, has come under increased scrutiny as a factor in the development of violence and other antisocial behaviors.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of adult antisocial behavior is made by a process of elimination. Mental disorder must be ruled out as a cause of or factor contributing to violence by the person. The mental disorders most frequently associated with violence are intermittent explosive disorder, antisocial personality disorder, borderline personality disorder, conduct disorder, schizophrenia, other psychotic disorders, bipolar I disorder, substance-related disorders, and cognitive disorders.

DIFFERENTIAL DIAGNOSIS

Intermittent Explosive Disorder Intermittent explosive disorder is defined in DSM-IV as an impulse-control disorder manifested by several discrete episodes of loss of control over aggressive impulses, resulting in serious assaultive acts or destruction of property. The degree of aggressiveness expressed during the episodes is grossly out of proportion to any precipitating psychosocial stressors. There are few or no signs of generalized impulsiveness or aggressiveness between episodes. The person usually feels remorseful and concerned about the consequences of the violence, such as divorce and loss of a job. The episodes of loss of control are not better accounted for by antisocial personality disorder, borderline personality disorder, conduct disorder, a psychotic disorder, a manic episode, and attention-deficit/hyperactivity disorder and are not due to substance intoxication or personality change due to a general medical condition (e.g., head trauma).

A 40-year-old male lawyer sought outpatient psychiatric help for a long-standing problem with his temper. He and his wife would have serious arguments every week or so. On a number of occasions he had pushed her out of his way as he attempted to leave the house and “cool off.” After arguments he would say how sorry he felt and plead with his wife to forgive him. The arguments had increased in number and intensity since the birth of their son 2 years earlier. A recurring theme in the arguments was his concern that his wife did not care about him and paid too much attention to the son. The wife felt she had sacrificed her professional career and that her physical appearance had deteriorated after the birth of the child. A few weeks before the first consultation the patient had pushed his wife, who sustained a laceration on her head. She insisted that he leave and that they separate legally.

The patient was an only child and felt resented by his mother, who had to stop working to take care of him. His father was a successful lawyer who spent little time with the family. The patient had reading problems throughout school. He was a bully since late childhood and would get into fights with other, usually smaller boys.

Over a year of psychotherapy exploring issues of self-esteem, in conjunction with a moderate dosage of clonazepam (Klonopin), he was able to reconcile with his wife without physical attacks on her.

Antisocial Personality Disorder Patients with antisocial personality disorder repeatedly get into fights or exhibit other assaultive behaviors, including spouse beating and child beating. The assaultive behavior is usually evident by age 15 years. Other behaviors that manifest by middle to late adolescence include unlawful behaviors such as destroying property, harassing others, stealing, and pursuing an illegal occupation. The person cannot sustain legitimate employment or academic performance. The person may fail to honor financial obligations, has no regard for the truth, cannot function as a responsible parent, has difficulty being monogamous, is reckless in regard to the safety of self or others, and is generally impulsive and fails to plan ahead. By definition, in DSM-IV conduct disorder is present before the age of 15 years and is associated with a number of problematic behaviors such as physical aggression, cruelty to other people or animals, destruction of property, fire setting, lying, truancy, stealing, and forcing someone to have sexual activity. The hallmark of antisocial personality disorder is lack of remorse in regard to the violence and other antisocial behaviors.

A 20-year-old unemployed man came to the emergency room at 3:00 AM expressing vague suicidal ideation. He had recently lost his job; after being accused of stealing money from the cash register in the hardware store where he worked, he had hit his supervisor with a wrench. The supervisor lost consciousness and sustained residual neurological dysfunction. The patient had worked at the hardware store for only 2 months. A fellow employee called the police, and the patient was arrested, released on bail, and given a trial date. He claimed to have no memory of the event.

During his emergency room visit the patient sought admission to the psychiatric unit for an evaluation to determine whether mental problems could have been responsible for his attack on his supervisor. He had been arrested twice for driving while intoxicated. He had been in numerous fights, usually in bars, and bragged about injuries he had inflicted on opponents. Since late childhood he had been in fights, and he had frequent behavioral and academic difficulties in school. He quit high school in his junior year and held many jobs for short periods of time.

The patient was admitted to the inpatient unit and was discharged 2 days later with no Axis I diagnosis. During his short stay on the unit, he created conflict among patients and staff.

Borderline Personality Disorder Borderline personality disorder includes frequent displays of intense anger and lack of control of anger, which results in recurrent physical fights. Violence is only one area of impulsiveness. Persons with borderline personality disorder attempt suicide, are self-mutilating, and are impulsive in regard to spending money, sexual activity, shoplifting, reckless driving, and other aspects of living. There are problems with identity disturbance and chronic feelings of emptiness. Severe violence targeted at one person may come from intense and unstable interpersonal relationships and fear of real or imagined abandonment by that person.

A 25-year-old attractive woman had been in weekly psychotherapy with a male therapist for 2 years. The content of the sessions involved her sense of loneliness and lack of direction in her professional life. She had numerous unsuccessful relationships with men, usually older than she and married or otherwise not available. The relationships ended with increasing demands by her for attention and, on several occasions, suicide attempts. She would break dishes, vases, and other objects during rages but had never attacked people.

Over the past 3 months she had begun to talk of increasing the frequency of psychotherapy sessions. Vague references by the patient to where the therapist lived and inquiries about his marriage turned to overt insistence during one session that they should have sex because he was the first man to really understand her. The therapist abruptly terminated the session by asking his secretary to come into the office. That night the patient appeared at the therapist's home with a gun. The police were called. Before they arrived she shot out a number of windows in the therapist's house and then shot herself.

Conduct Disorder Conduct disorders can involve violence to varying degrees of severity, ranging from use of a weapon that could seriously injure others, such as a gun, knife, bat, brick, or broken bottle, to less severe physical fights with other persons in the home. Conduct disorder precedes antisocial personality disorder. Children and young adolescents with conduct disorder are truant, destroy property, steal, are physically cruel to people or animals, force sex on others, or engage in sexual activity to obtain money, drugs, or other goods. Less severe manifestations of those antisocial behaviors can be found along with the violence in conduct disorder.

A 9-year-old girl in a private school was frequently sent to the principal's office. In class she would not remain quiet and often used obscene words. She touched other students in class and picked on certain students, whom she hit or poked with pencils. Those actions resulted in fights and name calling. She was caught stealing minor items from another girl's locker. She bragged that she had seen X-rated videotapes and seemed too sophisticated about sex for her age. The family dog had been injured a number of times, and her parents suspected that she was responsible. The girl would blame others for her behavior and frequently lied to her teachers and parents.

Psychotic Disorders Schizophrenia and other psychotic disorders are associated with an increased risk of violence, but violence is not found in most cases. Violence can result from the nature of the delusional system, from hallucinations, or as part of disorganized or agitated behavior. As part of paranoid delusional thinking and hallucinations, patients may believe that other persons are threatening, persecuting, or in some way trying to harm them; the violence is a reaction to the perceived threats. Often the theme of the delusion is consistent over time. In the disorganized and catatonic subtypes of schizophrenia violence comes from psychotic disorganization and extreme agitation and excitability. Some patients may have difficulty delaying gratification and a low tolerance for frustration, with violent consequences.

Sometimes violence is caused by factors other than the schizophrenic process per se. For example, this author has observed violence as an indirect complication of antipsychotic medication when patients with akathisia restlessly move around the inpatient unit. Moreover, schizophrenic patients may have another disorder primarily responsible for violence, such as neurological impairment, substance abuse, or mental retardation. Violence may not be related to any psychiatric disorder but rather to retaliation or self-defense. Patients with delusional disorder may take violent action to right what they believe to be a wrong. Systematized, prominent, and fixed delusions of a spouse's infidelity and persecutory delusions have occasionally resulted in severe violence, including homicides.

A 27-year-old woman with a diagnosis of paranoid schizophrenia had been admitted to a number of psychiatric hospitals. Her first episode occurred during her freshman year at college; she believed her food was being poisoned. Her parents denied the seriousness of her illness and persisted in trying orthomolecular treatment.

In the hospital the patient was very psychotic and had multiple delusions regarding food intake, including beliefs that the staff and other patients were trying to poison her. That belief led to several attacks on other patients, requiring physical restraints and emergency medication. The violence subsided when she was allowed to stay in her room for long periods of time, including mealtimes, until antipsychotic medication was effective in diminishing her delusions.

A 24-year-old man with a diagnosis of paranoid schizophrenia was attending a day hospital program and was taking oral antipsychotic medication, but compliance was problematic. He continued to hear the voice of a friend telling him to cut himself and to hurt others because he liked to see blood. On the day of admission he had threatened to kill a therapist at the day hospital with a razor blade; a razor blade was found in his pocket. In the hospital his psychosis responded to antipsychotic medication, and he was no longer delusional. Nevertheless, on two occasions he was violent. On one occasion he swung a curtain rod at a nurse in a menacing manner because he was denied a pass. The next day he sought out the nurse who told him he could not have a pass. He punched her in the face and loosened two of her teeth. When asked why he did it, he replied that he wanted to get even for the denial of a pass. He was not delusional. The nurse, with the support of the administration, pressed charges, and he was arrested by the police.

Bipolar I Disorder Manic patients may experience sudden, severe episodes of violence. Often violence erupts in the early phase of treatment of the manic episode. Such a patient may respond violently to any form of containment or limit setting, whether physical or otherwise, as when a nurse insists that the patient take a dose of medication or a clinician insists on hospitalization against the patient's will.

A 48-year-old woman came to the emergency room accompanied by her husband. She had a history of bipolar I disorder. That day in church she had been singing and talking very loudly when an usher asked her to be quiet; she got into a physical fight with him. In the previous week she had become increasingly euphoric and hypersexual. She bit her husband several times on the arm when they were having sexual relations.

The patient and her husband were escorted into the psychiatric examination room and seen by a female resident. About 5 minutes into the interview, as the husband was showing the resident the bite marks on his arm, the patient grabbed the resident by her hair and began to pound her head against the wall. The resident had left the door to the room open and had asked a security guard to sit outside the door. The security guard rushed in and restrained the patient. When asked a month later why she had attacked the resident, the patient said she felt restricted in the room and was angry that the resident was talking to her husband instead of to her.

Substance-Related Disorders Substance abuse is associated with violence. Alcohol-related disorders as a primary disorder or in conjunction with other psychiatric disorders can result in violence. The violence is often the result of disinhibition (particularly in the early phase of intoxication) and of emotional lability and impaired judgment. Stimulants, particularly cocaine, can be associated with violence. The early phase of cocaine intoxication is manifested by a euphoria that can turn to irritability, agitation, suspiciousness, and violence, especially with crack or intravenous cocaine. With continued use of cocaine in a binge and with long-term amphetamine use, suspiciousness can become paranoid delusional thinking and violence. Heavy cocaine use has caused manic-like delirium with severe violence.

The craving for cocaine may result in violence as the user desperately seeks more cocaine or the money to purchase it.

Hallucinogens and dissociative drugs, particularly phencyclidine, can cause violence, suicide, and bizarre behaviors. Hallucinogens can impair judgment and cause perceptual changes and delusional thinking. Benzodiazepines and sedatives do not commonly produce violence through intoxication. However, abrupt discontinuation of the short-half-life benzodiazepines may result in irritability and agitation, so that the risk of violence may be increased. Withdrawal from alcohol may produce violence by similar mechanisms or by the delirious state.

> A 27-year-old woman had a 2-year history of using crack cocaine. Prior to that she had abused alcohol and benzodiazepines. She had been hospitalized numerous times. Before the current hospitalization she had been serving as the cooker in a crack house, converting crystalline cocaine into crack. For that service she was provided with a continuous supply of crack over a number of hours. She recalled becoming increasingly irritable and drinking wine in an effort to calm down. She became suspicious that others in the group were taking more and more of the cocaine and that she would be left without an adequate amount and would “crash.” She had no recollection of what happened next, but was told by a friend that she suddenly picked up an ashtray and hit a man on his head several times. When examined at the time of admission she was delusional but responded well to a low dose of an antipsychotic agent, which was stopped after 1 week.

Cognitive Disorders Organic brain disease can result in violence. Violence in patients with temporal lobe epilepsy is not frequent but may occur during the ictal period and, if so, is often purposeless. Some researchers have maintained that more-subtle neurophysiological dysfunction is responsible for violence in some patients. They point out that the electroencephalogram (EEG) is an insensitive measure of subcortical activity and that some patients with episodic violence respond to anticonvulsants. Others stress the effect of head trauma and other neurological and medical disorders that can affect the brain and produce violent behavior. Infections of the brain, including viral encephalitis, AIDS, tuberculosis, fungal meningitis, syphilis, and infection with human herpesvirus, may be associated with violent behavior. Other diseases of the brain associated with violence include normal pressure hydrocephalus, cerebrovascular diseases, tumors, Huntington's disease, multiple sclerosis, Pick's disease, vascular dementia, Alzheimer's disease, Parkinson's disease, and Wilson's disease, as well as postanoxic or posthypoglycemic states with brain damage.

A wide range of medical illnesses have been associated with violent behavior. Many are treatable and reversible. Among those disorders are hypoxia, electrolyte imbalances, hepatic disease, renal disease, vitamin deficiencies (e.g., B_{12}, folate, thiamine), systemic infections, hypoglycemia, Cushing's disease, hyperthyroidism, hypothyroidism, systemic lupus erythematosus, poisoning (e.g., by heavy metals, insecticides) and porphyria.

> A 29-year-old recreational therapist was brought to the emergency room by her sister. A week earlier her sister noticed that the patient did not wake up at her usual time to go to work. When awakened by her sister, the patient seemed upset and confused and could not remember where she worked. Over the next few days she spent a great amount of time sleeping, alternating with sexually inappropriate behavior such as disrobing in public. She became physically violent toward her sister, who tried to prevent such behavior and who insisted she be evaluated by a psychiatrist. Laboratory tests, including a toxicology screening test, EEG, and magnetic resonance imaging (MRI), revealed no abnormalities except an elevated erythrocyte sedimentation rate and mild anemia. Before the episode the patient had been healthy, except that 3 weeks earlier she had flu-like symptoms with a runny nose, malaise, and elevated temperature. On the inpatient unit she was impulsive and unpredictably violent. She would punch other patients, pull their hair, and knock food trays from tables. She was delusional in that she believed she was on a maternity ward and pregnant, which she was not. She responded to a low dosage of an antipsychotic agent, and her violence subsided with carbamazepine (Tegretol) therapy. The most probable diagnosis was psychotic disorder due to postviral encephalitis.

LABORATORY EXAMINATION

Laboratory tests and imaging studies are aimed at detecting underlying brain pathology, which may be localized, particularly to the temporal and frontal lobes, as well as general neurological and medical disorders in which violence is part of general disorganization, dementia, or delirium. The usual routine laboratory tests are done, including a complete blood cell count, a blood chemistry panel (to determine electrolytes; blood urea nitrogen [BUN]; glucose, creatinine, calcium, and phosphate concentrations; and liver function), thyroid function tests (serum triiodothyronine [T_3], serum thyroxine [T_4], serum thyroid-stimulating hormone [TSH]), a syphilis tests, vitamin deficiency testing (B_{12}, folate, thiamine), urinalysis, electrocardiography, and chest X-rays. This author believes that EEG should be part of the routine evaluation of the violent patient. The routine EEG should be augmented by recording with nasopharyngeal leads and by a sleep and a sleep-deprived EEG. In addition, blood and urine specimens should be taken at the time of admission for alcohol and drug assays. Because cocaine has a short half-life—about an hour—blood should be drawn as soon as possible and placed in a container with fluoride to prevent subsequent breakdown of cocaine in the tube.

The author believes that MRI should be used routinely in assessing violent patients. MRI is better than computed tomography (CT) in detecting temporal lobe pathology. CT does not require the same degree of patient cooperation and therefore can be used more easily than MRI in agitated, hyperactive patients.

Other laboratory tests are used if a specific disorder is indicated. Such tests include arterial blood gas determinations (to test for hypoxia), an assay for human immunodeficiency virus (HIV) antibodies, (to test for HIV infection and AIDS), a glucose tolerance test (to test for hypoglycemia), determination of ceruloplasmin and copper concentrations in urine and serum (to test for Wilson's disease), urine porphobilinogen assay (to test for porphyria), a heavy metal screening test (to test for poisoning), antinuclear antibody assay (to test for systemic lupus erythematosus), and lumbar puncture if intracranial pressure is not elevated (to test for infection, multiple sclerosis, and hemorrhage).

PROGNOSIS

When a patient has exhibited violent behavior in the past or threatens violence, a decision must be made about the risk of violent behavior in the future. Decisions must be made about whether to hospitalize a potentially violent patient, when to discharge a previously violent patient, and whether to warn a potential victim. As with the evaluation of suicide potential, evaluation of violence potential includes assessment of how well planned the threat or violent ideation is. Vague threats of killing someone are not as serious as well-formulated threats against a specific person. As with suicide, the availability of means of inflicting injury is important. If the patient has recently purchased or owns a gun, the clinician should take the threat more seriously.

A history of violence or other impulsive behavior often predicts future violence. The clinician should assess the degree of past injuries, the person toward whom violence was directed, and the circumstances. Often there is a repetitive pattern of violent behavior and escalation. Alcohol and drug abuse should be assessed. Central nervous system disorders have been associated with violent behavior, as have some systemic disorders affecting the central nervous system. As with suicide, the clinician should take seriously threats of violence by a psychotic person and assess the potential for violence. Compliance with treatment, for example, a paranoid schizophrenia patient receiving depot antipsychotic agents or an alcoholic patient receiving disulfiram (Antabuse), is reassuring. All of those factors are weighed in the final assessment of whether the patient poses enough risk to others to require some action (Table 27.3-1).

High degree of intent to harm
Presence of a victim
Frequent and open threats
Concrete plan
Access to instruments of violence
History of loss of control
Chronic anger, hostility, or resentment
Enjoyment in watching or inflicting harm
Lack of compassion
Self-view as victim
Resentment of authority
Childhood brutality or deprivation
Decreased warmth and affection in home
Early loss of parent
Fire-setting, bed-wetting, and cruelty to animals
Prior violent acts
Reckless driving

Table 27.3-1 Commonly Cited Predictors of Dangerousness to Others

TREATMENT

Safety The immediate and long-term management of violent patients is often made more difficult by staff emotions and fears about safety. This author has developed a model to predict the short-term risk of violence by patients. If the risk is significant the clinician is expected to make efforts to protect potential victims from violence by the patient by changing the treatment plan, hospitalizing the patient, or warning the potential victim and the police. In such situations confidentiality can be breached, and the potentially violent patient should be informed that a warning will take place. It should be emphasized that warning serves both to protect the patient from the consequences of his or her violence, should it occur, and to protect the intended victim.

The safety of the clinician and the environment in which the treatment takes place should also be considered. In emergencies the clinician should not intervene unless it is safe to do so. Adequate manpower should be available to contain violence by patients. Attention to attire is important; clinicians should remove earrings, necklaces, and neckties before confronting potentially violent patients. Heavy or sharp objects in the office and intravenous poles in the emergency room are dangerous. The clinician must feel safe as the evaluation and treatment of the violent patient proceeds or else the process will be incomplete or distorted.

Emergency Management There are three ways to manage violence or impending violence: verbal deescalation, medication, and physical restraint or containment. A decision to attempt to deescalate aggression by talking to the patient is made on an individual basis, but a psychotic patient or one with a gross organic dysfunction as the cause of violent behavior rarely responds to verbal interventions. The options then are restraint, seclusion, and emergency medication. The choice is based on clinical needs in the individual case. If the cause of the violent behavior is unknown, restraint may be indicated to maintain the patient drug free for purposes of evaluation. A violent patient may be preferentially managed with seclusion and restraint because of medical illness or drug allergies that would preclude the use of certain medications to control violent behavior.

Seclusion and Restraint Seclusion and restraint should never be used for punishment, that is, as retribution for an act when no danger exists to the patient or others. Seclusion may be contraindicated because of the patient's clinical or medical condition. If the patient is suicidal or medically unstable so that close monitoring is necessary, seclusion is not appropriate. Restraint may be preferable for delirious patients, in whom the reduced sensory input of seclusion may lead to worsening of the clinical state. Seclusion is also relatively contraindicated in patients who have just taken overdoses and need close monitoring, in patients with symptoms of serious and uncontrollable self-mutilation, and when seclusion rooms cannot be sufficiently cooled on hot days for patients receiving drugs, such as antipsychotic agents, that impair thermoregulation.

Restraint and seclusion are emergency procedures that are usually initiated by nursing staff, but a physician should see the patient as soon as possible and write an order to continue the procedure. Orders are time-limited and must be renewed periodically, depending on the institution's guidelines. The American Psychiatric Association's *Task Force on the Psychiatric Uses of Seclusion and Restraint* outlined standards for the monitoring and care of a patient in restraint and seclusion by nursing staff to ensure the patient's safety.

Emergency Medication Emergency medication may be used instead of or with seclusion or restraint. Antipsychotic medication should be used primarily in the management of violent patients with psychotic symptoms. It may also be indicated for patients who are not psychotic but who are violent, as in the case of patients with brain dysfunction, whose condition may be worsened by benzodiazepines. The procedure of rapid tranquilization (i.e., administering low doses of antipsychotic medication over 30 to 60 minutes) with the specific goal of decreasing violence is effective. A range of antipsychotic agents can be used, and there is general safety and efficacy with rapid tranquilization. The approach is effective in the emergency management of schizophrenia, manic episodes, and other psychotic disorders.

There are two strategies of rapid tranquilization: the use of high-potency antipsychotic drugs with an anxiolytic drug or other drugs for sedation, if necessary, and the use of low-potency antipsychotic drugs with both sedative and antipsychotic effects. The strategy chosen should take into consideration the patient's history and physical status. The clinician should review the patient's previous responses to medication and any coexisting medical illnesses that may predispose to or be exacerbated by extrapyramidal side effects or orthostatic hypotension. If extrapyramidal adverse effects are less acceptable, low-potency antipsychotic drugs should be used; if orthostatic hypotension and related adverse effects are less acceptable, high-potency antipsychotic drugs are indicated. The clinician should feel free to change the strategy if the patient's medical condition changes.

The benzodiazepines can be used effectively in emergency situations. Benzodiazepines may be used with antipsychotic medications in patients with schizophrenia or manic patients and patients in other psychotic states, or they may be used alone in the management of nonpsychotic patients. For patients who appear to have some degree of control, they may be offered as an oral medication. However, most emergency situations require intramuscular medication. Lorazepam (Ativan) is preferred because, unlike diazepam (Valium) or chlordiazepoxide (Librium), it is reliably and rapidly absorbed from intramuscular injection. Clonazepam is more appropriate in the long-term management of violence than in emergency situations.

Long-Term Medication Medication is used to treat the underlying disorders (e.g., antipsychotic drugs to treat schizophrenia and lithium [Eskalith] to treat manic episodes). Anticonvulsants, b-adrenergic receptor antagonists, and lithium have been used to treat violence in patients who do not have classic seizure activity or bipolar disorders.

Anticonvulsants Carbamazepine is effective in managing aggression and irritability in patients with overt seizures, in schizophrenic patients with and without EEG abnormalities, and in other patients with episodic violence, particularly those with personality disorders without gross brain damage and mental retardation. Carbamazepine rarely causes agranulocytosis or aplastic anemia, but the patient's hematological status must be monitored. A gradual increase in dosage in 100- to 200-mg increments to 600 to 800 mg a day, or a therapeutic serum concentration of 8 to 12 ng/mL, is recommended. Valproic acid (Depakene, Depakote) may be used after an unsuccessful trial of carbamazepine or if the hematological status of the patient precludes use of carbamazepine. Therapeutic levels range from 50 to 100 g/ml. Hepatic function should be monitored with both medications.

b-Adrenergic Receptor Antagonists A number of reports have indicated the effectiveness of b-adrenergic receptor antagonists (also called beta-blockers), particularly propranolol (Inderal), in the management of aggressive behavior. Most patients who responded to propranolol were those with brain disease, often with gross impairment of function secondary to trauma, alcohol use disorders, encephalitis, Huntington's disease, dementia, Wilson's disease, and Korsakoff's syndrome. Some patients with minimal brain dysfunction or attention-deficit/hyperactivity disorder have also reportedly responded to propranolol. Nearly all patients in those studies were refractory to other medications, including antipsychotic drugs, anxiolytics, anticonvulsants, and lithium. In a number of instances concurrent antipsychotic medication was used. Before propranolol is prescribed, the patient should receive a thorough medical evaluation. Propranolol should not be used in patients with bronchial asthma, chronic obstructive pulmonary disease, insulin-dependent diabetes, cardiac diseases (including angina and congestive heart failure), diabetes mellitus, significant peripheral vascular disease, severe renal disease, and hyperthyroidism. Propranolol should be prescribed with caution for hypertensive patients because sudden discontinuation of propranolol may result in rebound hypertension.

The initial dosage of propranolol is 20 mg given three times a day; the dosage is increased by 60 mg every 3 to 4 days unless the pulse rate is below 50 beats per minute or the systolic blood pressure is less than 90 mm Hg. If the patient experiences severe dizziness, ataxia, or wheezing, the medication should be held and then reduced or discontinued if the symptoms persist. Dosages above 800 mg a day usually are not needed to control aggressive behavior. Other b-adrenergic receptor antagonists such as pindolol (Visken), nadolol (Corgard), and metoprolol (Lopressor) have been used successfully to treat aggressive behavior.

Lithium Lithium has been used to treat a number of different types of violent patients. In a double-blind trial lithium was effective in treating aggression in mentally retarded adult patients. The administration of lithium was the same as that for the management of bipolar I patients. Lithium has also been reported to be effective in patients with a brain syndrome and head injury; aggressive schizophrenic persons; nonpsychotic, aggressive prisoners and delinquents; and children with conduct disorders and attention-deficit/hyperactivity disorder.

Behavior Therapy Behavior therapy has been used in the management of violent behavior in severely impaired institutional populations. It may be used for patients with chronic schizophrenia or patients with mental retardation in conjunction with other treatment approaches. In setting up a behavior treatment program on an inpatient unit, the services of a trained, experienced behavior analyst should be obtained. The behavior analyst is responsible for planning the treatment program in conjunction with the interdisciplinary staff and training and supervising staff in the implementation of the treatment program. Behavior treatment programs should be reviewed periodically by persons other than the ward personnel to ensure their quality and ethical standards. In addition, the patient should be involved as much as possible in the formulation of the plan, and informed written consent should be obtained if the patient is competent to offer it.

Dynamic Psychotherapy Long-term psychotherapy can be effective for violent patients who are nonpsychotic and whose primary diagnosis is personality disorder or

intermittent explosive disorder. If spouse abuse is the problem, couples sessions may be indicated, provided the spouse's safety can be ensured. The first goal of therapy is to evaluate whether the patient is motivated, as in situations in which divorce or loss of a job is pending. Next, the patient must identify patterns of escalation in violence and learn to disengage at the early phase of escalation. The therapist should closely monitor transference and be prepared to deal promptly with negative thoughts or threats by the patient. Countertransference must also be monitored, because violence can evoke inappropriate feelings and reactions to patients. Psychotherapy should provide insight into why the patient feels compelled to use violence as a means of expression. Often such patients have difficulty expressing their feelings and resolving conflicts verbally. Issues common in the treatment of violent patients involve low self-esteem in terms of money, sex, and authority.

PREVENTION

Because antisocial behavior often begins during childhood, the clinician should focus on delinquency prevention. Any measures that improve the physical and mental health of socioeconomically disadvantaged children and their families are likely to reduce delinquency and violent crime. Because many recurrently violent persons have sustained many insults to the central nervous system, starting prenatally and continuing through childhood and adolescence, programs must be developed to educate parents about the dangers to their children of central nervous system injury, including the effects of psychoactive substances on the brain of the growing fetus. Public education regarding the releasing effect of alcohol on violent behaviors (not to mention its contribution to vehicular homicide) may also reduce crime.

Although people disagree about the contribution of violence in the media to violent crime, there is universal recognition that the media have propaganda potential. The extent to which the media, such as television, can be used to transmit positive social values has not yet been realized.

The most successful preventive measures within the field of medicine have come from community-wide public health programs (e.g., the campaign against smoking) and from programs that detect individual vulnerabilities (e.g., individual monitoring of blood pressure). Studies of adult antisocial behavior reveal the contribution of broad cultural factors and constellations of individual biopsychosocial vulnerabilities. Prevention programs must recognize and address both kinds of factors.

SUGGESTED CROSS-REFERENCES

Material relevant to adult antisocial behavior and criminality may be found in Chapter 2 on neuropsychiatry and behavioral neurology; Section 4.2 on sociology and psychiatry; Chapter 11 on substance-related disorders; Chapter 12 on schizophrenia; Chapter 14 on mood disorders; Chapter 24 on personality disorders; Chapter 40 on disruptive behavior disorders; and Section 54.1 on legal issues in psychiatry. Child and adolescent antisocial behavior is covered in Section 49.7.

SECTION REFERENCES

American Psychiatric Association. *Seclusion and Restraint: The Task Force Report*. American Psychiatric Association, Washington, DC, 1985.

American Psychiatric Association: *Clinician Safety: The Task Force Report*. American Psychiatric Association, Washington, DC, 1992.

Apperson LJ, Mulvey EP, Lidy CW: Short-term clinical prediction of assaultive behavior: Artifacts of research methods. Am J Psychiatry *150:*1374, 1993.

Beck JC, editor: *The Potentially Violent Patient and the Tarasoff Decision in Psychiatric Practice*. American Psychiatric Press, Washington, DC, 1985.

Blumstein A: Youth, violence, guns and the illicit-drug industry. J Crim Law Criminol *86:*10, 1995.

Davis S: Violence by psychiatric inpatients: A review. Hosp Community Psychiatry *42:*585, 1991.

Dickstein LJ, Nadelson CC, editors: *Family Violence: Emerging Issues of a National Crisis*. American Psychiatric Press, Washington, DC, 1989.

Eichelman BS, Hartwig AC, editors: *Patient Violence and the Clinician*. American Psychiatric Press, Washington, DC, 1995.

Evans RW, Gualtieri CT: Carbamazepine: A neuropsychological and psychiatric profile. Clin Neuropharmacol *8:*221, 1985.

Faulkner LR, Grimm NR, McFarland BH: Threats and assaults against psychiatrists. Bull Am Acad Psychiatry Law *18:*37, 1990.

*Federal Bureau of Investigation: FBI National Press Office—Uniform Crime Reporting (UCR) Program. Available at: http://www.fbi.gov/pressrel/prls4-98.htm. Accessed May 17, 1999.

*Flannery OJ, Huff CR, editors: Youth Violence: Prevention, intervention and social policy. American Psychiatric Press, Washington, DC, 1998.

Garza-Trevino ES: Neurobiological factors in aggressive behavior. Hosp Community Psychiatry *45:*690, 1994.

Garza-Trevino ES, Hollister LE, Overall JE: Efficacy of combinations of intramuscular antipsychotics and sedative-hypnotics for control of psychotic agitation. Am J Psychiatry *146:*1598, 1989.

Hillbrand M, Pallone NJ, editors: *The Psychobiology of Aggression*. Haworth Press, New York, 1994.

Howell JC, editor: *Guide for Implementing the Comprehensive Strategy for Serious, Violent and Chronic Juvenile Offenders*. U.S. Department of Justice, Washington, DC, 1995.

*Kurtz L, editor: *Encyclopedia of Violence, Peace and Conflict*. Academic Press, San Diego, CA, 1999.

Malmquist CP: Epidemiology of extreme violence. In *International Review of Psychiatry*, Mak FL, Nadelson CC, editors. American Psychiatric Press, Washington, DC, 1996.

Marzuk PM, Tardiff K, Hirsch CS: The epidemiology of murder-suicide. JAMA *267:*3179, 1992.

*Pinard G, Pagani L, editors: *Clinical Assessment of Dangerousness: Empirical Considerations*. Cambridge University Press, Cambridge, 1999.

Rose HM, Mc Clain PD: Race, place and risk revisited. Homicide Stud *2:*101, 1998.

Silver JM, Yudofsky SC: Propranolol for aggression: Literature review and clinical guidelines. Int Drug Ther Newsl *20:*9, 1985.

Swanson JW, Holzer CE, Ganju VK, Jono RT: Violence and psychiatric disorder in the community: Evidence from the Epidemiologic Catchment Area surveys. Hosp Community Psychiatry *41:*761, 1990.

*Tardiff K: *Assessment and Management of Violent Patients*, ed 2. American Psychiatric Press, Washington, DC, 1996.

Tardiff K: The current state of psychiatry in the treatment of violent patients. Arch Gen Psychiatry *49:*493, 1992.

*Tardiff K, editor: *Medical Management of the Violent Patient*. Marcel-Dekker, New York, 1999.

Tardiff K: Prediction of violence in patients. J Pract Psychiatry Beh Health *4:*12, 1998.

Tardiff K, editor: *The Psychiatric Uses of Seclusion and Restraint*. American Psychiatric Press, Washington, DC, 1984.

*Tardiff K: Unusual diagnoses among violent patients. Psychiatr Clin North Am *21:*567, 1998.

Tardiff K, Marzuk PM, Leon AC, Hirsch CS, Stajic M, Portera L, Hartwell N: Homicide in New York City: Cocaine use and firearms. JAMA *272:*43, 1994.

Thackrey M: *Therapeutics for Aggression*. Human Sciences Press, New York, 1987.

Turner JT: *Violence in the Medical Care Setting*. Aspen, Rockville, MD, 1984.

Stoff DM, Breiling J, Maser JD: *Handbook of Antisocial Behavior*. Wiley, New York, 1998.

Stoff DM, Cairns RF: *The Neurobiology of Clinical Aggression*. Erlbaum, Hillsdale, NJ, 1996.

Swarty MS, Swanson JW, Hiday VA, et al: Violence and severe mental illness: The effects of substance abuse and nonadherence to medication. Am J Psychiatry. *155:*226, 1998.

Van Praag HM, Plutchik R, Apter A, editors: *Violence and Suicidality: Perspectives in Clinical and Psychobiological Research*. Brunner/Mazel, New York, 1990.

*Volavka J: *Neurobiology of Violence*. American Psychiatric Press, Washington, DC, 1995.

Wilkinson CW, editor: Violence in the Workplace: Preventing, Assessing and Managing Threats at Work. Government Institutes, Rockville, MD, 1998.

27.4 BORDERLINE INTELLECTUAL FUNCTIONING AND ACADEMIC PROBLEM

JAMES J. MCGOUGH, M.D.

According to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) borderline intellectual functioning and academic problem are conditions, which may require clinical attention, but are not in themselves mental disorders. They nonetheless contribute substantially to utilization of mental health services. Borderline intellectual functioning and academic problem are usually recognized while individuals are in school. They can occur concurrently with a psychiatric disorder and adversely affect response to treatment or overall functioning. In these instances clinicians may devote more attention than if the psychiatric condition presented alone. Global Assessment of Functioning or V code conditions are occasionally noted when an evaluation for psychiatric causes of dysfunction has not been performed. Borderline intellectual functioning and academic problem can also present independently of any psychiatric disorder.

BORDERLINE INTELLECTUAL FUNCTIONING

Definition DSM IV describes borderline intellectual functioning as follows:

> This category can be used when the focus of clinical attention is associated with borderline intellectual functioning, that is, an I.Q. in the 71-84 range. Differential diagnosis between Borderline Intellectual Functioning and Mental Retardation (an I.Q. of 70 or below) is especially difficult when the coexistence of certain mental disorders (e.g., schizophrenia) is involved. Coding note: This is coded on Axis II.

History and Comparative Nosology The 1959 edition of Classification in Mental Retardation defined anyone with an I.Q. greater than one standard deviation below the mean (I.Q. below 85) as mentally retarded. This definition was applied without regard to functional impairment, and many individuals who fell within this range had no discernible adaptive difficulties. The 1973 edition of *Classification in Mental Retardation* redefined mental retardation as an I.Q. greater than two standard deviations below the mean (70 or below) with associated impairments in adaptive functioning. This definition of mental retardation has been retained in subsequent editions and is adhered to in DSM-IV and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). An I.Q. between 71 and 84 was “declassified” as mentally disordered and redefined as borderline intellectual functioning. Because individuals within this I.Q. range often have little impairment outside of educational settings, the diagnosis is often overlooked after completion of school. The condition may continue to be a focus of clinical attention if it impacts social functioning, vocational adjustment, or compliance with medical management. Approximately 14 percent of the general population has an I.Q. within this range; borderline intellectual functioning is similarly named and defined in ICD-10.

Etiology Performance on intelligence tests is a function of genetic and environmental factors. Twin and adoption studies support hypotheses that genes contribute to intellectual level. Early developmental experiences including infection, injury, poor nutrition, and emotional deprivation can adversely affect I.Q. No clinical syndromes have been specifically linked to borderline intellectual functioning.

Diagnosis and Clinical Features Mild developmental delays, persistent social immaturity, or pervasive academic underperformance can suggest borderline intellectual functioning. A detailed history is obtained with particular focus on developmental milestones, language, peer relations, and academic achievement. In suspicious cases administration of a standardized intelligence measure, such as the Wechsler Intelligence Scale for Children–Third Edition (WISC–III) or the Wechsler Adult Intelligence Scale–Revised (WAIS–R), confirms or rules out the condition.

A 20-year-old woman was poorly compliant with her diabetes management and suffered significant medical complications from poor blood sugar control. Her parents were increasingly angered by her unwillingness to comply with her insulin regimen, diet, and other household rules. The patient had completed high school in a special education class for the learning disabled and planned to attend junior college. Early developmental milestones were reportedly normal, but her parents described her as socially immature. Psychiatric evaluation revealed no significant depression, anxiety, or psychosis. Testing with the WAIS–R demonstrated borderline intellectual functioning with a verbal I.Q. of 80, performance I.Q. of 78, and full-scale I.Q. of 79. Feedback to her parents enabled them to set more realistic expectations for their daughter's behavior. The patient maintained improved blood sugar control after implementation of a behavioral approach to her diet and insulin management.

A 15-year-old boy was evaluated for “acting out and aggressive behavior.” Attention-deficit/hyperactivity disorder and a reading disorder were diagnosed in second grade. Since age 13 he exhibited increased difficulty with fire setting, lying, stealing, truancy, and public intoxication. Sleep and appetite were stable, and he denied feelings of depression or other mood symptoms. Stimulant medication decreased his symptoms of hyperactivity, but the patient expressed little motivation for improvement. Psychiatric evaluation was consistent with “attention-deficit/hyperactivity disorder,” conduct disorder, and reading disorder. Testing on the WISC-III revealed borderline intellectual functioning with a verbal I.Q. of 79, performance I.Q. of 77, and full-scale I.Q. of 78. The patient was admitted to a residential treatment center but his prognosis was considered to be poor.

Treatment Treatment for individuals with borderline intellectual functioning is focused on optimizing academic achievement and the capacity for successful adaptive living. Individuals in this group can have significant difficulties with school, work, and social relationships; proper school and vocational placement is essential. Children are generally placed in small, self-contained classrooms for the learning-handicapped. Interventions are student-focused and individualized. School work should emphasize mastery of academic skills appropriate to the student's cognitive capacity. Development of functional skills and guidance in vocational training are essential and behavioral interventions, such as social skills training, may be particularly helpful. Comorbid psychiatric conditions should be treated as indicated with medication or therapy; family therapy serves to educate and establish appropriate behavioral expectations in the home.

ACADEMIC PROBLEM

Definition DSM-IV describes academic problem as follows:

> This category can be used when the focus of clinical attention is an academic problem that is not due to a mental disorder or, if due to a mental disorder, is sufficiently severe to warrant independent clinical attention. An example is a pattern of failing grades or of significant underachievement in a person with adequate intellectual capacity in the absence of a Learning or Communication Disorder or another mental disorder that would account for the problem.

Academic problem is similarly named and defined in ICD-10.

Etiology Academic problem can result from factors intrinsic or external to the student. Psychiatric conditions such as anxiety or mood disorders can impair learning or lead to performance decline. Attention-deficit/hyperactivity disorder, chronic illness, and identity problems can lead to demoralization in school studies apart from any diagnosable mental condition. Family difficulties, social stressors, cultural deprivation, or poor fit between a student and teacher's temperamental style can also adversely affect scholastic performance.

Diagnosis and Clinical Features Academic problem is evidenced by a pattern of academic underachievement or a decline from a previous level of functioning. A comprehensive biopsychosocial assessment is fundamental to identification of causal factors. Particular focus should be given to past academic functioning and family and social stressors; concurrent psychopathology must be ruled out. Intelligence tests, measures of academic achievement, or language evaluation may be useful in differentiating academic problem from specific learning or communication disorders. Review of the medical history and physical examination may be useful in

identifying general medical conditions such as hearing loss, poor vision, and chronic illness that may adversely impact academic performance.

A 17-year-old student who had been getting As and Bs was active on the school football team and had a steady girlfriend. His parents were pleased when he obtained a part-time job, but after 3 months the student was failing Math. Psychiatric evaluation uncovered no evidence of depression, adjustment disorder, or other psychiatric condition. Family assessment revealed a history of financial difficulties and strong imperatives to earn money. The student modified his work schedule and repeated Math in summer school; he went on to attend college without difficulty.

A 27-year-old woman was unmotivated to complete her doctoral dissertation. Evaluation identified no symptoms of psychiatric disorder. A period of therapy uncovered fears of becoming more educated than her parents. The patient worked through this conflict and was able to obtain her degree.

A 16-year-old high school sophomore with a long history of attention-deficit/hyperactivity disorder wished to obtain a GED. He was failing most classes in spite of having average intelligence. Evaluation demonstrated demoralization about academics, but no anhedonia or other psychiatric symptoms. He responded well to individual tutoring and parental incentives to remain in school.

A 14-year-old honors student developed anhedonia, concentration difficulties, sleep disturbance, and weight loss. She failed several classes and was increasingly behind in her school work. She enrolled in remedial course work subsequent to her response to treatment for depression.

Treatment Treatment is student focused and based on the causes of the academic problem. Classroom-based social skills training and cognitive-behavioral therapies have been found to be effective for underachieving students. Students suffering from “poor fit” with a particular teacher may require placement in a new classroom. Academic difficulties may require remedial course work or tutoring. Comorbid psychiatric conditions and psychosocial difficulties should be treated with medication or psychotherapy, as necessary.

SUGGESTED CROSS-REFERENCES

Mental retardation is discussed in Chapter 34. Learning disorders, motor skills disorders, and communication disorders are discussed in Chapter 35, Chapter 36, Chapter 37, respectively.

SECTION REFERENCES

*Alderman MK: *Motivation for achievement: Possibilities for Teaching and Learning*. Erlbaum, Mahwah, NJ, 1999.

American Association on Mental Retardation: *Classification in Mental Retardation:* American Association on Mental Retardation, Chicago, 1993.

*Beal EW: Academic difficulties found within dysfunctional family relationships. Child Adolesc Psychiatric Clin N Am *6:*579, 1997.

*Brophy JE: *Teaching Problem Students*. Guilford Press, New York, 1996.

*Hagin RA: Psychological problems that present as academic difficulties. Child Adolesc Psychiatric Clin N Am *6:*473, 1997.

Helen H: Working with children who resist learning. In *Educational Therapy in Clinic and Classroom*, M Barrett, V Varma, editors. Whurr Publishers, London, 1996, p 3.

Knoff HM, Batsche GM: Project ACHIEVE: Analyzing a school reform process for atrisk and underachieving students. School Psychol Rev *24:*579, 1995.

Lerner JV, Lerner RM, Zabski S: Temperament and elementary school children's actual and rated academic performance: A test of a “goodness-of-fit” model. J Child Psychol Psychiatry *26:*125, 1985.

Sexson SB, Dingle AD: Medical problems that might present with academic difficulties. Child Adolesc Psychiatric Clin N Am *6:*509, 1997.

Sexson S, Madan-Swain A: The chronically ill child in the school. Special Series: Cognitive and academic issues related to chronic illness. School Psychol Quarterly *10:*359, 1995.

*Shapiro ES, Bradley KL: Treatment of academic problems. In *Cognitive Therapy with Children and Adolescents: A Casebook for Clinical Practice*, MA Reinecke, FM Dattilio, A Freeman, editors. Guilford, New York, 1996.

Smith JD: The revised AAMR definition of mental retardation: The MRDD position. Ed Train Ment Retard Dev Disabil *29:*179, 1994.

Smith JD, Hilton A: The preparation and training of the educational community for the inclusion of students with developmental disabilities: The MRDD position. Ed Train Ment Retard Dev Disabil *32:*3, 1997.

*Strohmer DC, Prout HT: *Counseling and Psychotherapy with Persons with Mental Retardation and Borderline Intelligence*. Clinical Psychology Publishing, Brandon, VT, 1994.

Westerman MA, La Luz EJ: Marital adjustment and children's academic achievement. Merrill-Palmer Q *41:*453, 1995.

Wittman JJ, Strohmer DC, Prout HT: Problems presented by persons of mentally retarded and borderline intellectual functioning in counseling: An exploratory investigation. J Appl Rehab Couns *20:*8, 1989.

27.5 OTHER ADDITIONAL CONDITIONS THAT MAY BE A FOCUS OF CLINICAL ATTENTION

LEAH J. DICKSTEIN, M.D.

The fourth edition of the American Psychiatric Association *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) includes 13 conditions that are not considered true mental disorders but make up the category of additional conditions that may be a focus of clinical attention. Four of these conditions: (1) occupational problem, (2) religious or spiritual problem, (3) acculturation problem, and (4) phase of life problem are discussed in this section. The remaining 9 conditions are discussed elsewhere in this textbook: (5) noncompliance with treatment, (6) malingering, (7) adult antisocial behavior, (8) child or adolescent antisocial behavior, (9) borderline intellectual functioning (coded on Axis II), (10) age-related cognitive decline, (11) bereavement, (12) academic problem, and (13) identity problem. All except borderline intellectional functioning are coded on Axis I.

DEFINITION AND NOSOLOGY

These conditions are not considered mental disorders by DSM-IV criteria, but are related to the mental disorders in one of two ways: (1) The condition may be the focus of clinical attention, diagnosis, or treatment in the absence of a mental disorder; for example, a 90-year-old recent widower who always believed he was self-sufficient realizes after his wife's death that his physical stamina is diminishing and that he will have to move to an assisted living environment (phase of life problem). (2) The individual has been diagnosed with a mental disorder but the additional condition is unrelated to the mental disorder; for example, a 43-year-old woman with a history of specific phobia of ants finds she is unable to advance at work despite excellent ongoing periodic reviews because of gender bias (occupational problem).

When these additional conditions have led to contact with the mental health care system, the person should undergo a complete neuropsychiatric evaluation to rule out and not overlook a current additional psychiatric disorder or prodromal symptoms and signs of underlying psychiatric illness. Clearly, the distinction between an additional condition and a DSM-IV diagnosis is not always clear-cut.

OCCUPATIONAL PROBLEM

DSM-IV includes the following statement about occupational problem:

> This category can be used when the focus of clinical attention is an occupational problem that is not due to a mental disorder or, if it is due to a mental disorder, is sufficiently severe to warrant independent clinical attention. Examples include job dissatisfaction and uncertainty about career choices.

Occupational psychiatry and its expansion to include organizational psychiatry are specifically focused on the psychiatric aspects of work problems, including vocational maladjustment. Symptoms of employment dissatisfaction are varied and include blatant work errors, perceived and verbalized unhappiness and disinterest, absenteeism and tardiness, and passive-aggressive behaviors including accidents and uncooperativeness. Psychiatric symptoms include general signs of distress, anger, resentment about most work assignments, lack of confidence, and disinterest in carrying out agreed upon and expected work responsibilities.

Occupational problems often arise during stressful changes in work, namely, at initial entry into the work force or when making job changes within the same organization to a higher position because of good performance or to a parallel position because of corporate need. Distress occurs particularly if these changes are not sought and no preparatory training has taken place, as well as during layoffs and at retirement, especially if retirement is mandatory and the person is unprepared for this event. Work distress can result if initially agreed to conditions change to work overload or lack of challenge and opportunity to experience work satisfaction, if persons feel unable to fulfill conflicting expectations or feel that work conditions prevent accomplishing assignments because of lack of legitimate power, and finally if persons believe they work in a hierarchy with harsh and unreasonable superiors.

Work Choices and Changes Currently, most workers expect to change occupations at least several times during their lifetimes. Many who enter the workforce are initially unsure of their interests and potential. They often do not seek vocational counseling in high school and college and accept the first available work opportunity. Others feel parental pressures to enter certain occupations (e.g., a family business, trade, or profession).

Young adults without role models or guidance from families or those in their communities too often underestimate their lifetime potential abilities to learn a trade or earn a college degree. In addition, women and members of minority groups often feel less prepared to accept work challenges, fear rejection, and do not apply for jobs for which they are qualified. As part of initial interviews for evaluation of occupational problems, patients should be encouraged to consider their talents, long-held dreams and goals in regard to work, successes in work and school, and motivation to risk learning what they would find satisfying.

Clearly, occupational distress is inevitable following unprovoked job loss, rejection, and discrimination. Minorities and those in low-paying and low-skilled jobs too often have less job security. Downsizing, factory closings, and moves affect many, often leaving these workers feeling hopeless and helpless about future employment, on welfare, and angry.

With ongoing and often sudden downsizing of corporations and businesses, men and women continue to struggle with unexpected job loss and premature retirement, even when finances are not an issue. In addition, men, in particular, define themselves by their work roles, and thus experience more occupational distress from these changes. Women may adjust faster to retirement, but they often have less financial security than men (they earn 76 cents on the dollar for comparable work); have generally been in lower-status work positions; find themselves widowed more often than men; and are more likely to be caring for children, grandchildren, and elderly relatives. Women represent more of the single working parent group and the working poor.

Stress and the Workplace Workplace distress is implicated in at least 15 percent of occupational disability claims. Expected distress follows recognized and uncontrollable work changes: downsizing, mergers and acquisitions, work overload, and chronic physical strains, including work noise, temperature, and injuries and strain due to computer work. According to a study reported in the June 1998 issue of *Working Woman*, the top ten most stressful jobs in 1998 were (1) the U.S. president, (2) firefighter, (3) senior corporate executive, (4) race car driver, (5) taxi driver, (6) surgeon, (7) astronaut, (8) police officer, (9) football player, and (10) air traffic controller.

Work frustration can also arise from an individual worker's unrecognized (and therefore unresolved) psychodynamic issues such as working appropriately with superiors and not relating to one's supervisor as a parent figure. Other developmental issues include unresolved problems with competition, assertiveness, envy, fear of success, and inability to communicate verbally in a constructive manner.

> A 38-year-old white male has outbursts of anger when told to complete a day's tasks in a construction company. He complains to his union officer who, after investigating the worker's complaints, states that the work assignment is fair. The worker has never taken pride in completing assignments and has moved from company to company on average every 9 months since resigning from the military 6 years earlier because he tired of taking orders.

Often work conflicts reflect similar conflicts in the worker's personal life and referral for treatment, unless there is insight, is in order.

Suicide Risk Some occupations both attract persons with a high suicide risk and involve increased chronic distress that may lead to higher suicide rates. Included are

health professionals, financial service workers, and police, the first and latter groups because of easier access to lethal drugs and weapons.

Career Problems of Women Most women work outside the home out of necessity to support themselves, or their dependents (whether children or adults) or as part of a working couple. With the divorce rate remaining at the 50 percent level, many women find themselves economically poorer after a divorce than they were when married, while divorced men usually find their economic status improved. Despite three decades of increasing knowledge about, and concern for, women in the work site, unique gender issues, bias, and lack of accommodation to their life cycle needs continue when they are responsible for dependents. Women were the largest group establishing new small businesses in the past decade. Many have left large corporations where they were not valued for their efforts because of their gender, termed contra culture. Furthermore, women still experience problems when they are the sole woman in a “man's field.” Despite increasing recognition of the need for men in relationships with women to assume home/family responsibilities, less than 25 percent of men do so equitably.

Women continue to find work options limited particularly in leadership roles, because of gender bias, not because of lack of competence. A recent catalyst study revealed that women held 11 percent of director seats at 500 firms despite several decades of available, well-trained, and interested women candidates. Men and women with power (i.e., “men and women of good conscience”) must help competent women fairly attain deserved work opportunities. This process is called “taking down the impenetrable Lexan Ceiling together,” rather than women being unable to break through the supposed glass ceiling.

Women of childbearing and child-rearing ages continue to find themselves in conflict with job expectations, opportunities, and personal responsibilities. Good-quality, on-site dependent-care facilities with extended hours are rare and often out of range financially. Major unresolved work issues that are unique for women at certain life stages include flextime and paid and unpaid dependent leave options. Gender differences in attitudes toward work, life priorities, and definition of a good job persist and must be understood and respected. Furthermore, when men assume responsibilities for dependents they are applauded, while women in similar circumstances are often denigrated and passed over for deserved work opportunities. However, an increasing number of large and successful corporations highly rated as sensitive to women employees' personal responsibilities and needs are listed annually in *Working Woman* and other publications. Beyond dependent care issues, women in the workforce continue to experience distress following sexual harassment, despite its illegality and media attention. Increasingly more women have travel responsibilities, work long hours, work shifts beyond daylight hours, and are victims of personal violence. Work sites, hotels, streets, and travel must be made safer.

Among dual-career families and partners, the woman is more likely to move when the man chooses to move for a work opportunity than vice versa. Consequently, a woman's career is interrupted. Fortunately, there is less reluctance to have the two members of a relationship work for the same organization than previously, albeit usually in different departments. Work distress may also stem from continuous miscommunication, especially that based on gender. A growing and useful literature is focused on understanding gender issues in miscommunication at work. Educating women and men about this miscommunication as a major stressor can decrease work distress and improve work output and attitudes.

Working Teenagers With unemployment at an all-time low, many teenagers work part time while attending high school. Stress can arise because of less parent-teenager interaction and constructive parental control issues about use of earnings and time spent away from home and consequent behaviors both in and outside the home. When both parents or a single parent works outside the home as well as the teenager, verbal communication must be proactive and clear.

Working Within the Home Although most women with children must work outside the home, at times they may be home full time or part time or work at home. When their husbands or significant others work full time outside the home, problems may develop from each one's perceived expectations of the other. Women who care for children and their home exclusively and their partners may particularly see the woman as not only economically dependent and inferior, but not as competent and not understanding of the man's stressors and needs. Ongoing respectful listening and verbal communication must be encouraged.

Children must also be included as appropriate by age in contributing to the cooperative efforts in maintaining the home, however small their efforts may be. Furthermore, repeated studies have shown that people in organizations increasingly take work home as the work expectation increases. This work-at-home experience can and does interfere with personal lives and satisfaction, which can then have further repercussions at work.

Chronic Illness As medical treatments for chronic diseases improve, businesses have been increasingly concerned about accommodating to patients with acquired immune deficiency syndrome (AIDS), diabetes mellitus, and other disorders. The issue of mandatory testing for AIDS and substance abuse (alcohol and other illegal substances) continues to be of concern. Employee assistance programs, including general health and education, have proved timely and cost-effective.

Domestic Violence Although occurring in the home, signs and symptoms that interfere with work often trigger identification of domestic violence victims. All employees experiencing work distress must be questioned about domestic violence by trained professionals.

Loss of a Job Regardless of the reason for job loss, most people experience distress, including symptoms of normal grief, loss of self-esteem, anger and reactive depressive and anxiety symptoms, as well as somatic symptoms and possibly substance abuse and increased or onset of domestic violence. Timely education, support programs, and vocational guidance should be instituted.

RELIGIOUS OR SPIRITUAL PROBLEM

According to DSM-IV:

> This category can be used when the focus of clinical attention is a religious or spiritual problem. Examples include distressing experiences that involve loss or questioning of faith, problems associated with conversion to a new faith, or questioning other spiritual values which may not necessarily be related to an organized church or religious institution.

In an increasingly scientific world, religion or spirituality are often not acknowledged to be of major importance in people's lives. Furthermore, with ever-increasing commitments to work and other activities, together with high divorce rates and geographic distances between nuclear and extended family members, religious commitment and participation are often minimal, missing, or (with conversion) conflicted events. Faith may be lost or questioned at life-cycle crises or events, which may evolve into problems. Psychiatrists must enable and assist patients to distinguish religious thought or experience from psychopathology and, if this is a problem, encourage patients to work through it, independently or with assistance.

Religious imagery may be recognized in mental illness when persons state they believe they have been commanded by God to take a dangerous or grandiose action.

> A young professional woman finds herself in conflict when she decides to marry outside her family's religious affiliation, despite being well aware that her family has neither participated in religious activities nor spoke of religious beliefs in their home and family life. She seeks help to resolve her relationship with her family and decide on the possibilities and repercussions of her impending marriage to someone she loves and respects whose nuclear family is of another religious denomination.

Cults Cults appear to be less popular recently and less attractive to naive late adolescents and young adults seeking assistance in discovering who they are and unsuccessfully struggling to develop more mature relationships with parents. Cults are led by charismatic leaders, often out of control themselves, with inappropriate and often unethical values but purporting to offer acceptance and guidance to followers. Acceptance into the cult usually depends on complete adoption of a leader's belief system with no allowance for individual thinking and behavior. Cult members are strongly controlled and forced to dissolve allegiance to family and others to serve the cult leader's directives and personal needs. These young members often come from educated families who seek professional help in persuading their children to leave the cult and deprograming therapy to restore personal psychological stability to the former cult members. Deprograming and adjustment back into family, society, and an independent life is time intensive and long term, with resultant posttraumatic stress disorder.

ACCULTURATION PROBLEM

In DSM-IV the following statement about the acculturation problem appears:

> This category can be used when the focus of clinical attention is a problem involving adjustment to a different culture (e.g., following migration).

Culture Shock Major cultural change can evoke severe distress, termed culture shock. This condition arises when individuals suddenly find themselves in a new culture in which they feel completely alien. They may also feel conflict over which lifestyles to maintain, change, or adopt. Children and young adult immigrants often adapt more easily than middle-aged and elderly immigrants. They learn the new language more easily, and they continue to mature in the new culture, whereas those more senior have had more stability and unchanging routines in their former culture. Culture shock from immigration clearly differs from psychiatric patients' restless and continuous moving secondary to their illness.

Culture shock may occur within one's own country with geographical, school, and work changes such as joining the military, experiencing school busing, or moving across country or to a vastly different neighborhood or from a rural area to a metropolis. Reactive symptoms are understandable and include anxiety, depression, isolation, fear, and a sense of loss of identity as one adjusts. If the person is part of a family or group making this transition and the move is positive and planned, stress can be lower. Furthermore, if selected cultural mores can be safely maintained as persons integrate into the new culture, stress is also minimized.

The constant geographical moves because of work opportunities and necessity involve a large proportion of U.S. workers. Joining activities in the new community and actively trying to meet neighbors and coworkers can lessen the culture shock.

> A 29-year-old single woman moved because of a work promotion to southern California from Indiana. She joined a volunteer group working with inner-city children to share her musical talents, visited several churches, then joined one, and made concerted efforts to greet neighbors on her way to and from work.

Brainwashing First practiced on American prisoners during the Korean War by the Chinese Communists, *brainwashing* is the deliberate creation of culture shock. Victims are isolated, intimidated, and made to feel different and out of place to break their spirits and destroy their coping skills. Once the person appears mentally weak and helpless, new ideas that they would never have accepted in their normal state are imposed upon them by the aggressors. Like cult victims, upon release and return to their homes, brainwashing victims with posttraumatic stress disorder require deprograming treatment, including reeducation and ongoing supportive psychotherapy both on an individual and group basis. Treatment is usually long-term to rebuild healthy self-esteem and coping skills.

Prisoners of War and Torture Victims Prisoners who survive war or torture experiences survive because of personal inner strengths developed in their earlier lives beginning within their emotionally strong and caring families; if from troubled families, they may more likely have suicided. Prisoners must constantly cope with continued anxiety, fear, isolation from known lives, and complete loss of all control over their lives. Those who appear to cope best believe they must survive for a reason (e.g., to tell others what they experienced or to return to loved ones). Prisoners who cope best describe living simultaneously on two levels: coping in the here-and-now to survive the situation while maintaining constant mental connections to their past values and experiences and those important to them.

Beyond the surviving prisoner's personal difficulties, including posttraumatic stress disorder, if and when their survival behavior continues, it can affect their consequent families, with inordinate fear of police and strangers, overprotection and overburdening of children to replace those lost, lack of sharing of the past, continued isolation from their current communities, or inappropriate expressed anger. Another generation can thus be affected in their personal development and psychological functioning and may require psychiatric evaluation and treatment.

> A 72-year-old Jewish woman survivor of the Auschwitz Nazi concentration camp stated that as a 19-year-old married Hungarian who wanted to become a lawyer, she accepted the leadership of a women's barracks because she knew she could defend them as her father had helped the poor in their Hungarian city. She encouraged others to care for their bodies, share food and songs and stories at night, and maintain the hope they would survive to find some family and make a new life for themselves after the war. At liberation after the death march, she helped American soldiers translate, found her sister and husband, and built a new life, continuing her leadership in her new community at a hospital in Jerusalem.

PHASE OF LIFE PROBLEM

The DSM-IV description of phase of life problem includes the following:

> This category can be used when the focus of clinical attention is a problem associated with a particular developmental phase or some other life circumstance that is not due to a mental disorder or, if it is due to a mental disorder, is sufficiently severe to warrant independent clinical attention. Examples include problems associated with entering school, leaving parental control, starting a new career, and changes involved in marriage, divorce, and retirement.

Although on some level adults recognize that in the course of a lifetime, life events intrude on expected plans, unexpected, multiple, major negative occurrences, especially if they are chronic, overwhelm a person's ability to recover and function constructively. Common phase of life problems include relationship changes such as a changed significant personal relationship, job crises, and parenthood.

Because of sex role socialization and consequent cultural expectations, men appear externally better able to handle these phase of life problems, and women, the poor, and minority group members appear more vulnerable to negative experiences, perhaps because they feel less empowered psychologically. Major life changes precipitate distress in the form of anxiety and depressive symptoms, inability to express reactive emotions directly, and often difficulties in coping with life responsibilities.

Persons best able to cope with phase of life problems appear to be those with positive attitudes and mature defense mechanisms and coping styles, including basic trust in self and others, good verbal communication skills, a capacity for creative and positive thinking, and the ability to be flexible, reliable, and energetic with strong family and personal relationships. Furthermore, a capacity for sublimation, adequate financial and work status, solid values, and healthy feasible goals can enable people to face, accept, and deal realistically with life problems and changes.

SUGGESTED CROSS-REFERENCES

Cross-cultural psychiatry is discussed in Section 4.4. Section 25.9 covers stress and psychiatry. Noncompliance with treatment appears in Section 27.1, malingering in Section 27.2, adult antisocial behavior in Section 27.3, child or adolescent antisocial behavior in Section 49.7, and borderline intellectual functioning and academic problem in Section 27.4. Age-related cognitive decline is discussed in Section 51.2b on the physiological aspects of aging. Bereavement appears in Section 28.5 on death, dying, and bereavement. Section 49.10 covers identity problems and borderline disorders. Cults are discussed in Section 28.9.

SECTION REFERENCES

Araki S, Kawakami N: Health care of work stress: A review. Jpn J Ind Health *35:*88, 1993.

Bellrosa C, Chen PY: The effectiveness and practicality of occupational stress management interventions: A survey of subject matter expert opinions. J Occup Health Psychol *2:*247, 1997.

Favaza AR: Psychotherapy: Diversity in psychotherapy: The politics of race, ethnicity, and gender. Am J Psychiatry *152:*807, 1995.

Goldenhar LM, Swanson NG, Hurrell JJ Jr, Ruder A, Deddens J: Stressors and adverse outcomes for female construction workers. J Occup Health Psychol *3:*19, 1998.

Greiner BA, Ragland DR, Krause N, Syme SL, Fisher JM: Objective measurement of occupational stress factors—an example with San Francisco urban transit operators. J Occup Health Psychol *2:*325, 1997.

Hauff E, Vaglum P: Organised violence and the stress of exile: Predictors of mental health in a community cohort of Vietnamese refugees three years after resettlement. Br J Psychiatry *166:*360, 1995.

Javier I, Escobar MD: Immigration and mental health: Why are immigrants better off. Arch Gen Psychiatry *55:*781, 1998.

Kageyama T, Nishikido N, Kobayashi T, Kurokawa Y, Kaneko T, Kabuto M: Self-reported sleep quality, job stress, and daytime autonomic activities assessed in terms of short-term heart rate variability among male white collar workers. Ind Health *36:*263, 1998.

*Kroll J: Religion and psychiatry. Curr Opinion Psychiatry *8:*335, 1995.

Levav I, Kohn R, Schwartz S: The psychiatric after-effects of the Holocaust on the second generation. Psychol Med *28:*775, 1998.

Levi L: Occupational mental health: Its monitoring, protection and promotion. J Occup Med *21:*26, 1979.

*Marshall NL, Barnett RC, Sayer A: The changing workforce, job stress, and psychological distress. J Occup Health Psychol *2:*99, 1997.

Melamed B: Psychological aspects of AIDS: What are friends for? Curr Opinion Psychiatry *8:*414, 1995.

Mickleburgh WE: Occupational mental health—a neglected service. Br J Psychiatry *148:*426, 1986.

Miller DM: Improving mental well-being in the workplace. Occup Med *47:*463, 1997.

Miller TQ, Markides KS, Chiriboga DA, Ray LA: A test of psychosocial vulnerability and health behavior models of hostility: Results from an 11-year follow-up study of Mexican Americans. Psychosom Med *57:*572, 1995.

Mor-Barak ME: Social support and coping with stress: Implications for the workplace. Occup Med *3:*663, 1988.

*Mumford DB: Cultural issues in assessment and treatment. Curr Opinion Psychiatry *8:*134, 1995.

Petterson IL, Arnetz BB: Measuring psychosocial work quality and health: Development of health care measures of measurement. J Occup Health Psychol *2:*229, 1997.

Phillips SD, Imhoff AR: Women and career development: A decade of research. Annu Rev Psychol *48:*31, 1997.

Piotrkowski CS: Gender harassment, job satisfaction, and distress among employed white and minority women. J Occup Health Psychol *3:*33, 1998.

Quershi B: Social psychiatry across cultures: Studies from North America, Asia, Europe and Africa. Br J Psychiatry *170:*487, 1997.

Rousseau C, Drapeau A: Parent-child agreement on refugee children's psychiatric symptoms: A transcultural perspective. J Am Acad Child Adolesc Behav *37:*629, 1998.

*Siddique CM, Aubry TD, Mulhall D: The burden of conditions not attributable to mental disorders. Am J Psychiatry *153:*489, 1996.

*Sloan RP, Bagiella E, Powell T: Religion, spirituality, and medicine. Lancet *353:*664, 1999.

Stutts JT, Kasdan ML: Psychosocial aspects of hand injuries and diseases. Occup Med *13:*513, 1998.

*Tannen D: *Talking from 9 to 5: Women and Men in the Workplace: Language, Sex and Power*. Avon, New York, 1995.

Van Tilburg MAL, Vingerhoets AJJM, Van Heck GL: Homesickness: A review of the literature. Psychol Med *2:*899, 1996.

Zilber N, Lerner Y: Psychological distress among recent immigrants from the former Soviet Union to Israel correlates of level of distress. Psychol Med *26:*493, 1996.

CHAPTER 28. SPECIAL AREAS OF INTEREST

28.1 PRIMARY CARE AND PSYCHIATRY

MACK LIPKIN, JR., M.D.

In the 1990s primary care reemerged as the keystone of the mental health system in the United States. Primary care accounts for the most patients seen, drugs prescribed, hospital admissions, and cases of mental disorder treated.

Primary care is dominant in mental health delivery for complex reasons. Most of the general population first seeks and receives care for mental disorders at the office of the primary care physician. This is because people usually do not perceive of their own problems as mental illness and because most people prefer to avoid the personal and public stigma of seeing a psychiatrist. Of course, some do see their problems as mental and present initially to psychiatrists, but they are and will remain in the minority. Others, while recognizing they have depression or anxiety, nonetheless see their primary care doctor because they are more accessable than psychiatrists.

Primary care physicians are increasingly, although variably, trained to care for mental disorders and the psychological distress that accompanies normal development, crises, and disasters. They will increasingly care for straightforward mental problems as psychiatrists, fewer in number, focus on the sickest patients with difficult mental disorders. This will be even more true as behavioral health care gets carved out and carved up by managed care and medical corporation and easy access to psychiatrists becomes increasingly restricted.

Medical practitioners must cope with an exploding knowledge base, increased technologization, and unbridled bureaucratization—factors that have transformed and dehumanized the practice of medicine and surgery. As a result, patients and physicians face both classic and new problems in managing disease and illness. Their solutions require knowledge of the relationship of medicine and psychiatry, which integrates information and skills from not only biological but also psychological and social or systems approaches.

Psychiatry and medicine are even more inseparable than mind and body. New data increasingly demonstrate that the simple notion of single organ, single function (e.g., glands only produce hormones, nerves only transmit impulses, the brain is the site of mental disorder) no longer holds. For example, the nervous system functions not only for sensory processing, motor action, and thinking and feeling, but also as an endocrine organ and an immune organ. Every physical illness and symptom has psychological effects on the patient. Some physical illnesses for which patients seek or require medical care have psychological causes; in some severe cases, medical management makes no difference, so practitioners use psychological management. Other illnesses are related indirectly to psychological causes, psychological matters being one of several risk factors. Some psychiatric illnesses have clear physical substrates that require medical treatment.

Mental health problems represent a major and growing burden worldwide. Mental health problems account for almost 10 percent of lost years of quality life and an additional third is related to behaviorally determined illness. Depression is as significant a problem as hypertension. Substance abuse annually accounts for hundreds of billions of dollars in lost productivity and direct costs. At the same time, the mental health sector of health care is facing stagnant or diminishing resources. These burdens fall by default in the laps of primary care physicians.

The value and importance of the role of primary care in mental health care are reinforced by some (perhaps dominant) forces within modern psychiatry, which is increasingly medical and biological in approach. Modern psychiatry risks abandoning psychiatry's humanistic origins and skills to primary care. Even as new research methods show the great importance of personal development, life context, and an ecological approach to the whole person, psychiatric training and perspectives exhibit a new orthodoxy of rigid classification and scientific explanation.

The biological and reductionist views of disease as resulting from infection, toxins, genetic defects, or metabolic derangement are too narrow. Such pioneers as George Groddeck, Sigmund Freud, Franz Alexander, and George Engel, who postulated that the predisposition and timing of physical disease relate to psychological and social factors, are proving correct. Psychological factors are relevant in understanding why people get sick when they do, why the illness takes the form it does, and why some patients heal better than others. The primary contributions responsible for progress in this area have come from four research areas. Behavioral immunologists are demonstrating the behavioral conditioning of the immune response, diminished B-cell and T-cell immune function after stressful life events such as the loss of a spouse, and alteration of susceptibility to disease through specific social or psychological changes. The discovery of receptors for endogenous morphinelike substances in the brain has created an explosion of new neuropharmacophysiology. The theoretical synthesis of pain data has led to the concept of a cascade of neural gates from the individual spinal levels where sensory input occurs through higher midbrain and cortical levels. New epidemiological studies have shown that depression is an independent risk factor for death after myocardial infarction, for development of chronic disease, and for illness-seeking behavior.

DEFINITIONS

The definitions of primary care and psychiatry are formal and informal, denotative and connotative.

Primary Care Primary care is a set of approaches and functions, not a set of problems. It is often embedded in an ideology concerning the role and purpose of medicine as serving the patient and the population in a comprehensive and humanistic manner to improve both illness and disease care and health and personal well-being. Primary care can be provided by practitioners from many specialties but usually involves generalist physicians. The Institute of Medicine recently defined primary care as:

the provision of integrated, accessible health care services by clinicians who are accountable for addressing a large majority of personal health care needs, developing a sustained partnership with patients, and practicing in the context of family and community.

What is new is the inclusion of integration of care by one clinician whose care is accessible, first contact, comprehensive, coordinated, and continuous in a sustained partnership. *Comprehensive* refers to inclusion of acute and chronic disease care, routine ongoing preventive care, psychiatric and psychosocial care, and referral (i.e., any problem at any stage of life). *Coordinated* refers to managing the complete problem list of the patient and serving as liaison between the various consultants and services for the patient. This includes providing quality control and help in understanding and decision making concerning what is going on and in some cases being a “gatekeeper,” monitoring access to specialist. The latter is an unpopular role for both parties. Advocacy for the patient is sometimes included as well. First contact means that a patient with a health problem first sees or calls the primary care physician. It is a necessary but not sufficient attribute of primary care practice (e.g., emergency medicine, a first-contact site, is an antipode of primary care).

A paramount attribute of primary care is that it is continuous, meaning that patient and doctor intend to stay in the relationship for several to many years. Because the average American moves every 3 to 4 years and the average doctor every 4 years, this relationship often stops for practical reasons. The most valuable feature of continuity is that each person gets to know the other and continues to learn and deepen the relationship. This permits the physician to interpret the patient's complaints accurately (“I am about to die” from a dramatic person with chest pain may be the equivalent of “nothing is really bothering me” from another more stoic person with exactly the same degree of ischemia). Continuity allows both to adapt to the other, which optimizes cooperation. This mutual adaptation makes possible the sustained partnership of a mature primary care relationship. Partnership also requires mutual respect and caring, the ability to negotiate, and an appropriate balance of power and decision making. Partnership and sharing of power are often absent in toto or in part in specialty care, including psychiatry. For example, although some psychodynamic therapists relate to their patients with respect, caring, and continuity, they seldom share power or decision making or negotiate. The average number of visits to a psychiatrist for a patient is four. Clearly, not all patients wish to make the major decisions, and not all doctors can allow their patients to

make decisions they disagree with without suffering and strenuous opposition.

Primary care theoreticians such as Ian McWhinney distinguish primary care from care of an episode of illness. Primary care focuses on patients in the context of their lives and ongoing health. It is a general, open-ended inquiry. Prevention is a priority, and helping the person is as important as managing the disease. These concepts are listed in Table 28.1-1.

Sustained partnership
Integrated
Accessible
Continuous
Comprehensive
Focused on the person rather than the episode of illness
Acute- and chronic-disease care
Preventive
First contact
Advocacy role
Care coordination role
Within context of family and community

Table 28.1-1 Definition and Functions of Primary Care

Several features of this definition pertain to mental health issues. In general, primary care doctors intend to care for most of a patient's problems, including the common mental disorders, especially those that present with physical symptoms, including depression, anxiety disorders, somatization, adjustment disorders, normal and abnormal grief reactions, substance abuse, and stress. A classification of the psychological and social issues seen in primary care is presented in Table 28.1-2. In dealing with these problems, primary care physicians to greater or lesser extents use talk therapies, psychotropic drugs, and therapies such as relaxation methods, meditation, and exercise. A complete empirical view of the extent to which primary care physicians use some or all of these methods is not available, but clearly they treat most patients with mental disorders and prescribe most psychotropic drugs. Primary care physicians refer when the diagnosis or effective response to therapy eludes them or specialized procedures or treatments are required.

Table 28.1-2 Classification of Psychological and Social Problems in Primary Health Care

Psychiatry The definition of psychiatry is both historical and functional. Formally, it is the branch of medicine concerned with diagnosis and treatment of mental disorders and related conditions. Societal attitudes about psychiatry and those of persons from other medical disciplines range from the respectful and worshipful to the magical and fearful. Unfortunately, in most parts of America and most world cultures, seeing a psychiatrist is stigmatized. People would usually rather see their own doctor, whom they trust and feel comfortable with, and neither have others know nor have to admit to themselves that they have a mental problem. Some cultures deny mental disorders to the extent that they lack words for craziness or depression (they are equally common in all major populations studied) or hide these issues in physical syndromes or euphemistic or vague terms, the equivalents of neurasthenia and anomie. This is even more true for problems that are not severe and those that are adjustments or adaptations to life and its crises and development.

The interface of primary care and psychiatry is complex and extensive. Darrell Regier has described the de facto mental health system in the United States in which 50 percent of persons with 1 of the 13 most common defined mental disorders are cared for in the primary care sector, and only about 20 percent are cared for in the mental health sector. In addition, there are more common cases in which presentations are somatized, disorders are “subsyndromal,” meaning they do not meet full criteria of the current classification.

The choices made concerning which sector of health care to use for mental-disorder–related problems are only beginning to be dispassionately studied. People with free care more often choose psychiatrists and use psychotropic drugs. Older patients, blacks, and less-educated people also use more psychotropic drugs but make fewer visits. Men use mental health specialists more, women use primary care generalists as often as they do psychological experts, and elderly persons most often see generalists. The reasons for these differences remain unknown.

The Epidemiologic Catchment Area (ECA) study and others reveal important relationships between mental health care and physical illness care. For patients with mental disorders, there is a twofold increase in the probability that they will visit a primary care physician. In fact, 58 percent of patients with mental disorders visit a generalist during any 6-month period. This author's own study found that patients visiting a mental health specialist had made 11 visits to a general internist in the previous year. In the subsequent year, they also made 11 visits, but in this case 7 of the visits were for mental health problems and 4 were for physical problems. The use of a mental health specialist did not decrease the number of visits, but the care was more appropriate. Because care was less technologically based, costs decreased.

Overall, 16 to 25 percent of all ambulatory visits in the sample of the ECA study were for a Diagnostic Interview Studies (DIS)–assessed disorder. An English study estimates that an additional 5 percent are for psychologically related or psychosomatic disorders. Although claims in the psychological literature suggest that up to 80 percent of medical visits are psychologically related, 15 to 20 percent seems relatively correct for those not in high-risk groups. For example, a comparison of surveys from developed countries shows a prevalence of psychosocial problems in primary care ranging from 4 to 32 percent of patients, with a mean of 16 percent. However, the prevalence of mental disorder and seeking care for psychological reasons goes up markedly in the lowest socioeconomic stratum and the physically ill.

INTERFACE BETWEEN PSYCHIATRY AND PRIMARY CARE

Certain conditions that are not experienced by patients as mental disorders such as somatization are especially problematic for both psychiatry and primary care. These conditions present as somatic signs or symptoms and are not perceived by the patient as mental. Thus, patients do not seek psychiatric care for them and they resist the suggestion they might be having a mental problem (“If one more doctor tells me it is all in my head I will scream!”). In other situations, somatic symptoms are masked expressions of mental disorders such as panic disorder and depressive disorder. Physical conditions (e.g., asthma, hypertension, and abdominal pain) or drug adverse effects or placebo response are precipitated, exacerbated, or extenuated by psychological considerations such as lack of social support, stress, personality style, or depression. Conversely, many medical disorders present as mental problems to psychiatrists.

Table 28.1-3 lists some of the most important situations in which primary care and psychiatry overlap and are likely to need each other's help.

Table 28.1-3 Examples of Common Interface Conditions Between Psychiatry and Primary Care

PSYCHIATRIC CONDITIONS WITH PHYSICAL CAUSE

Just as the internist who is presented with somatic complaints must consider the psychiatric or psychological causes, so must the mental health specialist or primary care practitioner consider physical causes when encountering a patient with psychiatric complaints (Table 28.1-4). Everything reasonable must be done to rule out the common physical causes of psychiatric illness. Every patient with a mood or psychotic disorder requires careful physical screening to be certain that the disorder is not based in correctable or ameliorable physical disease. When psychosis or a mood disorder is first recognized, patients must have a thorough medical evaluation by a physician aware of the disease entities that might be responsible.

Table 28.1-4 Medical Problems That Present With Psychiatric Symptoms

Mood disorders due to a general medical condition exhibit the full range of affective pathology. It is common after many viral illnesses to experience depression for days or weeks. These brief, self-limited illnesses may include suicidal ideation and vegetative signs, such as loss of appetite and libido and fatigue, which are transient and usually related to an obvious viral illness. The recently discussed chronic fatigue syndrome and Epstein-Barr virus (EBV) syndrome clearly exist in a few patients; but they are the latest examples of medicalization of adjustment disorders, depression, and somatization. A positive Epstein-Barr virus titre does not rule out a primary psychiatric problem.

Depressions related to endocrinopathies, such as hypothyroidism, may be much more extreme and dramatic; these conditions are the most often missed serious medical causes of depression. It is still possible to find patients with undiagnosed thyroid disease on the back wards of large psychiatric hospitals. Table 28.1-5 lists major categories and examples of medical conditions frequently associated with mood disorders.

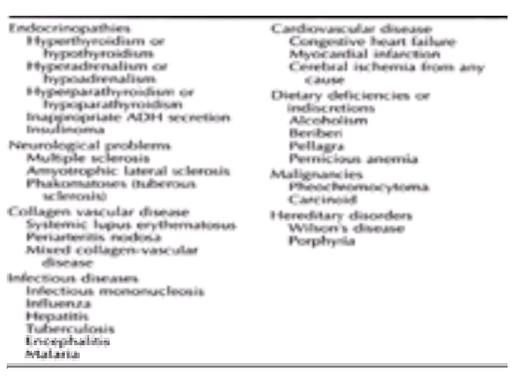

Table 28.1-5 Medical Conditions Associated With Mood Disorder

Among medications associated with the development of depression, propranolol hydrochloride (Inderal), one of the most commonly prescribed medications in the world, has an uncertain rate of adverse depressive effects but is often related to loss of libido and the development of impotence. The frequency of depression resulting from the other b-adrenergic receptor antagonists (beta-blockers) is uncertain. Many antihypertensive medications have clearer depressive effects, such as guanethidine (Esimil), reserpine and methyldopa (Aldomet). Hormones associated with depression may either cause depression directly or cause it on withdrawal. These hormones include adrenocortical steroids (cortisone and prednisone), estrogens and progestins (as in many birth control pills), and thyroid medications. Antiparkinsonian agents, including levodopa (Dopar), amantadine (Symmetrel), and bromocriptine (Parlodel), clearly cause depression. Equally, however, they allow expression of an already present depression, especially in recent-onset, severely incapacitated parkinsonian patients. The initial phases of prescription of these drugs are hazardous times, because suicide becomes physically possible for such patients. For example, one patient, upon being rescued from a decade of total immobilization in a hospital for the chronically ill, was able to leave the hospital and was driven home. One hour later, she got in her car and drove into a nearby lake. Withdrawal from stimulant drugs, such as amphetamines, cocaine, and methylphenidate (Ritalin), also may be associated with depression, as may withdrawal from adrenocorticosteroids such as cortisone and prednisone (Deltasone).

Many medical disorders are associated with psychotic states. Hyperparathyroidism and hypoparathyroidism, thyroid diseases, Cushing's and Addison's diseases, porphyria, and collagen vascular diseases, especially systemic lupus erythematosus and periarteritis nodosa, are among the most notable. Also important in this respect are seizure disorders (especially those involving the temporal lobes) and associated differential diagnostic entities, such as intracranial tumors and hemorrhage and atypical migraine. Posttraumatic conditions; sleep deprivation; the psychoses of neural degeneration, such as Alzheimer's and Pick's diseases, multiple sclerosis, and Huntington's disease; and any of the space-occupying and diffuse encephalitic conditions may be associated with, depression or psychoses. The number of conditions that can exhibit diffuse mental clouding, confusion, or delirium is enormous, but these states are less likely to be overlooked by the psychiatrist or mental health specialist because the mental status of the patient is clearly abnormal.

Drug-Related Psychosis Two main types of drugs are associated with the development of psychosis. One type is actually psychotogenic, causing frank psychosis that is related to the presence of a drug and associated with a clouded sensorium. In contrast, psychotomimetic drugs produce hallucinatory and other psychotic features while consciousness remains present and clear. Quinacrine (Atabrine) and related antimalarials may cause prolonged severe disorientation. Adrenocortical

steroids and adrenocorticotropic hormones are associated with euphoria, mania, and schizoaffective symptoms. When they are used in the treatment of systemic lupus erythematosus or other central nervous system (CNS) affecting collagen vascular diseases, the differentiation of steroid-induced psychotic disorder from CNS vasculitis challenges both psychiatric consultants and internists.

The most helpful approach is to attempt to establish baseline data that assess activity of the vasculitis process objectively. If possible, this attempt should be made premorbidly or when therapy with steroids begins rather than when the patient gets in trouble. This author has found it useful to compare the patient's electroencephalogram (EEG) findings with a baseline study; to measure complement concentration, consumption, and breakdown products; and to monitor the various antinuclear antibodies in the assessment of such cases. The psychotic collagen vascular patient often is too difficult to manage on a medical ward and requires the special services of psychiatric hospitalization. In these cases, continuity of care must be maintained by the internist or immunologist treating the vasculitis. Antipsychotic medication may be associated with a neuroleptic malignant syndrome with rapid onset of stupor, coma, delirium, mutism, rigidity, hypertension, tachycardia, and fever.

Endocrinopathies associated with psychosis include hypoparathyroidism and hyperparathyroidism and hypothyroidism and hyperthyroidism. Toxic agents are prominent offenders. Previously, mercury- and bromide-containing substances were important. Currently, those who inhale industrial hydrocarbons or abuse glue and cleaning fluids (especially adolescents) still present to emergency departments with psychoses. Many substances of abuse are associated with psychotic experiences, including all hallucinogens, (e.g., lysergic acid diethylamide [LSD] dimethyltryptamine [DMT]), phencyclidine, psilocybin, cocaine, and cannabis (marijuana). Alcohol withdrawal may result in alcohol withdrawal delirium (delirium tremens) usually easily detected because of the associated tremulousness, fever, and stigma of alcoholism. A delirium with hallucinations occurs with atropine- or belladonna-containing drugs and related compounds. When patients are taking antiparkinsonian medications, phenothiazines, or tricyclic drugs, belladonna-induced psychotic disorder may emerge. These medications also present a differential dilemma between psychosis that requires more medication and that produced by the drug itself. Here, the physical concomitants of atropinism, such as dry skin, mydriasis, tachycardia, urinary retention, and constipation, may be helpful.

The psychotomimetic drugs are associated with hallucinations, paranoia, restlessness, and altered associative processes without marked change of sensorium. They include amphetamines, LSD, cannabis, DMT, phencyclidine, and mescaline. Although any of these drugs may be suspected in adolescents with a substance abusing lifestyle, they may be missed in patients receiving these medications as part of another treatment. For example, amphetamine-like compounds are sometimes (needlessly) combined with bronchodilators in the treatment of asthma. Finally, antidepressants themselves may precipitate psychosis.

Substance Withdrawal Psychoses are associated with the withdrawal of some drugs; most prominent are the barbiturates, alcohol, and opioids. Treatment of these varied conditions requires specific knowledge of the condition or of the relevant pharmacology in the case of the substance-induced psychoses. Use of nonspecific antidotes in substance-induced psychoses may be fatal and should be avoided. Instead, creation of a supportive and reassuring environment, adequate monitoring of CNS depression and cardiorespiratory status, and time are often sufficient care. When a specific antidote is available, such as physostigmine (Antilirium) for anticholinergic-induced psychotic disorder or naloxone (Narcan) for opioid-induced psychotic disorder, it should be used.

DIFFERENCES IN THE EXPERIENCE OF MENTAL DISORDER IN PRIMARY CARE AND PSYCHIATRY

Several features distinguish the experience of the primary care physician from that of the psychiatric practitioner with respect to mental health. Most significant is that in choosing to see a psychiatrist patients must admit at some level that they have a psychiatric problem. Thus, mental health workers sometimes see persons at a higher level of awareness and acceptance. In contrast, patients can see their primary care doctor without acknowledging to themselves or the doctor that their problem is psychological. Thus, they may be less motivated, able, or willing to consider psychological explanations and treatments as relevant to their case and may strenuously reject such interpretations. While most mental health practitioners also describe problems of motivation and acceptance, these are strikingly more significant in primary care, where practitioners expect to work on these issues indirectly much of the time.

Psychological problems often present to primary care as physical complaints. Patients tend to bring all complaints to their primary care doctors initially and to focus primarily on physical complaints. Thus, of all the symptoms associated with a person's depression, the patient may highlight bodily distresses such as fatigue, impotence, or appetite or sleep change.

These considerations are especially salient with persons whose culture, religion, language, coping style, or personality are prone to deny psychological issues or to find them unacceptable, stigmatized, unsafe, or even taboo. For example, the Chinese culture lacks words for many common mental conditions and shuns discussion of negative affects. Many upper-class Anglo-Saxon communities also find such discussion distasteful. Many patients believe that if they are known to have a mental problem they will be teased or rejected and possibly lose their job or become unemployable. They fear that records are not confidential enough to prevent leakage of private information and resultant loss of privacy or insurance coverage.

Differences in Phenomenology Primary care doctors confront many psychological issues of adjustment or reaction that are not mental disorders, psychiatric illnesses, or signs of brain disease. Some of these relate to life-cycle stages—the birth of a child, puberty, adolescence, leaving home, having children, children leaving home, retirement, decline, senescence, and death—in the patient or others (e.g., death of spouse, the empty nest syndrome). Of course, mental health workers also deal with such problems, but perforce to a lesser extent.

Some common psychosocial problems presenting to primary care physicians are reactions to common events that are normal in populations of people but are experienced by the patient as unique, sometimes as disastrous (e.g., losing a job, developing a chronic illness, floods, war, or torture). Some are reactions to environmental problems such as floods or tidal waves, chronic violence, physical danger, or trauma to someone else. Here, the primary care physician is often in the best position to intervene before such issues evolve into full-fledged mental disorders (e.g., dissociative disorders, psychotic disorders, depressive disorder, posttraumatic stress disorder) or end in suicide. Recent data suggest that even 5 to 10 years after a major natural disaster, many patients still have significant depression, suicidality, or posttraumatic stress disorder.

At stressful times people develop physical symptoms, are more likely to get sick, and turn to their physicians for help. Most often, psychiatrists are not sought out in these situations because these persons do not regard themselves as mentally ill and the primary care doctor is more accessible. These conditions normally may have some anxious arousal, with symptoms of chest or abdominal pain, shortness of breath, agitation, sleeplessness or restlessness, among others.

The key to both diagnosis and treatment of defined mental disorders and psychological problems is the medical interview and the doctor-patient relationship that results. Extensive research has documented that in over 80 percent of diagnoses, the level of compliance, and patient and provider satisfaction are determined by the interview. Historically, deficiencies have existed in the teaching and monitoring of this core clinical skill. Most practicing doctors have never been observed doing a complete interview. Because how and what to teach is known and teaching is effective and durable, improving medical education in this area is a high priority.

Somatization Disorder Somatization disorder is a paradigm interface issue between the two disciplines. In this condition, patients experience psychological and social problems as physical symptoms and so seek medical care for them and resist psychological explanations for their problems. They are regularly told that their problem is mental and resent this suggestion, because they experience chest or back pain, shortness of breath, or similar symptoms. Patients with severe cases are high-stress, high-use, and high-frustration patients for both primary care physicians and those psychiatrists who do see them.

Somatization may be thought of as a transduction of psychological inputs into sensory ones, rather like synesthesia, in which sound may be experienced as smell or color (as is true for some musicians, e.g., in the blues). Such transductions occur in or accompany many mental disorders, particularly depression and anxiety. At least half of persons with major depressive disorder somatize, especially with fatigue, headache, and abdominal pain. By definition, persons with panic disorder have multiple symptoms such as shortness of breath, dizziness, chest pain or pressure, abdominal pain, and restlessness. More-disturbed persons (e.g., those with psychoses) have bizarre or distorted symptoms, so-called somatic delusions (e.g., “there is fire shooting out my ears.”).

Somatization is also a normal phenomenon in everyday life. Almost everyone does this in some ways at some times, often in culturally typical ways such as liver crisis *(crise du fois)* in France, heart symptoms in Germany, and abdominal pain *(agita)* in Italy. Somatization becomes a disorder when it leads the person to seek medical care and when it disrupts life or causes significant morbidity or disability. It is an enormous challenge for physicians because the symptoms that patients with somatization disorder experience are usually typical of other serious illnesses, and of course these patients get diseases related to physical pathology and die. One study suggested that of those diagnosed with somatization disorder about 25 percent eventually turn out to have demonstrable organic pathology in the affected system within 3 years:

Many disordered somatizers have related personality disorders, primary or secondary depression, or anxiety. Primary care doctors refer these patients to psychiatrists because they have ruled out organic problems, recognize obvious character and psychosocial pathology in the patient and the patient's illness-related behavior, or are at their wit's end. They find few psychiatrists even willing to see them in the first place and fewer who are willing to do more than rule out common mental disorders before returning the patient. Some providers get recruited in *splitting* (the behavior of some persons with character pathology to turn people they relate to against each other), in which the mental health provider accuses the primary care doctor of neglecting or failing the patient by not working up the patient's physical complaints. These patients are the best paradigm for the benefits of a team approach by mental health and primary care. When the primary care physician and the psychiatrist are working together, are physically near each other, and talk easily, the burden of these patients is lessened and splitting is less likely.

Adjustment Disorders If event or life cycle–related conditions are severe and do not resolve within a matter of weeks, they may be called adjustment disorders. These are named for the dominant psychological symptom (anxiety, depression), for example, adjustment disorder with anxious mood. If these persist for more than 6 months, they may be called chronic adjustment disorders. Some persons experience responses that do not resolve or that lead to activation or reactivation of a serious mental disorder. These are described, for example, as abnormal grief reaction (a frequent cause of somatization disorder) or adjustment disorder with depressed mood.

These conditions are more likely to be presented to the primary care physician than to a psychiatrist, but when they are severe, referral may be warranted or helpful. Making and managing such referrals must be done skillfully to avoid feelings of rejection, stigmatization, or both, and the no-show rates for such referrals are notoriously high. Here, again, physical proximity and even direct introduction of the patient to the mental health provider in the primary care physician's office or examination room can lessen losses to care. Effectively managing referrals to mental health also includes scheduling a visit back to oneself after the first visit to the mental health specialist, to emphasize that the referral is not a dismissal and to indicate an intention to participate actively in the ongoing care.

ADEQUACY OF PSYCHIATRIC TREATMENTS IN PRIMARY CARE

There is controversy about the adequacy of primary care treatment of mental disorders and of common psychological problems. Multiple data sets suggest that primary care physicians miss about 33 percent of major depressive disorder, over half of substance abuse, and at first pass, 90 percent of panic disorder. Giving the primary care doctor the results of simple screening devices used in advance of primary care visits, has little effect on the outcome of care or even in discussion of the detected problems.

Other studies suggest primary care physicians prescribe less medication, sometimes including substandard doses, for less time than do psychiatrists. This is taken as evidence of inadequate treatment. However, some data suggest that typical dosages (e.g., 150 mg of a tricyclic drug such as imipramine [Tofranil]) may not be needed in many, especially milder, cases. The cases primary care doctors miss are most often mild ones that often do not warrant drug treatment.

It is not surprising that mental disorder is missed in medical practice. One study showed that the average traditional internal medicine training program provides 2 hours a year of formal psychiatric education. Present-day primary care programs devote between 18 and 25 hours, and in some cases up to 80 hours a year, to training in psychosocial medicine. The general neglect of these issues persists despite initiatives of the American Board of Internal Medicine and other specialty boards mandating further training.

In parallel, some psychiatrists' care for primary care patients is variously criticized. Multiple studies suggest that much of the effect of psychiatric hospitalization is to increase medication use, invalidism, and time lost from work. Additional critiques have to do with a real or perceived lack of accessibility of psychiatrists. The primary care physician wants the psychiatrist to react and be helpful as soon as possible, not when there is an available opening. This is a significant cultural style issue-between the two disciplines. Psychiatrists maintain schedules and conditions of care that many patients and other doctors find impractical, which leads to patients dropping out of treatment or not making the visit. Psychiatrists, like many other types of practitioners, often find ways to eliminate from their practices the least attractive cases (e.g., unkempt, disruptive or manipulative persons) leaving the primary care system to cope with these taxing patients without backup.

Supportive Counseling in Primary Care The primary care physician facing a patient with a lifestyle or event-adjustment crisis normally uses an interview to elicit the nature and extent of the problem and then uses supportive counseling. The interview is the core tool of mental health care (Table 28.1-6). Much recent literature in primary care concerns the consensus that the interview can be conceptualized as having structures and functions. Each of the 14 structural elements (Table 28.1-7) and three functions (Table 28.1-8) is associated with specific behavioral skills that can be learned, improved, monitored, and taught.

Physician's Perspective	Patient's Perspective
Major medium of care	Major medium of care
Major time spent on interview	Major time spent on interview
Defines problem set	Defines helping set
Defines specific problems	Creates feeling of being understood
Provides context for understanding problems	Provides context for adapting to illness and understanding care
Establishes contract	Ensures goals are met
Medium of patient education	Learns how to care for self
Promotes compliance	Therapeutic alliance
Professional satisfaction	Health needs met: personally supported personal satisfaction
Interpersonal satisfaction	
Builds practice	Social needs facilitated
Facilitates healing	Comfort and cures

Reprinted with permission from Lipkin M Jr: In *Office Practice of Medicine*, W Branch, editor. Saunders, Philadelphia, 1987.

Table 28.1-6 Importance of the Medical Interview

Table 28.1-7 Structural Elements of the Medical Interview

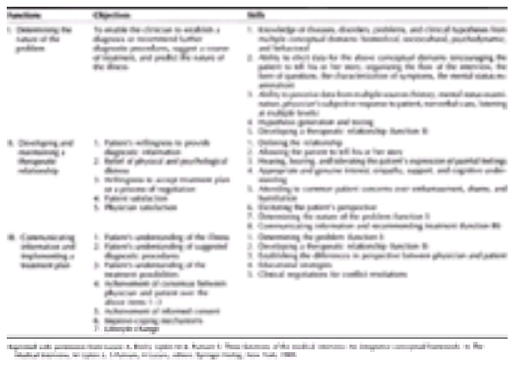

Table 28.1-8 Functions of the Medical Interview

The primary care physician in both developed and developing countries is faced with patients from a wide variety of cultural backgrounds because of differences in ethnicity, tribe, country of origin, language, and social class. It is often necessary to assess the patient's views of illness and healing to be effective in eliciting relevant information (it must be sought in culturally appropriate and sensitive ways) and in promoting patient adherence to agreed-upon plans. Arthur Kleinman has formulated a useful approach to elaborating a patient's health belief model, which is presented in Table 28.1-9 as a series of issues to consider and questions to use.

Prior illness episodes, especially of illnesses of standard severity (childbirth, renal stones, or surgery)
Cultural degree of stoicism
Cultural beliefs concerning the specific problem
Personal meaning or beliefs about the particular problem
Specific questions to ask to elicit the patient's explanatory model:
1. What do you call your problem? What name does it have?
2. What do you think caused your problem?
3. Why do you think it started when it did?
4. What does your sickness do to you? How does it work?
5. How severe is it? Will it have a short or long course?
6. What do you fear most about your sickness?
7. What are the chief problems that your sickness has caused for you?
8. What kind of treatment do you think you should receive? What are the most important results you hope to receive from treatment?
9. What have you done so far to treat your sickness?

Table 28.1-9 Assessment of Individual Illness Behavior in Becoming a Patient and Seeking Care

Normalizing the event and reaction is a first step. For example, at the loss of a spouse, one might say:

Your reactions are common. Many people feel the sorts of feelings you are having now and that feel so unique and terrible. And most people find that simply having the experiences, going ahead, and grieving, crying, or whatever works best for you, leads to gradually feeling better over weeks or months. It may take a year or two to feel yourself entirely again, and the special dates, anniversaries, birthdays, holidays, can be rough. You will never, of course, forget about your spouse, but over time the pain will lessen and the good memories will come to dominate.

Supportive counseling also involves such techniques as active listening, naming the feelings present (“sounds like you feel sad”), reflection, legitimization (“anyone who had been through that might feel as you do”), empathy (“I can really understand your pain”), and offers of support (“I will be here for you and go through this with you”). Often, the most useful act of the primary care doctor is to mobilize support, through family, neighbors, clergy, friends, employers, and nonmedical professionals such as social workers.

The primary care doctor may treat minor symptoms such as insomnia or anxiety with sedative hypnotics. Some believe that symptom-driven drug therapy interferes with the adjustment process, but such treatment is clearly indicated when the symptom interferes with vegetative functions (loss of sleep or more than transient loss of appetite), is incapacitating (“I can't do anything, I feel so agitated”), or extremely painful (Table 28.1-10).

Normalizing the feelings
Providing perspective
Suggesting a favorable resolution
Naming feelings
Reflection
Legitimization or validation
Expressing empathy
Expressing personal support
Mobilizing community and family support
Treating disruptive or damaging symptoms
Frequent visits

Table 28.1-10 Elements of Supportive Counseling

COLLABORATION OF PRIMARY CARE AND PSYCHIATRY

Psychiatry and primary care need effective partnership and mutual respect. The parallel sets of complaints highlight real problems on both sides of this critical interface. Although primary care doctors carry much of the mental health caseload in the nation, they do miss cases and often treat patients inadequately. They often find themselves in over their heads with nowhere to turn. Psychiatry as a discipline sometimes fails to care for nonremunerative patients, or to face dealing with the mental health of those in the ambulatory or medical hospital sectors. On the other side, managed care and other health insurance plans increasingly attempt to restrict direct access to psychiatrists, stipulate that care be given by social workers and psychologists, and limit the number of sessions with psychiatrists to as few as five. The two fields must cooperate. One way is to diminish the separation between the mental and the physical, so-called mind-body dualism, through mutual participation in training in both behavioral medicine and in primary care disciplines, at preclinical and clinical levels. A second technique is to have trainees at several stages of training (not just third-year clerkships) work on the other side of the fence. Of particular promise is having primary care residents do ambulatory consultations with mental health specialists and having psychiatry residents do real primary care training instead of the excessive inpatient medicine now typical.

Integrated Model of Education in Psychosocial Medicine The question of how to improve the recognition and treatment of mental disorders and psychological problems by primary care practitioners has been variously addressed. The American Academy on Physician and Patient has developed the most comprehensive approach. In this model, knowledge, skills, and attitudes are taught simultaneously, with a learner-centered style and a task orientation, in an attempt to change the behavior of the practitioner. The model has been shown to be effective and highly satisfying and to change behavior enduringly. Subjects taught have included the medical interview, communications skills, common aspects of psychiatry, somatization, pain, substance abuse, and teaching to teach, and the model has been used effectively with generalists and specialists.

Unfortunately, at this time most schools, residencies, care plans, and individual practitioners feel the time necessary to learn these complex matters is too great.

However, because missed cases generate needless testing, treatment, and iatrogenic harm, systems of care are increasingly interested in this area. For example, because each patient comorbid with depression who is hospitalized costs an additional $2500 in the ensuing year, medical care funders are interested. And because patients increasingly demand measurable effectiveness in meeting their needs, this will inevitably become a part of adequate primary care education.

Psychiatrists and Generalists Working Together What is not clear is how the various modern practice mixes of mental health and generalist practitioners will affect care outcomes and be accepted by patients. Funders want least-cost practitioners to care for mental disorders, but this is not popular with the public. Studies have suggested that having primary care and mental health practitioners physically in the same area improves referrals, outcomes, and patient and practitioner satisfaction. In one study, having selected patients alternate visits between primary care and psychiatry improved outcomes. When feedback is given concerning performance, when performance is tied to meaningful rewards, and when serious leadership attention is devoted to these issues, outcomes also improve.

There are practical and specific things to do to improve the effectiveness of individual collegial relationships between primary care and psychiatry. They boil down to good relationship skills and meticulous attention to the specifics, cultures, and feelings involved on each side. Some guidelines are outlined in Table 28.1-11.

Caring for the colleague
Cordial and polite demeanor
Intense and active initial listening
Developing a personal relationship
Taking the time to understand the other's world, preferably by training or spending time in it
Ascertaining the stated and implicit reasons for the contact
Negotiating differences
Agreeing on the mode(s) and frequency of contact
Avoiding use of jargon or insider talk
Agreeing on relative roles and responsibilities
Consulting on key decisions
Agreeing on who is to communicate what to whom (patient, family, staff)
Communicating in writing about consultations, discharges
Expressing gratitude for referral or consultation
Returning the favor of referral or consultation

Adapted from Lipkin M: Pulling together or falling apart. Prim Psychiatry, 4(1): 22, 1997.

Table 28.1-11 Guidelines for Effective Relationships Between Primary Care and Psychiatric Colleagues

When to Refer to Each Other One area in which specific guidelines may be helpful is when to call or involve someone from the complementary discipline (Table 28.1-12). Foremost, referral is warranted when the patient would benefit from partnership in the care. For primary care physicians, their ability to refer, their knowledge and skill levels, and their comfort with individual conditions each modify appropriate referral patterns. In general, referral is appropriate when the primary care doctor needs help. For psychiatrists, referral is always warranted when the patient does not have a primary care physician or when they need help with medical evaluation and treatment, especially when the patient may be somatizing.

From Primary Care to Psychiatry
When patient care would benefit from partnership
For therapeutically meaningful diagnostic uncertainty (e.g., is this patient psychotic or depressed?) but not for subtyping of depressions
When the patient does not get better on a full course of standard therapies
When the case is unduly complicated
When a specialized evaluation is needed (e.g., psychological testing)
When a specialized treatment is needed (e.g., antipsychotic medication initiation, desensitization therapy)
When the patient needs psychiatric hospitalization (e.g., decompensated, psychotic, suicidal)
From Psychiatry to Primary Care
When patient care would benefit from partnership
When the patient does not have a primary care doctor
When the patient needs a medical evaluation to complete the differential diagnosis
When the patient may have medical complications from the illness or treatment
To evaluate possible drug interactions
When the patient gets medically sick

Table 28.1-12 Referral Between Psychiatrists and Primary Care Physicians

The dominant mental health system is the primary care health system in all parts of the world including North America. This system has significant deficits in practitioner recognition and treatment of common mental disorders, psychological problems of the life cycle, and adjustment to life events including sickness and in interface areas such as somatization, psychophysiologic conditions, and stress-related disorders and complications. Remediation is possible and likely, given the existence of effective training models and social and health system imperatives that will demand more-effective care.

SUGGESTED CROSS-REFERENCES

Diagnosis, including signs and symptoms in psychiatry, is covered in Chapter 7. Clinical manifestations of psychiatric disorders is discussed in Chapter 8 and psychological factors affecting medical conditions are covered in Chapter 25. Other parts of Chapter 28 contain material related to primary care psychiatry such as chronic pain (Section 28.4), alternative medicine (Section 28.8), and death, dying and bereavement (Section 28.5).

SECTION REFERENCES

Barnes HN, Aronson MD, Delbanco TL: *Alcoholism: A Guide for the Primary Care Physician*. Springer-Verlag, New York, 1987.

*Clark W, Lipkin M, Graman H, Shorey J: Improving physicians' relationships with patients. J Gen Intern Med *14:*S45, 1999.

Fallowfield L, Lipkin M Jr, Hall A: Teaching senior oncologists communication skills: Results from phase I of a comprehensive longitudinal program in the United Kingdom. J Clin Oncol *16:*1961, 1998.

*Goold SD, Lipkin M: The doctor patient relationship: Challenges, opportunities, and strategies. J Gen Intern Med *14:*S26, 1999.

Kaplan C, Lipkin M Jr, Gordon G: Somatization in primary care: Patients with unexplained and vexing medical complaints. J Gen Intern Med *3:*177, 1988.

Katon W, Von Korff M, Lin E, Walker E, Simon GE, Bush T, Robinson P, Russo J: Collaborative management to achieve treatment guidelines. Impact on depression in primary care. JAMA *273:*1026, 1995.

Kleinman A: *The Illness Experience*. Basic Books, New York, 1988.

Levinson W, Roter D: The effects of two continuing medical education programs on communication skills of practicing primary care physicians. J Gen Intern Med *8:*318, 1993.

Lipkin M Jr: Pulling together or falling apart? Prim Psychiatry *4*(1):22, 1997.

*Lipkin M Jr: The Medical Interview. In *Behavioral Medicine in Primary Care: A Practical Guide*, M Feldman, J Christensen, editors. Appleton-Lange, Stamford, CT, 1997, p 1.

Lipkin M Jr, Kaplan C, Clark, W, Novack DH: Teaching medical interviewing: The Lipkin model. In *The Medical Interview: Clinical Care, Education, Research*, M Lipkin, SM Putnam, A Lazare, editors. Springer-Verlag, New York, 1995.

*Lipkin M Jr, Kupka K, editors. *Psychosocial Factors Affecting Health*. Praeger, New York, 1982.

*Lipkin M Jr, Putnam S, Lazare A, editors. *The Medical Interview: Clinical Care, Education, and Research*. Springer-Verlag, New York, 1995.

Novack D: Therapeutic aspects of the clinical encounter. Gen Intern Med *2:*346, 1987.

Paulsen RH: Psychiatry and primary care as neighbors: From the Promethean primary care physician to multi-disciplinary clinic. Int J Psychiatry Med *26:*113, 1996.

Weiner H: *Psychobiology of Human Disease*. Elsevier, New York, 1977.

28.2 PSYCHIATRY AND REPRODUCTIVE MEDICINE

SARAH L. BERGA, M.D., AND BARBARA L. PARRY, M.D.

Reproductive events and processes have psychological concomitants, but traditional medicine separates the treatment of reproductive events and psychological functioning and thus creates a problematic dichotomy between mind and body. This section addresses the effects of different reproductive events such as pregnancy, abortion, and menopause on endocrine and psychosocial functioning and highlights the bidirectional interaction between psyche and soma related to reproductive functioning. For instance, premenstrual syndrome—the mood, cognitive, and behavioral changes that occur in association with the menstrual cycle—exemplifies a somatopsychic disorder in which biological changes occurring in the soma trigger changes in psychological state. In contrast, hypothalamic or functional anovulation represents a psychosomatic illness that originates in the brain but alters somatic functioning. In keeping with a biopsychosocial model, key concepts about reproductive physiology are integrated into each section. The overall aim is to illustrate how the interaction between reproductive processes and psychosocial events manifests as psychiatric or gynecologic disorders.

REPRODUCTIVE DEVELOPMENT

Fetal and Neonatal Development The sex of an embryo is determined during the fertilization process. The normal chromosomal contribution of the oocyte (egg) is 23 (including an X chromosome), and the sperm contributes 23 (including an X or a Y chromosome). The translation of gonadal sex into phenotypic sex depends directly on the type and secretory activity of the gonad formed and on the responsiveness of target tissues in the developing fetus to gonadal products. The fetal testes begin to secrete testosterone and Muellerian-inhibiting substance (MIS) toward the end of the first trimester. If secretion is absent or compromised or if the fetal tissues are androgen insensitive, a spectrum of incomplete masculinization ensues. Also, excess androgen exposure at this time from maternal or fetal sources, such as occurs in fetal congenital adrenal hyperplasia, can partially masculinize a female fetus.

Sexual differentiation of the central nervous system (CNS) is believed to depend on the presence or absence of circulating levels of testosterone. The fetal testes begin to secrete testosterone in the late first trimester in response to placental human chorionic gonadotropin (hCG), while the fetal ovary does not. The organizational effects of testosterone on the developing CNS are thought to depend primarily upon in situ aromatization (the conversion of androgens to estrogens by the enzyme aromatase) of testosterone to estradiol. In contrast, estrogens from fetal or placental sources do not cross the blood-brain barrier and are not thought to imprint the developing CNS. Also, testosterone may bind directly (without conversion to estradiol) to androgen receptors in the CNS. The behavioral consequences that early exposure to testosterone has upon the developing brain are not clear, but brain areas with high aromatase activity (which thus can convert testosterone to estradiol) and androgen receptors in the nonhuman primate brain include the hypothalamus, amygdala, prefrontal visual, and somatosensory cortices. The asymmetry in exposure to testosterone also is present in early neonatal life. In late pregnancy, gonadotropin secretion is restrained by placental steroid production; when that restraint is lost at the time of birth, gonadotropin secretion rises dramatically in both sexes. In boys but not in girls, the gonadotropin rise is followed by an elevation of testosterone concentrations to adult levels. Thus, by 2 years of age, the brains and bodies of girls and boys have been exposed to dramatically different patterns of sex steroid secretion. The degree to which gender-related behavioral asymmetries are accounted for by those differences in hormone exposure is open to debate, but clearly a mechanism for inducing differences exists. In summary, the fetal sex steroid exposures exert primarily organizational effects upon the fetal CNS.

Puberty In humans, the postnatal gonadal hiatus involves the desynchronization of the hypothalamic gonadotropin-releasing hormone (GnRH) pulse generator by undetermined central mechanisms that are largely gonad independent. Puberty, the postnatal resumption of gonadal activity, is also a centrally initiated process that depends on synchronization of the pulsatile release of hypothalamic GnRH. GnRH neurons are endogenously pulsatile and active during fetal and neonatal life. In experimental conditions with primates, the entire pubertal process can be initiated and completed simply by providing exogenous pulses of GnRH or by stimulating the dormant GnRH neurons with excitatory amino acids. While prenatal exposure to sex steroids has organizational effects upon the developing brain, the reinitiation of gonadal activity during puberty is viewed as exerting activational influences upon the brain and behavior.

Gender The phenotypic sex of an individual depends upon chromosomal sex and on exposure and responsiveness of tissues to endogenous and exogenous sex steroids, but *gender* refers to the self-image and sex-role identity of that person. Gender is determined not only by exposure and responsiveness to sex steroids, but also by expectations and behavioral patterns that are learned in early childhood from parental, familial, and societal models. Gender develops across the life cycle in response to psychological and biological life events. Notions of gender direct marital, career, and other behavioral choices and may depart from biological options.

> A woman with *Muellerian agenesis* has ovaries but no uterus and vagina. Her childbearing and sexual options are constrained biologically, yet her sense of gender-appropriate behavior may drive her to pursue a heterosexual relationship and motherhood via surgical and technological advances.

Although gender is influenced by cultural expectations, each individual understands and responds to those expectations differently. To safeguard gender development, psychiatrists and psychologists must understand the medical and technological options available to individuals with disorders that alter phenotypic sexual development, while pediatricians, internists, and gynecologists must be mindful of the psychological effects of those disorders. Honesty is important, but physicians must convey information in a sensitive fashion. The clinical objective is to maintain a flexible mind-set to facilitate and support appropriate developmental choices.

Disorders of sexual development in phenotypic girls generally present at birth or at the time of expected puberty. *Adrenogenital syndrome* can be caused by congenital adrenal hyperplasia in female fetuses; the fetal adrenal secretes excess androgens, which then masculinize the external genitalia and possibly the brain. In female infants, the masculinized genitalia usually are recognized in the delivery room. Surgery to reduce clitoral size and create or widen the vaginal introitus may be necessary at a later date to restore the external appearance. However, the internal reproductive tract is normal and puberty generally occurs around the expected time if adrenal replacement therapy is adequate. Despite medical and surgical interventions, concerns over capacity for sexual function may emerge, particularly in late adolescence or young adulthood.

Turner's syndrome (XO gonadal dysgenesis) may be recognized at birth because of the associated physical stigmata, while XX gonadal dysgenesis usually presents at puberty. From a reproductive perspective, the two conditions are similar in that the ovaries do not contain responsive oocytes due to premature atresia of oocytes or failure of germ cell migration. Donor oocytes allow the option of pregnancy.

Other disorders of sexual development include *testicular feminization* and its variants; the gonads in that condition are testes, but the fetus is phenotypically female because a defect in the androgen receptor confers androgen insensitivity. Because the testes are normal, they secrete (MIS) and *Muellerian regression* occurs (regression of the anlage that would develop into the uterus and tubes). The vagina ends blindly, and the presenting complaint is usually primary amenorrhea. *Thelarche* (onset of breast development) occurs at the normal time because at puberty the testes secrete testosterone, which then is aromatized to estradiol that stimulates the growth of breast tissue. *Orchiectomy* (removal of the testes) is recommended after puberty to avoid the risk of gonadoblastoma, unless there is partial androgen sensitivity, in which case it is performed earlier to prevent the pubertal development of hirsutism and partial masculinization.

There are many types of androgen receptor defects and spectrum of clinical presentations. Some persons look phenotypically like men and may present with infertility secondary to azoospermia or oligospermia. A related disorder occurs when there is an inefficiency of the enzyme 5-a-reductase, which converts testosterone to dihydrotestosterone (DHT); DHT in turn masculinizes the external genitalia. Boys with a deficiency of that enzyme may look phenotypically female or incompletely masculinized at birth. At puberty, further masculinization, including phallic enlargement, may develop as the increased testicular secretion of testosterone partially

overcomes the enzyme deficiency and more dihydrotestosterone is made. Fertility is preserved in this condition, so it is prudent to rear the child as a male if possible. Neonatal treatment with dihydrotestosterone may help to masculinize the external genitalia.

XY gonadal dysgenesis is caused by a streak gonad that fails to secrete testosterone and MIS; generally the uterus, tubes, and vagina are present. This condition is detected when thelarche and menarche do not occur. A karyotype can confirm the XY chromosomal status. Because the gonads are inactive hormonally, they should be removed before puberty to avoid the risk of malignant degeneration. Exogenous hormone replacement is required to stimulate puberty and cause the development of secondary sexual characteristics. If donor oocytes are available, pregnancy is possible following in vitro fertilization and embryo transfer. Telling a young adolescent and her parents about the diagnosis can be difficult, but it is best to be honest, particularly since surgery will be required to remove the gonads.

A clinician who deals with patients with disorders of sexual development must have a clear understanding of the basic physiology of sexual differentiation, including that of the brain, as well as a sensitive perspective on the interactions among gender, sexuality, and reproductive capacity.

Early Sexuality The behavior of children can be understood by acknowledging the sexual motivations that, until Sigmund Freud's time, had been assumed to be quiescent until adolescence. The fact that incest, sexual abuse, and other sexual taboos that may occur in childhood have powerful after effects, even in adult life, bespeaks the importance of childhood sexuality in establishing healthy relationships.

In puberty, the overriding influence on sexuality is the dramatic change in the hypothalamic-pituitary-gonadal axis that occurs in both males and females at that time. Major hormonal changes, combined with the issue of identity diffusion, contribute to the characteristic tension in relationships, particularly between male and female peers and between adolescents and their parents.

In adulthood, one of the major indicators of maturity is the capacity to establish healthy sexual relationships. When in harmonious relationship to other components of the self and to others, sexuality is an integral component of the adult character.

Adolescence Carol Gilligan has noted in her longitudinal studies of girls before and during adolescence that there appears to be a marked decline in self-esteem occurring after puberty. Preadolescent girls exhibit a sense of self-confidence and a belief that they can influence the course of events in their lives. During adolescence, however, that self-esteem and belief in their own efficacy appears to diminish. Adolescent girls generally express less confidence in their abilities to change the world around them, accompanied by a sense of intimidation and passiveness. Indeed sex differences in depression, with female predominance, first appear at adolescence. Before then, boys are more vulnerable than girls to both medical and psychiatric illnesses.

One theory to account for this change in vulnerability to depressive illness in girls is the major hormonal changes in the reproductive neuroendocrine axis at puberty. Others have speculated that the increase in testosterone in males and the attendant aggressiveness, often directed at girls, contribute to girls' change in behavior. A large proportion of self-esteem in girls tends to be based on a sense of fulfillment in relationships, whereas self-esteem in boys generally is largely attributable to a sense of achievement in work, activities, and sports. Therefore, changes in relationships between the sexes during adolescence may affect girls and boys differently and have different consequences for their mental health. The need in boys for dominance in relationships, precipitated by surges in testosterone at puberty, may contribute to, and be a prerequisite for, a sense of self-esteem and self-worth. In contrast, passivity in girls relative to this dominance in boys may be required to maintain the relationship, which is girls' major source of self-esteem. Thus, changes in the reproductive neuroendocrine axis in girls and boys at puberty may set the stage for alterations in the power balance of relationships, which may affect the person's sense of self, self-esteem, and mental health. Interestingly, in a study of hormone replacement given to adolescent boys and girls with delayed puberty, testosterone precipitated aggression in boys while estrogen provoked aggression in girls. This study suggests that estrogen per se is unlikely to be the cause of the increased depression and loss of self-esteem observed in adolescent girls.

MENSTRUAL CYCLES

Menstrual Physiology Menstrual cyclicity results directly from ovarian cyclicity. Ovarian cyclicity starts each cycle with the development of a group or cohort of follicles, one of which becomes dominant. The follicles are composed of an oocyte surrounded by granulosa cells, which, in turn, are surrounded by theca cells.

As shown at the top of Fig. 28.2-1, follicular development is initiated by the hypothalamic release of GnRH at a pulse frequency of about once every 90 minutes. GnRH stimulates the release of the pituitary gonadotropins, luteinizing hormone (LH) and follicle-stimulating hormone (FSH). In turn, LH stimulates ovarian theca cells to synthesize and secrete androgens; FSH induces granulosa cell development, including the enzyme aromatase, which converts the thecally produced androgens to estrogens. In the presence of a constant GnRH pulse frequency of one pulse each 90 minutes, the secretion of LH and FSH will be regulated primarily by estradiol feedback at the level of the pituitary. Rising estradiol concentrations suppress FSH, thereby limiting the number of follicles that will develop mature oocytes.

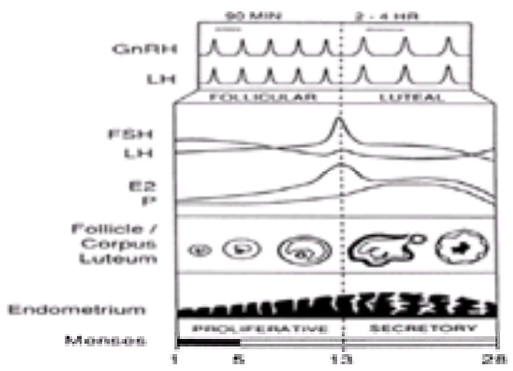

FIGURE 28.2-1 Schematization of the human menstrual cycle.

As illustrated in the middle panel of Figure 28.2-1, when estradiol concentrations exceed a critical threshold and remain elevated for at least 36 hours, which is the pattern one fully mature follicle produces, an LH surge is triggered, and ovulation ensues approximately 36 hours later. Thereafter, granulosa cells transform into progesterone-secreting luteal cells, and the ovulated follicle then is referred to as the corpus luteum.

The levels of LH, FSH, estradiol, and progesterone throughout the menstrual cycle and corresponding follicular events are schematized in the middle panels of Figure 28.2-1. The target tissues for ovarian steroids include the endometrium, whose developmental sequence is illustrated along the bottom right panel, and the hypothalamic GnRH pulse generator, whose frequency, as indicated in the top right panel, is slowed dramatically by the combination of estrogen and progesterone secreted during the postovulatory or luteal phase of the menstrual cycle. This inhibition of GnRH is followed by decreased secretion of LH and FSH, so that new follicular development is prevented until the corpus luteum regresses. As progesterone concentrations decline, GnRH pulsatility increases, and gonadotropin, especially FSH, secretion rises. The phases of the menstrual cycle can be termed follicular and luteal in reference to ovarian events or proliferative and secretory in reference to endometrial events.

It is generally assumed that regular menstrual cyclicity at intervals of 25 to 40 days signals ovarian cyclicity and ovulation, but as illustrated in Figure 28.2-2, that assumption is erroneous. A normal luteal phase has a length of more than 11 days and midluteal concentrations of progesterone secretion that exceed 10 ng/mL (30 nmol/L). The woman whose hormonal levels across a menstrual cycle are depicted in the bottom two panels of Figure 28.2-2 had a 39-day cycle. However, this cycle had a luteal phase whose adequacy with a length of 16 days and a peak progesterone concentration of 19.1 ng/mL cannot be doubted. In contrast, the woman whose hormonal concentrations are shown in the top two panels had an overall cycle length of 25 days. Although her estradiol levels rose to almost 300 pg/mL, the elevation in estradiol and LH was not followed by a rise in progesterone secretion. Thus, she had bleeding when her estradiol concentration fell, but the cycle was anovulatory.

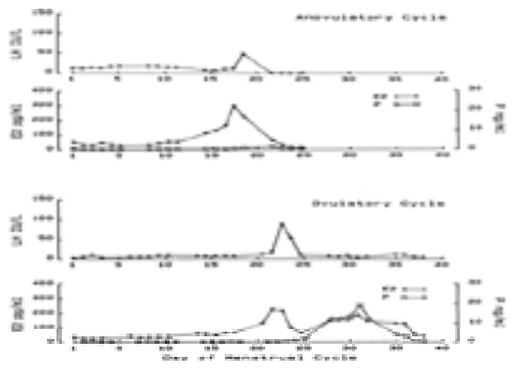

FIGURE 28.2-2 Daily concentrations of luteinizing hormone (LH), estradiol (E_2), and progesterone (P) obtained in two women from the onset of one episode of menstrual bleeding to the next. (By Tammy L. Daniels from original data.)

As these examples demonstrate, menses and menstrual cycle interval cannot be used to determine if a woman is or is not ovulatory or if the luteal phase was normal. The clinician who is linking a psychological state to a reproductive event must know how to determine the presence or absence of gonadal activity. Ovarian activity can be estimated by determining weekly estradiol and progesterone concentrations from the onset of one episode of menstrual bleeding to the next. Testicular activity can be estimated from one or two random determinations of testosterone. These assessments may not detect subtle compromise of gonadal function that may impair fertility, however, and further consultation with an appropriate specialist may be needed. Menstrual cycles can be expected to become more regular with advancing gynecological age (time since first menses) and, in the absence of pathology, generally remain regular until 5 to 10 years before menopause. During the perimenopausal years, ovarian function is characterized by higher estradiol and lower progesterone secretion. Ovarian function does not gradually decline but rather becomes erratic and unpredictable, with the potential for exposing the brain and soma to large fluctuations in hormone concentrations.

Premenstrual Syndrome Premenstrual syndrome (PMS), termed *premenstrual dysphoric disorder* in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), is a somatopsychic illness triggered by the expected excursions in sex steroids that accompany an ovulatory menstrual cycle. These somatic changes disturb psychological functioning in predisposed women. In contrast, functional hypothalamic anovulation arises in the psyche and affects the soma. In PMS, menstrual cycles are hormonally normal and the soma is preserved at the expense of the psyche. However, in a species that from an evolutionary perspective is accustomed to periodic pregnancy and lactation, there may be unappreciated physical and emotional consequences of incessant ovulation with attendant hormonal periodicity on the CNS that may place women with predisposing inherent vulnerabilities at high risk for mood disorders.

Functional Hypothalamic Anovulation Functional hypothalamic anovulation or amenorrhea (FHA) is a consequence of a nonorganic reduction in GnRH secretion that results in reduced pituitary secretion of gonadotropins and subsequent anovulation. Previously referred to as idiopathic or psychogenic amenorrhea in the psychiatric literature, FHA is a diagnosis of exclusion. That implies that the ovaries are capable of responding to appropriate gonadotropin input of either endogenous or exogenous origin; that there are no structural abnormalities of the thyroid, adrenal, pituitary, or brain; that the use of drugs, including antipsychotic medications, does not account for the suppression of GnRH; and that the patient is not pregnant. Neuroimaging may be needed to establish that there are no significant anatomical lesions of the brain or pituitary. Two major questions concerning the pathogenesis of FHA remain unresolved. First, the peripheral and central signals that disrupt GnRH pulsatility are poorly understood. Second, how do behavioral, cognitive, and personality variables activate the neural systems to disrupt GnRH secretion. Physicians and patients alike wonder what lifestyle variables provoke or contribute to this type of ovulatory dysfunction and what pharmacological and nonpharmacological treatment interventions should be considered.

The concept that psychogenic stress can induce reproductive dysfunction in women was formally introduced in 1946. The best biochemical evidence in support of the concept that stress impairs GnRH release in women with functional hypothalamic amenorrhea is the consistent demonstration that the activity of the hypothalamic-pituitary-adrenal axis is increased. Further, a prospective study found that young American women who developed transient amenorrhea while studying in Israel had higher urinary cortisol concentrations upon arrival than those whose menses remained regular.

There appears to be a dose-response relationship between the type, severity, and number of stresses on one hand and the proportion of women who develop anovulation on the other hand. Biological and psychological predispositions may confer resistance or sensitivity to various stressors. Exercise, low weight and weight loss, affective and eating disorders, various personality characteristics such as perfectionism and unrealistic expectations, drug use, and a variety of external and intrapersonal stresses have been linked to the development of anovulation. Most women with FHA, when carefully evaluated, display more than one of these traits or behaviors. Recent evidence suggests synergism between metabolic stressors, such as excessive exercise and nutritional restriction, that suppress the hypothalamic-pituitary-thyroidal axis and psychosocial challenges that activate the hypothalamic-pituitary-adrenal axis.

Recovery is possible if women with FHA develop response patterns to ongoing psychosocial demands that are less likely to activate the central and metabolic processes that disrupt pulsatile GnRH release. The current standard of practice other than observation is to offer pharmacological interventions such as oral contraceptives or hormonal replacement if fertility is not desired and pharmacological ovulation induction if fertility is sought. However, pharmacological intervention alone does not lead to spontaneous recovery and cannot be expected to ameliorate stress-induced alterations in central neurotransmission and hypothalamic function or to reverse ongoing metabolic derangements secondary to exercise or weight loss. For instance, bone accretion does not proceed apace, even if exogenous hormone replacement is given in supraphysiological doses, in the face of metabolic compromise. Although women with FHA frequently do not meet DSM-IV criteria for eating disorders or depressive disorders, all of these states are characterized by increased cortisol secretion, which alone can alter thyroid hormone secretion and action and induce metabolic adjustments. Thus, it is not surprising to find that women with amenorrhea due to decreased GnRH drive, regardless of cause, have lower bone mineral density. Further, pharmacological intervention alone may mask recognition of psychological dysfunction and forestall the development of more-effective response patterns. Also, ovulation induction may place low-weight women with functional hypothalamic amenorrhea who conceive at risk for premature labor and intrauterine growth retardation. If the parenting skills of women with the disorder are impaired by ongoing stress, their children may be at risk for poor psychosocial development. Clearly, treatment strategies need to consider that stress and mild psychological dysfunction can play an important role in the genesis of this form of anovulation. If an eating disorder is recognized, specialized psychiatric treatment is indicated.

Psychobiological characteristics can predispose a woman to chronic activation of central neural processes in response to commonplace events. Global hypothalamic dysfunction, including inhibition of GnRH pulsatility, can result. While identifying the neuromodulators that mediate the development and maintenance of the disruption of GnRH is of academic interest, clinicians must make sure that all other possible causes of anovulation and amenorrhea have been excluded before recommending psychological, psychiatric, or behavioral interventions.

FERTILITY AND INFERTILITY

Gametogenesis *Gametogenesis*, the production of oocytes and sperm, occurs as a result of puberty. Gametogenesis is the minimum physiological requirement for fertility. In women, the ovary must have oocytes that respond appropriately to gonadotropin input with the maturation of an oocyte that is ovulated. In men, the testes must be able to manufacture normal sperm capable of penetrating the layer of granulosa cells surrounding the oocyte and then fusing with the outer oocyte cell wall, the zona pellucida. The fallopian tube also must be capable of oocyte capture, and the density of sperm in the fallopian tube must be such that there is juxtaposition of sperm and oocyte. Thereafter, the tube must be capable of timely transfer of the fertilized oocyte or zygote to the primed endometrium of the uterus. Implantation must occur, and normal fetal development must ensue for the pregnancy to continue. In the absence of medical intervention, the customary manner in which ovulated oocytes are fertilized is via coitus or sexual intercourse.

Human Sexual Response It has been said that the most important sexual organ is the brain. This truism is meant to emphasize the essential role of desire and comfort in mediating sexual awareness and expression. Brain centers mediate libido in a general way. Other brain centers influence the selection of sexual partners and the circumstances under which sexual longings are expressed. Hormones such as estrogens and androgens made by the ovaries and testes bind in certain brain centers and increase sexual interest. Whether and how this sexual interest gets expressed as behavior depends upon a number of factors, including general health and well-being, the availability of a partner, a sense of what is appropriate, and prioritization of this interest above other interests and obligations. One of the major aims of social and cultural institutions is to channel or direct sexual interest in a manner that minimizes negative outcomes such as sexually transmitted diseases while ensuring that children will be born and prosper. Sexual interest can be inhibited by any number of negative emotions, including anger, fear, worry, and dislike,

and by a marked decline in health or hormones. Just like other personality factors, sexual interest or libido has a spectrum, with some people having more and others less. *Sexuality* is the term used to refer to a person's gender identity, feelings, sexual orientation, and attitudes, and it is distinct from the expression of sexuality and sexual behavior.

The human sexual response has been described as a cycle with four stages: excitement, plateau, orgasm, and resolution. The sexual response cycle is not simply a mechanical chain of events, however; it involves specific biological responses to psychological and sensory inputs. Thus the transition from one stage of the human sexual response to another is not automatic even if it is stereotypical. The human sexual response also depends on important biological events. First, there must be vasocongestion, a process in which an increased amount of blood concentrates in the tissues of the genitals and female breasts. Second, muscle tone must increase. For these two processes to take place, the nervous system must function appropriately and there can be no significant peripheral nerve impairments. If these biological processes cannot take place because of injury or illness, a person may have sexual feelings, but these feelings may not lead to a classical sexual response, such as vaginal lubrication or full penile erection.

Vaginal lubrication in women and penile erection in men are the most noticeable signs of the excitement phase. In women, this phase also involves internal vaginal expansion and nipple erection. In the plateau phase, the outer portion of the vagina swells, and the external aspects of the female genitalia, the labia, thicken and change to a darker color because of vasocongestion. In men, the testes increase in size and are pulled up against the body. A rash called a “sex flush” may develop on the upper torso. In both sexes, orgasm involves a series of muscular contractions that then diminish in intensity and rapidity. In women, there are muscular contractions of the outer portion of the vagina, the uterus, and the anal sphincter. In men, orgasm begins with contractions in the prostate and seminal vesicles, which cause ejaculation. During the resolution phase, the changes reverse to an unaroused state. Men, but not women, have a refractory period after orgasm in which ejaculation and orgasm cannot occur. In women, the orgasmic response is the same, regardless of whether there is manual stimulation of the clitoris or penile insertion, although the tempo of the sexual response may vary depending on the type of stimulation.

Aging may be associated with changes in sexuality. While sexual interest declines to some degree with aging, older men and women who live together are more sexually active than those who are not in a relationship. Because women tend to live longer than men, many elderly women will be without partners and will have limited opportunities for sexual expression, even if sexual drive is present. In men, significant changes in erections and ejaculation occur with aging. Usually it takes more stimulation to achieve an erection as a male ages. The plateau phase of sexual excitement is longer and may not end in ejaculation. When ejaculation occurs, the force of expulsion of semen and the intensity of ejaculation are lower. After ejaculation, the erection resolves more quickly and the refractory period increases. Many of these changes start to occur when men are in their 20s, and age-associated changes should be anticipated and viewed as normal.

Sexual Dysfunction in Women Sexual dysfunction can be due to psychological, hormonal, anatomical, and medication-related causes. These causes are not mutually exclusive. Psychological causes include inhibited sexual desire, which can be generalized or specific to one partner. If loss of libido is specific to one partner, then the prognosis for recovery is good. If libido always has been low and is generalized, then it is unlikely to change. Loss of interest in a given partner is difficult to restore, unless it is a temporary response to an identifiable event or issue that can be addressed in some mutually satisfactory way. Events such as childbirth that alter self-image and self-identity in both men and women may affect relationships libido, and the human sexual response in a variety of ways. As relationships are by their nature ever changing, it is to be expected that the sexual relationship will change with time; however, that change need not be for the worse. The intimacy inherent in sexual relationships requires continual renewal, so sexual relationships by nature may be vulnerable to the vicissitudes of life and aging. Decreased libido generally leads to less sexual arousal and unsatisfactory sexual contact for one or both partners. If neither partner views the sexual relationship as essential, then these developments may not threaten the stability of the overall relationship, but if there is a mismatch in outlook, alterations in libido and sexual expression may threaten the foundation of the relationship.

Hormonal changes can affect both libido and the reproductive tract. Loss of estrogen with menopause or from surgical removal of the ovaries can lead to loss of vaginal lubrication. Intercourse may be painful, or more typically, there may be less sexual sensation and decreased arousal. This may or may not decrease overall satisfaction within the relationship.

Anatomical reasons for sexual dysfunction include congenital malformations of the reproductive tract and results of surgery or childbirth. When extensive surgery is performed that involves the external female genitalia, sexual dysfunction may be due to loss of sensation from unavoidable injury to nerves, from pain at the surgical site, from loss of hormones if the ovaries were removed, from an altered sexual self-image, from fear or hesitation in the partner, or from any combination of these factors.

Organic conditions such as endometriosis, ovarian cysts, and pelvic infections can cause painful intercourse and sexual dysfunction. Diabetes can lead to nerve injury and sexual dysfunction due to reduced sensation. Other neurological conditions and alcohol use disorders can impair sexual responsivity.

Vaginismus refers to a condition in which the muscles around the outer third of the vagina have involuntary spasms in response to attempts at vaginal penetration. While this condition is generally thought to be psychological, women with vaginismus typically enjoy foreplay and can become sexually aroused and have orgasms. The male partner may develop erectile dysfunction in response, particularly if he becomes impatient or feels guilty.

Anorgasmia, or *orgasmic dysfunction*, refers to the inability to achieve an orgasm. This is not always viewed as a problem by women who report it. Anorgasmia is thought to be the most common sexual dysfunction, but it must be viewed in the context of each partner's desires and expectations. Therapy is generally not successful when the primary motivation for seeking medical intervention is to satisfy the male partner. The male partner may feel sympathetic, but he also may feel threatened, since many partners assume responsibility for the other's sexual satisfaction. A mismatch in libido or expectations can be more distressing than an identified problem.

Dyspareunia, or painful intercourse, can result from vaginal atrophy due to inadequate estrogen exposure, from endometriosis, from pelvic infections, from other anatomical conditions, or from lack of sexual lubrication.

Medications that may interfere with sexual responsivity or sexual interest include antidepressants, oral contraceptives, levonorgestrel implants (Norplant), medroxyprogesterone (Depo-provera), and GnRH-agonists. Contraceptive agents and GnRH-agonists likely act by altering hormone concentrations. In general, progestins diminish libido and sexual responsivity while estrogens increase both. Antidepressants decrease sexual drive and responsivity by acting upon important brain centers. Selective serotonin reuptake inhibitors (SSRIs) commonly lessen sexual interest. For SSRIs with shorter half-lives such as sertraline (Zoloft) and paroxetine (Paxil), it may be possible to reverse the effects on sexuality without interfering with the antidepressant effect by taking weekend drug holidays. Some antihypertensives also are associated with diminished sexual drive or satisfaction. If a close temporal link exists between the start of medication treatment and decreased sexual drive or responsiveness, it may be worth considering another medication, if a good alternative exists.

Appropriate treatment depends upon an accurate diagnosis. Organic or anatomical causes must be considered and may require medical or surgical remedies. The use of dilators may be helpful for women with vaginismus. Behavioral therapies such as sensate focus exercises, desensitization strategies for anxieties, relaxation training, assertiveness training, and directed masturbation have been successful, alone or in combination. Depending on the circumstances, couples counseling and interpersonal therapy may be recommended. The success of treatment depends upon the diagnosis.

Infertility *Infertility* is estimated to affect 10 to 20 percent of all couples. With the advent of sophisticated techniques for both diagnosing and treating the causes of infertility, many couples who in a previous era might have remained childless or adopted a child to build a family now can have biological children. Common causes of infertility include endometriosis, impaired spermatogenesis, damage to fallopian tubes or the vas deferens from infections or sterilization procedures, and ovulatory dysfunction or anovulation. The basic infertility evaluation can involve a semen analysis, hysterosalpingogram, hormonal studies, endometrial biopsy, and diagnostic laparoscopy to define possible contributing or causal factors (Table 28.2-1). Ovulation and ovulatory adequacy can be monitored by obtaining serial blood samples, performing an endometrial biopsy, or both. In contrast, spermatogenesis is difficult to assess because the process takes 70 to 90 days to complete and the morphological features of sperm that can be assessed by a standard semen analysis do not correlate well with function (i.e., ability to fertilize a human oocyte). Diagnostic uncertainty and a seemingly endless array of treatment approaches can easily provoke tension, indecision, and disagreement in the couple with infertility.

Possible Cause	Test	Comments
Anovulation	Basal body temperature chart	Patient must do each morning
	Endometrial biopsy	Office procedure in late luteal phase
	Serum progesterone	Single or multiple blood tests
	Urinary ovulation detection kit	Home use at midcycle to time intercourse
	Ultrasound	Visualize ovarian follicles and their rupture
Anatomical disorder	Hysterosalpingography	X-ray in proliferative phase to define intrauterine contours and tubal patency
	Diagnostic laparoscopy	View external surfaces of internal structures
	Hysteroscopy	Visualize endometrial cavity
Abnormal spermatogenesis	Semen analysis	Normal value >20 million per mL, 2 mL volume, 60% motility
	Postcoital test	Midcycle timing to look at sperm–cervical mucus interaction
Immunologic disorder	Antisperm antibodies	Aids semen test
Aspermia	Testicular biopsy	Determine eligibility for intracytoplasmic sperm injection
	Semen analysis for fructose	Determine if vas deferens are patent

Modified from Berkowitz GS, Lieg EW, Kennedy RH, Reiss GM, Herbert WNP: Lancet 1:1406, 1983; Ory SJ: Obstetrics and Gynecology for Office and Clinic, Philadelphia & Toronto, Saunders, 1978.

Table 28.2-1 Tests in the Infertility Workup

The sense that the window of opportunity is narrow can engender panic; the age of a woman correlates well with fecundity, and treatment success declines sharply after age 37 years, even when ovulation occurs regularly. Although fertility also declines with age in men, the decline is gradual until age 50 years, and fertility may persist beyond age 60. The offspring of older mothers have an increased risk of chromosomal trisomies, particularly trisomy 21. Recently, an increase in autosomal dominant conditions has been reported in the offspring of older fathers. Of the identifiable causes of infertility, about 40 percent are thought to be due to male factors, 40 percent to female factors, and 20 percent to both. Anovulation is the most common cause of infertility is women. About 20 percent of men with suboptimal semen parameters have microdeletions of the long arm of the Y chromosome in a region thought to play a critical role in spermatogenesis. About 10 to 20 percent of infertile couples have unexplained infertility, which means that there is no identifiable cause.

The degree to which psychogenic or lifestyle factors play a role in infertility is controversial. Couples should not be told their infertility is related to stress unless there is documented evidence of sperm dysfunction, ovulatory dysfunction, or sexual dysfunction. However, because lifestyle variables such as tobacco exposure, alcohol consumption, drug use, excessive exercise, weight loss or gain, and psychogenic stress can compromise reproduction, it is worth reviewing lifestyle factors with the couple and informing them of the potential adverse consequences of these factors in a supportive manner. Psychogenic variables can interfere with coital frequency and reduce the likelihood of conception, and they can suppress the central hypothalamic-pituitary drive to the gonad and compromise the quality and quantity of gametes produced.

Standard interventions were once limited to restorative surgery and gonadal stimulation, but they now include a variety of assisted reproductive techniques, including in vitro fertilization with embryo transfer (IVF-ET) and intracytoplasmic sperm injection (ICSI). In assisted reproductive technique procedures, the ovaries generally are hyperstimulated to produce at least three or four mature follicles from which the occytes can be retrieved surgically. Sperm and oocytes then are mixed and supported in vitro to promote fertilization. If fertilization occurs, the fertilized ovum or zygote is transferred to the recipient's uterus. The chance of implantation if three or more zygotes are transferred is generally around 30 percent but can be lower if fewer embryos are available or if the woman is over age 38. About 25 percent of implantations (as evidenced by a transiently elevated serum b-hCG concentration) are not sustained. The likelihood of a successful full-term pregnancy increases dramatically if fetal cardiac activity is observed about 2 weeks after missed menses (i.e., 4 weeks after the gamete or embryo transfer). The main indication for assisted reproductive procedures is failure to conceive with conventional interventions. In some instances, the reproductive impairment is so severe that conventional interventions are unlikely to offer success and an assisted reproductive procedure is recommended primarily. ICSI is used when there is failed fertilization during a previous IVF procedure or when too few sperm exist for conventional IVF.

Gamete intrafallopian transfer (GIFT) is a form of assisted reproduction with similar rates of success in which three to four oocytes are mixed with sperm and placed into the fallopian tube during a laparoscopy. Fertilization occurs in the tube and thus is not observed. GIFT is indicated in the absence of tubal disease or severe male factor, that is, conditions such as unexplained infertility or endometriosis unresponsive to superovulation alone. When it is not possible to correct underlying causes, substitutions are possible. Thus, not only are donor sperm available, so too are donor oocytes and surrogate mothers who will gestate the embryo of another couple. Embryos can be cryopreserved indefinitely and potentially transferred to a biologically unrelated recipient or to the woman herself years after menopause.

Treatments for infertility are expensive and consume much time and energy. A sense of psychological and physical invasion is common. Infertile men and women may feel angry, damaged, or guilty. Recent studies hinting at a link between ovarian cancer and infertility or infertility treatments further increase anxiety and psychological turmoil. The potential for husband and wife to respond differently to the experience of infertility and its treatments, particularly any of the assisted reproductive technologies, can overwhelm customary adaptation patterns and lead to psychological decompensation. Even successful infertility treatment poses challenges. Assisted reproductive technologies carry a 25 to 30 percent chance of twins and an almost 1 percent chance of triplets. If quadruplets or more result, the couple must grapple with the option of selective reduction. A multiple pregnancy increases fetal and maternal risks such as congenital abnormalities and preterm labor. Further, multiple gestation complicates postpartum adjustment and carries a lifelong parenting challenge.

Whether or not it is treated, infertility can produce a variety of psychological reactions. Responses depend on personality attributes, including adaptability, cultural expectations, support systems, knowledge about reproductive processes, and the attitudes of the involved clinicians. Infertility can damage self-image and impair sense of health and integrity. Some persons feel they are being made to suffer for past wrongs; others blame the medical profession for being inadequate or inattentive. A sense of panic, helplessness, and loss of control may compromise self-esteem. Women and men alike generally are overwhelmed by the complexity, cost, invasiveness, and uncertainty associated with medical intervention or adoption. Almost all patients who undergo infertility treatment (Table 28.2-1) experience a loss of privacy as they cope with physicians, nurses, insurance carriers, psychologists, hospitals, and laboratory personnel. The sense of intimacy that reinforces marriage and mutuality may be challenged when sex and procreation are separated by technology (Table 28.2-2). Ultimately, infertility presents the specter of a psychological death. Thus, it is not surprising that patients with infertility report as much anguish and distress as patients newly diagnosed with acquired immune deficiency syndrome (AIDS).

Method	Comments
Ovulation induction or augmentation (multiple agents, particularly clomiphene or highly purified gonadotropins)	Stimulates multifollicular development and ovulation; may produce multiple births; used in anovulation, luteal phase deficiency, unexplained infertility and assisted reproduction
Infection of spermatogenesis	Used in men with idiopathic hypothalamic hypogonadism of functional or organic nature
Artificial insemination	Donor sperm is injected into the uterine cavity or the fallopian tubes; the sperm of the husband may be used if healthy
Gamete intrafallopian transfer (GIFT)	Transfer of collected oocytes and sperm into the fallopian tubes; zygote may also be transferred (ZIFT); used for infertility from endometriosis and unexplained infertility
Intracytoplasmic sperm injection (ICSI)	Injection in vitro of sperm head or sperm DNA to cause fertilization and production of embryos for transfer to receptive endometrium; may be used even if sperm are barely viable; genetic cause of male infertility may be transmitted to offspring
Gamete donors	Donation of sperm or oocytes to another couple. Donation can include oocytes only, oocyte cytoplasm only to restore reproductive competence to aged oocytes, or donor sperm alone, or any combination
Surrogate mother	Surrogate mother receives donated embryo and carries the baby to term; a highly controversial technique with unresolved legal ramifications

Data in part from Berkowitz GS, Lieg EW, Kennedy RH, Reiss GM, Herbert WNP: Lancet 1:1406, 1983; Seidel RS, Soules MR: Obstetrics and Gynecology for Office and Clinic, Philadelphia & Toronto, Saunders, 1978.

Table 28.2-2 Assisted Reproduction Techniques

Ideally, psychological support services should be available for all individuals and couples undergoing active evaluation and intervention for infertility. Unlike the traditional medical model in which diagnostic testing precedes and directs the course of treatment, in infertility, empiric therapy may precede more invasive diagnostic testing. Cause and effect can be difficult to identify, and magical or superstitious thought processes are common in patients and practitioners alike. Infertility can threaten marital adjustment, particularly if the partners react differently to the challenges. Although fertility is an attribute of a couple, blame may be prematurely or wrongly assigned to one person in the dyad. In addition to individual or couples counseling, groups such as Resolve can offer support and information. When donor sperm, oocytes, or embryos or surrogate motherhood are being considered, a thorough psychological evaluation is recommended in advance of treatment, to ensure that all parties agree to the treatment plan and understand the potential implications and ramifications.

Reproductive technologies have made it possible to separate biological and social parenthood. A single woman may choose parenthood. The use of donor insemination or other fertility therapies in single women is controversial. Most women who seek such services have considered the consequences of single parenthood and feel that they will be able to rise to the challenges. Nonetheless, psychological evaluation should be made available to ensure that important topics and consequences have been considered. No long-term follow-up on the mental health and development of children born to single mothers has been done.

Dysfunctional attitudes and disharmony constantly threaten the infertile person. Prophylactic evaluation is preferred in all treatment settings, but specialized services may not always be immediately available. Stress can compromise gametogenesis and libido; infertility treatment alone can be stressful enough to activate the central mechanisms that compromise reproductive function. Psychological support is intended to lessen the likelihood of this effect and reverse it if it is already occurring.

Mood and cognition may be altered by pharmacological agents used to treat disorders of ovulation or to hyperstimulate the ovaries. The agents include clomiphene citrate (Clomid) and other antiestrogens, GnRH-agonists, human gonadotropins, hCG, progesterone, and bromocriptine (Parlodel). Danazol (Danocrine), a synthetic androgen, may be used to arrest the growth of endometriosis as a surgical or medical adjunct to later infertility therapy. Treatment of infertility itself is anxiety provoking because of the uncertainty and sense of expectation. However, most of the agents do alter estradiol and progesterone levels, and the resultant hormonal fluctuations may trigger mood lability in sensitive individuals. Agents such as clomiphene, danazol, and bromocriptine may also have direct effects on the brain. If the use of those agents is indicated, psychological support should be offered in preference to anxiolytics and antidepressants. A further concern for women with infertility has been introduced recently. Research has suggested that infertility treatments or infertility itself may confer an excess risk for ovarian cancer. The data, however, are far from conclusive. The uncertainty of this link only serves to exacerbate anxiety and guilt. Infertility is a far more devastating disorder than most lay persons and physicians realize.

Pregnancy Loss Unlike infertility, which can continue indefinitely, pregnancy loss is a defined event with an end. Acute reactions to an isolated miscarriage or perinatal loss such as blame, guilt, anger, and denial gradually lead to acceptance and resolution. A single miscarriage does not predict poor future reproductive performance. *Recurrent pregnancy loss*, defined as three or more losses before 28 weeks of gestation, carries a less unfavorable prognosis and resembles infertility in its chronicity and sense of lost potential. Although it might be assumed that grief and emotional turmoil would be related to gestational age, this assumption is not supported by observations. Women who miscarry have more than twice the relative risk of a major depression, compared with community women. More than half of women with a history of major depression suffered a recurrence following miscarriage. Because of the significant risk of depression, acute intervention is recommended for all women and couples who experience a miscarriage or perinatal loss; services should be extended beyond the acute event when indicated.

Pseudocyesis Pseudocyesis is the development of the classic symptoms of pregnancy—amenorrhea, nausea, breast enlargement and pigmentation, abdominal distention, and labor pains—in a nonpregnant woman. Pseudocyesis demonstrates the ability of the psyche to dominate the soma, probably via central input at the level of the hypothalamus. Predisposing psychological processes are thought to include a pathological wish for and fear of pregnancy; ambivalence or conflict regarding gender, sexuality, or childbearing; and a grief reaction to lost potential following a miscarriage, tubal ligation, or hysterectomy. The patient may have a true somatic delusion that is not subject to reality testing, but often a negative pregnancy test result or pelvic ultrasound scan leads to resolution. Psychotherapy is recommended during or after a presentation of pseudocyesis to evaluate and treat the underlying psychological dysfunction. A related event, couvade, occurs in some cultures; the father of the child undergoes a simulated labor, as though he were giving birth.

Fertility Control

Elective Abortion Elective abortion is distinct from spontaneous abortion (i.e., miscarriage) (Table 28.2-3). Elective abortion is the planned termination of a pregnancy. About 1.5 million abortions are performed in the United States each year—346 abortions for every 1000 live births. In Western nations, most women who obtain abortions are young, unmarried, and primiparous; in underdeveloped nations, abortion is most common among married women with two or more children.

Spontaneous	Spontaneous expulsion of the products of conception before viability (500 grams or about 24 weeks from last menses)
Habitual	Three or more spontaneous abortions
Missed	Expelling of products of conception; usually caused by the presence of a blighted ovum and lack of fetal development
Threatened	Uterine bleeding or cramping and positive pregnancy test; must be distinguished from an ectopic (usually tubal) pregnancy
Incomplete	Spontaneous passage of a portion of the products of conception and retention of placental fragments that result in ongoing bleeding
Elective	Induced by medical or surgical techniques before fetal viability; techniques include dilation, evacuation, and curettage; suction curettage; injection into the amniotic sac of saline or prostaglandins; hysterotomy; prostaglandins with antiprogestins (RU-486) or methotrexate; medical indications include the detection of fetal abnormalities by ultrasound or amniocentesis

Table 28.2-3 Types of Abortion

Fifty percent of abortions are performed under 8 weeks of gestation, 26 percent between 9 and 10 weeks, and 13 percent between 11 and 12 weeks. The remainder occur after 13 weeks, with 1 percent occurring after 21 weeks. Table 28.2-4 summarizes the most common abortion techniques. In contrast to what was previously thought, recent studies show no significant untoward psychological effects of elective abortion on the mental health of women. In fact, in one study comparing the outcomes of women who sought treatment in a clinic because they thought they were pregnant, women who did turn out to be pregnant and chose to abort were reported later to be doing better emotionally and in pursuing personal, educational, and work endeavors than either those who were not pregnant or those who chose to continue the pregnancy. Another study suggested untoward consequences on the mental health and well-being of women who were denied the option to abort and thus were required to unwillingly carry the pregnancy to term and then raise an unwanted child. That study suggested untoward consequences on the mental health and well-being of the child as well. However, termination of a wanted pregnancy because of an abnormal karyotype or fetal anomalies can be traumatic and supportive intervention is recommended.

Type	Benefits	Risks
Cervical dilation and evacuation of uterine contents by curettage or vacuum aspiration	Most commonly performed procedure for termination of pregnancy; can be done before 24 weeks gestation	All care Uterine perforation Adhesions Hemorrhage Infection Incomplete removal of fetus and placenta
Menstrual aspiration (minisuction)	Can be done within 1 to 3 weeks of missed period	Implanted/ectopic pregnancy was removed Uterine perforation (rare) Failure to recognize ectopic pregnancy
Medical: induction (cervical dilation with laminaria followed by high dose of IV oxytocin)	Can be used for second-trimester abortion	Water intoxication Uterine rupture Infection
Intrauterine: hyperosmotic solutions (saline)	Can be used for second-trimester abortions	Hypernatremic crisis Heart failure Peritonitis Hemorrhage Water intoxication Intrauterine necrosis
Prostaglandin (oral, intravaginal, cervical, or intramuscular)	Noninvasive procedure Can be used in conjunction with antiprogestins (RU 486) or methotrexate	Expulsion of live fetus Missed ectopic Hemorrhage Incomplete abortion
Antiprogestins (RU-486)	Nonsurgical First trimester only	Incomplete abortion Hemorrhage
Methotrexate	Nonsurgical First trimester only	Teratogenesis Hemorrhage Incomplete results

Table 28.2-4 Abortion Techniques

Techniques for inducing abortion (Table 28.2-4) have been known for at least 5000 years. In view of the risks of childbearing and the requirements of child rearing, induced abortion was viewed historically as a means of modulating the maternal and societal risks of unrestrained childbearing. Induced abortion affords women a measure of autonomy over a unique and psychologically powerful bodily function, but it arouses personal and public ambivalence and inspires controversy.

Historically, the concept of therapeutic abortion before *quickening*, generally defined as the time at which the mother perceives fetal movements, was well accepted. In

the United States, legislative regulations regarding induced abortion were introduced in the early 1800s, but their intent was to protect women from sepsis and other complications associated with the procedures. Induced abortion, whether by surgical or chemical means, now entails minimal medical risk and is clearly safer than carrying the pregnancy to term.

Current legislative attention has shifted from safeguarding maternal health to the ethics of abortion and the right of women to exercise control over their bodily functions. Traditionally, ethical deliberations regarding abortion were reserved for religions. Because of the diversity of religious perspectives on the role of women and on procreation, consensus is neither likely or necessarily desirable. Nonetheless, the current legislative debate centers on who has the authority to regulate induced abortion and to what degree. The 1992 Supreme Court ruling in *Planned Parenthood of Southeastern Pennsylvania v. Casey* upheld the basic right of women to elect pregnancy termination before fetal viability but also upheld the right of the state to regulate and restrict abortion and to define viability. Further, in some states (e.g., Pennsylvania), women who are pregnant lost the right to select whether they wish to receive advanced life support in life-threatening situations. The Freedom of Choice Act put before the U.S. Congress recognized the right of states to regulate the strictly medical aspects of abortion to safeguard maternal health. It restricts the freedom of a women to terminate a pregnancy after, but not before, fetal viability. This legislative effort represented a compromise between those who argue against any medical mediation of the process and those who would grant women complete autonomy. Application of this law would hinge on the interpretation of the phrase “fetal viability,” a concept that is not easily defined. Making the option of induced abortion contingent on fetal viability reveals continuing ambivalence about allowing women the autonomy to make decisions about physiological processes that directly shape their lives. It is difficult to appreciate how a legislative body can entrust the life of a child to the mother and simultaneously limit the mother's autonomy in making reproductive decisions. Current legislative efforts have focused on restricting the techniques used to perform late-term abortions by criminalizing physicians. This debate regarding abortion is far from resolution and is likely to intensify as medical advances such as RU-486, the antiprogestin that can chemically induce abortion in the early first trimester, increase therapeutic options.

Medical abortion soon will be available in the United States. In countries where this form of elective termination is currently available, acceptance is high but not universal. The medical risks of either surgical or medical abortion are few, but both procedures require physician input. Medical abortion, however, requires less physician input, although it does require careful monitoring to ensure that the abortion occurs in a timely fashion. Rarely, hemorrhage occurs that requires surgical intervention or a transfusion. Medical abortion can potentially increase a woman's autonomy of her reproductive capacity.

Contraception Methods of contraception fall into four main categories: barrier methods, hormonal agents, intrauterine devices, and natural family-planning methods that require abstinence around the time of ovulation (Table 28.2-5). The perfect contraceptive would be completely reversible, free of adverse effects, and completely efficacious. Since no perfect contraceptive is now available, people must select a method that best meets their needs. requently, the responsibility for choosing and using a contraceptive falls primarily on the woman. This can lead to resentment, lack of compliance, and marital disharmony. For those in a stable relationship, contraceptive counseling can include both partners.

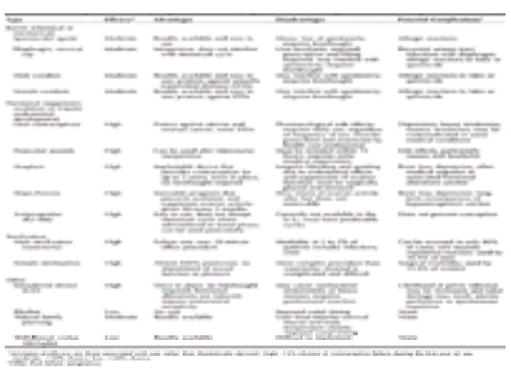

Table 28.2-5 Current Methods of Contraception

The assessment of the risks and benefits of a given contraceptive strategy should not be limited to a consideration of the biological risks alone but must take into account the mind-set of each party, particularly with regard to gender issues and cultural perspectives. If significant discord exists within a couple, counseling may be indicated. Counseling also can be considered for the person or couple who feel guilty about the idea of interfering with natural bodily functions for this purpose.

Clinicians must be mindful of the difference between actual and risk perception. Concerns of current and later bodily harm from use of contraception are common, and the risks of childbearing and child rearing frequently are underestimated. The mortality associated with childbearing is estimated to be 10 per 100,000 live births at ages 20 to 24, 44 per 100,000 between 35 and 39 years of age, and 71 per 100,000 after age 40. However, comparable mortality in nonsmoking women on oral contraceptives is 1.2 per 100,000 live births, 3.9 per 100,000 live births, and 6.6 per 100,000 live births, respectively. The potential risks of child rearing are difficult to quantitate, but they include role strain and stress, interrupted education or career advancement, depression, loss of independence, and compromise of the psychosexual relationship with the partner. This misperception of risk is particularly prevalent among adolescents, who may be uninformed and have unrealistic ideas about the rewards of pregnancy and the risks of sexually transmitted diseases. A teenager's aptitude for listening and making decisions may be limited by ambivalence and confusion. In short, despite the availability and relative affordability of contraceptive means, their use is constrained by the discrepancy between actual and imagined risk. Because coitus is not always anticipated there is a need for postcoital or so-called emergency contraceptives. Available methods include insertion of an intrauterine device (IUD), high doses of levonorgestrel, high doses of estrogen-progestin combinations, and antiprogestins. The use of levonorgestrel 0.75 mg taken twice 12 hours apart within 72 hours of intercourse was more effective (>99%) and associated with less severe adverse effects than the use of combined oral contraceptives. In the United States, only combined oral contraceptives (Preven) are explicitly marketed for use as postcoital contraception.

The mental and physical health risks of contraceptive use are significantly lower than those associated with pregnancy, parturition, and parenthood. However, contraception often gives an unwarranted illusion of control over a bodily process. No method is perfect, and unintended pregnancy can occur despite responsible use. Also, if a couple or person intends to have children, childbearing should not be postponed indefinitely in the hope that the perfect time will arrive. It can be difficult to conceive on demand.

Sterilization Sterilization can be accomplished by several methods of tubal ligation, by hysterectomy with or without oophorectomy in women, and by vasectomy in men. Hysterectomy for sterilization alone is associated with more morbidity and mortality than any method of tubal ligation; vasectomy is safer than any of the present methods of tubal ligation. Most studies do not support the contention that tubal ligation carries an increased incidence of gynecological sequelae, but failure rates are high in the first year following tubal ligation. In general, it is safest to perform a tubal ligation by laparoscopy at least several weeks following delivery. Tubal ligations immediately after delivery are done because of convenience or because of a woman's unwillingness or inability to return for a later tubal ligation.

It is difficult to determine how often reversal of a tubal ligation or vasectomy is sought or performed, and it is even more difficult to estimate the frequency of regret that the sterilization was performed. No formula can predict who will seek reversal, but factors such as remarriage, young age at time of sterilization, death of a child, and performing the sterilization during the postpartum period seem to play roles. Regret is high when sterilization is undertaken as a way to stabilize a marriage. The success of reversal depends on a variety of factors, including the method of sterilization, the ages of the partners, and the elapsed time since the sterilization procedure. Clinicians need to make couples and individuals aware of these considerations but should not use them as reasons to deny the planned procedure. Again, when a couple is involved, the mutuality of the decision should be discussed to avoid later resentment. Voluntary sterilization, especially vasectomy, has become the most popular form of birth control in couples married for more than 10 years.

SEXUALLY TRANSMITTED DISEASES

A *sexually transmitted disease* (STD) is a contagious disease acquired as a result of a physical sexual interaction. STDs always have been a reality, but from the 1950s through 1970s the infections were considered treatable and not life-threatening. AIDS, which is caused by infection with human immunodeficiency virus (HIV), is currently incurable, life threatening, and transmissible from mother to fetus. The specter of AIDS has captured the popular imagination. Although it was initially found in male homosexuals and intravenous drug abusers, HIV knows no boundaries. Sexual monogamy or abstinence has been advocated to ensure emotional health and a stable family arrangement for the potential outcome of sexual intercourse, children, but mutual sexual monogamy or abstinence also prevents STDs,

particularly AIDS.

Another sequela of STDs such as gonorrhea and chlamydia is pelvic inflammatory disease (PID). Untreated PID can develop into bilateral tuboovarian abscesses and necessitate hysterectomy and bilateral salpingo-oophorectomy. Early antibiotic treatment is advocated to prevent development of the abscesses and to reduce the likelihood of infertility, chronic pelvic pain, and ectopic pregnancy from tubal damage. These infections also can lead to obstruction of the vas deferens and chronic prostatitis in men and subsequent male infertility.

Another STD that can have serious consequences is venereal warts, or human papilloma virus (HPV). Genital infections with certain subtypes of HPV can lead to premalignant changes of the penis, vulva, vagina, and cervix and are thought to cause cervical cancer. Venereal warts can be removed chemically or surgically but are difficult to eradicate completely. Women who contract HPV are encouraged to have regular gynecological examinations and Papanicolaou smears to detect premalignant lesions.

Sexual monogamy and abstinence will prevent most STDs and are advocated as public health measures. However, libidinal impulses can be difficult to control and restrict. Therefore, measures such as condom use are recommended as an alternative public health measure. The ideal can be defined, but clinicians and policymakers need to remember that it cannot always be attained. Adolescents, in particular, need to know the potential consequences of sexual activity with regard to STDs and pregnancy. Admonishing teens to remain chaste is unlikely to be completely effective and may be counterproductive. To make responsible decisions, people need to understand the risks. Clinicians need to remember that the risks of sexual intercourse may be forgotten or seem minimal in comparison to the need for affection, contact, release, or escape. Persons who have low self-esteem or are stressed may view sex as a means of bolstering their self-image or escaping their stresses. The reinforcing properties of sex ensure that the problem of STDs will endure.

PELVIC PAIN

Pelvic pain can have many causes, including endometriosis, pelvic adhesions, ovarian or adnexal masses, hernias, and bowel or rectal disease. Pelvic pain can also be secondary to psychogenic causes, such as guilt, fertility or infertility fears, and the emotional disturbances associated with ongoing or past incest or sexual abuse. Pelvic pain should not be attributed to psychogenic causes unless a thorough evaluation has excluded organic causes. In most instances, the evaluation should include a diagnostic laparoscopy. Likewise, dyspareunia or pain with intercourse should not be assumed to have a psychogenic origin unless all anatomical causes have been excluded.

REPRODUCTION

Pregnancy, parturition (childbirth or delivery), and the postpartum period involve a series of tightly orchestrated hormonal events that create the potential for unique psychological states. As pregnancy advances, estrogen and progesterone concentrations rise, and a host of placentally elaborated polypeptide hormones induce metabolic adaptations. The effect of those hormonal excursions upon CNS activity and rhythms remains to be investigated; however, in the absence of predisposing psychological conditions or complications of pregnancy, most women adjust to the changing hormonal milieu without overt physical or psychological difficulty. Parturition is an intense biological event that is also tolerated by most, but it eventuates in the abrupt withdrawal of estrogens, progesterone, and other placental hormones. Parturition also triggers the neuroendocrine cascade that permits lactation. The metabolic demands of lactation, the need to deliver infant care across the usual circadian rest-activity cycle, and the hypoestrogenism that accompanies lactational amenorrhea present unique physiological challenges. Menses (and ovulation) generally resume in 6 to 12 weeks if the mother declines to nurse or when lactational frequency declines (usually after 4 to 6 months) if she does nurse.

Pregnancy Although pregnancy has dramatic emotional and psychological consequences, it generally protects women against the onset of mental illness unless there are underlying conditions. Teenage pregnancy appears to be a different story. For instance, pregnancy and motherhood protect against suicide, except in teenage mothers, who, despite their young age and negative personal or family histories of psychiatric illness, have an increased risk. Most likely that increased risk is attributable to the added stress felt by a young girl who is not yet equipped or ready to cope with the loss of social approval, the loss of control that pregnancy engenders, the demands of caring for a child, and the loss of future career or social options.

Pregnancy has not always been associated with low rates of suicide, however. Before the availability of contraception and legal abortion, pregnancy was often a cause of suicide, particularly if the pregnant woman was unwed and lacked social supports. In the past, punitive attitudes and adverse societal sanctions not only did not prevent unintended pregnancies, but were associated with increased maternal mortality from suicide and the complications of illegal abortions performed by unsafe techniques in unsterile settings.

Psychological Adaptation The stages of psychological adaptation to an intended pregnancy are not well understood. Both men and women who desire a pregnancy initially tend to feel empowered when it is achieved. However, as the pregnancy progresses and the realities of how life will be altered emerge, ambivalence develops. This is a normal psychological response to a major life transition. In general, most people adjust to the multiple bodily and psychological changes and develop effective coping strategies or attitudes. For instance, a husband may compliment his wife on her ever-changing physique as a means of reassuring them both that the child and parenthood are welcome, that the dyadic relationship is intact, and that he is sharing in the process. A woman may rejoice in her increasing girth and wear clothing that displays it, or she may wish to deemphasize the changes until they are dramatic. Medical advice and role models are helpful, but each woman must develop her own strategies to cope with the hormonal, metabolic, and psychological demands of pregnancy.

During the early stages of pregnancy, concerns tend to focus on bodily adjustment—how to cope with morning sickness, breast tenderness, changes in physique, sexuality, diet, and exercise. Depending on body habitus, increased abdominal size, breast size, or both may be apparent as early as 1 or 2 weeks after a missed menses or may not be apparent until 10 to 12 weeks later. Throughout pregnancy, the pregnant woman must adapt to changes in both physical and psychological boundaries. As the pregnancy advances, concerns shift to finances, the baby's room and material needs, employment issues, and capacity for parenthood. The extent to which the urge to prepare for the baby is biologically driven is not known, but in animals the preparatory efforts (nesting) appear to be hormonally initiated. In humans, this preparatory process does not seem to depend exclusively on hormones, as men also participate. Each partner may focus on a special concern, particularly if the marriage is based on discrete, stereotypic gender roles. However, throughout the pregnancy, each must share concerns and acknowledge ambivalence, to facilitate adjustment and mutuality. If this does not occur, the psychological and physical demands of pregnancy can lead to isolation and negative attitudes. Although parenting can be satisfying and deeply rewarding, it is a big responsibility; anticipatory anxiety is normal and focuses energies on important psychological tasks. The potential parents who have fantasies about the unlimited joys of parenting may not adjust well to the mundane, tedious realities associated with caring for babies and children. If the parents have other children, they must also develop strategies for coping with sibling adjustment.

The first pregnancy heralds a new life stage for both parents, but attention generally has focused on the developmental issues of the expectant mother. Pregnancy is seen as a maturational event that allows consolidation of gender identity and provides an opportunity to nurture the next generation. The psychological energy required to deal with issues of identification and differentiation in conjunction with profound metabolic and hormonal changes may lead to emotional lability, introspection, anxiety, or a desire for increased contact with one's parents. The age, life stage, and circumstances of the pregnant woman undoubtedly affect adjustment, but pregnancy is universally associated with concerns about adequacy during gestation and delivery and as a caregiver. In consolidating her identity, the woman must negotiate issues of autonomy, dependence, and sexuality. Past adjustment and current support play critical roles in determining the outcome of this pivotal developmental event. Although men do not have to confront the physical demands of pregnancy in themselves, the psychological adjustments required of men involve the same issues (i.e., consolidation of gender identity, attachment, autonomy, sexuality, and generativity).

Being able to anticipate facilitates adjustment. By definition, the first-time parent does not know fully what is required. The situation is made worse if the expectant mother or father has no guarantee of maternity or paternity leave. Allowing employers to make this decision on an individual basis is inappropriate and leaves too much room for employer self-interest. Standardized guidelines that guarantee a reasonable amount of financial and job security must be developed and applied uniformly. The needs of normal infants are predictable and do not vary by race, parental age, or socioeconomic status. Both parents need time and energy to initiate the psychological process of becoming a family. Obviously, the success of any economy depends on its human potential; strategies are needed that balance the short-term needs of business with the short- and long-term needs of children and parents. Working fathers have always been a reality; working mothers are now a reality and are likely to remain so. It is hypocritical and counterproductive to extol the virtues of parenthood but not develop reasonable provisions that allow working mothers and fathers to balance their commitments. As pregnancy can arouse primitive jealousies, working women who are pregnant may find that they are penalized in subtle ways for “neglecting” their job or other external responsibilities. Employment policies that protect pregnant women from undue pressure can help ameliorate that situation.

Psychological Tasks Associated With Assisted Reproduction Options and outcomes related to advanced reproductive technologies have introduced novel

psychological tasks that must be negotiated during and after pregnancy in addition to the more universal themes. The following situations are likely to be associated with enhanced anxiety and novel concerns: multiple gestation (Will the babies survive, and how can we parent so many at once?); surrogacy (Can I [or she] give up this baby?); donor oocytes or sperm (Who is the “real” mother or father?); selective fetal reduction (Did the “best” child survive?); single motherhood with donor sperm (Can I truly provide all that is needed? Who is the father?); in vitro fertilization or gamete intrafallopian transfer (Will the child be normal? How should I parent this special child? Who should know how this child was conceived?). These issues also arise in adoption and in single parenthood unrelated to assisted reproduction, but the elective decision to use reproductive technologies may make parents feel an enhanced sense of responsibility or predispose them to unrealistic fantasies about the child. The expectant parents in these situations may warrant prophylactic psychological intervention or, at the very least, especially attentive obstetricians and pediatricians.

Parturition Fears regarding pain and bodily harm during delivery are universal and to some extent, warranted. Preparation for childbirth affords a sense of familiarity and can ease anxieties, which facilitates delivery. Continuous emotional support during labor reduces the rate of cesarean section and forceps deliveries, the need for anesthesia, the use of oxytocin, and the duration of labor. A technically difficult delivery, however, does not appear to influence the decision to bear additional children.

Men's responses to pregnancy and labor have not been well studied, but the recent trend toward inclusion of fathers in the birth processes eases their anxieties and elicits a fuller sense of participation. If fathers are to be full partners in parenting, they must be given opportunities to participate and express themselves during pregnancy and parturition. Fathers do not parent the same way as mothers, and new mothers often need to be encouraged to respect these differences and view them positively.

Postpartum The postpartum period entails multiple physiological and psychological adjustments. Estrogen and progesterone concentrations decline precipitously, adrenal secretion of cortisol is altered, loss of placental hormones changes metabolism profoundly, and many new psychological challenges exist. Delivery itself also entails a metabolic challenge, especially if a cesarean section was performed. Maternal sleep deprivation is a common, if not universal, consequence of the infant's lack of circadian rest-activity cycle. Many of these factors can trigger postpartum depression and psychosis in biologically predisposed women. Recognizing postpartum depression, which occurs in as many as 10 percent of women, is a major problem, because screening and detection often fall to the obstetrician, who may only see the patient once at the postpartum checkup. It is critical for the obstetrician to ask about postpartum adjustment in a nonjudgmental manner and to indicate that the patient should call or return for help if she becomes overwhelmed. Optimally, the obstetrician should have an established referral process for new mothers who develop distressing psychological symptoms.

Lactation Lactation occurs because of a complex psychoneuroendocrine cascade that is triggered by the abrupt decline in estrogen and progesterone concentrations at parturition. Lactation becomes established because of the neurological stimulation transmitted by suckling. The composition and amount of breast milk changes as the infant grows. In general, babies should be fed as needed rather than by schedule. Breast-feeding has many benefits. The composition of breast milk supports timely neuronal development, confers passive immunity, and reduces food allergies. Psychological benefits accrue to both mother and infant. Every effort should be made by health professionals to convey these benefits. The workplace should include a place where women so interested can express breast milk.

Use of Psychotropic Medications In many instances the risk of psychiatric illness (e.g., depression or psychosis) is much worse for the brain and the interaction of mother and child than the potential adverse effects of psychotropic medication. If a woman becomes pregnant, however, medications should be evaluated immediately. The hormonal effects of pregnancy may change the course of her illness. However, it is not necessary to induce unwarranted guilt or to have the woman consider aborting if she does become pregnant while on psychotropic medications.

Definitive answers to the question of which psychotropic medications are safest during pregnancy and lactation are unlikely. While ethical considerations limit performing randomized, prospective interventional trials, observational data can and should be collected in registries. In patients with worsening psychiatric illness during pregnancy, outpatient psychotherapy, hospitalization, and milieu therapy should be attempted before routine use of psychotropic medication. Before a planned pregnancy, withdrawal of psychotropic medications should be attempted under close supervision. The importance of close rapport between the treating physician and the pregnant or breastfeeding patient cannot be overstated and will obviate or decrease reliance on psychotropic medication in many cases. The Food and Drug Administration (FDA) rates drugs in five categories of safety for use in pregnancy (Table 28.2-6).

Category	Definition	Drug Examples
A	No fetal risks in controlled human studies	Iron
B	No fetal risk in animal studies but no controlled human studies or fetal risk in animals but no risk in well-controlled human studies	Acetaminophen
C	Adverse fetal effects in animals and no human data available	Aspirin, haloperidol, chlorpromazine
D	Human fetal risk seen (may be used in life-threatening situation)	Lithium, tetracycline, ethanol
X	Proved fetal risk in humans (no indication for use, even in life-threatening situations)	Valproic acid, thalidomide

Table 28.2-6 FDA Rating of Drug Safety in Pregnancy

Antipsychotics No conclusive evidence indicates that antipsychotics are teratogenic, but they should not be used in the first trimester unless the patient presents a danger to herself, her unborn child, or others or hospitalization does not adequately control her psychosis. Use of these agents in the second and third trimesters is unlikely to cause fetal malformation, and there is no evidence of long-term sequelae in humans. The pediatrician should be alert to the possibility of behavioral disturbance in the neonate. In view of the potential for hypotension with aliphatic phenothiazines and thioridazine (Mellaril) and the possible increased risk of fetal malformation with chlorpromazine (Thorazine), use of high-potency agents appears preferable as first-line management; low-potency agents should be used only if unacceptable adverse effects occur with the use of other agents. Monitoring of uteroplacental sufficiency is indicated if severe hypotension occurs.

ANTIPSYCHOTICS IN LACTATION. The milk-to-plasma ratio of chlorpromazine is 0.3 to 1 and the ratio for perphenazine (Trilafon) is 1 to 1. The use of phenothiazines in lactating women has not been associated with serious consequences, and breast-feeding by mothers taking phenothiazines is not contraindicated.

Antiparkinsonism Agents Antiparkinsonian agents have not been adequately tested in pregnancy. In general, pregnant patients should not be routinely treated with antiparkinsonian drugs. The agents should be administered only to treat intolerable adverse effects associated with the use of antipsychotics.

Antidepressants Surprisingly little information exists on the use of tricyclic and other antidepressants in pregnancy, especially considering the frequency of major depression in women of reproductive age. Depression in the first trimester should be treated if possible by supportive measures. Cases of suicidality, incapacitating vegetative signs, or psychosis, warrant hospitalization. Use of antidepressants including SSRIs is warranted for vegetative signs accompanying a major depressive episode that do not resolve with supportive intervention. The available data show no statistical increase in the rate of congenital malformation in fetuses exposed to tricyclic drugs, but there are isolated case reports of limb deformities. Animal studies suggest possible teratogenic potential, but the studies are inconclusive. Amitriptyline (Elavil) and trimipramine (Surmontil) have been associated with adverse pregnancy outcomes in rats and rabbits. Trazodone (Desyrel) and amoxapine (Ascendin) have also been associated with poor fetal outcome in lower mammals. In general, therefore, while teratogenicity has not been proved, it is advisable to avoid using tricyclic drugs in the first trimester.

There has been concern about the effects of the use of tricyclic drugs in the second and third trimesters upon the neurological development of the fetus, particularly with respect to neurotransmitter systems. In addition, these agents may interfere with normal labor. Isolated case reports have documented a withdrawal syndrome in neonates, including cyanosis, difficulty breathing, and feeding difficulties after desipramine (Norpramin), urinary retention following nortriptyline (Pamelor), and dystonic movements and seizures after imipramine (Tofranil). These findings have led some authors to recommend a prelabor withdrawal period for patients on tricyclic drugs. However, no evidence indicates that intrauterine withdrawal is safer than extrauterine withdrawal. Since fetal hyperactivity and seizures are possible, extrauterine withdrawal may be safer. Like phenothiazines, tricyclic drugs can cause hypotension. A recent study by Irena Nulman found that the neurodevelopment of

children exposed to tricyclic drugs and fluoxetine (Prozac) during pregnancy and followed from birth until 86 months did not differ from that of the control group. Although another study reported more perinatal complications in women exposed to fluoxetine during pregnancy, the methods used to define the complications were not rigorous.

Whether intrauterine withdrawal is safer than extrauterine withdrawal is unclear, and the clinician should be guided by the mother's therapeutic response rather than assuming a benefit from withdrawing antidepressants prior to labor. If antidepressants are used in the third trimester, the pediatrician should be warned to anticipate the possibility of withdrawal symptoms in the infant, and any marked change in maternal blood pressure is an indication for obstetrical surveillance of uteroplacental sufficiency. Postpartum major depressive disorder should be treated first with supportive therapy, and transdermal estradiol, antidepressants, or both should also be offered to women who do not respond quickly. Estrogen therapy alone is usually not sufficient to treat major depressive disorder.

MONOAMINE OXIDASE (MAO) INHIBITORS The use of MAO inhibitors (MAOIs) is contraindicated in pregnancy for several reasons. First, studies found growth retardation in animals receiving doses in excess of the maximum human recommended dose. Second, pregnancy-induced hypertension is a common complication of pregnancy that often has an insidious onset; further exacerbation of that condition by use of MAOIs may lead to placental hypoperfusion and have serious fetal consequences. Third, if premature labor occurs, tocolysis with beta-mimetics may be obstetrically indicated but not possible because of potential interaction with the MAOI. Fourth, anesthetic management in labor may be complicated by the relative contraindication of opioids in patients taking MAOIs. When pregnancy occurs while the woman is taking MAOIs, the drug should be discontinued.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS A recent study showed that women taking fluoxetine during pregnancy do not have an increased risk of spontaneous pregnancy loss or major fetal anomalies, but women who take fluoxetine in the third trimester are at increased risk for perinatal anomalies. These investigators recommended that fluoxetine be discontinued if possible before the last trimester. However, other investigators found that the neurodevelopment of preschool children exposed in utero to tricyclic drugs or fluoxetine was unaffected. Specifically, they found no significant differences in temperament, mood, arousability, activity level, distractibility, global intelligence quotient (I.Q.), or language development in children exposed to SSRIs in fetal life.

ANTIDEPRESSANTS IN LACTATION. The use of antidepressants in the puerperium is a particularly important problem because of the clear increase in episodes of mood disorders during that time. The incidence of postpartum affective disturbance is reported to be between 10 and 15 percent. Factors predisposing to depression at that time include the rapid change in hormonal milieu which includes the abrupt withdrawal of estrogen and progesterone, decreasing serum cortisol concentration, changes in thyroid function, and the psychological and circadian stress of a newborn infant. Psychosis develops in a small percentage (0.1 percent) of new mothers. The evaluation of postpartum affective disorder should include thyroid function studies since these conditions are often associated. The treatment of postpartum depression should include immediate involvement in supportive psychotherapy. Often, the therapist can help the mother adapt to the stress of reorganizing the home to include the new baby and allay fears of the responsibilities of maternity and feelings of loss of her previous identity as a member of the workforce. However, in view of the lack of formal financial protection available to women in the workplace, concerns regarding employment and money may be well justified. Pharmacological intervention should be instituted early in the treatment. Alterations of sleep patterns are prominent symptoms resulting from both the affective disorder and the frequent awakenings by the newborn child.

The long-term effects on neurotransmitter development of antidepressants passed to the infant through the breast milk are unknown. Katherine Wisner and colleagues recently reviewed the literature and reported that amitriptyline, nortriptyline, desipramine, clomipramine (Anafranil), dothiepin, and sertraline were not found in quantifiable amounts in nurslings. Adverse effects were reported in some infants whose mothers were treated with doxepin (Adapin, Sinequan) and fluoxetine. While it is prudent for women to avoid antidepressant use during lactation, if the depression is severe, the alternative of a trial of an antidepressant should be presented in such a way that the new mother understands that her ability for effective interaction and bonding with the child may be enhanced by the increased emotional availability that may ensue. In almost all cases, patients who are not psychotic can decide for themselves whether to take antidepressants with or without discontinuation of breast-feeding or try supportive psychotherapy without pharmacological intervention. Transdermal estradiol may be used with antidepressants for women with postpartum depression without interfering significantly with lactation. Recurrence rates of postpartum depression are lower in women with a previous history of postpartum depression if antidepressant therapy is initiated within 12 hours after delivery.

Lithium and Antimanic Agents Lithium (Eskalith) is a suspected teratogen, and its use in the first trimester should be avoided. Its use was linked to development of Epstein's cardiac anomaly, although more-recent data suggest that this risk is lower than previously thought. Lithium crosses the placenta freely, and maternal and fetal plasma concentrations are similar. Reversible goiter from transplacental lithium poisoning can occur. Neonates exposed to lithium in utero have exhibited a variety of neurological toxic effects including muscle flaccidity, inhibition of normal neonatal reflexes, lethargy, and cyanosis and toxic cardiovascular effects, including atrial flutter, tricuspid regurgitation, and congestive heart failure.

Because of the consequences, lithium is best avoided throughout the first trimester if possible. When psychotic decompensation and destructive mania cannot be managed with antipsychotics alone, careful monitoring of lithium concentrations is required because of the dramatic shifts in fluid volume during pregnancy. A higher glomerular filtration rate, coupled with the increased plasma volume, often leads to a requirement for higher dosages in the gravid woman to achieve comparable serum lithium concentrations. However, dehydration often occurs during labor, especially in the hospital setting where oral intake may be restricted to guard the airway, and lithium toxicity may occur at this time. Moreover, a rapid loss of fluid volume after delivery may precipitate toxic effects. Similarly, because the newborn is subject to considerable volume fluctuations, especially with respiratory or gastrointestinal infection, toxic effects in the newborn are not uncommon. Despite these potential adverse events, however, a follow-up study of 60 babies exposed to lithium in utero failed to find more long-term sequelae in those infants without congenital malformations than in unexposed siblings, on the basis of maternal reports.

When lithium is used in the second and third trimesters, the clinician should monitor lithium concentrations at least every 2 weeks, perform an ultrasound on the fetus to exclude goiter before permitting a vaginal delivery, maintain adequate hydration during labor, decrease lithium to 50 percent of the usual dosage 2 weeks before the estimated date of confinement, and consider early delivery if decreasing the lithium dosage appears to adversely affect the maternal psychiatric condition or with evidence of pregnancy-induced hypertension.

CARBAMAZEPINE AND CLONAZEPAM Carbamazepine (Tegretol) and clonazepam (Klonopin) have not proved safer than lithium; however, in view of the high suspicion of lithium as a teratogen, the use of clonazepam instead of lithium is indicated in the first trimester if mania cannot be controlled with reasonable dosages of an antipsychotic. Anticonvulsives such as valproic acid (Depakene) and carbamazepine are used as mood stabilizers; their use has been associated with a 1 to 5 percent incidence of spina bifida or other neural tube defects. The risk is thought to be higher with valproic acid.

LITHIUM AND ANTIMANIC AGENTS IN LACTATION. The mean concentration of lithium in breast milk is roughly half the maternal serum concentration, with a range from 25 to 77 percent. During the first several months of life, the infant's kidney function is not fully developed, and the lithium excretion is slower than that of an older child. Lithium concentrations in the serum of breast-feeding infants are estimated to be a tenth to a half of those in maternal serum. Several reports of lithium toxicity in breast-feeding infants exist including cyanosis, T-wave abnormalities, and decreased muscle tone. Toxicity is especially likely during dehydration, such as during infection. Because of the toxicity, lithium is contraindicated while breast-feeding, because infants have greatly reduced renal clearance of lithium. Because of rapid metabolism, carbamazepine or valproate may be safer than lithium in the postpartum patient who cannot be adequately controlled on neuroleptics but who insists upon breast-feeding, providing the infant is carefully monitored.

Benzodiazepines Diazepam (Valium) crosses the placenta throughout gestation. Its use has been associated with cleft lip and palate. Whether diazepam is a teratogen in humans remains controversial. The likelihood is that diazepam is not teratogenic. Nonetheless, since its use is unlikely to be a medical necessity, it should be avoided in the first trimester. During the second and third trimesters, occasional use of benzodiazepines for discrete episodes of anxiety, indicated medical procedures, and severe insomnia is not likely to affect pregnancy outcome adversely.

BENZODIAZEPINES IN LACTATION. Diazepam and chlordiazepoxide (Librium) are excreted in breast milk. Therefore, nursing mothers should not use benzodiazepines.

Parenthood Parenthood is a major role that requires marked psychological adaptation. By necessity, a highly dependent human being becomes the focus of one's time and energy. Establishing an effective balance between the equally important roles of autonomy and dependence within each of the parents and between the spouses requires major shifts and reorganization.

In many cultures, additional family members live within or nearby the new family to provide support and share the time-consuming and energy-demanding burdens of the infant. Such extended families may contribute to additional role strain in the parents, but in toto they help stabilize the family unit. Increasingly in this country,

parents are being required to live up to the multiple roles of parent, spouse, friend, lover, and colleague in isolation. It often proves to be too much. Multiple competing responsibilities can precipitate child abuse, marital disruption, or mental illness. The old adage that it takes more than two people to raise a child seems well supported by experiences with parenthood. With all the stressors of parenthood, the situation of single parenthood (whether by choice or not) further accentuates these role demands. Often, single parents must seek substitute care in the form of baby sitters, live-in help, if affordable, and day care from schools or religious institutions. Since such resources are only minimally available and often very expensive, the role strain of parenthood without nearby family support systems may precipitate a state of crisis.

Even in modern society, women still have the primary responsibility for child rearing, so parenthood places particularly strong demands on the mother. Working and career-oriented mothers may be financially able to obtain other support systems to ease the demands of parenthood, but further role strain can be precipitated by demands and roles at work. Often women have multiple authoritarian demands at work but not enough power to carry out assignments. Psychological demand is further increased by the many responsibilities at home. Certainly men experience conflicts between family and work, but often their ability to use power in the workplace is balanced by their ability to receive nurturance at home, with relatively fewer demands and responsibilities and more ultimate power than women. The psychological strain on both men and women, particularly if it involves change and is not in keeping with their biological predispositions, exacerbates the customary demands of parenthood.

REPRODUCTIVE SENESCENCE

Menopause Menopause, the cessation of ovulation, is due to the depletion of responsive ovarian oocytes and generally occurs between ages 47 and 53 years. The hypoestrogenism that ensues can lead to hot flashes, sleep disturbances, vaginal atrophy and dryness, and cognitive and affective disturbances and predispose to osteoporosis, dementia, and cardiovascular disease. Menopause is generally heralded by 5 to 10 years of erratic ovarian function revealed by irregular uterine bleeding. Prior to the 1800s, menopause tended to occur just before death. Previous negative associations may reflect the nearly coincident timing of menopause and death at the turn of the last century; today's women, however, must adjust to the challenge of life after the loss of endogenous reproductive capacity.

Surgical menopause occurs when the ovaries are removed. That removal is often done in women older than age 35 years who need a hysterectomy to help protect them from future ovarian carcinoma. Prophylactic oophorectomy is a controversial procedure, and the symptoms that follow surgical castration before the age of natural menopause generally are more pronounced than those that follow natural menopause. In natural menopause, the ovaries remain intact and continue to secrete androgens, including testosterone and androstenedione which can be converted to estrone, a weak estrogen. Surgical menopause results in an abrupt and complete loss of the ovarian secretion of androgens, estrogens, and progesterone, often well before the age of menopause. Estrogen replacement therapy generally is indicated when bilateral oophorectomy is performed prior to natural menopause, but compliance may be compromised by ambivalence or a sense of premature loss if the decision to undergo oophorectomy was made without full consideration of all the risks and benefits. Androgen therapy is more controversial, in part because a tested, safe, physiological replacement product is not commercially available. The adrenal hormone dehydroepiandrosterone (DHEA) is an androgen currently being promoted as a dietary supplement to restore a sense of well-being, but few data are available on its long-term safety and effects.

Hormone Replacement Therapy The pros and cons of hormone replacement therapy have not been fully clarified; thus the question of whether to initiate or continue this therapy can pose a medical, financial, and philosophical challenge. Most studies show that estrogen replacement is associated with increased longevity, primarily because estrogen use confers cardioprotection. Estrogens also retard bone loss, retard urogenital atrophy, are mood stabilizing and neuroprotective, and retard the development of clinically apparent osteoporosis. Although women at midlife tend to gain weight, several studies have demonstrated that women who use hormone replacement therapy gain less weight and the weight gained is preferentially deposited in a gynecoid (hips and thighs) rather than android (abdominal) pattern. More-recent evidence suggests that ongoing estrogen use reduces the risk of Alzheimer's disease and other dementias by roughly 50 percent. Current evidence does not reveal a strong association between estrogen use and breast cancer. Although definitive studies have not been completed, evidence suggests that it is reasonable to permit women who have survived breast cancer to use hormone replacement therapy. Hypoestrogenism can impair sexual enjoyment and libido and cause hot flashes, night sweats, and profound fatigue. Those sequelae can generally be ameliorated or reversed with estrogen use. Progestin use is generally recommended for women with a uterus to guard against the development of endometrial hyperplasia and carcinoma; whether or to what degree progestin use compromises the beneficial effects of estrogen use needs clarification. The adverse effects that most limit progestin use are depression and bloating. Women who previously suffered from depression or PMS are most prone to mood disturbances with progestin use. To avoid or minimize the effects of progestin exposure, it is prudent to prescribe the smallest amount of progestin necessary to protect the endometrium from hyperplasia. In Europe, a progestin-containing IUD is used, which has the advantage that the progestin dose to the brain is minimized. In general, hormone replacement therapy enhances psychological functioning and promotes overall health. Recent data support the notion that antidepressant and antidementia agents are more efficacious in women when they are taking hormone replacement therapy. Since hormone replacement may last well over 30 years, it is worth titrating and individualizing route, dosage, type, and regimen with the goal of maximizing benefit and minimizing adverse effects.

The recent introduction of a wide range of estrogen and progestin products, including so-called designer estrogens or selective estrogen receptor modulators (SERMs) has expanded options. There is even less information about the long-term risks and benefits of partial estrogen agonists or SERMs than about more conventional estrogen preparations. Direct-to-consumer marketing of hormone preparations means that patients have many more questions, concerns, and ambivalence. To induce compliance, physicians must be able to explain what is and is not known. The same holds for phytoestrogens derived from soy and flax seed.

Both men and women age and experience an age-related decline in reproductive capacity, but only women experience complete gonadal cessation. Loss of reproductive capacity may present a psychological challenge to those who are not reconciled to the loss of fertility. However, even with gonadal failure, the availability of donor oocytes and sperm means that pregnancy can be initiated in a menopausal woman with an intact uterus who elects to pursue that option. Some women find these multiple options overwhelming. Menopause is not only a marker of life stage; it also presents biological and psychological challenges unique to women.

Psychophysiological Adaptation Depression at menopause was previously attributable to the empty-nest syndrome. In fact many women report an enhanced sense of well-being and enjoy the opportunity to pursue postponed goals. Recent acknowledgment that women's roles may encompass more than child rearing has helped facilitate this attitude. Thus, psychological factors may enhance well-being in older women.

In contrast, biological factors appear to predispose women to major mental illness at the time of menopause. A peak in the onset of bipolar illness in women around the time of menopause has been observed. One study reported induction of rapid mood cycling at the time of menopause in women. Yet another study found an increased number of affective episodes in women coinciding with the mean age of menopause at about 50 years.

Aging Further investigations are required to distinguish the psychological and biological effects of aging and menopause. Aging is associated with declined resiliency and circadian rhythmicity in both men and women. Desynchronized or reduced biological rhythms may compromise sleep, cognition, and mood. The brain is a target organ for gonadal steroids. Only women experience complete gonadal cessation. The loss of ovarian estrogen and progesterone exposure can precipitate hot flashes, decline in libido, sleep disturbances, affective complaints, and decreased memory, despite stable or improved life circumstances. However, both men and women experience a gradual decline in adrenal and gonadal androgen production. Androgens help maintain libido, muscle mass, strength, and energy. Androgen therapy has been advocated as a way to retard the aging process, but a tested, safe androgen replacement preparation is not commercially available for women. Many other chemotherapeutic agents have been advocated to reduce the burdens of aging. These include antioxidants such as vitamin E, phytoestrogen, and vitamin C, aspirin; alcohol; folate; vitamin D; calcium; herbal supplementals such as Ginko Biloba; and hormones such as DHEA, growth hormone, and melatonin. Menopausal women often expect that physicians who offer hormonal products at menopause to increase long-term health will also know whether other chemotherapeutic strategies have merit.

SUGGESTED CROSS-REFERENCES

Many of the reproductive transitions that may occur across a woman's life cycle are discussed elsewhere, including Section 27.5 on other conditions not attributable to a mental disorder, Section 13.4 on postpartum psychotic disorders, Section 28.3 on premenstrual dysphoric disorder, and Section 25.6 on endocrine and metabolic disorders. Chapter 19 on normal human sexuality and sexual disorders discusses important concepts regarding gender and sexuality. Chapter 20 discusses eating disorders, conditions in which ovarian cyclicity is often disrupted. Interactions between psyche and soma are discussed in Chapter 8 on clinical manifestations of psychiatric disorders, Section 25.9 on stress and psychiatry, and Section 1.11 on psychoneuroendocrinology. Interactions between self and society and the influences of culture are discussed in Section 4.1 on anthropology and psychiatry and Section 4.2 on sociology and psychiatry.

SECTION REFERENCES

*Adashi EY, Rock JA, Rosenwaks Z, editors: *Reproductive Endocrinology, Surgery, and Technology*, vol 1 and 2. Lippincott-Raven, Philadelphia, 1996.

Adler NE, David HP, Major BN, Roth SH, Russo NF, Wyatt GE: Psychological responses after abortion. Science *248:*41, 1990.

Aggleton P, O'Reilly K, Slutkin G, Davies P: Risking everything? Risk behavior, behavior change, and AIDS. Science *265:*341, 1994.

Appleby L: Suicide during pregnancy and in the first postnatal year. Br Med J *302:*137, 1991.

*Berga SL, guest editor: Brain, behavior, and reproductive function. Semin Reprod Endocrinol *15:*1, 1997.

Berga SL, Mortola JF, Girton L, Suh B, Laughlin G, Pham P, Yen SSC: Neuroendocrine aberrations in women with functional hypothalamic amenorrhea. J Clin Endocrinol Metab *68:*301, 1989.

Birge SJ: Is there a role for estrogen replacement therapy in the prevention and treatment of dementia? J Am Gerontol Soc *44:*865, 1996.

Chambers CD, Johnson KA, Dick LM, Felix RJ, Jones KL: Birth outcomes in pregnant women taking fluoxetine. N Engl J Med *335:*1010, 1996.

Cohen LS, Friedman JM, Jefferson JW, Johnson M, Weiner ML: A reevaluation of risk of in utero exposure to lithium. JAMA *271:*146, 1994.

Finkelstein JW, Susman EJ, Chinchilli VM, Kunselman SJ, D'Arcangelo MR, Schwab J, Demers LM, Liben LS, Lookingbill G, Kulin HE: Estrogen or testosterone increases self-reported aggressive behaviors in hypogonadal adolescents. J Clin Endocrinol Metab *82:*2433, 1997.

Giles DE, Berga SL: Cognitive and psychiatric correlates of functional hypothalamic amenorrhea: A controlled comparison. Fertil Steril *60:*486, 1993.

Gilligan C: *In a Different Voice. Psychological Theory and Women's Development*. Harvard University Press, Cambridge, MA, 1982.

Glasier A: Emergency postcoital contraception New Engl J Med *337:*1058, 1997.

Glasier A, Baird D: The effects of self-administering emergency contraception. N Engl J Med *339:*1, 1998.

Gregoire AJP, Kumar R, Everitt B, Henderson AF, Studd JWW: Transdermal oestrogen for treatment of severe postnatal depression. Lancet *347:*930, 1996.

Koren G, Pastuszak A, Ito S: Drugs in pregnancy. N Engl J Med *338:*1128, 1998.

*Knobil E, Neill JD, editors: *The Physiology of Reproduction*, vol 1 and 2, ed 2. Raven, New York, 1994.

Matthews KA, Wing RR, Kuller LH, Meilahn EN, Kelsey SF: Influences of natural menopause on psychological characteristics and symptoms of middle-aged healthy women. J Consult Clin Psychol *58:*345, 1990.

Neugebauer R, Kline J, Shrout P, Skodol A, O'Connor P, Geller P, Stein Z, Susser M: Major depressive disorder in the 6 months after miscarriage. JAMA *277:*383, 1997.

Nulman I, Rovet J, Stewart DE, Wolpin J, Gardner HA, Theis JGW, Kulin N, Koren G: Neurodevelopment of children exposed in utero to antidepressant drugs. N Engl J Med *336:*258, 1997.

Parry BL, guest editor: Women's disorders. In *Psychiatr Clin North Am*. Saunders, Philadelphia, 1989.

Person ES: *Dreams of Love and Fateful Encounters: The Power of Romantic Passion*. Norton, New York, 1988.

*Santoro N, Brown JR, Adel T, Skurnick JH: Characterization of reproductive hormonal dynamics in the perimenopause. J Clin Endocrinol Metab *81:*1495, 1996.

Schlesselman JJ: Net effect of oral contraceptive use on the risk of cancer in women in the United States. Obstet Gynecol *85:*793, 1995.

*Schmidt PJ, Nieman LK, Danaceau MA, Adams LF, Rubinow DR: Differential behavioral effects of gonadal steroids in women with and in those without premenstrual syndrome. N Engl J Med *338:*209, 1998.

Scolnik D, Nulman I, Rovet J, Gladstone D, Czuchta D, Gardner HA, Gladstone R, Ashby P, Weksberg R, Einarson T. Koren G: Neurodevelopment of children exposed in utero to phenytoin and carbamazepine monotherapy. JAMA *271:*767, 1994.

*Shawitz SE, Shawitz BA, Pugh KR, Fulbright RK: Effect of estrogen on brain activation patterns in postmenopausal women during working memory tasks. JAMA *28:*1197, 1999.

Sherwin BB: The impact of different doses of estrogen and progestin on mood and sexual behavior in postmenopausal women. J Clin Endocrinol Metab *72:*336, 1991.

Sourander L, Rajala T, Räihä I, Mäkinen R, Erkkola R, Helenius H: Cardiovascular and cancer morbidity and mortality and sudden cardiac death in postmenopausal women on oestrogen replacement therapy (ERT). Lancet *352:*1965, 1998.

Stadberg E, Mattsson LA, Milsom I: The prevalence and severity of climacteric symptoms and the use of different treatment regimens in a Swedish population. Acta Obstet Gynecol Scand *76:*442, 1997.

Stotland NL, editor: *Psychiatric Aspects of Reproductive Technology*. American Psychiatric Press, Washington, DC, 1990.

Sulak PJ, Haney AF: Unwanted pregnancies: Understanding contraceptive use and benefits in adolescents and older women. Am J Obstet Gynecol *168:*2042, 1993.

*van Hertzen H, van Look FA: Comparison of three single doses of mifepristone as emergency contraception: A randomized trial. *353:*697, 1999.

Wisner KL, Perel JM, Findling RL: Antidepressant treatment during breast-feeding. Am J Psychiatry *153:*1132, 1996.

CHAPTER 28. SPECIAL AREAS OF INTEREST

28.3 PREMENSTRUAL DYSPHORIC DISORDER

KIMBERLY A. YONKERS, M.D., AND LORI L. DAVIS, M.D.

A number of mood, behavioral, and somatic symptoms are reported by women during the late luteal phase of the menstrual cycle. For decades these signs and symptoms have been called *premenstrual syndrome*, a variably defined constellation of symptoms occurring during the premenstrual phase of the cycle. A subset of symptoms comprises *premenstrual dysphoric disorder*, which is found in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Premenstrual dysphoric disorder emphasizes mood symptoms and specifies that symptoms must lead to some degree of functional impairment, thus distinguishing it from milder premenstrual symptoms. Premenstrual dysphoric disorder is unique in that it is the only diagnostic category requiring prospective symptom confirmation. Despite skepticism about its existence, the verification process used to confirm this disorder is extremely rigorous and exceeds what is required for other diagnoses.

DEFINITION

The disorder is classified in DSM-IV as a mood disorder not otherwise specified. The clinical criteria for the disorder appear in an appendix. The initial five symptoms under Criteria A are listed according to the frequency with which they were endorsed by more than 600 premenstrual sufferers in five academic centers. At least one of the initial four psychological symptoms is required to confirm the diagnosis, thereby emphasizing the affective nature of the illness. Changes occurring in DSM-IV include the addition of the symptom, “a subjective sense of being overwhelmed or out of control,” and the stipulation of a symptom-free postmenstrual week. Although it is possible that a woman with premenstrual syndrome may have severe physical symptoms only, the diagnosis premenstrual dysphoric disorder can only be made if there are mood or other affective symptoms. A woman must experience symptoms during the luteal phase for the majority of menstrual cycles during the previous year. Finally, she must prospectively confirm symptoms with daily ratings over the course of two consecutive menstrual cycles.

HISTORY

Vivid descriptions and social stigma associated with menstrual-related mood and behavioral changes date back to Hippocrates, the Talmud, the Bible, and ancient African tribal culture. The feared contamination and dreaded power of the menstruating woman led to the custom of secluding a woman during her menstrual cycle in many of these ancient cultures. For centuries the mythical stigma associated with the menstruating woman has had an insidious influence on the development of a critical description and analysis of menstrualrelated disturbances even to the present day.

In 1873 Sir Henry Maudsley is credited with initiating a prospective view of the link between mental and somatic disturbance and the ovarian cycle. However, it was not until 1931 that Robert Frank published the first description of “premenstrual tension” and postulated a hormonal pathogenesis. In 1952 Katherina Dalton introduced the term *premenstrual syndrome*. Over the ensuing years, the condition was variably defined by different investigators. In an attempt to consolidate the definitions a consensus conference was held at the National Institute of Mental Health (NIMH) in 1983. This conference led to provisional research criteria for late luteal phase dysphoric disorder in the revised third edition of DSM (DSM-III-R). The criteria were included in an appendix because of controversy in the scientific as well as the public sectors. Proponents of its inclusion as a disorder maintained that it was a clinically significant condition that necessitated diagnostic criteria for clinical and research purposes; opponents feared the potential stigmatization and labeling of women. The DSM-IV task force recognized the need to re-evaluate this question and in 1988 appointed a DSM-IV Work Group to study these proposed diagnostic criteria. For more than 5 years this group of experts debated and reviewed the scientific data and concluded that evidence supported a link between late luteal phase dysphoric disorder and mood disorders. The rather cumbersome title of late luteal phase dysphoric disorder was replaced with *premenstrual dysphoric disorder* in DSM-IV. However, political groups continuing to fear the stigmatization of women persuaded the Board of Trustees to retain the criteria in an appendix.

COMPARATIVE NOSOLOGY

In the *Federal Register* premenstrual syndrome is defined as “a recurrent symptom complex which begins during the week prior to menstruation and disappears soon after the onset of the menstrual flow. This symptom complex consists predominantly of edema, lower abdominal pain (including cramps), breast tenderness, headache, abdominal bloating, fatigue, and feelings of depression, irritability, tension, and anxiety.” The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) includes a category for premenstrual tension, but no definition is associated with this diagnosis.

Other disorders that fall under the category of premenstrual syndrome yet remain distinct from premenstrual dysphoric disorder include various forms of premenstrual exacerbation of other psychiatric illness and severe physical premenstrual symptoms. The condition should be distinguished from dysmenorrhea, which is painful menstruation; genetic data show that the two conditions do not covary among twins.

EPIDEMIOLOGY

Two sources of data suggest that the prevalence rate for premenstrual dysphoric disorder is between 2 and 10 percent in the United States. First, retrospective community studies find rates for mild symptoms varying between 30 to 80 percent, while 2 to 10 percent of women experience severe or disabling symptoms. Second, smaller prospective community studies find rates of 3.4 and 4 percent.

Using retrospective reports, Epidemiological Catchment Area study (ECA) data show an equivalent rate between African-American women and white women. A study of 120 Nigerian women finds work impairment in 7.5 percent of women. Similarly, a Swedish study finds premenstrual complaints in 73 percent of women, although only 2 percent suffered from work absenteeism. Retrospective reports suggest an average age of onset in the middle to late 20s. This is distinct from the average age a woman seeks treatment, which is in the middle to late 30s.

Like patients suffering from other mood disorders, women with premenstrual dysphoric disorder tend to have a strong family history of major depressive disorder among first-degree relatives. One study found a history of major depressive disorder in one or more first-degree relatives among 65 percent of patients.

ETIOLOGY

The causes of premenstrual syndrome and premenstrual dysphoric disorder are unknown. A number of theories have been proposed but investigations have primarily included populations that experienced premenstrual syndrome and not premenstrual dysphoric disorder. It is possible that there are overlapping etiological factors between some forms of these two conditions.

The oldest theories postulate either an absolute or relative deficit of progesterone. Numerous older investigations have attempted to establish a difference in progesterone concentrations among symptomatic women versus controls and the results have been generally unsuccessful. Problems with these studies include the fact that progesterone is secreted in a pulsatile fashion and then metabolized in a matter of minutes. Thus, very frequent sampling or integrated plasma progesterone measures are required to more clearly define differences in the concentrations of this hormone. Similar problems beset studies attempting to find differences between estradiol concentrations or the progesterone-estradiol ratio in symptomatic versus control women. To date, it has not been clearly shown that either progesterone

concentrations or the ratio between progesterone and estradiol is aberrant or deficient in women suffering from either premenstrual syndrome or premenstrual dysphoric disorder. Similarly, several studies have failed to find differences in androgen concentrations between women who have premenstrual dysphoric disorder and controls, although one group finds higher free testosterone concentrations in the former set of women.

The timing of symptom onset, that is, during the middle to late luteal phase of the cycle certainly suggests a role for gonadal hormones. However, some work sheds doubt on pathophysiological attribution limited to ovarian hormones. In one study a progesterone antagonist administered during the midluteal phase of the cycle to women with premenstrual dysphoric disorder truncated the luteal phase, yet women continued to experience premenstrual symptoms. Further, casting doubt on the role of ovarian hormones are several studies that have shown that symptoms of premenstrual dysphoria improve in some women who have had a hysterectomy but no ovariectomy. An earlier study showed that sequential addition of estrogen and then progestin administered to postmenopausal women can provoke premenstrual-like symptoms. However, several subsequent reports attempting to replicate this found that when progesterone is administered in a double-blind placebo-controlled fashion, premenstrual symptoms are no more likely to occur with progesterone than placebo. On the other hand, in support of a pathophysiological role for ovarian hormones is work showing remission in women who are either (1) spontaneously anovulatory, (2) anovulatory due to suppression ovulation, or (3) anovulatory due to natural or surgical menopause.

Endogenous Opioid Peptides Some investigators suggest that premenstrual dysphoric disorder is triggered by a decline in endogenous opioid peptides such as b-endorphin and enkephalin. This is particularly intriguing since the symptoms of premenstrual dysphoric disorder are in some ways similar to opioid withdrawal. In addition, progesterone regulates the catabolism of enkephalin. Earlier work found that b-endorphin concentrations fluctuate during the menstrual cycle with peaks at ovulation and troughs during the luteal phase of the menstrual cycle. However, not all groups have replicated these findings and it is unclear how relevant peripheral b-endorphin concentrations are to processes occurring in the central nervous system (CNS). Using a different approach, several research groups have evaluated the pulsatile secretion of luteinizing hormone (LH), which reflects the secretion of gonadotropin-releasing hormone (GnRH) and is regulated by endogenous opioid peptides. Theoretically, if there was an abnormality in the level or rate of central endogenous opioid secretion, this may also be evident in the pulsatile secretion of LH. Several studies have found differences between premenstrual dysphoric disorder and control women in LH pulse frequency, but the cycle phase showing changes has differed among groups.

Neuroendocrine Markers Some researchers find higher rates of thyroid dysfunction in women with premenstrual syndrome or premenstrual dysphoric disorder. These inconsistence are perhaps explained by a report showing that women with the latter condition have greater variability in thyroid-stimulating hormone (TSH), triiodothryronine (T_3), uptake thyroxine (T_4), and the prolactin response to thyrotropin-releasing hormone. Transient alterations in hypothalamic-pituitary-adrenal axis function are also found in these patients. The luteal-phase response to corticotropin-releasing hormone is exaggerated in women with premenstrual symptoms versus controls. However, 24-hour urinary free cortisol excretion is not different between groups or cycle phase, suggesting that this change in responsiveness of the hypothalamic pituitary adrenal axis is brief and transient. This may explain why women with premenstrual dysphoric disorder continue to have normal menstrual cycles compared to patients with sustained hypercortisolism. Finally, in support of hypothalamic-pituitary-adrenal overactivity is the finding from a study that 62 percent of women with premenstrual dysphoric disorder failed to suppress cortisol after receiving dexamethasone (Decadron).

Catecholamines Some, but not all, studies find increased concentrations of the noradrenergic metabolite 3-methoxy-4 hydroxy phenyl-ethylene glycol in patients with premenstrual dysphoric disorder. Similarly, there is one report of a lowered homovanillic acid to 5-hydroxyindoleacetic acid ratio in women with this disorder. Early a- and b-adrenergic receptor binding studies largely conducted on women who retrospectively reported premenstrual syndrome find differences in either the number or affinity of these receptors in women with premenstrual syndrome versus controls. Several recent well-controlled studies on women who meet criteria for premenstrual dysphoric disorder find no differences between symptomatic women and controls in either a_2- or b-receptor binding.

Indolamines A large amount of work has investigated potential changes of serotonergic functioning in women with premenstrual dysphoric disorder. Whole blood concentrations of serotonin are lower in women with severe premenstrual syndrome compared to controls. However, because serotonin does not cross the blood-brain barrier this may not indicate CNS pathology. A number of groups find that luteal-phase platelet serotonin uptake is decreased in women with either premenstrual syndrome or premenstrual dysphoric disorder as compared with controls.

Administration of the challenge agent tryptophan produces a blunted growth hormone and cortisol response during both phases of the menstrual cycle in premenstrual dysphoric disorder women as compared with controls. The prolactin response to various challenge agents has been blunted only during the premenstrual phase, in the follicular phase, or not at all, depending on the agent used.

GABA Women with premenstrual dysphoric disorder have lower plasma concentrations of g-aminobatyric acid (GABA) during the luteal phase of the cycle than controls. However, women who had a previous history of major depressive disorder and a current diagnosis of premenstrual dysphoric disorder have lowered plasma GABA concentrations during both phases of the cycle, which illustrates that the state-trait differences in biology may also be related to other past illnesses.

Chronobiological Factors A chronobiological hypothesis of mood disorders has been explored in the context of melatonin, a hypothesized marker for the circadian oscillators. The offset of melatonin secretion and the integrated melatonin concentration are lower in women with premenstrual dysphoric disorder, which suggests an underlying chronobiological abnormality in these patients.

Other Biological Factors Several studies show that erythrocyte magnesium concentrations, a sensitive indicator of body stores of magnesium, are lower in premenstrual syndrome patients than controls. In addition, studies of zinc and copper concentrations during the luteal phase in these patients have been shown to be lower and higher, respectively, as compared to normal controls.

Psychological Factors Early psychoanalytic formulations of premenstrual distress proclaimed that premenstrual depression was the result of a woman's grief over her failure to conceive. Also included in these interpretations was a woman's repressed hostility towards and envy of the more favored males, with the monthly menstrual cycle serving as a brutal reminder of her femininity. These theories hold no validity because the symptoms of premenstrual dysphoric disorder begin to ameliorate with the onset of menses, contrary to the analytic theory.

Personality Factors Women with premenstrual dysphoric disorder have significantly higher scores for all three dimensions—Harm Avoidance, Reward Dependence, and Novelty Seeking—on the Tridimensional Personality Questionnaire (TPQ) compared to normative TPQ data. However, the Harm Avoidance scores are significantly lower in patients with premenstrual dysphoric disorder than those with major depressive disorder. A study evaluating the comorbidity of DSM personality disorders with premenstrual dysphoric disorder found that less than 10 percent of these women met the diagnostic criteria for a personality disorder on Structured Clinical Interview for DSM-III-R Axis II Personality Disorders (SCID-II).

Cognitive Style and Processing Some studies support a mild, phase-independent memory impairment and difficulty in learning new material in women with premenstrual dysphoric disorder as compared to controls. Enhanced performance on tests of frontal lobe function has been demonstrated in the follicular compared to the luteal phase in women with or without this disorder. Mood changes have not consistently accounted for these differences. Mood disorders are associated with a number of changes in cognitive processing, including selective attention to dysphoric stimuli. Paradigms evaluating whether women with premenstrual syndrome or premenstrual dysphoric disorder preferentially recall negative events or negative words are mixed.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of premenstrual dysphoric disorder is dependent upon certain core features: (1) a specific symptom constellation as outlined in DSM-IV, (2) the timing of symptom onset during the luteal phase of the cycle, (3) the absence of another underlying psychiatric or contributing general medical disorder, and (4) the severity of the disorder and its association with functional impairment. The symptoms of this disorder are primarily affective, with depressed mood constituting the most common symptom followed by anxiety or tension, mood swings, and anger or irritability; at least one of these four symptoms is required for the diagnosis. A minimum of five symptoms must be present and the other potential complaints, which are listed in Table 28.3-1, are somatic symptoms such as breast pain and bloating.

Table 28.3-1 DSM-IV Research Criteria for Premenstrual Dysphoric Disorder

DSM-IV does not stipulate at which point during the luteal phase symptoms must commence but it does include a criterion that the postmenstrual week be symptomfree. This was included to ensure that the constellation of symptoms experienced is not merely an exacerbation of an underlying disorder. DSM-IV criteria stipulate that luteal phase symptoms be prospectively confirmed during two consecutive cycles. This requirement arose from research showing that retrospective reports correlate poorly with the results of prospective ratings in many women. In one study of 82 women who sought treatment for clinically significant premenstrual symptoms, 14 percent did not show premenstrual mood changes, 56 percent had follicular phase symptoms, and only 30 percent of women demonstrated symptoms isolated to the luteal phase of the cycle. However, some data suggest that women with severe symptomatology are more likely to confirm prospective ratings. According to prospective ratings, the time of greatest well-being is just prior to ovulation, when estradiol is peaking. The most symptomatic interval is the few days immediately preceding menses, when both estradiol and progesterone are declining. There is variability in the duration of symptomatology, with some women experiencing a few days of premenstrual changes and other women having difficulty immediately after ovulation.

DIFFERENTIAL DIAGNOSIS

Countless symptoms have been associated with premenstrual syndrome, which increases the chances that another general medical disorder may be the problem. However, the symptom constellation for premenstrual dysphoric disorder is far more specific. The most significant disorders to rule out in the differential diagnosis are mild mood disorders, such as dysthymic disorder, or chronic anxiety disorders, such as generalized anxiety disorder. It may be that women who have exacerbation of chronic mild depression or generalized anxiety constitute one of the largest groups of women with premenstrual problems. Technically, these women may be considered to have premenstrual syndrome or premenstrual exacerbation of another disorder, but they do not have premenstrual dysphoric disorder. Daily ratings can be helpful in these cases but some women may report no dysphoric symptoms on a daily rating form and admit to chronic mild depression during a clinical interview. It is also important to rule out substance-related disorders. Substance use is not included in most daily rating forms and therefore requires independent assessment. In retrospective reports, women with intermittent depression or dysphoria, such as found in recurrent brief depressive disorder, may associate symptom expression only to the luteal phase of the cycle. A number of women experience dysphoria with the onset of menses. Retrospectively patients may attribute this to premenstrual syndrome, but daily ratings will show menstrual rather than premenstrual dysphoria.

Clinical studies commonly exclude women with a number of endocrine disorders such as Cushing's disease, hypothyroidism or hyperthyroidism, and endometriosis. It is unlikely that these disorders cause a syndrome with the symptoms of premenstrual dysphoric disorder, which has a time-limited symptom pattern. Nonetheless, these are conditions that may co-occur with premenstrual dysphoric disorder and could contribute to a woman's discomfort.

COURSE AND PROGNOSIS

A number of studies find that women with premenstrual dysphoric disorder commonly present for treatment at about age 36. Many women report that their symptoms began a decade earlier and have gradually worsened over time. This has been confirmed by several longitudinal studies but not others. Some work suggests that low parity and less frequent use of oral contraceptives are associated with more severe symptoms. Another study finds that a history of severe premenstrual syndrome in one's mother is a risk factor for severe premenstrual syndrome. Women with premenstrual dysphoric disorder are more likely than women without premenstrual difficulties to have had major depressive disorder. They are also at higher risk of eventually developing major depressive disorder, especially if they have symptoms that endure through the entire 2 weeks of the luteal phase. Thus, a consideration in the clinical course of premenstrual dysphoric disorder is not only the severity and cyclical recurrence of premenstrual symptoms, but the higher risk for other mood disorders.

If the age of onset is accurate, the cumulative time that symptoms are extant in women with premenstrual dysphoric disorder can be computed. For example, if symptoms begin at age 26 and continue monthly for approximately 7 to 10 days with every menstrual cycle until the average age of menopause (age 51), a woman with the disorder will experience approximately 1400 to 2800 symptomatic days (3 to 8 years).

TREATMENT

Substantial work has focused on the treatment of premenstrual syndrome. Difficulties with this body of literature include: (1) variable definitions, (2) a failure to stipulate that women keep prospective ratings prior to study entry, (3) a failure to exclude women with ongoing psychiatric illness, and (4) generally brief study periods, such as one cycle. Treatments that have been shown effective for premenstrual syndrome may not be palliative for premenstrual dysphoric disorder. For example, a treatment tested on women suffering from premenstrual bloating and breast pain may not be effective for the mood symptoms of premenstrual dysphoric disorder. Thus, one needs to be cautious when evaluating the literature on premenstrual syndrome and its applicability to women with premenstrual dysphoric disorder. Most treatment trials illustrate a high nonspecific (placebo) response rate, especially in studies where women had milder illness. Nonspecific response rates are somewhat lower with well-characterized premenstrual dysphoric disorder populations but remain substantial. Data suggest an average duration for nonspecific response of between one and three cycles.

Hormone Agonists and Antagonists The most widely studied hormone is progesterone, although a number of trials have employed synthetic progestins. To date there are 11 placebo-controlled randomized clinical trials using natural progesterone in dosages of about 200 to 400 mg per day during the luteal phase. In addition, there are four trials using dydrogesterone (the optical isomer of progesterone) and the synthetic progestin medroxyprogesterone (Provera). The results have been mostly negative, but most trials, including two large-scale studies of over 100 women with premenstrual dysphoric disorder show no greater benefit with a progestogen than placebo. However, milder symptoms and physical symptoms may improve more with progesterone or progestin treatment.

Estrogen has been evaluated in three randomized clinical trials and results are predominantly positive. Studies used daily ratings but none included clinical criteria for premenstrual dysphoric disorder. In one study of 68 women, a single-blind placebo run-in period identified a nonspecific response in 94 percent of women during the first two cycles.

Oral Contraceptives Most work evaluating the use of oral contraceptives for premenstrual symptoms used retrospective survey designs, with two exceptions. In the first, researchers compared a triphasic oral contraceptive with a monophasic one in 32 women. The triphasic preparation was superior to the monophasic one, but there was no placebo condition. The second study included 45 women and compared a triphasic oral contraceptive to placebo; this research group found that the oral contraceptive was superior for breast pain, tenderness, and bloating but not mood or irritability.

Danazol Danazol (Danocrine) is a synthetic partial androgen agonist that can induce anovulation. It has been used in dosages ranging between 100 mg a day and 400 mg a day. Although five out of the six danazol trials employed daily ratings, none stipulated criteria as rigorous as for premenstrual dysphoric disorder. Danazol was effective in the majority of these trials although the symptoms responding include breast pain and lethargy, and less often, mood symptoms. Unfortunately, the adverse affects associated with this medication, including the potential for hirsutism, limit its usefulness.

Gonadotropin-Releasing Hormone Agonists Like danazol, GnRH agonists induce an anovulatory state by downregulating pituitary gonadotrophs, thereby inducing a medical ovariectomy. Three of four double-blind placebo-controlled trials found that this class of medication is helpful for depressive symptoms occurring premenstrually. There are substantial risks associated with this treatment. It can only be continued for 6 months before progestin and estrogen are added again to

approximate a menstrual cycle. If hormone add-back therapy is not instituted, a woman runs a risk of bone loss and of increased lipid levels, with the subsequent risk of coronary artery disease and cerebral vascular disease.

Diuretics Several diuretics including spironolactone (Aldactone), ammonium chloride, and metaxalone (Skelaxin) have been used for the treatment of premenstrual syndrome. Spironolactone has been used most frequently, perhaps because of its other putative features, including its antiandrogenic properties as well as its capacity to treat mood swings. Only one clinical trial used daily ratings as entrance criteria and this study included women who met the criteria for premenstrual dysphoric disorder. Although the results from previous trials were mixed, this last study found that spironolactone (Aldactone) was helpful for somatic symptoms. This agent may be particularly beneficial when used in conjunction with psychiatric medications in premenstrual dysphoric disorder patients who have mood symptoms as well as the physical symptoms of headache and bloating.

Other Pharmacological Treatments Bromocriptine (Parlodel) is a dopamine agonist that binds to postsynaptic receptors and inhibits prolactin secretion. It has been used in randomized clinical trials, none of which stipulated criteria for premenstrual dysphoric disorder. The results are mixed, with the positive trials showing that bromocriptine is effective for breast pain. Medications whose use is supported by a single trial include naltrexone (ReVia), clonidine (Catapres), calcium (Tums), and doxycycline (Vibramycin). Again, the patients studied in this trial did not necessarily meet the criteria for premenstrual dysphoric disorder.

Mefenamic acid (Ponstel) and naproxen (Anaprox) have been used for the treatment of premenstrual syndrome. These agents appear to be helpful for pain associated with premenstrual symptoms.

Atenolol (Tenormin) blocks b-adrenergic receptors and suppresses melatonin; it has been evaluated in two randomized clinical trials. The first showed benefit for premenstrual irritability while the second found that atenolol was equal to placebo. Atenolol may be helpful as an adjunct treatment for women with premenstrual dysphoric disorder who have a partial response to another treatment.

Psychiatric Agents Given that premenstrual dysphoric disorder is characterized by affective symptoms such as low mood, mood swings, anxiety, and tension, a number of research groups have evaluated the efficacy of psychiatric drugs in this condition.

Lithium Given the periodic expression of the illness and the prominence of symptoms such as mood swings, several authors evaluated lithium (Eskalith) in the treatment of severe premenstrual syndrome. Unfortunately, placebo-controlled trials failed to find that it was superior to placebo.

Alprazolam Five placebo-controlled randomized controlled trials have used alprazolam (Xanax) to treat premenstrual dysphoric disorder. The dosages employed range between 0.75 mg a day to 4 mg a day during the luteal phase of the cycle over the course of two to four cycles. Four studies found alprazolam to be more effective than placebo. One study included women with premenstrual dysphoric disorder and premenstrual worsening of a mood disorder; alprazolam was not effective in this latter group. All studies showed a small effect size, which suggests that alprazolam may be more effective for women who have milder symptoms or as an adjunct to other treatments. Its benefit is that it can be given during the late luteal phase of the cycle, tapered when menstruation begins, and has an immediate onset of action.

Other Anxiolytic Agents One small preliminary study showed benefit for the serotonin (5-hydroxytryptamine [5-HT] subtype 1A) (5-HT $_{1A}$) partial agonist buspirone (BuSpar), when it was administered during the luteal phase of the cycle. The authors of this study believe that if this agent is given throughout the study it efficacy may be improved even further.

Antidepressant Agents Antidepressant drugs that have been evaluated in placebo-controlled randomized controlled trials include the clomipramine (Anafranil), fluoxetine (Prozac), fluvoxamine (Luvox), paroxetine (Paxil), sertraline (Zoloft), the serotonin agonist fenfluramine, the dual reuptake inhibitor venlafaxine (Effexor), bupropion (Wellbutrin), and maprotilin (Ludiomil). Nefazadone (Serzone) has only been evaluated in an open study and results from the venlafaxine multicenter trial are not yet available. The only placebo-controlled study on fluvoxamine failed to find superiority over placebo, but the methodology used in that study was problematic. The prototypical selective serotonin reuptake inhibitor (SSRI), fluoxetine, has been used in 6 small studies and one large multicenter study; the latter included over 200 women and compared a 20-mg-a-day and a 60-mg-a-day dosage to placebo. Self-ratings improved within the first cycle but attrition (11 percent) was high in the 60-mg-a-day group. In this trial treatment was administered throughout the cycle although the authors state that they continued a subgroup after the study on luteal phase treatment and that treatment response was maintained. A recent placebo-controlled study shows that fluoxetine is more effective than bupropion for premenstrual dysphoric disorder.

Paroxetine was shown to be effective for premenstrual dysphoric disorder in a small placebo-controlled study and one open trial. The placebo-controlled study is notable because paroxetine was compared to the noradrenergic antidepressant agent maprotiline. Although both agents were superior to placebo, paroxetine was significantly more effective. Similar findings were noted in a placebo-controlled study that compared bupropion to fluoxetine—fluoxetine was found to be more effective than bupropion even though bupropion was slightly more effective than placebo.

Sertraline has been evaluated in several small studies and a large multicenter study of more than 220 women. This trial, which was over three cycles, employed a flexible dosing pattern ranging between 50 mg a day and 150 mg a day. Sertraline was found to be more effective than placebo for nearly all symptoms but particularly for mood symptoms. In this multicenter trial, medication was prescribed continuously. Recent advances have been made in minimizing medication exposure to women with premenstrual dysphoric disorder who by definition are symptomatic during only the luteal phase of the cycle. The serotonergic tricyclic antidepressant agent clomipramine was effective in luteal-phase dosing. Sertraline also appears to be effective when administered only during the premenstrual phase of the cycle.

Dietary Modifications, Vitamin Supplements, Exercise, and Sleep Some researchers have suggested that pyridoxine (vitamin B_6) may be helpful for symptoms of premenstrual syndrome because it serves as a cofactor in a number synthetic reactions including those involved in the conversion of the amino acid tryptophan to serotonin. Of the seven randomized controlled trials using vitamin B_6, only one used daily ratings. The results are mixed with most showing vitamin B_6 equivalent to placebo. Similarly, studies using supplements such as evening primrose oil and optivite did not use rigorous entrance criteria and the results are variable.

Numerous uncontrolled reports, as well as the popular media, recommend dietary changes including elimination of refined sugar, salt, alcohol, and caffeine; unfortunately, these recommendations are not supported by data. In a double-blind, crossover study a specially formulated carbohydrate-rich beverage improved ratings of depression, anger, concentration, and memory compared to the isocaloric (placebo) beverage. The carbohydrate beverage may work by increasing the serum ratio of tryptophan to other large neutral amino acids. Studies of varying degree of scientific rigor show the positive effects of regular, moderate exercise on mood states and menstrual cycle symptoms. Most studies have been done on menstruating women who have PM premenstrual syndrome but not premenstrual dysphoric disorder. Sleep deprivation is palliative in major depressive disorder. Research shows that both early and late luteal phase sleep deprivation diminish symptoms in premenstrual dysphoric disorder.

Cognitive Psychotherapy, Education, and Support A number of cognitive-behavioral or relaxation therapies may be beneficial to the patient with premenstrual dysphoric disorders; however, further prospective placebo-controlled trials are needed. In one study both cognitive behavioral therapy and the comparison condition (called *information-focused therapy*, aimed at relaxation training, education, and nutrition) ameliorated the symptoms of anxiety, depression, physical changes, and negative thoughts in women with premenstrual dysphoric disorder. Another study found no difference in response to behavioral therapy or progesterone treatment in patients with prospectively diagnosed premenstrual dysphoric disorder. A third study found dydrogesterone treatment equivalent but not as longlasting as cognitive-behavioral therapy.

SUGGESTED CROSS-REFERENCES

Postpartum psychiatric disorders are discussed in Section 13.4, normal sexuality in Section 19.1, and chronobiology in Section 1.13. Monoamine neurotransmitters are discussed in Section 1.14 and amino acid neurotransmitters in Section 1.5. The biochemical aspects of mood disorders are discussed in Section 14.4. Chapter 31 presents in-depth discussions of many biological therapies.

SECTION REFERENCES

Bancroft J: The premenstrual syndrome—a reappraisal of the concept and the evidence. Psychol Med *24:*1, 1993.

Bancroft J, Cook A, Davidson D, Bennie J, Goodwin G: Blunting of neuroendocrine responses to infusion of L-tryptophan in women with perimenstrual mood change. Psychol Med *21:*305, 1991.

Berger CP, Presser B: Alprazolam in the treatment of two subsamples of patients with late luteal phase dysphoric disorder: A double-blind, placebo-controlled crossover study. Obstet Gynecol *84:*379, 1994.

Chuong CJ, Coulam CB, Kao PC, Bergstralh EJ, Go VLW: Neuropeptide levels in premenstrual syndrome. Fertil Steril *44:*760, 1985.

Endicott J, Halbreich U, Schact S, Nee J: Premenstrual changes and affective disorders. Psychosom Med *43:*519, 1981.

Eriksson E, Hedberg MA, Andersch B, Sundblad C: The serotonin reuptake inhibitor paroxetine is superior to the noradrenaline reuptake inhibitor maprotiline in the treatment of premenstrual syndrome. Neuropsychopharmacology *12:*169, 1995.

Eriksson E, Sundblad C, Lisjo P, Modigh K, Andersch B: Serum levels of androgens are higher in women with premenstrual irritability and dysphoria than in controls. Psychoneuroendocrinology *17:*195, 1992.

Freeman E, Sondheimer S, Rickels K, Polansky M: Ineffectiveness of progesterone suppository treatment for premenstrual syndrome. JAMA *264:*349, 1990.

Freeman EW, Weinstock L, Rickels K, Sondheimer SJ, Coutifaris C: A placebo-controlled study of effects of oral progesterone on performance and mood. Br J Clin Pharmacol *33:*293, 1992.

*Gehlert S, Chang CH, Harlage S: Symptom patterns of premenstrual dysphoric disorder as defined in the *Diagnostic and Statistical Manual of Mental Disorder-IV*. J Womens Health *8:*75, 1999.

Girdler S, Pedersen CA, Straneva PA, Leserman J, Stanwyck CL, Benjamin S, Light KC: Dysregulation of cardiovascular and neuroendocrine responses to stress in premenstrual dysphoric disorder. Psychiatry Res *81:*163, 1998.

Graham CA, Sherwin BB: A prospective treatment study of premenstrual symptoms using a triphasic oral contraceptive. J Psychosom Res *36:*257, 1992.

*Gurguis GN, Yonkers KA, Phan SP, Blakeley JE, Williams A, Rush AJ: Adrenergic receptors in premenstrual dysphoric disorder: I. Platelet alpha 2 receptors: Gi protein coupling, phase of menstrual cycle, and prediction of luteal phase symptom severity, Biol Psychiatry *44:*600, 1998.

Gurguis GN, Yonkers KA, Blakeley JE, Phan SP, Williams A, Rush AJ: Adrenergic receptors in premenstrual dysphoric disoder. II. Neutrophil beta2-adrenergic receptors: Gs protein coupling, phase of menstrual cycle and prediction of luteal phase symptom severity. Psychiatry Res *79:*31, 1998.

Halbreich U, Endicott J, Nee J: Premenstrual depressive changes: Value of differentiation. Arch Gen Psychiatry *40:*535, 1983.

Halbreich U, Tworek H: Altered serotonergic activity in women with dysphoric premenstrual syndromes. Int J Psychiatry Med *23:*1, 1993.

Halbreich U, Piletz JE, Carson S, Halaris A, Rojansky N: Increased imidazoline and a_2 adrenergic binding in platelets of women with dysphoric premenstrual syndromes. Biol Psychiatry *34:*676, 1993.

Halbreich U, Smoller JW: Intermittent luteal phase sertraline treatment of dysphoric premenstrual syndrome. J Clin Psychiatry *58:*1, 1997.

Hurt SW, Schnurr PP, Severino SK, Freeman EW, Gise LH, Rivera-Tovar A, Steege JF: Late luteal phase dysphoric disorder in 670 women evaluated for premenstrual complaints. Am J Psychiatry *149:*525, 1992.

Johnson SR: The epidemiology and social impact of premenstrual symptoms. Clin Obstet Gynecol *30:*369, 1987.

Kendler KS, Silberg JL, Neale MC, Kessler RC, Health AC, Eaves LJ: Genetic and environmental factors in the etiology of menstrual, premenstrual and neurotic symptoms: A population-based twin study. Psychol Med *22:*85, 1992.

*Le Melledo JM, Merani S, Koszycki D, Bellavance F, Palmour R, Gutkowska J, Steinberg S, Bichet DG, Bradwejn J: Sensitivity to CCK-4 in women with and without premenstrual dysphoric disorder (PMDD) during their follicular and luteal phases. Neuropsychopharmacology *20:*81, 1999.

Metcalf MB, Livesey JH, Wells JE, Braiden V, Hudson SM, Bamber L: Premenstrual syndrome in hysterectomized women: Mood and physical symptom cyclicity. J Psychosom Res *35:*555, 1991.

Metcalf MG, Braiden V, Livesey JH: Symptom cyclicity in women with the premenstrual syndrome: An 8-year follow-up study. J Psychosom Res *36:*237, 1992.

Mezrow G, Soupe D, Spicer D, Lobo R, Leung B, Pike M: Depot leuprolide acetate with estrogen and progestin add-back for long-term treatment of premenstrual syndrome. Fertil Steril *62:*932, 1994.

*Parry BL, Mostofi N, LeVeau B, Nahum HC, Golshan S, Laughlin GA, Gillin JC: Sleep EEG studies during early and late partial sleep deprivation in premenstrual dysphoric disorder and normal control subjects. Psychiatry Res *85:*127, 1999.

Pearlstein TB, Frank E, Rivera-Tovar A, Thoft JS, Jacobs E, Mieczkowski TA: Prevalence of Axis I and Axis II disorders in women with late luteal phase dysphoric disorder. J Affect Disord *20:*129, 1990.

Pearlstein TB, Stone AB, Lund SA, Scheft H, Zlotnick C, Brown WA: Comparison of fluoxetine, bupropion, and placebo in the treatment of premenstrual dysphoric disorder. J Clin Psychopharmacol *17:*261, 1997.

Perna G, Brambilla F, Arancio C, Bellodi L: Menstrual cycle-related sensitivity to 35% CO_2 in panic patients. Biol Psychiatry *37:*528, 1995.

Resnick A, Perry W, Parry B, Mostofi N, Udell C: Neuropsychological performance across the menstrual cycle in women with and without premenstrual dysphoric disorder. Psychiatry Res *77:*147, 1998.

Rivera-Tovar AD, Frank E: Late luteal phase dysphoric disorder in young women. Am J Psychiatry *147:*1634, 1990.

Roy-Byrne PP, Rubinow DR, Hoban MC, Grover GN, Blank D: TSH and prolactin responses to TRH in patients with premenstrual syndrome. Am J Psychiatry *144*(4):480, 1987.

Rubinow DR, Roy-Byrne P: Premenstrual syndromes: Overview from a methodologic perspective. Am J Psychiatry *141:*163, 1984.

Severino SK, Moline ML: *Premenstrual Syndrome: A Clinician's Guide*. Guilford, New York, 1989.

*Steinberg S, Annable L, Young SN, Liyange N: A placebo-controlled clinical trial of l-tryptophan in premenstrual dysphoria. Biol Psychiatry *45:*313, 1999.

Steiner M, Steinberg S, Stewart D, Carter D, Berger C, Reid R, Grover D, Streiner D: Fluoxetine in the treatment of premenstrual dysphoria. N Engl J Med *332:*1529, 1995.

Steiner M, Korzekwa M, Lamont J, Wilkins A: Intermittent fluoxetine dosing in the treatment of women with premenstrual dysphoria. Psychopharmacol Bull 1997.

Sundblad C, Hedberg MA, Eriksson E: Clomipramine administered during the luteal phase reduces the symptoms of premenstrual syndrome: A placebo-controlled trial. Neuropsychopharmacology *9:*133, 1993.

Warner P, Bancroft J: Factors related to self-reporting of the premenstrual syndrome. Br J Psychiatry *157:*249, 1990.

Yatham LN, Barry S, Dinan TG: Serotonin receptors, buspirone, and premenstrual syndrome. Lancet *6:*1447, 1989.

Yonkers KA: Treatment of premenstrual dysphoric disorder. Curr Rev Mood Anxiety Disord *1:*215, 1997.

*Yonkers KA: Assessing unipolar mood disorders in women. Psychopharmacol Bull *34:*261, 1998.

Yonkers KA, Halbreich U, Freeman E, Brown C, Endicott J, Frank E, Parry B, Pearlstein T, Severino S, Stout A, Stone A, Harrison W: Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment. A randomized controlled trial. Sertraline Premenstrual Dysphoric Collaborative Study Group. JAMA *278:*983, 1997.

Yonkers KA, Halbreich U, Freeman E, Brown C, Endicott J, Frank E, Parry B, Pearlstein T, Severino S, Stout A, Stone A, Harrison W: The Sertraline Premenstrual Dysphoric Collaborative Study Group: Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment. JAMA *278:*983, 1997.

Young SA, Hurt PH, Benedek DM, Howard RS: Treatment of premenstrual dysphoric disorder with sertraline during the luteal phase: A randomized, double-blind, placebo-controlled crossover trial. J Clin Psychiatry *59:*76, 1998.

Textbook of Psychiatry

28.4 GENETIC COUNSELING

KATE A. BERG, PH.D., AND DARRELL G. KIRCH, M.D.

New discoveries regarding the genetic basis of medical disorders have been appearing at an accelerating rate, creating broad public awareness that there are genetic determinants in a wide range of diseases. For the mental disorders, extensive family studies of heritability and recent reports of specific genetic linkage for Alzheimer's disease, schizophrenia, and bipolar disorder increasingly are leading patients and family members to request genetic counseling, directly or indirectly. Often their questions are posed to a psychiatrist. The following review of the terminology (Table 28.4-1) and key concepts in genetic counseling provides the psychiatric clinician with an approach to these questions regarding the inheritance of mental disorders.

Consultand	Person seeking genetic counseling
Genotype	Alleles at a specific locus
Informant	Person who gives information to make pedigree
Multifactorial	Traits resulting from the effect of multiple genes and environmental factors
Phenotype	Observed result of genotype and environmental factors
Polygenic	Trait produced by the action of many genes of small effect
Proband	Affected person through whom pedigree is investigated
Recurrence risk	Probability that a genetic disorder in one family member will occur in another member of the same or future generations

Table 28.4-1 Definitions of Terms

The first step in genetic counseling is to establish a diagnosis using a criteria-based system. Recurrence risk data are currently based on empirical information and may not be accurate regarding individual cases. However, when the data are used, they should be expressed in a manner that the consultand can comprehend, with minimum bias on the part of the physician toward a particular interpretation. Part of the adjustment to the information may lie in following up the family as new developments unfold. Persons who came to the session with unrealistically high estimates of risk may actually be relieved to learn that their risk was smaller than expected. When members of the family who are at high risk are discovered, intervention may even assist in altering the long-term course of the illness. The psychiatrist can help put the information in the context of the person's life and the family's needs.

Genes are not necessarily destiny. Even when the genetic contribution to a mental disorder is well defined and identifiable, the data suggest that modification of the phenotype occurs through as yet unidentified environmental factors. Undertaking the arduous scientific task of separating genetic and environmental components eventually should result not only in more effective treatment, but also in the possibility of prevention of some forms of mental illness.

HISTORY

Attitudes toward giving medical information on genetics have changed dramatically since the early part of the century. Early attempts at genetic counseling were largely prescriptive and often were implemented in the context of the eugenics movement; the current consensus is that genetic counseling should be nondirective. In the 1970s and early 1980s the primary focus of genetic counseling was the simple provision of medical facts. Successful counseling was frequently measured by the ability of the consultand to remember recurrence risk information accurately or to take a particular line of action in response to the information.

During the 1990s the emphasis has shifted to the impact that the genetic information may have on the individual and the family. Outcome is less likely to be measured by the accurate recitation of facts and is more likely to center on individual adjustment and decision making.

GENETIC FACTORS IN MENTAL DISORDERS

The evidence for a genetic contribution to the major mental disorders was initially derived from family analyses, including studies of twins and adoptees. The findings of these studies are summarized in Table 28.4-2, which shows that the recurrence risk for these disorders increases as the degree of genetic relatedness increases. The recent use of molecular genetic techniques to study families with multiple affected members has yielded a number of reports of linkage to possible chromosomal locations in different families with schizophrenia and bipolar disorder. The fact that more than one genetic locus has been identified for each of these disorders indicates that they are likely to be heterogeneous illnesses involving multiple susceptibility genes interacting with various environmental factors. Turning to substance abuse, family studies have long indicated that some forms of alcohol dependence have a genetic component, but molecular studies have failed to conclusively prove gene linkage.

	Schizophrenia[a]	Bipolar Disorder[b]	Depressive Disorder[c]	Panic Disorder[d]	Compulsive Disorder[e]
Lifetime prevalence[f]	0.8–1.9	0.4–1.1	1.7–4.7	1.4–1.5	1.9–3.0
Risk to first-degree relative		4.0–9.0	1.9–10.4	15.0–24.7	25
Both parents affected	46				
One parent affected	13				
Full sibling	9				
Risk to niece or nephew	4				
Concordance for monozygotic twins	48	67	34		

[a] Data from Gottesman II, Shields J: Schizophrenia: The Epigenetic Puzzle. Cambridge University Press, New York, 1982.
[b] Data from Baron M: Genetics of manic depressive illness: Current status and evolving concepts. In Grove, Brain, and Behavior. PR McHugh, VA Andreasen, editors. Raven, New York, 1991.
[c] Data from McGuffin P, Katz R: The genetics of depression and manic depression. Br J Psychiatry 155:294, 1989.
[d] Data from Crowe RR, Noyes R, Pauls DL, Slymen D: A family study of panic disorder. Arch Gen Psychiatry 40:1065, 1983.
[e] Data from Swedo SE, Rapoport JL, Leonard H, Lenane M, Cheslow D: Obsessive-compulsive disorder in children and adolescents. Arch Gen Psychiatry 46:335, 1989.
[f] Data from Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD, Regier DA: Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psychiatry 41:949, 1984.

Table 28.4-2 Recurrence Risk of Familial Clustering for Psychiatric Disorders (in Percentage)

Conclusive evidence does exist for a genetic contribution to other disorders listed in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). The gene responsible for Huntington's disease, an autosomal dominant disorder, and its accompanying dementia, has been localized to chromosome 4, and predictive testing is now possible for those at risk. Alzheimer's disease appears to be genetically heterogeneous. Relatively rare familial early-onset forms of Alzheimer's disease have been linked to either chromosome 14 or 21. More-recent studies have shown an association between some late-onset cases of Alzheimer's

disease and a specific allele of apolipoprotein E (ApoE) on chromosome 19. The potential clinical role of ApoE genetic screening is currently under close scrutiny.

BASIC STEPS IN GENETIC COUNSELING

The key steps in genetic counseling are outlined in Table 28.4-3. Genetic counseling is undertaken to establish a diagnosis, define the appropriate questions, provide information concerning the recurrence risk in an understandable form, give a framework for decision making concerning possible courses of action, and help the consultand make the best possible adjustment to the situation.

▸ Establish diagnosis and obtain accurate family history and pedigree
▸ Address issues and concerns raised by the consultand and by information gained through the genetic counseling session
▸ Assist with decision making and adjustment to information
▸ Provide adequate follow-up

Table 28.4-3 Steps in Genetic Counseling

In general, genetic counseling for psychiatric disorders is sought by the patients themselves, their families, their prospective spouses, and prospective adoptive parents. Geneticists and genetic counselors tend to refer consultands with questions about psychiatric disorders to psychiatrists because the factors influencing susceptibility to those disorders are complex, and there is a recognition that the presence of severe psychiatric disorders usually has serious consequences for the functioning of both the affected individual and the family. The psychiatrist may then be called upon to provide genetic information in addition to addressing the therapeutic issues that may arise. Levels of involvement by the psychiatrist may range from diagnosis and the provision of recurrence risk data to treating the primary psychiatric disorder.

As genetic knowledge expands and clinical use of genetic screening to assess individual vulnerability increases, psychiatrists may also find themselves increasingly called upon to provide assessment and treatment to consultands with nonpsychiatric disorders who are referred by other physicians or genetic counselors. Such individuals may require considerable psychological support in coping with the genetic information they have learned and in making major decisions relating to that information.

Step 1: Establishing Diagnosis and Obtaining Family History The diagnosis of the proband must be rigorously established according to recognized criteria, such as DSM-IV, before counseling takes place. Medical records should be acquired, and releases obtained to call or write treating physicians and other therapists. Since other conditions can mimic mental disorders, a thorough medical history and information about substance abuse are important. The realities of different situations will create great variability in the amount of reliable clinical information that is available.

When an accurate diagnosis has been established, the family history should be taken from the informant in the form of a pedigree using the symbols shown in Figure 28.4-1. It is customary to work up through the pedigree starting with the information given by the consultand and following the guidelines in Table 28.4-4.

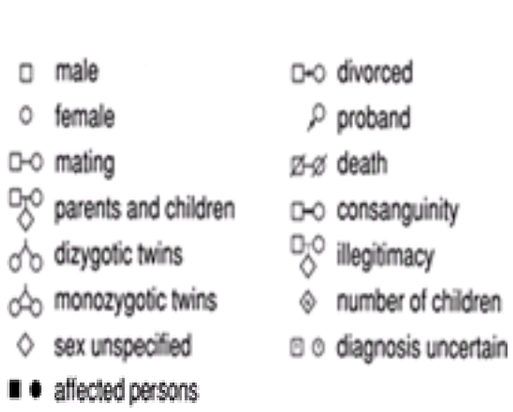

FIGURE 28.4-1 Symbols used in pedigree construction.

The informant should be asked the following questions about each person entered into the pedigree: What is the person's relationship to the informant (e.g., sibling, parent)? Is the person's relationship to the informant biological? Does the person have birth defects or a chronic disease? What is the ethnic background of the family? The following should also be determined: Marital status Maiden and married names of female relatives Occupation or educational status Address and telephone number (with assurance of confidentiality) The questions are relatively unintrusive and can, if necessary, be followed with more-sensitive questions (e.g., issues of paternity and consanguinity) if anything in the information given suggests that those topics should be pursued.

Table 28.4-4 Relevant Questions to Ask When Taking a Pedigree

If a relative is reported to be affected, the most satisfactory method of establishing the diagnosis is by direct interview or an examination of medical records. If that is not possible, as is often the case, an alternative may be to use the Family History–Research Diagnostic Criteria or similar instruments that attempt to establish a diagnosis through criteria-based questions.

Step 2: Provision of Information To provide information to the consultand, the psychiatrist should be familiar with the genetic concepts of complex disease and should determine (1) what information the consultand wants; (2) what is pertinent, but is not being asked; (3) at what pace and technical level the information should be given; and (4) what effect the information will have on the relationships within the consultand's family, since genetic information frequently has implications for the entire family.

Discussions of the genetic aspects of psychiatric disorders and the provision of recurrence risk information are complicated, not only because of diagnostic uncertainties, apparent etiologic heterogeneity, and comorbidity, but also because these disorders do not appear to involve straightforward mendelian genetics and presumably also have significant environmental components. In general, the multifactorial models used to describe complex inheritance of disorders like schizophrenia and bipolar disease are often highly technical and difficult for a layperson to conceptualize. Both the source and the magnitude of risk related to

recurrence data used in counseling need to be put in this perspective. It should be explicitly stated that the data apply to populations, not to individual persons, and thus are estimates at best.

In case A (Fig. 28.4-2) the unaffected husband (not the wife with a bipolar disorder) asked about the recurrence risk. The unaffected spouse of a person with bipolar disorder is often more concerned about the recurrence risk than is the affected person. The husband here might have been worried about the risk to his offspring, or he might actually be hoping to use information from genetic counseling to make decisions about his marriage to a person with a chronic illness or to deal with personal concerns about guilt and blame. Even for a disease whose genetic component is well understood, denial and a lack of understanding of true risk in carriers are common. In psychiatric disorders, dealing with the issue of risk in the context of other concerns may be exceptionally challenging.

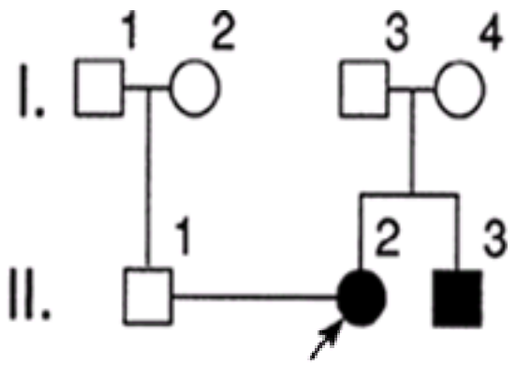

FIGURE 28.4-2 Case A. A couple in their mid-30s sought advice from the wife's psychiatrist on whether to have children. The wife (II-2), with a 10-year history of bipolar disorder and with bipolar disorder diagnosed in her brother (II-3), had had her symptoms controlled by the use of lithium carbonate. Previous attempts to function well without lithium had been unsuccessful. The couple expressed concerns about the effect of lithium on the pregnancy. In addition, the husband (II-1) wanted to know whether bipolar disorder is hereditary. (The recurrence risk of bipolar disorder is estimated to be up to 9 percent. Also, lithium has possible teratogenic effects, and the wife was entering the age range associated with increased chromosomal anomalies.)

The couple did not ask the physician about maternal age–related risk of chromosomal abnormalities in the fetus. It is standard practice to inform women who are 35 or older about the increased age-related risk of chromosomal anomalies and to discuss the availability of prenatal diagnosis. Although such discussion may fall outside the purview of psychiatric practice, this relevant clinical information should not be overlooked, even if it is handled by referral to a genetics clinic or to an obstetrician who specializes in prenatal diagnosis.

In case B (Fig. 28.4-3) the consultands were prospective adoptive parents. Their long history of infertility made them ready to look for mitigating information about recurrence risk. They also had had no direct experience with schizophrenia and had initially believed the recurrence risk to be even higher than it actually is for the offspring of either one or two affected parents.

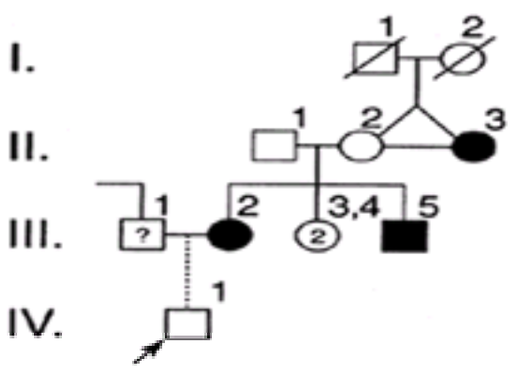

FIGURE 28.4-3 Case B. A childless couple in their late 30s, with a long history of infertility, were contemplating adoption of a newborn baby (IV-1). The baby was being placed for adoption because his mother (III-2) suffered from schizophrenia and was unable to care for him. The social worker in the case had comprehensive medical records for the biological mother and her family, including a brother (III-5) and a maternal aunt (II-3) with schizophrenia, but only hearsay reports (from the mother) that the putative father (III-1) also had schizophrenia. The adopting couple was concerned that the baby might be at increased risk for schizophrenia later in life. (The recurrence risk in the offspring of one affected parent is approximately 13 percent; for two affected parents the risk increases to 46 percent compared with a lifetime prevalence of 1 percent in the general population.)

Step 3: Decision Making and Adjustment to Information Most geneticists who perform genetic counseling take a nondirective approach to decision making. Such an approach should not be confused, however, with a lack of participation in comparing the risks and benefits of one course of action with those of others.

Each individual will view the same statistical risk from a different personal perspective. To aid the consultand, the data about statistical risk should be phrased in different ways. In case B, the risk could be described as “about 13 percent” or “one chance in eight of having schizophrenia,” and then also noted to be “an 87 percent chance of not developing schizophrenia.” The counseling clinician must take care to avoid offering personal judgments regarding the risk. Recent studies of genetic counseling and reproductive planning indicate that consultands are less focused on the absolute magnitude of the risk than on their personal desire to have children and their subjective interpretation of whether the risk is high or low.

After presenting the risk information, help in decision making can be given by exploring the consultand's feelings and reaction to the information; discussing any special cultural, ethnic, or religious factors that may be important; and using examples of options chosen by other families in similar situations. Inevitably physicians are asked what they would do in the same situation. Psychiatrists must avoid the temptation to say what they would do; answering the questions personally does not help patients think through the solution to a problem only they can answer.

Most geneticists gauge the effectiveness of counseling by their ability to help consultands meet certain goals. Primary among those goals are a diminution in any anxiety and guilt the consultands had experienced, an increase in their ability to understand the options available, and the formulation of effective plans to cope with the situation.

Step 4: Adequate Follow-Up An essential first follow-up step is a letter to the patient that recounts salient clinical issues, documents what the patient asked, recapitulates pedigree information the patient gave in narrative form, and outlines the plans, decisions, and any outstanding issues that remained unresolved at the end of the session. The letter is important not only to confirm what transpired during genetic counseling, but also for its use if the patient decides to seek a second opinion or further genetic counseling in the future. With written permission from the patient, a summary letter that includes the pedigree should also be sent to other physicians involved.

During the counseling session, the history of the patient may indicate other medical or psychiatric problems that need to be pursued, either with the consultand or with other family members. These situations are handled relatively simply through appropriate referral.

ETHICAL, LEGAL, AND SOCIAL CONSIDERATIONS

While psychiatrists always have dealt with the special considerations involved in receiving and communicating sensitive information, the new issues being raised by

advancing genetic knowledge are presenting major challenges. For example, an ethical and legal dilemma may arise when the physician receives information about untreated relatives that suggests that they are in need of medical intervention or at high risk for suicide. The consultand may not want other members of the family to know genetic information is being sought and may refuse permission for them to be contacted. This problem sometimes is compounded by situations such as that illustrated in case C (Fig. 28.4-4). Here, any information given to the consultand is also valid for the consultand's identical twin. While calling for careful attention to confidentiality, such a situation also clearly presents a need to encourage family discussion.

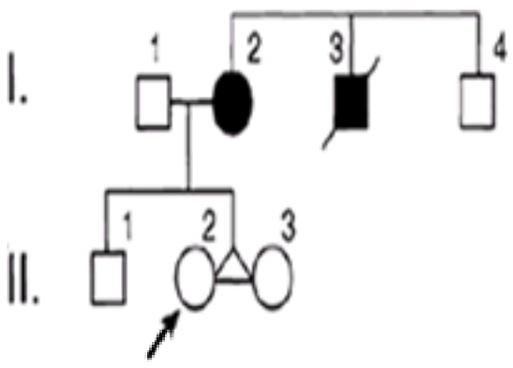

FIGURE 28.4-4 Case C. The consultand is interested in learning more about her own risk for Alzheimer's disease, a disease that affects her mother and maternal uncle. Her identical twin sister, however, vehemently denies that genetic factors could play a role in her relatives' illness. The dilemma for the physician is that when recurrence risk information is given to the consultand who is genetically identical to her sister, that information may also be conveyed to someone who has clearly expressed her wish not to know it.

As vulnerability genes for mental disorders are identified, policies and procedures for the appropriate protection of this information in relation to the decisions of insurers, employers, and similar entities are being developed by the medical profession, legislators, and others. As this occurs, psychiatrists who work with affected individuals and their families should take care always to adhere to the most rigorous ethical and legal standards of confidentiality regarding genetic information.

SUGGESTED CROSS-REFERENCES

The basic concepts of population genetics are reviewed in Section 1.17, and studies of genetic linkage of the psychiatric disorders are presented in Section 1.18. The genetics of specific disorders are discussed as follows: schizophrenia in Section 12.5, mood disorders in Section 14.3, and dementias in Chapter 10.

SECTION REFERENCES

Ashkenas J: Book Review of Nelkin D and Lindee MS: *The DNA mystique: the gene as a cultural icon*. Am J Hum Genet *60:*248, 1997.

*Baker DL, Schuette JL, Uhlmann WR, editors: *A Guide to Genetic Counseling*. Wiley-Liss, New York, 1998.

Baron M: Genetics of manic depressive illness: Current status and evolving concepts. In *Genes, Brain, and Behavior*, PR McHugh, VA McKusick, editors. Raven, New York, 1991.

Bernhardt BA: Empirical evidence that genetic counseling is directive: Where do we go from here. Am J Hum Genet *60:*17, 1997.

*Biesecker BB: Future directions in genetic counseling: Practical and ethical considerations. Kennedy Inst Ethics J *8:*145, 1998.

Bulik CM, Sullivan PF, Kendler KS: Heritability of binge-eating and broadly defined bulimia nervosa. Biol Psychiatry *44:*1210, 1998.

Cloninger CR, Kaufmann CA, Faraone SV, et al: Genome-wide search for schizophrenia susceptibility loci: The NIMH Genetics Initiative and Millenium Consortium. Am J Med Genet (Neuropsychiatric Genetics) *81:*275, 1998.

Collins FS: Sequencing the human genome. Hosp Pract *32:*35, 1997.

*Council on Ethics and Judicial Affairs: Multiplex genetic testing. Hastings Cent Rep *28:*15, 1998.

*Craddock N, Owen MJ: Modern molecular genetic approaches to psychiatric disease. Br Med Bull *52:*434, 1996.

Crowe RR, Noyes R, Pauls DL, Slyman D: A family study of panic disorder. Arch Gen Psychiatry *40:*1065, 1983.

Fletcher JC, Berg K, Tranoy KE: Ethical aspects of medical genetics. Clin Genet *27:*199, 1985.

Freedman R, Coon H, Myles-Worsley M, Orr-Urtreger A, Olincy A, Davis A, Polymeropoulos M, Holik J, Hopkins J, Hoff M, Rosenthal J, Waldo MC, Reimherr F, Wender P, Yaw J, Young DA, Breese CR, Adams C, Patterson D, Adler LE, Kruglyak L, Leonard S, Byerly W: Linkage of a neurophysiological deficit in schizophrenia to a chromosome 15 locus. Proc Natl Acad Sci USA *94:*587, 1997.

Frets PG, Niermeijer MF: Reproductive planning after genetic counseling: A perspective from the last decade. Clin Genet *38:*295, 1990.

*Gottesman II, Shields J: *Schizophrenia: The Epigenetic Puzzle*. Cambridge University Press, New York, 1982.

Hong Y, Pedersen NL, Brismar K, Faire U: Genetic and environmental architecture of the features of the insulin-resistance syndrome. Am J Hum Genet *60:*143, 1997.

*Kendler KS, Prescott CA: A population-based twin study of lifetime major depression in men and women. Arch Gen Psychiatry *56:*39, 1999.

Kessler S: *Genetic Counseling: Psychological Dimensions*. Academic Press, New York, 1979.

Levy PA, Dalton R, Shapira E: The awareness of chromosomal disarrangements in the evaluation of child psychiatry patients. Child Psychiatry Hum Dev *25:*281, 1995.

Mayeux R, Schupf N: Apolipoprotein E and Alzheimer's disease: The implications of progress in molecular medicine. Am J Public Health *85:*1280, 1995.

McGuffin P, Katz R: The genetics of depression and manic-depression. Br J Psychiatry *155:*294, 1989.

Michie S, Bron F, Bobrow M, Marteau TM: Nondirectiveness in genetic counseling: An empirical study. Am J Hum Genet *60:*40, 1997.

Monaghan KG, Van Dyke DL, Feldman G, Wiktor A, Weiss L: Diagnostic testing: A cost analysis for Prader-Willi and Angelman syndrome. Am J Hum Genet *60:*244, 1997.

Packer S: Family planning for women with bipolar disease. Hosp Community Psychiatry *43:*479, 1992.

Papolos DF, Faedda GL, Veit S, Goldberg R, Morrow B, Kucherlapati R, Shprintzen RJ: Bipolar spectrum disorders in patients diagnosed with velo-cardio-facial syndrome: Does a hemizygous deletion of chromosome 22q11 result in bipolar affective disorder. Am J Psychiatry *153:*1541, 1996.

Pelias MZ: Duty to disclose in medical genetics. Am J Med Genet *39:*347, 1991.

Pennington BF: Using genetics to dissect cognition. Am J Hum Genet *60:*13, 1997.

Pokorski RJ: Insurance underwriting in the genetic era. Am J Hum Genet *60:*205, 1997.

Reveley A: Genetic counseling for schizophrenia. Br J Psychiatry *147:*107, 1985.

Robins LN, Helzer JE, Weissman MM, Orvaschel H, Gruenberg E, Burke JD, Regier DA: Lifetime prevalence of specific psychiatric disorders in three sites. Arch Gen Psychiatry *41:*949, 1984.

Rothenberg K, Fuller B, Rothstein M, Duster T, Kahn MJE, Cunningham R, Fine B, Hudson K, King M-C, Murphy P, Swergold G, Collins F: Genetic information and the workplace: Legislative approaches and policy changes. Science *275:*1755, 1997.

Sharpe NF: Psychological aspects of genetic counseling: A legal perspective. Am J Med Genet *50:*234, 1994.

Swedo SE, Rapoport JL, Leonard H, Lenane M, Cheslow D: Obsessive-compulsive disorder in children and adolescents. Arch Gen Psychiatry *46:*355, 1989.

Thompson EA, Neel JV: Allelic disequilibrium and allele frequency distribution as a function of social and demographic history. Am J Hum Genet *60:*197, 1997.

Wachbroit R: The question not asked: The challenge of pleiotropic genetic tests. Kennedy Inst Ethics J *8:*131, 1998.

Weiner WJ, Lang AR, editors: *Behavioral Neurology of Movement Disorders*. Raven, New York, 1995.

28.5 DEATH, DYING, AND BEREAVEMENT

SIDNEY ZISOOK, M.D., AND NANCY S. DOWNS, M.D.

Death is a universal and unavoidable phenomenon. It arouses strong feelings of dread and fear in dying patients as well as in their families and health care providers. In clinical practice death is not the special province of any single discipline or the specialty of any one branch of medicine; rather, it is the universal reminder of life and its meanings. Most physicians, mental health workers, and clergy are involved in terminal-care situations and increasingly people die in hospitals or hospice settings. This brings the medical community to the forefront of decisions on when and how life should end. Similarly, the role of physicians in regard to their patients' suicidal impulses is becoming more complex. Rather than uniformly being on the side of preserving life and preventing suicide, physicians now are being urged in certain circumstances and in the name of compassion and dignity to help some patients carry out their suicidal wishes.

Once a patient dies, medical responsibility for the patient and their family is not over. Each person who dies leaves several loved ones, family, and friends behind. Grief is the general response to death. It is restorative for many, but the bereaved also are at elevated risk for general medical morbidity, mortality, and the onset or exacerbation of a number of mental disorders. Preventing or attenuating these adverse reactions is the task of the health care system.

DEATH AND DYING

History *Thanatology*, the study of death-related behaviors, thoughts, and feelings, is a relative newcomer in medical education curricula. In the early 1960s a few university texts on thanatology appeared, and professional journals such as *Omega* began publishing articles on death, dying, and bereavement. Largely due to Elizabeth Kubler-Ross's work on death and dying at the University of Chicago and to the hospice movement, death education grew through the 1970s. Eventually, popular interest in contemporary issues such as the right to die, organ transplantation, and euthanasia began to permeate the medical milieu; the study of death and dying was finally incorporated into most medical school programs by the 1980s.

Definitions The terms “death” and “dying” are not synonymous and have no unequivocal definitions. However, because medical, legal, and moral issues cluster around when and how death takes place, definitions are necessary. *Death* may be considered the absolute cessation of vital functions, while *dying* is the process of losing these functions. Dying may also be seen as a developmental concomitant of living, a part of the birth-to-death continuum. Living may entail numerous mini-deaths: the end of growth and its potential, health-compromising illnesses, multiple losses, decreasing vitality, growing dependency with aging, and finally dying. Dying and the person's awareness of it imbues humans with values, passions, wishes, and the impetus to make the most of time.

Advances in technology have changed the focus of death definitions. Since the late 1960s the definition of death has shifted from a focus on respiratory and circulatory function to a focus on brain activity. Brain functioning and the resuscitation potential of mechanically maintained patients are a current focus. Heart transplants in the 1960s caused “brain death” to gain in prominence. Viable, intact donor organs are needed for transplants, but because organs lose viability with prolonged use of artificial respirators physicians must be able to pinpoint when brain death occurs in such patients.

Uniform Determination of Death Act Responding to that need, the President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research published its definition of death in 1981. Working with the American Bar Association, the American Medical Association, and the National Conference of Commissioners on Uniform State Laws, the Commission established that one who has sustained either (1) irretrievable cessation of circulatory and respiratory functions, or (2) irretrievable cessation of all functions of the entire brain, including the brainstem, is dead. Determination of death must be in accordance with accepted medical standards. The provisions of this act are now law in several states, yet problems remain. For example, the concept of brain death may be criticized as extreme biomedical reductionism. Most people would agree the central issue is not whether a particular organ is functional, but whether the person is alive—a complex and subjective decision. Distinctions of this sort will inevitably evolve from the more difficult philosophical concept of what is meant by life and what death entails.

Personal Meanings of Death Avery Weissman distills different attitudes about death and its meaning into four categories: (1) Death is an illusion and an extension of life, a prologue to another form of life; death means transition, not extinction. (2) Death is an inevitable and inexorable fact of life, a confrontation with finitude; it is an endpoint in the “contract” between the living and life's parameters. (3) Death is an explanation and expiation of life, a final judgment on the life process; it offers retribution or reward, a release from mortality's constraints. (4) Death is an exigency and a defeat of life, a tragedy, negating life's values, signifying failure and futility. Each of these theories about death is in fact a belief about life. Is death a journey to another existence, a locked system with no exit, a chance for retribution and reward, or a futile exercise?

Vital Statistics Although everyone dies of something, the way one lives is often a prescription for how one dies. Most of the 10 leading causes of death can be linked to modern lifestyles. Controlling tobacco and alcohol, avoiding obesity, reducing stress and environmental carcinogens, limiting sale of firearms, and driving safely would substantially decrease age-adjusted death rates.

Table 28.5-1 shows the 1993 age-adjusted death rates for the 10 leading causes of death. In the United States over 2.2 million people die yearly, with a crude death rate of almost 900 deaths per 100,000. The two leading causes of death, heart disease and cancer, account for more than 50 percent of all deaths. Since 1980 deaths from cardiovascular and cerebrovascular diseases and accidents and homicide have declined, whereas deaths from diabetes, chronic lung disease, and human immunodeficiency virus (HIV) infection are on the rise.

Rank	Cause of Death	Age-Adjusted Death Rate per 100,000 Population	Change 1980–1995 (%)	Male/Female Ratio	Black/White Ratio
1	Diseases of the heart	138.2	−31.8	1.82	.79
2	Malignant neoplasms, including neoplasms of the lymphatic and hematopoietic tissues	129.8	2.2	1.2	.85
3	Accidents and adverse effects	29.2	−31.0	2.1	1.1
	Motor vehicle accidents	15.7	−35.4	2.3	1.0
	All other accidents and adverse effects	13.5	−102.7	2.2	1.2
4	Cerebrovascular diseases	26.7	−34.6	.68	.9
5	Chronic obstructive pulmonary diseases	21.2	33.3	1.2	.3
6	Human immunodeficiency virus (HIV) infection	16.2		5.4	.83
7	Diabetes mellitus	13.2	30.7	.81	1.4
8	Pneumonia and influenza	13.0	1.0	.88	.7
9	Suicide	11.0	3.3	4.9	.3
10	Homicide and legal intervention	8.8	18.3	4.9	7.0

Table 28.5-1 1995 Age-Adjusted Death Rates for the 10 Leading Causes of Death; Percentage Changes in Age-Adjusted Death Rates From 1980–1995 Sex- and Age-Specific Ratios for the United States

For the first time since 1980, the average life expectancy at birth declined (from 75.8 years to 75.5 years). The decline in life expectancy is thought to reflect increases in death rates for (1) chronic disease during influenza outbreaks, (2) pneumonia and influenza, and (3) infection and unintentional injuries. Race-specific ratios were greatest for homocide and suicide. Whites were 10 times more likely to die from suicide, but 7 times less likely to die from suicide than were African-Americans.

Sex-specific ratios were greatest for HIV infection, suicide, and homicide.

Unnatural causes of death are prevalent among the young. Between ages 14 to 24, the leading cause of death is accidents, mostly vehicular. Many fatal accidents, suicides, and homicides are alcohol related. Suicides, the third leading cause in this age group (particularly among white males), have doubled since 1950. Each year half a million suicides are attempted (using firearms, poison, or drug overdose) by American youths. Untreated depression puts this group at serious risk.

The high suicide rate in the elderly is seen predominantly in white males. The married elderly who maintain good health and are connected with extended families are at lowest risk.

The United States leads other civilized countries in homicides, the second leading cause of death in the 15- to 24-year age group and the leading cause among blacks aged 25 to 34 years. “Motiveless” murders or drug-related “gang” murders are now claiming statistical significance in the homicide category, but they are still overshadowed by the more traditional cases where victim and perpetrator know or are related to one another.

Stages of Death and Dying Kubler-Ross publicized and legitimized the study of death while promulgating more humane and sensitive treatment of the terminally ill. Her extensive interviews with 400 dying patients reveal that patients know without being told that they are dying; that they need to talk about it; and that they need to maintain hope, even if there is no hope of cure. Kubler-Ross's work reminds clinicians that fatal illness is not a sudden event, that dying patients retain their humanity, and that practitioners can provide support to patients by assuring them that they will maintain their relationship with them during the dying process.

Kubler-Ross postulates five stages that many dying patients pass through from the time they first become aware of their fatal prognosis to their actual death:

1. *Denial*. “No, not me!” is the dying patient's common initial response. If it does not interfere with treatment, denial can mitigate the initial overwhelming anxiety.
2. *Anger*. “Why me?” Indignation may surface when denial subsides. Patients are irritable, demanding, and critical; anger may be directed at themselves, caretakers, family and friends, or God.
3. *Bargaining*. “Yes me, but . . .” This stage entails promises to buy additional time. Patients may promise to donate their organs to research or they may reaffirm an earlier faith in God.
4. *Depression*. “Yes, me.” The patient comes to a full realization of what is going to happen and to whom. With the impending loss of life, a pervasive despondency may set in.
5. *Acceptance*. The patient begins to accept the inevitable. This stage need not constitute defeat or total surrender. “Yes, me, and I'm ready.”

These five stages are not all encompassing or prescriptive. Not everyone will reach these stages; perhaps only a few will reach acceptance. A patient may demonstrate aspects of all five stages in one interview or may fluctuate between stages. Moreover, patients may exhibit other coping methods—such as terror, humor, or compassion—to offset each stage.

Other researchers emphasize unique ways patients cope with dying. One posits a dynamic tension between acceptance and denial. In this model there is a clustering of intellectual and affective states, set against the backdrop of one's personality and philosophy. Between the constant ebb and flow of emotions lies an interplay of hope and disbelief. Acceptance and denial play a game of tug of war, with armies of emotions such as anguish, terror, acquiescence, surrender, rage, envy, disinterest, pretense, taunting, daring, and yearning for the deceased loved one.

Another model stresses coping styles and psychosocial issues brought into play in the intervals between the earliest phases of dying and death itself. The *prediagnostic stage* involves postponement and denial. During the *established-disease stage*, patients typically need to redefine themselves in terms of work, family, religion, economics, self-esteem, incapacity, and vulnerability. Illusions of survival and fears of extinction may fluctuate with varying degrees of denial or acceptance. The *decline and deterioration stage* is characterized by dependency. More medications and assistance, less autonomy and fewer options signal decline. Family, friends, and caregivers also experience emotional burdens as they face their own inadequacies and mortality. Thanatologists agree on the importance of a form of hope that does not avoid the truths of diagnosis, survival, or terminality. Hope for recovery can transform into hope for courage to confront death with peace of mind.

Near-Death Experiences Belief in an afterlife is common in spite of the absence of empirical validation. Afterlife phenomena have been reported throughout history. Apparition (ghost) experiences were investigated in early scientific inquiries; out-of-body experiences, messages from beyond the grave, and reincarnation memories are other “proofs” cited for life after death. One psychiatrist published data from 1000 cases of children who remember previous lives.

Recent informal evidence for an afterlife comes from the large number of Americans—up to 40 percent according to one estimate—who believe in near-death experiences. Gallup polls cited eight million such claims in the United States. The manner of near-death (illness, accident, or attempted suicide), the demographics (age, education, race, sex), and religious beliefs were not significantly related to frequency of a near-death report.

Near-death descriptions are often strikingly similar, involving an out-of-body experience of viewing one's body and overhearing conversations; feelings of peace and quiet; hearing a distant noise; entering a dark tunnel, leaving the body behind; meeting dead loved ones; witnessing beings of light; returning to life to complete unfinished business; and a deep sadness upon leaving this new dimension. This experience is usually described as peaceful and loving; it feels real to participants, who distinguish it from dreams or hallucinations. These experiences provoke sweeping lifestyle changes, such as fewer material concerns, heightened sense of purpose, belief in God, joy of life, compassion, less fear of death, enhanced approach to life and intense feelings of love. Although near-death experiences do not readily lend themselves to scientific investigation and thus are not “legitimized,” the popular press has made them common knowledge and patients may want to discuss them with clinicians.

Cultural and Religious Beliefs Death is universally regarded as an important event for the individual and the social unit. Every culture and religion provides ways for its members to think about death and respond to it. Rituals, elaborately organized beliefs, and viewing death as sacred are elements common to most cultural and religious responses to death and dying. A society's death beliefs and mourning practices reveal much about the structure, function, and substance of the social unit. American ways of dealing with death and dying are probably as heterogeneous as America; however, the scientific literature on American subcultural practices regarding death is scanty. Several investigators note that in the United States, death is probably one of the most intensely experienced of all human crises, yet it is rarely discussed openly—this reveals our culture's extreme anxiety about death.

Physicians must understand the sociocultural and religious context of the dying patient. Health care issues surrounding dying and death such as autopsy approval, routine lifesaving procedures, abortion, euthanasia, and suicide are embedded within a cultural context. This context influences the patient and the family as well as the physician. When patient and physician have cultural differences, the possibilities for poor communication, misunderstanding, incongruent expectations, and noncompliance are increased.

Death Beliefs All societies attempt to control death by symbolically and ritualistically assigning an order and meaning to life. Belief in an afterlife, communion with ancestors, and connection with the spirit realm are common in many cultures and religions. Several investigators note that symbolism and ritual serve the important function of sustaining the social group after its impermanent personnel (the individual) has left the group. Cultural or religious beliefs and practices provide a kind of immortality for the social group that is comprised of mortal members.

Mainstream religious groups in traditionally Protestant America have tended to view death with equanimity. Christianity is based on the alleged triumph of its leader over the death experience. In the invitation to participate in Christ's triumph over the grave, Christianity offers believers the hope of an eternal destiny through salvation. Belief in a future world is fundamental; death is the prelude to a more glorious future where the body is transformed and the soul lives on forever.

Catholics believe in life after death. However, before entering heaven the soul must be purified and the “mystery of love” perfected in purgatory. This is done by praying to God that purgatory will be brief and that God will expedite the soul's journey to heaven.

Like Christians, Muslims believe in life after death as another stage in God's plan. Thus, the death of a loved one is seen as a temporary separation and is accepted as part of God's will.

Judaism espouses less defined notions of immortality. Although Jewish literature and teachings speak of heaven as a reward for living in accordance with God's laws, the Old Testament contains no explicit reference to a hereafter. Hope of an afterlife is not at the center of modern Jewish life and thought; instead, Judaism emphasizes a belief in the timelessness of the human spirit. Individuals live on through their children, the survival of their memory, and the continuance of their

personality as a force among the living.

Buddhism views death as a normal process, a reality that is both inevitable and unpredictable. Instead of seeing death as an end or defeat, Buddhists liken death to changing clothes that are old and worn out. They consider the actual experience of dying as extremely significant because the state of mind at the time of death can influence the quality of the next rebirth.

Mourning Practices Religious and cultural views also strongly influence mourning practices and rituals. These practices provide a setting in which to verify death, say goodbye, and honor the deceased. Of all the expressions of grief associated with death, the most common universal expression is crying or wailing. Compared to other cultures, American's grief reactions are marked by emotional restraint.

The Church of Jesus Christ of Latter-Day Saints (Mormons) provides a strong supportive community within the church, and the belief in afterlife is so strong that bereaved survivors may be discouraged from active mourning.

Roman Catholic ritual begins with the wake, a 2-day vigil before the funeral. At this time members of the community, friends, and family come to pay their respects. The funeral is incorporated into the Mass. The casket is covered with a white pall, which symbolizes the white robe of baptism. The priest is also dressed in white—symbolic of the joy of faith, which overcomes the sadness and pain of death. A lighted candle representing the new life of Christ is placed next to the casket. A month later and then annually a special Mass is celebrated for the deceased.

Jewish Americans often observe specific rituals of grief and mourning, including rapid burials, laws against cremation, and prohibitions against autopsy. *Shiva*, a traditional 7-day period of full mourning after burial, provides a time for family and friends to mourn together. Next, a 1-month period of partial mourning ensues, and work and life gradually resume. Specific memorial prayers are repeated for 1 year, then the mourning period is terminated with a ceremonial unveiling of the tombstone.

Protestant mourning practices are varied, in part because of the numerous denominations. Usually the body is prepared by the undertaker, first viewed by the family, and then the public. The funeral is frequently held in the chapel of the funeral home on the afternoon of the third day after death. Prayers are said and scriptures are read that emphasize the Christian hope for resurrection and the sustaining power of God. A funeral sermon is preached that may refer to the deceased, followed by the procession to the cemetery and a brief committal service. Some communities serve a supper for friends and neighbors. An alternative to the traditional service is a memorial service in which the casket is not present.

Black Americans, traditionally more fervent than their white compatriots, frequently observe elaborate rituals to honor the dead and are likely to turn to religious leaders as well as family for support and comfort. Although blacks are more likely to die young than whites, blacks demonstrate less fear of dying. In general, blacks are more open towards dying and mourning, and receive more comfort from grieving.

Asian-Americans also value emotional expression during the mourning period. Generally an emotionally reserved cultural group, Asian-Americans are expected and encouraged to vent uninhibited emotion during the immediate postbereavement period. Further, grief is considered a lifelong process that is not expected to end or be resolved within any circumscribed period of time.

Wealthy Mexican-Americans display attitudes similar to those of wealthy whites, engaging in similar levels of denial and avoidance; in lower socioeconomic groups, acquiescence to one's own death is more common. Mexican-Americans mourn within their community and retain extended-family ties that provide a community context.

Native-American culture is comprised of many different tribes and tribal subcultures; thus, it is important to view grief and bereavement within the context of a given tribe. Viewing life and death as a continuous circle, most Native Americans see death as an integral part of the continuity of life. During the stress of bereavement, they rely on the support of family and tribe, often viewing life after death as a reunion with deceased ancestors. To cope with the death of a loved one, Native-Americans have developed a wide array of ceremonies and rituals. These tend to be social as well as personal events, and are important for the proper mourning of the deceased and for the restoration of balance and harmony. Aspects of normal grief and bereavement in Native-American cultures may be misinterpreted as pathological by outsiders. For example, the mourning practice of cutting the flesh may be misinterpreted as the self-mutilating behavior of a patient with borderline personality disorder. Several Native-American tribal cultures describe a phenomenon of ghost illness or ghost sickness that incorporates auditory and visual hallucinations and multiple somatic symptoms. Indeed, because Native-American cultures are so spirit centered, hallucinations, dreams, and visions, common throughout life, are often observed in periods of grief and bereavement.

African-, Asian-, Hispanic-, and Native-Americans are a handful of the varied racial and cultural subpopulations that define North American society. Clinicians who work with patients on the personal issues of death and dying must take into account the broader sociocultural context of individuals and their families.

Life Cycle Considerations The clinical diversity of death-related attitudes and behaviors between children and adults have their roots in age-dependent differences in the cause of death and in developmental factors. As opposed to adults, who usually die from chronic illness, children are apt to die from sudden, unexpected causes. Almost half of the children who die between 1 and 14 years of age and nearly 75 percent of those who die in late adolescence and early adulthood die from accidents, homicides, and suicides. With their characteristics of violence, suddenness, and mutilation, these unnatural causes of death are special stressors for the survivors. The bereaved parents and siblings of dead young children and teenagers often feel victimized and traumatized by their loss; their grief reactions resemble posttraumatic stress disorder. Devastating family disruptions can occur and the surviving siblings risk having their emotional needs delayed or unmet.

Children's attitudes toward death mirror their attitudes toward life. They share with adolescents, adults, and the elderly similar fears, anxieties, beliefs, and attitudes about dying. None welcome it without ambivalence, and all temper their acceptance with healthy doses of denial and avoidance. Even so, some universal concepts are relatively specific to age groups.

Children What we know about children's attitudes towards death come from surveying healthy children, relying on adult memories of childhood, or exploring fantasies and beliefs children express during bereavement. Systematic studies of dying children have been few because children generally experience sudden, violent deaths rather than lingering illnesses. Most late-childhood and adolescent deaths are unnatural, (accident, homicide, suicide); natural causes are generally malignancies or congenital abnormalities.

Dying children often are aware of their condition and want to discuss it. Research suggests that terminal and preterminal children 3 years of age and older are aware that they are dying without being told. They often have more sophisticated views about dying than their healthy counterparts, probably because of their own failing health, separations from parents, subjection to painful procedures, and deaths of hospital friends. Attendant fears and anxieties may be unexpressed or channeled in maladaptive directions.

Often unable to meet dying children's needs or to face their own feelings, parents and caretakers may tend to avoid afflicted children, even though human contact and communication are crucial. Perplexed parents may go to extremes—overindulging their dying children, withdrawing from them, or insisting that nothing is wrong. In such instances the child receives little help with the anxieties and fears of separation and abandonment or in developing the requisite hope and courage. Eventually sensing that they are dying, such children become more terrified than those who receive guidance and support from empathic, honest family members.

The affective and cognitive development of children colors their understanding of death and their subsequent fears about dying. For example, at the preschool, preoperational stage of cognitive development, death is seen as a temporary absence, incomplete and reversible, like departure or sleep. Separation from the primary caretaker(s) is the main fear of a preschooler. This fear surfaces as an increase in nightmares, more aggressive play, or concern about the deaths of others rather than in direct discourse. Regression to more infantile behaviors signals increasing dependence on parents. Such children may assume responsibility for their death, feeling guilty for dying. Hospitalizations, separations from parents, and being subjected to painful procedures may reinforce their guilt and their belief that they are being punished. They may be unable to relate the treatment to the illness, and may instead view treatment as punishment. Dying preschoolers need reassurance from their parents that they are loved, that they have done nothing wrong, that they are not responsible for their illness, and that they will not be abandoned.

School-aged children manifest concrete-operational thinking and recognize death as a final reality. However, they view death as something that happens to old people, not to them. Between the ages of 6 and 12 years children have active fantasy lives of violence and aggression, often dominated by themes of death and killing. Games of cops and robbers or cowboys and Indians involve causal mayhem in which the “dead” are routinely resuscitated. Death may be personified as a

skeleton or bogeyman that takes people away. Dying school-aged children ask questions about their illness if encouraged to do so; however, if they receive cues that the subject is taboo, they may withdraw and participate less fully in their own care. They need help to cope with their peers and the demands of school. Teachers should be informed and updated about the child's condition. Classmates may need education and assistance to help them understand and respond to what is happening to their friend.

Adolescents Capable of formal cognitive operations, adolescents understand that death is inevitable and final. Their major fears parallel those of all teenagers: loss of control, being imperfect, and being different. Concerns about body image, hair loss, or loss of bodily control may generate great resistance to continuing treatment. Alternating emotions of despair, rage, grief, bitterness, numbness, terror, and joy are common. An adolescent's cognitive capacity to understand death may not translate into an understanding that their own personal death is possible. The potential for withdrawal or isolation is great because teenagers may equate parental support with loss of independence or may deny their fear of abandonment by actually repulsing friendly gestures. Teenagers must be part of the decision-making process surrounding their death. Many are capable of great courage, grace, and dignity in facing death.

Adults Unlike children and teenagers, older adults often readily accept that their time has come. Elderly patients may talk or joke openly about dying, and sometimes welcome it. In their 70s or beyond, they no longer harbor illusions of indestructibility; most have already had several close calls, their parents have died, and they have gone to funerals for friends and relatives. Although they may not be happy to die, they can be reconciled to it.

According to Erik Erikson, the eighth and final stage in the life cycle brings either a sense of integrity or despair. As elderly adults enter the last phase of their lives, they reflect on their time and how it has been lived. When one has taken care of things and is relatively successful and adapted to the triumphs and disappointments of life, one can look back with satisfaction and only a few regrets; one experiences a sense of integrity about oneself, feeling that one has lived totally and well and that one's life has been meaningful. Integrity of the self allows an individual to accept inevitable disease and death without fear of succumbing helplessly. However, if a person looks back on life as a series of missed opportunities or as filled with personal misfortunes, the sense is of bitter despair, a preoccupation with what might have been if only this or that had happened; then death is viewed with fear because it symbolizes emptiness and failure.

Caring Caring for a dying patient is highly individual. Caretakers need to deal with death honestly, tolerate wide ranges of affects, connect with a suffering patient and bereaved loved ones, and resolve routine issues as they arise. Although each therapeutic relationship between a patient and health provider is unique because of the patient's and health provider's gender, constitution, life experience, age, stage of life, resources, faith, culture, and other considerations, there are major themes confronted by all health providers caring for dying patients.

Breaking the News After diagnosis and prognosis have been made, physicians need to tell the patient and families. Formerly, doctors subscribed to a conspiracy of silence, believing that the less patients knew, the better their chance for recovery; it was believed that news of impending death brought despair, so the truth was withheld. The current policy is honesty and openness; the question is not whether to tell the patient, but when and how.

The American Hospital Association in 1973 drafted the “Patients' Bill of Rights,” declaring that patients have a “right to obtain complete current information regarding diagnosis, treatment and prognosis in terms the patient can be reasonably expected to understand.” Full disclosure is mandated whether the patient requests it or not.

In breaking the news of impending death to the patient, diplomacy dictates the following guidelines: (1) The health team physician should participate. (2) The spouse or partner should be present if possible. (3) Relatives need comfort, as does the patient. (4) Use simple words, even with educated patients. (5) Show compassion and emotional support, avoiding bluntness or abruptness. (6) Guessing how long a patient has to live may be inaccurate and inadvisable. (7) Encourage and answer questions, signaling one's availability for honest communication. (8) Truth is not the enemy of hope. (9) Communicate a willingness to see the patient through until death occurs. (10) Explain the situation and introduce the next step. A gentle, sensible approach helps to modulate the patient's own denial and acceptance. The physician must choose how much information to give and when, on the basis of the patient's needs and capacities.

Fears Knowing that one's death is imminent can precipitate a number of fears. Suddenly, every bodily change is interpreted as ominous, and can provoke terror. Dying patients commonly express fears of disfigurement; loss of autonomy; being a burden; being physically repulsive; disappointing family, friends, and colleagues; and facing the unknown. One group of terminally ill cancer patients surveyed reported that their greatest fears were of abandonment, pain, and shortness of breath. Effective management programs must address patient concerns. Patients need to know that they will not be abandoned and that relief from severe pain is available. Patients benefit greatly from the reassurance that their providers will not allow them to live or die in pain.

Goals of Treatment Facing an illness that cannot be cured, patients enter a phase where comfort becomes the primary goal. As illness advances and physical debilitation progresses, the incidence of pain, depression, and delirium increases. As these complications are the antithesis of comfort, they must be carefully assessed and controlled as best as possible. For palliative care to be effective, suffering must be minimized so that the quality of the remaining life can be maximized. In her landmark publication, *On Death and Dying*, Kubler-Ross described the unfinished tasks of the dying: reconciliations, resolution of conflicts, and the pursuit of specific remaining goals. Kubler-Ross sees the clinician's major task as helping patients reach these targets.

The Institute of Medicine has defined a *good death* as one that is free from avoidable distress and suffering for patients, families and caregivers; in general accord with patients' and families' wishes; and reasonably consistent with cultural, clinical, and ethical standards. With such goals in mind, three principles have been described in the psychosocial management of dying patients: safe conduct, dignified dying, and appropriate death. *Safe conduct*—defined as acceptance, clarity, candor, compassion, and accessibility—is imperative in caring for terminal patients. The difference between having a healthy dependency and being a sick victim is the ability to trust and be trusted, so patients must feel that caretakers will be there and will treat them with respect and dignity. Safe conduct can help the patient achieve a *dignified dying*, which involves the patient being regarded as responsible and capable of clear perceptions, honest relationships, and purposeful behavior, consistent with the current level of ability. The physician helps the patient relinquish autonomy as needed, prescribes required medication, and encourages the patient to resolve any remaining conflicts. The third goal for the physician is to help the patient die an *appropriate death*, one that the person might choose if he or she had the choice. Suffering is at a low ebb, conflict is minimal, and behavior has been maintained on as high a level as is compatible with physical status. The patient both expects and accepts death, and is willing, albeit ruefully, to die.

Seven Cs of Care Ned Cassen suggests seven essential features in the management of the dying patient: (1) *Concern:* Empathy, compassion, and involvement are essential; concern is ranked as the quality most appreciated by patients. (2) *Competence:* Skill and knowledge can be as reassuring as warmth and concern. In particular, health care providers must adeptly manage the complications of terminal illness—pain, nausea, shortness of breath, and hopelessness. Patients benefit immeasurably from the reassurance that their providers will not allow them to live or die in pain. (3) *Communication:* Allow patients to speak their minds and get to know them. (4) *Children:* If children want to visit the dying, it is generally advisable; children bring consolation to dying patients. (5) *Cohesion:* Family cohesion reassures both the patient and family. The clinician who gets to know the family maximizes patient support and is prepared to help the family through bereavement. (6) *Cheerfulness:* A gentle, appropriate sense of humor can be palliative; a somber or anxious demeanor should be avoided. (7) *Consistency:* Continuing, persistent attention is highly valued by patients who often fear that they are a burden and will be abandoned; consistent physician involvement mitigates these fears.

Assisting the Family Family members may need help on how to best aid and relate to their dying loved one. The task of coping with major changes in roles and relationships associated with the terminal stage of illness is compounded by a lack of preparation on how to behave in the presence of someone who is dying. Fatal illness destroys future hopes and plans, educational ambitions, established marital and parent-child adjustments, financial stability, and lifestyle. Extended illness puts excessive pressure on existing social support systems within and around the family, depleting financial and social resources as well as causing interpersonal family relationships to become estranged. One researcher has offered the following suggestions to health professionals on how to assist the family: train family members to participate in treatment; encourage them to provide emotional support and to anticipate the special needs of the patient; allow unlimited visiting so that the whole family, including children, can spend time with the patient; and provide special hospital space for patient-family meetings as well as for family members to live when the patient's death is imminent. Many of these suggestions are incorporated in hospice care of the dying and their families.

Professional Supportive Programs Supportive interventions may enhance health and contribute to survivorship in some life-threatening illnesses. Several studies have shown that group therapy interventions enhance effective coping and increase well-being in cancer and acquired immune deficiency syndrome (AIDS) patients. One important study on the effects of group intervention for women with metastatic breast cancer found that women who receive supportive group therapy and self-hypnosis for pain alleviation have higher 10-year survival rates than those who do not.

Self-Help and Mutual Support Self-help programs provide a supportive milieu for patients to share their expertise and experience. After World War II the American Cancer Society began volunteer programs in which patients visited patients to offer practical help and support. The Society later sponsored volunteer programs for women with disfiguring surgical procedures (Reach to Recovery), for families dealing with childhood cancer (Candlelighters), and for other special interest groups.

Recently, volunteer support programs for serious medical problems, as well as for families who have survived potentially fatal illnesses, have proliferated. Such programs offer patients contact with others who have experienced similar fears and obstacles and who have found a way to face death courageously. Groups buffer the impact of illness and enhance coping by providing a sense of belonging, emotional support, access to information, social contact, and self-esteem. Reinforcement to maintain effort and face painful procedures is also offered. Most groups work in proximity to professional medical services as an adjunct to traditional care, and physicians are well advised to refer patients and family members to such programs.

Hospice The founding of St. Christopher's Hospice in England by Cicely Saunders in 1967 launched the modern hospice movement. Several factors in the 1960s propelled the development of hospices, including concerns about inadequately trained physicians, inept terminal care, gross inequities in health care, and neglect of the elderly. Life expectancy had increased and heart disease and cancer were becoming more common. Saunders emphasized symptom control, care of patient and family as a unit, an interdisciplinary approach, the use of volunteers, a continuum of care including home care, continuity of care among several settings, and follow-up with family members after a patient's death. By 1974 the first hospice in the United States, Connecticut Hospice, opened. As of 1998, there are about 3000 hospices in the United States. In 1997, they cared for more than 450,00 persons (about one fifth of Americans who died).

Round-the-clock pain control with opioids is an essential component of hospice management. In 1983 Medicare began reimbursing hospice care. Medicare hospice guidelines emphasize home care, with benefits for a broad spectrum of physician, nursing, psychosocial, and spiritual services at home, or if necessary, in the hospital or nursing home. To be eligible the patient must be physician-certified as having 6 months or less to live. By electing hospice care, patients indicate that they agree to receive palliative rather than curative care and waive their right to alternative medical treatments. Many hospice programs are hospital-based, sometimes in separate units and sometimes in the form of hospice beds interspersed throughout the facility. Other program models include free-standing hospices and programs, hospital affiliations, nursing home or extended-care facilities, and home care programs only.

Hospice families generally express satisfaction with their personal involvement in care. Savings with hospice are variable, but home care programs generally cost less than conventional care, particularly in the final months of life. Hospice patients are less likely to receive diagnostic studies or intensive therapy such as surgery, chemotherapy, or medication, and are more likely to die at home. However, research studies have failed to show that the quality of life for terminally ill patients served in the hospice setting is significantly different from the quality of life for patients served in conventional settings. This may be because most traditional programs incorporate modalities offered by hospice programs. Hospice care is a proven, viable alternative for patients who elect a palliative approach to terminal care. In addition, hospice goals—dignified, comfortable death for the terminally ill and care for patient and family together—have been increasingly adopted into mainstream medicine.

Alternatives to Mainstream Medicine Even many well-educated patients, once they are told they are terminally ill, do anything within their means to outrun the prognosis. They may seek alternative treatments, ranging from innocuous programs aimed at enhancing one's general health to more aggressive, harmful, or fraudulent treatments. The majority of unorthodox practitioners are physicians. Although most patients combine unorthodox with traditional treatment, a substantial number favor unorthodox approaches as the disease progresses.

Presently unproven methods to cure cancer emphasize a holistic approach, involving purification of the body, detoxification through internal cleansing, and attention to nutritional and emotional well-being. Despite their widespread appeal, none of these alternatives has been demonstrated to cure cancer, prolong life, or diminish suffering; yet all have strong followings bolstered by anecdotal accounts of their efficacy. The popular *metabolic therapy* attributes cancer to toxins and waste material accumulating in the body; treatment is based on reversing this process by diet, vitamins, minerals, enzymes, and colonic irrigations. *Nutritional approaches*, such as macrobiotic diets or megavitamins to enhance the body's capacity to destroy malignancy, also lack firm scientific support. In 1987 the National Research Council recommended minimizing carcinogenic substances and fat in the diet, and increasing whole-grain, fruit, and vegetable consumption as preventive guidelines. *Psychological approaches* cite maladaptive personality and coping styles as being cancer-causing; treatment consists of shaping a positive attitude. *Spiritual and mystical approaches* aim at harmony between the patient and nature; cancer may be seen as an external evil to be exorcised rather than treated. Finally, *immunotherapies* have gained popularity in recent years; cancer is attributed to a defective immune system and restoration of immunocompetency is seen as the cure. Substantive evidence of efficacy is lacking in all these approaches.

Society's present emphasis on health consciousness, mind-body relationships, and individual responsibility is reflected in the current holistic alternative treatments. Staff working with terminally ill patients, especially those with cancer, must be informed of these approaches and be prepared to discuss them with patients. They must assess the patient's wish for these therapies, especially because they reflect emotional needs unmet by conventional treatment. Until cancer and other progressive chronic illnesses are successfully treated, alternative cures will continue to surface.

Aggressive attempts also are made to forestall or overcome death. *Cryonic suspension* attempts to maintain patients in an indefinite hypothermic state until medical science develops means to resuscitate the body and cure the terminal illness. Several bodies are already cryonically suspended, awaiting the future, despite scientific skepticism about the procedure or its ultimate desirability. Others look to genetic engineering or cloning as potential antidotes to mortality. These approaches tend to see death as the enemy rather than as an integral part of life.

PHYSICIAN AID IN DYING

Until recently traditional medicine did not entertain the notion of allowing patients to choose when to die. Three major forces have coalesced to force the medical community to alter its position. First, more patients die in medical institutions, often from chronic conditions such as cancer, AIDS, or neurodegenerative processes. It is estimated that over 60 percent of all deaths occur in hospitals, with another 20 percent occurring in nursing homes. Of those dying at home, a growing minority do so with hospice care. Increasingly, health care providers have a significant and ongoing role in the way people live out their last days to weeks of life. Second, advances in modern medical technology can prolong an individual's biological processes such as respiration well beyond the point at which they have any reasonable hope for achieving an autonomous, meaningful existence. This capacity has created countless medical, legal, and ethical dilemmas that now face patients, their families, and society at large. Third, the trends toward increasing patient involvement in care, patient rights, and consumerism in medicine have forced the medical community to listen to patients' wishes and to offer them choices rather than to dictate solutions.

The current debate surrounding patient autonomy regarding life and death decisions was fueled by 3 pivotal court cases with variable outcomes. The most celebrated court case involved Karen Ann Quinlan, whose adoptive parents petitioned the court to disconnect her respirator after she had lapsed into a long, irreversible coma. The New Jersey Supreme Court, after much debate, ruled that a ventilator (presumed to be keeping her alive) could be turned off. To the surprise of all parties, the patient lived another 10 years without the respirator and died of pneumonia in 1985.

In a second landmark case, a nonvegetative patient, Elizabeth Bouvia, asked to be allowed to die. Suffering from cerebral palsy, with just enough muscular control to generate her electrically powered wheelchair and to chew and swallow if fed by someone else, she wanted to be admitted to a hospital and allowed to starve to death. The hospital force-fed her, which brought the issue to court. Although the judge thought that Bouvia was competent and sincere, he ruled in favor of the hospital, refusing to sanction passive euthanasia in a nondying patient, despite the patient's wishes and her chronic suffering. Ultimately, a California Court of Appeals ordered physicians to remove the nasogastric tube from Bouvia, rejecting the hospital's contention that removing the patient's tube would make it party to a suicide. Thus, the court did not require that a patient be comatose or terminally ill to refuse life-saving treatment.

In a third pivotal case the United States Supreme Court ruled in 1988 on an appeal regarding Nancy Cruzan. Cruzan had been in a persistent vegetative state since an accident in 1983, and her parents sought permission to withdraw her feeding tube and let her die. Citing the lack of evidence about what her wishes would have been, the Missouri Supreme Court denied the Cruzans' request, a decision that the United States Supreme Court upheld by a five-to-four vote. Although it affirmed the right of competent persons to refuse life-sustaining treatment, the court's decision retained the broad powers in individual states to rule on such matters without federal intervention. When originally proposed 30 years ago, advance directives were short statements by patients wishing to avoid “heroic measures.” Such directives have evolved into three types of advance directives: (1) living wills, (2) durable powers of attorney, and (3) value histories.

Advance Directives As the movement for the protection of health care consumers gains credibility and as patients begin to realize that they have rights, more individuals are leaving *advanced directives*. Such documents enable competent persons to state the type of medical care that they wish to receive should they become unable to participate in treatment decisions. A *living will* allows patients to express in writing whether they elect to receive or expressly refuse such procedures as cardiopulmonary resuscitation, artificial respiration, parental nutrition, and hydration. These documents may also have provisions regarding pain medications, even those that may shorten life. Living wills are legally enforceable in most states and are generally endorsed by the medical and legal professions. Table 28.5-2 provides an example of a living will. Another type of advanced directive, a *durable power of attorney*, designates someone to make medical decisions in the event that the person becomes incompetent or otherwise unable to make decisions. The third type of advanced directive, values histories, allows patients to

indicate their values and what gives their lives meaning, some for such as the so-called *Medical Directive* are a combination of all the types.

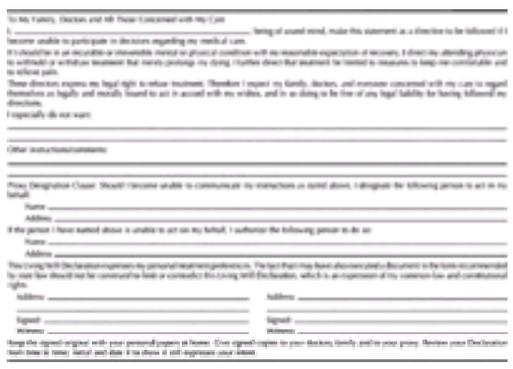

Table 28.5-2 Advanced Directive

The Uniform Rights of the Terminally Ill Act, drafted by the National Conference of Commissioners on Uniform State Laws was approved and recommended for enactment in all states at its annual conference in 1985. The act authorizes an adult to control the decisions regarding the administration of life-sustaining treatment by executing a declaration instructing a physician to withhold or to withdraw life-sustaining treatment if the person is in a terminal condition and is unable to participate in medical treatment decisions. In the early 1990s the Self-Determination Act became law in the United States and required that all health care facilities provide each patient admitted with written information about the right to refuse treatment, ask about advance directives, and keep written records of whether the patient has a directive or has designated a health care proxy. When such clarity is not present, common sense should prevail and the patient's next of kin should be asked to provide substitutive judgment about what the patient would want.

Euthanasia From the Greek term for good death, *euthanasia* means compassionately allowing, hastening, or causing the death of another. Generally someone resorts to euthanasia to relieve suffering, maintain dignity, and shorten the course of dying when death is inevitable. Euthanasia can be *voluntary* if the patient has requested it, *nonvoluntary* if the decision is made without the patient's consent, or *involuntary* if the decision is made against the patient's wishes. Euthanasia can be *passive*, simply withholding heroic lifesaving measures, or *active*, deliberately taking a person's life. Although the moral, ethical, medical, and legal issues involved in each form of euthanasia have not been fully worked out, most clinicians have little problem with passive, voluntary euthanasia if the patient's death is thought to be inevitable, close at hand, and accompanied by relentless suffering. However, few agree that involuntary, active euthanasia is acceptable under any circumstances. The ethics of other combinations of activity and consent are controversial.

Arguments for and against euthanasia have been articulated by well-intentioned, thoughtful individuals on both sides of the debate. Proponents of euthanasia argue that quality of life and patient dignity are more important than the number of hours or days a patient lives. Two of the strongest arguments revolve around patient autonomy and dignified dying. Respect for each patient's autonomy has become a guiding principle in medicine. One of the most dramatic ways patients may exercise their right to self-determination is by asking for life-sustaining treatment to be withdrawn. If the patient is mentally competent, physicians are honor bound to respect such wishes. Proponents of active, voluntary euthanasia argue that the same rights should be extended to patients who are not on life-sustaining treatment but who also choose to have their physicians help them die. The second argument rests on the knowledge that some deaths—especially from cancer, AIDS, or progressive neurological lesions—may be slow and miserable. Not all pain can be relieved, and often symptoms, such as dyspnea, nausea, and weakness are even more difficult to control. Furthermore, dying may bring great indignities and extended suffering. Thus, euthanasia may provide a dignified and peaceful alternative for some hopelessly miserable patients.

Opponents of euthanasia also provide strong ethical and medical justification for their position. First, active euthanasia, even if the patient voluntarily requests it, is a form of killing, and outside of war, self-defense, or perhaps capital punishment, killing should never be sanctioned. Second, euthanasia is almost never necessary because most suffering can be relieved with skillful and compassionate care. The fact that some physicians are poorly educated in pain control or other palliative care measures should be addressed by education rather than by legalizing euthanasia. Third, many patients who request aid in dying may have unrecognized or untreated depression, a difficult diagnosis to make in the context of chronic and painful medical illness. Core symptoms of major depressive disorder include the wish to die, compromised ability to think clearly, a sense of hopelessness about the future, a tendency to embellish pain and misery, and a view of the world that embraces pessimism and negativity. Yet, many terminally ill patients see primary care physicians who may be ill-equipped to diagnose depression in the elderly or other complex situations. Even well-trained psychiatrists have difficulty diagnosing depression in severely ill patients. Studies indicate that the will to live of many patients with life-threatening illnesses improves substantially after treatment of depression. Fourth, legalizing voluntary euthanasia will put us on a moral slippery slope, inexorably leading to involuntary euthanasia or to euthanasia for children with birth defects or adults with mental illness. Thus, once active euthanasia is tolerated or legalized, increasingly liberal interpretations of the guidelines will occur. As barriers are broken the line of acceptable cases will be redrawn and there will be little justification for limiting the practice to the terminally ill or to persons who are cognitively intact enough to request it. Why not extend one's “help” to persons in the earliest stages of HIV infection, nonterminal accident victims who envision a life of disability and dependence, patients suffering from chronic treatment-refractory depression, or illnesses such as Alzheimer's disease? Fifth, euthanasia would be a particular threat to the economically and socially vulnerable; and the poor, mentally ill, or disabled may be coerced to request it. In the present era of managed care, with its emphasis on cost containment and the economics of care, this argument is particularly disturbing. Patients may request euthanasia in order to not be a burden; overwrought family members might pressure vulnerable patients to request it, and cost-conscious doctors may see it as a solution to capitated care. Finally, euthanasia violates the time-honored mission of physicians to be unambiguously committed to preserving life. Legalizing euthanasia could lead to further distrust of doctors, and could keep otherwise needy patients from seeking care or consulting with physicians. Although many physicians would agree that patients have a right to die, and competent, rational patients may have a right to choose when to die, that does not translate into physicians having the right or obligation to kill.

Most medical, religious, and legal groups in the United States are more impressed with the arguments against euthanasia than for it and oppose legalization of euthanasia. Both the American Psychiatric Association and the American Medical Association condemn active euthanasia as illegal and contrary to medical ethics; however, few persons have been convicted of euthanasia.

The World Medical Association issued the following declaration on euthanasia in October 1987: “Euthanasia, that is, the act of deliberately ending the life of a patient, even at his own request or at the request of his close relatives, is unethical. This does not prevent the physician from respecting the will of a patient to allow the natural process of death to follow its course in the terminal phase of sickness.”

The New York State Committee on Bioethical Issues also opposes euthanasia. The committee stated the following:

> Physicians are obligated to relieve pain and suffering and to promote the dignity and autonomy of dying patients in their care. This obligation includes providing effective palliative treatment even though it may occasionally hasten death. However, physicians should not perform active euthanasia or participate in assisted suicide. Support, comfort, respect for patient autonomy, good communication, and adequate pain control may dramatically decrease the demand for euthanasia and assisted suicide. In certain carefully defined circumstances, it is humane to recognize that death is certain and suffering is great. However, the societal risks of involving physicians in medical interventions to cause patients' death is too great to condone active euthanasia or physician-assisted suicide.

Both proponents and opponents of euthanasia are closely watching the experience in the Netherlands, where euthanasia is officially a crime but its practice has been accepted for over 20 years, protected by a body of case law and strong public support. In 1984 the Royal Dutch Medical Association set out criteria for performing euthanasia or physician-assisted suicide: (1) the request for euthanasia must come only from the patient who must be entirely free, voluntary, and a mentally competent adult; (2) the patient's request must be considered, durable, persistent, and documented; (3) the patient must be experiencing intolerable (but not necessarily physical) suffering, with no prospect of improvement; (4) other alternatives must have been considered and found wanting; (5) the euthanasia must be performed by a physician; and (6) the physician must consult with an independent colleague who has experience in the field.

In 1990 the Dutch government appointed a commission, headed by Professor Jan Remelink, the Attorney General of the Dutch Supreme Court, to oversee a careful, nationwide study of the practice. From physician interviews and questionnaires, the commission reported that of the roughly 130,000 deaths in the Netherlands in

1990, 38 percent involved a medical decision concerning the end of life. Most of these were to allow someone with a terminal illness to die without heroics or to treat unbearable pain with medications in such doses that the patient's life may have been shortened. However, 2300 deaths, or 2 percent of the total, were the result of euthanasia. The most frequent reasons for requesting euthanasia included loss of dignity (57 percent), pain (46 percent), unworthy dying (46 percent), being dependent on others (33 percent), and tiredness of life (23 percent). Another 400 deaths (0.3 percent) were assisted suicides. In 1000 of the deaths (0.8 percent) the patient was not consulted when euthanasia was performed despite the clarity of the Dutch guidelines regarding this requirement. In response to these involuntary cases, in 1993 the national government and medical association released a statement condemning involuntary euthanasia, and reporting procedures were clarified by law. Still, only a minority of the physician-assisted deaths are reported as such on death certificates.

A study of Dutch nursing home physicians who administer euthanasia found that about 1 out of every 8 requests for euthanasia in nursing homes is granted. The most frequent reasons for euthanasia requests were fear of deterioration of condition (25 percent), fear of suffocation (15 percent), unbearable suffering (14 percent), and pain (11 percent). The period of time from the first discussion of the subject and the actual administration varied from more than a year (3 percent) to less than a day (7 percent). In 15 percent of the cases no consultation with another physician was requested, and in 15 percent of the cases it was someone other than the patient (e.g., child of patient, or physician) who first broached the subject. All guidelines were met only 41 percent of the time.

Both proponents and opponents of euthanasia point to the Dutch experiment to support their position. Those who favor some form of physician-assisted dying in the United States are reassured by the small fraction of total deaths accounted for by euthanasia or assisted suicide. Furthermore, it was performed in fewer than a third of those who requested it, even less so in nursing homes, indicating the reluctance of many doctors to honor such requests. And, euthanasia was usually performed on those who were terminally ill. On the other hand, those who oppose euthanasia or assisted suicide point to the abuses. In almost 1000 cases, or nearly 1 percent of all deaths, patients were euthanized without their request or consent. Accurate reporting remains a problem, and in nursing homes the topic of euthanasia may be first broached by physicians themselves who often carry it out almost immediately after initial discussion and often do not consult other physicians. The importance of waiting before carrying out euthanasia is underscored by the finding that 85 percent of the patients that initially request aid in dying in the Netherlands withdrew their requests once their symptoms were controlled. Furthermore, many argue that in the United States, which, unlike the Netherlands, does not have universal health insurance, and where large segments of the population are uninsured, these abuses are even more likely.

A recent survey of doctors and death in Australia found that medical decisions to end life are involved in about 30 percent of deaths in that country. Potentially life-prolonging treatment is withheld or withdrawn in 25 percent of deaths; in 24 percent of Australian deaths, a decision to use high doses of opioids to relieve pain may have hastened death without an explicit intent to do so; in another 7 percent of deaths, a decision to use high-dose opiods for pain relief was made with some intention to hasten death; and in 4 percent of deaths a decision was made to administer drugs that would end a patient's life without a specific request from the patient. Despite the controversy stirred by this survey, Australia's Northern Territory became the only country in the world where euthanasia and physician-assisted suicide are legal and practiced. Enacted in July 1996, Australia's Rights of the Terminally Ill Act mandates that the patient must be at least 18 years old and mentally and physically competent to make the request. At least three doctors must concur, including a specialist confirming that the patient is terminally ill and a psychiatrist who certifies that the patient does not suffer from a treatable depression. In addition, after all paperwork has been completed, there must be a waiting period of at least 9 days before the procedure can be administered. Since July 1996 four people died under the Act. However, by the end of the following year the Australian Parliament overturned the law.

Physician-Assisted Suicide In the United States most of the debate centers on physician-assisted suicide rather than euthanasia. Some have argued that physician-assisted suicide is a humane alternative to active euthanasia in that the patient maintains more autonomy, remains the actual agent of death, and may be less likely to be coerced. Others believe that the distinctions are capricious in that the intent in both cases is to bring about a patient's death. Indeed, it may be difficult to justify providing a lethal dose of medication to a terminally ill patient (physician-assisted suicide) while ignoring the desperate pleas of another patient who may be even more ill and distressed but who cannot complete the act because of problems with swallowing, dexterity, or strength.

There are several degrees to which a physician may assist in ending the life of a suicidal patient. Physician-assisted suicide may involve providing information on ways of committing suicide, supplying a prescription for a lethal dose of medication or a means of inhaling a lethal amount of carbon monoxide, or perhaps even providing a suicide device that the patient can operate.

Founded by Derek Humphry in 1965, the Hemlock Society is concerned with active euthanasia and assisted suicide. It boasts a national membership and broad recognition. The success of Humphry's book *Final Exi*, a how-to guide to suicide and a bestseller in 1991, attests to its popularity. The society supports physician-assisted suicide and active, voluntary euthanasia for the terminally ill but it does not encourage suicide in the absence of terminal illness. The Hemlock Society has led state initiatives to permit physicians to help terminally ill patients die in Washington in 1991 and again in California in 1992. The so-called death-with-dignity initiative did not pass, but in each case it received almost 50 percent of the popular vote. Since then, the Hemlock Society's efforts influenced the 1994 passage of a ballot initiative to legalize physician-assisted suicide in Oregon and in making public and medical opinions more accepting of physician-assisted suicide.

In 1991 the *New England Journal of Medicine* published the case of Diane, a woman with acute leukemia who elected not to undergo the recommended course of treatment, which had a 25 percent chance of success. This publication fueled the debate about the physician's role in assisted suicide. When Diane initially raised the possibility of taking her own life, her physician referred her to the Hemlock Society for information. When she subsequently requested enough barbiturates to kill herself, her physician gave her the prescription she eventually used to end her life. Despite her past history of depression and intermittent complaints of low energy, anger, trouble sleeping, and sadness, her physician felt that giving her the means to kill herself was appropriate, humane, and compassionate care. Others argued that the case highlights the potential pitfalls of legalizing physician-assisted suicide and the need for a comprehensive evaluation of occult depression in all patients who seriously consider suicide.

The controversy is further heightened by the publicity and court proceedings surrounding the activities of retired pathologist, Jack Kevorkian, who in 1989 provided his suicide machine to a 54-year-old woman with a diagnosis of probable Alzheimer's disease. After the woman killed herself with his device, Kevorkian was charged with first-degree murder. The charges were later dismissed because Michigan had no law against physician-assisted suicide. Since that case, Kevorkian has assisted in over 100 additional suicides, often for persons he had met only a few times and frequently for persons who did not have a terminal illness. Kevorkian's license to practice has been revoked and more murder charges have been leveled against him. The most recent charges against Dr. Kevorkian stem from a televised depiction of his first publicized procedure of clear-cut active euthanasia. He was convicted of manslaughter in 1999 and sentenced to prison. That an unliscenced physician could so readily move from assisted suicide to active euthanasia provides additional fuel to the slippery slope argument. His attorneys and followers applaud his courage in easing pain and suffering; his detractors counter that he has become a serial mercy killer. Several opponents of Kevorkian's methods charge that without safeguards, consultations, and thorough psychiatric evaluations, patients may seek suicide not because of terminal illness or intractable pain but because of untreated depressive disorders. They argue that suicide rarely occurs in the absence of psychiatric illness. Finding more effective treatments for pain and depression rather that inventing more sophisticated devices to help desperate patients kill themselves defines compassionate and effective physician care.

Despite the abhorrence that many physicians and medical ethicists express toward physician-assisted suicide, poll after poll shows that as many as two thirds of Americans favor the legalization of physician-assisted suicide in certain circumstances. There is even evidence that the formerly uniform opposition to physician-assisted suicide within the medical community has eroded. While one recent survey of hospice physicians and nurses found overwhelming opposition to assisted suicide, another 1992 survey reported that 21 percent of geriatric internists would consider assisting in the suicides of competent, nondepressed dementia patients. In a 1996 poll of Oregon psychiatrists' attitudes toward physician-assisted suicide, two thirds endorsed the view that a physician should be permitted, under some circumstances, to write a prescription for a medication whose sole purpose would be to allow patients to end their lives. The poll revealed that psychiatrists' moral beliefs influence how they might evaluate patients requesting assisted suicide and that their confidence in accurately assessing whether depression is impairing the judgment of a patient requesting assisted suicide was low. A 1997 poll of 118 physicians in San Francisco working with AIDS patients found that 53 percent of the physicians anonymously indicated that they had intentionally prescribed a deadly dose of narcotics to patients who wanted to die. A national survey published in 1998 polled over 1900 physicians and found that 18 percent had received requests for assistance with suicide and 11 percent had received requests for lethal injections. Furthermore, 3 percent of the physicians reported they had written at least one prescription to be used to hasten death, and 5 percent said they had administered at least one lethal injection. Finally, a 1998 telephone poll of oncologists found that 16 percent had participated in euthanasia or physician-assisted suicide. That poll also found that in only one third of cases do oncologists adhere to all of the main safeguards generally endorsed by proponents of euthanasia and physician-assisted suicide: Only 61 percent of patients both initiated and repeated their requests for aid in dying, and physicians sought consultation in 40 percent of cases (Table 28.5-3). About one quarter of the oncologists who performed euthanasia or assisted a patient with suicide regretted their participation. Consistent with their positions on active euthanasia, the American Medical Association, the American Psychiatric Association, and the American Bar Association continue to oppose physician-assisted suicide.

	Percentage adhering to safeguard
Patient request	
Patient initiated request	79
Patient repeated request	63
Patient initiated and repeated request	61
Patient distress	
Pain	84
Poor physical functioning	87
Pain or poor physical functioning	97
Second opinion	
Discussed with another physician	40
Evaluated by psychiatrist	5
Adhered to all 3 safeguards	34

Table 28.5-3 Adherence to Proposed Safeguards in the Practice of Euthanasia and Physician-Assisted Suicide Among Oncologists in the United States

Several court decisions are currently in various stages of review. In rejecting a constitutional right to physician-assisted suicide in 1997, the U.S. Supreme Court preserved the distinction between the withdrawal of life-sustaining treatments and assisted suicide or active euthanasia. At the same time, however, the Supreme Court may have blurred the distinction by endorsing terminal sedation. In 1994, Oregon narrowly passed a ballot initiative legalizing physician-assisted suicide (Table 28.5-4). A federal court judge subsequently struck down Oregon's Death with Dignity Law, in part because the law did not require evaluation by a mental health professional. That decision was later overturned by the Ninth Circuit Court of Appeals in February 1997, opening the way for the legalization of physician-assisted suicide in Oregon. Almost simultaneously, the Second Circuit Court of Appeals held that the New York law that bans assisted suicide is unconstitutional as applied to physicians and their dying patients. After an initiative to repeal the Oregon law was rejected by 60 percent of Oregon voters in November 1997 the law was enacted. Within a few months a 30-year-old grandmother in Portland, Oregon, became the first person disclosed to have died under this new suicide law. Shortly thereafter, Attorney General Janet Reno issued a statement that federal law does not authorize the prosecution of physicians who comply with the state law. More court's decisions may well shape the future direction of these important and hotly debated issues as the twenty-first century draws near.

- Oregon residents whose physicians determine they have less than six months to live are eligible to ask for suicide medication.
- A second doctor must determine if the patient is mentally competent to make the decision and is not suffering from mental illness such as depression.
- The law does not compel doctors to comply with patient requests for suicide medication.
- Doctors who agree to provide medication must receive a request in writing from the patient, signed by two witnesses. The written request must be made 48 hours before the doctor delivers the prescription. A second oral request is made just before the doctor writes the prescription.
- Pharmacists who are opposed to suicide may refuse to fill the prescriptions.
- The law does not specify which medication may be used. Supporters of the law say an overdose of barbiturates combined with antinausea medication would probably be used.

Table 28.5-4 Oregon's Assisted Suicide Law

To help guide clinicians facing requests for physician-assisted suicide, the vice president of the AMA's Institute for Ethics has proposed the following 8-step clinical protocol:

1. An evaluation of the patient for depression or other psychiatric conditions which could cause disordered thought;
2. An evaluation of the patient's “decision-making competence”;
3. A discussion with the patient about his or her goals for care;
4. Evaluation and response to the patient's “physical, mental, social, and spiritual suffering”;
5. Discussion with the patient about the full range of treatment and care options;
6. The consultation by the attending physician with other professional colleagues;
7. Assurance that care plans chosen by the patient are being followed, including removal of unwanted treatment and the provision of adequate pain and symptom relief; and
8. A discussion with the patient explaining why physician-assisted suicide is to be avoided and why it is not compatible with the principled nature of the care protocol.

Advances in technology bring more complex medical, legal, moral, and ethical controversies regarding life, death, euthanasia, and physician-assisted suicide. Some forms of euthanasia have found a place in modern medicine, and expansions in the boundaries of patient's rights and their ability to choose the way they live and die are inevitable. Both patients and physicians need to be better educated about depression, pain management, palliative care, and quality of life. Medical schools and residency training programs need to treat death, dying, and palliative care with the care and attention they deserve. Open debate and an honest exchange of viewpoints are needed. Society must ensure that economics, ageism, and racism do not get in the way of adequate and humane management of patients with a chronic terminal illness. Finally, national health care policy must provide adequate insurance, home care, and hospice services to all appropriate patients. If these mandates are followed, the argument for physician aid in dying will lose much of its impact.

BEREAVEMENT AND GRIEF

Definitions The terms bereavement and grief are not consistently applied. In 1982 the Institute of Medicine appointed a Committee on Health Consequences of the Stress of Bereavement, composed of multidisciplinary clinicians and researchers, to study bereavement factors and their impact on general and mental health. The definitions agreed upon by that committee include:

Bereavement—the fact of loss through death.
Bereavement reaction—any psychological, physiological, or behavioral response to bereavement.
Bereavement process—an umbrella term that refers to the emergence of bereavement reactions over time.
Grief—the feelings and associated behaviors, such as crying, accompanying the awareness of irrevocable loss (not necessarily, but including, loss through death).
Grieving process—the changing affective state over time.
Mourning—the social expressions of grief, including funerals, visitations, and rituals.

Normal Bereavement Reactions Sigmund Freud and other early psychoanalytic writers called the painful relinquishing of ties to the deceased the work of grief. Thus, grief was viewed as a restorative reaction to separation. Grieving is completed when the bereaved is free to invest emotional energy in new directions.

Influenced by attachment theory, recent theorists emphasize that the work of grief involves a series of attachment behaviors rather than simply separation. Attachment theory posits a human instinct to form strong, persistent affectional bonds. A natural response to the loss of an attachment bond is separation anxiety, which generates intense but predictable behavior geared to recoup or revive the lost relationship. Thus, the crying of a baby when its mother or father leaves is viewed as an attempt to retrieve the lost parent. In bereavement, these attempts fail to achieve their goal and attachment bonds must ultimately be realigned. The first response to loss, protest, is followed by a longer period of searching behavior. As hope to reestablish the attachment bond diminishes, searching behaviors give way to despair and detachment before bereaved persons eventually reorganize themselves around the recognition that the lost person will not return. While the bereaved ultimately learn to accept the reality of the death, they also find psychological and symbolic ways of keeping the memory of the deceased person very much alive. Grief work allows survivors to redefine their relationship to the deceased and to form new but enduring ties.

Grief and mourning may be best understood as attempts to solve the separation versus attachment dilemma. On the one hand, reality demands the bereaved person accept the loss in order to achieve a healthy adaptation to continued life in the real world—a life that is now missing a key person in a palpable way. On the other hand, psychological needs dictate a counterbalancing demand to maintain ties to the lost relationship, which although now technically illusory, are nevertheless quite real to the survivor. This dilemma requires working through for the survivor to regain emotional homeostasis, balancing the separation and attachment valences. The former demand—accepting the loss—is likely to generate intense anguish while the latter demand—maintaining the relationship—usually provides comfort. The typical course of bereavement includes some fluctuations between anguish and comfort, mediated by the shifting needs and capacities of the mourner.

Other Forms of Grief Much of the course, phenomenology, assessment, complications, and treatment of bereavement can be applied to other losses as well. Indeed, a growing body of literature on the psychological effects of divorce attests to the similarity between the grief associated with divorce and the grief associated with bereavement. Epidemiological studies suggest that divorce may be a greater risk factor for depression than widowhood. The loss of a pet may engender profound grief reactions in some persons. In medical settings the loss of youth, health, cognitive abilities, sensory acuity, body parts, relationships, roles, abilities, and potentials can and do trigger grief responses that often warrant clinical attention. A young woman facing a mastectomy, a middle-aged man about to retire, an old woman losing her eyesight—are all likely to grieve. In none of these instances does a set guideline prescribe what is considered normal or accepted grief. Rather, people cope with and respond to these losses in a variety of ways. An important component of understanding human behavior involves identifying the potential for persons to experience grief reactions to a multitude of everyday life events and appreciating the variability in the ways that those reactions are expressed.

Mourning From earliest history, every culture records its own beliefs, customs, and behaviors related to bereavement. Specific patterns include rituals for mourning (e.g., wakes or sitting shiva), for disposing of the body, for invocation of religious ceremonies, and for periodic official remembrances. The funeral is the prevailing public display of bereavement in contemporary North America. The funeral and burial service acknowledge the real and final nature of the death, countering denial; they also garner support for the bereaved, encourage tributes to the dead, unite families, and facilitate community expressions of sorrow. The burial is usually in a cemetery, which may serve as an ongoing serene focus for memories and tributes. If cremation replaces burial, ceremonies associated with dissemination of the ashes perform similar functions. Visits, prayers, and other ceremonies allow for continuing support, coming to terms with reality, remembering, emotional expression, and concluding unfinished business with the deceased. Several cultural and religious rituals provide purpose and meaning, protect the survivors from isolation and vulnerability, and set limits on grieving. Subsequent holidays, birthdays, and anniversaries serve to remind the living of the dead, and may elicit grief as real and fresh as the original experience; over time, these anniversary grievings become attenuated, but often remain in some form.

Phenomenology of Grief Bereavement reactions involve alterations in feeling states, coping strategies, interpersonal relationships, biopsychosocial functioning, self-esteem, and world view that may last indefinitely. Manifestations of grief reflect the individual's personality, previous life experiences, past psychological history, the significance of the loss, the nature of the bereaved's relationship with the deceased, the existing social network, intercurrent life events, health, and other resources. Despite individual variations in the bereavement process, investigators have proposed grieving process models, which include at least three partially overlapping phases or states: (1) initial shock, disbelief, and denial; (2) an intermediate period of acute discomfort and social withdrawal; and (3) a culminating period of restitution and reorganization. Like Kubler-Ross's stages of dying, the stages of grief do not prescribe a correct course of grief; rather, they are general guidelines that describe an overlapping and fluid process that varies with the survivors (Table 28.5-5).

Shock and denial (minutes, days, weeks)
Disbelief and numbness
Searching behaviors: pining, yearning, protest
Acute anguish (weeks, months)
Waves of somatic distress
Withdrawal
Preoccupation
Anger
Guilt
Lost patterns of conduct
Restless and agitated
Aimless and amotivational
Identification with the bereaved
Resolution (months, years)
Have grieved
Return to work
Resume old roles
Acquire new roles
Reexperience pleasure
Seek companionship and love of others

Table 28.5-5 Phases of Grief

In the *shock and denial* phase, disbelief and numbness predominate. The funeral, gathering of friends, and other mourning rites help survivors to accept the loss in a supportive environment. As numbness turns to intense feelings of separation, various searching behaviors (such as pining, yearning, and protest) take over.

The second phase, *acute anguish*, predominates once the searching proves fruitless. Erich Lindemann, in the first study of bereavement, described six components of acute grief: (1) Intense somatic distress, occurring in waves and lasting from 20 minutes to 1 hour, is manifested by a tight throat, choking and sighing, an empty feeling in the abdomen, weakness, tenseness, and mental pain. Because visits may provoke these somatic disturbances, withdrawal from friends, relatives, or others is common. (2) Thoughts of the deceased preoccupy the survivor. (3) Filled with guilt, survivors may accuse themselves of having mistreated or neglected the dead. (4) Irritation and anger are directed inward, or toward the deceased, friends, relatives, doctors, the world, or God. (5) Restlessness, agitation, aimlessness, and lack of motivation are accompanied by abandonment of usual habit patterns. (6) Identification phenomena, the adoption of traits and behaviors of the deceased (especially those of the final illness), which is often considered pathological, may occur. Acute anguish lasts weeks or months, gradually giving way to a return of well-being and the ability to go on.

In *restitution* (or *reorganization*), the bereaved recognizes the extent of the loss and that grieving has been accomplished. Attention shifts to life apart from the deceased. The hallmark of restitution is that survivors recognize that they can return to work, resume old roles, acquire new ones as necessary, experience pleasure, and seek companionship and love.

Duration of Grief Most societies mandate modes of bereavement and time periods for grieving. In contemporary America the bereaved is expected to return to work or school in a few weeks, to establish equilibrium within a few months, and be capable of pursuing new relationships within 6 months to a year. Ample evidence suggests that the bereavement process does not end within a prescribed interval; certain aspects persist indefinitely for many otherwise high-functioning, normal persons.

Perhaps the most lasting manifestation of grief, especially after spousal bereavement, is loneliness. Often present for years after the death of a spouse, for some loneliness may be a daily reminder of the loss. Other common manifestations of protracted grief occur intermittently. For example, a man who has lost his wife may experience elements of acute grief every time he hears her name or sees her picture on the nightstand. Usually these reactions become increasingly short-lived over time, dissipating within minutes, and may become tinged with positive and pleasant affects. Such bittersweet memories may last a lifetime. Thus, most grief does not fully resolve or permanently disappear; rather, grief becomes circumscribed and submerged only to reemerge in response to certain triggers.

Anticipatory Grief Grief may begin before the actual death. In *anticipatory grief*, grief reactions are brought on by the slow dying process of a loved one through injury, illness, or high-risk activity. Although anticipatory grief may soften the blow of the eventual death, it may also lead to premature separation and withdrawal while not necessarily mitigating later bereavement. At times the intensification of intimacy during this period may heighten the actual sense of loss even though it prepares the survivor in other ways.

Anniversary Reactions When the trigger for an acute grief reaction is a special occasion such as a holiday or birthday, the rekindled grief is called an *anniversary reaction*. It is not unusual for anniversary reactions to occur each year on the same day the person died, or in some cases, when the bereaved individual becomes the same age the deceased was at the time of death. Although these anniversary reactions tend to become relatively mild and brief over time, they can be experienced as the reliving of one's original grief and may prevail for hours or days.

Multidimensional Assessment of Bereavement and Grief Mourning phenomena are not sufficient to describe the full impact of losing a loved one to death. A multidimensional approach addresses seven features: (1) emotional and cognitive responses, (2) coping strategies, (3) continuing relationship with the deceased, (4)

changes in functioning, (5) alterations in existing relationships, (6) forming relationships, and (7) changes in identity. Assessing these features lends itself to rational management and treatment strategies.

The emotional and cognitive responses of shock, anguish, loss, anger, guilt, regret, anxiety, fear, intrusive images, depersonalization, feeling overwhelmed, loneliness, unhappiness, and depression may also be intermingled with positive feelings such as relief, joy, peace, or happiness, which emerge after the loss of a special person. To mitigate the anguish of acute grief, coping strategies help “dose” the pain a mourner can bear at any point. Strategies include suppression, intellectualization, rationalization, humor, avoidance, exposure, activity, involvement with others, passive distraction (e.g., immersion in television), or indulgence in food, alcohol, tobacco, or sex. Many persons say that their belief in God helps them cope and find meaning in death.

One powerful means of mitigation is the continuing relationship with the deceased. Faced with the dilemma of balancing reality's demands to accept life without their lost loved one and equally powerful inner forces dictating they maintain their attachments by retrieving what has been lost, the bereaved gradually learn to realign their attachment bonds. Survivors experience the transformation of a relationship that had heretofore operated on several levels of actual, symbolic, internalized, and imagined relatedness to one in which the actual (living and breathing) relationship has been lost—but other forms of the relationship remain and new ones may develop. Thus, it is not unusual for bereaved individuals to maintain continuing contact by dreaming of their deceased loved ones, looking for them in crowds, sensing their presence, feeling them watching out for or protecting them, reliving conversation or “speaking” with them, and even having auditory or visual hallucinations. In addition, symbolic representations (also known as linkage objects), such as the clothing, writings, favorite possessions or rings may be kept indefinitely, and living legacies such as identification phenomena, carrying out the deceased's mission, memorial donations, or seeing them live on in others through genetic endowments become other means of perpetuating the relationship. Periodically visiting the grave or lighting candles also keeps memories alive. Ultimately, memories become the most powerful means of continuing the relationship with the deceased. Bereaved individuals may get great comfort from hearing that the relationship has not been totally severed but that it is perfectly acceptable, and even normal for the relationship to endure indefinitely.

A fourth dimension of grief and bereavement, changes in functioning, involves an assessment of the ways general medical health, psychological health, social networks, and occupational functioning are affected. Grief may take its toll on any of these indices of biopsychosocial functioning. Even in the absence of severe medical or psychiatric complications, a diminished sense of well-being and vague somatic complaints are common. Sleeplessness is an enduring symptom of bereavement, lasting for over a year in a substantial percentage of bereaved individuals. Social inhibition or isolation occasionally occurs. Disability or impaired work performance and satisfaction are not uncommon. The challenges of taking on new roles or meeting new demands may appear overwhelming. For example, widows may need to learn for the first time how to service the car or maintain financial records and widowers may be challenged by cooking, shopping, housecleaning, or child-rearing tasks.

Bereavement may entail changes in ongoing relationships. Complex changes occur within the family. For example, when a spouse dies and there are grown children there may be conflicts in the expectations of the children and surviving parent over issues of emotional support, finances, decision-making, and future directions. Survivors may have to contend with the grief of their own parents, children, in-laws, and siblings and their efforts to enlist support for themselves. There is an opportunity for achieving greater intimacy, repairing old wounds and sharing grief; conversely, it may be a time of exacerbation, conflict, and disruption. Friends may be a major source of support for the bereaved, especially when empathy and sympathy are freely given and when they are accepting of the enormous fluctuations in the feelings, moods, and needs of the survivor. Friends can share the pain and allow its free expression. On the other hand, there are times when friends may feel threatened or overwhelmed by the intensity of the survivor's neediness and suffering.

A related dimension is the development of a new relationship, often a major concern yet a source of embarrassment and awkwardness. A bereaved spouse who attempts to begin a new romance may be challenged by continued devotion to the spouse, societal sanctions, fears of recurring loss, and perceived disloyalty of the children. Despite these challenges, many bereaved spouses, especially widowers, do remarry, often within a year of two of their bereavement and frequently are surprised at how well their new marriage works out. Remarriage has been found to be associated with well-being and a positive adaptation to bereavement. The strongest predictor for the remarriage of widows is age, with younger women being more likely than older widows to find a new mate. For widowers, income and education are most strongly associated with remarriage.

Finally, the multidimensional approach includes changes in identity. During initial grief, feelings of being overwhelmed and helpless may cause the bereaved to regress. Later, survivors may be surprised by their capacity to tolerate grief, carry on, and even find new approaches to life. Often there is an evolving sense of strength, autonomy, and independence. Having mastered acute grief, survivors experiences existential growth and may become more compassionate, patient, and balanced. In time, some survivors may transform their tragedies into new careers, relationships, and personal evolvement.

Bereavement Since bereavement often evokes depressive symptoms, the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) V code (no mental illness) of bereavement attempts to distinguish normal grief reactions from major depressive disorder (Table 28.5-6) According to DSM-IV, if the symptoms of a major depressive episode begin within 2 months of the loss of a loved one and do not persist beyond the 2 months, they are generally considered to result from bereavement unless they are associated with marked functional impairment or include morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms, or psychomotor retardation. DSM-IV recommends the following:

Bereavement	Major Depressive Disorder
Symptoms may meet syndromal criteria for major depressive episode, but survivor rarely has morbid feelings of guilt and worthlessness, suicidal ideation, or psychomotor retardation	Any symptoms as defined by DSM-IV
Dysphoria often triggered by thoughts or reminders of the deceased	Dysphoria often autonomous and independent of thoughts or reminders of the deceased
Onset is within the first 2 months of bereavement	Onset at any time
Duration of depressive symptoms is less than 2 months	Depression often becomes chronic, intermittent, or episodic
Functional impairment is transient and mild	Clinically significant distress or impairment
No family or past personal history of major depressive disorder	Family or past personal history of major depressive disorder

Table 28.5-6 Differentiating the Depressive Symptoms Associated with Bereavement from Major Depressive Disorder

> This category can be used when the focus of clinical attention is a reaction to the death of a loved one. As part of their reaction to the loss, some grieving individuals present with symptoms characteristic of a major depressive episode (for example, feelings of sadness and associated symptoms, such as insomnia, poor appetite and weight loss). The bereaved individual typically regards the depressed mood as normal, although the person may seek professional help for relief of associated symptoms such as insomnia or anorexia. The duration and expression of normal bereavement vary considerably among different cultural groups. The diagnosis of major depressive disorder is generally not given unless the symptoms are still present two months after the loss. However, the presence of certain symptoms that are not characteristic of a normal grief reaction may be helpful in differentiating bereavement from a major depressive episode. Those include (1) guilt about things other than actions taken or not taken by the survivor at the time of the death; (2) thoughts of death other than the survivor feeling that he or she would be better off dead or should have died with the deceased person; (3) morbid preoccupation with worthlessness; (4) marked psychomotor retardation; (5) prolonged and marked functional impairment; and (6) hallucinatory experiences other than thinking that he or she hears the voice of, or transiently sees the image of, the deceased person.

The depressive syndromes associated with bereavement are common; in many cases they are self-limited, require no active treatment, and are not associated with prolonged psychosocial impairment or reduction in life quality. However, clinicians must recognize that depressive syndromes, even in the context of bereavement, deserve careful attention and clinical evaluation. The DSM-IV designation of bereavement should not lull the clinician into a false sense of security or therapeutic nihilism. Research has shown that at least one third of persons who manifest a depressive syndrome within a few months of their loss continue to be depressed a full year later; further, their depressions often are associated with suicidal ideation, medical morbidity, and psychosocial disability. Thus, depressive syndromes, even if they can be understood in the context of bereavement, may herald a disabling and potentially life-threatening illness. When a depressive syndrome occurs early in the course of bereavement, the clinician should assess the severity, associated symptoms, persistence, pervasiveness, and the patient's past history when deciding whether to treat the syndrome. In general, the more severe the depressive symptomatology and the more the bereaved individual has a positive family or past personal history of major depressive disorder, the more likely it is that the early depressive syndromes of bereavement will persist. Effective treatment of depressive

syndromes facilitates the work of grief. The outworn notion that medications or psychotherapy impede the process of grief is unsubstantiated and may serve to prolong suffering and disability. Clearly, more research work needs to be done in this area.

Complicated Bereavement Several correlations between known risk factors and specific bereavement complications have been identified. These include multiple, sudden, or unexpected deaths, including suicide, murder, catastrophic circumstances, and stigmatized deaths (e.g., abortion, AIDS); preexisting personality in the bereaved marked by insecurity, low self-esteem, dependence, and characteristic difficulties coping with stress; a relationship with the deceased marked by ambivalence or undue dependency; loss of a child; poor health, depression, or prior substance abuse; multiple concurrent life stresses or other crises; and perceived lack of social support. Gender and age are controversial risk factors.

Complicated bereavement is described by a confusing array of terms: *abnormal, atypical, distorted, morbid*, or *unresolved*. Most concepts surrounding it can be viewed as either complicated processes of grieving, or as high-risk medical or psychiatric illnesses.

Complicated Processes Three patterns of complicated, dysfunctional grief syndromes have been identified: chronic, hypertrophic, and delayed grief. *Chronic grief*, the most common, is an unremitting grief that is often highlighted by bitterness and idealization of the dead. Chronic grief may be most likely to occur when the relationship between the bereaved and the deceased had been extremely close, ambivalent, or dependent, or when social supports are lacking and friends and relatives are not available to share the sorrow over the extended period of time needed for these mourners. In *hypertrophic grief*, seen most often after a sudden and unexpected death, bereavement reactions are extraordinarily intense. Customary coping strategies are ineffective in mitigating anxiety and withdrawal is frequent. When one family member is experiencing a hypertrophic grief reaction, disruption of family stability can occur. Hypertrophic grief frequently takes on a long-term course, but it may become attenuated over time. *Delayed grief* refers to absent or inhibited grief when one normally expects to find overt signs and symptoms of acute mourning. This pattern is marked by prolonged denial; anger and guilt may complicate its course. Although family members, friends, and clinicians are often alarmed when a bereaved person shows minimal overt symptoms of grief after losing a loved one, empirical evidence fails to support a relationship between inhibited grief and other negative outcomes.

Some investigators have used the *unresolved grief* to refer to arrests in the normal process of grief. Using operational criteria to define the arrests, one group of investigators found unresolved grief to be common, existing in 10 to 25 percent of the various populations studied; it is associated with depressive symptoms and syndromes, persistent somatic preoccupations and distress, pathological identification phenomena, and chronic mourning.

The more sudden, unexpected, and unnatural the death, the more likely the bereavement process is to overlap with traumatic stress reactions. Indeed, in the most traumatic situations, such as when an individual survives combat, natural disasters, or physical attacks while others succumb, both grief and posttraumatic stress disorders may co-occur. In other situations there may be more of a mingling of symptoms of grief and traumatic stress, resulting in what some have called *traumatic grief*. In such situations the grief tends to be both chronic and hypertrophic. Themes of violence, victimization, and volition (i.e., the choice of death over life, as in the case of suicide) are intermixed with other aspects of grief, and traumatic distress marked by fear, horror, vulnerability, and disintegration of cognitive assumptions may ensue. Disbelief, despair, anxiety symptoms, preoccupation with the deceased and the circumstances of the death, withdrawal, hyperarousal, and dysphoria are more intense as well as more prolonged than under nontraumatic circumstances, and there may be an increased risk for other complications. Deaths that are sudden and unanticipated but not necessarily unnatural are associated with grief reactions that involve greater levels of disbelief, anxiety, self-reproach, feelings of abandonment, and even suicidal ideation than grief following anticipated deaths. At least two groups of investigators recently have proposed operational criteria for complicated or traumatic grief. One of the groups also has found a strong association between traumatic grief and both physical and psychological morbidity, including substantial comorbidity with major depressive disorder with a fairly robust response to treatment with antidepressant medications.

High-Risk Medical or Psychiatric Illnesses Bereavement reactions also can be considered to be complicated if the medical or psychiatric outcomes are adverse. Medical complications include exacerbations of existing diseases and vulnerability to new ones; fear for one's health and more trips to the doctor; and an increased mortality rate, especially in men. Although early retrospective studies suggested that morbidity and mortality rates after bereavement were extremely high, more recent carefully controlled prospective studies reveal that such risks are not uniformly distributed. The highest relative mortality risk is found immediately after bereavement, particularly from ischemic heart disease. The greatest effect of bereavement on mortality is for men under age 65. Although greatest during the first year, the increased risk of mortality continues for many years unless the man remarries. Higher mortality rates in bereaved men than in bereaved women are due to increases in the relative risk of eath by suicide, accident, cardiovascular disease, and some infectious diseases. Evidence for increased mortality for women in the first year after bereavement is equivocal, but some evidence suggests increased mortality in the second year; in widows, the relative risk of death from cirrhosis and perhaps suicide increases. In both sexes bereavement appears to exacerbate health-compromising behaviors, such as alcohol consumption, smoking, and the use of over-the-counter medications.

Psychiatric complications of bereavement include increased risk for major depressive disorder, prolonged anxiety, panic, and a posttraumatic stress-like syndrome; increased consumption of alcohol and drugs; increased cigarette-smoking; and increased risk of suicide. The best studied of the potential pathological outcomes of bereavement is depression. Grief and depression are not synonymous, although symptoms and behaviors may overlap. Sleep disturbances and dysphoria are common to both, but not all bereaved persons manifest a full depressive syndrome. Studies show that about one third of all widows or widowers manifest a full depressive syndrome 1 month after the death of a spouse, approximately one fourth do so at 7 months, and approximately 15 percent at 1 and 2 years. Fifty percent will meet criteria for major depressive disorder during the first year; 10 percent may meet criteria for the entire year. Interestingly, antidepressant medications may ameliorate depression without necessarily affecting other aspects of grief.

The relationship between grief and depression is complex. One can grieve without being depressed, but loss may well precipitate depressive episodes in vulnerable individuals. Not only are bereaved persons at high risk for major depressive episodes, but they also appear to be at risk for lingering subsyndromal depressive symptoms. Such symptoms, even in the absence of full depressive disorders, may be associated with prolonged personal suffering, role dysfunction, and disability.

Until recently bereavement was excluded as a possible criterion for the diagnosis of posttraumatic stress disorder. However, DSM-IV allows the possibility for sudden and unexpected deaths of loved ones to precipitate posttraumatic stress disorder in vulnerable individuals. Indeed, a recent epidemiological study in Detroit found that news of a sudden and unanticipated death of a loved one accounted for 31 percent of all the posttraumatic stress disorders in that community. Others have found that even expected death following chronic illness can sometimes lead to posttraumatic stress disorder.

Because of their psychosocial, emotional, and cognitive immaturity, bereaved children may be especially vulnerable to psychopathology. Enduring psychological symptoms of anxiety, somatic conditions, and depression have been observed in community and patient samples of children who have lost a parent or a sibling, and some studies report a relation between childhood bereavement and later mental illness, especially depressive disorders.

Biological Perspectives Developments in bereavement biology have occurred on three major fronts: endocrinological responses, immunological changes, and sleep studies. Endocrinological studies suggest adrenocortical activation in response to bereavement, related to separation distress. Results of bereavement studies using the dexamethasone suppression test to measure adrenocortical responses have been equivocal, although one such study suggested a relationship between nonsuppression and separation anxiety. Similarly, the sudden loss of a partner, which evokes separation reactions, has been associated with increases in adrenal function and cortical secretion as well as elevated serotonin levels in the hypothalamus of squirrels and rhesus monkeys.

A variety of life stress events, including bereavement, have been found to alter humoral and cell-mediated immunity in humans and animals. Spousal bereavement has been associated with inhibited lymphocyte stimulation responses to several mitogens during the first several weeks after the death, and with impaired natural killer cell activity in bereaved widows with dysphoria. At least one monkey study also has demonstrated inhibited lymphocyte stimulation to various mitogens in infant monkeys after separation.

Sleep disturbances are among the most frequent and persistent bereavement-associated symptoms. Even if bereavement has not disrupted sleep per se, sleep electroencephalographic changes consistent with depression, such as decreased rapid-eye movement (REM) latency and deep sleep and increased total REM sleep, have been found in bereaved, depressed widows. Circadian rhythms and sleep pattern disturbances in infant monkeys also have been noted in response to separation.

Separation studies in primates may shed light on human grief and bereavement as well as on the pathogenesis of depression. Separation of infant monkeys from their mothers is highly traumatic, eliciting responses akin to protest followed by despair. The despair response can be attenuated if substitute caretakers are available. Further, short-term separations of this kind may have long-term effects and may predispose the infant monkeys to severe behavioral reactions to future separations. The protest-despair response is similar regardless of the cause of the separation. Developmental, neurochemical, and pharmacological variables can alter the

response. For example, if monkeys had been raised by their mothers in the first several months of life prior to life in peer groups, the response to later peer separation is less intense. Monkeys with lower concentrations of norepinephrine in cerebrospinal fluid responded to separation with more huddling and self-directed behaviors than those with higher concentrations. Imipramine (Tofranil) reverses peer separation reactions and prevents future reactions when the drug is maintained. Drugs that lower norepinephrine concentrations increase separation response severity whereas serotonin-blocking drugs have no effect, suggesting that the separation response is sensitive to catecholamine depletion.

LIFE CYCLE PERSPECTIVES

Bereavement During Childhood and Adolescence About 4 percent of North American children lose one or both parents by age 15; sibling death is the second most commonly experienced bereavement. Grief reactions are colored by development levels and concepts of death, and may not resemble adult reactions. Children may display minimal grief at the time of death and experience the full effect of the loss later. Indifference, anger, or misbehavior may be displayed rather than sadness; behaviors may be erratic and labile. Even older children frequently feel abandoned or rejected when a parent dies, and may show hostility toward the deceased or the surviving parent, now perceived as one who might also “abandon” them. They may feel responsible for the parent's death because of their own past misbehavior or because they said or wished that the parent would die at some time. Psychological tasks particularly relevant to bereaved children and adolescents have to do with their capacity to accept the tragedy of death, their capacity to tolerate the anguish that ordinarily accompanies grief, and, in the case of parental bereavement, their developmental need for the deceased parent in order to complete the goals of healthy maturation.

Children under the age of 2 may show loss of speech or diffuse distress. Children under the age of 5 may respond by developing eating, sleeping, bowel, and bladder dysfunctions. Strong feelings of sadness, fear, and anxiety can occur, but these feelings are not persistent and tend to alternate between longer-lasting “normal” states. School-aged children may become phobic or hypochondriacal, withdrawn, or pseudomature, and school performance and peer relations often suffer. Like adults, adolescents run the gamut of bereavement reactions, ranging from behavioral problems, somatic symptoms, erratic moods, to stoicism. Adolescent boys losing a parent may become deliquent; girls may turn to a sexual partner for comfort and reassurance. Behavioral disturbances and depressions are common at all ages. Although the surviving parent may underestimate their child's dysphoria, interviews of children and adolescents themselves reveal rates of depressive episodes that are at least as high as those of bereaved adults. Although several investigators suggest that bereaved children and adolescents have long-term vulnerability to a variety of physical and mental illness, the link between childhood bereavement and subsequent depressive disorders is strongest.

Bereaved children must be treated according to their levels of emotional and cognitive maturity. They need to be told that the death is real and irreversible and that they are blameless. Fairy tales and insincere theological explanations add to children's confusion and may reinforce notions that they have done something wrong. Feelings and concerns should be expressed, and questions should be invited and answered with simplicity, candor, and clarity. Children, like adults, need rituals to commemorate their loved ones; attending the funeral and participating in mourning may be beneficial first steps.

Bereavement During Adulthood There is no consensus on which type of loss is associated with the most severe reactions. Although the death of a spouse is often ranked as the most stressful life event, some have argued that losing a child is even more profound. The death of parents and siblings in adult life has not been systematically studied but is generally considered relatively mild compared to the loss of a spouse or child. The impact of therapeutic and spontaneous abortion generally has been underestimated—prenatal bonding may be intense but social support, mourning rituals, and other aspects of socially sanctioned grief have not been established for this loss. Thus, the bereaved parents may not have the opportunity to vent and share their misery. Grief appears most intense for the mother who suffers a late perinatal losses (stillbirths or neonatal deaths rather than miscarriages), and often is re-experienced during subsequent pregnancies. *Sudden infant death syndrome* is particularly problematic because the death is sudden and unexpected; parents may experience extra guilt or blame each other, often resulting in subsequent marital difficulties.

The surviving family members, friends, or lovers of individuals who have died from AIDS are uniquely challenged. AIDS carries with it the stigmata of the illness itself and of the gay community in general, caretakers' fear of contracting the illness; also, it is most prevalent in people who are in the prime of life. Asymptomatic infection may provide infected persons and those close to them time to adapt to the diagnosis. But when an HIV-positive person begins to manifest symptoms of opportunistic infection or associated cancer, the illness again becomes a threat and coping with its emotional reality is arduous and complex. Often caretakers as well as HIV-positive patients wish for death, which can evoke feelings of guilt. For bereaved lovers, their own HIV status, multiple losses, and other concurrent stressors may complicate recovery. Gay men who have lost lovers to AIDS may be more depressed, consider suicide more often, and be more vulnerable to illicit drug use than other bereaved individuals.

Despite stereotypes and popular mythology, the elderly do not have any more difficulty coping with the deaths of loved ones than their younger counterparts do. Indeed, their grief is often described as less intense and prolonged, and less likely to be associated with increased drinking, health concerns, depression, or anxiety. Yet seniors face more losses than individuals at other phases of the life cycle, and intense loneliness may be a lasting memorial to those who have died. Most studies of late-age bereavement have been done on community samples, thus bypassing the most frail senior citizens who live in nursing homes, tertiary care centers or those who are too physically or cognitively impaired to complete questionnaires. It stands to reason that for highly impaired seniors who lose a spouse they depended on for daily functions or who was their sole source of companionship, bereavement reactions may be profound.

Caring for the Bereaved Most bereaved individuals do fine without treatment. The support, reassurance, and information they require is usually provided by family and friends and sometimes by clergy, so that other professional help is unnecessary. When such support is not available, mutual support groups may help fill the gap. It is easy to underestimate the power most persons have to cope with distress and master adversity. However, there are times when formal intervention may facilitate the grieving process, prevent future problems, or treat ongoing complications. Various interventions are available. These include self-help approaches, such as the Widowed to Widowed model, They Help Each Other Spiritually (THEOS), or Compassionate Friends; group interactions, both mutual support types and professionally run groups; a variety of supportive, psychodynamic, short-term and grief-specific interventions; and medication. The specific indications for any one type are not always clear, and no one form of intervention has been proven more beneficial than others.

In general, for bereaved persons with moderate distress who are experiencing grief-process difficulties, a mutual-support group is often all that is necessary. It is efficient, cost-effective, and clinically effective. Such groups provide members with personal contact, information, an opportunity to share coping techniques, a sense of universality and belonging, increased self-worth, reinforcement for positive change, and an opportunity to help others. Several studies have confirmed the benefits of such mutual support or self-help programs.

Although natural, restorative healing processes help most people get through their grief without professional assistance, sometimes more formal psychotherapeutic interventions are warranted. This is particularly true for high-risk individuals, such as people who have a history of mental illness; who have lost someone to suicide or homicide and who are experiencing traumatic grief; who have experienced multiple simultaneous losses or adverse life events; who are bereaved as the result of an accident in which the survivor may have been at fault; or persons whose poor health, lack of social supports, or life stresses have overwhelmed their capacity to cope. Most advocated psychotherapeutic approaches are relatively brief and time-limited and have been shown to be beneficial. Certain common features can be identified: First, most therapies include an educational component, helping the bereaved know what to expect, offering perspective on their pain, and normalizing their confusing array of feelings and behaviors. Second, therapies encourage the expression of the full range of bereavement emotions and affects. Evocative techniques, such as role playing, writing letters to the deceased, looking at pictures, or even visiting the burial site can be effective. Third, most therapies attempt to help the bereaved come to peace with their new relationship to the deceased. This usually involves systematic exploration of all aspects of the relationship, both in the past over time and as it has developed since the death. Fourth, most therapies attend to the bereaved person's new identity, focusing on an integrated self-concept and a stable world view. The roles of the psychotherapist in the bereaved person's evolving identity are numerous. Therapists must recognize early regression for what it is, and they must be able to maintain a conviction about the bereaved person's adaptive capacities. Contained in such a conviction are both the belief and knowledge that the bereaved person can survive and will feel better. As survivors venture out to try something new, they may find themselves stymied by old conflicts that can be examined in a psychotherapeutic fashion. The therapist plays a powerful role in supporting the bereaved person's attempts to try out new behaviors, explore new relationships, expand attitudes, and grow through the experience.

When grief is complicated by a major depressive episode, psychotherapeutic techniques to treat major depressive disorder, such as interpersonal or cognitive-behavioral therapy, should be considered. An integrative approach that combines elements of psychoeducation, support, grief-specific interventions, and aspects of interpersonal and cognitive-behavioral therapy as appropriate for the person at any given time, are recommended. If the depression is mild, this may be all that is necessary. In more moderate to severe depressions, or when psychotherapy has not worked, psychopharmacological intervention should be considered. Medications are particularly indicated when there is a past history of severe depression, suicidal ideation, psychomotor retardation, morbid feelings of worthlessness or guilt, or when psychosocial impairment is substantial. When making the decision whether or how to treat a major depressive disorder associated with bereavement, time since death is only one of many considerations; equally important are such factors as past and family history of depression, severity of depressive symptoms, suicidality, general medical health, and functional impairment. Treatment does not impede the grieving process; rather, effective treatment facilitates adaptive

processes and prevents the distortion and coping interference brought on by depression.

Similarly, when grief is complicated by posttraumatic stress disorder, both conditions need to be addressed. In general, treatment of the traumatic distress takes precedence over the grief. Treating anxiety or sleeplessness pharmacologically is controversial. Prescriptions and over-the-counter preparations used by bereaved persons reflect their efforts to relieve suffering and limit dysfunction. However, no controlled studies have established the efficacy and safety of antianxiety or hypnotic agents, nor is there evidence to support claims that medication will suppress the bereavement process or interfere with an adaptive outcome. When grief-related symptoms of anxiety are expressed so continuously that they interfere with cognitive and other functions of living in a substantial way, anxiolytic medications and psychotherapeutic interventions should be considered. For persons who develop persistent sleep disorders, interventions with hypnotic agents can be both humane and beneficial. It is rare for the judicious use of anxiolytic or hypnotic agents in bereaved persons to lead to problems of drug abuse. However, when confronted with a bereaved person who complains of anxiety or poor sleep, the clinician should ascertain whether these symptoms are part of an occult depressive episode before treating the symptom.

SUGGESTED CROSS-REFERENCES

Related themes of loss are discussed in Chapter 26 on divorce and other relationational problems. Attachment theory is discussed in Section 44.1 on reactive attachment disorders of infancy or early childhood. Section 25.9 covers bereavement and loss within the broader context of stress theory. Dealing with potentially fatal or terminal-phase-illness is covered in Section 25.4 on cardiovascular disorders, Section 25.11 on psycho-oncology, and Section 2.8 on the neuropsychiatric aspects of AIDS. Uncomplicated bereavement is discussed in the differential diagnosis of major depression in Chapter 14.6 and again in Chapter 45 on childhood depression and suicide.

SECTION REFERENCES

Angell M: The Supreme Court and physician-assisted suicide—the ultimate right. N Engl J Med *336:*50, 1997.

Bowlby J: Processes of mourning. Int J Psychoanal *42:*317, 1961.

Berman CM, Rasmussen KL, Suomi SJ: Responses of free-ranging rhesus monkeys to a natural form of social separation. Child Dev *65:*1028, 1994.

Breibart W, Jacobsen PB: Psychiatric symptom management in terminal care. Clin Geriatr Med *12:*329, 1996.

Burnell: *Final Choices*. New York Insight Books, New York, 1993.

*Cassem NH: The dying patient. In *The Handbook of General Hospital Psychiatry*, Cassem NH, editor. St. Louis, Mosby, 1991.

Chochinov HM, Wilson KG, Enns M: Desire for death in the terminally ill. Am J Psychiatry *152:*1185, 1995.

Cleiren MPHD: *Adaptation After Bereavement*. Leiden University DSWO Press, 1991.

Emanuel EJ, Daniels ER, Fairclough DL, Clarridge BR: The practice of euthanasia and physician-assisted suicide in the United States: Adherence to proposed safeguards and effects on physicians. JAMA *6:*507, 1998.

Emanuel EJ, Emanuel LL: The promise of a good death. Lancet *2*(Suppl):S21, 1998.

Engel G: Is grief a disease? Psychosom Med *23:*18, 1961.

Foley KM: Competent care for the dying instead of physician-assisted suicide. N Engl J Med *336:*54, 1997.

Freud S: Mourning and melancholia. In *Standard Edition of the Complete Psychological Works*, vol 14. Hogarth Press, London, 1953–1966.

Ganzini MD, Fenn DS, Lee MA, Heintz RT, Bloom JD: Attitudes of Oregon psychiatrists toward physician-assisted suicide. Am J Psychiatry *153:*1469, 1996.

*Goodkin K, Blaney NT, Feaster DJ, Baldewicz T, Burkhalter JE, Leeds B: A randomized controlled clinical trial of a bereavement support group intervention in human immunodeficiency virus type 1-seropositive and -seronegative homosexual men. Arch Gen Psych *56:*52, 1999.

Hendin HS: Physician-assisted suicide and euthanasia: The human dimension of legalization. J Long Term Home Health Care *2:*29, 1998.

*Hendin H: *Seduced By Death: Doctors, Patients, and the Dutch Cure*. WW Norton, New York, 1997.

Holland JC: Cancer's psychological challenges. Sci Am *275:*158, 1996.

Humphry D, Wickett A: *The Right to Die*. Harper & Row, New York, 1986.

*Jacobs S: *Pathologic Grief*. American Psychiatric Press, Washington, DC, 1993.

*Kastenbaum RJ: *Death, Society and Human Experience*, ed 2. Macmillan, New York, 1991.

Kübler-Ross E: *On Death and Dying*. Macmillan, London, 1969.

Lantz PM, House JS, Lepkowski JM, Williams DR, Mero RP, Chen J: Socioeconomic factors, health behaviors, and mortality: Results from a nationally representative prospective study of U.S. adults. JAMA *21:*1703, 1998.

*Lindemann E: Symptomatology and management of acute grief. Am J Psychiatry *101:*141, 1944.

Meier DE, Emmons CA, Wallenstein S, Quill T, Morrison RS, Cassel CK: A national survey of physician-assisted suicide and euthanasia in the United States. N Engl J Med *17:*1193, 1998.

Middleton W, Raphael B, Burnett P, Martinek N: A longitudinal study comparing bereavement phenomena in recently bereaved spouses, adult children and parents. Aust NZ J Psychiatry *2:*235, 1998.

Muller MT, Van der Wal G, van Eijk JTM, Ribbe MW: Voluntary active euthanasia and physician-assisted suicide in Dutch nursing homes: Are the requirements for prudent practice properly met? J Am Geriatr Soc *42:*624, 1994.

O'Gorman SM: Death and dying in contemporary society: An evaluation of current attitudes and the rituals associated with death and dying and their relevance to recent understandings of health and healing. J Adv Nursing *6:*1127, 1998.

Osterweis M, Solomon F, Green M, editors: *Bereavement: Reactions, Consequences, and Care*. National Academy Press, Washington, DC, 1984.

Parkes CM: Bereavement in adult life. BMJ *7134:*856, 1998.

Parkes CM, Weiss RS: *Recovery from Bereavement*. Basic Books, New York, 1983.

Permut SR, Physician assisted suicide: Legal considerations. Hosp Med *4:*45, 1998.

Prigerson HG, Frank E, Kasl SV, Reynolds CF, Anderson B, Zubenko GS, Houck PR, George CJ, Kupfer DJ: Complicated grief and bereavement-related depression as distinct disorders: Preliminary empirical validation in elderly bereaved spouses. Am J Psychiatry *152:*22, 1995.

Raphael B: *The Anatomy of Bereavement*. Basic Books, New York, 1983.

*Reynolds CF III, Miller MD, Pasternak RE, Frank E, Perel JM, Cornes C, Houck PR, Mazumdar S, Dew MA, Kupfer DJ: Treatment of bereavement-related major depressive episodes in later life: A controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry *56:*202, 1999.

*Sanders CM: *Grief: The Mourning After Dealing With Adult Bereavement*, ed 3. Wiley & Sons, New York, 1999.

Shuchter S: *Dimensions of Grief: Adjusting to the Death of a Spouse*. Jossey-Bass, San Francisco, 1986.

Slome LR, Mitchell TF, Charlebois MPH, Benevedes JM, Abrams DI: Physician-assisted suicide and patients with human immunodeficiency virus disease. N Engl J Med *336:*417, 1997.

Stroebe MS, Stroebe W, Hansson RO, editors: *Handbook of Bereavement: Theory, Research and Intervention.* Cambridge University Press, New York, 1993.

Sullivan MD, Youngner SJ: Depression, competence, and the right to refuse lifesaving medical treatment. Am J Psychiatry *151:*7, 1994.

Van Der Maas PJ, Van Delden JJM, Pijnenborg L, Looman CWN: Euthanasia and other medical decisions concerning the end of life. Lancet *338:*669, 1991.

Weisman AD: *On Dying and Denying: A Psychiatric Study of Terminality*. Behavioral Publications, New York, 1972.

Worden JW, editor: Dealing with grief. Ses Psychother Pract *2*(Suppl):1, 1996.

*Yates P, Stetz KM: Families' awareness of and response to dying. Oncol Nurs Forum *26:*113, 1999.

Zisook S, Shuchter SR: Major depression associated with widowhood. Am Assoc Geriatr Psychiatry *1:*316, 1993.

28.6 CHRONIC PAIN AND THE PLACEBO EFFECT

JAMES C. EDMONDSON, M.D., PH.D.

Pain is an intrusive, unpleasant, sensory, and emotional experience. In most people's minds, pain is a reliable indicator of injury, and everyone has had experience with acute pain. The details of diagnosis and management of acute tissue injuries have been thoroughly studied for centuries by the spectrum of medical and surgical specialists. In most cases, acute pain stops when the tissues heal. However, when pain persists, its connection to actual tissue injury frequently becomes more tenuous, and psychological factors may begin to predominate.

In clinical practice chronic pain has two divergent sources, one resembling the acute pain experience and the other of more obscure origin. The first group contains patients with medical conditions that cause chronic tissue injury. Ongoing pathology may continuously trigger the acute pain experience, but because the tissues do not heal, the pain remains a daily annoyance for an extended time. Many patients with such conditions (Table 28.6-1) accept their plight and maximize their coping abilities; others may become devastated by their suffering and lose their jobs, family, and social supports. In the United States at least 30 million patients suffer from medical conditions associated with chronic pain.

Arthritis	Muscle strain
Osteoarthritis	Persistent overexertion
Rheumatoid arthritis	Inflammatory bowel disease
Cancer	Crohn's disease
Any histological type	Ulcerative colitis
Neuropathy	HIV
Diffuse	Sickle cell anemia
Focal	Endometriosis
Multifocal	Movement disorders
Brain injury	Parkinson's disease
Traumatic brain injury	Generalized dystonia
Thalamic stroke	Focal dystonia
Limb amputation	Huntington's disease
Phantom limb pain	Multiple sclerosis

Table 28.6-1 Medical Conditions Associated With Chronic Pain

In the second group of patients the acute pain experience persists beyond a reasonable time expected for tissue healing after an acute injury, thereby becoming chronic. These patients persistently perceive pain, with its associated affective distress, in the absence of demonstrable tissue pathology. This perplexing group of patients, by some estimates numbering another 40 million Americans, is not readily classified as to etiology or personality type.

Unlike the well-established body of knowledge for illnesses based on tissue pathology, classification and treatment of patients with chronic pain in the absence of obvious tissue pathology is a relatively young field. Because the condition defies medical logic, many such patients encounter a negative bias from caregivers, who may not acknowledge the patients' distress or who may suggest derisively that “it's all in your head.” Pain without injury, like the placebo effect, is discomforting for professionals taught to trust their senses. Nevertheless, caregivers must appreciate that their own smug confidence in appearances may prevent them from fully comprehending the cognitive and emotional limits of the pain experience. In fact, chronic pain without injury—like the placebo effect (healing without medication)—transcends evidence-based medicine and enters the realms of philosophy and politics. Although it is poorly addressed in traditional medical education, every practitioner will frequently encounter patients with chronic pain. In this chapter, the psychosocial and neuroscientific aspects of chronic pain are discussed.

History Aristotle considered pain to be an emotional experience distinct from the five primary senses, because it did not reflect a specific external stimulus. This view predominated until the seventeenth century when Rene Descartes distinguished the body from the mind and assigned a separate role to each in the perception of pain. Descartes likened the nerves to wires that took a tiny bit of a noxious stimulus and transmitted it faithfully to the brain, where the mind then examined, assessed, and reacted to it. The Cartesian dualistic view, which is highly intuitive, remains the view held by most people. However, Descartes' theory could not explain two paradoxical observations: severe tissue injury could occur without pain and pain could occur without injury. The first of these conundrums was addressed in the gate control theory of pain, introduced in 1965 by Patrick Wall and Ronald Melzack. The gate control theory took into account not only ascending signals, but also descending projections that determined which pain signals ascended to consciousness. In this theory, tissue injury without pain occurred because descending signals “closed the gate” to the perception of pain. Although the gate control theory explained injury without pain, it could not explain the perception of pain without injury. For example, amputees appeared to experience complex patterns of pain in portions of the limb distal to the site of amputation (i.e., in the complete absence of any peripheral tissue). This and other observations focused attention on the brain's internal representation of pain and how the brain attends to or ignores pain signals.

Sensory Experience Sensory systems create an internal representation of the external world. Projections of visual, sonic, and tactile input are organized into maps in subcortical nuclei and in patches of the cortex. Specifically, visual images are arranged visuotopically, sounds are represented tonotopically, and tactile sensations are faithfully displayed according to the body plan. For example, the somatotopic configuration of tactile sensations in the postcentral gyrus is called a homunculus (from the Latin for “little man”). The cortical representational code for olfactory and gustatory input, however, has not yet been deciphered. Relevant features are extracted from these primary sensory representations by secondary sensory areas and are compared with memories. Thus, at any one time a potentially huge volume of data about various sensory experiences can contribute to conscious perception. However, people can rarely attend to more than one or a few details at a time. People consciously perceive a meager image of the world and operate within a narrow range of available experience. How the brain determines which data reach consciousness and which are ignored is not well understood; yet the boundary zone between conscious and unconscious attention holds many clues to the experience of pain without injury.

A window into the activation of different brain regions during specific sensory operations is being opened by the techniques of functional neuroimaging. These methods permit identification in real time of the brain regions that are active during mental proceedings. They add the dimension of time to high-resolution neuroimaging. These powerful new techniques promise to help define the neuroanatomical pathways for learning and memory during the experience and recollection of pain. However, examination of brain activity, only some of which is conscious, may yield conclusions that run counter to common intuition. For example, patients with low back pain without an obvious pathological cause will likely be studied in the future by functional neuroimaging. If disabling low back pain is found to originate entirely from memory stores within the brain, without activation of pain fibers from the periphery, then the person experiencing the pain must indeed grapple with the counterintuitive notion that “it is all in the head.” The coming era promises to enthrall the theoreticians of pain and to raise profound questions about the reliability of conscious perception.

Clinical Aspects In current clinical practice, a myriad of pharmacological, exercise-based, cognitive, behavioral, and interventional treatments is available for chronic pain. The premises for these run the gamut from the conviction that nearly all pain faithfully represents some tissue damage to the notion that 95 percent of chronic pain results from psychological factors. Where the truth lies on this spectrum certainly differs for each patient. The placebo effect, frequently considered a nuisance to be explained away, in fact, likely underlies the success of many currently used treatments for chronic pain. Despite the prominent role of psychological factors in the

maintenance and remission of chronic pain, efforts to define a pain-prone personality or to devise a test battery that consistently predicts outcome have been largely disappointing. The most effective approach is a detailed, highly individualized assessment of physical, psychological, and supportive factors, and the steady application of a multifaceted treatment plan agreed upon by both the patient and the team of caregivers.

STIMULI THAT ELICIT ACUTE PAIN

Acute pain may be caused by a variety of stimuli. Tissue damage leads to the release of cellular components that bind to and activate nerve endings. Extremes of tactile pressure, heat, cold, or in fact any sensory input, if intense enough, may cause pain. Tissues heal themselves once the painful stimulus is removed, and the acute pain resolves simultaneously. Numerous medical conditions may cause persistent tissue damage and are thus associated with chronic pain (Table 28.6-1). Patients with these conditions deserve an exhaustive medical evaluation and prompt treatment of their conditions. Many of these conditions have a genetic contribution, a factor for which there are virtually no diagnostic tests at this time. Therefore, the statement that a person's complaints of pain are not supported by an objective assessment of disease or disability must be interpreted in light of the limitations of current diagnostic methods.

The pain receptors of different tissues respond to specific stimuli. Skin receptors respond to pricking, cutting, crushing, burning, and freezing. The visceral structures instead respond to mucosal engorgement or inflammation, smooth muscle spasm, or traction on the mesenteric attachment. Skeletal muscle pain receptors respond to ischemia, injuries to the connective tissue sheath, necrosis, hemorrhage, and injection of irritating solutions. Prolonged contraction of skeletal muscle causes an aching pain. Joints do not respond to pricking, cutting, or cautery but do respond to inflammation and hypertonic saline. Arterial pain receptors respond to piercing and inflammation.

NEUROBIOLOGY OF SOMATIC PAIN PERCEPTION

Skin and Peripheral Nervous System

Anatomy and Physiology External stimuli elicit nerve impulses at the peripheral nerve endings. The cell bodies of sensory fibers are located in the dorsal root ganglia. Despite the expectation that specific endings (i.e., free nerve endings, Meissner corpuscles, Pacinian corpuscles, and Ruffini endings) are associated uniquely with specific modalities of somatosensory perception (i.e., touch, pain, temperature, vibration, and pressure), studies have failed to confirm a clear correlation between histological morphology and function. For example, the cornea, which has only free nerve endings, can perceive all somatosensory modalities.

Cutaneous stimuli are classified as thermal, mechanical, and chemical. Individual nociceptive nerve endings may respond to any combination of sensitivities; most commonly they respond to all three classes. Thermal nociceptors are triggered by skin temperatures in excess of 45°C (113°F), a level of heating that elicits subjective pain. Mechanical nociceptors respond to as little as 10 micronewtons of force, which is too little force to elicit pain. Chemical nociceptors respond to bradykinin, serotonin, histamine, substance P, protons, and adenosine triphosphate (ATP), among other endogenous molecules.

Pain is transmitted from the skin to the spinal cord with a high degree of spatial resolution by both very fine, unmyelinated, slowly conducting C fibers and thinly myelinated A-delta fibers. Nonnociceptive (nonpainful) stimuli are transmitted by large, thickly myelinated, rapidly conducting A-beta fibers. Table 28.6-2 lists the fiber sizes and nerve conduction velocities of the three types of fibers. The cell bodies of the sensory neurons form the dorsal root ganglia. The diversity of dorsal root ganglion cell types, as defined immunocytochemically according to neurotransmitters and receptors, significantly exceeds the three defined sizes of nerve fibers.

Fiber Type	Fiber Diameter (μm)	Nerve Conduction Velocity (m/s)	Nociception
A-beta	6–20	30–80	No
A-delta	1–5	12–30	Yes
C	0.4–1	0.5–2	Yes

Table 28.6-2 Three Types of Sensory Nerve Fibers

Two classes of C fibers are distinguished by biochemical and anatomical characteristics. Whereas both fiber classes require nerve growth factor (NGF) during development, receptors for NGF persist into adulthood on only about half of the fibers. These NGF-responsive fibers contain substance P and calcitonin gene–related peptide (CGRP) and terminate in Rexed's lamina I and the outer part of lamina II of the dorsal horn of the spinal cord. Fibers that lose NGF responsivity acquire sensitivity to glial-derived neurotrophic factor (GDNF), do not contain substance P or CGRP, and project to the inner part of lamina II of the dorsal horn.

Chemistry Tissue damage causes the release of proteolytic enzymes, which liberate agonists for peripheral nociceptors, including histamine, prostaglandins, serotonin, bradykinin, substance P, protons, potassium ions, and ATP. These substances elicit pain and cause local vasodilation when injected into the skin. Substance P and CGRP are released into the skin from C-fiber terminals on stimulation of nociceptors, where they produce neurogenic inflammation: the dilation of cutaneous vessels and the release of histamine.

THERMAL SENSITIVITY Capsaicin, a product of *Capsicum* peppers, elicits the subjective sensation of heat. It has been used to identify a receptor, VR1, which contains a nonselective cation channel with high permeability for calcium ions. Sudden increases in temperature activate the VR1 ion channel. VR1 appears to be one member of a family of heat-sensitive ion channels, which are segregated in distinct neuronal populations.

MECHANICAL SENSITIVITY Little is known about the molecules that respond to mechanical stimuli in humans. In the nematode *Caenorhabditis elegans*, The *mec* gene family encodes mechanoreceptors, which have structural homology to sodium channels. Whether such a mechanism for transforming movement into electrical activity exists in humans is not known.

CHEMICAL SENSITIVITY The chemical responsivity of nociceptors is determined by the presence of specific receptor molecules (e.g., receptors for serotonin, histamine, bradykinin). Evidence for ATP as a nociceptive stimulus has appeared, including the identification of an ATP receptor, P2X3, which is highly abundant only in GDNF-responsive, non-peptide-containing C fibers.

Exposure to chemicals or infectious agents activates the immune system and triggers the release of cytokines that stimulate subdiaphragmatic vagal afferents. Activation of these fibers causes generalized heightened sensitivity to pain, which is experienced as “hurting all over.”

Spinal Cord

Anatomy and Physiology

INPUT In the spinal cord, pain fibers enter in the dorsolateral quadrant in a superficial layer called the dorsal root entry zone. Pain fibers tend to enter in the lateral part of the dorsal root entry zone, and many of the tiniest fibers ascend in Lissauer's tract. Most A-delta and C afferent pain fibers cross Lissauer's tract and enter the gray matter of the dorsal horn. Most A-delta fibers terminate in the most superficial layer of gray matter, Rexed's lamina I, and the remainder terminate in the middle layers of the dorsal horn, in lamina V. C fibers, as discussed above, terminate in laminae I and II. Lamina II is also called the substantia gelatinosa (Fig. 28.6-1). A smaller number of fibers decussate and terminate in the contralateral dorsal horn. Functional sensory units consisting of overlapping patches of skin are called dermatomes, and their corresponding sensory nerves with peripheral and central extensions emanate from the dorsal root ganglion and project to the corresponding

spinal segments, of which there are 31. Many fiber termini activate local somatic and autonomic reflexes entirely within the spinal cord.

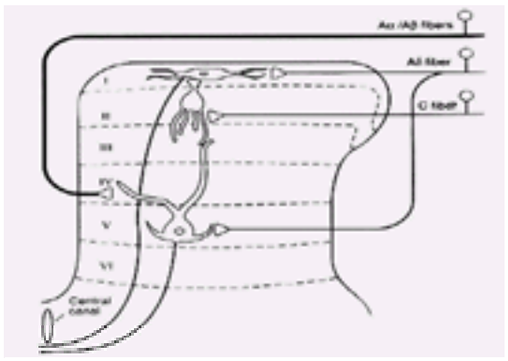

FIGURE 28.6-1 Termination of sensory afferent fibers in the dorsolateral quadrant of the spinal cord. The Roman numerals to the left denote the laminae of Rexed. A-delta (Ad) fibers project to both lamina I and lamina V, where they synapse on projection neurons. C fibers project to lamina II (shown) and lamina I (not shown). A-alpha and A-beta (Aa/Ab) mechanoreceptor fibers, which do not conduct pain impulses, terminate in lamina IV. (Reprinted with permission from Jessell TM, Kelly DD: Pain and analgesia. In *Principles of Neural Science*, ed 3, ER Kandel, JH Schwartz, TM Jessell, editors. Appleton & Lange, New York, 1991.)

LAMINAE I AND II Experimental ablation of substance P–containing, NGF-responsive pain fibers in lamina I does not affect responsiveness to painful mechanical or thermal stimuli, but it does abolish hypersensitivity to chemical stimuli produced by capsaicin. This suggests that lamina I may mediate hyperalgesic responses to inflammation. Disabling of processing by GDNF-responsive fibers in the inner part of lamina II by targeted knockout of the g isoform of protein kinase C in mice does not affect responsiveness to acute thermal or mechanical nociceptive stimuli; however, it does significantly reduce painful responses induced by nerve injury. These data suggest a model for the establishment of chronic pain states, in which NGF-responsive, peptidergic fibers mediate pain in inflammatory conditions, and GDNF-responsive, nonpeptidergic fibers mediate neuropathic pain.

DORSAL HORN CELL TYPES In lamina I, at least three histologically and functionally distinct neuronal cell types have been defined, each of which projects to a specific thalamic nucleus and thus activates a specific region of cerebral cortex. Nociceptive-specific cells respond to one or more types of nociceptors, a second cell type responds to innocuous cooling, and other cells respond to heat, pinch, and cooling (HPC cells). The ascending tracts projecting from these neurons maintain a strict somatotopic organization up to the thalamus.

CENTRAL PROJECTIONS The primary pain afferents synapse in the dorsal horn, and the secondary pain fibers decussate in the anterior commisure and ascend in the anterolateral fasciculus. The lateral fibers of this fasciculus form the spinothalamic tract, which rapidly transmits sensory-discriminitive aspects of pain to a lateral thalamic relay nucleus, from which a precise localization of nociceptive impulses is relayed to the mid/anterior insular cortex. Medial fibers of the anterolateral fasciculus contain the spinoreticulothalamic, or paleospinothalamic tract, which transmits impulses in a slower and less well localized fashion to the reticular core of the medulla and midbrain, which relay them to the medial and intralaminar nuclei of the thalamus. The medial thalamic area then projects to the anterior cingulate cortex. Other medial fibers of the anterolateral fasciculus project directly to the hypothalamus as the spinohypothalamic tract. The medial fibers with their diffuse projections and activation of the frontal lobes and limbic system are proposed to mediate the affective-motivational aspects of pain, or the unpleasant feelings associated with pain.

Painful thermal, electrical, and other stimuli travel in the medial and lateral ascending spinothalamic tracts. A parallel, lateral ascending pathway activated by nonpainful cold stimuli inhibits central processing of painful stimuli. Contact with a surface at temperatures above 15°C activates both systems and normally does not elicit pain because the inhibitory pathway suppresses the nociceptive pathway. If the inhibitory pathway is inactivated by peripheral nerve block of A-delta fibers, then stimuli up to 24°C are perceived as painful because of the unopposed action of the C fiber–activated nociceptive pathway.

Surgical sectioning of the ascending pain pathways may result in ablation of pain and thermal sense in the opposite side of the body. This effect is usually temporary, as auxiliary pathways are recruited over months to years to transmit nociceptive impulses. Therefore, partial cordotomy, or tractotomy, is generally an unsatisfactory approach to the treatment of chronic pain, because it is a permanent lesion that gives only temporary relief of pain.

Chemistry

NEUROTRANSMITTERS AND RECEPTORS C fiber terminals contain substance P, CGRP, glutamate, brain-derived neurotrophic factor (BDNF), and ATP, whose release is modified by presynaptic receptors for aminobutyric acid (GABA), μ opioids, adenosine, glutamate (N-methyl-D-aspartate [NMDA] and GluR5), and ATP (P2X family). Postsynaptic dendrites contain neurokinin-1 (NK-1), metabotropic glutamate, d-amino-3-hydroxy-5-methyl-4-isoxazole proprionate (AMPA) glutamate, NMDA glutamate, trkB, μ-opioid, and GABA receptors. C fiber inputs initiate long-term changes in synaptic excitability by promoting phosphorylation of the NMDA glutamate receptor. The resulting increase in response to any subsequent glutamatergic input underlies the phenomenon of central sensitization.

SUBSTANCE P AND OTHER NEUROKININS The neuropeptides substance P, neurokinin A, and neurokinin B bind most strongly to neurokinin receptors NK-1, NK-2, and NK-3, respectively. To discriminate the function of each neurokinin and each of the neurokinin receptors, targeted gene deletion (knockout) experiments have been done in mice. Two strains of mice were created in which the gene for preprotachykinin A, the precursor of substance P and neurokinin A, was deleted. These knockout mice displayed significantly reduced behavioral responses to moderately, but not mildly or intensely, painful stimuli. This reduction applied variously to thermal, mechanical, and chemical stimuli, but differences in the genetic backgrounds of the mice contributed to inconsistencies in their response in a variety of painful tests. These knockout mice required smaller dosages of opioids to achieve analgesia in the presence of moderately to intensely painful stimuli.

Early-generation NK-1 receptor inhibitors were inconsistently effective against pain, which implicated the NK-2 and NK-3 receptors in the mediation of substance P's effects. Several novel NK receptor antagonists were developed as adjuncts to opioids; however, available clinical trial data failed to demonstrate an analgesic action of NK-1 receptor antagonists. These studies indicated that substance P is not the sole pain substance; rather it appeared to be responsible for neurogenic inflammation in the periphery and may have modulated the overall sensitivity to pain. For example, the spinal cord action of substance P has greater intensity and domain in chronic inflammation than in acute pain.

Intense noxious stimulation sensitizes the central nervous system (CNS) and temporarily increases spinal reflexes, a phenomenon called wind-up. Wind-up is inhibited by antagonists of the NMDA and NK-1 receptors and is completely abolished in knockout mice that lack a functional NK-1 receptor. NK-1 receptor knockout mice can perceive acute pain of all intensities, but they are significantly impaired in their ability to suppress nociception after stress. The fight-or-flight response to stress involves temporary analgesia, probably mediated by descending pathways originating in the brainstem. This analgesia is adaptive in that it allows the animal to escape or defend itself without the distraction of pain. The stress-induced analgesic response is significantly reduced in NK-1 knockout mice.

INTERACTION OF NEUROKININS WITH GLUTAMATE Glutamate, acting through presynaptic NMDA receptors potentiates release of substance P, and substance P promotes the action of glutamate on postsynaptic NMDA receptors, the net effect of which is the enhancement of synaptic efficacy. Blockers of glutamate receptors of the NMDA and non-NMDA classes reduce pain-induced behaviors in laboratory animals. The NMDA glutamate receptor antagonist dizocylpine (i.e., MK-801) also reduces stress-induced analgesia.

SUBSTANCE P AND DEPRESSION Unexpectedly, a substance P receptor antagonist has shown promise as an antidepressant drug. The success of this agent has provided the first evidence of substance P receptor involvement in the regulation of mood. In this role, drugs of this class could become an important adjunctive treatment for chronic pain.

DESCENDING TRACT NEUROTRANSMITTERS In addition to substance P receptors, the dorsal horn contains binding sites for opioids, norepinephrine, and serotonin. The receptors, including serotonin (5-hydroxytryptamine [5-HT] type 2 [$5HT_2$]) receptors, are activated by descending fibers from the brainstem that reduce the sensitivity of the ascending pain system. Well-known examples of the potential descending influence of these corticospinal pathways include athletes or soldiers

with significant musculoskeletal injuries who nevertheless persist in their duties until the termination of the engagement, at which point they finally complain of pain.

Certain glutamatergic synapses in the dorsal horn are normally inactive ("silent") except in the presence of intense stimulation. Descending serotonergic input from the brainstem raphe nuclei activates 5-HT_1 receptors to transform these silent glutamatergic synapses into functional ones. In the early postnatal period, this stabilizes the serotonergic synapses and establishes the influence of the descending fibers. In the adult, recruitment of AMPA glutamate receptors may facilitate responses of pain pathways to nonpain input, which may promote chronic pain. This is only one of the effects of descending serotonergic input.

PROSTAGLANDINS Prostaglandins mediate pain perception at several points in the pain pathway. Aspirin, acetaminophen, and the nonsteroidal anti-inflammatory drugs all relieve pain by inhibiting cyclooxygenase-2, the synthesic enzyme for prostaglandins. In the spinal cord, prostaglandin E_2 increases the peak current through certain sodium channels, which contributes to the sensitization of nociceptors in inflammatory states. These particular sodium channels are found only on pain fibers.

NOVEL POTENTIAL THERAPEUTIC TARGETS A dorsal root ganglion acid-sensing ion channel (DRASIC) is unique to dorsal root ganglion neurons. DRASIC, the capsaicin receptor (VH1), and the specific sodium channel sensitized in inflammatory states are all unique to nociceptive neurons. Therefore, they each present a novel target for therapeutic intervention. At present, no agents acting at these sites are available for clinical use.

Plasticity of Peripheral Fibers and Spinal Cord

SENSITIZATION The terminal arborization and receptor densities of nociceptive fibers appear to be under tonic control by NGF. NGF appears to regulate the amplification of cutaneous stimuli both acutely and chronically. Acutely, increased concentration of NGF in the skin causes hypersensitivity to painful thermal stimuli, partially by triggering the release of histamine and serotonin from mast cells, and partially by direct action on nerve fibers. Conversely, blockade of NGF action prevents the sensitization of nociceptors to painful stimuli. In humans, administration of NGF produces prolonged muscle pain. Chronically, NGF regulates the densities of receptors for capsaicin, GABA, and bradykinin. NGF overexpression skews the population of A-delta fibers toward an increased percentage of nociceptors. NGF also controls the concentrations of substance P and CGRP, and the net effect of NGF is to increase excitability in the dorsal horn. Less is known about the role of GDNF in the maintenance of nonpeptidergic fibers. One role of GDNF analogous to that of NGF is to prevent the loss of nerve terminals resulting from nerve injury. As NGF can prevent the downregulation of substance P and CGRP caused by nerve section, so can GDNF maintain the levels of characteristic molecules of nonpeptidergic fibers after nerve injury.

Drugs that act on several cellular pathways have been tested in an effort to abort the persistence of established hypersensitivity. No approach has been fully successful in reversing the cascade of sensitization, at clinically tolerated doses. In contrast, in the case of postsurgical pain, techniques are under study to avoid triggering the sensitization responses to pain. For example, regional anesthesia (i.e., either spinal, epidural, or peripheral nerve blocks) is begun prior to the surgery and continued until tissue damage has begun to heal. Reliable clinical trials demonstrating the benefit of such "preemptive anesthesia" are difficult to carry out, but recent data show that rigorous regional anesthesia in the perioperative period reduces postoperative pain for weeks to months. These results bear on the use of anesthesia for painful experiences in newborns, such as circumcision. For example, newborns experience months of greater sensitivity in the heel pad that was lanced for blood samples than in the unlanced heel.

In one model of chronic pain, NGF increases the release of substance P in the spinal cord and triggers the sensitization of the spinal cord to painful stimuli (central sensitization) that characterizes states of persistent pain. Central sensitization is a heterosynaptic facilitation of dorsal horn neurons that results in the aberrant perception of pain during nonpainful stimulation. Overactivation of substance P–containing nerve termini resulting from NGF upregulation is thought to increase the simultaneous activation of pre- and postsynaptic elements in the dorsal horn and to allow "learning" of the pain response. Synapses thus stabilized will chronically lower the pain threshold (i.e., exaggerate the significance of painful stimuli).

Less can be said about modulation of GDNF in neuropathic pain states, other than the theoretical advantage of increasing peripheral concentrations of GDNF to limit loss of nociceptor synapses in lamina II. Many neurotrophic factors support the survival and function of A-delta and C fibers. As subsets of nociceptors are more fully characterized and new neurotrophic interactions are discovered, additional novel therapeutic targets may emerge.

NEUROPATHIC PAIN AND ALLODYNIA One proposed mechanism of neuropathic pain is that nerve damage reduces the number of nociceptive terminals in the spinal cord and permits the termini of innocuous mechanoreceptors to occupy the synapses normally reserved for pain responses. This results in the perception of pain in response to normal touch, a characteristic feature of chronic neuropathic states. Provision of NGF, GDNF, or both after nerve injury in mice maintains nociceptive termini in the spinal cord and prevents synaptic remodeling. The fact that nociceptors display receptors for a number of other known neurotrophic factors suggests that other trophic interactions will emerge to reveal further complexities of nociceptor synaptic plasticity.

Thalamus The spinothalamic tract fibers terminate on the ventrobasal, posterior, and intralaminar nuclei and the nucleus submedius. Each thalamic nucleus projects fibers to a distinct region of the cortex and contributes a distinct aspect of pain perception. As discussed above, two lateral thalamic relay nuclei—the ventral medial nucleus, posterior part and the ventral posterior inferior nucleus—receive nociceptive fibers from nocioceptive-specific and cooling-specific cells via the medial and lateral spinothalamic tracts and project to the middle and anterior insular cortex and the postcentral gyrus (Brodmann's area 3a). The medial thalamus—specifically the medial dorsal nucleus, ventral caudal portion—receives fibers from HPC cells via the inhibitory lateral spinothalamic tract and projects to the anterior cingulate cortex (area 24c). Whereas pain fibers are neatly segregated in the spinal cord, the response of a particular neuron becomes less predictable at the thalamic, limbic, and cortical levels. Thalamocortical projections mediating tactile and proprioceptive stimuli and the discriminative aspects of pain terminate in the somatosensory cortex. Fibers mediating the emotional aspects of pain probably terminate in the amygdala, hypothalamus, and limbic structures, including the anterior cingulate cortex.

Anecdotal evidence of one neurosurgeon collected during procedures to ablate tremors or anginal pain by cauterizing small regions of the thalamus showed that, in rare cases a small electrical stimulus immediately precipitated a classic panic attack. Cessation of the stimulus promptly terminated the attack. These observations of a psychological experience (usually viewed as a reaction to external danger) originating entirely within the brain resembled the phenomenon of pain sensations in the absence of tissue injury. A compelling body of evidence now supports a model in which people are aware only of the brain's representation of sensations, and that indistinguishable representations can rise to consciousness in response either to external or to internal stimuli.

Cerebral Cortex Nociceptive input rises to conscious perception in the cerebral cortex. Functional neuroimaging studies using functional magnetic resonance imaging (fMRI), positron emission tomography (PET) scanning, and electroencephalographic localization techniques have opened a new window on the higher processing of painful stimuli.

Thermal Pain Perception Behaviorally, cold environmental stimuli inhibit the perception of pain at the CNS level. Peripheral cold stimulation reduces the pain caused by electrical nerve stimulation, and cold ambient temperatures reduce pain behavior.

ANATOMY OF THERMAL RESPONSIVENESS Four brain regions of importance to thermal perception are the pericentral or primary somatosensory cortex (S1), the lateral opercular or secondary somatosensory cortex (S2), the mid/anterior insula, and the anterior cingulate cortex. S1 and S2 are differentially activated by cool (20°C) and warm (40°C) stimuli as follows. Cool stimuli activate only the mid/anterior insula and S2, while warm stimuli activate the insula, S1 and (more weakly) S2. The insula is activated by both innocuous and noxious stimuli, whereas the anterior cingulate is activated only by noxious stimuli. All four areas are activated by painfully hot (47°C) and painfully cold (5°C) stimuli, though cold activates S1 less intensely than does heat. The insula is activated by all thermal and painful stimuli, and it is proposed to maintain a general sense of the body's physiological condition. The anterior cingulate is activated only by thermal stimuli in the painful range. Digital subtraction of the activation patterns elicited by cool or warm stimuli from those elicited by cold or hot stimuli isolate only the anterior cingulate cortex as pain specific. This part of the anterior cingulate cortex is distinct from that activated by attention tasks, which indicates that the complex anterior cingulate cortex has a pain-responsive region that has been proposed to mediate the emotional and motivational aspects of pain. Separate studies have identified other cells in the anterior cingulate cortex that fire in anticipation of a painful stimulus. The patterns of activation of the anterior cingulate cortex, insula, and S1 and S2 in response to hot stimuli vary during the menstrual cycle. Activation in women in the midfollicular stage resembled that of men, but differed from that of women in the midluteal phase, although pain ratings did not vary across the phases.

THERMAL ILLUSION The pain-specific region of the anterior cingulate cortex is activated in a thermal grill illusion, in which a row of alternating warm and cool bars elicits pain. The warm (40°C) or the cool (20°C) bars in isolation are not painful, and neither stimulus alone activates the anterior cingulate pain locus. Interposing 1-cm–wide bars so that skin contacts an alternating pattern of warm and cold elicits pain and activates the anterior cingulate pain locus. The pattern of activation of the somatosensory cortices by the thermal grill is characteristic of cold pain. Cool bars alone activate both the medial nociceptive spinothalamic pathway and the

lateral inhibitory pathway. The inhibitory pathway normally predominates and suppresses the activity of the anterior cingulate pain region, and the cool bars are not perceived as painful. In the thermal grill illusion, the alternation of warm and cool bars reduces the lateral inhibitory activity by half, but does not affect the medial tract neurons. This inactivates the inhibitory signals and allows the medial pathway to activate the anterior cingulate region, which leads to the perception of pain.

Central Pain Syndromes Chronic pain can be caused by injury to the CNS itself (e.g., strokes, trauma, multiple sclerosis, and other structural lesions). Patients complain of deep aching or burning pain in the part of the body that projects to the injured region of the CNS. In addition, the affected body part has reduced pain sensitivity and thermal discrimination, and there may be allodynia, or “different pain,” in which an innocuous stimulus is perceived as painful.

The thermal grill illusion offers an explanation of the burning, ice-like pain and cold allodynia described by patients with central pain. Central pain syndromes are associated with damage to the lateral, inhibitory spinothalamic tract and the lateral thalamus and are not correlated with injury to the medial thalamus. In fact, one patient with lesion-induced central pain was unable to experience additional pain during the thermal grill illusion, suggesting that her lateral spinothalamic tract was fully inhibited by her injury and, therefore, not subject to further modification by the thermal grill. These data are consistent with a model in which inactivation of the inhibitory mechanism by a structural lesion aberrantly unmasks cold-stimulated pain perception associated with activation of the anterior cingulate cortex pain region.

Reciprocal Connections As in all sensory systems, there is an extensive projection from the higher to lower centers that modulates pain perception. The number of descending corticothalamic fibers is about ten times the number of ascending thalamocortical fibers. The cortex and hypothalamus project descending fibers to the periaqueductal gray matter of the midbrain, the ventromedial medulla, and to the posterior horn of the spinal gray matter. Other descending pathways originate in the locus ceruleus, dorsal raphe nucleus, and nucleus reticularis gigantocellularis.

Plasticity of Thalamocortical Loops The corticothalamic loop for nociception is not characterized, but data from other sensory systems provide a model for plasticity in a corticothalamic system. In the corticothalamic loop for vision, if the corticothalamic fibers are inactivated by treatment of the cortex with local anesthetics or NMDA receptor antagonists, the receptive fields of their corresponding thalamic neurons expand greatly, suggesting that corticothalamic fibers normally exert an inhibitory influence on expansion of the synaptic domain of thalamic neurons. In a complementary observation, if the cortical cells are electrically stimulated, they promote the sharpening of thalamic receptive fields and suppress responses of neighboring thalamic neurons. In this manner, descending corticothalamic fibers help to amplify specific thalamic inputs while simultaneously generating a surrounding field of suppression of thalamic activity. Conversely, synchronized firing of ascending thalamocortical cells increases the chance that their corresponding cortical cells will fire. It appears that thalamic maps are constantly remodeled by experience, as are cortical maps.

PHANTOM LIMB PAIN Phantom limb pain consists of the painful sensations consciously localized distal to the site of amputation. The absence of input from the amputated region leads in one of several ways to reorganization of sensory maps in the brain. Examples in which peripheral input molds cortical synaptic connectivity have been described in the somatosensory, visual, auditory, and olfactory systems. For example, in the somatosensory cortex of the adult, the homunculus redistributes itself over the remaining healthy cortex if part of the somatosensory cortex is damaged (e.g., by stroke).

Amputation deprives the CNS of sensory input from the distal limb, which would normally have been localized in sensory maps in the lower brainstem, thalamus, and cortex. Clinical experience suggests that reorganization may occur in any combination of the sensory maps. In some cases, input to thalamic neurons previously representing the hand may be reorganized so that they are stimulated by touching the arm stump, yet they may continue to signal touching of the hand in their connections to the cortex. Thus, upon contact of the arm stump, the subject is conscious of contact to the missing hand. Phantom limb pain presumably involves similar discrepancies between reorganization and respecification in pain pathways.

HYPNOSIS At the functional level, even in the absence of anatomical changes, alterations in mood, attention, or expectation can markedly influence the conscious perception of pain. For example, hypnosis is a state of heightened suggestibility attainable by a certain proportion of the population. Under a state of hypnosis, gross distortions of perception of any of the sensory modalities can be achieved on an instantaneous time scale. The anatomy of the sensory systems does not change, yet the same specific stimuli may be perceived with diametrically opposed emotional value before and after induction of the hypnotic state. For example, an onion may be savored as if it were a luscious chocolate truffle under hypnosis, and rejected as abhorrently pungent by the same subject only seconds later, when the hypnotic suggestion is reversed. How hypnosis redirects the flow of information in the cortex has not been determined, but it presumably involves sensory and association areas of the brain as well as the anterior cingulate gyrus. The unconscious changes in mental processing triggered by hypnosis are thought to be similar to those that underlie the effectiveness of placebo therapy for the treatment of chronic pain. Other techniques, such as biofeedback, attempt instead to establish conscious control over attention and allow purposeful diversion of psychic energy from pain pathways.

HEMISPHERIC LATERALIZATION OF PAIN PROCESSINC Recent functional neuroimaging studies found that painful and nonpainful somatosensory information was processed mostly in the right hemisphere. Although there was a homuncular representation in the primary somatosensory cortex of both hemispheres, painful and nonpainful thermal stimuli significantly activated the right thalamus, right dorsolateral prefrontal cortex, and right inferior parietal cortex, regardless of whether the stimuli were applied to the right or the left forearm. The dorsal prefrontal cortex is known to be densely connected with the posterior parietal cortex and is active during tasks requiring spatial working memory. In right-handed people, damage to the right posterior parietal cortex produces deficits in spatial attention and body awareness, which are not seen with the corresponding left cortical lesions. It is not known whether the same regions of the right hemisphere will turn out to be active during the experience of chronic pain.

ORGANIC CAUSES OF CHRONIC PAIN

Ongoing damage to any bodily tissues or to pain fibers at any point of the neural axis may elicit chronic pain. The mechanism is basically the same as that for acute pain, except that the injured tissues do not heal, and therefore the painful sensation continues without resolution. Analysis of the cause of pain and identification of appropriate treatment strategies are relatively straightforward in these cases, though many tens of millions of people in the United States with such conditions cannot achieve satisfactory pain relief with current analgesics. This major debilitating situation is due both to the limitations of current drugs and to prejudice on the part of caregivers, who often ration out an amount of analgesic quite inadequate for the person's subjective pain.

Physician prejudices play a large role in their habits of prescribing analgesic drugs. Many State Medical Boards have set up means of tracking prescriptions of controlled substances, and the threat of disciplinary action for overprescription worries many physicians. In 1997, in response to complaints that physicians were in fact undermedicating pain and causing needless patient suffering, the Federation of State Medical Boards published guidelines aimed at easing physicians' fears of the consequences of overly liberal use of controlled substances. Under these guidelines, physicians would not be punished if they were to take a proper medication history, outline a treatment plan, obtain informed consent from their patients, monitor the use of the drugs, and keep detailed records. As the population ages and more individuals develop conditions with chronic pain, physicians and patient advocate groups increasingly will seek clarification of the limits of governmental regulation of prescription of controlled analgesic drugs.

Any patient with chronic pain deserves a thorough evaluation seeking organic causes of pain, and if the first such evaluation does not yield positive findings, repeat evaluations should be performed at appropriate intervals. Establishment of a definite diagnosis guides the therapeutic approach, but equally important, it reassures the patient and family members that the suffering has a specific cause. Patients with chronic pain and their families must learn what will aggravate and what will mitigate the pain associated with a specific condition. Moreover, if no specific cause for the pain is uncovered despite a thorough investigation, the caregivers must be quite confident of this fact prior to discussing methods for accepting and coping with chronic pain. A careful, thorough medical evaluation should convince patients that the caregiver is dedicated to understanding their condition and suffering. A strong rapport between patient and provider is itself often a powerful analgesic.

Pain Due to Chronic Tissue Injury Many conditions with chronic tissue injury cause pain: musculoskeletal disorders, rheumatoid arthritis, cancer, acquired immune deficiency syndrome (AIDS), irritable bowel syndrome, peptic ulcer disease, inflammatory bowel disease, angina head trauma and movement disorders.

Other medical conditions causing chronic pain include sickle cell disease and endometriosis.

Pain Due to Neuropathy Stump pain is a common complaint after amputation of a limb. Stump pain is distinct from pain in a phantom limb, which presumably involves central projections of fibers severed in the amputation. Stump pain appears to result from the formation of painful neuromas at the severed ends of peripheral nerves.

CHRONIC PAIN WITHOUT EVIDENT ORGANIC CAUSE

Patients with certain types of chronic pain (e.g., chronic low back pain, angina pectoris, trigeminal neuralgia, migraine headache, tension-type headache, temporomandibular joint syndrome, postencephalitic myalgia syndrome, or fibromyalgia) should have a thorough neurological or orthopaedic evaluation, or both. However, in most cases, it will not be possible to ascribe the pain clearly to a specific lesion. It is possible that current diagnostic tools are insufficiently sensitive to detect certain subtle tissue injuries that could generate chronic pain. In fact, hypochondriacal patients who hold this conviction may spend a great deal of time pursuing a medical diagnosis while denying the possibility that the pain could be largely maintained by psychological conflicts.

After a thorough medical, neurological, and orthopaedic evaluation has failed to establish a physical reason for chronic pain, psychological factors must be considered. Chronic pain often appears at the location of a previous injury beyond a reasonable period of time for wound healing. The simplest explanation for this is that the memory of the painful injury retains its grip on the patient's attention, even after the painful stimuli cease.

PLACEBO EFFECT

The localization of pain without tissue pathology to the memory circuits of the CNS raises many of the same issues as an analgesic response to an inactive treatment—a placebo. In both cases, a subconscious mental operation either introduces or suppresses the experience of pain. In other words, the conscious brain accepts spurious input as real, based on unconscious motivations. This remarkable notion runs counter to the tenets of evidence-based medicine and tries the limits of common sense. It is instructive to examine the placebo effect in detail to evaluate the evidence for and against the ability of the brain to reconsider its own sensory input.

A placebo is an inactive treatment disguised as an active treatment. The success of the ruse depends on the health care setting, professional staff, complicated paraphernalia, and appearance of the medication. Whether a placebo response is elicited depends on the force of the provider's persuasion that a particular therapy will work and the readiness of the patient to ignore perceptable evidence that the intervention is a sham. Many psychological factors are thought to contribute to these tendencies, often in an ideosyncratic manner. Only some of these factors, however, have been confirmed experimentally. Ultimately, response to a painful stimulus is a synthesis of the physical sensation with internal memories, drives, and associations. The neuroanatomical terrain on which various sensations and recollections compete for conscious attention is being gradually and carefully mapped by functional neuroimaging.

The idea that conscious perception may have little or nothing to do with pathophysiology and that inactive remedies can take on a life of their own entirely within the brain is a discomforting thought. Yet, the placebo effect exerts a surprising and powerful impact not only on medical care, but more widely, in economics and politics.

Examples of the Placebo Effect The power of the placebo effect is tightly linked to the patient's perception of the caregiver. For example, in the eyes of many, the most impressive specialty is surgery. The surgeon is imbued with an aura of mystery and omnipotence. Surgical treatments are rarely tested against a placebo, and placebo-controlled trials that have been completed are generally compromised by manipulation of the assignments to active or placebo treatments for various reasons. Thus, in the absence of objective data, specific operations come into vogue on the basis of the well-meaning, but often subjective, convictions of influential surgeons.

For example, in the 1950s ligation of the internal mammary artery was commonly performed to treat angina pectoris, on the assumption that blood that would otherwise pass through the internal mammary artery would instead prevent ischemic heart injury by causing the coronary arteries to form extra blood vessels. Many patients whose internal mammary arteries were ligated experienced a reduction in angina after surgery. However, pathological examination of heart muscle failed to find any new vessels, which threw the premise of the procedure into question. The technique was finally subjected to a placebo-controlled trial. Some patients received a sham operation consisting of exposure of the artery without ligation; in the remainder, the vessel was ligated as usual. Unexpectedly, both groups had a comparable rate of success, as measured by reduced pain, increased walking distance, fewer medications, and some normalization of the electrocardiogram. Not only did these results indicate that placebo was as effective as the procedure, but more impressively, the benefits persisted in both groups for at least a 6-month period of observation. This was surprising, because the placebo effect was often considered to be temporary at best.

Other examples of therapies introduced widely, then later withdrawn because controlled studies failed to prove a specific action, include glomectomy for asthma, evamisole for herpes simplex, photodynamic inactivation for herpes simplex, organic solvents for herpes simplex, and gastric freezing for duodenal ulcers. Studies of these therapies, involving over 7000 patients, consistently showed good responses in 30 percent and excellent responses in 40 percent of the patients. In fact, clinical experience in general has confirmed that nonspecific but beneficial effects are commonly seen in the initial period after introduction of a therapy.

Another example of the power of an impressive machine and the influence of a white coat was the use of ultrasound therapy to reduce pain and jaw tightness following wisdom tooth extraction, a treatment initially reported to be as effective as steroids. To discriminate the effects of the ultrasound waves themselves from the remainder of the treatment experience, a study was designed so that the intensity of the ultrasound was set at different levels, including zero, without the knowledge of the patients or therapists. At no ultrasound output, there was actually a better outcome than when normal intensities were used. To rule out the possibility that the benefit was actually due to the massaging effect of the ultrasound probe, patients were taught to massage themselves with an inactive probe with the same movements used by the professionals. This produced no benefit. The only aspect of the ultrasound treatment associated with a good outcome was the presence of an expert-appearing therapist apparently administering invisible sound rays from a complex machine. The good outcome included not only pain reduction, but also increased jaw opening and reduction in swelling, two measures not usually considered purely psychological in origin.

These surgical and radiological examples were particularly dramatic, but similar examples too numerous to count have also been reported for pharmaceutical treatments. All new medications are subjected to placebo-controlled trials, in which some fraction of the subjects take a pharmacologically inactive pill identical in appearance to the active medicine. The percentage of patients taking placebo who nonetheless improve by the trial criteria generally hovers around 30 to 40 percent, but may range from nearly zero to almost 100 percent. A small number of novel treatments involve surgical implantation of genetically engineered cells, which the U.S. Food and Drug Administration (FDA) considers akin to drugs. The FDA is requiring sham operations, including drilling holes into the skull, prior to approval of the therapies. These unusual cases raise controversy, as patients may be subjected to the risks of anesthesia without receiving the active treatment.

Popular Perceptions of the Placebo Effect For centuries placebos were associated with quackery. In the absence of specific therapies, however, the power of persuasion was often the best treatment that could be offered. Nevertheless, more-astute observers recognized a fine line between benevolent and neutral or even malicious uses of placebos. With the discovery of effective, specific therapies, the ethics of dispensing a placebo became even more complicated, and the bias against understanding and using placebos strengthened. Presently, in the evaluation of new drugs, for example, the patient is asked to give “informed consent” to the possibility of being randomized to the placebo arm of a study. In practice, this smacks of coercion, as few patients are sophisticated enough to understand fully the comparisons being made. Moreover, the more thoroughly subjects are informed about the premises of the study, the lower the likelihood of a truly blinded comparison.

Today, deliberate deception in the form of prescribing a placebo is considered unethical outside institutionally sanctioned clinical trials. Within clinical trials comparison with placebo may be considered an irksome bureaucratic requirement for which investigators must hide their enthusiasm for their new agent. A considerable fraction of the $200 million to $500 million cost of bringing a drug from initial synthesis to market is spent on the comparison with placebos, mostly in phase III clinical trials. Yet, the existence of a fraction of patients who respond to placebo is generally deemed of no intellectual interest and an impediment to the discovery of the true mechanism of action of the drug. Despite this impatience with the placebo effect, the factors that determine the magnitude of the placebo effect are a key part of the art of medicine: the use of the provider-patient rapport to maximize compliance with and benefit from a prescribed therapy. In the case of established therapies that were not compared with placebo prior to their wide acceptance, it is often not considered ethical subsequently to assign patients to the placebo arm of a controlled trial. Thus, in the case of well-established treatments for certain emotionally charged ailments, the lack of an objective comparison with placebo has opened the door to unscrupulous manipulation by politically powerful groups.

To the medically unsophisticated, recognition of the placebo effect is counterintuitive and unsettling. A person who responds to a placebo may feel foolish and may become angry at the provider of the placebo.

Seven Myths Regarding the Placebo Effect Seven myths have arisen in traditional medicine to explain away the placebo effect.

MYTH ONE The first myth is that patients who respond to placebo were not actually sick in the first place but rather displayed symptoms of purely mental origin that happened to resemble a known organic condition. This notion, which basically supposes a hysterical origin for symptoms, is dangerous and cruel. It is contrary to true cases of conversion disorder, in which symptoms do not exactly mimic organic conditions but conform to the patient's own concepts of disease. It is too great a leap to

suppose that all placebo responders initially had a conversion disorder.

MYTH TWO The second myth is that the placebo is the equivalent of no treatment at all. This myth ignores the effect of the setting of drug administration. Further countering this myth is the fact that people can distinguish whether caregivers are actively giving them something or doing nothing for their illness.

MYTH THREE The third myth is that a fixed fraction of patients respond to placebos, usually about one third. In the literature, the placebo response rate varies from close to zero to almost 100 percent. The placebo response rate in clinical practice, where the patient knows the provider, is even higher than that in the detached, anonymous setting of clinical studies. Other factors that determine the placebo response have been identified, including the appearance of the dose of medication. The most effective formulation is a capsule with colored beads, followed in descending order of influence by colored tablets, then by white tablets with corners, and finally by round white tablets. Placebo response is increased if the pills are taken from a bottle with a commercial label rather than one with a typed label. In one example of the crossover experimental design, subjects who encountered active medicine first were more likely to respond to the placebo than those who encountered the placebo first. In fact, several studies have found that the sequence in which the pills are given can convert placebo nonresponders to responders.

MYTH FOUR The fourth myth is that people who respond to placebo are characteristic in certain psychological ways, variously described as suggestible, hypnotizable, neurotic, extroverted, introverted, acquiescent, desiring to please, lacking in sophistication, and accepting of authority. As with efforts to define a pain-prone personality, most studies have failed to find a placebo response–prone personality, and in those that reported a correlation, it was variously positive or negative.

MYTH FIVE The fifth myth is that pain is a multidimensional experience, and placebos affect only part of the pain. This supposes a simple Cartesian dualism of body and mind. While some experimental protocols have shown that opioids and benzodiazepines independently affect either the intensity or the unpleasantness of pain, respectively, clinical experience has shown that placebos may affect one, the other, or both dimensions. The exact effects depend on factors such as suggestion, expectation, and instruction.

MYTH SIX The sixth myth is that the placebo response is produced by endogenous opioids. With the description of endorphins and enkephalins (the endogenous ligands for the opioid receptors in the brain and spinal cord), a theory was soon proposed that the placebo effect was mediated by these molecules. This proposal captured the imagination of both scientists and the general population and gave some respectability to the distasteful business of placebos. Despite its popularity, however, there is no evidence to support this concept. Careful studies of the effect of the opioid receptor antagonist naloxone (Narcan) on the placebo response have shown either positive, neutral, or negative results. Actually, the endorphins and enkephalins are weak ligands for opioid receptors. However, this fact, does not change the idea that the opioid receptors are somehow involved. Indeed, the recent discovery of a potent endogenous ligand for the μ-opioid receptor, called endomorphin, would appear to throw a life preserver to the theory. Opioid receptors are likely to play some role in the placebo response, but it is quite unclear how critical their exact role is.

MYTH SEVEN The seventh myth is that statistical analysis can discriminate the pure therapeutic action from the placebo response. Since patients respond to many cues during treatment, including the actions of the drug itself, it is not possible to identify and correct for the placebo effect with certainty. An active treatment exerts physical and psychological effects that cannot be exactly duplicated by the placebo. Efforts to use a placebo whose adverse effects mimic those of the active drug are never as clean as desired, and this approach may simply muddle the conclusions by introducing additional variables. For some nonpharmacological treatments, such as acupuncture, an active placebo cannot be defined. It is nearly impossible not to tip the subject off to some expected effect of a new drug under evaluation, if only through the wording of the instructions to subjects or the subtle reactions of the person administering the therapy. For example, in 22 studies of analgesics covering a range of potencies from aspirin to morphine, the larger the therapeutic effect, the larger the placebo response. Whereas one would expect higher ratios of drug responders to placebo responders with stronger analgesics, in fact, over the entire spectrum of analgesic potencies, 55 to 60 percent more patients responded to the active drug than responded to the corresponding placebo.

The placebo response appears to be inextricably linked to the perception of pain. It has defied all attempts to wish or reason it away. On the contrary, the placebo effect is a powerful phenomenon, whose harnessing distinguishes master healers from technicians. Rational, evidence-based medicine has erected a barrier to the acceptance of both the placebo effect and chronic nonorganic pain. This defensive posture fits nicely with the needs of some patients with chronic pain to go from specialist to specialist in an endless search for the physical cause of their pain and to subject themselves to an array of potentially dangerous interventions in an effort to eradicate the source of their pain. A fuller appreciation of the placebo effect by providers and patients would often lead much more directly to control, if not elimination, of the pain. It is not wise to underestimate the power of the brain to legitimize a deceptive image.

Psychological Explanations of the Placebo Effect

Affective The idea that the placebo reduces the unpleasantness of pain or decreases anxiety is not supported by experimental data. Moreover, any affective changes associated with the placebo could be a consequence rather than a cause of the reduction in pain.

Cognitive The commonest explanation is that the placebo effect results from the expectations of the subject. Subjects will readily state whether they believe that a certain therapy will work, and this attitude faithfully predicts their response to placebo. The expectation of success has variously been called belief, faith, confidence, enthusiasm, bias, meaning, credibility, transference, and anticipation. Response to placebo requires sufficient learning ability to generate an expectation. The learning of expected effects depends on culture, education, experience, and personality. In turn, the intensity of the expectation is influenced by the provider's expectation, enthusiasm, and charisma. Many reports discuss the effects of the provider-patient relationship on the placebo response. It is a tragedy that the pressures of managed care to see many patients rapidly impair the development of provider-patient rapport and very likely undermine many intangible cues that contribute to the effectiveness of active therapies.

The clinical observations of the power of anticipation were supported by neurophysiological recordings of neurons in the frontal lobe of primates. Selected frontal lobe neurons began to fire when the animal expected a favorite type of food. If the food was presented, the neurons fired more rapidly; if food the animal disliked was presented instead, the animal became angry and the neurons ceased firing.

Behavioral Conditioning In classical Pavlovian conditioning, a subject learns to associate an unrelated stimulus with the expectation of a particular physical stimulus. In classical conditioning, a sound may elicit salivation, a drop in blood sugar, immune suppression, and other physiological responses, even in the absence of an appropriate causative stimulus. In the analogy with pain relief, a powerful analgesic reduces pain and serves as the unconditioned stimulus. If a uniformed professional administers the medication with a hypodermic needle, then the patient becomes conditioned to expect that pain relief will follow the appearance of the needle-wielding person in a white coat. Therefore, if the uniformed professional injects saline with the hypodermic needle, the patient may still achieve level of analgesia identical to that achieved with the active medicine.

Cognitive and conditioned factors may overlap in people. For example, it is possible to set up a paradigm of conditioned learning with a mild anesthetic, then to manipulate the expectation of potency of the medication and subsequently observe a greatly increased conditioned response to the same anesthetic.

Attention Appropriate to the Situation Anecdotes gathered in wartime and during sporting competitions describe individuals who have suffered severe tissue damage yet report pain only after the battle. It appears that their attention is focused on a primary goal, either leaving the deadly battlefield or winning the competition, and the somatic sensations of tissue injury are temporarily ignored. This disconnection between injury and pain is not seen in all victims, however; some report high levels of pain immediately. In general, fear reduces pain, while anxiety increases sensitivity to pain. Three behavioral responses are seen for pain: first, an attempt to avoid further injury; second, seeking aid and safety; and third, recovery from injury. The third phase includes immobilization of the painful part, protection of the injured area from contact, social withdrawal, and sleep.

NEUROANATOMICAL BASIS OF ATTENTION

In animals, stereotyped reflexive behavioral patterns are elicited in response to specific stimuli. The animal selects the stimulus to which it will respond and executes only one motor program at a time. Sensory input is used continuously to guide the movements toward the goal. In people, in the case of the placebo response, the conscious perception of pain is but one of several items that can grab the attention, and under certain circumstances the brain can ignore or exaggerate conscious awareness of somatosensory pain signals. The placebo taps into this capacity and diverts attention away from the perception of pain, directing attention instead to other nonpain concerns. Instead of operating at the site of tissue injury or on the ascending pain pathways, the placebo acts on the mechanism by which conscious

attention is directed. Experimental data are emerging to localize this function within the brain.

Somatic sensations first arrive in the cortex in the somatosensory areas. Simultaneously, other cortical regions are receiving visual, auditory, olfactory, and gustatory stimuli. Features extracted from these inputs are compared with memory stores and are routed through the amygdala, where an emotional significance is assigned to the internal representation of particular external stimuli. The amygdala and hippocampus consolidate new memories according to the goals and drives of the individual, and long-term memory stores are encrypted in the cortex. Separate experiments have determined a part of the anterior cingulate gyrus that is activated during all cognitive tasks, no matter in what sequence other brain regions are activated. This region is thought to play an important role in routing each particular event around the various areas of the brain.

Two complementary techniques have been used to map the pathway of neural impulses in humans in real time. Functional neuroimaging techniques (fMRI, PET scanning, and single photon emission computed tomography [SPECT] scanning) reveal all the brain regions activated in a brief mental task, but not the order in which they are activated. Electroencephalographic recordings of event-related potentials (ERPs) reveal the sequence of activation to millisecond accuracy. In combination, the imaging and the ERP data can be used to map a cognitive pathway from one brain region to the next. Improvements in fMRI techniques continue to increase the temporal and spatial resolution of the method.

Early studies to map changes in brain activity caused by changes in attention yielded interesting results almost immediately. For example, subjects were asked to attend to a specific feature in a scene, and this revealed one cascade of activation of different brain regions. The same subjects were then asked to attend to different details in the same scene, and a novel cascade of activations was seen instantaneously. In separate recent experiments designed to trace the brain activity required for sentence comprehension, change of a single word triggered an immediate and long-lasting divergence in cognitive pathways. In this protocol, subjects read sentences that were identical except that the first word was either *before* or *after*, which confer distinct meanings in discourse. Studies revealed a divergence in the ERP responses over the left frontal hemisphere within 300 ms, depending on whether the sentence began with *before* or *after*. Although the remainder of the sentence remained unchanged, the cognitive pathways continued to diverge over the course of the sentences. This study demonstrated that discrete high-level, real-world knowledge had immediate and sustained consequences on neural processing.

It is not difficult to envision that similar immediate and sustained changes in neural processing accompany each of the subtle cues triggered by a placebo and the context in which it is administered. Conversely, numerous contextual cues could influence the rate at which a painful experience recedes from attention. For example, if a painful experience distracted an individual and prevented recognition of a certain unconscious psychological conflict, then the memory of a painful experience might remain prominent to justify ignoring the unwanted problem.

A model for the activity of the placebo in the human brain can therefore be defined: a myriad of sensory data converges on the brain, including pain signals, antipain interventions, and placebos. These are compiled with memories and expectations based on neural pathways, orchestrated by the anterior cingulate gyrus. Subtle cues and stimuli may profoundly alter the particular pathways chosen. From this hodge-podge of competing events, a network (probably involving the prefrontal cortex and the caudate nucleus) assigns priorities by which individual events will rise to consciousness. For example, if painful sensations overwhelm other priorities, then the person will experience pain. Alternatively, if compliance with the suggestions of a respected caregiver is important, then pain may be ignored.

In the case of chronic pain in the absence of ongoing tissue damage, the memory of pain associated with a particular traumatic event may achieve priority. This will be experienced by the person as pain located in a specific part of the body. Moreover, actual ongoing nonpainful somatosensory stimuli from that body part (i.e., simple pressure) may be exaggerated and perceived as painful because of the associations in the spinal cord and brain with the memory of pain.

The pervasive influence of a painful memory clinically resembles posttraumatic stress disorder, in which the recollections of an extremely traumatic, but not necessarily directly injurious, event intrude continuously into every aspect of the subject's life. For example, there is a conceptual and biochemical similarity between the changes in the brain in posttraumatic stress disorder and the changes in the brain and spinal cord following intensely painful stimuli. In both cases, short-term, intense experience is “learned,” and this memory then exerts a pervasive long-term influence on subsequent experiences, whether painful or not. In both posttraumatic stress disorder and psychologically charged acute pain, a traumatic experience outside the usual range of experience is etched indelibly into the memory. Patients with posttraumatic stress disorder experience increased arousal and vigilance, intrusive recollections of the traumatic event, flashbacks, distortions of perception, and intense psychological distress and physiological reactivity in response to stimuli that resemble the traumatic event. In an analogous syndrome, the circuits of the spinal cord and brain respond to noxious environmental stimuli by establishing persistent hyperalgesia, in which patients experience chronic pain, preoccupation with painful sensations, and pain in response to even nonnoxious peripheral stimuli. Both posttraumatic stress disorder and sensitization of the spinal cord by noxious stimuli are ancient phenomena with a high degree of evolutionary conservation.

CLINICAL APPROACH TO PATIENTS WITH CHRONIC PAIN

Pain Behavior Pain behavior is the way individuals think, feel, and act in relation to their pain status. Pain behaviors are either adaptive or pathological, as when the pain behavior is disproportionate to the objective pathology. A full picture of an individual's pain behavior includes considerations of cognitive style and level of functioning, affective status, and overt behaviors.

Cognitive Aspects A person's individual conception of the cause and significance of pain influences mood and behavior. The conception of pain is formed from experience, observation, and teaching. Whether information is gathered reluctantly, eagerly, or ambivalently depends on cognitive style, developmental history, and personality type. Data can be gathered obsessively or only in overview, which determines how detailed the information is. Patients form opinions about the etiology, pathology, treatment, and prognosis of their condition. Patients may labor under a misperception caused by poor communication with a provider or through exaggeration of the significance of one or more factors, for a variety of reasons. Inaccurate information may then engender ineffective pain behavior.

Affective Aspects All features of an experience with pain have a symbolic significance derived from the person's unique background and sensitivities. The meaning of an individual's pain can only be discovered by patiently tracing associations evoked by the pain experience. Because pain is a somatic sensation, one must understand the patient's body image and its significance to the person's sense of self and place in society. Pain may be associated with ideas of punishment, guilt, loss, threat, or sexual gratification and may thus elicit corresponding emotions. The main affective states encountered with pain are depression, anxiety, and anger. These affects may lead to guilt and shame, and they may be intense enough to disrupt customary coping strategies.

If a patient with chronic pain is depressed or anxious, these conditions should be treated on their own merits with available methods. The “chicken-and-egg” question of whether depression causes chronic pain or vice versa does not have one answer. Each patient should be evaluated with a history and mental status examination to uncover both preinjury mood disturbances and current affect. A psychotherapist must ensure that patients have a realistic understanding of their illness. If patients resist challenge of distorted notions about their pathology, the therapist should provide reassurance that emotional support will be available at all stages of treatment.

Anger and resentment toward the provider may interfere with the provider-patient relationship, particularly if these feelings are not acknowledged and accepted by the patient. Patients need room to ventilate any feelings of anger, guilt, anxiety, and shame toward caregivers and reassurance that these feelings are normal and will not cause caregivers to punish or abandon them. Although it is only human for providers to react defensively to criticism, they must make every effort to establish any basis in reality for a patient's feelings of hostility and resentment.

Behavioral Aspects The wide range of individual reactions to pain, from minimal disruption of activities to the triggering of a calamitous collapse of family and occupational support, appears after injuries of many different types. Although one cannot predict a priori where an individual patient will fall on this spectrum, behaviorally oriented psychologists have studied a person's reaction to pain as a means of communicating their subjective experience. Pain behaviors include (1) verbal complaints of pain and suffering, (2) nonlanguage sounds (i.e., moans, sighs), (3) body posturing and gesturing (i.e., limping, rubbing a painful body part or area), and (4) maintenance of functional limitations (i.e., reclining for excessive periods of time). Efforts to distill the various possible combinations of each of the behaviors into a single index of subjective pain have been unsuccessful, perhaps for obvious reasons. For example, there is no systematic heirarchical scale for any of these pain behaviors individually, much less in combination. In addition, because pain behaviors are subject to voluntary modification, they cannot serve as the basis for an objective scale upon which to base predictions of outcome.

Patients' behaviors not only reflect what they think and feel, but also are constrained by social and medical boundaries. The practical limits of physical activities and recommendations for rehabilitative exercises should be detailed by a physician or rehabilitation therapist. Clear advice regarding the permissible range of active and passive movements permits the patient to exercise with less fear of further injury. Members of the patient's social setting should also be told what the patient is expected to do and how they can expedite recovery. Family members and friends may have their own unconscious conflicts and motives that lead them to perpetuate

the patient's illness role and undermine appropriate adaptation to injury.

Pain behavior varies from culture to culture, and whether it is abnormal depends on the culture. Pain behavior becomes abnormal if it is maladaptive, either by persisting longer than expected on the basis of the severity of injury or by being exaggerated and refractory to reasonable intervention by a qualified practitioner. Abnormal pain behavior is considered to be motivated by unconscious conflicts, as distinguished from malingering, in which the sick role is consciously manipulated for secondary gain. Unconscious conflicts may be neurotic, in which pain behavior is used as an immature defense against other fears, or psychotic, in which the patient's belief in the presence of pathology is delusional.

NEUROTIC ABNORMAL PAIN BEHAVIOR The two ends of the spectrum of neurotic abnormal pain behavior are hypochondriasis and conversion disorders. Hypochondriacal reactions are characterized by an overconcern for health disproportionate to the objective pathology. Hypochondriacal patients have anxiety about being ill, yet remain convinced of the presence of disease despite reassurance to the contrary. They are preoccupied with bodily symptoms, but usually they have insight into the irrational nature of their fears. These symptoms define hypochondriasis in the American Psychiatric Association's fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV). Hypochondriacal patients are more comfortable discussing their symptoms than any other problems in their lives, and they may deny any psychological stresses. Hypochondriasis permits the patient to withdraw from responsibilities, to become dependent on others, to punish the self and thereby assuage guilt, and to direct hostility toward others, including caregivers, by complaining but not responding to reassurance.

Patients with conversion reactions deny preoccupations or fears about illness but complain about pain and its consequences. These symptoms define pain disorder (DSM-IV), which has also been called somatoform pain disorder, psychogenic pain disorder, idiopathic pain disorder, and atypical pain disorder. The DSM-IV diagnostic criteria for pain disorder include (1) pain in one or more anatomical sites as the predominant focus of the clinical presentation and of sufficient severity to warrant clinical attention, (2) pain that causes clinically significant distress or impairment in social, occupational, or other important areas of functioning, (3) pain whose onset, severity, exacerbation, or maintenance is strongly influenced by psychological factors, (4) the symptom or deficit is not intentionally produced or feigned (as in factitious disorder or malingering), and (5) pain not better accounted for by a mood, anxiety, or psychotic disorder and not meeting criteria for dyspareunia.

Many patients with chronic pain display features of both hypochondriasis and conversion disorder, and these symptoms may be lumped together as abnormal pain behavior. The Illness Behavior Questionnaire (IBQ) can help to categorize a patient's neurotic tendencies.

BEHAVIORAL TREATMENTS The use of devices that have little or no practical effect on the source of pain, but which suggest comfort, may trigger subtle psychological reactions in the broad category of the placebo effect. For example, lumbar supports are often worn by workers suffering from low back pain. There is scant evidence that such supports confer sufficient immobility to reduce biomechanical loading of the trunk or to limit motion during manual lifting. On the other hand, wearing the supports may remind workers to use better lifting and safety techniques. Six large studies (each with some methodological limitations) have left unclear whether lumbar supports prevent low back injury. In contrast, the likelihood of reporting of low back pain is more clearly associated with prior back injury, dissatisfaction with the job, poor interactions between the worker and the supervisor, and poor employee appraisal rating. Thus, the benefit of lumbar supports, especially in the context of education in safe techniques, more likely is due to psychological factors and redirection of attention away from the dissatisfaction with the job than to actual physical factors.

EXERCISE General recommendations for exercise in patients with chronic pain cannot be defined; a specific program of stretching, mobilization, and gradual introduction of activities must be outlined individually for each patient, depending on the details of tissue injury and functional goals. Individuals' response to encouragement is influenced by their coping style. Patients with mature defenses will follow a graded exercise program faithfully and will tolerate a certain degree of discomfort with the expectation of improvement. Patients who derive secondary gain from illness, within the dynamics of their family or work situation, may find unconscious ways to sabotage an exercise program, either by avoidance or by overexertion that causes more injury.

Multidisciplinary Pain Clinics Abnormal pain behavior is best approached in a multidisciplinary pain clinic, in which the staff discusses and coordinates a multifaceted approach to helping the patients understand and treat their pain. Given the somatic nature of the complaints, caregivers must be able to move comfortably between somatic and psychological discussions. Credence must be given to the patients' complaints, and they must be fully involved in planning their treatment program. Treatment is most successful in an ongoing relationship with a primary caregiver who anticipates the inevitability of advances and setbacks and can help patients interact as needed with the multispecialty pain clinic.

MULTIAXIAL ASSESSMENT OF PAIN

Medical assessment and treatment of injuries is an essential first approach to a patient with pain, both for purposes of rehabilitation and to determine the degree of compensable disability. A thorough medical assessment may rule out organic pathology with confidence, yet delve into the patient's experience of pain only superficially or not at all. Many medical and surgical clinicians rely on subjective estimations of the level of pain, (i.e., mild, moderate, or severe) to follow a patient's progress. This approach introduces obvious variability from patient to patient and is an inadequate basis for emotional and psychological treatment approaches in patients who exhibit no obvious organic cause for pain.

The traditional medical conception of pain promotes a view that pain is of physical or psychological origin, or some combination thereof. In practice, there may be limits to how accurately clinical assessment and diagnostic tests can predict an expected level of pain or duration of disability. The importance of psychological factors may need to be inferred from the incongruity of the patient's complaints and disability compared with a best guess of the pain and disability expected from a specific injury. The presence of significant psychological conflicts does not guide the treatment but rather indicates a need for a comprehensive psychological assessment.

Rating Scales for Assessing Chronic Pain Numerous rating scales have been constructed to organize the diagnostic criteria for psychiatric disorders and for personality characteristics. Several psychiatric rating scales have been used to define personality profiles that predict outcome after injury or surgery, including the Minnesota Multiphasic Personality Inventory-II (MMPI II), Rorschach test, Symptom Checklist–90, as well as scales specifically developed for chronic pain patients, including the Pain Apperception Test, Low Back Cognitive Distortion Scale, low back (Lb) scale derived from the MMPI-II, Back Pain Classification Scale derived from the McGill Pain Questionnaire (MPQ), and the Mensana Clinic Test for Chronic Back Pain. Establishing the validity of a particular scale has generally included comparing answers given by a defined population of patients with those given by both a normal population and a population defined by an unrelated criterion. Not every scale has been thoroughly validated in clinical trials.

One early goal for which rating scales were developed was identification of a pain-prone personality. This elusive goal was based on certain assumptions, including the hypothesis that historical factors such as familial alcoholism, child abuse, or spouse abuse predisposed a patient to develop psychogenic pain as a masked form of depression. After several unsuccessful attempts, it became clear that empirical data were inadequate to demonstrate the existence of a pain-prone personality. Moreover, none of these rating scales, administered pretreatment, consistently and accurately predicted treatment outcome. The data from published tests of these rating scales also did not support similar attempts to define a “low back personality” apart from the emotional disturbance associated with chronic limitation and disruption of activity.

None of the attempts to discern a purely psychological etiology for chronic pain permitted formulation of an overarching theory of pain. In fact, close inspection of patients with chronic pain revealed as broad a range of contributory factors as there were different personalities and circumstances. This realization stimulated the development of comprehensive descriptive assessment instruments that sought to define an operative constellation of factors, including cognitive, affective, psychosocial, behavioral, and physiological components. This approach emphasized current functioning and did not attempt to pigeonhole patients into theoretical categories of personality or behavior.

Accurate characterization of patients' idiosyncratic views of their misfortune, their unique experience of pain, and their coping resources—personal, financial, familial, and social—were combined to outline a realistic treatment plan. It turned out that patients' motivation for treatment and how this was balanced against other demands on their time and energy was particularly important. A patient's perceptions of caregivers, employers, and family members were likely to influence the strength of the therapeutic bond and thus the success of a therapeutic regimen.

Tests developed to standardize the psychosocial and behavioral dimensions of chronic pain included the Sickness Impact Profile, the IBQ, the Behavioral Assessment of Pain (BAP), the Millon Behavioral Health Inventory (MBHI), and the West Haven–Yale Multidimensional Pain Inventory (MPI), as well as additional test batteries to assess coping repertoire. These provided information about one facet of the chronic pain experience, which was best used in combination with medical, affective, and cognitive assessments.

Application of scales that relied on self-reporting of distress and disability revealed that no person was a truly objective self-observer. For example, a group of asymptomatic occupational therapy students took three pain behavior rating scales with the instruction to imagine that they had chronic pain and to represent themselves as either coping well or coping poorly. The students achieved significantly different scores depending on how they imagined themselves coping. Students portraying poor coping emerged with scores indistinguishable from those of a comparison group of patients enrolled in a multidisciplinary pain clinic at a teaching hospital, showing that one could bias one's answers to self-assessment scales to appear either healthier or more ill. In conclusion, these studies reinforced the importance of a detailed, open-ended, unstructured interview, corroborated by significant others, as the only reliable way to reveal an individual's motivations.

Minnesota Multiphasic Personality Inventory The most widely used rating scale for chronic pain is the MMPI. Studies have focused on the “neurotic triad,” the first three of the 10 standard scales, in patients with chronic pain. Scores on the hypochondriasis (Hs) scale, which measures “abnormal, psychoneurotic concern over bodily health,” and the hysteria (Hy) scale, which measures the tendency toward conversion disorder and which shares 20 of its 60 items with the Hs scale, were both elevated in some patients with chronic pain. In contrast, scores on the depression (D) scale, which measures symptomatic depression, were low in many patients with chronic pain. These scores contributed to the “conversion V,” that is, high on scale 1 (Hs), low on scale 2 (D), and high on scale 3 (Hy).

People with high scores on the Hs scale are excessively concerned about bodily processes even though there is no organic basis to their problems. They have excessive bodily concerns, numerous vague somatic symptoms, and undefined complaints such as gastric upset, fatigue, pain, and physical weakness. They are likely to be selfish, self-centered, narcissistic, pessimistic, defeatist, cynical, and dissatisfied, and their complaining, whining, demanding, and critical behavior may make others miserable. They express hostility indirectly and are not very responsive to psychological therapy.

People with high scores on scale Hy tend to deny social anxiety, to deny psychological problems, to seek excessive attention and affection, and to resist gaining insight into their own behavior. They tend to react to stress by developing physical symptoms such as headaches, chest pains, weakness, and tachycardia. They have an expectant mood and are facile in social situations.

People with high scores on scale D are characterized by the features of depression, which include a generally negative frame of mind and a tendency to report poor morale, lack of hope in the future, dissatisfaction with life, or low mood. A patient with a high score on scale D is less likely to find pain self-reinforcing than a patient with a low score on scale D.

Among the numerous supplementary measures derived from the MMPI, subtypes of patients with chronic pain were defined in a variety of settings. In one attempt, four subtypes, called P, A, I, and N, were suggested. Type P appeared the most psychopathological, in that scores on nearly all scales were elevated. Type P patients were extreme in their claims of physical illness, psychological distress and social maladaptation. Demographic correlates included poor education, high rates of unemployment, and limited household income. Type A was defined by a conversion V on the neurotic triad scales. It had no unique demographic correlates. Type I had elevated scores on all of the neurotic triad scales and on no others. Type I patients appeared to be the most physically infirm, with multiple surgeries and hospitalizations. They often did not improve in physical status with treatment, but appeared to benefit psychologically. Type N profiles were normal in that no scale scores, except perhaps scale K (which measures defensiveness and a tendency to deny problems), were often elevated. Type N patients were moderate in their claims of ill health, often were better educated and employed, and appeared to respond well to treatment.

Despite the complex schemata suggested by these scales, application of MMPI criteria to additional groups of patients with chronic pain successfully predicted response to treatment only about half of the time. The MMPI was not specifically tailored for patients with pain, and its major weakness in this application is that it lacks questions to assess the presence of actual physical abnormalities that could cause pain. Its clinical usefulness is therefore limited compared with other, more specific rating scales. The lengthy MMPI should be used cautiously as the basis for treatment decisions in patients with chronic pain.

McGill Pain Questionnaire The MPQ consists of four parts (Fig. 28.6-2). First, patients indicate the site of their pain on a drawing of a human figure. Second, patients must choose between 78 adjectives divided into 20 groups, covering sensory qualities, affective states, and specific medical conditions. Third is a series of questions about prior pain experience, pain location, and use of pain medication. Fourth, patients rate their current pain on a pain intensity index (PPI), which is a visual scale from “no pain” (0) to “excruciating” (5).

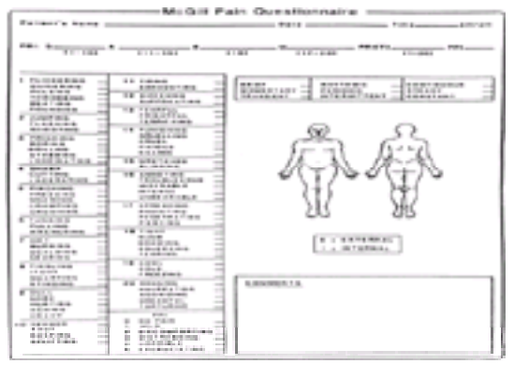

FIGURE 28.6-2 The McGill Pain Questionnaire. (Reprinted with permission from *Textbook of Pain*, ed 3, PD Wall, R Melzack, editors. Churchill Livingstone, Edinburgh, 1994.)

The MPQ has been applied in a wide range of settings in the assessment of both acute and chronic pain. The verbal pain descriptors circled are processed to generate ratings that aim to distinguish three components of the pain experience: the sensory-discriminative, the motivational-affective, and the cognitive-evaluative. However, empirical data do not support the proposed discriminations, but rather indicate that the MPQ scores are correlated only with the intensity of pain.

Emory Pain Estimate Model One of the earliest rating scales devised specifically for patients with chronic pain was the Emory Pain Estimate Model (EPEM), in which assessment was divided into “pathology” and “behavior.” The pathology score was derived from quantification of physical examination procedures (i.e., ratings of joint mobility and muscle strength) and abnormalities on radiological studies. The behavior score was a composite of activity levels, pain verbalizations, drug use, and abnormalities on certain MMPI scales. The EPEM defined four classes of patients with chronic pain. Class I, in which there were higher scores on the behavior scales and lower scores on the pathology scales, contained patients who displayed low levels of activity, high verbalizations of pain, prominent social and psychological malfunctions, and frequent misuse of medications. Class II, in which scores were low on both the behavior and pathology scales, contained patients who displayed highly dramatized pain complaints with ill-defined anatomical patterns, yet who did not have significant behavioral malfunctions. Class III, in which scores were high on both scales, contained patients with clearly demonstrable injury or illness and high-intensity illness behavior. Class IV, in which there were high scores on the pathology scale and low scores on the behavior scale, contained patients who demonstrated competent coping in the presence of a demonstrable pathological condition.

The EPEM was criticized for the arbitrary grouping of factors in both the behavior and pathology scales and for the uneven weighting of significant and incidental findings. It was more a projection of the authors' impressions than an empirically derived inventory. At the time it was devised, however, the EPEM was important in that it set out to compare medical and behavioral evaluations in the same patient. Subsequently, other, more robust, rating scales supplanted it.

Waddell Approach A comprehensive assessment of a series of patients with chronic pain was factor analyzed to determine the relative contributions to disability of physical impairment and psychosocial factors. Although the scales were not validated across several different groups of patients, conclusions were nonetheless drawn on a representative group of patients. Physical impairment was found to account for 40 percent of the variance on the disability index, while psychological distress (depression and somatic preoccupation) accounted for an additional 23 percent of the variance, and magnified illness behavior (cognitive distortion of symptoms) accounted for 8 percent of the variance. The methods in this report were not compiled into a clinically useful instrument, but the results represented one effort to discriminate the source of disability based on empirically derived criteria.

West Haven–Yale Multidimensional Pain Inventory The MPI characterized cognitive-behavioral factors, such as the impact of pain on the patients' lives, how

significant others responded to the patients' pain, and how active the patients were in common daily activities. The first of two parts assessed (1) reports of pain and suffering; (2) patients' perceptions of how pain interferes with their lives, including interference with family and marital functioning, work, and social-recreational activities; (3) dissatisfaction with their present levels of functioning in each of the areas listed in (2); (4) the support provided by spouses, family, and significant others; (5) perceived life control, incorporating the perceived ability to solve problems and feelings of personal mastery and competence; and (6) affective distress, including ratings of depressed mood, irritability, and tension. The second part consisted of ratings by patients of the frequency with which significant others respond to their expression of pain by (1) punishment (i.e., irritability, ignoring), (2) solicitousness (i.e., taking over their chores), and (3) distraction (i.e., encouraging them to work on a hobby).

Statistical analyses identified three distinct patient profiles on the MPI: (1) *Interpersonally Distressed*, patients with a common perception that significant others were not very understanding or supportive of their problems; (2) *Globally Dysfunctional*, patients who perceived the severity of their pain to be high, reported that pain interfered with much of their lives, reported a higher degree of psychological distress due to pain, and reported low levels of activity; and (3) *Adaptive Copers*, patients who reported high levels of social support, relatively low levels of pain and perceived interference, and greater activity. In two populations of patients with chronic pain, the three profiles represented 28, 43, and 29 percent of the patients, respectively.

Correlation with medical findings showed that the patients in the Globally Dysfunctional group had significantly more abnormal findings on medical assessment than the patients in either of the two other groups. Further analysis of the Globally Dysfunctional group permitted the definition of two subgroups of roughly equal size: a subgroup of Globally Dysfunctional patients with significant medical findings (a truly globally impaired group), and another subgroup with a high degree of psychological and behavioral disability but with relatively few abnormal physical findings.

The MPI was normed only on patients with chronic pain and it was validated against 10 other scaled measures of pain and cross-validation on additional groups of patients. For example, one comparison of MPI classification and abnormality on MMPI scales found that MPI Globally Dysfunctional and Interpersonally Distressed scores differed significantly from the MPI Adaptive Coper scores on MMPI scales 4 (psychopathic deviate), 6 (paranoia), 7 (psychasthenia [anxiety disorder with obsessive-compulsive features]), and 8 (schizophrenia). The Globally Dysfunctional and Adaptive Coper scores also significantly differed on MMPI scale 2 (depression). Some 79 percent of Globally Dysfunctional patients and 62 percent of Interpersonally Distressed patients displayed psychopathology as defined by significant two-point scale elevations on the MMPI, whereas only 23 percent of Adaptive Copers had significant two-point MMPI scale elevations.

A disadvantage of the MPI, found on wider use in the United States and Europe, was that the questions used to assess activity level, which were appropriate for middle-class suburban Connecticut, did not translate across socioeconomic and geographic boundaries.

Illness Behavior Questionnaire The IBQ consisted of a self-administered 62-item questionnaire developed from the Whiteley Index of hypochondriasis. Seven scales were derived. Scale 1, general hypochondriasis (GH), detected a fearful attitude toward illness, some degree of insight into the irrationality of these fears, and a high level of arousal or anxiety. Scale 2, disease conviction (DC), indicated strong affirmation of the presence of disease and resistance to reassurance by caregivers. Scale 3, psychological versus somatic focusing (P/S), defined a spectrum ranging from a tendency to blame oneself and accept the need for psychiatric help (psychological focusing) to a tendency to reject the possibility of psychological factors and instead focus on somatic problems (somatic focusing). Scale 4, affective inhibition (AI), indicated difficulty expressing personal feelings, especially negative ones, to others. Scale 5, affective disturbance (AD), indicated the presence of anxiety or depression. Scale 6, denial (D), indicated denial of current life problems and attribution of all problems to physical illness. Scale 7, irritability (I), indicated anger and awareness of interpersonal friction. Combinations of these primary scales yielded two additional second-order factors, affective state and disease affirmation, and an index of whether conversion disorder was present.

The profiles derived from the IBQ could clue the provider to the patient's pattern of resistance and could indicate where further questioning might uncover a psychological conflict at the root of neurotic abnormal pain behavior.

PHARMACOTHERAPY

Drugs appropriate for patients with chronic pain fall into three pharmacological classes: nonopioids, such as aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs), and acetaminophen; opioids and other narcotics, both natural and synthetic; and adjuvants, such as antidepressants, anticonvulsants, and antihistamines.

Analgesics Analgesics are most appropriate for the treatment of pain due to tissue injury. For patients with chronic pain but no demonstrated organic pathology, in whom psychosocial factors play a significant role in the maintenance of pain, long-term use of analgesics should be limited as much as possible. Such patients should be encouraged to increase their functioning while using a minimum of analgesics.

Aspirin and Nonsteroidal Anti-Inflammatory Drugs Aspirin and NSAIDs inhibit the cyclooxygenases COX-1 and COX-2 reducing the generation of prostaglandins such as prostaglandin E_2 which stimulate cyclic adenosine monophosphate (cAMP) and contribute to the sensitization of nociceptors. Inactivation of COX-2 is responsible for aspirin's anti-inflammatory, analgesic, antipyretic, and anticancer activities. Inactivation of COX-1 contributes to aspirin's antithrombotic, bronchospastic, and ulcerogenic effects, which significantly limit the use of aspirin and the NSAIDs for chronic pain. Aspirin irreversibly acetylates and inactivates both isoforms of COX but is 10 to 100 times more potent on COX-1.

The number of people in the United States with chronic pain not adequately treated with aspirin or NSAIDs (mostly osteoarthritis) is estimated at 30 to 40 million. Since the current treatment of chronic inflammatory conditions is often inadequate because NSAID dosing is limited by adverse effects, there is great interest in the recent discovery of derivatives of aspirin that selectively inhibit COX-2 and have little activity on COX-1 (Fig. 28.6-3). Because they are less limited by gastric adverse effects, selective COX-2 inhibitors can be used at concentrations capable of inactivating a greater percentage of the body's COX-2 than can aspirin or the NSAIDs. One such preclinical agent, *o*-(acetoxyphenyl)hept-2-ynyl sulfide (APHS), permanently inhibited COX-2 100 times more selectively than did aspirin. Reversible COX-2 inhibitors are at least as efficacious as NSAIDs for treatment of arthritis pain. Soon after it was marketed, the reversible COX-2 inhibitor celecoxib (Celebrex) became the fastest-selling new drug ever, reflecting the immense pent-up demand for novel analgeic drugs. In clinical experience, celecoxib and another reversible COX-2 inhibitor available in Great Britain, meloxicam, are associated with greatly reduced, but not absent, adverse gastrointestinal effects, including bleeding, ulcers, and perforations. It remains to be seen whether these adverse effects are directly due to the novel COX-2 inhibitors or are a carry-over from previous NSAID use. A second reversible COX-2 inhibitor, rofecoxib (Vioxx), is also approved for use in the United States.

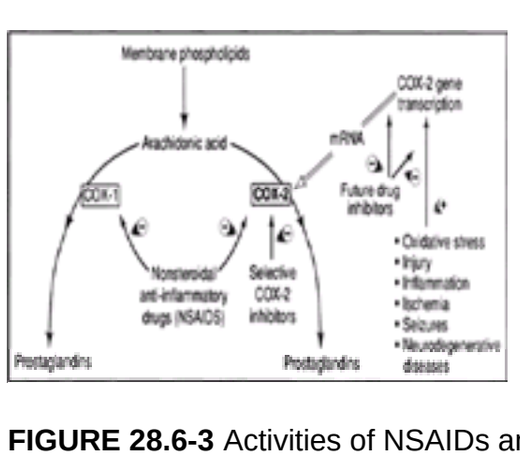

FIGURE 28.6-3 Activities of NSAIDs and COX-2 inhibitors. NSAIDs inhibit both COX-1 and COX-2, whereas selective COX-2 inhibitors only have significant inhibitory activity on COX-2. (Reprinted with permission from Pennisi E: Building a better aspirin. Science *280:* 1191, 1998.)

Clinical Guidelines for NSAIDs and Acetaminophen NSAIDs and acetaminophen are among the most widely used over-the-counter drugs (Table 28.6-3). These agents have a ceiling of therapeutic efficacy such that the pain in certain patients may be partially or not at all reduced by them. As doses reach the upper therapeutic range, each dose increment yields diminishing returns, and at higher doses, adverse effects predominate. Many patients with chronic pain control their pain with long-term use of these agents, and tolerance is not inevitable. Some patients, however, (especially those without demonstrable organic pathology) may eventually

cease to respond to these drugs and may then seek prescription analgesics.

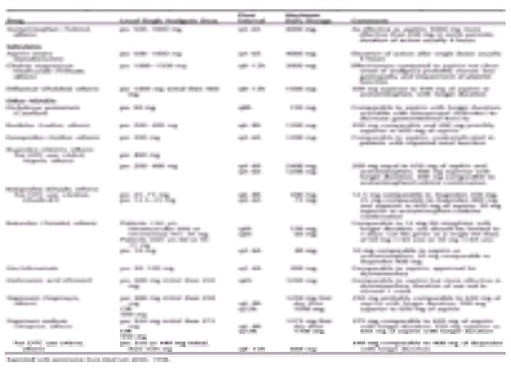

Table 28.6-3 Acetaminophen and NSAID Analgesics

ACETAMINOPHEN Unlike the NSAIDs, acetaminophen limits pain but does not reduce inflammation. There are almost no immediate adverse effects of acetaminophen, even at doses up to 1500 mg, but excessive doses taken either short- or long-term can cause potentially irreversible and even fatal hepatic injury. The threshold dose for hepatic damage is reduced in alcoholic patients, fasting patients, and those whose cytochrome P450 enzymes are induced. The maximum safe dosage for most patients is 4 g a day.

ASPIRIN The most prominent of aspirin's adverse effects are gastritis and (to a lesser degree) precipitation of asthma. Aspirin irreversibly inactivates platelet function and may prolong the bleeding time, which may be clinically relevant up to 10 days after the last dose, until bone marrow production replenishes the blood with active platelets. Long-term high dosages may cause salicylism, through a combination of CNS and peripheral effects. Many people take daily enteric-coated aspirin indefinitely as prophylaxis against heart attack and stroke.

OTHER NSAIDS Most NSAIDs are capable of higher levels of pain relief than are acetaminophen or aspirin. The NSAIDs comprise a heterogeneous group of chemical structures, and they have various specific toxicities in addition to their common gastric toxicity. Some NSAIDs are recommended for no more than 5 or 10 days of use, and these agents are unsuitable for chronic pain. Caution must be used in administering NSAIDs to patients with peptic ulcer disease, hypertension, alcohol abuse, congestive heart failure, ascites, volume depletion, and renal insufficiency and those who use diuretics. Some NSAIDs may cause drowsiness. NSAIDs cause a reversible inhibition of platelet function that ceases soon after the last dose is metabolized. Gastritis may be mitigated with concurrent administration of misoprostol (Cytotec), a prostaglandin E_1 analogue that compensates for the inhibition of prostaglandin synthesis resulting from inhibition of COX-1 by NSAIDs. Other agents used to reduce NSAID-induced gastritis include histamine type 2 (H_2) receptor antagonists, proton pump inhibitors, and antacids. Despite these countermeasures, NSAID toxicity can affect a significant percentage of elderly patients with chronic pain. Above the age of 60 years, the risk-benefit ratio tips in favor of opioids over NSAIDs.

Opioids The analgesic effect of exogenous opioids results primarily from binding to μ-receptors. Activation of k-opioid receptors may also produce analgesia, but a more significant function of k-receptors is to antagonize the actions of the μ-receptors. Both electrophysiological and behavioral studies in animals have shown that k-opioid agonists oppose the analgesic actions of μ-opioid agonists, probably at the brainstem level. μ-Opioid agonists produce euphoria and place preference, while k-opioid agonists cause dysphoria and aversion. μ-Opioid agonists are rewarding and increase release of dopamine in the mesolimbic system, whereas k-opioid agonists block the rewarding effects of μ-opioid agonists and decrease mesolimbic dopamine release. Mesolimbic dopamine activity is closely linked to reward behavior. Because morphine binds to both μ- and k-receptors, an imbalance between μ- and k-receptors may be the cellular basis for the tolerance that often develops to the effects of opioids. Clinical experience has repeatedly shown that opioid dosages must be gradually advanced to ever higher dosages to achieve the same subjective relief of pain.

μ-Opioid agonists bind to receptors in the spinal cord that act through inhibitory G proteins. Opioids inhibit neurotransmitter release at the first synapse encountered by incoming signals and inhibit sensitization of nociceptors by inflammatory stimuli. Opioids also reduce activation of the medial spinothalamic tract and increase activation of the inhibitory lateral spinothlamic tract. Thus, in sum, opioids act at the level of the spinal cord to reduce the activation of fibers that stimulate the anterior cingulate pain region and simultaneously potentiate the activation of fibers that suppress the anterior cingulate pain region.

At the supraspinal level, the principal antinociceptive pathways originate in the noradrenergic locus ceruleus and the serotonergic nucleus raphe magnus and descend into the spinal cord. Opioids activate these descending pathways. Thus, opioids act in concert on several neural tracts at both spinal and supraspinal levels, each of which independently reduces the intensity of pain perception.

Mice in which the μ-, k-, or d-opioid receptors have been experimentally deleted each display altered pain responses, though the abnormalities depend on the particular testing paradigm used. Further knockout studies targeting each of the various pharmacological subtypes of the μ-, k-, and d-receptors should permit finer dissection of the actions of each member of the opioid receptor family. Naturally occurring variations in pain sensitivity among different mouse strains have been associated with differences in the abundance of μ-receptors.

Clinical Guidelines for Opioids Naturally occurring opiates—morphine and other plant alkaloids—and the synthetic opioids are clinically the most effective analgesics across a wide range of conditions, including inflammatory and neuropathic states. Opioids are classified as full agonists, partial agonists, and mixed agonist-antagonists (Table 28.6-4). There does not appear to be a ceiling for the analgesic effects of full agonists, except that imposed by adverse effects. Injectable opioids have been used in numerous interventional treatments, including slowly releasing pumps that continuously inject tiny doses of opioids directly into the cerebrospinal fluid of the spinal cord. These have variably been shown to relieve certain types of chronic pain, but the effective dosage must be gradually increased each month. The clinical usefulness of opioids is limited by respiratory depression, constipation, development of tolerance, development of physical and psychological dependence, and perceived abuse potential, which is, however, rare above the age of 60. Recent clinical practice guidelines of the American Geriatrics Society in fact recommend use of opioids for treatment of chronic noncancer pain in patients over the age of 60, because the risk of tolerance and dependence with opioids is relatively less than the risk of gastric and renal toxicity due to chronic use of NSAIDs.

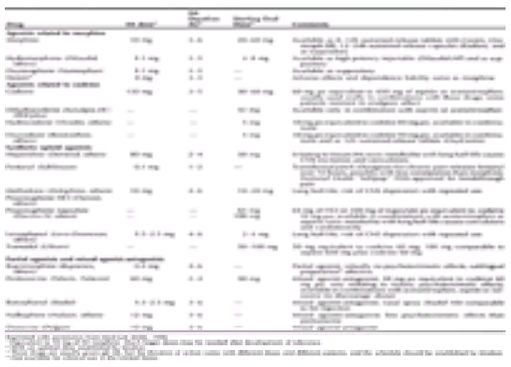

Table 28.6-4 Opioid Analgesics

TOLERANCE Tolerance to opioids develops with chronic use and is experienced first as a shorter duration of analgesia and later as a decreased effectiveness of each dose. Because tolerance to the CNS and respiratory depressive effects develops roughly simultaneously with the tolerance to the analgesic effects, the dosage

may safely be gradually escalated to achieve a constant degree of analgesia. Tolerance occurs independently for each opioid receptor, and clinically the cross-tolerance between full agonists may not be complete. The risk of tolerance may be reduced by using the lowest effective dosage of opioid in combination with nonopioids, and if tolerance develops, it may sometimes be overcome by changing to another full agonist, starting with half or less of the customary equianalgesic dosage.

DEPENDENCE Physical dependence is revealed by abrupt discontinuation of a strong opioid or by administration of an opioid antagonist. It may appear after only a few days of opioid use, but clinically relevant dependence results only from use of a morphinelike opioid for several weeks. Patients who use opioids for relief of acute pain or cancer pain rarely experience euphoria and only rarely develop dependence to the mood-altering effects of opioids.

ADVERSE EFFECTS The most common adverse effects include sedation, grogginess, dizziness, nausea, vomiting, and constipation. The most serious adverse effect is respiratory depression. Strategies include giving smaller, more frequent doses; using opioids with a shorter half-life; and augmentation with a stimulant such as dextroamphetamine (Dexedrine) or methylphenidate (Ritalin). Tolerance usually develops rapidly to the sedative and emetic effects, but constipation is persistent, and a bowel regimen should be given early in the course of treatment. Respiratory depression and apnea are more common in patients with chronic obstructive pulmonary disease, cor pulmonale, decreased respiratory reserve, or preexisting respiratory depression, even with usual doses of opioids. Tolerance to the respiratory depressive effects develops rapidly, but patients taking opioids are more sensitive to the respiratory depressant effects of general anesthetics, phenothiazines, sedative-hypnotics, tricyclic drugs, or other CNS depressants.

DOSAGE STRATEGIES The effective dosage of opioids varies widely among patients, and it is always necessary to titrate the dosage between the need for analgesia and the appearance of adverse effects. Once a strong opioid is chosen, its dosage should be increased to the point of toxicity prior to switching to another opioid. Once the adequate dosage has been determined, opioids should be given on a regular, around-the-clock schedule rather than on an as-needed basis. This schedule may reduce the overall dosage of opioids compared with that needed to treat the reemergence of pain. A supplementary dose should be available as needed between regular doses, for breakthrough pain.

Representative Opioid Drugs

MORPHINE Morphine was the first opiate to be discovered and is still the standard of efficacy for strong injectable analgesics. Oral formulations undergo extensive first-pass metabolism but are also highly effective, particularly in sustained-released preparations (MS Contin). Oral morphine is used widely for treatment of cancer pain, but fears about dose escalation necessitated by tolerance and about the development of dependence sharply limit its use for treatment of chronic, noncancer pain. Physician surveys have documented a solemn reluctance to prescribe high doses of opioids, which could cause death by respiratory suppression. This level of caution exceeds that for numerous other potentially hazardous therapies, and it shows the discomfort physicians have with pain management. When used appropriately, opioids rarely cause death. Morphine has no more liability for dependence than do equally effective dosages of any full opioid receptor agonist.

MEPERIDINE Because of the accumulation of a long-acting toxic metabolite, normeperidine, meperidine (Demerol) should be used only for the treatment of acute pain. Normeperidine may cause dysphoria, irritability, tremors, myoclonus, and seizures, especially with large repeated doses or in patients with renal insufficiency. Administration of meperidine to a patient using a monoamine oxidase inhibitor (MAOI) may cause severe encephalopathy and death.

METHADONE AND LEVORPHANOL Oral methadone (Dolophine) and sometimes levorphanol (Levo-Dromoran) are used to treat chronic cancer pain, especially in patients who are tolerant to morphine. Because of their long half-lives, 24 to 36 hours and 12 to 16 hours, respectively, significant amounts can accumulate with repeated dosing, especially in elderly or debilitated patients, which may cause CNS depression. Access to methadone is tightly regulated by state agencies because of its use as a treatment for opioid dependence.

TRAMADOL The oral opioid agonist tramadol (Ultram) also blocks the reuptake of norepinephrine and serotonin. Its effectiveness is comparable to that of combinations of aspirin or acetaminophen with codeine or propoxyphene. Tramadol may cause seizures, especially in patients concomitantly taking an antidepressant, an MAOI, or a dopamine receptor antagonist (typical antipsychotic). Tramadol is not scheduled as a controlled substance by the U.S. Drug Enforcement Agency, but opioid-type dependence has occurred.

PARTIAL AGONISTS AND MIXED AGONIST-ANTAGONISTS The partial agonist buprenorphine (Buprenex) and the mixed agonist-antagonists pentazocine (Talwin), butorphanol (Stadol), nalbuphine (Nubain), and dezocine (Dalgan) all have a therapeutic ceiling and are comparable to moderate dosages of full agonists. They can precipitate opioid withdrawal in patients physically dependent on full agonists. Their risk of dependence is less than that of full agonists but may be clinically relevant. Only pentazocine is available for oral use, in combination with acetaminophen or aspirin. Butorphanol is available in a nasal spray for treatment of acute migraine, but the development of dependence and drug abuse with this formulation is a growing problem.

Nicotinic Receptor Agonists Nicotine has antinociceptive properties, but its low activity and significant adverse effect profile make it unsuitable for this purpose in humans. Nicotinic acetylcholine receptors are assembled from various combinations of a, b, d, and g subunits. The receptor subtypes in the CNS responsible for pain relief differ from those that mediate the undesirable effects, which include those at the neuromuscular junction and in sympathetic ganglia. A nicotine analogue, epibatidine, isolated from the skin of wild frogs, is 200 times more potent an analgesic than morphine in animals, and its effects are not antagonized by opioid receptor antagonists, suggesting that it acts on a distinct neurotransmitter system. Epibatidine is not suitable for use in humans because it causes seizures and even death. Other adverse effects of epibatidine are mediated in part through non-CNS nicotinic acetylcholine receptors.

A derivative of epibatidine, ABT-594, was as effective an analgesic as morphine for both acute and chronic pain in early clinical trials and it lacked many of the undesirable effects of opioids. ABT-594 inhibits neuronal activity in the dorsal horn of the spinal cord produced by noxious stimuli, yet it has no effect on perception of innocuous pressure or thermal stimuli. Unlike opioids, which cause respiratory depression and sedation, ABT-594 has little effect on respiration and actually makes animals more alert. Both opioids and ABT-594 reduce the appetite, but ABT-594 causes fewer gastric adverse effects and has shown no withdrawal syndrome in at least one animal model of addiction. Whether the nicotine-like qualities of ABT-594 will produce addiction in humans similar to that produced by tobacco products is not yet known. Clinical trials of the safety of ABT-594 have begun in Europe. Testing of additional epibatidine-like drugs is a high priority of the pharmaceutical industry.

Cannabinoids D-9-Tetrahydrocannabinol (D-9-THC), the active ingredient of marijuana, has relieved pain inconsistently in humans. For example, marijuana smoking has long been used to treat chronic pain and nausea due to cancer. D-9-THC is available for prescription as dronabinol (Marinol). In animals, D-9-THC and other cannabinoid drugs cause significant motor impairment at analgesic doses. Cannabinoids occupy receptors also bound by endogenous cannabinoids. Endogenous cannabinoids share an analgesia-promoting brainstem circuit with opioids, which includes the rostral ventromedial medulla. Recently, inactivation of the rostral ventromedial medulla was shown to abolish the analgesic, but not the motor effects of cannabinoids. Cannabinoids and opioids modulate the rostral ventromedial medulla in pharmacologically dissociable ways to produce analgesia. Therefore, cannabinoid receptor agonist activity in the rostral ventromedial medulla could conceivably relieve pain without affecting movements, although further studies are needed to confirm this hypothesis. Based on this model, intensive drug development may eventually create a rostral ventromedial medulla–specific cannabinoid analgesic that is safe for humans.

Adjuvant Drugs The nociceptive element of chronic pain that is responsive to NSAIDs and opioids may be only a part of the entire pain experience. Neuropathic pain is generally less responsive to NSAIDs and opioids but may be effectively treated with antidepressants and anticonvulsants. The affective component of chronic pain may respond to antidepressant or anxiolytic agents. Any underlying psychiatric disorders should be treated appropriately.

Antidepressants Numerous studies have found an association between chronic pain and affective spectrum disorders such as depressive and anxiety disorders. Depression may exacerbate and distort the perception of nociceptive pain. Cognitive factors typical of depression, such as pessimism, learned helplessness, and negative automatic thoughts, may be ameliorated with antidepressant therapy. Antidepressant drugs are among the safest and most effective agents for the treatment of chronic pain. Sleep disturbances associated with chronic pain may respond to antidepressants with sedative properties, such as trazodone (Desyrel), nefazodone (Serzone), mirtazapine (Remeron), paroxetine (Paxil), or tricyclic drugs. Tricyclic antidepressants have been used extensively for treatment of constant neuropathic pain, but they have more serious adverse effects and offer no therapeutic advantage over newer antidepressants. Newer antidepressants may be more efficacious for neuropathic pain because higher dosages may be used without sedation or anticholinergic effects.

Anticonvulsants Lancinating pains typical of trigeminal neuralgia, postherpetic neuralgia, nerve trauma, or infiltration of nerves by cancer may respond to anticonvulsants, such as carbamazepine (Tegretol), phenytoin (Dilantin), clonazepam (Klonopin), and gabapentin (Neurontin).

Caffeine Caffeine is not an analgesic by itself, but in controlled trials it potentiated the analgesic effects of acetaminophen, aspirin, and ibuprofen for headache, oral surgery pain, and postpartum pain. A widely used migraine headache treatment contains a combination of butalbital (a barbiturate for sedation) and caffeine plus aspirin (Fiorinal) or acetaminophen (Fioricet, Esgic).

Antihistamines Hydroxyzine (Atarax, Vistaril) and other first-generation H_1 receptor blockers may increase the analgesic effect and reduce the nausea and vomiting of opioids in postoperative and cancer pain.

Corticosteroids In painful chronic inflammatory conditions such as arthritis or infiltration of nerves by tumor, the anti-inflammatory effects of corticosteroids may have analgesic effects. Long-term use of steroids is associated with weight gain, immunosuppression, peptic ulcer disease, skin changes, euphoria, insomnia, mood swings, personality changes, depression, psychosis, electrolyte abnormalities, loss of muscle mass, cataracts, glaucoma, and adrenocortical insufficiency (on precipitous withdrawal).

Antipsychotic Agents Chlorpromazine (Thorazine) and other low-potency dopamine receptor antagonists are effective in the acute management of migraine headache refractory to standard analgesics. This class of drugs is not commonly used for treatment of chronic pain, but it is useful for patients whose pain is part of a delusional complex.

SUGGESTED CROSS-REFERENCES

The neuropsychiatric aspects of musculoskeletal disorders and rheumatoid arthritis are discussed in Section 25.8. Management of cancer pain is discussed under psycho-oncology in Section 25.11. Pain syndromes associated with acquired immune deficiency syndrome (AIDS) are discussed in Section 2.7. Irritable bowel syndrome, peptic ulcer disease, and inflammatory bowel disease are discussed in Section 25.2. Angina is discussed in Section 25.4. Head trauma, which may lead to chronic headaches, is discussed in Section 2.5. Painful movement disorders are discussed in Section 2.6. Peripheral neuropathy is discussed in Section 2.11. HIV-associated neuropathy is discussed in Section 2.8. Central pain syndromes associated with multiple sclerosis are discussed in Section 2.7. Thalamic lesions due to stroke and tumors are discussed in Section 2.2.

SECTION REFERENCES

Adams RD, Victor M, Ropper AH: *Principles of Neurology*, ed 6. McGraw-Hill, New York, 1998.

*American Geriatrics Society Panel on Chronic Pain in Older Persons: The management of chronic pain in older persons. J Am Geriatr Soc *46:*635, 1998.

Amir M, Kaplan Z, Neumann L, Sharabani R, Shani N, Buskila D: Posttraumatic stress disorder, tenderness and fibromyalgia. J Psychosom Res *42:*607, 1997.

Bannon AW, Decker MW, Holladay MW, Curzon P, Donnelly-Roberts D, Puttfarcken PS, Bitner RS, Diaz A, Dickenson AH, Porsolt RD, Williams M, Arneric SP: Broad-spectrum, non-opioid analgesic activity by selective modulation of neuronal nicotinic acetylcholine receptors. Science *279:*77, 1998.

Beckham JC, Crawford AL, Feldman ME, Kirby AC, Hertzberg MA, Davidson JR, Moore SD: Chronic posttraumatic stress disorder and chronic pain in Vietnam combat veterans. J Psychosom Res *43:*379, 1997.

Butcher JN, Williams CL: *Essentials of MMPI-2 and MMPI-A Interpretation*. University of Minnesota Press, Minneapolis, 1992.

Catalano EM, Hardin KN: *The Chronic Pain Control Workbook*, ed 2. MJF Books, New York, 1996.

*Craig AD, Reiman EM, Evans A, Bushnell MC: Functional imaging of an illusion of pain. Nature *384:*258, 1996.

Crow CS: Childrens' Pain Perspectives Inventory (CPPI): Development assessment. Pain *72:*33, 1997.

Dolin SJ, Padfield NL, Pateman JA: *Pain Clinic Manual*. Butterworth-Heinemann, Oxford, 1996.

Fishbain DA, Cutler R, Rosomoff HL, Rosomoff RS: Chronic pain-associated depression: Antecedent or consequence of chronic pain? A review. Clin J Pain *13:*116, 1997.

Flor H, Elbert T: Maladaptive consequences of cortical reorganization in humans. Neurosci News *1:*4, 1998.

Galer BS, Schwartz L, Turner JA: Do patient and physician expectations predict response to pain-relieving procedures? Clin J Pain *13:*348, 1997.

Kaas JH, Jain N, Florence SL: The reactivation of somatosensory cortex after deactivation by peripheral nerve or spinal cord injury. Neurosci News *1:*12, 1998.

Kalgutkar AS, Crews BC, Rowlinson SW, Garner C, Seibert K, Marnett LJ: Aspirin-like molecules that covalently inactivate cyclooxygenase-2. Science *280:*1268, 1998.

Kerns RD, Turk DC, Rudy TE: The West Haven-Yale Multidimensional Pain Inventory. Pain *23:*345, 1985.

Kramer MS, Cutler N, Feighner J, Shrivastava R, Carman J, Sramok JJ, Reines SA, Liu G, Snavely D, Wyatt-Knowles E, Hale JJ, Mills SG, MacCoss M, Swain CJ, Harrison T, Hill RG, Hefti F, Scolnick EM, Cascieri MA, Chicci GG, Sadowski S, Williams AR, Hewson L, Smith D, Carlson EJ, Hargreaves RJ, Rupniak NMJ: Distinct mechanism for antidepressant activity by blockade of central substance P receptors. Science *281:*1640, 1998.

McCracken LM: Learning to live with the pain: Acceptance of pain predicts adjustment in persons with chronic pain. Pain *74:*21, 1998.

Melzack R, Wall PD: Pain mechanisms: A new theory. Science *150:*971, 1965.

Meng ID, Manning BH, Martin WJ, Fields HL: An analgesia circuit activated by cannabinoids. Nature *395:*381, 1998.

Portenoy RK: Chronic opioid therapy in nonmalignant pain. J Pain Symptom Manage *5*(Suppl):S46, 1990.

*Rosenfeld B, Breitbart W, Stein K, Funesti-Esch J, Kaim M, Krivo S, Galietta M: Measuring desire for death among patients with HIV/AIDS: The schedule of attitudes toward hastened death. Am J Psychiatry *156:*94, 1999.

Sarno JE: *Healing Back Pain*. Warner Books, New York, 1991.

Sarno JE: *The Mindbody Prescription*. Warner Books, New York, 1998.

Schwartz L, Slater MA, Birchler GR: The role of pain behaviors in the modulation of marital conflict in chronic pain couples. Pain *65:*227, 1996.

*Singer PA, Martin DK, Kelner M: Quality end-of-life care: Patients' perspectives. JAMA *281:*163, 1999.

*Tolle TR, Kaufmann T, Siessmeier T, Lautenbacher S, Berthele A, Munz F, Zieglgansberger W, Willoch F, Schwaiger M, Conrad B, Bartenstein P: Region-specific encoding of sensory and affective components of pain in the human brain: A positron emission tomography corrrelation analysis. Ann Neurol *45:*40, 1999.

Tollison CD, Satterthwaite JR, Tollison JW: *Handbook of Pain Management*, ed 2. Williams & Wilkins, Baltimore, 1994.

Turk DC, Okifuji A: What factors affect physicians' decisions to prescribe opioids for chronic noncancer pain patients? Clin J Pain *13:*330, 1997.

*Turk DC, Rudy TE: Towards a comprehensive assessment of chronic pain patients. Behav Res Ther *25:*237, 1987.

Wahlestedt C: Reward for persistence in substance P research. Science *281:*1624, 1998.

Waldman SD, Winnie AP: *Interventional Pain Management*. Saunders, Philadelphia, 1996.

*Wall PD, Melzack R: *Textbook of Pain*, ed 3. Churchill Livingstone, Edinburgh, 1994.

Warfield CA: *Principles and Practice of Pain Management*. McGraw-Hill, New York, 1993.

*Weiner RS: *Pain Management: A Practical Guide for Clinicians*, ed 5. St. Lucie Press, Boca Raton, FL, 1998.

Woolf CJ: Windup and central sensitization are not equivalent. Pain *66:*105, 1996.

Woolf CJ, Mannion RJ, Neumann S: Null mutations lacking substance: Elucidating pain mechanisms by genetic pharmacology. Neuron *20:*1063, 1998.

Zeltzer L, Bursch B, Walco G: Pain responsiveness and chronic pain: A psychobiological perspective. Dev Behav Pediatr *18:*413, 1997.

28.7 PHYSICAL AND SEXUAL ABUSE OF ADULTS

BESSEL A. VAN DER KOLK, M.D.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) specifies five problems related to abuse or neglect: (1) physical abuse of child, (2) sexual abuse of child, (3) neglect of child, all discussed elsewhere in this textbook, and (4) physical abuse of adult and (5) sexual abuse of adult, both discussed here. (Table 28.7-1). Physical abuse of adult includes spouse or partner abuse and abuse of elderly persons. Sexual abuse of adult includes rape, sexual coercion, and sexual harassment.

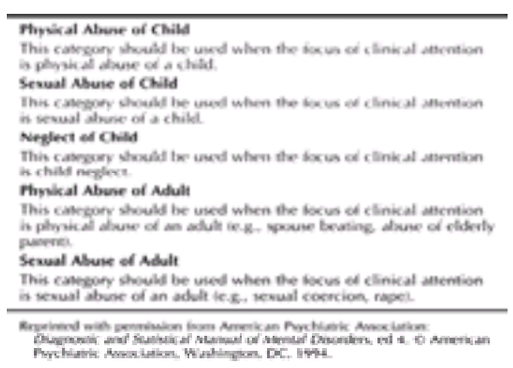

Physical Abuse of Child
This category should be used when the focus of clinical attention is physical abuse of a child.
Sexual Abuse of Child
This category should be used when the focus of clinical attention is sexual abuse of a child.
Neglect of Child
This category should be used when the focus of clinical attention is child neglect.
Physical Abuse of Adult
This category should be used when the focus of clinical attention is physical abuse of an adult (e.g., spouse beating, abuse of elderly parent).
Sexual Abuse of Adult
This category should be used when the focus of clinical attention is sexual abuse of an adult (e.g., sexual coercion, rape).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 28.7-1 DSM-IV Problems Related to Abuse or Neglect

Violence is probably the most important public health problem in the United States. Violent deaths, including those caused by accidents and interpersonal violence, are the third leading cause of death, after cardiovascular disease and cancer. However, when one considers the amount of expected life lost by specific causes of death, violent deaths are responsible for more years of expected life lost than any other cause of death. Accidents account for about 60 percent of violent deaths, suicides for 25 percent, and homicides for 15 percent. The morbidity resulting from violence is also considerable. This not only concerns the physical injuries and disabilities that result from interpersonal violence, but also the psychological sequelae that follow in its wake.

With an annual homicide incidence of 7.9 per 1000 citizens, the United States is considered the most violent industrialized nation in the world. One study estimated that 83 percent of Americans will fall victim to a violent crime (Table 28.7-2) at some point in their lives, and about 25 percent will become victims of three or more violent crimes. Overall, at least 15 percent of the population report having been molested, physically attacked, raped, or involved in combat. Men are physically assaulted more than women (11.1 vs. 10.3 percent), while women report higher rates of sexual assault (7.3 vs. 1.3 percent). More whites (18.4 percent) than African-Americans (11.6 percent) report having been physically assaulted, even though the death rate due to criminal assault is four to five times as high in black males and females than in whites. Teenagers and young adults are most vulnerable to becoming victims of violent crime; in 1996, about a third of all victims of violent crime were ages 12 to 19. Almost half of all victims of violence were under age 25. These different forms of violence have both short- and long-term sequelae, affecting both the physical and psychological well-being of the victim.

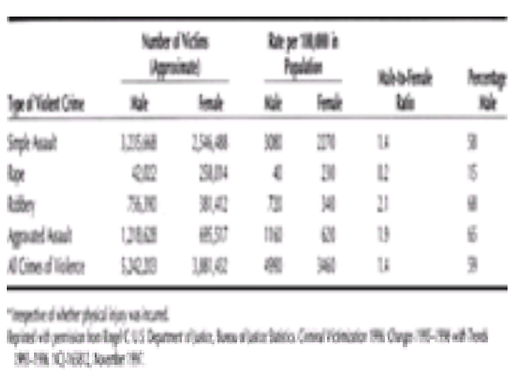

Type of Violent Crime	Number of Victims (Approximate)		Rate per 100,000 in Population		Male-to-Female Ratio	Percentage Male
	Male	Female	Male	Female		
Simple Assault	1,735,660	2,546,480	1080	1270	1.4	58
Rape	42,022	258,014	40	230	0.2	15
Robbery	756,390	381,432	720	340	2.1	68
Aggravated Assault	1,210,628	695,537	1160	620	1.9	65
All Crimes of Violence	3,742,203	3,881,432	4980	3460	1.4	59

*Irrespective of whether physical injury was incurred.
Reprinted with permission from Rand P. U.S. Department of Justice, Bureau of Justice Statistics. Criminal Victimization 1996: Changes 1995–1996 with Trends 1993–1996. NCJ-165812, November 1997.

Table 28.7-2 Victimization by Nonlethal Violent* Crime by Sex, 1996

High levels of violence are accompanied by high patterns of medical utilization; at least one in five women seen in emergency departments has symptoms related to abuse. In emergency rooms, patients are generally only symptomatically treated for their injuries, which are rarely recognized and treated as related to abuse. Even when recognized, medical staff are often at a loss about what to do and are unaware of resources to address the needs of traumatized individuals.

EPIDEMIOLOGY

Males account for the vast majority of both perpetrators and victims of violence. In 1997 the male rate of death by violence was 10.7/100,000; the female rate was 3.0 per 1000. The lifetime risk of becoming a victim of criminal homicide for men is 1 in 84, compared with 1 in 282 for women. The male suicide rate in 1995 for men was 19.8, for women, 4.4. Each year over a million completed or attempted robberies and between 4 and 5 million assaults are reported. These two crimes lead to over 75,000 injuries each year that are severe enough to require hospital admission. The male rate of victimization by these crimes in 1996 was 49.5 per 1000, compared with 32.3 per 1000 for women. Men sustain twice as many severe injuries as women do.

Despite the prevalence of male victimization, women are overrepresented in virtually all forms of research on victimization: research on the psychological response of male victims is virtually absent. This has a variety of causes, including the likelihood that men are more prone than women to regard the victim role as inconsistent with culturally defined self-perceptions. Research has shown that men are much less likely to report their victimization to either their families or the police and that women seek psychiatric help for the psychological effects of interpersonal victimization far more often than men do. James Gilligan has pointed out that “in forcing the victim to experience utter helplessness, victimization challenges a very basic component of male identity.” He describes how shame becomes the central psychological way in which male victims cope with the aftermath of interpersonal assault and victimization.

Domestic Violence For women (but not for men) violence perpetrated within relationships that are supposed to involve care and protection is a much more serious problem than violence perpetrated by strangers. In 1994 62 percent of the almost 3 million attacks on women were by persons whom they knew, while 63 percent of the almost 4 million assaults on males were by strangers. It is estimated that one in three Americans have witnessed an incident of domestic violence. Over a third of

the victims of domestic assault experienced serious injury, compared with a quarter of the victims of stranger assault. Domestic abuse and child abuse are closely related; in homes where spousal abuse occurs, children are abused at a rate 1500 percent higher than the national average.

Between 1993 and 1997 approximately 10,000 people over the age of 15 in the United States were killed by their domestic partners (including married, common-law, divorced, or dating partners); 70 percent of the victims were women and 30 percent were men. Approximately one half of women murdered in the United States are killed by a current or former male partner. It is estimated that about one third of emergency room visits by women are in response to partner violence. Even though in approximately half of violent couples the wives are also violent, the women virtually always are more terrified and more likely to be seriously injured. Women are more likely to be injured by their male partners than men are by their female partners. About one fifth of pregnant women are battered.

It is estimated that from 2 to 3 million women are assaulted by a male partner each year in the United States. Between a quarter and half of all women experience some form of physical violence from their partners during marriage. Victims of marital violence have a higher rate of unconsciousness and internal injuries than victims of stranger assault; in one study, over 80 percent of all assaults against spouses and ex-spouses resulted in injuries, compared with 54 percent of the victims of stranger violence.

After separation women continue to be at risk. One half of battered women who leave their relationships continue to experience violence from their estranged partners. Violence also is widespread in dating relationships. The prevalence of physical aggression is even higher among nonmarried couples than among married couples.

From 1976 through 1984 the overall numbers of women killing male partners decreased more than 25 percent. This drop was linked with the presence of domestic violence legislation and other services for abused women. This decrease was not matched by a similar decline in men killing female partners.

Rape *Rape* is defined as the perpetration of an act of sexual intercourse against the victim's will and consent, whether her will is overcome by force or fear resulting from threat of force or by drugs or intoxicants, or when mental deficiency renders her incapable of exercising rational judgment, or when she is below an arbitrary age of consent. Rape can occur between married persons and persons of the same sex. Legally, to be classified as rape, only slight penile penetration of the victim's outer vulva is required; full erection and ejaculation are not necessary. Forced acts of fellatio and anal penetration, although they frequently accompany rape, are legally considered sodomy.

In 1996, 197,000 American women were estimated to have been forcibly raped. Of these, only 16 percent were reported to the police. One of every eight adult women, or at least 12.1 million American women, will be the victim of forcible rape sometime in her lifetime. In the United States more than 6 out of 10 of all rape cases (61 percent) occur before victims reach age 18. Among U.S. adolescents aged 12 to 17, 8 percent are estimated to have been victims of serious sexual assault, 17 percent victims of serious physical assault, and 40 percent have witnessed serious violence. Twenty-nine percent of all forcible rapes occur before the age of 11. Among adult women, an estimated 32,101 pregnancies result from rape each year.

Sexual assault is also widely prevalent in domestic settings. James Kilpatrick found that only 22 percent of rapes were perpetrated by strangers. Husbands and boyfriends were responsible for 19 percent, and other relatives accounted for 38 percent.

Perpetrators Most rapists commit other violent crimes as well and use sex to dominate, hurt, and debase their victims. Sexual violence and physical violence often occur in the same individual. One survey found that almost 13 percent of the men in a large epidemiological sample had been physically aggressive toward their partners in the past year. More than 33 percent of those aggressive acts were severe, consisting of punching, beating, or threatening with a knife or gun. Gilligan has pointed out that the more hurt, humiliated, and downtrodden men feel, the more likely they are to perpetrate violence, particularly physical violence against intimates, the women and children in their life.

Men who assault female partners are more likely to have witnessed or experienced violence in childhood, to abuse alcohol, to be sexually assaultive to their wives, and to be at risk for violence against children than nonabusive males. Many jealous husbands and boyfriends are charming to everyone except their wives and girlfriends. They monitor the woman's every move and tend to prevent contact with other sources of support and protection. They are prone to interpret their behavior as betrayal and this infuriates them. When their anger explodes into violence, they lose control. In contrast with dependent men who are primarily afraid of being abandoned by their partners and try to ensure continued submission, others are sociopaths who become physiologically less aroused when their violence escalates. These sociopaths are cold and calculating con artists, relatively free of the trappings of emotional dependence, but with a high incidence of antisocial and criminal traits and sadistic behavior. They use aggression with the intent to control, subjugate, or intimidate their partners.

In these abusive relationships emotional abuse tends to continue even after physical attacks abate, eventually keeping the spouse intimidated without the use of force. Battered women often have complex relationships with their abusers. Deprived of alternative sources of care, they tend to seek refuge with the nearest available human being when frightened. When the perpetrator also is seen as the principal source of protection, victims often become protective of their batterers and form intense emotional attachments that may be hard to break, even when ample objective evidence shows that continued contact with the perpetrator constitutes a danger to the victim. Only about one third of battered women in one study were able to escape from their abusive relationships within a 2-year follow-up period. After 5 years, 65 percent of battered women had left. Battering rarely stops on its own. A combination of arrest, prosecution, fines, and counseling often works for nonsociopathic batterers.

ETIOLOGY

Age In the United States, most rape victims (61 percent) are below age 18; almost a third are below age 11. Since the mid-1980s the rate of violence among persons aged 15 to 17 has increased sharply, and, particularly among black youth in this age range has outpaced changes in violence in all other age groups. In 1996 about a third of all victims of violent crime were ages 12 to 19. Almost half of all victims of violence were under age 25. During a time span of 4 years, 94 percent of a group of men aged 20 to 29 had to go to an emergency room at least once. In 41 percent of the cases the cause was a violent encounter. Violent crime declined slightly during the 1990s.

Prior Exposure to Family Violence Witnessing domestic violence as a child is the single most important risk factor for the intergenerational transmission of family violence. A history of physical or sexual victimization in childhood is an important risk factor for later assault, regardless of the relationship to the perpetrator in these later assaults. Both victims of partner and stranger assault are nearly twice as likely to experience abuse in childhood than nonvictims. A survey of over 4000 adults found that witnessing marital aggression between parents was the single most important predictor of which adolescent girls would grow up to be victims of marital assault. Males who witnessed their fathers' violence as children were 10 times more likely to grow up to abuse their partners than boys not exposed to marital violence.

Sex Domestic abuse and sexual violence against women are enormous social problems. Most violence against women is perpetrated by an intimate partner; approximately one third of all women are assaulted by an intimate man during adulthood. Witnessing domestic violence markedly increases the sense of vulnerability in all children. However, gender differences exist in the ways that children respond to witnessing domestic violence; girls tend to become afraid of being attacked, while boys, though feeling fearful of being violated, tend to display behaviors indicating that they identify with the attackers.

Children who are traumatized by witnessing domestic violence tend to display pronounced, inappropriate attitudes about violence as a means of resolving conflict. They may come to expect it as a way of life. Witnessing the chronic helplessness of their mother and their father's alternating outbursts of affection and violence, they learn that people have no control over their impulses. As adults, they may attempt to undo the past by love, competency, and exemplary behavior. However, if this fails, their lack of experience with nonviolent resolution of differences makes them vulnerable to being unable to resolve differences by verbal communication. During periods of occupational stress or family transitions such as pregnancies and the birth of children, partners who come from violent homes may return to earlier coping mechanisms, such as self-blame, numbing (by means of emotional withdrawal or drugs and alcohol), or physical aggression.

Substance Abuse Between 5.9 and 6.9 million adult women in the United States are estimated to have a substance use disorder. Substance abuse is highly correlated with domestic assaults, both on the part of the victim and the perpetrator. Active substance use is associated with increased risk of victimization. Conversely, assault has been shown to lead to substance use. Risk of new victimization is greatest in women who already use drugs and who also have previous assault histories. Substance abuse use occurs in a context that increases the likelihood for victimization and is associated with a general propensity to affiliate with assault-prone individuals or to engage in risk-taking activities.

Socioeconomic Factors Poverty (but not race) is a strong risk factor for exposure to assaults and other traumas. For women, being with a partner is a risk factor for assault.

CLINICAL FEATURES, COURSE, AND PROGNOSIS

Victims of interpersonal assaults have many of the same reactions as victims of other types of trauma. During an assault, a victim's primary focus is on survival and self-protection. The subjective experience consists of a mixture of shock, confusion, numbness, withdrawal, and speechless terror. Some victims may put up fierce resistance, while others dissociate and offer no defense. The number of and severity of psychological symptoms is largely determined by the frequency and duration of the assault. The more prolonged the abuse, the more likely the victim is to react with numbing and dissociation. When the aggressor is someone the victim loves, trusts, and depends on (e.g., victims of intrafamilial abuse), the victim is likely to respond to assaults with increased dependence and paralysis in their decision-making processes. Unlike people who are brutalized by strangers, victims of marital violence have legal, financial, and role relationships with their assailants, which confounds their capacity to respond appropriately to the violence. They may feel particularly vulnerable, betrayed, and hopeless. They often become depressed and suicidal.

In the emergency room they display cuts, burns, bruises, sprains, black eyes, broken ribs, and lost teeth. In primary care settings they complain of chronic headaches, muscle aches, abdominal pains, recurrent vaginal infections, sleep and eating disorders, and depression.

When the perpetrator is a spouse, boyfriend, or acquaintance (Table 28.7-3), many victims are reluctant to label rape or physical assault as such. Many victims feel emotionally dependent on their assailant (generally, the greater the fear, the greater the feeling of dependence and fear of abandonment). Many victims are justifiably afraid of retribution by an assailant. Many victims are familiar with being blamed for the abuse, and they fear the stigma attached to being a victim of rape or domestic violence. Many victims have difficulty recalling all relevant aspects of violent interpersonal trauma. Indeed, the inability to recall important aspects of the event is part of the definition of posttraumatic stress disorder.

Table 28.7-3 Physician Reference Card

Like rape victims, women abused by male partners exhibit long-term reactions including fear, anxiety, fatigue, sleeping and eating disturbances, intense startle reactions, and physical complaints. They often continue to dissociate in the face of threat, appear spaced-out, and have difficulty imagining effective action. They often develop problem-focused coping, in which the goal is making the feelings go away rather than changing the objective situation. This emotion-focused coping accounts for the fact that persons who develop posttraumatic stress disorder are vulnerable to alcohol and other substance abuse. Between a quarter and a half of all patients who seek substance abuse treatment are comorbid with posttraumatic stress disorder. The relation between substance abuse and posttraumatic stress disorder is reciprocal: drug abuse leads to assault and assault leads to use. Research has shown that new assaults lead to significant role-escalating alcohol and drug use in women.

Not all victims of interpersonal assault develop posttraumatic stress disorder. Rape is a trauma that gives rise to one of the highest rates of any traumatic event, approximately 70 percent. After physical assaults approximately a quarter of victims develop posttraumatic stress disorder, and more than half of various shelter populations currently suffer from it. Once victims develop the disorder, the intense flashbacks and nightmares force them to cope with the constant recurrence of the memory of the trauma with little anticipation of relief. In contrast to the actual trauma, which is terrifying, the symptoms of posttraumatic stress disorder seem never ending. The recurrent intrusive recollections and the nightmares themselves become new triggers of panic, which leads to a variety of (usually maladaptive) avoidance maneuvers. Avoidance and numbing further alienates victims from contact with the social support essential for recovery. Thus the very attachments that can provide a powerful motivation for survival in the face of threat can be threatened by the secondary changes in the responsiveness of the victim with posttraumatic stress disorder. The disorder of attention and concentration means that they can no longer lock into their current environment with the same sense of energy and involvement; simple activities such as reading, conversing, and watching television are no longer reflex activities but demand greater effort. This further interferes with effective functioning and with the capacity to make the decisions necessary to put their life back on track. Once people develop posttraumatic stress disorder, they are most likely to suffer from depression as well. Sexual assault is associated with an increased lifetime rate of attempted suicide. A history of sexual trauma before age 16 years is a particularly strong correlate of attempted suicide.

Physical Health Consequences Approximately one third of sexual assault victims suffer some form of injury. Physicians frequently treat the injuries only symptomatically or fail to recognize the injuries as abuse. Given the high morbidity and mortality rates in domestic violence, failure to identify or intervene increases the likelihood of further physical injury or death.

Use of medical services increases sharply in the year following a rape. In one study, medical care was sought by 92 percent of crime victims during the first year following the crime and by 100 percent during the second year. There was an 18 percent increase in utilization over prerape levels during the first year, 56 percent during the second year, and during the third year service use was 31 percent above prerape levels.

Pregnancy as a Result of Rape The exact incidence of pregnancy after rape is unknown; however, an estimated 1 to 10 percent of rapes result in pregnancy. Not only is the report of rape often delayed, but reports of pregnancy are delayed even further. Many delayed reporters are adolescent victims, especially if the rape was incestuous. They conceal the pregnancy to protect the family from stress or because of fear of the abuser. Pregnancy occurs more often among incest victims than among women who have been raped by a nonfamily member, probably because the former may occur on a regular basis increasing the risk of pregnancy. About 10 to 20 percent of incest victims become pregnant.

Male Rape In some states, the definition of rape has become gender neutral: the word “person” has been substituted for the term “female”; homosexual rape is included in the definition, and the use of an object, in place of a penis, for penetration also constitutes sexual assault. Male rape is a significantly underreported crime.

Homosexual rape is much more frequent among men than among women, and it occurs regularly in closed institutions, such as prisons and maximum security hospitals. Estimates are that 0.5 to 3 percent of male inmates are sexually assaulted. No figures are available for nonincarcerated persons.

The attacks are violent, rather than sexual, and are used to express anger or power. The dynamics are identical to those involved in heterosexual rape. The crime enables the rapist to discharge aggression and to aggrandize himself. The victim is usually smaller than the rapist. He is always perceived as passive and unmanly (weaker) and is used as an object.

The rapist selecting a male victim may be heterosexual, bisexual, or homosexual. He forces his victim to have sex through entrapment (for example, by getting him drunk), through intimidation by threats with a knife or gun, or through the use of brute force. Male rape can involve one or several forced sexual acts: The most common act is anal penetration of the victim; the second most common act is forcing the victim to perform fellatio. Frequently, the rapist makes an effort to bring the victim to ejaculation by fellatio or masturbation. That condition is humiliating to the victim, and it reinforces the rapist's sense of conquest and his fantasy that the victim wanted or enjoyed the rape. The male rape victim often feels, as do the raped women, that he has been ruined. In addition, he often fears he will become

homosexual because of the attack.

Date Rape *Date rape* or *acquaintance rape* is a term applied to situations in which the rapist is known to the victim. The rape can occur with a new boyfriend, on a first date, on a ride home after a party, or when the man and woman have known each other for many months. Considerable data on this type of rape have been gathered from college populations, and many schools have set up counseling programs to deal with these assaults.

A survey of 500 students at one university found that 16 percent of the women had been raped by men they knew or were dating. Of the men interviewed, 11 percent stated they had committed rape. Those findings are reflected in studies at a number of institutions throughout the United States. In a survey including several universities, 38 percent of the male students said they would commit rape if they thought they could get away with it, and 50 percent of the male students subscribed to myths that predispose to rape tolerance—that women don't mean it when they say “no” to intercourse or that they want to be coerced into having coitus.

In addition to suffering the symptoms experienced by all rape survivors, victims of date rape berate themselves for exercising poor judgment in their choice of male friends. They are more likely to blame themselves for provoking the rapist (by saying they should have not worn a low-cut dress, for example) than are other victims. Ironically, self-blame aids the recovery of these and other rape victims. Focusing on something they could have done differently seems to restore an element of control to the victim who has been made to feel so helpless. Date rape victims are less likely than others to report the crime to authorities. Also, because victim and rapist frequently have friends in common, the woman fears being disbelieved and being ostracized by her social group if she talks about the attack.

Among very young teenagers who have had intercourse, the majority said they had done so at times involuntarily. Nearly three-quarters of women who had intercourse before age 14 said they had at some point been forced to have sex.

Statutory Rape Intercourse is unlawful between a male over 16 years of age and a female under the age of consent, which varies from 14 to 21, depending on the jurisdiction. Thus, if an 18-year-old man and a 15-year-old girl have consensual intercourse, the man may be held for statutory rape. That type of rape may vary drastically from the crimes described above in being nonassaultive and in being a sexual act, not a violent act. Nor is it a deviant act, unless the age discrepancy is sufficient for the man to be defined as a pedophile; that is, when the man is 10 years older than a young girl or the girl is 12 years old or less. Charges of statutory rape are rarely pressed by the consenting girl; they are usually brought by her parents.

Stalking *Stalking* is defined as a pattern of harassing or menacing behavior coupled with a threat to do harm. In 1990 California passed the first antistalking law and most states now prohibit stalking although some will not intervene unless an act of violence has occurred. In states with stalking laws the person can be arrested based upon a pattern of harassment and can be charged with either a misdemeanor or felony. Some stalkers continue that activity for years; others for only a few months. Stalkers may be mandated by the court to undergo counseling sessions. In all instances stalkers should be reported to law enforcement agencies as the best means of deterrent.

Sexual Harassment *Sexual harassment* refers to sexual advances, request for sexual favors, verbal or physical conduct of a sexual nature—all of which are unwelcomed by the victim. In over 95 percent of cases the perpetrator is a man, the victim, a woman. In the case of a man being harassed it is almost always by another man. A woman sexually harassing a man is an extremely rare event. The victim of harassment reacts to the experience in various ways. Some blame themselves and become depressed; others become anxious or angry. In general, harassment most commonly occurs in the workplace and many organizations have developed procedures to deal with the problem. All too often, however, the victim is unwilling to step forward and lodge a complaint because of fear of retribution, of being humiliated, of being accused of lying (which is exceedingly rare), or ultimately being fired from the job.

The types of behavior that make up sexual harassment are broad. They include abusive language, requests for sexual favors, sexual jokes, staring, ogling, and giving massages, among others.

To reduce harassment organizations may distribute material that is educational (Table 28.7-4). Employers are obligated to investigate every complaint, which most often are addressed to the Equal Employment Opportunity Commission. Appropriate organizational responses range from a written reprimand to firing the offender.

WHAT YOU SHOULD KNOW ABOUT SEXUAL HARASSMENT

WHAT IS SEXUAL HARASSMENT?

WHAT IS PROHIBITED?

The 1980 Equal Opportunity Employment Commission Guidelines for Sexual Harassment encompass:

1. Unwelcomed sexual advances.
2. Requests for sexual favors.
3. Verbal conduct of a sexual nature.
4. Physical conduct of a sexual nature.

UNDER WHAT CONDITIONS?

When such conduct has the purpose or effect of:

1. Unreasonably interfering with an individual's work performance.
2. Creating an intimidating, hostile or offensive working environment. (This can be interpreted to include the "terms, conditions or privileges of unemployment" such as the psychological and emotional work environment and subjecting female employees to anxiety and debilitation.)
3. When submission to or rejection of the conduct is made either explicitly or implicitly a term or condition of an individual's employment or
4. Is the basis for employment decisions affecting the individual.

WHO IS RESPONSIBLE?

1. The employer who is committing the act.
2. The employer, even if its agents and supervisors committed the act regardless of whether the acts were authorized or even forbidden.
3. The employer for acts committed by an employee's coworker.
4. The employer for conduct by a non-employee (such as a customer or supplier) if the employer "knows or should have known of the conduct and fails to take an immediate and appropriate action."
5. The employer, if timely corrective actions are not taken.

Reprinted with permission from Tyler Cloverleaf Associates, Philadelphia.

Table 28.7-4 Educational Material to Reduce Sexual Harassment

TREATMENT

The strongest predictors of psychological adjustment after an assault are the degree of injury, severity of life threat, a history of having received emotional support from a nonabusive adult during childhood, current social support, participation in psychotherapy, and educational background.

Early Intervention Assault victims require a multidisciplinary approach encompassing emotional, medical, and forensic care. Appropriate therapeutic action taken early may influence the prognosis for a significant proportion of assault victims. Psychological interventions must pay attention to the particular posttrauma phase of the victim.

Safety When their own resources are inadequate to deal with threat individuals need to rely on others to provide them with safety and care. After an assault victims must establish contact with their natural social support system. If this system is inadequate to ensure the safety of the patient, institutional resources must be mobilized to help the patient find a place to recover. After an assault the victim first needs to be in a place where she is physically safe and where she can rest, sleep, and eat.

Anxiety Management After the patient's safety has been ensured, variety of psychological interventions may be needed. Caregivers may need to help victims identify the problems they face and help them formulate appropriate solutions. Assault victims must learn to distinguish between the real-life threats that often persist and the haunting, irrational fears that are part of posttraumatic stress disorder. If anxiety dominates, victims need help in strengthening their coping skills. Practical anxiety-management skills training may include deep muscle relaxation, breathing control, role playing, covert modeling, thought stopping, and guided self-dialogue.

Emotional Processing To put the assault in perspective, some form of debriefing is often helpful. During this process victims are encouraged to recreate the event in words, individually or in a group. They are asked to articulate what they think happened and what led up to it—their own contributions to what happened, their thoughts and fantasies during the event, what the worst part of it was, and their reactions to it in detail, including how it has affected their perceptions of themselves and others. Such exposure therapy is thought to promote symptom reduction by allowing patients to realize that (1) remembering the trauma is not equivalent to experiencing it again; (2) the experience had a beginning, middle, and end; and (3) the event now belongs to their personal history.

In recent years a new technique, eye movement desensitization and reprocessing, a form of exposure therapy, has been shown to help desensitize patients with posttraumatic stress disorder without fully engaging them in a verbal reliving the traumatic experience. Although exposure therapy can help overcome traumatic intrusions, it needs to be applied cautiously. Some patients recalling their trauma may become flooded with both the traumatic memories and memories of previously forgotten trauma. Increased activation of traumatic memories may be associated with increased shame, guilt, anger, and alcohol and drug use. Patients with

posttraumatic stress disorder are a particularly avoidant group that presents a number of treatment obstacles, especially for exposure techniques.

Prevention Against Recurrence After gaining an understanding of their own vulnerability, victims need to devise a plan to prevent a recurrence. This may involve a geographic move or some other active measure that interrupts the assault–substance use cycle and may reduce the potential for both future victimization and substance abuse. Thus, interventions for women who experience assault should not be limited to attempts to reduce immediate and overt trauma-related symptomatology (e.g., anxiety and depression) but should also address development or exacerbation of substance use. Conversely, treatment of women with substance use disorders, particularly illegal substance abuse, should address the issues of prior victimization and how to reduce the likelihood of future victimization.

Group Treatment Group therapy, including twelve-step self-help programs, can be extremely useful in helping victims move beyond a helpless stance in which they are vulnerable to repeating the trauma, either as victim or as perpetrator. Such groups provide both human attachments and a meaningful cognitive frame for dealing with the sense of helplessness. The purpose of these groups is to help the member regain a sense of trust and belonging by making interpersonal commitments. As long as victims are dominated by feelings of shame and social isolation they are prone to regress to dealing with anxiety with anxious or dependent attachments and addictive behaviors. Groups emphasize living in the here and now and show that responsible adults are faced with a variety of choices in how to conduct their lives.

SUGGESTED CROSS-REFERENCES

Elder abuse is discussed in Section 51.6g; child abuse is discussed in Section 49.4; posttraumatic stress disorder is discussed in Chapter 15.

SECTION REFERENCES

*Acierno R, Resnick HS, Kilpatrick DG: Health impact of interpersonal violence 1: Prevalence rates, case identification, and risk factors for sexual assault, physical assault, and domestic violence in men and women. Behav Med *23:*53, 1997.

American Medical Association: Diagnostic and treatment guidelines on domestic violence. Arch Fam Med *1:*39, 1992.

Bachman R, Saltzman LE: *Violence Against Women: Estimates from the Redesigned Survey*. U.S. Department of Justice special report NCJ-154348, Office of Justice Programs, Bureau of Justice Statistics, Washington, DC, 1994.

Breslau N, Davis GC, Andreski P, Petersen E: Traumatic events and posttraumatic stress disorder in an urban population of young adults. Arch Gen Psychiatry *48:*216, 1996.

*Browne A: Violence against women: Relevance for medical practitioners. JAMA *267:*3184, 1992.

Browne A, Bassuk SS: Intimate violence in the lives of homeless and poor housed women: Prevalence and patterns in an ethnically diverse sample. Am J Orthopsychiatry *67:*261, 1997.

Burgess AW, Holmstrum LL: The rape trauma syndrome. Am J Psychiatry *131:*981, 1974.

Craven D: *Sex Differences in Violent Victimization*. Bureau of Justice Statistics, U.S. Department of Justice, Government Printing Office, Washington, DC, 1997.

Davidson JRT, Hughes DC, Blazer D, George LK: Posttraumatic stress disorder in the community: An epidemiological study. Psychol Med *21:*713, 1991.

*Dickinson LM, deRuy FV III, Dickinson WP, Candib LM: Health-related quality of life and symptom profiles of female survivors of sexual abuse. Arch Fam Med *8:*35, 1999.

Felitti VJ, Anla RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, Koss MP, Marks JS: Relationship of child abuse and household dysfunction to many of the leading causes of death in adults: The adverse child experiences (ACE) study. Am J Prev Med *14:*245, 1998.

Foa EB: Trauma and women: Course, predictors, and treatment. J Clin Psychiatry *58:*25, 1997.

Foa EB, Meadows EA: Psychosocial treatments for posttraumatic stress disorder. A critical review. Annu Rev Psychol *48:*449, 1997.

Gelles R, Journal, Straus MA: *Intimate Violence*. Touchstone Books, New York, 1988.

Gilligan J: *Violence: Our National Epidemic*. Putnam, New York, 1996.

Holmes MM, Resnick HS, Kilpatrick DG, Best CL, Moore JG, Moreno H: Rape-related pregnancy: Estimates and descriptive characteristics from a national sample of women. Am J Obstet Gynecol *175:*320, 1996.

Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB: Posttraumatic stress disorder in the national comorbidity survey. Arch Gen Psychiatry *52:*1048, 1995.

Kilpatrick DG, Aciemo R, Resnick HS, Saunders BE, Best CL: A 2-year longitudinal analysis of the relationships between violent assault and substance use in women. J Consult Clin Psychol *65:*834, 1997.

Kilpatrick DG, Edmunds CN, Seymour AK: *Rape in America: A Report to the Nation*. National Victim Center, Arlington, VA, 1992.

*Kilpatrick G, Resnick HS, Aciemo R: Health impact of interpersonal violence 3: Implications for clinical practice and public policy. Behav Med *23:*79, 1997.

Koop CE: Violence in America: A public health emergency. JAMA *267:*3075, 1992.

Koss MP, Goodman LA, Browne A, Fitzgerald LF, Keita GP, Russo NF: *No Safe Haven: Male Violence Against Women at Home, at Work, and in the Community*. American Psychological Association, Washington, DC, 1994.

Koss MP, Koss PG, Woodruff MS: Deleterious effects of criminal victimization on women's health and medical utilization. Arch Intern Med *151:*342, 1991.

Perkins CA: Age Patterns of Victims of Serious Violent Crime, U.S. Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC, 1997.

Resick PA: The psychological impact of rape. J Interpers Violence *8:*223, 1993.

Resnick HS, Acierno R, Kilpatrick DG: Health impact of interpersonal violence 2: Medical and mental health outcomes. Behav Med *23:*65, 1997.

Rothbaum BO, Foa EB, Riggs DS, Murdock T, Walsh W: A prospective examination of post-traumatic stress disorder in rape victims. J Trauma Stress *5:*455, 1992.

Straus MA, Gelles RJ: *Physical Violence in American Families: Risk Factors and Adaptation to Violence in 8,145 Families*. Transaction, New Brunswick, NJ, 1990.

*Thompson MP, Kaslow NJ, Kingree JB, Puett R, Thompson NJ, Meadows L: Partner abuse and posttraumatic stress disorder as risk factors for suicide attempts in a sample of low-income, inner-city women. J Traum Stress *12:*59, 1999.

Valentiner DP, Foa EB, Riggs DS, Gershuny BS: Coping strategies and posttraumatic stress disorder in female victims of sexual and non sexual assault. J Abnorm Psychol *105:*455, 1996.

van der Kolk BA, McFarlane AC, Weisaeth L: *Traumatic Stress: The Effects of Overwhelming Experience on Mind, Body and Society* . Guilford, New York, 1996.

van der Kolk BA, Pelcovitz D, Roth S, Mandel F, McFarlane AC, Herman J: Dissociation, somatization and affect dysregulation: The complexity of adaptation to trauma. Am J Psychiatry *153*(Suppl):83, 1996.

28.8 ALTERNATIVE AND COMPLEMENTARY HEALTH PRACTICES

THOMAS J. KIRESUK, PH.D. AND ALAN TRACHTENBERG, M.D., M.P.H.

OVERVIEW

The term *alternative and complementary health practices* covers a large and heterogeneous collection of health care systems and methods, including folk medicines, healing rituals, and self-care methods from cultures around the world. Recent information from the National Institutes of Health (NIH) National Center for Complementary and Alternative Medicine (NCCAM) (Table 28.8-1) states that the scope of alternative and complementary health practices includes 12 major health care systems, 26 categories of practice, over 350 methods, and 10,000 ways of using these methods. The term “alternative” has been defined in various ways. The World Health Organization defines alternative therapies as “usually lying outside of the official health sector.” David Eisenberg's definition stipulated, “neither taught widely in US medical schools nor generally available in US hospitals.” Thus, alternative therapies can be thought of as being alternative in origin (they come from somewhere else) rather than as necessarily being alternatives to standard therapies.

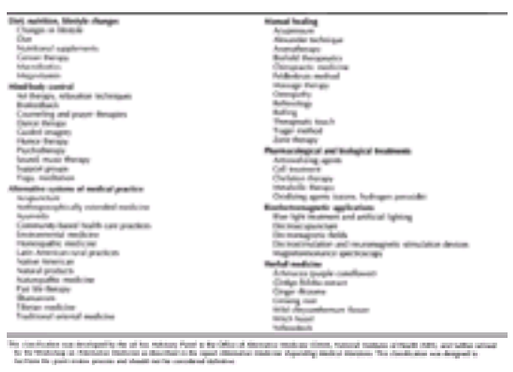

Table 28.8-1 Classification of Alternative Medical Practices From the NIH Office of Alternative Medicine

In contrast, conventional (standard, orthodox, or allopathic) medicine is the type of medicine taught in U.S. medical schools and practiced in U.S. hospitals. The term “allopathic” was coined by the inventor of homeopathy, Samuel Hahnemann (Fig. 28.8-1) to refer to the use of procedures or medications that counteract the signs and symptoms of disease. This term was used in contrast to his method, *homeopathy*, which used medications that could cause the same symptoms in a healthy person as those being treated in patients. Conventional medicine (including Western medical and surgical traditions) is based to a large extent on the results of scientific research and presumes that the body is a biological and physiological system. Disorders are believed to have causes that can be treated with medications, surgery, and complex technological methods to alleviate or cure.

FIGURE 28.8-1 Samuel Hahnemann. (Reprinted with permission from the New York Academy of Medicine, New York, NY.)

Alternative and complementary health practices typically use a holistic, health-oriented approach to the patient rather than a disease-oriented diagnostic and treatment model. The British have used the term “complementary” to denote that these therapies act by complementing, or building on, the body's own resources or defense systems. The treatment methods have not been developed through Western research methods but do have complex theoretical bases and centuries of reported experience with clinical applications. Modern medicine and psychiatry evolved out of a similar accumulation of empirically based knowledge, much of which has been, or is now being, subjected to modern research examination. For example, the origin of psychotherapy was based on nonsystematic case study reports, followed in the 1920s and 1930s by surveys of reports on improvement rates, then by studies that included control groups in their design, and from the 1950s to today, there has been exponential growth in research studies on psychological therapies with greater research rigor. It appears that alternative and complementary health practices are on a similar trajectory but, one hopes, at an accelerated rate, given the availability of contemporary research methodology and analyses.

Some treatment methods classified as alternative have been used in psychiatry for years, and psychiatrists may be surprised to see techniques such as biofeedback, guided imagery, psychotherapy, support groups, art therapy, and relaxation included as alternative medicine practices in the classification system produced by the NIH Office of Alternative Medicine. This classification difficulty points out the evolving, integrative process of medical education and practice. As research data accumulate, one might expect that other forms will find acceptance when it is merited.

The world-wide utilization rates of alternative and complementary health practices are extraordinary. The World Health Organization estimates that between 65 and 80 percent of the world's health care services are currently classified as traditional medicine. In many countries traditional medicine constitutes the major form of

treatment. According to Eisenberg's updated survey findings, Americans spent $27 billion, most of it not reimbursed by insurance, on alternative treatments in 1997. The use of these therapies increased by 25 percent during the seven-year period between surveys, and expenditures increased by 45 percent. Nearly 50 percent more people (a total of 83 million) visited alternative medicine practitioners in 1997 than in 1990. There are more professional visits for alternative and complementary health practices than for conventional primary care. Of Americans using alternative and complementary health practices, 60 percent do not tell their doctors. The authors expressed concern that as many as 15 million people take prescription drugs as well as herbs or high-dosage vitamins, perhaps placing these individuals at risk for harmful medication interactions.

Surveys suggest that U.S. consumers use alternative and complementary health practices for a wide variety of conditions and visit providers of alternative and complementary health practices more frequently than primary care medical doctors, some of whom have begun to refer to alternative therapy practitioners. Use of alternative and complementary health practices does not appear to be restricted by social class, educational level, or sex, although women and Americans with more education seem more inclined to use these practices. The Office of Alternative Medicine has identified more than 700 journals that publish research on alternative and complementary health practices topics. Many of these journals have been developed in Europe and Asia specifically for alternative and complementary health practices research. Approximately 40 percent are in languages other than English.

Patient and physician surveys are not the only indication of the increasing prevalence of alternative and complementary health practices. For instance, they are finding their way into medical school curricula (63 medical schools in a recent listing), and leaders in the health care market and legislators are debating coverage for alternative medicine. Changes in terminology from *unconventional* or *alternative* to *complementary* and *integrative* speak to the increasing collegiality between disciplines.

In 1991 Congress established the Office of Alternative Medicine (now the NCCAM) at the NIH to examine the broad range of therapeutic systems and evaluate their usefulness and safety, and provide scientific explanations for their effectiveness (if any). The NCCAM has been careful to point out that inclusion of treatments in their classification system and support of research on alternative and complementary health practices is not based on biomedical hypotheses about their mechanisms nor even necessarily a belief in their efficacy. It is driven, rather, by the public health imperative created by the fact that so many Americans are using alternative and complementary health practices in the absence of adequate data. Although many proponents of these practices are willing and even eager to submit their methods to objective scientific scrutiny, few of these proponents have adequate training or scientific resources themselves to conduct the necessary objective studies. The few proponents who do manage to perform an adequate study that suggests efficacy are simply disbelieved by those who choose to remain skeptical. The research task, therefore, has fallen to the public health community to bridge the gap between communities where alternative and complementary health practices are already culturally accepted and the orthodox biomedical community, who expect a higher quality of evidence despite continuing to use many regular medical interventions that have not yet met the modern standards of evidence-based medicine.

While the high prevalence of use of alternative and complementary health practices by the general public is well documented, the efficacy, effectiveness, safety, and validity of most of these therapies is not well established. The first task is to sort through the many claims, prevalent beliefs, and the evidence (which is often preliminary). However, many of the claims do not stand up to close scrutiny, and the rush to the marketplace by many advocates of particular treatments or methods greatly complicates the evaluation process. Certainly, popular publications and advertisements provide a flood of statements that have variable levels of formal evidence—most commonly none. In many cases alternative and complementary health practices research, while appropriately intended to elicit positive findings (if available) for closer review, often does not use standard research methods that are commonly accepted in conventional research. However, some therapies, through the process of evaluative review, do emerge as promising.

The following review lists some of the most visible alternative and complementary health practices that have some data on which to base claims or research for effectiveness in the treatment of (broadly defined) psychiatric conditions. Diagnostic systems such as the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the ninth revision of *International Statistical Classification of Diseases* (ICD-9) may or may not be used in these reports and claims; rather, terms such as *anxiety, insomnia, sense of well-being*, and *depression*, are used.

A general observation regarding the methodology of the research on alternative medicine applies to many forms of treatment research. The research is often reported without reference to the rationale for the choice of *P* value, sample size, estimates of statistical power (the probability of detecting a difference between treatments when in fact differences exist), the basis for estimating treatment effect size, and the rationale for the choice of controls. Without these basic determinations, much research in this area is inconclusive. If statistical differences are not found when sample sizes are inadequate, one cannot conclude that a lack of efficacy has been demonstrated (rather than a type II error). There is a definite role for case studies and preliminary reports in the development of new knowledge. Problems arise primarily when inappropriate conclusions or claims are made or when previously unknown adverse effects emerge.

ACUPUNCTURE AND ACUPRESSURE

Biomedical research on acupuncture has been evaluated by a recent NIH Consensus Development Conference. The best summary of this form of treatment can be obtained from the 1997 conference report:

The practice of acupuncture to treat identifiable pathophysiological conditions in American medicine was rare until the visit of President Nixon to China in 1972. Since that time there has been an explosion of interest in the United States regarding the application of the techniques of acupuncture to American medicine. In some European countries, such as France, acupuncture is already well accepted and regarded as just another medical specialty.

Acupuncture describes a family of procedures involving stimulation of anatomical locations on or beneath the skin by a variety of techniques. The most studied mechanism of stimulation of acupuncture points employs penetration of the skin by thin, solid, metallic needles (Fig. 28.8-2), which are manipulated manually or stimulated by electrical impulses. Stimulation of these areas by moxibustion (burning small cones of dried, powdered artemesia vulgaris (moxa) leaves held above the point to be warmed or placed on the skin and removed before overheating occurs), pressure, heat, and laser is also used in acupuncture practice, but due to the paucity of studies, these techniques are more difficult to evaluate. Thus, there are a variety of approaches to diagnosis and treatment in American acupuncture that incorporate medical traditions from China, Japan, Korea, France and other countries.

FIGURE 28.8-2 Acupuncture. Once in place, the needle may be manipulated either by twirling or gentle pumping action. (Reprinted with permission from Shealy CN, editor: *The Complete Family Guide to Alternative Medicine*. Barnes & Noble, New York, 1996.)

Acupuncture has been used by millions of American patients and performed by thousands of physicians, dentists, licensed acupuncturists, and other professionals for relief or prevention of pain and for a variety of health conditions. After reviewing the existing body of knowledge, the U.S. Food and Drug Administration recently removed acupuncture needles from the category of “experimental medical devices,” and now regulates them just as it does other clinical tools such as surgical scalpels and hypodermic syringes, under good manufacturing practices and single-use standards of sterility.

Over the years, the United States Public Health Service (NIH and particularly the former ADAMHA Institutes) has funded a variety of research projects on acupuncture, including studies on the mechanisms by which acupuncture may have its effects, as well as clinical trials and other studies. There is also a considerable

body of international literature on the risks and benefits of acupuncture, and the World Health Organization lists a variety of medical conditions that may benefit from the use of acupuncture or moxibustion. Such applications include prevention and treatment of nausea and vomiting; treatment of pain and addictions to alcohol, and illicit drugs; treatment of pulmonary problems such as asthma and bronchitis; and rehabilitation from neurological damage such as that caused by stroke.

The report concludes that positive results have been obtained regarding adult postoperative and chemotherapy nausea and vomiting and postoperative dental pain. Acupuncture may be a useful adjunct treatment or an acceptable alternative or be included in a comprehensive management program for conditions such as addiction, stroke rehabilitation, headache, menstrual cramps, tennis elbow, fibromyalgia, myofascial pain, osteoarthritis, low back pain, carpal tunnel syndrome, and asthma. Plausible mechanisms of action for acupuncture include the release of endogenous opioids and other neuropeptides in the central nervous system and the periphery and changes in neuroendocrine function.

The NIH consensus report does not refer to psychiatric conditions other than addiction. A recent review of acupuncture for schizophrenia concluded that research on acupuncture or low-power laser treatment was significantly or seriously flawed. While some studies (8 papers were reviewed) suggest that the treatments may be as effective as chlorpromazine (Thorazine) in schizophrenia, no scientifically sound conclusions can be drawn. The research is an example of preliminary research with findings that appear to be worthy of further study. One study provided preliminary information (*N* = 40) that acupuncture might reduce the dosage of antipsychotic medication for patients with schizophrenia, thereby reducing adverse effects without reducing clinical effectiveness. This is a potentially important role for acupuncture, as well as other unorthodox therapies, which may be used to complement standard therapies rather than as an alternative to replace treatments already proved effective.

There is also preliminary research regarding the treatment of depression. Two random clinical trials compared electroacupuncture with amitriptyline (Elavil) in the treatment of depressed patients. Both studies (sample sizes of 47 and 241) found significant improvement resulting from the treatments, with no significant difference between the two groups. Follow-up of 148 patients for 2 to 4 years revealed no significant difference in depression recurrence rates.

Another pilot trial of the effects of acupuncture for depression was conducted on a sample of 22 women with major depressive disorder. Significantly greater improvement was found in the 11 women randomized to receive an 8-week course of acupuncture specific for major depressive disorder as compared to the 11 controls who received non-specific acupuncture for the same period.

Animal research has been reported which indicates that acupuncture and electroacupuncture are capable of accelerating the synthesis and release of serotonin and norepinephrine in the central nervous system (CNS). The authors cite the above human subject research and stated that clinical data indicates that electroacupuncture is effective in treating depressive patients and is at least as effective and has a higher therapeutic index than tricyclic amitryptyline. Given the preliminary nature of the research and a plausible mechanism of action, there is a need for replication and placebo controls for these findings.

In an open trial report on the use of acupuncture for the treatment of psychogenic impotence, 20 of 29 subjects demonstrated improvement. A common observation made by acupuncturists, administrators, and many recipients of the treatment is that acupuncture has a soothing effect, inducing relaxation and greater amenability to treatment. In substance abuse treatment, acupuncture appears to be associated with reduction in alcohol abuse. Early research findings have led to current major research funded by National Institute on Alcohol and Alcoholism (NIAAA) and National Institute on Drug Abuse (NIDA). The effect of acupuncture on cocaine and heroin use remains to be demonstrated. An NIDA-supported study of acupuncture and cocaine use found no differences among different doses of acupuncture, true and sham acupuncture, and standard treatment. Patients receiving acupuncture perceived significant improvement, which was not related to placebo responsiveness measures. Current research methods have not examined this perceived efficacy. The finding does help explain the widespread popularity of the treatment in spite of the lack of consistent data supporting specific efficacy (in conditions other than nausea, vomiting, and some types of pain). A consortium that includes several federal agencies and the Conrad Hilton Foundation has funded a nationwide multisite study of acupuncture (among other treatments) for cocaine addiction, which may produce definitive data by the end of 1999.

The use of acupuncture in the treatment of smoking is another popular application that has been tested several times without consistent demonstration of benefit. The research on this topic is inconclusive because of the wide variation in the definition of the treatment and the variation in research methods. Animal research is being conducted which may shed light on whether acupuncture can have an effect on nicotine dependence. A recent study reported a randomized, single-blind comparison of two sets of acupuncture points administered to a sample of 46 subjects twice a week for three weeks. Test acupoints included 2 body electroacupuncture, 3 ear acupuncture, and 6 ear acupressure sites. Control acupuncture included 11 body points presumed to be noneffective for smoking cessation. Statistically significant differences were found in reduction of cigarette consumption, concentrations of cotinine and thiocyanate, but no significant changes in serum peroxides and plasma fibrinogen concentrations for either group. Taste aversion and desire to smoke fell significantly in both groups but the reduction was larger for the test group. This study did not have placebo control for the two sets of acupuncture treatments, the sample was small, duration of treatment effects was not measured, and the treatment points are multiple and require replication by others. However, this research is an example of the kind of carefully done, multimeasure study that can form the basis for further research. Even though the research evidence for smoking cessation is in an early stage of development, physicians will encounter many individuals who swear by this treatment.

HERBAL MEDICINE

Herbal medicine is probably the most widely used alternative and complementary health practice. Including all individual herbal remedies and their combinations, there are literally thousands of herbal treatments. Herbal medicine relies on plants to treat illness and maintain health. It is the oldest known system of medicine, with historical origins attributed to China about 4000 BC, but certainly plants were used medicinally worldwide throughout prehistory. Ancient texts of Asian medicine are still in use, and modern Chinese medicine relies on herbs in addition to other methods such as acupuncture, massage, diet, and exercise to correct “imbalances” in the body. A Greco-Roman medical text by Dioscorides, *De materia medica*, describes the use of over 500 plants to treat disease, beginning a tradition followed in European medicine to this day.

The decline of herbal medicine in Western countries in the twentieth century was related to scientific and technological advances that led to the use of synthetic and chemically purified pharmaceuticals; nevertheless, according to some estimates, at least 25 percent of current medicines are devised from the active ingredients of plants. The examples are many: digitalis from foxglove, ephedrine from “Mormon tea,” morphine from the opium poppy, taxol from yew, and quinine from the bark of the chinchona tree. In addition, many modern medicines are congeners of compounds from plants in which their biological actions were first discovered.

About $1.5 billion a year is spent in the United States on herbs that are legally classified as dietary supplements, and the quality control standards applied by the Food and Drug Administration (FDA) are those applied to foods rather than pharmaceutical or over-the-counter (OTC) quality standards. Concerns arise only when unsupported health claims are made for food or dietary supplements. Thus there is little reliability in content and quality control for most herbal preparations sold in health food stores, pharmacies, and supermarkets. Efforts to regulate quality are being made through organizations such as the American Botanical Council and the Herb Research Foundation, but according to the Federal Trade Commission (FTC), poor manufacturing practices and false advertising still exist. One useful approach to quality control conducts spectrographic and other pharmacological analyses to determine the herbs' identity and the concentration of specific compounds that may be associated with a particular herb's clinical activity. Germany and Britain have developed useful quality standards that can generally be depended on in products imported from European countries or claiming to meet those standards in U.S. products (such claims, if false, are actionable by the FTC). The possibility of allergic reactions or toxic interaction with medications should always be kept in mind when patients are eating plant substances of uncertain purity or identity.

St. John's Wort There are many herbal products that have been used legally or illegally for their mood-altering or sedative effects. Examples include the opium poppy (opium), *Atropa belladonna* (deadly nightshade), Indian hemp (hashish), hyoscyamine (henbane), scopolomine (thorn apple), and hypericum (St. John's wort). Of these, St. John's wort has received the greatest recent research attention and is currently the subject of an NIH-funded multisite trial of its effects on depression. Prior research with St. John's wort has used a variety of methods including comparison with placebo controls, comparison with imipramine (Tofranil), large-scale postmarket reporting, and animal research. Meta-analysis of 23 randomized trials that included a total of 1757 outpatients with mainly mild or moderate depressive disorders concluded that Hypericum extracts were significantly superior to placebo and as effective as standard antidepressants. The individual research studies reported very low (in most studies, none) rates of adverse effects, fewer and milder adverse effects compared with imipramine. Criticism of the research includes concerns regarding heterogeneity of patients, interventions, dosage levels, extract preparations, and diagnostic classifications.

There has been some concern regarding longer-term effects and possible enhancement of photosensitivity. In particular, it has been suggested that the use of St. John's wort should not be combined with use of a selective serotonin reuptake inhibitor (SSRI) (e.g., paroxetine [Paxil]). This example of potential undesirable consequences resulting from combinations of herbal and standardized pharmaceuticals is emblematic of a larger concern regarding all the known and unknown side

effects of drug and herbal product combinations.

A recent review of St. John's wort included a detailed discussion of the several proposed mechanisms of action and added concerns regarding adverse effects and drug interaction effects. Generally the authors felt that there were fewer adverse effects with hypericum than with conventional antidepressants. Potential adverse effects that were listed included photodermatitis, delayed hypersensitivity, gastrointestinal tract upset, dizziness, dry mouth, sedation, restlessness, and constipation. There do not appear to be significant adverse effects on cardiac conduction with hypericum extracts. Contraindications for the use of St. John's wort included pregnancy, lactation, exposure of strong sunlight, and pheochromocytoma. The potential for monoamine oxidase inhibitor-like drug interactions cannot be excluded since appropriate drug interaction studies are not available.

A recent review concluded that hypericum is at least as safe and possibly safer than conventional antidepressant drugs.

Kudzu (*Pueraria Lobata*) A Chinese herbal medicine (labeled NPI-028 by researchers in the Research Triangle) has been used historically in the treatment of alcohol intoxication. It is still prescribed in China and Southeast Asia, and its use was documented in early historical documents (Li Shizhen 1590–1596). NPI-028 consists of six Chinese herbs, including *Pueraria lobata* and *Citrus reticulata*. This preparation has suppressed alcohol drinking by three different strains of alcohol-preferring rats. Renee Lin and coworkers reported in 1995 that puerarin, daidzin, and daidzein, extracted from kudzu and given orally, suppressed voluntary alcohol consumption by alcohol-preferring P rats. NPI-028 was effective in lowering ethanol intake at lower doses in alcohol-preferring vervet monkeys. These effects of NPI-028, which contains kudzu, are comparable to recently reported effects of two extracts from the kudzu plant. Wing-Ming Keung and Bert Vallee at Harvard Medical School showed that daidzin and daidzein were the active herbal components from *Radix pueraria* (kudzu root) that suppressed alcohol intake in Syrian golden hamsters. Daidzin differs from disulfiram in its selective and reversible inhibition of aldehyde dehydrogenase-1 (ALDH-1). Daidzin also decreases blood alcohol concentrations and shortens sleep time induced by ethanol. These findings provide a basis for the possible use of this traditional Chinese herbal medicine in the treatment of alcohol abuse.

Kava The term *kava* refers to the Pacific Island plant, *Piper methysticum* Forst (Piperaceae) and the beverage prepared from it. The preparation is used in traditional island ceremonies and has therapeutic and mood-altering characteristics. There is a considerable body of literature regarding kava, including a bibliography containing approximately 800 references, articles concerning chemical and pharmacological studies, and a review written by a native of Fiji with 153 additional references. Kava was in an apparently successful treatment to facilitating smoking cessation in which it was combined with collaborative and village empowerment group and ceremonial methods.

Kava drinking has been introduced into Aboriginal communities in North Australia. Used at much higher consumption levels than in Oceania, several effects were noted: mental and physical relaxant effects, and anecdotally reported improvement in schizophrenic and mood disorders, resembling the effects of antipsychotic agents. There are also reports of abuse, and kava is one of the herbs sometimes recommended in the drug culture for a “legal and natural” high.

A recent review of kava human-trial research indicates that kava products may be beneficial in the management of anxiety and tension of nonpsychotic origin and does not adversely affect cognitive function, mental acuity, or coordination in comparison with oxazepam. Long-term use at higher daily dosages (400 mg of kavalactones) may result in scaling of the skin on the extremities. There is a possibility that kava may potentiate the action of other centrally mediated agents and interact with alcohol. However, in a placebo-controlled, double-blind study few adverse effects were found.

Valerian In the treatment of anxiety and sleep disorders, valerian is considered one of the most effective herbal preparations. It is obtained from the dried rhizome and roots of the perennial herb *Valeriana officinalis* and has been known for its sedative properties for 2000 years. Pharmacological tests have demonstrated the activity of valerian on the CNS. Biochemical receptor assays have demonstrated interaction with the inhibitory neurotransmitter g-aminobutyric acid (GABA). Several reviews of clinical investigations support its use as a calmative and sleep-promoting agent. A clinical report of an overdose (20 times the recommended therapeutic dose) stated that the symptoms were mild and resolved within 24 hours. Another study of overdose of a sleep-inducing preparation that included valerian found no evidence for hepatotoxicity, although longer-term studies would be required to establish risk for long-term users. The advantage over synthetic psychopharmacological agents is the low incidence of adverse effects.

A recent review of valerian included animal studies that showed results consistent with other hypnotic agents such as the benzodiazepines. Human clinical studies confirmed a mild sedative effect, but did not evaluate the efficacy of valerian as a treatment for primary or secondary insomnia. The authors concluded that valerian had not been demonstrated to be superior to existing hypnotic treatments or other treatments of insomnia. As with many herbal preparations, there is insufficient information to recommend valerian in pregnancy and during lactation. Valerian may potentiate the effects of other CNS depressants and the usual precautions taken with other sedating agents also apply to valerian.

Passion Flower In Britain, passion flower is one of the most popular herbal sedatives. It is derived from the dried flowering and fruiting top of a perennial climbing vine *Passiflora incarnata* L. However, in spite of its popularity, it is a relatively unproven minor tranquilizer. Animal research has indicated prolonged sleeping time and reduced amphetamine-induced hypermotility. Synergism with kava was also noted. The FDA prohibited the use of passion flower in OTC sedative preparations because it has not been proved safe and effective; however, the German Commission E authorized its use in the treatment of nervous unrest, on the basis of the results of animal research.

A recent summary stated that hypersensitivity vasculitis and “altered consciousness” have been reported with products containing passion flower. Interactions with other psychotropic medications have not been adequately studied. Because of its potential sedative effects, the usual precautions regarding operating motor vehicles or machinery should be made. The authors of this review concluded that excessive use during pregnancy and lactation should be avoided.

Gingko Biloba Probably one of the most studied of all traditional herbal medicines is gingko biloba. The gingko tree has bilobed leaves that resemble the maidenhair fern. It is believed to be the oldest living seed plant. Gingkolic acid and related alkylphenols constitute major components of the lipid fraction of the fruit pods of *Ginkgo biloba* L. This class of substances s present in ginkgo leaves, which are widely used to prepare extracts for the treatment of peripheral or cerebral circulatory disorders, as well as vascular dementia and dementia of the Alzheimer's type. Ginkgo is widely used in European medical practice to improve cognitive function in patients diagnosed with symptoms of “cerebral insufficiency,” with over 5 million prescriptions written annually in Germany alone. Most research studies use ginkgo leaf extract EGb 761 (Tavonin).

Pharmacological studies have shown that the therapeutic effect of ginkgo is based on several active constituents with vasoactive and free radical-scavenging properties. A recent review of studies of clinical improvement of memory and other cognitive functions by ginkgo biloba came to the following conclusions:

> The extract is used to alleviate symptoms associated with cognitive deficits, eg, decreased memory performance, lack of concentration, decreased alertness, tinnitus, and dizziness. The use of ginkgo extract in either dementias of the Alzheimer or multi-infarct type or in the case of cerebral insufficiency, a symptom complex related to age-dependent impairment of cerebral circulation, is based mainly on positive results from good-quality placebo-controlled studies that enrolled approximately 1200 patients with criteria established by International Classification of Diseases (9th and 10th revisions, ICD-9 and ICD-10) or the 3rd revision of the Diagnostic and Statistical Manual (DSM-III-R) (uncomplicated dementia). Effect on cognitive symptoms was within the range of a 25% improvement. Memory, concentration, and alertness were the first symptoms to be relieved, with tinnitus and dizziness improving somewhat later. A minimum of 4 to 6 weeks were needed before a pronounced effect could be expected. The pharmacologic advantage of ginkgo appears to be a very tolerable side-effect profile, with a side-effect frequency at the placebo level.

A recent multicenter placebo-controlled, double-blind, randomized trial of the extract of *Ginkgo biloba* for dementia concluded that EGb 176 was safe and appeared capable of stabilizing and (in a substantial number of cases) improving the cognitive performance and social functioning of demented patients for 6 months to 1 year. Although modest, the changes induced by EGb 176 were objective cognitive measures and were of sufficient magnitude to be recognized by the caregivers. Egb 761 and placebo did not differ significantly in the number of patients reporting adverse events or in the incidence and severity of these events.

A recent open trial of *Gingko biloba* extract found significant improvement in antidepressant-induced sexual dysfunction predominantly caused by SSRIs, with women being more responsive to the sexually enhancing effects than men. A related placebo-controlled study found EGb 761 effective against the congestive symptoms of premenstrual syndrome, with improvement also shown in neuropsychological symptoms.

There is evidence that ginkgo can serve as an augmenting agent with conventional antidepressants. In addition, animal studies indicate that ginkolide B may have a neuroprotective effect in brain injury. The most common adverse effect is headache, which can be avoided by individualized graduated dosage. While concern has

been expressed that ginkgo may potentiate other anticoagulants or increase bleeding time, the concern has not been substantiated. Nevertheless, a recent review concluded that caution should be exercised when ginkgo is combined with anticoagulant treatment, including aspirin, where there is a risk of bleeding as in peptic ulcer disease and subdural hematoma. Safety in pregnancy and lactation has not been established.

Yohimbine Yohimbine, an indole alkaloid obtained mainly from the bark of the yohimbe tree (*Pausinystalia yohimba*), a common tree in West Africa, is an a_2-adrenergic receptor antagonist that is marketed pharmaceutically for treatment of impotence. A wide range of animal and human research studies of yohimbine have used a variety of research designs and analyses. A recent meta-analysis of randomized, placebo-controlled trials of yohimbine monotherapy for erectile dysfunction examined the evidence for therapeutic efficacy and safety. The review found that yohimbine is superior to placebo and that serious adverse reactions were infrequent and reversible. Dietary supplements or other natural products of the yohimbe tree sometimes contain significant quantities of yohimbine, but this is unusual.

An earlier review of four meta-analyses integrated data from controlled and uncontrolled clinical trials of yohimbine used alone and controlled and uncontrolled trials of yohimbine used in combination with other drugs. The results indicated a consistent tendency for yohimbine and medications containing yohimbine to enhance erectile functioning, relative to placebo.

A cautionary note was provided by a study of the effects of yohimbine on blood pressure and plasma concentrations of catechols in patients with essential hypertension. The study concluded that yohimbine increases sympathetically mediated plasma norepinephrine concentrations, which in turn produce a pressor response. The authors stated that

> [y]ohimbine should be administered with caution to patients with high blood pressure, especially in individuals with evidence for increased basal sympathetic outflow or those undergoing concurrent treatment with tricyclic antidepressants or other drugs that interfere with neuronal uptake or metabolism of norepinephrine.

Preliminary studies indicate that yohimbine may be useful in the treatment of sexual adverse effects induced by SSRIs, and two case reports concerned clomipramine (Anafranil) (six patients) and fluoxetine (Prozac) (nine patients). A study of nine women with hypoactive sexual desire showed no improvement in sexual desire. A large-scale retrospective case review of 596 patients treated with SSRIs identified 97 patients who had experienced sexual dysfunction associated with SSRIs; 45 patients opted for treatment with trials of yohimbine, amantadine (Symmetrel), or cyprohepatidine (Periactin). All three antidotes were found to be safe and relatively effective, and yohimbine was significantly more effective in reversing SSRI-induced sexual dysfunction.

There are several animal studies using a variety of animal models. For example, yohimbine was shown to enhance sexual behaviors in male rats, including those that were castrated and aging male rats. Yohimbine influence in the absence of testosterone raises interesting questions about the role of a $_2$-receptors in sexual arousal and behavior.

A memorable single-blind study concerned the effect of yohimbine on the reproductive activity of 12-year-old male captive Nile crocodiles, 18 animals divided into 3 ponds. Each pond contained 6 males and 30 females. Yohimbine capsules of 30 mg, twice a day for 1 week were administered to all six in pond 1, two in pond 2, and none in the control pond. A significant increase in the frequency of headslap sexual display was noted as well as an appearance of this daily behavior 1 hour earlier than controls. While there was no apparent increase in copulatory frequency in treated animals, probable increase in successful copulations was inferred because of the prolonged reproductive period and a significant increase in the fertility percentage of eggs in the treatment ponds. While not easily replicated, these results are compatible with findings in other animal research.

A recent review of yohimbine stressed the psychopathology that may be associated with its use. The authors indicate yohimbine is commonly used to provoke panic attacks and anxiety in studies of the pathophysiology, psychopharmacology, and treatment of anxiety disorders. They point out that tricyclic drugs, medications with central a-adrenergic blocking properties, centrally acting sympathomimnetics, monoamine oxidase inhibitors, and antimuscarinic agents potentiate the action of yohimbine. They conclude their summary with cautionary statements, which stress the potential of yohimbine to produce psychiatric symptoms, primarily anxiety or panic, especially in patients with preexisting panic disorder. Once again, however, they point out that many commercial products containing yohimbine bark actually have little or no yohimbine.

A summary of herbal remedies that affect psychiatric symptoms is given in Table 28.8-2.

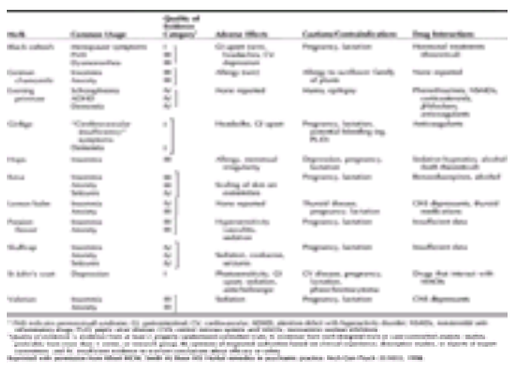

Table 28.8-2 Herbal Remedies Commonly Used to Treat Psychiatric Symptoms*

OTHER HERBAL REMEDIES

There are several other herbal preparations that have been mentioned or reviewed with regard to their psychiatric relevance. They are mentioned here briefly in order to alert the reader to their presence in the literature and because of the likelihood that increasing numbers of patients may be using them or asking about them. The summaries are derived from a review article appearing in the November 1998 *Archives of General Psychiatry*.

Capsicum Capsicum (*Capsicum annuum* L) commonly appears in the form of chili or cayenne pepper. A number of studies support its use for pain relief where it is applied in the form of capsaicin cream. Other uses, such as treatment for colic and dyspepsia, and for improvement of peripheral circulation in patients with cardiovascular conditions, have not been investigated scientifically. Capsicum is mentioned in this section because it may elevate the secretion of catecholamines and therefore should be used with caution concurrently with monoamine oxidase inhibitors.

Chaste Tree Chaste tree (*Vitex agnuscastus* L) has been used traditionally for symptoms of prementrual syndrome, mastodynia, menopause, hyperprolactinemia, and menstrual irregularity. Because an affinity for dopamine type 1 (D_1) and receptors D_2 has been reported, there is a potential for interaction with other dopaminergic drugs, such as the antipsychotics and metoclopramide (Reglan).

Evening Primrose There is little evidence to support the use of evening primrose (*Oenothera biennis* L) for psychiatric conditions such as schizophrenia, childhood hyperactivity, and dementia. Case reports indicate potential harmful effects in mania and epilepsy. Phenothiazines, nonsteroidal anti-inflammatory drugs, corticosteroids, b-adrenergic receptor antagonists, and anticoagulants may interact adversely with evening primrose.

German Chamomile This herb (*Matricaria recutita* L) has a mild hypnotic effect, and has been used in a variety of gastrointestinal, and mouth and skin irritations, pediatric colic and teething, and mild insomnia and anxiety. There have been no randomized or controlled clinical studies. Adverse reactions are rare and mainly allergic in nature.

Hops The female flowers of the plant Humulus Lupulus L have been used as mild sedative and hypnotic agents. There are no clinical studies of its effects on

insomnia or anxiety disorders. The authors caution against its use in depression, pregnancy, and during lactation. Possible potentiation effects may exist when used with sedative hypnotic agents and alcohol.

Lemon Balm No clinical studies support the use of lemon balm (*Melissa officinalis* L) for its reputed hypnotic or anxiolytic effects in humans. No adverse effects have been reported from the use of lemon balm. However, the authors note that safety in pregnancy and lactation has not been established. The herb may potentiate other CNS depressants including alcohol and may interact with thyroid medications or thyroid disease.

Siberian Ginseng Siberian ginseng (*Eleuthrococcus senticosus* [Rupr and Maxim]) is readily confused with plants of the genus Panax which include Panax ginseng CA Meyer (i.e., ginseng, Chinese ginseng, Korean ginseng) and Panax quinquefolius L (i.e., Canadian ginseng, American Ginseng). Siberian ginseng is believed to help fatigue and stress, to improve endurance, and to have immunostimulatory properties. It is mentioned here because some studies report alteration of barbiturate-induced sleeping time and therefore should be used with caution with sedative-hypnotic agents.

Panax Ginseng Panax ginseng is commonly used in the belief that it can treat stress and fatigue and improve endurance. Common adverse effects include insomnia, hypertension, diarrhea, restlessness, anxiety, and euphoria. The authors recommended caution with regard to use in patients with hypertension and diabetes and in conjunction with centrally acting medications. They also note that ginseng may potentiate the effect of monoamine oxidase inhibitors, sympathomimetics and haloperidol (Haldol).

Skullcap While the ingredients and pharmacology of Skullcap (*Scutellaria laterifolia* L) are not well documented, and the existing research may not apply to the preparations currently in use, the reviewers describe its use as a sedative and anticonvulsant. Giddiness, confusion, sedation and seizures are potential adverse effects. The authors advise avoiding skullcap use in pregnancy and lactation and caution that there may be interaction with other CNS drugs.

It should be noted that while the authors of this review provided thorough critique of extant research of herbal preparations and were consistent in their examination of claims for herbal products and potential hazards, they concluded their review with the following statements:

> However, the experience of other cultures, whether in less developed countries or in history, should not be ignored. Contemporary medical research may finally allow us to separate the traditional remedies that can effectively treat disease from those that are superstition and myth. In addition, research into the biochemical and pharmacological effects of these herbs may uncover novel treatments for psychiatric illness or yield fresh insights into basic disease mechanisms.

A summary of herbal remedies that affect psychiatric symptoms is given in Table 28.8-2.

AROMATHERAPY

Aromatherapy is the therapeutic use of aromatic plant oils. Named by the French chemist René-Maurice Gattefosse in 1928, aromatherapy is one of the fastest growing alternative therapies in the United States and Europe. The essential oils of plants are organic compounds with specific cyclic or aromatic components. Substances with strong odors were used in ancient civilization as both medicines and perfumes. Today, plant oils are inhaled using atomizers or absorbed though the skin during massage (aromatherapy massage). Plant oils are said to have many therapeutic effects—analgesic, psychological, antimicrobial—some of which have been demonstrated in various ways. For examples, oil of cloves is a commonly used dental analgesic. Aromatherapy is said to relieve stress and anxiety and alleviate gastrointestinal and musculoskeletal disorders, among others. In psychiatry, olfactory stimulation has been used to elicit feeling tones, memories, and emotions during psychotherapy. The essential oils used in aromatherapy can cause skin irritation or allergic reactions in some people. Table 28.8-3 lists some essential oils and their purported effects.

Oil	Effect
Chamomile	Tranquilizing
Eucalyptus	Antiseptic
Geranium	Astringent
Jasmine	Stimulant
Lavender	Analgesic, tranquilizing
Neroli	Sedative-hypnotic
Peppermint	Stimulant
Rosemary	Stimulant
Sandalwood	Antiseptic, stimulant

Table 28.8-3 Essential Oils in Aromatherapy and Their Purported Effects

One example of aromatherapy research that pertains to psychiatry is an open-trial, small-sample (*N* = 15) study of depressed patients conducted in Japan, which reported that the application of citrus fragrance could markedly reduce necessary doses of antidepressants. While replication, larger samples, and placebo controls would be required to substantiate this finding, the result is consistent with reports of users of aromatherapy. Previous animal research indicated that citrus fragrance could alleviate stress-induced immunosuppression, and various scents have been used in nursing homes to improve the mood of severely demented (and other) patients.

TRANSCRANIAL NEUROELECTRIC STIMULATION

The typical application of transcranial neuroelectric stimulation involves placing surface electrodes in the mastoid region. The electrodes are stimulated with low-amperage and low-frequency alternating current. No seizures are induced, and there is little (if any) resemblance between transcranial neuroelectric stimulation and electroconvulsive therapy (ECT). The method was originally used in the 1950s to treat insomnia and has been used in the treatment of depression and anxiety. Studies of the effectiveness of this method in treatment of substance abuse have not shown a consistent treatment effect, although greater comfort during detoxification has been reported. Similar methods have been called Limoge and Russian Electrosleep therapy.

A related method, rapid-rate transcranial magnetic stimulation, applies electromagnetic stimulation to the left dorsolateral prefrontal cortex of patients with drug-resistant depression. Only a small sample (*N* = 17) of patients was used in a multiple-crossover, randomized, placebo-controlled trial in which sham stimulation was applied to different cortical areas. Eleven of the 17 patients showed significant improvement on the Hamilton Rating Scale for Depression, and no adverse effects were reported. It was suggested that this method might serve as a nonconvulsive alternative to ECT.

THERAPEUTIC TOUCH

Therapeutic touch, a theory of healing with the hands, was developed by Dolores Krieger who studied Theosophists and other healing practitioners who healed by “laying on of hands.” The first major publication appeared in 1972, and since that time the method has been taught in more than 80 colleges and universities in the United States and in 67 other countries. Therapeutic touch has gained popularity in the nursing profession as well as among some physicians and less orthodox healers. Recent television and news reports describe the use of therapeutic touch during surgical procedures.

An initial stage in both therapeutic touch and many forms of massage is placing the hands at a distance ranging from less than an inch up to several inches above the body. It is believed that something unspecified (e.g., somatic information, heat, “energy,”) can be transferred and sensed by the therapist and sometimes the patient and that the healing process is thereby enhanced. Krieger reported four studies stating that increase in hemoglobin values indicate the biological effect underlying the laying-on of hands. An intriguing double-blind study also suggested that therapeutic touching speeds the healing of small, standardized skin wounds.

A review of the research in this area concluded that some evidence exists to support the practice of therapeutic touch for the treatment of pain and anxiety but that the 11 studies lacked conceptual clarity and operational definition of the treatment method and findings. A recent study in *JAMA* presents the view of skeptics of this and other alternative and complementary health practices. This study tested the assertion of some practitioners that they could sense the effects of human bodies on their hands but did not test any of the therapeutic claims made for therapeutic touch. A more recent clinical trial reported efficacy for therapeutic touch in the treatment of pain from osteoarthritis of the knee. There may be several aspects of the interaction between the therapeutic touch practitioner and patient that affect a patient's state of mind or other determinants of pain and healing.

HOMEOPATHY

Homeopathy was developed by Hahnemann 200 years ago and was introduced to the United States in the early nineteenth century. It was believed to be effective in cholera epidemics against which the standard treatments of the day (bleeding and purging) were ineffective or harmful. By 1900 there were 22 homeopathic medical schools, more than 100 hospitals, and approximately 15,000 practitioners in the United States. In the 1920s, with the introduction of modern pharmacotherapies and the standardization of conventional physician training programs using funding guidelines favoring American Medical Association (AMA)-approved institutions, homeopathic hospitals and schools (including almost all of the women's medical schools of the time) disappeared or were converted to allopathic medicine. However, the practice of homeopathy is increasing in this country and around the world. Homeopathy has been popular in Europe throughout the twentieth century, and in India it has retained enormous popularity since its introduction by the British Empire. Homeopathic medicines are sold over the counter in the United States; in 1992 sales were $200 million, and they have increased each year since then.

Hahnemann based homeopathic drug treatment on the following assumptions: disease symptoms result from the body's own efforts to throw off the illness; medical substances elicit a standard array of signs and symptoms in healthy people; the medicine whose effect in normal people closely resembles the symptoms of the illness being treated is the one most likely to initiate the curative response by helping the body complete its own curative efforts. This law of similars— *Similia similibus curantur* (“Let likes be cured by likes”)—led to the term “homeopathy” (“similar experiences”). The process begins with administering small doses of a substance to healthy volunteers to determine the agent's symptom profile. This is called a homeopathic “proving.” The patient is then given the substance whose symptom profile most closely matches the patient's symptom profile. Classical homeopathy develops individualized treatment for each patient; however, a variety of other approaches exist. Homeopathic remedies sold in the United States must meet the standards of monographs in the *Homeopathic Pharmacopoeia of the U.S*. (HPUS), which was recognized in the Food, Drug and Cosmetic Act with authority equivalent to that of the *United States Pharmacopoeia* (USP).

The HPUS is unique in several ways. First, there are over 2000 medications, including plants (e.g., aconite and hellebore), minerals (e.g., copper, gold, and iodine), and animal products (e.g. snake and jellyfish venom and tissue extracts). Second, medications are prepared as tinctures, that is, mixed with 95 percent grain alcohol, or as pills with lactose fillers. Last, medications are dispersed in infinitesimally dilute solutions, from 1:10 to 1:10 ^{1000}, far beyond the point at which any molecules of the medicine can be expected to still be found in the solution. To date, no one has been able to provide possible mechanism for homeopathy. Meta-analyses and reviews of research have come to different conclusions regarding the efficacy of homeopathy. While homeopathy is being used in the United States and the United Kingdom, little research has been published in standard peer-reviewed journals regarding its clinical application. However, a vast homeopathic medical literature exists, some of which has been published continuously (and with consistent standards) since the nineteenth century. While “classical” homeopathy is said by its proponents to be most effective in treating mental symptoms, the literature on randomized blinded trials is currently limited mainly to gastrointestinal and respiratory problems.

A case report on 12 outpatient adults who had major depressive disorder, social phobia, or panic disorder illustrates the use of homeopathy in psychiatry. Duration of treatment was 7 to 80 weeks. Response rates were 50 to 58 percent, depending on the outcome measure. While the authors concluded that homeopathy might be useful in the treatment of affective and anxiety disorders in patients with mildly to severely symptomatic conditions, the study is clearly an early case study that requires replication, larger samples, and placebo or treatment comparison controls.

> A patient with paranoid schizophrenia had been completely noncompliant with medications and had been continuously hospitalized for over 4 years. In the week following one oral administration of an individually chosen homeopathic remedy, this patient began to completely cooperate with the ingestion of antipsychotic medications that then served to clear his psychosis.

While anecdotal, this case is not isolated. An open minded physician cannot see many of these impressive changes without feeling that this is something worthy of further study.

Another intriguing randomized blinded trial conducted on postconcussive head injury suggested that moderately disabled patients could significantly improve their level of function, even years after their original injury.

The homeopathic evaluation process includes holistic consideration of all physical and psychological characteristics of the patient and does not necessarily depend on a diagnosis. As in many forms of alternative and complementary health practices, restriction to single therapeutic agents for particular diagnoses is believed to be an inappropriate way to evaluate the treatment. There are research methods to deal with this issue, but they are complex and more expensive than usual, and they require preliminary findings to justify the investment.

Bach Flower Essence Therapy A related method, Bach flower essence therapy, first described by a British homeopath named Edward Bach and named for him, involves elixirs produced by flowers floating in a bowl of water in sunlight. These elixirs are consumed, and practitioners claim that the treatment produces gentle healing and is free of adverse effects. While there are anecdotal reports of the use of these modalities for the treatment of addictive disorders and posttraumatic stress disorder, efficacy remains to be demonstrated.

MASSAGE

Massage involves manipulation of the soft tissues and surfaces of the body. It was prescribed for the treatment of disease over 5000 years ago by Chinese physicians, and Hippocrates considered it an important method of healing.

Massage may affect the body in several ways: it is said to increase blood circulation, improve the flow of lymph, soothe sore muscles, and have a tranquilizing effect on the mind. Many mothers and pet-owners have noted the calming effects of massage on children or animals.

Various massage techniques have been described: stroking, kneading, pinching, rubbing, knuckling, tapping, or applying friction. Massage is most often done with the hands and fingers, but vibrating machines and electrical stimulation are also used. The different types of massage therapy that have evolved over the years are more similar than different. These include Swedish, Oriental, Shiatsu, and Esalen massages. Most people who experience massage find it physically and mentally refreshing. A few preliminary studies have supported the perceived benefits of the treatment.

In a study of 11 patients with human immunodeficiency virus (HIV) infection, using an interrupted time trial design, massage therapy increased natural killer cell number, natural killer cell cytotoxicity, and soluble cluster of differentiation 8 (CD8) and the cytotoxic subset of CD8 cells. There was a major decrease in cortisol, significant decreases in anxiety, and increases in relaxation. The authors concluded that apparent clinical improvement is associated with massage.

A review of massage therapy for infants and children with various medical conditions showed that the treatment resulted in lower anxiety and stress hormones and improved clinical status. In a related study, 52 hospitalized depressed and adjustment-disordered children were treated with 30-minute back massage given daily for a 5-day period. Compared with subjects viewing relaxing videotapes, massaged subjects were less depressed and anxious and had lower saliva cortisol levels after the massage. Nurses rated the subjects less anxious and more cooperative on the last day of the study.

AYURVEDA

Asian Indian medicine has been traced to sacred writings dating to the second millennium BC. This system of medicine evolved through early periods of magical practices and use of a wide variety of herbal medicines. Medical treatises written during a period around the first century AD form the basis for most later writings. Current forms of this medical system demonstrate both Buddhist and Hindu contributions and are highly respected, involving considerable medical training from

recognized ayurvedic medical schools all over India.

Recent writings pertaining to psychiatry include treatises explaining achievements in the field of mental health, basic ayurvedic concepts, and their relationship to the modern scientific world. Ayurvedic concepts may be related to conditions such as anxiety, neurosis, and depressive disorders. Medical literature reviews have compared ayurvedic concepts to the works of Freud, Kurt Lewin's field theory, homeostatic imbalance as a source of mental illness, and personality types and their correspondence to types of mental disorders.

BIOENERGETICS

Bioenergetics, based on the belief that dammed-up “vital” energy produces maladaptive behavioral patterns, evolved from the work of the Austrian psychoanalyst Wilhelm Reich (1897–1957), who studied with Freud. Reich believed that energy fields were propelled by sexual impulses called ergs and that satisfactory orgasms indicated healthy bodily functioning. Modern-day practitioners look for areas of muscular tension in the body that are thought to be associated with repressed memories and emotions. Therapists try to bring these repressions to consciousness through a variety of relaxation techniques, including massage or noncontact methods resembling therapeutic touching. The research published in this area includes topics such as the effects of bioenergetics and progressive relaxation on self concept, locus of control, and social behavior of hospitalized psychiatric patients.

Related to these concepts are the use of deep breathing and focusing on attendant bodily feelings, together with facilitated fantasy development. Four types of related body psychotherapies are Gestalt therapy, Reichian therapy, Lowenian bioenergetics, and primal therapy. All share the belief that psychopathological functioning is rooted in rigidification of behavior and experience that are biologically and somatically, or physically remembered, in contrast to the reasoning of mental consciousness, or CNS memory. Several therapies use various kinds of mental and physical manipulations to “dislodge” traumatic memories of painful events stored in physical locations of the body.

BIOFEEDBACK

Biofeedback refers to a therapeutic process in which information about an individual's physiological processes such as blood pressure, heart rate, brain waves, temperature of extremities, or muscle tension is monitored electronically and “fed back” by means of sounds, lights, or electronic gauges. Using these techniques individuals can control a variety of physiological responses formerly thought to be completely nonvoluntary and thereby deal more effectively with stress reactions such as anxiety and pain. Feedback regarding status of brain waves has also been found useful—alpha waves being associated with relaxation and meditation, theta waves with focused attention. Herbert Benson has described a common “relaxation response” visible in most people as a result of alpha wave biofeedback, transcendental meditation (TM), or other reproducible meditation techniques.

Some early findings indicate that alpha or theta feedback may have significant effects on substance abuse. Several studies show positive effects on presurgical anxiety. Enough studies exist to permit meta-analyses of various forms of anxiety reduction methods, including biofeedback. In substance abuse studies, the various therapeutic components of the biofeedback process have not been separated. Number of sessions, attention effects, relationship with the therapist, and situational demand are some of the variables that could be involved in producing the treatment effects.

A review of biofeedback for mental disorders concluded that the method can help selected patients modify specific responses or response patterns, but it does not appear to be a treatment of choice for any mental disorder.

CHELATION

Chelation therapy is a conventional medical procedure used to treat accidental poisoning with heavy metals, such as lead, arsenic, and mercury. A chelating agent (ethylenediamminetetraacetic acid [EDTA]) is infused into the bloodstream and binds to the metal, which is then excreted renally. As an alternative medical practice, chelation therapy is used as a form of preventive medicine to remove lead, cadmium, and aluminum from the body. These substances are believed to be associated with hypertension, premature aging, memory loss, and the symptoms of Alzheimer's disease. Chelation therapy is popular as a treatment for atherosclerosis, but the randomized trial data do not support its efficacy for indications other than heavy metal toxicity.

An example of research on this method is a study of 83 male and 56 female patients (mean age, 63 years) who were treated with EDTA infusions with supportive multivitamin and trace mineral supplementation over approximately 60 days. Improvement was shown on the Cornell Medical Index emotional symptomatology. The research in this area, as with atherosclerosis, is inconclusive but does provide an insight into the popularity of the method and the loyalty of its users.

CHIROPRACTIC

Chiropractic is primarily concerned with the diagnosis and treatment of disorders of the musculoskeletal system, especially those of the spine. It was developed by a Canadian, Daniel David Palmer (1845–1913), who moved to the United States in 1895. He believed that spinal misalignment leading to abnormal nerve transmission is the cause of all illness (Fig. 28.8-3).

FIGURE 28.8-3 Daniel David Palmer (1845-1913), the founder of chiropractic. (Reprinted with permission from Shealy CN, editor: *The Complete Family Guide to Alternative Medicine*. Barnes & Noble, New York, 1996. Credited to British Chiropractic Association.)

Chiropractors diagnose illness by clinical examination and X-ray. Doctors of chiropractic (D.C.s) are trained in accredited chiropractic colleges and licensed in all 50 states. Procedures include the adjustment and manipulation of the articulations and adjacent tissues of the human body, particularly of the spinal column, and sometimes related therapies such as heat therapy, traction, and nutritional counseling. The latter procedures are included in the practices of the “mixers” but eschewed by the “straight” chiropractors who strictly limit their practice to spinal manipulation. Chiropractic is the largest independent alternative health profession in the Western world, and the United States is home to 52,000 of the world's approximately 56,000 chiropractors. Chiropractors are recognized by many government and insurance agencies and treat approximately 20 million patients in the United States annually.

There is little available research regarding chiropractic treatment of psychiatric illnesses. One small-sample (*N* = 21) study of patients with elevated blood pressure involved randomization to active treatment, placebo treatment, or no treatment. Active treatment was found to be related to reduced blood pressure while state anxiety was reduced in the active and placebo treatment. This is an early preliminary study that could form the basis for more-definitive studies.

EXERCISE

A recent review of exercise as a treatment for depression included reference to more than 1000 trials and a meta-analysis of 80 studies. Generally, the evidence indicated that a statistically significant treatment effect favored the exercise groups over that found for the comparison groups. This effect was found for all forms of

regular exercise, was independent of age or sex, and increased with the duration of therapy.

DANCE THERAPY

Dance therapy was formally recognized in 1942, when pioneer dance therapist Marian Chase (1896–1990) was hired at St. Elizabeth's Hospital in Washington, D.C. Although within the profession dance and movement are widely used synonymously, each term actually describes a point of view; *movement* encompasses the whole world of physical motion, whereas *dance* is a specific creative act within that world, usually involving music. The American Dance Therapy Association defines dance therapy as “the psychotherapeutic use of movement which furthers the emotional and physical integration of the individual.” Dance therapy sessions have four basic goals: the development of body awareness; the expression of feelings; the fostering of interaction and communication; and the integration of the physical, emotional, and social experiences that result in a sense of increased self-confidence and contentment.

There are several hundred references to dance therapy, many of which are descriptive and theoretical. The method has been applied to a wide range of psychiatric conditions. A meta-analysis of 23 studies with a total sample of 781 subjects, concluded that dance and movement therapy can be effective for treatment of a variety of symptoms, particularly anxiety. However, the research is characterized as having methodological problems such as inadequate control groups and lack of use of standardized measures.

EYE MOVEMENT DESENSITIZATION AND REPROCESSING

Eye movement desensitization and reprocessing is a method discovered by Francine Shapiro, who reported reduced levels of distress associated with traumatic memories if the recall of these memories could be accompanied with a certain form of rapid and rhythmic eye movements. While the nature of the research and the necessary components of the treatment are controversial it is generally conceded that the procedure is useful in the treatment of traumatic stress reactions in some patients. Criticisms generally concern the fidelity to standards of treatment; the necessity of eye movements; the use of appropriate control groups, research designs and measures; consideration of comorbidity; and consideration of other concurrent treatment. It has been suggested that as the quality of the research increases, there is less support for the efficacy of the technique beyond its imaginal exposure component. While initial reports concerning this technique have been encouraging, additional research is recommended. However, the current popularity of the method is notable.

CULTURALLY SPECIFIC THERAPIES

It is particularly difficult to summarize culturally specific therapies because of the extent and variety of these healing practices associated with religious beliefs and experiences that are applied worldwide. These “treatments” are not limited to geographic areas such as ayurveda in India or traditional Chinese medicine in China. Adding to the difficulty is the fact that they are not linked to written documents that expound their basic tenets or standards of practice. The writings that do exist are typically descriptive, often the publications of anthropologists or individuals who had personal experiences with these “treatments” or processes. Nevertheless they must be represented in this summary of alternative and complementary health practices because of their importance and prevalence, and their use worldwide to heal mental and physical illnesses that are inextricably intertwined with spiritual “illnesses.”

Shamanism A shaman is an individual who is believed to have the power to heal the sick and communicate with the spirit world, including primeval forces and spirits of the dead. The origin of the term appears to be from the Siberian and Ural-Altaic region, but individuals having this designation can be found in many parts of the world, including American aboriginal groups (first nations, Indians, and Alaskan natives). Qualifications of a medicine man (or woman) are determined by a series of initiatory trials and teaching and “certification” by qualified, recognized elders. These often include cleansing ceremonies (fasting, sweat lodge) and a “vision quest.”

Through the experience of trance-like states, the shaman conducts healing ceremonies that deal with the individual's loss and recovery of the soul. The ceremony is sometimes facilitated by rhythmic sounds, dancing, physical pain or privation, and the use of “spiritual herbs.” Through this process the shaman escorts the soul of the dying to the afterlife or to spiritually determined solutions to insoluble personal or social problems.

Research on this topic is generally descriptive and philosophical, although a few studies use clinical research methodology. One study reports on shamanistic contributions to the treatment of schizophrenia and other mental illness in Japan. Another study examined the outcomes and satisfaction of 188 patients treated by shamans in Taipei, Taiwan, with matched samples of patients treated by physicians. Five distinctive sickness types, including psychiatric problems, were involved. Both patient and experimenter reports indicated 75 percent improvement but greater dissatisfaction with shamanistic treatment, which is often difficult or uncomfortable.

Generally, the treatment research on shamanism can be characterized as early and preliminary. The difficulties in applying standard biomedical research designs to indigenous therapeutic practices are great, so studies are primarily naturalistic and observational. Even in the absence of efficacy data, however, the inclusion of nonharmful cultural rituals in the community mental health setting can be justified by concerns about cultural competence, the importance of community traditions, contributions to self-image, and the maintenance of socially positive behavior.

Peyotism or Peyote Religion, Native American Church Peyotism, or Peyote religion, is a Native American pantribal religious system that uses *Lophophora williamsi* (peyote), a substance obtained from the top of the peyote cactus. The substance contains mescaline, which is noted for its hallucinogenic and other mental effects. The use of peyote is part of a religious ritual that contains Native American and Christian components and promotes brotherly love, family care, self-support, and avoidance of alcohol. Peyote, like any drug, can be used to produce states of intoxication without accompanying religious intent. This is a major reason for the controversy surrounding the legalization of the substance for religious purposes. The Native American Church was formalized in 1918 to seek exemption from government prohibition of the use of peyote for sacramental purposes. Several states have supported the constitutional freedom of religion for this organization. The Supreme Court has established a legal precedent that has significant implications for the religious use of peyote and for sacramental use of all traditional substances other than alcohol. It has been estimated that there are more than 225,000 members of the Native American Church.

Research on this topic is generally difficult because of the multicomponent nature of the “treatment,” the self-selection of the church members, lack of explicit standardization of the treatment, and other factors such as the formation of social networks to help prevent relapse and the participation by many in dance ceremonies held several times a year in various parts of the country. “Treatment” may be a social system that has peyote as only one of its probable “active” ingredients. The range of related practices includes sweat lodges, other herbs, cultural reeducation, and sun dances, among others. Research results have been contradictory, and recommendations drawn from historical and epidemiological data have not been investigated in controlled studies. Several factors probably contribute to outcome success in this area as they do in all substance abuse research. Many health workers in the Indian Health Service of the U.S. Public Health Service have seen long-term sobriety associated with Native American Church attendance by former public inebriates.

Even with this embryonic state of scientific information, this alternative and complementary health practice has great momentum, and participants and believers have a powerful conviction of its benefits and personal evidence for its salutary effects on not only substance abuse but physical health, mental health, and spiritual wholeness as well.

PRAYER AND SPIRITUALITY

The many forms of prayer, the number and variety of participants, and the theological and cultural contexts within which prayer is conducted make this topic extraordinarily difficult to summarize. The broadest definition of spiritual healing includes many of the topics covered above, since practices or treatments have accompanying philosophical, theological, and supernatural concepts (viz., the concept of “qi” in traditional Chinese medicine, the Vedic origins of ayurvedic medicine, the “other-worldly” concepts of shamanism). There is considerable interest in defining the professionally appropriate place of prayer and spirituality in the practice of medicine, psychology, nursing, and other health professions. The pervasive interest in faith healing, the curative anecdotes of television evangelists, and the millions of hopeful individuals visiting religious shrines in search of relief give witness to the continuing interest in and prevalence of prayer and spirituality in the process of healing. Some religions specifically recommend against standard psychiatric therapies and offer their own approach as the only valid alternative for mental and spiritual health.

The authors believe that no responsible physician should recommend replacing safe and efficacious medical practices with various forms of faith healing or religious treatment. However, words in the title of this section—complementary and alternative health practices—were chosen deliberately to indicate that these forms of intervention may include personal practices that may facilitate, enhance, or impede standard interventions. When demonstrated to be safe and effective, some of

these may be appropriate for addition to the psychiatric armamentarium. However, no matter how safe or even effective, some may never be appropriately included for medical treatment or insurance reimbursement. Some of the uneasiness that surrounds this topic is related to reports of the withholding of medical treatments with known efficacy because of a religious disbelief in the physical reality of the symptoms or illness or the spiritual dangers of otherwise safe practices. On the other hand, there should be no opposition to the appropriate recognition of a patient's religious beliefs and behaviors in American medicine. This summary considers the health effects of spirituality and describes the research that deals with a component of spirituality, prayer.

Prayer The traditional research approach to topics of this kind is to select one component of a complex form of intervention and manipulate it in an effort to demonstrate its independent or interactive contribution to the healing process. In this case, the act of prayer can be considered a form of distant healing defined by Elizabeth Targ as any purely mental effort undertaken by one person with the intention of improving the physical or emotional well-being of another (intercessory prayer). In a similar manner, the act of prayer can be self-administered and studied as part of the therapeutic process (personal or group prayer). An additional classification provided by Jeffrey S. Levin lends itself to research strategies that can reduce the total complexity of the topic. The classification consists of a pair of dichotomous concepts that are based on whether healing has naturalistic or supernatural origins and whether it operates locally or nonlocally.

Several lines of research evaluate prayer and spirituality. A few are listed here to indicate the variety of approaches and reported effects. Several clinical reports and reviews advocate the use of shared prayer, silent prayer, and distant prayer in nursing care. One review concerned the use of guided imagery and prayer and reported the benefits experienced by both nurses and patients dealing with terminal medical conditions.

Surveys indicate that 92 percent of a sample of inner city homeless women reported one or more spiritual or religious practices. Some 48 percent reported that prayer was significantly related to less use of alcohol or street drugs or both and fewer perceived worries and depression.

Experimental research included a large randomized control trial that investigated the effects of directed and nondirected prayer and a control condition randomly assigned to 496 volunteers. The measures included self-esteem, anxiety, and depression. Agents (those praying) improved significantly on 10 measures. A significant positive correlation was found between the amount of prayer the agents did and their scores on the five objective tests. Agents had significantly better scores than did subjects on all objective measures. Patient improvement was related to the subject's conviction concerning whether he or she had been assigned to a control or experimental group.

Wound healing, the rate of reepithelialization of standardized full-thickness human dermal wounds surgically administered, was used as the outcome measure to compare therapeutic touch, Reiki, LeShan, and intercessory prayer. The five experiments conducted produced inconsistent results. The complexity of the design and treatments made this research difficult to interpret.

A randomized, double-blind, within-subject crossover research design was used to examine the effect of nontraditional distant prayer upon autonomic and CNS measures using multisite surface electromyographic recordings. Other measures included hand temperature, heart rate, skin conductance, blood volume, and pulse. Two of the four muscle regions that were monitored showed significant reduction in electromagnetic readings during and following the distant healing intervention, for most subjects. While some scant evidence supports the possibility of distant mental effects on humans, animals, and other biological systems, the confounding effects of hope, expectation, relaxation, investigator bias, and other participation activities need to be taken into account in future research. In human substance abuse research, for instance, it remains to be determined whether prayer has efficacy in addition to its nonspecific, personal treatment effects.

One randomized blinded trial of intercessory prayer (distant prayer by others) for patients who were identified by name and study number showed no benefit to those randomly assigned to be prayed for. However, the notion of praying for an unknown party by study number has been ridiculed by many critics. Also, several drug treatment officials have claimed that they “would pray for the control group.” This negative result does not disprove previous published work claiming evidence of benefit to cardiac care unit patients who were randomized to receive intercessory prayer or not receive it. In that study, complication rates and survival were significantly better in myocardial infarct patients who were prayed for.

Spirituality A large body of epidemiological research indicates that religious beliefs and practices are negatively correlated with substance abuse and positively correlated with health status. However, the investigation of these findings by clinical studies is just beginning.

One study combined nearly all the desirable aspects of research design and measurement: a homogeneous baseline group including careful selection and exclusion criteria; use of accepted diagnostic and outcome measures; random assignment to treatments; waiting list controls; posttreatment, 3-month, and 2-year follow-up; control for significant therapist variables; multivariate analyses; monitored treatments; and separate analysis of therapist religiousness and the religiously based cognitive-behavioral therapy. The clinically depressed patients were self-identified Christians. The religiously based cognitive therapy and the pastoral counseling patients obtained significantly improved posttreatment depression and adjustment scores, compared with those of the nonreligious cognitive therapy group and the waiting list controls. One interesting finding was the difference that was attributed to the superior performance of the nonreligious therapists (with values dissimilar to those of the patient) in the religious therapy group. Improvement in all three groups was equal at follow-up and greater than the posttreatment waiting list group effect. Generally, the findings indicated that religious subjects receiving religiously based cognitive-behavioral therapy did better at posttreatment regardless of the religiosity of the therapist conducting the therapy. Relevant to the discussion of this topic were the enduring effects of the pastoral counseling treatment that was used as one of the experimental controls.

Nocebo Effects Probable effects can be associated with prayer and interventions having spiritual components without knowing the causal components or mechanisms for these effects. If a therapy and therapeutic process are harmonic with the patient's value and religious system, there are likely to be enhanced treatment outcomes, even if these effects can be understood as placebo effects. However, if there is efficacy in such a process, there are also likely to be risks. The negative placebo outcomes associated with a medical or religious intervention are termed *nocebo effects*, results that worsen mental or physical health and intensify suffering.

No discussion of alternative and complementary health practices in mental health would be complete without a description of the nocebo effect. Just as positive expectations and a good doctor-patient relationship can have a positive effect on health outcomes (the placebo effect); it is reasonable to assume that negative expectation or a noncommunicative, distrusting, or hostile relationship can have negative effects on the patient.

A large number of case histories describe intensification of illness or production of death in various cultures, including the United States. Sudden death, hexes, curses, the consequences of taboo violations, voodoo death, and the production of aversive symptoms have been considered from various perspectives—anthropological descriptions, conditioned reflex learning, illusion of control studies, cerebral neurophysiology, and sympathetic adrenal responses leading to cardiovascular collapse. The point here is that the effects exist, and it is conceivable that negative expectations generated by any respected figure of medical authority could be associated with ill effects.

Obviously, all physicians and health practitioners should strive to obtain the most positive possible placebo effect in our clinical relationships, to add as much as possible to the specific efficacy of our therapies. The placebo effect was most of what physicians had to offer patients for most of medical history. The loss of doctor-patient relationships with optimal placebo effects has been blamed, rightly or wrongly, for much of the current popularity of alternative practitioners in the United States. Physicians should never worsen outcomes by painting an unnecessarily bleak picture for patients and their families, no matter how strong the medicolegal incentives are for doing so.

SUGGESTED CROSS-REFERENCES

Psychoneuroendocrinology is covered in Section 1.11 and the immune system and its interaction with the central nervous system is discussed in Section 1.12. Cross-cultural psychiatry is covered in Section 4.4, Chapter 25 discusses psychological factors affecting medical condition, and Section 25.9 discusses stress and psychiatry. Chronic pain and the placebo effect is covered in Section 28.6 and cults, quarks, and nonprofessional therapies are discussed in Section 28.9. Transcranial magnetic stimulation is also discussed in Section 31.3, other biological therapies which also contains an extensive table on herbal preparations.

The views expressed in this section are those of the authors, not those of the United States Public Health Service.

SECTION REFERENCES

*Acupuncture—NIH consensus development panel on acupuncture. JAMA *280:*1518, 1988.

Atkinson JM: Shamanisms today. Annu Rev of Anthropol *21:*307, 1992.

Bach E, Wheeler FJ: *The Bach Flower Remedies*. Keats, New Canaan, CT, 1979.

Balodhi JP: Constituting the outlines of a philosophy of ayurveda: Mainly on mental health import. Indian J Psychiatry *29:*127, 1987.

Binik YM: Psychosocial predictors of sudden death: A review and critique. Soc Sci Med *29:*667, 1985.

Calabrese JD: Spiritual healing and human development in the Native American church: Toward a culture psychiatry of peyote. Psychoanal Rev *84:*237, 1997.

Cohen AJ, Bartlik B: Ginkgo biloba for antidepressant-induced sexual dysfunction. J Sex Marital Ther *24:*139, 1998.

*Eppley KR, Abrams AI, Shear J: Differential effects of relaxation techniques on trait anxiety: A meta-analysis. J Clin Psychology *45:*957, 1989.

Ernst E, Pittler MH: Yohimbine for erectile dysfunction: A systematic review and meta-analysis of randomized clinical trials. J Urol *159:*433, 1998.

Ernst E, Rand JI, Stevinson C: Complementary therapies for depression. Arch Gen Psychiatry *55:*1026, 1998.

Fugh-Berman A: *Alternative Medicine: What Works*. Williams & Wilkins, Baltimore, 1997.

Hahn RA: The nocebo phenomenon: Concept, evidence, and implications for public health. Prev Med *26*(5[1]):607, 1997.

Keller AA, Hamer R, Rosen RC: Serotonin reuptake inhibitor-induced sexual dysfunction and its treatment: A large-scale retrospective study of 596 psychiatric outpatients. J Sex Marital Ther *23:*165, 1997.

Krieger D: *Accepting Your Power to Heal: The Personal Practice of Therapeutic Touch*. Bear, Santa Fe, 1993.

*Manukulasuriya U: Alternative medicine. NZ Med J *112:*103, 1999.

*Newall CA, Anderson LA, Phillipson JD: *Herbal Medicines: A Guide for Health-care Professionals*. Pharmaceutical Press, London, 1996.

O'Laoire S: An experimental study of the effects of distant, intercessory prayer on self-esteem, anxiety, and depression. Altern Ther Health Med *3:*38, 1997.

**PDR for Herbal Medicines*. Medical Economics, New Jersey, 1999.

Propst, LR, Ostrom R, Watkins P, Dean T: Comparative efficacy of religious and nonreligious cognitive-behavioral therapy for the treatment of clinical depression in religious individuals. J Consult Clin Psychol *60:*94, 1992.

Ritter M, Low KG: Effects of dance/movement therapy: A meta-analysis. Arts Psychother *23:*249, 1996.

Shapiro F: Errors of context and review of eye movement desensitization and reprocessing research. J Behav Ther Exp Psychiatry *23:*313, 1996.

Singh YN: Kava: An overview. J Ethnopharmacol *37:*13, 1992.

Soholm B: Clinical improvement of memory and other cognitive functions by Ginkgo biloba: Review of relevant literature. Adv Ther *15:*54, 1998.

Targ E: Evaluating distant healing: A research review. Altern Ther Health Med *3:*74, 1997.

Turner JG, Clark AJ, Gauthier D, Williams M: The effect of therapeutic touch on pain and anxiety in burn patients. J Adv Nursing *28:*10, 1998.

Tyler ET: *Herbs of Choice: The Therapeutic Use of Phytomedicinals*. Pharmaceutical Products, New York, 1994.

Walsh RN: What is a shaman? Definition, origin and distribution. J Transpers Psychol *21:*1, 1989.

*Wong AHC, Smith M, Boon HS: Herbal remedies in psychiatric practice. Arch Gen Psychiatry *55:*1033, 1998.

Warren KS, Mosteller F: Doing more good than harm: The evaluation of health care interventions. Ann NY Acad Sci *703:*1, 1993.

*Wirth DP, Cram JR: The psychophysiology of nontraditional prayer. Int J Psychosom *41:*68, 1994.

Yasgur J: *Homeopathic Dictionary and Holistic Health Reference*. Van Hoy, Greenville, PA, 1998.

28.9 NONPROFESSIONAL THERAPIES, QUACKS, AND CULTS

LOUIS J. WEST, M.D.

In 1910 Robert Means Lawrence wrote:

> It is worthy of note that the most notorious quacks, often men of genius and education, though mentally ill-balanced, and morally of low standards, have been great travellers and shrewd observers of human nature. When such an one becomes ambitious to acquire wealth, he likely to prove a dangerous person in the community.

Nonprofessional modes of therapy for psychiatric disorders have multiplied rapidly in the United States and other developed countries during recent years, along with the general explosion of alternative medicine. There is now a bewildering variety of such methodologies. Quacks, who have always touted cures for physical diseases, today widely extend their claims to include mental and emotional illnesses as well. Meanwhile, there has been an alarming growth in the number of totalist cults, many of which claim to offer their followers health benefits—physical, mental, emotional, and spiritual. Physicians, other health professionals, and the public must be aware of these developments and recognize both their attractions and their dangers.

NONPROFESSIONAL THERAPIES

Some nonprofessional approaches to psychological self-improvement are essentially benign carried out by organizations engaged in procedures that may benefit some of their members, clients, followers or customers. However, many enterprises pose definite risks to those lured by their promises. All represent potential hazards if pursued as substitutes for professional care needed by a patient. A real danger exists that serious symptoms—physical, mental, or emotional—will be misunderstood, ignored, or masked during the course of nonprofessional interventions. Symptoms may even be temporarily relieved through the power of suggestion or the placebo effect but to the dangerous neglect of underlying pathology, which patients often erroneously attribute nonexistent to their nonprofessional therapists. Patients also are likely to expect professional responsibility, confidentiality, and other ethical standards or commitments that are legally not required of nonprofessionals and, in fact, usually do not obtain.

Ruthless cult leaders and venal quacks may offer psychiatric benefits in the cynical knowledge that they are perpetrating hoaxes and swindles. Even sincere nonprofessional healers tend to assume that their methods possess wide or universal applicability and to offer them to all comers with uncritical enthusiasm. Indications and contraindications for treatment are usually lacking in the formulations of such self-styled healers. Many of them assume that anyone is bound to benefit from their procedures, regardless of the sufferer's specific complaint. Unlike such persons, ethical practitioners of health-related professions realize that risk is inherent in any specific treatment. Untoward reactions inevitably occur in some patients. A qualified professional must be able to recognize latent and manifest complications of therapy and to take prompt and appropriate action if needed to help or protect the patient. When a given treatment produces dangerous side effects, fails to produce the desired therapeutic benefit, or (as may happen) fails to prevent—or even causes—exacerbation of the patient's illness, health professionals understand, but nonprofessionals often do not realize, that such treatment must be discontinued or modified. Patients seeking help from amateurs, quacks, cult leaders, gurus, faith healers, and assorted alternative practitioners are likely to believe that these nonprofessionals possess the knowledge, skills, clinical judgment, and ethical values of physicians. Disaster may result.

Contemporary quacks, while thriving, have been subject to surprisingly little scrutiny. On the current scene, with its myriad commercial health-related promoters, it is extremely difficult to differentiate the traditionally defined quack from the nonprofessional or alternative therapist, let alone the genuine professional practitioner. For example, all are now free to advertise. Many individuals, organizations, agencies, sects, and cults tout nonprofessional mental healing, psychotherapy, and self-improvement to the public. These vendors are a large and constantly changing group. Some purport to be purely spiritual, mystical, or religious. Others profess to be totally scientific, secular, and rational. Between these extremes is a spectrum of enterprises that partake, to various degrees, of both mystique and pseudoscientific rationale.

Another spectrum applicable to nonprofessional therapeutic enterprises is the specificity of goals. Some operators and organizations offer procedures that are directed precisely at highly defined problems or symptoms, such as smoking, overeating, alcohol abuse, drug addiction, and sexual inadequacies. Others strive to generally restore or maintain health, regardless of the disease or complaint in the particular case. Still others promote a general state of self-improvement, going far beyond the health orientation proper but including it. The widely accepted idea (especially popular in Western societies) that everyone needs self-improvement, better health, greater longevity, and development of untapped potential makes such approaches essentially universal in their stated applicability and seductively attractive.

MYSTIQUES

The past few years have seen a renewed and unabashed interest by some psychiatrists in parapsychology, including telepathy, psychokinesis, clairvoyance, and prescience. One prominent colleague has written positively on reincarnation, another on mediums, a third on thought transference in dreams. Experiences with hallucinogenic drugs have led others to formulate different, perhaps mystical, ways of knowing reality. Even biological scientists and astronauts have involved themselves in parapsychological experiments that would have seemed outrageously mystical 20 years ago but are now taken as a matter of course. Astrology and prophecy seem to be as much in vogue today as they were in the fifteenth century. Many citizens eagerly accept the idea that the earth is being visited regularly by benign denizens of other solar systems, ferried by spacecraft known as unidentified flying objects (UFOs). A professor of psychiatry has treated a number of people who claim to have been abducted and manipulated by these aliens; he believes and publishes their stories.

The extent to which mental illness today is recognized as bona fide disease may be reflected by the fact that faith healers of this day seem to be concentrating more of their attention on it. In the distant past the healing of mental disorders remained largely the province of various priesthoods. *Madness*, defined as demonic possession, was cast out (exorcised) through the use of interpersonal or group ceremonies that were given religious definitions. Today some of the most successful nonprofessional healing and self-improvement programs offer a combination of old-fashioned religious and modern pseudoscientific (or science fiction) formulations. The reasons for such developments probably include the power of suggestion derived from the mysteries of both science and religion and a general groundswell of antiintellectual sentiment within the New Age (actually medieval) fashion of thought.

QUACKS

Voltaire commented that the charlatan was born when the first knave met the first fool. Quacks probably deserve to be considered charlatans, but their victims (or patients) are not necessarily fools. They may be merely ignorant, or superstitious, or naive, or suggestible, or misled by sincere friends, or misguided by certain religious teachings, or frightened, or even desperate enough to try anything when other remedies have failed or when more responsible helpers can offer no hope. Such patients and their loved ones may be vulnerable to the false or exaggerated claims of the quack, whose primary motives are usually financial gain, prestige, power, or all three. The quack thrives by claiming to help people with ailments for which cures or treatments have not yet been satisfactorily developed by scientific medicine; for diseases that are particularly feared such as cancer, acquired immune deficiency syndrome (AIDS), sexually transmitted disease, tuberculosis, or epilepsy; with chronic, recurrent, or intermittent syndromes such as arthritis, multiple sclerosis, or vascular headaches; and today, for nervous and mental disorders of all kinds.

Famous quacks, both men and women, have often been striking in appearance, personality, or both. They usually attach phony degrees and titles to their names and

may display false or misleading credentials on their office walls. Testimonials by persons allegedly helped or cured are commonly used, not only by ordinary quacks but by religious faith-healers as well, to enhance their prestige and to increase the suggestibility of prospective patients. One of the most famous quacks in history was a nineteenth century Englishman named John St. John Long. While he specialized in curing (usually healthy) patients of imaginary tuberculosis, he also ventured into psychiatry. He stated, “The great power I possess in extracting fluid from the brain, has enabled me to reach the very seat of disease.” He was a handsome and dynamic man who was said to be particularly successful in treating unhappy female patients.

An important element of the continuing success of quackery, and of healing cults, especially in relationship to psychiatric disorders, is mistrust or resentment toward the established practice of medicine. Physicians have always been vulnerable to attack. Keeping mankind healthy and out of the grave is an impossible task. Health professionals are likely to be prosperous (evoking envy); they are busy, and their professional manner may be brusque; the therapy they prescribe may be unpleasant, protracted, and without guarantees of cure. Quacks have traditionally enhanced their own credibility by disparaging reputable physicians and legitimate medical practice. They play upon public ambivalence toward the medical establishment, which has led in recent times not only to the thriving of quacks, but also to the spread and acceptance of alternative medicine in all its manifestations. Quacks deliberately misrepresent themselves as possessing medical or other health-professional qualifications and skills that they do not posses. They are active in all aspects of the health field, preying on patients with both physical and mental illnesses and on patients' families as well. Even licensed health professionals may properly be termed quacks if they dishonestly or fraudulently claim to offer cures through knowledge that they lack or use techniques or medications that they tout as beneficial or curative when they have no known value (other than the placebo effect and the powerful effect of deliberate suggestion) for the condition being treated. Spiritual healing is, of course, extremely difficult to attack under laws concerned with pure foods and drugs or governing the practice of legitimate health professions, and it is not surprising that many quacks operate under this umbrella.

Despite government regulations, agencies to control fraud and misrepresentation, and widely available educational materials on health matters, quackery may have prospered in the last few years partly because of the high cost and incomprehensibility of modern biotechnical medicine; partly because so many persons lack health insurance; partly because HMOs demand rapid, high-volume, impersonal health care delivery by legitimate practitioners; partly because the growing numbers of elderly adults are vulnerable to extravagant claims of relief or cure for their chronic ills; and partly because many contemporary young persons are open to antiestablishment attitudes that leave them vulnerable to the blandishments of persuasive quacks. In their skepticism about large commercial and governmental institutions, including universities and large research organizations, numerous young people have been caught up in a wave of antiintellectualism that began in the 1960s and has not yet abated. Meanwhile the quacks are constantly at work to exploit every advantage, legally and culturally, in their untiring efforts to make money at the suffering public's expense.

CULTS

Periods of great turbulence in human history are often accompanied by the emergence of unusual numbers of cults, religious and otherwise. Some cultlike activities and the beginnings of a counterculture (beatniks and the beat generation) emerged in America in the 1940s and 1950s following World War II and the Korean War. A new set of disturbances in the American culture welled up during the 1960s with the expansion of an unpopular war in Southeast Asia, massive upheavals over civil rights, and a profound crisis in values defined by unprecedented affluence on one hand, potential thermonuclear holocaust on the other. Young people were caught up in three rebellions: red (New Left) against inequities of political and economic power; black, against racial injustice; and green (the counterculture) against materialism in all its manifestations including individual and institutional struggles for power. Drug abuse and violent predators took an awful toll among the counterculture's pilgrims (hippies) in the late 1960s. Many fled to form colonies, generally called communes. Others turned to the apparent security of paternalistic religious and secular cults, which have been multiplying at an astonishing rate ever since.

Definitions of “cult” in *Webster's Third New International Unabridged Dictionary* include “a religion regarded as unorthodox or spurious (the exuberant growth of fantastic cults); also, a minority religious group holding beliefs regarded as unorthodox or spurious,” a system for the cure of disease based on the dogma, tenets, or principles set forth by its promulgator to the exclusion of scientific experience or demonstration,” and “great or excessive devotion or dedication to some person, idea, or thing.” Some of Webster's other definitions are even more benign (e.g., America's cult of home fixer-uppers.)

The term “cult” is used here primarily to describe groups that satisfy one or another of Webster's definitions but that also can be properly described as totalist, after Robert J. Lifton, who derived his concept of totalism from Erik Erikson's contribution to *Totalitarianism*. Totalism describes a tendency to “all-or-nothing emotional alignment” that can be exploited by “those ideologies which are most sweeping in content and most ambitious—or Messianic—in their claims, whether religious, political, or scientific. And where totalism exists, a religion, or political movement, or even a scientific organization becomes little more than an exclusive cult.”

The 1985 Wingspread Conference report provided the following definition (slightly modified):

> Cult (totalist type): a group or movement exhibiting a great or excessive devotion or dedication to some person, idea, or thing, and employing unethical, manipulative or coercive techniques of persuasion or control (e.g., isolation from former friends and family, debilitation, use of special methods to heighten suggestibility and subservience, powerful group pressures, information management, promotion of total dependency on the group and fear of leaving it, suspension of individuality and critical judgment designed to advance the goals of the group's leaders, to the possible or actual detriment of members, their families, or the community. There are many religious sects, new and old, that are not cults as defined above.

Totalist cults are likely to exhibit three characteristics: (1) excessively zealous, unquestioning commitment to the group and its leadership by the members, (2) manipulation and exploitation of members, and (3) harm or the danger of harm. Many groups do not fit neatly into a category of sect, commune, or cult. Furthermore, groups may change their characteristics over time, becoming more or less like cults, totalist or otherwise. Terms such as “new religious movement” have been used by some to describe certain cultic groups. A problem with this approach is that it may lend unwarranted respectability to a less-than-respectable enterprise.

Persons between the ages of 15 and 30 are especially subject to cult recruitment. A recent study of students in San Francisco found that half were open to accepting an invitation to attend a cult meeting, and approximately 3 percent reported they already belonged to cultic groups. No single personality profile characterizes those who join cults. Many well-adjusted, high-achieving persons from intact families have been successfully recruited by cults, as have individuals with varying degrees of psychological impairment. However, to the extent that predisposing factors exist, they may include naïve idealism, situational stress (frequently related to normal crises of adolescence and young adulthood, such as romantic disappointment or school problems), emotional dependency, disillusionment, or an excessively trusting nature. Ignorance of the ways in which groups can manipulate individuals is a relatively general characteristic of cult victims—until it is too late. Many contemporary cults in the United States, while not defining themselves primarily in terms of some healing system, do nevertheless claim extravagant healing powers, whether their shibboleths be religious, psychological, commercial, mystical, or related to self-improvement.

Estimates suggest that there are now more than 5000 cults in the United States. Many (but not all) of these are religious and, of course, they are not all alike. Some are very small, between 10 and 50 people; others are large, powerful, wealthy, and international. By definition cults pose significant threats to the personal freedom and well-being of their members. Regrettably, despite all the evidence of misdeeds perpetrated by cults in the name of religion, these threats are largely covered up, minimized, and obscured by the cults, their apologists and protectors, and even the media. Nevertheless, much information has accumulated from exposes of various cult-related scandals; from cult refugees; from families, relatives, and friends of cult victims; and from a few objective studies. Unfortunately, it is difficult to obtain accurate data by direct investigations of cults because they systematically deceive the public and, of course, would-be investigators. They conceal facts; they harass their critics, and they intimidate and dominate their members. All of this is designed to prevent a free flow of information. It is hard to carry out good scientific research when genuine direct access to or truly unrestricted observation of, the phenomena or subject material is forbidden and when the cult leadership controls or can influence the circumstances of scrutiny, the choice of subjects, or the nature of the responses. Even so, existing data now suffice to convince any reasonable person that the claims of harm done by cults are bona fide. Many people are already dead or dying, ill or malfunctioning, crippled or (if young) developing improperly as a result of their involvement in cults. They are exploited; they are used and misused; their health suffers; and they are made to commit improprieties ranging from lying to murder. Their families are often devastated and increasingly seek psychiatric advice. Estimates run as high as 10 million Americans who have been at least briefly involved with cultic groups during the past 30 years.

Many apologists—individuals and organizations—contribute to the cults' veneer of respectability behind which ugly and harmful things are happening. Some apologists appear to be romantics who project onto the cults some of their own hopes for religious reform, spiritual rebirth, rejection of materialism, or even escape from the dangers of the thermonuclear age. Other apologists, including some civil libertarians, take a more pragmatic stand, shrugging off whatever abuses the cults may perpetrate and pointing out that any countermeasures would be unacceptable because they violate freedom of religion as guaranteed by the First Amendment to the Constitution of the United States. Other apologists appear to have been successfully deceived by charismatic cult leaders or their representatives. Certain politicians fall into this category, in some cases after the cult has contributed money to their election campaigns. Some armchair academicians who either have never seen the destructive effects of cults or prefer to deny their reality will happily discuss cults (perhaps using another definition) at length and in elaborate terms as an interesting new ferment in contemporary society or in some other uncritical way. If the apologists are church officials, physicians, or behavioral scientists, the grateful

cults have been known to reward them with grants, awards, published praise, and even “research” opportunities.

The Wingspread Conference pointed out that cults arouse concern because of their unethical or manipulative practices and their lack of consideration for the individual's needs, goals, and social attachments. Of even greater concern is that these practices often result in harm to persons, families, and society at large. A number of specific harms are outlined in the Wingspread Conference report, broken down into five general areas: (1) individuals and families, (2) government and law, (3) business, (4) education, and (5) religion. Although the Wingspread outline of harms is fairly comprehensive, it does not include a classification for certain major crimes, such as fraud, rape, battery, torture, child abuse, and murder, all of which have been committed at the instigation of cult leaders or in cult settings.

There are occasional flurries of public alarm about some cults' methods of recruitment, exploitation of members, restriction on members' freedom, retaliation against defecting members, struggles with members' families engaged in rescue operations (including so-called deprogramming), dubious fiscal practices and the like. Sometimes the public's concern has been increased by death threats against investigative reporters, leaked internal memoranda justifying violence, the discovery of weapons caches and attacks on opponents.

Some cultic organizations are relatively subtle about recruitment, albeit harsh when it comes to defections. Others are tireless in their recruitment activities. Many use techniques that in some respects resemble the political indoctrination methods prescribed by Mao Tse-tong during the communist revolution and its aftermath (1945–1955) in China. These techniques, described by the Chinese as “thought reform” or “ideological remolding,” were labeled “brainwashing” in 1950 by the American journalist Edward Hunter. Such methods were subsequently studied in depth after the Korean War by a number of Western scientists including Lifton, Edgar H. Schein, Albert Biderman, Bernard Siegel, Margaret T. Singer, and this author. Many have noted similarities between the Maoist techniques and those of contemporary totalist cults.

Recruitment and Indoctrination Successful indoctrination of recruits by a typical totalist cult is likely to include many if not most of the following elements:

1. Isolation of recruits and manipulation of their environment.
2. Control over channels of communication and information.
3. Debilitation through inadequate diet and fatigue.
4. Degradation or diminution of the self.
5. Induction of insecurity, fear, and confusion, with joy and certainty through surrender to the group as the goal.
6. Peer pressure, often applied through ritualized “struggle sessions” generating guilt and requiring open confessions.
7. Insistence by seemingly all-powerful hosts that the recruits' survival—physical or spiritual—depends on identifying with the group.
8. Assignment of monotonous tasks or activities, such as chanting or copying written materials or mind-numbing repetitious lessons in so-called classes.
9. Acts of symbolic betrayal or renunciation of self, family, and previously held values, designed to increase the psychological distance between recruits and their previous way of life.

Of course a cult that uses such drastic techniques to control and exploit hundreds or thousands of members does not begin by doing so with each novice. The cult's initial approach is often made by an attractive peer, maybe of the opposite sex, to an individual who seems an appropriate target, perhaps one who is vulnerable through unchanneled idealism, homesickness, uncertainty of purpose; who is traveling or otherwise away from usual surroundings; or who simply happens to be alone. In a friendly way, with much smiling and casual touching, a conversation is generated during which the recruit's interests and background are elicited. The scout then happily announces that some friends who are interested in the same things (e.g., peace, poverty, altruism, racism, self-improvement, the meaning of life) are meeting that very evening, and there will be a free meal and a guest speaker. By such means, aided by intense eye contact, flattery, and personal attention, the recruit is maneuvered into attending the first meeting, study group, or the like. There may be no religious discussion. Instead, there might be a rousing but nebulous lecture followed by cheery talk, sincere smiles, much eye contact, hand-holding, and general expressions of great affection. Toward the evening's end the recruit may be encouraged to attend a 3-day workshop at an attractive home or a ranch located in the country, “to learn more about us and to allow us to get to know you better.” Manipulation, sensory barrage, and indoctrination begin in earnest at the retreat. Long hours, continuous frantic pace, cheering, chants, and constant stimulation by the group are likely to render the young person fatigued, suggestible, and compliant.

Gradually, over a period of days, weeks, or even months, information about the group's identity and the demands it places on members are more fully revealed. But recruits, isolated from their usual surroundings and overwhelmed by the group, lose the ability to evaluate these items according to their usual frame of reference. Perhaps the recruits want to leave, but the cars have been moved. Everyone insists that they should stay. They are wanted. They are loved. They yield. Sometimes they actually become entranced, entering a dissociated state of altered consciousness, with subsequent amnesia for significant portions of these crucial early hours and days.

In the weeks that follow, the recruits are constantly accompanied by members of the group who surround them with love, affection, and camaraderie and increasingly exhort them to surrender their autonomy, relish self-sacrifice, embrace the cult's discipline, and do unselfish work. In return they will receive health benefits, expanded consciousness, mental tranquility, a divine message, inspiration, a new way of life, a sense of absolute security through identification with the all-encompassing group and its omnipotent leader, or the like. When the adaptation process has progressed far enough, the leaders may judge the recruit ready to assume the duties of full membership. Depending on the cult these may include recruiting others, office work, internal housekeeping tasks, menial labor, raising money on street corners or airports, parading and chanting, even scavenging for edible garbage. The new cultist is progressively influenced to turn away from friends, family, and career and to donate material goods and earnings to the organization. Some will be urged to make a will in favor of the leadership, to agree to carry out all commands given by those in charge, to make new contracts for advanced courses, or otherwise to deepen their bondage. As time passes some cult members psychologically deteriorate. They can become incapable of complex, rational thought. Their responses to questions become stereotyped and they find it difficult to make even simple decisions unaided. Their judgments about events in the outside world are impaired. At the same time their insight is so reduced that they fail to realize how much they have changed.

After months or years of membership, such a former recruit may emerge from the cult, maybe rescued by friends or family or perhaps having escaped following prolonged exploitation and suffering that culminated either in revolt against a particular outrage or in gradual, painful disillusionment. Many such refugees appear dazed and confused, unable to resume their previous way of life, and fearful of being captured, punished, and returned to the cult. Those who have worked with these refugees from cults find that many experience thinking difficulties and marked impairment of self-expression. *Floating* is a frequent phenomenon, with the excultist experiencing dissociative symptoms, a resurgence of the cult-derived pseudoidentity, and an impulse to return to the cult and beg forgiveness for having escaped. Other frequent symptoms of these refugees include depression, indecisiveness, a general sense of disorientation, patchy amnesia, flashbacks, emotional numbing, nightmares, and other symptoms of posttraumatic stress disorder. The following criteria apply to most of the cults from which such refugees are likely to come:

1. The venture was initiated by a self-proclaimed leader who claims to have been chosen by a “higher power” to lead such a group or to have discovered some great cosmic or ultrascientific secret justifying this leadership. (Leaders of outer space cults often claim flying saucer and UFO personnel have assigned them their mission.)
2. The cult has developed a double set of ethics to promote the group welfare: one for use within the cult (e.g., to be open and honest with each other and with the leader) and another for use in dealing with nonmembers (e.g., deceive and manipulate). Because of the double standard of ethics in the cult, duplicity is often practiced in fund-raising, the declared premise being that anything that advances the group's glorious ultimate purpose is justified.
3. The cult has required the recruit to undergo a major change in lifestyle.

The United States has at least 10 major types of totalist cults, and within each type the cult exhibits its own peculiar beliefs, practices, and social mores. The list below is not exhaustive, but most of the current American cults can be classified under one of the following headings:

1. Neo-Christian religious cults
2. Hindu and Eastern religious cults
3. Occult, witchcraft, and satanism cults
4. Spiritualist cults
5. Zen and other Sino-Japanese philosophical-mystical cults
6. Race cults
7. Flying saucer and outer space cults
8. Psychological cults
9. Political cults

10. Communal, self-help, or self-improvement groups that have evolved into cults

Cult Indoctrinee Syndrome Various signs and symptoms observed by mental health professionals over the past 25 years suggest that a cult indoctrinee syndrome may be defined. Features include:

1. Drastic alteration of the victim's value hierarchy, including abandonment of previous academic or career goals. The changes are often sudden and catastrophic, rather than the gradual change that might result from maturation or education.
2. Reduction of cognitive flexibility and adaptability. Victims answer questions mechanically, substituting stereotyped cult-specific responses for what their own previously reasoned answers might have been.
3. Narrowing, blunting, or artificiality of affect. The victim may appear emotionally flat and lifeless or almost frantically cheerful and ebullient.
4. Regression. Victims become childishly dependent on the cult leaders, who are expected to make decisions for them.
5. Sometimes there are physical changes, such as pallor or weight loss, with deterioration in the victim's physical appearance, strange or masklike facies, blank stare or darting, evasive eyes.
6. Clear-cut psychopathological changes may appear, including depression, dissociative phenomena, anxiety states, obsessional ruminations, delusional thinking, hallucinations, and various other psychiatric signs and symptoms.

If the opportunity occurs, proper evaluation of a cult indoctrinee should include a thorough history obtained from at least two persons—parents, siblings, spouse, or friends who knew the person well in precult days and (when possible) have had contact with the victim during the cult experience. Special attention should be given to ascertaining the indoctrinee's range of precult intellectual and other interests and accomplishments and to assessing how these presently appear to the informant. Comparison of precult and present behavior is important, even though the family members or friends may not be able to label what they notice with technical terms. Observers often use such descriptions as “spacey,” “zombie,” “programed,” “not herself anymore,” “he's a different person now,” “they changed her somehow,” and “he has a ten-thousand yard stare.” Several examples or instances of each aspect of change should be elicited from the informant, because this gives the professional interviewer a better appreciation of what is being reported.

Cult Refugees, Survivors, or Veterans Mental health professionals are seeing a growing number of excultists as patients. These individuals are returning to society or otherwise trying to resume a normal life after months or years in cults. Some studies report that certain of those who become involved in cults were maladjusted, troubled, or even psychiatrically ill beforehand. However, clinicians with the largest experience agree that most adolescents and young adults who join cults were not seriously disturbed to begin with. Many are fairly well educated. Especially at times of dislocation, disorientation, loneliness or existential crisis, they may become vulnerable to approaches by cults looking for just such persons to recruit. The larger cults instruct recruiters to seek those traveling alone; newly arrived foreign youths; students entering university life for the first time; persons who seem lonely, alienated, or displaced; those seeking companionship and a sense of meaning; or just anyone else who might be a good catch. Even if 9 out of 10 slip through the net, the tenth may be a keeper, and the net can be cast again and again. Any open-minded, honest, idealistic young person who is interested in self-improvement, spiritual development, or making the world a better place is vulnerable. While most contemporary cults recruit mainly young adults, a few also seek elderly adults or entire families. Family life can be profoundly disturbed by cults in many ways. Sometimes a whole family becomes a group of refugees or the family is fragmented when some members escape and others remain in the cult. Certain family backgrounds have been suggested to render young people vulnerable to the lures of cults, and some families do foster a combination of indecisiveness and rebelliousness that could make a cult seem attractive to a young person seeking a more perfect family. Some cultists do come from disturbed family backgrounds; however, most seem to come from the same variety of families as their demographically matched contemporaries. Too much blame is often placed on cult victims or their families for their plight (“blaming the victim”) while the ruthless, cunning, and often sophisticated techniques used by cults to attract, capture, and retain their member-victims are ignored.

Each cult purports to offer an improved or healthier state of mind, an expanded state of being, a moral or spiritual state of righteous certainty, and the privilege of belonging to a superior elite. This supposedly beneficial state can be reached only by following the narrowly prescribed pathways of the particular group. To grasp this approach to life, the new recruit must surrender a critical mind, must yield to the flow of force, must have childlike trust and faith. Numerous methods are used to manipulate recruits into such an accepting state of mind. They are lured, cajoled, induced to suspend the use of rational judgment and urged to limit their attention to what is at hand, to drop ties with the past (in some cases to leave families, friends, jobs, or schools), to shut out all competing ideologies, to block off memories of the past, and to erase all doubt.

These ends require great mental avoidance. Attention must be constricted and contact with families and past life must be minimized or abandoned. Any sense of personal independence or rebellion against the cult leader must be suppressed and an accepting, noncritical mental stance must be attained and maintained. This state of mind is induced by hours of incessant, mind-numbing solitary or group chanting; long repetitive lectures, guilt-inducing dogma, endless classes, and “training” such as prolonged staring at one another; and interpersonal and group methods of every kind. When the desired results are achieved the new cultists do not necessarily appear mentally ill or bizarre. However, family and friends who previously knew them well may observe distressing changes. Attention is now constricted. New information is avoided. “She doesn't seem to register what we say.” “His mind seems closed.” Conversation is monotonous and repetitious, filled with the dogma, catchwords, and phrases of the cult. Dissociation, denial, suppression, and regression are implicated in the victim's maintenance of the cult's formulations of life. Of course such maintenance is reinforced. The more the person conforms to cult requirements, the more approval and affection is bestowed by the group.

By regressing mentally and socially, conforming to the leader's dictates, and identifying with the leader's power, new cultists may feel secure. They are assured that henceforth they need never feel lonely and uncertain. The suspension of critical, rational thought and even verbal intelligence is urged or even required. The guru, prophet, cult leader, or surrogates persuade the follower to accept, surrender, melt, and obey. While maintaining such a state, even for years, some cultists may be laboring up to 20 hours a day in fields or shops, fund-raising, baby sitting, or performing various chores. In some cults they wear used clothing and eat cheap food while the cult leaders wear jewels, drive expensive cars, live in palatial homes, use followers as servants, and control all of the money and resources that the members turn over to the cult. These contrasts are justified by the leaders as being proper, even necessary, for the good of the group and are passively—even gladly—accepted by the followers.

Pseudoidentity in Cult Victims Families and friends of cult victims regularly complain that their loved one has profoundly changed and has “become a different person.” Prolonged clinical experience has led to formulation of a previously undefined alteration of identity, called “pseudoidentity,” to account for this phenomenon. Dissociative features of various types and degrees may be components of various psychopathological entities. However, the distortion or alteration of a person's identity and the appearance of a new and different persona can involve denying the meaning of a great deal of previous experience, cutting off many emotional associations that connect individuals to their personal histories, and distorting recall (if not suppressing portions of the past). Certain types of trauma in childhood are believed to sow the seeds of later malfunctions in the victim's associative resources, with a resulting clinical picture known as dissociative identity disorder. Since the early descriptions by Morton Prince (1906) and Pierre Janet (1907) there have been many psychodynamic theories that largely explain dissociative identity disorder and other major dissociative disorders (e.g., fugue, amnesia, somnambulism) in terms of inner conflicts. Not well studied, however, is how particular forms of environmental stress in adults can disrupt the normally integrative function of identity. Under such conditions, victims may use dissociative mechanisms to change their usual or normal expressions and behaviors that compromise personality and then assume an altered persona or pseudoidentity that facilitates coping with the situation they are in, regardless of how they got there.

Conditions of brutal captivity, such as those experienced by prisoners of war or civilian victims of hostage taking, in which the captor seeks to force a false confession or induce compliant behavior, can generate posttraumatic stress disorder with prominent dissociative features. During the Korean War the Chinese Communists' success in eliciting false confessions of war crimes (e.g., germ warfare), self-denunciations, and participation in anti-American propaganda activities by some captured American flyers was in great part achieved because the captors' absolute control over the prisoners' environment enabled them to inflict a variety of stresses that induced state of debility, (especially through sleep deprivation), dependency, and chronic apprehension or dread. In this state suggestibility is heightened, and identification with the aggressor (captor) may occur.

Clinical and experimental experience is replete with examples of increased suggestibility or controllability of individuals during altered states of awareness induced by hypnosis, psychotropic substances, environmental manipulation, sleep deprivation, sensory isolation, powerful emotions (especially in large groups), and other special circumstances. But pseudoidentity is more than a temporary role assumed by the subject in a laboratory exercise or a transient period of intoxication. It resembles the alter in a genuine case of dissociative identity disorder. However, the pseudoidentity differs from the alter in many important respects.

Pathogenesis Dissociative identity disorder is most likely a consequence of early childhood trauma, with symptoms appearing later in life as a result of inner conflicts combined with some interpersonal or situational crisis. A pseudoidentity is usually generated by external stress originating in the environment of a person who previously may have been quite free of any signs or symptoms of personality malfunction and for whom the new persona represents a transformation required to meet

the demands of a new life situation (e.g., membership in a totalist cult) dramatically different from the previous one.

Psychopathology The genuine dissociative identity disorder patient may have more than one alter (although in the author's opinion, this is rare); but those with pseudoidentity change (swiftly or gradually) to display a single different personality. Under certain conditions the patient may abruptly switch back and forth between behaviors characteristic of the two personalities but without the typical dissociative identity disorder boundaries between the two and without the patient's awareness of one self being separate from the other. In dissociative identity disorder the different alters or personalities primarily reflect facets of the original character; in pseudoidentity the new personality primarily reflects the new situational forces and requirements. In dissociative identity disorder the original identity is usually unconscious of the existence of others as they shift back and forth; in pseudoidentity the original identity remains but is overlaid or enveloped by the new.

Treatment Therapy of both syndromes requires appreciation of the mental mechanisms involved, the reality of traumata or stress—however subtle—in pathogenesis, and the techniques proved useful in managing dissociative disorders. While in psychotherapy for dissociative identity disorder the usual goal is reconciliation and integration of the alters into a new and healthier whole, the therapeutic goal for the patient with pseudoidentity is restoration of the original identity. However, patients then usually require treatment for the residual posttraumatic stress syndrome that is the legacy of the stress that produced and maintained the pseudoidentity syndrome and for the special problems of the restored but battered original identity.

Prognosis Dissociative identity disorder is difficult to treat, and untreated cases rarely remit. The outlook is generally better for the patient with pseudoidentity, although the syndrome may become chronic like any dissociative disorder or (in the old terminology) monosymptomatic major hysteria. Sometimes merely returning patients to their original life situations (or even a neutral environment in which information is freely and honestly exchanged and nonexploitive people are available for support) will result in the abrupt or gradual disappearance of the pseudoidentity. However, these patients then face resuming many long-neglected functions and working through the complex emotional aftermath of having—for whatever period of time and to whatever degree—been a different person. This process usually requires well-informed professional treatment.

Reentry Counseling Perhaps because of the contemporary ubiquity of computers, the terms “programing” and “deprograming” have become part of the cult-relevant language. The social and interpersonal influences and indoctrination methods used by cultists produce the stereotyped responses described above, resembling computer programs. The term “deprogramer” is now used by some to describe a person who gives cultists who have been forcibly extracted or have escaped from the cult information that might cause them to reconsider their commitment to the cult, to leave it or (having left) to understand what happened and to make a new start. Those who do this helping work (many themselves former cult victims) are now usually called “reentry counselors.” They provide information to persons who are considering joining cults, who have already joined cults, or who have left cults. The reentry counselors know how they themselves were persuaded to join, how the cult indoctrinated them to its ways, and what their mental states were while they were members. Some have collected considerable data (e.g., clippings from newspapers, books, videotapes, recorded interviews). Cult members appreciate the facts about their former cults and understand feelings of awakening as they begin to assimilate information previously avoided or denied them by the cults. Reentry counselors and support groups of former cult members can be helpful allies to the mental health professional who is treating a cult victim. Just as there is no single pattern of precult experiences, there is no uniform set of postcult problems. However, in addition to the identity problems and dissociative symptoms described above, three major difficulties are faced by most persons who emerge from cults: feelings of loneliness, difficulty with decision making, and a mood of depression or a pervasive sense of existential meaninglessness. Eventually, ex-cultists remark about great improvements in their mental states after being out of the cult for a time, especially if appropriate treatment, reentry counseling, or both have been provided. Observers often note that cult survivors seem brighter, more resourceful, and more talented than when they first escaped.

FUTURE DIRECTIONS

Some assume that people who enter cults do so (1) as the culmination of a genuine religious pilgrimage, (2) in search of relief from symptoms of psychopathology, (3) as an escape from a bad family situation, or (4) to find an oasis of peace or spiritual serenity. *Ideological hunger* has been described as a common psychosocial characteristic of individuals who were members or former members of a variety of religious cults. Cult recruits are commonly told that they had such a need but did not realize it until their indoctrinators revealed it to them. However, in the clinical experience of many investigators, most former cult members recall having no such ideological hunger before recruitment into the cult, nor do their parents or other informants.

Treatment of cult victims is not easy. They (and often members of their families) have been subjected to peculiar and complex stresses, sometimes for many years. Psychiatric care will of course be dictated by the nature of the illness, the premorbid personality, and the patient's remaining available strengths. However, the treating physician must know as much as possible about cults, especially the one that victimized the patient. The *Cultic Studies Journal*, published in Boston by the American Family Foundation, is a valuable source of objective information. Former cult members now working as exit counselors or reentry counselors may be extremely useful collaborators. Self-help groups composed of former members or professionally directed group therapy involving several former cultists at a time have also proved most worthwhile.

The yellow pages of the telephone directory in any American city with a population of a quarter-million or more (and in many smaller ones as well) will reveal advertisements by a variety of putative mental healers, counselors, hypnotists, spiritual health advisors and self-styled healers of every imaginable stripe. Other practitioners, while not advertised so blatantly, nevertheless attract large followings of those who seek peace of mind, relief from stress, or simply answers to worrisome questions about their health or family problems or about the meaning of life. There are more nonprofessional therapies than can be catalogued, and new ones appear all the time. The health professional must be generally aware of the ubiquity and appeal of cults, quacks, and “alternative” or nonprofessional therapies; to understand their seeming benefits—usually temporary—to some individuals (through the power of suggestion and the placebo effect) but their equally inevitable hazards and harm to others. True professionals must not shrink from warning the vulnerable about these risks and must conduct themselves so that even a confused and bemused public can clearly differentiate scientific medicine and its ethical practitioners from their myriad dubious alternatives.

SUGGESTED CROSS-REFERENCES

Alternative medicine is discussed in Section 28.8. Posttraumatic stress disorder is covered in Chapter 15 on anxiety disorders. Dissociative disorders are presented in Chapter 18. Hypnosis is covered in Section 30.3.

SECTION REFERENCES

Barrett S, Jarvis WT, editors: *The Health Robbers: A Close Look at Quackery in America*. Prometheus Books, Buffalo, NY, 1993.

Benson H: *The Relaxation Response*. Morrow, New York, 1975.

Butler K: *A Consumer's Guide to “Alternative Medicine.”* Prometheus Books, Buffalo, NY, 1992.

Cialdini R: *Influence: The New Psychology of Modern Persuasion*. Quill, New York, 1984.

*Galanter M: Cults and zealous self-help movements: A psychiatric perspective. Am J Psychiatry *145*:543, 1990.

Hassan S: *Combatting Cults' Mind Control*. Park Street Press, Rochester, VT, 1988.

Langone MD, editor: *Recovery from Cults: Help for Victims of Psychological and Spiritual Abuse*. Norton, New York, 1993.

Lawrence RM: *Primitive Psycho-therapy and Quackery*. Constable & Co., London, 1910.

Lifton RJ: *Thought Reform and the Psychology of Totalism*. Norton, New York, 1961.

Lifton RJ: Cults: Religious totalism and civil liberties. In *The Future of Immortality and Other Essays for a Nuclear Age*. Basic Books, New York, 1987.

Pankratz L, editor: Unvalidated, fringe, and fraudulent treatment of mental disorders. Int J Ment Health *19*(Suppl):1,1990.

Raso J: *“Alternative” Healthcare: A Comprehensive Guide*. Prometheus Books, Buffalo, NY, 1994.

*Singer MT, Lalich J: *Cults in Our Midst*. Jossey-Bass, San Francisco, 1995.

Singer MT, Temerlin MK, Langone MD: Psychotherapy cults. Cultic Stud J *7:*101, 1990.

Stotland NL: When religion collides with medicine. Am J Psychiatry *156:*304, 1999.

Tobias ML, Lalich J: *Captive Hearts, Captive Minds: Freedom and Recovery from Cults and Other Abusive Relationships*. Hunter House, Alameda, CA, 1992.

*West LJ: Persuasive techniques in contemporary cults: A public health approach. In *Cults and New Religious Movements*, M Galanter, editor. American Psychiatric Press, Washington, DC, 1989.

West LJ: A psychiatric overview of cult-related phenomena. J Am Acad Psychoanal *21:*1, 1993.

West LJ: A psychiatric implications of stressful methods employed by totalist cults. In *Stress in Psychiatric Disorders*, RP Liberman, J Yager, editors. Springer, New York, 1994.

West LJ, Allen JR: Three rebellions: red, black, and green. In *Progress in Psychoanalysis*, vol XIII, SJ Lynn, JW Rhue, editors. Guilford, New York, 1994.

Young JH: *American Health Quackery*. Princeton University Press, Princeton, NJ, 1992.

*Zimbardo P: What messages are behind today's cults. Am Psychol Monit *28:*14, 1997.

CHAPTER 29. PSYCHIATRIC EMERGENCIES

29.1 SUICIDE

ALEC ROY, M.D.

Suicide is a major public health problem: approximately 0.9 percent of all deaths are the result of suicide. About 1000 persons are estimated to commit suicide each day worldwide. In the United States suicide ranks as the eighth leading cause of death, and there are approximately 75 suicides per day, or one every 20 minutes, and more than 30,000 each year. The suicide rate in the United States has averaged 12.5 per 100,000 in the twentieth century, with a high of 17.4 per 100,000 during the Great Depression. From 1983 to 1998 the overall suicide rate has remained relatively stable whereas the rate for 15- to 24-year-olds has increased two to threefold. The number one suicide site in the world is the Golden Gate Bridge in San Francisco.

RISK FACTORS

Suicide is usually most accurately viewed as a multidetermined act. Risk factors for suicide include psychiatric disorder, social factors, psychological factors, biological factors, genetic factors, and physical disorder.

Psychiatric Disorders Important knowledge about why people commit suicide has been obtained from psychological autopsy studies of suicide victims in the general population. For example, the first such report studied the lives of 134 persons on whom the medical examiner of St. Louis City and County returned a suicide verdict during a 1-year period. Ninety-eight percent were clinically ill (94 percent with psychiatric disorders) and 68 percent suffered from one of two psychiatric disorders—45 percent from mood disorder and 23 percent from alcoholism (alcohol abuse or dependence). This pattern of results has been replicated many times (Table 29.1-1).

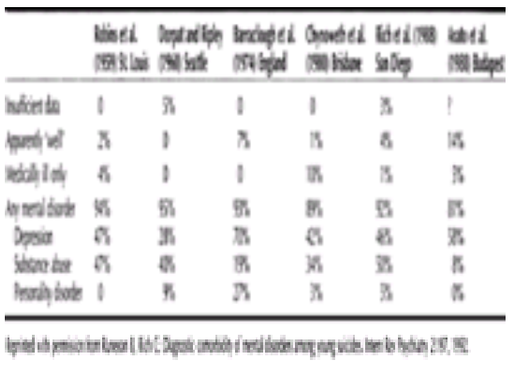

Table 29.1-1 Systematic Postmortem Interview Studies of Consecutive Suicides

There are significant differences between young and old suicide victims. Studies in San Diego and Rochester show that substance use disorders and antisocial personality disorder were found more often among suicide victims under 30 years of age whereas mood and cognitive disorders were found more often among suicide victims aged 30 and over. Precipitant stressors associated with suicide in those under 30 were separation, rejection, unemployment, and legal troubles; physical illness stressors were found more often among suicide victims over 30.

Depressive Disorders Mood disorder is the psychiatric diagnosis most commonly associated with suicide, and thus depression is one of the most important risk factors for suicide. For example, in one general population study, 45 percent of suicide victims were diagnosed as having had a primary depressive disorder, as were 30 percent in a second study and 64 percent in a third study. Follow-up studies report that about one in six depressed patients (15 percent) dies by committing suicide, although many of these studies were carried out before lithium prophylaxis was available.

The characteristics of patients with major depressive disorder who commit suicide was examined by the Finnish Suicide Research Center. Males were overrepresented as 45 of the 71 depressed suicide victims were male (63 percent) and 26 were female (37 percent). Thus, male sex is a risk factor for suicide among depressed subjects. The mean age was 50 years. The majority (85 percent) were complicated cases of depression with comorbid diagnoses, and comorbidity varied with sex and age. For example, males were overrepresented among the third of the depressed suicides with comorbid substance abuse (Table 29.1-2). Older suicide victims had more comorbid physical illness and fewer personality disorders, which were found more among younger suicide victims.

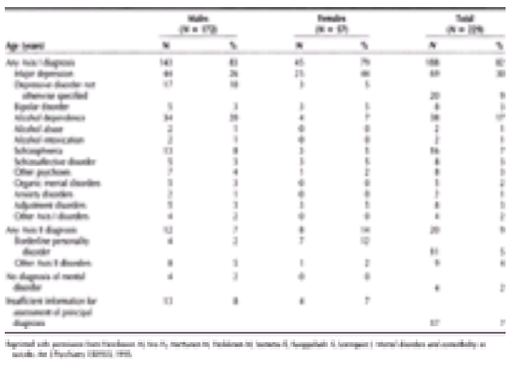

Table 29.1-2 Principal DSM-III-R Axis I and II Diagnoses of 229 Male and Female Suicide Victims

As regards treatment, 75 percent of the depressed suicide victims had a history of psychiatric treatment and two thirds had had psychiatric treatment during the last year. However, only 45 percent were receiving psychiatric treatment at the time of suicide. The mean time from last contact with health care was 39.0 ± 89.8 days. Eighteen percent had had face-to-face contact with a health care professional on the day they committed suicide, 39 percent had visited a professional within the last week, and 66 percent had done so within the last 3 months before suicide. In 60 percent of patients the reason for the visit was psychiatric and in 40 percent it was somatic. However, suicidal thoughts were communicated at only 29 percent of the last visits.

Most depressed suicide victims had received no treatment for depression. The antidepressant treatment of depression before death was either absent or inadequate. Only 3 percent had received antidepressants in adequate dosages, 7 percent had received weekly psychotherapy, and 3 percent had undergone electroconvulsive therapy (ECT) (Table 29.1-3). None of the 24 depressed patients with psychotic features had received adequate psychopharmacological treatment. Only 8 percent

had used antidepressants to commit suicide. Men had received less treatment for depression than women and more often used violent suicide methods.

	Males (*N* = 42)		Females (*N* = 24)		Total (*N* = 66)	
Antidepressant	*N*	%	*N*	%	*N*	%
None	33	79	11	46	44	67
Tricycle equivalent dose (mg)						
0–74	3	7	6	25	9	14
75–149	5	12	6	25	11	17
150–300	1	3	1	4	2	3

$x = 5.97$, $df = 1$, $p = 0.02$ (with Yate's correction)
Reprinted with permission from Isometsa E, Henriksson M, Aro H, Heikkinen M, Kuoppasalmi K, Lonnquist J: Suicide in major depression. Am J Psychiatry 151:530, 1996.

Table 29.1-3 Antidepressant Treatment Received by 66 Male and Female Suicide Victims with Major Depressive Disorder

Thus, the Finnish workers found that although about half of the depressed suicide victims were receiving psychiatric care at the time of death, few were receiving adequate treatment for depression. They concluded that for suicide prevention in major depressive disorder, great improvements in treatment and follow-up are required, particularly for males with depression.

Schizophrenia The suicide risk is high among schizophrenia patients: up to 10 percent die by committing suicide. Most persons with schizophrenia who commit suicide do so during the first few years of their illness. Thus, schizophrenic suicides tend to be relatively young and about 75 percent are unmarried males; approximately 50 percent have made a previous suicide attempt. Depressive symptoms are closely associated with their suicide; studies have reported that depressive symptoms were present during the last period of contact in up to two thirds of schizophrenia patients who committed suicide; only a small percent commit suicide because of hallucinated instructions or in order to escape persecutory delusions. Up to a third of schizophrenic suicides occur during the first few weeks and months following discharge from hospital; another third commit suicide while they are inpatients.

Alcohol Use Disorders Alcoholic persons have an increased risk of suicide, with a lifetime suicide risk of 2.2 to 3.4 percent. More men than women are found among alcoholic suicide victims. Alcoholic persons usually commit suicide after years of alcohol abuse. Comorbidity plays an important role; persons with alcoholism who have comorbid depressive disorders are at particularly high risk. Workers in St. Louis examined the relation of alcoholic suicide to specific life events. Among 31 alcoholic suicides, 48 percent had lost a loved one during the year before they committed suicide, and 32 percent had experienced such a loss during their last 6 weeks. The St. Louis group examined the lives of 50 other alcoholic suicide victims. Based on the hypothesis that suicide among alcoholics may represent a reaction to life events, the researchers recorded their opinions about the most important reason for suicide in each case (Table 29.1-4). Loss of a close relationship was the most frequently cited precipitating event; other events included job trouble, financial difficulties, and being in trouble with the law. For only 1 of the 50 suicide victims could no precipitating event be identified.

Presumed Reason	Number
Marital separation/divorce	8
Friction with spouse/lover	9
Expectation of loss (realistic)	5
Estrangement from family	2
Bereavement	2
Friction with family	1
Job trouble	4
Financial trouble	3
Trouble with the law	2
Depressed (as principal or only reason)	6
Feeling of disgrace	2
Feared rehospitalization	2
Inability to control drinking	1
Other	2
No provocation identified	1
Total	50

Reprinted with permission from Murphy G: Suicide in alcoholism. In Suicide, Roy A, editor. Williams & Wilkins, Baltimore, 1986.

Table 29.1-4 Presumed Most Important Reason for Suicide Among 50 Alcoholic Suicide Victims

Because only 25 to 40 percent of alcoholic suicide victims suffered a recent loss, other predictors of suicide were searched for. Sixty-seven alcoholic suicide victims were compared with a community sample of 106 male alcoholics on six putative suicide risk factors. These included recent heavy drinking, presence of major depressive disorder, suicidal thoughts, poor social support, living alone, and unemployment. The alcoholic suicide victims differed significantly from the living alcoholics on all six of these risk factors (Table 29.1-5). Every one of the suicide victims had at least one these risk factors; three victims had one factor, three had two factors, and eight had all six factors. By comparison, only one living alcoholic had four risk factors and none had more than four (Table 29.1-6). The researchers concluded that the risk of suicide increases as the number of risk factors increases. The presence of as few as four of these factors may identify four fifths of persons with alcohol dependence at highest risk for suicide.

	Number and Percentage of Subjects	
Risk Factor	Alcoholics Who Committed Suicide (*N* = 67)	Living Alcoholics (*N* = 106)
Recent heavy drinking	65 (97%)	44 (42%)
Talk or threat of suicide	53 (79%)	24 (23%)
Little social support	50 (75%)	28 (26%)
Major depressive disorder	39 (58%)	5 (5%)
Unemployed	36 (54%)	19 (18%)
Living alone	30 (45%)	18 (17%)

Reprinted with permission from Murphy G, Wetzel R, Robins E, McEvoy L: Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiatry 49: 459, 1992.

Table 29.1-5 Comparison of Frequency of Risk Factors in White Male Alcoholics Who Committed Suicide and in Living Alcoholic Controls

Number of Risk Factors per Subject	Number of Alcoholics Who Committed Suicide (*N* = 67)	Number of Living Alcoholics (*N* = 106)
0	0	26
1	}	31
2	}	34
3	15	14
4	19	1
5	19	0
6	8	0

The risk factors are current heavy drinking, major depression, poor or no social support, unemployment, living alone, and suicidal thoughts or communication.
Reprinted with permission from Murphy G, Wetzel R, Robins E, McEvoy L: Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiatry 49: 459, 1992.

Table 29.1-6 Distribution of Risk Factors in White Male Alcoholics Who Committed Suicide and in Living Alcoholics

OTHER SUBSTANCE USE DISORDERS DRUG DEPENDENCE There is an increased risk of suicide among substance abusers. For example, the suicide rate of heroin addicts is about 20 times greater than that of the general population. The availability of lethal amount of drugs, intravenous use, associated antisocial personality disorder, chaotic lifestyle, and impulsivity are some of the factors that predisposes drug-dependent persons to suicidal behavior, particularly when they are dysphoric, depressed, or intoxicated.

The Rochester group found that comorbidity of substance abuse and mood disorders was common among suicides. Among suicide victims with substance abuse, older age at death predicted major depressive disorder, among victims with mood disorders, younger age at death predicted comorbid substance abuse (Table 29.1-7).

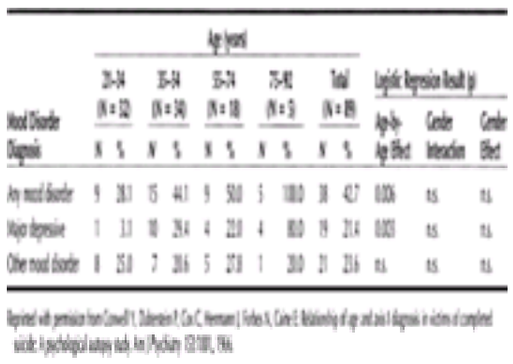

Table 29.1-7 Comorbidity of DSM-III-R Mood Disorders Among 89 Victims of Completed Suicide With Substance Use Disorders

Personality Disorder It is recognized that patients with borderline and antisocial personality disorders have an increased risk of suicide (Table 29.1-8). Recently the Finnish group reported a psychological autopsy study of suicide victims with personality disorders and found that they were almost always (95 percent) associated with current Axis I depressive disorders, substance use disorders, or both.

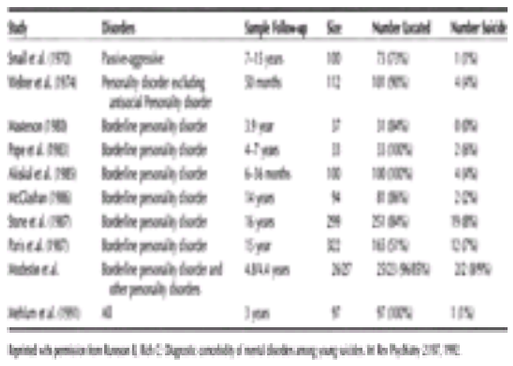

Table 29.1-8 Suicides in Follow-up Studies of Patients With Personality Disorders

The Finnish group also reported that 67 of a random sample of 229 suicide victims had an Axis II personality disorder. About one fifth (*N* = 43, 29 percent) of all the 229 suicides had a cluster B diagnosis (dramatic, emotional, or erratic), compared to the estimated prevalence of 4 to 5 percent in the general population. Ten percent of the sample (*N* = 23) had a Cluster C diagnosis (anxious or fearful), and only 1 person had a Cluster A diagnosis (odd or eccentric).

The next compared the personality-disorder suicide victims with sex- and age-matched suicide victims without personality disorder. Suicides with Cluster B personality disorders were more likely than comparison subjects to have substance use disorders (79 percent versus 40 percent) and previous nonfatal suicide attempts (70 percent versus 37 percent) and were less likely to have Axis III physical disorders (29 percent versus 50 percent). Suicide victims in Cluster B had almost always (98 percent) either comorbid depressive disorders (74 percent), substance use disorders (79 percent), or both (55 percent). They found no evidence of impulsive suicides that would have occurred without Axis I disorders. In contrast, subjects with Cluster C personality disorders did not differ from their controls on any variable.

Panic Disorder Data from the National Institute of Mental Health (NIMH) Epidemiologic Catchment Area (ECA) Study showed that 20 percent of individuals with panic disorder had made a suicide attempt at some time. This high rate was similar to the rate for individuals with major depression. When patients with panic disorder without comorbidity were examined, the lifetime rate of suicide attempts remained raised at 7 percent. Similarly, individuals with social phobia have increased rates of suicidal ideation and suicide attempts. The Finnish group found that current panic disorder was rare among completed suicides, being found in only 1.2 percent of all suicides. However, panic disorder suicide victims had superimposed major depression and substance abuse, and associated personality disorders. Thus, clinical assessment of suicide risk in panic and phobia patients should include determination of the presence or absence of major depressive disorder, substance abuse, and personality disorder.

Comorbidity Comorbidity is common among suicide victims. For example, the Finnish group reported that while 93 percent of a random sample of 229 suicide victims had an Axis I psychiatric disorder, two or more diagnoses on Axis I were present for 44 percent of the suicide victims. The most prevalent disorders were depressive disorders (59 percent) and alcohol dependence or abuse (43 percent). A diagnosis on Axis II was made for 31 percent and at least one diagnosis on Axis III was made for 46 percent of the suicides. Only 12 percent of the suicide victims received only one Axis I diagnosis without any comorbidity.

Social Factors Emile Durkheim was the first to examine how social and cultural influences impact on the risk of suicide. He found that suicide rates in European countries differed in relation to various demographic and social factors, and concluded that “the suicide rate varies inversely with the integration of social groups of which the individual forms a part.” Egoistic suicide is determined by a lack of meaningful family ties or social interactions. *Anomic* suicide occurs when the relationship between an individual and society is broken by social or economic adversity. For example, suicide rates rose during the Great Depression of the 1930s and fell in nearly all European countries during both world wars. Altruistic suicide results from excessive integration in society (e.g., harakiri, sati).

National Suicide Rates These vary from below 10 per 100,000 (Ireland, Egypt) to over 35 per 100,000 (Baltic states—the highest). Variations between countries reflect cultural and social factors as well as differences in the methods used to report suicide. In general the suicide rate is low in less prosperous countries and higher in more affluent countries; it is high in all Eastern European nations. The United States has a rate in the middle range, around 12 per 100,000.

Age Suicide rates increase with age. Among men, suicides peak after age 45; among women, the greatest number of completed suicides occurs after age 55. Rates of 40 per 100,000 population are found in men aged 65 and older. The elderly attempt suicide less often than do younger people but are successful more often. The elderly account for 25 percent of the suicides, although they make up only 10 percent of the total population. The rate of suicide for those 75 or older is more than three times the rate among the young.

Sex Males at all ages commit suicide more often than females and male-to-female suicide ratios range from 2 to 1 to 7 to 1. Also, males use more violent methods, like hanging, shooting, and jumping; females more often overdose or drown.

Race Ethnic and minority groups tend to be more cohesive and have lower suicide rates. In the United States, blacks have lower suicide rates than whites. The rate of suicide among whites is nearly twice that among nonwhites, but the figures are being questioned because the suicide rate among blacks is increasing. In 1989 the suicide rate for white males (19.6 per 100,000 persons) was 1.6 times that for black males (12.5), 4 times that for white females (4.8), and 8.2 times that for black females (2.4). Among inner-city youth and certain Native American and Inuit groups, suicide rates have greatly exceeded the national rate. Suicide among immigrants is higher than in the native-born population. Immigrants, who are likely to feel alienated, have higher rates than natives of either their adopted country or their country of origin.

Marital Status Married persons have the lowest suicide rates. The suicide rate for single persons is twice that of married persons and that of the divorced, separated, or widowed, four to five times higher (Table 29.1-9). Neighborhood and families ties and religious affiliations affect the risk of suicide. Suicide rates are lower among Jews and Catholics than among Protestants.

	Suicide rate per 100,000	
Marital status	Males	Females
Single	29.44	6.42
Married	18.94	4.68
Divorced	72.61	16.47
Widowed	92.13	8.85

Reprinted with permission from Tsuang M, Simpson L, Fleming J: Epidemiology of suicide. Int Rev Psychiatry 92:117, 1992.

Table 29.1-9 Suicide Rates by Marital Status for Males and Females in 1988

Employment Work, in general, protects against suicide. The unemployed have higher suicide rates, probably owing to an interaction of socioeconomic circumstances, psychological vulnerability, and stressful life events. During economic recessions and depressions and times of high unemployment, the suicide rate increases; during times of high employment and during war, the rate decreases.

Psychological Factors

Freud The first important psychological insight into suicide came from Sigmund Freud. In his 1917 paper “Mourning and Melancholia,” Freud stated that suicide represented aggression turned inward against an introjected, ambivalently cathected love object. Freud doubted that there would be a suicide without the earlier repressed desire to kill someone else.

Menninger Building on Freud's concepts, Karl Menninger in *Man Against Himself* conceived of suicide as a retroflexed murder, an inverted homicide, as a result of the patient's anger toward another person, which is either turned inward or used as an excuse for punishment. He also described a self-directed death instinct (Sigmund Freud's concept of *thanatos*). He described three components of hostility in suicide: the wish to kill, the wish to be killed, and the wish to die.

Recent Theories Contemporary suicidologists are not persuaded that there is a specific psychodynamic or personality structure associated with suicide. However, they have suggested that much can be learned about the psychodynamics of suicidal patients from their fantasies as to what would happen and what the consequences would be if they were to commit suicide. It is thought that the suicidal patients who are most likely to act out suicidal fantasies are those who have suffered the loss of a love object or have sustained a narcissistic injury, who experience overwhelming moods like rage and guilt, or who identify with a suicide victim. Aaron Beck has demonstrated the importance of hopelessness while Edward Shneidman has observed cognitive constriction of choices among the suicidal. Group dynamics underlie mass suicides like those at Masada and Jonestown.

Biological Factors Diminished central serotonin plays a role in suicidal behavior. A group at the Karolinska Institute were the first to note that low concentrations of the serotonin metabolite 5-hydroxy-indoleacetic acid (5-HIAA) in the lumbar cerebrospinal fluid (CSF) were associated with suicidal behavior. This finding has been replicated many times and in different diagnostic groups. Postmortem neurochemical studies have reported modest decreases in serotonin itself, or 5-HIAA, in either the brain stem or frontal cortex of suicide victims. Postmortem receptor studies have reported significant changes in presynaptic and postsynaptic serotonin binding sites in suicide victims. Taken together, these CSF, neurochemical, and receptor studies support the hypothesis that reduced central serotonin is associated with suicide. Recent studies also report some changes in the noradrenergic system of suicide victims.

Low concentrations of CSF 5-HIAA also predicts future suicidal behavior. For example, the Karolinska group examined completed suicide in a sample of 92 depressed patients who had attempted suicide. They found that 8 of the 11 patients who committed suicide within 1 year belonged to the subgroup with below-the-median concentrations of 5-HIAA in CSF. The suicide risk in that subgroup was 17 percent as compared with 7 percent among those with above-the-median concentrations of 5-HIAA in CSF (Fig. 29.1-1). Also, the cumulative number of patient-months survived during the first year after attempted suicide was significantly lower in the subgroup with low 5-HIAA concentrations. The Karolinska group concluded that low CSF 5-HIAA concentrations predict short-range suicide risk in the high-risk group of depressed patients who have attempted suicide.

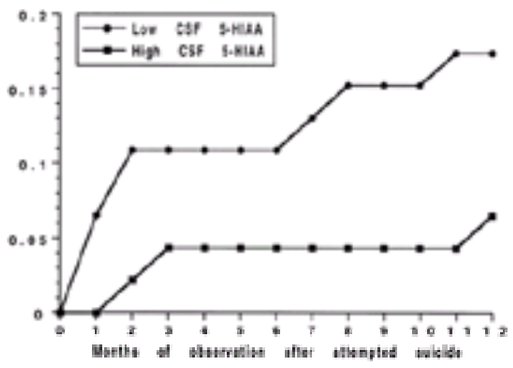

FIGURE 29.1-1 Cumulative suicide risk during first year after attempted suicide in patients with low versus high CSF concentrations of 5-HIAA. Filled circles indicate CSF 5-HIAA concentrations below the sample median and filled squares indicate concentrations above the sample median (87nM). (Reprinted with permission from Nordstrom P, Samuelsson M, Asberg M, Traskman-Bendz L, Aberg-Wistedt A, Nordin C, Bertilsson L: CSF concentrations 5-HIAA predicts suicide risk after attempted suicide. Suicide Life Threat Behav *24:*1, 1994.)

Genetic Factors Suicidal behavior, like other psychiatric disorders, tends to run in families. For example, Margaux Hemingway's 1997 suicide was the fifth suicide among four generations of Ernest Hemingway's family. In psychiatric patients a family history of suicide increases the risk both of attempted suicide and of completed suicide in most diagnostic groups. In medicine the strongest evidence for the possibility of genetic factors comes from twin and adoption studies and from molecular

genetics. Such studies in suicide are reviewed below.

Twin Studies In 1991, 176 twin pairs in which one twin had committed suicide were investigated. In nine of these twin pairs, both twins had committed suicide. Seven of these nine pairs concordant for suicide were found among the 62 monozygotic pairs, while two pairs concordant for suicide were found among the 114 dizygotic twin pairs. This twin group difference for concordance for suicide (11.3 percent versus 1.8 percent) is statistically significant ($P < .01$).

In another study a group of 35 twin pairs of which one twin had committed suicide was collected and the living co-twin was interviewed. It was found that 10 of the 26 living monozygotic co-twins had themselves attempted suicide, compared with 0 of the 9 living dizygotic co-twins ($P < 0.04$). Although monozygotic and dizygotic twins may have some differing developmental experiences, these results show that monozygotic twin pairs have significantly greater concordance for both suicide and attempted suicide, which suggests that genetic factors may play a role in suicidal behavior.

Danish-American Adoption Studies The strongest evidence suggesting the presence of genetic factors in suicide comes from the adoption studies carried out in Denmark. A screening of the registers of causes of death revealed that 57 of 5483 adoptees in Copenhagen eventually committed suicide. They were matched with adopted controls. Searches of the causes of death revealed that 12 of the 269 biological relatives of these 57 adopted suicide victims had themselves committed suicide compared with only 2 of the 269 biological relatives of the 57 adopted controls. This is a highly significant difference for suicide between the two groups of relatives. None of the adopting relatives of either the suicide or control group had committed suicide.

In a further study of 71 adoptees with mood disorder, adoptee suicide victims with a situational crisis or impulsive suicide attempt or both particularly had more biological relatives who had committed suicide than controls. This led to the suggestion that a genetic factor lowering the threshold for suicidal behavior may lead to an inability to control impulsive behavior. Psychiatric disorder or environmental stress may serve "as potentiating mechanisms which foster or trigger the impulsive behavior, directing it toward a suicidal outcome."

Molecular Genetic Studies Tryptophan hydroxylase (TPH) is an enzyme involved in the biosynthesis of serotonin. A polymorphism in the human *TPH* gene with two alleles—U and L—has been identified. As low concentrations of 5-HIAA in CSF are associated with suicidal behavior, it was hypothesized that such individuals may have alterations in genes controlling serotonin synthesis and metabolism. It was found that impulsive alcoholics, who had low CSF 5-HIAA concentrations, had more LL and UL genotypes. Furthermore, a history of suicide attempts was significantly associated with *TPH* genotype in all the violent alcoholics (Fig. 29.1-2); 34 of the 36 violent subjects who attempted suicide had either the UL or LL genotype. Thus, it was concluded that the presence of the L allele was associated with an increased risk of suicide attempts. Also, a history of multiple suicide attempts was found most in subjects with the LL genotype and to a lesser extent among those with the UL genotype (Fig. 29.1-3). This led to the suggestion that the L allele was associated with repetitive suicidal behavior. The presence of one TPH*L allele may indicate a reduced capacity to hydroxylate tryptophan to 5-hydroxytryptophan in the synthesis of serotonin, producing low central serotonin turnover and thus low concentrations of 5-HIAA in CSF.

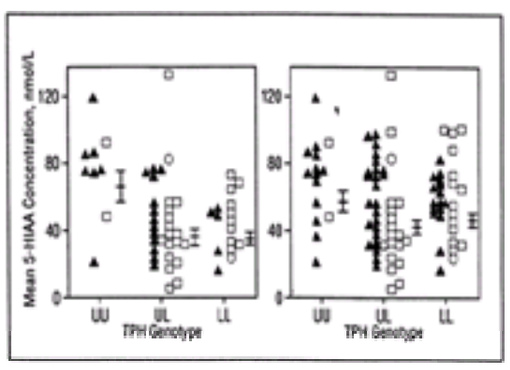

FIGURE 29.1-2 Relationship between tryptophan hydroxylase *(TPH)* genotype, mean (±SEM) 5-hydroxyindoleacetic acid (5-HIAA) concentration, and history of suicide attempts. The 5-HIAA concentrations of the subjects are plotted against their *TPH* geotypes for impulsive subjects (left) and all subjects (right). Triangles represent subjects who had never attempted suicide; squares, suicide attempters; and circles, those who had committed suicide. (Reprinted with permission from Nielsen D, Goldman D, Virkkunen M, Tokola R, Rawlings R, Linnoila M: Suicidality and 5-hydroxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry *51:*34, 1994.)

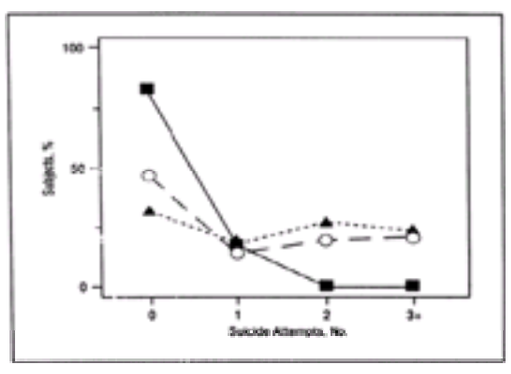

FIGURE 29.1-3 Relation between *TPH* genotype and lifetime history of multiple suicide attempts. For each genotype, the fraction of subjects having each genotype (UU, squares; UL, circles; LL, triangles) is plotted against the number of suicide attempts they have made in their lives. (Reprinted with permission from Nielsen D, Goldman D, Virkkunen M, Tokola R, Rawlings R, Linnoila M: Suicidality and 5-hydroxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry *51:*34, 1994.)

Physical Disorder Postmortem studies show that a physical illness is present in 25 to 75 percent of all suicide victims; a physical illness is estimated to be an important contributing factor in 11 to 51 percent of suicides. In each instance, the percentage increases with age.

Fifty percent of men with cancer who commit suicide do so within a year of receiving the diagnosis. Cancer of the breast or genitals is found in 70 percent of women with cancer who commit suicide. There are seven diseases of the central nervous system that increase the risk of suicide: epilepsy, multiple sclerosis, head injury, cardiovascular disease, Huntington's disease, dementia, and acquired immune deficiency syndrome (AIDS). All are disease in which an associated mood disorder is known to occur. Persons with epilepsy also have available barbiturates and other medications with which to kill themselves.

Four endocrine conditions are associated with increased suicide risk: Cushing's disease, anorexia nervosa, Klinefelter's syndrome, and porphyria. Mood disorders also attend these disorders. The two gastrointestinal disorders with an increased suicide risk are peptic ulcer and cirrhosis, both physical disorders found among alcoholic persons. The two urogenital problems with an increased suicide risk are prostatic hypertrophy and hemodialysis, both problems in which changes in mood occur. Because of the increased incidence of major depressive disorder in patients with cerebral disease, the association of suicide with cerebrovascular disorders and progressive organic deterioration leading to dementia is hardly surprising.

SUICIDE IN THE ELDERLY

Compared to other age groups, those 65 years and older have the highest risk of committing suicide. For example, the suicide rate for the elderly is 50 percent higher than the rate for teenagers or the U.S. national rate. In the United States 18 elderly persons commit suicide each day, 1 every 80 minutes. The majority of elderly suicides are committed using a firearm. Some older individuals have a higher suicide risk than others: most at risk are males, whites, the recently widowed, and those

aged 75 years or more.

The two psychiatric conditions most associated with suicide in the elderly are depression and alcoholism. Psychological autopsy studies show that approximately 70 percent of elderly suicide victims suffered from depression in the weeks and months before their suicide. Approximately 20 percent of elderly suicides meet the criteria for a substance abuse disorder, usually alcohol abuse or dependence.

Studies show that loss and stress are major etiologic factors in the depression and alcohol abuse found among elderly suicide victims. These include physical losses resulting from poor health, painful illness, sensory deficits, and cognitive decline; social losses like death of a spouse and loss of the work role; and income losses associated with retirement and medical expenses. Such losses may lead to reduced social networks and social isolation, and can produce feelings of despair, loneliness, demoralization, dependency on others, helplessness and hopelessness, as well as suicidal ideation (Table 29.1-10).

Vital Loses	Stress	Despair	Alcoholism
Physical	More vunerable	Helplessness	Helplessness
Social	Less resistant	Hopelessness	Hopelessness
Financial	Failure in coping	Anxiety	Anxiety
Personal	Mechanisms	Decreased self-concept	Decreased self-concept
Emotional	Self-concept	Lowered self-esteem	
Cognitive	Lowered self-esteem	Loneliness	
	Loneliness		

Reprinted with permission from Osgood N: Suicide in the elderly: Etiology and assessment. In the Psychiatry. 1237, 1992.

Table 29.1-10 A Model of the Aging Process and Suicide

SUICIDE AMONG PSYCHIATRIC INPATIENTS

The great majority of psychiatric patients who commit suicide do so as outpatients. Many of these outpatients suicides occur within the first few weeks or months after discharge from a psychiatric institution, which highlights the fact that the postdischarge period is a period of increased suicide risk for psychiatric patients.

However, up to 15 percent of psychiatric patients who commit suicide do so as inpatients. Many of these suicides do not actually occur on the ward itself but in stairwells, other hospital buildings, on the grounds, near railway lines, or in rivers. Other inpatients commit suicide at home when out on a pass, on weekend leave, or when absent without leave.

The commonest diagnosis among inpatient suicides in the acute psychiatric ward is depression and the suicide risk is highest in the first few days of the admission. Times of staff rotation, particularly of psychiatric residents, are periods associated with inpatient suicides. Clusters of inpatient suicides tend to be associated with periods of ideological change on a ward, staff disorganization, or staff demoralization.

However, suicides also occur among long-term psychiatric inpatients. In hospitals the diagnosis most often associated with suicide is schizophrenia. Surprisingly, inpatients with schizophrenia may commit suicide after having been in the hospital for more than a year. Often, they have previously exhibited suicidal behavior and have been admitted involuntarily (Table 29.1-11). Suicide in such schizophrenia patients may be precipitated by attempts at rehabilitation or discharge.

Patient Characteristics	Suicide Victims N = 37	Controls N = 37	Significance
Past suicide attempt	23	13	P<0.02
Involuntary last admission	26	16	P<0.02
Living alone	26	16	P<0.02
Single	21	20	NS
Separated/divorced/widowed	11	9	NS
Married	5	8	NS
Unemployed	34	30	NS
Alcohol and/or drug abuse	16	15	NS
Received antidepressants	16	13	NS
Received lithium	6	6	NS
Received ECT	13	9	NS

Reprinted with permission from Roy A, Draper R: Suicide among psychiatric hospital inpatients. Psychol Med 29:1199, 1999.

Table 29.1-11 Comparison of Inpatient Suicide Victims and Inpatient Controls

MANAGEMENT OF SUICIDAL PATIENTS

Many suicides among psychiatric patients are preventable. There are occasional patients whose suffering is so great and intense or so chronic and unresponsive to treatment that their eventual suicide may be perceived as inevitable; fortunately, such patients are relatively uncommon. Some have severe personality disorders or are highly impulsive; such patients commit suicide when dysphoric, intoxicated, or both.

Risk factors are useful in assessing the short-term risk of suicide of a patient in an acute crisis. The psychiatric and social risk factors for suicide include a past attempt, a chronic psychiatric disorder, recent hospital discharge, living alone, being unemployed, being unmarried, and being liable to develop depressive symptoms. The immediate postdischarge period is a time of increased risk. However, these risk factors identify too many false-positive and false-negative patients to be useful in the long-range prediction of suicide.

The evaluation for suicide potential involves a complete psychiatric history, a thorough examination of the mental state, and inquiry about depressive symptoms, suicidal thoughts, intents, plans, and attempts. Lack of future plans, giving away personal property, making a will, or having recently experienced a loss imply increased risk for suicide. The decision to admit the patient to a psychiatric facility depends on the diagnosis, the severity of depression and suicidal ideation, the patient's and family's ability to cope, the living situation, the availability of social support, and the absence or presence of risk factors for suicide. Precautions against suicide include removing potentially lethal drugs and firearms from the home and prescribing medication only in weekly supplies.

Useful measures for managing a depressed suicidal inpatient include searching the patient's belongings and person on arrival on the unit for objects that might be used for suicide, and repeating the search at times of exacerbation of suicidal ideation. Ideally, the suicidal depressed inpatient should be managed on a locked unit with shatterproof windows, and the patient's room should be located near the nursing station to maximize observation by the nursing staff. The treating team will have to assess how much to restrict the patient and whether there should be regular checks or continued direct observation. Vigorous treatment with antidepressant medication should be initiated, and if there is no response within a reasonable period of time, or if the suicide risk is too great, a course of ECT should be considered.

LEGAL AND ETHICAL CONSIDERATIONS

Liability issues stemming from suicides in psychiatric hospitals frequently involve questions about the patient's rate of deterioration, the presence during hospitalization of clinical signs indicating risk, and the psychiatrist's and the staff members' awareness of and response to those clinical signs. What the courts require is not that suicide never occur but that the patient be periodically evaluated for suicidal risk, that a treatment plan with a high level of security be formulated, and that

the staff members follow that treatment plan. The role of an aider and abettor in suicide adds another dimension to the legal morass.

ATTEMPTED SUICIDE

Persons who attempt suicide also pose a major health problem. In the United Kingdom, there are 100,000 hospital admissions each year because of deliberate self-poisoning (2.5 percent of all admissions). In England and Wales, self-poisoning is the most common diagnosis of admitted patients under 50 years of age; among women it is the most common reason for emergency medical admission whereas among men it is second only to heart attacks.

Attempted suicide is in many ways an unsatisfactory term. For example, most attempters do not actually wish to commit suicide; their motives are different. Thus, the Edinburgh group introduced the term *parasuicide* in an effort to signify that suicide attempts are not just failed suicides but are a very different behavior.

Rates and Demographics The ECA Study found that 21 percent of adults reported that there had been a period of 2 weeks or more at some time during their lives when they thought about their own or another's death; 7.1 percent said they felt so low they wanted to die; 10.2 percent said that they had thought about committing suicide; and 2.9 percent reported that they had attempted suicide at some time in their lives.

A review of the world literature showed that attempted suicide rates varied between 100 and 300 per 100,000; that there was a preponderance of females in all countries; that about 50 percent of attempters were under 30 years of age; there was an excess of divorced persons; the lower social classes were overrepresented; and that a diagnosis of depression was made in between 35 to 79 percent of cases. Females aged 15 to 19 years have the highest rate of suicide attempts. Up to 1 in 100 in this group may be expected to attempt suicide each year. The highest rate for males is found among those aged 25 to 29, of whom 1 in 200 attempts suicide each year.

Acute Problems The circumstances surrounding a suicide attempt invariably involve recent life change, particularly interpersonal stress. Suicide attempters report five particular events significantly more frequently than controls: serious arguments with the spouse, having a new person in the home, serious illness of a family member, serious personal physical illness, and having to appear as a defendant in court. When event categories are examined, suicide attempters have more undesirable exit-and-entrance events (e.g., birth, marriage) and interpersonal arguments than controls.

Chronic Problems Persons who attempt suicide also tend to have chronic problems with marriage, children, work, finances, health, or alcohol. The Oxford group found that two thirds of married attempters had chronic marital problems and one half had been involved in a recent extramarital affair. One fourth of patients had problems with their children; parents who neglect or abuse their children are particularly likely to attempt suicide. Up to one third of attempters have had a medical or surgical admission in the previous year. Persons with epilepsy are overrepresented among attempters.

Among adolescents who attempt suicide, three quarters report difficulties with one or both parents, one half report problems with their school work or teachers, and one half report difficulties with a boyfriend or girlfriend; others have work problems or experience social isolation.

Psychiatric Disorders Hospital studies show that about 40 percent of those who attempt suicide have a history of psychiatric treatment. Psychiatric assessments reveal that about 50 percent have a personality disorder, and up to 40 percent have other psychiatric disorders. The most common diagnoses that are not personality disorders are depressive disorders (up to 40 percent of women and 30 percent of men).

Studies using observer-rated instruments find that although symptoms like tension, depressed mood, hopelessness, irritability, worrying, and poor concentration are present in 40 to 75 percent of attempters, a definite psychiatric disorder can be diagnosed at admission in only about 30 percent of attempters; this percentage decreases rapidly over subsequent weeks. However, workers in Christchurch, New Zealand, have shown that among serious suicide attempters 90.1 percent had a psychiatric disorder with high rates of mood disorder, substance abuse, and antisocial personality disorder. The incidence of comorbidity was high: 56 percent had two or more disorders. The risk of a serious suicide attempt increased with increasing psychiatric comorbidity.

Motivations, Explanations, and Consequences The majority of suicide attempts are impulsive. Two thirds of such persons report that they had thought about the act for less than an hour beforehand. There is often a discrepancy between the motivations for the attempt as perceived by the medical staff and the explanations given by the patient. In one study, over one third of patients stated that they wanted to die, but their psychiatrists agreed with this in only 50 percent of cases. The most common reason, given by up to 40 percent of both patients and their psychiatrists, was that the suicide attempt was made to get relief from a terrible state of mind. The great difference between the reports of patients and their psychiatrists was in identifying the reason for the attempt as being to frighten, get back at, or make someone feel sorry for the way they had been treated. This hostile or manipulative reason was chosen by about 10 percent of patients but by 70 percent of their psychiatrists.

Erwin Stengel recognized the appeal function of attempted suicide, which he saw as a cry for help. Although there is no specific personality trait, attempters are often immature, egocentric, anxious, dependent, hostile, impulsive, and have difficulties in relationships. Follow-up studies show that in up to three quarters of cases the suicide attempt is followed by improvements in both social problems and psychiatric symptoms. This improvement is largely because responses occur and may partly determine whether there will be another attempt.

Risk of Repetition About 40 percent of attempters have made a previous attempt. Follow-up studies show that between 13 and 35 percent will repeat the attempt during the next 2 years. During this time up to 7 percent will make two or more attempts, 2.5 percent three or more attempts, and 1 percent, five or more attempts. Thus, there appear to be three subgroups of repeaters: the very occasional repeater, the person who repeats several times within a short period, and the chronic on habitual repeater.

There are seven main items that may be helpful in identifying the patient at risk of making another suicide attempt: problems with alcohol, antisocial personality disorder, impulsivity, previous inpatient psychiatric treatment, previous outpatient psychiatric treatment, previous attempt that led to admission, and living alone.

Relationship to Suicide It is recognized that those who attempt and those who commit suicide represent different populations with some overlap. Approximately 1 percent of persons who attempt suicide will commit suicide during the following year. For 8 to 50 suicide attempters, 1 will eventually commit suicide. The risk of subsequent suicide varies with sex and age. For example, at the Karolinska Hospital country in Stockholm the suicide risk over the next 5 years after attempting suicide among men (8.3 percent) was nearly twice the suicide risk among women (Fig. 29.1-4). Both older and younger male suicide attempters are at high risk of suicide (7 percent and 10 percent, respectively), and older women are at higher risk than younger women (6 percent versus 2 percent) (Table 29.1-12). The suicide risk was particularly high during the first year after the suicide attempt.

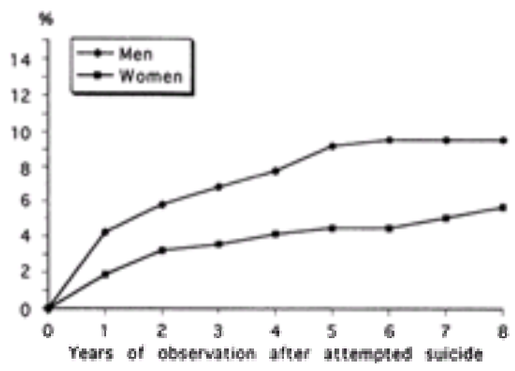

FIGURE 29.1-4 Cumulative suicide risk after attempted suicide in male and female patients referred to the psychiatric emergency room. Filled circles indicate men and filled squares indicate women. (Reprinted with permission from Nordstrom P, Samuelsson M, Asberg M, Traskman-Berdz L, Aberg-Wistedt A, Nordin C, Bertilsson L: Survival analysis of suicide risk after attempted suicide. Acta Psych Scand *91:*336, 1995.)

	Women		Men	
Age	<35 years	>35 years	<35 years	>35 years
Suicide	12	31	26	21
Not suicide	471	490	241	281
Suicide risk	2.5%	6%	9.7%	7%

Reprinted with permission from Nordstrom P, Asberg M: Survival analysis of suicide risk after attempted suicide. Acta Psych Scand 91:336, 1995.

Table 29.1-12 Male and Female Suicide Risk by Age After Attempted Suicide

Follow-up studies show that other factors associated with subsequent suicide include being unemployed or retired; being separated, divorced, or widowed; living along; having poor physical health; having received medical treatment within the last 6 months; having a psychiatric disorder, including alcoholism; having made many previous attempts by violent methods; the presence of a suicide note; and a history of previous attempts. A subgroup of suicide attempters who have severe personality disorder and interpersonal conflicts, and who often had alcohol or other substance dependence, commit suicide while acutely depressed.

SUGGESTED CROSS-REFERENCES

Mood disorders are discussed in Chapter 14. Anxiety disorders are the subject of Chapter 15. Personality disorders are the focus of Chapter 24. Chapter 11 discusses substance-related disorders; of particular interest is Section 11.2, alcohol-related disorders. Psychotherapies are the subject of Chapter 30. Section 28.5 covers death, dying, bereavement, and physician-assisted suicide.

SECTION REFERENCES

*Amir M, Kaplan Z, Efroni R, Kotler M: Suicide risk and coping styles in posttraumatic stress disorder patients. Psychother Psychosom *68:*76, 1999.

*Angst J, Angst F, Stassen HH: Suicide risk in patients with major depressive disorder. J Clin Psychiatry *60:*57, 1999.

Beautrais A, Joyce R, Mulder R, Fergusson D, Deavoll B, Nightingdale S: Prevalence and comorbidity of mental disorders in persons making serious suicide attempts. Am J Psychiatry *153:*1009, 1966.

Blumenthal SK, Kupfer D, editors: *Suicide Over the Life Cycle: Risk Factors, Assessment and Treatment of Suicidal Patients*. American Psychiatric Press, Washington, DC, 1990.

Brent D, Bridge F, Johnson B, Connolly J: Suicidal behavior run in families. A controlled family study of adolescent suicide victims. Arch Gen Psychiatry *53:*1145, 1996.

Gould M, Fisher P, Parides M, Flory M, Shaffer D: Psychosocial risk factors of child and adolescent completed suicide. Arch Gen Psychiatry *53:*1155, 1996.

Conwell Y, Duberstein P, Cox C, Hermann J, Forbes N, Caine E: Relationship of age and axis I diagnoses in victims of completed suicide: A psychological autopsy study. Am J Psychiatry *153:*1001, 1996.

Hawton KC, editor: *Attempted Suicide*. Oxford University Press, Oxford, 1982.

Henriksson M, Aro H, Marttunen M, Heikkinen M, Isometsa E, Kuoppasalmi K, Lonnquist J: Mood disorders and comorbidity in suicide. Am J Psychiatry *150:*935, 1993.

Henriksson M, Isometsa E, Kuoppasalmi K, Heikkinen M, Marttunen M, Lonnquist J: Panic disorder in completed suicide. J Clin Psychiatry *57:*275, 1996.

Isometsa E, Henriksson M, Aro H, Heikkinen M, Kuoppasalmi K, Lonnquist J: Suicide in major depression. Am J Psychiatry *151:*530, 1994.

Isometsa E, Henriksson M, Heikkinen M, Aro H, Marttunen M, Kuoppasalmi K, Lonnquist J: Suicide among subjects with personality disorder. Am J Psychiatry *153:*667, 1996.

Kreitman NE: *Parasuicide*. Wiley, New York, 1977.

Mann J, Stoff D, editor: *Psychobiology of Suicidal Behavior*. Ann N Y Acad Sci, New York, 1997.

Mann J, Malone K, Sweeney J, Brown R, Linnoila M, Stanley B, Stanley M: Attempted suicide characteristics and cerebrospinal fluid amine metabolites in depressed inpatients. Neuropsychopharmacology *15:*576, 1996.

*Muller-Oerlinghausen B, Berghofer A: Antidepressants and suicidal risk. J Clin Psychiatry *60:*94, 1999.

Murphy G, Armstrong J, Hermele S, Fisher J, Clendenin W: Suicide and alcoholism. Arch Gen Psychiatry *36:*65, 1979.

Murphy G, Wetzel R, Robins E, McEvoy L: Multiple risk factors predict suicide in alcoholism. Arch Gen Psychiatry *49:*459, 1992.

Nielsen D, Goldman D, Virkkunen M, Tokola R, Rawlings R, Linnoila M: Suicidality and 5-hydroxyindoleacetic acid concentration associated with a tryptophan hydroxylase polymorphism. Arch Gen Psychiatry *51:*34, 1994.

*Nilsson A: Lithium therapy and suicide risk. J Clin Psychiatry *60:*85, 1999.

Nordstrom P, Samuelsson M, Asberg M, Traskman-Bendz L, Aberg-Wistedt A, Nordin C, Bertilsson L: CSF 5-HIAA predicts suicide risk after attempted suicide. Suicide Life Threat Behav *24:*1, 1994.

Osgood N: Suicide in the elderly: Etiology and assessment. Intern Rev Psychiatry *4:*217, 1992.

Pfeffer C, editor: *Suicide Among Youth: Perspectives on Risk and Prevention*. American Psychiatric Press, Washington, DC, 1989.

Rich C, Young D, Fowler R: San Diego Suicide Study, I: Young vs old subjects. Arch Gen Psychiatry *43:*577, 1986.

Robins E: *The Final Months: A Study of the Lives of 134 Persons who Committed Suicide*. Oxford University Press, New York, 1981.

Roy A: Family history of suicide. Arch Gen Psychiatry *40:*971, 1983.

Roy A: Risk factors for suicide in psychiatric patients. Arch Gen Psychiatry *39:*1089, 1982.

Roy A: Self-destructive behavior. Psychiatr Clin North Am *8:*215, 1985.

Roy A, editor: Suicide. Int Rev Psychiatry *4:*1, 1992.

Roy A: Suicide in chronic schizophrenia. Br J Psychiatry *141:*171, 1982.

Roy A, editor: *Suicide*. Williams & Wilkins, Baltimore, 1986.

Roy A, DeJong J, Linnoila M: Cerebrospinal fluid monoamine metabolites and suicidal behavior in depressed patients: A five-year follow-up study. Arch Gen Psychiatry *46:*609, 1989.

Roy A, Linnoila M: Alcoholism and suicide. Suicide Life Threat Behav *16:*244, 1986.

*Simpson SG, Jamison KR: The risk of suicide in patients with bipolar disorders. J Clin Psychiatry *60:*53, 1999.

Suominen K, Henriksson M, Suokas J, Isometsa E, Ostama A, Lonnquist J: Mental disorders and comorbidity in attempted suicide. Acta Psych Scand *94:*234, 1996.

*Urbaitis JC: Psychiatric treatment helps survivors of suicide attempts. Crit Care Med *27:*458, 1999.

CHAPTER 29. PSYCHIATRIC EMERGENCIES

29.2 OTHER PSYCHIATRIC EMERGENCIES

BEVERLY J. FAUMAN, M.D.

DEFINITION

A psychiatric emergency is any unusual behavior, mood, or thought, which if not rapidly attended to may result in harm to a patient or others. Psychiatric emergencies arise in the context of chronic psychiatric illness; however, they may also occur as the consequence of medical illness that presents with psychiatric symptoms, or as a result of an adverse drug reaction or intoxication or a drug-drug interaction. Other psychiatric crises may occur when a patient is the victim of severe physical or emotional trauma, and is unable to respond adequately without professional intervention. The social milieu contributes to the definition of a psychiatric crisis by its limitations in tolerating disruptive behavior. A psychiatric emergency may occur in a home, on the street, in an outpatient department, on a psychiatric unit, on a medical or surgical unit of a general hospital, or in the emergency department.

HISTORY

Extensive literature describing psychiatric emergencies and their treatment first appears around the turn of the twentieth century, when psychiatric disturbances occurred during wars and it was necessary to treat soldiers rapidly and effectively in order to maintain troop strength. Battle fatigue, shell shock, or more recently traumatic stress reaction were observed to respond to treatment more quickly when applied as soon as possible after the onset of symptoms, rather than upon evacuation from the front and during extended bed rest. By World War II the United States Armed Forces adopted this approach, significantly decreasing the severe and prolonged psychiatric disturbances that plagued soldiers previously, affecting their quality of life and resulting in multiple and protracted psychiatric hospitalizations. Emergency treatment and the expectation that battle weary soldiers would recover quickly and be able to return to fight improved the response rate fourfold.

It was during the same time period that Erich Lindemann conducted his classic study on the survivors of a tragic nightclub fire in Boston. One significant finding was that the survivors who did well were able to grieve for those who died, and for the fright and physical injuries they had incurred, and that therapy directed towards encouraging grief increased the likelihood of recovery. In addition, patients who seemed to grieve effectively also appeared to have generally gained in coping skills thereafter.

The subspecialty of emergency psychiatry has developed in parallel to the specialty of emergency medicine, as a response to deinstitutionalization and the identification of a large homeless population, the availability of increasingly safer and more effective neuroleptic agents, and the new safer and more rapidly acting antidepressant drugs. A large body of literature attesting to the efficacy of interventions in the emergency department, a formal organization, established protocols and philosophies, and the availability of fellowships for subspecialty training has grown up since the 1960s. The official designation in 1979 of emergency medicine into a specialty transformed emergency rooms around the country from way stations to the morgue into emergency departments, staffed by board-certified emergency physicians and providing which can provide definitive treatment for multiple disorders. The American Association for Emergency Psychiatry was established in 1988. Emergency psychiatric treatment now addresses problems not only in general hospital emergency departments, but all urgent and critical-care settings, including consultation services in hospitals and in outpatient clinics, both psychiatric and nonpsychiatric. In addition, as medical, surgical and obstetrical care become increasingly non-hospital based, psychiatric treatment of even very disturbed patients is being provided by mobile treatment and assertive community treatment programs, in homes, homeless shelters, and at other sites where such patients are causing disturbances. Unique emergency psychiatry facilities have been developed in the past few years that allow clinicians to institute a fairly thorough workup and initiation of treatment for newly diagnosed patients with major psychiatric disorders, often providing from 1 to 3 days of observation, acute treatment, and intensive outpatient programs.

In response to disasters, field hospitals that provide acute care sometimes include psychiatric personnel, not only to treat victims of psychological trauma, but also their families, and to consult with, debrief, and provide emotional support for the rescue workers, who often experience acute stress reactions because of the nature of their work.

Social factors have also demanded that psychiatry develop techniques that produce results rapidly. Health maintenance organizations and utilization review of treatment occurred in response to rising health care costs. It became difficult to justify the position that long-term treatment or long hospitalizations had any advantage over brief outpatient interventions or crisis hospitalization when society or an employer was paying the costs and when patients did appear to respond to many emergency interventions. The problems of homelessness and poverty, diminishing state programs for inpatient treatment of psychiatric problems, substance abuse, and the continuing problems produced by human immunodeficiency virus and related illnesses have increased the role of the emergency department as the primary provider of care for those patients. Although welfare and Medicare programs are being privatized, and the Veteran's Administration also has begun to seek care from private systems in order to cut costs, psychiatric access remains so limited that emergency departments continue to house patients for extended periods, and placement into hospitals or treatment programs becomes a frustrating exercise in intimidation.

DEMOGRAPHICS

The psychiatric emergency patient can be conceptualized in three broad categories, with somewhat different but overlapping demographics. The one that first comes to mind is the patient with an acute episode of a recurrent or chronic psychiatric illness, for whom the usual outpatient facilities are not available. The second is the patient with an underlying medical basis for psychiatric symptoms:intoxication, delirium, drug-drug interactions, or otherwise medically induced psychiatric symptoms. The third group of patients is essentially normal psychologically, but an overwhelming event, such as an accident, crime, or sudden onset or discovery of serious physical illness has strained their psychological capacity. Multiple demographic studies by university-based emergency services tend to measure the first group disproportionately, and find in that population a heavy overrepresentation of abuse of alcohol or illicit drugs as primary or complicating factors. The modal patient is a single female under the age of 40, in contrast to the population at large and also the usual age of a psychiatric outpatient. The elderly tend to underutilize psychiatric emergency services, although when they do present to emergency departments they tend to use a higher amount of medical, psychiatric, and social resources. The proportion of patients who are single (widowed, divorced, separated, or never married) is higher in the emergency population than in the population at large or in the population of outpatient clinic users.

Racial and ethnic characteristics of an emergency department population partially mirror the demographics of the region served. In general, however, African-Americans and those in low socioeconomic classes are overrepresented, while Hispanics and Asians are somewhat underrepresented. City dwellers use the emergency department more frequently than rural citizens, which appears to be more related to accessibility than to any diagnosis or other demographic feature. Most patients come on their own, more frequently after-hours. The most frequent presentations are agitation or anxiety, followed by suicidal or assaultive threats or behavior, and then psychotic behavior. Physical illness or an inadvertent drug reaction accounts for a substantial portion of the rest. Chronic psychiatric illness, with or without substance abuse is commonly determined to be the etiology of these symptoms. A small subset of patients *(repeaters)* habitually use the emergency department for definitive treatment, never following up on referrals to outpatient clinics or drug treatment programs.

BASIC PRINCIPLES OF EMERGENCY PSYCHIATRY

Approach Several aspects of the psychiatric emergency distinguish it from the usual psychiatric contact and dictate management. Such patients do not have an appointment, may appear at any hour, often appear threatening, and frequently do not want treatment. Also, particularly in a very busy emergency service, outpatient clinic, or mobile crisis center several patients may require treatment simultaneously, so a prioritizing mechanism needs to exist.

Initial Evaluation and Management *Triage*, a term derived from military medical practice, means sorting; it is the process of prioritizing patients according to how ill they are and how quickly their problems must be addressed in order to maximize their chance for recovery. Strong consideration is given to the determination of whether delay in treatment will adversely affect the patient's outcome. Traditionally, these decisions have been based on the patient's vital signs, status of airway, breathing, and circulation, which must be attended to promptly in order to preserve the patient's life. For the psychiatric patient, the discriminators are usually more subtle. Furthermore, when the patient presents with disordered behavior, these discriminators are sometimes ignored. Because there is a general scarcity of psychiatrists available to emergency departments, it is especially necessary to triage psychiatric emergency patients accurately. When a patient is evaluated at a freestanding psychiatric facility instead of a comprehensive emergency department, the possibility of an underlying medical condition can be missed. Table 29.2-1 illustrates a method of triage for the psychiatric emergency.

Category	Rationale	Initial Management
Emergency	Abnormal vital signs Threatening behavior	Assess medically Assess control needs Sitter, quiet room, seclusion, restraints
Acute	Overwhelming anxiety Suicidal ideation Patient is unable to explain problem coherently Family expresses extreme concern Patient does not meet criteria for emergency	Emergency pharmacotherapy Assess psychiatrically
Nonacute	Patient requests psychiatric help, but does not meet criteria for emergency or immediate response	Assess psychiatrically, consider referral to psychiatrist or psychiatric clinic

Table 29.2-1 Triage of the Emergency Psychiatric Patient

Somewhat different guidelines for triage apply for a mobile treatment team; in this instance the decision to send a team out must consider the likelihood that on-site intervention will avert the need for hospital treatment, as opposed to guidelines for a paramedic or ambulance call, in which the task is primarily to stabilize the patient in order to allow transport to a hospital. Commonly, the task in the first instance is facilitated by the fact that the patient seen by a mobile team or in an outpatient clinic is known to the team or therapist.

Environmental Measures In the emergency department the physician frequently has to evaluate an unwilling or uncooperative patient, in contrast to the private practice of psychiatry, where considerations such as holding a patient for an adequate evaluation is rare; even if the patient should decide to walk out of an evaluation prematurely, the obligation of the physician is minimal. In addition, the stimulation of a busy emergency department can escalate the patient's symptoms. Environmental controls may be needed to protect the other patients as well as to ensure that the patient does not leave the emergency department unnoticed before the evaluation is complete.

Decreasing the noise, the activity, and the anxiety exhibited by other patients or by medical personnel who are uncomfortable around psychiatric patients can reduce the patient's symptoms. A “quiet room” or a seclusion room is useful in this regard. The patient's vital signs should be obtained before he or she is placed in a quiet room, and then the patient should be monitored visually if possible. Restraints may be needed to hold a patient long enough for a proper evaluation, particularly when the patient is unknown to the staff and the cause of the symptoms is uncertain, the patient is unable to cooperate with the examination, the patient needs a secure environment, or if the patient cannot be left unattended safely. When a disturbed patient leaves the emergency department before an adequate assessment can be done, malpractice actions can ensue, especially if it can be shown that the patient was in an altered mental state when he or she left. In national surveys on the use of restraint in psychiatric facilities, four-point leather restraints remain the most used method of containing an out-of-control patient. Their advantage over medication is that they are immediately reversible, and are an obvious reminder of the patient's condition. Once a diagnosis is clearer, pharmacological management may be used to control the patient's symptoms and behavior and to initiate or resume treatment. Emergency department staff should regularly practice using restraints, and the restraints should be maintained and reviewed regularly, much as hospital crash cart. Seclusion, restraints, and psychopharmacologic management should not be construed as sufficient treatment, nor should these measures ever be regarded as punitive.

Protection of Others There is a growing literature on identification of the potentially violent patient, on staff training in the management of violent or disruptive patients, and on the social network and environment that increases or decreases violent behavior. In most studies, substance abuse, severe psychiatric disorders, and especially the combination of the two are disproportionately disruptive and such patients are more likely to become violent in the emergency department. The presence of uniformed city police and hospital security personnel stationed in the emergency department appears to reduce the incidence of violence and of harm to health care workers. It is not clear how helpful metal detectors are. Willingness to use physical restraints for the patient who has been identified as potentially violent is reassuring to the staff, which in turn sets an atmosphere of calm and a sense that things are under control, which also decreases the incidence of patient outbursts. Other measures that increase the safety and effectiveness of staff are self-defense courses, a protocol for handling violent patients, good lighting, video monitoring of all areas of the emergency department, panic buttons, and secure handling of addictive drugs, syringes, and needles.

Secondary Evaluation of the Psychiatric Emergency Patient Once life-threatening medical conditions have been addressed or ruled out and the patient has been stabilized, the specific nature of the psychiatric problem can be addressed. The decisions to be considered are whether it can be treated definitively in the emergency department, requires a referral for follow-up, or whether more acute measures such as hospitalization, admission to a crisis facility, intensive outpatient treatment, management by a mobile treatment team, or assertive community treatment are necessary. In a traditional outpatient setting, a psychiatric evaluation includes the most likely diagnosis and a treatment plan, and the patient's cooperation is obtained in instituting that plan. The emergency assessment is often quite different. The most important priorities are preservation of the patient's life and well-being, and secondarily, whether definitive care can be completed within the psychiatric emergency system. The modal psychiatric patient presents a systems issue rather than a diagnostic one. In the emergency assessment making a diagnosis may be less significant and more difficult to do, or in the chronic psychiatric patient all too obvious, given the constraints of time and lack of adequate information, than addressing the relationship of the patient to primary support systems. For the chronically mentally ill patient or the substance abuser, the foremost question is not the diagnosis but where the primary support system has failed. A chronically mentally ill patient who decompensates because he has stopped his medication may not need hospitalization if he or she is not suicidal, is aware of the need for treatment, and has a supportive environment such as a community mental health program which will follow him. On the other hand, the acutely psychotic patient who has no obvious social support, is unwilling to acknowledge psychiatric illness, and may be dangerous to others may require hospitalization each time his illness reaches a peak.

An effective treatment plan depends upon the availability of resources, and on the time frame for emergency treatment. Although extended observation in the emergency department may be useful in some instances, in practice this is often a way of postponing a difficult decision. Crisis beds and mobile treatment teams allow for repeated observation of the patient over time, ascertaining whether the initial treatment is working or needs modification. In general, however, the dispositional decision for a psychiatric emergency patient must be made within a few hours. This decision is influenced by the availability of beds or alternative treatment options such as day treatment or crisis intervention programs, the patient's own resources and motivation, and whether any dispositions must be preapproved by a third-party payer.

The disposition becomes more self-evident when the focus of the evaluation is the immediate reason for the use of emergency services, especially when the patient has a long-standing psychiatric illness. If the onset of the illness was acute, an adequate medical assessment as well as a psychiatric one is needed. In a patient who has no known psychiatric history, this is crucial, but even when there is a previous psychiatric history, an acute medical cause can precipitate the disordered behavior. Medical assessment of a psychiatric patient in the emergency department is often problematic. The physical examination may seem magical or intrusive, and is commonly avoided by the psychiatrist, for whom it is not generally routine. There is often a hope that the patient will have been medically assessed and cleared prior to the psychiatric consultation. However, the reverse is often true: a patient exhibiting disordered behavior is frightening to nonpsychiatric physicians, who are reluctant to approach such a patient, let alone complete an adequate evaluation to rule out medical emergencies. It is not uncommon for triage nurses or paramedics to forgo blood pressure, pulse, and temperature measurements on such patients. Some psychiatric services may refuse to evaluate a patient until the patient is

medically clear, and hospitals often will not admit a psychiatric patient without medical clearance. However, every medical condition cannot be ruled out in an emergency department evaluation. The emergency physician generally performs only a focused evaluation, addressing those complaints that are of immediate concern to the patient; most such evaluations take only a few minutes.

Most people find it difficult to believe that a patient is not able to control his or her behavior. Patients with psychiatric symptoms commonly walk into an emergency room (the vertical emergency) rather than being wheeled in or carried in on a stretcher. This makes it difficult to appreciate that they may not be in control of their behavior, and also tends to create the impression that they are less ill. Frequently, the nurse forgets to obtain vital signs. The overall impression is that there is something wrong with the patient's mind, not his body, and that a freestanding psychiatric facility is the appropriate site for treatment. Whereas medical emergencies have clearly recognizable characteristics, such as an obvious injury, the acute onset of pain, seizures, or a fever, the psychiatric patient may be well groomed, physically intact, and ambulatory.

> A 43-year-old physician entered the emergency department, having driven himself to the hospital. He reported to the triage nurse that he had taken 100 lithium tablets and 40 imipramine tablets within the last 15 minutes. He also gave her the name of his treating psychiatrist. Because the patient was so coherent, and perhaps because he identified himself as a physician, standard triage protocols were ignored. The patient's vital signs were not recorded, and the nurse reported the information she had obtained to the emergency physician. The emergency physician also overlooked normal protocols, and proceeded to call the patient's psychiatrist. While he was on the telephone awaiting a response, the patient collapsed and was resuscitated only after strenuous efforts by the emergency department staff.

Patients with chronic psychiatric illnesses who also have a medical illness may interpret their symptoms in psychotic ways, and will request to see a psychiatrist rather than an emergency physician. Chronically ill psychiatric patients also may have learned from experience that they will not be taken seriously by anyone other than a psychiatrist. Even if they do realize that they may have a serious medical illness, they will ask to see a psychiatrist. In addition, their psychiatric symptoms may be exacerbated by the medical illness.

One aspect of psychiatric illness is that the patient frequently does not acknowledge being sick. The patient may actually be experiencing the sudden onset of serious psychiatric illness; not uncommonly, however, symptoms are long-standing but are undiagnosed, until some event brings the patient's behavior to the attention of others. At that point the patient is brought on an emergency basis. The patient may be quite resistant to the notion that he has any psychiatric disorder and will not cooperate with any examination or treatment.

Medical Assessment Four elements in the traditional physical examination will help to rule out significant medical illness: specific historical data, vital signs, certain physical signs, and the mental status examination. The following findings should direct attention to medical evaluation, and carry particular weight when there is no additional medical assessment, as in a psychiatric clinic or with a mobile treatment team.

HISTORY The presence of medical illness should be strongly considered when psychiatric symptoms appear suddenly in a previously well-functioning person. Symptoms significantly different from a patient's usual ones in a patient with long-term mental illness also suggest medical illness or drug reaction. A patient over the age of 30, with psychotic symptoms appearing for the first time, an awareness or conviction that these symptoms are foreign, and especially with concomitant symptoms of cognitive dysfunction should be considered to have a possible organic illness. Clinicians should consider medical and substance–related causes when a patient has a history of a recently diagnosed medical illness, a new prescription, or a change in dosage. A personality change or marked lability of mood noticed by friends or relatives also suggest the onset of a serious medical condition.

PHYSICAL FINDINGS Vital signs should be obtained on admission to the emergency department, even when the patient appears physically healthy or is intimidating because of disturbed behavior. Psychiatric disorders do not affect vital signs to a significant degree; therefore, abnormal vital signs must be further evaluated by a history, physical examination, and appropriate laboratory tests. Elevation of temperature, pulse, or respiratory rate suggests an underlying medical condition. Certain physical findings strongly suggest an organic disorder, including lateralizing neurological symptoms, confusion, and incontinence. Table 29.2-2 lists common medical conditions that frequently present as psychiatric emergencies and may be marked by abnormal vital signs. Table 29.2-3 lists those conditions that, although not common, may also appear initially as psychiatric problems. Delay in recognizing them can be life-threatening. The laboratory tests listed in Table 29.2-4 may clarify the diagnosis; the psychiatric evaluation can proceed while these tests are pending.

Hypothyroidism	Myocardial infarction
Hyperthyroidism	Alcohol intoxication
Diabetic ketoacidosis	Alcohol withdrawal
Hypoglycemia	Chronic obstructive pulmonary disease
Urinary tract infection	Acute liver disease
Pneumonia	Substance withdrawal

Table 29.2-2 Common Medical Illnesses That Often Present as Psychiatric Emergencies

Myocardial infarction	Hypokalemia
Pulmonary embolism	Hypercalcemia
Subarachnoid hemorrhage	Hypocalcemia
Epidural hemorrhage	Splenic rupture
Cocaine intoxication	Subacute bacterial endocarditis
Amphetamine intoxication	Steroid-induced psychotic disorder
Encephalitis	Phencyclidine-induced psychotic disorder
Malignant hypertension	

Table 29.2-3 Less Common but Potentially Life-Threatening Conditions That Present Frequently With Psychiatric Symptoms

Complete blood count
Electrolytes
Blood alcohol concentration or breathalyzer
Blood glucose concentration
Calcium level
Urine assay for substances of abuse

Table 29.2-4 Laboratory Studies to Assess Potential Organic Causes of Acutely Altered Behavior

MENTAL STATUS EXAMINATION Several mental status examination formats have been used widely to determine the presence of organic disorder. The most widely used is the mini-mental status. Regardless of the instrument used, the important point is that some assessment of the patient's mental status should be obtained consistently and recorded. The patient's appearance, ability to cooperate in the examination, level of consciousness and whether it changes during the evaluation, the lability of affect, cognition, memory, judgment, and the presence of agitation are important in determining an underlying organic basis for disordered behavior. Table 29.2-5 provides a brief mental-status examination for use in the emergency department. The first five questions relate to the patient's orientation and the rest to the patient's memory, judgment, ability to abstract, and interest in the outside world.

1. What is your name?
2. Do you know where you are?
3. Can you tell me how old you are?
4. What is today's date?
5. How did you get here?
6. Can you tell me why you are here?
7. What do you think you need from us today?
8. Tell me something important that you have heard on the news in the past few days.
9. I am going to give you a five-digit number. (Do so). Please repeat it back to me.
10. Spell the word "marble." Now spell it backwards.

Table 29.2-5 Brief Mental Status Assessment for the Emergency Department

Psychiatric Assessment A psychiatric assessment in the emergency department focuses on four specific questions. What is wrong? Why is the patient seeking treatment now? What was the patient like before? What does the patient expect? It is important to know whether there is a previous psychiatric history in the context of these four questions.

WHAT IS WRONG? A patient experiencing a psychiatric emergency may not be able to tell you what is wrong and may not believe that anything is wrong. Because psychiatric patients do not have obvious injuries or physical illness, and may have normal vital signs, they frequently are asked to wait while other, more obvious emergencies are addressed. One result of this inaccurate triage is that others who might have accompanied the patient and could give useful information may leave before they are seen. Police officers often take these patients to jails or detention centers because they can be relieved of responsibility for them much more quickly than in an emergency department. Always consider that the patient may have an underlying medical condition even in the absence of initial findings.

A 74-year-old widow was brought to the emergency department by a social worker who wanted the patient hospitalized until an alternative placement could be arranged as the patient had been evicted from her apartment. Questioning determined that the patient had been evicted from 30 or more residences in the previous five years for creating disturbances that upset other tenants. The patient explained that she stayed awake every night in order to bang on the walls with a broom to keep a gypsy from entering her apartment and spraying her with a poisonous substance. She stated that this harassment had been going on since her husband died, and that she knew it was continuing because when she did fall asleep, she would awaken in the morning to find her legs swollen. A mental status examination was remarkably normal except for that delusion. Medical evaluation revealed congestive heart failure, which quickly responded to digitalis. The patient was placed in a senior citizens' apartment complex without further problems.

A 39-year-old patient was brought to the emergency department stating that she was the “bionic woman.” She did not further clarify that description, but said that her primary concern was her difficulty unlocking the door when she arrived at work that morning. She thought that she should be evaluated by a doctor. Her vital signs were reported as within normal limits, including a temperature of 35°C, blood pressure of 100/40 mm Hg, and a respiratory rate of 15 per minute. Because she looked somewhat disheveled and seemed to be an unreliable historian, the decision was made to hospitalize her on the psychiatric unit of the general hospital, where she received her first physical examination. She was noted to have sparse body hair (she had been wearing a wig), her vital signs were as reported, and she had distinctly hypoactive reflexes. She also had a 14-inch well-healed horizontal scar at the base of her neck, which she said was a result of a thyroidectomy some 20 years earlier. Laboratory studies confirmed that she had virtually no endogenous thyroid hormone. Thyroid replacement led to complete recovery.

WHY NOW? What has changed? Even though it is very helpful to ask this question in a usual outpatient evaluation, in a psychiatric emergency it is essential to know the specific precipitant for the visit at that particular time in order to make a reasonable recommendation for further treatment. For example, a patient may present to the emergency department distraught and depressed. If the psychiatrist has not asked the relevant questions that reveal that the patient's therapist has just left for vacation, a lot of time may be spent in trying to arrange a disposition to outpatient treatment that is not appropriate and will not be effective because it does not address the acute symptoms. A long-term mentally ill patient might appear to need hospitalization when actually the patient has undergone a recent trauma, and needs some crisis intervention in the form of understanding about how difficult the experience was. The insidious onset of a chronic mental illness may only come to the attention of professionals when something untoward has occurred in the environment; the letter carrier, a visiting relative, or a teacher notices that something is amiss and forces the family, who had been ignoring bizarre or isolative behavior for several weeks, to take action. A patient who is known to have a psychiatric illness or who has been prescribed psychotropic medication may exhibit new psychiatric symptoms that are related to a change in the medication or to a drug-drug interaction rather than to the illness.

A familiar experience for house officers in the emergency department is a sudden surge in psychiatric emergencies in July of each year, when residents complete training and transfer their patients to new therapists. The phenomenon of a crisis reaction in response to the loss of a therapist is generalizable to the loss of any empathic caretaker, including an internist, nurse, or relative. Asking why the patient did not contact the previous therapist at this time can quickly get to the root of the present crisis, and often point the way to a more effective intervention.

The time of day that patients come to the emergency department is often an implied expression of their wish for a particular disposition. Patients and families of psychiatric patients who arrive after normal business hours often know that the community mental health clinic or the admitting office to the state hospital is closed, and that the records from a therapist might be more difficult to obtain, forcing an admission that in working hours could be averted.

A 70-year-old woman was depressed following the death of her husband. She lived alone, and although some family members were concerned that she had a significant mental problem, they lived in another city and could not convince nearby relatives and friends that there was something seriously wrong. When her visiting nurse went on vacation, the woman did not recognize the substitute who was sent. She became paranoid and threatened suicide. The substitute visiting nurse immediately arranged to have the patient brought to an emergency department. Thus, although the illness and deterioration had been going on for over a year, it was only the change of the evaluator that defined the situation as an emergency.

A 16-year-old boy became reclusive, and for several months refused to attend school or to participate with his family in any activities, including eating meals or watching television with them or accompanying them on visits to relatives. The family was dysfunctional and ignored the son's abnormal behavior, even allowing him to eat his meals in his room. When a letter carrier arrived with a special delivery letter and found the doorbell answered by a naked, disheveled, and bizarre-looking young man (the rest of the family was out), he notified the city's Department of Social Services. The agency responded immediately by sending an ambulance to bring the boy into the emergency department for evaluation. Although the disorder, paranoid schizophrenia, was well established by that point, it was perceived as an emergency only because a new observer had witnessed the strange behavior.

WHAT WAS THE PATIENT LIKE BEFORE? The primary goals for the emergency intervention is to reestablish the patient's baseline, to place the patient appropriately within the system of health care delivery and to minimize future use of the emergency services. It may be very difficult to determine a baseline when a patient is floridly psychotic, especially if no background information is available. Police or paramedics who brought the patient in or a mobile treatment team that has seen the patient's current living situation may be very helpful in this regard. How the patient is dressed, the level of language used, or the patient's description of daily activities are clues to how well the patient was functioning before the onset of the present illness.

WHAT DOES THE PATIENT EXPECT? Patients have an idea of what will happen when they seek treatment, which may be positive or negative, based on previous experiences or on what they have been told. Generally, these expectations are appropriate or at least directly related to their presenting complaint, for instance to be cured, to achieve symptomatic relief, or to get a prescription. However, when a recommendation has been made or a treatment plan is begun and the patient still seems very distressed, or when the patient returns to the emergency department within the next few hours or days with a fairly similar complaint, it is possible that there were unspoken expectations that had not been considered initially. For instance, the patient may expect to be hospitalized, even arriving with a packed suitcase, when the presenting complaint did not warrant hospitalization. The patient may want a prescription, a letter permitting him to miss work, or a call made to a girlfriend or parents “explaining” some inappropriate behavior or failure. As described earlier, the patient may have the magical expectation that in every medical evaluation, the doctor will have had the foresight to look for and rule out acquired immune deficiency syndrome (AIDS) or cancer. Patients are generally reluctant to ask directly for anything that feels like an excessive demand, especially when it is out of the boundaries of direct medical care. While they may have no trouble requesting, or even demanding a prescription, or an additional lab test, they may find it harder to request a letter for an employer or completion of an insurance form; they may expect physicians to read their minds. These issues appear in most doctor-patient interactions, but are more apparent an emergency because there has not been an established relationship in which these elements have been worked out.

A 47-year-old woman with schizophrenia, chronic paranoid type, presented to the emergency department incoherent and physically out of control. She walked in a bizarre fashion, and periodically her knees would appear to collapse. The psychiatrist recognized her, and was aware that there were no beds available that day on the psychiatric unit. She told this to the patient, and offered intensive outpatient treatment instead. At that point, the patient fled into the main hospital lobby reciting the Lord's Prayer loudly, attracting attention from passersby. The psychiatrist followed her and reiterated that no beds were available, but offered the patient some additional neuroleptic that evening, promising that she would be hospitalized on the following day if it was necessary. The patient became quite coherent and cooperative then, explaining that her son was graduating from high school the next day, and that she was overwhelmed at the thought of trying to provide some sort of celebration for him. The psychiatrist offered to write a doctor's excuse insisting that the patient's medical condition mandated bed rest, and prohibiting a celebration at home. With this note the patient was able to leave the emergency department, and was seen the next day in outpatient follow-up with marked clearing of her inappropriate behavior.

A 50-year-old man with schizophrenia, chronic paranoid type, well-known to the community mental health clinic, staggered into the waiting room clutching his chest, from which serosanguinous fluid was oozing. Clinic personnel sat him in a wheelchair, and ran to call the police, assuming that he had been stabbed. The psychiatrist approached the patient and asked what the problem was. It was determined that the patient had recently had surgery for an emphysematous bleb, and when the wound began to ooze he was reluctant to return to the hospital, choosing instead to come to his familiar clinic for help, knowing he would be cared for and listened to there.

Management of the Psychiatric Emergency Patient After the initial evaluation, when it has been determined that the problem is primarily psychiatric, the physician needs to decide how extensive an intervention to implement. Practical techniques to facilitate recovery focus on diminishing the emergency and defusing the problem sufficiently to allow the patient to utilize more traditional treatments. The goal is not necessarily complete resolution of the patient's problem, but enough clarification of the problem that the solution is evident. This may be accomplished over several visits.

Crisis Intervention During a crisis, the fluid nature of events offers the opportunity to effect major changes in problem-solving that would not occur during more stable periods. If maturation is thought of as a process of mastering small crises over time, then crises are the windows in which growth occurs; the patient does not simply return to a prior baseline when the crisis is over. The emergency psychiatrist provides stability for the distressed patient, and within that relationship helps to identify the elements of the problem, assess the critical issues, consider more effective options, and arrive at a solution.

One group of patients that has been found to respond unexpectedly well to crisis intervention is the seriously or chronically mentally ill, who, in the words of an old joke, may be crazy but are not stupid. It has been shown that addressing the specific emergency with the patient often results in rapid clearing of psychotic symptoms.

In general, the criteria for recommending crisis intervention or brief therapy are a clear precipitant for the patient's decompensation, strong social supports, and good premorbid function. The premise on which the following form of crisis intervention is based is that the reason the patient has been unable to cope with the precipitant was unpreparedness. In brief therapy, a temporary therapeutic relationship with the psychiatrist or other mental health professional will be provided, which will then terminate. However, this termination will be defined either by a termination date or a specific number of sessions so that it can be anticipated, talked about, and mastered. In brief therapy the patient is first encouraged to talk about the acute event, repeatedly and in detail, and then, in the context of planning for the termination of therapy, to discuss how life will go on hereafter.

During brief treatment two lines of thought will be evident: first, the patient will talk about all aspects of the precipitating event, such as the patient's ambivalence, fantasies the patient may have had about anticipating or foretelling the event; and guilt, embarrassment, or subsequent symptoms. Ideally, the patient will then be able to express some anger, (e.g., “Why did she die and leave me!”) More often the anger comes in the form of a metaphor, expressed as abandonment by the therapist, especially around termination. The therapist can point this out as being a displacement from the precipitating event, and continue to facilitate the patient's expression of understandable feelings. When the total treatment time has been defined at the outset, it appears that patients go through the same process and with nearly the same results even in therapy lasting one emergency department visit.

The application of crisis intervention is limited by the ability of the physician who sees the patient initially to see patients in follow-up visits. Relative contraindications to brief therapy are overwhelming suicidal thoughts, poor premorbid functioning, or continuing deterioration over the first two or three sessions. Ideally, if medication seems necessary, it should not be prescribed until the patient has been seen and reassessed in a second session, enabling the psychiatrist to determine the patient's capacity to engage in and profit from the therapeutic alliance.

A 23-year-old building and grounds engineer came to the emergency department complaining of nightmares and panic attacks. He dated the onset of these symptoms to an assault at work at an apartment complex; a man who appeared intoxicated came running at him waving a handgun. The patient fled in his truck and returned to the office of the city housing department, his employer. He described the above episode, gave a detailed description of the assailant, and went home. Over the next several days, he called the office each morning to find out what action they had taken on his complaint. When he met repeatedly with an apparent lack of response, especially coupled with a threat that he would lose his job if he didn't return to work soon, he began to have difficulty sleeping. On the morning that his sick leave expired, he got into his truck to return to work, but within minutes experienced tightness in his chest, shortness of breath, and profound sweating. As unlikely as he knew that was, he feared that he was having a heart attack and drove himself to the hospital. On the initial visit he was evaluated for a cardiac event; findings were negative and he was discharged without follow-up. After the fourth panic attack he decided to seek psychiatric consultation. Therapy, which was completed in four 45-minute sessions in the emergency department, consisted of detailed inquiry into the episode, followed by a similarly detailed discussion of his attitudes about his job. The nightmares changed character from his view of himself as a passive victim to viewing himself as empowered, and then stopped altogether. He also explored other job possibilities, and within one month began work at a higher paying and more challenging job, much more suited to his educational background and abilities.

Emergency Pharmacotherapy Emergency pharmacotherapy is an important adjunct for stabilization of the psychiatric emergency. However, there are several good arguments for caution in the use of medications in the emergency department. If the patient has a chronic psychiatric illness, medical management should be entirely handled by the system that oversees the patient's care. The emergency management of such a patient should be directed towards determining what went wrong within the system that allowed the patient to run out of medications, decompensate, or otherwise fail treatment. Any response that does not take the system failure into account ensures the likelihood of a repetition. If the patient is new to the mental health system or has an acute, new condition, the ideal management would be to enroll the patient with a therapist or clinic as soon as possible, possibly observing him for 24 to 72 hours or having him return to the emergency department several times over the next few days to determine whether the medication is working. If emergency hospitalization is required, administering medication before transferring the patient to the hospital may obscure the presenting symptoms.

A patient should be medicated only for definite target symptoms. Just as morphine can obscure the diagnostic landmarks in an acute abdomen, antipsychotic medication will modify psychotic symptoms caused by a general medical condition just as effectively as in a functional psychosis, without clarifying the diagnosis. Overmedication may delay evaluation by several hours. Acute depressive symptoms might respond to the interpersonal interaction with a psychiatrist alone; prescribing an antidepressant prematurely may blur both the diagnosis and the need for medication.

Under the following circumstances, medication should be administered to the patient in the emergency department.

1. The patient is out of control, and less intrusive means, including a quiet room, the presence of staff members, or restraints, have been used with minimal effectiveness.
2. The patient's behavior is beginning to compromise evaluation or the care of other patients in the area.
3. Medication is being used to control psychotic symptoms that would otherwise interfere with further workup.
4. The cause of the patient's symptoms is established and medication is indicated to gain the patient's cooperation for further care.
5. Medication increases the patient's ability or willingness to cooperate for transfer to a psychiatric facility that has already agreed to accept the patient.
6. The patient is enrolled in a care system that will be responsible for follow-up.

Noncompliance with medications is the major factor in decompensation of the chronically mentally ill patient. Although hospitalization may be averted by resuming the regularly prescribed medication, in order to prevent subsequent decompensations it is very important to determine why the patient had discontinued medication or missed appointment.

RAPID TRANQUILIZATION An acutely agitated or psychotic patient can be calmed in a very short period of time. For several years various protocols for rapid neuroleptization were promulgated; however, it was never demonstrated that rapid relief of acute psychosis led to shorter hospitalizations or that it was in the best interests of the patient to use this treatment. A common practice in the United States is to administer high-potency dopamine receptor antagonist, intramuscularly if necessary, and a benzodiazepine concurrently. This allows for a lower dose of the antipsychotic, and a decreased likelihood of extrapyramidal adverse effects. Although there are as yet no controlled studies regarding the use of the newer seroton-dopamine antagonists or anticonvulsants in emergency room settings, their efficacy and relative lack of extrapyramidal symptoms make them potentially useful alternatives.

Decision to Hospitalize A great deal of time is spent in the emergency department to determine if hospitalization is indicated for a given patient. Locating a bed, obtaining the patient's consent or completing paperwork for certification of the need for hospitalization, procuring a bed, arranging transportation, and verifying the patient's insurance status, especially when the hospitalization must be preauthorized, triples the amount of time spent on psychiatric patients in comparison to the general emergency department caseload. Although alternatives to hospitalization may appear to provide more humane care for patients, they are not well developed yet in most cities.

The efficacy and cost-effectiveness of alternative interventions remain to be proved. Professional involvement is intense, and is not necessarily rewarded by an improved outcome. For patients with a history of prior hospitalizations, costs, outcome and subsequent hospitalization rates are not improved by aggressive outpatient crisis management. The decision to hospitalize a patient is influenced more by the inability of social support systems to deal with disturbed and disturbing behavior than on the patient's diagnosis or severity of symptoms. Enrollees in prepaid health plans are less likely to be hospitalized than are patients with conventional insurance, but this may have more to do with the continuity of care than with restrictions of capitated plans. Ideally, an emergency psychiatric hospitalization should be reserved for patients whose long-term outcome will be otherwise compromised, such as by suicide, drug reactions, or other life-threatening situations.

SPECIAL PROBLEMS

Violence Violent and potentially violent patients evoke anxiety, with the consequence that they may be treated punitively. This includes prescribing neuroleptic drugs before the patient has been adequately evaluated, and in doses that may be dangerously high. Predictors of violence are unreliable, although some protocols have been developed to identify warning signs and useful responses to them, and thus theoretically to minimize violence in the emergency department. Although psychiatric disorders account for a significant number of violent incidents, notable contributors to violence are intoxication, brain injury, or a metabolic disturbance. The environment of the emergency department, which often is noisy, bright, loud, and intense, may drive patients who are already struggling with control over their behavior to become violent.

Management Prevention of violence is easier than regaining control after a patient has caused harm. A protocol for control of violence is effective not only for managing the patient but also for reassuring the staff. At the first indication of threatening behavior, the patient needs to be confronted with the fact that such behavior is frightening and may interfere with provision of care. Ideally, one member of the staff should be the designated spokesperson in these situations. If patients are not able to demonstrate an ability to conform their behavior to the needs of the emergency department, they must be restrained in a humane manner until an adequate evaluation of the reason for the emergency department visit can be completed. Leather restraints and an isolated, quiet area are essential to the management of such patients. Short-term sedation, generally with benzodiazepines, is the common practice in the United States, particularly when there is a suspicion that substance abuse underlies the behavior. Consequences for belligerent, violent, and destructive behavior may include pressing assault charges after the patient's medical needs have been addressed.

Acute Psychosis An acutely psychotic patient, possibly responding to auditory or visual hallucinations, may have a chronic psychiatric illness, or the psychotic symptoms may be an expression of a life-threatening medical condition, injury, or intoxication.

Management Psychotic symptoms may be aggravated by the excitement of the emergency department; the patient may also be paranoid and may interpret offers of help as threats and intrusions. Such patients should be moved from a more stimulating to a less stimulating environment, where they can be evaluated for the presence of an underlying treatable medical condition. If this is a first episode of psychotic symptoms or if these symptoms represent a significant change for a psychiatric patient, this workup is crucial. In addition, serious medical illness may precipitate relapse of a long-standing psychiatric illness.

Unrecognized Depression Multiple vague somatic complaints are common in the emergency department. These are the patients who frequently irritate the nonpsychiatrist because nothing seems to be wrong. Backache, headache, sleep disturbance, stomach upsets, and fatigue do not lend themselves to resolution by laboratory tests, X-rays, or specific diagnosis. Because women are more likely to present with such complaints than are men, they are also regarded less seriously by physicians. When the psychiatrist is consulted it is more often because nothing can be found than because the emergency physician believes the patient has an underlying psychiatric problem. The most likely diagnosis is a depressive disorder, although multiple somatic complaints are also an early harbinger of schizophrenia. Victims of domestic violence and parents who are physically abusing their children also may seek treatment in emergency departments, unwilling or unable to state the true problem but hoping that the physician will recognize their distress. Patients who also fear that they have a very serious underlying medical problem such as cancer or AIDS, but have no specific symptoms of those diseases seek multiple medical evaluations believing that one of these visits will confirm their worst fears. These patients may be depressed, hypochondriacal, psychotic, suffering from somatization disorders or personality disorders, or they may have experienced the death of someone close to them recently, in particular from cancer or AIDS.

Management A patient with multiple or nonspecific physical complaints often experiences relief when the problem is identified as depression. However, depressive disorders are frequently chronic or recurrent illnesses, and should not be treated as an emergency. After determining that there is no threat to life, such as from suicidal ideation or intent, the patient should be referred to an appropriate clinic or therapist for definitive treatment. In terms of the long-term relationship with the emergency physician, the psychiatrist should provide some feedback about the features that distinguish the various possible diagnoses, hopefully minimizing the laboratory tests that reinforce the patient's belief in the existence of a serious medical problem.

Adjustment Disorder Adjustment disorders arising from change and conflict are common occurrences. Although the most common precipitants of an adjustment disorder are changes in relationships, such as through divorce or death, patients who endure a severe medical illness or physical trauma will frequently develop significant psychiatric symptoms, also generally in the form of an adjustment disorder. Sometimes the symptoms arise months after the event, and the patient is unaware of the connection. The patient presents to an emergency department with symptoms of acute anxiety or symptoms of the earlier illness that no longer is present.

The friends and relatives of patients who are being treated for severe injuries or illnesses, or who die in a disaster or after resuscitative efforts in the emergency department, and the survivors of a disaster who suffer minimal or no physical sequelae, experience acute stress reactions. When the symptoms become disabling,

resulting in significant disruption of daily life, physical symptoms, or excess substance use, the patient may seek treatment.

Management Professional intervention is generally not needed for acute stress reactions, and premature intervention of a prophylactic nature may even interfere with a person's ability to master the events. Brief psychotherapy is a very effective treatment for uncomplicated adjustment disorders, and can be administered in an emergency department or in more traditional psychiatric treatment settings.

Substance Abuse Substance abuse as the presenting problem or as a complication of other emergency department presentations is frequent; it is also frequently undetected. Over half of the patients seen in the emergency department because of traumatic injuries sustained in fights or vehicular accidents or because of self-inflicted wounds have positive results to an alcohol or drug screen. The estimates of intoxication in psychiatric emergency patients ranges from 4 to 20 percent. Homelessness and unemployment are more frequent among intoxicated patients, who are predominantly male. Comorbid psychiatric diagnoses seen commonly in the emergency department are panic attacks, anxiety disorders, and depressive disorders. Although alcohol is still the most commonly abused substance, stimulant abuse and intoxication is also frequent.

Management A high index of suspicion and drug and alcohol screens will detect many patients who deny substance abuse or who do not volunteer this information to the physician. Although patients commonly will not follow through on referrals to substance abuse treatment programs, confrontation when there is laboratory confirmation and consistent referrals to established treatment offer the best outcome. Most patients who are not identified or confronted as substance abusers and are not referred to appropriate treatment centers will continue their substance abuse behavior.

Anxiety Disorders and Panic Attacks Patients experiencing anxiety disorders and panic attacks commonly perceive them as acute medical episodes, believing that they are having a heart attack, pulmonary embolism, or other life-threatening event. Hyperventilation may also induce symptoms similar to panic attacks, including lightheadedness, dizziness, shortness of breath, paresthesias, and tingling about the mouth. Patients who have begun to abuse minor tranquilizers for relief of anxiety symptoms seek those medications from many sources, including the emergency department. Anxiety is also a prominent symptom of withdrawal from a wide variety of substances, including nicotine and caffeine.

Management When the true cause of the symptoms is recognized, expensive and unnecessary medical tests, which are often countertherapeutic, can be avoided. The emergency treatment of hyperventilation is straightforward. The cause of the patient's symptoms are explained as rapid breathing, which can be treated by rebreathing into a paper bag. That maneuver causes carbon dioxide to build up in the patient's circulation, which will abort the symptoms. Patients often experience great psychological as well as physical relief from their symptoms by this treatment because it is so concrete and effective. Long-term treatment of anxiety disorders and panic attacks should not be instituted in the emergency department, because the abuse potential for the medications used in treatment of these problems is so high. Such patients should be referred to a psychiatrist or psychiatric clinic for more definitive care of these illnesses.

AIDS AIDS remains the modern-day plague. Because of the high incidence of substance abuse among emergency department patients, there is an increased potential for exposure of emergency personnel to the human immunodeficiency virus (HIV). Universal precautions, including gowns, gloves, and goggles, have become the standard in most emergency departments.

Whatever the initial presentation, AIDS ultimately affects the central nervous system, either directly or secondarily, and so psychiatrists are frequently consulted. Sometimes the first symptoms of the illness are psychiatric, ranging from depression, confusion, or psychosis to sudden delirium. In addition, AIDS patients are especially vulnerable to acute grief reactions to the initial diagnosis, sometimes confounded by the fact that they have seen others who died from AIDS. Encephalopathy is diagnosed before death in as many as 30 to 40 percent of AIDS patients; at postmortem up to twice that number have evidence of central nervous system involvement. Many patients who are at risk may present to emergency departments wanting to be examined or tested because they fear that they have acquired the virus.

Management A careful history may rule out the possibility of exposure to HIV, but patients may remain anxious until there is verification by laboratory testing. Now that home tests are available, this may become less of a problem, but the underlying issue is still the patient's anxiety. This may respond to an emergency intervention, but may require referral to a psychiatric follow-up. For the patient who has an AIDS-related psychiatric disorder, treatment is specific to the presenting symptoms and the overall medical needs of the patient.

SPECIAL POPULATIONS

Trauma Victims Victims of assault are generally brought to the emergency department if there are physical injuries. In a large majority of patients who have been mugged or raped, the physical injuries are minor but the psychological wounds are profound and long-lasting. When the psychological reactions to such traumas are recognized early, treatment can prevent years of dysfunction, nightmares, flashbacks, and substance abuse, and even save marriages and lives. Long-term follow-up of rape victims demonstrates persistence of guilt feelings, low self-esteem, phobias, and irrational lifestyle changes years after the event. The emergency physician sometimes realizes that a traumatic injury was a suicide attempt, either intentional or covertly, and consults a psychiatrist.

Domestic violence is a serious emergency department problem. Some studies suggest that 20 percent of emergency department patients are there because of injuries inflicted by a spouse. Victims of domestic violence may have classic injuries of domestic abuse, or may present with vague complaints such as headache, backache, fatigue, depression or “nerves.”

Management Brief psychotherapy techniques enable many trauma patients to resume good functioning quickly. Victims of rape or domestic violence can be referred to advocacy groups that provide counseling, legal advice, and support for the prosecution of the offender. A few programs provide financial compensation to victims of crimes. Potentially suicidal or self-destructive behavior should be evaluated carefully.

Much like the battered child syndrome, the constellation of symptoms that point to underlying domestic violence may be missed unless there is a high index of suspicion. When the physician fails to identify the domestic violence that underlies the patient's injuries, the patient may be subjected to further injury upon returning home. The abusive spouse, enraged by the prospect of being discovered, may beat the spouse or lover again upon finding out that she has been to an emergency department, even though she did not divulge the fact of abuse to the physician. A patient who has been the victim of domestic violence severe enough to warrant emergency department treatment should be encouraged to take legal action, enter a safe house, or otherwise take steps to be protected from further abuse. Like the patient with a primary substance abuse problem, the victim of domestic violence may require repeated recommendations over a number of emergency department visits to acknowledge that this is a serious life-threatening problem, and that some action needs to be taken to change this.

Emergency Room Repeaters A familiar group of patients, well known to emergency department personnel, show up for a host of problems ranging from trivial complaints to serious ones on such a regular basis that they are known as *repeaters* or even as “groupies.” The staff may regard them with bemused concern, or may feel irritated and taken advantage of when the patient's complaints do not appear to warrant emergency care. Repeaters tend to have psychiatric diagnoses, and in comparison to other patients are significantly more likely to have homicidal or suicidal thoughts, to be intoxicated, and to have a primary diagnosis of substance abuse or schizophrenia. They are also more likely to have had a history of previous psychiatric treatment and psychiatric hospitalizations.

Management When a patient is recognized as a repeater in that visits are frequent, complaints appear trivial, and the patient has not followed any of the instructions for follow-up from previous visits, the emergency department staff needs to discuss their own behavior with regard to previous management of that patient. Potential issues include determining whether they have overlooked an underlying problem (such as organic brain disorder or substance abuse), why there might have been a breakdown in the prior treatment plan or whether the referral was not feasible for the patient, (e.g., because of cost, distance, or unavailability), or whether the patient has been made to feel special in some way. Otherwise, repeaters may evoke frustration that can lead to inappropriate management, which in turn may be life-threatening. The staff must be consistent in their approach, exploring the present cause for the visit and making an appropriate recommendation for follow-up. Mobile treatment programs and integration of emergency services with community mental health treatment programs may minimize repeated visits to the emergency department by chronically mentally ill patients who are in crisis, by identifying crises very early or even preventing them by aggressive case management.

Homeless Mentally Ill It is estimated that 25 to 35 percent of homeless persons suffer from serious, chronic mental illness. Lack of an address interferes with the ability of a homeless individual to receive ongoing care in a psychiatric clinic; furthermore, when there is a daily struggle to find food and shelter a scheduled clinic appointment may not be a priority. In the past few years medical assistance programs in many states have been privatized, meaning that patients must obtain referrals for any specialty care from a health maintenance organization, as well as preapproval for emergency department visits. Whether this will be beneficial or detrimental

to the homeless mentally ill remains to be seen.

Management Treatment in the emergency department should focus on the presenting problem, but follow-up is problematic. In recent years programs in several major cities have addressed the medical problems of the homeless by initiating contact with them on the street or in shelters, acknowledging their need for flexibility in outpatient care, and obtaining medical and other benefits on their behalf to improve access to care and to fill prescriptions. When such patients are able to obtain funds through welfare or other social programs and obtain housing, it is important that such artificial boundaries as catchment areas not interfere with the continuity of care they may have been receiving. There is some evidence that assertive case management improves the patient's long-term quality of life, decreases thought disorder, and increases appropriate interaction with others in the community.

Chronically Ill Young Adults These patients are often brought to emergency departments by families who have become frustrated by the cycle of improvement in the hospital, followed by decompensation when the patient is discharged and discontinues medication. Another large group of chronically ill young adults have lost contact with their families, live a somewhat nomadic existence, and like the homeless mentally ill struggle daily for survival, rather than for good medical care. Many have developed mental disorders due to a general medical condition (e.g., head trauma) and substance-induced disorders.

Management On each visit to the emergency department, the patient must be evaluated for any change in symptoms and for threats to life, such as suicidal ideation. Subsequently, good communication with the appropriate community mental health center, psychosocial rehabilitation program, day treatment program, or sheltered workshop can minimize inappropriate use of the emergency department. An assertive community treatment team and safe, supervised housing can also be helpful for this group of patients.

Elderly Patients Patients over 75 are over one and a half times more likely to seek treatment in an emergency department than the average patient. The usual geriatric patient is a single white female. Entry to the emergency department is more likely to be driven by the frustration of caregivers than by self-referral. In the emergency department the elderly use medical, social, and psychiatric resources disproportionately, and are more likely to need hospitalization because they are more likely to have multiple medical problems, a prior psychiatric history, be on multiple medications, have limited financial resources, and be socially isolated. The most common reason for bringing an elderly patient to the emergency department is unusual or disturbing behavior, which has been present for 6 months or more; the most frequent diagnosis is major depressive disorder. An often-overlooked additional factor to the deterioration of a geriatric patient's function is the overuse of alcohol. Adverse effects or complications from medication are also a frequent contribution.

Management Treatment is targeted to the patient's presenting problem. Because of the greater likelihood of medical illness, drug reaction, or combinations of these producing or contributing to the psychiatric picture, it is strongly recommended that elderly patients be evaluated for medical illness and substance-related causes whenever there is a behavioral change. The brief mental status evaluations referred to earlier should be incorporated into every evaluation of a geriatric patient. Because a change in environment may be confusing or even dangerous for an elderly patient who may have impaired vision and hearing, hospitalization should be avoided. A better alternative is brief, intensive treatment in the emergency department or a short-stay unit, perhaps combined with home visits by nursing, social services, or mobile treatment teams. Family consultation is useful. Coordination of treatment by a generalist, nurse, or social worker maximizes the quality of life for the elderly and for their families.

Children and Adolescents Unlike the modal adult patient in the emergency department, children and adolescents generally reside with one or more other members of a family, and it is commonly within the context of a family crisis that the emergency arises. When children or adolescents come to emergency departments unaccompanied, it is likely that physical or sexual abuse is occurring in the family. Over the past several years the suicide rate for adolescents has gone up, and suicide attempts, gestures, and ideation occur about twice as often as in an adult emergency population. Substance abuse and intoxication are frequent precipitants for emergency department visits.

Management Treatment is focused on the presenting problem. The problems of children and adolescents who present as emergencies are frequently attributable to family crises, and in those instances are best dealt with as system issues in the family. Unless it is obvious that serious physical or sexual abuse is occurring within the family, the child or adolescent patient should be treated like a soldier in wartime would be; rapid intervention as close to the “front line” as possible has the greatest chance of being effective, in terms of the immediate result and of longer-term general functioning. Removing a child from a family abruptly results in many long-term problems of stigmatization, loss of self-esteem, and subsequent difficulties in reintegrating with the family and the community. Adolescents who have separated from their families, emotionally if not physically, can be managed in their own right. This may require attention to problems such as locating housing, funding, and arranging follow-up care.

ETHICAL ISSUES

Because decisions in emergency psychiatry often need to be made quickly and with limited information, it is imperative that the physician considers life-threatening illness, the stigma associated with a diagnosis of mental illness and with hospitalization in a psychiatric facility, and the interests of the society at large. The person making these decisions must be free of institutional pressures, such as filling hospital beds or meeting financial goals. Legal considerations should not divert the psychiatrist from making a medically or psychiatrically appropriate recommendation. In general, there is a wide legal latitude when medical decisions are made in an emergency situation. Emergency psychiatry is often relegated to very junior faculty or relatively inexperienced psychiatry residents. In light of the weight of the decisions to be made and the complexity of many emergency patient presentations, there should be sufficient opportunity for supervision and consultation from experienced clinicians.

RESEARCH AND EVALUATION

Research to date has focused on demographics, decision-making, and the emergency use of medication. The prediction and management of violence, recidivism, comorbidity, and staffing have been addressed in some studies. Specific demographic populations, such as the elderly, children and adolescents, and the homeless have been the subjects of several studies. Current and future research directions include exploration of the resources of electronic information storage and transmission for rapid retrieval of updated information in the emergency setting while preserving confidentiality. Measurements of the outcome of emergency interventions and seeing what types of information improve the patient's outcome and the cost-effectiveness of various emergency interventions are also important areas for future research. Effective research in these areas will be enhanced by standardizing clinician variables in diagnosis, decision-making, and treatment. The changing face of health care delivery, with the privatization of medical assistance and veteran patient populations, will further drive research to look at cost-effective emergency treatments; because of capitation in general medical service provision it will be important to study utilization of primary care services for patients who access the emergency department. Further research is needed also on the cost and impact of mobile crisis intervention programs, continuous-care teams, and the use of newer neuroleptic agents in the emergency setting. For the dual-diagnosis patient, research needs to be conducted on integrated treatment efforts, and on such patient's recidivism, number of crises, and use of the emergency department.

Table 29.2-6 lists in alphabetical order common psychiatric emergencies.

Table 29.2-6 Common Psychiatric Emergencies

SUGGESTED CROSS-REFERENCES

Mood disorders are discussed in Chapter 14, panic disorders in Section 17.1, generalized anxiety disorder in Chapter 15, adjustment disorders in Chapter 23, and the psychiatric aspects of AIDS are discussed in Section 2.8. Adult antisocial behavior, criminality, and aggression are discussed in Section 27.3. The traditional psychiatric interview is described in Section 7.1; medical assessment and more exhaustive laboratory testing relevant to the psychiatric patient are covered in Section 7.7. Section 30.8 provides an in-depth discussion of brief dynamic and crisis therapy, and psychopharmacology pertinent to use in the emergency department is discussed in Chapter 31. More detailed information about the presentation and management of the disruptive child and adolescent patient may be found in Chapter 40. A thorough discussion of the geriatric patient may be found in Chapter 51, and special issues, including abuse of the elderly and medical problems that complicate or otherwise contribute to their psychiatric problems, are discussed in Section 51.6g and Section 51.3a. Physical abuse of the adult, including domestic violence, is discussed in Section 28.7. Chapter 11 covers all aspects of substance abuse. Chapter 52 discusses the economic and community aspects of treating the psychiatric patient.

SECTION REFERENCES

Allan MH: Definitive treatment in the psychiatric emergency service. Psychiatr Q *67:*247, 1996.

Bell CC, Jenkins EJ, Wolanyo K, Rhodes H: Response of emergency rooms to victims of interpersonal violence. Hosp Community Psychiatry *45:*142, 1994.

*Bengelsdorf H, Church JO, Kaye RA, Orlowski B, Alden DC: The cost effectiveness of crisis intervention. J Nervous Ment Disorder *181:*757, 1993.

Borum R, Deane MW, Steadman HJ, Morrisey J: Police perspectives on responding to mentally ill people in crisis: Perceptions of program effectiveness. Behav Sci Law *16:*393, 1998.

*Bremer JD: Acute and chronic responses to psychological trauma: Where do we go from here? Am J Psychiatry *156:*349, 1999.

Citrone L, Volavka J: Psychopharmacology of violence, Part I: Assessment and acute treatment. Psychiatr Ann *27:*691, 1997.

*Colenda CC, Greenwald BS, Crossett JH, Husain MM, Kennedy GJ: Barriers to effective psychiatric emergency services for elderly persons. Psychiatr Serv *48:*321, 1997.

Claassen CA, Gilfillan S, Orsulak P, et al: Substance abuse among patients with a psychotic disorder in a psychiatric emergency room. Psychiatr Serv *48:*353, 1997.

*Crenshaw WB, Cain KA, Francis PS: An updated national survey on seclusion and restraint. Psychiatr Serv *48:*395, 1997.

Cuffel B: Disruptive behavior and the determinants of costs in the public mental health system. Psych Serv *48:*1562, 1997.

*Fauman MA, Fauman BJ: The differential diagnosis of organic-based psychiatric disturbance in the emergency department. JACEP *6*:315, 1977.

Fichtner OG, Flaherty JA: Emergency psychiatry training and the decision to hospitalize. Acad Psychiatry *17:*130, 1993.

Fisher WA: Restraint and seclusion: A review of the literature. Am J Psychiatry *151:*1584, 1994.

Fisher WH, Geller JL, Altaffer F, Bennett MB: The relationship between community resources and state hospital recidivism. Am J Psychiatry *149:*385, 1992.

Foa E: Trauma and women: Course, predictors, and treatment. J Clin Psychiatry *58:*25, 1997.

Folstein MF, Folstein SE, McHugh PR: “Mini-mental state” a practical method for grading the cognitive state of patients for the clinician. J Psychiat Res *12:*189, 1975.

*Fortney JC, Owen R, Clothier J: Impact of travel distance on the disposition of patients presenting for emergency psychiatric care. J Behav Health Serv Res *26:*104, 1999.

Gilfillan S, Claassen CA, Orsulak P, Carmody TJ, Sweeney JB, Battaglia J, Rush AJ: A comparison of psychotic and nonpsychotic substance abusers in the psychiatric emergency room. Psychiatr Serv *49:*825, 1998.

Goenjian AK, Karayan I, Pynoos RS, Minassian D, Najarian LM, Steinberg AM, Fairbanks LA: Outcome of psychotherapy among early adolescents after trauma. Am J Psychiatry *154:*536, 1997.

Gustafson DH, Sainfort F, Johnson SW, et al: Measuring quality of care in psychiatric emergencies: Construction and evaluation of a Bayesian index. Health Serv Res *28:*131, 1993.

*Haywood TW, Kravitz HM, Grossman LS, Cavanaugh JL, Davis JM, Lewis DA: Predicting the “revolving door” phenomenon among patients with schizophrenic, schizoaffective, and affective disorders. Am J Psychiatry *152:*856, 1995.

Hillard JR: Emergency psychiatry: The past and future of psychiatric emergency services in the U.S. Hosp Community Psychiatry *45:*541, 1994.

Hughes DH: Implications of recent court rulings for crisis and psychiatric emergency services. Psychiatr Serv *47:*1332, 1996.

Kahn RL, Goldfarb AL, Pollack M, Peck A: Brief objective measures for the determination of mental status in the aged. Am J Psychiatry *117:*326, 1960.

Kaplan HI, Sadock BJ, editors: *Pocket Handbook of Emergency Psychiatric Medicine*. Williams & Wilkins, Baltimore, 1993.

Lindemann E: Symptomatology and management of acute grief. Am J Psychiatry *101:*141, 1944.

*Mann J, Waternaux C, Haas GL, and Malone KM: Toward a clinical model of suicidal behavior in psychiatric patients. Am J Psychiatry *156:*181, 1999.

Morse G, Calsyn R, Klinkenberg W, Trusty M, Gerber F, Smith R, Tempelhoff B, Ahmad L: An experimental comparison of three types of case management for homeless mentally ill persons. Psychiatr Serv *148:*497, 1997.

Mulvey EP, Lidz CW: Measuring patient violence in dangerousness research. Law Hum Behav *17:*277, 1993.

*Pfefferbaum B, Flynn BW, Brandt EN, Lensgraf SJ: Organizing the mental health response to human-caused community disasters with reference to the Oklahoma City bombing. Psychiatr Ann *29:*109, 1999.

Rhodes LA: *Emptying Beds: The Work of an Emergency Psychiatric Unit*. University of California Press, Berkeley, CA, 1991.

Slagg NB: Characteristics of emergency room patients that predict hospitalization or disposition to alternative treatments. Hosp Community Psychiatry *44:*252, 1993.

Solomon P, Gordon B: Outpatient compliance of psychiatric emergency room patients by presenting problems. Psychiatr Q *59:*271, 1988.

P>Sullivan PF, Bulik CM, Forman SD, Mezzich JE: Characteristics of repeat users of a psychiatric emergency service. Hosp Community Psychiatry *44:*376, 1993.

Szuster RR, Schanbacher BL, McCann SC: Characteristics of psychiatric emergency room patients with alcohol- or drug-induced disorders. Hosp Community Psychiatry *41:*1342, 1990.

Torrey EF: Violent behavior by individuals with serious mental illness. Hosp Community Psychiatry *45:*653, 1994.

Weissberg M: Chained in the emergency department: The new asylum for the poor. Hosp Community Psychiatry *42:*317, 1991.

CHAPTER 30. PSYCHOTHERAPIES

30.1 PSYCHOANALYSIS AND PSYCHOANALYTIC PSYCHOTHERAPY

GLEN O. GABBARD, M.D.

Most psychotherapy conducted in the United States derives from psychoanalytic principles. Formal psychoanalysis, usually involving four or five sessions a week with the patient reclining on a couch, is taught in psychoanalytic institutes throughout the country. During his life, Sigmund Freud (1856–1939), the founder of psychoanalysis (Fig. 30.1-1), predicted that the principles of the "talking cure" would ultimately be adapted into forms of psychotherapy. With some trepidation, he recognized that the "pure gold" offered by psychoanalysis would ultimately be "alloyed with the copper of suggestion."

FIGURE 30.1-1 Sigmund Freud in his Vienna office. (Etching by Max Pollak. Courtesy of Menninger Foundation Archives, Topeka, KS.)

Developments in the field since Freud's death suggest that his trepidation was perhaps unwarranted. Psychoanalytic (sometimes called psychodynamic) psychotherapies have been demonstrated to be highly effective and useful treatments for a wide range of disorders. Recent surveys suggest that psychoanalytic psychotherapy is still the major form of psychotherapy taught in training programs and the form of therapy that mental health professionals choose when they seek their own treatment. This modality usually takes place anywhere from one to three times per week, with the patient and therapist sitting face to face. This type of psychotherapy goes by a variety of names, including expressive psychotherapy, expressive-supportive psychotherapy, insight-oriented psychotherapy, intensive psychotherapy, and exploratory psychotherapy. In addition, a variation known as psychoanalytically or psychodynamically informed supportive psychotherapy is also widely practiced. All three of these modalities are considered in this section. Freud and some of his collaborators are pictured in Figure 30.1-2.

FIGURE 30.1-2 Freud with several of his early collaborators. *Top, left to right:* Otto Rank, Karl Abraham, Max Eitingon, Ernest Jones; *bottom, left to right:* Freud, Sandor Ferenczi, Hanns Sachs. (Courtesy of Louis Linn, M.D.)

HISTORY OF THE PSYCHOANALYTIC METHOD

Psychoanalysis is a treatment derived from psychoanalytic theory, the most comprehensive conceptual model of mental life available. It is based on the notion that early childhood experiences are internalized and, in concert with biological urges from within, form a dynamic unconscious that influences behavior, thoughts, and feelings. The term *psychoanalysis* is used to refer to a theory of the mind, a procedure or method of inquiry, and a type of treatment. The last of these three uses is addressed in this section.

Evolution of Freud's Technique Freud's early use of hypnosis was profoundly influenced by his friend and older colleague, Josef Breuer, a Viennese internist (Fig. 30.1-3). In his work with Anna O., Breuer had discovered that when his patient was placed in a hypnotic trance, her symptoms disappeared as a consequence of their verbal expression. Anna O. dubbed this treatment the talking cure. Freud felt he should be more fully trained in hypnotic techniques, so he spent the year 1885 to 1886 with the French neurologist Jean-Martin Charcot at the Salpetrière in Paris. When Freud opened his practice back in Vienna in 1887, he used hypnosis to remove symptoms through suggestion. His results with this technique were less than encouraging, so he shifted to the cathartic method based on Breuer's account of his treatment of Anna O.

FIGURE 30.1-3 Josef Breuer.

Freud's use of this method led to a coauthored work on hysteria in the mid-1890s, called *Studies in Hysteria*. In this contribution Breuer and Freud discussed their theory that repressed ideas outside of conscious awareness were transformed or converted into symptoms. Through a process of recovering and verbalizing affects with which these repressed ideas were associated, they sought to remove symptoms through abreaction.

In the famous case of Frau Emmy von N., Freud attempted to use the cathartic method in conjunction with hypnosis to identify the etiological origins of hysterical symptomatology. She had consulted him for a variety of hysterical complaints, including phobias, delirium, hallucinations, anesthesia and pain in her leg, and an ovarian neuralgia. He suspected that inhibited sexuality played a role in causing her illness. However, he felt the limitations of his cathartic method prevented fuller exploration.

Freud's abreactive strategy was limited; therapeutic benefits did not persist after the physician-patient contact was terminated. He began to recognize that the personal relationship with the physician might be more important to the therapeutic action of the treatment than the specialized hypnotic techniques he had developed. This hypothesis was further confirmed when one of his female patients woke up from a hypnotic trance and threw her arms around him. This incident led Freud to believe that erotic attachment to the physician might potentially be a major therapeutic force in psychoanalytic treatment.

These observations led Freud to his discovery of transference, which remains a key principle of psychoanalytic theory and technique today. *Transference* is generally defined as displacement onto the analyst of the thoughts, feelings, and behavior originally associated with significant figures from the past. He also had an inkling of the dark side of transference in a passage from *Studies in Hysteria*, where he noted that the patient might be “seized by a dread of becoming too much accustomed to the physician personally, of losing her independence in relation to him, and even of perhaps becoming sexually dependent on him.”

Freud's discovery of transference led him to abandon hypnosis, which he viewed as concealing aspects of the transference. In addition, Freud noted that many patients were not amenable to hypnosis, so he developed the *concentration method* to replace it. This technique was developed during the treatment of Elizabeth von R., who was quite resistant to the abreactive technique. Recalling an observation by Hippolyte Bernheim, a hypnotist who was an adversary of Charcot, Freud believed that forgotten memories could be retrieved when the physician asked the appropriate leading questions. Thus he asked Elizabeth von R. to lie down on the couch and close her eyes. Freud pressed his hand on her forehead and told her that she would recall relevant memories when he questioned her about them.

This approach, in turn, led to Freud's discovery of free association, which continues to be a major principle of psychoanalytic technique in modern practice. Freud noted that his interruptions of Elizabeth von R. and his constant injunctions to remember interfered with her spontaneous flow of thought. Elizabeth von R. informed Freud that she could have told him her thoughts all along but she was not sure he wanted to hear them. This comment showed Freud that his own forceful direction had led the patient to censor her thoughts. As a result, he began instructing his patients to say whatever came to mind, without censoring their thoughts. No matter how unpleasant, unacceptable, or seemingly trivial an association seemed, Freud was convinced that these thoughts should be verbalized and understood. At the same time he abandoned his technique of touching the forehead, allowing the patient to lie on a couch and free associate in an unfettered and nondirected way.

These clinical investigations led to two other fundamental psychoanalytic concepts. Freud noted that a number of patients could not be hypnotized, while others could not recall memories or free associate. Freud labeled this unconscious thwarting of the therapeutic technique *resistance*. He then recognized that some patients who were extremely distressed were also resistant to treatment. Freud concluded that active unconscious forces were excluding unacceptable or unpleasant thoughts as well as distressing memories from conscious awareness. Freud called this active force *repression*.

Freud considered repression crucial to the process of symptom formation. He believed that a traumatic experience or series of experiences, often sexual in nature, occurred in the childhood of neurotic patients. These memories were then repressed because they were painful and unacceptable, but the excitement associated with the traumatic events could not be entirely extinguished, and traces persisted in the unconscious as repressed memories. This original excitement eventually is triggered by a contemporary event, such as a disturbing love affair, leading to the return of the repressed in the form of a neurotic symptom.

Post-Freud: Early Twentieth Century to the Present The history and growth of the psychoanalytic movement from classic psychoanalysis to current practices were marked by repeated revisions by Freud himself, as well as by reappraisals and rebellions by neo-Freudian and non-Freudian contemporaries and descendants, as chronically manifest in internecine conflicts over diverse and intricate points of theory and practice (Table 30.1-1). From the beginning, as described by T. Byram Karasu, some analysts have argued against nearly all of Freud's basic conceptual premises, from the strictly sexual etiology of the neuroses to his views on female psychology. Others have reacted more directly to their clinical concerns that too few patients meet the rigorous requirements associated with orthodox psychoanalysis, and they have attempted to make the treatment more widely applicable, affordable, and terminable.

Table 30.1-1 Historical Development of Psychoanalytic Psychotherapy

In the 1920s and 1930s such analysts as Sandor Ferenczi, Otto Rank, and Wilhelm Stekel tried to broaden the applicability of psychoanalysis to a larger clinical spectrum by shortening treatment time (Rank was the first to propose an end-setting time limit) and emphasizing a more active, affective, caretaking approach, particularly with the use of the therapist as a substitute primary object in the treatment of young children. Stekel's Institut für Aktive Analyse (Institute of Active Analysis) was the first organized attempt at psychoanalytically based brief psychotherapy. Those technical innovations culminated in the establishment in the late 1940s of the school of Franz Alexander and Thomas French, the first group to experiment directly with influencing the transference by varying the frequency of sessions and to emphasize face-to-face interviews. Alexander and French defined the major role of the analyst as providing a corrective emotional experience to compensate for early developmental experiences by being different from the patient's parents. However, those therapies soon sparked a major controversy over whether they could still be considered psychoanalysts. Threatened by the widening scope of psychoanalytic technique, the conservative therapists claimed that any departure from a strictly neutral-interpretive stance would drastically compromise treatment.

Secession From Freud With the widespread dissemination of psychodynamic concepts and applications after World War II, a major turning point in psychoanalytic development occurred in the 1950s. As the practice of psychoanalysis broadened to include a great variety of clinical conditions, especially with the extension of classical analysis to psychoses and character disorders, analysts were obliged to alter both their strategies and the theories they used to explain the changes in perspective. Initially, especially in the earlier era of postwar psychiatry, psychoanalysis was applied almost indiscriminantly to a wide range of mental disorders reflecting every degree of severity of illness; during that experimental period analysts began more actively to recognize the limitations of standard methods.

Those critical years saw the expansion in interpersonal, cultural, and phenomenological directions of the psychoanalytic movement. European and American analytic theorists—including Alfred Adler, Otto Rank, Carl Jung, Erich Fromm, Karen Horney, Harry Stack Sullivan, and Medard Boss—separately seceded from Freud (some to form their own schools) and wrought sufficient theoretical changes to challenge and extend the early boundaries of psychoanalysis.

Karasu points out that Horney's rejection of Freud's libido theory of neurosis in favor of a more social orientation led to an interpersonal approach in the analysis of neurotic patients that drew less distinction between analysis and psychotherapy, gave the analyst a more active role, and dispensed with free association and the couch. Sullivan's pioneering work with a population of schizophrenic patients drew attention to their distorted parataxic thinking (illogical causal connections between concurrent events) in adult interpersonal relationships, and he influenced treatment techniques by defining the face-to-face psychiatric interview that fostered the role

of the analyst as participant-observer. Within the framework of existential analysis (e.g., Boss' *Daseinanalysis*) the role of the analyst was further altered by emphasizing the real, here-and-now encounter. Others, who were engaged in the treatment of patients with character problems, began to acknowledge the increasing need to enlarge the scope of treatment by introducing “parameters” (Kurt Eissler's term) that would help such patients in analysis. That technical direction was influenced by the growing effects of ego psychology on the understanding of ego defects and associated problems. Wilhelm Reich's analysis of character armor, as well as Otto Fenichel's investigations of problems of psychoanalytic technique, expanded the analysis of resistances in determining pathological character traits.

Sigmund Freud's daughter Anna had a major influence in the evolution of ego psychology. In her famous observation that “there is depth in the surface,” she stressed that much of the analyst's work involves analyzing defenses against drive derivatives. These defenses typically manifest themselves as resistances in the analytic process. Anna Freud stressed that analysts tended to overvalue a search for deeply repressed unconscious drive derivatives while not recognizing the complexity of the patient's characteristic defenses. Her father had primarily emphasized repression in his discussion of defenses, but Anna Freud elaborated on his work by providing detailed descriptions of a number of additional defense mechanisms, including regression, undoing, identification, projection, introjection, reaction formation, sublimation, reversal, and turning against the self.

Heinz Hartmann further expanded ego psychology by focusing on adaptation. He stressed that a line of ego development existed independent of instinctual drives. Certain primary autonomous ego functions, such as perception, motility, memory, and intelligence, were thought to arise from a conflict-free sphere of the ego. Edith Jacobson integrated ego psychology with object relations theory. Her views were later supported by Margaret Mahler's research on infants and their mothers during the separation-individuation stages of child development.

As ego psychology continued to develop, a number of clinicians felt that indications for analysis could be broadened to include more difficult-to-treat patients. Edward Bibring, Leo Stone, and Eissler were instrumental in defining the margins of the “widening scope” of psychoanalytic technique. In addition, Merton Gill emphasized the crucial importance of analyzing transference in the here and now.

Interest also grew in the real relationship between analyst and patient. Elizabeth Zetzel and Ralph Greenson argued for a nontransferential element in the relationship between analyst and patient, often referred to as the *therapeutic alliance* or *working alliance*. This construct referred to the relatively nonconflictual aspects of the collaboration between analyst and patient toward commonly held treatment goals.

Parallel to these developments of ego psychology in America, object relations theory was evolving in the United Kingdom. Whereas classical Freudian theory from which ego psychology was derived emphasized drive discharge as the most important motivator for the infant, the British object relations theorists viewed object seeking as the fundamental motivation. However, the original object relations theory evolved from the work of Melanie Klein, who linked internal object relations to drives. Relying heavily on the death instinct for her theory, she believed that the infant's fear of annihilation from the death instinct was responsible for a defensive splitting process early in life. Derivatives of the death instinct, such as hatred, sadism, aggression, and envy, are disavowed by the infant and projected into the mother. This internal splitting process leaves the “goodness” within the infant, while the mother becomes a terrifying “bad object” experienced by the infant as persecutory. This paranoid-schizoid position is ultimately resolved when the infant recognizes that the “bad mother” and the “good mother” are the same person. At this point the split-off bad and good qualities are integrated into an ambivalently held whole object. This integration marks the entry into the depressive position, in which children worry that their own destructiveness might damage the mother.

Influenced by Klein, Wilfred Bion emphasized the role of the mother in containment of affects that the infant found difficult to manage. In his view, psychotic anxieties of infancy are metabolized and detoxified by the mother and returned to the infant in a more tolerable form. Through this process of projection and introjection, the infant eventually builds internal structure.

Influenced by Klein and Bion, the so-called independent, or middle group, of object relations theorists evolved in a direction substantially removed from drives. This group, composed of W. R. D. Fairbairn, D.W. Winnicott, Michael Balint, and others, emphasized a theory of deficit rather than conflict in Kleinian theory. In their view insufficient maternal nurturance created what Balint called the basic fault, a feeling in many patients that something is missing. Winnicott felt that just as the mother provided a holding environment to help the child develop awareness of separateness, the analyst needed to provide a holding environment for the patient.

Personality Disorders The patients described by Sigmund Freud generally fell into the category of symptomatic neurotics. Over time psychoanalysts began to see fewer and fewer patients with hysterical, phobic, or anxiety symptoms and more patients who exhibited character neuroses. These patients presented generalized complaints about problems in living, such as difficulties at work or in intimate relationships. The analysis of character began to take precedence over the analysis of a neurotic symptom.

In the late 1960s and early 1970s, a burgeoning literature developed on narcissistic and borderline personality disorders. As a result, the term *character neurosis* became subsumed under a broader category of personality disorders. Character neuroses represent the neurotically organized variants such as obsessive-compulsive and hysterical personality disorders. Narcissistic and borderline personality disorders, representing much of the spectrum commonly seen by psychoanalysts and psychoanalytic therapists, reflect a level of ego organization intermediate between neurosis and psychosis. Otto Kernberg was extremely influential in his systematic descriptions of borderline personality organization. He delineated the following structural characteristics: (1) nonspecific manifestations of ego weakness, including poor anxiety tolerance, poor impulse control, and poorly developed subliminatory channels; (2) a shift toward primary process thinking, particularly under the pressure of strong affects or in unstructured situations, such as projective psychological testing; (3) a pattern of specific defensive operations, including splitting, projective identification, primitive idealization, denial, omnipotence, and devaluation; (4) pathological internalized object relations based on splitting, so that representations of self and others are alternatingly all good or all bad.

Kernberg regarded narcissistic personality disorder as a variant of the borderline condition. He felt that these patients had the same underlying structural organization as the borderline patient but smoother ego functioning related to an integrated and pathologically grandiose self. Heinz Kohut, on the other hand, viewed narcissistic patients as developmentally arrested because of parental failures of empathy in response to the child's phase-appropriate needs for idealization and mirroring. His ideas evolved into self psychology, which involves a different developmental line than classical theory. Kohut regarded narcissistic personality disorders as requiring selfobject responses from others. In other words, these patients require others to perform certain functions for them to make them feel complete. In this regard, the other person is used as an extension of the patient's self. In the psychoanalytic situation, these needs develop into selfobject transferences of mirroring and idealizing. Kohut believed that the analyst must create an empathic atmosphere to repair the damage from empathic failures in parents. He also stressed a process of transmuting internalization in which the patient gradually internalizes characteristics of the analyst as a way of strengthening the patient's self.

Short-Term (Brief) Therapy The emergence of a variety of short-term (brief) psychodynamic psychotherapies was also a hallmark of the 1970s and 1980s. They are now primary, preferred modalities in their own right. Like other forms of psychoanalytic psychotherapy, they meet several contemporary needs: socioeconomic constraints that make protracted and costlier treatment less viable, the wish to serve larger numbers of patients, and the current pressures for accountability (which calls for systematic studies of psychotherapy outcomes based on specific formats and circumscribed time frames). As a further attraction the earlier fears of symptom substitution as an inevitable consequence of short-term treatment have essentially been quelled, allowing the enduring strengths of the Freudian legacy to be retained as the foundation for newer, more expedient techniques.

Peter Sifneos's anxiety-provoking psychotherapy, David Malan's intensive brief psychotherapy, Habib Davanloo's broad-focus short-term dynamic psychotherapy, James Mann's time-limited psychotherapy, and Lewis Wolberg's eclectic short-term psychotherapy are examples of the range of brief treatments that derive from the psychoanalytic approach. These combined interpretive and noninterpretive techniques, active approaches, and abbreviated time requirements are all accepted features of current psychoanalytic psychotherapy.

Operational guides for psychoanalytic practice have been developed. Hans Strupp and Jeffrey Binder's time-limited dynamic psychotherapy (TLDP) and Lester Luborsky's principles of psychoanalytic psychotherapy, a manual for supportive-expressive treatment, are two examples of the systematic approach to psychodynamic treatment.

A major theoretical trend in recent years that has profoundly affected psychoanalytic technique is the growing recognition that psychoanalysis is both a two-person psychology and a one-person psychology. The image of the analyst as a detached “blank screen” figure dispassionately observing the patient is regarded largely as a mythical construct that is misleading in terms of the realities of psychoanalytic work. In fact, the analyst's own subjectivity has a continuing influence on the patient, just as the patient's subjectivity influences the analyst's perceptions in an ongoing way. This recognition of two psychologies in mutual interaction with one another is often subsumed under the term *intersubjectivity*.

Social Constructivism The perspective of Gill and his collaborator Irwin Hoffman, often referred to as *social constructivism,* has been highly influential in the evolution of the notion of intersubjectivity. Central to the social constructivist view is the notion that the analyst's real behavior significantly affects the shape of the patient's transference. Also, analysts are considered unable to grasp fully how their irreducible subjectivity influences the patient in the analytic process. This point of view owes a debt to the constructivist view within philosophy, which suggests that reality is to some extent a construction influenced by the observer's subjectivity rather than simply an objective set of facts “out there” that are passively recorded by the individual's perceptual apparatus. Hence two individuals are present in the psychoanalytic consulting room, and together they construct a mutually negotiated reality. Another basic tenet of this perspective is that the patient may see something that the analyst is resisting just as the analyst observes resistance in the patient. Consequently, the analyst needs to take seriously the patient's interpretations of the analyst's conscious and unconscious influences on the patient. The social constructivist view emphasizes both spontaneity and personal engagement with the patient that fully acknowledges the contributions of the analyst to the transference.

Analysts who practice from this perspective do not deny that an intrapsychic world of the patient exists that is a valid focus of the psychoanalytic process. The patient arrives at the analysis with a preexisting set of conflicts, internal object relations, and self-structures. There is still a collaborative search for specific truths about the patient's unconscious that forms the basis of psychoanalytic treatment. However, there is now an awareness that the analysand's unconscious is perceived by, and filtered through, the analyst's subjectivity.

Countertransference This emphasis on social constructivism, intersubjectivity, and two-person aspects of the psychoanalytic process occurred in parallel with an increasing awareness of the role played by countertransference in psychoanalytic theory and technique. Terms such as projective identification, role responsiveness, and countertransference enactment have entered common usage and reflect a convergence in thinking among diverse psychoanalytic groups.

Projective Identification Projective identification originated in the work of Klein, who used the term to describe an intrapsychic fantasy of evacuating part of the self and putting it in someone else. Subsequent Kleinian thinkers have expanded the intrapsychic fantasy to an interpersonal process. Bion, for example, explicitly stressed that the analyst actually feels coerced by the patient into playing a role in the patient's fantasy. Thomas Ogden has been the most significant contributor to the contemporary usage of projective identification. In his early writings he described three aspects of the process: (1) an aspect of the patient's self is projectively disavowed by unconsciously placing it in the analyst; (2) the patient then exerts interpersonal pressure that coerces the analyst to experience or identify unconsciously with what has been projected; and (3) the analyst processes and contains the projected contents, leading them to be reintrojected by the patient in a modified form. As his thinking evolved, Ogden stressed that these aspects should not be construed as a rigid, linear sequence of steps, but rather as creating a dialectic in which analyst and patient enter into a relationship. In this relationship, the two members of the analytic dyad are simultaneously “at one” with each other and separate.

Ogden, then, provided a bridge between projective identification and intersubjectivity. Projective identification serves to create an interpersonally decentered subject, what he calls an *analytic third,* created by the two subjectivities. Ogden argues that the totalistic view of countertransference as referring to everything the analyst thinks or feels is not meaningful. He believes that countertransference must be viewed as constituting a dialectic between the analyst as a separate entity and the analyst as a joint creation of the intersubjectivity of the analytic process. He stresses that actually three subjectivities are involved in psychoanalytic work: the subjectivity of the analyst, the subjectivity of the analysand, and that of the analytic third. Projective identification, then, occurs as a mutual process in both parties and creates a newly integrated “subject” in the interpretive space between analysand and analyst.

The patient should not be “blamed” for all the feelings experienced by the analyst. Preexisting feelings in the analyst may interact with the feelings projected by the patient, and ongoing psychological work by the analyst is necessary to differentiate feelings that originate in the analyst from those evoked by the patient.

Patients do not project into an empty container. The process of projective identification requires a “hook” in the internal world of the analyst. Some projections are a better “fit” with the preexisting intrapsychic conflicts and representations within the analyst. Projective identification may feel like an alien force taking control of the analyst, and the analyst's subjective experience may be something like, “I'm not acting like myself.” However, the interaction can best be understood as the activation of repressed self- or object representations within the analyst that have been activated by interpersonal pressure from the patient. Hence the analyst's usual sense of conscious identity is disrupted by the unexpected emergence of these repressed aspects of the analyst's internal world. If a good fit exists between the analyst's internal world and that of the patient, the analyst becomes engaged as a participant in the externalization of the patient's internal object relationship.

Heinrich Racker found it useful to subdivide the analyst's countertransference reactions into concordant and complementary. *Concordant countertransference reactions* involve a situation in which the analyst identified with a self-representation within the patient. *Complementary countertransference* implies that the analyst has identified with a projected internal object representation of the patient. Racker viewed this as consistent with the projective identification model in which the analyst's own conflicts are activated by the patient's projections.

Role Responsiveness Joseph Sandler coined the term *role responsiveness* to describe a phenomenon very similar to projective identification. He made the following observation:

> Very often the irrational response of the analyst, which his professional conscience leads him to see entirely as a blind spot of his own, may sometimes be usefully regarded as a compromise formation between his own tendencies and his *reflexive acceptance of the role which the patient is forcing on him*.

Sandler regarded this process as one in which a patient unconsciously actualizes an internalized object relationship in the transference to the analyst, and the analyst then feels compelled to assume a role derived from the patient's intrapsychic world. He regarded this form of identification as fundamentally similar to Racker's notion of complementary countertransference. Sandler distinguished it from a process of *primary identification,* an automatic mirroring process that underlies analytic empathy. He provides an everyday example of this phenomenon. If one is following a man down the street who suddenly stumbles and leans over to the right to brace his fall, one has a natural tendency to lean slightly to the right as well.

Sandler underscored that one must be wary of assuming a one-to-one correspondence between what goes on in the analyst's mind and what resides in the patient's mind. He said it is erroneous to assume that any intense emotional reaction by the analyst to the patient's words or behavior is most likely projective identification. He argued that the term should be reserved for instances in which the patient unconsciously intends to evoke a specific emotional reaction in the analyst. In other words, he considered it erroneous to assume that all intense countertransference feelings are induced by the patient.

Countertransference Enactment The emphasis on the two-person aspects of psychoanalysis has entered into the literature of American ego psychological analysts through the concept of countertransference enactment. Theodore Jacobs introduced the term as a reference to subtle examples of interlocking transference-countertransference phenomena that operate unconsciously. These enactments may be as subtle as a changed body posture or increased muscle tension. They also may be more flagrant, (e.g., when analysts blurt out something that they wish or that they regret). James McLaughlin observed that the roots of the word *enactment* involve the notion of simulating or playing a part. He argued that both analyst and patient are constantly reacting to each other's behavior. Judith Chused defined countertransference enactment as an instance in which a transference fantasy is actualized and evokes a countertransference reaction in the analyst.

The similarities between Chused's definition and the contemporary view of projective identification are striking. Chused emphasized that individual variations in the analyst might well result in different countertransference enactments in different analysts, even if they treated the same patient.

In general, the way that American ego psychological analysts have written about countertransference enactments suggests a greater emphasis on the narrow view of countertransference. In other words, like Freud, they often conceptualize their emotional reactions to their patients as reflecting experiences from their own past that are revived in the interaction with the patient. Nonetheless, most contemporary analysts affiliated with ego psychology would agree that the analyst's countertransference may convey important information about the patient's inner world.

A common ground has emerged in recent years in the understanding of countertransference. Both the contemporary use of projective identification among analysts influenced by Klein and the British School of object relations and the use of countertransference enactment by ego psychological analysts converge around a view that the analyst's countertransference is a joint creation by both patient and analyst. In other words, the responses evoked in the analyst by the analysand work in concert with the analyst's own internal self and object relations to determine the final shape of the countertransference response.

The wide acceptance of the inevitability of various countertransference enactments in the course of an analysis have led psychoanalytic clinicians to recognize that such enactments can be analyzed and understood in a way that advances the analytic work. Countertransference awareness may only emerge after some

countertransference has been enacted. Owen Renik has argued that analysts must allow for spontaneity so that attenuated enactments can be analyzed and understood within the process. Similarly, members of the American relational school, such as Stephen Mitchell and Lewis Aron, have stressed that the analyst cannot avoid becoming embedded within the analysand's relational matrix. Analysts will play assigned roles even if they desperately attempt to stand outside the patient's internal cast of characters.

Pluralism Another trend in contemporary psychoanalytic discourse is a movement toward pluralism. The hegemony of ego psychology in American psychoanalysis has given way to a point where many analysts regard themselves as eclectic. The way analysts work in the privacy of their consulting rooms often differs rather dramatically from their publicly held theoretical positions. Most analysts probably borrow useful aspects of a variety of theories and develop their own integration with a patient whose psychopathology is multifaceted. Fred Pine has illustrated a sophisticated synthesis of four different theoretical perspectives: drive theory, ego psychology, object relations theory, and the psychology of the self. He stresses that one does not have to adopt the entire edifice of theory to borrow useful aspects of different perspectives that address different clinical phenomena. One theory is unlikely to address all of the complexities one encounters in the clinical situation.

In concert with the new emphasis on pluralism has been a movement away from a number of traditional polarizations within the field of psychoanalysis. These include one-person versus two-person psychology, intrapsychic versus interpersonal domains, drive theory versus relational perspectives, and oedipal versus preoedipal understanding. Many analysts recognize that a dialectical tension must exist between these apparent polarities and that both poles are necessary for a complete understanding of the human psyche.

Karasu has constructed a synthetic approach that attempts to locate generic factors in all psychotherapy. From this overarching principle, he has evolved a developmentalist metatheory that tries to bring deficit and conflict viewpoints under one umbrella. The conflicts and deficits in adult psychopathology are viewed as having origins in different phases of development. In this regard, he has constructed a fourfold matrix that includes dyadic deficit, dyadic conflict, triadic deficit, and triadic conflict. This matrix can be applied to a variety of psychiatric disorders encountered by the analyst or psychoanalytic therapist as a conceptual model for understanding the developmental origins of the difficulties.

Another form of pluralism is the attempt to integrate psychoanalytic thinking with phenomenological and philosophical perspectives. Richard Chessick, for example, has questioned the implicit assumptions upon which the understanding of patients rest. He has incorporated the views of Jean-Paul Sartre, Jacque Lacan, Martin Heidegger, and others who have tried to bring traditional philosophical perspectives to psychoanalytic data.

PSYCHOANALYTIC TREATMENT

Indications and Analyzability Determining which patients are suitable for psychoanalysis has undergone an evolution that parallels the historical changes in theory and technique. The original indications for psychoanalytic treatment were relatively straightforward—only those patients suffering from symptomatic neuroses of a hysterical, obsessive-compulsive, or phobic nature were suitable for psychoanalytic treatment. Freud also believed that as patients age, their lack of psychological malleability makes them less amenable to psychoanalysis. Hence only young patients, generally those in their 20s and 30s, were viewed as ideal candidates for analysis.

Those rather narrow indications are now considered limiting if not archaic. Few patients present to clinicians with complaints of discrete neurotic symptoms as they did in Freud's day. Manifestations of pervasive character pathology, such as difficulties in a patient's ability to love and to work, are likely to be the presenting complaints today. The changes in the conditions that patients bring to psychoanalysis have been mirrored in the advances in technique that are based on developments in self psychology and object relations theory. It is now common to refer to the *widening scope* of psychoanalytic treatment in the context of the expanded indications for such treatment.

Moreover, the age restrictions advocated by Freud have largely been discounted, as increasing research and clinical data attest to the malleability of middle-aged and elderly patients. Adult developmental studies indicate that significant changes in defenses, internal object relations, and self-representations occur throughout the adult life cycle. Analysts now commonly treat patients in their 50s, 60s, and even 70s.

Conditions considered amenable to analysis today include, in addition to the symptomatic neuroses, certain anxiety disorders, highly perfectionistic depressed individuals, some sexual disorders, certain personality disorders, including obsessive-compulsive, histrionic, avoidant, and narcissistic, selected patients at the upper end of the spectrum of borderline personality disorder, self-defeating personality disorder, and many cases of mixed personality disorder. In addition, many patients who do not fit the fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV) categories may experience significant distress and frequently seek psychoanalytic treatment. Common problems that bring such individuals to analysis include difficulties with intimacy, relatedness, and commitment; lack of assertiveness; avoidant tendencies; self-defeating behavior; problems with authority; shyness; unresolved grief; or problems related to separation or rejection.

The preceding list of indications for psychoanalysis is incomplete, however, without also considering generalized criteria for analyzability that apply regardless of the diagnosis. Significant suffering must be present, so that the patient is motivated to sacrifice the time and financial resources required for psychoanalysis. Also, patients who enter analysis must have a genuine wish to understand themselves, rather than a desperate hunger for symptomatic relief. They must be able to withstand the frustration, anxiety, and other strong affects that emerge in analysis without fleeing or acting out their feelings in a self-destructive manner. Other ego functions that must be intact, in addition to the ability to delay impulses and control affects, are reality testing and self/object differentiation. Without reasonably good ego functioning in those areas, a stable, workable transference will not be established.

Reasonably mature superego development is also necessary for a patient to be analyzable. Patients who are pathological liars, who are morally corrupt, or who commit criminal acts are rarely suitable for psychoanalytic treatment. At least average, if not superior, intelligence is required for analytic work, as is a well-developed capacity to verbalize internal experiences. The patient must have some capacity to regress in the service of the ego but also an ability to reinstitute defenses by the end of the session. One other important criterion of analyzability is a form of thinking that is usually referred to as *psychological mindedness*. Patients who possess this trait can think abstractly and symbolically about the unconscious meanings of their behavior. They tend to look inward for explanations about things that happen to them, rather than externalize responsibility. They also can think by analogy—that is, something that happens in outside relationships is analogous to similar developments in the transference relationship to the analyst.

The psychoanalyst must also take into account certain external life situations in determining the patient's suitability for analysis. It is generally better not to begin analysis in the midst of a major upheaval or life crisis, such as a job loss or a divorce. Sufficient stability in one's work situation is necessary to ensure that a patient will be in the same geographical location for a period of several years. Serious physical illness may interfere with the patient's ability to invest in a long-term treatment process. The patient must have sufficient financial resources so that analytic treatment does not present an unbearable economic hardship. The final decision must take into account those external life factors in addition to a careful psychodynamic evaluation of the patient's ability to use psychoanalysis productively, based on the criteria noted above.

One broadly based indication for psychoanalysis is a history of unsuccessful previous treatments. A survey of 580 analytic patients found that 82 percent of the patients currently in psychoanalysis had undergone other forms of psychotherapy in the past. The mean age at the beginning of the analysis of the sample was 36.2 years, and 59 percent of the patients were female. Of the 580 patients, 71.4 percent had at least one personality disorder diagnosis. Of the Axis I diagnoses, a mood disorder was the most common, and anxiety disorder was second. A significant number of the patients, 27 percent of the total, had been sexually abused or assaulted. High levels of sexual dysfunction, 43.4 percent of the total, were also found in the group.

Analytic Process In formal psychoanalysis, the process is conducted four or five times a week in sessions that last either 45 or 50 minutes. The patient usually reclines on a couch while the analyst sits behind the couch out of the patient's view. The patient is invited to say whatever comes to mind without censoring it or passing judgment on it. This technique, known as *free association,* promotes a useful regression in the patient that encourages attunement to internal processes and establishment of a transference attachment to the analyst. In addition to the process of free association, the patient is expected to mobilize basic ego resources to make connections between the associations that emerge and to gain mastery through insight. In that regard, the patient works in a collaborative process with the analyst, rather than taking a passive position in which the analyst does all the work.

When Freud shifted from the use of hypnosis and suggestion to the technique of free association, his goal was to bring repressed material into conscious awareness, so that the patient gains insight through understanding how dynamically repressed childhood events contributed to the symptoms experienced as an adult. Freud noted that the regression produced by the analytic setting allows reemergence of infantile conflicts in the form of a *transference neurosis,* in which all the feelings and

fantasies involved in the oedipal constellation of childhood are reactivated in the relationship with the analyst.

Psychoanalysis as a therapeutic procedure, then, was originally based on the notion of uncovering and reconstructing the analysand's past through re-creation of that past in the transference neurosis with the analyst. However, the understanding of the analytic process has expanded considerably as a result of increasing interest in the analyst's role in the analytic relationship. The reexperiencing, or reenactment, of the past is now viewed as influenced by the current analyst-patient relationship and the contributions of the analyst to that relationship. Hence the meaning of analytic data in a particular hour is gleaned not exclusively from the patient's historical past but also from the interplay between the patient and analyst in the present. The past is influenced by the present just as the present is influenced by the past in the analytic setting.

Phases of the Analytic Process The analytic process has traditionally been divided into three phases, although the gradations are acknowledged to be somewhat arbitrary and, to some extent, overlapping. The first, or opening, phase involves the patient's establishment of a therapeutic alliance with the analyst. That alliance involves developing trust and a willingness to work collaboratively with the analyst in the pursuit of understanding. The opening phase also provides a chance for patients to tell their story. As this narrative unfolds, the analyst becomes familiar with the details of the patient's personal history and the patient's conceptualization of the origin of the problems that brought the patient to analysis. The second, or middle, phase is characterized by the appearance of a transference neurosis, which is then systematically analyzed and worked through in the analyst-patient relationship. The issues that arise in the middle phase often stem from a mixture of oedipal and preoedipal conflicts. The new emphasis on adult development is accompanied by a growing awareness that issues deriving from latency, adolescence, and adult development also require attention. The third, or termination, phase of analysis requires that the patient mourn and give up infantile attachments to parental figures at the same time that the loss of the analyst as a real person is being mourned. Autonomy, independence, and development of the capacity for continuing self-analysis are also prominent issues to be addressed during the termination phase. Analysts do not typically view an analysis as finished or complete nowadays. All analyses end with a certain degree of unfinished business that the analysand continues to analyze by identifying with the analyst's role and continuing the process of self-analysis.

Principles of Technique Psychoanalytic technique has undergone rather profound changes since its development by Sigmund Freud. Considerable controversy exists about so-called classical technique. Many aspects of the psychoanalytic approach designated by that term appear to distort how Freud himself really practiced. Moreover, Freud clearly ignored some of the recommendations in his technique papers in his actual technique with his patients. Another difficulty lies in how certain of his technical prescriptions became cornerstones of psychoanalytic technique while others were relatively neglected. For example, his statements about abstinence, neutrality, and anonymity are often viewed as inviolate precepts, while his other statements on the importance of flexibility, spontaneity, tact, and openness are omitted from discussions of technique.

Another dimension of this difficulty is that Freud actually wrote a rather limited number of contributions that focus specifically on psychoanalytic technique. Of 300 scientific publications, only about 20 papers focused on how the analyst should conduct psychoanalytic treatment. Most of these were published relatively early in his psychoanalytic career. He also emphasized that rigid rules were not of much use in the practice of analysis. In a 1910 paper on “wild analysis,” he made the following observation:

> It is not enough, therefore, for a physician to know a few of the findings of psychoanalysis; he must also have familiarized himself with its technique if he wishes his medical procedure to be guided by a psychoanalytic point of view. This technique cannot yet be learnt from books, and it certainly cannot be discovered independently without great sacrifices of time, labour, and success. Like other medical techniques, it is to be learnt from those who are already proficient in it.

Freud also made it clear that the technique he evolved was one suited to his own individuality. He allowed for the possibility that other analysts might adopt different technical strategies based on what was most clinically useful to a particular analyst and a particular patient.

Freud's technique was guided by his understanding of the primary task of psychoanalytic treatment, which in his view was to make the unconscious conscious. Hence free association and dream analysis were given privileged status within his psychoanalytic method. The analysis of transference and neutrality were not cornerstones of Freud's clinical work. Because of Freud's emphasis on making the unconscious conscious, he placed great weight on what he designated the *fundamental rule:* the patient must agree to be completely honest with the analyst and say whatever comes to mind. Freud was quite insistent to his patients that they must follow the fundamental rule and overcome their resistances to doing so. This authoritarian stance by the analyst is quite different from the way most analysts practice today. In an international sample of eminent analysts surveyed, only 37 percent followed Freud's recommendations about instituting free association as the fundamental rule of treatment. Many contemporary analysts are concerned that issuing an authoritarian injunction that tells analysands exactly what to do might increase their resistance to the process.

The ensuing discussion of technical principles puts Freud's views into context with more recent innovations in technique.

Resistance One of Freud's greatest insights into the human psyche was that most people, no matter how extensive their suffering, are highly ambivalent about changing. Powerful internal forces oppose the analyst's efforts to produce insight and promote new modes of adaptation. Indeed, those internal forces, which Freud termed *resistance*, were crucial in the evolution of psychoanalytic technique. Hypnosis and abreaction did not produce lasting change, Freud learned, because forces of resistance tenaciously cling to the neurotic compromise formations. In one of his 1912 papers on technique, Freud observed that “The resistance accompanies the treatment step by step. Every single association, every act of a person under treatment must reckon with the resistance and represents a compromise between the forces that are striving toward recovery and the opposing ones.”

Resistances are ubiquitous in psychoanalysis. They may be viewed as the analyst's daily bread-and-butter work. They manifest themselves in every aspect of the patient's behavior and associations and may take a variety of forms, including falling asleep during sessions, forgetting the analyst's observations, withholding important thoughts and feelings from the analyst, demanding symptom relief instead of understanding from the analyst, falling silent during the sessions, failing to pay the bill, and even free associating so literally that none of the material is connected or used constructively to produce understanding.

Although resistance may be either conscious or unconscious, resistance to the analytic process arises automatically as soon as one is invited to free associate. The exercise of a patient's will power is often insufficient to overcome these powerful forces. In Freud's day, he never referred to “analyzing” resistance. Rather, his view was that resistances needed to be overcome. In his well-known technique paper “Remembering, Repeating and Working Through,” published in 1914, he offered the following advice:

> The first step in overcoming the resistances is made, as we know, by the analyst's uncovering the resistance, which is never recognized by the patient, and acquainting him with it. Now it seems that beginners in analytic practice are inclined to look on this introductory step as constituting the whole of their work. I have often been asked to advise upon cases in which the doctor complained that he had pointed out his resistance to the patient and that nevertheless no change had set in; indeed, the resistance had become all the stronger, and the whole situation was more obscure than ever. The treatment seemed to make no headway. This gloomy foreboding always proved mistaken. The treatment was as a rule progressing most satisfactorily. The analyst had merely forgotten that giving the resistance a name could not result in its immediate cessation. One must allow the patient time to become more conversant with his resistance with which he has now become acquainted, to *work through* it, to overcome it, by continuing, in defiance of it, the analytic work according to the fundamental rule of analysis.

In other words, Freud's emphasis was on overcoming the resistance with the assistance of the analyst, who uncovered the resistance for the patient. In Freud's recommendation analysts are not advised to analyze the resistance in terms of its meaning or its complex defensive function. Instead, patients are supposed to push it aside or plow through it while forcing themselves to say what comes to mind. In accounts by his own analysands of how Freud worked, they paint a portrait of Freud as an analyst who tried to convince his patients through appeals to their willpower that they could decide to give up their resistance.

Modern analysts approach resistances with the idea that they must be analyzed. Through the patient's associations to them and the analyst's observations of the setting in which they occur, resistances are investigated with the idea of learning more about their structures, their sources, and the specific affects with which they are associated. Moreover, the form that a patient's resistances take provides a gold mine of information about that particular patient's characteristic defensive operations. Resistance may be conceptualized as the manifestation of the patient's typical defense mechanisms as they appear in the analytic setting. Another way of making the distinction is to say that defenses, which operate intrapsychically, are translated into resistances, which appear interpersonally, when the patient enters an analytic relationship with the analyst. To be sure, some resistance continues to be intrapsychic as well, but even this will become manifest within the relationship of

the analyst and analysand.

Resistance takes many forms. One particular subgroup of special importance is *transference resistance,* which includes fantasies, thoughts, feelings, and wishes that directly involve the analyst. Loving or sexual wishes directed toward the analyst, for example, may serve the function of avoiding the analytic task of developing insight. Transference resistance may also take the form of defending against the awareness of feelings toward the analyst. Also, one set of transference feelings or wishes may defend against awareness of another, more disturbing set. Transference itself may be regarded as a resistance when it interrupts the process of analysis, and the systematic working through of intense feelings is a major task of psychoanalytic treatment.

Since Freud's time, the understanding of resistance has evolved considerably. Freud conceptualized analysis as primarily an effort to allow memories to surface and to make the unconscious conscious. A more contemporary view is that resistance actually resists the analyst rather than the emergence of memories. Moreover, analysands seek to repeat their past experiences in action rather than to reflect and remember them. Hence Freud's definition of resistance as a blockage of emerging memories is considered limiting by contemporary standards. Analysts today would view the natural flow of the patient's thoughts as heading in the direction of unmodified enactments rather than memory per se, and the resistance would be viewed as a revelation of an important desire or wish in a complex object relationship that is manifested in the transference. In working with the resistance, the analyst wants to facilitate both the conscious activation of repressed wishes and a cool contemplation of their significance.

Resistance tends to change character during an analysis. In the opening phase, resistance may take the form of avoiding a meaningful attachment to the analyst. In the middle phase, it may involve intense and irrational feelings toward the analyst. In the termination phase resistance may manifest itself as a return of symptoms that were present at the beginning of the analysis or a reluctance to give up the passive and dependent regressed relationship with the analyst.

Analyses that are conducted from a self-psychological perspective approach resistance in a manner radically different from the classical or ego psychological model described here. As a consequence of the deemphasis on the ego and its defense mechanisms in self-psychology, the role of those defenses as they emerge in the analysis is also minimized. The view of the analytic process is not one of an inexorable unraveling of defenses. Rather, the analyst attempts to empathize with the core subjective experience of the patient in the interest of reactivating growth in the domain of the self. From that standpoint, both defenses and resistances are viewed as healthy intrapsychic activities that safeguard and preserve the integrity of the self. Kohut preferred to speak of the defensiveness of patients, rather than their resistances, and he stressed that defensive attitudes are both psychologically valuable and adaptive. In his supervision of colleagues, he stressed that the patient's material should be taken at face value rather than inverted or investigated for the “real meaning” beneath the surface. Similarly, he stressed that the analyst's technique cannot bypass the patient's subjective experience.

Object relations theory provides yet another perspective on resistance. Particular self-object affect units within the patient gain prominence for compelling intrapsychic reasons, and the patient only reluctantly gives up those object relations paradigms in the interest of developing new forms of object relatedness. The patient's internal object relations are repeated in the analytic setting with the hope that they will be altered in some manner in the context of a new relationship. At the same time, however, relinquishing internal objects and their corresponding self-representations involves feelings of loss and grief. Those feelings are often defended against by the patient's strong propensity to cling tenaciously to familiar self-object affect units.

Transference *Transference* refers to the displacement onto the analyst of attitudes and feelings originally experienced in relationships with people from the past. Transference patterns appear automatically and unconsciously in the analytic relationship. Analysands suddenly find themselves reacting to the analyst with intense feelings that are inappropriate, at least in part, to the current situation. Patients unconsciously reenact a past relationship instead of remembering and verbalizing it. Transference is not unique to the analytic setting. Every significant relationship in adult life is a new addition of the original attachments of childhood. The chief difference between transference in the analytic relationship and transference as it occurs in outside relationships is that transference is analyzed in the analytic setting.

Freud deemphasized analysis of the transference in his own practice because his principal goal was to bring unconscious material into the conscious awareness of the analysand. He felt that certain aspects of transference were the vehicle of cure and should be left alone. Specifically, what he termed the *unobjectionable positive transference* was the perception by patients that the analyst was a helpful clinician trying to collaborate with them in finding a cure for their problems. This type of transference should not be analyzed, in Freud's view, because it does not interfere with the unfolding of the associations. However, Freud identified two other forms of transference that required analysis—the erotic and the negative. In other words, a basic principle in Freud's approach was that the transference did not require analysis unless it was serving as a resistance to free association. Critics of Freud have pointed out that he may have neglected positive transferences that served to conceal negative feelings such as anger and contempt.

Freud held to the view that transference was basically a template existing in the patient's intrapsychic world that derived from meaningful childhood relationships. He was convinced that this transference would necessarily emerge if the patient simply followed the method of free association. He did not distinguish between the real relationship and transference, and he was not concerned that real aspects of his behavior might influence the patient's transference to him.

In the contemporary view, transference is regarded as a mixture of the new relationship with the analyst and a reenactment of a past relationship with a significant childhood figure. The analyst is regarded as an active participant in the therapeutic relationship. In the current conceptual framework, the analyst's personal characteristics exert a powerful influence on the specific shape and intensity of the patient's transference. The notion of opacity is considered a myth by contemporary analysts, and the analyst's real characteristics must be taken into account when the analyst formulates therapeutic interventions.

In the classical view, the transference neurosis is the distinct and pivotal entity that develops in the middle phase of the analysis as a consequence of the replication of the childhood neurosis. The patient's attachment to the analyst is so intense that emotional satisfaction from the analyst becomes more important than the original therapeutic goals. The contemporary perspective on transference is that object relations ideas are more significant than drive-derived aims aroused by the analytic regression. The term *transference neurosis* is seen as lacking specific meaning or utility in the analytic setting; the term has been replaced to a large extent by references to the various transference responses as shifting self-representations and object representations arise in the crucible of the analytic dyad.

Analysts influenced by self psychology and intersubjectivity theory point out that transference has a bidimensional quality. One dimension involves a repetition of past experiences with old objects. The other involves a quest for a new object or selfobject experience that will be reparative and corrective.

Therapeutic Alliance The controversial term *therapeutic alliance,* used interchangeably with *working alliance,* refers to the conscious aspect of the relationship between the analyst and the patient involving the collaborative pursuit of common therapeutic goals. In the therapeutic alliance the patient identifies with the general aims and strategies of psychoanalytic treatment and cooperates with the analyst in the search for understanding and insight. To form such an alliance, the patient must be able to split the ego into a functioning component that is involved in free associating and an observing component that is identified with the analyst's task of analyzing.

Several elements contribute to establishing a viable therapeutic alliance: shared arrangements regarding fees, payment, appointments, and confidentiality; the mutually agreed on goals for the analysis; and a shared understanding regarding the roles and responsibilities of each member of the analytic relationship. Although the therapeutic alliance is often regarded as the patient's ability to collaborate with the analyst, most clinicians would agree that the analyst's contributions to the formation of the alliance are just as important as the patient's.

Many analysts have been critical of the validity of the concept. One source of criticism stems from the seminal work of Zetzel on the concept, in which she linked certain technical recommendations with efforts to enhance the collaboration of the patient. She suggested that the analyst should not be too silent for too long without making supportive comments that acknowledge the patient's anxiety and distress. She was writing in a context of the expansion of psychoanalytic technique to include the treatment of borderline and narcissistic patients. Nevertheless, some critics have viewed her contribution as implying too much reliance on supportive and nonanalytic strategies.

Another source of criticism of the concept is that the distinction between transference and therapeutic alliance is illusory. Just as Freud's notion of the unobjectionable positive transference might lead the analyst to neglect certain concealed feelings behind the overall positive relationship of the analysand to the analyst, the therapeutic alliance construct might also be misused in that way. Some critics would argue that the relatively collaborative and positive relationship between patient and analyst must itself be regarded as a transference that has layers of complexity. If the analyst does not analyze the apparent alliance, the patient may terminate with one sector of the personality largely unanalyzed.

Countertransference The term *countertransference* was coined by Freud in 1910. He actually wrote very little about the phenomenon in the course of his career, but his general use of the term suggested that it was tantamount to transference in the reverse direction—the unconscious conflicts from the analyst's past are displaced onto the patient. In short, countertransference is the analyst's transference to the patient. Freud viewed it as interfering with the analyst's optimal functioning, and he issued a command that the "analyst shall recognize this counter-transference in himself and overcome it."

Freud was greatly concerned about how a number of his colleagues appeared to succumb to the siren song of "love" in the analytic relationship. In a letter dated 31 December 1911, Freud wrote to Jung the following concern:

> Frau C. has told me all sorts of things about you and Pfister, if you can call the hints she drops "telling"; I gather that neither of you has yet acquired the necessary objectivity in your practice, that you still get involved, giving a good deal of yourselves and expecting the patient to give something in return. Permit me, speaking as the venerable old master, to say that this technique is invariably ill-advised and that it is best to remain reserved and purely receptive. We must never let our poor neurotics drive us crazy. I believe an article on "countertransference" is sorely needed; of course we could not publish it, we should have to circulate copies among ourselves.

Freud saw the ravages of countertransference first hand, and his papers on technique sometimes read as a version of the ten commandments (i.e., as an appeal to his colleagues to control their emotional reactions to their patients).

The conceptualization of countertransference as stemming from the analyst's unconscious conflicts displaced onto the patient is often referred to as the *narrow,* or classical, view of countertransference. In modern psychoanalytic theory, the term has undergone considerable revision. A *broad* view is now in wide use among analysts to connote the analyst's total emotional reaction to the patient. That shift was fueled by Klein, Winnicott, and Kernberg—all of whom viewed countertransference reactions, particularly in work with seriously disturbed patients, as stemming more from the patient than from the analyst. In other words, through projective identification the analyst begins to feel powerful affects that are associated with the patient's internal self-representations and object representations. Winnicott referred to one variant of such phenomena as *objective hate* to reflect the fact that certain patients cause everyone to hate them; therefore, the analyst's feelings do not reflect specific unconscious conflicts from the analyst's past.

Another implication of the broadened conceptualization of countertransference introduced by the object relations theorists is that the analyst's feelings provide important information about the analysand's internal world. Moreover, by reacting to the patient in a role-responsive manner, the analyst gains greater understanding about the typical problems that occur in the patient's other relationships. Hence countertransference is regarded not so much as an obstacle to the analyst's functioning but as an indispensable aid to the analyst's ability to understand and to intervene effectively with the patient.

As noted above, each of these views tends to undervalue the fact that both analyst and patient generally contribute to the construction of the countertransference. Psychoanalysts today generally view countertransference as a joint creation involving contributions from both members of the dyad. One of the analyst's principal tasks is to engage in a continuing process of self-scrutiny in which the two components are distinguished and understood.

The contemporary focus on intersubjectivity has made the term *countertransference* somewhat murky. Some intersubjectivists have suggested that countertransference simply cannot be differentiated from the rest of the analyst's subjectivity and that there is no way to sort out which element in the countertransference belongs to the patient and which to the analyst. Others have suggested that subjectivity is a fundamental aspect of the analyst that is present independent of the patient. Countertransference is a phenomenon that evolves in response to the patient. Still others argue that countertransference is basically a component of the totality that constitutes the subjectivity of the analyst.

Interpretation Interpretation is one of the analyst's primary therapeutic instruments in psychoanalysis. It involves an explanatory statement by the analyst that links a symptom, thought, feeling, or behavior to its unconscious meaning. It is the vehicle by which the analyst delivers understanding to the patient. In the ideal situation, interpretation is designed to make the patient consciously aware of unconscious (or preconscious) material that is close to the surface of consciousness.

Freud himself appeared to vacillate between two positions informing the use of interpretation. In some of his writings, he suggested that analysts postpone interpretation until the unconscious material is close to the patient's conscious awareness. At other times, however, he would offer his view of what was going on in the patient's unconscious even though it was quite distant from the patient's consciousness. He often expected his patients to face unpleasant truths about themselves without regard to the readiness of the patient's ego to process the interpretation.

Genetic interpretations are statements that connect thoughts, conflicts, and feelings in the present with their childhood antecedents. The analyst reconstructs the patient's past on the basis of memories, dreams, free associations, and transference distortions to shed light on how the present situation repeats the past.

Transference interpretations, often viewed as the analyst's central mutative activity, explain to patients how their perception of the analyst is a reenactment of a past significant relationship. The historical conceptualization of transference interpretation generally involved genetic reconstructions, but the contemporary view deemphasizes reconstruction and helps the patient see how current wishes and expectations in the analytic setting are influenced by the past. Moreover, analysts today enlist the patient's collaboration in understanding the current interpersonal situation with the analyst and how it is influenced by past relationships. In this regard, an attitude of pathological certainty or extreme authoritarianism is discouraged in the delivery of an interpretation. Rather, an interpretive intervention is presented as a hypothesis about the patient's behavior and its meaning. The analyst is then interested in hearing the patient's thoughts about the hypothesis.

Transference interpretation has two broad categories. The first involves explanations of resistance to the resolution of the transference. The explanations may involve genetic elements or may focus on the here-and-now situation. An example of a genetic interpretation is: "You may not want to give up your erotic attachment to me because you hope that it will lead to getting a response from me that you never received from your father." An example of a here-and-now interpretation is: "You may not want to give up your erotic attachment to me because you have the fantasy that you can control me with your sexual demands." The second category of transference interpretation involves explications of resistance to the awareness of transference. An example is: "I wonder if your sudden need to launch into an extramarital affair is related to the development of sexual feelings for me that you cannot yet acknowledge."

A single interpretation does not resolve either transference or resistance in one fell swoop. The analyst's initiation of an interpretive strategy marks the beginning of a prolonged process that Freud referred to as *working through*. Freud commented that "one must allow the patient time to become more conversant with his resistance with which he has now become acquainted, to work through it, to overcome it, by continuing, in defiance of it, the analytic work according to the fundamental rule of analysis." Similar transference and resistance themes are encountered in a variety of different situations, and the repetitive interpretation process each time those familiar patterns emerge eventually leads to mastery and insight.

From an ego psychological perspective, working through is necessary because of the tenacity of ego defenses. A single demonstration of the impulse-defense configuration does not suffice for the patient to relinquish those defenses and allow the warded off drive components to enter conscious awareness. Repeated confrontations of the defensive strategies are necessary in a variety of situations to accomplish a major shift in the intrapsychic relationships of ego, superego, and id. Paul Gray has stressed that the analyst's principal task is *resistance analysis,* in which the analyst pays close attention to defensive shifts away from sexuality or aggression. Over time patients learn to monitor these shifts themselves and recognize their flight from particular issues they find disconcerting. Gray's vision is one of enlisting the patient's ego in a partnership with the analyst. He advises against making interpretations that attempt to bypass ego functions and speak directly to the patient's underlying drive derivatives.

Object relations theorists view the working-through process somewhat differently than the ego psychologists. Proponents of the British middle group, for example, such as Fairbairn and Winnicott, stress that the analytic setting provides a new object for the patient. Working through involves attachments to the new object, internalizing the new object relationship, and relinquishing the old internal object relations units as the new relationship takes center stage in the patient's psyche.

Self psychologists do not conceptualize the working-through process as involving resolution of intrapsychic conflict. Instead, they regard it as the resumption of previously arrested developmental processes. The patient's relationship with the analyst in the present reactivates that frozen development so that selfobject transferences can evolve. Through the repetitive examination of self-selfobject disruptions in the transference, the patient gradually internalizes the empathic interactions with the analyst to repair defects in the self.

DREAM INTERPRETATION Freud's classic work, *The Interpretation of Dreams* and subsequent papers on the mental phenomena of everyday sleeping and waking

demonstrate that dreams, parapraxes, jokes, puns, and metaphors are not chance or senseless events but are meaningful acts subject to the basic laws governing unconscious processes. Freud believed early on that the analysis of dreams, such as he had conducted on himself, was sufficient for a full self-analysis. As the fastest and surest route to the origins of psychic conflicts—the royal road to the unconscious—dreams remain for many analysts the perfect paradigm of the interpretive process and often supply a source of information that cannot be derived in any other way.

Dreamwork is a distorting process by which latent dream thoughts are condensed, displaced, and converted through symbolic substitution into manifest dream material of which the recollection, upon wakening, is only an unintelligible facade or disguise. The dream is generally looked upon as a microcosm of the major concerns of the patient, especially as a commentary on the events of the preceding day (day residue). For the dreamer it serves a largely wish-fulfilling function behind which resides the original unsatisfied instinctual impulse. Current views also suggest a prospective function of dreams, as a direct attempt to adapt to present stresses or to solve future problems. Despite its salient role in analysis, however, dream interpretation should not be pursued for its own sake but be subject to the same rules that govern all analysis. Excessive attention to dreams can become a form of resistance by drawing the patient away from more spontaneous concerns.

Neutrality, Abstinence, and Anonymity Some of the most striking contradictions between the so-called classical model of technique and the way Freud actually behaved involves the principles of neutrality, abstinence, and anonymity. Part of the confusion clearly relates to the difference between the way that Freud conducted himself in his analyses and some of the statements he made in his papers on technique. For example, he suggested that the analyst should be like a mirror that reflects back to the patient only what is shown to the analyst. In other words, the analyst should eschew too much self-disclosure. At other points in his technique papers, he advises the analyst to proceed with “emotional coldness” and to put all personal feelings aside, “even his human sympathy.” The image projected in these papers is clearly that of a detached scientist dispassionately examining a specimen without any personal or human involvement with the patient. The thinking behind this approach was that by providing a “blank screen” or mirror, the analyst could facilitate the emergence of the patient's transference.

Two sources of information suggest that Freud did not follow these prescriptions in his actual analytic work. Both written reports from his own analysands and his published case material indicate that Freud's personality was very much involved in the analytic process. He gossiped with his patients; he offered his opinions about theories, works of art, and other people, and certainly did not attempt to conceal his political and moral value system. He was transparent regarding his mood and loved to talk with his patients about topics of mutual interest. He was real, often authoritative, and certainly spontaneous.

In fact, Freud did not actually use the word “neutrality” in his writings. James Strachey, his principal translator, translated the German word *indifferenz* as “neutrality” even though another German word carries that meaning more precisely. The German word actually implies an underlying emotional participation by the analyst rather than the detachment that is often connected with the term *neutrality*.

Fundamentally the issues of neutrality, abstinence, and anonymity were irrelevant for Freud. He desired to facilitate the patient's free associations so that the unconscious would emerge. His analysands report that he was far from cold or detached in his own practice. Freud's statements about technique appear to relate more to his concerns about the vulnerability of his colleagues to countertransference acting out and the tendency of some analysts to misuse the analytic situation to talk about themselves.

The fact that *neutrality* is occasionally misunderstood to suggest that analysts should remain cold and aloof does not mean that the term is meaningless. The most widely accepted contemporary meaning is a nonjudgmental stance regarding the patient's feelings, wishes, and behaviors. Anna Freud suggested that the analyst must remain equidistant from the id, the ego, the superego, and the demands of external reality. By maintaining a neutral position, the analyst eventually helps patients become aware that their own critical attitudes are being attributed to the analyst.

Concern, empathy, and emotional warmth are all part of effective analytic work, and a cold, rejecting attitude in the analyst may become an aspect of the real relationship in the present that is incorporated into the patient's transference. The influence of object relations theory on psychoanalytic technique has redefined the analyst's role as one involving participation in the process. When analysts allow themselves to respond in a disciplined yet spontaneous manner to the patient's projections, they find that their role-responsive reactions provide useful information and an empathic sense of the patient's internal world.

The contemporary models of technique have generally argued for a disciplined spontaneity, much like the way Freud himself practiced. Analysts today are aware that anonymity is largely a mythical construct. Various kinds of personal information are available in the analyst's office. Photographs of the family, artwork, books, and other articles of personal interest are fully visible to the patient. The analyst's reactions to material the patient brings to the process are also revealing. Indeed, one could argue that every intervention the analyst makes reveals something about the analyst's personal proclivities. Hence one cannot avoid self-disclosure. The issue becomes one of selecting what is useful to disclose in an explicit verbal mode versus what is preferable to keep to oneself. Analysts still recognize a value in analytic restraint. Although it is impossible to be entirely neutral in the sense of being nonjudgmental, analysts can nevertheless choose how much is verbally revealed about their judgments.

Freud attempted to distinguish between the transference and a personal, real relationship. In Freud's view only the transference required analysis. Contemporary analysts recognize the difficulty in trying to separate out the real relationship from the transference and recognize that each infiltrates the other. Hence the entire relationship with the analyst must be subject to analytic scrutiny.

Abstinence was a technical principle designed by Freud under the same cloud of sexual boundary violations by some of his early colleagues and disciples. He was mainly concerned that extremes of feelings must be kept in check. In a 1919 paper, “Lines of Advance in Psycho-Analytic Therapy,” Freud clarified that:

> by abstinence, however, is not to be understood doing without any and every satisfaction—that would of course not be practicable; nor do we mean what it popularly connotes, refraining from sexual intercourse; it means something else which has far more to do with the dynamics of falling ill and recovering.

What Freud meant in essence was that the analyst withholds gratification of transference wishes so that they can be analyzed rather than relieved. Analysts today recognize that partial transference gratifications occur throughout an analysis. The analyst's warmth, humanity, laughter in response to a joke, and empathic listening are all gratifying. Richard Chessick has described it as follows: “The skill of the therapist is to do just enough so that the patient has some sense of the therapist's recognition that words alone will not suffice, but not to afford so much gratification that the patient has no motivation to change, which would engender an artificial need for more and more gratification.” Indeed, the concept of analytic boundaries has arisen in part to clarify the role of gratification versus abstinence in modern technique.

Analytic Boundaries The notion of analytic or professional boundaries is a relatively recent development in psychoanalysis. The early history of psychoanalysis is replete with phenomena that would now be viewed as unethical or incompetent technique. Freud analyzed Ferenczi, for example, while walking through the countryside. He regularly introduced his patients to his family and often had meals with them. Winnicott broke confidentiality by revealing information about his patients to other patients. Ferenczi engaged in mutual analysis, a process in which after hearing the patient's associations for an hour, he traded places with the patient and free associated himself for an hour. There are numerous cases of analysts falling in love with their patient and interrupting the analysis so the two could embark on a love affair.

In the last two to three decades, a good deal of writing has appeared in this area. Analytic boundaries connote an edge, or limit of acceptable professional behavior. They are not designed to be an arbitrary set of rules by which ethics committees and licensing boards determine whether or not disciplinary measures are needed. They are simply guidelines to facilitate the analytic process. In a 1927 letter to Ferenczi, Freud indicated his intent regarding such guidelines:

> I consider the most important thing was to emphasize what one should not do and to point out the temptations and directions contrary to analysis. Almost everything positive that one should do I have left to “tact,” the discussion of which you are introducing. The result was that the docile analyst did not perceive the elasticity of the rules I have laid down, and submitted to them as if they were taboos. Sometimes all that must be revised, without, it is true, doing away with the obligations I had mentioned.

Indeed the analytic boundaries must build in flexibility. They provide an envelope that creates an optimal environment for the emergence of the analytic process. They must be flexible because different patients require adjustments in the boundaries. Similarly, different patient-analyst dyads find their own optimal conditions under which analysis can be conducted. The relative degree of gratification versus frustration varies given the subjectivities of the two parties. Some patients may benefit from direct self-disclosure of the analyst's feelings about a specific situation, while others may deeply resent it and close down as a result. Nevertheless, a variety of boundary considerations serve as guidelines to assist the analyst in maintaining a professional rather than a personal relationship.

One boundary is that of role. The analytic role allows the analyst to be subjectively involved while also maintaining a certain degree of relative objectivity as a participant in a professional relationship. While libidinal, or sexual, demands cannot be gratified, growth needs, such as empathic understanding, caring, and concern, are often an integral part of the analytic role. In this regard, abstinence can be redefined as the construction of an analytic setting that serves as a holding environment for the containment and processing of powerful feelings. Analysts do not act like lovers or parents, but they try to understand the way in which the patient attempts to re-create the parent-child relationship in the analytic dyad and the similarities that can be identified between difficulties in love relationships and those in the transference.

Time is another boundary in analysis. The usual session lasts 45 or 50 minutes, and the time constraints in the process emphasize that there are limits in the relationship. These may be productively analyzed in terms of how they frustrate the patient's wishes. Extraanalytic contacts are viewed as analytic material that can be productively discussed and understood.

Place and space are also inherent to the analytic boundaries. By meeting in the analyst's office, the patient is clearly shown that the relationship is confined to the analyst's professional workplace. Similarly, a fee is paid for the analyst's service, which makes clear that analysis is work rather than a social hour. A clinician who fails to charge or to collect from a patient may unwittingly give a message to the patient that the relationship is a special one that exists outside the realm of professional services.

Lavish or expensive gifts from analyst to patient or from patient to analyst may also violate a professional boundary. Such gifts may serve as an unconscious bribe designed to suppress anger in either party. Small and inexpensive gifts may be accepted by the analyst when the analyst feels that the acceptance might enhance the process. Even when such gifts are accepted, however, the meaning of the gift to the patient in most cases will be analyzed as an important part of the work.

Clothing and language represent boundaries that are often neglected in the analytic work. An analyst who dresses in a highly revealing or frankly seductive way may convey a misleading message to the patient. Similarly, obscene or profane language may offend some patients and feel like a verbal form of sexual violation.

As noted previously, self-disclosure is an important boundary issue that analysts must attend to in their decisions about interventions. Analysis is by nature asymmetrical. Although two subjectivities are in the room and strong feelings are stirred in both parties, the process is designed to focus primarily on the person who is paying for the service. Self-disclosure by the analyst is inevitable, of course, and the analyst is always making decisions about how useful it would be to disclose certain aspects of the analyst's subjectivity to the patient. It may be extraordinarily useful in some cases to talk about areas of common interest with the patient in the service of building a therapeutic alliance. However, excessive self-disclosure of one's personal problems may unduly burden the patient. Analysts who talk about their own problems misuse the patient's time and money. In some cases, the disclosure of here-and-now countertransference feelings may advance the process in a constructive way. However, telling patients, “I have sexual feelings for you” may shut down the process and make patients feel that they must set the boundaries in the relationship. Similarly, disclosing that one hates a patient or is bored by a patient is rarely useful.

Sexual contact of any form between analyst and patient is unacceptable. This constitutes a firm, inflexible boundary. In the transference the analyst becomes a parent to the patient so that any form of sexual relations is symbolically incestuous. Moreover, because of the transference and the power differential, the patient cannot provide informed consent to such a relationship, even if the patient is an adult and consciously attracted to the analyst. This absolute abstinence regarding sexual contact facilitates frank and detailed discussion of the patient's sexual desires in a safe environment.

Physical contact such as hugs or pats on the back may also be problematic. In general, a handshake is probably the ordinary limit of physical contact between analyst and analysand. An occasional hug in the midst of a personal tragedy, such as the loss of a child, spouse, or parent, might be appropriate if initiated by the patient. However, when the analyst initiates a hug, the patient may readily misconstrue the analyst's motives. Another basic guideline regarding touch is that the impact of such behavior on the patient may be dramatically different from the analyst's intent. Patients may ask for hugs or even demand them, but psychoanalysis is about the wish to be held or hugged rather than the concrete enactment of such wishes. The patient must be engaged in a mourning process to deal with the grief and resentment about the deprivations of childhood, the frustrations in the present, and the insistence of the analyst that the relationship remain an analytic one. When actual physical contact occurs, especially on a regular basis, the distinction between the symbolic and the concrete is lost, and the patient may feel that powerful childhood longings will finally be satisfied by the person of the analyst. This situation will likely evoke false hopes in the patient that can never be gratified.

When one gets outside the realm of extreme boundary violations, such as financial exploitation or sexual involvement between analyst and patient, there are gray areas where certain transgressions of boundaries may actually be helpful in an analytic process. Thomas Gutheil and the author distinguish between a boundary *violation,* which exploits the patient and is harmful or potentially harmful, and a boundary *crossing* that may be benign or even helpful to the analytic process. With the contemporary understanding that countertransference enactments are inevitable, it is sometimes hard to differentiate which enactment is a crossing and which is a violation. Often the meaning of a boundary crossing may be the only real clue to understanding whether the enactment is a violation or not. In general, a countertransference enactment that turns into a useful boundary crossing is attenuated, does not occur on a regular or repetitive basis, and is subject to analytic scrutiny. Boundary violations, on the other hand, are generally part of an ongoing, repetitive pattern of boundarylessness and are rarely subject to detailed analytic scrutiny. Finally, in a boundary violation the analyst's behaviors cause harm to the patient and destroy the viability of the analytic process (Table 30.1-2). Sexual relations between analyst and patient are an example of a boundary violation. The following vignette is an example of a boundary crossing.

Dimensions of analytic boundaries	Professional role, time, place and space, money, gifts, business transactions, clothing, language, confidentiality, excessive self-disclosure of personal problems, physical contact, sexual relations
Boundary violations	Egregious enactments that are often repetitive, not subject to analytic scrutiny, pervasive, and harmful to the patient while also destroying the viability of the analysis
Boundary crossings	Benign, and even helpful, countertransference enactments that are attenuated, occur in isolation, are subject to analytic scrutiny, and extend the analytic work in a positive direction

Data from T. Gutheil, M.D., G. Gabbard, M.D., and J. Lester, M.D.

Table 30.1-2 Analytic Boundaries and Boundary Violations

At the end of a long and frustrating session, the therapist felt like nothing she had said had been helpful. She had been preoccupied with the fact that her own mother was ill, and she had not been as available to the patient as usual. As the therapist and patient got up from their chairs, the patient asked, “So where are you going next week?” Without thinking, the therapist responded, “I'm going to visit my mother in Colorado. She's very ill.” The patient said, “Oh, I'm sorry,” and left the office looking worried. The therapist instantly recognized that she had revealed information that might have burdened the patient with her own problems. As she reflected on her enactment, she recognized that she has been feeling guilty about not having been very helpful because she was so preoccupied and was unconsciously trying to gain some sort of absolution or forgiveness from the patient by explaining the situation to him.

After her 1-week absence, she brought up what had happened at the end of the session with the patient and explained that she probably should not have burdened him with the information she offered. She went on to suggest that they might beneficially explore his reactions to it. In the course of sharing some of his thoughts about what happened, the patient noted that his mother had always confided her problems to him, and he thought something like that was repeating itself in the therapeutic relationship. In this instance no lasting harm was done to the patient, and the event turned out to be a useful focus for further exploration. An important reason why this enactment was constructive was that it was discussible between therapist and patient.

Goals and Therapeutic Action It is customary in most of medicine to speak of cure as a treatment goal, but one must be considerably more cautious in discussing the outcome of psychoanalysis. A key aspect of the analyst's position is avoiding therapeutic zeal. The analyst's task is to analyze, interpret, and provide understanding. The patient's task is to decide how much and in what direction to change. When analysts start to have a specific therapeutic agenda for their patients, they are at risk of creating a situation in which the patient changes to please the analyst or rebels to defy the analyst, a re-creation of common parent-child situations. Hence the goals of analysis emanate as much from the patient as from the analyst. Conversely, without clear therapeutic goals for the analysis, it can meander

aimlessly indefinitely, with both parties ultimately feeling frustrated.

The goals of psychoanalysis are intimately linked to conceptualizations of the therapeutic action of the treatment. The mechanism by which psychoanalysis works is a source of controversy in the field. When Freud first began treating hysterical patients in the 1890s, the therapeutic action was simple and straightforward in his view: through abreaction the analysis would help the patient release dammed up affects and recover previously repressed fantasies, memories, and feelings. After Freud developed the structural model, though, abreaction became less significant, and he regarded the therapeutic action of psychoanalysis as involving a redistribution of the interrelationships among the intrapsychic agencies. In Freud's words, “Where id was, there ego shall be.”

Freud always thought that transference work was crucial to effecting change in the patient. However, Strachey in 1934 most clearly articulated an understanding of how transference interpretation produces change in psychoanalysis. Strachey believed that a harsh superego was at the core of most neuroses and that modification of the superego through interpretation is the essential therapeutic action of psychoanalysis. Strachey identified a vicious cycle that takes place in analysis in which the patient's id impulses and harsh superego are transferred to the analyst. The analyst is then experienced as a dangerous and frightening object. Seeing the analyst as a threat produces further destructive impulses toward the analyst; those impulses are then projected to create an even more hostile object.

That vicious cycle is broken, in Strachey's view, through mutative transference interpretations. After interpretation of the transferred id impulses to the analyst, patients become aware that their hostile impulses are directed toward internal parental objects, rather than real objects in the present, such as the analyst. As those interpretations are repeated in the context of the process of working through, patients realize the discrepancy between their own perceptions of the analyst and the real characteristics of the analyst, who then becomes a much less frightening object. The superego is correspondingly modified.

Mechanisms of change and analytic goals vary according to one's theoretical perspective. Analysts working within the framework of the British School of object relations theory believe that an analysis must do more than alter interrelationships between intrapsychic structures. They stress the need for change in mental representations of self and object and in the affective linkage between those representations. They believe that such changes are brought about partly through interpretation but also through the analytic relationship itself. The provision of a holding environment, the analyst's persistent curiosity and interest in the internal experience of the patient, and the analyst's ability to endure attacks or seductions from the patient, provided a new emotional experience for internalization.

Kleinian analysts emphasize the process of containment in producing change. Through projective identification, the analyst becomes a container of the patient's self- and object-representations, often in association with powerful affective states. Before those projected contents are returned to the patient by the process of reintrojection, the analyst psychologically processes (and thereby modifies) the patient's representations. For example, if a patient projects an abusive internal object representation into the analyst, the analyst feels pressured into behaving abusively toward the patient, but instead attempts to process and understand the projected contents. When the patient reintrojects the abusive internal object representation, it has been modified by the analyst's containment process. The modified object representation then modifies the corresponding internal self-representation. In other words, a new mode of object relatedness has been internalized through the projective identification and containment.

Some object relations analysts, such as Winnicott, view the patient as developmentally frozen. The experience of the analyst as a new object allows the patient to resume development and to experience previously thwarted positive relational and emotional capacities. Some analysts have compared it to a reparenting model in which the analyst conveys acceptance and validation of the patient's uniqueness and autonomy.

Self-psychology views the goal of psychoanalytic treatment as the strengthening of a weakened self so that it can move from a dependence on archaic selfobject experiences to a position of greater self-cohesion that allows for reliance on mature selfobjects. Kohut asserted that the goal was accomplished by laying down psychic structure through optimal frustration and transmuting internalizations. He noted that failures of empathy are unavoidable in analysis and that the systematic interpretation and understanding of those empathic failures provide optimal frustration and the gradual internalization of the analyst as a new object. The essential curative aspect of the analytic process, however, is the establishment of empathic attunement between the self and the selfobject on a mature level. The experience of empathy allows the patient to overcome enslavement to archaic selfobjects in the service of maintaining self-cohesion.

A trend that cuts across theoretical models is the recognition that multiple modes of therapeutic action come into play in most patients. Both insight and a new relationship experience, for example, may be influential in producing change in the same patient. Moreover, different groups of patients may change in different ways. Sidney Blatt identified two types of patients—introjective and anaclitic—who appear to respond differently to psychoanalytic treatment. The introjective patients are ideational and preoccupied with establishing and maintaining a viable self-concept rather than with establishing intimacy. They appear to be more responsive to insight through interpretation than to the relationship with the therapist. On the other hand, anaclitic patients are more concerned with issues of relatedness than self-development and gain greater therapeutic value from the quality of the therapeutic relationship than from interpretation.

Many models of therapeutic action emphasize the analyst's efforts to provide symbolic meaning to the patient's thoughts, feelings and behaviors. The patient's internal object relations, based on early attachment experiences, are encoded in implicit or procedural memory. In psychoanalytic treatment, these patterns appear in the way the patient relates to the analyst. The analyst points out these patterns as a way of helping the patient develop a sense of mentalization or reflective function, that is, the capacity to recognize that behavior grows out of internal feeling states and representations that serve as motivating sources. From a developmental perspective, this mode of therapeutic action is analogous to a parent reflecting on a child's behavior. Children perceive themselves in their parents' minds and develop a sense of self from such perceptions. Similarly, patients may gain a greater sense of who they are from observing their therapists' reflections about them. These interactions also enable the patient to appreciate more fully the separate subjectivity of the analyst.

Conscious mastery of repetitive relational modes is accompanied by nonconscious affective and interactive connections. Many former patients remember moments of emotional engagement with the analyst far more clearly than the content of any specific interpretation. Some changes occur in the realm of implicit or procedural knowledge involving how to act, feel, and think in a particular relational context. Psychoanalysis can be conceptualized as a new attachment relationship that restructures attachment-related implicit memory. Stored prototypes are modified by new interactions with an affectively engaged therapist. At the same time, explicit memory involving a conscious narrative is altered by interpretive understanding.

Regardless of which theoretical school the analyst embraces, patients at the termination of a successfully conducted analysis feel certain predictable changes. They often have a sense of being freed up from previously mystifying internal constrictions. They may feel more creative and productive so that they are able to pursue work and leisure time activities with more enthusiasm. They are also likely to experience a much greater sense of mastery over their internal states. Greater tolerance of anxiety, for example, allows them to use that disturbing affect as a signal to reflect on what internal and external circumstances are causing concern and thereby to understand the source of the anxiety. Most patients at the end of analysis also experience a much expanded capacity to love others without as much contamination by conflicted relationships from the past. Through identification with the analyst, patients have internalized the self-analytic process that can be used to understand and master new situations as they arise throughout the life cycle. All of these changes can be used by analyst and patient to assess the readiness for termination.

Psychoanalytic Psychotherapy Psychoanalytic psychotherapy is based on fundamental dynamic formulations and techniques that derive from psychoanalysis and is designed to broaden its scope. Psychoanalytic psychotherapy in its narrowest sense is the use of insight-oriented methods only. However, as generically applied today to an ever-larger clinical spectrum, it incorporates a blend of uncovering and suppressive measures.

The strategies of psychoanalytic psychotherapy currently range from expressive (insight-oriented, uncovering, evocative, or interpretive) techniques to supportive (relationship-oriented, suggestive, suppressive, or repressive) techniques. Although those two types of methods are sometimes regarded as antithetical, their precise definitions and the distinctions between them are by no means absolute.

Thirty years after delineating four intrinsic criteria of psychoanalysis (centrality of transference, neutral analyst, regression to transference neurosis, interpretation as sole instrument), Gill revised them in 1984 to provide the backbone of modern psychoanalytic psychotherapy: (1) the centrality of transference is broadened to encompass nontransference manifestations; (2) the guiding principle of neutrality is loosened to acknowledge the real personality and attitudes of the analyst; (3) the regressive transference neurosis is supplanted or replaced by less frustration, fantasy, and focus on the past (and greater gratification, reality, and focus on the present); and (4) interpretation is extended to include more nonexploratory methods (with genetic interpretations expanded by here-and-now interpretations. Those intrinsic changes accompany three major extrinsic changes: reduced session frequency, reduced or more flexible treatment duration, and abandonment of the couch.

Psychoanalytically informed psychotherapies are often divided, somewhat arbitrarily, into psychoanalysis, expressive psychotherapy, and supportive psychotherapy. These three modalities actually reside on a continuum with significant overlap and blending occurring among the three categories. An overview of this continuum is

presented in Table 30.1-3.

Table 30.1-3 Psychoanalytically Informed Psychotherapies: A Clinical Continuum*

In general, psychoanalysis and expressive therapy are oriented to analyzing transference and defenses and uncovering unconscious material. Supportive psychotherapy, while psychoanalytically informed as well, is geared to suppressing unconscious conflict and bolstering defenses. These distinctions have been used traditionally to differentiate expressive from supportive modalities. In recent years, however, there is a growing consensus that the traditional dichotomy is misguided. In the Menninger Foundation Psychotherapy Research Project, Robert Wallerstein studied the treatment of 42 patients in great detail and had access to long-term follow-up data. He concluded that all forms of psychoanalytically informed psychotherapy, from the most expressive (e.g., psychoanalysis) to the most supportive contain a mixture of supportive and expressive elements. Commenting on the tendency in the literature to view supportive therapy as inferior to expressive therapy and analysis, Wallerstein also stressed that the changes accomplished by supportive interventions appear to be just as durable as those produced by expressive strategies. He argued that psychotherapy should be viewed as taking place on an expressive-supportive continuum and that the key point in every psychotherapy process is to determine how and when to be expressive and how and when to be supportive. This shift in the way that psychoanalytically informed psychotherapies are conceptualized has led to a tendency in the recent literature to acknowledge the blending of strategies with terms such as expressive-supportive or supportive-expressive psychotherapy. The effective psychoanalytic psychotherapist will ultimately shift flexibly back and forth along the expressive-supportive continuum, depending on the needs of the patient at a given moment in the psychotherapy process.

Despite this tendency for therapists to switch back and forth across the expressive-supportive continuum, many therapies are predominantly expressive or predominantly supportive. This designation is largely determined by two factors: the use of specific therapeutic interventions and the degree to which the transference is analyzed. Supportively weighted psychotherapy is much more likely to be characterized by a greater number of extratransference interventions than are focused on the therapeutic relationship itself. By contrast, expressive therapy attends in much greater detail to the analysis of transference. The therapeutic interventions themselves can be categorized as residing on an expressive-supportive continuum as illustrated in Figure 30.1-4.

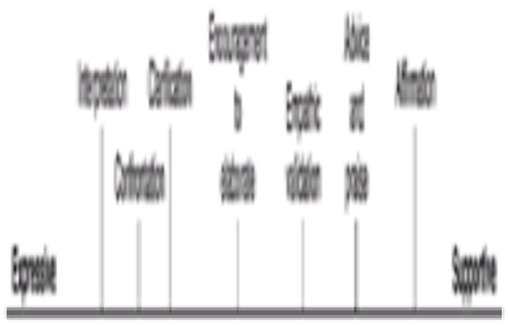

FIGURE 30.1-4 An expressive-supportive continuum of interventions.

Considered the primary intervention in psychoanalysis and highly expressive psychotherapy, interpretation has traditionally designated an intervention designed to make something previously unconscious available to the conscious awareness of the patient. As the concept has evolved, however, the criterion that something unconscious is made conscious has become less absolute. Interpretations often involve explanatory statements that link behaviors, thoughts, fantasies, feelings, or symptoms to their unconscious meanings or origins. They may also link experiences in the transference to parallel situations in current extratransference settings or to the past. For example, the psychotherapist might say to a patient who has forgotten to bring his check so that he could pay for his session, “I wonder if the reason you forgot to bring the check is that you are dissatisfied with the kind of treatment you're receiving here.” Interpretations can also be made on the basis of extratransference phenomena. For example, “It may be that you cannot confront your colleague at work about his unpleasant comments because he reminds you of your father and it revives the same set of conflicts about being forthright with your dad.” Interpretation of unconscious content is generally postponed until the unconscious material is close enough to conscious awareness that the interpretation can be heard nondefensively by the patient.

Confrontation, the next most expressive intervention, generally calls the patient's attention to something the patient is trying to avoid or minimize. Confrontations frequently reflect back to patients the aspects of their behavior that affect others in specific ways. They may also help patients come to terms with disavowed feelings. Although confrontation may have an aggressive connotation, this type of intervention can also be gentle and delivered with sensitivity. For example, “I'm not sure if you noticed, but after I asked you to tell me more about your father, you changed the subject to the car trouble that you've been dealing with recently.”

Clarification is lumped together with interpretation and confrontation as the principal interventions in expressive psychotherapy. Clarification involves an attempt to pull together a number of diverse things the patient has been saying and repackage as a summary of the principal points the patient is making. For example, after listening to the patient for 30 minutes, the therapist might observe: “It sounds like what you're saying is that you repeatedly try to make yourself understood in a variety of situations—with your boss, with your husband, and with your father—and in each situation you have the experience that nobody really grasps what you're trying to say.” As this example indicates, the therapist may help patients articulate something that they have difficulty putting into words.

Encouragement to elaborate, an intervention that lies directly in the middle of the expressive-supportive continuum, may be generally defined as a statement or question that facilitates further elaboration on material the patient is currently discussing. Typical examples are such statements as the following: “I'm interested in hearing more about your feelings about the divorce,” “Does anything come to mind about the dream?” or “How long would you say you felt this way about your dad?” Such questions and statements are used equally in predominantly expressive psychotherapies and predominantly supportive ones.

Empathic validation is a comment designed to help the patient feel understood by demonstrating that the therapist is attuned to the patient's internal state. The therapist may wish to validate the patient's specific feelings being addressed at a moment in the therapy: “I can certainly understand why you would feel outraged at having been treated that way.” The therapist may also acknowledge the adaptive value of the patient's behavior: “I can appreciate from your point of view that the way you behaved seemed like the only viable alternative.” Some self psychologists have suggested that empathic immersion in the patient's internal experience is essential, regardless of where the therapy resides on the expressive-supportive continuum, and that all interventions should be delivered from an empathic perspective.

Advice and praise involve two types of interventions linked by the fact that both reinforce certain activities and proscribe others. Advice, of course, can be defined as direct suggestions to the patient regarding what to do or how to behave. Praise is designed to reinforce certain kinds of feelings, behaviors, or thoughts by expressing overt approval of them. These types of comments are used frequently in supportive therapy. For example, when a patient avoids facing a self-destructive situation, the

therapist might say, “I think it's important that you stop seeing this man in light of his abusive behavior toward you.” If the same patient stopped seeing the man on her own, the therapist might say, “I'm very pleased that you made the decision to stop seeing him because I think it represents a wise choice.” Therapists who use these interventions are departing from a more traditional view of neutrality by offering a value judgment of a particular situation. However, even in certain instances of expressive therapy, therapists often make no effort to completely mask their feelings. Hence advice and praise are used occasionally in expressive therapy as well.

Affirmation is a simple intervention designed to support or affirm the patient's behaviors or comments. In that sense it is linked to advice and praise but is more succinct. Examples are “You're absolutely right,” or “I certainly see exactly what you mean.” Even responses such as “yes” or “uh-huh” may be affirming to the patient.

Expressive and supportive psychotherapy are considered separately below, even though most psychotherapies involve some mixture of expressive and supportive interventions. Indeed, with patients who have limited ego weaknesses, such as those with borderline personality disorder, transference interpretations, which are the most expressive of all interventions, may only be effective after the way is paved with a string of supportive interventions.

Expressive Psychotherapy

Indications and Contraindications According to Karasu expressive psychotherapy is indicated under the following circumstances. (1) Psychoanalysis is diagnostically or clinically contraindicated because the patient lacks sufficient ego strength to tolerate the regression, frustration, or suspension of reality that is required; the patient is deficient in the cognitive resources necessary to achieve deep insight; the patient is not sufficiently motivated because of competing family, social, or cultural influences; or the problem is so pressing that lengthy treatment is unsuitable. (2) Practical or logistical considerations prevail whereby patients otherwise suited for a more lengthy psychoanalysis are unavailable for long-term treatment for reasons of time, money, or geography. For that group the techniques used are closer to those of psychoanalysis proper, because extrinsic rather than intrinsic factors are largely responsible for the change of approach.

Diagnostically, psychoanalytic psychotherapy expressive mode is suited to a range of psychopathology with mild to moderate ego weakening, including neurotic conflicts, symptom complexes, reactive conditions, and the whole realm of nonpsychotic character disorders. The last includes those disorders of the self that are among the more transient and less profound on the severity-of-illness spectrum, such as narcissistic behavior disorders and narcissistic personality disorders. It is also the treatment recommended for some patients with borderline personality disorders, although special variations may be required to deal with the associated turbulent personality characteristics, primitive defense mechanisms, tendencies toward regressive episodes, and irrational attachments to the analyst.

The persons best suited for the expressive psychotherapy approach have fairly well integrated egos and the capacity both to sustain and to detach from a bond of dependency and trust. They are, to some degree, psychologically minded and self-motivated, and they can, at least temporarily, tolerate doses of frustration without decompensating. Patients must have some capacity for introspection and impulse control, and they should be able to recognize the cognitive distinction between fantasy and reality.

They must experience suffering significant enough to motivate them to persevere in the face of frustration, and they should have some meaningful relationships in their lives. The capacity to sustain productive work also reflects the ego strength necessary to engage in expressive treatment. Patients should also be able to regress in the service of the ego and bounce back following the regression. Reflective responses to trial interpretations and the capacity to think in terms of analogy and metaphor also indicate the patient's suitability for expressive work. These characteristics are summarized in Table 30.1-4.

Expressive	Supportive
Strong motivation to understand	Significant ego defects of a chronic nature
Significant suffering	Severe life crisis
Ability to regress in the service of the ego	Low anxiety tolerance
Tolerance for frustration	Poor frustration tolerance
Capacity for insight (psychological mindedness)	Lack of psychological mindedness
Intact reality testing	Poor reality testing
Meaningful object relations	Severely impaired object relations
Good impulse control	Poor impulse control
Ability to sustain work	Low intelligence
Capacity to think in terms of analogy and metaphor	Little capacity for self-observation
Reflective responses to trial interpretations	Brain-based cognitive dysfunction
	Tenuous ability to form a therapeutic alliance

Reprinted with permission from Gabbard GO: Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition. American Psychiatric Press, Washington, DC, 1994.

Table 30.1-4 Indications for Expressive or Supportive Emphasis in Psychotherapy

Goals The overall goals of expressive psychotherapy are to increase the patient's self-awareness and to improve object relations through exploration of current interpersonal events and perceptions. In contrast to psychoanalysis, major structural changes in ego function and defenses are modified in light of patient limitations. Instead of systematically resolving the unconscious nuclear conflict, the therapist may opt to resolve some conflict areas and undo specific resistances, intentionally overlooking or reinforcing others. The aim is a more limited, and thus select and focused, understanding of one's problems. Rather than uncovering deeply hidden and past motives and tracing them back to their origins in infancy, the major thrust is to deal with preconscious or conscious derivatives of conflicts as they became manifest in present interactions. Although insight is sought, it is less extensive; instead of delving to a genetic level, greater emphasis is placed on clarifying recent dynamic patterns and maladaptive behaviors in the present. Symptom relief may be an acceptable aim rather than simply a concomitant of the better resolution of conflicts. In more-integrated patients who have the capacity for greater insight, it may be a prelude to further analytic work.

Limitations A general limitation of expressive psychotherapy, as of psychoanalysis, is the problem of emotional integration of cognitive awareness. The major danger for patients who are at the more disorganized end of the diagnostic spectrum, however, may have less to do with the overintellectualization sometimes seen in neurotic patients than with the threat of decompensation from, or acting out of, deep or frequent interpretations that the patient cannot integrate properly.

Some therapists fail to accept the limitations of a modified insight-oriented approach and so apply it inappropriately to simulate the techniques and goals of psychoanalysis. Overemphasis on dreams and fantasies, zealous efforts to use the couch, indiscriminant deep interpretations, and continual focus on the analysis of transference may have less to do with the patient's needs than with those of a therapist who is unwilling or unable to be flexible.

Ms. B., a 34-year-old intelligent and verbal divorced woman, presented with complaints of being unappreciated at work. Always angry and irritable, she considered quitting her job and even leaving the city. Her social life was being negatively affected; her boyfriend had threatened to leave her because of her extremely hostile, clinging behavior (the same reason her ex-husband had given when he left her 9 years earlier after only 16 months of marriage).

Her past included promiscuity and experimentation with various drugs, and currently she indulged in heavy drinking on weekends and occasionally smoked marijuana. She had held many jobs and had lived in various cities. The eldest of three children of a middle-class family, she came from an unhappy and unstable home: her brother had been in and out of psychiatric hospitals: her sister had left home at the age of 16 upon becoming pregnant and being forced to marry; and her overly controlling parents had subjected their children to psychological (and occasionally physical) abuse, alternating between heated arguments and passionate reconciliations.

Initially, Ms. B. attempted to contain her rage in treatment, but it frequently surfaced and alternated with childlike helplessness; she interrogated the psychiatrist regarding his credentials, ridiculed psychodynamic concepts, constantly challenged statements, and would demand practical advice but then denigrate or fail to follow the guidance given. The psychiatrist remained unprovoked by her aggression and explored with her the need to engage him negatively. Her response was to question and test his continued concern.

When her boyfriend finally left her, she attempted suicide (she cut her wrists superficially), was briefly hospitalized, and on discharge was placed on monoamine oxidase inhibitors (MAOIs) for 6 months for her minor but protracted depression. The psychiatrist maintained their regular frequency of sessions despite her greater demands. Although she was puzzled by the steadiness of his interest, she gradually felt safe enough to express her vulnerabilities. As they explored her lack of full commitment to work, friends, and the therapist, she began to understand the meaning of her anger in terms of the early abusive relationship with her parents and her tendency to bring it into contemporary relationships. With the psychiatrist's encouragement she also began to seek work and relationship-oriented minigains. By the end of the second year of treatment, she had decided to remain in the city to stay at her place of employment and continue therapy. She needed to experience and practice her somewhat fragile new self, which included greater intimacy in relationships, additional mastery of work skills, and a more cohesive sense of self. (Courtesy of T. Byran Karasu, M.D.)

Supportive Psychotherapy Supportive psychotherapy aims at the creation of a therapeutic relationship as a temporary buttress or bridge for the deficient patient. As described by Karasu it has roots in virtually every therapy that recognizes the ameliorative effects of emotional support and a stable, caring atmosphere in the management of patients. As a nonspecific attitude toward mental illness, it predates scientific psychiatry, with foundations in eighteenth-century moral treatment, whereby for the first time patients were treated with understanding and kindness in a humane interpersonal environment free from mechanical restraints.

That perspective underlies such diverse developments as milieu therapy for rehabilitating chronic hospitalized patients; crisis intervention to assist otherwise functioning persons through sudden periods of great turmoil or stress; and guidance or counseling practices for children, ex-patients, and nonpatients who need interim help in social, academic, or vocational areas but do not warrant a long-term or in-depth exploration. Supportive psychotherapy has been the chief form used in the general practice of medicine and rehabilitation, frequently to augment extratherapeutic measures such as prescriptions of medication to suppress symptoms, of rest to remove the patient from excessive stimulation, or of hospitalization to provide a structured therapeutic environment, protection, and control of the patient. It may be applied as primary or ancillary treatment. The global perspective of supportive psychotherapy (often part of a combined-treatment approach) places major etiological emphasis on external rather than intrapsychic events, particularly on stressful environmental and interpersonal influences on a severely damaged self.

As a viable modality within the psychoanalytic movement, supportive psychotherapy has been described as the most ill-defined and nebulous of all psychotherapies. It is the treatment to which very difficult, characterologically or intellectually limited patients were referred when no other modality, particularly insight-oriented psychotherapy, seemed suitable. In the late 1940s Alexander and French noted the persistent tendency to differentiate between two main categories of treatment—primarily supportive methods and primarily uncovering methods. The oppositional posture in direct relation to psychoanalysis may be ascribed to its long-standing secondary place within the analytic framework, in which any deviation from the original pure form constituted a compromised or inferior approach.

Role in Analysis With the widening scope of psychoanalysis, supportive psychotherapy has emerged as a specific body of techniques. Today supportive psychotherapy is the preferred modality when primarily supportive (suppressive) strategies, such as suggestion, persuasion, encouragement, reassurance, advice, reality testing, or environmental manipulation, are used (instead of or in addition to insight-oriented strategies) and when caretaking in the form of a positive relationship provides the basic context or milieu for listening to and understanding patients (no matter what specific strategies are used). Not only the actual methods but the entire atmosphere and nature of communication are characterized by support and safety. Its modus operandi has been considered a substitute form of treatment, which temporarily gives patients what they lack or have been deprived of, and which can be therapeutically supplied or at least strengthened. The therapist, as auxiliary or substitute ego, thus offers the patient a secure and nurturing interpersonal relationship as well as specific control, direction, and counsel.

In the psychoanalytic spectrum supportive psychotherapy represents not so much a separate entity as an important emphasis within its expressive-supportive boundaries. Instead of being negatively viewed as a compromised approach for patients unsuitable for analysis, it is positively viewed as effective treatment for a broad clinical range of patients. Concern and empathy are integral to all therapeutic endeavors, but the focus here is on a stance or interpersonal bond that is affectively responsive and oriented to present reality, in which interpretive work in the broadest sense can more readily occur.

Most recent examinations of the subject explicitly recognize the complementarity of expressive and supportive modes of treatment, which can be combined, modified, and individualized according to the needs of the patient. By placing the therapeutic focus more directly on the positive or caring doctor-patient relationship (without analyzing it as a past transference manifestation) and by modifying the analytic technique and verbal communication (less silent and neutral), a supportive philosophy that emphasizes nurturance, real object gratification, and present reality can be used to enhance an analytic structure founded in psychodynamic concepts and expressive techniques.

Indications and Contraindications According to Karasu, supportive psychotherapy is generally indicated for those patients for whom psychoanalysis or insight-oriented psychoanalytic psychotherapy is typically contraindicated—those who have poor ego strength and whose potential for decompensation is high. Amenable patients fall into the following major areas: (1) persons in acute crisis or a temporary state of disorganization and inability to cope (including those who might otherwise be well functioning), whose intolerable life circumstances have produced extreme anxiety or sudden turmoil (e.g., persons going through grief reactions, illness, divorce, or job loss or who were victims of crime, abuse, natural disaster, or accident); (2) patients with chronic, severe pathology, with fragile or deficient ego functioning (e.g., those with latent psychosis, impulse disorder, or severe character disturbance); (3) patients whose cognitive deficits and physical symptoms make them particularly vulnerable and so unsuitable for an insight-oriented approach (e.g., certain psychosomatic or medically ill persons); (4) patients with low anxiety tolerance and difficulty dealing with frustration; (5) patients who lack psychological mindedness and are therefore concrete in their cognitive functioning; (6) those who have difficulty distinguishing external from internal reality; (7) patients with severely impaired interpersonal relationships; (8) those who suffer from poor impulse control and are therefore action prone; (9) patients with low intelligence and little capacity for self-observation; and (10) patients who have an extremely limited ability to form a therapeutic alliance with the therapist. These indications are contrasted with the indications for expressive therapy in Table 30.1-4.

Supportive psychotherapy is usually used when other complications enter the clinical picture, as when the primary diagnosis is physical illness but the patient needs help in dealing with disability, or for long-term mental illness, in which intermittent support and guidance in everyday living augment the control of symptoms through pharmacotherapy and provide out-of-hospital management. Regardless of the diagnosis, a supportive approach is also used in the early phases of virtually all treatments in preparation for the establishment of a firm therapeutic alliance (even when expressive treatment is the predominant mode) or intermittently whenever the patient is in danger of excessive regression.

Because support forms a tacit part of every therapeutic modality, it is rarely contraindicated as such. The typical attitude regards better-functioning patients as unsuitable, not because they will be harmed by a supportive approach, but because they will not derive sufficient benefit from it. In aiming to maximize the patient's potential for further growth and change, supportive therapy tends to be regarded as relatively restricted and superficial, and thus it is not recommended as the treatment of choice if the patient is available for, and capable of, a more in-depth approach.

Goals Supportive psychotherapy has more limited objectives than does insight-oriented psychotherapy because the patient's psychopathology and diminished psychic resources restrict the potential for major intrapsychic change or growth. Because the patient is usually in a decompensated state, all efforts (at least initially) aim at reconstituting and stabilizing the patient's function. Although the individual aims of supportive treatment vary, the general thrust is toward the amelioration or relief of symptoms through behavioral or environmental restructuring within the existing psychic framework. That often means helping patients to adapt better to problems and to live more comfortably with their psychopathology. To restore the disorganized, fragile, or decompensated patient to a state of relative equilibrium, the major goal is to suppress or control symptomatology and to stabilize the patient in a protective and reassuring benign atmosphere that militates against overwhelming

external and internal pressures. The ultimate goal is to maximize the integrative or adaptive capacities so that the patient increases the ability to cope, while decreasing vulnerability by reinforcing assets and strengthening defenses.

Major Approach and Techniques The techniques of supportive psychotherapy are designed to restore or enhance the patient's ego strength by helping the patient to control impulses through direct limit setting, to gain a more accurate picture of reality through the clarification and testing of perceptions, to sustain the adaptive structure by accepting (not analyzing or confronting) defensive maneuvers, and to develop better coping skills through direct teaching and practical advice and through the use of the therapist as a role model and as a constantly reassuring figure on whom the patient can rely. The active therapist serves as a substitute or auxiliary ego until patients can incorporate those assets in which they are deficient, either acutely or characterologically. The safety and security of a controlled and consistent therapeutic atmosphere also act as a holding environment to contain the patients' aggression and dangerous impulses as they learn better to repress or sublimate the tendencies to act out feelings rather than to verbalize them.

Transference management also differs from that of classic psychoanalysis and expressive psychoanalytic psychotherapy. Supportive psychotherapy focuses on the fostering and maintenance of a positive transference at all times and on vitiating the effects of negative transference, should it arise. Instead of facilitating a regressive transference neurosis, complete or partially controlled transference regression is contraindicated, as intense transferences that are allowed to develop in ego-impaired patients can produce sudden, turbulent reactions that have disruptive effects on both the patient and treatment. Although all attempts are made to ensure a positive transference, it is rarely analyzed as such; exceptions are those instances when positive transference may lead to acting out, which can be forestalled by drawing attention to the patient's distorted perceptions and projections. The development of negative transference is minimized by the use of highly structured interventions, with the therapist's assuming a much more direct and active approach than in expressive psychotherapy. That includes discouraging free associations, fantasies, or primary process magical thinking and continually bringing the patient back to reality and the immediate situation.

Techniques are thus designed to focus on conscious external events and on the therapist as a largely nontransferential figure. To accomplish such aims, therapeutic neutrality is judiciously suspended, with much greater direction, disclosure, and gratification offered than would be appropriate in more uncovering approaches. The therapist as a real object serves to validate the patient's reality testing, diminishes the breakthrough of regressive negative transference, and provides a reliable companion to sustain patients until they can function independently.

Limitations To the extent that much supportive therapy is spent on practical, everyday realities and on dealing with the external environment of the patient, it may be viewed as more mundane and superficial than depth approaches: because those patients are seen intermittently and less frequently, the interpersonal commitment may not be as compelling on either the patient's or the therapist's part. Greater severity of illness (and possible psychoses) also makes such treatment potentially more erratic, demanding, and frustrating. The need for the therapist to deal with other family members, caregivers, or agencies (auxiliary treatment, hospitalization) can become an additional complication, as the therapist comes to serve as an ombudsman to negotiate with the outside world of the patient and with other professional peers. Finally, the supportive therapist needs to be able to accept personal limitations and the patient's limited psychological resources and to tolerate the often unrewarded efforts until small gains are made.

Mr. C., a 50-year-old married man with two sons and owner of a small construction company, was referred by his internist after recovery from bypass surgery because of frequent, unfounded physical complaints. He was taking minor tranquilizers in increasing doses, not complying with his daily regimen, and avoiding sexual contact with his wife and had dropped out of group therapy for postsurgical patients after one session.

He came to his first appointment 20 minutes late, after having “forgotten” two previous ones. He was extremely anxious, often lost in his train of thought, and was semidelusional about his wife and sons, suggesting that they might want to have him locked up. He briefly told his life history, which included his coming from a strict and hard-working but caring middle-class family and the death of his mother when he was only 11 years old. He had joined his father's business (taking over on his father's death 2 years earlier), with both of his sons as associates. Describing himself as successful in work and marriage, he claimed that “the only test I ever failed was the stress test.”

Mr. C. explained his lack of compliance with diet restrictions as a lack of will and his constant contact with the internist as his having real physical problems, not yet diagnosed; he rejected the idea of addiction to tranquilizers, insisting that he could quit any time. He had no fantasy life, remembered no dreams, made it clear that he had entered treatment on his internist's instruction only, and started each session by stating that he had nothing to talk about.

After suggesting that Mr. C. was coming to sessions just to pass the “sanity test” and that there was no reason to have him locked up, the psychiatrist encouraged the patient to join him in figuring out the real reasons for his anxiety. Initial sessions were devoted to discussing the patient's medical condition and providing factual information about heart and bypass surgery. The therapist likened the patient's condition to that of an older house getting new plumbing, trying to allay his unrealistic fears of impending death. As Mr. C.'s anxiety declined he became less defensive and more psychologically accessible. As the therapist began to explore his difficulty in accepting help, Mr. C. was able to talk about his inability to admit problems (weaknesses). The therapist's explicit recognition of the patient's strength in admitting his weaknesses encouraged the patient to reveal more about himself—how he had welcomed his father's death and his belief that perhaps his illness was punishment. The psychiatrist also encouraged him to speak about his unrealistic guilt, at the same time helping him to recognize his suspicion of his sons as the reflection of his own wishes concerning his father and his lack of commitment to his medical regimen as a wish to die to expiate guilt. After steady urging by the therapist, Mr. C. returned to work. He agreed to meet monthly with the psychiatrist and to taper off his use of phenothiazines. He even agreed that he might see the psychiatrist for analysis in the future because his wife now jokingly complained of his obsessive dieting, his uncompromising exercise regimens, and his regularly scheduled sexual activities. (Courtesy of T. Byram Karasu, M.D.)

Empirical Research Abundant evidence indicates that psychotherapy is effective. At the end of psychotherapy, the average treated patient is better off than 80 percent of untreated patients. Moreover, the magnitude of the effect of psychotherapy is equivalent to a level that justifies the interruption of clinical trials on the grounds that it would be unethical to withhold such a highly effective treatment from patients.

When the efficacy of psychodynamic therapies is compared with that of other forms of psychotherapy, the results are generally nonsignificant. In other words, overall, psychodynamic psychotherapy is equivalent in efficacy to cognitive-behavior therapy, behavior therapy, and other standardized modalities. Dynamic therapy of depressed patients has been tested in head-to-head comparisons with cognitive-behavior therapy and has been found equally effective. Expressive-supportive therapy of opiate-dependent methadone-maintenance patients has been compared with standard drug counseling. At 6-month follow-up, patients who received the expressive-supportive therapy were doing better on all outcome measures than those with standard drug counseling. Specifically, the psychotherapy patients maintained gains that had been established during the therapy, while those who had standard drug counseling had lost many of those gains. In a large retrospective study at the Anna Freud Centre involving children with anxiety and depressive disorders, intensive psychoanalytic treatment four to five times per week was found to be more efficacious than less-intensive therapy one to three times per week. Treatment length was also positively correlated with better outcomes. Another study showed that intensive psychoanalytic psychotherapy four times a week was more effective in treating children with learning difficulties than once-a-week sessions.

In one controlled outcome study of psychoanalysis, 11 diabetic children were treated with intensive psychoanalytic psychotherapy or psychoanalysis three to four times a week during hospitalizations that lasted an average of 15 weeks. A comparison group was given only standard inpatient medical intervention. At the beginning of the study, all the children had grossly abnormal blood glucose profiles that resulted in repeated hospitalizations. The treatment group showed dramatic improvements that were maintained at 1-year follow-up. Patients in the comparison group returned to their prehospitalization level of blood glucose profiles within 3 months after discharge. In addition, the children treated with intensive psychoanalytic therapy required much less rehospitalization than the comparison group.

Other studies have looked at specific interventions made by the therapist and the link between these interventions and the outcome. Three separate studies have shown that accurate interpretations of the patient's core conflicts predict positive treatment outcome. This finding applies to outcomes within the session of the therapist's interpretation as well as those with short- and long-term follow-up.

Data are accumulating that patients with personality disorders may require more-extended psychoanalytically oriented psychotherapy to achieve significant results. A Norwegian study examined the outcomes of 45 outpatients treated with brief to moderate-length psychodynamic therapy. The patients in the sample had between 9 and 53 sessions of psychotherapy by experienced psychodynamic psychiatrists. For those patients who suffered from personality disorders (but not those without Axis II diagnoses), the number of sessions of treatment directly correlated with the degree of insight they attained 2 years after therapy and with the degree of psychodynamic change 4 years after termination of the therapy.

In another study, 25 patients with personality disorders received extended psychodynamic psychotherapy for a mean period of 25.4 months. They were then followed

up for a mean period of 5.2 years. At the termination of psychotherapy, 72 percent of the patients no longer qualified for a personality disorder diagnosis. At 5-year follow-up, 68 percent did not meet criteria for an Axis II diagnosis, suggesting that the vast majority of patients maintained treatment gains.

In an Australian study of 30 patients with borderline personality disorder, the 12-month period prior to psychotherapy was compared with the 12-month period after the patients received psychotherapy. The psychotherapy itself consisted of twice-weekly psychodynamic treatment for 12 months. There was marked clinical improvement following the psychotherapy period. Medical visits dropped from 3.5 visits per patient per month to 0.5 visits. Incidents of self-harm fell from 3.8 episodes per year to 0.8 episodes. Hospital admissions fell from 1.8 per year to 0.7, while months spent as an inpatient dropped from 2.9 to 1.5 months.

While psychoanalytic psychotherapy is often thought to be prohibitively expensive, this study of patients with borderline personality disorder suggests that in many cases providing such treatment is highly cost-effective. Borderline patients tend to use up a great many health care dollars in visits to emergency rooms, visits to other medical specialists, psychiatric hospitalization, and various diagnostic workups. There is considerable cost savings in providing regular therapy over an extended period of time. Almost all of the outcome measures included a reduction in overall expense as well.

The notion that providing extended psychotherapy reduces the need for hospital treatment is amply supported by other data as well. In 1989 the managed care organization working with CHAMPUS, which covers U.S. military dependents, expanded outpatient psychotherapy coverage under its new utilization review system. The expansion of the availability of psychotherapy resulted in a net savings of $200 million through reductions in psychiatric hospitalization over 3 years. For every $1 spent on psychotherapy, $4 were saved. Similarly, a comparison of the health care expenditures in Australia and New Zealand reflects the cost-effectiveness of having extended psychotherapy available. Unlimited outpatient psychotherapy benefits, including analysis, in Australia led to lower overall mental health care expenditures than in New Zealand, which had limited outpatient coverage. Even though there were twice as many psychiatrists in Australia as in New Zealand, Australia spent $5.7 million per 100,000 population on mental health care expenditures, while New Zealand spent $7 million per 100,000.

The early literature on the economic impact of psychotherapy stressed a concept known as *cost offset*. The notion inherent in this construct was that the addition of psychotherapy would reduce costs of other medical care. Currently, the emphasis in the mental health field has shifted away from this concept to more a complex understanding of how psychotherapy may be beneficial from an economic standpoint. One factor in this shift has been the growing awareness that the narrowly construed concept of cost offset tends to ignore the effects of psychiatric illness on the quality and quantity of life. In other words, implied in the cost-offset notion is that treating mental illnesses is beneficial primarily because it reduces overall costs of other medical care. In this regard it can contribute to the stigma associated with mental illness by suggesting that psychiatric disorders are not devastating in and of themselves and do not deserve treatment.

Another factor in the shift is that many health economists have argued that an exclusive focus on savings in treatment costs may be misplaced. Many discussions of cost-effectiveness stress costs while undervaluing effectiveness. The notion of cost effectiveness should not be regarded as meaning cheap. Rather, it is more accurately defined as meaning high value. In other words, investing a little money in psychotherapy often results in greatly improved productivity on the job and reduced absenteeism from work. Of course an insurance company or managed care company may not view disability improvements as saving them any money. However, there is a broad consensus among health economists that cost-effectiveness can only be defined in terms of the larger societal perspective rather than the narrow interests of a particular business.

In summary, the current thinking on the economic impact of psychotherapy is that it must take into account two different types of costs: (1) the direct costs accruing from actual expenditures for the delivery of the treatment and (2) indirect costs associated with disability, reduced productivity on the job, and mortality costs (i.e., suicide).

A recent German study examined treatment costs connected with 666 patients in either dynamic psychotherapy or psychoanalysis. These treatments decreased medical visits by one third, reduced lost work days by two fifths, and lowered hospital days by two thirds. Successful outcomes were linked to longer duration of treatment. A 1995 *Consumer Reports* survey of 2900 readers who had gone to mental health professionals recorded a similar finding on duration of treatment. Patients who had extended psychotherapy did considerably better than those in short-term treatment, and those whose length of therapy or choice of therapist was limited by managed care interference did worse. One of the repeatedly voiced concerns of managed care and insurance companies regarding the coverage of psychotherapy is that overutilization will result in exorbitant expenditures. Data do not support that speculation. In the most methodologically sophisticated study of the use of psychotherapy under several different copayment arrangements, the Rand Corporation found that even when psychotherapy is offered for free, only about 4.3 percent of the covered population use it and that the average number of sessions is 11.

Several efficacy studies of psychoanalysis have been done in the last half century. These studies, conducted at the Menninger Foundation, Columbia University, the Boston Psychoanalytic Institute, and the New York Psychoanalytic Institute, all suggest that patients who are suitable for psychoanalysis derive substantial therapeutic benefit. The improvement rates recorded in these studies typically range from 60 to 90 percent, and the effect sizes are significant. Three of the studies also found a correlation between length of treatment and therapeutic benefit. Like many of the personality disorder studies mentioned earlier, this psychoanalytic research typically does not use randomized assignment of patients to a treatment group and a control group. Hence one cannot state with certainty that the improvements observed in these studies are definitely related to the treatment provided, because the passage of time may have resulted in some improvements even without treatment.

While a randomized, controlled trial would be ideal, considerable methodological problems exist in trying to use randomization and controls in a study in which the treatment takes 3 to 7 years to complete. These problems include dropout rates decimating the sample, finding a suitable control group, the occurrence of life events that may influence outcome, and the artificiality in the treatment method by the introduction of a treatment manual. Because of these difficulties, a great many research efforts in the area of psychoanalysis in recent years have focused on process measures. Often these involve measuring the impact of the therapist's or analyst's intervention on certain aspects of the patient's characteristics or the process itself. In addition, some highly sophisticated single-case designs have been developed to measure not only process but minioutcomes within sessions.

The editors wish to thank T. Byran Karasu, M.D., for the material on expressive and supportive psychotherapy that appeared in the sixth edition of the *Comprehensive Textbook of Psychiatry*.

SUGGESTED CROSS-REFERENCES

The basic theory of psychoanalysis is also discussed in Section 6.1 on psychoanalysis. As a subspecialty the history and basic tenets of child psychoanalysis are detailed in Section 48.1 on individual psychotherapy of children. Greater elaboration of time-limited dynamic psychotherapy and short-term psychoanalytic approaches appears in Section 30.10 on recent methods of psychotherapy, and more specifically with regard to particular diagnoses in Section 14.9 on psychotherapy of mood disorders. Extending beyond the field of psychoanalytic psychotherapy per se, the relationship of psychoanalysis to other theories or conceptual frameworks is discussed in Section 6.4 on schools derived from philosophy and psychology, including comparison and rapprochement of Kohut and Rogers, and a discussion of integrated views of personality and psychotherapy (i.e., behaviorism and psychoanalysis).

SECTION REFERENCES

*Alarcon RD: Kandel's challenge to psychoanalysts (Letter). Am J Psychiatry *156:*665, 1999.

Brenner C: *Psychoanalytic Technique and Psychic Conflict*. International Universities Press, New York, 1976.

Chessick RD: *The Technique and Practice of Listening in Intensive Psychotherapy*. Aronson, Northvale, NJ, 1989.

Chessick RD: *Dialogue Concerning Contemporary Psychodynamic Therapy*. Aronson, Northvale, NJ, 1996.

Freud A: *The Ego and Mechanisms of Defense*. International Universities Press, New York, 1966.

*Freud S: *Standard Edition of the Complete Psychological Works of Sigmund Freud,* 23 vols. Hogarth Press, London, 1953–1964.

Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

Gabbard GO: Countertransference: The emerging common ground. Int J Psychoanal *76:*475, 1995

Gabbard GO, Lazar SG, Hornberger J, Spiegel D: The economic impact of psychotherapy: A review. Am J Psychiatry *154:*147, 1997.

Gabbard GO, Lester EP: *Boundaries and Boundary Violations in Psychoanalysis*. Basic Books, New York, 1995.

Garfield SL: *The Practice of Brief Psychotherapy*, ed 2. Wiley, New York, 1998.

Gill MM: Psychoanalysis and psychotherapy: A revision. Int Rev Psychoanal *11:*161, 1984.

Gill MM: *Psychoanalysis in Transition: A Personal View*. Analytic Press, Hillsdale, NJ, 1994.

Gray P: *The Ego and Analysis of Defense*. Aronson, Northvale, NJ, 1994.

*Greenson R: *The Technique and Practice of Psychoanalysis*, vol 1. International Universities Press, New York, 1967.

Gutheil TH, Gabbard GO: Misuses and misunderstanding of boundary theory in clinical and regulatory settings. Am J Psychiatry *155:*409, 1998.

Hartmann H: *Ego Psychology and the Problem of Adaptation*. International Universities Press, New York, 1959.

Hoffman IZ: *Ritual and Spontaneity in the Psychoanalytic Process*. Analytic Press, Hillsdale, NJ, 1998

*Kandel ER: Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited. Am J Psychiatry *156:*505, 1999.

*Karasu TB: *Wisdom in the Practice of Psychotherapy*. Basic Books, New York, 1992.

*Kernberg OF: *Object Relations Theory and Clinical Psychoanalysis*. Aronson, New York, 1976.

Klein M: *Contributions to Psychoanalysis, 1921–45*. Hogarth Press, London, 1948.

*Kohut H: *Analysis of the Self*. International Universities Press, New York, 1984.

Lohser B, Newton PM: *Unorthodox Freud: The View from the Couch*. Guilford, New York, 1996.

Luborsky L: *Principles of Psychoanalytic Psychotherapy*. Basic Books, New York, 1984.

Ogden TH: *Subjects of Analysis*. Aronson, Northvale, NJ, 1994.

Pine F: *Diversity and Direction in Psychoanalytic Technique*. Yale, New Haven, CT, 1998.

Reich W: *Character Analysis*. Touchstone Books, New York, 1974.

Sandler J: Countertransference and role-responsiveness. Int J Psychoanal *3:*43, 1976.

Stevenson J, Meares R: An outcome study of psychotherapy for patients with borderline personality disorder. Am J Psychiatry *149:*358, 1992.

Strupp H, Binder JL: *Psychotherapy in a New Key*. Basic Books, New York, 1984.

Wallerstein RW: *Forty-two Lives in Treatment: A Study of Psychoanalysis and Psychotherapy*. Guilford, New York, 1986.

30.2 BEHAVIOR THERAPY

ROLF G. JACOB, M.D., AND WILLIAM H. PELHAM, FH.D.

The term “behavior” in *behavior therapy* is defined as anything a person does. Behavior therapy involves changing what patients do, to improve their health. Behavior therapy includes a methodology, referred to as behavior analysis, for the strategic selection of behaviors to change and a technology to bring about behavior change, such as modifying antecedents or consequences or giving instructions. Today, behavior therapy has not only pervaded mental health care but under the rubric of behavioral medicine it has also made inroads into other medical specialties.

Behavior therapy represents clinical applications of principles developed in behavioral science. Behavioral psychology, or behaviorism, arose a decade after the turn of the century, in reaction to the method of introspection that dominated psychology at the time. John B. Watson, the father of behaviorism, had initially studied animal psychology. This background made it a smaller conceptual leap to argue that psychology should concern itself only with publicly observable phenomena (i.e., overt behavior). According to behavioristic thinking, since mental content is not publicly observable, it cannot be subjected to rigorous scientific inquiry. Consequently, behaviorists developed a focus on overt behaviors and their environmental influences.

Today different behavioral schools continue to share a focus on verifiable behavior. Behavioral views differ from cognitive views in holding that physical rather than mental events control behavior. However, current behavioral orientations differ in how mental phenomena are accommodated. For example, according to methodological behaviorism, mental phenomena or speculations about them have no scientific interest. According to molar or teleologic behaviorism, what seems mental becomes public once the context is considered. Such contexts can be appreciated after the individual has been observed over long enough time. Radical behaviorism treats verbal behavior as a class of behaviors with unique properties; mental events such as thinking are equivalent to the verbal behaviors that describe them.

As new behavioral principles are discovered, new clinical behavioral applications follow, and existing clinical applications may need to be reconceptualized. For example, developments in Pavlovian conditioning require evaluation of treatments developed from earlier Pavlovian models. Advances in the understanding of verbal behavior led to new approaches to treatment and increased the interface with both philosophy and social psychology. Philosophy has a long tradition in the study of logic, which is self-conscious verbal behavior. Social psychology has examined our verbal conventiens with respect to goal setting and emotions.

Behavior therapy uses a precise terminology and represents a specific view of how behavior is influenced. To understand and employ behavioral techniques, the reader needs to learn how to think in a behavioral mode. This mode supplements the normal diagnostic and psychopharmacological mode in which psychiatrists usually function.

BASIC BEHAVIORAL ANALYSIS AND TERMINOLOGY

Behavior therapy uses a precise terminology and represents a specific view of psychopathology. To understand and use behavioral techniques, the therapist must learn to think in a behavioral mode. Much of the terminology used by behavior analysts today is based on the well-known animal learning experiments conducted by B.F. Skinner and Ivan Pavlov that became prototypes for the study of operant and classical conditioning.

Operant Conditioning In Skinner's prototypical experiment, the rate of a specific behavior of a laboratory animal is modified by planned environmental manipulations. Consider the preparation in which a laboratory rat is taught to activate a lever to receive food. Even before being brought into the experimental chamber, the rat is deprived of food or water. This act is called an “*establishing operation*.” After the animal is placed in the chamber, food is delivered repeatedly in a dispenser that opens with a distinctive sound. The rat quickly learns to approach the dispenser upon hearing the sound. Next, food delivery is made contingent on successive approximations to the target behavior, lever pressing (*shaping*). Presenting a food pellet contingent on a target behavior such as lever pressing is a *consequent operation*. If the lever-pressing behavior increases because of the presentation of food, the experimenter concludes that presentation of the food pellets served as a *reinforcer* of the behavior. The process of learning from the consequences of one's behavior is *operant conditioning* or *instrumental conditioning*. The experimental setup can be further modified. For example, the experimenter can arrange to have lever pressing result in access to food only when a light in the chamber is turned on. After some time, the rat will press the lever only when the light is on. Thus, the behavior has come under *antecedent control*. Specifically, it is a form of antecedent control called a *signaling operation*–the signal being that food reward now is available.

Operations in *establishing operations, antecedent operations*, and *consequent operations* refer to behaviors performed by the experimenters rather than the subjects. Thus, these terms were constructed with the notion that the experimenters are the agents and what they do are the operations. This perspective differs markedly from lay descriptions of behavior, in which the subject is the agent, as illustrated in the following two examples:

(1) The rat presses the lever because lever-pressing behavior is reinforced by the delivery of food according to a specified schedule. The efficacy of food as a reinforcer has been established by food deprivation.
(2) The rat is hungry. It presses the lever because it believes that doing so will result in the delivery of food, and it wants to eat the food.

The lay description (2) is mentalistic (i.e., it explains the behaviors in terms of the desires and beliefs of an agent or subject. The mentalistic idiom is characterized by a mental verb such as *knowing* or *believing* followed by the word “that.” This is a grammatical form similar to indirect speech, and logicians who distinguish between meaning as sense and meaning as reference tell us that expressions following the *that* of indirect speech lose their ordinary reference (because one can no longer substitute an expression for another one referring to the same entity). Thus, the philosopher W. V. Quine cautions against the use of mentalistic language in science because it is referentially opaque. Mentalistic talk is used in everyday discourse, however, because it can be shorthand for describing complex behaviors and because such descriptions convey a sense of empathy. It can be used as a means of generating hypotheses about environmental causes. However, the lack of clarity of reference of mentalistic talk provides subterfuge for the recalcitrant witness under cross-examination. Similarly, irresolvable disagreements over the intention expressed in a behavior can contribute to marital distress. As Quine stated in his book *Roots of Reference:* “One must not mistake the familiarity of mentalistic talk for clarity.”

The increasing cultural emphasis on inner-directed goal pursuit during the recent decades, however, the behavioristic idea that an outside agent might control a person's behavior became uncomfortable. Consistent with this *Zeitgeist,* psychological terminology has become cognitive. Below are contrasting examples of cognitive and behavioral language. The cognitive statement is from *Learned Helplessness* by C. Peterson and coworkers:

(3) Individuals decide among goals and choose the most highly preferred.

The behavioral statement is from the glossary in A. C. Catania's textbook, *Learning* (4th edition):

(4) Preferences quantify the relative effectiveness of different consequences as reinforcers

Statement (3) is similar to the mentalistic statement (2). This becomes more obvious if one considers that the verbs that correspond to cognitive entities such as beliefs govern indirect “that” clauses (e.g., “belief” equals “he believes that” or “memory” equals “she remembers that”). Thus, these terms are referentially opaque. Donald Davidson, in his essay “Psychology as Philosophy,” states that mentalistic statements cannot be subject to physical laws as pursued in science: “it is only the events as described in the vocabulary of thought and action that resist incorporation into a closed deterministic system. These same events, described in appropriate physical terms, are as amenable to prediction and explanation as any.”

Pavlovian Conditioning The second prototypical learning experiment on which behavioral terminology is based is Pavlov's famous experiment in which a food-deprived dog is presented with food after sounding a bell. When the food is presented, the dog begins to salivate; this response is called the *unconditional* (or *unconditioned*) *response* to the food, the *unconditional* (or *unconditioned*) *stimulus*. When the sound from the bell is presented before the food for the first time, the dog responds with an *orienting response*. The orienting response is an alerting response reflexively elicited by new stimuli that has the effect of directing the dog's attention toward the stimulus. After repeated pairings of bell and food, the animal will begin to salivate in response to the bell. The sound of the bell, originally a *neutral stimulus*, has become a *conditional* (or *conditioned*) *stimulus* that elicits a conditional response. To determine if a conditional response has developed, a *test trial* is given in which the tone is presented but not accompanied by food. Compared with Skinner's rat, the important difference is that Pavlov's dog is receiving food because of the bell, not because of salivating. The experimenter's actions in Pavlov's preparation included presenting the conditional stimulus and the unconditional stimulus. Presentation of the unconditional stimulus is a *stimulus presentation operation*, whereas presentation of the conditional stimulus is a *signaling operation*—the bell signals the imminent presentation of food.

The process of developing the conditional response is called *classical conditioning*, or *Pavlovian conditioning*. The above example of conditional response (salivation) may give the impression that conditional responses are always the same as, or similar to, the unconditional response. In fact, early Pavlovian theory postulated that the function of the conditional stimulus was to substitute for the unconditional stimulus. However, later observations indicated that this is not true. Table 30.2-1 lists examples of other unconditional and conditional responses. For example, when the food (unconditional stimulus) in Pavlov's experiment is paired with a light rather than a buzzer, the dog often responds by licking the light source. The conditional response to stimuli signaling an aversive unconditional stimulus, such as electric shock or a predator, is freezing.

Species	Unconditional Stimulus	Conditional Stimulus	Conditional Response
Pigeon	Food	Light	Pecking at light (autoshaping)
Rat		Person feeding animals entering room	General increase in activity
		Tone	Lateral head movement (head jerking) in food area.
		Light	Standing on hind legs (rearing)
Dog	Light	Licking light source	
		Tone	Salivation
Rat	Shock	Light, tone	Conditioned suppression of positively reinforced behavior
		Light, tone	Freezing, analgesia
	Nausea, malaise induced by lithium chloride injection	Taste	Rejection of taste
Human	Nausea from cancer chemotherapy	Clinic environment	Fear, nausea
		Taste of food ingested before tx	Taste aversion
	Opioid drug injection	Drug taking environment	Hypoanalgesia

Table 30.2-1 Examples of Conditional Responses

In general, Pavlovian conditioning occurs if a personally relevant unconditional stimulus is surprising. Acquisition of conditional responses is affected by many parameters. Stimuli that were present right before the unconditional stimulus are potential indicators of future occurrences of the event. With repeated presentations, stimuli consistently present before the event but absent without the event are likely to become elicitors of conditional responses. Thus, teleologically, conditional reflexes seem to be designed to help an individual anticipate future important events.

Respondent and Operant Behaviors The behaviors elicited by the conditioned and unconditioned responses are referred to as *respondent behaviors*. Respondent behaviors often studied were glandular or autonomic responses, responses not normally considered to be under voluntary control. Behaviors that changed in response to previous consequences, *operant behaviors*, tend to involve the skeletal musculature. Operant behaviors are often considered to be under voluntary control. In ordinary language, operant behaviors are referred to as *acts*.

The distinction between operant and respondent behaviors just presented, however, is an oversimplification. For example, respondent behaviors such as freezing in response to danger involve the skeletal musculature. Similarly, operant behaviors are supported by appropriate autonomic changes; for example, animals trained to run in a treadmill for food can be expected to show the changes in cardiovascular parameters typical for exercise. The distinction between operant and respondent behaviors remains clinically useful, however, because it directs attention to the primary variables affecting a behavior. Most behaviors have both operant and respondent elements. For example, emotions are evoked by personally important antecedent events; therefore they are respondent. However, the display of emotions is also affected by consequences. An angry individual who wins a fight is likely to show anger and pick fights more often in the future. Furthermore, in some situations, specific emotions may be expressed because it is socially appropriate to do so. Medical personnel who learn about behavior analysis and behavior therapy find operant behavioral principles a major source for revised thinking.

Antecedent Operations Antecedent operations involve controlling a behavior by changing something that precedes the behavior, including establishing operations, presentation of eliciting stimuli, and signaling operations.

Establishing Operations Measures taken to increase or decrease the value (reinforcing capacity) of a particular consequence are called *establishing operations*. Food deprivation increases the reinforcing value of food-related reinforcers, and satiation reduces them. Other deprived states include thirst (water deprivation) and fatigue (sleep deprivation). Aversive stimulation is an important type of establishing operation. The presentation of aversive stimuli establishes the rewarding value of opportunities to escape. Through classical conditioning, stimuli that signal imminent presentation of an aversive event can themselves evoke escape behaviors. Presentation of the conditional stimulus is a *conditional establishing operation*. Escaping from the conditional stimulus may allow the individual to avoid the unconditional stimulus altogether. Behaviors having the latter effect are called *avoidance behaviors*. The behavioral view of anxiety disorders is largely based on what is known about escape and avoidance behaviors.

In everyday life, anything that affects the value of a consequence, even if not deliberately produced, is an establishing condition. For example, the crash of an individual's personal computer in the middle of an important piece of work establishes the value of help from the local computer expert. The computer crash is called an establishing “condition” rather than “operation” because, presumably, it was not deliberately initiated by somebody. The term *setting events* is an alternative term for establishing conditions. (However, historically, the concept of setting events was formulated before “establishing operations,” which has been in general use only for the last 10 to 15 years. “Setting event” is also used synonymously with “conditional discriminative stimuli” or “context,” described below).

In the clinical setting, symptoms from medical disorders establish reinforcers for acts to reduce them in a manner similar to the presentation of aversive stimuli. Furthermore, certain drug treatments function as establishing operations because they can change the value of a reinforcer. For example, stimulants tend to increase reinforcer efficacy, whereas antipsychotic drugs such as chlorpromazine (Thorazine) decrease reinforcer efficacy.

INDIVIDUAL DIFFERENCES AND LEARNING HISTORY Individuals differ in how much they value a specific reinforcer. Thus, individual differences affect behavior in a manner similar to establishing operations. Individual differences are related to genetic factors, learning histories, and their interaction. For example, genetically transmitted differences in the case of forming and extinguishing a conditional reflex to an aversive unconditional stimulus may manifest as *anxiety proneness*. Anxiety-prone individuals are more likely to act to escape or avoid aversive stimulation. Developmental factors also introduce variability in the way behavior is influenced. For example, children are relatively more sensitive to delay in reinforcement than adults. Certain reinforcers (e.g., toys) have greater value for children than adults. Genetic factors can influence the learning history that an individual will acquire. For example, the sex of a newborn baby instantly prompts differential child rearing practices in the parents. More subtly, certain genetically determined temperaments (e.g., “the difficult child”) may affect a parent's child-rearing practices in ways that can create problems later in life. Such a mechanism has been proposed for the etiology of borderline personality disorder.

Certain learning histories result in enduring “learning sets” that influence an individual's learning strategy in future situations. A history of unpredictable aversive stimulation from which the individual could not escape is conducive to *learned helplessness,* in which individuals respond to new aversive situations with passivity. Rats previously exposed to contingencies in which they could successfully escape from shock are relatively resistant to developing learned helplessness (“learned mastery”). More generally, a history of overcoming difficulties leads to *learned industriousness*. A history of overcoming intermittent reinforcement, extinction, or other forms of “frustrative nonreward,” is conducive to greater persistence in new learning situations. In the same vein, a learning history of overcoming delay of reward establishes an increased value for delayed consequences in the future. This is useful in the development of self-control.

Learning histories do not necessarily reflect exposure to actual contingencies. *Observational learning* (i.e., learning by observing the behavior of others) can affect the variables that reinforce behavior. Culturally transmitted values often result from such learning. Furthermore, modeling and observational learning are the primary method is for learning the academic and professional behavioral repertoires required for independent functioning. Obviously, reading and mathematical skills are essential for gainful employment. Lack of social skills can limit the rehabilitation of schizophrenics or social phobics. One of the first questions considered in behavior analysis is whether the individual has acquired the repertoire of behaviors needed to address a problem. Lack of appropriate skills is referred to as a *behavioral deficit*.

Stimulus Presentation Operations, Habituation, and Sensitization Stimulus presentations elicit various (unconditional) reflexive responses. Responses of this kind that are important for the behavior therapist include orienting responses, startle responses, and defensive or emotional behavior. Repeated presentation of the same stimulus can have two consequences: sensitization and habituation. The former involves an increase in responding, and the latter, a decrease. Slow rates of stimulus presentation are relatively more likely to result in sensitization, and rapid rates, in habituation. Habituation does not occur to highly aversive stimuli. After an intervening longer period without a stimulus, a response that had habituated is again elicited (*spontaneous recovery*).

Signaling Operations In signaling operations a stimulus is presented that on the basis of the individual's learning history anticipates (signals) the presentation of another stimulus or consequence. Signaling operations affect both operant and respondent behaviors. First consider these operations in the context of classical conditioning. The relevant operations are summarized in Table 30.2-2. Simple classical conditioning refers to the case already discussed in which a conditional stimulus (A) precedes the unconditional stimulus (US), the latter causing an unconditional response (UR). The term *reinforcement* means that the unconditional stimulus actually follows the conditional stimulus. A test for the formation of a conditional response requires presentation of the conditional stimulus without the unconditional stimulus (i.e., an unreinforced presentation). If conditioning has occurred, a conditional response (CR) is observed. An arrangement in which the unconditional stimulus follows only some of the conditional stimuli is called *partial reinforcement*. Conversely, presentation of the unconditional stimulus without a preceding conditional stimulus is called *noncontingent presentation of the unconditional stimulus*. Both partial and noncontingent reinforcement reduces conditioning. However, noncontingent unconditional stimulus presentations tend to reduce the strength of the conditional response more than a comparable partial reinforcement schedule does.

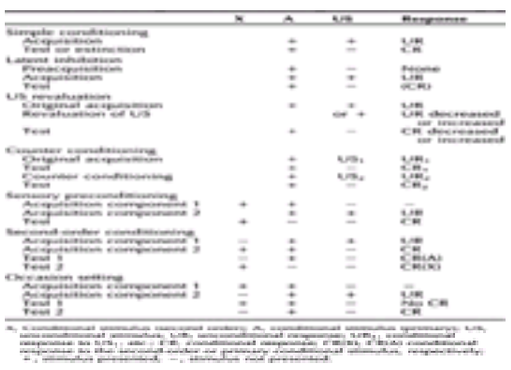

Table 30.2-2 Signaling Operations for Respondent Conditioning

INTEROCEPTIVE CONDITIONING Internal stimuli can become conditional stimuli. For example, if unconditional shock is differentially paired with two different levels of food deprivation (i.e., high deprivation vs. low deprivation), conditional freezing is observed only to the deprivation level that was conditioned. Interoceptive conditioning has a role in the genesis of panic attacks. Internal sensations such as heart palpitations or dizziness that previously preceded panic attacks can acquire the ability to trigger full-blown panic attacks.

EXTINCTION Once established, conditional responses can be modified in several ways: extinction, revaluation of the unconditional response, and counterconditioning. Extinction occurs after repeated unreinforced presentations of the conditional stimulus. After such presentations, the conditional response gradually diminishes and ultimately fails to be elicited. However, even after complete extinction, the conditional response has not been completely unlearned; it may recover spontaneously in a different context or after a longer time interval (*spontaneous recovery*). Furthermore, if the pairings between the conditional stimulus and the unconditional stimulus are resumed, reacquisition occurs more easily.

LATENT INHIBITION Preexposure to a neutral stimulus before conditioning represents a learning history in which this stimulus was present but the unconditional stimulus was absent. If an individual has been preexposed in this manner, a conditional response develops less readily to this stimulus. This phenomenon is called *latent inhibition*. Latent inhibition can be used to prevent adverse reactions to medical treatments. For example, children scheduled for their first surgery do better if they are shown a video of a model undergoing similar procedures. Similarly, preexposure of cancer patients to the chemotherapy environment may reduce the later incidence of conditioned aversive responses to this environment.

POSTCONDITIONING REVALUATION OF THE UNCONDITIONAL STIMULUS Revaluation of the unconditional stimulus involves changing the magnitude of the unconditional response. For example, treatment with opioid decreases unconditional responses to painful stimulation (resulting in devaluation of the unconditional stimulus). In this case, conditional responses to stimuli anticipating the unconditional stimulus also diminish. Conversely, the unconditional response might be increased by increasing the intensity of painful stimulation (inflation of the unconditional stimulus). In this case, conditional responses also increases. For example, patients with spider phobias avoid places in which they are likely to encounter spiders. If they overcome their spider phobia with treatment, devaluation of the spider–unconditional stimulus, should also diminish the fear of places where spiders might reside.

COUNTERCONDITIONING The conditional response can be modified by replacing the unconditional stimulus with another unconditional stimulus that elicits unconditional responses incompatible with that of the original one. For example, an aversive unconditional stimulus might be replaced with an appetitive one. In the treatment of conditional fears in children, presentation of the feared object can be paired with presentation of candy. This method is called *counterconditioning*. The new unconditional stimulus is called a *counterconditioning stimulus*. Counterconditioning thus differs from extinction in that the conditional response not only disappears, it is replaced by a new conditional response appropriate for the new unconditional stimulus. The distinction between extinction and counterconditioning is important for an understanding of various exposure treatments.

SENSORY PRECONDITIONING Signaling operations can involve more than one stimulus. The various permutations of the timing of two conditional stimuli that introduce stimulus configurations and responses can be quite complex; nevertheless, they are important for an understanding of anxiety disorders and their behavioral treatment. Depending on the details of timing, introduction of a second conditional stimulus can lead to second-order conditioning, sensory preconditioning, and occasion setting. In sensory preconditioning, two neutral stimuli, *X* and *A,* are paired without any reinforcement by an unconditional stimulus. The unconditional stimulus is then paired with *A*. If sensory preconditioning has occurred, *X* will elicit a conditional response, even though *X* has never been paired with the unconditional stimulus. Sensory preconditioning may explain why patients with a history of traumatic conditioning develop fears that seem unrelated to the original conditioning event.

SECOND-ORDER CONDITIONING In second-order conditioning, a neutral stimulus (*A*) is first paired with an unconditional stimulus, resulting in a conditional response to *A* (*CR-A*). Thereafter, during some trials, a new stimulus (*X*) is presented before *A*, the unconditional stimulus being omitted during these trials. (The unconditional stimulus is omitted to ensure that *X* is paired with *A* but not with the unconditional stimulus). After a number of such trials, presentation of *X* alone begins

to elicit a conditional response (*CR-X*). (The topography of *CR-X* is not necessarily the same as that of *CR-A*). If the conditional response to *A* is extinguished, the conditional response to *X* is too. Second- and higher-order conditioning may be a mechanism for making phobic reactions persistent. That is, a stimulus (*X*) may have become second-order conditioned to an aversive conditional stimulus (*A*). If the patient fails to escape from *X*, the first-order conditional stimulus *A* will appear and cause escape behaviors from it and in the process reinforce the second-order conditional stimulus *X*.

OCCASION-SETTING STIMULI A stimulus can also be conditioned to *moderate* the circumstances under which the conditional stimulus will elicit a conditional response. To do this, the conditional stimulus (*A*) is preceded by another stimulus (*X*) during some trials, whereas during other trials, *A* is presented alone. The unconditional stimulus occurs only in the *A*-alone trials. For example, suppose the unconditional stimulus was an aversive shock. If *A* is preceded by *X, A* signals "safety." If *A* is presented alone, it signals "danger." The conditional fear response will be elicited only in the latter context. Stimuli such as *X* that moderate the conditional response to a conditional stimulus are called *occasion-setting stimuli*. Occasion setting markedly increases an individual's ability to adapt to new environments. It may be a major mechanism for exposure-based interventions with phobias. Strictly speaking, fears are never unlearned. What is learned with treatment is the safety of the conditional fear stimulus *A* in the context of a new stimulus, *X*, where *X* might represent the delivery of treatment.

CONTEXTUAL STIMULI The function of occasion setting can involve the entire setting rather than just a particular stimulus. For example, if a conditional stimulus is paired with a shock in one setting (setting *A*) but not reinforced in another setting (setting *B*), the conditional response will develop only in setting *A*. The stimuli that define the situation are called *contextual stimuli*. For example, conditional responses may be elicited in the experimental cage but not in the home cage. A human analogue is a resident physician hearing the telephone ringing. In the context of being on call, a conditional stress response will probably be elicited. In the context of vacation, however, the stress response does not occur (except perhaps the first few times).

SIGNALING OPERATIONS IN OPERANT CONDITIONING In operant conditioning, a stimulus that marks the availability of a specific reinforcer is called a *discriminative stimulus* (often abbreviated as S_D; a stimulus signaling that reinforcement is not available is called an S_D, "S-delta"). For example, ringing of the telephone signals that picking it up would result in a conversation.

Signaling operations for operant conditioning can be as complex as those for respondent conditioning. In Table 30.2-2, analogous operant situations can be derived by placing discriminative stimuli in the "X" column, behavior in the "A" column (e.g., key pecking), and the presence or absence of reinforcement in the "US" column. The careful reader will realize that in this arrangement, the S_D or S_d is functionally analogous to occasion setting in classical conditioning; that is, discriminative stimuli set the occasion for reward of the behavior.

The operant contigency that involves a discriminative stimulus, behavior, and consequence is called a *three-term contingency*. The three-term contingency represents the traditional A-B-C (antecedent-behavior-consequence) of behavior analysis. For example, the antecedent might be a red traffic light; the behavior might be driving through the intersection, and the consequence, a traffic ticket. Adding additional levels of conditional stimuli, adds nuance to the stimulus control of behavior. The three-term contingency can be placed under conditional stimulus control, resulting in a *four-term contingency*. For example, only in certain states does a red traffic light cause a right turn to be punished. The discriminative stimulus (the red light) is now the second term in the contingency. The first term (the state) contextually defines the meaning of the discriminative stimulus and is called a *conditional discriminative stimulus*.

DISCRIMINATION, GENERALIZATION, AND CONCEPT FORMATION In both operant and respondent conditioning, stimuli similar to a signaling stimulus can elicit the responses called for by the signal. The more closely the stimulus resembles the signal, the more effective it is in this regard. The relationship between effectiveness and degree of similarity is called a *generalization gradient*. With repeated nonreinforcement of the alternative stimulus, generalization gradients tend to narrow, that is, previously effective alternative stimuli lose their effectiveness. This process is referred to as *discrimination*.

In some discrimination experiments, the alternate stimulus is so similar to the original signal that the difference can barely be detected. When this happens, the laboratory animal may display highly emotional behaviors that persist after termination of the experiment (experimental neurosis). A behavior therapy technique called systematic desensitization was originally developed on cats that had developed experimental neurosis in this fashion.

Generalization and discrimination learning have implications for treatments involving nonreinforced exposure to feared stimuli. For the treatment to have enduring effects, it is considered important to include exposure to the most significant fear (the *core fear*). If the core stimulus is not included, an apparent improvement may just represent discrimination training, leaving the core fear untreated.

In a naturalistic setting, signaling stimuli are rarely identical each time they are presented. In this case, the combined processes of generalization and discrimination may lead to an outcome similar to *concept formation*. Such learning enables us to respond appropriately to new objects. For example, based on past discriminations between many different chairs and nonchairs, one can identify a particular object as a chair, without having seen it before. Humans and animals tend to differ with respect to the information used to form concepts. Animals use principles of similarity (feature learning), whereas humans are likely to form concepts based on semantic (verbal) relations.

Consequent Operations Consequent operations involve presenting or removing a stimulus contingent upon a behavior. If, after this operation, the rate increases, the consequent stimulus is said to be a reinforcer. Reinforcers can be positive or negative. In *positive reinforcement*, a stimulus is added (e.g., food, water, or a drug). In *negative reinforcement* a stimulus that had been previously presented as an establishing operation is removed.

The definition of *punishment* is analogous to that of reinforcement, except that the direction of the behavior change is reversed. "Positive punishment" refers to a consequence that involved adding a stimulus that decreased the rate of the behavior. Negative punishment refers to the operation of removing access to a positive reinforcer. Alternative terms for this latter type of punishment is "fine," or *timeout from positive reinforcement*. This mode of punishment is the one most often used in clinical practice.

Knowing What Will Be Reinforcing Ahead of Time According to the definitions just presented, a reinforcer can be identified only after the behavior has increased. It is tempting to try to avoid this circularity by finding an alternative definition that does not refer to behavior change. One such alternative is to consider "positive reinforcement" as "something pleasurable" or "rewarding." With this definition, one could argue that depriving an animal of enough food to result in significant weight loss introduces an aversive state from which presentation of food would provide relief. Contingent food could be considered equivalent to negative reinforcement (i.e., removal of hunger). Positive reinforcement, in this new context, would be equivalent to gaining access to dessert after completing the main course. Recent neurobiological findings suggest that the brain circuitry involved in food reward differs depending on whether the animal is deprived or nondeprived. The same differences are observed in the case of reward by opiate drugs in nondependent and dependent organisms. To use a mentalistic analogue, "needing" and "liking" may be different phenomena.

Even if the circular definition is retained, certain principles can be used to predict what consequences are likely to be reinforcing. These include the transsituational principle, response deprivation, natural reinforcement, and conditional reinforcement.

TRANSSITUATIONAL REINFORCERS A consequence that has been reinforcing in one situation might be reinforcing in another. Such transsituational reinforcers include food, certain tastes, affection, toys, and devices producing sensory stimulation (such as video games). Attention from others is an important transsituational reinforcer. Attention can serve as a reinforcer even if the resulting interaction is negative.

RESPONSE DEPRIVATION Another principle involves providing or restricting access to another behavior to decrease it below its spontaneous rate. A common example is providing access to TV after homework is completed. For this to work, TV watching must first be restricted to a level below the child's unrestrained TV-watching habits. Therefore, this principle is referred to as response deprivation, or the *Premack principle,* after the originator of a similar method.

NATURAL REINFORCERS As part of social practice, certain consequences regularly follow certain behaviors. For example, a natural reinforcer for artistic behaviors might be public display of a work of art. Natural reinforcers are often contrasted with contrived reinforcers; for example, earning points for art work is a contrived reinforcer.

CONDITIONAL REINFORCERS Stimuli that signal the availability of a reinforcer (discriminative stimuli or conditional stimuli) can themselves become reinforcers. Promises, points, or tokens exchangeable for good or activities are examples of conditioned reinforcers from the clinical arena or everyday life. Their value depends upon their relationship with the primary reinforcer that they signal. Certain conditional reinforcers signal the availability of a number of different primary reinforcers.

Such reinforcers are called *generalized reinforcers,* of which money is one.

Contingency Learning: Contacting New Contingencies How does change in reinforcement produce a change in behavior? The mechanism by which behavior changes after reinforcement is considered analogous to the evolution of species by natural selection; that is, a reinforcing consequent event will select the behavior that preceded it. The initial contact with a contingency may be quite accidental and involves temporal contiguity between the response and the consequence. However, contiguity is not equivalent to causation; an organism's adaptation to the new situation should be enhanced if it can discriminate between a mere contiguity and a causal relationship. Some organisms, including human newborns emit behaviors that have the effect of differentiating between causal and mere contiguous relations between the behavior and the consequence—*contingency-detecting behaviors*. Essentially, contingency-detecting behaviors involve increasing the variability of the behavior to include both very high and very low rates. The low rate tests whether the consequence can occur in the absence of the behavior; the high rate tests whether the behavior is sometimes not followed by the consequence. Increasing the variability of behavior by generating and examining the consequences of a wide range of behaviors is the foundation of a clinical technique called *problem solving*, discussed below.

Even after a contingency has established itself and remains constant, changes in rate of behavior during a session will vary systematically, with an initial increasing trend followed by a decreasing trend. These variations are attributed to the combined effects of habituation and sensitization to the reinforcer. Sensitization tends to take place early in the session, resulting in an increase in behavior, whereas habituation becomes maximal later in the session. A high rate of reinforcement favors habituation; and a low rate favors sensitization.

Intermittent Reinforcement Schedules It is usually easier to shape a new behavior if every instance of the target behavior, or approximations thereof, is reinforced (continuous reinforcement). However, once established, the reinforcement schedule can be thinned (intermittent reinforcement). Behaviors maintained on an intermittent reinforcement schedule tend to be more resistant to extinction than those on a continuous schedule. A large number of different reinforcement schedules have been studied. The most common schedules involve presenting the reinforcer after a specific number of responses (ratio schedule) or after a specific amount of time (interval schedules). These schedules can be either variable or fixed. The variable interval schedule, for example, results in a steady rate of behavior without pauses. The variable ratio schedule often results in very high rates of behavior. In addition to the basic schedules of reinforcement, certain specialized schedules have clinical applications. A schedule in which reinforcement is withheld if the response rate exceeds a certain value is called *differential reinforcement of low rate*. An intervention in which reinforcement is withheld for one behavior and allocated to a different behavior is called *differential reinforcement of other behavior*.

Schedules can be combined to run concurrently or sequentially. They can be signaled by discriminative stimuli, which makes it possible to arrange for an animal to "choose" between different schedules. Reinforcers known to maintain a behavior can also be presented without first requiring a response. Such noncontingent reinforcement decreases the probability of the response.

Multiple Behaviors Competing for a Reinforcer The relative rate of one behavior compared with another tends to be a function of the relative amount of reinforcement allocated to each. This principle is referred to as the "matching law." Of particular interest is the rate of a target behavior relative to the rate of all alternative behaviors. For behaviors maintained by a variable interval schedule, this can be predicted by the formula

$$B = K \times R/(R + R_e)$$

where B is the rate of the behavior (e.g., lever pressing), K is the maximum possible rate of the behavior, R is the rate of contingent reinforcement, and R_e is the rate of reinforcement behaviors other than B. Figure 30.2-1 shows a family of functions relating B to R. Each curve represents different values of R_e or K. As can be seen, each curve shows that increases in the contingent reinforcement rate R are associated with negatively accelerated increases in B, until an asymptote K is reached. The parameter K often reflects specifics of the behavior and recording device (e.g., lever pressing vs. key pecking) and the amount of effort involved in producing B.

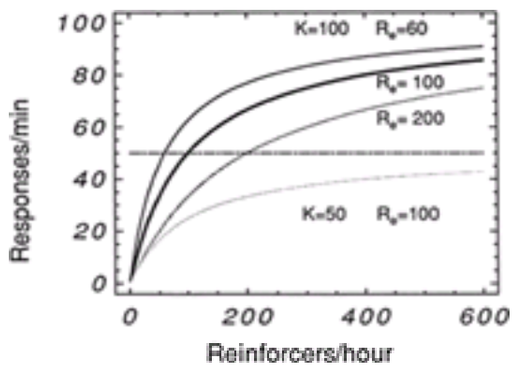

FIGURE 30.2-1 Rate of behavior as a function of R (reinforcement rate), R_e (noncontingent reinforcement), and K (maximal response rate).

The effect of the reinforcement rate R is moderated by R_e, the frequency of noncontingent reinforcement. The larger the R_e, the larger R must be to maintain a given rate of behavior. Environments have been characterized as rich (R_e large) or lean (R_e small). In rich environments, changing the amount of contingent reward, R, will have less effect than in lean environments.

Obviously, one way of increasing a behavior is to increase the reinforcement rate R. If we cannot change R, we can instead increase R_e (i.e., make the environment leaner). A third way of increasing a behavior is to increase K. This can be accomplished by reducing the effort needed to produce the response. The formula above also points toward ways of decreasing a behavior. One way is to stop providing contingent reinforcement ($R = 0$, i.e., extinction). If the reinforcer is stopped abruptly, however, the behavior may increase before it decreases. During this phase, the individual may exhibit emotional behavior indicating frustration. This phenomenon is an example of "frustrative nonreward." An alternative to extinction is increasing the rate of noncontingent reinforcement (increasing R_e), that is, creating a richer environment. Another way of decreasing behavior is to make it more effortful (decreasing K). For example, self-injurious face slapping has been reduced by attaching a weight to the person's wrist.

Another approach to decreasing a behavior is punishment, that is, presenting an aversive stimulus contingent on the behavior. However, like frustrative nonreward, punishment tends to have a number of undesirable effects, including aggression and suppression of concomitant positively reinforced behaviors. Furthermore, punishment attempted for defensive behaviors (i.e., behaviors already under aversive control by negative reinforcement) often further increases the behavior. The punishment method associated with the fewest problems is timeout from positive reinforcement.

Immediate and Delayed Reinforcement In general, the more closely they follow the act the more effective reinforcers tend to be. The strength of a behavior is inversely related to passage of time to the reinforcer. One reason why video games are so reinforcing is that the consequences of the individual's actions appear within fractions of seconds. Certain problems can be attributed to patients being unduly influenced by immediate small rewards at the expense of larger delayed rewards. Examples of such behaviors include overeating, binge eating, risky sexual behaviors, and addictive behaviors.

One puzzling aspect of such impulsive behaviors is that individuals may firmly decide not to engage in them, only to change their mind when the opportunity becomes imminent. The mirror image of impulsive acts is impulsive procrastination. Here the short-term reward comes from escaping or avoiding an unpleasant or difficult task or from timeout from other more rewarding activities in which the individual could have engaged. Why do people change their mind at the last minute and perform impulsive acts that they had resolved to eliminate? Why do people tend to make New Year's resolutions and not keep them? A behavioral answer to this question is shown in Figure 30.2-2. The reinforcing effect of a reward diminishes proportionally with the inverse of the time delay, a hyperbolic function. Figure 30.2-2 shows two hypothetical operant discount functions with identical parameters; one for a large later reward and the other for a small reward occurring relatively soon.

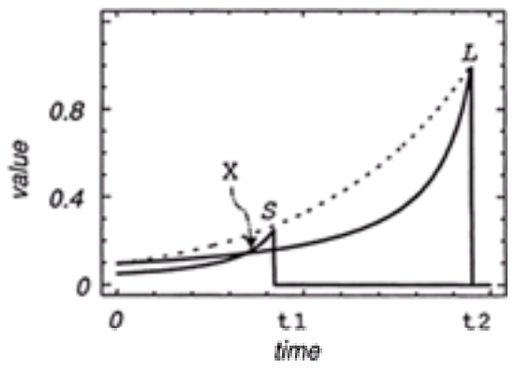

FIGURE 30.2-2 Hyperbolic discount functions associated with reinforcement delay.

Assume that the individual has a choice of two mutually exclusive behaviors. One leads to a relatively small reward *(S)* after a relatively short delay (time t_1). The other leads to a larger reward *(L)* after a longer delay (time t_2). For example, for someone who is reporting an intention to stop smoking, *S* might represent smoking a cigarette at a party scheduled to occur at t_1, and *L* the sum of all long-term benefits of not smoking. At time 0, the discount curve from *L* has a higher value than the one from *S*. Consequently, the person reports an intention not to smoke. As time approaches t_2, however, the discount function for *S* increases faster than the one for *L*, and they cross over at point X. At this point, the person changes his or her mind and smokes, if a cigarette is available, perhaps rationalizing that "I will stop at the next party."

L and *S* can also represent aversive consequences. For example, *L* might be a deadline for a term paper or book chapter and reflect the aversive consequences of missing the deadline. *S* represents the rewarding value of escaping from a difficult task, or the cost of giving up other rewarding activities (e.g., watching television) to start working on the project. Again, as the procrastinator approaches t_1, the motivation to procrastinate becomes more salient. At point X, the motivation to delay becomes stronger than the motivation to start, and the procrastinator puts the project off for the next day. Other examples of impulsive behaviors or delays include staying in bed after the alarm clock has rung, going to a movie instead of studying for finals, or making excessive credit card purchases.

Discount reversals are not limited to operant contingencies. To understand them better, consider some examples from nature. Alert students of high school physics should remember that the apparent brightness of a shining object is related to the square of its distance. Therefore, the brightness of a nearby object increases faster when approached than that of a distant object. If a strong light is far away and a weak one close by, at some point the latter may appear brighter than the former. Similarly, suppose you drive on a road in a valley and see a telephone pole in the distance, in front of a mountain. The top of the telephone pole projects somewhere onto the greenery of a foothill in front of the mountain. As you get closer, however, the pole's visual angle (and apparent size) increases more quickly than that of the mountain behind it. At a specific point close to the pole, the top of the pole will project onto the sky above the mountaintop behind it. This point is the equivalent of point X in the figure, the point of discount reversal.

Discount reversals do not normally occur in economic decision making, because the curve (e.g., compound interest) relating future economic value to the passage of time is exponential. Exponential curves that reflect the same interest rate starting at various time points do not cross over. Figure 30.2-2 also shows an exponential curve that connects *L* with the value of the operant curve from *L* at time zero. An exponential curve from *S* would run parallel to the one from *L* rather than crossing it. For this reason, rational economic decisions do not change as a function of time. It is the discount reversal that occurs with the hyperbolic function that explains many impulsive behaviors.

The model just outlined suggests several approaches to impulsive behaviors. First, any plan to deal with the problem should be made before the crossover point. Reversals of decisions made after that point are suspect. Efforts to prevent relapse of alcoholics include developing the patient's ability to identify such postreversal-point decisions. Such decisions have the appearance of being quite innocent. For example, a compulsive gambler decided to enter a casino "to obtain change for the parking meter." This decision resulted in a relapse of gambling behavior. The term "apparently irrelevant decisions" has been coined to help patients avoid such traps.

Second, measures can be taken to eliminate access to *S*, a method called *precommitment*. This method involves varying degrees of restraint. For example, the smoker discussed earlier may avoid temptation by not going to the party. Patients in weight reduction programs are instructed to plan their meals when they are not hungry and to keep the amount of food in the house at a minimum. Precommitment can also involve physical restraint. This was the method used by Odysseus, when he had himself tied to the mast before sailing past the sirens. Legal restraint was used in a program for drug addicts. Patients wrote incriminating letters and deposited them with the staff. In case of a positive drug screen, the staff was authorized to send the letter to authorities who would have interest in the matter.

Third, *S* could be delayed to a later time (to the right in Fig. 30.2-2). This is the natural consequence of procrastination. When t_1 is moved to a later date by procrastination, the escalating fear of missing the deadline (i.e., the curve from *L*) may finally completely overshadow the aversive consequences associated with getting started. Because of this phenomenon, procrastination tends to be self-limiting. Impulsive behaviors, on the other hand require specific action. Impulsive buyers may reduce their purchases if they establish a habit of first "consulting the pillow" (i.e., delaying the purchase by 1 day by getting a good night's sleep). Similarly, patients with bulimia are advised to remove food from the table as soon as the meal is finished, thus providing a small delay before food is accessible again.

Fourth, the value of *S* relative to *L* could be reduced by modifying its establishing conditions. Binge eating, for example, is in part related to self-imposed dietary restrictions that induce a food-deprived state. These patients can reduce *S* by normalizing their diet. Another way to decrease the net effect of *S* is to increase the noncontingent reinforcement rate, thus creating an environment rich in reinforcement (see section on consequent control). For example, treatment of binge eating includes advising patients to expand their repertoires of rewarding activities, including recreation (i.e., increase those behaviors they like to do rather than ought to do). In the case of procrastination, the punishing value of *S* might be reduced with a technique similar to imaginal exposure called *process simulation*. This involves the patients imagining, or mentally rehearsing, the specific activities involved in working on the project.

An examination of study behaviors of college students before a test found that students assigned to 5 minutes of process simulation started studying earlier, felt less anxious before the test, and had better scores than students assigned to outcome simulation. Process simulation involved the students imagining the specific steps needed to study for the test, including preparing a suitable work area, turning off extraneous stimulation, and going through the study materials.

A final set of possibilities for interventions can be derived from the properties of the delayed consequence, *L*. For example, patients might behave as if they do not sufficiently value *L*. To remedy this, measures are taken to increase the value of *L. L* could be small because the patient has not been informed of its value. Some patients may simply be misinformed, for example, of the value of mammograms or the dangers of smoking or drinking alcohol. An obvious way to increase *L* is to provide appropriate information. Besides increasing *L*, the discount function can be modified, which is mathematically equivalent to reducing the delay in *L*. Obviously, if the function falls less rapidly, the value of *L* at the starting point will be higher. The value of the discount function can be changed by experience. Learning to "delay gratification" in this way is a part of child rearing with demonstrated implications for later adjustment. Even among adults, older individuals tend not to discount delayed rewards as highly as younger individuals. The delay of *L* can be reduced by formulating subgoals or segmenting the time in other ways. In the laboratory, the delay between *S* and *L* can be effectively shortened by the use of conditioned reinforcement or other stimuli that mark, or bridge, the delay interval. This is what may be involved in mentoring. Mentors provide bridging stimuli that help guide their younger colleagues toward distant professional rewards.

Modeling and Observational Learning Learning new behaviors by means of observations is called "observational learning." The individual being observed is called the "model." The model does not have to be a live person; written or verbal instructions represent a special case of modeling sometimes called "symbolic modeling." Modeling can also be done with videotaped or filmed models. In behavior therapy, modeling is the primary method of teaching new behaviors such as social skills. Combined with contingency management, modeling has been used successfully to teach language skills to autistic children. Videotapes of a child undergoing preoperative preparations can be used to decrease preoperative distress in children.

As a consequence of observational learning, the individual matches or imitates the modeled behavior. Given the central importance of modeling in the teaching of new behaviors in humans, it may seem surprising that early behaviorists did not show much interest in it. One reason may have been that Watson, the father of American Behaviorism, during his early career as an animal psychologist, spent two unsuccessful years trying to train cebus monkeys to imitate such behaviors as using a rake

to pull a grape closer to the cage. The monkeys failed to imitate behaviors modeled either by another monkey or by Watson himself:

“Finally, I began showing the monkeys how to draw in the food. I would wait patiently until I apparently had their attention, then slowly hook the T-piece around the grape and slowly draw in the food. The moment the grape rolled near, the animal began to strain at the tether and to attempt to grasp the grape with the paw. The situation was again immediately as before and the rake left near. Never once did any one of the animals push out the rake, hook the blade around the grape, and pull them in, nor did they in any other way ever show any signs of perceiving the relation which ought to exist between the two objects.”

Vicarious Conditioning Besides providing information about how to perform a new behavior, modeling can also provide information on the consequences of behaviors already in an individual's repertoire. After observing the consequences of the model's behavior, the observer is more likely to engage in the same behavior, if the model was reinforced. Such changes in behavior in response to consequences observed with others are said to be a result of “vicarious conditioning.” Thus, there is “vicarious extinction” “vicarious punishment,” “vicarious reinforcement,” etc.

For example, in a street crossing in which the traffic light indicates “don't walk,” one individual walking across the street may prompt others to follow suit. Aggression is another example. The fact that modeling of unpunished aggressive behavior increases aggressive behaviors in the observers continues to fuel the debate about the role of violence on television or in the movies in modeling violent behaviors.

Behavior Systems Principles of operant or classical conditioning are based on the assumption that behaviors and reinforcers can be arbitrarily paired. However, in actuality, response-consequence relations are subject to a number of task-specific constraints. For example, it would be difficult to train a hungry cat to emit grooming behaviors with food reward, but quite easy to shape begging behavior. Similarly, it is easy to condition fear responses in monkeys to neutral stimuli resembling snakes but difficult to do so to neutral stimuli resembling flowers. The principles of Pavlovian conditioning also do not easily predict the nature of the conditional response that will occur for a given conditional or unconditional stimulus (Table 30.2-1).

The topography of the conditioned response can vary with the conditional stimulus–unconditional stimulus interval. In an experiment with laboratory rats, the conditional stimulus consisted of a rolling ball that signaled presentation of a food unconditional stimulus. When the interval between the two was short, the conditional response involved approaching the food tray and digging in it. When the interval was long, however, the conditional responses resembled the behaviors a rat normally emits after it has caught an insect of prey. The conditional responses included chasing the ball, seizing it, carrying it around, gnawing at it, and only reluctantly giving it up to go to the feeder.

Ethological observations suggest that certain biologically important establishing conditions (e.g., hunger) lead to specific combinations of observing and motor responses that combine into fixed sequences of behavior (e.g., looking for food, chasing, eating). These sequential behavioral patterns are called “behavior systems.” More specifically, W. Timberlake defines a behavior system as

> An organized set of sensory perceptual mechanisms, related response components, and motivational processes that evolved because they were correlated with better service of a particular function. In combination with current stimulus conditions, a behavior system can be used to predict and account for the form of responding that emerges.

A feature of this approach that has not been encountered above is stimulus-selection behavior. For example, a hungry traveler will scan the environment for restaurant signs.

The major behavioral systems are organized around the general functions of feeding, reproduction, and defense. In addition, a dominance-submission dimension organizes the social behavior of primates. In humans, the dominance-submission system, along with the reproductive system, has evolved to form various modes of affiliative behavior. A behavioral system has a temporal dimension and a hierarchical dimension. The temporal dimension reflects the specific sequence of behaviors involved in fulfilling the function of the system. For example, appetitive systems, such as feeding or mating, begin with a general search for a suitable object, pursuing and engaging the object, and ultimate consummation or consumption. For the defensive behavior system, the temporal dimension reflects the degree to which danger is imminent, that is, the likelihood of being eaten by a predator. The dominance system might involve behavioral modes of threat and appeasement.

The hierarchical dimension reflects the degree of abstraction used to describe the behavior. The most concrete level contains specific *action patterns*. Sequences of action patterns combine into *modules*, and sequences of modules make up behavioral *modes* (e.g., the general search, chase, and consumption modes of the predatory food acquisition subsystem). The most general level, the *system level*, defines the overall biological or evolutionary task pursued. Table 30.2-3 shows a list of modules and action patterns in the predatory feeding subsystem of the rat. It begins with a travel module, in which environmental stimuli elicit various acts of exploratory locomotion. It normally ends with the consumption module, in which the sensory aspects of food elicit eating behavior.

Module	Activating Stimuli	Act
Travel	Features of general environment	Locomoting
Chase	Distal moving object	Tracking Chasing
Capture	Proximal moving object	Grabbing with mouth and paws
Test	Tactile stimuli	Holding Gnawing
Ingest	Food, taste	Chewing Swallowing
Reject	Bad taste	Spitting out

Adapted from Timberlake W, Silva F: Observation of behavior, inference of function and the study of learning. Psychonom Bull Rev 1:73, 1994.

Table 30.2-3 Examples of Modules Within the Predatory Subsystem of the Rat

The behavioral systems approach can explain the effect of the conditional stimulus–unconditional stimulus interval on the differing topographies of the conditional responses that emerge. Presentation of food defines the endpoint of the feeding system. A long interval elicits conditional responses that are characteristic of a module located relatively early in the temporal dimension of the system. A short interval can be expected to prime behaviors characteristic of later modules of the system. Consistent with this notion, in the case of the rat and the rolling ball conditional stimulus, the chase behavior occurred with the long interval, whereas digging in the food tray and gnawing occurred with the short interval.

Defensive Behavior System The modes in the defensive behavior system include the preencounter, postencounter, and circa-strike modes (Table 30.2-4). These are important for our understanding of phobias and posttraumatic stress reactions.

Mode	Activating Stimuli	Responses
Preencounter	Potential danger	Reorganization of meal patterns
Postencounter	Detecting a predator	Freezing, potentiated startle reflex, opioid analgesia
Circa-strike	Contact with predator	Threat, attack, escape, nonopioid analgesia

Adapted from Fanselow MS, Kim JJ: Neural organization of the defensive behavior system responsible for fear. Psycho Bull Rev 1:429, 1994.

Table 30.2-4 Modes Within the Defensive Behavior System

The preencounter mode is invoked when the organism enters a context that previously was associated with danger but which now lacks signals of imminent danger. That is, the environment is one in which predators are likely but none has actually been identified. The behaviors in this mode include avoidance behaviors, which in the rat lead to alterations in meal patterns. The postencounter mode occurs when a conditional stimulus for an aversive event is presented. In the natural environment, this would correspond to the moment when a predator has been identified. The characteristic behavior for this mode is freezing. Opiate-mediated analgesia is a component of the freezing response. The circa-strike mode is triggered by actual physical contact with the predator. In the laboratory, the circa-strike mode can be evoked by aversive stimulation, such as electric shock. Circa-strike behaviors include threat, counterattack, or escape. The animal that survives the circa-strike experience usually goes back into postencounter mode. Thus, freezing also occurs after a traumatic event.

Behavior Systems, Posttraumatic Stress, and Phobias The behavior systems approach provides a useful perspective for anxiety disorders in humans. For example, in war settings, behaviors appropriate for combat have been described as governed by a survival mode that episodically reappears in the civilian lives of veterans with combat-related posttraumatic stress disorder, manifesting as inappropriate aggressive behaviors. Survival mode behaviors may be equivalent to the circa-strike behaviors of the defensive system. Other symptoms of posttraumatic stress disorder (i.e., freezing, numbing, or dissociation) seem to be reminiscent of the postencounter mode. Parallels have also been drawn between the defensive behavior system and animal phobias. However, some animal phobias (e.g., fear of spiders, cockroaches, rats, maggots) go with the emotion of disgust rather than fear. Disgust fits better with rejection of food (i.e., poison). This function is better included in the appetitive behavior system as a postingestion module (Table 30.2-3).

As anticipated by William James, agoraphobic avoidance might be functionally equivalent to preencounter behaviors.

> Again, take the strange symptom which has been described of late years by the rather absurd name of agoraphobia. The patient is seized with palpitations and terror at the sight of any open place or broad street which he has to cross alone. He trembles, his knees bend, he may even faint at the idea. Where he has sufficient self command he sometimes accomplishes the object by keeping safe under the lee of a vehicle going across, or joining himself to a knot of other people. But usually he walks around the sides of the square, hugging the houses as closely as he can. The emotion has no utility in a civilized man, but when we notice the chronic agoraphobia of our domestic cats, and see the tenacious way in which many wild animals, especially rodents, cling to cover, and only venture on a dash across the open as a desperate measure—even then making for every stone or bunch of weeds which may give a momentary shelter—when we see this we are strongly tempted to ask whether such an odd kind of fear in us be not due to the accidental resurrection, through disease, of a sort of instinct which may in some of our ancestors have had a permanent and on the whole a useful part to play?

In primates, a *dominance-submission social system* determines an individual's social status within a group. Social phobic symptoms appear to be related to this system. The main unconditional stimulus eliciting social phobic responses, including shame and embarrassment, is a prolonged stare by a more dominant member of the group.

Implications for Behavior Therapy The behavior systems approach has implications for behavior therapy that would not be immediately obvious from an operant or Pavlovian perspective.

MODE-SPECIFIC STIMULUS CONSTRAINTS A characteristic of behavior systems is that eliciting stimuli selection depends on the prevailing behavioral mode. Such constraints tend to be more pronounced late in the temporal dimension and seem to be particularly prominent in the sexual behavior system. The human counterpart to sexual behavior systems has been called a sexual script; *scripts* refer to plans or guidelines of what to do when being sexual. If scripts differ too much between partners, sexual dysfunction may develop.

PROGRESSING ALONG THE TEMPORAL DIMENSION Gaining access to the next module serves as a natural reinforcer for the behavior in the current module. For example, catching the prey is the natural reward for the chase. A high degree of deprivation is not needed for such natural reinforcement, as owners of satiated bird-chasing cats can verify. Incidental teaching is a method in which a patient or child's access to the next module is contingent on completing some small task.

The temporal dimension can be violated, either by moving along too quickly or by moving in the reverse direction. Such violations cause problems. A human example of an endpoint being scheduled too soon is proposing marriage on the first date. Conversely, forcing a move from a later module to an earlier one (i.e., further away from the ultimate goal) would be going against nature and the equivalent of an aversive event. For example, few individuals would enjoy having to go to the grocery store (a general search act) just after having served (but not yet eaten) a meal.

PRIMING A DIFFERENT SYSTEM Some of the most dramatic behavior changes occur when a switch is induced from one system to another. This can be done by presenting a salient stimulus (priming) from the other system. For example, a barking dog can be quieted by presenting food—the food being a priming stimulus that evokes the feeding system. A human example of this might be spending the proverbial first date buying a dishwasher or having a parent enter the room during a romantic interaction. Similarly, presenting stimuli priming an entertainment mode of affiliative behavior may derail development of aggressive dominant behaviors.

Behavior Systems and Molar Action Patterns Describing behavior as involving sequences of subroutines organized into larger functional units is not unique to the behavioral systems approach. In a form of contemporary behaviorism called *molar behaviorism*, behavior descriptions can be arranged on a continuum from “molecular” to “molar.” Sequences of molecular actions combine into “action patterns,” which can be part of still larger units of behavior described at increasingly molar levels of abstraction. At the highest level, such patterns can be described as transitioning into “values.”

A behavioral fragment can have different functions depending on the overall context in which it is embedded. Identifying this requires extending the observation time. For example, “swinging a hammer” can be part a number of different contexts (e.g., “hammering a nail,” “smashing a radio,” or “engaging in aerobic exercise”). Hammering a nail, in turn, can be part of “building a house” or “putting a sign on a tree.” “Building a house” can be an instance of “providing shelter,” whose significance might be “supporting one's family.” The latter might be regarded as one of the functions of “being a good parent.”

The molar behavioral view includes an interesting approach to the phenomenon of time. Acts such as singing a song or going to a movie tend to have natural endpoints, and such acts can represent one unit of behavioral time. The occurrence of reinforcement or other natural endpoints segments a person's time into units of behavioral time. For example, molar behavioral analysis of gambling behavior suggests that the stream of gambling behaviors is punctuated by periodic “settling of accounts.” “Compulsive” gamblers tend to settle accounts only after winning a bet, never after loosing a bet. For this reason, they will leave the gambling table only after a win. It can be shown mathematically that this pattern of stopping after a win increases losses. An approach in which the gambler stops after a certain number of bets is not associated with a similar bias toward loss.

A consequence of the molar behavioral analysis of time is that a behavior described at a molar level involves fewer behavioral time units than the same behavior considered at a molecular level. At very high molar levels (e.g., “being free from sin”), accounts may never be settled, in which case a unit of behavioral time is lifelong. The self-control paradigm discussed above described how long-term desirable consequences are offset by immediate but incompatible reinforcers. In a molar approach to this problem, the duration to L, in behavioral time, can effectively be functionally shortened by considering the molar (meaningful) aspects of the behaviors that precede L. That is, the patient needs to engage in behaviors that represent alternatives to—and, ideally, are incompatible with—the impulsive behavior. Furthermore, these alternative behaviors need to be considered at a molar level.

A patient with an alcohol problem joined Alcoholics Anonymous and successfully completed a year's abstinence. A former amateur pilot, the patient had resumed a hobby of building airplanes in his garage. Because drinking was not compatible with building and flying airplanes, the molar airplane-building pattern may have contributed to his success.

VERBAL BEHAVIOR: MAJOR DETERMINANT OF HUMAN BEHAVIOR

It has been speculated that phylogenetically, brain structures supporting speech and sequential motor behaviors tended to develop in synchrony and were adjacent to each other. The ability to manufacture and make use of tools preceded or developed in tandem with speech, both requiring sequencing of behaviors in time. The

sequencing of behavior can also occur across individuals (e.g., taking turns at communal meals). Complex cooperative behaviors (e.g., having one watch out for enemies while the other gathers food, or even mass migration) could be coordinated with the help of verbal behavior defining immediate or extended time frames. The relevant verbal behaviors may have included assertions (e.g., the location of food or sources of danger), commands, and promises (e.g., “I will share the food if you follow the command”). From assertions and promises, mentalistic concepts of beliefs and intentions may have evolved.

With (perhaps) the exception of observational learning, the principles of behavior discussed above apply equally to animals and humans. What is unique to humans is their facility with a special form of communication, verbal behavior. A verbal description of a situation transforms a multichannel, instantaneous sensory experience into output involving a single channel extended in time. The fact that verbal behavior is extended in time may be why humans developed the capability to plan sequences of behavior. It has been speculated that phylogenetically, verbal behavior emerged together with tool use and group living. Both of these activities require sequencing of behaviors, such as taking turns when eating a communal meal, or using tools sequentially (hammer, anvil, and raw material) when building a shelter or making hunting gear. Brain structures involving speech and sequential motor behavior are closely linked.

Communication via Evocative Signals In primates, communication occurs via vocalization visual displays, and facial emotional expressions are part of this communication system. This mode of communication can elicit sympathic emotional states in the recipient. For example, monkeys who have been reared apart from other monkeys tend not to be fearful of snakes but can develop such fears after observing other monkeys acting fearfully with snakes. In humans too, the visual and vocal channels are important for transmission of emotionally tinged information. For example, observation of facial expressions and other nonverbal behaviors constitute one channel for developing empathy with a patient. (The other channel for empathy is derived from information about a patient's situation, i.e., contingencies of reinforcement).

Verbal Behavior The main communication system in humans is verbal behavior. Compared with evocative communication, verbal communication is less prone to sympathic emotional responses in the recipient. The difference between the two modes can be appreciated by inspecting advertisement for cigarettes. The unobtrusively displayed warning label required by federal law represents transcribed verbal information, whereas the advertisement itself contains evocative information. In everyday situations, both verbal and evocative modes of communication are used. The content of the verbal and evocative modes can be inconsistent at times. For example, sarcasm involves positive verbal content but negative vocal content, whereas friendly banter may include verbal insults in the congenial context of the locker room. Verbal behaviors can be classified according to their function (reinforcing consequences). Among the verbal behavior categories proposed by B. F. Skinner, the best known are the “mand” (request) and the “tact” (description of an event). An important component of treating individuals with mental retardation is teaching them to mand, thus obviating the need for other attention-getting behaviors.

The relation between semantic and phonic components of a word is arbitrary, differing across languages, and under reinforcement control to conform to the norms of the verbal community. Ignoring developmental issues, verbal expressions are learned in three ways: by observing someone pointing to an object while naming it (ostension) by explicit definition (a synonymous phrase), and by implicit definition, which involves explaining a variety of sentences in which the word is used.

Behavioral Characteristics of Verbal Stimuli Research findings over the last two decades indicate that verbal stimuli have one unique property, the potential to form a bidirectional relation with the object denoted. As Steven C. Hayes and Linda J. Hayes said in *Rule Governed Behavior,* “The relation between a word and a referent is bidirectional. A word “stands for” the referent only if the referent ‘is called' the word.”

Bidirectionality thus means that a word can evoke the responses that the object would have elicited and the object prompts utterance of the word. The word and the object are in an *equivalence relation* with each other. For example, when people look at a lemon, they may think or say “lemon.” Conversely, when they hear the word “lemon,” they may experience sympathic salivation, especially if they elaborate by describing its color, its shape, the aroma of lemon zest, and the typical sensations that follow squirting lemon juice on the tongue. Similarly, patients with panic disorder are often reluctant to give detailed descriptions of their symptoms, lest such descriptions operating in the reverse direction evoke symptoms of panic. Bidirectionality at times causes confusion between “use” and “mention” of a sentence.

The behavioral concept of bidirectionality comes from the “matching to sample” experimental paradigm. In a paradigmatic experiment, a laboratory pigeon faces three different response keys aligned horizontally. The middle key displays a sample stimulus, either a square or a circle. The lateral keys are the matching keys, which present a color stimulus, either red or green. An experimental trial consists of the sample key lighting up, presenting the sample stimulus. If the sample stimulus is a square, the pigeon is reinforced for pecking the red matching key.

After many successful trials, the roles of the sample and matching stimuli are reversed; that is, when red is presented on the sample key, the correct response is choosing the matching key depicting a square. A bidirectional relation between square and red could be inferred, if the pigeon picked the square immediately, without reinforcement. Pigeons cannot do this, but verbally competent human subjects are. Perhaps the verbal behavior of naming enables bidirectional relations to emerge without explicit training.

Equivalence relations are not limited to pairs of stimuli. An equivalence class refers to several equivalent stimuli. The relations among the members of an equivalence include reflexivity ($A = A$), symmetry (if $A = B$, then $B = A$), and transitivity (if $A = B$ and $B = C$, then $A = C$). Furthermore, if the experimental design is modified to reinforce matching based on criteria other than identity, verbal stimuli become related on the basis of the logical relation reinforced (e.g., larger than, opposite).

Verbal Relations and Logic The recent behavioral insights into verbal behavior constitute a paradigm shift, whose end result is similar to and was anticipated by the linguistic turn of analytic philosophy more than a century ago (although the turn came from a different direction). For this reason, although the behavioral analysis of verbal behavior is not the same as logic, the study of philosophy has acquired renewed interest for the behavior therapist.

Quine's behaviorally inspired philosophical writings on verbal behavior are particularly relevant. Among categories of verbal sentences, an important one is the “observation sentence,” the description of a verifiable stimulus event of limited duration (e.g., the patient is screaming). The “standing sentence” reflects the truth of an event independent of subsequent stimulation (e.g., “the Sunday paper has arrived”). The “theoretical sentence” is a standing sentence referring to theoretical relations. The *observation categorical* is an observation sentence implied by another observation sentence (“if Bob is teased he gets angry”) or theoretical sentence (“whenever uranium's critical mass is exceeded, it will detonate”).

The total of all verbal relations can be conceptualized as a semantic network. At the periphery, the observation sentences constitute the interface with the world. Centrally in the network are the theoretical sentences that may concern abstract objects (e.g., molecules), dispositions (e.g., solubility), and logical truths (e.g., something cannot be both true and false). The relations between the nodes of the network are subject to change, either from direct experience or indirectly through changes in the relations of the central units. The individual's verbal community exerts its standardizing effect only at discrete points in the network, leaving much to be undefined. At the interface between the network and the world, observation sentences can be either affirmed or refuted. When an observation categorical is refuted, its premises must be reconsidered, and changes in these premises require reconsideration of the sentences that implied them. Sentences that are “logically true” (tautological) provide other anchors for the verbal community. For example, when a woman introduces a man as “my husband, the bachelor,” a listener suspects that the woman's use of language was idiosyncratic.

Verbally Defined Stimuli and Verbal Reasoning Because of their capacity to form bidirectional and other relations, verbal stimuli not only can become functionally equivalent to actual stimuli, but also can serve to create new entities (e.g., numbers, square strawberries, the king of U.S.A, metaphysics). For example, when novelists develop a fictitious character, they create a large number of new relations among the verbal stimuli. A well-developed, seasoned character may appear to us as real as a live person. Thus, the role of verbal behavior is not limited to the function of communication, it also serves to create new verbal realities that can be as stimulus-like and as unquestioned as physical objects. Such entities are called *verbally defined stimuli*. In the clinical arena, verbally defined stimuli may evoke maladaptive behaviors. For example, suicidal behavior is made possible by grasping the verbally defined constructs of “future” and “death.” Interestingly, animals do not commit suicide; the behavior of suicide requires a verbally competent agent.

A further example of verbally defined stimuli is the concept of the “self.” According to one behavioral analysis, the self is a verbally defined stimulus complex that develops as a result of concept formation from sentences that include first person pronoun, the word “I.” The section on stimulus presentation operations discussed how the concept of a “chair” developed as the intersection of repeated discriminations and generalizations from chairlike and nonchairlike stimuli. The concept of self might develop similarly, as the intersection of multiple I-statements. This process involves mastering the semantics of the first person pronoun. First person pronoun statements include “I am here,” “I am thinking about I statements,” “I am not over there,” “yesterday I had pain in my shoulder,” “If only I did not have to take that exam next month,” “I remember the bicycle I got as a child,” “When I am 65 I will retire at the beach. (The selected examples concern different time frames, which illustrates how, as the common denominator in these and other examples, the concept of “self” tends to acquire a sense of timelessness.)

Verbally defined stimuli also support the behavior of thinking or verbal reasoning. An important aspect of verbal reasoning is that it follows logical rules, in contrast with nonverbal reasoning, which (based on the principles of conditioning) is governed by perceptual similarity and learned discrimination. Nonverbal reasoning is also called “intuitive,” or “emotional,” reasoning. The two modes of reasoning sometimes lead to different conclusions; for example, patients with panic disorder report that “intellectually” they know that panic attacks are not dangerous, but “in their heart” they fear them anyway. Logical reasoning sometimes leads one astray because of the paradoxical contradictions inherent in self-referential logical situations. Here, it is advantageous to have the nonverbal mode available for “common sense” guidance.

Verbally Governed and Contingency-Shaped Behavior In nonverbal organisms, behavior change is mainly brought about by the shaping of operant responses. In humans, however, instructional control of behavior is the dominant mode of behavior change. Behavior controlled by verbal instructions is called *verbally governed behavior,* whereas behavior controlled by physical contingencies of reinforcement is called *contingency-shaped behavior*.

Verbally governed behavior is classified into several categories. In *tracking*, the instruction is followed because the behavior results in favorable adaptations to the environment, for example, putting a piece of prefabricated furniture together by following an instruction manual. A second category, *pliance* (derived from “compliance”) contains behavior affected by consequences for complying with the instruction itself, regardless of whether or not the instruction is useful. Some law abiding behaviors are of this type. A third category, *augmenting,* contain verbal establishing operations. For example, the statement “a drink sure would be nice” increases the reinforcing value of a drink.

Behaviors that look identical may be functionally different depending on the type of controlling stimuli. For example, putting on a sweater may be either verbally governed or contingency based. In the latter case, reinforcement is based on removal of the aversive effects of cold weather. If verbally governed, putting on the sweater may constitute a response to a mother's command, “put your sweater on” (i.e. a ply) or it could be a track (i.e., a response aimed at not getting cold in response to information that it is cold outside). Finally, it might be augmental, in response to hearing that a certain famous person wears this type of sweater. In many situations, it is preferable to arrange for tracks than for plies. This is especially true in conflictual relationships in which deliberate disobedience might be reinforced. This can result in self-destructive behaviors (e.g., a child failing a test to upset the parent.

Verbally governed behaviors are better than contingency-shaped behaviors when the individual is faced with a new situation. Verbally governed behavior also helps individuals pursue their long-term goals. Verbal stimuli can bring distant goals closer in time and thus enable the individual to choose a larger delayed reward over the immediate gratification from an impulsive act.

On the other hand, verbally governed behaviors can also be associated with problems. Verbally governed behavior is often described in terms of behavioral performance, whereas contingency-shaped behaviors are described in terms of their consequences. Instructions therefore tend to focus on the topography of a behavior rather than the result. The expression “going through the motions” illustrates the case in which a contingency-shaped behavior is being replaced by an instruction-following behavior that replicates its topography but not the accomplishments.

“Expertise” in an area implies exposure to the direct consequences of one's behavior. To generate such expertise without an initial tedious trial and error period, instructions are given to generate a behavioral performance that is good enough to be reinforced by natural contingencies. As instructions are later withdrawn, the behavior comes under the control of the natural contingencies and develops further because of shaping by these contingencies. The process of coaching in sports training represents a careful balance between instructions and behavioral practice in contact with natural contingencies. Excessive instructional control may reduce performance. (This is why compliments can be used to interrupt an opponent's winning streak.)

It may be particularly difficult to change a verbally governed behavior to a contingency-shaped one when the behavior is performed primarily to please the instructor (i.e., can be categorized as a ply). This may be one reason why some verbally governed behaviors are reported to be relatively insensitive to changes in the natural contingencies. For example, a therapist providing treatment according to a detailed manual is engaging in verbally governed behavior. In a research setting (and perhaps future managed care settings), the therapist may be subject to additional contingencies from treatment fidelity assessments. Such contingencies would discourage the therapist from engaging in contingency-detecting behaviors. Under such circumstances, therapists seem to run the risk of becoming immunized against the possibility of learning from direct experience. On the other hand, some interventions (e.g., exposure and contingency management) require strong compliance from the patient. In this case, it may be an advantage for the therapist to be bound by a treatment manual, since it leaves no room to negotiate with a patient who tries to change the treatment program.

Another problem arises when instructional control is attempted but is not possible. For example, spontaneous social behaviors are contingency shaped. In social skills training, patients learn many rules of interpersonal conduct. Such patients may display a robotlike quality in their learned social behaviors (e.g., too much eye contact rather than too little). Thus, a social behavior should be practicing under naturalistic conditions so that contingency control can be established.

Private Events Private events refer to stimuli or actions accessible only to the individual. Private events elicit ambivalent feelings among behaviorists. B.F. Skinner succinctly stated the problem as follows:

> In a sense, verbal behavior which cannot be observed by others is not properly part of our field. It is tempting to avoid the problems it raises by confining ourselves to observable events, letting anyone extend the analysis of his own covert behavior who wishes to do so. But there would then be certain embarrassing gaps in our account. In intraverbal chaining, for example, necessary links are sometimes missing from the observable data. When someone solves a problem in “mental arithmetic,” the initial statement of the problem and the final overt answer can often be related only by inferring covert events.

Private events can be divided into those that seem similar to stimuli and those that are similar to responses. Private stimuli are interoceptive sensations, such as proprioceptive or visceral stimuli, which are important in our understanding of panic disorder. In panic disorder, fear behavior can become conditioned to interoceptive stimuli that regularly precede a panic attack.

During childhood, the individual is taught to name private stimuli. Because the private events are, strictly speaking, not accessible to the teacher, names for these stimuli are learned by indirect means. For example, a report of “nausea” may be anchored by visible physiological responses. In other cases, extending observation time tends to clarify the nature of a private event to a bystander. For example, a toothache may be validated by a later finding of tooth decay; the feeling of hunger can be inferred from the individual's consumption of a large meal. Metaphorical extension of public stimuli to private events gives a listener a tangible impression of the quality of the sensation. For example, pain can be sharp, dull, burning, etc.

Private behaviors can be categorized along a dimension of observing versus acting. Mental behaviors similar to observing include attending, imagining, or recalling. Mental behaviors similar to acts include thinking, reasoning, planning and goal setting. Still others have elements of both, such as categorizing, reformulating, or calculating.

One distinction of clinical importance is the difference between imagining and thinking. Psychophysiological studies indicate that imagining is associated with autonomic responses appropriate for the image being imagined. For example, imagined fear-inducing consequences are associated with increases in heart rate or changes in skin conductance. Worries, on the other hand, are mostly verbal. Thinking worrisome thoughts, unlike imagining, is not associated with pronounced physiological reactions. However, if one specific worry is pursued beyond its normal duration (i.e., if the patient is prevented from changing the subject of worry), a fearful image may appear concomitantly with escalating anxiety levels. Patients with generalized anxiety disorder normally switch to a different worry before this happens. Thus, although experienced as anxiety, worrying may have the function of avoiding even greater fears.

A conceptually tricky issue is whether mental behaviors can be verbally governed (i.e., deliberately changed). Successfully instructing someone to attend to a particular stimulus seems possible, at least temporarily. Similarly, subjects can be asked to set certain immediate goals (i.e., intending, the mental counterpart of promising) or to apply certain strategic modes of thinking (problem solve). However, this control is only partial. For example, it is easier to maintain attention if the stimulus is changing, than if it is constant.

Verbal instructions to change a verbally defined mental behavior risk paradoxical failure. For example, when the change attempt involves not attending to a stimulus or not thinking about something, the verbal instruction provides the very stimulus that is to be avoided. It is impossible to follow the instruction “don't think about pink elephants” and impossible to report seriously that the change attempt was successful. At other times, the mental experience targeted for change becomes inextricably confounded with the instruction of changing the experience. This occurs when a person attempts to change a behavior defined as spontaneous. For example,

deliberately trying to increase one's love for a person is a case of “wanting to love,” which is the anathema of romantic love. (That is, being told “I want to love you” is not as pleasing as being told “I love you.”) Similar difficulties are inherent in wanting to believe, wanting to want, or wanting to desire. For example, patients with hypoactive sexual desire disorder may inadvertently aggravate the problem by recognizing it as a problem. A therapeutic approach to all these paradoxical situations is finding a way to extricate the target behaviors from verbal control. Acceptance and commitment therapy was developed for this very purpose.

Emotions as Descriptors of Contingencies of Reinforcement Recent social psychological research findings suggest that verbal reports of specific emotions suggest specific antecedent conditions, action tendencies and reinforcing consequences. Emotions can be differentiated by whether the eliciting event represents something desirable or undesirable and is located in the past or in the future (e.g., fear). Furthermore, if the event was in the past, a further consideration is whether it was caused by a deliberate act by another person, was accidental, or was a self action (e.g., anger vs. frustration vs. guilt, shame, or embarrassment). Furthermore, emotional behaviors are followed by natural consequences similar to the endpoints of a behavior systems mode. For example, the action tendencies of anger converge on a goal of retribution. The simplified account below should help the therapist to understand how a patient's verbal report of emotional responses reflects that patients life situations.

Fear and Relief Verbal descriptions of fear-inducing situations indicate that the unconditional stimulus for fear is an undesirable event in the future signaled by a conditional warning stimulus in the present. Characterized by the action tendencies of avoiding, freezing, or running away, fear is accompanied by autonomic symptoms such as heart palpitations. The emotional expression of fear (e.g., fear vocalizations and facial expressions) tends to elicit vicarious fear reactions in others—including the therapists when accompanying patients are undergoing exposure treatments. Successful escape behaviors tend to be accompanied by a feeling of relief. From an operant perspective, the feeling of relief is a marker for negative reinforcement. Therefore, for patients with agoraphobia, a history of many episodes of relief suggests a history of frequent negatively reinforced escape behaviors. The experience of relief may lead to sensitization rather than desensitization.

Sadness and Joy Joy marks the attainment of an endpoint of an appetitive behavioral mode. Depending on its intensity, joy tends to be signaled by the behavior of smiling, laughing, jumping up and down, or dancing. (Such behaviors are commonly displayed or enacted by the winners in game shows on television.) Conversely, sadness implies a situation in which the positive reinforcer is no longer available. Behavioral models of depression postulate that depressed individuals do not receive enough positive reinforcement. The action tendency for sadness is giving up. Display of sadness tends to elicit helping behaviors that may restore the lost reinforcer.

Social psychological research indicates that a person's goals are hierarchically organized, with intentions at the lowest levels, and life goals, or ideal selves, at the highest. (This goal system is yet another way of describing of hierarchically organized sequences of behavior.) Two types of life goals have been described, positive (ideal self) and negative (feared self, e.g., teenagers goal of differing from their parents). Depending on the distance from the goal, the affect associated with the pursuit of the ideal self tends to vary along the positive affect dimensions, (i.e., progress from sadness to elation). Affects associated with the avoidance of the feared self, on the other hand, tend to vary along the negative affect dimension (i.e., progress from anxiety to relief).

The difference between the positive affect and negative affect dimensions is reflected in our language. The positive affect dimension reflects things we want to do, whereas the negative affect dimension reflects things we ought to (or should) do. (Note that the issue of free will mostly tends to be raised with respect to the latter.) The distinction between wants and shoulds has been used therapeutically to increase the rate of positive reinforcement in a patient's life, thus creating an environment richer in positive reinforcement. As discussed in the consequent operations section, this enrichment reduces the effect of the reinforcer maintaining dysfunctional behaviors, such as alcohol or drug intake or bulimic behaviors.

Anger Past undesirable events deliberately caused by another person, such as receipt of punishment, is the primary stimulus for anger. The difference between anger and frustration is that anger is evoked by a deliberate act by another person, whereas with frustration, the event was an accident. Patients with chronic medical disorders should have ample reasons to report frustration. In addition, they may be angry with their physicians for not helping them enough or at the opposing party in a workman's compensation scenario. The action tendency for anger is hitting. The goal associated with anger is punishment of the offender. (However, this punishment is likely to be reciprocated in a coercive escalation.)

Pigeons exposed to schedules of reinforcement involving frustrative nonreward or aversive stimulation will work for a reinforcer that is an opportunity to attack an innocent pigeon in an adjoining cage. That is, for both pigeons and humans, innocent bystanders can become targets for retaliatory behaviors. For example, when in the survival mode, patients with combat-related posttraumatic stress disorder often act abusively toward their spouses. Attacking (or expressing anger) is equivalent to an appetitive or approach behavior. This means that the individual experiences the opportunity to punish the perpetrator as desirable. The appetitive nature of anger was confirmed in a recent brain imaging study in which anger induction resulted in changes typical of appetitive behaviors rather than defensive behaviors.

Some behavior therapists recommend the use of unreinforced exposure to treat anger. As we will see later, the purpose of exposure is to extinguish conditional responses to signals for undesirable future events. However, in anger the unconditional stimulus has already occurred; therefore, there are no antecedent conditional stimuli to extinguish. Exposure to past anger-inducing events (e.g., circumstances of a patient's hostile divorce) may even result in increased brooding over this event. Such brooding increases, rather than decreases, anger. Furthermore, behavioral treatment of posttraumatic stress disorder in rape victims was less effective for patients who reported much anger. However, exposure to anger-inducing stimuli may be helpful if the problem is one of fear of anger.

Modification of excessively angry behaviors is difficult. Specific techniques might include validation, education, reevaluation of the offense, avoiding escalation, avoiding alcohol or recreational drugs, derailment by the priming of other behavioral systems, and finding alternative means of resolution. When angry individuals come for psychiatric or behavioral help, the behavioral system view predicts that they initially try to enlist the therapist as an “ally” to punish the offender. This type of alliance, however, is not what is normally considered a therapeutic alliance. Nevertheless, the appropriate therapeutic response in this situation is validation, which means that the therapist agrees with the patient's point of view. The next step might be clarifying the nature of the initial offense. Perhaps the offense was less deliberate than the patient reports. If the source of the anger involves a parent, for example, one strategy might be to perform a behavioral analysis of the parent. This might lead to an understanding of the contingencies controlling the parent's behavior and make it seem less deliberate.

Another approach is to educate the patient about the general nature of anger. Often it is helpful to point out that a particular situation (e.g., chronic innitus) is destined to make the patient frustrated or angry. This intervention tends to be more effective if done even before the patient mentions any such feelings. Explain to the patient that individuals in an angry mode selectively look for things to be angry about and respond more quickly and decisively (but also more impulsively) to small frustrative events. They are irritable and lash out at innocent bystanders. All these effects will be amplified if the patient is intoxicated. A final target for anger treatment is to modify the goal of the anger action tendency. The goal of punishment is replaced by one of forgiveness or one of negotiated resolution. The latter seems to be a favored method among elementary school teachers, who encourage the use of words rather than fists. Assertiveness training is often used for patients who lack an adequate repertoire of negotiating behaviors.

When the patient is in a state of acute anger (in survival mode), a partner or bystander must not escalate with counteraggression. It may be best for the partner to leave the situation altogether until things cool down. Patients can be taught self-control measures that involve the priming method already discussed. For example, watching a comedy show (entertainment), going fishing (general search behavior), or seeking out pleasant odors (e.g., baking a cake) may derail the anger mode.

Disgust Disgust is a specific emotional response to decayed animal matter or body waste. It seems to be most closely associated with the olfactory sensory modality. Disgust thus may serve a principal function of avoiding ingestion of harmful substances, and it supports hygienic behaviors of various kinds. A variant of disgust is also involved in a sexual rejection system in which an individual does not want to be touched by a pursuer. Disgusting stimuli follow a special logic called the law of sympathetic magic. For example, the disgust-eliciting function from a stimulus can be transferred to other stimuli that were in physical contact with it or are similar to it in other ways. This logic is also observed in many patients with obsessive-compulsive disorder who avoid touching door knobs touched by others and engage in excessive cleaning behaviors. Blood and injury phobias and phobias of animals of the creeping or crawling type constitute disgust responses rather than fear responses. The distinction has practical importance in that the eliciting stimuli are unconditional, rather than conditional, stimuli.

Shame, Embarrassment, and Guilt Shame, embarrassment, and guilt are three self-conscious emotions for which the individual's own behavior is the antecedent condition. They can be conceptualized as modes in the dominance/submission system or a larger affiliative behavior system. These emotions reflect various ways of regaining acceptance by the group after an infraction. In guilt, the stimulus is a specific act that caused harm to another person. The action tendency for guilt is undoing and perhaps self-punishment. The undoing aspect can be constructive and lead to conflict resolution. Show of guilt might have a function of preventing punishment by others. For example, criminals may get lighter sentences if they show remorse.

Shame occurs in response to humiliation, or publicly revealed substandard performance. The expression of shame, such as hiding one's face, reflects an action tendency of submitting or hiding. Shame is also related to anger; that is, having one's shameful acts exposed may belong to the class of anger-inducing events.

Individuals who report shameful experiences (i.e., shame-prone individuals) tend to be more aggressive in anger-inducing situations than individuals who are guilt prone; the latter tend to use conflict-resolution strategies instead.

Embarrassment occurs when an individual violates a social convention in front of an audience. Stimuli that reliably elicit embarrassment include seeing oneself singing on a video tape, seeing someone else get embarrassed, being the target for teasing, and being praised too effusively. The expression of embarrassment includes a controlled smile, averting one's gaze, and blushing and is thought to have an affiliative function. For example, the blushing response, being only skin deep, is physiologically similar to the blush that occurs in sexual presentations. The natural consequences for embarrassment include sympathy, humor, and forgiveness.

Surprise Surprise occurs when something that was expected did not occur or when something that was not expected did occur. As discussed above in the section on Pavlovian conditioning, such surprisingness is necessary for conditioning or learning to occur. For example, if an agoraphobic entering a feared situation is unpleasantly surprised by the high intensity of the anxiety response, sensitization might occur. Pleasant surprises, on the other hand, are conducive to desensitization.

Feelings About Feelings Emotional responses can serve as stimuli for secondary emotional responses. For example, an individual who has a childhood history of losing fights may be fearful of getting angry. Fear of fear tends to develop when the fear responses interfere with task performance. For example, musicians with stage fright fear the effects that being fearful will have on their performance. Similarly, fear of blushing may contribute to the development of social avoidance behaviors. Individuals often attempt to cope with fears of emotions by suppressing the emotional complex. However, this only leads to frustration that further compounds the problem. The reader will recognize this scenario as one leading to emotional avoidance, a target behavior suggesting a treatment-approach based on acceptance.

FUNCTIONAL BEHAVIORAL ANALYSIS EXAMPLES

The following sections contain functional analyses of three prototypical clinical problems. The first involves self-injurious behaviors in patients with mental retardation and focuses mainly on the consequences of the behavior. The second problem is a geriatric inpatient with dementia who is “resistant to care,” in which analysis emphasizes establishing conditions and contextual control of behavior. The third problem is parasuicidal behaviors in patients with borderline personality disorder. The behavior analysis of this problem led to the design of dialectic behavior therapy and considers early learning history, current behavioral deficits and contingencies, and principles of choosing among multiple target behaviors.

Self-Injurious Behaviors in Patients With Mental Retardation Self-injurious behaviors are common among individuals with mental retardation. Single-subject designs can be used to identify the relevant controlling variables for the individual patient. In a study summarizing 152 functional analyses, the most common behaviors were hitting head (58 percent), head banging (43 percent), biting (37 percent), hitting body (16 percent), and scratching (16 percent).

The functional analyses were conducted as single subject designs, which varied the consequences of a self-injurious behavior. One consequence, negative attention, involved the experimenter reprimanding the patient and physically interrupting the self-injurious behavior. Another experimental condition, escape from demand, first engaged the patient in a task (as an establishing operation). Contingent on a self-injurious behavior, task material was removed, thus permitting the patient to escape from the task. Other conditions, called “Play” and “Alone” were included in which self-injurious behavior led to no specific externally administered consequence. In the play condition, the experimenter would interact with the patient in a friendly manner on a predetermined fixed time schedule (such as every 10 or every 30 seconds). Some play conditions were modified to follow a differential-reinforcement-of-other-behavior schedule, in which no self-injurious behavior could have occurred for 5 seconds prior to initiation of a play interaction. In the alone condition, the patient was in the therapy room without any play material, this condition represented a test for automatic reinforcement (sensory stimulation).

Contingent escape from task demands accounted for the occurrence of self-injurious behavior in most of the patients. Negative attention (contingent reprimands) came second. The most common cause for self-injurious behavior was contingent escape from task demands. The second most common cause was negative attention (contingent reprimands). In 20 percent of the patients, no environmental cause was identified. In these patients, the self-injurious behavior was hypothesized to be maintained by self-stimulation. Treatments that targeted the reinforcers identified by the functional analysis were more effective than treatments that targeted contingencies irrelevant for the patient's self-injurious behavior. Treatments involving timeout and extinction worked for patients whose self-injurious behavior was maintained by negative attention but not for patients whose self-injurious behavior was maintained by escape from either task demands or automatic reinforcement. Treatments that worked relatively well for all three subtypes of self-injurious behavior included noncontingent play or play interaction during the differential-reinforcement-of-other-behavior schedule. That noncontingent play worked may seem surprising, but it provides an example of how a behavior can be decreased by introducing a rich reinforcing environment.

Aggressive Behaviors in Inpatients With Dementia

> The patient was a 72-year-old female with dementia, who also suffered from urinary incontinence and an arthritic condition in her shoulder. She was able to walk only with support. She had been admitted to a psychogeriatric inpatient unit for being “resistive to care.” On the unit, the patient communicated mostly by nodding or shaking her head or with occasional one-word answers, at times smiling and at other times looking fearful, visually scanning the room. Mental status examination indicated that she knew her own name and that of her daughter, but she was never able to learn the names of her caregivers. She was disoriented to time and place. Specifically, she did not identify her current environment as a hospital. The patient mostly sat quietly in a chair in the dayroom. The target behavior, “resistance to care” occurred in the morning when getting out of bed, in the evening when getting back to bed, or during bathroom visits. These activities required two or three staff members. To reduce the incidence of incontinence, her daily routine included regular visits to the bathroom. The following event constitutes an example of “resistance to care.” Two male staff members approached her for her scheduled bathroom visit. She motioned them away. Ignoring this, the staff members began to lift the patient by her upper arms. The patient responded by scratching one of the staff members. The staff members terminated the interaction. A “level II” was called, and three safety officers in identical blue jackets appeared.

Establishing Conditions The verbal deficits in patients with dementia deprive them from a primary mode of communication. This makes it difficult for the patient to make her needs known. She may have been unable to communicate, for example, that certain ways of lifting resulted in pain in her shoulder. The lack of verbal repertoire also results in deficits in verbally governed behavior, which includes sequences of behavior directed toward a goal. The dementia also eradicates the patient's learning history. This may contribute to the emergence of previously punished behaviors (disinhibition), including aggressive or intrusive behaviors. Impaired hearing or vision reduce access to positive reinforcers, setting the stage for depression. Sensory deprivation also may establish self-stimulatory behaviors such as repetitive screaming. Fatigue may establish requests for various activities as aversive. Medical conditions associated with pain, if accentuated by staff interventions, may lead to reinforcement of escape behaviors.

Interventions aimed at modifying establishing conditions include a careful medical evaluation, which may identify sources of distress that can be relieved by situational adjustments or specific treatment. Hearing aids or intervention to improve vision (e.g., cataract surgery) may restore access to lost reinforcers. Treatment with antidepressants may increase reinforcer efficacy. Fatigue can be addressed by scheduling the activities around the patient's natural diurnal cycle. For example, aggressive behaviors occurring primarily in the morning after the patient has been awakened can be ameliorated simply by letting the patient remain asleep until she wakes up naturally.

Contextual Control The fragmentation of the patient's learning history also affects antecedent control of behavior, particularly by the complex stimulus constellations involved in contextual control. Being disoriented to place, the patient may not respond appropriately to the context of hospital, where behaviors appropriate for “being a patient” include tolerating having one's privacy invaded. The prevailing custom for nursing staff in psychiatric hospitals to wear street clothes rather than nursing uniforms further masks the hospital context. In fact, the only uniforms the patients see are worn by safety personnel and may be reminiscent of police uniforms. What remains is a public place populated by strangers, a stimulus configuration consistent with the context of perhaps a hotel lobby or a railroad station. In these contexts, when a stranger violates an individual's personal space or disrespects his or her physical integrity, aggressive responses can be seen as adaptive.

Treatment interventions might be aimed at restoring the salience of the “hospital” context. For patients with reasonably intact verbal repertoires, frequent reminders that “you are in a hospital” may work. It would be interesting to examine whether the incidence of aggressive behavior would be reduced in an environment where the staff routinely wears white nursing uniforms.

Stimulus Control The most common antecedent of aggressive behavior involves requests from a staff member for patients to cooperate in the execution of activities of daily living. Treatment includes appropriate use of touch and slowing the pace of the staff member's verbal behavior. The principle of *behavioral momentum* can be used to prime a set of complying with requests. This principle states that a behavior is likely to continue once it has started (gained momentum). Thus, compliance

with a request should be more likely if patients have just complied with another, perhaps smaller, request. Specifically, patients might be asked for “a small favor,” one that they are unlikely to refuse (such as helping the nurse with something). After one or more such granted requests, patients should be more likely to comply with the targeted request (e.g., taking medication).

Consequent Control The behavior of patients with dementia is highly affected by reinforcer delay. Thus, it is unlikely that aggressive behavior is reinforced by increased attention from staff in most patients. Screaming behaviors often appear to be self-stimulatory and are difficult to modify with behavioral techniques. The behavior of grabbing at a staff member's body parts may be reinforced by the sensory consequences of grabbing. Such behavior can be modified by promoting other behaviors that accomplish similar sensory consequences. Patients could be given an equivalent object to hold, such as a softball, pillow, or stuffed animal.

Suicidal Behaviors in Patients With Borderline Personality Disorder The functional analysis of borderline behaviors illustrates the behavioral analysis of complex patients. In such cases, behavior therapists attempt to identify functional classes of target behaviors and their dynamic interrelationships. Certain problems may be caused by the consequences of other more primary problems. For example, in many patients, symptoms of generalized anxiety disorder are driven by a perfectionistic way of responding to contingencies. A decision is also made as to what target behaviors need to be addressed first. The initial treatment efforts may need to be focused on target behaviors that would interfere with future treatment of the primary targets.

A systematic behavioral analysis of borderline personality disorder began as an analysis of suicidal behaviors and evolved into a treatment called dialectic behavior therapy. The term *dialectic* was chosen to reflect the approach of finding a third way (synthesis) to resolve opposing contingencies affecting these patients.

Establishing Conditions in Early Childhood: An Invalidating Environment The section on verbal behavior above, reviewed how a sense of self can be seen as a result of a contextual definition of the word “I.” Thus, a concept of “I-ness” is formed by abstracting over many sentences that contain the word “I.” Problems in the development of a sense of self can arise when other persons assume control over children's labeling of their private events. This reasoning has led to the inclusion of the important therapeutic principle of validation into dialectic behavior therapy and other recently developed treatments.

Besides the basic problem with invalidation, patients may have been raised in either chaotic, abusive families, or in families that had low tolerance for emotional, expressive behaviors. Expecting more or different behaviors than a child can emit, the parent resorts to coercive control. This leads to problems inherent in such control, including diminished exploratory behaviors, restricted problem-solving abilities, and skills deficits. In addition, unavoidable punishment leads to learned helplessness, anxiety depression, and posttraumatic stress symptoms.

Current Contingencies

ESTABLISHING CONDITIONS Behavioral deficits in regulating emotional behaviors, tolerating distress, interacting with others, and problem solving reduce access to positive reinforcers. This contributes to a lean reinforcement environment that increases dependency on the few reinforcers available to the patient. Consequently, the patient may be unable to leave current chaotic or abusive relationships.

A history of physical abuse establishes rewarding consequences for avoidance and escape behaviors. Self-mutilatory behaviors may function as escape behaviors. For example, physical abuse administered on a noncontingent time schedule (e.g., whenever the parent is drunk) may establish “passage of time without punishment” as a conditional aversive establishing stimulus. In such a schedule, the lowest probability of punishment occurs immediately after an abusive episode. In adult life, the patient may functionally “reset” the clock with an episode of self-mutilatory behavior, thus shortening the agony of “waiting for the other shoe to drop.”

ANTECEDENT CONTROL As mentioned above, borderline patients are often subject to conflicting contingencies. For example, 2 patients who rely on instructions from others are often placed in conflict by inconsistencies in the standards advocated by different people. This results in a problematic degree of inconsistency in their behaviors. As treatment progresses, the therapist becomes an important source of stimulus control. For example, these patients frequently show increases in borderline behaviors when the therapist is unavailable (e.g., vacations).

CONSEQUENT CONTROL Parasuicidal behaviors, unsafe sexual behaviors, and aggressive behaviors are impulsive behaviors. Such behaviors are maintained by certain immediate reinforcers and are insufficiently affected by undesirable but delayed consequences. In addition, the patient may respond to idiosyncratic, verbally constructed, future realities that serve as reinforcers for avoidance or escape behaviors.

Consequences implemented by the therapist become an important countermeasure. Withdrawal of attention can serve as a timeout from positive reinforcement. This can be applied contingent on a parasuicidal behavior; such withdrawal also is a useful strategy to counteract the therapist's natural tendency to increase attention for (and thus reinforcing) such behaviors. On the other hand, positive reinforcement is needed to retain the patient in treatment. Validation by the therapist is an important positive reinforcer. Therapists can also serve as “mentors,” functionally providing bridging stimuli to maintain behaviors toward delayed larger rewards. Verbally constructed realities can be addressed by cognitive techniques or by recontertualization (see section on acceptance and commitment therapy).

SELECTION OF TARGET BEHAVIORS Dialectic behavior therapy begins with efforts to address current problems; and among these, the first efforts are focused on behaviors that directly interfere with further treatment. Such therapy-interfering behaviors include suicidal behaviors, substance abuse, high-risk sexual behaviors, criminal behaviors, dysfunctional relationship behaviors, vocational behaviors, and housing-related behaviors (e.g., living in shelters). The second stage of treatment addresses symptoms related to posttraumatic stress. The third stage seeks to help the patient move toward individual goals and may include clarification of values or other desired long-term life patterns. In dialectic behavior therapy, patients attend group sessions in which specific skills are taught, as well as individual sessions with a primary therapist. The individual therapy component was inspired by a related treatment program called “functional analytic therapy.”

FUNCTIONAL ANALYTIC THERAPY. Functional analytic therapy is a behavioral analogue to insight-oriented individual therapy. No empirical outcome data have been published for this treatment program, but its system of target behaviors is of interest. Functional analytic program therapists use the term “clinically relevant behavior” (CRB) for a target behavior that occurs within the session. CRBs are of three types (CRB1–3). CRB1s are within-session behaviors representative of the patient's presenting problem. For example, for a patient with “difficulty making friends,” CRB1s might include poor eye contact, poor listening skills, difficulties in progressing with treatment because of imperative crises of various kinds taking precedence. CRB2s are new within-session behaviors indicating improvement in CRB1s. Such behaviors are reinforced by the therapists. Because the patients are adults, exuberant statements that often work with children, such as “wonderful” or “that's terrific” are ineffective. Finding effective social reinforcers or bridging stimuli for adult clients showing signs of improvement is an art. Straightforward praise or overly enthusiastic demeanor often is counterproductive, perhaps because it evokes embarrassment (see section on emotions) or because the patient fears that improvement will lead to an undesired increase in responsibilities. The first author, who is not a functional analytic program therapist, highlights the positive by using such nonverbal behaviors as leaning forward, note taking, and vocalizations. When patients take a particularly significant step, he might interview them about their strategy of overcoming the difficulties, using a model of the television sportscaster interviewing someone who just won a challenging race. Functional analytic program therapists use natural reinforcers that are not specified in advance. For example, the therapists might allow themselves to become more active in moving the session along. Because the CRB2s are affected by similar eliciting and consequent stimuli both within and outside a therapy session, changes in CRB2s are assumed to generalize to the patient's daily life.

CRB3s are the patients' description of contingencies affecting their behavior. CRB3s represent the behavioral analogue to insight. The therapist initially models CRB3s by describing or interpreting relevant behavioral processes as they unfold in the session (e.g., “I've noticed that each time you started talking about your spiritual beliefs, I've changed the topic, and you no longer bring it up.”). Behaviors similar to CRB3s are also a mainstay dialectic behavior therapy, in which patients perform detailed behavioral analyses of crisis events as chains of behavior.

In functional analytic therapy patients with a poor sense of self because of invalidating childhood environments may be asked to engage in free association exercises, in which they are asked to verbalize anything that comes to mind, with minimal cues from the therapists. Such exercises set the stage for I statements, which (as mentioned) form the basis for development of a sense of self. Patients are instructed to tell the therapist anything that enters their mind, without censoring. Basically the patients are encouraged to observe their private experiences and describe them as “I” statements. Patients with a poor sense of self become highly anxious during such exercises.

BEHAVIORAL DEFICITS AND SOCIAL SKILLS TRAINING

The treatment of behavioral deficits does not differ in principle from educational activities that take place in any school setting. However, behavioral therapists typically focus on deficits in social behavior rather than academic skills, hence the term *social skills training.* Variants of social skills training include assertiveness

training, generalized social skills training, communication skills training, and general strategies for problem solving.

Assertiveness Training Assertiveness training was first used in behavior therapy to countercondition social anxiety. With the publication of *Your Perfect Right* by R. E. Alberti and M. L. Simmons in 1970, learning assertive behavior became a goal in its own right. Since then, assertiveness training has been extensively covered in the popular media and has become part of our culture. Assertiveness is defined as follows:

> Assertive behavior enables a person to act in his or her own best interest, to stand up for herself or himself without undue anxiety, to express honest feelings comfortably and to exercise personal rights without denying the rights of others.

The goal of standing up for oneself in the definition above overlaps with that for aggressive behaviors, but aggressive behaviors (e.g., bullying or intimidating) deprive others of their rights, whereas assertive behaviors reflect a spirit of cooperation in which the rights of both parties are respected.

Two types of situations frequently call for assertive behaviors, setting limits on pushy friends or relatives and such commercial situations as countering a sales pitch or being persistent when returning defective merchandise. Early assertiveness training programs tended to define specific behaviors as assertive or nonassertive. For example, individuals were encouraged to assert themselves if somebody got in front of them in a supermarket checkout line. Increasing attention is now given to context, that is, what would be assertive behavior in this situation depends on circumstances.

Social Skills Training Children learn a number of social skills. Deficits in social skills make children vulnerable to adverse overtures from peers. Table 30.2-5 shows examples of social skills deficits often found in children with impulsive disorders or learning disabilities. Social skills can be enhanced by the parent reviewing (perhaps at night after the child has gone to bed) problematic situations observed during the day. In a coaching mode, the parent can later discuss or suggest possible alternative behaviors. Educational social skills training groups or interventions are available in some settings.

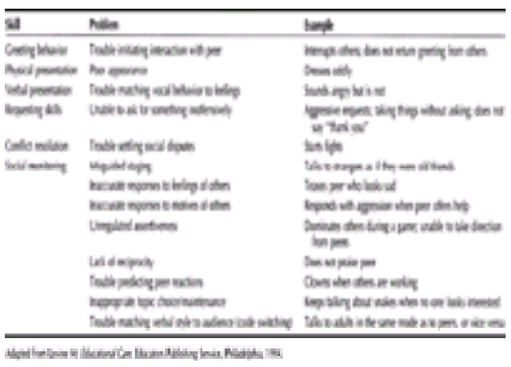

Skill	Problem	Example
Greeting behavior	Trouble initiating interaction with peer	Interrupts others; does not return greeting from others
Physical presentation	Poor appearance	Dresses oddly
Verbal presentation	Trouble matching vocal behavior to feelings	Sounds angry but is not
Requesting skills	Unable to ask for something inoffensively	Aggressive requests; taking things without asking; does not say "thank you"
Conflict resolution	Trouble settling social disputes	Starts fights
Social monitoring	Misguided staging	Talks to strangers as if they were old friends
	Inaccurate responses to feelings of others	Teases peer who looks sad
	Inaccurate responses to motives of others	Responds with aggression when peer offers help
	Unregulated assertiveness	Dominates others during a game; unable to take direction from peers
	Lack of reciprocity	Does not praise peer
	Trouble predicting peer reactions	Clowns when others are working
	Inappropriate topic choice/maintenance	Keeps talking about snakes when no one looks interested
	Trouble matching verbal style to audience (code switching)	Talks to adults in the same mode as to peers, or vice versa

Adapted from Gresham M (Educational Care. Education Publishing Service, Philadelphia, 1994).

Table 30.2-5 Examples of Social Skills Deficits in Childhood

After social skills are learned, they can also be lost. This is particularly true for patients with schizophrenia. The negative symptoms of schizophrenia essentially involve social skills deficits.

The negative symptoms in schizophrenic patients constitute behavioral deficits that go beyond difficulties with assertiveness. These patients have both inadequate expressive behaviors and inappropriate stimulus control of their social behaviors (i.e., they do not pick up social cues). Similarly, patients with depression often experience a lack of social reinforcement because of a lack of social skills, and social skills training is efficacious in the treatment of depression. Patients with social phobia similarly often have not acquired adolescents' social skills. In fact, their social defensive behaviors (e.g., avoiding eye contact, making brief statements, and minimizing self-disclosure) increase the probability of the rejection that they fear.

Social skills training programs for schizophrenics cover conversation skills, conflict management skills, assertiveness skills, community living skills, friendship and dating skills, work and vocational skills, and medication management skills. Each of these skills has several components. For example, assertiveness skills include making requests, refusing requests, making complaints, responding to complaints, expressing unpleasant feelings, asking for information, making apologies, letting someone know you are afraid, and refusing alcohol and street drugs. Each component involves specific steps. For example, conflict management includes skills in negotiating, compromising, tactful disagreeing, responding to untrue accusations, and leaving overly stressful situations. A situation in which conflict management skills might be used is when the patient and a friend decide to go to a movie and their choice of movie differs.

Negotiating and compromising, for example, involves the following steps:

1. Explain your viewpoint briefly
2. Listen to the other person's viewpoint
3. Repeat the other person's viewpoint
4. Suggest a compromise

Communication Skills Training The courses in interviewing techniques that mental health professionals take in their training are examples of training in communication skills. The communication skills of summarizing, reflecting, validating, and using “I” statements and eye contact are also taught in the context of marital therapy. However, a recent prospective study found that such communication patterns had no predictive value with respect to either the future duration of a marriage or happiness in a marriage. Instead, predictors of successful marriages included the husband's willingness to give up power to his wife, and the wife's capacity to use humor when providing feedback to her husband.

Physician-Patient Communication Physicians' communication behaviors with their patients are not always characterized by high levels of social skills. Based on the first author's experiences in a Psychiatric Medicine clinic, this seems to be particularly true when a patient's medical condition is not improving and the patient persists in complaining. Poor communication and bad treatment contribute equally to a patient's decision to file a malpractice suit. Consequently, efforts are under way to improve physician communication patterns.

A recent study showed that residents in internal medicine could learn such skills in a structured program that involved a 1½-day workshop followed by six small group sessions and six one-to-one feedback sessions during the following 6 to 8 months. The residents were taught how to greet patients, acknowledge the patient's chief complaint, be responsive to the patient's emotional expressions, and clarify the purpose of the consultation. They were taught to specify the amount of time available in advance and signal a change of interview mode ahead of time, (e.g., when a formal review of systems is to begin). Introducing the mind-set of treating the patient as an equal partner in decision making, the residents were taught to provide detailed information of assessment results, preliminary diagnosis, etiology, and treatment options and prognosis, while at the same time remaining sensitive to patient's emotional responses.

Parent-Child Communication Certain ways that a parent phrases a request to a child (particularly one with problems of conduct or attention) are likely to result in noncompliance. Table 30.2-6 shows that these include multiple commands, vague commands, commands phrased as questions, repeated commands, and commands that require compliance over an extended time.

Table 30.2-6 Effective and Ineffective Parental Commands

Giving effective commands requires getting the child's attention (e.g., asking the child to look at you). The parent's tone of voice should be neutral but slightly louder than usual. A positive tone of voice is more effective than a negative one. Commands should be limited to a few manageable steps. Finally, the same command should not be given repeatedly (i.e., nagging); rather, the parent should define the time interval for compliance (e.g., by the count of five) and apply predetermined consequences for noncompliance.

Problem Solving Learning problem-solving strategies benefits most psychiatric patients, particularly patients with social phobia, generalized anxiety disorder, and depression. It is an integral part of dialectic behavior therapy and of behavioral marital therapies. Problems can be classified as changeable or unchangeable. Unchangeable problems include natural disasters and socioeconomic trends beyond the control of the individual (e.g., unemployment due to plant closing). Problem solving is a systematic way of verbally generating new behaviors that can then be applied to contact new contingencies. The problem-solving heuristics developed by d'Zurilla and coworkers are widely used and involve the following steps.

1. Problem Orientation. Patients are taught that good problem solvers respond to problems as normal events that constitute challenges rather than threats. They realize that problem solving may take time and involve effort and do not give up when no quick fix is available. A problem-solving mind-set is primed by first discussing problems in society and their possible solutions. Patients are given a homework assignment—keep a diary of problems as they occur so that one or two can be selected for detailed analysis.
2. Problem Definition and Formulation. The problem is described in behavioral detail, including what happened, where and when it happened, who was involved, what feelings were involved, what goal or current concern was involved, and how much distress was experienced. Problems are defined in terms of specific behaviors rather than undesirable personality traits. Obstacles for earlier attempts to solve the problem are reviewed.
3. Generation of Alternative Solutions. The patient brainstorms a number of different potential solutions. For example, patients might be asked to generate at least five potential solutions. The solutions proposed should not be variations on one theme or be limited to those that worked in the past. The deferment of judgment and variety principles guide this process. If the problem is not changeable, solutions are selected involving emotion-focused coping. These may include behavioral techniques to modify anxiety or distress, or interventions aimed at cultivating acceptance.
4. Deciding on a Solution. A determination is made whether enough information is available to decide on a specific solution. If there is, the list is screened clearly inappropriate. The patient is asked to anticipate the outcome for each remaining solution in terms of short-term possibilities and long-term consequences.
5. Solution Implementation and Verification. The specific actions required for the chosen approach are listed, along with obstacles or barriers to their implementation. List making and self-monitoring may facilitate adherence. Environmental changes to facilitate the solution may be considered, such as placing reminder notices at strategic places.

CONTINGENCY MANAGEMENT

Contingency management is the use of principles of operant conditioning to change the patient's behavior by modifying discriminant stimuli or consequences. Contingency management techniques are implemented by someone other than the patient. Because they require precise control over the patient's sources of reinforcement, operant procedures are particularly used in the treatment of children, institutionalized individuals, and drug addicts on replacement therapy.

Paradigmatic examples of operant behavior therapy below focus on the treatment of an attention-deficit hyperactivity disorder. The techniques are representative of those used for the other disruptive disorders of childhood such as conduct disorder and oppositional disorder. Although the underlying principles are simple, contingency management interventions not only require careful attention to detail, but also necessitate skills in interacting with teachers and other individuals in the patient's larger system. Hence contingency management procedures are discussed in more detail than other behavioral techniques.

Classroom Contingency Management in Specialized Settings In special classrooms, contingency management procedures include point, or token economy, reward systems, timeout, and response cost (losing tokens for undesirable behavior). The procedures are implemented by a consulting professional, an expert teacher, or a trained paraprofessional. Patients treated in this setting include children with severe problems (e.g., comorbid aggression). The question has been raised whether interventions based on positive reinforcement alone are sufficient or whether punishment procedures are needed such as reprimands, timeout, and response cost (loss of privileges). Research has shown that with children with conduct disorder, positive reinforcement alone is insufficient. The same conclusion applies to the treatment of attention-deficit/hyperactivity disorder. Positive consequences alone are neither sufficient nor necessary, but prudent negative consequences are essential. Nevertheless, most behavioral interventions programs for attention-deficit/hyperactivity disorder include both positive and negative consequences.

Home-Based Reinforcement Programs The behavioral intervention most commonly used in clinical settings is best described as outpatient-based, clinical behavior therapy. This approach involves training parents to implement behavior management programs with their children and concurrently training teachers to use classroom management strategies with the patient.

Parent Training Parents are taught learning principles and standard behavioral techniques during 8 to 16 educational group sessions in the skills training format described in the previous section. A typical list of all topics appears in Table 30.2-7. One or more of these sessions focus on how to give effective commands (i.e., principles discussed in the previous section).

1. Review of attention-deficit/hyperactivity disorder and its treatment
2. Overview of social learning and behavior management principles
3. Home/school daily report card
4. Giving effective commands and reprimands
5. Time out
6. Attending, rewarding, and ignoring skills
7. Establishing and enforcing rules—when-then contingencies
8. Home point systems—reward and response cost
9. Home reward and response cost
10. Planned activities and setting generalization outside of the home
11. Homework problems
12. Facilitating peer relationships
13. Level systems
14. Dealing with the school
15. Parental stress, anger, and mood management

Table 30.2-7 Topics in Parent Training Sessions

Working With the School System Federal law requires that children with attention-deficit/hyperactivity disorder must be accomodated in these classroom settings. The Individuals with Disabilities Education Act (IDEA) requires that these children be evaluated for special services and receive an Individualized Education Plan

(IEP), including special education placement if indicated. The Rehabilitation Act of 1973, section 504, requires that special accommodations be made in the classroom setting of a child with attention-deficit/hyperactivity disorder (generally labeled a 504 Plan) even if the child is not eligible for special education.

Given these laws, one might assume that schools would automatically institute appropriate interventions, but more often than not, this is not the case. Thus it is incumbent upon the therapist to ensure that the child receives an adequate school-based intervention. The specific procedures that implement the regulations vary across states and school districts, and clinicians therefore must become familiar with how these regulations are implemented locally. Parents must know their rights under these laws, so that they can remind school administrators to support appropriate interventions with their children.

Before contacting school officials, the therapist needs to be aware of any special considerations involving the target child, parents, school, or district. These include the parents' relations with the school, the child's status regarding special education placement, and district policies regarding attention-deficit/hyperactivity disorder, the IDEA, and section 504 of the Rehabilitation Act.

The initial contact between the therapist and school should always begin with a phone call informing the school principal that the child's parents have solicited the therapist's services. The therapist briefly describes the goals and nature of the classroom program that will be developed with the teacher, obtains the principal's permission and support, and establishes a mechanism for keeping the principal informed. The principal also needs to be aware that treatment will likely continue in subsequent school years, so that plans for long-term continuation need to be built into the program.

Working With the Teacher The classroom-based interventions for attention-deficit/hyperactivity disorder involve standard contingency management procedures. Regular or special education classroom teachers, with help from school support staff or outside consultants, can implement most of these programs. Teachers can be taught contingency management techniques using handbooks, texts, or training programs. Table 30.2-8 shows a sequence of topics typically included in such training. A significant component of the program is a home-based reinforcement system involving a daily report card. The report card provides feedback from the teacher to the parents on the child's school performance, and parents provide a reinforcing consequence at home. The report card is discussed in detail below.

1. Introduction to attention-deficit/hyperactivity disorder, rationale for and overview of treatment (assess teacher knowledge and use of behavioral procedures, and design content of subsequent sessions accordingly)
2. Home/school daily report card
3. Classroom rules and structure; instructional modification for an individual child
4. Ignoring mild, inappropriate behaviors that are not reinforced by peer attention and praising appropriate behaviors
5. Giving effective commands and reprimands
6. When-then contingencies
7. Classwide interventions
8. Group contingencies
9. Response cost/reward point or token system for the target child
10. Time out (classroom, office, systematic exclusion)
11. Discussion of special services or special class placement

Table 30.2-8 Topics in Behavioral Teacher Training

In mental health settings the therapist commonly has a single meeting or telephone contact with a teacher. However, research clearly shows this is inadequate for the program to be implemented and maintained successfully. Rather, consultations with the teacher might involve up to 10 contacts—weekly at first and tapered over time, depending on success. In the initial contact, the therapist's main task is obtaining the teacher's cooperation, keeping in mind that unless the teacher referred the child, the teacher has not asked for this assistance. Some teachers are willing and even eager to cooperate, but more commonly, teachers have misgivings. For example, they may be concerned about the amount of time required or express doubt about the validity of attention-deficit/hyperactivity disorder as a disorder and the efficacy of behavior modification or medication as treatments.

Table 30.2-9 lists important topics covered during meetings with the teacher. The therapist should develop the program in collaboration with the teacher rather than unilaterally imposing a particular approach; the teacher's involvement in treatment is essential for success. The teacher is being asked to make changes that will last for the entire school year; therefore, it is important to address concerns the teacher may have regarding time commitments. Teachers might be reminded that they are already spending a great deal of time managing the target child. Granted the behavioral program requires an initial investment of extra time, but if the treatment is successful, the teacher will have saved time in the end. The therapist also assesses the teacher's current methods of classroom management and knowledge of behavior management principles. Many teachers, especially the younger ones, have had formal training in a behavioral approach. If explanation is necessary, it is most efficiently offered in terms of functional analysis of behavior. Thus, target behaviors are defined and relevant antecedents, consequences, and setting events are identified.

Introduce information about attention deficit/hyperactivity disorder, including prognosis without treatment, common treatments
Ascertain the teacher's commitment and build rapport; if the meeting is live (vs. phone), observe child and teacher informally before or after meeting as part of assessment
Discuss the teacher's current teaching practices (e.g., classroom structure and classroom management procedures in use)
Ensure that teacher has appropriate list of class rules, as "follows rules" will almost always be a target behavior
Discuss problematic child behaviors with goal of identifying specific target behaviors for modification; form a preliminary list of target behaviors
Discuss basics of observing and recording with teacher and develop baseline tracking procedures for target symptoms

Table 30.2-9 Items Covered in Initial Meetings Between Therapist and Teacher

Subsequent teacher contacts involve discussions of ways to implement behavioral procedures, specifically with the patient, and the patient's response to treatment. The therapist and teacher work together to develop procedures to modify the antecedents, consequences, and setting events identified in the functional analysis. Procedures must be implemented consistently by both teachers and parents. The behavioral program is set up in the shaping paradigm, initially targeting small improvements and gradually going for more. Treatment responses are monitored continually, and interventions are modified when they fail to have sufficient impact or are no longer needed.

Daily Report Card Daily report cards are the most basic and necessary component of classroom interventions with attention-deficit/hyperactivity disorder children. The report card serves several purposes. First, it can be used to provide feedback to the child. Second, it informs parents of the child's behavior during the school day. Third, it establishes a channel for regular communication between parents and teachers. Finally, it provides a way for parents to reinforce the child daily for improvement in target behaviors. The steps involved in establishing a daily report card–based intervention are the following:

1. **SELECT THE GOALS TO BE ACHIEVED** Treatment goals are determined at a molar level based on examination of rating scales, testing, and teacher/parent interviews. Examples of goals are improvement of academic grades, classroom behavior, or relationships with peers or adults. The goals must be socially valid, that is, chosen so that reaching them is likely to result in reduced impairment and improved outcome for the child.
2. **FORMULATE TENTATIVE TARGET BEHAVIORS** Once selected, a goal is operationalized in terms of a set of specific target behaviors. Target behaviors are molecular representations of the molar goals previously chosen. They need to be defined in a manner that permits objective and reliable measurement (i.e., resulting in a quantitative score). The teacher, child, parents, and therapist must all be able to agree whether or not a target behavior has occurred.

To avoid loopholes, the language that describes treatment goals as sets of target behaviors may have a legalistic flavor. For example, the general goal of improved academic grades might be broken down into specific behaviors that include (1) having the materials and assignments necessary to do tasks, (2) completing assigned academic tasks, (3) being accurate on assigned tasks, and (4) completing and returning homework. Similarly, the general goal of improving relationships with teachers and peers can be broken down into such target behaviors as (1) does not interrupt other children during their work time, (2) does not tease other children, (3) plays without fighting at recess, (4) obeys the teacher when commands are given, and (5) refrains from talking back to adults.

3. **GATHERING BASELINE DATA FOR TENTATIVE TARGET BEHAVIORS** Baseline recording involves daily observing or monitoring of the target behaviors for a week or more until a stable pattern is obtained without any intervention. The data collected provide objective information about the validity of the goal selection. If valid, they are also used to define and titrate the criteria for reward used in the initial steps for change.

   Developing methods of quantifying behavioral observations is an area of methodological research in behavior therapy. For the applied setting, the principle the simpler the better holds. One avenue for quantifying behavior involves measures of achievement or performance aspects of a behavior (i.e., the visible traces of a behavior). For example, a student's performance can be measured as school grades, and a salesman's performance, as number of dollars earned. For the child with attention-deficit/hyperactivity disorder, an achievement measure of classroom seatwork might involve recording the percentage of assignments completed during each day and the accuracy of this work. In applied work involving the daily report card, the number of yes responses (see below) is a convenient achievement measure.

   Another approach records the performative (movement) aspects of the behavior. Many behaviors have individual instances (bouts) that can be identified and counted; these can be monitored as frequencies (i.e., how many times the behavior occurred during the school day). Counting individual bouts of behavior is impractical when the target behavior is frequent (e.g., child interrupts 100 times a day). In this case, recording could be done for the portion of the day that is especially problematic. Another (methodologically less respectable) solution is to monitor the behavior on an interval basis (e.g., proportion of 15-minute intervals in which the child interrupted at least once).

4. **DECIDING ON FINAL TARGET BEHAVIORS** Using the baseline data, the therapist and teacher decide which target behaviors to include on the daily report card. There should be no more than five to eight target behaviors—fewer for an elementary school child. The initial criteria defining success for each target behavior are specified.

   Figure 30.2-3 shows an example of a daily report card. The criteria for success (earning a “Yes”) in each target behavior determine the level of performance required for home reinforcement. Initially, these criteria are targeted to require the child to improve a certain amount over the baseline level. A good criterion is one that the child fulfills 70 percent to 90 percent of the time, or a 20 percent improvement over baseline. A success rate below 70 percent indicates that the criteria are too difficult.

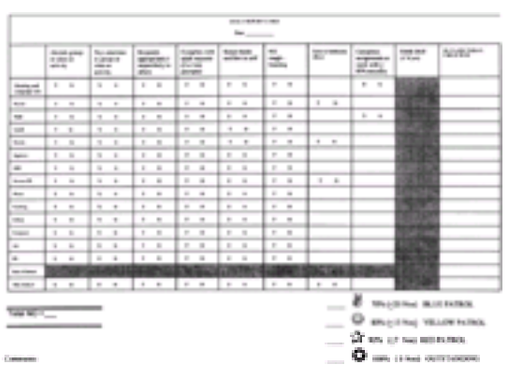

FIGURE 30.2-3 Daily report card for a child with attention-deficit/hyperactivity disorder. The daily report card provides a tabular framework reflecting days, times of day, subject areas (or if the child has different teachers, for each subject by teacher). For each target behavior, the child can earn a “yes” or “no” in each of the daily subdivisions. The patients using this card, a first and second grader, had problems in classroom behavior, relations with peers and the teacher, and academic performance. Each of these domains is reflected in the student's card.

5. **EXPLAINING THE PROGRAM TO CHILDREN** The daily report card is presented to children as a way of focusing during the day on things that are problems and learning how to overcome those problems. Children earn rewards at home for their behavior and performance at school. This is done to help them learn new ways of getting along with others and doing well in the classroom and on schoolwork. The daily report card is reviewed in detail with the children in age-appropriate language, including the target behaviors and their exact definitions.

   For example, for the target comply with adult requests (Fig. 30.2-3) the list included Do what grown-ups ask you to do. Do not say “why should I” or “no” and Stop when grown-ups tell you to stop. Children are also informed how the target behaviors will be monitored, exactly what criteria will mean a “Yes,” what the rewards are, and what their responsibilities are.

   Children are asked to recite back the information. They will be responsible for presenting the daily report card to each teacher for evaluation and a signature and for taking the card home. Once the program has begun parents or teachers or both review the daily report card targets with the children at the beginning of each day.

6. **ESTABLISHING THE HOME-BASED REINFORCEMENT SYSTEM** The parents' job is to reward the children for attaining the goals that have been established on the daily report card. Both daily and weekly rewards are established. Daily rewards provide immediate feedback to the children and increase their motivation to make the desired behavioral changes. Weekly rewards (e.g., for having three or four positive daily report cards during the week) further this goal, begin to develop the children's ability to delay gratification, and establish the efficacy of delayed reinforcement.

   Parents sometimes raise objections about giving their children rewards for work that they “should be doing anyway.” The therapist might counter that although the children should be doing their work, they are not (because of their disorder). This is why the intervention is necessary. In brainstorming for possible reinforcers, the principles outlined above for knowing what is reinforcing ahead of time are useful heuristics. Rewards are set up as a menu of choices. The possibility of choice is important because of the possibility of reinforcer habituation (this is particularly true for the daily reinforcers). Weekly rewards might include going to a favorite restaurant or renting a movie video. The rewards on the menu must only be available to the children when they earn them. At the same time, once the children have earned a reward, it must be given. Specifically, it cannot be withheld in response to misbehavior at home occurring after the children have earned their reward.

   The child in Figure 30.2-3 had a cumulative daily reward structure. For the lowest level of reinforcement (yellow patrol, 70 to 80 percent Ys), the reinforcement menu included the privilege of riding in a preferred seat in the car. (Because the child rode home with his parent, this reinforcer had a relatively short time delay). The next level (blue patrol, 89 to 90 percent Ys) qualified for a yellow patrol reinforcer plus the added privilege of watching TV. Red patrol (90 to 99 percent) was rewarded with either of these plus ice cream. Finally, the case of the child bringing home all Ys (100 percent) evoked a highly celebratory mood. In addition to the lower level rewards, the child became eligible for staying up 30 minutes past the regular bedtime. Even years later, this child would occasionally come home after a difficult day asking for a reward (“I need a reward for not hitting when my friend teased me”). An example of weekly rewards, earned if all five school days had been at least a yellow patrol, was a trip to the bookstore to buy a book.

   Some teachers worry about causing the child to be punished if they send home a negative report. Such worries can be allayed by the fact that the parents have specific instructions for this event. When the child brings home a daily report card meeting criteria for a certain reward level, parents praise the child for the behavior. On the other hand, when a child brings home a card on which criteria are not met, the parents respond in a neutral, matter-of-fact, nonpunitive manner. They explain that since the goals were not reached, there will be no reward. The parents also express confidence that the child will be able to do better

the next day.

Children with attention-deficit/hyperactivity disorder may at times forget to bring the daily report card home. If this is a major problem, parents can provide a reward for this behavior alone. Once the behavior for bringing the daily report card home has been established, if it does not come home for any reason, parents assume that the child received all Nos and do not reward the child.

7. **MONITORING AND MODIFYING THE DAILY REPORT CARD** The key to the success of the daily report card is that it must be monitored continually. Changes in target behaviors are tallied and criteria for reward are adjusted. If a criterion is consistently reached (e.g., at a 90 percent rate), the criterion is changed to require an increased level of performance. At the same time, criteria must be increased gradually enough to maintain the child's success a relatively high proportion of the time (shaping). Once the criterion for a target is consistently at an acceptable level (i.e., a level within the range shown by nonproblematic children in the classroom), the target behavior is dropped from the daily report card. The child is told that he or she is doing so well that the target isn't needed any longer. Similarly, new targets may be added if the child exhibits new behavioral problems.
8. **TROUBLE SHOOTING** After establishment and implementation of the daily report card, the therapist's further interactions with the teacher include anticipating and addressing problems. Trouble-shooting evaluations are indicated if, for example, the child's success rate, during a 10-day window, falls below an average of 70 percent in any one of the target behaviors. Table 30.2-10 lists a set of guidelines that can be used to trouble shoot the daily report card. The therapist updates the functional analysis, evaluating each treatment component with respect to fidelity (adherence to intervention plan) and efficacy. This might include direct observation in the classroom, which can be compared with the teacher's reports on the daily report card. The therapist discusses the results of the observations with the teacher and provides feedback. If necessary, additional teacher-administered interventions are added, such as contingent praise or timeout from positive reinforcement.

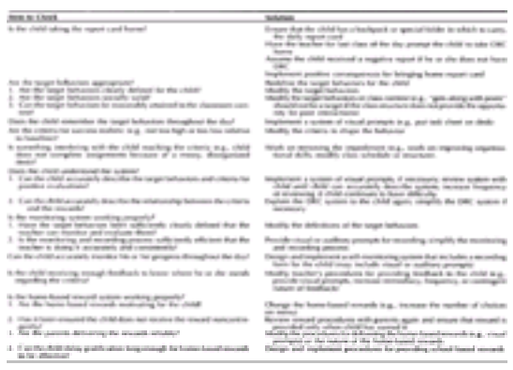

Table 30.2-10 Trouble-shooting a Daily Report Card

9. **TEACHING PARENTS HOW TO WORK WITH SCHOOLS** The need to taper therapist-teacher contact over time places great importance on teaching the parents how to work with the schools. This process begins with the therapist meeting with the parents and teacher simultaneously, thus modeling the behavior required. Over several months, the major responsibility for contact with teachers is shifted from the therapist to the parents. The therapist may provide outlines or scripts to the parents and correct feedback of specific interactions. In addition, if parents experience problems with cooperation in the school, the therapist provides guidance in trouble-shooting. The parents of the child whose daily report card is shown in Figure 30.2-3, were able to develop and monitor the daily report card for 2 years without contact with a therapist.
10. **TERMINATING THE DAILY REPORT CARD** The therapist's contact with the teacher and school should not be terminated prematurely. Toward the end of the school year, the therapist conducts programing for maintenance and relapse prevention, including procedures for bridging from one school year to the next, involving both the child's current and future teachers.

When all daily report card targets are consistently (e.g., during a 10 day window) at or above 90 percent criterion, and no other target behavior remains to be shaped, the daily report card can be phased out. Several procedures are possible. For example, the daily report card could be overall implemented only in the morning, gradually reducing the time covered, until the child is functioning without it. Alternatively, the daily report card could be eliminated on some days, and then used only on alternate days, every third day, and so on up to a week. With either method, card use is phased out as long as the 90 percent or better criterion continues to be reached for all targets over the 10-day window. The child in Figure 30.2-3 self-recorded his “Ys” for a period, with his reliability checked by the teacher.

Teachers and parents should maintain at least a weekly report of some type for the child with attention-deficit/hyperactivity disorder. For example, the daily report card may need to be used again after a hiatus, if gains are not maintained or new problems arise. Furthermore, school problems for these children often worsen in middle school and junior high school, and weekly feedback systems may be quite useful in those grades.

Intensive Treatment Programs Outpatient behavioral treatment, even at the intensity described above, may not suffice for many children with attention-deficit/hyperactivity disorder, particularly those with concomitant aggressive behaviors. For a number of years, the second author has conducted the Summer Treatment Program, an 8-week partial treatment for children with attention-deficit/hyperactivity disorder, located in a summer camp–like setting. This program has been named a Model Service Delivery Program in Child and Family Mental Health by a task force of the American Psychological Association.

The program runs for 9 hours a day, 5 days a week, for 8 weeks. Children are placed in age-defined groups of 12 children and 5 staff members, and group members stay together throughout the summer. Each group spends 2 hours daily in the classroom. The classes are either led by a teacher or involve individualized paper and pencil or computer-assisted instructions. The remainder of each day consists of recreationally based group activities, during which an integrated behavioral program is implemented. The backbone of the behavioral intervention in the Summer Treatment Program is a point system, backed up by both rewards and costs. The system is in effect in every setting throughout the day. Points are exchanged for privileges (e.g., weekly field trips), and social honors. Children also receive a daily report card on which their progress in individualized target behaviors is tracked and rewarded by the parents. Children are disciplined for certain behaviors, with discipline in the form of loss of privileges or time out from ongoing activities. If the standard interventions do not produce the desired behavior change for a child, a functional analysis of the problematic behavior is conducted, and an individualized program is developed.

The Summer Treatment Program was expressly developed to focus on treating difficulties in peer relationships, an area in which clinical behavior therapy and medication have not had major impact. Deficits in peer relationships are arguably the most severe area of impairment for children with attention-deficit/hyperactivity disorder, and they predict a poor long-term outcome. These children are annoying and aggressive toward peers, and so, they are actively rejected. The primary technique in treating peer relations are dyadic interactions, cooperative talks, social skills training, and group problem solving, all integrated as part of the system and recreational activities. The Summer Treatment Program integrates treatment for peer difficulties into daily activities in which children engage with peers (e.g., soccer, basketball), thus allowing observation and direct modification of a child's dysfunctional peer interactions in a setting close to the natural contexts in which children interact with peers.

Summer Treatment Program staff members expend considerable effort on positive social interactions with the children, frequently praising the children's behavior and their academic and athletic efforts. A 2:1 ratio of point awards and social reinforcement to point loss is targeted, which results in a positive atmosphere. Children like the program in spite of the rigidly applied point system and the ubiquitously implemented rules.

Combined Pharmacological and Behavioral Interventions As discussed above, treatment with certain drugs, particularly those affecting dopaminergic neurotransmission, serves as an establishing condition for the efficacy of reinforcers. It is therefore not surprising that combining behavioral and pharmacological treatments for attention-deficit/hyperactivity disorder has several advantages. First, the behavioral component of treatment can usually be reduced in scope and complexity. Second, the cost-effectiveness of treatment is improved. Third, drug and behavioral treatments often have complementary effects. Stimulant medication can reduce problematic behaviors that are difficult or impossible to treat with practical behavioral programs (e.g., low-rate, peer-directed aggression and stealing that occurs in the absence of adult authority). Finally, if the addition of stimulant medication reduces the complexity of the behavioral intervention, parents and teachers may be more likely to maintain it.

Conversely, adding behavioral treatment to medication could improve long-term outcome. Medication effects on antisocial behaviors are enhanced when behavioral contingencies supporting prosocial behavior are in effect. Furthermore, children with attention-deficit/hyperactivity disorder lack cognitive and behavioral skills that are necessary for academic and social adjustment. To the extent that these skills must be acquired for successful long-term outcome, medication alone would not be

expected to be a sufficient treatment. Studies conducted in the Summer Treatment Program suggest that more than 40 percent of children maintained on medication only would show incremental benefit from the addition of behavioral treatment. Conversely, close to 80 percent of children treated with behavior therapy can be expected to benefit from added stimulant treatment.

Behavioral recordings can be used to titrate the optimum stimulant medication dosage. In addition, such recordings can identify the treatment response profile of a child. Individual responses to combined-treatment approaches differ widely. These differing responses are illustrated in the three cases depicted in Fig. 30.2-4. All three were boys with attention-deficit/hyperactivity disorder participating in the Summer Treatment Program. Beginning on the tenth day of the program, all received 0.3 mg/kg methylphenidate (Ritalin) twice a day or placebo, randomized over days. The figure shows the percentage of time each boy followed activity rules (Cases 1 and 2) or rate of noncompliance (Case 3). The figures graph the children's behavior separately for placebo days and methylphenidate treatment days.

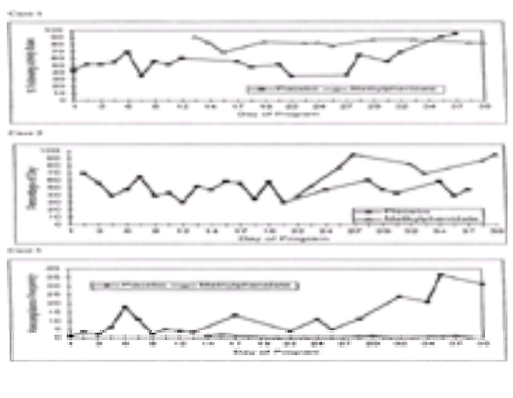

FIGURE 30.2-4 Data for responders to behavioral treatment (Case 1), behavioral and drug treatment combined (Case 2), and drug treatment with methylphenidate (Case 3). The dependent variable for Case 1 and Case 2 is the percentage of daily activities during which the child followed rules. For Case 3, the dependent variable is the daily frequency of noncompliance with commands.

The first boy (age 10) had attention-deficit/hyperactivity disorder only, no comorbid externalizing disorders. The data for behavior therapy plus placebo show that his compliance with activity rules improved dramatically throughout the summer. Addition of methylphanidate resulted in high compliance that remained. By the end of the program, his behavior on placebo days was slightly better than that on medication days, and combined treatment was no longer necessary. Thus, medication was incrementally effective for this child but it did not interact with the behavioral treatment.

The second boy (age 8) had a comorbid diagnosis of oppositional-defiant disorder. Behavioral treatment alone did not sufficiently improve his rate of following activity rules, as evidenced by the lack of an improving trend over the days on behavior therapy plus placebo. However, on days when a 0.3 mg/kg dose of methylphenidate was added to his treatment regimen, he showed gradual improvement, and he reacted good levels of rule following by the end of the summer on medication days only. This boy thus exhibited improved behavior only when medication was combined with behavioral treatment.

The third boy (age 9) had conduct disorder as well as attention-deficit/hyperactivity disorder. The data displayed for him indicate noncompliance (thus, in contrast with the first two cases, lower values are better). His graph shows that noncompliance behaviors steadily worsened over the course of treatment. However, he showed an immediate, positive response to 0.3 mg/kg methylphenidate. Moreover, the drug apparently prevented the deterioration in behavior that the child showed when unmedicated. This boy showed considerable variability in treatment response across domains of dependent variables. Rule-following behaviors responded better to treatment than the noncompliance data shown. This type of finding suggests the effects of a behavioral intervention need ongoing monitoring across relevant behavioral domains and not just on a single target behavior.

These examples show that some children derive benefit from a combined treatment regimen while others do not. This raises an important question: how does a clinician select among different doses and combinations of medication and behavioral treatments to ensure the best possible treatment for an individual child? The answer is through systematic assessment, using ecologically valid individualized measures of behavioral functioning. The daily report card is such a measure and is a sensitive method of assessing whether medication yields additional benefits beyond behavioral interventions.

MODIFICATION OF AUTOMATIC BEHAVIORS (HABITS)

As noted repeatedly above, behavior tends to be organized hierarchically and sequentially. This is true in terms of life goals, intermediate goals, and subgoals and in terms of molar acts versus molecular action patterns. With repeated practice, execution of a lower-level act tends to require less and less attention (i.e., it becomes more and more automatic). For example, learning a new dance involves steps that become more automatic with practice; eventually one can attend to other things, such as conversing with the dance partner.

Early behavioral theorists, such as E. R. Guthrie, postulated that behaviors become linked to each other by movement-produced stimuli, and so, habits develop as chains of behavior. (This view has not fared well in the face of experiments in which curare was used to block such movement-produced stimuli; however the model is still useful for generating ideas for the change of automatic behaviors.) The behaviors in a behavioral chain are prompted both by external stimuli and by stimuli from the previous link. With practice, the movement-produced stimuli become more important, and the external stimuli less so. Finally, response-response chains are created that require very little external input. Concomitantly, the associated movements become simpler, more localized, graceful, and fluent.

Once established, automatic behaviors may continue even without reinforcement by the stimuli that established them, because they are still cued by their own movement-produced stimuli. Therefore, some habits, such as nail biting or hair pulling, may persist indefinitely.

Because they do not require attention, ingrained habits can escape detection even by individuals highly sophisticated in the psychology of habits. Guthrie reported the following event:

> Most smokers, while busily engaged in activities not associated with smoking, can go for long periods with no craving. Others find that the craving is strictly associated with such things as the end of a meal, if it has been their practice to smoke at that time. I once had a caller to whom I was explaining that the apple I had just finished was a splendid device for avoiding a smoke. The caller pointed out that I was smoking at the moment.

Learning New Habits Many preventive health behaviors need to become regular occurrences (i.e., occur automatically) to make a difference. Such behaviors include using seat belts, using condoms, flossing or brushing teeth, or taking prescribed medication. To build a new habit, one can identify regularly occurring behaviors and use the stimuli from these behaviors to prompt the new habit.

In interventions to increase medication compliance, patients are asked to identify regularly occurring behaviors, the settings in which they occur, and the time of day when they occur. Examples of such events include brushing teeth, drinking coffee, or taking another medication. The patients are then asked to identify habits that could be paired with an additional step of medication taking, considering how regularly the habit is performed (e.g., coffee drinking on weekends vs. weekdays) and how easily it might be disrupted. For example, brushing teeth tends to be a regularly performed habitual behavior. Once the existing habit has been chosen, the medication bottle is placed adjacent to the habit-evoking stimulus (e.g., the toothbrush). The patient is advised to take the medication contingent on this stimulus (i.e., prior to the habitual behavior).

Unlearning Habits A habit can be unlearned by interfering at multiple points in the behavioral chain or, more easily, by sidetracking the habit by attending to events at the beginning of the chain.

As an example of sidetracking, consider the famous problem of the child who has developed a habit of entering the house with dirty shoes. The child left muddy tracks through the hallway into the kitchen. In response to the mother's admonition, the child would take the shoes off in the kitchen. However, despite repeated sequences that end in the kitchen, the child continued to enter the house with dirty shoes. As a sidetracking intervention, this mother told the child entering the kitchen in dirty

shoes to (1) go outside, (2) reenter the house, and (3) take the shoes off immediately at the designated proper location. The maladaptive habit was sidetracked because the entry point into the house now prompted the desired behavior.

Habit Reversal Training Habit reversal training is a comprehensive approach toward eliminating dysfunctional habits that can be used to treat nail biting, hair pulling, and tics, including the verbal tics of Tourette's syndrome. Research indicates that the essential components of habit reversal training are awareness training and the contingent use of a competing response. If applied to children, positive reinforcement by the parent is valuable. Habit reversal can be taught in a few sessions.

AWARENESS TRAINING The patient is asked to describe the habit verbally in great detail and to demonstrate it to the therapist. To increase awareness further, the therapist might ask the patient to interrupt the behavioral chain at any point. Attention is also given to attempts to conceal the behavior (e.g., pretending to clean the space between teeth while actually nail biting).

Awareness training also involves identifying the triggering stimuli. For example, the patient's hands must be unoccupied for hair pulling or nail biting to occur. Stuttering occurs particularly when a patient is speaking on the telephone or interacting with unfamiliar people; conversely, singing or shouting tends to prevent stuttering. Behaviors that represent the beginning of the dysfunctional habit are also identified. For example, patients may breathe shallowly or hold their breath before they begin to stutter. Touching the face in some manner tends to precede nail biting or hair pulling. As a homework assignment, patients can be asked to write a brief report of the steps involved in the habit and the circumstances leading up to it. They also begin to self-monitor the frequency of the habit using golf counters or self-recording forms.

COMPETING RESPONSES AND REPAIRING BEHAVIORS Competing responses are designed to be incompatible with the habits. Patients apply the competing response when they enter a habit-prone situation or as soon as they find themselves engaging in a prehabit or habit behavior and maintain it for 1 to 3 minutes, or until their subjective urge diminishes. Thereafter patients engage in repairing behaviors, to remove stimuli that might prompt future incidences.

For hair pulling and nail biting, the competing responses include a fist-clenching maneuver, followed by a repairing response and (if necessary) additional clenching. For nail biting, repairing involves smoothing rough edges with an emery board. For hair pulling, repairing involves combing or brushing the hair. In the competing response, clenching, the patient grasps something or simply makes a fist, with the fingers closing over the thumb with enough tension to feel pressure in the fingers but less than is required to whiten the knuckles. Care is taken to keep the competing response from being obvious to others. For example, while reading a book, the patient might hold the open book firmly with both hands.

For stuttering, the competing response involves altering the pattern of respiration before and while speaking. This includes prolonging the inspiration preceding speech, avoiding breath holding, and exhaling about one third the air before beginning to speak. Also, vowel sounds are lengthened slightly. Importantly, the flow of air during exhalation should not stop at the onset of speech. Instead, patients are instructed to “breathe into” the word and continue to exhale past the end of the phrase spoken. Phrases should be kept short so patients do not run out of breath. Patients first practice with vowels, then consonants, and then by reading aloud. If they have trouble with particular sounds (such as /P/ or /K/), the dictionary can be consulted and reading exercises can focus on those sounds. In situations likely to lead to stuttering, patients should think about what to say before speaking. If a stutter occurs, they immediately stop speaking and use the competing response.

Initial research studies found these techniques quite effective. Table 30.2-11 lists a number of these studies, selected to illustrate the variety of competing responses. However, a recent clinical replication series obtained disappointing results. Among patients treated with a 9-session program in which one of the sessions was devoted to habit reversal, 12 of 14 patients who completed the treatment responded, but after follow-up, most of them had relapsed. However, since only one of the sessions was devoted to the component that had proved essential in other studies, perhaps these results can be improved.

Table 30.2-11 Overview of Habit-Reversal Procedures

RELAXATION THERAPY

Muscle relaxation is used either as a component of treatment programs (e.g., systematic desensitization) or as treatment in its own right (relaxation therapy). Relaxation is characterized by (1) immobility of the body, (2) control over the focus of attention, (3) low muscle tone, and (4) cultivation of a specific frame of mind, described as contemplative, nonjudgmental, detached, or mindful.

In the United States the first proponent of relaxation therapy was Edmund Jacobson, who created progressive muscular relaxation at the beginning of this century. His book *Progressive Relaxation* was published in 1929. In Germany, Johannes H. Schulz created a relaxation method based on autosuggestion, called *autogenic training*. The first edition of his book *Das autogene Training,* appeared in 1932. Relaxation therapy as practiced today also includes elements from Eastern meditative practices. Assimilation of these practices into the Western clinical arena occurred in several waves and continues today. For example, the 1970s saw the clinical applications of techniques based on mantra yoga (i.e., transcendental meditation or the Benson One). More recently, mindfulness meditation has entered the clinical arena.

The physiological changes that take place during relaxation are the opposite of those induced by the adrenergic stress responses that are part of many emotions. Muscle tension, respiration rate, heart rate, blood pressure, and skin conductance decrease. Finger temperature and blood flow to the finger usually increase. Relaxation increases respiratory heart rate variability, an index of parasympathetic tone. Recent findings suggest that opioid antagonists can block the effect of relaxation on stress responses.

Indications Because the effects of relaxation seem to be the opposite of those of acute stress, relaxation therapy has been advocated for conditions thought to be related to adrenergic stress responses. Empirical findings indicate that relaxation therapy is effective for migraine and tension headaches and insomnia. Relaxation therapy has also been advocated as a treatment for hypertension. In the first author's experience, however, relaxation has not been more effective than attention placebo for this condition. Relaxation therapy reduced blood pressure during relaxation sessions and “white-coat” hypertension but not ambulatory blood pressure. Relaxation or temperature biofeedback has also been advocated for Raynaud's disease. However, a recently completed multicenter trial showed that the effects of temperature biofeedback did not surpass those of pill placebo.

In psychiatry, relaxation therapy is mainly used as a component of multifaceted broad-spectrum programs. Its use in desensitization is mentioned above. Relaxing breathing exercises (see below) are often helpful for patients with panic disorder, especially that considered to be related to hyperventilation. In the treatment of patients with anxiety disorders, relaxation can serve as an occasion-setting stimulus (i.e., as a context of safety in which other specific intervention can be confidently tried [see section on exposure]).

Contraindications and Side Effects Relaxation therapy has few contraindications. Some claim that patients with severe asthma should use caution when practicing relaxation because the decrease in sympathetic activity could increase airway resistance. In addition, the authors would not recommended relaxation therapy for acutely psychotic patients, patients with agitated depression, or those prone to dissociation because the lack of proprioceptive cues during prolonged immobility might facilitate outer body experiences. For certain indications, self-report of mild depressive symptoms seems to be a negative prognostic sign. For example, in our hands,

relaxation therapy for tension headaches mainly helped patients with a Beck Depression Inventory score of 4 or less.

When patients perform a relaxation exercise for the first few times, they sometimes report feelings of increased anxiety. Some may even have panic attacks. This phenomenon is called relaxation-induced anxiety, or relaxation-induced panic. Relaxation-induced anxiety may be elicited by exposure to feared bodily cues, letting go, intrusive images or thoughts, remembering traumatic events, or discomfort with the interpersonal aspects of the relaxation session. The latter may involve paranoid or sexual ideation. In patients with panic disorder, an increase in carbon dioxide tension may be the cause of relaxation-induced panic.

Ironic Processes and Relaxation Simply telling a patient to relax does not ensure that a relaxed state will be achieved. In fact, trying too hard to relax can have paradoxical effects, as demonstrated in the following experiment conducted by D. M. Wegner and co-workers:

In a recent study, skin conductance measures were used as an index of stress arousal in two groups of normal subjects. Subjects in one of the groups were instructed to relax in the following manner. “Relax as much as you can. Using any technique, I want you to remove all anxiety, reduce all tension in your muscles, and become completely and totally relaxed. It is very important that you feel thoroughly relaxed. Please become relaxed as quickly as possible, but I will give you 5 minutes to achieve deep relaxation.” Subjects in the other group were told to “just sit there” for a 5-minute baseline. Then both groups were required to answer questions from a “new I.Q. test.” The results showed that subjects instructed to relax had higher levels of arousal than subjects told to just sit. This difference was apparent at baseline. In addition, during the quiz the subjects who had been instructed to relax showed larger increases in skin conductance. Thus, deliberate relaxation made the subjects more tense.

In clinical practice, the paradoxical effects of trying too hard are often addressed by referring to the difference between active and passive volition. *Passive volition* refers to a mind-set of letting, rather than making, things happen. For patients who seem to be overly focused on achieving immediate results, the first author has found it helpful to share his observation that the effect will happen “when they least expect it.” Practitioners of mindfulness meditation cultivate nonjudging, patience, beginner's mind, self-trust, nonstriving, acceptance (of what is already a fact), and letting go.

Progressive Muscular Relaxation Progressive muscular relaxation involves sequential tension and relaxation of individual muscles. Tension is included to help the patient focus attention on specific muscles or identify the sensations of a tense muscle. Variants of progressive muscular relaxation differ with respect to how many muscle groups are practiced in a session. Some work with one or a few muscles during a session; others focus on total body relaxation during the first session.

Jacobson's Method Early in this century, a frequently prescribed remedy for various ailments was protracted rest. In Jacobson's own words, progressive relaxation was developed to “seek by scientific means, the most effective form of rest.” Jacobson observed that:

> When an individual lies “relaxed,” in the ordinary sense, the following clinical signs reveal the presence of residual tension: respiration is slightly irregular in time or force; the pulse-rate, although often normal, is in some instances moderately increased as compared with later tests; voluntary or local reflex activities are revealed in such slight marks as wrinkling of the forehead, frowning, movements of the eye balls, frequent or rapid winking, restless shifting of the head, a limb or even a finger; finally, the mind continues to be active, and once started, worry or oppressive emotion will persist. It is amazing that a faint degree of tension can be responsible for all this.

Learning relaxation, therefore, involves cultivating a muscle sense. To develop the muscle sense further, patients are taught to isolate and contract specific muscles or muscle groups, one at a time. For example, patients flex the forearm while the therapist holds it back to observe tenseness in the biceps muscle. (Jacobson used the word “tenseness” rather than “tension” to emphasize the patient's role in tensing the muscles.) Once this sensation is reported, Jacobson would say, “this is you doing! What we wish is the reverse of this—simply not doing ... relaxation is simply a negative.” Patients are repeatedly reminded that relaxation involves no effort. In fact “making an effort is being tense and therefore is not to relax.” As the session progresses, patients are instructed to let go further and further, even past the point when the body part seemed perfectly relaxed.

Patients would work in this fashion with different muscle groups, often over more than 50 sessions. For example, an entire session might be devoted to relaxing the biceps muscle. Another feature of Jacobson's method was that instructions were given tersely so they would not interfere with the patients' focus on muscle sensations; suggestions commonly used today (e.g., “your arm is becoming limp”) were avoided. Patients were also frequently left alone, while the therapist attended to other patients.

Later Adaptation of Progressive Muscular Relaxation Joseph Wolpe chose progressive relaxation as a response incompatible with anxiety when designing his systematic desensitization treatment (see next major section below). For this purpose, Jacobson's original method was too lengthy to be practical. Wolpe abbreviated the program to 20 minutes during the first six sessions (devoting the remainder of these sessions to other things such as behavioral analysis). In a later modification of progressive relaxation, patients completed work with all the principal muscle groups in one session. The specific muscle groups and instructions for this type of progressive relaxation are listed in Table 30.2-12. Once the patients have mastered this procedure (typically after three sessions), these groups are combined into larger groups. Finally, patients practice relaxation by recall (i.e., without tensing the muscles).

Muscle Group	Instruction
Dominant hand and forearm	Make a tight fist, now
Dominant biceps, triceps	Make your upper arm tense by countersensing muscles
Nondominant arm, forearm	Make a tight fist, now
Nondominant biceps	Make your upper arm tense by countersensing muscles
Forehead	Lift eyebrows
Orbital and nose muscles	Squint and wrinkle your nose
Lower cheeks and jaws	Bite your teeth together and pull the corners of mouth back
Neck and throat	Pull your chin toward your chest but prevent it from happening by countersensing muscles in front and back
Chest, shoulders, upper back	Take a deep breath, hold it, and pull the shoulder blades upward (if sitting) or backward (if supine)
Abdominal or stomach region	Make your stomach hard, as if you were going to hit yourself
Dominant thigh	Counterpose extensors and flexors
Dominant lower leg	Dorsiflex foot
Dominant foot	Curl toes upward (not down to avoid cramps)
Nondominant thigh	Counterpose extensors and flexors
Nondominant calf	Dorsiflex foot
Nondominant foot	Curl toes upward (not down to avoid cramps)

Adapted from Bernstein DA, Borkovec TD: Progressive Relaxation Training: A Manual for the Helping Professions. Research Press, Champaign IL, 1973.

Table 30.2-12 Outline of Initial Progressive Relaxation Session, All Muscle Groups

Autogenic Training Autogenic training is a method of self-suggestion that originated in Germany. It involves the patients directing their attention to specific bodily areas and hearing themselves think certain phrases reflecting a relaxed state. In the original German version patients progressed through six themes over many sessions. The six areas are listed in Table 30.2-13 along with representative autogenic phrases.

Theme	Examples of Self-Statements
Heaviness	“My left arm is heavy”
Warmth	“My left arm is warm”
Cardiac regulation	“My heart beat is calm and regular”
Breathing adjustment	“It breathes me”
Solar plexus	“My solar plexus is warm”
Forehead	“My forehead is cool”

Table 30.2-13 Sample Autogenic Phrases

Autogenic relaxation is an American modification of autogenic training, in which all six areas are covered in one session. Autogenic training became popular in the United States in the 1970s. It has been used particularly in the treatment of migraine headache.

Biofeedback Biofeedback has received much media attention and thus is well known. Briefly, biofeedback involves the recording and display of small changes in the feedback parameter. The display may be visual, such as a big meter or a bar of lights, or auditory. Patients are instructed to change the levels of the parameter, using the feedback from the display as a guide. This treatment was based on a discovery that laboratory animals could control their autonomic function using biofeedback under stringent conditions that appeared to eliminate possible mediation by skeletal muscle activity. However, these relevant experiments have not been replicated, a circumstance that removes the scientific basis for biofeedback as a method of control over autonomic function. Biofeedback can be used by itself or in combination with relaxation. Patients with urinary incontinence use biofeedback alone to regain control over the pelvic musculature. Biofeedback is also used in the rehabilitation of neurological disorders.

Biofeedback-Assisted Relaxation The most common biofeedback modalities used in biofeedback-assisted relaxation include frontalis electromyographic feedback or feedback to increase finger temperature. Temperature feedback is particularly interesting as a therapeutic modality because ironic effects from trying too hard are particularly salient with this treatment. Temperature feedback generates tangible events that can be used to discuss such effects with patients.

Mindfulness Practice When relaxation therapy is used as a single treatment, the purpose is to produce some benefit in the patient's everyday life; that is, the effects generated during a relaxation session should eneralize to the patient's everyday activities. Two quite different approaches to ensure such generalization are mindfulness and applied relaxation. Mindfulness training is a component of dialectic behavior therapy. In dialectic behavior therapy, patients are taught the construct of the wise mind, which provides context for the mindfulness attitude. Individuals are instructed to cultivate mindfulness by taking a nonjudgmental stance, focusing on one thing in the moment (one-mindfulness), and doing what works. Mindfulness is taught as having three components: observing, describing, and participating.

Applied Relaxation Applied relaxation involves eliciting a relaxation response in the stressful situation itself. Discussion above showed that this is not advisable right away because of the possible ironic effects of relaxation. Therefore, patients should first practice relaxation in nonstressful circumstances. The method developed by Lars-Göran öst and coworkers in Sweden has been proved efficacious for panic disorder and generalized anxiety disorder. Establishing the relaxation response in the patient's natural environment consists of seven phases of one to two sessions each: progressive relaxation, release-only relaxation, cue-controlled relaxation, differential relaxation, rapid relaxation, application training, and maintenance. Details are provided in Table 30.2-14.

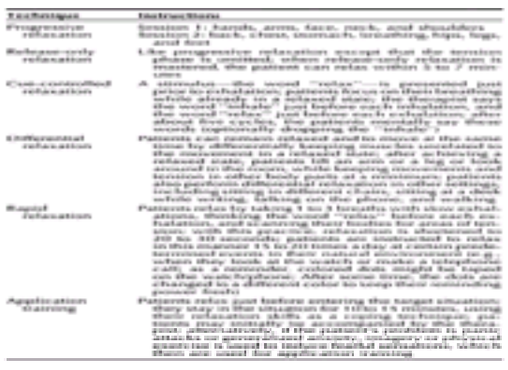

Table 30.2-14 Steps in Applied Relaxation

Applied Tension Patients with blood and injury phobia show a unique, biphasic response when exposed to a phobic stimulus. The first phase is associated with increased heart rate and blood pressure. In the second phase, however, blood pressure suddenly falls and the patient faints. Isometric muscle contractions increase blood pressure. A technique that is the opposite of relaxation, applied tension, can be used to counteract the fainting response.

The treatment extends over four sessions. In the first session patients learn to tense the muscles of the arms, legs, and torso for 10 to 15 seconds (as if they were bodybuilders). The tension is maintained long enough for a sensation of warmth to develop in the face. The patients then release the tension but do not progress to a state of relaxation. The maneuver is repeated five times at half-minute intervals. This method can be augmented with feedback of the patient's blood pressure during the muscle contraction; increased blood pressure suggests that appropriate muscle tension was achieved. The patients continue to practice the technique five times a day. A side effect of treatment that sometimes develops is headaches. In this case, the intensity of the muscle contraction and the frequency of treatment are reduced.

During the next two sessions, the patients are shown a series of slides of mutilated bodies. They are coached in identifying early warning signs for fainting, such as queasiness, cold sweats, or dizziness, and in applying the learned muscle tension response quickly, contingent on these warning signs. During the final two sessions, patients perform applied tension while donating blood or watching a surgical operation.

Breathing Retraining Breathing exercises are among the most ubiquitously prescribed interventions within and outside behavior therapy. Given this, surprisingly little empirical information is available to evaluate their effects critically. For example, the common advice for patients to frequently take a deep breath when symptomatic does not make sense from a physiological perspective.

Physiological Effects The breathing exercises used in behavior therapy were originally associated with Eastern meditative practices. They typically involve slow breathing through the nostrils, with exhalation longer than inhalation and an emphasis on diaphragmatic, as opposed to thoracic breathing. Changes in breathing produce a number of physiological effects. Switching from thoracic to abdominal breathing decreases both cardiac output and peripheral resistance. Switching from mouth breathing to breathing through the nostrils should increase dead space and hence alveolar carbon dioxide tension. It also leads to greater swings in intrathoracic pressure, which in turn affects the venous blood flow into the thoracic cavity and the heart. This effect may be even more pronounced in alternate-nostril breathing, in which the patient occludes one nostril at a time with the finger.

Slowing the respiration rate to eight cycles per minute increases arterial carbon dioxide tension, reduces cardiovascular reactivity to stress, and increases respiratory sinus arrhythmia, an index of parasympathetic tone. Parasympathetic tone has been of interest recently, because reduced respiratory heart rate variability (consistent with decreased vagal tone) may constitute a cardiovascular risk factor, at least in post–myocardial infarction patients. Breathing more deeply also increases respiratory sinus arrhythmia but may increase heart rate and muscle blood flow. Holding one's breath substantially decreases the heart rate.

Clinical Methods of Breathing Retraining The principal methods of breathing retraining are natural breathing and paced breathing.

NATURAL BREATHING Involves abdominal breathing through the nostrils. Upper chest breathing (the heaving bosom) is avoided. Patients let their breathing occur at its own natural rate. Before natural breathing can occur, however, faulty breathing habits must be corrected. Patients suffering from anxiety often hyperventilate—their respiration tends to be either too deep or too rapid. Even so, they often complain of air hunger. This may be related to a habit of not allowing complete exhalation, as they begin the next inhalation too soon. In addition, such patients may previously have been told to take deep breaths as a way to relax; however, because deep breathing increases dead space ventilation, carbon dioxide tension (PCO_2) should decrease, resulting in the symptoms of hyperventilation (e.g., dizziness).

Patients can learn abdominal breathing by placing one hand on the abdomen and the other on the upper chest and watching the degree of excursion. Patients may be encouraged to let the rate and volume of breathing decrease by prolonging the exhalation phase and allowing a brief pause at the end of exhalation. Exhaling against a partially closed glottis (making the sound of the letter H but with mouth closed) is a way of prolonging the expiration. Since this involves exhaling against slight resistance, making the H sound should be accompanied by an increase in intrathoracic pressure during exhalation similar to (but much less than) what occurs during the Valsalva maneuver. This might lead to an increase in vagal tone. A breathing technique used in the rehabilitation of patients with chronic obstructive pulmonary disease is exhalation through pursed lips. Pursed-lips breathing tends to prolong exhalation and increase expiratory load. This technique increases arterial oxygen

saturation in pulmonary patients. (It is possible that the H-sound technique has similar effects, although no data seem to be available.)

PACED BREATHING Involves inhaling and exhaling for specific durations. Patients may be instructed to inhale while counting to 2, then exhale while counting to 4 (one count per second). They then work slowly on prolonging the cycle. Breath holding may be added between the inhalation and exhalation (e.g., inhaling for 2 seconds, holding breath for 4 seconds, exhaling for 4 seconds). Paced breathing is useful for patients who feel they have no sense of control over their breathing, including patients who hyperventilate. It might work by providing a therapeutic context for respiration-produced stimuli that had been anxiogenic.

THERAPEUTIC EXPOSURE

Along with contingency management procedures, the treatment of phobias is one of the most important contributions of behavior therapy. The behavioral treatments for phobia have one feature in common, namely, exposure to the feared stimulus. Historically, exposure-based treatments were based on principles of Pavlovian conditioning as they were formulated in the late 1940s. Below, exposure techniques are introduced in this context; in later sections, these and other techniques are discussed from the perspective of newer theoretical developments.

Behavioral methods to treat anxiety were developed on the assumption that stimuli induced fear because they served as warning stimuli, or conditional stimuli, for the future occurrence of an aversive unconditional stimulus. It was further assumed that by the time the patient sought treatment, the conditional stimulus–unconditional stimulus contingency was no longer operative. However, because of persistent avoidance or escape behaviors, the patient had not been able to contact the new contingencies, which resulted in maladaptive fears. Behavioral methods, in essence, seek to bring the patient in contact with the prevailing contingencies. This means that the individual must be brought in contact with (i.e., be exposed to) the warming stimulus, to learn firsthand that no dangerous consequences will ensue.

Exposure-based treatments were developed according to two competing rationales, counterconditioning and extinction. Methods based on counterconditioning are called *desensitization*. The presentation of the fear-evoking stimulus is paired with a behavior that induces an emotional state incompatible with anxiety (i.e., counterconditioning; see section on signaling operations). In the first reported application of counterconditioning by Mary Cover Jones in 1924, a child's fear of rabbits was treated using food as a counterconditioner, but food is not a good universal counterconditioner for adults. Another way of creating an anxiety-inhibiting state is through deep muscle relaxation. In this case, patients attend to the phobic stimulus while they are in a state of deep muscle relaxation. If anxiety develops, the exposure episode is terminated. Less common counterconditioners include assertive behaviors, relief from an aversive stimulus such as mild electric shock or holding one's breath, or the therapist's encouragement and support.

For counterconditioning to work, the anxiety level experienced with the phobic stimulus has to be minimal. In addition to using the counterconditioner, treatment is arranged to begin with stimuli that induce minimal levels of anxiety. Only after exposure to this stimulus occurs without anxiety does treatment progress to a stimulus with higher anxiety potential. To this end, a number of anxiety-evoking stimuli are collected and ordered by their anxiety-inducing potential. Treatment is then implemented on this hierarchy of situations. For example, in the child treated by Jones, a rabbit was placed closer and closer to the child (little Peter) while he was eating.

In methods using the extinction rationale, the aim is to arrange for unreinforced exposures to the conditional stimulus; that is, the signal is presented but the feared consequence does not appear. These methods involve prolonged exposure while escape behaviors are being blocked. In this case, anxiety first increases and then diminishes when the feared consequences do not materialize. These methods, associated with high levels of anxiety, are referred to as “flooding.”

The methods just discussed can be applied to imagined stimuli or real stimuli (in vivo). This results in four basic approaches: in vivo exposure, including in vivo flooding and in vivo desensitization, and imaginal exposure, including imaginal desensitization and imaginal flooding. A related procedure is “implosion,” which is similar to imaginal flooding, but the therapist takes greater liberty in embellishing the stimuli to come closer to the inferred core fear in the patient's condition. “Graduated in vivo exposure” means exposure to real-life situations arranged in a hierarchy but without an emphasis on counterconditioning.

Treatments also differ in being self-administered or requiring the participation of a therapist. Thus potentially, eight different treatment approaches exist. “Programed practice,” for example, refers to self-administered, graduated, in vivo exposure. Not all of the eight permutations are viable treatments, however; self-directed flooding is one example.

The most commonly used treatments are the in vivo exposure methods. Imaginal exposure is used when the scenarios feared by the patient cannot easily be created at will. This tends to be the case in posttraumatic stress disorder and some cases of obsessive-compulsive disorder or social phobia. In general, these treatments are highly effective in the treatment of phobias. In fact, one-session exposure treatments lasting half a day have been used for some specific phobias such as fear of spiders. Imaginal flooding and some vivo methods may induce very high degrees of distress that make compliance problematic. In agoraphobic patients undergoing in vivo exposure, heart rates of 160 to 180 beats per minute are common. It is therefore prudent to inquire about the patient's cardiovascular health before embarking on such treatment.

Imaginal Systematic Desensitization Imaginal systematic desensitization was pioneered by Wolpe, a founding father of behavior therapy. Comparative outcome studies have indicated, however, that imaginal desensitization is not as effective as in vivo exposure, particularly with agoraphobic patients. Wolpe himself argued that agoraphobia is a heterogeneous condition requiring a broader behavioral approach that includes targets such as marital conflict, dependency, and medical conditions (e.g., seizures and Mènière's disease. Imaginal desensitization also requires more sessions than in vivo procedures and tends to be tedious to administer. For these reasons, imaginal desensitization is not commonly used in the clinical treatment of phobias today. However, it remains an option for patients who require a low-anxiety approach, such as children or medically compromised patients.

Patients who are candidates for systematic desensitization are first taught deep muscle relaxation. They also receive instruction in rating their current anxiety levels numerically; these numbers are referred to as “subjective units of distress.” The subjective units of distress scale has predetermined anchor points (e.g., 0 = no anxiety, 100 = maximal anxiety). Information elicited about the patient's fears is catalogued and organized into groups unified by a theme. Within a theme, scenes are rank-ordered according to anxiety level. During the first desensitization session, patients initially practice imagining a pleasant scene. This enables the therapist to determine the patients' ability to visualize and (incidentally) to screen for relaxation-induced anxiety. The treatment then proceeds to the scene lowest in the hierarchy. The scene is presented verbally by the therapist. The patient is instructed to raise an index finger when the image is clear. The therapist then lets the patient imagine the scene for 5 to 7 seconds. At this time, the patient provides a subjective unit of distress rating and then focuses on cues associated with relaxation. After about 20 seconds, the next scene is presented. Depending on the rating, this scene is either the same or the next one in the hierarchy. The desensitization program involves about 10 scene presentations in early sessions and more in later sessions. The duration of treatment ranges from 6 to more than 100 sessions.

Imaginal Flooding In contrast to systematic desensitization, imaginal flooding involves only one scene. This scene corresponds to the one on top of a systematic desensitization hierarchy. Below is an imaginal flooding scene from S. M. Turner and D. C. Beidel for a female patient with social phobia.

Jim has invited you to attend Christmas dinner at his family's home. You are very nervous about this because you know there will be a lot of people there. You get to Jim's house and walk in the front door. Everyone turns to look at you and you introduce yourself as Jim's girlfriend. At the dinner table everyone is sitting rather quietly when Jim's father looks straight at you and says, “Why don't we have Jim's girlfriend say the grace for us tonight.” You feel the adrenaline flow as the center of attention is focused on you. You want to run away from the table but you know that you can't. Your mouth is very dry and your heart is pounding. Your face feels very hot and you are sure that it is red. Just then one of the children points at you laughing and says, “Look how red she is getting!” You try desperately to regain your train of thought but you can't. Your mind is blank. All you can focus on is your pounding heart and your red hot face. Everyone is looking at you, intently waiting for you to start the grace. ou manage to recall a very simple dinner prayer that you used to say as a child but as you begin to speak you hear that your voice sounds very shaky. You finish and you notice that many of the people around the table are now staring at you in surprise and pity. They think you are the most nervous individual they have ever seen. You can hear the whispers around the room as Jim's family members are astonished at how shy you are. You hear things like, “Poor girl. I really feel sorry for her,” and “Boy, Jim sure got himself a quiet one this time.” One of the children blurts out, “What's wrong with her, mommy?” and the child's mother says, “She's awfully shy, honey.” All you can do is look down at your food and avoid eye contact until this nightmare is over, but you continue to hear the comments during the entire dinner and you can feel the stares from the people around the table.

A scene like this is constructed on the basis of material obtained in interviews with the patient. Care is taken to include (1) the sensory qualities of the scene, (2) the patient's action tendencies and responses, and (3) the patient's interpretations (i.e., the personal significance of the scene). Examples of sensory statements are “You go in the front door” and “At the dinner table, everyone is sitting rather quietly,” and “Jim's father is looking straight at you.” Examples of responses or action tendencies include “All you can focus on is your pounding heart,” and “You look down at your food and avoid eye contact.” Examples of significance descriptions

include “People are looking at you in surprise and pity,” “What's wrong with her,” and “They think that you are the most nervous individual they have ever seen.” These significance descriptions are related to the essence, or core, of the fear. The duration of flooding treatment depends on the target problem. The treatment program for social phobia from which the flooding scene came extends over 12 weekly sessions. Flooding sessions tend to be more successful if scheduled more often than once a week.

Imaginal flooding sessions last for 90 to 120 minutes. Patients are seated comfortably, with tight clothing loosened. The room is dimly lit and quiet. The patients are instructed to close their eyes and imagine the scene as described, as clearly as possible. The scene is read slowly and clearly, and the reading is repeated after a few minutes. The flooding session may involve several repetitions of the scene. The interscene interval can gradually be lengthened to 5 to 10 minutes or more. During the session, the therapist stays in touch by periodically asking the patients about their success in imagining the scene. Patients may be asked to describe details of the scene visualized. In addition, the patients are asked to perform subjective units of distress ratings once every 10 to 15 minutes (some therapists ask for ratings every 5 minutes and also ask for ratings of the clarity of the image). Typically, the subjective units of distress levels rise rapidly during the first 10 minutes, peak after 40 minutes, and then slowly decline. The session is terminated after 90 minutes or if the ratings have declined below 50 percent of the peak level. If this has not happened, the session is continued. (This lack of fit with the “50-minute hour” is a major impediment to the use of flooding treatments in fee-for-service settings and has contributed to the current popularity of cognitive-behavioral methods.)

The report of anxiety indicates that the patient has successfully imagined the scene. Patients are not allowed to engage in any coping methods to reduce anxiety. Care is taken so that the patient does not dissociate during the scene, a problem occurring in individuals with posttraumatic stress disorder. Dissociation functionally represents an escape behavior and is managed by the therapist staying in verbal communication with the patient and providing reorienting cues.

In Vivo Exposure Methods In vivo exposure methods are among the most effective behavioral mental health treatments. In vivo exposure is the treatment of choice for phobias and obsessive-compulsive disorder (particularly when symptoms are largely limited to washing compulsions). The two main approaches are flooding and graduated exposure.

Flooding and graduated exposure can be compared with two ways of entering a swimming pool of cold water. Flooding would be equivalent to jumping off the diving board. Graduated exposure, on the other hand, involves repeated immersions of increasingly larger body areas, alternating with strategic retreats. The advantage of the latter method is its more humane nature. For example, mothers teach their children to play in the water using this method, often using facial and vocal cues of support as a counterconditioner. The disadvantage, however, is that the individual might decide that the water is too cold, after all, and therefore fail to complete the procedure.

In Vivo Flooding In vivo flooding involves exposure to a maximally fear-inducing situation, which can be real or contrived. Contrived situations are feasible in anxiety clinic settings with sufficient staff or student resources and can be structured to resemble the core situation avoided by the patient.

> The patient was a 33-year-old female with social fears of eating in public. In particular, she was afraid of being observed by others when chewing and swallowing, particularly at dinner parties. A contrived situation was arranged in which the patient came to the session with a prepared meal and drink. She entered a conference room in which five people in professional attire were already seated along a table. The patient was instructed to eat her meal in front of these individuals. Between bites, she was instructed to look at them often, and they had been instructed to avoid staring contests. She was not to distract herself from her anxiety symptoms. She was to eat her meal slowly, paying attention to the behavior of the observers and to her anxiety symptoms. (e.g., dry mouth or difficulty swallowing). No conversation between the patient and observers was permitted. The observers would look at her and observe her chewing and swallowing behaviors, at times writing comments in a notebook. Occasionally observers would communicate by whispering to each other, exchanging written notes, or giving knowing glances and smiles. The only other communication occurred between the patient and therapist, and this was limited to providing her subjective units of distress rating. The session lasted 90 minutes.

Note that this situation may seem quite traumatizing. However, because the exposure session is long and continued until ratings decline, the patient does not become sensitized.

Graduated In Vivo Exposure Often used for specific phobias, agoraphobia, and obsessive-compulsive disorder, this treatment involves exposure to situations organized along a hierarchy of gradually more fear-inducing situations. When the anxiety is pervasive, several hierarchies might be constructed. The hierarchy must involve only small steps. If patients anticipate that the next situation on the hierarchy involves a large increase in fear, they might engage in avoidance behaviors. They might drop out of treatment before reaching this step; if they go through with the step and experience a surprisingly large increase in anxiety, sensitization may occur, even of the steps already completed.

However, it is sometimes advisable to schedule the most difficult situations in the middle of treatment. This has been advocated in the treatment of patients with obsessive-compulsive disorder so that more time is available for this situation. In addition, because there is less time to build up anticipatory anxiety, approaching the peak in the middle of treatment may make patients less likely to drop out of treatment.

Self-directed Exposure (Programed Practice) During programed practice, patients are on their own or with a safe person other than the therapist. For example, for a patient with fear of driving through tunnels, the tunnels in the area are surveyed and rated for the amount of anxiety they generate. Further permutations can be created by considering whether the patient is riding, driving accompanied, or driving unaccompanied. Again, the steps between successive items should be relatively small. For example, a patient with fear of bridges was treated with a programed practice hierarchy of situations involving decreasing distances to a local bridge. The patient performed all the steps but dropped out before the crucial one, driving across the bridge.

The length of the session depends on how rapidly anxiety levels decline. Patients can be instructed to identify the peak subjective units of distress rating and terminate the session when the maximal rating within a run has declined to 50 percent of baseline. Alternatively, patients who are not quantitatively inclined can monitor whether coming out of the tunnel is associated with a feeling of relief. Relief suggests negative reinforcement of leaving the tunnel, which means that further trials are needed. On the other hand, if the patient begins to feel bored, it suggests that the patient's anxiety has decreased sufficiently and the session can be terminated.

Adaptations of Exposure Methods This section discusses adaptations of classical exposure methods, including response prevention and contamination of the patient's home setting.

Response Prevention Compulsive behaviors in patients with obsessive-compulsive disorder are considered functional neutralizing behaviors that temporarily reduce anxiety. With neutralizing behaviors, patients can quickly undo the effect of formal exposure sessions. Therefore, for exposure to be effective, these patients need *response prevention,* that is, they are kept from performing their compulsive behaviors.

If the patient's compulsions are pervasive or highly imperative, hospital admission is indicated. However, for the admission to be of optimum benefit, the staff must be trained in carrying out exposure and response prevention. For this reason, such patients are best treated in programs with a programatic focus on obsessive-compulsive disorder. During the hospital stay, daily sessions involve imaginal or in vivo exposure to individualized anxiety-inducing stimuli. These are accompanied by response prevention on a 24-hour basis. Such treatment programs are typically completed in 2 to 3 weeks.

Exposure and response prevention often create extreme distress and feelings of revulsion in the patients, making it difficult to ensure their cooperation. In the original inpatient program designed in England, patients were physically restrained when necessary. However, restraints have been replaced by patient cooperation ensured by an adequate treatment rationale, reminders, and persuasive efforts.

Patients with obsessive-compulsive disorder often engage in subtle avoidance behaviors that require alertness and experience on the part of the therapist. For example, one inpatient with contamination fears undergoing response prevention showed remarkably little distress during her inpatient stay. The lack of anxiety suggested that the stimuli driving her anxiety had not been found. One day, clean clothes were discovered sealed in a plastic bag in her dresser. When this bag was opened and the clothes were contaminated, she showed considerable distress. Thereafter, progress was made in treatment.

Contamination of the Home Environment Over the years, patients have often trained family members to help them with their compulsive behaviors. To keep family

members from enabling the patient's compulsive behaviors, they need to be coached in limit setting and other assertive behaviors.

Some patients with obsessions related to contamination have succeeded in creating a virtually sterile home environment; successful response prevention requires that the therapist visit the patient's home to restore a normal degree of contamination. The therapist deliberately contaminates all conceivable areas that the patient has managed to keep clean. This is easier said than done. An excerpt from Turner and Beidel of a patient's report of a home-contamination session illustrates the comprehensive nature of the intervention.

[The therapist] started touching everything very slowly. I was told to follow behind and touch everything she touched. It was like we were spreading the contamination. She touched doorknobs, light switches, walls, pictures, and woodwork. She opened drawers in each bedroom and touched the contents. She opened closets and touched clothes hanging on the rods. She touched the towels and sheets in the linen closet. She went through the children's rooms, touching dolls, stuffed animals, models, Star Wars figures, transformers, and books.

[The therapist] kept talking to me quietly and calmly all the time we went along. I had been anxious when we started, but, as we continued, my anxiety level decreased. At one point, when I had begun to think the worst was over, she pointed to the attic door and said we were going inside. I said “No, that's where the mice were.” She told me I didn't want to have a place in my home that was off limits. I agreed but became very anxious. It was very hard for me to go inside. I began touching the boxes too, but I was very upset. Then, she put her hands down on the floor and wanted me to do the same. I said, “I can't. I just can't.” Julie said “Yes you can.”

[The therapist] spent several hours with me that day. Before she left, she made a list of things for me to do by myself. Twice a day I was to go through the house touching everything the way she had done with me. I was to invite a friend of mine who had a pet to come and visit and also friends of my children who had pets.

Exposure to Stimuli Presented in Virtual Reality Advances in computer technology have made it possible to present environmental cues in virtual reality for exposure treatment. Beneficial effects have been reported with virtual reality exposure of patients with height phobia, fear of flying, spider phobia, and claustrophobia.

Exposure for Problems Other Than Anxiety Although mostly used for anxiety-related problems, exposure-based treatments can be used for other problems as well.

Bulimia Nervosa Two methods of exposure are used in the treatment of bulimia nervosa. One involves exposure to cues that have triggered binge-eating episodes. Of course, patients are not allowed to binge during such exposure (response prevention). Another method involves exposure to stimuli that occur after a meal (e.g., feeling bloated), with response prevention for purging. Exposure treatments are included in multicomponent treatment programs. Outcome data for exposure treatment of bulimia nervosa are variable but suggest that for exposure to be effective, several sessions of 90 to 150 minutes must be included.

Alcohol Craving Exposure with response prevention has also been used for alcohol-related cues. For example, T. Sidhartan and coworkers showed that this treatment decreased drinking behavior in volunteers with acknowledged problems with alcohol use. Cue exposure was provided over six weekly 90-minute group sessions for 3 to 4 clients. Cue exposure resulted in greater reduction in both the amount and frequency of drinking than a standard cognitive-behavioral intervention.

A treatment rationale was provided consistent with scientific knowledge of stimulus-elicited cravings (cue reactivity). For example, the context of a bar, the presence of drinking buddies, and the smell of alcohol or priming doses of alcohol can all be expected to increase the probability of drinking behavior. During the next 30 minutes, subjects were asked to consume priming doses of their favorite drink. Male subjects consumed three, and female subjects consumed two. After these priming drinks, the subjects poured another drink. This time, however, they had to refrain from drinking during exposure to alcohol-related cues. These cues included (1) looking at the drink for 3 minutes, (2) holding the glass in their hands for 3 minutes, and (3) holding the glass near their mouths for 3 minutes, sniffing the content regularly. After a 5-minute break, the exposure sequence was repeated. To prevent cognitive avoidance behaviors, the subjects were encouraged to share their thoughts about drinking, including their desire to drink and any change therein. However, they could not assist other group members.

Between sessions, the subjects practiced this method in their normal drinking situations, after first being introduced to the principles of graded in vivo exposure. They surveyed the coming week with respect to potential drinking-related situations and were instructed to have the same number of priming drinks and then refrain from further drinking. During the next session, the clients shared the results of their homework assignments with the group.

Exposure to Private Events In the historical development of exposure techniques, the treatments were targeted at stimuli that were external to the patient (i.e., phobic stimuli). Although in imaginal desensitization or flooding the actual stimuli are private, they were seen as representations of the external world. However, even with optimal treatment, some patients with agoraphobia relapsed in response to panic attacks. One unexpected panic attack could undo many months of treatment. For this reason, current treatment programs for panic attacks include exposure to the internal sensations that characterize panic attacks. Exposure can also be directed at remembered events, as in the treatment of posttraumatic stress disorder.

Interoceptive Exposure Interoceptive exposure is part of many treatment programs for panic disorder. Interoceptive stimuli are created by maneuvers such as deliberate hyperventilation, spinning in an office chair, or performing jumping jacks. The therapist discourages subtle avoidance behaviors detected during these exercises. Further on in treatment, patients are instructed to arrange exposure through naturalistic activities (e.g., having a cup of coffee, going to aerobics class, eating heavy food, wearing a tight collar, relaxing in a sauna, or driving in a hot car).

Implosion In implosion, exposure is arranged to both remembered events and hypothesized events. Imagined scenes are embellished by the therapist, often along psychodynamic lines, in a successive approximation toward the patient's core fear.

In a newer variant of implosion, the therapist lets the patient take the lead by successions of revised scripts that include material recalled during the presentation of previous scenes. The material is selected on the basis of similarities in exteroceptive cues (e.g., objects such as knives), response-produced cues (e.g., eating, drinking, sex), thoughts (e.g., revenge or death), and emotional stimuli (e.g., guilt, shame). As the scripts evolve, the patient may remember new, previously avoided (repressed) material. The visualization of such scenes evokes intense anguish, described as protopathic pain. Often the event ultimately recalled is one involving severe abuse. The patients may assume childlike behaviors as if they were reliving the scene.

Eye Movement Desensitization and Reprocessing The exposure component of eye movement densensitization and reprocessing involves attending to a recalled traumatic event. During exposure, therapists move their fingers back and fourth in front of the patients; the patients keep their eyes open and follow the therapist's fingers with their eyes. Whether the inclusion of eye movements and other features of this techniques is associated with superior efficacy than that of other forms of exposure treatments is controversial. Although eye movement desensitization and reprocessing has shown efficacy compared with control conditions, superiority over other exposure methods has not been proved. The originators of this technique advise prospective therapists to obtain formal training in the method; elaboration of scenes can take directions that the naive therapist does not anticipate and may find frightening.

In the initial step of eye movement desensitization and reprocessing, patients identify a scene to be worked on (i.e., the remembered event currently causing the most distress) and describe the image, emotions, and body sensations associated with the memory. They also identify a negative thought associated with the memory and consider an alternative positive thought. They then perform a subjective units of distress rating and rate the believability of the alternative positive thought.

During the desensitization part of the treatment patients concentrate on the images, emotions, body sensations, and negative thoughts associated with a traumatic event. At the same time, they perform the rapid-tracking eye movements. The therapist's fingers are kept about 18 inches from the patient's eyes. During the visualization, which lasts 12 to 30 seconds, 24 to 60 eye traverses are done. After the visualization, patients briefly report any changes in the images or concurrent experiences. Then a new visualization cycle is begun, during which patients focus on any new spontaneously generated associations. Similar cycles are repeated until the original scene no longer elicits discomfort. They then perform a body scan; residual sensations suggest that something is left unprocessed and needs continued work. Otherwise, they proceed with the next step, reprocessing. While patients imagine the original scene, they mentally rehearse the positive thought, performing the eye movements as before.

Recent Theoretical Perspectives on Exposure-Based Methods The treatments discussed thus far have been based on models in which a conditional anxiety response was reduced by counterconditioning or extinction of conditional stimulus. These models have several weaknesses. First, conditioning events are less common in patients with phobias than seems reasonable if these methods work primarily by extinguishing conditional anxiety responses. Second, because conditional stimuli are readily extinguished in the laboratory, simple conditioning models do not explain well why avoidance behaviors in phobias are so persistent without treatment. These findings suggest that often times the fear-inducing stimulus should be thought of as an unconditional stimulus rather than a conditional stimulus.

Another issue is whether individuals actually unlearn their fear. Animal experiments suggest that fear conditioning is not unlearned but contextually controlled.

Devaluation of the Unconditional Stimulus The situation described above involving the social phobic patient visiting her boyfriend's family for the holidays would induce social distress in most people. It contains a priming stimulus of the dominance-submission system—the dominant person in the group (the boyfriend's father) staring at the patient—who being an outsider is at the lowest end of the dominance hierarchy. It is hard to imagine how this scene could ever be a neutral stimulus. It makes more sense to consider this scene an imagined unconditional stimulus (which probably is in a bidirectional relation with a real situation and thus is functionally equivalent). The way exposure reduced anxiety to this scene can be described as habituation of an unconditional stimulus, not extinction. The result would be equivalent to devaluation of the unconditional stimulus, a phenomenon described above. As another example, for individuals with fear of spiders, the spider is not a warning signal for something else; it is a primary, unconditional stimulus that belongs to the food rejection module of the feeding system. Spiders often elicit disgust rather than anxiety in these patients.

If exposure is used for habituation, the general idea is for the patient to get used to the unconditional stimulus. The success of exposure for habituation depends on the nature of the unconditional stimulus. Not all stimuli habituate. Patients with chronic pain or tinnitus are exposed to pain or auditory stimuli for prolonged periods without habituation. On the other hand, habituation to disgust appears to occur rather easily.

Habituation is just one way of devaluing an unconditional stimulus; other methods might thus represent alternative approaches to treatment. In humans, additional information about the unconditional stimulus can devalue it. In fact, Wolpe always emphasized that correcting misinformation is an important part of treatment. For example, a gun aimed at you normally evokes a fear response, which would quickly dissipate upon discovery that the “gun” is a toy.

Besides habituation and instructions, an unconditional stimulus can be devalued a third way, based on the intensity of the unconditional response (response feedback). The way this feedback is framed may affect the value of the unconditional stimulus. For example, a social comparison can be made showing that the experience the patient is having is normal and common. In the case of a first panic attack, it might help a potential panic patient to know that onclinical panics occur in one third of the general population. However, framing the feedback as a social comparison may not work, if the unconditional response is highly aversive.

Patients with vestibular dysfunction sway more than normal comparison subjects in response to certain characteristics of the visual environment exemplified by heights. Patients with vestibular dysfunction also tend to have a fear of heights. The increased sway in high places constitutes response feedback that suggests an increased risk of falling, and therefore leads to unconditional stimulus inflation of height stimuli in these patients. However unconditional stimulus devaluation may occur if patients learn how this effect develops and how to cope accordingly.

Research in the authors' laboratory suggests that the fear of heights is an expression of space and motion discomfort that occurs because individuals with vestibular dysfunction prefer visual information over vestibular spatial information. The long visual distances that characterize heights reduce the visual feedback contingent on body sway that a person would normally obtain in closer surrounds. Therefore, even normal persons show increased body sway in high places, and the visually dependent person is even more vulnerable. When patients find themselves swaying for no apparent reason, they may interpret their responses as having an urge to jump. The authors tell patients that their height vertigo has a physiological explanation. It will diminish if they sit down or focus on proprioceptive cues from their feet and ankles.

Debriefing of future unconditional stimuli can also be used to prevent anxiety. For example, riding in the space shuttle during launch and subsequent microgravity presents many fear-inducing stimuli to an inexperienced passenger. Yet astronauts are trained and debriefed for years before launch. A report by an astronaut of his subjective experiences suggested (to the first author) that his lack of fear might have been a consequence of his training, which was geared to minimizing the possibility of unpleasant surprises.

DEVALUING WORRISOME IMAGES BY COGNITIVE EXPOSURE Research reviewed above showed that fearful images of an unconditional stimulus can elicit fear responses and that these may habituate upon prolonged exposure. However, if the patient cognitively escapes from the image (e.g. by directing attention to another thought) the fear remains. This type of reasoning has been used to develop new treatments for generalized anxiety disorder.

When a situation arises that is ambiguous, an individual responds with an increase in vigilant behaviors. Increasing vigilance may result in obtaining information that decreases the ambiguity. However, the increased vigilance comes at the expense of increasing somatic sensations, which may be accompanied by frightening, intrusive images. Verbal worries are less arousing than the imaged fearful events; thus verbal worries may represent escape or avoidance of imaged fearful events. The images can be conceptualized as unconditional stimuli. The purpose of treatment is to devaluate them by prolonged attention to the frightening image. As this is done without interruptions, physiological responses to the image should habituate. This is the principle for a technique called *functional cognitive exposure*.

This treatment involves identifying the most threatening images among the patient's worries, and arranging the images in a hierarchy. Care is taken to identify the most frightening image. One method of finding it is asking serial “what if—” questions (i.e., sequentially asking what will happen after a feared event has occurred). The patient's description of the scary image is recorded on tape. Using a Walkman-type tape recorder, the patient uses the tape for exposure. During the exposure, the patient engages in *cognitive response prevention,* in which methods that the patient uses to reduce fear are identified and proscribed. This technique of cognitive exposure and response prevention is also applied to patients with obsessive-compulsive disorder with pure obsessions.

Worry prolongation, another exposure approach to worries, does not require tape loop technology. This technique is part of a comprehensive cognitive-behavioral program for worry. In preparation, patients first make a short list of their major worries, identifying the worst possible outcome in each of them, regardless of whether or not the outcome seems realistic. They write the worry down or record it on tape, making sure to include the scariest consequence of the event worried about. Two worries are identified in this fashion. The patient will start exposure to the one that is less anxiety provoking.

The actual worry exposure session lasts about 30 minutes. It begins by the patient visualizing the chosen worry image, taking care to include contextual sensory cues in different sensory modalities, as well as action tendencies and feelings. Image formation tends to be easier the scarier the worry is. Thus, lack of success in visualization could mean that the patient chose an aspect of the worry that is not scary enough or that the worry is depicted too abstractly. Once the image is reasonably clear, visualization continues for 25 to 30 minutes. During the visualization, anxiety levels decrease; if they do not, the patient may have inadvertently switched to another worry. Alternatively, the worst consequence of the target worry may not have been included, because it still is being avoided by the patient. The designers of worry prolongation treatment also included a problem-solving task of generating alternatives for the worst possible outcome.

Occasion Setting and Contexts of Safety According to current Pavlovian understanding, neither extinction nor counterconditioning of a fear means that the fear is unlearned; rather, the behavior is brought under the influence of new contextual cues. In a typical animal experiment, a conditioned fear response is first established by classical conditioning. Then, treatment is provided in the form of extinction (presentation of the conditional stimulus without the unconditional stimulus) or counterconditioning (presentation of food after the conditional stimulus). The treatment may occur in the same setting as the original fear learning or in a different setting. Finally, the effect of treatment is tested by examining fear responses to the conditional stimulus in various contexts or by examining how quickly a fear is relearned.

After treatment is completed, typically the conditional stimulus no longer elicits a fear response. However, findings of contextual control on renewal, reinstatement, and reacquisition of the conditional response indicate that the conditional stimulus still has fear-evoking capabilities. *Renewal* means the extinguished fear behavior immediately recovers if the conditional stimulus is presented in a context different from the one in which the treatment occurred. *Reinstatement* means that the fear behavior to the conditional stimulus reemerges after the unconditional stimulus is presented, without a warning signal; importantly, reinstatement occurs only in the context in which the original learning took place.

When the original conditional stimulus–unconditional stimulus contingency is reinstituted, conditional responses are learned much faster than the first time. Reacquisition is especially easy if the second set of pairings is presented in the setting of the original conditioning. Conversely, reacquisition is slowed down if (1) treatment is provided in a setting different from the one in which the original learning occurred; (2) the reacquisition trials are held in a setting that resembles the one in which the treatment was provided, or (3) if cues from the treatment setting are introduced into the reacquisition setting. Research has also shown that fear acquisition generalizes more easily across contexts than does extinction. To compensate for this effect, the treatment effect should be more stable if treatment is provided in multiple contexts.

These findings imply that what is learned is what specific contexts are safe. To obtain robust treatment effects, treatment should be conducted in multiple contexts, including those resembling previous unsafe contexts. When implementing treatment in a previously unsafe context, a salient cue available from a safe context is

helpful.

METHODS OF INTRODUCING NEW CONTEXTS One way of introducing a new context is to provide a coherent rationale for treatment that suggests that “this time it is different.” Practitioners have long been intuitively aware of the importance of having the patient understand the treatment rationale. However, they probably did not think of their rationale-giving behavior as equivalent to providing cues for a new context or as having the effect of devaluing the unconditional stimulus.

Other techniques may be analogous to introducing new contexts or occasion-setting stimuli. In treatment programs for posttraumatic stress disorder, patients are frequently asked to write down their experiences in detail to provide material for the imaginal exposure scenes. However, the writing also provides a new context in which the traumatic experiences are viewed. Incidentally, writing one's deepest thoughts surrounding a trauma can have beneficial health effects even when no other treatment is provided.

In a technique called deliteralization (or semantic satiation), borrowed from acceptance and commitment therapy, a particular word is repeated rapidly for about 1 minute. As the word is repeated, the meaning tends to fade away and attention becomes focused on the sound of the word. Thus, if reading about the lemon exercise in the beginning of this section had sensory effects on the reader, this effect can now be undone by saying “lemon” 40 times within 1 minute. Using the same method for the word “fish” should also be informative.

The first author tried deliteralization as a context-introducing device in the treatment of a 38-year-old female advertising professional. In the past, she had received training as an actress. The patient had a nearly 20-year history of panic disorder with agoraphobia and posttraumatic stress disorder. She also had a 14-year history of medical problems including vestibular dysfunction, chemically induced porphyria, chronic fatigue syndrome, asthma, seizure disorder, and benzodiazepine dependence. She had known the therapist for years. Because she had to take a trip out of town for business reasons, the session was spent preparing her for the trip. During the session, she was asked to list thoughts that she would have in an agoraphobic situation. These were first written on the blackboard in the therapist's office and then transcribed and printed out in a font typical for greeting cards (Fig. 30.2-5). This format was used to increase the salience of a new context for these statements.

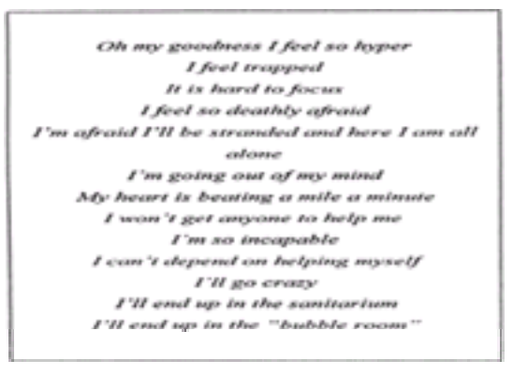
Oh my goodness I feel so hyper
I feel trapped
It is hard to focus
I feel so deathly afraid
I'm afraid I'll be stranded and here I am all alone
I'm going out of my mind
My heart is beating a mile a minute
I won't get anyone to help me
I'm so incapable
I can't depend on helping myself
I'll go crazy
I'll end up in the sanitarium
I'll end up in the "bubble room"

FIGURE 30.2-5 Agoraphobic sentences in a new context.

She was asked to read this list aloud. However, she was afraid that the events described in the statements would actually happen if she read them and was quite reluctant to do this. She began to show signs of dissociation approaching this task. The patient agreed, however, to read the statements, using a Southern dialect. She would repeat the reading several times during the rest of the session, experimenting with various intonations. Along with the humorous responses this produced, the change to Southern dialect made the new acoustic context more salient. The same technique had been used successfully in a previous session involving elevators. It probably also evoked a molar action pattern of acting, which she had done with some success earlier in life. After this session, the patient completed her trip and reported that for the first time in years, she had been able to ride in a taxi by herself. She had made plans to go to a Halloween party dressed up as a southern belle. After the session the patient continued to carry the greeting card with these sentences in her purse. This can be interpreted as a reminder stimulus of the deliteralized context created in the session that facilitated the transfer of safety learning to other contexts.

Guided Mastery Treatment Among the exposure-based treatments in current use, guided mastery treatment, although developed with a self-efficacy rationale, seems to provide the best conceptual fit with the mechanisms involved in therapeutic exposure. This approach has been used successfully with various phobic conditions and may be particularly successful for refractory cases of agoraphobia. However, no empirical data are available for its effectiveness for posttraumatic stress disorder or obsessive-compulsive disorder.

The aim of guided mastery is to enhance self-efficacy. In the context of exposure treatments for agoraphobia, self-efficacy refers to how certain the patients feel that they will be able to stay in the situation for a specified time. Increased self-efficacy after an exposure situation seems to be related to the pleasant surprise response discussed in the section on emotions above as necessary for conditioning to lead to therapeutic changes. Agoraphobic patients were asked to predict how fearful they would be during a 90-minute session of in vivo exposure they were about to start. A situation consistent with pleasant surprise would have patients overpredict their fear (i.e., the fear actually experienced is less than predicted). Results indicated that those who had overpredicted their fear showed the greatest increase in self-efficacy after the session.

In general, the most powerful way of changing self-efficacy is through the success or failure of relevant behaviors. Guided mastery treatment thus creates success experiences in phobic patients through change in behavior. Successful coping with an unconditional stimulus should devalue it. In patients with multiple phobias, the one in which performance is the most impaired provides the most room for improvement. Therefore, such phobias are treated first; generalization of the treatment effect may automatically improve the minor fears as well.

The goal of mastery treatment is to increase the patient's performance in phobic situations. By framing the goal that way, the context is changed from eliminating fear to gaining mastery. This distinction is followed through even in minor details of the treatment. For example, rather than rating their anxiety levels subjective units of distress at regular intervals, patients rate their confidence on a scale from 0 to 100. Traditional exposure treatments emphasize enduring high anxiety levels and allowing the anxiety to diminish with time. In guided mastery, patients learn to manage more effectively in the presence of scary stimuli, regardless of whether they are anxious or not. Nevertheless, as patients learn to function in the presence of high anxiety, their fear responses tend to decrease incidentally.

The role of the therapist, or field helper, is to promote the patient's proficiency, flexibility, and independence when carrying out the exposure task. Flexibility means that the patient can carry out the exposure task in several different ways. Independence is promoted by providing less help as the patient's proficiency increases. In the early stages of treatment, the therapist often stays closely involved. If a task becomes too difficult for a patient, the therapist can improvise and create tasks at levels that the patient can do.

For some patients, touching a feared object such as a spider may be impossible. Remember that the fear in such patients is highly mixed with disgust (i.e., law of sympathetic magic). The method used for such patients is *contact desensitization*. In this method, the patient touches the therapist, who is touching the spider. The first step in a typical hierarchy might be having the spider in a glass jar and the therapist touching it with a pencil. The patient, wearing rubber gloves, is asked to touch the therapist's elbow. Next, the patient and therapist might both hold the pencil touching the spider. The next item could be the same except the patient now takes off the gloves.

Later in treatment the therapist enhances the flexibility of performance, by having patients abandon defensive rituals, including self-protective coping behaviors. Information about coping can be elicited directly from patients. One way of setting the stage for greater behavioral fluency is to try a more difficult task first and then return to an easier task. Jokingly called the “piece of cake” principle, this appears to be a good way of setting up pleasant surprises. Patients are also encouraged to vary their behaviors during both the exposure sessions and their everyday lives. For example, a patient who coped with agoraphobia by parking in a specific place at work might be advised to vary where she parks. This approach appears to approximate learning safety in multiple contexts (discussed above).

Finally, the patient's independence is fostered by the therapist withdrawing assistance as soon as possible. For example, a height phobic proceeding from a lower floor, where he had received the instructions described in the last vignette, might be told, “Now do the same thing here that you did on the previous floor.” In addition,

learning the underlying rationales for treatment means the patient can ultimately assume greater self-direction and better maintain treatment gains.

MODIFICATION OF VERBAL RELATIONS—ACCEPTANCE AND COMMITMENT THERAPY

Our search for manipulable causes of behavior has also identified variables that are not manipulable. For example, in the section on private behaviors, several important nonmanipulable variables covaried with behavior, including wanting, believing, feeling, and some forms of thinking and attending. The section on emotions described patients with feelings about feelings. Attempts to eliminate such behavior not only are certain to be unsuccessful, but also are compounded by the additional aggravation from frustrative nonreward. The behavioral mode primed by such frustrative nonreward does not differ fundamentally from that of someone who tries to fix a broken piece of equipment by kicking it. In both cases, the problem is worsened and the individual fails to do what is called for, backs off and searches for controlling events that can be changed.

Acceptance and commitment therapy was developed to address problems that stem from such misguided attempts at self-control, collectively labeled “emotional avoidance.” It capitalizes on behavioral insights into the nature of verbal behavior described above. Acceptance and commitment therapy seeks to countermand learning histories indicating that problematic behaviors are initiated by thoughts or feelings. Patients are taught to look beyond their thoughts and feelings and focus their efforts on altering manipulable causes of behavior.

The word *acceptance* in acceptance and commitment therapy is synonymous with comprehensive distancing. The distancing involves patients dropping their futile attempts to control their mental events. *Commitment* means patients successfully getting on with their lives, in accordance with their values, by selecting only targets that can be deliberately changed. For this purpose, all the behavior therapy techniques described above are at their disposal.

Some of the components of acceptance and commitment therapy can be found in earlier therapeutic modalities, including existential, gestalt and brief strategic therapy. However, these imports have been honed by theoretical insights, especially with respect to verbal behavior. The resulting treatment strategies aim to modify deep-seated verbal conventions, conventions of which most of us would not be aware.

The need for acceptance is introduced to patients by pointing out the futility of trying to reduce their mental distress by willpower. Various metaphors illustrate this. For example, the “struggle” is compared with being trapped in Chinese handcuffs. Another metaphor involves the patient, blindfolded, trying to get out of a large hole. The method used—digging with a shovel—only makes the hole larger.

The patients' mental control behaviors are identified and labeled (e.g., digging). An emotional avoidance–type problem can be summarized succinctly by the maxim, “If you aren't willing to have it, you've got it.” Because of these discussions, patients may experience feelings of hopelessness framed as creative. At the same time, they may feel validated by having their intuition about the futility of all previous rational solutions confirmed.

Acceptance committment therapy includes procedures aimed at undermining, or deconstructing, unhelpful verbal constructs. To facilitate the development of a questioning attitude with respect to verbal stimuli, patients might be told not to believe what the therapist says. Misleading features inherent in our language are exposed, such as the failure to distinguish between valuations and physical attributes. For example, the expressions “a wooden chair” and “a bad chair” have the same semantic structure, yet the attribute “wooden” would persist even in the absence of humans, while the value “bad” would not.

Deliteralization is used in acceptance and commitment therapy to help deconstruct the semantic context of language, not to reduce symptoms. Deliteralization shows a phonetic context for words as an alternative to the usual semantic context. Another technique of language deconstruction involves having patients deliberately change the way they refer to mental events. For example, instead of saying “X,” patients are taught to say “I have a thought that X”; instead of “this bad chair,” patients might say, “I have an evaluation that this chair is bad.” These reformulations prompt patients to find out where did the thought (evaluation) come from (i.e., antecedent and consequent conditions affecting the thought).

Patients are also taught to differentiate between observing and acting, which contribute to time-varying and time-invariant aspects of the experience of self. Exercises are performed in which patients observe the flow of their thoughts like a spectator watching a parade, noting how long it takes until they are hooked by a particular thought and leave the spectator role.

Values are constructs developed in acceptance and commitment therapy for the commitment aspect. The goal of therapy is described as “putting values into action.” The construction of values involves distinguishing between various aspects of the values concept. To introduce a long-term time perspective, patients might be asked what they would like their epitaphs to say. Distinctions are made between values as feelings (not modifiable) and values as acts (modifiable).

Another distinction contrasts chosen and imposed values. Imposed values are pliance based (i.e., maintained by approval from the patient's culture, parent, or therapist), and the patient can provide verbal reasons for them. Values based on choice, on the other hand, are contingency based and so are not easily justified. For example, most people cannot explain why they chose a particular flavor of ice cream.

The patient is engaged in discussion about ideal values in various social roles. Patients are asked (without being encumbered by reality constraints) to specify their values in ten areas: marriage, family relations, friendships, employment, education, recreation, spirituality, citizenship, physical well-being, and therapy. For example, for family relations, patients are asked to write vignettes about the kind of brother, sister, child, or parent they would like to be. These vignettes are reviewed at a later session. Care is taken to identify pliance-based values. Additional distinctions involve achievable versus unachievable goals, and interventions are targeted at the latter. As the values discussions continue, new suppressed values may become apparent. Patients may also be asked if certain values could be further improved.

If the patient has followed a practice of waiting for the right feeling before acting on a value, this pattern is questioned. Acts of love are possible, even when the person does not presently feel love. Acting according to a value is sometimes referred to “valuing with one's feet.” For example, the patient might be asked, “In the direction of what valued end are your feet taking you?”

OUTCOMES WITH BEHAVIOR THERAPY

Standards for Treatment Efficacy In the field of psychotherapy outcome research, recent efforts have focused on identifying treatments for which effectiveness is supported by empirical data. These efforts required a reexamination of what the standards for such empirical data should be for behavioral interventions.

One set of such standards concerns internal validity. For a treatment to be considered efficacious, the outcome data must show statistically significant superiority of treatment date. The outcome measures of the study have to be reliable, and minimum design standards must be met. A single-group design or single-subject design limited to pre- and post-measures has insufficient internal validity. For a treatment to be considered efficacious, a group design must include a minimal intervention control group, including waiting list or assessment only. The most stringent design is one that permits factoring out of improvement related to nonspecific effects of therapist's attention. The minimum requirement for a treatment to be considered specific is a design that permits comparisons with a placebo.

A second set of standards concerns external validity. For example, the results obtained by a treatment must be assessed by relevant measures. External validity also includes the expectation that a result can be replicated, methods must be specified in detail. Treatments are usually effective only for particular problems, thus inclusion and exclusion criteria must be specified and assessed reliably. Finally, since investigators with a special allegiance to a particular treatment tend to obtain the best outcomes with that treatment, the treatment effect must be replicated by an independent investigating team. For a treatment to be considered efficacious or specific, all internal and external validity standards have to apply. A possibly efficacious treatment is one shown to be efficacious by only one investigating team.

Childhood Disorders Other Than Attention-Deficit/Hyperactivity Disorder Important empirically supported treatments for childhood problems (other than attention-deficit disorder) are listed in Table 30.2-15. Cognitive-behavioral treatments are efficacious for internalizing disorders such as childhood anxiety disorders and depression. For externalizing disorders, the efficacy of parent-administered contingency management is well established. One of the most outstanding accomplishments of behavior therapy is the result of intensive operant treatment of children with autistic disorder. Compared with normal children, autistic disorder children tend to show less transfer of learning to previously unlearned behaviors. This means that each new social behavior must be laboriously shaped. Nevertheless, the outcome seems well worth the effort: Compared with control children, treated children showed markedly higher I.Q. scores and were significantly more commonly placed in regular classrooms.

	Contingency Management	Skills Training	Cognitive and Behavioral	Exposure	Specialized live remarks	Relaxation	Self-Control	Remarks
Anxiety disorders			E	X				
Depression		P						
Oppositional and aggressive behaviors	E	X	E					Skills problem solving parenting skills
Autism	P							Spectacular results
Obesity	P							
Preparation for medical and dental procedures					X			

Data from Kazdin AE, Weisz JR: Identifying and developing empirically supported child and adolescent treatments. J Consult Clin Psychol 66: 19, 1998.
E, efficacious; P, probably efficacious; X, treatment is part of a multicomponent program.

Table 30.2-15 Empirically Supported Treatments for Childhood Mental Disorders Other Than Attention-Deficit Disorder

Attention-Deficit/Hyperactivity Disorder A task force of the American Psychological Association to evaluate empirical support for psychosocial treatments recently reviewed all the treatment outcome studies for attention-deficit/hyperactivity disorder. This review used criteria different from those discussed above for empirically supported treatments. Clear empirical support was shown for the efficacy of behavioral approaches to parent training and classroom interventions for this disorder. However, there were no empirical studies of efficacy for many of the psychosocial treatments for attention-deficit/hyperactivity disorder that are most commonly used in mental health settings, such as individual therapy and play therapy. This finding is consistent with extensive meta-analyses showing that there is little reason to believe that nonbehavioral psychotherapeutic interventions have any impact on disruptive behavior disorders.

COGNITIVE INTERVENTIONS Many different types of cognitive-behavioral treatments have been applied to children with attention-deficit/hyperactivity disorder, including verbal self-instructions, problem-solving strategies, cognitive modeling, self-monitoring, self-evaluation, self-reinforcement, and others. Outcome studies are remarkably consistent in showing that cognitive-behavioral interventions do not produce clinically important changes in the behavior and academic performance of children with this disorder. Limited evidence indicates that other types of cognitive-behavioral interventions—such as social skills training and problem solving—may have some clinical efficacy in the treatment of attention-deficit/hyperactivity disorder when they are combined with intensive multicomponent behavioral treatment programs. It might reasonably be argued that these interventions could have enhanced value in maintenance and generalization as adjuncts to a lengthy and intensive behavioral treatment program. No studies have addressed this issue.

DIRECT CONTINGENCY MANAGEMENT IN THE CLASSROOM Numerous studies of classroom response cost programs reveal that they yield dramatic improvement in disruptive behavior and academic productivity, often raising treated children to normative levels of functioning. The acute effects of contingency management, although substantial, are typically not as large or as comprehensive as the acute effects of medication while medication is active.

CLINICAL BEHAVIOR THERAPY Behavioral interventions have been used for children with the specific diagnosis of attention-deficit/hyperactivity disorder for more than 20 years, and more than 60 studies have demonstrated efficacy (Table 30.2-16). Studies have consistently revealed improvement in both classroom and home settings. Thus, clinical behavior therapy of the sort that is likely to be implemented by therapists in community mental health, primary care, and private practice settings yields clinically important improvement on multiple measures in home and school settings. These treatments therefore meet the American Psychological Association task force criteria for a well-established treatment.

Table 30.2-16 Categorization of Studies Reviewed

Although the short-term improvements obtained with clinical behavioral interventions are substantial, they are typically not as large as those obtained in short-term studies with medication, as long as medication is still being administered. However, the skills taught in behavioral interventions are maintained if parents and teachers continue the programs, while the effects of medication disappear when medication is discontinued.

INTENSIVE TREATMENT PROGRAMS The summer treatment program discussed above results in large treatment effects as shown by parent, teacher, and staff ratings of improvement; standardized parent ratings of disruptive behavior; and direct observations of social and academic behaviors, self-esteem, and parental satisfaction. Pretreatment to posttreatment changes in self-concept and disruptive behavior were twice the size of those reported in studies using clinical behavior therapy. Notably, these positive treatment effects were also present for children with characteristics indicating poor prognosis with clinical behavior therapy, including those with comorbid aggression, from single-parent or from low income families. The dropout rate was 3 percent, compared with rates up to 50 percent that characterize other treatment approaches in outpatient settings.

An area that remains to be addressed concerning behavioral treatment of attention-deficit/hyperactivity disorder is cost-effectiveness. On one hand, behavioral treatments are much less cost-effective than stimulant medication. The essential question, however, is the long-term cost-benefit ratio. Medication is inexpensive but has no long-term impact on outcome; that is, it does not reduce costs associated with such outcomes as special education, dropping out of school, delinquency or criminal behavior, and alcohol or substance abuse. The long-term impact of psychosocial treatments has not yet been evaluated. Nevertheless, the summer treatment program cost of $2500 a child (in 1997) might enable a child to stay in the regular classroom who otherwise would have been placed in special education (about $5000 a year), residential treatment ($25,000 a year), or juvenile incarceration ($50,000 a year).

COMBINED BEHAVIORAL AND STIMULANT TREATMENT In 1986 a review was conducted of the 19 studies in which a combination of behavioral and stimulant treatments was used in children with attention-deficit/hyperactivity disorder. Most studies showed superiority of combined treatment on at least one essential dependent measure. For the average child treated in these studies, a combined intervention resulted in greater improvement than did either treatment alone. Studies conducted in the decade after that seminal review have supported this conclusion, routinely demonstrating the superiority of a combined approach to treatment.

One study compared (1) an 8-week clinical behavioral intervention (mainly parent rather than teacher training), (2) stimulant medication with relatively high doses of methylphenidate, and (3) both treatments combined. Outcome was measured after the behavioral treatments had ended, but children in medication groups were still medicated. All groups showed substantial improvement over baseline. However, on a variety of measures the combined-treatment group demonstrated improvement over those receiving either treatment alone. Figure 30.2-6 illustrates this outcome for the three groups on psychiatrist, parent, and teacher ratings of global improvement. Further, the combined-treatment group was superior to both the behavioral and methylphenidate groups in bringing treated children to a normative range.

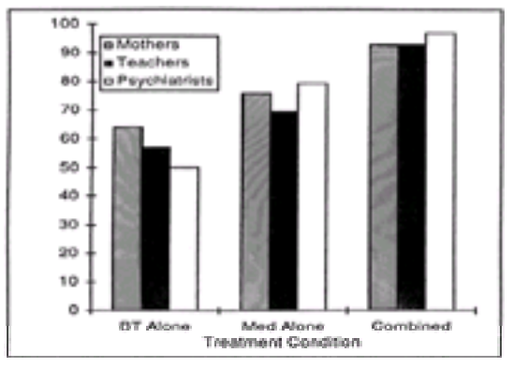

FIGURE 30.2-6 Percentage of patients rated improved by mothers, teachers, and psychiatrists as a function of treatment. (Data from Klein RG, Abikoff H: Behavior therapy and methylphenidate in the treatment of children with ADHD. J Attention Disord *2:*89, 1997.)

A study conducted in a summer therapy program classroom environment showed that both the behavioral intervention and 0.3 mg/kg methylphenidate (slightly less than 10 mg per dose) had salutary effects. However, the combination of the two resulted in behavioral improvement equal to that obtained with a dose of methylphenidate twice as high. This means that it should be possible to reduce the effective dose of methylphenidate—from a high dose to a low-to-moderate dose—by adding a behavioral program in the classroom. This would be particularly valuable for children who develop adverse effects (tics, slowed growth) on high dosages of psychostimulant medications.

Psychiatric Disorders in Adults Table 30.2-17 summarizes recently published reviews applying the standards discussed above for empirically supported treatments for adult mental disorders. The table is limited to those that included a behavioral component. Because early behavior therapists were relatively unconcerned about psychiatric diagnoses, a number of older studies are not included because they were conducted before the need for careful specification of inclusion or exclusion criteria was recognized.

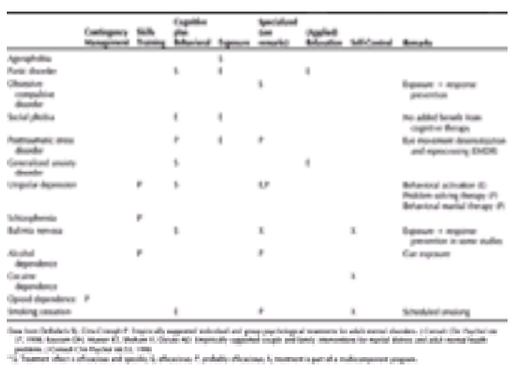

Table 30.2-17 Empirically Supported Behavioral Treatments for Adult Psychiatric Disorders*

Six targeted interventions are considered both efficacious and specific. These include exposure therapy for agoraphobia, exposure and response prevention for obsessive-compulsive disorder, and cognitive behavioral treatment for panic disorder, depression, and bulimia nervosa. The cognitive behavioral treatments represent combinations of behavior therapy and cognitive therapy, as described in the chapter on cognitive therapy. Attempts have been made to identify the active components of these combinations. In our view, the preponderance of the evidence suggests that at least for anxiety disorders, exposure is the most essential component. Exposure plus response prevention has also been used in bulimia nervosa. In those studies in which exposure proved to be an important component, the session length was 2 hours, and exposure was included in most of the sessions. There is room for improvement over the treatments currently available; only 55 percent of patients are in remission at follow-up. The cognitive treatment of depression includes an initial behavioral component aimed at activating the patient. A recent study comparing the full cognitive behavioral program with a condition involving only the behavioral activation component showed no differences between these treatments.

Eight treatments are considered efficacious but not specific. These include cognitive-behavioral treatment for social phobia and cognitive-behavioral treatment for smoking cessation; the behavioral component in these treatments is exposure and self-control procedures, respectively. The list of efficacious treatments also includes exposure therapy for panic disorder, social phobia, and posttraumatic stress disorder, and applied relaxation for panic disorder and generalized anxiety disorder. In general, treatments for generalized anxiety disorder leave room for improvement. Finally, the list includes behavior therapy for depression. Behavior therapy for depression aims at increasing pleasant events during the patient's daily life. The sources for the table did not mention social skills training for depression, although it has been found to be as effective for depression. The reason for its lack of inclusion may have been the absence of treatment manuals (these studies were conducted before the need to “manualize” research treatments had been uniformly recognized.)

Ten treatments are considered possibly efficacious. For anxiety disorders, these include cognitive (nonbehavioral) therapy for obsessive-compulsive disorder, cognitive-behavioral treatment for posttraumatic stress disorder, and eye movement desensitization and reprocessing for posttraumatic stress disorder. For depression, they include problem-solving training and behaviorally based marital therapy. For alcohol or drug abuse, possibly efficacious treatments include social skills training and cue exposure for alcohol dependence, relapse prevention (education in self-control procedures) for cocaine dependence, and contingency management for opiate dependence. Of these, much debate has centered on eye movement desensitization and reprocessing. Does the evidence support claims that it is more effective for posttraumatic stress syndrome than other exposure-based treatments? Most evidence indicates that this therapy can help patients with various anxiety disorders, including posttraumatic stress disorder. However, claims of superior effectiveness compared with other treatments cannot be supported at this time.

A problem with compiling empirically supported treatments is that newer treatments are not well represented, simply because less time has been available to perform the relevant studies. Therefore, two newer treatments are briefly discussed, dialectic behavior therapy and acceptance and commitment therapy.

EFFICACY OF DIALECTIC BEHAVIOR THERAPY In a study evaluating the effect of a 1-year treatment program for patients with borderline personality disorder, dialectic behavior therapy was compared with usual care. Overall, dialectic behavior therapy subjects showed improved behavioral adjustment, although they continued to report feeling distressed. Specifically, they had a lower dropout rate (16 percent), a lower incidence of parasuicidal behaviors, less severe suicidal acts, lower admission rates and inpatient days (8 versus 39), and greater reductions in self-report of angry affect. At a follow-up assessment 18 to 24 months later, dialectic behavior therapy patients showed superior social adjustment and better work performance. They also continued to show fewer parasuicidal behaviors, anger, and fewer inpatient days. In a second cohort, dialectic behavior therapy patients again showed lower anger ratings and better social adjustment. On the other hand, in both cohorts, symptoms involving depressive affect were not differentially affected.

ACCEPTANCE AND COMMITMENT THERAPY Acceptance and commitment therapy was compared with cognitive behavior therapy and a treatment called “partial cognitive therapy.” The latter treatment was included because the complete cognitive therapy program includes elements consistent with distancing. In cognitive therapy, distancing is derived from reframing facts as beliefs; that is, rather than responding to their negative cognitions as facts, patients respond to them as beliefs. The results of this study indicated equivalent improvements in all three groups, with 7 of the 10 to 11 patients in each group classified as treatment responders. Unlike acceptance and commitment therapy, cognitive therapy decreased scores on questionnaire measures of dysfunctional attitudes, suggesting that the effects of these two therapies were not mediated by the same process.

Recently, data have been published concerning the use of acceptance and commitment therapy in an outpatient managed-care setting, using patient measures of problem severity, coping, and acceptance. Eight masters-level psychologists were trained in acceptance and commitment therapy. In addition, 10 therapists did not participate in the training but had taken workshops in other treatment modalities. The patient sample consisted of more than 200 individuals with a variety of

psychiatric disorders or relationship problems. About half were treated before the experimental therapist had been trained in acceptance and commitment therapy, and half were treated after the training. After training, the trained therapists produced better coping outcomes among their patients and significantly improved an acceptance index compared with pretraining outcomes, but the severity of the problem was not affected. These trained therapists also initiated fewer referrals for medication. Interestingly, 80 percent of patients who received acceptance and commitment therapy completed their treatment by 5 months and agreed with their therapists that the treatment was complete. By comparison, after 5 months, 50 percent of patients of untrained therapists thought their treatment was still in progress, although only 22 percent of their therapists thought so.

Behavioral Medicine Treatments The Introduction mentions that behavioral treatments have made inroads into clinical medicine. Table 30.2-18 lists the empirically supported indications for behavioral treatments in the management of medical disorders. An operant-based multicomponent inpatient program has been shown to specifically reduce dysfunctional pain behaviors in low-back pain patients. Biofeedback of the back muscle electromyographic activity is considered a possibly efficacious treatment. Relaxation therapy, at times augmented with cognitive therapy or biofeedback, is a common denominator of the remaining multicomponent interventions. Cognitive-behavioral treatment is rated efficacious and specific for rheumatic diseases and efficacious for reducing distress in cancer patients and reducing the symptoms of irritable bowel syndrome. With respect to the latter indication, however, in one study, cognitive-behavioral treatment did not surpass the effect of a sophisticated attention placebo procedure. Temperature or frontalis electromyographic biofeedback-assisted relaxation is considered efficacious for migraine headaches. Finally, relaxation therapy and systematic desensitization have been used to reduce anxiety or nausea associated with cancer treatments.

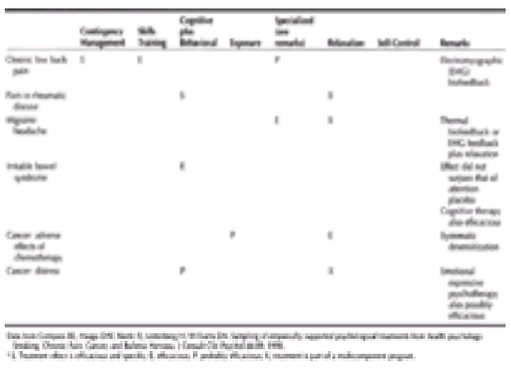

Table 30.2-18 Empirically Supported Behavioral Medicine Treatments*

SUGGESTED CROSS-REFERENCES

For further information on the theoretical foundations for behavior therapy, see Section 3.3 on learning theory. Behavior therapy with adults is often administered concomitantly with techniques of cognitive therapy, which are covered in Section 30.6. Behavioral techniques are used in the service of psychiatric rehabilitation, covered in Section 48.7 and Section 48.8. Behavior therapy also shares features with brief psychotherapy covered in this section. Outcome research is discussed in Section 5.3. Further information on many clinical disorders discussed in this chapter can be found in chapters focusing on them, including attention-deficit/hyperactivity disorder (Chapter 39), tic disorders (Chapter 42), stuttering (Section 37.4), mental retardation (Chapter 34), personality disorders (Chapter 24), eating disorders (Chapter 20), and anxiety disorders (Chapter 15).

SECTION REFERENCES

Acceptance and Commitment Therapy Project: *Acceptance and Commitment Therapy. A Working Manual for the Treatment of Emotional Avoidance Disorders*. Context Press, Reno, NV, 1995.

Bellack AS, Mueser KT, Gingerich S, Agresta J: *Social Skills Training for Schizophrenia: A Step-by-Step Guide*. Guilford, New York, 1997.

Bermúdez JL: *The Paradox of Self-Consciousness*. MIT Press, Cambridge, MA, 1998.

Catania AC: *Learning*, ed 4. Prentice Hall, Upper Saddle River, NJ, 1998.

Craske MG, Barlow DH, O'Leary T: *Mastery of Your Anxiety and Worry*. Graywind Publications, Albany, NY, 1992.

Craske MG, Meadows EA, Barlow DH: *Mastery of Your Anxiety and Panic, II, and Agoraphobia Supplement. Therapist Guide*. Graywind Publications, Albany, NY, 1994.

Chemtob CM, Novaco RW, Hamada RS, Gross GM: Cognitive-behavioral treatment of severe anger in postraumatic stress disorder. J Consult Clin Psychol *65:*184, 1997.

Coie JD, Dodge KA: Aggression and antisocial behavior. In *Handbook of Child Psychology, vol 3*, W Damon, N Eisenberg, editors. Wiley, New York, 1998.

Davidson D: *Essays on Actions and Events*. Clarendon Press, Oxford, 1980.

D'Zurilla TJ: Problem solving therapies. In *Handbook of Cognitive-Behavioral Therapies,* KS Dobson, editor. Guilford, New York, 1988.

Enright RD, North J, editors: *Exploring Forgiveness*. University of Wisconsin Press, Madison, Wisconsin, 1998.

*Esterling BA, L'Abate L, Murray EJ, Pennebaker JW: Empirical foundations for writing in prevention and psychotherapy: Mental and physical health outcomes. Clin Psychol Rev *19:*79, 1999.

Feske U: Eye movement desensitization and reprocessing treatment for post traumatic stress disorder. Clin Psychol Sci Pract *5:*135, 1998.

Fisher JE, Swingen DN: Contextual factors in the assessment and management of aggression in dementia patients. Cogn Behav Pract *4:*171, 1997.

Foa EB, Rothbaum BO: Treating the trauma of rape. *Cognitive-Behavioral Therapy for PTSD*. Guilford, New York, 1998.

Gottman JM, Coan J, Carrere S, Swanson C: Predicting marital happiness and stability from newlywed interactions. J Marriage Fam *60:*5, 1998.

Greenfield PM: Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. Behav Brain Sci *14:*531, 1991, and *21:*159, 1998.

Gunther LM, Denniston JC, Miller JR: Conducting exposure treatment in multiple contexts can prevent relapse. Behav Res Ther *36:*75, 1998.

Guthrie ER: *The Psychology of Learning,* ed 2. Peter Smith, Gloucester, MA, 1960.

Hayes SC, editor: *Rule-Governed Behavior*. Plenum, New York, 1989.

*Hayes SC, Gifford EV. The trouble with language: Experiential avoidance, rules, and the nature of verbal events. Psychol Sci *8:*170, 1997.

Hibbs E, Jensen P, editors: *Psychosocial Treatments for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*. American Psychological Association Press, New York, 1996.

Iwata BA, Pace GM, Dorsey MF, Zarcone JR, Vollmer TR, Smith RG, Rodgers TA, Leannar DC, Shore BA, Goh H-L, Cowdery GE, Kalsher MJ, McCosh KC, Willis KD: The functions of self-injurious behavior. An experimental-epidemiological analysis. J Appl Behav Anal *27:*215, 1994.

Jacob RG, Redfern MS, Furman JM: Optic flow-induced sway in anxiety disorders associated with space and motion discomfort. J Anxiety Disord *9:*411, 1995.

Jacob RG, Shapiro AP, O'Hara P, Portser S, Kruger A, Gatsonis C, Ding Y: Relaxation therapy for hypertension: Setting-specific effects. Psychosom Med *54:*87, 1992.

Jacobson E: *Progressive Relaxation: A Physiological and Clinical Investigation of Muscular States and Their Significance in Psychology and Medical Practice*, ed 2. University of Chicago Press,

Chicago, 1938.

James W: *Principles of Psychology*, vol 2. Dover, New York, 1950 [originally published in 1890].

Kozak MJ, Foa EB: *Mastery of Obsessive Compulsive Disorder*. Graywind Publications, Albany, NY, 1997.

Ladouceur R, Freeson MH, Gagnon F, Thibodeau N, Dumont J: Cognitive-behavioral treatment of obsessions. Behav Modif *19:*247, 1995.

Langewitz WA, Eich P, Kiss A, Wössmer B. Improving communication skills—a randomized controlled behaviorally oriented intervention study for residents in internal medicine. Psychosom Med *60:*268, 1998.

Lerner J, Frankel ME, Meadows E, Hanbree E, Foa EB: Effectiveness of a cognitive behaviors treatment program for trichotillomania. An uncontrolled evaluation. Behav Ther *29:*157, 1998.

Linehan MM: *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. Guilford, New York, 1993.

Linehan MM, Tutek DA, Heard HL, Armstrong HE: Interpersonal outcome of cognitive behavioral treatment for chronically suicidal borderline patients. Am J Psychiatry *151:*1771, 1994.

Lonigan G, Elbert JC, Johnson SB: Empirically supported psychosocial interventions for children: An overview. J Clin Child Psychol *27:*138, 1998.

*Nader K, Bechara A, van der Kooy D: Neurobiological constraints on behavioral models of motivation. Annu Rev Psychol *48:*85, 1997.

Nijenhuis ERS, Spinhoven P, Varnerlinden J, van Dyck R, van der Hart O: Somatoform dissociative symptoms as related to animal defensive reactions to predatory imminence and injury. J Abnorm Psychol *107:*63, 1998.

*O'Donahue W, Krasner L, editors: *Theories of Behavior Therapy—Exploring Behavior Change*. American Psychological Association, Washington, DC, 1995.

*Pelham WE, Waschbusch DA: Behavioral intervention in ADHD. In *Handbook of Disruptive Behavior Disorders*. H Quay, A Quay, editors. Plenum, New York, 1999.

Peterson C, Maier SF, Seligman MEP: *Learned Helplessness. A Theory for the Age of Control*. Oxford University Press, New York, 1993.

Quine WV: *The Roots of Reference*. Open Court, LaSalle, IL, 1973.

Rachlin H: Self-control: Beyond commitment. Behav Brain Sci *18:*109, 1995.

Rusting CL, Nolen-Hoeksema S: Regulating responses to anger: Effects of rumination and distraction on angry mood. J Pers Soc Psychol *74:*790, 1998.

*Sawchuk CN, Lohr LM, Lee TC, Tolin DF: Exposure to disgust-evoking imagery and information processing biases in blood-injection-injury phobia. Behav Res Ther *37:*249, 1999.

Schmajuk NA, Lamoureux JA, Holland PC: Occasion setting: A neural network approach. Psychol Rev *105:*3, 1998.

*Shipherd J, Beck JG: The effects of suppressing trauma-related thoughts on women with rape-related posttraumatic stress disorder. Behav Res Ther *37:*99, 1999.

*Sidman M: *Equivalence Relations and Behavior: A Research Story*. Authors Cooperative, Boston, 1994.

Sidtharthan T, Sitharthan G, Hough MJ, Kavanagh DJ: Cue exposure in moderation drinking: A comparison with cognitive-behavior therapy. J Consult Clin Psychol *65:*878, 1997.

Skinner BF: *Verbal Behavior*. Prentice Hall, 1957 (reprinted in 1992 by the B. F. Skinner Foundation, Cambridge, MA).

Strohsahl KD, Hayes SC, Bergan J, Romano P: Assessing the field effectiveness of acceptance and commitment therapy: An example of the manipulated training method. Behav Res Ther *29:*35, 1998.

Timberlake W, Silva F: Observation of behavior, inference of function and the study of learning. Psychonom Bull Rev *1:*73, 1994.

Turner SM, Beidel DC: *Treating Obsessive-Compulsive Disorder*. Pergamon, New York, 1988.

Turner SM, Beidel DC, Cooley MR: *Social Effectiveness Therapy: A Program for Overcoming Social Anxiety and Social Phobia. A Therapist Guide* . Turndel, Charleston, SC, 1994.

*Walker HM, Walker JE: *Coping with Noncompliance in the Classroom: A Positive Approach for Teachers*. ProEd, Austin, TX, 1991.

Watson JB: Imitation in monkeys. Psychol Bull *5:*169, 1908.

Wegner DM, Broome A, Blumberg SJ: Ironic effects of trying to relax under stress. Behav Res Ther *35:*11, 1997.

Wegner DM, Pennebaker WJ, editors: *Handbook of Mental Control*. Prentice Hall, Englewood Cliffs, NJ, 1993.

Williams SL: Guided mastery treatment of agoraphobia: Beyond stimulus exposure. In *Progress in Behavior Modification*, vol 26, M Hersen, RM Eisler, PM Miller, editors. Sage, Newbury Park, CA, 1990.

30.3 HYPNOSIS

HERBERT SPIEGEL, M.D., MARCIA GREENLEAF, PH.D., AND DAVID SPIEGEL, M.D.

Imagination, a faculty highly developed in humans, facilitates cognitive restructuring, planning, initiation, and implementation of behavior change. Hypnosis, once dismissed as “nothing but heated imagination,” is in fact a powerful means of directing imagination, imagery, and attention to control physical response to stress and pain, change habits, manage dissociative symptoms, and enhance control over psychological and somatic function. With appropriate therapeutic design, this variant of imagination can be activated, identified, measured, controlled, and used for specific therapeutic purposes. Hypnosis can be understood as a form of controlled imagination.

Hypnosis is a useful instrument for the psychotherapist, much as the scalpel is for the surgeon. The hypnotic trance state per se is not therapeutic, but rather it allows the clinician to make systematic use of a person's psychological resources for change; it also facilitates and accelerates the impact of a psychotherapeutic intervention.

As a phenomenon, hypnosis has held an important place in the study of human psychology, ranging from Jean Martin Charcot's and Pierre Janet's early work on through psychoanalysis, through cognitive psychology, to brain imaging. As this body of research continues to grow, hypnotic techniques are becoming more understandable and useful. It is now understood that trance states and hypnotic phenomena occur spontaneously and frequently, with or without formal induction. When persons with trance capacity act with high motivation, concentrate intensely, or find themselves under duress, they tend to shift into spontaneous trance states. Learning to recognize trance phenomena is important in clarifying diagnoses and making treatment choices, even if hypnosis is not used in the formal sense.

The word “hypnosis” (from the Greek root *hypnos*, meaning sleep) is misleading in some ways because the phenomenon to which it refers is not a form of sleep; rather, it is a complex process of attentive, receptive concentration. Although peripheral awareness is reduced in both sleep and hypnosis, focal attention, which is diffuse in sleep, is heightened during the hypnotic trance.

HISTORY

Eastern and Western philosophy, literature, and religion are replete with descriptions of various trance states, ecstatic states, and spontaneous dissociation among both healers and those being healed. These phenomena were first formally described as therapeutic instruments in the eighteenth century with Franz Anton Mesmer's controversial theory that magnetic energy or an invisible fluid could be channeled from a therapist or an object to correct imbalances and restore health to an individual with illness. His unorthodox methods and explanations of magnetic forces attracted negative attention from the scientists of the day and the French government. In 1784 a panel of experts that included Benjamin Franklin, the famous chemist Anton Laurent Lavoisier, and the infamous Joseph Ignace Guillotin met in Paris at the behest of the king. They concluded that the phenomenon was nothing but “heated imagination.”

James Braid, a physician and surgeon in England during the 1840s, observed phenomena similar to what Mesmer had reported. He found he could produce trance states using eye fixation and eye closure. In 1847, Braid departed from the discredited magnetic influence theory and created a psychological concept he called “monoideism”—mental concentration on a single dominant idea. In this state subjects were highly suggestible and could focus their attention on specific ideas that would influence behavior.

Charcot considered the hypnotic state a neurophysiological phenomenon (*sommeil nerveux,* nervous sleep) and a sign of mental illness, “ *un etat* nerveux *artificiel ou* experimental” (an artificial or experimental nervous state). This position about hypnosis was supported by Janet, but both were opposed by Hippolyte Bernheim, who believed that hypnosis was a function of the normal brain. Sigmund Freud studied hypnosis with Jean Martin Charcot, and it was central to his classical work on hysteria with Breuer. Freud documented in his autobiography the moment that he gave up the formal use of hypnosis. A patient threw her arms around his neck during a hypnotic trance: “I was modest enough not to attribute the event to my own irresistible personal attractions, and I felt that I had now grasped the nature of the mysterious element that was at work behind hypnotism. In order to exclude it or at all events to isolate it, it was necessary to abandon hypnotism.” In Freud's mind, the transference underlay hypnosis, and to control or eliminate it, he decided to stop formally using hypnotic inductions, but he retained the couch and turned his attention to the process of free association rather than trance as a therapeutic technique. However, this formal change did not eliminate the occurrence of spontaneous trance in the course of psychotherapy.

Because of the high incidence of shell-shock during World War I, Ernst Simmel, a German psychoanalyst, became interested in hypnosis for the treatment of war neurosis. He developed a technique for accessing repressed material, which he called *hypnoanalysis*. During World War II, hypnosis played a prominent part in the treatment of pain, combat fatigue, and neurosis. Some of the first research on memory recall and control of physiological activity with hypnosis was carried out at the U.S. Army School of Military Psychiatry.

In 1955 the British Medical Society formally recognized hypnosis and recommended that it be taught in medical schools. In 1958, the American Medical Association and American Psychiatric Association followed this example. About this time, the scientific study of hypnosis was made possible by the development of scales for measuring hypnotic capacity.

DEFINITION AND THEORY

Hypnosis can be understood as attentive, receptive focal concentration with diminished peripheral awareness. It is also typified by a feeling of involuntariness; movements seem automatic, and suggested perceptions can alter or replace ordinary ones. The hypnotic state can be entered and exited in a matter of seconds. All hypnosis is, in essence, self-hypnosis, but when a person allows this form of concentration to be structured by another, the hypnotic experience is also characterized by an intense and sensitive interpersonal relatedness between the two, with a relative suspension of critical judgment. This intense concentration can be actively initiated and structured to achieve agreed-upon goals.

Hypnosis is currently understood in a manner consistent with the conceptualization of Bernheim of the Nancy, as a normal activity of a normal mind. The trancelike experience perhaps most familiar in everyday life is absorption in a good novel, play, or movie to the extent that distracting stimuli are ignored and involvement in the activity is intense. It resembles the suspension of disbelief that typifies a good theater experience. Research has demonstrated that persons who report having such intense absorbing experiences can be shown by formal measurement to be relatively highly hypnotizable. In a therapeutic relationship, this facilitates change by teaching an individual to become so focused on a specific theme or goal that they ignore old associations and can more easily accept new thoughts and feelings. Thus the use of the hypnotic state can accelerate therapeutic change.

Laboratory and clinical researchers have demonstrated that hypnotizability is a stable and measurable trait. While it does vary somewhat throughout the life cycle, peaking during the late childhood years and declining for some during adolescence, it remains relatively stable throughout maturity, and then declines further during senescence. A longitudinal study demonstrated a remarkable .7 test-retest correlation for scores on the Stanford Hypnotic Susceptibility Scale over a 25-year interval.

SPECTRUM THEORY OF HYPNOSIS

Hypnotizability has implications beyond the choice of hypnosis to facilitate treatment. Hypnotizability represents a convergence of biopsychosocial phenomena. The hypnotic experience is a process that transforms a trait into a state. The measurement of hypnotizability can be effectively used in the clinical setting to assess and initiate hypnotic experience while making clinical decisions about the type and direction of treatment. The degree of hypnotizability provides useful information about the way in which an individual relates to the self and the social environment.

Components of Hypnotizability The range of ability to experience an attentive, receptive state of concentration with a relative decrease in peripheral awareness and cognitive evaluation is based on sensitivity to inner cues as well as cues from the hypnotist, the external environment, or both. Experiencing hypnotic concentration requires a convergence of three essential components—all of which are necessary to some degree—absorption, dissociation, and suggestibility (Fig. 30.3-1).

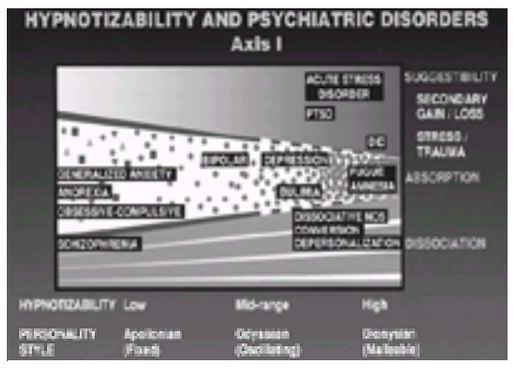

FIGURE 30.3-1 Model of the shift from normal to hypnotic attention. Hypnosis involves narrowing the focus of attention, with concomitant increases in dissociation of thoughts, perceptions, and feelings at the periphery and increased suggestibility. (See Color Plate 8.)

1. *Absorption*—an ability to reduce peripheral awareness to facilitate greater focal attention, analogous to looking through a telephoto lens. The object of attention is seen with great detail, but relatively devoid of context. The shift into the hypnotic state is something like a psychological zoom lens that can shift toward highly focused attention. As attention becomes more intense and focused, awareness of orientation in time and space decreases.
2. *Dissociation*—a functional separation of elements of identity, memory, perception, consciousness, or motor response from the mainstream of conscious awareness. The more absorbed one is in focal attention, the more likely information at the periphery of awareness is to slip out of consciousness. Thus components of self-awareness, temporal orientation, perception, and physical activity may occur without consciousness, or seem involuntary.
3. *Suggestibility*—a tendency to perceive and accept signals and information with a relative suspension of customary critical judgment. The intensity of focus in hypnosis tends to suspend (but not eliminate) the evaluative component that allows us to judge our responses. Motivation, secondary gain or loss, and the degree to which one can suspend cognitive process affect suggestibility. For the highly hypnotizable person, response to input can seem almost compulsive. Less hypnotizable individuals may still respond with a sense of automaticity.

NEUROPHYSIOLOGIC CORRELATES OF HYPNOSIS

Power Spectral Analysis of Electroencephalograms Since the beginnings of hypnosis scientists have been trying to learn more about its neurological basis. Electroencephalographic (EEG) studies and power spectral analyses of brain electrical activity have provided relatively little help in understanding the physiology of the hypnotic state. Even though initially described as a form of nervous sleep, today hypnosis and sleep are known to share no psychological or neurological characteristics. Actually the brain electrical pattern of a hypnotized subject is more similar to that of a fully awake and attentive individual than to the pattern of a sleeping person. Hilgard Ernest and colleagues reported increased alpha activity among highly hypnotizable individuals, which is present whether or not the subjects were in trance. They also found an alpha laterality difference favoring the left hemisphere among highly hypnotizable persons. This difference suggests that hypnosis may preferentially involve the right cerebral hemisphere, since alpha is the noise the brain makes when resting but alert. Thus relatively less alpha on the right suggests more activity. More-recent studies suggest that theta power, especially in the right frontal region, best differentiates high from low hypnotizable individuals. However, no clear signature of the hypnotic state has been found.

Event-Related Potentials Event-related potential (ERP) studies of the effects of hypnosis on perceptual processing have been more productive. ERP studies indicate that alteration of perception under hypnosis changes portions of the waveform affected by inattention and meaning. In particular, a hypnotic hallucination designed to obstruct perception of visual and somatosensory stimuli results in reduction of amplitude of early (P100) and late (P300) components of the waveform. Thus the subjectively reduced perception typical of, for example, hypnotic analgesia, is accompanied by reduced ERP amplitude to perception in that modality. In the visual system this finding has been localized to the left occipital cortex, consistent with work indicating imagery generation in that region. More-recent investigation has identified both common reductions in components such as P100, which are attention-related, and others that are not (P200; Fig. 30.3-2). This finding means that hypnotic alteration of perception differs from simple inattention to the stimulus.

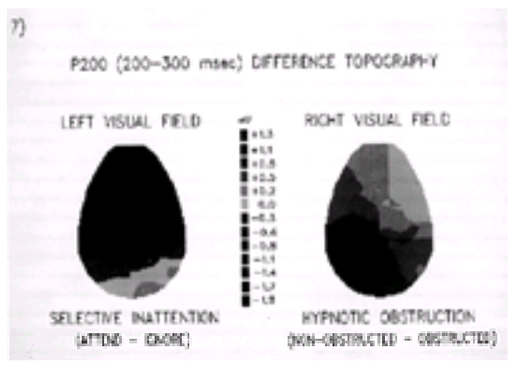

FIGURE 30.3-2 Brain electrical activity mapping of visual event-related potentials comparing the effects of selective inattention to a visual stimulus (attending to the other visual hemifield) and hypnotically hallucinating an obstruction to that stimulus. Selective inattention involves increased amplitude anteriorly, while hypnotic hallucination produces decreased amplitude in the occipital cortex. (See Color Plate 8.)

Similarly, hypnotic amnesia affects attentional components of the ERP and a late negative component (N400) but not late positive components, consistent with intact implicit (but not explicit) memory processing. Recent position emission tomography (PET) studies show that response to true and false recollections can be distinguished through activation of parietal auditory association cortex. Thus modulation of perceptual response to stimuli in the primary sensory association cortices could affect storage and retrieval of memories of these events, potentially linking dissociative responses during trauma that might modulate perceptual encoding to dissociative amnesia afterward.

Similar changes have been found in a small study of ERPs of patients suffering from conversion disorder. The literature has described the effect of selective attention to somatosensory eventrelated potentials (SERPs) and the presence of a characteristic signal appearing about 140 msec after the somatic stimulus is given, known as the N140. It has been suggested that if indeed these patients are constantly monitoring the affected body part or side, they should show the classic N140 signal on the unaffected side, compared with what he calls the enhanced N140 on the affected side. In fact, so far, the small sample in this study has proved this to be the case. Even more interesting, in at least one subject the difference in N140 amplitude completely reverted to normal under hypnosis after suggestions for normal sensation

and functioning of the affected side were given. This normalization disappeared when the subject was allowed to come out of the hypnotic trance with his deficit, as the patient was not ready to give up his symptoms.

Positron Emission Tomography Two PET studies have demonstrated effects of hypnosis in the region of the anterior cingulate gyrus, especially during hypnotic analgesia. This makes sense since that portion of the brain regulates attention focusing along with the frontal cortex. This suggests that during hypnosis there is specific activation of the anterior attentional systems involving focusing (anterior cingulate) and arousal (frontal, especially on the right). Therefore, these imaging techniques may prove useful in further identifying subsystems within the brain devoted to specific types of perceptual and cognitive processing involved in the hypnotic process. Indeed, hypnotizability is correlated with homovanillic acid concentrations in the cerebrospinal fluid (CSF), suggesting the role of the dopamine system (homovanillic acid is a dopamine metabolite), which is heavily represented in the frontal cortex. Homovanillic acid concentration in the CSF primarily reflects activity in the frontal cortex and basal ganglia, which are rich in dopaminergic synapses. Leo Hollister suggested that administration of dopaminergic agents such as amphetamine, may enhance hypnotizability. The automaticity observed in hypnotic motor behavior could represent an activation of the basal ganglia, which is involved in both implicit memory and routine motor activity.

MISCONCEPTIONS ABOUT HYPNOSIS

Despite the aphorism that myths are beliefs that never were true and always will be, there have been so many misconceptions and misunderstandings about hypnosis that it is worth examining and clarifying some of the most prominent ones.

Myth 1: Hypnosis Is Sleep On the contrary, hypnosis is aroused, attentive concentration. EEG studies demonstrate that the hypnotic trance state is consistent with a state of resting alertness and inconsistent with sleep by EEG criteria.

Myth 2: Hypnosis Is Projected Onto the Patient Hypnosis is best understood as the activation of an inner capacity to experience a focused hypnotic state. The use of a hypnotizability test early in treatment can reinforce the idea that the therapist's job is to evaluate the patient's responsiveness, rather than project or force a particular mental state onto the patient. The role of the therapist is to provide an occasion during which persons may identify, explore, and mobilize their own trance capacity. This understanding helps both clinician and patient maintain a proper perspective on the role of hypnosis as an adjunct to a primary psychotherapeutic strategy, whether or not the patient can experience many hypnotic phenomena.

Myth 3: Only Weak or Sick People Are Hypnotizable This tried and untrue proposition has been held at least since the time of Charcot and Janet. It is true that the hysteric patients studied by these men often had symptoms such as conversion paralysis, fugue states, and abreactions, that could be understood as spontaneous, undisciplined trance states. However, the vast majority of highly hypnotizable persons do not have mental disorders.

The capacity to experience hypnosis has, in general, been associated in recent research with relative mental health. Germaine Lavoie and Michel Sabourin noted the absence of highly hypnotizable persons among a population of schizophrenic patients. The authors have found significantly lower hypnotizability among patients with thought, character, and affective disorders. The clinician must bear in mind, however, that about 5 percent of mentally healthy persons are not hypnotizable. On the other hand, some disorders have been found to be associated with high hypnotizability, including posttraumatic stress disorder and dissociative disorders. These observations offer the possibility of a clinical measurement of the capacity for focused attention and treatment responsiveness, which can also be helpful in differential diagnosis. For example, finding that a patient is highly hypnotizable makes a diagnosis of schizophrenia unlikely and dissociative identity disorder or brief reactive psychosis more likely. There are no gender differences in relation to hypnotizability.

Myth 4: Hypnosis Is Therapy The trance state is like an intensely focused beam of light that can illuminate something and allow the subject to absorb it. However, while it may be a pleasant and relaxing state, it is not in and of itself therapy. Indeed, there is no hypnotherapy. Rather, hypnosis is best used to facilitate a primary treatment strategy.

Myth 5: Hypnosis Is Dangerous Nothing in the trance state itself is inherently dangerous, but exploitation by the operator and abuse of the patient using hypnosis can be harmful. This issue is not unique to hypnosis. No one has ever been lost in a trance state or been psychologically damaged merely from entering a trance state under the guidance of a competent therapist. Compared with other psychiatric interventions, which include medication adverse effects and potential for misuse, hypnosis is a benign and safe facilitator of treatment. Furthermore, hypnosis does not occur only when a doctor decides to use it. When concentrating intently in certain private reveries or interpersonal situations and in various responses to stress, approximately 75 percent of the population can spontaneously enter trance states to varying degrees. If the therapist does not recognize spontaneous trance, the therapist-patient interaction can be impaired.

Myth 6: Symptom Removal Is Dangerous Because hypnosis is often used in a brief, symptom-oriented treatment strategy, the theoretical issue of symptom substitution needs to be addressed. Some segments of the psychiatric community still believe that the removal of a psychologically determined symptom before the development of insight regarding the meaning of the symptom by working it through in transference leaves the original conflict unresolved and predisposes the patient to development of a new and possibly more serious symptom. Symptom-oriented treatment is also seen from this classical closed-energy psychoanalytic perspective as undermining the process of psychological growth by making the therapist an ally with the forces of repression of unconscious conflict.

There are two major problems with this position—one theoretical, the other empirical. First, it is difficult to be certain that what is called insight regarding the cause of the symptom is not really a rationalization constructed by the therapist and the patient to explain the conflict. An explanation or narrative truth is not necessarily an insight or the truth. Psychotherapists of various theoretical persuasions may come up with different but equally plausible explanations for the same symptom. Furthermore, the conflict that may have given rise to the symptom may well have long since receded in importance when the patient presents for treatment. The factors that perpetuate a symptom—including secondary gain and loss, habit, and humiliation—may take precedence over the original conflict in importance. Human beings do not fit the paradigm of a closed hydraulic energy system. The therapeutic problem becomes one of breaking the habit and providing a socially face-saving way for patients to change their behavior. The few existing reports of harmful symptom substitution in the literature are less than compelling and do not provide data about patient, doctor expectancies, or use of coercion.

CONTRAINDICATIONS

In general, formal hypnosis is remarkably safe when used with sound clinical judgment in a goal-directed setting. However, clinical hypnotic mischief is more prevalent and occurs when the therapist is not aware of the patient's hypnotic proneness and unwittingly generates iatrogenic phenomena by the way questions are asked or statements are made. Spontaneous trance, without formal induction, can readily occur in patients with midrange to high hypnotic capacity when they are under stress or simply focusing attention. Thus, with trance proneness and formal inductions, some precautions need to be observed.

The clinician should explain briefly and directly the nature of hypnosis, emphasize the importance of hypnotizability as a trait to reduce anxiety about performance or coercion, state that the patient may discontinue the trance experience at any time, and clarify the goals of the hypnotic intervention. From a clinical perspective, hypnosis should not be used in an atmosphere that is threatening or coercive. Patients still may perceive the therapist as exerting considerable power over them, and awareness of and willingness to discuss and diffuse such concerns can be helpful.

Certain types of patients should be approached with caution. Suspicious or paranoid patients usually avoid or resist efforts at hypnosis. Such patients, in whom unconscious fears and suspicions are evoked, may reduce their anxiety by refusing to cooperate with a hypnotic induction. An occasional patient may respond to a trance induction with a spontaneous abreaction. If this occurs, it is important to explore and restructure the patient's experience calmly, using it to enhance the patient's access to and control over upsetting memories, fears, and fantasies. Certain fragile patients who have already suffered many painful failures may be vulnerable if their expectations regarding the hypnotic experience are unrealistic. Discovering that they are not hypnotizable or that they do not achieve symptomatic relief by using hypnosis may constitute an additional failure. Depressed patients usually fail to respond to their full capacity until the depression is treated with more-traditional psychotherapy, medication, or both.

Because perceptions can be temporarily altered by hypnotic signals and attentional focus is narrowed, hypnotized persons may to some extent suspend their usual critical judgment within the trance and depart substantially from their usual conduct in compliance with the hypnotist's signal, despite some resistance and discomfort about it. The more hypnotizable the person is, the greater the likelihood that this can happen, especially if pressed.

In the clinical situation the therapist must accept responsibility for helping to structure the setting appropriately. This means being clear and structured, indicating the beginning, assessment, hypnotic work, and end of the hypnotic experience precisely. More important, highly hypnotizable persons must be taught about their own

vulnerability to exploitation, so that they may take steps to protect themselves from it. The authors have seen more than a few patients who have been manipulated by persons who wittingly or unwittingly used hypnotic techniques to get the patient to comply with their wishes, be they sexual or financial, during a relative suspension of the patient's otherwise intact critical judgment.

One successful businessman found himself signing over an interest in his business to a contractor who, it was later discovered, used hypnosis to get the businessman to abandon his usual business judgment. The examination of hypnotizability helped him to learn a rather expensive lesson regarding his own vulnerabilities.

ASSESSMENT OF HYPNOTIC CAPACITY

A clinical scale to measure hypnotizability must meet many criteria. On one hand, it must be reliable and valid and, on the other hand, it needs to be brief and aesthetically acceptable in the clinical setting. The Hypnotic Induction Profile, a 5-minute clinical assessment procedure, was standardized on a clinical population in a treatment context and meets these criteria (Table 30.3-1). This clinical scale has a moderate positive correlation with the longer laboratory-based Stanford Hypnotic Susceptibility Scale. A derivative of the Stanford Hypnotic Susceptibility Scale developed for clinical use requires about 20 minutes for administration. It was standardized on a nonclinical student population and is highly correlated with the parent scale. The Stanford scales measure behavioral compliance and suggestibility. The Hypnotic Induction Profile measures the convergence of suggestibility with absorption and dissociation and enables the discovery and observation of the degree of involuntariness. Both tests, developed from different perspectives, represent an attempt to make measurement of hypnotizability clinically relevant.

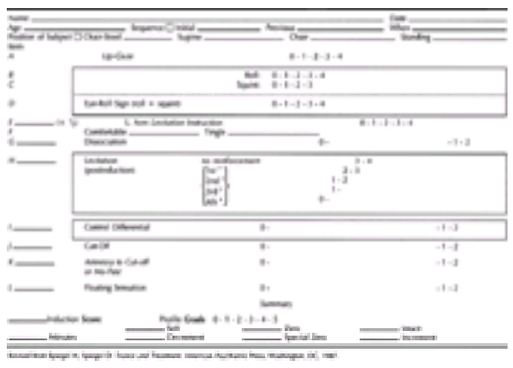

Table 30.3-1 Hypnotic Induction Profile Evaluation Sheet

Hypnotic Induction Profile Administration of the Hypnotic Induction Profile can be a routine part of the initial visit and evaluation. The test begins with the eye-roll sign, a presumptive measure of biological ability to experience dissociation. In the test procedure for eye-roll sign measurement, the patient is told “Hold your head looking straight forward; while holding your head in that position, look upward, toward your eyebrows—now toward the top of your head [up-gaze]. While continuing to look upward, close your eyelids slowly [roll].

The up-gaze and roll are scored on a 0 to 4 scale (Fig. 30.3-3) by observing the amount of sclera visible between the lower eyelid and the lower edge of the cornea. If an internal squint occurs, the degree is scored on a 1 to 3 scale (Fig. 30.3-3). The squint score is added to the roll score. This procedure takes about 5 seconds. The eye-roll is a part of the hypnotic induction, which is also scored as an initial indicator of the potential for hypnotic experience. The eye-roll provides a rapid hypnotic induction that gives initial information that can be compared with more traditional perceptual and motor items that follow.

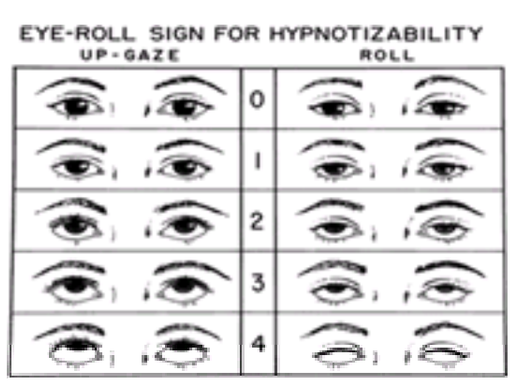

FIGURE 30.3-3 Scoring of eye-roll response to hypnotic induction (© Spiegel 12/76).

Subjects are then instructed to keep their eyes closed but let them relax and concentrate on a subjective experience of floating. As they do that, they are told that their left arm will feel light and buoyant and will float into the air, even after the formal hypnotic experience is ended until a cutoff signal (touching the left elbow) is given.

There are seven scoreable items:

1. Speed with which the subject learns to produce lightness and buoyancy in his left hand and arm
2. Degree of dissociation experienced after the first arm levitation
3. Amount of verbal reinforcement needed to get the arm to rise after its been pulled down by the examiner
4. Absence or presence of involuntariness
5. Speed of response to the cut-off signal
6. Presence or absence of spontaneous amnesia
7. Alteration of somatosensory awareness during the hypnotic experience

The degree of capacity and responsivity as well as the resonance or dissonance between the two correlate with (1) personality style; (2) presence or absence of cognitive flow, implying diagnoses on a health-illness spectrum; and (3) degree of hypnotizability. The degree of hypnotizability provides an initial prediction of the utility of techniques involving hypnosis and likely responses to different treatment approaches.

In conducting research on hypnotizability with the Hypnotic Induction Profile, a population of 6522 consecutive patients in one sample and 2274 consecutive in a second sample (a total of 8796 consecutive patients) were assessed with the Hypnotic Induction Profile by two different therapists. The analysis of data showed the same distribution within each patient population in each therapist's sample. There was confirmation and agreement between the samples that about 75 percent of the subjects were hypnotizable (20 percent low, 48 percent midrange and 7 percent high) and about 25 percent were not. Most of this latter group showed evidence of cognitive impairment because of various psychiatric disorders. These findings are remarkably similar to the Manhattan Project which found that 23.5 percent of the city's population revealed marked to severe psychiatric problems.

The total sample of 8796 out-patients in the study is large enough to suggest that the findings provide stable values for generalization: the results support the hypothesis that hypnotizability can be measured on a continuum, indicating a probable presence or absence of psychopathology. Instructions for the Hypnotic Induction Profile are presented in Table 30.3-2.

Table 30.3-2 Instructions for the Hypnotic Induction Profile

Clinical Correlates of Hypnotic Capacity From a clinical point of view, there are two broad types of response to the test: intact and nonintact. When the sensorimotor responses are approximately consistent with the biological indicator eye roll, the response is intact, which indicates usable hypnotic capacity. When the sensorimotor responses are considerably below the biological prediction, the response is classified as nonintact (soft or decrement). The biological potential for hypnosis as measured by the eye roll can be interfered with by some psychological, neurological, or pharmacological problem and, therefore, is not behaviorally expressed. The more disturbed population—those with thought, character, affective, and some neurological disorders—prove to be considerably less hypnotizable or nonhypnotizable. Approximately one quarter of the adult population has little or no usable hypnotic capacity. In some cases, this is a product of normal development while in others, it is associated with some disorder that impairs concentration. An absence of usable hypnotic capacity indicates that other forms of treatment, either psychotherapeutic or psychopharmacological, are likely to be more effective.

Within the range of intact hypnotic capacity, as measured by the Hypnotic Induction Profile, it is clinically useful to distinguish three broad types of personality style. Those with low hypnotizability, in the 6 to 9 range on the induction score, seem to have Apollonian features. These persons are cerebral and not affect oriented, tend to be inflexible, organized, and somewhat controlling in relationships, have a heightened sense of responsibility, are relatively untrusting, tend to live more in the past that is usually divided among different issues or points in space at the same time. They excel at contrasting and comparing. They often prefer the written to the spoken word. Should they decompensate, they tend toward obsessional symptoms.

Those with high hypnotizability have Dionysian features and reveal an induction score in the 14 to 16 range on a 0 to 16 scale on the Hypnotic Induction Profile. They are relatively heart-oriented and are prone to such intense absorption when they concentrate that they become momentarily disoriented when the task, play, or movie is over. They tend to relinquish control in interpersonal relationships, to be trusting and to value the spoken over the written word, and to learn by affiliating with experience, rather than by critically examining it. They often have exceptionally good memories, and if they decompensate, it is usually in the direction of hysteria or depression. Their trance proneness becomes a vehicle for them to experience and express intrapsychic and interpersonal stress.

Those in the midrange group (scores of 10 to 13) have characteristics that represent a compromise between the two extremes, reminiscent of the compromise made by Homer's wandering hero, Odysseus. The Odyssean time orientation is more a balance of the past, present, and future. They do not have the immersion in the present that characterizes the Dionysians nor the avoidance of the present characteristic of the Apollonians. They tend to have balanced relationships in terms of control and trust, and they learn by accommodating to the new material, rather than by either assimilating it or affiliating with it. If these persons decompensate, they are prone to periods of depression, a syndrome of alternating action and despair.

Psychopathology When a patient presents with a psychiatric problem, the provocation may be related to internal factors (e.g., genetics, drugs, biological deficits) or external factors (e.g., stress, deprivation, trauma), which result in an impairment of psychological function. The authors hypothesize that these factors interact in identifiable patterns consistent with the correlation of hypnotizability and personality style and manifest as predictable Axis I clinical syndromes or Axis II personality disorders (Fig. 30.3-4 and Fig. 30.3-5).

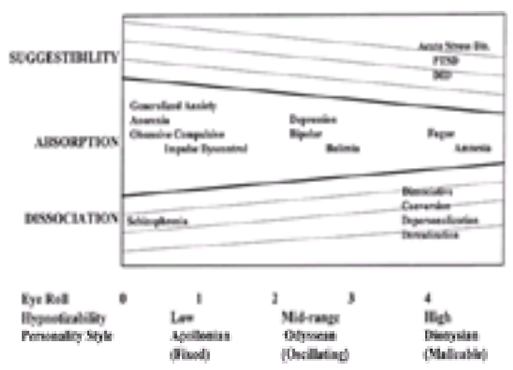

FIGURE 30.3-4 Relation between hypnotizability and Axis I psychiatric disorders.

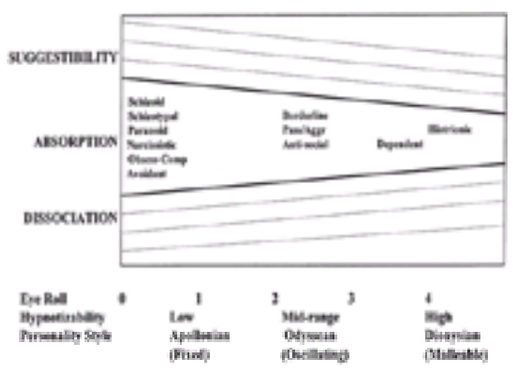

FIGURE 30.3-5 Relation between hypnotizability and Axis II disorders.

For example, the “cerebral” Apollonian type may develop cognitive impairments, with avoidant interpersonal styles and proneness to despair—for example, obsessive-compulsive disorder, anorexia nervosa, generalized anxiety disorder, and schizophrenia (Axis I) and schizoid, paranoid, and avoidant personality disorders (Axis II). The “oscillating” Odyssean type may develop problems of intimacy, fluctuating assumptions, and beliefs with resultant confusion and is subject to mood swings—for example bipolar disorders, major depressive disorder, bulimia nervosa (Axis I), and borderline, passive-aggressive and antisocial personality disorders (Axis II). The “ecologically sensitive” Dionysian is prone to experience disruptions of self-integration and dependency to the point of helplessness and is vulnerable toward major depression—for example acute stress disorder, dissociative identity disorder, dissociative fugue, dissociative amnesia, somatization disorder, conversion disorder, depersonalization disorder, posttraumatic stress disorder (Axis I), and histrionic and dependent personality disorders (Axis II). There is an accumulating body of evidence to support that characteristic coping skills and psychopathology correspond with the different degrees of hypnotizability (Table 30.3-3).

	Apollonian	Odyssean	Dionysian
Primary conflict area	Cognition (Intrusions)	Interpersonal Space (Intimacy)	Self-Concept (Ego Integration)
Interpersonal style	Avoidant	Oscillating	Dependent
Affect			
Anger	Diffuse	Other-directed	Self-directed
Depression	Despair	Bipolar disorder Dysthymic disorder Major depressive disorder	Major depressive disorder

Table 30.3-3 Characteristic Psychopathology

CHOICE OF PSYCHOTHERAPY

Clinical generalizations have proved useful because they help the therapist select the most appropriate problem-solving approach or choice of therapy. Apollonians, in general, prefer intellectual, introspective, insight-oriented therapy, including that of the psychoanalytic type. In contrast to this why-oriented therapeutic approach, Dionysians with psychological problems require a more structured psychotherapy that focuses on what they should do with their lives rather than why they should do it. Their tendency to comply with signals from others is presented to them as a problem in an interpersonal sense and is also used by the therapist to help them structure their approaches to problems. Those in the midrange seem prone to make good use of existentially oriented therapy as they fluctuate between periods of intense immersion in the world and times of withdrawal and questioning. A psychotherapy that helps them work through various disappointments and problems frees them to resume more-active involvement in life.

Highly hypnotizable persons are particularly vulnerable to inappropriate therapy, especially because they tend to affiliate with new premises while suspending critical judgment. The introspective, analytic approach can lead to confusion for such patients because they emphasize compliance with the process, rather than developing insight.

> T.S. was a 22-year-old law student with an excellent academic record who sought psychiatric help after becoming profoundly depressed when someone at a party flippantly implied that he was homosexual. His anxiety impaired his ability to study and he took a leave of absence from school. On the advice of the school physician, he entered analysis and became progressively more depressed during 7 months of therapy. Despite his anxiety about coping during his analyst's 1-month vacation, he noted that his depression improved. When the analysis resumed, his symptoms again increased, and his parents insisted on an outside consultation.
>
> The Hypnotic Induction Profile revealed that he was highly hypnotizable (16/16 on the induction scale). He stopped his analysis and began a psychotherapy that emphasized ego support and guidance. His homosexual anxiety was treated not as a profound truth demanding unconscious exploration, but rather as a function of his proneness to affiliate with the premises of others, no matter how irrelevant. His depression resolved, and he graduated with honors. He has since married, has two children, and is in a successful law practice.

Role of the Therapist Hypnotizability assessment can help the therapist to understand and respond to a given patient's intrapsychic and interpersonal style. A given individual presents with a “cluster of both conscious and unconscious awareness of self, beliefs, assumptions, values, judgments, fantasies, myths, philosophies, spiritual and social striving, and goals.” This includes “the concepts of ego, self, and identity” and can be considered the individual's myth-belief constellation—the private filter system through which persons view themselves, the world, and the flow between the self and the world. This postmodern perspective emphasizes understanding the form more than the content of the particular person's belief system at the time of the treatment intervention. Clinicians choose diagnostic approaches that enable them to learn enough about the person's interpersonal and intrapersonal processes to formulate the problem at hand, respect the aesthetics of the individual, and establish realistic and meaningful therapeutic goals.

Individuals at the low end of hypnotizability tend to be relatively rigid and controlling, and react with an internally driven compulsivity. A person in the midrange of hypnotizability is less rigid and more moderate and most likely to respond with a combination of internal and external processing. Persons at the high end of hypnotizability are so extremely flexible that their processes could be considered an externally driven compulsivity.

At both ends of the spectrum, there appears to be a narrowing of individual choice. However, at the fixed end of the spectrum (low hypnotizability), the person's attachment to content and lack of flexibility in process makes it most difficult to intervene. It is at the most flexible end of the spectrum (midrange to high hypnotizability) that the most promise for change is found.

The task is to identify which aspects of the person are relatively fixed (ecologically insensitive) and which aspects are relatively malleable (ecologically sensitive). With the Hypnotic Induction Profile, these different features are identified and the clinician can observe where the person places on the fixed-flexible continuum. By assaying the components of hypnotizability and identifying personality style and a person's myth-belief constellation, therapists can sharpen their judgment in deciding on the most effective treatment strategy for the person in the given clinical context. This is a nonlinear, multilevel process, with each aspect of information interacting with each other and the whole.

RESTRUCTURING THERAPY: LIFESTYLE CHANGES AND SYMPTOM MANAGEMENT

Since the days of Hippocrates, medicine's main accomplishment has been symptom relief and containment more often than cure. Not only may the goals be more practical, but symptom-oriented work helps build a therapeutic alliance. The starting point of therapy is easily identified by both therapist and patient.

Traditional therapeutic approaches that analyze why a problem exists or explore developmental dynamic interactions may be unnecessary to treat habitual Axis I problems such as smoking or overeating, phobias, anxiety, conversion symptoms, and chronic pain control. These problems can respond quickly, many times in a single session, when patients are taught self-hypnosis with a strategy to take charge of their lives and develop a new perspective on the problem. Regardless of the problem and the basic therapeutic approach, it is useful to teach the patient self-hypnosis to use outside of the therapeutic dyad. The authors use a method derived from the Hypnotic Induction Profile. The patient is taught:

> One, look up toward your eyebrows, all the way up; two, close your eyelids and take a deep breath; three, exhale, let your eyes relax and let your body float.
>
> As you feel yourself floating, you permit one hand or the other to feel like a buoyant balloon and allow it to float upward. As it does, your elbow bends and your forearm floats into an upright position. When your hand reaches this upright position, it becomes a signal for you to enter a state of meditation and increase your receptivity to new thoughts and feelings.
>
> In this state of meditation, you concentrate on this feeling of imaginary floating and at the same time concentrate on (see 3 critical points to stop smoking below).
>
> Reflect on the implications of these critical points and then bring yourself out of this state of concentration called self-hypnosis by counting backward in this manner: Three, get ready; two, with your eyelids closed, roll up your eyes (do it now); and, one, let your eyelids open slowly. Then when your eyes are back in focus, slowly make a fist with the hand that is up; and, as you open your fist slowly, your usual sensation and control returns. Let your hand float down. That is the end of the exercise, but you can retain a general overall feeling of floating.
>
> By doing this exercise 10 different times each day, you can float into this state of buoyant repose. Give yourself this island of time, 20 seconds, 10 times a

day, in which to use this extra receptivity to reimprint these critical points. Reflect upon them, then float back to your usual state of awareness, and then continue with what you ordinarily do.

To make the technique more useful, the patient is taught a camouflaged method to enter the self-hypnotic state with the eyelids closed and bringing one hand up to the forehead. To an outsider, it looks like the person is in deep thought.

The impact of a variety of strategies is intensified with hypnotic concentration to internalize new perspectives on old problems. The principle of restructuring is to help a person master something new rather than fight against an old problem or pattern. The therapist and patients form a collaborative relationship, with the therapist as a teacher, encouraging patients to develop their own executive control. This sets the stage for the next phase of treatment, if necessary, to deal with comorbid psychiatric disorders.

Keeping in mind that human beings use their imaginations to create explanations and meaningful narratives, formal hypnosis helps the person discover relevant control of unconscious process and cognitive self-awareness to develop mastery over anxiety, confusion, panic, and unhealthy life choices. Critical points and suggestions are reinvoked by patients with a method of self-hypnosis that they have learned in the session.

INDICATIONS

Smoking Cessation Hypnosis has proved surprisingly effective in helping people stop smoking. There are several mechanisms by which it may contribute to smoking cessation. The ritual of the hypnotic exercise may provide a substitute for the “breathing exercise” that accompanies the act of smoking; the positive affirmations often used in self-hypnosis provide positive reinforcement for behavior change and promote positive self-image; its use enhances self-observation and self-monitoring; and it can facilitate cognitive restructuring of the smoking habit.

It is not most effective to tell the patient who wants to stop smoking “don't smoke.” This can have the paradoxical effect of directing more attention toward the smoking habit. Rather, the patient is presented with a paradigm shift from “smoke/don't smoke” to “respect my body.” Many points are made to emphasize the power of taking a stand for one's own body rather than against smoking. Emphasis is placed on the significance of learning to use one's mind to protect one's body from the damage inflicted by smoking. The hypnotic state is used to illustrate the importance of focusing on a mental theme that has more power than a physical urge. The patient is taught a method of self-hypnosis (see above) and these three critical points:

1. For my body, smoking is a poison.
2. I need my body to live.
3. I owe my body this respect and protection.

Patients reflect on the personal meaning of this for themselves and then exit the trance state. The self-hypnosis exercise takes 20 seconds and is prescribed to be practiced 10 times a day (until it becomes an integral part of the person's mind-body pattern) as a private meditation to reinforce independence and mastery with an ability to transform an urge to smoke into an act to protect the body. The success of this single-session intervention compares favorably with other techniques, including pharmaceutical treatments.

A number of studies demonstrates the efficacy of hypnosis as a tool to facilitate control of smoking habits. These studies show success rates in cigarette abstinence after treatment with hypnosis ranging from 13 to 64 percent. In these studies abstinence is defined as no smoking during a follow-up time of at least 6 months. The first author of this section developed a single-session approach for smoking cessation that is widely used. His results have been replicated and shown to produce outcomes of 20 to 35 percent long-term complete abstinence. Others have reported abstinence rates as high as 40 percent at 6 months. These numbers are better than the rates of unassisted quitting. Studies have also shown that higher hypnotizability predicts better outcome. While there is no evidence that treatments using hypnosis are more effective than other interventions for smoking cessation, they may be more efficient because patients can use a self-administered treatment strategy (self-hypnosis) to reinforce a more-adaptive cognitive restructure, while providing them with an exercise on physical relaxation.

Weight Control Typical of the generally modest results seen in weight-control regimens, rarely is hypnosis alone considered the treatment of choice for weight problems. It is usually employed as an adjunct to a comprehensive dietary and exercise control program for weight reduction and management. Also, the concept of going on a diet leads to long-term failure despite short-term success, since few people are able to live indefinitely on a diet, and failure to stay on a prescribed plan becomes self-defeating. The cycle of losing, gaining, and losing weight can lead to medical complications, and periodic starvation actually further dysregulates response to hunger signals and therefore control of eating behavior.

Like the use of self-hypnosis in the control of smoking described above, the purpose is to restructure the patient's relationship to food. In trance, patients are taught to (1) develop a personal plan and focus on eating behavior, (2) clarify and differentiate eating from overeating for their particular body type and goals, and (3) learn to eat on the basis of respect and protection for their bodies.

The hypnotic modality is used to develop a new perspective, provide a way to initiate feelings of affirmation and control, and simultaneously avoid the traps of deprivation, failure, and self-deprecation. Patients are instructed to learn to eat with respect for their bodies, to focus on what they are for rather than what they are against. An important component of this approach consists of teaching the patient to use self-hypnosis training to control the urge to overeat. This is better accomplished by preparing a list of foods that constitutes eating with respect and then comparing an urge with the list. If the desired food is on the list, patients are encouraged to eat it as a gourmet would, focusing intently on all aspects of the eating experience and enjoying it. If the food is not on the list, patients are asked to recognize the desire, rather than fight the urge. Then patients are encouraged to use self-hypnosis to compare this urge with their overall commitment to protect and treat their body with respect and therefore to eat with respect. By using this method patients can see their desire to eat not as an occasion to feel deprived, but rather as one in which they are enhancing their mastery of the urge by choosing to protect their body.

This technique was examined in a randomized trial with 45 subjects. The addition of hypnosis to a behavioral intervention resulted in significantly greater weight loss at 3-month follow-up (6.4 kg vs. 1.3 kg, $H_c = 4.7$, $P < .05$). Furthermore, hypnotizability as measured by the Stanford Hypnotic Susceptibility Scale, form C, and weight loss were significantly correlated ($r = –.56$, $df = 13$, $P < .03$). A meta-analysis performed by Irving Kirsch on 18 studies in which adjunctive use of hypnosis was compared with other psychotherapeutic treatments demonstrated that hypnosis substantially improved treatment outcome. This hypnotic supplementation effect was strongest for the treatment of obesity. However, the magnitude of this effect was questioned by D.B. Allison and M.S. Faith because of computational inaccuracies and methodological problems in one study. Clinical experience suggests that those within 20 percent of their ideal body weight may obtain some benefit from such restructuring techniques with self-hypnosis, when combined with a regimen of a balanced diet and exercise.

Enhancing Medical Care In medicine, hypnosis is becoming a more accepted modality to help patients with symptom control (pain, nausea and vomiting, itching, edema), anxiety secondary to illness and procedures, altering the course of physiological responses, and reversing the helpless and hopeless feelings that frequently occur during and after hospitalization. Issues of transference that frequently need to be worked through in long-term psychotherapy can create a powerful sense of support for the patient in brief medical psychotherapy, which adds to the patient's feelings of self-worth and, in turn, can support independent action. The use of the hypnotic modality and the relationship that ensues between the therapist and the patient, can be used to identify individual capacity for increased self-care and control over responses to illness and symptoms. The hypnotic procedure can become a bridge to transmit caring, comfort, and respect; maximize patient participation, and enhance feelings of independence and mastery over somatic responses to anxiety, fear, and anger. Such interventions may facilitate medical recovery, since patients who have the most stable blood pressure and respiratory function along with minimal bleeding and a perception of pain control recover more quickly and at less risk than those who are unstable in these areas.

Psychological interventions involving hypnosis have been successful in alleviating physical symptoms due to asthma, allergies, dermatological problems, blood pressure instability, bleeding, burns, wound debridement, warts, gastrointestinal disturbances, impotence, contractures of the hand, migraine and tension headaches, enuresis, and obstetrics in both adult and pediatric populations. Hypnosis is useful in rehabilitation after accidents and stroke as well as with psychogenic symptoms. In diabetes, cancer, heart disease, renal failure, orthopaedics, and somatoform disorders, the use of hypnotic interventions has been associated with better symptom management, decreased depression, less pain, and improved psychophysiological function. Approaches vary according to the presenting symptoms, problem formulation, degree of hypnotizability, motivation, an individual's belief system, mutually acceptable therapeutic goals, and the type and stage of medical problem.

The various techniques of self-hypnosis become an adjunct to other, more primary psychotherapeutic interventions. In support group settings, where the primary emphasis is on supportive/expressive group psychotherapy, self-hypnosis has been helpful in decreasing pain that amplified depression. Approaches for habit control,

pain management, phobias, and some manifestations of anxiety tend to be rapid and direct, with strategies taught to be under the control of the patients who initiate their own entry into trance states and review the therapeutic strategies on their own, outside the therapeutic dyad.

Relevant strategies are as crucial in dealing with the psychological sequelae of medical illness as they are in psychotherapy. The phase of illness, the nature of the problem, and the patient's individual resources are important factors in selecting the treatment strategy. If patients have the mental and physical capacity to initiate and use their own mental, imaginative processes, hypnotic ceremonies involving long induction procedures and techniques that emphasize the direction of the therapist can have a counterproductive effect by creating yet another situation of dependent and passive behavior. Yet if the patient is in the terminal stages of illness that require considerable mental and physical exertion, direct suggestions that do not involve patient initiation can be useful.

A hospitalized 56-year-old woman with metastatic breast cancer was anhedonic, refusing food, and lying in a fetal position and had revealed suicidal ideation to the nurse in charge of her care. In the initial interview, her response to the Hypnotic Induction Profile identified her as highly hypnotizable (induction score of 16). She was taught to enter a hypnotic state and was given the following instruction: “Imagine you are safely surrounded in every direction by luminous serenity and a protective calmness. You can be safely breathing and comfortably breathing in this energy as though surrounded by your own private buffer zone. Inside this buffer zone, you are free to use your memory in a selective manner. You can choose those people, places and circumstances that give you a profound sense of satisfaction and accomplishment, while letting everything else fade away. When you choose to come out of the trance state, give yourself a posthypnotic suggestion to maintain your protective buffer zone and allow calmness and serenity to stay within you. When you are out of the formal trance state you can continue to feel calm and serene.” Within hours of learning this technique and sharing her experiences with the therapist, she was sitting up in bed and enthusiastically requesting food. Follow-up sessions reviewed the self-hypnosis strategies and reinforced her feelings of control, vitality, and sense of meaning. Her improvements in mood continued, and she became an active participant in her own care.

A hospitalized 61-year-old woman with metastatic breast cancer (previously assessed as moderately hypnotizable—induction score of 11) had become almost totally immobile, was suffering with severe edema, and was struggling to breathe with the help of an oxygen mask. With little ability to focus her own attention, a simple and directive approach was used. The therapist used her voice as a metronome to influence the breath rhythm of the patient. The words “calm” and “easy,” timed with the patient's exhalations, were used as direct suggestions to enable the patient to breathe without gasping for air. Suggestions were then given: “Allow the bed to work with you as your partner in care. It can hold you safely and securely from underneath while your mind is free to travel through time and space in whatever direction increases your sense of well-being. You can feel the love of your family and friends all around you, giving you additional support, comfort, and strength.” Staff and family members were taught to repeat these suggestions to the patient. The benefits were twofold: the caregivers felt they had something to contribute, and the patient was comforted in the dying process.

Hypnosis has proved helpful in the consultation setting for problems including needle phobias, appetite loss, treatment refusal, respiratory problems, acute and chronic pain, preparation for surgery and invasive diagnostic procedures, and nausea and vomiting related to chemotherapy. In other specialties, interventions with hypnosis helped resolve phantom limb pain, fear of anesthesia, elevated heart rate and blood pressure, and somatoform disorders. In one well-described case by Joseph Dane, with 8-year follow-up, a patient with multiple sclerosis was treated using hypnosis and experienced return to ambulatory status within 2 weeks after beginning the use of hypnosis. She had reduced pain, balance problems, and diplopia, and her improvement persisted.

Surgical Preparation Three approaches selected on the basis of the patient's hypnotizability are useful in using hypnosis to help patients prepare for surgery. Those who test low in hypnotizability prefer to focus predominantly on the specific physiological benefits of less bodily tension and concrete outcome, those who test in the midrange of hypnotizability require more attention to the emotional component of their response to the up-coming procedure, and those who test at the high end of hypnotizability do best with minimal details and firm, clear, simple directions on how to conduct themselves.

Most patients accept that learning to minimize fear, anxiety, and anger reduces their state of physiological arousal. The general principles are to teach patients to enter their own trance state and focus on (1) allowing and welcoming the skilled medical team to do what they know best; (2) choosing images, metaphors, and body language (floating, muscles soft, thoughts of being someplace safe and secure); (3) suggestions on how they want their body to react before, during, and after surgery; and (4) the benefits of the treatment for psychophysiological functioning. In some cases, it is desirable to decrease or eliminate anesthesia. Suggestions are then geared to images and instructions to produce imagined numbness.

Side Effects of Chemotherapy The use of hypnosis to alleviate nausea and vomiting associated with chemotherapy has been reported in the literature for more than 20 years. There are three major facets to target:

1. The anxiety component—patients are taught to enter a hypnotic state and use their imagination to produce generalized relaxation
2. Conditioned responses to hospital cues—patients are to enter hypnosis and entertain suggestions to create comfortable responses to specific environmental cues
3. Physiological responses to the chemotherapy agent—patients are taught to enter hypnosis and focus on suggestions to reverse the usual peristaltic activity that leads to vomiting; it is useful to reframe the chemotherapeutic substance as the person's agent of care and to create a therapeutic relationship that emphasizes self-mastery

The therapist's manner should encourage patients to define themselves as collaborative participants in treatment with a focus on a commitment to living rather than on sickness and death.

The highly hypnotizable patient can frequently control nausea and vomiting by hallucinating the taste of orange or mint and dissociating from negative environmental cues; the midrange patient will benefit from guidance to contain emotions (e.g., the screen technique); and the low hypnotizable patient will benefit from distraction, frequently needing help selecting memories or imaginative projections with sufficient interest to encourage absorption. Patients in each of the three categories need additional help to manage negative cues from friends, family, and hospital staff as well as guidance in creating a secure support system at home, at work, and in their place of treatment. A growing body of data supports the clinical observations of the authors that hypnotic interventions reduce nausea and vomiting.

Pain Although one of the best-established therapeutic effects of hypnosis is analgesia, this technique is underused for analgesia. Since the mid-nineteenth century hypnosis has been known to be effective in controlling even severe surgical pain. Hypnosis seems to work through three primary mechanisms: muscle relaxation, perceptual alteration, and cognitive distraction. Pain is often accompanied by reactive muscle tension. Patients frequently splint the part of their body that hurts. Yet because muscle tension can by itself cause pain in normal tissue and because traction on a painful part of the body can generate more pain, techniques that induce greater physical relaxation can reduce pain in the body. Thus, the first step after patients enter a state of hypnosis is to have them concentrate on an image that connotes physical relaxation such as floating or lightness. This often produces physical relaxation and reduces pain.

The second major component of hypnotic analgesia is perceptual alteration. Patients can be taught to imagine that the affected body part is numb. This is especially useful for extremely hypnotizable individuals who can, for example, relive an experience of dental anesthesia and reproduce the drug-induced sensations of numbness in their cheek, which they can then transfer to the painful part of their body. Some can also simply switch off perception of the pain with surprising effectiveness. More commonly, subjects can substitute a different sensation for the painful one. Temperature metaphors are often especially useful, which is not surprising given the fact that pain and temperature sensations are part of the same sensory system, the lateral spinothalamic tract. Thus imagining that an affected body part is cooler or warmer using an image of dipping it in ice water or heating it in the sun often helps patients transform pain signals. Some patients prefer to imagine that the pain is a substance with dimensions that can be moved or can flow out of the body as if it were a viscous liquid. Others like to imagine that they can step outside their body, for example, to visit another room in the house. The use of hypnosis to enhance involvement in imagery that reduces the pain and related anxiety provides direct and indirect benefits to many pain patients and those with other medically related difficulties.

Less-hypnotizable individuals often do better with distraction techniques that help them focus on competing sensations in another part of the body. Painful stimuli tend to attract, indeed coerce, attention to them. It is possible to use hypnosis to help even low hypnotizable subjects to acknowledge the pain but maintain focus on sensations in other parts of their bodies, often coupled with various anxiety-control techniques.

A 34-year-old patient with Hodgkin's disease had been put on meperidine (Demerol) to control abdominal pain. The patient was referred for self-hypnosis to manage the pain to maintain a clear sensorium. After the hypnotic induction, the patient was given the following hypnotic suggestions: “As you imagine yourself floating, make it more vivid by imagining you are floating in an icy stream or lake. Make it so icy, that as you imagine the water getting colder and colder, you feel an imaginary tingling numbness coming from the cold water into your body. This tingling numbness serves as a protective coating around your abdomen, so you learn to filter the hurt out of the pain. Practice this exercise every 1 to 2 hours—and each time give yourself a posthypnotic signal to retain this feeling of numbness, even when you are out of the formal trance state. As you retain this imaginary numbness, you can begin to extend it so it stays with you around the clock. By constantly making this commitment to impose numbness, even though you know the pain is there, you feel the numbness more than the pain.” This patient developed an ability to function without Demerol.

One of the primary goals is to teach the patient to stop fighting the pain, decreasing psychophysiological tension and discomfort while simultaneously increasing mastery and control with imagined tingling numbness. Thus patients learn to modulate the pain perception and the attention they pay to it by filtering the hurt out of the pain. The heightened awareness of the trance state provides the patient with a powerful tool to respond to signals that have been coming from the body upward to the brain and, through training, practice, and personal choice, influence signals from the brain downward to the body. In trance, imagined activity is heightened, maximizing selective mind-body effects.

Before introducing an induction technique and therapeutic strategy, the clinician first makes the distinction between acute and chronic pain. In the acute situation of trauma, injury, or physical insult, the person may have already entered a spontaneous trance state that maximizes hypnotic responsiveness. Motivation to receive help and comfort creates a powerful transference. In the acute situation, direct suggestions tend to produce immediate results, even with low hypnotizable persons who in other circumstances prefer explanations and collegial discussions. In an acute medical emergency, the most immediate need of the patient is to turn control over to an expert in caring for the problem at hand.

A 34-year-old woman was referred for hypnosis when she was hospitalized for the delivery of her fifth child. She was reliving memories of four extremely painful deliveries and had expressed a desire to kill herself rather than try and live through a fifth delivery. Using the eye roll sign as an emergency indicator, she was identified as potentially highly hypnotizable. She rapidly went into trance as evidenced by stabilization of her blood pressure, heart rate, and respiratory function. Within minutes, she accepted the following suggestions: “Imagine yourself in the delivery room, letting yourself float outside your body, able to watch the delivery. By being outside your body, you have the freedom to come and go as you please, but you always maintain the right to pay attention to the delivery of your child. You have the freedom to participate in every way you wish, but there is no need to feel any of the physical sensations that would cause discomfort. You can respond to verbal cues from the medical team to help you know when to push and when to rest. You do your part, the team does their part, and you can experience the joy of the arrival of your child without any unnecessary discomfort.” The patient readily accepted these instructions and was able to give birth to a healthy baby girl without any further undue stress or complications.

For chronic pain, assessment includes the impact on lifestyle, relationships, ability to maintain previous levels of activity, issues of secondary gain-loss, and prognosis. The success of the hypnotic intervention correlates with the degree of patient motivation, the ability of the therapist to formulate an acceptable rationale, and how well the therapeutic strategy matches the personality style and hypnotic capacity.

As with any pain treatment technique, including analgesic medications, hypnosis is more effective when used early in the pain cycle, before the central pain response component has developed and the pain has become so overwhelming it impairs concentration. Patients should be encouraged to use this technique early and often because it is simple and effective and has no adverse effects. A randomized trial comparing group therapy plus hypnosis with routine medical care demonstrated that metastatic breast cancer patients taught hypnosis had half the pain experienced by the control group at 1-year follow-up. In a randomized trial in interventional radiology, hypnosis was shown by Elvira Lang and colleagues to produce better analgesia than patient-controlled analgesia with midazolam (Versed) and fentanyl (Duragesic), producing less anxiety, fewer adverse effects, less cardiovascular instability, and fewer procedural interruptions.

Hypnosis is especially effective in comforting children who are in pain. Several good studies have shown greater efficacy than placebo attention control in randomized trials. This is likely because children as a group are more hypnotizable than adults. Their imaginative capacities are so intense that separate relaxation exercises are usually unnecessary. Children naturally relax when they mobilize their imagination during the sensory alteration component of hypnotic analgesia.

Recent research indicates cortical effects of hypnotic analgesia exercises, which include reduced ERP amplitude in response to somatosensory stimuli and reduced activity in the anterior cingulate gyrus. Thus hypnotic alteration of nociception seems to involve cortical modulation of pain perception. In other words, the strain in pain lies mainly in the brain.

Anxiety Disorders Anxiety disorders are among the most widely prevalent psychiatric disturbances, affecting approximately 15 percent of the population. Anxiety can be understood as a state of hyperarousal experienced as both emotional and psychosomatic discomfort. Patients often describe their anxiety in physical terms, such as palpitations, GI discomfort, chest pain, sweating, and motor restlessness. Hypnosis can be a helpful adjunctive tool for treating anxiety disorders because of its ability to help patients control their physical reaction to anxiety-provoking stimuli, thereby dissociating somatic response from psychological distress. This enables them to attend to the stimuli long enough to restructure their point of view and achieve a sense of mastery.

Most strategies for the treatment of anxiety disorders that use hypnosis combine instructed physical relaxation with a restructuring of cognition, using imagery coupled with physical relaxation. As in the treatment of anxiety disorders by systematic desensitization or progressive muscle relaxation, patients are instructed to maintain a physical sense of relaxation (e.g., using an image of floating) while at the same time picturing the anxiety-producing stimulus. The relaxation instruction must use an image that connotes reduced somatic tension, such as “floating” or “lightness,” rather than a direct instruction to “relax.” This latter, more-cognitive term may actually induce more anxiety, arousing concerns about adequate performance, while by contrast, affiliation with a somatic metaphor usually produces some reduction in tension. Unlike systematic desensitization, hypnosis can be used to produce a physically relaxed state quite rapidly without going through a series of graded muscle relaxation and imagery steps. Also unlike systematic desensitization, the coupling of relaxation to a fearful stimulus does not require the development or working through of a hierarchy.

A typical self-hypnosis induction can be quite rapid. For example, a patient can be told: Now just get as comfortable as you can. There are many ways to enter a state of self-hypnosis. One simple but useful method is to count to yourself from one to three. On one, do one thing: look up. On two, do two things: slowly close your eyes and take a deep breath. On three, do three things: let your eyes relax but keep them closed, let your breath out, and let your body float. Then let one hand or the other float up into the air like a buoyant balloon. This is your signal to yourself and to me that you are ready to concentrate.

Initially, the use of hypnosis in the session can help demonstrate to patients that they have more control over somatic responsiveness than they had imagined. It is often useful to begin by teaching patients to create a place in their “mind's eye” where they feel safe and secure. On occasion, it helps subjects to learn how to project their image onto an imaginary screen. Later they can learn to manipulate the screen by either making it bigger or smaller, having the screen being nearer or farther away, as needed.

> Just allow your body to float, as if you were floating in a bath, a lake, or a hot tub. Enjoy this sense of floating lightness. Now, picture in your mind's eye an imaginary screen. It might be a movie screen, a television screen, or a piece of clear blue sky. First picture a pleasant scene, somewhere you enjoy being.
>
> Allow the patient to experience this state for a minute or two, then inquire about the experience:
>
> With your eyes closed and remaining in this state of concentration, describe how your body is feeling right now. What image are you picturing?
>
> After receiving the answers, add:
>
> Notice how you can use your store of memories and fantasies to help yourself and your body feel better.

One method to help patients shift gears and restructure their relationship to thoughts and feelings is the screen technique. The patient is told: “You can enter the trance state by permitting your body to float. By permitting your body to float, your mind is free to concentrate on an imaginary screen in front of you. While in trance you have the power to teach your body to maintain the physical sensation of floating and, at the same time, use the imaginary screen to become the director, script writer, actor, organizer, and balancer of your own thoughts and feelings. First, let the central part of the screen serve as a receiving screen for all your thoughts and

feelings. Now imagine a split screen on the right for your creative thoughts and a split screen on the left (the sinister side) for your worry thoughts. Place your thoughts and feelings of anxiety, fear, and worry on the left; place your creative thoughts, feelings, and visions of refuge on the right. After all, you are more than worries and fears. You are now learning to use your mind as a psychological zoom lens, with the right to revisit any worries and fears, if you wish—you don't have to fight them. By putting your thoughts and feelings on an imaginary screen and learning to let your body float, your body is no longer a battleground for your emotions."

For insomnia, the initial concept is the screen technique (see above) with an additional emphasis on switching from sympathetic to parasympathetic function. The patient is told: "Muscle tension is an enemy of sleep. The autonomic nervous system has a day and a night shift. Muscle tension is a barrier that impairs the transition. By feeling a sense of floating while projecting your thoughts onto the screen, tension dissolves, allowing the day shift to yield to the night shift to let sleep come like a welcome friend."

Patients may also use the trance state as a means of facing their concerns more directly. They can be taught to use the screen technique on their own. They can achieve this by placing an image of an upcoming performance or fearful situation on one side of the screen; on the other side, testing out various strategies for mastering the situation. The screen becomes a brainstorming and problem-solving technique that allows them to face anxiety-provoking situations more directly, but seeing them from a different perspective. This helps them move from a passive to an active coping mode, not merely avoiding but responding to the feared stimulus, by acknowledging and working out responses to it while managing somatic reaction to the anxiety.

Other approaches using hypnosis have included instructing patients in a trance to imagine that they are literally somewhere else, away from the fearful stimulus, thus separating themselves from the anxiety-producing experience. Positive reinforcement, or ego-strengthening techniques, have also been used; for example, giving hypnotic instructions to patients suggesting that their capacity to master the situation and their response to it will improve. There is little reason to use uncovering techniques seeking to link the complaint of anxiety to some early traumatic experience in cases of phobia or generalized anxiety disorders. This is different though in cases of posttraumatic stress disorder, where more work may be needed to confront the traumatic experience.

Treatment for performance anxiety, generalized anxiety, and insomnia follows the restructuring approach. Instead of fighting against fear, conflicts, stressful events, or problems, patients focus on physical sensations and learn to distance themselves from the issues, to develop a sense of freedom, clarity of thought, and relief.

Phobias The underlying challenge of a phobic response (irrational fear) is to clarify the confusion between a possibility and a probability. Distinctions are made between anticipatory anxiety for likely and unlikely events, with additional differentiation between useful, protective fear and unnecessary, maladaptive fear. With these clarifications, the therapeutic strategy is to teach the patient to focus on what one is for rather than the fear of what one is against. This provides the momentum for rapid change.

For flying phobia, a prevalent clinical condition, the patient is asked to consider the difference in safety probabilities between an amateur pilot and a professionally licensed pilot on a commercial plane. Anticipatory anxiety is described as a healthy response to mobilize one's resources for a new phase of action. Patients are taught that if they interpret anticipatory anxiety as panic they are perverting a natural resource. Patients are then taught to focus on the many instruments that have been developed to act as an extension of the body. "For example, a hammer is an extension of your hand; a bicycle is an extension of your legs; an airplane is an extension of your arms. When you choose to fly on a scheduled airplane, in one direct move, you can accept the pilot as your agent, and through this choice you are rationally controlling the plane. By choosing to allocate the technology to someone with the proper skills and training, the control is still yours." Patients are taught a self-hypnosis exercise in which they concentrate on floating with the plane, using the pilot as their agent of control.

OUTCOME A number of studies demonstrate the utility of hypnosis in the treatment of anxiety and phobias. A study by the first author and colleagues with 178 patients treated with a single session of self-hypnosis for flying phobia indicated that 52 percent were either improved or cured at 7-year follow-up. There has been considerable interest in hypnosis as an adjunct to dental procedures and in the treatment of dental phobia. The success of hypnotic techniques has created a new interest in teaching the usefulness of hypnotically assisted dental work in dental schools across the country. Similarly, hypnotic techniques have been successfully used to assist phobic patients undergoing a number of medico/surgical and diagnostic procedures, thus diminishing the need for excessive anesthesia or antianxiety medication, improving compliance, and eliminating trauma to patients. Hypnosis may be especially helpful as an adjunctive tool for treating these anxiety disorders because of the ability of the hypnotized person to control somatic response.

Posttraumatic Stress Disorder The most distressing thing about a traumatic event is the sense of absolute helplessness that it engenders. This helplessness is reenacted in both acute and posttraumatic stress disorder through loss of control over the state of mind in the aftermath of trauma, with spontaneous dissociative states, startle reactions, or intrusive recollections of the event. Having regained control of their bodies, many trauma victims feel they have lost control over their minds as they are subjected to nightmares, flashbacks, and intrusive memories. The therapist can be contaminated with association to the trauma. Such patients may tend to identify the therapist with the assailant and feel that the therapy amounts to reinflicting the trauma. The process of the therapy, especially when a technique such as hypnosis is used, must be structured so that it enhances patients' sense of control. This approach can allow patients to integrate the image of themselves as victims with the ongoing, more-global image of themselves as persons coping effectively with severe stress, making the repressed material conscious and therefore less powerful, and enabling them to establish a new, more congruent self-image, absorbing the loss into the ongoing flow of their lives. The psychotherapy of posttraumatic stress disorder contains elements of desensitization, in which reexposure to the traumatic stimulus may gradually deprive it of some of its emotional power. Indeed, the use of hypnosis in the psychotherapy of trauma was initially thought to be limited to abreaction, based on Freud's cathartic method. The idea was that some intense affect associated with the traumatic event needed to be released and that simple repetition of the memory of the event with its associated emotion in the trance state would discharge the energy producing the symptoms. However, it became clear to Freud that conscious, cognitive work must be done on the material for it to be successfully worked through. Indeed, it is now clear that cognitive restructuring of the meaning of the traumatic event coupled with a continued sense of being cared about by the therapist enhances the effect of hypnosis in psychotherapy.

Treatment using hypnosis involves not merely abreaction of trauma, but working through it by assisting with the management of uncomfortable affect, enhancing patients' control over it, and enabling them to cognitively restructure its meaning. Catharsis is a beginning but not an end in itself, and it can lead to retraumatization if not accompanied by support in managing affective response, control over the accessing of traumatic memories, and help in working them through. F. Lindemann's grief work model is useful. Observations of normal grief after trauma have led to recognition that a certain amount of emotional discomfort and physical restlessness and hyperarousal is a natural and indeed necessary part of acknowledging, bearing, and putting traumatic memories into perspective. This is often facilitated by using a hypnotic imagery technique, the split screen, in which patients are asked to picture some aspect of the trauma on one side of the screen, bearing the associated uncomfortable affect, and then picture on the other side of the screen something they did for self-protection or to aid others. In this way the traumatic memory is acknowledged but restructured to encompass efforts at mastery as well as the experience of helplessness.

Hypnosis can be used to provide controlled access to the dissociated or repressed memories of the traumatic experience and then help patients restructure their memories of the events. The unusual characteristics of the hypnotic state provide reassurance that the distress associated with the traumatic memories can to some extent be put aside when the hypnotic state is ended. Also, the dissociation typical of hypnosis can be used to separate psychological from somatic distress. Patients can then find a condensation image that symbolizes some aspect of the trauma. It is often helpful to have them do this on an imaginary screen, giving them some sense of distance from the event. It is also useful to divide the screen in half, having the patient picture on one side some aspect of the event (e.g., a rape victim's image of the assailant) and on the other side of the screen, some self-protective action (e.g., struggling with the assailant, talking with him, running away). This lets the patients restructure their view of the assault, facing it, but not simply in the familiar terms of the humiliation, pain, and fear with which it was initially associated. Victims can better acknowledge their helplessness when they also recognize their efforts to protect themselves. Bereaved individuals can picture themselves at the graveside on one side of the screen and at an earlier moment of joy with the deceased on the other side. They can then be taught to practice a self-hypnosis exercise in which they grieve and work through traumatic memories while enhancing their sense of control over the process.

PRINCIPLES OF PSYCHOTHERAPY WITH HYPNOSIS The principles of psychotherapy for acute and posttraumatic stress disorder using hypnosis can be summarized with the following eight Cs:

Confrontation. The patient should confront the traumatic events directly rather than attribute the symptoms to some long-standing family or personality problem.

Confession. The therapist should help trauma survivors discuss deeds or emotions that are embarrassing to them and at times repugnant to the therapist. It is important to help these patients distinguish between inappropriate guilt and real remorse. Such patients may well be telling the therapist about aspects of the traumatic event that they have discussed with no one else.

Consolation. The intensity of traumatic experiences requires an actively consoling approach from therapists, lest they be perceived as being judgmental or as inflicting

rather than treating trauma-induced emotional pain. Appropriate expressions of sympathy and concern can be helpful in acknowledging and diffusing this common traumatic transference reaction.

Condensation. An image that condenses a crucial aspect of the traumatic experience should be found. This representation can make the overwhelming aspects of the trauma more manageable by putting it in concrete, symbolic form. Furthermore, it can facilitate restructuring of the experience by joining previously disparate images; for example, linking the pain associated with the death of a friend in combat with the happiness experienced during some earlier shared time. This allows patients to alter the pain of the loss by attending to positive aspects of the lost relationship that remain in memory as well as negative ones.

Consciousness. Previously dissociated traumatic memories should be made conscious in a gradual manner that does not overwhelm the patient.

Concentration. The intense, focused concentration characteristic of the hypnotic state should be used to reinforce the boundaries of the traumatic experience and the painful affect associated with it. Directing sharply defined attention toward the loss also implies that when the hypnotic state is ended, attention can be shifted away from the traumatic experience. This counters the common fear of “opening Pandora's box,” and unleashing feelings not easily contained.

Control. Because the most painful aspect of severe trauma is the absolute sense of helplessness, the loss of control over one's body and the course of events, the process by which the therapeutic intervention is conducted must enhance the patient's sense of control over the traumatic memories. Structure the experience so that patients are given the opportunity to stop working through when they feel they have had enough, can remember as much from the hypnosis as they care to, and feel they are in charge of the self-hypnosis experience. They should learn to use it on their own as a self-hypnosis exercise as well as with the therapist. Such procedures help patients imbue traumatic memories with a greater sense of control and mastery.

Congruence. The goal is to help patients integrate dissociated or repressed traumatic material into conscious awareness in such a way that they can tolerate experiencing the memories as part of their continuous life history. In this way the traumatic past is not incompatible with their present xperience. Patients should emerge from therapy having reviewed not only what was done to them but what they did to protect themselves, not only what they lost, but what they had valued and why.

OUTCOME There are numerous clinical reports of the effectiveness of hypnosis as an adjunct to treatment of posttraumatic stress disorder but few controlled studies of the efficacy of hypnosis in treating this disorder. However, one controlled study by D. Brom and colleagues of 112 subjects with posttraumatic stress disorder demonstrated that psychotherapy with hypnosis was superior to a control condition and was equivalent to both psychodynamic therapy and systematic desensitization. Hypnosis was particularly effective in reducing intrusion rather than avoidance symptoms.

Dissociative Disorders Dissociation is best understood as part of a continuum of phenomena involving ongoing fragmentation and coalescence of conscious and unconscious associations. The ongoing process of dissociation and reassociation leads to a variety of patterned integrations of emotional and cognitive material that previously existed as fragments of information.

Dissociation of identity (dissociative identity disorder), memory (dissociative amnesia and fugue), or consciousness (depersonalization disorder, dissociative trance disorder) result in an array of symptoms that affect intrapsychic and interpersonal functioning. A central feature of the dissociative disorders is a loss of conscious access to some parts of experience, even though the individual always maintains some connection to a main psychological orientation and an underlying intentionality.

An important development in the modern understanding of dissociative disorders is the establishment of a clear link between trauma and dissociation. While the role of traumatic stress in eliciting dissociative symptoms was a part of Janet's early thinking as well as Freud's, recent work has examined in more detail the proximate role of trauma in eliciting dissociative symptoms. However, not enough critical data has been studied to establish a causative relation among trauma, its timing, its intensity, and its meaning in the emergence of subsequent dissociative disorders. Secondary gain and malingering are critical factors that cannot be ignored in the genesis and maintenance of these syndromes. Often the Axis II diagnosis provides the most significant elements of the clinical picture.

It is important to note the differential diagnosis between pathological dissociation, which is the primary result of an interaction between stressful events and high hypnotizability, and psychosis which is the result of a internal cognitive break, which is associated with low or no hypnotizability. Edward Frischholz and colleagues have found that patients with dissociative disorders have a mean eye roll sign of 3.38 (on a 0 to 4 scale) and have high scores on the Hypnotic Induction Profile, while patients with schizophrenia have a mean eye roll sign of 1.73 and have low scores on the Hypnotic Induction Profile. The difference in both hypnotic capacity and responsivity between these two diagnoses makes the use of formal hypnosis (including testing) a primary aspect in the diagnosis and treatment of dissociative disorders but not in the treatment of schizophrenia. In dissociative disorders, whether the dissociations are spontaneous or conscious, they are always potentially reversible.

Symptomatology and Treatment Dissociative disorders involve a failure in the customary integration of identity, memory, or consciousness. Loss of access to parts of experience to consciousness can intensify many preexisting problems. Formal hypnotic inductions with appropriate treatment strategies are used to teach an individual to access dissociated feelings and memories and to control dissociative process. Conceptualizing dissociative identity disorder as a chronic posttraumatic stress disorder, the psychotherapeutic strategy also involves working through traumatic memories.

To the extent that psychosocial stress triggers dissociation, resolution of that stress can help resolve and reduce the likelihood of further symptoms. Highly hypnotizable individuals who have difficulty asserting their own point of view in difficult interpersonal relationships or who may be dealing with severe situational stressors are prone to extreme dissociative symptoms. Psychotherapy can assist such individuals to recognize and modify their tendency to uncritical compliance with others and their extreme sensitivity to rejection and disapproval.

Memory Retrieval Since memory loss in dissociative disorders is often complex, its retrieval is an integral part of the psychotherapeutic process. In general, recall of recent trauma is more feasible and more accurate than recall of remote trauma. The therapeutic strategy is designed to use memory recall to counter the dissociative fragmentation and to work toward integration.

Patients with dissociative disorders can be helped with a psychotherapeutic approach that facilitates conscious control of dissociated memories and behavior previously experienced as automatic and unwilled. Clinicians must stay alert to the extreme malleability and the uncritical acceptance of new information by the highly hypnotizable person. Excessive exploration may further increase disintegration, leading to confusion or retraumatization. The goal is to garner only the most relevant data that will enable the patient to establish more effective control for present and future living.

Outcome While there are no controlled trials of the efficacy of hypnosis in treating dissociative disorders, the clinical literature indicates that it is highly useful. A recent survey of 305 clinicians indicated that individual psychotherapy facilitated by hypnosis twice a week was the primary treatment for dissociative identity disorder patients, with anxiolytics and antidepressants being used as secondary adjunctive tools.

HYPNOSIS IN FORENSIC PSYCHIATRY

Recent controversies surrounding the so-called false memory syndrome have aroused concern about the use of hypnosis in the clinical setting and have raised questions about the validity of the material recovered by the use of hypnosis. Hypnosis had been used not infrequently to refresh recollection of witnesses and victims of crimes. Despite concern that hypnosis inevitably contaminates memory, the research literature on hypnosis and memory indicates that hypnosis affects belief more than content. Any memory retrieval technique that increases the production of memories may lower the response criterion (i.e., the willingness of a subject to report a thought as a memory). Thus the mere act of trying harder to remember something about an event can always convert thoughts, fantasies, or leading questions into memories. Any special effects of hypnosis may be more feared because they have been more thoroughly examined. Despite these concerns, there are examples of the use of hypnosis in the forensic setting which led to correct new recall. A well-known example is the case involving the driver of a hijacked school bus in Chowchilla, California. Under hypnosis, the bus driver was able to recall the license plate on the car driven by the kidnappers. This information, not available to consciousness prior to hypnotic intervention, led to the arrest and conviction of the criminals.

No memory, whether retrieved with or without the use of hypnosis, can be determined to be true without independent confirmation. However, the courts have become especially concerned about possible contamination of memory by hypnosis. As a result, judges are increasingly unwilling to admit the testimony of a person hypnotized while testifying and recently have also begun to exclude testimony of witnesses who have previously been hypnotized about the event in question. This

dilemma is confounded by the fact that motivated or emotionally charged persons, particularly those with midrange to high hypnotizability, can shift into a spontaneous trance state while trying to recall an event without a formal hypnotic induction. These concerns are based on fears of confabulation and “concreting.” In some circumstances, hypnotized individuals can make up information and believe their newly created story is real. In the phenomenon of concreting subjects emerge from hypnosis with an enhanced conviction that their memories are correct. This makes witnesses appear more convincing to a jury than they really are about the facts. Indeed several states (e.g., New Jersey and New York) restrict the testimony of victims or witnesses who have used hypnosis to refresh their recollection. The reason for the courts' objection to the use of hypnosis is a combination of real and exaggerated dangers of hypnosis. After much legal battle, some courts now allow witnesses to testify after the use of hypnosis provided certain guidelines are followed that relate primarily to the training and independence of the professional doing the hypnotic interrogation and the electronic recording of the entire process.

Therapists treating trauma or crime victims must be aware that the use of hypnosis may compromise a witness's ability to testify in court. If civil or criminal legal issues are pending, it is advisable to ask the patient to discuss the situation with their personal attorney or the district attorney and obtain written consent for the use of hypnosis from both the attorney and the patient. If hypnosis is to be used and there will be subsequent legal involvement of the patient, all contact with the patient should be electronically recorded, preferably on videotape. Because hypnosis involves a suspension of critical judgment and thus a state of heightened suggestibility or responsivity to social cues, the interview must be conducted with a minimum of inserted information. To avoid the risk of contaminating the subject's memories the authors recommend using open-ended questions such as “what happens next?” rather than “how did he sexually abuse you?”

Hypnosis should never be used to replace standard investigative work. Hypnosis is most likely to be worth the risk when traumatic amnesia exists for the events of a crime or when all other avenues of exploration have been exhausted.

Hypnosis is a state of mind and brain that involves intensity of focus, dissociation, and heightened responsiveness to social cues. It is a measurable trait not equally shared, and differences in hypnotizability reflect developmental stages, the presence or absence of psychopathology, and personality style. It is a highly useful adjunctive tool in the management of somatic symptoms such as pain, habit control, and anxiety disorders. It is also useful in the treatment of posttraumatic and dissociative disorders. Fears about its effects are generally due to misinformation, and given its favorable risk/benefit profile, hypnosis deserves more general application in psychiatry and the rest of medicine.

SUGGESTED CROSS-REFERENCES

The biology of memory is discussed in Section 3.4 and psychoanalysis in Section 6.1. Anxiety disorders are the subject of Chapter 15 and dissociative disorders are presented in Chapter 18. Chapter 25 discusses psychological factors affecting a medical condition. Evaluation of psychotherapy is discussed in Section 30.11.

SECTION REFERENCES

Allison DB, Faith MS: Hypnosis as an adjunct to cognitive-behavioral psychotherapy for obesity: A meta-analytic reappraisal. J Consult Clin Psychol *64:*513, 1996.

Barabasz M, Spiegel D: Hypnotizability and weight loss in obese subjects. Int J Eat Disord *8:*335, 1989.

Breuer J, Freud S: Studies in hysteria. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 2, J Strachey, editor. Hogarth Press, London, 1995.

Brom D, Kleber RJ, Defare PB: Brief psychotherapy for post-traumatic stress disorder. J Consult Clin Psychol *57:*607, 1989.

Butler L, Spiegel D: Trauma and memory. In *Review of Psychiatry*, vol 16, L Dickstein, M Riba, J Oldham, editors. *American Psychiatric Press Review of Psychiatry*. American Psychiatric Press, Washington, DC, 1997.

Butler LD, Duran REF, Jasiukaitis P: Hypnotizability and traumatic experience: A diathesis-stress model of dissociative symptomatology. Am J Psychiatry *153:*41, 1996.

Crawford HJ, Gruzelier JH: A midstream view of the neuropsychophysiology of hypnosis: Recent research and future directions. In *Contemporary Hypnosis Research*, E Fromm, M Nash, editors. Guilford, New York 1992.

*DeBetz B, Stern DB: Factor analysis and score distributions of the HIP: Replications by a second examiner. Am J Clin Hypn *22:*95, 1979.

*dePascalis V: Psychophysiological correlates of hypnosis and hypnotic susceptibility. Int J Clin Exp Hypn *47:*117, 1999.

Frischholtz EJ, Lipman LS, Braun BG, Sachs RG: Psychopathology, hypnotizability and dissociation. Am J Psychiatry *149:*1521, 1992.

Greenleaf M: Clinical implications of hypnotizability: Enhancing the care of medical patients and surgical patients. Psychiatr Med *10:*77, 1992.

Greenleaf M, Fisher S, Miaskowski C, DuHamel K: Hypnotizability and recovery from cardiac surgery. Am J Clin Hypn *35:*119, 1992.

Hargadon R, Bowers K, Woody E: Does counterpain imagery mediate hypnotic analgesia? J Abnorm Psychol *104:*508, 1995.

Hilgard ER, Hilgard JR, editors: *Hypnosis in the Relief of Pain*. Brunner/Mazel New York, 1994.

Holyroyd J: Hypnosis treatment of clinical pain: Understanding why hypnosis is useful. Int J Clin Exp Hypn *4:*417, 1996.

Jasiukaitis P, Nouriani B, Spiegel D: Left hemisphere superiority for event-related potential effects of hypnotic obstruction. Neuropsychologia *34:*661, 1996.

Kihlstrom JF: The cognitive unconscious. Science *237:*1445, 1987.

Kirsch I, Montgomery G, Sapirstein G: Hypnosis as an adjunct to cognitive-behavioral psychotherapy: A meta-analysis. J Consult Clin Psychol *63:*214, 1995.

Lang E, Joyce J, Spiegel D, Hamilton D, Lee K: Self-hypnotic relaxation during interventional radiological procedures: Effects on pain perception and intravenous drug use. Int J Clin Exp Hypn *44:*106, 1996.

*Loewenstein RJ: Psychogenic amnesia and psychogenic fugue: A comprehensive review. In *American Psychiatric Press Review of Psychiatry,* vol 10, A Tasman, SM Goldfinger, editors. American Psychiatric Press, Washington, DC, 1991.

McConkey KM, Sheehan PW: *Hypnosis, Memory, and Behavior in Criminal Investigation*. Guilford, New York, 1995.

Piccione C, Hilgard ER, Zimbardo PG: On the degree of stability of measured hypnotizability over a 25-year period. J Pers Soc Psychol *56:*289, 1989.

Posner MI, Petersen SE: The attention system of the human brain. Annu Rev Neurosci *13:*25, 1990.

Rainville P, Duncan GH, Price DD, Carrier B, Bushnell MC: Pain affect encoded in human anterior cingulate but not somatosensory cortex. Science *277:*968, 1997.

Sabourin ME, Cutcomb SD, Crawford HJ, Pribram K: EEG correlates of hypnotic susceptibility and hypnotic trance: Spectral analysis and coherence. Int J Psychophysiol *10:*125, 1990.

Schacter D, Reiman E, Curran T: Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. Neuron *17:*267, 1996.

*Scheflin A, Spiegel H, Spiegel D: Forensic uses of hypnosis. *Handbook of Forensic Psychology*, ed 2, AK Hess, IB Weiner, editors. Wiley, New York, 1999.

*Spiegel D: Healing words: Emotional expression and disease outcome. JAMA *281:*1328, 1999.

*Spiegel D: Hypnosis and implicit memory: Automatic processing of explicit content. Am J Clin Hypn *40:*231, 1998.

Spiegel D, Bierre P, Rootenberg J: Hypnotic alteration of somatosensory perception. Am J Psychiatry *146:*749, 1989.

Spiegel D, Cardena E: Disintegrated experience: The dissociative disorders revisited. J Abnorm Psychol *100:*366, 1991.

Spiegel D, Frischholz EJ, Fleiss JL, Spiegel H: Predictors of smoking abstinence following a single-session restructuring intervention with self-hypnosis. Am J Psychiatry *150:*1090, 1993.

Spiegel D, Hunt T, Dondershine HE: Dissociation and hypnotizability in posttraumatic stress disorder. Am J Psychiatry *145:*301, 1988.

Spiegel D, Moore R: Imagery and hypnosis in the treatment of cancer patients. Oncology *11:*1179, 1997.

Spiegel H: The grade 5 syndrome: The highly hypnotizable person. Int J Clin Exp Hypn *22:*303, 1974.

*Spiegel H: Silver linings in the clouds of war: A five decade retrospective. In *American Psychiatry: 1944–1994*, R Menninger, J Nemiah, editors. American Psychiatric Press, Washington, DC, 1999.

Spiegel H, Greenleaf M: Personality style and hypnotizability: The fix-flex continuum. Psychiatr Med *10:*13, 1992.

*Spiegel H, Spiegel D: *Trance and Treatment: Clinical Uses of Hypnosis*. American Psychiatric Press, Washington, DC, 1987.

Spiegel H, Spiegel D, Frischholz EJ, Maruffi BL: Hypnotic responsivity and the treatment of flying phobia. Am J Clin Hypn *23:*239, 1981.

Szechtman H, Woody E, Bowers KS, Nahmias C: Where the imaginal appears real: A positron emission tomography study of auditory hallucinations. Proc Nat Acad Sci USA *95:*1956, 1998.

30.4 GROUP PSYCHOTHERAPY, COMBINED INDIVIDUAL AND GROUP PSYCHOTHERAPY

ANNE ALONSO, PH.D.

People thrive best in a community that values their participation and protects their dignity. The stresses that impinge on individuals are defined in part by their biology, family dynamics, and culture. The interplay of these factors shapes how people live and work with others, and how they feel about themselves is also in part shaped by how others experience them and respond to them.

Today's culture has many fewer stable and dependable sources of community and fewer stable and long-lasting ties with institutions or with individuals; thus the individual is left without many of the historically dependable sources of identity. On these shifting sands, individuals must somehow achieve adequate individuation from their immediate family and learn how to manage intimacy with yesterday's stranger.

The integration of mind, body, and social context is vulnerable to assault from problems in any one of these dimensions. When problems occur, people suffer and seek our help for their suffering. Group psychotherapy offers the opportunity for purposefully created, closely observed, and skillfully guided interpersonal interaction, which can positively influence the countless varieties of human distress and malfunction. Distorted perceptions of others, insufficient communication, inadequately discharged affects, stereotyped behaviors, impulsive actions, and alienation can all be addressed and modified within the therapeutic group.

HISTORY

Early Beginnings Group psychotherapy originated in America in 1905 and was first conducted by an internist, Joseph H. Pratt, who led groups composed of tubercular patients. His groups bore remarkable similarity to those currently offered by cognitive-behaviorists and self-help groups to alleviate the distress caused by commonly shared physical or psychological problems. In 1919 L. Cody Marsh applied the group method of treatment to institutionalized mental patients. About the same time E. W. Lazell used group methods in treating psychiatric patients. In 1925 Trigant Burrow coined the term *group analysis* and conducted group therapy for noninstitutionalized patients. In the early 1930s Louis Wender introduced the notion of the group as a re-created family, applied psychoanalytic concepts to group therapy, and began the use of combined therapy. Paul Schilder, one of Wender's contemporaries, elaborated on Wender's approach and encouraged the use of free association.

Coming of Age Samuel Slavson, an engineer by profession, founded the American Group Psychotherapy Association in 1948 and is regarded as having had the most influence on the development of American group psychotherapy. He referred to his method as analytic group psychotherapy and concentrated on the individual in the group, rather than on the group itself. His greatest contribution is acknowledged to be the development of group psychotherapy with children. During the 1930s and 1940s Alexander Wolf actively applied the principles of psychoanalysis to groups of adults. In contrast to Slavson, Wolf believed that in analytic group psychotherapy, the psychoanalytic work was as deep as in the individual psychoanalytic setting and could even be more profound. It remained for current research to buttress Wolf's claims about the effectiveness of group psychotherapy.

European Influence Group psychotherapy also created interest in Europe. The Romanian psychiatrist Jacob L. Moreno (1898–1974), primarily identified with psychodrama borrowed from his experience with the Stegreiftheater (Theater of Spontaneity) in Vienna. Thus his introduction of group psychotherapy in 1910 included role-playing and role-training methods. Moreno may have coined the term *group therapy* in 1931. Although he never formally conducted group psychotherapy, Sigmund Freud called attention to group psychology in *Group Psychology and Analysis of the Ego,* published in 1921, Freud's insights into group formation and transference reactions form the underpinnings of modern psychoanalytically oriented group psychotherapy. Alfred Adler's major contribution to group psychotherapy was his explanation of the processes that account for the therapeutic influence of the group; in particular, he examined the relation of the group to the individual and the individual to the group. Kurt Lewin introduced his field theory concepts, emphasizing that the group differs from the simple sum of its parts. Lewin coined the term *group dynamics* in 1939 and is considered the pioneer of viewing the group as a whole.

Impact of the British School In England, group therapy evolved largely as a result of the pressure to treat a large number of psychiatric casualties during World War II. Wilfred R. Bion, an analysand of Melanie Klein's, offered the hypothesis that the group has a separate mental life, with its own dynamics and complex emotional states, which he termed *basic assumption cultures*. His basic assumptions paralleled Klein's stages of infant development. Bion is best known for his work with the *group as a whole*, and his ideas have gained immense popularity in the United States. Henry Ezriel in 1950 and John D. Sutherland in 1952 were also active in group therapy, emphasizing the “here-and-now” and “then-and-there” perspectives. Ezriel is also credited as the first person to describe the transferences between the members themselves and between the members and the group as a whole.

Mainstream Group Theories and Therapies in the United States The field of group therapy has incorporated the historic influences into a solid clinical modality articulated by the stellar work of a number of theorists in the United States and their followers. Helen Durkin in 1964 integrated principles of systems theory with group dynamics into the field of group therapy. Along with Durkin, Henrietta Glatzer became a prime force in the application of psychodynamic principles to group psychotherapy. She is known primarily for the concept of “the working alliance.” “She used the concept to emphasize the importance of the member-to-member interaction as a principal component of the working alliance in therapy groups.” Saul Scheidlinger's prolific output set the stage for a broad-based view of group development, including concepts such as the “mother-group” and the regressive pull in groups that activates early childhood patterns.

The wisdom of the early theorists and the visionary quality of their insight about group treatment is reflected and revitalized in the work that many clinicians are applying to current problems of epidemic proportion such as breast cancer and acquired immune deficiency syndrome (AIDS).

THEORETICAL ISSUES

While group therapy was originally based on psychoanalytic principles, a broad range of theories now informs the practice of group psychotherapy.

Psychodynamic Theorists Most current analytic theorists define the psychodynamic umbrella as consisting of four major branches: classical theory, object relations theory, ego psychology, and self psychology. In addition, the analytic umbrella covers interpersonal theory, humanistic-existential theories, and feminist-analytic thought.

Classical Analytic The classical analytic emphasis on libido and aggression finds expression in the unconscious forces that propel the group as a whole along its epigenetic trajectory. In this model the group develops along the same epigenetic trajectory as the individual, namely oral, anal, phallic, and genital imperatives dictating the movement from early to mature group stages.

Object Relations Object relations theory finds in groups a natural environment for the projections of internal part-objects onto the other people in the room (members and leader) and the gradual reintrojections of the split-off aspects of the self within the containment of the group envelope.

Ego Psychology Anna Freud's interest in ego defenses by which an individual copes with potentially devastating anxiety influences a host of group interventions

having to do with recovery from physical or mental illness.

Self Psychology Self psychologists recognize the mirroring and empathic possibilities among committed members who can serve as selfobject functions for one another. All of the analytic group therapies rely on interpretation of resistances, defenses, and the multiple transferences that occur between members, between member and leader, and to the group as a whole.

Related Analytic Theorists Sullivanians and other interpersonalists such as Irvin Yalom stress the healing that comes from the real relationships among the members, and the feminists see in groups the opportunity to inspect the impact of gender on environment and vice versa. Existentialists depend on the group to explore, verify, and accept the individual's personal myth and intentionality in the world as they contribute to the individual's self authenticity.

Nonanalytic Theorists In addition to the psychodynamic theorists who have dominated the field of group therapy, a host of other theoreticians have contributed to the field of group psychotherapy. Some of them include the cognitive-behaviorists, the gestalt theorists, transactional analysts, and the redecision therapists.

Cognitive-Behavioral Theorists Cognitive-behaviorists see in the group model the opportunity to rethink old cognitive schemas and to question some of the prior cognitive distortions. Behavioral clinicians address social and psychiatric problems within a testable conceptual framework; thus the social influence that occurs in groups has a pivotal role in altering maladaptive responses by the use of modeling and reinforcement of new behaviors.

Gestalt Theorists Many of the theories that developed in the 1960s and 1970s used group models. The theory of gestalt therapy is defined by two central ideas: first, that the proper focus of psychology is the experiential present moment and second, that it is only possible to know the self in relation to other things in the interactive field. Gestalt theory as developed by Fritz Perls, Paul Goodman, and others from the Cleveland Gestalt Institute, brought the field theory of Kurt Lewin and Kurt Goldstein to bear on patients' attempts to integrate action, affect, and cognition. While working with an individual protagonist within the group, these thinkers used the whole group as a supportive container and reactor to the protagonist's clinical development. In gestalt theory the focus is on organizing the accumulation of past experience into the present reality of the person. Thus, the balance between figure and ground (or content and process) is the direction toward health.

Transactional Analysis Transactional analysis divides personality into three phenomenological components or ego states, colloquially termed *parent, adult,* and *child*. Eric Berne developed a group therapy approach that understands social behavior in terms of strokes or transactions exchanged by people in maladaptive patterns called games.

The redecision school of transactional analysis focuses on “ego state patterns” that may be thought of as character patterns, as expressed in the individual's relationship with others; these clinicians seek to resolve maladaptive early life decisions through contracts for change. Eugene Kerfoot and John Gladfelter describe redecision methods in a brief group therapy format and concern themselves with training models.

Self-Help Groups Alongside group therapy offered by the professional community, a large array of self-help groups have emerged and flourished. The first of these was Alcoholics Anonymous, which became a model for many other problem-specific support groups led by their own members. These groups have become a major source of recovery and cure and illustrate the value of community and group cohesion for managing a range of human distress. In particular, they focus on compulsive problems, such as addictions of all sorts, self-mutilation, and gambling.

All of the modalities listed above share certain curative factors. Table 30.4-1 contains a comprehensive listing of the range of therapeutic factors in group therapy.

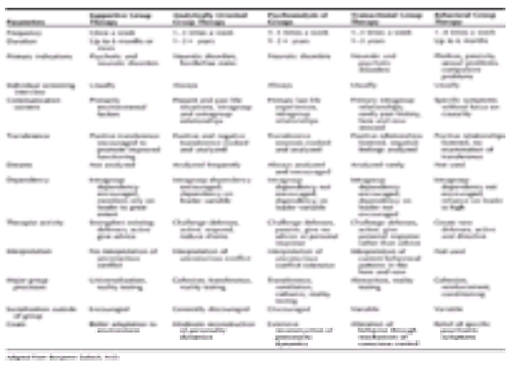

Table 30.4-1 Comparison of Different Types of Group Psychotherapy

THERAPEUTIC FACTORS

Whatever the model, group therapy rests on the assumptions that there are healing factors that emerge and operate in all groups and that some of these can be brought into play to allow the individuals in the group to grow and develop beyond the constrictions in life that brought them into treatment. While some group theorists rely on a cluster of factors relevant to their models of the mind and of pathology, all use some of the whole group of therapeutic factors identified in Table 30.4-2.

Table 30.4-2 Twenty Therapeutic Factors in Group Psychotherapy

It follows that the analysts will probably rely on the recapitulation of the family of origin, and on the way the transferences develop relative to those early family experiences; the cognitive behaviorists will exploit the factors that most rely on examining and clarifying schemas of more or less adaptive cognitions, to offer a simple example. At the same time, all of the theories rely on some universal and powerful factors that cut across all models.

Universal Healing Factors Going from the specific theories and their concepts of clinical change, processes common to all groups move the group on its developmental trajectory and enable the members to use the therapeutic factors delineated above. In what ways are groups therapeutic, and what promotes the healing forces in a group?

Cohesion Primary among universal healing factors is cohesion, which can be defined as a sense of belonging to a stable and accepting group, that relates to the attractiveness of the group to its members. The sense of we-ness seems to allow patients a high level of self-disclosure, to accept one another, to form meaningful

relationships, and to be supportive despite the obvious interpersonal difficulties experienced in the process. As these problems are worked through in the group, the gains in self-awareness and empathy toward others carry over into relationships outside the group, as do all alterations in the dyadic treatments.

Managing Shame Clinicians frequently worry about the shaming aspects of the disclosure that are part and parcel of the work in the group. It is true that going public with one's problems can be a daunting goal. If individuals in the group can manage to do so, paradoxically shame is reduced, patients begin to understand that they are part of the human condition and not subhuman or suprahuman, and development can resume. The group therapist sets the stage for the work of the group by developing a climate that is respectful and avoids undue shaming of any one member by underscoring the universal quality of all human pain and by encouraging frankness and empathic confrontation of the members by one another.

Amplification and Contagion Some of the earlier writers such as G. LeBon and W. MacDougall understood the effectiveness of the intragroup processes along with the dimensions of contagion and amplification. *Amplification* is a process whereby the individual's dimly felt or secretly held problems are spoken by another member or members in the group as centers for their own concern. The constrained members then have their affects amplified and universalized by hearing them as belonging to others in the group and feel a deepening awareness of the troublesome affects that can now be approached in the accepting community of group colleagues. Another factor is that of contagion. At its worst, of course, this is what turns a group into a mob, as has been seen all too often in history. But at its finest, it is a process that enables individuals to rise beyond their solitary life and that makes the group more powerful than the sum of its parts. Some examples of benevolent groups helping in times of natural disasters come to mind. Together with cohesion and the management of shame, these constitute the four fundamental cornerstones by which group treatment affects its members. All four basic curative processes (cohesion, shame reduction, amplification, and contagion) rest on the basic assumptions first mentioned, namely, that all people want to belong and to fit into a community. The power of that force propels each person beyond the constraints of personal inhibitions or cultural habits to achieve that sense of we-ness with the group.

A group of patients have been meeting in a large general hospital for a number of years. These are very distressed people who live isolated lives and exist on the bare margin of society. Most have been hospitalized; some many times. Still, they attend the group as faithfully as they can. When they do make it to the meetings, they speak of the agonizing effort it took to get them out of their cocoons and into the room, but they also say such things as “I set my week by this group.” The wish to join the group “family” can override the schizoidal defenses that so compromise their lives. Notably, the sense of family and community serves to keep these patients out of the hospital and able to function more or less autonomously.

SETTING THERAPEUTIC GOALS

Planning and Organizing a Group Before approaching the concrete work of planning and organizing a psychotherapy group, the goals of the group must be clearly understood and developed by the leader. These in turn depend on the setting, the population, the time available for treatment, and the training and capacity of the leader or leaders. Therapy groups can be organized with the following goals in mind:

1. Character change
2. Supporting homogeneous patient populations
3. Targeting certain symptoms
4. Reestablishing premorbid levels of functioning

The goal of the group informs the structure, membership, duration, and leadership of the group. Also derived from this decision is a rational group contract that can provide a focus to help contain the members of the group. These variables together become the context for the work. For example, if the group is an outpatient, open-ended group with the goal of character change and the resolution of life-long disabling patterns of feelings and behaviors, the contextual factors will differ from those of a group that is organized on an in-patient unit.

Character Change Character difficulties are tenacious for all human beings, from the healthiest neurotic to the most regressed patient that is treated. For all persons character problems are:

1. Outside the patient's awareness
2. Syntonic and perceived as “who I am” when brought into awareness
3. Resistant to change even when the patient wants to make a change
4. Repeated compulsively until worked through; that is, they are robbed of some of their power with each experience of successful change to better alternatives
5. Difficult to change without strong motivation to overcome psychological inertia

The neurotic patient finds intrapsychic conflicts emerging in the interpersonal field of the group and makes use of the group's curative factors to overcome the resistances to newer intimacies. Some of the earlier literature on group treatment expressed doubts about the appropriateness of this “uncovering” kind of group for sicker patients. More recently, group theorists have argued that in fact the distributed transferences in the group seem to mitigate an overly threatening regression for such patients. Character problems have a tenacity that is difficult to reorganize in a brief treatment. Instead the process requires frequent regressions in the service of the ego and attempts to work through the same characterologic habits again and again. At the same time, brief analytic group models have been used to address sectors of the personality difficulties to good effect.

Supporting Targeted Patient Populations Since the time that Pratt offered his “classes” for tubercular patients at the Massachusetts General Hospital in 1905, people have come together to commiserate with one another around common problems, to share information, and to learn how to deal with the impact of those problems on their lives. Groups have been organized around medical illnesses such as cancer, diabetes, and AIDS; around psychological problems such as bereavement; and around psychosocial sequelae of trauma such as war or natural disasters. The goal of such groups is to provide support and information embedded in a socially accepting environment with persons who can really understand what one another is going through. The treatment may emerge from cognitivebehavioral principles, psychodynamic principles, or psychoeducational ones. Frequently these groups are time limited, and members often join at the same time and terminate together. The problems addressed in these groups are found in a broad variety of patients from the very healthy to the more distressed and cut across other demographic variables such as age and culture. Research indicates that involvement in these groups can have remarkably positive effects on extending survival and quality of life for the severely ill patient such as women with end-stage breast cancer.

Targeting Certain Symptoms Group treatment may target specific symptoms. This approach to psychopathology is congruent with categorical nosological systems such as the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Diagnosis here is seen as symptomatic rather than developmental, and the treatment goal is to alleviate the symptoms. For example, patients with eating disorders or specific phobias are clustered in groups that promote skills for self-monitoring and for replacing the automatic symptom with a more adaptive set of behaviors and cognitions. These groups again may include members from a broad range of intrapsychic development, which is not the primary focus of the group. At the same time, some people who work successfully in these groups may want to continue the work of personality change in open-ended dynamic groups when the symptom is relieved.

Reestablishing Premorbid Levels of Functioning People in acute and immediate distress may find support in groups that have reestablishment of a person's equilibrium as their main goal. Patients who have suffered a breakdown of their lives and have needed to be hospitalized can use groups within the inpatient or partial hospitalization setting. These groups focus primarily on restructuring the patient's sensorium, managing the acute distress, and planning for a return to the community. These patients also need help in dealing with the shameful consequences of hospitalization and the sometimes elusive process of establishing outpatient treatment to support them upon discharge. A benevolent inpatient or partial hospital group experience is of special value in the latter problem, since affordable treatment will be primarily offered in group therapy for the foreseeable future.

TECHNIQUES

Developing the Group Contract All groups need a contract, or a set of agreements about process, content, and goals. The more explicit this contract is in the mind of the leader, the more a prospective patient can feel a full partnership in the decision to join a particular group. The contract provides a structure and lends meaning that advances the work of the group and provides the safe, predictable maturational environment for each member's progress. Nonetheless, each aspect of the group agreement will probably generate internal resistances that are part and parcel of all human change. The specificity of the agreements allows each person a baseline from which to examine and overcome these resistances. The agreement or contract needs to be congruent with the kind of group the leader will conduct, taking into account the multiple variables explored above.

Sample Contract An open-ended psychodynamic group may organize around the following contract in which each member agrees to:

- Attend all meetings for the whole time. Regularity of attendance keeps each member in the process and helps the group cohere, without having to worry about consistency of the participants.
- Be responsible for the fee. Groups offer a rare opportunity to deal with the management of money in a public arena. To that end, a discussion of the fee in the group can be most helpful.
- Put into words the problems, feelings, and thoughts that bedevil them. Safety in a group is preserved by an agreement to express feelings in words rather than in acts, but a major advantage is also the value of sublimating in language what is heretofore felt remotely or is out of awareness.
- Keep the relationships primarily therapeutic, not social. While this is not a universally accepted aspect of group agreements, most psychodynamic leaders try to avoid siphoning the conflicts by segregating certain interactions from the whole group. Some will ask that the members bring to the group interactions that may have occurred among them in outside meetings.
- Protect the identities of the other group members. Confidentiality is crucial in the exposed setting of the therapy group. Members may want to discuss what happened in their group, but the identity of the other members must be protected if the group is to meet its goals.
- Stay until they have finished what they came to do. In the course of the group, people typically become aware of the work needed to overcome the problems that brought them into treatment. In psychodynamic groups, this is less well defined than in groups that are symptom focused, and it must be continually assessed by the individual, the group, and the leader.
- When they think the time has come, they allow the other members of the group to participate in the decision to terminate. Learning to collaborate with important others is a major advantage to working in a group. Thus, allowing others' opinions to influence, but not dictate, one's decisions is a major accomplishment.
- Leave enough time to say good-bye. For many, saying good-bye has historically been so painful that it has been avoided or shortened to formulaic rituals. Taking the time to say good-bye to the group is another chance to symbolically say good-bye to other loves and other homes.

Selecting Patients The question of who should be treated in a group is as elusive as the question of who should be treated in psychotherapy. Perhaps a more salient question is that of therapeutic fit for a particular group.

Indications A psychodynamic group in which early developmental conflicts and relationships are assumed to interfere with the here-and-now of the patient's life should be reasonably homogeneous for the level of ego development and heterogeneous in every other regard. Mixing persons of different genders, cultures, and ages can be extremely useful so long as they emerge from a similar developmental spectrum. The differences among them can then be addressed and exploited to the advantage of the group members. However, when patients diverge sharply in levels of ego development, group cohesion and universality are compromised. For example, a group of patients who all experience severe anxieties about loss consequent to serious abandonment throughout their lives will work together in a group. On the other hand, mixing two or three such patients in a group of persons who are conflicted about intimacy and sustained relationships may well result in two subgroups that lack easy empathic rapport with the other's internal dilemmas. The specific symptom or population designation of other groups defines patient selection.

Contraindications Some patients cannot make good use of group therapy without former clinical intervention. For example, an actively manic patient may be more overstimulated than helped in a group. Another category of patients who frequently are referred to groups are severely schizoidal persons who have really never developed any kind of sustained human relationship. Placing these patients in a group overrides their capacity and may set them up for early failure. Acutely disturbed patients may need and deserve individual attention prior to entry into an ongoing therapy group. In all of these cases prior treatment—psychopharmacological, individual supportive therapy, or both—may enable the patient eventually to make good use of group therapy.

Preparing Patients for the Group Both anecdotal and empirical evidence indicates that investing significant time in preparing a patient for group therapy improves the chances of a successful entry into a group. In addition to the usual history taking, it is helpful to examine patients' fantasies and biases about groups and to collect the history of their participation in all kinds of groups such as family, school, sports, work, and friendships. This is the time to discuss the group agreements and the rationale that underlies them and to elicit the patient's collaboration in the enterprise by making as much information available as possible. Patients are helped by knowing how the group works, what the leader's role might be, and what they might expect for themselves.

WHAT HAPPENS IN GROUP THERAPY?

Basic Early Group Experience A major advantage in all psychotherapy groups is that patients enter a group and promptly “have” their problems instead of just talking about them. In attempting to get to know the other members of the group to feel a sense of belonging, people inevitably experience the emergence of the very conflicts that brought them to treatment in the first place. Thus, the phobic patient may find it difficult to come to the meetings, and the very needy patient may find that time sharing feels intolerable.

Most people enter therapy groups feeling that their problems are uniquely shameful and that their pathology sets them apart from competent human beings. Group members come to recognize that their problems are more similar than different from those of others in the group. As members increasingly feel accepted and cared about in the group, cohesion and universality promote the beginning of the therapeutic work for each patient.

Psychodynamic Group Therapy The fundamental principles of all psychodynamic treatment rest on eliciting and interpreting the transference and providing an opportunity to remember, repeat, and work through difficult experiences and introjects. It is assumed that these conflicts are primarily unconscious and that the unconscious is made conscious through the interpretation of the transference and replacement of the old bad objects with new good objects. In terms of ego psychology, archaic defenses are brought to light and interpreted so that new and more adaptive defenses may develop. Self psychology would suggest that the group serves as a self-object that offers a mirroring environment that can help each individual cope with empathic disappointments and emerge with a more cohesive sense of self.

Developmental Trajectory of a Group Bion and others since have been interested in mapping out the expected developmental trajectory of a therapy group and the expectable stages that evolve during that development. While numerous models have been offered, variations of the model set forth by Bion offer a clarifying way to understand the needs of the members in the early-, middle-, and late-stage group and to anticipate the developmental imperatives attendant to each stage. Since this is an epigenetic model, it applies most clearly to psychodynamic groups, but it may also be used to think about what happens in any group given the stage of engagement and the relative health and pathology of its members.

Group theorists have relied on certain developmental principles to organize and map out the trajectory of a group. Table 30.4-3 illustrates some of the developmental maps.

Early	Middle	Mature
Imperatives		
Establishing trust and dependency	Dealing with conflict: fight/flight	Establishing intimacy
Requisite processes		
1. Developing mourning	1. Splitting, projection, and projective identification	1. Increased capacity for empathy
2. Protection from undue humiliation	2. Reworking characterological problems	2. Capacity for concern
3. Nurturance	3. Managing interpersonal distance	3. Revisiting unresolved grief generated by healthy group termination
4. Early cohesion	4. Reworking guilt-ridden aspects of self	4. Hope and creativity increase
5. True cohesion	5. Moving from external to internal locus of control	5. Termination accompanied by sadness, excitement, and renewed vitality

Courtesy of Alonso H.

Table 30.4-3 Trajectory of Group Development

STAGE 1 The developmental chart represents the ideal group development, in which early cohesion is established by a clear sense of meaning, often imbedded in

the group agreements. As the leader encourages a sense of respect, the members feel protected from undue humiliation or assault and can begin to experience nurturance from the group. The leader's group-as-a-whole interventions aid in this awareness of universality. It is common in the early group for members to say to a virtual stranger, “Why, we are really alike. I know exactly what you mean!” The early group senses that the power of the group is greater than the sum of its parts, and the group can proceed to the next stage of development.

STAGE 2 As the group begins to feel too much like a tight little island, relief from the fantasy of merger may come in the way of a split in which negative transference is directed at the leader. If the leader can welcome this development of the terrible 2s stage, the group can move on to the work of dealing with dissent, difference, and the resolution of archaic characterological ways of struggling with disappointment and bias. This is the stage of projective identifications and external locus of control, where the impulse to flee is most prevalent.

Projective Identification Recent theorists have emphasized the value and ubiquitousness of a primitive defense initially defined by Klein. Projective identification has come to be seen as one of the most powerful curative forces in a therapy group that provides the opportunity for each member to temporarily disown the more noxious aspects of the self by eliciting them in another. That other is a willing participant in this process, in exchange for similar favors. For example, a couple may be locked together in an illusion that one person is vulnerable and ill while the other person is only strong, helpful, and long-suffering. Each member in a sense is disowning the half of the self that is now invested in the other. In a sense this is a form of transiently psychotic transference in which the other is perceived as the container of badness and an impediment to one's growth and well-being. Each unconsciously splits off that unacceptable part of the self, thereby reducing each person's anxiety. In a therapy group, this illusion is difficult to maintain, since other competing attachments and perspectives from the members and the leader challenge the tenacity of the split. At best, this allows the work of analysis to proceed until persons are ready to reown the split-off parts of their personalities. This process of oscillating projections and introjections promotes increasing reintegration that smoothes the way for resolution of unconscious conflicts and neurotic traits.

Transference distortions give way to clearer relatedness with others in the here-and-now. The individual is now available for true intimacy with a separate though related other.

> Mr. R. and Ms. G. were members of a group that was deeply locked in fight-flight stages of development. Week after week, Mr. R. would rail at Ms. G. for whining instead of getting a job and, indeed, a life. She would typically burst into tears and finally threatened to leave the group if Mr. R. stayed in the group. This cycle was repeated so many times that the other members finally pointed out that for all they hated each other, they seemed unable to avoid locking horns week after week instead of dealing with other more sympathetic partners in the group. Eventually they called a truce and felt bereft, to their great surprise. Slowly they grew to know and recognize in each other what had been heretofore split off and denied in the self.

Over time, as group members reown their projections in the ensuing relatedness among members, their differences seem less threatening and actually come to represent a chance for expansion of the capacity to empathize with others without requiring that they be identical to the self.

STAGE 3 The mature group is characterized by a mutual concern among the members. As the members begin to accept the concept of good enough to replace the demands of perfection for self or others, intimacy is now possible in new measure. Usually, a member terminates at this stage, and that departure is an opportunity for all to revisit the areas of unresolved grief in their own lives. The movement from melancholia to greater completion of mourning results, as always, in a renewed sense of hope and increased creativity. Terminations are accompanied by a combination of sadness, excitement, and renewed vitality.

Resistance to Change The scenario above is idealized. In fact, many impediments arise in a group, which take the form of avoidance, resistance, and even early dropouts. However motivated, change is painful for all persons and it mobilizes the defenses in the service of maintaining the status quo, as the familiar devil we know. People mobilize their defenses and demonstrate their symptoms in the group. In the era of categorical diagnostic systems, it is difficult to remember that symptoms are compromise solutions to deeper problems, by which patients have tried to avoid being overwhelmed over the course of their lives. What has come to be called acting-out, with all its imbedded judgment, is nothing more than an opportunity to help patients in the group develop a deepening awareness of the cost of the archaic solution, as members confront each other while empathizing with the tempting solution of replacing words or feelings with distracting action. The pull to continue to feel cohesive with the peers in the group motivates the patient to tolerate the challenges and to seek new ways of dealing with difficult feelings and impulses.

Transference in Groups The standard target for psychoanalytic group treatment is the reawakening of early neurotic and characterological problems in the transference that emerges in the clinical hour. The emergence of repressed and avoided conflicts is facilitated by the multiplicity of transference targets and the powerful impact of the above-mentioned curative factors. As transferences develop from member to member, from member to leader, and from member to the group as a whole, people reexperience some of their earliest conflicts.

Defenses as Adaptation Resistance emerges and can be explored to illuminate the underlying defensive structure of each individual. With increasing safety, members regress to earlier defensive modes and arrive at increasing insight into the underlying personality dilemmas. These can then be experienced and analyzed in the group, by both the other members and the leader. Initially this process can feel shaming and threatening in a number of dimensions. However, with each event all members become more deeply aware of their own projections and displacements and can gradually reown those split-off parts of the self. When this happens the splits that had depleted the ego are healed, and psychic energy is made available for the conscious goals of the ego.

Self-Help Groups The emergent self-help movement has generated groups that differ from the above in a number of ways. Self-help groups are egalitarian and leaderless, focused on the principles of universality and unconditional positive regard, and open-ended and variable in their membership. Their goal is not necessarily to promote interaction among members; indeed, most discourage the kind of interpersonal discourse commonly found in the group therapies described above. The common enemy is the disease, and the common bond joins the members in resistance to succumbing to the ravages of that disease.

The Role of the Leader in The Group Before the group begins, the leader must make several decisions that have major implications for the whole enterprise. Beyond the obvious focus on the kind, duration, and theoretical underpinnings, the leader must then decide on matters of:

1. Membership
2. Logistics, such as place, time, and fees
3. Whether to work alone, or with a cotherapist
4. Whether patients will be treated in group therapy alone or in some combination of group therapy, individual therapy, pharmacotherapy, or self-help group
5. Managing records and protecting confidentiality

In addition to the logistical decisions listed above, the leader's stance must be consistent with the goals of the group. Leaders of a psychodynamic open-ended group are probably more likely to sit back and allow the group associations to lead the way for the group's work while they comment, like a critic at a concert. On the other hand, this stance makes little sense for the leader of a cognitive-behavioral group, who will be engaged in conducting desensitization exercises and providing cognitive restructuring, including homework exercises to meet the goals of that therapeutic endeavor. The range of clinical stances related to leader and group variables is illustrated in Table 30.4-4.

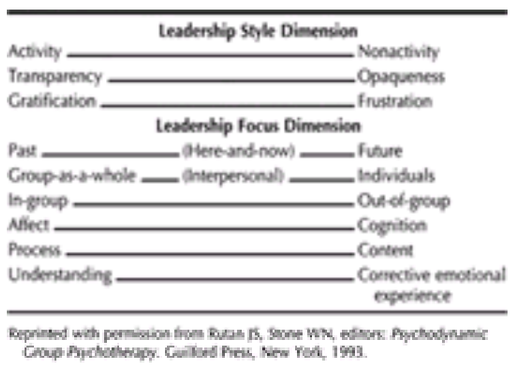

Leadership Style Dimension		
Activity		Nonactivity
Transparency		Opaqueness
Gratification		Frustration
Leadership Focus Dimension		
Past	(Here-and-now)	Future
Group-as-a-whole	(Interpersonal)	Individuals
In-group		Out-of-group
Affect		Cognition
Process		Content
Understanding		Corrective emotional experience

Reprinted with permission from Rutan JS, Stone WN, editors: *Psychodynamic Group Psychotherapy*. Guilford Press, New York, 1993.

Table 30.4-4 Leader Dimensions of the Group Therapist

Authority and Group Leadership A group leader must exercise authority about all the above factors if the group is to be safe and containing for its members. Whether a member who is difficult in the group stays or leaves or whether a new member enters must not be left to a vote, no more than it is in a family. The privilege and burden of administrative and matters of inclusion and exclusion is a serious responsibility of the group leader as is the question of single or cotherapy leadership. The leader is not a member of the group, despite the ambivalent entreaties of the members to bring the leader into the group, any more than the group is a democracy. The clearer the leader is about the boundaries, the safer are the members to indulge all the fantasies of wanting to corrupt the process or overcome the leader's authority.

> Dr. R.'s group had recently been decimated by the premature departure of a very difficult patient who, it turned out, had also been attending another group at the same time and decided to go with “the better one.” The group was furious and told her so, but she left after citing their anger as proof of the inferiority of the group and the leader. The following week, the group learned that Mr. T., a long-standing and much loved member, had cancer and would have to be away periodically for treatments. The group was adamant that the leader should not fill the departed patient's chair, and in fact they voted to make the group a six-member group, challenging the leader to deal with his distress in the group as well, if indeed he was not a “cold fish.” He resisted their entreaties and at the same time remained empathic toward their wishes. Eventually, he did introduce a new member who settled in nicely, and the group proceeded. Had the leader succumbed to the break in the group frame, this group would have felt deeply threatened and ineptly held.

The fiduciary responsibilities of the leader are better maintained to the extent that the leader exercises restraint and relative neutrality in the sense of nonjudgmental listening and responding to the patient's struggles. By remaining warm and neutral, the leader can listen nonjudgmentally to all aspects of the whole group's impulses and resistance, without taking sides or carrying the burden of somehow policing the group and deciding which are good feelings and interactions and which are not.

Incorruptibility of the Leader In the therapy group, as in any therapeutic work, the leader must remain as pure as Caesar's wife. No dual roles, no overly familiar incursions into the members' personal lives (nor theirs into the leader's), and no special fee arrangements that are not in the awareness of the group can be tolerated without damaging the integrity of the group boundaries. In short, all group business that cannot be conducted in the group ought not to be conducted at all. Refusal to hold secret any extragroup contacts sets a fine model that says the group is the safe therapeutic agent.

Question of Cotherapy As in many clinical decisions, the question of leading alone or with a colleague depends in part on the model, in part on the context and setting, in part on the availability of an appropriate cotherapist, and the system's support, administrative and otherwise, for committing two professionals to the same task at the same time. Inpatient or partial hospitalization groups tend to meet several times a week, and for those groups, cotherapy is a useful way to ensure continuity of leadership. On the other hand, some analytic group leaders avoid cotherapy because of the splitting of the patient-to-leader transferences. There are no rigid rules, but certain caveats must be observed.

Cotherapists work best when:

1. They are truly cotherapists, of relatively equal status and experience. When a student and a supervisor work together, it is useful to acknowledge this reality.
2. They share a common theory base.
3. They are willing to dedicate an hour or so per week to working out their collaborative problems and their perceptions of the group.
4. They are comfortable in sharing the fee.

Failure to observe these agreements may leave the patients low on the priority list of therapeutic concern while the cotherapy pair compete or otherwise undercut one another. On the more positive side, when cotherapy works well, both clinicians and patients have the advantage of two professional heads and hearts working in concert for the benefit of all.

Benefits for the Group Leader Groups can function as a hall of mirrors, reflecting some aspects of each member including those that are out of awareness. Benefits that accrue to the members from this expansion of self-awareness also accrue to the leader. If knowledge is power and psychological knowledge empowers the individual to lead a better life, then the knowledge of self can increase the leader's personal and professional depth as well. It remains a truism that it is always a race between the patient and the therapist as to who cures whom first. This is eight times truer in a cohesive and well-functioning group. The leader of a group is in a position to develop vicariously each time a member works through another of the universal dilemmas that make up the human experience. In the same vein, the not-so-central role of the leader in a group can be a wholesome and humbling reminder of the innate capacity of the patients to find healing in themselves and in one another, thereby mitigating godlike fantasies in the group toward the leader, and in leaders toward themselves.

Stumbling Blocks for the Group Leader Given the clear clinical advantages of leading therapy groups in the fiscal climate of the late 1990s, it is intriguing to wonder why the option remains difficult for many clinicians. On closer scrutiny, however, a number of potentially daunting reasons can explain this reluctance. Some have to do with the leadership process itself, others concern training for conducting groups, and still others relate to the context of clinicians, and whether they can gather enough referrals to sustain a group practice. The countertransferential pulls on the leader are geometrically increased in a group, just as the power of the multiple transferences have a major impact on the members.

WITHIN THE GROUP The group leader's mistakes are always open to the scrutiny of the whole group. When the members are encouraged to be honest and critical toward one another, the leaders can hardly do less about their personal or professional gaffes.

> Dr. X. announced to her group that they had about 5 minutes left today. The members pointed out that in fact there were 35 minutes left to the session. Dr. X. had nowhere to hide and needed to explore, in the group and later with her consultants, what had led her to want to curtail the session. She came to realize that the group was deeply involved in matters of grief that were currently relevant to her own life and that she had suppressed during the meeting. She then needed to find ways to say some of this to the group without generating a situation in which they felt they needed to care for the leader before they could go back to expecting care for their own concerns.

HELPLESSNESS Confronted with the sense of despair that may be resonating in several members of a troubled group, leaders may find themselves feeling helpless and of little use to the group. For example, some abbreviated treatments mandated for seriously disturbed patients may challenge any clinician's hope and optimism. If the group leader is also personally threatened by rescue fantasies, the group can only amplify the sense of clinical futility.

A group leader is always the target of a set of transferences, whether or not the work is psychodynamic and whether or not the group focuses consciously on managing and interpreting the transferences. It may be more challenging to avoid taking personally the cumulative mistrust, anger, or idealization for that matter, especially at times in the leader's life when loved ones are missing or when work is a source of difficulty. The pull of the group cohesion is powerful, and it is not uncommon for leaders to look at a departing, well-functioning group and wish that they were a member of that group, rather than the leader.

Another set of internal problems arise when the therapist needs to feel in control of the treatment situation. The dyadic model allows at least the fantasy of the therapist being in control, whereas in the group, especially an open-ended psychodynamic group, the members and the group as a whole move the process along, while the leader accompanies and shepherds it, at most.

Inadequacy of Training In most training programs, group training remains minimal, the common assumption being that clinical training is somehow generic and what works for individual therapy will serve as well for group treatment. This failure to take seriously a whole body of theory and research can leave the clinician at sea about how to even think about initiating a group, let alone learning how to manage impasses when they occur.

In fact, many training programs have curtailed or even eliminated the experience group for their trainees that gives them a taste for group process. Indeed, in some cases the training group may be conducted in ways that traumatize rather than support the trainees; when that happens the likelihood of trainees moving on to conduct their own groups is low.

Since most clinicians still enter individual treatment first for their own problems and self-awareness, the stage is set early for perceiving group therapy as second-rate treatment. At the same time they experience the pressures from the work environment to care for their patients in groups, which tends to generate anxiety and

disparage the modality, with little good experience on which to base those judgments.

Opportunities for Further Training The whole field of mental health has so expanded that typically, even the best training programs afford far too little time to learn to be an expert in anything. Across the United States and in other countries as well, training programs have sprung up that provide full postgraduate training, and some meet the criteria for membership in the National Registry for the Certification of Group Therapists. Affiliates of the American Group Psychotherapy Association typically offer training opportunities, as do other continuing education centers, some free standing and some affiliated with teaching hospitals. Given the parsimony of group treatment and the pressures on the mental health dollar, such training can more than pay for itself in many dimensions, both professionally and financially.

Work Context Leaders of therapy groups need to ensure enough referrals to keep the groups populated, with enough confidence to remain calm in the face of real or fantasied departure of one or more members. Thus, the need to join organizations that acknowledge the clinician as a competent group therapist or otherwise place the group in the awareness of potential patients and referrers is of paramount importance. In addition, it is most useful to meet regularly with a group of like-minded colleagues who will help each clinician stay invested in the difficult early stages of group formation.

GROUP THERAPY COMBINED WITH INDIVIDUAL THERAPY

Occasionally, patients are treated in simultaneous individual and group therapy, either by the same therapist or by two different ones. This option is useful for a variety of patients.

Overintellectualized Patient For some patients, insight becomes a way to avoid feeling. When it does, placing these patients in a therapy group lets them see how they interact with others who offer feedback and affective resonance.

> Mr. N. was referred for group treatment by his analyst, who felt the analysis was stalled. Indeed, Mr. N. appeared to know himself quite well, at least intellectually. In the group, he presented a seamless “case history” of himself, retelling a tragic childhood with calm and equanimity. One of the other patients began to weep as he spoke, recognizing similarities in her own life. Another expressed doubt about how cool Mr. N. was when presenting apparently very private material to a group of strangers and doubted that he was affectively present in the narrative. Mr. N. was at first startled, then reacted by assuming that they had not yet worked through their feelings as he had. The group gently confronted him with this assumption, and eventually he became more in touch with his intense affect, which had been blunted in the dyadic transference lock with his analyst. Both treatments proceeded successfully to a happy termination.

Patients who Cannot Tolerate the Dyadic Transference of Individual Treatment Dyadic treatment can threaten the fragile ego boundaries of patients who are either very needy or who are overstimulated by the apparent promises of the individual work. These patients often flee treatment or regress to terrifying actions that can be life threatening in near psychotic levels of transference. Adding group therapy can distribute the transferences across all the members of the group and the group leader and may enable the individual and the group treatment to proceed more safely and more productively.

> Ms. B. was referred to group treatment after her fifth serious suicide attempt; each of these had occurred during the absence of her individual therapist. She came to the group most reluctantly, fearful that she was going to be dropped off in the group and abandoned. She viewed the group leader with much distrust but found two or three members of the group who could empathize with her plight, many of whom had also suffered similar self-destructive incidents and could help her move away from taking action. Her presence in the group was stormy and embattled for years, finally ending with her becoming a strong proponent of group treatment. Notably, she was never hospitalized after joining the group. Finally, she concluded her individual work and remained in the group with the full agreement of both therapists.

Collaboration Between the Two Therapists In the case of Ms. B., it was crucial that the two therapists collaborate by frequent phone calls and by avoiding her many attempts to split them. This was possible because they had a similar level of experience and shared a parallel view of her strength and pathology. When two therapists do not agree or do not respect one another's work, patients are potentially at great risk of harm to themselves or at least in a stalemate of iatrogenic origin.

Combined Treatment With the Same Therapist Patients are sometimes seen both in individual and group therapy with the same therapist. This treatment plan has many advantages and, as always, some costs.

It is illuminating to deal with the intrapsychic dimensions of a patient in the individual hour and then observe this patient express those internal dilemmas and live them out in interactions with members of the group. For example, mild and extremely gentle individual patients might startle their therapists by launching an aggressive attack on one or more members of the group when the shy facade is challenged.

For some patients, however, sharing the therapist's attention can be so distressing that the work of therapy is stalled. It may be far better for that patient to be referred to another leader's group or to defer group treatment to a more secure time.

Boundaries in Combined Treatment Clinicians struggle with how much to preserve the privacy of what they know about the patient from the dyadic hour when the patient enters the group and how much to disclose. While there are no hard and fast rules here, what matters is consistency and the prior agreement with the patient about this matter. Many clinicians opt to protect the information while urging the patient to bring the problems into the group. One major exception is the case where the group or one of its members is at risk; as usual, the rules of confidentiality are suspended when there is any threat to the safety of any of the participants. It is useful to agree that all treaters will be in regular contact with one another to work for the patient's advantage.

LEGAL AND ETHICAL ISSUES

Confidentiality and the Leader The group leader is bound by the usual rules of confidentiality in the group as in any other clinical encounter with the patient. With the exception of a threat to a person or persons, this confidentiality holds and is usually elaborated in the code of ethics of the therapist's professional organization. The same obtains for any conduct of the therapist that violates the code of ethics relative to sexual or other extraprofessional contact with a patient.

Confidentiality and Group Members The bigger challenge is confidentiality among members. Aside from stressing the importance of protecting the identities of patients in the group, the leader can do little to ensure compliance, nor is it against the law for members to break confidentiality. Group therapists worry about whether the group can become a pool of witnesses in the case of a subpoena. Some states extend the same protection to the members that they do to the leader (namely, the patient-therapist privilege), but this has not been tested and may not apply with all professionals across the disciplines.

RESEARCH AND EVALUATION

Group therapy research has tended to focus on outcome and prediction. More recently, measures have been developed that seek to relate the patient's sense of belonging and feeling valued in the group with the effectiveness of the treatment. Studies continue to support the importance of group cohesion on group effectiveness, and feeling valued is seen as a statement of cohesion. The research literature indicates greater confidence in the efficacy of group treatment, showing no appreciable difference between individual, group, or pharmacotherapy. But these studies remain problematic, given the problems that bedevil most social science research: it is difficult to control therapist differences, and attempts to do so by providing manuals for intervention become different models than those in real life. The nonspecific factors are elusive, but they still seem to indicate that patients progress when they feel cared about, when the leader is warm and somewhat structured, when the match with colleagues in the group is appropriate, and when the goal and direction of the group are clear and consistent.

Researchers have tended to move beyond the question of do groups work to a finer look at how they work, in what circumstances, and for whom. Proper and careful screening and otherwise preparing a patient to enter a group results in a greater chance of success in entering and staying. As stated earlier, groups in which the members are at about the same level of ego development do better than groups with a large disparity of ego development levels. Short-term, focused groups are more successful if leaders have structured agendas and time boundaries and if the patients are more homogeneous in regard to the problem being addressed.

Research instruments are proving useful for measuring patient satisfaction and self-reports of increased well-being. The Clinical Outcome Results battery developed by the American Group Psychotherapy Association is but one example and it uses such measures as the Symptom Checklist 90-Revised (SCL-90-R), the Social

Adjustment Scale Self Report (SAS-SR), the Multiple Affect Adjective Check List-Revised (MAACL-R) and the Global Assessment Scale (GAS). More recently, measures such as the Structured Analysis of Social Behavior and the Group Climate Questionnaire (GCQ) have been used extensively by MacKenzie and others working with patients in structured, time-limited groups.

A major shortcoming in group therapy research stems from the pragmatics of conducting research over a long time, with more amorphous goals. Thus, most data are derived from research on time-limited groups, usually within the cognitive-behavioral or interpersonal model. While those findings are important, they have limited applicability to the more open-ended dynamic models of group treatment. That research remains to be enlarged.

CONSULTATION AND SUPERVISION FOR GROUP THERAPY

It is often difficult for group leaders to ask for help, as it is difficult for most professional helpers, especially beyond the formal training years. But failure to find help in group leadership can expand the strain on the leader exponentially, given the number of people in the consultation room and the multiplicity of counter-transference vectors. A well-running group can look deceptively autonomous of the leader's impact, but the leader's calmness and full attention is the platform on which the group grows. Occasional consultations, ongoing peer supervision, or both, are safe, judicious practices. It also allows leaders to take advantage of their affiliative needs, to avoid using the patient group for dealing with the loneliness of the well-functioning group leader.

SUGGESTED CROSS-REFERENCES

Additional discussion of the conceptual models used in group psychotherapy appears in Chapter 6 on theories of psychoanalysis and on other psychodynamic schools and on approaches derived from psychology and philosophy. More information about the clinical application of those models is located in Section 30.1 on psychoanalysis and psychoanalytic psychotherapy, Section 30.2 on behavior therapy, Section 30.6 on cognitive therapy, and Section 30.8 on brief psychotherapy. Further elaboration about psychiatric emergencies is found in Chapter 29. Detailed descriptions and backgrounds about the special patient populations and self-help groups are in Chapter 11 on substance-related disorders, Chapter 12 on schizophrenia, Section 13.2 on delusional disorders, Chapter 14 on mood disorders, Chapter 15 on anxiety disorders, and Chapter 16 on somatoform disorders. Elaboration about other population groups for which group therapy has been used can be found in Chapter 25 on psychological factors affecting medical condition (psychosomatic disorders), Chapter 24 on personality disorders, Section 51.2c on the psychological aspects of aging, Section 51.3 on the psychiatric disorders of late life, and Section 51.4l on group therapy with the elderly.

SECTION REFERENCES

Alonso A, Rutan JS: Activity/nonactivity and the group therapist: “Don't just do something, sit there.” Group *20:*43, 1996.

*Alonso A, Swiller HI, editors: *Group Therapy in Clinical Practice*. American Psychiatric Press, Washington, DC, 1993.

Bion WR: *Experiences in Groups*. Tavistock, London, 1961.

Bloch S, Crouch E: *Therapeutic Factors in Group Psychotherapy*. Oxford University Press, New York, 1985.

*Brabender V, Fallon A: *Models of Inpatient Group Psychotherapy*. American Psychiatric Press, Washington, DC, 1993.

*Cohen B, Ettin M: Self-structure and self-transformation in group psychotherapy. Int J Group Psychother *49:*61, 1999.

Durkin H: *The Group in Depth*. International Universities Press, New York, 1964.

Ezriel H: Psychoanalytic group therapy. In *Group Therapy: 1973, An Overview*, LR Wolberg, EK Schwartz, editors. Intercontinental Medical Book, New York, 1973.

Foulkes SH: Group process and the individual in the therapeutic group. Br J Med Psychol *34:*23, 1961.

Freud S: Group psychology and analysis of the ego. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*. Hogarth, London, 1962.

Frost J: Countertransference considerations for the gay male when leading psychotherapy groups for gay men. Int J Group Psychother *48:*3, 1998.

Gans JS: Broaching and exploring the question of combined group and individual therapy. Int J Group Psychother *40:*123, 1990.

Gans J, Alonso A: Difficult patients: Their construction in group therapy. Int J Group Psychother *48:*311, 1998.

Ganzarain RC, Buchele BJ: *Fugitives of Incest: A Perspective from Psychoanalysis and Groups*. International Universities Press, Madison, CT, 1988.

Glatzer H: The working alliance in analytic group psychotherapy, Int J Group Psychother *28:*147, 1978.

*Jackson D: The team meeting on a rapid turnover psychiatric ward: Clinical illustration of a model for stages of group development. Int J Group Psychotherapy *49:*41, 1999.

*Janoff DS, Schoenholtz-Read J: Group supervision meets technology: A model for computer-mediated group training at a distance. Int J Group Psychother *49:*255, 1999.

Kanas N, Cox P: Process and content in a therapy group for bipolar outpatients. Group *22:*39, 1998.

Kaplan HI, Sadock BJ, editors: *Comprehensive Group Psychotherapy,* ed 3. Williams & Wilkins, Baltimore, 1993.

Kelly JA, Murphy DA, Bahr GR, Kalichman SC, Morgan MG, Stevenson LY, Koob JL, Brasfield TL, Bernstein BM: Outcome of cognitive-behavioral and support group brief therapies for depressed, HIV-infected persons. Am J Psychiatry *150:*1679, 1993.

*Klein RH, Bernard HS, Singer DL, editors: *Handbook of Contemporary Group Psychotherapy*. International Universities Press, Madison, CT, 1992.

Leszcz M: The interpersonal approach to group psychotherapy. Int J Group Psychother *42:*37, 1992.

Leiberman MA, Yalom I, Miles M: *Encounter Groups: First Facts*. Basic Books, New York, 1973.

Leszez M, Goodwin P: The rationale and foundations of group psychotherapy for women with metastatic breast cancer. Int J Group Psychother *48:*245, 1998.

*MacKenzie KR, editor: *Classics in Group Psychotherapy*. Guilford, New York, 1992.

Malan DH, Balfour FHG, Hood VG, Shooter AMN: Group psychotherapy: A long-term follow-up study. Arch Gen Psychiatry *33:*1303, 1976.

*McCallum M, Piper WE: Personality disorders and response to group-oriented evening treatment. Group Dynamics *3:*3, 1999.

Moreno JL: *Psychodrama*. Beacon Press, Beacon, NY, 1947.

Motherwell L, Prudent S. Childlessness and group psychotherapy: Psychological and sociological perspectives. Group *22:*145, 1998.

Pam A, Kemper S: The captive group: Guidelines for group therapists in the inpatient setting. Int J Group Psychother *43:*419, 1993.

Rice C: Group therapists, poets and other artists: Reflections on God, the devil, and projective identification. Int J Group Psychother *48:*107, 1998.

Riester AE, Kraft IA, editors: *Child Group Psychotherapy*. International Universities Press, Madison, CT, 1986.

*Rutan JS, Stone WS, editors: *Psychodynamic Group Psycotherapy*. Guilford, New York, 1993.

Scheidlinger S: On the concept of “mother-group.” Int J Group Psychother *24:*417, 1974.

Weiner MF: *Techniques of Group Psychotherapy*. American Psychiatric Press, Washington, DC, 1984.

Wong N: Clinical considerations in group treatment of narcissistic disorders. Int J Group Psychother *29:*325, 1979.

30.5 FAMILY THERAPY AND COUPLE THERAPY

ALAN S. GURMAN, PH.D. AND JAY L. LEBOW, PH.D.

Family and couple therapy is a rubric that subsumes a large number of therapeutic forms, varying widely in their theoretical foundations, structure and process, and treatment aims. Family-couple therapy has been defined as “any psychotherapeutic endeavor that explicitly focuses on altering the interactions between or among family members and seeks to improve the functioning of the family as an amount, or its subsystems, and/or the functioning of individual members of the family.” These family and couple therapies have been used in the treatment of disordered individuals as well as dysfunctional relationships, across all developmental levels, and with a broad range of family structures. *Couple therapy* has recently come to replace the more familiar and limiting term *marital therapy* because of its emphasis on the link and bond between two people, without the associated judgmental tone of social value implied by the more traditional term. Although couple therapy is generally viewed as a subtype of family therapy, the majority of influential clinical theoreticians in the family realm have in fact infrequently focused on the treatment of conflictual couples' marriages.

Table 30.5-1 presents the major approaches to family and couple therapy in terms of central concepts, typical goals, and commonly used strategies and techniques and offers brief comments on each method's current status. While the schools of family and couple therapy are at times subdivided into problem-solving and transgenerational types, there has been so much cross-fertilization among them since the late 1980s that such distinctions are now more apparent than real. Also, there is significant historical and conceptual overlap among many of these thirteen approaches. Nonetheless, each therapy approach holds a unique place in the broader field. Some family and couple therapy methods involve logical extensions of therapeutic work with individuals whereas others clearly either have anti-individual therapy origins, or at least arose outside traditional treatment settings. Family and couple therapy approaches are at least as varied as individual approaches in terms of such central organizing principles as length, temporal focus, intrapersonal versus interpersonal emphasis, typical goals, and views on the role and stance of the therapist. In addition, highly specialized family and couple intervention methods have been developed for clinical populations with a wide range of common presenting problems, for example, infidelity, divorce and remarriage, abuse and violence, sexual dysfunction, and chronic illness.

Table 30.5-1 Major Approaches to Family and Couple Therapy

HISTORY

Contemporary family and couple therapy has varied historical roots, including the fields of group dynamics, family life education, gynecological medicine, sexology, social casework, and psychoanalysis. Couple or marital therapy traces its roots rather independently to the earlier marriage counseling tradition, spearheaded in the late 1920s and 1930s by such famous clinics as the Marriage Council of Philadelphia, the Marriage Consultation Center in New York, the American Institute of Family Relations in Los Angeles, and the Marriage Guidance Council centers in the United Kingdom. Emerging in the next two decades were the *collaborative* (two therapists) and *concurrent* (same therapist) husband-and-wife psychoanalysis and the more radical *conjoint marital therapy* (the partners meet with the same therapist together). Such inclusion of spouses in treatment showed the earliest recognition that problematic family interaction patterns maintain what is purportedly individual psychopathology, and that intervention with symptomatic individuals is often inadequate to the therapeutic task.

The major accelerant of the dominant theories of modern family therapy, however, was the rapid and independent emergence in the late 1940s and 1950s of clinical and research interest in the multigenerational transmission and reinforcement processes involved in various psychiatric disturbances, especially psychoses and schizophrenia. Nathan Ackerman and Salvador Minuchin's groundbreaking family work with child- and adolescent-focused problems began in the same general period, and Ivan Boszyormenyi-Nagy and James Framo brought the ideas of intergenerational psychodynamics to the forefront. Although a large number of the early family therapy pioneers were psychiatrists, innovators from other professions were also highly visible in this rapidly growing field, for example, John Bell and Framo (clinical psychology), Haley (communications), and Virginia Satir (social work). Even anthropological (Gregory Bateson) and hypnotherapeutic (Milton Erickson) perspectives helped to establish the early foundations of the family therapy movement. Interestingly, a great deal of early conceptual work in family therapy occurred in clinical research settings. Regardless of substantive focus or work context, however, what all the first-generation family therapists had in common were a passionate questioning of the dominant individual models (primarily psychoanalytic) of problem development, maintenance, and treatment, and an assertion of the explanatory power of thinking in terms of circular versus linear causality. The conceptual fervor that characterized such systemic truth-seeking was often associated with a rather close-minded dismissal of extant views of mental health intervention.

The 1960s saw several groundbreaking professional conferences in the field and the founding of what were to become influential training institutes. The first family therapy journal, *Family Process*, was inaugurated in 1962, opening a new era for the field by giving it a home journal whose board of directors and editorial board included almost all the luminaries of the day.

The 1970s was a period of major expansion in the field, both conceptually and organizationally. By this time, the vast majority of currently influential approaches to family and couple therapy had become established in a variety of training centers, both free-standing and university-based. Schools of therapy proliferated, and some of the energy that had previously been put into the conceptual battle with mainstream mental health perspectives was redirected across scholastic lines. Organizationally, two major events in American family and couple therapy took place. In 1978, the American Association of Marriage and Family Counselors (AAMFC), formerly the American Association of Marriage Counselors, changed its name to the American Association for Marriage and Family Therapy (AAMFT) and soon became the official accrediting and credentialing body in the field. In 1978 the American Family Therapy Association, now American Family Therapy Academy (AFTA), was founded by most of the early pioneers in the field and originally consisted of many senior clinicians and researchers.

The 1980s and 1990s have witnessed four major patterns of continuing interest and influence in the field of family and couple therapy. First, there has been an extraordinary explosion in the professional literature, with the appearance annually of dozens of new books, and the creation of more than three dozen journals spanning all five continents and published in almost a dozen different languages. Second, research activity has rapidly evolved due in part to the growth of family psychology, focusing a good deal on comparative treatment outcome studies and the treatment of diagnosable psychiatric disorders as well as explicitly relational difficulties. Third, the field has begun to embrace a variety of perspectives on relational functioning, such as those based in feminism and multiculturalism, that have broadened the conceptual and perceptual lenses of the human condition beyond those typically set forth by the dominant and traditional mental health points of view. Also, having established a pervasive footing throughout the world of psychotherapy, the field of family and couple therapy has become increasingly integrative. This integration movement, which followed a parallel movement in the domain of individual psychotherapy, has brought together purportedly incompatible theories of family or couple interaction and intervention and has coordinated interventions at multiple levels of behavior and experience, for example, intrapsychic, interpersonal, and

biological. In these ways the field has become more consistently and genuinely systemic and contextual.

Theoretical Issues The defining and distinguishing conceptual characteristics of family-couple therapists, regardless of their theoretical allegiances (Table 30.5-1) is their emphasis on the context in which clinically relevant behavior occurs. While more strident family therapy theoreticians and clinicians may restrict their vision to the realm of human interaction, an ecumenical attitude is more typical of systemically oriented clinicians. As a result, intrapsychic processes (especially conscious and preconscious) of individuals are not seen as beyond the purview of family-oriented therapists, and clinically relevant biological phenomena are seen as factors that are potentially as relevant to treatment as any others in the given clinical case. In contrast to the conceptual extremism that has at times characterized some leaders in the family and couple field, the arbitrary punctuation of understanding human experience is not consonant with effective and responsible systems-oriented psychotherapy.

Family Systems Theory Much of the newer orientation to a family systems view of clinical problem formation and treatment derives from a mixture of ideas from *general systems theory* (the study of the relationships between and among interacting components of a system that exists over time) and *cybernetics* (the study of systemic regulatory mechanisms that operate via feedback loops). A third major orienting perspective in family systems theory is that of *family development* (the family life cycle). The first two of these frameworks did not originate in the study of human behavior per se, and so are typically used metaphorically in discussing clinical events and processes. More human-level theories, especially those emphasizing psychodynamic and social learning principles, have also had a pervasive influence on the field.

CLINICAL IMPLICATIONS OF FAMILY SYSTEMS THEORY Family systems thinking is *antireductionistic* (wholeness, organization). The notion that the whole is greater than the sum of its parts (*nonsummativity*) implies the inappropriateness of trying to understand individuals (e.g., symptoms, problems, behavior) apart from their relational situation. Thus, “crazy,” “maladaptive,” or “dysfunctional” behavior often makes sense when viewed contextually within the framework of the social (family) system at hand. As a result, the role of the identified (or index) patient in the family can be varied. One member's difficulties may regulate problematic underlying family patterns, may itself be a major source of stress on the family, or may be a consequence of relationship conflict. Moreover, since there is no objective truth that best describes any clinical situation (“the map is not the territory”), multiple vantage points are valued (*constructivism*). Direct observation of family and couple interaction by the therapist is therefore complementary to patient reports. Although individual psychotherapy within a family systems perspective is certainly done, the preferred clinical mode of interviewing and intervening is with all (sub) “systems of import.”

Family life is viewed as a continuous balanced (*homeostatic*) process of adaptive change (*morphogenesis*) and maintenance of stability (*morphostasis*). To maintain its organizational identity, the structure of this social system must be flexible (*open system*) yet predictable (e.g., hierarchies, boundaries), so that the most salient requirements of the group family roles are satisfied. *Normality* and *pathology* are relativistic concepts in family systems thinking, and are contingent upon particular cultural practices and beliefs and the requirements of both individuals and other family units (*subsystems*) at different and relatively predictable stages of development (*family life cycle*).

In family systems thinking, focusing on the current organization of the family is often though not always viewed as a sufficient data base with which to intervene (“the system is its own best explanation”). Since it is assumed that an individual and a family or couple may have traveled a variety of etiological routes to the same point (*equifinality*), history-taking during the assessment phase of therapy is often minimal, and exploration of the past is not a dominant facet of many courses of family and couple therapy (although some methods do emphasize such investigation— Table 30.5-1). “What?” questions are as likely as “why?” questions. Family members and others may unwittingly reinforce the very problematic patterns that are of concern (collusion, quid pro quo, family rules) and the therapist's strategic and technical options are increased because change in any one part of the system may lead to change in other parts (feedback loops, circular causality). The essential intervention principle derived from family systems thinking is to place primary effort into blocking, disrupting, and replacing repeated sequences of interaction that maintain the clinical problems of interest, whether of an individual or relationship nature.

Techniques As is true of individual psychotherapy, the field of family and couple therapy includes an extraordinary array of both disparate and overlapping strategies and techniques. This vast landscape of clinical approaches can be addressed in two complementary ways. First, the variety of salient ideas and methods is sampled, and then descriptions are offered of typical or modal styles of clinical practice. Four aspects of the therapeutic experience are considered in this manner: the structure and composition of therapy, the role of the therapist, goal-setting and assessment, and specific clinical techniques.

Structure and Composition of Therapy In family therapy the question of who participates produces many answers. On the patient or client side, the complainant or “customer” (who may or may not be a patient in psychiatric or diagnostic terms) is always included. And while the patient (note that one may not be designated when a relationship is “the patient”) is typically present, this is not always the case, for example, a child-focused family treatment may regularly exclude the child from sessions with the parents that are focused on parenting skills, marital conflict, and other areas of concern. Family therapy may be conducted with one patient or client present (e.g., Bowen family systems therapy, Mental Research Institute brief therapy), may be limited to the nuclear family (e.g., behavioral, psychoanalytic), or may even be transgenerational (e.g., structural, symbolic-experimental): combined therapies (e.g., individual and family) are more common than is reflected in the literature. In everyday practice those included (other than the therapist) are whoever is thought to constitute the *systems of import,* which sustain the problem or may help to resolve it; the problem is usually operationally defined as the “minimum sufficient network” to be likely to achieve the desired therapeutic goals. Relevant systems may include nonfamily members such as other health care providers, school teachers, and representatives of various community agencies.

Most family therapy is conducted by a single therapist, though co-therapy (especially of one male and one female therapist) is somewhat common in training centers, and is even preferred in some family therapy approaches (e.g., contextual, symbolic-experiential). Moreover, therapeutic teams of several clinicians, usually comprised of the actual treating therapist and some number of colleague observer-consultants (behind a one-way mirror), are highly valued in certain models (e.g., Milan systemic, strategic, structural), especially in training institutes.

The length of family therapy ranges from ideal courses of fewer than ten visits (e.g., Mental Research Institute brief therapy, solution-focused) to up to dozens of visits (e.g., Bowen family systems) or even hundreds (e.g., psychoanalytic) over several years. Family therapy is generally quite brief by traditional standards of length in individual therapy, even when it is not explicitly time-limited (which is not the usual case), with most family and couple therapy lasting 10 to 20 sessions over a period of less than 6 months. Family therapy is typically conducted in private offices and clinics but not uncommonly takes place on hospital wards and even in family homes.

Role of the Therapist The therapist's role in family and couple therapy varies along several dimensions, most noticeably in terms of emotional closeness or distance to the family or couple. Three gross categories of the therapist's emotional proximity can be discerned: the educator-coach, the perturbator, and the healer. These fundamental relational stances vary as a function, for example, of the degree to which therapists intentionally and systematically uses his or her “self” (e.g., by self-disclosure of fantasy material, personal or countertransfertential reactions, or factual information), and explicitly addresses the nature and meaning of the therapist-family member relationship (e.g., transference analysis). The therapist as educator-coach (e.g., behavioral, Bowen Family Systems, Psychoeducational) sees him/herself as possessing expert, professional knowledge about human relationships and change processes, and attempts to impart such knowledge to the couple and families with whom he or she works as a basis for inducing change. The family therapist as perturbator (e.g., Mental Research Institute brief therapy, narrative, strategic, solution-focused) possesses expert understanding of problematic family processes, but tends to use this awareness from an outside stance to induce change in the family system without giving family members information, concepts, or methods they can take away from therapy for future use. The family therapist as healer (experiential-humanistic, psychoanalytic, transgenerational) places special value on the transformative power of personal relationships.

Despite such a divergence of views on the stance of the family therapist, it is widely agreed that a working therapeutic alliance is essential for effective therapy even though such a collaborative relationship is established in different ways in different therapeutic approaches. In everyday practice it is also often the case that therapeutic alliances need to be established in different ways with different family members, for example, one may respond well to empathic understanding from the therapist, whereas another may not begin to form a working alliance unless the therapist offers guidance and concrete suggestions. The therapist needs to track recurrent family and couple interaction patterns that maintain relational or symptomatic problems without inflexibly siding with any individual or other subsystem against others, although temporarily taking sides to unbalance fixed relational structures is often necessary.

Assessment and Goal-Setting Predictably, different family theoretical orientations emphasize different treatment goals, and thus involve different assessment foci. Since all family and couple therapies are concerned with interactional problem-maintaining cycles, the issue raised earlier about who should be included in ongoing treatment should always be assessed early. It is not unusual for initial assessment to include people who may be discharged from most or all of ongoing treatment, especially in child- or adolescent–focused cases. Marital therapy cases less commonly involve multiple generations, although family-of-origin sessions with the partners' parents are a possibility, and the partners' children are occasionally included, for example, if parenting problems coexist with couple problems.

ASSESSMENT METHODS The methods of family therapy assessment vary. Self-report paper-and-pencil methods are commonly used in behavior therapy for standardized measures (e.g., child behavior checklists) and for case-specific monitoring related to particular patient or family treatment aims (e.g., establishing

baseline measures of the frequency of problem behavior). Couple therapists not infrequently also use self-report inventories to help assess overall marital satisfaction as well as common areas of couple conflict. Genograms are often used in multigenerationally oriented methods to help the clinician and the family discern cross-generational patterns of maladaptive functioning. Finally, direct observational methods involving various microanalytic coding schema are occasionally used, especially in academic settings, but are too labor intensive to be used regularly in clinical practice.

The clinical interview is the major vehicle for assessment. It is often difficult for even a reasonably sophisticated observer to identify the theoretical orientation of family therapists during the opening stage of therapy in that therapists' interventions are often less discrepant than their conceptualizations of what they experience and observe. Nonetheless, therapists of different theoretical schools can be rather reliably differentiated in terms of the levels of assessment upon which they focus. Two such levels may be identified: the unit level, and the experiential level. The *unit level* is the composition of the psychosocial unit of evaluative interest. The individual (identified patient or others), the marriage or couple, the parental subsystem, the sibling subsystem, the whole family, and the family-plus-nonnuclear family social entities (e.g., grandparental subsystem, school system) may all be given careful attention. More psychodynamic, experiential-humanistic, and intergenerationally inclined therapists tend to be interested in assessing the potential treatment planning role (even if only by reference, rather than face-to-face) of a larger number of units whereas more pragmatic orientations focused on resolution of presenting problems (e.g., behavioral, psychoeducational, strategic) tend to assess a less complex array of these units. The *experiential level* refers to the level of organization at which assessment occurs (e.g., molecular, biological, unconscious, conscious, interpersonal, transpersonal). Therapists also differ on the related dimension of past versus present centeredness. The more pragmatic therapists (who focus more on presenting problems and teachable therapeutic techniques such as Behavioral, Psychoeducational, Mental Research Institute brief therapy, strategic, and structural tend to show little to no significant interest in either unconscious psychological processes or the family's or its individual members' past. More aesthetically oriented therapists (e.g., psychoanalytic, contextual, humanistic, symbolic-experiential), by contrast, who tend to espouse a more relationship-based style of intervention in which the “real” problem is believed initially to be hidden are more attuned to psychological events that are not so immediate. Indeed, such therapists' assessments tend to emphasize inference whereas the more pragmatic therapists' assessments tend to emphasize observation.

GOAL-SETTING Different theoretical orientations to family and couple therapy will emphasize different types of typical goals (Table 30.5-1) in line with their differing views of symptom-formation and symptom-maintenance and the role of the therapist. Most family- and couple-oriented therapists endorse many or most of the following goals, regardless of the particular nature of the presenting problem: (1) reduction of psychiatric symptoms, or when such symptoms are not present or dominant, reduction of other presenting problem behavior or experience, especially in relation to interactional patterns that maintain the problem(s); (2) increased family and couple resourcefulness, for example, improved communication, problem-solving and conflict resolution skill, enhanced coping skill and adaptability; (3) improvement in the fulfillment of individual psychological needs for attachment, cohesion, and intimacy; increased trust and equitability; enhanced capacity to foster the development of individual family members; (4) increased family ability to interact effectively with important larger social systems; and (5) increased awareness and understanding of how couples' and families' patterns of interaction influence the group's everyday effectiveness in living as well as how such patterns affect and are affected by the psychological health and satisfaction of individuals.

Within some influential approaches to family therapy (e.g., Bowen family systems, contextual, symbolic experiential), certain goals are considered virtually universal, that is they are relevant in all cases regardless of differences between families or couples. In several other widely practiced approaches (e.g., Mental Research Institute brief therapy, strategic, narrative, solution-focused) such universal treatment aims are almost never invoked, as each clinical case is viewed as a sort of unique experiment. In the former case symptoms and presenting problems are often seen as signs of and metaphors for dysfunction elsewhere in the family system whereas in the latter such difficulties and concerns tend to be taken more or less at face value.

The treatment goals largely involve the *ultimate goals* of treatment, that is, desired end-states. In addition, a variety of *mediating goals* arise in the various family therapy approaches. *Mediating* goals are of a shorter-term nature and include change in psychological processes that an individual, couple, or family must undergo in order to reach their own or their therapist's treatment objectives. These are sometimes referred to as process goals. Common forms of such mediating or process goals are achievement of interactional (dyadic) or genetic (multigenerational) insight; the teaching of various interpersonal skills, for example, communication and problem-solving; and the disruption of interlocking pathologies or blocking of rigid symptom- or problem-maintaining patterns of behavior to allow opportunities to experiment with more adaptive responses. Mediating goals may also be more abstract and are not necessarily made explicit to families by the therapist.

Finally, it is important to appreciate that treatment goals are not always explicitly negotiated in family or other methods of therapy. Mediating goals in a wide variety of approaches are particularly unlikely to be discussed between the family and therapist, and ultimate goals also vary in the degree of overt, concrete discussion they receive across the more than one dozen influential methods of family and couple therapy.

Techniques of Family and Couple Therapy The techniques of family and couple therapy cast at least as broad a net as those of individual psychotherapy. While there are well over a dozen discernibly separate schools of family and couples therapy (Table 30.5-1), there is a good deal of overlap within several subgroups of them because of some common origins (e.g., Mental Research Institute brief therapy, strategic therapy, solution-focused, Milan systemic). Moreover, since the 1980s there has been a strong movement within the field toward combining elements of different methods, which has led to the increased borrowing of techniques across scholastic lines. Some of this borrowing has been in the form of technical eclecticism, that is using techniques presumed to be relevant and effective without regard to the originating theories' basic assumptions or the contradictions therein contained. Other borrowing has grown out of the search for the so-called common ingredients of effective therapy, and has shown considerable attention to matters of conceptual clarity and coherence. In addition, the general practice of family and couple therapy has become increasingly comprehensive and less doctrinaire, for example, in the use of individual therapy plus couple therapy, couple therapy plus (child-focused) family therapy, and other methods. Moreover, the field's early disdain for psychiatric and psychodiagnostic perspectives and practices has largely eroded as clinicians increasingly coordinate the use of psychopharmacological agents with flexible psychosocial treatment plans.

Therapeutic Techniques A representative sample of treatment techniques in the field of family and couple therapy is shown in Table 30.5-2.

Techniques Focused on In-Session Experience (Orientation)	Techniques Focused on Out-of-Session Experience (Orientation)
Circular questioning (Milan systemic)	Communication training (behavioral)
Communication training (behavioral)	Family management skills (psychoeducational)
Family sculpting (experiential)	Letters, therapeutic (narrative)
Family reconstruction (Satir model)	Ordeal prescription (strategic)
Genogram construction (Bowen family systems)	Paradoxical injunctions (Mental Research Institute)
Interpretation of unconscious processes (psychoanalytic)	Parent management training (behavioral)
Positive connotation (Milan systemic)	Problem-solving training (behavioral)
Problem-solving training (behavioral)	Pretending, prescription of (strategic)
Reframing (strategic)	Rituals, prescription of (Milan systemic) (e.g., invariant prescription)
Scaling, miracle, and exceptions questions (solution-focused)	Task assignment (behavioral, structural)
Teaching the emotional functioning of systems (Bowen family systems)	

Table 30.5-2 Representative Treatment Techniques Across the Range of Family-Couple Therapies

Therapeutic techniques are created and used to draw upon a theory's presumed mechanisms of change, that is, those causal processes by which clinically relevant change occurs. In addition, each therapy approach refers to a *dominant site of change,* which is a biopsychosocial space in which the clinically most salient change is believed to occur.

Table 30.5-2 reveals that behavior change is the dominant mode of change induction in family and couples therapy in contrast to insight and reflection. *Behavior-change techniques* refer to any therapeutic techniques that are called upon to modify observable behavior, whether at the level of the individual, the dyad, or the whole family whereas insight-oriented techniques refer to those that lead to change in awareness or perhaps affective experience without any automatic change in overt behavior. In contrast to much traditional individual psychotherapy in which insight is generally assumed to precede therapeutic change, in most family and couples therapy the opposite sequence is generally preferred. In addition, family therapists are usually more bi-directional in their thinking, that is, they believe that change can be initiated in any domain of psychosocial organization. For pragmatic reasons, however, initial change is more often sought at the interactional, public level of experience.

Table 30.5-2 distinguishes between family and couples therapy techniques that focus on in-session versus out-of-session experience. The wide use of techniques that emphasize patients' experiences away from the consultation room reflects family therapists' respect for the healing power of families and their understanding that

change that endures and generalizes to everyday life is not achieved primarily in the substitutive relationship between therapists and their patients, but between and among family members in their natural environment. What is especially striking about the centrality of out-of-session techniques in family and couple therapy is that it also reflects the modal family therapist's view that the dominant site of action in therapy change is within the family or couple relationship.

Therapeutic techniques in family and couples therapy are dominated by those focused on cognitive dimensions of experience, such as meaning and attribution, and those that focus on action. The former, the more cognitive of these techniques, may emphasize a therapist's attempts to change meaning, to discover meaning, or to cocreate meaning. Such efforts, can range from the cognitive-behavioral therapist's attempts to persuade family members that their partners general inexpressiveness reflects not a lack of love but rather internal discomfort with intimate conversation to the strategic therapist's positive reframing of such inexpressiveness as an understandable attempt to maintain a tolerable level of affective arousal in a marriage to a highly expressive mate, even with the unfortunate self-sacrifice that it requires. Some meaning-oriented interventions in family and couple therapy assume that the therapist's meaning is “correct” and that these interventions reflect a knowable reality and psychological truth (e.g., behavioral, Bowen family systems, psychoanalytic) whereas others are markedly different and believe that since there is no knowable external reality, all of therapy involves the making of meanings (“co-construction of reality”) rather than their discovery (e.g., Narrative, Solution-Focused, Strategic). For these latter approaches truth is pragmatic, that is it is a meaning or explanatory framework that leads to clinically relevant change but is not presumed to reflect any knowable truth.

Action-oriented techniques can be further meaningfully divided into those that assume that family members already have the requisite behaviors in their repertoire, and those that assume that family members presently lack such skills or knowledge. In learning theory terms, this is the important distinction between problems of performance and problems of acquisition. Action-oriented techniques involve either therapeutic directives or skill training. Directives can involve either in-session or out-of-session (often referred to as “homework tasks”) actions, for example, asking a father to not allow his 8-year-old daughter to divert his attention when he seems uncomfortable with his wife's assertiveness toward him, or assigning the couple the task of seeing to it that their noncompliant child goes to bed at an appropriate predetermined time so that they are guaranteed “couple time” together several evenings per week. The same couple may also need specific training in communication and problem-solving skills in order to carry out the out-of-therapy assignment.

Don and Kris, the 44-year-old parents of three children ages 19, 17, and 12, began contact with the therapist during Don's week-long hospitalization because of his intensifying depression and increased use of anxiolytics. Although he had had two brief psychiatric hospitalizations in his 20s and early 30s Don had generally functioned quite well for several years. But over the preceding several months, he had become easily fatigued, generally pessimistic and irritable, and “clingy.”

These symptoms, among others, seemed to have arisen as Kris became increasingly less involved with Don's life. Kris, the eldest of three siblings, had a long history of self-sacrifice, originating in her family of origin, in which she was the “parentified child” of an alcoholic father and a passive, depression-prone mother. Although her father's alcohol problem had largely abated over the previous decade, the marriage between Kris' parents had never progressed much beyond becoming a “pseudo-civil, polite but distant” relationship. They had retired to the Sun Belt, hoping the change would allow them to “start over,” but Kris' mother had become increasingly depressed of late as marital conflicts became more frequent.

Kris had recently become extremely involved in her parents' difficulties, especially as her mother's confidante, and less involved in her own family's life. In addition to this loss, Don, a generally avoidant man, felt more and more isolated from meaningful relationships, now that his 19-year-old son was away at college, and his 17-year-old son was “never around.” Don was very dependent upon and angry at Kris, but feared voicing his concerns about their disconnected lives. This dilemma was reminiscent of his relationship to his father, who had essentially abandoned the family when Don was 7 years old.

Over the next 5 months couples therapy addressed problematic patterns at several levels. Structural interventions included dislodging Kris from her therapist-like role with her parents, in part by helping her to help her parents identify a skilled couples therapist for them to see in their new location and by supporting her in setting and keeping appropriate boundaries in her conversations with her mother.

Cognitively oriented work on Kris' guilt about putting her own needs higher on her priority list, for example, for connectedness to her husband and children, occurred in the sessions, aided by bibliotherapy and other homework assignments. Don's normal adult attachment needs were validated, and he was also encouraged to explore new peer relationships in his community. A good deal of therapy time was devoted to communication and problem-solving training. In this, Don had to address his anxiety about assertiveness and fear of conflict and abandonment, and Kris had to address her anxiety-guilt about “saying no” and expressing her own needs.

Clinical Issues

Indications There are two overarching sets of indications for family and couple therapy: those that involve the therapist's mindset and those based on empirical validation or support. With regard to the former, it has been said that the indications for family therapy are to be found in the therapist. That is, just as behavior therapy is not equivalent to the practice of systematic desensitization, family therapy is not merely a collection of interpersonal techniques. It is a therapeutic collage of ideas about both family and individual stability and change, the evaluation and maintenance of psychopathology and problems-in-living, and relational ethics. Family therapy might better be called systemically sensitive therapy and in this sense reflects a basic world view as much as clinical treatment methodology. For therapists thus inclined, all clinical problems involve salient interactional components and so some kind of family or interpersonally functionally significant others' involvement in therapy is always called for even in the systemically aware treatment of an individual.

Although there are some likely contraindications to family and couple therapy, there is now an impressive array of common clinical disorders and problems for which various family or couple treatment methods have been empirically demonstrated to be effective. In a few cases, family interventions are probably even the treatment of choice. These disorders span both adult and child adolescent difficulties.

ADULT DISORDERS *Schizophrenia* Research has demonstrated unequivocally the effectiveness of a range of psychoeducational and behavioral treatment methods that teach family members about schizophrenia, and emphasize the reduction of expressed emotion and the enhancement of skills for coping with stress. These psychosocial methods are used in conjunction with medication. Some of them include multiple-family group therapy and all include interpersonal skills training. The inclusion of even a few family sessions often has powerful posthospitalization effects on both symptoms and recidivism, with rehospitalization rates regularly reduced by over 50 percent. Moreover, these psychoeducational approaches appear to be enormously cost-effective as compared to standard care. This body of evidence supports a handful of particular family intervention models, and the practice of dramatic, affect-arousing, and symbolically interpretive styles of work with these families, once fashionable in some influential quarters of the family field, is no longer empirically warranted.

Mood Disorders Psychoeducational methods have recently shown promise in the family treatment of bipolar I disorder as well as schizophrenia. Among the adult mood disorders, however, major depressive disorder has been the most responsive to certain systems-oriented therapeutic approaches. Specifically, behavioral marital therapy has been compared to individual cognitive therapy, the individual therapy treatment of choice for major depressive disorder. Both therapies lead to significant improvement in depression, but only behavioral marital therapy leads to an increase in marital satisfaction, attributable to its communication-skill and problem-solving skill components. It has been found that (1) couple therapy is not an effective treatment for depressed persons whose marriages are not in serious distress; (2) treating major depressive disorder with individual therapy does not help to resolve marital difficulties; and (3) individual therapy for depressed women who see their marital conflict leading to their depression may actually increase marital discord. Given the central role of relationship problems in predicting the relapse of depression in married women, the research offers strong support for behavioral marital therapy for women who are both depressed and maritally distressed.

Substance Abuse Earlier major outcome studies of structural family therapy for opiate addicts showed that this approach retained the users and their families in treatment longer than standard interventions, led to a decrease in drug use and related behavior problems, and increased social functioning. More recent studies favor family therapy over individual therapy and show family therapy to be a cost-effective adjunct to methadone maintenance. In the realm of adult alcoholism, a number of models of family involvement have been shown to have strong positive effects in terms of helping to motivate alcoholic individuals to enter treatment. On the other hand, the long-term outcomes (over a year) for the main rehabilitation phase have been only marginally more positive than individual alcoholism treatment (which itself is not especially effective).

Anxiety Disorders The inclusion of the spouse of the patient with agoraphobia in therapy sessions, often called spouse-involved treatment (not marital therapy), appears to lead consistently to lower dropout rates and more reliable practicing of exposure. In addition, spouse involvement in cases that are accompanied by marital distress may help to reduce marital tension, which in turn may make symptomatic spouses less anxious overall.

Marital Discord Troubled marriages (there have been no treatment studies of dissatisfied nonmarried heterosexual or same-sex couples) not only bode poorly for recovery from certain individual psychiatric disorders, but are also emotionally painful and stress-inducing in themselves. Since as many as 40 to 50 percent of the reasons for adults seeking mental health services may involve marital dissatisfaction, marital conflict constitutes a significant public health concern, despite the psychiatric status of marital discord as a mere other condition that may be a focus of clinical attention. There is compelling evidence that conjoint methods of marital therapy are the only validated approaches to date, and no individually oriented treatment methods have been shown to be effective. Thus far, three distinct methods have received either promising or very substantial empirical support. Emotionally focused therapy, grounded in attachment theory and espousing an experiential style, has shown considerable promise in several studies. Insight-oriented marital therapy, a psychodynamic method that emphasizes dealing with conflictual emotional processes both within and between marital partners, has been found to be very helpful at both the conclusion of therapy and at long-term follow-up. Finally, Behavioral Marital Therapy is by far the most frequently investigated couple therapy approach. Findings from about three dozen studies reveal it to have the largest effect size of all couple treatments. The central change-promoting treatment components of this method are behavioral exchange and communication and problem-solving skills training, recently enhanced by attribution-shifting and empathy-based acceptance training.

Sexual Dysfunction Although sex therapists generally do not travel in the same professional circles as family and couple therapists, there is every reason to consider sex therapy as a bona fide systemic therapy approach because of its conjoint nature. Drawing upon learning theory principles used in behavioral exposure therapies and communication training approaches, among others, conjoint couples sex therapy is well established as the treatment of choice for such common sexual dysfunction as premature ejaculation, vaginismus, primary female orgasmic disorder, and secondary erectile dysfunction. On the other hand, ejaculatory inhibition, primary erectile disorder, secondary orgasmic dysfunction, and problems of inhibited sexual desire are not so effectively dealt with by conjoint therapy.

CHILD AND ADOLESCENT DISORDERS ***Childhood Conduct Disorders*** Behavioral Parent Training (also known as parent management training or social learning family therapy) has a long history of demonstrated effectiveness in helping parents deal with aggressive and oppositional children and preadolescents. The intervention model (which typically excludes the child from treatment sessions) places central emphasis on learning–theory-based contingency management procedures, in which parents are taught essential parenting skills for providing appropriate consequences for desirable and undesirable child behaviors. In addition, effective parent training appears to lead to increased marital satisfaction (which in turn probably leads to greater parental unity, which leads to better enforcement of consequences).

Delinquency and Substance Abuse Three types of family therapy have been shown to be as effective as and at times more effective than standard care or individual psychotherapy in the treatment of acting-out adolescents, including those involved in drug abuse. These therapies are Functional Family Therapy, a variant of Behavioral Family Therapy; multisystemic therapy, an approach that targets an array of social systems (e.g., schools, per groups) in addition to the nuclear family; and a modified version of structural therapy. These highly structured treatments actively engage clinically relevant systems, enhance communication skills, and set clear behavioral and generational limits. They consistently yield positive clinical changes in recidivism, substance use, and impulse-control disorders, in addition to fostering higher rates of case retention than traditional forms of mental health intervention.

Other Indications In addition to the empirically justified indication for one or more forms of family and couple therapy, systems-oriented therapy is also commonplace in many other clinical situations. Dealing with the process of marital separation and divorce, addressing the complexities of remarriage and stepfamily life, crises of mid-life adaptation, extramarital affairs, crosscultural transitions, and chronic nonpsychiatric conflict constitute a large share of a family therapist's clinical practice, with or without an attendant diagnosable psychiatric disorder in one or more family members.

Contraindications There are no absolute contraindications to the use of family therapy, although clinical experience suggests caution in certain common circumstances. A therapist's lack of training in family therapy and a disinclination to view psychopathology and psychotherapy interpersonally and systemically are probable contraindications. Training in individual psychotherapy alone does not prepare adequate conceptual and technical therapist skills for engaging in family and couples therapy. Moreover, a conceptual bias toward intrapsychic experience may lead to a reductionistic, noninteractional focus and perhaps inevitably to a side-taking and blaming posture.

Some patient characteristics may be relative contraindications to family and couple therapy. Patients who refuse to divulge material essential to the harmony and smooth functioning of marriage and family living may thereby impose artificial limits on the success of treatment. A serious lack of commitment to a marital or couple relationship is probably a counterindication for systems therapy. A separated or divorced marital status does not preclude systems-oriented intervention, but directs it toward a style and focus of therapy attuned to the couple's relationship status, that is, one in which issues of individual adjustment and development may be as central as those of the family unit. In addition, the presence of ongoing family violence or sexual abuse is believed by many practitioners to preclude conjoint family or couple therapy because of the possibility of uncontrollable affective arousal during and especially outside of treatment sessions. Moreover, the courts typically remove alleged offenders from the family home, and legally allowed contact between the offender and the alleged victim may be limited to meetings with mental health professionals. Such common restrictions usually preclude clinically meaningful change in family structural-interactional terms, as the emphasis is often put on preventing continuing abuse, healing from the trauma of the disclosure of the relational betrayal, and restructuring the family-minus-offender in terms of its everyday functioning. Offenders are usually ordered into individual or specialized offender-group therapy and are regularly reintegrated into the family therapy much later on.

Finally, certain patient characteristics that have traditionally been seen as relative contraindications for uncovering individual psychotherapy seem not especially salient in the family therapy context, that is, age, intelligence, verbal ability, and psychological-mindedness, unless the therapy itself involves a long-term insight-oriented approach.

Limitations The main limitation in the practice of family and couple therapy is that systemically oriented treatment is not the sole treatment of choice for certain kinds of psychiatric and psychological disorders and problems, although it may deserve to play an adjunctive role. For some problems alternative treatments are clearly called for and are reliably effective even without family involvement, for example, individual behavioral treatment of many adult anxiety disorders.

Another limitation on family and couple therapy is involved when one or more family members resist genuine involvement, for example out of fear of being blamed for family difficulties or being exposed in some manner. The skilled family therapist must come prepared to cope with such possibilities.

Complications and Ethical Issues Complications in the practice of family therapy arise most often in terms of maintaining confidentiality, obtaining adequate diagnostic information, and receiving payment for professional services.

In conjoint family therapy confidentiality between or among family members does not exist. When individual family members seek to reveal to the therapist outside the conjoint sessions material that is of a confidential nature, anti-therapeutic side-taking could occur. Moreover, unlike therapists, family members are in no way legally bound to maintain confidentiality about other family members with people outside the family. These concerns, although probably not widespread, may certainly limit some patients' therapeutic involvement and raise sensitive strategic dilemmas for therapists.

Since a great many family and couples therapists focus mostly on those aspects of family life that seem essential for resolving presenting problems, and because most family and couple therapy is of relatively short duration, the therapist may not become aware of individuals' psychiatric histories, parts of which could be pertinent to the current course of treatment. Even in the most problem-centered, present-centered of treatments, it is always appropriate to carry out a brief exploration of potentially relevant psychiatric material, including the symptom or problem, and treatment histories of family members other than the identified patient or the member in whose name professional fees are submitted for payment. Doing so is not inherently inconsistent with interactionally focused treatment because such information, especially if it refers to current conditions, may set limits on the amount or rate of change possible via an interactionally based treatment approach.

A potential complication encountered with unfortunate frequency in family therapy is that what may be reasonably considered an appropriate treatment, or even the treatment of choice, may not allow therapist reimbursement or may require therapists to force-fit standard diagnostic criteria to one or more family members. Ironically, although the ascendance of managed-care–based health delivery systems has generally constrained psychotherapists' decision-making options, some managed-care arrangements, for example, the closed panel health maintenance organization, have actually expanded family therapists' treatment planning alternatives in that the professional service providers are the insurers and because managed-care organizations favor clinical methods that are problem focused and relatively brief.

Ethically, family therapists seem to face two major recurrent issues. The first involves the dilemma of whether to adhere strictly to “the systems perspective” even when the clinical situation suggests that a more linear, noninteractional point of view is not only arguable, but perhaps even the only ethical one to adopt in terms of likely clinical effectiveness. The best interests of patients must always supercede the best interests of therapists' conceptual frameworks.

There is an inherent tension in many family and couple therapy cases between balancing the best interests of individual family members with those of the group as a whole. For example, helping a family minimize its pain in the process of a divorce hardly fosters family cohesion, although it may reduce individual suffering and foster individual growth on certain members' parts while simultaneously increasing the stress felt by other individuals. Even when major overhauling of the basic family structure is not at issue, it is regularly a challenge of some proportion to therapists who must seem to support the treatment goals of two (or more) family members whose individual treatment goals appear to be largely antithetical, for example, a wife wants her husband to assume more household responsibilities as she returns to school or the workforce while her husband insists on maintaining his “freedom” to spend a significant amount of time in activities outside the home, arguing that they help him “relax” in the face of very stressful job or career conditions.

Emerging Issues and Trends Family therapy began to a significant degree as a clinical and conceptual movement in rebellion against the dominant belief systems in the mental health professions, as was true of behavior therapy. As the twentieth century closes, both these iconoclastic clinical forces have become assimilated into the mainstream of mental health treatment and training. Most family therapies have moved away from earlier radical systemic positions that endorsed allowing in multiple vantage points but arbitrarily excluded certain views. Thus, adopting a nonlinear explanatory framework, for example, is no longer believed to be antithetical to attributing and accepting responsibility for one's own actions, as in the case of marital and family violence.

As the concepts and practices of family therapy have penetrated the traditional mental health professions, more integrative therapies have evolved. These integrative approaches draw together varied theoretical orientations (e.g., psychodynamic and behavioral) and foster combined clinical practice at several levels of biopsychosocial organization (e.g., individual, family, larger systems). Just as the biological and the psychosocial are being addressed together, so is there a trend toward returning the psyche to the system by attending to levels of organization within the individual in concert with levels of multiperson systems.

There is also an increasing push toward developing and refining more specific variants of family and couple treatment for application to specific kinds of cases, such as particular diagnostic groups (e.g., psychoeducational programs for schizophrenia, cognitive-behavioral couple therapy for depression, and multisystemic treatment for adolescent delinquency). As family therapy has become a mainstay of the dominant mental health professions, it has adopted some of the world views and operations of those domains, including the operationalization and manualization of therapies. Although this trend may appear to some observers to narrow the scope and purview of family therapy, it has moved clinical theorizing in practice in quite the opposite, broadening, direction. Thus, gender and culture have received increasing attention as essential forces to be attended to in clinical practice. Contrary to the uniformity myth identified in the late 1960s by psychotherapy researchers, contemporary family therapy does not assume that the same treatment will be appropriate across different patient groups.

Finally, although certain commonly used family therapy approaches have shown very impressive treatment results, other historically influential approaches lag far behind in demonstrating rather than merely asserting their effectiveness. Thus, there is virtually no empirical evidence of the effectiveness of Bowen family systems therapy, symbolic-experiential therapy, Milan systemic therapy, Mental Research Institute brief therapy, or contextual therapy. Similarly, some recently popular approaches have yet to amass any compelling evidence of positive treatment outcome, for example, narrative therapy and solution-focused therapy. The family and couples approaches that have presented a reasonable body of empirical support for their methods (e.g., cognitive-behavioral marital therapy, psychoeducational family therapy, functional family therapy, emotionally focused couple therapy, behavioral parent training, insight-oriented marital therapy, and structural therapy) have evolved largely within academic settings and are thus more connected to an empirical, hypothesis-testing tradition. Nonetheless, other influential approaches are not likely to be sustained or to expand their influence if they remain too far outside the mainstream of mental health practice and do not become accountable for their effectiveness.

SUGGESTED CROSS-REFERENCES

Psychosocial treatments for mood disorders are presented in Section 14.9. Psychosocial treatment of schizophrenia is discussed in Section 12.9. Substance abuse and substance-related disorders are discussed in Chapter 11. Relational problems are covered in Chapter 26. Sexual dysfunctions are presented in Section 19.1a. Behavior therapy is covered in Section 30.2 and cognitive therapy in Section 30.6. Family therapy with children is further discussed in Section 48.5 and interventions and consultations with families of the elderly in Section 51.4k.

SECTION REFERENCES

Alexander J, Parsons BV: *Functional Family Therapy*. Brooks/Cole, Pacific Grove, CA, 1982.

Anderson CM, Reiss D, Hogarty B: *Schizophrenia and the Family*. Guilford, New York, 1986.

Anderson H: *Conversation, Language and Possibilities: A Postmodern Approach to Therapy*. Basic Books, New York, 1997.

Boszormenyi-Nagy I, Spark GM: *Invisible Legalities: Reciprocity in Intergenerational Family Therapy*. Harper & Row, New York, 1973.

Broderick CB, Schrader SS: The history of professional marriage and family therapy. In *Handbook of Family Therapy*, ed 2, AS Gurman, DP Kniskern, editors. Brunner/Mazel, New York, 1991.

Bowen M: *Family Therapy in Clinical Practice*. Aronson, Northvale, NJ, 1978.

Carlson J, Sperry L: *The Disordered Couple*. Brunner/Mazel, New York, 1998.

*Christenson A, Heavey CL: Interventions for couples. Ann Rev Psychol *50:*165, 1999.

Dattilio F: *Case Studies in Couple and Family Therapy: Systemic and Cognitive Perspectives*. Guilford, New York, 1998.

de Shazer S: *Keys to Solutions in Brief Therapy*. Norton, New York, 1985.

*Fincham F, Beach S: Conflict in marriage: Implications for working with couples. Ann Rev Psychol *50:*47, 1999.

Framo JL: *Family-of-Origin Therapy: An Intergenerational Approach*. Brunner/Mazel, New York, 1992.

Gottman JM: Psychology and the study of marital processes. Ann Rev Psychol *49:*169, 1998.

Greenberg LS, Johnson SM: *Emotionally Focused Therapy for Couples*. Guilford, New York, 1988.

Gurman AS: Integrative Marital Therapy: Toward the Development of an Interpersonal Approach. In *Forms of Brief Therapy,* S Budman, editor. Guilford, New York, 1981.

*Gurman AS, Kniskern DP, editors: *Handbook of Family Therapy*. Brunner/Mazel, New York, 1991.

*Gurman AS, Kniskern DP, Pinsof WM: Research on the process and outcome of marital and family therapy. In *Handbook of Psychotherapy and Behavior Change*, ed 3, SL Garfield, AE Bergin, editors. Wiley, New York, 1986.

Haley J: *Problem-solving Therapy*. Harper & Row, New York, 1976.

Haley J: *Leaving Home: The Therapy of Disturbed Young People*, ed 2. Brunner/Mazel, New York, 1997.

Jacobson NS, Gurman AS: *Clinical Handbook of Couple Therapy*. Guilford, New York, 1995.

Kaslow NJ, Celano MP: The family therapies. In *Essential Psychotherapies*, AS Gurman, SB Messer, editors. Guilford, New York, 1995.

*Lebow JL, Gurman AS: Research assessing couple and family therapy. Ann Rev Psychol *46:*27, 1995.

Minuchin S: *Families and Family Therapy*. Harvard University Press, Cambridge, MA, 1974.

Neill J, Kniskern DP, editors: *From Psyche to System: The Evolving Therapy of Carl Whitaker* . Guilford, New York, 1982.

Nichols MP, Schwartz RC: *Family Therapy: Concepts and Methods*, ed 3. Allyn & Bacon, Boston, 1995.

Patterson GR: *Families: Application of Social Learning Theory to Family Life*. Research Press, Champaign, IL, 1975.

*Patterson J, Williams L, Grauf-Grounds C, Chamow L: *Essential Skills in Family Therapy*. Guilford, New York, 1998.

Pinsof WM: *Integrative Problem-Centered Therapy*. Basic Books, New York, 1995.

Pinsof WM, Wynne LC, editors: Special Issue—The Effectiveness of Marital and Family Therapy. J Marital Fam Ther *21:*1, 1995.

Roberts LG: *Transgenerational Family Therapies*. Guilford, New York, 1992.

Satir VM: *Conjoint Family Therapy*. Science and Behavior Books, Palo Alto, CA, 1964.

Scharff D, Scharff JS: *Object Relations Family Therapy*. Aronson, Northvale, NJ, 1987.

Selvini-Palazzoli M, Boscolo L, Cecchin GF, Prata G: *Paradox and Counterparadox*. Aronson, New York, 1978.

Stanton MD, Shadish WR: Outcome, attrition, and family-couples treatment for drug abuse: A meta-analysis and review of the controlled, comparative studies. Psychol Bull *122:*170, 1997.

Walsh F, editor: *Normal Family Processes,* ed 2. Guilford, New York, 1993.

Watzlawick P, Weakland JH, Fisch R: *Change: Principles of Problem Formation and Problem Resolution*. Norton, New York, 1974.

White M, Epston D: *Narrative Means to Therapeutic Ends*. Norton, New York, 1990.

Wynne LC, editor: *The State of the Art in Family Therapy Research: Controversies and Recommendations*. Family Process Press, New York, 1988.

30.6 COGNITIVE THERAPY

A. JOHN RUSH, M.D., AND AARON T. BECK, M.D.

Cognitive therapy is based on a theory of psychopathology that draws on concepts from cognitive and social psychology as well as from information-processing and psychoanalytic theory. Therapeutic principles and techniques are derived partly from both these theoretical frameworks and from clinical empirical investigations. Cognitive theory posits that the way individuals interpret experience determines how they feel and behave. For example, if persons view situations as dangerous, they experience anxiety and want to escape. The four basic emotions of sadness, elation, anxiety, and anger, respectively, are evoked by perceptions of loss, gain, danger, and wrongdoing by others.

Biased information processing causes distorted interpretations in the form of verbal or pictorial cognitions as well as cognitive deficits. Examples of such biases include incorrectly evaluating specific life situations as hostile or dangerous. Cognitive deficits are represented by failing to generate adequate strategies to solve problems.

Specific psychiatric disorders are associated with characteristic cognitive biases that result from and contribute to impaired corrective function of higher-level cognitive processes. In depression, for example, patients see themselves, their experiences, and their futures in negatively biased ways, which in turn sustains or magnifies depressive symptomatology.

Cognitive therapy employs specific treatment strategies to correct these habitual thinking errors found in different psychopathological states. Treatment employs a combination of verbal interventions and behavior modification techniques designed to help patients identify their dysfunctional cognitions, test whether they are based on logic and reality, and correct the distorted conceptualizations and dysfunctional beliefs that underlie their dysfunctional thinking. A wide range of therapeutic techniques consistent with the cognitive model of psychopathology may be used. The therapy itself is active, structured, time limited, and focused on current problems. As patients begin to think and act more realistically, their symptoms and behavior improve.

Cognitive therapy has been used successfully in the treatment of depressive disorders, anxiety disorders (e.g., phobias, panic disorder, obsessive-compulsive disorder) eating disorders, substance abuse, and personality disorders; as well as a number of psychophysiological disorders. Cognitive therapy may be used in individual, group, couple, and family formats. Recently, adaptations for treatment of the elderly, depressed children and adolescents, and patients with bipolar I disorder or schizophrenia have been prepared. Finally, cognitive techniques have been specified and evaluated to increase patients' compliance to medication regimens.

HISTORY

The phenomenological approach to psychology, psychoanalytic psychology, and cognitive psychology has contributed to the theoretical foundations of cognitive therapy as formulated by Aaron T. Beck. The concept that individuals' views of themselves and their personal world determine behavior is derived originally from the Greek stoic philosophers. The philosophical emphasis on conscious subjective experience stems from the works of Immanuel Kant, Martin Heidegger, and Edmund Husserl. The theories of these phenomenologists were made operational by psychologists such as Alfred Adler, Otto Rank, and Karen Horney, who were influenced by them. Many of the basic theoretical or therapeutic concepts of cognitive therapy can be found in their writings. Adler, for example, emphasized both the importance of understanding patients within the framework of their own conscious experience and the need to determine in therapy the conscious meaning that patients attach to their experiences.

The structural theory and depth psychology of Sigmund Freud contributed the theory of hierarchical structuring of cognition into primary and secondary processes and the concept that symptoms and affect are based upon “pathogenic ideas.” The major contributions of cognitive psychologists include George Kelly's concept of “personal constructs” as a determinant of behavior and the cognitive theories of emotion formulated by Magda Arnold and Richard Lazarus.

The emphasis of cognitive therapy on finding solutions to conscious, definable, here-and-now problems has been paralleled by rational-emotive therapy, developed by Albert Ellis. The therapeutic style of gentle questioning and unconditional acceptance owes much to the client-oriented therapy of Carl Rogers. The procedure of developing common themes across the patient's emotional reactions to diverse experiences is related to the neopsychoanalytic approach of Leon Saul. The behavior therapy movement contributed several crucial therapeutic strategies: the structuring of the interview, preparing agendas for individual sessions, setting goals, eliciting feedback, operationalizing the procedures, testing hypotheses, formulating and testing problem-solving techniques, and assigning homework. Finally, the empirical thrust of cognitive therapy has been influenced by the work of cognitively oriented writers such as Albert Bandura, Jerome Frank, Marvin Goldfried, Arnold Lazarus, Michael Mahoney, Donald Meichenbaum, and G. Terence Wilson.

MODEL OF PSYCHOPATHOLOGY

Relation of Cognition to Emotions The basic cognitive themes can be derived from the examination of four basic emotions. Thus, sadness is evoked when there is a perception of *loss*—a defeat or deprivation—frequently in the form of unfulfilled or disconfirmed positive expectations or the disruption of a valued personal relationship. The usual consequence is to withdraw the investment in the particular source of disappointment. Elation, in contrast, follows from a perceived gain and thus is likely to reinforce activity toward achieving a goal.

In contrast to sadness and gladness, which relate to positive goals, anxiety and anger are usually responses to threats to the self or valued attachments. Anxiety is evoked when persons are concerned about their vulnerability—for example, to being hurt, killed, or devalued—and are consequently impelled to withdraw or to appease. Anger focuses on the offensive qualities of the threat, with the behavioral consequence of aggressive self-defense or counterattack.

Cognition and Psychiatric Disorders The cognitive and emotional concomitants of the psychopathological syndromes can be viewed as exaggerated, persistent forms of the kind of normal emotional and behavioral reactions described above. Thus, in depression, sadness and withdrawal of interest in previous goals are intensified and prolonged beyond the usual sense of defeat or deprivation of normal reactions. In anxiety disorders, a generalized and intensified sense of vulnerability to a wide variety of threats stimulates anxiety and avoidance.

Each of the disorders is characterized by idiosyncratic beliefs about the self or others that distinguish them from the other disorders. Panic disorder, for example, is characterized by a catastrophic misinterpretation of bodily sensations or mental experiences. In obsessions the content revolves around a warning or doubt: the compulsion is prompted by an urge to dispel the danger through a repetitive act (Table 30.6-1).

Disorder	Idiosyncratic Cognitive Content
Depressive disorders	Negative view of self, experience, and future
Hypomanic or manic episodes	Inflated view of self, experience, and future
Suicidal behavior	Hopelessness and debased self-concept
Generalized anxiety disorder	Fear of physical or psychological danger
Phobia	Fear of danger in specific, avoidable situations
Panic disorder	Fear of immediate physical or mental disaster
Paranoid state	View of others as biased, manipulative, and devious
Conversion disorder	Concept of motor or sensory abnormality
Obsessive-compulsive disorder	Continual thoughts about safety; repetitive acts to ward off threat
Anorexia nervosa or bulimia nervosa	Fear of being fat and unattractive
Hypochondriasis	Concern about serious insidious medical disorder

Table 30.6-1 Cognitive Profile of Psychiatric Disorders

In delusional disorder, the perception of being mistreated is generalized and leads to anger and an inclination to counterattack. The specific cognitive profile of each disorder provides the framework for specific cognitive interventions.

Each of the personality disorders is characterized by a set of dysfunctional beliefs. The dependent personality, for example, is driven by beliefs such as “I cannot function without somebody to lean on.” The avoidant personality is motivated by beliefs such as “If people are aware of my performance, they will be critical and rejecting.” The beliefs of the histrionic personality disorder center around the theme, “I need to entertain people in order to be accepted.”

Cognitive Processing Both normal and pathological emotions are mediated by primitive cognitive processes (analogous to Freud's primary processing). These processes support conceptualization of situations that are global and relatively crude, while higher levels of cognitive processing (analogous to Freud's concept of secondary process) are more refined and, when the latter function properly, provide reality testing and correction of the global, primal conceptualizations. In psychopathological states, however, these higher-level cognitive functions are impaired and lead to extreme levels of anger, anxiety, elation, or sadness in an ever-widening range of inappropriate situations. That is, malfunctioning secondary process thinking leaves primary responses unmodulated in full-blown psychiatric disorders.

Although these pathological and psychological processes clarify the role of cognitive mechanisms in the development of depression, anxiety, and other disorders, they do not address the issue of causality. The ultimate causes of these disorders are not necessarily found in the cognitive structures or processes even though cognitive therapy may be an effective intervention. Causes are best understood in terms of the interaction of innate biological, developmental, and environmental factors.

In the various disorders certain dysfunctional beliefs relevant to helplessness, danger, and unlovability play an important role in information processing. Embedded in cognitive structures *(schemas),* these dysfunctional beliefs mold the individual's thinking and contribute to the cognitive errors typical of psychopathology.

Logical Errors in Thinking *Arbitrary inference* is the process of drawing a conclusion in the absence of evidence supporting it or even in the face of evidence to the contrary. *Selective abstraction* refers to focusing on a detail taken out of context, ignoring other, more salient, features of the situation, and conceptualizing the whole experience on the basis of this element. *Overgeneralization* refers to drawing a general conclusion across all situations on the basis of a single incident. *Magnification* and *minimization* represent distorted evaluations of the relative importance of particular events. *Personalization* describes the tendency to relate external events to oneself when there is no basis for making such a connection. *Dichotomous thinking* occurs when the patient tends to make interpretations on an either-or, black-or-white basis (Table 30.6-2).

Arbitrary inference	Selective abstraction
Overgeneralization	Magnification and minimization
Personalization	Dichotomous thinking

Table 30.6-2 Cognitive Distortions in Psychopathology

PRINCIPLES OF THERAPY

Therapeutic Relationship As in other therapies, the therapist-patient relationship is important in cognitive therapy and provides the medium for improvement. Therapists function as guides to enable their patients to acquire the understanding that will help them to cope better with their problems—the process of guided discovery—and also as catalysts to promote the kind of corrective experiences outside of therapy that will enhance the patients' adaptive skills. The style of the therapist in manifesting genuine warmth and nonjudgmental acceptance is similar to that of the client-centered or Rogerian therapist. In contrast to Rogerian and psychoanalytic therapists, however, the cognitive therapist plays an active role in helping to pinpoint present problems, focusing on important areas, proposing and rehearsing specific cognitive and behavioral techniques, planning homework assignments, and re-evaluating the experiences during therapy.

Most of the therapist's verbal statements are in the form of questions, reflecting the basic empirical orientation and the immediate goal of converting the patient's closed belief system into an open system. The therapist actively engages the patient in working out the agenda for each session and elicits feedback from the patient regarding the therapist's suggestions and behavior during the session. Although an attempt is made to maintain an optimal level of warmth and rapport during the therapy, this effort includes the uncovering of patients' negative reactions to the therapist and resistance to assignments. Transference reactions are often valuable in demonstrating to patients their interpersonal distortions. Similarly, resistance is dealt with in terms of the underlying dysfunctional beliefs.

Because the standard duration of cognitive therapy for depression is 15 to 20 visits over a 12-week period, there is considerable pressure on both patient and therapist to make the best possible use of the available time. Thus, a substantial effort is made to induce patients to carry out homework assignments designed to help them recognize and respond to negative cognitions, master the cognitive and behavioral skills learned during the sessions, and to test dysfunctional beliefs about themselves and others. This time-limited format requires that patients rather readily form and maintain the therapeutic alliance. Patients for whom formation of the alliance is troublesome (e.g., patients with borderline personality disorder) are generally unsuitable for this time-limited structure and will require much longer treatment. Panic and generalized anxiety disorders often improve in ten sessions or less.

The proximal goal is to promote behavioral activation, cognitive restructuring, which involves modification of the patient's systematic bias in interpreting personal life experiences and making future predictions. In sum, cognitive therapy is a learning experience in which the therapist plays an active role in helping the patient to uncover and modify cognitive distortions and dysfunctional beliefs.

Case Conceptualization To prepare a treatment plan, a tentative conceptualization of the case is needed. Ideally, this formulation includes an understanding of the person's idiosyncratic dysfunctional beliefs, specific vulnerabilities, and how the particular stresses impinge upon these vulnerabilities to activate the present symptoms. Most patients have dysfunctional core beliefs, such as “I am helpless” or “I am unlovable,” upon which dysfunctional disorder-specific (idiosyncratic) beliefs

are superimposed—such as “If I don't have love, I am nothing” (depression) or “A disaster could occur at any time” (anxiety disorder).

Each personality disorder is also characterized by a specific set of dysfunctional beliefs. The cross-sectional analysis delineates the dysfunctional core beliefs and idiosyncratic beliefs and the dysfunctional strategies developed by the patient in reaction to these beliefs (e.g., the excessive attachments of the dependent personality in response to beliefs of helplessness or unlovability).

Case formulation rests on both a historical and prospective evaluation of the pattern of symptom exacerbation in relation to situations and the patient's conceptualization of them. By searching for common denominators among symptoms, the patient's views of diverse situations and the underlying cognitive pattern can be identified.

A woman was handicapped by fears of elevators, tunnels, hills, closed spaces, riding in open automobiles, flying in an airplane, swimming, walking fast, running, strong winds, and hot muggy days. A careful review of these problems revealed that the common denominator was a fear of suffocation—she had the idiosyncratic belief that she could suffocate in one of these situations in which there could conceivably be a deficit of air: stalled elevators, tunnels, closed car, and closed spaces. Her expectation of riding in an open car was that the “wind would be swept out of my mouth.” Fear of riding in an airplane was expressed in a fantasy of a leakage or a sudden drop of oxygen occurring. Initially, treatment consisted of inducing the patient to pay more attention to her respiratory cues. As she began to realize that she felt a tightening of her chest muscles and a sense of suffocation even when she only thought about being in one of these situations, she began to recognize that her fears originated in her imagination and did not represent realistic dangers.

STRUCTURE OF THERAPY PROGRAM

Behavioral Techniques These are prominently used at the beginning of therapy with profoundly depressed patients who have a limited capacity for introspection and abstraction needed to identify and evaluate automatic thoughts or assumptions. Often, the first step is to obtain baseline information by asking patients to complete the Daily Schedule of Activities form. This report forms the basis by which the therapist and patient design and schedule homework that can vary substantially depending on the phase of therapy. Because a significant part of the positive effect of cognitive therapy results from the patients' application of the principles outside the therapy session, it is important to help patients structure their days and weeks such that the impact of the therapy is maximized.

Initially, therapist and patient agree on scheduling activities that help to mobilize the patient and counteract the inertia often present, especially in depression. Because most patients need to proceed in small steps, a graded-task assignment is developed to enable them to have progressively greater success experiences without overextending themselves. Typical behavioral techniques are listed in Table 30.6-3.

Behavioral Techniques	Cognitive Techniques
Mastery and pleasure ratings	Identifying automatic thoughts
Activity scheduling	Testing and correcting automatic thoughts
Graded task assignment	Reattribution techniques
Behavioral rehearsal	Identifying and testing imperatives—“shoulds”
Role-playing	Cognitive rehearsal through imagery
Testing dysfunctional beliefs	Imaginal recreation of pathogenic events

Table 30.6-3 Behavioral and Cognitive Techniques

The more severely retarded patients, and particularly those who are hospitalized, may have a wide variety of negative thoughts regarding the activities, such as “I won't be able to do it,” “I would feel silly,” or “It won't do any good.” These cognitions are “grist for the mill” and can be formulated as predictions or hypotheses that the patient can then test as part of the graded task assignment.

Cognitive Techniques When the patient is already active, more purely cognitive procedures may be used. The patients track their automatic thoughts, particularly when they precede or accompany a negative feeling. An automatic thought occurs spontaneously, is very rapid, and represents an immediate interpretation of a situation. A depressed patient, for example, on seeing a lose friend had the automatic thought, “She won't want to talk to me.” These patients are asked to fill out the Daily Record of Dysfunctional Thoughts and are trained to give reasonable or rational responses to their negative automatic thoughts.

During the course of cognitive therapy the therapist and patient review the relationship between the automatic thoughts and an objective description of a disturbing event. Logical errors are identified and more realistic interpretations of the event are considered, thereby correcting the automatic thoughts.

Reality testing of automatic thoughts and assumptions is carried out by treating the belief or thought as a hypothesis to be tested; for example, discussion of its validity with the therapist using information already provided by the patient in prior sessions or through prior homework assignments. Alternatively, the therapist and patient may design a homework assignment (“experiment”) to more directly test the notion in question by gathering new evidence.

As therapy progresses, attention is focused on the patient's underlying beliefs, such as “If I'm not successful, then I am a failure” or “If somebody doesn't like me, it means I'm socially undesirable.” These beliefs are re-evaluated in the same way as automatic thoughts; namely, in terms of the evidence supporting them, the logical basis on which they rest, and empirical testing.

Automatic Thoughts Most people are unaware that automatic thoughts precede unpleasant feelings or automatic inhibitions. These thoughts occur rapidly, are at the fringe of awareness, and include both specific verbal ideas and images. With some training, however, patients increase their awareness of these thoughts and are able to pinpoint them with a high degree of regularity. It is possible to perceive a thought, focus on it, and evaluate it just as it is possible to identify and reflect on a sensation, such as pain.

Automatic thoughts are distinguished from the ordinary flow of thoughts observed in reflective thinking or free association. For a specific form of psychopathology, a large proportion of automatic thoughts reflect the idiosyncratic cognitive theme specific to the syndrome (Table 30.6-1). Although these thoughts are often very rapid and generally are at the fringe of consciousness, they may be identified retrospectively because they generally precede some emotion such as anger, sadness, or anxiety: the content of the automatic thought is consistent with that emotion. These thoughts are generally plausible to the patient, who takes it for granted that they are accurate. Finally, they have an imperative quality such that they recur even though the person tries to block them out.

In therapy, patients learn to identify their automatic thoughts and their reactions to them. Patients may report, for example, a number of circumstances in which they feel unaccountably upset. Usually, a gap may be identified between the stimulus situation and the emotional response to it. An excessive or inappropriate emotional response becomes more comprehensible when patients recollect the thought that occurred during this gap. Albert Ellis has referred to the sequence as the ABC sequence. *A* is the activating stimulus or event and *C* is the excessive inappropriate response. *B* is the blank in the patient's mind that must be filled by identifying the fringe automatic thought.

Depressed and anxious patients predictably interpret many situations in systematically biased ways, even though more plausible interpretations are available. When asked to reflect on alternative explanations, patients may realize that their initial interpretations are erroneous or rest on unlikely inferences. They may then be able to recognize that they tailored the facts to fit their performed negative conclusions. When one compares what actually appened in a particular situation with the patient's automatic thoughts in that situation, a number of logical errors are often apparent (Table 30.6-2).

Cognitively oriented techniques (Table 30.6-3) are applied as patients become less symptomatic and thereby become more objective about themselves and their thinking. As patients are able to view their thoughts objectively, these thoughts become hypotheses that are amenable to experimental testing and revision using

cognitive techniques. They thereby unravel, clarify, and modify the meanings they have assigned to upsetting events. Cognitive techniques, such as identifying automatic thoughts, recognizing and correcting cognitive distortions, and identifying broad beliefs and assumptions that underlie the dysfunctional thoughts, are used to clarify patients' problems.

Through a process of collaborative empiricism, the therapist and patient may produce shifts in thinking to a more realistic level. Structural changes may be achieved through the analysis of basic beliefs and imperatives that have governed the person's responses. Structural change extends beyond modifying the cognitive errors associated with the specific syndrome to changing the underlying organization of beliefs and assumptions that misclassify events as threatening or upsetting.

Homework The design of appropriate homework assignments to test specific dysfunctional notions requires ingenuity on the part of the therapist. For instance, many persons with anxiety or depression believe that their behavior is consistently being scrutinized and judged adversely by other people. They view themselves as vulnerable, rejected, and ridiculed. These patients may believe that if they deviate even slightly from conventional expectations or from their usual rather restricted roles, other people will notice and disapprove. Homework assignments consist not only of completing the various forms mentioned previously but also of testing these kinds of dysfunctional beliefs.

APPLICATION TO SPECIFIC CONDITIONS

Cognitive therapy has been applied to a variety of psychopathological conditions (nonpsychotic major depressive disorder, generalized anxiety disorder, panic disorders or phobias, obesity, alcohol or other substance use disorders, chronic pain). Different techniques may be used depending on the cognitive profile of the disorder, the phase of therapy, and the specific cognitive formulations of a given case.

Depression

Cognitive Triad Depression entails the activation of three major cognitive patterns that lead to individual patients viewing themselves, their experiences, and their future in negative biased ways (“the cognitive triad”). The first component consists of patients' negative concept of themselves, present in nearly all depressed patients. They regard themselves as deficient, helpless, or unlovable, and they tend to attribute their unpleasant experiences to their presumed physical, mental, or moral defects. In their opinion, they are undesirable and worthless because of their presumed defects, and they tend to reject themselves because of them. Furthermore, they may regard themselves as lacking in those attributes that they consider to be essential for the attainment of happiness or contentment.

The second component is the patients' predominantly negative interpretations of day-to-day and past experiences. They tend to see their personal world as exorbitantly demanding, as presenting insuperable obstacles to achieving their life goals, or as being devoid of pleasure or gratification. The third component, a negative view of the future, is present in nearly all depressions. Depressed patients anticipate that their current troubles will continue indefinitely and they see a future life of unremitting hardship, deprivation, and frustration.

Motivational and behavioral symptoms of depression are derived from, maintained, or exacerbated by negative cognitive patterns. Paralysis of the will results from the depressed patients' pessimism and hopelessness. Expecting their efforts to end in failure, they are reluctant to commit themselves to any goal, and their activity level drops. Believing they cannot affect the outcome of various situations, they experience a desire to avoid such situations. Suicidal wishes are an extreme expression of the desire to escape from problems that appear to be uncontrollable, interminable, and unbearable.

Regarding themselves as inept and undesirable, patients tend to overestimate the difficulty of normal tasks in life and expect things to turn out badly. They may unnecessarily seek help from others (increased dependency). Similarly, indecisiveness may derive from patients' beliefs that they are incapable of making the right decision. Even the subjective loss of energy, easy fatigability, and inertia seem to be derived from the negative expectancies.

Predisposition The cognitive model posits that certain individuals are predisposed to depression because of beliefs developed early in their lives. These beliefs are shaped by personal experiences, and stem from identification with significant others and from their perceptions of attitudes of other people toward them. Once a particular basic belief is formed, it may influence subsequent concept formations and if it persists, it is incorporated into an enduring cognitive structure or schema.

Although latent or inactive at given times, schemas are activated by particular situations analogous to those early experiences that shaped the development of the schema. When these schemas are activated, depression may develop. For instance, if a person had lost a parent during childhood, the disruption of a close interpersonal adulthood relationship may activate the concept of irreversible loss implanted by earlier experiences. Other types of precipitating events include not performing as expected on an examination, being demoted, acquiring a disease, or encountering serious difficulties or frustrations in meeting important life goals. Depressions do not always occur in relation to specific stress situations, but may be reactions to a series of traumatic experiences.

Although such events may be painful to the average person, they would not produce depression unless the person was especially sensitive to the situation because the specific predisposition was embedded in an underlying schema or belief. When such traumas are encountered, depression-prone persons experience a negative shift in their views of every aspect of their lives. As the depression deepens, their thinking is increasingly saturated with typical depressive themes, regardless of the immediate situation until they gradually lose their ability to view their negative thoughts objectively.

The activation of these schemas interferes with the capacity for objective appraisal of all events, thoughts, and emotions and reasoning is impaired. Systematic cognitive errors (e.g., arbitrary interpretation and selective abstraction) occur as dysfunctional schemas and become more activated. The patient becomes less and less objective in evaluating current situations.

The patients' cognitive and emotional responses themselves further serve to reinforce these negative views. As the unpleasant life situation activates cognitive schemas related to defeat or deprivation, negative expectancies and self-blame ensue. These in turn produce sadness, apathy, loneliness, and reduced performance. The emotions and inertia are interpreted as yet another sign of loss or failure—“I'm feeling bad, so things must be bad”—which further reinforces negative attitudes. Automatic thoughts evolve from these schemas.

Ms. M., a 68-year-old widowed mother of three, was self-referred for the chief complaint of “I am depressed,” following her husband's death. The current depressive episode had lasted 6 months and included a marked weight loss, anorexia, insomnia, easy fatigability, reduced capacity for enjoyment, social withdrawal, suicidal ideation, guilt, and difficulty concentrating. Her mood seemed reactive to daily stresses and events.

Stressors around the onset of the current episode included a gallbladder operation 3 months prior to the onset, and a visit to her daughter's house, during which the patient learned her daughter and son-in-law might be transferred out of state due to his position with a major corporation. Treatment of the current depressive episode with three different antidepressant drugs led only to a partial response.

Basically, four techniques were used to change cognition during twice-weekly cognitive therapy over a 6-week period.

Step 1 The first step consisted of scheduling activities each hour of each day and asking the patient to record the relative degree of mastery and pleasure experienced with each. This exercise provided a basis for examining how the patient evaluated herself with respect to different activities. Activities that she appeared to underrate for mastery and pleasure formed the basis for eliciting negative automatic thoughts associated with each activity. For example, she cleaned her living room and kitchen, but rated her sense of mastery as 0. Her thoughts at the time were: “Babs (her daughter) won't be here anymore” and “I used to enjoy this and now I hate it.”

Step 2 In addition to the thoughts collected through schedule-keeping (step 1), a number of other cognitions were collected using the Dysfunctional Thought Record. The patient was asked to record the events, her depressed or upsetting feelings, and the associated thinking in each instance when she felt a drop in her mood. The cognitive errors of overgeneralization and selective abstraction were identified and corrected by having the patient practice answering her automatic thoughts using a notebook and tape recorder.

Step 3 The thoughts collected in Steps 1 and 2 contained the theme of “Without Babs I have nothing.” This dysfunctional belief was logically refuted with several pieces of evidence provided by the patient herself. First, she was shown that in her 68 years, her daughter had been around for only 24 of them, yet the other 44 years had been very rewarding. Secondly, her daughter had not been home much of the time since she had attended college and then worked; yet, during the period that her daughter had been away, the patient continued to enjoy and participate meaningfully in a variety of activities.

Finally, the patient's own list of activities for the previous several weeks provided data that revealed to the patient that she had initially enjoyed and felt satisfied about volunteer work in a hospital, church social gatherings, and shopping when Babs was not around at all. So in fact without her daughter the patient still experienced substantial degrees of pleasure and mastery. Ms. M. came to see that although interaction with her daughter had provided a great deal of enjoyment and satisfaction, there was no hard evidence that without her daughter her life was meaningless.

Step 4 Finally, alternatives for maintaining contact with her daughter by phone, letter, or audiorecording were explored. Thus, even if her daughter did move out of state, some contact time nearly equivalent to the rather modest time during which the patient and her daughter had interacted after her daughter's marriage might be feasible. As the patient explored these alternatives, she began to see that her daughter had developed her own life. While her daughter might have to move away with her husband, the daughter was not disappearing altogether. Further, the potential move could be viewed as a testimony to the patient's successful completion of the mothering role.

The patient responded rather rapidly so that her depression ratings had normalized by the ninth session. Nine months following termination, the patient reported doing volunteer work, experiencing no symptoms of the syndrome of depression and scored a 0 on the Beck Depression Inventory (BDI).

Ms. M. initially recorded her activities and rated them for the sense of mastery and pleasure associated with each. This technique not only helped her to become objective about her present day-to-day activities, but also helped to elicit thoughts that occurred in specific situations that led her to devalue herself. Her preoccupation with the anticipated loss of her daughter became apparent. In addition, evidence was accrued that later was used to refute one of the themes in her thinking—namely, that without her daughter, nothing was worthwhile.

Suicidal Behavior Suicidal behavior has specific cognitive, motivational, and behavioral features that lead to particular cognitive techniques. Suicidal individuals evince high levels of hopelessness—the greater the hopelessness, the more likely it is that patients will commit suicide and their thinking often has a rigid all-or-none quality that precludes objective problem-solving. Hopelessness further impairs poor problem-solving; conversely, the difficulties in coping with life situations independently contribute to suicidal potential.

When suicidal individuals are unable to generate alternative solutions to solve current problems (i.e., their accustomed way of dealing with a problem does not work), they seem unable to think of alternatives to rectify the situation. Thus, much of therapy is devoted to problem-solving techniques, such as defining the problem, generating alternative ways of solving the problem, and implementing alternative solutions.

In treating suicidal patients it is important to promote a good relationship from the beginning and to establish bridges from one therapy session to the next so as to discourage suicide attempts between sessions. The therapist immediately focuses on the patients' hopelessness and applies a variety of cognitive techniques to increase the patient's objectivity about current dilemmas. The therapist teaches patients to identify and change their all-or-nothing thinking style, thereby opening up alternative solutions previously not considered.

Suicidal patients seem to operate according to certain specific beliefs, such as “If my girlfriend doesn't respond to me, it means she doesn't like me and I will always be alone.” By addressing both the hopelessness and the rigid self-defeating beliefs, the therapist helps patients to recognize a variety of alternative solutions to life's problems (i.e., suicide to escape these problems is not necessary).

COGNITIVE MODEL OF ANXIETY DISORDERS

Various anxiety syndromes may be considered as expressions of a malfunction of normal survival mechanisms. As in depression, the various symptoms are the manifestation of the operation of the specified systems: cognitive, affective, behavioral, motivational, and physiological. The basic mechanism for coping with a threat is the same for the anxious patient as for the normal person. The difference between the two is that the anxious patient's perception of danger is either incorrect or excessive and is based on incorrect or inaccurate false premises whereas the normal response to threat involves a reasonably accurate assessment of the possibility of harm.

Although normal persons may incorrectly identify situations as dangerous (and thereby become immobilized to deal with them), this misinterpretation is readily amenable to objective reappraisal following which the response returns to a baseline level. The anxious patient, in contrast, consistently misperceives danger and is relatively insensitive to cues indicative of safety. Moreover, the patient's capacity for objective reappraisal becomes impaired.

The thinking of patients with clinical anxiety revolves around the notion of danger and how it should be dealt with. When contemplating a problematic or ambiguous situation, anxious patients automatically anticipate the most negative consequences. Although it may be adaptive under some circumstances to anticipate the worst possible case, patients with clinical anxiety are consistently fixated on anticipating extreme outcomes and therefore are overprepared to deal with physical or social threats.

The cognitive appraisal of danger triggers (1) the emotional component, (2) the motor components specifically designed for coping with a threat (e.g., fight, flight, freeze, or faint), and (3) the physiological components (i.e., activation of the autonomic nervous system), which facilitates the action of the specific motor component. Anxiety symptoms reflect the subjective experience of the specific systems: nervousness (affective); muscular tension (motor); wish to flee (motivational); inhibition of speech and movement (behavioral); and multiple fears and inhibition of memory and selective thinking (cognitive).

Subjective Anxiety Subjective anxiety prods an organism to take protective action in response to perceived danger. The immediate response to danger (e.g., freezing), occurs almost instantaneously. However, anxiety also impels the person to select an appropriate strategy after appraising the dangerous situation. Subjective anxiety generally increases as one approaches a dangerous situation and decreases as one withdraws. Thus, subjective anxiety reinforces the motivation to reduce the danger by withdrawing. Furthermore, anxiety also stimulates the person to mobilize active coping mechanisms to reduce the danger.

When anxiety is continuous yet without objective threat, as in the anxiety disorders, it may be incorrectly attributed to ongoing situations perceived as dangerous. When a person incorrectly conceives of problematic situations as dangerous, there is no opportunity to develop or apply coping skills because there is no objective danger for which the coping is appropriate.

Autonomic activity facilitates the motor activity through regulating blood supply, metabolism, and temperature control (e.g., through sweating). A person manifests the same type of physical response in response to a psychosocial threat—for instance, negative evaluation by a superior—as to a physical threat (e.g., being injured or killed). The same autonomic-motor pattern, such as defensive stiffening that is manifested by sweating and pulse and blood pressure changes, occurs in response to either type of threat. Inhibition of action (muscular rigidity, blocking, and muteness) is expressed in tonic immobility associated with sympathetic dominance. Demobilization (muscular flaccidity and loss of postural control) is embodied in atonic immobility associated with parasympathetic dominance. Muscular and vocal inhibition reflects the impulse for self-protection and self-regulation.

This mechanism involved in anxiety disorders is counterproductive in the usual situation of psychosocial threat. A person with test anxiety or public speaking anxiety becomes fixated on the ultimate disaster (e.g., failing or being rejected). Thoughts of failure or rejection, although extraneous to the immediate performance task, are central to the danger and are consequently enhanced. Concentration, planning, and recall regarding the task are extraneous to the danger and are thus blocked.

Cognitive Therapy For intensely anxious patients a cognitive groundwork must be laid before any behavioral tasks are introduced. Patients are given a framework to integrate the experiences aimed at correcting their exaggerated fears. Thus, patients who have strong fears and inhibitions relevant to asserting themselves would be encouraged to explore all the presumed catastrophic occurrences that would result from unsuccessful self-assertion. They could then weigh the magnitude of pain resulting from a reversal and compare it with the potential gain from self-assertion.

In depression, the main thrust is towards disconfirming a strong belief, such as “I can't do anything” or “Nothing brings me pleasure.” A single disconfirming experience (e.g., in which patients do something they expected to fail at) can substantially undermine this belief. In contrast, in anxiety a series of positive outcomes do not in themselves prove that a negative outcome might not occur at another time. Thus, the initial work with anxiety patients consists of reducing the catastrophizing regarding the importance of a negative outcome and teaching patients coping techniques in order to improve the chances of positive outcomes and to learn how to deal with negative outcomes should they occur.

Many anxious individuals view their anxiety as being constantly at a high level, but typically, anxiety occurs in waves. Self-monitoring anxiety symptoms helps patients to recognize that anxiety has a beginning, a peak, and then tapers off, and that they are able to handle it successfully. Thus, when in an anxiety-provoking situation, the person merely has to “wait it out” until the wave of anxiety passes.

In the early stages various imagery techniques are helpful. Patients can be encouraged to “live through” a frightening situation by imagining that it is occurring right in the therapist's office (*cognitive rehearsal*). Patients can then be asked to imagine the worst and best possible outcomes. This technique gives them greater perspective of the feared situation and constitutes a learning experience in itself.

After preliminary cognitive restructuring using Socratic questioning, reevaluation of anticipated catastrophes, and imagery techniques, patients are encouraged to try out what they have learned in actual threatening situations. These tasks are assigned in a graded way so patients are only gradually exposed to threatening situations. In these situations they continue to monitor automatic thoughts, and when possible, to re-evaluate them.

Panic Disorder The treatment of panic disorder focuses on the patient's tendency to make catastrophic misinterpretations about bodily sensations or mental experiences. The first one or two sessions are devoted to determining the precise nature of the patients' symptoms and how they misinterpret them. Careful attention is paid to the specific automatic thoughts that occur during the panic attack, the situation in which the panic attack is likely to occur, and the prodromal symptoms of the attack. By the third session the patient is generally ready to consider a different interpretation of the cause of the symptoms leading to the panic attack (*reattribution*).

It is usually possible to produce a mild panic-like attack through hyperventilation, exercise, use of imagery, or other stimuli, depending on the specific case. Patients are also taught distraction techniques to demonstrate to them that panic attacks are controllable. Following this, patients are encouraged to enter situations in which panic attacks can occur and to apply the techniques or rational restructuring based on what they have learned. It is often desirable for patients to go through at least one more panic attack from beginning to end, but without using rational responses or distractions to convince them that the panic attack in itself does not constitute a serious danger.

Obsessive-Compulsive Disorder Initial treatment focuses on increasing patients' awareness of automatic thoughts and images at the time of urges to carry out compulsive acts (e.g., handwashing or checking). These cognitions often portray the extent of patients' fears that motivate these redundant activities. In most cases individuals respond to these catastrophic ideations (or images) as though there is a high probability of their occurring. Thus, decatastrophizing methods are important. It is also crucial to steer the patient to recognize the excessive sense of responsibility in thoughts such as “If I don't do these compulsions and something bad happens, it would be my fault” or “It is my responsibility to protect my own life and those of my family.”

Typical behavioral methods include *response prevention,* in which the patient is prevented from engaging in the compulsive rituals; *exposure,* in which patients are exposed to threatening material (e.g., a pile of dirty towels). These behavioral methods are used to elicit the patient's thinking and images. By testing the reality of these thoughts and images, patients are able to cognitively restructure their unrealistic fears of some disaster.

Obsessions without compulsive behavior are sometimes more difficult to treat. The therapist focuses not only on the obsessional thought, for example, “bloody vagina” or “God is dead,” but on the self-evaluative cognitions that follow the thought, such as “I must be going mad” or “I'm God.” The therapist demonstrates that attempts by the patient to neutralize or suppress the offensive thought only increases its intensity. The patient is encouraged to allow these thoughts to occur and the therapist may attempt to desensitize patients by having them listen to taped recordings of the verbal thoughts.

Paranoid States To treat paranoid states, a strong collaborative alliance with the patient is essential and the patient must trust the therapist. In applying cognitive methods, the therapist must be aware that some of the patient's beliefs may be so highly charged that they are better left unexplored until a later date.

Initial interventions entail having patients monitor their automatic thoughts regarding other people's presumed biases towards them, with subsequent encouragement to look for alternative explanations for others' presumed noxious behavior. Next, certain basic beliefs, such as “I can't trust anybody” or “Everybody is out to get me,” can be subjected to a logical and empirical analysis. The therapist should be alert to finding at least a single episode that can disconfirm an absolute statement (e.g., “Everybody mistreats me”) since a single disconfirmation, if accepted by the patient, can help to undermine the absolute nature of such thoughts.

Decatastrophizing methods are also useful to demonstrate to patients that the consequences of being mistreated, abused, or persecuted are not necessarily as bad as they think. Further, it is important to have the patients learn that by attaching too much value to certain rules they impose on others (such as that other people must show deference towards them), they are continually setting themselves up for feeling infringed upon.

Somatoform Disorders Cognitive therapists occasionally see patients with a somatoform disorder, often after some type of traumatic physical episode that the patient believes caused damage to the brain, the spinal cord, the nerves, or some other part of the body. The therapist can devise the cognitive formulation for the particular patient by pinpointing exactly what the patient thinks has happened, the patient's theory as to what damage was caused, and the particular symptom presumably caused by that damage. The therapist can then reeducate the patient about the improbability of a real physical problem and can give the patient a reattributing type of explanation for the symptom.

Eating Disorders Anorexia nervosa and bulimia nervosa are characterized by a constellation of maladaptive beliefs that generally revolve around a few central assumptions, such as, “My weight and shape determine whether I am worthwhile or socially acceptable.” Associated with these assumptions are beliefs such as “I will look ugly if I gain more weight,” “If I don't starve myself, I will let go completely and become enormous,” “The only thing in my life that I can control is my weight.”

For the most part therapists initially give graded task assignments to attempt increased food intake (anorexia) or decreased binge eating, purges, and vomiting (bulimia). Fairly rapidly, however, discussion focuses on specific faulty beliefs regarding eating and its effect on body weight, the importance of appearance, and the patient's insistence on self-control. As therapy progresses, the cognitive analysis spreads to various other dysfunctional beliefs such as the patient's negative

self-concept (helpless, unlovable, undesirable), hypersensitivity to control by others, and wishes for approval.

Personality Disorders A complete elaboration of cognitive therapy of the personality disorders has been published. The focus of treatment is directed toward formulating the patient's basic beliefs (e.g., “Since I am helpless, I need to have a stronger person available at all times”) and demonstrating how those beliefs feed into the patient's dysfunctional behavior.

In contrast to the treatment of Axis I disorders, treatment of Axis II disorder requires a long period of therapeutic work—often 1 or more years. Also, there is a greater therapeutic focus on transference issues, exploring childhood patterns and even revivifying pathogenic childhood experiences, an increasing convergence with psychodynamic therapy. The major difference remains that the cognitive therapist is more active and directive, the therapeutic sessions are more structured, the content is based on conscious rather than unconscious or repressed memories, on exploring and testing cognitive distortions and basic beliefs, and patients are expected to carry out homework assignments.

M.K., a 40-year-old, hard-driving director of a research institute, was admitted because of a major depressive episode and generalized anxiety disorder. During the depressive episode, he had strong beliefs such as “I am worthless,” “I am useless,” and “I am a burden to my family.” With therapy for the acute depressive episode, these ideas were transformed into testable hypotheses about himself. Through Socratic questioning, application of logic, and empirical testing, the patient was able to regard these notions as testable beliefs rather than as reality. The cognitive interventions effectively contradicted his notions and, as these beliefs were attenuated, his depressive and anxious symptoms dissipated. Both patient and therapist recognized, however, that his personality problems (a mixture of compulsive and narcissistic features), led him to overemphasize achievement, productivity, efficiency, and perfection.

While initially regarding these beliefs as normal, healthy, and adaptive, M.K. came to see himself as being driven by these beliefs. When these traits were translated into beliefs, he began to see how dysfunctional they were. For him to understand and gain perspective on these beliefs, the therapist explored their origin during the developmental period of his life. M.K. began to see that these ideas were not based on rational decision-making, but rather they were imprinted on him by his parents.

The patient, an only child, saw his father as tough, demanding, and having very high expectations for him. The patient observed his father withdraw approval from him when he did not perform extremely well in school, and he observed his mother criticize his father for his supposed inadequacies. As a child, M.K. learned, “I must do extremely well to be accepted or else I risk others' disapproval and withdrawal of their respect and caring.” At times, for example when his best friend was promoted ahead of him, he had the thought “I'm inadequate.”

The patient coped with the painful beliefs of inadequacy and fear of rejection by developing and living up to extremely high standards, crystallized in a cluster of beliefs: “I must be the best at everything I do. It isn't enough to excel at one thing. If I have the potential in any area, I must live up to that potential in each of these areas.” For many years he fulfilled his professional potential as a biochemical researcher. This success strengthened his belief that “I can and should excel at everything now and in the future.”

Problems inherent in this belief, however, included tremendous personal sacrifice for overvaluing professional achievements. He had no time to “stop and smell the roses,” to nurture his wife and children, to appreciate the joys of nature, or to replenish himself physically and emotionally. He suffered from chronic anxiety, sleep problems, psychosomatic symptoms, and he was unable to increase his scientific achievements over a long period of time. External factors interfered with achievement: increased administrative demands, scarcer funding, changes in the attitude of faculty who wished to go elsewhere, and increased requirements from a new president. His time and energy were finite, yet the external factors over which he had no control continued to escalate. He could not accept these real limitations on his high-level productivity; instead, he blamed himself—“I'm inadequate.”

He developed anxiety symptoms when he perceived that he might not have the resources to cope with situations in which achievement was at risk. He was vulnerable to depression when he was forced to take time out due to overwork. When anxious, he felt as if he was on a tightrope, fearing that with one misstep, he would fall all the way down. When depressed he saw himself as being inadequate and a failure. “If I were really adequate, I could continue to achieve. Since I can't achieve to the same level as previously, it means I'm inadequate.”

Therapy consisted of delineating and then modifying these dysfunctional beliefs. The therapist explored with him his perfectionistic, dichotomized view of success, “Either I live up to my potential in all areas or I've failed,” whereby one misstep felt fatal to him. He believed he had to continuously drive himself to stay at the top because even one step below the top was equal to failure in his mind. In addition, he was hypersensitive to and became either anxious or irritated at perceived obstacles to smooth functioning: signs of physical or emotional illness in himself, a secretary's mistake, the tardiness of others' reports; in short, anything that could interfere with his achieving optimally.

The therapist helped the patient to fashion more functional beliefs: (1) “It is rewarding to succeed highly, but lesser success is rewarding also and has no bearing on my adequacy or inadequacy. I am adequate no matter what;” (2) “Scientific achievement is important but not vital;” (3) “What is vital is a reasonable mix of achievement; emotional and physical well-being; an enriching family and social life; joy in recreation, relaxation, and nature; and spiritual and personal growth;” and (4) “It is impossible for anyone to achieve this vital balance if any one area is overvalued.”

Therapy extended over the course of 1 year. The patient's dysfunctional attitudes gradually attenuated and new attitudes were incorporated. He continued to have a few brief spells of anxiety and depression but without full-blown recurrences of either disorder.

Schizophrenia Four recent well-controlled British studies indicate that cognitive or cognitive-behavioral therapy was more effective than the control groups in the treatment of acute schizophrenia (along with medication and routine hospital care) and also in chronic hospitalized patients with treatment-resistant schizophrenia. In two 18-month follow-up studies of medication-resistant patients with schizophrenia, the cognitive therapy groups showed a significant and continuing improvement on the Brief Psychiatric Rating Scale (BPRS) whereas the control groups were no better at 18 months than at baseline. Further, the cost of the cognitive therapists was offset by reduced service utilization and associated costs during treatment and subsequent follow-up.

Other Applications Cognitive therapeutic methods have been developed for application in a variety of formats, clinical contexts, and use in special populations, including depressed patients seen in primary-care settings, depressed adolescents, and adults with bipolar I disorder. Cognitive therapy has been used successfully in group, family, and couple therapy formats. Cognitive therapy has also been shown to be effective with marital problems. Some of the newer applications of cognitive therapy have been with posttraumatic stress disorders, sexual offenses such as exhibitionism and incest, alcohol abuse, hypertension, and dissociative disorders. In addition, cognitive therapy has recently shown promise in the treatment of smoking cessation. Several studies have shown that cognitive therapy is effective in the treatment of cocaine abuse. One study found that cognitive therapy provided better long-term outcome than clinical management, while the other found that participants in cognitive-behavioral therapy were significantly more likely to achieve abstinence than those who participated in a 12-step facilitation program.

Particularly notable is the development of an interactive, video-based, computer-assisted adaptation of cognitive therapy for depressed patients. This approach does not eliminate the therapist, but does appear to save therapist time and have high patient acceptance.

In sum, cognitive therapy appears to be applicable to an increasingly large number of disorders, formats, populations, and clinical contexts. Age, sex, socioeconomic status, and racial or ethnic background do not appear to be important factors, but the cognitive therapist needs to adapt the theory and principles to the special features associated with specific subgroups, and empirical evaluations of these adaptations must be provided.

OUTCOME STUDIES

Table 30.6-4 summarizes systematically controlled efficacy studies of cognitive therapy for various disorders. The findings have been supportive of cognitive therapy for patients with major depressive, generalized anxiety, panic, and somatoform disorders; heroin dependency; as well as anorexia nervosa, bulimia nervosa, and schizophrenia.

Disorder	Conclusions
Major depressive disorder (outpatients)	Meta-analysis Cognitive therapy >other psychotherapies Cognitive therapy >pharmacotherapy
Generalized anxiety disorder	Cognitive therapy >behavior therapy
Panic disorder	Cognitive therapy >supportive therapy Cognitive therapy >behavior therapy, imipramine (Tofranil)
Hypochondriasis	Cognitive therapy >standard treatment
Inpatient depression	Cognitive behavioral + standard >standard treatment alone
Bulimia nervosa	Cognitive therapy >interpersonal therapy
Heroin dependency	Cognitive therapy + supportive-expressive >methadone group counseling
Schizophrenia	Cognitive therapy >treatment as usual or befriending therapy (all treatments included antipsychotic medications)

Table 30.6-4 Clinical Outcome Studies of Cognitive Therapy

A meta-analysis of studies of cognitive therapy for major depressive disorder showed the superiority of this treatment over pharmacotherapy, behavior therapy, and other forms of psychotherapy. Another finding was that the average rate of relapses across four published studies was around 30 percent in cognitive therapy as compared to in excess of 60 percent in pharmacotherapy (without maintenance).

A report showed that cognitive therapy when combined with standard hospital treatment was more effective than standard hospital treatment (usually involving antidepressant drugs) followed by a period of outpatient therapy. Also, contrary to expectations, cognitive therapy has been found to be as effective in endogenous as in nonendogenous depressed patients. Of particular interest is the finding of an excellent response rate of 81 percent for cognitive therapy for hospitalized patients with depression who received no medication.

Cognitive therapy has been used with all age groups from children through geriatric populations. Several controlled studies have shown it to be at least as effective as antidepressant medication in the treatment of elderly depressed patients.

INDICATIONS

Cognitive therapy can be used to (1) remove or moderate the symptoms of the disorder as a sole treatment or in combination with medication; (2) reduce the likelihood of relapse or recurrence; (3) increase adherence to recommended medication treatment; (4) address specific psychosocial difficulties (e.g., marital discord, low self-esteem) that may either have preceded or been caused by the disorder; or (5) modify underlying beliefs (schemas) that contribute to dysfunctional personality trends or disorders. The indications for cognitive therapy are determined more by patient and therapist variables than by the nature of the disorder.

Patients The ideal patients are psychologically minded; able to recognize and label their emotions; to become aware of their automatic thoughts; and to see the connection between thoughts, feelings, and behaviors. The degree of fit between the patient's own personal notions of psychology and the basic cognitive model is important. Patients who adhere to such popular notions as the relation between stress and psychological disorders and the importance of self-control seem to benefit more from cognitive therapy than those wedded to Freudian concepts such as the unconscious and infantile fixations. High intelligence is not a prerequisite. Motivation for therapy is important but not initially crucial. Some hopeless, unmotivated, or depressed patients become highly motivated once they experience improvement.

Therapists As in any therapy, therapist characteristics are important. The ideal therapists are psychologically minded, versatile, attentive, empathic, and uncritical. They do not bring their own “personal baggage” (such as the need to control or show off) into the therapy session. A high motivation to “rescue” patients may be an asset with very sick patients but may be counterproductive for the less acutely ill. Skill in conducting cognitive therapy is obviously important for successful treatment. Several studies have shown a surprisingly high correlation between therapists' competency and successful outcome. Therapists who are skilled at educating patients about the cognitive nature of their problems and how to resolve them seem to get the best results. At least 1 year of supervised cognitive therapy following training and experience in basic interviewing skills is recommended.

Disorders Systematic data are lacking regarding the utilization of cognitive therapy for severe, psychotic, or melancholic depression, and for depressive episodes in bipolar I disorder. Since medications are of established efficacy in these conditions, cognitive therapy can be used to enhance the effectiveness of medications and has a specific role in promoting adherence to medication and training the patient to cope more successfully with life stressors and to integrate a more adaptive self-concept. Cognitive therapy without medication may be successful even with depressed inpatients who are refractory to medication or in whom medication is medically contraindicated. In view of the limited number of studies, however, cognitive therapy is not recommended as the initial treatment for depressed inpatients.

Most outpatients with either depressive or panic disorders who ultimately respond will do so within 5 to 7 weeks of once- or twice-weekly treatment with cognitive therapy alone, although a few patients have a more gradual response. If a symptom reduction of 50 percent according to a standard rating scale for depression is not achieved in 14 sessions, it might be useful to reconsider the diagnosis and treatment plan.

Maintenance treatment (booster sessions once or twice a month for 6 or 12 months after a course of more intensive once- or twice-weekly acute-phase treatment) may be useful in those with recurrent depressions; however, empirical evaluation of this notion is yet to be completed.

Clinical experience suggests that for those with double depression (recurrent major depressive episodes without full interepisode recovery superimposed on dysthymic disorder), the combination of cognitive therapy and medication may be the optimal choice. Clinical experience also suggests that major depressions that accompany medical disorders respond rapidly to cognitive therapy, especially in those without a history of preexisting psychopathology. For example, patients with their first myocardial infarction, with some forms of cancer, or with physical injuries that require marked psychological readjustments (e.g., blindness, loss of limb) often profit from this approach.

LIMITATIONS

Notwithstanding frequent successes with different individuals, for no obvious reason, some patients do not respond to cognitive therapy. Some with longstanding chronic conditions who have been treated by many professionals without lasting results may not do well in time-limited cognitive therapy. Sometimes an appropriate pharmacological agent added to cognitive therapy will produce a synergistic effect. In refractory cases, transfer to another therapist may be helpful. The therapist should also consider a change of strategy; for example, spending more time on empathic listening and less on exploration or becoming more active and directive. In any event, cognitive therapy is not a panacea even though its success rate is respectable.

Adverse reactions are difficult to differentiate from lack of efficacy. For instance, suicide attempts as well as premature terminations may be evidence of either adverse reactions or lack of efficacy. Some studies suggest that cognitive therapy is associated with lower premature termination rate than antidepressant pharmacotherapy. Perhaps the structured, planned, directive nature of this approach helps retain depressed outpatients in treatment. If so, cognitive therapy might be particularly useful for outpatients of low socioeconomic class, whose dropout rate from psychotherapy is particularly high.

Several limitations to this treatment are suggested by clinical experience. Patients with severely impaired reality testing (e.g., fixed delusions) or impaired reasoning abilities or memory function (e.g., organic brain syndromes) do not appear to respond well to cognitive therapy. However, cognitive methods may have a place even in those conditions if integrated into a total therapeutic regimen.

Despite our attempts to delineate the indications and contraindications for cognitive therapy, the fact remains that it may be difficult to predict outcome in an individual case. To some extent, the application is empirical (i.e., trial and error) for a given case. In some instances of refractory depression, for example, cognitive therapy has been used as a last resort and has been successful. For the most part, however, the professional evaluating a case can reasonably consider cognitive therapy, either alone or in combination with pharmacotherapy, for any of the conditions previously listed for treatment.

SUGGESTED CROSS-REFERENCES

Learning therapy is discussed in Section 3.3, sociology in Section 4.2, psychoanalysis in Chapter 6, other types of psychotherapy in the other sections of this chapter and the biological therapies in Chapter 31.

SECTION REFERENCES

Basco MR, Rush AJ: *Cognitive-Behavioral Therapy for Bipolar Disorder*. Guilford Press, New York, 1996.

*Beck AT: *Cognitive Therapy and the Emotional Disorders*. International Universities Press, New York, 1976.

*Beck AT, Emery G, Greenberg RL: *Anxiety Disorders and Phobias: A Cognitive Perspective*. Basic Books, New York, 1985.

*Beck AT, Freeman A, Associates: *Cognitive Therapy of Personality Disorders*. Guilford Press, New York, 1990.

*Beck AT, Rush AJ, Shaw BF, Emery G: *Cognitive Therapy of Depression*. Guilford Press, New York, 1979.

Beck AT, Sokol L, Clark DA, Berchick RJ, Wright FD: A crossover study of focused cognitive therapy for panic disorder. Am J Psychiatry *149:*778, 1992.

Beck JS: Cognitive approaches to personality disorders. In *American Psychiatric Press Review of Psychiatry*, vol 16. LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1997.

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, Iyengar S, Johnson BA: A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry *54:*877, 1997.

*Bruce TJ, Spiegel DA, Hegel MT: Cognitive-behavioral therapy helps prevent relapse and recurrence of panic disorder following alprazolam discontinuation: A long-term follow-up of the Peoria and Dartmouth studies. J Consult Clin Psychol *67:*151, 1999.

Butler G, Fennell M, Robson P, Gelder MP: Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. J Consult Clin Psychol *59:*167, 1991.

Clark DM, Wells A: Cognitive therapy for anxiety disorders. In *American Psychiatric Press Review of Psychiatry,* vol 16, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1997.

*Curtis D: Intensive cognitive behaviour therapy for chronic schizophrenia. Specific effect of cognitive behaviour therapy for schizophrenia is not proved. BMJ *318:*331, 1999.

*Dobson K: A meta-analysis of the efficacy of cognitive therapy of depression. J Consult Clin Psychol *57:*414, 1989.

Drury V, Birchwood M, Cochrane R, MacMillan F: Cognitive therapy and recovery from acute psychosis: A controlled trial. I: Impact on psychotic symptoms. Br J Psychiatry *196:*593, 1996.

Drury V, Birchwood M, Cochrane R, MacMillan F: Cognitive therapy and recovery from acute psychosis: A controlled trial. II: Impact on recovery time. Br J Psychiatry *196:*602, 1996.

Durham RC, Murphy I, Allan T, Richard K, Treliving LR, Fenton GW: Cognitive therapy, analytic psychotherapy and anxiety management training for generalized anxiety disorder. Br J Psychiatry *165:*315, 1994.

Fairburn CG, Jones R, Peveler RC, Carr SJ, Solomon RA, O'Connor ME, Burton J, Hope RA: Three psychological treatments for bulimia nervosa. Arch Gen Psychiatry *48:*463, 1991.

Garety P, Fowler D, Kuipers E, Freeman D, Dunn G, Bebbington P, Hadley C, Jones S: London-East Anglia randomised controlled trial of cognitive-behavioral therapy for psychosis. II: Predictors of outcome. Br J Psychiatry *171:*420, 1997.

Hollon SD, Najavits L: Review of empirical studies on cognitive therapy. In *American Psychiatric Press Review of Psychiatry,* vol 7, AJ Frances, RE Hales, editors. American Psychiatric Press, Washington, DC, 1988.

Kuipers E, Garety PA, Fowler D, Dunn G, Bebbington P, Freeman D, Hadley C: London-East Anglia randomised controlled trial of cognitive-behavioural therapy for psychosis. I. Effects of the treatment phase. Br J Psychiatry *171:*319, 1997.

Maude-Griffin PM, Hohenstein JM, Humfleet GL, Reilly PM, Tusel DJ, Hall SM: Superior efficacy of cognitive-behavioral therapy for urban crack cocaine abusers: Main and matching effects. J Consult Clin Psychol *66:*832, 1998.

Miller IW, Normal WH, Keitner GI, Bishop SB, Dow MG: Cognitive-behavioral treatment of depressed inpatients. Behav Res Ther *20:*25, 1989.

Power KG, Simpson RJ, Swanson V, Wallace LA, Feistner ATC, Sharp D: A controlled comparison of cognitive-behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalized anxiety disorder. J Anxiety Disord *4:*267, 1990.

Scott J, Wright JH: Cognitive therapy for chronic and severe mental disorders. In *American Psychiatric Press Review of Psychiatry,* vol 16, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1997.

*Segal ZV, Gemar M, Williams S: Differential cognitive response to a mood challenge following successful cognitive therapy or pharmacotherapy for unipolar depression. J Abnorm Psychol *108:*43, 1999.

*Sloan G: Anorexia nervosa: A cognitive-behavioural approach. Nurs Stand *13:*43, 1999.

*Tarrier N, Pilgrim H, Sommerfield C, Faragher B, Reynolds M, Graham E, Barrowclough C: A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. J Consult Clin Psychol *67:*13, 1999.

Tarrier N: Coping and problem solving in the treatment of persistent psychotic symptoms. Presented at the Second International Conference on Psychological Treatments for Schizophrenia. Oxford, United Kingdom, October, 1997.

Thase ME, Bowler K, Harden T: Cognitive behavior therapy of endogenous depression. Part 2: Preliminary findings in 16 unmedicated inpatients. Behav Res Ther *22:*469, 1991.

Thase ME, Simons AD, Cahalane JF, McGeary J: Cognitive behavior therapy of endogenous depression. Part 1: An outpatient clinical replication series. Behav Ther Res *22:*457, 1991.

Warwick HMC, Salkovskis PM: Hypochondriasis. In *Cognitive Therapy in Clinical Practice,* J Scott, JMG Williams, AT Beck, editors. Routledge, London, 1989.

*Wilhelm S, Otto MW, Lohr B, Deckersbach T: Cognitive behavior group therapy for body dysmorphic disorder: A case series. Behav Res Ther *37:*71, 1999.

Wilkes TCR, Belsher G, Rush AJ, Frank E, Associates: *Cognitive Therapy for Depressed Adolescents*. Guilford, New York, 1994.

Woody GE, McLellan AT, Luborsky L, O'Brien CP, Blaine J, Fox S, Herman I, Beck AT: Severity of psychiatric symptoms as a predictor of benefits from psychotherapy: The Veterans Administration-Penn study. Am J Psychiatry *141:*1172, 1984.

Wright JH, Thase ME, Beck AT, Ludgate JW, editors: *Cognitive Therapy with Inpatients. Developing a Cognitive Milieu*. Guilford, New York, 1993.

30.7 INTERPERSONAL PSYCHOTHERAPY

MYRNA M. WEISSMAN, PH.D., AND JOHN C. MARKOWITZ, M.D.

DEFINITION

Interpersonal psychotherapy, a time-limited treatment for major depressive disorder, was developed in the 1970s, defined in a manual, and tested in randomized clinical trials by the late Gerald L. Klerman and collaborators. Its success in research trials has led to its subsequent modification for different age groups, for subtypes of mood and non–mood disorders, and for use as a long-term treatment, in couples and group formats, over the telephone, and in a patient guide. The interpersonal psychotherapy manual has been translated into Italian, German, and Japanese. Having begun as a research intervention, interpersonal psychotherapy is only recently being disseminated among clinicians and in residency training programs. Increasing requests for training have followed the publication of interpersonal psychotherapy efficacy data, the promulgation of practice guidelines that include interpersonal psychotherapy among antidepressant treatments, interest in defined and proved treatments for managed care, and the endorsement of the 1995 *Consumer Guide*. A host of new research applications are also under study.

In 1993 practice guidelines appeared for mental health professionals and primary care practitioners. The two sets of guidelines differ considerably in scope, audience, and level of scientific basis for treatment recommendation. Neither claims to define the standard of care for any individual patient. Each discusses interpersonal psychotherapy as a short-term and maintenance treatment for depression, used alone or in combination with medication. The American Psychiatric Association (APA) *Practice Guideline for Major Depressive Disorder in Adults* describes interpersonal psychotherapy as useful for patients who face conflicts with significant others or who are having difficulty adjusting to a life transition. The clinical practice guidelines for treatment of depression in primary care settings recommended interpersonal psychotherapy for short-term treatment of nonpsychotic depression, to remove symptoms, prevent relapse and recurrence, correct causal psychological problems with secondary symptom resolution, and correct secondary consequences of depression. The guidelines state that medication alone may suffice to prevent relapse or recurrence and maintain remitted recurrent depression. The primary care guidelines describe interpersonal psychotherapy, cognitive-behavioral therapy, and behavior therapy as effective in most cases of mild-to-moderate depression but characterize indications for continuation-phase psychotherapy as unclear.

HISTORY AND THEORETICAL ISSUES

Interpersonal psychotherapy was initially formulated as a time-limited, weekly, outpatient treatment for depressed patients. Based on the ideas of Harry Stack Sullivan and the interpersonal school, interpersonal psychotherapy makes no etiological assumptions but uses the connection between onset of depressive symptoms and current interpersonal problems as a treatment focus. Interpersonal psychotherapy generally deals with current rather than past interpersonal relationships, focusing on the patient's immediate social context. It attempts to intervene in symptom formation and social dysfunction associated with depression rather than addressing enduring aspects of personality, which is difficult to assess during an episode of an Axis I disorder.

The theoretical foundation of interpersonal psychotherapy can be found in the writings of Sigmund Freud on grief and mourning and on attachment theory as developed by John Bowlby, who proposes that humans have an innate tendency to seek attachments. These attachments contribute to the survival of the species and to individual satisfaction. The threat of loss of important attachments creates anxiety and sadness, and frequent threats of such loss may predispose to depression. Attachment theory has stimulated a strong body of empirical research on the relationship between interpersonal relationships and depression. This research includes studies of childhood experiences and later adult depression, studies of the offspring of depressed parents, the childhood of adult depressives, social stress and life events, the role of social supports, lack of intimacy, marital discord and onset of depression.

Based on these theoretical and empirical sources, interpersonal psychotherapy assumes a connection between the onset of mood disorders (and perhaps other psychiatric disorders) and the interpersonal context in which they occur. The problem areas on which interpersonal psychotherapy for depression focuses were defined by psychosocial research on depression that suggested that complicated bereavement, role disputes, and role transitions were often associated with depressive episodes, either as precipitant or consequence. The progress of interpersonal psychotherapy has been defined by the sequential adaptation and empirical testing of the approach for particular psychiatric diagnoses.

TECHNIQUES

Phases of Treatment Interpersonal psychotherapy as a short-term treatment has three phases. The first, usually one to three sessions, includes diagnostic evaluation and psychiatric history and sets the framework for the treatment. The therapist reviews symptoms, diagnoses the patient as depressed by standard criteria (e.g., the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV]), and gives the patient the sick role. The sick role may excuse the patient from overwhelming social obligations but requires that the patient work in treatment to recover full function. The psychiatric history includes the *interpersonal inventory*, a review of the patient's current social functioning and close relationships, their patterns, and mutual expectations. Changes in relationships proximal to the onset of symptoms are elucidated (e.g., death of a loved one, children leaving home, worsening marital strife, or isolation from a confidant). This review provides a framework for understanding the social and interpersonal context of the onset of depressive symptoms and defines the focus of treatment.

The therapist assesses the need for medication on the basis of symptom severity, past history and response to treatment, and patient preference and then educates the patient about depression by explicitly discussing the diagnosis, including the constellation of symptoms that define major depression and what the patient might expect from treatment. The therapist next links the depressive syndrome to the patient's interpersonal situation within the framework of one of four interpersonal problem areas: (1) grief, (2) interpersonal role disputes, (3) role transitions, or (4) interpersonal deficits.

Strategies In the middle phase, the therapist pursues strategies specific to the chosen interpersonal problem area. For *grief*, defined as complicated bereavement following the death of a loved one, the therapist facilitates mourning and gradually helps the patient find new activities and relationships to compensate for the loss. *Role disputes* are conflicts with a significant other: a spouse, other family members, coworker, or close friend. The therapist helps the patient explore the relationship, the nature of the dispute, and options to resolve it. If these fail, they may conclude that the relationship has reached an impasse and consider ways to change the impasse or to end the relationship. *Role transition* includes change in life status (e.g., the beginning or end of a relationship or career, a move, promotion, retirement, graduation, or diagnosis of a medical illness). Patients learn to deal with the change by recognizing positive and negative aspects of the new role they are assuming and assets and liabilities of the old role it replaces. *Interpersonal deficits*, the fourth interpersonal psychotherapy problem area, defines the patient as lacking social skills, including having problems initiating or sustaining relationships.

Interventions Interpersonal psychotherapy sessions address present here-and-now problems rather than childhood or developmental issues. Sessions open with the question “How have things been since we last met?” This focuses the patient on recent interpersonal events and recent changes in mood, which the therapist attempts to link. Therapists take an active, nonneutral, supportive, and hopeful stance. Important interventions include helping the patient identify desires for resolution of interpersonal problem areas and options for achieving those desires. The therapist emphasizes the need to test new skills and change the interpersonal environment and often uses role playing to help the patient prepare for such life changes. Interpersonal psychotherapy focuses on what happens in the patient's life outside the office rather than on processes such as transference within it.

The final phase of interpersonal psychotherapy, occupying the last few weeks of treatment, supports the patient's sense of independent competence by recognizing and consolidating therapeutic gains. The therapist also helps the atient anticipate and develop ways of identifying and countering depressive symptoms should they

arise in the future.

CLINICAL ISSUES

Indications Indications for interpersonal psychotherapy have been defined by efficacy trials. Interpersonal psychotherapy has been one of the most carefully studied psychotherapies for mood disorders and the only psychotherapy tested in a maintenance treatment study. Its efficacy has been established for various subtypes of nonpsychotic depressive disorders and for bulimia nervosa.

Limitations Interpersonal psychotherapy showed no benefit in two studies of patients with substance abuse. It has not yet been definitively tested for other psychiatric disorders, such as anxiety disorders and borderline personality disorder, although studies are planned or under way.

Complications Interpersonal psychotherapy is an uncomplicated treatment that has met with high levels of patient acceptance and excellent clinical results in most trials. Attrition has been lower than that in competing therapies in several studies. Not all patients benefit from interpersonal psychotherapy, but no particular complications have been reported. Because the treatment is time-limited, nonresponse to interpersonal psychotherapy should lead to reevaluation and prescription of an alternate treatment after 12 to 16 weeks, minimizing the risk of complications.

Contraindications Interpersonal psychotherapy is not intended to be a treatment for delusional depression. It showed no benefit in two studies of patients with opioid dependence and cocaine abuse.

Ms. A., a 29-year-old never-married vice president in a successful business, presented with a 9-month history of depression complicated by panic attacks. She described her involvement in a relationship with her superior at work, which they both desired but which work policy forbade. Her symptoms emerged under the pressure she felt to resolve this situation, from which she saw no way out. Neither she nor her boyfriend Bob wanted to leave their jobs, and neither wanted to end the relationship. Yet exposure of their secret threatened both their jobs.

In this context Ms. A. developed the full spectrum of symptoms of major depressive disorder, with a Hamilton Rating Scale for Depression score of 28 on presentation, including intermittent panic attacks and suicidal ideation without attempts. Her work and mood suffered. She had a family history of depression and alcoholism but had no prior symptoms, and she did not drink. She had been in twice-weekly psychodynamic psychotherapy for 2 years, starting after the breakup of a previous relationship. She felt this therapy had increased her self-understanding but had not helped her symptoms. She had considered but declined to take antidepressant medication, feeling it would not resolve her dilemma.

Although quite depressed, Ms. A. quickly agreed that she was in a depressive episode and was able to relate it to her interpersonal situation. This had some aspects of a role transition—a change in her personal and work roles—but was best characterized as a role dispute with Bob: what did she want to happen, and what options did she have to negotiate its happening? A 12-week time limit was set for treatment. Her interpersonal therapist focused on helping her explore various options. She developed a resume and went to a headhunter to seek new job possibilities, although she was ambivalent at leaving her hard-earned and promising post. Role-playing helped her discuss the situation and its possibilities with her lover. Subsequent encounters with Bob better defined the nature of their agreements and differences.

Simply defining the problem and mobilizing Ms. A. to constructive action yielded an improvement in her mood and a greater sense of mastery of her situation. Interpersonal psychotherapy also helped her to develop interpersonal skills: although Ms. A. was appropriately assertive in her work role, she was far more pliant and submissive in her personal relationships, a problem that emerged in reviewing her interpersonal inventory of past romances. Depressive symptoms steadily subsided, but the turning point came in the sixth week of interpersonal psychotherapy when, after considerable exploration and role-playing in therapy, Ms. A. told her boyfriend that she had decided she could not give up her job; either they would have to end the relationship, or he would have to move. He initially said that his own job came first. But several days later he recanted, saying how much he loved her and that he would try to find a new job himself. Her symptoms resolved. Shortly thereafter, a wonderful prospect appeared for Ms. A. in effect giving her a promotion at a different, related company, and allowing the relationship to continue overtly. She attributed her symptomatic improvement to her newly firm stand, not to the job change. On 6-month follow-up, she remained euthymic without panic attacks and felt she was holding her own in a relationship that was going more smoothly than previous relationships ever had.

Goals of Treatment As illustrated above, interpersonal psychotherapy has two goals: (1) to help the patient solve an interpersonal problem and thereby (2) to alleviate the patient's psychiatric syndrome (e.g., depression).

ETHICAL ISSUES

In encouraging depressed patients to recognize and enact their own goals and wishes, therapists must be careful not to impose their own values (e.g., whether the patient stays in a marriage). Although interpersonal therapists take a relatively relaxed and informal, conversational approach to therapy, they must know appropriate boundaries regarding self-disclosure and focus treatment on benefit for the patient.

RESEARCH AND EVALUATION

Interpersonal psychotherapy has been widely tested in outcome research. The range of this research and its encouraging results indicate the excitement surrounding the treatment.

Acute Treatment of Major Depression The efficacy of interpersonal psychotherapy as a treatment for acute depression was first demonstrated by Klerman, the first author, and colleagues in a four-cell, 16-week randomized trial of interpersonal psychotherapy and amitriptyline (Elavil), alone and in combination, and a nonscheduled control treatment for 81 outpatients with major depressive disorder. There were no significant differences between interpersonal psychotherapy and amitriptyline in symptom reduction by the end of treatment, although the effects of amitriptyline appeared earlier. Each active treatment reduced symptoms more effectively than the nonscheduled control group, and combined amitriptyline-interpersonal psychotherapy was more effective than either active monotherapy. A 1-year naturalistic follow-up found that many patients sustained benefits from the brief interpersonal psychotherapy intervention. Patients who received interpersonal psychotherapy developed significantly better psychosocial functioning, whether or not they received medication. This effect on social function did not occur with amitriptyline alone and had not been evident at the end of the 16-week trial. Many patients from all treatments, however, reported requiring additional treatment over the follow-up year. This suggested that short-term treatment was insufficient for persistent response, a fact now recognized in many studies that include a maintenance phase.

The most ambitious short-term treatment study to-date, the multisite National Institute of Mental Health Treatment of Depression Collaborative Research Program (NIMH TDCRP), randomly assigned 250 depressed outpatients to 16 weeks of imipramine (Tofranil), interpersonal psychotherapy, cognitive-behavioral therapy, or placebo plus clinical management. Most subjects completed at least 15 weeks or 12 treatment sessions. Less-symptomatic patients improved in all treatments, including placebo. Imipramine induced the most rapid response and was most consistently superior to placebo. Interpersonal psychotherapy was comparable to imipramine on several outcome measures and showed a mean outcome superior to that with placebo for the more severely depressed patients. Cognitive-behavioral therapy was not superior to placebo for this group.

A reanalysis of the efficacy data from the NIMH TDCRP by sophisticated statistical techniques ordered treatment efficacy; medication was superior to psychotherapy, and the psychotherapies were superior to placebo, particularly among the symptomatic and impaired patients. This reassessment found cognitive-behavioral therapy to be inferior to interpersonal psychotherapy for patients with Beck Depression scores above approximately 30, which is generally considered the boundary between moderate and severe depression. These findings are consistent with those originally reported, but they sharpen differences between treatments.

An 18-month naturalistic, follow-up study of TDCRP patients found no significant difference in recovery among remitters (defined by the presence of minimal or no symptoms following the end of treatment and sustained during an 18-month follow-up) among the four treatment groups: 30 percent for cognitive-behavioral therapy 26 percent for interpersonal psychotherapy 19 percent for imipramine, and 20 percent for placebo. Among patients who had remitted by the end of the 16-week study, relapse over the 18-month follow-up was 36 percent for cognitive-behavioral therapy 33 percent for interpersonal psychotherapy, 50 percent for imipramine, and 33

percent for placebo. The authors concluded that 16 weeks of specific treatments were insufficient to achieve full and lasting recovery for many patients.

Hoencamp and colleagues are undertaking a study at the Hague in the Netherlands comparing interpersonal psychotherapy and nefazadone (Serzone) alone and in combination, for the acute treatment of major depression.

Continuation and Maintenance Treatment Interpersonal psychotherapy was first developed and tested in an 8-month, six-cell trial in 1974. Today this study would be considered a continuation rather than a maintenance treatment, as the concept of long-term treatment of depression has changed. Acutely depressed outpatient women (*N* = 150) who responded (>50 percent symptom reduction rated by a clinical interviewer) to a 4- to 6-week acute phase of amitriptyline were randomized to receive 8 months of treatment with weekly interpersonal psychotherapy alone, amitriptyline alone, combined interpersonal psychotherapy amitriptyline, interpersonal psychotherapy-placebo alone, or no pill. Randomization to interpersonal psychotherapy or a low contact condition occurred at entry into the continuation phase, whereas randomization to medication, placebo, or no pill occurred at the end of the second month of continuation. Maintenance pharmacotherapy was found to prevent relapse and symptom exacerbation, whereas interpersonal psychotherapy improved social functioning. The effects of interpersonal psychotherapy on social functioning were not apparent for 6 to 8 months. No negative treatment interactions were found, and combined psychotherapy-pharmacotherapy yielded the best outcome.

In the longest maintenance trial, Ellen Frank and colleagues studied 128 outpatients with recurrent depression, defined as a history of at least three acute major depressive episodes, who responded to 12 weeks of combined imipramine and interpersonal psychotherapy for the acute episode and 20 weeks of continuation treatment. Interpersonal psychotherapy was modified for this maintenance trial (interpersonal psychotherapy-M), primarily by emphasizing problems that persist or develop as a consequence of remission. Because of the length of treatment, multiple problem areas were addressed.

Following continuation treatment, patients were randomly assigned to receive 3 years of maintenance in one of five cells: monthly maintenance interpersonal psychotherapy, either (1) alone, (2) combined with imipramine, or (3) combined with placebo, or medication clinic visits for either (4) continuing high-dosage imipramine or (5) placebo, without further psychotherapy. This study was unique in using the highest dosage of medication ever in a maintenance trial (mean, 200 mg a day) and the lowest dose of interpersonal psychotherapy ever (monthly). This is also the only maintenance psychotherapy study of major depressive disorder to date.

Findings showed the highest significant effect for maintenance imipramine in preventing recurrence (survival time) and a modest but significant effect for interpersonal psychotherapy. After 1 year, major depression had recurred in 65 percent of placebo patients, 46 percent of those on interpersonal psychotherapy without imipramine, 17.9 percent of those on imipramine alone, and only 8 percent of those receiving interpersonal psychotherapy-imipramine, a trend supporting the superiority of combined treatment. At 3 years, mean survival without recurrence of depression was 45 weeks on placebo, 74 weeks on interpersonal psychotherapy plus placebo, 82 weeks for interpersonal psychotherapy alone, 124 weeks for imipramine alone, and 131 weeks for interpersonal psychotherapy-imipramine. Survival without recurrence was longer in patients who received interpersonal psychotherapy only than in those who did not (*P* = .043). The authors concluded that imipramine maintenance treatment at a dosage of 200 mg a day effectively prevents recurrence of major depressive episodes in patients with a history of recurrent episodes and that monthly interpersonal psychotherapy lengthens the time between episodes in patients not receiving medication. The 82-week survival time without recurrence with interpersonal psychotherapy alone for patients with a history of recurrence would suffice to protect many women, with recurrent depression, through pregnancy and nursing without medication.

Attempting to determine factors contributing to the efficacy of interpersonal psychotherapy, the investigators found that the quality of interpersonal psychotherapy, as measured by the interpersonal specificity and purity of the sessions (i.e., specific focus on interpersonal themes) but unrelated to the individual clinicians per se, was associated with longer time to recurrence. Audiotapes of therapy sessions were rated blind to the patient's clinical state and time since entry into the study. High specificity and purity of sessions were associated with a delay in recurrence of depression more than four times as long as would have been expected on the basis of the patient's history of recurrences. Patients whose therapy sessions were rated above the median on specificity of interpersonal psychotherapy had a median survival time of almost 2 years, whereas those below relapsed in less than 5 months.

A cost-effectiveness analysis based on this study showed that maintenance medication or interpersonal psychotherapy, either alone or in combination gave better expected lifetime health than placebo as measured by Quality Adjusted Life Years. A similar conclusion was reached comparing combined imipramine and interpersonal psychotherapy to interpersonal psychotherapy alone. In neither case were health care costs reduced.

The first phase of the maintenance trial was short-term–treatment study that used interpersonal psychotherapy for 12 to 24 weeks. Remitters, defined as having a Hamilton Rating Scale for Depression (HAM–D) score of 7 or below were entered into the maintenance study. Patients who failed to remit on interpersonal psychotherapy alone experienced higher levels of somatic anxiety and were more likely to meet diagnostic criteria for life time panic disorder. The authors concluded that concomitant anxiety can adversely affect the outcome of interpersonal psychotherapy for depression.

Nonremitters also had significantly worse subjective sleep quality and other indicators of disturbed sleep.

Further study required to determine whether the efficacy of interpersonal psychotherapy differs in patients receiving newer medications (e.g., selective serotonin reuptake inhibitors). Additional questions, including the effectiveness of more frequent than monthly interpersonal psychotherapy need to be addressed. A study of differing schedules of interpersonal psychotherapy for depressed women is under way in Pittsburgh.

Geriatric Depressed Patients The first use of interpersonal psychotherapy in geriatric depressed patients was not an efficacy study but an addition to a 6-week medication trial, to enhance compliance and to provide some treatment in the placebo control group. Noting that grief and role transition specific to life changes were the prime focus of treatment, the authors suggested interpersonal psychotherapy modifications including more flexible duration of sessions, greater use of practical advice and support (e.g., arranging transportation, calling the physician), and recognition that major role changes may be impractical and detrimental (e.g., divorce at age 75).

One 6-week clinical trial comparing interpersonal psychotherapy and nortriptyline (Aventyl) in 30 geriatric depressed patients showed some advantages for interpersonal psychotherapy, largely because medication adverse effects produced higher attrition in the medication group. Interpersonal psychotherapy was not modified for this study.

Charles Reynolds and colleagues in Pittsburgh conducted a 3-year maintenance study for geriatric patients with recurrent depression using a design similar to the study by Frank and coworkers. Interpersonal psychotherapy was modified for geriatric patients as interpersonal psychotherapy late-life maintenance treatment. During the acute phase, patients received combined nortriptyline and monthly interpersonal psychotherapy. This study then compared interpersonal psychotherapy late-life maintenance treatment alone, nortriptyline alone, interpersonal psychotherapy late-life maintenance treatment plus nortriptyline, interpersonal psychotherapy late-life maintence treatment and placebo, and placebo alone as randomized discontinuation maintenance therapies. The manual has been modified to allow more flexible session length, since elderly patients may not tolerate 50-minute sessions. Some of the authors found that older patients need to address early life relationships in their psychotherapy, distinguishing interpersonal psychotherapy late-life maintence treatment from the typical here-and-now focus of interpersonal psychotherapy. As in the earlier geriatric study, they found that therapists needed to help patients solve practical problems and should share awareness that some problems may not be amenable to resolutions (e.g., existential late-life issues or life-long psychopathology). They found that elderly depressed patients whose sleep quality normalized with early continuation of treatment had an 80 percent chance of remaining well during the first year of maintenance treatment. Response rates were similar in patients receiving nortriptyline or interpersonal psychotherapy. The final results in 187 patients over the age of 59 with recurrent depression found that maintenance treatment with nortriptyline or monthly interpersonal psychotherapy was superior to placebo in preventing or delaying recurrence. Combined treatment using both medication and psychotherapy was the optimal strategy in preserving recovery.

Bereavement-Related Depression A study comparing interpersonal psychotherapy and nortriptyline for short-term treatment of bereavement-related major depression is ongoing in Pittsburgh. The interpersonal psychotherapy modification includes more-detailed information gathering in the initial phase of treatment on the quality of earlier and current relationships and roles and determines available social supports for spousal bereavement. Detailed information on practical quality-of-life issues includes bill payment, financial burden, leisure activities, and children.

Depressed Adolescents Laura Mufson and coworkers modified interpersonal psychotherapy to incorporate adolescent developmental issues, adding as a fifth problem area the single parent family, an interpersonal situation found frequently among their adolescents. Parents are also involved in the initial phase, and their permission is obtained for treatment. Telephone contacts are readily used, and the school is involved when appropriate. In an open trial, 14 patients were significantly

less depressed and had improved functioning after 12 weeks of treatment. At 12 weeks, none met criteria for revised third edition of DSM (DSM-III-R) major depression. A follow-up of 10 patients found that only 1 met criteria for an affective disorder at 1 year. There were no hospitalizations, suicide attempts, or pregnancies, and all remained in school, even though about a third had experienced serious negative life events since completing treatment: violence, physical or sexual abuse, parental psychopathology, or the death of a parent. The small sample of adolescents of low socioeconomic status limits generalizability. A larger, controlled clinical trial recently has been completed and has demonstrated the efficacy of interpersonal psychotherapy in symptom reduction of improving social function.

Depressed Adolescent Mothers Laune Gillies at Clark Institute, Toronto, is completing a pilot study of depressed pregnant adolescents (ages 15 to 19). Thirty patients, all scoring 14 or higher on the Beck Depression Inventory (BDI), were randomly assigned to 12 weekly sessions of arts and crafts, a psychoeducational group, or individual interpersonal psychotherapy. Eleven subjects terminated early, with attrition approximately equal across groups. Results are pending.

Depressed Human Immunodeficiency Virus (HIV)-Positive Patients (IPT-HIV) The second author and coworkers modified interpersonal psychotherapy for depressed HIV patients, emphasizing common issues in this population including concern about illness and death, grief, and role transitions. In a pilot open trial, 21 of the 24 depressed patients responded with symptom reduction. Most patients found brief therapy suitable to their situation, in which time was precious and quick results needed. A 16-week randomized clinical trial has been completed comparing interpersonal psychotherapy with the modified cognitive-behavioral therapy, supportive psychotherapy, and imipramine plus supportive psychotherapy. Treatment of 101 patients showed that while symptoms decreased across therapies, interpersonal psychotherapy and imipramine resulted in differential improvement relative to supportive psychotherapy or cognitive-behavioral therapy. These results resemble outcomes for the more symptomatic patient subgroup in the TDCRP.

Depressed Primary Care Patients H.C. Schulberg and colleagues completed a clinical trial comparing interpersonal psychotherapy with pharmacotherapy for depressed ambulatory medical patients in a primary care setting. The interpersonal psychotherapy manual was not modified, but interpersonal psychotherapy conformed with practices of the primary care center; for example, nurses took vital signs prior to each session. If a patient was medically hospitalized, interpersonal psychotherapy continued in the hospital when possible.

Patients with current major depressive disorder (*N* = 276) were assigned to either interpersonal psychotherapy nortriptyline, or primary care physicians' usual care. They were seen weekly for 16 weeks and monthly thereafter for 4 months in interpersonal psychotherapy. Depressive severity declined more rapidly with either nortriptyline or interpersonal psychotherapy than in usual care. Among treatment completers, approximately 70 percent receiving nortriptyline or interpersonal psychotherapy but only 20 percent in usual care were judged recovered at 8 months. Patients with lifetime comorbid panic disorder had poorer recovery than those with major depressive disorder alone, regardless of treatment. Interestingly, a post hoc analysis found that severely depressed patients (Hamilton Rating Scale for Depression Score ^{3} 20) did equally well on interpersonal psychotherapy and nortriptyline. However, less severely depressed patients responded more rapidly to medication as compared to interpersonal psychotherapy during the first 3 months.

Conjoint Interpersonal Psychotherapy for Depressed Patients With Marital Disputes Since marital conflict, separation, and divorce have been associated with the onset and course of depressive episodes and individual psychotherapy for depressed patients in marital disputes may lead to termination of some marriages, a manual was developed for conjoint therapy of depressed patients with marital disputes. This modified interpersonal psychotherapy focuses on the current marital dispute and involves the spouse in all sessions. Eighteen patients with major depressive disorder linked to the onset or exacerbation of marital disputes were randomly assigned to 16 weeks of individual interpersonal psychotherapy or conjoint interpersonal psychotherapy. Patients in both treatments showed the same reduction in depressive symptoms. Patients in conjoint interpersonal psychotherapy had significantly better marital adjustment, greater marital affection, and better sexual relations than did patients receiving interpersonal psychotherapy alone. These preliminary findings require replication with a larger sample and other control groups.

Antepartum and Postpartum Depression Margaret Spinelli is using interpersonal psychotherapy to treat women with antepartum depression at Columbia University. Examination of this role transition focuses on the pregnant woman's self-evaluation as a parent, physiological changes of pregnancy, and altered relationships with spouse or significant other and other children. A fifth problem area, complicated pregnancy, has been added. Timing and duration of sessions require flexibility because of bed rest, delivery, obstetrical complications, and child care. Young children may be brought to sessions, where postpartum mothers may breast feed. Telephone sessions and hospital visits may be necessary. Spinelli has published results of an open trial and is undertaking a controlled clinical trial comparing interpersonal psychotherapy with didactic parent education groups in depressed pregnant women over a 16 week period of treatment and 6 monthly follow-up sessions.

Scott Stuart and Michael O'Hara have a trial under way comparing interpersonal psychotherapy and a waiting list control for women with postpartum depression. They assess both the mothers' symptom state and interactions with their infant.

Dysthymic Disorder (IPT-D) In a modification of interpersonal psychotherapy for dysthymic disorder, patients are encouraged to reconceptualize what they had seen as their lifelong character flaws as ego-dystonic, chronic mood-dependent symptoms. Contrary to expectation, open treatment found that dysthymic disorder patients with lifelong chronicity had a reduction of depressive symptoms in 16 weekly interpersonal psychotherapy sessions. Markowitz treated 17 pilot subjects; none worsened, and 11 remitted. Medication benefits roughly half of dysthymic patients, but nonresponders may need psychotherapy, and even medication responders may benefit from combined treatment. Based on these pilot results, a comparative study of 16 weeks of modified interpersonal psychotherapy for dysthymic disorder alone, supportive psychotherapy, sertraline plus clinical management, and interpersonal psychotherapy and sertraline, is under way at Cornell Medical Center.

Gina Browne and Meu Steiner at McMaster University in Hamilton, Canada have a 20-month study of approximately 700 dysthymic disorder patients receiving unmodified interpersonal psychotherapy or sertraline alone or in combination for 1 year. Preliminary results at 6 months showed that sertraline (Zoloft) alone had better compliance than interpersonal psychotherapy alone and was equally effective to sertraline plus interpersonal psychotherapy and more effective than interpersonal psychotherapy alone, as measured by symptom reduction. However, patients who received interpersonal psychotherapy either alone or in combination with sertraline had economically important containment of or reduction in expenditures from use of health and social services and lower expenditures for social assistance. One and 2-year follow-up studies are planned to assess the enduring nature of these effects and the impact, if any, on family functioning.

A trial is under way in Toronto comparing interpersonal psychotherapy with the short-term psychodynamic therapy of Lester Luborsky in 72 patients with dysthymic disorder with or without major depressive disorder (double depression) for 12 weekly sessions followed by 4-month monthly sessions. Initial results of interpersonal psychotherapy indicate that most patients reported a reduction of symptoms. The authors found that dysthymic disorder patients responded to interpersonal psychotherapy just as did patients with major depressive disorder, but that because of the longstanding nature of their illness, patients with dysthymic disorder were more difficult to treat.

Bipolar Disorder E. Frank and colleagues in Pittsburgh are assessing the benefits for patients with bipolar disorders of adjunctive interpersonal psychotherapy modified by social zeitgeber theory—behavioral scheduling of daily sleep patterns—as maintenance treatment of lithium-stabilized bipolar disorder patients. Comparing bipolar disorder interpersonal psychotherapy with medication alone, the 3-year maintenance treatment study will initially include biweekly interpersonal psychotherapy visits, tapering to monthly sessions in the final 2 years.

Substance Abuse and Dependence Interpersonal psychotherapy has not been shown to be effective in two clinical trials with substance abuse patients. One study, complicated by high attrition, found no benefit in reducing psychopathology for adjunctive interpersonal psychotherapy added to standard treatment, compared with the standard program alone, for 72 methadone (Dolophine) maintained opioid abusers. The same team in a separate trial found 12 weeks of interpersonal psychotherapy ineffective and marginally worse than behavioral treatment for 42 cocaine abusers attempting to achieve abstinence. The two negative studies suggest limits to the utility of interpersonal psychotherapy but do not necessarily doom its use for substance abuse. For example, interpersonal psychotherapy might be useful for newly abstinent, recovering alcohol-dependent patients, who face numerous psychosocial stressors that precipitate relapse.

Bulimia Nervosa Christopher Fairburn and coworkers altered interpersonal psychotherapy for two studies of bulimia nervosa patients. They used fewer, differently scheduled sessions, primarily to allow comparison with cognitive-behavioral therapy. Except during the assessment phase, when a history of the eating disorder is taken, the content of the interpersonal psychotherapy approach resembles that for depression. Discussion of the patient's eating disorder is avoided during the course of treatment. In fact, interpersonal psychotherapy therapists were instructed not to let talk of bulimia nervosa continue for more than 10 seconds before refocusing the patient on interpersonal issues.

Thus unlike interpersonal psychotherapy for depression, which emphasizes a medical model of depression as illness, interpersonal psychotherapy for bulimia nervosa

actively avoids the diagnosis. This makes clinical sense; depressed patients tend not to focus on depression as an illness, typically confusing it with their personal failings, whereas bulimia nervosa patients are all too symptomatically aware and tend to talk of nothing else. It may help bulimia nervosa patients to shift focus from their symptoms to address their current life situation, stressors, and interpersonal interactions.

The initial study randomized 24 patients to 19 sessions of either cognitive-behavioral therapy or interpersonal psychotherapy over 18 weeks and demonstrated the comparability of the treatments. In a second study, 75 patients with bulimia nervosa were randomly assigned to cognitive-behavioral therapy, interpersonal psychotherapy, or behavioral treatment for 19 sessions over 18 weeks, with a 1-year follow-up. Few patients in behavioral treatment ceased binge eating and purging. Patients in cognitive-behavioral therapy and interpersonal psychotherapy made equivalent, substantial changes across all symptom domains, with interpersonal psychotherapy taking longer to achieve its effects.

A 6-year follow-up study was conducted of 99 patients from both trials. Patients who received behavioral treatment did poorly (86 percent still had an eating disorder), whereas those who received either interpersonal psychotherapy or cognitive-behavioral therapy had better prognoses: 28 to 37 percent met eating disorder criteria. The authors are now experimenting with interpersonal psychotherapy supplemented by a cognitive-behavioral therapy self-help program for bulimia nervosa. A replication study is comparing cognitive-behavioral therapy and interpersonal psychotherapy in bulimia nervosa at Stanford and Columbia in collaboration with investigators at Oxford University and Rutgers University.

Group Format for Bulimia Nervosa Drawing on the work of Christopher Fairburn, Denice Wilfley and coworkers modified interpersonal psychotherapy in a 16-weekly session group format and compared it with group cognitive-behavioral therapy and a waiting list control for 56 women with nonpurging bulimia nervosa. At termination, both group interpersonal psychotherapy and cognitive-behavioral therapy significantly reduced binge eating, but the waiting list did not. These results persisted at 1-year follow-up. A randomized clinical trial of 162 women is now comparing group interpersonal psychotherapy and cognitive-behavioral therapy for 20 sessions over 20 weeks. The initial phase of interpersonal psychotherapy, in which the problem area is identified and the interpersonal psychotherapy concepts and contract presented, is conducted individually. Groups meet for 90 minutes.

Anxiety Disorders The interpersonal aspects of social phobia make it a natural starting place for testing interpersonal psychotherapy as a treatment of anxiety disorders. Interpersonal psychotherapy is being modified for social phobia independently by Joshua Lipsitz at Columbia University and Stuart and O'Hara at the University of Iowa. Open trials are in progress at both sites. Lipsitz, having completed 9 pilot cases, reports that the standard interpersonal psychotherapy ingredients, including the medical model, provision of the sick role, and the supportive therapeutic stance, clearly appear to benefit most patients. The interpersonal formulation, linking anxiety symptoms to interpersonal context, appeals more to some patients than others. For example, a patient with prominent blushing insisted her social anxiety was due to the blushing and unrelated to her interactions with other people or her ability to express emotion freely. Interpersonal psychotherapy seemed to help, nonetheless.

Many social phobic patients present in role transitions: geographic moves, career changes, change in family constellation due to separation from significant others, or the onset of a major illness. Social anxiety may increase in intensity or may simply pose a greater problem in the context of these transitions.

Because social phobia often has an early onset and chronic course, lacking acute precipitants, Lipsitz borrowed an approach from the work of the second author on interpersonal psychotherapy for dysthymic disorder. Providing the medical model of social phobia, offering reassurance that the patient has a disorder that is treatable, and instilling hope for change, the therapist initiates a therapeutic role transition. Patients recognize shyness, passivity, and social awkwardness as symptoms of a disorder, rather than their true personality. This formulation creates a positive stressor that serves as a frame for exploring current interpersonal experiences (“It sounds like you can be quite assertive with your staff now that you're not letting the social phobia get in the way”).

Lipsitz has categorized as “role insecurity” subtler role difficulties than those defined by role deficit—lack of assertiveness, conflict avoidance, difficulty expressing anger, rejection sensitivity—often seen in social phobic patients. This formulation emphasizes the patient's potential for more-competent behavior once social phobia is mitigated. Role playing may be helpful in teaching or reinforcing social skills.

Stuart notes two groups of socially phobic patients: those with avoidant features who desire relationships but have difficulty maintaining them and those with schizoid traits who want to minimize social interactions. The latter seek treatment only when circumstance requires them to interact more frequently; for example, a librarian with no psychiatric history who sought treatment upon receiving a promotion. Her new job required supervising other employees, whereas before she had contentedly shelved and catalogued books with minimal social contacts. Addressing this as a role transition helped her through the acute stress, although Stuart was skeptical of the long-term benefit of treatment.

Social Phobia in a Group Format Weissman and Bonnie Jacobson have adapted interpersonal psychotherapy in a group format for patients with shyness. The patients had social phobia in unstructured interpersonal situations (e.g., parties, intimate discussions with significant others) but not in defined work situations. Most patients were highly successful in professional or business careers despite their phobias.

The 10-session, time-limited group focused on defining and describing the diagnosis, giving the patient the sick role, and finding practical strategies for dealing with shyness in specific situations (e.g., developing scripts to initiate a more personal conversation with an estranged father, or a discussion with a spouse about having a baby). As noted by Lipsitz, the chronicity of the disorder led to a focus on a therapeutic role transition from an impaired to a less-impaired state. The group format seemed to provide a safe haven in which patients could interact with others who had similar symptoms.

Panic Disorder A manual for interpersonal psychotherapy in panic disorder is being developed by Arzt and Meriam van Rijsoort in Maastricht, the Netherlands.

Other Axis I Disorders David Veale in London is conducting a 15-week clinical trial comparing cognitive-behavioral therapy with interpersonal psychotherapy for patients with body dysmorphic disorder and a preoccupation with an imagined defect in appearance that causes distress in social functioning.

George Ikkos has modified interpersonal psychotherapy to manage patients with chronic somatization in primary care. This adaptation adds a fifth problem area, the patient's relationship with health professionals and the pursuit of medical care. The interpersonal psychotherapy medical model works readily with these patients because of their inappropriate use of the health care system. Treatment seeking is reconceived as an interpersonal issue. These patients are easily recruited for interpersonal psychotherapy treatment; however, 12 sessions may be insufficient to engage patients and develop a working alliance. Therapists emphasize that they are not trying to modify the patient's experience of pain, but to help the patient deal more effectively with the problem. An open trial of 20 to 30 patients is planned.

Borderline Personality Disorder Interpersonal psychotherapy has been adapted for patients with borderline personality disorder. The focus in the initial phase is on assessment of symptom patterns related to the disorder, such as anger and impulsivity in interpersonal relations. A fifth problem area, self-image, has been added to address the identity disturbance that is central to borderline personality disorder. Telephone contact is used during crises but not encouraged in lieu of regular scheduled sessions. A pilot randomized trial on 24 patients is under way comparing 12 weekly sessions of interpersonal psychotherapy with relationship management therapy, with monthly follow-up for 6 months. The trial uses a sequential treatment/therapist crossover design to control for therapist effects.

Initial results show lower attrition in interpersonal psychotherapy (10 percent) than in the control group (50 percent). Comparison of prescores and postscores of patients in interpersonal psychotherapy with borderline personality and those with dysthymia or double depression found similar and significant symptomatic improvement on self-report scales across diagnostic groups. It will be interesting to see whether time-limited psychotherapy alleviates a time-resistant disorder such as borderline personality.

Other Disorders

Insomnia A modification for interpersonal psychotherapy for patients with insomnia has been developed by Elizabeth Schramm in Hamburg, Germany. This research arose from the observation that insomnia is often associated with stressful interpersonal life events. The approach emphasizes management of insomnia and regularizing social rhythms adapted from work with bipolar disorder patients. The initial phase presents information on sleep hygiene and rhythm, and patients keep a sleep diary.

Metastatic Breast Cancer Interpersonal psychotherapy is being delivered over the telephone to housebound patients who have metastatic breast cancer by the group at Memorial Sloan-Kettering Cancer Center in New York (see below). A similar study is under way at Clarke Institute, Toronto.

Post-Myocardial Infarction S. Stuart has modified interpersonal psychotherapy for patients with depression after myocardial infarction. The adaptation includes careful initial medical evaluation to ensure that symptoms of insomnia, anergia, and other vegetative symptoms represent clinical depression rather than cardiac illness. Stuart notes that care must be taken to ensure that the physical symptoms exceed those expected during this time period. Medication monitoring must ensure both that cardiac medications are not producing adverse psychiatric effects and that psychotropic medications do not have adverse cardiac effects.

Interpersonal psychotherapy confronts grief over loss of the healthy state, struggles with mortality, remorse over past unhealthy behavior, and the effects of the illness on marital relations and work. Another theme is being forced into the role of receiving care after a life of independence. The authors note that no other trials of counseling with post-myocardial infarction patients have focused on depression. They report that although these patients acknowledge depression, they often politely refuse psychological intervention, even if treatment is labeled “stress reduction” or “coping skills.”

Interpersonal Counseling Many patients presenting to primary care practices report psychiatric symptoms but do not meet full criteria for a psychiatric disorder. These symptoms can be debilitating and result in fruitless, expensive workups and high utilization of medical procedures. Interpersonal counseling based on interpersonal psychotherapy was designed for distressed primary care patients who do not meet syndromal criteria for psychiatric disorders. Interpersonal counseling is administered by health care professionals, usually nurse practitioners without formal psychiatric training, for a maximum of six sessions. The first session can last as long as 30 minutes; subsequent sessions are briefer.

Interpersonal counseling therapists assess the patient's current functioning, recent life events, occupational and familial stress, and changes in interpersonal relationships. They assume that such events may provide the context in which emotional and bodily symptoms occur. Some 128 patients presenting to a primary care clinic and scoring 6 or higher on the Goldberg General Health Questionnaire (GHQ) were randomized either to interpersonal counseling or to usual care without psychological treatment. Over an average of 3 months, often involving only one to two sessions, interpersonal counseling subjects showed significantly greater symptom relief than controls on the GHQ, particularly improvement in depressed mood. Interpersonal counseling led to greater use of mental health services by patients newly attuned to the psychological source of their symptoms.

Subsyndromally Depressed Hospitalized Elderly Patients Observing that depressive symptoms that did not reach criteria for major depressive episode nonetheless impeded the recovery of hospitalized elderly patients, Jana Mossey and coworkers conducted a 10-session trial of interpersonal counseling with elderly hospitalized medical patients who had depressive symptoms. They modified the technique by increasing the number of sessions from 6 to 10, their duration from 30 to 60 minutes, and increasing the flexibility of scheduling from once weekly to a schedule that reflected the patient's medical status. They randomized 76 hospitalized patients over age 60 who did not meet criteria for major depression but had depressive symptoms on two consecutive assessments, to receive either interpersonal counseling administered by clinical nurse specialists or usual care. A nondepressed control group was also followed. Patients found interpersonal counseling feasible and acceptable. Assessment at 3 months showed a nonsignificantly greater reduction in depressive symptoms and greater improvement on all outcome variables for interpersonal counseling than for usual care, versus a slight symptomatic increase among controls. Rehospitalization in the interpersonal counseling and nondepressed control groups was virtually identical (11 to 15 percent), and significantly below that in the subsyndromally depressed group receiving usual care (50 percent). At 6 months, differences between the treatment groups reached statistical significance in reduction of depressive symptoms and self-rated health but not in physical or social functioning. A 1-year evaluation is pending. The investigators felt that 10 sessions were insufficient for some patients and that a maintenance phase might have been useful. They judged the clinical nurse specialists acceptable therapists.

Interpersonal Psychotherapy by Telephone Because many patients avoid or have difficulty reaching an office for face-to-face treatment, interpersonal counseling is being tested as a telephone treatment. An open trial under way at Memorial Sloan-Kettering Cancer Center in New York offers interpersonal psychotherapy to severely medically, but not psychiatrically, ill patients with metastatic breast cancer who are receiving high-dosage chemotherapy. Both the patient and partner receive weekly 30-minute sessions for approximately 15 weeks. All patients have metastatic breast cancer considered incurable by standard treatments. Their chemotherapy regimen is unusually disabling, and patients must move to within minutes of the hospital for the 2 to 3 months of treatment, causing major disruptions for themselves and their families. They must take extended leaves from jobs and leave children and spouses at home. The purpose of the study was to adapt interpersonal psychotherapy to relieve cancer-related stress rather than depression, but prevalence of depression and depressive symptoms is high. The investigators report that interpersonal psychotherapy fits a wide range of the patients' needs well.

The first author and Lisa Miller at Columbia University have a pilot telephone trial under way comparing interpersonal psychotherapy with no treatment in 30 patients who have had recurrent depression but have not received regular treatment.

Interpersonal Psychotherapy Patient Guide The first author has developed a user-friendly interpersonal psychotherapy patient guide with accompanying worksheets. The guide is designed for patients who want to learn about or are receiving interpersonal psychotherapy. It leads the patient through treatment in simple language. Worksheets can be used to facilitate sessions or to monitor problem areas after treatment. Formal testing to determine whether the patient book facilitates treatment has not been done. Informal reports by Gillies and the second author note that patient response has been positive, and that some therapists have found the patient guide useful during training in interpersonal psychotherapy.

Training in Interpersonal Psychotherapy Until recently, interpersonal psychotherapy practitioners were few and almost exclusively limited to participants in research studies. Interpersonal psychotherapy training is now increasingly included in professional workshops and conferences, with training courses conducted at university centers in Canada, Europe, Asia, and New Zealand. A training videotape (Kingsley Communications, Houston, Texas) describes interpersonal psychotherapy and demonstrates the initial assessment phase of treatment.

Training workshops for mental health professionals from a variety of disciplines have been held at Cornell Medical School, New York; the University of Pittsburgh; and in Toronto. Interpersonal psychotherapy is taught in some (but not most) residency training programs and has been included in family practice and primary care training. Interpersonal psychotherapy clinics have been established at the Clarke Institute in Toronto and at New York Hospital-Cornell Medical Center in New York City.

Although the principles of interpersonal psychotherapy are straightforward, training requires more than reading the manual. Candidates should have a graduate clinical degree (M.D., Ph.D., M.S.W., R.N.) and several years of experience conducting psychotherapy. Interpersonal psychotherapy training programs are designed to help experienced therapists refocus their treatment by learning new techniques, not to teach novices psychotherapy. The model program includes a brief didactic phase in which the manual is reviewed, and a longer practicum during which the therapist treats two to three patients under close supervision, monitored by videotapes of the sessions. Therapists who performed well on a first supervised case often did not require further intensive supervision, and experienced therapists committed to the approach required less supervision.

Future Directions Evidence from controlled clinical trials suggests that interpersonal psychotherapy is a reasonable alternative or adjunct to medication as an acute, continuation, or maintenance treatment for patients with major depression; for acutely depressed HIV-positive patients, and for patients with bulimia nervosa. It is a promising, but still not fully tested, treatment for depressed adolescent and geriatric patients, patients with dysthymic disorder, and primary care patients with mild depression and a couples treatment for depressed patients with marital disputes. More efficacy data are needed before stronger claims can be made. Open trials are encouraging but cannot be considered evidence of efficacy. Interpersonal psychotherapy is not effective (compared with standard treatment) for opioid- and cocaine-dependent patients. Several interpersonal psychotherapy clinical trials based on modified treatment manuals are under way or planned, and several manuals are under development. The manuals differ in the depth of their modifications, and some adaptations for particular disorders have not been formalized. There is little agreement on what a manualized interpersonal psychotherapy adaptation should include.

The development of psychotherapy manuals represents progress. They will meet growing requests from third-party payers, managed care, and patients for accountability and specification of treatment. Manuals are a necessary first step in testing efficacy. In the authors' opinion psychiatric and other mental health treatment training programs should include clinical instruction in time-limited, manual-defined psychotherapies in addition to exposing trainees to long-term psychotherapy.

Even for experienced therapists, interpersonal psychotherapy training programs are still not widely available, but access is growing. The established training criteria for research have produced reliable therapists for clinical trials. Yet the educational process for interpersonal psychotherapy in clinical practice requires further study. What educational level and experience does a clinician require to learn interpersonal psychotherapy? How much supervision does an experienced therapist require? Will reading the manual suffice?

Interpersonal psychotherapy research to date has focused largely on outcome. With some effective outcomes now demonstrated, process research is worthwhile to determine specific ingredients of the treatment. Dismantling studies might be useful. For example, the initial diagnostic and assessment phase of interpersonal psychotherapy is essential to the clinical management of all depressed patients. The efficacy of using only the initial interpersonal psychotherapy phase might be compared with the full treatment. Data from the study by Frank and coworkers: showed the value of monthly maintenance interpersonal psychotherapy for patients with recurrent depression, but a series of studies might compare weekly, biweekly, and monthly frequency of sessions. Such studies would be comparable to phase II or III pharmacotherapy trials.

It is unclear how efficacy data for psychotherapies like interpersonal psychotherapy will be used in managed care reimbursement decisions. Will standardization of treatment, training credentials, and evidence of effectiveness and cost offset be required? Since clinical trials usually exclude the patients with complex comorbidity seen in clinical practice, reimbursement policy based on data from clinical trials should adjust for this contingency. It will be important to learn how the efficacy of interpersonal psychotherapy in clinical research translates into effectiveness when used by general clinicians. It is unclear whether untested psychotherapies will be reimbursed, whether psychotherapy will be reimbursed at all, and if so with what limits on intensity and duration of treatment. Strict limits on psychotherapy reimbursement would harm many patients (e.g., recurrent depressive patients unresponsive to medication, or depressed pregnant or nursing women). The efficacy of antidepressant pharmacotherapy is unequivocal, yet medicating depressed patients without psychological management risks poor compliance and persistent unresolved social stressors. The best interests of all psychiatric patients are ensured by access to a range of treatment modalities whose efficacy has been established in controlled clinical trials.

SUGGESTED CROSS-REFERENCES

Substance-related disorders are covered in Chapter 11, mood disorders in Chapter 14, anxiety disorders in Chapter 15, somatoform disorders in Chapter 16, eating disorders in Chapter 20, personality disorders in Chapter 24, and relational problems in Chapter 26. Evaluation of psychotherapies is discussed in Section 30.11. Combined psychotherapy and pharmacotherapy is covered in Section 30.12. Tricyclic drugs are discussed in Section 31.30.

SECTION REFERENCES

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB: National Institute of Mental Health treatment of depression collaborative research program: General effectiveness of treatments. Arch Gen Psychiatry *46:*971, 1989.

Fairburn CG, Norman PA, Welch SL, O'Connor ME, Doll HA, Peveler RC: A prospective study of outcome in bulimia nervosa and the long-term effects of three psychological treatments. Arch Gen Psychiatry *52:*304, 1995.

Foley SH, Rounsaville BJ, Weissman MM, Sholomskas D, Chevron E: Individual versus conjoint interpersonal psychotherapy for depressed patients with marital disputes. Int J Fam Psychiatry *10:*29, 1989.

Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ: Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry *47:*1093, 1990.

Frank E, Kupfer DJ, Wagner EF, McEachran AB, Cornes C: Efficacy of interpersonal psychotherapy as a maintenance treatment of recurrent depression. Arch Gen Psychiatry *48:*1053, 1991.

*Klerman GL, Budman S, Berwick D, Weissman MM, Damico-White J, Demby A, Feldstein M: Efficacy of brief psychosocial intervention for symptoms of stress and distress among patients in primary care. Med Care *25:*1078, 1987.

Klerman GL, DiMascio A, Weissman MM, Prusoff BA, Paykel ES: Treatment of depression by drugs and psychotherapy. Am J Psychiatry *131:*186, 1974.

*Klerman GL, Weissman MM: *New Applications of Interpersonal Psychotherapy*. American Psychiatric Press, Washington, DC, 1993.

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES: *Interpersonal Psychotherapy of Depression*. Basic Books, New York, 1984.

*Markowitz JC: Interpersonal psychotherapy. In *Review of Psychiatry* vol 17, JM Oldham, MB Riba, editors, American Psychiatric Press, Washington, DC, 1998.

Markowitz JC: Psychotherapy of dysthymia. Am J Psychiatry *151:*1114, 1994.

Markowitz JC: *Interpersonal Psychotherapy for Dysthymic Disorder*. American Psychiatric Press, Washington, DC, 1998.

Markowitz JC, Klerman GL, Clougherty KF, Spielman LA, Jacobsberg LB, Fishman B, Frances AJ, Kocsis JH, Perry SW: Individual psychotherapies for depressed HIV-positive patients. Am J Psychiatry *152:*1504, 1995.

Markowitz JC, Kocsis JH, Fishman B, Spielman LA, Jacobsberg LB, Frances AJ, Klerman GL, Perry SW: Treatment of HIV-positive patients with depressive symptoms. Arch Gen Psychiatry *55:*452, 1998.

Markowitz JC, Svartberg M, Swartz HA: Is IPT time-limited psychodynamic psychotherapy? J Psychother Pract Res *7:*185, 1998.

Mossey JM, Knott KA, Higgins M, Talerico K: Effectiveness of a psychosocial intervention, interpersonal counseling, for subdysthymic depression in medically ill elderly. J Gerontol *51A:*M172, 1996.

Mufson L, Moreau D, Weissman MM, Klerman GL: *Interpersonal Therapy for Depressed Adolescents*. Guilford, New York, 1993.

Reynolds CF, Frank E, Kupfer DJ, Thase ME, Perel JM, Mazumdar S, Houck PR: Treatment outcome in recurrent major depression: A post hoc comparison of elderly (“young old”) and midlife patients. Am J Psychiatry *153:*1288, 1996.

*Reynold CF III, Frank E, Perel JM, Imber SD, Cornes C, Miller MD, Mazumdar S, Houck PR, Dew MA, Stack JA, Pollock BG, Kupfer DJ: Nortriplyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: A randomized controlled trial in patients older than 59 years. JAMA *281:*39, 1999.

Rounsaville BJ, Glazer W, Wilber CH, Weissman MM, Kleber HD: Short-term interpersonal psychotherapy in methadone-maintained opiate addicts. Arch Gen Psychiatry *40:*629, 1983.

Schulberg HC, Block MR, Madonia MJ, Scott CP, Rodriguez E, Imber SD, Perel J, Love J, Houck PR, Coulehan JL: Treating major depression in primary care practice. Arch Gen Psychiatry *53:*913, 1996.

Shea MT, Elkin I, Imber SD, Sotsky SM, Watkins JT, Collins JF, Pilkonis PA, Beckham E, Glass DR, Dolan RT, Parloff MB: Course of depressive symptoms over follow-up: Findings from the National Institute of Mental Health Treatment for Depression Collaborative Research Program. Arch Gen Psychiatry *49:*782, 1992.

Spinelli MG: Interpersonal psychotherapy for depressed antepartum women: A pilot study. Am J Psychiatry *154:*1028, 1997.

Stuart S, Cole V: Treatment of depression following myocardial infarction with interpersonal psychotherapy. Ann Clin Psychiatry *8:*203, 1996.

Weissman MM: *Mastering Depression through Interpersonal Psychotherapy: A Patient Guide Psychotherapy*. Psychological Corporation, San Antonio, TX, 1-800-228-0752, 1995.

*Weissman MM, Markowitz JC, Klerman GL: *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books, New York, 1999.

Weissman MM, Prusoff BA, DiMascio A, Neu C, Goklaney M, Klerman GL: The efficacy of drugs and psychotherapy in the treatment of acute depressive episodes. Am J Psychiatry *136*:555, 1979.

Wilfley DE, Agras WS, Telch CF, Rossiter E, Schneider J, Cole AC, Sifford L, Raeburn S: Group cognitive-behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A controlled comparison. J Consult Clin Psychol *61:*296, 1993.

30.8 BRIEF PSYCHOTHERAPY

ROBERT J. URSANO, M.D., AND ANN E. NORWOOD, M.D.

Psychotherapy is the generic term for a large number of treatment techniques that are designed to change behavior through verbal interchange (talking). The psychodynamic psychotherapies—the brief psychodynamic psychotherapies, interpersonal psychotherapy and cognitive psychotherapy—aim to change behavior through the reorganization of mental structures and processes. These changes result from new understandings and the new recognition of old patterns of behavior that have been long enacted but not observed. Through this process an individual's perception, expectation, and behavior can change.

In the past, the terms “brief psychotherapy” and “long-term psychotherapy” were frequently used synonymously with *supportive* and *explorative* psychotherapy, respectively. However, this is no longer accurate. *Brief* and *long-term* describe only the duration, not the technique, focus, or goal. The brief psychotherapies have developed their own theory of technique and goals. Brief psychotherapy is primarily distinguished from longer-term treatments by the time limits placed upon it. These time limits give the brief psychotherapies unique characteristics and distinguish them from long-term psychotherapy and psychoanalysis. The time limits also mean that brief psychotherapists must confront their own ambitiousness and perfectionism as well as any exaggerated ideal of personality structure and function.

Interest in brief psychotherapy has flourished in recent years. Psychotherapy is a verbal interchange between two individuals, one an expert and the other a help-seeker, in this case, the patient. The goal of their work together is to identify the patient's characteristic problems in living so they could be changed. As in all of medicine, both nonspecific and specific factors related to the treatment affect the outcome of this work. The nonspecific curative factors of abreaction, the provision of new information, and maximizing success experiences, guide all forms of medical treatment including the brief psychotherapies. These principles increase the probability that patients will experience relief of pain and suffering. The individual psychotherapies also have specific technical interventions and procedures above and beyond the nonspecific curative factors. As in other medical treatments, there are contraindications and dangers in the use of psychotherapy.

It is not sufficient in the present-day era of managed care and closely watched outcome measures to state that psychotherapy was a success but no change occurred. Although the therapist in the individual psychodynamically derived brief psychotherapies does not require behavioral change, the end result of the therapist's technical expertise is to achieve behavioral change, including changes in well-being, physical health, social supports, and societal productivity as well as symptomatic relief.

HISTORY

Following World War II, interest in psychoanalysis resulted in a rapid growth in the demand for psychotherapy. This considerably increased the pressure on psychiatrists to develop briefer forms of psychotherapy. In the mid 1940s, Franz Alexander and Thomas French made important contributions to the development of brief dynamic psychotherapy. In their report on the Chicago Institute of Psychoanalysis's research project, they advocated shortening treatment and seeing patients less frequently to minimize regression. They proposed focusing treatment on the present rather than the past, using historical conflicts to inform the therapist in providing the best corrective emotional experience for the patient in the present. The goal of the treatment was to provide a complete psychoanalysis in a shorter period of time. Alexander's work, in particular, has been perhaps the most influential in the development of brief psychodynamic psychotherapy.

More recently, the community mental health movement of the 1960s, the increasing cost of mental health care, and the current rise of managed care have stimulated efforts to find briefer forms of psychotherapy. Brief psychotherapy is a necessary part of the modern psychiatrist's armamentarium. The brief psychodynamic psychotherapies of the past three decades have not attempted a complete analysis but concentrate on the analysis of a focal conflict to the exclusion of others. Contemporary brief psychodynamic psychotherapy was heavily influenced by the British School's development of brief focal psychotherapy. Michael Balint sponsored a workshop of experienced psychoanalytic psychotherapists that initially focused on clinical evaluation and attempted to determine which patients might be suitable for deviations from the standard psychotherapy of the time toward a briefer treatment. After Balint's death, David Malan carried on the work of the group. Malan developed and applied the principles of psychodynamic treatment to brief treatment, delineating methods for evaluating process and outcome variables. He emphasized the importance of therapeutic planning and the identification of a focal conflict.

At the same time David Malan was undertaking his research at the Tavistock Clinic, Peter Sifneos was studying brief psychotherapy at the Massachusetts General Hospital in Boston. Many of his conclusions are similar to Malan's. However, there are some differences. Sifneos developed what he called short-term anxiety-provoking psychotherapy as a technique and theory. He emphasized strict inclusion and exclusion criteria for choosing patients. Habib Davanloo broadened the focus of the brief psychodynamic psychotherapies to include more than one conflict. He also advocated the use of psychodynamic psychotherapy, actively confronting resistances, and including other patient populations such as individuals who suffered from character pathology and chronic phobic and obsessional neuroses. As in the other psychodynamic psychotherapies, James Mann's time-limited psychotherapy developed a central issue as the focus of the treatment. However, he introduced the unique contribution of focusing on the meaning of time. Mann related this to the patient's difficulties in confronting loss and separation—the reality of time and death.

In the last two decades, brief psychotherapy has been increasingly research based. Hans Strupp, Lester Luborsky, and Mardi Horowitz have all introduced manualized psychodynamic treatments with substantial contributions to our research understanding of this treatment modality.

Additionally, cognitive therapy and interpersonal psychotherapy represent brief psychotherapeutic techniques greatly influenced by psychoanalysis but with distinctly different approaches. They share many common elements. Both originally were developed as short-term individual psychotherapies for the treatment of depression but have evolved significantly from this initial focus. They have been termed *reeducative psychotherapies* because they emphasize model-specific case formulations, procedurally guided interventions, and education. They also differ from psychodynamic psychotherapy in that they do not center work on transference, favoring an active, collaborative role.

Cognitive Therapy Cognitive therapy was developed by Aaron Beck and colleagues. Cognitive psychotherapy is a brief method of psychotherapy developed over the last two decades by Beck and his colleagues at the University of Pennsylvania. It is similar to behavior therapy in that it aims at direct removal of symptoms rather than resolution of underlying conflicts, as in psychodynamic psychotherapies. Unlike traditional behavioral approaches, however, the subjective experience of the patient is a major focus of attention in cognitive psychotherapy, as in the psychoanalytically oriented treatments. Beck traces his interest in the topic to 1956, when he became curious about unpredicted findings in his empirical studies of the relationship between specific psychodynamic factors and depression. When psychodynamic concepts did not adequately explain his clinical findings about depression, he explored alternative hypotheses. Beck was struck by the distortions of depressed patients in their negative views of themselves and the world. Based on his observations and research by others, Beck set out to develop a treatment to diminish depressive symptoms by correcting patient distortions. Behavior therapy substantially influenced the development of cognitive therapy. In particular, methodological behaviorism's focus on specifying discrete goals and the steps needed to attain them accompanied by prompt feedback were incorporated into cognitive therapy. Beck developed the concept of “collaborative empiricism,” a method of correcting cognitive distortions by testing the patient's negative beliefs in real life situations.

Interpersonal Psychotherapy Interpersonal psychotherapy is a focused, short-term therapy developed by Gerald Klerman and colleagues. Originally, interpersonal psychotherapy was developed as a time-limited (generally 16 sessions), focused individual psychotherapy for nonpsychotic, unipolar depression. Interpersonal psychotherapy's focus is on interpersonal events rather than intrapsychic or cognitive events. Its theoretical underpinnings are based on interpersonal psychiatry, which, in turn, was heavily influenced by the sociological and philosophical views of the time. The first psychiatrist to articulate the importance of psychosocial factors in the development of mental illness was Adolph Meyer. Harry Stack Sullivan expanded the emphasis on psychosocial factors, defining psychiatry as the field of

interpersonal relations. The use of interpersonal theory in treating patients with depression, as distinguished from an exclusively psychodynamic or biological approach, began with Sullivan and Frieda Fromm-Reichmann's work in the 1930s and 1940s. Interpersonal psychotherapy for depression grew out of the work of the New Haven-Boston Collaborative Depression Research Project on outpatient psychotherapy of patients with nonpsychotic unipolar depression, which began in the late 1960s. The research project's work involved testing psychotherapy with depressed patients either alone or in conjunction with pharmacotherapy. In the early phase of the research, the major interest was in pharmacotherapy as a maintenance treatment for depression. When it was included as a treatment to be examined under the National Institute of Mental Health (NIMH) Collaborative Study on the Treatment of Depression, interpersonal psychotherapy was refined further and training was developed. Interpersonal psychotherapy has since been adapted for use with couples and groups, has been used for a broader group of psychiatric disorders, and can be a longer-term treatment.

THEORETICAL ISSUES

Although at times we speak of psychotherapy as beginning as soon as the physician sees the patient, this hyperbole is used primarily to underscore the importance of interpersonal and transferential elements in the initial meeting with the patient. It is extremely important to distinguish the diagnostic interviews from the ongoing treatment. The evaluation process is distinct from, although related to, psychotherapeutic technique. The evaluation is particularly important in brief psychotherapy because of the need for rapid and accurate assessment without the luxury of time to reevaluate and correct mistakes afforded in the longer-term treatments. The interventions and technical procedures performed during the evaluation phase differ substantially from the technical aspects of psychotherapy itself. The necessity for diagnosis, treatment recommendation, and consideration of the interaction among the patient's ego strength, physical health, and selection variables and the different treatment options (including “no treatment indicated”) is a primary part of the physician's responsibility during the evaluation phase for brief psychotherapy, as in any treatment decision.

Through negotiation with the patient, a treatment decision is reached, and the psychotherapy begins. Many patients do not make it through the evaluation phase of seeking help. Repeatedly, it has been shown that more than 50 percent of patients drop out by five sessions. Setting a time limit appears to decrease the dropout rate. It is inappropriate to consider those who drop out as having been in psychotherapy. Clearly, some of these patients benefit during their short contact with mental health professionals, some through the nonspecific curative factors of help seeking and others through guidance and crisis intervention. However, many may drop out because they did not receive what they were looking for. Despite the therapist's interest in targets of treatment, what is dealt with in treatment is what patients can bring into focus, what patients can tolerate talking about, and what they can tolerate the therapist talking about.

Brief Psychodynamic Psychotherapy Theory Primary to the brief psychodynamic treatments is having the patient rapidly feel engaged and involved in the work. The brief psychodynamic psychotherapies are more focused than the extensive reworking of personality undertaken in longer-term psychotherapies and more here-and-now oriented, with less attempt to reconstruct the developmental origins of conflicts. Through the exploration of the patient's metaphors and symbols, defensive patterns and disturbances in present interpersonal relations can be identified in both the treatment setting and in the patient's life. Being able to hear what the patient has to say and understand its meaning remains central to all psychoanalytically oriented treatments. The therapist's task is to identify the continuity present, but hidden, between each session. The therapist operates on the hypothesis that each session is related to the previous one. The establishment of the therapeutic alliance through identification of the patient's initial anxieties related to beginning therapy remains an important technique in the early phases of the psychodynamic treatments. The therapist must establish conditions under which the patient can favorably hear and deal with the interpretations that the therapist plans to give.

The psychodynamic psychotherapies share an emphasis on understanding (1) the mechanisms of defense used by the patient to decrease anxiety and other uncomfortable feelings associated with areas of conflict which are out of awareness and (2) the characteristic transference relationships that distort the patients' response to their adult world. Typically these two areas, defense and transference, create the world of meaning in which the patient lives. The techniques of the brief psychodynamic psychotherapies are directed toward clarifying these areas and presenting them to the patient to increase understanding and thus change symptoms and behavior.

The goal of treatment is to clarify defenses and the transference in the patient's present life, but particularly as they present in the treatment setting itself, where they can be most clearly seen and are available for the patient to learn from immediately. Participating in this process requires certain capacities in the patient. As in all of medicine, medical treatments are given to patients not to diseases, so patients must be evaluated to determine whether or not they can successfully use the treatment. All medical treatments, including psychoanalysis, make certain requirements upon the patient, which must be taken into account when the treatment is prescribed. The psychodynamic psychotherapies generally require individuals who can access their fantasy life in an active and experiencing manner (i.e., psychological mindedness) and can get up and leave it behind at the end of a session. This does not necessarily mean high intelligence quotient (I.Q.) which, as is often true, when accompanied with rigidity, intellectualization and rumination, can be a contraindication to a brief psychodynamic treatment, since these defenses can be quite formidable. The availability of interpersonal support in the patient's real environment and the patient's ability to experience and simultaneously observe highly charged affective states are also necessary for successful treatment.

The brief psychodynamic psychotherapies focus on the recovery of childhood experiences as they appear in the relationship with the analyst and defense analysis. The re-creation in the doctor-patient relationship of a conflicted relationship with a childhood figure is called the *transference neurosis*. Frequently, the transference is paternal or maternal, but it need not be. Sibling, aunt, uncle, and grandparent transferences are all important parts of psychoanalytic work. When transference neurosis is present, the patient experiences the therapist in a very alive manner as the significant figure from the past. Frequently, other elements of the past are simultaneously experienced in the patient's life. Usually in brief psychotherapy the transference is not as intense as in a longer-term treatment or a psychoanalysis. However, a recent precipitant to the patient's problems (as is usually the case in beginning brief psychodynamic psychotherapy) can intensify any transference responses considerably.

The structure of the psychodynamic therapies is directed toward setting up the therapeutic situation and the patient's observing capacity so that the transference can be interpreted. The hope is to be able to interpret both the triangle of anxiety (wish–defense–anxiety) and the triangle of insight (transference figure–relationship in the present to which it relates–the transference element evident in the therapeutic relationship). Interpretations rarely reach the ideal of being given in one sentence in one session. More frequently, interpretations occur over a period of time during which past, present, and transference experiences are linked together. Individuals in an emergent crisis (e.g., imminently suicidal, psychotic, recent major life trauma) and, therefore, very concerned and focused on the real events in their life cannot enter into a brief psychodynamic psychotherapy without first having supportive treatment. A true life crisis does not allow the patient the opportunity to explore fantasies.

Part of the specific treatment effects of the brief psychodynamic psychotherapies result from the interpretation of the transference situation. The transference is best understood as a reawakening of clusters of effects, cognitions, and behaviors that are linked to significant individuals in the patient's past. Behavioral change occurs in the context of the arousal associated with these figures and the simultaneous understanding of the experience. The patient's ability to perceive previously hidden feelings and relationships and the view of the future and the past can change.

Countertransference is an important element in brief psychodynamic psychotherapy as in other psychodynamic treatments. *Countertransference* is narrowly defined as the analyst's transference responses to the patient. More broadly defined, it is the total emotional reaction to the patient, including feelings unconsciously induced in the analyst by the interpersonal pressure of the patient in the clinical setting. In other words, the analyst will respond to the patient as many other persons in the patient's life respond. In this regard, countertransference provides valuable information about the patient's internal world and the patient's characteristic problems in interpersonal relationships. Countertransference is an ubiquitous occurrence that is increased by life stress and unresolved conflicts in the analyst. It can appear as either an identification with, or a reaction to, the patient's conscious and unconscious fantasies, feelings, and behaviors. Analysis of countertransference reactions may help the therapist recognize subtle aspects of the transference relationship and better understand the patient's experience. Because of the more active stance, the brief psychodynamic psychotherapist can be particularly prone to countertransferences that show up as overinvolvement or aggression. In addition, the brief time available for treatment can make recovery from countertransference errors quite difficult.

Free association is part of the technique of treatment. However, the brief psychotherapist also asks questions, directs the patient's attention, and uses benign neglect; that is the therapist avoids some areas of conflict that cannot be dealt with at this time or in a short period of time. Freud described free association by analogy with a train ride. He suggested that if the doctor and the patient were riding together on a train and the doctor was blind, the patient would not forget to describe the beautiful mountains or the ugly coal slags. This analogy is meant to convey the sense of the patient reporting all thoughts that come to mind without censorship and without thinking them too trivial. In point of fact, free association is difficult to attain. The therapist's task is to identify the spots at which free association breaks down (the occurrence of a defense, clinically experienced by the therapist as resistance). Often when patients can talk freely, with a coherent narrative about their conflicts, the work of the treatment is completed.

Early in treatment, the therapist establishes a therapeutic alliance with the patient that allows reality-based consideration of the demands of the treatment and a working collaboration between analyst and analysand directed to understanding the patient. The therapist focuses on analyzing the defenses the patient uses to minimize conflict and disturbing affects. Defenses such as intellectualization, reaction formation, denial, repression, and other neurotic cognitive mechanisms are identified and repeatedly interpreted to the patient. Through the analysis of defense, the transference grows in a manner that can be analyzed. Dreams, as well as slips of the tongue and symptoms, provide an avenue to the understanding of unconscious conflict. Often a given brief treatment concentrates on only one or two defenses. As these are clarified, the transference relationship may clarify.

Medications are frequently used in conjunction with brief psychodynamic psychotherapy. This can complicate the treatment and its progress as well as aid in symptom recovery. The therapist must explore the meaning of the medication and its role in the patients' view of their self and interpersonal strengths and vulnerabilities. The brief psychodynamic psychotherapies can also serve as an alternative to medication treatment for less severe symptoms or when medication is contraindicated.

The brief psychodynamic psychotherapist operates under several rules. These include the rule of neutrality, by which the therapist favors neither the patient's wishes (id) nor the condemnations of these wishes (superego), and the rule of abstinence, whereby the analyst does not provide gratification to the patient similar to that of the wished-for object. Transference has also been described as the role pressure placed upon the analyst to conform to the behaviors of the significant individual in the past. Therefore, transference is experienced by the therapist as an expectation from the patient that the therapist will behave in a certain manner. This interpersonal pressure is sometimes referred to as *projective identification,* or *role responsiveness. Abstinence* is the avoidance of becoming this figure in reality and gratifying these wishes. However, the therapist may frequently be drawn in by the patient to behave in a manner that is reminiscent of the transference figure, and this enactment of the transference wish may provide valuable information about what others experience with the patient.

Interpersonal Psychotherapy Theory Interpersonal psychotherapy (IPT) is a short-term psychotherapy developed by Klerman and colleagues. Interpersonal psychotherapy is brief—12 to 16 weeks in duration—and focuses on current interpersonal problems in outpatient nonbipolar, nonpsychotic, depressed patients. Interpersonal psychotherapy has often been used in combined psychotherapy and pharmacological treatment studies. Interpersonal psychotherapy has also been used in treating drug abuse; however, it did not significantly improve outcome when patients were already in a well-run treatment program that included weekly group psychotherapy. Interpersonal psychotherapy derives from the interpersonal school of psychiatry that originated with Meyer and Sullivan. The understanding of social supports and of attachment provides further theoretical underpinning for this form of psychotherapy. IPT focuses on reassurance, clarification of feeling states, improvement in interpersonal communication, perception testing, and interpersonal skills rather than personality reconstruction.

Interpersonal psychotherapy is a focused individual psychotherapy. The therapist focuses on the current social functioning. A complete inventory of current and past significant interpersonal relationships, including the family of origin, friendships, and relations in the community, is a part of the evaluation phase. Patterns of authority, dominance and submission, dependency and autonomy, intimacy, affection, and activities are observed. Cognitions are generally seen as beliefs and attitudes about norms, expectations, roles, and role performance. Defense mechanisms may be recognized, but they are explored in terms of interpersonal relations. Similarly, dreams may be examined as a reflection of current interpersonal problems. The interpersonal therapist may explore distorted thinking by comparing what patients say with what they do or by defining their view of an interpersonal relationship.

Interpersonal psychotherapy has been primarily used in the treatment of depressed patients. In its opening phase, a detailed symptom history is taken, usually with a structured interview. The symptoms are reviewed with the patient and the patient receives explicit information about the natural course of depression as a clinical condition. There is an emphasis on legitimizing the patient in the sick role. A second major task of this phase is the assessment of interpersonal problem areas. There is an attempt to identify one or more of four problem areas: grief reactions, interpersonal role disputes, role transition, or interpersonal deficits. Each of these areas is considered to be related to depression. The middle phase of treatment is directed toward resolving the problem area or areas. Clarifying positive and negative feeling states, identifying past models for relationships, and guiding and encouraging the patient in examining and choosing alternative courses of action are the basic techniques for handling each problem area. The focus is kept on current dilemmas, not past interpersonal relationships. Interpersonal events rather than intrapsychic or cognitive events are the focus of interpersonal psychotherapy.

Cognitive Psychotherapy Theory Cognitive psychotherapy was first developed for the treatment of depression, and this remains its most fully developed theoretical area. The central principle of the cognitive theory of depression is that symptoms result from characteristic distortions in thinking. The individual's reports of conscious ideas, feelings, and wishes are taken as data. This position contrasts with that of classical behaviorists, who might view them as indicating deeper meanings.

To the cognitive therapist, depressed persons interpret life differently from those who are not depressed. Depressed people have negative interpretations of the world, themselves, and the future (the negative cognitive triad). To depressed individuals, events reflect defeat, deprivation, or disparagement, and their lives are filled with obstacles and burdens. They view themselves as unworthy, deficient, undesirable, or worthless and expect the future to continue the miseries of the present. These evaluations are products of the negative biases inherent in depressive thinking and are applied regardless of the objective nature of the individual's circumstances. For example, a depressed man might think that the checkout person at the grocery store said “Hello” to be nice to him because the person felt sorry for him and knew he was ill. On the other hand, failure to say “Hello” might be interpreted as showing that the checkout person disliked the patient.

In cognitive theory, the affective and motivational aspects of depression are secondary to the negative interpretations of events; that is, how one views an experience determines one's emotional (affective) response to it. For example, a man traveling in a strange city finds himself sitting alone at a conference. If he interprets the experience as a loss of companionship, he is likely to report that he feels sad; if instead he interprets the empty row next to him to mean he is unattractive to people, he may feel shame; if he sees himself sitting alone because the local attendees are unfriendly, he may experience anger. Conversely, a man who sees the same situation as evidence of his independence may feel proud and strong. Persons who anticipate failure, opposition, or humiliation as a result of their efforts may respond with decreased motivation and active avoidance or escapist wishes. Beck has speculated that the physical symptoms of depression may also reflect one's thought processes. For example, psychomotor retardation may be the physical expression of a patient's passive, helpless attitude.

Other psychiatric conditions have their own characteristic cognitive patterns that determine the nature of the symptoms. Anxious persons continually anticipate personal danger or impending disaster, but since this is potential rather than actual (as in depression), their self-esteem remains intact. Paranoid individuals view the world as responding negatively to them, but unlike depressive individuals, they place the blame with the world rather than with themselves.

The thinking disorder of depression as seen in cognitive theory includes arbitrary inference, selective overgeneralization, magnification and minimization, and inexact labeling. These cognitions (verbal thoughts) often feel involuntary and automatic. This thinking is so automatic and the resultant cognitions so fleeting that persons are virtually unaware of them. Such automatic thoughts differ from unconscious thoughts in that they can easily be made fully conscious if the person's attention is directed to them. (Thoughts of this type would be described as “preconscious” in traditional psychoanalytic theory.) A large portion of the work of cognitive psychotherapy deals with training patients to observe and record their automatic thoughts in response to the events of daily life.

Cognitive theory postulates a chronic state of depression proneness that may precede the actual illness and remain after the symptoms have abated. Depression-prone individuals have relatively permanent depressive cognitive structures that determine how new stimuli are perceived and conceptualized. Such structures are called *cognitive schemas.* The schemas of depression may include ideas such as “I am stupid” or “I cannot exist without the love of a strong person.”

Unlike with automatic thoughts, patients are not typically aware of their underlying general assumptions. Rather, they must be deduced from many specific examples of distorted thinking. Schemas such as “I am stupid” may lie dormant much of the time only to be reactivated by a specific event such as difficulty in accomplishing a task. These enduring self-concepts and attitudes are assumed to have been learned in childhood on the basis of the child's experiences and reactions of important family members. Once formed, such attitudes can be self-perpetuating.

Just as depressive thoughts can be triggered by events, episodes of depressive illness may, from the cognitive perspective, be triggered by sufficient stress. Such stresses may be specific to individuals and the particular sensitivities they developed in childhood. Alternatively, sufficient nonspecific stress may precipitate depression in vulnerable individuals. Loss, a setback in a major goal, a rejection, or an unsolvable dilemma are especially common precipitants of depression. The onset of medical illness, with its attendant limitations and associated meanings, is also seen as likely to trigger depression in many people.

TECHNIQUES

Brief Psychodynamic Psychotherapy

Malan and the Tavistock Group: Focal Psychotherapy Focal psychotherapy, which developed from the workshops of Balint and Malan, is an example of applied psychoanalysis. Malan has carried on Balint's earlier work. Previous attempts to develop brief forms of psychoanalytic psychotherapy primarily involved the use of “activity.” However, Malan emphasized the importance of choosing and maintaining a narrow focal area to be dealt with in a brief period of time. Rather than increasing activity, which was frequently equated with manipulation, Malan stresses the importance of finding the appropriate focus in the patient's story and consistently interpreting the focal problem area. Through selective attention and neglect, the therapist maintains the focus and completes a brief psychotherapy. The importance of determining the focus underscores the value of the diagnostic process prior to the initiation of psychotherapy, including the psychodynamic assessment of the patient.

Malan identifies the following factors as leading to the lengthening of treatment: resistance, overdetermination, a need for working through the roots of conflict in early childhood, transference, dependence, negative transference connected with termination, and the transference neurosis. In addition, some therapist characteristics may lengthen treatment, including a tendency toward passivity, a sense of timelessness conveyed to the patient, therapeutic perfectionism, and a preoccupation with deeper earlier experiences. All of these factors must be dealt with to maintain a brief therapy. For Malan, identifying a focal conflict acceptable to the patient is critical to a successful outcome (Table 30.8-1). In addition, the patient must be able to think in feeling terms, demonstrate high motivation, and exhibit a good response to trial interpretations made during the evaluation phase. Patients who have had serious suicidal attempts, drug addiction, long-term hospitalization, more than one course of electroconvulsive therapy, chronic alcoholism, incapacitating severe chronic obsessional symptoms, severe chronic phobic symptoms, or gross destructive or self-destructive acting out are excluded from treatment. The patient is also excluded from focal psychotherapy if the therapist anticipates any of the items listed in Table 30.8-2.

Goal	Identify the defense, the anxiety, and the impulse Link the present, the past, and the transference
Focus	Internal conflict present since childhood Link the present, the past, and the transference
Selection criteria	Patient can think in feeling terms Highly motivated Good response to trial interpretation
Duration	Up to 1 year Mean, 20 sessions
Termination	Set definite termination date at beginning of treatment

Table 30.8-1 Key Elements of Malan and the Tavistock Group's Brief Focal Psychotherapy

1. Therapist cannot make affective contact with the patient during the evaluation
2. Therapist anticipates needing extended work
   - To generate motivation
   - To decrease rigid defenses
   - To reach complex or deep-seated issues
   - To resolve unfavorable, intense transference, or dependence that may develop
3. Depressive or psychotic disturbance may intensify and place the patient at risk

Table 30.8-2 Exclusion Criteria for Malan's and the Tavistock Group's Brief Focal Psychotherapy

For Malan, the criteria in Table 30.8-2 represent specific dangers. If the therapist cannot make contact with the patient or the patient exhibits low motivation or rigid defenses, forming an effective therapeutic working alliance within a short time will be difficult. Complex or deep-seated issues that must be dealt with to resolve a conflict area require longer treatment. Difficult transference relationships may also prevent timely termination or lead to premature termination. Severe depressive or psychotic episodes during treatment can endanger the patient and require adjunctive treatments. Thus, Malan takes the time limitation in brief therapy seriously, which requires rapid establishment of a therapeutic alliance and the ability to terminate therapy without the development of unexpected serious symptoms.

Malan, in contrast to other practitioners, does not necessarily exclude patients with serious psychopathology. In fact, several patients that he presents exhibit significant pathology. Rather, he sees the balance between motivation and focality as of primary importance. A patient with only moderate motivation but a highly focal conflict might be accepted into treatment. Similarly, a patient with high motivation but not as focal a conflict might also be accepted into treatment with the hope that the focus would clarify in a short period of time.

Identifying the precipitating factors, early traumatic experiences, or repetitive patterns can indicate the area of internal conflict present since childhood and the focus for treatment. The therapist should assess the congruence between the current conflict and the nuclear, or childhood, conflict during the evaluation phase. The patient's response to interpretations about aspects of this conflict may lead to acceptance into treatment. According to Malan, the greater the probability that the conflict area will manifest itself in the transference, the more positive the outcome will be. Furthermore, he reported that transference interpretations correlated with character change and that character change endured 2 to 10 years later.

Malan is less concerned with technique than with the importance of choosing the focus. He uses all the usual technical procedures of psychoanalytic psychotherapy and emphasizes the importance of interpreting the transference and connecting it to current and past relationships. This triangle of insight (the transference, the current relationship, and the past relationship) leads to the patient's cure. Overall, the goal is to clarify the nature of the defense, the anxiety, and the impulse that the patient is experiencing and to link these to the present, the past, and the transference. Once the defense and the anxiety are clarified, the link to the past can be made. The interpretation that links to the past may be reassuring to the patient because it emphasizes that the conflict belongs to the world of fantasy rather than the world of the present. Malan emphasizes transference interpretations as the most therapeutically effective interpretations because of their here-and-now character.

In the brief therapy unit at the Tavistock Clinic, a time limit was almost always given when treatment began. For trainees this was usually 30 sessions. However, in his publications, Malan indicates a mean of 20 sessions for patients with favorable outcomes. The longer time for trainees gives an opportunity to correct mistakes that might occur. In some published cases, therapy was extended up to 1 year (46 sessions). In general, Malan advocates the importance of a definite date rather than a number of sessions. Practically speaking, this eliminates the need for the patient and therapist to keep count of the number of sessions and deviates concern about whether or not to make up missed sessions. Such a time limit gives a definite beginning, middle, and end to the therapy; helps to concentrate the patient's material and the therapist's work; maintains the focus; and decreases diffuseness that might lead into long-term work.

Sifneos: Short-Term Anxiety-Provoking Psychotherapy Sifneos emphasizes the importance of patient selection because of the anxiety-provoking nature of his brief psychotherapy techniques (Table 30.8-3). He distinguishes anxiety-provoking therapy from anxiety-suppressing therapy, commonly referred to as *supportive psychotherapy*. For short-term anxiety-provoking psychotherapy (STAP), patients must have above average intelligence and have had at least one meaningful relationship with another person. Patients who have had such a relationship will be able to withstand the anxiety produced by the therapy and rapidly develop a mature collaborative relationship with the therapist. This criterion tends to exclude narcissistic disorders. In addition, the patient must be highly motivated for change, not only for symptom relief. Sifneos also identifies several criteria for patient selection on the basis of the presentation of the patient during the evaluation. Patients must have a specific chief complaint. If the patient has a number of complaints, Sifneos asks the patient which complaint is of top priority. The patients' ability to identify one conflict area and to postpone work on others is considered an indication of their ability to tolerate anxiety. Sifneos looks for patients with anxiety,

depression, phobias, conversion, and mild obsessive-compulsive features or personality disorders involving clear-cut interpersonal difficulties. During the evaluation, patients must show an ability to interact with the evaluating psychiatrist, to express feelings, and to show some flexibility.

Goal	Resolution of oedipal conflict
Focus	Oedipal (triangular) conflict/competition
Selection criteria	Above-average intelligence At least one past meaningful relationship High motivation Specific chief complaint Able to interact with evaluator Able to express feelings Flexible
Duration	A few months Average, 12–16 sessions
Termination	No specific date given

Table 30.8-3 Key Elements of Sifneos' Short-Term Anxiety-Provoking Psychotherapy (STAP)

Sifneos is one of the few authors who clarifies his assessment of motivation. He defines *motivation* as including the patient's ability to recognize symptoms as psychological, a tendency to be introspective and honest about emotional difficulties, and a willingness to participate in the treatment situation. In addition, motivation includes curiosity, a willingness to change and to make reasonable sacrifices, and a realistic expectation of the results of psychotherapy.

Sifneos focuses on the oedipal conflict and does not expect a good outcome in dealing with other than oedipal conflict areas. Most failures using STAP have occurred in patients who complained of reactive depression following the loss of a loved one. He believes that this failure is due to the nontriangular (nonoedipal) origins of the ambivalent feelings in some patients. In such cases, when the issue of termination arises, the patient regresses and an impasse is reached.

During the initial phase of psychotherapy, the therapist must establish good rapport with the patient to create a therapeutic alliance. The therapist uses anxiety-provoking confrontations to clarify issues around the patient's early life situation and present-day conflict. The therapist avoids areas such as passivity, dependence, and acting out, which might lead to extensive regression. The use of anxiety-provoking confrontations in a direct attack on the patient's defenses distinguishes STAP from other brief psychotherapies. Although it is made clear to patients during their evaluation that the psychotherapy is expected to last only a few months, no specific number of sessions or termination date is given. Interviews are held weekly and are 45 minutes long. The vast majority of treatments last from 12 to 16 sessions, and none go beyond 20 sessions. The aggressive, confrontational style of this treatment underscores the importance of excluding preoedipal problems and the importance of countertransference reactions in the therapist related to being too aggressive.

Mann: Time-Limited Psychotherapy Mann has focused on the specific limitation of time in brief psychotherapy. He sees the time variable as a specific operative factor in psychotherapy as well as an element in its curative effect. The experience of the timelessness of treatment and of the treatment's termination are significant elements in Mann's view of the psychotherapeutic process.

Usually two to four evaluation meetings occur before psychotherapy begins. Mann limits psychotherapy to a total of 12 treatment hours, distributed according to patient need. This may result in weekly half-hour sessions for 24 weeks or twice-weekly hour sessions for 6 weeks. In practice, however, nearly all patients are seen in once-weekly, 45- or 50-minute sessions for 12 weeks. Mann admits having chosen the number 12 somewhat arbitrarily; however, his clinical experience indicates that between 10 and 14 sessions suffice. He emphasizes the importance of using a uniform number of sessions in evaluating the psychotherapeutic process among different therapists. This facilitates study of the relation between the patient's presenting problems and psychotherapeutic technique. Also, the patient accepts provision of a specific number of sessions as a typical medical “prescription.” Finally, the setting of a specific last session in the initial contract with the patient allows the therapy to have a clear beginning, middle, and end (Table 30.8-4).

Goal	Resolution of the present and chronically endured pain Change in the patient's negative self-image
Focus	Present and chronically endured pain Particular image of the self Loss, time, and termination
Selection criteria	High ego strength Able to engage and disengage Therapist quickly able to identify a central issue Excludes serious depression, acute psychosis, and borderline personality organization
Duration	12 Treatment hours Usually 12 sessions
Termination	Set date of last session at beginning of treatment

Table 30.8-4 Key Elements of Mann's Time-Limited Psychotherapy

Mann, to some extent, minimizes selection as a central issue for brief psychotherapy. He does indicate a number of exclusionary criteria: serious depression, acute psychosis, borderline personality organization, and the inability to identify a central issue. Mann sees Sifneos' criteria as primarily excluding borderline patients. He does not agree with Sifneos' emphasis on superior academic or work performance.

More recently Mann has expanded his selection criteria by emphasizing the importance of the patient's ego strength as measured by prior work performance and past relationships. Patients who may have difficulty engaging and disengaging rapidly from treatment are excluded. This group includes schizoid patients, certain obsessional patients, patients with strong dependency needs, some narcissistic patients, some depressive patients who will not be able to form a rapid therapeutic alliance, and some patients with psychosomatic disorders who do not tolerate loss well.

According to Mann, selecting the central issue for the psychotherapy is a critical event. It is the vehicle for engaging the patient in the work of therapy and the basis for a successful outcome. Mann looks for a central issue that is developmentally and adaptively relevant and has recurred over time. He describes this issue as the patient's “present and chronically endured pain” and characterizes it as preconscious. Mann has further described the central issue as including a particular image of the self. The central issue formulated in terms of time, affect, and an image of the self is the “paradigm of the transference” expected to emerge in treatment. The therapist's statement of the central issue is a clarification that can be readily recognized, felt, and held onto by the patient. Time-limited psychotherapy is intended to resolve this present and chronically endured pain and the patient's “negative self-image.” The therapist frames the central issue to the patient in terms of a general statement about feelings.

In their volume *A Casebook in Time Limited Psychotherapy* Mann and Robert Goldman described in detail the phrasing of the central issue to the patient. It is the central issue that specifies the therapeutic contract and the goal of the therapy. For a 41-year-old depressed woman who was preoccupied with her husband and children being even 1 minute late, Mann suggested the central issue “You've encountered extreme life situations and have managed them remarkably well... yet you fear and have always feared that despite your best efforts you will lose everything.” For a 31-year-old married man attempting to get a college degree who was consumed with a fear of failing, Mann suggested the central issue. “Because there have been a number of sudden and very painful events in your life, things always seem uncertain, and you are excessively nervous because you do not expect anything to go along well. Things are always uncertain for you.”

Mann uses the usual psychoanalytic psychotherapy techniques: defense analysis, transference interpretation, and genetic reconstruction. Transference is interpreted from within the central identified conflict area and in terms of the adaptive processes of the patient. However, Mann does not confront the patient. In general, his

interventions are very close to the conscious material provided by the patient. Mann identifies specific dynamic events that unfold during the 12 sessions. The opening sessions are understood as filled with the unconscious magical expectation that past pain will now be resolved. During the initial phase, the therapist makes few comments and accepts the positive transference of the patient. Important aspects of the current problem, defense mechanisms, coping styles, and genetic roots of the central issue become clearer during this phase. In the middle four sessions, resistance is likely to appear, as well as the negative transference. The patient experiences the frustration that all of the wished for changes may not occur. In the ending phase of treatment, termination and the patient's resistances to termination in the face of unresolved problems in other areas of life are prominent. Mann sees confronting separation and termination issues as critical to the success of brief psychotherapy. Frequently, patients unconsciously reveal an awareness that the midpoint of treatment has come. They experience separation from the transference-invested therapist as a separation from an ambivalently experienced person from the past, without having achieved the fantasized magical resolution. The goal is to enable patients to separate from the transference-invested therapist less ambivalently than they did from this earlier important figure. Consequently, both the resolution of the central issue and the unfolding of an attachment/separation process in the 12-session treatment contract are intimately related through the development and interpretation of the transference.

Davanloo: Broad-Focus Short-Term Dynamic Psychotherapy Davanloo writes about broad-focus, short-term dynamic psychotherapy. His selection criteria include patients with an oedipal focus, those with a loss focus, and those with multiple foci. Davanloo is particularly interested in patients suffering from long-standing obsessional and phobic character styles presently called mild-to-moderate obsessive-compulsive personality disorder, avoidant personality disorder, and dependent personality disorder. His research data indicate that 30 to 35 percent of the psychiatric outpatient population can benefit from this mode of therapy. Most information about his technique is derived from the publication of cases, presentations, and brief descriptions of his research that accompany case presentations.

The initial evaluation is a specific, focused interview in which the patient's defenses against true feelings are gently but consistently confronted. Davanloo says that this is not a universal technique for the initial interview and cautions on its use with patients with severe psychopathology. Selection is based on psychological-mindedness, the quality of the patient's interpersonal relations, and, in particular, the presence of at least one meaningful relationship in the patient's past. The patient's ability to tolerate and experience anxiety, guilt, and depression are important (Table 30.8-5). Patients must be motivated to complete the treatment process and to resolve neurotic problems. Their ability to respond to interpretation is an important selection criteria. In particular, response to transference interpretations that link the transference with the present and the past is a critical feature in the assessment for broad-focus, short-term dynamic psychotherapy. Davanloo finds no value in criteria based on severity and duration of illness. Finally, the presence of flexibility in the ego's defensive pattern and a lack of use of the primitive defenses of projection, splitting, and denial are important factors in selecting patients.

Goal	Resolution of oedipal conflict,
Focus	Loss focus or multiple foci
Selection criteria	Psychological mindedness At least one past meaningful relationship Able to tolerate affect Good response to trial transference interpretation High motivation Flexible defenses Lack of projection, splitting, and denial
Duration	5 to 40 Sessions, usually 5 to 25 sessions Longer duration for more seriously ill patients
Termination	No specific termination date set Patient is told the treatment will be "short"

Table 30.8-5 Key Elements of Davanloo's Broad-Focus Short-Term Dynamic Psychotherapy

The technique Davanloo uses in therapy is a continuation of that used in the initial interview. The emotional experience of the patient in the transference is emphasized. Patients are "gently but relentlessly" confronted about their defenses against feelings in the transference relationship and in the past. All the usual techniques of psychoanalytic psychotherapy are used: defense analysis, transference interpretations, and genetic reconstruction. Dreams and fantasy material are also used. Transference interpretations tend to be made early. Because of the confrontative style, a strong therapeutic alliance is necessary. Patients frequently experience hostile, angry feelings toward the therapist because of being confronted. Davanloo actively pursues the patient's defenses against recognizing the anger and its transference elements. He warns therapists that passive dependent and obsessional characters may develop a symbiotic transference relationship. This may be avoided through active confrontation and selection of patients. The active confrontation of defenses and early transference interpretations tend to mobilize powerful affects and memories early in treatment.

Davanloo recommends 5 to 40 sessions, depending on the patient's conflict area (oedipal versus multiple foci) and other selection criteria. In general, his treatments fall between 15 and 25 sessions. He does not recommend setting a specific termination date but rather makes clear to the patient that treatment will be short. Shorter time periods (5 to 15 sessions) are chosen for patients with a predominantly oedipal focus, longer durations (20 to 40 sessions) for the more seriously ill group.

COMPARISON OF THE BRIEF PSYCHODYNAMIC PSYCHOTHERAPIES

The work of Malan, Sifneos, Mann, and Davanloo shows substantial overlap in each author's goals, selection criteria, technique, and duration of treatment. These authors would agree that the goals of brief psychotherapy include facilitating health-seeking behaviors and mitigating obstacles to normal growth. From this perspective, brief psychotherapy focuses on the patient's continuous development throughout adult life and the context-dependent appearance of conflict, depending on environment, interpersonal relationships, biological health, and developmental stage. This picture of brief psychotherapy supports modest goals that require the therapist to refrain from perfectionism. Malan, Sifneos, Mann, and Davanloo also seem to agree that brief psychotherapy uses the propitious moment and long-term treatment primarily uses the shared past of the therapist and patient. Both the propitious moment and the shared past carry psychotherapeutic advantages and disadvantages, emphasizing certain technical possibilities and limiting others.

Many of the selection criteria emphasized by Malan, Sifneos, Mann, and Davanloo are common to all kinds of psychodynamic psychotherapy. However, unique selection criteria are necessitated by the brief duration of treatment. Patients in brief psychodynamic psychotherapy must be able to engage quickly with the therapist, to terminate in a short period of time, and to carry on much of the working through and generalizing of the treatment effects on their own.

The necessity for greater independent action by the patient mandates high levels of ego strength, motivation, and responsiveness to interpretation. Sifneos's rather unique emphasis on intelligence as a criterion may be related to his anxiety-provoking interpretations, which require a broader educational context to be understood. The importance of rapid establishment of the therapeutic alliance underlies a substantial number of the selection and exclusion criteria.

The central importance of a focus in brief psychotherapy is mentioned by all authors. They also emphasize the importance of the evaluation sessions to determine this focus. Mann formulates the focus to the patient in terms of the patient's fears and pain. However, he would probably agree with Malan, Davanloo, and Sifneos on the importance of constructing the psychodynamic focus at a deeper level in one's own understanding of the work being done. Maintaining the focus is the primary task of the therapist. This enables the therapist to deal with complicated personality structures in a brief period of time. Resistance is limited through benign neglect of potentially troublesome but nonfocal areas of the personality. The elaboration of techniques of establishing and maintaining the focus of treatment is critical to all brief individual psychodynamic psychotherapies.

The authors also discuss the importance of transference interpretations. However, the manner and rapidity of dealing with transference vary considerably. Malan takes a more typical psychoanalytic approach of waiting for transference to become resistance before it is interpreted. Sifneos, in his emphasis on the oedipal relationship, is more aggressive in handling the deep conflictual areas of transference material. Davanloo is confrontational in developing a transference experience. This confrontational style may at times confuse the patient's experience of the real and the transferential therapist. However, Davanloo's focus on the treatment of severe obsessional disorders, in which the need to increase affective awareness is high, may be where this particular technique is most useful. Aggressive, competitive, and hostile feelings that might otherwise remain firmly defended may thus become available to these patients.

The role of countertransference in brief psychotherapy is as complicated as it is in long-term treatment. Countertransference issues related to the aggressive techniques used by Sifneos and Davanloo have been observed. Countertransference experiences related to termination and loss can also be prominent. The

goal-directed techniques of brief psychotherapy limit the development of regressive countertransference responses.

There is remarkable agreement on the duration of brief psychotherapy. Although the duration ranges from 5 to 40 sessions, authors generally favor 10 to 20 sessions. The duration of treatment is critically related to maintaining the focus within the brief psychotherapy. In Rogerian therapy a positive correlation was found between the number of sessions and recovery. In general, an increasingly successful outcome (measured by the patient's self-concept) is seen up to about 20 sessions. In meta-analyses of a number of brief psychotherapies, 75 percent of patients show some improvement by 26 sessions. However, this study included a wide range of types of treatment. When treatment extends beyond 20 sessions, therapists may frequently find themselves enmeshed in a broad character analysis without a focal conflict. Change after 20 sessions may be quite slow. Clinical experience generally supports the idea that brief individual psychodynamic psychotherapy should involve 10 to 20 sessions, although more complicated cases will require longer treatment. Often, extending treatment beyond 20 sessions is recognition that treatment will involve 40 or 50 sessions.

Brief psychodynamic psychotherapy for depression and narcissistic disturbances has been described. These authors perform an important role by describing the techniques most effective for specific diagnostic categories. Horowitz has described brief psychotherapy focused on the stress responses evidenced by various personality styles. He emphasizes that this psychotherapy is directed toward dealing with the process of the stress response and not character change. However, his outcomes indicate that selected character changes are possible in some areas. The distinction between recovery from a disruption in homeostatic balance, reconstitution of self-esteem and self-concept, and changes in character structure requires further exploration.

The identification of critical points during brief psychotherapy, when the danger of becoming a long-term treatment is most acute, will further clarify the technical handling of brief psychodynamic psychotherapy. Characteristic phenomena have been identified in the interlude between a short- and a long-term treatment: an increasing vagueness with respect to the goals of the treatment, decreased activity by the therapist, and emergence of the transference as a central element. These variables indicate the potential for a short-term psychotherapy to become a long-term treatment. The fourth to sixth hours of once-a-week therapy are when incipient or potential regression may suddenly appear. At this time the patient is testing the boundaries of the treatment. Action by the therapist is required if a brief psychotherapy is to remain exactly that—brief. The study of technical interventions that occur at these critical moments will further elucidate the technical handling of limited regression in brief psychodynamic psychotherapy.

Interpersonal Psychotherapy In interpersonal psychotherapy the therapist's attitude is one of exploration, similar to that in other insight-oriented psychotherapies when applied in a medical model. Applying the dictum of working “from the surface to the depths” results in much of psychodynamic psychotherapy resembling interpersonal psychotherapy. However, Klerman and colleagues have found it useful to highlight the differences between these approaches to standardize a psychotherapeutic technique.

Interpersonal psychotherapy focuses on the patient's interpersonal relationships at the time that psychiatric symptoms began to develop. It emphasizes current rather than past interpersonal relationships and makes no assumption about the cause of the disorder. Interpersonal psychotherapy draws heavily upon psychodynamic theory. In contrast to psychodynamic psychotherapy, the therapist's role is one of active advocacy and the therapeutic relationship is not interpreted as transference. Its goals are to diminish symptoms and maximize social function, rather than to alter personality structure (Table 30.8-6).

Goal	Improvement in current interpersonal skills
Focus	Four problem areas: grief, interpersonal role disputes, role transitions, and interpersonal deficits
Selection criteria	Outpatient, nonbipolar, nonpsychotic depression
Duration	12 to 16 Weeks, usually once weekly
	Three phases: diagnostic evaluation, implement strategies related to one of the four problem areas, consolidate gains
Technique	Education
	Reassurance
	Clarification of feeling states
	Remediation of interpersonal communication
	Testing perceptions
	Development of interpersonal skills
	Medication

Table 30.8-6 Key Elements of Interpersonal Psychotherapy

As short-term treatment for depression, interpersonal psychotherapy consists of three phases. The first phase generally consists of the first one to three sessions. Its goals are to gather psychiatric history, establish a diagnosis, and introduce the framework for treatment. The psychiatric history explores current social functioning. Particular attention is paid to gathering extensive information about interpersonal events that may have precipitated the depressive episode. The patient's current and past significant interpersonal relationships, including the family of origin, friendships, and community relations, are also reviewed during this first phase. The data gleaned from this review are used to identify one of four problem areas that will guide the therapy: unresolved grief, social role disputes, social role transitions, and interpersonal deficits. Psychiatric diagnosis is informed by standard criteria. Decisions about the concomitant use of medication are based on the severity of the symptoms, past response to interventions, and patient preference. The patient is placed in the sick role; the therapist explicitly discusses the diagnosis of depression and its attendant symptoms as well as explaining what the patient can expect from treatment. The depressive syndrome is then related to the patient's main interpersonal theme, which related to the syndrome's onset.

The middle phase of treatment is directed toward resolving the problem area. Specific goals and strategies are used for each of the four areas. For a patient whose main problem area is unresolved grief, the goals of treatment are to facilitate the mourning process and assist the patient in finding new activities and relationships to offset the loss. In the treatment of interpersonal role disputes, the dispute is identified, a plan of action is chosen, and a satisfactory resolution is sought through the modification of expectations or improved communication. If resolution proves impossible, clients are encouraged to consider terminating that plan of action in favor of finding better ones. For depression associated with role transitions, the patient is helped to mourn and accept the loss of the old role. Efforts are directed to helping the patient regard the new role as more positive than the old one. Self-esteem is enhanced by focusing on the new skills that are mastered in learning the new role. Finally, when interpersonal deficits are the core theme, the therapist encourages the patient to establish relationships and diminish social isolation.

Each session begins with the question “How have things been since we last met?” to focus on current mood states and their association with recent interactions. The basic techniques for handling each problem area are clarifying positive and negative feeling states, identifying past models for relationships, and guiding and encouraging the patient in examining and choosing alternative courses of action.

Interpersonal psychotherapy has been used across a variety of depressed populations: geriatric, adolescent, HIV-infected, dysthymic disorder, bipolar disorder, and depressed patients with marital problems. It has also been used for nonmood psychiatric conditions such as substance abuse and bulimia nervosa.

Cognitive Psychotherapy Cognitive therapy aims to correct habitual patterns of distorted thinking. It uses both behavioral modification techniques and verbal interventions geared at empirically testing erroneous beliefs. At inception it was a brief, individual treatment for depression. It has been used for a wide variety of psychiatric conditions and has been adapted for marital and group formats (Table 30.8-7).

Goal	Identify and correct cognitive distortions and dysfunctional beliefs that maintain symptoms
Focus	Beliefs about oneself, the world, and the future
Selection criteria	Primarily used in unipolar, nonpsychotic depression Nonendogenous depression Symptoms not sustained by pathological family
Duration	Time limited, usually 15 to 25 weeks, once-weekly meetings
Techniques	Collaborative empiricism Identification of irrational beliefs and automatic thoughts Identification of attitudes and assumptions underlying negatively biased thoughts Education Structured and directive Assigned readings Homework and behavioral techniques

Table 30.8-7 Key Elements of Cognitive Psychotherapy

Cognitive therapy as it is applied in the treatment of outpatient depression is a brief psychotherapy. Generally, marked improvement is seen between 5 and 7 weeks of once-weekly or twice-weekly therapy. If patients haven't achieved a 50 percent reduction in depressive symptoms using a standard rating scale for depression within 14 sessions, the patient's diagnosis and treatment plan should be reevaluated. After the primary therapeutic course is completed, once- or twice-monthly booster treatments are often given for the next 6 to 12 months.

Cognitive psychotherapy is a directive, time-limited, multidimensional psychological treatment that derives from the cognitive theory of psychiatric disorders. Cognitive theory posits that negative biased thoughts predispose to depression. These thoughts are often unconscious but can become conscious, and if the patient can be induced to become more logical and realistic, the depressed mood can be alleviated.

This cognitive reorientation is accomplished with a collaborative empirical approach. Together patients and therapist discover the irrational beliefs and illogical thinking patterns associated with the patients' depressive affects. They then devise methods patients can use to test the validity of their thinking. The therapist helps patients become aware of their irrational beliefs and distorted thinking (“automatic thoughts”) and to see, indeed, whether their ideas are objectively true or logical.

Cognitive psychotherapy is generally conducted over 15 to 25 weeks in once-weekly meetings. With more severely depressed patients, two or three meetings per week are recommended for the first several weeks. Although cognitive psychotherapy was developed, and is usually administered, as an individual treatment, its principles have been successfully applied to group settings.

Cognitive psychotherapy was developed for unipolar, nonpsychotic, depressed outpatients and has been shown effective in this subcategory of depression. The presence of bipolar illness, delusions or hallucinations, or extremely severe depression are contraindications for cognitive psychotherapy as the sole or primary treatment modality. Other contraindications include the presence of medical illness or use of a medication that may cause depression, the presence of an organic brain syndrome, or an ongoing problem of substance abuse. In addition, cognitive psychotherapy may not be indicated for endogenous depression, which may be accompanied by endocrine, sleep, or other biological abnormalities, or for patients who are enmeshed in a family system that maintains a fixed view of them as helpless and dependent. Cognitive psychotherapy may also be useful in patients who refuse to take, fail to respond to, or are unable to tolerate, medication as well as those who prefer a psychological approach in the hope of greater long-term benefits.

Cognitive psychotherapy proceeds in a succession of regular stages. The first stage is devoted to introducing the patient to the procedures and rationale of the therapy and setting goals for the treatment. The therapist may assign reading material on the cognitive theory of depression. The therapist devotes particular attention to eliciting a full description of the presenting problems and difficulties from the patient. With the patient, the therapist decides on the immediate goals of treatment based on the patient's priorities and the amenability of the symptoms to treatment. These procedures foster the therapeutic alliance by establishing a collaborative tone and socialize the patient to some of the basic therapeutic procedures, such as homework assignments.

In the next stage, the therapist begins to demonstrate to the patient that cognitions and emotions are connected. Patients are taught to become more aware of their negatively biased automatic thoughts and to recognize, both during and outside the psychotherapy hours, that negative affects are generally preceded by such thoughts. In addition, in this stage of cognitive psychotherapy, therapists attempt to find areas in which patients can be helped to demonstrate to themselves rather quickly that they are not as helpless or hopeless as they had thought, often accomplished by graded behavioral assignments (see below), especially with more severely depressed patients. This serves to help break the cycle of hopelessness and passivity.

In the next phase, which normally comprises most of the work, the emphasis shifts to a detailed exploration of the patient's cognitions and their role in perpetuating depressive feelings. The therapist helps patients clarify their beliefs and find ways of empirically testing their validity. Together the therapist and patient explore the content of the patient's automatic thoughts, pointing out instances of illogic and helping the patient consider alternative points of view that are not negatively biased.

In the final stage of psychotherapy, patients have had a great deal of experience in recognizing their habitual thought patterns, testing their validity, and modifying them when appropriate with resultant substantial symptomatic relief. Psychotherapy then focuses on the attitudes and assumptions that underlie the patient's negatively biased thinking. Generally, patients are not immediately conscious of such attitudes, as they would be of automatic thoughts, and they must be identified by careful examination of specific cognitions. An example of such an underlying assumption is the idea “If I'm nice, bad things won't happen to me.” A logically equivalent assumption is then “If bad things happen to me, it is my fault because I am not nice.” Specific automatic thoughts might result from such an assumption: “I caused my husband to treat me badly” or “I never have good times because I'm not nice.” In helping the patient successfully challenge and modify such assumptions, the therapist hopes to diminish the long-term frequency and intensity of negatively biased automatic thoughts and their concomitant negative affects. Cognitive therapists consider an increased understanding of the underlying assumptions about oneself and one's life to be an important element in increasing the patient's own ability to recognize and correct illogical, negatively biased thinking.

At all stages, cognitive psychotherapy is structured and directive. Each session follows a plan that is jointly formulated by the patient and therapist. Typically, a session begins with a review of the patient's reactions to the previous psychotherapy session, homework assignments, and the patient's salient experiences in the interim. Patient and therapist may then discuss possible topics and goals for the current session. Often, depressed patients feel overwhelmed and need help breaking down their problems into individual target symptoms that can be prioritized and dealt with one at a time. Target symptoms that might be the focus of a session include intense sadness, pervasive self-criticism, passivity and avoidance, sleep disturbance, or other affective, motivational, or cognitive manifestations of depression. The therapist repeatedly formulates the patient's beliefs and attitudes as testable hypotheses and helps the patient devise and implement ways of verifying them. The therapist maintains an inquiring attitude toward the patient's reactions to the therapist and the therapeutic procedures. Such reactions are explored for evidence of misunderstanding and distortion, which are then dealt with in the same way as the patient's other cognitions.

Comparison of Psychodynamic, Cognitive, and Interpersonal Brief Psychotherapies Because cognitive and interpersonal psychotherapies derive from the psychodynamic model, there is a high degree of overlap in the problem areas identified in any given patient. The conceptualization of the problem, however, differs. In many ways it is complementary rather than mutually exclusive. The psychoanalytic, interpersonal, and cognitive psychotherapies share explorative and change-oriented goals. Cognitive psychotherapy focuses on the patient's thinking; interpersonal psychotherapy, on the patient's interpersonal relations and social supports; psychoanalytically oriented treatments, on the internal experience of the patient and its relationship to past experience. Cognitive and interpersonal psychotherapies are most frequently used to treat depression rather than to treat an entire range of psychopathology. No studies using well-defined psychodynamic psychotherapies and medication have been performed. In this area, interpersonal psychotherapy and cognitive psychotherapy have been more closely studied.

Interpersonal psychotherapy cognitive psychotherapy, and the more traditional psychodynamic psychotherapies can be compared and contrasted. All three modalities are complex methods of treatment that must be tailored to the individual patient. All demand a high degree of clinical judgment, and the therapist needs a considerable amount of time to acquire competency. The relationship between the therapist and patient and establishment of a therapeutic alliance are essential in interpersonal psychotherapy, cognitive psychotherapy, and the psychoanalytically oriented treatments. Extensive exploration of the patient's thoughts and feelings, including those involving the therapist, are a major portion of the work. In addition, in all three, the therapist attempts to maintain an investigative, collaborative, and nonjudgmental stance (Table 30.8-8).

Time limited
Based on theoretical models derived from psychodynamic models: Recognize defenses (schema and cognitive distortions), transference (interpersonal patterns) and feelings and behaviors out of awareness
Published manuals specify the treatment
Can be used concurrently with medication
Empirical research indicates efficacy

Table 30.8-8 Common Features of the Brief Psychodynamic, Cognitive and Interpersonal Psychotherapies

In practice, cognitive psychotherapy is similar to the analysis of defense in the psychodynamic approaches. Understanding defenses focuses the patient and therapist on the hidden cognitive distortions that result in the patient's faulty perception of both the internal and external world. In the dynamic model, defense mechanisms are directed toward the control of anxiety resulting from conflict. The defenses, however, distort perception and cognition—similar to the distortions that are the focus of cognitive psychotherapy. In cognitive psychotherapy, cognitions are seen as causing the patient's distress. Much of the work in identifying these cognitions and alerting the patient to them is similar to the understanding and interpretation of defenses in the psychodynamic psychotherapies. The schemata underlying the faulty cognitions of cognitive psychotherapy are unconscious assumptions, which the psychodynamic model sees as derived from earlier experience. Both treatments share the importance of identifying these unconscious patterns of behavior and revealing them to the patient. To the extent that a psychodynamic psychotherapy focuses on the here-and-now experience of the patient rather than on the reconstruction of past experience, the similarity to cognitive psychotherapy increases. Frequently understanding a defensive pattern used by a patient to handle an ongoing conflict can be the endpoint of a well-conducted psychodynamic psychotherapy. In such a case, the outcome for cognitive and psychodynamic individual psychotherapy might be quite similar.

In the degrees of structure and directiveness, however, the two types of psychotherapy are different. In psychodynamic psychotherapy, the structure of the session is largely determined by the flow of the patient's thoughts and their interaction with the therapist's interpretive comments. In contrast, cognitive psychotherapy sessions are structured by an agenda that focuses the patient's thoughts and activities. In psychodynamic psychotherapy, the role of the therapist is limited to that of an empathic interpreter and sharer of the patient's experiences; in cognitive psychotherapy, the therapist may direct, prescribe, enjoin, educate, train, or role play as well. Furthermore, cognitive psychotherapy places much more emphasis on directly altering psychopathology rather than facilitating its alleviation through insight and resolution of inferred underlying conflicts.

Interpersonal psychotherapy is most closely related to the psychodynamic object relations' perspective; understanding internal objects rests upon understanding the actual interpersonal relationships of the patient, including the relationship with the physician. Both interpersonal and psychodynamic psychotherapy share a focus on identifications and transference, although IPT does not interpret transference. (Interpersonal psychotherapy defines a *paradigm* similar to *transference* as “past models for relationships.”) In addition, interpersonal psychotherapy pays particular attention to withdrawal and detachment, areas related to defenses in the psychodynamic model and faulty cognitions in the cognitive model. Interpersonal psychotherapy identifies interpersonal, rather than intrapsychic or cognitive, events. This frequently means that the interpersonal psychotherapy therapist's attention is directed to the same area of disturbance as the cognitive or psychodynamic psychotherapist's. However, the identified problem—interpersonal deficits, faulty cognitions, or intrapsychic conflict—is different.

Differences among the interventions of these psychotherapies are more striking than the differences in their goals or identified problem areas. To what extent it is the differences in psychotherapeutic treatments and to what extent it is their similarities that produce behavioral change is less clear. Both cognitive and interpersonal psychotherapy use more directive and behavioral interventions than psychodynamic approaches. Interpersonal psychotherapy and cognitive psychotherapy make more use of teaching new behavioral skills. More than the cognitive and interpersonal treatments, the psychodynamic psychotherapies rely on the patient to activate and practice new behaviors without direction. The briefer psychotherapies interpersonal psychotherapy, cognitive psychotherapy, and brief psychodynamic psychotherapy lack the extended working-through and application period of psychoanalysis and intensive (long-term) psychodynamic psychotherapy.

EFFICACY: RESEARCH AND EVALUATION

The brief psychodynamic treatments have a small empirical database, but much further research is needed. In general, these studies have supported the efficacy of this treatment approach. However, methodological issues are prominent in most research in this area. The development of handbooks for treatment has gone far in improving research in the psychoanalytically oriented brief treatments.

The effectiveness of psychotherapy is not argued as in the past. The brief psychodynamic psychotherapies have an effect size similar to many other medical treatments. However, which psychotherapy should be used for which patient by which therapist is still unclear. The cost-effectiveness of psychotherapeutic treatment remains hotly debated and is a focus of much current research. Individual psychotherapy results in fewer days of hospitalization for patients on medical or surgical services of a general hospital. In health clinics or health maintenance organizations, brief psychotherapy decreases the number of visits to primary health care providers, reduces the number of laboratory and X-ray studies, decreases the number of prescriptions given, and overall reduces direct health care costs. Recent summaries of the cost-offset effects of outpatient mental health treatment, most of which was short term, are hopeful but not unambiguous. One study found that outpatient psychotherapy reduced medical care utilization by an average of 33 percent. Furthermore, these reductions occurred mostly in the more expensive, inpatient medical services. In another study, 72 patients with significant emotional problems, treated only by internists in a general medical clinic, were compared with 62 patients who, in addition to being treated by internists for medical problems, received 10 weekly psychotherapy visits. The groups had approximately the same degree of emotional disturbance. At 4-month and 1-year follow-ups, the brief psychotherapy group reported significantly more global improvement than did the nonpsychotherapy group. Also, more patients in the brief psychotherapy group were employed at 1-year follow-up than in the nonpsychotherapy group. This study suggests specific beneficial effects of brief psychotherapy used in a medical setting by skilled psychotherapists.

Malan's finding of the importance of making the transference/parent link for successful outcome of treatment is significant and requires further exploration. Recently, a replication of his finding has been published, although others have found that too often interpretation of the transference may be related to a negative outcome. More recently, studies have also shown that overuse of transference interpretations may lead to poorer outcome. Measures of the therapeutic alliance, particularly from the patient's perspective, contribute to outcome, usually modestly. In another study attempting to replicate Malan's work, a patient's ability to relate to another person correlated with outcome 2 and 5 years after completion of treatment. Dynamic changes evident at 2 years were also present after 5 years.

Strupp's studies of college students in brief psychotherapy tend to confirm the importance of interactional variables. He found that the quality of the therapeutic interaction and the handling of the transference and countertransference were critical to success or failure in treatment. Strupp's studies indicate that patients treated by nonprofessionally trained therapists, on average, are as much improved as patients treated by professional therapists. However, they also show that such nonexperienced therapists run out of relevant material and soon become unwilling to continue to treat patients over an extended period of time. One of the important tasks of training in psychotherapy may be developing the ability to endure with the patient and, over time, with numbers of patients. Technical training and a theoretical framework may allow the therapist to maintain a sense of competence, direction, and interest in the work that the nonprofessional therapist cannot maintain.

IPT combined with pharmacotherapy is more effective than IPT or amitriptyline (Elavil) given alone. As single therapeutic modalities, IPT and antidepressant therapy are roughly comparable in appropriate mild-to-moderate depression. Maintenance IPT without active medication has resulted in a significantly greater time without relapse than supportive care and placebo.

Beck's cognitive psychotherapy model has been studied the most extensively of any of the psychotherapeutic modalities. Cognitive therapy is more efficacious for acute-phase treatment than waiting-list control conditions. In combination with antidepressant medication, cognitive therapy has proved superior to low-contact medication control groups. In general, cognitive therapy has been comparable to other active psychotherapies (e.g., behavior therapy, brief dynamic therapy, pastoral counseling). It is the only individual psychotherapy that has been formally adapted for treatment of depressed psychiatric inpatients. Cognitive therapy has not yet

been adequately tested for use as a continuation, or maintenance-phase, therapy.

In contrast to most other psychotherapies, there is a small, growing literature on the efficacy of cognitive psychotherapy. Although the number is still relatively small, all studies examining the outcome of cognitive psychotherapy have found it an effective treatment. Cognitive psychotherapy is more effective than no psychotherapy in treating both depressed volunteers and psychiatric patients with diagnoses of depression. The results are somewhat more mixed when cognitive psychotherapy is compared with other forms of treatment. For example, cognitive psychotherapy has been found superior to relaxation training, traditional insight-oriented psychotherapy, behavioral therapy, and client-centered therapy, whereas social skills' training, scheduled pleasant activities, and behavior therapy have been found to be just as effective.

The pattern of results is similar when cognitive psychotherapy is compared with standard antidepressant treatment. Some studies have found that depressed patients do better with cognitive psychotherapy than antidepressants, while others have found the two types of treatment to be equally effective. Although patients may not have been receiving adequate doses of medication in some of the studies that found cognitive psychotherapy to be superior, no study has yet found cognitive psychotherapy less effective than medications in unipolar, depressed outpatients. In these studies, depression (mild to moderate) was diagnosed by standard criteria. Therapy was carried on for 12 weeks (20 sessions of cognitive psychotherapy). Whether some patients benefit more from cognitive therapy or from medication treatment is unclear. Two studies compared patients with endogenous or nonendogenous depression and found no difference in response to cognitive psychotherapy or antidepressants. Older patients with nonendogenous depression respond better to cognitive (or behavioral) psychotherapy than do those with endogenous depression. Private practice patients have responded better to cognitive psychotherapy than to antidepressants, while hospital clinic patients fared equally well with the two treatments. The reasons for this difference are not apparent.

Some evidence indicates that when cognitive psychotherapy is found superior to tricyclic drug treatment, the advantage is in the speed of recovery rather than the eventual outcome. Patients given imipramine (Tofranil) or cognitive psychotherapy improved to about the same extent by 1 year after treatment, although the cognitive psychotherapy group was less symptomatic immediately following treatment. Similarly, patients who received cognitive psychotherapy did about as well 3 months after treatment as those who received treatment as usual (variable doses of antidepressants plus supportive psychotherapy), although those in the cognitive group were more improved at the end of treatment. Since depression is generally a self-limited disease, it is not surprising to see recovery rates equalizing after longer periods of follow-up. The more interesting question of whether cognitive psychotherapy protects against recurrences of depression over longer periods of follow-up has not been addressed in the literature.

Cognitive techniques and perspectives have been applied to conditions including (but not limited to) eating disorders, alcohol abuse, control of smoking, pain management, panic disorder, and tolerance of invasive hospital procedures. Data on the efficacy of such treatments are more scattered than for depression, but the pattern of findings is similar.

In summary, research to date has demonstrated that cognitive psychotherapy is about as effective as antidepressants in unipolar, depressed outpatients. As with antidepressants, however, not all patients respond well, and there are no reliable predictors of treatment response. Similarly, it is not clear that there are any target symptoms or behaviors that are specifically affected by cognitive therapy compared with other treatments. The active ingredients of cognitive psychotherapy in the treatment of depression remain to be identified. Focal, directive psychotherapies, in general, appear to be more effective than traditional, unstructured psychodynamic psychotherapy for depressed patients, but no clear superiority of cognitive psychotherapy over other directive therapies has been demonstrated.

FUTURE DIRECTIONS

The brief psychotherapies are an important component in the treatment of numerous disorders, primarily the adjustment, anxiety, and mood disorders. Both alone and in combination with medication they are an effective part of the treatment armamentarium. Clinicians should be trained in both brief and longer-term treatments and their use in brief, intermittent, and maintenance therapy. Skill in the longer-term psychotherapies is important to developing skill in the brief psychotherapies, in which rapid establishment of the therapeutic alliance and the accurate assessment of transference and defense patterns are important.

Empirical studies comparing well-defined brief psychodynamic psychotherapy with cognitive and interpersonal psychotherapies are limited. Future research must address which form of psychotherapy may be most helpful for which patient. In developing research strategies, it would be helpful to relate this question to the mode or modes through which any particular patient can most effectively learn new behaviors. An individual's available learning path (e.g., through the study of cognitions, interpersonal relations, or subjective experience) is influenced by state, trait, and contextual variables. The process of learning in the psychotherapies and the brief psychotherapies in particular, a process of altering neuronal organization through behavioral (primarily verbal) means, may be influenced by the patient's diagnosis, medications, past history, cognitive style, developmental stage, affective availability, doctor-patient match, and other variables. The effectiveness of the therapist in any given modality is certainly a critical variable. The differences and similarities in the outcomes of the various brief psychotherapies also require study.

SUGGESTED CROSS-REFERENCES

Learning theory is discussed in Section 3.3, and cognitive therapy in Section 30.6. Interpersonal therapy is covered in Section 30.7.

SECTION REFERENCES

*Balint M, Ornstein P, Balint E: *Focal Psychotherapy*. Lippincott, Philadelphia, 1972.

Barbar JP, Crits-Cristhoph P, Luborsky L: Effects of therapist adherence and competence on patient outcome in brief dynamic therapy. J Consult Clin Psychol *64:*619, 1996.

*Beck AT, Rush AJ: Cognitive therapy. In *Comprehensive Textbook of Psychiatry,* ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

*Crits-Christoph P: The efficacy of brief dynamic psychotherapy: A meta-analysis. Am J Psychiatry *149:*151, 1992.

Crits-Christoph P, Barber JP, editors: *Handbook of Short-Term Dynamic Psychotherapy*. Spectrum, New York, 1991.

Davanloo H, editor: *Short-Term Dynamic Psychotherapy*. Aronson, New York, 1980.

Frances A, Perry S: Transference interpretations in focal therapy. Am J Psychiatry *140:*405, 1983.

Gabbard GO: Mind and brain in psychiatric treatment. Bull Menninger Clin *58:*427, 1994.

Gabbard GO, Lazar SG, Hornberger J, Speigel D: The economic impact of psychotherapy: A review. Am J Psychiatry *154:*147, 1997.

Hoglend P: Long-term effects of transference interpretations: Comparing results from a quasi-experimental and a naturalistic long-term follow-up study of brief psychotherapy. Acta Psychiatr Scand *93:*205, 1996.

Holmes J: Psychodynamics narrative and intentional causality. Br J Psych *173:*279, 1998.

Horowitz MJ, Marmar C, Krupnick J, Wilner K, Kaltreider N, Wallerstein R: *Personality Styles in Brief Psychotherapy*. Basic Books, New York, 1984.

Howard K, Kopta S, Krause M: The dose-effect relationship in psychotherapy. Am Psychol *41:*159, 1986.

Hoyt M, Farrell D: Countertransference difficulties in a time-limited psychotherapy. Int J Psychoanal Psychother *10:*191, 1984.

Kandel G: A new intellectual framework for psychiatry. Am J Psychiatry *135:*457, 1998.

Klerman GL, Weissman MM, Rounsaville BJ, Chevron ES: *Interpersonal Psychotherapy of Depression*. Basic Books, New York, 1984.

Lazar SG, editor: *Extended Dynamic Psychotherapy: Making the Case in an Era of Managed Care, Psychoanalytic Inquiry Supplement*. Analytic Press, Newark, NJ, 1997.

Luborsky L: *Principles of Psychoanalytic Psychotherapy: A Manual for Supportive Expressive Treatment*. Basic Books, New York, 1984.

Luborsky L, Auerbach AH: The therapeutic relationship in psychodynamic psychotherapy: The research evidence and its meaning for practice. In *Psychiatry Update: The American Psychiatric*

Association Annual Review, vol 4, RE Hales, AJ Frances, editors. American Psychiatric Press, Washington, DC, 1985.

MacKenzie KR, editor: The time-limited psychotherapies. In *Review of Psychiatry,* LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1996.

Malan DH: *A Study of Brief Psychotherapy*. Plenum, New York, 1975.

Malan DH: *Toward the Validation of Dynamic Psychotherapy*. Plenum, New York, 1980.

Mann J: *Time-Limited Psychotherapy*. Harvard University Press, Cambridge, MA, 1973.

Mann J, Goldman R: *A Casebook in Time-Limited Psychotherapy*. McGraw Hill, New York, 1982.

Markowitz JC, Svartberg M, Swartz HA: Is IPT time-limited psychodynamic psychotherapy? Psychother Pract Res *7:*185, 1998.

Marziali EA: Prediction of outcome of brief psychotherapy from therapist interpretive interventions. Arch Gen Psychiatry *41:*301, 1984.

Miller NE, Luborsky L, Barber JP, Docherty JP, editors: *Psychodynamic Treatment Research*. Basic Books, New York, 1993.

Rush AJ, Giles DE: Cognitive therapy: Theory and research. In *Short-term Psychotherapies for Depression*, AJ Rush, editor. Guilford, New York, 1982.

Sifneos PE: *Short-Term Psychotherapy and Emotional Crisis*. Harvard University Press, Cambridge, MA, 1972.

Sifneos PE: The current status of individual short-term dynamic psychotherapy and its future: An overview. Am J Psychother *37:*472, 1984.

Sledge WH, Moras K, Hartley D, Levine M: Effect of time-limited psychotherapy on patient dropout rates. Am J Psychiatry *147:*1341, 1990.

*Strupp HH, Binder J: *Psychotherapy in a New Key: Time-Limited Dynamic Psychotherapy*. Basic Books, New York, 1984.

Ursano RJ, Silberman EK: Individual psychotherapies. In *American Psychiatric Press Textbook of Psychiatry*, JA Talbott, RE Hales, SC Yudofsky, editors. American Psychiatric Press, Washington, DC, 1988.

*Ursano RJ, Sonnenberg S, Lazar S: *Concise Guide to Psychodynamic Psychotherapy in the Era of Managed Care: Principles and Techniques*. American Psychiatric Press, Washington, DC, 1998.

Weissman M, Klerman GL, editors: *New Applications of Interpersonal Psychotherapy*. American Psychiatric Press, Washington, DC, 1993.

*Weissman MM, Markowitz JC, Klerman GL: *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books, New York, 1999.

Wiborg IM, Dahl AA: Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? Arch Gen Psychiatry *53:*689, 1996.

30.9 ERIKSONIAN CLINICAL THEORY AND PSYCHIATRIC TREATMENT

EDWARD R. SHAPIRO, M.D., AND M. GERARD FROMM, PH.D.

Erik Erikson (Fig. 30.9-1) made significant contributions to a broad range of disciplines, but the application of his theories to clinical practice has not received proper recognition. Although considered among the group of ego psychologists that transformed psychoanalytic theory in the 1950s and 1960s, Erikson's emphasis on how the individual is inextricably bound to sociocultural and historical forces has broad and largely underdeveloped clinical significance.

FIGURE 30.9-1 Painting of Erik Erikson (by Norman Rockwell, Courtesy of Edward R. Shapiro, M.D.).

David Rapaport described Erikson's work as the culmination of ego psychology, the first psychoanalytic theory of the person's relationship to social reality with which the ego is always engaged. Rapaport noted, “[Erikson's] concept of mutuality specifies that the crucial coordination is between the developing individual and his human (social) environment.” This concept extends clinical work beyond the boundaries of the individual and opens it to larger scrutiny. From his clinical study of patients, Erikson recognized how the pathology and recovery of patients are linked to how they engage and are engaged by their world.

THEORETICAL ISSUES

Erikson noted that people get as invested in ideas as they do in people. His clinical and theoretical approach included questions of commitment, engagement, and integrity. Integrity was central to psychological health. He defined it as “man's obligation to the most mature meaning available to him, even if this should bring discomfort to himself, deprivation to his mate and offspring, and the loss of friends, all of which must be imagined and endured in order not to be exposed to a final sense of disgust and despair.” Integrity transcends the pair, or dyad, and requires discovery of larger social tasks to which the individual can become committed. This formulation illuminates Erikson's clinical perspective of humans as social beings, linked to larger historical and cultural developments. Integrity and psychological health require discovering this connection.

Erikson had a broad view of the task of psychoanalysis. He considered it a tool for examining “that aspect of man which in a given historical period was being neglected, exploited, suppressed by prevailing technology and ideology.” Given the immediate dependency on the human environment in early life, he considered the study of children and adolescents a crucial window for developing psychoanalytic theory. Erikson conceptualized the identity crisis that occurred during adolescence as a necessary integration of the adolescents' social reality with their history and new biological capacities. Identity formation is an activity of the ego that derives from the relation of individuals' self-definition to developing social definitions of those individuals as they mature. This notion opened the study of interaction, family process, and social change as mediators of clinical phenomena.

TECHNIQUE

According to Erikson, the ability of the patient's ego to mediate between the demands of unconscious internal pressures (i.e., instincts and affects) and the pressures of the external world (i.e., relationships, social reality) was a primary focus of treatment. In Erikson's thinking, the surface (i.e., the ego) revealed the depth. When he listened to patients, he listened with an artist's ear—he heard the words, saw the spatial relationships, and made temporal links between the somatic, affective, and interpersonal parts of the personality. He listened to texture and saw configuration. He paid attention to the whole life context of any immediate situation. He asked a number of questions. What is the immediate stimulus for the patient's reaction? What is the acute life conflict, the current developmental stage, the issues that are manifest? In what developmental context did the patient's reaction first occur? Is it now manifest in the relationship to the therapist, in a repetitive conflict, in a characteristic way that the individual solved earlier developmental struggles? In what social context is the individual embedded, what roles are available? What are the typical stereotypes, particular opportunities, and barriers? What are the characteristic ways in which the individual takes in information? What defenses does the individual use? Where are the individual's deepest psychological investments?

Like his contemporary, Donald W. Winnicott, Erikson had a family perspective. He recognized that children react to the shared family unconscious, reading their parents' vulnerabilities and unexpressed wishes with clarity. Studying seriously disturbed patients, he saw that their breakdowns occurred at times of developmental separation or individuation (e.g., physical intimacy, occupational choice, or identity formation). Erikson's study of identity revealed the consistently accruing social value of increasing commitment to and from others. He saw social recognition as providing defenses against impulses, and he saw work as a means to consolidate conflict-free achievements. Work provided an opportunity to resynthesize childhood identifications in new ways in accord with available roles.

Role of the Therapist Erikson felt that, in the consulting room, therapists had to be aware of both their own and the patient's obligations to others. Therapists have methodological responsibilities and must understand both their role and their motivations. Self-observational vigilance is essential, and Erikson believed that therapists cannot see in another what they have not discovered in themselves. For Erikson, the therapeutic contract with the patient is the essence of study. As new data emerge from the patient, the therapist must be committed to continuous conceptual revision. Establishing a state of trust between doctor and patient is a basic requirement. Erikson felt that the therapist needed a sense of personal trustworthiness that could be transmitted to the patient.

CLINICAL ISSUES

Although no independent Eriksonian psychoanalytic school exists in the same way as the Freudian and Jungian schools, Erikson made many important contributions to the therapeutic process, many of which can be gleaned from examining his rare reports of patients he treated. One such case was reported in his seminal paper “The Nature of Clinical Evidence,” published in 1957. In that paper, Erikson writes that the concepts of “repression,” “regression,” “transference,” and “libido” are essential to his clinical way of thinking. He adds that he tries “to keep each linked with the observation and experience of the clinical encounter as a new event in the patient's life history.” Examination of unpublished aspects of this case (available to the authors) reveals a complex clinical theory that was ahead of its time.

The patient, a seminarian in his 20s, presented with anxiety, hopelessness, “psychic pain” in his body, and suicidal thoughts. He was distressed by impulses to hurt others and obsessively turned to prayer for help, feeling increasingly alone and that his impulses were terribly incompatible with his religious aspirations. These symptoms were the latest attack in a series of such symptoms, beginning when the patient was a senior in high school, following his maternal grandfather's death.

The oldest of five children, the patient was unusually close to his mother who raised him according to the rigid schedule then in vogue. He idealized his father, who was largely absent from the home and “unable to deal with feelings.” He had significant separation problems in kindergarten. Throughout his childhood, he found his mother to be alternatively happy and depressed, loving and withdrawn, nagging and martyred. The patient was uncomfortable with her physical attention to him and her overprotectiveness. Most of the patient's comments concerning family discord centered around his mother's changing moods and his feeling of impotent rebellion against her. The patient felt he was a great deal more like his mother in personality than his father, which was a source of concern to him. He was very close to his mother's father, with whom he spent summers in a rural community. Mother was very supportive of their relationship. Shortly before grandfather's death during his adolescence, the patient was rebellious to him. Grandfather's death led to a significant depression in the mother. Shy and a good student, the patient was preoccupied with issues of power, prestige, and acceptance by others.

Psychological testing confirmed borderline psychotic features in an inhibited, obsessive-compulsive personality, struggling to maintain control over aggressive impulses but also able to distance himself from these feelings well enough to manage them and examine them in therapy.

Erikson's published summary of this case reports one of the patient's dreams and the clinical crisis it suggested: “There was a big face sitting in a horse and buggy. The face was completely empty, and there was horrible, slimy, snaky hair all around it.” The patient stated anxiously, “I'm not sure it wasn't my mother.”

The patient's associations indicated that the “horse and buggy” referred to his beloved grandfather. In the patient's history, he had shown delinquent tendencies, defying his grandfather just before his death. Erikson wondered whether the patient felt that his anger had destroyed his identity supports. His mother had been frightened by his rebelliousness, and he had come to feel controlled by her overprotectiveness. Erikson thought of the Medusa-like image in the dream as the “female void” and considered the meanings of “loss of face” and lack of identity. The general theme of the dream seemed to Erikson to be “when I have faith in another's love and strength, angry feelings come up and I end up mistrusting, empty and full of despair.”

Dreams and Free Association Like Freud, Erikson worked with the patient's associations to the dream as the “best leads” to understanding its meaning. He valued the first association to the dream, which he believed to be powerful and important. Ultimately, Erikson listened for “a central theme which, once found, gives added meaning to all the associated material.”

Erikson believed interpretation to be the primary therapeutic agent, sought as much by the patient as by the therapist. He emphasized free-floating attention as the method that enabled discovery to occur. Erikson once described this attentional stance by commenting that in clinical work, “You need a history and you need a theory, and then you must forget them both and let each hour stand for itself.” This frees both parties from counterproductive pressures to advance in the therapy and allows them both to notice the gaps in the patient's narrative that signal the unconscious.

In relation to the patient described above, Erikson noted “the seemingly paradoxical fact that during his prior hour [before the dream] the patient had spoken of an increased well-being” or again, “Although the patient acted as if he were close to a breakdown, I had the impression that, in fact, there was a challenge in all of this, and a rather angry one.” Beside the central dream image of a horrifying empty face, Erikson notes “omission of important items present in most dreams: motion, action, people, spoken words.” He surveys this surface of the dream as a way of placing it in context; each omission or inclusion has potential significance.

Repression Erikson wrote of his patient: “I have to assume that the patient is (to varying degrees) unconscious of the meaning which I discern in his communications, and that I am helping him by making fully conscious what may be totally repressed, barely conscious, or simply cut off from communication.” He described his patient—and by extension all patients—as a split or conflicted subject, whose self-esteem, stability, and relationships were precariously maintained by an enduring, unconscious obliteration of some core emotional experience. For Erikson, “the traumatic past is of course a present frontier, perceived as an acute conflict.” He saw the ego as an integrative force directed toward overcoming repression.

Regression and the Setting Erikson also described a “regressive trend, a going back to earlier failures in order to solve the past along with the present.” This point of view locates Erikson among the developmental psychoanalysts, like Winnicott, who view regression as a persistent, unconscious effort toward reintegration. They thus follow and extend Freud's later theoretical revisions that explore the compulsion to repeat as an effort to master trauma.

In this positive and optimistic view of regression, Erikson does not neglect his clinical responsibility to the patient's functional regression. “Our patient's behavior and report confront me with a therapeutic crisis, and it is my first task to perceive where the patient stands as a client, and what I must do next.” Erikson felt the patient's panic in and about his dream. He wrote that “even in the hour of reporting, the dream-state seemed still vivid enough to threaten the patient's sense of reality.”

His assessment of the patient's regression not only took into account the emotional tone of the patient's narrative in the session, but “first and last depends, of course, on the *setting* of his work.” Since Erikson's patient was in an open residential treatment center (i.e., the Austen Riggs Center), he was already familiar with the setting and its “calculated risks” and “special opportunities.” He knew that the patient's adaptation to Riggs had been assessed and established. He could thus conclude that “a dream-report of the kind just mentioned, in a setting of this kind, will first of all impress the clinical observer as a diagnostic sign. This is an ‘anxiety dream'.” This statement requires emphasis because along with a number of other comments, Erikson pointed to the setting as a “third” element to the patient-therapist dyad. The setting has symbolic meaning that influences the psychotherapeutic work and gives it special meaning. Context organizes meaning, and Erikson expanded the notion of the relevant context to include multiple levels: environment, family, society. Regarding the patient's dream reported on above, Erikson stated: “whether this dream is the sign of an impending collapse, or. . . a potentially beneficial clinical crisis” depends on its context. In this case, the patient's successful adaptation to the environment pointed Erikson toward the latter interpretation (i.e., potential benefit).

Erikson believed that emotional conflicts of childhood were played out again as part of current developmental tasks. Dream analysis led Erikson to recognize that “these patients [i.e., borderline patients] are deeply, if unconsciously, convinced that they have caused a basic disturbance in their mothers.” The “violated' mother” and the associated “images of guilt” become “an obstacle in the resolution of adolescence—as if a fundamental and yet quite impossible restitution were a condition for adulthood.” To move forward, Erikson's patient must first move backward; this is regression in the service of the ego.

Erikson elaborated a clinical point of view that considered depending on the analytic setting and the relationship to the therapist central to change. Describing the patient above, he wrote:

> [H]e had taken a real chance with himself and with me. Under my protection and the hospital's he had hit bottom by chancing a repetition of his original breakdown. He had gone to the very border of unreality and had gleaned from it a highly condensed and seemingly anarchic image. Yet the image, while experienced as a symptom, was in fact a kind of creation, or at any rate a condensed and highly meaningful communication and challenge, to which my particular clinical theory had made me receptive.

Erikson's view of regression was ultimately an optimistic one. He says, “I would not give the past a kind of fatalistic dominance over the present.” He believed that “we all relive earlier and earliest stages of our existence in dreams, in artistic experience, and in religious devotion, only to emerge refreshed and invigorated.”

Libido Erikson considered libidinal attachments and experiences of dependence and abandonment of paramount importance. Erikson's concept “specifies that the crucial coordination is between the developing individual and his human (social) environment, and that this coordination is mutual.” His theory is not one of quanta of energy but of relationships. Erikson focused on the power of abandonment and dependence. He emphasized and examined attachment, separation, and mutuality in the multiperson system.

Erikson's patient associated the horse and buggy on which the horrifying empty dream-face was seated with his beloved maternal grandfather and, by extension, with the loss of the rural caretaking culture of his childhood. The nostalgia for his grandfather turned out to be the prevailing mood of his mother and the pervasive, unnamed emotional communication from her to her son. Simultaneously, she unconsciously communicated that anything aggressively individuating in him would prove deadly, or at least crushingly disappointing to her. This is a familiar shared fantasy in families of borderline patients, in which the child's separation represents and requires a hateful devaluation of the family. In Erikson's patient, the mother's traumatic loss of her father pushed her unconsciously toward an irrational claim on

her son; his necessary developmental process threatened to reopen her repressed grief. Erikson described this as missed mutuality. Although Erikson did not develop a family group formulation, he did highlight the absent father who might have provided both concretely and symbolically some “third” perspective on this conflicted dyad.

In his 1950 paper “The Theory of Infantile Sexuality,” Erikson outlined his thinking about “libidinal disbalance” along with a hint of his clinical approach. His crucial leap beyond Freud was to give full recognition to the idea that erogenous zones are the sites of highly charged exchanges and interactions with the environment, another aspect of mutuality. Anality thus represents not only a particular site of stimulation, but the action modes of holding onto or letting go and the relational modes of submission or defiance. Erikson thus described a rudimentary action language that is embedded in the body and that helps structure the psyche. His therapeutic approach requires that clinicians pay attention to the patient's actions, particularly those that have not yet been put into words (e.g., the affects surrounding the dream).

Transference Erikson described transference as an important part of treatment. “I would also acknowledge the power of *transference*, that is, the patient's transfer to me of significant problems in his past dealings with the central people in his life; but I would know that only by playing my role as a new person in his present stage of life can I clarify the inappropriateness of his transferences from the past.” He suggests that the context in which transference can be interpreted is the real therapeutic role in the relationship between doctor and patient. Inquiry and care, not indulgence or exploitation, is a new experience for these patients, which contributes to therapeutic change.

Erikson felt that clinical psychoanalysis relied methodologically on “disciplined subjectivity.” By that he meant that he not only attended to what patients said about their dreams and how they behaved in the telling, but he attended also to what he himself thought and felt as he sat with the patient. In focusing on the transference, Erikson considered whether the patient was breaking down or trying to communicate. He noted that “the first would mean that he is slipping away from me. . . the second, that he is reaching out for me with an important message which I must try to understand and answer.” Erikson was completely attentive to the patient's actions in relation to him, and he believed that the therapist had a role-related responsibility for answering or interpreting.

Erikson considered his own reactions to the horrifying empty face of his patient's dream. As the patient reported the dream, Erikson wrote that his “facial and tonal expression reminded me of a series of critical moments during his treatment when he was obviously not quite sure that I was ‘all there' and apprehensive that I might disapprove of him and disappear in anger.” In analyzing dreams, Erikson believed that the therapist “does well to raise discreetly the masks of the various dream persons to see whether he can find his own face or person or role represented.” By so doing, Erikson finds not only transference as distortion but transference as the actual current version of an earlier, but still powerful issue. Indeed, Erikson considers how the patient's transference may be “right” in some aspects. His patient's dream, for example, included both the recent sense that the patient trusted him and an earlier tenuousness in their relationship in response to Erikson's going off for an emergency operation months before. He described that experience as “inflicting his mortality” on the patient. “At the time of this dream-report I still was. . . mildly uncomfortable.” Thus the patient experienced “conflict between his sympathy, which makes him want to take care of me, and his rightful claim that I should take care of him.” Like the dream figure, Erikson had wavy hair.

Erikson comes to a partial interpretation that translates the dream event into a sentence:

> I concluded that the empty face had something to do with a certain tenuousness in our relationship, and that one message of the dream might be something like this: “If I never know whether and when you think of yourself rather than attending to me, or when you will absent yourself, maybe die, how can I have or gain what I need most—a coherent personality, an identity, a face?”

Whatever other meanings there might be to the dream, whatever other central life figures might eventually show through the empty face, Erikson emphasized transference analysis. “I put first (the patient's) fear that he may yet lose himself by losing me too suddenly or too early.”

Countertransference Erikson did not explicitly name countertransference, yet he was aware of it and used it technically. Recognizing the protection given by his professional role, Erikson noted, “The clinician. . . finds himself part of another man's most intimate life history. Luckily he also remains the functionary of a healing profession.” With regard to using the countertransference or, as Erikson puts it, “the disposition of the clinician's ‘mixed' feelings,” Erikson believed that “the therapist must use his own emotional responses during a clinical encounter as an evidential source and as a guide in intervention. . . Patients of the type of our young man, still smarting in his 20s under what he considered his mother's strange emotions in his infancy, can learn to delineate social reality and to tolerate emotional tension only if the therapist can juxtapose his own emotional reactions to the patient's emotions.”

Erikson concluded about countertransference that [a]ny psychotherapist. . . who throws out his ethical sentiments with his irrational moral anger, deprives himself of a principal tool of his clinical perception. . . [O]ur indignation, admitted and scrutinized for flaws of sulkiness and self-indulgence, is, in fact, an important tool both of therapy and of theory.” The therapist's reaction represents, for Erikson, an important aspect of “available emotion and responsive thought.” The therapist must maintain a “disciplined subjectivity.” He must also, “while facing most intimate and emotional matters. . . maintain intellectual inner contact with. . . conceptual models.”

For Erikson, countertransference is used to make sense of the transference and also help the patient consolidate reality testing. Erikson described the previous hours and the way in which the patient presented the dream as feeling like an angry challenge. He noted that “the patient's main resistance consisted of testing me by a whining tone of voice, a child's voice saying, ‘I don't want you to go away, but I know I will make you go away by my not wanting you to.' “He associated the patient's reaction to his operation and the psychological testing that had revealed the patient's precarious control over his aggression. He communicated this reaction to the patient in the form of an interpretation.

Interpretation Erikson told his patient “without anger, but not without some honest indignation, that my response to his account included a feeling of being attacked.” He suggested that the patient was protesting, “Why are you not doing for me what my grandfather did? Why is my father not like my grandfather? Why do all of you father figures abandon me to my mother's masochistic rage? Why is there death, mutilation and rage in the world? How can I sustain love, i.e., *keep your face alive in me*, when my mother's face always disintegrated, making me feel that I had destroyed her?” Erikson goes on to describe feeling worry and pity. He said that he had been touched by the patient but also challenged to prove something. He felt imposed upon with the “burden” of the patient's “future which (the patient) could well learn to manage.” Erikson's final interpretation went something like this: “I feel attacked, worried, pitying, touched, and challenged to prove the goodness of mothers, the immortality of grandfathers, my own perfection, and God's grace.”

This interpretation, a playful tour de force, took into account the manifest dream, the patient's affective presentation and associations, the transference repetition of the patient's rage at the mother and at the grandfather's death, the narcissistic wishes for an idealized figure around whom to build an identity, and Erikson's countertransference irritation. Erikson says, “By relating the fact that his underlying anger aroused mine, and that I could say so without endangering either myself or him, I could show him that in his dream he had also confronted anger.” This is a way of facing the patient. Erikson is describing the therapist surviving the patient's aggression, without retaliation or capitulation, a capacity now considered essential in work with patients with personality disorders. Erikson says of his patient, “[W]hile accepting his transferences as meaningful, I had refused to become drawn into them.” Instead, he compensates for this therapeutic refusal by “offer[ing] him my help as defined by my professional status in attempting to understand what was behind his helplessness.”

“Clinical evidence,” Erikson writes, “is characterized by an immediacy which transcends formulations.” Erikson's patient was “amused, delighted and encouraged” by his confrontative interpretation. Erikson points out that “[t]he patient left the hour—to which he had come with a sense of dire disaster—with a broad smile.” He concludes that the “proof (of an interpretation) lies in the way in which the communication between the therapist and patient ‘keeps moving,' leading to new and surprising insights and to the patient's greater assumption of responsibility for himself.”

The interpretation condensed Erikson's irritation at the patient's “whining” demand and his “desperate insistence on finding security,” while overcoming his countertransference wish to take care of the patient. The interpretation revealed Erikson's confidence in his patient's capacity to deal with confrontation. It affirmed the patient's inner resources to continue his development, and indicated Erikson's appreciation of the risk his patient had taken by symbolically repeating the original trauma in the dream. The “new relationship” captured by the interpretation is between the patient and an adult in a therapeutic role, in which understanding is possible.

Action and Actuality Not only did Erikson pay attention to action as communication, he also supported it as potentially important engagement. He affirmed the healing role of play and its derivative, work. Erikson defined *actuality* as “the world of participation, shared with a minimum of defensive maneuvers and a maximum of mutual activation.” He emphasized how an individual's maturing capacities evoke cooperative responses from others that are needed for joint survival.

Erikson's formulation about actuality derived from his study of Freud's Dora case, during which the treatment foundered on Freud's uncertainty about what Dora needed from him. Erikson argued that Dora needed to have Freud recognize the historical truth of her family's infidelities. Recognition of potential historical truth is not the same as the practice of “validating” historical truth (e.g., childhood abuse). Erikson knew that the therapist could not know the truth in some objective way. That kind of presumed validation can represent a defensive collusion between patient and therapist to ward off the patient's current aggression. To suggest that patient and therapist can know the whole truth collapses the space for inquiry, for not yet understanding, and for perspective. Erikson means to hold onto perspective. The recognition of potential historical truth has to do with the young person's need to have an adult affirm the reality of family dynamics and the possible accuracy of her interpretation of those dynamics. This testing of reality could be approached another way. The therapist could conduct a family meeting in which the patient's interpretation of her family could be explored and tested. Erikson suggests that Dora required an external affirmation of reality. Only then could Dora consider the role of her own impulses in these events. Dora's family was an example of failed mutuality. Erikson grasped the way adolescents need help in recognizing and coming to terms with this shared failure before daring to venture beyond.

In the informal clinical notes on his patient, Erikson explored the world of action in some detail. Because his patient felt unconsciously that he had damaged his mother, he found himself preoccupied with sexuality as a means of both separation and reparation. Successful sexual intercourse, he felt, would “break through to reality, establish himself as a man, [and develop] a lasting ability to think clearly.” There were certainly women patients at Austen Riggs willing to help him, but in line with his commitment to the patient as embedded in the larger community with its values, this evoked in Erikson the need to “firmly insist on some. . . precautions for the protection of Riggs, of the lady (who might become) involved, and of himself.” Erikson described these potential partners as “uprooted.” They could not be expected to keep faith with the patient nor he with them. He stated, “Neither damnation nor salvation will result from these experiences.”

This prohibition served two purposes: it provided a frame within which the patient's anxieties could be contained, and it evoked the patient's transference fantasies that Erikson wanted all the women for himself or, as mother figure, that he would tear out his hair in response to the patient's exploration. When the patient developed sexual fantasies about these women, who were older, Erikson felt he “should not shy away from such material but should rather delineate its reality in an infantile context and its symptomatic nature in the present adult context.” By this, he was referring to incestuous fantasies. Erikson encouraged therapists not to avoid guiding patients; he believed that therapists must offer patients both prohibitions and permissions.

Work In his clinical notes, Erikson added to his ideas about identity the relationship to work and the development of a professional role. He noted that the patient's breakdown occurred in the context of an urgent professional dilemma: wanting to become a monk was “the last step in a series of attempts at becoming somebody who was ‘to help weak people,' while recanting all competition.”

Joan Erikson, Erik's wife, developed an activities program at the Austen Riggs Center as an “interpretation-free zone” where patients could take up work roles or function as students with artists and craftspeople, without the burden of the patient role. This work space encouraged the play and creativity required for the patients' work development to parallel the process of their therapy. This patient moved from pottery in the program, to musician, to bartender in a local inn, to laborer with a construction business, to his professional career. Erikson notes that the pottery preserved the patient's capacity to function as a “good worker,” though he felt the patient was in danger of “becoming a slave” to the clay. As a musician, to his surprise, Erikson's patient felt the fusion of his mother's loving voice with his own noisy infantile one, allowing him to feel a safe connection to her and to realize he had not destroyed her. Once he took a real-world job, he found himself struggling with what Erikson called, “the boss problem. . . the co-worker problem and. . . the work-role problem,” all of which held him safely in a role but pushed him to face his fear of loss of ego boundaries, for example, by reacting to work problems in an all-or-nothing way and ragefully ruining his work in the process. Erikson noted that:

> Impulse control is different in a bartender than in a minister, and so are the small social crises—or so it seems, until one tries it out. Then it becomes apparent that positions of trust, intermediary positions between owner and employees, positions highly susceptible to the weaknesses (and confessions) of others, have a lot in common. . . . It is such recognition of the *principles* of a work situation which. . . gives the patient eventually the freedom to choose his career not on the basis of role-appearance, but on that of inner compatibility of activity, role and personality.

Erikson's patient eventually became a doctor. Erikson commented:

> The patient's medical identity was not the sum of all his identifications with doctors, but. . . the inner coherence of his gifts and trends quite independently pointed in this direction:. . . the wish to help, now relatively freer from soul-saving fervor; the technical gifts of using his hands, now relatively freer from masturbation guilt, and tried on machines; psychological empathy, now relatively freer from over-identification, and tried in an industrial work-situation; an excellent mind for study;. . . liking for being depended on and for wearing a halo, now relatively freer from megalomanic and despotic trends; a freer wish to compete, even with his father for whom he felt a new respect; and behind it all, I assume, the age-old (unconscious) wish of doctors: wanting to cure mother, who seemed so incurable when he was a baby.

Further Clinical Considerations Erikson did not see himself as founding his own school of psychoanalytic psychotherapy nor indeed as departing from Freud in his clinical approach. Though his developmental road map enriches the Freudian psychosexual categories enormously—and offers the possibility of facilitating understanding and change throughout the life cycle—Erikson's clinical theory remained grounded in core psychoanalytic concepts creatively adapted to work with more severely disturbed patients.

Erikson considered the wish to get well, the synthesizing capacity of the ego, and a capacity for mutuality to be important indications that a successful treatment might be possible. He saw developmental trauma and conflicts associated with dependency, separation, and aggression as indications that such treatment might be needed.

APPLICATION TO PSYCHOPATHOLOGY

Erikson described sequential stages of the life cycle, each representing a success or failure of the mutual coordination between the child's development and the responses of the human environment. He related failures of each stage to particular pathological outcomes.

Basic trust versus mistrust—This stage marked the developmental struggle of early infancy. Failures contribute to depression, hopelessness, difficulties in engaging others, suspiciousness, and withdrawal (characteristics seen in disorders in the schizotypal and major depressive spectrum).

Autonomy versus shame or doubt—This characterized the period of separation-individuation and toilet training. Failures in this negotiation contributed to narcissistic vulnerability, delinquency, and obsessive behavior.

Initiative versus guilt—This struggle characterized the outcome of the oedipal conflict. Erikson saw failures here contributing to generalized anxiety, phobias, inhibitions, impotence, or psychosomatic illness.

Industry versus inferiority—Failures in this negotiation during the ages 7 to 11 contributed to work inhibitions, feelings of inadequacy, or a compensatory and defensive drive for money, power, and prestige.

Identity versus identity diffusion—Characterizing the adolescent struggle, failures here contribute to difficulties leaving home, prolonged dependence, the resurgence of borderline and narcissistic phenomena, and manifestations of role diffusion in work or career.

Intimacy versus isolation—Erikson saw the formation of healthy adult relationships as manifesting successful resolution of this stage.

Generativity versus stagnation—This dimension of mature adult life, which involves ongoing growth and commitment to what one has created, Erikson related to midlife crises.

Integrity versus despair—This late-life phase, having to do with the fulfillment of identity, Erikson connected to the depression-related episodes of old age.

Though Erikson's stage theory was revolutionary in the 1960s, subsequent study of parent-child relations have revealed that these phenomena are infinitely more complex. These stages repeat throughout the life cycle. Residues of earlier difficulties recur, and development has a much less linear shape, as do the influences of

the environment. Eriksonian diagnosis reflects this complexity. Rather than insisting on a one-to-one correspondence between developmental phase failure and a specific disorder, Erikson thought that a true developmental diagnosis would describe the range between the initial fixation and the later developmental arrest. In other words, diagnosis has to do with when the trouble started and how far the patient got in development before breaking down.

Goals of Treatment Erikson underlined missed mutuality as a quieter, yet equally traumatic warping of development and considered the restoration of mutuality to be a goal of treatment, as both a way of working and as an outcome. Additional goals included both the developmental advance that follows the resolution of particular life crises (e.g., a restored capacity for trust, the development of more-flexible autonomy, or the discovery of less-inhibited initiative), and the integration of self-awareness and relatedness to the world that combines to make up a coherent and recognizable identity. Erikson's treatment was aimed at helping the patient's ego become stronger and heal itself. A goal of therapy is to recognize how patients have passed or not passed through the various stages of the life cycle and how the various crises have or have not been met. He believed that psychological growth and development can occur throughout the life cycle.

Future Directions Erikson's clinical concepts suggest that learning is a two-way street. Crucial information is gained by “listening” therapeutically to how the younger generation responds to its interpersonal environment. Erikson adds that “it is in this mutuality of the development of the older and younger generations that certain and universal values such as love, faith, truth, justice, order, work, etc. in all of their defensive strength, compensatory power, and independent creativity become and remain important joint achievements of the individual ego development and of the social process.”

Drawing again on the clinical context, Erikson believed that “if ‘psychoanalyzed' man learns to recognize the fact that even his previously repudiated or denied impulses may be ‘right' in their refusal to be submerged without a trace (the trace being his symptoms), so he must also learn that his strongest ethical judgments are right in being persistent even if modern life may not consider it intelligent or advantageous to feel strongly about such matters.” This brings Erikson to an important conclusion: “this means that we somehow harbor a model of man which could serve as a scientific basis for the postulation of an ethical relationship of the generations to each other.”

Erikson's gift was to take the psychoanalytic lens, initially developed to look inward, and use it to look outward as well—to play, creativity, the lived and examined life, the social playground, adult development, collective conflicts, and organizational structures. His grasp of the inextricable links between the individual and the context, his recognition of “the interdependence of individual aspiration and social striving,” and his ability to hold the interests of community and society while strengthening the individual remain as challenges to the future development of clinical work.

SUGGESTED CROSS-REFERENCES

Erik Erikson's life and ideas are further discussed in Section 6.3, and Sigmund Freud's ideas are discussed most fully in Section 6.1. Other theories of personality and psychopathology are discussed in Section 6.3 and Section 6.4. Schizophrenia is discussed in Chapter 12, mood disorders in Chapter 14, personality disorders in Chapter 24, and psychosomatic disorders in Chapter 25. Normal child development and adolescent development are discussed in Section 32.2 and Section 32.3, respectively; adulthood is discussed in Chapter 50, normal human sexuality in Section 19.1, and normal aging in Section 51.2c. Psychoanalysis and psychoanalytic psychotherapy are discussed in Section 30.1.

SECTION REFERENCES

Erikson EH: *Childhood and Society*. Norton, New York, 1950.

*Erikson EH: The dream specimen of psychoanalysis. J Am Psychoanal Assoc *2:*5, 1954.

Erikson EH: The problem of ego identity. J Am Psychoanal Assoc *4:*54, 1956.

Erikson EH: The first psychoanalyst. Yale Rev *46:*40, 1956.

*Erikson EH: Case report discharge note. Austen Riggs Center, 1957.

Erikson EH: *Young Man Luther: A Study in Psychoanalysis and History*. Norton, New York, 1958.

*Erikson EH: Reality and actuality. J Am Psychoanal Assoc *10:*451, 1962.

*Erikson EH: The nature of clinical evidence. In *Insight and Responsibility*. Norton, New York, 1964.

Erikson EH: *Toys and Reasons: Stages in the Ritualization of Experience*. Norton, New York, 1977.

Freud S: Beyond the pleasure principle. In *Standard Edition of the Complete Psychological Work of Sigmund Freud*, vol 18. Hogarth Press, London, 1966.

Freud S: The ego and the id. In *Standard Edition of the Complete Psychological Work of Sigmund Freud*, vol 19. Hogarth Press, London, 1966.

*Friedman LT: *Identity's Architect: A Biography of Erik H. Erikson*. Scribner, New York, 1999.

Gill M: *Analysis of Transference*. International Universities Press, New York, 1982.

Muller JP: *Beyond the Psychoanalytic Dyad*. Routledge, New York, 1996.

Rapaport D: A historical survey of psychoanalytic ego psychology. Identity and the life cycle. Psychol Issues Monogr *1:*5, 1959.

Scharf JS, editor: *Foundations of Object Relations Family Therapy*. Aronson, New York, 1989.

Shapiro ER, Carr AW: *Lost in Familiar Places: Creating New Connections between the Individual and Society*. Yale University Press, New Haven, 1991.

Shapiro ER, Shapiro RL, Zinner J, Berkowitz DA: The borderline ego and the working alliance: Indications for individual and family treatment in adolescence. Int J Psychoanal *58:*77, 1977.

Shapiro ER, Zinner J, Shapiro RL, Berkowitz DA: The influence of family experience on borderline personality development. Int Rev Psychoanal *2:*399, 1975.

Shapiro RL: Identity and ego autonomy in adolescence. In *Science and Psychoanalysis*, J Masserman, editor. Grune & Stratton, New York, 1966.

Wallerstein RS, Goldberger L: *Ideas and Identities: The Life and Work of Erik Erikson*. International Universities Press, New York. 1998.

Winnicott DW: The use of an object. Int J Psychoanal *50:*711, 1969.

Winnicott DW: *Playing and Reality*. Basic Books, New York, 1971.

Zinner J, Shapiro RL: Projective identification as a mode of perception and behavior in families of adolescents. Int J Psychoanal *53:*523, 1972.

30.10 OTHER METHODS OF PSYCHOTHERAPY

KENNETH Z. ALTSHULER, M.D.

This section reviews several psychotherapies, some with a particular political slant or dealing with a special population, and some older therapies that continue to have discrete followings. New therapies usually begin with charismatic therapists attracting disciples to their particular theories and teachings. The group so formed refines and promulgates the master's principles and tactics. As the word is spread, a school and often a movement is formed. Proponents of the treatment undertake research and generally show the therapy to be powerful in its scope and effectiveness.

When the founder leaves the scene, the group continues to augment the therapy's principles of practice and extend its range. As generations of adherents follow, what is useful or novel in the original tends to become incorporated into the mainstream. The movement may continue as a splinter group, shrink, or disappear, or have another incarnation and extension as it is rediscovered by a different discipline with a therapeutic aim.

Over the years the parent disciplines of dynamic and behavioral therapy have drawn closer and blurred their boundaries, sharing certain characteristics that form the basis of all psychotherapies. Such common elements include: a healer-patient relationship, with the different roles carrying different expectations and a differential balance of power; a nonjudgmental, supportive acceptance of the patient; and a comforting opportunity for emotional catharsis and the owning up to unpleasant wishes. All also have rituals, that is, things to do together that maintain a focus and keep the parties interested, and a theory that informs the therapist and patient and is a tool with which to reframe experience and enable behavioral change.

The universality and considerable power of these common elements are factors to bear in mind when claims of distinction and unique effectiveness are offered under the banner of any therapy, new or old.

PARADOXICAL AND STRATEGIC THERAPIES

History According to the principle of paradoxical intention fighting the symptoms of performance anxiety increases their severity, but invoking them deliberately may yield the opposite. Milton Erickson elaborated these ideas into paradoxical therapy, and the method was further refined and promulgated widely by Jay Haley.

Erickson was a hypnotherapist whose prowess in the use of paradoxical techniques was reputed to be extraordinary. While working on the double-bind theory of schizophrenia, Haley became interested in the possibility of therapeutic double binds as well as pathological double binds. Haley attended one of Erickson's hypnosis workshops in 1953 and, with John Weakland, began a regular series of visits that lasted for 20 years. Haley and his colleagues observed and learned from Erickson and published what they had learned. Others made similar visits or were attracted by the ideas or the stories they heard, and they added to the reputation and the literature of the growing movement. Haley later developed strategic therapy, essentially a method that applied Ericksonian paradoxical principles, along with other directive techniques, to family therapy.

Interest in paradoxical therapy grew over the years and reached its height in the late 1970s and early 1980s. By this time the method claimed success in treating individuals, couples, and families and for a range of problems, including phobias, compulsions, insomnia, schizophrenia, Tourette's disorder, urinary retention, transvestic fetishism, anorexia nervosa, agoraphobia, childhood fecal incontinence, alcohol use disorders, depressive disorders, fire setting, asthma, constipation, truancy, and a variety of sexual performance and relationship difficulties.

Erickson died in 1980 and a foundation was formed to continue the work he started. He was one of the first to show that brief therapy can be an effective and primary approach. He advocated that therapists should be not just passive listeners but active interventionists guiding the therapy; focusing on symptoms, rather than on dynamics, and on action, rather than on understanding; prescribing as part of the treatment the very resistances and symptoms the patient brought to it; and using indirection and metaphor to achieve the goals of therapy.

Theoretical Issues The overarching idea in paradoxical therapy is to embrace and amplify the notion of paradoxical intention and thereby reduce performance anxiety or destabilize a fixed internal or social system. The paradox and the destabilization may be effective in themselves or may cause secondary reverberations, such as increased self-awareness and the evocation of new behavior or emotions, which lead to helpful change.

Paradoxical therapy is relatively atheoretical about cause and not much concerned with it. Behavioral prescription is emphasized, with the idea that altering internal and social feedback loops can alter the system. Symptoms that block change can be made vehicles for it if they are embraced and prescribed or altered in their timing or the function they serve and if change is restrained. Instructing an impotent patient under no circumstances to try intercourse, for example, may dissipate anxiety and allow him to do so; telling an insomniac to get up and polish the floors for the rest of the night may similarly invite sleep. Paradoxical therapy also recommends defining the problem clearly, establishing clear goals for the treatment, offering a plan, and enlisting the collaboration of the patient or the family in the suggested strategy.

Techniques Some of the principles of paradoxical therapy can be directly translated into techniques: using patients' tendencies toward compliance or defiance—for example, by overtly expecting a relapse and being surprised about improvement; reframing (i.e., relabeling behavior and emotions so that their experience is changed from negative to positive or vice versa); using metaphor, indirect directives, and cryptic paradox; and distancing patients from their symptoms or challenging patients by humor or sarcasm. These techniques and the extent of the therapist activity encouraged in paradoxical therapy are illustrated in the following cases of Erickson and his followers.

Making the Activity Deliberate

An alcoholic wife would secretly hide her bottles and her husband would search for them with increasing anger. The relationship was at a breaking point. Erickson prescribed that she deliberately hide the bottles and that he search for them. If he found one, he would keep it. If it eluded him, she would be allowed to drink in peace. He also suggested that they buy a trailer and go fishing, an activity that they both disliked. The reported results were that the drinking abated and the couple later happily used the trailer to go camping. This conscious game apparently subverted the purpose of the prior secret activity, and forcing the couple to share a disliked recreation brought them into close contact (with alcohol unavailable) and pressured them to communicate, enjoy each other, and make pleasant vacation plans.

Prescribing the Symptom

Instead of telling her husband what she felt, a woman would unconsciously start a fight with him when she wanted affection. She was instructed to do so on purpose and intensely the next time she felt empty or unloved. Doing so, she noted the connection and realized that the attack was so forceful that her partner could not possibly respond in a supportive way. She also felt ashamed, and the experience helped her acknowledge and express her wishes more directly than in the past.

Aversive Conditioning

A woman consulted Erickson about her son who had schizophrenia, who did nothing but sit around the house and moan. Erickson recommended taking him for a drive; she was to drop him off, drive 3 miles farther, then sit and read while she waited for him. After doing that a few times, the son volunteered to walk and later asked if he could go bowling while she read, since he liked bowling more than walking.

A second patient told Erickson that he could drive only on certain streets and that he would become nauseated, vomit, and faint if he tried to drive outside the city limits. Erickson required him to drive to the edge of town in his best clothes, park and lie down in a sandy ditch at the roadside, and lie there for at least 15 minutes. After that, he was to drive one or two car lengths farther and repeat the procedure. He was to do that until he could drive from one telephone pole to the next. He was also told to stop immediately and get into the ditch at the first sign of any symptoms. The patient tried but reported: “The more I did it the madder I got. So I just quit and began to enjoy driving.” Thirteen years later, Erickson reports, the former patient still had no trouble driving. Anger and defiance allegedly brought about a cure.

Flooding

A young male patient of Erickson's avoided a particular restaurant because of his fear of fainting; he also avoided socializing and women. Erickson asked how he would feel about going to dinner at the restaurant; Erickson also described various types of attractive women. The young man said that he would inevitably faint in the restaurant and that he was afraid of women, particularly of attractive divorcèes. Erickson arranged for an attractive divorcee to join them and said that they would go to dinner; he would drive, so as not to worry if the patient should faint. As they parked the car and approached the restaurant, Erickson suggested various places to faint on and discussed the relative merits and dangers of each place. At the restaurant Erickson ignored the young man in conversation and, alerting the management that it was a joke, made an unpleasant scene. The waitress joined in, slamming dishes, and the divorcee added to the discomfort of the scene by urging the patient: “Clean your plate. The fat is good for you!” According to the report, the patient called a friend of his the next night and took him out to the same place without incident, and the therapeutic dinner laid the groundwork for getting him over other fears. Apparently, the flooding experience showed him he could survive the worst, and Erickson's half-mocking concern about his fainting was supportive.

In another example of flooding, a married couple who were both enuretic came to Erickson with great anxiety. He prescribed deliberately wetting the bed together, then sleeping in it; the enuresis disappeared in 2 weeks.

Indirection and Metaphor

A young man feared the open streets but was really afraid of women. Erickson talked to him about his manly biceps and his brains and about improving both. He also suggested that the patient move away from his mother. Erickson labels the technique the use of metaphor and reports that the young man moved, dated, and was soon married.

A more likely use of metaphor is a case in which Erickson, treating a couple with a sexual problem, addressed them in terms of food and preparing dinner. He spoke of how different people have different tastes, how dallying over appetizers can be as pleasant as diving quickly into the main course, how some like soft candles and others like bright light to see by, how a wife can show her appetite, and how a husband can dine in a way that pleases her. Presumably influenced outside the range of consciousness, the couple went home and had a pleasant dinner and successful sex.

In a cryptic use of metaphor, Erickson hypnotized a man who suffered from impotence and talked of large buildings built on a firm structure and fully functional. That patient too apparently overcame his problem.

A man who was hypochondriacally concerned about his heart was seen as heartsick about his job and his wife. He stayed home, never raised his voice, and avoided the exertion of sex because he worried about having a heart attack. The therapist told the wife to take her husband's symptoms seriously and each time he had a pain to offer to call an ambulance, get literature about funeral costs, and inquire about burial plots. The husband became angry. Expressing his anger started a dialogue with his wife about the true nature of his complaints, which allowed his depression to lift and the problems between them to be resolved.

Destabilization

A 10-year-old boy engaged in compulsive, public masturbation; the compulsion had endured for years, despite many treatment efforts. Taking the behavior seriously the therapist asked the boy to keep a list of when he masturbated and to note when he found it most pleasurable. With the idea that he should do it when he found it most pleasurable (on weekends), he was assigned to do it eight times on Sunday but not on Monday. When the boy failed and masturbated on Monday, the disappointed therapist prescribed that he do it 12 times the next Sunday. When the boy was not up to the task, the therapist suggested that, as punishment, the boy masturbate once daily in public and eight times on Sunday. Masturbation was also prescribed as a punishment for thumb sucking and as an alternative to homework. Gradually, the boy stopped and even enjoyed his homework.

Examples reported as successful include telling a woman with headaches that were symptomatic of family relationship problems, that the headaches were probably incurable, and that therapy should aim at helping her live with them; asking the parents of a child with severe and long-standing encopresis whether they could tolerate being normal and being better parents than their mothers (the father then threatened the child with laxatives and the problem disappeared); in a combination of reframing and negative emphasis or challenge, telling a family with a school-phobic child: “Maybe it's better she stay at home. It keeps the family from being normal, and being normal may not be tolerable.” (The therapist reported that the family did not get insulted and leave, but changed spontaneously to prove him wrong.)

Clinical and Ethical Issues Paradoxical therapy is clearly not a psychotherapy itself but a set of techniques usable in therapy. Making explicit the assumptions that underlie the techniques reveals their range of usefulness and their pitfalls. In addition to the idea that prescribing or restraining the symptoms can reduce performance anxiety, the assumptions are that symptoms always reflect a problem in a social context or relationships; that they are somehow and ultimately volitional, even if unconsciously motivated; and that they will therefore respond to being emphasized or manipulated.

Such assumptions are probably acceptable in selected cases in which emphasizing the symptom amounts to flooding, an accepted and well-demonstrated technique of behavioral therapy, and in relationship problems in which prescribing the symptom or altering the context of its appearance may help the participants recognize their inappropriate emotions or behavior patterns.

Using paradoxical therapy in illnesses in which a biological element has been recognized is more of a problem. Expecting a patient with Tourette's disorder to respond to such techniques alone suggests a volitional element greater than what is present. The biologically driven symptoms of schizophrenia are unlikely to respond to paradoxical therapy, and although behaviorally conditioned symptoms may respond, the case histories show the difficulty of distinguishing such boundaries. Most of the literature about paradoxical therapy is anecdotal, and the rare research study is less enthusiastic than authors who suggest that success is the norm.

Most importantly, many of the case histories in the literature include manipulations of the patient or the families that may cover or even invite aggressive acting-out behavior by the therapist, and some case histories suggest a simplistic conception of illness in which punishment is curative. When the operating field permits behavior that is less than open and straightforward, therapists need even more vigilance and self-awareness than when procedural safeguards protect the patient.

CLIENT-CENTERED THERAPY

History Client-centered therapy was created in 1940 by Carl Rogers, a psychologist. Believing that diagnosis, planning, and interpretation manipulate the patient to conform to the therapist's ideas. Rogers tried only to listen, understand, and reflect what he perceived the patient to feel. When Rogers did so, he found that the patient often responded with a deep feeling of personal involvement and the production of new and important material. That passive or nondirective therapy was later

named *client-centered therapy* and was published as such in 1951. Rogers' ideas found wide acceptance, and a movement around the principles of the therapy gathered momentum and grew over the next three decades. Rogers himself continued to develop the method and its applications at the University of Wisconsin, University of Chicago, and Ohio State University and in La Jolla, where he died in 1987.

Client-centered therapy is based on the ideas that each person has the ability to solve psychological problems and that feeling understood and highly valued aid the forces within to do so. Ideas like an instinct toward self-realization and limitless inner powers had enormous appeal to the humanistic idealism that was developing about the time the method was published. A focus on the value of the individual and Rogers' later precept that long and particular training was not necessary for becoming a good therapist were similar rallying points for the antiestablishment, antiauthority emotions of the 1960s and the 1970s.

Rogers played a major role in wresting the practice of psychotherapy from psychiatry alone and distributing it into the hands of psychology and other disciplines. He is also credited with being the first to research his results, the first to record sessions word by word in movies and on videotape, the first to publish transcripts of his failures to encourage the study of what goes wrong, and the first to put forward the thesis that one does not need academic degrees to become a therapist. All but the last idea have become well-accepted elements in the fields.

Client-centered therapy has been reported to be useful for children and adults, problems of neuroses, situational problems, speech difficulties, psychosomatic problems (e.g., allergies), and to some extent psychosis. It has been practiced with individuals and groups, and its principles have been applied in industry, education, child rearing, and intensive encounter groups. Just before his death in 1987 Rogers was training policymakers and leading communication groups in South Africa and workshops in Hungary and the Soviet Union.

Theoretical Issues The overarching presupposition of client-centered psychotherapy is that a basic forward-moving force exists within, a tendency to grow, actualize, and integrate. A number of other principles flow from that idea: all people are believed to have the capacity to understand what factors cause them unhappiness and to overcome those factors; to do so, they must bring what is going on inside into awareness; and to effectuate the circumstances that make awareness possible, the therapist must establish a relationship that is warm and accepting.

Rogers believed that human beings immediately attach meaning to whatever is perceived, so that perception becomes reality. A group of perceptions are organized as the self-concept—people's perceptions of their own characteristics, their relationships with others, and the value they place on those perceptions. Experiences contradicting the preferred perceptions of that organization cause anxiety or maladjustment and must be shut out. In therapy the structure of the self is relaxed so that the patient can drop the barrier and admit, “Yes, I do have those parts of myself that are unacceptable.” When this happens, the previously denied parts are accepted and then integrated, and the result is a better-functioning person, one less tied up in denial and conflict, more effective, and more tolerant of others than in the past.

The aim of the therapy is to create an interpersonal environment in which the self-structure can relax and material can come into consciousness. The therapist sets the three conditions that foster relaxation and the personal growth that follows. The three conditions are *congruence* (or *genuineness*), the ability to be present as a person, to be real in the relationship with the client; *unconditional positive regard*, which conveys to the client, “I accept you as you are, without making judgments or evaluations”; and real *empathic understanding*, a sense of standing in the client's shoes, seeing the world from that vantage point, and then communicating what is understood from that view to the client. When these conditions are met, the client feels released and understood and is able to accept and integrate, view the world differently, and continue the process by bringing out new material.

Client-centered therapy was first practiced in a course of 5 or 6 hours, then 15 or 20 hours, and still later 100 or more hours. Whether the shift was caused by clients' seeking greater levels of change, the enlistment in treatment of more difficult cases, an evolution in technique, or a fall-off in results after the initial enthusiasm is unclear.

Techniques The techniques of client-centered psychotherapy serve the goal of having the therapist convey a sense of the worth and significance of the client, and the idea that the therapist is trying to understand. Far from being laissez-faire, the therapist needs to actively and empathetically help clarify what the patient feels. To be entirely nondirective, the therapist restates and reflects a close approximation of the client's feelings. When this is done correctly, clients feel as if the therapist were almost a part of them, working on their problems as they want to work on them; they feel accepted warmly, without any judgmental overtones. The acceptance and empathy are themselves therapeutic.

Rogers cited an example of an explosive 13-year-old boy who was changed and calmed by sessions in which he and the therapist played tic-tac-toe or in which the therapist allowed silence while the patient sat at the window with his back to the therapist, quietly counting cars. The therapist conveyed respect and belief in the child and the idea that the hour belonged to the child, and that the adult would not be shocked by anything he did. The only limitation was that anger was allowed no physical expression, only symbolic and verbal expression.

The therapy aims not to interpret the client's unconscious motivations or conflicts but to reflect what the client feels. As the counselor and the client progressively clarify their feelings within the warm and accepting relationship, denied feelings come to light and can be dealt with and integrated.

Rogers cited a male patient who had considerable anger. As the therapist accepted the client's feelings and clarified at whom they were aimed, the client came to see which parts were rational and which were not. He was able to say, “If you are like my father or I only see you that way, I hate you as I hated him.” When the unrealistic aspects were expressed openly, they dissolved, leaving the real relationship in place.

The theory also claims that resistance is overcome rather than interpreted. This is accomplished by the consistent acceptance and valuing of the client, the thoroughly undemanding nature of the relationship, and the empathic effort to clarify the client's feelings.

A woman with denied anger at her husband embarrassed him with dizziness and faintness at gatherings that were important to him. The therapist consistently reflected her feelings to her with such questions as, “You seem to feel ...,” “You're confused by...,” and “It seems to you that...” This clarification helped her to move from affirmations of a good relationship with her husband to recollections of old resentments and the recognition of present ill will. Keeping the focus on her also aborted any tendencies for her to project or to accuse the therapist of anything but interest.

Transference is allegedly avoided in a similar fashion. Rogers gave examples of an angry client:

Client:	Am I supposed to tell you everything?
Therapist:	It's up to you; you can talk about anything you want to.
Client:	I had a big fight with you in my mind as I came here. I thought the whole thing was a fraud—I'm being very frank—you can talk like we do down here almost anywhere if you get someone to listen.
Therapist:	It seemed it was kind of a cheat, something you could get anywhere.
Client:	It's like advertisements you see where for a few dollars people will sit and listen to you.
Therapist:	Just listen to your troubles.
Client:	Of course, they don't have your extensive background, but they do the same thing. Then after I talked it out with myself, I realized that the reason I was mad at you is because I feel nervous and jittery coming here (pause) because I'm facing something I don't like to, talking things out myself.
Therapist:	So the feelings about someone listening to you being a fraud you recognize as being connected with your fear and irritation at having to face things in yourself?
Client:	That's right. That's the real logic.

In another case with intense and chaotic sexual and angry feelings, the patient said that she felt the doctor was the most wonderful person she had ever known and that the feeling was more than merely sexual. The therapist reflected back, “You really feel deeply attached to me.” A few sessions later, the dialogue went like this:

Client:	The thing I want is to have sexual intercourse with you. I don't dare ask; I'm afraid you'll be nondirective.

Therapist: You have this awful tension, and want so much to have relations with me.

Client: [after a bit more] Can't we do something about it? Will you relieve the tension... give me a direct answer?

At that point the therapist gently said that the answer was “No,” and the client was relieved. In a subsequent session the client was angry, and the following dialogue took place:

Client: I hate you! I wish you never were born!

Therapist: You just hate me bitterly.

Client: I think I'll throw you in the lake. I'll cut you up! You think people like you, but they don't. I wish you were dead!

Therapist: You detest me, and you'd like to get rid of me.

Client: You think my father did bad things with me, but he didn't! You think he wasn't good, but he was! You think I want intercourse, but I don't.

Therapist: You feel I absolutely misrepresent all your thoughts.

Through this type of work the client later confessed a hallucination and later said: “I had to get rid of this somewhere. I knew you'd understand. It's true I hate myself, but I couldn't say it, so I thought of all the ugly things I could say to you instead.”

Rogers wrote that these interviews showed how feelings of transference were dealt with in the same way as other feelings, and how a consistent reflection of the client's feelings essentially dissolved a transference issue by helping the client to recognize her irrational emotions, after which she could deal with them and what they hid.

Indications The general indications for therapy are when clients feel that their organized self-structures are no longer effective in meeting their needs in the reality situation or when they perceive discrepancies in themselves or see that their behavior is out of control. The description generally covers problems in the neurotic range, but a client-centered approach has been used with 2-year-olds and adults of 65, with people who have mild adjustment problems and disorders diagnosed as psychotic, and with persons of all social classes and levels of intelligence. The theory affirms that a trial of the therapy is unlikely to do harm.

Early in the development of the method, Rogers believed that diagnosis was inadequate and prejudicial. Placing the locus of evaluation in the hands of the therapist rather than the client had undesirable consequences, undermining the client's power of self-determination and raising the specter of misuse for social control. Nosological limitations may have supported this early position, and the social circumstances of the 1960s may have lent it credence, but Rogers' followers believed in it as late as 1987.

Ethical Issues, Research, and Evaluation Rogers' ideas sparked and rode the crest of the humanistic movement in psychology. His emphasis on the value of clients and on the need to accept and respect them in therapeutic relationships were timely and have had a refreshing influence on child rearing, education, and group relations as well as on psychotherapy. His attempts to train nonprofessionals were less well advised and the attempts foundered; the method became so simplified that blank verbal repetitions too easily replaced a true empathic congruence and effort to understand.

Practitioners of client-centered psychotherapy did not believe that diagnosis was particularly useful and did not follow the medical model of illness. That a warm and genuine relationship is sufficient to dissolve resistances that are dynamically motivated and tenaciously held is another tenet open to question. Reflecting what the client feels without interpreting where it comes from or the purposes it may serve is uninformative and could drive the client to action as a further defense. Rogers admitted as much when he noted that the method is frequently unsuccessful with clients who feel incapable of managing themselves and who insist that the therapist take over; a different approach interprets the client's insistence and resentment, rather than just reflecting them. Similarly, reflection in the face of other intense transferences seems to be less effective than a focus on the fantasies that lend force to the transference.

In its contributions to psychotherapy research, client-centered therapy has been outstanding. Rogers and his followers were the first to record and review transcripts of sessions, the first to document the process of treatment, and the first to do follow-up studies of results. They developed such innovations as the Q-sort and scales to objectify the Rorschach test. They showed that in successful therapy negative attitudes about the self are replaced by positive attitudes, that acceptance of others occurs in parallel, that psychological tension decreases, and that plans take the place of old complaints. Although initial claims about success in treating schizophrenia patients had to be dropped, and although Rogers and his colleagues did not do controlled comparisons of their method with others, they were able to document that client-centered therapy is effective. By clearly delineating that a successful outcome is associated with genuine acceptance and an empathic congruence that facilitates the identification and the expression of emotion, they recognized two of the features that empower psychotherapy generically. Perhaps most importantly their emphasis on progressive clarification of the client's feelings sounded an insistent cautionary note about the need for clear understanding instead of a facile interpretation based on the assumptions of one's theory.

TRANSACTIONAL PSYCHOTHERAPY

History Transactional analysis or psychotherapy was the creation of Eric Berne, a psychiatrist with psychoanalytic training who parted with the Freudian movement to focus on social interaction as the unit of analysis. *Transactional Analysis in Psychotherapy* was published in 1961 and *Games People Play* in 1964; *Games People Play* became a record bestseller, with sales of more than 2.5 million copies. Berne's thesis was that in social interactions people seek intimacy, comfort, or recognition, which they get through a series of mutual exchanges or *strokes*. If the ego states of the participants are adult and appropriate, the transaction is harmonious and realistic. Ulterior motives, sponsored by other than an adult ego state, cause discomfort in the transaction, and transactions habitually colored by such motives may become ritualized and predictable.

Written with wit and humor, Berne's books inspired considerable controversy. Some accused them of being facile psychologizing better left to cocktail parties rather than true contributions to psychotherapeutic theory and practice. The fact that the movement has endured, inspiring a journal and an accrediting body and an occasional article in the psychiatric literature, suggests that a number of its ideas have merit.

Theoretical Issues Transactional analysis rests first on the principle that human beings have a hunger for stimulation and contact in the infantile style. When such strivings are blocked, efforts are made to regain the intimate contact. Compromises and sublimations are made in that pursuit so that a hint of the original intimacy may suffice, and the stimulus hunger may be satisfied symbolically. Thus, the original physical strokes of caress and comfort are replaceable by metaphorical stroking by which the person is recognized, flattered, or supported and even by acknowledgments that may appear to bring discomfort. A stroke is a fundamental unit of interaction. An exchange of strokes is a *transaction*, the unit of social intercourse; a stimulus from one person produces a corresponding response in another person. Because social stroking satisfies the original infantile need, it follows that any social intercourse is better than none.

In addition, social interactions fill time, provide structure, and avoid boredom. Interactional programs, all of which contain stroking opportunities, may be *material*, dealing with reality issues; *social*, which constitutes the norms and manners of a social order; or *individual*, by which incidents are sneaked into the other programs in order to get extra, personalized strokes. The sequences of a person's program are developed in line with that person's *life script*, a set of overall goals and patterns, and these sequences are the individual *games*. The games range from harmless and pleasantly satisfying to dangerous and destructive. What develops in a game is assumed to be the purpose of the game and what eventuates as its end result are the strokes that constitute the aim and the satisfaction.

To understand and diagram the interactions in a game, the theory posits three ego states, the parent, the child, and the adult. In the parental ego state, the person reacts either as the parent or as the parent would have directed. In the child state, one may be adaptively compliant or whiny and withdrawing or, as the natural child, rebellious, creative, or impulsive. The adult state appraises reality and acts in tune with it; the adult also moderates between the parent and the child. All three states are present and potentially active at all times.

When the adult in one person addresses the adult in another and is responded to in adult terms, the stimulus and the response are complementary and the transaction is smooth. The same is true if the child addresses the parent and the parent responds appropriately—for example: Child (to parent): “Water!” and parent (to child): “Here, baby.” Such transactions are normal, two-stroke harmonious social interactions.

But if one partner in a couple asks, “Where are my cuff links?” and the other responds, “You're always losing things!” the lines of communication are crossed; the query may be Adult to Adult, but the response is Parent to Child, and the situation gets sticky. A comparable Child-to-Parent cross in response to an Adult-to-Adult

stimulus may occur when one notes to a person driving, “You missed the turnoff,” and the driver responds, “You're always criticizing me!”

Such ulterior transactions are the basis of games, which are ongoing series of complementary ulterior transactions progressing to a predictable defined outcome that satisfies some aspect of the person's life-script strivings. On the surface the interactions may appear plausible, but their recurrent nature, concealed snares, and predictable payoffs define them as parts of a neurotic game.

The test of a game is that a respondent's refusal to play causes increasingly intense efforts to force its continuation; continued refusal may elicit despair or bewilderment. The cure for a game is recognition of its presence, its rules of play, and its motivation. A cure is signified when the initiator can note with a laugh, “There I go again,” and desist volitionally.

Techniques Because games generally involve at least two parties they reveal themselves and are best treated in couples or groups. Transactional analysis has therefore had its most widespread application in couples and group therapies, although it has also been useful in individual treatment. The techniques aim to illustrate the game, dissect the transactions, and frustrate players in their pursuit of strokes. In so doing, the players may learn a considerable amount about their life scripts and how they evolved, although that is not necessarily the goal of treatment.

Games are named for convenience of recognition, with the name usually connoting something about the ends sought. Thus, in “If it weren't for you,” a woman (usually) picks a domineering man to marry, who then forbids her to do things she allegedly wants. She blames him for her unhappiness; they quarrel; their sex life deteriorates; he gives her gifts to make up; and the cycle goes on. In reality, the woman fears the activities she says she longs for, has a sexual inhibition (which the husband may share), and is pleased by the opportunity to prove that men are tyrannical. In the game of “Schlemiel,” a man (usually) messes up or breaks things in order to be castigated and then forgiven.

Other life games include “Debtor,” “Alcoholic,” and “See what you made me do.” The aim of each is to obtain one or several ulterior satisfactions without accepting conscious responsibility. In “Alcoholic” the players may include the alcoholic, a persecutor (usually a spouse), a rescuer (often a health professional), a patsy (who gives money or sympathy), and sometimes a connection (who may tempt the alcoholic to drink). Each player may also be in the game for idiosyncratic reasons, games, or strokes. For the alcoholic person the payoff is the hangover and its chance to indulge in psychological torment and suffering; for the rescuer it is an opportunity to play “I'm only trying to help you”; for the persecutor, it is “Look what you've done to me.” The antithesis, the goal of treatment, is to have each one recognize the role played and to frustrate the primary player (and the other persons addressed as well) by giving up the role assigned.

Groups may be randomly selected or homogeneous. In treatment sessions the theory of transactional analysis is taught assiduously and with blackboard illustrations, the aims of treatment are outlined in operational terms, and contracts are made about individual large goals and are redrawn for small, particular steps in the course of therapy. The therapist's role is active and confrontational but egalitarian.

Research and Ethical Issues Interest in transactional analysis peaked long before the advent of criterion-based diagnosis and controlled clinical trials, and little research on it can stand up to today's critical scrutiny. It was claimed to be effective for persons with neuroses, psychoses, personality disorders, sexual disorders, and mental retardation; for individuals and groups; and as a means of self-help for young children and teenagers. Like most therapies, however, its usefulness rests on clinical lore and a certain common-sense validity. What seems sensible includes an emphasis that is typical of later cognitive-based treatments—namely, identifying elements of behavior that reflect neurotic problems so that their maladaptiveness can be acknowledged, their driving emotions and underlying assumptions recognized, and the pattern of behavior altered. Psychoanalytic therapies share the same aims but include the additional goal of genetic uncovering. More like cognitive therapies also are transactional analysis' emphases on confrontation, contracts, homework, active diagramming of conflict, and teaching about behavior.

The weakness in the system and the potential it has for abuse in work with patients lie in its theoretical assumptions. Neurotic behavior by definition is not realistically aimed and is often maladaptive. To assume that all bad consequences are deliberately sought is incorrect and unfair. To the extent that such a view is held, important motivations may be missed and therapy tainted by a tendency to level accusations. If this position is avoided and complaints are not trivialized by being seen routinely as goals, the dissection of social transactions and the focus on personal responsibility offered by transactional therapy conform to what is generally considered to be of therapeutic value.

GUIDED IMAGERY THERAPY

History The use of guided imagery or guided affective imagery as a form of therapy was systematized by Hanscarl Leuner in 1954. Leuner considered himself an analyst, and he was influenced by Carl Gustav Jung's interest in active imagination and Johannes Heinrich Schultze's autogenic training. Having confronted himself with disturbing emotions and images during his self-analysis, Jung thought that spontaneous, inner, dreamlike images were translations of unconscious emotions and that observing and interpreting them could bring unconscious contents to consciousness. He called the technique *active imagination*, and he made it a part of what clinicians following his analytical tradition would use. Schultze framed the principles of *autogenic training* as a form of self-hypnosis and relaxation training, a mind- and organ-strengthening system based on expansive powers alleged to reside in the mind.

Leuner found that resting comfortably with the eyes closed for some time led to the appearance of images and that an exploration of the images with a guide was a pleasant and beneficial experience. He induced the images actively to shorten the interval before their appearance, added analytic insight to the guided exploration, and so framed *guided affective imagery*.

The ideas of guided affective imagery caught on in the United States and other Western countries, and was reported to be useful in treating in chronic neuroses, personality disorders, depression, phobias, obsessive-compulsive disorders, sexual disorders, hyperhidrosis, substance abuse, anorexia nervosa, ulcerative colitis, and hives. The niche it filled in organized psychiatry, however, was never more than a small one. The ancillary disciplines of medicine, particularly nursing, embraced it with more enthusiasm than did psychiatry, and positive claims have been made regarding its use in postsurgical pain control, in general adaptation to illness, and particularly in coping and dealing with incurable, progressive disease.

Theoretical Issues Guided affective imagery attempts systematically and psychotherapeutically to use the human ability to create images. It holds that so-called *catathymic images*, images appearing before one's closed eyes as if on a screen and ultimately coming together and taking on movement to tell a story, represent a person's interpersonal world in autosymbolic imaginal form. The images include color, plasticity, and action, and the more one is absorbed in them the richer, more colorful, and more subtle they become. The symbols and the stories project the person's conflicts. Psychoanalytically informed guidance as the person recounts the experience can materially enhance conflict resolution by manipulating the symbols or assisting the story toward a successful outcome. The crucial idea is that influencing the symbols and the story influences the conflict.

Techniques The patient is placed in a comfortable chair and taught to practice relaxation, with suggestions of warmth, heaviness, and regular breathing. The patient is then invited to imagine a harmless, pleasant image, like a flower and is asked to describe it in detail, to look in its center, or to touch or smell it. Involvement in the images heightens the relaxation and vice versa; the patient soon becomes able to induce a readiness for imagery quickly. The hypnoidal state is ended with a command, such as, “Clench your fists, bend your arms, open your eyes, and you are back, awake and refreshed.”

When such learning is well established, the therapist relaxes the patient once more and suggests an image to be used in the therapy—for example, a meadow. The patient is told to nod when it comes into view and then to describe the scene.

Specific images used in the therapy are classified as basic, intermediate, or advanced, according to the level of conflict they are likely to involve and the skill and training required by the therapist. Basic images include a meadow, a brook, a mountain, a house, and a wood; intermediate images include a car ride and a rosebush (representing women's and men's sexual feelings toward the opposite sex); and advanced images involve such things as a swamp, a volcano, and a lion.

Descriptions of a meadow or a brook (the patient may see a brook in a meadow spontaneously or be invited to imagine one) reflect current psychological issues. The meadow may be rich and green or scorched to stubble; the brook may be cool and inviting or a raging torrent. The therapist follows and guides, with such questions as, “Can you describe that?” “What do you think of it?” or “What do you see as you follow the brook?”

The therapist's understanding is informed by knowledge and theory and determines whether the therapist interprets to patients or leads them to helpful activity. The therapist may suggest, for example, that a darkening scene reflects the patient's inner mood or that too perfect a picture shows an effort to mask distress. If patients

encounter a fence in the meadow or steep rocks as they follow the brook, the therapist may suggest that the patient climb over or find a way around the obstacle.

A young male patient saw a mountain from which rose another mountain and an observation tower. The young man had problems with authority and would stutter in the presence of his father or his boss. The therapist interpreted the scene to himself to mean that the young man felt watched and that too much was expected of him. The therapist guided him to ascend the tower, which the patient did successfully but amid images of fearful wandering and threats of thunder; the symptoms resolved.

The suggested motifs represent different aspects of the patient's psychology: the house and the edge of the wood—one's own person; the brook—vitality and energy; the mountain and its panorama—the patient's view of the past, present, and future. The therapist looks to see whether supplies and a warm fire are in the house, what comes out of the wood, and whether the brook is cool and refreshing or an icy harbor of dangerous creatures. The management principle is to resolve the conflict symbolically.

One young student got over a block about taking an oral examination by being led in fantasy to feed a dangerous animal (the professor) coming out of a wood; those fearful of their own impulses might do well to approach such animals, pat them, and make friends with them. The guide conducts the exploration, using the symbols as a protective disguise, with empathy and warmth, positive feedback, praise, and at times direct suggestion.

Clinical and Ethical Issues Guided affective imagery combines hypnotherapy with certain psychoanalytic ideas. Its major assumption is that manipulation of the symbols of conflict can resolve the conflict, a tenet that is open to question. The supportive relationship, the sense of involvement and of effectuating something through the process, and the direct suggestions must contribute considerably to whatever effectiveness the technique has.

Interest in guided-image therapy as a formal, psychoanalytically based psychotherapy has waned. Truncated and modified, however, guided imagery has gained a following in pain control, holistic medicine, and the treatment of incurable illness. In pain control it has been reported to be useful as a relaxation and distraction technique, with additional suggestions of finding a second and then several seconds of decreased pain and so extending the period of pain relief. The appeal in holistic medicine is to the limitless powers of the mind, a hopeful approach that has found favor in the nursing of patients with cancer and other terminal illnesses.

FEMINIST PSYCHOTHERAPY

History Feminist psychotherapy is a product of the women's movement. It is an eclectic mix of therapies rather than a particular form of treatment. Its uniting aims, shared with the feminist movement, are to empower women and to raise their self-esteem. The influence of militant feminism in its development and its goal of correcting what are seen as injustices to oppressed women lend a stridency of tone to writings on feminist psychotherapy that may turn some readers away from its important elements.

Theoretical Issues The overarching principles guiding feminist psychotherapy are that the personal is political and that the millennia of sexist attitudes in the socialization of women have harmed them. The aims of treatment are to recognize the harmful effects of the sexist society, to explore inherent contradictions in the roles prescribed for women, to support women in the exploration of their inner resources and their capacities for being nurturing, and to demystify and rectify power relationships.

Its subsidiary ideas are that oppression leads to feelings of impotence and to true impotence, and to low self-esteem and conflicts. Traditional therapy is seen as being a mechanism of social control in the service of the status quo and as being based on an imbalance, even in the therapeutic relationship. Feminist psychotherapy rejects such a model and rejects the idea of conflict developing and residing within the person without regard to the socioeconomic system. It tries therefore to equalize power by modeling, self-disclosure, and disclaiming expertise. It focuses on how the stresses of a patriarchal society cause symptoms—for example, on how depression, a lack of assertiveness, and job difficulties may grow from assigned sex roles and sexual harassment, with inner conflict being only secondary to exterior conditions; on how a lesbian life-style may be a choice to love and commit to a woman and have nothing to do with an unresolved dislike of men; and on how, if sexual intercourse is male aggression, a woman's best defense may be in symptoms such as dyspareunia and vaginismus.

Feminist psychotherapy believes that women of necessity teach their daughters to forgo their own needs and to nurture and please others, as mothers do themselves in childrearing. Conflicts between autonomy and dependence result and are not resolved by therapies that help women endure their suffering, mask or suppress their anger, or rely on antidepressants and sedatives.

Techniques Techniques vary with the type of therapy being adapted for feminist use. Psychodynamic therapy and expressive and supportive therapies are used, and guided imagery has been invoked. Most of the literature, however, places a heavy emphasis on the cognitive approach. Steps include redefining power and reducing guilt; defining one's own feelings and respecting them and so defining one's own personal reality; emphasizing responsibility for one's choices and actions; working through; and terminating therapy when the patient accepts her own needs as legitimate, understands the power differentials that hinder her, can assert herself appropriately, and needs the therapist less than in the past.

Cognitive beliefs common to women (which are also common to men) include the idea that a woman must be loved by everyone, that others' needs count more than her own, that she needs the support of someone strong, and that she has no control over her emotions. Recognizing these beliefs, disputing them, and practicing behavior based on alternative views are part of the treatment; assertiveness training is also commonly recommended.

Feminist psychotherapy believes that women are better therapists for women than are men. A female therapist's empathy and experience are most suited to help a woman become the best expert for herself, and male therapists are considered less likely to treat female patients as equals.

Research and Evaluation Because feminist psychotherapy is more a part of a social movement than a particular therapy, little research has been done on its effectiveness. Many of the ideas it has spawned are provocative and merit consideration. Feminists call attention to the facts that women have suffered from unequal power relationships; that male dominance in a family is taken for granted, but a special note in the case history is likely if the wife has higher status than the husband; that the description “absent father” is a less extreme judgment of the person than is the description “maternal deprivation”; and that the term “dominant mother” carries a similar pejorative quality. Other ideas, although more extreme, are piquant enough to stimulate a close look—for example, that the family is a system that perpetuates male dominance and that at the junior high school level boys do not so much catch up in school performance as that girls fall behind because their type of understanding—feelingful and intuitive—is less valued than male understanding in a structure that is focused on adapting to male-dominated ideas. The rhetoric that societal oppression is the cause of all the ills women suffer, however, ignores the possibility that women may have personal, internal conflicts. By encouraging self-justification and rationalization, such a position may serve the health of women poorly.

PSYCHOTHERAPY FOR THE DEAF

History Psychotherapy for the deaf is not a particular type of psychotherapy but a group of therapies that are adapted to the needs of a special population—people who develop a profound hearing impairment before speech is established. The impairment is defined operationally as one that renders audition useless as a means of intercourse with the hearing world, and it generally involves a hearing loss of 80 decibels or more across the speech frequencies, even with the aid of amplification. Its most obvious consequence is to interfere with the ability to learn language by simple auditory imitation and at the normal time and pace in development.

Psychiatric interest in the deaf began with the work of Franz Kallmann, John Rainer, and Kenneth Altshuler in the mid-1950s. Then the country's most prominent geneticist in psychiatry, Kallmann was approached by a branch of what is now the Department of Health and Human Services for assistance in obtaining case closure among psychiatrically ill deaf clients. Kallman recruited Rainer and Altshuler and began a series of normative studies and studies of deaf persons with mental illness. The first effort established an outpatient psychiatric clinic, and subsequent projects developed an inpatient unit for statewide services, a vocational rehabilitation liason and halfway house program, and consultation programs at a school for the deaf. Centers focused on psychiatry and the deaf were subsequently established in Chicago, San Francisco, Washington, and several other cities and countries around the world. At present, a small caucus of psychiatrists interested in working with the profoundly deaf meet regularly at annual meetings of the American Psychiatric Association.

Theoretical Issues To adapt a given psychotherapy to the needs of persons with profound early deafness demands that the therapist understand what the handicap means to one's life experience and to take into account the limitations this experience places on the therapeutic exchange.

A first guiding principle is that the simple absence of audition has no relevance to innate intelligence potentials; intelligence quotients (I.Q.s) among the deaf follow the same bell-shaped curve as in the hearing. Among the deaf, performance scores are generally higher than verbal scores because of the lag in language development. Deaf children of deaf parents develop vocabulary at the same rate as their hearing peers, albeit in sign language, because their parents are able to communicate with them freely. Occasionally, an enlightened hearing parent learns sign language to provide the only free, two-way accessibility that the handicap allows.

Too often, however, parents are caught up in an educational controversy that has gone on since Alexander Graham Bell and the 1880s about whether signing discourages the learning of speech. Evidence suggests that an educational approach that embraces all avenues of communication—combining oralism, speech and lipreading instruction, amplification, and sign language—gives the best results. Moreover, the combination supports the exercise of age-appropriate interests and language skills in accord with their natural developmental timetable. In the absence of a widespread combined approach, studies in the 1970s showed that the average deaf adolescent left high school with about fourth- and fifth-grade English and mathematical skills.

As a result of these findings and the urging of mental health professionals, since the 1970s there has been a major shift in educational practices from a near-exclusive emphasis on oralism toward some variety of the combined mode of instruction. Pockets of resistance exist, and the recent availability of cochlear implants at a relatively early age may recast expectations and heat up the issue again.

A factor contributing to the endurance of the controversy is that parents are understandably hurt and saddened when the handicap is recognized and they wish that it would go away. Motivated partly by denial and partly by the commendable desire that their child's life be like any other child's life, parents resist sign language as the visible symbol of the handicap and the harbinger of a separate and sequestered existence. What is lost in the face of these pressures is the recognition that language, generally mediated through audition, brings a tremendous amount of information and experience. Without hearing or the effective alternative of sign language, uncountable hours of speech and lipreading instruction can only provide language late and often incompletely.

Sound also brings nuance and music, and it is the sole distance receptor that scans 360 degrees of the environment at all times, perhaps even in sleep. What may be associated with missing those elements of experience is uncertain.

A final concomitant of deafness, at least in the 90 percent of those born to hearing parents, is that the parents of deaf children feel confused, uncertain, and often guilty. Their own ambivalence, the conflicting advice they may receive, and the stranger to communication in their midst often makes them overprotective, inconsistent in discipline, and erratic in their hopes and approach.

As a result of all these factors but probably influenced primarily by the language lag and the lack of general experience, deaf persons have been described as more naive, impulsive, concrete, and immature than their hearing counterparts. Such descriptions apply only as a composite, and the deaf should not be stereotyped. In a large enough number, more deaf persons may show more of those traits than their hearing peers, but the presence and the extent of the traits in a particular person are not predictable.

Techniques Communication is the sine qua non of psychotherapy. The first requirement with the deaf is to understand what life experiences the handicap entails; the second requirement is to interact freely with patients so as to apprehend their personal and particular experiences and to exchange understanding about the details. By adulthood, virtually every person with early profound deafness is comfortable and fluent in sign language (barring an accompanying organic problem that interferes with language acquisition, regardless of mode). Having the therapist equally versed in manual language would be ideal, but that is rarely possible, despite the fact that a growing number of psychiatrists and others have learned to sign. The problem is lessened to the extent that the patient's language, speech, and lipreading skills are developed. Often these skills are insufficient to allow the conduct of meaningful psychotherapy, even with exchanges in writing. In such instances an interpreter is necessary. Deaf persons often rely on interpreters, and the therapist rather than the patient may be the one who is uncomfortable with the third party's presence. One must be alert, however, to the interpreter's tendency to interpret when the therapist wants an exact translation; for example, when the patient's thoughts are to be evaluated, the interpreter's efforts to make sense of them may be counterproductive. Supportive psychotherapy can be conducted with the aid of an interpreter, but substantial insightful work probably demands a subtlety of understanding only available with manual skill on the part of the therapist or adept speech and lipreading skills on the part of the patient.

Clinical Issues Less than profound deafness and deafness after speech is well established are problems of a different nature and require different adaptations. To the extent that language has been developed before the loss of hearing and to the extent that the loss is remediable by amplification, the patient's life experiences are less affected, and therapy requires less in the way of modification. However, the loss itself, its cause, and the patient's changed life circumstances because of it are factors that affect the person's problems and the therapy.

A couple of additional points must be kept in mind in counseling the hearing parents of deaf children. Eager for their child to have a normal life, the parents may be impatient with modest results in attempts to remediate the handicap and its effects. Complaints of bad treatment at the hands of many institutions or agencies must be evaluated with that possibility in mind. Also, denial on the part of a family may be brought to an abrupt end with the child's impending adolescence. When parallel play and participation in simple games give way and the horizons of social interaction loom large, the effects of the handicap become unavoidable, and parents may become depressed.

Research The educational, developmental, and psychiatric literature on early profound deafness is considerable, but no controlled studies of psychotherapy have been reported. The psychiatric literature, mostly from the 1970s and before, suggests that major mental illness is no more prevalent among the deaf than in the hearing, that impulsive or aggressive behavior may be more common in the presentation of illness among the deaf, and that guilt and psychomotor retardation may be less frequent in cases of depressive disorders in the deaf. Anecdotal reports also indicate that various forms of psychotherapy are as useful in the deaf as in the hearing when modifications in view of the handicap are made.

COMPUTERS AND PSYCHOTHERAPY

History and Techniques Psychotherapy by means of a computer, psychotherapy by nonprofessionals, and brief psychotherapy under managed care are not particular schools of treatment. What they share are the aims of cutting costs, thereby making therapy more widely available. They lower costs by abbreviating treatments or by substituting less expensive providers for more expensive ones.

Computer programs have been written in the last few years to provide educational materials to patients or to interact with them in a more truly therapeutic way, for example, aiming at enhanced self-awareness or self-control. Psychotherapy by nonprofessionals utilizes common sense and supportive approaches that derive from general experience. Brief therapy under managed care rests on the thesis that dynamic understanding can be utilized episodically, and that widely spaced sessions providing insight and support can help most patients to overcome their acute problems in living.

Theoretical Issues Each approach raises challenging issues of theory. Is the comfort and the flexibility of a real person requisite for effective psychotherapy, or can a computer be individually responsive enough, and can patients develop transferences and a sense of safety relating to it? Is professional training necessary to obtain satisfactory results, or can well-adjusted and mature adults do just as well? Are regular meetings, weekly or more often, cogent therapeutic agents or can monthly meetings be equally efficacious?

Clinical and Ethical Issues As with all therapies, these novel approaches have been reported to be successful and have gained a following. They pose considerable challenge to traditional therapies, calling into question many of the basic and revered assumptions on which the longer-established methods rest. Since the treatments are new and research on any of them is in its infancy, it is likely that early claims will moderate over time as research establishes their limits. Meanwhile, the fact that they are experimental and unproven should be kept in mind and clearly communicated to patients. There is little danger in the computer-driven and the nonprofessional efforts because their innovative nature makes informed consent and scrutiny by institutional review boards mandatory before treatment under such conditions is initiated. The current emphasis on cost containment, however, does not afford the same protection in the case of the minimalist approaches being instituted by enthusiastic managers of care. In such cases, patients may need better information than is now offered if they are to make an informed decision.

TRANSPERSONAL PSYCHOTHERAPY

History Transpersonal psychology grew out of the humanistic movement of the 1960s. Defining itself as “the personal, plus,” it means to include all facets of personal

psychology and psychiatry *plus* deeper or higher transcendent experiences. It thus embraces a range of religious and spiritual experiences and parapsychological phenomena such as conversion; crises of faith; ecstasy, peak and near-death experiences; spiritual mystical experiences; and past life experiences. While the term *transpersonal* was first used by William James and later expanded on by Jung, it was such writers as Abraham Maslow, Robert Assagioli, Ken Wilber, and Stanislav Grof who generated its theoretical underpinnings. Each of these authors developed theories of needs and motivations or of stages of development that encompassed and were roughly modeled on traditional developmental theory, but added the extra dimensions required by the transpersonal emphasis. The *Journal of Transpersonal Psychology* started publication in 1969 and continues to exist. The Menninger Clinic sponsored the first Council Grove Conference at that time, focused on such topics as meditation, biofeedback, yoga, and psychedelic agents. Since then, several transpersonally oriented institutions were founded, mostly in California, and for the past several years about fifteen dissertations dealing with transpersonal experiences are written yearly. The fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) includes religious and spiritual problems under other conditions that may be a focus of clinical attention.

Theoretical Issues In the free-for-all of theory building, several authors have contributed to transpersonal psychology and psychiatry but no single contribution dominates. Thus, the theory is an overlapping pastiche rather than a coherent whole: Abraham Maslow studied a series of self-actualized persons and found them similar to traditional descriptions of enlightened persons. He concluded that human beings have a biologically based instinctive nature that is fulfilled in spiritual self-actualization, and he and Stanislov Grof coined the term and founded the field of *transpersonal psychology*. Roberto Assagioli suggested that each person has many subpersonalities, and that a spiritual psychosynthesis is needed to unleash the transforming and creative energies of a *super conscious*, which also exists. Ken Wilber proposes prepersonal, personal, and transpersonal developmental stages; disorders of the prepersonal developmental level result in infantile psychoses or narcissistic or borderline personality disorders; disturbances at the personal level give rise to neuroses; and spiritual emergencies, Kundalani crises, and dark nights of the soul rest on disturbances at the transpersonal level. Stanislov Grof, during 17 years as scholar in residence at the Esalen Institute in California, experimented widely with psychedelics in psychotherapy and proposed that in psychedelic and other altered states of consciousness, patients reexperience the trauma of their birth. He suggested the existence of four basic perinatal matrices (BPMs) representing experiences from the entry into the intrauterine state (BPM I), through entry to the birth experience (BPM II), passage toward a light with intense anxiety and near-death experiences (BPM III), and culminating in total surrender, a sense of overwhelming and unending love, light, expansion, and forgiveness (BPM IV). After experiencing or reexperiencing BPM IV, in adulthood further transpersonal expansions may occur in two domains: expansion of ego boundaries in ordinary real terms and extensions beyond the usual framework. The first includes a sense of identification with the earth and all of life in the universe and the second includes spiritual communication; out-of-body experiences; and telepathic, precognitive, and other paranormal experiences. These experiences were often seen as psychotic by traditional psychologists, and therefore constitute spiritual emergencies that require special help. Grof urged mental health professionals to learn to distinguish these nonordinary, transpersonal states from mental illness.

Techniques Technique in transpersonal psychotherapy is an amalgam of the generally accepted and the esoteric, depending on the world view of the practitioner and the issues being addressed. The range is from the psychodynamic and behavioral through meditative approaches or the recommendation of spiritual texts, to rapid-breathing techniques, guided imagery, and the use of psychedelic drugs to promote altered states of consciousness. The teachings are variously applied to a similarly wide range of issues.

Defensive spirituality, in which religious or spiritual beliefs are used to deny feelings or conflicts, or *offensive spirituality*, in which there is an obnoxious, self-reassuring but intrusive stance of “having the one true path” may be amenable to dynamic techniques. Similarly, loss of faith or a conversion experience may be dealt with by a sensitive therapist to determine what factors are at work in the patient's life and whether the conversion, for example, is a legitimate search for truth and the comfort of faith or a magical denial doomed to fail. Expressive explorations of hallucinations or delusions or of a psychotic religious experience after the psychosis is controlled are recommended by some as a means of finding personal insights.

Transpersonal psychotherapists also try to distinguish genuine religious experiences from psychoses, the latter being often more terrifying, accompanied by deterioration in other areas of life, and containing delusions of being addressed by specific religious figures. The boundaries, however, are hard to define in *spiritual emergencies*, when the individual feels dazzled by truths too powerful to grasp, or in *Kundalani awakenings*, states in which energies expand upward from the base of the spine and explosively alter consciousness. Supportive therapy is recommended in these cases, along with education about the experience. Meditation and the use of such mantras as “May I be happy” “May I be peaceful” are also reported useful if the patient is having anxiety over a crisis of faith. Near-death experiences may also lead to a transformed view of the world, and support is recommended along with help in “those parts of the ego that may have died.” The traditional psychotherapist and psychiatrist may have difficulty in areas of transpersonal psychiatry that aim to help individuals recognize and realize a presumed “ultimate state that transcends our usual state of functioning, a state of mystical illumination or cosmic unity.” Some of these include experiences such as channeling, encounters with unidentified flying objects (UFOs), possession states, and past-life experiences. In these areas the transpersonal literature is ambiguous, on the one hand cautiously attesting that such experiences are meaningful to the person and embody content relevant to the individual's personality, and on the other accepting them as normative spiritual experiences, an ultimate state of mystic illumination or cosmic unity reflective of a reality beyond the norm.

Research and Evaluation Research on the paranormal while always of interest because of its supernatural allure, has failed to confirm the legitimacy of past-life experiences, precognition, psychokinesis, and similar phenomena. Also, because transpersonal psychotherapy is not a single discipline there have been no systematic studies of its efficacy. The field emphasizes the personally meaningful and possibly transformative effects of such things as conversions or near-death experiences; the importance of religion and mysticism in the lives of many; the difficulty of teasing apart the boundaries between psychiatric illness, personal mystique, and cultural differences in expected experience; and the need for sensitivity in working with patients whose life and experiences are different from one's own.

SUGGESTED CROSS-REFERENCES

Psychoanalysis is discussed in Section 6.1 and Section 30.1, behavior therapy in Section 30.2, hypnosis in Section 30.3, group therapy in Section 30.4, cognitive therapy in Section 30.6, couples therapy in Section 30.5, and brief psychotherapy in Section 30.8. Learning theory is discussed in Section 3.3.

SECTION REFERENCES

*Altshuler KZ: Studies of the deaf: Relevance to psychiatric theory. Am J Psychiatry *127:*1521, 1971.

Altshuler KZ: Will the psychotherapies yield differential results: A look at assumptions in therapy trials. Am J Psychother *43:*310, 1989.

Beck AT, Sokol L, Clark DA, Berchick R, Wright F: A crossover study of focused cognitive therapy for panic disorder. Am J Psychiatry *149:*778, 1992.

Beitman BD: *The Psychiatrist's Guide to Cost Containment: How to Survive and Thrive in an Age of Managed Care* . Sage Publications, Thousand Oaks, CA, 1998.

Berne E: *Games People Play: The Psychology of Human Relationships*. Castle, New York, 1964.

Berne E: *Transactional Analysis in Psychotherapy*. Grove, New York, 1961.

Boorstein S: *Clinical Studies in Transpersonal Psychotherapy*. State University of New York Press, Albany, 1997.

Bozarth JD: Empathy from the framework of client-centered therapy and the Rogerian hypothesis. In *Empathy Reconsidered: New Directions in Psychotherapy*, AC Bohart, LS Greenberg, editors. American Psychological Association, Washington, DC, 1997.

*Butler M: Guidelines for feminist therapy. In *Handbook of Feminist Therapy*, LB Rosewater, LEA Walker, editors. Springer, New York, 1985.

*Christensen A, Jacobson NS: Who (or what) can do psychotherapy: The status and challenge of nonprofessional therapies. Psychol Sci *5:*8, 1994.

Dolezal-Wood S, Belar CD, Snibbe J: A comparison of computer-assisted psychotherapy and cognitive-behavioral therapy in groups. J Clin Psychol Med Settings *5:*103, 1998.

Dutton-Douglas MA, Walker LE: *Feminist Psychotherapies: Integration of Therapeutic and Feminist Systems*. Ablex, Norwood, NJ, 1988.

Esplen MJ, Garfinkel PE: Guided imagery treatment to promote self-soothing in bulimia nervosa: A theoretical rationale. J Psychother Pract Res *7:*102, 1998.

Fiese BH: Family context in pediatric psychology from a transactional perspective: Family rituals and stories as examples. J Pediatr Psychol *22*(2):183, 1997.

*Gerson SN: When should managed care firms terminate private benefits for chronically mentally ill patients? Behav Healthcare Tomorrow *3:*31, 1994.

Greyson B: The near-death experience as a transpersonal crisis. In *Textbook of Transpersonal Psychiatry and Psychology*, JR Battista, AB Chinen, BW Scotton, editors. Basic Books, New York, 1996.

*Haley J: *The Uncommon Therapy*, ed 2. Norton, New York, 1986.

Jaffee DT, Bressler D: Guided imagery: Healing through the mind's eye. In *Imagery: Its Many Dimensions and Applications*, JE Shorr, GE Sobel, P Robin, JA Connella, editors. Plenum, New York, 1981.

Jones L, River EM: Current uses of imagery in cognitive and behavioral therapies. In *Innovations in Clinical Practice: A Source Book*, L VandeCreek, S Knapp, TL Jackson, editors. Professional Resource Press, Sarasota, FL, 1997.

Kadis LB, McClendon R: *Concise Guide to Marital and Family Therapy*. American Psychiatric Press, Washington, DC, 1998.

Leuner H: *Guided Affective Imagery: Mental Imagery in Short-Term Psychotherapy*, WA Richards, editor. Thieme-Stratton, New York, 1984.

Lytle LR, Lewis JW: *Culturally Affirmative Psychotherapy with Deaf Persons*, NS Glickman, MA Harvey, editors. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

Meadow KP: *Deafness and Child Development*. University of California Press, Berkeley, 1980.

*Meredith LS, Rubenstein LV, Rost K, Ford DE, Gordon N, Nutting P, Camp P, Wells KB: Treating depression in staff-model versus network-model managed care organizations. J Gen Intern Med *14:*39, 1999.

Rainer JD, Altshuler KZ: *Comprehensive Mental Health Services for the Deaf*. New York State Psychiatric Institute, New York, 1966.

Rainer JD, Altshuler KZ, Kallman FJ, editors: *Family and Mental Health Problems in a Deaf Population*. Springer, New York, 1963.

*Rogers CR: *Client Centered Psychotherapy*. Houghton Mifflin, Boston, 1951.

*Rogers WH, Wells KB, Meredith LS, Sturm R, Burnam MA: Outcomes for adult outpatients with depression under prepaid or fee-for-service financing. Arch Gen Psychiatry *50:*517, 1993.

Rohrbaugh M, Tennen H, Eron JB: Paradoxical interventions. Curr Psychiatr Ther *21:*67, 1982.

Schultz JH, Luthe W: *Autogenic Training: A Psychophysiologistic Approach in Psychotherapy*. Grune & Stratton, New York, 1959.

*Seligman MEP, Levant RF: Managed care policies rely on inadequate science. Prof Psychol Res Pr *29:*211, 1998.

Selmi PM, Klein MH, Greist JH, Sorrell SP, Erdman HP: Computer-administered cognitive-behavioral therapy for depression. Am J Psychiatry *147:*51, 1990.

Teasell RW, Shapiro AP: Strategic-behavioral intervention in the treatment of chronic nonorganic motor disorders. Am J Phys Med Rehab *73:*44, 1994.

Troesch LM, Rodehaver CB, Delaney EA, Yanes B: The influence of guided imagery on chemotherapy-related nausea and vomiting. Oncol Nurs Forum *20:*1179, 1993.

Weeks GR, L'Abate L: *Paradoxical Psychotherapy: Theory and Practice with Individuals, Couples, and Families*. Brunner/Mazel, New York, 1984.

Wright CI, Fish LS: Feminist family therapy: The battle against subtle sexism. In *Subtle Sexism: Current Practice and Prospects For Change*, NV Benokraitis, editor. Sage Publications, Thousand Oaks, CA, 1997.

30.11 EVALUATION OF PSYCHOTHERAPY

KENNETH I. HOWARD, PH.D., RONALD F. KRASNER, M.D. AND STEPHEN M. SAUNDERS, PH.D.

Psychotherapy is the most prevalent outpatient treatment for psychiatric disorders. Unlike other medical interventions, however, psychotherapy entails a particular kind of conversation between the therapist and the patient, and is not generally dependent on tangible technical interventions such as setting a bone or suturing a wound. Given the intangible nature of psychotherapy, it has been difficult to establish its scientific validity. To further complicate matters, since the introduction of psychotherapy a plethora of competing schools have emerged. Unlike many other areas of medicine, however, new psychotherapeutic techniques and ideologies were adopted before research evidence had been produced to support their superiority over current practices. Thus, the acceptance of a particular psychotherapeutic approach was dependent on the influence and charisma of the inventor of that approach.

Given this ambiguity and the need for scientific evidence, the new field of psychotherapy research emerged.

HISTORY

Psychotherapy research is focused on the empirical investigation of the processes and outcomes of psychotherapy. It aims to increase our knowledge regarding the nature of therapeutic interventions, the patients who will most benefit from those interventions, and the outcomes expected from those interventions. It is now well established that psychotherapy achieves significant positive effects over and above control conditions. However, in order to determine how such effects are achieved, how to investigate specific therapies for specific disorders, and the variables that account for positive outcome, new research paradigms and enhanced collaboration between clinicians and researchers are necessary. The overarching goals of psychotherapy research are to improve the practice of psychotherapy, inform public policies regarding psychotherapy, and streamline the provision of mental health care.

The history of psychotherapy research can be understood by considering a sequence of developmental phases, each with its own central issues and achievements. The earliest scientific investigations of psychotherapy began in the 1920s as the first clinical researchers began to document their treatment results. Studies of nonpsychoanalytic treatments appeared in the 1930s, but there was little effort through the 1940s to study either psychoanalytic or nonpsychoanalytic treatments. However, competitors to the psychoanalytic paradigm made their appearance, and researchers such as Carl Rogers brought psychotherapy from private offices to be scientifically studied.

When Hans Eysenck's landmark 1952 review resulted in the claim that 67 percent of outpatients “spontaneously” improved in 2 years without treatment, psychotherapy researchers became even more motivated to search for scientific evidence regarding the efficacy of psychotherapy. Objective methods for measuring recorded events during therapy and controlled comparative outcome studies were developed utilizing Fisherian statistical methodology (random assignment to treatment conditions, null hypothesis testing with *t*-tests, analysis of variance, and correlations). The 1960s to the early 1980s saw the significant expansion and organization of psychotherapy research efforts. New methods were employed, most significantly the results of meta-analysis (an assessment of treatment effectiveness through averaging and combining results across studies). The first of these meta-analyses, presented by Mary L. Smith, Gene Glass, and Thomas Miller in 1980, showed a mean effect size for psychotherapy of 0.85, indicating that psychotherapy was very effective indeed. Finally, from 1984 to the present there has been a consolidation and reformulation of psychotherapy research that has begun to accept eclecticism and the relevance of models, stages, and averaged growth curves that in turn yield assessments of patients' progress leading to beneficial outcome.

Psychotherapy research remains bedeviled by the diversity of the variables investigated, the varying methods of appraisal, the heterogeneity of the patients studied, the differences in therapist training and skill, and the variations in clinical settings. However, there is now a substantial body of evidence that shows that: (1) there are effective psychological treatments for a large number of psychiatric disorders; (2) psychotherapeutic approaches either alone or in combination with psychotropic medications are more effective than placebo; and (3) psychotherapeutic treatments may be at least as effective as medications and may enhance the effects of medication.

The question of whether psychotherapy works has become as useless as the question of whether surgery or antibiotics work. The main goal is to match the appropriate intervention applied by the appropriately trained practitioner to the appropriate pathological condition. Psychotherapy research can provide some guidance in this regard and can forge a link with providers of psychotherapy by furnishing information relevant to the current case in treatment.

CONSUMERS OF PSYCHOTHERAPY

A substantial proportion of the population (about 25 percent) meet the criteria for a psychiatric disorder in any given year, but the vast majority of those who do so (over 80 percent) do not get help from a mental health specialist. When evaluating the effectiveness of psychotherapy, an important issue is who receives such help, why they seek it, and how they get it.

Utilization of Psychotherapy Examining the characteristics of psychotherapy users through the use of a single national survey is difficult since only 3 to 5 percent of the general population will visit a mental health practitioner in a given year. However, based on the combined information from several large-scale, national surveys conducted in the 1980s, it has been shown that two thirds of those who make at least one mental health visit are female, and that 90 percent are white. The most educated are more likely to make a visit; about 50 percent of psychotherapy patients have had at least some college education. Age is also related to the probability of making a mental health visit: the youngest and oldest are the least likely to make an initial visit and over 50 percent of patients are between 21 to 40 years of age. Surveys indicate that income is not related to the likelihood of seeking mental health care. Having a diagnosable mental illness significantly increases the likelihood that a person will seek mental health care, and having multiple diagnoses increases the likelihood further. Patients with diagnoses such as schizophrenia, somatization, panic disorder, and major depressive disorder are more likely to seek professional help than are those with diagnoses such as obsessive-compulsive disorder, substance use disorders, dysthymic disorder and phobias. However, almost half of those seeking such care do not meet the criteria for a psychiatric diagnosis, and research indicates that the best predictor of seeking mental health care is level of distress, whether from a psychiatric illness, an interpersonal problem, or inadequate coping in a particular situation. A study in the 1990s of 500 persons seeking psychological treatment found that the most common patient complaints were interpersonal problems, depression, uncontrolled behavior, and anxiety.

Help-Seeking Behavior Given the frequent finding that most persons who need psychiatric help do not get it, it is important to establish how persons go about seeking psychotherapy. Models of help-seeking behavior focus on the series of decisions that must be made, such as recognizing that a problem exists, deciding that seeking psychotherapy is appropriate, and contacting a professional helper. Research indicates that problem recognition is the most difficult and time-consuming step to achieve, and that some individuals have a significantly easier time accepting the need for psychotherapy than do others. However, the help-seeking process is complicated. For example, it has been found that adolescents experiencing suicidal ideation realized that they needed help, but were less likely than nonideating peers to obtain it. Most persons who seek professional care will first seek help from their family, friends, acquaintances, and others outside the mental health profession (such as primary care physicians and clergy). Others will look to nonprofessional sources, such as self-help groups (e.g., Alcoholics Anonymous). This help may reduce distress to the extent that professional care becomes unnecessary. The social network might either promote or discourage the individual from seeking professional mental health care; friends and family may be able to identify psychiatric problems and provide information about and encourage the use of such care, but they may also transmit attitudes that make formal help-seeking less likely.

Utilization Because research has established that the median effective dose of therapy is between 6 and 8 sessions, an important issue is whether a person who seeks therapy actually engages in treatment (defined here as at least 8 visits). This is related to the issue of equity and cost, as it has been shown that 44 percent of

patients make less than 4 visits and account for 6.7 percent of outpatient psychotherapy expenditures, whereas 16.2 percent of patients make more than 24 visits and account for 57.4 percent of expenditures. Many possible factors have been investigated, including patient income, level of education, age, sex, race, and socioeconomic variables. Controlling for their increased likelihood of making an initial visit, females are not more likely than males to continue in therapy once they have made that initial visit. The most educated are more likely to make an initial visit as well as to enter therapy given such a visit; nonwhites are significantly less likely either to make an initial visit or to continue treatment after the initial visit. The youngest are the most likely to enter therapy given a visit, whereas those 61 and older are the least likely to enter therapy after the first visit. Income is not related to the likelihood of making at least one visit for mental health care but is positively related to the likelihood of engaging in therapy. Thus, research has shown that there is a frequent but not invariable relation between specific patient variables and length of treatment. In contrast, preparing patients for psychotherapy (e.g., via role induction interview) has not made a discernible difference in treatment engagement. Similarly, attempts to predict continuation in psychotherapy using psychological tests, measures of patient expectations, presenting problems, and social support have not been successful.

PATIENT, THERAPIST, AND TREATMENT VARIABLES RELATED TO OUTCOME

Conceptions of Outcome Reviews of studies of psychotherapy have identified at least 800 different outcome measures, the most frequent of which, the Beck Depression Inventory, has been used in only 8 percent of these studies. Obviously, many outcome measures have only been used in a single study. The plethora of outcome measures illustrates three outcome issues: (1) selection of outcome measures varies according to the beholder; (2) outcome measures are of mixed psychometric quality; and (3) outcomes measures can be grouped into categories by their similarities. Each of these points will be considered in turn.

Stakeholders The enterprise of psychotherapy involves a variety of interested parties or stakeholders. The patient is most concerned about relief from personal distress. The therapist attends most closely to symptom amelioration and to correction or removal of the pathogen or causal condition thought to underlie the patient's symptomatic condition. The client—who often is the patient, but could also be a spouse, employer, or parent—usually is most concerned with better functioning. The purchaser attends most closely to the cost efficiency of methods and outcomes. Outcome has different meanings for each of these parties.

Psychometric Properties Psychometric investigations of outcome measures indicate that many measures tap a common domain of emotion and behavior. For example, studies of self-reported depression and anxiety inventories indicate that most measure a common domain of negative affect. Nevertheless, outcomes researchers continue to use inventories and measures as if they measure the variables indicated by their label; reliability and validity studies are clearly needed in this area.

Outcome Measures Most of the outcome assessments used in controlled studies of psychotherapy have used one or more symptom measures. This focus on symptoms as the category of choice in outcomes represents the focus of the clinician and contributes to the efforts of psychotherapy advocates to achieve parity with clinical trials of medications for similar disorders. The second most frequently studied domain of outcomes, although it can be found only in a minority of published studies, is functioning. Social functioning can include leisure, intimacy, and work functioning; each of these subdomains has been studied separately. The domain of well-being (or freedom from subjective distress) has received the least attention, although it is clearly of paramount importance to the patient.

Patient Variables A number of patient variables, such as education level and race, are related to the probability both of seeking and obtaining psychotherapy. Of even more interest to clinicians and researchers is predicting who will benefit from treatment. The relationship between the ultimate effectiveness of therapy and pretherapy patient characteristics, such as age, sex, social class, intelligence, personality, and diagnosis, has been extensively researched but few consistent relationships have been uncovered. This is partly because of the complexity and related difficulty of construct measurement. For example, patient expectations about treatment have been measured in naturalistic settings and experimentally manipulated in the laboratory, and some studies suggested that patients who expect so will ultimately do better. Other research has not found this to be the case, however, and the validity of the measures and the manipulation of expectancy have been questioned. At first glance, it would seem that patient variables would be relatively easy to assess, but in fact most research in this area has had similar problems (e.g., only recently has the diagnosis of personality disorder achieved even marginally acceptable reliability).

Patient Characteristics Despite numerous studies, research has not shown a consistent relationship between treatment outcome and patient age, gender, or sociodemographic variables. A meta-analysis of over 500 studies found no correlation between age and psychotherapy outcome, and patient gender has also been shown to be unrelated to treatment effectiveness. As noted, other sociodemographic variables, such as education and race, are related to the probability both of seeking and obtaining psychotherapy, but there is no clear relationship between socioeconomic status and treatment outcome.

Research over the past 10 years has shown that there is a frequent but weak relationship between social class and length of treatment. However, the variables of sex, age, and diagnosis do not show this relationship. Although more black than white patients tend to terminate treatment early, there is no consistent pattern. Research aimed at investigating the preparation of patients for psychotherapy has not been successful in showing that patients so prepared (e.g., by a role induction interview) remain in treatment longer. Similarly, psychological tests and the study of patient expectations to predict continuation in psychotherapy have not been successful. However, the study of the interactional variables between the patient and therapist have revealed that those who dropped out liked the clinician less, felt less respected, experienced a weaker therapeutic alliance, viewed the psychotherapist as more passive, and viewed psychotherapy as being less potent that other helping interventions.

The relationship of patient characteristics to the outcome of psychotherapy has become one of the most critical areas in psychotherapy research. The patient as a factor in this interaction is complicated enough, and there are additional problems resulting from the wide variation in conceptions and assessments of outcome. Besides these, researchers have also been faced with the complications introduced by variations in the type of therapy offered, in the training and skill of the therapists studied, and in the kinds of patients treated. Based on these considerations the most recent research has shown no relations between outcome and the social class, age, and sex of the patient.

The degree of patient disturbance has been widely studied and has revealed a relation between more serious disturbances and poorer outcomes. Many other patient variables in relation to outcome have also been studied including ego strength; the personality attributes of affiliation, motivation, and intelligence; patient expectations; patient attractiveness; and patient-therapist similarity. However, none appear to be as important as in-therapy process variables such as patient openness during therapy, which has been significantly associated with better outcomes. The continuation or outcome of psychotherapy clearly cannot be predicted based on patient variables alone.

Numerous patient variables have been proposed to be prerequisites for productively engaging in the work of psychotherapy, including ego strength, motivation, and intelligence. For example, it has long been presumed that psychoanalytic psychotherapy required intelligent, verbally facile, insightful patients. In like manner, most therapists would agree that motivation for treatment is essential. There is limited empirical evidence regarding the relationship between these variables and outcome, however, because they have been relatively infrequently studied and because, with the exception of intelligence, these variables tend to change as treatment progresses (i.e., they may be the focus of treatment). The available evidence suggests that the association between ego strength and outcome is positive but unimpressive and that the association between intelligence and therapeutic outcome appears to be positive but fairly small. Finally, there is not even a clear relation between treatment effectiveness and patient motivation.

Investigations of the relation between patient disturbance and outcome has yielded the most consistent findings and a relatively consistent relation between severity of disturbance and poorer outcome has been reported. Studies converge to suggest that “the rich get richer,” that is, patients with less disturbance at the beginning of treatment continue to be relatively healthier at the end of treatment than their more distressed peers. Regarding particular mental illnesses, the efficacy of specific psychotherapies for specific conditions has been established. For example, psychosocial family education has been shown to reduce familial distress and risk for relapse in schizophrenic patients receiving medication. There is evidence supporting the effectiveness of psychosocial interventions in the treatment of depressive and anxiety disorders as, for example, exposure-based procedures have been shown to be consistently effective in the treatment of panic disorder. An exciting area of research concerns psychotherapy with individuals with personality disorders. Although individuals with comorbid personality disorders and Axis I disorders tend to have poorer outcomes in general, relatively effective treatments are being developed.

Some of the best research regarding patient variables concerns the interaction of patient and therapist and therapy variables. The idea of matching patient and therapist on either demographic or personality characteristics has begun to be examined, but theoretical reasoning is still more plentiful than are empirical findings. The research into matching patients and therapists has been too recent and inconsistent to draw firm conclusions or, more importantly, to make recommendations, but some important findings are emerging. A 1995 study in Los Angeles County examined the effect of matching patient and therapist ethnicity in the treatment of over 13,000 Asian-Americans, Mexican-Americans, African-Americans, and whites. In all ethnic groups, matched patients stayed longer in therapy, but matching was

positively related to outcome only for the Mexican-American group.

Other studies have attempted to match patient characteristics and therapy intervention. By definition, such research requires standardized therapies so research is this area is still fairly new and conclusive statements cannot be made. One study examined patient variables of coping style (i.e., externalization and defensiveness) as patients interacted with different types of therapy for major depression. Results suggested that externalizing patients and patients low in defensiveness did better in cognitive therapy, whereas internalizing patients and patients high in defensiveness did better in supportive or self-directed therapy.

Methodological problems associated with research into patient variables are complex, but the field is improving because of the introduction of standardized treatments and the development of more reliable and valid measures. Particularly interesting and exciting are efforts to match patient variables with treatment processes to optimize outcome.

Diagnosis Psychotherapy for individual conditions continues to be studied and includes a variety of psychotherapeutic modalities for a variety of psychopathological conditions. Research emphasis has focused on “empirically validated” psychotherapies, which include cognitive and behavioral psychotherapy, interpersonal psychotherapy, and short-term dynamically oriented psychotherapy.

SCHIZOPHRENIA Psychosocial family education has been shown to reduce the risk for relapse in schizophrenic patients receiving medication. Familial distress can also be reduced by this intervention.

MOOD DISORDERS The empirical evidence supporting the role of psychotherapy and psychosocial interventions in the treatment of patients with depression has become increasingly established. Based partially on improved research design and data analysis, this body of research has established the efficacy of psychosocial treatments of depression in its acute phase. The role of cognitive therapy in the treatment of depression also has become firmly established by meta-analyses of the results of psychotherapy research and in a review of the American Psychiatric Association's *Practice Guidelines on Manic Depressive Disorder and Bipolar Disorder*. According to other researchers this form of psychotherapeutic intervention has even been shown to be superior to pharmacotherapy in many cases. This treatment has also been studied and judged to be effective in its application to some specific populations, for example, to treat depressed patients infected with human immunodeficiency virus (HIV). It has been suggested that alone or in combination with medication, psychotherapy is probably the most effective treatment for many individuals suffering from depression because of its lack of side effects, because of its acceptance by certain patients, and because it deals with some of the social issues that may prolong depression or lead to relapse.

ANXIETY DISORDERS There is now evidence that psychosocial treatments are effective in the treatment of every anxiety disorder when compared to no treatment. However, many of these psychosocial interventions, which can be characterized as cognitive-behavioral with contributions from interpersonal and dynamic approaches, have limited clinical applicability and further research is needed. In panic disorder exposure-based procedures have been shown to be consistently effective. Meta-analyses of treatment outcome of panic disorder have also supported these findings. In addition, empirical studies of defense mechanisms in panic disorders have also been undertaken and have proven useful in the psychotherapy of this condition. Further research has been undertaken recently on the psychotherapeutic treatment of social phobia, generalized anxiety disorder, obsessive-compulsive disorder, specific phobia, and posttraumatic stress disorder. Combinations of exposure and cognitive therapies have shown some initial promise, and comparisons to pharmacotherapy have continued to show that psychotherapeutic treatments (most often cognitive-behavior therapy) demonstrate superior results.

P>**SUBSTANCE ABUSE AND DEPENDENCE** There are only a few well-controlled studies on the psychosocial treatment for substance abuse, but the existing data suggest that treatment for substance abuse should involve multiple modalities targeted to various specific problems in this population, including comorbid psychiatric disorders. Studies have shown that psychosocial treatments for substance abuse seem to be beneficial. In recent research supportive-expressive psychotherapy for patients with opioid dependence during methadone (Dolophine) maintenance treatment in community programs resulted in longer-lasting gains than did drug counseling. Specific recommendations based on empirical research indicate the usefulness of community reinforcement approaches with a subset of patients with cocaine dependence, family therapy for adolescents with substance abuse problems, and equivocal results with relapse prevention therapy, which perhaps is more effective for cocaine-dependent than for marijuana-dependent individuals.

PERSONALITY DISORDERS The psychotherapeutic treatment of individuals with personality disorders remains challenging and complex because of the heterogeneity and the variable severity of these disorders, and the observation that personality traits and their corresponding disorders are resistant to change and very difficult to modify. Individuals with a personality disorder in addition to an existing Axis I disorder have been shown to have poorer outcomes in general. In one study patients with major depression with a comorbid personality disorder were found to have more severe psychiatric disturbance at intake and they did not improve as much as those without a personality disorder. Although research findings support a poorer response to treatment for those with a personality disorder, the limited data that are available suggest that the presence of depression may be a positive prognostic indicator for patients with antisocial personality disorder. Here too the efficacy of short-term psychotherapy is an area of increasing interest. In a 1994 study two forms of short-term psychotherapy were employed to treat patients with personality disorders and patients in both therapies improved significantly as compared to patients who were not treated in this way.

Therapist Variables Therapist variables—the qualities or characteristics of psychotherapists that contribute to the process and outcome of psychotherapy—have been frequently studied and most empirical studies have failed to demonstrate a significant correlation between therapist demographic variables and outcome. There is a modest relation between age similarity of therapist and patient and outcome. Most reviews of research on the topic of therapist gender have failed to support the concern that male therapists may inhibit the progress of female patients. There is some evidence that patients of female therapists may have greater symptomatic improvement, but most studies do not indicate an effect on treatment outcome based on the sex of the therapist or on a match between the sex of the therapist and the patient. The opinion that ethnic similarity and ethnic sensitivity of the therapist result in enhanced outcomes has only received limited support from the empirical literature.

Therapist characteristics such as personality and coping patterns, emotional well-being, values and beliefs, and cultural attitudes are difficult to study experimentally because they are not amenable to experimental control. Recent research has focused on therapist traits such as dominance, locus of perceived control, and therapist conceptual level. Some findings suggest that similarity of cognitive style and level may facilitate patients remaining in therapy and experiencing speedier improvement. The emotional well-being of the therapist has been studied in terms of degree of perceived therapist distress on disturbance and whether the therapist has undergone personal therapy. Positive therapist mental health can enhance treatment effectiveness, especially in high-functioning patients, and a therapist's emotional problems can impact negatively on therapeutic progress. However, because the reasons why patients go into psychotherapy are so varied, the role of personal therapy on effectiveness is also quite varied. The therapist's values, attitudes, and beliefs have come under scrutiny because of the concern that these traits may exert an unwanted influence on the practice of psychotherapy. Religious beliefs and other general attitudes and values have been studied. Many studies have found that psychotherapy improvement may be augmented by a complex pattern of similarity and dissimilarity. No reliable conclusions have been drawn to date regarding attributes such as gender, lifestyle, and socioeconomic background.

The theoretical orientation of psychotherapists has been shown to have differing effect sizes in a number of meta-analytic studies. However, since the effects of specific interventions have been difficult to separate from therapist orientation, conclusions about any particular therapeutic orientation are as yet premature. Three interrelated variables have been studied in relation to the therapist's professional background and the effectiveness of therapy: level of professional training, amount of experience, and professional discipline. Research in these areas has yielded equivocal and contradictory results. Recently, however, a meta-analysis of graduate training in psychotherapy has found that treatment outcomes and treatment duration are associated with more training. Therapeutic styles are inexorably intertwined with therapeutic interventions. Interpersonal styles, verbal styles, nonverbal styles, and combined verbal and nonverbal patterns have all been studied, but no stable conclusions have emerged. Therapist interventions or technical procedures designed to bring about psychotherapeutic results relate to both therapist qualities and therapy processes. One such device for attaining therapeutic efficacy is the therapy or treatment manual. Because therapies that use therapy manuals can be empirically distinguished, they are often used in psychotherapy research. Results from studies employing therapy manuals show that a significant positive relation exists between compliance with a treatment manual and outcome of treatment. Additional research also shows that therapists' relationship skills and supportiveness may be diminished because of the therapists' focus on compliance with a manual so the use of these treatments may only be effective on certain types of patients. Specific therapist interventions have also been studied and include therapist directiveness; therapist self-disclosure; and therapist interpretations on transference, motives, and resistance, but no clear trends have emerged from these studies.

Treatment Process Variables: The Therapeutic Alliance Studying the effectiveness of psychotherapy entails identifying its curative elements, that is, determining what process variables are related to patient improvement. One aspect that has been repeatedly and consistently identified is the *therapeutic alliance*, sometimes called *therapeutic relationship, working alliance*, or *therapeutic bond*. For example, studies have shown that patients who dropped out of therapy prematurely liked the clinician less, felt less respected, viewed the psychotherapist as more passive, and generally experienced a weaker therapeutic alliance. Research reviews also have consistently concluded that a positive therapeutic relationship, whether measured from the perspective of nonparticipant observers or from the perspective of the

patient or the therapist, is related to therapeutic effectiveness across a wide variety of therapeutic modalities and patient problems. A meta-analysis of 24 studies of the relation between working alliance and treatment outcome concluded that there is a consistent relation between the quality of the alliance and positive outcome. The quality of the alliance appears to be influenced by a variety of interpersonal and intrapersonal patient characteristics. Patients who have a better history of and capacity for social relationships tend to develop a better alliance than defensive and negativistic patients. Interestingly, problem severity does not appear to affect the quality of the alliance. An exciting area of research concerns tracking the alliance across therapy, including monitoring therapist interventions when there are problems in the alliance. There is convergent evidence that measures of the alliance early in treatment are more strongly related to eventual outcome than later measures. This appears to reflect the fluctuating nature of the quality of the relationship across therapy and the frequent finding that during longer-term treatments the alliance is often disrupted and must be attended to in the course of successful treatment.

Current research is investigating whether the alliance mediates or moderates change. Carl Rogers proposed that the alliance was curative in and of itself, theorizing that the patient's experience of a nonjudgmental, warm, genuinely caring attitude on the part of the therapist mediated positive personality change. In contrast, Freud postulated that the relationship was a necessary precondition to or moderator of the work of psychoanalysis, which was responsible for positive change. Whether the relationship between the patient and the therapist is real or distorted has to do with the analytic concept of *transference*, which holds that the feelings that patients have toward therapists are always a mixture of attributes displaced onto the therapist from figures in the patient's past and the real characteristics of the therapist. Further research into the causes, uses, and effects of the alliance is clearly needed.

Patient Treatment Matching The aptitude-by-treatment-interaction approach to psychotherapy research attempts to identify the treatment strategies and procedures that produce optimal outcomes for individual patients. The aptitude-by-treatment interaction approach describes patients and therapists dimensionally with quantitative differences (e.g., resistive, impulsive, extroverted, directive, supportive). The questions addressed in aptitude-by-treatment interaction research include: (1) What procedures are important to effective treatment in different contexts? (2) Which patients are likely to respond therapeutically to therapist/therapy characteristics? and (3) How can patients be effectively assigned to therapies and their response monitored within those therapies?

The aptitude-by-treatment interaction model has been used to show that highly anxious, repressed patients are more responsive to emotion clarification on supportive approaches whereas externally cued patients respond better to more directive therapies such as cognitive behavioral psychotherapy. Also, supportive therapy has been shown to be more effective with patients having higher levels of distress; cognitive therapy has been shown to be more effective with patients presenting with lower levels of subjective distress. Research has provided evidence that depressed caregivers of older adults are differentially responsive to psychodynamic and cognitive behavior psychotherapy depending on the length of time in the caregiving role. Although largely unreplicated and few in number, these aptitude-by-treatment interaction studies support the principle of differential treatment selection.

Psychotherapy Combined With Pharmacotherapy A large proportion of patients who obtain treatment for mental disorders receive a combination of psychotherapy and medication. This blending of treatments is used widely for the treatment of depression, anxiety disorders, substance use disorders, schizophrenia, and some personality disorders. The American Psychiatric Associations Practice Guidelines on the treatment of depression and substance use disorders have embodied this combination of therapeutic approaches. Since the 1990s research on the relative efficacy of pharmacotherapy and psychotherapy, both in comparison and in combination with one another, has greatly expanded. Double-blind, controlled clinical trials have been the standard research design employed to study medications alone. However, newer strategies, such as the four-group factorial design in which both treatments are evaluated individually against one another, against a control group, and against their combination and the six-group design in which the placebo effect is evaluated, have been developed. The addition of a supplemental treatment to a proven standard treatment is another research design useful in assessing the value of combined treatments.

There are three possible outcomes of comparative and combined psychotherapy-drug treatments: (1) no therapeutic effect (i.e., combined treatment provides the same effect as individual treatment; (2) positive effects (i.e., combined treatment is better than individual treatment either by an additive and synergistic effect)—greater than the sum of the two treatments—or by a facilitative interaction—the treatment is only effective when combined; and (3) negative effects, (i.e., combined treatment shows less improvement than individual treatments). The hypothesized clinical mechanisms of positive action of medications on combined treatments are that medications facilitate psychotherapeutic accessibility, medications influence ego functions for participation in psychotherapy, medications promote psychotherapeutic abreaction, and medications have a positive effect on attitude and expectations. The hypothesized clinical mechanisms of positive action of psychotherapy on pharmacotherapy are that psychotherapy facilitates compliance with medications, that psychotherapy itself is rehabilitative, and that psychotherapy teaches skills that can be used to prevent relapse. However, it has also been hypothesized that combined psychotherapy and pharmacotherapy may have some general negative effects, but little empirical research has been done in this area. In relation to psychotherapy, medications may produce a placebo effect that promotes authoritarianism, reduces symptoms such that psychotherapy is prematurely terminated, and may undermine useful defenses. On the other hand, in relation to pharmacotherapy, psychotherapy may undercut compliance.

A number of different strategies have been identified for combining treatment modalities. The combination of either individual, group, or family psychotherapy with psychopharmacotherapy has been studied. In addition, psychoeducational approaches, self-help groups, and medication groups have proved useful in some instances. These strategies have been shown to improve psychotherapy process and outcome as well as to enhance medication compliance, treatment adherence, the prevention of relapse, and the recurrence of illness.

EFFICACY, EFFECTIVENESS, AND EFFICIENCY

Psychotherapy has been evaluated from three research perspectives: efficacy, effectiveness, and efficiency. Each of these perspectives has a distinct purpose, methodology, and interpretative context.

Efficacy Efficacy studies typically address the sufficiency of a particular highly controlled treatment in a specific setting with a carefully selected set of patients having a specific disorder. Efficacy studies are designed to control extraneous mediating factors such as therapist training (e.g., degree of treatment structure, protocol compliance), other treatments (single, combined, or multiple), the roles of the therapist (e.g., assessment and selection, monitoring of patient progress), participant selection (homogeneous or comorbid presentations), and treatment parameters (duration, dose, modality). The *sine qua non* of efficacy research is random assignment to treatment and control conditions. What constitutes an appropriate control condition has received considerable attention. Among the control group alternatives are (1) the no-treatment control, (2) the waiting-list control, (3) the placebo-attention control, and (4) the “usual care,” “best alternative treatment” control, also labeled the “minimal-treatment” control). Confidence in treatment efficacy increases as study comparisons reveal that a particular psychotherapy is better than no therapy, better than a nonspecific therapy, and better than an alternative therapy.

Over the past 15 years controlled clinical trials of psychotherapy have attempted to standardize therapy structure through the use of treatment manuals that provide a treatment rationale, goals, specification of treatment processes, and a sequence of action for conducting the psychotherapy. These manuals are efforts to standardize and operationalize treatments with greater specificity.

Prominent book-length examples of the treatment manual include David Barlow's *Anxiety and Its Disorders;* Aaron Beck, John Rush, Brian Shaw, and Chad Emery's *Cognitive Therapy of Depression;* Gerald Klerman, Myrna Weissman, Bruce Rounsaville, and Eve Chevron's *Interpersonal Psychotherapy of Depression;* Marsha Linehan's *Cognitive Behavioral Treatment of Borderline Personality Disorder;* Lester Luborsky's *Principles of Psychoanalytic Psychotherapy: A Manual for Supportive-Expressive Treatment;* and Hans Strupp and Jeffrey Binder's *Psychotherapy in a New Key: A Guide to Time-Limited Dynamic Psychotherapy*. Many more textbook and manual approaches take the form of methodological manuscripts in scientific journals. In 1995 the Task Force on Promotion and Dissemination of Psychological Procedures of the American Psychological Association drafted and published the empirical evidence to support these therapies; this list is periodically updated as new studies are published.

Psychotherapy manuals promote the use of a single conceptual framework for a therapy, and a set of prescribed techniques within that framework. Psychotherapy manuals are not so specific as to dictate the therapist's actions at each moment during the therapeutic interaction. Rather, they supply guidelines for the application of specific interventions and goals for sessions. Reactions to treatment-manual therapies include claims of excessive rigidity, of overlooking patient and therapist variability, and of irrelevance to the clinical practice of psychotherapy. Defenders of treatment-manual therapies have cited the importance of codifying effective behaviors for treating specific problems, the manuals' value in training new therapists, and the development of even more improved psychotherapies. There is some evidence that studies using treatment manuals had larger effect sizes than studies that did not use them.

The specification of therapies in manuals led to an interest in the evaluation of adherence to those prescriptions. Efficacy studies typically have included a compliance review through video tapes or audiotapes of therapy sessions. This review often uses a scale developed to assess that particular textbook psychotherapy. Scales are

available for cognitive therapy, interpersonal psychotherapy, time-limited dynamic psychotherapy, and supportive-expressive psychotherapy, among others.

Meta-analyses of psychotherapies collectively and of types of psychotherapy individually yield effect sizes of 0.75 to 0.85. Regardless of the outcome measure used, the means of the treated groups are at about three fourths of a standard deviation better than the means of the untreated groups. Meta-analyses of types of psychotherapy generally have not established the advantages of one type of therapy over another when considered across contexts, patients, and presenting problems.

Effectiveness Effectiveness studies are concerned with the delivery of psychotherapy in applied clinical settings. Psychotherapy may be only one component of an integrated intervention effort involving other health and social services. The experimental controls used in efficacy studies are usually absent in effectiveness studies. Patients may satisfy diagnostic criteria for multiple disorders. Moreover, the length of treatment may be determined more by patient motivation or resources or by the limitations of health care financing than by treatment prescription. Paradoxically, it seems that time-limited, planned short-term treatments are actually of longer median duration than naturally occurring traditional psychotherapies. Also, in effectiveness research, mental health professionals are likely to come from several disciplines and to vary in levels of training and experience. The goal of effectiveness research is to describe the response to the typical delivery of psychotherapy for patients who seek it, and to identify the treatment parameters that best predict patient response.

Some effectiveness studies begin with a program audit of a treatment package. These studies may include an assessment of the characteristics of settings and providers, the representativeness of the treatment samples, the treatment actually delivered, the barriers to treatment, and the costs of treatment. Examples include clinic studies of psychotherapy with children and adolescents, and a combination of medication, social skills training, and family education for patients with severe mental illness. Effectiveness studies often employ quasi-experimental designs and involve nonequivalent comparison groups in an attempt to make findings most relevant for actual practice.

Another approach to effectiveness studies is to describe the patterns of mental health service delivery assuming that efficacy studies have established the causal relation linking psychotherapy and outcome. This approach makes use of repeated measurements (multiwave data) in the course of psychotherapy, and uses the tenets of measurement reliability such as the “reliable change index” and “clinically significant change” to show how patients change over time. This has produced several models of patients' change in the course of psychotherapy, including the explication of a dose-response relation between the logarithm of the number of sessions and the normalized probability of improvement. A dose response of 50 percent (of patients improving) has been observed with approximately 8 sessions of psychotherapy; a dose reponse of 75 percent occurred with approximately 26 sessions; and improvement reached asymptope at approximately 85 percent after a little over a year of psychotherapy. The phase model has identified categories of change—subjective well-being, symptoms, and life functioning—that improve at different rates and in a probabilistically necessary causal sequence. Patients must first be remoralized (improvement in subjective well-being) before symptoms can improve, and symptoms must improve before life functioning can improve. Clinical trials of the treatment of depression with psychotherapy have shown that work functioning does not improve until the symptoms of depression have ameliorated. Other research using this approach to effectiveness research has shown that symptoms can be categorized as *acute, moderate*, and *chronic*, each with a different dose-response pattern.

The results of effectiveness studies are more generalizable to actual clinical practice than are the results of efficacy studies. However, because of the use of random assignment the results of efficacy studies may be less confounded by other influences than are the results of effectiveness studies. In defense of the onsumer Reports study, Martin Seligman has persuasively argued the relative merits of these approaches. That study showed that in actual practice psychotherapy was a very effective intervention for a wide variety of psychiatric conditions.

Efficiency The identification of patterns of change for aggregated cases of individual psychotherapy patients has provided a context for the study of individual patients and an assessment of the efficiency of treatment for each individual. Hierarchical linear modeling techniques have been used to show that a set of variables collected at intake could be used to predict the pattern of response of an individual psychotherapy patient. For some patients, the expected pattern of response is rapid and likely to require relatively few sessions of psychotherapy; for other patients the expected pattern of improvement is gradual and likely to require many sessions of psychotherapy to reach clinically significant change. Identification of these predictive patterns takes into account the variables—severity, previous experience with psychotherapy, difficulty of attending sessions—that mediate and moderate patient change. A particular treatment can be judged against its expected pattern of change, and a therapist's performance with severe or less severe cases can be more fairly judged against expected patterns of change. The monitoring of progress in psychotherapy and the allocation of therapeutic resources (e.g., sessions) also are made more rational; patients are judged against an expected pattern of change instead of against a prototype or ideal pattern.

The field of psychotherapy research is represented by the international interdisciplinary Society for Psychotherapy Research, which has over 1300 members from about 40 countries. The Society holds an annual meeting, and individual chapters (e.g., North America, Europe, United Kingdom, South America) also hold regular meetings. These meetings provide a forum for the exchange of ideas and findings. Research methods and findings are published in the Society's journal, *Psychotherapy Research* and in the several editions of the *Handbook of Psychotherapy and Behavior Change*, which is edited by Alan Bergin and Sol Garfield.

On the basis of over 1000 controlled studies, it can be stated unequivocally that psychotherapy is efficacious for a broad range of psychiatric disorders. However, efficacy and effectiveness studies have not yet clearly shown one kind of psychotherapy to be consistently superior to another; the average outcomes of treatments (across therapists, patients, and settings) are fairly equivalent. Comparable average effects over a range of patients would also be true of antibiotics and antidepressants. The task remains to show differential outcomes for different treatments for different patients. Until specific treatments that are consistently superior for homogeneous groups of patients can be found, the focus must remain on concurrent assessments to provide feedback regarding a patient's response to current treatment. As in the rest of medicine, the idea here is to conduct research that will yield information that can be used as systematic feedback to shape treatment in a way that indicates the need for a different intervention when the patient is not benefiting and indicates the continuation of a beneficial course.

SUGGESTED CROSS-REFERENCES

Psychotherapy is discussed in Chapter 30, schizophrenia in Chapter 12, mood disorders in Chapter 14, anxiety disorders in Chapter 15, substance-related disorders in Chapter 11, and personality disorders in Chapter 24.

SECTION REFERENCES

Antonuccio D, Dandon W, DeNelsky G: Psychotherapy versus medication for depression: Challenging the conventional wisdom with data. Prof Psychol Res Pr *26:*574, 1995.

Barchas J, Marzuk P, Beutler L: Introduction to the special section on the contribution of psychotherapy and pharmacology research to national mental health care. J Consult Clin Psychol *64:*635, 1996.

Barlow D, Lehman C: Advances in the psychosocial treatment of anxiety disorders: Implications for national health care. Arch Gen Psychiatry *53:*727, 1996.

Beitman B, Klerman G, editors: *Integrating Pharmacotherapy and Psychotherapy*. American Psychiatric Press, Washington, DC, 1991.

*Bergin AE, Garfield SL, editors: *Handbook of Psychotherapy and Behavior Change*, ed 4. Wiley, New York, 1994.

Bruce T, Spiegel D, Gregg S, Nuzzarello A: Predictors of alprazolam discontinuation with and without cognitive behavior therapy in panic disorder. Am J Psychiatry *152:*1156, 1995.

Busch F, Shear K, Cooper A, Shapiro T: An empirical study of defense mechanisms in panic disorder. J Nerv Ment Dis *183:*299, 1995.

Carroll K, Rounsaville B, Nich C, Gordon L, Wirtz PW, Gawin F: One-year follow-up of psychotherapy and pharmacotherapy for cocaine dependence: Delayed emergence of psychotherapy effects. Arch Gen Psychiatry *51:*989, 1994.

Clarkin J, Pilkonis P, Magruder K: Psychotherapy of depression: Implications for reform of the health care system. Arch Gen Psychiatry *53:*717, 1996.

Crits-Christoph P, Siqueland L: Psychosocial treatment for drug abuse: Selected review and recommendations for national health care. Arch Gen Psychiatry *53:*749, 1996.

Diguer L, Barber J, Luborsky L: Three concomitants: Personality disorders, psychiatric severity, and outcome of dynamic psychotherapy of major depression. Am J Psychiatry *150:*1246, 1993.

Docherty JP, Streeter MJ: Progress and limitations in psychotherapy research: A focus on depression. J Psychother Res Pract *2:*100, 1993.

Feske U, Chambless D: Cognitive behavioral versus exposure-only treatment for social phobia: A meta analysis. Behav Therapy *26:*695, 1995.

Frueh C, Turner S, Beidel D: Exposure therapy for combat-related PTSD: A critical review. Clin Psychol Rev *15:*799, 1995.

Goldstein MJ: Psychoeducation and relapse prevention. Int Clin Psychopharmacol *9:*59, 1995.

Gould R, Otto M, Pollack ML: A meta-analysis of treatment outcome for panic disorder. Clin Psychol Rev *15:*819, 1995.

Grawe K: Research-informed psychotherapy. Psychother Res *7:*1, 1997.

Hollon SD, DeRubeis RJ, Evans MD, Weimer MJ, Garvey MJ, Grove WM, Tuason VB: Cognitive therapy and pharmacotherapy for depressions: Singly and in combination. Arch Gen Psychiatry *49:*774, 1992.

Hollon S, Fawcett J: Combined medication and psychotherapy. In *Treatment of Psychiatric Disorders*, ed 2, G Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

*Hollon S: The efficacy and effectiveness of psychotherapy relative to medication. Am Psychol *51:*1025, 1996.

Horvath AO, Luborsky L: The role of the therapeutic alliance in psychotherapy. J Consult Clin Psychol *62:*561, 1993.

*Howard KI, Moras K, Brill PL, Martinovich Z, Lutz W: The evaluation of psychotherapy: Efficacy, effectiveness, patient progress. Am Psychol *51:*1059, 1996.

Jacobson N, Hollon S: Cognitive-behavior therapy versus pharmacotherapy: Now that the jury has returned its verdict, it is time to present the rest of the evidence. J Consult Clin Psychol *64:*74, 1996.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen H, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the national comorbidity survey. Arch Gen Psychiatry *51:*8, 1994.

*Kopta M, Howard K, Lueger R, Sanders S: Individual psychotherapy outcome and process research: Challenges leading to greater turmoil or a positive transition. Ann Rev Psychol *50:*441, 1999.

Kranzler H, Burleson J, Korner P, Del Boca F: Placebo-controlled trial of flouoxetine as an adjunct to relapse prevention in alcoholics. Am J Psychiatry *152:*391, 1995.

*Lipsey MW, Wilson DB: The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. Am Psychol *48:*1181, 1993.

Markovitz J, Klerman G, Clougherty K, Spielman L, Jacobsberg LB, Fishman B, Frances AJ, Kocsis JH, Perry SW: Individual psychotherapies for depressed HIV-positive patients. Am J Psychiatry *152:*1504, 1995.

Persons J, Thase M, Crits-Christoph P: The role of psychotherapy in the treatment of depression: Review of two practice guidelines. Arch Gen Psychiatry *53:*283, 1996.

Russell R, Orlinsky D: Psychotherapy research in historical perspective: Implications for mental health care policy. Arch Gen Psychiatry *53:*708, 1996.

Sanders S: Psychotherapy is effective: A brief review of empirical literature supporting the effectiveness of psychotherapy. In *Controversies in Psychotherapy and Counseling*, C Feltham, editor. Sage Publications, London, 1998.

Saunders SM: Applicant's experience of the process of seeking psychotherapy. Psychotherapy *30:*554, 1993.

*Seligman MEP: The effectiveness of psychotherapy: The *Consumer Reports* study. Am Psychol *50:*965, 1995.

Shea M, Widiger T, Klein M: Comorbidity of personality disorders and depression: Implications for treatment. J Consult Clin Psychol *60:*857, 1992.

Simon GE, VonKorff M, Durham ML: Predictors of outpatient mental health utilization by primary care patients in a health maintenance organization. Am J Psychiatry *151:*908, 1994.

Smith ML, Glass GV, Miller TI: *The Benefits of Psychotherapy*. The Johns Hopkins Press, Baltimore, 1980.

Sperry L, Brill P, Howard K, Grissom G. *Treatment Outcomes in Psychotherapy and Psychiatric Interventions*. Brunner/Mazel, New York, 1996.

Vessey JT, Howard KI: Who seeks psychotherapy? Psychotherapy *30:*546, 1993.

Wampold B: Methodological problems in identifying efficacious psychotherapies. Psychother Res *7:*21, 1997.

Weissman M, Klerman G: Interpersonal psychotherapy for depression. In *Integrating Pharmacotherapy and Psychotherapy*, B Beitman, G Klerman, editors. American Psychiatric Press, Washington, DC, 1991.

Winston A, Laikin M, Pollack J, Samstag LW, McCullough L, Muran JC: Short-term psychotherapy of personality disorders. Am J Psychiatry *151:*190, 1994.

Woody GE, McLellan AT, Luborsky L, O'Brien CP: Psychotherapy in community methadone programs: a validation study. Am J Psychiatry *152:*1302, 1995.

30.12 COMBINED PSYCHOTHERAPY AND PHARMACOTHERAPY

GLEN O. GABBARD, M.D.

Psychiatry is the medical specialty that integrates the biological and the psychosocial perspectives in both diagnosis and treatment. The provision of optimal clinical care requires avoiding biological or psychological reductionism. Often the best psychiatric treatment plan for a given patient involves a combination of medication and psychotherapy. Of all medical specialists and mental health professionals, the psychiatrist is uniquely positioned to administer both treatments.

The challenge of integrating a psychosocial and biological understanding of the patient is a formidable one. The breadth of psychiatry as a specialty requires psychiatrists to be knowledgeable about neurotransmitters, the latest psychopharmacological agents, and the interface between genetics and environment. At the same time, they must be familiar with intrapsychic conflicts, patterns of relationships, and psychological meanings of symptoms. This interface between “mind” and “brain” makes psychiatry one of the most intellectually stimulating of all pursuits. Moreover, if the specialty were reduced to the prescription of medication, other medical specialists could easily replace psychiatrists. Similarly, if psychiatry were confined to psychotherapeutic interventions, other mental health professionals could perform similar functions and make psychiatrists unnecessary. Combined psychotherapy and pharmacotherapy is the daily bread and butter of the psychiatric clinician.

MIND-BRAIN INTERFACE

The notion that mind and brain are separate entities, often referred to as *Cartesian dualism*, has long since given way to the recognition that mind and brain are inextricably linked. Nevertheless, the task of integrating what is psychological and what is biological can be extraordinarily complex. Much of this difficulty arises from the irreducible subjectivity of consciousness that makes the reduction of the psychological to the biological highly problematic. William James suggested that consciousness should be regarded as a process rather than a substance. Subsequent philosophers of the mind-body problem have stressed that even though all mental states are embodied and therefore coexist with physical states, neuroscientific explanations of psychological states fail to capture the indispensable nature of first-person subjectivity.

One investigation used positron emission tomography (PET) to measure cerebral blood flow in seven healthy subjects while at rest. The patterns of the blood flow were then compared to measures of the same subjects when they imagined or recalled a sad event. When the subjects were sad, the researchers identified significant differences in the regional cerebral blood flow in the inferior and orbitofrontal cortices. This investigation illustrates that the psychological and biological, while interconnected, are different domains with different languages. The subjective state of sadness has physiological correlates in the brain that are measurable. However, the memories and personal meanings associated with the state of sadness cannot simply be reduced to changes in cerebral blood flow in specific areas of the cortex.

There was a time in psychiatry when medication would be prescribed for disorders regarded as having a biological etiology, and psychotherapy would be employed for psychological disturbances, but recent research has rendered this dichotomy highly questionable. Studies using PET have demonstrated that behavior therapy and cognitive therapy have effects similar to those of fluoxetine (Prozac) on the cerebral metabolic rates in the head of the right caudate nucleus in patients with obsessive-compulsive disorder. The most biological disorder in psychiatry, schizophrenia, is the one with perhaps the most persuasive evidence that psychotherapy can prevent relapse; indeed, some studies suggest that family therapy reduces relapse of schizophrenia to the same extent as antipsychotic medication does.

Recent investigation with cancer patients provides further evidence of the impact of psychotherapy on patient physiology. Metastatic breast cancer patients experiencing group psychotherapy lived an average of 18 months longer than did a randomly assigned control group who did not receive such therapy. Similarly, patients with malignant melanoma who attended a support group had more lengthy remissions and more favorable mortality rates than a control group who was not given the support group protocol.

The evidence accumulating from these studies suggests that psychotherapy must work by changing the brain. If psychotherapy is a form of learning, then studies on animals who are lower on the phylogenetic ladder may be specially relevant. Scientific studies on rats and marine snails have demonstrated that the number of synapses per neuron actually increases in response to learning based on environmental experiences and social contacts with others of the same species; in other words, neuroanatomical changes accompany learning and group interaction.

Also, the impact of neurotransmitters may change according to psychosocial influences. A neuron in the crayfish has a dramatically different response to serotonin depending on the animal's social status. In dominant animals, serotonin makes the neuron fire; in subordinate creatures, serotonin suppresses firing. This neuron that controls the tail-flip reflex (involved in the fight-or-flight-response) can change from firing to suppressing within the same animal. When two previously subordinate crayfish are put together, one becomes dominant and the neuron starts firing in response to serotonin as a result.

Although the leap from lower animals to human beings is substantial, the data are highly suggestive of a complex interaction between environmental and innate biological factors. Similarly, it is now known that environment has a substantial effect on gene expression. Less than 2 percent of all diseases in medicine follow familiar Mendelian patterns of inheritance; most involve some degree of genetic diathesis triggered by environmental stressors. Studies of depression, for example, have shown that genetic factors appear to influence the vulnerability of the individual to react to psychosocial stressors with a major depressive episode.

The interface of mind and brain was further complicated by research demonstrating that there are robust genetic and biological influences on what have often been considered to be personality traits. Shyness, reward-dependence, novelty-seeking, introversion-extroversion, persistence, and other features of personality appear to be influenced by genetic factors. In a similar vein, recent psychopharmacologic studies demonstrate that psychotropic medications may alter traits that have long been considered characterological.

This abbreviated overview of some of the recent research on the mind-brain interface and the interaction of genes and environment underscores the complexity of the conceptual underpinnings of combining medication and psychotherapy. Clearly, there is no simple one-to-one correlation between psychotherapeutic interventions and the so-called psychological aspects of illness. Neither can it be assumed that psychotropic agents affect only “hardwired” dimensions of the illness. What is fairly certain is that psychotherapy and medication often work synergistically to provide the best possible treatment for the patient.

PSYCHOTHERAPEUTIC MANAGEMENT

In conceptualizing how psychotherapeutic interventions and medication work together, two general strategies can be delineated. One is the psychotherapeutic management inherent in skilled pharmacotherapy practice; the other is the combination of formal psychotherapy and the prescription of medication.

In considering the former, a good starting point is the American Psychiatric Association “Practice Guidelines for Treatment of Major Depressive Disorder in Adults,” which stress that psychotherapeutic management is an essential component of every medication-based treatment plan. One way of understanding the recurrent finding that a placebo condition in a controlled trial is often an effective treatment for a significant number of patients is that good clinical management has psychotherapeutic effects. Simply by asking about symptoms and taking a history, the clinician often helps patients to become aware of connections between external events, the meaning of those events, and symptoms that lead them to insight about their illness. The fact that a caring physician is listening and providing help may be a powerful corrective emotional experience for some patients. Another mechanism of action may involve the so-called transference cure, in which the patient gets

better to please the physician.

Many psychodynamic principles derived from psychotherapy apply equally to pharmacotherapy practice. A pharmacotherapeutic alliance is essential to ensure that the patient understands the reasons for the medication and complies with the treatment plan, just as a psychotherapeutic alliance is essential to enlist the patient as a collaborator in psychotherapy.

This principle was persuasively demonstrated in an analysis of the role of the therapeutic alliance in the outcome of 225 depressed outpatients enrolled in the Treatment of Depression Collaborative Study sponsored by the National Institute of Mental Health. The patients were given one of four treatments: (1) 16 weeks of cognitive-behavior therapy, (2) 16 weeks of interpersonal therapy, (3) imipramine (Tofranil) plus clinical management, or (4) placebo plus clinical management.

Using videotapes and an instrument designed to measure patient and therapist behaviors, trained raters scored the therapeutic alliance for all patients in the project. Outcome was also assessed using standard depression-rating measures. The investigators found that the therapeutic alliance was just as important for drug therapy as for psychotherapy. Also, the alliance continued to be important whether it was measured in early sessions or later ones. The nature of the therapeutic alliance accounted for more of the variance in treatment outcome (21 percent) than the treatment method itself (1 percent). The patient's contribution to the alliance, rather than the therapist's, seemed to be the significant factor related to outcome. The investigators speculated that the importance of a good therapeutic alliance in the outcome of pharmacotherapy may partially explain why depressed patients often do worse in primary-care settings. Physicians average only about 8 minutes per patient in primary-care practice, and this constraint may interfere with the time and the psychotherapeutic management necessary to establish a good alliance.

Other dynamic principles, such as transference, resistance, and countertransference, also are an integral part of pharmacotherapy practice. Many problems with noncompliance can be traced to these principles. The application of psychodynamic constructs to compliance problems in pharmacotherapy is often referred to as *dynamic pharmacotherapy*.

Transference is not a phenomenon limited to psychotherapy. The attribution of qualities to the prescribing physician that stem from figures in the patient's past occurs routinely in clinical practice. A patient may perceive the psychiatrist as authoritarian and refuse to cooperate with the prescribed treatment plan because it reminds him of how his father used to bark orders at him and try to control him. If the psychiatrist then feels irritated at the patient for not cooperating and becomes more insistent, the problem may worsen because the clinician is behaving in the authoritarian manner that the patient fears. In this case, the psychiatrist's countertransference has entered into the equation and affected the patient's capacity to collaborate with a good alliance.

Resistance to taking medication may relate to issues of transference and countertransference in many cases. In others, the resistance stems from a fundamental ambivalence about getting better. In cases of depression, in particular, some patients may feel that they have committed such sinful and evil acts that they deserve to be punished for them by remaining depressed. An antidepressant medication may have a specific meaning for such patients,—the potential to relieve their suffering—so they may not fill their prescription or take the tablets as prescribed.

Medication may be imbued with a myriad of other meanings. Many patients view medication as a crutch. They may make no distinction between being addicted to a narcotic and taking maintenance dosages of antidepressant medication. Some patients may interpret needing medication as a sign of weakness, so they will often discontinue it on their own as soon as their symptoms start to abate. In still other cases, the medication has a meaning connected with a family member.

A woman who was not responding optimally to desipramine (Norpramin, Pertofrane) was prescribed lithium (Eskalith) in addition to the desipramine as an augmentation strategy. She took the prescription from her psychiatrist, but when she returned the following week, she acknowledged when asked that she had not filled the prescription. She promised that she would. At her appointment the following week, her psychiatrist asked her if she would come in to have her blood drawn for a lithium level. She sheepishly acknowledged that she had not been taking her lithium. When the psychiatrist inquired why she had chosen not to, she responded with intense emotion in her voice: “I will not be genetically tied to my mother.” Although the statement was absurd at one level, the psychiatrist understood what she meant. Her mother had been diagnosed as having bipolar disorder, had been prescribed lithium, and had ultimately committed suicide. To this particular patient, taking lithium meant that she would end up exactly like her mother.

Optimal psychotherapeutic management of a pharmacotherapy patient involves attention to these psychological dimensions of the relationship and to the medication itself. In addition, the clinician must empathize with the patient's perception of medication and try to appreciate the meaning it has for that patient. Good listening skills, attention to rapport, and a systematic effort to establish a good alliance based on careful explanations are all integral parts of this approach to pharmacotherapy.

MEDICATION IN COMBINATION WITH FORMAL PSYCHOTHERAPY

History In the 1940s and 1950s when psychoanalysis dominated the landscape of psychiatric practice, there was a reluctance to combine medication with psychotherapy. One of the general concerns about combined treatment was the fear that psychotropic agents might alleviate the anxiety or depression that motivated the patient to tolerate the frustrations and deprivations of psychotherapeutic exploration. Another concern was that the prescription of medication by the analyst or therapist would interfere with the development of the transference to the clinician and the subsequent understanding of that transference.

This belief regarding the inherent incompatibility of psychotherapy and medication persisted into the 1960s and 1970s, despite a number of studies showing no sign of the kinds of problems that clinicians had feared; in fact, research showed just the opposite. In many cases medication and psychotherapy working in concert produced superior results to either modality alone. In a 1993 review of 13 studies comparing dynamic psychotherapy versus other forms of psychotherapy, the investigators found the main trend of all the comparative studies to be nonsignificant differences in outcome. However, their conclusion of the nonsignificant-difference effect did not hold when applied to combined treatments versus single treatments. The advantage for combined treatments was striking. Psychotherapy plus pharmacotherapy was clearly superior to psychotherapy or pharmacotherapy alone.

In clinical practice today there is a widespread acceptance of the combined use of medication and psychotherapy. Even among psychoanalysts, who were once the most vocal critics of combined treatment, prescribing is commonplace. In a survey of members of the American Academy of Psychoanalysts, 90 percent of respondents said they were prescribing medications. In a study of psychoanalytic candidates' training cases at the Columbia University Center for Psychoanalytic Training and Research, medication was combined with psychoanalysis in 29 percent of the cases, reflecting the fact that medication is no longer seen as interfering with certification or graduation.

Many analysts and other therapists have noted that medication and psychotherapy work synergistically to improve outcomes in a wide variety of illnesses. Investigations of outcome geared to specific disorders consistently demonstrate advantages to combined approaches.

Specific Diagnostic Categories

Schizophrenia An extensive series of studies have examined the rate of relapse when a specific form of psychoeducational family therapy is combined with antipsychotic medication. This intervention is based on the observation that high levels of expressed emotion in the families of patients with schizophrenia predicted relapse following hospital discharge. High-expressed-emotion families have been characterized as excessively intrusive, critical, and overinvolved with the patient who suffers from schizophrenia. Psychoeducational family therapy is aimed at helping the family reduce the factors that constitute expressed emotion and at educating the family about schizophrenia and the need to continue antipsychotic medication indefinitely.

In a 1991 study by a team of investigators led by Gerard Hogarty, the addition of psychoeducational family therapy to antipsychotic medication reduced the relapse rate of patients with schizophrenia by half. The relapse rate was even further reduced at 1-year follow-up when social skills training was added to family therapy and medication. A similar study in 1994 found that there was a 77 percent hospital readmission rate in an 18-month period when patients with schizophrenia were given no medication and no family intervention. Either medication alone or family intervention alone reduced the rate of schizophrenia by approximately 50 percent. However, when the family intervention and medication were combined, only 10 percent of patients required readmission during the 18-month period.

An innovative form of individual therapy, termed personal therapy, has also shown promise in combination with antipsychotic medication. A 3-year controlled trial using random assignment, reported in 1997 that combining personal therapy and medication (most commonly fluphenazine or haloperidol [Haldol] administered intramuscularly) was more effective than family and supportive therapies in preventing psychotic and affective relapse as well as noncompliance. This effect, however,

was demonstrated only for those patients with schizophrenia who were living with their families. The personal therapy focused on the patient's characteristic response to stress and avoided symbolic interpretation or exploration of unconscious issues. The approach encouraged the patient to identify the cognitive, physiological, and affective experience of stress. Techniques drawn from social skills training encouraging the resumption of expected roles, and the provision of social and avoidance techniques were also part of the therapy.

Psychoeducational measures about the prodromes of psychosis, adaptive strategies, and successful rehabilitation were included. In all cases, selected principles of personal therapy were carefully tailored to the patient's individual needs and the stage of the patient's illness.

Personal therapy appeared to have other positive effects on the patients beyond the prevention of relapse. There were pervasive improvements in the area of social adjustment. While supportive therapy had some impact on adjustment, the improvement peaked at 12 months and leveled off thereafter. By contrast, patients receiving personal therapy continued to improve in social adjustment in the second and third years after discharge.

Major Depressive Disorder The differential effects of medication versus psychotherapy have been studied more extensively for depression than for other disorders. Whereas antidepressant medication seems to have its greatest effects on anhedonia, psychomotor retardation, sleep disturbances, appetite disturbances, delusions, and hallucinations, psychotherapy seems to act preferentially on interest, mood, social relationships, and vocational issues. Yet the landmark National Institute of Mental Health (NIMH)-Treatment of Depression Collaborative Research Project did not include a combined-treatment cell in its investigation of treatments for depression. The study clearly demonstrated that cognitive and interpersonal therapy are effective treatments for depression, but whether combined treatments had an advantage was not possible to discern from the study design. Fortunately, other studies have investigated the impact of combining psychotherapy and medication.

A study of interpersonal therapy (IPT) plus medication included comparison groups of medication alone and IPT alone. After 16 weeks, there was a clear advantage for combined treatments over either single modality. As predicted, the two approaches appeared to target different symptoms preferentially; medication seemed most useful for vegetative symptoms, and psychotherapy had a greater impact on interest and mood. Also, medication acted more rapidly than psychotherapy.

Studies of cognitive therapy have also shown some modest evidence of advantages for the combined therapy. Adding cognitive therapy to antidepressants appears to be more effective in preventing relapse than medication alone. Studies combining behavior therapy with antidepressants have methodological problems, but the combined treatment does appear to work more rapidly than behavior therapy alone. Combinations of psychodynamic therapy and antidepressants have not been tested in randomized, controlled trials.

Investigations of the combination of couples and family therapy with medication suggest that combined treatment produces greater improvements in the overall quality of the marital and family relationships while medication produces more rapid symptom change. Hence the combination enhances the breadth of the response, because both sets of target symptoms are improved when the combined modalities are part of the treatment plan.

Studies of the combination of medication and psychotherapy in the maintenance phase of major depressive disorder also show superiority of combined treatment. In a randomized controlled trial involving 107 elderly patients who had fully recovered from a major depressive episode, patients were assigned to one of four maintenance therapy conditions: (1) medication clinic with nortriptyline (Pamelor), (2) medication clinic with placebo, (3) monthly maintenance interpersonal therapy and nortriptyline, or (4) monthly maintenance interpersonal therapy with placebo. The recurrence rates over the 3-year period of the study were only 20 percent when nortriptyline and interpersonal therapy were combined, compared with 43 percent in the group who received nortriptyline and medication clinic visits and 64 percent in the group that received interpersonal therapy and placebo. In this elderly population there was a clinically significant advantage of combining medication and psychotherapy compared to either treatment alone, and the investigators concluded that combined treatment was the optimal clinical strategy in preventing relapse.

In elderly depressed patients, there is some research that allows the clinician to determine whether maintenance psychotherapy or antidepressant medication is indicated. Recovery of good subjective sleep quality during early continuation treatment is associated with the maintenance of a good remission with only monthly maintenance interpersonal psychotherapy. Patients who do not have such rapid recovery of good subjective sleep quality may be vulnerable to recurrence of major depressive episodes if the antidepressant medication is withdrawn.

In summary, psychotherapy and medication work on different target symptoms in patients with recurring major depressive disorder and therefore combine to enhance the overall breadth of response. Medication works more rapidly than psychotherapy and may provide more reliable relief from acute distress, but psychotherapeutic modalities enhance social functioning and appear to extend the relapse-free period. A 1997 “mega-analysis” comparing nonbipolar depressed patients treated with psychotherapy alone to those receiving combined treatment found that the advantage of the combination is particularly striking in more severe depressions that are recurrent.

Bipolar I Disorder Much less controlled research has been conducted on combined treatments of bipolar disorders, although there is a growing consensus that psychosocial interventions are essential in the treatment of the majority of patients. A German study looked at relapse rates in 20 cases of bipolar I disorder and 10 cases of schizoaffective disorder before and after treatment with systemic family therapy in conjunction with medication. The average duration of the treatment was 14.7 months with a range of 0 to 35 months, and the average number of sessions was 6.60 with a range of 1 to 19. The relapse rate was measured by the number of hospitalizations during the period of observation. Following family therapy, there was a 77.6 percent reduction of relapse in the total sample (67.8 percent for the patients with bipolar I disorder and 89.8 percent for patients with schizoaffective disorder).

This statistically significant reduction in relapse rate was accompanied by far lower rates of hospitalization. Before family therapy, only 1 of 30 cases required no hospitalization whatsoever; after family therapy, 14 of the 30 cases required no hospitalization. Only an average of 6.6 sessions of family therapy was required, so this intervention was also highly cost effective because of the reduction in hospital costs.

Several characteristics were noted in the families with improved relapse rates. Most significantly, the patients were no longer viewed by themselves or by their family members as victims of an illness beyond their control; family embers and patients felt empowered and gained a sense of active mastery over the illness.

Although this study did not utilize random assignment or control groups, the results nevertheless suggest that some family interventions that have proven so effective in prevention of relapse in schizophrenia may also be useful in the treatment of bipolar disorder. An American study at Cornell University Medical Center, utilizing both random assignment and a control group, has lent further support to the value of combining family therapy and medication. In this study 60 patients with a major mood disorder were randomly assigned to an inpatient family intervention plus medication in the context of standard hospital treatment, or to medication and standard hospital treatment without family intervention. Twelve patients who received the family therapy and 9 in the control group were diagnosed as having bipolar I disorder. At 18-month follow-up, significantly fewer patients from the family intervention group had been rehospitalized as compared with the group who did not receive the family treatment. Similarly, on measures of global outcome and work or primary role functioning, the patients who received the family intervention were doing significantly better than those who did not.

With the increasing awareness that lithium alone does not constitute effective prophylaxis for many patients with bipolar I disorder, individual psychotherapy has been added to increase occupational and social functioning, encourage compliance with lithium or other mood stabilizers, and cut through the denial so common in these patients. Many patients deny that their manic or hypomanic episodes are part of their illness, insisting that it is simply part of who they are; others will manifest a form of psychic discontinuity in which the manic “self” is split off from the euthymic “self,” as though the two are in no way connected. Some clinicians will use videotapes or audiotapes of the patient during a manic episode to help them integrate that aspect of themselves into the illness and overcome their denial. Patients with bipolar I disorder may also need assistance in mourning losses they have incurred through their erratic behavior during manic phases. Finally, individual psychotherapy may assist these patients in identifying the stressors that are instrumental in producing relapses. Research on patients with bipolar I disorder who are followed longitudinally and are systematically interviewed regarding life events reveals that those with the most frequent relapses are also subject to higher rates of major stressors.

At least one controlled study has demonstrated that the addition of individual cognitive therapy to lithium treatment improves compliance with the treatment and prevents relapse. Many patients have acknowledged that psychotherapy has been an important adjunct in their overall treatment.

Panic Disorder and Phobias Meta-analyses of the existing studies on the combination of behavior therapy and imipramine have consistently shown that greater improvement occurs when the treatments are combined than when either is used alone. Combined treatment shows superior efficacy for reducing specific phobia, social phobia, and functional impairment; it is roughly equivalent to medication in terms of reducing the frequency of panic attacks.

The advantage of combined treatments does not appear to hold when a combination of benzodiazepines and behavior therapy is compared to either modality alone. Some research suggests that high doses of benzodiazepines may undermine positive posttreatment outcomes with cognitive-behavior therapy, especially when the medication has been discontinued after the completion of the therapy. There may be some exceptions, however, when the combination may be useful. The addition of situational exposure therapy for patients with clinically significant agoraphobia being treated with benzodiazepines may be one such example; multimodal cognitive-behavior therapy may be useful when benzodiazepines are being withdrawn over time.

A Scandinavian study compared clomipramine (Anafranil) alone for patients with panic disorder with clomipramine plus 15 weekly sessions of dynamic psychotherapy. All patients in both treatment groups were panic-free within 26 weeks. On termination of clomipramine after 9 months, the relapse rate was significantly lower in the clomipramine-plus-psychotherapy group. The investigators concluded that dynamic therapy reduces psychosocial vulnerability associated with panic disorder.

In selected cases, couples or family therapy may be necessary in combination with medication or behavior therapy.

> A middle-aged housewife was virtually housebound because of her anxiety about having a panic attack in the shopping mall. Her husband had adapted to her disorder to a large extent, and he frequently performed routine tasks like grocery shopping for her. When she was successfully treated with exposure plus imipramine, she described deterioration in her marital relationship. The treating psychiatrist asked her to bring her husband into the sessions with her. It soon became apparent that her husband was highly ambivalent about her improvement because he was convinced that she would be attracted to another man when she went out shopping and have an extramarital affair. His jealous rage had led him to become extremely controlling of his wife at home, and she was wondering whether it would be better if she just returned to being housebound again.

This case vignette reflects how the marital couple may reach a stable equilibrium around an illness. Without attention to the marital issues that support and maintain the disorder, there is little likelihood that improvement will be lasting. One 3-year study of the marital interaction between 36 married agoraphobic women and their husbands found that 7 of the husbands reacted with abnormal jealousy when their wives responded to treatment. In each case the husband's reaction adversely influenced the treatment outcome. Even when the husband does not react with a jealous rage, there may be strong clinical reasons to involve the spouse in the behavioral treatment of a phobia. In one study of 66 agoraphobic patients, those in a home-based program that used the husband as a cotherapist fared better than patients whose husbands were not involved in the treatment. Including the husband encourages the couple to collaborate in strategies that improve the agoraphobic patient's ability to leave the home.

Obsessive-Compulsive Disorder Treatment of obsessive-compulsive disorder usually takes one of two directions. Treatment with selective serotonin reuptake inhibitors (SSRIs) often results in patients with obsessive-compulsive disorder making significant symptomatic improvement. While clomipramine is also useful, its adverse effects make SSRIs the preferred agents. In addition, about three quarters of patients with obsessive-compulsive disorder who comply with behavior therapy and who consciously apply the techniques show sustained improvement in symptoms.

Because the improvements effected by SSRIs are limited in terms of overall symptom reduction in patients who relapse rapidly when the agents are discontinued, there is a strong case to be made for combining behavior therapy and SSRIs. Some studies suggest both short-term and long-term improvements as well as more rapid response when both treatments are used. Behavior therapy also holds out the possibility that the medication can be discontinued without risking patient relapse.

Despite some suggestive data, however, the reports published so far on the advantages of combined medication and behavior therapy are not entirely convincing. Overall, combining treatments appears to improve the outcome as compared to medication alone, but this is not necessarily true for behavior therapy alone. Nevertheless, for behavior therapy to maintain robust results, the patient needs to invest time, energy, and money. Usually, daily therapy sessions are conducted for several weeks, after which these patients are expected to devote a good deal of time to exposure work. Hence it may not always be practical to provide behavior therapy at that intensity, and many patients may benefit with a less intense exposure accompanied by the use of SSRIs.

Some patients with obsessive-compulsive disorder become highly invested in their symptoms. Although psychodynamic therapy has not been shown to be effective in directly reducing the obsessions and rituals of obsessive-compulsive disorder, it may be useful in helping these patients to understand their resistance to receiving effective treatments like behavior therapy and SSRIs. Often the patient has involved the whole family system in a variety of cleaning rituals, and family therapy may be helpful as well in assisting family members to reduce their overinvolvement with the patient. When family members hear that they would be more helpful by refusing to collude with the patient's rituals and demands and by encouraging exposure (e.g., allowing the patient to touch doorknobs), they often feel a great sense of relief. Many feel obligated to collude with the patient, and only family intervention can give them permission to try other approaches.

Substance Dependence Both psychotherapy and methadone (Dolophine) have shown encouraging success with patients with opioid dependence. Data are also accumulating that the combination of the two approaches may achieve even better outcomes. In a randomized, controlled trial comparing treatments, opioid-dependent patients were assigned to one of three groups. One group had only methadone and virtually no psychotherapy. The second group had methadone plus meetings with a counselor that were oriented towards behavioral interventions. The third group had enhanced services involving the same dose of methadone and the same form of counseling, but also received additional resources, including a half-time employment counselor, a full-time psychiatrist, and a half-time family therapist.

Analysis of the results of this trial showed that the groups receiving psychotherapy had greater earning power, less welfare income, and strikingly lower rates of hospitalization as compared to the group that did not receive psychotherapy. In addition, a beneficial impact on costs could be inferred because of the reduced rate of hospitalization when psychotherapy was added. The investigators also noted that the stepwise, incremental value of the enhanced services over simple counseling in the second group indicated that family therapy, the presence of a psychiatrist, and employment counseling were highly useful interventions.

Personality Disorders In all of psychiatry, psychology and biology may come together most clearly in the realm of personality disorders. Personality appears to be composed of two broad components: *temperament*, which is constitutional and based on heritable traits, and *character*, which is primarily influenced by experiences in the environment. The seven-factor psychobiological model of personality developed by C. Robert Cloninger and his colleagues lends itself to treatment strategies that incorporate both medication and psychotherapy (Fig. 30.12-1).

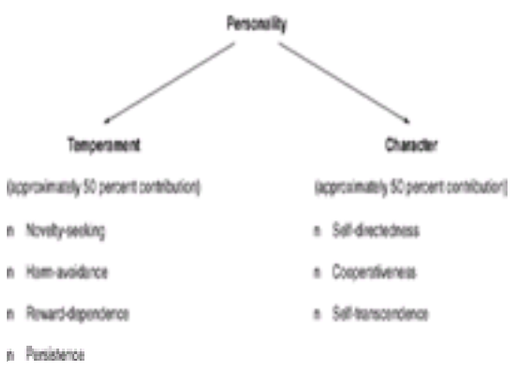

FIGURE 30.12-1 A psychobiological model of personality.

Temperament is composed of four heritable traits that are genetically homogeneous and inherited independently from one another. These include harm-avoidance, novelty-seeking, reward-dependence, and persistence. These traits are observable early in childhood and are relatively stable throughout the life cycle. These temperament traits account for approximately 50 percent of personality. The other half of personality is accounted for by character, within which three major character traits can be distinguished: self-directedness, cooperativeness, and self-transcendence. Character is heavily influenced by environmental experiences and matures in a stepwise manner throughout development. The DSM-IV personality disorders are all characterized by low ratings on self-directedness and cooperativeness. Self-transcendence does not appear to discriminate personality disorders from normal personality.

One clinical implication of the growing understanding of the relative contributions of biology and psychology to personality disorder is that psychopharmacological intervention may target temperament traits while psychotherapeutic approaches address character difficulties. It is an oversimplification to think of medication as the exclusive treatment for Axis I disorders and psychotherapy as the exclusive treatment for Axis II disorders. Even in normal personality functioning, serotonergic function appears to modulate aspects of affective experience and social behavior. In a placebo-controlled study of normal volunteers, paroxetine (Paxil) was administered for 4 weeks, and changes in personality variables were studied. Compared to placebo, paroxetine decreased negative affect and increased social affiliation.

Although no systematic studies have been reported on the combination of psychotherapy and medication compared to either treatment alone in patients with personality disorders, recent investigations of the role of psychopharmacological intervention on personality disorder patients suggest that the effect of such agents on temperament may facilitate psychotherapeutic work with character. Double-blind controlled studies have demonstrated that monoamine oxidase inhibitors (MAOIs) may improve atypical depression in patients with borderline personality disorder, particularly if symptoms such as hyperphagia and hypersomnia are present. Carbamazepine (Tegretol) has also been shown to decrease behavioral dyscontrol in patients with borderline personality disorder. Brief psychotic episodes in this patient group have been effectively treated by low-dose antipsychotic medication. However, much recent attention has been directed to the role of SSRIs. One study using fluoxetine (Prozac) showed that patients with borderline personality disorder showed improvement in depression, anxiety, paranoia, interpersonal sensitivity, obsessionality, hostility, and global functioning compared to a control group who received placebo. However, higher daily dosages of fluoxetine, averaging 80 mg, were necessary to produce these changes.

A second controlled study using fluoxetine found that intense anger of patients with borderline personality disorder can be significantly reduced through the use of this medication. The reduction of hostility and anger does not appear to be confined only to patients with borderline personality disorder. In a third study, patients with a variety of personality disorders were treated with an average daily dosage of 20 mg to 40 mg of fluoxetine. Compared to the control group on placebo, these patients showed statistically significant reductions in impulsive aggressive behavior. The medication seemed to be particularly useful for verbal aggression and aggression against objects rather than on assaultive behavior.

Hence the rationale for combining medication and psychotherapy is in part the fact that negative affects can be reduced by the administration of agents such as SSRIs. When rage and aggression are lowered, the patient may be more accessible to psychotherapeutic interventions. The decreased intensity of affect may allow the patient to reflect on and contemplate internal states and problems that occur in relationships.

Aggression is not the only focus of pharmacotherapy. Indeed, there are four general symptom domains that may benefit from targeted psychopharmacological intervention in personality disorders: (1) aggression and behavioral dyscontrol, (2) affective instability and mood symptoms, (3) anxiety, and (4) psychotic symptoms and other cognitive perceptual distortions. The empirical research suggesting that several different agents are of benefit with patients who suffer from borderline personality disorder reflects the fact that these different domains can be targeted by different agents.

Several types of psychotherapy have been effective in treating the various personality disorders. A randomized controlled trial of dialectical behavior therapy with borderline personality disorder found that substantial improvements in suicidal and parasuicidal behavior as well as reductions in the need for hospitalization are possible with 1 year of that modality. Similarly, longitudinal studies of dynamic psychotherapy suggest that improvements in personality disorders can be expected with long-term dynamic psychotherapy. While clinical trials involving combined therapies are not yet available to provide definitive evidence, a strong argument can be made for the combination in clinical practice based on the independent findings of psychopharmacological and psychotherapeutic studies.

Eating Disorders For many years the typical treatment for bulimia nervosa has been a combination of psychotherapy and an antidepressant. A 1997 study using a randomized, controlled design has demonstrated that combined treatment is better than either modality alone. A total of 120 women with bulimia nervosa were assigned to one of five conditions: (1) cognitive-behavior therapy plus placebo, (2) cognitive-behavior therapy plus an antidepressant, (3) dynamically oriented supportive therapy plus placebo, (4) dynamically oriented supportive therapy plus an antidepressant, and (5) an antidepressant alone. There was a two-stage medication intervention in which a second antidepressant, fluoxetine, was employed if the first antidepressant, desipramine, was either ineffective or poorly tolerated.

This study yielded a number of significant results. First, cognitive-behavior therapy was found to be superior to supportive therapy in reducing the behavioral symptoms of bulimia nervosa such as vomiting and bingeing. Patients who received medication in combination with psychotherapy experienced greater improvement in binge eating and depression than did patients receiving placebo and psychological treatment. Cognitive-behavior therapy plus an antidepressant was also superior to medication alone, but supportive therapy plus the antidepressant was not. The investigators concluded that cognitive-behavior therapy was the treatment of choice for patients with bulimia nervosa but that the two-stage antidepressant intervention added modestly to the benefit of the psychotherapy.

CLINICAL CONSIDERATIONS

In considering various combinations of pharmacotherapy and psychotherapy, there are two models commonly used in clinical practice today. The one-person model involves a psychiatrist conducting the psychotherapy and prescribing medication for the same patient at the same time. The two-person model divides the functions so that the clinician conducting the psychotherapy and the physician prescribing the medication are two separate individuals. Practical considerations determine which model is used. If a psychotherapist does not feel competent in prescribing certain medications, a psychopharmacology consultant must be involved. Conversely, if a psychopharmacologist does not feel sufficiently skilled to provide the psychotherapy, the patient must be referred to a separate psychotherapist. For financial reasons, a managed-care company may stipulate that a psychiatrist can only see the patient for 15 minutes to prescribe medications and may assign the psychotherapy to a less expensive nonmedical therapist. Another common situation occurs when psychotherapy begins with a nonmedical therapist and the patient becomes depressed; a consultation with a psychopharmacologist is then requested. If a psychiatrist is the primary treater and is competent at both treatment modalities, there are several factors that must be taken into account in deciding which approach to recommend.

One-Person Model Psychiatrists who decide to employ the one-person model must accustom themselves to thinking about both mind and brain in their regular sessions with the patient. Just like the physicist who simultaneously thinks in terms of particles and waves, the psychiatrist must think about a dysfunctional brain and a distressed mind as part of a unified whole. In practice this may require a flexible shifting back and forth between an empathic, introspective subjective approach versus a more objective descriptive approach. While psychotherapy often encourages verbalization as opposed to action, prescribing medication requires clear action on the part of the clinician (and the patient). Moreover, the clinician must ask direct questions designed to elicit information about symptoms and adverse effects rather than the more open-ended approach of allowing the patient to set the agenda in a dynamic therapy process.

Also, whereas a dynamic therapist may not assume an authoritarian posture toward the patient, this position may shift when the clinician discusses medication and cites the literature on such things as maintenance dosages and prevention of relapse. Both patient and clinician alike may need to shift mental sets in the course of a session.

One strategy for managing these difficulties is to set aside a few minutes at the beginning or the end of each session to review how the medication is affecting the patient's symptoms and to write a prescription, if necessary. During the remainder of the session, the medication may be discussed more from the standpoint of its meaning to the patient. Indeed, one of the advantages of this arrangement is that the prescribing and the psychotherapy do not get split off from one another in a way that fragments the treatment. All transferences and resistances are dealt with by one clinician. Compliance problems may be readily linked to specific transference paradigms because of the clinician's intimate knowledge of the patient's internal object relations.

Two-Person Model Surveys have shown that approximately 65 percent of psychiatrists have prescribed medication to patients who are in psychotherapy with other clinicians. In some cases two clinicians have to be involved because one is a nonmedical therapist. In other cases, however, psychiatrists who are conducting psychotherapy do not feel comfortable with the kind of bimodal thinking necessary to prescribe the medication as well and they may prefer to keep the transference uncontaminated by feelings about the medication; they may also not feel qualified to prescribe because they have not kept up with psychopharmacology data.

Although the two-person model may be effective for some patients, it also runs the risk of serving as a nidus for splitting, particularly but not exclusively with patients with borderline personality disorder.

A 27-year-old woman with a diagnosis of borderline personality disorder had been seeing a psychotherapist once a week for 1 year. She and her therapist had felt that they were at an impasse, and the therapist suggested that she consult a local psychopharmacologist. She saw the psychopharmacologist for 30 minutes, and he diagnosed her as having a major depression. He explained all about the illness to her, and he prescribed fluoxetine.

The patient felt an almost immediate positive response to the medication, and at her next psychotherapy appointment, she blasted her psychotherapist with a long tirade about her inadequacies: “You sit there for 1 year, just listening and trying to understand me. All this time I had a depression that needed drug treatment. Dr. A (her psychopharmacologist) was so helpful. He took my complaints seriously, and for God's sake, he took some action! All you do is talk, talk, talk. Why didn't it occur to you sooner that something needed to be done?”

In this brief vignette the psychopharmacologist becomes the idealized object who was responsive to the patient's needs, in contrast to the psychotherapist, who is seen as the bad object, almost sadistic in her perceived refusal to “do something.” The individual prescribing the medication may contribute to the split by implying either directly or indirectly that the psychotherapist has been derelict in her duty by not sending the patient for medication sooner.

The cleavage may occur along opposite lines as well. Some patients with borderline personality disorder feel that the prescribing psychiatrist is eager to get rid of them because the appointment is scheduled for only 15 minutes. They may object to being “thrown out” of the doctor's office without the opportunity to talk more about what they are experiencing. In this case, the prescriber becomes the bad object, while the therapist is idealized for taking the time to listen and express concern about the patient's internal experience.

Splitting of this nature cannot be entirely prevented, but there are steps that can be taken to minimize the potential destructiveness of this form of behavior. From the outset the patient should understand that the psychotherapist and the prescribing psychiatrist must be given permission to communicate about diagnosis and treatment. They should consider themselves as part of a treatment team rather than isolated individuals. Discussing the patient's perceptions about the other clinician openly and honestly reduces a great deal of the tension created by such splits. In addition, the two clinicians should agree on which one of them is responsible for making decisions about hospitalization, vacation coverage, changes in medication, and the investigation of any potential medical problems.

Although communication between the two is an ideal arrangement with which few would disagree, studies suggest that frequently such communication is rather minimal. In some two-person treatments, the contacts between the two clinicians may be limited to an as-needed basis, which in practice means that almost no communication occurs unless there is a major crisis. In other situations prescribing psychiatrists also supervise the psychotherapy of patients for whom they prescribe medication. Physicians who prescribe for patients whose therapy they are supervising should recognize that they are potentially liable for the acts of those hierarchically below them, a situation referred to as *vicarious liability*. They can also be directly liable for failure to adequately supervise the psychotherapist.

Although it is rarely done in practice, some forensic experts have suggested that a written three-way agreement among psychotherapist, pharmacotherapist, and patient should be drawn up and signed at the beginning of such an arrangement. The details of the agreement would include the purpose of each treatment, the relative roles of each of the clinicians, policies for communication between the clinicians, any supervisory responsibilities entailed in the arrangement, and who should be called in an emergency.

The frequency and duration of the communication between clinicians cannot be arbitrarily established, but it certainly must take place often enough so that each clinician has a clear understanding of the treatment approach of the other. Often a critical factor in the evolution of splitting is that the patient distorts to one clinician what the other is saying or doing. If the clinician hearing this report simply takes it at face value without calling and checking out the veracity of the account with the other clinician, the situation may rapidly deteriorate. The clinician hearing the report may subsequently collude with the patient's feeling of being victimized by the other clinician. When communication finally occurs, the colluding clinician may accuse the colleague, which results in greater defensiveness. A tactful inquiry about whether the patient's account is accurate is generally much more productive.

Other Clinical Issues When pharmacotherapy is added to dynamic psychotherapy or psychoanalysis, several responses are possible. Patients often respond by assuming that the analyst or therapist believes all change should come about through psychological interventions, so that a particular patient may view the addition of medication as something of a failure. In one report, a patient was convinced that his analyst regarded the positive effects of taking medication as primarily psychological. This patient, on the other hand, thought 90 percent of the effect was pharmacological. Turning treatment over to medication in the context of a well-established psychotherapeutic process will mean different things to individual patients, and the clinician conducting the analysis or therapy should be alert to these interpretations.

Patients who are in both treatments are often in a quandary about the relative roles of medication and psychotherapy. Are their problems caused by an alteration in brain chemistry for which they have no responsibility? Or are their feelings of incapacitation and disability secondary to psychological issues that they should be able to master? These questions may blend into concerns about identity and self. Patients may wonder who they actually are after their chronic depression improves substantially with medication. They may need psychotherapy to adjust to a new identity based on being euthymic.

COST EFFECTIVENESS

Many managed-care companies insist that psychiatrists do not need to conduct psychotherapy because nonmedical therapists can do just as well at a cheaper price. Hence they allow the psychiatrist to see the patient for a 15-minute medication check while referring the psychotherapy to someone whose time is much less expensive, often a counselor with a Bachelor's degree who is really not well trained in psychotherapy. While this may appear on the surface to be more cost effective, the treatment may become fragmented to the point where the patient's condition deteriorates and hospitalization is necessary. In that event, the hospital costs far outweigh the added expense of one clinician serving both functions. Moreover, the division of labor between two clinicians requires extremely close collaboration. The time the two clinicians spend communicating with each other is often a hidden cost, one not generally reimbursed, that must ultimately be factored into a comprehensive consideration of cost effectiveness. Meanwhile, data need to be collected on the two alternative arrangements to systematically study whether one is ultimately more cost effective than the other.

In the absence of definitive data, clinical experience suggests that in a number of situations, there are clear advantages to a psychiatrist performing both roles. A few examples include the following: (1) for patients with schizophrenia and other psychotic disorders who are not compliant with prescribed medication; (2) patients with bipolar I disorder who deny illness and do not cooperate with the treatment plan; (3) patients with serious or unstable medical conditions, where the psychiatrist's medical knowledge is important in the overall management; (4) patients with severe personality disorders, such as borderline personality disorder, where the treatment can be potential for being disrupted by splitting maneuvers; (5) impulsive and severely suicidal patients who are likely to require hospitalization in the course of outpatient treatment; (6) patients with eating disorders who present complicated management problems; and (7) patients who present a clinical picture in which the need for medication is unclear, and an ongoing assessment of that need is an integral part of the treatment (Table 30.12-1). The psychiatrist is in a position to evaluate the need for medication by following the patient over time. Just as a surgeon is trained to know when to operate and when to refrain from surgery, the psychiatrist is trained to assess when it is best not to prescribe medication. In all these clinical situations one central clinician who works with the patient and possibly the family on both medication and psychological issues may ensure an overall integration of the treatment plan that may lead to better compliance.

- Patients with schizophrenia and other psychotic disorders who are not compliant with prescribed medication.
- Patients with bipolar I disorder who deny illness and do not cooperate with the treatment plan.
- Patients with serious or unstable medical conditions.
- Patients with severe personality disorders.
- Impulsive and severely suicidal patients who are likely to require hospitalization.
- Patients with eating disorders who present complicated management problems.
- Patients who present a clinical picture in which the need for medication is unclear, thus requiring ongoing assessment.

Table 30.12-1 Clinical Situations Where It Is Especially Advantageous for One Psychiatrist to Provide Both Medication and Psychotherapy.

In many discussions of the value of combined psychotherapy and pharmacotherapy, dissenters argue that the combination of the two treatments may be unduly expensive and without benefits commensurate with the additional cost. This perspective has led some to suggest innovative means of gaining benefit from both modalities. For example, for agoraphobic patients, some have suggested that the prescription of imipramine plus the provision of instructions for systematic self-directed exposure may be a cost-efficient way to derive benefits from both treatments. Similar efforts have been made to assist obsessive-compulsive disorder patients with self-administered exposure in vivo practice.

Generalizations about the expense of combined treatments must be made with caution, however. In a well-designed study of the delivery of treatment for depression in primary-care settings, the Rand Corporation found that in the general practitioner's office, where the average appointment lasted 8 minutes, the diagnosis of depression was frequently missed. Also, when medication was prescribed, either benzodiazepines or suboptimal dosage of antidepressants were frequently prescribed. When mental health professionals became involved with the patients, they frequently delivered a combination of psychotherapy and appropriate dosages of antidepressant medication. The specialist treatment was somewhat more expensive to deliver, but the small increase in money quadrupled the effectiveness of the overall treatment in terms of vastly improved functional outcomes for the depressed patients. The investigators pointed out that too often the emphasis is placed only on cost rather than being placed equally on effectiveness. The concept of cost effectiveness should not be synonymous with cheap rather, it should be viewed as high value.

The combination of antipsychotic medication and family therapy in the treatment of schizophrenia prevents relapse and avoids the need for further inpatient care to the extent that the additional cost of the combined treatments is more than offset by the savings in fewer inpatient days. Similarly, in patients with bipolar and schizoaffective disorder, an average of six family therapy sessions significantly reduced the need for hospitalization and thus more than paid for itself. When cost effectiveness is considered from the standpoint of all costs, including the direct costs of treatment and the indirect costs of work disability and absenteeism, the combined use of pharmacotherapy and psychotherapy is often the most cost-efficient intervention available.

SUGGESTED CROSS-REFERENCES

Schizophrenia is covered in Chapter 12, mood disorders in Chapter 14, and anxiety disorders in Chapter 15. The other sections of Chapter 30 cover various schools and approaches of psychotherapy. Chapter 31 extensively covers psychopharmacology.

SECTION REFERENCES

American Psychiatric Association: Practice guideline for major depressive disorder in adults. Am J Psychiatry *150*(Suppl): 1993.

Baxter KR, Schwartz JM, Bergman KS, Szuba MP, Guze BH, Mazziotta JC, Alazraki A, Selin CE, Ferng H-K, Munford P, Phelps ME: Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. Arch Gen Psychiatry *49:*618, 1992.

Beitman BD: Pharmacotherapy and the stages of psychotherapeutic change. In *Review of Psychiatry*, vol 12, JM Oldham, MB Riba, A Tasman, editors. American Psychiatric Press, Washington, DC, 1993.

*Cloninger CR, Svrakic DM, Pryzbeck TR: A psychobiological model of temperament and character. Arch Gen Psychiatry *50:*975, 1993.

Coccaro EF, Kavoussi RJ: Fluoxetine and impulsive aggressive behavior in personality-disordered subjects. Arch Gen Psychiatry *54:*1081, 1997.

Cowdry RW, Gardner DL: Pharmacotherapy of borderline personality disorder: Alprazolam, carbamazepine, trifluoperazine, and tranylcypromine. Arch Gen Psychiatry *45:*111, 1988.

Elkin I, Shea T, Watkins JT, Imber SD, Collins JF, Glass DR, Pilkonis PA, Leber WR, Docherty JP, Fiester SJ, Parloff MB: National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. Arch Gen Psychiatry *46:*971, 1989.

Falloon IRH, Boyd JLJ, McGill CW, Williamson M, Razani J, Moss HB, Gilderman AM, Simpson GM: Family management in the prevention of morbidity of schizophrenia: Clinical outcome of a two-year longitudinal study. Arch Gen Psychiatry *42:*887, 1985.

Fawzy FI, Fawzy NW, Hyun CS, Elashoff R, Guthrie D, Fahey JL, Morton DL: Malignant melanoma: Effects of an early structured psychiatric intervention, coping, and affective state on recurrence and survival 6 years later. Arch Gen Psychiatry *50:*681, 1993.

Hogarty GE, Cornblith SJ, Greenwald D, DiBarry AL, Cooley S, Ulrich RF, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, I: Description of study and effects of relapse rates. Am J Psychiatry *154:*1504, 1997.

Hammen C, Gitlin M: Stress reactivity in bipolar patients and its relation to prior history of disorder. Am J Psychiatry *154:*856, 1997.

Hogarty GE, Greenwald D, Ulrich RF, Cornblith SJ, DiBarry AL, Cooley S, Carter M, Flesher S: Three-year trials of personal therapy among schizophrenic patients living with or independent of family, II: Effects on adjustment of patients. Am J Psychiatry *154:*1514, 1997.

Gabbard GO: Mind and brain in psychiatric treatment. Bull Menninger Clin *58:*427, 1994.

Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice: The DSM-IV Edition*. American Psychiatric Press, Washington, DC, 1994.

*Gabbard GO, Goodwin FK: Integrating biological and psychosocial perspectives. In *Review of Psychiatry*, vol 15, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1996.

Gabbard GO, Lazar SG, Hornberger J, Spiegel D: The economic impact of psychotherapy: A review. Am J Psychiatry *154:*147, 1997.

Glick ID, Clarkin JF, Goldsmith SJ: Combining medications with family psychotherapy. In *Review of Psychiatry*, vol 12, JM Oldham, MB Riba, A Tasman, editors. American Psychiatric Press, Washington, DC, 1993.

Greist JH, Jefferson JW: Obsessive-compulsive disorder. In *Treatments of Psychiatric Disorders: The Second Edition*, vol 2, GO Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

Gunderson J, Links P: Borderline personality disorder. In *Treatments of Psychiatric Disorders: The Second Edition*, vol 2, GO Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

Gutheil TG: The psychology of psychopharmacology. Bull Menninger Clin *46:*321, 1982.

*Hogarty GE Anderson CM, Reiss DJ, Kornblith SJ, Greenwald DP, Ulrich RF, Carter M: The Environmental-Personal Indicators in the Course of Schizophrenia (EPICS) Research Group: Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia, II: Two-year effects of a controlled study on relapse and adjustment. Arch Gen Psychiatry *48:*340, 1991.

Hollon SD, Fawcett J: Combined medication and psychotherapy. In *Treatments of Psychiatric Disorders: The Second Edition*, vol 2, GO Gabbard, editor. American Psychiatric Press, Washington, DC, 1995.

Knutson B, Wolkowitz OM, Cole SW, Chan T, Moore EA, Johnson RC, Terpstra J, Turner RA, Reus VI: Selective alteration of personality and social behavior by serotonergic intervention. Am J Psychiatry *155:*373, 1998.

Kopta M, Lueger R, Sanders S, Howard K: Individual psychotherapy outcome and process research: Challenges leading to greater turmoil or a positive transition? Ann Rev Psychol *50:*441, 1999.

*Krupnick JL, Stosky SM, Simmens S, Moyer J, Watkins J, Elkin I, Pilkonis PA: The role of therapeutic alliance in psychotherapy and pharmacotherapy outcome: Findings in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. J Consult Clin Psychol *64:*532, 1996.

Leff JP, Kuipers L, Berkowitz R, Sturgeon D: A controlled trial of social intervention in the families of schizophrenic patients: A two-year follow-up and issues in treatment. Br J Psychiatry *146:*594, 1985.

Linehan MM, Armstrong HE, Suarez A, Allmon D, Heard HL: Cognitive-behavioral treatment of chronically parasuicidal borderline patients. Arch Gen Psychiatry *48:*1060, 1991.

Luborsky L, Diguer L, Luborsky E, Singer B, Dickter D, Schmidt KA: The efficacy of dynamic psychotherapies: Is it true that “everyone has won and all must have prizes”? In *Psychodynamic Treatment Research: A Handbook for Clinical Practice*, NE Miller, L Luborsky, JP Barber, JP Docherty, editors. Basic Books, New York, 1993.

Mavissakalian MR: Combined behavioral and pharmacological treatment of anxiety disorders. In *Review of Psychiatry*, vol 12, JM Oldham, MB Riba, A Tasman, editors. American Psychiatric Press, Washington, DC, 1993.

McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP: The effects of psychosocial services in substance abuse treatment. JAMA *269:*1953, 1993.

Retzer A, Simon FG, Webber G, Stierlin H, Schmidt G: A follow-up study of manic-depressive and schizoaffective psychoses after systemic family therapy. Family Process *30:*139, 1991.

Reynolds CF, Frank E, Houck PR, Mazumdar S, Dew MA, Cornes C, Buysse DJ, Begley A, Kupfer DJ: Which elderly patients with remitted depression remain well with continued interpersonal psychotherapy after discontinuation of antidepressant medication? Am J Psychiatry *154:*958, 1997.

*Reynolds CF, Frank E, Perel JM, Imber SD, Cornes C, Miller MD, Mazumdar S, Houck PR, Dew MA, Stack JA, Pollock BG, Kupfer DJ: Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: A randomized controlled trial in patients older than 59 years. JAMA *281:*39, 1999.

Spiegel D, Bloom J, Kraemer HC, Gottheil E: Effect of psychosocial treatment on survival of patients with metastatic breast cancer. Lancet *2:*888, 1989.

Spiegel DA, Bruce TJ: Benzodiazepines and exposure-based cognitive-behavior therapies for panic disorder: Conclusions from combined treatment trials. Am J Psychiatry *154:*773, 1997.

Stevenson J, Meares R: An outcome study of psychotherapy for patients with borderline personality disorder. Am J Psychiatry *149:*358, 1992.

Thase ME, Greenhouse JB, Frank E, Reynolds CF III, Pilkonis PA, Hurley K, Grochocinski V, Kupfer DJ: Treatment of major depression with psychotherapy or psychotherapy-pharmacotherapy combinations. Arch Gen Psychiatry *54:*1009, 1997.

Walsh BT Wilson GT, Loeb KL, Devlin MJ, Pike KM, Roose SP, Fleiss J, Waternaux C: Medication and psychotherapy in the treatment of bulimia nervosa. Am J Psychiatry *154:*523, 1997.

Wiborg IM, Dahl AA: Does brief dynamic psychotherapy reduce the relapse rate of panic disorder? Arch Gen Psychiatry *53:*689, 1996.

Woodward B, Duckworth KS, Gutheil TG: The pharmacotherapist-psychotherapist collaboration. In *Review of Psychiatry*, vol 12, JM Oldham, MB Riba, A Tasman, editors. American Psychiatric Press, Washington, DC, 1993.

Zhang M, Wang M, Li J, Phillips M: Randomized controlled trial of family interventions for 78 first-episode male schizophrenic patients: An 18-month study in Suzhou, Jiangsu. Br J Psychiatry *165*(Suppl):96, 1994.

CHAPTER 31. BIOLOGICAL THERAPIES

31.1 GENERAL PRINCIPLES OF PSYCHOPHARMACOLOGY

JACK A. GREBB, M.D.

The use of drugs to treat psychiatric disorders is often the foundation for a successful treatment approach that can also include other types of interventions such as psychotherapy or behavioral therapies. As knowledge about the biology of normal and abnormal brain function continues to grow, the practice of clinical psychopharmacology continues to evolve in scope and effectiveness. Those involved in the prescribing and clinical follow-up of psychiatric drug treatments must remain current with the research literature, including the emergence of new agents, the demonstration of new indications for existing agents, and the identification and treatment of drug-related adverse effects. The emergence of new drugs and new indications is one of the most exciting areas of psychiatry.

The practice of pharmacotherapy in psychiatry should not be oversimplified—for example, it should not be reduced to a one-diagnosis-one-drug approach. Many variables affect the practice of psychopharmacology, including drug selection and administration; the psychodynamic meaning to the patient; and family and environmental influences. Some patients may view drug treatment as a panacea and others may view it as an assault. The patient, the patient's relatives, and the nursing staff must be instructed on the reasons for the drug treatment as well as the expected benefits and potential risks. In addition, the clinician may also find it useful to explain the theoretical basis for pharmacotherapy to the patient and other involved parties.

Drugs must be used in effective dosages for sufficient periods, as determined by previous clinical investigations and clinical experience. Subtherapeutic doses and incomplete therapeutic trials should not be given to a patient simply because the psychiatrist is excessively concerned that the patient will develop adverse effects. The use of dosages that are too low or durations that are too short merely exposes the patient to some risk, without providing the patient the maximum chance of therapeutic benefit. Treatment response and the emergence of adverse effects must be monitored closely; drug dosage should be adjusted accordingly, and appropriate treatments for emergent adverse effects must be instituted as quickly as possible.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics is the study of how the body handles a drug; *pharmacodynamics* is the study of the effects of the drug on the body. Pharmacokinetic drug interactions are the effects of drugs on the plasma concentrations of each other, and pharmacodynamic drug interactions are the effects of drugs on the biological activities of each other.

Pharmacodynamics The major pharmacodynamic considerations include receptor mechanisms; the dose-response curve; the therapeutic index; and the development of tolerance, dependence, and withdrawal phenomena. The receptor for a drug can be defined generally as the cellular component to which the drug binds and through which the drug initiates its pharmacodynamic effects on the body. A drug can be an agonist for a receptor, thus stimulating the specific biological activity of the receptor; or an antagonist, thus inhibiting the biological activity. Some drugs are classified as partial agonists because they are not capable of fully activating a specific receptor. The receptor site for a psychopharmacological drug is often the receptor for an endogenous neurotransmitter. For example, most antipsychotic medications are receptor antagonists at the dopamine type 2 (D_2) receptor. However, this may not be true for other psychotherapeutic drugs (e.g., lithium [Eskalith], which may act by inhibiting the enzyme inositol-1-phosphatase).

The dose-response curve plots the drug concentration against the effects of the drug (Fig. 31.1-1). The potency of a drug refers to the relative dose required to achieve certain effects. Haloperidol (Haldol), for example, is more potent than chlorpromazine (Thorazine) because approximately 5 mg of haloperidol is required to achieve the same therapeutic effect as 100 mg of chlorpromazine. However, both these drugs are equal in their clinical efficacy—that is, the maximum clinical response achievable by administration of a drug.

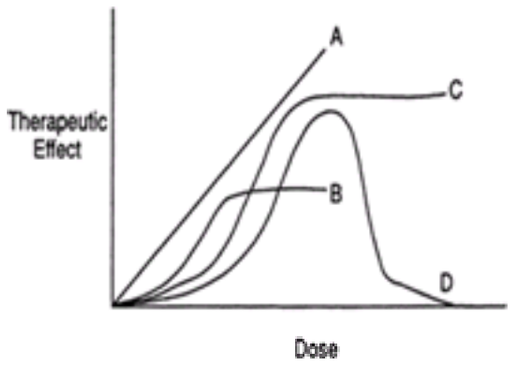

FIGURE 31.1-1 Dose-response curves plotting the therapeutic effect as a function of increasing the dose, often calculated as the log of the dose. Drug A has a linear dose response, drugs B and C have sigmoidal curves, and drug D has a curvilinear dose-response curve. Although doses of drug B are more potent than are equal doses of drug C, drug C has a higher maximum efficacy than does drug B. Drug D has a therapeutic window, such that both low and high doses are less effective than are midrange doses.

The adverse effects of most drugs are often a direct result of their primary pharmacodynamic effects. *Therapeutic index* is a relative measure of the toxicity or safety of a drug and is defined as the ratio of the median toxic dose to the median effective dose. The *median toxic dose* is the dose at which 50 percent of patients experience a specific toxic effect, and the *median effective dose* is the dose at which 50 percent of patients have a specified therapeutic effect. The therapeutic index for haloperidol and cardiovascular effects is quite high, as evidenced by the wide range of dosages in which haloperidol is prescribed. Conversely, the therapeutic index for lithium is quite low, thus requiring careful monitoring of serum lithium levels in patients for whom the drug is prescribed. Both interindividual and intraindividual variations can affect the response to a specific drug. An individual patient may be hyporeactive, normally reactive, or hyperreactive to a drug. For example, some patients require 150 mg a day of imipramine (Tofranil) whereas others may require 300 mg a day. Idiosyncratic drug responses occur when a patient experiences a particularly unusual or rare effect from a drug. For example, some patients become quite agitated when given a benzodiazepine, such as diazepam (Valium).

A person may become less responsive to a particular drug as it is administered over time, a process that is referred to as the development of tolerance. The development of tolerance may sometimes be associated with the appearance of physical dependence on the drug, which can be seen with drugs such as opioids or benzodiazepines. A patient who has developed physical dependence to a drug will experience withdrawal symptoms if the drug is discontinued. Withdrawal symptoms can be mentally and physically quite uncomfortable, and in some cases, even fatal.

CLINICAL GUIDELINES

The practice of clinical psychopharmacology requires skill as both a diagnostician and as a psychotherapist, knowledge of the available drugs, and the ability to plan a pharmacotherapeutic regimen. The selection of a drug and the initiation of treatment should be based on the patient's history, the patient's current clinical state, and the treatment plan. The treating clinician should know the purpose or the goal of a drug trial, the length of time that the drug needs to be administered to assess its efficacy, the approach to be taken to reduce any adverse effects that may occur, alternative drug strategies to be used should the current one fail, and whether long-term maintenance of the patient on the drug is indicated. In almost all cases the treating clinician should explain the treatment plan to the patient and often to the family and other caretakers. The patient's reaction to and ideas about a proposed drug trial should be considered; however, if the clinician believes that accommodating the patient's wishes would hinder treatment, this should also be clearly explained to the patient.

Choice of Drug The first two steps in selecting a specific drug for the treatment of a patient are diagnosis and the identification of target symptoms. Although this process is often best accomplished when the patient is not receiving any drug treatments, it is often not possible to evaluate patients in a drug-free state because of severity of symptoms or insurance coverage that strictly limits the amount of time a patient can be hospitalized. If a patient presents to the clinician and is receiving a suboptimal dosage of an otherwise appropriate drug, the clinician may decide to continue the current drug at a more appropriate dosage. In many cases, however, deciding how effective the drugs that the patient is currently taking are is a very difficult clinical decision. In such cases it is often advisable to taper off and stop the current drugs and then initiate new therapy that can be evaluated more accurately and completely.

From among the drugs appropriate to a particular diagnosis, the specific drug should be selected according to the patient's history of drug response (compliance, therapeutic response, adverse effects), the patient's family history of drug response, the profile of adverse effects for that drug with regard to the particular patient, and the prescribing clinician's usual practice. If a drug has previously been effective in treating a patient or a family member, the same drug should be used again unless there is some specific reason not to do so. A history of severe adverse effects from a specific drug is a strong indicator that the patient would not be compliant with that particular drug. It is unfortunate that patients and their families are often quite ignorant about which drugs have been used before, in what dosages, and for how long. This ignorance may reflect the tendency of clinicians not to explain drug trials to their patients, and the low frequency of the practice of giving patients and their caregivers a written record of drug treatments for their personal medical records. One caveat to obtaining a past history of drug response is that because of their mental disorders patients may report the effects of a previous drug trial inaccurately. If possible, patients' medical records should be obtained to confirm their reports. Most psychotherapeutic drugs of a single class have been demonstrated to be equally efficacious; however, the drugs do differ in their adverse effects on individual patients. A drug should be selected that minimally exacerbates any pre-existing medical problems that a particular patient may have.

Nonapproved Dosages and Uses Under the federal Food, Drug, and Cosmetic (FDC) Act, the Food and Drug Administration (FDA) has the authority to control the initial availability of a drug by approving only those drugs that demonstrate both safety and effectiveness, and then to ensure that the drug's proposed labeling is truthful and contains all information pertinent to the safe and effective use of that drug. An additional level of government regulation is directed by the Drug Enforcement Agency (DEA), which has classified drugs according to their abuse potential (Table 31.1-1), and clinicians are advised to exercise increased caution when prescribing controlled substances (control level II-IV).

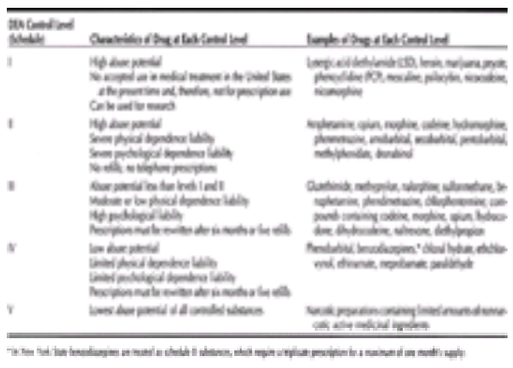

DEA Control Level (Schedule)	Characteristics of Drug at Each Control Level	Examples of Drugs at Each Control Level
I	High abuse potential No accepted use in medical treatment in the United States at the present time and, therefore, not for prescription use Can be used for research	Lysergic acid diethylamide (LSD), heroin, marijuana, peyote, phencyclidine (PCP), mescaline, psilocybin, nicocodeine, nicomorphine
II	High abuse potential Severe physical dependence liability Severe psychological dependence liability No refills, no telephone prescriptions	Amphetamine, opium, morphine, codeine, hydromorphine, phenmetrazine, amobarbital, secobarbital, pentobarbital, methylphenidate, dronabinol
III	Abuse potential less than levels I and II Moderate or low physical dependence liability High psychological liability Prescriptions must be rewritten after six months or five refills	Glutethimide, methyprylon, nalorphine, sulfomethane, benzphetamine, phendimetrazine, chlorphentermine, compounds containing codeine, morphine, opium, hydrocodone, dihydrocodeine, naltrexone, diethylpropion
IV	Low abuse potential Limited physical dependence liability Limited psychological dependence liability Prescriptions must be rewritten after six months or five refills	Phenobarbital, benzodiazepines,* chloral hydrate, ethchlorvynol, ethinamate, meprobamate, paraldehyde
V	Lowest abuse potential of all controlled substances	Narcotic preparations containing limited amounts of noncontrolled active medicinal ingredients

*In New York State benzodiazepines are treated as schedule II substances, which require a triplicate prescription for a maximum of one month's supply.

Table 31.1-1 Characteristics of Drugs at Each DEA Level

Once a drug is approved for commercial use, the physician may, as part of the practice of medicine, lawfully prescribe a different dosage for a patient or otherwise vary the conditions of use from what is approved in the package labeling without notifying the FDA or obtaining its approval. Specifically, the FDC Act does not limit the manner in which a physician may use an approved drug. Although physicians may treat patients with an approved drug for unapproved purposes—that is, for indications not included on the drug's official labeling—without violating the FDC Act, the patient retains the right to redress for possible medical malpractice. This is a significant concern because the failure to follow the FDA-approved label may create an inference that the physician was varying from the prevailing standard of care. Although the failure to follow the information on the drug label does not impose liability per se and should not preclude a physician from using good clinical judgment in the interest of the patient, the physician should be aware that the label presents important guidelines regarding the safe and effective use of the drug.

Clinicians may prescribe medication for any reason that they believe to be medically indicated for the welfare of the patient. However, when using a drug for an unapproved indication or in a dose outside the usual range, it is good clinical practice to document the reasons for such treatment decisions in the patient's chart, as well as the explanation to the patient regarding this treatment recommendation. If clinicians are in doubt about a treatment plan, they should consult a colleague or suggest that the patient obtain a second opinion.

Therapeutic Trials The therapeutic trial of a particular drug should last for a previously determined length of time, unless abbreviated because of the appearance of adverse effects. Because the symptoms of psychiatric disorders are more difficult to assess than other physiological symptoms, such as hypertension, it is important for specific target symptoms to be identified at the initiation of a drug trial. The clinicians and the patient can then assess the target symptom over the course of the trial to help determine whether the drug has been effective; the use of patient or clinician rating scales might be useful for this purpose. If a drug has not been effective in reducing target symptoms in the specified length of time, and if other reasons for the lack of response can be eliminated, the drug should be tapered and stopped.

The brain is not a group of on-off neurochemical switches; rather, it is an interactive network of neurons in a complex homeostasis. Thus, the abrupt discontinuation of virtually any psychiatric drug is likely to disrupt the functioning of the brain further. Another common clinical mistake is the routine addition of medications without the discontinuation of a prior drug. Although this practice is indicated in specific circumstances, such as lithium augmentation of an unsuccessful trial of antidepressant agents, it often results in increased noncompliance and adverse effects and in the clinician's not being able to determine whether it was the second drug alone or the combination of drugs that resulted in a therapeutic success or a particular adverse effect.

Therapeutic Failures The failure of a specific drug trial should prompt the clinician to ask a number of questions regarding the patient, the illness, and the treatment. First, was the original diagnosis correct? In answering this question, the clinician should include the possibility of a mental disorder due to an undiagnosed medical condition, and also the possibility of illicit drug abuse. Second, are the observed symptoms related to the original disorder or are they actually adverse effects of the drug treatment? Some antipsychotic drugs, for example, can produce akinesia, which resembles psychotic withdrawal, or akathisia and neuroleptic malignant syndrome, which resemble increased psychotic agitation. Third, was the drug administered at an appropriate dosage for a sufficient length of time? Because patients can vary greatly in their absorption and metabolism of drugs, the clinician may need to measure plasma levels of a drug to ensure a sufficient dose of the drug. Fourth, did a pharmacokinetic or pharmacodynamic interaction with another drug the patient was taking reduce the efficacy of the newly prescribed drug? Fifth, did the patient take the drug as directed? Drug noncompliance is a common clinical problem that arises as a result of complicated drug regimens (more than one drug in more than one daily dosage), adverse effects (especially if unnoticed by the clinician), and poor patient education about the drug treatment plan (Table 31.1-2).

Excessively complex regimen (multiple agents, multiple small doses)
Early onset and persistence of side effects
Slow onset of beneficial effects
Low apparent relapse risk experienced if treatment is interrupted
Psychosis, confusion, dementia, pseudodementia, low intelligence, impaired hearing or vision, illiteracy
Simple lack of information, need for patient education
Financial hardship, conflicting obligations of time or money
Resentment, lack of confidence or trust
Specific psychopathology: paranoid delusions, hopelessness, masochism, anxiety and fear, ambivalence, control, splitting, passive aggression, passive dependence, denial, sociopathy, substance abuse
Involvement of multiple clinicians
Poor clinician-patient relationship
Inevitable human error

Adapted from Baldessarini RJ, Cole JO: Chemotherapy. In *The New Harvard Guide to Psychiatry*, AM Nichols, editor. Belknap Press, Cambridge, MA, 1988.

Table 31.1-2 Conditions That May Reduce Adherence to Recommended Treatment

COMBINED PSYCHOTHERAPY AND PHARMACOTHERAPY

Using drugs that affect the brain in combination with psychotherapy is standard therapy in contemporary psychiatry. When pharmacotherapy and psychotherapy are used together, the approach should be coordinated, integrated, and synergistic (Table 31.1-3). In many cases it has been demonstrated that the results of combined therapy are superior to those of either type of therapy alone. The term *pharmacotherapy-oriented psychotherapy* is used by some practitioners for this combined approach. When the psychotherapy and the pharmacotherapy are directed by two separate clinicians, the clinicians must communicate clearly with each other.

Stage	Engagement	Pattern Search	Change	Termination
Goals	Trust Credibility Self-observer alliance	To define problem patterns that, if changed, would lead to a desirable outcome	1. Relinquish old pattern(s) 2. Initiate new pattern(s) 3. Practice new pattern(s)	To separate efficiently
Techniques	Convey empathic understanding Effective suggestions Effective medications	Questions/queries Homework—idiosyncratic meanings ascribed to medication	Interpretation Reframing Behavioral suggestion Medication-induced change	Mutually agreed Patient initiates Therapist initiates Medication influenced
Content	Medication responsive Diagnosis	Does response to medication reflect a problem pattern?	Medication effects or insight around medication use accelerates change	Medications may prolong termination
Resistance	Are excessive side effects resistant to treatment?	Does pattern of nonadherence to medication regimen reflect a problem pattern?	Do new side effects suggest resistance to change?	Symptom recurrence not necessarily indication for medication change
Transference	Physician seen as malevolent or all-powerful	A key interpersonal pattern reflected in meaning of medication?	Unresolved distortions may be signaled by a new medication issue inhibiting change	Desire for new or more medication reflects desire to hold therapist
Countertransference	Physician failure to prescribe appropriately	Medication prescription reflects disbelief response to patient	Sudden change in regimen reflects an attempt to undermine change	New medication reflects desire to keep contact

Reprinted with permission from Reiman ED: Is Integrating Pharmacotherapy and Psychotherapy. ED Reiman, GL Klerman, editors. American Psychiatric Press, Washington, 1991.

Table 31.1-3 Stages of Individual Psychotherapy: A Medication Emphasis

Countertransference As in all types of psychotherapy, clinicians must be aware of their conscious and unconscious feelings toward their patients as well as of their own psychological attitudes towards psychotherapeutic drugs. Medications cannot replace a strong therapeutic alliance between a clinician and a patient. Therapists who are pessimistic about the value of psychotherapy or who misjudge the patient's motivation may prescribe medications out of their own nihilistic beliefs; others may withhold medication if they overvalue psychotherapy or undervalue pharmacological treatments. Each patient must be evaluated individually without clinician bias, and each treatment should be evaluated on the basis of a risk-benefit ratio.

Combined Therapy in Specific Disorders

Depressive Disorders Some patients and clinicians fear that medication will mask the depression and that psychotherapy will thus be impeded. However, that medication can be viewed as a facilitator in overcoming the anergia that may inhibit communication between the clinician and patient. The clinician can point out to the patient that the medication may have helpful effects on some of the associated symptoms of depression such as poor sleep and appetite, and even on such symptoms as irritability or social withdrawal.

Suicidal Behavior The possibility of suicide must be considered in treating patients with schizophrenia, bipolar I disorder, depressive disorders, severe personality disorders, and anxiety disorders (especially those who have panic attacks). If the clinician decides that the patient is at imminent risk for suicidal behavior, hospitalization is always indicated. If the patient can be managed outside a hospital, the patient's medication supply should be given to a responsible family member who can monitor the dosage and frequency of the prescribed medication. As a further precaution, the clinician may prescribe the patient a drug that is known to have little or no lethal potential when taken in overdose; however, even these drugs can become lethal when a patient takes multiple different medications in an overdose.

Bipolar I Disorder Most psychotherapists insist that patients with bipolar I disorder be medicated before they start any insight-oriented therapy. Without such medication, most patients with bipolar I disorder are unable to establish the necessary therapeutic alliance with their therapists. When such patients are depressed their abulia seriously disrupts their flow of thoughts, and the sessions are generally nonproductive. When they are manic, their flow of associations can be so rapid and their speech so pressured that the therapist may be flooded with material and unable to make appropriate interpretations or to assimilate the material into the patient's disrupted cognitive framework.

Anxiety Disorders Many drugs are effective in managing the signs and symptoms associated with the wide range of anxiety disorders. As symptoms are controlled by medication, patients are reassured and develop confidence that they will not be incapacitated by the disorder. This effect is particularly relevant to panic disorder, which is often associated with significant anticipatory anxiety about the attack. Depression may also complicate the symptom picture in patients with anxiety disorders and has to be addressed both pharmacologically and psychotherapeutically.

Schizophrenia and Other Psychotic Disorders Drug therapy is almost always used in the treatment of these disorders. In attempting individual psychotherapy with these patients, the therapist must work especially hard to establish an effective treatment relationship and therapeutic alliance with the patient. The schizophrenia patient generally defends against closeness and trust, and often becomes suspicious, anxious, hostile, or regressed in therapy.

Substance Abuse Patients who abuse alcohol or other substances present special challenges to therapists who wish to combine psychotherapy with pharmacotherapy. Some clinicians rarely use medications in this patient group, although treatment of underlying anxiety or depression might help with the abuse in particular patients. The clinician conducting psychotherapy with a patient who abuses substances should have no reservations about requesting random urine toxicological tests; as in all forms of insight-oriented psychotherapy, the psychological significance of such tests should be examined.

SPECIAL TREATMENT CONSIDERATIONS

Children Special care must be taken when administering psychotherapeutic drugs to children. Although the small volume of distribution suggests the use of lower doses than in adults, a child's higher rate of metabolism suggests that a higher ratio of mg of drug to kg of body weight should be used. In practice it is best to begin with a small dose and to increase it until clinical effects are observed. However, the clinician should not hesitate to use adult dosages in children if these dosages are effective and the adverse effects are acceptable.

Geriatric Patients The two major concerns when treating geriatric patients with psychotherapeutic drugs are that elderly persons may be more susceptible to adverse effects (particularly cardiac effects) and may metabolize and excrete drugs more slowly (Table 31.1-4), thus requiring lower dosages of medication. Another concern is that because geriatric patients often take more than one medication the clinician needs to consider the possible drug interactions. In practice clinicians should begin treating geriatric patients with a small dose, usually about one-half of the usual starting dose. The dosage should be raised in small increments more slowly than for middle-aged adults until either a clinical benefit is achieved or unacceptable adverse effects appear. Although many geriatric patients require a small dosage of

medication, many others require the usual adult dosage.

Phase	Change	Effect
Absorption	Gastric pH increases Decreased surface villi Decreased gastric motility and delayed gastric emptying Intestinal perfusion decreases	Little overall change Absorption is slower but just as complete
Distribution	Total body water and lean body mass decrease Increased total body fat, more marked in women Albumin decreases, gamma globulin increases, alpha$_1$ glycoprotein unchanged	Volume of distribution (Vd) increases for lipid-soluble drugs, decreases for water-soluble drugs The free or unbound percentage of albumin-bound drugs increases
Metabolism	Renal: renal blood flow and glomerular filtration rates decrease Hepatic: decreased enzyme activity and perfusion	Decreased metabolism leads to prolonged half-lives, if Vd remains the same
Total body weight	Decreases	Think on a mg-per-kg basis
Receptor sensitivity	May increase	Greater effect

Reprinted with permission from Ehrlinger HP: Clinical Guide to Somatic Therapies in Psychiatry. American Psychiatric Press, Washington, DC, 1988.

Table 31.1-4 Pharmacokinetics and Aging

Pregnant and Nursing Women The basic rule is to avoid administering any drug to a woman who is pregnant (particularly during the first trimester) or who is breastfeeding a child. This rule, however, occasionally needs to be broken when the mother's psychiatric disorder is severe. If psychotherapeutic medications need to be administered during pregnancy, the possibility of therapeutic abortion should be discussed. The two most teratogenic drugs in the psychopharmacopeia are lithium and the standard anticonvulsant agents. Lithium administration during pregnancy is associated with a high incidence of birth abnormalities, including Ebstein's malformation, a serious abnormality in cardiac development. Other psychoactive drugs (antidepressants, antipsychotics, and anxiolytics) are less clearly associated with birth defects but should also be avoided during pregnancy if at all possible. The most common clinical situation occurs when a pregnant woman becomes psychotic. If a decision is made not to terminate the pregnancy, treatment with antipsychotic drugs or electroconvulsive therapy (ECT) is preferable to lithium.

The administration of psychotherapeutic drugs at or near delivery may cause the baby to be overly sedated at delivery, thus requiring a respirator, or to be physically dependent on the drug, requiring detoxification and the treatment of a withdrawal syndrome. Virtually all psychiatric drugs are secreted in the milk of a nursing mother; therefore, mothers on those agents should be advised not to breastfeed their infants.

Medically Ill Patients Considerations when administering psychiatric drugs to medically ill patients include a potentially increased sensitivity to adverse effects, either increased or decreased metabolism and excretion of the drug, and interactions with other medications. As with children and geriatric patients, the most reasonable clinical practice is to begin with a small dosage, increase it slowly, and watch for both clinical benefit and adverse effects. Determining the plasma drug concentrations may be helpful for such patients.

COMMON ADVERSE EFFECTS

Patients generally have less trouble with adverse effects if they have been told to expect them. It is reasonable to explain the appearance of adverse effects as evidence that the drug is working, especially if the receptor mechanism for the adverse effects is the same as for the therapeutic effect. However, clinicians should distinguish between probable or expected adverse effects and those that are rare or unexpected.

Most psychotherapeutic drugs neither affect a single neurotransmitter system nor are their effects localized to the brain. The effects of psychotherapeutic drugs on neurotransmitter systems result in a wide range of adverse effects associated with their use. For example, some of the most common adverse effects of psychotherapeutic drugs are caused by the blockade of muscarinic acetylcholine receptors (Table 31.1-5). Many psychotherapeutic drugs antagonize dopaminergic, histaminergic, or adrenergic receptors, resulting in the adverse effects listed in Table 31.1-6.

Blurred vision
Constipation
Decreased salivation
Decreased sweating
Delayed or retrograde ejaculation
Delirium
Exacerbation of asthma (through decreased bronchial secretions)
Hyperthermia (through decreased sweating)
Memory problems
Narrow-angle glaucoma
Photophobia
Sinus tachycardia
Urinary retention

Table 31.1-5 Potential Adverse Effects Caused by Blockade of Muscarinic Acetylcholine Receptors

Antidopaminergic
Endocrine dysfunction
Hyperprolactinemia
Menstrual dysfunction
Sexual dysfunction
Movement disorders
Akathisia
Dystonia
Parkinsonism
Tardive dyskinesia
Antiadrenergic (primarily α_1)
Dizziness
Postural hypotension
Reflex tachycardia
Antihistaminergic
Hypotension
Sedation
Weight gain
Multiple neurotransmitter systems
Agranulocytosis (and other blood dyscrasias)
Allergic reactions
Anorexia
Cardiac conduction abnormalities
Nausea and vomiting
Seizures

Table 31.1-6 Potential Adverse Effects of Psychotherapeutic Drugs and Associated Neurotransmitter Systems

Treatment of Common Adverse Effects Many adverse effects are seen in patients who take psychotherapeutic drugs. The management of the adverse effects is similar regardless of which psychotherapeutic drug the patient is taking.

Dry Mouth Dry mouth is caused by the blockade of muscarinic acetylcholine receptors. When patients attempt to relieve the dry mouth by sucking on sugar-containing hard candies, they increase their risk of dental caries. They can avoid the problem by chewing sugarless gum or sucking on sugarless hard candies. Some clinicians recommend the use of a 1-percent solution of pilocarpine (Salagen), a cholinergic agonist, as a mouthwash three times daily. Other clinicians suggest 10 to 30 mg of bethanechol (Urecholine, Myotonachol) tablets, another cholinergic agonist, once or twice daily. It is best to start with 10 mg once a day and to increase the dosage slowly. Adverse effects of cholinomimetic drugs, such as bethanechol, include tremor, diarrhea, abdominal cramps, and excessive eye watering.

Blurred Vision The blockade of muscarinic acetylcholine receptors causes mydriasis (pupillary dilation) and cycloplegia (ciliary muscle paresis), resulting in presbyopia (blurred near vision). The symptom can be relieved by cholinomimetic eye drops. A 1-percent solution of pilocarpine can be prescribed as one drop in

each eye four times daily; bethanechol can be used for dry mouth as an alternative.

Urinary Retention The anticholinergic activity of many psychiatric drugs can lead to urinary hesitation, dribbling, and retention, as well as an increased rate of urinary tract infections. Elderly male patients with enlarged prostate glands are at increased risk for such adverse effects; 10 to 30 mg of bethanechol three to four times daily is usually effective in the treatment of the adverse effects on urination.

Constipation The anticholinergic activity of psychotropic drugs can result in the adverse effect of constipation. The first line of treatment involves the prescribing of bulk-forming laxatives, such as psyllium (Metamucil, Fiberall). If that treatment fails, cathartic laxatives, such as milk of magnesia, can be tried. Prolonged use of cathartic laxatives can result in a loss of their effectiveness. Bethanechol, 10 to 30 mg three to four times daily, can also be used.

Orthostatic Hypotension Orthostatic hypotension is caused by the blockade of a_1-adrenergic receptors. It is necessary to warn patients of this possible adverse effect, particularly if patients are elderly. The risk of hip fracture from falls is significantly elevated in patients who are taking psychotropic drugs. With patients at high risk of experiencing orthostatic hypotension, the clinician should choose a drug with low a_1-adrenergic activity. The patient can be instructed to get up slowly and to sit down immediately if dizziness is experienced. The patient can also try support hose to help reduce venous pooling of blood.

Sexual Dysfunction The use of psychiatric drugs can be associated with sexual dysfunction—decreased libido, impaired ejaculation and erection, and inhibition of female orgasm. Although warning patients about these adverse effects may increase their concern, they are not likely to report sexual dysfunction spontaneously to the physician. Also, some sexual dysfunctions may be related to the primary psychiatric disorder. Nevertheless, if sexual dysfunction emerges after pharmacotherapy has begun, it may be worthwhile to attempt to treat the symptoms. Neostigmine (Prostigmin), 7.5 to 15 mg orally 30 minutes before sexual intercourse, may help alleviate impaired ejaculation. Impaired erectile function may be helped with bethanechol given regularly, or possibly yohimbine (Yocon). Cyproheptadine (Periactin), 4 mg every morning, can be used to treat inhibited female orgasm, or 4 to 8 mg orally 1 to 2 hours before anticipated sexual activity for the treatment of inhibited male orgasm secondary to serotonergic drugs. Sildenfinil (Viagra) may also help in both men and women. The clinician and patient should consider switching treatment to another drug less or not at all associated with sexual dysfunction if this adverse effect is not acceptable to the patient.

Weight Gain Weight gain accompanies the use of many psychotropic drugs as a result of retained fluid, increased caloric intake, decreased exercise, or altered metabolism. Weight gain is a very common reason for noncompliance with the drug treatment regimen. Edema can be treated by elevating the affected body parts or by administrating a thiazide diuretic. If the patient is taking lithium or cardiac medications, the clinician must monitor blood concentrations, blood chemistries, and vital signs carefully. The patient should also be instructed to minimize the intake of fats and carbohydrates and to exercise regularly. If patients have not been exercising, the clinician should recommend that they start an exercise program at a modest level of exertion.

OVERDOSES

Clinicians must always be aware of the risk that a psychiatric patient will use the prescribed psychotherapeutic medication in an overdose in an attempt to commit suicide. It is good clinical practice to write nonrefillable prescriptions for small quantities of drugs when suicide is a consideration. In extreme cases an attempt should be made to verify that patients are taking the medication and not hoarding it for a later overdose attempt. Patients may attempt suicide just as they are beginning to get better. Clinicians therefore should continue to be careful about prescribing large quantities of medication until recovery is almost complete. Another consideration for clinicians is the possibility of accidental overdose, particularly by children in the household. Patients should be advised to keep psychotherapeutic medications in a safe place. A guide to the symptoms and treatment of overdoses from psychotherapeutic drugs is given in Table 31.1-7.

Table 31.1-7 Intoxication and Overdose With Selected Psychotherapeutic Drugs

PSYCHOTHERAPEUTIC DRUGS

The numerous pharmacological agents used to treat psychiatric disorders are referred to by three general terms that are used interchangeably: *psychotropic drugs, psychoactive drugs*, and *psychotherapeutic drugs*. Traditionally, these agents were divided into four categories: (1) antipsychotic or neuroleptic drugs used to treat psychosis; (2) antidepressant drugs used to treat depression; (3) antimanic drugs or mood stabilizers used to treat bipolar I disorder; and (4) antianxiety or anxiolytic drugs used to treat anxious states, although those same drugs were also effective as hypnotic agents in higher doses. The division, however, is less valid now than it was in the past for the following reasons: (1) Many drugs of one class are used to treat disorders previously assigned to another class. For example, many so-called antidepressant drugs are used to treat anxiety disorders, and some so-called antianxiety drugs are used to treat psychosis, depression, or bipolar I disorder. (2) Drugs from all four categories are used to treat disorders not previously treated with drugs (e.g., eating disorders, panic disorders, and impulse-control disorders). (3) Such drugs as clonidine (Catapres), propranolol (Inderal), and verapamil (Isoptin) can effectively treat a variety of psychiatric disorders and do not fit easily into the aforementioned classification of drugs.

Guide to Use An alphabetical list of generic drug names is presented in Table 31.1-8, with cross-references to the sections in which they or their class are discussed. A list of therapeutic indications and the drugs commonly used for those indications is presented in Table 31.1-9, also with cross-references to the sections in which they are discussed.

Table 31.1-8 Cross-References by Generic Name of Drug

Generic Name	Brand Name	Section Title	Section Number
Acebutolol	Sectral	b-Adrenergic Receptor Antagonists	31.5
Acetophenazine	Tindal	Dopamine Receptor Antagonists	31.17
Alprazolam	Xanax	Benzodiazepine Agonists and Antagonists	31.10
Amantadine	Symmetrel	Anticholinergics and Amantadine	31.6
Amobarbital	Amytal	Barbiturates and Similarly Acting Substances	31.9
Amoxapine	Asendin	Tricyclic and Tetracyclic Drugs	31.30
Aprobarbital	Alurate	Barbiturates and Similarly Acting Substances	31.9
Atenolol	Tenormin	b-adrenergic Receptor Antagonists	31.5
Benztropine	Cogentin	Anticholinergics and Amantadine	31.6
Biperden	Akineton	Anticholinergics and Amantadine	31.6

Brofaromine	Consonar	Monoamine Oxidase Inhibitors	31.20
Bromocriptine	Parlodel	Other Pharmacological and Biological Therapies	31.33
Buprenorphine	Buprenex	Opioid Agonists	31.23
Bupropion	Wellbutrin	Bupropion	31.11
Buspirone	BuSpar	Buspirone	31.12
Butabarbital	Butisol	Barbiturates and Similarly Acting Substances	31.9
Butaperazine	Repoise	Dopamine Receptor Antagonists	31.17
Carbamazepine	Tegretol	Anticonvulsants	31.7
Carisprodol	Soma	Barbiturates and Similarly Acting Substances	31.9
Carphenazine	Proketazine	Dopamine Receptor Antagonists	31.17
Chloral hydrate	Noctec	Chloral Hydrate	31.14
Chlorpromazine	Thorazine	Dopamine Receptor Antagonists	31.17
Chlorprothixene	Taractan	Dopamine Receptor Antagonists	31.17
Citalopram	Celexa	Selective Serotonin Reuptake Inhibitors	31.25
Clomipramine	Anafranil	Tricyclics and Tetracyclics	31.30
Clonazepam	Kionopin	Benzodiazepine Receptor Agonists and Antagonists	31.10
Clonidine	Catapres	Clonidine	31.16
Clorgyline	—	Monoamine Oxidase Inhibitors	31.20
Clozapine	Clozaril	Serotonin-Dopamine Antagonists	31.26
Cycrimine	Pagitane	Anticholinergics and Amantadine	31.6
Cyproheptadine	Periactin	Antihistamines	31.8
Dantrolene	Dantrium	Other Pharmacological and Biological Therapies	31.33
Desipramine	Norpramin, Pertofane	Tricyclics and Tetracyclics	31.30
Dexfenfluramine	—	Other Pharmacological and Biological Therapies	31.33
Dextroamphetamine	Dexedrine	Sympathomimetics	31.27
Diazepam	Valium	Benzodiazepine Receptor Agonists and Antagonists	31.10
Diltiazem	Cardizem	Calcium Channel Inhibitors	31.13
Disulfiram	Antabuse	Other Pharmacological and Biological Therapies	31.33
Diphenhydramine	Benadryl	Antihistamines	31.8
Divalproex	Depakote	Anticonvulsants	31.7
Donepezil	Aricept	Cholinesterase Inhibitors	31.15
Doxepin	Adapin, Sinequan	Tricyclics and Tetracyclics	31.30
Droperidol	Inapsine	Dopamine Receptor Antagonists	31.17
Estazolam	ProSom	Benzodiazepine Receptor Agonists and Antagonists	31.10
Ethinamate	Valmid	Barbiturates and Similarly Acting Substances	31.9
Ethopropazine	Parsidol	Anticholinergics and Amantadine	31.6
Ethchlorvynol	Placidyl	Barbiturates and Similarly Acting Substances	31.9
Fenfluramine	Pondimin	Other Pharmacological and Biological Therapies	31.33
Flumazenil	Mazicon	Benzodiazepine Receptor Agonists and Antagonists	31.9
Fluoxetine	Prozac	Selective Serotonin Reuptake Inhibitors	31.25
Fluphenazine	Prolixin, Permitil	Dopamine Receptor Antagonists	31.17
Flurazepam	Dalmane	Benzodiazepine Receptor Agonists and Antagonists	31.10
Fluvoxamine	Luvox	Selective Serotonin Reuptake Inhibitors	31.25
Gabapentin	Neurontin	Anticonvulsants	31.7
Glutethimide	Doriden	Barbiturates and Similarly Acting Substance	31.9
Halazepam	Paxipam	Benzodiazepine Receptor Agonists and Antagonists	31.10
Haloperidol	Haldol	Dopamine Receptor Antagonists	31.17
Hydroxyzine	Atarax, Vistaril	Antihistamines	31.8
Imipramine	Tofranil	Tricyclic and Tetracyclic Drugs	31.30
Isocarboxazid	Marplan	Monoamine Oxidase Inhibitors	31.20
Labetalol	Normodyne, Trandate	b-adrenergic Receptor Antagonists	31.5
Lamotrigine	Lamictal	Anticonvulsants	31.7
Levodopa	Larodopa	Levodopa	31.33
Levomethadyl acetate	ORLAAM	Opioid Antagonists	31.23
Levothyroxine	Levoxine, Levothroid, Synthroid	Thyroid Hormones	31.28
Liothyronine	Cytomel	Thyroid Hormones	31.28
Lithium	Eskalith, Lithobid	Lithium	31.18
Lorazepam	Ativan	Benzodiazepine Receptor Agonists and Antagonists	31.10
Loxapine	Loxitane	Dopamine Receptor Antagonists	31.17
Mephobarbital	Mebaral	Barbiturates and Similarly Acting Substances	31.9
Meprobamate	Miltown	Barbiturates and Similarly Acting Substances	31.9
Mesoridazine	Serentil	Dopamine Receptor Antagonists	31.17
Methadone	Dolophine, Methadose	Opioid Agonists	31.23
Metharbital	Gemonil	Barbiturates and Similarly Acting Substances	31.9
Methohexital	Brevital	Barbiturates and Similarly Acting Substances	31.9

Methylphenidate	Ritalin	Sympathomimetics	31.27
Methypryton	Noludar	Barbiturates and Similarly Acting Substances	31.9
Metoprolol	Lopressor, Toprol	b-Adrenergic Receptor Antagonists	31.5
Midazolam	Versed	Benzodiazepine Receptor Agonists and Antagonists	31.10
Mirtazapine	Remeron	Mirtazapine	31.19
Moclobemide	Aurorix	Monoamine Oxidase Inhibitors	31.20
Molindone	Moban	Dopamine Receptor Antagonists	31.17
Nadolol	Corgard	b-Adrenergic Receptor Antagonists	31.5
Nefazodone	Serzone	Nefazodone	31.22
Nifedipine	Adalat, Procardia	Calcium Channel Inhibitors	31.13
Nimodipine	Nimotop	Calcium Channel Inhibitors	31.13
Nortriptyline	Pamelor, Aventyl	Tricyclics and Tetracyclics	31.30
Olanzapine	Zyprexa	Serotonin Dopamine Antagonists	31.26
Orphenadine	Norflex, Dispal	Anticholinergics and Amantadine	31.6
Oxazepam	Serax	Benzodiazepine Receptor Agonists and Antagonists	31.10
Paraldehyde	Generics	Barbiturates and Similarly Acting Substances	31.9
Paroxetine	Paxil	Selective Serotonin Reuptake Inhibitors	31.25
Pemoline	Cylert	Selective Sympathomimetics	31.27
Phenelzine	Nardil	Monoamine Oxidase Inhibitors	31.20
Phenobarbital	Solfoton	Barbiturates and Similarly Acting Substances	31.9
Pentobarbital	Nembutal	Barbiturates and Similarly Acting Substances	31.9
Perphenazine	Trilafon	Dopamine Receptor Antagonists	31.17
Pimozide	Orap	Dopamine Receptor Antagonists	31.17
Pindolol	Visken	b-Adrenergic Receptor Antagonists	31.5
Piperacetazine	Quide	Dopamine Receptor Antagonists	31.17
Prazepam	Centrax	Benzodiazepine Receptor Agonists and Antagonists	31.10
Prochlorperazine	Compazine	Dopamine Receptor Antagonists	31.17
Procyclidine	Kemadrin	Anticholinergics and Amantadine	31.6
Promazine	Sparine	Dopamine Receptor Antagonists	31.17
Promethazine	Phenergan	Antihistamines	31.8
Propranolol	Inderal	b-Adrenergic Receptor Antagonists	31.5
Protriptyline	Vivactil	Benzodiazepine Receptor Agonists and Antagonists	31.10
Quazepam	Doral	Benzodiazepine Receptor Agonists and Antagonists	31.10
Quetiapine	Seroquel	Serotonin-Dopamine Antagonists	31.26
Reboxetine	Edronax, Norebox	Other Pharmacological and Biological Therapies	31.33
Reserpine	Diupres	Dopamine Receptor Antagonists	31.17
Risperidone	Risperdal	Serotonin-Dopamine Antagonists	31.26
Secobarbital	Seconal	Barbiturates and Similarly Acting Substances	31.9
Selegiline	Eldepryl	Monoamine Oxidase Inhibitors	31.20
Sertindole	Serlect, Serdolect	Serotonin-Dopamine Antagonists	31.26
Sertraline	Zoloft	Selective Serotonin Reuptake Inhibitors	31.25
Sulpiride	Dogmatil, Sesif	Dopamine Receptor Antagonists	31.17
Tacrine	Cognex	Cholinesterase Inhibitors	31.15
Temazepam	Restoril	Benzodiazepine Receptor Agonists and Antagonists	31.10
Thiopental	Pentothal	Barbiturates and Similarly Acting Substances	31.9
Thioridazine	Mellaril	Dopamine Receptor Antagonists	31.17
Thiothixene	Navane	Dopamine Receptor Antagonists	31.17
Topiramate	Topamax	Anticonvulsants	31.7
Tiagabine	—	Anticonvulsants	31.7
Tranylcypromine	Parnate	Monoamine Oxidase Inhibitors	31.20
Trazodone	Desyrel	Trazodone	31.29
Triazolam	Halcion	Benzodiazepine Receptor Agonists and Antagonists	31.10
Trihexyphenidyl	Artane	Anticholinergics and Amantadine	31.6
Trifluoperazine	Stelazine	Dopamine Receptor Antagonists	31.17
Triflupromazine	Vesprin	Dopamine Receptor Antagonists	31.17
Trimipramine	Surmontil	Tricyclics and Tetracyclics	31.30
L-Tryptophan	—	Other Pharmacological and Biological therapies	31.33
Valproate	Depakene	Anticonvulsants	31.7
Valproic Acid	Depakene	Anticonvulsants	31.7
Venlafaxine	Effexor	Selective Serotonin-Noradrenaline Reuptake Inhibitors	31.24
Verapamil	Calan, Isoptin	Calcium Channel Inhibitors	31.13
Vigabatrin	Sabril	Anticonvulsants	31.7
Yohimbine	Yocon	Other Pharmacological and Biological Therapies	31.33
Ziprasidone	Zeldox	Serotonin-Dopamine Antagonists	31.26
Zolpidem	Ambien	Benzodiazepine Receptor Agonists and Antagonists	31.10

Table 31.1-9 Major Mental Disorders and the Most Common Drugs and Classes of Drugs Used in Their Treatment

Disorder	Section Number
Aggression and agitation (see Intermittent explosive disorder)	
Akathisia (see Medication-induced movement disorders)	
Alcohol use disorders	
b-adrenergic receptor antagonists	31.5
Benzodiazepines	31.10
Carbamazepine	31.7a
Disulfiram	31.33
Lithium	31.18
Naltrexone	31.21
Anorexia nervosa (see Eating disorders)	
Anxiety (see also specific anxiety disorders)	
Antihistamines	31.8
Benzodiazepines	31.10
Buspirone	31.12
Bipolar I disorder	
Anticonvulsants	31.7
Benzodiazepines (especially clonazepam)	31.10
Calcium Channel inhibitors	31.13
Dopamine receptor antagonists	31.17
Lithium	31.18
Serotonin-dopamine antagonists	31.26
L-tryptophan	31.33
Bulimia nervosa (see Eating disorders)	
Cyclothymic disorder (see Bipolar I disorder)	
Delusional disorder (see Schizophrenia)	
Dementia of the Alzheimer's type (cognitive symptoms)	
Cholinesterase inhibitors	31.15
Depressive disorders	
Anticonvulsants	31.7
Benzodiazepines (especially alprazolam)	31.10
Bromocriptine	31.33
Bupropion	31.11
Calcium channel inhibitors	31.13
Lithium	31.18
Mirtazapine	31.19
Monoamine oxidase inhibitors	31.20
Nefazodone	31.22
Reboxetine	31.33
Serotonin-noradrenaline reuptake inhibitors	31.24
Selective serotonin reuptake inhibitors	31.25
Sympathomimetics	31.27
Thyroid hormones	31.28
Trazodone	31.29
Tricyclic and tetracyclic drugs	31.30
L-Tryptophan (withdrawn from U.S. market)	31.33
Dysthymic disorder (see depressive disorders)	
Dystonia (see Medication-induced movement disorders)	
Eating disorders	
Lithium	31.18
Monoamine oxidase inhibitors	31.20
Selective serotonin reuptake inhibitors	31.25
Trazodone	31.29
Tricyclics and tetracyclics	31.30
Generalized anxiety disorder	
b-adrenergic receptor antagonists	31.5
Barbiturates and similarly acting drugs	31.9
Benzodiazepines	31.10
Buspirone	31.12
Selective serotonin-noradrenaline reuptake inhibitors	31.24
Selective serotonin reuptake inhibitors	31.25
Trazodone	31.29
Tricyclics and tetracyclics	31.30
Intermittent explosive disorder	

COMBINATION DRUGS

In addition to drugs that contain a single active ingredient, a small number of combination drugs are available in the United States (Table 31.1-10). It is possible that the use of such drugs may increase the patients' compliance by simplifying the drug regimen. A problem with combination drugs, however, is that the clinician has less flexibility in adjusting the dosage of one of the components; that is, the use of combination drugs may cause two drugs to be administered when only one drug continues to be necessary for therapeutic efficacy.

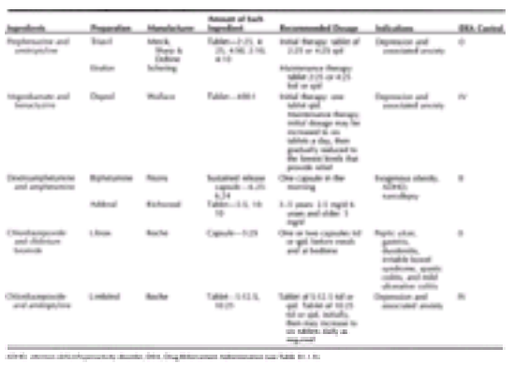

Table 31.1-10 Combination Drugs Used in Psychiatry

SUGGESTED CROSS-REFERENCES

The process by which new drugs are approved in the United States is explained in Section 31.3; pharmacokinetics and drug interactions are described in Section 31.2; and combined psychotherapy and pharmacotherapy is explained in Section 30.12. Medication-induced movement disorders are discussed in Section 31.4.

SECTION REFERENCES

Altshuler LL, Cohen L, Szuba MP, Burt VK, Gitlin M, Mintz J: Pharmacologic management of psychiatric illness during pregnancy: Dilemmas and guidelines. Am J Psychiatry *153:* 592, 1996.

Barki ZHK, Kravitz HM, Berki TM: Psycotropic medications in pregnancy. Psychiatr Ann *28:* 486, 1998.

*Beitman BD, Klerman GL, editors: *Integrating Pharmacotherapy and Psychotherapy*. American Psychiatric Press, Washington, DC, 1991.

*Gabbard GO: *Psychodynamic Psychiatry in Clinical Practice*. American Psychiatric Press, Washington, DC, 1990.

Glick ID, Lecrubier Y, Montgomery SA, Vinar O, Klein DF: Efficacious and safe psychotropics not available in the United States. Psychiatr Ann *26:* 354, 1996.

*Kane JM, Lieberman JA, editors: *Adverse Effects of Psychotropic Drugs*. Guilford Press, New York, 1992.

*Mahoney MC, Connolly BF, Smith CM: A clozapine overdose with markedly elevated serum levels. J Clin Pharmacol *39:* 97, 1999.

Pincus HA, Tanielian TL, Marcus SC, Olfson M, Zarin DA, Thompson J, Zito JM: Prescribing trends in psychotropic medications. Primary care, psychiatry, and other medical specialties. JAMA *279:* 526, 1998.

*Prien RF, Robinson DS, editors: *Clinical Evaluation of Psychotropic Drugs*. Raven, New York, 1994.

Shen WW: The metabolism of psychoactive drugs: a review of enzymatic biotransformation and inhibition. Biol Psychiatry *41:* 814, 1997.

*Stahl SM: *Essential Psychopharmacology*. Cambridge University Press, Cambridge, UK, 1996.

*Tallman JF: Neuropsychopharmacology at the new millennium: New industry directions. Neuropsychopharmacology *20:* 99, 1999.

31.2 PHARMACOKINETICS AND DRUG INTERACTIONS

PHILIP G. JANICAK, M.D., AND JOHN M. DAVIS, M.D.

Pharmacokinetics describes the various phases that a drug undergoes when it enters the body, including absorption, distribution, metabolism, and elimination. Pharmacokinetics influences the process by which critical drug concentrations reach the sites of action (typically a neuron or group of neurons). Important related pharmacokinetic factors are described in Table 31.2-1. *Pharmacodynamics* describes what a drug does in the body, which depends on an agent reaching its site or sites of action (e.g., neurotransmitters, receptors, second messengers) in the appropriate concentration. Interaction with the targeted site then produces a series of changes in subcellular neuronal activity that culminates in clinically relevant behavioral effects.

Absorption	Process by which a drug proceeds from the site of administration to the site of measurement (generally plasma or whole blood)
First-pass effect	Hepatic extraction of orally administered drugs prior to reaching the systemic circulation
Volume of distribution (V_d)	How much drug is distributed throughout the body
Steady-state concentration (C_{ss})	Drug concentration achieved when the amount administered per unit time equals the amount eliminated per unit time
Biological half-life ($t_{1/2}$)	Time required for the drug concentration in plasma (or blood) to fall by one-half
Elimination rate constant (K_e)	Percentage of drug in the body eliminated per unit time
Clearance (Cl)	A measure of a drug's elimination from the body (i.e., the amount of drug in the body times the elimination rate constant)
First-order kinetics	Amount of drug eliminated per unit time is directly proportional to its plasma concentration
Zero-order kinetics	A fixed amount of drug is eliminated per unit time regardless of plasma concentration

Reprinted with permission from Janicak PG, Davis JM, Preskorn SH, Ayd FJ Jr: *Principles and Practice of Psychopharmacotherapy*, ed 2. Williams & Wilkins, Baltimore, 1997.

Table 31.2-1 Pharmacokinetic Factors

Drug interactions change the clinical effect of one or more concurrently administered agents by altering their respective pharmacokinetics or pharmacodynamics. Recently discontinued agents with long elimination half-lives (e.g., fluoxetine [Prozac] and its metabolite norfluoxetine) can also be involved in drug interactions.

PRIMARY PHARMACOKINETIC PHASES

Absorption Most agents are administered orally and absorbed through the small bowel. They then enter the portal circulation and pass through the liver, where they may undergo significant presystemic metabolism (i.e., first-pass effect) before reaching the systemic circulation. Since most psychotropic agents are highly lipophilic, they quickly cross the blood-brain barrier and are then delivered to the site of action via the cerebral circulation. Because of their lipophilicity, these agents are usually rapidly and completely absorbed, have a high first-pass effect, have a large volume of distribution, and are readily available to the central nervous system (CNS).

Bioavailability is the portion of a drug that is absorbed from the site of administration. Intravenous administration presumably produces 100 percent absorption. Various factors that attenuate a drug's bioavailability when given orally or intramuscularly include its physicochemical properties, specific formulation, precipitation at the site of injection, and disease states that alter gastrointestinal function or first-pass effect. A shorter time to peak concentration (T_{max}) translates to a more rapid onset of a drug's clinical effect (Fig. 31.2-1). While this is often desirable (e.g., for a sedative-hypnotic), at times too-rapid absorption may result in a toxic peak concentration (C_{max}). A drug's lipophilicity and polarity are also pertinent. For example, more-polar compounds penetrate the gastrointestinal tract and blood-brain-barrier slowly. Thus, oxazepam (Serax), the most polar benzodiazepine, is a poor sedative because of its delayed absorption into the systemic and cerebral circulations.

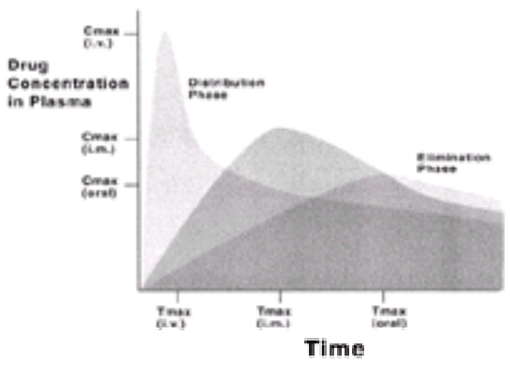

FIGURE 31.2-1 Single-dose plasma drug concentration versus time curves for same drug given intravenously, intramuscularly, or orally.

Different formulations (e.g., two different generic antipsychotic agents) can have substantially different bioavailabilities. While the Food and Drug Administration (FDA) requires that a generic formulation differs from a brand name product by no more or no less than 30 percent in bioavailability, two different generic formulations can differ by up to 60 percent. Thus switches from one generic product to another can involve a substantial difference in efficacy or toxicity. Route of administration can affect absorption as well as the ratio of parent compound to its various metabolites. While the intramuscular route usually results in more-rapid absorption, this is not always true. For example, certain benzodiazopines (e.g., diazepam [Valium]) tend to crystallize after injection and are actually less available than their oral counterpart because of the slower, more erratic, and less complete release from muscle tissue sites.

The first-pass effect can be altered by various diseases (e.g., cirrhosis, congestive heart failure) and by a number of substances ingested concurrently (e.g., alcohol, selective serotonin reuptake inhibitors [SSRIs]). In turn, this influences the C_{max} and ratio of parent compound to metabolites. Avoiding hepatic metabolism may enhance drug potency. For example, parenteral administration avoids the first-pass effect, so many agents have greater potency when given intramuscularly than when given orally. Diseases such as cirrhosis can lead to portacaval shunting; thus drugs can bypass the hepatic metabolism and directly enter the systemic circulation. Finally, acute alcohol intoxication can, in the absence of liver dysfunction, increase absorption and decrease first-pass metabolism of agents such as tricyclic drugs, contributing to the greater toxicity seen when they are combined (e.g., in an overdose).

Distribution Distribution depends on an agent's lipophilicity as well as an organ's fat and protein content. The vascularity of an organ determines the rate of accumulation. After a single oral dose, drug concentration reaches C_{max} and then undergoes an initial, relatively rapid decline as it is distributed into other body compartments (e.g., adipose tissue, aqueous, bone) rather than elimination (Fig. 31.2-1). In addition, most psychoactive drugs are highly protein bound, and only a small amount of an unbound drug (i.e., free fraction) is available at the site of action. Even a small shift in the bound-unbound ratio can substantially increase the free-drug fraction, at times increasing toxicity. Malnutrition, aging, and other drugs (e.g., divalproex [Depakote], fluoxetine) that compete for protein binding sites will

affect this ratio. Alterations in the relative fat, protein, and total body water content in elderly patients partially explain why this age group often has a longer drug effect than younger adults.

Metabolism Most psychotropic drugs are hepatically transformed to more-polar metabolites that are then excreted. Biotransformation may involve hydroxylation, demethylation, oxidation, and sulfoxide formation. While most agents undergo these phases, some (e.g., lorazepam [Ativan]) only undergo simple conjugation with glucuronic acids or are excreted without being transformed (e.g., lithium [Eskalith]). Since conjugation can occur in organs other than the liver, hepatic compromise usually does not substantially affect the rate of clearance of agents metabolized in this manner. Various metabolites are formed in biotransformation, which may significantly determine the ultimate behavioral effects (intended or adverse) of a drug. For example, fluoxetine's primary metabolite, norfluoxetine, has actions similar to those of the parent compound but is cleared more slowly. Over time this metabolite accumulates and becomes the primary acting agent (rather than fluoxetine). Alternatively, certain metabolites are less effective and more toxic than the parent compound (e.g., 2-hydroxyimipramine versus imipramine [Tofranil]). Thus, a shift in the ratio of parent compound to metabolite may decrease the antidepressant effect and increase toxicity.

Enzyme Induction and Inhibition An important related issue is the alteration of activity in the cytochrome P450 enzyme system. These isoenzymes are primarily located in the endoplasmic reticulum of hepatocytes and the gut wall. They make up a family of more than 30 known and related enzymes responsible for the oxidative metabolism of many drugs, including most psychotropic agents. Certain substances may induce this system and thus accelerate the clearance of agents that are substrates of these enzymes; examples include alcohol, smoking (polyaromatic hydrocarbons), barbiturates, and anticonvulsants. Indeed, some are such potent inducers that they can increase their own metabolism to a clinically significant degree (e.g., carbamazepine [Tegretol]).

Other substances may inhibit this system and diminish the metabolism of coadministered substrates, leading to potential problems with toxicity. Examples include SSRIs; quinidine (Quinidex), a potent inhibitor used as a reference standard for inhibition studies of the cytochrome P450 (CYP) 2D6 (CYP 2D6) isoenzyme; certain antipsychotic drugs; and grapefruit juice (flavonoids, furanocoumarins). One of the best-known examples of this phenomenon is the decrease in terfenadine (Seldane) metabolism by agents that inhibit the CYP 450 3A3/4 isoenzyme (e.g., ketoconazole [Nizoral], erythromycin [E-Mycin]). As a result, there may be an increase in the concentration of the prodrug parent compound, which lacks the antihistaminic properties of its major metabolite. More importantly, increased prodrug concentrations are potentially cardiotoxic (e.g., torsades des pointes).

Various diseases that directly (e.g., cirrhosis, viral infections) or indirectly (e.g., cardiac failure) compromise hepatic function can also alter the rate of drug metabolism. This may result in toxicity because of increased concentration of a drug (e.g., cardiac arrhythmias due to elevated concentrations of a tricyclic drug).

Elimination Most psychoactive agents are eliminated by biotransformation to one or more metabolites with greater polarity. The increased water solubility and decreased lipid solubility of these polar metabolites then causes a distribution shift from various organs to plasma and increases renal clearance. This is represented on the single-dose, concentration-time curve as the terminal elimination phase (Fig. 31.2-1). Typically, there is a gradual but steady decline in plasma concentration over time. Renal insufficiency, dehydration, changes in plasma pH, and certain drugs may affect the kidney's ability to excrete a given agent. For example, certain diuretics and nonsteroidal anti-inflammatory drugs (NSAIDs) can interfere with renal excretion of lithium.

For many of the factors discussed, therapeutic drug monitoring has become the standard of care for a variety of psychotropics under certain clinical circumstances (e.g., tricyclic drugs; lithium, divalproex and carbamazepine; some antipsychotic agents).

DRUG INTERACTIONS

Drug combinations are being used increasingly to enhance efficacy (e.g., lithium plus an antidepressant) or to counteract adverse effects of a drug (e.g., antiparkinsonian agents plus neuroleptic drugs). Fortunately, most interactions between concurrent drug therapies are not clinically relevant, but some may interfere with drug efficacy or cause serious morbidity or mortality. Elderly persons are usually at greatest risk for adverse drug interactions since they often receive multiple drugs, often self-prescribed, and use more over-the-counter (OTC) agents; tend to see several physicians for a variety of complaints; and are especially vulnerable to CNS sedation because of increased sensitivity and alterations in pharmacokinetics. Drug interactions may be secondary to pharmacokinetic factors, pharmacodynamic factors, or both. Problems can arise in any or all pharmacokinetic phases (Table 31.2-2).

Phase	Interaction
Absorption	Antacids or anticholinergic agents may impede absorption of other agents
Distribution	Agents with higher protein-binding affinity may displace other drugs, thus increasing their free fraction (e.g., divalproex plus carbamazepine)
Metabolism	Drugs that induce or inhibit CYP 450 isoenzymes can increase or decrease metabolism of coadministered agent
Elimination	Sodium bicarbonate or ascorbic acid can increase excretion of various agents; angiotensin-converting enzyme (ACE) inhibitors may increase lithium concentrations

Table 31.2-2 Drug Interactions at Various Pharmacokinetic Phases

Pharmacokinetic Interactions CYP 2D6 isoenzyme inhibition or induction can alter drug metabolism. This has taken on increased importance with the advent of potent inhibitors (e.g., fluoxetine and paroxetine [Paxil]) and potent inducers (e.g., carbamazepine), many of which are commonly used for various psychiatric disorders. In addition, 5 to 10 percent of white persons possess an autosomal recessive genetic defect in the expression of the isoform CYP 2D6, making them poor metabolizers. Individuals can be genotyped for CYP 2D6 activity by use of polymerase chain reaction (PCR) techniques. Thus, drug inhibition or genetic polymorphism may cause toxic increases of drug in the concentration such as certain tricyclic agents. Fluvoxamine (Luvox) inhibits CYP 1A2 and has resulted in a twofold increase in concentrations of substrates such as theophylline (Theo-Dur) when coadministered. CYP 2C19 also exhibits genetic polymorphism in Japanese (20 percent), African-Americans (2 to 20 percent), Africans (4 to 8 percent), and whites (3 to 5 percent), who have been reported to be poor metabolizers. Important potential interactions may occur when inhibitors (e.g., fluvoxamine, fluoxetine, and sertraline [Zoloft]) and substrates of this isoenzyme (e.g., diazepam, clomipramine [Anafranil]) are administered. CYP 3A4 is possibly the most important isoenzyme, accounting for up to 60 percent of the total CYP 450 content. Potent inhibitors of CYP 3A4 (e.g., nefazodone [Serzone], fluvoxamine, norfluoxetine-fluoxetine, paroxetine, and sertraline) raise concern about the potential for cardiotoxic effects when they are combined with the nonsedating antihistamines terfenadine or astemizole (Hismanal). This does not appear to be an issue, however, when they are used in combination with the nonsedating antihistamines loradatine (Claritin) or citrizine (Zyrtec). Further, this may only have clinical relevance when more than one CYP 3A4 inhibitors are combined or when a single agent is taken in large amounts, such as in an overdose.

Extrapolation from in vitro studies of the CYP 450 enzyme system to the clinical setting is presently littered with possible confounds. Further, other factors are almost certainly involved, since most patients receiving such drug combinations do not exhibit cardiotoxicity. In part, this is due to wide interindividual variability in drug metabolism resulting from genetic polymorphism in the functional expression of various CYP 450 isoenzymes. Thus, while in vitro evidence allows a ranking of various agents in terms of their inhibition potency for various isoenzymes, the relative lack of direct comparison between in vitro data and clinically significant drug interactions does not permit ranking most agents in this latter respect.

Given our present state of knowledge about the CYP 450 enzyme system, the following recommendations seem appropriate.

CYP 2D6 isoenzyme: Antidepressants and antipsychotic drugs should be used cautiously with fluoxetine and paroxetine. Type 1C antiarrhythmic agents should not be used in combination with certain SSRIs.

CYP 3A4 isoenzyme: Fluvoxamine, nefazodone, and fluoxetine should be used with caution when combined with terfenadine, astemizole, certain benzodiazepines (e.g., triazolam [Halcion]), and carbamazepine.

CY 1A2 isoenzyme: Fluvoxamine should be avoided with patients on theophylline or clozapine (Clozaril).

CYP 2C9/10 and CYP 2C19 isoenzymes: Combined fluoxetine plus phenytoin (Dilantin), sertraline plus tolbutamide (Orinase), and fluvoxamine plus warfarin (Coumadin) should be used cautiously.

CYP 2B6 isoenzyme: While its clinical relevance is unclear, it is known that bupropion (Wellbutrin) is converted to its major metabolite by this isoenzyme.

Finally, even after discontinuing fluoxetine administration, given the prolonged elimination half-life of norfluoxetine, administration of substrates with the potential for adverse interactions should be carefully monitored for several weeks.

Flavin-containing Monooxygenases (FMO) The FMO system is also responsible for drug metabolism. While it is probably not an inducible system like the cytochrome P450, its metabolic capacity may be inhibited. Presently, there are only five known FMO families (FMO 1–5), which may work in conjunction with the cytochrome P450 system. For example, olanzapine (Zyprexa) is metabolized by CYP 1A2, CYP 2D6, and FMO 3. Thus, alterations in the metabolism of either system may alter the metabolism of the other, with the subsequent potential for clinically relevant drug interactions.

Pharmacodynamic Interactions Pharmacodynamic interactions may cause changes in efficacy or toxicity in the absence of changes in pharmacokinetics. These may occur because of CNS changes that affect the sites of action. These types of interactions have not been studied as extensively as pharmacokinetic interactions and thus are not as well understood. One example is the antagonism that may occur when parkinsonian patients on levodopa (Levodopa) or carbidopa (Lodosyn) are given an antipsychotic drug to control associated psychotic symptoms, often leading to a recurrence of parkinsonian symptoms (e.g., akinesia, increased rigidity).

Another is the potential for developing the serotonin syndrome when SSRIs are combined with monoamine oxidase inhibitors (MAOIs). This syndrome carries the potential for significant morbidity and mortality; it is characterized by symptoms such as confusion, myoclonus, hyperreflexia, diaphoresis, and possibly cardiovascular compromise.

Clinicians should always anticipate the potential for adverse interactions, which are best managed preventively. In this context, the following steps are appropriate:

Obtain a detailed medication history, including the use of other prescribed drugs, OTC drugs, illicit drugs, and alcohol

Identify high-risk groups such as elderly and very young patients and those with certain medical disorders such as renal or hepatic disease

Educate patients, including written instructions when appropriate (e.g., MAOI diet)

Avoid polypharmacy when possible

Keep detailed, updated references on important potential drug interactions

Table 31.2-3, Table 31.2-4, Table 31.2-5, Table 31.2-6, Table 31.2-7, Table 31.2-8, Table 31.2-9, Table 31.2-10, Table 31.2-11 and Table 31.2-12 summarize some of the more common clinically relevant potential drug interactions, the type of data that supports these interactions, and their suggested clinical management.

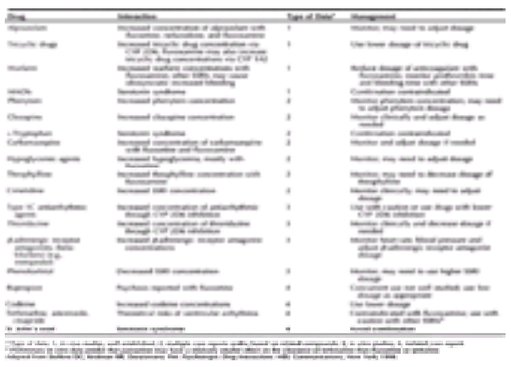

Table 31.2-3 Selected SSRI Drug Interactions

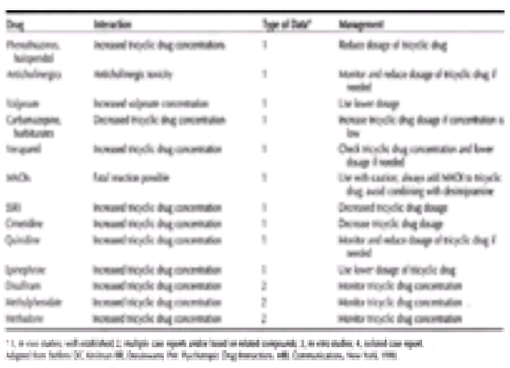

Table 31.2-4 Selected Tricyclic Drug Interactions

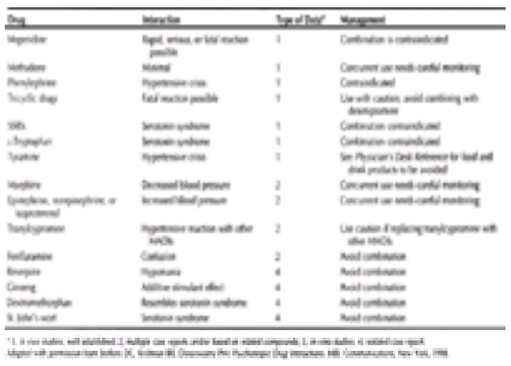

Table 31.2-5 Selected MAOI Drug Interactions

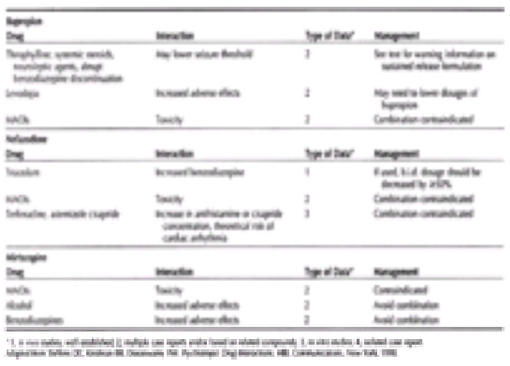

Table 31.2-6 Drug Interactions of Other Antidepressants

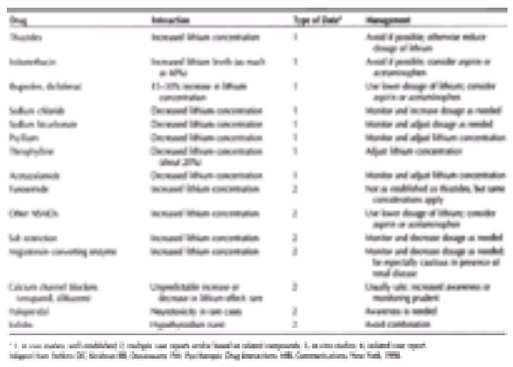

Table 31.2-7 Selected Lithium Drug Interactions

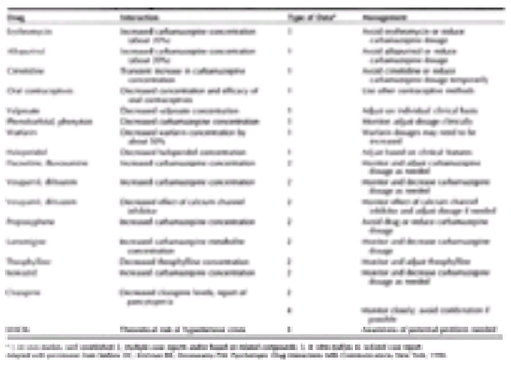

Table 31.2-8 Selected Carbamazepine Drug Interactions

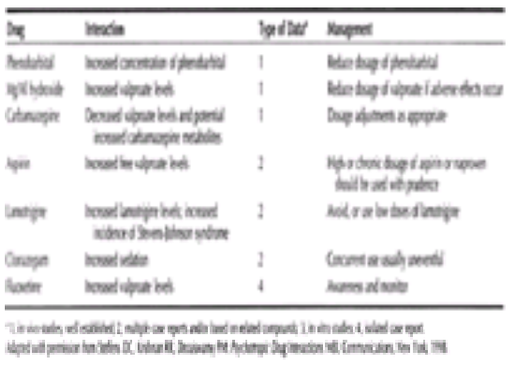

Table 31.2-9 Selected Valproate Interactions

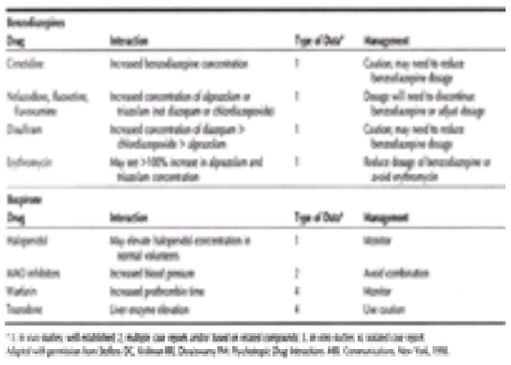

Table 31.2-10 Selected Anxiolytic Drug Interactions

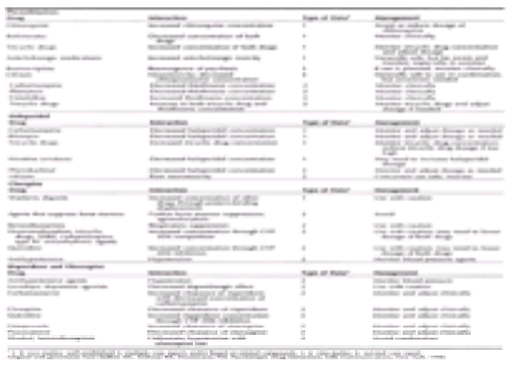

Table 31.2-11 Selected Antipsychotic Drug Interactions

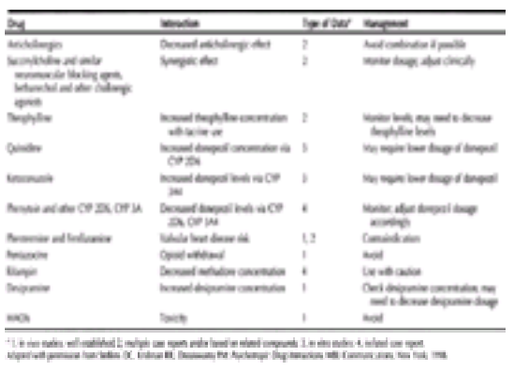

Drug	Interaction	Type of Data*	Management
Anticholinergics	Decreased anticholinergic effect	2	Avoid combination if possible
Succinylcholine and similar neuromuscular blocking agents, bethanechol and other cholinergic agonists	Synergistic effect	2	Monitor dosage; adjust clinically
Theophylline	Increased theophylline concentration with fluoxetine use	2	Monitor levels; may need to decrease theophylline levels
Quinidine	Increased donepezil concentration via CYP 2D6	3	May require lower dosage of donepezil
Ketoconazole	Increased donepezil levels via CYP 3A4	3	May require lower dosage of donepezil
Phenytoin and other CYP 2D6, CYP 3A	Decreased donepezil levels via CYP 2D6, CYP 3A4	4	Monitor; adjust donepezil dosage accordingly
Pimozide and fluvoxamine	Valvular heart disease risk	1, 2	Contraindication
Pentazocine	Opioid withdrawal	1	Avoid
Rifampin	Decreased methadone concentration	4	Use with caution
Desipramine	Increased desipramine concentration	1	Check desipramine concentration; may need to decrease desipramine dosage
MAOIs	Toxicity	1	Avoid

*1, in vivo studies: well-established; 2, multiple case reports and/or based on related compounds; 3, in vitro studies; 4, isolated case report.
Adapted with permission from Jenkins DC, Krishnan RR, Drug Interactions With Psychotropics. Drug Interactions 1988. Eisenmann Lecture, New York, 1998.

Table 31.2-12 Miscellaneous Drug Interactions

SUGGESTED CROSS-REFERENCES

Chapter 11 discusses substance-related disorders as well as commonly used substances such as alcohol, caffeine, and nicotine, all of which may affect drug metabolism. Section 51.2b and Section 51.4b discuss the physiological aspects of aging and the psychopharmacology of elderly patients, both of which consider related alterations in drug metabolism.

SECTION REFERENCES

Ameer B, Weintraub RA: Drug interactions with grapefruit juice. Clin Pharmacokinet *33:* 103, 1997.

*Barone JA, Byerly WG: Determination of bioequivalence of psychotropic drugs and concerns involving product interchange. J Clin Psychiatry *47* (Suppl):28, 1986.

Bertilsson L, Dahl ML, Tybring G: Pharmacogenetics of antidepressants: Clinical aspects. Acta Psychiatr Scand *391* (Suppl):14, 1997.

Finley PR, O'Brien JG, Coleman RW: Lithium and angiotensin-converting enzyme inhibitors: Evaluation of a potential interaction. J Clin Psychopharmacol *16:* 68, 1996.

Fraser AG: Pharmacokinetic interactions between alcohol and other drugs. Clin Pharmacokinet *33:* 79, 1997.

Goff DC, Baldessarini RJ: Drug interactions with antipsychotic agents. J Clin Psychopharmacol *13:* 57, 1993.

Greene DS, Barbhaiya RH: Clinical pharmacokinetics of nefazodone. Clin Pharmacokinet *33:* 260, 1997.

*Gregg CR: Drug interactions and anti-infective therapies. Am J Med *106:* 227, 1999.

Hansten PD: Understanding drug-drug interactions. Sci Med January/February:16, 1998.

Janicak PG: The relevance of clinical pharmacokinetics and therapeutic drug monitoring: Anticonvulsant mood stabilizers and antipsychotics. J Clin Psychiatry *54* (Suppl):35, 1993.

*Janicak PG, Davis JM, Preskorn SH, Ayd FJ Jr: *Principles and Practice of Psychopharmacotherapy*, ed 2. Williams & Wilkins, Baltimore, 1997.

Janicak PG, Javaid J, Leach A, Sharma RP, Dowd S, Davis JM: A two phase double-blind randomized study of three haloperidol plasma levels for acute psychosis with reassignment of initial nonresponders. Acta Psychiatr Scand *95:* 343, 1997.

Janicak PG, Winans B: The laboratory in clinical psychiatry. In *Review of Psychiatry*, vol 16, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1997.

Jefferson JW: Drug interactions-friend or foe? J Clin Psychiatry *59* (Suppl):37, 1998.

*Ketter TA, Flockhart DA, Post RM, Kenicoff K, Pazzaglia PJ, Marangell LB, George MS, Callahan AM: The emerging role of cytochrome P450-3A in psychopharmacology. J Clin Psychopharmacol *15:* 387, 1996.

*Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry *153:* 311, 1996.

*Preskorn S: Clinically relevant pharmacology of selective serotonin reuptake inhibitors. An overview with emphasis on pharmacokinetics and effects on oxidative drug metabolism. Clin Pharmacokinet *32:* 1, 1997.

*Riesenman C: Antidepressant drug interactions and the cytochrome P450 system: A critical appraisal. Pharmacotherapy *15:* 845, 1995.

Shader RI, Moltke L, Schmider J, Harmatz J, Greenblatt D: The clinician and drug interactions: Update. J Clin Psychopharmacol *16:* 197, 1996.

Shoaf SE, Linnoila M: Interaction of ethanol and smoking on the pharmacokinetics and pharmacodynamics of psychotropic medications. Psychopharmacol Bull *27:* 577, 1991.

Taylor D, Lader M: Cytochromes and psychotropic drug interactions (Editorial). Br J Psychiatry *168:* 529, 1996.

*Tseng AL, Foisy MM: Significant interactions with new antiretrovirals and psychotropic drugs. Ann Pharmacother *33:* 461, 1999.

Winans E, Cohen L: Assessing the clinical significance of drug interactions in psychiatry. Psychiatr Ann *28:* 399, 1998.

31.3 DRUG DEVELOPMENT AND APPROVAL PROCESS IN THE UNITED STATES

PAUL LEBER, M.D.

Before a new drug may be legally marketed in the United States, it must, under the requirements of the Federal Food, Drug and Cosmetic (FDC) Act, a national drug-regulatory law, be evaluated in applicable preclinical and clinical tests and be found by the Food and Drug Administration (FDA), based upon a review of the results of those tests, to be both “safe for use” and “effective in use” under the conditions of use that are recommended in its proposed or approved labeling. This chapter reviews selected aspects of the history, requirements, and enforcement of the Act.

FEDERAL GOVERNMENT INVOLVEMENT IN THE REGULATION OF DRUGS INTENDED FOR HUMAN USE

Government regulation of drug development and marketing is vital to the safety and effectiveness of the armamentarium. However, this was not always the case. Well into the twentieth century the pharmaceutical industry and its political allies were able to defeat or delay the passage of measures granting the federal government the authority to regulate the marketing of new drugs. A historical account of the development of the government's regulatory powers not only serves to explicate the aims of drug regulation, but testifies compellingly to the necessity for a federal drug regulatory system.

Although a federal law banning the importation of adulterated drugs was enacted in 1848, historians generally view the federal government's regulation of the drug industry to have begun with the passage of the Pure Food and Drugs Act of 1906. Enacted in an era of public concern that followed a series of sensationalistic revelations in the press about unscrupulous practices and deplorable conditions extant in the food and drug industries, the Pure Food and Drugs Act gave the federal government the authority to take legal action against the marketers of adulterated or misbranded drug products.

Enforcement of the 1906 Act proved difficult, however, because it required the government to assume the burden of proving that a product was adulterated or that its labeling was “false or misleading in some particular.” The latter proved exceedingly difficult to do in any case that turned on the opinions of experts rather than on findings of fact.

One factor undermining the utility of the 1906 Act as a tool to control false and misleading health claims was the conservative manner in which the judiciary of the time interpreted its provisions. In a celebrated Supreme Court case involving a product advanced as a cure for cancer, Chief Justice Oliver Wendell Holmes, Jr., offered the view that Congress could not have intended the Act's ban against the marketing of products bearing labeling that was “false or misleading in any particular” to apply to more than the accuracy of statements concerning the identity, purity, and strength of the product's chemical components. Holmes justified his interpretation on the grounds that an application of the Act's provisions to matters involving the accuracy of a claimed therapeutic use would require that the staff of the FDA be competent to assess medical claims, and this, Holmes opined, was so far beyond their talents and competency that Congress, in its wisdom, could not possibly have intended it.

Although Congress subsequently attempted to overcome this limitation by explicitly amending (Sherley Amendment of 1911) the 1906 Act to include fraudulent therapeutic claims among those that would cause a product to be misbranded, the action had little practical effect. Under the amended 1906 Act, the government was obliged to establish the existence of fraudulent intent on the part of a sponsor advancing a false and misleading therapeutic claim and this proved to be even more difficult than documenting that a claim was false. Thus, the 1906 Act, a relatively weak measure to begin with, turned out to be a far less effective antiquackery tool than those who had fought for its passage intended.

However, subsequent efforts of FDA officials and their allies to obtain the passage of more potent drug legislation met with repeated failures, in no small measure because of the political power of the regulated industry. Indeed, it is widely believed that despite the sweeping social and political changes wrought by Roosevelt's New Deal, the original version of the current drug regulatory law, the FDC Act, would have been unlikely to have been enacted into law had it not been for the disastrous consequences of an error made in 1937 by a chemist working for a respected, long-established, pharmaceutical house.

In an attempt to satisfy an unmet market demand for a liquid preparation of a sulfanilamide antibiotic, the firm's chief chemist prepared a formulation that employed diethylene glycol as the means to solubilize the antibiotic. Unknown to the chemist, diethylene glycol was a renal toxin, a fact that might have been discovered had the sponsor bothered to carry out even rudimentary premarket safety tests. Ironically, the product had undergone premarket testing, but only for fragrance, appearance, and palatability. Thus, when the elixir was put into commercial distribution, the consequences were calamitous.

In retrospect, it was estimated that upwards of 107 individuals died needlessly and painfully as a direct consequence of the firm's negligence. The toll in lives might have been far greater had not the FDA been able to use the limited powers it had under the 1906 law to recover all but 0.8 percent of the 240 gallons of the elixir that had been distributed throughout the country. Although the FDA's success in limiting the extent of the injury done by the firm's recklessness was widely acclaimed, so too was the irony that the FDA's authority to intervene had been based on a technicality. Under the 1906 act, the FDA had no authority to prevent the interstate distribution of a manifestly dangerous drug. Accordingly, it had only been able to intervene because the so-called elixir was misbranded. Elixirs are solutions in which alcohol serves as the solvent; accordingly, the product being distributed by Massengill which employed diethylene glycol as the solvent was misbranded under the provisions of the 1906 Act.

The story of the elixir of sulfanilamide tragedy illustrates how an unexpected event that heightens public anxiety, even if only transiently, can have far-reaching and long-lasting consequences for federal law and government policy. In 1938 public outrage about the elixir disaster, coupled with knowledge that the entire episode could have been prevented and fear that it might be repeated virtually compelled Congress to replace the 1906 act with one that had been pending before it and bottled up in committee, for more than 5 years.

Law-making is not an entirely coherent nor logical process, turning as much if not more on the political exigencies of the moment as on philosophical convictions, scientific principles, or sound analyses. As a wag once advised, “those who have a liking for either sausage or law should avoid observing either being made.”

It follows that authority gained as a consequence of public anxiety can just as readily be lost because of it, a point that has not been missed by those who are opposed to federal controls over drug development and marketing. Much of current antiregulatory rhetoric strives, for example, to heighten public anxieties by linking the current lack of effective treatments for a wide variety of life-threatening and devastating conditions (e.g., acquired immune deficiency syndrome [AIDS], cancer, dementia, schizophrenia) to the adverse effects of government regulation on incentives for new drug development.

In any event, the net value of drug regulation notwithstanding, the FDC Act of 1938 was seminal, creating a system for the premarket testing and clearance of new drugs that endures, albeit in slightly modified form, to the present time.

In passing the act, Congress relied upon the federal government's authority to control interstate commerce. The 1938 act obliged those planning to market new drugs to submit their products for pre-market clearance by making it a violation under the act to introduce into interstate commerce a new drug that was not the subject of an effective New Drug Application (NDA). To obtain an effective NDA, a sponsor of a new drug was to submit an application containing reports of all tests reasonably applicable to show that the drug product met the requirements established by the act for a marketed drug. Among other requirements, the act specified that the reports would have to show that the drug was not only what its labeling claimed it to be in regard to its strength, identity, and purity, but that it would be “safe for use”

under the conditions of use recommended in its proposed or approved product labeling.

On its part, the FDA was obliged by the act to accept a sponsor's application and to allow it to become effective, unless upon review, the agency determined that reports it contained failed to show that the drug met the act's requirements.

The 1938 act provided an exemption, now known as an Investigational New Drug (IND) application, to allow experts “qualified by training and experience to assess the safety of drugs” to obtain supplies of unapproved new drugs to conduct clinical experiments. An exemption from the act's provisions was required if there was to be research with unapproved new drugs because the Act banned the introduction into interstate commerce of any drug for which an NDA was not in effect. This investigational exemption was essentially available for the asking and drug distribution under the exemption was not closely monitored by the FDA, a fact that was to have important political consequences 25 years later.

Although the 1938 act was a major advance, it was not without limitations. One glaring weakness was its failure to require that marketed drugs be shown to be effective in use. In fairness, in 1938 a regulatory demand that a product be shown effective prior to marketing would have been exceedingly difficult to implement. Although individual academics of the time were mindful that uncontrolled clinical observations were an unreliable basis upon which to assess the efficacy of drug products, especially those that lacked an immediate and dramatic effect (e.g., treatments for tuberculosis or cancer), the medical community at large would have been unlikely to have shared that view. Moreover, from a practical perspective there was no clear consensus extant in 1938 as to the methods that might be employed to assess drug efficacy. It was not until the mid-1950s that prominent authorities in pharmacology and medical therapeutics began more or less uniformly to assert that reliable evidence of drug efficacy could only be gained in valid controlled clinical trials.

The shift in views among medical authorities about the nature of the evidence required to document the effectiveness of drugs did not, however, lead to an immediate change in the 1938 act's requirements. Advocates seeking to strengthen the provisions of the 1938 act did not meet with success until the political equilibrium was again perturbed by public alarm arising in the wake of yet another drug-caused calamity.

In 1961 thalidomide, a drug that had been widely and successfully marketed as a hypnotic in Europe since 1957, was determined to be the cause of limb reduction defects in children born to mothers who had used it during their pregnancies. The news about thalidomide's teratogenic potential broke during the time a series of Congressional hearings were being held on the drug industry and its alleged price-fixing practices. Although thalidomide had not been marketed in the United States, it had been distributed relatively widely to physicians under the Act's investigational exemption provisions. The distribution was more promotional than investigatory, intended largely to draw attention to thalidomide's pending marketing. The marketing of thalidomide had been delayed for several months, however, because Frances Kelsey, the FDA medical officer assigned to the NDA, had unresolved concerns about the drug's potential to cause peripheral neuritis.

Publicity about thalidomide's capacity to cause limb reduction defects and the agency's refusal to allow the product to be marketed because of concerns about its toxicity clearly affected Congressional attitudes about the need for reform and revision of the provisions of the 1938 act. Although only a handful of American children actually suffered injury as a direct result of the domestic distribution of thalidomide, the narrowness of the nation's escape from yet another drug-linked disaster is widely credited with securing the passage of the 1962 Kefauver-Harris amendments.

The 1962 amendments revised the 1938 act in several important ways, two of which are of major interest. First, under the revised act, marketed drugs were required not only to be safe for use, but to be effective in use. Specifically, the amended act instructed the FDA to disapprove an NDA if, upon review, the agency determined that there was a lack of “substantial evidence” of the drug's effectiveness in use.

Second, the 1962 amendments gave the agency explicit authority to control and monitor the investigational use of unapproved new drugs. Under the FDC Act's revised IND provisions, the Secretary was allowed to link the granting of investigational exemptions to (1) the adequacy of preclinical tests justifying the use of the investigational drug in humans, (2) the adequacy of procedures that would be employed to control the distribution and disposition of supplies of the investigational drug, (3) the promise of IND holders to maintain records of the results of their clinical investigations and the making of reports of these results to the agency, and (4) the promise of investigators to obtain informed consent from research subjects, except in those circumstances in which “. . . [the sponsor or his representative]. . . deem it not feasible or, in their professional judgment, contrary to the best interests of such human beings.

Although the FDC Act has subsequently been amended on a variety of occasions and in a number of important ways (e.g., the Patent Term Restoration and Generic Drug Act, the Prescription Drug User Fee Act, the FDA Modernization Act of 1997) the 1962 amendments remain the most important set of amendments to the 1938 act. Indeed, with minor exceptions, the structure, function, and operation of the current drug regulatory system is based upon those amendments.

IMPLEMENTATION AND ENFORCEMENT OF THE ACT

The text of federal laws typically provide little more than an outline of Congressional aims and intent. Accordingly, the agency that has responsibility for the enforcement of a federal law must develop and promulgate regulations, that is, detailed rules and procedures, for its actual enforcement.

ENFORCEMENT OF THE ACT'S INVESTIGATIONAL NEW DRUG REGULATIONS

The conditions under which INDs may be granted, and the requirements that must be met by sponsors and sponsor-investigators to conduct clinical trials under their provisions are detailed in the agency's IND regulations, which appear in Title 21 in the Code of Federal Regulations (21 CFR 312).

An IND is required of any person or persons intending to conduct clinical research with an unapproved new drug within the United States. An IND is ordinarily not required for clinical investigations of marketed drug products, however, unless the sponsor of the study intends to rely upon its results for some regulatory purpose (e.g., to seek a new claimed use for the product.).

Agency regulations on INDs are written as if all who conduct clinical research with unapproved new drugs do so with the intent to develop the information and evidence necessary to gain approval of an NDA for the product. Accordingly, the regulations describe the requirements for clinical testing in reference to three more or less sequentially ordered, phases of premarket testing. Phase 1 is intended to study a drug's clinical pharmacology and to identify its common toxicities. In phase 2 it is contemplated that the sponsor will conduct carefully controlled clinical studies to document the product's effectiveness. In phase 3 clinical testing is to be conducted under conditions that more closely mimic those likely to obtain when the drug is marketed. Throughout each phase of a drug's development, the requirements that must be met to support clinical testing are intentionally limited to those deemed essential to protect human subjects from unreasonable, unnecessary, and potentially avoidable risks.

The preclinical tests required to support initial, short-term studies of a drug in humans are limited to those necessary to establish that the drug, administered at the levels and for the duration proposed, will be unlikely to cause death, catastrophic events, or serious and lasting damage to organ structure or function. Although a more comprehensive set of preclinical toxicity tests must ordinarily be completed prior to the conduct of large-scale phase 2 and 3 trials, their purpose is not to guarantee the safety of human subjects participating in these trials, but to identify and develop the means to monitor the function and integrity of organ systems that may be uniquely susceptible to injury by the drug.

Positive findings of in vitro and in vivo animal teratogenicity and carcinogenicity studies are an exception. Because clinical experience is almost never a reliable means to assess a drug's oncogenic and teratogenic risk, the positive results of such studies are ordinarily considered definitive. However, evidence that a drug is teratogenic or tumorigenic in animals only rarely leads to a decision to preclude further development. Typically, such findings affect how and in whom the drug is tested and, if and when the drug is marketed, are invariably described in its product labeling. Preclinical safety tests serve primarily to guide clinical drug development; regulatory conclusions about the safety or lack of safety of an investigational drug are based almost entirely on clinical experience.

However, serious toxicity findings in animal tests are not ignored, and typically lead to the (1) conduct of additional preclinical studies (i.e., to determine the mechanism by which the drug causes injury—it may not be operative in humans), and (2) intensified monitoring of human subjects during and immediately following their exposure to the investigational drug.

Although the sponsor and the agency often agree about the nature of the remedial steps that should be taken when a potential drug-related risk is identified in either preclinical or clinical tests, disagreements can arise. If a disagreement cannot be resolved through negotiations with a sponsor, agency officials are authorized to

place a clinical “hold” on the sponsor's ongoing studies with the drug. However, such holds may only be imposed under certain specified conditions. Specifically, before imposing a clinical hold an agency official must conclude that the circumstances extant meet one or more of the conditions described at 21 CFR 312.42 (b) of the IND regulations. Of note, all but one of the five enumerated conclusions are linked to concerns about the safety of human subjects. The sole exception applies when a clinical trial intended to gain evidence of a drug's effectiveness in use is determined to be “. . . clearly deficient in design to meet its stated objectives.”

To ensure that agency scientific staff have sufficient time to evaluate all information relevant to the first use of a new investigational drug, sponsors who apply for an IND are obliged to agree not to initiate clinical testing under their IND until 30 days elapse after the time of its receipt by the agency. If agency officials do not impose a hold within these 30 days, the sponsor may initiate clinical testing. Even if it does not impose a hold, the agency may offer suggestions and comments about the conduct of clinical testing; typically, these are communicated to the sponsor in writing.

IND regulations also impose ongoing monitoring and reporting requirements and obligations in regard to clinical research conducted under active INDs. Not only must sponsors of IND make periodic reports to the agency about the status of their work with the investigational drug, but they must report, in a timely manner as prescribed by regulation, new information bearing on adverse preclinical and clinical experiences that have occurred in association with the investigational agent's use. Sponsors are also obliged to submit, in advance of the initiation of any new clinical investigation, a protocol describing how that trial will be carried out. It is through the operation of these and other requirements that the agency monitors, and, as required, can act to modify the course of a new investigational drug's testing and development.

ENFORCEMENT OF THE ACT'S NDA REGULATIONS

In passing the FDC Act Congress sought to ensure that every drug lawfully marketed within the United States would, prior to marketing, have been shown to be both safe for use and effective in use under the conditions of use described in its product labeling. To secure this result, Congress instituted the system of premarket clearance administered by the FDA. The act, however, describes in only very broad brush strokes, how that system is to operate.

Although the act enumerates four findings of an agency review of an NDA that justify a conclusion that a drug has not been shown to be “safe for use,” (i.e., the extent of testing for safety is inadequate, the findings of tests performed show the drug to be unsafe, the findings of tests performed fail to show that the drug is safe, there is insufficient information in general to reach a conclusion about the drug's safety), the act provides no guidance at all concerning the identity of the specific tests that sponsors should employ to assess the safety of a drug.

While this might seem at first sight to be a shortcoming, it is not. To the contrary, Congress would have been ill-advised to offer instruction about a technical subject lying outside its expertise. Accordingly, in its wisdom, Congress delegates to the FDA the task of identifying the tests most appropriate for securing its intent. This strategy not only ensures that the choice of methods is guided by appropriately informed experts, but that the methods used are those consistent with the best, most up-to-date scientific practice. Had Congress specified precise methods for the assessment of drug safety set by statute, the requirements of the act would have rapidly become outdated.

There are instances, however, where Congress is specific about its intent. The definition of substantial evidence provided in the FDC Act is a case in point:

> . . . evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or the proposed labeling thereof.

The definition makes clear that regulatory determinations concerning the effectiveness of a drug must derive from evidence adduced in valid experiments, in particular, clinical experiments. Admittedly, the definition allows a role for medical experts, but only to the extent that their expertise in the management of patients with a disease speaks to their capacity to offer not only a fair and responsible but also an informed opinion about the meaning of the evidence adduced. The definition of substantial evidence gives no weight to the views of experts in regard to their beliefs or their impressions gained in the course of clinical practice.

Although the definition of substantial evidence provided in the Act espouses the highest principles of epistemology and the scientific method, the definition is not without limitations from a regulatory enforcement perspective. What, precisely, is intended by the phrase “adequate and well-controlled clinical investigations?” What did Congress intend, if anything, by its use of the plural of the word “investigations?” The agency has long interpreted it to reflect Congressional intent to require independent replications and corroborations of experimental results bearing on drug efficacy. Although this interpretation is entirely consistent with the use of the word, it is hardly as explicit a defense of the principle as those who champion the scientific method would prefer. In 1997 Congress clarified its intent, amending the FDC Act (FDA Modernization Act of 1997) to allow the agency in selected circumstances to conclude that substantial evidence exists when results from but one adequate and well-controlled clinical investigation are provided. The agency has yet to revise its regulations to conform to this change, but a document promulgated in March of 1997 (Evidence document), indicates that reliance on the results of a single adequate and well-controlled clinical investigation will continue to be the exception rather than the rule. In large part, an agency's regulations are intended to clarify potential ambiguities of the sort just described. Unfortunately, it is not always evident to those drafting regulations, especially when drafting those written in the immediate aftermath of an act's passage, just what in the words of a statute will or will not prove ambiguous.

To illustrate, it soon became apparent following the passage of the 1962 amendments that the agency would have to enumerate explicitly what attributes a clinical trial must possess to be considered adequate and well controlled. The result of that effort now appears at 21 CFR 314.126.

There are still other cases, however, where the need for clarifying regulations would clearly be helpful but are not provided. One reason is that it is often impossible to draft meaningful regulations. The set of adverse effects reported in association with the use of a new drug during its premarketing development that would regularly lead to a regulatory determination that the drug is “unsafe for use” is a case in point. The findings that lead to a regulatory determination that a drug is safe or unsafe for use derive much less from objective evidence-based judgments than from opinions, albeit those of medical professionals reasonably experienced and competent in the management of the condition that the investigational drug is intended to treat. In contrast to a regulatory effectiveness determination that turns on the finding of a statistically significant difference favoring a new drug over a control in more than one valid experiment, a conclusion that a drug is safe for use derives from a complex and inchoate process that requires experts to weigh several factors simultaneously, including not only the severity and incidence of the harms and injuries associated with the use of a drug, but the magnitude of the drug's therapeutic effects (which are, at best, only imprecisely and imperfectly estimated), the severity and the natural history of the condition under treatment and the availability, if any of alternative treatments.

It is also important to be mindful that agency regulations are written at a level that applies to all therapeutic drug classes. To be clear, agency policy and guidance on specific psychotherapeutic indications are available, but they do not have the force of regulations; however, this does not present a practical problem. At least in regard to issues of clinical trial design, conduct, and analysis, all product classes can be considered identical. Similarly, the methods and strategies used to assess the safety of a drug in use are indistinguishable in form, whatever the therapeutic class involved. Admittedly, the methods and techniques used to assess, collect, and classify mental phenomena are discipline specific, but the strategies for their analysis are not.

To the extent that discipline-specific enforcement of the FDC Act is necessary, it is available because the authority to make decisions regarding the development of psychotherapeutic drug products is delegated to clinicians and scientists trained and experienced in the neurosciences, psychiatry, and psychology as well as to those with expertise in matters pharmacological. Thus, the assessment of INDs and NDAs for products with psychotherapeutic indications are carried out by reviewers and supervisors with expertise and skills relevant to the task.

COMMERCIAL DEVELOPMENT OF PSYCHOTHERAPEUTIC DRUG PRODUCTS UNDER AN IND

A sponsor seeking to develop a commercial drug product for a psychotherapeutic indication submits an IND to the Division of Neuropharmacological Drug Products. The information provided in the initial IND is assessed by a division team comprised of chemists, pharmacologists, and clinicians and their supervisors. Assisted as required by biopharmacokineticists and statisticians, the team seeks not only to determine whether the information submitted in the IND is sufficient to support the initiation of the clinical study proposed, but provides feedback to the sponsor about relevant issues and findings potentially affecting the product's further development. A decision is required within 30 days; accordingly, division scientists reviewing the IND have less than 20 working days to complete their initial assessment.

Throughout the subsequent course of a drug product's development under an IND, the Division staff and consultants seek to ensure that the extent and scope of clinical testing is appropriately supported by the extent and adequacy of preclinical test and their results. This effort is facilitated by agency regulations that require the sponsor to make periodic reports concerning the progress of the development effort. In particular, sponsors are obliged to report, within a time frame dictated by their severity, information concerning drug-associated adverse events. Although there is no formal requirement that a sponsor make reports to the IND of analyses of efficacy data obtained from clinical trials, protocols for all clinical trials are submitted to the agency prior to their initiation. Because the ultimate goal of a commercial sponsor is to collect evidence that can be used to gain approval of an NDA, sponsors typically not only keep the Division apprised of the results of clinical studies completed under their IND, but seek the Division's advice in regard to the design, conduct, and plans for analyses of all major clinical trials that might serve as a source of support for approval of an NDA. The extent of agency-sponsor cooperation in the 1990s is quite extensive, extending to almost every aspect of a drug's preclinical and clinical evaluation including, in particular, the characterization of its pharmacokinetic and pharmacodynamic attributes.

It is noteworthy that, despite the joint efforts of agency and sponsor, only a small proportion of drug products for which INDs are submitted eventually become the subject of an NDA submission. This is not surprising because any number of events may occur that cause a sponsor to abandon a drug development effort. A drug being developed may prove to be too toxic or too difficult to use. Still, in other instances a drug may be found, the positive findings of preclinical models and earlier clinical reports notwithstanding, to be ineffective or effective, but only at doses that prove intolerable or unsafe. Not uncommonly, a drug development program is abandoned for reasons that have nothing whatsoever to do with the drug's performance or its pharmacological properties. A commercial sponsor may determine that a changing market has removed virtually all likelihood of the drug product being marketed profitably or a firm may simply lack the funds to pursue development.

As drug development proceeds, the Division of Neuropharmacological Drug Product and firm continue in their efforts to ensure that by the completion of the testing program, all information necessary to support an approvable NDA will be available. The division's staff and the sponsor's representatives typically work out the substantive details of the format and organization of an NDA well before its time of submission.

REVIEW AND EVALUATION OF A COMMERCIAL SPONSOR'S NDA

Once an NDA for a psychotherapeutic drug product has been submitted to the FDA, the Division has 60 days to determine whether or not it is sufficiently complete from a technical perspective to warrant review. This preliminary inspection period is intended to ensure that the FDA's limited resources are not wastefully expended in efforts to review an application that will have to be rejected. Why the review of an application that is disorganized or lacking an accurate index would pose a problem becomes obvious once its physical size is appreciated.

The typical NDA is comprised of hundreds of 3– to 4–inch-thick volumes of standard-sized pages containing a diversity of test results, reports, tables, and analyses. The information contained within these volumes deals with everything from the source of the raw materials used to manufacture the drug substance, to the specifications of the limits on impurities allowed in the drug product, to statistical analyses of clinical data, to descriptions of individual patients and the adverse events that befell them in the course of their exposure to the drug or the agent used as a comparative control.

The formal goal of an FDA review is to determine whether or not the reports submitted show that the drug, on whose behalf the NDA was submitted, meets the standards set by the FDC Act for a marketed drug product. Much of the review, accordingly, actually involves the concurrently conducted systematic review of each report contained in the application by a scientist or clinician who is qualified to evaluate them.

A substantive proportion of the Division's review effort, for example, is carried out by industrial chemists on reports about drug chemistry, manufacturing, and controls. Likewise, reports of preclinical pharmacology and toxicity tests are examined by staff pharmacologists and toxicologists. A team of clinicians and statisticians jointly reviews the reports of all controlled clinical trials capable by design of assessing the efficacy of the drug for the claim or claims advanced by the sponsor. The review of each such study entails a systematic assessment of issues ranging from the means used to evaluate and diagnose subjects, the fidelity to which the study protocol was followed in regard to subject selection and exclusion criteria, randomization, blinding, use of proscribed concomitant medications, and methods of outcome assessment. Subjects prematurely discontinued from clinical trials are closely examined because such “censoring” may, depending upon its kind and causes, so bias estimates of treatment effect adduced in a trial as to render them worthless or uninterpretable. Clinicians, including psychiatrists and internists, also scrutinize reports of abnormal laboratory tests and adverse clinical events in an effort to determine whether these findings are likely to be caused by the new drug, and if so to what extent. The clinical team also seeks to determine the conditions and regimen under which the drug can be used with minimum risk and maximum benefit.

Once these discipline-specific reviews are completed, all the diverse findings of the review team's effort must be assessed to determine whether the benefits likely to be gained from the use of the drug are sufficient to outweigh the risks likely to be associated with its use. This assessment entails a number of subjective judgments and hence any close call on an application is invariably arguable.

In general, current agency policy favors the approval of NDAs for products of documented effectiveness, even in the face of serious risks, provided product labeling can describe those risks fully and accurately. Thus, a major part of a division's review team's effort is devoted to the development of labeling under which it can be concluded, as required by law, that the drug will be both safe for use and effective in use. In this regard it is worth noting that at the time of a product's initial approval, its full panoply of risks, in particular those that occur at relatively low incidence, are often unrecognized.

This limitation in the knowledge of a drug's risks of use is not attributable to the shortcomings of the efforts of either the FDA or the sponsor, but to the relatively small number of patients exposed to a drug during the course of a typical commercial development program. Experience gained with even a few thousand patients is inadequate to ensure that a single example of an event caused by the drug infrequently will be observed. For example, there is a 5 percent chance that an event that affects 1 in every 1000 patients will not be observed even once in a clinical development program involving 3000 patients. While a risk of 0.1 percent may seem small enough from the perspective of the individual patient, it is quite large from a public health perspective. For example, a drug that causes hemolytic anemia at that rate would, if administered to 100,000 patients, be expected to cause 100 cases. In theory, depending on the nature of the illness being treated, this incidence might be deemed acceptable from a regulatory perspective so long as the product labeling gives adequate warning of the risk. Regulatory acceptance of such a product is not enough, however; a sponsor might, the medical value of the drug notwithstanding, be obliged to withdraw such a product from marketing because of the tort liability considerations.

DRUG REGULATION AFTER THE APPROVAL OF AN NDA

Those who drafted the Act were obviously mindful of the fact that the risks of a drug might not be discovered until after it has been marketed. Accordingly, the sponsor of a marketed drug continues to have an obligation to report all new information bearing on the safety of the product to the NDA file. This requirement of the Act is the basis for the FDA's authority to require sponsors to collect, classify, and summarize reports made to them of adverse events associated with the use of their drugs to the FDA. Physician and other health care practitioners may report adverse clinical experiences directly to the FDA as well.

The role a drug product actually plays in the genesis of an untoward clinical or laboratory event reported in association with its use is exceedingly difficult to ascertain with any degree of certainty, a fact often overlooked by the lay media and even by medical professionals. Nevertheless, although it is far from a perfect science, the analysis of spontaneous reports made by practitioners is an FDA responsibility, one in which the division staff participates in association with colleagues from the agency's Office of Epidemiology and Biostatistics. The FDA is responsible for ensuring that marketed drug products are, as required by the FDC Act, safe for use and effective in use under the conditions of use recommended in their approved labeling and the federal government bears the responsibility for ensuring the quality and reliability of drugs, and how that responsibility is met through the operation of the Act.

SUGGESTED CROSS-REFERENCES

Section 31.1 discusses the general principles of psychopharmacology. Drugs and pregnancy are discussed in Section 28.2 on psychiatry and reproductive medicine. The biological therapies are discussed throughout Chapter 31.

SECTION REFERENCES

*Jackson CO: *Food and Drug Legislation in the New Deal*. Princeton University Press, Princeton, NJ, 1970.

*Leber PD: FDA: The federal regulation of drug development. In *Psychopharmacology: The Third Generation of Progress*, HY Meltzer, editor. Raven, New York, 1987.

Leber PD: Hazards of inference: The active control investigation. Epilepsia *30* (Suppl):S57, 1989.

Leber PD: Postmarketing surveillance of adverse drug effects. In *Adverse Effects of Psychotropic Drugs*, J Lieberman, J Kane, editors. Guilford, New York, 1992.

*Leber PD: The role of the regulator in the evaluation of the acceptability of new drug products. In *Psychotropic Drug Development*, D Healy, DJ Doogan, editors. Chapman & Hall, London, 1996.

*Leber PD: Slowing the progression of Alzheimer's disease: Methodological issues. Alzheimer Dis Assoc Disord *11*(Suppl):S10, 1997.

*Silverman M, Lee PR: *Pills, Profits and Politics:* University of California Press, Berkeley, 1974.

*Wermeling DP: Clinical research: Regulatory issues. Am J Health Syst Pharm *56:* 252, 1999.

31.4 MEDICATION-INDUCED MOVEMENT DISORDERS

EDMOND HSIN-TUNG PI, M.D., AND GEORGE M. SIMPSON, M.D.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) included a new diagnostic category, medication-induced movement disorders, and listed research criteria for further study in an appendix. This recommendation should promote early detection and help to develop well-defined and effective clinical psychopharmacological approaches for these disorders.

Since the first neuroleptic, chlorpromazine (Thorazine), was introduced in 1952, a variety of neuroleptic drugs have been discovered. While their relative potency to block the dopamine type 2 (D_2) receptors differed, they nonetheless all possess an antipsychotic property and affect the D_2 receptors in the central nervous system, particularly the basal ganglia. Thus they are associated with extrapyramidal adverse effects as well as tardive dyskinesia. These adverse effects may be confounded with negative symptoms of psychosis (e.g., blunted affect, avolition, alogia, and asociality), which can further impair patients' psychosocial function and can also interfere with medication compliance and other aspects of treatment. Medication noncompliance secondary to extrapyramidal averse effects is a major cause of relapse of psychoses and adversely affects the prognosis. Thus, these adverse effects represent a significant clinical issue that limits the effectiveness of antipsychotic drugs in the treatment of psychiatric patients, more so the seriously and persistently ill.

The term *neuroleptic* (“to take the neuron”) has been used broadly to refer to medications including both the “typical” and “atypical” neuroleptic agents. Typical conventional neuroleptics (dopamine receptor antagonists) refers to those medications (e.g., chlorpromazine, fluphenazine [Prolixin], haloperidol [Haldol]) that diminish or abolish positive psychotic symptoms more than negative symptoms and also produce extrapyramidal effects. Atypical neuroleptics (predominantly serotonin-dopamine antagonists) refers to those medications (e.g., clozapine [Cloxaril]) that diminish or abolish psychotic symptoms but do not produce or produce fewer extrapyramidal effects and do not fit classic behavioral or biochemical models for neuroleptic drugs such as catalepsy and stereotypic behavior in animals. The absence of or presence of few extrapyramidal adverse effects is perhaps related to improvement in both negative and positive symptoms of psychosis. Serotonin (5-hydroxytryptomine [5-HT]) type 2 [5-HT_2] receptor antagonistic effects are thought to play a major role in mitigating extrapyramidal adverse effects with atypical neuroleptic drugs including clozapine, olanzapine (Zyprexa), quetiapine (Seroquel), risperidone (Risperdal), and sertindole (Serlect). Risperidone has dose-related extrapyramidal adverse effects above the dosage range of 2 to 6 mg a day. Other medications that also have dopamine-blocking properties and can cause extrapyramidal adverse effects are dopamine receptor antagonists drugs used in the treatment of nausea, such as prochlorperazine (Compazine) or metoclopramide (Reglan). Amoxapine (Asendin), which is marketed as an antidepressant, is chemically related to loxapine (Loxitane) and has D_2 receptor blocking properties. Amoxapine can produce akathisia, cogwheel rigidity, and tardive dyskinesia.

The most common neuroleptic-induced movement disorders are parkinsonism, acute dystonia, and akathisia. In terms of reversibility, the extrapyramidal adverse effects are, in large part, diminished or abolished by administering antiparkinsonian agents and discontinuing or lowering the dose of the neuroleptic. Tardive dyskinesia is a late-onset movement disorder associated with neuroleptic treatment and can be irreversible.

NEUROLEPTIC-INDUCED PARKINSONISM

DSM-IV describes the disorder as:

> [p]arkinsonian tremor, muscular rigidity or akinesia developing within a few weeks of starting or raising the dose of a neuroleptic medication (or after reducing a medication used to treat extrapyramidal symptoms).

The essential feature of neuroleptic-induced parkinsonism is the presence of muscular rigidity, akinesia, and tremor. Muscular rigidity and akinesia occur more frequently than tremor. Associated symptoms and signs may include worsening of negative symptoms of schizophrenia, depressive-like manner, dysphoria, slow speech, bradyphrenia (slow thinking), excessive salivation, drooling, difficulty in swallowing or drinking, shuffling gait, micrographia, seborrhea, and positive glabellar tap. The clinical picture is indistinguishable from that of idiopathic parkinsonism. Differential diagnosis includes idiopathic or other neurological condition–induced parkinsonism and other psychiatric conditions with bradykinetic features.

Muscular rigidity may affect any skeletal muscle, for example, the extremities, neck, shoulders. There are two types of muscular rigidity: continuous (“lead-pipe”) and cogwheel (rhythmic, ratchet-like resistance). Akinesia is a lack of spontaneous motor activity, difficulty in initiating movements, a stooped and shuffling gait, and slow monotonous speech. Depressive-like manner with dysphoria described as akinetic depression should be differentiated from primary depression, demoralization, and residual schizophrenic defect. Tremor is a steady, rhythmic oscillation, 3 to 6 cycles per second, more prominent in the resting position. Tremor may affect extremities (unilateral or bilateral), head, jaw, mouth, tongue, or lips (“rabbit syndrome”). Objectively defined rating scales (e.g., Simpson's Neurological Rating Scale) have helped clinicians to assess extrapyramidal effects. At least 50 percent of outpatients receiving long-term neuroleptic therapy develop parkinsonian symptoms or signs at some point in the treatment course. These symptoms usually develop within a few days to a few weeks of starting or increasing the dosage of a neuroleptic medication. Symptoms may also develop insidiously or exacerbate after the reduction or withdrawal of an antiparkinsonian medication (e.g., benztropine [Cogentin], diphenhydramine [Benadryl], trihexyphenidyl [Artane], amantadine [Symmetrel]). Severity of the symptoms tends to remain stable or gradually decrease when maintained on the same dosage of neuroleptic and symptoms are reversible after discontinuation of the neuroleptic. Elderly patients are at the highest risk for neuroleptic-induced parkinsonism. Other risk factors are a coexisting neurological condition and past history of neuroleptic-induced parkinsonism.

The neuropathophysiology of the condition is related to the postsynaptic blockade of D_2 receptors in the basal ganglia, which causes an imbalance of dopamine and acetylcholine. The occurrence of extrapyramidal adverse effects with neuroleptic drugs is related to dosage and relative occupancy of D_2 and muscarinic receptors. It has been reported that extrapyramidal adverse effects are associated with 80 percent or greater blockade of D_2 receptors. On the other hand, only 50 to 75 percent occupancy of D_2 receptors is necessary to attain antipsychotic efficacy.

The prophylactic administration of antiparkinsonian agents remains controversial. For patients with known risks of developing parkinsonism, an antiparkinsonian agent may be prescribed concomitantly with the neuroleptic drug. For others, when signs of parkinsonism appear, first consider reducing the neuroleptic dosage, or start concurrent administration of a low dosage of an antiparkinsonian agent. Assess patients periodically and attempt to taper then discontinue the antiparkinsonian agent. In most cases, antiparkinsonian agents can be withdrawn without recurrence of symptoms. Abrupt withdrawal of antiparkinsonian agents may lead to the recurrence of extrapyramidal adverse effects; it is therefore always recommended that antiparkinsonian agents be decreased gradually. If parkinsonism symptoms do not respond totally, then switch to a neuroleptic that produces less (e.g., thioridazine [Mellaril]) or no (e.g., clozapine) parkinsonism symptoms. Also, serotonin-dopamine antagonists, such as olanzapine, quetiapine, risperidone, and sertindole, which possess a reduced risk for extrapyramidal adverse effects, may be prescribed. A clear relation between high risperidone dosage and the occurrence of extrapyramidal adverse effects has been reported. The DSM-IV criteria for neuroleptic-induced parkinsonism are listed in Table 31.4-1.

A. One or more of the following signs or symptoms has developed in association with the use of neuroleptic medication:
   (1) parkinsonian tremor (i.e., a coarse, rhythmic, resting tremor with a frequency between 3 and 6 cycles per second, affecting the limbs, head, mouth, or tongue)
   (2) parkinsonian muscular rigidity (i.e., cogwheel rigidity or continuous "lead-pipe" rigidity)
   (3) akinesia (i.e., a decrease in spontaneous facial expression, gestures, speech, or body movements)
B. The symptoms in criterion A developed within a few weeks of starting or raising the dose of a neuroleptic medication, or of reducing medication used to treat or prevent acute extrapyramidal symptoms (e.g., anticholinergic agents).
C. The symptoms in criterion A are not better accounted for by a mental disorder (e.g., catatonic or negative symptoms in schizophrenia, psychomotor retardation in a major depressive episode). Evidence that the symptoms are better accounted for by a mental disorder might include the following: the symptoms precede the exposure to neuroleptic medication or are not compatible with the pattern of pharmacologic intervention (e.g., no improvement after lowering the neuroleptic dose or administering anticholinergic medication).
D. The symptoms in criterion A are not due to a nonneuroleptic substance or to a neurological or other general medical condition (e.g., Parkinson's disease, Wilson's disease). Evidence that the symptoms are due to a general medical condition might include the following: the symptoms precede exposure to neuroleptic medication, unexplained focal neurological signs are present, or the symptoms progress despite a stable medication regimen.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 31.4-1 DSM-IV Diagnostic and Research Criteria for Neuroleptic-Induced Parkinsonism

NEUROLEPTIC MALIGNANT SYNDROME

DSM-IV defines this disorder as:

> [s]evere muscle rigidity, elevated temperature, and other related findings (e.g., diaphoresis, dysphagia, incontinence, changes in level of consciousness ranging from confusion to coma, mutism, elevated or labile blood pressure, elevated creatine phosphokinase [CPK]) developing in association with the use of neuroleptic medication.

Neuroleptic malignant syndrome has been reported since 1960, but it is still poorly defined. The syndrome may represent a heterogeneous group of neuroleptic-induced extrapyramidal adverse effects with concurrent fever. One should not automatically diagnose extrapyramidal adverse effects and fever as neuroleptic malignant syndrome. Estimates of the 1-year prevalence of this syndrome in individuals exposed to neuroleptic medications range from 0.02 to 2.4 percent. It is a rare but potentially fatal disorder; mortality rates reported are in the 10 to 20 percent range. No clear evidence implicates any particular typical neuroleptic agent as more or less likely to cause neuroleptic malignant syndrome; atypical neuroleptics (e.g., clozapine and risperidone) have also been reported to be associated with neuroleptic malignant syndrome, but with clozapine it is thought to be rare.

The syndrome tends to develop when neuroleptic treatment is initiated or the dosage is increased, particularly when the dose is high or parenteral. Patients have been reported more likely to be agitated or dehydrated and often to need restraint or seclusion before development of neuroleptic malignant syndrome. The syndrome most often affects young males. Characteristic symptoms include muscular rigidity with elevated body temperature and serum CPK and, in severe cases, myoglobinuria and acute renal failure. The syndrome usually lasts 5 to 10 days after discontinuation of neuroleptic treatment. Differential diagnoses include catatonia, medical condition–induced movement disorders, malignant hyperthermia, and heat stroke.

Prevention is an essential part of managing this heterogeneous condition. The lowest dosage of neuroleptic is recommended, with monitoring of the onset of extrapyramidal adverse effects. Early detection and prompt intervention to eliminate extrapyramidal adverse effects, particularly severe muscular rigidity, may prevent its progress and development of neuroleptic malignant syndrome and its serious complications. Once the syndrome is confirmed, initiate supportive measurements to stabilize autonomic dysfunction, discontinue neuroleptic treatment, continue administering an anticholinergic agent to alleviate muscle rigidity if cardiac status is stable, and institute necessary measurements to lower fever. b-Adrenergic receptor antagonists (beta-blockers) or benzodiazepines may be indicated for akathisia or agitation if there is no clinical contraindication. Amantadine, bromocriptine [Parlodel], and dantrolene [Dantrium] have also been used in some cases. Dopamine agonists are important in patients with a fever over 103°F when a reduction or withdrawal of the anticholinergic agent may be advisable.

Two weeks after the resolution of neuroleptic malignant syndrome, rechallenge with a neuroleptic may be considered. If a neuroleptic drug is clinically indicated, prescribe the lowest dosage with gradual incremental increases plus anticholinergic therapy or prescribe a neuroleptic drug with more anticholinergic property (e.g., thioridazine or an atypical neuroleptic agent). Adequate hydration and well-controlled ambient temperature are effective prophylactic measurements to lower the risks of recurrent neuroleptic malignant syndrome. The DSM-IV criteria for neuroleptic malignant syndrome are presented in Table 31.4-2.

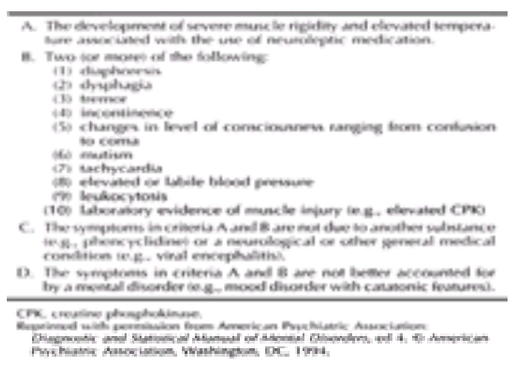

A. The development of severe muscle rigidity and elevated temperature associated with the use of neuroleptic medication.
B. Two (or more) of the following:
   (1) diaphoresis
   (2) dysphagia
   (3) tremor
   (4) incontinence
   (5) changes in level of consciousness ranging from confusion to coma
   (6) mutism
   (7) tachycardia
   (8) elevated or labile blood pressure
   (9) leukocytosis
   (10) laboratory evidence of muscle injury (e.g., elevated CPK)
C. The symptoms in criteria A and B are not due to another substance (e.g., phencyclidine) or a neurological or other general medical condition (e.g., viral encephalitis).
D. The symptoms in criteria A and B are not better accounted for by a mental disorder (e.g., mood disorder with catatonic features).

CPK, creatine phosphokinase.
Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 31.4-2 DSM-IV Research Criteria for Neuroleptic Malignant Syndrome

NEUROLEPTIC-INDUCED ACUTE DYSTONIA

DSM-IV defines this condition as an:

> [a]bnormal positioning or spasm of the muscles of the head, neck, limbs, or trunk developing within a few days of starting or raising the dosage of a neuroleptic medication (or after reducing a medication used to treat extrapyramidal symptoms).

These are acute, dramatic, bizarre muscular spasms (tonic reaction) mainly affecting the head, neck (torticollis), facial muscles (grimacing), eye muscles (oculogyric crises), jaw (trismus), and upper extremities. In rare instances the muscles of the back are affected (opisthotonos). Onset usually occurs within the first 24 to 48 hours after initiating neuroleptic therapy or after increases of the neuroleptic dosage. Risk for this disorder factors include being young, male, taking a high-potency antipsychotic, taking large dosages of an antipsychotic, or receiving parenteral administration of an antipsychotic. These spasms are usually dramatically relieved by parenteral administration of antiparkinsonian agents, followed by oral antiparkinsonian treatment, which should be assessed periodically.

Acute dystonia involving the laryngeal and pharyngeal musculature has been associated with neuroleptic medication. Characteristic symptoms of acute laryngeal dystonia include dyspnea and gestures indicating subjective respiratory distress, such as pointing to or clutching the throat. Immediate intravenous administration of antiparkinsonian agents (e.g., benztropine or diphenhydramine HCl) provides rapid relief of laryngeal dystonia and may be life saving for this potentially lethal condition. The DSM-IV criteria for neuroleptic-induced acute dystonia are listed in Table 31.4-3.

A. One (or more) of the following signs or symptoms has developed in association with the use of neuroleptic medication:
   (1) abnormal positioning of the head and neck in relation to the body (e.g., retrocollis, torticollis)
   (2) spasms of the jaw muscles (trismus, gaping, grimacing)
   (3) impaired swallowing (dysphagia), speaking, or breathing (laryngeal pharyngeal spasm, dysphonia)
   (4) thickened or slurred speech due to hypertonic or enlarged tongue (dysarthria, macroglossia)
   (5) tongue protrusion or tongue dyskinesias
   (6) eyes deviated up, down, or sideward (oculogyric crisis)
   (7) abnormal positioning of the distal limbs or trunk
B. The signs or symptoms in criterion A developed within seven days of starting or rapidly raising the dose of neuroleptic medication, or of reducing a medication used to treat (or prevent) acute extrapyramidal symptoms (e.g., anticholinergic agents).
C. The symptoms in criterion A are not better accounted for by a mental disorder (e.g., catatonic symptoms in schizophrenia). Evidence that the symptoms are better accounted for by a mental disorder might include the following: the symptoms precede the exposure to neuroleptic medication or are not congruous with the pattern of pharmacological intervention (e.g., no improvement after neuroleptic lowering or anticholinergic administration).
D. The symptoms in criterion A are not due to a nonneuroleptic substance or to a neurological or other general medical condition. Evidence that the symptoms are due to a general medical condition might include the following: the symptoms precede the exposure to the neuroleptic medication, unexplained focal neurological signs are present, or the symptoms progress in the absence of change in medication.

Table 31.4-3 DSM-IV Research Criteria for Neuroleptic-Induced Acute Dystonia

NEUROLEPTIC-INDUCED ACUTE AKATHISIA

According to DSM-IV this condition is associated with:

> [s]ubjective complaints of restlessness accompanied by observed movements (e.g., fidgety movements of the legs, rocking from foot to foot, pacing, or inability to sit or stand still) developing within a few weeks of starting or raising the dose of a neuroleptic medication (or after reducing a medication used to treat extrapyramidal symptoms).

The reported prevalence of akathisia has varied between 20 and 75 percent. Its onset is within a few days of initiation of medication, but it can also occur later in neuroleptic treatment.

Akathisia is more likely to occur in recent-onset psychosis, associated with subjective feelings of dysphoria, anxiety, or restlessness; inability to relax, stand still, or stay still; and pacing and shifting of the legs in the sitting position. Many individuals, particularly patients with a chronic psychotic disorder, do not report any subjective complaints, and in the authors' opinion, the objective component is the more important one in all but those with new-onset of acute psychotic disorder. Patients with severe cases may present with agitated, importunate, demanding, or almost histrionic behavior. The condition may be misdiagnosed as an exacerbation of the psychotic condition, resulting in increased dosage of neuroleptic drug, which further worsens akathisia. Akathisia may result in medication noncompliance because of uncomfortable adverse effects. Recently, it was reported that akathisia observed at any time (whether treated or not) was associated with a poor treatment outcome. The reported prevalence of akathisia associated with atypical neuroleptic drugs, including risperidone and olanzapine, is considerably lower than that with conventional neuroleptic drugs.

Treatment strategies for akathisia are similar to those for neuroleptic-induced parkinsonism. b-Adrenergic receptor antagonists have been found to be efficacious, and benzodiazepines may be useful. Recently, the 5-HT_2 antagonist cyproheptadine (Periacitin) was reported to reduce akathisia that did not respond to standard pharmacotherapy. The DSM-IV criteria for neuroleptic-induced acute akathisia are presented in Table 31.4-4.

A. The development of subjective complaints of restlessness after exposure to a neuroleptic medication.
B. At least one of the following is observed:
   (1) fidgety movements or swinging of the legs
   (2) rocking from foot to foot while standing
   (3) pacing to relieve restlessness
   (4) inability to sit or stand for at least several minutes
C. The onset of the symptoms in criteria A and B occur within four weeks of initiating or increasing the dose of the neuroleptic, or of reducing medication used to treat (or prevent) acute extrapyramidal symptoms (e.g., anticholinergic agents).
D. The symptoms in criterion A are not better accounted for by a mental disorder (e.g., schizophrenia, substance withdrawal, agitation from a major depressive or manic episode, hyperactivity in attention-deficit/hyperactivity disorder). Evidence that symptoms may be better accounted for by a mental disorder might include the following: the onset of symptoms preceding the exposure to the neuroleptics, the absence of increasing restlessness with increasing neuroleptic dose, and the absence of relief with pharmacologic interventions (e.g., no improvement after decreasing the neuroleptic dose or treatment with medication intended to treat the akathisia).
E. The symptoms in criterion A are not due to a nonneuroleptic substance or to a neurological or other general medical condition. Evidence that symptoms are due to a general medical condition might include the onset of the symptoms preceding the exposure to neuroleptics or the progression of symptoms in the absence of a change in medication.

Table 31.4-4 DSM-IV Research Criteria for Neuroleptic-Induced Acute Akathisia

NEUROLEPTIC-INDUCED TARDIVE DYSKINESIA

DSM-IV defines this disorder as:

> [i]nvoluntary choreiform, athetoid, or rhythmic movements (lasting at least a few weeks) of the jaw or extremities developing in association with the use of neuroleptic medication for at least a few months (may be for a shorter period of time in elderly persons).

Tardive dyskinesia is sometimes associated with rocking movements, and truncal muscles and respiratory muscles may be involved. Differential diagnoses include spontaneous dyskinesias (which were reported during the preneuroleptic era), mannerism or stereotyped motor behavior, senile dyskinesia, Huntington's disease, Tourette's disorder, Wilson's disease, Meige's syndrome, nonneuroleptic medication-induced dyskinesia (e.g., with tricyclics, antihistamines, phenytoin [Dilantin], sympathomimetics, steroids, selective serotonin reuptake inhibitors), dental problems, mouth sores, somatoform disorder, or factitious disorder.

Tardive dyskinesia is a well-recognized adverse effect associated with the long-term use of neuroleptic therapy and usually occurs insidiously, but early and more rapid onset sometimes occurs. During the past four decades, continuous efforts have been made to study the epidemiology, to identify the risk factors involved, and to develop effective treatment for tardive dyskinesia. The reported cumulative incidence is 5 percent after 1 year, 10 percent after 2 years, 15 percent after 3 years, and 19 percent after 4 years, up to about 25 percent. The overall incidence among young persons ranges from 3 to 5 percent per year. In individuals over age 45 years with a median duration of 21 days of total lifetime neuroleptic treatment, the cumulative incidence of tardive dyskinesia was 26 percent after 1 year, 52 percent after 2 years, and 60 percent after 3 years of neuroleptic treatment. The prevalence of spontaneous dyskinesia in patients with schizophrenia who have never been exposed to neuroleptic treatment is 2 to 4 percent. The reported prevalence of neuroleptic-induced tardive dyskinesia in individuals who have received long-term neuroleptic treatment (including cross-cultural studies) ranges from 0.5 to 65 percent, with an average prevalence of 20 percent. Such a wide variation in the reported prevalence is perhaps due to a lack of a precise objective definition of tardive dyskinesia, different examination techniques, variable diagnostic criteria for this movement disorder as well as training of the examiners. The increased use of standardized rating scales (e.g., Abnormal Involuntary Movement Scale [AIMS]), and more rigorous diagnostic criteria should eliminate these problems.

The disorder has expanded from the initial buccolingual-masticatory syndrome (pouting, puckering, and smacking of lips; tongue protrusion, curling worm-like movements and lateral movements of the tongue, which thrusts against the cheeks to produce the “bonbon” sign; plus chewing movement of the lower jaw) to include various additional abnormal movements and signs. These include facial grimacing; blinking of the eyes, movement of the eyelids; choreoathetoid movements of the fingers, wrists, ankles and toes; axial hyperkinesia, rocking or torsion movement of neck and trunk or both; diaphragmatic involvement resulting in grunting (glottal dyskinesia) and difficulty breathing. The abnormal tongue movements are best detected by having the patient stand with the mouth open and eyes closed; to activate the condition further or manifest it, the patient is asked to tap the thumbs against the fingers while the tongue is observed inside the mouth. As a rule, patients with tardive dyskinesia either are not aware of or do not present subjective complaints of the abnormal movement except in severe cases. Dyskinetic movements disappear during sleep.

Age (i.e., being elderly) and sex (i.e., being female) are the two most consistent risk factors identified in tardive dyskinesia, and advanced age is more likely to be

associated with persistent tardive dyskinesia. A history of extrapyramidal adverse effects, particularly with development early in the course of neuroleptic treatment or severe effects such as akathisia resistant to anticholinergic treatment, is reported to be significantly associated with tardive dyskinesia. The risk ratio for those with extrapyramidal adverse effects is 2.32 times that of those without. Antiparkinsonian agents are not thought to be a risk factor, but they can temporarily worsen tardive dyskinesia. Duration of neuroleptic treatment and intermittent or interrupted neuroleptic treatment are also reported to increase the risk of tardive dyskinesia. Other possible risk factors include individual susceptibility, ethnicity (being African-American, organicity, having negative symptoms of psychosis or deficit states, having diabetes mellitus, and history of alcohol-related or mood disorder). The total cumulative dosage of neuroleptic drugs administered may relate to increased severity of tardive dyskinesia. All dopamine receptor antagonists have been associated with tardive dyskinesia. At present, there are insufficient data to affirm that serotonin-dopamine antagonists neuroleptic drugs are not associated with tardive dyskinesia. Clozapine, an atypical neuroleptic drug, may not cause tardive dyskinesia, and at the least, the risk is likely to be lower with clozapine than with typical neuroleptic drugs. Although the early findings with other atypical neuroleptic drugs suggest that these agents also possess a low risk, this needs to be evaluated in prospective studies.

Tardive dyskinesia can be classified into persistent (irreversible) and withdrawal (reversible) types; reversibility following neuroleptic medication discontinuation has been reported to be from 0 to 90 percent. In general, when individuals with tardive dyskinesia are off neuroleptic medication, in one-third of patients dyskinetic movements become reversible within 3 months, and in more than 50 percent of patients movements remit within 12 to 18 months. Reversibility is lower in elderly than younger groups. It has been estimated that 5 to 40 percent of all tardive dyskinesia cases remit, and 50 to 90 percent of mild cases remit after discontinuation of neuroleptic medication. Although it is impossible to predict reversibility precisely, these findings suggest a better prognosis for patients treated with neuroleptic drugs at lower doses or for shorter durations, or both. Both patients removed from neuroleptic medication as well as those maintained on low-dose neuroleptic drugs appear to have continuous improvement or even resolution of tardive dyskinesia over many years. Therefore clinicians must recognize tardive dyskinesia as early as possible and assess the indications for prolonged neuroleptic treatment. In summary, a favorable long-term outcome correlates with younger age; early recognition; lower doses, shorter duration of neuroleptic treatment, or both; and increased duration of follow-up.

Although the exact neuropathophysiology of tardive dyskinesia remains to be clarified, the most accepted theory involves supersensitivity of dopamine receptors in the basal ganglia following neuroleptic treatment. Various pharmacological agents including both agonists and antagonists for different central nervous neurotransmitters (dopamine, acetylcholine, norepinephrine, serotonin, g-aminobutyric acid), as well as vitamin E have been tried, but none shows evidence of consistent efficacy. Thus, the primary focus of the clinical management of tardive dyskinesia is prevention. The most important point in prevention is prescribing neuroleptic agents only when they are clinically indicated. Neuroleptic drugs should not be prescribed for disorders that can be treated effectively with other medications or alternative treatments. For patients who need neuroleptic treatment, the benefit-to-risk consideration dictates the treatment decision. Whenever neuroleptic dosages are being titrated upward or downward, the risks of psychosis must be balanced against the risk of tardive dyskinesia for each patient. When neuroleptic treatment is indicated, prescribe the lowest possible dosage of neuroleptic and antiparkinsonian agents for the shortest possible period of time. Then assess the patient's psychiatric status and carefully evaluate the need for continual or maintenance neuroleptic treatment, and periodically (e.g., every 3 to 6 months or more frequently if necessary) assess for abnormal movements. Document the presence or absence of tardive dyskinesia. If the patient's psychiatric status is unstable, discuss with the patient, family, or both as appropriate the necessity and possible adverse effects of continual neuroleptic treatment. If clinically possible (i.e., the psychiatric condition is stable), a careful trial of discontinuation of neuroleptic medication should be made with close follow-up. Tardive dyskinesia may appear or worsen following the discontinuation or lowering of neuroleptic doses (withdrawal dyskinesia). In mild situations, withdrawal of the neuroleptic drug should be slowed and benzodiazepines may temporarily be prescribed (as muscle relaxants and to alleviate the patient's anxiety) or both. If withdrawal dyskinesia is severe enough to interfere with daily routine activities such as eating and sleeping, then pure dopamine receptor antagonists (e.g., fluphenazine, haloperidol) may be prescribed and titrated to suppress the movements. This should be followed by a gradual taper of the neuroleptic agent at a later date. If the psychotic symptoms recur or the psychosis worsens, neuroleptic treatment must be reinstituted, and the procedures described above should be followed. Atypical drugs such as clozapine are indicated for patients suffering from both tardive dyskinesia and schizophrenia, especially those with negative symptoms. Clozapine improves and may also prevent tardive dyskinesia. The DSM-IV criteria for neuroleptic-induced tardive dyskinesia are presented in Table 31.4-5.

A. Involuntary movements of the tongue, jaw, trunk, or extremities have developed in association with the use of neuroleptic medication.
B. The involuntary movements are present over a period of at least four weeks, and occur in any of the following patterns:
(1) choreiform movements (i.e., rapid, jerky, nonrepetitive)
(2) athetoid movements (i.e., slow, sinuous, continual)
(3) rhythmic movements (i.e., stereotypies)
C. The signs or symptoms in criteria A and B develop during exposure to a neuroleptic medication or within 4 weeks of withdrawal from an oral (or within 8 weeks of withdrawal from a depot) neuroleptic medication.
D. There has been exposure to neuroleptic medication for at least three months (one month if age 60 or older).
E. The symptoms are not due to a neurological or general medical condition (e.g., Huntington's disease, Sydenham's chorea, spontaneous dyskinesia, hyperthyroidism, Wilson's disease); ill-fitting dentures; or exposure to other medications that cause acute reversible dyskinesia (e.g., L-dopa, bromocriptine). Evidence that the symptoms are due to one of these etiologies might include the following: the symptoms precede the exposure to the neuroleptic medication or unexplained focal neurological signs are present.
F. The symptoms are not better accounted for by a neuroleptic-induced acute movement disorder (e.g., neuroleptic-induced acute dystonia, neuroleptic-induced acute akathisia).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 31.4-5 DSM-IV Research Criteria for Neuroleptic-Induced Tardive Dyskinesia

MEDICATION-INDUCED POSTURAL TREMOR

According to DSM-IV, this disorder is characterized by:

> [f]ine tremor occurring during attempts to maintain a posture that develops in association with the use of medication (e.g., lithium, antidepressants, valproate).

In addition to neuroleptics, other medications well documented to induce postural tremor include adrenocorticosteroids, alcohol, amphetamines and other sympathomimetics agents, bronchodilators, caffeine, hypoglycemic agents, levodopa (Dopar, Larodopa) lithium salts, monamine oxidase inhibitors, thyroid hormone, tricyclic drugs, and valproate (Depakene).

In general, treatment strategies include lowest possible dosage of psychotropic medication, monitoring and adjusting the dosage of other agents or decreasing the use of agents (e.g., caffeine, alcohol) that induce or worsen tremor. b-Adrenergic receptor antagonists help to reduce tremor. Lithium-induced hand tremor is a common nuisance rather than a serious adverse effect unless it is associated with lithium neurotoxicity. Tremor occurs with therapeutic lithium levels but may be related to dosage and serum lithium concentrations. Lowering the concentration of lithium in serum within the therapeutic range can decrease the tremor and b-adrenergic receptor antagonists can reverse lithium-induced postural tremor. Tremor occurs as part of sedative-hypnotic withdrawal and may signal the impending onset of more serious and severe withdrawal reaction including seizures; thus tremor needs careful evaluation and treatment. The DSM-IV criteria for medication-induced postural tremor are listed in Table 31.4-6.

A. A fine postural tremor that has developed in association with the use of a medication (e.g., lithium, antidepressants, valproic acid).
B. The tremor (i.e., a regular, rhythmic oscillation of the limbs, head, mouth, or tongue) has a frequency between 8 and 12 cycles per second.
C. The symptoms are not due to a preexisting nonpharmacologically induced tremor. Evidence that the symptoms are due to a preexisting tremor might include the following: the tremor was present before the introduction of the medication, the tremor does not correlate with serum levels of the medication, and the tremor persists after discontinuation of the medication.
D. The symptoms are not better accounted for by neuroleptic-induced parkinsonism.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 31.4-6 DSM-IV Research Criteria for Medication-Induced Postural Tremor

MEDICATION-INDUCED MOVEMENT DISORDER NOT OTHERWISE SPECIFIED

A variety of medications other than neuroleptic agents have been associated with movement disorders that do not meet the criteria for any of the specific disorders listed in DSM-IV (Table 31.4-7). These medications include anticholinergic drugs; antihistaminic drugs; lithium; monamine oxidase inhibitors; oral contraceptives; phenytoin and other anticonvulsants (when toxicity presents); reserpine (Diupres, Serpasil); tetrabenazine; tricyclic drugs; and dopamine agonists such as apomorphine, bromocriptine, or levodopa.

This category is for medication-induced movement disorders not classified by any of the specific disorders listed above. Examples include: (1) parkinsonism, acute akathisia, acute dystonia, or dyskinetic movement that is associated with a medication other than a neuroleptic; (2) a presentation that resembles neuroleptic malignant syndrome that is associated with a medication other than a neuroleptic; or (3) tardive dystonia.

Table 31.4-7 DSM-IV Diagnostic Criteria for Medication-Induced Movement Disorder Not Otherwise Specified

An akathisia-like syndrome associated with abrupt withdrawal of imipramine (Tofranil) has been reported. Dyskinetic movements were attributed to the anticholinergic action from long-term treatment with tricyclic drugs in a few cases. Although no definite association between selective serotonin reuptake inhibitors (SSRIs) and extrapyramidal adverse effects has been established, reports have described their association. The most common adverse effect is akathisia, followed by dystonia, parkinsonism, and tardive dyskinesia–like states.

DSM-IV lists tardive dyskinesia as a distinct diagnostic entity but only mentions one other tardive condition (tardive dystonia) under medication–induced movement disorder not otherwise specified. Other tardive conditions, so-called tardive syndrome, have been proposed and include many subtypes: tardive dystonia, tardive akathisia, tardive Tourette's syndrome, tardive myoclonus, tardive parkinsonism. Also, there are four proposed subtypes of severe tardive dyskinesia: tardive dystonia, blepharospasm, choreoathetosis, and tardive akathisia.

Tardive dystonia is a late-onset (frequently of earlier onset than other tardive conditions), persistent dystonia associated with neuroleptic treatment, which usually affects younger male patients. It commonly involves the muscles of the neck, shoulders, and trunk, causing opisthotonos. It is characterized by sustained abnormal posture with torticollis (abnormal position of the head caused by spasmodic contraction of neck muscles), hemiballismus (jerking and twitching movements of one side of the body), truncal torsion (twisting of the trunk of the body (e.g., Pisa syndrome caused by flexion of the trunk to one side accompanied by slight backward rotation), blepharospasm (twitching or spasmodic contraction of muscles of the eyelids), and grimacing (distortion of the face). The prevalence is given as 1 to 2 percent. As with tardive dyskinesia, there is no effective treatment for this sometimes incapacitating condition. In contrast to tardive dyskinesia, increasing the dosage of neuroleptic drugs may not suppress the dystonic movements; anticholinergic agents in high dosages and tetrabenazine may be helpful. Only few of the reported patients in the literature achieved full recovery. Injection of botulinum toxoid every few months is frequently helpful for severe eye blinking and even when larger muscles are affected.

Tardive Tourette's syndrome appears either during neuroleptic treatment or following discontinuation of treatment. It can usually be suppressed by increasing neuroleptic doses and is worsened by neuroleptic withdrawal. There is clear need for further study to gain more understanding about these tardive conditions.

ADVERSE EFFECTS OF MEDICATION NOT OTHERWISE SPECIFIED

This category allows clinicians to record the adverse effects of medications, other than movement symptoms, that become a focus of treatment (Table 31.4-8). Examples of such adverse effects include priapism, severe hypotension, and cardiac abnormalities.

This category is available for optional use by clinicians to code side effects of medication (other than movement symptoms) when these adverse effects become a main focus of clinical attention. Examples include severe hypotension, cardiac arrhythmias, and priapism.

Table 31.4-8 DSM-IV Diagnostic Criteria for Adverse Effects of Medication Not Otherwise Specified

FUTURE DIRECTIONS

Because no adverse effect–free psychotropic medication is available, clinicians must follow the basic principles of rational psychopharmacotherapy (viz., to maximize the therapeutic effects and minimize the adverse effects). Higher dosages lead to more adverse effects but do not necessarily lead to more therapeutic effects. Therefore, the general recommendation is to use the lowest possible amount of neuroleptic drug necessary to produce clinically desirable benefits and at the same time lower the risk of extrapyramidal adverse effects and tardive dyskinesia. However, there is still a need to define how low a dosage should be to effectively treat active psychopathology and to minimize the risks of relapse in maintenance treatment. Meantime, the effort to develop new neuroleptic drugs that are therapeutically equal but carry low or no risk of causing these adverse effects is encouraging and exciting. These so-called atypical agents will inevitably replace typical neuroleptic drugs for the treatment of psychotic state.

SUGGESTED CROSS-REFERENCES

Neuroleptics are discussed in Chapter 31, dopamine antagonists other than neuroleptics are discussed in Section 31.17, lithium is discussed in Section 31.18, b-adrenergic receptor antagonists are discussed in Section 31.5, and cholinesterase inhibitors are discussed in Section 31.15.

SECTION REFERENCES

*Adler CH: Differential diagnosis of Parkinson's disease. Med Clin North Am *83:* 349, 1999.

Adler LA, Edson R, Lavori P, Peselow E, Duncan E, Rosenthal M, Ratrosen J: Long-term treatment effects of vitamin E for tardive dyskinesia. Biol Psychiatry *43:* 868, 1998.

*American Psychiatric Association: Practice guideline for the treatment of patient with schizophrenia. Am J Psychiatry *154* (Suppl):1, 1997.

Casey DE: Extrapyramidal syndrome: Epidemiology, pathophysiology and the diagnostic dilemma. CNS Drugs *5* (Suppl):1, 1996.

*Farde L, Nordstrom A-L, Wiesel F-A, Pauli S, Halldin C, Sedvall G: Positron emission tomographic analysis of central D_1 and D_2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine: Relation to extrapyramidal side effects. Arch Gen Psychiatry *49:* 538, 1992.

Gardos G, Cole JO, Salomon M, Schniebolk S: Clinical forms of severe tardive dyskinesia. Am J Psychiatry *144:* 895, 1987.

*Giladi N, Melamed E: Levodopa therapy can ameliorate tetrabenazine-induced parkinsonism. Mov Disord *14:* 158, 1999.

Gleason PP, Conigliaro RL: Neuroleptic malignant syndrome with risperidone. Pharmacotherapy *17:* 617, 1997.

Hasan S, Buckley P: Novel antipsychotics and the neuroleptic malignant syndrome: A review and critique. Am J Psychiatry *155:* 1113, 1998.

Jeste DV: Neuroleptic-associated tardive syndrome. Psychiatr Clin North Am *9:* 183, 1986.

*Jeste DV, Caligiuri MP, Paulsen JS, Heaton RK, Lacro JP, Harris MJ, Bailey A, Fell RL, McAdams LA: Risk of tardive dyskinesia in older patients: A prospective longitudinal study of 266 outpatients. Arch Gen Psychiatry *52:* 756, 1995.

*Kane JM, Woerner M, Borenstein M, Wegner J, Lieberman J: Integrating incidence and prevalence of tardive dyskinesia. Psychopharmacol Bull *22:* 254, 1986.

Kane JM, Woerner MG, Pollack S, Safferman AZ, Lieberman JA: Does clozapine cause tardive dyskinesia? J Clin Psychiatry *54:* 327, 1993.

Kapur S, Remington G, Jones C, Wilson A, Dasilva J, Houle S, Zipursky R: High levels of dopamine D_2 receptor occupancy with low-dose haloperidol treatment: A PET study. Am J Psychiatry *153:* 948, 1996.

Koek RJ, Pi EH: Acute laryngeal dystonic reactions to neuroleptics. Psychosomatics *30:* 359, 1989.

Kurz M, Hummer M, Oberbauer H, Fleischhacker WW: Extrapyramidal side effects of clozapine and haloperidol. Psychopharmacology *118:* 52, 1995.

Kurzthaler I, Hummer M, Kohl C, Miller C, Fleischhacker WW: Propranolol treatment of olanzapine-induced akathisia (Letter). Am J Psychiatry *154:* 1316, 1997.

Leo RJ: Movement disorders associated with the serotonin selective reuptake inhibitors. J Clin Psychiatry *57:* 449, 1996.

Levinson DF, Simpson GM: Neuroleptic-induced extrapyramidal symptoms with fever. Arch Gen Psychiatry *43:* 839, 1986.

Marder SR, Meibach RC: Risperidone in the treatment of schizophrenia. Am J Psychiatry *151:* 825, 1994.

Morgenstern H, Glazer WM: Identifying risk factors for tardive dyskinesia among long-term outpatients maintained with neuroleptic medications. Results of the Yale Tardive Dyskinesia Study. Arch Gen Psychiatry *50:* 723, 1993.

Pelonero AL, Levinson JL, Pandurangi AK: Neuroleptic malignant syndrome: A review. Psychiatric Services *49:* 1163, 1998.

Pi EH, Gutierrez MA, Gray GE: Tardive dyskinesia: Cross-cultural perspectives. In *Psychopharmacology and Psychobiology of Ethnicity*, KM Lin, RE Poland, G Nakasaki, editors. American Psychiatric Press, Washington, DC, 1993.

Pi EH, Simpson GM: Atypical neuroleptics: Clozapine and the benzamides in the prevention and treatment of tardive dyskinesia. Mod Prob Pharmacopsychiatry *21:* 80, 1983.

Sachdev P, Kruk J, Kneebone M, Kissane D: Clozapine-induced neuroleptic malignant syndrome: Review and report of new cases. J Clin Psychopharmacol *15:* 365, 1995.

Sachdev P, Mason C, Hadzi-Pavlovic D: Case-control study of neuroleptic malignant syndrome. Am J Psychiatry *154:* 1156, 1997.

Simpson GM, Angus JWS: A rating scale for extrapyramidal side effects. Acta Psychiatr Scand *212:* 11, 1970.

Simpson GM, Pi EH: What's happening to the neuroleptic malignant syndrome? Psychiatr Ann *26:* 172, 1996.

*Simpson GM, Pi EH, Sramek JJ Jr: Neuroleptics and antipsychotic drugs. In *Meylers Side Effects of Drugs*, ed 13, MN Dukes, editor. Elsevier Science, Amsterdam, 1996.

Shulman LM, Singer C, Weiner WJ: Improvement of both tardive dystonia and akathisia after botulinum toxin injection. Neurology *46:* 844, 1996.

Small JG, Hirsch SR, Arvanitis LA, Miller BG, Link CG: Quetiapine in patients with schizophrenia: A high- and low-dose double blind comparison with placebo. Arch Gen Psychiatry *54:* 549, 1997.

Tollefson GD, Beasley CM Jr, Tran PV, Street JS, Kruger JA, Tamura RN, Graffeo KA, Thieme ME: Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: Results of an international collaborative trial. Am J Psychiatry *154:* 457, 1997.

Weiss D, Aizenberg D, Hermesh H, Zemishlany Z, Munitz H, Radwan M, Weizman A: Cyproheptadine treatment in neuroleptic-induced akathisia. Br J Psychiatry *167:* 483, 1995.

Woerner MG, Sheitman BB, Lieberman JA, Kane JM: Tardive dyskinesia induced by risperidone? Am J Psychiatry *153:* 843, 1996.

Zimbroff DL, Kane JM, Tamminga CA, Daniel DG, Mack RJ, Wozniak PJ, Sebree TB, Wallin BA, Kashkin KB: The Sertindole Study Group: Controlled, dose-response study of sertindole and haloperidol in the treatment of schizophrenia. Am J Psychiatry *154:* 782, 1997.

31.5 b-ADRENERGIC RECEPTOR ANTAGONISTS

GEORGE M. SIMPSON, M.D., AND CALVIN J. FLOWERS, M.D.

HISTORY

Most organs in the body are controlled to some extent by the two types of adrenergic receptors (a and b). The b-adrenergic receptors are further subdivided into b1 and b2, based on target tissue effects. Both selective and nonselective antagonists of b-receptors have been developed and are used to control hypertension, cardiac arrhythmias, hyperthyroidism, glaucoma, and tremor; for migraine prophylaxis; and for various neuropsychiatric purposes. The first b-adrenergic receptor antagonist (beta-blocker) was dichlorisoproterenol, which was abandoned because of partial agonist properties, which at the time were thought to be clinically unacceptable. This was followed by pronethalol, which produced thymic tumors and was discarded but prompted the discovery of propranolol (Inderal) the prototype b-adrenergic receptor antagonist. The most common b-adrenergic receptor antagonists used in psychiatry are propranolol, nadolol (Corgard), pindolol (Visken), labetalol (Normodyne, Trandate), metoprolol (Lopressor, Toprol), atenolol (Tenormin), and acebutolol (Sectral), though a multitude of others exist and theoretically could be used.

CHEMISTRY

Non-Subtype-Selective b-Adrenergic Antagonists Propranolol is chemically described as 1-(isopropylamino)-3-(1-naphthyloxy)-2-propanol hydrochloride; its molecular structure is shown in Figure 31.5-1. Propranolol hydrochloride is a stable, white, crystalline solid that is readily soluble in water and ethanol. Its molecular weight is 295.81.

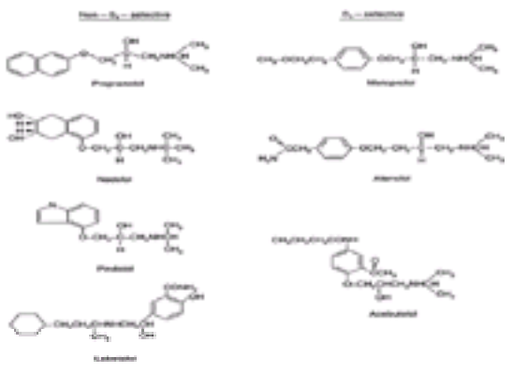

FIGURE 31.5-1 Molecular structures of b-adrenergic receptor antagonist.

Nadolol is chemically designated 1-(*tert*-butylamino)-3-[(5,6,7,8-tetrahydro-*cis*-6,7-dihydroxy-1-naphthyl)oxy]-2-propanol. Its molecular structure is shown in Figure 31.5-1. The molecular weight is 309.40. Nadolol is a white crystalline powder. It is freely soluble in ethanol, soluble in hydrochloric acid, slightly soluble in water and in chloroform, and very slightly soluble in sodium hydroxide.

Pindolol is chemically described as 1-(indol-4-yloxy)-3-(isopropylamino)-2-propanol. Pindolol is a white to off-white odorless powder soluble in organic solvents and aqueous acids. Its structural formula is shown in Figure 31.5-1. Its molecular weight is 248.33.

Labetalol is chemically designated 5-[1-hydroxy-2-[(1-methyl-3-phenylpropyl)amino]ethyl]salicylamide monohydrochloride. Labetalol HCl has the empirical formula $C_{19}H_{24}N_2O_3$.HCl and a molecular weight of 364.9. It has two asymmetrical centers and therefore exists as a molecular complex of two diasteroeisomeric pairs. Dilevalol, the *R,R*' stereoisomer, makes up 25 percent of racemic labetalol. Labetalol is a white or off-white crystalline powder, soluble in water. The molecular structure is shown in Figure 31.5-1.

Subtype-Selective b-Adrenergic Antagonists Metoprolol is chemically described as 1-(isopropylamino)-3-[*p*-(2-methoxyethyl)phenoxy]-2-propanol (2:1) *dextro*-tartrate salt. Metoprolol is a white, practically odorless, crystalline powder with a molecular weight of 684.82. It is very soluble in water; freely soluble in methylene chloride, in chloroform, and in alcohol; slightly soluble in acetone; and insoluble in ether. The molecular structure is shown in Figure 31.5-1.

Atenolol is chemically described as benzeneacetamide, 4-[2'-hydroxy-3'-[(1-methylethyl)amino]propoxy]. The molecular formula is $C_{14}H_{22}N_2O_3$, and the structural formula is shown in Figure 31.5-1. Atenolol (free base) has a molecular weight of 266. It is a relatively polar hydrophilic compound with a water solubility of 26.5 mg/mL at 37°C and a log partition coefficient (octanol/water) of 0.23. It is freely soluble in 1N HCl (300 mg/mL at 25°C).

Acebutolol has the molecular formula $C_{18}H_{28}N_2O_4$·HCl. Its molecular weight is 372.9. Acebutolol HCl is a white or slightly off-white powder freely soluble in water and less soluble in alcohol. Chemically it is defined as the hydrochloride salt of butanamide, *N*-[3-acetyl-4-[2-hydroxy-3-[(1-methylethyl)amino]propoxy]phenyl], (+/–)- or (+/–)-3'acetyl-4'-[2-hydroxy-3-(isopropylamino)propoxy]butyranilide.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption After oral administration, propranolol, pindolol, labetalol, acebutolol, and metoprolol are almost completely absorbed from the gastrointestinal tract. Nadolol and atenolol are incompletely absorbed. The general pharmacological properties of commonly used b-adrenergic receptor antagonists are given in Table 31.5-1.

	Selectivity	Lipid Solubility	Half-Life (hr)	Bioavailability (%)
Propranolol	No	High	3–6	30
Nadolol	No	Low	14–24	35
Pindolol	No	Moderate	3–4	80
Labetalol	No	Moderate	4–6	30
Metoprolol	Yes	High	3–4	50
Atenolol	Yes	Low	5–8	40
Acebutolol	Yes	Low	3–4	50

Table 31.5-1 Pharmacological Characteristics of b-Adrenergic Receptor Antagonists

Distribution The b-adrenergic receptor antagonists have markedly different lipophilic properties. The least lipophilic of them, nadolol and atenolol, cross the blood-brain barrier poorly and thus have a higher ratio of peripheral to central effects. The highly lipophilic b-adrenergic receptor antagonists, propranolol and metoprolol, have potent central and peripheral effects. The others have intermediate lipophilic properties and a more balanced central and peripheral effect. Certain b-adrenergic receptor antagonists (e.g., pindolol), have intrinsic sympathomimetic effects as a result of their partial agonist properties. Most b-adrenergic receptor antagonists with the exception of nadolol and atenolol, are extensively metabolized in the liver, with only 25 to 45 percent reaching the systemic circulation. Nadolol and atenolol are not extensively metabolized and are excreted largely unchanged in the urine. All currently available b-adrenergic receptor antagonists are excreted in breast milk and should be administered with caution to nursing women.

Metabolism The elimination of b-adrenergic receptor antagonists from the body is quite variable. Nadolol and atenolol have relatively long half-lives, which allows once-a-day dosing. The other b-adrenergic receptor antagonists have shorter half-lives and require multiple daily dosing, with the exception of propranolol and metoprolol, which are available in long-acting forms. Table 31.5-2 presents the half-lives of commonly used b-blockers.

Definitely effective
Performance anxiety
Lithium-induced tremor
Neuroleptic-induced akathisia
Probably effective
Adjunctive therapy for alcohol withdrawal and other substance-related disorders
Adjunctive therapy for aggressive or violent behavior
Possibly effective
Antipsychotic augmentation
Antidepressant augmentation

Table 31.5-2 Psychiatric Uses for b-Adrenergic Receptor Antagonists

Pharmacodynamics

Mechanism of Action The neurotransmitters norepinephrine and epinephrine share two different types of adrenergic receptors called a and b. a-Receptors are divided into type 1 and type 2. Type 1 a (a_1)-receptors are located postsynaptically and produce vasoconstriction, whereas a_2-receptors are generally presynaptic inhibitory autoreceptors. There are three types of postsynaptic b receptors. b_1-activation produces increased heart rate, force of contraction and conduction; b_2-receptors cause bronchodilation. The exact function of b_3-receptors is unclear.

In the central nervous system, the major noradrenergic nucleus is in the locus ceruleus, located in the dorsal pons. The noradrenergic neurons in the brain are responsible for central regulation of the sympathetic nervous system, modulation of pain, and hormone release. b-adrenergic receptor antagonists produce their effects by competitive inhibition of the b-receptors, thus they interfere with, or block, the effect of natural agonists.

Blood Concentrations and Relation to Action The high individual variation in b-adrenergic receptor antagonists concentration is a function of many factors including extent of hepatic metabolism, coadministered medications, food ingestion, and length of therapy. The lower lipophilic nature of nadolol and atenolol, with their lower hepatic clearance results in reduced interindividual variation. There currently is no application in psychiatry for the use of plasma concentrations of b-adrenergic receptor antagonists.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

b-Adrenergic receptor antagonists have been used to treat a variety of psychiatric conditions (Table 31.5-2). There have been reported studies of their effective use in anxiety disorder, panic attacks, alcohol withdrawal, aggression, and neuroleptic-induced akathisia.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

b-Adrenergic receptor antagonists generally have minimal effects on healthy individuals when taken at doses commonly used in psychiatry. They do have a marked blood pressure–lowering effect in patients with hypertension. Likewise, patients with airway disease may experience severe bronchoconstriction. b-adrenergic receptor antagonists also affect carbohydrate metabolism.

THERAPEUTIC INDICATIONS

Neuroleptic-Induced Akathisia Akathisia produced by neuroleptic agents is a common, distressing adverse effect that has significant implications for treatment. It is most commonly described as an objective and subjective restlessness that occurs to some degree in a large proportion of patients treated with neuroleptics. It is often underdiagnosed and not treated or misdiagnosed and inappropriately treated as increased psychosis. Of the various agents used to treat this condition, the b-adrenergic blocking drugs are the best studied and generally accepted. Propranolol is the prototype agent and was first used for this purpose in the early 1980s. In case studies and both open-label and controlled trials, propranolol has produced fairly rapid, well-tolerated relief from akathisia induced by neuroleptics. In several randomized, double-blind studies propranolol yielded significant decreases in subjective and objective akathisia compared with placebo, anticholinergics, and benzodiazepines. Generally low doses of propranolol are effective, usually less than 100 mg per day, in divided doses. The effectiveness of propranolol in treating neuroleptic-induced akathisia, prompted considerable interest in other b-adrenergic receptor antagonists as antiakathisia agents to determine whether the lipophilic properties and subtype-selective properties produced any differential efficacy. The question has not been clearly answered, but results thus far indicate a trend toward greater efficacy with lipophilic agents, probably indicating a central cause for neuroleptic-induced akathisia. The results of nonselective versus selective b-adrenergic receptor antagonists in treating neuroleptic-induced akathisia are mixed, with studies reporting effectiveness with members of each group.

Lithium-Induced Tremor Lithium-induced tremor occurs in most patients taking lithium (Eskalith). It is more common in men than women and occurs early in lithium therapy and in patients taking multiple other psychotropic medications. It presents as a fine, 5 to 12 cycles-a-second action tremor, but may occur at rest. b-adrenergic receptor antagonists are effective in controlling lithium-induced tremor; however, other pharmacological strategies must always be attempted first, including using the lowest therapeutically effective dose, reducing consumption of agents such as caffeine that worsen tremor, and trials of bedtime-dosing strategies. Once these

attempts have failed, administering b-adrenergic receptor antagonists is usually effective both subjectively and objectively. No well-controlled, head-to-head trials exist regarding differential effectiveness of b-adrenergic receptor antagonists. However, low-dose propranolol is well established, and several case series describe adequate control of lithium-induced tremor with atenolol and metoprolol in patients with asthma.

Anxiety Disorders (Performance Anxiety, Panic Disorder, Social Phobia, Posttraumatic Stress Disorder, Generalized Anxiety Disorder Not long after their development, b-adrenergic receptor antagonists were shown to possess anxiolytic properties in patients treated for cardiac dysfunction. This led to multiple comparisons with other anxiolytics including benzodiazepines for the treatment of different anxiety syndromes. Early studies focused on acute stress reactions such as performance anxiety, with claims for their effectiveness in nervous musicians and in subjects with excess anxiety before public speaking. Studies show improved performance on test scores of students taking 40 mg of propranolol prior to the examination. The logical extension of this research was the use of b-adrenergic receptor antagonists in patients with panic disorder. Unfortunately results have been inconsistent, and comparisons with benzodiazepines reveal the latter to be generally more effective. A limited number of studies indicate that atenolol and propranolol decrease some of the somatic symptoms of social phobia and generalized anxiety disorder. Two uncontrolled studies exist of b-adrenergic receptor antagonists in patients with posttraumatic stress disorder, one in war veterans and one in childhood sexual abuse victims; both showed propranolol effective in reducing symptoms, but more studies are needed.

Substance-Related Disorders Because b-receptor antagonism tends to reduce some of the withdrawal symptoms of alcohol and benzodiazepines such as tachycardia, hypertension, and tremor, their use in these syndromes continues to be investigated. b-Adrenergic receptor antagonists will not prevent or treat all alcohol withdrawal symptoms, especially delirium. Their main function is as an adjunct to benzodiazepines, to relieve the somatic sympathetic symptoms of sedative-hypnotic withdrawal. Several small, open-label studies found propranolol helpful in opioid and nicotine withdrawal. Some evidence indicates that b-adrenergic receptor antagonists may help reduce opioid craving in heroin users.

Aggressive or Violent Behavior b-Adrenergic receptor antagonists have gained considerable acceptance as adjunctive pharmacotherapy for patients with behavioral syndromes, including aggression, agitation, and violent outbursts. Most studies have included patients with a number of associated neuropsychiatric conditions (e.g., dementia, traumatic brain injuries, severe developmental disorders, and other brain syndromes). Generally, propranolol has been effective in case series and in open trials; large double-blind controlled studies are rare. Furthermore, in several studies patients received multiple other psychotropic medications in addition to b-adrenergic receptor antagonists, which makes it difficult to draw conclusions. b-Adrenergic receptor antagonists do possess serotonergic properties, and decreased central serotonin concentrations have been associated with increased aggressive and impulsive behavior; this may explain their effectiveness. Most studies indicate that high dosages of propranolol (up to 500 mg per day or more) decrease the frequency and intensity of violent behavior syndromes.

Communication Disorders (Stuttering) Stuttering is a speech disorder that has been attributed to many different causes, including incomplete language lateralization and conflicting cerebral dominance. It is also an uncommon adverse effect of antidepressant and antipsychotic medication. Whatever the cause, it clearly worsens during periods of anxiety. Several recent case reports describe improvement in stuttering with b-adrenergic receptor antagonists and while controlled trials are needed, low-dose propranolol may be a useful adjunct to speech therapy.

Antidepressant Augmentation Recently interest has developed in the use of b-adrenergic receptor antagonists in antidepressant augmentation or adjunctive therapy strategies. The rationale for this involves the ability of many b-adrenergic receptor antagonists to inhibit the serotonin (5-hydroxytryptamine [5-HT]) subtype 1A (5-HT_{1A}) presynaptic receptors. Several case reports and open-label trials have shown pindolol to be safe and effective in combination with selective serotonin reuptake inhibitors (SSRIs) and moclobemide (Aurorix) in untreated depressed patients and treatment-resistant patients. A recent double-blind, controlled trial of fluoxetine (Prozac) plus pindolol or placebo failed to detect any differential efficacy.

Antipsychotic Augmentation Beta-adrenergic receptor antagonists may produce some improvement in patients with agitated psychosis, but it is doubtful that this is an antipsychotic effect. Possibly concomitant use of b-adrenergic receptor antagonists and antipsychotics raises the antipsychotic concentration, which may explain the improvement in some studies of patients with schizophrenia. Studies comparing propranolol with chlorpromazine (Thorazine) in the treatment of schizophrenia have generally used subtherapeutic doses of neuroleptic.

PRECAUTIONS AND ADVERSE REACTIONS

The b-adrenergic receptor antagonists have a number of important adverse effects on different organ systems including cardiovascular (e.g., hypotension, bradycardia), respiratory (e.g., exacerbation of asthma), metabolic (e.g., exacerbation of diabetes), sexual dysfunction, neuropsychiatric (e.g., fatigue and irritability), and gastrointestinal (nausea and diarrhea). The adverse effects and toxicity of these drugs are summarized in Table 31.5-3.

Cardiovascular
Hypotension
Bradycardia
Dizziness
Congestive failure (in patients with compromised myocardial function)
Respiratory
Asthma (less risk with β_1-selective drugs)
Metabolic
Worsened hypoglycemia in diabetic patients on insulin or oral agents
Gastrointestinal
Nausea
Diarrhea
Abdominal pain
Sexual Function
Impotence
Neuropsychiatric
Lassitude
Fatigue
Dysphoria
Insomnia
Vivid nightmares
Depression (!)
Psychosis (rare)
Other (rare)
Raynaud's phenomenon
Peyronie's disease
Withdrawal syndrome
Rebound worsening of preexisting angina pectoris when β-adrenergic receptor antagonists are discontinued

Table 31.5-3 Adverse Effects and Toxicity of b-Adrenergic Receptor Antagonists

Historically the possible link between the use of b-adrenergic receptor antagonists and the development of depression has been quite controversial. While it is difficult to completely rule out, a recent large retrospective study provided little evidence to support a causal association between administration of b-adrenergic receptor antagonists and the development of depressive symptoms.

Much of the caution in b-adrenergic receptor antagonist use centers around their possible exacerbation of two conditions: asthma and chronic obstructive pulmonary disease. In both cases, this is a function of b_2-adrenergic receptor antagonism causing bronchial smooth muscle contraction and increased airway resistance. In patients with no history of pulmonary disease b_2-receptor blockade has little influence on pulmonary function. The newer subtype-selective b-adrenergic receptor antagonists theoretically provide some safety margin, but at higher dosages, subtype-selective agents become increasingly nonselective. Since b-adrenergic receptor antagonists tend to interfere with the peripheral manifestations of hypoglycemia they must be used with caution in diabetic patients prone to low blood glucose concentrations. As with any medication administered to pregnant women, the potential benefits and risks to both mother and fetus need careful consideration. To date no adequate, well-controlled studies of b-adrenergic receptor antagonists in pregnant women exist. Some b-adrenergic receptor antagonists were toxic to embryos in animal studies at dosages much higher than the maximum human dosage. Most b-adrenergic receptor antagonists are excreted in human milk, and thus they should be used with caution in nursing mothers. The safety and effectiveness of these drugs in treating neuropsychiatric conditions in children and adolescents has not been established.

DRUG INTERACTIONS

Cholesterol-lowering agents such as cholestyramine (Questran) and colestipol (Colestid) decrease absorption of b-adrenergic receptor antagonists. Certain medications (e.g., phenytoin [Dilantin] and phenobarbital [Luminal]) induce the hepatic enzymes responsible for their biotransformation and lower the plasma level of b-adrenergic receptor antagonists, as does smoking. b-Adrenergic receptor antagonists reportedly increase theophylline concentration when administered concomitantly and may increase concentrations of certain antipsychotics and antidepressants. Combined use of b-adrenergic receptor antagonists and calcium channel blockers has an additive effect on cardiac conduction.

DOSAGE AND ADMINISTRATION

Propranolol is available in 10-, 20-, 40-, 60-, 80-, and 90-mg tablets and 4-, 8-, and 80-mg/mL solutions, as well as 60-, 80-, 120, and 160-mg sustained-release capsules. Nadolol is available in 20-, 40-, 80-, 120-, and 160-mg tablets. Pindolol has 5- and 10-mg tablets and labetalol comes in 100-, 200-, and 300-mg tablets. Metoprolol is available in both 50- and 100-mg tablets and 50-, 100-, and 200-mg sustained-release tablets. Atenolol is available in 25-, 50-, and 100-mg tablets, and acebutolol in 200- and 400-mg capsules. Atenolol, labetalol, metoprolol, and propranolol are also available in parenteral forms.

For the treatment of chronic disorders, propranolol treatment is usually initiated at 10 mg given orally three times a day or 20 mg given orally twice a day. The dosage can be raised by 20 to 30 mg a day until a therapeutic effect begins to emerge. The dosage should be leveled off at the appropriate range for the disorder under treatment. The treatment of aggressive behavior sometimes requires dosages up to 800 mg a day, and therapeutic effects may not be seen until the patient has been receiving the maximal dosage for 4 to 8 weeks. The patient's pulse and blood pressure should be taken regularly, and the drug should be withheld if the pulse is below 50 or the systolic blood pressure is below 90. The drug should also be temporarily withheld if the patient has severe dizziness, ataxia, or wheezing. Treatment with b-adrenergic antagonists should never be discontinued abruptly. Propranolol should be tapered at 60 mg a day until a dosage of 60 mg a day is reached, after which the drug should be tapered by 20 mg a day every 3 or 4 days.

Lithium can be associated with a tremor even when lithium concentrations are within normal therapeutic ranges. Propranolol (20 to 160 mg a day, two or three times a day) is effective treatment for lithium-induced tremor. Whether nonlipophilic, peripherally acting b-adrenergic antagonists are as effective as propranolol is controversial.

Some studies indicate that b-adrenergic antagonists are effective in treating the aggressiveness and violent behavior that can be associated with schizophrenia and organic mental diseases such as trauma, tumor, anoxic injury, encephalitis, alcoholism, and Huntington's disease. Most of the patients studied in the reports have not been responsive to antipsychotics, lithium, anticonvulsants, and benzodiazepines. Approximately 50 percent of the patients in the studies improved with b-adrenergic antagonist treatment. The range of propranolol dosages for aggression and violent behavior is from 50 to 960 mg a day. Two lipophilic, centrally acting drugs, pindolol and metoprolol, have also been used successfully for this indication in case reports and open trials.

Propranolol has been reported useful in reducing the peripheral manifestations of anxiety (e.g., tremor, tachycardia) associated with social phobia and the anxiety associated with performance, such as examinations and musical recitals. Propranolol (10 to 40 mg) taken 20 to 30 minutes before the performance is the usual dose.

Propranolol has been reported useful as an adjuvant to benzodiazepines but not as a sole agent in the treatment of alcohol withdrawal. One study used the following dose schedule: no propranolol for a pulse less than 50; 50 mg propranolol for a pulse between 50 and 79; 100 mg propranolol for a pulse of 80 or more. The patients who also received propanolol had less severe withdrawal symptoms, more stable vital signs, and a shorter hospital stay than patients who received only benzodiazepines.

SUGGESTED CROSS-REFERENCES

Dopamine antagonist treatment and complications are discussed in Section 32.15, treatment of lithium-induced adverse-effects are discussed in Section 32.16, and anticholinergics are discussed in Section 31.6.

SECTION REFERENCES

Berman RM, Darnell AM, Miller HL, Anand A, Charney DS: Effect of pindolol in hastening response to fluoxetine in the treatment of major depression: A double-blind, placebo-controlled trial. Am J Psychiatry *154:* 37, 1997.

Blaisdell GD: Akathisia: A comprehensive review and treatment summary. Pharmacopsychiatry *27:* 139, 1994.

Border R, Thomas P, Dupuis B: Effect of pindolol on onset of action of paroxetine in the treatment of major depression intermediate analysis of a doublel-blind, placebo-controlled trial. Am J Psychiatry *155:* 1346, 1998.

*Bright RA, Everitt DE: b -Blockers and depression. JAMA *267:* 1783, 1992.

Concores JA, Dackis CA, Davies RK, Gold MS: Propranol and stuttering. Am J Psychiatry *143:* 1071, 1986.

Dave M: Treatment of lithium induced tremor with atenolol. Can J Psychiatry *34:* 132, 1989.

Granville-Grossman KL: Propranolol, anxiety, and the central nervous system. Br J Clin Pharmacol *1:* 361, 1974.

Greendyke RM, Shuster DB, Wooton JA: Propranolol in the treatment of assaultive patients with organic brain disease. J Clin Psychopharmacol *4:* 282, 1984.

Kathol RG, Noyes R Jr, Slymen DJ, Crowe RR, Clancy J, Kerber RE: Propranolol in chronic anxiety disorders: A controlled study. Arch Gen Psychiatry *37:* 1361, 1981.

Kraus ML, Gottlieb LD, Horwitz RI, Anscher M: Randomized clinical trial of atenolol in patients with alcohol withdrawal. N Engl J Med *313:* 905, 1985.

*Lader M: b-Adrenoceptor antagonists in neuropsychiatry: An update. J Clin Psychiatry *49:* 213, 1988.

Lipinski JF, Zubenko GS, Cohen BM, Barreira PJ: Propranolol in the treatment of neuroleptic-induced akathisia. Am J Psychiatry *143:* 412, 1984.

Little KY, Clark TB, Ranc J, Duncan GE: Beta-adrenergic receptor binding in frontal cortex from suicide victims. Biol Psychiatry *34:* 596, 1993.

Manchanda R, Hirsch SR: Does propranolol have an antipsychotic effect? A placebo-controlled study in acute schizophrenia. Br J Psychiatry *148:* 701, 1986.

Mattes JA: Metoprolol for intermittent explosive disorder. Am J Psychiatry *142:* 1108, 1985.

*McAllister-Williams RH, Young AH: Pindolol augmentation of antidepressant therapy. Br J Psychiatry *173:* 536, 1999.

Perez V, Gilaberte I, Faries D, Alvarez E, Artigas F: Randomised, double-blind, placebo-controlled trial of pindolol in combination with fluoxetine antidepressant treatment. Lancet *349:* 1594, 1997.

Polakoff S, Lindem KJ, Bemporad JR, Richardson L, Rosenfeld B: Nadolol to treat aggression and psychiatric symptomatology in chronic psychiatric inpatients: A double-blind, placebo-controlled study. J Clin Psychiatry *53:* 41, 1992.

*Ratey JJ, Prough EE: The current status of b-blockers in psychiatric practice. Directions in Psychiatry *16:* 16, 1996.

*Ratey JJ, Sorgi P, O'Driscoll GA, Sands S, Daehler ML, Fletcher JR, Kadish W, Spruiell G, Polakoff S, Lindem KJ: Nadolol to treat aggression and psychiatric symptomatology in chronic psychiatric inpatients: A double-blind, placebo-controlled study. J Clin Psychiatry *53:* 41, 1992.

Silver JM, Yudofsky SC, Kogan M, Katz BL: Elevation of thioridazine plasma levels by propranolol. Am J Psychiatry *143:* 1290, 1986.

Tyrer P: Current status of b-blocking drugs in the treatment of anxiety disorders. Drugs *36:* 773, 1988.

Wells BG, Cold JA, Marken PA, Brown CS, Chu C, Johnson RP, Nasdahl CS, Auuybi MA, Knott DH, Arheart KL: A placebo-controlled trial of nadolol in the treatment of neuroleptic-induced akathisia. J Clin Psychiatry *52:* 255, 1991.

Yudofsky SC, Silver JM, Hales RE: Pharmacologic management of aggression in the elderly. J Clin Psychiatry *51* (Suppl):22, 1990.

31.6 ANTICHOLINERGICS AND AMANTADINE

JONATHAN M. MEYER, M.D., AND GEORGE M. SIMPSON, M.D.

HISTORY

Although the sedating effects of plants in the *Solanaceae* family were known for centuries, it was the vagolytic properties that prompted Jean-Martin Charcot to use these compounds therapeutically in Parkinson's disease, a disorder then ascribed to parasympathetic overactivity. Under his aegis, a treatise was published in 1867 documenting the effectiveness of belladonna alkaloids in Parkinson's disease, and these compounds continued as a mainstay of Parkinson's disease treatment until the development of levodopa (Larodopa) a century later. At present, anticholinergic medications are primarily used for the prophylaxis and treatment of extrapyramidal adverse effects induced by postsynaptic dopaminergic blockade from antipsychotic medications; however, the increasing use of serotonin-dopamine antagonists (atypical antipsychotics) and an appreciation of therapeutic efficacy with lower doses of dopamine receptor antagonists (typical antipsychotics) will undoubtedly contribute to a decline in their routine use for extrapyramidal adverse effect prophylaxis. Important issues in the prophylactic use of anticholinergics include low rates of acute dystonia in most patients, the negative impact on cognition and memory produced by anticholinergics, the relative risk of exacerbating and possibly producing tardive dyskinesia when anticholinergic use is associated with long-term neuroleptic treatment, and short-term risks of increased extrapyramidal adverse effects produced by abrupt withdrawal of anticholinergics during neuroleptic treatment. The available anticholinergic agents appear equally effective for the treatment of medication-induced dystonia, rabbit syndrome, and parkinsonism but less so for akathisia.

Amantadine (Symmetrel), initially marketed as an antiviral agent for the prophylaxis and treatment of influenza A, was fortuitously noted in 1969 to improve the symptoms of Parkinson's disease. While the mechanism of action is unclear, amantadine appears to act presynaptically, where it blocks reuptake and facilitates release of stored dopamine from terminals, and postsynaptically by direct agonism and receptor modulation. There is also pharmacologic evidence for noncompetitive antagonism of *N*-methyl-d-aspartate (NMDA) glutamate receptors. Amantadine largely lacks the cognitive impairment associated with anticholinergic administration and treats extrapyramidal effects effectively but has enjoyed only modest use in the United States.

CHEMISTRY

Selected anticholinergics are shown in Figure 31.6-1. The molecular structure of amantadine is shown in Figure 31.6-2. Benztropine (Cogentin), and trihexyphenidyl (Artane) are tertiary amines with significant muscarinic cholinergic blockade, but benztropine also possesses antihistaminic activity.

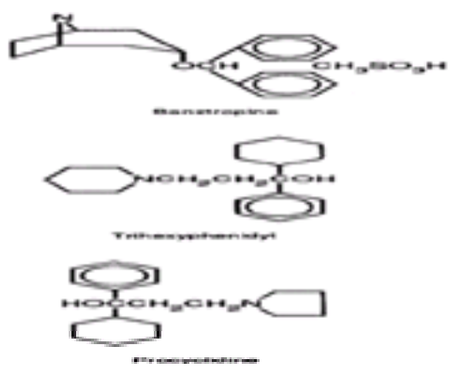

FIGURE 31.6-1 Molecular structures of selected anticholinergic drugs.

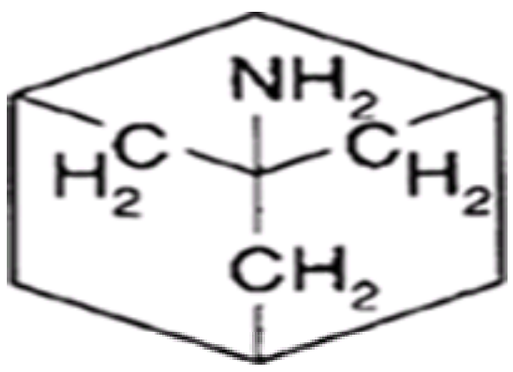

FIGURE 31.6-2 Molecular structure of amantadine.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Amantadine Amantadine is well absorbed after oral administration, with peak levels achieved 1 to 4 hours after ingestion. More than 90 percent of a single dose can be recovered unmetabolized in the urine. The plasma half-life is 12 to 18 hours in healthy young adults but sufficiently longer in those with renal impairment and elderly persons that a dose reduction of 50 percent in elderly adults will achieve adequate concentrations in serum. A dose of 100 mg twice a day yields peak concentrations in plasma of 0.5 to 0.8 μg/mL, and plasma trough levels of 0.3 μg/mL. Toxicity is seen at serum levels of 1.0 to 5.0 μ/mL and may be associated with delirium, hallucinations, seizures, or arrhythmias. The neurological symptoms of overdose have been reported to be reversible by physostigmine (Antilirium, Eserine). At low dosages, the side-effect profile is comparable to that of placebo, but continued usage has been occasionally associated with reversible effects on the cardiovascular system including livedo reticularis, peripheral edema, and orthostatic hypotension. Atropine-like adverse effects such as dry mouth, blurred vision, and palpitations have also been reported, usually at higher dosages.

Animal studies demonstrate evidence of teratogenicity but typically at doses many times the human dose, on a milligram per kilogram basis. There are a limited number of cases in the literature of in utero amantadine exposure during the first trimester. One woman used amantadine to prevent relapse of multiple sclerosis

throughout two of her pregnancies and delivered two normal infants, yet there is a report of a woman treated with amantadine in the first trimester who gave birth to an infant at 29 weeks gestation with tetralogy of Fallot and tibial hemimelia. Given the lack of extensive experience in pregnant women, this drug should be avoided in the first trimester unless the physician feels that the benefits to the mother outweigh the risks to both mother and fetus. Amantadine has a large volume of distribution and is found in breast milk and in nearly serum concentrations in saliva.

Anticholinergics The piperidine compound trihexyphenidyl, one of the first synthetic anticholinergic agents available, replaced the belladonna alkaloids for treatment of Parkinson's disease in the 1940s because of its more favorable adverse-effect profile. Biperiden (Akineton) is also in this class, along with procyclidine (Kemadrin) and cycrimine (Pagitane). The anticholinergic potency of trihexyphenidyl exceeds that of the antihistamines but is less than that of benztropine. It is considered the most stimulating anticholinergic agent and therefore has a concomitant higher risk of abuse than the antihistamines or benztropine. Trihexyphenidyl has good gastrointestinal absorption, achieving peak plasma concentration in 1 to 2 hours, with a serum half-life of approximately 10 to 12 hours, generally necessitating thrice-daily dosing to achieve satisfactory clinical results. The dosage range is 2 to 5 mg thrice daily.

For parenteral administration, diphenhydramine (Benadryl) and benztropine are most commonly used. Diphenhydramine is an antihistamine with sufficient anticholinergic properties to be useful for extrapyramidal adverse effect treatment. For intramuscular administration in patients with acute dystonia dosages of 25 to 50 mg are typically used. Benztropine was synthesized in the 1950s in an attempt to use the effective moieties of the belladonna alkaloids and the antihistamines. It thus combines a benzhydryl group with a tropine group to create a compound that is more anticholinergic than trihexyphenidyl but less antihistaminic than diphenhydramine. The clinical effect of a single dose appears to last 5 hours, thereby necessitating two or three daily doses. Dosing usually starts at 0.5 to 1 mg twice a day, with a daily maximum of 6 mg, although slightly higher dosages are used in rare circumstances. For treatment of acute dystonia intramuscular doses are 1 to 2 mg.

While no significant pharmacokinetic interactions exist between anticholinergics and other drugs, the studied effect on levodopa kinetics reveals some findings worth examining. In one group of six Parkinson's disease patients receiving levodopa with and without anticholinergic therapy, plasma levodopa kinetics were unchanged, but delayed absorption occurred in one patient, and reduced peak plasma concentrations in two. Impaired absorption of chlorpromazine (Thorazine) was reported in rats administered trihexyphenidyl. Presumably, reduced gastrointestinal motility caused these findings.

Pharmacodynamics Cholinergic receptors are divided into two broad classes: nicotinic and muscarinic, with the former located in striated muscle and autonomic ganglia. Muscarinic receptors are the primary type within the central nervous system (CNS) and are also present on smooth muscle, cardiac muscle, and other tissue with parasympathetic innervation. Five subtypes of muscarinic cholinergic receptors have been cloned (numbered M_1 through M_5), but only 4 distinct pharmacological types can be distinguished M_1, M_2, M_3 and M_4. All anticholinergic drugs used for treatment of extrapyramidal adverse effects are competitive muscarinic antagonists with good CNS penetration. Affinity studies using cloned human muscarinic receptors show little selectivity among the receptor subtypes for the commonly used anticholinergic agents (Table 31.6-1). The parasympatholytic activity of these drugs produces the systemic adverse effects such as dry mouth, cycloplegia, tachycardia, delayed gastrointestinal transit, and urinary hesitancy. Recent data demonstrate that benztropine may also produce its therapeutic effect by the inhibition of dopamine reuptake.

	M_1	M_2	M_3	M_4
Atropine	1	1	1	1
Benztropine	0.46	1.56	1.0	1.83
Biperiden	0.96	7.0	3.54	4.0
Trihexyphenidyl	3.2	7.78	5.82	4.33
Procyclidine	9.2	27.78	11.27	11.67
Diphenhydramine	200.0	133.33	208.2	186.7
Chlorpromazine	50.0	166.67	60.91	66.67
Risperidone	22000	4111	11818	4833
Clozapine	6.2	53.3	18.18	18.33

Adapted from Bolden C, Cusack B, Richelson E: Antagonism by antimuscarinic and neuroleptic compounds at five cloned human muscarinic cholinergic receptors in Chinese hamster ovary cells. J Pharmacol Exp Ther 260:576, 1992.

Table 31.6-1 Relative Antagonism of Human Muscarinic Receptor Subtypes

The importance of central cholinergic activity in parkinsonism was clarified in 1967 by an investigator who demonstrated that peripherally acting agents (agonists and antagonists) had no effect on existing parkinsonism, while physostigmine-induced exacerbation of symptoms was reversible by centrally acting anticholinergics. A concept of cholinergic-dopaminergic balance within the basal ganglia has evolved in which a hypodopaminergic state from Parkinson's disease or neuroleptic use increases relative cholinergic activity. Cholinergic agonism alone does not suffice to produce extrapyramidal adverse effects; however, in the context of ongoing dopaminergic blockade, cholinergic rebound may result in extrapyramidal effects where none had previously existed. (Withdrawal phenomena are discussed below in the section on parkinsonism and long-term usage.)

The propensity of the various antipsychotics to cause extrapyramidal effects is due in part to differences in antagonism of type 2 dopamine (D_2) receptors and inherent anticholinergic activity. High-potency drugs such as haloperidol (Haldol) lack significant muscarinic antagonism and have increased liability to extrapyramidal effects, while thioridazine (Mellaril) and other low-potency antipsychotics are strongly anticholinergic with proportionally fewer extrapyramidal effects. Clozapine (Clozaril) and olanzapine (Zyprexa) are among the most anticholinergic antipsychotics, although their unique properties are felt in part to derive from low affinity for dopamine type 2 (D_2) receptors and significant serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) antagonism. While published binding affinities indicate relative in vitro muscarinic activity, clinical pharmacological equivalence data yield different relative activities, which should be kept in mind when switching between various drugs (Table 31.6-2).

Drug	Pharmacological Equivalent
Atropine	0.5
Benztropine	1
Biperiden	2
Trihexyphenidyl	3.5
Diphenhydramine	30
Other drugs	
Amitriptyline	50
Imipramine	100
Thioridazine	200
Chlorpromazine	300
Clozapine	375

Reprinted with permission from de Leon J, Canuso C, White AO, Simpson GM: A pilot effort to determine benztropine equivalents of anticholinergic medications. Hosp Community Psychiatry 45:606, 1994.

Table 31.6-2 Clinical Benztropine Equivalents (mG)

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Both anticholinergics and amantadine are effective for the prophylaxis and treatment of extrapyramidal adverse effects with the exception of tardive phenomena and akathisia. Akathisia does not respond robustly to these drugs, and other treatment strategies are required including the use of benzodiazepines or b-adrenergic receptor antagonists, dose reduction, or switching to a less potent or atypical antipsychotic. While they possess similar effectiveness, the difference in adverse effect profiles between anticholinergics and amantadine has been demonstrated in several studies. One of the few double-blind, controlled studies noted comparable efficacy between amantadine and benztropine in a group of 44 schizophrenic patients, regardless of extrapyramidal effect severity. While amantadine produced fewer adverse effects, benztropine produced a statistically significant improvement in rigidity. When amantadine or biperiden were blindly substituted for trihexyphenidyl in

one group of 26 chronic schizophrenics no differential effects were seen on parkinsonism or existing tardive dyskinesia severity. Studies in the Parkinson's disease population have also demonstrated comparable efficacy of amantadine and anticholinergic therapy in decreasing symptoms such as tremor, but a controlled 2-week trial comparing trihexyphenidyl and amantadine in Parkinson's disease patients with unilateral tremor revealed greater reduction in tremor amplitude with trihexyphenidyl.

Amantadine's general lack of significant anticholinergic adverse effects in therapeutic doses, especially on memory, constitutes its major advantage over anticholinergic agents for the treatment or prophylaxis of extrapyramidal effects. In one group of 26 schizophrenic patients, significantly worse visual recall was noted in those given biperiden, with poorer performance on all recall tests, but no memory effects were found among those on amantadine. This potential benefit is somewhat mitigated by clinical concerns that the dopaminomimetic properties may exacerbate underlying psychosis in some individuals. Although the aforementioned study revealed no exacerbation of positive symptoms among the amantadine patients, an open study of 30 patients with chronic schizophrenia described worsening of positive symptoms among the more severely ill patients after amantadine was substituted for benztropine.

Prophylaxis Studies The overuse of prophylactic anticholinergic medications has been the subject of numerous studies, prospective and retrospective. Recent prospective studies using lower doses of antipsychotics demonstrate a 20.9 to 33 percent incidence of acute dystonic reactions in patients receiving no prophylaxis (Table 31.6-3). With increased dosages of higher-potency antipsychotics, the incidence of dystonia rose to 47 percent. Regardless of antipsychotic dosage, over 90 percent of all dystonic reactions reported in various studies occur within the first 3 days of treatment. The trends toward higher incidence among younger patients and those of male sex was observed in most studies. In particular, one retrospective study revealed a markedly higher incidence of acute dystonia in the 10- to 19-year-old group (65 percent) than in 20- to 29-year-olds (46 percent) and 30- to 39-year-olds (32 percent). These combined data suggest that only when sufficient dosages (>12 mg a day of haloperidol or its equivalent) of high-potency antipsychotics are required in a group at high risk by age and sex should anticholinergic prophylaxis be entertained on a routine basis. The anticholinergic dosage should be tapered during the second week of treatment unless the emergence of further extrapyramidal effects warrants continuation.

Study	Treatment Group	N	Daily Dosage: Chlorpromazine Equivalent	Dystonia (%)
Kenyon et al. (1983)				
Retrospective	Prophylaxis	80	713	22
	No prophylaxis	135	719	68
Stanek et al. (1986)				
Prospective	Prophylaxis	176	1317	12.9
	No prophylaxis	86	739	20.9
Wirshing et al. (1986)				
Prospective	Prophylaxis	22	1290	0
	No prophylaxis	17	1013	47
			Mean dosages of each drug: thiothixene 32 mg, haloperidol 20 mg, trifluoperazine 25 mg, fluphenazine 23 mg	
Bauer et al. (1987)				
Retrospective	Prophylaxis	7	23 mg haloperidol (range: 8–40 mg)	0
	No prophylaxis	56	18 mg haloperidol (range: 5–35 mg)	93
Goff et al. (1991)				
Prospective	Prophylaxis	84	13.8 mg haloperidol	14
	No prophylaxis	93	14.3 mg haloperidol	33

Table 31.6-3 Anticholinergic Prophylaxis Studies

A consensus statement of the World Health Organization (WHO) published in 1990 summarizes this position:

> On the basis of these considerations, the prophylactic use of anticholinergics in patients on neuroleptic treatment is not recommended, and may be justified only early in treatment (after which it should be discontinued and its need should be reevaluated). As a rule, these compounds should be used only when Parkinsonism has actually developed, and when other measures, such as the reduction on neuroleptic dosage or the substitution of the administered drug by another less prone to induce Parkinsonism, have proven ineffective.

Among the concerns addressed in the WHO consensus statement are the deleterious effects of anticholinergics on both cognition and the development of tardive dyskinesia. Several studies convincingly demonstrate the negative effects of various anticholinergic agents on the cognition of normal volunteers and individuals with schizophrenia. Among patients with schizophrenia, these deficits appear to be independent of neuroleptic dose, intelligence, or clinical status, and reliable improvement in memory occurs upon anticholinergic discontinuation. Whether anticholinergics play a causal role in the development of tardive dyskinesia is unclear, but their long-term use has been associated with an increased risk of tardive dyskinesia. One 7-year prospective trial of 50 patients maintained solely on depot neuroleptics noted a significant ($P < .01$) positive association between anticholinergic use and change in tardive dyskinesia rating.

THERAPEUTIC INDICATIONS

Dystonia While rarely a life-threatening event, development of an acute dystonic reaction can be a frightening, painful experience for the patient that potentially undermines future compliance. Dystonia can be caused by any agent with sufficient dopamine blockade, including low-potency drugs such as prochlorperazine (Compazine). Acute dystonia occurs in the early hours and days of treatment, more often with the administration of high-potency neuroleptics, in younger males, and especially during initial treatment in neuroleptic-naive subjects or after significant dosage increases. Dystonic symptoms may appear in a stuttering or discontinuous fashion, so a patient may complain of tongue thickness or deviation for a minute that resolves, only to later return. This picture often confuses clinicians and staff and may lead to an erroneous conclusion that the patient is malingering or fabricating symptoms. Similarly, anxiety can exacerbate movement disorders arising from the basal ganglia, so that reassurance or the administration of a benzodiazepine may significantly relieve milder dystonias, at times leading clinical staff to attribute the symptoms to other causes.

Depending on severity, benztropine (1 to 2 mg intramuscularly) or diphenhydramine (25 to 50 mg intramuscularly) are most commonly used. Clinical onset is relatively rapid, with full effect seen in 10 to 15 minutes. If symptoms persist, another dose may be administered at that time or sooner with severe reactions or rare instances of airway compromise. Some advocate the addition of small amounts of lorazepam (Ativan) (1 to 2 mg) for those requiring an additional parenteral dose; however, one must monitor the combined sedative effects of multiple drugs on respiration. When the dose of an antipsychotic is subsequently discontinued, significantly reduced, or switched to a low-potency or atypical neuroleptic, there may be no need to continue anticholinergic therapy. If the patient continues to receive the current dosage of neuroleptics, short-term prophylaxis against future episodes should be instituted with benztropine (1 to 2 mg orally twice a day), trihexyphenidyl (2 mg orally three times a day), or amantadine (100 mg orally twice a day).

Parkinsonism and Issues Regarding Long-Term Use The clinical syndrome of neuroleptic-induced parkinsonism is virtually identical to idiopathic Parkinson's disease. In its milder forms, however, it may be difficult to istinguish iatrogenic parkinsonism from the negative symptoms of schizophrenia. The clinician should routinely assess patients taking neuroleptics with a standardized rating scale sensitive to parkinsonism (e.g., the Simpson-Angus Neurological Rating Scale) and maintain a low threshold for treatment. Parkinsonism typically develops after 1 to 3 weeks of treatment and involves the classic symptoms of rigidity, tremor, and bradykinesia. Elderly adults are particularly prone to the parkinsonian effects of all dopamine receptor antagonists. Female sex and use of high-potency neuroleptics are also associated with higher risk for this complication. Clozapine does not induce parkinsonism and was the first drug of choice for the treatment of levodopa-induced psychosis in Parkinson's disease prior to the release of quetiapine. Clozapine's effect in this group is seen at very low dosages, in the range of 6.25 to 25 mg a day.

Once diagnosed, dose reduction is the first option, as parkinsonism responds slowly and at times incompletely to anticholinergics and amantadine. These drugs should be considered temporizing measures only, to alleviate symptoms while modifying the antipsychotic regimen. Elderly persons are particularly sensitive to the cognitive effects of anticholinergic medications, and these drugs should be avoided in this population. Amantadine treatment can be started at 100 mg twice a day and titrated upward over a period of days to a maximum of 400 mg a day as necessary. It is also available as a 50 mg/5 mL syrup to permit lower starting dosages in elderly patients or those with renal dysfunction. Any of the common antiparkinsonian medications may also be used, starting at the equivalent of 1 mg benztropine twice daily. If antipsychotic dose reduction is not possible or effective, one can switch to any of several available atypical antipsychotics that exert significantly less dopamine D_2 blockade.

Another important issue regarding continued treatment with combined anticholinergics and antipsychotics is the induction of cholinergic sensitivity induced by chronic muscarinic blockade. Schizophrenic patients maintained on anticholinergics have rates of extrapyramidal effects as high as 68 percent when these medications are

abruptly discontinued. These high rates of extrapyramidal adverse effects can be understood as the result of disturbed cholinergic-dopaminergic balance in the basal ganglia, which drives the system toward a hypodopaminergic state because of cholinergic overactivity. Even residual serum concentrations of competitive dopamine antagonists suffice to cause extrapyramidal effects in this context. This mechanism of withdrawal extrapyramidal adverse effects is illustrated by a study of individuals with chronic schizophrenia who were titrated on trifluoperazine (Stelazine) to produce a predefined level of extrapyramidal effects. No significant symptoms occurred upon abrupt withdrawal of trifluoperazine alone; however, patients subsequently given the same trifluoperazine dose plus benztropine for 4 weeks showed a marked increase in both extrapyramidal effects and other physical symptoms of cholinergic rebound when administration of both was abruptly discontinued. The availability of serotonin-dopamine antagonists with low D_2 binding should obviate the need for long-term combined anticholinergic-antipsychotic therapy in the near future. Patients currently on such regimens should be slowly tapered off the anticholinergic medication at a rate of 0.5 mg of benztropine equivalent per week. More-precipitant withdrawal may cause extrapyramidal effects, other physical symptoms, and possible exacerbation of underlying psychosis.

Akathisia Akathisia is a neuroleptic-induced syndrome of restlessness with both mental and physical components, associated with the use of antipsychotics and other drugs including serotonin reuptake inhibitors. These patients experience an intense subjective sense of anxiety that is extremely distressing. The complications of akathisia include noncompliance, frank agitation, and (rarely) suicidal ideation or behavior in an attempt to relieve the intense dysphoria that accompanies this feeling of restlessness. Akathisia must be sought when examining patients on neuroleptics for adverse effects, as behavioral manifestations may be subtle or absent in milder forms. Reliable, standardized akathisia scales are available to document the progress of interventions. While differentiating akathisia from anxiety due to underlying psychosis can be difficult, worsening after the administration of neuroleptics, particularly high-potency agents, should raise the level of suspicion. Among the antipsychotics, dopaminergic blockade is essential for development of akathisia, with an increased incidence noted when potent antipsychotics are used and near absence with clozapine administration. Nonetheless, anticholinergics and amantadine are only modestly successful in relieving akathisia, implying central mechanisms involving other subcortical structures and possibly spinal dopamine or norepinephrine systems.

For milder symptoms in relatively stable patients, anticholinergics alone can be used and are sometimes quite effective. Benzodiazepines and the more lipophilic b-adrenergic receptor antagonists have documented efficacy and should be used for more severe symptoms. Experience with amantadine in treating akathisia is limited. Once akathisia is diagnosed, dose reduction strategies should be used if possible, as symptoms may be ongoing. The experience with clozapine also points to favorable outcomes with serotonin-dopamine antagonists.

Direct Effects on Positive and Negative Symptoms Involvement of cholinergic mechanisms in the pathophysiology of schizophrenia has gained increasing popularity in recent years, especially in regard to clozapine's postulated mechanism of action. Interpretation of data from add-on studies is problematic, given the inherent difficulty in distinguishing the improvement in negative symptoms due to the anticholinergic medication from a decrease in parkinsonism. Attempts at antipsychotic-free trials have been performed with mixed results. In one double-blind, placebo-controlled crossover study of 17 stable outpatients with schizophrenia taking no antipsychotic medication for 2 months, no significant differences were noted in levels of psychosis, depression, Brief Psychiatric Rating Scale (BPRS) score, or negative symptoms after administration of trihexyphenidyl for 4 weeks. When 40 schizophrenic inpatients, medication-free for 2 weeks, were given 4 mg twice a day of biperiden for 2 days, investigators noted both an increase in positive symptoms ($P < .001$) and a decrease in negative symptoms ($P < .01$).

PRECAUTIONS AND ADVERSE REACTIONS

Mild systemic symptoms can be seen with therapeutic dosages of any anticholinergic medication, and the additive effects of various medications may give rise to symptoms of toxicity. Elderly persons are especially prone to develop memory impairment or frank delirium when using even modest dosages of drugs with anticholinergic properties. All such agents should be eschewed as much as possible in this age group. Urinary retention is a specific concern in older males. Care should also be exercised in patients with angle-closure glaucoma, those with known cognitive impairment, and those with poorly compensated heart disease who may not tolerate tachycardia.

Toxicity from accidental or intended overdose produces symptoms including tachycardia, dry mucosa and skin, and confusion or delirium, at times with hallucinations. When severe, the patient may be comatose with autonomic instability, cardiac arrhythmias, and a flushed appearance (“red as a beet”) with elevated body temperature. The latter is more common in the very young because of impaired cholinergic-mediated sweating. Anticholinergic overdose should always be treated in an emergency room setting where proper life support measures are available. Physostigmine, an acetylcholinesterase inhibitor with good CNS penetration, should be administered intravenously in doses of 0.5 to 2.0 mg in more serious cases. This medication must be given with continuous cardiac monitoring, as it may induce unstable cardiac rhythms. Parenterally administered physostigmine is metabolized within 2 hours, so repeat doses may be necessary.

Epidemiologically based estimates of anticholinergic abuse vary widely (1 to 17 percent), but certain patterns emerge from these data. In general, these agents are used for their direct euphoriant effects or to potentiate the effects of alcohol or psychoactive drugs. The more-stimulating anticholinergic trihexyphenidyl has the highest potential for abuse and can be purchased on the street. Some schizophrenic patients may misuse their anticholinergic medications because of the subjective feeling of improvement in negative symptoms at higher doses, although no euphoria is experienced. For this reason or possibly because of previous experiences with extrapyramidal effects weaning certain patients from their anticholinergic regimen is sometimes very difficult.

LABORATORY INTERFERENCES

The anticholinergic medications and amantadine are not known to affect directly the results of urinalysis, routine serum chemistries, or blood counts.

DOSAGE AND ADMINISTRATION

Amantadine is available as 100-mg capsules, typically initiated at 100 mg twice daily and titrated over several days to a maximum of 400 mg a day as necessary. The typical dosage is 100 mg thrice daily. For elderly persons or those with renal dysfunction, it is also available as a 50-mg/5-mL syrup to permit lower starting dosages. Table 31.6-4 outlines the dosage forms and average daily oral dosage for the common anticholinergic medications available in the United States.

Generic Name	Brand Name	Preparations	Daily Oral Dosage
Benztropine	Cogentin	Tablets: 0.5, 1.0, 2.0 mg Injection: 1 mg/mL	0.5–2.0 mg BID-TID
Biperiden	Akineton	Tablets: 2 mg Injection: 5 mg/mL	2 mg BID-QID
Trihexyphenidyl	Artane	Tablets: 2, 5 mg	2–5 mg TID
Diphenhydramine	Benadryl	Capsules: 25, 50 mg Injection: 25–50 mg/mL Elixir: 12.5 mg/5 mL	25–50 mg BID, 25 mg QID

Table 31.6-4 Anticholinergic Dosages and Preparations

Acute dystonic reactions require the use of a parenteral agent, with benztropine (1 to 2 mg intramuscularly) or diphenhydramine (25 to 50 mg intramuscularly or intravenously) being the drugs most readily available. For prophylaxis of extrapyramidal effects or treatment of parkinsonism, the lower dosage range is used to initiate therapy with all agents. Subsequent dosage is titrated on the basis of the tolerance of adverse effects or the emergence of extrapyramidal effects.

SUGGESTED CROSS-REFERENCES

Detailed information on medication-induced movement disorders can be found in Section 31.4. Further characteristics of dopamine receptor antagonists are extensively covered in Section 31.17. Section 31.15 on cholinesterase inhibitors provides interesting background on cholinergic pathways within the CNS.

SECTION REFERENCES

Baker LA, Cheng LY, Amara IB: The withdrawal of benztropine mesylate in chronic schizophrenic patients. Br J Psychiatry *143:* 584, 1983.

Bergen J, Kitchin R, Berry G: Predictors of the course of tardive dyskinesia in patients receiving neuroleptics. Biol Psychiatry *32:* 580, 1992.

*Blaisdell GD: Akathisia: A comprehensive review and treatment summary. Pharmacopsychiatry *27:* 139, 1994.

Bolden C, Cusack B, Richelson E: Antagonism by antimuscarinic and neuroleptic compounds at five cloned human muscarinic cholinergic receptors in Chinese hamster ovary cells. J Pharmacol Exp Ther *260:* 576, 1992.

Curran HV, Pooviboonsuk P, Dalton JA, Lader MH: Differentiating the effects of centrally acting drugs on arousal and memory: An event-related potential study of scopolamine, lorazepam and diphenhydramine. Psychopharmacology *135:* 27, 1998.

de Leon J, Canuso C, White AO, Simpson GM: A pilot effort to determine benztropine equivalents of anticholinergic medications. Hosp Community Psychiatry *45:* 606, 1994.

DiMascio A, Demirgian E: Antiparkinson drug overuse. Psychosomatics *11:* 596, 1970.

DiMascio A, Diosdado BL, Greenblatt DJ, Marder JE: A controlled trial of amantadine in drug-induced extrapyramidal disorders. Arch Gen Psychiatry *33:* 599, 1976.

Duvoisin R: Cholinergic-anticholinergic antagonism in parkinsonism. Arch Neurol *17:* 124, 1967.

*Elliott KJ, Lewis S, el Mallakh RS, Looney SW, Caudill R, Bacani Oropilla T: The role of parkinsonism and antiparkinsonian therapy in the subsequent development of tardive dyskinesia. Ann Clin Psychiatry *6:* 197, 1994.

Factor SA, Molho ES, Brown DL: Acute delirium after withdrawal of amantadine in Parkinson7apos;s disease [see comments]. Neurology *50:* 1456, 1998.

Goff DC, Amico E, Dreyfuss D, Ciraulo D: A placebo-controlled trial of trihexyphenidyl in unmedicated patients with schizophrenia. Am J Psychiatry *151:* 429, 1994.

Goff DC, Arana GW, Greenblatt DJ, Dupont R, Ornsteen M, Harmatz J, Shader RI: The effect of benztropine on haloperidol-induced dystonia, clinical efficacy and pharmacokinetics: A prospective double-blind trial. J Clin Psychopharmacol *11:* 106, 1991.

Hayden FG: Antimicrobial agents: Antiviral agents. In *The Pharmacological Basis of Therapeutics*. ed 9, LS Goodman, A Gilman, JG Hardman, AG Gilman, editors. McGraw-Hill, New York, 1996.

Heinik J: Effects of trihexyphenidyl on MMSE and CAMCOG scores of medicated elderly patients with schizophrenia. International Psychogeriatrics *10:* 103, 1998.

Katz IR, Sands LP, Bilker W, DiFilippo S, Boyce A, D'Angelo K: Identification of medications that cause cognitive impairment in older people: The case of oxybutynin chloride. J Am Geriatr Soc *46:* 8, 1998.

Keepers GA, Clappison VJ, Casey DE: Initial anticholinergic prophylaxis for neuroleptic-induced extrapyramidal syndromes. Arch Gen Psychiatry *15:* 1113, 1983.

Koller W: Pharmacologic treatment of parkinsonian tremor. Arch Neurol *43:* 126, 1986.

Land W, Pinsky D, Salzman C: Abuse and misuse of anticholinergic medications. Hosp Community Psychiatry *42:* 580, 1991.

Manos N, Gkiouzepas J, Logothetis J: The need for continuous use of antiparkinsonian medication with chronic schizophrenic patients receiving long-term neuroleptic therapy. Am J Psychiatry *138:* 184, 1981.

McInnis M, Petursson H: Withdrawal of trihexyphenidyl. Acta Psychiatry Scand *71:* 297, 1985.

Meszaros K, Lenzinger E, Hornik K, Schonbeck G, Hatzinger R, Langer G, Sieghart W, Aschauer HN: Biperiden and haloperidol plasma levels and extrapyramidal side effects in schizophrenic patients. Neuropsychobiology *36:* 69, 1997.

Ordenstein L: *Sur la paralysie et la sclerose en plaque generalise*. Martinet, Paris, 1867.

Purkerson SL, Potter LT: Use of antimuscarinic toxins to facilitate studies of striatal m4 muscarinic receptors. J Pharmacol Exp Ther *284:* 707, 1998.

Roldan G, Bolanos Badillo E, Gonzales Sanchez H, Quirarte GL, Prado Alcala RA: Selective M1 muscarinic receptor antagonists disrupt memory consolidation of inhibitory avoidance in rats. Neurosci Lett *230:* 93, 1997.

Silver H, Geraisy N: Effects of biperiden and amantadine on memory in medicated chronic schizophrenic patients. A double-blind cross-over study. Br J Psychiatry *166:* 241, 1995.

Silver H, Geraisy N, Schwartz M: No difference in the effect of biperiden and amantadine on parkinsonian- and tardive dyskinesia-type involuntary movements: A double-blind crossover, placebo-controlled study in medicated chronic schizophrenic patients. J Clin Psychiatry *56:* 167, 1995.

Simpson GM, Amin M, Kunz E: Withdrawal effects of phenothiazines. Compr Psychiatry *6:* 347, 1965.

Sramek JJ, Simpson GM, Morrison RL, Heiser JF: Anticholinergic agents for the prophylaxis of neuroleptic-induced dystonic reactions: A prospective study. J Clin Psychiatry *47:* 305, 1986.

*Sweeney JA, Keilp JG, Haas GL, Hill J, Weiden PJ: Relationships between medication treatments and neuropsychological test performance in schizophrenia. Psychiatry Res *37:* 297, 1991.

Tandon R, DeQuardo JR, Goodson J, Mann NA, Greden JF: Effect of anticholinergics on positive and negative symptoms in schizophrenia. Psychopharmacol Bull *28:* 297, 1992.

Tune LE, Strauss ME, Lew MF, Breitlinger E, Coyle JT: Serum levels of anticholinergic drugs and impaired recent memory in chronic schizophrenic patients. Am J Psychiatry *139:* 1460, 1982.

*Ungvari GS, Chiu HF, Lam LC, Pang AH, Chung DS, Li S-W, Chiu S-N, Lum FC, Leung T: Gradual withdrawal of long-term anticholinergic antiparkinson medication in Chinese patients with chronic schizophrenia. J Clin Psychopharmacol *19:* 141, 1999.

*Vaughan RA, Agoston GE, Lever JR, Newman AH: Differential binding of tropane-based photoaffinity ligands on the dopamine transporter. J Neurosci *19:* 630, 1999.

*Vitiello B, Martin A, Hill J, Mack C, Molchan S, Martinez R, Murphy DL, Sunderland T: Cognitive and behavioral effects of cholinergic, dopaminergic, and serotonergic blockade in humans. Neuropsychopharmacology *16:* 15, 1997.

Wilcox JA, Tsuang J: Psychological effects of amantadine on psychotic subjects. Neuropsychobiology *23:* 144, 1990.

Winslow GS, Stillner V, Coons DJ, Robinson MW: Prevention of acute dystonic reactions in patients beginning high-potency neuroleptics. Am J Psychiatry *143:* 706, 1986.

World Health Organization: Prophylactic use of anticholinergics in patients on long-term neuroleptic treatment. Br J Psychiatry *156:* 412, 1990.

Yang CC, Deng JF: Anticholinergic syndrome with severe rhabdomyolysis—an unusual feature of amantadine toxicity (Letter). Intensive Care Med *23:* 355, 1997.

Young BK, Camicioli R, Ganzini L: Neuropsychiatric adverse effects of antiparkinsonian drugs. Characteristics, evaluation and treatment. Drugs Aging *10:* 367, 1997.

31.7 ANTICONVULSANTS

31.7A CARBAMAZEPINE

CARLOS A. ZARATE, JR., M.D., AND MAURICIO TOHEN, M.D., DR.P.H.

Carbamazepine (Tegretol) is an iminostilbene drug with a chemical structure similar to that of the tricyclic drug imipramine (Tofranil). Carbamazepine was developed in the late 1950s in the laboratories of J. R. Geigy in Basel, Switzerland. Its antiepileptic properties were first reported in 1963. Carbamazepine was approved for use in the United States for the treatment of trigeminal neuralgia in March 1968 and for temporal lobe epilepsy in August 1974. Carbamazepine was first synthesized as a potential antidepressant but was later found to have a beneficial effect in the treatment of a variety of medical conditions, including seizure disorders and paroxysmal pain syndromes. Carbamazepine is particularly effective in the treatment of complex partial seizures and generalized seizures thought to arise from the temporal lobe and limbic systems. Additionally, carbamazepine has been increasingly recognized as an effective alternative to lithium (Eskalith, Lithobid) and valproate (Depakene) in the treatment of bipolar disorder.

CHEMISTRY

Carbamazepine's chemical name is 5-*H*-dibenz[*b,f*]azepine-5-carboxamide. When a carbamyl (carboxamide) group is added to the 5 position of iminodibenzyl, considerable antiepileptic activity is conferred. However, a more intense antiepileptic effect occurs when the carbamyl side chain is combined with iminostilbene to make carbamazepine; the structure is similar to that of iminodibenzyl except for a double bond between the 10 and 11 positions (Fig. 31.7a-1). Carbamazepine is oxidized by cytochrome P450 oxidase to an active 10,11-epoxide metabolite that possesses both anticonvulsant and antinociceptive properties. The epoxide is further hydrolyzed to a dihydroxy compound that is inactive. A congener of carbamazepine includes a keto compound called oxcarbazepine, which also has anticonvulsant and antinociceptic properties. Carbamazepine is highly insoluble in water and is difficult to dissolve in most preparations. Because of its insolubility, no intravenous formulation is available for human use.

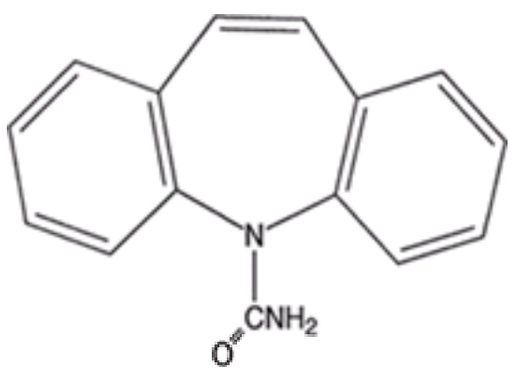

FIGURE 31.7a-1 Molecular structure of carbamazepine.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Carbamazepine is absorbed slowly and erratically through the gastrointestinal tract. Peak plasma concentrations are generally attained 2 to 8 hours after oral ingestion. However, the suspension is absorbed somewhat faster, and the carbamazepine extended-release (Tegretol-XR) tablet slightly slower, than the conventional tablet. Following a twice-a-day dosage regimen, the suspension provides higher peak concentrations and lower trough concentrations than those obtained with conventional tablets. Carbamazepine extended-release tablets offer the advantage of providing steady-state plasma concentrations comparable to those obtained with conventional carbamazepine tablets given four times a day. Carbamazepine distributes rapidly into all tissues, with 70 to 80 percent bound to plasma proteins. Concentrations of carbamazepine in cerebrospinal fluid correspond with concentrations of free drug in plasma and range from 17 to 31 percent of those in plasma. The average half-life after a single dose is 26 hours (18 to 54 hours), but with long-term treatment, carbamazepine may induce its own metabolism (a phenomenon called autoinduction), and its half-life may be decreased to 5 to 26 hours. Carbamazepine's irregular absorption has been attributed to the anticholinergic properties of the drug modifying gastrointestinal transit time and to a very slow dissolution rate in gastrointestinal fluid.

Elimination Carbamazepine is metabolized in the liver, mainly by the microsomal cytochrome P450, and excreted by the kidneys. Only 1 percent is eliminated by biliary excretion. The predominant pathway of metabolism in humans involves conversion to the 10,11-epoxide by the cytochrome P450 (CYP) isozyme 3A4 (CYP 3A4). This metabolite is as active as carbamazepine and has neurotoxic effects. Its concentrations in plasma and brain may reach 50 percent of those of carbamazepine. The keto derivative of carbamazepine, oxcarbazepine, is metabolized only minimally by CYP 3A4, so that autoinduction and heteroinduction are less clinically problematic than with carbamazepine. The half-life of the 10,11-epoxide is much shorter than that of the parent compound (6 to 7 hours). Its metabolism results in the inactive monohydroxy derivative rather than the active metabolite of carbamazepine. After oral administration of [^{14}C]-carbamazepine, 72 percent of the administered radioactivity was found in the urine and 28 percent in the feces. The urinary radioactivity was composed largely of hydroxylated and conjugated metabolites, with only 3 percent of unchanged carbamazepine. Younger age groups (below the age of 15) metabolize carbamazepine to carbamazepine 10,11-epoxide more rapidly than adults. The effects of race and sex on the pharmacokinetics of carbamazepine have not been systematically studied.

Pharmacodynamics

Mechanism of Action Despite the widespread clinical use of carbamazepine, the molecular mechanisms underlying both its anticonvulsant and mood-stabilizing effects have not been identified. Carbamazepine's antiepileptic action is believed to occur by two basic mechanisms: (1) effects on neuronal ion channels, and (2) effects on synaptic and signal transduction pathways. The first mechanism has been attributed to either its ability to reduce high-frequency neuronal discharge by binding and inactivating voltage-sensitive sodium channels and decreasing sodium influx, or its effect on potassium channels to increase potassium conductance, or both. However, because carbamazepine's effects on sodium channels are acute, and because its antiepileptic and antinociceptic effects are more rapid in onset than its antimanic or antidepressant effects, some authors have speculated that the drug's sodium channel effects do not account for its mood-stabilizing properties.

The anticonvulsant effects of carbamazepine on amygdala-kindled seizures have been attributed to its ability to decrease release of the excitatory amino acid aspartate and its effects on a_2-adrenergic receptors. Also, it has been suggested that carbamazepine's antiepileptic actions result from its acute binding and antagonistic action at the peripheral-type benzodiazepine receptors rather than central-type benzodiazepine receptors, (which are linked to chloride channels and related to the antiepileptic effects of diazepam [Valium], clonazepam [Klenopin], and lorazepam [Ativan]. The anticonvulsant actions of carbamazepine have also been

linked to a variety of other neurotransmitter systems including complex effects on norepinephrine and decreases in g-aminobutyric acid (GABA) turnover as well as an ability to decrease calcium influx through *N*-methyl-d-aspartate (NMDA) receptors. A recent study failed to find that the inhibition of glutamate release was responsible for carbamazepine's antiepileptic action.

The antinociceptive effects of carbamazepine have been linked to its effects on the type B ($GABA_B$) receptors, as the antinociceptive effects of both carbamazepine and baclofen (Lioresal) are inhibited by $GABA_5$ antagonists.

Long-term administration of carbamazepine is associated with increases in adenosine receptors, substance P levels, and sensitivity; decreased somatostatin concentration in the cerebrospinal fluid; and greater decreases in GABA turnover. Other actions reported to occur with carbamazepine include complex effects on second messenger systems such as decreased activity of both basal cyclic adenosine monophosphate (cAMP) and forskolin-stimulated cAMP production. It appears that the action of carbamazepine on cAMP is likely mediated through an inhibitory G-independent protein mechanism. To date, the data suggest that carbamazepine inhibits cAMP production by acting directly on adenylyl cyclase, through factors that are associated with adenylyl cyclase, or both. In addition, carbamazepine decreases some aspects of phosphoinositide turnover that are not dissimilar from those of lithium carbonate. Carbamazepine has been reported to stimulate, not inhibit, inositol monophosphatase.

The antimanic effects of carbamazepine show a lag in onset. Carbamazepine has been shown to attenuate forskolin-induced c-*fos* (an immediate-early gene). c-*fos* and other immediate-early genes are known to be involved in a number of long-term neuronal responses. Thus, the inhibitory effects of carbamazepine on adenylyl cyclase can potentially bring about long-term changes in gene expression and thereby regulate cellular function. This may explain in part the lag in carbamazepine's antimanic effects. Recent studies have attributed the efficacy of mood stabilizers to a protein kinase substrate, the myristoylated alanine-rich C kinase substrate (MARCKS) protein. MARCKS binds calmodulin, which has been implicated in cellular processes associated with cytoskeletal restructuring and neuroplasticity. While lithium and valproate have been found to produce a significant time-dependent downregulation of MARCKS expression, carbamazepine has not. These data suggest that the mechanism mediating the clinical efficacy of carbamazepine in the treatment of mania may be independent of that of lithium and valproate.

None of these mechanisms have been definitely linked with the psychotropic effects of carbamazepine, and it remains unknown whether the actions underlying the drug's antiepileptic effects are also responsible for its mood-stabilizing properties.

CLINICAL DRUG STUDIES

Bipolar Disorder

Manic Episodes Over a dozen double-blind studies comparing carbamazepine with a placebo, with lithium, or with antipsychotic agents have been published. Most studies indicate positive antimanic effects for carbamazepine and find it superior to placebo and comparable to lithium and antipsychotic agents for the treatment of acute mania, with approximately two-thirds of patients showing improvement. However, only a few of these studies are unconfounded by concomitant medications. Pooled results from these latter studies reveal an overall response rate for carbamazepine in manic episodes of 50 percent, compared with 56 percent for lithium monotherapy and 61 percent for dopamine receptor antagonist monotherapy (differences not significant). Its antimanic effects may be augmented by, or synergistic with concurrent administration of other mood-stabilizing agents (i.e., lithium and valproate) and dopamine receptor antagonists.

Bipolar Depression Only three controlled studies have found carbamazepine effective for the depressive episodes in bipolar disorders, the response rate approximating 30 percent. Some studies have reported that its antidepressant effects may be augmented by concurrent administration of lithium. Factors suggesting a positive antidepressant response to carbamazepine during the depressed phase of bipolar I disorder include a greater thyroid hormone decrement thyroxine [T_4] or free T_4).

Prophylaxis Several controlled long-term prophylaxis studies suggest that carbamazepine is effective in preventing the recurrences of both manic and depressive episodes. A meta-analysis of eight randomized double-blind studies that compared carbamazepine with lithium for the maintenance treatment of bipolar I disorder found both drugs to be comparable in reducing relapses and prolonging euthymic intervals over an average of 1.7 years of treatment. Also, carbamazepine's mood-stabilizing effects are reportedly augmented by the concurrent administration of lithium, valproate, thyroid hormones, antipsychotic agents, and antidepressants. As with other mood-stabilizing drugs (i.e., lithium and valproate), carbamazepine appears better for preventing a recurrence of manic than depressive episodes. Some authors report that tachyphylaxis may occur in patients maintained on carbamazepine. This apparent loss of efficacy has been hypothesized to be due to tolerance or to progression of the illness. Although the available partially controlled, long-term, prophylactic studies have been criticized for various methodological flaws, they are highly consistent with a much larger uncontrolled case series indicating carbamazepine's efficacy and tolerability.

Predictors of Response Initial studies reported that a history of poor response to lithium, rapid cycling, severe mania, mixed or dysphoric mania, and a lower prevalence of familial bipolar disorders predicted a favorable outcome to carbamazepine therapy. More-recent studies indicate that decreasing or stable episode frequencies and decreasing mania severity correlate with favorable carbamazepine response.

Blood Concentrations and Relation to Action Therapeutic blood concentrations for carbamazepine as an anticonvulsant are between 4 and 12 µg/mL; however, the ideal range for carbamazepine as a mood stabilizer is unknown. Standard clinical practice is to achieve a range of 8 to 12 µg/mL. In general, monitoring of plasma concentrations is recommended until stability, monthly for the next 6 months, and every 6 to 12 months thereafter (Table 31.7a-1).

	Baseline	Weekly to Stability	Monthly for 6 Mo	6–12 Mo
CBC	+	+	+	+
Bilirubin	+		+	+
Alanine aminotransferase	+		+	+
Aspartate aminotransferase	+		+	+
Alkaline phosphatase	+		+	+
Carbamazepine level		+		+

CBC, complete blood count.

Table 31.7a-1 Laboratory Monitoring of Carbamazepine for Adult Psychiatric Disorders

After approximately 2 to 3 weeks of administration, carbamazepine induces hepatic enzymes that facilitate metabolism of the drug (autoinduction of metabolism.) Thus, a patient who is unable to tolerate 800 or 1000 mg a day in the first several weeks of treatment may not experience adverse effects at these or higher dosages after the third or fourth week of treatment.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The most common adverse effects of carbamazepine administration are neuromuscular. Adverse events may be divided into those that are largely dosage dependent and those that are idiosyncratic (Table 31.7a-2). Adverse events from carbamazepine that are dosage dependent may be minimized by either dosage reduction or by more gradual titration of the dosage.

Dosage-Related Adverse Effects	Idiosyncratic Adverse Effects
Double or blurred vision	Agranulocytosis
Vertigo	Stevens-Johnson syndrome
Gastrointestinal disturbances	Aplastic anemia
Task performance impairment	Hepatic failure
Hematological effects	Rash
	Pancreatitis

Table 31.7a-2 Adverse Events Associated With Carbamazepine

Central Nervous System The most common neurological adverse effects include dizziness, sedation, ataxia, and diplopia. These effects can largely be minimized or avoided with slow, gradual titration of the dosage. Carbamazepine may also produce mild peripheral polyneuropathies and involuntary movements. Carbamazepine's effects on memory and psychomotor performance have been reported to be minimal in comparison with those of the more behaviorally toxic anticonvulsants, such as phenobarbital. Several studies suggest that carbamazepine's effect on cognition (as determined by neuropsychiatric tests) was not significantly different from that observed with lithium. Initial studies of carbamazepine in normal volunteers and in epileptic patients suggest that it has little adverse effect on cognitive processes, especially compared with the anticonvulsant phenytoin (Dilantin) and phenobarbital. A recent report suggests that the effects of carbamazepine and valproate monotherapy on cognitive functioning are similar and that both drugs produce minimal effects compared with pretreatment baseline performance.

Blood Carbamazepine appears to cause a mild leukopenia in approximately 1 to 2 percent of patients. In most patients treated, this leukopenia is mild and is usually without clinical impact as the lower white blood cell count is usually within normal range. The leukopenia primarily involves granulocytes but does not predispose patients to infection. Reports of agranulocytosis in aplastic anemia associated with carbamazepine continue to appear in the literature. The most recent estimate of incidents is eight patients per million treated, or approximately 1 in 125,000. The benign suppressive effects on granulocytes are thought to be exerted by an inhibitory action at the level of colony-stimulating factor in the bone marrow. In the event of asymptomatic leukopenia or thrombocytopenia, the carbamazepine dosage can be reduced or (in patients with severe changes) the drug discontinued. There is considerable disagreement in the neurological and hematological literature as to the recommended frequency of hematological monitoring to deal adequately with this rare but potentially fatal complication of carbamazepine therapy. The most conservative recommendations are to obtain a complete blood count (CBC) at baseline, weekly to stability and 6 to 12 months thereafter (Table 31.13-1). In addition, patients should be counseled to contact the physician immediately in the event of fever, sore throat, malaise, petechiae, or other signs of hematological dysfunction.

Hormones Many studies have reported that carbamazepine suppresses circulating concentrations of the thyroid hormone T_4, free T_4, and triiodothyronine (T_3) without significantly increasing thyroid-stimulating hormone (TSH) concentrations; however, the methods used in those studies involved dilution of serum. A recent study suggests that carbamazepine-treated patients initially exhibit a transient increase in free T_4, a transient decrease in serum TSH, decrease in total T_4, and normal free T_4 in the serum. However, in patients receiving long-term treatment with carbamazepine, the concentrations of free T_4 and free T_3 in serum are within the normal range.

Based on the drug's ability to stimulate vasopressin receptor function directly or indirectly, carbamazepine on occasion has been associated with hyponatremia and (rarely) with water intoxication. This retention of water in dilutional hyponatremia thus has an effect opposite to that of lithium, which impairs the suppressant function and is associated with nephrogenic diabetes insipidus. Carbamazepine is an alternative to lithium for patients with severe diabetes insipidus. However, carbamazepine does not reverse the effects of lithium-induced diabetes insipidus when the two drugs are used concurrently. Patients on high dosages of carbamazepine and elderly patients appear more prone to hyponatremia. Checking serum electrolytes would appear warranted, as it would when any patient develops a confusional syndrome or a sign that could be associated with hyponatremia (such as severe weakness or headache).

A recent 5-year prospective follow-up study of 36 consecutive patients with recently diagnosed idiopathic epilepsy treated with carbamazepine found changes in serum lipid levels. Serum total cholesterol and high-density lipoprotein cholesterol concentrations increased after 2 months of treatment with carbamazepine and remained high after 1 and 5 years. However, concentrations of low-density lipoprotein cholesterol and triglycerides in serum increased only initially and did not remain elevated by 1 and 5 years follow-up. The authors suggest that the change in lipid metabolism may be associated with induction of liver enzymes during carbamazepine medication. Carbamazepine increases total serum cholesterol largely by increasing the high-density lipoproteins, so that the effect is unlikely to have deleterious cardiovascular consequences. Other endocrinological changes include decreases in free androgen concentrations and increases in the excretion of urinary free cortisol.

Liver Occasional mild elevations in liver function tests are relatively frequent in many series of psychiatric and neurological patients treated with carbamazepine. They do not appear to warrant special attention or drug discontinuation. With more persistent enzyme elevation, drug discontinuation would appear indicated. Rarely, carbamazepine has been associated with severe cases of hepatitis, which on pathological examination showed either an inflammatory or a toxic-metabolic picture. Liver function should therefore be monitored at 6- to 12-month intervals.

Heart Cardiac toxic reactions related to carbamazepine occur but are infrequent. Conduction disturbances (sometimes resulting in bradycardia or Stokes-Adams syndrome) have been reported.

Skin Carbamazepine-induced rashes are common in psychiatric patients. Although most reports in the literature indicate that rashes are observed in only a small percentage of patients treated with carbamazepine, certain authors suggest that rashes occur in 12 to 16 percent of patients. The risk of having a serious adverse skin reaction appears to be 3 cases per million users per week. The rashes induced by carbamazepine appear to be pleomorphic. The most common is a macular papular rash associated with pruritus emerging usually within the first 3 weeks after initiation of drug use. In most cases rashes resulting from carbamazepine are benign and self-limiting; however, more severe cases may occur such as exfoliative dermatitis, erythema multiforme, Stevens-Johnson syndrome, or toxic epidermal necrolysis (Lyell's syndrome). Therefore, most authors recommend discontinuing carbamazepine with the onset of rash, because exfoliate reactions and Stevens-Johnson syndrome can occur but are relatively uncommon. Some patients have been reexposed to carbamazepine without the reemergence of a rash, but in most cases the rash returns, usually with a rapid onset. Until recently, valproic acid had often been viewed as safer with respect to cutaneous reactions. However, one study suggests that the crude relative risk of severe adverse cutaneous reactions including Stevens-Johnson syndrome and toxic epidermal necrolysis is similar for carbamazepine and valproic acid if the drugs are used for short periods (<2 months) but is much higher for carbamazepine when the drugs are used for longer periods.

Weight In contrast to other psychotropic agents used in the treatment of mood disorders, such as lithium and valproic acid, carbamazepine does not appear to be associated with significant weight gain.

Other Adverse Effects Rare idiosyncratic adverse effects reported include psychological disturbances (e.g., sporadic cases of mania and psychosis), pancreatitis, and renal complications (e.g., renal failure, oliguria, hematuria, and proteinuria).

THERAPEUTIC INDICATIONS

Carbamazepine is approved by the Food and Drug Administration (FDA) for use in the following seizure types: (1) partial seizures with complex symptomatology (psychomotor, temporal lobe, (2) generalized tonic-clonic seizure (grand mal), and (3) mixed seizure patterns that include the above. It is also approved for use in a variety of paroxysmal pain syndromes, including trigeminal neuralgia. However, carbamazepine is not FDA approved for the treatment of bipolar disorders. Nevertheless, carbamazepine is being widely used in both short-term and long-term treatment of mood disorders. However, because of the availability of lithium and divalproex (Depakote) and their FDA approval for bipolar I disorder, carbamazepine should be considered a second-line agent for this indication. Carbamazepine is often used as a substitute to lithium or valproate or as an adjunct to lithium for the partial responder. In patients with refractory seizure disorders and refractory mood disorders, it is occasionally used in combination with valproate. In these instances, the dosage of carbamazepine should be decreased (as valproic acid displaces carbamazepine from its protein-blind sites and also increases the concentrations of the 10,11-epoxide metabolite), and valproate dosages may need to be increased

to maintain therapeutic concentrations.

A 2- to 3-week trial of carbamazepine in the treatment of a manic episode usually suffices to ascertain whether a patient is going to respond, although longer periods may be required to assess antidepressant efficacy. Some 50 to 70 percent of patients show a reduction in manic and depressive episodes with the use of carbamazepine for prophylaxis. In general, lithium may be considered as an adjunctive treatment in patients who have achieved a partial response on carbamazepine alone. There have been occasional reports of neurotoxicity with this combination; however, the two drugs appear to be well tolerated in combination by most patients. Thus, in the management of acutely severe manic patients administration of the drugs may be started simultaneously. When used in this combination, carbamazepine, like an antipsychotic agent, may help control the extreme elements of excitation and aggression prior to the onset of the inadequate response to lithium. If control is not adequate, antipsychotics may be added to this combination. Therefore, it is quite reasonable to treat the most severely manic patient with usual therapeutic dosages of several agents together, rather than to resort to extremely high dosages of lithium or antipsychotic agents and the risk of toxicity.

Some patients show a selective response to carbamazepine but not to other anticonvulsants such as valproate or phenytoin (Dilantin). In addition, some patients appear to respond to valproate when they do not respond adequately to carbamazepine alone. Thus, valproate may be considered an alternative to carbamazepine in the management of severe manic episodes that are refractory to traditional treatment.

It was previously suggested that carbamazepine need not be tapered for discontinuation, as nonepileptic patients showed no apparent risk for seizures after withdrawal. However, a recent study suggests that 12 epilepsy patients from whom carbamazepine was withdrawn rapidly (over 4 days) experienced significantly more generalized tonic-clonic seizures than those from whom carbamazepine was withdrawn more slowly (over 10 days). Manic or psychotic symptoms may reemerge rapidly however.

Schizophrenia and Schizoaffective Disorder Several double-blind studies comparing the addition of carbamazepine or placebo with ongoing treatment with antipsychotic agents suggests that the adjunctive use of carbamazepine may be effective in some patients with schizophrenia or schizoaffective disorder, particularly those with florid positive symptoms or problems with aggressive behavior.

Impulse-Control Disorders Carbamazepine has been reported to have a role in a variety of impulse-control disorders including borderline and other personality disorders.

Other Uses Several studies suggest that carbamazepine and benzodiazepines have comparable efficacy in the management of acute alcohol withdrawal syndromes. Since carbamazepine does not have the same abuse potential as benzodiazepines, it may have a role in the long-term treatment of alcoholic patients. However, further studies are required to clarify its exact role in this population. Carbamazepine has also been reported to have a role in a variety of pain syndromes including diabetic neuropathy and migraine prophylaxis. However, high-quality studies are needed to determine the effects of carbamazepine in treating chronic pain syndromes and for comparisons of antidepressants with anticonvulsants.

Contraindications Carbamazepine should be avoided in patients with a history of bone marrow suppression, narrow-angle glaucoma, hypersensitivity to the drug, or known sensitivity to any of the tricyclic compounds, such as amitriptyline (Elavil) or imipramine. It should be avoided in patients taking a monoamine oxidase inhibitor. Before administration of carbamazepine, monoamine oxidase inhibitors should be discontinued for a minimum of 14 days, or longer if the clinical situation permits. Carbamazepine should be used with caution in patients with preexisting significant cardiovascular disease because it can slow the atrioventricular system.

Safety in Overdose In determining the overall risk-benefit ratio of a medication effectively for ill patients who are at substantial risk for suicide, the relative safety of the agent in an overdose situation is an important consideration. Although uncommon, fatal carbamazepine toxicity does occur. Carbamazepine overdose is characterized by neurological symptoms such as diplopia, dysarthria, ataxia, vertigo, nystagmus, hypoflexia or hyperflexia, and coma. In addition, seizures, respiratory depression, hypothermia, oral dyskinesias, and cardiac conduction defects may occur. Plasma concentrations of carbamazepine are only moderately correlated with severity, 10,11-epoxide serum concentrations have been reported to correlate better with carbamazepine intoxication. Supportive techniques including gastric lavage (which should be undertaken up to 12 hours after ingestion) with activated charcoal (50–100 grams, then 12.5 grams an hour until the patient is symptom free) have been attempted in such cases. Early charcoal hemoperfusion before the patient becomes hypotensive and administration of flumazenil (Ranazicon) to block central-type benzodiazepine receptor effects for coma, have also been tried. The benefit from plasmapheresis remains unclear. Forced diuresis, peritoneal dialysis, and hemodialysis, however, are not recommended. Previous reports indicated that many patients have survived large overdoses. Few fatalities have occurred, and people have been reported to survive ingested doses of carbamazepine as high as 20 grams (i.e., 100 capsules).

Use During Pregnancy Carbamazepine has been reported to have teratogenic effects. First-trimester exposure may be associated with an increased risk of neural tube defects, craniofacial defects, fingernail hypoplasia, and developmental delay. Exposure to antiepileptic drugs during pregnancy is associated with increased risk of congenital anomalies in offspring. The incidence of malformations in infants of mothers with epilepsy who are treated with antiepileptic drugs is two or three times that of infants of mothers without epilepsy. The incidence of malformations associated with carbamazepine has been estimated to be approximately 7.3 percent. A recent prospective case-control study examining the risk of fetal death and anomalies from antiepileptic drugs compared the outcomes of pregnant women with epilepsy treated with antiepileptic drugs (N = 174) and their offspring with a control group of 355 healthy women and their offspring. They found that the offspring of women with epilepsy who were exposed to antiepileptic drugs had statistically significant higher rates of fetal death and anomalies than the control group. This risk of an abnormal outcome (death and anomalies) was highest with phenobarbital, followed by phenytoin and carbamazepine. The risk of neural tube defects associated with in utero antiepileptic agent exposure may be reduced by prophylactic treatment with high doses of folate—ideally before conception.

Thus, while the conservative approach to medication should be taken during pregnancy (i.e., no medications unless they are absolutely required), the initial data from the epilepsy literature speaks to the relative safety of carbamazepine, particularly in relation to lithium. Both carbamazepine and its epoxide metabolite have been reported to transfer to breast milk.

LABORATORY INTERFERENCES

Carbamazepine may interfere with the dexamethasone-suppression test and with some pregnancy tests. Carbamazepine increases the metabolism of sex hormones used in birth control preparations, so higher-dose formulations of those preparations may be required for the oral contraceptives to maintain their efficacy.

As listed in Table 31.7a-3, clinically meaningful drug interactions have occurred with concomitant medications. Most drug interactions result from carbamazepine's induction of hepatic microsomal CYP 2D6, 1A2, 3A4, and 2C9/10, which leads to an accelerated elimination of drugs normally metabolized by this system, including the barbiturates and oral contraceptives. Thus, carbamazepine may decrease the blood levels and efficacy of a variety of drugs. Most notably, carbamazepine may interact with commonly prescribed agents, such as antidepressants, antipsychotic drugs, oral contraceptives, and anticoagulants. Plasma concentrations of these coprescribed agents may decrease to a clinically significant degree, requiring dosage adjustments to compensate for the lowered plasma concentrations. Carbamazepine has been reported to decrease levels of haloperidol by 40 to 60 percent. While administering carbamazepine to refractory excited psychotic patients taking neuroleptic drugs, such as haloperidol is generally effective, occasional exacerbations may be associated with reduction of serum neuroleptic levels to the undetectable range.

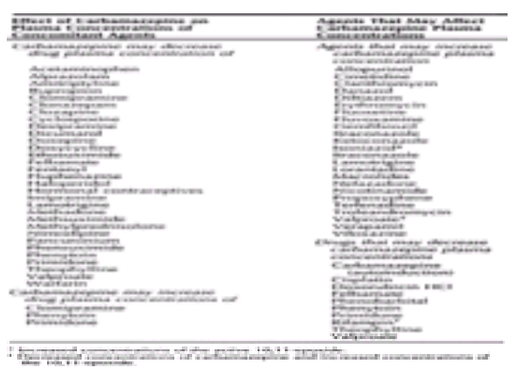

Table 31.7a-3 Carbamazepine-Drug Interactions

Several drugs have been reported to markedly increase concentrations of carbamazepine, sometimes leading to toxic symptoms in patients otherwise well maintained. CYP 3A4 inhibitors that inhibit carbamazepine metabolism can increase plasma carbamazepine concentrations. These drugs include the commonly used antibiotic erythromycin and its congeners. Epoxide produces the same effect, although this has not been reported with other monamine oxidase inhibitors. Calcium channel inhibitors can increase carbamazepine concentrations, with verapamil (Cardizem) having a greater effect than nifedipine (Procardia). The concentration of carbamazepine in blood can also be significantly (up to 40 percent) and rapidly increased with erythromycin. Symptoms of carbamazepine toxicity (confusion, sedation, and ataxia) may develop. Reduction of the carbamazepine dosage is required when administration of erythromycin is started. Other types of antibiotics are preferred, if possible. Verapamil and diltiazem in particular can significantly increase carbamazepine concentrations in blood. Similarly, concentrations of carbamazepine are increased by divalproex, selective serotonin reuptake inhibitors (especially fluoxetine [Prozac] and fluvoxamine [Luvox]), erythromycin, propoxyphene (Darvon), cimetidine (Tagamet), and isoniazid (Cortinazin). Valproate produces slight increases in the concentrations of the 10, 11-epoxide metabolite but no substantial toxicity is associated with this combination.

CYP 3A4 inducers can increase the rate of carbamazepine metabolism. Drugs that have been shown, or that would be expected, to decrease the concentration of carbamazepine in plasma include barbiturates, phenytoin, and primidone (Mysoline).

Both carbamazepine and clozapine (Clozaril) can suppress bone marrow leukocytogenesis. Therefore their combination should be avoided. In one study, neutropenia (<1500 neutrophils per µL) was more common in 14 patients given both agents than in 133 patients not given the anticonvulsant.

DOSAGE AND ADMINISTRATION

Carbamazepine is available for oral administration as chewable tablets of 100 mg, tablets of 200 mg, sustained-release tablets of 100, 200, and 400 mg, and a suspension of 100 mg per 5 mL. When prescribing carbamazepine in the treatment of manic episodes, initial dosages may be 400 to 600 mg a day in divided doses (three or four times a day), with daily increments of 200 mg a day every few days until subjective adverse effects or clinical response is reached. Until tolerance to adverse effects develops, titration many need to be slowed or even reversed. Lower initial dosages should be used in the euthymic or depressed patient. Response is often observed in a dosage range of 1000 to 1200 mg a day, although increases to 1600 mg a day may be indicated in some patients in the absence of adverse effects. Because of carbamazepine's short half-life, the drug may need to be given in divided doses three or four times a day, with the bulk of the dose administered at bedtime to take advantage of its sedating effects. Recently, an extended-release carbamazepine tablet was developed, permitting patients to take carbamazepine only twice daily. The long-acting carbamazepine would help improve compliance by avoiding the inconvenience of multiple-daily dose therapy. Patients may be switched on a mg-per-mg basis from multiple-daily-dose-form of carbamazepine to the extended-release form twice daily from one day to the next without retitrating the daily dosage. Carbamazepine should be stored in a cool dry place, not in humid places such as the bathroom, because tablets may absorb considerable moisture and lose up to one-third of their activity. The recommended laboratory monitoring of carbamazepine for adult psychiatric disorders is described in Table 31.7a-1.

SUGGESTED CROSS-REFERENCES

Epilepsy is discussed in Section 2.4, schizophrenia in Chapter 12, schizoaffective disorder in Section 13.1, mood disorders in Chapter 14, impulse control disorders in Chapter 22 and lithium in Section 38.18.

SECTION REFERENCES

Anderson GD: A mechanistic approach to antiepileptic drug interactions. Ann Pharmacother *32:* 554, 1998.

Ballenger JC, Post RM: Therapeutic effects of carbamazepine in affective illness: A preliminary report. Commun Psychopharmacol *2:* 159, 1978.

Ballenger JC, Post RM: Carbamazepine in manic-depressive illness: A new treatment. Am J Psychiatry *137:* 782, 1980.

*Bernus I, Dickinson RG, Hooper WD, Eadie MJ: The mechanism of the carbamazepine-valproate interaction in humans. Br J Clin Pharmacol *44:* 21, 1997.

Bertilsson L, Tomson T: Clinical pharmacokinetics and pharmacological effects of carbamazepine and carbamazepine 10,11-epoxide: An update. Clin Pharmacokinet *11:* 177, 1986.

Blackburn SC, Oliart AD, Garcia Rodriguez LA, Perez Gutthann S: Antiepileptics and blood dyscrasias: A cohort study. Pharmacotherapy *18:* 1277, 1998.

*Denicoff KD, Smith-Jackson EE, Disney ER, Ali SO, Leverich GS, Post RM: Comparative prophylactic efficacy of lithium, carbamazepine, and the combination in bipolar disorder. J Clin Psychiatry *58:* 470, 1997.

Fritze J, Beneke M, Lanczik M, Schneider B, Walden J: Carbamazepine as adjunct or alternative to lithium in the prophylaxis of recurrent affective disorders. Pharmacopsychiatry *27:* 181, 1994.

Greil W, Kleindienst N, Erazo N, Muller-Oerlinghausen B: Differential response to lithium and carbamazepine in the prophylaxis of bipolar disorder. J Clin Psychopharmacol *18:* 455, 1998.

*Greil W, Ludwig-Mayerhofer W, Erazo N, Schochlin C, Schmidt S, Engel RR, Czernik A, Giedke H, Muller-Oerlinghausen B, Osterheider M, Rudolf GA, Sauer H, Tegeler J, Wetterling T: Lithium versus carbamazepine in the maintenance treatment of bipolar disorders—a randomized study. J Affect Disord *43:* 151, 1997.

Hojer J, Malmlund H-O, Berg A: Clinical features in 28 consecutive cases of laboratory confirmed massive poisoning with carbamazepine alone. Clin Toxicol *31:* 449, 1993.

Hough C, Irwin RP, Gao X-M, Rogawski MA, Chuan D-M: Carbamazepine inhibition of NMDA-stimulated calcium influx in cerebellar granule cells. Neurosci Abstr *19:* 1781, 1993.

Isojärvi JIT, Pakarinen AJ, Myllylä VV: A prospective study of serum sex hormones during carbamazepine therapy. Epilepsy Res *9:* 139, 1991.

Isojärvi JIT, Pakarinen AJ, Myllylä VV: Serum lipid levels during carbamazepine medication. A prospective study. Arch Neurol *50:* 590, 1993.

Jung H, Noguez A, Mayet L, Fuentes I, Gonzalez-Esquivel DF: The distribution of 10-hydroxy carbazepine in blood compartments. Biopharm Drug Dispos *18:* 17, 1997.

Kale P, Thompson P, Provenzaro R, Higgins M: Evaluation of plasmapheresis in the treatment of acute overdose of carbamazepine. Ann Pharmacother *24:* 866, 1993.

Kaneko S, Otani K, Kondo T, Fukushima Y, Nakamura Y, Ogawa Y, Kan R, Takeda A, Nakane Y, Teranishi T: Malformation in infants of mothers with epilepsy receiving antiepileptic drugs. Neurology *42* (Suppl):68, 1992.

Keck PE Jr, McElroy SL, Nemeroff CB: Anticonvulsants in the treatment of bipolar disorder. J Neuropsychiatry Clin Neurosci *4:* 395, 1992.

Ketter TA, Post RM: Clinical pharmacology and pharmacokinetics of carbamazepine. In *Anticonvulsants in Psychiatry*, RT Joffe, JR Calabrese, editors. Marcel Dekker, New York, 1995.

*Ketter TA, Post RM, Worthington K: Principles of clinically important drug interactions with carbamazepine. Part 1. J Clin Psychopharmacol *1:* 198, 1991.

*Ketter TA, Post RM, Worthington K: Principles of clinically important drug interactions with carbamazepine. Part 2. J Clin Psychopharmacol *11:* 306, 1991.

Lenox RH, McNamara RK, Watterson JM, Watson DG: Myristoylated alanine-rich C kinase substrate (MARCKS): A molecular target for the therapeutic action of mood stabilizers in the brain? J Clin Psychiatry *57* (Suppl):23, 1996.

Malow BA, Blaxton TA, Stertz B, Theodore WH: Carbamazepine withdrawal: Effects of taper rate on seizure frequency. Neurology *43:* 2280, 1993.

*Manji HK, Chen G, Hsiao JK, Risby ED, Masana MI, Potter WZ: Regulation of signal transduction pathways by mood stabilizing agents: Implications for the delayed onset of therapeutic efficacy. J Clin Psychiatry *57* (Suppl):34, 1996.

McElroy SL, Keck PE Jr: Antiepileptic drugs. In *The American Psychiatric Press Textbook of Psychopharmacology*, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1995.

McQuay H, Carroll D, Jadad AR, Wiffen P, Moore A: Anticonvulsant drugs for management of pain: A systematic review. Br Med J *311:* 1047, 1995.

Meyer MC, Straughn AB, Mhatre RM, Shah VP, Williams RL, Lesko LJ: The relative bioavailability and in vivo-in vitro correlations for four marketed carbamazepine tablets. Pharm Res *15:* 1787, 1998.

Nau H, Kuhnz W, Egger H-J, Rating D, Helge H: Anticonvulsants during pregnancy and lactation: Transplacental, maternal and neonatal pharmacokinetics. Clin Pharmacokinet *7:* 508, 1982.

Okuma T: Effects of carbamazepine and lithium on affective disorders. Neuropsychobiology *27:* 138, 1993.

Okuma T: Use of antiepileptic drugs in schizophrenia: A review of efficacy and tolerance. CNS Drugs *1:* 269, 1994.

Pollock BG: Recent developments in drug metabolism of relevance to psychiatrists. Harvard Rev Psychiatry *2:* 204, 1994.

Post RM: Mechanism of action of carbamazepine and related anticonvulsants in affective illness. In *Psychopharmacology: A Generation of Progress*, H Meltzer, WE Bunney Jr, editors. Raven, New York, 1987.

Post RM, Weiss SRB: The neurobiology of treatment-resistant mood disorders. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Prevey ML, Delaney RC, Cramer JA, Cattanach L, Collins JF, Mattson RH. Effect of valproate on cognitive functioning. Comparison with carbamazepine. The Department of Veterans Affairs Epilepsy Cooperative Study 264 Group. Arch Neurol *53:* 1008, 1996.

Roujeau J-C, Kelly JP, Naldi L, Rzany B, Stern RS, Anderson T, Auquier A, Bastuji-Garin S, Correia O, Locati F, Mockenhaupt M, Paoletti C, Shapiro S, Shear N, Schöpf E, Kaufman DW: Medication use and the risk of Stevens-Johnson syndrome or toxic epidermal necrolysis. N Engl J Med *333:* 1600, 1995.

Schweizer E, Rickels K, Case WG, Greenblatt DJ: Carbamazepine treatment in patients discontinuing long-term benzodiazepine therapy. Arch Gen Psychiatry *48:* 448, 1991.

Stoll AL, Severus WE: Mood stabilizers: shared mechanisms of action at postsynaptic signal-transduction and kindling processes. Harv Rev Psychiatry *4:* 77, 1996.

Surks MI, DeFesi CR: Normal serum free thyroid hormone concentrations in patients treated with phenytoin or carbamazepine. A paradox resolved. JAMA *75:* 1495, 1996.

Tennis P, Stern RS: Risk of serious cutaneous disorders after initiation of use of phenytoin, carbamazepine, or sodium valproate: A record linkage study. Neurology *49:* 542, 1997.

Tohen M, Castillo J, Baldessarini RJ, Zarate C Jr, Kando JC: Blood dyscrasias with carbamazepine and valproate: A pharmacoepidemiological study of 2,228 patients at risk. Am J Psychiatry *152:* 413, 1995.

Waldmeier PC, Baumann PA, Wicki P, Feldtrauer JJ, Stierlin, Schmutz M: Similar potency of carbamazepine, oxcarbazepine and lamotrigine in inhibiting the release of glutamate and other neurotransmitters. Neurology *45:* 1907, 1995.

Waters CH, Belai Y, Gott PS, Shen P, De Giogio CM: Outcomes of pregnancy associated with antiepileptic drugs. Arch Neurol *51:* 250, 1994.

*Woolston JL: Case-study: Carbamazepine treatment of juvenile-onset bipolar disorder. J Am Acad Child Adolesc Psychiatry *38:* 335, 1999.

Yukawa E, Honda T, Ohdo S, Higuchi S, Aoyama T: Detection of carbamazepine-induced changes in valproic acid relative clearance in man by simple pharmacokinetic screening. J Pharm Pharmacol *49:* 751, 1997.

31.7 ANTICONVULSANTS

31.7B VALPROATE

SUSAN L. MCELROY, M.D., HARRISON G. POPE, ,R., M.D., AND PAUL E. KECK, ,R., M.D.

The indications for valproate (Depakene) currently recognized by the U.S. Food and Drug Administration (FDA) are (1) the treatment of the manic episodes associated with bipolar disorder; (2) sole and adjunctive therapy for complex partial seizures that occur either in isolation or in association with other types of seizures, sole and adjunctive therapy for simple and complex absence seizures, and adjunctive therapy for multiple seizure that includes absence seizures; and (3) the prophylaxis of migraine headaches. Controlled studies have shown valproate to be effective in other primary generalized seizures occurring in idiopathic epilepsies, including generalized tonic-clonic and myoclonic types, and in infantile spasms and photosensitive epilepsy. In addition, growing reports suggest that valproate may be helpful in other neurological and psychiatric conditions, including some chronic pain and movement disorders as well as depressive, schizoaffective, anxiety, eating, alcohol use, substance use, impulse control, personality, and cognitive disorders. Reports concerning the efficacy of valproate in schizophrenia, however, have not been encouraging.

HISTORY

Valproic acid, also known as valproate, (Depakene), and divalproex (Depakote), was originally developed in Europe after its antiepileptic properties were serendipitiously discovered in 1963 while it was being used as a vehicle for other drugs that were being screened for antiepileptic activity. It was first used clinically in France in 1964 and was first approved for use in the United States as an antiepileptic drug for absence seizures in 1978.

As early as 1966 in France researchers reported that valpromide (the amide prodrug of valproate) might be effective in the treatment of bipolar disorders. After that report, numerous open and controlled studies found that valproate had short-term and long-term mood-stabilizing effects in some patients with bipolar and schizoaffective disorders, with particularly robust efficacy in the manic phase of bipolar I disorder. In 1996, valproate received approval in the United States for use in acute mania. Later that same year, the drug received approval for use in complex partial seizures and migraine headache prophylaxis.

CHEMISTRY

Valproate, also known as *n*-dipropylacetic acid or 2-propylpentanoic acid (Fig. 31.7b-1), is a simple branched-chain carboxylic acid and thus is chemically different from other available antiepileptic drugs. Four oral preparations of valproate are currently marketed in the United States: valproic acid; sodium syrup; divalproex, an enteric-coated stable coordination compound composed of sodium valproate and valproic acid in a 1 to 1 molar relationship; and divalproex sprinkle capsules, which may be opened and sprinkled on food. Of note, divalproex tablets may be administered rectally. An intravenous form (Depacon) is also available for use in the United States. Valpromide, the amide precursor of valproate, is marketed in Europe.

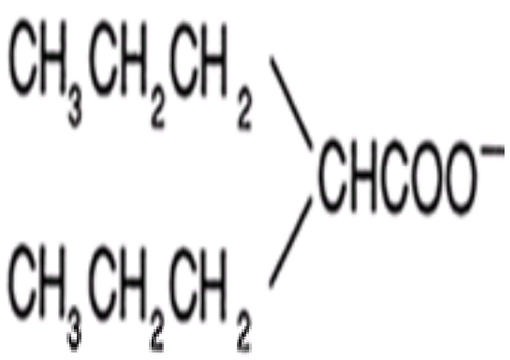

FIGURE 31.7b-1 Molecular structure of valproate.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics All valproate preparations are rapidly and completely absorbed after oral administration. Approximately 90 percent of the drug is bound to plasma proteins. However, the unbound portion of valproate increases as serum concentrations increase, apparently because the binding sites on plasma proteins become saturated as the serum concentration of the drug approaches the therapeutic range, which is generally believed to be between 50 and 125 or 150 µg/mL in both neurological and psychiatric applications. Free valproate in the blood crosses readily into the cerebrospinal fluid by passive diffusion and possibly by carrier-mediated transport.

Valproate is metabolized almost entirely by the liver, primarily by conversion to a glucuronide conjugate and by mitochondrial b-oxidation. Several of valproate's metabolites, such as 2-propyl-2-pentenoic acid, have potent anticonvulsant effects. Valproate has a half-life of 8 to 17 hours, with a shorter half-life in patients taking drugs that accelerate hepatic metabolism, including other antiepileptic drugs.

Pharmacodynamics Valproate has been hypothesized to mediate both its antiepileptic effects and its psychiatric effects in part through potentiation of central nervous system (CNS) g-aminobutyric acid (GABA) function. GABA is the major inhibitory neurotransmitter in the mammalian CNS. Valproate inhibits the catabolism of GABA, increases its release, decreases GABA turnover, increases GABA type B ($GABA_B$) receptor density, and may also enhance neuronal responsiveness to GABA. Studies have suggested that valproate-induced increased concentrations of GABA in brain and improved neuronal responsiveness to GABA are associated with seizure control. Other research, however, suggests that valproate exerts its antiepileptic effects by direct neuronal effects (i.e., reducing sodium influx and increasing potassium efflux) or by interactions with gamma hydroxybutyrate (GHB), a product of GABA metabolism that may have epileptogenic properties.

Although the role of GABA in mood disorders is not well understood, several lines of evidence have implicated GABA in mood disorders. Many drugs known to be effective in the treatment of manic episodes—including lithium (Eskalith), antipsychotics, and carbamazepine (Tegretol)—influence GABA transmission. Also, GABA transmission may modulate circadian rhythms in the suprachiasmatic nucleus of the hypothalamus, and some have suggested that valproate exerts its antimanic effects by normalizing an abnormally long endogenous circadian period. Moreover, GABA concentrations in the CNS are low in some animal models of aggression. Irritability and aggression are common manic symptoms. Also, the administration of GABA antagonists induces killing behavior in rats, whereas the administration of

various GABAergic compounds (including valproate) inhibits aggressive behavior in rats and mice.

Other effects of valproate that might contribute to its mood-stabilizing properties include decreased dopamine turnover, altered serotonin function, decreased *N*-methyl-d-aspartate (NMDA)-mediated currents, decreased aspartate release, and decreased cerebrospinal fluid (CSF) somatostatin concentrations.

Growing evidence indicates that the serum concentrations of valproate required to treat manic episodes exceed 50 µg/mL, the same concentrations believed effective for epilepsy. This concentration may represent the point at which protein-binding sites are saturated and substantial free drug becomes available.

Few studies in either neurological or psychiatric patients have routinely used valproate serum concentrations above 100 µg/mL, although a few patients respond to concentrations of 100 to 200 µg/mL after failing to respond to lower concentrations. However, adverse effects are more common with valproate concentrations above 100 µg/mL.

CLINICAL DRUG STUDIES

Acute Mania In psychiatry, valproate has been studied most extensively in the manic phase of bipolar I disorder and in schizoaffective disorder, bipolar type. To date, more than 16 uncontrolled studies and 7 controlled studies have assessed the efficacy of valproate (or valpromide) in manic patients. The first 16 uncontrolled studies reported on 663 manic patients, of whom approximately 419 (63 percent) showed a moderate-or-marked response to valproate.

Three early controlled studies (in 1985 or before) compared valproate with a placebo in a total of 14 patients; of those, 10 (71 percent) were described as markedly better after taking valproate than after a placebo. Since 1991, four larger controlled studies of valproate monotherapy have appeared. In the first study, 36 inpatients meeting the criteria for bipolar disorder, manic phase, in the revised third edition of Diagnostic and Statistical Manual of Mental Disorders (DSM-III-R), were randomized to valproate (*N* = 17) or placebo (N = 19) for 7 to 21 days. Valproate serum concentrations were maintained at 50 to 100 µg/mL by an unblinded investigator. Valproate proved significantly superior to placebo at endpoint on measures of manic symptoms, overall psychiatric symptoms, global assessment of functioning, and the need for supplemental dosages of lorazepam (Ativan). Nine (53 percent) of the 17 valproate-treated patients had more than 50 percent improvement in ratings of mania, compared with 2 (11 percent) of the 19 placebo-treated patients (Fig. 31.7b-2). Patients responding to valproate displayed a prompt response, with a sharp reduction in manic symptoms occurring 1 to 4 days after achieving valproate serum concentrations of at least 50 µg/mL (Fig. 31.7b-3).

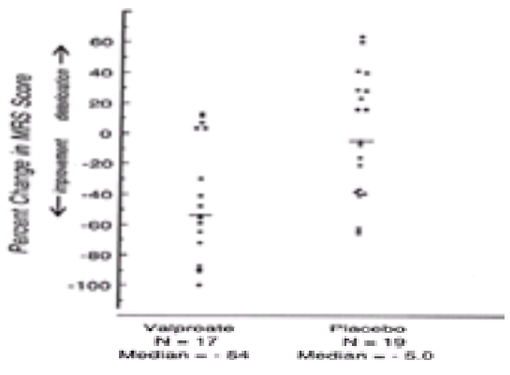

FIGURE 31.7b-2 Percentage change in Manic Rating Scale scores in 17 patients receiving valproate versus 19 patients receiving placebo. (Reprinted with permission from Pope HG Jr, McElroy SL, Keck PE Jr, Hudson JI: Valproate in the treatment of acute mania: A placebo-controlled study. Arch Gen Psychiatry *48:* 62, 1991.)

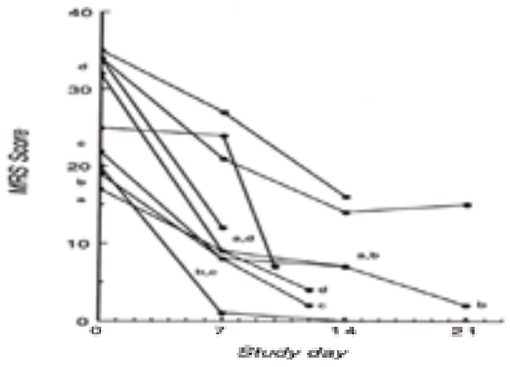

FIGURE 31.7b-3 Change in Manic Rating Scale scores by study day in nine bipolar patients who responded to valproate. (Reprinted with permission from Pope HG Jr, McElroy SL, Keck PE Jr, Hudson JI: Valproate in the treatment of acute mania: A placebo-controlled study. Arch Gen Psychiatry *48:* 62, 1991.)

In the second study, 27 acutely manic patients were randomly assigned to valproate or lithium. No group received a placebo. Twelve (92 percent) of 13 patients assigned to the lithium group were rated responders, compared with 9 (64 percent) of the 14 patients assigned to the valproate group. Although the response rate to lithium exceeded that to valproate in the study, the difference was not statistically significant (*P* = .20 by Fisher's exact test, two-tailed). Also, the response rate to lithium was unusually high in the study.

In the third study, 68 acutely manic patients meeting Research Diagnostic Criteria for manic disorder were randomized to valproate, 35 to lithium, and 73 to placebo. Premature termination due to lack of efficacy occurred in 30 percent of the patients assigned to valproate, 33 percent of patients assigned to lithium, and 51 percent of patients assigned to placebo. The difference between valproate and placebo on this index was significant (*P* = .017). The proportion of patients exhibiting at least 50 percent improvement was higher for both valproate and lithium than for placebo: 48 percent of the patients assigned to valproate, 49 percent with lithium, and 25 percent with placebo. The difference between valproate and placebo was highly significant (*P* = .004). No significant difference was observed between valproate and lithium in this study.

In the fourth study, 36 inpatients with bipolar I disorder, manic or mixed phase with psychotic features by DSM-III-R criteria were randomized to receive either valproate (20 mg/kg a day) or haloperidol (Haldol) (0.2 mg/kg a day) in single-blind fashion for 6 days. There was no placebo group. Valproate and haloperidol were equally effective in acutely reducing manic and psychotic symptoms. Ten (48 percent) of 21 patients receiving valproate and 5 (33 percent) of 15 patients receiving haloperidol were classified responders. The greatest rate of improvement for both drug regimens occurred over the first 3 days of treatment. Adverse effects were infrequent and minor for both drugs, except for extrapyramidal side effects, which were significantly more common with haloperidol.

In total, of 134 manic patients receiving valproate under blind conditions in seven studies, approximately 52 percent were rated responders. Moreover, in a recently completed multicenter, parallel group, double-blind, 3-week study from Europe, 136 patients with acute mania receiving open-label antipsychotics were randomized to treatment with valproate or placebo. Compared with placebo-treated patients, valproate-treated patients displayed a significantly greater decline in concomitant antipsychotic treatment at the end of the study.

Several preliminary findings emerge from the studies of valproate in acute mania. First, the time course of response is relatively rapid, with the onset of benefit usually evident in 1 to 2 weeks and often in several days after achieving therapeutic serum concentrations. Indeed, four recent reports have shown that valproate may be rapidly effective in acute mania when given via the oral loading strategies of 20 to 30 mg/kg a day, which produce serum concentrations of approximately 80 µg/mL and 100 to 120 µg/mL, respectively, after 1 day of treatment. In two of these studies, 50 percent of 40 acutely manic patients responded to valproate monotherapy within 5 days. Second, anecdotal reports suggest that valproate's short-term antimanic effects may be augmented by lithium, carbamazepine, typical antipsychotics, clozapine (Clozaril), and gabapentin (Neurontin). Third, several factors may be associated with a more favorable antimanic response to valproate than to lithium,

including rapid cycling (the occurrence of four or more mood episodes in 1 year) and possibly ultrarapid cycling; mania accompanied by mild, moderate, or severe depressive symptoms, including mixed or dysphoric mania; and organic or complicated mania (mania caused by, or associated with, medical or neurological illness or drugs). Recent controlled data suggest that acutely manic patients with mixed features or rapid cycling are just as likely to display an antimanic response to valproate as patients with pure mania or slow cycling. Other possible predictors of a better response to valproate than to lithium include comorbid substance abuse or panic attacks. Prior response to lithium or other antiepileptic drugs is not associated with valproate response.

More studies comparing valproate with reference agents such as lithium and carbamazepine are needed. Also needed are studies in which valproate or placebo are added to an existing medication regimen (e.g., lithium) in patients with persistent manic symptoms, and studies in prepubertal, adolescent, and elderly patients with mania. Nonetheless, valproate is now considered by many authorities to be a first-line mood stabilizer in patients of all ages with bipolar disorders (with the exception of children under 3 years of age because of hepatotoxicity).

Acute Bipolar Depression No controlled studies have assessed the efficacy of valproate in the depressed phase of bipolar I or II disorders, but uncontrolled data suggest that valproate is less effective in the short-term treatment of bipolar depression than in the short-term treatment of acute mania. In an open-label trial of valproate in 101 patients with rapid-cycling bipolar I or bipolar II disorder (43 of whom received valproate monotherapy and 58 of whom received valproate in combination with lithium, carbamazepine, antipsychotics, antidepressants, or a combination), 21 percent of patients who were depressed at the time valproate treatment was initiated displayed a short-term response to the drug, compared with 64 percent and 87 percent of the patients who were acutely manic or mixed, respectively. However, in a recent report of open-label valproate monotherapy in 6 patients with bipolar I depression, 4 were considered treatment responsive after approximately 4 weeks. At present, controlled studies comparing valproate with placebo, mood stabilizers, and standard antidepressants in acute bipolar depression are needed.

Acute Major Depressive Disorder Although no controlled studies have assessed the efficacy of valproate in acute major depressive disorder, five uncontrolled studies have provided data on the treatment of unipolar depression with valproate. In those studies, only about 58 (30 percent) of 195 patients with acute major depressive disorder receiving valproate showed a moderate-to-marked antidepressant response—a rate far below the rate of at least 60 percent in patients with acute mania. More recently, two patients with agitated major depression were described whose agitation responded when valproate was added to their antidepressant regimens.

Prophylactic Treatment More than 10 open studies and four comparison trials (two of which have not yet been published in total) have assessed valproate in the prophylactic treatment of bipolar or major depressive disorders. The open-label studies, which extended for up to 10 years, and three of the four comparison trials, which extended for 1 to 2 years, have reported significant reductions in the number, the duration, and the severity of recurrent manic and depressive episodes with prophylactic valproate treatment.

One open-label study compared 32 patients (17 with bipolar disorder, 10 with depressive disorders, 4 with schizoaffective disorder, and 1 with mixed bipolar disorder) treated for 1 year without valproate followed by 1 year with valproate. The average number of hospitalizations in that cohort decreased from 0.76 a year without valproate to 0.18 a year with valproate, and the mean duration of hospitalization decreased from 35 days to 7 days. Open-label findings on a cohort of 101 rapid-cycling bipolar patients at another center continue to suggest that valproate has marked prophylactic antimanic properties but poor-to-moderate prophylactic antidepressant properties.

In the first comparison trial, 150 patients with various bipolar (*N* = 121) or unipolar (*N* = 29) disorders were randomly assigned to open-label maintenance treatment with valpromide (*N* = 78, 1200 mg a day) or lithium (*N* = 72) for an average of 18 months. Valpromide and lithium were equally effective in reducing the frequency of recurrent manic and depressive episodes. Specifically, the number of episodes per patient per month during the 2 years preceding admission to the study was decreased by 84 percent (from 0.172 to 0.028) with valpromide treatment, and by 79 percent (from 0.163 to 0.034) with lithium treatment. Both drugs were slightly better in preventing manic episodes than depressive episodes. Valpromide was slightly better tolerated than lithium: none of the valpromide-treated patients discontinued valpromide due to intolerance, whereas six (8 percent) of the lithium-treated patients did so.

In the second comparison trial, a double-blind, placebo-controlled, study of 12 bipolar I disorder patients treated for 14 weeks to 1 year, patients receiving the combination of lithium and valproate were significantly less likely to suffer a relapse compared with those receiving lithium and placebo.

In the third comparison trial, a 2-year study of valproate versus lithium in 26 children, adolescents, and young adults (ages 11 to 21 years) with bipolar I disorder, valproate was associated with greater compliance, fewer treatment-emergent adverse effects, and a lower rate of relapse than lithium. The difference in relapse rates was attributed to better compliance with valproate.

The fourth comparison trial was a randomized, double-blind, parallel-group, multicenter study comparing the efficacy of valproate, lithium, and placebo as prophylactic therapy during a 52-week maintenance period. Patients with bipolar I disorder who met recovery criteria within 3 months of the onset of a manic episode were randomized to maintenance treatment with valproate, lithium, or placebo in a 2 to 1 to 1 ratio. Psychotropic medications were discontinued before randomization, except for open-label valproate or lithium, which were gradually tapered over the first two weeks of maintenance treatment. The main outcome measure was time to a manic episode; secondary measures were time to a depressive episode, time to either a manic or a depressive episode, and average change from baseline in scores for depression, mania, and global assessment of function. Neither the valproate- nor the lithium-treated group differed significantly from the placebo-treated group in time to a manic episode. However, valproate was superior to placebo in terms of lower rates of discontinuation for a recurrent mood episode. Also, valproate was superior to lithium regarding longer duration in the study, and less deterioration in depression and global assessment of function scores.

Several preliminary impressions emerge from these studies. First, valproate has generally appeared more effective for the prophylaxis of manic episodes than for depressive episodes (except in the fourth comparison trial in which neither valproate nor lithium separated from placebo and valproate was more effective for depressive symptoms). Second, valproate appears to be as effective as lithium in reducing recurrent affective episodes and is possibly associated with fewer adverse effects and better compliance. Third, valproate may be particularly useful in the long-term treatment of bipolar I and II disorders associated with mixed episodes, rapid cycling, or both—conditions that respond poorly to lithium prophylaxis. Fourth, valproate may be of value when used in conjunction with other agents for prophylactic treatment, such as lithium, carbamazepine, antipsychotics (including clozapine), antidepressants, gabapentin, topiramate (Topamax), and lamotrigine (Lamictal). Fifth, although there are no controlled studies of valproate in children, adolescents, or elders with bipolar disorder, valproate may be particularly indicated for young and elderly patients with acute mania (including those with mixed features, rapid cycling, cognitive or neurological abnormalities, and lithium resistance or intolerance) because of the drug's broad spectrum of efficacy in mania; its relatively benign cognitive, dermatological, thyroid, and renal adverse effect profiles; and the extensive clinical experience accumulated from its use in young and elderly patients with epilepsy.

Schizoaffective Disorder No controlled studies of valproate in schizoaffective disorder are available. Open studies suggest that in general, valproate is less effective in schizoaffective disorder than in bipolar disorders. In an open, retrospective study of 56 valproate-treated patients who all had manic symptoms, 24 (67 percent) of 36 patients with bipolar disorders responded favorably, compared with 9 (45 percent) of 20 patients with schizoaffective disorder, bipolar type. Thus, although perhaps not as effective, valproate may be a useful agent in some schizoaffective patients, particularly in those with the bipolar type of the disorder and when used adjunctively with lithium, carbamazepine, or an antipsychotic, including clozapine, risperidone (Risperdal), olanzapine (Zyprexa), quetiapine (Seroquel), and sertindole (Serlect).

Schizophrenia Valproate appears to be of little value in schizophrenia. In one analysis of 99 putative schizophrenic patients in seven studies, 27 (27 percent) reportedly showed a moderate-or-marked response. However, on closer inspection, among the patients with classic schizophrenia not accompanied by symptoms of major mood disorders or of epilepsy, the response rate to valproate was probably much lower than 27 percent. For example, two reports described 14 patients with diagnoses of schizophrenia by the Research Diagnostic Criteria (8 of them were treated on an open-label basis and 6 with a placebo control), and one report described 8 patients meeting the third edition of DSM (DSM-III) criteria for schizophrenia. Among those 22 rigorously diagnosed patients, 2 were minimally improved, 15 were unchanged, and 5 were described as showing qualitatively worsened symptoms.

Panic Disorder Three case reports, three small open-label trials, and one placebo-controlled, crossover study have all suggested that valproate may be effective in the treatment of panic disorder. However, in a double-blind, placebo-controlled, parallel-group study with 35 patients, valproate was not superior to placebo in reducing the number and duration of panic attacks (possibly because of the high rate of response to placebo).

In the controlled crossover study, valproate significantly decreased the length and the frequency of panic attacks and produced significant improvement on the

Clinical Global Impressions scale and the Hamilton Rating Scale for Anxiety in 12 patients. The effect on the actual frequency of attacks, however, was not reported.

In one open-label study, 16 patients with panic disorder were given a lactate infusion followed by a 28-day period of treatment with valproate, followed by rechallenge with lactate. Of the 14 patients completing the trial, 10 (71 percent) experienced more than 50 percent reduction in the weekly frequency of their panic attacks, and 6 (43 percent) described a complete remission. Valproate blocked reinduction of the panic attack on lactate rechallenge in 10 (83 percent) of the 12 patients who had initially displayed a panic attack with lactate infusion.

In a recent open-label study, all 12 patients with DSM-III-R panic disorder treated with valproate for 6 weeks displayed moderate-or-marked improvement. Measures of panic attacks and anxiety improved more quickly and more robustly than measures of phobic avoidance. Eleven of the 12 patients elected to remain on valproate. All 11 patients displayed sustained improvement at 6-month follow-up.

Whether the efficacy of valproate is greater in patients with panic disorder who have electroencephalogram (EEG) abnormalities than in those with normal EEGs is not known. Carbamazepine, for example, may be more effective in the former group, but not the latter. EEGs were not assessed in the open or controlled valproate studies. Further investigations of the role of EEG abnormalities and response to antiepileptic drugs are therefore needed in patients with panic disorder.

Posttraumatic Stress Disorder Two case reports of three patients and one open clinical trial with 16 patients have all reported valproate to be effective in the treatment of combat-related posttraumatic stress disorder: In two cases, valproate reduced irritability and aggression. In the third case, valproate decreased flashbacks, anxiety, and hypervigilance. In the larger series, 10 of 16 Vietnam veterans meeting DSM-III-R criteria for posttraumatic stress disorder were rated significantly improved with valproate, particularly in symptoms of insomnia and hyperarousal-hyperactivity. Avoidant-withdrawal symptoms also improved significantly in most of the patients.

Obsessive-Compulsive Disorder One case series described 8 of 10 patients with obsessive-compulsive disorder, previously intolerant to serotonin reuptake inhibitors, who were subsequently able to tolerate rechallenge with these agents after pretreatment with valproate. In that series, it was speculated that obsessive-compulsive disorder and complex partial seizures have similarities and that valproate was helpful because of its anticonvulsant effects. However, the authors have observed the same phenomenon in patients with obsessive-compulsive disorder and a comorbid bipolar disorder. In such patients, serotonin-reuptake inhibitors induce hypomanic, manic, or mixed symptoms that subsequently respond to valproate and other mood stabilizers.

Alcohol and Sedative-Hypnotic Withdrawal At least five open studies and one controlled trial suggest that valproate may exert a palliative effect in withdrawal from alcohol. The controlled study found that patients with alcohol withdrawal who received standard detoxification regimens augmented by valproate required less than the usual dosage of the standard medication and improved more rapidly than with standard agents alone. However, no formal statistical analysis of those results was presented. Another controlled study of 16 patients with alcohol withdrawal randomized to valproate versus placebo ($N = 8$) or valproate ($N = 8$), found no significant difference in outcome between the two treatments.

In one small series and two case reports, valproate was reported to ameliorate withdrawal symptoms from benzodiazepines. Two of the patients in those studies also showed improvement in concomitant symptoms of panic disorder.

Substance Abuse One case report described the successful use of valproate in the short-term treatment of cocaine dependence.

Eating Disorders One open-label study of valproate describes three women with bulimia nervosa and comorbid bipolar disorders who all were previously unresponsive or partially responsive to lithium and antipsychotics and who all had mildly abnormal EEGs with focal or generalized slowing. All three patients displayed a marked decrease in both mood and bulimic symptoms in response to valproate. One patient twice experienced a recurrence of mood and bulimic symptoms when her valproate level fell below 50 µg/mL. On both occasions, her symptoms promptly remitted again when serum valproate concentrations were raised above that threshold.

Neither case series nor controlled studies of valproate in patients with eating disorders without comorbid bipolar disorders have yet appeared.

Impulse-Control Disorders Several case reports describe the successful treatment of intermittent explosive disorder or kleptomania accompanied by bipolar disorder with valproate. Below, the authors describe a woman with severe kleptomania and mixed mania, both of which worsened with fluoxetine but responded to valproate.

> Ms. A., a 36-year-old married mother employed full time as an office manager, was hospitalized in the spring of 1995 for recurrent shoplifting and severe depression. She described a near-constant urge to steal that was impossible to resist, stealing an average of four times every day (prior to admission). She also experienced very rapid (ultradian) and severe mood and energy swings, meeting DSM-III-R criteria for kleptomania and bipolar disorder, mixed, respectively. Ms. A. described the course of both her kleptomania and mixed mania as chronic—with near-daily symptoms since the ages of 15 and 14 years, respectively.
>
> It was recommended that she begin taking valproate for treatment of her mixed mania. Upon discharge, however, she was treated intermittently with fluoxetine (Prozac). At presentation to the center in the fall of 1995, after 1 month of continuous fluoxetine treatment (20 mg a day), Ms. A. continued to meet DSM-III-R criteria for kleptomania and mixed bipolar disorder. Moreover, she reported that the frequency and severity of her stealing impulses and behavior, as well as the rapidity and severity of her mood swings, had increased since beginning fluoxetine treatment (she was stealing 7 to 8 times every day). Valproate was added to her fluoxetine regimen. Within 1 month, on valproate 1500 mg a day and fluoxetine 20 mg a day, Ms. A. reported a significant reduction in the frequency and intensity of her stealing impulses and episodes as well as improvement in her mood swings. Valproate was increased to 2000 mg a day and fluoxetine was discontinued. One month later, Ms. A. described complete resolution of stealing impulses, continued remission of stealing behavior, and a further decrease in mood swings. After 8 months of valproate treatment, Ms. A. reported continued resolution of stealing impulses, no further stealing, and mild mood swings that were manageable.

Behavioral Agitation (or Disturbance) of Dementia Two small uncontrolled studies, one retrospective chart review, and one case report have examined the efficacy of valproate for behavioral agitation in elderly patients with dementia. Of 38 demented patients in these four reports, 21 (61 percent) displayed moderate-or-marked reduction in agitation in response to valproate treatment. One study noted that physical aggression, verbal aggression, and restlessness were the types of behaviors that responded best to valproate.

Several tentative impressions emerge from these studies regarding the use of valproate in dementia-related behavioral agitation. First, valproate's antiagitation effects are often apparent within 1 to 3 weeks of treatment. Second, valproate may be effective as monotherapy and in patients whose agitation has failed to respond to (or has been worsened by) antipsychotics and benzodiazepines. Third, valproate may be effective in reducing behavioral agitation in a variety of types of dementia, including dementia of the Alzheimer's type, vascular dementia, and the combination of the two. Fourth, some patients who respond acutely continue to display reduced behavioral agitation over extended periods of time with maintenance treatment. Fifth, valproate may effectively reduce behavioral agitation in some elderly demented patients at doses and serum concentrations below those generally required for efficacy in epilepsy and mania (e.g., 125 to 750 mg a day and 15 µg to 50 µg/mL). However, some patients respond only when standard doses and serum concentrations are used. Sixth, valproate was generally well tolerated, with only 5 (13 percent) of the 38 patients reported to have adverse effects.

Borderline Personality Disorder Two open-label studies have examined valproate in borderline and other personality disorders. In the first study, 11 outpatients with borderline personality disorder according to DSM-III-R criteria and without current major depression or a history of a bipolar or a psychotic disorder were given valproate for 8 weeks after an 8-week medication-free baseline period. Of the 8 patients who completed the trial, 4 (50 percent) were considered responders and were rated as displaying “less” or “much less” overall pathology and mood symptoms. Valproate was found most helpful for irritability, anxiety, anger, rejection sensitivity, and impulsivity. To confirm these findings, the authors of this study are now conducting a double-blind, placebo-controlled, parallel-group trial of valproate in 25 patients with borderline personality disorder.

In the second trial, 10 patients meeting DSM-IV criteria for at least one personality disorder (5 with borderline type) received valproate for 8 weeks for the treatment of impulsive aggressive behavior. All patients had failed a trial of a selective serotonin reuptake inhibitor. Response to valproate was assessed with the modified Overt Aggression Scale. Six of 8 completers reported significant decreases in irritability ($P = .003$) and impulsive aggressive behavior ($P = .017$). Response to valproate was

apparent by the end of 4 weeks of treatment and continued to occur through week 8 of treatment.

Other Behavioral Dyscontrol Syndromes Isolated case reports have described the successful treatment with valproate of rage attacks associated with mania and epilepsy (*N* = 1), posttraumatic stress disorder (*N* = 2), and severe closed head injury (*N* = 1). Further research on the efficacy of valproate in behavioral dyscontrol seems warranted, since several controlled studies have found the antiepileptic drugs carbamazepine and phenytoin (Dilantin) effective in patients with those symptoms. Also, valproate was shown to suppress aggressive behavior in animals.

Affective Symptoms in Persons With Mental Retardation One study examined 209 persons with mental retardation who were referred for evaluation of behavioral symptoms. Patients were assigned to valproate if they exhibited three of the following four symptoms: irritability, sleep disturbance, aggressive or self-injurious behavior, and behavioral cycling. Of 18 patients who met enrollment criteria and were followed for 2 years on drug treatment, 14 (78 percent) were described as responding favorably to treatment. A history of epilepsy or a suspicion of seizures was strongly associated with a favorable response to valproate in this study (*P* < .005).

THERAPEUTIC INDICATIONS

Neurological Conditions Valproate is approved by the U.S. Food and Drug Administration for use in simple and complex absence seizures, in multiple seizures including absence seizures as a component, in complex partial seizures, and in the prophylactic treatment of migraine headaches. Valproate continues to be increasingly used as monotherapy for primary tonic-clonic seizures and in the treatment of myoclonic seizures. It is also used in the treatment of chronic (especially neurological) pain syndromes, where it may ameliorate the lancinating or shooting component, and in certain movement disorders, such as nonepileptic myoclonus, clozapine-induced myoclonus, hemiballismus, Sydenham's chorea, posttraumatic choreoathetosis, and dyssynergia cerebellaris myoclonica. However, valproate does not appear to be effective in tardive dyskinesia or Huntington's disease.

PSYCHIATRIC CONDITIONS

Divalproex is currently approved by the FDA for the treatment of the manic episodes associated with bipolar disorder. It is the only drug other than lithium approved for mania. In particular, valproate is widely used in the short-term treatment of the manic phase of bipolar I disorder and in the prophylactic treatment of patients with recurrent manic and depressive episodes. The drug is also useful in the short-term and prophylactic treatment of schizoaffective disorder, bipolar type, although it may not be efficacious in that disorder as frequently as it is in patients with classic bipolar disorder. Moreover, accumulating evidence suggests that valproate may be useful in many other psychiatric conditions (Table 31.7b-1).

Strong support
Short-term treatment of mania
Moderate support
Short-term treatment of the manic phase in schizoaffective disorder, bipolar type
Prophylactic treatment of bipolar disorders
Weak or preliminary support
Short-term treatment of major depressive disorder
Prophylaxis for recurrent major depressive disorder
Panic disorder
Posttraumatic stress disorder
Eating disorders, obsessive-compulsive disorder, and impulse-control disorders (when accompanied by symptoms of bipolar disorders)
Alcohol and sedative-hypnotic withdrawal
Borderline personality disorder
Behavioral agitation in dementia
Episodic behavioral dyscontrol or impulsive aggression

Table 31.7b-1 Psychiatric Indications for Valproate

Valproate may be particularly useful in patients with bipolar I or II disorder or schizoaffective disorder, bipolar type who (1) respond poorly to lithium or carbamazepine or are unable to tolerate their adverse effects; (2) have a partial response to lithium, carbamazepine, or antipsychotics and require an additional medication to achieve an adequate response; (3) have mania accompanied by mild, moderate, or severe depressive symptoms, including mixed or dysphoric mania; (4) have rapid-cycling bipolar disorder (four or more episodes of mania, hypomania, or depression in a year); (5) have neurological abnormalities or disorders (e.g., seizures, EEG abnormalities, head trauma, migraine, headaches); and (6) have comorbid substance use, anxiety, eating, impulse control, or attention-deficit/hyperactivity disorders. Valproate may also be indicated for persons who develop manic symptoms in the context of a neurological or medical disease. Because of the drug's efficacy in mixed mania, rapid cycling, and organic or complicated mania, together with the extensive experience accumulated from its use in children and elderly adults with epilepsy and its favorable side-effect profile, valproate may also have advantages over lithium in young and elderly patients with bipolar disorder.

Valproate may be of value in the short-term and prophylactic treatment of some patients with major depressive disorder, but present evidence suggests reserving it for patients who have clearly failed standard short-term and prophylactic antidepressant therapies. Valproate appears to be of little value for schizophrenia unaccompanied by a major mood syndrome.

Valproate may also be useful as monotherapy or adjunctively in patients with panic disorder; posttraumatic stress disorder; psychoactive substance use, eating, obsessive-compulsive, or impulse control disorders accompanied by symptoms of bipolar disorder; withdrawal from alcohol, sedative-hypnotic drugs, or cocaine; behavioral agitation associated with dementia; borderline personality disorder; and impulsive aggression (or behavioral discontrol) due to a variety of disorders.

PRECAUTIONS AND ADVERSE REACTIONS

The adverse effects of valproate are summarized in Table 31.7b-2.

Common
Gastrointestinal irritation
Nausea
Sedation
Tremor
Weight gain
Hair loss
Uncommon
Vomiting
Diarrhea
Ataxia
Dysarthria
Persistent elevation of hepatic transaminases
Rare
Fatal hepatotoxicity (primarily in pediatric patients)
Reversible thrombocytopenia
Platelet dysfunction
Coagulation disturbances
Edema
Hemorrhagic pancreatitis
Agranulocytosis
Encephalopathy and coma
Respiratory muscle weakness and respiratory failure

Table 31.7b-2 Adverse Effects of Valproate

Serious Adverse Reactions In very rare cases, valproate causes fatal hepatotoxicity, hemorrhagic pancreatitis, agranulocytosis, encephalopathy with coma, and skeletal muscle weakness with respiratory failure. To the authors' knowledge, no fatalities attributable to any of those disorders have been reported among more than 1000 patients described in the psychiatric literature. In the neurological literature, the great majority of valproate-associated fatalities are attributable to hepatotoxicity.

Valproate-associated fatal hepatotoxicity is idiosyncratic and not related to valproate dosage. It usually occurs within the first 3 months of valproate treatment and is

characterized by decreased alertness or lethargy, anorexia, nausea and vomiting, jaundice, hemorrhage, edema, ascites, and (in epileptic patients) increased seizures. Liver function test results, including those for hepatic transaminase levels and bilirubin levels, often but not invariably are elevated. The cause of the syndrome is unknown; however, risk factors for the syndrome have been clearly identified: young age (especially less than 2 years); the use of multiple antiepileptic drugs; and the presence of developmental delay or a metabolic disorder.

With the identification of these risk factors, the rate of fatal hepatotoxicity with valproate has decreased from 1 in 10,000 in 1978 to 1984 to 2.5 in 100,000 in 1985 to 1986 and 2.6 in 100,000 in 1987 to 1993. For patients receiving valproate as their sole antiepileptic drug during the latter period, the rate of fatal hepatotoxicity was 1.0 in 100,000. In patients over the age of 10 years receiving valproate monotherapy from 1978 to 1993, only one case of fatal hepatotoxicity was reported to Abbott Laboratories (the drug's manufacturer). In this case, however, the hepatotoxicity may have been due to other complicating factors. Thus, in psychiatric patients who are over the age of 10 years and are not receiving multiple antiepileptic drugs, the risk of fatal hepatotoxicity is extremely low.

Valproate may cause coma and death when taken in overdose, although recovery was described in a patient whose serum concentrations exceeded 2000 µg/mL. Valproate-induced coma has been reversed with naloxone (Narcan), and elevated serum valproate concentrations have been decreased by hemodialysis and hemoperfusion.

Valproate is well known to be teratogenic; when administered during the first trimester of pregnancy, it causes an increased incidence of neural tube defects such as spina bifida. One recent analysis estimated the rate to be 1 to 2 percent in women receiving valproate during pregnancy for the treatment of epilepsy. One recent study of 17 infants born to mothers receiving sodium valproate found that 9 had minor abnormalities and 5 of those had major malformations, described as the fetal valproate syndrome. The most frequent malformation in this series was congenital heart disease. Nine of the infants showed evidence of valproate withdrawal, including irritability, tone abnormalities, or seizures. The study found a tendency for the frequency of both minor and major abnormalities to be associated with the valproate dosage administered during the first trimester.

Valproate is excreted in breast milk at a level of 1 to 10 percent of serum concentrations. No data are available on the risk, if any, that it poses to nursing infants.

No evidence is available for a mutagenic effect of valproate. In long-term toxicity studies with rats, valproate produced a significant increase in the prevalence of pulmonary adenomas and fibrosarcomas. However, in extensive clinical experience with valproate, no evidence indicates a carcinogenic effect in humans.

Minor Adverse Effects The most frequent side effects reported with valproate are gastrointestinal (nausea, dyspepsia, vomiting, and diarrhea) and neurological (sedation, ataxia, dysarthria, and tremor). Gastrointestinal symptoms are most frequent at the beginning of treatment and often decrease with time. They tend to be more common with valproic acid and sodium valproate than with the enteric-coated formulation of divalproex sodium, although in a few cases the reverse is true.

In the double-blind study of valproate versus a placebo in manic patients described above, gastrointestinal discomfort or nausea was reported in 5 (25 percent) of 20 manic patients receiving valproate, with frank vomiting in one additional patient (5 percent). However, of 23 manic patients assigned to a placebo in the study, 5 (22 percent) had gastrointestinal discomfort or nausea, and 2 more (9 percent) had vomiting.

Persistent gastrointestinal adverse effects can be reduced by treatment with histamine type 2 (H_2) receptor antagonists, such as famotidine (Pepcid) and cimetidine (Tagamet), or sucralfate (Carafate), or by using divalproex sprinkle capsules in place of divalproex sodium tablets.

Neurological side effects, such as ataxia and sedation, are best controlled by dosage reduction. Tremor is an occasional problem with valproate alone and is a frequent complaint when valproate is combined with lithium or serotonin reuptake inhibitors. Tremor can often be controlled by b-adrenergic receptor antagonists, such as propranolol (Inderal), and sometimes, in the authors' experience, by gabapentin.

Other bothersome side effects of valproate are weight gain, reported in up to 44 percent of patients on long-term treatment, and hair loss, reported in 3 to 12 percent of patients. Although the hair loss is often transient, total alopecia has been reported in rare cases. Supplemental treatment with multivitamins containing zinc and selenium may minimize hair loss. The authors have had preliminary success treating valproate—induced weight gain with topiramate (a new antiepileptic associated with anorexia and weight loss). Thrombocytopenia may occur on initiation of valproate treatment but is usually transient (despite continuation of drug), always reversible upon drug discontinuation, and rarely clinically significant. Reversible platelet dysfunction, coagulopathies, and edema are uncommonly reported. Other infrequent minor adverse effects include headache, diplopia, and dizziness.

Valproate frequently produces modest elevations in hepatic transaminases. Those abnormalities have been described in 5 to 40 percent of patients, but are rarely associated with clinically significant hepatic dysfunction. The elevations usually subside promptly after discontinuation of valproate administration or dosage reduction, and often eventually subside despite the continuation of valproate treatment at the same dosage or even higher dosages.

Reports that women treated with valproate for epilepsy had higher rates of polycystic ovaries or hyperandrogenism than epileptic women not treated with valproate led to the speculation that valproate might cause polycystic ovaries. Polycystic ovaries are often part of a syndrome that includes hirsutism, menstrual disturbances, obesity, and elevated androgens. Reanalysis of these reports, however, suggest that severe obesity was responsible for the polycystic ovaries, not the valproate itself.

DRUG INTERACTIONS

The interactions of valproate with other drugs are summarized in Table 31.7b-3.

Drug	Interactions Reported With Valproate
Lithium	Increased tremor
Antipsychotics	Increased sedation; increased extrapyramidal effects; delirium and stupor (single report)
Clozapine	Increased sedation; confusional syndrome (single report)
Carbamazepine	Acute psychosis (single report); ataxia, nausea, lethargy (single report); may decrease valproate serum concentrations
Antidepressants	Amitriptyline and fluoxetine may increase valproate serum concentrations
Diazepam	Serum concentration increased by valproate
Clonazepam	Absence status (rare; reported only in patients with preexisting epilepsy)
Phenytoin	Serum concentration decreased by valproate
Phenobarbital	Serum concentration increased by valproate; increased sedation
Other CNS depressants	Increased sedation
Anticoagulants	Possible potentiation of effect

Table 31.7b-3 Interactions of Valproate With Other Drugs

Valproate and Other Psychotropic Drugs Valproate is commonly combined with other antimanic, antidepressant, antipsychotic, and antiepileptic agents in the short-term and prophylactic treatment of bipolar disorders. In general, combinations of valproate with lithium, antipsychotics, antidepressants, carbamazepine, gabapentin, lamotrigine, topiramate, and benzodiazepines are well tolerated and often appear to be more effective than valproate used alone.

The combination of valproate with lithium is especially common and rarely produces difficulty when both drugs are maintained in their respective therapeutic ranges. Since both drugs may precipitate tremor, however, their additive effect may produce an unacceptable degree of tremor, necessitating treatment with b-adrenergic receptor antagonists such as propranolol, or, in the authors' experience, gabapentin.

The combination of valproate and typical antipsychotics is also commonly used, especially in acutely manic patients. Although valproate and antipsychotics do not appear to affect the plasma concentrations of one another significantly, the combination may evoke increased sedation and has been reported to increase antipsychotic-induced extrapyramidal effects, which may be treated by antipsychotic dosage reduction or by the addition of an antiparkinsonian agent. One report

described delirium and stupor from the combination of valproate and antipsychotics, accompanied by diffuse EEG slowing down to the delta range. Thus, the mental status of patients receiving the combination should be monitored, and an EEG should be obtained if significant cognitive changes appear.

Clozapine may also be combined with valproate. In general, this combination is well tolerated, with sedation the most frequently reported adverse reaction. There is one report of the combination producing a confusional state with sedation, functional impairment, and slurred speech. In view of the increased risk of seizures with high dosages of clozapine, however, some authorities have suggested deliberately adding valproate when high dosages of clozapine are used to profit from valproate's antiepileptic properties. Increasing clinical experience suggests that valproate may also be safely combined with other atypical antipsychotics, including olanzapine, risperidone, quetiapine, and sertindole.

The combination of valproate and carbamazepine is occasionally administered to patients with bipolar disorders. Two reports described possible adverse effects from this combination. The first case involved acute psychosis in a patient with epilepsy in whom carbamazepine was added to a stable regimen of valproate. The authors speculated that the phenomenon was attributable to the fact that valproate increases serum concentrations of a metabolite of carbamazepine, carbamazepine-10,11-epoxide. Another report described ataxia, nausea, and daytime lethargy in a manic patient who received subtherapeutic dosages of valproate in conjunction with previously well tolerated levels of carbamazepine.

In a review of 26 consecutive patients with bipolar disorders treated with valproate plus carbamazepine, none of the patients had significant toxicity from the combination, and several appeared to respond better to the combination than to a single anticonvulsant. Also, open studies examining the efficacy and the tolerability of valproate plus carbamazepine in a substantial number of patients with refractory epilepsy indicate that the combination may be effective and is often well tolerated.

However, carbamazepine may decrease serum concentrations of valproate (since carbamazepine increases hepatic microsomal enzyme activity), valproate may increase the free carbamazepine fraction (by displacing carbamazepine from plasma proteins), or valproate may increase carbamazepine concentrations (by inhibiting its metabolism), depending on which interaction predominates.

In the authors' experience, the combination of valproate with the new antiepileptic drugs gabapentin, topiramate, and lamotrigine is generally well tolerated. However, because valproate potently inhibits the metabolism of lamotrigine, lower lamotrigine doses must be used when the two drugs are coadministered. Also, the risk of rash with lamotrigine is slightly increased when it is combined with valproate.

Antidepressants—including tricyclic drugs, monoamine oxidase inhibitors, selective serotonin uptake inhibitors, and other agents—are frequently combined with valproate, usually without difficulty. However, the tricyclic amitriptyline (Elavil) has been reported to increase valproate serum concentrations, possibly by increasing the binding of valproate to tissue. Conversely, valproate has been reported to increase the levels of tricyclic drugs, presumably by inhibiting their metabolism. Addition of fluoxetine to valproate has also been reported to increase serum valproate concentrations significantly.

When combined with benzodiazepines or carbamazepine, valproate may compete for protein-binding sites. Valproate has been reported to increase serum concentrations of unbound diazepam (Valium) and carbamazepine for that reason. However, lorazepam is apparently not displaced in that manner. The combination of valproate and clonazepam (Klonopin) has been reported to produce absence status. However, that effect appears to be rare and is presumably even more rare in psychiatric patients with no history of epilepsy.

Valproate With Other Drugs Unlike many other antiepileptic drugs, valproate does not potentiate hepatic metabolism and hence does not decrease plasma concentrations of other concomitant drugs. Rather, valproate tends to inhibit the clearance of other drugs metabolized in the liver. However, valproate has been reported to decrease phenytoin concentrations, leading in some cases to breakthrough seizures, possibly as a result of competition for protein-binding sites.

Conversely, valproate may increase the concentration of phenobarbital (Luminal) and probably other barbiturates by impairing nonrenal clearance. Even in the absence of concentration changes, the combination may result in increased sedation. Similarly, valproate may potentiate the effects of other CNS depressants, such as alcohol.

Because valproate is highly protein bound, serum valproate free fraction concentrations can be increased and valproate toxicity precipitated by coadministration of other highly protein-bound drugs (e.g., aspirin, carbamazepine, diazepam) that can displace valproate from its protein-binding sites. Serial monitoring of valproate concentrations, concentrations of concomitant drugs, and sometimes the free fractions of those drugs is particularly important when such combinations are used. However, since some of the interactions, especially CNS depression, have been described even with apparently normal concentrations of both valproate and other drugs, monitoring the patient's clinical status is also essential.

Since valproate inhibits the secondary phase of platelet aggregation, it should be used cautiously in combination with other drugs that affect coagulation, such as warfarin (Coumadin) and aspirin.

LABORATORY INTERFERENCES

Valproate rarely interferes with the results of laboratory tests. However, the drug has been reported to cause overestimation of serum-free fatty acids by about 40 percent. Serum triglyceride levels are not affected. Also, valproate metabolites may produce a false-positive result in a urinary test for ketones.

DOSAGE AND ADMINISTRATION

The pretreatment evaluation for valproate should include a medical evaluation (with focus on hepatic, hematological, and pancreatic function) and baseline hepatic and hematological laboratory tests. The latter should include a complete blood cell count (CBC) (with differential and platelet count) and serum concentrations of lactic dehydrogenase (LDH), serum glutamic oxaloacetic transaminase (SGOT), serum glutamic pyruvic transaminase (SGPT), bilirubin, and alkaline phosphatase. Other baseline measures may include bleeding time, prothrombin time (PT), and partial thromboplastin time (PTT) for patients with bleeding abnormalities; g-glutamyltransferase (GGT), PT, PTT, and plasma protein level in patients with liver disease; renal and thyroid function tests in patients who may have renal or thyroid dysfunction; serum amylase in patients with pancreatic disease; and a pregnancy test in females who might be pregnant. Before valproate treatment is begun (or, in some instances, after some patient response), patients should be warned of the risk of hepatic, hematological, and pancreatic reactions; told about the signs and symptoms of these reactions; and instructed to report them should they occur. Patients should also be cautioned about valproate's adverse effects of appetite stimulation, weight gain, and hair loss and should be considered for prophylactic zinc and selenium supplementation to reduce any hair loss experienced. Postpubertal girls and women must be warned of the drug's teratogenic effects and should be considered for prophylactic folate supplementation.

In acutely manic adolescents and adults, valproate treatment may be begun via the oral-loading strategy of 20 to 30 mg/kg a day so that therapeutic concentrations can be achieved on the first day of treatment, often with rapid onset of response and minimal adverse effects. However, in patients who are euthymic, hypomanic, depressed, very young, or elderly, valproate treatment should be initiated in lower, divided doses (e.g., 250 to 750 mg a day in adolescents and adults, and 125 to 250 mg a day in children and elderly adults) to minimize gastrointestinal and neurological toxicity. Valproate loading has not been tested in children with mania (although it has been used in children with epilepsy), and it should be avoided in elderly patients because of the risk of neurotoxicity. Patients who cannot swallow pills, particularly children and elderly adults, can use divalproex sprinkle capsules, which can be pulled apart and sprinkled on food. The valproate dose is then titrated upward (e.g., by 125 to 750 mg every several days) according to response and adverse effects, generally to a serum concentration between 50 and 125 or 150 µg/mL. Occasional patients may respond best to valproate concentrations from 150 to 200 µg/mL (although adverse effects are more frequent with concentrations above 100 µg/mL). Once patients are stabilized, valproate dosage regimens can often be simplified to one daily dose, usually taken at night, to enhance convenience and compliance.

Most authorities recommend that hepatic and hematological parameters be regularly monitored in bipolar patients receiving valproate (i.e., weekly to every several months during the initiation of treatment, and, once stable, every 6 to 24 months while the patient is taking the drug). It has been suggested that liver function tests be monitored more frequently in children younger than 10 years.

Careful clinical monitoring is probably superior to routine laboratory screening in detecting valproate-induced life-threatening hepatic, hematological, or pancreatic reactions. Indeed, the physician must be aware that routine monitoring of hematology and blood chemistry levels does not necessarily predict the development of

severe hepatotoxicity, blood dyscrasias, or pancreatitis. Thus, clinicians must instruct patients to notify them if signs or symptoms of those effects are experienced.

If mild hepatic transaminase elevation is noted, valproate treatment may be continued at the same dosage or a lower dosage with periodic monitoring of liver function tests until the enzymes return to normal levels. If transaminase increases are pronounced (more than three times the upper limit of the normal range), valproate should be reduced in dosage or discontinued. If discontinued, rechallenge with valproate may be appropriate once the abnormalities have returned to normal. If abdominal pain occurs, a serum amylase level should be obtained to rule out pancreatitis. Guidelines for these laboratory tests are summarized in Table 31.7b-4.

Prior to treatment
Standard chemistry screen with special attention to liver function tests
Complete blood count, including white cell and platelet count
During treatment
Liver function tests monthly for 3 months, then every 3 to 24 months if no abnormalities are found
Complete blood work with platelet count monthly for 3 months, then every 3 to 24 months if findings are normal
If liver function test results become abnormal
Mild transaminase elevation (less than three times normal): monitoring every 1 to 2 weeks; if stable and patient is responding to valproate, results are monitored monthly
Pronounced transaminase elevation (more than three times normal): dosage reduction or discontinuation of valproate; increase dose or rechallenge if transaminases normalize and if the patient is a valproate responder

Table 31.7b-4 Recommended Laboratory Tests During Valproate Therapy

Valproate treatment is commonly initiated in patients already receiving lithium or antipsychotics. In such cases, valproate may usually be added to an existing regimen without difficulty, although the clinician should monitor the patient for drug interactions. If the patient responds well to the addition of valproate, the other medications can be systematically subtracted from the regimen with the hope that the patient can be successfully maintained on valproate alone.

Table 31.7b-5 summarizes the preparations of valproate currently marketed in the United States.

Generic Name	Trade Name, Form (doses)	Time to Peak
Valproate sodium injection	Depacon Injection (100 mg valproic acid/ml)	1 hr
Valproic acid	Depakene, capsules (250 mg)	1–2 hr
Sodium valproate	Depakene, syrup (250 mg/5 mL)	1–2 hr
Divalproex sodium	Depakote, delayed-release tablets (125, 250, 500 mg)	3–8 hr
Divalproex sodium coated particles in capsules	Depakote, sprinkle capsules (125 mg)	Compared with divalproex tablets, divalproex sprinkle has earlier onset and slower absorption, with slightly lower peak plasma concentration

Table 31.7b-5 Valproate Preparations Available in the United States

SUGGESTED CROSS-REFERENCES

The neuropsychiatric aspects of epilepsy are discussed in Section 2.4. Substance-related disorders are discussed in Chapter 11, schizophrenia in Chapter 12, schizoaffective disorder in Section 13.1, mood disorders in Chapter 14, anxiety disorders in Chapter 17, disruptive behavior in children in Chapter 40, and eating disorders in Chapter 20. Other biological therapies are discussed in other sections of Chapter 31.

SECTION REFERENCES

*Bowden CL, Brugger AM, Swann AC, Calabrese JR, Janicak PG, Petty F, Dilsaver SC, Davis JM, Rush AJ, Small JG: Efficacy of divalproex vs. lithium and placebo in the treatment of mania. Depakote Mania Study Group. JAMA *271:* 918, 1994.

Brady KT, Sonne S, Anton R, Ballenger JC. Valproate in the treatment of acute bipolar affective episodes complicated by substance abuse: A pilot study. J Clin Psychiatry *56:* 118, 1995.

Bryant AE, Dreifuss FE: Valproic acid hepatic fatalities. III. U.S. experience since 1986. Neurology *46:* 465, 1996.

Calabrese JR, Rapport DJ, Kimmel SE, Reece B, Woyshville MJ: Rapid cycling bipolar disorder and its treatment with valproate. Can J Psychiatry *38:* 557, 1993.

Dreifuss FE: Valproic acid. Toxicity. In *Antiepileptic Drugs*, ed 4, RH Levy, RH Mattson, BS Meldrum, editors. Raven, New York, 1995.

Eberle AJ: Valproate and polycystic ovaries. J Am Acad Child Adolesc Psychiatry *37:* 1009, 1998.

Fesler FA: Valproate and combat-related post-traumatic stress disorder. J Clin Psychiatry *52:* 361, 1991.

Freeman TW, Clothier JL, Pazzaglia P, Lesem MD, Swann AC: A double-blind comparison of valproate and lithium in the treatment of acute mania. Am J Psychiatry *149:* 108, 1992.

Freeman MP, Stoll AL: Mood stabilizer combinations: A review of safety and efficacy. Am J Psychiatry *155:* 12, 1998.

Geracioti TD: Valproic acid treatment of episodic explosiveness related to brain injury [Letter]. J Clin Psychiatry *55:* 416, 1994.

Hollander E, Grossman R, Stein DJ, Kwon J: Borderline personality disorder and impulsive-aggression: The role for divalproex sodium treatment. Psychiatric Ann *26* (Suppl):464, 1996.

Hudson JI, Pope HG Jr: The role of anticonvulsants in the treatment of bulimia. In *Use of Anticonvulsants in Psychiatry: Recent Advances*, SL McElroy, HG Pope Jr, editors. Oxford Health Care, Clifton, NJ, 1988.

Kastner T, Finesmith R, Walsh K: Long-term administration of valproic acid in the treatment of affective symptoms in people with mental retardation. J Clin Psychopharmacol *13:* 448, 1993.

Kavoussi RJ, Coccaro EF: Divalproex sodium for impulsive aggressive behavior in patients with personality disorder. J Clin Psychiatry *59:* 676, 1998.

Keck PE Jr, McElroy SL: Drugs for treatment of bipolar disorder: Antiepileptic drugs. In *The American Psychiatric Association Textbook of Psychopharmacology*, ed 2, AF Shatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1998.

Keck PE Jr, McElroy SL, Thienhaus OJ, Faedda GL: Antiepileptic drugs in the treatment of withdrawal and detoxification states. In *Anticonvulsants in Psychiatry*, K Modigh, OH Robak, P Vestergaard, editors. Wrightson Biomedical Publishing, Petersfield, UK, 1994.

Keck PE Jr, Taylor VE, Tugrul KC, McElroy SL, Bennett JA: Valproate treatment of panic disorder and lactate-induced panic attacks. Biol Psychiatry *33:* 542, 1993.

Kmetz GF, McElroy SL, Collins DJ: Response of kleptomania and mixed mania to valproate [Letter]. Am J Psychiatry *154:* 580, 1997.

Kusumakar V: Prophylaxis of early onset bipolar I disorder: A two-year follow-up study of lithium versus divalproex sodium [Abstract]. In *Child and Adolescent Psychopharmacology News*, S Kutcher,

editor. Guilford, New York, 1996.

Lambert PA, Carraz G, Borselli S, Carrel MS: Action neuropsychotrope d'un nouvel anti-épileptique: Le deépamide. Ann Med Psychol *1:* 707, 1966.

Lambert PA, Venaud G: Comparative study of valpromide versus lithium as prophylactic treatment in affective disorders. Nerv J Psychiatr *7:* 1, 1995.

*Lapierre O, Dubreucq J-L, Beauchemin M-A, Vinet B: Valproic acid intoxication in a patient with bipolar disorder and chronic uremia. Can J Psychiatry *44:* 188, 1999.

Levy RH, Matson RH, Meldrum BS, editors: *Antiepileptic Drugs*, ed 4. Raven Press, New York, 1995.

McElroy SL, Keck PE Jr, Stanton SP, Tugrul KC, Bennett JA, Strakowski SM: A randomized comparison of divalproex oral loading versus haloperidol in the initial treatment of acute psychotic mania. J Clin Psychiatry *57:* 142, 1996.

*McElroy SL, Weller E: Psychopharmacologic treatment of bipolar disorder across the life span. In *American Psychiatric Association Annual Review of Psychiatry*, vol 16, LJ Dickstein, MB Riba, JM Oldham, editors. American Psychiatric Press, Washington, DC, 1997.

McElroy SL, Soutullo CA, Beckman DA, Taylor P Jr, Keck PE Jr: DSM-IV intermittent explosive disorder: Report of 27 cases. J Clin Psychiatry *59:* 203, 1998.

Müller-Oerlinghausen B, Retzow A: Valproate as adjunct to neuroleptic medication in the treatment of acute episodes of mania. Pharmacopsychiatry *30:* 202, 1997.

*Pope HG Jr, McElroy SL, Keck PE Jr, Hudson JI: Valproate in the treatment of acute mania: A placebo-controlled study. Arch Gen Psychiatry *48:* 62, 1991.

Sachs G: Is divalproex sodium an appropriate initial treatment for bipolar depression? Psychiatr Ann *26* (Suppl):454, 1996.

Schatzberg AF, De Battista C, DeGolia S: Valproate in the treatment of agitation associated with depression. Psychiatr Ann *26* (Suppl):471, 1996.

Settle EC Jr: Valproic-acid associated encephalopathy with coma [Letter]. Am J Psychiatry *152:* 1236, 1995.

*Shulman KL, Tohen M, Kutcher SP, editors: *Mood Disorders Across the Life Span*. Wiley, New York, 1996.

Solomon DA, Ryan CE, Keitner GI, Miller IW, Shea T, Kazim A, Keller MB: A pilot study of lithium carbonate plus divalproex sodium for the continuation and maintenance treatment of patients with bipolar I disorder. J Clin Psychiatry *58:* 95, 1997.

Stoll AL, Banov M, Kolbrener M, Mayer PV: Neurologic factors predict a favorable valproate response in bipolar and schizoaffective disorders. J Clin Psychopharmacol *14:* 311, 1994.

*Swann AC, Bowden CL, Morris D, Calabrese JR, Petty F, Small J, Dilsaver SC, Davis JM: Depression during mania: Treatment response to lithium or divalproex. Arch Gen Psychiatry *54:* 37, 1997.

*Tohen M, Grundy S: Mangement of acute mania. J Clin Psychiatry *60:* 31, 1999.

Zajecka J: Panic disorder and posttraumatic stress disorder. Psychiatr Ann *26* (Suppl):480, 1996.

31.7 ANTICONVULSANTS

31.7C OTHER ANTICONVULSANTS

NORMAN SUSSMAN, M.D.

Three of the newer antiepileptic drugs—gabapentin (Neurontin), lamotrigine (Lamictal), and topiramate (Topamax)—appear to have a therapeutic spectrum that extends to conditions seen in psychiatric practice. Like valproate (Depakene) and carbamazepine (Tegretol), they represent possible alternatives or adjuncts for the treatment of bipolar disorders, anxiety disorders, agitation, and substance abuse. Support for use of the newer antiepileptic drugs comes primarily from clinical observation, retrospective reviews, and prospective open-label trials. The quality and amount of available data are better for gabapentin and lamotrigine than topiramate, owing to their longer availability in clinical practice. There are few well-controlled studies.

The newer antiepileptic agents are structurally diverse (Fig. 31.7c-1) and have multiple central nervous system effects. They differ in metabolism, drug interactions, and adverse effects. The clinical significance of the neurochemical mechanisms associated with these drugs is not fully understood. None of the drugs has an identical combination of neurochemical actions. Each agent has one or more of the following cellular effects: (1) enhancement of inhibitory g-aminobutyric acid (GABA)-mediated processes, (2) reduction of processes mediated by excitatory amino acids such as glutamate, and (3) antagonistic effects on voltage-gated ion channels, such as calcium channels and sodium channels. In addition to these cellular effects, there may be other mechanisms, such as changes in monamines, adenosine, or *N*-methyl-D-aspartate (NMDA) responses. Table 31.7c-1 compares mechanism-related actions of these anticonvulsants.

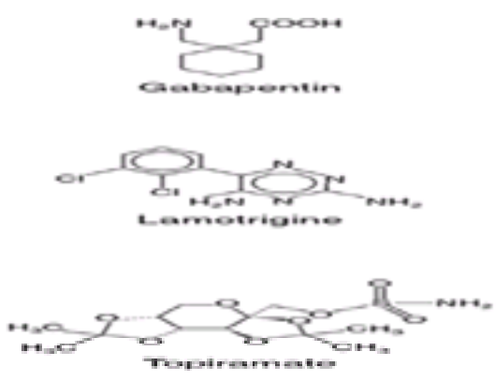

FIGURE 31.7c-1 Molecular structure of other anticonvulsants used in psychiatry.

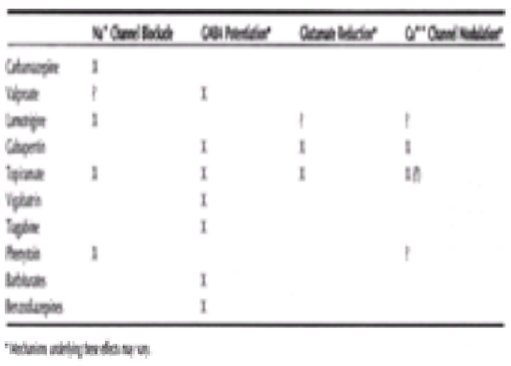

Table 31.7c-1 Comparison of Mechanism-Related Actions of Anticonvulsants

Nearly all data pertaining to the incidence of adverse events comes from experience obtained studying new anticonvulsants in combination with other antiepileptic drugs, since monotherapy is rarely attempted at the outset of epilepsy clinical trials. When used alone (and without pharmacokinetic and pharmacodynamic interactions), the type and incidence of adverse effects may vary.

GABAPENTIN

Gabapentin, approved for use by the Food and Drug Administration (FDA) in December 1993, is indicated for use as adjunctive treatment for partial seizures (both with and without secondary generalization) in patients older than 12 years of age. Gabapentin produces improvement in mood and quality of life, even when seizure control is not improved. This observed sense of well-being associated with gabapentin originally called attention to the potential value of gabapentin in the management of common psychiatric disorders.

Pharmacological Actions

Pharmacokinetics Gabapentin is poorly absorbed in the colon, with about 60 percent of a 600-mg dose being absorbed. Along with other amino acids, gabapentin is a substrate of the L-amino acid transport system found in the small intestine and on neurons and astrocytes. Thus gabapentin can easily cross cell membranes such as the blood-brain barrier and extracellular space (ECS) to glia and neuronal cytoplasm. However, this system becomes saturated at higher doses of gabapentin, which prevents a proportional increase in the amount of drug absorbed. Food does not impair the absorption of gabapentin; in fact high-protein meals significantly increase its maximal peak concentration. The time to maximum plasma concentration after ingestion is 2 to 4 hours.

Gabapentin is water soluble and can be given intravenously. It has a volume of distribution of 0.7 L/kg. Gabapentin has a half-life of 5 to 10 hours. It is not metabolized and does not have hepatically mediated drug interactions. The absorbed drug is eliminated unchanged in the urine, mainly by glomerular filtration.

Pharmacodynamics Gabapentin is an amino acid that is structurally similar to GABA. It was synthesized as a GABA analogue in the hope that it would act as a GABA receptor agonist. While gabapentin increases the synthesis and concentration of GABA, thus enhancing GABAergic activity, it does not affect GABA receptors but does cause dose-related elevations in brain GABA concentrations.

Gabapentin has been reported to alter the metabolism or concentrations of glutamate and glutamine and to bind with high affinity to a novel binding site—the

gabapentin receptor—that is associated with voltage-sensitive calcium channels. Other reported actions of gabapentin include blockade of sodium channels, interaction with an auxiliary subunit of voltage-gated Ca^{2+} channels, and increases in plasma serotonin concentrations.

The clinical effects of gabapentin are more pronounced at higher doses, but there are no established plasma concentrations that correlate with clinical response.

Therapeutic Indications

Bipolar Disorder Case reports and uncontrolled trials suggest that gabapentin facilitates stabilization of mood cycling and helps control manic episodes. In almost all reports, gabapentin is used adjunctively. These reports involve patients with different bipolar disorders (bipolar I disorder, bipolar II disorder, cyclothymic disorder, and bipolar disorder not otherwise specified) who have failed to achieve adequate control with lithium (Eskolith, Lithobid), valproate, or carbamazepine. While there are reports that gabapentin may treat the depressive phase of bipolar disorder with a lower liability for induction of mania or mood cycling than with an antidepressant, apparent mania or cycling has been reported after initiation of gabapentin treatment. Some studies suggest that gabapentin may be more useful in patients with bipolar II disorder than in those with bipolar I disorder.

In contrast to an unusually large amount of highly positive open-label data and many spontaneous case reports, no placebo-controlled trials confirm the efficacy of gabapentin for bipolar disorders as either monotherapy or adjunctive therapy. It has been suggested that the observed benefits of gabapentin therapy in bipolar patients may differ qualitatively from those associated with conventional mood-stabilizing agents, perhaps reflecting secondary anxiolytic or antiagitation properties.

Other Disorders In addition to bipolar disorders, gabapentin appears to have multiple uses in psychiatry and neurology. Patients with refractory behavioral dyscontrol have been reported to have decreases in the frequency and intensity of explosive outbursts when treated with gabapentin. Gabapentin has also been used to treat behavioral dyscontrol in children, in patients with dementia, and in patients following head injury. Animal models of anxiolytic activity strongly suggest that gabapentin has antianxiety activity. Anecdotal reports led to placebo-controlled double-blind studies that found gabapentin effective in patients with social phobia and in those with significant panic symptoms. Animal studies consistently find that gabapentin relieves induced causalgia and prevents hyperalgesia. Gabapentin has become widely used by pain specialists for the management of neuropathic pain, postherpetic neuralgia, thalamic pain, and erythromelalgia. It has been reported to relieve zoster pain, diabetic neuropathy, and intractable chronic regional pain syndrome. Gabapentin is widely used to treat reflex sympathetic dystrophy. A double-blind, randomized, placebo-controlled study that examined the effectiveness of gabapentin in the treatment of postherpetic neuralgia found that after 8 weeks in the study, the average daily pain score was significantly lower with gabapentin than with placebo. Possible effectiveness of gabapentin augmentation in obsessive-compulsive disorder patients with incomplete response to fluoxetine (Prozac), monotherapy has been reported. A mean dosage of about 2500 mg a day of gabapentin was used. Gabapentin is mildly sedating and normalizes sleep. It can be given at bedtime as an alternative to benzodiazepine agonists or other hypnotic drugs. Withdrawal symptoms and craving that accompany discontinuation of benzodiazepines, alcohol, and cocaine may be helped by gabapentin. Abrupt discontinuation of gabapentin does not cause a withdrawal syndrome.

Precautions and Adverse Reactions The most frequent side effects of gabapentin are sedation, dizziness, and ataxia, which tend to be mild and transient. Lower-extremity edema has been noted. Because gabapentin is almost exclusively eliminated through the kidneys, patients with renal impairment should be monitored closely. There is no serious toxicity with gabapentin overdose.

Drug Interactions Gabapentin does not interact with hepatic enzymes and neither inhibits nor induces them. The antacid aluminum hydroxide-magnesium hydroxide (Maalox) causes about a 25 percent reduction in absorption of gabapentin. Drugs that cause sedation may cause increased sedation when combined with gabapentin.

Laboratory Interferences Gabapentin does not interfere with any laboratory tests.

Dosage and Administration The optimal dosage of gabapentin in the management of psychiatric disorders has not been established. Both case reports and studies show extreme variability in patient response to different dosages of gabapentin. In published reports, the effective daily dosage ranges from 600 to 6000 mg. The single-dose threshold at which gabapentin may begin to interfere with its own absorption is 1200 mg; thus dosages of 3600 mg or more a day should be divided. Because this effect is not always noted and some patients derive further clinical benefit from doses as high as 6000 mg a day, when higher dosages are given, determining gabapentin concentrations can be useful.

LAMOTRIGINE

Lamotrigine, a phenyltriazine derivative, was approved in the United States in 1994. It is indicated for adjunctive treatment of partial seizures both with and without secondary generalization seizures in adults and adolescents. Both anecdotal reports and controlled trials suggest that lamotrigine may possess activity in treating both mania and depression, suggesting that it may be a truly bimodal (i.e., effective in both depression and mania) mood stabilizer.

Pharmacological Actions

Pharmacokinetics Lamotrigine is rapidly and completely absorbed. There is virtually no first-pass effect. The mean elimination half-life of lamotrigine used alone is about 30 hours, but there may be considerable individual variation—from approximately 15 to 70 hours. Maximum concentrations are achieved 2 to 4 hours after ingestion. The volume of distribution is 1.1 L/kg. It is weakly bound (55 percent) to plasma protein, and there are no clinically significant effects on hepatic drug-metabolizing enzymes. Lamotrigine is metabolized mainly by glucuronic conjugation and is eliminated almost entirely in the urine.

Pharmacodynamics Lamotrigine acts at voltage-sensitive sodium channels. Reportedly, it indirectly inhibits the presynaptic release of glutamate and aspartate and attenuates calcium entry via an effect on voltage-sensitive calcium channels. It may also alter calcium concentrations through non-voltage-dependent calcium-sensing cation channels.

Clinical Drug Studies

Bipolar Disorders Anecdotal reports, case series, prospective case studies, and controlled trials suggest utility of lamotrigine in treating bipolar disorder in treatment-refractory patients. Bipolar patients with mania, depression, mixed states, and rapid cycling have experienced significant improvement with maintenance of effect. In some cases, patients have been tapered from all other psychiatric medications.

A 7-week, multicenter, randomized, double-blind, placebo-controlled trial involving nearly 200 patients with a diagnosis of bipolar I disorder, current episode depressed, found that patients receiving lamotrigine demonstrated significant improvement in depression by week 3. The dosing schedule for lamotrigine during this trial is given in Table 31.7c-2. Significant antidepressant effects were found with lamotrigine 50 mg a day and 200 mg a day, with a greater response at the higher dose. Seven patients receiving lamotrigine (5.4 percent) and three patients receiving placebo (4.6 percent) developed manic, hypomanic, or mixed episodes.

Group	Weeks 1–2	Week 3	Week 4	Weeks 5–7
50 mg	25 mg q.d.	25 mg b.i.d.	25 mg b.i.d.	25 mg b.i.d.
200 mg	25 mg q.d.	25 mg b.i.d.	50 mg b.i.d.	100 mg b.i.d.

Table 31.7c-2 Lamotrigine: Dosing for Double-Blind Monotherapy Trial

A double-blind, placebo-controlled trial that compared gabapentin, lamotrigine, or placebo monotherapy for patients with treatment-refractory mood disorders found a statistically significant difference between lamotrigine and placebo for patients experiencing a positive response, improvement was evident by the third week of treatment for patients receiving lamotrigine and by the first week of treatment for patients receiving gabapentin. Patients responding to gabapentin experienced an apparent loss in antidepressant effect by week 5. The response rate by diagnostic subtypes (bipolar I, bipolar II, and unipolar) were, respectively, lamotrigine, 73, 47, and 29 percent; gabapentin, 42, 29, and 0 percent.

A 12-month prospective, open-label, multicenter trial evaluated the use of lamotrigine as adjunctive or monotherapy in 75 patients with bipolar disorder. All patients were either nonresponsive or intolerant to ongoing pharmacotherapy. Lamotrigine dosing was based on whether patients were receiving other drugs during the study, as summarized in Table 31.7c-3. Marked improvement was noted in 45 percent of those who entered the study in the depressed phase; 60 percent of patients who entered the study in hypomanic, manic, or mixed states, and 50 percent of patients who met criteria for rapid cycling. Three patients experienced exacerbation of mania and were hospitalized. Four patients switched from depression to mania and were hospitalized.

Treatment	Weeks 1–2	Weeks 3–4	Weeks 4–5
Lamotrigine monotherapy	25	50	100–200 (500 maximum)
Lamotrigine plus carbamazepine	50	100	200–500 (700 maximum)
Lamotrigine plus valproate	25 every other day	25	50–200 (200 maximum)

Data from 12-month prospective, open-label, multicenter trial as add-on or monotherapy in 75 patients with bipolar disorder who were either nonresponsive or intolerant of ongoing pharmacotherapy.

Table 31.7c-3 Lamotrigine Dosing (mg/day)

Other Disorders Lamotrigine can be of benefit in the treatment of psychosis arising in the course of epilepsy. This is a potentially important property since neuroleptic drugs, the usual treatment, can lower the seizure threshold. Lamotrigine possesses antinociceptive activity in animal models, although this has not been noted clinically. There have been reports of lamotrigine providing benefit to patients with restless legs syndrome.

Precautions and Adverse Reactions The most commonly observed adverse effects associated with use of lamotrigine are dizziness, ataxia, somnolence, headache, diplopia, blurred vision, nausea, vomiting, and rash. Approximately 10 percent of individuals in premarketing clinical trials discontinued treatment because of an adverse event. Lamotrigine therapy is not associated with changes in weight.

The frequency of adverse events in a placebo-controlled study of lamotrigine in bipolar disorder was similar for patients receiving lamotrigine or placebo. The only exception was headache, which occurred more frequently among patients taking lamotrigine.

The product information for lamotrigine contains a boxed warning about severe, potentially life-threatening rashes associated with use of the drug. These rashes occur in adults at an estimated rate of 1 in 1000 and include Stevens-Johnson syndrome and toxic epidermal necrolysis. Rare deaths have been reported, but their numbers are too few to permit calculation of risk. In pediatric patients, the incidence of severe rash is very much higher than that reported in adults, with reports from clinical trials suggesting an upper frequency estimate of 1 in 50 to 1 in 100 pediatric patients who may develop a potentially life-threatening rash.

Why age is a risk factor for rash is unknown. Other than age, the risk of rash may be increased by (1) combining lamotrigine with valproic acid, (2) exceeding the manufacturer's recommended starting dosage of lamotrigine, and (3) exceeding the recommended rate of dose escalation for lamotrigine. Nearly all life-threatening rashes associated with lamotrigine have occurred within 2 to 8 weeks of treatment initiation. Rash can occur after prolonged treatment (e.g., 6 months) as well. Lamotrigine should be discontinued at the first sign of rash, unless the rash is clearly not drug related. In some clinical trials the rate of rash has been significantly higher among placebo-treated patients than those being treated with lamotrigine. Routinely informing patients of the need to self-monitor for rash, while necessary, apparently leads to overrecognition in some instances.

Drug Interactions Drugs that inhibit the hepatic cytochrome P450 (CYP) system, particularly the CYP 3A4 isozyme, increase lamotrigine concentrations. Carbamazepine and other hepatic enzyme stimulators increase the metabolism of lamotrigine.

Lamotrigine does not affect the metabolism of other drugs; however, combined use of lamotrigine with other anticonvulsants can alter its pharmacokinetics. Valproate inhibits the metabolism of lamotrigine, increasing its half-life to 70 hours. Phenytoin (Dilantin), carbamazepine, phenobarbitone (Donnatal, Quadrinal), and primodone (Mysoline), hepatic enzyme–inducing agents, enhance metabolism of lamotrigine, reducing its half-life to about 13 hours.

Laboratory Interferences Lamotrigine does not interfere with common laboratory tests.

Dosage and Administration Because serious rash is an early adverse effect that is linked to plasma concentrations of lamotrigine, it is recommended that lamotrigine treatment be initiated at a dosage of 25 mg a day and increased slowly. When used as monotherapy, the suggested dosing schedule should be that indicated in Table 31.7c-3. Dosage should be adjusted for patients taking valproate or carbamazepine. The maximum daily dose varies, depending on concurrent medications. When used as monotherapy the maximum dosage is 500 mg a day in divided doses.

TOPIRAMATE

Topiramate was synthesized in an attempt to develop an analogue of fructose that could block gluconeogenesis and be used as a hypoglycemic agent. During animal testing, it failed to exhibit hypoglycemic activity but was noted to have antiseizure activity. It gained FDA approval in December 1996. It is indicated as adjunctive therapy for the treatment of adults with partial-onset seizures.

Pharmacological Actions

Pharmacokinetics Absorption of topiramate is rapid and almost complete. It has about 80 percent bioavailability and is not affected by food. Peak plasma concentrations occur about 2 hours following a 400-mg oral dose. The pharmacokinetics of topiramate are linear, with dose-proportional increases over the standard dosage range of 200 to 800 mg a day.

The mean plasma elimination half-life is 21 hours. Steady state is reached in 4 days. Topiramate is poorly bound to plasma proteins (13 to 17 percent) and thus is unlikely to displace highly protein-bound drugs.

Topiramate is not extensively metabolized, and 70 to 85 percent of an administered dose is eliminated unchanged in urine. Compared with those with normal renal function, the clearance of topiramate is reduced by about 40 percent in those with moderate renal impairment and by about 55 percent in severely renal-impaired persons. Topiramate is cleared by hemodialysis at a rate 4 to 6 times greater than that of a normal individual. The weight-adjusted clearance of topiramate is higher in children than adults. Although the mechanism is not understood, clearance of topiramate is decreased in patients with hepatic impairment.

Pharmacodynamics Three properties may contribute to the pharmacology of topiramate: sodium channel blocking, potentiation of GABA activity at the $GABA_A$ receptor, and antagonism of kainate-induced activation of the kainate–adenosine monophosphate (AMP) subtype of glutamate receptor. Topiramate also inhibits

some carbonic anhydrase isoenzymes, which may lead to formation of kidney stones but does not appear to affect topiramate's antiepileptic or mood activity.

Bipolar Disorder Open-label studies and case series reports have described the successful use of topiramate as a treatment for bipolar disorder. Populations described include both inpatients and outpatients with bipolar mood disorders who are unresponsive to conventional therapy. Too few patients have been reported to permit conclusions about the effectiveness and tolerability of topiramate as a mood stabilizer.

Precautions and Adverse Reactions The most common adverse events associated with the use of topiramate are central nervous system related. The most significant adverse effects can be classified in two broad categories: (1) psychomotor slowing, difficulty with concentration, and speech or language problems (in particular, word-finding difficulties) and (2) somnolence or fatigue. Dizziness or imbalance, confusion, memory problems, and irritability and depression are occasionally observed. The incidence of psychomotor slowing is only marginally dose related, but both language problems and difficulty with concentration or attention clearly increase in frequency with higher dosages. Detrimental effects on cognition tend to be associated with high initial dosages and rapid dosage escalation; they may develop insidiously and are often (but not always) time limited. About 1.5 percent of patients reported renal calculi during clinical trials, a tenfold increase over normal rates. There is no organ toxicity.

Unique among the anticonvulsants, topiramate use may be accompanied by significant weight loss. Of patients taking topiramate for a mean duration of 148 days, 72.3 percent lost a mean of 11 pounds. Because of this side effect, topiramate could in theory be combined with other anticonvulsants to counteract weight gain.

In reports of topiramate use to treat mood disorders, adverse effects have included anorexia and weight loss, paresthesia, agitation, confusion, hallucinations, somnolence, fatigue, and impaired concentration and memory. These effects were generally dose related. Symptoms of overdose include confusion, ataxia, hyperreflexia, and lethargy.

Drug Interactions When combined with other anticonvulsant drugs, topiramate has no effect on concentrations of carbamazepine, phenobarbital, or primodone. There is little or no increase in phenytoin concentration and a minimal decrease in valproic acid concentration. Carbamazepine and valproic acid lower concentrations of topiramate. Topiramate has no clinically relevant effect on the plasma levels of classical neuroleptic drugs, tricyclic drugs, theophylline (Theo-Dur) and coumarin. Concomitant use of topiramate with central nervous system (CNS) depressants can cause excessive sedation. Combined use with acetazolamide (Diamox) or other carbonic anhydrase inhibitors can increase the risk of renal stones. Topiramate decreases ethinyl estradiol (Brevicon, Demulen) concentrations by about one-third, possibly interfering with the effectiveness of contraceptive medication.

Laboratory Interferences Apart from laboratory values related to carbonic anhydrase inhibition, no laboratory interference has been associated with topiramate use.

Dosage and Administration Initial doses of topiramate should be as low as 25 mg a day. It can be titrated upward as tolerated to a maximum of 400 mg a day in two divided doses. In an open-label inpatient study of bipolar patients, dosages ranged from 50 to 1300 mg a day, with a mean of 614 mg a day.

TIAGABINE, VIGABATRIN, AND FELBAMATE

Three other new anticonvulsants, tiagabine (Gabatril), vigabatrin (Sabril), and felbamate (Felbatol), are available or waiting FDA approval. Unlike gabapentin, lamotrigine, and topiramate, these drugs have not been reported to have potential indications as treatments for conditions seen in psychiatric practice. Felbamate and vigabatrin are associated with serious adverse effects that have deterred casual off-label use. Because each of the newer compounds possesses mechanisms that are similar to agents shown to have behavioral and mood properties, these drugs may eventually be found to benefit patients with mental disorders.

Tiagabine Tiagabine was approved by the FDA in 1997 and is indicated as adjunctive therapy in adults and children 12 years and older in the treatment of partial seizures. It is believed to block reuptake of GABA into the presynaptic neurons, thus increasing extracellular concentrations of GABA. At concentrations 20 to 400 times those inhibiting the uptake of GABA, tiagabine binds to histamine type 1, (H_1), 5-hydroxytryptamine type 1B, ($5\text{-}HT_{1B}$), and benzodiazepine receptors and chloride channels. There have been no reports of tiagabine use in the treatment of mental disorders.

Tiagabine is rapidly and completely absorbed, with peak plasma concentrations occurring about 45 minutes after ingestion. The metabolism of tiagabine has not been fully elucidated but it is likely to involve the cytochrome P450 CYP 3A isozyme. Based on studies with anticonvulsants, tiagabine does not appear to stimulate or inhibit the metabolism of other drugs.

The most common side effects of tiagabine are CNS related: dizziness or lightheadedness, lack of energy, somnolence, nervousness, irritability, tremor, and difficulty with concentration and attention. Nausea and abdominal pain may also occur.

Vigabatrin Vigabatrin, also called g-vinyl GABA, is available in Canada and has been on the market in Europe since the late 1980s. It is awaiting approval in the United States.

Vigabatrin irreversibly inhibits GABA-transaminase, the GABA degradation enzyme, resulting in enhanced release of GABA through the GABA transporter and increased brain GABA concentrations. It has also been noted to inhibit striatal dopamine release. Vigabatrin has little protein binding, has a long half-life, and is mainly excreted unchanged by the kidneys.

Preliminary studies have shown vigabatrin to be as effective as benzodiazepines in facilitating alcohol and cocaine withdrawal. Doses used for alcohol withdrawal were 2000 mg a day. Vigabatrin attenuates cocaine-induced increases in extracellular dopamine in the striatum and nucleus accumbens of animals. Vigabatrin inhibition of GABA-transaminase may account for these effects, since that enzyme regulates dopamine concentration. In animal studies, vigabatrin blocked cocaine-related drug-seeking behavior. Imaging studies showed that it decreased in dopamine following administration.

In contrast to topiramate, vigabatrin does not have detrimental effects on cognitive function. In rare instances, when used by patients with seizure disorder, vigabatrin caused such adverse behavioral effects as agitation, irritability, depression, and psychosis.

Visual field constriction and blurring has been reported during vigabatrin therapy. These visual symptoms may be severe and persistent. They appear to be associated with retinal cone system dysfunction. Some patients may experience modest weight gain. Should vigabatrin be found to be useful as a treatment for substance abuse or mood or anxiety disorders, there is a possibility that the drug may be prone to cause rebound or withdrawal symptoms. Some epilepsy patients taking vigabatrin have experienced increased frequency or severity of seizures (compared with baseline) when the drug was stopped abruptly. These symptoms improved when vigabatrin treatment was restarted and gradually tapered.

Felbamate The FDA approved felbamate, a derivative of the antianxiety agent meprobamate (Equanil, Miltown), in 1993. It is indicated as both adjuvant therapy and monotherapy for adults with partial seizures with or without generalization and in children with partial and generalized seizures associated with the Lennox-Gastaut syndrome (i.e., epilepsy and mental retardation in children). Felbamate interacts as an antagonist at the strychnine-insensitive glycine recognition site of the NMDA-receptor complex. Felbamate reduces glutamate activity, potentiates GABA, and has sodium channel–blocking properties.

Felbamate is not generally used to treat psychiatric disorders for several reasons. There is no literature indicating that it offers special benefits. In fact, felbamate use has been associated with adverse behavioral effects. These adverse effects, noted when felbamate was used as monotherapy in patients with epilepsy, include induction of mania and stimulant-like effects. Possibly serious adverse effects also represent a deterrent to casual, off-label use; these rare but potentially serious adverse effects are aplastic anemia and hepatic failure. Because of these adverse effects the use of felbamate is restricted to patients with seizure disorder who have not responded to, or have not tolerated, other available treatments. The rate of aplastic anemia is estimated to be between 1 in 3000 and 5000 patients. Rates are highest among those with a past history of aplastic anemia or disorders of the immune system. Felbamate product information contains a special high caution “black box” warning about use of the drug.

Common adverse effects of felbamate are dose related and include insomnia, somnolence, headache, nausea, vomiting, anorexia, and dizziness. Felbamate use is further complicated by the potential for pharmacokinetic interactions.

FUTURE DIRECTIONS

The newer anticonvulsants show preliminary evidence of efficacy for mental disorders. The comparative efficacies of the newer drugs and standard agents, both for short-term treatments and prophylactic therapy, have not been established. Effectiveness for these drugs for monotherapy or adjunctive therapy has not been determined. Dose requirements for off-label uses and efficacy and safety in children and elderly adults need to be clarified. Any differences in the therapeutic spectrum of specific drugs are unknown. For example, do some drugs more effectively treat manic episodes than depressive episodes? Are there differences in effectiveness for mixed episodes or rapid cycling? Until more is known about the effectiveness of these drugs in their off-label uses, they should be reserved for patients who are unresponsive or intolerant of conventional agents. When using these drugs, inform the patient about their off-label use for medicolegal reasons. Discuss and document the rationale for use of these drugs, especially if approved treatments have not been tried first. Review risks that may be serious (e.g., the rash with lamotrigine and the visual disturbances with vigabatrin). Evidence supporting the effectiveness of topiramate beyond epilepsy is too preliminary to permit meaningful comparisons with lamotrigine and gabapentin.

SUGGESTED CROSS-REFERENCES

Bipolar disorders are discussed in Chapter 14 on Mood Disorders. Section 31.7a discusses carbamazepine, and Section 31.7b discusses valproate. Seizure disorders are covered in Section 2.4 on neuropsychiatric aspects of epilepsy.

SECTION REFERENCES

*Ben-Menachem E: Topiramate: Current status and therapeutic potential. Exp Opin Invest Drugs *6:* 1085, 1997.

*Boyd RA, Turck D, Abel RB, Sedman AJ, Bockbrader HN: Effects of age and gender on single-dose pharmacokinetics of gabapentin. Epilepsia *40:* 474, 1999.

*Cabras PL, Hardoy MJ, Hardoy MC, Carta MG: Clinical experience with gabapentin in patients with bipolar or schizoaffective disorder: Results of an open-label study. J Clin Psychiatry *60:* 245, 1999.

Calabrese JR, Fatemi S, Woyshville M: Antidepressants effects of lamotrigine in rapid cycling bipolar disorder. Am J Psychiatry *153:* 1236, 1996.

*Cora-Locatelli G, Greenberg BD, Martin J, Murphy DL: Gabapentin augmentation for fluoxetine-treated patients with obsessive-compulsive disorder. J Clin Psychiatry *59:* 480, 1998.

Fatemi S, Rapport D, Calabrese J, Thuras P: Lamotrigine in rapid-cycling bipolar disorder. J Clin Psychiatry *58:* 12, 1997.

Fogelson D, Sternbach H: Lamotrigine treatment of refractory bipolar disorder. J Clin Psychiatry *58:* 271, 1997.

Ghaemi SN, Katzow JJ, Desai SP, Goodwin FK: Gabapentin treatment of mood disorders: A preliminary study. J Clin Psychiatry *59:* 426, 1998.

Hardin CL, Pick LH: Alterations in mood and anxiety in epilepsy patients treated with gabapentin. Epilepsia *37:* 137, 1996.

Hill RR, Stagno SJ, Tesar GE: Secondary mania associated with the use of felbamate. Psychosomatics *36:* 404, 1995.

Ketter TA, Malow BA, Flamini RKD, White SR, Post RM, Theodore WH: Felbamate monotherapy has stimulant-like effects in patients with epilepsy. Epilepsy Res *23:* 129, 1996.

*Knoll J, Stegman K, Suppes T: Clinical experience using gabapentin adjunctively in patients with a history of mania or hypomania. J Affect Disord *49:* 229, 1998.

Kotler M, Matar M: Lamotrigine in the treatment of resistant bipolar disorder. Clin Neuropharmacol *21:* 65, 1998.

*Kushnir MM, Crossett J, Brown PI, Urry FM: Analysis of gabapentin in serum and plasma by solid-phase extraction and gas chromatography-mass spectrometry for therapeutic drug monitoring. J Anal Toxicol *23:* 6, 1999.

Kusumakar V, Yatham L: Lamotrigine treatment of rapid cycling bipolar disorder. Am J Psychiatry *154:* 1171, 1997.

Kusumakar V, Yatham L: An open study of lamotrigine in refractory bipolar depression. Psychiatry Res *72:* 145, 1997.

Labbate L, Rubey R: Lamotrigine for treatment-refractory bipolar disorder. Am J Psychiatry *154:* 1317, 1997.

*Martin R, Kuzniecky R, Ho S, Hetherington H, Pan J, Sinclair K, Gilliam F, Faught E: Cognitive effects of topiramate, gabapentin, and lamotrigine in healthy young adults. Neurology *52:* 321, 1999.

*Petroff OA, Hyder F, Mattson RH, Rothman DL: Topiramate increases brain GABA, homocarnosine, and pyrrolidinone in patients with epilepsy. Neurology *52:* 473, 1999.

*Pollack MH, Matthews J, Scott EL: Gabapentin as a potential treatment for anxiety disorders. Am J Psychiatry *155:* 992, 1998.

*Ryback RS, Brodsky L, Muasifi F: Gabapentin in bipolar disorder. J Neuropsychiatry Clin Neurosci *9:* 301, 1997.

Sporn J, Sachs G: The anticonvulsant lamotrigine in treatment-resistant manic-depressive illness. J Clin Psychopharmacol *17:* 185, 1997.

Stanton SP, Keck PE Jr, McElroy SL: Treatment of acute mania with gabapentin (Letter). Am J Psychiatry *154:* 287, 1997.

*Stephen LJ, Maxwell JE, Brodie MJ: Transient hemiparesis with topiramate. Br Med J *318:* 845, 1999.

Walden J, Hesslinger B, van Calker D, Berger M: Addition of lamotrigine to valproate may enhance efficacy in the treatment of bipolar affective disorder. Pharmacopsychiatry *29:* 193, 1996.

31.8 ANTIHISTAMINES

LAWRENCE S. GROSS, M.D., AND GEORGE M. SIMPSON, M.D.

INTRODUCTION AND HISTORY

Histamine is a low-molecular-weight amine that interacts with specific receptors in various target tissues. In addition to its role in allergic reactions and gastric acid secretion, there is considerable evidence that histamine functions as a central nervous system (CNS) neurotransmitter. Histamine-blocking agents were first discovered in the 1930s, introduced into clinical medicine in the 1940s, and in common use by the 1950s. Historically, antihistamines have been used in psychiatry for nonspecific sedation, as hypnotics, and to treat anxiety, childhood behavior problems, and the adverse effects of antipsychotic medication. The antihistamines most commonly used in psychiatric practice block histamine at the H_1 receptor and are considered traditional (first-generation) antihistamines. They are listed in Table 31.8-1 and will be the focus of this discussion.

Generic Name	Brand Name
Diphenhydramine	Benadryl
Hydroxyzine	Atarax
Hydroxyzine	Vistaril
Promethazine	Phenergan
Cyproheptadine	Periactin

Table 31.8-1 Antihistamines Used in Psychiatry

A new class of drugs that blocked histamine-induced gastric acid secretion was discovered in the 1970s, leading to the development of the histamine type 2 (H_2) receptor antagonists. In the 1980s nonsedating (second-generation) histamine type 1 (H_1) receptor antagonists were developed to treat allergic diseases. Although not used as psychiatric treatments, selected drugs from these categories are listed in Table 31.8-2 and are included because their potential toxicity and interactions with other drugs is of importance to psychiatrists.

Class	Generic Name	Brand Name	Comments
H_2 receptor antagonists	Cimetidine	Tagamet	Widely prescribed for treatment of ulcers and gastroesophageal reflux. All have the potential for CNS toxicities, including psychosis and delirium
	Ranitidine	Zantac	
	Famotidine	Pepcid	
	Nizatidine	Axid	
Second-generation (nonsedating) H_1 receptor antagonists	Terfenadine	Seldane	May prolong QT interval and induce serious cardiac arrhythmia, particularly when co-administered with drugs inhibiting hepatic cytochrome P450 3A4. Production of terfenadine stopped February, 1998; use of astemizole not recommended with SSRIs or nefazodone.
	Astemizole	Hismanal	
	Loratadine	Claritin	No apparent cardiotoxicity, but increased levels can occur when co-administered with CYP 3A4 inhibitors.
	Cetirizine	Zyrtec	
	Fexofenadine	Allegra	

Table 31.8-2 Other Antihistamines With Relevance to Psychiatric Practice

CHEMISTRY

All available H_1 antagonists are reversible, competitive inhibitors of histamine at the H_1 receptors. Like histamine, they contain a substituted ethylene moiety, but, unlike histamine, they have other groupings that may be used to broadly classify the traditional (first-generation) antihistamines. The classes of the agents most widely used in psychiatry include the ethanolamines (diphenhydramine [Benadryl]), the piperazines (hydroxyzine [Vistaril], hydrochloride [Atarax]), and the phenothiazines (promethazine [Phenergan]). Cyproheptadine (Periactin) contains a tricyclic nucleus with a piperadine side ring. The chemical structures of these compounds are shown in Figure 31.8-1.

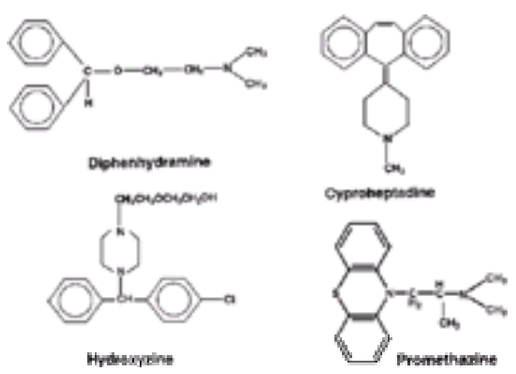

FIGURE 31.8-1 Molecular structures of antihistamines used in psychiatric practice.

PHARMACOLOGICAL ACTIONS

Most H_1 antagonists have similar pharmacological properties and can be discussed together.

Pharmacokinetics

Absorption The first-generation antihistamines are well absorbed from the gastrointestinal tract. Following oral administration, effects begin within 15 to 30 minutes and maximal blood concentrations are reached in 1 to 3 hours. Duration of action is generally 4 to 6 hours, but some compounds are much longer acting. Parenteral injection produces a more rapid onset of action but may cause hypotension, particularly with rapid intravenous administration.

Distribution Antihistamines such as diphenhydramine are widely distributed throughout the body and are contraindicated in nursing mothers. These drugs cross the blood-brain barrier and have both central and peripheral effects.

Metabolism and Elimination The first-generation antihistamines are extensively metabolized by the liver, with most eliminated as metabolites in the urine; little, if any antihistamine is excreted unchanged. Elimination is more rapid in children and slower in patients with severe hepatic disease. H_1 receptor antagonists induce hepatic microsomal enzymes and therefore may facilitate their own metabolism.

Pharmacodynamics

Mechanism of Action The clinical effects of antihistamines are the result of their blockade of histamine receptors. H_1 receptors are present throughout the CNS, with dense concentrations found in the hypothalamus. Histamine increases wakefulness via H_1 receptors, which may explain the sedative potential of the traditional antihistamines.

Many first-generation H_1 antagonists, including diphenhydramine, hydroxyzine, promethazine, and cyproheptadine, also block responses to acetylcholine mediated by muscarinic receptors. In addition to its antihistaminic and anticholinergic effects, cyproheptadine is a potent serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptor antagonist.

Blood Levels Blood levels of antihistamines are not routinely obtained in clinical practice.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

There are no recent studies clearly demonstrating the safety and efficacy of traditional antihistamines in psychiatric disorders. The original literature addressing the sedative and hypnotic uses of these agents is decades old, limiting comparison to research using current diagnostic criteria. More recent studies of antihistamines as sleep aids demonstrated significantly more improvement in several sleep parameters compared with placebo, but these effects were less pronounced than the effects of the benzodiazepines.

An open clinical trial has indicated that cyproheptadine may be useful in the treatment of neuroleptic-induced akathisia. A double-blind, placebo-controlled study of the treatment of chronic schizophrenia with cyproheptadine showed no improvement in negative symptoms and exacerbation of some positive symptoms. Preliminary findings from open-label studies and case reports suggest that famotidine (Pepcid), an H_2 receptor antagonist, may be beneficial as an adjunctive treatment for schizophrenia.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Target organ effects of the traditional antihistamines result from the direct action of their blockade of histamine receptors or their associated anticholinergic activity; they are associated with both CNS depression and excitation. Sedation is common at therapeutic dosages, but these agents vary in their ability to cause CNS depression and patients vary in their response to individual drugs. The ethanolamines, such as diphenhydramine, are particularly sedating. Stimulation occasionally occurs in patients at therapeutic dosages, and central excitation may be a dramatic feature of poisoning. Some H_1 antagonists, particularly promethazine, also have antiemetic effects and local anesthetic activity.

H_1 antagonists block the action of histamine on capillary permeability and respiratory and vascular smooth muscle, which helps to suppress the vasodilation, edema, and pruritis associated with various allergic responses. These drugs inhibit histamine-mediated bronchoconstriction in challenge tests but, at usual dosages do not inhibit bronchoconstriction in patients with asthma or other hypersensitivity reactions.

Unlike the H_2 antagonists, the H_1 antagonists do not inhibit gastric secretion at all; they suppress histamine-mediated salivary, lacrimal, and other exocrine secretions to varying degrees. The anticholinergic properties of these agents may reduce secretions in cholinergically innervated glands and in tissues such as the respiratory system.

THERAPEUTIC INDICATIONS

Historically, H_1 antagonists have been widely used for nonspecific sedation and as mild hypnotics, particularly in over-the-counter sleep aids. They have largely been replaced in psychiatric practice by benzodiazepines, which have fewer side effects and are more effective for the treatment of anxiety and for the induction and maintenance of sleep. Antihistamines may have a role in the treatment of patients with histories of benzodiazepine abuse or other contraindications to benzodiazepine use. Although antihistamines are prescribed for the treatment of childhood behavioral problems, a recent review of the literature found no clear indication for their use in the treatment of child and adolescent anxiety disorders. They may have some anxiolytic effect, but their use is limited by lowered seizure threshold, sedation, anticholinergic adverse effects, and the potential for abuse in adolescents.

Diphenhydramine Diphenhydramine is indicated for the treatment of allergic reactions and motion sickness. It is also indicated for treating parkinsonism (including neuroleptic-induced parkinsonism) in elderly patients unable to tolerate more potent agents, other age groups with mild parkinsonism, or patients requiring treatment in combination with centrally acting anticholinergic agents. The injectable form of diphenhydramine is effective when oral treatment is impractical (e.g., acute allergic or dystonic reaction). Although probably not as effective as the more specific anticholinergic agents in the treatment of drug-induced extrapyramidal side effects, diphenhydramine could be useful to treat extrapyramidal adverse effects in patients with acute psychotic agitation when additional sedation may be desirable. When administered intravenously, diphenhydramine is the drug of choice for the treatment of dystonic reactions induced by dopamine-blocking agents; however, it is not effective in relieving akathisia. Diphenhydramine is also found in a variety of nonprescription cold and allergy preparations and in sleep-inducing aids.

Hydroxyzine Hydroxyzine is the only antihistamine with Food and Drug Administration (FDA) approval for use in the symptomatic relief of anxiety and tension associated with psychiatric and medical conditions. Its effectiveness as an antianxiety agent for periods exceeding 4 months has not been systematically assessed. It is also indicated for the management of pruritis caused by allergic conditions, and it is used as a premedication sedative and following general anesthesia.

Promethazine As one of the more sedating and anticholinergic H_1 antagonists, promethazine is indicated in the treatment of a variety of conditions, including allergic reactions and motion sickness. It is used in the surgical setting as an antiemetic; as an adjunct to analgesics for postoperative pain control; and as a preoperative, postoperative, or obstetric sedative. It is also indicated for sedation, relief of apprehension, and production of light sleep in both children and adults.

Cyproheptadine Cyproheptadine has FDA approval for use in the treatment of vasomotor rhinitis and a variety of allergic reactions. Because of its serotonin-blocking effects in addition to its H_1–histamine-receptor antagonism, cyproheptadine has been used in the treatment of antidepressant-induced sexual dysfunction; however, adverse effects have been reported and results have been inconsistent. Systematic studies in this area are lacking.

An open clinical trial of 4 mg of cyproheptadine four times a day by mouth reported improvement in 15 of 17 patients with neuroleptic-induced akathisia. Case reports

of other conditions indicating improvement following cyproheptadine treatment include recurrent nightmares with posttraumatic themes and one case of suspected serotonin syndrome.

PRECAUTIONS AND ADVERSE REACTIONS

The most common adverse effects of first-generation H_1 antagonists involve the nervous system and the gastrointestinal tract. These drugs have a high incidence of sedation, which may be a desirable effect in some patients but may interfere with the patient's daily activities. Other nervous system reactions include dizziness, tinnitus, incoordination, fatigue, blurred vision, insomnia, and tremors. Promethazine, a phenothiazine derivative, can cause extrapyramidal symptoms. Antihistamines may be associated with CNS-related emotional reactions such as lassitude, nervousness, and euphoria. Gastrointestinal effects, which may be reduced by giving the drug with meals, include loss of appetite, epigastric distress, nausea, vomiting, and constipation or diarrhea. Untoward anticholinergic effects include dryness of the mouth, thickening of respiratory secretions, urinary frequency, dysuria, and urinary retention.

Because of their anticholinergic actions, these antihistamines should be used with great caution in patients with narrow-angle glaucoma, stenosing peptic ulcer, pyloroduodenal obstruction, symptomatic prostatic hypertrophy or bladder neck obstruction. Caution should also be exercised in the treatment of patients with hyperthyroidism, cardiovascular disease, hypertension, increased intraocular pressure, and history of bronchial asthma. Elderly patients may be more likely to experience sedation, confusion, dizziness, and hypotension; young children may experience paradoxical excitation.

Although uncommon, first-generation H_1 antagonists have the potential for abuse and dependence. They may be fatal in overdose, with excitatory effects posing the greatest danger in acute poisoning. The syndrome includes hallucinations, excitement, ataxia, incoordination, and seizures. Along with dilated pupils, sinus tachycardia, urinary retention, and fever, there may be deepening coma progressing to cardiorespiratory collapse and death; treatment is generally symptomatic and supportive.

There are no adequate and well-controlled studies of these antihistamines in pregnant women. The drugs fall into the FDA's Pregnancy Categories B (no evidence of risk in humans) and C (risk cannot be ruled out). Results are similarly lacking regarding the excretion of these drugs in breast milk, but their use is generally contraindicated in nursing mothers. Clinical studies and experience with these antihistamines do not suggest carcinogenicity or teratogenicity in humans.

Despite not being clinically used in psychiatry, it is important to note that H_2 antagonists, widely prescribed for the treatment of upper gastrointestinal disorders, have all been associated with CNS toxicities. Reactions reported have included psychosis, agitation, hallucinations, delirium, disorientation, confusion, obtundation, and hostility. A review of the literature indicated that the symptoms usually emerged during the first 2 weeks of treatment and most resolved within 3 days of discontinuing therapy; all the reviewed reactions resolved within 7 days. Cimetidine (Tagamet) was most associated with CNS reactions, but there was no clear evidence to indicate that one H_2 blocker was more likely than another to cause these symptoms. Although the estimated incidence of CNS toxicities was low (0.2 percent in outpatients versus 1.6 to 1.9 percent in hospitalized patients), consulting psychiatrists should be aware of the potential CNS toxicities of these widely used medications.

DRUG INTERACTIONS

Monoamine oxidase inhibitors intensify and prolong the anticholinergic effects of antihistamines, so these drugs should be used with caution. The sedative effects of traditional H_1 antagonists are additive with those of alcohol and other CNS depressants. Hydroxyzine and promethazine may potentiate the actions of narcotics, nonnarcotic analgesics, and barbiturates, necessitating dosage reductions of these CNS depressants during coadministration. In case reports, the antidepressant and antibulimic effects of fluoxetine (Prozac) were reversed in a few patients when cyproheptadine was added to treat fluoxetine-induced anorgasmia.

The second-generation H_1 antagonists astemizole (Hismanal) and terfenadine (Seldane) have the potential to cause prolongation of the QTc interval, resulting in polymorphic ventricular tachycardia (*torsades de pointes*). This can occur when these drugs are taken in higher-than-recommended doses, in patients with liver disease, or by coadministration of drugs that inhibit the hepatic cytochrome P_{450} (CYP) 3A4 isoenzyme, thought to be primarily responsible for terfenadine and astemizole metabolism. Of importance to psychiatrists is the fact that the serotonergic antidepressants nefazodone (Serzone) and fluvoxamine (Luvox), inhibitors of CYP 3A4, should not be used with terfenadine and astemizole because of potentially fatal cardiac arrhythmias. The manufacture of terfenadine (Seldane) in the U.S. was stopped in early 1998 because of cardiac risks. The drug fexofenadine (Allegra) provides essentially the same antihistaminic effects without apparent cardiac risks.

Because of minor metabolic pathways of astemizole (Hismal) involving CYP 1A2 and 2D6, the list of substances not recommended for coadministration with astemizole has been expanded to include all of the serotonin selective reuptake inhibitors as well as nefazodone. Loratadine (Claritin), cetirizine (Zyrtec), and fexofenadine, nonsedating antihistamines not associated with this cardiotoxic interaction, are considered safer choices for patients taking these antidepressants. However, levels of these antihistamines can be increased in the presence of CYP 3A4 inhibitors, and recent reports have suggested closer inspection of their safety profiles.

Clinicians should inquire about the use of over-the-counter sleep preparations, many of which contain antihistamines and may be involved in interactions with other medications.

LABORATORY INTERFERENCES

H_1 antagonists can interfere with allergy skin tests and should be discontinued at least 1 week before any such testing is performed. Promethazine has been reported to interfere with pregnancy tests and increase blood glucose levels. When taken within 2 weeks of delivery, promethazine may inhibit platelet aggregation in the newborn. One report found that large quantities of diphenhydramine resulted in a false-positive urine screen for phencyclidine.

DOSAGE AND ADMINISTRATION

The recommended preparations and dosages for children and adults for the antihistamines most commonly used in psychiatry are listed in Table 31.8-3.

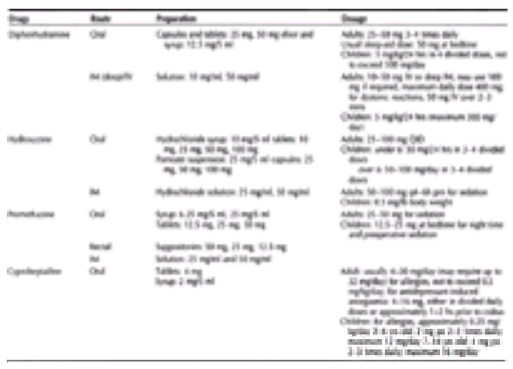

Table 31.8-3 Dosage and Administration of Traditional Antihistamines

SUGGESTED CROSS-REFERENCES

Treatment of extrapyramidal side effects is discussed in Section 31.4, benzodiazepines are discussed in Section 31.9, treatment of anxiety and anxiety disorders is discussed in Section 15.7 and Section 15.8, hypnotic medications and the treatment of insomnia are discussed in Chapter 21, pharmacotherapy in children is discussed in Section 48.6, antidepressant-induced sexual dysfunction is discussed in Section 19.1a, and antidepressant-related metabolic effects and drug

interactions are discussed in Section 31.2.

SECTION REFERENCES

*Aizenberg D, Zemishlany Z, Weizman A: Cyproheptadine treatment of sexual dysfunction induced by serotonin reuptake inhibitors. Clin Neuropharmacol *18:* 320, 1995.

Arnott S, Nutt D: Successful treatment of fluvoxamine-induced anorgasmia by cypraheptadine. Br J Psychiatry *164:* 838, 1994.

*Babe KS Jr, Serafin WE: Histamine, bradykinin, and their antagonists. In *Goodman and Gilman's The Pharmacological Basis of Therapeutics*, ed 9, JG Hardman, LE Limbird, PB Molinoff, RW Ruddon, AG Gilman, editors. McGraw-Hill, New York, 1996, p 581.

Brophy MH: Cyproheptadine for combat nightmares in post-traumatic stress disorder and dream anxiety disorder. Military Med *156:* 100, 1991.

*Cantu TG, Korek JS: Central nervous system reactions to histamine-2 receptor blockers. Ann Intern Med *114:* 1027, 1991.

Decastro RM: Reversal of MAOI-induced anorgasmia with cyproheptadine (Letter). Am J Psychiatry *142:* 783, 1985.

De Nesnera AP: Diphenhydramine dependence: A need for awareness (Letter). J Clin Psychiatry *57:* 136, 1996.

Drug Evaluations Annual 1994. American Medical Association, Chicago, 1993.

Feder R: Reversal of antidepressant activity of fluoxetine by cyproheptadine in three patients. J Clin Psychiatry *52:* 163, 1991.

Goldbloom DS, Kennedy SH: Adverse interaction of fluoxetine and cyproheptadine in two patients with bulimia nervosa. J Clin Psychiatry *52:* 261, 1991.

Harsch HH: Cyproheptadine for recurrent nightmares (Letter). Am J Psychiatry *143:* 1491, 1986.

Himmel MH, Honig PK, Worobec AS: Dangers of non-sedating antihistamines (Letter). Lancet *350:* 69, 1997.

Kahn DA: Possible toxic interaction between cyproheptadine and phenelzine, letter to editor. Am J Psychiatry *144:* 1242, 1987.

*Kutcher SP, Reiter S, Gardner DM, Klein RG: The pharmacotherapy of anxiety disorders in children and adolescents. Psychiatr Clin North Am *15:* 41, 1992.

Lappin RI, Auchincloss EL: Treatment of the serotonin syndrome with cyproheptadine (Letter). N Engl J Med *331:* 1021, 1994.

Levine BS, Smith ML: Effects of diphenhydramine on immunoassays of phencyclidine in urine (Letter). Clin Chem *36:* 1258, 1990.

Lindquist M, Edwards IR: Risks of non-sedating antihistamines (Letter). Lancet *349:* 1322, 1997.

Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry *153:* 311, 1996.

Nightingale SL: From The Food and Drug Administration. JAMA *277:* 370, 1997.

Oyewumi LK, Vollick D, Merskey H, Plumb C: Pamotidine as an adjunct treatment of resistant schizophrenia. J Psychiatry Neurosci *19:* 145, 1994.

**Physicians' Desk Reference*, 53 ed. Medical Economics Company, Montvale, NJ, 1999.

Prell GD, Rosse RB, Deutsch SI: Apparent absence of famotidine-antipsychotic drug interactions in patients with chronic schizophrenia (Letter). J Psychiatry Neurosci *21:* 61, 1996.

Riley AJ, Riley EJ: Cyproheptadine and antidepressant-induced anorgasmia (Letter). Br J Psychiatry *148:* 217, 1986.

Silver H, Blacker M, Weller MPI, Lerer B: Treatment of chronic schizophrenia with cyproheptadine: A double-blind, placebo-controlled study. Biol Psychiatry *30:* 523, 1991.

*Weiss D, Aizenberg D, Hermesh H, Zemishlany Z, Munitz H, Radwan M, Weizman A: Cyproheptadine treatment in neuroleptic-induced akathisia. Br J Psychiatry *167:* 483, 1995.

*Yap YG, Camm AJ: The current cardiac safety situation with antihistamines. Clin Exp Allergy *29:* 15, 1999.

31.9 BARBITURATES AND SIMILARLY ACTING SUBSTANCES

LAWRENCE A. LABBATE, M.D., GEORGE W. ARANA, M.D., AND JAMES C. BALLENGER, M.D.

Adolph von Baeyer discovered the precursor to the barbiturates, barbituric acid, in 1864. Barbituric acid was the product of urea and malonic acid. Although various hypotheses exist regarding the origin of the name “barbituric,” it was likely derived from the word “urea” and the name “Barbara,” though to whom it refers is uncertain. Barbital, the first drug of the class with hypnotic action, was discovered in Berlin, Germany, in 1892 by Josef von Mering and Emil Fisher. Barbital was introduced to clinical practice as Veronal in 1903 and became widely used for its sedative and hypnotic properties. Numerous other barbiturates were subsequently synthesized, including phenobarbital (Solfoton, Luminel) in 1912.

Although the barbiturates were an improvement over the bromine salts, opium, or ethanol for conditions such as anxiety or insomnia, their usefulness was limited by their narrow therapeutic range and low therapeutic index (LD_{50}/ED_{50}), the ratio of lethal to effective dose. Some barbiturates may cause respiratory compromise or arrest at several times the effective dose. Although many benefited from the sedative and hypnotic qualities, this narrow therapeutic index led to many completed suicides. Several studies in the 1970s found barbiturates to be the most common drugs taken in overdose, and they accounted for a large percentage of suicides and hospitalizations from self-poisoning. Barbiturates have largely been replaced in current practice by the benzodiazepines which have a high therapeutic index and a relatively lower incidence of other complicating features (hepatic enzyme induction, development of life-threatening dependence and tolerance, and others). The primary uses of barbiturates in current psychiatric practice are anesthesia in electroconvulsive therapy (ECT), the differential diagnosis of catatonia, and detoxification of patients dependent on alcohol or other sedative hypnotics. In neurological practice barbiturates are used in the treatment of epilepsy or status epilepticus and to determine cerebral dominance for language (Wada test), discussed below.

CHEMISTRY

The barbiturates are the class of compounds centered on the core molecule barbituric acid (Fig. 31.9-1). Barbiturates in which the oxygen at C2 is replaced by a sulfur molecule are known as thiobarbiturates, the most common of which is thiopental (Pentothal). Compounds with an oxygen at C2 are properly called oxybarbarbiturates, but are commonly called barbiturates. It is, however, common practice to call both groups barbiturates. Although barbituric acid has no effects on the central nervous system (CNS), an alkyl or aryl group at C5 confers sedative-hypnotic properties. Large aliphatic groups at C5 increase lipid solubility and generally decrease latency to onset of action and duration of action. Methohexital (Brevital) is an example of a compound with a large aliphatic side chain.

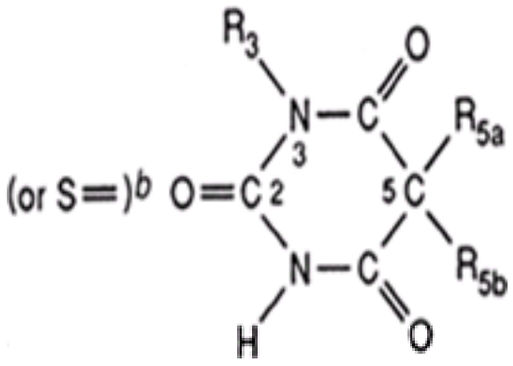

FIGURE 31.9-1 Molecular structure of barbiturates currently available in the United States.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption and Administration The barbiturates are routinely administered orally when used as hypnotic agents but may be given intravenously for seizures or anesthetic use. Intramuscular use is discouraged because soluble preparations cause pain and sometimes necrosis at the injection site. Although food in the stomach may delay absorption of barbiturates, bioavailability is unaffected, with most of the drug absorbed in the small intestine. This class of drugs has a broad range of lipophilicity, an important factor influencing absorption; the more lipophilic compounds are more rapidly absorbed (Table 31.9-1). Certain preparations are available as rectal suppositories, particularly for use in infants.

Barbiturate	Onset (min)	Duration (h)	Elimination Half-Life (h)
Amobarbital	60	10–12	8–42
Aprobarbital	45–60	6–8	14–34
Butabarbital	45–60	6–8	34–42
Mephobarbital	60+	10–12	11–67
Pentobarbital	10–15	3–4	15–48
Phenobarbital	60+	10–12	80–120
Secobarbital	10–15	3–4	14–40

Table 31.9-1 Half-Lives, Onset, and Duration of Action of Selected Barbiturates

Distribution Binding of barbiturates to plasma proteins is based on lipid solubility and ranges from lipophilic compounds with a highly bound fraction (80 percent), such as thiopental or methohexital, to phenobarbital (20 to 45 percent), which is less lipophilic. Weak acids such as warfarin (Coumadin) or aspirin may displace barbiturates from albumin and heighten barbiturate CNS effects. Highly soluble barbiturates such as secobarbital (Seconal) and pentobarbital (Nembutal) are rapidly absorbed into the brain according to vascular distribution and undergo redistribution to less vascular areas of the brain and other tissues within 30 minutes. This rapid

redistribution of barbiturates based on lipophilicity accounts for their relatively brief duration of action, which is predominantly influenced by redistribution of the compound into fat stores rather than the elimination half-life, which determines duration of action in most other classes of drugs. Barbiturates distribute to all tissues, cross the placenta, and are distributed to fetal tissues; they are secreted in breast milk of women treated with these compounds.

Elimination Barbiturates are primarily metabolized in the liver by P450 microenzymes. The barbiturates may undergo oxidation at C5, N hydroxylation, or N alkylation. Although most metabolized barbiturates are inactive, the N-methyl congeners produce active metabolites. For example, mephobarbital (Mebaral) gives rise to phenobarbital. Many inactivated barbiturates also undergo glucuronidation to allow renal elimination. Although most compounds undergo hepatic metabolism, the more water soluble compounds such as phenobarbital and aprobarbital (Alurate) are excreted partly unchanged in the urine. With impaired renal function, concentrations of these drugs will be higher, leading to CNS and cardiovascular depression. More lipophilic barbiturates such as secobarbital or pentobarbital are much less affected by renal function.

The elimination half-lives of select oral-use barbiturates (Table 31.9-1) are long enough that drug will accumulate with repeated daily dosing. Repeated use induces hepatic microenzymes and thus shortens elimination half-life over the initial 2 to 4 weeks of treatment. This calls for caution in the first month of treatment because a dosage that appears adequate in the first week may be insufficient in the fourth week. This increased metabolism affects other drugs that depend on this system, for example, carbamazepine (Tegretol). Elimination half-life is shorter in young persons than in elderly individuals or infants. Half-life may also be increased in chronic liver disease and cirrhosis. Barbiturate dosage should be lowered or the drug avoided completely in cirrhosis.

Pharmacodynamics

Mechanism of Action Although the exact mechanism of action is unknown, barbiturates likely potentiate g-amino butyric acid (GABA)-induced increases in chloride conductance (at postsynaptic GABA Type A [$GABA_A$] receptors) and reduce glutamate-induced depolarization (at a-amino-3-hydroxy-a-methyl-4-isoxazole propionate [AMPA] glutamate receptors) within the CNS. Thus, barbiturates increase the effects of the primary brain inhibitory transmitter, GABA, and decrease the effects of the primary excitatory transmitter, glutamate. $GABA_A$ receptors are a heterogeneous group of receptors composed of five subunits coupled to a chloride ionophore. The variety and heterogeneity of the subunits allow differential effects within the brain and multiple sites of action. Barbiturates resemble benzodiazepines in their mechanism of action, though barbiturates and benzodiazepines act at separate sites on the $GABA_A$-chloride-ionophore complex. Barbiturates do not displace benzodiazepines from their binding sites; they enhance binding of benzodiazepines at the receptor complex, magnifying the effects of the benzodiazepines. The effects of barbiturates depend largely on the presence of chloride or similarly acting anions. Picrotoxin, a drug that limits chloride flow through the ionophore, antagonizes the effects of barbiturates. At anesthetic or toxic doses, however, barbiturates directly increase chloride channel opening, even in the absence of GABA. This is an important difference between barbiturates and benzodiazepines; benzodiazepines do not directly increase chloride channel opening. This difference likely accounts for the relative safety of the benzodiazepines compared with the barbiturates.

Although barbiturates act as depressants at all levels of the neuraxis, the reticular activating system is particularly sensitive to their effects, accounting for their efficacy as sedative-hypnotics. Other brainstem effects of barbiturates are dose dependent and include induction of anticonvulsant effects with respiratory depression at high doses and anesthesia at even higher doses. The mechanism of action for anticonvulsant effects is not certain. It is postulated that barbiturates exert their anticonvulsant effects by decreasing neuronal membrane excitability (primarily by enhancing $GABA_A$ receptor–mediated inhibition) and depressing monosynaptic and polysynaptic transmission.

Blood Concentrations Plasma concentration is directly related to level of CNS depression, though because of neuroadaptation, long-term barbiturate users may have relatively little sedation despite high plasma concentration. On the other hand, naive users may experience toxicity when concentration is in the therapeutic range. The typical therapeutic range for phenobarbital is 10 to 40 µg/mL, though patients may require higher or lower dosages depending on individual differences such as length of treatment, severity of illness, or concurrent use of other anticonvulsants.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

The barbiturates were widely used before the Food and Drug Administration (FDA) established standards for new drugs. Although many barbiturates have been approved for chronic anxiety and short-term treatment of insomnia, few controlled studies have compared barbiturates with placebo or benzodiazepines, and no controlled studies exist for the treatment of syndromes such as anxiety or panic disorders using the currently accepted nosology of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). The barbiturates, especially amobarbital (Amytal) and secobarbital, appear to be effective hypnotics. One double-blind study found that a combination of amobarbital (100 mg), and secobarbital (100 mg) was more effective than flurazepam (Dalmane) for sleep induction and maintenance and did not lose efficacy after 2 weeks. (Some studies have found loss of efficacy of barbiturates after two weeks.) Although the highly lipid soluble barbiturates may effectively treat insomnia, the authors do not recommend their use. Complications that may arise from the development of tolerance, the narrow therapeutic index, drug-drug interactions, and hepatic microsomal induction render this class of compounds less desirable than benzodiazepines or newer nonbenzodiazepine sedative-hypnotics.

The barbiturates may provide some relief from anxiety, but little evidence from controlled studies or uncontrolled reports suggest that they are useful in the treatment of panic disorder, social phobia, other phobias, posttraumatic stress disorder, or obsessive-compulsive disorder. Barbiturates carry a higher risk for precipitating respiratory or cardiac failure than the benzodiazepines or selective serotonin reuptake inhibitors (SSRIs); their use for the treatment of anxiety or anxiety disorders is limited to patients with known sensitivity to benzodiazepines or SSRIs because of adverse effects or allergic phenomena. Because of their known potential for physical dependence, barbiturates carry a high risk for abuse or addiction and have been classified by the FDA as schedule II, III, or IV compounds.

Two anesthetic agents have been studied in controlled trials. Methohexital appears to be well tolerated and effective in the induction of anesthesia for ECT and is widely used for this purpose. Methohexital was compared with propofol (Diprivan) and found to be equally efficacious, even though seizures last longer with methohexital. Further, depression treatment outcome is equivalent with methohexital or propofol. Thiopental was assessed in controlled studies for the treatment of increased intracranial pressure following head injury and appears to be effective for intracranial hypertension, providing long-term improvement in morbidity and mortality of severely brain injured patients. Thiopental also appears to limit neuropsychiatric sequelae of patients undergoing cardiac surgery.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System Barbiturates depress all levels of the CNS, inhibiting afferent connections from the reticular activating system to the thalamus causing drowsiness; reducing transmission to the cerebral hemispheres; and at sufficiently high doses, inducing dose-dependent anesthesia. The sensitivity of the reticular activating system to this class of medication is responsible for the sedation, and depressed transmission in the cerebellum is thought to cause ataxia. Barbiturates also cause more-subtle effects on learning, memory, and fine motor skills, detectable on neuropsychological tests. Barbiturates reduce rapid eye movement (REM) sleep with short-term use, but tolerance develops. Barbiturates may also cause subtle effects on mood, ranging from depression to euphoria, and long-acting compounds have prolonged effects including irritability and drowsiness.

Respiratory System Barbiturates decrease respiratory drive and depress the rhythmic quality of respiration. Respiratory drive is only slightly diminished at hypnotic doses. At doses only several times higher than that for sleep induction, however, neurogenic drive is nearly eliminated. Higher doses reduce hypoxic and chemoreceptor drive, and doses used for deep anesthesia completely eliminate respiratory drive. The narrow therapeutic window makes barbiturates potentially dangerous to use in any patient with pulmonary insufficiency. Cough and laryngeal reflexes are generally preserved up to doses that limit respiration. Laryngospasm, a troublesome complication of barbiturates, may result from intravenous use of these compounds, and coughing, hiccoughs, and sneezing may also be induced in the early stages of barbiturate-induced anesthesia.

Cardiovascular System In normal subjects only suprapharmacological doses appreciably affect cardiac output or blood pressure. At hypnotic doses there are no significant cardiovascular effects. In patients with heart failure or shock, barbiturates may cause marked reduction in blood pressure. In general, barbiturate anesthesia is less likely to produce cardiac problems than volatile anesthetics. Cardiac decompensation is largely related to barbiturate poisoning. Selective b_1-adrenergic receptor agonists can overcome myocardial depression caused by poisoning.

Intestines Barbiturates used in hypnotic doses have little effect on smooth muscle motility; at higher doses, gastrointestinal motility is slowed. Symptomatic improvement from functional gastrointestinal syndromes is thought to be related to the CNS effects of barbiturates.

Liver The barbiturates increase production of cytochromes and lipid content of the hepatic smooth endoplasmic reticulum. This increases metabolism of steroid hormones, cholesterol, bile salts, vitamin K, and a number of drugs, including barbiturates, warfarin, tricyclic antidepressants, propanolol (Inderal), quinidine (Duraquin), and phenytoin (Dilantin). Barbiturates may also increase the production of other metabolic enzymes, including b-aminolevulinic acid (ALA) synthetase and aldehyde dehydrogenase. The increase in ALA may exacerbate porphyria in patients with intermittent porphyria; thus barbiturates are contraindicated in these patients.

Renal System Renal function and tubular transport may be decreased by very high doses of barbiturates. In general, barbiturates do not affect renal function, nor do they depend on renal function for elimination. Doses of phenobarbital and aprobarbital, drugs excreted through the kidneys, need to be lowered in renal failure.

INDICATIONS AND THERAPEUTIC USES

The following barbiturates have FDA approval for the treatment of anxiety and apprehension or insomnia: amobarbital, secobarbital, pentobarbital, butabarbital (Butisol), and aprobarbital. Amobarbital has also been approved for use in narcoanalysis, though this use is controversial. The FDA has approved methohexital for use in ECT anesthesia.

Electroconvulsive Therapy Methohexital is commonly used as an anesthetic agent for ECT. It has lower cardiac risks than other barbiturate anesthetics. Used intravenously, methohexital produces rapid unconsciousness and because of rapid redistribution has a brief duration of action (5 to 7 minutes). Methohexital, like other barbiturates, provides no analgesia before unconsciousness, so unconsciousness must be ensured before paralytic agents are administered. Typical dosing for ECT is 0.7 to 1.2 mg/kg. A number of studies comparing methohexital with propofol have demonstrated that seizures are shorter with propofol, but efficacy appears equivalent, and cognitive effects may be worse with methohexital. Methohexital may also be used to abort prolonged seizures in ECT or to limit postictal agitation.

Seizures The barbiturates have a long history of use in short-term and prophylactic treatment of seizures. Phenobarbital, the most commonly used barbiturate for seizures has indications for the treatment of generalized tonic-clonic and simple partial seizures. Parenteral barbiturates are used in the emergency management of seizures independent of cause, including meningitis, eclampsia, or tetanus. Intravenous phenobarbital should be administered slowly, 10 to 20 mg/kg for status epilepticus.

Wada Test To localize speech before resection of brain tissue associated with intractable seizures, intravenous barbiturates (most commonly amobarbital, occasionally methohexital) are injected into the carotid artery of the suspected side of interest. Although protocols vary by center, usually 100 to 200 mg of amobarbital is injected intravenously via a transfemoral catheter placed above the level of the carotid bifurcation. Barbiturates injected this way cause temporary contralateral limb paresis and language defects if the injected hemisphere is responsible for language. Memory and naming tests are performed during the Wada test and assessed after neurological function returns to normal. Good memory test results suggest that the unaffected side supports memory function and surgical ablation may leave little deficit. The Wada test may also be used in conjunction with functional neuroimaging studies to study memory and hippocampal perfusion, though this testing is not yet of regular clinical use.

Narcoanalysis Sodium amobarbital has been used historically as a diagnostic aid in a number of clinical conditions including conversion reactions, catatonia, hysterical stupor, and unexplained muteness and to differentiate stupor of depression, schizophrenia, and structural brain lesions. Much of this literature is anecdotal, poorly controlled, and difficult to interpret. Nevertheless, a limited number of these reports present compelling case material, particularly in various differential diagnostic situations. Some of this literature observed that amobarbital decreases arousal and cognitive function in patients with structural brain lesions, facilitates speech in depressed patients, and causes psychotic patients to disclose delusional material. Catatonia was reportedly responsive in patients given 600 to 900 mg intravenously; subjects moved freely and responded appropriately shortly after amobarbital administration.

The “amytal interview” is performed by placing the patient in a reclining position and administering amobarbital intravenously, 50 mg a minute. Infusion is continued until lateral nystagmus is sustained or drowsiness is noted, usually at 75 to 150 mg. Following this, 25 to 50 mg may be administered every 5 minutes to maintain narcosis. Interviews should be gentle, and the patient should be allowed to rest for 15 to 30 minutes after the interview before attempting to walk.

Sleep The barbiturates reduce sleep latency and the number of awakenings during sleep, though tolerance to these effects generally develops within 2 weeks. Slow wave sleep (stages 3 and 4) is shortened in general, and time spent in REM sleep is diminished. Sleep body movements decrease during barbiturate use for sleep induction. On electroencephalograms (EEG), barbiturates decrease low-frequency activity (alpha rhythm) and increase high-frequency activity (beta rhythm). This activation may evoke feelings of euphoria in some patients. Discontinuation of barbiturates for insomnia often leads to rebound increases on EEG measures.

Withdrawal From Sedative Hypnotics Barbiturates are sometimes used to determine extent of tolerance to barbiturates or other hypnotics to guide detoxification. Once intoxication has resolved, a test dose of pentobarbital (200 mg) is given orally. An hour later the patient is examined. Tolerance and dose requirement are determined by the degree to which the patient is affected. Unaffected patients have marked tolerance; patients with mild cognitive or cerebellar impairment have moderate tolerance; and patients who are put to sleep have minimal or no tolerance. If the patient is not sedated, another 100 mg of pentobarbital may be administered every 2 hours, up to three times (maximum, 500 mg over 6 hours). The amount needed for mild intoxication corresponds to the approximate daily dose of barbiturate used. Phenobarbital (30 mg) may then be substituted for each 100 mg of pentobarbital. This daily dose requirement may be administered in divided doses and gradually tapered by 10 percent a day, with adjustments made according to withdrawal signs.

PRECAUTIONS AND ADVERSE REACTIONS

Overdose Barbiturates depress the CNS. In overdose or in combination with other CNS depressants (especially alcohol or benzodiazepines), barbiturates may be lethal. Overdose may lead to respiratory depression or arrest, absent deep tendon reflexes, hypothermia, coma, and death. EEG may reveal loss of electrical activity in severe overdose, which may be reversible with detoxification. One case report noted 28 hours of isoelectric recording with subsequent full recovery.

Gastrointestinal and Nutritional Barbiturates may be toxic or lethal in patients with cirrhosis or acute intermittent porphyria and should not be used in these patient groups. Long-term barbiturate use may interfere with vitamin D metabolism and result in osteopenia or rickets. Patients may require vitamin D supplements.

Pregnancy Some evidence indicates that barbiturates cause fetal malformations, and use in the first trimester of pregnancy is contraindicated. A recent study also suggests that in utero exposure to phenobarbital is associated with intelligence deficits in adult men. Caution should be taken with barbiturates at any time in pregnancy. Barbiturate use late in pregnancy may result in neonatal dependence and a withdrawal syndrome upon delivery. Use during pregnancy has been associated with neonatal coagulopathy because barbiturates accelerate metabolism of vitamin K. Barbiturates pass into breast milk and cause CNS depression in breast-feeding newborns.

CNS Excitement Some patients exhibit paradoxical excitement rather than sedation; phenobarbital has been most often implicated. Pediatric, geriatric, or medically ill patients are at greatest risk for this phenomenon.

Addiction Potential Highly lipid soluble, short-acting agents such as pentobarbital (“yellow jackets”) or secobarbital (“red devils”) may have greater abuse potential than the other barbiturates. Barbiturates may also be used by drug addicts dependent on cocaine, opioids, or alcohol. Barbiturates should be avoided in patients with a history of alcohol or drug dependence. Moreover, because of the low therapeutic index, accidental poisoning is not uncommon. It was once thought that persons taking barbiturates took them automatically, their memory clouded by intoxication, but this seems not to be the case. Rare overdose attempts were unintentional in studies examining this so-called automatism phenomenon. These drugs should not be used in suicidal patients. Barbiturates may be lethal when used in combination with alcohol or other sedative hypnotics. Severe barbiturate addicts may use up to 2 or 3 grams of pentobarbital a day, and these patients commonly experience severe withdrawal following abrupt discontinuation, including seizures, delirium, and autonomic arousal similar to that with alcohol withdrawal.

Hypersensitivity Allergic reactions may occur, especially in patients with a history of atopy or asthma. They may include hives, wheezing, facial swelling, and (rarely) exfoliative dermatitis or the Stevens-Johnson syndrome, for which fatalities have been reported with phenobarbital. Maculopapular rashes are common, and some reports note an incidence of 1 to 3 percent among patients treated with phenobarbital.

Cardiopulmonary Rapid intravenous use of barbiturates may cause a catastrophic drop in blood pressure, even before CNS depression. Hypotension may be

reversed by dopamine. Pulmonary complications and renal failure account for most poisoning deaths. Patients with preexisting pulmonary compromise are at risk for respiratory embarrassment. Rapid intravenous use may cause laryngospasm or bronchospasm. Intravenous use may cause local irritation or thrombosis. Using large veins and administering drugs slowly may ameliorate this problem.

Hematopoietic Rarely, megaloblastic anemia, thrombocytopenia, or agranulocytosis may occur. Barbiturates may increase folate metabolism, and patients may need folate supplements.

DRUG INTERACTIONS

Use of barbiturates in conjunction with other medications may be complicated by additive effects with CNS depressants or compounds known to cause hepatic microsomal enzyme induction. Barbiturates used in combination with ethanol, opiates, benzodiazepines, or monoamine oxidase inhibitors may increase CNS depression. Drug interactions result from barbiturate-induced microsomal enzyme activity with induction related to long-term use rather than single doses or use for short periods (fewer than 7 days). Many drugs may be affected by this nonspecific increase in metabolism, which may cause loss of efficacy of the metabolized drug, though patients vary in the degree to which other agents are affected. If barbiturates are added to a drug regimen, the clinician must be alert to decreased efficacy of other compounds and make appropriate dosage adjustments. The tricyclic antidepressants carbamazepine, haloperidol (Haldol), or paroxetine (Paxil) may require dosage increases if barbiturates are added. Table 31.9-2 list drugs that may be affected by chronic barbiturate use. Plasma concentrations of other drugs that are unstudied may also be reduced. On the other hand, phenobarbital concentration may be increased by concurrent use of valproic acid (Depakene) or methylphenidate (Ritalin), and limited data suggest that fluoxetine (Prozac) and possibly other serotonin reuptake antidepressants may inhibit barbiturate metabolism. Chronic barbiturate use induces metabolism of gaseous anesthetics, and the increased metabolism may increase hepatotoxicity risk associated with enflurane or halothane.

The metabolism of the following drugs has been reported to be increased with long-term use of barbiturates. Others unlisted may also be affected
Analgesics—acetaminophen, fenoprofen
Antiarrhythmics—digitalis, lidocaine, mexiletine
Antibiotics—chloramphenicol, metronidazol, rifampin, tetracycline, griseofulvin
Anticoagulants—warfarin
Anticonvulsants—carbamazepine, phenytoin
Antidepressants—amitriptyline, desipramine, paroxetine, protriptyline
Antihypertensives—methyldopa
Antipsychotics—haloperidol, thioridazine, loxapine
β-Adrenergic receptor antagonists—labetalol, propranolol, metoprolol
Benzodiazepines—clonazepam, diazepam
Contraceptives—all containing estrogens
Immunosuppressants—corticosteroids, cyclophosphamide, cyclosporin, dacarbazine
Xanthines—aminophylline, caffeine, theophylline

Table 31.9-2 Drug Interactions

LABORATORY INTERFERENCES

Barbiturates may interfere with the metyrapone test by increasing metyrapone metabolism and decreasing the response to the metyrapone test. Bilirubin concentration may be reduced in neonates or patients treated with chronic barbiturates because barbiturates induce glucuronyl transferase, the enzyme responsible for bilirubin conjugation. Barbiturates may cause a false positive phentolamine (Regitine) test result or interfere with absorption of radioactive cyanocobalamin.

DOSAGE AND ADMINISTRATION

Dosage and routes of administration are shown in Table 31.9-3. Table 31.9-4 lists barbiturates contained in mixed preparations.

Drug	Available Preparations	Daily Dosage Range[a] (sedative[b]/hypnotic)	Anticonvulsant	Pediatric
Amobarbital	200 mg	100–200 mg/50–300 mg	65–500 mg IM	2 to 6 mg/kg up to 100 mg
Aprobarbital	40 mg/5 ml oral	40–160 mg/40–120 mg	not established	not established
Butabarbital	30 mg/5 ml oral; 15, 30, 50, 100 mg	45–120 mg/50–100 mg	not established	2–6 mg/kg (max 100 mg)
Mephobarbital	32, 50, 100 mg	64–300 mg/100 mg	200–600 mg	16–32 mg three to four times daily (1 mo–age 5); 32–64 mg three to four times daily (> age 5)
Methohexital	500 mg/50 cc (i)	0.7–1.2 mg/kg for ECT	not established	not established
Pentobarbital	30, 100 mg; 50 mg/ml (i); 20 mg/ml (e); 50, 60, 120, 200 mg (r)	60–100 mg/100–150 mg	100 mg intravenous at one minute intervals up to 500 mg	2 to 6 mg/kg up to 100 mg
Phenobarbital	8, 15, 30, 60, 100 mg; 30, 60, 65, 130 mg/ml (i); 20 mg/5 ml (e)	30–120 mg/100–300 mg	100–300 mg intravenous, up to 600 mg/day; 60–250 mg/d oral	1–3 mg/kg
Secobarbital	100 mg; 50 mg/ml (i)	100–300 mg/100 mg	5.1 mg/kg intravenous, may repeat every 1–4 hours	3–3 mg/kg

[a] Dosages are oral form unless specified: (i)=injection, (r)=rectal suppository.
[b] Sedative dosage is equal to the hypnotic dosage, but should be given 3 to 4 times daily.

Table 31.9-3 Barbiturates Dosages

Product Name	Barbiturate	Other Content
Fioricet with Codeine	Butalbital, 50 mg	Caffeine, 40 mg; codeine, 30 mg; acetaminophen, 325 mg
Fiorinal with Codeine	Butalbital, 50 mg	Caffeine, 40 mg; codeine, 30 mg, aspirin, 325 mg
Cafatine-PB	Phenobarbital, 30 mg	Caffeine, 100 mg; ergotamine tartate, 1 mg; alkaloid of belladona, 0.125 mg
Barbidonna	Phenobarbital, 32 mg	Atropine sulfate 0.025 mg; scopolamine, 0.0074 mg; hyoscyamine, 0.1286 mg
Butibel	Butabarbital, 15 mg	Belladonna extract, 15 mg
Donnatal Extentabs	Phenobarbital, 48.6 mg	Atropine sulfate 0.0582 mg; scopolamine, 0.0195; hyoscyamine, 0.311 mg
Phenerbel-S	Phenobarbital, 40 mg	Alkaloid of belladona, 0.2 mg; ergotamine tartate, 0.6 mg

Table 31.9-4 Barbiturates Contained in Combination Products

NONBARBITURATE, NONBENZODIAZEPINE SEDATIVE-HYPNOTIC AGENTS

A number of agents were widely used after the advent of the barbiturates and before the use of benzodiazepines in the treatment of anxiety and insomnia. Four such available drugs are paraldehyde, ethchlorvynol (Placidyl), meprobamate (Miltown) and glutethimide (Doriden). These drugs are not advised for use because of their low therapeutic index. Unfortunately, following the requirement of triplicate copy prescriptions for benzodiazepines in New York state, two studies found marked increases in several of these drugs.

Paraldehyde Paraldehyde is a cyclic ether, first used in 1882 as a hypnotic. It has also been used historically to treat epilepsy or delirium tremens. Because of its low therapeutic index it has been supplanted by the benzodiazepines and other anticonvulsants. Its malodorous scent is remembered by all practitioners who prescribed

it.

Chemistry The chemical name is 2,4,6 trimethyl-1,3,5-trioxane. The structure is shown in Figure 31.9-2.

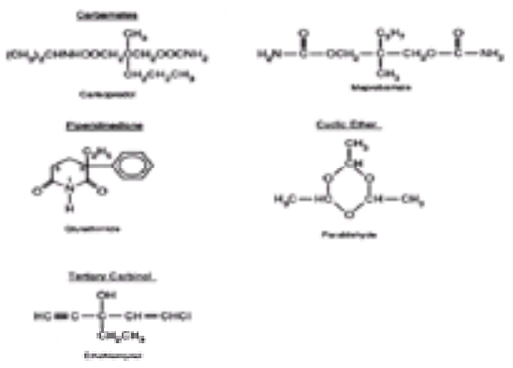

FIGURE 31.9-2 Molecular structures of other, similarly acting drugs.

Pharmacological Actions

PHARMACOKINETICS Paraldehyde is rapidly absorbed from the gastrointestinal tract and from intramuscular injections. It is primarily metabolized to acetaldehyde by the liver, and unmetabolized drug is expired by the lungs. Reported half-lives range from 3.4 to 9.8 hours. Onset of action is 15 to 30 minutes.

PHARMACODYNAMICS The mechanism of action is unknown.

Clinical Drug Studies Paraldehyde was commonly used before controlled studies were required for regulatory approval. No placebo-controlled studies exist.

Effects on Specific Organ Systems Paraldehyde may cause CNS depression, respiratory depression in the newborn, and hepatitis with long-term use. The drug passes into the fetus, but there are no studies in humans or animals regarding teratogenicity.

Therapeutic Indications Although paraldehyde is indicated by the FDA in the treatment of seizures induced by tetanus or eclampsia or status epilepticus, it should be used only when safer agents fail. Paraldehyde is not indicated as an anxiolytic or hypnotic and has little place in current psychopharmacology.

Precautions and Adverse Reactions Paradehyde frequently causes foul breath because of expired unmetabolized drug. It may inflame pulmonary capillaries and cause coughing. It may also cause local thrombophlebitis with intravenous use. Patients may experience nausea and vomiting with oral use. Overdose leads to metabolic acidosis and decreased renal output. There are no specific studies in children or elderly patients. There is risk of abuse among drug addicts.

Drug Interactions Disulfiram (Antabuse) inhibits acetaldehyde dehydrogenase and reduces metabolism of paraldehyde, leading to possible toxic concentration of paraldehyde. Paraldehyde has additive sedating effects in combination with other CNS depressants such as alcohol or benzodiazepines.

Laboratory Interferences Paraldehyde, like the other agents in this section, may interfere with the metyrapone, phentolamine, or urinary 17-hydroxycorticosteroid tests.

Dosage and Administration Paraldehyde is available in 30-mL vials for oral or intravenous use. It should not be placed in plastic containers. Unused drug exposed to the air rapidly deteriorates to acetaldehyde and acetic acid and should be discarded after 24 hours. For seizures in adults, up to 12 mL (diluted to a 10 percent solution) may be administer by gastric tube every 4 hours. For children the oral dose is 0.3 mg/kg.

Meprobamate Meprobamate, a carbamate, was introduced shortly before the benzodiazepines specifically to treat anxiety. It is uncommonly used today, though it once had great popularity. Following a triplicate benzodiazepine prescription requirement in New York, however, meprobamate use more than doubled. There is little rationale for meprobamate use in current psychopharmacology.

Chemistry The chemical name is 2-methyl-2-propyl-1,3-propanediol dicarbamate. The structure is shown in Figure 31.9-2.

Pharmacological Actions

PHARMACOKINETICS Meprobamate is rapidly absorbed from the gastrointestinal tract and from intramuscular injections. It is primarily metabolized by the liver, and a small portion is excreted unchanged in urine. Plasma half-life is approximately 10 hours.

PHARMACODYNAMICS The mechanism of action is unknown, though it appears to affect the limbic and reticular activating systems.

Clinical Drug Studies Meprobamate was approved for use before controlled clinical trials examined efficacy and safety in specific diagnoses. Meprobamate has been compared with barbiturates, diazepam (Valium), and chlordiazepoxide (Librium) for the treatment of anxiety. In mixed groups of patients with anxiety, meprobamate has only rarely fared better than placebo or barbiturates. These studies were performed before the currently accepted diagnoses of panic disorder, social phobia, obsessive-compulsive disorder, generalized anxiety disorder, or posttraumatic stress disorder, and no data exist about treating these conditions with meprobamate. One study found that patients with drug addiction preferred meprobamate to lorazepam (Ativan), suggesting that this preference would give meprobamate a high risk for abuse by patients with drug abuse histories.

Effects on Specific Organ Systems Meprobamate may cause CNS depression, especially when combined with other CNS depressants, and because it passes into breast milk, infants may be sedated by mothers using meprobamate and breast-feeding. Limited data suggest that higher rates of fetal malformations result from meprobamate use in the first trimester of pregnancy. Prolonged use may decrease saliva production and lead to dental caries, oral candidiasis, or periodontal disease. Elderly patients are more sensitive than younger adults to effects of meprobamate.

Therapeutic Indications Meprobamate is indicated for short-term treatment of anxiety disorders. This indication, however, is not based on current diagnoses or controlled trials. The drug has been supplanted by the benzodiazepines and antidepressants in the treatment of anxiety. It has been used as a hypnotic, but it is not approved for this use.

Precautions and Adverse Reactions Meprobamate may cause CNS depression and death in overdose and carries the risk of abuse by patients with drug or alcohol dependence. Abrupt cessation following long-term use may lead to a withdrawal syndrome including seizures and hallucinations. Meprobamate may exacerbate acute intermittent porphyria. Hypersensitivity reactions including wheezing and hives are uncommon. Paradoxical excitement or leukopenia is rare. It should not be used in patients with hepatic compromise.

Drug Interactions Meprobamate has additive sedating effects in combination with other CNS depressants such as alcohol, barbiturates, or benzodiazepines.

Laboratory Interferences Meprobamate, like the other agents in this section, may interfere with the metyrapone, phentolamine, or urinary 17-hydroxycorticosteroid tests.

Dosage and Administration Meprobamate is available in 200-, 400-, and 600-mg tablets as well as in a combination of aspirin, 325 mg, and 200 mg of meprobamate

(Equagesic) for oral use. For adults, the usual dosage is 400 to 800 mg twice daily. Elderly patients and children, age 6 to 12, require half the adult dose. The authors do not recommend use of meprobamate in any age group.

Ethchlorvynol Ethchlorvynol, a tertiary carbinol, was historically widely used to treat insomnia and anxiety. It has since been replaced by the benzodiazepines and other safer agents in the treatment of anxiety and sleep disturbance. It has a low therapeutic index. There is no rational use for ethchlorvynol in current psychopharmacology.

Chemistry The chemical name is 1-chloro-3-ethyl-1-penten-4-yn-3-ol. It is a liquid. The structure is shown in Figure 31.9-2.

Pharmacological Actions

PHARMACOKINETICS Ethchlorvynol is rapidly absorbed from the gastrointestinal tract. It is primarily metabolized by the liver, and a small portion is metabolized by the kidney. Metabolites are excreted in the urine. Elimination half-life is approximately 10 to 20 hours.

PHARMACODYNAMICS The mechanism of action is unknown, though it is a potent CNS depressant.

Clinical Drug Studies Ethchlorvynol was approved for use before controlled clinical trials examined efficacy and safety in specific diagnoses. It has anecdotal effectiveness for brief treatment of insomnia.

Effects on Specific Organ Systems Ethchlorvynol may cause CNS depression. Overdose may cause respiratory depression, bradycardia, or death. It passes into the placenta, and while there are no human studies of teratogenicity, animal studies suggest higher rates of stillbirth. It should not be used in pregnancy. Use late in pregnancy may result in neonatal CNS depression or a withdrawal syndrome. Higher rates of lung tumors have been reported in mice; and there are no human studies regarding carcinogenicity. Abuse of ethchlorvynol may cause pulmonary edema or effusions. Elderly patients are more sensitive than younger adults to effects of ethchlorvynol for treatment of anxiety.

Therapeutic Indications Ethchlorvynol is indicated for up to 1-week treatment of insomnia. It has been used to treat anxiety in the past, but it has no FDA indication.

Precautions and Adverse Reactions Ethchlorvynol may cause CNS depression, slurred speech, double vision, confusion, and death in overdose. There is risk of abuse by patients with drug or alcohol dependence. Abrupt cessation following long-term use may lead to a withdrawal syndrome including seizures and hallucinations. Ethchlorvynol may exacerbate acute intermittent porphyria or cause cholestatic jaundice. Hypersensitivity reactions are uncommon. Paradoxical excitement is rare. It should not be used in patients with hepatic compromise.

Drug Interactions In combination with other CNS drugs including tricyclic drugs, alcohol, barbiturates or benzodiazepines, ethchlorvynol has additive sedating effects. Ethchlorvynol stimulates hepatic microenzymes and may increase metabolism of many drugs, most notably warfarin.

Laboratory Interferences Ethchlorvynol may cause a false-positive phentolamine test result.

Dosage and Administration Ethchlorvynol is available in 200-, 500-, and 750-mg capsules. For adults, the usual dose is 500 to 750 mg at bedtime. It is best taken with food to limit rate of onset for patients in whom ataxia is of concern. Patients requiring withdrawal should be switched to a barbiturate such as phenobarbital and gradually tapered. Safety and efficacy in children or elderly patients has not been established. The authors do not recommended ethchlorvynol use in any age group.

Glutethimide

Chemistry The chemical name is 3-ethyl-3-phenyl-2,6-piperidinedione glutarimine. The structure is shown in Figure 31.9-2.

Pharmacological Actions

PHARMACOKINETICS Glutethimide is absorbed erratically from the gastrointestinal tract, is metabolized by the liver, and has an onset of action of 30 minutes. Its elimination half-life is 10 to 12 hours, and it is known to induce hepatic microsomal enzymes.

PHARMACODYNAMICS The mechanism of action is unknown, though it has potent anticholinergic activity.

Clinical Drug Studies Glutethimide has anecdotal effectiveness for brief treatment of insomnia but is not FDA approved for this use.

Effects on Specific Organ Systems Glutethimide may cause CNS depression. Overdose may cause respiratory depression, bradycardia, or death. It passes into the placenta and breast milk. There are no human or animal pregnancy studies. Use late in pregnancy may result in neonatal CNS depression or a withdrawal syndrome. There are no human studies regarding carcinogenicity. Elderly patients are more sensitive than younger adults to sedative and anticholinergic effects of glutethimide.

Therapeutic Indications Glutethimide has been used to treat insomnia but without FDA indication.

Precautions and Adverse Reactions Glutethimide may cause CNS depression, slurred speech, double vision, confusion, and death in overdose. There is risk of abuse by patients with drug or alcohol dependence. Abrupt cessation following long-term use may lead to a withdrawal syndrome including seizures and hallucinations. Glutethimide may exacerbate acute intermittent porphyria. Because of its potent anticholinergic effects it should be used cautiously in patients with prostatic hypertrophy or narrow-angle glaucoma. Hypersensitivity reactions are uncommon. Paradoxical excitement is rare. It should not be used in patients with hepatic compromise.

Drug Interactions Glutethimide has additive sedating effects in combination with other CNS drugs including alcohol, barbiturates, and benzodiazepines. Glutethimide induces hepatic microenzymes and may increase metabolism of many drugs, most notably warfarin.

Laboratory Interferences Glutethimide may cause a false-positive phentolamine test result or interfere with urine assays for 17-ketosteroids.

Dosage and Administration Glutethimide is available in 500-mg tablets. For adults, the usual dose for insomnia is 250 to 500 mg at bedtime. For elderly patients the recommended dose is 250 mg. Safety and efficacy in children has not been established. Patients requiring withdrawal should be switched to a barbiturate such as phenobarbital and gradually tapered. The authors do not recommend using glutethimide in any age group.

SUGGESTED CROSS-REFERENCES

Anxiety disorders are discussed in Section 15, substance-related disorders in Chapter 11. Chloral hydrate is covered in Section 31.14. Sleep disorders are discussed in Chapter 21.

SECTION REFERENCES

Almeyda J, Levantine A: Drug reactions. XVII. Cutaneous reactions to barbiturates, chloral hydrate and its derivatives. Br J Dermatol *86:* 313, 1972.

Breimer DD: Clinical pharmacokinetics of hypnotics. Clin Pharmacokinet *2:* 93, 1977.

Dundee JW, McIlroy PDA: The history of the barbiturates. Anaesthesia *37:* 726, 1982.

*Estes JW: The road to tranquility: The search for selective anti-anxiety agents. Synapse *21:* 10, 1995.

Haider I, Oswald I, Matthew H: E.E.G. signs of death. Br Med J *3:* 314, 1968.

Hobbs WR, Rall TW, Verdoorn TA: Hypnotics and sedatives: Ethanol. In *Goodman & Gillmans's The Pharmacological Basis of Therapeutics*, ed 9, JG Hardman, LE Limbird, PB Molinoff, RW Ruddon, A Goodman Gilman, editors. Macmillan, New York, 1996.

Johns MW: Sleep and hypnotic drugs. Drugs *9:* 448, 1975.

*Linnoila M, Erwin CW, Logue PE: Efficacy and side effects of flurazepam and a combination of amobarbital and secobarbital in insomniac patients. J Clin Pharmacol *20:* 117, 1980.

Macdonald RL, Kelly KM: Antiepileptic drug mechanisms of action. Epilepsia *36* (Suppl):S2, 1995.

Matthew H: Barbiturates. Clin Toxicol *8:* 495, 1975.

*McNutt LA, Coles FB, McAuliffe T, Baird S, Morse DL, Strogatz DS, Baron RC, Eadie JL: Impact of regulation on benzodiazepine prescribing to a low income elderly population, New York State. J Clin Epidemiol *47:* 613, 1994.

Miller LG: Herbal medicinals: Selected clinical considerations focusing on known or potential drug-herb interactions. Arch Intern Med *158:* 2200, 1998.

Nussmeier NA, Arlund C, Slogoff S: Neuropsychiatric complications after cardiopulmonary bypass: Cerebral protection by a barbiturate. Anesthesiology *64:* 165, 1986.

Pons G, Rey E, Matheson I: Excretion of psychoactive drugs into breast milk. Principles and recommendations. Clin Pharmacokinet *27:* 270, 1994.

Reinisch JM, Sanders SA, Mortensen EL, Rubin DB: In utero exposure to phenobarbital and intelligence deficits in adult men. JAMA *274:* 1518, 1995.

Roberts I, Schierhout G, Alderson P: Absence of evidence for the effectiveness of five interventions routinely used in the intensive care management of severe head injury: A systematic review. J Neurol Neurosurg Psychiatry *65:* 729, 1998.

Swanson RA, Seid LL: Barbiturates impair astrocyte glutamate uptake. Glia *24:* 365, 1998.

Tietjen CS, Hurn PD, Ulatowski JA, Kirsch JR: Treatment modalities for hypertensive patients with intracranial pathology: Options and risks. Crit Care Med *24:* 311, 1996.

Trenerry MR, Loring DW: Intracarotid amobarbital procedure. The Wada test. Neuroimag Clin North Am *5:* 721, 1995.

*United States Pharmacopeial Convention: Barbiturates. In *United States Pharmacopeial Dispensing Information–Drug Information for the Health Care Professional*, ed 19. United States Pharmacopeial Convention, Rockville, MD, 1999.

31.10 BENZODIAZEPINE RECEPTOR AGONISTS AND ANTAGONISTS

JAMES C. BALLENGER, M.D.

INTRODUCTION AND HISTORY

In the mid-1950s Roche Laboratories synthesized two compounds with anxiolytic and sedative properties, which in laboratory tests appeared to be superior to meprobamate (Miltown) and chlorpromazine (Thorazine), the principal agents available at that time. The work of Leo Sternbach and Earl Reeder led to development of the first benzodiazepine, chlordiazepoxide (Librium), in 1959. This was followed in 1963 by the release of an even more potent agent, diazepam (Valium). Since then, this class of medications has become one of the most popular in the world because of its effectiveness and safety. There are now more than three dozen benzodiazepines on the market. They have become the mainstay of treatment for anxiety, insomnia, alcohol withdrawal, and several other conditions and have largely replaced meprobamate and the barbiturates because of their increased efficacy and safety.

Various studies have documented that approximately 10 to 15 percent of the populations of developed nations have taken a benzodiazepine in the past year. However, between 2 and 8 percent have taken them on a daily basis for a month. Those using benzodiazepines appear to need them: they generally have several medical conditions as well as psychiatric conditions, are distressed and older, and appear to benefit from them. Despite worry in the general public that the benzodiazepines are overused, most patients who would probably benefit from treatment do not receive it. The evidence suggests that, in fact, undertreatment is much more frequent than overtreatment.

Some have suggested that more is known about how benzodiazepines work than perhaps any other medications used in neuropsychiatry. This area continues to evolve rapidly with new molecular biological techniques elucidating a tremendous heterogeneity of receptors involved in the action of the benzodiazepines. There is considerable promise that even more effective agents will be developed from this increased understanding.

Ligands of Benzodiazepine Receptor There are a series of ligands that interact with the g-aminobutyric acid–benzodiazepine (GABA-BZ) receptor but which have quite differing, even opposing, effects (i.e., either reducing or increasing chloride channel openings). The traditional agonists increase the frequency of the chloride ionophore openings and are anxiolytic, sedating, and anticonvulsant. Compounds with essentially the opposite action (i.e., decreasing the chloride ionophore openings and producing anxiety and convulsions) have also been recognized, and these have been termed *inverse* agonists. The third type, classified benzodiazepine antagonists, block the action of agonists and inverse agonists but have, in the main, no intrinsic activity themselves.

Other benzodiazepine direct agonists include zopiclone (Imovane), some triazolopyridazines, zolpidem (Ambien), and some b-carbolines. Although some bind to specific receptor subtypes (e.g., zolpidem), they all have the basic action of increasing the opening of the chloride channel.

Inverse agonists do the opposite and decrease the frequency of the opening of the chloride channel. They include other b-carboline derivatives and a compound named the diazepam-binding inhibitor (DBI). These agents have been observed to create anxiety in humans as well as the secondary physiological effects of anxiety (increased heart rate, blood pressure, cortisol, etc.).

The benzodiazepine antagonists block or reverse the effect of agonists or inverse agonists. The imidazobenzodiazepine derivative flumazenil (Romazicon) was the first and most widely studied benzodiazepine antagonist. In animals and humans, flumazenil can reverse all benzodiazepine actions and is marketed to reverse benzodiazepine overdose.

Partial Agonists Recently, agents that have less-functional effects, despite occupying most or all of their receptors have been termed *partial agonists*. Partial agonists have been synthesized not only for agonists, but also for inverse agonists and even antagonists. Some compounds can function as partial agonists at some receptors and full agonists at others.

The discovery of partial agonists has stimulated the search for an agent that would not be as sedating, amnestic, or prone to withdrawal symptomatology but would still have sufficient anxiolytic properties. The medication most extensively studied has been abecarnil in generalized anxiety disorder. Partial agonism may also explain certain features of the traditional benzodiazepines. Clonazepam (Klonopin) has been recognized to function as a partial agonist at certain receptors. This may explain why clonazepam use is apparently associated with fewer withdrawal difficulties.

Endogenous Ligands As with the opioids, the presence of a receptor suggests that endogenous ligands may exist for the benzodiazepine receptor. Such an endogenous compound could, theoretically, be either an anxiogenic inverse agonist or an anxiolytic agonist. Although no ligands have been found in living humans, several candidates have been put forward, including DBI, thought to have anxiogenic properties. Interestingly, desmethyldiazepam, the metabolite of diazepam and chlordiazepoxide, was isolated from brains of rats never exposed to benzodiazepines. Similarly, Sandra File and colleagues have found these same metabolites in human brains stored in the 1940s, long before the chemical synthesis of benzodiazepines two decades later. This also could have come from dietary sources, in that diazepam and lorazepam (Ativan) have been found in various foods.

BENZODIAZEPINES

Chemistry The basic structure of all benzodiazepines involves 2-aminobenzodiazepine-4-oxidates. Modifications to the structure at the R1 position increase the potency as attachments to R1 with greater electron properties increase potency (Fig. 31.10-1). Other modifications leading to drugs with differing properties have primarily been to the ring systems. Although controversial, it is thought that the fused triazolo ring in positions 1 and 2 in the diazepine ring is what confers special potency on the triazolobenzodiazepines such as triazolam (Halcion), alprazolam (Xanax), and estazolam (Prosom). In a similar fashion, a fused imidazo ring at the 1 and 2 position confers considerable potency on the preanesthetic midazolam (Versed).

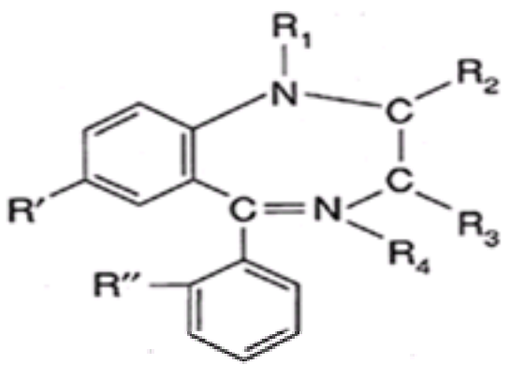

FIGURE 31.10-1 Molecular structure of benzodiazepine nucleus.

Pharmacological Actions

Pharmacokinetics

ABSORPTION AND DISTRIBUTION Benzodiazepines are readily absorbed from the gastrointestinal tract, reaching peak levels within 30 minutes to several hours. Clorazepate (Tranxene) is metabolized to its active metabolite in the stomach and then absorbed. Onset of action is primarily related to differences in lipid solubility, which affects both absorption and movement into the brain. Only lorazepam and midazolam are reliably absorbed after intramuscular injection.

Most of the benzodiazepines are highly protein bound, although this does vary somewhat. Benzodiazepines like diazepam with high lipid solubility are both quickly absorbed and quickly pass into the brain. Conversely, benzodiazepines with lower lipid solubility (e.g., lorazepam) are less rapidly absorbed and thus have a slower onset of action. Lipid solubility also obviously affects duration of action, in that their action depends not only on entry to the brain, but also clearance from the brain and other areas. Drugs with lower lipid solubility will be cleared more slowly and thus have a longer duration of action.

The varying half-lives of the benzodiazepines are also responsible for differences in duration of action and other clinical features, but this often depends more on the half-life of active metabolites. For instance, the half-life of lorazepam is 2 hours, but that of its principal metabolite, desalkyllorazepam, is more than 25 times that. Also, the principal metabolite of diazepam has a half-life of almost 100 hours. For the hypnotic flurazepam (Dalmane), this can be a negative feature in that it can lead to drowsiness the day after it is taken for insomnia. When diazepam is used for anxiety, this may be a positive feature in that missed doses are less important and discontinuation is perhaps easier. Conversely, benzodiazepines with a shorter half-life require more-frequent dosing (e.g., three or four times a day). This can be reassuring to many anxious patients but can have negative connotations for some patients and pose practical problems for others. Also, anxiety symptoms return in some patients using benzodiazepines with a shorter duration of action (e.g., alprazolam).

Most benzodiazepines cross the placental barrier and are also found in breast milk. Newborns of mothers who have been receiving benzodiazepines can be lethargic, have breathing problems, or even withdrawal symptoms postnatally.

METABOLISM Almost all benzodiazepines are metabolized primarily through the liver, and although some are excreted without being metabolized, most undergo multiple biotransformations and form active metabolites. Most are biotransformed by oxidation (also known as phase 1 metabolism), including diazepam, chlordiazepoxide, and chlorazepate, each forming multiple active metabolites. A smaller number of benzodiazepines undergo biotransformation by glucuronidation (phase 2) to inactive glucuronides or other sulfated or acetylated substances. Some agents such as diazepam are metabolized through both phase 1 and 2 processes.

The type of metabolism has some clinical importance in that patients with impaired liver function or elderly patients (at least theoretically) better tolerate those metabolized by phase 2, which involves simpler biotransformation to inactive metabolites. This is most important in elderly persons, especially those who smoke, both of which increase half-life, or those with significant hepatic impairment (e.g., alcoholics with cirrhosis).

Pharmacodynamics

Mechanism of Action Recent research has demonstrated that the action of the benzodiazopines is intimately connected to the g-aminobutyrate (GABA) system, which represents 30 percent of the cortical and thalamic inhibitory system. Specific benzodiazepine binding in the central nervous system (CNS) was first found in 1977. Further research has demonstrated that benzodiapine binding to these sites increases the activity of the GABA receptor, but only in the presence of GABA. This occurs at some, but not all, CNS GABA receptors. This has led to this receptor being referred to as the BZ-GABA receptor, although it is clear that other agents bind to this receptor as well, including picrotoxin, barbiturates, anesthetic-derivatives, alcohol, and penicillin.

Benzodiazopines and GABA both increase the propensity of the other to bind to their receptor sites. Benzodiazepines do not have any direct effect on the action of the receptor unless GABA is present. GABA's principal action is to open the chloride ionophore permitting chloride ions to move inward. Benzodiazepines increase the frequency and number of openings of the chloride channel, thus decreasing cellular excitability. This is thought to be the principal mechanism, although others have been suggested (e.g., increase in calcium-dependent potassium conductance).

The discovery of high-affinity, saturable, and stereo-specific benzodiazepine receptors by several groups in 1977 allowed the understanding of the various actions of benzodiazepine. Ataxia could be understood via the effects of GABA neurons in the cerebellum, sedation in the reticular formation, memory effects in the hippocampus, and muscle relaxant properties in the spinal cord. The potency of benzodiazepines also closely paralleled their ability to displace tritiated benzodiazepines in vitro.

The GABA-BZ receptor is an oligomeric glycoprotein with four membrane-expanding regions with 20 to 30 hydrophobic amino acids in each subunit. More recently, molecular biological studies have demonstrated at least 15 different proteins in this receptor, with at least 5 units each of 1 to 6 variants. Other studies suggest 15 subunits. Theoretically, there could be more than 500 variations in the benzodiazepine receptor, with differing pharmacology depending on the makeup of the subunits.

These receptors are localized throughout the brain, with some types clustering in one region and others in differing regions. Although some controversy remains, benzodiazepine receptors have been classified as type 1 (BZ_1) (with a high affinity for the triazolopyridazines) or type 2 (BZ_2) (with low affinities for the triazolopyridazines). Type I receptors are the most common and are found in the cerebellum and cortex. There are high numbers of type 2 receptors in the hippocampus, striatum, and spinal cord. Both groups are found in high concentrations in the cortex.

This is a highly promising area for future drug development, given the remarkable heterogeneity of the subunits. Because these receptors are also differentially localized in various parts of the CNS with differing projection patterns and differing functional effects, a variety of potential new compounds that interact with the GABA-BZ receptor is possible.

Therapeutic Indications As mentioned, because of their widespread efficacy and minimal adverse effects, as well as their rapid onset of action, the benzodiazepines have replaced predecessors for multiple indications.

Insomnia The most common use of benzodiazepines, certainly in primary care and general medicine, is for insomnia. Flurazepam is one of the most frequently prescribed hypnotics, with few adverse effects. The very short half-life benzodiazepine triazolam enjoyed transient popularity because its short half-life allowed it to be cleared before morning, and it was not associated with daytime drowsiness. However, serious concern and controversy have arisen concerning its apparent potential for serious adverse effects at higher doses (see below).

Seizures Benzodiazepines, particularly clonazepam, have anticonvulsant properties and are used for that indication. Clonazepam is used for petit mal in children.

Alcohol Withdrawal The benzodiazepines have become the treatment of choice for alcohol withdrawal in most countries, certainly in the United States. Traditionally, they are first given in higher doses (40 to 60 mg per day), and then tapered over several days.

Anxiety Disorders Generalized anxiety disorder, adjustment disorder with anxiety, but not necessarily pathological anxiety associated with life events (for example, after an accident) are the major clinical applications for benzodiazepines in psychiatry and general medical practice. Most patients should be treated for a predetermined, specific, and relatively brief period. Some patients with generalized anxiety disorder may warrant maintenance treatment with benzodiazepines. The serotonin-specific reuptake inhibitors are effective antianxiety agents that lack abuse potential, although their antianxiety effects require 2 to 4 weeks to develop.

For two anxiety disorders, panic disorder with or without agoraphobia and social phobia, the two high-potency benzodiazepines, alprazolam and clonazepam, are effective. The Food and Drug Administration (FDA) has approved the use of alprazolam for the treatment of panic disorder. The dosage guidelines for the use of alprazolam in panic disorder are similar to those for depression, as already discussed. Paroxetine (Paxil) and sertraline (Zoloft) have also been approved by the FDA for treatment of panic disorder. Because paroxetine and sertraline may not be fully effective for 2 to 4 weeks after initiation, coadministration of a high-potency

benzodiazepine for the first 2 to 4 weeks of use of paroxetine or sertraline can provide rapid control of anxiety, before the long-term benefits of paroxetine or sertraline emerge.

Benzodiazepines, especially clonazepam, which has serotonergic properties, may treat the anxiety component of obsessive-compulsive disorder. Clonazepam may be effective for certain patients who do not respond to clomipramine. Benzodiazepines may also be used to augment clomipramine or the serotonin-specific reuptake inhibitors. Benzodiazepines may help reduce hyperarousal in posttraumatic stress disorder.

Depression Unique among the benzodiazepines, alprazolam has antidepressant effects equal to those of the tricyclic drugs, but alprazolam is not effective with seriously depressed inpatients. The efficacy of alprazolam in depressive disorders may reflect its potency; the antidepressant effects of other benzodiazepines may be evident only at doses that also induce sedation or sleep.

Other Conditions They are also widely used by nonpsychiatrists to treat situational anxiety, especially that associated with medical treatment, and as muscle relaxants. Lorazepam and clonazepam are often used alone or adjunctively in the treatment of mania. The benzodiazepines (especially intramuscularly lorazepam) are used to manage both substance-induced (except amphetamine) and psychotic agitation in the emergency room. A few studies report the use of high dosages of benzodiazepines in patients with schizophrenia who had not responded to antipsychotics or who were unable to take the traditional drugs because of adverse effects. The successful use of lorazepam intramuscularly for the transient resolution of catatonia has been reported. Benzodiazepines have also been used instead of amobarbital (Amytal) for drug-assisted interviewing. Several studies have found that benzodiazepines are also effective in treating some cases of akathisia.

Precautions and Adverse Effects

Central Nervous System Certainly, the most prominent and common adverse effects of benzodiazepines are those related to their principal actions in the CNS. Almost all benzodiazepines are significantly sedating, although most patients develop rapid tolerance to this side effect within several days. Patients can experience significant daytime sleepiness, even falling asleep during daytime activities or exhibiting other signs of overdosage (e.g., ataxia or slurred speech). The ataxia associated with the benzodiazepines has particular importance for elderly persons, who are more at risk for overdose because of the reduced metabolic rate associated with aging. This has particular importance with the increased risks in elderly persons for ataxia, falls, and hip fractures. In a related fashion, younger individuals have always been cautioned about operating heavy machinery during benzodiazepine use. More recently, benzodiazepine interference with the ability to drive automobiles has been studied. This bears close following, as research suggests significant impairment of driving ability.

Certainly, the cognitive effects of benzodiazepines have been and continue to be an important issue. The anterograde amnestic effects of benzodiazepines are used widely in anesthesia induction to produce anterograde amnesia. Intravenous administration appears to be required to produce this effect, but it also has been reported with oral dosing, especially with triazolam. This effect may be increased if alcohol is also consumed.

Probably more important are the observations that long-term benzodiazepine therapy interferes with concentration and memory of new material. However, it does appear that the ability to remember and recall information learned prior to benzodiazepine therapy is not compromised. Although this can be a particular problem in elderly patients and may even lead to confusional, delirious, and even pseudodementia-type pictures, these effects on memory are generally subtle. These adverse effects, in particular, must be monitored with each individual patient to balance these relatively minor memory disturbances (if they occur) with clinical efficacy.

An important issue has been the use of benzodiazepines to induce amnesia in date rape. Although it is generally believed that flunitrazepam (not sold legally in the United States) is especially effective in inducing amnesia, all full agonist benzodiazepines are, essentially, equally capable of inducing anterograde amnesia.

A further issue is whether behavioral disinhibition occurs with the benzodiazepines. This has been reported as an increase in anxiety and, more commonly, in hostility, impulsive thefts, acts of aggression, and (particularly) verbal hostility. It is unclear how common these reactions are, and some patients believably describe that they had meant to say certain angry things for many years, but were too anxious to do so. Therefore, it is unclear whether this represents disinhibition or not.

Rarely, incontinence has been reported with patients receiving benzodiazepines. Also, certain benzodiazepines, especially at higher doses, have been associated with sexual dysfunction (primarily decreased libido or inability to maintain an erection) and are common enough that clinicians should inquire about changes in sexual function.

Certainly, intravenous benzodiazepines can lead to pulmonary depression or even arrest. This can even occur with oral doses in patients with chronic obstructive pulmonary disease or sleep apnea.

There is some concern that benzodiazepine use is associated with the emergence of depression. This is theorized to occur by either direct action of the drug or by treating anxiety in a mixed anxiety-depressive disorder and thereby uncovering the depression.

Flumazenil has been associated with seizures in epileptic patients who have been treated with benzodiazepine anticonvulsants, in patients on long-term treatment with benzodiazepines, and in patients who have also overdosed with tricyclic drugs. Occasionally, cardiac arrhythmias have also been observed with flumazenil use.

Abuse Liability In this author's experience, concern over the so-called abuse liability of the benzodiazepines in the general public and even the medical profession is the most controversial area of clinical medicine. Despite considerable scientific evidence that the risk of drug abuse with benzodiazepines is low, there is tremendous prejudice against their use in many individuals, in certain treatment settings, and even countries (e.g., United Kingdom). This area has had excellent review, and it is actually much more common for anxious patients to take lower doses than prescribed, and routinely they do not escalate dosage, even over prolonged use. Dose increases are generally appropriate and are not an indication of abuse.

Despite this research literature and clinical experience, many people and certainly the media believe that benzodiazepines are widely and routinely associated with abuse. This seems to stem from at least two sources. First, there is a general belief that patients who have abused other drugs are at high risk for abusing benzodiazepines. Although some data support this view and there are certainly patients who abuse benzodiazepines in context with other illegal drugs of abuse or even by themselves, this is a controversial area. Second, conventional wisdom suggests that benzodiazepine use by alcoholics be generally avoided. However, recent analyses of even these data suggest that abuse of benzodiazepines (even by alcoholic individuals) may not be as certain in all alcoholics as believed.

Certainly, primary abuse of benzodiazepines does occur. Abuse of oral benzodiazepines has been relatively infrequent, but there has been a sharp increase in other abuse of benzodiazepines, particularly in Britain. Temazepam (Restoril) is being increasingly abused through intravenous use, with extraordinary vascular side effects, even resulting in loss of limbs, and some evidence suggests that flunitrazepam has been abused by heroin addicts more than other benzodiazepines.

Confusion over the meaning of the terms *addiction, dependence*, and *abuse* has significantly contributed to the continued confusion and controversy in this area. Recent efforts to clarify the semantic confusion, especially differentiating the routine withdrawal symptom liability observed in 30 to 90 percent of patients from terms such as *addiction* and *abuse* have been helpful.

Pregnancy Complications from use of benzodiazepines during pregnancy are somewhat unclear. However, some data suggest palate abnormalities; although in general, use during pregnancy is relatively safe if clinical issues outweigh the risks. If a mother is taking a benzodiazepine at birth, children can have difficulties with the drug in their bloodstream.

Discontinuation of Treatment One of the more important areas of benzodiazepine use centers around the discontinuation of effective or ineffective treatment. Pressure to discontinue often arises from the patient, patient's family, or even the medical professional. This is often related to exaggerated fears of becoming addicted or an attitude that use of benzodiazepines is a sign of weakness rather than a treatment for a medical condition. There are some good reasons to discontinue effective pharmacotherapy, including determining if the patient still needs the medication as mentioned above. Others include serious adverse effects or the wish to conceive a child.

Successful discontinuation involves several activities that although simple are often not performed. Patients and their families need to be told what to expect with discontinuation, because it is often associated with anxiety before (and certainly if) withdrawal symptomatology occurs.

Withdrawal symptoms occur in 30 to 90 percent of patients, depending on the definition of symptoms, length and dosage of previous treatment, and (most importantly) the rate of taper. Most of the symptoms are mild to moderate and easily tolerated, although abrupt discontinuation of a relatively high dose of benzodiazepines has been (rarely) associated with convulsions. Although benzodiazepines, particularly those with a long half-life, can be abruptly discontinued, this is not recommended in any case. In fact, recent clinical experience suggests that the withdrawal syndrome can be largely avoided with an appropriately slow taper. In generalized anxiety disorder patients, this might well be 3 to 6 weeks, and in panic disorder patients 8 to 24 weeks. This is less necessary with shorter periods of treatment or low doses of benzodiazepines.

When patients discontinue taking benzodiazepines, several types of symptoms occur. First, patients can relapse with return of their original symptoms. Second, patients can experience withdrawal symptoms, which generally occur within several days of discontinuation but generally decrease and disappear within 2 to 3 weeks. With high-potency benzodiazepines, withdrawal symptoms can be observed even with decreases from one dosage level to another. *Rebound* refers to the sudden onset of more severe symptoms than those in the original anxiety syndrome within days of decreasing or discontinuing benzodiazepine use. These generally clear very rapidly. *Relapse* refers to the return of the original anxiety syndrome. Since all of these syndromes overlap considerably and are often difficult to distinguish, the term *relapse* is generally reserved for stable or increasing symptoms present 3 to 4 weeks after medication use is discontinued that represent return of the original syndrome.

Proper management of discontinuation includes educating patients and their family, support, easy availability of the clinician, providing other strategies for dealing with anxiety or sleep difficulties, and the passage of time. Some recent studies suggest that formal cognitive-behavioral group treatment is also helpful in helping patients discontinue long-term benzodiazepine use (Table 31.10-1).

Anxiety
Irritability
Insomnia
Fatigue
Headache
Muscle twitching or aching
Tremor, shakiness
Sweating
Dizziness
Concentration difficulties
Nausea, loss of appetite*
Observable depression*
Depersonalization, derealization*
Increased sensory perception (smell, sight, taste, touch)*
Abnormal perception or sensation of movement*

* Symptoms likely to represent true withdrawal, rather than an exacerbation or return of original anxiety.

Reprinted with permission from Roy-Byrne, PP, Hommer D: Benzodiazepine withdrawal: Overview and implications for the treatment of anxiety. Am J Med 84(3)(1A), 1988.

Table 31.10-1 Commonly Observed Withdrawal Symptoms (Benzodiazepine Withdrawal Syndrome)

DRUG INTERACTIONS

The most common drug-drug interaction occurs when benzodiazepines are combined with other CNS depressants (e.g., alcohol, barbiturates, opioids and antihistamines). Patients should be cautioned about potential potentiation of the depressant effects (Table 31.10-2).

Decrease absorption		
Antacids		
Increase CNS depression		
Antihistamines		
Barbiturates and similarly acting drugs		
Cyclic antidepressants		
Ethanol		
Increase benzodiazepine levels (compete for microsomal enzymes; probably little or no effect on lorazepam, oxazepam, temazepam)		
Cimetidine	Erythromycin	Fluoxetine
Disulfiram	Estrogens	Isoniazid
Decrease benzodiazepine levels		
Carbamazepine (possibly other anticonvulsants)		

Reprinted with permission from Arana GW, Hyman SE: *Handbook of Psychiatric Drug Therapy*, ed 2. Little, Brown, Boston, 1991.

Table 31.10-2 Interactions of Benzodiazepines With Other Drugs

Cimetidine (Tagamet) and disulfiram (Antabuse) decrease metabolism of benzodiazepines, prolonging their effects. This can be clinically important with estrogen and oral contraceptives. Diazepam and gallamine or succinylcholine can result in paralysis. Fluvoxamine (Luvox) and other inhibitors of the cytochrome P450 enzyme CYP 3A4 can also inhibit metabolism of benzodiazepines and greatly increase their effects. The half-life of digoxin (Lanoxin) is increased by benzodiazepines by an unknown mechanism. With the resurgence of tuberculosis, particularly in acquired immune deficiency syndrome (AIDS)-related problems, antibiotics can also affect benzodiazepine metabolism. Rifampin (Rifadin) increases hepatic metabolic enzymes, and isoniazid decreases them.

Relationship of Blood Concentration to Clinical Efficacy The clinical effects of the benzodiazepines depend on the time course of receptor occupancy, which itself depends on dose and plasma concentration. However, controversy continues about whether significant correlations exist between dose and plasma concentration and clinical effects, because research results are mixed. Several early studies failed to find a correlation, although others did find positive correlations between plasma concentration and clinical improvement.

More recently, positive correlations between alprazolam plasma concentration and clinical response were observed in panic disorder patients. Although there was wide variability between patients given identical doses of alprazolam, reductions in panic and phobic symptomatology and drug concentration in blood were observed. Adverse effects also increased with increased plasma concentration at a certain point, and both studies suggested a range of plasma concentration that were associated with positive clinical response and relatively few adverse effects.

Dosage and Administration Table 31.10-3 summarizes the benzodiazepine receptor agonists and antagonists and presents dosage and preparations.

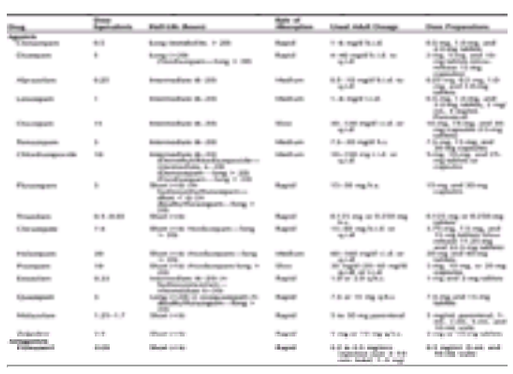

Table 31.10-3 Half-Lives, Doses, and Preparations of Benzodiazepine Receptor Agonists and Antagonists

The starting dosage of alprazolam for the treatment of depression should be 0.5 mg two or three times a day and should be raised in 0.5 mg a day intervals every 3 or 4 days. The maximal dosage is usually 4 mg a day, although some investigators and clinicians have used dosages as high as 10 mg a day. The use of high dosage is controversial because of the possibility of withdrawal symptoms. Clinicians must taper, rather than abruptly stop, alprazolam use, usually at the rate of 0.5 mg a day every 3 to 4 days or slower.

Length of Treatment Much of the formal research with benzodiazepines has involved short-term use, and they are generally labeled as such. However, accumulated clinical experience and recent research has demonstrated that, in fact, most anxiety conditions are quite long term. The long-term nature of panic disorder has recently been established. In the early trials when benzodiazepines were discontinued blindly after 6 weeks, approximately half of the patients relapsed with return of anxiety. It would seem prudent for clinicians to taper and try to discontinue benzodiazepines in generalized anxiety disorder patients after 6 weeks to see whether they can remain off medications. Current practice with panic disorder patients is to continue treatment for 6 to 18 months before tapering and attempting discontinuation. Although longer treatment has become commonplace in anxiety, benzodiazepine use in the treatment of insomnia should be restricted to the short term in most cases.

ZOLPIDEM

Zolpidem is a hypnotic that acts at the g-aminobutyric acid (GABA)-benzodiazepine complex as the benzodiazepines do, but it is not itself a benzodiazepine. The only indication for zolpidem at this time is as a hypnotic. The drug lacks the muscle-relaxant effects that are common to the benzodiazepines. Zolpidem is an imidazopyridine, and its chemical structure is shown in Figure 31.10-2.

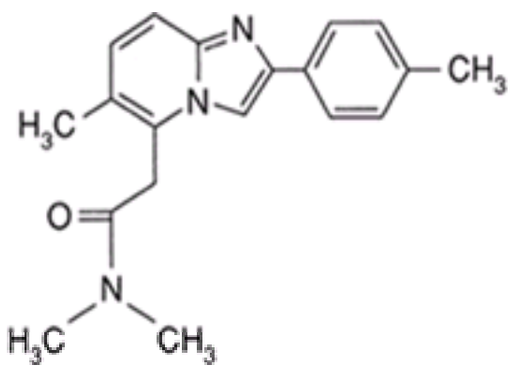

FIGURE 31.10-2 Molecular structure of zolpidem.

Pharmacological Actions Zolpidem is rapidly and well absorbed after oral administration, and it reaches peak plasma levels in about 2 to 3 hours. Zolpidem has a half-life of about 2½ hours and is metabolized primarily by conjugation. Zolpidem does not have any active metabolites. Zolpidem has a much higher affinity for BZ_1 receptors than for BZ_2 receptors. Zolpidem is also more specific for the CNS benzodiazepine receptors than for the peripheral benzodiazepine receptors. These pharmacodynamic properties are consistent with the drug's efficacy as a hypnotic in the absence of significant anticonvulsant or muscle-relaxant properties. The binding site of zolpidem is likely to be similar to that of the benzodiazepines, as the effects of zolpidem can be prevented or reversed by the benzodiazepine receptor antagonist flumazenil.

Therapeutic Indications The sole indication at the time for zolpidem is as a hypnotic. Several studies have found an absence of rebound REM after the use of the compound for the induction of sleep. The comparatively few data available indicate that zolpidem may not be associated with rebound insomnia after the discontinuation of its use for short periods.

Precautions and Adverse Reactions Because of the short half-life of zolpidem, clinicians may reasonably evaluate a patient for the possibility of anterograde amnesia and anxiety the day after its administration, although neither of these adverse effects has been reported. Emesis and dysphoric reactions have been reported as adverse effects. Tolerance and dependence have been reported in less than 1 percent of patients, and the withdrawal symptoms are similar to those described for benzodiazepines. Zolpidem is secreted in breast milk and is, therefore, contraindicated for use by nursing mothers. The dosage of zolpidem should be reduced in patients with renal and hepatic impairment.

Drug Interactions and Laboratory Interferences Information on drug interactions and laboratory interferences is limited. Therefore, clinicians should consider the possibility of such an interaction or interference in a patient who is being treated with zolpidem.

Dosage and Administration A single 10 mg dose is usual for the treatment of insomnia. For patients under age 65, the initial dose of 10 mg can be increased to 15 to 20 mg if necessary. For patients over age 65, an initial dose of 5 mg may be advised. Prolonged use of zolpidem or any hypnotic is not recommended.

FLUMAZENIL

Flumazenil is a benzodiazepine receptor antagonist. It reverses the psychophysiological effects of the benzodiazepine agonists (for example, diazepam). The use of flumazenil is limited to emergency rooms and other emergency settings. The molecular structure of flumazenil is based on a benzodiazepine nucleus (Fig. 31.10-3).

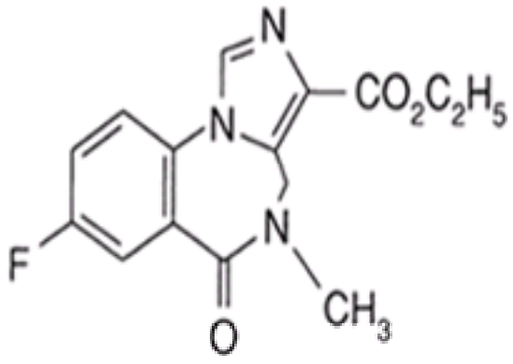

FIGURE 31.10-3 Molecular structure of flumazenil.

Pharmacological Actions After intravenous administration, flumazenil has a half-life of 7 to 15 minutes. Protein binding is about 50 percent. Clearance of flumazenil occurs primarily by hepatic metabolism. The major metabolites of flumazenil are the deethylated free acid and its glucuronide conjugate, which are excreted in the urine. Elimination of the drug is essentially complete in 72 hours. The pharmacokinetics of flumazenil are not significantly affected by gender, age, renal failure, or hemodialysis but are affected by hepatic impairment, which prolongs the half-life.

Flumazenil antagonizes sedation, impairment of recall, and psychomotor retardation produced by benzodiazepine receptor agonists. But flumazenil does not reverse the effects of other CNS depressants, even if they also act partly on the $GABA_A$ receptor (for example, ethanol and barbiturates). Flumazenil is also ineffective in

reversing the effects of opioids.

Therapeutic Indications Flumazenil is used to reverse the effects of benzodiazepine receptor agonists that have been used for clinical indications (for example, sedation and anesthesia) or in overdose.

Precautions and Adverse Reactions The most common adverse effects of flumazenil are nausea, vomiting, dizziness, agitation, emotional lability, cutaneous vasodilation, injection-site pain, fatigue, impaired vision, and headache. The most common serious adverse effect of the use of flumazenil to reverse benzodiazepine overdose is the precipitation of seizures, which is especially likely to occur in patients with seizure disorders, those who are physically dependent on benzodiazepines, or those who have ingested very large quantities of benzodiazepines. Flumazenil alone may impair memory retrieval.

Drug Interactions and Laboratory Interferences No deleterious drug interactions have been noted when flumazenil is administered after narcotics, inhalation anesthetics, muscle relaxants, and muscle-relaxant antagonists administered in conjunction with sedation or anesthesia. In mixed-drug overdose the toxic effects (for example, seizures and cardiac arrhythmias) of other drugs (for example, tricyclic drugs) may emerge with the reversal of the benzodiazepine effects of flumazenil. For example, seizures caused by an overdose of tricyclic drugs may have been partially treated in a patient who had also taken an overdose of benzodiazepines. With flumazenil treatment, the tricyclic-induced seizures or cardiac arrhythmias may appear and may result in a fatal outcome.

No laboratory interferences have been associated with the use of flumazenil.

Dosage and Administration For the initial management of a known or suspected benzodiazepine overdose, the recommended initial dose of flumazenil is 0.2 mg (2 mL) administered IV over 30 seconds. If the desired level of consciousness is not obtained after waiting 30 seconds, a further dose of 0.3 (3 mL) can be administered over 30 seconds.

Most patients with a benzodiazepine overdose respond to a cumulative dose of 1 to 3 mg of flumazenil; doses beyond 3 mg of flumazenil do not reliably produce additional effects. If a patient has not responded 5 minutes after receiving a cumulative dose of 5 mg flumazenil, the major cause of sedation is probably not due to benzodiazepine receptor agonists, and additional flumazenil is likely to have no effect.

SUGGESTED CROSS-REFERENCES

Disorders for which benzodiazepine therapy may be considered are discussed in Section 11.2 on the treatment of alcohol-related disorders, Chapter 15 on anxiety disorders, and Chapter 21 on sleep disorders. Section 3.1 takes up perception and cognition; Section 3.4 is devoted to the biology of memory. Section 31.12 discusses buspirone, and Section 31.25 discusses the selective serotonin reuptake inhibitors.

SECTION REFERENCES

Ballenger JC: Long-term pharmacologic treatment of panic disorder. J Clin Psychiatry *52:* 18, 1991.

*Ballenger JC, Fyer AJ: Examining criteria for panic disorder. Hosp Community Psychiatry *44:* 226, 1993.

Balter MB, Levine J, Manheimer DI: Cross-national study of the extent of anti-anxiety/sedative drug use. N Engl J Med *290:* 769, 1974.

Bennett JA, Moioffer M, Stanton SP, Dwight M, Keck PE Jr: A risk-benefit assessment of pharmacological treatments for panic disorder. Drug Safety *18:* 419, 1998.

Blanchard JC, Boireau A, Garret C, Julou L: In vitro and in vivo inhibition of zopiclone of benzodiazepine binding to rodent brain receptors. Life Sci *24:* 2417, 1979.

*Carpenter T Jr, Buchanan W, Kirkpatrick B, Breier AF: Diazepam treatment of early signs of exacerbation in schizophrenia. Am J Psychiatry *156:* 299, 1999.

Chudnofsky C: Safety and efficacy of flumazenil in reversing conscious sedation in the emergency department. Acad Emerg Med *4:* 944, 1997.

Ciraulo DA, Sands BF, Shader RI: Critical review of liability for benzodiazepine abuse among alcoholics. Am J Psychiatry *145:* 1501, 1988.

Cohn JB, Wilcox CS: Long-term comparison of alprazolam, lorazepam and placebo in patients with an anxiety disorder. Pharmacotherapy *4:* 93, 1984.

Costa E, Guidotti A: Endogenous ligands for benzodiazepine recognition sites. Biochem Pharmacol *34:* 3399, 1985.

DeRobertis E, Pena C, Paladini AC, Medina JH, de Stein ML, DeRobertis E: New developments in the search for endogenous ligand(s) of central benzodiazepine receptors. Neurochem Int *13:* 1, 1988.

Dorow R, Horowski R, Paschelke G, Amin M: Severe anxiety induced by FG 7142, a beta-carboline ligand for benzodiazepine receptor function. Lancet *2:* 98, 1983.

File SE: The benzodiazepine receptor and its role in anxiety. Br J Psychiatry *152:* 599, 1988.

Flint AJ: Epidemiology and comorbidity of anxiety disorders in later life: Implications for treatment. Clin Neurosci *4:* 31, 1997.

Greenblatt DJ, Divoll M, Abernethy DR, Ochs HR, Shader RI: Clinical pharmacokinetics of the newer benzodiazepines. Clin Pharmacokinet *8:* 233, 1983.

Greenblatt DJ, Harmatz JS, Shader RI: Plasma alprazolam concentrations: Relation to efficacy and side effects in the treatment of panic disorder. Arch Gen Psychiatry *50:* 715, 1993.

Haefely W: The biological basis of benzodiazepine actions. In *The Benzodiazepines: Current Standards for Medical Practice*, DE Smith, DR Wesson, editors. MTP Press, Hingham, MA, 1985.

Kunovac JL, Stahl SM: Future directions in anxiolytic pharmacotherapy. Psychiatr Clin North Am *18:* 895, 1995.

Lader M: Clinical pharmacology of anxiolytic drugs: Past, present and future. In *GABA Receptors and Anxiety: From Neurobiology to Treatment*, G Biggio, E Sanna, E Costa, editors. 1995.

Lesser IM, Lydiard RB, Antal E, Rubin RT, Ballenger JC, DuPont R: Alprazolam plasma concentrations and treatment response in panic disorder and agoraphobia. Am J Psychiatry *149:* 1556, 1992.

Luddens H, Korpi ER: Biological functions of $GABA_A$ benzodiazepine receptor heterogeneity. J Psychiatry Res *29:* 77, 1995.

Lydiard RB, Laraia M, Ballenger JC, Howell EF: Emergence of depressive symptoms in patients receiving alprazolam for panic disorder. Am J Psychiatry *144:* 664, 1987.

Malizia A, Nutt DJ: Psychopharmacology of benzodiazepines—an update. Hum Psychopharm *10* (Suppl):1, 1995.

Medina JH, Viola H, Wolfman C, Marder M, Wasowski C, Calvo D, Paladini AC: Overview—flavonoids: A new family of benzodiazepine receptor ligands. Neurochem Res *22:* 419, 1997.

Mohler H, Okada T: Benzodiazepine receptors: Demonstration in the central nervous system. Science *1989:* 849, 1977.

Norman TR, Ellen SR, Burrows GD: Benzodiazepines in anxiety disorders: Managing therapeutics and dependence. Med J Aust *167:* 490, 1997.

Nowell P, Mazumdar S, Buysse D, Dew M, Reynolds C, Kupfer D: Benzodiazepines and zolpidem for chronic insomnia: A meta-analysis of treatment efficacy. JAMA *278:* 2170, 1997.

*Nutt D: Selective ligands for benzodiazepine receptors: Recent developments. In *Current Aspects of the Neurosciences*, NN Osborne, editor. Macmillan, New York, 1989.

Patterson DR, Ptacek JT, Carrougher GJ, Sharar SR: Lorazepam as an adjunct to opioid analgesics in the treatment of burn pain. Pain *72:* 367, 1997.

Pecknold JC: A risk-benefit assessment of buspirone in the treatment of anxiety disorders. Drug Safety *16:* 118, 1997.

Reddy DS, Kulkarni SK: Neurosteroid coadministration prevents development of tolerance and augments recovery from benzodiazepine withdrawal anxiety and hyperactivity in mice. Methods Find Exp Clin Pharmacol *19:* 395, 1997.

*Rickels K, Case WG, Schweizer E, Garcia-Espana F, Fridman R: Long-term benzodiazepine users 3 years after participation in a discontinuation program. Am J Psychiatry *148:* 757, 1991.

*Romach MK, Somer GR, Sobell LC, Sobell MB, Kaplan HL, Sellers EM: Characteristics of long-term alprazolam users in the community. J Clin Psychopharmacol *12:* 316, 1992.

Rothschild AJ: Disinhibition, amnestic reactions, and other adverse reactions secondary to triazolam: A review of the literature. J Clin Psychiatry *53:* 69, 1992.

*Shader RI, Greenblatt DJ: Use of benzodiazepines in anxiety disorders. N Engl J Med *328:* 1398, 1993.

Sieghart W, Eichinger A, Riederer P, Jellinger K: Comparison of benzodiazepine receptor binding in membranes from human or rat brain. Neuropharmacology *24:* 751, 1985.

Squires RF, Braestrup CL: Benzodiazepine receptors in rat brain. Nature *266:* 732, 1977.

Study RE, Barker JL: Diazepam and pentobarbital: Fluctuation analysis reveals different mechanisms for potentiation of gamma-aminobutyric acid responses in cultured central neurons. Proc Natl Acad Sci USA *78:* 7180, 1981.

Woods J, Winger G: Abuse liability of flunitrazepam. J Clin Psychopharmacol *17* (Suppl):1S, 1997.

Young WS, Kuhar MJ: Radiohistochemical localization of benzodiazepine receptors in rat brain. J Pharmacol Exp Ther *212:* 337, 1980.

Zorumski CF, Isenberg KE: Insights into the structure and function of GABA-benzodiazepine receptors: Ion channels and psychiatry. Am J Psychiatry *148:* 162, 1991.

31.11 BUPROPION

ROBERT N. GOLDEN, M.D., AND LINDA M. NICHOLAS, M.D.

HISTORY

Thirty years ago, a group of pharmacologists decided to develop a novel antidepressant compound. They devised a research plan in which they would create a molecule that was active in conventional antidepressant screening models yet differed from the current available antidepressants in five specific ways: (1) it would possess different chemical, pharmacological, and biochemical properties from tricyclic drugs; (2) it would not inhibit monoamine oxidase (MAO) activity; (3) it would not possess sympathomimetic activity; (4) it would not exhibit anticholinergic effects; and (5) it would not depress myocardial activity. The group believed that the first two of these properties would enhance their chances of developing a unique approach to treating depression, while the other three properties would evoke an adverse-effect profile that was far superior to that of existing antidepressant medications. Their approach led to the discovery of bupropion (Wellbutrin).

As predicted, early clinical trials demonstrated that bupropion had efficacy comparable to that of conventional antidepressants and an apparently more benign side-effect profile, with a striking lack of significant anticholinergic toxicity and minimal effect on cardiovascular systems. However, shortly after the drug was shipped to pharmacies in 1986, seizures were reported in a small number of nondepressed, bulimia nervosa patients treated with bupropion. The manufacturer immediately withdrew the drug and launched extensive clinical investigations focusing on bupropion's effects on seizure threshold. Finally, in 1989, bupropion was released for clinical use in this country. In 1996, a slow-release preparation (Wellbutrin-SR) became available for clinical use, and bupropion (Zyban) was recognized by the Food and Drug Administration as a safe and effective aid, in combination with a behavioral modification program, in smoking cessation treatment.

CHEMISTRY

Bupropion is a unicyclic aminoketone (Fig. 31.11-1), with a chemical structure that is unique among antidepressants. The lack of certain common functional groups (e.g., *N*-methylpiperazine) often found in neuroleptics and the absence of complex heterocyclic fused rings probably contribute to bupropion's lack of the type of adverse-effect profile often seen in tricyclic and tetracyclic drugs. Bupropion's chemical structure is similar to that of psychostimulants, including amphetamines and diethylpropion (Tenuate), which may account for certain shared characteristics.

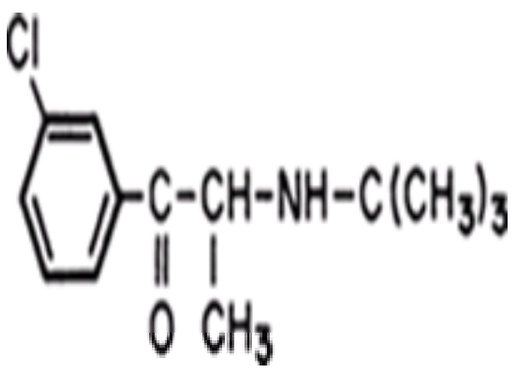

FIGURE 31.11-1 Molecular structure of bupropion.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption The bupropion is rapidly absorbed following oral administration, with peak blood concentrations reached within 2 hours. Mean protein binding in healthy subjects is 85 percent. The elimination is biphasic, with an initial phase of approximately 1.5 hours and a second phase of about 14 hours.

The sustained-release formulation of bupropion is bioequivalent to the immediate-release formulation with respect to the parent compound, active metabolites, and a pharmacological activity–weighted composite. Peak plasma concentrations of bupropion are reached approximately 3 hours after oral administration of the sustained-release form, followed by a biphasic decline, that is less pronounced than that seen following administration of the immediate-release formulation. At equal doses, the total amount of bupropion absorbed from the immediate-release and sustained-release formulations is about the same. Peak and trough steady-state plasma concentrations of bupropion following oral administration of 150 mg of the sustained-release formulation twice a day are about 85 and 107 percent respectively, of the concentrations achieved following 100 mg of the immediate-release formulation thrice daily. Bupropion areas under the concentration time curve (AUCs), peak plasma concentration, and AUCs for all three of the detectable active metabolites are equivalent for the immediate-release and sustained-release formulations.

Distribution Protein binding for bupropion ranges from 82 to 88 percent at concentrations of 84 to 239 ng/mL. It is widely distributed to tissues. The cerebrospinal fluid to plasma ratio of bupropion is 0.43, and the ratios for the three active metabolites range from 0.04 to 0.50. Bupropion crosses the placental barrier, although fetal tissue concentrations are lower than those in placental and other maternal tissues. Bupropion is excreted in breast milk.

Metabolism and Elimination Bupropion undergoes extensive hepatic metabolism, including a pronounced first-pass effect. At steady state, the three active metabolites, hydroxybupropion, erythyrohydrobupropion, and threohydrobupropion, predominate over the parent compound in plasma and cerebrospinal fluid and may play an important role in determining clinical response. Less than 1 percent of the drug is excreted unchanged, and approximately 87 and 10 percent of radioactively labeled drug dose is recovered in the urine and feces, respectively.

Pharmacodynamics

Mechanism of Action The mechanism of action of bupropion remains unclear. Early investigations focused on enhanced dopaminergic activity because of reuptake inhibition by the parent compound as a likely mechanism of action. Some of the behavioral effects of bupropion in classical animal models of antidepressant activity, including stimulation of locomotor activity and effects on the Porsolt forced swim test, are abolished by destruction of dopaminergic neurons by 6-hydroxydopamine. Also, bupropion-induced behavioral sensitization in the rat is accompanied by a selective potentiation of the effects of the compound on interstitial dopamine concentrations in the nucleus accumbens. In a clinical study, plasma concentrations of the major metabolite of dopamine, homovanillic acid, increased in patients who

failed to respond to bupropion treatment but not in those who achieved clinical response.

More recently, the role of norepinephrine in the mechanism of action of bupropion has received considerable attention. In hospitalized depressed patients, bupropion, along with the norepinephrine reuptake inhibitor desipramine and MAO inhibitors, enhanced noradrenergic function, as measured by 24-hour excretion of 6-hydroxymelatonin, a physiological gauge of noradrenergic activity. At the same time, these antidepressants reduced “whole body norepinephrine turnover,” that is, the 24-hour excretion of norepinephrine and its metabolites. These observations suggest that bupropion increases the functional efficiency of noradrenergic systems. This effect may be mediated by bupropion's active metabolites, including hydroxybupropion, which inhibit the reuptake of norepinephrine into rat cortical tissue.

The three known active metabolites, hydroxybupropion, erythyrohydrobupropion, and threohydrobupropion, may play important roles in the mechanism of bupropion's antidepressant activity. In animal models, hydroxybupropion possesses more-potent antidepressant properties than the parent compound, and threohydrobupropion also demonstrates some antidepressant activity. In depressed patients, the steady-state plasma and cerebrospinal fluid concentrations of the active metabolites actually exceed those of bupropion itself.

Blood Levels and Relation to Action No consistent, clear relationship between steady-state bupropion plasma concentrations and clinical response has emerged. One study reported a curvilinear relation between antidepressant efficacy and trough bupropion plasma concentrations, with greatest efficacy associated with bupropion plasma concentrations between 25 and 100 ng/mL. In contrast, another study found greater clinical response associated with trough levels below 30 ng/mL. A study examining both parent and metabolite concentrations found no relation between steady-state plasma bupropion concentrations and clinical response, but a significant relation between metabolite concentrations and clinical outcome. For each of the three active metabolites, higher concentrations were significantly related to poor clinical outcome; this effect was most striking for hydroxybupropion. The interpretation of this observation is unclear. It may indicate toxic effects that emerge at higher metabolite concentrations. On the other hand, the metabolites may be responsible for bupropion's therapeutic effects, and a curvilinear plasma concentration-response relation (like that seen with nortriptyline) may exist.

CLINICAL DRUG STUDIES

Double-blind, placebo-controlled studies have clearly demonstrated that bupropion is superior to placebo and comparable to conventional tricyclic drugs in the treatment of major depressive disorder. In four multicenter studies, bupropion was consistently superior to placebo in treating depressed outpatients and inpatients, with 60 to 70 percent efficacy, compared with 30 percent with placebo. At the same time, bupropion's side-effect profile did not differ substantially from that associated with placebo treatment, with minimal cardiovascular changes. In controlled comparisons to standard antidepressants, including amitriptyline (Elavil), imipramine (Tofranil), doxepin (Sinequan), and nortriptyline (Aventyl), bupropion has generally demonstrated similar efficacy. When a sedating tricyclic drug such as doxepin or amitriptyline is used as the gold standard treatment, greater or more rapid improvement in sleep or both can yield more impressive results in the insomnia component of depression rating scales. On the other hand, in such comparisons bupropion appears superior in terms of diminished anticholinergic adverse effects and unwanted weight gain.

A few studies have compared bupropion with other second-generation antidepressants. In a two-center double-blind study, outpatients with major depressive disorder were treated with either 225 to 450 mg of bupropion (N = 63), or 150 to 400 mg of trazodone (Doseryl) (N = 61). Following 6 weeks of treatment 58 percent of bupropion-treated patients were considered much or very much improved compared with 46 percent of trazodone-treated patients. In another study, fluoxetine (Prozac) and bupropion demonstrated similar efficacies in treating major depressive disorder and accompanying anxiety symptoms in a double-blind controlled trial. Both drugs provoked minimal adverse effects.

Psychiatrists are increasingly called upon to treat depressed patients who have been refractory to one or more antidepressant treatments. Under double-blind conditions, bupropion was superior to placebo in treating hospitalized patients who were refractory to tricyclic drugs. In an open-label study, depressed outpatients with histories of either nonresponse or nonresponse plus intolerance to tricyclic drugs showed marked improvement on bupropion. In a recently published open study of 41 depressed patients who had failed to respond to well-documented tricyclic treatment, about half responded to bupropion.

Bupropion's adverse-effect profile is relatively mild and benign, and the medication is generally better tolerated than tricyclic drugs. While considerable attention has been focused on the issue of treatment-associated seizures, when bupropion is prescribed according to the current recommended guidelines regarding patient selection and dosage, the risk is similar to that of tricyclic and tetracyclic antidepressants. In contrast to serotonin selective reuptake inhibitors (SSRIs), bupropion does not appear to impair psychosexual function.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System Bupropion pharmacotherapy has been linked to seizures, although the relative risk is probably not substantially greater than that with other antidepressants when the current prescribing guidelines are followed. Other, more common adverse effects include agitation, insomnia, and restlessness. Psychosis, confusion, delirium, and other altered mental status effects have been described but probably occur in a very small minority of patients.

Cardiovascular System Clinical trials of potential new antidepressants exclude patients with substantial coexistent medical conditions, for obvious safety and ethical considerations. In this population, bupropion did not demonstrate cardiac adverse effects, and in contrast to tricyclic drugs, it did not affect cardiac conduction or induce orthostatic hypotension. Since becoming available for clinical use, bupropion has demonstrated a lack of impairment on cardiac output in patients with preexisting heart failure, and its benign profile in terms of orthostatic blood pressure effects has been confirmed in patients with preexisting cardiac disease. Bupropion has been well tolerated by patients who had previously developed orthostatic hypotension in conjunction with tricyclic antidepressant pharmacotherapy. Occasional patients with preexisting hypertension may exhibit increased blood pressure following initiation of bupropion treatment. Because of a lack of reported clinical experience in patients with a history of recent myocardial infarction or unstable heart disease, the manufacturer recommends exercising care in prescribing bupropion to these patient populations.

Blood Bupropion does not appear to affect the hematopoetic system. Rare instances of lymphadenopathy, anemia, and pancytopenia have been reported but without clear causal relationship to bupropion exposure.

Liver Although scattered abnormalities in liver function test results were reported during the premarketing clinical trials, there is no clinical evidence that bupropion has hepatotoxicity in humans. Because bupropion and its active metabolites undergo hepatic metabolism, patients with substantial liver disease and impairment should receive reduced dosages and should be carefully supervised and monitored for possible toxic effects from elevated blood concentrations of the parent compound and active metabolites.

Sexual Functioning In sharp contrast to SSRIs and some other antidepressants, bupropion does not appear to be associated with psychosexual dysfunction, including decreased libido, anorgasmia, and erectile problems. In fact, several published reports describe successful use of bupropion in patients who had stopped treatment with other antidepressants because of psychosexual adverse effects.

Weight In contrast to tricyclic drugs and several other antidepressants, bupropion causes appetite suppression more often than appetite enhancement. Weight loss of more than 5 pounds occurred in 28 percent of patients treated with bupropion, about twice that seen with placebo or tricyclic antidepressants. In addition, less than 10 percent of bupropion-treated patients gained weight versus nearly 35 percent of patients receiving tricyclic drugs.

Skin Bupropion has been associated with rashes and pruritis in a small minority of patients.

THERAPEUTIC INDICATIONS

The primary indication for bupropion is the treatment of major depression. Bupropion has been shown superior to placebo and as effective as both standard tricyclic antidepressants and the SSRIs in treating both ambulatory and hospitalized depressed patients. Under double-blind conditions, bupropion was superior to placebo in treating hospitalized patients who were refractory to tricyclic pharmacotherapy. In an open-label study, outpatients with a history of either nonresponse or nonresponse coupled with intolerance to tricyclic drugs showed marked improvement when switched to bupropion treatment. Additionally, in a recent open study of 41 depressed patients who had failed to respond to well-documented tricyclic treatment, about half responded to bupropion.

The role of bupropion in the treatment of bipolar disorder depression is unclear. Most clinicians recognize the potential danger of switching a depressed bipolar disorder patient into a manic state or inducing a rapid-cycling pattern associated with the use of antidepressants alone without concurrent mood-stabilizing agents. Initially, bupropion was considered especially helpful in these situations, with a putatively lower propensity for inducing the switch process. However, the past several years have seen several case reports of possible bupropion precipitation of mania and a mixed affective state. In addition, in a published series of 11 consecutive bipolar disorder patients, 6 experienced hypomanic or manic symptoms when bupropion was added to their treatment regimen. On the other hand, a recent report suggests that bupropion may be particularly promising in treating rapid-cycling bipolar II disorder patients. In that case series, all six patients showed improvement, with “dramatic” improvement in four that was sustained after an average of 2 years of treatment. In addition, none of the patients developed hypomania or rapid cycling, in contrast to the experience with conventional antidepressants. Furthermore, a recent prospective, double-blind trial found bupropion to be less likely to induce hypomania or mania in depressed bipolar disorder patients than desipramine (Norpramin). Thus, the most valid data suggest that bupropion may be quite useful in the treatment of bipolar disorder depression, but caution is still advised.

Because bupropion enhances dopaminergic neurotransmission, it makes sense that its potential efficacy has been examined in other psychiatric illnesses in which psychostimulant-like properties could be useful. Bupropion has been found effective in the treatment of attention-deficit disorders in two double-blind, placebo-controlled trials in children and one open trial in adults. In a double-blind crossover study, bupropion and methylphenidate had similar efficacies (Ritalin) in children with attention-deficit/hyperactivity disorder. However, bupropion may exacerbate tics in children with comorbid Tourette's disorder and attention-deficit disorders. Bupropion's activating properties have encouraged open trials of its potential use in the treatment of chronic fatigue syndrome and fatigue associated with multiple sclerosis.

In a placebo-controlled, double-blind trial, bupropion was superior to placebo in reducing episodes of binge eating and purging in patients with bulimia nervosa. However, because four subjects experienced grand mal seizures during treatment, the use of bupropion in the treatment of bulimia nervosa should be avoided. In a recent double-blind comparison, the efficacy of bupropion was intermediate between that of fluoxetine and placebo in the treatment of premenstrual dysphoric disorder but not significantly different from either. Uncontrolled case reports have described the use of bupropion in the treatment of social phobia.

In a recent randomized, double-blind study, 42 male veterans who smoked at least one pack of cigarettes daily received either bupropion (300 mg daily) or placebo for 3 months as an adjunct to group smoking-cessation counseling. Using salivary nicotinine concentrations to confirm smoking cessation, bupropion was significantly superior to placebo on all a priori defined measures of smoking cessation. A second study, examining 160 nondepressed male subjects and 30 female subjects, the sustained-release preparation of bupropion was again found superior to placebo as an adjunct for smoking cessation in patients who had failed multiple prior attempts to stop smoking. In a 3-month double-blind controlled trial of group smoking-cessation counseling, smoking abstinence of at least 1 week was achieved by 60 percent of the patients in the bupropion group, compared with 35 percent of patients in the placebo group.

In the largest published study to date, 615 cigarette smokers were assigned, under double-blind conditions, to treatment with either placebo or 100, 150, or 300 mg of sustained-release bupropion as an adjunct to brief counseling sessions for smoking cessation. After 7 weeks of treatment, the rates of smoking cessation (confirmed by carbon monoxide measurements in expired air) were 19 percent in the placebo group, 29 percent in the 100-mg bupropion group, 39 percent in the 150-mg bupropion group, and 44 percent in the 300-mg bupropion group. At 1-year follow-up, the respective cessation rates had dropped to 12, 20, 23, and 23 percent. Subjects who were continuously abstinent through the end of treatment exhibited a mean absolute weight gain that showed a significant inverse correlation with bupropion dose.

PRECAUTIONS AND ADVERSE REACTIONS

Much attention has understandably focused on the rare, but serious, adverse effect of bupropion-induced seizure. A careful retrospective review found that the incidence of seizures in patients receiving bupropion at dosage of 450 mg a day or less ranged from 0.33 to 0.44 percent, depending on the method used to complete this calculation. The cumulative 2-year risk in patients receiving 450 mg a day or less was 0.48 percent. This is comparable to the risk of seizures associated with other tricyclic or tetracyclic drugs. For example, the estimated frequency of seizures in outpatients with no predisposing factors, receiving tricyclic drugs at modest dosages (e.g., 150 mg a day or less) is 0.1 percent, and the frequency increases to 0.6 to 0.9 percent at higher dosages (e.g., 200 mg a day or more). In a prospective study of 3100 patients receiving the sustained-release formulation of bupropion, the incidence of seizures was 0.1 percent at dosages of 300 mg a day or less. To minimize the risk of seizure induction, patients should be carefully evaluated for potential risk factors, including past history of seizures, recent withdrawal from alcohol or anxiolytic drugs, concomitant therapy with other drugs that lower the seizure threshold, and history of organic brain disease or abnormal electroencephalogram (EEG). In addition, conservative dosage titration with a maximum dose of 400 to 450 mg a day and use of divided-dose schedules (e.g., thrice daily for the immediate-release formulation and twice daily for the sustained-release formulation) should further minimize the risk for bupropion-related seizures.

Bupropion's profile of common adverse reactions and side effects is strikingly different from that of conventional tricyclic drugs. It lacks anticholinergic effects, is clearly not sedating, and suppresses appetite in some patients. Unlike several other second-generation antidepressants, including the serotonin selective reuptake inhibitors, bupropion does not cause psychosexual dysfunction. A recent report documents that patients who experience psychosexual dysfunction in conjunction with fluoxetine treatment report substantially greater satisfaction with their sexual function while receiving bupropion.

Bupropion's cardiovascular profile is especially favorable. It does not appear to cause electrocardiographic ECG changes, and orthostatic hypotention is rare, even in patients with preexisting heart disease. In a direct comparison, bupropion was found safer than nortriptyline in regard to cardiovascular adverse effects.

Bupropion appears to be relatively less lethal following overdose than tricyclic drugs and certain other antidepressants. However, as with all medications, it should be distributed with care to patients who are at risk for suicide attempt. In a review of 58 cases of bupropion overdose and 9 cases of combined bupropion-benzodiazepine ingestion, bupropion did not appear to possess major cardiovascular toxicity. Neurological toxicity was described in that case series, including tremors, lethargy, and seizures; the seizures responded well to treatment with either benzodiazepines or phenytoin (Dilantin). Two fatal overdoses have been reported, associated with peripheral blood bupropion concentrations of 4.0 to 4.2 mg/L and total metabolite concentrations of 15 and 16.6 mg/L. By history, the estimated lethal doses were less than 10 grams.

Bupropion is associated with side effects that are linked to its pharmacological profile of enhanced dopaminergic neurotransmission. Bupropion can have activating effects similar to those seen with conventional psychostimulants, which are often helpful in patients with psychomotor retardation. On the other hand other patients can experience such effects as troubling agitation or insomnia. Appetite suppression is also perceived as an advantage in some clinical situations but a disadvantage in others. Psychotic symptoms, including hallucinations and delusions, have been associated with bupropion treatment. Since the initial description of bupropion-related psychoses, numerous case reports have described similar toxic reactions, including organic mental disorders, delirium, and catatonia. These relatively rare psychotic reactions may reflect overstimulation of dopaminergic systems, since one series of patients with bupropion-associated psychoses displayed increased plasma concentrations of homovanillic acid, a major metabolite of dopamine.

DRUG INTERACTIONS

Since bupropion undergoes extensive hepatic metabolism, other medications that inhibit or induce hepatic microsomal enzyme systems may affect its metabolism. The combined use of bupropion with MAO inhibitors should be avoided because of the risk of a hypertensive reaction if bupropion is added to ongoing MAO inhibitor. Treatment with MAO inhibitors should be discontinued at least 2 weeks prior to initiating bupropion pharmacotherapy.

Because bupropion increases synaptic availability of dopamine by blocking its reuptake, patients with coexisting Parkinson's disease and depression can often decrease their dose of antiparkinsonian medication when bupropion is added to their regimen. However, the combination of bupropion and antiparkinsonian medication should be administered cautiously, since the emergence of hallucinations, confusion, and dyskinesia has been described following the addition of bupropion to previously therapeutic dosages of levodopa (Larodopa).

Case reports have described the onset of delirium and a grand mal seizure following the addition of bupropion to fluoxetine treatment. Although the combination of lithium (Eskalith) and bupropion has been generally safe and well tolerated in a small number of published studies, a case report suggests that bupropion may affect lithium serum concentrations, and combined bupropion-lithium therapy has been linked to seizures in three patients. Two case reports suggest that carbamazepine (Tegretol) may decrease bupropion blood concentration and increase hydroxybupropion concentrations, while bupropion may increase valproate (Depakene) concentration.

LABORATORY INTERFERENCES

Bupropion pharmacotherapy may be associated with false positive test results for amphetamine in urine toxicology screening tests. In one report, the white blood cell count decreased by 10 to 14 percent during the first 2 months of therapy. However, there is no need to monitor hematological indexes routinely during bupropion therapy.

DOSAGE AND ADMINISTRATION

Bupropion is available in 75- and 100-mg tablets; the sustained-release formulation is available in 100- and 150-mg tablets. For the immediate-release preparation, the usual adult dosage is 100 mg thrice daily. Dosing should begin, in a healthy adult, at 100 mg twice a day and may be increased to 100 mg thrice daily after a minimum of 3 days. An increase in dosage, up to a total of 450 mg a day, in divided doses that do not exceed 150 mg each, can be pursued for patients who fail to respond to lower doses.

The usual target adult dosage of the sustained-release formulation is also 300 mg, but given as 150 mg twice a day. Dosing with the sustained-release formulation, in a healthy adult, is recommended to begin at 150 mg a day, given as a single daily dose in the morning. If the 150-mg dose is well tolerated, the 150-mg twice-daily target dose can be administered as early as the fourth day of treatment. At least 8 hours should elapse between successive doses. When indicated, the total dosage may be increased to a maximum of 400 mg a day, given as 200 mg twice a day.

The manufacturer recommends close adherence to the recommended dosing guidelines to minimize the risk of seizure. As with other psychotropic medications, particular attention should be given to dosage for elderly adults, children and adolescents, and patients with coexistent liver or renal disease.

SUGGESTED CROSS-REFERENCES

Mood disorders are discussed in Chapter 16, nicotine-related disorders in Chapter 13, and other antidepressant pharmacotherapies in other sections of Chapter 32.

SECTION REFERENCES

Apter JT, Woolfolk RL: Lithium augmentation of bupropion in refractory depression. Ann Clin Psychiatry *2:* 7, 1990.

*Ascher J, Cole JO, Colin J, Feighner JP, Ferris RM, Fibiger H, Golden R, Martin P, Potter W, Richelson E, Sulser F: Bupropion: A review of its mechanism of antidepressant activity. J Clin Psychiatry *56:* 395, 1995.

Ashton AK, Rosen RC: Bupropion as an antidote for serotonin reuptake inhibitor-induced sexual dysfunction. J Clin Psychiatry *59:* 112, 1998.

Barrickman LL, Perry PJ, Allen AJ, Kuperman S, Arndt SV, Hermann KJ, Schumacher E: Bupropion versus methylphenidate in the treatment of attention-deficit hyperactivity disorder. J Am Acad Child Adoles Psychiatry *34:* 649, 1995.

*Berigan TR, Deagle EA III: Treatment of smokeless tobacco addiction with bupropion and behavior modification (Letter). JAMA *281:* 233, 1999.

Bupropion (Zyban) for smoking cessation. Med Lett Drugs Ther *39:* 77, 1997.

Carnive JM, Clark RD, Calais LA, Qualls C, Tuason VB: Bupropion treatment in veterans with posttraumatic stress disorder: An open study. J Clin Psychopharmacol *18:* 379, 1998.

Conners CK, Casat CD, Gualtieri CT, Weller E, Reader M, Reiss A, Weller RA, Khayrallah M, Ascher J: Bupropion hydrochloride in attention deficit disorder with hyperactivity. J Am Acad Child Adolesc Psychiatry *34:* 1314, 1996.

Davidson J: Seizures and bupropion: A review. J Clin Psychiatry *50:* 256, 1989.

Davidson J, Miller R, Fleet JVW, Strickland R, Manberg P, Allen S, Parrot R: A double-blind comparison of bupropion and amitriptyline in depressed patients. J Clin Psychiatry *44* (5, sect 2):115, 1983.

Feighner J, Hendrickson G, Miller L, Stern W: Double-blind comparison of doxepin versus bupropion in outpatients with a major depressive disorder. J Clin Psychopharmacol *6:* 27, 1986.

*Feighner JP, Gardner EA, Johnston JA, Batey SR, Khayrallah MA, Ascher JA, Lineberry CG: Double-blind comparison of bupropion and fluoxetine in depressed outpatients. J Clin Psychiatry *52:* 329, 1991.

Ferguson J, Cunningham L, Merideth C, Apter J, Ionescu-Pioggia M, Samara B, Johnston JA, Ascher J: Bupropion in tricyclic antidepressant nonresponders with unipolar major depressive disorder. Ann Clin Psychiatry *6:* 153, 1994.

Ferry LH, Burchette RJ: Evaluation of bupropion versus placebo for treatment of nicotine dependence. Presented at the 147th annual meeting of the American Psychiatric Association, Philadelphia, PA, May 22–26, 1994.

Fogelson DL, Bystritsky A, Pasnau R: Bupropion in the treatment of bipolar disorders: The same old story? J Clin Psychiatry *53:* 443, 1992.

Gardner EA, Johnston A: Bupropion: An antidepressant without sexual pathophysiological action. J Clin Psychopharmacol *5:* 24, 1985.

Golden RN, Dawkins K, Nicholas L, Bebchuk J: Trazodone, nefazodone, bupropion, and mirtazapine. In *Textbook of Psychopharmacology*, ed 2, A Schatzberg, C Nemeroff, editors. American Psychiatric Press, Washington, DC, 1998.

*Golden RN, DeVane L, Laizure SC, Rudorfer MV, Sherer M, Potter WZ: Bupropion in depression: The role of metabolites in clinical outcome. Arch Gen Psychiatry *45:* 145, 1988.

Golden RN, James S, Sherer M, Rudorfer MV, Sack D, Potter WZ: Psychoses associated with bupropion treatment. Am J Psychiatry *142:* 1459, 1985.

Golden RN, Markey SP, Risby ED, Cowdry RW, Potter WZ: Antidepressants reduce whole-body norepinephrine turnover while enhancing 6-hydroxymelatonin output. Arch Gen Psychiatry *45:* 150, 1988.

*Golden RN, Rudorfer MV, Sherer M, Linnoila M, Potter WZ: Bupropion in depression: Biochemical effects and clinical response. Arch Gen Psychiatry *45:* 139, 1988.

Goodnick PJ: Blood levels and acute response to bupropion. Am J Psychiatry *149:* 399, 1992.

Goodnick PJ, Dominguez RA, De Vane CL, Bowden CL: Bupropion slow-release response in depression: Diagnosis and biochemistry. Biol Psychiatry *44:* 629, 1998.

Goodnick PJ, Sandoval R, Brickman A, Klimas NG: Bupropion treatment of fluoxetine-resistant chronic fatigue syndrome. Biol Psychiatry *32:* 834, 1992.

*Hebert S: Bupropion (Zyban, sustained-release tablets): Reported adverse reactions. CMAJ *160:* 1050, 1999.

Horne RL, Ferguson JM, Pope HG Jr, Hudson JI, Lineberry CG, Ascher J, Cato A: Treatment of bulimia with bupropion: A multicenter controlled trial. J Clin Psychiatry *49:* 262, 1988.

Hurt RD, Sachs DP, Glover ED, Offord KP, Johnston JA, Dale LC, Khayrallah MA, Schroeder DR, Glover PN, Sullivan CR, Corghan IT, Sullivan PM: A comparison of sustained-release bupropion and placebo for smoking cessation. N Engl J Med *337:* 1195, 1997.

Kiev A, Masco HL, Wenger TL, Johnston JA, Batey SR, Holloman LC: The cardiovascular effects of bupropion and nortriptyline in depressed outpatients. Ann Clin Psychiatry *6:* 107, 1994.

Labbate LA: Bupropion-SR-induced increased libido and spontaneous orgasm (Letter). Can J Psychiatry *43:* 644, 1998.

Masand P, Stern TA: Bupropion and secondary mania. Is there a relationship? Ann Clin Psychiatry *5:* 271, 1993.

*Namerow LB: Seizure associated with bupropion and guanfacine. J Am Acad Child Adolesc Psychiatry *38:* 2, 1999.

Pearlstein TB, Stone AB, Lund SA, Scheft H, Zlotnick C, Brown WA: Comparison of fluoxetine, bupropion, and placebo in the treatment of premenstrual dysphoric disorder. J Clin Psychopharmacol

17: 261, 1997.

Pitts WM, Fann WE, Halaris AE, Dressler DM, Sajadi C, Snyder S, Ilaria RL: Bupropion in depression: A tri-center placebo-controlled study. J Clin Psychiatry *44* (5, sect 2):95, 1983.

Pollock BG, Reynolds CF III, Kirshner M, Sweet RA: Bupropion plasma levels and CYP2D6 phenotype. Ther Drug Monit *18:* 581, 1996.

Popli AP, Tanquary J, Lamparella V, Masand PS: Bupropion and anticonvulsant drug interactions. Ann Clin Psychiatry *7:* 99, 1995.

*Raymond E, Bahdai Z, Waters H, Kinzie S, Chan WL, Blake E, Evans MF, Chan D, Shaw E: What's new in smoking cessation: Zyban. Can Fam Physician *45:* 633, 1999.

Riggs PD, Leon SL, Mikulich SK, Pottle LC: An open trial of bupropion for ADHD in adolescents with substance use disorders and conduct disorder. J Am Acad Child Adolesc Psychiatry *37:* 1271, 1998.

Roose SP, Dalack GW, Glassman AH, Woodring S, Walsh BT, Giardina EGV: Cardiovascular effects of bupropion in depressed patients with heart disease. Am J Psychiatry *148:* 512, 1991.

Sachs GS, Lafer B, Stoll AL, Banov M, Thibault AB, Tohen M, Rosenbaum JF: A double-blind trial of bupropion versus desipramine for bipolar depression. J Clin Psychiatry *55:* 391, 1994.

Spencer T, Biederman J, Steingard R, Wilens T: Bupropion exacerbates tics in children with attention-deficit hyperactivity disorder and Tourette's syndrome. J Am Acad Child Adolescent Psychiatry *32:* 211, 1993.

Spiller HA, Ramoska EA, Krenzelok EP: Bupropion overdose: A 3-year multi-center retrospective analysis. Am J Emergency Med *12:* 43, 1994.

*Walker PW, Cole JO, Gardner EA, Hughes AR, Johnston JA, Batey SR: Improvement in fluoxetine-associated sexual dysfunction in patients switched to bupropion. J Clin Psychiatry *54:* 459, 1993.

Weisler RH, Johnston JA, Lineberry CG, Samara B, Branconnier RJ, Billow AA: Comparison of bupropion and trazodone for the treatment of major depression. J Clin Psychopharmacol *14:* 170, 1994.

Wender PH, Reimherr FW: Bupropion treatment of attention-deficit hyperactivity disorder in adults. Am J Psychiatry *147:* 1018, 1990.

31.12 BUSPIRONE

OLGA BRAWMAN-MINTZER, M.D., R. BRUCE LYDIARD, FH.D., M.D., AND JAMES C. BALLENGER, M.D.

Buspirone (BuSpar) is one of the azaspirone class of clinical compounds and differs both structurally and pharmacologically from benzodiazepines, barbiturates, or other sedative-anxiolytic compounds. Unlike benzodiazepines, buspirone does not exert anticonvulsant, muscle-relaxant, or hypnotic effects. The precise mechanism of action of buspirone is not known but appears to be complex, probably involving several brain neurotransmitter systems. Pharmacologically, buspirone exhibits a high affinity for serotonin (5-hydroxytryptamine [5-HT]) subtype in (5-HT_{1A}) receptors and moderate affinity for dopamine type 2 (D_2) receptors in vitro. The main active metabolite, 1-pyrimidinylpiperazine (1-PP), posesses a $_2$adrenergic receptor antagonism. Buspirone has no appreciable affinity for the g-aminobutyric acid (GABA)-benzodiazepine-chloride receptor complex and does not affect binding of benzodiazepines at the receptor complex in vitro.

CHEMISTRY

Buspirone (Fig. 31.12-1) is chemically designated 8-[4-[4-(2-pyrimidinyl)-1-piperazinyl]butyl]-8-azaspiro[4,5]decane-7,9-dione monohydrochloride. The heteroaryl piperazine nature of buspirone appears to be important for the anxiolytic activity, serotonergic effects, and dopamine antagonist properties.

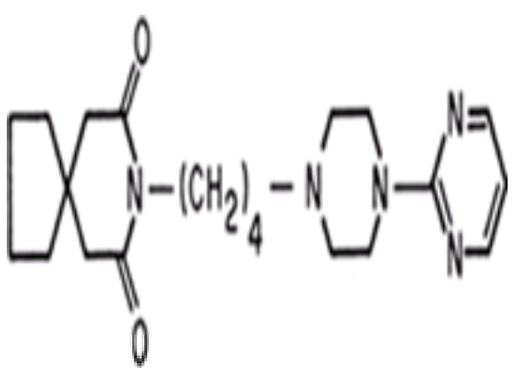

FIGURE 31.12-1 Molecular structure of buspirone.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Buspirone is rapidly and almost completely absorbed after oral administration. However, the drug undergoes extensive first-pass metabolism, with approximately 4 percent of the parent drug reaching the systemic circulation. Peak plasma buspirone levels of 1 to 6 ng/mL occur within 40 to 90 minutes after a single oral dose of 20 mg. In healthy adult volunteers, the relation between plasma buspirone concentration and the oral dose appeared linear at single doses ranging from 10 to 40 mg. Following a multiple-dose oral administration, buspirone may have nonlinear pharmacokinetics, that is, dose increases, multiple doses, or both may lead to higher plasma concentrations than those predicted from single-dose studies. Ingestion of food delays gastrointestinal absorption of buspirone and increases the amount of unchanged buspirone reaching the systemic circulation. However, neither its efficacy nor the incidence of adverse effects appear to be affected by whether buspirone is administered with or without food.

Distribution At present, data regarding distribution of buspirone into body tissues and fluids are available primarily from animal models. In animals, buspirone is extensively distributed following intravenous administration, achieving higher concentrations in lung, kidney, and adipose tissue than in plasma. The concentration of buspirone in the brain parallels that in the plasma, but the active metabolite (1-PP) accumulates in the brain at concentrations 4 to 5 times higher than those in plasma. The volume of distribution of buspirone after intravenous administration averages 5.3 L/kg in healthy adults.

Approximately 95 percent of buspirone is bound to plasma proteins (albumin and a $_1$-acid glycoprotein). However, buspirone does not displace other highly protein bound drugs such as phenytoin (Dilantin), propranolol (Inderal), and warfarin (Coumadin). In vitro, buspirone does displace digoxin (Lanoxin).

Metabolism and Elimination After undergoing extensive hepatic metabolism, buspirone and its metabolites are excreted primarily in the urine (29 to 63 percent) and to a lesser extent in feces (18 to 38 percent). The major metabolites of buspirone are the inactive 5-hydroxybuspirone (5-HB) (via hydroxylation) and the pharmacologically active metabolite 1-PP (via oxidative dealkylation). The oxidation process has been shown to be mediated by cytochrome P450 isoenzyme 3A4 (CYP 3A4) in vitro. In animals, 1-PP is 20 to 25 percent less potent than buspirone, but brain concentrations are 15- to 30-fold greater than those of buspirone despite relatively lower blood concentrations.

The elimination half-life of unchanged buspirone following a single dose of 10 to 40 mg ranges from 2 to 11 hours. The average elimination half life of 1-PP following 20 mg of buspirone is 6 hours. The average elimination half life of 1-PP following 20 mg of buspirone is 6 hours. The elimination half-life of buspirone may be prolonged in patients with renal impairment and those with cirrhosis. Such patients may require lower doses of buspirone. The extent of secretion of buspirone and metabolites in human milk is unknown. However, in animal models buspirone and its metabolites appear in measurable amounts.

Pharmacodynamics

Mechanism of Action The precise mechanism of pharmacological action of buspirone is not known but probably involves several brain neurotransmitter systems.

BENZODIAZEPINE-GABA SYSTEM Unlike benzodiazepines, buspirone has no affinity for the GABA-benzodiazepine-chloride receptor complex. Buspirone does not appear to stimulate or inhibit binding of benzodiazepines at the receptor complex in vitro, although binding may be enhanced in vivo. Buspirone does not directly affect GABA binding. Further, benzodiazepine antagonists (such as flumazenil [Romazicon]) do not block the anxiolytic activity of buspirone. However, it has been suggested that buspirone may possibly indirectly affect GABA transmission via other mechanisms.

DOPAMINE SYSTEM Buspirone has a moderate affinity for dopamine D_2 receptors in the central nervous system (CNS). Buspirone appears to function as an antagonist of D_2 presynaptic autoreceptors, thus increasing dopaminergic neurotransmission and dopamine turnover. Some evidence in animal models indicates that buspirone may, in a limited fashion, block the postsynaptic D_2 receptors. It may also possess some dopamine agonist activity. However, the clinical significance of

these dopamine-related effects of buspirone has not been established, and they are probably clinically unimportant.

SEROTONIN SYSTEM Serotonin function has been thought to be important in various psychiatric disorders including anxiety and depression. Thus, it has been suggested that buspirone's effects on the serotonin system may contribute to its anxiolytic and possibly antidepressant effects. Specifically, buspirone was found to have high affinity for the 5-HT_{1A} receptor subtype in the CNS. The 5-HT_{1A} receptors are present in high concentrations in the dorsal raphe nucleus, which contains the cell bodies of the ascending serotonin neurons (i.e., somatodendritic-presynaptic receptors), and in the hippocampal and cortical regions, where the ascending 5-HT neurons terminate (i.e., postsynaptic receptors). Buspirone acts as a partial agonist or mixed agonist/antagonist at the 5-HT_{1A} receptor. The agent appears to have predominantly agonistic activity at the presynaptic 5-HT_{1A} receptors and partial agonistic activity at the postsynaptic receptors. Buspirone's ability to inhibit spontaneous firing of serotonergic neurons in the dorsal raphe nucleus, which probably accounts for its anxiolytic activity, has be suggested to be mediated through the somatodendritic serotonin autoreceptors. It has also been suggested that the agonist activity of buspirone at the postsynaptic 5-HT_{1A} receptors may be associated with its antidepressant activity.

NORADRENERGIC SYSTEM Buspirone was initially reported to be inactive at CNS adrenergic receptor sites. However, both preclinical and human data indicate that 1-PP, the main metabolite of buspirone, has central and peripheral a_2-adrenergic receptor blocking activity. Buspirone and 1-PP increase neuronal activity in the locus ceruleus and increase norepinephrine metabolism. Buspirone also decreases striatal concentrations of norepinephrine and increases the concentrations of 3-methoxy-4-hydroxyethylenephenylglycol (MHPG). Buspirone may also downregulate b_1-adrenoreceptors.

CHOLINERGIC SYSTEM The effects of buspirone on the cholinergic neurotransmitter system have not been fully elucidated. However, buspirone does not seem to interact directly with cholinergic receptors. In animal models buspirone administration increases the acetylcholine efflux in hippocampal neurons. Buspirone also decreases acetylcholine concentrations in the striatum.

Relation of Clinical Effects and Blood Concentrations No clear relation exists between plasma buspirone concentrations and clinical efficacy. Thus, there are no clinical indications for determining buspirone plasma concentrations.

CLINICAL STUDIES

Generalized Anxiety Disorder The efficacy of buspirone as an anxiolytic has been established in several multicenter, double-blind placebo-controlled trials, primarily for the treatment of generalized anxiety disorder. The published studies compared buspirone with another anxiolytic agent, placebo, or both. Buspirone has been compared with diazepam (Valium), clorazepate (Tranxene), alprazolam (Xanax), lorazepam (Ativan), and oxazepam (Serax). Most studies had a 1-week placebo run-in phase followed by double-blind randomization to a 4-week treatment phase (buspirone, a benzodiazepine, or placebo). The Hamilton Rating Scale for Anxiety (HAM-A) was the most widely used rating scale for clinical assessments. The mean doses of buspirone ranged from 7.5 to 42.3 mg daily.

Most of the available studies used the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) criteria for generalized anxiety disorder, which required only a 1-month duration of anxiety symptoms. Three studies used the revised third edition (DSM-III-R) or the fourth edition (DSM-IV) diagnostic criteria, which extended the required duration of anxiety symptoms to a minimum of 6 months. Thus, fewer data are available on the efficacy of buspirone in subjects with chronic symptoms of anxiety. Of the 13 peer-reviewed, placebo-controlled studies, 6 showed no benefit of buspirone over placebo following 4 to 6 weeks of treatment. In 6 of the 10 peer-reviewed, placebo-controlled studies that compared buspirone with a standard benzodiazepine, buspirone was statistically significantly better than placebo and comparable to diazepam, alprazolam, and lorazepam. However, 2 of those 10 showed no benefit for diazepam and buspirone over placebo, and 2 showed no benefit of buspirone over placebo. The therapeutic response to buspirone was usually evident by 2 to 3 weeks, compared with a more rapid (e.g., 1 to 2 weeks) response with benzodiazepines. Benzodiazepines may also be slightly more effective in the treatment of somatic symptoms of anxiety, but no significant differences appear to exist between buspirone and benzodiazepines in measures of psychic anxiety. Buspirone may be more effective in the treatment of anger and hostility symptoms than benzodiazepines. Finally, some investigators suggested that prior treatment with benzodiazepines may decrease the efficacy of buspirone. However, available data indicate that only very recent (i.e., within 4 weeks) treatment with benzodiazepines tends to reduce the anxiolytic benefit of buspirone.

One study examined the effects of long-term treatment (6 months) of generalized anxiety disorder symptoms with buspirone or clorazepate. Results suggest that buspirone was as effective as chlorazepate in the treatment of anxiety after the second week of treatment. Further, in contrast to clorazepate, abrupt discontinuation of buspirone treatment was not accompanied by withdrawal symptoms. At 40-month follow-up, 30 percent of patients who received 6-months of clorazepate treatment but none of patients who received buspirone had returned to regular use of anxiolytic medications (benzodiazepines in most cases). The authors concluded that long-term buspirone therapy for generalized anxiety disorder may have produced more lasting improvement than therapy with the benzodiazepines.

The efficacy and tolerability of buspirone in the treatment of anxiety symptoms was also evaluated in elderly patients. One 4-week, placebo-controlled trial examined the effects of buspirone (at doses ranging from 15 to 30 mg a day) in the treatment of anxiety symptoms in 40 elderly patients who had both anxiety and depressive symptoms. Patients receiving buspirone exhibited significantly lower scores in the HAM-A, Hamilton Rating Scale for Depression (HAM-D), and Clinical Global Impression (CGI) rating scales than to placebo recipients. Buspirone treatment was generally well tolerated by elderly patients at doses similar to those used to treat younger patients; in fact, most (75 percent) buspirone-treated patients reported no adverse experiences.

Other Anxiety Disorders The effects of buspirone have been evaluated in the treatment of panic disorder, social phobia, obsessive-compulsive disorder, and posttraumatic stress disorder. In two double-blind, placebo-controlled studies (one with imipramine [Tofranil] and another with alprazolam), buspirone did not differ from placebo in the treatment of patients with panic disorder, but the active agents did. In a recent controlled study in patients with panic disorder, buspirone enhanced the effects of cognitive behavioral therapy on measures of generalized anxiety and agoraphobia; however this enhanced effect did not persist at follow-up.

Two open trials found buspirone only modestly effective in the treatment of patients with social phobia (predominantly the generalized subtype). One double-blind, placebo-controlled study evaluating the effects of buspirone and cognitive-behavioral therapy in the treatment of musicians with performance anxiety who met DSM-III-R criteria for social phobia found buspirone ineffective in the treatment of performance anxiety. Finally, in a recent double-blind, placebo-controlled study in the treatment of patients diagnosed with DSM-IV social phobia patients, buspirone did not differ from placebo.

Several case reports or series and one open trial suggest that buspirone may have some efficacy in the treatment of posttraumatic stress disorder symptoms, such as increased arousal and reexperiencing of the traumatic event.

The existing data regarding the efficacy of buspirone in the treatment of obsessive-compulsive disorder are limited and conflicting. For example, in one open trial buspirone was ineffective in the treament of obsessive-compulsive disorder symptoms. Conversely, in the only controlled study comparing the effects of buspirone, clomipramine (Anafranil), and placebo in the treatment of 18 patients with obsessive-compulsive disorder, buspirone and clomipramine produced significant and comparable improvement. Buspirone is also the most studied augmenting agent in obsessive-compulsive disorder patients. Adding buspirone to ongoing fluoxetine (Prozac) treatment was reported to diminish obsessive-compulsive disorder symptoms in open trials, but controlled studies examining the usefulness of adding buspirone to fluoxetine, fluvoxamine (Luvox), and clomipramine treatment in obsessive-compulsive disorder patients reported that buspirone was ineffective in enhancing the therapeutic outcome in most patients. Patients who do benefit appear to be those with a partial response; patients with no response appear to obtain no further benefit from the addition of buspirone.

Major Depressive Disorder Following an early open-label study that suggested the possible effectiveness of buspirone in the treatment of nonmelancholic major depression three placebo-controlled studies (varying considerably in their diagnostic/severity criteria), one meta-analysis of eight studies, and one composite of five parallel group placebo-controlled studies, examined the effects of buspirone treatment in patients with major depression and anxiety symptoms or primary generalized anxiety disorder with depressive symptoms. Overall, these studies found buspirone (at relatively high mean dosages ranging from 30 to 56 mg daily) to be superior to placebo in the treatment of depressive and anxiety symptoms (as measured by the HAM-D and HAM-A rating scales and CGI ratings). In some studies patients with more-severe depressive symptoms or melancholic subtype appeared to respond better to buspirone than patients with mild or nonmelancholic depression. In several studies a considerable number of patients (up to 64 percent of buspirone-treated patients and up to 74 percent of placebo-treated subjects) discontinued the study before 8 weeks.

Open trials have tested buspirone augmentation in patients with major depression who had failed to respond to standard trials of antidepressants. For example, in one open trial, seven of eight nonresponders (six patients receiving fluoxetine and one patient receiving imipramine) reported a full or partial antidepressant response after the addition of buspirone, 30 mg daily. In another open trial, 17 of 25 patients initially receiving fluoxetine or fluvoxamine were reported to improve (using

four-point CGI ratings) following the addition of buspirone (at doses of 20 to 50 mg daily). However, in a recent double-blind, placebo-controlled study that evaluated the efficacy and safety of buspirone augmentation in the management of depressed patients who failed to respond to the selective serotonin reuptake inhibitors (SSRIs) paroxetine or citalopram, no significant between-group differences in response rates were found. It should be noted that study results may be inconclusive due to the high placebo response rates observed (47 percent for placebo versus 51 percent for buspirone). In summary, the efficacy of buspirone in the treatment of depression remains uncertain at this time.

Other Uses

Alcohol Use Disorders The effects of buspirone in treating relapsed alcoholics with concurrent anxiety symptoms have been mixed. Four double-blind, placebo-controlled studies examined the effects of buspirone treatment of anxious alcoholics. In one study buspirone treatment was associated with significantly greater treatment retention and greater decreases in alcohol craving, anxiety, and depression than placebo. A study of alcoholics with comorbid generalized anxiety disorder reported a similar advantage of buspirone over placebo treatment. Another study found buspirone therapy to be associated with reduced anxiety, a slower return to heavy alcohol consumption, and fewer drinking days during the follow-up period. In contrast, one study of male alcoholic veterans in aftercare showed no differences between groups of patients treated with buspirone or placebo on measures of anxiety and alcohol consumption.

Psychosis Its activity at the D_2 receptor created some interest in the potential antipsychotic effects of buspirone. Although buspirone exhibited some evidence of antipsychotic potential in several animal models and reportedly has transient antipsychotic effects when administered in large dosages (mean dosage of 1470 mg daily) to patients with schizophrenia, the drug has no established antipsychotic activity in humans at usual dosages when administered alone. Several open trials that evaluated the efficacy of buspirone as an adjunct to antipsychotics in schizophrenia patients found no clear efficacy in decreasing psychosis or negative symptoms.

Aggression Some data indicate that buspirone may be an effective antiaggressive agent. Several animal studies have demonstrated potential anticonflict properties for buspirone. In humans, a small, placebo-controlled trial found buspirone to be effective in reducing aggression and anxiety in five of six developmentally disabled and mentally retarded subjects. Others have also suggested that buspirone may be effective in the treatment of agitation in brain-injured and demented patients. For example, buspirone treatment decreased scores on the Neurobehavioral Rating Scale in seven traumatically injured patients with behavioral problems. In a retrospective evaluation of the effects of buspirone treatment (for at least 3 months) in 10 patients with organically related psychiatric diagnosis and aggressive behaviors, 9 of the 10 patients treated with buspirone showed some improvement in their behavioral target symptoms. Patients with at least 50 percent reduction in their symptoms (6 of 10 patients) received relatively high doses of buspirone (an average dosage of 50 mg a day). Buspirone was well tolerated and caused minimal sedation and cognitive changes in these reports. In a recent open trial, buspirone was effective as an adjunct in the pharmacotherapy of oppositional defiant disorder symptoms in children with comorbid DSM-III-R defined attention-deficit hyperactivity disorder and oppositional defiant disorder whose attention-deficit disorder symptoms were significantly helped by standard treatments.

Pervasive Developmental Disorders A case report described improvement of hyperactive behaviors in two or four autistic disorder children receiving buspirone treatment. Additional open-label study reported moderate to marked improvement in 16 of 22 children diagnosed with pervasive developmental disorders who received buspirone 15 to 45 mg daily for 6 to 8 weeks.

Premenstrual Disorder Buspirone has also been investigated in the treatment of premenstrual syndrome. Two controlled studies (one using parallel design and one using crossover design) found buspirone to be more effective than placebo in reducing physical and psychic symptoms associated with premenstrual syndrome.

Somatoform Disorders Augmentation with buspirone had modest therapeutic effects in nonresponders and partially responsive patients with body dysmorphic disorder receiving fluoxetine or clomipramine.

Sexual Dysfunction Finally, limited data indicate that buspirone improved sexual function in generalized anxiety disorder patients, which was not correlated with anxiety reduction. Further, in one retrospective study, buspirone was found to alleviate SSRI-induced sexual dysfunction in 11 of 16 patients who received treatment with SSRIs. Failure to respond to buspirone treatment was most common among patients who received either high dosages of SSRIs or low dosages of buspirone (15 mg daily).

Nicotine Dependence The effects of buspirone in smoking cessation have been evaluated. The results of the available studies have been inconsistent and contradictory. The rate of smoking abstinence has been reported to range from 36 to 88 percent and 16 to 89 percent in buspirone and placebo treatment groups respectively. A controlled trial using strict methodology is needed to further evaluate buspirone efficacy in smoking cessation and the reduction of nicotine withdrawal symptoms.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Nervous System The most frequent adverse effects associated with buspirone treatment are CNS effects. The most common effects include dizziness, drowsiness, headaches, and light-headedness. Adverse effects associated with CNS depression, such as sedation, psychomotor dysfunction, or fatigue, occur less frequently with buspirone than with benzodiazepines. The potential of buspirone for causing adverse effects associated with its activity in the dopaminergic system is minimal. However, symptoms such as akathisia, dystonia, tremor, muscle stiffness/rigidity, or involuntary movements have occasionally been associated with buspirone treatment.

Gastrointestinal System Nausea has been reported in patients receiving buspirone. Less frequent adverse gastrointestinal effects include dry mouth, abdominal distress, diarrhea, constipation, and vomiting.

Cardiovascular System Although tachycardia, palpitations, and nonspecific chest pain have been reported in some patients receiving buspirone, the frequency of these symptoms does not appear to differ from that in patients receiving placebo. Buspirone has minimal, if any, effect on blood pressure.

Dermatological Effects Adverse dermatological effects such as rash or pruritus occur occasionally in patients receiving buspirone.

Endocrine System Buspirone produces a dose-dependent stimulation of prolactin secretion. Menstrual irregularities, galactorrhea, and thyroid abnormalities have occurred occasionally in patients receiving buspirone, but no causal relationship has been established.

Other Systems Although some side effects or adverse events have been reported in other organ systems, buspirone has little, if any, effect on the respiratory system, platelets, smooth muscles, or autonomic nervous system function.

THERAPEUTIC INDICATIONS

Buspirone is indicated by the Food and Drug Administration (FDA) for the treatment of “anxiety disorders or the short-term relief of the symptoms of anxiety,” a description that corresponds roughly to generalized anxiety disorder. No FDA approval for other indications has been granted.

PRECAUTIONS AND ADVERSE REACTIONS

The main contraindication to buspirone therapy is a history of a hypersensitivity reaction to buspirone. Buspirone should be used with caution in patients with impaired renal function, since the drug and its metabolites are excreted primarily by the kidneys. It should also be used cautiously in patients with hepatic dysfunction.

The combined use of monoamine oxidase inhibitors (MAOIs) and buspirone is not recommended because of the potential for hypertensive crisis. A seizure was reported in one obsessive-compulsive disorder patient who received buspirone augmentation of SSRI treatment. Hypomania was reported in a patient with a bipolar disorder who was receiving buspirone, lithium (Eskalith) tranylcypromine (Parnate), and alprazolam. Buspirone does not exhibit cross-tolerance with benzodiazepines, other sedative-hypnotic drugs, or alcohol. Some studies found that concomitant administration of buspirone during benzodiazepine taper may reduce the development of rebound anxiety or benzodiazepine-withdrawal symptoms. However, abrupt substitution of buspirone for benzodiazepines in chronic benzodiazepine users is not recommended.

Animal studies have found no evidence of buspirone-induced mutagenicity, carcinogenicity, or teratogenicity. The FDA has designated buspirone as a category B (no evidence of risk in humans) agent for pregnancy. Buspirone is secreted in the milk of rats, and it is recommended that buspirone be avoided in lactating women.

DRUG INTERACTIONS

Administration of buspirone to patients receiving MAOIs has been reported to be associated with hypertensive crisis. Concomitant administration of buspirone and haloperidol (Haldol) has resulted in increased serum haloperidol concentrations, suggested to be due to competitive inhibition of the oxidative dealkylation of haloperidol. Because buspirone is highly protein bound, it could theoretically displace other protein-bound drugs. Although in vivo studies failed to reveal substantial evidence of displacement by buspirone of phenytoin, propranolol, or warfarin, in vitro evidence suggests that buspirone may displace less firmly bound drugs such as digoxin. However, those effects have not been reported to be clinically important. Concurrent use of buspirone and cimetidine (Tagamet) is unlikely to result in a clinically important interaction. Buspirone does not appear to alter blood alcohol concentrations nor to potentiate alcohol-induced impairment of psychomotor and cognitive performance.

Although drug interaction studies suggest that buspirone does not potentiate benzodiazepine-induced CNS depression to a clinically important degree, the effects of combined therapy have not been fully evaluated. Buspirone potentiated the effects of diazepam on several tests of psychomotor function and has subjectively potentiated diazepam-induced sedation, but it also counteracted some diazepam-induced attention and learning acquisition impairments. Buspirone did not substantially alter hypnotic or psychomotor effects of flurazepam (Dalmane) and triazolam (Halcion). In general buspirone can be safely combined with benzodiazepines if it is indicated. Other CNS depressants such as analgesics, antihistamines, and sedative-hypnotics generally have little or no effect on the frequency or severity of adverse effects associated with buspirone administration.

Buspirone may interact with drugs that inhibit CYP 3A4. For example, it has been shown that potent CYP 3A4 inhibitors, such as erythromycin, itraconazole (Sporanox), nefazodone (Serzone), and grapefruit juice, increase buspirone plasma concentrations. Thus, lower dosages of buspirone are recommended, if the drugs are to be used in combination.

LABORATORY TESTS

Buspirone has not been reported to interfere with clinical laboratory values.

DOSAGE AND ADMINISTRATION

Buspirone is prepared as buspirone hydrochloride and is available in 5-, 10-, and 15-mg (Dividose) tablets. The usual initial adult dosage is 10 to 15 mg daily in three divided doses. Recently, the manufacturer recommended use of a twice-a-day dosing schedule (i.e., 7.5 mg two times a day), which appears to be equally effective and well tolerated. To achieve an optimal therapeutic response, the dosage may be increased by 5 mg a day at 2- to 3-day intervals. Patients may benefit from dosage titration (using twice-a-day dosing) to 30 to 60 mg a day. A lower initial dose of 5 mg twice a day may be necessary to avoid adverse effects in some patients.

SUGGESTED CROSS-REFERENCES

Anxiety disorders are discussed in Chapter 15, mood disorders in Chapter 14, and benzodiazepines in Section 31.9.

SECTION REFERENCES

Algeri S, De Luigi A, De Simoni MG, Imeri L, Marconi M, Nava S, Perego C, Sacchetti G: Multiple and complex effects of buspirone on central dopaminergic system. Pharmacol Biochem Behav *29:* 823, 1988.

*Apter JT, Allen LA: Buspirone: Future directions. J Clin Psychopharmacol *19:* 86, 1999.

Berlin I, Chalon S, Payan C, Schollnhammer G, Cesselin F, Varoquaux O, Puech AJ: Evaluation of the alpha $_2$-adrenoceptor blocking properties of buspirone and ipsapirone in healthy subjects. Relationship with the plasma concentration of the common metabolite 1-(2-pyrimidinyl)-piperazine. Br J Clin Pharmacol *39:* 243, 1995.

Bohm C, Robinson DS, Gammans RE, Shrotriya RC, Alms DR, Leroy A, Placchi M: Buspirone therapy in anxious elderly patients: A controlled clinical trial. J Clin Psychopharmacol *10:* 47S, 1990.

Buitelaar JK, van der Gaag RJ, van der Hoeven J: Buspirone in the management of anxiety and irritability in children with pervasive developmental disorders: Results of an open-label study. J Clin Psychiatry *59:* 56, 1998.

Chiaie RD, Pancheri P, Casacchia M, Stratta P, Kotzalidis GD, Zibellini M: Assessment of the efficacy of buspirone in patients affected by generalized anxiety disorder, shifting to buspirone from prior treatment with lorazepam: A placebo-controlled, double-blind study. J Clin Psychopharmacol *15:* 12, 1995.

Clark DB, Agras WS: The assessment and treatment of performance anxiety in musicians. Am J Psychiatry *149:* 278, 1992.

Cottraux J, Note ID, Cungi C, Legeron P, Heim F, Chneiweiss L, Bernard G, Bouvard M: A controlled study of cognitive behavior therapy with buspirone or placebo in panic disorder with agoraphobia. Br J Psychiatry *167:* 635, 1995.

Dimitriou EC, Dimitriou CE: Buspirone augmentation of antidepressant therapy. J Clin Psychopharmacol *18:* 465, 1998.

Enkelmann R: Alprazolam versus buspirone in the treatment of outpatients with generalized anxiety disorder. Psychopharmacology *105:* 428, 1991.

Fabre LF: Buspirone in the management of major depression: A placebo-controlled comparison. J Clin Psychiatry *51:* 55, 1990.

Farid P, Abate MA: Buspirone use for smoking cessation. Ann Pharmacother *32:* 1362, 1998.

Fichtner CG, Crayton JW: Buspirone in combat-related posttraumatic stress disorder. J Clin Psychopharmacol *14:* 79, 1994.

Gammans RE, Stringfellow JC, Hvizdos AJ, Seidehamel RJ, Cohn JB, Wilcox CS, Fabre LF, Pecknold JC, Smith WT, Rickels K: Use of buspirone in patients with generalized anxiety disorder and coexisting depressive symptoms. Neuropsychobiology *25:* 193, 1992.

Goff DC, Midha KK, Brotman AW, McCormick S, Waites M, Amico ET: An open trial of buspirone added to neuroleptics in schizophrenic patients. J Clin Psychopharmacol *11:* 193, 1991.

Grady TA, Pigott TA, L'Heuraux F, Hill JL, Bernstein SC, Murphy DL: Double-blind study of adjuvant buspirone for fluoxetine treated patients with obsessive compulsive disorder. Am J Psychiatry *150:* 810, 1993.

Gross MD: Buspirone in ADHD with OCD. J Am Acad Child Adolesc Psychiatry *34:* 1260, 1995.

Kranzler HR, Burleson JA, Del Boca FK, Babor TF, Korner P, Brown J, Bohn MJ: Buspirone treatment of anxious alcoholics. Arch Gen Psychiatry *51:* 720, 1994.

Laakmann G, Schule C, Lorkowski G, Baghai T, Kuhn K, Ehrentraut S: Buspirone and lorazepam in the treatment of generalized anxiety disorder in outpatients. Psychopharmacol Ser *136:* 357, 1998.

Lader M, Scotto JC: A multicentre double-blind comparison of hydroxyzine, buspirone and placebo in patients with generalized anxiety disorder. Psychopharmacology *139:* 402, 1998.

*Landen M, Bjorling G, Agren H, Fahlen T: A randomized, double-blind, placebo-controlled trial of buspirone in combination with an SSRI in patients with treatment-refractory depression. J Clin Psychiatry *59:* 664, 1998.

*McEvoy GK, editor: *AHFS Drug Information*. American Society of Hospital Pharmacists, Bethesda, MD, 1996.

Norden MJ: Buspirone treatment of sexual dysfunction associated with selective serotonin re-uptake inhibitors. Depression *2:* 109, 1994.

Phillips KA: An open study of buspirone augmentation of serotonin-reuptake inhibitors in body dysmorphic disorder. Psychopharmacol Bull *32:* 175, 1996.

Rickels K, Amsterdam JD, Clary C, Puzzuoli G, Schweizer E: Buspirone in major depression: A controlled study. J Clin Psychiatry *52:* 34, 1991.

*Rickels K, Schweizer E, Csanalosi I, Case WG, Chung H: Long-term treatment of anxiety and risk of withdrawal: Prospective comparison of clorazepate and buspirone. Arch Gen Psychiatry *45:* 444, 1988.

Rickels K, Freeman E, Sondheimer S: Buspirone in treatment of premenstrual syndrome. Lancet *1:* 777, 1989.

*Sheehan DV, Raj AB, Sheehan KH, Soto S, Knapp E: The relative efficacy of high dose buspirone and alprazolam in the treatment of panic disorder: A double-blind placebo controlled study. Acta Psychiatr Scand *88:* 1, 1993.

Sramek JJ, Tansman M, Suri A, Hornig-Rohan M, Amsterdam JD, Stahl SM, Weisler RH, Cutler NR: Efficacy of buspirone in generalized anxiety disorder with coexisting mild depressive symptoms. J Clin Psychiatry *57:* 287, 1996.

Stanislav SW, Fabre T, Crimson ML, Childs A: Buspirone's efficacy in organic-induced aggression. J Clin Psychopharmacol *14:* 126, 1994.

Tollefson GD, Nabtague-Clouse J, Tollefson SL: Treatment of comorbid generalized anxiety in a recently detoxified alcoholic population with a selective sertonergic drug (buspirone). J Clin Psychopharmacol *12:* 19, 1992.

*Van Vliet IM, den Boer JA, Westenberg HG, Pian KL: Clinical effects of buspirone in social phobia: A double-blind placebo-controlled study. J Clin Psychiatry *58:* 164, 1997.

Wilkinson LO, Middlemiss DN, Hutson PH: 5-HT $_{1A}$ receptor activation increases hippocampal acetycholine efflux and motor activity in the guinea pig: Agonist efficacy influences functional activity in vivo. J Pharmacol Exp Ther *270:* 656, 1994.

Yocca FD: Neurochemistry and neurophysiology of buspirone and gepirone: Interactions at presynaptic and postsynaptic 5-HT $_{1A}$ receptors. J Clin Psychopharmacol *10:* 6S, 1990.

CHAPTER 31. BIOLOGICAL THERAPIES

31.13 CALCIUM CHANNEL INHIBITORS

ROBERT M. POST, M.D.

HISTORY

Calcium dysregulation has been linked to psychiatric illness and to mood disorders in particular, with a high incidence of mood disorders accompanying calcium dysregulation in hyperparathyroidism and hypoparathyroidism. A high incidence of depression accompanies the increases of serum calcium in hyperparathyroidism and mood instability accompanies the decreases in calcium in hypoparathyroidism as well. Early preclinical work identified major disruptions in motor activity and other physiological endpoints with specific alterations of the parathyroid glands.

Whereas evidence of alterations in plasma and cerebrospinal fluid (CSF) calcium in patients with primary mood disorder has remained inconsistent and controversial, more recent evidence of increased intracellular calcium in blood elements (both platelets and lymphocytes) has been widely documented and multiply replicated. This increased intracellular calcium occurs in both the basal condition and following stimulation with serotonin, thrombin, or related agents. Steven Dubovsky, who pioneered these initial investigations, has also gone on to study the potential therapeutic implications of the calcium channel inhibitors, particularly their ability to act as antimanic agents (Table 31.13-1). Bolstering this empirical and theoretical framework implicating calcium dysregulation in affective illness are the data showing that many of the other mood-stabilizing agents effective in the treatment of bipolar I disorder, such as lithium (Eskalith), carbamazepine (Tegretol), divalproex (Depakote) or valproate (Depakene), and perhaps lamotrigine (Lamictal), also appear to exert major effects on different aspects of intracellular calcium metabolism.

Table 31.13-1 Calcium Channel Inhibitors in Mood Disorders

Because calcium is a critical intracellular transduction mechanism and is highly regulated this system becomes a key target for therapeutics (Fig. 31.13-1). Moreover, the 50 to 100 nanomolar intracellular calcium concentrations are at risk from the 10,000 times greater concentration of calcium in the extracellular fluid. With neuronal depolarization and firing, not only is there a major sodium influx but also a rapid calcium influx through voltage-gated calcium channels as well as through the ligand-gated *N*-methyl-D-aspartate (NMDA) receptor. Thus, restoring calcium homeostasis requires active calcium extrusion mechanisms, and a variety of calcium buffering and intracellular sequestration mechanisms, including those of mitochondria and endoplasmic reticulum. Recent data suggest that the mood stabilizers lithium, carbamazepine, and perhaps valproate and lamotrigine may block calcium cation (Ca^{++}) influx through the NMDA receptor.

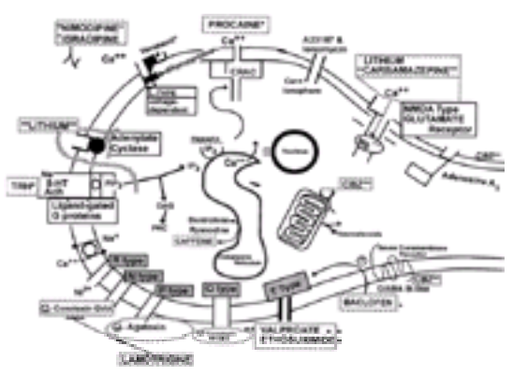

FIGURE 31.13-1 Calcium channel diversity and drug targets. This figure outlines the multiple extracellular and intracellular membrane effects of different agents (** = commonly used drugs; * = drugs explored in some psychiatric patients) and the diversity of other calcium channel regulatory sites that could be amenable to new drug development. Carbamazepine is shown to have four putative actions on different aspects of cellular biochemistry; two of lithium's many actions are highlighted, including adenylate cyclase and PI turnover.

CHEMISTRY

The L-type (long-lasting) calcium channel inhibitors have been most widely studied and have the greatest current potential for therapeutics. This channel requires neuronal depolarization for calcium entry. However, it should be noted that a variety of other calcium channels have been identified and these may ultimately also be differentially targeted for clinical therapeutics (Fig. 31.13-1). The N-type (neither L nor T) calcium channel is blocked by omega-conotoxin GVIA, the P-type (Purkinje cell) by omega-agatoxin, and the Q-type by omega-conotoxin MVIIC. T-type (transient) calcium channels have been implicated in the therapeutic action of ethosuximide and potentially of valproate in their actions on spike-and-wave absence epilepsy. Ligands acting at G-proteins hydrolyze phosphatidylinositol 4,5-bisphosphate (PIP_2) to diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP_3), the latter of which acts at a heparin-sensitive calcium channel on the endoplasmic reticulum. This action is different from the calcium-related ranodyne receptor, at which caffeine and dantrolene (Dantrium) act. Dantrolene is also used as a part of the emergency medical regimen for the treatment of malignant hyperthermia.

As noted in the *Physicians' Desk Reference* for the L-type calcium channel inhibitors, verapamil (Calan) is an almost white, crystalline powder that is practically free of odor and has a bitter taste. It is soluble in water, chloroform, and methanol and is not chemically related to other cardioactive drugs. Chemically, verapamil is

benzeneacetonitrile, a-[3-[[2-(3,4-dimethoxyphenyl)ethyl]-methylamino]propyl]-3,4-dimethoxy-a-(1-methyl-ethyl)hydrochloride. The chemical formula of the phenylalkylamine is $C_{27}H_{38}N_2O_4$ (Fig. 31.13-2) and its molecular weight is 454.59.

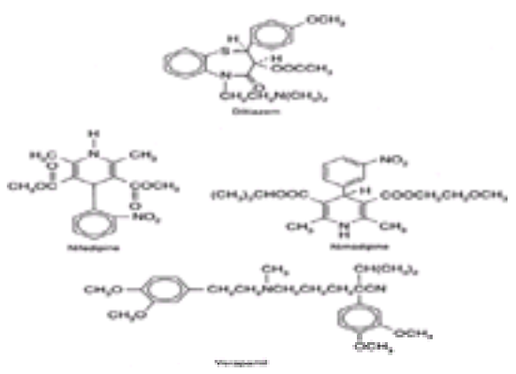

FIGURE 31.13-2 Molecular structures of calcium channel inhibitors

Diltiazem (Cardizem) is a white to off-white crystalline powder with a bitter taste. It is soluble in water, methanol, and chloroform. Chemically, diltiazem is 1,5-benzothiazepin4(5H)-one,3-(acetyloxy)5 - [2 - (dimethylamino) - ethyl] - 2,3 - dihydro - 2 - (4-methoxy-phenyl) - , monohydrochloride,(+)- *cis*. Its chemical formula is $C_{22}H_{27}ClN_2O_4S$ (Fig. 31.13-2). This benzothiazepine has a molecular weight of 450.98.

Nifedipine (Procardia) is 3,5-pyridinedicarboxylic acid, 1,4-dihydro-2,6-dimethyl-4-(2-nitrophenyl)-, dimethyl ester. The chemical formula is $C_{17}H_{18}N_2O_6$ (Fig. 31.13-2). It is a yellow crystalline substance that is practically insoluble in water, but soluble in ethanol. It has a molecular weight of 346.34.

Nimodipine (Nimotop) is isopropyl (2-methoxyethyl) 1,4-dihydro-2,6-dimethyl-4-(3-nitrophenyl)-3,5-pyridine-dicarboxylate. It has a molecular weight of 418.5, and its chemical formula is $C_{21}H_{26}N_2O_7$ (Fig. 31.13-2). Nimodipine is a yellow crystalline substance that is practically insoluble in water. Its closely related dihydropyridine congeners include isradipine (DynaCirc) and amlodipine (Norvasc).

Amlodipine is (RS) 3-ethyl-5-methyl-2-(2-amino-ethoxymethyl)-4-(2-chlorophenyl)-1,4-dihydro-6-methyl-3,5-pyridine dicarboxylate benzenesulphonate. It has a molecular weight of 567.1, and its chemical formula is $C_{20}H_{25}ClN_2H_6O_3S$. Amlodipine is a white crystalline powder, slightly soluble in water, and sparingly soluble in ethanol.

Flunarizine is (E)-1-[bis (4-fluorophenyl)methyl]-4-(3-phenyl-2-propenyl)-piperazine. Its chemical formula is $C_{26}H_{26}F_2N_2$ (Fig. 31.13-2) and its molecular weight is 404.51.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Calcium channel inhibitors are well absorbed by the gastrointestinal tract, each being substantially metabolized by a hepatic first-pass effect.

Distribution As with all hepatically metabolized drugs, there are considerable intraindividual and interindividual variations in the plasma concentrations of the calcium channel inhibitors. Verapamil crosses the placenta and can be detected in umbilical vein blood at delivery.

Studies in rats have shown significant concentrations of calcium channel blockers in maternal milk following oral administration. It is recommended that women who wish to breastfeed should not take these drugs.

Elimination The half-life of verapamil after the first dose is 2 to 8 hours, but it increases to 5 to 12 hours after the first few days of therapy. Verapamil enters the central nervous system (CNS) poorly and attains concentrations approximately 0.05 percent those in plasma. The half-lives of the dihydropyridines are short for nimodipine (1 to 2 hrs) and isradipine (DynaCirc) (1½–2 hrs), but longer for amlodipine (30 to 50 hrs), making once-daily or twice-daily dosing more feasible with the latter compound (Table 31.13-2).

	Phenylalkylamine (on outside of Ca^{++} channel)	Dihydropyridines (inside Membrane and Ca^{++} Channel)		
	Verapamil (Calan, Isoptin)	Nimodipine (Nimotop)	Isradipine (DynaCirc)	Amlodipine (Norvasc)
Half-life	Short (5–12 hr)	Short (1–2 hr)	Short (1 1/2–2 hr)	Long (30–50 hr)
Starting dosage	30 mg t.i.d.	30 mg t.i.d.	2.5 mg b.i.d.	5 mg q.a.
Peak daily dosage	480 mg	240–480 mg	15 mg	10–15 mg
Antimanic	++	++	(++)	*
Antidepressant	±	+	(+)	*
Antischizoaffective	±	++	(++)	*
Anticocaine (Dopamine)	-	++	++	*

*No systematic studies, only case reports and clinical observations

Table 31.13-2 Half-Lives, Dosages, and Effectiveness of L-Type Calcium Channel Inhibitors in Mood Disorders

Pharmacodynamics

Mechanism of Action There are several subtypes of these voltage-dependent blockers of calcium influx that modulate the relatively long opening times of these channels. These compounds include the widely recognized phenylalkylamine type characterized by verapamil, the benzothiazepine type characterized by diltiazem, and the dihydropyridine type characterized by nifedipine, nimodipine, isradipine, amlodipine, nicardipine (Cardene), and nitrendipine. Even though these agents all potently bind to this voltage-dependent calcium channel and act to inhibit calcium influx, their biochemical and physiological effects are remarkably different (Table 31.13-3). In addition, not all dihydropyridines are equivalent. For example, nimodipine, but not nifedipine, abolishes calcium oscillations in hypothalamic neurons, and amlodipine, but not nimodipine, has prominent antioxidant properties.

Effect	Dihydropyridine Nimodipine, Isradipine	Phenylalkylamine Verapamil	References
Lipid Solubility	+ +	±	Freedman & Waters, 1987
Anticonvulsant:			
Animals:			
Kindling	+ +	±	Hunjel & Iyer, 1994 Vezzani et al., 1988
Reperfusion	+ +	0	Meyer et al., 1986a
Kainic	+	0	Vezzani et al., 1988 Pazzagli et al., 1992
Patients	+ +	NA	Brandi et al., 1988 Lakin et al., 1991 de Falco et al., 1992
Block cocaine-induced:			
Hyperactivity	+ +	0	Pani et al., 1990, 1994
Sensitization	+ +	NA	Weiss et al., unpublished
Dopamine overflow	+ +	0	Pani et al., 1990, 1991
Positive in kindled-mouse model	+ +	0	Czuczi et al., 1989 de Jonge et al., 1991 Mogilnicka et al., 1987, 1988
Positive in learned-helplessness model	+ +	0	Cienfuegos et al., 1988 Martin et al., 1989 de Jonge et al., 1991
Like antidepressants:			
Glycine inhibition of ^{3}[H] 5,7-DCKA Binding	+ +	NA	Nowak et al., 1992
Inositol cycling (electrical fibers)	+ +	±	Pazzaglia et al., 1991

Table 31.13-3 Effects of Dihydropyridine L-Type Calcium Channel Blockers Versus Verapamil

Verapamil and the phenylalkylamines are charged and gain access to the inside of the calcium channel less readily than most of the dihydropyridines, which are uncharged (except amlodipine). These different membrane properties and binding characteristics result in a profile of effects for the dihydropyridines compared with the phenylalkylamines, including increased lipid solubility; anticonvulsant effects on kindling and hyperfusion models and clinical seizures; the ability to block cocaine hyperactivity, behavioral sensitization, and its associated dopamine overflow in the nucleus accumbens (whereas verapamil is inactive); positive effects in two animal models of depression, including the forced swim test and learned helplessness model; decreased glycine inhibition of ^{3}[H] 5,7-DCKA binding similar to that of other antidepressant modalities; and possibly a different effectiveness on mood in some patients with bipolar disorder. Blood concentration and relation to action have not been clearly established.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Design Verapamil has been reported effective in the treatment of acute mania. Open case studies were followed with a small series of double-blind studies as well as seven studies comparing verapamil to placebo or to lithium. The early studies were generally positive in acute mania, showing effects equal to those of lithium, and in several studies, better than placebo. However, more recent studies have revealed more equivocal results, and despite the generally positive controlled literature, verapamil remains rarely used in the general therapeutics of bipolar disorders. Moreover, in a placebo-controlled study, verapamil (like placebo) was distinctly less effective than either standard antidepressant treatment or amitriptyline (Elavil) in acute depression.

Given continuing uncertainties regarding verapamil's place in the clinical therapeutics of the mood disorders, other groups have begun to study the dihydropyridines more intensively because of their potentially different pharmacological and therefore therapeutic profiles. One study at the National Institute of Mental Health (NIMH) reported improvement in six mania patients in an open trial, and another study reported improvement in approximately one-third of a heterogeneous group of treatment refractory patients with recurrent mood disorders using double-blind, placebo substitution in B-A-B-A trials to reconfirm responsiveness. In the latter study, several of the best responders had ultra–ultra-rapid (or “ultradian”) fluctuations in mood (i.e., multiple switches within a 24-hour period).

Interpretation In an updated series of the previous NIMH study, 10 of 30 patients were reported to show a clinically relevant degree of response to nimodipine monotherapy, as rated by a marked or moderate response on the Clinical Global Impressions (CGI) scale. One such patient is illustrated in Figure 31.13-3A in which prospective daily life chart ratings of mild, moderate, and severe mania and depression are illustrated by nursing ratings blind to treatment with nimodipine or placebo. As illustrated, gradual titration of nimodipine to average dosages of 630 mg a day resulted in the complete amelioration of a moderate to severe depressive episode. Response was verified with placebo substitution, which resulted in a pattern of ultra-rapid mood fluctuations that were again ameliorated by the institution of nimodipine on a blind basis. The patient was able to be transferred from one dihydropyridine to another (nimodipine to isradipine) without loss of efficacy.

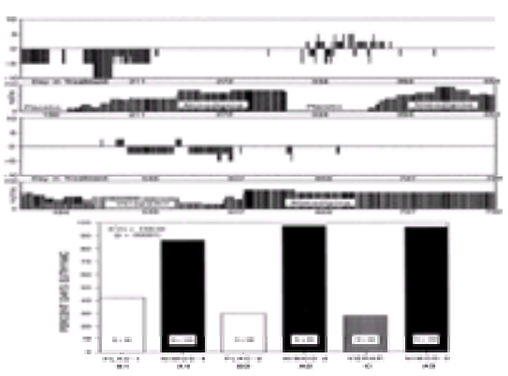

FIGURE 31.13-3 A. Prospective daily life chart ratings of mood (above) by nurses blind to medication status (below). Boxes above the horizontal line (euthymia) represent mild, moderate, or severe mania. Boxes below the line represent mild, moderate, or severe depression. A complete record of all her medications for each day of the study is depicted below each life chart line. Each line represents 1 day. Numbers indicating the day of the patient's clinical trial are in 2-month intervals. **B**. Efficacy of nimodipine in a female with bipolar II disorder. Daily life chart ratings of the same patient as illustrated in **A**. The numbers in each bar are the number of days during that phase of the study that were included in data analyses. For example, medication phases did not include periods of titration or overlap with another psychotropic medication. The statistic represents the percentage of days euthymic during each of the three double-blind nimodipine phases compared to the percentage of days euthymic during the two placebo phases as well as the blind verapamil phase.

Also demonstrated in this patient was the inability of verapamil at maximally tolerated dosages of 320 mg a day to maintain clinical improvement, which was reestablished with the third institution of nimodipine treatment on a blind basis. This patient was euthymic 85 to 95 percent of her days on full dosages of nimodipine compared with being euthymic 30 to 40 percent of her days during periods of placebo or verapamil administration (Fig. 31.13-3B). The patient had previously received multiple clinical trials with a variety of antidepressant agents and lithium without success; her depression had been complicated by comorbid symptoms of generalized anxiety and panic disorder. Her mean Hamilton Depression ratings decreased from 28 during the placebo phase to the 10 to 15 range during the latest nimodipine phase. Similarly, on this blind therapy, her Beck Depression Inventory scores were halved, from approximately 10 to under 5. The patient has continued to remain markedly improved on isradapine 3 years after discharge from the inpatient and outpatient programs at the NIMH.

This patient's detailed case response, performed on a blind basis, as well as the second patient described below, suggest the possibility that the same degree of clinical responsivity achieved by the dihydropyridine L-type calcium channel blockers may not be readily substituted for by the more widely used phenylalkylamine L-type calcium channel blocker, verapamil.

Figure 31.13-4A illustrates a patient with marked ultradian mood fluctuations in whom significant amplitude decreases in manic and depressive oscillations were observed when placebo (B) was substituted for nimodipine (A). This decrease was verified as not being attributable to spontaneous improvement in the course of illness because a second phase of placebo substitution resulted in a renewal of both manic and depressive severity, which again improved with the blind reintroduction of nimodipine. Only a 1-month sampling of the last ratings in the trial is illustrated in the figure, and a more complete picture of the entire trial is summarized in Figure 31.13-4B.

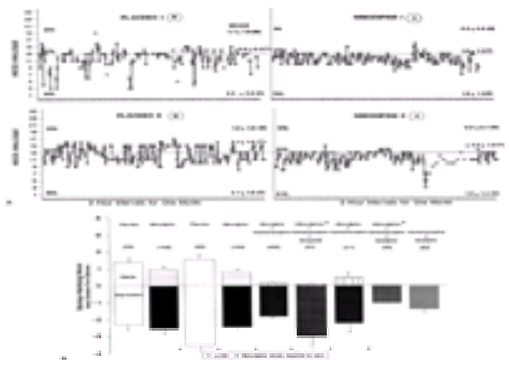

FIGURE 31.13-4 A. Effect of nimodipine in an ultra–ultrarapid (ultradian) cycling patient. Twice-daily to two-hourly mood analogue ratings show a partial response to nimodipine, relapse after placebo substitution, and renewed response in this female with bipolar II disorder. Only the last month of each double-blind phase is illustrated. Mean depression and mania ratings (as a difference from euthymia) are given for mania (M) above the euthymia mood line or for depression (D) below the line. **B**. Efficacy of dihydropyridine L-type calcium channel inhibitors in a female with bipolar II disorder (the same patient illustrated in **A**). Mean deviation from euthymia ratings (number of days in parentheses) in a bipolar II ultra–ultra-rapid cycling patient showing efficacy of nimodipine monotherapy; efficacy of nimodipine-carbamazepine combination therapy; unsuccessful transition from nimodipine to verapamil; successful reinstitution of nimodipine-carbamazepine combination therapy; and finally, successful transition to isradipine-carbamazepine combination therapy.

This figure not only demonstrates the improvement in mania and depression observed during the trial, but also additional improvement when carbamazepine (Tegretol) was added to the incomplete response to nimodipine. At this time, the patient was able to resume her full-time employment as an artist and teacher whereas she had been unable to do so in the prior 4 years, during which time she was treated with lithium as well as the lithium-carbamazepine combination.

Interactions In an attempt to ascertain whether this patient would respond equally well to the related phenylalkylamine L-type calcium channel inhibitor verapamil, a slow, blind crossover was attempted to achieve maximally tolerated doses of this agent (320 mg a day). This crossover resulted in mood destabilization, and the patient was again successfully reintroduced to nimodipine in combination with carbamazepine on a blind basis (Fig. 31.13-4B). The substitution of nimodipine for another dihydropyridine L-type calcium channel inhibitor, isradipine, was successful and the patient was discharged on this combination. This patient's clinical response indicates a partial amelioration of ultradian cycling amplitude with the dihydropyridine L-type calcium channel inhibitor and the need for further augmentation with another mood stabilizer, such as carbamazepine. However, such augmentation with carbamazepine was effective in only 4 of 14 patients overall.

Carbamazepine blocks calcium influx through the NMDA receptor, and it remains to be determined whether two different mechanisms converge on calcium dysregulation or whether some other mechanism of action of carbamazepine could account for its additive effects in this patient. An open prophylactic study reported that patients with bipolar disorders responded better to the combination of nimodipine plus lithium than either agent alone. Which of lithium's panoply of actions account for this additive effect also remains to be determined.

Recurrent Brief Depressive Disorder Recent studies have described a syndrome, recurrent brief depressive disorder characterized by severe, rapidly recurring, brief episodes of depression that do not meet the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria for major depressive disorder because of their brevity. These episodes average 3 days or less, are associated with considerable morbidity and some mortality through suicide, and appear more common in women than men (although they are not linked to the menstrual cycle). The incidence of recurrent brief depression in the general population has been reported to be as high as 8 to 10 percent, although the interpretation of this number is obscured by the fact that the pattern of recurrent brief depression can be to some extent interchangeable with the pattern of more consistent periods of major depressive disorder. Nonetheless, initial clinical trials suggest the lack of adequate responsivity of the recurrent brief depression pattern to standard antidepressant treatments, including the tricyclics, monoamine oxidase inhibitors (MAOIs), selective serotonin reuptake inhibitors (SSRIs), and lithium.

Several patients with recurrent brief depression studied appeared to show an excellent response to nimodipine. One such patient is illustrated in Figure 31.13-5, and her clinical response at average maintenance doses of 270 mg a day was confirmed by exacerbation of her recurrent brief depression pattern upon placebo substitution and then renewed responsivity to blind reinstitution of nimodipine. It is particularly noteworthy, however, that in contrast to the bipolar I disorder patients illustrated in Figure 31.13-3 and Figure 31.13-4, this patient with recurrent brief depression did sustain her clinical responsivity during a blind crossover to verapamil and was able to be discharged on this compound. These preliminary but double-blind controlled observations suggest the potential utility of L-type calcium channel inhibitors in recurrent brief depressive disorder and large, controlled clinical trials are clearly indicated.

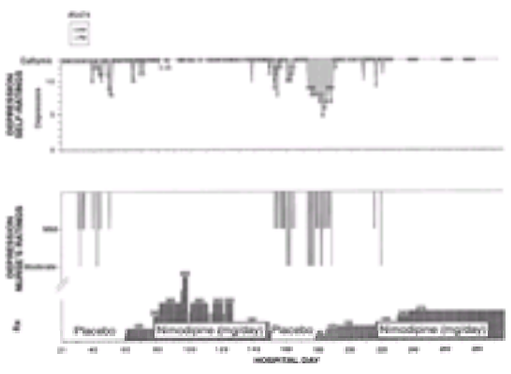

FIGURE 31.13-5 Response to nimodipine in recurrent brief depression. Patient's depression rating (**top**) and nurse's rating (**bottom**) reflect improvement in recurrent brief depression from placebo to nimodipine, with a recurrence of symptoms upon placebo substitution and re-response in the last phase.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Although no well-controlled studies of the calcium channel inhibitors in large numbers of pregnant women exist, no clearcut teratological effects on the human fetus have been reported, and their relative safety in pregnancy has been suggested. The ultimate role for the dihydropyridines and other L-type calcium channel inhibitors as alternatives to lithium (with its low risk for Ebstein's anomaly) or the mood-stabilizing anticonvulsant drugs carbamazepine and valproate (with their risk of spina bifida) during pregnancy remains to be systematically investigated.

THERAPEUTIC INDICATIONS

Calcium channel blockers are effective antihypertensive agents and are widely used for that purpose as well as in the treatment of ischemic heart disease and migraine. However, nimodipine is approved for use only in subarachnoid hemorrhage. The major use of the calcium channel inhibitors in psychiatry is currently in the area of treatment of bipolar disorders with limited use in depression and recurrent brief depressive disorder.

PRECAUTIONS AND ADVERSE REACTIONS

The most common adverse effects associated with calcium channel inhibitors are associated with vasodilation and include dizziness, headache, skin flushing, tachycardia, nausea, dysaesthesias, and peripheral edema. Verapamil and diltiazem in particular can induce bradycardia and atrioventricular heart block. In older patients the hypotensive, bradycardic, and conduction effects of calcium channel inhibitors require careful monitoring by the clinician. Other symptoms commonly reported include constipation, fatigue, rash, coughing, and wheezing. Rare adverse effects that have been reported with diltiazem include hyperactivity, akathisia, and

parkinsonism; with nifedipine (Procardia), depression; with verapamil, delirium, hyperprolactinemia, and galactorrhea; and with flunarizine (Sibelium), parkinsonism.

The calcium channel inhibitors are relatively well tolerated and considerable clinical experience has been gained by using them in the treatment of hypertension, arrhythmias, and other medical syndromes. The most common dose-limiting adverse effects of nimodipine include lightheadedness, gastrointestinal upset, subjective sense of tightness in the chest (unaccompanied by electrocardiographic changes), and flushing. The sense of warmth associated with erythema has not been uncommon with nimodipine treatment, particularly on the anterior aspects of the lower extremities.

Some investigators have advocated the potential use of the calcium channel blockers as a substitute for lithium in pregnancy in light of the lack of evidence for calcium channel inhibitor teratogenicity. Lithium has a potential for major cardiac anomalies such as Ebstein's anomaly in pregnancy, and valproate and carbamazepine have risks for spina bifida. There is recent evidence that an increased incidence of Ebstein's anomaly with lithium may have been overestimated, and the safety of continued use of lithium during pregnancy is being re-evaluated.

Recent evidence widely cited in the lay press has suggested some reasons for caution in routinely using L-type calcium channel inhibitors for hypertension and other indications, because of the potential for increasing the risk for reinfarction compared with b-adrenergic receptor antagonists, and increasing the risk of a variety of malignancies. Others argue that the data are far from conclusive and in many instances opposing views have been suggested that calcium channel inhibitors are able to decrease cancer-inducing mechanisms. Much of the initial work used case-control methods, not prospective randomized studies. Moreover, the general perspective in cardiology is that the short-acting nifedipine compounds should be relatively avoided, but not necessarily the longer-acting preparations. In support of this view is a recent positive study of the efficacy of the longer-acting amilodipine on blood pressure and on increased long-term survival. Thus, many patients continue to remain on calcium channel inhibitors for blood pressure and various other cardiological purposes, and it would appear reasonable to continue to pursue these agents for their potential role in the mood disorders, both as lithium substitutes and as alternatives with a different potential range of therapeutic effects.

DRUG INTERACTIONS

b-adrenergic receptor antagonists in combination with verapamil and diltiazem have been reported to cause cardiac depression. Caution should also be used with inhaled anesthetics, digoxin (Lanoxin), quinidine (Duraquin), cimetidine (Tagamet), and other antihypertensive agents. Neurotoxicity, choreoathetosis, parkinsonism, and bradycardia have been reported with use in combination with lithium.

LABORATORY INTERFERENCES

Verapamil and diltiazem (but not the dihydropyridines) increase carbamazepine concentrations and potential resultant neurotoxicity.

DOSAGE AND ADMINISTRATION

Verapamil is available in 40-, 80-, and 120-mg tablets. The starting dosage is 80 mg given orally three times a day; it can be raised in increments every 4 to 5 days, up to 80 to 120 mg three to four times a day. The patient's blood pressure, pulse, and electrocardiogram (in patients over 40 years old or with a history of cardiac illness) should be followed routinely. Diltiazem is available in 30-, 60-, 90-, and 120-mg tablets. The starting dosage is 30 mg given orally four times a day and can be increased to a maximum of 360 mg a day. Nifedipine is available in 10-, 20-, 30-, 60-, and 90-mg tablets. The starting dosage is 10 mg given orally three or four times a day and can be increased to a maximum dosage of 180 mg a day. Nimodipine is available in 30-mg capsules. The oral dosage is 30 mg three times a day with slow increases in dosage to 60 mg given every 4 hours or higher if tolerated. Isradipine is available in 2.5-, 5-, and 10-mg capsules or tablets. The starting dosage is 2.5 mg a day and can be increased to a maximum dosage of 15 mg a day in divided doses. Amlodipine is available in 2.5-, 5-, and 10-mg tablets. The starting dosage is 5 mg a day at night with a maximum dosage of 10 to 15 mg a day.

Investigational Approaches and Prediction of Response With the panoply of drugs now available for the short-term and long-term treatment of bipolar disorders, one major problem is determining how to sequence these drugs in the absence of clinical and biological markers that might enhance the rapidity with which a given drug is optimally matched to a given patient. There are few clinical or biological markers of response to the calcium channel blockers. Case series have suggested that lithium responders were most likely to also respond to verapamil, but some patients have responded to nimodipine after failing lithium. Among the nimodipine responders were patients with patterns of illness that are often refractory to conventional treatment, including those with rapid, ultra-rapid, and ultradian presentations as well as recurrent brief depression.

A promising area of investigation is work using brain imaging that found that patients with mood disorders with the classical pattern of relative frontal hypometabolism in major depressive disorder were among those who responded to nimodipine whereas equally depressed patients with hypermetabolism, particularly in frontal areas and the left insula, responded to carbamazepine. These data raise the possibility that regional topographies of blood flow or metabolism might ultimately help to identify a subgroup of patients who are more responsive to the calcium channel inhibitors. Responders to nimodipine had significantly lower concentrations of somatostatin in their cerebrospinal fluid (CSF) compared with nonresponders. Since nimodipine significantly increases somatostatin in the CSF, it is possible that those who have low basal concentrations are helped more by this drug action than those with normal or high concentrations.

Given the substantial percentage of depressed patients who show increased intracellular calcium alterations during or in bipolar disorder patients even between depressive episodes, it remains to be seen whether this neurobiological alteration, consistently observed on a statistical basis across many studies compared with normal volunteers, will provide a basis for helping to select those patients who respond to interventions that block calcium influx such as that of the L-type calcium channel inhibitors. In addition, exploring the molecular basis for the increased intracellular calcium in depressed patients may also help to locate new targets for therapeutic medications.

SUGGESTED CROSS-REFERENCES

Basic science discussions of electrophysiology are found in Section 1.8 and Section 1.9; the biochemical and clinical aspects of the mood disorders are discussed in Section 14.3 and Section 14.6 respectively. Section 2.1 addresses the clinical aspects of neuropsychiatric disorders.

SECTION REFERENCES

Angst J, Merikangas K, Scheidegger P, Wicki W: Recurrent brief depression: A new subtype of affective disorder. J Affect Disord *19:*87, 1990.

Brunet G, Cerlich B, Robert P, Dumas S, Souetre E, Darcourt G: Open trial of a calcium antagonist, nimodipine in acute mania. Clin Neuropharmacol *13:*224, 1990.

Carman JS, Wyatt ES, Smith W, Post RM, Ballenger JC: Calcium and calcitonin in bipolar illness. In *Neurobiology of Mood Disorders*, RM Post, JC Ballenger, editors. Williams & Wilkins, Baltimore, 1984.

Delisi SM, Konopka LM, O'Connor FL, Crayton JW: Platelet cytosolic calcium responses to serotonin in depressed patients and controls: Relationship to symptomatology and medication. Biol Psychiatry *43:*327, 1998.

de Vry J, Fritze J, Post RM: The management of coexisting depression in patients with dementia: Potential of calcium channel antagonists. Clin Neuropharmacol *20:*22, 1997.

Dubovsky SL: Calcium antagonists for manic-depressive illness. Neuropsychobiology *27:*184, 1993.

Dubovsky SL: Calcium channel antagonists as novel agents for manic-depressive disorder. In *Textbook of Psychopharmacology*, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1995.

*Dubovsky SL, Buzan RD: Novel alternatives and supplements to lithium and anticonvulsants for bipolar affective disorder. J Clin Psychiatry *58:*224, 1997.

Dubovsky SL, Franks RD, Allen S: Calcium antagonists in mania: A double-blind study of verapamil. Psychiatry Res *18:*309, 1986.

Dubovsky SL, Murphy J, Thomas M, Rademacher J: Abnormal intracellular calcium ion concentration in platelets and lymphocytes of bipolar patients. Am J Psychiatry *149:*118, 1992.

Dunn RT, Frye MS, Kimbrell TA, Denicoff KD, Leverich GS, Post RM: The efficacy and use of anticonvulsants in mood disorders. Clin Neuropharmacol *21:*215, 1998.

Emamghoreishi M, Schlichter L, Li PP, Parikh S, Sen J, Kamble A, Warsh JJ: High intracellular calcium concentrations in transformed lymphoblasts from subjects with bipolar I disorder. Am J Psychiatry *154:*976, 1997.

Garza-Trevino ES, Overall JE, Hollister LE: Verapamil versus lithium in acute mania. Am J Psychiatry *149:*121, 1992.

Giannini AJ, Loiselle RH, Price WA, Giannini MC: Comparison of antimanic efficacy of clonidine and verapamil. J Clin Pharmacol *25:*307, 1985.

Giannini AJ, Taraszewski R, Loiselle RH: Verapamil and lithium in maintenance therapy of manic patients. J Clin Pharmacol *27:*980, 1987.

Grunze H, von Wegerer J, Greene RW, Walden J: Modulation of calcium and potassium currents by lamotrigine. Neuropsychobiology *38:*131, 1998.

Grunze H, Walden J: Reduction of the frequency of occurrence of low magnesium induced field potentials in the hippocampus slice preparation of guinea pigs: A good screening tool for calcium antagonistic effects of anticonvulsant and antipsychotic drugs. Magnes Res *10:*119, 1997.

Hoschl C, Kozeny J: Verapamil in affective disorders: A controlled, double-blind study. Biol Psychiatry *25:*128, 1989.

Hough CJ, Irwin RP, Gao X-M, Rogawski MA, Chuang D-M: Carbamazepine inhibition of N-methyl-D-aspartate-evoked calcium influx in rat cerebellar granule cells. J Pharmacol Exp Ther *276:*143, 1996.

Jacobson SJ, Jones K, Johnson K, Ceolin L, Kaur P, Sahn D, Donnenfeld AE, Rieder M, Santelli R, Smythe J, Pastuszak A, Einarson T, Koren G: Prospective multicentre study of pregnancy outcome after lithium exposure during first trimester. Lancet *339:*530, 1992.

Janicak PG, Sharma RP, Pandey G, Davis JM: Verapamil for the treatment of acute mania: A double-blind, placebo-controlled trial. Am J Psychiatry *155:*972, 1998.

Jimerson DC, Post RM, Carman JS, Van Kammen DP, Wood JH, Goodwin FK, Bunney WE Jr: CSF calcium: Clinical correlates in affective illness and schizophrenia. Biol Psychiatry *14:*37, 1979.

Kaplan NM: Do calcium antagonists cause cancer? (Letter; comment). Lancet *348:*541, 1996.

Ketter TA, Kimbrell TA, George MS, Stein RM, Willis MW, Benson BE, Frye MA, Cora-Locatelli G, Post RM: Baseline hypermetabolism may predict carbamazepine response, and hypometabolism nimodipine response in mood disorders. Abstracts of the XXth CINP Congress: 10, 1996.

*Magee LA, Schick B, Donnenfeld AE, Sage SR, Conover B, Cook L, McElhatton PR, Schmidt MA, Koren G: The safety of calcium channel blockers in human pregnancy: A prospective, multicenter cohort study. Am J Obstet Gynecol *174:*823, 1996.

Manna V: Bipolar affective disorders and role of intraneuronal calcium. Therapeutic effects of the treatment with lithium salts and/or calcium antagonist in patients with rapid polar inversion. Minerva Med *82:*757, 1991.

*Mason RP, Leeds PR, Jacob RF, Hough CJ, Zhang KG, Mason PE, Chuang DM: Inhibition of neuronal apoptosis by the charged calcium antagonist amlodipine and antioxidants in rat cerebellar granule cells. J Neurochem, 1999, in press.

Mikuni M, Kagaya A, Takahashi K, Meltzer HY: Serotonin but not norepinephrine-induced calcium mobilization of platelets is enhanced in affective disorders. Psychopharmacology *106:*311, 1992.

Montgomery SA, Montgomery D, Baldwin D, Green M: The duration, nature and recurrence rate of brief depressions. Prog Neuropsychopharmacol Biol Psychiatry *14:*729, 1990.

Mrsic J, Zupan G, Erakovic V, Simonic A, Varljen J: The influence of nimodipine and MK-801 on the brain free arachidonic acid level and the learning ability in hypoxia-exposed rats. Prog Neuropsychopharmacol Biol Psychiatry *21:*345, 1997.

Nonaka S, Hough CJ, Chuang DM: Chronic lithium treatment robustly protects neurons in the central nervous system against excitotoxicity by inhibiting *N*-methyl-D-aspartate receptor-mediated calcium influx. Proc Natl Acad Sci USA *95:*2642, 1998.

Pazzaglia PJ, Post RM, Ketter T, Callahan AM, Marangell LB, Frye MA, George MS, Kimbrell TA, Leverich GS, Cora-Locatelli G, Luckenbaugh D: Nimodipine monotherapy and carbamazepine augmentation in patients with refractory recurrent affective illness. J Clin Psycho Pharmacol *18:*404, 1998.

*Pazzaglia PJ, Post RM, Ketter TA, George MS, Marangell LB: Preliminary controlled trial of nimodipine in ultra-rapid cycling affective dysregulation. Psychiatry Res *49:*257, 1993.

*Post RM, Pazzaglia PJ, Ketter TA, Denicoff K, Weiss SRB, Hough C, Chuang D-M, Stein R, Frye M: Carbamazepine and nimodipine in affective illness: Efficacy, mechanisms of action, and interactions. In *Pharmacotherapy of Mood and Cognition*, vol 4, S Montgomery, U Halbreich, editors. American Psychiatric Press, Washington, DC, 1999.

Post RM, Weiss SRB, Chuang D, Ketter TA: Mechanisms of action of carbamazepine in seizure and affective disorders. In *Anticonvulsants in Mood Disorders*, RT Joffe, JR Calabrese, editors. Marcel Dekker, New York, 1994.

*Post RM, Weiss SRB, Clark M, Chuang DM, Hough C, Li H: Lithium, carbamazepine, and valproate in affective illness: Biochemical and neurobiological mechanisms. In *Mechanisms of Action of the Mood Stabilizers Lithium, Carbamazepine, and Valproate in Affective Illness*, H Manji, editor. American Psychiatric Press, Washington, DC, 1999.

Richter CP: Two day cycles of alternating good and bad behavior in psychotic patients. Arch Neurol Psychiatry *39:*587, 1938.

Schumacher TB, Beck H, Steinhauser C, Schramm J, Elger CE: Effects of phenytoin, carbamazepine, and gabapentin on calcium channels in hippocampal granule cells from patients with temporal lobe epilepsy. Epilepsia *39:*355, 1998.

*Snyder SH, Reynolds IJ: Calcium-antagonist drugs. Receptor interactions that clarify therapeutic effects. N Engl J Med *313:*995, 1985.

Sze KH, Sim TC, Wong E, Cheng S, Woo J: Effect of nimodipine on memory after cerebral infarction. Acta Neurol Scand. *97:*386–392, 1998.

von Wegerer J, Hesslinger B, Berger M, Walden J: A calcium antagonistic effect of the new antiepileptic drug lamotrigine. Eur Neuropsychopharmacol *7:*77, 1997.

Walden J, Grunze H, Bingmann D, Liu Z, Dusing R: Calcium antagonistic effects of carbamazepine as a mechanism of action in neuropsychiatric disorders: Studies in calcium-dependent model epilepsies. Eur Neuropsychopharmacol *2:*455, 1992.

Walden J, Wegerer JV, Roed I, Berger M: Effects of the serotonin-1A agonists buspirone and ipsapirone on field potentials in the hippocampus slice: Comparison with carbamazepine and verapamil. Eur Neuropsychopharmacol *5:*57, 1995.

Walton SA, Berk M, Brook S: Superiority of lithium over verapamil in mania: A randomized, controlled, single-blind trial. J Clin Psychiatry *57:*543, 1996.

Wang SJ, Tsai JJ, Gean PW: Lamotrigine inhibits depolarization-evoked Ca++ influx in dissociated amygdala neurons. Synapse *29:*355, 1998.

Zanchetti A: The calcium channel blocker controversy in perspective. Cardiology *88:*66, 1997.

CHAPTER 31. BIOLOGICAL THERAPIES

31.14 CHLORAL HYDRATE

LAWRENCE A. LABBATE, M.D., GEORGE W. ARANA, M.D., AND JAMES C. BALLENGER, M.D.

HISTORY

Chloral hydrate was discovered by Justus von Liebig in 1831 and was determined to have hypnotic properties by Rudolph Buchheim in 1861. It was introduced into clinical use in the late nineteenth century, at first as an anesthetic and subsequently as a hypnotic, for which it had widespread use. Although initially thought to induce its effects by metabolism to chloroform, chloral hydrate's active metabolite turned out to be trichloroethanol. Until the recent common use of benzodiazepines it was often used as a hypnotic in medical and surgical practice. Further, until more effective and specific agents were developed, it was also used to treat many psychiatric conditions including anxiety, psychosis, mania, and delirium. Chloral hydrate has been largely replaced by the benzodiazepines and zolpidem (Ambien) in the clinical treatment of insomnia. The benzodiazepines have a much higher therapeutic index and are therefore preferred. Chloral hydrate is sometimes used as a hypnotic agent, although it is not the standard of care in psychiatric practice. It is currently used more commonly in pediatrics as a sedative for procedures, including imaging studies or dental work.

Chemistry Chloral hydrate is formed by adding water to the carbonyl group of chloral, 2,2,2-trochloroacetaldehyde. Chloral hydrate serves largely as a pro-drug. It is metabolized by alcohol dehydrogenase to its active metabolite trichloroethanol (CCl_3CH_2OH).

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Chloral hydrate is completely absorbed in the small intestine, rapidly reduced to trichloroethanol, and effects are generally manifest in 30 to 60 minutes. Food in the stomach slows onset of action.

Distribution Chloral hydrate and trichloroethanol are lipid soluble and enter all tissues, including breast milk, although concentration in breast milk is usually low and not clinically significant.

Elimination Chloral hydrate is rapidly metabolized in the liver by alcohol dehydrogenase. Ethanol accelerates metabolism because its oxidation reduces chloral hydrate and thus leads to more rapid central nervous system (CNS) effects. The active drug, trichloroethanol, has a plasma half-life ranging from 4 to 8 hours. Trichloroethanol is primarily metabolized in the liver by conjugation with glucuronic acid to form urochloralic acid and is then excreted into urine and bile. To a lesser extent, trichloroethanol and chloral hydrate are oxidized to trichloroacetic acid and excreted in urine. Trichloroacetic acid is not an active drug and has a half-life of approximately 4 days.

Pharmacodynamics

Mechanism of Action Pharmacological effects are attributable to trichloroethanol. Limited study in animals suggests that, like the barbiturates trichloroethanol exerts g-aminobutyric acid A (GABA) receptor channel, and potentiates GABAergic transmission. Chloral hydrate thus serves to increase chloride ion flow and depress excitable cells.

Blood Levels Because chloral hydrate is a prodrug, plasma level of trichloroethanol are measured. Typical clinical doses, 500 to 1500 mg in adults or 15 mg/kg in children, are associated with peak plasma concentrations of trichloroethanol of 7 to 12 µg/mL.

CLINICAL DRUG STUDIES

Interpretation and Design Chloral hydrate was introduced to clinical practice long before Food and Drug Administration (FDA) requirements were in place for new drugs. As a result there were no placebo-controlled studies for its use, even though it has historically been used as an antianxiety or hypnotic agent or in the treatment of alcohol withdrawal. Long after its introduction a number of double-blind studies compared the use of benzodiazepines with chloral hydrate for the treatment of insomnia. A number of open-label studies have also evaluated chloral hydrate in children undergoing imaging, medical, or surgical procedures. Chloral hydrate, however, does not have an FDA indication for pediatric sedation.

Double-blind studies find that chloral hydrate is as effective as the benzodiazepines in the short-term treatment of insomnia. However, people may become tolerant to these effects after 2 weeks. Chloral hydrate appears in this respect similar to the barbiturates clinically and is not an effective agent for chronic insomnia. It also appears to be an effective agent for sedation in young children undergoing minor surgery and imaging studies (computerized tomography, magnetic resonance imaging, electroencephalography [EEG], vesicoureterogram). Respiratory compromise limits its use. Chloral hydrate may have some utility as an adjunct in the treatment of alcohol withdrawal, and a few controlled studies have found it effective for this purpose.

Effects on Specific Organ Systems

Central Nervous System Chloral hydrate has little if any analgesic properties and may induce excitement or agitation in the presence of pain. It also has anticonvulsant properties, although the ratio of anticonvulsant to sedative properties is low. Because of its anticonvulsant properties chloral hydrate has been used historically in the treatment of eclampsia or tetanus. The benzodiazepines or barbiturates are far superior to chloral hydrate as anticonvulsants. Chloral hydrate's most common CNS effects are due to depression at all levels of the neuraxis associated with high dose or abuse. High doses (over 4 grams) may be associated with stupor, confusion, ataxia, falls, or coma. Some individuals have survived doses over 30 grams. In general, appropriate doses are not associated with day-after hangover effects.

Respiratory System In typical clinical doses chloral hydrate has little effect on respiration. Toxic doses, however, may cause respiratory compromise and the drug should be used cautiously in patients who have obstructive pulmonary disease or asthma.

Cardiovascular System In oral therapeutic doses there are no effects on blood pressure, although higher doses or overdoses may result in lowered blood pressure because chloral hydrate depresses cardiac contractility. Patients with cardiac disease should probably not be given chloral hydrate.

Gastrointestinal System Chloral hydrate does not have specific effects on smooth muscle and does not induce hepatic microsomal enzymes the way the barbiturates do. The drug is irritating to mucous membranes and may cause unpleasant taste, epigastric distress, vomiting, or diarrhea.

Pregnancy There are no studies on the effects of chloral hydrate on the fetus. Use during pregnancy may precipitate a withdrawal syndrome in the newborn. Chloral hydrate crosses into breast milk and may cause drowsiness in nursing infants.

Therapeutic Indications

Insomnia Relatively less costly than other sedative-hypnotics, chloral hydrate is as an inexpensive drug for the treatment of insomnia although it is seldom prescribed for this purpose. Effective doses range from 500 to 1500 mg, although some patients require 2 grams. It has largely been supplanted by the benzodiazepines and zolpidem in the treatment of insomnia. Chloral hydrate is sometimes helpful for patients who do not benefit from benzodiazepines or for elderly patients with agitation associated with insomnia. The benzodiazepines, however, appear superior to chloral hydrate for sedation and sleep as reported in various studies that compare them directly.

Sedation for Pediatric Procedures Oral administration or rectal suppositories are used for children undergoing minor procedures or imaging studies. This is not an FDA-approved indication, although it is commonly used in pediatrics. The indicated dose is 50 mg per kg of body weight for oral or rectal use, with usual dose not higher than 1 gram; however, to ensure sedation, some practitioners use 80 to 100 mg per kg of body weight and up to 3 grams. Several recent controlled clinical trials evaluating chloral hydrate for various tests, especially imaging studies, have supported the clinical practice of using higher than the recommended dose of 50 mg per kg, and the dose of 80 to 100mg per kg may be optimal. Higher doses are more likely to ensure completion of the study without awakenings, and complications are uncommon even at doses of 100 mg per kg. Doses should not be repeated, and children should be monitored and not released until they are awake and appropriate for discharge.

Preoperative Use Chloral hydrate may be used preoperatively, 1 gram orally, to allay anxiety or produce sleep even though the benzodiazepines are preferable in this setting.

Alcohol Withdrawal Chloral hydrate has historically been used to treat alcohol withdrawal although it is rarely used now and has been replaced by the benzodiazepines, which are safer.

Precautions and Adverse Effects

Central Nervous System Like other CNS depressants, use of chloral hydrate may be associated with sedation; tiredness the morning after use; dizziness or lightheadedness; slurred speech; and less commonly confusion, hallucinations, or paradoxical excitement.

Abuse and Dependence Like the barbiturates, chloral hydrate has abuse potential, primarily by individuals with a history of alcohol or other drug abuse. Because tolerance can develop, addicts may take large doses to maintain their addiction. Sudden cessation of chloral hydrate in addicts leads to a withdrawal syndrome similar to barbiturate withdrawal, including autonomic instability, seizures, delirium, and possibly death if untreated.

Overdose Toxic doses generally result from use of 10 grams or more in adults, although elderly patients may be more sensitive. Poisoning resembles barbiturate poisoning, including loss of respiration, pinpoint pupils, and loss of deep tendon reflexes. Supportive treatment is the same as for other CNS depressants, and hemodialysis or hemoperfusion may promote trichloroethanol clearance. Deaths have been reported in children being sedated for procedures, and repeated doses should not be used for procedure sedation. Accumulation of trichloroethanol may cause excessive CNS depression and respiratory arrest.

Allergic Reactions Patients occasionally develop rash, erythema, urticaria, eosinophilia, or leukopenia.

Gastrointestinal Tract Because of its irritating properties, oral chloral hydrate should be avoided in patients with esophagitis, gastritis, gastric or duodenal ulcers, proctitis, or colitis. Chloral hydrate has been reported to precipitate attacks of intermittent porphyria and should not be used in patients with advanced liver disease.

DRUG INTERACTIONS

Patients receiving intravenous furosemide (Lasix) with chloral hydrate may experience vasodilation and flushing, tachycardia, hypertension or hypotension, and sweating. The symptoms are thought to be caused by displacement of thyroxine from the bound state. Similarly, chloral hydrate may displace oral anticoagulants from plasma proteins and increase the effects of a stable dose of warfarin (Coumadin) for several weeks. Monitoring of coagulation must be frequent if the combination is to be used, but because many other hypnotic agents are available patients taking warfarin should not be given chloral hydrate. Chloral hydrate combined with alcohol is known popularly as a “Mickey Finn.” The combination causes a combined and possibly synergistic effect of these two CNS depressants. Possible synergistic effects may be explained by the inhibition of metabolism of alcohol by chloral hydrate and the enhancement of production of trichloroethanol, the active drug of chloral hydrate, by ethanol. The two should not be used together.

LABORATORY INTERFERENCE

Chloral hydrate may interfere with urine glucose tests, such as Clinitest, giving false-positive results. Glucose enzymatic tests are not affected. Chloral hydrate also interferes with fluorometric tests of urine catecholamines and some assays of urine 17-hydroxycorticosteroids. Chloral hydrate may also cause false-positive results of the phentolamine test.

DOSAGE AND ADMINISTRATION

Chloral hydrate comes as capsules, syrup, or rectal suppositories. Capsules come in 250-mg and 500-mg strengths. Syrup comes in concentrations of 500 mg per 5 mL. Oral solution should be kept sealed and away from light to prevent deterioration. Oral solution may be irritating to the stomach and should be mixed in 4 to 8 ounces of water or juice. Suppositories come in strengths of 325 mg, 500 mg, and 650 mg.

Use in Insomnia Although insomnia is a common complaint in psychiatric practice, it is often symptomatic of underlying mood, anxiety, psychotic, or substance abuse disorders. If the primary psychiatric syndrome is targeted by other treatment, then adjunctive treatment with chloral hydrate may be appropriate. However, insomnia is often better treated with benzodiazepines or zolpidem because of their lower potential for toxicity. Some clinicians prefer to use chloral hydrate to treat insomnia in geriatric patients although this is controversial and research findings are mixed. If chloral hydrate is used it should be given 15 to 30 minutes before bedtime at doses of 500 or 1000 mg; some patients may require up to 2 grams. The dose for geriatric patients is 500 mg.

Sedation for Diagnostic Tests in Children For EEG sedation, 25 mg per kg of body weight is suggested for oral or rectal use. For other procedures the manufacturer-recommended dose is 50 mg per kg body weight, by oral capsule or rectal suppository, with a maximum dose of 1 gram. For many children the 50-mg per kg dose may be inadequate, and doses up to 100 mg per kg of body weight (up to 3 grams) are required. Because an insufficient dose may lead to agitation and inability to complete a procedure, underdosing is a clinical problem and it is suggested that children over age 5 not be given chloral hydrate for procedure sedation.

SUGGESTED CROSS-REFERENCES

Barbiturates are discussed in Section 31.8, and benzodiazepines are discussed in Section 31.9. Substance-related disorders are presented in Chapter 11, and sleep disorders are presented in Chapter 21.

SECTION REFERENCES

Devarajan S: Interaction of fluoxetine and chloral hydrate. Can J Psychiatry *37:*590, 1992.

*Greenberg SB, Faerber EN, Aspinall CL, Adams RC: High-dose chloral hydrate sedation for children undergoing MR imaging: Safety and efficacy in relation to age. Am J Roentgenol *161:*639, 1993.

Kaplan HS, Sadock BJ: Chloral hydrate. In *Kaplan and Sadock's Synopsis of Psychiatry*, ed 8. Williams & Wilkins, Baltimore, 1998.

Holister LE: The pre-benzodiazepine era. J Psychoactive Drugs *15:*9, 1983.

Johns MW: Sleep and hypnotic drugs. Drugs *9:*448, 1975.

Lambert GH, Muraskas J, Anderson CL, Meyers TF: Direct hyperbilirubinemia associated with chloral hydrate in the newborn. Pediatrics *86:*277, 1990.

Linoila M, Viukari M, Numminen A, Auvinen J: Efficacy and side effects of chloral hydrate and tryptophan as sleeping aids in psychogeriatric patients. Pharmacopsychiatry *15:*124, 1980.

*Lovinger DM, Zimmerman SA, Levitan M, Harrison NL: Trichloroethanol potentiates synaptic transmission mediated by gamma-aminobutyric acid. A receptors in hippocampal neurons. J Pharmacol Exp Ther *264:*1097, 1995.

Malach M, Berman N. Furosemide and chloral hydrate: Adverse drug interaction. JAMA *232:*638, 1975.

Marti-Bonmati L, Ronchera-Oms CL, Casillas C, Poyatos C, Torrijo C: Randomised double-blind clinical trial of intermediate versus high-dose chloral hydrate for neuroimaging of children. Neuroradiology *37:*687, 1995.

*McEvoy GK, editor: *AHFS 99 Drug Information*. American Society of Health-System Pharmacists, Bethesda, MD, 1999.

*Napoli KL, Ingall CG, Martin GR: Safety and efficacy of chloral hydrate sedation in children undergoing echocardiography. J Pediatr *129:*287, 1996.

Oaks DL, Robertson A, Brodersen R: The effect of chloral hydrate and its metabolites, trichloroethanol and trichloracetic acid, on bilirubin albumin binding. Pharmacol Toxicol *71:*196, 1992.

*Sourkes TL: Early clinical neurochemistry of CNS-active drugs: Chloral hydrate. Molec Chem Neuropathol *17:*21, 1992.

United States Pharmacopeial Convention: Chloral hydrate. In *United States Pharmacopeial Dispensing Information—Drug Information for the Health Care Professional*, ed 15. The United States Pharmacopeial Convention, Rockville, MD, 1995.

Vade A, Sukhani R, Dolenga M, Habisohn-Schuck C: Chloral hydrate sedation of children undergoing CT and MR imaging: Safety as judged by American Academy of Pediatrics guidelines. Am Roentgenol *165:*905, 1995.

31.15 CHOLINESTERASE INHIBITORS

STEVEN C. SAMUELS, M.D., AND KENNETH L. DAVIS, M.D.

Alzheimer's disease is a progressive neurodegenerative disease that gradually destroys brain function. Symptoms of the disease include cognitive deterioration involving memory, language, praxis, higher executive functioning and behavioral changes. In the United States, 4 million patients are afflicted with the disease, with direct and indirect costs for the disease approaching 100 billion dollars annually. The disease is age related with the prevalence of 3.0 percent at age 65, approaching 50 percent by age 85 years and leveling off at 40 percent by age 95. As the population ages, increased numbers of people will develop Alzheimer's disease. Cautious optimism stems from the recent availability of cholinesterase inhibitors that may slow down the progression of this terrible disease, but much more research is required before the disease can be arrested or reversed or those at risk for the disease can be reliably identified and prevented from developing it.

Several lines of evidence suggest that cholinergic function has a substantial role in Alzheimer's disease. Cholinergic transmission is involved in learning and memory. Anticholinergic agents cause memory and attention deficits in animals and humans. Destruction of the basal forebrain in animals or cholinergic circuitry in humans results in cognitive deficits. In clinical trials, cholinergic agents have restored cognitive function and decreased decline in Alzheimer's disease patients. The correlation between cholinergic abnormalities and the degree of cognitive impairment in Alzheimer's disease is a consistent finding in a number of studies. The significance of the cholinergic changes in Alzheimer's disease was eloquently described in a comprehensive examination of neuropathological and neurochemical changes in brains of those afflicted. The strongest correlations between clinical measures of dementia severity and neurochemical changes involved choline acetyltransferase, the enzyme necessary for the formation of acetylcholine.

The cholinergic hypothesis in Alzheimer's disease led to research and development of a number of potential therapeutic agents. These agents are in various stages of preclinical and clinical development in animals and humans. The agents currently approved by the Food and Drug Administration (FDA) for Alzheimer's disease are predominantly cholinesterase inhibitors. Attempts to alter cholinergic transmission have included increasing presynaptic precursors to acetylcholine, stimulating muscarinic receptors with selective muscarinic agonists and decreasing the enzymatic degradation of acetylcholine by blocking acetylcholinesterase or butylcholinesterase with cholinesterase inhibitors.

Lecithin and choline, the precursors to acetylcholine, have been extensively studied and shown not to be beneficial in Alzheimer's disease. Muscarinic agonists have demonstrated benefit in preclinical studies, and some compounds are now in clinical trials. Although tonic stimulation may be relevant in arousal and attention, continuous stimulation of receptors by cholinergic agonists may result in receptor downregulation because the physiological stimulation is not pulsatile. This downregulation may limit the effectiveness of the muscarinic agonists. Advancing age and Alzheimer's disease progression are associated with a decrease in number and function of cholinergic neurons, possibly explaining the failure of precursors affect the disease. Moreover, this decline in cholinergic neuron number and function may also decrease the effectiveness of the muscarinic agonists and the cholinesterase inhibitors.

TACRINE

Tacrine (Cognex), the first FDA-approved cholinesterase inhibitor for the treatment of dementia of the Alzheimer's type, is a centrally acting, noncompetitive, reversible inhibitor of acetylcholinesterase and butyrylcholinesterase. Originally synthesized in 1945, tacrine was not recognized as a cholinesterase inhibitor until 1953. Tacrine was initially used to attenuate morphine withdrawal in patients with cancer and pain syndromes. Tacrine is effective in slowing the apparent progression of dementia in a subgroup of patients with Alzheimer's disease who can tolerate the medication for sufficient duration. Tacrine's use in Alzheimer's disease is limited by four-times-a-day dosing, gastrointestinal adverse effects, the potential for hepatic toxicity, and the need for frequent serum monitoring.

Chemistry The molecular structure of tacrine (9-amino-1,2,3,4-tetrahydroacridine) is shown in Figure 31.15-1. Tacrine is metabolized to a number of hydroxylated products, including 1-OH-tacrine (which itself was being developed for the treatment of dementia of the Alzheimer's type) as velnacrine.

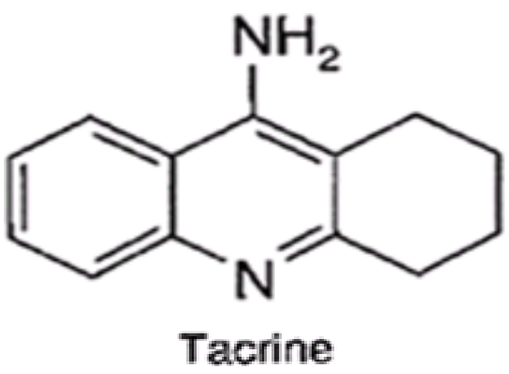

FIGURE 31.15-1 Molecular structure of tacrine.

Pharmacological Actions

Pharmacokinetics Tacrine is well absorbed from the gastrointestinal tract; however, intake with food reduces its bioavailability by 30 to 40 percent. Peak plasma concentration is reached 1 to 2 hours after oral dosing. Steady-state tacrine concentrations are reached after dosage initiation or change in 24 to 36 hours. The steady-state volume of distribution is about 300 L. The elimination half-life of tacrine is 2 to 3 hours. Tacrine is hepatically metabolized by the cytochrome P450 (CYP) is 0 enzymes 1A2 (CYP 1A) and CYP 2D6 and undergoes hydroxylation and conjugation. A small percentage of the drug is excreted in the urine, and dose adjustments are not necessary in patients with renal impairment.

Pharmacodynamics Tacrine's mechanism of action in Alzheimer's disease appears related to acetylcholinesterase inhibition resulting in increased acetylcholine availability. The relation between plasma concentration and oral dosing is nonlinear. Plasma concentrations of tacrine in women are double those of men, possibly related to decreased CYP 1A2 activity in women. Cigarette smokers have up to one-third lower serum tacrine levels than nonsmokers, probably related to induction of CYP 1A2 in smokers. Age and tacrine clearance are not directly related.

Design and Interpretation of Clinical Drug Studies Clinical trials using tacrine for Alzheimer's disease have variable methodological strength. Early studies demonstrating remarkable promise of tacrine for dementia were not well controlled. Additional studies in the 1980s had mixed results and continued methodological problems including inadequate dosing strength and duration of treatment. These concerns limited generalizations that could be made about effectiveness. Improved methodology in one 12-week study and 30-week study were pivotal in the ultimate FDA approval of tacrine.

Tacrine and Functional Brain Studies Brain single photon emission computed tomography (SPECT) and xenon (Xe)-inhalation have been used to examine brain function in Alzheimer's disease patients who received tacrine. These studies revealed correlations between clinical response to tacrine and functional imaging

changes. Tacrine may improve technetium-99 (^{99m}Tc)-labeled ethylene dicysteinate retention abnormalities (a measure of cerebral blood flow) in patients with mild-to-moderate dementia of the Alzheimer's type. The results were dose dependent; patients who received 75 mg of tacrine a day demonstrated improvement, while those receiving 25 mg of tacrine a day did not show functional change in brain SPECT. In an open-label 14-month trial of tacrine with maximum dosage of 125 mg a day, Xe-inhalation methodology was used to measure regional cerebral blood flow (rCBF) in probable Alzheimer's disease patients before and after tacrine treatment. Patients who responded favorably to tacrine showed improvement or stabilization in rCBF compared with those who did not receive tacrine. The results support a dosage-dependent response to tacrine. The possibility of using functional measures as predictors of clinical response to pharmaceuticals is intriguing, and further research is required in this area.

Pharmacoeconomic Considerations Tacrine has been estimated to generate potential savings up to 17 percent of the current costs of Alzheimer's disease, or a total 3.6 billion dollars annually. These calculations were based upon an improvement of 1 point on the Mini-Mental State Examination (MMSE) by patients tolerating 80 to 160 mg a day and a 2 point increase in patients who tolerated 160 mg a day. Patients with 1 to 2 points of improvement on the MMSE are estimated to have 9.5 to 12.1 months of reduced community and institutional costs. Industry-sponsored studies examining the effects of various agents on the direct and indirect costs of Alzheimer's disease are ongoing, and justification for choosing one pharmacotherapeutic agent over another may become closely tied to a drug's pharmacoeconomic profile.

Quality of Life In addition to pharmacoeconomic considerations, tacrine's effects on quality-of-life measures have been examined. The effect of tacrine on mortality and placement into a long-term care facility was examined in about 90 percent of the 663 patients from the 30-week clinical trial. Patients were followed for 2 years or until they died or were placed in a nursing home. Patients receiving more than 80 mg of tacrine daily were less likely to die or be placed in a nursing home than patients on a lower tacrine dose (odds ratio > 2.7). The absence of a control group and the retrospective nature of the study limit generalizations, but the dose-response nature of the findings is intriguing.

Effects on Specific Organs and Systems Tacrine dosages of 80 to 160 mg a day appear necessary for clinical response. At least 12 weeks of dose titration are required to achieve a dosage of 120 mg a day. More time is often required because the patient has gastrointestinal side effects or elevated hepatic transaminases. Transaminase elevations exceeding five times the upper limit of normal should result in tacrine discontinuation. Even with transaminase elevations, patients may still be rechallenged with tacrine after the transaminase levels normalize. In the clinical trials, no deaths associated with hepatitis in these patients were reported. A significant number of patients may tolerate rechallenge with tacrine, often at higher dosages than those that initially led to temporary discontinuation of the drug. Tacrine may have vagotonic effects on the heart rate and should be used with caution in patients with supraventricular cardiac arrhythmias. Gastrointestinal effects secondary to the parasympathetic effects of tacrine necessitate caution in patients at risk for peptic ulcer disease.

Therapeutic Indications Tacrine is indicated for the treatment of mild-to-moderate Alzheimer's disease.

Precautions and Adverse Reactions The clinical trials described above reported frequent adverse effects from tacrine. Hepatic abnormalities, although most often asymptomatic, require frequent monitoring of alanine transaminase (ALT) in serum. Dyspepsia, nausea, vomiting, diarrhea, anorexia, and abdominal pain are the most common adverse gastrointestinal effects. Adverse effects occurring in at least 2 percent of patients in clinical trials are reported in Table 31.15-1. Although many of the adverse effects occurred at similar rates in the tacrine and placebo groups, withdrawal from clinical trials occurred at a much higher rate in the tacrine group.

Adverse Event	Tacrine (%) (N = 634)	Placebo (%) (N = 342)
Laboratory interferences		
Elevated transaminase	29	2
Digestive system		
Nausea and/or vomiting	28	9
Diarrhea	16	5
Dyspepsia	9	6
Anorexia	9	3
Abdominal pain	8	7
Flatulence	4	2
Constipation	4	2
Body as a whole		
Fatigue	4	3
Weight loss	3	1
Musculoskeletal system		
Myalgia	9	5
Nervous system		
Dizziness	12	11
Ataxia	6	5
Insomnia	6	5
Somnolence	4	3
Tremor	2	<1
Psychobiological function		
Anxiety	3	2
Respiratory system		
Rhinitis	4	4

Adapted from tacrine package insert.

Table 31.15-1 Adverse Events Associated With Tacrine and Placebo Based Upon Clinical Trial Data

Hepatic Monitoring and Rechallenge Current recommendations for patients on tacrine treatment who tolerate the drug without significant elevations in hepatic transaminase concentrations are serum ALT monitoring every other week for 16 weeks, then monthly for 2 months, then once every 3 months. Patients may continue according to the titration schedule with up to twice the upper limit of normal ALT; weekly transaminase monitoring is necessary in patients with two to three times the upper limit of normal. Patients with ALT values three to five times the upper limit of normal should reduce the tacrine dosage by 40 mg a day and begin weekly transaminase monitoring. The dosing titration may resume with every-other-week transaminase monitoring when the ALT concentrations normalize. Patients with ALT concentrations exceeding 5 times the upper limit of normal should stop taking tacrine and be carefully monitored for signs and symptoms of hepatitis. If there is evidence of clinical jaundice, total bilirubin above 3 mg/dl, or clinical signs and symptoms of hypersensitivity such as rash or fever, patients should permanently stop taking tacrine and not be rechallenged. In one study that led to modification of the prescribing guidelines for tacrine, almost 88 percent of patients tolerated tacrine rechallenge, and 72 percent of the patients tolerated higher tacrine dosages than the initial dosage that resulted in discontinuation.

Weekly transaminase monitoring is required for all patients rechallenged with tacrine. After ALT values return to normal, patients may resume taking tacrine 10 mg four times a day. The titration schedule resumes after 6 weeks if the patient tolerates the drug and transaminase levels are acceptable (less than 3 times the upper limit of normal). Patients with initial elevations in ALT up to 10 times the upper limit of normal may be rechallenged after transaminase values normalize, although there is limited clinical experience with rechallenging these patients. Hypersensitivity reactions to tacrine resulting in eosinophilia or granulomatous hepatitis have been reported, and rechallenge of these patients is not recommended.

Drug Interactions Patients on theophylline (Theo-Dur) who also take tacrine show a doubling of theophylline concentration, as both drugs are metabolized by CYP 1A2. Cimetidine (Tagamet) is also metabolized by the same isoenzyme as tacrine, and coadministration of these agents requires judicious monitoring, as both drug concentrations increase. Tacrine prolongs the effect of succinylcholine by inhibiting butylcholinesterase, the enzyme responsible for succinylcholine degradation. There have been no reported drug interactions with digoxin (Lanoxin), warfarin (Coumadin), or diazepam (Valium).

Laboratory Interferences Tacrine does not interfere with urinalysis, spectroscopy, or other laboratory tests. Elevation of transaminases is associated with tacrine use.

Dosage and Administration Tacrine is available in 10-mg, 20-mg, 30-mg, and 40-mg capsules. After a thorough physical examination and review of transaminase levels, tacrine administration should be started at 10-mg four times a day for 6 weeks. The dosage may then be increased by 10 mg four times a day every 6 weeks until the target dosage of 40 mg four times a day is reached. Titration should proceed unless transaminases are elevated or there is other evidence of intolerance. If transaminase elevations occur, tacrine dosage is maintained at the current level or reduced. Patients who discontinue tacrine for 4 weeks must resume taking it at initial dosages of 10 mg four times a day, with attendant transaminase serum monitoring. Patients with gastrointestinal intolerance of tacrine may benefit from coadministration of tacrine with meals, although the drug's bioavailability may decrease by 30 to 40 percent.

DONEPEZIL

In December 1996, donepezil (Aricept) became the second cholinesterase inhibitor FDA-approved for treating dementia of the Alzheimer's type. This compound has once-daily dosing, lacks significant hepatotoxicity, and does not require serum monitoring. Prolonged dosage titration of the drug is not required. Donepezil was in development for close to four decades; in vitro selectivity of donepezil for acetylcholinesterase compared to butylcholinesterase was studied first in 1961.

Chemistry Donepezil, known chemically as (±) -2,3-dihydro-5,6-dimethoxy-2-1 [[1-(phenylmethyl)-4-piperidinyl]methyl]-1 *H*-inden-1-one hydrochloride, is a

piperidine-based acetylcholinesterase inhibitor. The empirical formula is $C_{24}H_{29}NO_3HCl$, and the molecular weight is 415.96. The molecular structure is shown in Figure 31.15-2.

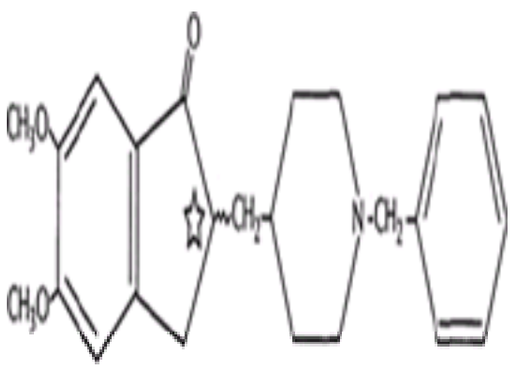

FIGURE 31.15-2 Molecular structure of donepezil.

Pharmacological Actions

Pharmacokinetics Donepezil is well absorbed in the gastrointestinal tract, and has an oral availability of 100 percent. Administration time and intake with food does not influence absorption. Peak plasma concentration is reached 3 to 4 hours after oral dosing. The steady-state volume of distribution is 12 L/kg. Donepezil is 96 percent bound to human plasma proteins, predominantly albumin (75 percent) and a_1 acid glycoprotein (21 percent). Multiple-dose administration leads to steady state in 15 days with a four- to sevenfold increase in donepezil concentration. The elimination half-life of donepezil is 70 hours. Donepezil is excreted in the urine and hepatically metabolized by CYP 3D4 and CYP 2D6 and undergoes glucuronidation. There are four major metabolites, two of which are active, and a number of minor metabolites. The package labeling reports that patients with hepatic disease (stable alcoholic cirrhosis) had 20 percent less hepatic clearance of donepezil than healthy controls. Patients with renal disease did not differ from healthy subjects in their clearance of donepezil.

Pharmacodynamics The drug reversibly and noncompetitively inhibits hydrolysis of acetylcholine by acetylcholinesterase in the brain, which results in increased availability of acetylcholine in the synaptic cleft. The inhibitor dissociation constant for donepezil is one or two orders of magnitude lower than that for tacrine, indicating a stronger inhibitory effect by donepezil on acetylcholinesterase. The selective inhibition of donepezil is 1250 times greater for acetylcholinesterase than for butylcholinesterase. The plasma concentration is linearly correlated with daily oral dose in the range of 1 to 10 mg.

Design and Interpretation of Clinical Drug Studies In a double-blind placebo-controlled study, probable Alzheimer's disease patients were treated with donepezil in dosages of 1 mg, 3 mg, and 5 mg a day or placebo for 12 weeks. The donepezil-treated patients demonstrated dosage-related improvements in the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-cog) and the MMSE. The 5-mg donepezil group demonstrated statistically significant improvement in the ADAS-cog compared with the placebo group. Donepezil was not associated with hepatic toxicity and was well tolerated in this study.

In another 12-week double-blind placebo-controlled study with a 3-week placebo washout period, patients received either placebo or 5 mg or 10 mg a day of donepezil. Patients randomized to 10 mg a day received this dosage after 1 week at 5 mg a day. At 12 weeks, the means in ADAS-cog scores in the donepezil-treated groups and placebo group differed significantly. The differences between the 5-mg and 10-mg groups were not statistically significant. After 3 weeks of placebo washout, the donepezil-treated groups demonstrated loss of treatment effect. The difference in ADAS cog scores between donepezil-treated patients and the placebo group was about 3 units at 12 weeks. Studies have reported an expected change of 9 points on the ADAS cog in untreated probable Alzheimer's disease patients. At 12 weeks, patients receiving donepezil also demonstrated statistically significant improvement compared with the placebo group in Clinician's Interview-Based Impression of Change Plus Caregiver (CIBIC-plus) scores, a clinician-based assessment of change after interview with the patient and caregiver.

A 30-week study also used the ADAS-cog and CIBIC-plus Assessment as outcome measures. After 24 weeks of active double-blind placebo-controlled treatment, patients had 6 weeks of single-blind placebo-controlled washout. Patients were randomized to receive donepezil at 5 mg a day, 10 mg a day (after 1 week at 5 mg a day), or placebo. After 24 weeks, ADAS-cog scores were higher in donepezil-treated patients than in those in the placebo group, although there was no statistically significant difference between the 5- and 10-mg groups. CIBIC plus scores at 24 weeks also statistically favored the donepezil-treated groups. There was no difference between the groups receiving 5 and 10 mg a day. After the 6-week placebo washout, the placebo groups and donepezil-treated groups were indistinguishable in ADAS-cog scores, suggesting that donepezil treatment did not change the underlying disease. There have been no direct comparisons of tacrine and donepezil. However, the change in the ADAS-cog scores for donepezil at its best has not been as large as the change in ADAS-cog scores for tacrine at its best.

Effects on Specific Organs and Systems Donepezil does not appear to cause hepatic injury. The drug may have vagotonic effects on the heart rate and should be used cautiously in patients with supraventricular cardiac arrhythmias including sick sinus syndrome. Gastrointestinal effects are secondary to the parasympathetic effects of donepezil. Gastric acid secretion may increase with donepezil. Patients with or at risk of peptic ulcer disease and patients taking nonsteroidal anti-inflammatory drugs (NSAIDs) should be carefully monitored for evidence of gastrointestinal bleeding while receiving donepezil. Nausea, diarrhea, and vomiting may occur more commonly with the 10 mg than the 5 mg a day dosage.

Therapeutic Indications Donepezil is indicated for the treatment of mild-to-moderate dementia of the Alzheimer's type. Other forms of dementia may have some cholinergic depletion but are not currently indicated conditions for prescribing donepezil.

Precautions and Adverse Reactions Table 13.15-2 includes all adverse events reported in clinical trials of donepezil that occurred at a rate of at least 2 percent and at greater frequency than in placebo-treated patients. The most common adverse effects associated with donepezil use were diarrhea, nausea, insomnia, vomiting, muscle cramps, fatigue, and anorexia. These events were frequently mild and transient, resolving during continued treatment with donepezil. Adverse events occurred more frequently in women and the elderly. The most common adverse effects associated with donepezil that led to discontinuation of clinical trials were nausea, diarrhea, and vomiting. Patients who were titrated up to 10 mg a day after 1 week had a higher discontinuation rate than those taking 5 mg a day. An open-label study of the placebo-treated groups revealed that the adverse event rates were lower in the patients titrated up to 10 mg over 6 weeks than in patients titrated to 10 mg a day in 1 week. The rates of adverse effects in the slower-titration group were similar to those in the group receiving 5 mg a day of donepezil.

Adverse Event	Placebo (N = 333)	Donepezil (N = 747)
Percentage of patients with any adverse event	72	74
Body as a whole		
Headache	9	10
Pain, various locations	8	9
Accident	6	7
Fatigue	3	5
Cardiovascular system		
Syncope	1	2
Digestive system		
Nausea	6	11
Diarrhea	5	10
Vomiting	3	5
Anorexia	2	4
Hemic and lymphatic system		
Ecchymosis	3	4
Metabolic and nutritional systems		
Weight decrease	1	3
Musculoskeletal system		
Muscle cramps	2	6
Arthritis	1	2
Nervous system		
Insomnia	6	9
Dizziness	6	8
Depression	<1	3
Abnormal dreams	0	3
Somnolence	<1	2
Urogenital system		
Frequent urination	1	2

Table 31.15-2 Adverse Events Reported in Clinical Trials Comparing Donepezil With Placebo

Drug Interactions The package insert for donepezil reports that the drug is highly protein bound. In vitro displacement studies with donepezil did not demonstrate

displacement of other tightly protein-bound drugs (furosemide [Lasix]), warfarin, digoxin). However, no in vivo drug displacement studies have been performed. Such studies in animals, with rapid infusion of drugs through the intravenous route as might occur under some clinical circumstances, would be revealing. The implications of tight protein binding for donepezil will remain unclear until in vivo studies are performed and potential drug interactions evaluated, especially since the average patient with Alzheimer's disease is taking multiple medications. The company reports that donepezil binding to albumin was not affected by furosemide, warfarin, or digoxin. The effect of donepezil on malnourished or cachectic patients has not been examined. Although there have been no formal studies examining drug interactions, the company reports that donepezil has no significant pharmacokinetic effects on warfarin, theophylline, cimetidine, and digoxin. Donepezil may increase succinylcholine effects on muscle relaxation. Agents that inhibit CYP 2D6 or CYP 3A4 may inhibit donepezil metabolism. Inducers of CYP 2D6 and CYP 3A4 may increase the elimination rate of donepezil. Cholinesterase inhibitors may also interfere with the effects of anticholinergic agents. This concern may be more academic than practical, as anticholinergic agents should generally be avoided in patients with Alzheimer's disease.

Laboratory Interferences There is no current evidence that donepezil interferes with urinalysis, spectrometry, or other laboratory tests.

Dosage and Administration Donepezil is supplied as a round tablet containing either 5 or 10 mg of donezepil hydrochloride. The 5-mg tablets are white, and the 10-mg tablets are yellow. The recommended starting dose is 5 mg daily. The medication is usually administered in the evening so the peak plasma concentration occurs when the patient is sleeping, minimizing adverse effects. The clinical trials demonstrated a dose trend favoring 10 mg over 5 mg a day, although the results were not statistically significant. Based on the clinical trials, it is a matter of prescriber and patient collaboration whether to increase from 5 to 10 mg a day. Open trials with 10 mg titration occurring at 6 weeks demonstrated no significant difference in adverse effects in the 5-mg and 10-mg groups. These authors recommend that patients be given a trial at 10 mg a day after 4 to 6 weeks if they are tolerating the 5 mg a day dosage. The half-life of donepezil is reported to be 70 hours on the basis of studies with younger patients. No half-life studies have been performed with elderly patients. Pharmacokinetic and pharmacodynamic changes in elderly persons may lead to an increased half-life, and it may be preferable to use dosages of 5 mg a day.

INVESTIGATIONAL DRUGS

Physostigmine Physostigmine's molecular structure is shown in Figure 31.15-3. Physostigmine (Antilirium, Eserine), a short-acting reversible cholinesterase inhibitor, has been extensively studied in Alzheimer's disease. Physostigmine improves recognition memory and enhances long-term memory processing. The response is dose dependent with a significant amount of individual variability, requiring a dose-finding phase of treatment to determine a patient's best dosage. Physostigmine is available in parenteral and oral preparations. Studies of oral and parenteral physostigmine in Alzheimer's disease patients have demonstrated variable responses. The drug's effectiveness in dementia appears related to cholinergic inhibition in serum and cerebrospinal fluid (CSF), and drug concentrations in serum and CSF fluctuate significantly with either parenteral or oral administration. Physostigmine is also associated with peripheral cholinergic side effects that often lead to discontinuation of the drug. More recently, a longer-acting oral preparation, physostigmine salicylate, has demonstrated some benefit in Alzheimer's disease patients in phase III clinical trials.

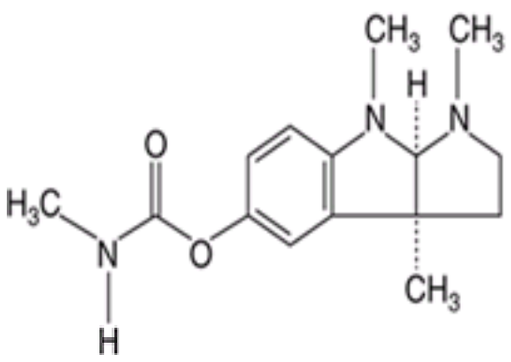

FIGURE 31.15-3 Molecular structure of physostigmine.

Eptastigmine Eptastigmine's molecular structure is shown in Figure 31.15-4. Eptastigmine, or heptylphysostigmine, is another longer-acting preparation of physostigmine. One 13-week controlled clinical trial of eptastagmine included 103 patients with probable Alzheimer's disease. The first 4 weeks were randomized, double-blind placebo-controlled, followed by 1 week of washout and an 8-week open-label phase. Subjects had MMSE scores between 10 and 26. Uniform dosing resulted in nausea and vomiting in some patients, and the methodology was changed to dosing the drug by weight. Outcome ratings included both physician and caregiver global ratings, independent activities-of-daily-living measure, and tests of logical memory, semantic word fluency, and executive functioning. One hundred three patients entered the double-blind phase (81 eptastigmine, 22 placebo). After 4 weeks, 94 remained in the study (74 eptastigmine, 20 placebo). Cholinergic adverse effects, uncooperation, protocol violation, and clinical worsening were all cited as reasons for dropout. At completion of the double-blind phase, the physician global measure and independent activities-of-daily-living measure favored the eptastigmine group. The tests of memory, word fluency, and higher executive functioning demonstrated no between-group differences. Dosage response followed an inverted U shape. Agranulocytosis has been reported with high doses of this drug and may limit its use. More clinical trial data are required with this agent.

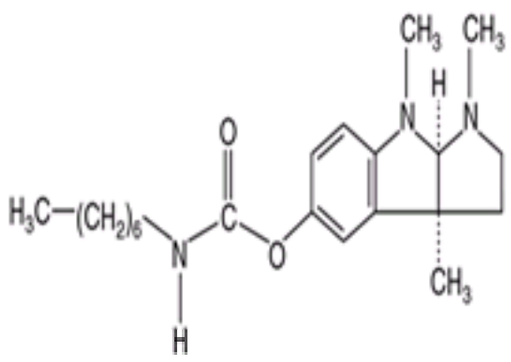

FIGURE 31.15-4 Molecular structure of eptastigmine.

Rivastigmine Rivastigmine's molecular structure is shown in Figure 31.15-5.

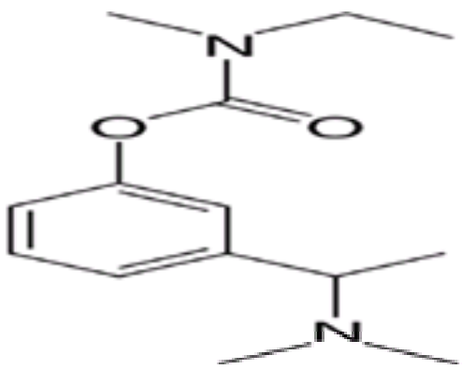

FIGURE 31.15-5 Molecular structure of rivastigmine.

Rivastigmine is a pseudoirreversible carbamate acetylcholinesterase inhibitor that is selective for the hippocampus and cortex. The drug is not metabolized by the liver, and no hepatic toxicity has been reported. It is metabolized by cholinesterase and is not protein bound. Rivastigmine was demonstrated to improve acquisition and retention memory in animals with selective basal forebrain lesions. Phase III clinical trials have been completed, and significant benefits have been reported in the ADAS, CIBIC Assessment and activities-of-daily-living measures. Concern about the possibility of increased deaths in the high-dosage group receiving rivastigmine compared with placebo led to inquiry of the Executive Director of Novartis. He responded that 62 patients died in the clinical trials with rivastigmine (including deaths occurring in patients treated up to 5 years prior to June 30, 1997). No statistical significant difference in the incidence of mortality was seen in placebo-controlled studies in phase I (rivastigmine, 0.8 percent, placebo, 0.5 percent) or phase III (rivastigmine –0.3 percent; placebo –0.1 percent). In phase 3 trials, no difference was found in mortality rates (Kaplan-Meier survival analysis) for patients treated long term with rivastigmine, which was assessed by comparing mortality in patients who were randomized to rivastigmine (13.0 deaths per 1000 patient-years) and treated long term with rivastigmine, with patients randomized to placebo for the first six months and treated long term with rivastigmine (13.5 deaths per 1000 patient-years). Nested case-control analyses were conducted to explore a dosage-response relationship for mortality, based on last prescribed dosage of rivastigmine and post hoc assignment of patients to nonoverlapping dosage ranges of 1 to <4, 4 to 6, >6 to 9, and >9 to 12 mg a day. Analyses comparing the highest to the lowest dose categories yielded a rate ratio of 0.8, indicating the absence of increased risk of mortality at higher dosages of rivastigmine. The causes of death in the rivastigmine clinical trials were numerous and similar to those reported in the literature for an Alzheimer's disease population.

Galanthamine Galanthamine's molecular structure is shown in Figure 31.15-6. It is a tertiary amine of the phenthrene group that competitively inhibits acetylcholinesterase but not butyrylcholinesterase. The compound is plant derived, naturally occurring, and an allosteric agonist at nicotinic sites without producing desensitization, as well as a cholinesterase inhibitor. This ability to enhance the sensitivity of the acetylcholine receptor is possessed only by galanthamine and physostigmine. The mechanism of action is analogous to that of benzodiazepines at the g-aminobutyric acid (GABA) A site. Consequently, the drug has an in vivo cholinomimetic activity that is even stronger than its acetylcholinesterase inhibitory properties alone would predict. Galanthamine's properties are interesting because in animals the combination of agonist and cholinesterase inhibitor may have an additive effect with fewer adverse events. Placebo-controlled clinical trials have demonstrated benefit in ADAS, and activities-of-daily-living measures. The usual effective dosage of galanthamine is 20 to 40 mg in divided doses. The drug has been approved in Austria, and multicenter clinical trials are under way in Europe and the United States.

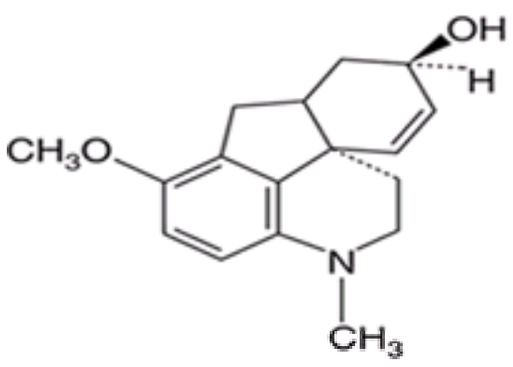

FIGURE 31.15-6 Molecular structure of galanthamine.

Metrifonate Metrifonate's molecular structure is shown in Figure 31.15-7. It is an irreversible cholinesterase inhibitor that is currently used to combat schistosomes. Metrifonate preferentially inhibits butyrylcholinesterase over acetylcholinesterase. Metrifonate is the prodrug of dichlorvos, the long-acting irreversible organic cholinesterase inhibitor. Although metrifonate has a short half-life, it achieves long-acting cholinesterase inhibition. Animal and human studies support metrifonate's memory-enhancing effects. Double-blind placebo controlled studies with Alzheimer's disease patients have demonstrated benefits for metrifonate-treated patients in 12- and 26-week trials. The benefits have been in cognitive measures (Alzheimer's Disease Assessment Scale-cog), behavioral measures (Neuropsychiatric Inventory), and global function (Clinician's Interview-Based Impression of Change-plus). The Food and Drug Administration is reviewing metrifonate, but there are no imminent plans for release of this medication because of safety concerns after some patients demonstrated prolonged muscle weakness after receiving the drug.

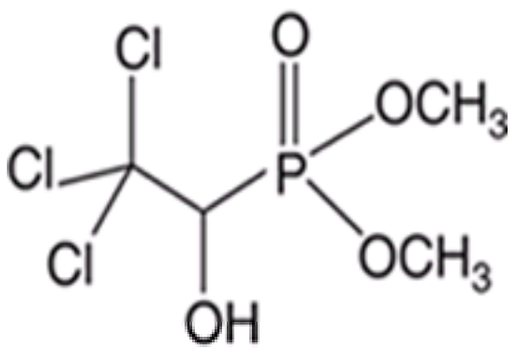

FIGURE 31.15-7 Molecular structure of metrifonate.

FUTURE DIRECTIONS

The concept of a polypharmaceutical approach to Alzheimer's disease has a theoretical rationale, as multiple domains appear to be involved in its pathogenesis. Future therapeutic approaches for Alzheimer's disease may be similar to hypertensive or neoplastic treatment, in which combination therapy is commonplace. Combination treatment with cholinesterase inhibitors and other agents has not been prospectively evaluated in Alzheimer's disease. A retrospective analysis to assess the effects of estrogen replacement therapy on tacrine response was performed using data from the 30-week tacrine study. Women receiving estrogen replacement therapy and tacrine had better measures of cognition and overall function than those not taking estrogen. Although these results are intriguing, prospective studies using cholinesterase inhibitors in combination with other treatments are required prior to recommending combination treatment as a standard of care.

SUGGESTED CROSS-REFERENCES

The neural sciences are discussed in Chapter 1 and dementia in Chapter 10 and Section 51.3e.

SECTION REFERENCES

Bierer LM, Haroutunian V, Gabriel S: Neurochemical correlates of dementia severity in Alzheimer's disease: Relative importance of the cholinergic deficits. J Neurochem *64:*749, 1995.

Canal N, Imbimbo BP: Relationship between pharmacodynamic activity and cognitive effects of eptastigmine in patients with Alzheimer's disease. Clin Pharmacol Ther *60:*218, 1996.

Corey-Bloom J, Anand R, Veach J: A randomized trial evaluating the efficacy and safety of ENA-713 (rivastigmine tartrate), a new acetylcholinesterase inhibitor, in patients with mild to moderately severe Alzheimer's disease. Int J Ger Psychopharmacol *1:*55, 1998.

Cummings JL, Cyrus PA, Bieber F, Mas J, Orazem J, Gulanski B: Metrifonate treatment of the cognitive deficits of Alzheimer's disease. Metrifonate Study Group. Neurology *50:*1214, 1998.

*Davis KL, Thal LJ, Gamzu ER, Davis CS, Woolson RF, Gracon SI, Drachman DA, Schneider LS, Whitehouse PJ, Hoover TM: A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. The Tacrine Collaborative Study Group. N Engl J Med *327:*1253, 1992.

*Knapp MJ, Knopman DS, Solomon PR, Pendlebury WW, Davis CS, Gracon SI: A 30-week randomized controlled trial of high-dose tacrine in patients with Alzheimer's disease. The Tacrine Study Group. JAMA *271:*985, 1994.

*Knopman D, Schneider L, Davis K, Talwalker S, Smith F, Hoover T, Gracon S. Long-term tacrine (Cognex) treatment: Effects on nursing home placement and mortality. Tacrine Study Group. Neurology *47:*166, 1996.

Mohs RC, Davis BM, Johns CA, Mathe AA, Greenwald BS, Horvath TB, Davis KL: Oral physostigmine treatment of patients with Alzheimer's disease. Am J Psychiatry *142:*28, 1985.

Morris JC, Cyrus PA, Orazem J, Mas J, Bieber F, Ruzicka BB, Gulanski B: Metrifonate benefits cognitive, behavioral, and global function in patients with Alzheimer's disease. Neurology *50:*1222, 1998.

Rogers SL, Doody RS, Mohs RC, Friedhoff LT: Donepezil improves cognitive and global function in Alzheimer disease: A 15-week, double-blind, placebo-controlled study. Donepezil Study Group. Arch Intern Med *158:*1021, 1998.

*Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT: A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Donepezil Study Group. Neurology *50:*136, 1998.

Summers WK, Majovski LV, Marsh GM, Tachiki K, Kling A: Oral tetrahydroaminoacridine in long-term treatment of senile dementia, Alzheimer type. N Engl J Med *315:*1241, 1986.

Thal LJ, Schwartz G, Sano M, Weiner M, Knopman D, Harrell L, Bodenheimer S, Rossor M, Philpot M, Schor J, Goldberg A: A multicenter double-blind study of controlled-release physostigmine for the treatment of symptoms secondary to Alzheimer's disease. Physostigmine Study Group. Neurology *47:*1389, 1996.

*van Gool WA: Efficacy of donepezil in Alzheimer's disease: Fact or artifact? Neurology *52:*218, 1999.

*Wallace W, Haroutunian V: Using the subcortically lesioned rat cortex to understand the physiological role of amyloid precursor protein. Behav Brain Res *57:*199, 1993.

Watkins PB, Zimmerman HJ, Knapp MJ, Gracon SI, Lewis KW: Hepatotoxic effects of tacrine administration in patients with Alzheimer's disease. JAMA *271:*992, 1994.

31.16 CLONIDINE

NORMAN SUSSMAN, M.D.

INTRODUCTION AND HISTORY

Clonidine (Catapres) is a presynaptic a_2-adrenergic receptor agonist that is approved for use as an antihypertensive agent. Stimulation of a_2-adrenergic receptors reduces the firing rate of noradrenergic neurons and reduces plasma concentrations of norepinephrine. Because of the widespread actions of the noradrenergic system, clonidine has also been adopted for use as a psychopharmacological agent. The most important clinical applications in psychiatry are as therapy for attention-deficit/hyperactivity disorder, opioid-withdrawal, Tourette's disorder, and posttraumatic stress disorder. It is also used as a pharmacological probe to assess central a_2-receptor sensitivity in psychiatric disorders.

Clonidine is clearly a useful drug. The improved symptomatic control provided by clonidine may enable the patient to function more effectively and to benefit from other treatment modalities. Nevertheless, its role as a treatment for selected mental disorders is generally limited to instances when other interventions have failed to ameliorate symptoms adequately. Although many reports exist of clonidine being used for a range of disorders, there are few double-blind controlled studies to support its efficacy for these uses. Clonidine also has significant limitations and risks. The frequent development of tolerance tends to limit its long-term effectiveness, making it more useful as short-term therapy. There is also uncertainty about cardiovascular risks with its use in children. Guanfacine (Tenex), another a_2-adrenergic receptor agonist, appears to offer some advantages over clonidine. It is less sedating than clonidine and has a longer half-life. There is, however, less clinical experience with guanfacine, and fewer controlled studies involve its use in psychiatric disorders than use of clonidine.

CHEMISTRY

Clonidine is an imidazoline compound whose molecular structure is shown in Figure 31.16-1.

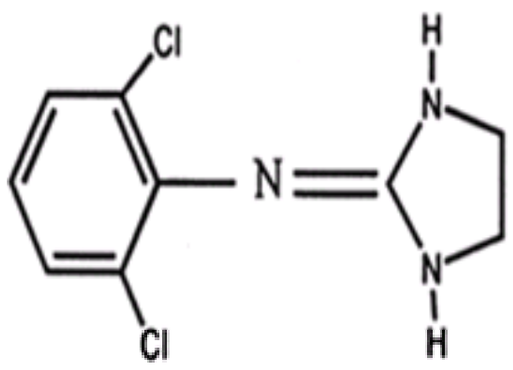

FIGURE 31.16-1 Molecular structure of clonidine.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Clonidine is well absorbed from the gastrointestinal tract and reaches peak plasma levels 1 to 3 hours after oral administration. It acts rapidly, decreasing blood pressure within 30 to 60 minutes after an oral dose.

Bioavailability The bioavailability of clonidine in healthy adults averages about 75 percent and may approach 100 percent in some patients.

Distribution Clonidine is lipophilic. It rapidly penetrates the brain, its primary site of action. It is also extensively distributed, which shortens its effective half-life.

Elimination In adults, about 50 percent of the drug is metabolized in the liver. Between 40 and 60 percent of the drug is excreted unchanged in the urine. The plasma half-life of the parent compound is 12 to 16 hours. There are no active metabolites. In persons with severely impaired renal function, the half-life increases up to 40 hours. The rate of clonidine metabolism is highest in young children (under age 6). In prepuberty it is still twice the adult rate, which is reached in midadolescence. The clinical significance of this rapid metabolism is the shortening of clonidine's behavioral activity and an increased risk of interdose withdrawal symptoms. In some cases in children it may need to be given four times a day. Both children and adults demonstrate considerable individual variation in the elimination half-life of clonidine.

Pharmacodynamics

Mechanism of Action Short-term administration of clonidine stimulates presynaptic a_2-autoreceptors. This effect is believed to account for the sedative and behavioral effects of the drug. Clonidine has additional central nervous system (CNS) activity that accounts for its overall clinical effects. It is a postsynaptic a_1-adrenergic receptor antagonist and it interacts with nonadrenergic binding sites termed imidazoline receptors. Clonidine has a high affinity for the I1 imidazoline receptor subtype. Clonidine stimulates locus coeruleus neurons through imidazoline receptors. It has been postulated that the interaction with imidazoline receptors contributes to clonidine's hypotensive effects. Both a_2-receptors and imidazoline receptors have also been found to be dysregulated in depression. Both receptors are also involved in mediating the therapeutic effects of some antidepressant drugs.

Administration of clonidine induces secretion of growth hormone through activation of a_2-receptors in the hypothalamus. The growth hormone response is extensively used as a probe to evaluate central noradrenergic function in both psychiatric and neurological disorders. It is presumed to represent an indirect measure of disturbed central adrenergic receptor function and a possible biologic marker of these disorders. Blunted growth hormone response to clonidine has been reported in some studies of patients who are depressed or have panic disorder. This finding among patients with major depressive disorder is presumed to reflect decreased responsiveness of postsynaptic a_2-receptors. Clonidine also fails to increase growth hormone concentration in rat strains that may represent a genetic model of depression. Long-term treatment with antidepressants restores clonidine's effect on growth hormone secretion in these animals. Untreated depressed patients show a

higher affinity of platelet a $_2$-receptors for clonidine than do healthy subjects. Studies also suggest that imidazoline receptors are dysregulated in depression. Chronic treatment with antidepressants normalizes platelet I1 binding sites.

Clonidine has also been used to measure stress response. Clonidine produces greater vasoconstriction and growth hormone release in males. This is of interest because it is consistent with observations that males exhibit more pronounced activation of the sympathoadrenal system after exposure to physiological stressors.

Nocturnal urinary growth hormone concentrations after a clonidine test is used to screen children with short stature for growth hormone deficiency.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Clonidine reduces peripheral sympathetic tone, lowers diastolic and systolic blood pressure, and causes bradycardia. Activation of central a $_2$-receptors results in a sleep-like state in animal studies. Clonidine causes sedation and sleep in humans. It increases slow-wave sleep, reduces rapid eye movement (REM) sleep time and percentage, and increases REM latency. The effects of clonidine on the gastrointestinal tract are minimal. It produces some reduction in basal gastric acid secretion. Clonidine has little effect on renal function.

THERAPEUTIC INDICATIONS

Clonidine is useful in a broad range of conditions seen in psychiatric and neurological practices. As a rule, clonidine is most beneficial in conditions that are associated with a hyperadrenergic state, such as treatment of opioid withdrawal and acute posttraumatic stress disorder. Tic disorders and attention-deficit/disorder represent other common uses in psychiatry.

Opioid Withdrawal Clonidine reduces the manifestations of adrenergic hyperactivity that accompany abrupt opioid withdrawal, including withdrawal from methadone. Signs and symptoms of opioid withdrawal include hypertension, tachycardia, dilated pupils, sweating, lacrimation, and rhinorrhea. The effectiveness of clonidine in treating opiate withdrawal is presumed to derive from its blockade of noradrenergic neurons of the locus coeruleus. Clonidine is best used in an inpatient setting, in part because side effects can be more readily managed. Hypotension can be particularly problematical and requires close monitoring.

Several oral clonidine protocols for opioid withdrawal are used (Table 31.16-1). The doses used to treat withdrawal symptoms are usually higher than those used to treat hypertension. Patients should be started on clonidine before discontinuation is begun, usually at a dosage of 0.1 to 0.2 mg four times a day. After stabilization for 2 to 3 days, the clonidine is tapered over 5 to 10 days. Clonidine can be used alone or combined with naltrexone for opioid detoxification. Other protocols use the transdermal clonidine patch (Table 31.16-1). Before being given the patch, patients should receive a test dose of oral or sublingual clonidine.

Clonidine 0.1–0.2 mg p.o. 4 times a day; hold for systolic blood pressure <90 mm Hg or bradycardia; stabilize for 2–3 days, then taper over 5–10 days

OR

Clonidine 0.1–0.2 mg p.o. q4–6h as needed for withdrawal signs or symptoms; stabilize for 2–3 days, then taper over 5–10 days

OR

Test dose with clonidine 0.1–0.2 mg p.o. or sublingually (for patients weighing over 200 lbs); check blood pressure after 1 h. If diastolic BP >70 mm Hg and no symptoms of hypotension, begin treatment as follows:

Weight (lb)	Number of clonidine patches
<110	2–3 TTS-1 patches
110–160	1 TTS-2 and 2 TTS-1 patches
160–200	2 TTS-2 patches
>200	2–3 TTS-2 patches

OR

Test dose of oral clonidine 0.1 mg; check BP after 1 h (if systolic blood pressure >90, do not give patch)

Place 2 TTS-2 clonidine patches (or 3 patches if patient weighs >150 lb) on hairless area of upper body; then

For first 24 h after patch application, give oral clonidine 0.2 mg q6h; then

For next 24 h, give oral clonidine 0.1 mg q6h

Change patches weekly

After 2 weeks of 2 patches, switch to 1 TTS-2 patch (or 2 TTS-2 patches if patient weighs >150 lb)

After 1 week of 1 patch, discontinue patches

Reprinted with permission from Kosten [footnote text illegible]

Table 31.16-1 Oral Clonidine Protocols for Opioid Detoxification

Clonidine can reduce neonatal opioid abstinence at doses of 3 to 4 µg/kg. In contrast to acute withdrawal, the use of clonidine as an adjunct to gradual opioid withdrawal has not been established. It may have limited value in alleviating abstinence symptoms during the transitional period between opioid cessation and naltrexone.

Alcohol Withdrawal Some studies suggest that clonidine may reduce the severity of acute alcohol withdrawal. Symptoms of anxiety, diarrhea, headache, lightheadedness, and fatigue of the alcohol withdrawal syndrome may be decreased. However, both clinical experience and research data are limited. Most reported patients were in mild withdrawal. Its effects on reducing delirium and seizures are not known. Thus, clonidine may be useful as adjunctive therapy, mainly with benzodiazepines, but it is not recommended as monotherapy. A recent survey of inpatient alcohol treatment programs in the United States found that of all inpatients treated for alcohol withdrawal, 7 percent received clonidine. About 60 percent of these patients, by comparison, were prescribed benzodiazepines. Clonidine may have some limited value in the management of benzodiazepine withdrawal.

Acute Nicotine Withdrawal Clonidine is a possible alternative to nicotine replacement for smoking cessation. Anxiety, irritability, and craving associated with nicotine withdrawal have been reported to be decreased by clonidine. The transdermal patch is associated with fewer adverse effects and greater compliance.

Tourette's Disorder Although the results of clonidine in treating Tourette's disorder are mixed, some clinicians use the drug as a first-line agent for the treatment of Tourette's disorder because it does not have the serious side effects associated with the standard agents, haloperidol (Haldol) and pimozide (Orap). These include tardive dyskinesia and neuroleptic malignant syndrome. A double-blind placebo-controlled study has found clonidine alone to be of some value. Clonidine can decrease tic severity and frequency, impulsivity, and hyperactivity. The starting child dosage is 0.05 mg a day; it can be raised to 0.3 mg a day in divided doses. Three months are needed before the beneficial effects of clonidine can be seen in Tourette's disorder. Clonidine discontinuation may be associated with rebound agitation and tics.

Autistic Disorder Clonidine reduces irritability, impulsivity, and hyperactivity in children with autistic disorder, and a modest effect has been reported in aggressive children and hyperactive and impulsive children with autistic disorder. A modest effect has been reported when clonidine was used in conjunction with other medications in an adult with autism and intermittent aggressive behavior. Many patients developed tolerance to the therapeutic effects of clonidine after several months of treatment.

Attention-Deficit Disorders and Tic Disorder Children who are hyperactive, highly aroused, defiant, and labile have been reported to show improvement in mood, activity level, cooperation, and frustration tolerance. Sleep disturbances are also common in attention-deficit/hyperactivity disorder whether or not psychostimulants are used. Most children treated with clonidine for attention-deficit/hyperactivity disorder–associated sleep disturbances show significant improvement in the quality of their sleep. When used alone, clonidine is not effective in the treatment of inattention. Clonidine is typically used in combination with stimulants when there is nonresponse, partial response, or negative response to stimulants. It is also used when side effects limit optimal doses of a stimulant.

A typical total daily oral dosage is 0.3 mg as tolerated in divided doses, two to four times a day. One suggested approach is to start with a daily dosage of 0.05 mg and increase the dosage on a three-times-a-day or four-times-a-day basis until a response is obtained. Another strategy is to start with 0.025 mg four times a day. Most patients respond to dosages below 5 mg daily. Transdermal clonidine is a useful alternative when compliance is an obstacle to treatment.

The combination of clonidine with sympathomimetics has not been studied for safety and efficacy, but clinical experience suggests that use of these combinations benefits children with attention-deficit hyperactivity disorder who are unresponsive to other treatments. A potential benefit of combining clonidine with stimulants is improvement of sleep difficulties caused by the activating compounds late in the day. In published reports of dosing for this purpose, doses ranged from 50 to 800 µg.

Instances of sudden death in children taking both clonidine and methylphenidate (Ritalin) have been reported and extensively discussed in the scientific literature. The mechanism causing these deaths has not been identified. Evidence that these deaths were related to this combination of drugs is still viewed as tenuous.

Nevertheless, vital signs should be monitored whenever clonidine is used, because of its hypotensive effects. Electrocardiography is also reasonable, when clonidine is used alone or in combination with other drugs if abnormalities are found during the physical examination or if there is evidence or a history of preexisting heart disease. Alternatives to the clonidine-methylphenidate combination may also be considered. Dextroamphetamine can be substituted for methylphenidate, and guanfacine can be substituted for clonidine.

Posttraumatic Stress Disorder Hyperarousal is a main feature of posttraumatic stress disorder. The effectiveness of clonidine in this disorder is best established during short-term treatment. Clonidine alleviates exaggerated startle response and insomnia, nightmares, and other manifestations of hyperarousal in patients with posttraumatic stress disorder. It is used in both adults and children, and has been studied in children as young as age 3. When used in children, it also improves aggressiveness and other behavioral disturbances. Most children can be given an initial oral dose of 0.05 mg in the morning and, if this is well tolerated an additional 0.05 mg in the evening. Once the patient is stabilized, the daily dosage and the timing of each dose needs to be individualized. The usual total daily dosage is less than 2 mg.

Cognition Clonidine is reported to enhance cognition in young adult volunteers and patients with schizophrenia and Korsakoff's syndrome, possibly through enhancement of CNS noradrenergic responsiveness. Cognitive tasks associated with prefrontal cortex functions show the most improvement.

Mood Disorders Clonidine has a limited role in treating mania. Although open-label studies suggest that clonidine may be useful, controlled trials have failed to confirm these results and have been associated with a high dropout rate because of adverse effects, mainly rash and hypotension. When reported, the observed benefits have been short-lived. In animal models, abrupt withdrawal resulted in behaviors believed to be predictive of clinical depression in humans.

Anxiety Disorders Other potential indications for clonidine include the anxiety disorders. Panic disorder is postulated to involve central noradrenergic dysregulation. The a_2-adrenergic receptor antagonist yohimbine provokes apprehension and panic attacks, while clonidine antagonizes these anxiogenic effects, if only briefly. Clonidine partially attenuates lactate-induced panic attacks.

While better than placebo in some studies of panic disorder patients, clonidine has been shown to be less effective than standard antipanic agents. Limited evidence indicates that clonidine may produce short-term improvement of symptoms in obsessive-compulsive disorder, posttraumatic stress disorder, and generalized anxiety disorder.

Schizophrenia Clinical studies of clonidine monotherapy have produced conflicting results. One study, involving a small number of patients, found haloperidol combined with clonidine to have greater efficacy than haloperidol alone in the treatment of schizophrenia.

Anesthesia and Analgesia Intravenous clonidine is used as a preanesthetic medication. It has anesthetic properties—potentiating morphine-induced antinociception—and is used in association with local anesthetics in epidural anesthesia. Clonidine also provides a sedative effect.

Other Disorders Isolated reports describe successful use of clonidine in treating mental and neurological disorders including premenstrual syndrome, restless leg syndrome, tardive dyskinesia, and smoking cessation. Clonidine has been used for the treatment of aggressiveness in children.

Clonidine can reduce the excessive salivation associated with clozapine treatment. In these patients, use of the transdermal patch is recommended, since it provides clonidine at the time when hypersalivation is most prominent, during naps and nighttime sleep. Transdermal clonidine has been noted to be effective in the prophylaxis of episodic cluster headaches.

PRECAUTIONS AND ADVERSE REACTIONS

Hypotension and sedation are the two side effects that most often limit the use of clonidine, regardless of the condition being treated. No matter what the indication for clonidine use, it should be withheld if a patient becomes hypotensive (blood pressure less than 90/60).

Clonidine can worsen arrhythmias, so electrocardiogram (ECG) abnormalities or bradycardia should prompt a cardiac consultation before clonidine is administered. Sedation is especially problematical, since tolerance to this side effect does not typically occur, which limits dosage increases when tolerance to the therapeutic effects of the drug occurs. Sedation may also lead to paradoxical dyscontrol of aggression and impulsivity. Other common adverse effects associated with clonidine are dry mouth and eyes, dizziness, nausea, and constipation. Some patients experience sexual dysfunction. Uncommon CNS adverse effects include insomnia, anxiety, and depression; rare CNS adverse effects include vivid dreams, nightmares, and hallucinations. Clonidine may interfere with motor recovery after ischemic lesions.

Fluid retention may occur and can be treated with diuretics.

Patients who overdose on clonidine can present with coma and constricted pupils, symptoms that are also seen in an opioid overdose. Other symptoms of overdose are decreased blood pressure, pulse, and respiratory rates and irritability, lethargy, and seizures.

Clonidine should be used with caution for patients with heart disease, renal disease, Raynaud's syndrome, or a history of depression. Clonidine should be avoided during pregnancy and by nursing mothers because of its cardiovascular effects. There is no evidence that clonidine is associated with teratogenicity in animal studies. Elderly persons are more sensitive to the drug than are younger adults, and children and adults are susceptible to the same side effects. Although rare, instances of clonidine abuse occur.

Localized skin irritation may develop with use of the transdermal patch. Moving the patch from one site to another can minimize irritation. Some patients who developed localized contact sensitization to the patch developed a generalized rash when they were switched to oral clonidine.

WITHDRAWAL

Discontinuation of clonidine, especially if it is abrupt and follows long-term use, can result in a severe withdrawal syndrome characterized by apprehension, restlessness, abdominal pain, sweating, tremor, palpitations, and headache. There may be a rapid and severe rise in blood pressure. Withdrawal symptoms typically occur about 20 hours after cessation of use. Clonidine use must always be tapered when it is discontinued. Because of its short half-life, some patients experience withdrawal symptoms between doses.

DRUG INTERACTIONS

Coadministration of clonidine and tricyclic drugs can inhibit the hypotensive effects of clonidine. Trazodone (Desepyl) has the potential to produce hypotension and sedation when combined with clonidine. Any antihypertensive agent or drug that causes hypotension as a side effect may amplify blood pressure drops if coadministered with clonidine. Clonidine may also enhance the CNS depressive effects of barbiturates, alcohol, and other sedative-hypnotic agents. Concomitant use of b-adrenergic receptor antagonists can increase the severity of rebound phenomena, including hypertension, when clonidine use is discontinued. Yohimbine, an a_2-adrenergic receptor antagonist, blocks the pharmacological effects of clonidine. The reported sudden deaths of children taking concurrent methylphenidate and clonidine remains unexplained. They may represent isolated instances of children with preexisting cardiovascular abnormalities who either experienced adverse events unrelated to medication or had reactions to the clonidine alone.

LABORATORY INTERFERENCES

Clonidine is not known to interfere with any commonly used clinical laboratory tests.

DOSAGE AND ADMINISTRATION

Clonidine is available in 0.1-, 0.2-, and 0.3-mg tablets. The usual starting dosage, 0.1 mg orally twice a day, can be raised by 0.1 mg a day to an appropriate level.

Therapeutic dosages for hypertension commonly range between 0.2 and 1.2 mg a day. The maximum recommended daily dosage is 1.2 mg. When tolerance to the hypotensive effects of clonidine develops, the dosage should be increased or a diuretic be added. There is a dose-response curve.

The transdermal patch delivers 0.1, 0.2, or 0.3 mg every 24 hours and is applied every 7 days. Transdermal clonidine is sometimes useful to improve compliance. It also reduces fluctuations in clonidine plasma concentrations. The recommended starting dose for all adults is the lowest dose. The initial patch contains a dosage of 0.1 mg a day for both children and adults. The maximum dosage, which is determined by both clinical response and tolerability, is about 4 mg a day. The maintenance dosage for a child is 4.5 µg/kg daily, and for an adult is 8 µg/kg daily. The patch is effective for only 5 days in children—as opposed to 7 days in adults—because of more rapid hepatic metabolism.

SUGGESTED CROSS-REFERENCES

Alcohol-related disorders are discussed in Section 11.2, nicotine-related disorders in Section 11.4, and opioid-related disorders in Section 11.9. The biochemical aspects of mood disorders are covered in Section 14.3. Attention-deficit disorders are discussed in Chapter 39, and tic disorders in Chapter 42.

SECTION REFERENCES

Ahmed I, Takeshita J: Clonidine: A critical review of its role in the treatment of psychiatric disorders. CNS Drugs *6:*53, 1996.

*American Society of Addiction Medicine: Detoxification: principles and protocols. In *The Principles Update Series: Topics in Addiction Medicine*. American Society of Addiction Medicine, Chevy Chase, MD, 1997.

Aulakh CS, Mazzola-Pomietto P, Murphy DL: Long-term antidepressant treatment restores clonidine's effect on growth hormone secretion in a genetic animal model of depression. Pharmacol Biochem Behav *55:*265, 1996.

*Bremner JD, Krystal JH, Southwick SM, Charney DS: Noradrenergic mechanisms in stress and anxiety, II. Clinical studies. Synapse *23:*39, 1996.

Cantwell DP, Swanson J, Connor DF: Case study: Adverse response to clonidine. J Am Acad Child Adolesc Psychiatry *36:*539, 1997.

El-Kadi AOS, Sharif SI: The influence of chronic treatment with clonidine, yohimbine and idazoxan on morphine withdrawal. Psychopharmacology *132:*67, 1997.

Harmon RJ, Riggs PD: Clonidine for posttraumatic stress disorder in preschool children. J Am Acad Child Adolesc Psychiatry *35:*1247, 1996.

Holloway M, King WD, James LP: Clonidine poisoning in Jefferson County Alabama. Ann Emerg Med *29:*511, 1997.

Kellner M, Yassouridis A, Jahn H, Wiedemann K: Influence of clonidine on psychopathological, endocrine and respiratory effects of cholecystokinin tetrapeptide in patients with panic disorder. Psychopharmacology *133:*55, 1997.

*Oesterheld J, Tervo R: Clonidine: A practical guide for usage in children. SD J Med *49:*234, 1996.

Piletz JE, Halaris AE, Chikkala D, Qu Y: Platelet I1-imidazoline binding sites are decreased by two dissimilar antidepressant agents in depressed patients. J Psychiatr Res *30:*169, 1996.

*Prince JB, Wilens TE, Biederman J, Spencer TJ, Wozniak JR: Clonidine for sleep disturbances associated with attention-deficit hyperactivity disorder: A systematic chart review of 62 cases. J Am Acad Child Adolesc Psychiatry *35:*599–605, 1996.

*Sallee FR, Richman H, Sethuraman G, Dougherty D, Sine L, Altman-Hamamdzic S: Clonidine challenge in childhood anxiety disorder. J Am Acad Child Adolesc Psychiatry *37:*655, 1998.

*Schmidt ME, Matochik JA, Goldstein DS, Schouten JL, Zametkin AJ, Potteer WZ: Gender differences in brain metabolic and plasma catecholamine responses to alpha $_2$-adrenoreceptor blockade. Neuropsychopharmacology *16:*298, 1997.

*Tallarida RJ, Stone DJ Jr, McCary JD, Raffa RB: Response surface analysis of synergism between morphine and clonidine. J Pharmacol Exp Ther *289:*8, 1999.

Tancer ME, Stein MB, Black B, Uhde TW: Blunted growth hormone responses to growth hormone releasing factor and to clonidine in panic disorder. Am J Psychiatry *150:*336, 1993.

31.17 DOPAMINE RECEPTOR ANTAGONIST (TYPICAL ANTIPSYCHOTICS)

STEPHEN R. MARDER, M.D., AND DANIEL P. VAN KAMMEN, M.D., IH.D.

Before 1990 the traditional antipsychotic drugs were the only specific agents available in the United States for the treatment of psychosis. These agents are highly effective for nearly all disorders that result in psychotic thought processes. However, these drugs have important limitations: they can cause acute and chronic neurological symptoms that significantly limit their effectiveness; they are ineffective or only partially effective for a substantial proportion of psychotic patients; and they are relatively ineffective for some of the most important signs and symptoms of the most common psychotic illness, schizophrenia. A new generation of antipsychotic agents that includes clozapine (Clozaril), risperidone (Risperdal), sertindole (Serlect), olanzapine (Zyprexa), and other drugs may be less severely limited in all three of these areas. These agents have in common similar or greater activity at serotonin (5-hydroxytryptamine [5-HT]-type 2A) (5-HT $_{2A}$) receptors than at dopamine type 2 (D_2) receptors. In contrast, the therapeutic response to the traditional antipsychotics is thought to be associated with their antagonism at D $_2$ receptors. For this reason, the two groups of drugs can be designated dopamine receptor antagonists and serotonin-dopamine antagonists (SDAs). (The term *dopamine receptor antagonist* is a relatively new designation for these agents and is not widely accepted at the time of this publication.) Other terms, which are used interchangeably, are *traditional, conventional*, and *typical antipsychotic*. Since the SDAs have only recently become available, the relative indications for these two groups of agents are not firmly established. However, it remains likely that the use of dopamine receptor antagonist will decrease as the use of SDAs increases.

Despite their limitations, dopamine receptor antagonists have been important in the management of psychosis. The modern era of psychopharmacology began with the discovery of chlorpromazine (Thorazine), the first effective antipsychotic, in the early 1950s. The annual increase in state and county mental hospital beds in the United States that had occurred since the beginning of the twentieth century reversed shortly after the arrival of antipsychotics, because psychotic patients with schizophrenia and other illnesses could be discharged from hospitals after they received drug treatment. Moreover, patients with stabilized illnesses became candidates for psychosocial treatments and rehabilitation. Evidence also suggests that treatment with antipsychotics can improve the long-term course of schizophrenia.

The dopamine receptor antagonists are antipsychotic rather than antischizophrenic. That is, they are effective for treating psychosis, regardless of its cause. As a result, antipsychotics are standard treatments for many patients with secondary psychotic disorders, bipolar disorders, major depressive disorder with psychotic features, and schizoaffective disorder.

Although the introduction of the SDAs will reduce the proportion of patients receiving dopamine receptor antagonists, these agents will continue to be important for the treatment of psychosis. Under a number of conditions these agents will continue to be preferred for antipsychotic treatment. For example, currently, none of the SDAs is available as a long-acting agent or a short-acting intramuscular drug. Also, patients who respond well to dopamine receptor antagonists without discernable adverse effects may do as well on the less costly older drugs. Others may find that adverse effects of SDAs such as ejaculatory disturbances in men and weight gain are of greater concern than the extrapyramidal adverse effects associated with dopamine receptor antagonists. For these reasons psychiatric clinicians must remain familiar with this group of agents.

HISTORY

The first antipsychotic drugs were the phenothiazines. Drugs from this class were originally used as antihelminthics in veterinary medicine and as urinary antiseptics in humans. In 1950 Paul Charpentier at the Rhône-Poulenc Laboratories in Paris synthesized chlorpromazine, a mild antihistaminic that appeared notable as a sedating agent. Henri Laborit, a surgeon interested in drugs that would decrease preoperative anxiety and prevent postsurgical shock, convinced the Rhône-Poulenc laboratories of the importance of identifying compounds that would relax patients and make surgical shock less likely. Chlorpromazine appeared to fulfill these criteria, and also appeared qualitatively different from other sedating drugs. Patients appeared indifferent toward the environment and more tranquil. These properties led Laborit to attempt to convince psychiatrists, including Jean Delay and Pierre Deniker, to administer chlorpromazine to patients with mania and schizophrenia. In 1952 they reported that chlorpromazine was effective in treating mania and schizophrenia. Within a year, the atmosphere in Paris psychiatric hospitals had improved substantially. The discovery of chlorpromazine's effectiveness for psychosis marked the beginning of modern psychopharmacology by showing that drugs could have specific effects on the symptoms of mental disorders that went beyond sedation. Moreover, the availability of drugs with direct effects on mental disorders became a powerful tool for laboratory studies with animals; that is, understanding how drugs affect the brain could provide information on the illnesses that were affected.

The use of chlorpromazine was reported by Heinz Lehman from Montreal in 1954, and it rapidly captured the interest of U.S. psychiatrists. In 1955 Henry Brill, the assistant state mental health commissioner in New York, initiated the general use of chlorpromazine and reserpine in state hospitals in that state. Although reserpine was also an effective antipsychotic, its adverse effects led clinicians to prefer chlorpromazine, and as a result, it is rarely used for treating psychosis.

In 1953 reports surfaced indicating that chlorpromazine caused parkinsonian symptoms. Although some individuals warned that these symptoms were evidence of neural toxicity, others suggested that the dosage should be increased until patients experienced a mild form of parkinsonism. Later that year, clinicians noted that chlorpromazine could also elicit forms of persistent dyskinesia, a syndrome later called *tardive dyskinesia*.

The introduction of chlorpromazine was followed by the introduction of other phenothiazines including perphenazine (Trilafon) and fluphenazine (Prolixin). In 1958 the first effective butyrophenone, haloperidol (Haldol), was introduced by Paul Janssen from Belgium. The first thioxanthene antipsychotics were introduced the same year by P.V. Peterson and his coworkers.

Although many other dopamine receptor antagonists were introduced in subsequent years, they were all similar in effectiveness and differed only in their adverse-effect profiles. The first long-acting antipsychotic, fluphenazine enanthate was introduced in the early 1960s. Although this and subsequent long-acting agents were similar to other antipsychotics in effectiveness and adverse effects, they provided a way to administer an antipsychotic without requiring patients to take daily oral medication.

CHEMISTRY

The dopamine receptor antagonists have been referred to by a number of names including neuroleptics, antipsychotics, and major tranquilizers. The designation *neuroleptic* was suggested because of the tendency of these agents to have neurological adverse effects and thus to appear to “seize the neuron.” This term is based on the observation that neurological adverse effects are inevitably associated with the antipsychotic activity of these agents. However, the effectiveness of drugs such as clozapine suggests that neurological adverse effects are not necessary to treat psychosis; thus, referring to an adverse effect of these agents rather than the primary clinical effect may be misleading. The term *major tranquilizer* referred to the strange quietness or blandness (or ataraxia) that had been associated with these agents. However, the suggestion that these agents function by their tranquilizing effects is also misleading. These agents, as a group, can be most economically called *antipsychotic drugs*, thus referring to their target symptoms. Differentiating two classes of antipsychotics, the dopamine receptor antagonists and the SDAs, appears to be meaningful since it suggests a mechanism that underlies important clinical differences between these two groups of agents.

The dopamine receptor antagonists may be subclassified according to either their chemical structure or their clinical effects. In the section that follows these agents

will be classified according to their chemical structure. Another system for subclassifying this group uses the antipsychotic potency of these agents (Table 31.17-1). According to this system, agents are classified as low-, mid-, and high-potency agents. This method is probably more useful to clinicians because it provides information about the amount of drug required for a clinical effect and the likelihood of important adverse effects.

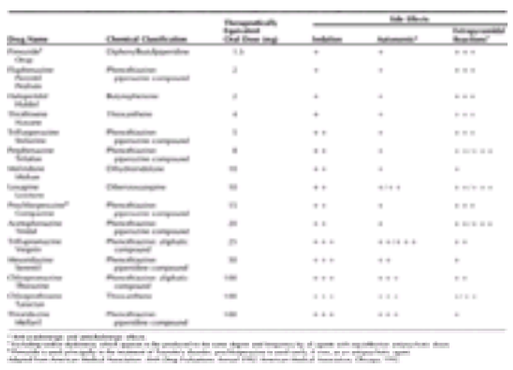

Table 31.17-1 Dopamine Receptor Antagonists: Potency and Adverse Effects

Phenothiazines All phenothiazines have the same three-ring structure with different side chains joined at the nitrogen atom of the middle ring (Fig. 31.17-1). Their activity can be affected by substitutions at positions 2 or 10. The phenothiazines are usually divided into three classes on the basis of substitutions at position 10. These substitutions have important effects on the pharmacological characteristics of the phenothiazine. Substituting an aliphatic side chain results in a drug with relatively low antipsychotic potency such as chlorpromazine. These agents also tend to be sedating and to cause anticholinergic and hypotensive effects at their effective dosages. Chlorpromazine has a chlorine atom at position 2. Eliminating the chlorine atom yields promazine, a weak antipsychotic. Substituting a piperidine ring at position 10 leads to an effective group of antipsychotics that includes thioridazine (Mellaril) and mesoridazine (Serentil). These drugs and the aliphatic phenothiazines have similar potencies and adverse-effect profiles. Fluphenazine and trifluoperazine (Vesprin) are representative antipsychotics with a piperazine group substituted at the 10 position. Piperazine phenothiazines have fewer anticholinergic and autonomic effects but greater affinity for D_2 sites, and as a result they produce more extrapyramidal adverse effects. A number of piperazine phenothiazines have been esterified at a free hydroxyl group with ethanoic or decanoic acid to produce injectable, long-acting antipsychotics.

Figure 31.17-1 Molecular structure of dopamine receptor antagonists and reserpine.

Thioxanthenes The thioxanthenes share a similar three-ring structure with the phenothiazines but with a carbon atom substituted for nitrogen at position 10. Chlorprothixene is an effective low-potency aliphatic thioxanthene agent with antipsychotic potency and an adverse-effect profile similar to those of its phenothiazine counterpart, chlorpromazine. Thiothixene (Navane), clopenthixol (not marketed in the United States), and flupenthixol (not marketed in the United States) are thioxanthenes with a piperazine substitution. Again, their clinical characteristics resemble those of their phenothiazine counterparts. As a result of their double bond at position 10, these agents have geometric isomers; the *cis* isomers demonstrate greater antipsychotic activity.

Butyrophenones Butyrophenones are characterized by a substituted phenyl ring attached to a carbonyl group that is linked by a 3-carbon chain to a tertiary amino group. Most of the clinically useful butyrophenones have a piperidine ring attached to the tertiary amino group. Haloperidol is the representative drug for this class. It and other butyrophenones tend to be potent D_2 antagonists with minimal anticholinergic and autonomic effects. Haloperidol, a substituted piperidine, is the most commonly used drug from this class. Other important butyrophenones include droperidol (Inapsine), a short-acting, sedating compound commonly used to control agitation, and spiperone, an agent with a very high affinity for D_2 receptors, commonly used to label receptors in animal studies and in positron emission tomography (PET) scanning.

Dibenzoxazepines Dibenzoxazepine antipsychotic drugs have a three-ring structure with a seven-member center ring. Loxapine (Loxitane) is the only drug from this group that is available in the United States. Clozapine, a dibenzodiazepine, serotonin-dopamine antagonist, differs from loxapine in having a nitrogen instead of an oxygen atom in the middle ring, as well as differences in the side chains.

Dihydroindoles The dihydroindoles are structurally related to serotonin, melatonin, and indole hallucinogens such as dimethyltryptamine. The only dihydroindole available in the United States is molindone (Moban).

Diphenylbutylpiperidines The diphenylbutylpiperidines are similar in structure to the butyrophenones. Pimozide (Orap), the only diphenylbutylpiperidine available in the United States, is derived from the butyrophenone benperidol. Penfluridol and fluspirilene are antipsychotics from this group that are available outside of the United States.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics The antipsychotic drugs come from a number of chemical classes, each with somewhat different pharmacokinetic properties (Table 31.17-2 and Table 31.17-3), which can have important effects on the clinical use of the drugs. The phenothiazines and the thioxanthenes share structural similarities and have a number of shared metabolic pathways. The same is true for the butyrophenones and the diphenylbutylpiperidines. This overview focuses on chlorpromazine and haloperidol as representative compounds. Since the pharmacokinetics of the long-acting agents differ substantially from those of the short-acting agents, they are discussed separately.

Drug	Bioavailability (%)	Protein Binding (%)	Vd (L/kg)	$t_{1/2}$ (h)	Active Metabolites	Plasma Concentration 12 Hours Postdose (ng/mL)
Chlorpromazine	≤10–15	90–95	7–20	8–15	7-Hydroxy, others	100–300
Thioridazine	25–35	99	–	5–10	Mesoridazine, sulforidazine	200–800
Perphenazine	~25	90–95	10–15	8–12	None known	0.5–5.0
Fluphenazine	20–50	90–95	–	14–24	7-Hydroxy*	0.2–2.0
Thiothixene	~30	90–95	–	34	None known	1.0–5.0
Haloperidol	45–70	92	10–15	12–36	Reduced haloperidol	2.0–10.0
Fluspirilene	45–60	90–93	12–24	22–36	None known	1.0–10.0

*Questionable activity.
Reprinted with permission from Baldessarini RJ. Pharmacokinetics and drug interactions. Update for new antipsychotics. J Clin Psychiatry. 57(suppl):C:1996.

Table 31.17-2 Comparison of the Pharmacokinetics of Different Antipsychotics

Age
Elderly patients demonstrate greater variability in drug clearance
The "super-elderly" (those greater than 80 years of age) demonstrate the most consistent reduction in clearance rates
Genetics
Polymorphism at CYP 2D6 and CYP 2C19
Wide intrapatient variability at CYP 3A
Ethnic differences in metabolism
Substance abuse behaviors
Cigarette smokers demonstrate increased metabolic capacity, CYP 1A2
Ethanol abusers can demonstrate increased or diminished metabolic capacity dependent on the extent of their drinking behavior and their hepatic and nutritional status
Medical conditions
Decreased hepatic blood flow can reduce clearance, e.g., congestive heart failure
Hepatic disease such as cirrhosis and hepatitis can decrease clearance
Enzyme inducers
Carbamazepine, phenytoin, ethanol/etoh, barbiturates
Clearance inhibitors
Examples include: serotonin selective reuptake inhibitors, tricyclic antidepressants, cimetidine, beta-blockers, isoniazid, methylphenidate, erythromycin, triazolobenzodiazepines, chloramphenicol, ciprofloxacin, ketoconazole, and many others
Changes in binding proteins
Stress (acute phase reactants increase such as α_1 acid glycoprotein increase)
Hypoalbuminemia can occur with malnutrition or hepatic failure
Changes in protein binding occur due to hepatic or renal failure

Reprinted with permission from Baldessarini RJ. Pharmacokinetics and drug interactions. Update for new antipsychotics. J Clin Psychiatry. 57(suppl):C:1996.

Table 31.17-3 Variables That Influence the Pharmacokinetics of Antipsychotics

Plasma concentrations of short-acting drugs rise rather rapidly during the so-called absorption phase and then decline during the distribution, metabolic, and elimination phases. Each of these three latter phases is influenced by different factors related to the chemical structure of the agent.

Absorption In general, the dopamine receptor antagonists are well absorbed when they are administered orally or parenterally. As with most drugs, oral administration results in less-predictable absorption than parenteral administration. Liquid concentrates are absorbed slightly more rapidly than pills. Plasma concentrations of the drugs usually peak 1 to 4 hours after ingestion and 30 to 60 minutes after intramuscular administration. In general, intramuscular preparations reach their peak concentrations faster than oral drugs and thus have an earlier onset of action. For example, intramuscular administration of most antipsychotics results in peak plasma concentrations in about 30 minutes, with clinical effects apparent within 15 to 30 minutes. Most orally administered dopamine receptor antagonists result in peak plasma concentrations 1 to 4 hours after administration. A number of factors can interfere with gastrointestinal absorption of these drugs, including antacids, coffee, smoking, and food.

Drug concentrations usually reach steady state in about 3 to 5 times their half-lives. Thus, steady-state levels for chloropromazine, haloperidol, and most other dopamine receptor antagonists are reached in about 3 to 5 days, since their half-lives are about 24 hours. The bioavailability (i.e., the amount of drug reaching the systemic circulation) increases substantially—as much as 10-fold—when dopamine receptor antagonists are administered parenterally. This difference may reflect incomplete absorption of the drug in the gastrointestinal tract and extensive metabolism of oral drugs during the first pass through liver and gut.

Distribution The initial decline in the plasma concentration of an antipsychotic (or any drug) results primarily from drug distribution into various bodily compartments. Since antipsychotics are highly lipophilic, they tend to accumulate in tissues such as fat, lungs, and brain. Brain concentrations of antipsychotics tend to be substantially higher than those in plasma. Since plasma and brain concentrations are in equilibrium, measuring plasma levels concentrations suffices for estimating concentrations in brain.

Most dopamine receptor antagonists are highly protein bound. For example, more than 90 percent of drugs such as fluphenazine and haloperidol is bound to plasma protein. The remaining unbound portion is the drug available to pass through the blood-brain barrier. In theory, conditions that alter the amount of plasma protein (e.g., malnutrition) alter the amount of bioavailable antipsychotic drug.

Metabolism and Elimination The metabolism of antipsychotic drugs is largely hepatic and occurs through conjugation with glucuronic acid, hydroxylation, oxidation, demethylation, and sulfoxide formation. The metabolism of the phenothiazines and thioxanthenes is particularly complex. For example, chlorpromazine has more than 100 different potential metabolites, with some metabolites having significant amounts of pharmacological activity. For thioridazine, a substantial amount of the drug's activity is from metabolites that may be more active than thioridazine itself. One metabolite, mesoridazine, is also marketed as an antipsychotic. On the other hand, haloperidol has only one major metabolite, reduced haloperidol, which has substantially less antidopaminergic activity than the parent compound. However, reduced haloperidol is converted back to the parent compound and thus may contribute to antipsychotic activity.

Most dopamine receptor antagonists are metabolized by the cytochrome P450 (CYP) 2D6 and CYP 3A isoenzymes. Since these same isoenzymes also metabolize a number of drugs that are commonly used in combination with antipsychotics, a number of important drug-drug interactions are possible.

Patients demonstrate substantial differences in the extent to which they metabolize these compounds, which explains, in part, the marked differences among patients in the oral dosages required. This, along with differences in absorption, may also explain the 20-fold variation in plasma concentrations found among patients treated with the same dosage of an antipsychotic.

The systemic clearance of dopamine receptor antagonists is high because of a high hepatic extraction ratio. As a result, only negligible amounts of the unchanged drug are excreted by the kidneys. Phenothiazines, thioxanthines, and their metabolites are excreted in the urine and the feces.

Pharmacokinetics of Long-Acting Antipsychotics The pharmacokinetics of long-acting injectable antipsychotics differ markedly from those of short-acting oral and injectable drugs. Long-acting fluphenazine and haloperidol are administered as esters dissolved in sesame oil. The oil is injected into a muscle, and the drug gradually diffuses from the oily vehicle into the surrounding tissues. The rate-limiting step appears to be the rate of diffusion; once the drug enters the tissue, it is rapidly hydrolyzed and the parent compound is released. Long-acting compounds, on the other hand, are absorbed continually during the interval between injections. Moreover, patients who have received multiple injections absorbed the drug from multiple injection sites simultaneously. As a result, it takes long-acting compounds much longer to reach steady state, and they are eliminated much more slowly. For example, the decanoate forms of haloperidol and fluphenazine require about 3 months to reach steady state, and substantial plasma concentrations can be detected months after therapy has been discontinued.

Elderly Patients Elderly patients require lower-than-usual dosages of dopamine receptor antagonists for a number of reasons, including decreased renal clearance, decreased cardiac output, decreased liver size, and decreased P450 activity. In addition, elderly patients may be more sensitive to the parkinsonian adverse effects of the drugs.

Pharmacodynamics The pharmacodynamics of an antipsychotic describe its activity at important sites of action, particularly at its brain receptors. Neuroscience research has recently provided important information about the effects of dopamine receptor antagonists at a number of different receptor sites, how these effects influence groups of neurons, and how activity at receptors is translated into intracellular events.

Mechanism of Action of Dopamine Receptor Antagonists

Dopamine Hypothesis The dopamine hypothesis states that antipsychotics reduce psychotic symptoms by decreasing dopamine activity. It was originally proposed by Arvid Carlsson from Sweden and other basic scientists, based on the observation that haloperidol and chlorpromazine increased the concentration of dopamine metabolites in dopamine-rich areas of the mouse brain and had inconsistent effects on other neurotransmitters. They interpreted these findings as indicating that these two agents were acting as dopamine antagonists and that self-regulating systems were compensating by increasing dopamine production. Other investigators noted that dopamine receptor antagonists inhibited amphetamine-induced activation and stereotypical behaviors of rodents. This effect has been found to be a reasonably reliable predictor of antipsychotic activity in humans. Moreover, the dose-response relation in animals provides information about the likely clinical dosage range for patients. These behaviors are thought to be mediated by dopamine and have been used to screen for compounds that are likely to have clinical efficacy as antipsychotics. Other evidence supporting a role of dopamine in psychosis comes from the observation that all drugs—at least to date—capable of reducing psychosis are also dopamine receptor antagonists. Moreover, agents associated with increased dopamine activity such as amphetamines, methylphenidate (Ritalin), or cocaine tend to increase dopamine neurotransmission or are themselves dopamine agonists.

A role for dopamine in the activity of antipsychotics was given strong support by the work of Solomon Snyder from Johns Hopkins University and Philip Seeman from the University of Toronto in the mid 1970s. Using similar receptor binding methods, both found that the antipsychotic potency of a number of drugs was highly correlated with the drug's affinity for D_2 receptors. Recent studies using highly selective D_2-receptor ligands and PET indicate that dopamine receptor antagonists are effective when approximately 80 percent of D_2 receptors in the brain are occupied (Table 31.17-4). Higher occupancy rates may be associated with more adverse effects without greater effectiveness. In contrast clozapine is effective when only 40 to 60 percent of D_2 receptors are occupied. Although why clozapine is effective at lower levels of D_2 occupancy is unclear, it has been suggested that activity at other receptors, particularly 5-HT_{2A} receptors, may contribute to clozapine's clinical activity (Fig. 31.17-2).

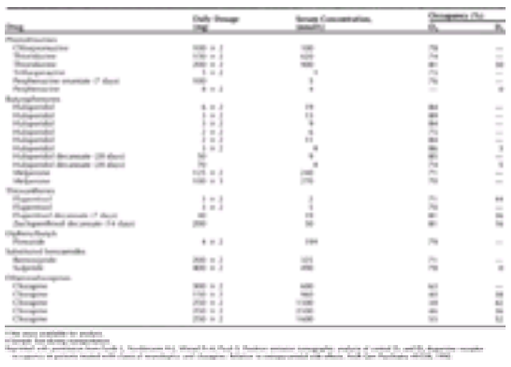

Table 31.17-4 Dopamine Receptor Occupancy of Antipsychotics in Recent Studies

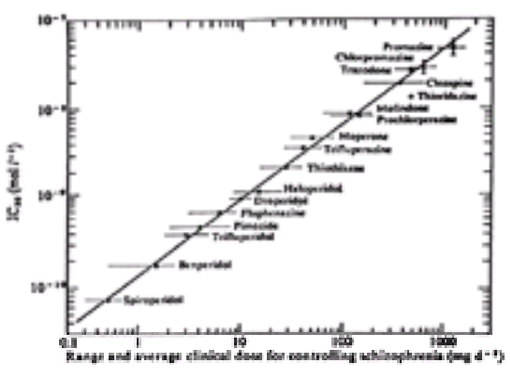

Figure 31.17-2. Blockade of dopamine release. (Reprinted with permission from Seeman P, Lee T: Antipsychotic drugs: Direct correlation between clinical potency and presynaptic action on dopamine neurons. Science *188:*1217, 1975.)

Other studies using PET scanning have found no difference in D_2 receptor occupancy in the striatum between individuals who respond to an antipsychotic and those who do not. This finding suggests that processes beyond the dopamine receptor affect response or that dopamine blockade in the striatum does not mediate behavior change.

Effects of Dopamine Receptor Antagonists on the Dopamine System Dopamine neurons with cell bodies in the ventral tegmental area (the A_{10} cell group in rat brain) project via the mesolimbocortical dopamine system to the nucleus accumbens, amygdala, and the neocortex. Dopamine cells in the substantia nigra (the A_9 cell group) project to the caudate putamen via the nigrostriatal system. It has been proposed that activity in the mesolimbocortical system explains the antipsychotic activity of drugs whereas neurological adverse effects such as extrapyramidal effects and tardive dyskinesia result from activity in the nigrostriatal system. Thus drugs with preferential activity at mesolimbocortical sites would exhibit clinical effectiveness with fewer neurological adverse effects.

Recent advances in molecular biology have permitted cloning of a number of dopamine receptor subtypes. The discovery of these subtypes has raised the question of their possible role in the treatment of psychosis. Dopamine receptors can be classified into two groups, D_1-like (D_1 and D_5) and D_2-like receptors (D_2, D_3, D_4). Moreover, each of these receptor subtypes may have variants, which can potentially explain heterogeneity in drug responses. The antipsychotic potency of dopamine receptor antagonists is highly correlated with their affinity for D_2 receptors. This relationship has not been established with other dopamine receptor subtypes. Nevertheless, certain characteristics of receptor subtypes suggest that they may contribute to clinical responses. For example, a role for D_4 receptors has been suggested by clozapine's high affinity for this subtype. Also, the distribution of D_1 receptors in cortical areas and the localization of D_3 receptors in limbic areas suggest that these subtypes may be clinically important for antipsychotic effects.

Recent studies of the dopamine system in animals and man have suggested important interactions between the subcortical and prefrontal dopamine systems that may affect responses to dopamine receptor antagonists. Lesions in the prefrontal cortex of rats increase dopamine activity in the striatum, suggesting important functional linkages. Similarly, increases in dopamine activity in the prefrontal cortex are associated with decreases in striatal activity. A number of investigators have proposed that negative symptoms of schizophrenia are associated with decreased dopamine receptor antagonist prefrontal activity, while positive symptoms are generated by subcortical increases in dopamine receptor antagonist activity. In this model, an effective antipsychotic would increase dopamine receptor antagonist activity in prefrontal areas and decrease activity in subcortical areas.

Other clinical evidence for a role of dopamine in the antipsychotic effects of dopamine receptor antagonists comes from studies of homovanillic acid, a metabolite of dopamine. In these investigations, changes in plasma concentrations of homovanillic acid in patients taking antipsychotics are assumed to reflect changes in dopamine activity in the brain. These studies indicate that plasma homovanillic acid concentration rises soon after patients begin treatment with an antipsychotic and then falls below the pretreatment concentration in patients who improve. Moreover, a number of studies indicate that patients with a high concentration of homovanillic acid in plasma prior to drug treatment are more likely to improve than those with low concentrations. These findings suggest that improvement on an antipsychotic agent is associated with an overall reduction in dopamine activity. Since subcortical areas contribute most to dopamine activity in the brain, this finding may not reflect cortical dopamine activity.

Studies that monitor the intracellular firing rates of dopamine neurons have also improved understanding of the effects of antipsychotics. L.A. Chiodo and Benjamin S.

Bunney reported that administration of haloperidol to rats results in a short-term increase in the firing of midbrain dopamine neurons followed by a prolonged decrease in firing, referred to as “depolarization inactivation.” It has been proposed that the antipsychotic effects of these drugs is associated with the decreased firing. Taken together with the studies of plasma homovanillic acid concentration, these findings suggest that an antipsychotic response requires the blockade of D_2 receptors followed by a decreased dopamine activity.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Design The dopamine receptor antagonists were the first specific compounds for treating a major mental illness. Their effectiveness challenged the views of some clinicians who believed that psychosis was a psychological reaction related to early experiences and that only psychological treatment could have antipsychotic effects. This skepticism helped inspire the development of controlled clinical drug trials aimed at demonstrating the effectiveness of antipsychotics. Early studies, such as those developed by the Veterans Administration and the National Institute of Mental Health (NIMH) were influential in establishing objective research methods. From those trials a number of principles emerged that have guided the study of antipsychotics and other drugs in psychiatry and medicine.

One principle is that to be considered effective, an antipsychotic or any other drug should yield a better outcome than a placebo. That principle controls for nonspecific improvement that occurs because of such factors as (1) spontaneous fluctuations in symptoms; (2) the rate of spontaneous improvement for the illness; (3) the expectation of improvement, which is related to starting treatment with a new drug; (4) the subject's protection from environmental stresses that may occur with hospitalization; and (5) other environmental factors such as the hospital milieu and psychological treatments that can lead to improvement.

Two other principles that reduce sources of bias are random assignment to a treatment group and double-blind methods. Random assignment ensures that differences in outcome between groups (usually between those receiving a drug and those receiving a placebo or between two different active-treatment groups) is not attributable to patients with different expected outcomes or drug responses being assigned to one group. (For example assigning more severely ill patients or those predicted to improve spontaneously to a particular group.) Double-blind methods were developed to ensure that bias by patients or by those who evaluate the outcome does not influence the study. (For example, a person who performs a clinical rating might rate patients receiving a drug as more improved than those receiving a placebo.) Double-blind methods also protect against bias in the way treatment is administered. (For example, a treating clinician might give more attention to a patient who is receiving a certain drug.)

A number of modifications in trial methods have been developed since the first studies of dopamine receptor antagonists, although the principles of random assignment and blindness remain important. For example, treating acutely ill patients with schizophrenia with dopamine receptor antagonists has become a standard of care and has led some to question the use of placebos. Alternative methods for demonstrating the efficacy of a compound include showing that it is as effective as an antipsychotic agent with proved efficacy or showing that a dosage thought to be subtherapeutic is less effective than a higher dosage. Since blindness is often difficult to preserve when a drug has known adverse effects, active placebos with side effects but without central nervous system (CNS) effects have been used.

Interpretation

Acute Psychosis Numerous double-blind studies have demonstrated that dopamine receptor antagonists are more effective than placebos for relieving symptoms of psychosis. The most compelling studies were carried out in the 1960s by the NIMH and have been replicated by many research groups since then. These studies also demonstrated that antipsychotics are effective for every subtype and subgroup of patients with schizophrenia. In 1995 the Schizophrenia Patient Outcomes Research Team (PORT) updated prior reviews and concluded that dopamine receptor antagonists are clearly effective in reducing positive symptoms during acute psychotic episodes. Donald Klein and John Davis reviewed studies that compared more than one antipsychotic and found that with the exception of mepazine and promazine, all of these agents were equally effective.

Controlled clinical trials have demonstrated that antipsychotic medications are effective in diminishing most symptoms of schizophrenia. Thus, symptoms such as hallucinations and paranoid ideation are improved by these agents whereas anxiety and depression are relatively unaffected. The extent to which patients improve on these agents varies considerably. Dopamine receptor antagonists can lead to complete remission of psychotic symptoms. Studies from the NIMH indicate that 60 percent of drug-treated patients achieve a complete remission or experience only mild symptoms; the remaining 40 percent continue to experience psychotic symptoms. Approximately 8 percent of patients do not improve with drug treatment. The 1995 PORT found that dopamine receptor antagonists can induce remission of positive symptoms in about 70 percent of patients.

The so-called positive or psychotic symptoms—which include hallucinations and delusions—are more responsive to dopamine receptor antagonists than are negative symptoms such as blunted affect, emotional withdrawal, and lack of social or vocational interest or disorganized symptoms such as impaired attention and disorganized thought processes. Although the latter two groups of symptoms are less responsive than psychotic symptoms, double-blind studies indicate that they do improve to some degree. However, the responses are complex. Extrapyramidal effects can cause negative symptoms or worsen them. Patients with parkinsonism may demonstrate fewer expressive gestures and diminished interest in their environment. Anticholinergic antiparkinson medications may impair memory in patients who require these drugs to treat parkinsonism and other extrapyramidal adverse effects.

All forms and types of schizophrenia improve with dopamine receptor antagonists. Moreover, no evidence suggests that any particular subtype of schizophrenia responds better to a particular type of antipsychotic. Although it was proposed that low-potency, sedating dopamine receptor antagonists were more effective for agitated patients and high-potency, nonsedating drugs were more effective for more-withdrawn patients, this was never substantiated in controlled trials. As a group, female patients tend to respond better and to require lower dosages than male patients. Unfortunately, the only indicator that can be used to predict patient response to an antipsychotic is the patient's prior responses.

Maintenance Therapy A number of studies have compared the ability of dopamine receptor antagonists and placebos to prevent relapse in patients who have been stabilized following a psychotic episode. These studies provide strong support for continued dopamine receptor antagonist treatment after patients have recovered from a psychotic episode. A review of 24 double-blind studies of maintenance antipsychotic therapy found that in all studies many more patients relapsed on placebo compared to those who received an antipsychotic drug. A meta-analysis that included most of these studies reported that approximately 72 percent of patients relapsed in a year when given a placebo compared with only 23 percent of those given a conventional antipsychotic. A more recent review cited 66 studies in which antipsychotic use was discontinued. The mean relapse rate was 53 percent for patients withdrawn from medication and 16 percent for patients maintained on antipsychotic treatment.

A number of recent investigations have focused on the usefulness of continuing medication for patients who have recovered from a first episode of schizophrenia. Again, maintenance medication bestowed a substantial advantage, although relapse rates were lower for first-episode patients than for multiepisode patients. Other studies indicate that patients who have been well and stable on antipsychotic therapy for years have a high likelihood of relapse when their medications are discontinued.

Studies by Donald Johnson suggest other advantages for maintenance antipsychotics in schizophrenia. He found that patients who relapsed when they were receiving dopamine receptor antagonists had episodes that were less severe than those in patients who had discontinued taking their drugs. Drug-maintained patients were less likely to have episodes with self-destructive behaviors, violence, and antisocial acts than patients who were not receiving medication. Patients who were not receiving medication were also more likely to require involuntary hospitalization. Patients who discontinued their medications ended up, on average, receiving more total medication because the dosage needed to treat acute relapses is much higher than the dosages usually prescribed for preventing relapse.

A number of statistics also indicate the limitations of pharmacotherapy for chronic schizophrenia. Because medication compliance is a serious problem for patients with schizophrenia, up to 50 percent of patients fail to comply with recommended treatments. Even when patients respond to treatment, 10 percent are permanently disabled. In addition, more than 80 percent of patients are not fully employed and require public support. In large cities, individuals with schizophrenia comprise 14 to 50 percent of the homeless. The rate of completed suicide is about 10 percent, and the mortality rate is increased because of greater accident proneness.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Effects on Motor Systems Animals given high doses of a dopamine receptor antagonist develop a syndrome with immobility, increased muscle tone, and abnormal postures called catalepsy. In addition, these agents decrease spontaneous motor activity in rodents. These motor effects in animals and humans are caused by dopamine receptor blockade in the striatum and inactivation of dopamine neurons in the substantia nigra.

In humans, all of the traditional antipsychotic medications produce extrapyramidal adverse effects including parkinsonism (i.e., stiffness, tremor, and shuffling gait), dystonia (i.e., abrupt onset, sometimes bizarre muscular spasms affecting mainly the musculature of the head and neck), and akathisia (i.e., objective and subjective restlessness). In addition, prolonged treatment with dopamine receptor antagonists can lead to the development of late-appearing abnormal movements called tardive dyskinesia.

Neuroendocrine Effects Antipsychotic drugs influence the secretion of hormones in the pituitary and elsewhere, mainly because of their blockade of dopamine receptors. Dopamine—either through the blockade of receptors in the hypothalamus or in the pituitary itself—affects the secretion of a number of hormones, particularly prolactin. Prolactin secretion is tonically inhibited by dopamine. As a result, dopamine receptor antagonists increase serum prolactin concentration in the usual clinical dose range, and this ability is correlated with their antipsychotic potency. Increased prolactin secretion sometimes produces gynecomastia and galactorrhea. Dopamine receptor antagonists also suppress luteinizing hormone (LH) and follicle-stimulating hormone concentrations. These changes can lead to amenorrhea and inhibition of orgasms in women. Evidence also indicates that dopamine receptor antagonists can inhibit the release of growth hormone (GH).

THERAPEUTIC INDICATIONS

Indications for dopamine receptor antagonists are shown on Table 31.17-5.

Acute psychotic episodes in schizophrenia and schizoaffective disorder
Maintenance treatment in schizophrenia and schizoaffective disorders
Mania
Depression with psychotic symptoms
Delusional disorder
Borderline personality disorder
Substance-induced psychotic disorder
Delirium and dementia
Mental disorders due to a medical condition
Childhood schizophrenia
Pervasive developmental disorder
Tourette's disorder
Huntington's disease

Table 31.17-5 Indications for Dopamine Receptor Antagonists

Acute Psychotic Episodes in Schizophrenia and Schizoaffective Disorder Schizophrenia is usually characterized by periods of relapse and remission. Acute psychotic episodes are periods during which patients demonstrate the emergence or worsening of positive or disorganized symptoms of schizophrenia such as hallucinations, delusions, and disorganized thoughts.

Maintenance Treatment in Schizophrenia and Schizoaffective Disorders Dopamine receptor antagonists are effective in decreasing the risk of psychotic relapse in patients who have recovered from a psychotic episode. An international consensus conference on guidelines for relapse prevention recommended that patients who had multiple prior psychotic episodes continue maintenance therapy for at least 2 to 5 years and often indefinitely. One to 2 years of maintenance treatment was recommended for patients following a first episode of psychosis.

Mania Dopamine receptor antagonists are effective in reducing excitement and psychotic symptoms in mania. These agents tend to have a more rapid onset of action than antimanic drugs including lithium (Eskalith), valproic acid (Depakene), and carbamazepine (Tegretol). This observation has led to the widespread practice of using combined treatment with antipsychotic and antimanic drugs during the first days of treatment of severe excited states, before the antimanic compound has its onset of action. When the antimanic compound has become effective, the dosage of antipsychotic can usually be reduced or use of the drug can be discontinued. Although only limited evidence supports the use of antipsychotics for long-term management of bipolar disorder, it is common clinical practice to manage some patients with severe illness with maintenance antipsychotics combined with antimanic agents. Some studies indicate that patients with mood disorders are more apt to develop tardive dyskinesia than patients with schizophrenia, suggesting the importance of using these drugs for as short a time as possible and only when clearly indicated.

Depression With Psychotic Symptoms Patients with major depressive disorder may develop psychotic symptoms such as hallucinations or delusions. These patients tend to exhibit faster improvement when an antipsychotic is added to an antidepressant agent. Although most studies have focused on perphenazine, fluphenazine, and haloperidol, no evidence suggests that any dopamine receptor antagonist is more effective than any other for this indication. The added benefits of dopamine receptor antagonists are most apparent for patients tormented by severe delusions. When the psychotic symptoms have remitted, antipsychotic use should be discontinued.

Delusional Disorder Delusional disorder is characterized by nonbizarre delusions, that is delusions within the realm of possibility for the patient. Moreover, patients with this disorder seldom hallucinate, and their behavior is seldom bizarre. Patients with delusional disorder often improve substantially when they are treated with a dopamine receptor antagonist.

Borderline Personality Disorder Borderline personality disorder is characterized by a pervasive pattern of instability that affects mood, interpersonal relationships, and self-image. Some patients with severe forms of this disorder may experience transient psychotic states, particularly when they are under stress. These states may be characterized by suspiciousness, ideas of reference, impulsiveness, and aggressiveness.

Low dosages of thiothixene and haloperidol have been found effective in some of these patients, and no evidence indicates that other dopamine receptor antagonists are less effective. Unfortunately, most trials have been relatively brief, and it is unclear how medications affect the overall course of this disorder or how they interact with the psychotherapeutic management these patients usually require.

Substance-Induced Psychotic Disorder A number of drugs including amphetamines, cocaine, alcohol, and phencyclidine can cause psychotic symptoms while the patient is intoxicated or during drug withdrawal. These symptoms are usually time limited and remit within hours or days, depending upon the concentration of the drug ingested and the rate of elimination of the agent. Although dopamine receptor antagonists may reduce psychotic symptoms, most of these episodes do not require treatment with antipsychotics unless the symptoms cause dangerous behaviors or unusual suffering. Benzodiazepines may help calm the patient until psychotic symptoms remit. Some authors recommend prescribing benzodiazepines rather than dopamine receptor antagonists for phencyclidine intoxication. In some circumstances it may be helpful to follow the patient's untreated course, to differentiate drug-induced syndromes from schizophrenia or bipolar I disorder.

Dopamine receptor antagonists are also useful for patients with hallucinations or delusions that occur with alcoholism, particularly chronic alcoholic hallucinosis. They should be prescribed carefully since these drugs may increase the likelihood of seizures that occur during alcohol withdrawal. As a result, benzodiazepines are more appropriate than antipsychotics for uncomplicated alcohol withdrawal.

Delirium and Dementia Low dosages of high-potency dopamine receptor antagonists can be useful in the treatment of the psychotic symptoms and agitation associated with delirium that may be due to organic conditions. Interactions with the anticholinergic properties of low-potency drugs may exacerbate or prolong toxic deliriums caused by anticholinergic agents.

Dopamine receptor antagonists can be useful for reducing agitation and psychotic symptoms related to dementia, Wernicke's encephalopathy, and Korsakoff's syndrome. A number of careful double-blind studies have found these drugs superior to placebo for treating agitated elderly patients. A review by Philip Janicak found that 66 percent of geriatric patients with organic psychosis improved when given an antipsychotic (versus 35 percent of those given a placebo). Low dosages of high-potency dopamine receptor antagonists are usually preferable for elderly patients with dementia. Low-potency medications can cause orthostatic hypotension and falling episodes. In addition, the anticholinergic effects of low-potency drugs can worsen cognitive and memory impairments. The treatment of elderly patients is complicated since many receive multiple medications and are vulnerable to drug-drug interactions. Nevertheless, the adverse effects of these drugs should be

weighed against the dangers of agitation in elderly patients.

Mental Disorders Due to a Medical Condition Dopamine receptor antagonists are effective for a broad range of medical conditions that result in psychosis. This category includes endocrinopathies and psychoses associated with temporal lobe epilepsy. Dopamine receptor antagonists are also helpful for patients with psychotic symptoms that result from steroid use.

Early-Onset Schizophrenia Schizophrenia infrequently has its onset during childhood. Children with schizophrenia may have symptoms similar to those in adults with schizophrenia. Although clinical experience indicates that children improve with antipsychotics, few controlled trials are available to demonstrate this effect. Children must be treated with the lowest effective dosage of an antipsychotic, since these drugs may impair learning.

Pervasive Developmental Disorders Patients with pervasive developmental disorders (e.g., autistic disorder) may display periods of hyperactivity, screaming, and agitation with combativeness. Although few controlled studies exist for this disorder, patients tend to improve with high-potency agents. Since these agents may impair learning, there is concern that they are overprescribed in some settings.

Impulse-Control Disorders Prescribing dopamine receptor antagonists for individuals with extremely poor impulse control and a propensity for violent behavior is a long-standing, but controversial, practice. This practice is not supported by controlled clinical trials and may be appropriate only when other measures such as the use of carbamazepine, lithium, or b-adrenergic receptor antagonist (beta-blockers) has failed.

Tourette's Disorder Tourette's disorder is a neurobehavioral disorder associated with motor and vocal tics. Motor tics may involve a single muscle group or several, thus appearing as purposeful movements. Similarly, vocal tics may involve an inarticulate sound or actual words. Often the words are socially inappropriate or even obscene. Milder tics may not require pharmacotherapy. However, when the tics are disabling, an antipsychotic medication can reduce the severity of both motor and vocal tics. Although haloperidol and pimozide are the most commonly used agents, no evidence indicates that other agents are less effective.

Huntington's Disease Genetically transmitted, Huntington's disease follows an autosomal dominant type of heritability with 100 percent penetrance. The median age of onset is 35 to 40 years, but childhood onset and much later onsets also occur. The most common mental symptom is depression, which may progress to a psychotic delusional state, sometimes followed by manic or hypomanic episodes. Individuals may exhibit symptoms that closely resemble schizophrenia with delusions and hallucinations.

Antipsychotics are useful during the early stages of the illness to reduce chorea. They are also effective in managing irritability, paranoia, hallucinations, violent tendencies, and bizarre behaviors. High-potency agents are usually preferred, but all agents appear to be effective. Clinicians should be aware of acute extrapyramidal symptoms, particularly in the rigid forms of the disorder.

Other Indications Antipsychotics are also effective for treating ballismus, an illness characterized by abnormal activity of the axial and proximal musculature. Because of their antiemetic effects, dopamine receptor antagonists are sometimes prescribed for severe emesis associated with chemotherapy. Patients with intractable hiccups are sometimes helped by dopamine receptor antagonists. Metoclopramide (Reglan), a dopamine receptor antagonist, is commonly prescribed for patients with gastroesophageal reflux and delayed gastric emptying.

PRECAUTIONS AND ADVERSE REACTIONS

Neurological Adverse Effects The most serious adverse effects of dopamine receptor antagonists are neurological and largely confined to the extrapyramidal motor system (Table 31.17-6). These extrapyramidal effects are, in turn, separated into acute effects that occur in the first days or weeks of treatment and chronic effects that occur after patients have received medications for months or years.

Acute extrapyramidal syndromes
Akathisia
Acute dystonia
Parkinsonism
Neuroleptic malignant syndrome
Chronic extrapyramidal syndromes
Tardive dyskinesia
Perioral tremor

Table 31.17-6 Neurological Adverse Effects of Dopamine Receptor Antagonists

Acute Extrapyramidal Syndromes All of the dopamine receptor antagonists are associated with extrapyramidal adverse effects. These effects—particularly parkinsonism—are attributed to decreased dopamine activity in the basal ganglia, which suggests that they are inescapable, since blockade of D_2 receptors is associated with the antipsychotic effect of these agents. This relation also explains why the potency of dopamine receptor antagonists for causing extrapyramidal adverse effects is related to their affinity for D_2 receptors in the basal ganglia. Acute extrapyramidal effects, which include dystonia, akathisia, and drug-induced parkinsonism, have their onset relatively soon after the initiation of antipsychotic drug treatment and remit soon after drug use is discontinued. In contrast, tardive extrapyramidal syndromes occur after months of treatment and commonly persist after medication use is discontinued.

AKATHISIA Akathisia comes from the Greek word *akathizin*, which means inability to sit still. It is the most common acute extrapyramidal adverse effect and often the most distressing. A recent study found that 41 percent of patients treated with an antipsychotic had mild akathisia, and an additional 21 percent had moderate-to-severe akathisia. The most common clinical manifestations consist of shifting the weight from foot to foot, walking in place, inability to keep the legs still, feelings of inner restlessness, and shifting of body positions in a chair. Patients with akathisia may describe a compelling urge to walk or initiate movement. In mild cases, patients may experience a subjective feeling of restlessness but not show increased motor activity. Akathisia may appear in the second or third day of antipsychotic treatment but more frequently has its onset after 5 days.

Differentiating akathisia from behaviors related to the psychotic illness may be difficult. Feelings of restlessness may result in irritability or anxiety. Psychotic patients with akathisia may also appear more hostile or belligerent. This could lead clinicians to conclude incorrectly that the patient requires an increase in medication, which in turn could worsen the extrapyramidal effect. Akathisia is also thought to be a correlate of poor antipsychotic drug response.

ACUTE DYSTONIA Acute dystonias consist of intermittent and sustained spasms of the muscles of the trunk, head, and neck, leading to involuntary movements. They can be experienced as the most frightening adverse effect of dopamine receptor antagonists, particularly when patients experience the sudden onset of what may appear to be a form of paralysis.

Ten percent of drug-induced dystonias occur during the first hours of drug treatment, and 90 percent occur within the first 3 days. Common types of dystonia include opisthotonos, a rigid contraction of the back muscles with arching; retrocollis and torticollis of the neck; oculogyric crisis, a spasm in which an eye or both eyes are turned upward; macroglossia and tongue protrusion, which can lead to choking; and laryngeal dystonias. Rarely, dystonias of laryngeal or pharyngeal muscles can lead to sudden death.

Younger patients, particularly young males, are more likely to develop dystonias. They are more common when patients are treated with large doses of high-potency dopamine receptor antagonists. Dystonias almost always respond rapidly to antiparkinson medications and can usually be prevented by either pretreatment with antiparkinson medications or by limiting the antipsychotic dosage prescribed. Dystonic reactions occur in about 40 percent of patients treated with high-potency drugs

without prophylactic antiparkinsonian medications.

PARKINSONISM Patients taking dopamine receptor antagonists may experience all of the common motor symptoms of idiopathic parkinsonism including rigidity, bradykinesia, shuffling gait, and tremor. These adverse effects commonly occur during the first 5 to 30 days of treatment and may persist until the dosage is lowered or use of the drug is discontinued. Examination usually reveals a positive glabella tap. This motor disturbance affects about 30 percent of patients treated long-term with dopamine receptor antagonists. The first evidence of drug-induced parkinsonism may be a diminished arm swing or decreased facial expressiveness.

Risk factors for drug-induced parkinsonism include increasing age, dose, a history of parkinsonism, and underlying basal ganglia damage (e.g., vascular insult).

In its milder forms, neuroleptic-induced parkinsonism may appear as a decrease in spontaneous gestures, a masked facial expression, apathy, unspontaneous speech, and difficulty in initiating usual activities—so-called akinesia. These symptoms may be difficult to distinguish from the negative or deficit symptoms of schizophrenia. Patients with akinesia may also appear to be depressed. The tendency of akinetic patients not to cross their legs can be helpful in diagnosing this adverse effect.

MANAGEMENT OF ACUTE EXTRAPYRAMIDAL SYNDROMES Anticholinergic agents are the most effective treatment for extrapyramidal adverse effects. The acute dystonias are alleviated by intramuscular or intravenous diphenhydramine (Benadryl) (50 mg, 25 mg in children), intramuscular or intravenous benztropine (Cogentin), or other anticholinergics within minutes after the drug enters the bloodstream. Oral anticholinergic agents are usually prescribed for parkinsonism and often for akathisia. Reducing the dosage of the antipsychotic is often effective, but patients commonly experience extrapyramidal adverse effects at the minimum dosage needed to manage psychosis. Amantadine (Symmetrel), an antiviral agent that affects dopamine release, also effectively reduces most extrapyramidal adverse effects. Direct dopamine agonists (e.g., bromocriptine [Parlodel]) and indirect dopamine agonists (e.g., levadopa [Larodopa]) are not preferred, as they may aggravate psychotic symptoms.

Anticholinergics are frequently, but not always, effective for treating akathisia. b-Adrenergic receptor antagonists, particularly propranolol (Inderal; 10 to 80 mg a day), have been found effective for treating akathisia and are preferred by many clinicians. b-Adrenergic receptor antagonists may be effective because akathisia is a disturbance in the normal balance between dopamine and norepinephrine.

Anticholinergic antiparkinson medications should be prescribed with caution in combination with low-potency dopamine receptor antagonists and some antidepressants because the combined anticholinergic activity may lead to a number of adverse effects including hyperthermia and delirium. Sustained tachycardia may also occur. Further caution is warranted in the use of these agents in individuals with glaucoma, ileus, urinary retention, or prostatic hypertrophy. Because many over-the-counter cold remedies, sedatives, and sleep aids contain anticholinergic agents and antihistamines, patients should be warned about the interaction between these agents and anticholinergic antiparkinson drugs.

Some controversy surrounds the appropriateness of prescribing antiparkinson medications prophylactically when treatment with dopamine receptor antagonists is initiated. Reports indicate that 30 to 50 percent of patients receiving long-term treatment do not need anticholinergics. In 1990 the World Health Organization (WHO) published a consensus statement on the use of anticholinergics in patients receiving long-term antipsychotic treatment. The statement argued against the prophylactic use of these agents because of some of their adverse effects (e.g., possible worsening of tardive dyskinesia, risk of anticholinergic toxicity, impaired memory, and possible pharmacokinetic interactions with the primary treatment agent). The consensus group concluded that anticholinergics are overprescribed, that they should be given only as needed, and that adverse effects should first be treated by lowering the antipsychotic dosage or by using other agents. However, reports indicate that anticholinergics may be helpful in managing forms of extrapyramidal adverse effects that the clinician may not recognize as adverse side effects (i.e., milder forms of akathisia may appear as anxiety or irritability and milder forms of akinesia may appear as negative symptoms such as decreased motivation and a lack of affect). In these individuals, long-term anticholinergics may be useful for optimal pharmacotherapy.

Prophylactic antiparkinson medications may be helpful in other circumstances. Because of the discomfort and anxiety associated with acute dystonias, prophylaxis is prudent when there is a substantial risk of this effect, for example when young patients are treated with high-potency dopamine receptor antagonists. Some patients may interpret a dystonia as evidence that they are allergic to a particular drug, which may lead to drug refusal. Prudent practice may include using prophylactic antiparkinson drugs in high-risk patients but gradually lowering their dosage, with regular evaluations for obvious and subtle manifestations of extrapyramidal adverse effects.

SDAs are an important alternative for patients with extrapyramidal adverse effects. Patients with extrapyramidal effects that cannot be adequately controlled with dosage reduction or an adjunctive medication will probably improve if they are given a newer drug.

NEUROLEPTIC MALIGNANT SYNDROME Neuroleptic malignant syndrome is an uncommon but potentially fatal complication of dopamine receptor antagonists. Its main clinical features include (1) hyperthermia; (2) severe muscular rigidity; (3) autonomic instability including hyperthermia, tachycardia, increased blood pressure, tachypnea, and diaphoresis; and (4) changing levels of consciousness. It usually presents as muscular rigidity and progresses to elevated temperature, fluctuating consciousness, and unstable vital signs. These symptoms are often associated with increased creatine phosphokinase (CPK) and aldolase activity. Less common are increased liver transaminase activity, leukocytosis, myoglobinemia and myoglobinuria. Acute renal failure may also occur. Mortality in well-developed cases has been reported to range from 20 to 30 percent and may be higher when depot forms are used. Recent improvements in the early recognition of neuroleptic malignant syndrome have substantially improved the mortality rate.

The neuroleptic malignant syndrome is more common when high-potency dopamine receptor antagonists are prescribed at high dosages and when dosage is escalated rapidly. The syndrome is twice as common in males as in females and is more likely to occur in younger patients. Clinicians should be concerned about any patient who displays severe muscular rigidity and a rising body temperature, since early diagnosis and treatment can be life saving.

Management of Neuroleptic Malignant Syndrome When neuroleptic malignant syndrome is diagnosed or suspected, antipsychotic use should be discontinued and supportive and symptomatic treatment should begin. This may include treating extrapyramidal adverse effects with antiparkinson medications, correcting fluid and electrolyte imbalances, treating fever, and managing cardiovascular symptoms such as hypertension or hypotension. Recent studies indicate that dantrolene (Dantrium) may be effective for treating severe neuroleptic malignant syndrome. Treatment begins with intravenous administration of 0.8 to 2.5 mg/kg every 6 hours, with a maximum of 10 mg/kg daily. When the symptoms subside and the patient can swallow, dantrolene is administered orally in dosages of 100 to 200 mg daily. Bromocriptine can then be added at dosages of 20 to 30 mg daily in four divided doses. Amantadine may be helpful if the other agents do not suffice. The course of treatment is commonly 5 to 10 days unless a long-acting injectable agent has been used.

Although neuroleptic malignant syndrome is considered by some to be an extreme extrapyramidal adverse effect, its cause and biological risk factors are unknown. Most patients require antipsychotic medications after they recover from neuroleptic malignant syndrome, and usually either the same dopamine receptor antagonist or a different one can be used. Low-potency drugs may be preferable for patients who have previously had neuroleptic malignant syndrome. However, even clozapine can cause the syndrome. At this time there are no data to indicate if clozapine or other SDAs are associated with a lower risk of neuroleptic malignant syndrome than traditional antipsychotics.

Chronic Extrapyramidal Syndromes

TARDIVE DYSKINESIA Tardive dyskinesia is a movement disorder that may occur following long-term treatment with antipsychotic medications. Patients with tardive dyskinesia may have any or all of a number of abnormal movements, which frequently consist of mouth and tongue movements, such as lip smacking, sucking and puckering as well as facial grimacing. Other movements may include irregular movements of the limbs, particularly choreoathetoid-like movements of the fingers and toes and slow, writhing movements of the trunk. Younger patients tend to develop slower athetoid movements of the trunk, extremities, and neck. Although seriously disabling, dyskinesia is uncommon, in a small proportion it may affect walking, breathing, eating, and talking.

The differential diagnosis of tardive dyskinesia includes a number of disorders of the basal ganglia (including Huntington's disease, Wilson's disease, and Sydenham's chorea), hyperthyroidism, hypoparathyroidism, tardive Tourette's disorder, and dyskinesias related to other drugs, such as levodopa and amphetamines. Abnormal movements were also described in patients with schizophrenia before the introduction of antipsychotic medications.

Although antipsychotic medications frequently suppress the movements of tardive dyskinesia, increasing the medication dosage is only appropriate when the movements are severe and incapacitating. When patients have their dosages of antipsychotic medication either decreased or discontinued, abnormal movements

may worsen temporarily or appear for the first time and then diminish. Abnormal movements that occur in this context have been referred to as *withdrawal emergent dyskinesias*. Antiparkinson medications often worsen the movements of tardive dyskinesia. However, when patients suffer from combined tardive dyskinesia and acute extrapyramidal adverse effects, antiparkinson medication is often unavoidable.

Although the biological mechanism underlying tardive dyskinesia is still controversial, a number of investigators have hypothesized that the disorder is related to increased sensitivity of dopamine receptors in the basal ganglia. This is supported by the observation that drugs that block dopamine such as antipsychotic medications suppress tardive dyskinesia, whereas dopamine agonists worsen the disorder. These observations are consistent with findings from preclinical laboratories indicating that blocking dopamine receptors can result in their upward regulation.

At least 10 to 20 percent of patients treated with dopamine receptor ntagonists for more than a year develop tardive dyskinesia. In long-term institutionalized patients the prevalence is between 15 and 20 percent. Recent prospective studies indicate that approximately 4 to 5 percent of patients who are undergoing antipsychotic drug treatment will develop tardive dyskinesia each year, and certain populations are at a greater risk than others. Increasing age increases risk, with elderly women being particularly vulnerable. A recent prospective study by Margaret Woerner and coworkers found that cumulative rates of tardive dyskinesia in an elderly cohort were 25 percent, 34 percent and 53 percent after 1, 2, and 3 years of treatment with an antipsychotic. Moreover, tardive dyskinesia in elderly adults is less likely to remit when use of dopamine receptor antagonists is discontinued. Patients with affective disorders may also be at greater risk for tardive dyskinesia when they are treated with dopamine receptor antagonists. Other possible risk factors for tardive dyskinesia are the dosage of antipsychotic medications and the duration of treatment with these drugs.

Early observations of the course of tardive dyskinesia suggested that the disorder was inevitably progressive and irreversible. In other words, patients who developed even mild dyskinesias were likely to progress toward severe tardive dyskinesia. More-recent evidence indicates otherwise. When dopamine receptor antagonist therapy is discontinued, a substantial proportion of tardive dyskinesia patients enjoy a remission, particularly those with tardive dyskinesia of rather recent onset. Sometimes this remission will take several months to occur.

Tardive dyskinesia does not appear to be a progressive disorder for most patients. The disorder seems to develop rather rapidly and then stabilize and often improve. A number of studies followed the course of tardive dyskinesia in patients who continued to take drugs for several years and found that the severity of tardive dyskinesia improves in most patients, even if treatment with dopamine receptor antagonists continues. Moreover, this improvement can be clinically meaningful in some patients. Individuals who are most likely to remit are those with a recent onset and those with an onset prior to age 40. This remission may take months or longer to occur.

Perioral Tremor This is an uncommon side effect of dopamine receptor antagonists that appears after months or years of drug treatment. The peculiar tremors of the mouth have led to this disorder being called “rabbit syndrome.”

Management of Tardive Dyskinesia The authors recommend that all patients with schizophrenia who receive dopamine receptor antagonists for more than 6 months have regular, systematic evaluations for tardive dyskinesia. Although instruments such as the Abnormal Involuntary Movement Scale (AIMS) were originally developed as research tools, they also provide a structured system for evaluating and recording abnormal movements. Thus, performing an AIMS test once every 3 to 6 months is a reasonable way to monitor patients.

Once tardive dyskinesia is diagnosed, the clinician should evaluate the need for continued antipsychotic treatment. This decision requires considering the dangers of both the psychotic illness and the tardive dyskinesia. Whenever possible this decision should be made in collaboration with both patient and family, and the results of this discussion should be documented in the patient's medical record. For most patients with chronic schizophrenia, the best decision will probably involve continuing the antipsychotic. If this is the case, some evidence suggests that using the lowest effective dosage may minimize the risk of tardive dyskinesia.

Patients with severe tardive dyskinesia may also be candidates for a trial of clozapine. Most clinical reports indicate that clozapine can suppress tardive dyskinesia. In addition, substantial data indicate that clozapine is less likely to cause the disorder than other antipsychotic medications and therefore is promising for the long-term management of patients who have developed tardive dyskinesia. This potential advantage of clozapine should be considered tentative, since data on the incidence of tardive dyskinesia in patients on clozapine are inadequate. In addition, studies using animal models indicate that clozapine use does not evoke the increased sensitivity of striatal D_2 receptors that is associated with typical neuroleptics and has been theorized to result in tardive dyskinesia. At this time data are inadequate to indicate whether newer SDAs are effective in suppressing tardive dyskinesia.

The American Psychiatric Association task force on tardive dyskinesia recommends a number of steps for the prevention and management of the disorder, including (1) establishing objective evidence that antipsychotic medications are effective; (2) using the minimum effective dosage for long-term treatment; (3) exercising caution with children, elderly adults, and patients with mood disorders; (4) examining patients on a regular basis for evidence of dyskinesia and noting the results of the examination in the medical record; (5) if tardive dyskinesia is diagnosed, considering alternatives to antipsychotics, obtaining informed consent, and considering dosage reduction; and (6) if the dyskinesia worsens, considering discontinuation of the antipsychotic, use of a different drug, or a trial of clozapine.

Cardiovascular Effects

Cardiac Adverse Effects Animal studies have shown that antipsychotic agents may decrease cardiac contractility, disrupt enzyme activity in cardiac cells, decrease tissue concentrations and increase circulating concentrations of catecholamines, and prolong atrial and ventricular conduction time and refractory periods. Low-potency agents—particularly chlorpromazine and thioridazine—can induce electrocardiographic (ECG) abnormalities such as prolongation of the QT and PR intervals, blunting of the T waves, and depression of the ST segment. Malignant arrhythmias (including torsade de pointes) have been associated with thioridazine. Clinicians should consider obtaining an ECG before and during treatment with thioridazine when patients have preexisting cardiac pathology or prolonged QT intervals. Patients with a history of cardiac disease are probably most safely managed with a high-potency antipsychotic such as haloperidol.

Orthostatic (Postural) Hypotension Low-potency dopamine receptor antagonists (e.g., chlorpromazine and thioridazine) may cause orthostatic hypotension, particularly when these agents are first administered. This adverse effect can be dangerous when it results in fainting or falls. Both elderly patients and individuals with impaired balance should be monitored carefully for orthostatic changes. Since tolerance may develop gradually, patients should initially be treated with a relatively low dosage of these agents (e.g., 25 mg orally twice daily) and the dosage should be titrated upward over several days until a target dosage is reached. When appropriate, patients should be warned of this adverse effect and told to sit or stand up slowly, to sit with their feet outside the bed when arising, to wait for as long as a minute, and to sit or lie down if they feel faint. Some patients—particularly elderly adults—may have persisting orthostatic hypotension and thus may benefit from a change to high-potency antipsychotic treatment.

Orthostatic hypotension is more likely when low-potency agents are administered intramuscularly. Thus, clinicians should consider measuring the patient's blood pressure (both lying and standing) before and 1 to 2 hours after the first dose and before subsequent doses during the first days of treatment.

If hypotension does occur, even with higher-potency agents, it can usually be managed by having patients lie down and instructing them to arise slowly. Some patients may benefit from stockings, volume expansion, or pressor agents. The use of epinephrine paradoxically worsens antipsychotic-induced hypotension through b-adrenergic stimulation and is thus contraindicated. Pure a-adrenergic agents such as metaraminol (Aramine) and norepinephrine (Levophed) are the drugs of choice for this condition.

Sudden Unexplained Death In spite of reports associating dopamine receptor antagonists with sudden unexplained death, no evidence indicates that the incidence of such deaths changed in psychiatric patients after introduction of antipsychotic therapy. Similarly, the incidence of sudden unexplained death does not differ in patients on high-potency drugs and those on low-potency drugs. An American Psychiatric Association task force report concluded:

> [Al]though a relationship between the use of antipsychotic drugs and sudden death has not been firmly established, it has also not been disproved. From a neurocardiologic perspective, these drugs have the potential for both increasing and decreasing the risk of sudden death. Ultimate outcome is probably determined by a multitude of interacting factors, and the role played by the drug in a given individual is difficult, if not impossible to determine.

Reports of sudden death in patients treated with antipsychotics could be explained by cardiac arrhythmias or other causes. For example, seizures, asphyxiation, heat stroke, malignant hyperthermia, and neuroleptic malignant syndrome could result in relatively sudden death. Individuals who may be more vulnerable to sudden death

include elderly adults, physically debilitated individuals, and agitated patients (particularly those who are physically restrained).

Gastrointestinal Tract The most common gastrointestinal adverse effects of dopamine receptor antagonists are related to their peripheral anticholinergic effects. These include dry mouth, constipation, occasional diarrhea, and urinary retention. Nausea and vomiting are less common. Anticholinergic effects are most prevalent with low-potency dopamine receptor antagonists such as chlorpromazine or thioridazine and are uncommon with high-potency drugs. These effects can be exacerbated when other anticholinergic drugs such as tricyclic drugs are combined with a low-potency antipsychotic.

Patients who experience a severely dry mouth should be advised to rinse their mouth frequently and to chew sugarless gum or candy. Sugar-containing candies may lead to fungal infections in the mouth and increase the risk of dental caries. Constipation can be managed with stool softeners or laxatives. Clinicians should manage severe constipation carefully, since it can progress to paralytic ileus.

Liver Dopamine receptor antagonists may cause transient abnormalities in liver function test results. These are seldom of serious medical concern. Obstructive or cholestatic jaundice, a serious condition that was commonly associated with chlorpromazine treatment in the past, is currently uncommon.

Blood Dopamine receptor antagonist therapy may be associated with temporary and transient depression of leukopoiesis. Mild leukopenia may occur for prolonged periods, although more commonly, the leukocyte count eventually returns to normal. Agranulocytosis is rare with dopamine receptor antagonists, occurring in fewer than 1 in 10,000 cases. Thrombocytopenia and pancytopenia may also occur.

Kidney and Urinary Function Chlorpromazine has weak diuretic effects. The anticholinergic effects of low-potency drugs may lead to urinary retention.

Skin and Eye Effects Cutaneous reactions seen during the first weeks of antipsychotic drug treatment include urticarial, maculopapular, petechial, and edematous reactions. These reactions usually improve when drug use is discontinued and sometimes clear even when drug use continues. Patients receiving low-potency dopamine receptor antagonists (particularly chlorpromazine) may develop photosensitivity reactions consisting of severe sunburn or rash. Patients should be instructed to avoid direct sunlight or use sunscreens.

Low-potency dopamine receptor antagonists, especially chlorpromazine, are associated with an uncommon discoloration of the skin. Skin areas that are exposed to sunlight, particularly the face and neck develop blue-gray, metallic discoloration. This skin reaction is usually associated with long-term treatment with high drug dosages. Changing to treatment with a high-potency drug usually leads to gradual improvement.

Patients receiving long-term treatment with chlorpromazine may develop granular deposits in the anterior lens and posterior cornea. These deposits—visualized on slit lamp examination—seldom affect the patient's vision. Changing to another drug usually yields gradual improvement in the condition.

High dosages of thioridazine—above 1000 mg daily—can cause retinal pigmentation, which can lead to serious visual impairment or blindness. Moreover, this condition may not remit when thioridazine therapy is discontinued. As a result, thioridazine should not be prescribed at dosages above 800 mg daily.

Endocrine Effects Dopamine receptor antagonists increase circulating prolactin concentrations. In women, increased prolactin concentration can lead to breast enlargement, galactorrhea, and irregular menses including anovulatory cycles and infertility, menses with abnormal luteal phases, or frank amenorrhea and hypoestrogenemia. Increased prolactin concentration associated with dopamine receptor antagonists may also suppress testosterone in men. Prolactin concentration elevation results directly from D_2 blockade in the tuberoinfundibular dopamine pathway and the pituitary. Since prolactin concentration is under tonic inhibitory control by dopamine, blockade of dopamine receptors increases prolactin concentration. Some patients—but not all—develop tolerance to the effects of these drugs on prolactin concentration after several weeks. Dosage reduction may relieve symptoms related to increased prolactin concentration.

Effects on Sexual Function Dopamine receptor antagonists can have important effects on sexual functioning in both men and women. In men use of these agents can decrease libido and cause ejaculatory and erectile disturbances. Erectile dysfunction reportedly occurs in 23 to 54 percent of males receiving an antipsychotic and can be an important cause of medication noncompliance in men. Retrograde ejaculation has been reported with some dopamine receptor antagonists, particularly thioridazine. In addition, these agents can affect ejaculatory latency and the quantity of ejaculate. Since sexual functioning is affected by dopamine, serotonergic, and adrenergic systems it is unclear how sexual dysfunction resulting from dopamine receptor antagonists is mediated. Dosage reduction may be helpful in some individuals.

Women who receive dopamine receptor antagonists may experience decreased libido, anorgasmia, and decreased lubrication. Although a number of adjunctive medications have been reported, there is little evidence that any are effective. As with men, reducing the dosage of the antipsychotic may be helpful.

Pregnancy Although dopamine receptor antagonists cross the placenta, little evidence suggests any association between prenatal exposure to an antipsychotic and increased incidence of congenital malformations. A recent meta-analysis by Lori Altshuler and colleagues found that the baseline incidence of congenital anomalies is 2.0 percent and the incidence with antipsychotics is 2.4 percent. This suggests that these agents add a small additional risk to the fetus. This meta-analysis found no evidence that high-potency drugs increase the risk to the fetus.

Prenatal exposure of rats to dopamine receptor antagonists may lead to a decreased number of brain dopamine receptors. Other animal studies indicate that prenatal haloperidol exposure may raise cholesterol concentrations in offspring. It is unclear if these animal findings indicate risks in humans.

In developing a treatment for pregnant patients with schizophrenia, the relatively low risk of congenital abnormalities associated with antipsychotics should be weighed against the dangers of psychotic episodes. Clinicians may attempt to discontinue these medications during the first trimester if this is feasible and should consider carefully whether the risks of taking these drugs during the remainder of the pregnancy are justified by the likely benefits. If patients have a history of relapse when treatment with dopamine receptor antagonists is discontinued, the drugs are relatively safe. If dopamine receptor antagonists are prescribed, high-potency compounds are probably safer.

Lactation Dopamine receptor antagonists are secreted in low concentrations with lactation, suggesting that patients should be advised against breast-feeding during antipsychotic treatment.

Seizure Threshold Dopamine receptor antagonists use tends to lower the seizure threshold. Thus they are associated with an increased risk of seizures. This risk is usually clinically important in individuals with a prior history of seizures. Low-potency drugs have a greater tendency to lower the seizure threshold. Therefore, high-potency dopamine receptor antagonists may be preferable for seizure-prone individuals.

Temperature Regulation Effects Some patients become poikilothermic (i.e., they become hypothermic in the winter and overheat in the summer) when taking chlorpromazine or other low-potency agents. Special precautions should be taken for these patients during exposure to temperature extremes.

Overdosage Overdosages with dopamine receptor antagonists are usually characterized by exaggeration of their usual adverse effects. Thus, high-potency drugs are associated with severe extrapyramidal reactions, including dystonias and severe rigidity, as well as hypotension and sedation. Lower-potency drugs are more likely to be associated with central nervous system (CNS) depression, sedation, anticholinergic effects, and hypotension. The clinical picture may also include agitation, restlessness, convulsions, fever, autonomic reactions such as dry mouth and ileus, ECG changes, and cardiac arrhythmias. In severe overdoses with dopamine receptor antagonists the pupils are mydriatic, deep tendon reflexes decrease and areflexia may occur, tachycardia and hypotension are present, and the electroencephalogram (EEG) shows diffuse slowing and low voltage. This picture may progress to delirium and coma, with respiratory depression and hypotension.

The management of dopamine receptor antagonist overdose is symptomatic and supportive, since overdoses with these agents are seldom lethal. However, the outcome may be unfavorable when these agents are combined with others, particularly CNS depressants such as alcohol, barbiturates, and benzodiazepines. If the overdose is recent, gastric lavage is helpful. If the patient can swallow, activated charcoal is indicated. Emetics should not be administered because their effects may be attenuated by the antipsychotic and because they are associated with a risk of aspiration pneumonia. Aspiration may be a serious risk if the patient develops a dystonic reaction of the head or neck. Forced diuresis and hemodialysis are not effective because dopamine receptor antagonists are highly protein bound and lipophilic.

When hypotension occurs, standard measures for managing circulatory shock should be implemented. The hypotension may occur when b-adrenergic receptors are unopposed during a-receptor blockade. Therefore, the clinician should consider norepinephrine or dopamine. Epinephrine is contraindicated.

Because of their quinidine-like effects, overdoses with thioridazine and mesoridazine may lead to heart block and ventricular fibrillation. Since these agents may have long half-lives, patients recovering from substantial overdoses may require cardiac monitoring for several days.

Tolerance, Dependence, and Withdrawal Symptoms Some tolerance develops to the sedative, anticholinergic, hypotensive, and some other effects of dopamine receptor antagonists. Tolerance does not develop to the antipsychotic effects. Some patients experience a decrease in extrapyramidal adverse effects, suggesting tolerance to the motor effects of these agents.

Dopamine receptor antagonists are not addicting. As a result, most patients can be withdrawn from use of these agents without difficulty. Rapid withdrawal of low-potency dopamine receptor antagonists can evoke symptoms related to their anticholinergic effects, commonly manifestations of “cholinergic rebound” including nausea, vomiting, and diarrhea. Other reported withdrawal symptoms include headache, sweating, restlessness, and insomnia. These symptoms tend to occur within 2 to 3 days of drug withdrawal and may last as long as 2 weeks. Most patients who stop taking high-potency dopamine receptor antagonists do not show withdrawal symptoms. These observations suggest that withdrawal from low-potency agents should occur gradually over one to 2 weeks or longer.

DRUG INTERACTIONS

Clinicians should be aware of several types of interactions. Drugs may interact with the metabolism of dopamine receptor antagonists; age and gender may affect metabolic pathways; medical conditions may affect protein-binding, absorption, and fecal or renal clearance. Individuals who are dehydrated or malnourished or who have impaired liver and kidney function are at risk for altered sensitivity and metabolism.

Metabolism Dopamine receptor antagonists may have clinically important interactions with drugs that are metabolized by the same cytochrome P450 enzymes. For the conventional antipsychotics, particularly haloperidol, perphenazine, and thioridazine, this involves cytochrome P450 (CYP) isoenzyme 2D6 (CYP 2D6). Since some selective serotonin reuptake inhibitor (SSRIs)—particularly fluoxetine (Prozac) and paroxetine (Paxil)—are potent inhibitors of this enzyme, plasma concentration of the dopamine receptor antagonist is likely to increase when an SSRI is added, which may increase or worsen extrapyramidal adverse effects. Other drugs that are substrates for CYP 2D6—including tricyclic and tetracyclic drugs, b-adrenergic receptor antagonists, and cimetidine (Tagamet)—may also increase antipsychotic plasma concentrations.

Other drugs—particularly phenytoin (Dilantin), carbamazapine, and barbiturates—can reduce antipsychotic plasma concentrations by increasing the metabolism of the dopamine receptor antagonist. Reports indicate that carbamazapine can reduce the plasma concentration of haloperidol by 50 percent. Phenobarbital increases urinary excretion and thus decreases the plasma concentration of chlorpromazine.

Other common interactions should concern clinicians. Antacids, activated charcoal, cimetidine, cholestyramine (Questran), kaolin, and pectin can decrease the absorption of antipsychotics from the gut. Increased absorption may occur with drugs that decrease gastrointestinal motility such as digoxin (Lanoxin) and steroids.

Coffee and tea may precipitate dopamine receptor antagonists in vitro, but the clinical significance of this is unclear. Trifluoperazine and chlorpromazine have been reported to precipitate with lithium citrate.

Antipsychotic drugs antagonize the effects of dopamine agonists or levodopa when these drugs are used to treat parkinsonism. Chlorpromazine, haloperidol, and thiothixene can block the antihypertensive effects of guanethidine (Ismelin). Dopamine receptor antagonists may also enhance the effects of CNS depressants such as analgesics, anxiolytics, and hypnotics.

Dopamine receptor antagonists may enhance the effects of oral hypoglycemics. Phenothiazines enhance the effects of oral anticoagulants, although haloperidol may decrease these effects. Cigarette smoking may lower dopamine receptor antagonist blood concentrations by enzyme induction.

Adverse Effect Interactions Opioids, captopril (Capoten), trazodone (Desyrel), tricyclic and tetracyclic drugs, hydralazine (Apresoline), and minoxidil (Loniten) may increase hypotension. Combination of a low-potency dopamine receptor antagonist and another sedating drug may lead to excessive sedation. Combining low-potency dopamine receptor antagonist with other anticholinergic compounds can lead to peripheral anticholinergic symptoms such as dry mouth and urinary retention and central effects such as cognitive deficits and delirium (atropine psychosis). Patients should be warned about the use of over-the-counter medications that contain anticholinergics.

Amoxapine (Asendin), fluoxetine, chloroquine (Aralen), and lithium may increase extrapyramidal symptoms. Hydroxyzine (Vistaril) potentiates thioridazine and mesoridazine toxicity. Diuretics and corticosteroids enhance cardiotoxic adverse effects through hypokalemia, and quinidine enhances the quinidine-like effects of these agents.

LABORATORY INTERFERENCES

Chlorpromazine and perphenazine may give false-positive or false-negative results on pregnancy tests. These drugs may also give inaccurate reports of bilirubin (with reagent test strips) and urobilinogen (with Ehrlich's reagent test). Phenothiazines may produce false-positive results in tests for phenylketonuria. Phenothiazines may also interfere with 17-ketosteroid assays (with the Haltorff-Koch modification of the Zimmermann reaction) and 17-OH corticosteroid assays (with the modified Glenn-Nelson technique).

DOSAGE AND ADMINISTRATION

Dopamine receptor antagonists are available in a wide range of formulations and dose sizes (Table 31.17-7).

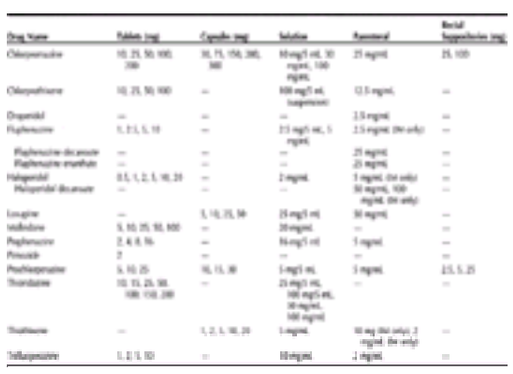

Table 31.17-7 Dopamine Receptor Antagonist Preparations

Treatment of Acute Psychosis

Initiation of Treatment Treatment with a dopamine receptor antagonist is usually initiated soon after a diagnosis is made. Whenever possible patients should receive a psychiatric evaluation including a history, a physical examination, and a laboratory evaluation that includes complete blood count (CBC); blood electrolytes and glucose determinations; liver, renal and thyroid function tests; and other appropriate evaluations (including ECG, pregnancy tests, and screens for alcohol and drugs).

The results of this evaluation can be useful in confirming the diagnosis and establishing pretreatment baselines. Usually it is not necessary to delay treatment while awaiting the results from these tests. Moreover, since there are few contraindications to dopamine receptor antagonists, these evaluations can be postponed when patients refuse an evaluation or when treatment is urgent.

Choice of Agent No evidence indicates that any dopamine receptor antagonist is more effective than any other. The choice among these agents is based on prior responses (both objective improvement and the patient's subjective response to the drug), available routes of administration and dosage forms, adverse-effect profiles, and cost.

The introduction of the SDAs (clozapine, risperidone, olanzapine, and sertindole at this time) has complicated the decision-making process because these agents have certain advantages over the dopamine receptor antagonists. The most important is that SDAs cause substantially less severe extrapyramidal adverse effects. Patients who experience these effects at dosages of conventional antipsychotics needed to treat their psychosis will probably not experience them with an effective dosage of an SDA. Clozapine is effective for patients who remain symptomatic when taking a dopamine receptor antagonist. Some evidence indicates that other SDAs are also effective for this treatment-refractory population. Substantial evidence indicates that SDAs are more effective than dopamine receptor antagonists for the treatment of negative symptoms. An additional advantage of some SDAs (including clozapine, olanzapine, and sertindole) is that they are less likely to increase prolactin concentration. As a result, these agents may be helpful for women who experience galactorrhea and irregular menses with dopamine receptor antagonist. All of these advantages of SDAs indicate that the dopamine receptor antagonists are likely to play a declining role in the treatment of psychosis. For example, a recent schizophrenia treatment algorithm from Texas recommends that patients receive trials with three newer antipsychotics, risperidone, olanzapine, and quetiapine (Seroquel) before being treated with a dopamine receptor antagonist. Treatment algorithms from the American Psychiatric Association and the Department of Veterans Affairs, on the other hand, recommend dopamine receptor antagonists as first-line antipsychotics, along with newer drugs. Populations that may continue to receive dopamine receptor antagonists include patients who require treatment with a long-acting depot drug and those who experience minimal adverse effects. Dopamine receptor antagonists are less expensive than SDAs; thus, certain payers for medical care may only permit changing to an SDA under certain circumstances.

High-potency dopamine receptor antagonists such as haloperidol and fluphenazine are more commonly prescribed than low-potency compounds. High-potency drugs are more strongly antidopaminergic and thus more likely to cause extrapyramidal adverse effects than the low-potency medications such as chlorpromazine or thioridazine. However, the extrapyramidal effects may be easier to manage than the sedation and orthostatic hypotension associated with low-potency medications. High-potency medications can be administered intramuscularly more readily because they seldom cause problems with blood pressure. In addition, because of sedation, orthostatic hypotension, and lethargy, the dosage of low-potency dopamine receptor antagonists should be gradually titrated upward; patients on high-potency medications can usually be treated with a therapeutic dosage within a day or two. Midpotency drugs such as loxapine and molindone are options for patients who have difficulty tolerating the adverse effects of high- or low-potency drugs.

Route of Administration Dopamine receptor antagonists can be administered orally or in short- or long-acting injectable forms (Table 31.17-7). Under most conditions, the oral forms of these agents are preferred by patients. Once patients are stabilized on a dosage, most patients can be managed with a single daily dose. Short-acting injectable forms are administered intramuscularly to calm agitated patients or for patients who refuse oral medication. The clinical effects of an intramuscular dose may appear as soon as 15 minutes after the injection. Since the dopamine receptor antagonists are most effective when patients have been in a steady state for several days or even weeks, patients should be given the drugs by the oral route as soon as possible.

Dopamine receptor antagonists can also be administered in long-acting injectable form. In the United States, only two antipsychotics, haloperidol and fluphenazine, are available in long-acting formulations. Long-acting antipsychotics are usually prescribed for maintenance treatment of stabilized patients. These drugs are usually poor choices for the treatment of acute psychosis because of their pharmacokinetics. Since it takes months to reach steady state or to eliminate drugs that are administered in this manner, the clinician cannot titrate dosage against clinical effectiveness. Moreover, if patients develop adverse effects from a depot drug, a period of time will separate the reduction of drug dosage and decreased the plasma concentration and adverse effects. Long-acting depot drugs are much better suited to long-term maintenance, after the effectiveness and sensitivity to adverse effects have been established during short-term treatment.

Dosage Patients demonstrate marked differences in the dosage of a dopamine receptor antagonist that is optimal (Table 31.17-8). For example, some patients find it difficult to tolerate the adverse effects from 1 mg of haloperidol whereas others tolerate 100 mg without suffering any adverse effects. Finding the best dosage for a dopamine receptor antagonist is both important and difficult. The importance stems from the tendency of these drugs to cause adverse effects at their effective doses. Often the clinician must weigh the therapeutic advantages of a particular dosage against discomforting or disabling adverse effects. Determining the best dosage is difficult because often one cannot titrate dosage against clinical response because there is a delay—often of several weeks—between a dosage change and the clinical response. Thus the most reasonable practice may be to administer a dosage that is likely to be effective and then wait an adequate amount of time (usually 4 to 6 weeks) to determine if the patient is responding.

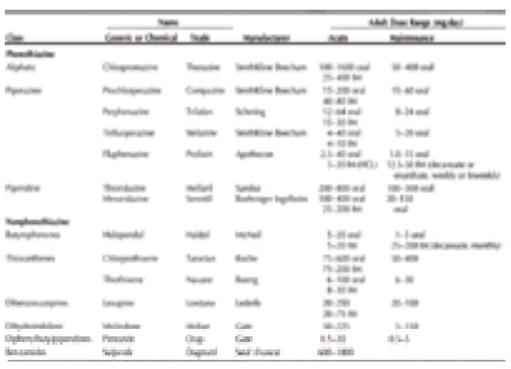

Table 31.17-8 Dopamine Receptor Antagnoists

Dosage comparison studies provide guidance to clinicians who are making decisions regarding the dosages of dopamine receptor antagonists for most patients with schizophrenia or schizoaffective disorder. Dosages below 300 mg a day of chlorpromazine (or 5 mg of fluphenazine or haloperidol) are likely to be too low for many psychotic patients. At the same time dosages above 1000 mg a day of chlorpromazine (or 20 mg of haloperidol or fluphenazine) are seldom necessary and may lead to substantial adverse effects. Most patients respond to daily dosages between 5 and 20 mg of fluphenazine or haloperidol or equivalent amounts of another dopamine receptor antagonist. Children and elderly patients should be given much lower dosages.

Some patients can tolerate very high dosages of dopamine receptor antagonists, particularly high-potency drugs. This observation led some clinicians to raise the prescribed dosages in hope that more is better, which resulted in a substantial increase in the average dose of antipsychotic prescribed in the United States during the 1970s and 1980s. However, dosage comparison studies failed to support the routine use of higher doses; when groups of patients were given daily dosages above 2000 mg of chlorpromazine or 40 mg of haloperidol, the rate of improvement and the amount of improvement was no greater than for those given more-moderate doses. Moreover, these higher dosages were associated with greater extrapyramidal adverse effects. Clinicians are sometimes impressed by individuals who require these higher dosages, suggesting that a small group of patients may benefit from high dosages.

As-Needed Medication Prescribing medication as needed is recommended only for the initial phases of treatment. Although dopamine receptor antagonists can result in nearly immediate calming of agitated individuals, this response appears to differ from the true antipsychotic effect of these medications, which has a delayed onset. Treating excitement and agitation with the addition of benzodiazepines such as lorazepam (Ativan) has the advantage of calming patients without causing extrapyramidal adverse effects.

Blood Concentrations and Clinical Activity Finding the best dosage of an antipsychotic for an individual patient often requires weeks or months of treatment. The clinician usually selects a drug dosage that is effective for most patients with the same illness or a dosage that was effective for the patient at an earlier time. Once the dosage is selected and administered, the clinician must wait days or weeks until it is clear whether it is effective, toxic, or ineffective. If clinicians could select a

drug dosage on the basis of a patient's plasma or blood concentration it would decrease the amount of time it takes to arrive at an appropriate dosage.

Early studies showed that plasma concentrations in patients who received the same dosage of a drug varied widely, which suggested that studying individual differences in plasma concentrations would provide information that could be useful to clinicians. However, early studies were limited to drugs such as chlorpromazine with a complex metabolism, so that some of the antipsychotic activity could result from metabolites rather than the parent drug. This and other problems (including flawed research methods) may explain why early studies failed to find a reliable relation.

A number of more recent studies have focused on other dopamine receptor antagonists and have shown more promising results. Haloperidol has received the most attention for the study of the relation between plasma concentration and clinical response. Haloperidol has a single important metabolite, reduced haloperidol, which does not have significant antipsychotic activity. Thus, the plasma concentration of haloperidol may accurately represent the effective amount of the drug. At least five research groups, but not all, have reported a relation between haloperidol concentration and clinical response. Table 31.17-9 gives reported plasma concentration ranges for four commonly prescribed dopamine receptor antagonists. Although all of these ranges are supported by one or more high-quality studies, the strongest evidence is for haloperidol.

Drug	Therapeutic Plasma Concentration (ng/mL)
Chlorpromazine	30–100*
Fluphenazine	0.2–2.0†
Haloperidol	2–15*
Perphenazine	0.8–2.4†

* Consider dosage reduction in patients with plasma concentration above upper limit.
† Range of concentration in which good response without debilitating adverse effects has been found; no evidence that reduction of plasma concentration above the upper limit improves the antipsychotic effect.
Adapted from Van Putten T, Marder SR, Wirshing WC, Aravagiri M, Chaber N: Neuroleptic plasma levels. Schizophr Bull 17:197, 1991.

Table 31.17-9 Reported Therapeutic Plasma Concentration Ranges for Antipsychotic Drugs

At this time, empirical studies do not support routine measurement of plasma concentration. However, the data suggest certain circumstances in which antipsychotic plasma concentrations can be helpful to clinicians. A concentration measurement may be helpful in the patient who is receiving an apparently adequate dosage but not responding. A low concentration may suggest that the patient is not complying or that poor absorption or extensive drug metabolism is interfering. In other cases, the clinician may find it difficult to interpret anxiety or agitation and may consider whether psychotic agitation could come from an inadequate dosage or akathisia result from an excessive dosage. A high or low plasma concentration could provide valuable information in these circumstances.

Plasma concentrations should be determined only after steady state has been reached, preferably during the late phase of the plasma concentration curve. This usually means after 5 days on a stable oral dosage and 10 to 12 hours after the last dose.

Adjunctive Medications Antipsychotic medications are often supplemented by other psychotropic medications to manage comorbid conditions that frequently occur with schizophrenia. Substantial evidence indicates that antidepressants are effective for treating depression in patients with schizophrenia and schizoaffective disorder that remains after psychosis is adequately treated. Mood-stabilizing drugs including lithium, carbamazepine, and valproate are effective for patients with unstable moods. Benzodiazepines are often effective for treating anxiety as well as agitation and excitement.

Maintenance Treatment

Duration of Treatment During the maintenance phase of treatment the treatment goals include preventing psychotic relapse and improving the patient's level of functioning and quality of life. Although dopamine receptor antagonists are effective in delaying or preventing relapse, they are also associated with short- and long-term adverse effects. The analysis of costs and benefits usually favors continuing to treat patients with schizophrenia or schizoaffective illness with an antipsychotic. Patients who have recovered from a single psychotic episode have a substantial risk of relapse if medication is discontinued. As a result, an international consensus conference recommended that first-episode patients receive 1 to 2 years of maintenance antipsychotic therapy. The same group recommended that patients who have had multiple episodes should continue to receive an antipsychotic for at least 2 to 5 years. Patients with a history of violent, aggressive behavior or serious suicide attempts should receive antipsychotics indefinitely. If a decision is made to discontinue medications, the dosage should be reduced gradually and the patient should be carefully observed.

Long-Acting Injectable Antipsychotics Substantial evidence indicates that long-acting injectable antipsychotics have advantages for patients during the maintenance stage of treatment, partly because these agents help solve the problem of poor medication compliance in schizophrenia. Clinicians can administer drugs without depending upon patients taking pills. In addition, because depot drugs bypass variations in drug absorption and first-pass hepatic and gut metabolism, drug concentrations may be more consistent. A number of studies have compared the relative effectiveness of long-acting and oral drugs for preventing relapse in stabilized patients. Studies that closely resemble treatment in the community have found that patients are better protected by depot medications. A number of more carefully controlled studies have compared the risk of psychotic relapse in patients randomly assigned to receive either oral or depot medication. The only comparison that lasted more than 1 year found an advantage for depot drugs. A meta-analysis of the studies comparing oral and depot agents found a significantly lower relapse rate in patients who received depot treatment ($P < .0002$).

Treatment Strategies During Maintenance Concerns about the adverse effects of antipsychotics and the risk for tardive dyskinesia have led to a number of proposed strategies for safely managing patients on the lowest possible dosage of a dopamine receptor antagonist. A number of studies have found that patients can be safely managed with substantially lower dosages of a long-acting depot antipsychotic than are commonly prescribed. These studies found that most patients did well when their dosages were 80 percent lower than those usually prescribed for short-term treatment. Patients treated with lower dosages had a small increase in the risk of relapse that was balanced in some studies by reductions in adverse effects, improved compliance, and suggestions of improved community adjustment.

Other studies have focused on a strategy named *targeted* or *intermittent therapy* in which antipsychotic dosage is gradually reduced to zero. Patients are monitored carefully for early, or prodromal, signs of relapse. Antipsychotic treatment is resumed when these signs appear. A number of studies, including a recent NIMH collaborative study, found that this strategy is associated with high relapse rates, indicating that it is probably not advisable for most patients. Another study found that early intervention was helpful for patients who were already receiving a low dose of a depot dopamine receptor antagonist. When these individuals demonstrate prodromal signs of relapse, their depot medication is supplemented with oral medication.

Management of Psychosis and Disruptive Behavior in Elderly Patients with Dementia Psychomotor agitation and aggressive behaviors are relatively common in patients with Alzheimer's Disease and other dementing illnesses. Dopamine receptor antagonists and SDAs are commonly prescribed for these conditions and have been found to be moderately effective when compared with placebos. Although thioridazine was frequently prescribed for agitated elderly patients in nursing homes, its adverse effects on blood pressure and memory have led many clinicians to prescribe high-potency agents, particularly haloperidol. A recent report from the New York State Psychitaric Institute found that treating agitated Alzheimer's Disease patients with 2 to 3 mg of haloperidol daily was more effective than .5 to .75 mg daily or placebo. The authors recommended starting these patients on 1 mg daily and gradually increasing the dosage as tolerated.

SUGGESTED CROSS-REFERENCES

Serotonin-dopamine antagonists are discussed in Section 31.26, and other biological therapies are discussed in other sections of Chapter 31. Psychiatric rating scales are discussed in Section 7.8. Cognitive disorders are discussed in Chapter 10, substance-related disorders in Chapter 11, schizophrenia in Chapter 12, other psychotic disorders in Chapter 13, and mood disorders in Chapter 14. Schizophrenia and delusional disorders in the elderly are discussed in Section 51.3f, and the use of antipsychotics in the elderly is discussed in Section 51.4e. Medication-induced movement disorders are covered in Section 31.4.

SECTION REFERENCES

*American Psychiatric Association: Practice guideline for the treatment of patients with schizophrenia. Am J Psychiatry *154*(Suppl):1, 1997.

Baldessarini RJ, Cohen BM, Teicher MH: Significance of neuroleptic dose and plasma level in the pharmacological treatment of psychoses. Arch Gen Psychiatry *45:*79, 1988.

*Chiles JA, Miller AL, Crismon ML, Rush AJ, Krasnoff AS, Shon SS: The Texas Medication Algorithm Project: Development and implementation of the schizophrenia algorithm. Psychiatr Serv *50:*69, 1999.

Chiodo LA, Bunney BS: Population response of midbrain dopaminergic neurons to neuroleptics: Further studies on time course and nondopaminergic neuronal influences. J Neurosci *7:*629, 1987.

Coryell W, Miller DD, Perry PJ: Haloperidol plasma levels and dose optimization Am J Psychiatry *155:*48, 1998.

Creese I, Burt DR, Snyder SH: Dopamine receptor binding predicts clinical and pharmacologic potencies of antischizophrenic drugs. Science *192:*481, 1976.

De Leon J, Simpson GM: Assessment of neuroleptic-induced extrapyramidal symptoms. In *Adverse Effects of Psychotropic Drugs*, JM Kane, JA Lieberman, editors. Guilford, New York, 1992.

Devanand DP, Marder K, Michaels KS, Sackheim HA, Bell K, Sullivan MA, Cooper TB, Pelton GH, Mayeux R: A randomized, placebo-controlled dose-comparison trial of haloperidol for psychosis and disruptive behaviors in Alzheimer's Disease. Am J Psychiatry *155:*1512, 1998.

*Dixon LB, Lehman AF, Levine J: Conventional antipsychotic medications for schizophrenia. Schizophr Bull *21:*567, 1995.

Farde L, Nordstrom AL, Wiesel FA, Pauli S, Halldin C, Sekaall G: Positron emission tomographic analysis of central D and D_1 and D_2 dopamine receptor occupancy in patients treated with classical neuroleptics and clozapine. Arch Gen Psychiatry *49:*534, 1992.

Gilbert PL, Harris MJ, McAdams LA, Jeste DU: Neuroleptic withdrawal in schizophrenic patients. Arch Gen Psychiatry *52:*173, 1995.

*Janicak PG, Davis JM, Preskorn SH, Ayd FJ: *Principles and Practice of Psychopharmacology*. Williams & Wilkins, Baltimore, 1993.

*Jeste DV, Lacro JP, Palmer B, Rockwell E, Harris MJ, Caligiuri MP: Incidence of tardive dyskinesia in early stages of low-dose treatment with typical neuroleptics in older patients. Am J Psychiatry *156:*309, 1999.

Johnson DAW, Pasterski JM, Ludlow JM, Street K, Taylor RDW: The discontinuance of maintenance neuroleptic therapy in chronic schizophrenic patients: Drug and social consequences. Acta Psychiatr Scand, *67:*339, 1983.

Kahn RS, Davis KL: New developments in dopamine and schizophrenia. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Kane JM, Lieberman J: Tardive dyskinesia. In *Adverse Effects of Psychotropic Drugs*, JM Kane, JA Lieberman, editors. Guilford, New York, 1992.

Kissling W, editor: *Guidelines for Neuroleptic Relapse Prevention in Schizophrenia*. Springer-Verlag, Berlin, 1991.

Marder SR, Hubbard JW, Van Putten T, Midha KK: The pharmacokinetics of long-acting injectable neuroleptic drugs: Clinical implications. Psychopharmacology *98:*433, 1989.

Sachdev P, Kruk J: Clinical characteristics and predisposing factor in acute drug-induced akathisia. Arch Gen Psychiatry *51:*963, 1994.

Seeman P, Lee T, Chau-Wong M, Wong K: Antipsychotic drug doses and neuroleptic/dopamine receptors. Nature *261:*717, 1976.

*Woerner MG, Alvir JM, Saltz BL, Lieberman JA, Kane JM: Prospective study of tardive dyskinesia in the elderly: Rates and risk factors. Am J Psychiatry *155:*1521, 1998.

Wyatt RJ: Neuroleptics and the natural course of schizophrenia. Schizophr Bull *17:*325, 1991.

31.18 LITHIUM

JAMES W. JEFFERSON, M.D., AND JOHN H. GREIST, M.D.

HISTORY

A report appeared in 1800 describing the discovery of two new minerals, petalite and spodumene, on an island near Stockholm, Sweden. When the former was analyzed, a small percent remained unidentified until 1817 when Johan Arfwedson discovered a new alkali that was named lithion by Jons Jacob Berzelius, his laboratory chief. The next year, Humphrey Davy was the first to isolate lithium metal.

In 1843 Alexander Ure introduced lithium into medicine when he showed in vitro that a uric acid bladder stone lost weight in a lithium carbonate solution. By the early 1860s Sir Alfred Garrod had discovered that gouty uric acid deposits in finger joints were also soluble in vitro in a lithium solution. In the second half of the nineteenth century the notion of uric acid imbalances as the cause of disease expanded far beyond bladder stones and gout to include many ills. A. Trousseau in France and Alexander Haig in England incorporated mania and depression into the uric acid diathesis. In 1886, Karl Lange, a Dane, reported that a mixture containing lithium carbonate was effective in the prophylactic treatment of depression, and in 1894 his brother, Fritz Lange, described the effectiveness of lithium carbonate for treating acute depression. In the United States, William A. Hammond (a former Surgeon General of the Army) wrote in 1871: “Latterly I have used the bromide of lithium in cases of acute mania, and have more reason to be satisfied with it than any other medicine calculated to diminish the amount of blood in the cerebral vessels, and to calm any nervous excitement that may be present.” For mania, Hammond preferred lithium bromide to other bromide preparations and suggested rather generous amounts: “The doses should be large—as high as 60 grains or even more—and should be repeated every two or three hours till sleep be produced, or at least till half a dozen doses be taken.” Since this represents a more than generous amount of both lithium and bromide, Hammond's patients may well have been toxic rather than tranquilized.

From the late 1880s through the early 1900s, lithium was embraced by the general public in the form of mineral spring waters. Taking the waters was reputed to cure just about everything, as evidenced by an advertisement for Bear Lithia Water: “Cures kidney and bladder troubles, uric acid, gout, and rheumatism, phosphoric deposits, inflammation of the bladder, dropsical affections, brick-dust deposits, and all forms of dyspepsia. . .” (Fig. 31.18-1). A testimonial to Buffalo Lithia Water stated: “The water from spring No. 1 is the most powerful restorative of the broken-down enfeebled human system that I have known.” In Wisconsin lithia spring water was used to brew Lithia Beer, a regional favorite for decades. As late as 1920 the White Rock Mineral Spring Company was bottling 120,000 bottles of lithia water daily. Nonetheless, the lithia water bubble had burst many years earlier. In 1896 the U.S. Bureau of Chemistry found that nearly all lithia waters on the United States market contained at most only spectroscopic traces of lithium, and in the early 1900s the Supreme Court of the District of Columbia concluded: “For a person to obtain a therapeutic dose of lithium by drinking Buffalo Lithia Water he would have to drink from one hundred and fifty thousand to two hundred and twenty-five thousand gallons of water per day.” It was difficult not to conclude that the only condition that lithia water cured was thirst.

FIGURE 31.18-1 Advertisement for Bear Lithia Water.

Nevertheless, lithium remained in the public eye. Lithium tablets became available, and by 1907 the *Merck Index* listed 43 medicinal preparations containing lithium. With the tablets came an awareness of the toxic potential of lithium—generalized weakness and tremor in patients was reported in 1898 and diarrhea, vomiting, and death were described in animals in 1903. These early lessons were apparently forgotten, for when lithium reemerged into popularity in 1948 in the form of a 25 percent solution of lithium chloride for use as a salt substitute in patients on low-sodium diets, the dangers were not appreciated. By 1949, however, reports of severe lithium intoxication and death resulted in the removal of lithium products from the market and delayed the acceptance of lithium by American psychiatry for many years.

At almost the same time that lithium was poisoning cardiac patients in the United States, the Australian psychiatrist, John F.J. Cade was observing its antimanic effect in psychiatric patients. To explore the hypothesis that mania might be “a state of intoxication by a normal product of the body circulating in excess, whilst melancholia is the corresponding deprative condition,” Cade injected urine from patients into guinea pigs. He noted that lithium urate modified the toxicity of urea in a favorable direction which, he stated, “argues for a strong protective function for the lithium ion against the convulsant mode of death produced by toxic doses of urea.” Based on the use of lithium salts in the nineteenth century and particularly because “single and repeated doses of lithium citrate and lithium carbonate in the doses contemplated produced no discernible ill effects on the investigator himself,” Cade concluded that there were “no ethical contraindications to using them in mania.” In 1949 he described strikingly favorable results in 10 patients with mania including the first who had been hospitalized for 5 years in a state of manic excitement—”He had enjoyed preeminent nuisance value in a back ward all those years and bid fair to remain there for the rest of his life.” Instead, the illness went into full remission, the patient was discharged and returned to work. The story does not end here, however, for euthymia led to noncompliance which in turn reignited the mania with resultant hospitalization. Lithium was reinstituted successfully, but eventually (22 months after first taking lithium) he died from“[-T]oxaemia due to lithium salts therapeutically administered.” Cade's success with lithium led him to investigations of other ions such as rubidium, cesium, lanthanum, neodymium, and strontium, but the remarkable effects of lithium were never duplicated.

Later, the clinical observations of Cade were substantiated in more rigorous clinical trials conducted by Mogens Schou and others who firmly established the effectiveness of lithium for mania and for the prophylactic treatment of manic-depressive disorder. Lithium use in the United States increased gradually in the late 1960s, but not until 1970 did the Food and Drug Administration (FDA) approve its labeling for the treatment of mania. The only other approved indication was in 1974 for maintenance therapy in patients with a history of mania.

In 1975 the Lithium Information Center was established at the University of Wisconsin in Madison to collect, catagorize, and disseminate information on lithium and its roles in medicine. The Center, now at the Madison Institute of Medicine in Madison, Wisconsin, contains an ever-increasing database of over 28,400 articles.

CHEMISTRY

Lithium follows hydrogen and helium on the periodic table, making it the third simplest element (atomic number 3, atomic weight 6.94) and the first solid element. It is a member of the group IA alkaline metals together with sodium, potassium, rubidium, cesium, and francium. Among them, lithium has the smallest atomic radius and the highest melting point, boiling point, and ionic charge density. In nature lithium exists as two quadrupolar isotopes, Li^{6} (7.42 percent) and Li^{7} (92.58 percent). Most of the lithium used in the United States is obtained from spodumene ore in North Carolina with lesser amounts from brines in Nevada. Its applications extend far beyond psychiatry, including the manufacture of aluminum, ceramics, and synthetic rubber; uses in batteries, greases, air conditioning, industrial drying, welding, brazing, and as a carbon dioxide absorbent; and as a vehicle for delivering chlorine to swimming pools (lithium hypochlorite). Trace amounts of lithium are present in food and water and hence in all living tissue. Whether lithium is essential to normal human growth and development has not been established, and there is no evidence of a lithium deficiency state in humans. Experimentally produced lithium deficiency in goats and rats led to reduced fertility, retarded development, and reduced longevity.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics Lithium is rapidly and completely absorbed, with serum concentrations peaking in 1 to 1½ hours with standard preparations and in 4 to 4½ hours with the slow and controlled-release forms. Unlike most psychiatric drugs, lithium has no clinically important protein-binding properties and no metabolites. It is excreted almost entirely by the kidneys, although small amounts are also lost in sweat and feces. A substantial amount of filtered lithium is reabsorbed (primarily in the proximal tubules), so that renal lithium clearance is about one-fifth of creatinine clearance. The elimination half-life of lithium is about 18 to 24 hours, although it is considerably longer in the elderly because of the age-related decrease in glomerular filtration rate (GFR). Consequently, elderly patients tend to need lower than usual doses to reach a given serum concentration and require longer than usual to reach steady-state. Duration of treatment may also affect half-life as evidenced by a small study that found it to be 1.3 days in those just starting treatment but 2.4 days in those treated for over a year. Obesity may alter lithium kinetics based on the observation that clearance was almost 50 percent higher in obese subjects than in normal-weight volunteers. Lithium is distributed widely throughout the body, although the rate and extent of entry into tissues varies. For example, thyroid and renal concentrations exceed serum level, whereas red blood cell, spinal fluid, and brain concentrations do not. Lithium enters and leaves the central nervous system (CNS) slowly, which may explain why acute overdoses with relatively high blood levels are sometimes well tolerated and why clinical manifestations of chronic intoxications often persist long after blood levels have decreased.

Pharmacodynamics “Because the key is so small, it fits into many locks.” Thus did Mogens Schou describe the difficulty defining the mechanism of action of lithium. There is no evidence for lithium deficiency being the cause of bipolar disorders. Indeed, the amount of lithium required for a clinical effect dwarfs the naturally occurring levels in the body by a factor of about 100. A monovalent cation, lithium has chemical similarities to sodium and potassium as well as to the divalent cations, calcium and magnesium. Consequently, ion substitution has been evoked to explain its clinical effects with considerable attention focused on ion pumps and channels, particularly sodium-lithium counterexchange. Sodium, potassium-adenosine triphosphatase (Na, K-ATPase) has been implicated in the cause and treatment of bipolar disorder. Altered levels of Na, K-ATPase activity have been found in subjects with manic and depressed bipolar disorder compared to euthymic subjects with bipolar disorders in some but not all studies. Since neuronal transmembrane potential differences are maintained by the Na, K-ATPase pump (also known as the sodium pump), perturbations of this system are felt to lead to neurotransmitter aberrations that translate into the clinical manifestations of bipolar disorders. Lithium crosses cell membranes by various mechanisms including the sodium pump, the sodium leak canal, a lithium-bicarbonate exchange, and the sodium-lithium counterexchange system. By substituting for sodium in these transmembranal ion exchanges, lithium is believed to exert its therapeutic effect. Although this theory has been examined and refined since the late 1970s, it continues to be challenged by experimental inconsistencies.

No neurotransmitter has been shown to be unaffected by lithium. As a result, serotonin, norepinephrine, dopamine, acetylcholine, g-aminobutyric acid (GABA), and a variety of neuropeptides have been implicated both singularly and in combination to explain lithium's diverse effects.

Other theories involve effects on neuroendocrine systems, kindling and behavioral sensitization, biological rhythms, viruses, and a variety of signal transduction pathways. Both the adenylate cyclase and phosphoinosotide second-messenger systems are altered profoundly by lithium in a variety of ways, including effects on guanine nucleotide-binding proteins (G proteins) and protein kinase C. The complexity of the situation was summarized by Husseini Manji and Robert Lenox:

> Chronic lithium exerts significant transcriptional and posttranscriptional effects, and that those actions of lithium may be mediated via protein kinase C (PKC)-induced alterations in nuclear transcription regulatory factors responsible for modulating the expression of proteins involved in the long-term neural plasticity and cellular response.

The observation that lithium inhibited inositol monophosphatase, reduced brain inositol levels, and inhibited inositol 1,4,5-trisphosphate (IP_3) accumulation in rat and mouse cerebral cortex led to the inositol depletion hypothesis of its mechanism of action. Further studies, however, using guinea pig, rabbit, and monkey cerebral cortex found that lithium actually increased IP_3 concentrations (the same was seen in rat and mouse if supplemental inositol was provided). Subsequently it was shown that the accumulation of IP_3 was promoted by lithium-induced glutamate release activating the *N*-methyl-D-aspartate (NMDA) receptor. If glutamate release was inhibited by carbetapentane or if the ionotropic NMDA receptor was blocked by selective antagonists, lithium failed to stimulate IP_3 accumulation. Interestingly, in the same model valproate also stimulated glutamate release and increased IP_3 concentrations, although the mechanism appeared to be different from that of lithium. At concentrations that were maximally effective when used alone, the combination of the two drugs had additive effects on glutamate release (an observation that may explain the clinically observed benefits of combining these drugs in treatment-resistant patients.)

G proteins have also been targeted as sites of lithium activity. Studies of long-term lithium administration have found altered G protein function (such as stabilization of subunit dissociation) suggesting that transmembranal signal transduction is affected. Because G proteins are involved in conveying information from numerous neurotransmitters, neuromodulators, and hormones to intracellular effectors, lithium's effects at this level can cascade widely and deeply within cells.

In addition to lithium altering G-protein function, it has also been shown that lithium decreases concentrations of some of the PKC isozymes and of myrisoylated alanine-rich C kinase substrate (MARCKS), a protein involved in neurotransmission; similar effects were also seen with valproate. Furthermore, therapeutically equivalent concentrations of lithium in cultured rat neurons increased transcriptional factor binding to both activated protein-1 (AP-1) and adenosine monophosphate cAMP-responsive element (CRE). It should be obvious that signal transduction pathways provide numerous interrelated targets for lithium's actions. Which of these, if any, would explain its mechanism of action in bipolar disorder remains to be determined.

In keeping with the “one key fits many locks” observation, lithium has also been shown to inhibit glycogen synthase kinase-3 (GSK-3), a protein kinase involved neuronal in cytoskeletal development. Among its many activities, GSK-3 regulates the phosphorylation of tau, the microtubule-associated proteins involved in the paired helical filaments of Alzheimer's disease. It has been speculated that lithium may even play a role in the treatment of Alzheimer's disease.

The bipolarity virus theory arose from the observation that lithium is an effective treatment for two episodic illnesses; namely, genital herpes simplex and bipolar disorder. Its inhibitory effect on herpes simplex viral replication is well established, but whether it also inhibits an unidentified virus that causes bipolar disorders is a matter of speculation. Investigators have yet to explain how this simple ion with its ubiquitous neurobiological effects manages to resolve mania and depressions and provide long-term mood stabilization.

Design and Interpretation of Clinical Drug Studies The history of lithium encompasses most of the history of clinical drug trials in psychiatry. The research that established its antimanic activity to the satisfaction of the FDA was meager compared to today's large, double-blind, placebo-controlled parallel design, multi-center trials. Although the response rate consistently favored lithium over placebo, the studies were small, short-term, and of crossover design. Nonetheless, the recent large (*N* = 179) multicenter, randomized, double-blind, parallel design, 3-week comparison of lithium, divalproex (Depakote), and placebo for the treatment of acute mania found lithium to be twice as effective as placebo (49 percent versus 25 percent) and the equal of divalproex (49 percent versus 48 percent). Since the study design excluded patients previously treated with divalproex or valproic acid but included numerous patients in whom lithium had been previously ineffective or not well tolerated, to conclude that the two drugs have equal efficacy for the treatment of mania may be premature.

In the 1960s open studies by experienced clinicians comparing episode frequency before and during lithium treatment did much to establish the long-term effectiveness of the drug. Citing lack of scientific rigor in the design of these studies, critics cast doubt on the conclusion and stimulated a plethora of double-blind, placebo-controlled studies (both prospective and discontinuation) that convincingly established the long-term prophylactic efficacy of lithium in bipolar disorder. Currently in the United States divalproex appears to be evolving over a similar course with an initial ground swell of clinical enthusiasm now being subjected to the

rigors of more scientific scrutiny.

THERAPEUTIC INDICATIONS

Bipolar Disorders

Mania Lithium is accepted universally as an antimanic drug with effectiveness greater than placebo. Its onset of action is relatively slow, with clinical improvement usually occurring over the first 1 to 3 weeks of treatment. Although one small open study suggested that oral lithium loading might produce improvement in a few days, this approach has yet to be substantiated. Instead, the concomitant use of a benzodiazepine and an antipsychotic drug is often necessary during the initial phase of treatment. Since disturbed sleep may ignite or fuel a manic episode, early treatment of insomnia with a sedative such as clonazepam (Klonopin) or lorazepam (Ativan) may be useful.

Not all patients with acute mania respond equally well to lithium. The response rate of 78 percent in the initial research trials contrasts with findings of less success when dealing with patients with mixed or dysphoric mania, rapid cycling, comorbid substance abuse, or organicity. For mania patients unresponsive to lithium or intolerant of it, alternative treatments include valproate, carbamazepine (Tegretol), antipsychotics drugs, electroconvulsive therapy (ECT), and a variety of less well established pharmacological agents. Whether another mood-stabilizer would be preferable to lithium as the initial treatment of mania, or at least for some subtypes of mania, has not been established in well-designed studies. The effectiveness of lithium maintenance therapy is far better established than for any other drug, which may weigh in favor of using lithium to treat the initial manic episode.

Depression Lithium is not approved by the FDA as an antidepressant medication, nonetheless, research supports its effectiveness, at least for depression in bipolar disorders. In placebo-controlled studies 79 percent of depressed bipolar disorder patients and 36 percent of major depressive disorder patients had a partial or complete response to lithium. Four of five double-blind comparisons found lithium just as effective as tricyclic drugs (the data did not distinguish between bipolar and unipolar depression). Using a conventional antidepressant as the sole treatment for bipolar disorder depression appears to increase the risk of drug-induced mania and rapid cycling, a risk established most clearly with tricyclic drugs. Consequently, initial treatment with lithium alone is often preferred for bipolar depression, although for more severe depressions, some clinicians prefer to combine lithium with an antidepressant drug. Therapeutic blood concentrations have not been well established for the antidepressant effect of lithium, but appear to be in the same general range as for mania. If an antidepressant is combined with lithium to treat depression, the antidepressant should be withdrawn gradually soon after remission to minimize the risk of antidepressant-induced mania or rapid cycling. In practice, however, patients often seem to require and tolerate the long-term use of both lithium and an antidepressant. The role of carbamazepine and valproate as monotherapies for bipolar depression has not been adequately studied.

If breakthrough depression occurs during lithium maintenance, lithium-induced hypothyroidism, noncompliance, and substance abuse must be considered as possible causes. Often, however, no apparent cause can be found and it must be assumed that the pressure of the illness overcame the protective effect of lithium. With that in mind, increasing the serum lithium level to 1 or 1.2 mEq/L may be all that is necessary to terminate the episode. Even if the patient's thyroid function is normal, supplementation with thyroid hormone in the form of 25 µg a day of liothyronine (Cytomel) may be beneficial. Milder episodes of depression may resolve spontaneously or with brief psychotherapy. For moderate to severe depression, it may be necessary to add an antidepressant drug. Although tricyclic medications have been studied most extensively and are more effective than placebo, evidence is growing to support the preferential use of bupropion (Wellbutrin) or a selective serotonin reuptake inhibitor (SSRI) in part because antidepressant-associated mania may be less frequent, milder, less dysphoric, and easier to manage. In a comparative study the monoamine oxidase inhibitor tranylcypromine (Parnate) was more effective and better tolerated than imipramine (Tofranil) in bipolar depressed outpatients who were not receiving a mood stabilizer. The role of other antidepressants such as venlafaxine (Effexor), nefazodone (Serzone), and mirtazapine (Remeron) remains to be established. Augmenting lithium with either valproate or carbamazepine may have a beneficial antidepressant effect without increasing the risk of antidepressant-induced mania or rapid cycling.

ECT is an effective treatment for bipolar depression, although it tends to be used only after medications have failed. Concern has been expressed that the presence of lithium during ECT increases the risk and severity of cognitive dysfunction, but whether this is true remains controversial.

Maintenance Therapy Long-term treatment with lithium is an effective way to reduce the frequency, severity, and duration of manic and depressive episodes in patients with bipolar disorder. In 10 double-blind studies the relapse rate when patients were receiving lithium was 34 percent versus 81 percent on placebo. Lithium was somewhat more effective against mania than against depression. Treatment response to lithium monotherapy is usually less than complete and supplemental pharmacotherapies are often necessary either intermittently or continuously. Nonetheless, there is growing evidence for a major benefit from long-term lithium therapy: namely, reduced mortality. In general, bipolar disorder carries with it an increased risk of premature death, particularly but not entirely from suicide. Studies comparing mortality rates in lithium-treated patients to the general population, to patients following discontinuation of lithium, and to patients treated with alternative prophylactic medications all support the mortality-lowering and suicide-reducing effect of lithium.

Work by Paul Grof and others to predict a favorable response to lithium maintenance found that the diagnosis of a mood disorder, an episodic course with euthymic intervals, and the absence of rapid cycling were the main predictors. The following factors have been associated with a less than ideal response to lithium maintenance: mixed or dysphoric mania, rapid cycling (four or more episodes per year), many prior episodes, poor interepisode functioning, an episode pattern of depression-mania-euthymia, comorbid substance abuse, and comorbid personality disorder. Of great interest and as yet unresolved by properly designed studies is whether alternatives to lithium such as valproate and carbamazepine will yield better results under similar circumstances.

Just exactly when to begin long-term lithium treatment is not fully resolved, although most recommendations suggest earlier rather than later—possibly after only one episode of mania and certainly after two. There is growing concern that episodes beget episodes (perhaps a form of kindling or behavioral sensitization), and that each episode increases the likelihood of the next and may shorten the interval to the next, hence the recent trend toward earlier initiation of maintenance therapy. Obviously, many individual factors must be taken into consideration, including the abruptness, severity, and frequency of episodes; the presence of risk factors; and attitudes toward medication. Early initiation of maintenance therapy may benefit many patients at the expense of some patients being treated unnecessarily.

> A patient had two manic episodes in two years, one resolving spontaneously and one with ECT, followed by 18 euthymic years without any treatment before the next episode occurred. Had she been treated with lithium after the second manic episode, the subsequent 18 years would have been considered a testimonial to the therapeutic effectiveness of lithium.

The maximum benefits of lithium maintenance may not be immediate; with continued treatment relapses sometimes become less severe and less frequent. Some patients appear to develop a tolerance to lithium after several years of successful use; this tolerance can sometimes be overcome by the addition of carbamazepine or valproate. Similarly, patients who are only partial responders to lithium maintenance may benefit from the addition of one of these drugs. Just as breakthrough depression may respond to a temporary increase in serum lithium concentration so may breakthrough mania. At such times, however, supplementing lithium with an antipsychotic drug or a benzodiazepine may be necessary.

Maintenance lithium concentrations are usually lower than those used for the treatment of acute mania. Package inserts for lithium products refer to 0.6 to 1.2 mEq/L as desirable for long-term control; in general, higher serum lithium concentrations within this range are more likely to be effective but they cause more adverse effects and noncompliance.

Long-term treatment does not cure bipolar disorders. Following discontinuation of successful lithium therapy, the risk of recurrence increases substantially. For example, a review of discontinuation studies of patients with bipolar I disorder found the monthly risk of recurrence to be 28 times higher after lithium was stopped. In addition, rapid discontinuation of lithium appears to be associated with a rebound phenomenon characterized by increased risk of early recurrence of manic and depressive episodes. Gradual withdrawal of lithium seems to delay but not prevent recurrent mania and it may both delay and reduce the risk of recurrent depression.

Another phenomenon of concern has been termed lithium discontinuation-induced refractoriness. Some patients, approaching 15 percent, who had responded well to lithium prophylaxis may not respond again when lithium is reintroduced after a failed discontinuation trial. A decision to discontinue successful lithium maintenance should not be taken lightly, but rather must be weighed carefully against the continued risk of adverse effects and toxicity.

Major Depressive Disorder Although major depressive episodes may respond to lithium alone, traditional antidepressant medications are considerably more effective and are the preferred treatment. The major value of lithium in patients with major depressive disorder is as an augmenting agent when antidepressants alone have been ineffective. On average, 50 to 60 percent of patients respond when lithium carbonate, usually 300 mg three times daily (less in the elderly), is added to a

wide variety of antidepressant drugs. Occasionally, a dramatic response is noted in 24 to 48 hours but improvement is more likely to occur gradually over several weeks. Favorable responses have been described in both psychotic and nonpsychotic depression. While alternative explanations may apply in some situations (spontaneous remission, the longer duration of antidepressant treatment, or an independent effect of lithium), research support for a true augmenting interaction is quite persuasive. Experts disagree, however, as to when lithium augmentation should be introduced during the course of dealing with treatment resistance.

Mogens Schou states: “Doubt about the efficacy of prophylactic lithium treatment in unipolar illness seems almost exclusively a U.S. phenomenon.” Indeed, most studies have found lithium to be more effective than placebo and as effective as antidepressant drugs for the long-term management of major depressive disorder. Lithium has not, however, demonstrated clear advantages over conventional antidepressant drugs, and since long-term treatment is most likely to involve the drug that ended the acute episode (an antidepressant), lithium maintenance for major depressive disorder is likely to be reserved for treatment failures.

Schizoaffective Disorder and Schizophrenia In general, the less affective and the more schizophrenic an illness is, the less likely it is to respond to lithium. The same cannot be said of an episode because the acute manifestations of mania and schizophrenia may be indistinguishable. Although not approved by the FDA for schizoaffective disorder and not specifically studied for the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) version of this disorder, lithium is generally accepted to be of value for treating this condition, especially in combination with antipsychotic drugs.

Antipsychotic drugs remain the treatment of choice for schizophrenia, although patient response is often less than ideal. The addition of lithium to an antipsychotic agent is one of several reasonable treatment strategies for antipsychotic-resistant schizophrenia, although the availability of serotonin–dopamine antagonists (atypical antipsychotics), such as clozapine (Clozaril), risperidone (Risperdal), olanzapine (Zyprexa), and others, make this less necessary. Improvement of schizophrenia with lithium augmentation of an antipsychotic medication does not depend on the presence of symptoms of a mood disorder.

Aggression Lithium appears to have anti-aggressive effects independent of its mood-stabilizing action. Several controlled studies have found reduced aggression in subjects recruited from prison populations. Behavioral dyscontrol and self-mutilation in mentally retarded patients have also been successfully treated with lithium. However, lithium has not been used extensively to treat aggression associated with head trauma or epilepsy, and the results have been mixed. No work has been done to evaluate the effect of lithium in those patients defined by DSM-IV as having intermittent explosive disorder. Aggressive outbursts are also treated with medications such as anticonvulsants, b-adrenergic receptor antagonists, benzodiazepines, buspirone (BuSpar), antipsychotic drugs, and trazodone (Desyrel). The relative ranking of lithium among those other drugs must be determined on an individual basis.

Alcohol Use Disorders Lithium's potential for treating alcohol dependence and abuse is based in part on the close association between mood disorders and alcohol use disorders and in part on animal research that found reduced alcohol intake in rodents receiving lithium. Efforts to establish lithium as a useful treatment for alcoholism have been largely unsuccessful. A large (*N* = 457), multicenter, double-blind, placebo-controlled 1-year study found that lithium did not alter the course of alcoholism in depressed or nondepressed alcoholics. Although not all research reports have been that discouraging, lithium is unlikely to play a major role in the treatment of alcoholism. Some patients may be exceptions, but clinicians have found no way to identify them prospectively.

Other Indications The philosophy that “lithium is good for what ails you” led to the treatment of a wide variety of psychiatric (Table 31.18-1) and nonpsychiatric medical conditions (Table 31.18-2) with the drug. In general, reports in these areas involve small numbers of patients in studies often lacking in scientific rigor. In such situations both positive and negative outcomes may be inconclusive and even misleading. For example, two double-blind, placebo-controlled trials found lithium ineffective in obsessive-compulsive disorder. Only 16 patients were involved, however, and the treatment duration of 2 weeks was insufficient to evaluate the drug adequately. Two recent well-designed, placebo-controlled studies did find lithium to be ineffective in augmenting conventional drugs for obsessive-compulsive disorder in treatment-resistant patients.

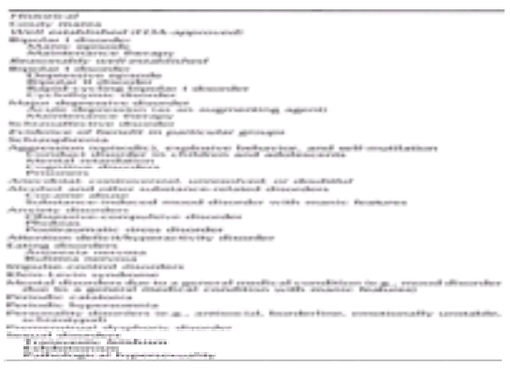

Table 31.18-1 Psychiatric Uses of Lithium

Table 31.18-2 Nonpsychiatric Uses of Lithium*

Efforts to establish lithium as a useful treatment for personality disorders have been largely unsuccessful. When favorable outcomes do occur, it is likely that a comorbid mood disorder has responded followed by indirect improvement in “personality.” In days of the first edition of DSM (DSM-I), lithium was used successfully to treat inpatients with emotionally unstable character disorder. The findings were never replicated or extended to outpatients, and the disorder failed to appear in subsequent editions of DSM. Today, the substantial mood component of borderline personality disorder suggests a role for lithium in its treatment, but controlled trials have not been conducted.

There has been considerable interest in lithium's potential for treating a wide variety of nonpsychiatric medical conditions (Table 31.18-2). The granulocyte-stimulating effect of the drug has been used to increase white blood cell count in conditions such as Felty's syndrome, idiopathic neutropenia, chemotherapy-induced granulocytopenia, and immunosuppressant drug-induced leukopenias. Lithium's antithyroid effects have been utilized investigationally to treat hyperthyroidism. Neurologists have found lithium useful as a treatment for chronic and episodic cluster headaches. The observation that lithium inhibits replication of deoxyribonucleic acid (DNA) viruses such as human herpesvirus 1 and 2 led to successful double-blind, placebo-controlled trials of both topical and oral preparations to treat herpes genitalis. Seborrheic dermatitis has also been shown in controlled studies to respond to a lithium-containing ointment that is marketed outside the United States. The therapeutic promise of lithium extends beyond medicine as evidenced by its fungicidal effects on powdery mildew, a serious affliction of cucumbers.

PRECAUTIONS

Children and Adolescents At least 20 to 30 percent of patients with bipolar I disorder have their first episode before the age of 20 years. This percentage may be even higher if age of onset is dated from first appearance of symptoms rather than from first well-defined episode or first hospitalization. Although adolescent onset is

common, prepubertal onset is far less so, but its prevalence may be underestimated because of atypical presentations and reluctance to make the diagnosis in children.

In general, diagnostic criteria for bipolar disorder in the young are the same as in adults; however, atypical presentations and comorbidity, especially with attention-deficit/hyperactivity disorder, may make recognition more difficult. Lack of a well-established longitudinal history further complicates diagnosis and may contribute to miscategorizing first-episode psychotic mania as schizophrenia.

Another diagnostic dilemma is whether a first episode depression will follow a bipolar or unipolar course. The following factors may be predictive of a bipolar outcome and influence early treatment with lithium or another mood stabilizer: (1) rapid onset with psychosis and psychomotor retardation, (2) family history of bipolar disorder, and (3) antidepressant-induced hypomania.

Despite a substantial literature on the use of lithium to treat children and adolescents with conditions such as bipolar disorder, aggression, and behavioral problems associated with mental retardation and developmental disorders, there remains a paucity of well-controlled studies. The range of serum lithium concentrations in adolescents is similar to that in adults and the likelihood of responding appears the same. The adverse effect profile of lithium is also the same across age groups, although in view of adolescent concern about body image, adverse events such as acne and weight gain may be poorly tolerated. Also, nontoxic cognitive dulling induced by therapeutic amounts of lithium may impact negatively on academic performance.

When and if to begin long-term lithium therapy is an even more difficult decision when treating young patients than when treating adults. Because of concern about possible long-term adverse effects, some experts think that children and young adolescents should not be treated for longer than 6 months. At the same time, they acknowledge that untreated bipolar disorders can have devastating effects on development.

Elderly Patients In the absence of organicity and substance abuse, most evidence suggests that advanced age alone does not compromise responsiveness to lithium. However, the use of lithium in the elderly is complicated by factors that include associated medical illnesses and medications, special diets, age-related reduction in GFR, and increased sensitivity to adverse effects.

Whether the elderly as a group respond to lower serum concentrations of lithium than do their younger counterparts is not known. Increased sensitivity to adverse effects and toxicity confine many elderly patients to relatively low doses and blood concentrations, which may prevent them from benefiting from the drug. However, this is not always the case:

> After ten hospitalizations for mania and depression beginning at age 20, a 53-year-old woman was stabilized on 1500 mg a day of lithium carbonate and remained on this regimen for the next 15 years with serum concentrations ranging between 0.6 and 0.75 mEq/L. It was only at age 68 that an increase in serum concentration led to a dose reduction to 1200 mg daily. The patient continues to take this dose at age 74 and remains in remission and is free of adverse effects.

In general, the elderly should be started on lower-than-usual dosages, with dosage changes occurring less frequently than in younger patients. The elimination half-life of lithium increases with age, and the time required to reach steady state is much longer in the elderly. If lithium is stopped, serum levels fall more slowly and the resolution of adverse effects and toxicity may be prolonged.

> A 60-year-old man was treated with 900 mg a day of lithium carbonate. The dosage was continued unchanged for 10 years, despite laboratory evidence of gradually increasing serum lithium and creatinine concentrations. Even as the clinical symptoms of toxicity were being reported, a thiazide diuretic was added to treat hypertension. Three weeks later, the patient was hospitalized with a serum lithium concentration of 4.2 mEq/L and marked neurological impairment that never fully resolved.

With appropriate monitoring, lithium use in the elderly can be both safe and effective. Direct comparisons with other mood stabilizers in this age population are lacking.

Pregnant Patients Lithium is in FDA pregnancy category D, which includes drugs for which there is evidence of human fetal risk, but whose potential benefits may outweigh the risk in some pregnant women. The International Register of Lithium Babies reported an 11 percent incidence of major malformations in babies exposed to lithium during the first trimester, with a disproportionate representation of cardiovascular malformations, especially Ebstein's anomaly of the tricuspid valve. These data may exaggerate risk because of the greater likelihood of reporting abnormal outcomes, failing to control for other medications and medical conditions, and failing to control for the effect of the illness itself on outcome. Recent studies have not confirmed such a high incidence of malformations and the link with cardiovascular anomalies. Also, extensive investigations in a variety of animal species have not found lithium to be a cardiac teratogen. Consequently, the teratogenic risk of lithium appears to be lower than was once believed, and exposure per se is not considered grounds for a therapeutic abortion. Nonetheless, the teratogenic risk of lithium still appears to be greater than that found in the general population (4 to 12 percent versus 2 to 3 percent). Fetal echocardiography is advised to screen for cardiovascular malformations in women exposed to lithium during the first trimester of pregnancy.

Physiological changes accompanying pregnancy alter maternal lithium metabolism. The GFR increases 30 to 50 percent over baseline and plasma volume increases by 50 percent. The filtered sodium load increases markedly (5000 to 10,000 mEq daily), as does the renal tubular reabsorption of sodium, resulting in a cumulative sodium retention of close to 1000 mEq. Complications of pregnancy, such as hypertension and edema and their treatments may further complicate lithium therapy. Following delivery, the altered renal physiology of pregnancy returns rapidly to normal. These factors and their variability among individual patients dictate that lithium use during pregnancy be closely supervised. Fetal and maternal blood concentrations are similar, so women should receive only minimum effective dosages. To reduce the risk of toxicity in the newborn, clinicians should markedly reduce or possibly temporarily discontinue the drug shortly before delivery. In view of evidence that abrupt or rapid discontinuation of lithium is associated with a high risk of relapse (which may be increased further postpartum), dosage reduction rather than discontinuation may be more appropriate as delivery approaches. Full prophylactic concentrations should be reestablished following delivery because this is a time of great vulnerability to recurrence. Because lithium appears in breast milk and in the blood of breast-fed babies, the American Academy of Pediatrics Committee on Drugs feels that lithium is contraindicated during breastfeeding. On the other hand, the merits of breastfeeding are considerable, and in some situations the benefits may outweigh the risks.

Medically Ill Patients The patient with a bipolar disorder and coexisting medical illnesses presents a particularly difficult therapeutic challenge. Untreated manic or depressive disorder can adversely affect a medical illness (e.g., mania and congestive heart failure) and lithium may be more difficult or impossible to use in the presence of a medical illness (e.g., severe renal disease). Also, because medically ill patients are receiving other medications, the risk of adverse drug interactions is increased. Consequently, treatment of such patients requires close collaboration between psychiatric and nonpsychiatric physicians.

Rapid-Cycling Patients Rapid cycling has been defined as four or more mood disorder episodes per year. Up to 20 percent of bipolar disorder patients are rapid cyclers and the majority of them are women. Risk factors associated with rapid cycling include a depression-hypomania course, antidepressant drug therapy, thyroid abnormalities, and neurological disease. Ultra-rapid cyclers have 24 to 48 cycles, and the concept of ultra-ultra rapid cycling has been introduced to describe patients with frequent mood shifts within a 24-hour period.

Overall, rapid-cycling patients are less responsive to lithium than are those who cycle more slowly, although both full and partial responses are common. Prior treatment with antidepressant drugs may predict a poorer response to lithium. Also, because antidepressant drugs may induce rapid cycling, merely discontinuing the antidepressant agent may be sufficient to slow the cycling pattern. Correcting thyroid abnormalities is an important aspect of treating rapid cycling. Preliminary research suggests that hypermetabolic doses of thyroid hormone may be useful in dealing with treatment-resistant rapid cycling. Valproate or carbamazepine alone or in combination with lithium are the most viable alternatives to lithium alone. Approaches that are less well established include calcium-channel inhibitors (particularly nimodipine [Nimotop] for ultrarapid cycling), choline, clozapine, olanzepine (Zyprexa) risperidone, newer anticonvulsants such as gabapentin (Neurontin) and lamotrigine (Lamictal), and ECT.

ADVERSE REACTIONS

Adverse effects are to be expected during lithium therapy. Fewer than 20 percent of patients have no adverse effects, but only about 30 percent have more than minor complaints. The incidence of severe adverse effects may be underestimated, however, because patients who do not tolerate therapy may not be available for

comment. Recognizing and minimizing adverse effects can do much to enhance compliance with lithium treatment (Table 31.18-3).

Neurological
Benign, nontoxic: dysphoria, lack of spontaneity, slowed reaction time, memory difficulties
Tremor: postural, occasional extrapyramidal
Toxic: coarse tremor, dysarthria, ataxia, neuromuscular irritability, seizures, coma, death
Miscellaneous: peripheral neuropathy, benign intracranial hypertension, myasthenia gravis–like syndrome, altered creativity, lowered seizure threshold
Endocrine
Thyroid: goiter, hypothyroidism, exophthalmos, hyperthyroidism (rare)
Parathyroid: hyperparathyroidism, adenoma
Cardiovascular
Benign T-wave changes, sinus node dysfunction
Renal
Concentrating defect, morphological changes, polyuria (nephrogenic diabetes insipidus), reduced GFR, nephrotic syndrome, renal tubular acidosis
Dermatological
Acne, hair loss, psoriasis, rash
Gastrointestinal
Appetite loss, nausea, vomiting, diarrhea
Miscellaneous
Altered carbohydrate metabolism, weight gain, fluid retention

Table 31.18-3 Adverse Effects of Lithium

Neurological Reactions Patients often complain of dysphoria, lack of spontaneity, slow reaction times, intellectual inefficiency, and spotty impairment of memory. Although these complaints may be subtle, they are a common cause of poor compliance. The following must also be considered as possible causes of these complaints: breakthrough depression, lithium-induced hypothyroidism or hypercalcemia, other illnesses, and other drugs. Single daily dosing at bedtime or a dosage reduction may be helpful, although switching to another mood stabilizer may be necessary.

> After 19 years of successful lithium maintenance, a 37-year-old woman was switched to carbamazepine because of polyuria and proteinuria. Only in retrospect did she and her physician become aware of the cognitive dulling and impaired concentration that had been present during her many years on lithium.

Tremor The most common tremor associated with lithium use is a benign postural tremor with a frequency of 8 to 12 Hz in the hands (which is similar to the tremor seen with valproate). Because the tremor worsens during activities requiring fine motor control, it can be socially embarrassing and occupationally troublesome. A nontoxic tremor often improves spontaneously, but if it does not, benefit may be obtained from dose reduction, use of a slow-release lithium preparation, elimination of dietary caffeine, discontinuation of other medications, and treatment of associated anxiety. Medications useful in treating lithium tremor include primidone (Mysoline) and b-adrenergic receptor antagonists such as propranolol (Inderal). With long-term lithium therapy, a tremor with parkinsonian characteristics may occur occasionally. A severe tremor at any time during the course of lithium therapy may be an indication of lithium intoxication.

Other Nontoxic Effects Lithium use has been associated rarely with peripheral neuropathy, downbeat nystagmus, benign intracranial hypertension (pseudotumor cerebri), a myasthenia gravis-like syndrome, and lowering of the seizure threshold. Creativity has been variously enhanced, impaired, and unaltered by lithium therapy.

Lithium Intoxication Lithium intoxication is primarily a neurotoxicity that can lead to death or permanent neurological damage (often cerebellar). Cardiovascular, gastrointestinal, and renal manifestations may also be present. Factors associated with toxicity include excessive intake (accidental or deliberate), reduced excretion (kidney disease, low-sodium diet, drug interaction), reduced volume of distribution (dehydration), and individual sensitivity (the elderly and the organically impaired).

> A woman with lithium-induced nephrogenic diabetes insipidus was hospitalized for treatment of a manic episode. Antipsychotic drug-induced sedation impaired her ability to replace fluids, and she quickly became lithium toxic and hypernatremic. Permanent cerebellar damage was the result.

There is no magic serum lithium concentration below which intoxication never occurs and above which it is inevitable. Severe toxicity has been described in patients whose serum concentrations were within the so-called therapeutic range of 0.5 to 1.5 mEq/L. Other patients have tolerated high concentrations after deliberate overdoses without noticeable clinical consequences. In general, however, the higher the concentration and the longer the exposure, the more serious the intoxication. Early clinical manifestations include dysarthria, ataxia, and coarse tremor. Ominous findings include impaired consciousness, neuromuscular irritability (fasciculations, myoclonus), seizures, and coma. Some cases of lithium intoxication have resembled Creutzfeldt-Jakob disease or neuroleptic malignant syndrome.

Treatment is directed at the removal of lithium from the body. In mild cases, nothing more may be necessary than discontinuation of intake. After an overdose, gastric lavage is indicated and may have to be repeated. Although activated charcoal is not helpful following an acute overdose with lithium, preliminary reports using polystyrene sulfonate (Kayexalate), a cation exchange resin, or whole bowel irrigation with polyethylene glycol solution (GoLYTELY) have been promising. In the presence of normal renal function, mild to moderate toxicity often responds to correcting dehydration and maintaining proper fluid and electrolyte balance. Whether forced diuresis provides additional benefit is open to debate. Hemodialysis is the treatment of choice for severe intoxication. Lithium is removed effectively by dialysis because it is water soluble, unbound to plasma protein, without metabolites, and of simple structure (an element). Redistribution of lithium from tissues to blood after dialysis usually results in a rebound increase in its blood level; this may necessitate further dialysis. Because of the drug's tissue-toxic effects and its gradual removal from the CNS, clinical improvement may lag many days behind the lowering of serum concentration.

Thyroid Reactions Transient mild abnormalities in thyroid function testing are common early in the course of lithium treatment, but are usually of little or no clinical consequence. Some patients, however, develop goiter or clinical hypothyroidism sometime during the course of treatment. Women and those with pre-existing thyroid dysfunction are more than usually susceptible. Among lithium's many effects on thyroid function, it impedes the release of hormone from the gland. The goiter that has been described in about 5 percent of patients taking lithium is rarely of cosmetic or obstructive importance; however, a recent ultrasound study suggested that thyroid gland enlargement is considerably more common than 5 percent.

Clinical hypothyroidism occurs in at least 4 percent of patients taking lithium. Once diagnosed, it can be treated with supplemental levothyroxine (Synthroid) at a dosage that returns the thyroid-stimulating hormone (TSH) concentration to normal. Chemical hypothyroidism (abnormal thyroid function tests without clinical manifestations) as determined by elevated serum TSH concentrations is considerably more common than clinical hypothyroidism. Substantial elevations of TSH are likely to progress to clinical hypothyroidism and thus should be treated with exogenous thyroid hormone. Minor elevations, on the other hand, especially early in the course of lithium therapy are likely to normalize without treatment.

Both exophthalmos and hyperthyroidism have been described during lithium therapy. One small study found long-term lithium use to be associated with thyrotoxicosis at a rate three times higher than that found in the general population. Monitoring of thyroid function (usually measurement of serum TSH concentration with or without triiodethyronine [T_3] and thyroxine [T4] concentrations) is an integral part of lithium therapy. Whether it is done routinely (every 6 or 12 months) or only when clinical suspicions are aroused is a matter of preference.

Cardiovascular Reactions Benign, reversible T-wave changes on the electrocardiogram (ECG) are common during lithium therapy and are of no clinical consequence. Varying degrees of heart block have been reported during lithium intoxication and occasionally during the therapeutic use of lithium. Electrophysiological studies show that lithium can impair sinus node function, and sinus dysrhythmias, sometimes associated with syncopal episodes, have been reported in a few patients. Lithium does not have clinically important effects on blood pressure.

Although the drug is not contraindicated in the presence of cardiovascular disease, it should be used with caution. A pacemaker may be necessary if lithium is used in the presence of clinically significant sinus node disease. Low-salt diets, certain diuretics, angiotensin-converting enzyme inhibitors, fluid-electrolyte imbalances, and impaired renal function all predispose to lithium toxicity.

Renal Reactions The great kidney debate over the nephrotoxicity of lithium is likely to resist full resolution, despite voluminous writings on the subject. In the majority of patients lithium impairs renal concentrating ability, which in itself is of no clinical importance. The concentrating defect is not always reversible after lithium

discontinuation, suggesting that both functional and structural changes have occurred in the distal tubules and collecting ducts.

Polyuria is the most clinically troublesome renal effect of lithium. Figures vary considerably, but a 24-hour urine volume of more than 3 liters (1 to 2 L is normal) has been reported in as many as 35 percent of patients taking lithium. Lithium increases urine volume primarily by inhibiting the effect of antidiuretic hormone on the kidneys, although additional renal and CNS mechanisms have been identified. When severe, polyuria can be socially and occupationally compromising; a cause of insomnia, weight gain (through consumption of high-calorie beverages), poor nutrition, and noncompliance; and potentially dangerous if dehydration occurs. Treatment considerations include (1) adequate fluid replacement, (2) using the lowest effective dosage, (3) and counteractive medications such as thiazides or potassium sparing diuretics, or indomethacin (Indocin). Amiloride (Midamor) or an amiloride-hydrochlorothiazide combination (Moduretic) is preferred to minimize the risk of hypokalemia. Whether single-daily dosing is beneficial remains controversial, as do treatments such as inositol, potassium supplementation, and desmopressin (DDAVP).

Lithium can also alter kidney morphology. The most common finding is a nonspecific interstitial fibrosis, although a lithium-distinctive tubular lesion has also been described. In the absence of lithium intoxication, the morphological changes attributed to lithium tend to be mild, especially when compared with appropriately selected controls.

Most studies of GFR find few or no reductions that could be attributed to lithium. Although reassuring, these observations do not exclude the possibility that a small number of patients are susceptible to lithium-induced reductions in GFR. Clinicians have described patients whose creatinine clearance has crept downward over the years to an extent that exceeds that expected from age alone. A study of 142 patients on lithium for a minimum of 15 years found reduced glomerular filtration rates in 21 percent and increased serum creatinine in 12 percent (after correcting for age).

A 55-year-old man, successfully stabilized on lithium for 22 years, experienced a gradual increase in serum creatinine to 2.4 mg/dL (range of normal 0.6 to 1.3). A thorough evaluation by a nephrologist produced no another explanation, and lithium was felt to be the most likely cause.

There have been occasional well-documented reports of lithium-related nephrotic syndrome and of incomplete distal renal tubular acidosis.

Regular monitoring of renal function is a necessary part of long-term lithium therapy. For most patients, the periodic measurement of serum creatinine, a urinalysis, and a clinical estimate of urine volume will suffice. If appropriate, a 24-hour urine volume, protein, and creatinine clearance will provide more comprehensive information. Because there are many causes of impaired kidney function, consultation with a nephrologist is appropriate if abnormalities are found.

Other Reactions Weight gain is a common adverse effect of lithium and may be due to the drug's complex effects on carbohydrate metabolism. Other possible causes include lithium-induced hypothyroidism and increased caloric intake from thirst-quenching beverages.

Gastrointestinal adverse effects—including nausea, appetite loss, and loose stools—are common but usually mild. Severe gastrointestinal symptoms may portend impending lithium intoxication.

The most common hematological effect of lithium is a benign, reversible, and usually mild granulocytosis induced by stimulation of granulocyte production. An increase in platelet count is noted occasionally, but there are no clinically important effects on red blood cell count.

Dermatological adverse effects include the first occurrence or worsening of acne and psoriasis, scattered reports of rashes of various types, and hair loss that is only occasionally related to lithium-induced hypothyroidism.

Alterations in calcium metabolism and parathyroid function have also been described. Lithium decreases the renal excretion of calcium, with an associated increase in serum calcium concentration, which usually does not exceed normal. Over time, parathyroid hyperplasia or a parathyroid adenoma may occur with elevated parathyroid hormone concentrations and hypercalcemia.

Sexual dysfunction, including erectile disorder, has been attributed to lithium, but documented cases are quite uncommon.

There is an inverse relation between dietary sodium intake and serum lithium concentration. Low-sodium diets result in increased serum lithium concentrations and may cause lithium intoxication unless the dosage of lithium is reduced. Diets that severely restrict fluid intake place patients at risk for dehydration and subsequent lithium intoxication.

DRUG INTERACTIONS

Antipsychotics Lithium and antipsychotic drugs have been used together safely and effectively for years; however, concern has been expressed that the combination may cause an encephalopathy in certain persons. Although the interaction has been most frequently reported with haloperidol (Haldol), a number of other antipsychotic drugs have also been implicated. A clear cause-and-effect relationship has not been established; nonetheless, caution is advised, and the simultaneous use of high dosages of the two drugs should be avoided.

Anticonvulsants In treatment-resistant cases, lithium is often combined with carbamazepine, clonazepam, or valproate and to a lesser extent with gabapentin and lamotrigine. Neurotoxic interactions have been described, but are usually avoidable with careful clinical management. When two drugs are used together, shared adverse effects may be more pronounced; for example, tremor and weight gain from lithium and valproate. In addition, lithium offers some protection against carbamazepine-induced granulocytopenia although whether it does the same for the rare agranulocytosis and aplastic anemia is unknown.

Antidepressants In general, lithium and antidepressant combinations are safe and effective, although there have been occasional reports of a serotonin-like syndrome when lithium was combined with potent serotonin reuptake inhibitors. Serum lithium concentrations are not altered in any substantial way by antidepressant drugs.

Anti-inflammatory Drugs Most nonsteroidal anti-inflammatory drugs reduce renal lithium clearance and increase the serum lithium concentration in a way that is clinically important and potentially dangerous. Among the drugs that cause this interaction are indomethacin (Indocin), phenylbutazone, diclofenac (Voltaren), ketoprofen (Orudis), oxyphenbutazone, ibuprofen (Motrin, Advil, Nuprin), piroxicam (Feldene), and naproxen (Naprosyn). Aspirin and sulindac (Clinoril) appear to be exceptions. Risk factors include high doses of the anti-inflammatory drug, increased age, and renal impairment. Because many of these drugs are now available without prescription, it is especially important that patients be aware of the possibility of an interaction.

Diuretics Thiazide diuretics decrease renal lithium clearance and increase the serum lithium concentration. Lithium toxicity has been known to occur. If a thiazide is prescribed, lithium dosage reduction is often necessary. If a thiazide is discontinued, lithium dosage may need to be increased to avoid subtherapeutic levels. Although less well established, potassium-sparing diuretics may also cause lithium retention. On the other hand, loop diuretics such as furosemide (Lasix) do not reduce and may actually increase renal lithium clearance. Osmotic and xanthine diuretics (caffeine, theophylline [Theo-dur], and aminophylline) also increase lithium clearance. Because diuretics are often given to patients who are medically ill with unstable fluid electrolyte status, interactions with lithium may not be predictable, and close monitoring is advised.

Angiotensin Converting Enzyme (ACE) Inhibitors ACE inhibitors are used extensively to treat cardiovascular disease. A growing number of case reports suggest that these drugs may cause lithium retention and toxicity in some patients. In contrast, losartan (Cozaar), an angiotensin II receptor (type AT_1) antagonist, does not appear to have this effect.

Other Drug Interactions Adverse interactions with varying degrees of substantiation have been reported between lithium and a number of other drugs (Table 31.18-4). These interactions may involve increases or decreases in serum lithium concentrations or neurotoxic symptoms in the presence of unchanged serum levels. Questions about potential interactions should be addressed by consulting a handbook on drug interactions, a regional drug information center, or the special lithium information center.

Table 31.18-4 Drug Interactions with Lithium

LABORATORY INTERFERENCES

Lithium does not appear to interfere in any clinically important way with other laboratory tests. A distinction must be made, however, between interfering with laboratory tests and causing alterations in laboratory tests. Lithium does cause alterations in a number of areas, including white blood cell count (increase), thyroid function (decrease), and serum calcium (increase). Those changes are induced by lithium rather than by an artifactual interference with the tests themselves.

Laboratory Monitoring Lithium has been measured in virtually every ody fluid, but only blood (serum or plasma) is used in clinical practice. Despite considerable research, red blood cell, saliva, and tear lithium measurements and ratios such as red blood cell to plasma have failed to gain widespread acceptance. Methods used to measure serum lithium include flame photometry, atomic absorption spectrophotometry, and ion-selective electrode analysis. Any of these methods in competent hands yields results that are acceptable for both therapeutic and toxic clinical situations. More recently, a colorimetric assay has been developed. The analyzers used in clinical laboratories are usually not capable of measuring the low levels of naturally occurring lithium in the blood, but more sensitive measures such as flameless atomic absorption spectroscopy and mass spectrometry are available for that purpose. Promising research has been conducted using Li^{7} magnetic resonance spectroscopy to measure lithium concentrations in human brain.

The periodic measurement of serum lithium concentration is an essential aspect of patient care, but it should always be combined with sound clinical judgment. A laboratory report listing the therapeutic range as 0.5 to 1.5 mEq/L may lull a clinician into disregarding early signs of lithium intoxication in patients whose levels are less than 1.5 mEq/L. Clinical toxicity, especially in the elderly, has been well documented as occurring within this so-called therapeutic range.

The 1996 United States package inserts for lithium products list effective serum concentration for mania between 1.0 and 1.5 mEq/L (usually achieved with 1800 mg of lithium carbonate daily) and for long-term maintenance between 0.6 and 1.2 mEq/L (usually achieved with 900 to 1200 mg of lithium carbonate daily). The dose-blood level relationship may vary considerably from patient to patient. The likelihood of achieving a response at levels above 1.5 mEq/L is usually outweighed greatly by the increased risk of toxicity, although occasionally a patient may both require and tolerate a higher-than-usual blood concentration.

What constitutes the lower end of the therapeutic range remains a matter of debate. A prospective 3-year study found that patients maintained at between 0.4 and 0.6 mEq/L (mean 0.54) were 2.6 times more likely to relapse than those maintained at between 0.8 and 1.0 mEq/L (mean 0.83). However, the higher blood concentrations produced more adverse effects and were less well tolerated. A recent reanalysis of these data concluded that lower maintenance levels per se were not responsible for the higher relapse rate, but rather that the risk of relapse was increased only in patients who were abruptly randomized from higher pre-study lithium levels to the lower level group. An international group of experts concluded recently that a good prophylactic effect can usually be obtained within the range of 0.4 to 0.8 mEq/L; unfortunately, trial and error is the only way to establish the minimum effective level for a given patient.

Essential to interpreting serum lithium levels is knowing the sampling interval (the time between the last dose and the drawing of blood), the dose form, and the dosage schedule. Blood concentrations obtained an hour or so after a dose are considerably higher than those obtained after a long sampling interval. The 12-hour interval that has been adopted as standard has been defined as follows: (1) the blood should be drawn in the morning, 12 hours (±30 minutes) after the last dose; (2) a multiple-dose regimen should be used; and (3) a steady-state condition should exist (skipped or extra doses within 4 or 5 days should be avoided). Although the effectiveness of lithium does not depend on the daily dosing schedule (every other day dosing, however, has been shown to be less effective than daily dosing), 12-hour serum concentrations will be higher with single as opposed to multiple daily dosing. This difference may range between 10 and 30 percent and cannot be easily predicted. Consequently, treatment is usually initiated with at least twice-daily dosing, although switching later to single-daily dosing may have the advantage of improved compliance and fewer adverse effects.

Even when these guidelines are followed, the serum level may fluctuate from measurement to measurement, reflecting factors such as dietary sodium intake, mood state, activity level, body position, and improper sample collection (use of a lithium-heparin anticoagulant will increase the concentration by more than 1 mEq/L).

Dosage Prediction In an effort to make lithium as therapeutically efficient as possible, investigators have developed several approaches to predicting the daily dosage required to produce the desired serum concentration. Dosage prediction techniques include giving a single dose of lithium carbonate and using the serum concentration after a given interval (usually 24 hours) to predict the maintenance dose; measuring renal lithium clearance; and the use of mathematical models. Thus far, prediction techniques have not been widely accepted and most clinicians continue to adjust the dosage based on the patient's clinical condition and previous blood concentrations. Nonetheless, “loading” a drug to produce a more rapid onset of action has become more appealing because of economic pressures to minimize length of hospital stay. Although one study found that a 30 mg/kg loading dose of slow-release lithium carbonate was well tolerated and achieved a therapeutic concentration in blood in 12 hours, it remains to be determined whether lithium loading can reduce time to remission or duration of hospitalization.

DOSAGE AND ADMINISTRATION

The extent of medical evaluation necessary before starting lithium is determined by individual patient characteristics such as age, associated illnesses, and the presence of other drugs; consequently, a thorough medical history is essential. The history determines the need for and the extent of physical and laboratory evaluations. All patients, however, should have renal and thyroid function tested before beginning treatment.

In the United States lithium preparations include immediate-release (Eskalith, Lithonate, Lithotabs), slow-release (Lithobid), and controlled-release (Eskalith CR) forms of lithium carbonate and a lithium citrate syrup. There are no parenteral forms. Both 300 mg of the carbonate and 5 ml of the citrate contain about 8 mEq (mmol) of lithium (Table 31.18-5). In view of the increasing prominence of valproate divalproex (Depakote) for the treatment of manic episode, the existence of a lithium valproate compound in Russia is of interest. Elsewhere in the world lithium acetate, glutamate, gluconate, orotate, and sulphate preparations have been available or are currently available. Clinicians should be aware that *lithium* and *lithium carbonate* are not interchangeable (300 mg of lithium is equivalent to 1597 mg of lithium carbonate).

Lithium carbonate capsules	150, 300, 600 mg
Lithium carbonate tablets	300 mg
Lithium carbonate controlled-release tablets	450 mg
Lithium carbonate slow-release tablets	300 mg
Lithium citrate syrup	8 mEq/5 mL

Table 31.18-5 Lithium Preparations Available in the United States

Starting Dosage Lithium therapy is initiated in divided doses. Once a patient is stabilized, single daily dosing is sometimes convenient. In the presence of normal kidney function, a total daily dose of 1200 to 1800 mg of lithium carbonate generally produces an antimanic serum concentration of 0.8 to 1.2 mEq/L. Maintenance levels of 0.6 to 1 mEq/L can usually be attained with 900 to 1200 mg daily. In general, a conservatively low dose is started, perhaps 300 mg twice or three times daily, a serum concentration is obtained after steady-state is reached (in 4 or 5 days), and the dose is adjusted accordingly. Patients vary widely in the dose required to reach a desired blood concentration although the elderly usually require lesser amounts. Consequently, these figures must be considered guidelines rather than absolutes.

Maintenance Dosage During maintenance therapy, patients must be evaluated clinically, lithium levels determined periodically, and appropriate laboratory tests performed at regular intervals. Monthly visits are common early in treatment if the clinical course is uncomplicated. Patients who have been stable for extended periods may be seen at intervals of 3, 4, or even 6 months. Patient education, including written materials and support programs, such as those by the National Depressive and Manic Depressive Association (NDMDA, 730 N. Franklin Street, Chicago, IL 60610), plays an important role in ensuring safe, effective, and compliant long-term treatment.

Discontinuation Lithium is discontinued if it is ineffective or not tolerated. Patients may stop the drug for other reasons, such as a perceived or real loss in creativity, feeling cured, or a dislike for feeling controlled by a medicine. After a period of stability with maintenance therapy, a trial off lithium may be considered, although the risk of recurrence is considerable (especially if there have been several prior episodes), and there have been reports of failure to respond to lithium when treatment is reinstituted. Discontinuation should be gradual over many weeks, because more abrupt discontinuation appears to be associated with a higher likelihood of early recurrence of mania or depression. Teaching patients and significant others to recognize early signs of recurrence is an important part of the discontinuation process.

SUGGESTED CROSS-REFERENCES

Intraneuronal signaling pathways are discussed in Section 1.8; mood disorders are discussed in Chapter 14. Chapter 31 presents biological therapies, particularly calcium channel inhibitors, carbamazepine, valproate, and antidepressant medications. Child psychiatry is discussed in Chapter 32 and Section 51.3d discusses mood disorders in the elderly. Schizophrenia is presented in Chapter 12 and other psychotic disorders are discussed in Chapter 13.

SECTION REFERENCES

Alessi N, Naylor MW, Ghaziuddin M, Zubieta JK: Update on lithium carbonate therapy in children and adolescents. J Am Acad Child Adolesc Psychiatry *33:*291, 1994.

American Psychiatric Association: Practice guideline for the treatment of patients with bipolar disorder. Am J Psychiatry *151*(Suppl):1, 1994.

*Baldessarini RJ, Tondo L: Recurrence risk of bipolar manic-depressive disorders after discontinuing lithium maintenance treatment: An overview. Clin Drug Invest *15:*337, 1998.

Baldessarini RJ, Tondo L, Floris G, Rudas N: Reduced morbidity after gradual discontinuation of lithium treatment for bipolar I and bipolar II: A replication study. Am J Psychiatry *154:*551, 1997.

Bauer M, Ahrens B: Bipolar disorder: A practical guide to drug treatment. CNS Drugs *6:*35, 1996.

*Bauer MS, Callahan AM, Jampala C, Petty F, Sajatovic M, Schaefer V, Wittlin B, Powell BJ: Clinical practice guidelines for bipolar disorder from the department of Veterans Affairs. J Clin Psychiatry *60:*9, 1999.

Bendz H, Aurell M, Balldin J, Mathé AA, Sjödin I: Kidney damage in long-term lithium patients: A cross-sectional study of patients with 15 years or more on lithium. Nephrol Dial Transplant *9:*1250, 1994.

Bendz H, Sjödin I, Aurell M: Renal function on and off lithium in patients treated with lithium for 15 years or more. A controlled, prospective lithium-withdrawal study. Nephrol Dial Transplant *11:*457, 1996.

Bowden CL, Brugger AM, Swann AC, Calabrese JR, Janicak PG, Petty F, Dilsaver SC, Davis JM, Rush AJ, Small JG, Garze-Treviño ES, Risch SC, Goodnick PJ, Morris DD: Efficacy of divalproex vs lithium and placebo in the treatment of mania. JAMA *271:*918, 1994.

Braden GL: Lithium-induced renal disease. In *Primer on Kidney Diseases*, A Greenberg, editor. Academic Press, San Diego, 1998.

Cade JFJ: The story of lithium. In *Discoveries in Biological Psychiatry*, FJ Ayd, B Blackwell, editors. Lippincott, Philadelphia, 1970.

Cohen LS, Friedman JM, Jefferson JW, Johnson EM, Weiner ML: A reevaluation of risk of in utero exposure to lithium. JAMA *271:*146, 1994.

Cookson J: Lithium: Balancing risks and benefits. Br J Psychiatry *171:*120, 1997.

DasGupta K, Jefferson JW: Treatment of mania in the medically ill. In *Psychotropic Drug Use in the Medically Ill*, PA Silver, editor. Basel, Karger, 1994.

Dixon JF, Hokin LE: The antibipolar drug valproate mimics lithium in stimulating glutamate release and inositol 1,4,5-trisphosphate accumulation in brain cortex slices but not accumulation of inositol monophosphates and biphosphates. Proc Natl Acad Sci USA *94:*4757, 1997.

El-Mallakh RS: *Lithium Actions and Mechanisms*. American Psychiatric Press, Washington DC, 1996.

Finley PR, Warner MD, Peabody CA: Clinical relevance of drug interactions with lithium. Clin Pharmacol *29:*172, 1995.

Gelenberg AJ, Jefferson JW: Lithium tremor. J Clin Psychiatry *56:*283, 1995.

Gelenberg AJ, Kane JM, Keller MB, Lavori PW, Rosenbaum JF, Cole K, Lavelle J: Comparison of standard and low serum levels of lithium for maintenance treatment of bipolar disorder. N Engl J Med *321:*1489, 1989.

Goodnick PJ, editor: *Mania: Clinical and Researchers Perspectives*. American Psychiatric Press, Washington DC, 1998.

*Goodwin FK, Jamison KR: *Manic-Depressive Illness*. Oxford University Press, New York, 1990.

Hong M, Chen DCR, Klein PS, Lee VM-Y. Lithium reduces tau phosphorylation by inhibition of glycogen synthase kinase-3. J Biol Chem *272:*25326, 1997.

Jefferson JW: Lithium still effective despite its detractors. Br Med J *316:*1330, 1998.

Jefferson JW: Lithium. In *Predictors of Treatment Response in Mood Disorders*, PJ Goodnick, editor. American Psychiatric Press, Washington, DC, 1996.

Jefferson JW, Greist JH: Lithium in psychiatry: A review. CNS Drugs *1:*448, 1994.

*Jefferson JW, Greist JH, Ackerman DL, Carroll JA: *Lithium Encyclopedia for Clinical Practice*, ed 2. American Psychiatric Press, Washington, DC, 1987.

Johnson FN: *The History of Lithium Therapy*. Macmillan, London, 1984.

Johnson FN: *Lithium and the Endocrine System*. Karger, Basel, 1988.

*Jope RS: A bimodal model of the mechanism of action of lithium. Molecular Psychiatry *4:*21, 1999.

Kallner G, Petterson U: Renal, thyroid and parathyroid function during lithium treatment: Laboratory tests in 207 people treated for 1–30 years. Acta Psychiatr Scand *91:*48–51, 1995.

Lenox RH, Manji HK: Lithium. In *The American Psychiatric Drug Textbook of Psychopharmacology*, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1995.

Maj M, Pirozzi R, Magliano L, Bartoli L: Long-term outcome of lithium prophylaxis in bipolar disorders: A 5-year prospective study of 402 patients at a lithium clinic. Am J Psychiatry *155:*30, 1998.

Manji HK, Chen G, Hsiao JK, Risby ED, Masana MI, Potter WZ: Regulation of signal transduction pathways by mood-stabilizing agents: Implications for the delayed onset of therapeutic efficacy. J Clin Psychiatry *57* (Suppl):34, 1996.

McClellan J, Werry J: Practice parameters for the assessment and treatment of children and adolescents with bipolar disorder. J Am Acad Child Adolesc Psychiatry *36:*138, 1997.

*Merrill W: Case 35–1998: Use of lithium to prevent corticosteroid-induced mania. N Engl J Med *340:*1123, 1999.

Moncrieff J: Lithium: Evidence reconsidered. Br J Psychiatry *171:*113, 1997.

Mukherjee S: Combined ECT and lithium therapy. Convuls Ther *9:*274, 1993.

*Müller-Oerlinghausen B: Drug interactions with lithium—a guide for clinicians. CNS Drugs *11:*41, 1999.

Ozaki N, Chuang D-M: Lithium increases transcription factor binding to AP-1 and cyclic AMP-responsive element in cultured neurons and rat brain. J Neurochem *69:*2336, 1997.

Post RM, Leverich GS, Altshuler L, Mikalauskas K: Lithium-discontinuation-induced refractoriness: Preliminary observations. Am J Psychiatry *149:*1727, 1992.

Price LH, Heninger GR: Lithium in the treatment of mood disorders. Drug Therapy *331*(9):591, 1994.

Sanborn K, Jefferson JW: Everyman's guide to the fluctuating lithium level: Obvious and obscure reasons why serum lithium levels change. Ann Clin Psychiatry *3:*251, 1991.

*Schou M: Forty years of lithium treatment. Arch Gen Psychiatry *54:*9, 1997.

Schou M: Phases in the development of lithium treatment in psychiatry. In *The Neurosciences: Paths of Discovery II*, F Samson, G Adelman, editors. Birkhauser, Boston, 1992.

Silverstone T, Romans S: Long term treatment of bipolar depression. Drugs *51:*367, 1996.

Solomon DA: Polypharmacy for maintenance treatment of patients with bipolar I disorder. Essential Psychopharmacol *2:*287, 1998.

Suppes T, Baldessarini RJ, Faedda GL, Tondo L, Tohen M: Discontinuation of maintenance treatment in bipolar disorder: Risks and implications. Harvard Rev Psychiatry *1:*131, 1993.

Tondo L, Baldessarini RJ, Hennen J, Floris G, Silvetti F, Tohen M: Lithium treatment and risk of suicidal behavior in bipolar disorder patients. J Clin Psychiatry *59:*405, 1998.

Tondo L, Baldessarini RJ, Hennen J, Floris G: Lithium maintenance treatment of depression and mania in bipolar I or bipolar II disorders. Am J Psychiatry *155:*638, 1998.

Yonkers KA, Little BB, March D: Lithium during pregnancy. Drug effects and their therapeutic implications. CNS Drugs *9:*261, 1998.

Zornberg GL, Pope HG: Treatment of depression in bipolar disorder: New directions for research. J Clin Psychopharmacol *13:*397, 1993.

CHAPTER 31. BIOLOGICAL THERAPIES

31.19 MIRTAZAPINE

JAMES L. CLAGHORN, M.D.

Mirtazapine (Remeron) is the first among a new class of antidepressants described as noradrenergic and specific serotonergic antidepressants. Mirtazapine was first marketed in the Netherlands in 1994; it has since been marketed worldwide for the treatment of depression. Mirtazapine was introduced in the United States in 1996 for the treatment of depression.

CHEMISTRY

Mirtazapine (1,2,3,4,50,14b-hexahydro-2-methylpyrazino [2,1-a]pyrido[2,3-c][2]benzazepine) is a novel antidepressant. It belongs to the chemical class of compounds known as the piperazinoazepines and is not related to any known class of psychotropic drugs. It has the empirical formula $C_{17}H_{19}N_3$, and its molecular weight is 265.36 mol. Mirtazapine is marketed as a racemic mixture (50/50) of the (+) and (–) enantiomers, both of which possess pharmacological activity essential for the overall antidepressant effects of the racimate (Fig. 31.19-1).

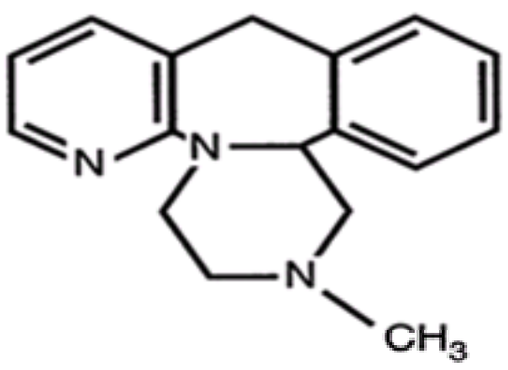

FIGURE 31.19-1 Molecular structure of mirtazapine.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Mirtazapine is well absorbed from the gastrointestinal tract after single or multiple oral doses and reaches peak plasma concentrations within 2 hours. Both the rate and extent of absorption of mirtazapine are minimally affected by the presence of food in the stomach, so patients may take mirtazapine without regard to meals. The absolute bioavailability of mirtazapine after single and multiple oral dosing is approximately 50 percent.

Distribution Human plasma protein binding is aspecific and reversible, and mirtazapine is approximately 85 percent protein bound at therapeutic dosages.

Metabolism Orally administered mirtazapine is extensively metabolized by the liver via demethylation, hydroxylation and N-oxidation, followed by conjugation with glucuronic acid. In vitro data from human liver microsome studies indicate that mirtazapine is metabolized by multiple cytochrome P450 (CYP) isoenzymes, namely CYP 2D6, CYP 3A4, and possibly CYP 1A2 to produce the 8-hydroxy, *N*-desmethyl, and *N*-oxide metabolites. Demethyl-mirtazapine is the only metabolite that demonstrates pharmacological activity. However, it is five to ten times less active than the parent compound, and it has been estimated that the contribution of the total pharmacodynamic profile of mirtazapine is 3 to 6 percent. Mirtazapine does not appear to have any inducing or inhibiting effects on the cytochrome P450 system, and its disposition appears to be independent of polymorphic CYP 2D6 activity. It appears unlikely that mirtazapine would inhibit the metabolism of drugs metabolized by CYP 1A2, CYP 2D6, or CYP 3A4. Potent inhibitors of CYP 2D6 would not affect the metabolism of the major enantiomer of mirtazapine, though its effect on the minor enantiomer cannot be predicted.

Elimination Mirtazapine follows linear pharmacokinetics within a dose range between 15 and 75 mg and has a mean elimination half-life of 20 to 40 hours after oral administration. Steady-state plasma concentrations of mirtazapine are attained within 5 days, with about 50 percent accumulation. It is eliminated primarily in the urine (80 percent), with small amounts in the feces (15 percent), up to 4 percent of the orally administered drug can be eliminated unchanged in the urine. Elderly persons and patients with liver or renal disease may require cautious dose increases.

Pharmacodynamics

Mechanism of Action As with all antidepressants, the exact mechanism of action of mirtazapine is unknown. Its pharmacological effects can best be described as enhancing noradrenergic activity and having specific serotonergic activity. Preclinical studies demonstrated that mirtazapine does not inhibit norepinephrine reuptake but preferentially antagonizes central presynaptic a_2-adrenergic autoreceptors or heteroreceptors, which respectively control norepinephrine and serotonin release. Mirtazapine blocks presynaptic a_2-adrenergic autoreceptors, leading to an increase in the release of norepinephrine. Mirtazapine's affinity in vitro for central presynaptic a_2-autoreceptors is approximately 30 times higher than that for central and peripheral a_1-adrenoreceptors. The low affinity of mirtazapine for central a_1-adrenoreceptors located on serotonin cell bodies allows for unopposed stimulation of a_1-receptors by released norepinephrine; this increases the firing rate of serotonergic neurons. Mirtazapine also blocks a_2-adrenergic heteroreceptors located on serotonin terminals, thus enhancing serotonin release; therefore the net effect of mirtazapine is an increase in both noradrenergic and serotonergic neurotransmission.

The a_2-adrenergic autoreceptors and heteroreceptors and serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) blocking effects predominate in the (+)-enantiomer; a_2-heteroreceptor blockade and 5-HT_3-blocking effects predominate in the (–)-enantiomer.

The stimulation or blockade of several postsynaptic serotonin receptor subtypes is responsible for both clinical and adverse effects. Stimulation of 5-HT receptors is thought to mediate the therapeutic effects of antidepressants, whereas stimulation of 5-HT_2 and 5-HT_3 receptors is thought to be involved with both clinical and adverse effects. Mirtazapine has a low affinity for 5-HT_{1A} and 5-HT_{1B} subtypes, thus allowing unopposed stimulation of postsynaptic 5-HT_{1A} receptors. In addition, the drug is a potent antagonist of 5-HT_2 and 5-HT_3 receptors, which may account for low incidences of serotonin-related adverse effects such as anxiety, insomnia, nausea, and sexual dysfunction reported during the use of mirtazapine. Blockade of these receptors has also been associated with deep-sleep–promoting and

anxiolytic activity. Mirtazapine reduces rapid eye movement (REM) sleep, increases REM sleep latency, and improves deep sleep.

Mirtazapine has little affinity for dopamine type 1 (D_1) and D_2 receptors and has only moderate affinity for muscarinic and cholinergic receptors, which may explain the low incidence of anticholinergic and dopaminergic adverse effects. Mirtazapine is a potent antagonist of histamine type 1 (H_1) receptors, which may explain its somnolence-inducing effects. Although mirtazapine can cause daytime somnolence, this effect is thought to be partially offset by its action on the noradrenergic system at usual therapeutic dosages.

Blood Concentrations and Relation to Action The available data demonstrate that mirtazapine has linear pharmacokinetics within a dose range of 15 to 75 mg. As with other new antidepressants, there are currently no data to support routine blood concentration measurements as predictors of clinical response.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

In published reports of clinical trials, mirtazapine has been compared with placebo, amitriptyline (Elavil), fluoxetine (Prozac), trazodone (Desyrel), doxepin (Adapin, Sinequan), and clomipramine (Anafranil) in the treatment of depression (Table 31.19-1). Both inpatients and outpatients who met the criteria for major depression in the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) participated in these studies. The 17-item Hamilton Rating Scale for Depression (HAM-D) total score was used to assess baseline severity of depression. Patients ranging from 18 to 65 years of age and over with HAM-D scores of 18 or above (moderate to severe depression) participated in the studies.

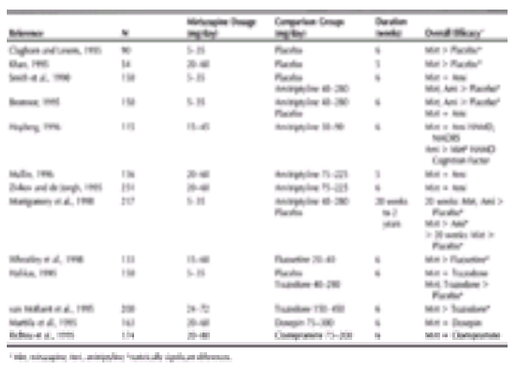

Table 31.19-1 Selected Studies of Mirtazapine in the Treatment of Depression

Antidepressant efficacy was measured using the 17-item HAM-D, Montgomery-Aåsberg Depression Rating Scale (MADRS), Beck Depression Inventory, and Zung's Self-Rating Depression Scale (SDS). The Brief Psychiatric Rating Scale (BPRS), the Clinical Global Impression (CGI) Scale, the Visual Analogue Mood Rating (VAMR) Scale, Quality of Life Enjoyment and Satisfaction Questionnaire (QLESQ), and the Global Assessment Scale (GAS) were used to assess general psychopathology, clinical impression, and global functioning. Efficacy analyses were performed on an intent-to-treat basis, thus including all patients who had received at least one dose of study medication and had at least one postbaseline efficacy assessment: efficacy analysis was based on changes from baseline in rating scales scores during treatment. In addition to weekly observed-case analysis, a last-observation-carried-forward analysis was performed until week 6. A positive response to treatment was defined as a 50 percent or more reduction in the 17-item HAM-D score from baseline.

Most clinical studies allowed dosages to be tailored to individual patient needs and to be titrated on the basis of patient response, which reflects most clinical practice. Dosages ranged from 5 to 60 mg a day in placebo-controlled studies and 5 to 80 mg a day in the active-controlled studies.

Mirtazapine's use in treating a subset of patients with sleep disturbances, anxiety, and cognitive impairment was evaluated using the HAM-D Sleep Disturbance factor IV (sum of items 4, 5, and 6), HAM-D anxiety/somatization factor (sum of items 10, 11, 12, 13, 15, and 17), the HAM-D anxiety/agitation factor (sum of items 9, 10, and 11) and HAM-D Cognitive Disturbance factor III.

Major Depressive Disorder

Comparisons of Mirtazapine and Placebo Ninety outpatients between the ages of 18 and 65 years with moderate-to-severe depression were randomized to either mirtazapine (5 to 35 mg a day) or placebo in a 6-week, double-blind trial. Mirtazapine was significantly superior to placebo (*P* £ .05) for patients with moderate-to-severe depression in weeks 1 to 4 and at endpoint for the following efficacy parameters: 17-item HAM-D, MADRS, and CGI.

Statistically significant (*P* £ .05) between-treatment differences were observed that favored mirtazapine at weeks 1 to 5 and at endpoint in the reduction of the anxiety/somatization factor, at weeks 2 to 3 and at endpoint for the retardation factor, weeks 2 to 4 and endpoint for the cognitive disturbance factor, and weeks 1 to 4 and endpoint for the sleep disturbance factor. The dropout rates were 40 percent for mirtazapine-treated patients and 58 percent for placebo-treated patients. The percentage of dropouts for drug-related adverse effects or lack of efficacy was 27 percent for mirtazapine-treated patients and 40 percent for placebo-treated patients. Significantly more placebo-treated patients (40 percent) than mirtazapine-treated patients (18 percent) discontinued treatment because of a lack of efficacy. Mirtazapine was well tolerated, with somnolence and increased appetite the only adverse effects significantly different from those with placebo.

The efficacy of mirtazapine (20 to 60 mg a day) was compared with that of placebo in a 5-week, randomized, double-blind, placebo-controlled study in 54 inpatients and outpatients with a diagnosis of major depressive disorder. Mirtazapine-treated patients had statistically significant and clinically relevant differences at weeks 4, 5, and at endpoint compared with placebo-treated patients on the 17-item HAM-D and MADRS scales. At the end of the study 56 percent of mirtazapine-treated patients and 36 percent of placebo-treated patients responded to therapy. A similar percentage of patients withdrew from the study, and there was no statistically significant difference between mirtazapine-treated and placebo-treated patients in reported/adverse clinical effects.

In addition to the results of these placebo-controlled trials, a meta-analysis was performed on efficacy data from three U.S. and two European placebo-controlled studies. A total of 495 patients (mirtazapine, 249; placebo, 246) were available for evaluation. The results of the meta-analysis confirmed the efficacy findings of the individual clinical trials: mirtazapine-treated patients showed a significantly greater reduction from baseline in the mean group 17-item HAMD scores than placebo-treated patients. The difference was seen as early as week 1 of treatment and continued throughout the treatment period. Mirtazapine-treated patients also demonstrated significantly greater improvements in MADRS scores at all weekly assessments and at endpoint. At endpoint approximately 51 percent of mirtazapine-treated patients responded to treatment (versus 33 percent of placebo-treated patients) as assessed by HAM-D criterion (*P* £ .05). Also, an analysis of dropout due to lack of efficacy indicated that mirtazapine-treated patients (13.7 percent) were less likely to drop out because of lack of efficiency (*P* £ .05) than placebo-treated patients (21.8 percent).

Comparison of Mirtazapine and Amitriptyline The efficacy of mirtazapine was compared with that of amitriptyline in several placebo-controlled studies. In one 6-week, double-blind study 150 outpatients were randomized to receive mirtazapine (5 to 35 mg a day), amitriptyline (40 to 280 mg a day), or placebo. Both active treatment groups showed statistically significant improvement at weeks 1, 2, 4, and endpoint in HAM-D scores compared with placebo. These findings were further supported by results on the MADRS, SDS, and CGI Scale. Somnolence, constipation, and weight gain were the only three adverse effects reported by more mirtazapine-treated patients than placebo-treated patients, although they were not substantially different from that reported in the amitriptyline group. However, significantly more amitriptyline-treated patients reported decreased visual accommodation, dry mouth, headache, dyspepsia, constipation, tachycardia, hypertension, hypotension, discoordination, dizziness, and tremor than mirtazapine-treated or placebo-treated patients. In this study mirtazapine was more effective than placebo and as effective as amitriptyline in treating depression, with a significantly better tolerability profile than amitriptyline.

Another trial compared the efficacy of mirtazapine (5 to 35 mg a day) with amitriptyline (40 to 280 mg a day) in a 6-week, double-blind placebo-controlled study in 150 outpatients. Statistically significant HAM-D score reductions (*P* £ .05) were seen in the mirtazapine-treated and amitriptyline-treated patient groups from week 1 through endpoint, compared with placebo-treated patients. Mirtazapine-treated patients demonstrated statistically significant (*P* £ .05) reductions at weeks 1 to 6 and endpoint for the HAM-D item 1 (depressed mood) compared with placebo-treated patients. Amitriptyline-treated patients demonstrated statistically significant (*P* £ .05)

reductions at weeks 2, 3, and 4 on the same item, compared with patients treated with placebo. Both active-treatment groups showed significantly greater improvements than placebo-treatment groups on the MADRS, SDS, and CGI Scales. This was further evidenced by the percentage of patients (responders) who had at least a 50 percent reduction in HAM-D total score from baseline to endpoint. Significant differences were noted in responder rates after 6 weeks of treatment: 70 percent of the patients receiving mirtazapine responded, compared with 58 percent of amitriptyline-treated patients, and 33 percent of placebo-treated patients. Mirtazapine was well tolerated, with dry mouth and somnolence occurring more frequently than in placebo-treated patients. Treatment with amitriptyline was associated with significantly higher rates of somnolence, dry mouth, constipation, and dyspepsia than treatment with mirtazapine and placebo.

Some 251 inpatients with major depressive disorder were treated with either mirtazapine (20 to 60 mg a day) or amitriptyline (75 to 225 mg a day) in a 6-week, randomized, double-blind, multicenter study. Most patients in both treatment groups were severely depressed as determined by a group mean 17-item HAM-D score of 28 for mirtazapine-treated patients and 27.6 for amitriptyline-treated patients. Both treatment groups demonstrated equal improvement at all scheduled assessments and at the end of the trial (72 percent response rate). Similar improvements were noted on BPRS and GAS measurements. There were no statistically significant differences between the treatment groups on any scale used. Both treatments were generally well tolerated; however, dry mouth and sweating were reported significantly more frequently after 42 days of treatment with amitriptyline than with mirtazapine (*P* £ .05).

A total of 217 patients who participated in a short-term, double-blind comparison of mirtazapine (5 to 35 mg a day), amitriptyline (40 to 280 mg a day), and placebo were enrolled in a long-term extension for up to 2 years. Efficacy and safety analyses were performed on the first 20 weeks of data. Significantly (*P* < .05) fewer mirtazapine-treated patients relapsed than placebo-treated patients, and mirtazapine-treated patients had a significantly longer time to relapse as well. Similar results were found comparing amitriptyline-treated patients with placebo-treated patients. Mirtazapine also held an advantage in preventing relapses, compared with amitriptyline (*P* < .05). Similar results were found in the prophylactic phase of treatment after 20 weeks. Amitriptyline was better than placebo; however, there was no difference in the number of patients with a sustained response. Mirtazapine was well tolerated, with fewer mirtazapine-treated patients reporting adverse effects than amitriptyline-treated patients.

A meta-analysis was performed on the efficacy and safety data from four randomized, double-blind, 6-week, single-center studies comparing mirtazapine (5 to 35 mg a day) with amitriptyline (40 to 280 mg a day) and placebo in 580 outpatients with DSM-III diagnoses of major depression. Both active treatments were equally effective and significantly (*P* £ .05) more effective than placebo at weekly assessments, as assessed by reductions from baseline on the 17-item HDRS, MADRS, CGI scores, HDRS item 1 (depressed mood), and HDRS factors (anxiety/somatization, sleep disturbance, and melancholia). Both treatment groups demonstrated statistically significant (*P* £ .05) greater improvement and significantly greater percentages of responders and remitters than placebo. Mirtazapine was better tolerated than amitriptyline. Amitriptyline-treated patients reported significantly more dry mouth, dyspepsia, blurred vision, constipation, tremor, vertigo, and tachycardia than mirtazapine-treated patients and placebo-treated patients.

In another study, 115 elderly patients with moderate-to-severe depression were randomized to receive either mirtazapine (15 to 45 mg a day) or amitriptyline (30 to 90 mg a day) in a double-blind, flexible-dose, 6-week study. Both treatments were equally effective, as assessed by the 21-item HRSD and MADRS scales at weeks 2, 4, 6, and endpoint, with no statistically significant differences between the treatment groups at any assessment time or endpoint. Minor, but statistically significant (*P* £ .05), changes favoring amitriptyline were noted at endpoint on GCI-Global Improvement Scale, at all assessments on the HRSD cognitive factor, and at 6 weeks for the HRSD retardation factor. However, an equally high percentage of mirtazapine- (74 percent) and amitriptyline-treated patients (81 percent) responded to therapy according to CGI criteria. Both drugs were well tolerated, and there were no statistically significant differences in reports of adverse reactions.

A meta-analysis was performed on pooled data from randomized, double-blind comparisons of mirtazapine and amitriptyline; 732 patients were available for efficacy analysis. The two treatments were found equally effective at all weekly assessments and at endpoint, as assessed by reductions from baseline in total HAM-D and HAM-D item 1 scores. A similar percentage of mirtazapine-treated patients (70 percent) and amitriptyline-treated patients (73 percent) responded to treatment. Analysis of patients with moderate-to-severe depression indicated that reductions from baseline were statistically significantly larger (*P* £ .001) for mirtazapine-treated patients than for placebo-treated controls and equivalent to those of amitriptyline-treated patients. Also, the percentage of mirtazapine-treated patients (9 percent) who dropped out because of a lack of efficacy was comparable to that of amitriptyline-treated patients (6.6 percent).

Another meta-analysis was performed on pooled data from 405 severely depressed patients from randomized, double-blind, amitriptyline-controlled trials. The treatments were equally effective at all weekly assessments and at endpoint, as assessed by reductions from baseline in total HAM-D and HAM-D item 1 scores. Mirtazapine and amitriptyline were equally effective in treating depressed mood and symptoms more commonly associated with depression, such as anxiety, sleep, and vegetative disturbances.

Comparison of Mirtazapine and Fluoxetine The efficacy of mirtazapine (15 to 60 mg a day) was compared with that of fluoxetine (20 to 40 mg a day) in 133 depressed (mean HAM-D_{17}: 26.0, mirtazapine and 26.1, fluoxetine) inpatients and outpatients in a multicenter, 6-week study. Patients from both treatment groups responded to therapy, although the decrease from baseline on the HAM-D was larger in the mirtazapine group than the fluoxetine group throughout the treatment period. The mirtazapine-treated patients had statistically significant (*P* £ .05) improvement at days 21 and 28 compared with the fluoxetine group. At the assessments from day 21 forward the absolute difference favored mirtazapine over fluoxetine by between 3.7 and 4.2 points on the HAM-D scale, which is comparable to the difference observed in a comparison of antidepressant therapy and placebo. These findings were supported by the results of the CGI Scale. Patients from both treatment groups became calmer and more alert, energetic, relaxed, and tranquil, as assessed by the VAMR Scale. A statistically significant difference favored mirtazapine at day 7 for the Calm/Excited, Lethargic/Energetic, and Agitated/Tranquil items. Both treatment groups demonstrated an improvement in quality of life, as determined by the QLESQ Scale. Both treatments were well tolerated. More fluoxetine-treated patients complained of headache (17.9 percent) and nausea (10.4 percent), while more mirtazapine-treated patients complained of dry mouth (18.2 percent) and blurred vision (7.6 percent).

Comparisons of Mirtazapine and Other Antidepressants The efficacy of mirtazapine was also assessed in comparative trials with trazodone, doxepin, and clomipramine. In one study, mirtazapine (5 to 35 mg a day) was compared with trazodone (40 to 280 mg a day) and placebo in 150 outpatients 55 years of age or older in a 6-week, randomized, double-blind trazodone- and placebo-controlled trial. Statistically significant differences between both active compounds and placebo were noted at weeks 2, 3, and 4 on the MADRS. Mirtazapine (but not trazodone) showed a significant (*P* £ .05) improvement compared with placebo, as assessed by reductions from baseline in the 21-item HAM-D total score at weeks 2, 3, 4, and 6. Mirtazapine and trazodone were associated with a significantly higher frequency of somnolence and dry mouth than placebo; treatment with trazodone resulted in significantly higher frequencies of dizziness and blurred vision than treatment with placebo.

Two hundred inpatients with moderate-to-severe depression were randomized to receive either mirtazapine (24 to 72 mg a day) or trazodone (150 to 450 mg a day) in a 6-week, double-blind trial. Mirtazapine was clinically more efficacious on day 42 as assessed by changes in HAM-D, BPRS, GAS, and Beck Depression Inventory scores and percentage of responders (*P* £ .05). Mirtazapine was well tolerated. A greater percentage of trazodone-treated patients experienced somnolence and postural hypotension, although the differences were not statistically significant. Mirtazapine demonstrated significant clinical advantages over trazodone in terms of clinical efficacy and tolerability.

Mirtazapine (20 to 60 mg a day) was compared with doxepin (75–300 mg a day) in a 6-week, randomized, double-blind clinical trial. The drugs demonstrated equal efficacy in 163 hospitalized patients or outpatients with depression. Approximately 65 percent of mirtazapine-treated patients and 62 percent of doxepin-treated patients responded to treatment at endpoint; doxepin-treated patients complained more frequently of dry mouth and movement disorders.

Another study compared mirtazapine (20–80 mg a day) with clomipramine (50–200 mg a day) in a multicenter, double-blind, 6-week trial. A total 174 hospitalized, endogenously depressed patients were randomized to receive either mirtazapine or clomipramine. At the end of the study, 80 percent of mirtazapine-treated patients and 86 percent of clomipramine-treated patients responded to therapy, as assessed by the HAM-D criterion. Depressive symptoms assessed by the MADRS, BPRS, and GAS also improved significantly in both treatment groups. Dry mouth, constipation, tremor, vertigo or dizziness, faintness on rising, and nausea were reported more frequently by clomipramine-treated patients. Mirtazapine appears to be as effective as clomipramine in the treatment of depression and is better tolerated.

Major Depressive Disorder With Sleep Disturbances The effect of mirtazapine on sleep disturbances in depressed patients has been examined using HAM-D factor IV (sum of items 4, 5, and 6) in placebo-controlled clinical trials. One placebo-controlled study also showed a significant improvement (*P* £ .05) in sleep disturbances in mirtazapine-treated patients beginning in the first week of treatment, compared with placebo-treated patients. Similarly, the second study showed significantly greater improvement in sleep disturbances (*P* £ .05) in the mirtazapine and amitriptyline groups, compared with placebo. The third study showed significant (*P* £ .05) improvement in sleep disturbances from the second week onward in the mirtazapine and trazodone treatment groups compared with the placebo group. The results of two meta-analyses confirm the efficacy findings of the individual clinical trials. A meta-analysis of 364 mirtazapine-treated patients and 368 amitriptyline-treated patients showed that mirtazapine (5 to 60 mg a day) was significantly better than placebo (*P* £ .05) and comparable to amitriptyline (75 to 280 mg a day) for improving sleep disturbances during treatment and at the endpoint analysis. Another meta-analysis of three U.S. and two European placebo-controlled

studies evaluated the efficacy of mirtazapine for improving sleep disturbances in a total of 249 mirtazapine-treated and 246 placebo-treated patients. Significant (*P* £ .05) improvement favoring mirtazapine was shown for the overall population as well as for a subgroup of depressed patients with prominent sleep disturbances at baseline.

Major Depressive Disorder With Anxiety Mirtazapine's potential value in treating patients with major depressive disorder with anxiety features has also been studied. Mirtazapine-treated patients experienced significantly greater improvements from baseline in somatic anxiety, psychic anxiety, and agitation than placebo-treated patients. In one study, the mirtazapine and amitriptyline groups had significantly (*P* £ .05) greater improvement in the HAM-D Anxiety factor scores. Similar results were noted in another study in which the reductions in the mean HAM-D Anxiety factor scores were significantly greater (*P* £ .05) in the mirtazapine and trazodone groups than in the placebo group. A meta-analysis of 364 mirtazapine-treated (5 to 60 mg a day) and 368 amitriptyline-treated (75 to 280 mg a day) patients demonstrated that mirtazapine was significantly better than placebo and similar to amitriptyline in reducing HAM-D Anxiety factor scores.

A meta-analysis was performed to compare the efficacy data of mirtazapine with that for placebo and amitriptyline in patients with major depressive disorder and either Anxiety/Agitation (sum of HAM-D items 9, 10, 11) or Anxiety/Somatization (mean of HAM-D items 10, 11, 12, 13, 15, 17). Efficacy data from eight randomized, double-blind, placebo-controlled clinical trials were evaluated. A total of 385 patients (132 placebo-treated, 161 mirtazapine-treated and 92 amitriptyline-treated) were available for evaluation. Mirtazapine-treated patients demonstrated a statistically significant (*P* £ .05) reduction in the Anxiety/Agitation factor compared with placebo-treated patients at weeks 1, 2, 4, 6, and endpoint. There was no statistically significant difference between mirtazapine-treated patients and amitriptyline-treated patients at weeks 1, 3, 4, 5, 6, and endpoint. Similar results were found for the Anxiety/Somatization factor. Mirtazapine-treated patients demonstrated statistically significant (*P* £ .03) greater reductions in Anxiety/Somatization throughout weeks 1 through 6 and were comparable to amitriptyline-treated patients during this same time period.

Major Depressive Disorder With Cognitive Disturbances The effect of mirtazapine on cognitive disturbances in patients with major depressive disorder was evaluated using HAM-D factor III in several studies. In one study mirtazapine-treated patients and amitriptyline-treated patients demonstrated significantly greater reductions from baseline than placebo-treated patients. In a placebo-controlled study, mirtazapine-treated patients demonstrated statistically significant between-treatment differences compared with placebo for cognitive disturbances. These differences were seen at weeks 2 through 4 and at endpoint. Lastly, both mirtazapine (20 to 80 mg a day) and clomipramine (50 to 200 mg a day) were evaluated for their effects on cognitive disturbances; both demonstrated similar improvements in the HAM-D cognitive factor.

ANXIETY DISORDERS

Mirtazapine's potential value for other indications has been studied. The efficacy of 15 to 25 mg of mirtazapine a day was evaluated in 40 patients with a primary diagnosis of anxiety states, according to the World Health Organization's ninth revision of the *International Statistical Classification of Diseases* (ICD-9). Anxiety symptoms were evaluated using the Hamilton Anxiety Scale, Zung Anxiety Scale, and the Global Assessment Scale. Mirtazapine-treated patients experienced significantly greater improvements from baseline in overall anxiety symptoms, psychic anxiety, and global functioning than placebo-treated patients. Further studies are required before mirtazapine can be recommended as primary therapy for generalized anxiety disorder.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System Mirtazapine is a potent antagonist of histamine H_1 receptors, which may explain its somnolence-inducing effects. In placebo-controlled studies in the United States, somnolence was reported in 54 percent of mirtazapine-treated patients, 18 percent of placebo-treated patients, and 60 percent of amitriptyline-treated patients. Somnolence resulted in the discontinuation of therapy in 10.4 percent of all mirtazapine-treated patients and 2.2 percent of all placebo-treated patients in the U.S. controlled studies. Somnolence appears to be transient and to wear off within a few days of starting treatment. In one study, sedation generally appeared during the first week of treatment and then decreased in intensity during further dose titrations. In another 6-week study, reported somnolence typically occurred during the first week of therapy, gradually diminished, and disappeared by the end of week 4. In European placebo-controlled studies, the incidence of somnolence was 14.8 percent in mirtazapine-treated patients and 10.3 percent in placebo-treated patients. The lower incidence of somnolence in placebo-controlled European studies might be attributable to higher starting dosages (15 to 20 mg a day) than those in U.S. studies (5 to 10 mg a day). Mirtazapine's somnolence-inducing effects at higher dosages appear to be less pronounced than anticipated in relation to its potent antihistaminergic properties, probably because of the activating effects of increased noradrenergic neurotransmission that occurs within the recommended dose range (15 to 45 mg a day).

Dizziness was reported in 7 percent of mirtazapine-treated patients, 3 percent of placebo-treated patients, and 14 percent of amitriptyline-treated patients in the United States controlled studies.

Cardiovascular System Mirtazapine has little or no effect on heart rate or blood pressure. In one review individual adverse effects were assigned to different symptom cluster categories (sedation, anticholinergic effects, heart rhythm effects, hypotension/vertigo, typical selective serotonin reuptake inhibitor [SSRI]-related effects, and sexual dysfunction). The percentages of patients with at least one symptom from each of the symptom clusters were calculated on a treatment group basis. There were no differences in the incidences of heart rhythm cluster (arrhythmia, palpitations, tachycardia) between mirtazapine-treated patients (7 percent) and placebo-treated patients (10 percent). The incidence of heart rhythm cluster was 16 percent for amitriptyline, which was statistically significantly higher than in either mirtazapine-treated or placebo-treated patients. The incidence of the hypotension-vertigo symptom cluster (dizziness, postural hypotension, faintness, hypotension, orthostatic hypotension, lightheadedness, syncope, postural syncope, vertigo) was 5 percent for both mirtazapine-treated and placebo-treated patients and 11 percent for amitriptyline-treated patients, which was statistically significant versus both.

Gastrointestinal System Increased appetite and weight gain have been reported more frequently with mirtazapine use than with placebo or amitriptyline. Appetite increase was reported in 17 percent of mirtazapine-treated patients, 2 percent of placebo-treated patients, and 6 percent of amitriptyline-treated patients in controlled studies in the United States. A weight gain of 7 percent or more of body weight was reported in 7.5 percent of mirtazapine-treated patients, 0 percent of placebo-treated patients, and 5.9 percent of amitriptyline-treated patients.

Since 5-HT_3 antagonists have been shown to reduce nausea and other gastrointestinal symptoms, the low incidence of nausea and other gastrointestinal symptoms in mirtazapine-treated patients may be attributable to its blockade of postsynaptic 5-HT_3 receptors. There have been three case reports of patients receiving mirtazapine to alleviate SSRI-related nausea. Mirtazapine (15 mg) was added to the SSRI regimen, and the nausea completely resolved. In each instance, the nausea returned upon discontinuation of mirtazapine. Mirtazapine administration was restarted, and the nausea completely resolved once more.

THERAPEUTIC INDICATIONS

Mirtazapine is currently indicated for the treatment of depression. Its efficacy in the treatment of depression was established in 5- or 6-week randomized, double-blind, placebo-controlled clinical trials. The study populations consisted of inpatients and outpatients whose diagnoses corresponded most closely to the category of major depression as defined by the third editional DSM (DSM-III), with a minimum baseline score of 18 or above on the first 17 items of the HAM-D scale.

Mirtazapine's potential value in treating depressed patients with either sleep or anxiety disturbances has also been studied. Mirtazapine appears to be significantly better than placebo and similar to amitriptyline or trazodone in treating sleep and anxiety disturbances and may be of significant benefit for patients with major depression with symptoms of Anxiety/Agitation or Anxiety/Somatization.

PRECAUTIONS AND ADVERSE REACTIONS

Overall, mirtazapine appears to be well tolerated. Approximately 16 percent of mirtazapine-treated patients (*N* = 453) and 7 percent of placebo-treated patients (*N* = 361) discontinued treatment in U.S. 6-week clinical trials because of adverse effects. The most commonly reported adverse effects for mirtazapine-treated patients included somnolence (10.4 percent) and nausea (1.5 percent).

Mirtazapine's favorable adverse-effect profile may be a result of its pharmacological effects. Its ability to selectively block 5-HT_2 and 5-HT_3 receptors may account for a low incidence of such symptoms as anxiety, insomnia, nausea, and sexual dysfunction. Headache was reported at a significantly (*P* £ .05) higher incidence in placebo-treated patients. The typical SSRI symptom cluster (e.g., nausea, vomiting, diarrhea, headache, agitation, and insomnia) was reported significantly (*P* £ .05)

less frequently in mirtazapine-treated patients than in placebo-treated patients.

Blockade of 5-HT_2 receptors may protect against acute akathesia but may result in slowly developing restless leg syndrome similar to that induced by mianserin. There have been two case reports of restless leg syndrome occurring after 5 to 6 weeks of mirtazapine treatment. In each case, the symptoms of restless legs disappeared after discontinuation of mirtazapine therapy. In one case the patient was rechallenged with mirtazapine and again developed restless leg syndrome. Mirtazapine was discontinued and the symptoms disappeared.

Mirtazapine has little affinity for D_1 and D_3 receptors and moderate affinity for muscarinic and cholinergic receptors, which may explain its low incidence of dopaminergic and anticholinergic adverse effects.

Mirtazapine has not been associated with sexual dysfunction in controlled clinical trials. In U.S. clinical trials, impotence was reported infrequently (1 of 100 to 1 of 1000) and abnormal ejaculation was reported rarely (<1 of 1000). Decreased libido was reported at a similar low incidence in mirtazapine- and placebo-treated males (4 percent and 7 percent, respectively) and female patients (4 percent for both) but occurred at a higher rate in males treated with amitriptyline (14 percent). Interpretation of these data is difficult since reports of decreased libido may be associated with either depressive illness or treatment.

Drowsiness, excessive sedation, dry mouth, increased appetite, and weight gain appear to occur more frequently with mirtazapine than with placebo. These complaints were typically mild and transient, did not require dosage adjustments, and may improve with continued use.

Anticholinergic symptoms (dry mouth, constipation, taste perversion, and visual disturbances), cardiovascular symptoms (palpitations and tachycardia), and neurological symptoms (tremor and vertigo) occurred more frequently with amitriptyline than with mirtazapine.

Mirtazapine has little or no effect on heart rate, blood pressure, or body temperature.

Rare adverse reactions such as mania or seizures have been reported in a few mirtazapine-treated patients. Hypomanic or manic switch occurred in 3 of 1299 (0.2 percent) of patients treated with mirtazapine in U.S. studies. In premarketing trials, only one seizure was reported among 2796 patients receiving mirtazapine. No controlled trials have been conducted in patients with a history of seizure, and caution should be exercised when mirtazapine is used in these patients.

In premarketing clinical trials, 2 of 2796 patients receiving mirtazapine developed severe symptomatic neutropenia (agranulocytosis) and a third patient developed severe neutropenia; this hematological disorder was reversible in all three patients. One of the patients who developed severe symptomatic neutropenia had Sjögren's syndrome, an autoimmune disease frequently associated with hematological disorders, including neutropenia. The other patient was concomitantly taking ibuprofen (Motrin) and aspirin, both of which can cause agranulocytosis. There have been five reported cases of severe white blood cell (WBC) disorders in mirtazapine-treated patients worldwide from September 1994 through May 1998. The precise cause of the WBC disorders in these cases is unknown because of the possible association of agranulocytosis in each case with known risk factors.

Data regarding the effects of overdose with mirtazapine are accumulating, demonstrating a benign profile. Mirtazapine appears to be safe in a limited number of overdose cases, which sharply contrasts with tricyclic drug overdose. Six patients took a substantial overdose of mirtazapine in multiples of 10 to 30 times the maximum recommended daily dose. The characteristic symptoms were transient without any clinically relevant changes in vital signs or electrocardiogram (ECG) readings. ECG abnormalities have been observed in 70 to 90 percent of patients overdosing on tricyclic drugs, which differs significantly from mirtazapine. Mirtazapine appears to be safe in overdose of substantial multiples of the recommended daily dose.

Clinically significant but reversible elevations (³3 times the upper limit of normal) in alanine aminotransferase (ALT) were noted in 2 percent (8 of 424) of mirtazapine-treated patients in a pool of short-term U.S. controlled trials, compared with 0.3 percent (1 of 328) of placebo-treated patients and 2 percent (3 of 181) of amitriptyline-treated patients. ALT concentrations returned to normal both in patients who discontinued mirtazapine and in those who continued mirtazapine. The incidence of clinically significant elevations in ALT, aspartate transferase (AST), and g-glutamyltransferase was similar in mirtazapine- and placebo-treatment groups (0.10, 0.12, and 0.072, respectively) in terms of patient exposure years. There have also been reports of elevated nonfasting cholesterol levels and nonfasting triglyceride levels in patients receiving mirtazapine. However, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) ratio data collected at one site in an uncontrolled study of 49 nonfasting patients showed that lipid metabolism did not change significantly from visit to visit, suggesting that the changes from baseline may be related to increased dietary intake associated with increased appetite.

There is currently no experience with mirtazapine in pregnant women. Women who plan to become pregnant or who are pregnant should be informed that the risk to the fetus is unclear and that the drug should only be used if clearly needed. It is not known whether mirtazapine is excreted into human breast milk.

DRUG INTERACTIONS

Formal studies have been performed on pharmacodynamic drug interactions between mirtazapine and other drugs commonly used in patients with depression. A study in healthy volunteers showed that mirtazapine can have additive subjective and objective effects on cognitive and motor performances when given with either alcohol or diazepam (valium). Caution should be used when mirtazapine is combined with other central nervous system–depressant drugs, and patients should be warned about the concomitant use of alcohol.

Mirtazapine has minimal inhibitory effects on CYP 2D6, CYP 1A2, and CYP 3A4 in vitro. The lack of cytochrome P450 inhibition suggests that mirtazapine is not likely to have a clinically significant inhibitory effect on the metabolism of any drugs that are substrates for these cytochrome P450 enzymes. However, further research is needed before any definitive conclusions can be made about the risks associated with coadministration of mirtazapine with other drugs metabolized by the cytochrome P450 subfamilies.

Although there have been no reports of an interaction between mirtazapine and the monamine oxidase inhibitors (MAOIs), it is recommended that mirtazapine not be used in combination with an MAOI or within 14 days of initiating or discontinuing MAOI therapy.

LABORATORY INTERFERENCES

There are currently no data that indicate that mirtazapine interferes with urinalysis, spectrometry, or other laboratory tests.

DOSAGE AND ADMINISTRATION

The average adult starting dosage of mirtazapine is 15 mg a day administered as a single oral dose, preferably at bedtime. Controlled clinical trials have established an effective dosage range of 15 to 45 mg a day. The pharmacokinetics of mirtazapine indicate that once-daily dosing is sufficient and that dose increases can be made at 4- to 5-day intervals. Elderly patients or those with renal or hepatic disease may develop higher serum concentrations than younger patients; no specific dosage reduction appears necessary, but dosage should be increased with caution.

Mirtazapine is available as 15-mg, oval, scored, yellow tablets and 30-mg oval, scored, red-brown tablets.

SUGGESTED CROSS-REFERENCES

The biochemical aspects of mood disorders are reviewed in Section 14.3 and their clinical features are reviewed in Section 14.6. Other pharmacological treatments of depressive disorders appear in Section 14.7.

SECTION REFERENCES

Bremner JD, Wingard P, Walshe TA: Safety of mirtazapine in overdose. J Clin Psychiatry *59:*233, 1998.

*Carpenter LL, Jocic Z, Hall JM, Rasmussen SA, Price LH: Mirtazapine augmentation in the treatment of refractory depression. J Clin Psychiatry *60:*45, 1999.

*Claghorn JL, Lesem MD: A double-blind placebo-controlled study of org 3770 in depressed outpatients. J Affect Disord *34:*165, 1995.

Cohen M, Panagides J, Timmer CJ, Huisman JAM: Pharmacokinetics of mirtazapine from orally administered tablets: Influence of a high-fat meal. Eur J Drug Metab Pharmacokin *22:*103, 1997.

*Connor KM, Davidson JRT, Weisler RH, Ahearn E: A pilot study of mirtazapine in posttraumatic stress disorder. Int Clin Psychopharm *14:*29, 1999.

Dahl ML, Voortman G, Alm C, Elwin CE, Delbressine L, Vos R, Bogaards JJP, Bertilsson L: In vitro and in vivo studies on the disposition of mirtazapine in humans. Clin Drug Invest *13:*37, 1997.

de Boer T: The pharmacologic profile of mirtazapine. J Clin Psychiatry *57:*19, 1996.

Delbressine LPC, Moonen MEG, Kaspersen FM, Wagenaars GN, Jacobs PL, Timmer CJ, Paanakker JE, van Hal HJM, Voortman G: Pharmacokinetics and biotransformation of mirtazapine in human volunteers. Clin Drug Invest *15:*45, 1998.

Delbressine LPC, Vos RME: The clinical relevance of preclinical data: Mirtazapine, a model compound. J Clin Psychopharmacol *17:*29S, 1997.

Fawcett J, Barkin RL: A meta-analysis of eight randomized, double-blind, controlled clinical trials of mirtazapine for the treatment of patients with major depression and symptoms of anxiety. J Clin Psychiatry *59:*123, 1998.

*Halikas JA: Org 3770 (mirtazapine) versus trazodone: A placebo controlled trial in depressed elderly patients. Hum Psychopharmacol *10:*S125, 1995.

Heimpel H: Drug-induced agranulocytosis. Med Toxicol *3:*165, 1988.

Hoes MJ, Zeijpveld JH: First report of mirtazapine overdose. Int J Clin Psychopharmacol *11:*147, 1996.

Hoyberg OJ, Maragakis B, Mullin J, Norum D, Stordall E, Ekdahl P, Ose E, Moksnes KM, Sennef C: A double-blind multicentre comparison of mirtazapine and amitriptyline in elderly depressed patients. Acta Psychiatr Scand *93:*184, 1996.

Kasper S, Zivkov M, Roes KCB, Pols AG: Pharmacological treatment of severely depressed patients: A meta-analysis comparing efficacy of mirtazapine and amitriptyline. Eur Neuropsychopharmacol *7:*115, 1997.

Kasper S: Clinical efficacy of mirtazapine: A review of meta-analyses of pooled data. Int Clin Psychopharmacol *10:*25, 1995.

*Khan MC: A randomized, double-blind, placebo-controlled, 5-weeks' study of org 3770 (mirtazapine) in major depression. Hum Psychopharmacol *10:*S119, 1995.

Kuitunen T: Drug and ethanol effects on the clinical test for drunkenness: Single doses of ethanol, hypnotic drugs, and antidepressant drugs. Pharmacol Toxicol *75:*91, 1994.

Markkula J, Lauerma H: Mirtazapine-induced restless legs. Hum Psychopharmacol *12:*497, 1997.

Mattila M, Jaaskelainen J, Jarvi R, Romanov M, Miettinen E, Sorri P, Ahlfors U, Zivkov M: A double-blind study comparing the efficacy and tolerability of mirtazapine and doxepin in patients with major depression. Eur Neuropsychopharmacol *5:*441, 1995.

Mattila M, Mattila MJ, Vrijmoed-de Vries M, Kuitunen T: Actions and interactions of psychotropic drugs on human performance and mood: Single doses of org 3770, amitriptyline, and diazepam. Pharmacol Toxicol *65:*81, 1989.

Montgomery SA: Safety of mirtazapine: A review. Int Clin Psychopharmacol *10:*37, 1995.

Montgomery SA, Reimitz PE, Zivkov M: Mirtazapine versus amitriptyline in the long-term treatment of depression: A double-blind placebo-controlled study. Int Clin Psychopharmacol *13:*63, 1998.

Moutsopoulos HM: Sjögren's syndrome. In *Harrison's Principles of Internal Medicine*, ed 13, KJ Isselbacher, E Braunwald, JD Wilson, editors. McGraw Hill, New York, 1994.

Mullin J, Lodge A, Bennie E, McCreadie R, Bhatt GS, Fenton G: A multicentre, double-blind, amitriptyline-controlled study of mirtazapine in patients with major depression. J Psychopharmacol *10:*235, 1996.

Nelson JC: Safety and tolerability of the new antidepressants. J Clin Psychiatry *58:*26, 1997.

Nutt DJ: Efficacy of mirtazapine in clinically relevant subgroups of depressed patients. Depression Anxiety *7:*125, 1998.

Nutt D: Mirtazapine: Pharmacology in relation to adverse effects. Acta Psychiatr Scand *96:*31, 1997.

Owen JR, Nemeroff CB: New antidepressants and the cytochrome P450 system: Focus on venlafaxine, nefazodone, and mirtazapine. Depression Anxiety *7:*24, 1998.

Pedersen L, Klysner R: Antagonism of selective serotonin reuptake inhibitor-induced nausea by mirtazapine. Int Clin Psychopharmacol *12:*59, 1997.

Richou H, Ruimy P, Charbaut J, Delisle JP, Brunner H, Patris M, Zivkov M: A multicentre, double-blind, clomipramine-controlled efficacy and safety study of org 3770. Hum Psychopharmacol *10:*263, 1995.

Ruigt GSF, Kemp B, Groenhout CM, Kamphuisen HAC: Effect of the antidepressant org 3770 on human sleep. Eur J Clin Pharmacol *38:*551, 1990.

Sitsen JMA, Moors J: Mirtazapine, a novel antidepressant, in the treatment of anxiety symptoms results from a placebo-controlled trial. Drug Invest *8:*339, 1994.

*Smith WT, Glaudin V, Panagides J, Gilvary E: Mirtazapine vs amitriptyline vs placebo in the treatment of major depressive disorder. Psychopharmacol Bull *26:*191, 1990.

Stahl S, Zivkov M, Reimitz PE, Panagides J, Hoff W: Meta-analysis of randomized, double-blind, placebo-controlled, efficacy and safety studies of mirtazapine versus amitriptyline in major depression. Acta Psychiatr Scand *96:*22, 1997.

Stimmel GL, Sussman N, Wingard P: Mirtazapine safety and tolerability: Analysis of the clinical trials database. Prim Psychiatry *82:*1997.

Timmer CJ, Lohmann AAM, Mink CPA: Pharmacokinetic dose-proportionality study at steady state of mirtazapine from Remeron ™ tablets. Hum Psychopharmacol *10:*S97, 1995.

van Moffaert M, de Wilde J, Vereecken A, Dierick M, Evrard JL, Wilmotte J, Mendlewicz J: Mirtazapine is more effective than trazodone: A double-blind controlled study in hospitalized patients with major depression. Int Clin Psychopharmacol *10:*3, 1995.

Voortman G, Paanakker JE: Bioavailability of mirtazapine from Remeron ™ tablets after single and multiple oral dosing. Hum Psychopharmacol *10:*S83, 1995.

*Wheatley DP, van Moffaert M, Timmerman L, Kremer CME, and the Mirtazapine-Fluoxetine Study Group: Mirtazapine: Efficacy and tolerability in comparison with fluoxetine in patients with moderate to severe major depressive disorder. J Clin Psychiatry *59:*306, 1998.

Zivkov M, Roes KCB, Pols AG: Efficacy of org 3770 (mirtazapine) vs amitriptyline in patients with major depressive disorder: A meta-analysis. Hum Psychopharmacol *10:*S135, 1995.

31.20 MONOAMINE OXIDASE INHIBITORS

SIDNEY H. KENNEDY, M.D., KEVIN F. MCKENNA, M.D., ıH.D. AND GLEN B. BAKER, PH.D.

HISTORY

It is now 40 years since the introduction of monoamine oxidase inhibitor (MAOI) antidepressants. Using leftover hydrazine V-2 rocket fuel from World War II, scientists developed isonicotinic acid 2-isopropylhydrazine (iproniazid) as an antituberculous agent. Following reports of unexpected euphoria among tuberculous patients, investigators evaluated this “psychic energizer” in depressed psychiatric patients, and in 1958 one author wrote: “the best news for psychiatry in 1957 was the discovery that iproniazid protects serotonin from monoamine oxidase. Serotonin, allowed free activity in the brain is perhaps the most energetic releaser of reserve power in the human machine.”

The MAOI honeymoon proved to be relatively brief. Iproniazid was ultimately withdrawn because of its toxic effects on the liver; a fatal hypertensive reaction to its less hepatotoxic nonhydrazine successor, tranylcypromine (Parnate), was attributed to the drug's interaction with tyramine in mature cheese, “the cheese reaction,” and several multicenter trials involving phenelzine (Nardil) failed to demonstrate its clinical efficacy.

Nevertheless, a small but convinced group of psychopharmacologists in both the United States and the United Kingdom continued to evaluate the safety and efficacy of tranylcypromine, phenelzine, and isocarboxazid (Marplan) as the three surviving representatives of what have come to be known as the classic MAOIs. Several developments are responsible for renewed interest in the field of MAOI therapies:

1. Evidence of a preferential clinical response to MAOIs in patients with atypical and anergic bipolar depressions
2. Reduced concerns about food and drug interactions
3. Availability of rapidly acting effective antidotes such as the calcium channel inhibitor nifedipine (Adalat, Procardia) to treat hypertensive reactions
4. Discovery that monoamine oxidase (MAO) has at least two isoenzyme forms (A and B) that metabolize different neurotransmitters; serotonin and norepinephrine are preferred substrates for the A form, dopamine is a substrate for both A and B forms, and phenylethylamine is metabolized by the B-isoenzyme
5. The development of selective A and B inhibitor drugs that can be prescribed without dietary restrictions. As well as selectivity, new members of the MAOI drug class are also reversible inhibitors of MAO, which has important clinical implications. Instead of a further 2-week washout after the discontinuation of phenelzine, tranylcypromine, or isocarboxazid to allow the body to replace irreversibly bound MAO with a new supply of active enzyme, reversible inhibitors such as moclobemide (Manerix) leave no lasting effects on MAO beyond 48 hours.

CHEMISTRY

In 1997, only irreversible MAOIs had Food and Drug Administration (FDA) approval in the United States. In addition to phenelzine and tranylcypromine, selegiline (Eldepryl) had approval for the treatment of Parkinson's disease. Pargyline was approved for treatment of severe hypertension but is no longer manufactured. Although, manufacturing of the antidepressant isocarboxazid was discontinued in 1990 in the United States and 1992 in Canada, it was relaunched in the United States in 1998. Moclobemide, a reversible inhibitor of MAO type A (MAO $_A$) (RIMA) has been available in Europe, Canada, Australia, and elsewhere since the early 1990s but is not available in the United States. Brofaromine, also a RIMA with demonstrated antidepressant efficacy, is not currently under active investigation. However, a new RIMA compound, befloxatone, is undergoing clinical trial evaluation in the United States and Europe.

The MAOIs may be classified according to chemical structure (hydrazine versus nonhydrazine); by affinity for substrate site (reversible versus irreversible) or by degree of selectivity for the A or B isoenzyme (A or B inhibitor) (Table 31.20-1).

Generic Name	Brand Name or Company Code	Presently Marketed in North America	Reversible/Irreversible Selectivity
Iproniazid	Marsilid	No	Irreversible, nonselective
Isocarboxazid, INN	Marplan	Yes, as antitubercular	Irreversible, nonselective
Phenelzine	Nardil	Yes	Irreversible, nonselective
Isocarboxazid	Marplan	Yes*	Irreversible, nonselective
Tranylcypromine	Parnate	Yes	Irreversible, nonselective
Clorgyline	—	No	Irreversible, MAO_A selective
Pargyline	Eutonyl	Recently withdrawn as an antihypertensive	Irreversible, MAO_B selective
Selegiline	Eldepryl, (–)-Deprenyl	Yes	Irreversible, MAO_B selective
Moclobemide	Aurorix, Manerix	Not in U.S. but available in Canada	Reversible, MAO_A selective
Brofaromine	Consonar	No	Reversible, MAO_A selective
Lazabemide	Ro 19-6327	No	Reversible, MAO_B selective
Befloxatone	—	No	Irreversible, MAO_B selective
Milacemide	—	No	Partially reversible, relatively nonselective
Toloxatone	—	No	Reversible, MAO_A selective
Befloxatone	—	No	Reversible, MAO_A selective

* Isocarboxazid was reintroduced for clinical use in the United States in 1998.

Table 31.20-1 Some MAO Inhibitors and Their Actions

Hydrazine Status Both phenelzine and isocarboxazid are hydrazine derivatives, with a nitrogen-nitrogen bond in their side chains. The first generation of hydrazine compounds (iproniazid and isoniazid) was withdrawn because of hepatotoxicity. Tranylcypromine and selegiline represent nonhydrazine arylalkylamines. These compounds are structurally related to amphetamines, and in a single case report of tranylcypromine overdose, detectable levels of amphetamine and its metabolites were found, although this finding has not been replicated. Some have suggested that this may explain why tranylcypromine has a more rapid therapeutic effect than nonhydrazines. However, euphoric effects of amphetamine levels are not detected at therapeutic doses.

Selectivity and Reversibility Pargyline was the first selective MAOI to receive FDA approval. Although it proved selective for the B isoenzyme and did not require dietary restriction of tyramine intake (the gastrointestinal tract contains predominantly MAO $_A$), it was not an effective antidepressant. Selegiline is also a B inhibitor at low doses, but even at therapeutic doses in antidepressant trials, the selectivity is diminished. Despite their selectivity, both pargyline and selegiline represent examples of irreversible selective inhibitors, in contrast to moclobemide with both reversibility and selectivity (Fig. 31.20-1).

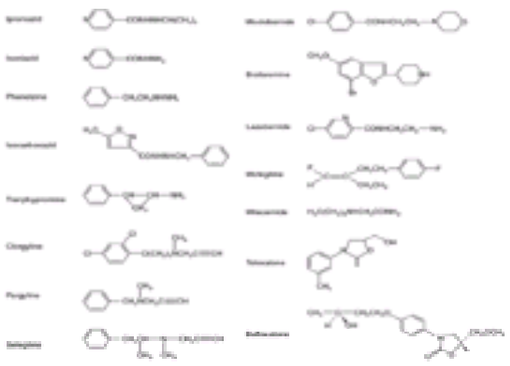

FIGURE 31.20-1 Molecular structures of some MAOI.

PHARMACOLOGICAL ACTIONS

Characteristics of the MAO Enzyme System The MAOs are flavoproteins found on the outer membranes of mitochondria. They catalyze the oxidative deamination of a variety of alkylamines, resulting in the formation of hydrogen peroxide and an aldehyde derivative of the amine (Fig. 31.20-2). The aldehyde is then further reduced or oxidized to the corresponding alcohol or carboxylic acid, respectively.

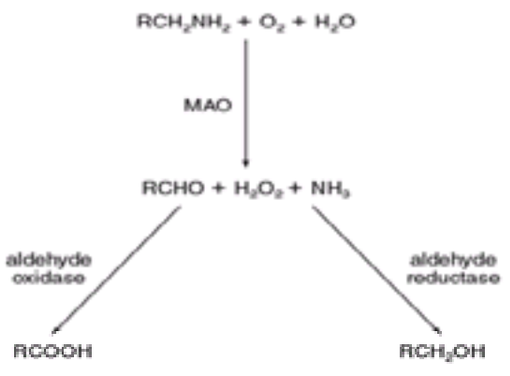

FIGURE 31.20-2 The reaction catalyzed by MAO.

Free cytoplasmic and extraneuronal neurotransmitter amines would be most susceptible to MAO metabolism. The presence of high concentrations of MAO in the blood-gut and blood-brain barriers supports the idea that MAO serves a protective or detoxifying role.

Irreversible inhibitors attach to the flavin–adenine dinucleotide group, decreasing breakdown of the substrate neurotransmitters and denaturing the original MAO, which results in a deficiency of MAO to denature dietary tyramine or amine medications. In contrast, the reversible inhibitors such as moclobemide have a direct inhibitory action on MAO_A and may be competitively displaced by tyramine, substantially reducing the likelihood of food and drug interactions.

The enzyme is expressed widely among eukaryotic organisms and in all mammals studied to date. Many cells express both forms of MAO, in differing proportions, but some tissues contain only one form, which has expedited purification of the enzymes. In human tissue, MAO_A is found in the placenta while MAO_B is found in platelets. As platelets are relatively easy to procure, the vast majority of human research on MAO has examined MAO_B. The human brain expresses both enzymes, but MAO_B predominates (80 to 95 percent), while in the rat brain MAO_A predominates over MAO_B. Immunohistochemical studies using polyclonal or monoclonal antibodies in postmortem human brain have demonstrated that serotoninergic neurons contain predominantly MAO_B, while catecholaminergic neurons contain MAO_A. Both forms of MAO are present in glial and endothelial cells. These findings suggest that the two forms of MAO are independently regulated and perform different functions. The presence of MAO_B in serotonergic neurons is surprising, since serotonin is preferentially deaminated by MAO_A; this finding suggests that MAO_B may function to degrade low levels of dietary and endogenous amines such as phenylethylamine or tyramine, with little effect on serotonin unless concentrations of this amine are elevated. Under normal conditions the classic neurotransmitters metabolized by MAO (norepinephine, dopamine, and serotonin) are preferentially stored in vesicles where they are not exposed to MAO.

Homology between the amino acid sequences of A and B isoforms is 70 percent and they are encoded by separate genes located on the short arm of the X chromosome. During development, the MAO_A form is expressed first, followed by MAO_B, which increases in proportion through life. Complete absence of MAO and its corresponding messenger ribonucleic acid (mRNA) has been described in a family of male cousins with a submicroscopic deletion of Xp21-p11, an area that includes the Norrie disease gene. Norrie disease is a rare X-linked recessive neurological disorder characterized by retinal dysplasia with blindness, mental retardation, and progressive hearing loss. Clinical features in these patients that may be related to complete absence of MAO include somatic growth failure, abnormal sexual maturation, autonomic nervous system dysfunction with hypotension, sleep disturbances with a marked reduction in the amount of rapid eye movement (REM) sleep, flushing, atonic seizures, motoric hyperactivity, and hyperreflexia. Large reductions in the urinary excretion of the norepinephrine metabolites, 3-methoxy-4-hydroxyphenyl glycol (MHPG) and vanillylmandelic acid (VMA), as well as the dopamine metabolite, homovanillic acid (HVA), were observed in these individuals. There was also a marked increase in the urinary excretion of phenylethylamine (PEA), a substrate of MAO-B (Table 31.20-2). The neurochemical changes are dramatic and much more pervasive than those following drug-induced inhibition of MAO.

MAO_A	Mixed	MAO_B
Epinephrine	*m,p*-Tyramine	β-Phenylethylamine
Norepinephrine	Dopamine	Phenylethanolamine
Metanephrine	Octopamine	*o*-Tyramine
Serotonin (5-HT)	Synephrine	Benzylamine
Sumatriptan	Tryptamine	1-Methyl-4-phenyl-1,2,3,6-tetrahydropyridine
	N-methyltryptamine	
	N,N-dimethyltryptamine	

Table 31.20-2 Some Common Substrates of MAO

Numerous attempts have been made to correlate hereditary variations in MAO activity with human disease. Wide interindividual variations in platelet and fibroblast MAO activity, as much as 50-fold, have been described in studies of normal volunteers and those with psychiatric disorders. Low MAO_B activity has been associated with stimulus-seeking and suicidal behavior, alcoholism, and bipolar disorder. MAO_B activity increases with age in animals, elderly healthy humans, and (to a greater extent) patients with neurodegenerative disease. The increased activity of MAO_B with age may reflect the glial cell proliferation that is linked to neuronal loss during

aging, as the concentration of the enzyme is thought to be higher in glial than in neuronal cells.

Platelet MAO Patients with more than 80 percent inhibition of platelet MAO after 2 weeks of treatment were reported to have a better antidepressant response than patients with less than 80 percent inhibition. However, other studies have not demonstrated a significant relation between clinical response and percentage of platelet MAO inhibition with phenelzine, tranylcypromine, or selegiline. Platelet MAO, of course, is exclusively MAO_B and does not necessarily reflect the central effect of the drugs on MAO_A. Also no role exists for platelet MAO inhibition measurement with RIMAs.

Effects of MAOIs on Brain Amine and Acid Levels The inhibition of MAO by drugs such as phenelzine and tranylcypromine results in an often dramatic increase in the concentration of a number of brain amines termed “trace amines,” for example, PEA, *meta*- and *para*-tyramine, and octopamine. In addition, PEA is a metabolite of phenelzine. These trace amines can markedly affect uptake, release, or both of catecholamines and serotonin at nerve endings. They may also act as neuromodulators through direct actions on receptors for the catecholamines, serotonin, or both.

Several groups of researchers have noted that administering phenelzine to rats increases brain concentrations of g-aminobutyric acid (GABA). Long- and short-term studies suggest that this elevation results, at least in part, from an inhibition of the catabolic enzyme GABA-transaminase. Increased brain concentrations of the amino acid alanine and inhibition of alanine transaminase have also been observed. Increasing evidence indicates that GABA is involved in the action of antipanic drugs, and phenelzine is commonly used to treat panic disorder. Although the N^2-acetyl analogue of phenelzine is also an effective MAOI, it neither alters GABA levels nor produces anxiolytic effects in the rat. Phenelzine's inhibition of GABA transaminase and elevation of GABA can be dramatically reduced by pretreating rats with another MAOI. This suggests that a metabolite produced by the action of MAO on phenelzine is ultimately responsible for the increase in brain GABA concentration. In theory this would make phenelzine a preferred antidepressant in the treatment of depressed patients with seizure disorders.

The structures of phenelzine and tranylcypromine are similar to those of PEA and amphetamine, and not surprisingly, they affect the uptake, release, or both of dopamine, norepinephrine, and, to a lesser extent, serotonin. At therapeutically equivalent doses of phenelzine and tranylcypromine in animal studies, concentrations of these drugs in the brain high enough to affect the uptake and release of these neurotransmitters could be attained. Several reports indicate that high doses of tranylcypromine (1.3 to 3.0 mg/kg a day) are effective in treating patients who suffer from refractory depression. Since these doses are well above those reported to inhibit MAO by more than 90 percent, effects of tranylcypromine other than MAO inhibition may contribute to its antidepressant effects at the high dosages. Free tranylcypromine concentration may also contribute to the adverse effects of this drug, including mean orthostatic drop of systolic blood pressure and increased pulse rate. Blood pressure elevations have also been significantly correlated with the dose of tranylcypromine; one group of investigators hypothesized that the initial hypertensive response to tranylcypromine is mediated by norepinephrine and that the orthostatic hypotensive effect is mediated by a direct interaction between tranylcypromine and a-adrenergic receptors. Of the RIMA compounds so far investigated, moclobemide appears to have few effects beyond MAO inhibition. In contrast, brofaromine is also a relatively potent inhibitor of serotonin reuptake, which may contribute to its therapeutic efficacy.

Changes in several presynaptic and postsynaptic receptors may follow the increase in concentrations of the amines, amino acid neurotransmitters, or both caused by the MAOIs. These delayed effects may be associated with the lag between administration of the MAOI and onset of antidepressant effect. Downregulation of b- and a_2 adrenergic, serotonin (5-hydroxytryptame [5-HT]) type 2 ($5\text{-}HT_2$), and tryptamine receptors occurs after long-term administration of phenelzine or tranylcypromine.

Phenelzine has been reported to inhibit tyrosine amino transaminase, aromatic amino acid decarboxylase, and dopamine b-hydroxylase, in addition to inhibiting GABA transaminase and alanine transaminase. Aromatic amino acid decarboxylase, GABA transaminase, and tyrosine amino transferase are pyridoxal-dependent enzymes, and phenelzine may deplete blood pyridoxal-5-phosphate concentrations in humans. In vitro studies of rats suggest that phenelzine increased phenylethanolamine *N*-methyltransferase and catechol *O*-methyltransferase activities above control values. Long-term administration of tranylcypromine is also reported to increase activity of aromatic amino acid decarboxylase. Reports also indicate that both phenelzine and tranylcypromine interact with enzymes involved in drug metabolism. Patients prescribed phenelzine or tranylcypromine may be taking other drugs concomitantly; thus metabolic drug-drug interactions may occur. MAOIs have been reported to inhibit the degradation of such drugs as hexobarbital, ethylmorphine, aminopyrine, meperidine (Demerol), and antipyrine (Auralgan). Recent studies indicate that tranylcypromine is a relatively potent inhibitor of cytochrome P450 (CYP) isoenzyme 2C19 (CYP 2C19).

Another factor that must be considered in the actions of tranylcypromine is the presence of chiral centers. Tranylcypromine is used clinically as the racemate but the (+) enantiomer is more potent than (–)-tranylcypromine in inhibiting MAO, whereas (–)-tranylcypromine is more effective than (+)-tranylcypromine in inhibiting uptake of catecholamines. The two enantiomers also differ in their interaction with $5\text{-}HT_1$ receptors in human postmortem frontal cortex, with (–)-tranylcypromine displaying a higher affinity than the enantiomer. (+) Results to date from studies on laboratory animals and humans indicate marked pharmacokinetic differences between the two enantiomers in brain and plasma.

METABOLISM

Phenelzine Although both phenelzine and tranylcypromine are absorbed rapidly after oral administration and have short elimination half-lives, much is still unknown about the metabolism of these two drugs and the contribution of metabolites to their overall pharmacological profiles, even though both drugs have been commercially available for many years. Numerous studies have been carried out on the relation between acetylator status of patients and response to treatment with phenelzine. These investigations were based on the assumption that phenelzine is acetylated because it is similar in structure to drugs such as isoniazid, which are known to be acetylated. In fact, the existence of *N*-acetyl phenelzine as a metabolite of phenelzine was adequately demonstrated only recently, and current indications are that it is only a minor metabolite. Phenelzine is an unusual drug in that it not only inhibits MAO but is also apparently a substrate for MAO. Studies with radiolabeled phenelzine suggest that phenylacetic acid and *p*-hydroxyphenylacetic acid are metabolites of phenelzine. *p*-Hydroxyphenylacetic acid is also of interest because it is a metabolite of the endogenous amine *p*-tyramine (*p*-hydroxy-b-phenylethylamine). b-PEA is also a known metabolite of phenelzine and evidence now indicates formation of *p*-hydroxyphenelzine from phenelzine. These observations raise the question of the route of formation of *p*-hydroxyphenylacetic acid. Possible routes are as follows: Phenelzine ® phenylacetic acid ® *p*-hydroxyphenylacetic acid; phenelzine ® PEA ® *p*-tyramine or phenylacetic acid or both ® *p*-hydroxyphenylacetic acid and phenelzine ® *p*-hydroxyphenelzine ® *p*-hydroxyphenylacetic acid (with or without *p*-tyramine intermediate). These routes have not been studied thoroughly, nor is there detailed information on the pharmacological activity of metabolites such as *p*-hydroxyphenelzine. Phenylethylidene hydrazine ($PhCH_2CH = N\text{-}NH_2$) or 1-(2-phenylethyl) diazene ($PhCH_2CH_2N = NH$) may be intermediates in going from phenelzine to phenylacetic acids, and hydrazine (H_2NNH_2) itself is a possible endproduct of phenelzine metabolism. Another possible route of metabolism of phenelzine is N-methylation. No information is available about involvement of cytochrome P450 isozymes in the various steps of PHE metabolism.

Tranylcypromine Although several popular textbooks mention that amphetamine is a metabolite of tranylcypromine, the bulk of evidence in the literature indicates that this is not the case. Despite a case report on the presence of amphetamine in the plasma of a patient who had overdosed on tranylcypromine, other studies conducted on humans and rats have not revealed amphetamine in human urine or rat brain after administration of pharmacologically relevant doses of tranylcypromine. The presence of the N-acetyl and ring hydroxylated metabolites of tranylcypromine (Fig. 31.20-3) have been demonstrated in rats and microbes. Little information is available about involvement of specific cytochrome P450 (CYP) isozymes in the metabolism of tranylcypromine, but tranylcypromine itself is a relatively potent inhibitor of CYP 2C19.

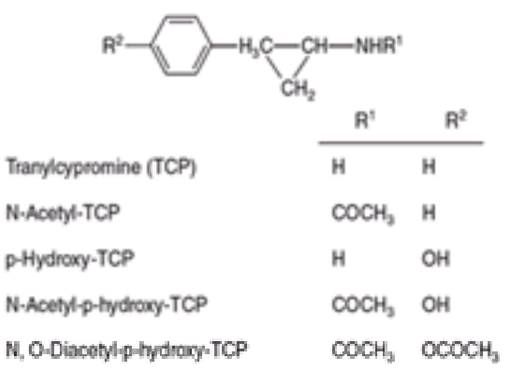

	R^1	R^2
Tranylcypromine (TCP)	H	H
N-Acetyl-TCP	$COCH_3$	H
p-Hydroxy-TCP	H	OH
N-Acetyl-p-hydroxy-TCP	$COCH_3$	OH
N, O-Diacetyl-p-hydroxy-TCP	$COCH_3$	$OCOCH_3$

FIGURE 31.20-3 Structures of identified metabolites of tranylcypromine.

Selegiline [(–)-Deprenyl] Selegiline is extensively metabolized to (–)-methamphetamine, (–)-amphetamine and (–)-*N*-desmethyldeprenyl (*N*-propargylamphetamine) (Fig. 31.20-4). Cytochrome P450 isozymes may be involved in these metabolic pathways. The relative contribution of these metabolites to the therapeutic actions and side effects of selegiline remains controversial.

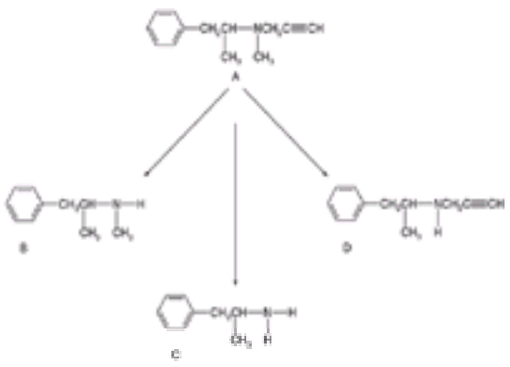

FIGURE 31.20-4 Principal metabolites of deprenyl (selegiline) **A, B**, methamphetamine; **C**, amphetamine; **D**, *N*-desmethylselegiline (*N*-propargylamphetamine).

Moclobemide and Brofaromine The two RIMAs moclobemide and brofaromine demonstrate similar T_{max} values, but brofaromine is more highly protein bound and has a longer elimination half-life than moclobemide. Both are extensively metabolized (Fig. 31.20-5 and Fig. 31.20-6). Little information is available on involvement of cytochrome P450 isozymes in these metabolic routes, although it has been reported that moclobemide is a substrate of CYP 2C19 and an inhibitor of CYP 2C19, CYP 2D6, and CYP 1A2 and that CYP 2D6 contributes to O-demethylation of brofaromine. It is suggested that the moclobemide dosage should be reduced by one-half when it is given in combination with the potent CYP 2D6 inhibitor, cimetidine (Tagamet). It has been suspected for some time, but not yet definitely shown, that an active metabolite may contribute to the MAOI activity of moclobemide.

FIGURE 31.20-5 Metabolism of brofaromine in humans. (Adapted from Waldmeier PC, Amrein R, Schmid-Burgk W: Pharmacology and pharmacokinetics of brofaromine and moclobemide in animals and humans. In *Clinical Advances in Monoamine Oxidase Inhibitor Therapies*, SH Kennedy, editor. American Psychiatric Press, Washington, DC, 1994.)

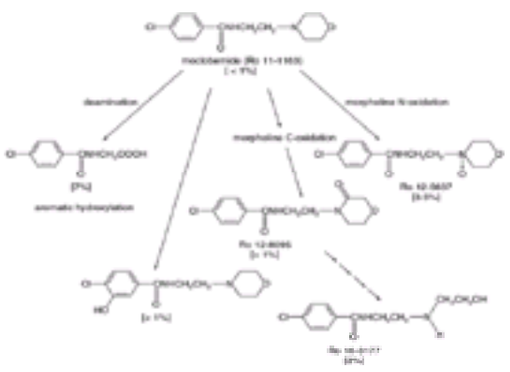

FIGURE 31.20-6 Metabolism of moclobemide in humans. (Adapted from Waldmeier PC, Amrein R, Schmid-Burgk W: Pharmacology and pharmacokinetics of brofaromine and moclobemide in animals and humans. In *Clinical Advances in Monoamine Oxidase Inhibitor Therapies*, SH Kennedy, editor. American Psychiatric Press, Washington, DC, 1994.)

THERAPEUTIC INDICATIONS

The MAOIs have found an important niche and are often the drugs of choice for atypical depression and depression associated with anxiety, panic, and phobias. However, they may be more useful in treating major depressive disorders than is generally realized, and tranylcypromine in daily doses of 100 mg or more is reportedly effective treatment for previously treatment-resistant depression.

The concept of “atypical depression” has been applied to different subtypes of depression; in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), it refers to a depression characterized by mood reactivity, extreme sensitivity to interpersonal loss or rejection, prominent anergia, and increased appetite, weight, and sleep. Previously, atypical depression was subclassified into anxious (type A) and reversed vegetative (type V) types. The Columbia University definition of atypical depression combined reversed vegetative symptoms with mood reactivity and intense rejection sensitivity. This group has demonstrated a better response to phenelzine than to imipramine (Tofranil) in patients with “Columbia University atypical depression.” Considerable overlap appears to exist between the rejection sensitivity predisposition and Axis II personality disorders, especially borderline personality disorder. Both phenelzine and tranylcypromine have been shown superior to comparative drugs in borderline personality disorder, and a relation between favorable response to tranylcypromine and history of childhood attention-deficit disorder was noted. To date little evidence suggests superior efficacy of RIMA agents such as moclobemide in atypical depression or borderline personality disorder. Preliminary reports suggest that selective serotonin reuptake inhibitors (SSRIs) may be as effective as phenelzine in atypical depression.

A 38-year-old divorced mother of two teenage daughters presented at the Mood Disorders Clinic with a history of failed antidepressant treatment trials. Two years earlier she had become increasingly irritable, exhausted, and tearful following the breakup of her marriage. She could not tolerate fluoxetine (Prozac) or sertraline (Zoloft) and discontinued each of these SSRIs after only a few days because of gastrointestinal symptoms. Although she could tolerate desipramine (Norpramin) up to 300 mg daily, she felt no benefit after 4 months and was reluctant to continue this tricyclic antidepressant medication because of adverse effects (weight gain and sweating in particular). Because of her increasing fatigue, increasing sleep needs, and generally “atypical” profile, she was offered a trial of phenelzine. After 4 weeks, at a dosage of 60 mg daily she reported significant improvement in her energy level, mood, and sleep pattern. Two years later she continues to take phenelzine 45 mg daily with significant adverse effects (occasional insomnia and anorgasmia).

A group at the University of Pittsburgh has emphasized the presence of type V (reversed vegetative) symptoms in what they describe as “anergic bipolar depression.” Although not replicated elsewhere, this group has consistently demonstrated superiority of tranylcypromine over imipramine in bipolar depression. The extent to which findings from the Columbia group are specific to phenelzine and those of the Pittsburgh group are specific to tranylcypromine remains unknown.

Many patients with atypical depression also suffer from panic attacks and anxiety symptoms. Some authors consider phenelzine the treatment of choice for panic disorder. Limited evidence of efficacy exists in other disorders that overlap with the depressive and anxiety disorders, including bulimia nervosa, anorexia nervosa, social phobia, and posttraumatic stress disorder.

MAOIs have also been used in the therapeutic management of various conditions including anginal pain, atypical facial pain, refractory thought disorders, neurodermatitis, treatment-resistant narcoleptic states, migraine, narcolepsy, attention-deficit disorder, and idiopathic orthostatic hypotension.

> A 33-year-old anesthesiology resident developed symptoms of major depression. He had always been a socially anxious and avoidant student with few close friends. He reluctantly tried a series of SSRI, tricyclic, and MAOI antidepressants, on each occasion claiming that adverse effects prevented him from continuing. He reluctantly agreed to try moclobemide because of its apparently favorable side-effect profile. He took moclobemide to a maximum dosage of 600 mg daily for 1 year, successfully completed his residency, and no longer takes antidepressant medication.

Moclobemide and brofaromine are effective therapeutic agents with more favorable side-effect profiles than the classical irreversible inhibitors. Both have been reported to be effective antidepressants, and recent reports indicate that they may also be useful in the treatment of panic disorder and social phobia. Selegiline, an irreversible selective MAO_B inhibitor is not a particularly effective antidepressant at doses at which only MAO_B is inhibited but has shown some efficacy in depressive disorders at higher dosages at which both MAO_B and MAO_A are inhibited. Selegiline has been used extensively, usually in combination with levodopa (Larodopa), for treatment of Parkinson's disease. The reasons for its effectiveness in this disorder are not entirely clear, but suggested mechanisms include reduced metabolism of dopamine reduced oxidative stress, increased expression of aromatic amino acid decarboxylase, and neurotrophic actions (by selegiline, its metabolites, or both). Because of selegiline's reported neuroprotective and neuronal rescue properties, it has been proposed that this drug may also be useful in treatment of other neurodegenerative disorders, such as Alzheimer's disease.

Continuation and maintenance treatments with MAOI and RIMA therapies have been understudied. Three published randomized, placebo-controlled trials exist involving long-term MAOI therapy, all with phenelzine, which proved superior to nortriptyline (Aventyl, Pamelor) in a maintenance trial in elderly depressive patients and superior to placebo and imipramine in adult depressed populations.

PRECAUTIONS AND ADVERSE REACTIONS

The most feared adverse effect of irreversible MAOIs is the tyramine-induced hypertensive crisis. Once it is decided to initiate a trial of an irreversible MAOI, the patient should be instructed to avoid foods with substantial tyramine content (other sympathomimetic amines such as PEA may also be culprits in this regard). Tyramine, a substrate of MAO present in significant concentrations in certain fermented foodstuffs including red wine, tap beer, cheese, yeast extracts, and pickled fish, causes a vasopressor response, an effect dramatically accentuated in patients taking an MAOI. This response is partly due to the MAO-inhibited patient's inability to deaminate tyramine (which is normally broken down by MAO_A in the gut), resulting in displacement of intracellular stores of norepinephrine. When dietary restrictions are followed, the risk of a serious hypertensive problem is curtailed. Although lists of foods that should be avoided by patients taking MAOIs are available, many of these are generally considered too extensive, and recent years have seen a trend to develop more user-friendly MAOI diets. Case reports exist of rare patients who developed idiosyncratic, spontaneous, hypertensive episodes shortly after taking initial doses of tranylcypromine and (more rarely) phenelzine. Reports of minimal blood pressure rises in response to intravenous and oral tyramine administration in patients taking RIMAs (moclobemide and brofaromine) suggest that these drugs are safer and can be used without stringent dietary control. However, it is suggested that they be given after meals, to reduce the likelihood of causing any inhibition of tyramine metabolism. At usual therapeutic doses, the cheese effect should not be a problem with the selective MAO_B inhibitor selegiline, but at higher doses this drug exhibits some inhibition of MAO_A, and dietary restrictions should be followed.

> A 57-year-old male had been successfully treated with tranylcypromine for 10 years after failing to respond to previous antidepressants. At his follow-up appointments he admitted to being somewhat complacent about following a low-tyramine diet, since there appeared to be no hazardous sequelae. A confirmed hypertensive crisis occurred sometime later and was successfully treated with sublingual nifedipine (10 mg). The cause of this reaction was ultimately traced to a change in his beer-drinking habit. For years he had consumed one or two standard domestic beers every week. Just prior to the reaction he sampled a microbrewery bottled draft beer that on subsequent analysis turned out to have a high tyramine content.

Because of its a-adrenergic receptor blocking properties, phentolamine (Regitine) (5 mg intramuscularly or intravenously) has traditionally been the recommended treatment for an MAOI-induced hypertensive crisis. Onset of action is rapid, usually within 5 minutes, and a single dose is generally effective. A diuretic such as furosemide (Lasix) may be given intravenously to avoid fluid retention, and if necessary a b-adrenergic receptor antagonist (beta-blocker) such as propranolol (Inderal) may be given to control tachycardia. If necessary, additional doses of phentolamine can be given over several hours. A recent alternative treatment is sublingual nifedipine (10 mg with a repeat dose after 20 minutes if required). As with phentolamine, the onset of action occurs within 5 minutes.

Although a hypertensive crisis is the most feared reaction with classical MAOIs, paradoxically hypotension is a more common cardiovascular adverse effect, leading to dizziness and fainting after sudden postural changes. In fact this may be a useful clinical sign for determining therapeutic dosage; that is, the dosage may be titrated upward cautiously until some postural hypotension is observed. If postural hypotension becomes a problem in otherwise responsive patients, mineraloglucocorticoids may be given while monitoring serum potassium concentration to avoid hypokalemia. Insomnia can be problematic with MAOIs, ultimately necessitating discontinuation of the drug in a small number of persons; this adverse effect can often be managed by lowering the drug dosage or adding a sedative drug (e.g., a benzodiazepine or trazodone [Desyrel]) to the therapeutic regimen at bedtime. Paradoxically, some patients are sedated on phenelzine and take the medication prior to sleep. The irreversible MAOIs may cause a marked suppression of REM sleep, which is less likely to be experienced with RIMAs. Other common adverse effects of MAOIs include weight gain, peripheral edema, and sexual dysfunction. A peripheral neuropathy resulting from pyridoxine deficiency produced by phenelzine may occur in some patients. In comparison, there is a reduced incidence of sleep disorders, sexual dysfunction, and weight gain with the RIMAs. There are virtually no reports of sexual dysfunction associated with moclobemide.

A discontinuation syndrome consisting of arousal, mood disturbance, and somatic symptoms has been reported with sudden cessation of phenelzine or tranylcypromine administration. As with other classes of antidepressants, it is prudent to taper the dosage of MAOIs and RIMAs over several weeks. With MAOIs (but not with RIMAs) one must avoid administering other antidepressants for at least 10 to 14 days after drug discontinuation to allow a full return to normal MAO activity before starting treatment with a new drug.

DRUG-DRUG INTERACTIONS

The major drug-drug and food-drug interactions (including the so-called “cheese effect”) involving MAOIs are summarized in Table 31.20-3.

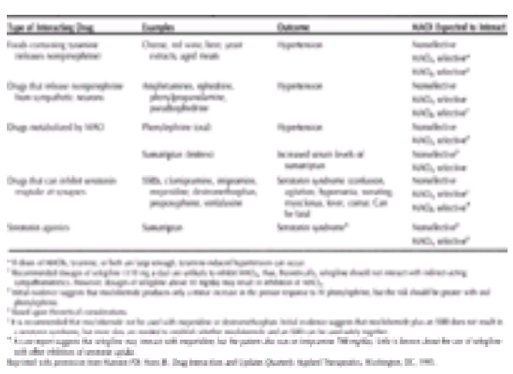

Table 31.20-3 Examples of Interactions of MAOIs with Drugs and Food

Although coadministration of certain MAOIs with tricyclic drugs, SSRIs and psychostimulants such as amphetamines are listed in Table 31.20-4, all three of these combinations have been used successfully in the treatment of refractory depression, but they should only be used with extreme caution. Although MAOIs have been used effectively in combination with amitriptyline (Elavil), imipramine, doxepin (Sinequan), and trazodone (Desyrel), their use with SSRIs or clomipramine (Anafranil) may lead to the potentially fatal serotonin syndrome. Combinations involving tranylcypromine and clomipramine are among the most dangerous and should be avoided. When switching treatment from irreversible MAOIs to tricyclic drugs or SSRIs, a minimum washout of 2 weeks is required to allow complete recovery of MAO activity. Moclobemide and SSRI or tricyclic drug combinations have been reported in the literature to be safe and efficacious, but this work is preliminary and close observation of patients is important. When moclobemide or brofaromine treatment stops, patients completely recover MAO activity in only 24 to 48 hours. When switching from administration of tricyclic antidepressants or SSRI antidepressants to an MAOI, a washout period of 10 to 14 days suffices, except in the case of fluoxetine, which requires a washout period of 5 weeks, primarily because of the long elimination half-life of its active metabolite, norfluoxetine.

Drug	Usual Dose (mg/day)	Usual Maximum Dose (mg/day)	Dosage Form
Isocarboxazid (Marplan)	20–40 mg	60	Oral 10-mg tablets
Phenelzine (Nardil)	30–60	90	Oral 15-mg tablets
Tranylcypromine (Parnate)	20–60	60	Oral 10-mg tablets
Selegiline or (−)deprenyl (Eldepryl)	10	30	Oral 5-mg tablets
Moclobemide (Manerix)	300–600	600	Oral 100- or 150-mg tablets

Table 31.20-4 Typical Dosage Forms and Recommended Dosages for Currently Available MAOIs

Effective lithium (Eskalith) augmentation of MAOIs was first reported in 1972. In countries where it is available, L-tryptophan may also be used in augmentation strategies, but treatment should start at low dosages (250 to 500 mg daily) and be gradually titrated upward to reduce the risk of developing the serotonin syndrome. This syndrome, which results from excessive increases in serotonergic tone, is characterized by tremor at rest, hypertonicity, myoclonus, and autonomic signs. Hallucinations may also result, and a life-threatening hyperthermia may occur.

Phenelzine and tranylcypromine have been reported to inhibit the activity of a number of enzymes involved in drug metabolism. Tranylcypromine has been identified as a relatively potent inhibitor of CYP 2C19, and moclobemide has been reported to be a substrate for CYP 2C19 and to inhibit CYP 2D6, CYP 2C19, and CYP 1A2. Thus the potential for metabolic drug-drug interactions between these MAOIs and coadministered drugs that are substrates for or inhibitors of these isozymes should be considered.

Interactions with prodrugs of amino acids or valproic acid (Depakene), which are converted in the body by MAO_B to amino acids or valproic acid, respectively, may occur. Several such drugs are under investigation as anticonvulsants, and the potential for a serious drug-drug interaction with MAOIs must be considered, as the MAOIs will interfere with metabolism of the prodrug to the active drug of interest.

LABORATORY INTERFERENCES

Interference with laboratory tests is not a major problem with the MAOIs. However, these drugs tend to lower blood glucose values. Acutely elevated blood glucose concentrations, such as that produced during a glucose tolerance test, may decrease the activity of platelet MAO, resulting in a possible synergism with the activity of nonselective or type-B MAOIs.

MAOIs may also be associated with a minimal false elevation in thyroid function test results. Since the nonselective and MAO_A-selective MAOIs cause marked increases in metanephrine and nonmetanephrine, there may be an interference with urinary measurements of these amines when testing for pheochromocytoma.

DOSAGE AND ADMINISTRATION OF MAOIs

Table 31.20-4 summarizes dosage forms and recommended dosages for MAOIs.

RECENT ADVANCES IN MAO AND MAOI RESEARCH

Several novel MAOIs are currently under investigation (Tables 31.20-1 and Fig. 31.20-1) as potential therapeutic agents in depression, anxiety, and neurodegenerative disorders. The involvement of imidazoline receptors (which are intimately associated with MAO) in the etiology of depression and the actions of antidepressant drugs is another recent area of interest. There appears to be an increased density of brain and platelet imidazoline receptors in depressed subjects, and irreversible MAOIs downregulate type 2 imidazole (I_2)-receptors in rat brain. Active research continues on possible components of the endogenous inhibitor of MAO, tribulin. Determination of the structure and function of such components may be useful in developing future MAOIs. For several years, there has been keen interest in the possible effects of cigarette smoking on MAO activity. This interest arose from observations that smoking is very prevalent in people with psychiatric disorders and is associated with a decreased risk of Parkinson's disease. This area of research received increased stimulus from the recent observation that brains of living smokers show 40 percent less MAO_B than brains of nonsmokers or former smokers.

SUGGESTED CROSS-REFERENCES

Section 1.4 provides information on monoamine neurotransmitters. Section 7.7 deals with medical assessments and laboratory testing in psychiatry that might help determine where patients require pharmacological treatment. Chapter 14 provides information on mood disorders. Chapter 15 on anxiety disorders provides information on another condition treated by MAOIs—panic disorder. Section 31.1 gives the general principles of psychopharmacology.

SECTION REFERENCES

*Baker GB, Coutts RT, McKenna K, Sherry-McKenna RL: Insights into the mechanisms of action of the MAO inhibitors phenelzine and tranylcypromine: A review. J Psychiatr Neurosci *17:*206, 1992.

*Baker GB, Urichuk LJ, McKenna KF, Kennedy SH: Metabolism of monoamine oxidase inhibitors. Cell Mol Neurobiol *19:*411, 1999.

Blackwell B: Hypertensive crisis due to monoamine-oxidase inhibition. Lancet *ii:*849, 1963.

Ciraulo DA, Shader RI, Greenblatt DJ, Creelman WL: *Drug Interactions in Psychiatry*, ed 2. Williams & Wilkins, Baltimore, 1995.

Crane GE: Further studies on iproniazid phosphate. J Nerv Dis *124:*322, 1956.

Curet O, Damoiseau G, Aubin N, Sontag N, Rovei V, Jarreau F-X: Befloxatone, a new reversible and selective monoamine oxidase-A inhibitor. I. Biochemical profile. J Pharmacol Exp Ther *277:*253,

1996.

Fahn S, Chouinard S: Experience with tranylcypromine in early Parkinson's disease. J Neural Transm *52*(Suppl):49, 1998.

Fowler JS, Volkow ND, Wang GJ, Pappas N, Logan J, MacGregor R, Alexoff D, Shea C, Schlyer D, Wolf AP, Warner D, Zezulkova I, Cilento R: Inhibition of monoamine oxidase B in the brains of smokers. Nature *379:*733, 1996.

Freedman M, Rewilak D, Xerri T, Cohen S, Gordon AS, Shandling M, Logan AG: L-deprenyl in Alzheimer's disease: Cognitive and behavioral effects. Neurology *50:*660, 1998.

*Fulton B, Benfield P: Moclobemide. An update of its pharmacological properties and therapeutic use. Drugs *52:*450, 1996.

Gardner DM, Shulman KI, Walker SE, Tailor SAN: The making of a user friendly MAOI diet. J Clin Psychiatry *57:*99, 1996.

Gram LF, Geuntert TW, Grange S, Vistisen K, Brosen K: Moclobemide, a substrate of CYP 2C19 and an inhibitor of CYP 2C19, CYP 2D6 and CYP 1A2: A panel study. Clin Pharmacol Ther *57:*670, 1995.

Healy D: *The Antidepressant Era*. Harvard University Press, Cambridge, MA, 1997.

Himmelhoch JM, Detre T, Kupfer DJ, Swartzberg M, Byck R: Treatment of previously intractable depression with tranylcypromine and lithium. J Nerv Ment Dis *155:*216, 1972.

Jefferson JW: Is tranylcypromine really metabolized to amphetamine. J Clin Psychiatry *53:*450, 1992.

*Kennedy SH: Continuation and maintenance treatments in major depression: The neglected role of monoamine oxidase inhibitors. J Psychiatry Neurosci *22:*127, 1997.

Kennedy SH, Glue P: MAOIs: Past, present, future. In *Clinical Advances in Monoamine Oxidase Inhibitor Therapies*, SH Kennedy, editor. American Psychiatric Press, Washington, DC, 1994.

Kennedy SH, Ralevski E, Davis C, Neitzert C: The effects of moclobemide on sexual desire and function in healthy volunteers. Eur Neuropsychopharmacol *6:*177, 1996.

Krishnan KRR: Monoamine oxidase inhibitors. In *The American Psychiatric Press Textbook of Psychopharmacology*, ed 2, AF Schatzberg, CB Nemeroff, editors. American Psychiatric Press, Washington, DC, 1998.

Lenders JWM, Eisenhofer G, Abeling NGGM, Berger W, Murphy DL, Wagemakers LMB, Kopin IJ, Karoum F, vanGennip AH, Brunner HG: Specific genetic deficiencies of the A and B isoenzymes of monoamine oxidase are characterized by distinct neurochemical and clinical phenotypes. J Clin Invest *97:*1010, 1996.

*Lotufo-Neto F, Trivedi M, Thase ME: Meta-analysis of the reversible inhibitors of monoamine oxidase type A moclobemide and brofaromine for the treatment of depression. Neuropsychopharmacology *20:*226, 1999.

Mann JJ, Aarons SF, Wilner PJ, Keilp J: A controlled study of the antidepressant efficacy and side effects of (–)-deprenyl: A selective monoamine oxidase inhibitor. Arch Gen Psychiatry *46:*45, 1989.

Medical Research Council: Clinical trial of the treatment of depressive illness. Br Med J *1:*881, 1965.

Pande AC, Birkett M, Fechner-Bates S, Haskett RF, Greden JF: Fluoxetine versus phenelzine in atypical depression. Biol Psychiatry *40:*1017, 1996.

Piletz JE, Halaris AE, Chikkala D, Qu Y: Platelet I $_1$-imidazoline binding sites are decreased by two dissimilar antidepressant agents in depressed patients. J Psychiatr Res *30:*169, 1996.

*Quitkin FM, Rothschild R, Stewart JW, McGrath PJ, Harrison WM: Atypical depression: A unipolar depressive subtype with preferential response to MAOIs (Columbia University Depressive Studies). In *Clinical Advances in Monoamine Oxidase Inhibitor Therapies*, SH Kennedy, editor. American Psychiatric Press, Washington, DC, 1994.

Robie TR: Marsilid in depression. Am J Psychiatry *114:*936, 1958.

Robinson DS, Cooper TB, Satya PJ, Corcella J, Lutz T: Metabolism and pharmacokinetics of phenelzine: Lack of evidence for acetylation pathway in humans. J Clin Pscychopharmacol *5:*333, 1985.

Shulman KI, Tailor SA, Walker SE, Gardner DM: Tap (draft) beer and monoamine oxidase inhibitor dietary restrictions. Can J Psychiatry *42:*310, 1997.

Stewart JW, Tricamo E, McGrath PJ, Quitkin FM: Prophylactic efficacy of phenelzine and imipramine in chronic atypical depression: Likelihood of recurrence on discontinuation after 6 months' remission. Am J Psychiatry *154:*31, 1997.

*Thase ME, Trivedi MH, Rush AJ: MAOIs in the contemporary treatment of depression. Neuropsychopharmacology *12:*185, 1995.

Volz H-P, Gleiter CH: Monoamine oxidase inhibitors: A perspective on their use in the elderly. Drugs & Aging *13:*341, 1998.

Volz HP, Gleiter CH, Moller HJ: Brofaromine versus imipramine in in-patients with major depression—a controlled trial. J Affect Disord *44:*91, 1997.

Waldmeier PC, Amrein R, Schmid-Burgk W: Pharmacology and pharmacokinetics of brofaromine and moclobemide in animals and humans. In *Clinical Advances in Monoamine Oxidase Inhibitor Therapies*, SH Kennedy, editor. American Psychiatric Press, Washington, DC, 1994.

Youdim MBH, Aronson JK, Blau K, Green AR, Grahame-Smith DG: Tranylcypromine (‘Parnate') overdose: Measurement of tranylcypromine concentrations and MAO inhibitory activity and identification of amphetamine in plasma. Psychol Med *9:*377, 1979.

Yu PH, Boulton AA: A comparison of the effect of brofaromine, phenelzine and tranylcypromine on the activities of some enzymes involved in the metabolism of different neurotransmitters. Res Commun Chem Pathol Pharmacol *16:*141, 1992.

31.21 NALTREXONE

STEPHANIE S. O'MALLEY, PH.D., SUCHITRA KRISHNAN-SARIN, PH.D. AND BRUCE J. ROUNSAVILLE, M.D.

Naltrexone (Revia), an opioid antagonist, was synthesized in 1965 and proved to have a number of advantages over earlier opioid antagonists including cyclazocine, nalorphine, and naloxone (Narcan). Specifically, naltrexone had a relatively long half-life, was orally effective, was not associated with dysphoria, and could be administered once a day. While earlier opioid antagonists, principally naloxone, were used to reduce narcotic overdose, these characteristics of naltrexone made it suitable for use in preventing relapse to opioid use in detoxified opioid addicts. In 1984 naltrexone was approved by the Food and Drug Administration (FDA) for providing blockade of the effects of exogenously administered opioids in detoxified opioid addicts. Over the ensuing years, naltrexone has been tried for the treatment of a wide range of psychiatric disorders, including eating disorders, autistic disorders, self-injurious behavior, nicotine dependence, cocaine dependence, and alcohol dependence. On the basis of extensive preclinical literature that linked endogenous opioids to alcohol drinking behavior and demonstrated that opiate antagonists reduced alcohol self-administration in animals, studies of the efficacy of naltrexone in the treatment of alcohol dependence were undertaken. Naltrexone was approved by the FDA for this use in alcohol dependence in 1994.

Naltrexone is a nonspecific opioid antagonist that binds to all three types of opioid receptors in a dose-dependent manner. It has the highest antagonistic potency at the μ-receptors, which is 5 to 10 times greater than that at d-receptors and 13 to 33 times greater than that at k-receptors. New formulations, including depot versions (Biotek, Woburn), are being developed with the goal of minimizing compliance problems associated with opioid antagonist therapy and are currently in phase-two testing. Recently, opioid antagonists have been developed with affinities for specific opioid receptor types, which have enormous use as selective tools to identify specific receptor types involved in specific pharmacological endpoints. Most of the new specific opioid antagonists are nonpeptides that can easily penetrate the blood-brain barrier and therefore can be administered via peripheral routes. All of these are in the preclinical phase of testing.

CHEMISTRY

Naltrexone, or 17-(cyclopropylmethy)-4,5a-epoxy-3,14-dihydroxymorphinan-6-one hydrochloride, is a competitive opioid antagonist that was developed by substituting an *N*-methyl group on the opioid agonist oxymorphone with a methylcyclopropyl moiety. Naltrexone is available as a white crystalline hydrochloride salt that is soluble in water, with a melting point of 275°C and a pK_a of 8.13 at 37°C. Its molecular formula is $C_{20}H_{23}NO_4 \cdot HCl$ (Fig. 31.21-1).

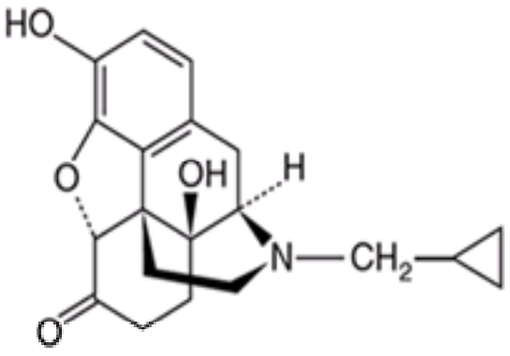

FIGURE 31.21-1 Molecular structure of naltrexone hydrochloride.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics Naltrexone is rapidly and completely absorbed following oral administration of a 100-mg dose, with peak plasma concentrations of 19 to 44 mg/L achieved within 1 hour. Oral bioavailability of naltrexone is 60 percent, with 20 percent of the drug being bound to plasma proteins. A single dose of naltrexone has a volume of distribution of 16.1 L/kg, which decreases to 14.2 L/kg following repeated administration. Naltrexone is almost completely metabolized in the liver. Plasma concentrations of naltrexone exhibit an initial biphasic elimination phase of 3.9 to 10.3 hours' duration, followed by a terminal elimination phase observed 24 hours following administration, with a half-life of 96 hours.

Naltrexone is primarily recovered in plasma, urine, and feces as conjugated naltrexone and conjugated and unconjugated 6b-naltrexol along with minor quantities of 2-hydroxy-3-*O*-methyl-6b-naltrexol and 2-hydroxy-3-*O*-methylnatrexone. Approximately 60 percent of the drug is recovered in the urine in 48 hours following oral administration and in 72 hours following intravenous administration. Naltrexone has a plasma half-life of 10.5 hours regardless of whether it is administered via the oral or intravenous route. Conjugated 6b-naltrexol is a weaker antagonist that has a plasma half-life of 14 to 19 hours and has been proposed to contribute to the longer duration of action of naltrexone. Recent studies indicate that after a 100-mg dose, naltrexone can be detected in plasma for 30 hours, while 6-b-naltrexol can be detected for up to 125 hours.

Pharmacodynamics Naltrexone is a nonspecific opioid antagonist that binds to all three opioid receptor sites (μ, d, k) as a function of dose administered. Naltrexone has minimal opioid agonistic effects, and its relative opiate antagonistic potency is 12 times that of nalorphine and 2.5 times that of naloxone or cyclazocine. The time course of action of oral naltrexone in rodents is about 3 times longer than that of naloxone. No significant species differences in the binding of [^{3}H]naltrexone have been documented. An oral 100-mg dose of naltrexone almost completely antagonized the subjective and objective responses of 25 mg of intravenous heroin. Administration of 50 mg of naltrexone precipitates an abstinence syndrome in morphine-dependent subjects, an effect that decreases with chronic use. This decrease is believed to be due to increases in the number and functional activity of brain opioid receptors with long-term naltrexone administration. Naltrexone increases secretion of gonadotropin-releasing hormone and corticotropin-releasing factor and significantly increases plasma concentrations of b-endorphin, cortisol, lutenizing hormone, and follicle-stimulating hormone.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Design Naltrexone has demonstrated efficacy in blocking the effects of exogenously administered opiates and as an aid in preventing relapse in recently detoxified patients with alcohol dependence. It has been investigated as a key element in several protocols for rapid opioid detoxification and as a treatment for self-injurious and other autistic behavior in patients with mental retardation or autistic disorder. The use of naltrexone as a maintenance treatment for opioid dependence, as treatment for alcohol dependence, and in rapid opioid detoxification is supported by studies using experimental designs. Support for use in mentally retarded and autistic patients is based only on small studies and case reports that have not been confirmed by larger double-blind clinical trials. Open-label and randomized controlled trials with relatively small samples have examined the efficacy of naltrexone in reducing binge eating and obesity.

Interpretation

Alcohol Dependence The efficacy of naltrexone was originally demonstrated in two double-blind placebo-controlled trials of naltrexone in alcohol-dependent patients. In each of these studies, treatment with 50 mg daily was initiated following a minimum of 7 days of abstinence or completion of medical detoxification and was provided with concurrent psychosocial interventions. In a combined analysis of these studies, naltrexone-treated patients were more likely to remain abstinent (54 versus 31 percent), to experience fewer days of drinking (3.1 versus 7.5 percent), and to report lower rates of relapse to heavy drinking (25 versus 52 percent). Furthermore, participants who lapsed (i.e., were nonabstinent) were less likely to experience heavy drinking if they were taking naltrexone rather than placebo (57 versus 77 percent). Reductions in craving were apparent at weeks 3 and 9. Failure to find more consistent differences in craving may be related to the use of single-item analog scales to measure craving.

As naltrexone is used in general practice, less robust effects can be expected in more-heterogeneous samples of patients. While some recent clinical trials have confirmed the overall efficacy of naltrexone, other studies have found that the differences between naltrexone- and placebo-treated patients emerge only when the sample is stratified on the basis of compliance. Thus, strategies to improve compliance are likely to be important. Patients who take their medication in the morning or have an established routine for medication dosing have higher compliance rates.

Current research is examining the potential effectiveness of higher dosages of naltrexone (100 to 150 mg of naltrexone daily) particularly among patients with polysubstance abuse problems. Preclinical research and small human pilot studies suggest that greater suppression of drinking results with concurrent treatment with naltrexone and a serotonin reuptake inhibitor. Larger-scale clinical trials are under way to follow up on these preliminary findings.

Opioid Dependence As an opioid antagonist, naltrexone selectively competes with exogenously ingested opioids for central nervous system (CNS) and non-CNS opioid receptors and blocks their activity. This blockade lasts for up to 72 hours after naltrexone administration and prevents patients from achieving sought-after reinforcement from opioid use. This blockade can have positive psychological consequences including reduced craving related to reduced immediate availability of opioid effects and deconditioning of cue-related craving as patients are exposed to people, places, and things formerly associated with opioid use while protected from opioid effects. The opioid antagonist properties of naltrexone do not appear to be subject to development of tolerance, as potency has been demonstrated after maintenance for 1 year or more. At the dosage used, naltrexone does not have opioid agonist properties, and patients do not obtain a psychoactive reward (high). Neither is it associated with significant withdrawal symptoms if use is discontinued. However, patients may be more sensitive to lower dosages of opioids once naltrexone is discontinued, because of the loss of tolerance to exogenous opioids.

In carefully selected patients who are motivated to take naltrexone, maintenance on naltrexone can be used as an alternative to maintenance treatment with opioid agonists, such as methadone or levomethadyl acetate (ORLAAM). As a nonscheduled medication without psychoactive or addictive properties, naltrexone can be administered as a maintenance treatment in a range of medical settings without any special licensing requirements. Because diversion is not an issue, take-home medications can be provided with no special precautions. However, lack of initial patient acceptance and subsequent medication noncompliance limits the utility of naltrexone for opioid dependence; here the strengths of naltrexone are also its weaknesses. As a drug without psychoactive effects which blocks effects of opioids, naltrexone can be used only after completion of a 7- to 10-day period of complete opioid abstinence or after a rapid-detoxification procedure (see below). Hence, patients must start treatment with a goal of achieving total opioid abstinence and a willingness to undergo detoxification. Unlike patients maintained on opioid agonists, patients on naltrexone achieve no rewarding effects of taking medications and experience no withdrawal effects if doses are missed. Thus, there are no pharmacological barriers to dropping out of treatment. In addition, some patients continue to complain of at least mild adverse effects or opioid withdrawal effects throughout the first several weeks of treatment. Because of these factors, most patients from unselected populations drop out within the first 3 months of treatment. Greater success has been achieved with patients motivated by powerful contingencies, such as opioid-dependent health care professionals for whom opioid abstinence is a condition for regaining medical licensure or incarcerated addicts for whom treatment is a requirement for probation or parole. A treatment plan involving naltrexone dosage monitored by either a medical professional or a significant other also enhances compliance.

Opioid Detoxification Protocols Successful outpatient protocols for rapid detoxification have been developed combining naltrexone with clonidine (Catapres), a selective a_2-adrenergic receptor antagonist initially used as an antihypertensive agent. This combined treatment allows patients to be completely opioid-free from the first day of treatment, to have completed detoxification within 48 to 72 hours, and to be able to begin naltrexone maintenance at that time. For most patients, withdrawal symptoms are moderate to severe on the first day of treatment and drop off sharply thereafter. Because blood pressure may be lowered substantially by clonidine administration, patients must be monitored for up to 8 hours on the initial detoxification day, limiting this procedure to settings in which treatment rooms and adequate nursing staff are available. These protocols are particularly useful in decreasing the need for inpatient treatment and reducing absence from work. Also, as a method of initiating maintenance naltrexone treatment, this mode of detoxification removes the need for an unprotected 7- to 10-day opioid-free period prior to starting naltrexone administration, a time when vulnerability to relapse is high.

A variant method of ultrarapid inpatient detoxification involves naltrexone administered to opioid-dependent patients who are placed under light sedation or general anesthesia (including mechanical ventilation) followed by detoxification using full doses of naltrexone. Subsequently, patients continue to receive naltrexone to prevent relapse. While this method eliminates patients' conscious awareness of withdrawal discomfort, the added risks and expenses of general anesthesia outweigh this benefit. In addition, studies are yet to be conducted comparing relapse rates following this ultrarapid detoxification procedure with the safer clonidine-naltrexone detoxification protocol. As a result, ultrarapid detoxification cannot be recommended at this time.

Nicotine Dependence Several lines of evidence suggest that the opioid system may be involved in nicotine dependence. Nicotine administration is known to stimulate release of endogenous opioids and to attenuate perception of pain via opioid mechanisms. However, the potential clinical utility of naltrexone for smoking cessation remains to be established. Clinical laboratory studies examining the effects of opioid antagonists, including naloxone and naltrexone, on smoking behavior have been inconclusive, with some studies showing reductions in smoking and others finding no effect. Pilot studies suggest that naltrexone in combination with nicotine replacement therapy may have some value in smoking cessation. Given the preliminary stage of this research, naltrexone cannot be recommended for use in smoking cessation.

Self-Injurious Behavior Naltrexone has been proposed as a treatment of self-injurious behavior and of autistic behaviors in patients with mental retardation or autistic disorder. Clinically significant symptom reduction has been reported from numerous open-case studies and trials with small sample sizes. Larger studies using more rigorous experimental designs have failed to replicate these reports. Hence, naltrexone's utility for these symptoms is not established.

Eating Disorders and Obesity Numerous studies in rodents have suggested an association between altered endogenous opioid activity and obesity and other eating disorders. Exogenous and endogenous opioid administration increases food intake, while opioid antagonists block the increase in food intake. The clinical data on the use of opioid antagonists in eating disorders are controversial, and the literature is replete with studies that demonstrate differential effects of these agents based on the type of eating disorder. For example, a laboratory study demonstrated that naloxone reduces the consumption of sweet high-fat foods by both obese and lean binge eaters compared with that by obese and lean nonbingers, suggesting a role of the opioid system in binge eating but not in obesity. Controlled clinical trials in obese individuals have failed to document a significant effect of naltrexone on weight loss. A few small-sample controlled clinical trials have been conducted on bulimia nervosa, with some reporting reductions in binge symptomatology. With regard to anorexia nervosa, naloxone infusion produced a 10-fold increase in weight gain per week, an effect that ended when the naloxone infusion was stopped in patients who were concurrently treated with amitryptyline (Elavil). In a separate study of eight subjects, naltrexone also seemed to have a small nonsignificant beneficial effect on weight gain. These results are preliminary and have not been extended to a larger controlled trial. Therefore, the role for naltrexone and other opioid antagonists in the treatment of eating disorders remains unclear, and larger, more definitive controlled clinical trials are needed.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Respiratory System In opioid-dependent individuals, administration of naltrexone produces a variety of nasal symptoms such as nasal congestion, sneezing, rhinorrhea, sore throat, cough, and shortness of breath. Studies in normal opioid-free individuals have not documented any nasal symptoms. Some subjects have also reported blurred eyes and increased light sensitivity as well as tinnitus.

Heart Naltrexone increases blood pressure in opioid-dependent individuals and can also produce palpitations and tachycardia. Naltrexone also decreases fibrinogen levels in opiate addicts.

Smooth Muscles Naltrexone produced muscle tension, tremors and twitching in opioid-dependent subjects. Individuals who are not dependent on opioids have not

displayed similar changes.

Liver While naltrexone has not been shown to be hepatotoxic when administered in the recommended dose of 50 mg a day, hepatotoxic potential has been documented for a fivefold higher dosage of 300 mg a day in obese subjects and in patients with Alzheimer's disease. At the dosages currently recommended, naltrexone was not associated with increased risk of hepatotoxicity in over 500 alcohol-dependent patients treated for up to 12 weeks. Nonetheless, naltrexone is contraindicated in subjects with acute hepatitis or liver failure. However, mild-to-moderate transaminase elevations associated with alcohol or other substance use are not a contraindication.

Endocrine System Changes in baseline levels of some hypothalamic, pituitary, and gonadal hormones have been reported with naltrexone. Naltrexone significantly increases plasma testosterone luteinizing hormone, adrenocorticotropic hormone (ACTH), and cortisol concentration in opioid-dependent subjects and to reverse heroin-induced decreases in the same. In healthy adults, naltrexone has been shown to increase plasma prolactin, luteinizing hormone, testosterone, ACTH, follicle-stimulating hormone (FSH), and cortisol concentrations.

Gastrointestinal Symptoms Abdominal pain and cramps have been documented following naltrexone administration in both opioid-dependent, alcohol-dependent and nondependent persons. Some persons also report nausea, vomiting, loss of appetite, diarrhea, or constipation.

Carcinogenesis and Mutagenesis Carcinogenicity studies in rats have documented a small increase in mesotheliomas and vascular tumors. No evidence of mutagenicity has been documented.

Pregnancy Naltrexone administered at doses that were 140 times the therapeutic dose in humans decreased the pregnancy rate in rats and had an embryocidal effect in rats and rabbits; 100-mg/kg doses in rats and 60-mg/kg doses in rabbits had similar effects. Preclinical studies showed that naltrexone passes through the placenta to the fetus. There are no adequate or controlled studies of naltrexone in pregnant or nursing women.

Genitourinary Less than 10 percent of subjects report increased frequency of urination, discomfort during urination, increased frequency of erections, or alteration in sexual interest.

Psychiatric Symptoms Non-opioid-dependent individuals have also reported feeling fatigued, tired, and listless following administration of naltrexone. Both opioid-dependent and nondependent individuals report increased irritability, restlessness, mental confusion, and slowing following naltrexone administration. Substance abusers have an increased risk of suicide, and this risk is not reduced by naltrexone.

THERAPEUTIC INDICATIONS

FDA-approved indications for naltrexone include the maintenance treatment for detoxified patients with opioid dependence and relapse prevention in patients with alcohol dependence. Naltrexone has established utility when combined with clonidine as part of outpatient and inpatient detoxification protocols.

PRECAUTIONS AND ADVERSE REACTIONS

As an opioid antagonist, naltrexone can precipitate an acute withdrawal syndrome in persons who are dependent on opioids. As a result, patients should be abstinent from short-acting opioids (e.g., heroin) for at least 5 days and from longer-acting opioids (e.g., methadone [Dolophine]) for 10 days or more as judged by self-report and urine toxicology screening test. If there is a question of occult dependence in an individual with a negative urine toxicology screen for opioids who shows no signs of withdrawal, a naloxone challenge test is recommended in which the patient is observed for signs and symptoms of opioid withdrawal for 20 minutes following a subcutaneous injection of 0.8 mg of naloxone. Otherwise, naltrexone is generally well tolerated, and most patients who have no reaction to a test dose of naloxone report no adverse effects. The profile of reported adverse effects has considerable similarity to the symptoms of opioid withdrawal and includes gastrointestinal distress (nausea, vomiting, diarrhea, abdominal pain), anxiety and restlessness, dysphoria, mild hypertension, headache, and insomnia. These adverse effects are most prominent in the first several days of use and improve rapidly for patients remaining in treatment. However, a substantial minority of those with a diagnosis of opioid dependence (5 to 10 percent) may refuse a second dose, and an additional 10 to 20 percent may discontinue medication in the first month of treatment, often citing continued discomfort reminiscent of opioid withdrawal throughout the first 1 to 3 weeks of treatment. Adverse effects can be minimized by ensuring an adequate interval between cessation of opioid use and initiation of naltrexone therapy and by a graduated initial dosage schedule starting with 12.5 or 25 mg daily for the first several days.

Patients should be advised that taking large quantities of exogenous opioids in an attempt to overcome the blockade provided by naltrexone can be extremely dangerous. Life-threatening opioid intoxication may result, including respiratory arrest or circulatory collapse. In addition, patients should be advised that once naltrexone use is discontinued, they may be more sensitive to the effects of opioids because of loss of tolerance and may be at greater risk for overdose.

Among alcohol-dependent patients, nausea is the most common adverse reaction to naltrexone; it can be mild to severe and typically occurs within 90 minutes of administration. In a large multisite safety study, adverse effects that occurred in more than 2 percent of patients included nausea (10 percent), headache (7 percent), dizziness (4 percent), nervousness (4 percent), fatigue (4 percent), insomnia (3 percent), vomiting (3 percent), anxiety (2 percent), and somnolence (2 percent). The effect of naltrexone on acute alcohol withdrawal has not been investigated, and the studies demonstrating the efficacy of naltrexone in alcohol dependence required a minimum of 7 days of abstinence or medical detoxification from alcohol prior to initiating treatment. The rate of early termination due to adverse events has been substantially higher in studies in which naltrexone was administered to actively drinking alcoholics. As a result, naltrexone should probably be administered only after the risk of alcohol withdrawal has passed (4 to 5 days). Women and younger patients appear to be at somewhat increased risk of adverse effects.

The potential for hepatotoxicity at high doses has been raised as a more serious concern. However, convincing reports of elevated liver function test results have been limited to patients treated for eating disorders with doses of 250 to 300 mg daily, five to six times higher than the recommended maintenance dosage for opioid or alcohol dependence. As a precaution, patients should receive a full battery of liver function tests prior to receiving naltrexone, and naltrexone is contraindicated in patients with liver failure or acute hepatitis. The moderate transaminitis often seen in alcohol and opioid dependence is not a contraindication; however, baseline bilirubin and serum transaminases and periodic monitoring of transaminases are indicated. As a guideline, liver function tests should be repeated monthly for the first 3 months and every 3 to 6 months thereafter, if there is no evidence of rising enzyme levels. If persistent elevations in liver enzymes occur, naltrexone should be discontinued. The possibility of medication noncompliance and continued alcohol or drug use should be ruled out when mild transaminase elevations (threefold or less) occur.

Because naltrexone is metabolized by the liver and excreted through the kidneys, caution should be used in patients with severe liver disease or renal impairment. Patients with compensated and decompensated liver cirrhosis show a 5- to 10-fold increase in naltrexone levels relative to subjects with normal liver function. The increased systemic availability of naltrexone, which is believed to be due to lower rates of conversion of naltrexone to 6-b-naltrexol, appears to be related to the severity of liver disease.

Naltrexone has not been studied in pregnant or nursing women, but an embryocidal effect was demonstrated in the rat and rabbit given doses approximately 140 times the recommended human dose. As a result, naltrexone should be used during pregnancy only when the potential benefit clearly outweighs the potential risks to the fetus. Intensive psychosocial methods to support abstinence should be considered first-line therapy in pregnant women.

DRUG INTERACTIONS

Because naltrexone is an opioid antagonist, it blocks the pain-relieving effects of analgesics. In patients maintained on naltrexone, nonopioid pain medications or approaches should be tried first when possible. Thus, the use of nonsteroidal anti-inflammatory drugs should be considered the first line of pharmacological therapy for pain management. For elective surgery, cessation of naltrexone treatment 72 hours prior to surgery is recommended. Naltrexone does not reverse effects of nerve root blocks or spinal anesthesia and does not antagonize agents used for general anesthesia. If opioid medications are necessary for treatment of severe pain, the blockade can be overridden with a rapidly acting analgesic, but treatment requires close monitoring in a setting equipped and staffed for cardiopulmonary resuscitation, because the resulting respiratory depression may be deep and prolonged. Patients should be given a card stating that they are on naltrexone to alert medical personnel in case of an emergency.

The safety and efficacy of combined use of disulfiram (Antabuse) and naltrexone are unknown. Given that both medications are potentially hepatotoxic, their concomitant use is not routinely recommended at this time. However, some clinicians use the two drugs concomitantly for brief periods with more-frequent monitoring of hepatic function. Although formal drug interaction studies have not been conducted with antidepressant medications, the incidence of adverse events reported by patients receiving naltrexone in addition to antidepressants was similar to that of patients receiving naltrexone alone in a large safety study. Increased lethargy and somnolence have been reported with the combined use of naltrexone and thioridazine (Mellaril).

LABORATORY INTERFERENCES

Thin-layer, gas-liquid, and high-pressure liquid chromatographic methods used for the detection of opioids in the urine are not interfered with by naltrexone. The potential for a false-positive urine result for opioids using less-specific urine screening tests such as enzyme multiplied immunoassay technique (EMIT) may exist, because naltrexone is a derivative of oxymorphone.

DOSAGE AND ADMINISTRATION

Naltrexone is prepared in scored 50-mg capsule-shaped tablets. Blister packages of 28 tablets are available as a means of enhancing compliance. The recommended dosage for adults is 50 mg a day for alcohol or opioid dependence. Lower initial dosages can be considered over the first few days for patients at increased risk of adverse effects with the dosage increased as tolerated. The efficacy of doses ranging from 25 mg to 150 mg daily is being investigated. Morning administration is recommended for improved medication compliance. To supervise the ingestion of naltrexone in the treatment of narcotic addiction, flexible dosing can be used, including 100 mg every other day or 150 mg every third day.

Naltrexone should not be administered until there is no reasonable possibility of opiate use within the past 7 to 10 days or unless the patient has completed a clonidine (Catapres)-naltrexone detoxification, and abstinence has been verified by urine testing for opiates. If there is a question about continued opiate use, opiate-free status should be confirmed by a negative response to an intramuscular naloxone test dose of 0.8 mg. If substantial opioid withdrawal symptoms do emerge, initiation of naltrexone should be delayed at least 1 additional day. Graduated daily dosing beginning with 25 mg on the first day is recommended during the initial 1- to 2-week period prior to initiating the above thrice weekly schedule.

For rapid detoxification from opioids, patients begin the first opioid-free day with clonidine 0.2 mg orally given every 2 hours for a daily total of up to 1.8 mg, and blood pressure is checked every 30 to 60 minutes as needed. Naltrexone (12.5 mg) is administered 1 to 3 hours after the first clonidine dose. To reduce muscle cramps and later insomnia, oxazepam (Serax), 30 to 60 mg is administered orally at the same time as the initial clonidine dose, and 15 to 30 mg is given every 4 to 6 hours as needed, with a maximum daily dose of 180 mg. Patients should remain at the clinic with a bed available for up to 8 hours on this day and be taken home by a reliable escort. On day 2, similar doses of clonidine and oxazepam are provided, but naltrexone (25 mg) is administered at the onset of the day. Relatively asymptomatic patients can return home after 3 to 4 hours. On days 3 to 5, 50 mg of naltrexone is administered, and doses of clonidine and oxazepam are tapered such that these are discontinued by 5 to 10 days after beginning detoxification while naltrexone administration (50 mg daily) is continued.

SUGGESTED CROSS-REFERENCES

Issues related to specific disorders are in the following sections: alcohol-related disorders (Section 11.2), opioid-related disorders (Section 11.9), and eating disorders (Chapter 20). Clonidine is reviewed in Section 31.16.

SECTION REFERENCES

Bodnar RJ: Opioid receptor subtype antagonists and ingestion. In *Drug Receptor Subtypes and Ingestive Behaviors*, SJ Cooper, RG Clifton, editors. Academic Press Limited, London, 1996.

Croop RS, Faulkner EB, Labriola DF: The safety profile of naltrexone in the treatment of alcoholism—results from a multicenter usage study. Arch Gen Psychiatry *54:*1130, 1997.

de Zwaan M, Mitchell JE: Opiate antagonist and eating behavior in humans: A review. J Clin Pharmacol *32:*1060, 1992.

*Farren C, O'Malley SS, Rounsaville BJ: Naltrexone and opiate abuse: Who benefits from it and how to increase its effectiveness. In *New Treatments for Opiate Dependence: What Treatments For Which Patients*, S Stine, T Kosten, editors. Guilford, New York, 1997.

Froehlich JC, Wand G: The neurobiology of ethanol-opioid interactions in ethanol reinforcement. Alcoholism. Clin Exp Res *20(Suppl):*181A, 1996.

Gonzales JP, Brogden RN: Naltrexone. A review of its pharmacodynamic and pharmacokinetic properties and therapeutic efficacy in the management of opioid dependence. Drugs *35:*192, 1988.

Herz A: Endogenous opioid systems and alcohol addiction. Psychopharmacology *129:*99, 1997.

Howlett TA, Rees LH: Endogenous opioid peptides and hypothalamic-pituitary function. Annu Rev Physiol *48:*527, 1986.

Jaffe AJ, Rounsaville B, Chang G, Schottenfeld RS, Meyer RE, O'Malley SS: Naltrexone, relapse prevention, and supportive therapy with alcoholics: An analysis of patient treatment matching. J Consult Clin Psychol *64:*1044, 1996.

Luby ED, Marrazzi MA, Kinzie J: Case reports-treatment of chronic anorexia nervosa with opiate blockade. J Clin Psychopharmacol *7:*52, 1987.

Marrazzi MA, Bacon JP, Kinzie J, Luby ED: Naltrexone use in the treatment of anorexia nervosa and bulimia nervosa. Int Clin Psychopharmacol *10:*163, 1995.

Mitchell JE, Christenson G, Jennings J, Huber M, Thomas B, Pomeroy C, Morley J: A placebo-controlled, double-blind crossover study of naltrexone hydrochloride in outpatients with normal weight bulimia. J Clin Psychopharmacol *9:*94, 1989.

O'Connor PG, Carroll KM, Shi JM, Schottenfeld RS, Kosten TR, Rounsaville BJ: Three methods of opioid detoxification in a primary care setting. Ann Intern Med *127:*526, 1997.

*O'Connor PG, Kosten TR: Rapid and ultrarapid opioid detoxification techniques. JAMA *279:*229, 1998.

*O'Malley SS, Jaffe AJ, Chang G, Rode S, Schottenfeld R, Meyer RE, Rounsaville B: Six-month follow-up of naltrexone and psychotherapy for alcohol dependence. Arch Gen Psychiatry *53:*217, 1996.

Oslin D, Liberto JG, O'Brien J, Krois S, Norbeck J: Naltrexone as an adjunctive treatment for older patients with alcohol dependence. Am J Geriatr Psychiatry *5:*324, 1997.

*Oslin DW, Pettinati HM, Volpicelli JR, Wolf AL, Kampman KM, O'Brien CP: The effects of naltrexone on alcohol and cocaine use in dually addicted patients. J Subst Abuse Treat *16:*163, 1999.

Spanagel R, Zieglgansberger W: Anti-craving compounds for ethanol: New pharmacological tools to study addictive processes. Trends Pharmacol Sci *18:*54, 1997.

Takemori AE, Portoghese PS: Selective naltrexone-derived opioid receptor antagonists. Annu Rev Pharmacol Toxicol *32:*239, 1992.

*Volpicelli JR, Rhines KC, Rhines JS, Volpicelli LA, Alterman AI, O'Brien CP: Naltrexone and alcohol dependence. Role of subject compliance. Arch Gen Psychiatry *54:*737, 1997.

Weinrieb RM, O'Brien CP: Naltrexone in the treatment of alcoholism. Annu Rev Med *48:*477, 1997.

*Willemsen-Swinkels SHN, Buitelaar JK, Nijhof GJ, van Engeland H: Failure of naltrexone hydrochloride to reduce self-injurious and autistic behavior in mentally retarded adults. Arch Gen Psychiatry *52:*766, 1995.

Yeomans MR, Gray RW: Selective effects of naltrexone on food pleasantness and intake. Physiol Behav *60:*439, 1996.

Zubin RS, Tempel A: Neurochemical correlates of opiate receptor regulation. Biochem Pharm *35:*1623, 1986.

31.22 NEFAZODONE

STEVEN J. GARLOW, M.D., PH.D., MICHAEL J. OWENS, PH.D. AND CHARLES B. NEMEROFF, M.D., PH.D.

Nefazodone (Serzone) is an antidepressant that was introduced in the United States in 1995. Nefazodone is chemically unrelated to selective serotonin reuptake inhibitors (SSRIs), tricyclic drugs, tetracyclics drugs, or monamine oxidase inhibitors. Nefazodone exerts multiple effects on serotonergic neurons and their receptors, which are believed responsible for its antidepressant action. Nefazodone has a pharmacological profile that overlaps significantly with that of trazodone (Desyrel), with some notable differences. This is not unexpected because nefazodone resulted from a drug development program that sought to improve on the actions of trazodone by maintaining its efficacy in the treatment of depression while minimizing some of its adverse effects. The only current Food and Drug Administration (FDA) indication for the use of nefazodone is in the treatment of major depressive disorder.

Nefazodone is a phenylpiperazine with some structural similarity to trazodone. Like trazodone, nefazodone has diffuse and complex pharmacological actions that include antagonism at the serotonin (5-hydroxytryptamine [5-HT]) type 2A (5-HT $_{2A}$) receptor and inhibition of the serotonin transporter. This combination of actions is hypothesized to account for the antidepressant action displayed by nefazodone. The structural differences between nefazodone and trazodone account for the differences in their respective side-effect profiles. In particular, nefazodone is less sedating than trazodone, and to date no cases of nefazodone-induced priapism have been reported. This particular difference is thought to be due to nefazodone's relatively low affinity for peripheral a $_1$-adrenergic receptors compared with trazodone. While lower than that of trazodone, nefazodone does have some antagonist action at peripheral a $_1$-receptors, and this is thought to mediate the orthostatic hypotension that it can cause.

CHEMISTRY

Nefazodone hydrochloride is a synthetically derived phenylpiperazine antidepressant (Fig. 31.22-1). Its formal chemical designation is 2-[3-[4-(3-chlorophenoxyl)-1-piperazinyl]-propyl]-5-ethyl-2,4-dihydro-4-(2-phenoxyethyl)-3 *H*-1,2,4-triazol-3-one monohydrochloride, with a generic formula of $C_{25}H_{32}ClN_5O_2 \cdot HCl$, and a molecular weight of 506.5.

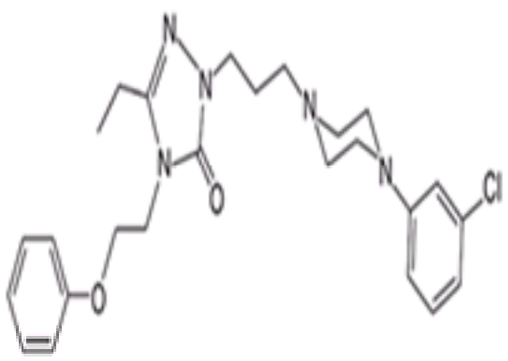

FIGURE 31.22-1 Molecular structure of nefazodone.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics

Absorption Nefazodone is rapidly and completely absorbed after oral administration. Peak plasma concentrations occur within 1 hour of oral dosing. Ingestion of nefazodone with food slows the rate of absorption from the gastrointestinal tract; whether this has clinically meaningful effects is unclear.

Distribution Nefazodone is widely distributed in the body after oral dosing. The compound is distributed into all tissues and compartments and readily crosses the blood-brain barrier. Nefazodone is extensively (99 percent) but loosely protein bound in serum. It does not displace other protein-bound drugs including chlorpromazine (Thorazine), desipramine (Norpramin, Pertofrane), diazepam (Valium), phenytoin (Dilantin), lidocaine (Xylocaine), prazosin (Minipress), propranolol (Inderal), verapamil (Calan, Isoptin), or warfarin (Coumadin).

Metabolism Nefazodone is extensively metabolized in the liver via n-dealkylation and aliphatic and aromatic hydroxylation. Nefazodone is a substrate for the cytochrome P450 (CYP) 3A4 (CYP 3A4) isoenzyme. Three major pharmacologically active metabolites result from this hepatic biotransformation, hydroxynefazodone, triazoledione, and *m*-chlorophenylpiperazine (m-CPP). Less than 1 percent of an oral dose of nefazodone is excreted unchanged in urine. The half-life of nefazodone after oral administration is 2 to 4 hours. Nefazodone and hydroxynefazodone display nonlinear pharmacokinetics, with serum concentrations exhibiting greater than expected increases with dosage increases. Thus the increases in area under the curve (AUC) and maximum concentration (C_{max}) are on the order of threefold to fourfold in response to doubling of the dosage. Steady-state serum levels are achieved at 4 to 5 days with twice-daily dosing regimens or with changes in previous regimens. In older patients, single oral doses of nefazodone resulted in AUCs and C_{max}s as much as twice as high as in nonelderly patients. With a multiple-dose regimen, serum concentrations in elderly patients were on the order of 10 to 20 percent higher than in nonelderly persons. There was also a sex-based difference in serum concentrations, with females having higher AUCs and C_{max}s than male subjects after a single oral dose; no such differences were noted on a multiple-dose schedule. This concatenation of results suggests that elderly patients (particularly, elderly female patients) should initially receive half the normal dosage of nefazodone. However, the final therapeutic range is the same in elderly and nonelderly patients and in female and male patients. Patients with liver disease (in particular, cirrhosis) exhibited final serum concentrations of nefazodone that were 25 percent higher than those in normal volunteers. Therefore, nefazodone dosage should be adjusted downward for patients with liver disease. Renal disease seems to have no effect on serum concentrations of nefazodone. The CYP 2D6 is responsible for the metabolism of a large number of drugs including tricyclic drugs and SSRIs but not nefazodone. Thus patients with low CYP 2D6 activity, also referred to as “poor metabolizers,” do not require dosage adjustments of nefazodone.

Pharmacodynamics Nefazodone and its metabolites display a diverse and complex pharmacological profile. Nefazodone is a potent antagonist of the 5-HT $_{2A}$ receptor. This has been demonstrated both in vitro and in vivo. This antagonist action is hypothesized to be central to the antidepressant action of nefazodone. The antagonist action at the 5-HT$_{2A}$ receptor is one of the areas of overlap between nefazodone and trazodone. Nefazodone inhibits the serotonin transporter, which likely also contributes to this drug's antidepressant actions. Nefazodone displays moderate antagonism of the serotonin transporter in vitro, but at clinically relevant doses it does antagonize the serotonin transporter in vivo. Nefazodone also displays moderate inhibition of the norepinephrine transporter, which may contribute to its antidepressant action; in contrast it has little affinity for the dopamine transporter. Nefazodone has low-to-moderate antagonist affinity for a $_1$-receptors, much less than trazodone. Presumably this accounts for the differences between the two compounds in the incidence of both orthostatic hypotension and priapism. While both cause

some orthostatic hypotension due to antagonism of the peripheral a_1-receptors, only trazodone has been reported to cause priapism, presumably because of its higher affinity to this receptor. Otherwise, nefazodone appears to have little or no affinity for other receptor types including muscarinic cholinergic, adenosine, dopamine, g-aminobutyric acid (GABA), opioid, glutamate, glycine, and other serotonin receptors.

Hydroxynefazodone displays a pharmacological profile almost identical to that of the parent compound, with similar binding affinities. Hydroxynefazodone is a potent antagonist of the 5-HT_{2A} receptor and an inhibitor of the serotonin transporter. This metabolite has moderate inhibitor affinity for the nonepinephrine transporter. Given the receptor affinities and pharmacokinetics of hydroxynefazodone, it seems reasonable to conclude that this metabolite contributes significantly to the therapeutic efficacy of nefazodone. The trizalodione metabolite has a range of receptor activities similar to those of nefazodone and hydroxynefazodone but has much lower affinity for the various transporters and receptors than the other two compounds. While the hydroxynefazodone metabolite almost certainly participates in the antidepressant action of the agent, it is not clear whether the trizalodione compound does so.

The third pharmacologically active metabolite of nefazodone, m-CPP, results from hepatic transformation of both nefazodone and trazodone. This compound has a number of actions; most notably, it is a weak antagonist of the 5-HT_{2A} receptor but a moderately potent agonist of the 5-HT_{2C} and 5-HT_{1B} receptors. Administration of m-CPP to experimental animals results in behavioral activation and potentiation of anxiogenic stimuli. Administration of m-CPP to human subjects has anxiogenic effects and when given to anxiety-prone subjects can trigger significant symptoms of anxiety and panic. Both antidepressants, nefazodone and trazodone, are reported to cause symptoms of anxiety in some patients at the beginning of treatment, which is probably due to the m-CPP metabolite. m-CPP and trazodone have also been reported to provoke migraine headache in susceptible individuals.

THERAPEUTIC INDICATIONS

Nefazodone became available in the United States in early 1995. Most of the clinical-trials literature about nefazodone supports its action as an antidepressant. Because this agent is newly approved, relatively few studies have been published on its use in other disorders. In contrast, trazodone has been available in the United States since 1981, and a large body of information exists related to its use in a wide range of psychiatric disorders and clinical settings.

Major Depressive Disorder Nefazodone was evaluated as an antidepressant in four phase III clinical trials. These double-blind, placebo- and active treatment (imipramine [Tofranil])-controlled trials evaluated the efficacy of nefazodone in the treatment of outpatients with major depressive disorder. Treatment response was quantified with the 17-item Hamilton Rating Scale for Depression (HAM-D-17) and the Clinical Global Improvement (CGI) scale. Nefazodone showed greater efficacy than placebo and equal efficacy to imipramine in the treatment of depression in three of the four trials. In the fourth trial, neither nefazodone nor imipramine were superior to placebo. Both agents continued to maintain efficacy during long-term (continuation) treatment. The endpoint dosage for patients receiving nefazodone was in the 300 to 500 mg a day range. Patients treated with nefazodone reported fewer adverse effects and had fewer premature treatment terminations than patients treated with imipramine. The principal conclusion drawn from these trials is that nefazodone has efficacy equal to that of imipramine in the treatment of major depressive disorder. Additionally, nefazodone is well tolerated, with fewer reported adverse effects than with imipramine.

Nefazodone has been directly compared with sertraline (Zoloft) and paroxetine (Paxil) in the treatment of major depressive disorder. In the trial comparing nefazodone with sertraline, 166 patients that fulfilled the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) criteria for nonpsychotic major depression were randomized to treatment with either nefazodone (100 to 600 mg a day) or sertraline (50 to 200 mg a day). Patients were treated for 6 weeks with active agent and were assessed with the HAM-D-17, the CGI, and a sexual function questionnaire. The mean daily dosage for each agent at the end of the study was 148 mg for sertraline and 456 mg for nefazodone. Both agents were judged efficacious in the treatment of depression, with no statistically significant difference in their antidepressant action. Treatment with sertraline was reported to have a negative impact on sexual functioning and satisfaction in men and women; patients treated with nefazodone did not report adverse sexual consequences during the trial.

The trial comparing nefazodone with paroxetine was carried out in 20 centers in the United Kingdom and Ireland. In this trial, 206 patients that fulfilled DSM-III-R criteria for moderate-to-severe nonpsychotic depressive disorders were randomized to treatment with either nefazodone or paroxetine. One limitation of this trial is that patients deemed to be at risk for suicide were specifically excluded from the study. The battery of outcome measures was similar to that used in the sertraline comparator study cited above, though sexual functioning was not assessed directly. Both treatment groups had virtually identical outcomes—significant improvement in depressive symptoms. Both agents were concluded to possess similar efficacy in the treatment of major depression. Both treatment groups had virtually identical rates of discontinuation due to adverse events; 15 (14 percent) of the nefazodone-treated patients, and 13 (13 percent) of the paroxetine-treated patients dropped out of the trial. These studies indicate that nefazodone is as effective in the treatment of major depression as the two well-studied SSRIs sertraline and paroxetine. All three agents carried similar rates of adverse events, but nefazodone seems to cause fewer adverse sexual effects than the other agents, especially sertraline.

Nefazodone has also been compared directly with amitriptyline (Elavil) in the treatment of depression. In this trial, 106 inpatients with a diagnosis of major depression were randomized to treatment with nefazodone or amitriptyline in a 6-week, double-blind trial. Treatment response was measured with multiple instruments including the Montgomery-Asberg scale, Hamilton Rating Scale for Depression (HAM-D), CGI, and the Patient Global Assessment scale. Amitriptyline demonstrated significant superiority to nefazodone in the treatment of depression. However, this study differed from those cited above demonstrating that nefazodone had efficacy equal to that of imipramine in terms of the final modal dosage of nefazodone. In this study comparing nefazodone with amitriptyline, the modal dosage of nefazodone was 242 mg a day, while in the studies comparing nefazodone with imipramine, the dosage range of nefazodone was 300 to 500 mg a day. Based on these data, it seems reasonable to conclude that the lack of efficacy reported in this study is merely due to inadequate dosage of nefazodone.

The efficacy of nefazodone as a treatment for various syndromal types of depression was assessed by analyzing the pooled data from four placebo-controlled trials. The efficacy of nefazodone in the treatment of severe depression, melancholic depression, and recurrent depression was determined with this analysis. Patients were defined as having severe depression by the CGI. Patients were defined as having melancholic features or recurrent depression by DSM-III-R criteria. This pooled analysis revealed that nefazodone was effective in the treatment of all three depressive subtypes. The mean endpoint dose of nefazodone for the patients in this analysis was 379 mg a day.

The cost-effectiveness of nefazodone as a treatment for depression has been analyzed in two different studies. Nefazodone was compared with either imipramine, fluoxetine (Prozac) or a step regimen in which patients who failed on imipramine were treated with nefazodone, with a model that represented “ideal” primary care practice. The measures that were compared for the three regimens included total lifetime medical cost, quality-adjusted life years, and cost per quality-adjusted life years gained. The lifetime medical cost of treatment with nefazodone was $16,669, for imipramine $15,348, for the step regimen $16,061, and for fluoxetine, $16,998. Treatment with nefazodone resulted in the largest gain of in quality-adjusted life years 14.64; treatment with imipramine yielded 14.32, the step regimen 14.40, and fluoxetine 14.58. Thus, nefazodone is a cost-effective treatment for depression in comparison to imipramine or fluoxetine. The analysis also reveals that fluoxetine treatment is cost-effective, though slightly more expensive and slightly less effective than nefazodone. In a second analysis, the costs of nefazodone treatment were compared with those of imipramine over the course of a 1-year continuation-phase of a double-blind trial comparing the efficacy of the two agents. The clinical trial was conducted in the United Kingdom. The average cost per successfully treated patient with nefazodone was 254 British pounds; the cost for imipramine was 323 British pounds. Thus, for patients carried through the continuation phase of the protocol (who by definition are those who responded to the particular agent during the initial 8-week phase), the cost of treatment with nefazodone is lower than that of imipramine.

Panic and Other Anxiety Disorders Nefazodone has been evaluated as a treatment of panic and anxiety symptoms both in the setting of panic and anxiety symptoms secondary to or comorbid with a diagnosis of major depressive disorder and as syndromal noncomorbid entities. The results of two randomized, placebo-controlled studies of nefazodone in the treatment of depression were retrospectively analyzed for specific actions on symptoms of anxiety. The patients enrolled in the trials had primary diagnoses of major depressive disorder, and the trials were carried out in both primary care and psychiatric clinic settings. Nefazodone was found effective in treating the anxiety symptoms associated with depression. The active comparator, imipramine, was also effective in treating these symptoms, though not as effective as nefazodone. Retrospective, blinded record review revealed that a subgroup of the patients treated in the psychiatric clinic was suffering from comorbid panic disorder. Patients from this subgroup who received nefazodone had significant improvement in their panic symptoms including relief of panic and phobic anxiety. Patients from the panic-prone subgroup who received imipramine did not differ from those that received placebo. Thus in this particular cohort of patients with comorbid major depressive disorder and panic disorder, nefazodone was superior to both placebo and imipramine, particularly in the relief of panic symptoms.

The results of six randomized, double-blind trials of nefazodone as a treatment for major depressive disorder were subjected to a meta-analysis to assess its efficacy as a treatment for anxiety symptoms associated with depression. The patients fulfilled criteria for major depression and had HAM-D scores of 18 or above. The retrospective analysis was carried out on 817 patients: 345 received placebo, 288 received imipramine, and 184 received nefazodone. Treatment with either imipramine or nefazodone was associated with statistically significant improvement in the Hamilton Rating Scale for Anxiety (HAM-A), the HAM-D anxiety factor,

HAM-D psychic anxiety item, and the HAM-D agitation item. The patients who received nefazodone experienced significant improvement in somatic anxiety (HAM-D, item 11), while those that received imipramine did not differ from those that received placebo. Nefazodone treatment was associated with more-rapid improvement in agitation (HAM-D, item 9) than treatment with either imipramine or placebo. The improvement in this item was discernible after as little as 1 week of nefazodone treatment.

Nefazodone has been evaluated in open-label trials as a treatment for panic disorder, generalized anxiety disorder, and obsessive-compulsive disorder. Fourteen patients who fulfilled DSM-III-R criteria for panic disorder were entered in an 8-week open-label trial of nefazodone. Patients were not excluded for having comorbid diagnoses of major depression, dysthymia, generalized anxiety disorder, or depression not otherwise specified. A flexible dosage regimen was used with a range of 200 to 600 mg a day. Panic attack frequency and severity and the CGI were used to assess outcome, and 10 of 14 patients were rated as much or very much improved after 8 weeks of treatment. The frequency of panic attacks decreased from a mean of 5.4 pretreatment to 2.1 after 8 weeks of treatment. Improvements were reported in panic attack severity, HAM-D, HAM-A, CGI, and the Sheehan Disability Scale after 4 weeks of active treatment. Six of the patients initially enrolled in the study had “pure” panic or only minor depressive symptoms; five improved. Eight of the patients had diagnoses of comorbid major depression; five improved. Five patients had comorbid generalized anxiety disorder; three improved. No patient prematurely terminated treatment because of adverse reactions.

Twenty-one patients who fulfilled DSM-IV criteria for generalized anxiety disorder were enrolled in an open-label trial of nefazodone. The patients were treated for 8 weeks, and outcome measures included the HAM-A and CGI scales. Of the 21 patients initially enrolled, 15 completed the entire 8-week trial. Of these, 12 of 15 were rated very much or much improved, 1 minimally improved, and 2 unchanged; no patients were rated worse. The results of the open-label trials and the post hoc analyses of the original clinical trials data sets indicate that nefazodone may be effective in the treatment of panic, generalized anxiety disorder, and other anxiety disorders. This seems to be true in the setting of panic and anxiety comorbid to major depression and in anxiety disorders occurring independently of depression. These preliminary observations need to be confirmed with double-blind, placebo-controlled trials specifically designed to test nefazodone as a treatment for panic disorder, generalized anxiety disorder, and other anxiety disorders.

There is one small open-label trial of nefazodone in the treatment of obsessive-compulsive disorder. This study was conducted in depressed patients with comorbid obsessive-compulsive disorder. Of the 15 patients who completed the depression trial, 9 had comorbid obsessive-compulsive disorder as defined by DSM-III-R criteria. The patients who completed the 8-week trial showed significant improvement in depressive and anxiety symptoms, regardless of whether they had comorbid obsessive-compulsive disorder. The author reported a trend toward efficacy in reducing obsessional thoughts in patients with obsessive-compulsive disorder, but this did not attain statistical significance. The results of this study do not support the use of nefazodone in the treatment of obsessive-compulsive disorder; whether or not this agent has such efficacy must await the completion of double-blind, placebo-controlled studies.

Premenstrual Syndrome Nefazodone has been evaluated as a treatment for premenstrual syndrome (PMS), premenstrual dysphoric disorder, and depression comorbid with PMS. In this open-label trial, patients with PMS without comorbid depression and patients with PMS with comorbid depression or dysthymia were treated with nefazodone. The patients were treated for 8 weeks with a flexible dosage range from 100 to 600 mg a day with twice-daily dosing. Both groups experienced significant improvement in premenstrual symptoms by the end of the first treated menstrual cycle. The average daily dosage at the end of the first cycle was 245 mg (range, 100 to 400 mg). Of the 54 patients initially enrolled in the study, 47 completed the full 8 weeks of treatment, with a mean daily dosage at the end of 8 weeks of 319 mg. This promising preliminary result serves as an impetus for placebo-controlled, double-blind studies specifically designed to test the efficacy of nefazodone as a treatment for PMS as well as PMS comorbid with depression.

Pain Although there are no references to the clinical utility of nefazodone in the treatment of pain, there are preclinical data that bear on this issue. Given that nefazodone has some structural and neurochemical similarity to trazodone, it might be expected to have similar antinocioceptive properties. Nefazodone demonstrates analgesic properties in certain animal models but not in others. It potentiates morphine-induced analgesia in a number of animal models. Interestingly, nefazodone does not affect the lethal dose (LD_{50}) for morphine, nor does it affect the morphine's action on gastrointestinal motility. This concatenation of observations suggests that nefazodone potentiates the analgesic action of morphine without enhancing its adverse effects. Nefazodone may be a useful adjunct in the treatment of chronic pain. Clearly controlled clinical trials are needed.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

General Nefazodone has been reported to cause relatively few adverse events. Approximately 3500 patients have received nefazodone in controlled clinical trials, including 1300 patients in long-term (1 year or more) continuation treatment. The incidence of adverse effects and the frequency of their occurrence in patients receiving nefazodone can be directly compared with those in patients receiving other antidepressants (tricyclic drugs, SSRIs) and placebo. In comparison to patients receiving placebo, the most frequently reported adverse effects of patients being treated with nefazodone are nausea (21 versus 14 percent), somnolence (19 versus 13 percent), dry mouth (19 versus 13 percent), dizziness (12 versus 6 percent), constipation (11 versus 7 percent), asthenia (11 versus 6 percent), lightheadedness (10 versus 4 percent), and amblyopia or blurred vision (6 versus 3 percent).

P>**Central Nervous System** The adverse events most frequently referable to the nervous system experienced by patients receiving nefazodone are somnolence, dizziness, lightheadedness, asthenia, and amblyopia. While potentially bothersome to individual patients, these side effects are considered benign. Nefazodone does not appear to alter the seizure threshold, and it is not associated with an increase in the incidence of seizures. There are no reports of its use in patients receiving electroconvulsive therapy (ECT).

Nefazodone has interesting effects on sleep architecture. Nefazodone treatment tends to increase the density of rapid eye movement (REM) sleep in both healthy volunteers and depressed patients. In a study of eight healthy volunteers, subjects received 100 mg of nefazodone twice daily for 7 days. Polysomnographs were recorded after 1 and 7 days of treatment. Treatment with nefazodone increased REM sleep time in both the 1-day and 7-day recordings. The largest increase in REM sleep time occurred after 7 days of treatment. Treatment with nefazodone did not alter other parameters of REM sleep (i.e., REM latency or number of REM periods). Treatment with nefazodone was also associated with a decrease in the wake-after-sleep-onset measure and an increase in the percentage of actual time asleep. In a double-blind, placebo-controlled, crossover designed study, 12 healthy volunteers were treated with either trazodone, buspirone (BuSpar), or nefazodone and subjected to multiple sleep studies. Nefazodone treatment increased REM sleep time, while the other two agents suppressed REM sleep. All three agents had minimal impact on the other sleep stages. Nefazodone was reported to have no impact on sleep architecture or REM sleep time in a larger study. In this double-blind, placebo-controlled study, 37 healthy volunteers were randomized to receive either paroxetine (30 mg a day) or nefazodone (400 mg a day) for a total 16 days of treatment. Polysomnographs were recorded after 1 and 16 days of treatment. Treatment with paroxetine reduced total REM sleep and increased REM latency. Paroxetine treatment was also associated with reduced actual sleep time and sleep efficiency. Treatment with nefazodone resulted in no change in REM sleep or any of the measures of sleep continuity. Thus in healthy subjects, nefazodone has little impact on REM sleep and may actually increase the total REM time in comparison to other antidepressant agents, which tend to suppress REM density and increase REM latency.

Nefazodone has been reported to increase REM sleep time in patients with major depressive disorder. In an open-label trial, 10 patients who were suffering from major depression were treated with nefazodone (400 to 600 mg a day) and subjected to polysomnographic study at baseline and during the treatment phase of the trial. All of the patients enrolled in the study were deemed to be nefazodone responders because they all experienced a 50 percent reduction in HAM-D scores from pretreatment baselines. Treatment with nefazodone neither increased REM latency nor suppressed REM sleep. There was a trend toward increased REM sleep during the second REM period and a trend toward decreased REM during the third period. Treatment with nefazodone was associated with improved sleep continuity as measured by decreased wake and movement time and increased stage 2 sleep. Nefazodone clearly improves sleep continuity in patients suffering from major depression, has no effect on REM latency, does not suppress REM sleep, and may even increase total REM time.

The mechanism by which nefazodone increases REM sleep remains obscure. Generally, agents that potentiate cholinergic functioning increase REM sleep, but nefazodone does not have cholinergic agonist properties. One possible explanation is that nefazodone affects the cells of the dorsolateral and pedunculopontine nuclei. These cholinergic cells are intimately involved in generating the pontine-geniculate-occipital waves associated with REM sleep. These cell groups also have a high density of $5\text{-}HT_{2A}$ receptors. Antagonizing the $5\text{-}HT_{2A}$ receptors on these cells may potentiate their cholinergic outflow, resulting in increased REM sleep.

The effect of nefazodone on daytime sleepiness, highway driving, and cognitive functions was determined in healthy adult and elderly volunteers. This was a double-blind, placebo-controlled, crossover designed study in which the subjects received either placebo, imipramine (50 mg twice daily), or nefazodone (100 or 200 mg twice daily). The subjects were tested with an “over-the-road” driving test, a battery of psychomotor tests, and multiple sleep latency tests. The most prominent result of this study was that the subjects treated with imipramine had decreased lateral position control in the driving test after the first dose. This effect disappeared after 7 days of treatment. In comparison, 1 day of nefazodone administration at either dosage had no effect on the driving performance test. Seven days of nefazodone treatment at the 200-mg dosage was associated with slight impairment in the lateral position control in the driving test. Imipramine treatment was

associated with minor impairment in the psychomotor test battery after 1 and 7 days. Nefazodone treatment for 7 days at the higher dosage was associated with minor impairment in cognitive and memory function. Neither agent was noted to increase daytime sleepiness. Two main conclusions seem warranted: (1) neither drug has a deleterious impact upon driving or cognitive performance and (2) although somnolence is a frequently reported side effect of nefazodone, its use is not associated with an actual increase in daytime sleepiness as revealed by the multiple sleep latency test.

The effect of nefazodone on a number of psychomotor and cognitive performance measures was determined in a study of 12 healthy volunteers. In this study, patients were randomized to receive either nefazodone (200 or 400 mg a day), imipramine (150 mg a day), or placebo for 8 days. The subjects were tested on days 1, 7, and 8 of treatment. Nefazodone treatment was associated with a dose-dependent improvement in psychomotor performance and complex memory tasks. In contrast, treatment with imipramine impaired performance of these tasks. Subjects who received either active compound reported subjective changes in alertness and other bodily symptoms. The subjects received a measured dose of alcohol on the eighth day of treatment, followed by the battery of tests. Treatment with nefazodone did not potentiate the sedative-hypnotic effects of alcohol, whereas imipramine tended to enhance the impact of alcohol on psychomotor performance and memory tasks. The lack of impact of nefazodone on the psychomotor and memory tasks is hypothesized to be due to its lack of anticholinergic activity.

Genitourinary System Unlike trazodone, nefazodone has not been reported to cause priapism. The reason for this is not known, but one hypothesis attributes it to the fact that nefazodone has a lower affinity for peripheral a_1-receptors than does trazodone. Trazodone is thought to cause priapism through direct action on a_1-receptors on the vascular tissues in the penis and clitoris. Nefazodone does increase total nighttime tumescence in patients while they are asleep, but this is apparently due to its impact on REM sleep and not due to peripheral adrenergic activity. In a study in which the effects of nefazodone on sleep were compared with those of trazodone and buspirone, nefazodone was noted to increase REM sleep time. Subjects treated with nefazodone exhibited a concomitant increase in nighttime tumescence that correlated directly with REM sleep time. In subjects treated with trazodone, tumescence was increased because of an increase in the time to detumescence. Trazodone-treated subjects had normal initiation of erection with the onset of REM sleep, but the erection persisted past the point of termination of REM sleep because of a failure of detumescence. This increased time to detumescence caused by trazodone is likely due to its direct action on peripheral a_1-receptors.

The incidence of sexual adverse effects of any kind is low for nefazodone in comparison to other agents. In at least one clinical trial in which nefazodone was compared with an SSRI, sertraline, patients who received nefazodone reported significantly better and more satisfying sexual functioning than did those receiving the SSRI. This was a consistent finding, reported by both male and female patients.

Cardiovascular System Treatment with nefazodone is associated with an increased risk for the occurrence of hypotensive episodes. A pooled analysis of the vital sign data generated during the placebo-controlled premarketing studies revealed that 5.1 percent of nefazodone-treated patients experienced a hypotensive episode, compared with 2.5 percent of the placebo-treated patients. Interestingly, there was no difference between nefazodone and placebo-treated patients in the report of syncopal episodes. The incidence of postural hypotension recorded in the premarketing data base was 2.8 percent of nefazodone-treated patients, 10.9 percent of those treated with tricyclic drugs, 1.1 percent of the SSRI treated, and 0.8 percent of placebo-treated patients. Care should be used when prescribing nefazodone to patients with a condition that might be worsened by hypotensive episodes. This includes patients with known cardiovascular and cerebrovascular disease, those at risk for dehydration or hypovolemia, and those taking antihypertensive agents. Treatment with nefazodone during the premarketing trials was not associated with the development of clinically significant electrocardiogram (ECG) abnormalities. The nefazodone-treated patients had a slightly higher incidence of sinus bradycardia than those who received placebo. Sinus bradycardia was defined as a heart rate below 50 beats per minute (bpm) and a drop of 15 bpm from pretreatment baseline. This event occurred in 1.5 percent of the nefazodone-treated patients and only 0.4 percent of those receiving placebo. Because patients with recent myocardial infarction or unstable heart disease were excluded from the premarket clinical trials, it is not possible to assess the risks of nefazodone in this patient population.

Other Systems Nefazodone has no carcinogenic or mutagenic activity in an array of preclinical assay systems. Nefazodone slightly decreased fertility in a rat assay system at doses three times the maximal human dose but had no effect at dosages that approximated the therapeutic range in humans. Nefazodone's teratogenic potential has been evaluated in rabbit and rat systems. Treatment at doses five to six times the maximal human dose resulted in no malformations. In the rat, dosages five times the human dose increased early pup mortality and decreased pup weights. This effect disappeared at a dose 1.3 times the maximal human dose. There are no systematic studies of nefazodone use in pregnant women. This agent is classified in the category C risk group, indicating that it should be used in pregnancy only if its potential benefits outweigh the potential risk to the fetus. It is not known whether nefazodone or its metabolites enter human breast milk.

Studies in both monkeys and humans indicate that nefazodone has no abuse liability.

PRECAUTIONS AND ADVERSE EFFECTS

Overdose The experience with nefazodone in overdose during the premarketing trials was very limited. A total of seven overdose incidents were reported during these trials. There were no fatalities in patients after nefazodone overdose alone or in combination with other agents. The reported dose range of nefazodone ingested in overdose was 1000 to 11,200 mg. Symptoms of nefazodone overdose include nausea, vomiting, sedation, and somnolence. There are no specific treatments for nefazodone overdose. Treatment should be supportive and directed at particular symptoms, specifically at maintaining an adequate airway and cardiovascular function.

Use in Medically Compromised Patients The pharmacokinetics of nefazodone have been evaluated in patients who suffered from hepatic impairment. In one study, the metabolism of nefazodone in patients with biopsy-proven hepatic cirrhosis was compared with that of healthy controls matched by age, sex, and weight. Acute and chronic exposure paradigms were used to determine the effects of hepatic cirrhosis on the immediate and steady-state concentrations of nefazodone and its metabolites. The concentration of nefazodone and hydroxynefazodone after a single oral dose in subjects with cirrhosis was approximately twice that in controls. At steady state, the concentration of nefazodone and hydroxynefazodone in the subjects with cirrhosis was approximately 25 percent greater than that in the controls. The concentration of m-CPP in the subjects with cirrhosis was two- to threefold higher than that in controls, while there was no difference in the concentration of the triazoledione metabolite between the two groups. In a separate study, the kinetics of nefazodone, hydroxynefazodone, and m-CPP were compared in healthy subjects and those with mild and severe hepatic impairment. The results indicated that mild liver disease does not affect the kinetics of nefazodone metabolism, but severe impairment does. The presence of liver disease is clearly not a contraindication to nefazodone use. The starting dose should be lowered as should the steady-state dose in patients with hepatic insufficiency.

The kinetics of nefazodone, hydroxynefazodone, and m-CPP in healthy elderly subjects (over 65 years) were compared with those in healthy younger subjects (18 to 40 years). The study included four separate groups: younger men, younger women, elderly men, and elderly women. The CYP 2D6 activity was determined for all of the subjects, and all of the study participants were classified as extensive metabolizers. The peak plasma concentration of nefazodone and hydroxynefazodone was approximately twofold higher in the elderly subjects than in the younger subjects. The levels of nefazodone and hydroxynefazodone at steady state were approximately 50 percent higher in the elderly women than in the other three groups. The premarketing trials included 488 elderly individuals. These patients did not experience increased numbers or types of adverse events. In particular, the elderly group had the same rate of cardiovascular events as did the nonelderly group. Based on these results, nefazodone appears to be quite safe to use in elderly patients. The manufacturer recommends initiating therapy in elderly persons at half the dose recommended for nonelderly persons. The dosage in elderly persons should be increased into the same range recommended for nonelderly patients, but the usual precautions observed in treating elderly patients (slow dosage escalation) should be observed.

Because patients with previous myocardial infarction or with unstable cardiac disease were excluded from the premarket clinical trials, there is no information available on the use of nefazodone in this population. In a study of patients suffering from renal disease (with creatinine clearances from 7 to 60 mL a minute per 1.73 m^2), renal impairment had no effect on steady-state concentrations of nefazodone. No information is available on the use of nefazodone in patients with other chronic or systemic diseases (neurological, metabolic, oncological). Given the lack of systematic study in these patients populations, caution should be exercised when considering the use of nefazodone or other antidepressant agents in medically ill patients.

Drug-Drug Interactions The identification and characterization of specific isoenzymes of the cytochrome P450 system has greatly improved the ability to predict drug-drug interactions. Knowledge of which drugs are metabolized by specific isoenzymes, coupled to the knowledge of which drugs inhibit specific isoenzymes allows treatment regimens to be tailored to the needs of the individual patient. This knowledge also helps prevent adverse events arising from pharmacokinetic interactions between different therapeutic agents.

Nefazodone, hydroxynefazodone, and *p*-hydroxynefazodone all inhibit the CYP 3A4 in vitro and in vivo. Nefazodone and its metabolites are very weak inhibitors of CYP 2D6 and have no activity on CYP 1A2. Caution must be exercised when using nefazodone in combination with other drugs metabolized by CYP 3A4. Specifically, nefazodone should not be used in combination with terfenadine (Seldane), astemizole (Hismanal), or cisapride (Propulsid), as all three of these agents

are substrates of CYP 3A4. Increased concentrations of these agents are associated with torsade de pointes (QT prolongation and potentially lethal ventricular arrhythmias). The interaction of nefazodone with these three agents has not been directly observed in human subjects, but the interaction between other inhibitors of CYP 3A4 (ketoconazole [Nizarol]; erythromycin) and these agents has been shown to cause this type of ventricular tachyarrhythmia.

The specific in vivo interaction between nefazodone and a number of other drugs has been directly assayed in a series of clinical trials. The interaction between nefazodone and triazolam (Halcion), which is metabolized by CYP 3A4, was assessed in a placebo-controlled, double-blind study in healthy male volunteers. The immediate and long-term administration of triazolam did not affect the pharmacokinetics of nefazodone. Twice-daily administration of nefazodone for 7 days significantly affected the pharmacokinetics of triazolam, resulting in a 60 percent increase in circulating triazolam concentrations, associated with pronounced alterations in measures of psychomotor performance and sedation. Similarly designed studies were used to assess the interaction between nefazodone and alprazolam (Xanax), which is also metabolized by CYP 3A4. Coadministration of the two agents did not affect the pharmacokinetics of nefazodone, hydroxynefazodone, or triazoldione, but the steady-state concentration of m-CPP was increased threefold and the half-life ($t_{1/2}$) of m-CPP was increased twofold; the steady-state concentration of alprazolam was increased twofold. Coadministration of nefazodone and alprazolam resulted in poorer psychomotor performance and increased sedation, compared with alprazolam alone. The interaction between nefazodone and lorazepam was evaluated in similarly designed studies. Lorazepam (Ativan) is metabolized via conjugation, not hydroxylation, and is probably not a substrate of CYP 3A4. Coadministration of lorazepam and nefazodone had no effect on the pharmacokinetics of either agent. Steady-state levels of m-CPP in the coadministration group were slightly decreased. There were no impairments in psychomotor performance, memory, or sedation in the subjects who received nefazodone and lorazepam together. Thus, when nefazodone is coadministered with benzodiazepines that are metabolized via the CYP 3A4 pathway (triazolam, alprazolam), the dosage of the benzodiazepine should be decreased, whereas no dosage adjustment is necessary when benzodiazepines metabolized via direct conjugation (lorazepam) are coadministered with this antidepressant.

The pharmacokinetics and pharmacodynamics of theophylline, which is metabolized via the CYP 1A2 pathway, were not affected by coadministration of nefazodone. Likewise, there were no changes in the pharmacokinetics of nefazodone (Serzone) in the coadministration group. Coadministration of warfarin (which is metabolized via CYP 2C) and nefazodone resulted in a 12 percent decrease in warfarin C_{max}, but no change in prothrombin ratios or bleeding times. Therefore this appears to be a clinically insignificant interaction. The pharmacokinetics of nefazodone were not affected by the coadministration of warfarin. A modest pharmacokinetic interaction was observed between nefazodone and haloperidol (Haldol), which is metabolized by CYP 2D6. In these studies the AUC for a 5-mg oral dose of haloperidol was increased 36 percent in subjects with a steady-state plasma concentration of nefazodone. Coadministration of haloperidol did not affect the pharmacokinetics of nefazodone or its metabolites. It does appear that nefazodone and haloperidol are safe to use in combination, but in this situation consideration should be given to lowering the dose of haloperidol, especially in patients known to be prone to antipsychotic drug–induced adverse effects. There is no apparent pharmacokinetic interaction between nefazodone and the histamine type 2 receptor (H_2) inhibitor cimetidine (Tagamet).

Coadministration of nefazodone and digoxin (Lanoxin) resulted in a 29 percent increase in C_{max} and a 15 percent increase in AUC for digoxin. There was no effect on the pharmacokinetics of nefazodone. These agents are considered safe to use in combination, but downward adjustment in digoxin dosage is recommended as is monitoring of digoxin concentrations in plasma. Coadministration of propranolol and nefazodone resulted in a 29 percent reduction in C_{max} and 14 percent reduction in AUC for propranolol (Inderal). There were no clinically significant effects on vital signs or ECG results. The changes in propranolol pharmacokinetics are considered clinically insignificant; no particular dosage adjustment is recommended for this combination. Coadministration of propranolol did not affect the pharmacokinetics of nefazodone.

The interaction between nefazodone and monamine oxidase inhibitors (MAOIs) has not been systematically studied in experimental animals or human subjects. Significant interactions between MAOIs and both SSRIs and tricyclic drugs have been reported. The serotonin syndrome, characterized by hyperthermia, autonomic instability, rigidity, myoclonus, agitation, delirium, and coma has been reported to occur from the interaction of SSRIs and MAOIs. Additionally, severe hyperthermia and potentially fatal seizures have been reported to occur from the combination of tricyclic drugs and MAOIs. The manufacturer of nefazodone recommends that this agent not be used in combination with MAOIs, that at least 14 days elapse between terminating treatment with an MAOI and initiating treatment with nefazodone, and that at least 1 week elapse between terminating treatment with nefazodone and initiating treatment with an MAOI.

DOSAGE AND ADMINISTRATION

Nefazodone is indicated for the treatment of major depressive disorder. Labeling for nefazodone suggests initiating therapy in healthy adults at a dosage of 100 mg twice a day. These authors recommend 50 mg twice a day as a starting dosage to diminish adverse events. The effective dosage range from all of the premarket trials is 300 to 600 mg a day. The dosage should be titrated upward in increments of 100 mg a day at weekly intervals. Dosages above 600 mg a day are not recommended. As with all antidepressants, several weeks of treatment in the therapeutic dose range may be required to realize a complete response. There are no data to suggest an optimal maintenance dose of nefazodone once a therapeutic response has been realized. Standard practice with other antidepressants suggests that the therapeutic dose should be continued as the maintenance dosage. The premarket data indicate that nefazodone at therapeutic dosages is safe for at least 1 year.

Treatment of elderly patients should also be initiated at 50 mg twice a day. Upward dosage titration should be adjusted to the individual needs of the elderly patient, generally with a slower rate of dosage escalation. The final therapeutic range for treatment of elderly patients may be similar to that in nonelderly patients, but dosage adjustment should be based on clinical assessment of the individual patient. Similar recommendations pertain to treating medically debilitated patients and patients with liver disease.

Nefazodone tablets are supplied in 50-mg, 100-mg, 150-mg, 200-mg, and 250-mg sizes. The 100-mg and 150-mg tablets are scored so that they can be easily broken in half. This allows finer adjustments in the dosing regimen.

SUGGESTED CROSS-REFERENCES

Section 1.4 discusses monoamine neurotransmitters. Chapter 14 discusses mood disorders, including somatic treatments in Section 14.7. Chapter 15 discusses anxiety disorders. Section 31.11 discusses bupropion; Section 31.19 discusses mirtazapine; Section 31.20 discusses MAOIs; Section 31.24, SSRIs; and Section 31.29 discusses trazodone.

SECTION REFERENCES

Armitage R, Yonkers K, Cole D, Rush AJ: A multicenter, double-blind comparison of the effects of nefazodone and fluoxetine on sleep architecture and quality of sleep in depressed patients. J Clin Psychopharmacol *17:*161, 1997.

Baldwin DS, Hawley CJ, Abed RT, Maragakis BP, Cox J, Buckingham SA, Pover GH, Ascher A: A multicenter double-blind comparison of nefazodone and paroxetine in the treatment of outpatients with moderate-to-severe depression. J Clin Psychiatry *57*(Suppl):46, 1996.

Feiger A, Kiev A, Shrivastava RK, Wisselink PG, Wilcox CS: Nefazodone versus sertraline in outpatients with major depression: Focus on efficacy, tolerability, and effects on sexual function and satisfaction. J Clin Psychiatry *57*(Suppl):53, 1996.

Feighner J, Targum SD, Bennett ME, Roberts DL, Kensler TT, D'Amico MF, Hardy SA: A double-blind, placebo-controlled trial of nefazodone in the treatment of patients hospitalized for major depression. J Clin Psychiatry *59:*246, 1998.

Goldberg RJ: Antidepressant use in the elderly. Current status of nefazodone, venlafaxine and moclobemide. Drugs Aging *11:*119, 1997.

*Greenberg WM: Visual field “shimmering” associated with nefazodone. J Clin Psychiatry *60:*124, 1999.

*Greene DS, Barbhaiya RH: Clinical pharmacokinetics of nefazodone. Clin Pharmacokinet *33:*260, 1997.

Lader MH: Tolerability and safety: Essentials in antidepressant pharmacotherapy. J Clin Psychiatry *57*(Suppl):39, 1996.

Marcus RN, Mendels J: Nefazodone in the treatment of severe, melancholic, and recurrent depression. J Clin Psychiatry *57*(Suppl):19, 1996.

Montgomery SA, Brown RE, Clark M: Economic analysis of treating depression with nefazodone v. imipramine. Br J Psychiatry *168:*768, 1996.

*Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry *153:*11, 1996.

Owens MJ, Morgan WN, Plott SJ, Nemeroff CB: Neurotransmitter receptor binding profile of antidepressants. J Pharmacol Exp Ther *283:*1305, 1997.

Revicki DA, Brown RE, Keller MB, Gonzales J, Culpepper L, Hales RE: Cost-effectiveness of newer antidepressants compared with tricyclic antidepressants in managed care settings. J Clin Psychiatry *58:*47, 1997.

*Rickels K, Robinson DS, Schweizer E, Marcus RN, Roberts DL: Nefazodone: Aspects of efficacy. J Clin Psychiatry *56*(Suppl):43, 1995.

*Robinson DS, Marcus RN, Archibald DG, Hardy SA: Therapeutic dose range of nefazodone in the treatment of major depression. J Clin Psychiatry *57*(Suppl):6, 1996.

*Robinson DS, Roberts DL, Smith JM, Stringfellow JC, Kaplita SB, Seminara JA, Marcus RN: The safety profile of nefazodone. J Clin Psychiatry *57*(Suppl):31, 1996.

*Van Ameringen M, Mancini C, Oakman JM: Nefazodone in social phobia. J Clin Psychiatry *60:*96, 1999.

31.23 OPIOID AGONISTS

RICHARD S. SCHOTTENFELD, M.D., AND HERBERT D. KLEBER, M.D.

Methadone (Dolophine, Methadose), an orally effective and long-acting agonist at the μ-opioid receptor, was first synthesized at the end of the Second World War as a substitute analgesic for morphine. Methadone has been in use as a detoxification treatment from heroin or other opiate-like (opioid) drugs since the late 1940s and for maintenance treatment of opioid dependence since the mid 1960s. More recently, a second orally effective, synthetic μ-agonist, levomethadyl acetate (ORLAAM) has gained Food and Drug Administration (FDA) approval for the maintenance treatment of opioid dependence. Because of the long half-lives of its active metabolites, levomethadyl acetate (also known as L-a-acetylmethadol [LAAM] can be administered on a three times per week schedule, rather than daily, as is necessary for methadone. A third medication, buprenorphine (Buprenex), a long-acting partial μ-agonist, is currently under investigation for use in opioid detoxification and for the maintenance treatment of opioid dependence. Buprenorphine's profile of pharmacologic effects suggests that there may be a number of advantages to its use, including a decreased risk of overdose or abuse, greater ease of withdrawal, and the possibility of less than daily administration.

HISTORY

Harris Isbell and his colleagues at the Addiction Research Center in Lexington, Kentucky, were the first to establish the safety and efficacy of methadone for detoxification from heroin and other opioids. Substitution of the longer-acting opioid (methadone) for a shorter-acting opioid (e.g., heroin), followed by gradual tapering of methadone, led to fewer symptoms and less severe withdrawal than abrupt discontinuation or gradual tapering of a shorter-acting opioid. Because methadone is orally effective, additional advantages of its use include the ease and safety of its administration and the absence of reinforcement of injection drug use. While newer treatments for opioid withdrawal (e.g., clonidine [Catapres] and clonidine plus naltrexone [Revia]) can decrease the time needed to complete withdrawal and to initiate maintenance treatment with the opioid antagonist, naltrexone, methadone substitution and tapering remains the gold standard against which all other opioid withdrawal treatments continue to be measured.

The use of methadone for the maintenance treatment of opioid dependence was pioneered by Vincent Dole and Marie Nyswander at the Rockefeller Institute in New York City. These researchers established that once-a-day oral dosing with methadone reduced heroin craving associated with opioid withdrawal but did not lead to the fluctuations in mood or physical state associated with maintenance on heroin or other opioids with shorter half-lives. Prolonged daily maintenance at sufficiently high doses led to development of cross-tolerance to other opioids and attenuation or blocking of the euphoric effects of self-administered heroin or other illicit opioids. After the initial clinical trials demonstrated that methadone maintenance led to substantial and sustained reductions in illicit opioid use and criminal activity and dramatic improvement in social and vocational functioning, methadone was approved for the maintenance treatment of opioid dependence in carefully regulated treatment settings.

Despite more than three decades of scientific investigation and clinical use conclusively demonstrating its safety and efficacy, however, methadone maintenance continues to arouse community and political opposition and controversy. As a result, methadone maintenance is not available in many communities, funding for it has been limited, and program philosophy and policies of some funding and regulatory agencies limit the dosage and duration of treatment to less than optimal levels. Concerns about potential misuse of methadone maintenance or diversion of prescribed methadone for illicit use have also contributed to a regulatory approach that has limited access to methadone maintenance and kept this treatment out of traditional medical practice settings and confined to a relatively small number of specialized treatment programs.

While methadone maintenance has proved to be remarkably effective, development and approval of alternatives to methadone have been delayed. Levomethadyl acetate was extensively evaluated in the mid-1970s as an alternative to methadone for maintenance treatment. Despite promising results, its approval by the FDA was delayed until 1993, and it is still not available in many areas. Buprenorphine appears to offer some advantages over methadone for detoxification and maintenance treatment, but buprenorphine has not yet received FDA approval for either of these uses. Advantages and disadvantages and specific indications for use of buprenorphine compared with methadone or levomethadyl acetate have not been fully investigated.

Key Issues Key issues for agonist maintenance treatment of opiate dependence include choice of agent, optimal dose and duration of treatment, and the relevance to treatment outcome of program features such as treatment setting (e.g., traditional maintenance program, primary care clinic, physician office), counseling, and psychiatric, medical, and vocational services. Patients entering treatment often experience a wide range of medical, social, legal, and other drug-related problems, and high rates of psychiatric comorbidity and concurrent alcohol, cocaine, or other substance abuse or dependence also pose special challenges for treatment. The use of maintenance treatment during pregnancy, as well as its role in limiting transmission of human immunodeficiency virus (HIV) and facilitating medical treatment of acquired immune deficiency syndrome (AIDS), tuberculosis, endocarditis, or other severe complications of injection drug use, also merit special attention. Finally, many federal and state regulations govern agonist maintenance treatment programs, and there is considerable interest in revising these regulations to expand treatment access and permit greater program flexibility and tailoring of treatment to individual needs. Restrictions on take-home doses of methadone need to be considered in relation to their impact on treatment and the potential for diversion of methadone for illicit use. Although the success of maintenance treatment for heroin addiction generally depends on the ability to engage patients in a comprehensive rehabilitation program, some patients may not require additional services to benefit fully from agonist maintenance.

CHEMISTRY

The structural formulas of methadone (4,4-diphenyl-6-dimethylamino-heptanone-3hydrochloride), levomethadyl acetate (L-a-6-dimethylamine-4,4-diphenyl-3-heptyl acetate hydrochloride), and buprenorphine (17-(cycloprophylmethyl)-a-(1,1-dimethylethyl)-4,5-epoxy-18,19-dihydro-3-hydroxy-6-methoxy-a-methyl-6,14-etheneomorphinan-7-methanol hydrochloride) are presented in Figure 31.23-1. Methadone and levomethadyl acetate are chemically only remotely related to morphine and other opium alkaloid derivatives, but their pharmacological properties are qualitatively and quantitatively similar to those of morphine. Buprenorphine is a semisynthetic derivative of thebaine, which is one of the naturally occurring phenanthrene alkaloids derived from the poppy plant and contained in opium.

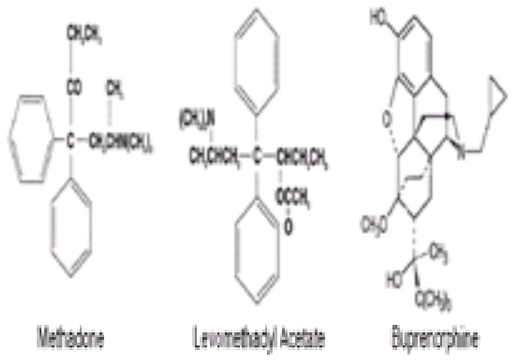

FIGURE 31.23-1 Molecular structures of opioid agonists.

Methadone and levomethadyl acetate act primarily as pure agonists at the μ-receptor. Methadone activity at the μ-receptor results almost entirely from the levorotatory isomer, which is 8 to 50 times more potent at these receptors than dextrorotatory isomer (D-methadone). Its analgesic potency when administered parenterally is approximately equivalent to that of morphine. Oral administration delays the onset of analgesia, reduces peak analgesia, and increases duration of analgesia. Levomethadyl acetate is extensively metabolized to nor-methadyl acetate and dinor-methadyl acetate. Both of these metabolites have longer half-lives and are more potent at the μ receptor than the parent compound. Buprenorphine is a high-affinity partial agonist at the μ-receptor and also acts as a potent k-antagonist. The slow rate of dissociation of buprenorphine from the μ-receptor may account for its prolonged duration of action.

PHARMACOLOGICAL ACTIONS

More than 90 percent of methadone administered orally is rapidly absorbed from the gastrointestinal tract. It is detectable in plasma within 30 minutes, and peak concentrations are reached within 2 to 6 hours. About 90 percent of methadone is reversibly bound to plasma and tissue proteins, including proteins in the brain and liver. Methadone is highly lipid soluble, crosses the blood-brain barrier, and is widely distributed throughout the body. Protein binding as well as the large volume of distribution of methadone result in shortened duration of analgesic activity (about 4 hours) following single-dose administration. Analgesic effects of methadone in nontolerant individuals are associated with methadone plasma concentrations above 30 to 60 ng/mL, whereas much higher concentrations (150 to 300 ng/mL) may be required for chronic pain analgesia. Following repeated administration of large doses of methadone (in patients who have developed tolerance), methadone accumulates in tissues, and reversible protein binding leads to relatively limited fluctuations in methadone levels throughout a 24-hour period. In patients maintained on once-daily administration of methadone (80 to 120 mg), peak concentrations typically range from 400 to 1600 ng/ml, while trough concentrations range from 150 to 1000 ng/ml.

The primary route of elimination of methadone is through biotransformation in the liver to inactive metabolites. Unbound methadone is extracted by the liver and undergoes *N*-demethylation, through the action of cytochrome P450 isoenzyme 3A4 (CYP 3A4), followed by immediate cyclization. Methadone metabolites as well as small amounts of unchanged methadone are excreted in the urine and bile. Acidification of urine increases urinary excretion of methadone and its major metabolites. The elimination half-life of methadone ranges from 24 to 36 hours.

Levomethadyl acetate is also rapidly absorbed from an oral solution, with peak blood concentrations occurring 1.5 to 2 hours after ingestion. Demethylation of levomethadyl acetate to nor-levomethadyl acetate occurs during first-pass liver metabolism; subsequent *N*-demethylation of nor-levomethadyl acetate leads to the formation of dinor-levomethadyl acetate. Nor-levomethadyl acetate is estimated to be 6 to 12 times and dinor-levomethadyl acetate 1.5 to 3 times as potent as at the μ-receptor. Thus, peak opioid effects following oral administration of levomethadyl acetate are delayed until it is metabolized to its more active metabolites, and these metabolites contribute to the efficacy and prolonged duration of activity of levomethadyl acetate. During steady-state administration of levomethadyl acetate (3 mg/kg a week on a three times per week schedule), observed half-lives of methadyl acetate, nor-levomethadyl acetate, and dinor-levomethadyl acetate were approximately 2.6 days, 2 days, and 4 days, respectively. Peak plasma concentrations for levomethadyl acetate and its active metabolites are at least five times higher than trough levels 72 hours after dose administration.

Buprenorphine has a variable bioavailability following sublingual administration but an extremely low bioavailability following oral administration, because of extensive intestinal and first-pass liver metabolism. It is currently available in the United States only in a formulation suitable for injection use, but sublingual liquid and tablet formulations have been investigated for treatment of opioid withdrawal and for maintenance treatment of opioid dependence. A sublingual tablet combining buprenorphine and naloxone (Narcan) is expected to have decreased abuse liability and may thus require less stringent regulations than those governing either methadone or methadyl acetate. Evaluation of the pharmacokinetics and metabolism of buprenorphine has been limited somewhat because of difficulties developing sufficiently sensitive analytic methods to measure the low plasma concentrations associated with its pharmacological effects. Following sublingual administration, there is a two-phase elimination profile, with plasma half-lives estimated at 3 to 5 hours and more than 24 hours. Because buprenorphine dissociates slowly from brain opioid receptors, it appears feasible to administer it during maintenance treatment on an alternate-day or thrice-weekly schedule.

Like other opioids, methadone, levomethadyl acetate, and buprenorphine cross the placental barrier, and long-term administration during pregnancy can result in opioid withdrawal in the neonate. Methadone maintenance during pregnancy, however, appears to be a safer option than continued heroin use by the pregnant addict. The severity of neonatal withdrawal is related to maternal methadone dose and duration of treatment. Use of methadyl acetate during pregnancy is not recommended because its safety has not been extensively studied, and there is a risk of delayed emergence of neonatal withdrawal. There is interest in the use of buprenorphine during pregnancy, because of the possibility that it may lead to less severe neonatal withdrawal, but its safety for use during pregnancy has not been sufficiently evaluated.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Early Studies of Maintenance Treatment Morphine maintenance clinics for the treatment of opioid-dependent patients that opened in many cities in the United States following passage of the Harrison Narcotics Act in 1914 were closed by the U.S. Treasury Department in the 1920s. Approximately 25,000 physicians were indicted for violating the Harrison Narcotic Act; an estimated 20,000 of them paid substantial fines, and 3000 were jailed. Opioid agonist maintenance treatment was not reinstated in the United States until the 1960s. The initial studies of methadone for maintenance treatment were conducted by Dole and Nyswander in collaboration with Mary Jeanne Kreek at Rockefeller Hospital. In six hospitalized narcotic addicts, stable doses of methadone (80 to 120 mg a day) resulted in a gradual cessation of the patients' fluctuations in mood and clinical state and cessation of their obsessive preoccupation with obtaining and administering drugs. Patients did not experience euphoria or sedation following methadone dose administration and appeared to function normally, including successful completion of complex tasks testing vigilance and concentration. With long-term maintenance, these investigators documented the development of tolerance to the euphoric effects of usual street doses of heroin.

Subsequently, 128 carefully selected patients were inducted onto high-dose methadone in the hospital setting and then maintained on methadone in the community following discharge. At 2 years, retention was excellent, with only 13 patients discharged for drug abuse or antisocial behavior. Of the 107 patients remaining in treatment, 71 percent were steadily employed or attending school, and none had relapsed to narcotic use. These findings were extraordinary compared with results of nonmethadone treatment of heroin addicts, whose relapse rates typically exceed 90 percent in the first year after hospital treatment. Similar improvements were reported for heroin addicts maintained on methadone after prison discharge, compared with those placed on waiting lists for methadone.

The initial success of methadone maintenance led to the widespread implementation of methadone maintenance programs nationally, greater variability in treatment outcome, and considerable controversy about both the ethics of substituting one opiate for another and the effectiveness of methadone maintenance programs. By 1993 an estimated 115,000 patients were receiving methadone maintenance treatment at any time in 759 programs in the United States.

Controlled Clinical Trials of Methadone, LAAM, and Buprenorphine Maintenance Three placebo-controlled random assignment clinical trials have established the efficacy of methadone maintenance treatment, and several additional clinical trials have demonstrated that its efficacy is dose dependent. In one study, 100 heroin addicts in Hong Kong were initially stabilized on methadone (60 mg a day), with half the patients undergoing blinded withdrawal from methadone at the rate of 1 mg a day followed by maintenance on placebo for the duration of the 32-week study. Dosage adjustments were allowed for the methadone-treated subjects, and both

groups had equivalent access to counseling and other services. Treatment retention at 32 weeks was 76 percent for the methadone group and 10 percent for the placebo group, and great reductions in illicit heroin use were found for subjects who remained in treatment. A second randomized study, in Sweden, comparing methadone maintenance with outpatient drug-free treatment also found significant benefits of methadone maintenance: at 2 years, 12 of 17 methadone patients were free of illicit drug use and employed or in school, while only 1 of 17 patients allocated to drug-free outpatient treatment was abstinent, 12 were addicted to heroin (including 3 who suffered major medical problems resulting from drug dependence), 2 were in prison, and 2 had died. The most recent placebo-controlled double blind clinical trial, comparing daily placebo and daily maintenance on methadone (20 or 50 mg) in 247 opioid-dependent patients, found significant dosage-dependent effects on both treatment retention and illicit opioid use during the 14-week stable maintenance period. Additional double-blind clinical trials comparing higher (50 to 80 mg) and lower (20 to 40 mg) daily methadone dosages generally also show superiority of the higher dosages with regard to treatment retention and reductions in illicit opioid use. These studies support recommendations that daily methadone doses exceeding 60 mg generally yield better results.

A number of clinical trials of levomethadyl acetate and buprenorphine also demonstrate the dosage-dependent efficacy of these medications for maintenance treatment. In one study, patients randomly assigned to the highest Monday-Wednesday-Friday levomethadyl acetate dosage (100, 100, and 140 mg) had the lowest rates of heroin use during treatment, patients assigned to the lowest dosage (25, 25, and 35 mg) had the highest rates of heroin use, and patients assigned to a mid-range dosage (50, 50, and 75 mg) had intermediate rates of use. Daily buprenorphine dosages of 8 to 16 mg sublingual liquid formulation have been found to reduce heroin use more than lower daily dosages. In meta-analyses of randomized clinical trials comparing methadone and levomethadyl acetate and comparing methadone and buprenorphine, outcomes for methadone-treated patients have generally been somewhat better than for patients treated with either levomethadyl acetate or buprenorphine, suggesting that methadone will continue to be the first-line treatment for most patients except in select situations.

Effectiveness of Methadone Maintenance Programs The results of two nationwide evaluation studies of drug abuse treatment supported by the National Institute on Drug Abuse (NIDA) provide additional evidence for the effectiveness of methadone maintenance treatment. In the Drug Abuse Reporting Project, 1-year retention of methadone maintenance patients entering treatment between 1969 and 1974 exceeded 50 percent, and marked decreases in illicit drug use and criminality as well as increased employment were observed after 3 years for all those entering methadone maintenance. The subsequent Treatment Outcome Prospective Study conducted on patients entering drug treatment programs from 1979 to 1981 found similar benefits of methadone maintenance.

Following the closing of a county methadone program, more than half of the 94 involuntarily discharged clients reverted to heroin abuse within 2 years. Compared with a matched sample from another county clinic, involuntarily discharged patients had nearly double the arrest or incarceration rate. Similar adverse outcomes were noted in methadone-maintained patients who were unable or unwilling to transfer to private, fee-based programs following closure of a publicly supported program in San Diego.

Health status also improves markedly as a result of methadone maintenance. Untreated heroin dependence is associated with substantially increased mortality and morbidity. During methadone maintenance treatment, however, the annual mortality risk averages 25 percent of the risk for untreated opioid-dependent persons, largely as a result of reductions in deaths due to heroin overdose or suicide. Substantial reductions in injection drug use associated with methadone maintenance treatment also lead to reductions in HIV transmission and improvements in immune system function and other health measures. In one study, the prevalence of HIV infection among untreated heroin addicts during an 18-month period was 22 percent, compared to a prevalence of 3.5 percent among methadone maintained patients in the same city.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The relative safety of long-term methadone maintenance has been established through prospective and retrospective studies. No evidence of methadone toxicity for organ systems has been found during the course of three decades of widespread clinical use. Less information is available regarding the long-term safety of levomethadyl acetate or buprenorphine. Levomethadyl acetate has been found to prolong the QT interval in some patients, so careful monitoring of patients with cardiac conduction deficits is recommended. Although patients entering methadone maintenance programs suffer from high rates of a number of chronic diseases, including (most commonly) HIV infection, hepatitis, tuberculosis, other infectious diseases, and cardiac or renal disorders, methadone maintenance is associated with improved health condition and decreased risk of HIV transmission, largely as a result of reducing injection drug use and facilitating appropriate use of medical services.

THERAPEUTIC INDICATIONS

Methadone is used for the short-term detoxification (30 days), long-term detoxification (180 days), and maintenance of opiate and opioid addicts. Methadone is a schedule II drug; its administration is tightly governed by specific federal laws and regulations.

The highly effective analgesic effects of methadone are used in the management of chronic pain when less potentially addictive agents are inadequate.

Levomethadyl acetate, a control level (schedule) II drug, is used only for maintenance treatment of opioid dependent patients. It is not used for detoxification treatment or for analgesia. Levomethadyl acetate can only be dispensed by approved narcotic treatment programs, and its administration is tightly governed by specific federal and state laws and regulations. Buprenorphine is currently approved only for treatment of moderate to severe pain and is a control level V analgesic. Its use outside of experimental programs for maintenance or detoxification treatment of opioid dependence will require federal approval.

PRECAUTIONS AND ADVERSE EFFECTS

Overdose The acute effects of opioid agonists include miosis, decreased gastrointestinal motility, sedation, hypothermia, bradycardia, hypotension, and respiratory depression. The risk of overdose is greatest during induction onto maintenance treatment, prior to the development of tolerance. During induction, deaths related to methadone overdose have been reported for patients given initial methadone doses of 50 to 60 mg daily. Deaths typically occur 2 to 6 days after initiation of methadone treatment, and evidence of chronic persistent hepatitis is often found on postmortem examination, suggesting that decreased methadone elimination may contribute to overdose. The slow onset of effects of levomethadyl acetate may also increase the risk of opioid overdose during methadyl acetate induction if patients also continue to use illicit drugs, and overdose can also result from too frequent (e.g., daily) dosing. Although buprenorphine may also produce respiratory depression, the risk of overdose appears to be lower with buprenorphine than with either methadone or methadyl acetate. Deaths have been reported in association with overdose on buprenorphine in combination with benzodiazepines.

Adverse Effects Tolerance to most of the pharmacological effects of opioid agonists develops during long-term maintenance, and relatively few adverse effects are reported by patients. The most common adverse effects of methadone include increased sweating, constipation, decreased libido or interference with orgasm, weight gain, and insomnia or sleep irregularities. Menstrual irregularities are reported frequently by women, especially during the initial 6 months on methadone. Occasional adverse effects include peripheral edema, urinary retention, rash, and arthralgia. Adverse effects of methadyl acetate are generally similar to those of methadone and most commonly include anxiety, insomnia, sweating, and constipation.

The effects of long-term methadone maintenance on neuroendocrine and immune system functioning have been the subject of intensive evaluation. Although tolerance develops to many of the endocrine effects of methadone typically found during the first 12 months of treatment (decreased follicle-stimulating hormone [FSH] and luteinizing hormone [LH] levels, abnormalities of cortisol response to stress), several neuroendocrine abnormalities persist. Patients on long-term maintenance may have elevated prolactin concentrations and diminished prolactin response to hypoglycemia and hypothalamically mediated impaired ability to concentrate urine during dehydration. While the results of some in vitro studies suggest that methadone may impair natural killer cell cytotoxicity at high dosages, careful clinical evaluations have not detected methadone-induced impairments of immunological functioning in methadone-maintained patients. Natural killer-cell cytotoxicity activity has been found to be significantly lower in active heroin addicts without HIV infection or other substance abuse than in long-term methadone patients or healthy, non–drug-abusing controls. No differences were noted between methadone patients and healthy controls.

PREGNANCY

Methadone maintenance combined with the provision of adequate psychosocial support and medical services can lead to marked improvements in obstetrical outcome for pregnant heroin addicts. Heroin addiction is associated with increased rates of spontaneous abortion, abruptio placenta, toxemia, eclampsia, infection, premature labor, and septic thrombophlebitis. Improved maternal nutrition and health and increased use of prenatal care among pregnant women treated in

comprehensive methadone programs are associated with improved obstetrical outcomes.

The methadone dosage during pregnancy is a critical issue. Lower methadone dosages can prevent opioid withdrawal in the pregnant woman and reduce the likelihood of opioid withdrawal in the neonate. Higher methadone dosages are associated with higher rates of neonatal withdrawal but may be needed to prevent continued heroin abuse. Especially during the third trimester, methadone metabolism increases so that lower doses may not hold the patient, and it may be necessary to use higher doses or split the daily dose and administer it twice daily. Abnormalities in fetal activity have been noted when methadone is administered in a single daily dose (increased fetal activity prior to dose administration and depressed activity subsequent to methadone administration), and split dosing administered twice daily leads to stabilization of fetal activity. The clinical significance of these fetal movement abnormalities has not been documented.

Because the elimination half-life of methadone is prolonged in the newborn, neonatal withdrawal symptoms may be delayed in onset and last for many weeks. Signs of opioid withdrawal in the neonate include tremor, high-pitched cry, increased muscle tone and activity, poor sleep and eating, mottling, yawning, sweating, and skin excoriation. Withdrawal-related convulsions can be life threatening. Phenobarbital and paregoric can be used to control hyperactivity, prevent convulsions, and relieve gastrointestinal problems. Some studies suggest that there is an increased rate of sudden infant death syndrome among children exposed prenatally to methadone. The decreased effects of buprenorphine on respiratory depression compared with methadone suggest an additional rationale for evaluating buprenorphine maintenance during pregnancy.

DRUG INTERACTIONS

The licit or illicit use of many drugs, including barbiturates, phenytoin (Dilantin), carbamazepine (Tegretol), and rifampin (Rifadin), as well as heavy alcohol use may induce hepatic microsomal enzymes and markedly increase the rate of methadone metabolism. In patients maintained on methadone, enhanced metabolism can lead to opioid withdrawal symptoms and craving, associated with methadone trough concentrations below 100 ng/mL, at previously adequate methadone dosages. The methadone dosage may need to be increased in these patients. For patients maintained on methadyl acetate, however, liver enzyme induction may lead to increased conversion of methadyl acetate to its even more active metabolite, nor-methadyl acetate, and consequently to higher peak activity levels or possibly a shortened duration of action. Acute opioid withdrawal may also be precipitated in methadone- or levomethadyl acetate-maintained patients by ingestion of narcotic antagonists (e.g., naltrexone), partial μ-agonists (e.g., buprenorphine) or analgesics with mixed agonist-antagonist properties, such as pentazocine (Talwin). Methadone- or levomethadyl acetate-maintained patients with illicit use of nonopioid drugs may also experience opioid withdrawal and request a methadone dosage increase, but in the latter situation, appropriate management consists of interventions to decrease illicit drug use rather than necessarily increasing methadone dosage.

Methadone, buprenorphine, and levomethadyl acetate may have important drug interactions with medications that are also metabolized by hepatic cytochrome CYP 3A4. Competitive inhibition of methadone metabolism following acute use of alcohol or administration of cimetidine (Tagamet), erythromycin, ketoconazole (Nizoral), and fluvoxamine (Luvox) may lead to higher plasma concentrations or prolonged duration of effects of methadone. Inhibition of metabolism of methadyl acetate, however, may slow the onset, lower the activity, or increase the duration of action of levomethadyl acetate. Methadone excretion may also be reduced by medications that alkalinize the urine.

Methadone maintenance may also interfere with the metabolism of medications metabolized by cytochrome P450 and lead to elevated plasma concentrations of desipramine (Norpramine, Pertofrane) and fluvoxamine at usual therapeutic dosages. Reductions in dosage of antidepressant medications is warranted if patients are symptomatic, show signs of toxicity, or have markedly elevated plasma concentrations. Recent reports suggest that methadone may also lead to higher zidovudine (Retrovir) concentrations and the possibility of increased zidovidine toxicity at standard dosages. Several new protease inhibitors, including ritonavir (Norvir), indinavir (Crixivan), and saquinavir (Fortovase) are also metabolized by cytochrome P450 and when studied in vitro in human liver microsomes competitively inhibit methadone and buprenorphine demethylation. The clinical implications of interactions between methadone and other antiretroviral agents, including protease inhibitors, have not been extensively evaluated, and there is also a lack of information regarding their potential interactions with either levomethadyl acetate or buprenorphine.

LABORATORY INTERFERENCES

No known laboratory interferences are associated with methadone, levomethadyl acetate, or buprenorphine treatment.

DOSAGE AND ADMINISTRATION

Methadone is supplied in dispersible tablets of 5, 10, and 40 mg; in oral solutions of 5 mg/5 ml, 10 mg/5 ml, and 10 mg/ml; and in a parenteral form of 10 mg/ml.

Levomethadyl acetate is supplied only as an oral solution (10 mg/mL). Buprenorphine is currently supplied only in 1 mL ampules containing 0.3 mg buprenorphine. Sublingual tablet formulations of buprenorphine containing buprenorphine only or buprenorphine combined with naloxone in a 4:1 ratio are in use outside of the United States and are being investigated in the United States for maintenance treatment of opioid dependence.

Variability of Outcome in Clinical Practice Despite clear evidence of the overall effectiveness of methadone maintenance treatment, wide variability in treatment outcome has been observed in different programs. An evaluation of six methadone programs in three cities in the Northeast showed the magnitude of these differences. Current intravenous drug use ranged from 9.8 percent of the patients in one program to 57.1 percent in another program. Methadone dose, program features (philosophy, amount and quality of counseling services, effectiveness of program director, and stability of staff), and patients' characteristics (concomitant psychiatric disorders, polydrug abuse) account for many of these differences.

Methadone Dosage Dole and Nyswander used methadone doses of 80 to 120 mg daily to prevent craving and produce tolerance to usual street doses of opiates so that the effects of illicit opioid use would be blocked or attenuated. Despite compelling evidence that the efficacy of methadone is dosage dependent, with higher daily dosages leading to higher retention and greater reductions in illicit opioid use, many programs prescribe inadequate dosages. In a 1988 survey of methadone programs, the average methadone dosage in two-thirds of the programs was 50 mg a day or less, which is considerably lower than the recommended average dosage exceeding 60 mg a day, and almost 25 percent of methadone maintenance patients were maintained on less than 40 mg of methadone daily. In the study of six methadone clinics, methadone dosage was one of the most important factors associated with illicit opioid use. More than one-third of patients maintained on a methadone dose below 35 mg used heroin at least weekly, while there was no illicit use of opioids among patients maintained at a dosage of 80 mg a day or more. Methadone dosage was not associated with other outcomes, however, including cocaine use or criminal behavior.

In clinical practice, no single methadone dosage will be optimal for all patients, and maintenance dosage should be based on patient response and also possibly guided by assessment of trough methadone concentrations in plasma. Some experts have suggested that an optimal daily methadone dosage will result in trough plasma concentrations above 150 to 200 ng/mL. Methadone plasma concentrations may not correlate directly with the concentration of unbound methadone, thus limiting their clinical utility. On the basis of these considerations, it is clinically appropriate to plan to increase methadone dosage over several weeks to 70 mg a day for all patients. It is important to start at a lower daily dosage (25 to 30 mg) to avoid the possibility of producing a lethal overdose in a nontolerant patient. Patients who experience sedation or other methadone adverse effects as dosage is increased may be maintained at a lower dosage if they remain abstinent from all illicit opiates. Patients who continue opiate use at any given dosage may benefit from an increased methadone dose, and methadone dosage should be increased gradually until patients eliminate or reduce illicit opiate use. The methadone dosage is not likely to affect use of nonopiate illicit drugs.

Program Features Program features, including a program philosophy emphasizing social rehabilitation and long-term maintenance, provision of more and higher quality counseling services, high staff morale and low staff turnover, and presence of an effective program director, are associated with more-favorable treatment outcomes. The magnitude of the effects of program features is illustrated in the results of a large multisite evaluation of methadone and levomethadyl acetate. Despite identical study protocols and medication dosage, rates of early termination among the clinics differed considerably. In two clinics, 30 percent of patients or less terminated treatment prematurely, while 75 percent or more of the patients in seven other clinics left the study before 40 weeks.

A number of controlled clinical trials have demonstrated the importance of psychosocial and behavioral interventions in improving the effectiveness of methadone maintenance. In one study, patients maintained on a fixed methadone dosage were randomly assigned to one of three levels of counseling: minimal (once a month), basic (weekly), or enhanced (weekly plus on-site medical and vocational services). Minimal services were effective for a minority of patients. Outcomes were significantly improved for patients receiving basic or enhanced counseling. Although enhanced services led to the best outcomes, basic services were overall more cost-effective. More recently, another study found that enhanced counseling services were as effective, and more cost-effective, than an intensive day treatment program for newly admitted methadone maintenance patients.

The efficacy of specific counseling approaches and of contingency management has also been evaluated in methadone programs. In one recent study, behavioral counseling based on the community reinforcement approach led to higher rates of abstinence than standard drug counseling. Several studies have shown the benefits of giving patients contingent rewards (e.g., voucher-based incentives with a monetary value or take-home bottle privileges) for maintaining abstinence or complying with treatment plan recommendations. Aversive contingencies for continued illicit drug use, such as methadone dosage reduction or detoxification, are more controversial since they may lead to increased illicit opioid use. Rebound of drug use following discontinuation of contingency management may also be a problem.

Treatment outcome is also associated with program features, including consistent and fair implementation of program policies and rules and the quality of the counseling services. Studies suggest that the most effective counselors are those who (1) establish a supportive relationship with patients, (2) develop and formulate a coherent rehabilitation plan based on discussions with the patient and other staff, (3) follow program rules closely, (4) meet more than once a month with the patient, and (5) provide sensible advice and referral to special resources when needed. The best counselors are those who can anticipate the patient's problems and discuss in advance strategies to prevent or cope with these problems. The success of methadone treatment quite clearly depends on the ability to recruit, train, supervise, and retain effective counselors.

Length of Treatment Virtually all studies of methadone maintenance indicate that longer time in treatment is associated with improved outcome and that involuntary termination of methadone treatment is associated with subsequent high rates of relapse to illicit opiate use. Based on these findings, a consensus recommendation was established by drug abuse researchers and clinicians convened by NIDA that no arbitrary limit should be placed on the duration of methadone treatment and that the decision to terminate treatment should be made jointly by the patient and staff. The findings of the six-site evaluation of methadone treatment programs support these recommendations. Programs emphasizing long-term maintenance and social rehabilitation were found more effective than programs emphasizing short-term treatment. Despite the consensus recommendations, in actual practice, long-term methadone treatment is either discouraged or not available in many programs. In some areas, long-term methadone maintenance is also prohibited by policy makers or not covered by medical insurance. Half of the methadone programs surveyed in 1988 encouraged patients to detoxify in less than 6 months.

ALCOHOL AND OTHER DRUG ABUSE DURING METHADONE MAINTENANCE

Although the pharmacological actions of maintenance on methadone reduce opioid craving, methadone has little direct beneficial effect on other substance use, and abuse of or dependence on alcohol, benzodiazepines, cocaine, marijuana, and other substances often complicates treatment outcome. Although some patients may increase their use of alcohol or other nonopioid substances after entering methadone maintenance programs, methadone maintenance is associated with an overall decrease in rates of nonopioid substance use, especially for patients who cease heroin use.

Alcohol, Benzodiazepine, and Sedative Abuse Alcohol, benzodiazepine, and sedative abuse continues to be a major problem among patients maintained on methadone. Rates of heavy alcohol consumption range from 17 to 30 percent or more of methadone patients, and alcoholism is a leading cause of death among methadone maintenance patients. Some studies suggest that most patients who experience difficulties with alcohol while maintained on methadone had such problems before starting on methadone. High doses of benzodiazepines are often abused by methadone patients to boost their methadone dose or experience a high. Therapeutic approaches to the abuse of alcohol or benzodiazepines in methadone programs include medical withdrawal; behavioral monitoring (daily breathalizer to detect alcohol use, increased frequency of urine monitoring to detect benzodiazepine or sedative use); intensified individual, group, or family counseling; and use of contingency contracting (providing positive incentives such as take-home bottles for abstinence, and aversive consequences such as loss of bottle privileges for continued use). Referral to Alcoholics Anonymous (AA) and supervised disulfiram (Antabuse) administration have also been beneficial in ameliorating alcohol problems. Because many AA programs are inhospitable to methadone patients and discourage them from remaining on the medication, it is important to refer to AA programs that support rather than undermine methadone treatment.

Cocaine Abuse In recent years, cocaine abuse has become a major problem among methadone-maintained patients. Rates of cocaine abuse averaged 58 percent among patients entering the six-site study, and rates among those in treatment range as high as 40 percent or more in many programs. Ethnographic studies suggest that cocaine use has become part of the drug use and social life of patients in some programs, with use not only by the most deviant methadone patients but also by a substantial proportion of patients who are otherwise compliant with program rules.

Cocaine abuse is of particular concern because it negates many of the benefits of methadone maintenance with regard to reducing criminal activity or the spread of HIV infection. Behavioral interventions, including implementing positive or aversive contingencies for abstinence or continued use, have been shown to have short-term beneficial effects. One recent study found that 9 of the 19 patients (47 percent) randomly assigned to voucher-based reinforcement for cocaine-negative urine tests achieved 7 to 12 weeks of sustained abstinence from cocaine during a 12-week study, while only 1 of 18 patients (6 percent) assigned to a noncontingent, yoked voucher-control condition achieved more than 2 weeks of sustained abstinence. Voucher rewards for abstinence had broad beneficial effects in addition to decreasing cocaine use. In another study, use of vouchers with a monetary value to reward participation in planned activities outside of the clinic was more effective than either voucher rewards for abstinence or a no-voucher control in reducing rates of cocaine use. The value and schedule of rewards also affects the efficacy of contingency management. Relapse following cessation of contingency management is often a problem, however, and additional strategies are needed to encourage sustained or long-term abstinence. Although some studies initially suggested that buprenorphine maintenance would be associated with lower rates of cocaine use than methadone maintenance, more-recent double-blind random-assignment studies have not found such a difference. Controlled studies of pharmacological adjuncts, such as desipramine, carbamazepine, mazindol (Sanorex), amantadine (Symmetrel, Symadine), or bromocriptine (Parlodel), have not consistently documented beneficial effects of these agents with patients in methadone programs. Because alcohol is often used along with cocaine and studies suggest that disulfiram pretreatment leads to increased anxiety and dysphoria following cocaine administration, disulfiram is currently being evaluated for the treatment of cocaine dependence in agonist-maintained patients. Two recent double-blind clinical trials found greater reductions in cocaine use among methadone-maintained patients and buprenorphine-maintained patients randomly assigned to disulfiram than in those assigned to placebo.

CONCOMITANT PSYCHIATRIC DISORDERS

High rates of psychiatric disorders, most notably depressive, anxiety, and personality disorders, have been consistently documented in opioid addicts. In one recent study, 47 percent of patients entering a methadone treatment program had at least one lifetime psychiatric diagnosis other than a substance use disorder. While psychiatric comorbidity is associated with poorer treatment outcome overall, professional psychotherapy as an adjunct to traditional counseling and cautious use of psychotropic medications can improve outcome for psychiatrically impaired patients. A recent double-blind study found that depressed methadone-maintained patients randomly assigned to imipramine (Tofranil) had greater improvements in depression and reduced illicit drug use more than patients allocated to placebo.

TERMINATION OF TREATMENT

Whether and when to terminate methadone maintenance treatment remains controversial. The benefits of methadone maintenance are most apparent while patients remain in treatment, and high rates of relapse to illicit drug use with accompanying increases in criminal activity and deteriorating medical, vocational, and social functioning occur after withdrawal from methadone maintenance. After termination of methadone maintenance, 70 to 80 percent of patients return to illicit drug use. The highest rates of relapse occur in patients who have been administratively discharged because of program noncompliance. Even the most successful, socially rehabilitated methadone maintenance patients who have maintained abstinence and voluntarily terminate treatment exhibit relapse rates during the first 3 years ranging from 20 to 50 percent. The fragility of many patients and their need for continued social services and program supports is illustrated by the results of an experimental program of medical maintenance. In this program, patients who had been successfully treated in a methadone maintenance clinic for at least 5 years, as evidenced by stable employment and the absence of criminal activity or use of illegal drugs or excessive alcohol, were offered the opportunity to receive methadone prescriptions from a primary care physician. At follow-up visits from 42 to 111 months, 72 of the first 100 patients remained in medical maintenance; 15 had unfavorable discharges for cocaine use, repeatedly losing medications, or repeatedly missing appointments; 8 had voluntarily withdrawn from methadone or left the program; and 5 had been discharged for medical reasons.

Because relapse is so common following methadone withdrawal, it is unrealistic to set withdrawal from methadone as a goal of treatment for all patients. Patients who decide voluntarily to withdraw from methadone should be offered a choice of withdrawal protocols as well as the opportunity to increase their methadone dosage or return to methadone maintenance if relapse to illicit opioid use seems imminent. Withdrawal protocols may include gradual, blinded tapering of methadone dosage, clonidine, combined clonidine and naltrexone, or buprenorphine substitution and discontinuation. Because protracted abstinence symptoms may be present for up to 1 year following cessation of methadone, it is advisable for patients to remain in treatment following completion of withdrawal and for drug-free patients to have facilitated reentry to a methadone program if relapse seems imminent or has occurred.

REGULATORY AND POLICY ISSUES

The use of opioid agonists for detoxification or maintenance treatment of opioid dependence is tightly regulated by federal, state, and local regulations. Special licenses issued by the Drug Enforcement Agency (DEA) are required for physicians and clinics to prescribe or dispense methadone for detoxification or methadone or levomethadyl acetate for maintenance treatment. Federal rules issued jointly by the FDA and NIDA define standards for admission to maintenance treatment and minimum requirements regarding drug abuse testing during treatment, govern the provision of take-home doses of methadone, and prohibit take-home doses of levomethadyl acetate. State and local regulations and policies often place additional restrictions on the provision of maintenance treatment. While the federal regulations and policies were intended primarily to ensure program quality and prevent improper use or abuse of opioid agonists prescribed for detoxification or maintenance treatment, there has been growing concern that the regulations and policies have prevented needed expansion of maintenance treatment and have inadvertently contributed to stigmatization of patients enrolled in methadone treatment. These issues were addressed in a 1997 National Institutes of Health (NIH) Consensus Statement calling for increased funding, expanded access to and availability of agonist maintenance treatment, and revision of current laws and regulations.

Two policy issues continue to arouse considerable controversy. One concerns take-home doses of methadone or other opioid agonists. Take-home doses are a source of controversy because of the potential for their diversion to street use or accidental ingestion by children or nontolerant individuals. In methadone maintenance programs, take-home bottles are often necessitated by programmatic and fiscal constraints that limit the ability of some clinics to dispense 7 days a week to all patients. When take-home bottles are routinely dispensed to patients, diversion is quite likely. Take-home privileges, however, can be an essential element of the overall rehabilitation program. They provide one of the strongest positive incentives for compliance with the therapeutic goals of treatment and constitute an essential component of contingency contracting in many methadone programs. Additionally, take-home doses allow patients who are engaged in successful rehabilitation efforts to wean themselves from daily contact with the program and from other patients in drug treatment. Monitoring procedures in methadone programs, including routine and random bottle recalls and limitation of take-home bottles to patients who have progressed in treatment and have refrained from illicit drug use, will limit (although not entirely prevent) diversion. The tablet combining buprenorphine and naloxone may also be of sufficiently low abuse or overdose potential to permit its being prescribed routinely for patient self-administration rather than requiring that it be dispensed and ingested under direct observation.

A second area of controversy concerns the need for ancillary services in agonist maintenance programs and the possibility of providing maintenance treatment in primary care clinics or physicians' offices rather than maintenance clinics. Because the costs of full-service methadone programs are in part responsible for the limitation on the number of available methadone treatment slots, no-frills or low-service methadone has been evaluated as a less expensive means of making methadone treatment more widely available. Results of studies evaluating programs that dispense methadone but offer limited or no additional services to new entrants to treatment have been disappointing. Although rates of continued heroin use decrease markedly for patients in these programs, rates of cocaine and other illicit drug use and risk of HIV infection remain high. There is evidence, however, that agonist-maintained patients in more advanced stages of recovery do not necessarily require additional counseling or ancillary services and may be treated successfully in a physician's office. Strategies for providing counseling services along with opioid agonist maintenance medications in primary care settings or physicians offices may also permit expansion of maintenance treatment into these settings while also retaining its effectiveness. Integration of addiction treatment into medical practices may help to expand the availability of services, improve access, and reduce the stigma of addiction treatment.

SUGGESTED CROSS-REFERENCES

Substance-related disorders are discussed in Section 13.10, opioid-related disorders). Noncompliance with treatment is discussed in Section 27.1, adult antisocial behavior and criminality in Section 27.3, and public psychiatry in Section 52.1. Section 31.1 discusses the general principles of psychopharmacology.

SECTION REFERENCES

Abbott PJ, Weller SB, Delaney HD, Moore BA: Community reinforcement approach in the treatment of opiate addicts. Am J Drug Alcohol Abuse *24:*17, 1998.

Anglin MD, Almog IJ, Fisher DG, Peters KR: Alcohol use by heroin addicts: Evidence for an inverse relationship. A study of methadone maintenance and drug-free treatment samples. Am J Drug Alcohol Abuse *15:*191, 1989.

Anglin MD, Speckart GR, Booth MW, Ryan TM: Consequences and costs of shutting off methadone. Addict Behav *14:*307, 1989.

*Ball JC, Ross A: *The Effectiveness of Methadone Maintenance Treatment*. Springer-Verlag, New York, 1991.

Bell J, Bowron P, Lewis J, Batey R: Serum levels of methadone in maintenance clients who persist in illicit drug use. Br J Addict *85:*1599, 1990.

Bickel WK, Amass L, Higgins ST, Badger GJ, Esch RA: Effects of adding behavioral treatment to opioid detoxification with buprenorphine. J Consult Clin Psychol *65:*803, 1997.

Bickel WK, Rizzuto P: The naturalistic oscillating patterns of alcohol consumption in alcoholic methadone patients. J Stud Alcohol *52:*454, 1991.

*Brooner RK, King VL, Kidorf M, Schmidt DW: Psychiatric and substance use comorbidity among treatment-seeking opioid abusers. Arch Gen Psychiatry *54:*71, 1997.

Caplehorn Jr, Dalton MS, Haldar F, Petrenas AM, Nisbet JG: Methadone maintenance and addicts' risk of fatal heroin overdose. Substance Use Misuse *31:*177, 1996.

Chatham LR, Rowan-Szal GA, Joe GW, Simpson DD: Heavy drinking, alcohol-dependent vs. nondependent methadone-maintenance: A follow-up study. Addict Behav *22:*69, 1997.

D'Aunno T, Vaughn TE: Variations in methadone treatment practices. JAMA *267:*253, 1992.

Dole VP: Implications of methadone maintenance for theories of narcotic addiction. JAMA *260:*3025, 1988.

Dole VP, Nyswander M: A medical treatment for diacetylmorphine (heroin) addiction. JAMA *193:*80, 1965.

Dole VP, Nyswander ME, Kreek MJ: Narcotic blockade. Arch Intern Med *118:*304, 1966.

Dole VP, Nyswander ME, Warner A: Successful treatment of 750 addicts. JAMA *206:*2708, 1968.

Effective Medical Treatment of Opiate Addiction. NIH Consensus Development Statement *15:*1, 1997.

Eisenberg T, Bigelow GE, Strain EC, Walsh SL, Brooner RK, Stitzer ML, Johnson RE: Dose-related efficacy of levomethadyl acetate for treatment of opioid dependence: A randomized clinical trial. JAMA *277:*1945, 1997.

Finnegan LP: Effects of maternal opiate abuse on the newborn. Fed Proc *44:*2314, 1985.

Glanz M: Methadone vs. L-a-acetylmethadol (LAAM) in the treatment of opiate addiction. Am J Addict *6:*339, 1997.

Grella CE, Anglin MD, Wugalter SE: Patterns and predictors of cocaine and crack use by clients in standard and enhanced methadone maintenance treatment. Am J Drug Alcohol Abuse *23:*14, 1997.

Gronbladh L, Gunne L: Methadone-assisted rehabilitation of Swedish heroin addicts. Drug Alcohol Depend *24:*31, 1989.

Gronbladh L, Ohlund LS, Gunne LM: Mortality in heroin addiction: Impact of methadone treatment. Acta Psychiatr Scand *82:*223, 1990.

Iguchi MY, Belding MA, Morral AR, Lamb RJ, Husband SD: Reinforcing operants other than abstinence in drug abuse treatment: An effective alternative for reducing drug use. J Consult Clin Psychology *65:*421, 1997.

Inturrisi CE, Colburn WA, Kaiko RF, Houde RW, Foley KM: Pharmacokinetics and pharmacodynamics of methadone in patients with chronic pain. Clin Pharmacol Ther *41:*392, 1987.

Iribarne C, Berthou F, Carlhant D, Dreano Y, Picart D, Lohezic F, Riche C: Inhibition of methadone and buprenorphine N-dealkylations by three HIV-1 protease inhibitors. Drug Metab Dispos *26:*257, 1998.

Jaffe JH, Martin WR: Opioid analgesics and antagonists. In *The Pharmacological Basis of Therapeutics*, ed 7, AG Gilman, LS Goodman, TW Rall, F Murad, editors. MacMillan, New York 1985.

Joe GW, Simpson DD, Hubbard RL: Unmet service needs in methadone maintenance. Int J Addict *26:*1, 1991.

*Kaltenbach K, Berghella V, Finnegan L. Opioid dependence during pregnancy: Effects and management. Obstet Gynecol Clin North Am *25:*139, 1998.

Kraft MK, Rothbard AB, Hadley TR, McLellan AT, Asch DA: Are supplementary services provided during methadone maintenance really cost-effective? Am J Psychiatry *154:*1214, 1997.

Kreek MJ: Using methadone effectively: Achieving goals by application of laboratory, clinical, and evaluation research and by development of innovative programs. In *Improving Drug Abuse Treatment*, RW Pickens, CG Leukefeld, CR Schuster, editors. National Institute on Drug Abuse Research Monograph 106, Rockville, MD, 1991.

Ling W, Klett CJ, Gillis RD: A cooperative clinical study of methadyl acetate. Arch Gen Psychiatry *35:*345, 1978.

Ling W, Charuvastra C, Kaim SC, Klett CJ: Methadyl acetate and methadone as maintenance treatments for heroin addicts. Arch Gen Psychiatry *33:*709, 1976.

Marsch LA: The efficacy of methadone maintenance interventions in reducing illicit opiate use, HIV risk behavior and criminality: A meta-analysis. Addiction *93:*515, 1998.

*McLellan AT, Arndt IO, Metzger DS, Woody GE, O'Brien CP: The effects of psychosocial services in substance abuse treatment. JAMA *269:*1943–1959, 1993.

Moody DE, Alburges ME, Parker RJ, Collins JM, Strong JM: The involvement of cytochrome P450 3A4 in the N-demethylation of L-alpha-acetylmethadol (LAAM), norLAAM, and methadone. Drug Metab Dispos *25:*1347, 1997.

National Institute on Alcohol Abuse and Alcoholism: Methadone maintenance and patients in alcoholism treatment. Alcohol Alert *1:*1, August 1988.

Novick DM, Joseph H, Salsitz EA, Kalin MF, Keefe JB, Miller EL, Richman BL: Outcomes of treatment of socially rehabilitated methadone maintenance patients in physicians' offices (medical maintenance): Follow-up at three and a half to nine and a fourth years. J Gen Intern Med *9:*127, 1994.

Novick DM, Ochshorn M, Ghali V, Croxson TS, Mercer WD, Chiorazzi N, Kreek MJ: Natural killer cell activity and lymphocyte subsets in parenteral heroin abusers and long-term methadone maintenance patients. J Pharmacol Exp Ther *250:*606, 1989.

Novick DM, Pascarelli EF, Joseph H, Salsitz EA, Richman BL, DesJarlais DC, Anderson M, Dole VP, Nyswander ME: Methadone maintenance patients in general medical practice. JAMA *259:*3299, 1988.

Nunes EV, Quitkin FM, Donovan SJ, Deliyannides D, Ocepek-Welikson K, Koenig T, Brady R, McGrath PJ, Woody G: Imipramine treatment of opiate-dependent patients with depressive disorders: A placebo-controlled trial. Arch Gen Psychiatry *55:*153, 1998.

Rettig RA, Yarmolinsky A, editors: *Federal Regulations of Methadone Treatment*. Institute of Medicine, National Academy Press, Washington, DC, 1995.

Schottenfeld RS, Pakes JP, Oliveto A, Ziedonis D, Kosten TR: Buprenorphine vs. methadone maintenance treatment for concurrent opioid dependence and cocaine use. Arch Gen Psychiatry *54:*713, 1997.

Senay EC: Methadone maintenance. In *American Psychiatric Association Task Force on Treatments of Psychiatric Disorders*, vol 2, chap 136. American Psychiatric Association, Washington, DC, 1989.

Silverman K, Higgins ST, Brooner RK, Montoya ID, Cone EJ, Schuster CR, Preston KL: Sustained cocaine abstinence in intravenous polydrug abusers through voucher-based reinforcement therapy. Arch Gen Psychiatry *53:*409, 1996.

Stitzer ML, Bigelow GE: Contingent methadone take-home privileges: Effects on compliance with fee payment schedules. Drug Alcohol Depend *13:*395, 1984.

Stitzer ML, Kirby KC: Reducing illicit drug use among methadone patients. In *Improving Drug Abuse Treatment*, RW Pickens, CG Leukefeld, CR Schuster, editors. National Institute on Drug Abuse Research Monograph 106, Rockville, MD, 1991.

*Strain EC, Bigelow GE, Liebson LA, Stitzer ML: Moderate versus high-dose methadone in the treatment of opioid dependence: A randomized trial. JAMA *281:*1000, 1999.

Wesson DR: Revival of medical maintenance in the treatment of heroin dependence (Editorial). JAMA *259:*3314, 1988.

*White WL: *Slaying the Dragon: The History of Addiction Treatment and Recovery in America*. Chestnut Health Systems/Lighthouse Institute, Bloomington, IL, 1998.

Willenbring ML, Morley JE, Krahn DD, Carlson GA, Levine AL, Shafer RB: Psychoneuroendocrine effects of methadone maintenance. Psychoneuroendocrinology *14:*371, 1989.

Woody G: Psychotherapy in community methadone programs. Am J Psychiatry *152:*1302, 1995.

Woody GE, McLellan AT, Luborsky L, O'Brien CP: Psychotherapy and counseling for methadone-maintained opiate addicts: Results of research studies. In *Psychotherapy and Counseling in the Treatment of Drug Abuse*, LS Onken, JD Blaine, editors. National Institute on Drug Abuse Research Monograph 104, Rockville, MD, 1990.

Yancovitz SR, DesJarlais DC, Peyser NP, Drew E, Friedmann P, Trigg HL, Robinson JW: A randomized trial of an interim methadone maintenance clinic. Am J Public Health *81:*1185, 1991.

31.24 SELECTIVE SEROTONIN-NORADRENALINE REUPTAKE INHIBITORS

LINDA BEAUCLAIR, M.D., DENISA RADOI-ANDRAOUS, M.D., AND GUY CHOUINARD, M.D., M.SC.

The second generation of antidepressant drugs includes the selective serotonin reuptake inhibitors (SSRIs) and bupropion (Wellbutrin), which are safer and better tolerated than and equally effective as the tricyclic drugs and monamine oxidase inhibitors (MAOIs), the first antidepressants. Another class of antidepressant drugs has emerged, the selective serotonin-noradrenaline reuptake inhibitors (SNRIs), which have also been shown to be effective and safe antidepressants. (These drugs have also been called the serotonin-norepinephrine reuptake inhibitors.) Three compounds that inhibit serotonin and noradrenaline reuptake in a selective manner have been studied. These compounds are venlafaxine (Effexor), and milnacipran and duloxetine. A faster onset of antidepressant action in patients with major depressive disorder has been proposed with venlafaxine and, to a lesser extent, with milnacipran, but there are methodological difficulties in measuring the onset of antidepressant action. Because of their lack of affinity for muscarinic, histaminergic, and a $_1$-adrenergic receptors, SNRIs do not cause the unwanted adverse effects common to tricyclic drugs and MAOIs.

CHEMISTRY

Venlafaxine [R/S-1[2-(dimethylamino)-1-(4-methoxyphenyl) ethyl cyclohexanol hydrochloride] and milnacipran [1-phenyl-1-diethyl-aminocarbonyl-2-aminomethyl-cyclopropane (Z) hydrochloride] are tertiary amines. Duloxetine [(+)-*N*-methyl-d-(1-naphthalenyloxy)-2-tiophenepropanamine hydrochloride] is a secondary amine. Their chemical structures are shown in Figure 31.24-1. Unlike milnacipran which shares similarities with the nonselective MAOI tranylcypromine (Parnate), venlafaxine and duloxetine are structurally distinct from tricyclic drugs and MAOIs. Only duloxetine has a structure partially similar to that of fluoxetine (Prozac).

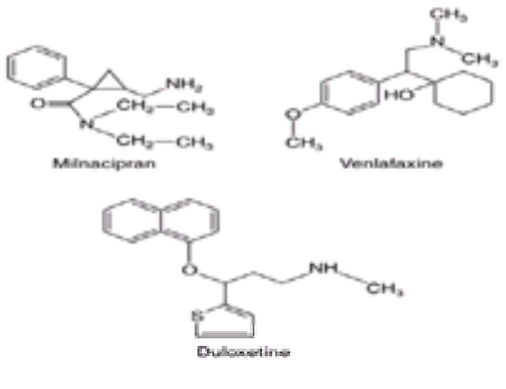

FIGURE 31.24-1 Molecular structure of selective serotonin-noradrenaline reuptake inhibitors.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics Venlafaxine is well absorbed in the gastrointestinal tract, with peak plasma concentrations (t_{max}) occurring at 2 to 2.5 hours for single doses of 25 to 150 mg. The presence of food delays the t_{max} by 20 to 30 minutes but does not affect the peak concentrations or the area under the concentration-time curve of either venlafaxine or its metabolites. Venlafaxine is metabolized by cytochrome P450 (CYP) isoenzyme 2DG (CYP 2D6) to an active metabolite (*O*-desmethylvenlafaxine) and by C4P 3A/4 to an inactive metabolite (*N*-desmethyvenlafaxine). Although venlafaxine is metabolized by these two cytochrome isoenzymes, the presently available data suggest that it does not interact with other drugs metabolized by these hepatic isoenzymes, because venlafaxine does not significantly inhibit or weakly inhibits the activity of isoenzymes CYP 2C9, CYP 2D6, CYP 1A2, or CYP 3A3/4; however, potential drug interactions associated with those isoenzymes do exist. Venlafaxine has an elimination half-life of about 3.5 hours, and that of the active metabolite is in the range of 9 to 11 hours. Venlafaxine and its metabolites are primarily excreted by the kidney. Venlafaxine is poorly dialyzed and is less than 35 percent plasma protein bound. It reaches its steady state within 3 days of multiple-dose therapy and exhibits linear pharmacokinetics.

Milnacipran is well absorbed and is not metabolized by the hepatic cytochrome 450 system. The t_{max} is 1.7 to 2.1 hours. The average plasma elimination half-life ($t_{1/2}$) is 5.8 to 8.1 hours. After a dose, 50 percent of milnacipran is excreted in the urine unchanged and 40 percent as inactive metabolites. Administration of doses of 25 to 200 mg twice daily for 14 days to 10 healthy volunteers resulted in rapid achievement of steady-state concentrations (2 to 3 days).

Duloxetine hydrochloride is acid labile. An enteric-coated formulation has been developed to avoid problems with acid degradation in the stomach. Duloxetine is metabolized by demethylation to a primary amine and a hydroxylated metabolite inducing moderate hepatic induction of both cytochrome P450 $_{1B}$ and P450$_{1A}$. However, hepatic enzyme induction did not alter plasma concentrations of duloxetine and its *N*-desmethyl metabolite. These two metabolites have the same potency in inhibiting the reuptake of serotonin as the parent compound, but the demethylated metabolite is less active in inhibiting the reuptake of noradrenaline than the parent compound. Duloxetine is well absorbed and extensively metabolized, 77 percent of the drug is recovered in urine and 15 percent in feces. The average time to reach maximum plasma concentrations (t_{max}) is about 5 hours, depending on the dose. The average plasma elimination half-life ($t_{1/2}$) is about 13 hours. It is estimated that with daily dosing, a pharmacokinetic steady state would be achieved in 3 to 4 days. Many studies have shown an apparent effect of food on the absorption of duloxetine: a significant delay in the time to maximum concentrations (t_{max}) and a significant increase in the extent of absorption, but there was a consistent t_{max} of 4 to 6 hours in the fasted state.

Pharmacodynamics Venlafaxine, milnacipran, and duloxetine inhibit the reuptake of noradrenaline, serotonin, and (to a lesser extent) dopamine. Venlafaxine and duloxetine are three- to fourfold more potent inhibitors of serotonin than noradrenaline uptake, whereas milnacipran is twice as potent at inhibiting noradrenaline than serotonin. Venlafaxine, milnacipran, and duloxetine lack significant affinity for cholinergic, adrenergic, histaminergic, and opioid receptors. Venlafaxine and milnacipran do not inhibit MAO activity.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Venlafaxine and milnacipran were studied in double-blind, placebo-controlled and active-controlled, parallel-group randomized clinical trials that were designed according to the rules on conducting antidepressant trials. Data from those trials show venlafaxine and milnacipran to be safe and effective antidepressants.

Venlafaxine A 4-week, multicenter, placebo-controlled trial was conducted to evaluate the efficacy and short-term safety of venlafaxine in hospitalized patients (*N* = 93) who met the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) criteria for major depression and melancholia. This study

demonstrated an early benefit from therapy in venlafaxine-treated patients (*N* = 46), related to rapid onset of the antidepressant efficacy of venlafaxine with relatively high daily doses (150 to 375 mg a day) and mean dosages in the range of 350 mg a day. Venlafaxine benefit was statistically significant (*P* £ .01) as reflected in the mean change from baseline for the Montgomery-Asberg Depression Rating Scale (MADRS) total score as early as day 4, and improvement on the Hamilton Depression Rating Scale for (HAM-D) after week 1.

The efficacy and safety of venlafaxine were evaluated in a randomized, double-blind, parallel-group comparison of venlafaxine, imipramine (Tofranil), and placebo in outpatients (*N* = 224) who met the DSM-III-R criteria for major depression for a minimum duration of 1 month. The maximum mean daily doses administered were 182 ± 48 mg of venlafaxine and 176 ± 56 mg of imipramine during the 6-week study. Venlafaxine was found statistically superior to placebo, was well tolerated, and showed a significant advantage over imipramine according to the HAM-D at week 6.

In these two studies, venlafaxine was shown to be effective in both older, severely ill hospitalized patients and younger, moderately to markedly ill outpatients. The results of these two studies suggest the utility of venlafaxine for patients with depression of varying severity.

More recently, a 12-week, double-blind, placebo-controlled study was conducted in 384 outpatients with major depression and major depression associated with anxiety. Fixed total doses of 75, 150, and 200 mg were administered twice daily. This study found that twice-daily doses of 75 to 200 mg a day of venlafaxine produced a dose-related improvement in the primary efficacy parameters and in the onset of significant antidepressant effects, which was noted at weeks 1 to 2 with the highest dosage tested (200 mg a day). The study also found these dosages of venlafaxine safe and effective for the treatment of major depression and depression associated with anxiety.

A multicenter, double-blind, placebo-controlled trial compared the safety and efficacy of venlafaxine and trazodone (Desyrel) in outpatients (*N* = 225) during an initial 6-week treatment study. The mean daily doses after titration were 156 to 160 mg of venlafaxine and 294 to 300 mg of trazodone. The maximum dose for each patient was achieved by day 15. Both active treatments were significantly superior to placebo at week 4 and at final evaluation. Venlafaxine produced more improvement in the cognitive disturbance and retardation factors on the HAM-D, and trazodone was more effective against the sleep disturbance factor. Venlafaxine also produced higher response rates on the Clinical Global Impression (CGI) Scale at the final evaluation, whereas trazodone did not. Some of venlafaxine's advantage over trazodone includes a better side-effect profile and better long-term tolerability.

Venlafaxine was also compared with fluoxetine, an SSRI, in a multicenter inpatient trial. Total scores for both MADRS and HAM-D were lower (*P* > .05) in the venlafaxine group than in the fluoxetine group at weeks 4 and 6. Overall tolerance was similar for both treatments. Venlafaxine was significantly more effective than fluoxetine in a sample of severely depressed patients with melancholia.

In a multicenter double-blind study, the safety and efficacy profile of venlafaxine (37.5 mg twice daily) was compared with that of fluoxetine (20 mg once daily) in outpatients (*N* = 314) with major depression for a maximum of 8 weeks of the study. The dosage of venlafaxine could be increased to 75 mg twice daily. Among patients who increased their dosage at week 2, venlafaxine was superior (*P* < .05) from week 3 onward on the HAM-D.

In a multicenter inpatient study, a full or partial response was achieved after 12 weeks of treatment with venlafaxine (150 to 375 mg a day) in approximately one-third of 70 patients who were refractory to either three different antidepressants from at least two different classes, or electroconvulsive therapy and two different antidepressants.

An open pilot study of 15 patients with primary dysthymia treated with venlafaxine for 12 weeks with a dose range of 75 mg to 225 mg daily showed significant changes of HAM-D from pretreatment to postreatment ($p < 0.001$). These results suggest that venlafaxine could be useful in the treatment of primary dysthymia but further studies are needed to confirm them.

Venlafaxine was also investigated in the treatment of bipolar II disorder in a double-blind six-week study comparing once versus twice daily venlafaxine dosing starting at 37.5 mg daily and increasing up to 225 mg daily. The efficacy and safety of venlafaxine were assessed in 17 bipolar II versus 31 patients with major depression. The efficacy was similar in both groups with no episode of induced mania in both groups. It appeared that a 6-week venlafaxine treatment may be safe and effective in bipolar II major depression. However, further controlled and long-term studies are required to confirm these results.

Two double-blind studies investigated the efficacy of venlafaxine in the treatment of generalized anxiety disorder (GAD), one comparing venlafaxine with placebo and one comparing venlafaxine with placebo and buspirone. These studies suggest that venlafaxine may be a useful treatment of generalized anxiety disorder without concomitant depression.

Some other studies, open and double-blind, suggested that venlafaxine could have potential application in the treatment of obsessive-compulsive disorder, panic disorder, agoraphobia, social phobia and attention deficit disorder.

Milnacipran Milnacipran was studied in double-blind, placebo-controlled and active controlled, parallel-group randomized clinical trials. These studies suggest that milnacipran is a safe and effective antidepressant with faster onset of action, better tolerability than tricyclic drugs, and better efficacy than SSRIs. In these studies, all patients (*N* = 2462) met the DSM-III-R criteria for major depressive disorder. At the optimal dosage (50 mg twice daily), milnacipran was more efficacious than placebo on all outcome measures. Data from these studies that compared three dosages of milnacipran (25, 50, and 100 mg twice daily) with placebo in 112 patients with moderate-to-severe depression showed that after 8 weeks, milnacipran (50 mg twice daily) was superior to placebo as demonstrated by a 50 percent reduction in total HAM-D. Milnacipran (200 mg a day) reduced the MADRS total score more than placebo but did not produce significant changes in the HAM-D score or in the proportion of CGI responders. Milnacipran (25 mg twice daily) was less effective than the other two dosages and its effects did not differ from those of placebo.

Seven randomized, double-blind trials with a comparable design compared the efficacy of milnacipran with that of amitriptyline (Elavil), imipramine (Tofranil), and clomipramine (Anafranil) in patients with major depression. Milnacipran was better tolerated than, and as effective as, the tricyclic drugs in patients with major depression. In a double-blind study comparing the efficacy and safety of milnacipran and imipramine in elderly patients with major depressive episode both drugs were equally efficacious, but milnacipran had a lower incidence of adverse effects and did not impair cognitive functioning.

Comparisons between milnacipran and SSRIs are limited in number. The results of these studies suggest that the tolerability and efficacy of milnacipran are comparable to those of the SSRIs fluoxetine and fluvoxamine (Luvox).

Duloxetine Duloxetine was studied in a 6-week open-label uncontrolled multicenter design at a dosage of 20 mg daily in 79 inpatients and outpatients. Clinical response (defined as a 50 percent reduction in HAM-D total) occurred in 78.2 percent of patients, whereas remission (defined as an HAM-D total of 6 or less) occurred in 60.3 percent of patients. Duloxetine was found to be safe and well tolerated in this patient population.

Another open-label study using three different doses of duloxetine (10–30 mg a day) was designed to evaluate the safety and efficacy of duloxetine in patients with major depression and depressive state. After 6 weeks of administration, the final global improvement rate improved moderately in 52.2 percent of patients. The HAM-D total changed from 24.8 to 9.4, and the response rate (defined as a 50 percent reduction in HAM-D total) was 70.8 percent. Remission rate (defined as a HAM-D total below 7) was 33.3 percent.

In a single-center, double-blind, placebo-controlled study in outpatients who met the DSM-III-R criteria for major depression, patients were randomized to treatment with either duloxetine (20 mg a day) or placebo. The mean change in both the HAM-D and MADRS totals was numerically greater in the duloxetine-treated group than in the placebo-treated group, but these differences were not statistically significant. Statistically significantly more duloxetine-treated patients achieved remission (HAM-D below 7) than placebo-treated patients.

A multicenter, double-blind, placebo-controlled 10-week study in outpatients who met the DSM-III-R criteria for major depression found no significant differences between duloxetine-treated and placebo-treated patients, even though duloxetine-treated patients presented significant improvement in CGI-Severity and CGI-Improvement (*P* = .001 and .002, respectively). Duloxetine-treated patients experienced a significant increase in blood pressure compared with the placebo group, but the magnitude of change was not considered clinically significant. In general, duloxetine was safe and well tolerated.

In a double-blind noncontrolled study using imipramine as comparator, conducted to find the appropriate dosage of duloxetine for patients with major depression, patients were given duloxetine (10 to 30 mg a day) or imipramine (50 to 150 mg a day) for 6 weeks. Response rate (defined as a 50 percent reduction in HAM-D total)

was 44 percent for duloxetine and 40 percent for imipramine, remission rate was 29 percent for duloxetine and 25 percent for imipramine. Adverse effects were reported in 49 percent of patients on duloxetine and 64 percent of those on imipramine (*P* = .08).

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

All the clinical trials done to assess the safety and efficacy of venlafaxine, milnacipran, and duloxetine excluded any serious concomitant systemic illness. Those studies showed, however, that these antidepressants appear safe and well tolerated. However, venlafaxine has been associated with sustained hypertension, especially at dosages above 225 mg a day. It is unknown whether unstable, elevated blood pressure or presenting cardiovascular disease are serious risk factors, since these studies were done on healthy subjects. Further studies are necessary to assess this effect on blood pressure. In a small number of patients, modest sustained elevations in diastolic blood pressure were observed during duloxetine treatment.

THERAPEUTIC INDICATIONS

Sufficient data have shown that venlafaxine and milnacipran are efficacious in relieving symptoms of major depressive disorder, and some clinical data suggest their usefulness in patients with severe depression and in elderly patients who are not responding to SSRIs. According to current data, duloxetine appears to be efficacious in the treatment of major depression. However, as the effectiveness of milnacipran and duloxetine in long-term use has not yet been demonstrated, their use for extended periods should be periodically reevaluated. In long term treatment, venlafaxine reduced the frequency of relapses in comparison with placebo treatment. In 1999 venlafaxine was also approved by the Food and Drug Administration (FDA) for generalized anxiety disorder.

PRECAUTIONS AND ADVERSE REACTIONS

Published studies and unpublished data (from pharmaceutical companies) suggest that venlafaxine, milnacipran, and duloxetine are safe and well tolerated. However, caution is necessary, since the number of patients who have taken these drugs is limited. Also, during these studies, persons with any serious diseases or concomitant medications were excluded.

The adverse reactions most commonly observed (occuring in at least 5 percent of patients) with venlafaxine but not seen at an equivalent incidence among placebo-treated patients (i.e., incidence for venlafaxine at least twice for that of placebo) are asthenia, sweating, nausea, constipation, anorexia, vomiting, somnolence, dry mouth, dizziness, nervousness, anxiety, tremor, blurred vision, abnormal ejaculation or orgasm, and impotence in men. Data from placebo-controlled studies suggest dose dependency for several adverse effects. There is a significant association between higher drug dosage and greater incidence of chills, hypertension, anorexia, nausea, agitation, dizziness, somnolence, tremor, yawning, sweating, and abnormal ejaculation.

Treatment with venlafaxine is associated with sustained hypertension, especially in patients treated with more than 300 mg a day. Sustained hypertension is defined as treatment-emergent supine diastolic blood pressure of 90 mm Hg or above and 10 mm Hg or more above baseline for three consecutive visits. During the clinical trials, most of the blood pressure increases were in the range of 10 to 15 mm Hg for supine diastolic blood pressure. About 3 percent of venlafaxine-treated patients experienced an increase in blood pressure, but less than 1 percent of the venlafaxine-treated group (*N* = 19) discontinued treatment because of hypertension. The incidence of sustained hypertension is dose related: less than 100 mg a day, 3 percent of patients 100 to 200 mg a day, 5 percent; 200 to 300 mg a day, 7 percent; and above 300 mg a day, 13 percent. Either dosage reduction or discontinuation should be considered for patients with sustained hypertension.

A venlafaxine withdrawal syndrome was described by A.K. Louie and colleagues in two patients carrying the diagnosis of major depression and one patient with dysthymia, receiving venlafaxine. The third author and Abdullah Dallal also described three patients who suffered venlafaxine withdrawal reactions involving mainly gastrointestinal and central nervous system symptoms. Because of the severity of venlafaxine withdrawal reactions, it is recommended that the drug be gradually tapered over 2 to 4 weeks and over a longer period when required. Angelink and colleagues also reported a venlafaxine withdrawal syndrome after abrupt discontinuation in two elderly patients treated with 300 mg a day for 4 to 6 weeks.

A possible explanation for the discontinuation symptoms seen with venlafaxine is that it reflects a combined functional deficiency of serotonin and norepinephrine at central norepinephrine and serotonin receptors. In the presence of venlafaxine, the uptake of released serotonin and norepinephrine is continuously blocked, resulting in exposure of the receptors to increased concentrations of serotonin and norepinephrine. This in turn could lead to desensitization or downregulation of these receptors. When venlafaxine is withdrawn, the concentrations of serotonin and norepinephrine fall, providing insufficient functional agonist stimulus for the downregulated receptors.

The events most frequently causing withdrawal of milnacipran have been dysuria, palpitations, hypotension, tachycardia, gastrointestinal events (e.g., constipation, nausea, vomiting), dry mouth, headache, tremor, sweating, dizziness, and nervousness. Dysuria was reported in patients receiving milnacipran and occurred more frequently with this drug than with tricyclic drugs. Precautions and dosage reduction are recommended in elderly patients or in those with renal impairment but not in those with hepatic dysfunction. C_{max}, area under the concentration-time curve, and $t_{1/2}$ values were increased in patients with renal diseases because of significant reductions in apparent total and renal clearance of drug. Changes in total clearance paralleled those in renal clearance, and both parameters were correlated with the decreased glomerular function.

The most common adverse events with duloxetine are insomnia, nausea, headache, diarrhea, dry mouth, somnolence, weakness or fatigue, heartburn, and a flulike syndrome with rhinitis and pharyngitis. A small number of patients have shown sustained elevations in diastolic blood pressure.

The effect of venlafaxine, milnacipran, and duloxetine in pregnancy, or during breast-feeding, is not documented. Therefore, their use is not recommended in those situations until more data are available.

In overdose, according to present data, venlafaxine and milnacipran appear safe. No death was reported, and all patients recovered without sequelae.

Venlafaxine, milnacipran, and duloxetine should not be used in combination with MAOIs or within 2 weeks of terminating treatment with MAOIs. Treatment with MAOIs should be not be started until 2 weeks after discontinuation of these antidepressants.

DRUG INTERACTIONS

Presently available data indicate no drug interaction between venlafaxine and lithium (Eskalith), diazepam (Valium), or alcohol. However, cimetidine (Tagamet) increases the pharmacological activity of venlafaxine and its metabolite. Therefore, caution is advised with the use of cimetidine and venlafaxine in patients with preexisting hypertension, in elderly patients, and in patients with hepatic or renal dysfunction. The same precautions as with cimetidine apply to patients treated with erythromycin and derivatives (e.g., clarithromycin [Biaxin]).

Venlafaxine is a relatively weak inhibitor of CYP 2D6; however, the potential exists for drug interaction between venlafaxine and drugs that inhibit CYP 2D6 metabolism. The general recommendation is to start on one-half to one-quarter of the usual dose when switching from fluoxetine to venlafaxine and after that to titrate the dose of venlafaxine upward as tolerated over a few weeks as the norfluoxetine levels gradually decrease. The reason for titrating the venlafaxine dose when switching from fluoxetine to venlafaxine is that fluoxetine inhibits both CYP 2D6 and CYP 3A3/4 and how venlafaxine and its major metabolite, *O*-desmethylvenlafaxine, are metabolized in the liver by these enzymes may reduce the clearance of both the parent drug and its active metabolite, leading to a higher concentration of venlafaxine in blood.

O-Desmethylvenlafaxine is metabolized by CYP 3A3/4, and its hepatic clearance can be reduced by coadministration of drugs that inhibit this enzyme (ketoconazole and related antifungal agents; fluoxetine, fluvoxamine, nefazodone) or increased by drugs that induce this enzyme (e.g., anticonvulsants such as carbamazepine [Tegretol]). Hence it is preferable to adjust the dosage when using venlafaxine and these other drugs in combination or for a while after discontinuation of fluoxetine, because of its long half-life. Therefore dosage reduction is necessary in older patients and patients with hepatic impairment.

Because milnacipran is not metabolized by the hepatic cytochrome P450 system, it is less likely to interact with drugs that are metabolized by this system. In volunteers, coadministration of lithium or lorazepam (Ativan) did not change the pharmacokinetics of any drug. However, carbamazepine decreased the concentration

of milnacipran in plasma and methotrimeprazine (Levoprome) increased it. The potential for drug-drug interaction is unknown for duloxetine.

LABORATORY INTERFERENCES

The studies of venlafaxine and milnacipran reported no interferences with any specific laboratory tests. A few patients taking duloxetine have developed modest elevations in liver enzymes for which duloxetine treatment has been stopped, even though the drug may not have caused the elevations. No permanent liver injury has been seen.

DOSAGE AND ADMINISTRATION

Venlafaxine is available in 37.5-mg and 75-mg tablets. The usual recommended starting dosage is 75 mg a day, divided in two or three doses, to be taken with food. If there is no clinical improvement after 4 weeks, the dose can be increased by a maximum of 75 mg a day at an interval of no less than 4 days. For moderately depressed outpatients, there is no evidence of usefulness of dosages above 225 mg a day. However, more severely depressed patients responded to higher dosages, between 350 and 375 mg a day. The maximum dose is 375 mg a day. It is recommended that the total daily dose be decreased by 50 percent for patients with hepatic or renal impairment. Elderly patients who are healthy require no dose modification. However, since elderly patients may show greater sensitivity to medication, they should be followed with caution. Because there are no data available, the dosage of venlafaxine is unknown in children.

Milnacipran is most commonly administered at a dosage of 50 or 100 to 150 mg twice daily for the treatment of major depressive disorder. Improvement usually occurs within 2 weeks of treatment initiation, but some patients respond sooner. On the basis of clinical trials, the optimal daily dosage of milnacipran seems to be 100 mg in two divided doses. At a dosage of 50 mg twice daily, patients with melancholia were particularly responsive. Limited data suggest that milnacipran could prevent relapse at a dosage of 50 to 200 mg twice daily. Caution and dose reduction are required in elderly patients or those with renal impairment but not in those with hepatic dysfunction.

The dosage of duloxetine is not yet determined. It appears that 20 mg a day would have an antidepressant effect. Further clinical trials are needed to assess the optimum dosage and whether this new drug requires dose titration or not.

SUGGESTED CROSS REFERENCES

Treatment of depressive disorders is covered in Section 14.7. Pharmacokinetics is covered in Section 31.2. Monamine neurotransmitters are described in Section 1.4.

SECTION REFERENCES

Amsterdam J: Efficacy and safety of venlafaxine in the treatment of bipolar II major depressive episode. J Clin Psychopharmacol *18:*414, 1998.

Angelink MW, Zitzelsberger A, Klieser E: Withdrawal syndrome after discontinuation of venlafaxine (Letter). Am J Psychiatry *154:*1473, 1997.

Artigas F: Selective serotonin/noradrenaline reuptake inhibitors (SNRIs). CNS Drugs *4:*79, 1995.

Ballenger J: Clinical evaluation of venlafaxine. J Clin Psychopharmacol *16*(Suppl):29S, 1996.

Berk M, du Plessis AD, Birkett M, Richard D: An open-label study of duloxetine hydrochloride, a mixed serotonin and noradrenaline reuptake inhibitor, in patient with DSM-III-R major depressive disorder. Lilly Duloxetine Depression Study Group. Int Clin Psychopharmacol *12:*137, 1997.

Bloch M: Severe psychiatric symptoms associated with paroxetine withdrawal (Letter). Lancet *346:*57, 1995.

*Burnett FE, Dinan TG: The clinical efficacy of venlafaxine in the treatment of depression. Rev Contemp Pharmacother *9:*303, 1998.

*Caccia S: Metabolism of the newer antidepressants. An overview of the pharmacological and pharmacokinetic implications. Clin Pharmacokinet *34:*281, 1998.

Clerc GE, Ruimy P, Verdeau-Paillè J: A double-blind comparison of venlafaxine and fluoxetine in patients hospitalized for major depression and melancholia. Int Clin Psychopharmacol *9:*139, 1994.

Cunningham LA, Borison RL, Carman JS, Chouinard G, Crowder JE, Diamond BI, Fischer DE, Hearst E: A comparison of venlafaxine, trazodone, and placebo in major depression. J Clin Psychopharmacol *14:*99, 1994.

Dallal A, Chouinard G: Withdrawal or rebound symptoms associated with abrupt discontinuation of venlafaxine (Letter). J Clin Psychopharmacol *18:*343, 1998.

Derivan A, Entsuah R, Haskins T, Rudolph R, Venlafaxine XR 214 Study Group: Double-blind, placebo-/comparator-controlled study of once daily venlafaxine XR and buspirone in outpatients with generalized anxiety disorder. Eur Neuropsychopharmacol *8*(Suppl):S26, 1998.

*Dewan MJ, Anand VS: Evaluating the tolerability of the newer antidepressants. J Nerv Ment Dis *187:*96, 1999.

Dierick M, Ravizza L, Realini R, Martin A: A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. Prog Neuropsychopharmacol Biol Psychiatry *20:*57, 1996.

Entsuah AR, Rudolph RL, Hackett D, Miska S: Efficacy of venlafaxine and placebo during long-term treatment of depression: A pooled analysis of relapse rates. Int Clin Psychopharmacol *11:*137, 1996.

Ereshefsky L: Drug-drug interactions involving antidepressants: Focus on venlafaxine. J Clin Psychopharmacol *16*(Suppl):375, 1996.

Guelfi JD, Ansseau M, Corruble E, Samuelian JC, Tonelli I, Tournoux A, Pletan Y: A double-blind comparison of the efficacy and safety of milnacipran and fluoxetine in depressed inpatients. Int Clin Psychopharmacol *13:*121, 1998.

Guelfi JD, White C, Hackett D, Guichoux JY, Magni G: Effectiveness of venlafaxine in patients hospitalized for major depression and melancholia. J Clin Psychiatry *56:*450, 1995.

Haskins T, Rudolph R, Pallay A, Derivan A, for the Venlafaxine XR 210 Study Group: Double-blind, placebo-controlled study of once daily venlafaxine XR in outpatients with generalized anxiety disorder (GAD). Eur Neuropsychopharmacol *8*(Suppl):S26, 1998.

Kasper S, Pletan Y, Solles A, Tournoux A: Comparative studies with milnacipran and tricyclic antidepressants in the treatment of patients with major depression: A summary of clinical trials results. Int Clin Psychopharmacol *11*(Suppl):35, 1996.

Kelsey JE: Dose-response relationship with venlafaxine. J Clin Psychopharmacol *16*(Suppl):21S, 1996.

Khan A, Upton GV, Rudolph RL, Entsuah R, Leventer SM, Velanfaxine Investigator Study Group: The use of venlafaxine in the treatment of major depression and major depression associated with anxiety: A dose-response study. J Clin Psychopharmacol *18:*19, 1998.

Klamerus KJ, Maloney K, Rudolph RL, Sisenwine SF, Jusko WJ, Chang ST: Introduction of a composite parameter to the pharmacokinetics of venlafaxine and its active O-desmethyl metabolite. J Clin Pharmacol *32:*716, 1992.

Lecrubier Y, Pletan Y, Solles A, Tournoux A, Magne V: Clinical efficacy of milnacipran. Placebo-controlled trials. Int Clin Psychopharmacol *11*(Suppl):29, 1996.

Lopez-Ibor J, Guelfi JD, Pletan Y, Tournoux A, Prost JF: Milnacipran and selective serotonin inhibitors in major depression. Int Clin Psychopharmacol *11*(Suppl):41, 1996.

Louie AK, Lannon RA, Kiesch MA, Lewis TB: Venlafaxine withdrawal reactions (Letter). Am J Psychiatry *153:*1652, 1996.

Montgomery SA, Prost JF, Solles A, Briley M: Efficacy and tolerability of milnacipran: An overview. Int Clin Psychopharmacol *11*(Suppl):47, 1996.

Nierenberg AA, Feighner JP, Rudolph R, Cole JO, Sullivan J: Venlafaxine for treatment-resistant unipolar depression. J Clin Psychopharmacol *14:*419, 1994.

*Nutt D, Johnson FN: Potential applications of venlafaxine. Rev Contemp Pharmacother *9:*321, 1998.

*Puech A, Montgomery SA, Prost JF, Solles A, Briley M: Milnacipran, a new serotonin and noradrenaline reuptake inhibitor: An overview of its antidepressant activity and clinical tolerability. Int Clin Psychopharmacol *12:*99, 1997.

Puozzo C, Leonard BE: Pharmacokinetics of milnacipran in comparison with other antidepressants. Int Clin Psychopharmacol *11*(Suppl):15, 1996.

Ravindran AV, Charbonneau Y, Zaharia MD, Al-Zaid K, Wiens A, Anisman H: Efficacy and tolerability of venlafaxine in the treatment of primary dysthymia. J Psychiatry Neurosci *23:*288, 1998.

Rudolph RL, Deriran AT: The safety and tolerability of venlafaxine hydrochloride: Analysis of the clinical trials database. J Clin Psychopharmacol *16*(Suppl):54S, 1996.

*Sambunaris A, Hesselink JK, Pinder R, Panagides J, Stahl SM: Development of new antidepressants. J Clin Psychiatry *58:*40, 1997.

Schweizer E, Feighner J, Mandos LA, Rickels K: Comparison of venlafaxine and imipramine in the acute treatment of major depression in outpatients. J Clin Psychiatry *55:*104, 1994.

*Silverstone PH, Ravindran A: Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. Venlafaxine XR 360 Study Group. J Clin Psychiatry *60:*22, 1999.

Spencer CM, Wilde MI: Milnacipran: A review of its use in depression. Drugs *56:*405, 1998.

Tignol J, Pujol-Domenech J, Chartres JP, Leger JM, Pletan Y, Tonelli I, Tournoux A, Pezous N: Double-blind study of the efficacy and safety of milnacipran and imipramine in elderly patients with major depressive episode. Acta Psychiatr Scand *97:*157, 1998.

Wong DT, Bymaster FP, Mayle DA, Reid LR, Krushinski JH, Robertson DW: LY248686, a new inhibitor of serotonin and norepinephrine uptake. Neuropsychopharmacology *8:*23, 1993.

Szabadi E: Fluvoxamine withdrawal syndrome. Br J Psychiatry *160:*283, 1992.

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25A INTRDUCTION AND OVERVIEW

JEFFREY E. KELSEY, M.D., FH.D., AND CHARLES B. NEMEROFF, M.D., FH.D.

The first of the selective serotonin reuptake inhibitors (SSRIs) (also called serotonin-specific reuptake inhibitors), fluoxetine (Prozac) was introduced in the United States in 1988. Since that time, sertraline (Zoloft) and paroxetine (Paxil) were approved for the treatment of depression and subsequently for panic disorder and obsessive-compulsive disorder. In 1998 citalopram (Celexa), a very selective SSRI, was introduced in the United States. Fluvoxamine (Luvox) is approved in the United States for the treatment of obsessive-compulsive disorder and in other countries for depression. The availability of these agents as well as other newer-generation antidepressants such as venlafaxine (Effexor) nefazodone (Serzone), and mirtazepine (Remeron) has markedly changed the treatment of depression and certain anxiety disorders. Currently, fluoxetine is the second most widely prescribed medication in the United States. Members of this class are characterized by a generally more acceptable adverse effect burden than both tricyclic drugs and monoamine oxidase inhibitors (MAOIs) and a much higher safety margin. None of the SSRIs act as fast as sodium channel inhibitors and thus they do not display the cardiac toxicity seen in tricyclic drug overdose. In addition, SSRIs are not active at a-adrenergic receptors, obviating the orthostatic hypotension associated with the use of older antidepressants including tricyclic drugs, MAOIs, and trazodone (Desyrel).

Since the introduction of fluoxetine, SSRIs have been used successfully to treat an expanding variety of psychiatric disorders. The widespread use of the SSRIs has also provoked a public debate about the widespread use of psychotropic medications in general and antidepressants in particular.

The therapeutic activity of the SSRIs is believed to result from their ability to inhibit the reuptake of serotonin into presynaptic serotonergic nerve terminals, though SSRIs do exert effects on other neurotransmitter systems as well. Studies comparing one SSRI with another have failed to demonstrate any difference in efficacy in the treatment of major depression. One must be cautious in extrapolating the lack of differences in group populations to individual patients. It has been shown repeatedly that a sizable percentage of depressed patients who fail an adequate trial of one SSRI responds to a second SSRI, likely because SSRIs do differ from one another. Each SSRI is structurally unique, and differences exist in the selectivity of reuptake inhibition, medication interactions, and pharmacokinetics of drugs in this class. For example, at clinically effective dosages paroxetine likely inhibits not only serotonin reuptake but norepinephrine reuptake as well. Similarly, sertraline likely blocks both serotonin reuptake and dopamine reuptake. Reuptake potencies of the SSRIs based on in vitro assays for serotonin, norepinephrine, and dopamine are shown in Table 31.25a-1.

	K_i (nM)			
Drug	Serotonin	Norepinephrine	5-HT:NE*	Dopamine
Citalopram	1.3	4,000	3078	28,000
Fluoxetine	6	1,100	183	>10,000
Fluvoxamine	25	500	20	4,200
Paroxetine	1	350	350	2,000
Sertraline	7	1,400	200	230

* Ratio of serotonin to norepinephrine. Not all values are from the same study, so the relative ratios of selectivity (5-HT:NE) are the representative values.

Table 31.25a-1 Relative Selectivity of Serotonin, Norepinephrine or Dopamine Reuptake Inhibition in an In Vitro Assay

PHARMACOKINETICS

The SSRIs as a class exhibit a wide range of serum half-lives ranging from hours to days. Table 31.25a-2 summarizes the pharmacokinetic characteristics of this class of agents. SSRIs as a class demonstrate high protein binding. The significance of this in terms of potential drug-drug interactions remains relatively unexplored. The compartment associated with the biological activity of a drug is the free drug plasma concentration. For fluoxetine, paroxetine and sertraline this represents less than 6 percent of the total drug present in serum, based on in vitro calculations. However the free drug can also be metabolized. The question remains whether or not clinically significant displacement of protein binding occurs with these compounds and drugs of other therapeutic classes such as warfarin (Coumadin). Both albumin and a_1-acid-glycoprotein are involved, though the relative proportions are often not well characterized. Extrapolation of this to clinical practice suggests that other therapeutic compounds with high protein binding and low therapeutic indexes should be monitored when they are coprescribed with SSRIs.

Drug	Time to Peak Plasma Concentration	Half-Life	Half-Life Metabolite	Time to Steady State (days)	Plasma-Protein Binding (%)
Fluoxetine	6-8 h	4-6 days	4-16 days	28-35	95
Fluvoxamine	3-8 h	15 h	–	3-7	80
Paroxetine	5-6 h	21 h	–	5-10	95
Sertraline	4.5-8.5 h	26 h	62-104 h	3-7	95
Citalopram	4	35 h	35 h	7	80

Table 31.25a-2 Pharmacokinetic Profiles of the SSRIs

DRUG INTERACTIONS

Drug-drug interactions among SSRIs and other therapeutic drugs have received increasing scrutiny in the past few years; however, firm recommendations for clinical practice remain less than ideal. The two major classes of drug-drug interactions involving SSRIs are protein binding and enzyme inhibition through the cytochrome P450 (CYP) isoenzyme system. Indications that a drug-drug interaction is occurring, either by protein-binding displacement or inhibition of metabolism include increased adverse effects at a lower-than-expected dosage or therapeutic efficacy at a lower antidepressant dosage than expected. However some drug-drug interactions have only silent side effects that may go unnoticed. Ultimately, the pharmacokinetic interactions must be considered in terms of pharmacodynamic

consequences.

Inhibition of the cytochrome P450 system by antidepressants is receiving increasing study. The cytochrome P450 system is a collection of more than 30 different hepatic isoenzymes that show peak absorbance at 450 nm in a spectrophotometer after exposure to carbon monoxide. These isoenzymes metabolize endogenous substances, substances found in food and the environment, and drugs. Although a number of isoenzymes have been identified by molecular biological techniques, only a small proportion have been well characterized. Each isoenzyme has a three-character name: a number that denotes the family, a letter denoting the subfamily, and another number, which specifies the gene. This is seen in the names of the most commonly described isoenzymes such as CYP 1A2, 2D6, and 3A4. Genetic polymorphism of certain of these isoenzymes results in some individuals who are totally or relative deficient (i.e., poor metabolizers) in a particular isoenzyme. The extent of isoenzyme polymorphism across various ethnic groups has not been extensively studied.

Antidepressants are typically divided into groups with high, medium or low potential isoenzyme inhibition, as shown in Table 31.25a-3. A number of commonly prescribed medications and the isoenzymes that metabolize them are shown in Table 31.25a-4. Note that some medications such as the tertiary tricyclic drugs have multiple pathways of metabolism. Examples of medication interactions include fluvoxamine and theophylline (Theo-Dur) via the CYP 1A2 isoenzyme; clozapine (Clozaril) and fluvoxamine (CYP 1A2); sertraline, fluoxetine, or paroxetine (CYP 2D6); and paroxetine, fluoxetine, and sertraline with tricyclic drugs through the CYP 2D6 isoenzyme.

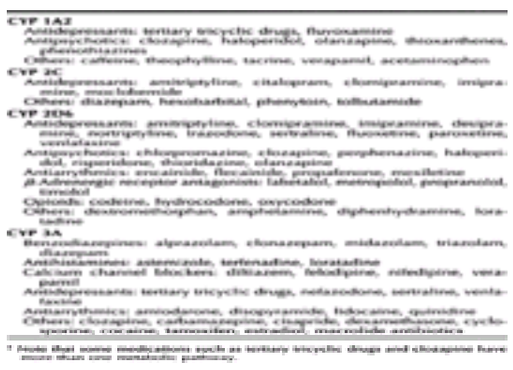
CYP 1A2
Antidepressants: tertiary tricyclic drugs, fluvoxamine
Antipsychotics: clozapine, haloperidol, olanzapine, thioxanthenes, phenothiazines
Others: caffeine, theophylline, tacrine, verapamil, acetaminophen
CYP 2C
Antidepressants: amitriptyline, citalopram, clomipramine, imipramine, mirtazapine, nortriptyline, trazodone, sertraline, fluoxetine, paroxetine, venlafaxine
Antipsychotics: chlorpromazine, clozapine, perphenazine, haloperidol, risperidone, thioridazine, olanzapine
Antiarrhythmics: mexiletine, flecainide, propafenone, encainide
β-Adrenergic receptor antagonists: labetalol, metoprolol, propranolol, timolol
Opioids: codeine, hydrocodone, oxycodone
Others: dextromethorphan, amphetamine, diphenhydramine, loratadine
CYP 3A
Benzodiazepines: alprazolam, clonazepam, midazolam, triazolam, diazepam
Antihistamines: astemizole, terfenadine, loratadine
Calcium channel blockers: diltiazem, felodipine, nifedipine, verapamil
Antidepressants: tertiary tricyclic drugs, nefazodone, sertraline, venlafaxine
Antiarrhythmics: amiodarone, disopyramide, lidocaine, quinidine
Others: clozapine, carbamazepine, cisapride, dexamethasone, cyclosporine, cerivastatin, tamoxifen, estradiol, macrolide antibiotics

† Note that some medications such as tertiary tricyclic drugs and clozapine have more than one metabolic pathway.

Table 31.25a-3 Commonly Prescribed Medications That Are Substrates of the Cytochrome P450 1A2, 2C, 2D6, or 3A4 Isoenzymes

Relative Rank	CYP 1A2	CYP 2C	CYP 2D6	CYP 3A
Higher	Fluvoxamine Grapefruit	Fluoxetine Fluvoxamine	Paroxetine Fluoxetine	Fluvoxamine Nefazodone Troleandomycin Grapefruit
Moderate	Tertiary amine tricyclics Fluoxetine	Sertraline	Secondary amine tricyclics Sertraline	Fluoxetine Sertraline
Low or minimal	Paroxetine Venlafaxine Nefazodone Sertraline Bupropion Mirtazapine Citalopram	Venlafaxine Paroxetine Citalopram	Fluvoxamine Venlafaxine Nefazodone Bupropion Mirtazapine Citalopram	Venlafaxine Paroxetine Mirtazapine Citalopram

Table 31.25a-4 Interaction Potentials of Antidepressant Drugs at Cytochrome 450

Whether or not isoenzyme inhibition has clinical relevance depends on a number of factors. Patients taking multiple medications are at risk if those medications have a low therapeutic index or do not have alternative pathways of metabolism or if the patients have clinically significant renal or hepatic disease. Other risk factors include the presence of highly potent isoenzyme inhibitors such as quinidine (Duraquin) at the 2D6 isoenzyme or ketoconazole (Nizoral) at the 3A4 isoenzyme. An informed awareness of potential drug-drug interactions and monitoring for them is the best way to maximize clinical care and minimize avoidable risks.

The SSRIs are safe, well-tolerated medications that have benefited millions of patients worldwide. Dosing is less complicated than with tricyclic drugs, and initial dosing at a low level with gradual escalation is a successful and well-tolerated strategy for most patients. Administration with meals at initial dosing helps minimize nausea, which is perhaps the most common early adverse effect of the SSRIs. Sexual dysfunction was poorly recognized during early clinical trials but is now recognized to affect a significant number of patients. Taken together, the risk-benefit relationship with this class of medications is favorable in the vast majority of patients.

SUGGESTED CROSS-REFERENCES

Additional information regarding monoamines and their receptors appears in Section 1.4. The basic electrophysiology of neurons is discussed in Section 1.9. The neurobiology of mood disorders is discussed in Section 16.4, and biochemical aspects of anxiety disorders are covered in Section 15.3.

SECTION REFERENCES

*Alfaro CL, Lam YMF, Simpson J, Ereshefsky L: CYP2D6 status of extensive metabolizers after multiple-dose fluoxetine, fluvoxamine, paroxetine, or sertraline. J Clin Psychopharmacol *19:*155, 1999.

Benfield P, Heel RC, Lewis SP: Fluoxetine: A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in depressive illness. Drugs *32:*481, 1986.

Bertilsson L, Dahl ML, Tybring G: Pharmacogenetics of antidepressants: Clinical aspects. Acta Psychiatr Scand Suppl *391:*14, 1997.

Brosen K: Are pharmacokinetic drug interactions with the SSRIs an issue? Int Clin Psychopharmacol *11*(Suppl):23, 1996.

DeVane CL: The place of selective serotonin reuptake inhibitors in the treatment of panic disorder. Pharmacol Ther *17:*282, 1997.

*Dewan MJ, Anand VS: Evaluating the tolerability of the newer antidepressants. J Nerv Ment Dis *187:*96, 1999.

Haffmans PM, Timmerman L, Hoogduin CA: Efficacy and tolerability of citalopram in comparison with fluvoxamine in depressed outpatients: A double-blind, multicentre study. The LUCIFER Group. Int Clin Psychopharmacol *11:*157, 1996.

Hollander E: Obsessive-compulsive disorder-related disorders: The role of selective serotonergic reuptake inhibitors. Int Clin Psychopharmacol *11*(Suppl):75, 1996.

Kasper S, Fuger J, Moller HJ: Comparative efficacy of antidepressants. Drugs *43*(Suppl):11, 1992.

*Lane R, Baldwin D: Selective serotonin reuptake inhibitor-induced serotonin syndrome: Review. J Clin Psychopharmacol *17:*208, 1997.

Lejoyeux M: Use of serotonin (5-hydroxytryptamine) reuptake inhibitors in the treatment of alcoholism. Alcohol Alcoholi *31*(Suppl):69, 1996.

Leonard BE: Pharmacological differences of serotonin reuptake inhibitors and possible clinical relevance. Drugs *43*(Suppl):3, 1992.

*Leonard HL, March J, Rickler KC, Allen AJ: Pharmacology of the selective serotonin reuptake inhibitors in children and adolescents. J Am Acad Child Adolesc Psychiatry *36:*725, 1997.

Margolese HC, Assalian P: Sexual side effects of antidepressants: A review. J Sex Marital Ther *22:*209, 1996.

Milne RJ, Goa KL: Citalopram. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depressive illness. Drugs *41:*450, 1991.

*Montejo-Gonzalez AL, Llorca G, Izquierdo JA, Ledesma A, Bousono M, Calcedo A, Carrasco JL, Ciudad J, Daniel E, De la Gandara J, Derecho J, Franco M, Gomez MJ, Macias JA, Martin T, Perez V, Sanchez JM, Sanchez S, Vicens E: SSRI-induced sexual dysfunction: Fluoxetine, paroxetine, sertraline, and fluvoxamine in a prospective, multicenter, and descriptive clinical study of 344 patients. J Sex Marital Ther *23:*176, 1997.

Montgomery SA, Rasmussen JG, Tanghoj P: A 24-week study of 20 mg citalopram, 40 mg citalopram, and placebo in the prevention of relapse of major depression. Int Clin Psychopharmacol *8:*181, 1993.

Nemeroff CB: Paroxetine: An overview of the efficacy and safety of a new selective serotonin reuptake inhibitor in the treatment of depression. J Clin Psychopharmacol *13*(Suppl):10S, 1993.

*Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system [see Comments]. Am J Psychiatry *153:*311, 1996.

Newhouse PA: Use of serotonin selective reuptake inhibitors in geriatric depression. J Clin Psychiatry *57*(Suppl):12, 1996.

Rechlin T: The effect of amitriptyline, doxepin, fluvoxamine, and paroxetine treatment on heart rate variability. J Clin Psychopharmacol *14:*392, 1994.

Richelson E: Pharmacokinetic drug interactions of new antidepressants: A review of the effects on the metabolism of other drugs. Mayo Clin Proc *72:*835, 1997.

Samoil D, Grubb BP: Neurally mediated syncope and serotonin reuptake inhibitors [Review]. Clin Autonom Res *5:*251, 1995.

Schatzberg AF: Treatment of severe depression with the selective serotonin reuptake inhibitors. Depression Anxiety *4:*182, 1996.

Shader RI, von Moltke LL, Schmider J, Harmatz JS, Greenblatt DJ: The clinician and drug interactions—an update [Editorial]. J Clin Psychopharmacol *16:*197, 1996.

*Sheehan DV, Harnett-Sheehan K: The role of SSRIs in panic disorder. J Clin Psychiatry *57*(Suppl):51, 1996.

Stokes PE, Holtz A: Fluoxetine tenth anniversary update: The progress continues. Clin Ther *19:*1135, 1997.

von Moltke LL, Greenblatt DJ, Schmider J, Harmatz JS, Shader RI: Metabolism of drugs by cytochrome P450 3A isoforms. Implications for drug interactions in psychopharmacology. Clin Pharmacokinet *29*(Suppl):33, 1995.

Wade AG, Lepola U, Koponen HJ, Pedersen V, Pedersen T: The effect of citalopram in panic disorder. Br J Psychiatry *170:*549, 1997.

Wilde MI, Plosker GL, Benfield P: Fluvoxamine. An updated review of its pharmacology, and therapeutic use in depressive illness. Drugs *46:*895, 1993.

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25B CITALOPRAM

JEFFREY E. KELSEY, M.D., FH.D., AND CHARLES B. NEMEROFF, M.D., FH.D.

HISTORY

Citalopram (Celexa) is the most recently approved selective serotonin reuptake inhibitor (SSRI) in the United States, though it has been available for a number of years in over 50 other countries worldwide. Currently, citalopram is approved by the Food and Drug Administration (FDA) for the treatment of major depressive disorders but as with the other SSRIs, data suggest efficacy in dysthymic disorder, anxiety disorders, and bipolar disorder depression.

CHEMISTRY

The molecular formula of citalopram is $C_{20}H_{22}BrFNO_2$, with a molecular weight of 405.35. The chemical designation is (±)-1-(3-dimethylaminopropyl)-1-(4-fluorophenyl)-1,3-dihydroisobenzo-furan-5-carbonitrile, and citalopram is a racemic mixture of the hydrogen bromide salt. The structure of citalopram is shown in Figure 31.25b-1.

FIGURE 31.25b-1 Molecular structure of citalopram.

PHARMACOLOGICAL ACTIONS

Absorption Citalopram demonstrates good absorption orally, with approximately 80 percent of a 40-mg dose being absorbed systemically relative to an intravenous dose. Food does not interfere with the absorption of citalopram, and peak plasma concentrations are achieved approximately 4 hours following oral dosing. Citalopram is available in the United States as tablets and is also available as an intravenous solution in other parts of the world.

Distribution The volume of distribution of citalopram is about 12L/kg. In vitro, citalopram and two metabolites, demethylcitalopram (DCT) and didemethylcitalopram (DDCT) are approximately 80 percent bound to serum proteins. The rate of plasma clearance is about 330 mL/per hour for citalopram, of which 20 percent of the clearance is due to renal excretion of the citalopram itself.

Metabolism and Elimination Citalopram is a racemic mixture of *R*-citalopram and *S*-citalopram in equal proportions. The *S*-enantiomer of citalopram as the primary component that demonstrates inhibition of serotonin reuptake and appears to be metabolized at a slightly faster rate than the *R*-enantiomer. After intravenous administration of the racemic mixture, citalopram is metabolized to DCT, DDCT, and citalopram-*N*-oxide. There is also a pharmacologically inactive propionic acid derivative that is presumably formed by deamination of citalopram, DCT, or DDCT. The *N*-demethylation of citalopram occurs in vitro via the cytochrome P450 (CYP) 3A4 and CYP 2C19 pathways. The parent compound, citalopram, is the predominant molecule found in the systemic circulation, with concentrations of DCT and DDCT being about 50 and 10 percent of those of citalopram, respectively. Citalopram is eight times more potent than its metabolites in inhibiting serotonin reuptake.

The elimination half-life of citalopram is close to 35 hours, with hepatic metabolism as the primary pathway. The pharmacokinetics of citalopram are linear and dosage-proportional over the usual dosing range of 10 to 60 mg a day. At steady state, mean plasma concentrations of citalopram are approximately 120 and 367 nmol/L for subjects receiving 20 and 60 mg a day, respectively. The expected steady-state concentration of citalopram in plasma is approximately 2.5 times that following a single dose.

Geriatric patients, defined in the pharmacokinetic studies as patients 60 years of age or older, show reduced clearance of citalopram. The citalopram area under the concentration-time curve (AUC) and half-life were increased by 23 to 30 percent and 30 to 50 percent, respectively, across single-dose and multiple-dose studies in geriatric patients. The recommended dosage for most geriatric patients is 20 mg a day. One small study ($N = 32$) suggested that women have a greater AUC for citalopram than men; however, this finding was not reproduced in larger studies ($N = 625$). Impairment of hepatic function is associated with impaired metabolism of citalopram, with a resultant 37 percent decrease in clearance of orally administered citalopram and a twofold increase in the half-life. In contrast, mild-to-moderate renal impairment resulted in only a 17 percent decrease in citalopram clearance. There are in sufficient data to determine the extent of changes in clearance in patients with more severe renal impairment. Dosages of 20 mg a day are usually considered to be the maximum dosage for geriatric patients or those with hepatic impairment; a dosage adjustment is usually not necessary for mild-to-moderate renal impairment.

PHARMACODYNAMICS

Mechanism of Action The therapeutic activity of citalopram is believed to be due to specific serotonin reuptake inhibition at the presynaptic serotonergic nerve terminal. Citalopram, the most selective of the SSRIs, shows more than a 3000-fold higher potency for blocking serotonin reuptake than norepinephrine reuptake, with Kis in vitro of 1.3, 4000, and 28,000 nmol/L for serotonin, norepinephrine, and dopamine, respectively. Citalopram lacks any clinically significant receptor affinity for muscarinic cholinergic, histamine type 1 (H_1), a_1-a_2-, and b-adrenergic, dopamine type 1 (D_1) or D_2, serotonin (5-hydroxytryptamine [5-HT]) subtype 1a ($5\text{-}HT_{1A}$) or $5\text{-}HT_{2A}$, g-aminobutyric acid (GABA), and benzodiazepine receptors.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

A number of clinical trials have investigated the efficacy of citalopram in major depression, in which thousands of patients have received the compound. Citalopram is available as both an intravenous infusion and tablets for oral administration.

Major Depressive Disorder A fixed-dose, double-blind, placebo-controlled 6-week trial of citalopram at 10, 20, 40 or 60 mg in 650 outpatients with revised third edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM III-R) major depression showed superiority of the active drug at both 40 and 60 mg a day. Study participants all had a 1-week placebo run-in period and then, if still eligible, were randomized to placebo or one of the citalopram dosing arms. Subjects in the 10- and 20-mg a day group started treatment at that dosage, those in the higher dosage groups were titrated from a starting dosage of 20 mg a day for days 1 to 3, and 40 mg a day for days 4 to 6. Study participants randomized to the 60 mg a day dosage had the final dosage increase on day 7. Using a last observation carried forward (LOCF) analysis, the two lower dosages, 10 and 20 mg a day, did not separate from placebo by week 6 in Hamilton Rating Scale for Depression (HAM-D) scores. The two higher dosages, 40 and 60 mg a day, were both statistically superior to placebo at endpoint but did not differ from each other.

A second recent outpatient study used a flexible dosing strategy and raised the dosage of citalopram to 80 mg a day, or if that was not tolerable, to the highest tolerated dose. In this study, the average dosage of citalopram was 30 mg a day at week 1 and 52 mg a day at week 4. There was an overall response rate of 80 percent in the citalopram-treated group.

Citalopram prevented relapse following improvement in a short-term trial. In one study, patients who responded to open-label treatment with citalopram (20 to 60 mg a day) as evidenced by a score on the Montgomery Asberg Depression Rating Scale (MADRS) of 12 or less, were randomized to either placebo or ongoing citalopram. After the 24-week double-blind phase, 18 of 74 (24.3 percent) patients receiving placebo had relapsed, compared with only 21 of 152 (13.8 percent) who continued to receive citalopram. Relapse was defined as a score of 25 or above on the MADRS. A second 24-week maintenance study comparing citalopram at doses of 20 or 40 mg a day demonstrated less relapse with active drug than with placebo treatment.

A comparison study of citalopram and sertraline (Zoloft) in a general practice setting suggested no differences in efficacy between the two drugs. The mean doses of citalopram and sertraline were 34 and 82 mg a day, respectively. Both drugs were well tolerated. Studies comparing citalopram with either fluvoxamine (Luvox) or fluoxetine (Prozac) in depressed outpatients have also suggested equal efficacy and tolerability. An inpatient study conducted by the Danish University Antidepressant Group suggested that the tricyclic antidepressant clomipramine (Anafranil), which inhibits both norepinephrine and serotonin reuptake, had better efficacy than citalopram in an inpatient setting when the criterion for response was a HAM-D score of 7 or less. In contrast to this finding, a comparison of citalopram, an SSRI, to maprotiline (Ludiomil), a norepinephrine-selective tricyclic drug, found no difference in efficacy between the two in 29 hospitalized patients.

Comparison of orally or intravenously administered citalopram in a small study (*N* = 30 per group) suggested that intravenous administration for the first 10 days, at a dosage of 40 mg a day, produced a more rapid response than oral administration at the same dosage. Both routes of delivery were equally well tolerated.

Trichotillomania A small open-label study in patients with trichotillomania showed only a 38.5 percent response rate (Clinical Global Impression [CGI] of 2 or less) after 12 weeks of treatment. The mean dosage of citalopram in the 11 of 12 subjects who completed the trial was 36.2 mg a day.

Obsessive-Compulsive Disorder A small (*N* = 29) open-label study examined the efficacy of citalopram in obsessive-compulsive disorder. Dosages of 40 to 60 mg a day were reported to have efficacy in about 75 percent of subjects.

Panic Disorder Citalopram treatment of eight patients at a dosage of 20 mg a day significantly decreased the frequency of panic attacks induced by a cholecystokinin tetrapeptide challenge.

Social Phobia An open-label trial of 22 patients with social phobia found an 86 percent response rate among study participants after 12 weeks of treatment at 40 mg a day.

Substance Dependence A placebo-controlled comparison trial of citalopram (20 mg a day) and fluvoxamine (150 mg a day) in detoxified alcoholics reported them to be equipotent in maintaining abstinence when combined with cognitive behavioral therapy. Citalopram-treated patients also reported reduced craving for alcohol. Another study, however, found citalopram to be better than placebo only in a subgroup of males below the group median of alcohol consumption. In this group of 30 men with mean alcohol consumption of 111 ± 51 grams a day of pure alcohol, only those with the lower daily consumption (mean 85 ± 15 grams a day of pure alcohol) derived more benefit from citalopram than from placebo.

Fibromyalgia A placebo-controlled trial of citalopram, up to 40 mg a day, revealed no superiority of the SSRI over placebo treatment. Twenty-two patients with fibromyalgia were randomized to 20 mg a day of citalopram or placebo, with the option to increase to 40 mg a day after 4 weeks if there was no response. A variety of outcome measures including self-report and physician assessment failed to differentiate the two groups.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Respiratory System No adverse respiratory effects noted with citalopram have differed significantly from those observed with placebo in premarketing trials. Allergic reactions to citalopram have been reported with the potential for respiratory involvement.

Cardiovascular System The premarketing clinical trials database showed a heart rate 1.7 beats per minute lower in subjects receiving citalopram than in those receiving placebo. Such clinically insignificant decreases in heart rate have been seen after treatment with other SSRIs. There were no changes in cardiac conduction as measured by electrocardiographic (ECG) changes associated with citalopram treatment. Animal studies in beagle dogs reported sudden death in 5 of 9 dogs treated with citalopram at a dosage of 8 mg/kg a day for 17 to 31 weeks. Elevated DDCT concentrations were considered responsible for QTc prolongation and subsequent fatal arrhythmias. The metabolism of citalopram differs to a great extent in beagles and humans with the former exhibiting markedly higher levels of DDCT. Specific analysis of the QTc in premarketing clinical trials failed to reveal any abnormalities associated with citalopram treatment.

No significant changes in supine or orthostatic blood pressure have been observed in association with citalopram treatment. However, premarketing clinical trials generally exclude patients with preexisting cardiac disease.

An analysis reported from Sweden suggests that with a single drug overdose of citalopram, some patients who ingest more than 600 mg exhibit ECG conduction changes. Doses exceeding 1900 mg invariably produce ECG changes.

Blood Citalopram has been rarely associated with reports of abnormal bleeding. Medications that inhibit serotonin reuptake are believed to inhibit platelet aggregation, which may produce bleeding or bruising in susceptible individuals.

Gastrointestinal System Citalopram, like the SSRIs, has the potential to produce nausea. The incidence of nausea in the premarketing clinical trials was 21 percent in the citalopram-treated group and 14 percent in the placebo-treated group. In 1036 patients receiving citalopram, nausea was the most common adverse event associated with discontinuation, with a rate of 4 percent. Nausea with this class of medications is usually dose related and transient. Dosing citalopram with meals initially or starting at a reduced dosage of 10 mg a day often reduces the incidence of nausea. Short-term treatment trials with citalopram (6 to 8 weeks duration) were associated with a weight loss of 1.1 pounds on average. The long-term effects of citalopram on body weight are not known.

Skin Rashes or pruritus were reported in the premarketing clinical trials at a rate of at least 1 percent.

Central Nervous System Tremor was reported by 8 percent of clinical trial participants receiving citalopram and 6 of those who received placebo. Dizziness was cited as the reason for study discontinuation by 2 percent of study participants receiving citalopram in premarketing trials.

THERAPEUTIC INDICATIONS

Citalopram currently has been approved by the FDA for the treatment of major depressive disorder. A variety of clinical experiences ranging from placebo-controlled, double-blind trials to case reports points to the efficacy of citalopram in other affective and anxiety disorders.

PRECAUTIONS AND ADVERSE REACTIONS

The most prevalent adverse events associated with citalopram treatment, mostly reported from non-U.S. trials, include nausea, dry mouth, somnolence, sweating, tremor, diarrhea, and ejaculatory dysfunction. This side-effect profile is typical of SSRIs, and in general, citalopram is quite well tolerated.

Sexual dysfunction was reported in the premarketing trials at a low rate, but the actual incidence in clinical practice is thought to be higher, similar to that with the other SSRIs.

DRUG INTERACTIONS

Evidence indicates that citalopram has the lowest potential for untoward drug-drug interactions of the SSRIs, probably because of its benign effects on the cytochrome P450 isoenzyme system. Citalopram is considered to be a weak inhibitor of the CYP 1A2, 2C19 and 2D6 isoenzymes, with no appreciable activity on the CYP 3A4 isoenzyme. Nevertheless, drug-drug interactions have been reported with cimetidine (Tagamet) and metoprolol (Lopressor). Cimetidine increases the AUC and maximum serum concentration of citalopram. In contrast, citalopram, 40 mg a day for 22 days, approximately doubled the serum concentration of metoprolol, a b-adrenergic receptor antagonist CYP that is metabolized by the CYP 2D6 isoenzyme. Citalopram displays lower protein binding (80 percent) than the other SSRIs but not as low as venlafaxine (Effexor) (27 percent). Good clinical care requires being aware of the potential for drug-drug interactions, especially when a new medication is added to a treatment regimen.

Citalopram should not be combined with monoamine oxidase inhibitors (MAOIs), because of the potential for developing the serotonin syndrome. The manufacturer recommends a washout period of 14 days between citalopram and MAOIs when switching from one to the other.

LABORATORY INTERFERENCES

Citalopram does not appear to interfere with any laboratory testing.

DOSING AND ADMINISTRATION

Citalopram is available in 20- and 40-mg scored tablets. The tablets contain copolyvidone, cornstarch, crosscarmellolose sodium, glycerin, lactose monohydrate, magnesium stearate, hydroxypropyl methyl cellulose, microcrystalline cellulose, and polyethylene glycol. The 20-mg tablets also contain titanium dioxide and iron oxides as coloring agents.

The suggested starting dosage of citalopram in patients with major depression is 20 mg a day. Effective dosages for the treatment of major depression have ranged from 20 to 80 mg a day in dosage-finding trials, though there were reports of failure at 20 mg a day in some trials in which higher doses were successful. In general, if a nongeriatric patient tolerates a dosage of 20 mg a day and does not demonstrate the desired response, it is reasonable to increase the dosage to at least 40 mg a day, assuming intact hepatic function. Some patients with sensitivity to citalopram or other SSRIs may do best at a starting dosage of 10 mg a day for 4 to 6 days before increasing the dosage. Steady-state concentrations of citalopram are achieved after about 1 week of treatment. There are insufficient clinical trial data to suggest target dosages for other disorders. SSRIs as a class are generally used at higher dosages to treat obsessive-compulsive disorder and panic disorder than to treat major depression. Because patients with panic disorder are generally more sensitive to adverse effects of medications than those with major depression, gradual dosage titration is of paramount importance in this group.

Geriatric patients, age 65 or older, composed 24 percent of the patient population (1034 of 4422 subjects) in citalopram premarketing trials. The most common dosing range in this population was between 20 and 40 mg a day. Citalopram clearance is reduced in the geriatric patient. Insufficient clinical data exist to suggest dosing strategies in children, though one might expect dosages to be similar to those for healthy adults, but adjusted for body mass.

SUGGESTED CROSS-REFERENCES

Additional information regarding monoamines and their receptors appears in Section 1.4. The basic electrophysiology of neurons is discussed in Section 1.9. The classification of mental disorders is discussed in Chapter 9, and mood disorders in geriatric patients in Section 51.3d. The neurobiology of mood disorders is discussed in Section 14.4. Lithium therapy is discussed in Section 31.18 and the biological treatment of mental illness in elderly adults in Section 51.4a to Section 51.4g.

SECTION REFERENCES

Angelone SM, Bellini L, Di Bella D, Catalano M: Effects of fluvoxamine and citalopram in maintaining abstinence in a sample of Italian detoxified alcoholics. Alcohol Alcohol *33:*151, 1998.

*Bajo S, Battaglia M, Pegna C, Bellodi L: Citalopram and fluvoxamine in Tourette's disorder. J Am Acad Child Adolesc Psychiatry *38:*230, 1999.

Balldin J, Berggren U, Engel J, Eriksson M, Hard E, Soderpalm B: Effect of citalopram on alcohol intake in heavy drinkers. Alcohol Clin Exp Res *18:*1133, 1994.

Baumann P, Nil R, Bertschy G, Jecker A, Brandli H, Morand J, Kasas A, Vuagniaux O, Ramseier F: A double-blind double-dummy study of citalopram comparing infusion versus oral administration. J Affect Disord *49:*203, 1998.

Bjerkenstedt L, Edman G, Flyckt L, Hagenfeldt L, Sedvall G, Wiesel FA: Clinical and biochemical effects of citalopram, a selective 5-HT reuptake inhibitor—a dose-response study in depressed patients. Psychopharmacology *87:*253, 1985.

Bouwer C, Stein DJ: Use of the selective serotonin reuptake inhibitor citalopram in the treatment of generalized social phobia. J Affect Disord *49:*79, 1998.

Ekselius L, von Knorring L, Eberhard G: A double-blind multicenter trial comparing sertraline and citalopram in patients with major depression treated in general practice. Int Clin Psychopharmacol *12:*323, 1997.

*Favre MP, Sztagzel J, Bertschy G: Bradycardia during citalopram treatment: A case report. Pharmacol Res *39:*149, 1999.

Fuglum E, Rosenberg C, Damsbo N, Stage K, Lauritzen L, Bech P: Screening and treating depressed patients. A comparison of two controlled citalopram trials across treatment settings: Hospitalized patients vs. patients treated by their family doctors. Danish University Antidepressant Group. Acta Psychiatr Scand *94:*18, 1996.

Gravem A, Amthor KF, Astrup C, Elgen K, Gjessing LR, Gunby B, Pettersen RD, Kyrdalen L, Vaadal J, Ofsti E, Aarvold A: A double-blind comparison of citalopram (Lu 10-171) and amitriptyline in depressed patients. Acta Psychiatr Scand *75:*478, 1987.

*Haffmans PM, Timmerman L, Hoogduin CA: Efficacy and tolerability of citalopram in comparison with fluvoxamine in depressed outpatients: A double-blind, multicentre study. The LUCIFER Group. Int Clin Psychopharmacol *11:*157, 1996.

Koponen H, Lepola U, Leinonen E, Jokinen R, Penttinen J, Turtonen J: Citalopram in the treatment of obsessive-compulsive disorder: An open pilot study. Acta Psychiatr Scand *96:*343, 1997.

*Montgomery SA, Rasmussen JG, Tanghoj P: A 24-week study of 20 mg citalopram, 40 mg citalopram, and placebo in the prevention of relapse of major depression. Int Clin Psychopharmacol *8:*181, 1993.

Norregaard J, Volkmann H, Danneskiold-Samsoe B: A randomized controlled trial of citalopram in the treatment of fibromyalgia. Pain *61:*445, 1995.

Patris M, Bouchard JM, Bougerol T, Charbonnier JF, Chevalier JF, Clerc G, Cyran C, Van Amerongen P, Lemming O, Hopfner Petersen HE: Citalopram versus fluoxetine: A double-blind, controlled, multicentre, phase III trial in patients with unipolar major depression treated in general practice. Int Clin Psychopharmacol *11:*129, 1996.

*Personne M, Sjoberg G, Persson H: Citalopram overdose—review of cases treated in Swedish hospitals. J Toxicol Clin Toxicol *35:*237, 1997.

*Price PL: Citalopram: Another SSRI antidepressant SDJ Med *52:*99, 1999.

*Robert P, Montgomery SA: Citalopram in doses of 20–60 mg is effective in depression relapse prevention: A placebo-controlled 6-month study. Int Clin Psychopharmacol *10*(Suppl):29, 1995.

Shlik J, Aluoja A, Vasar V, Vasar E, Podar T, Bradwejn J: Effects of citalopram treatment on behavioural, cardiovascular and neuroendocrine response to cholecystokinin tetrapeptide challenge in patients with panic disorder. J Psychiatry Neurosci *22:*332, 1997.

Sidhu J, Priskorn M, Poulsen M, Segonzac A, Grollier G, Larsen F: Steady-state pharmacokinetics of the enantiomers of citalopram and its etabolites in humans. Chirality *9:*686, 1997.

Stein DJ, Bouwer C, Maud CM: Use of the selective serotonin reuptake inhibitor citalopram in treatment of trichotillomania. Eur Arch Psychiatry Clin Neurosci *247:*234, 1997.

*Timmerman L, de Beurs P, Tan BK, Leijnse-Ybema H, Sanchez C, Hopfner Petersen HE, Cohen Stuart MH: A double-blind comparative clinical trial of citalopram vs maprotiline in hospitalized depressed patients. Int Clin Psychopharmacol *2:*239, 1987.

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25C FLUOXETINE

JEFFREY E. KELSEY, M.D., FH.D. AND CHARLES B. NEMEROFF, M.D., FH.D.

HISTORY

Approximately 10 years ago, fluoxetine (Prozac), a selective serotonin reuptake inhibitor (SSRI) was introduced into the United States. Since that time, fluoxetine has become the most commonly prescribed antidepressant in the United States. Although initially approved for the treatment of major depression, fluoxetine currently also has Food and Drug Administration (FDA) approval for the treatment of obsessive-compulsive disorder and bulimia nervosa. Evidence suggests that it is effective in the treatment of dysthymic disorder, panic disorder, social phobia, premenstrual syndrome, and bipolar disorder depression.

CHEMISTRY

Fluoxetine's empirical formula is $C_{17}H_{18}F_3NO \cdot HCl$ and it has a molecular weight of 345.79. Its chemical designation is (±)*N*-methyl-3-phenyl-3[(a,a,a-trifluoro-*p*-toylyl)-oxy]-propylamine hydrochloride and its structure is shown in Figure 31.25c-1.

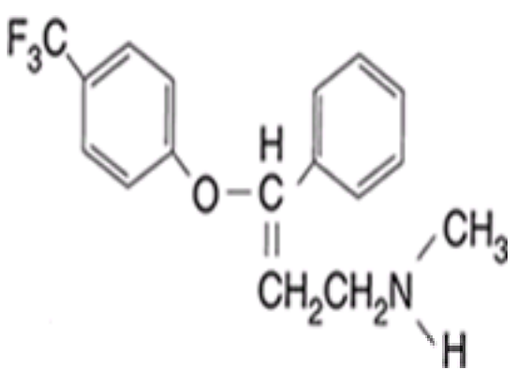

FIGURE 31.25c-1 Molecular structure of fluoxetine.

PHARMACOLOGICAL ACTIONS

Absorption Fluoxetine is well absorbed orally, with 72 to 90 percent systemic availability reflecting the combination of almost complete absorption and first-pass metabolism by liver hepatocytes. Maximum plasma concentration is achieved within 6 to 8 hours of ingestion following a 40-mg dose and is not affected by food. Fluoxetine is available as either pulvules (10 or 20 mg) or oral concentrate. Both forms have equal rates of absorption.

Distribution In vitro, fluoxetine is approximately 95 percent bound to serum proteins (albumin and a_1-acid-glycoproteins) at concentrations of 200 to 1000 ng/mL. The volume of distribution for fluoxetine and norfluoxetine, the biologically active metabolite, range from 20 to 45 L per kg. The rate of plasma clearance is 20 and 9 liters per hour for fluoxetine and norfluoxetine, respectively. Fluoxetine is widely distributed, including excretion into breast milk.

Metabolism and Elimination Fluoxetine is a racemic mixture of *R*-fluoxetine and *S*-fluoxetine in equal proportions. Both are approximately equipotent in serotonin reuptake inhibition activity, though the *S*-fluoxetine enantiomer is more slowly eliminated and is therefore the predominant form in plasma at steady state. Fluoxetine undergoes demethylation in the liver to norfluoxetine and a number of other unidentified minor metabolites of fluoxetine, though norfluoxetine is the only metabolite identified that possesses biological activity. In animals, the *S*-enantiomer of norfluoxetine is essentially equivalent in activity to *R*- or *S*-fluoxetine, but *R*-norfluoxetine has significantly less activity. Inactive metabolites produced by hepatic metabolism are excreted by the kidneys.

The elimination half-life of fluoxetine is 1 to 3 days after short-term administration and 4 to 6 days following long-term administration. Norfluoxetine has an elimination half-life of 4 to 16 days, independent of the duration of administration. Steady-state levels of fluoxetine are higher than would be predicted on the basis of single-dose administration. This is likely due to saturation of the cytochrome P450 (CYP) 2DG isoenzyme and subsequent metabolism by alternate nonsaturable pathways. Steady state is achieved because of these nonsaturable pathways. Norfluoxetine in contrast, demonstrates a linear dose-proportion relationship. The long half-lives of both fluoxetine and norfluoxetine indicate the possibility of drug-drug pharmacokinetic interactions for a sustained period of time after treatment with the drug is discontinued.

Impaired hepatic function is associated with impaired metabolism of fluoxetine. In a series of subjects with cirrhosis of the liver, the mean half-life of fluoxetine was increased to 7.6 days, compared with 2 to 3 days for normal subjects. The half-life of norfluoxetine was similarly increased to a mean of 12 days, compared with 7 to 9 days for normal subjects. This suggests the necessity of dosage reduction in individuals with significant liver disease.

Renal disease does not appear to impair the metabolism of either fluoxetine or norfluoxetine in single-dose studies. However, data on repeated dose administration as occurs in clinical practice are not available and a reduction in dosing size or frequency may be indicated for patients with renal impairment.

PHARMACODYNAMICS

Mechanism of Action The therapeutic activity of fluoxetine is believed to be due to specific serotonin reuptake inhibition at the presynaptic serotonergic nerve terminal. Using an in vitro synaptosomal assay system, fluoxetine is 23 times more potent in blocking serotonin uptake than in blocking norepinephrine uptake, with K_{is} of 12 and 278 nmoles, respectively. In contrast to the tricyclic drugs, fluoxetine essentially lacks any significant affinity at the muscarinic cholinergic, type 1 (H_1), histamine, a_1-adrenergic and serotonin (5-hydroxytryptamine [5-HT]) type 1 ($5\text{-}HT_1$) or $5HT_2$ subtypes. This translates into a significant reduction in adverse effects produced by receptor blockade, compared with the older antidepressants. Fluoxetine, unlike tricyclic drugs, also lacks affinity for cardiac fast sodium channels and thus demonstrates a superior safety profile with regard to cardiac toxicity. Fluoxetine has no effect on monoamine oxidase (MAO) activity.

Blood Concentrations and Relation to Action In contrast to many tricyclic drugs, there appears to be no significant relation between blood concentrations of fluoxetine and subsequent therapeutic response in major depressive disorder. A study of patients with obsessive-compulsive disorder failed to reveal any relation between fluoxetine or norfluoxetine plasma concentrations after 20, 40, or 60 mg a day of fluoxetine and clinical response. However, the assays used failed to differentiate between *S*-norfluoxetine, the more potent 5-HT reuptake inhibitor, and the less active *R*-norfluoxetine.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Major Depressive Disorder Thousands of patients have participated in dosage-ranging studies and active comparator trials of fluoxetine. Most studies have involved outpatients between 18 and 65 years of age, the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) or the revised third edition (DSM-III-R) at least moderate depression severity with major depression meeting criteria of Hamilton Rating Scale for Depression [HAM-D] ³ 20). Few placebo-controlled trials have studied the geriatric and inpatient populations.

The early dose-finding studies compared 20, 40, or 60 mg a day of fluoxetine with placebo. There were 363 randomized subjects, of whom 61 percent were female, and the average baseline HAM-D score was approximately 25 for all groups. All three dosages of fluoxetine showed approximately equal efficacy, with separation from placebo at weeks 4 and 6, as assessed by Clinical Global Improvement Scale. Patients treated for at least 3 weeks exhibited response rates (50 percent or more decrease in HAM-D) of 33, 72, 71, and 64 percent in the placebo, 20, 40, and 60 mg a day groups, respectively. The 60 mg a day dosage had nearly three times the dropout rate due to adverse events (25 vs. 8 percent) as the 20 mg a day group. A second dosage-finding study using lower doses (5, 20, or 40 mg a day) found equal efficacy for all three dosages of fluoxetine, and each was superior to placebo. The 5 mg a day dosage was superior to placebo at weeks 1 to 6 and the 20 or 40 mg a day groups were superior at weeks 3 to 6. These studies suggest that a wide range of dosages is effective in the treatment of moderate or severe depression. The effects of fluoxetine in milder depression (HAM-D scores of 15 to 19) were investigated using doses of 20, 40, or 60 mg daily. Endpoint analysis failed to reveal any difference between fluoxetine and placebo. Completer analysis revealed improvement for the 40 and 60 mg a day groups compared with the placebo group, assessed by clinical global improvement only.

Comparator trials with fluoxetine have included amitriptyline (Elavil), doxepin (Sinequan) imipramine (Tofranil), moclobemide (Aurorix), paroxetine (Paxil), and venlafaxine (Effexor). Doxepin has been compared with fluoxetine in geriatric and nongeriatric patient groups. In the geriatric population, fluoxetine (20 to 80 mg a day) demonstrated efficacy equivalent to that of doxepin (50 to 200 mg a day). Scores on the HAM-D decreased from 25.2 to 16.3 (fluoxetine) and 25.9 to 16.6 (doxepin) in a 6-week trial, a trial length many investigators consider too short to demonstrate maximal efficacy in a geriatric population.

A preliminary report with relevance to this population suggests that in geriatric females, estrogen replacement therapy in conjunction with fluoxetine treatment is more efficacious than fluoxetine alone (HAM-D score improvement of 40.1 and 17 percent, respectively). This retrospective analysis of a relatively small number of study participants did not control for factors that may have led to hormone replacement.

In the nongeriatric study (average age, 41.9 years for the fluoxetine treatment group; 44.4 years for the doxepin treatment group) doxepin and fluoxetine displayed equal efficacy, with baseline HAM-D scores of 26.1 reduced to 16.7 after fluoxetine treatment, and baseline HAM-D scores of 25.5 reduced to 13.4 after doxepin treatment. In general, the comparison trials with amitriptyline and imipramine show efficacies equal to that of fluoxetine, but fluoxetine clearly possesses a superior side-effect profile. Fluoxetine and paroxetine have similar efficacy in outpatients with major depressive disorder of at least moderate severity. Fluoxetine (20 to 40 mg a day) and moclobemide (300 to 600 mg a day), the reversible inhibitor of monoamine oxidase (MAO) type A (RIMA), demonstrated equal efficacy in 122 outpatients with major depressive disorder.

Too few inpatient comparator studies exist to permit firm conclusions. Fluoxetine was compared with imipramine with median dosages of 80 mg a day (fluoxetine) and 200 mg a day (imipramine). There was a nonsignificant trend ($P = .063$) toward greater improvement with imipramine. When fluoxetine was compared with moclobemide in inpatients, both agents demonstrated equal efficacy. One multicenter inpatient trial in France compared fluoxetine and venlafaxine; venlafaxine showed greater efficacy at weeks 4 and 6. Unfortunately, none of the comparator inpatient trials to date have been placebo controlled.

Limited data exist to provide general guidance on how to proceed after nonresponse to fluoxetine. When should a practitioner change the treatment (e.g., by increasing the dose)? Another issue is how to deal with the loss of an initial therapeutic response. In an open-label study, 108 patients with major depressive disorder were treated with fluoxetine (20 mg a day). Patients who failed to respond after 3 weeks were randomized to continue treatment with the same dosage or at 60 mg a day for an additional 5 weeks. At the end of 8 weeks of treatment, both groups demonstrated improvement, with no statistically significant differences between the two groups that were randomized or the group that showed an initial response in the first 3 weeks of treatment. Another study was designed to address the issue of predicting outcome after 8 weeks of fluoxetine treatment, based on responses to the drug at 2, 4, or 6 weeks. In this open-label study, 143 outpatients with DSM-III-R major depression received a fixed dosage of 20 mg a day of fluoxetine. The proportion of patients who failed to show a 20 percent decrease in their HAM-D scores by weeks 2, 4, or 6 but went on to have a 50 percent or greater decrease at week 8 were calculated by survival analysis. The proportion of those who did not demonstrate a 20 percent improvement by weeks 2, 4, or 6 but did respond by week 8 was 36.4, 18.9, and 6.5 percent, respectively, suggesting that there is less than a 20 percent chance of responding to fluoxetine in those who have not responded at all by 4 weeks of treatment. In yet another study, comparing fluoxetine and moclobemide, patients (both inpatients and outpatients) showed equal response rates after 6 weeks of treatment (fluoxetine, 20 to 40 mg a day or moclobemide, 300 to 600 mg a day). Those who failed to respond to the lower dosage by 3 weeks of either drug had their dosage doubled. The higher dosage of moclobemide (600 mg a day) was nearly twice as effective as the higher dosage of fluoxetine (40 mg a day) as assessed by the Clinical Global Impression Scale. The issue of what to do if an initial response is lost has also been examined. Eleven of 18 patients who relapsed on fluoxetine (20 mg a day) (61 percent) exhibited a salutary response during follow-up after the dosage was increased to 40 mg a day.

A common dilemma for the clinician is deciding what course of action to take if a first adequate antidepressant drug trial fails. Unfortunately, limited data address the question of whether switching to a second SSRI after failing a first is indicated. In an open-label, non-placebo-controlled trial of treatment failures or nonresponders to sertraline (Zoloft) who were then switched to fluoxetine, 40 to 70 percent responded. In the absence of placebo-controlled trials, it is difficult to determine how switching within a class compares with the efficacy of switching to a different class.

The efficacy of fluoxetine in the treatment of major depression in children and adolescents (ages 7 to 17 years) was evaluated in a double-blind, placebo-controlled trial. Using the intent-to-treat sample, 56 percent of the patients treated with fluoxetine and 33 percent of those receiving placebo were rated as “very much” or “much” improved. Complete symptom remission was observed in 31 and 23 percent of the fluoxetine and placebo-treated patients, respectively.

Bipolar Disorders Fluoxetine, like other antidepressants, can induce mania in patients with bipolar disorders. Induction of hypomania or mania was reported at a rate of approximately 1 percent in the premarketing major depressive disorder and obsessive-compulsive disorder trials, a rate comparable to that with other antidepressants. A treatment trial comparing 6 weeks of fluoxetine, imipramine, or placebo in depressed patients with bipolar disorder showed improvement in 86, 57, and 38 percent of those receiving fluoxetine, imipramine, and placebo, respectively. During the 6-week trial and extension phase, none of the subjects receiving fluoxetine switched into mania. In this same trial, 4 of 25 patients who were treated with fluoxetine after failing to respond to imipramine or placebo treatment experienced mania.

Comparison of antidepressant-induced mania and spontaneous mania suggested that the latter is more severe. Another group, analyzing treatment-emergent switches into mania by bipolar patients found that SSRIs were associated with a 3.7 percent rate, compared with 11.2 percent for tricyclic drugs and 4.2 percent for placebo.

Dysthymic Disorder The data on treatment of dysthymia with fluoxetine are relatively sparse. In a placebo-controlled trial, 35 subjects with dysthymic disorder (primary dysthymia without a current major depression episode) exhibited response rates of 62.5 percent (10 of 16) in the fluoxetine group and 19 percent (3 of 16) in the placebo group. Response was defined as a 50 percent or greater decrease in the HAM-D scale and a Clinical Global Improvement Scale improvement of 1 or 2. No difference, however, was seen on the Hopkins Symptom Checklist or Cornell Dysthymia Rating scale. In this study, the mean patient age was 36.2 years, and 62.5 percent reported the onset of depressive symptoms in childhood. A similar 60 percent response rate was found in an open-label trial of fluoxetine in elderly patients with dysthymic disorder in an 11-week trial of active medication following a 2-week placebo lead-in period. Data on the effectiveness of fluoxetine after long-term treatment in dysthymia are needed.

Seasonal Affective Disorder The efficacy of fluoxetine in winter seasonal affective disorder (mood disorder with seasonal pattern) was evaluated in 68 outpatients with DSM-III-R recurrent major depression, seasonal (winter) type. Participants were randomly assigned to fluoxetine (20 mg a day) or placebo for 5 weeks. Both

groups exhibited clinical improvement, though the fluoxetine group had a higher rate of clinical response (defined as a 50 percent or greater reduction in the 29-item HAM-D scale between initiation and termination), 59 percent compared with 34 percent for placebo. Those who did not demonstrate some improvement by weeks 2, 4, or 6 but did respond by week 8 were 36.4, 18.9, and 6.5 percent of the patient population, respectively, suggesting that similar to observations with major depression, additional fluoxetine treatment after 8 weeks does not lead to subsequent recovery.

Trichotillomania Two placebo-controlled, crossover studies have failed to demonstrate efficacy of fluoxetine over placebo. The longer was a 31-week trial with an initial 2-week placebo washout period, followed by a 12-week treatment period with either fluoxetine or placebo, a 5-week washout with no placebo capsules, followed by a crossover for 12 additional weeks with the other treatment.

Obessive-Compulsive Disorder Fluoxetine was the first SSRI approved for the treatment of obsessive-compulsive disorder in the United States. Two multicenter, placebo-controlled trials in 355 outpatients between the ages of 15 and 70 years with DSM-III-R obsessive-compulsive disorder demonstrated the efficacy of fluoxetine. A 1-week placebo lead-in period was followed by randomization to placebo or 20, 40, or 60 mg of fluoxetine daily; patients in the 60 mg a day group received 40 mg a day for the first week. All three dosages of fluoxetine produced significant improvement over placebo by week 5. Only the 40 and 60 mg a day dosages demonstrated significant improvement over placebo at all weeks from 5 onward. The 20 mg a day group was superior to placebo at weeks 5, 11, and 13. Improvement in the Yale Brown Obsessive Compulsive Scale (Y-BOCS) from baseline to week 13 was –0.7, –4.7, –5.4, and –7.6 for the placebo, 20, 40, and 60 mg a day groups, respectively. There was a trend toward greater improvement with the higher dosages, though there was also a higher incidence of adverse events. In the authors' experience, patients with obsessive-compulsive disorder tend to have a higher tolerance for adverse effects than those with affective or other anxiety disorders. Completion rates for subjects who completed week 1 were 86, 87, 78, and 76 percent for the placebo, 20, 40, and 60 mg a day groups, respectively. All three dosages were effective, and a reasonable approach may be to initiate therapy at 20 mg a day and increase the dosage as needed.

The rate of therapeutic response to fluoxetine and other SSRIs by patients with obsessive-compulsive disorder is considerably slower than that by patients with depression. This issue was addressed by a continuation study of fluoxetine after the initial 13 weeks of double-blind treatment. Treatment responders continued double-blind treatment, and nonresponders had their dosage increased in an open-label fashion. Those in the blinded condition (receiving a dose of fluoxetine or placebo equal to that in the initial 13 weeks) continued to improve. Nearly two-thirds of open-label participants demonstrated a clinical response in the additional 24 weeks of treatment.

Fluoxetine has been compared with the tricyclic drug clomipramine (Anafranil), also approved for the treatment of obsessive-compulsive disorder. In a double-blind, randomized trial, fluoxetine (40 mg a day) and clomipramine (150 mg a day) demonstrated comparable efficacy in 55 patients with DSM-III-R obsessive-compulsive disorder.

Bulimia Nervosa Daily dosages of 60 mg (but not 20 mg) of fluoxetine, have been shown effective in the treatment of women with bulimia nervosa. An 8-week double-blind, placebo-controlled trial in 387 women with bulimia nervosa aged 18 years or older demonstrated that fluoxetine (60 mg a day) is superior to placebo in decreasing the frequency of binge-eating and vomiting episodes. Fluoxetine treatment also decreased depression, carbohydrate craving, and pathological eating attitudes and behaviors. Baseline HAM-D scores were not associated with outcome, suggesting that improvement was not simply an epiphenomenon of treating depression. Although the response rates (³50 percent improvement) to 60 mg a day were 63 and 57 percent for binge-eating episodes per week and vomiting episodes per week, respectively, less than 25 percent of subjects exhibited complete remission.

Panic Disorder Published studies examining the efficacy of fluoxetine in the treatment of panic disorder suggest that fluoxetine, like other SSRIs, is effective in the treatment of panic disorder. Panic disorder often requires smaller initial dosages than those used in depression, and many of the reports used starting dosages of 5 mg a day. This can be accomplished by using the liquid preparation (20 mg/5 mL) or by dissolving the contents of the pulvules in juice and having the patient drink an aliquot of the juice each day. Individuals with panic disorder treated with fluoxetine at a dosage of 20 mg a day often report worsening of panic attacks initially and exhibit a correspondingly high discontinuation rate.

Social Phobia Fluoxetine has been suggested to be effective in treating social phobia in a series of letters and one published open-label trial. No controlled trial data have been published.

Substance Dependence Fluoxetine was found effective in preventing relapse in male veterans with severe alcohol dependence in a small placebo-controlled trial. In another study investigating the use of fluoxetine in crack cocaine dependence, 32 subjects were randomized to either fluoxetine (40 mg a day) or placebo in addition to standard outpatient treatment. Subjects in the fluoxetine group stayed in treatment longer (median 11 weeks versus 3 weeks for placebo), but there was no difference in cocaine use or craving between the two groups in weeks 1 to 6. Because of poor retention of subjects in the placebo group, no longer-term comparisons could be made.

Premature Ejaculation Sexual dysfunction is a common adverse effect of medications that inhibit serotonin reuptake into the presynaptic neuron. This potential adverse effect has been put to therapeutic use in the treatment of males with premature ejaculation. In an 8-week open-label study, 11 male patients were treated with fluoxetine (20 to 60 mg a day) following a 2-week placebo washout. A within-subjects repeated measures design demonstrated increased intravaginal ejaculation latency as well as improved self-reported and partner's reported satisfaction. A larger double-blind, placebo-controlled trial comparing fluoxetine (20 to 40 mg a day) with placebo in 17 males with premature ejaculation demonstrated greater improvement (assessed by latency to intravaginal ejaculation) in the fluoxetine treated group.

Chronic Fatigue Syndrome A study of fluoxetine treatment of chronic fatigue syndrome with 96 patients randomized to placebo or fluoxetine (20 mg a day) and stratified by presence or absence of depression showed no between-group differences after 8 weeks of treatment.

Fibromyalgia A small-sample (*N* = 42), double-blind, placebo-controlled study that compared fluoxetine (20 mg a day) with placebo in the treatment of women with fibromyalgia found the SSRI to have no efficacy after 6 weeks of treatment.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Respiratory System No adverse respiratory effects noted with fluoxetine have differed significantly from those seen with placebo in premarketing trials. Occasionally, allergic reactions may involve the respiratory system, and dyspnea is a prominent symptom. Anaphylactoid events have also (rarely) been reported after fluoxetine treatment.

Heart and Blood Pressure Fluoxetine was not found to have any significant effects on either systolic or diastolic blood pressure. There may be a minor decrease in heart rate, with an average slowing of about 3 beats per minute observed in premarketing trials. In depressed patients without cardiac disease, fluoxetine elicited no changes in cardiac conduction in dosages up to 80 mg a day, in comparison to amitriptyline. Another study compared fluoxetine with doxepin; fluoxetine produced no adverse electrocardiographic effects. In contrast, doxepin increased the heart rate an average of 12 beats per minute and prolonged the corrected QT interval from 417 ± 36 to 439 ± 28 msec. These studies were all conducted in individuals without cardiac disease.

In one small study (*N* = 16) fluoxetine was used to treat resistant vasovagally mediated episodes of hypotension and bradycardia in individuals who are tilt-table positive. Of the 13 patients who tolerated fluoxetine, 7 (53 percent) had a negative repeat tilt-table test result and were asymptomatic for 19 ± 9 months of follow-up.

Blood Fluoxetine has rarely been reported to produce bleeding and bruising in some individuals. SSRIs as a class are believed to inhibit platelet aggregation, which may underlie this effect. Prospective small-sample studies of a number of hematological parameters following administration of fluoxetine have failed to show any differences. However, because individuals with fluoxetine-induced bruising or bleeding were not included in these studies, it is difficult to know how to interpret the results. The ability of SSRIs to reduce platelet aggregation may be an important intervention for patients with occlusive coronary and cerebrovascular artery disease and deserves study.

Gastrointestinal System Fluoxetine, like other SSRIs, can produce nausea. This is usually dose related and, in most patients, subsides with time. Initial administration of fluoxetine with meals, or in a sensitive individual, at a lower starting dosage often reduces the incidence. Diarrhea and anorexia are also experienced more often with fluoxetine than with placebo. Occasionally weight loss is seen with fluoxetine, and a minority of patients will report weight gain.

Skin Approximately 4 percent of more than 5600 subjects in premarketing trials developed either a rash or urticaria. Nearly a third of these individuals were withdrawn

from treatment because of the rash, urticaria, or both. Excessive sweating has been reported with fluoxetine at a rate about double that seen with placebo.

Central Nervous System (CNS) Potentially important CNS adverse effects seen with administration of fluoxetine include nervousness, insomina, drowsiness, anxiety, tremor, dizziness, and fatigue. Less than 7 percent of patients in premarketing clinical trials in the United States discontinued treatment because of CNS adverse effects of nervousness, anxiety, insomnia, or dizziness. Abnormal or more vivid dreams are occasionally reported by patients taking fluoxetine, as with the other SSRIs.

In the early 1990s a debate arose about whether fluoxetine treatment was associated with increased suicidal ideation and suicide attempts in some patients. A comparison of the emergence of suicidal ideation in patients receiving fluoxetine, tricyclic drugs, or placebo suggested that fluoxetine-treated patients had less emergent suicidal ideation than those treated with tricyclic drugs. Both antidepressant groups showed less emergent suicidal ideation than the group treated with placebo. Fluoxetine produces symptoms resembling akathisia in some patients, often described as restlessness by patients.

THERAPEUTIC INDICATIONS

Fluoxetine currently has FDA approval to be marketed for the treatment of major depressive disorder, obsessive-compulsive disorder, and bulimia nervosa. A variety of clinical experiences ranging from placebo-controlled, double-blind trials to case reports points to the efficacy of fluoxetine in dysthymic disorder, atypical depression, bipolar disorder depression, panic disorder, social phobia, and other anxiety disorders as well as obsessive-compulsive spectrum disorders.

PRECAUTIONS AND ADVERSE REACTIONS

The most prevalent CNS effects include headache, nervousness, insomnia, drowsiness, anxiety, and tremor. Akathisia, though qualitatively different from that observed with antipsychotic drugs, has occasionally been reported. In premarketing testing, the seizure incidence was 0.2 percent, which is similar to the rate seen with other antidepressants.

Common gastrointestinal system adverse effects include nausea, diarrhea, dry mouth, anorexia, and dyspepsia. In clinical practice, less nausea is often seen than was reported in the premarketing trials, and it can be minimized by coadministering the drug with food and reducing the initial dosage in sensitive individuals. Tolerance to nausea usually develops over time.

Sexual dysfunction was reported to occur at a low rate in the premarketing trials, but this was based on spontaneous patient reports. The actual incidence in clinical practice is much higher and consists of delayed ejaculation or impotence in men, delayed orgasm or anorgasmia in women, and decreased libido in both sexes. A number of strategies have been suggested to deal with this effect including decreased dosage, drug holidays for shorter-half-life SSRIs, or addition of buspirone (BuSpar), yohimbine (Yocon), amantadine (Symmetrel), cyproheptadine (Periactin), or bupropion (Wellbutrin). No controlled studies of the effects of any of these strategies are available.

DRUG INTERACTIONS

Fluoxetine may interact pharmacodynamically or pharmacokinetically with other medications. The potential for the serotonin syndrome characterized by autonomic instability, abdominal pain, diarrhea, confusion or delirium, increased motor activity, and myoclonus, or in severe cases, the potential for hyperpyrexia, cardiovascular shock, or death, is perhaps greatest when fluoxetine is combined with MAO inhibitors (MAOIs). The prudent washout period between starting fluoxetine treatment after an MAOI is at least 2 weeks. Five weeks is the minimum recommended interval between stopping fluoxetine treatment and administering an MAOI, because of the long half-life of norfluoxetine. l-Tryptophan in combination with SSRIs also can produce the serotonin syndrome.

Potential pharmacokinetic interactions involve protein binding (poorly characterized as to clinical significance) and the cytochrome P450 system. The manufacturer warns that fluoxetine may displace other tightly protein-bound drugs such as warfarin (Coumadin) or digitoxin (Crystodigin). Good clinical care calls for careful monitoring of potential increased concentrations of agents when tightly protein-bound drugs are coadministered, especially if one agent has a low therapeutic index.

In the cytochrome P450 isoenzyme system, fluoxetine is considered to be a relatively potent inhibitor of the CYP 2C and 2D6 isoenzymes, with no appreciable activity in the CYP 3A4 system, though norfluoxetine does demonstrate some inhibition of CYP 3A4 in vitro. Table 31.25a-4 contains a partial listing of cytochrome isoenzyme substrates.

Fluoxetine may produce hypoglycemia in patients with diabetes who are treated with insulin or oral hypoglycemic drugs. Dosage reduction of the hypoglycemic drug or insulin may be required.

LABORATORY INTERFERENCES

Fluoxetine does not appear to interfere with any laboratory testing.

DOSING AND ADMINISTRATION

Fluoxetine is available in 10- or 20-mg pulvules. The pulvules also contain F D & C Blue No. 1, gelatin, iron oxide, silicon, starch, titanium dioxide, and other inactive ingredients. A liquid solution with a concentration of 20 mg/5 mL is also available.

The suggested starting dosage of fluoxetine in patients with major depression is 20 mg a day. Effective dosages for the treatment of major depression have been reported to range from 5 to 80 mg a day in dosage-finding trials. Because of its long half-life, achieving a new steady state after dosage adjustment takes longer than with most other antidepressants. Decreasing the dosage to treat emerging adverse effects, particularly several weeks after initiation of treatment or the last dosage adjustment, is a reasonable approach.

Preclinical trials showed fluoxetine (60 mg a day) to have greater efficacy in bulimia nervosa than placebo. Treatment of obsessive-compulsive disorder also commonly uses higher dosages than those used in major depressive disorders with dosages of 20 to 80 mg a day most common. The long half-lives of fluoxetine and norfluoxetine make interpretations of the least effective dosage somewhat complicated. Fortunately, because fluoxetine is generally well tolerated, this is less of an issue than it may be with older antidepressants such as tricyclic drugs.

Treatment of panic disorder is often initiated with less than 5 to 10 mg a day, with a gradual upward dosage titration. Treatment of social phobia or premenstrual syndrome is often started at 10 to 20 mg a day. Dosage is adjusted upward as needed.

Starting dosages of 10 mg a day are often used in treating children, adolescents, or elderly patients. Adjustment upward is based on clinical response and tolerance of adverse effects.

SUGGESTED CROSS-REFERENCES

Depression is discussed in Chapter 14, anxiety disorders in Chapter 15, and eating disorders in Chapter 20. Behavior therapy is discussed in Section 30.2, and tricyclic drugs in Section 31.30.

SECTION REFERENCES

*Anonymous: Fluoxetine in the treatment of bulimia nervosa. A multicenter, placebo-controlled, double-blind trial. Fluoxetine Bulimia Nervosa Collaborative Study Group. Arch Gen Psychiatry *49:*139, 1992.

*Beasley CM Jr, Dornseif BE, Bosomworth JC, Sayler ME, Rampey AH Jr, Heiligenstein JH, Thompson VL, Murphy DJ, Masica DN: Fluoxetine and suicide: A meta-analysis of controlled trials of

treatment for depression Br Med J *303:*685, 1991 [published erratum appears in Br Med J *303:*968, 1991].

Benfield P, Heel RC, Lewis SP: Fluoxetine. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic efficacy in depressive illness [Review]. Drugs *32:*481, 1986.

Berman RM, Darnell AM, Miller HL, Anand A, Charney DS: Effect of pindolol in hastening response to fluoxetine in the treatment of major depression: A double-blind, placebo-controlled trial. Am J Psychiatry *154:*37, 1997.

Bowden CL, Schatzberg AF, Rosenbaum A, Contreras SA, Dessain E, Sayler M: Fluoxetine and desipramine in major depressive disorder. J Clin Psychopharmacol *13:*305, 1993.

Clerc GE, Ruimy P, Verdeau-Pailles J, on behalf of the French Inpatient Study Group: A double-blind comparison of venlafaxine and fluoxetine in patients hospitalized for major depression and melancholia. Int Clin Psychopharmacol *9:*139, 1994.

Cohn JB, Collins G, Ashbrook E, Wernicke JF: A comparison of fluoxetine imipramine and placebo in patients with bipolar depressive disorder. Int Clin Psychopharmacol *4:*313, 1989.

Diaz-Martinez A, Benassinni O, Ontiveros A, Gonzalez S, Salin R, Basquedano G, Martinez RA: A randomized, open-label comparison of venlafaxine and fluoxetine in depressed outpatients. Clin Ther *20:*467, 1998.

Dornseif BE, Dunlop SR, Potvin JH, Wernicke JF: Effect of dose escalation after low-dose fluoxetine therapy. Psychopharmacol Bull *25:*71, 1989.

*Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T: A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. Arch Gen Psychiatry *54:*1031, 1997.

Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Carmody T, Mayes TL: Fluoxetine in child and adolescent depression: Acute and maintenance treatment. Depression Anxiety *7:*32, 1998.

Fava M, Rosenbaum JF, McGrath PJ, Stewart JW, Amsterdam JD, Quitkin FM: Lithium and tricyclic augmentation of fluoxetine treatment for resistant major depression: A double-blind, controlled study. Am J Psychiatry *151:*1372, 1994.

Goldenberg D, Mayskiy M, Mossey C, Ruthazer R, Schmid C: A randomized, double-blind crossover trial of fluoxetine and amitriptyline in the treatment of fibromyalgia. Arthritis Rheum *39:*1852, 1996.

*Goldstein DJ, Wilson MG, Thompson VL, Potvin JH, Rampey AH Jr: Long-term fluoxetine treatment of bulimia nervosa. Fluoxetine Bulimia Nervosa Research Group. Br J Psychiatry *166:*660, 1995.

Lam RW, Gorman CP, Michalon M, Steiner M, Levitt AJ, Corral MR, Watson GD, Morehouse RL, Tam W, Joffe RT: Multicenter, placebo-controlled study of fluoxetine in seasonal affective disorder. Am J Psychiatry *152:*1765, 1995.

*Leon AC, Keller MB, Warshaw MG, Mueller TI, Solomon DA, Coryell W, Endicott J: Prospective study of fluoxetine treatment and suicidal behavior in affectively ill subjects. Am J Psychiatry *156:*195, 1999.

Lopez-Ibor JJ Jr, Saiz J, Cottraux J, Note I, Vinas R, Bourgeois M, Hernandez M, Gomez-Perez JC: Double-blind comparison of fluoxetine versus clomipramine in the treatment of obsessive compulsive disorder. Eur Neuropsychopharmacol *6:*111, 1996.

Michelson D, Lydiard RB, Pollack MH, Tamura RN, Hoog SL, Tepner R, Demitrack MA, Tollefson GD: Outcome assessment and clinical improvement in panic disorder: Evidence from a randomized controlled trial of fluoxetine and placebo. Am J Psychiatry *155:*1570, 1998.

*Montgomery SA, McIntyre A, Osterheider M, Sarteschi P, Zitterl W, Zohar J, Birkett M, Wood AJ: A double-blind, placebo-controlled study of fluoxetine in patients with DSM-III-R obsessive-compulsive disorder. The Lilly European OCD Study Group. Eur Neuropsychopharmacol *3:*143, 1993.

Nierenberg AA, McLean NE, Alpert JE, Worthington JJ, Rosenbaum JF, Fava M: Early nonresponse to fluoxetine as a predictor of poor 8-week outcome [see comments]. Am J Psychiatry *152:*1500, 1995.

Pearlstein TB, Stone AB: Long-term fluoxetine treatment of late luteal phase dysphoric disorder. J Clin Psychiatry *55:*332, 1994.

Rickels K, Smith WT, Glaudin V, Amsterdam JB, Weise C, Settle GP: Comparison of two dosage regimens of fluoxetine in major depression. J Clin Psychiatry *46:*38, 1985.

Riddle MA, Scahill L, King RA, Hardin MT, Anderson GM, Ort SI, Smith JC, Leckman JF, Cohen DJ: Double-blind, crossover trial of fluoxetine and placebo in children and adolescents with obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry *31:*1062, 1992.

Roose SP, Glassman AH, Attia E, Woodring S, Giardina EG, Bigger JT Jr: Cardiovascular effects of fluoxetine in depressed patients with heart disease. Am J Psychiatry *155:*660, 1998.

Schneider LS, Small GW, Hamilton SH, Bystritsky A, Nemeroff CB, Meyers BS: Estrogen replacement and response to fluoxetine in a multicenter geriatric depression trial. Fluoxetine Collaborative Study Group. Am J Geriatr Psychiatry *5:*97, 1997.

Schweizer E, Rickels K, Amsterdam JD, Fox I, Puzzuoli G, Weise C: What constitutes an adequate antidepressant trial for fluoxetine? J Clin Psychiatry *51:*8, 1990.

Steiner M, Steinberg S, Stewart D, Carter D, Berger C, Reid R, Grover D, Streiner D: Fluoxetine in the treatment of premenstrual dysphoria. Canadian Fluoxetine/Premenstrual Dysphoria Collaborative Study Group [see comments]. N Engl J Med *332:*1529, 1995.

Streichenwein SM, Thornby JI: A long-term, double-blind, placebo-controlled crossover trial of the efficacy of fluoxetine for trichotillomania. Am J Psychiatry *152:*1192, 1995.

Teicher MH, Glod C, Cole JO: Emergence of intense suicidal preoccupation during fluoxetine treatment. Am J Psychiatry *147:*207, 1990.

Thase ME, Blomgren SL, Birkett MA, Apter JT, Tepner RG: Fluoxetine treatment of patients with major depressive disorder who failed initial treatment with sertraline. J Clin Psychiatry *58:*16, 1997.

Tollefson GD, Bosomworth JC, Heiligenstein JH, Potvin JH, Holman S: A double-blind, placebo-controlled clinical trial of fluoxetine in geriatric patients with major depression. The Fluoxetine Collaborative Study Group. International Psychogeriatrics *7:*89, 1995.

Tollefson GD, Rampey AH Jr, Potvin JH, Jenike MA, Rush AJ, Kominguez RA, Koran LM, Shear MK, Goodman W, Genduso LA: A multicenter investigation of fixed- dose fluoxetine in the treatment of obsessive-compulsive disorder [published erratum appears in Arch Gen Psychiatry *51:*864, 1994]. Arch Gen Psychiatry *51:*559, 1994.

Vercoulen JH, Swanink CM, Zitman FG, Vreden SG, Hoofs MP, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G: Randomised, double-blind, placebo-controlled study of fluoxetine in chronic fatigue syndrome. Lancet *347:*858, 1996.

Wood SH, Mortola JF, Chan YF, Moossazadeh F, Yen SS: Treatment of premenstrual syndrome with fluoxetine: A double-blind, placebo-controlled, crossover study. Obstet Gynecol *80:*339, 1992.

CHAPTER 31. BIOLOGICAL THERAPIES

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25D FLUVOXAMINE

JEFFREY E. KELSEY, M.D., FH.D. AND CHARLES B. NEMEROFF, M.D., FH.D.

HISTORY

Fluvoxamine is currently approved in the United States for the treatment of obsessive-compulsive disorder but has been used for many years in other countries to treat depression. The efficacy of fluvoxamine is believed to be due to the reuptake inhibition of serotonin into the presynaptic serotonergic nerve terminal.

CHEMISTRY

Fluvoxamine, a structurally unique antidepressant, is a member of the 2-aminoethyloxime ethers of the aralkylketone family. The empirical formula of fluvoxamine is $C_{15}H_{21}O_2N_2F_3 \cdot C_4H_4O_4$, and it has a molecular weight of 434.4. The chemical designation is 5-methoxy-4'-(trifluoromethyl) valerophenone-(*E*)-*O*-(2-aminoethyl) oxime maleate (1:1), and the structure is shown in Figure 31.25d-1.

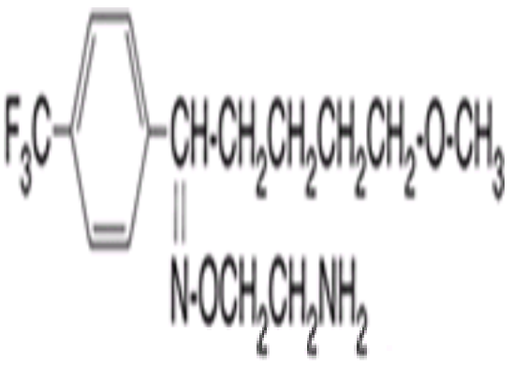

FIGURE 31.25d-1 Molecular structure of fluvoxamine.

PHARMACOLOGICAL ACTIONS

Absorption The bioavailability of fluvoxamine maleate is 53 percent, and its absorption is not significantly affected by food. At steady state, maximal plasma concentrations are achieved 3 to 8 hours after dosing.

Distribution In vitro, fluvoxamine is approximately 80 percent bound to serum proteins, predominantly albumin, at concentrations of 20 to 2000 ng/mL. The volume of distribution for fluvoxamine is approximately 25 L/kg. The extent of excretion of fluvoxamine into breast milk is unclear at this time.

Metabolism and Elimination Fluvoxamine is primarily metabolized by oxidative demethylation and deamination in the liver. The main human metabolite, as determined by radiotracer studies, is fluvoxamine acid, which along with its *N*-acetylated analogue accounted for about 60 percent, of urinary products. In this same assay, fluvoxethanol made up about 10 percent. Fluvoxamine acid is 10 to 100 times less potent in an in vitro assay of serotonin reuptake inhibition.

Fluvoxamine does not demonstrate linear pharmacokinetics. In 30 normal volunteers receiving fluvoxamine at 100, 200, or 300 mg a day for 10 days, the maximum plasma concentrations were 88, 283, and 546 ng/mL, respectively. This represents increases of 160 and 207 percent for the 200- and 300-mg dosages, respectively, over what would have been predicted from the lowest dose. The mean plasma half-life at steady state is 15.6 hours in young healthy volunteers receiving 100 mg a day and 13.6 hours for a 50-mg daily dosage. After repeated dosing, the half-life increases to 17 to 22 hours. Steady state is typically achieved in about 7 days.

Elderly patients show a prolonged half-life of fluvoxamine. Compared with younger patients, in elderly patients the half life at 50 mg a day goes from 13.6 to 17.4 hours and from 15.6 to 25.9 hours for 100 mg a day dosing.

Hepatic dysfunction is associated with a 30 percent decrease in fluvoxamine clearance compared with that in healthy controls. In contrast, renal impairment (creatinine clearance of 5 to 45 mL per minute) was not associated with decreased clearance.

CLINICAL STUDIES

Obsessive-Compulsive Disorder In a multisite, double-blind placebo-controlled trial of fluovoxamine in obsessive-compulsive disorder, 169 patients were randomized to fluvoxamine (100 to 300 mg a day) or placebo. Efficacy evaluations were available for 78 patients in each group, and fluvoxamine was more effective than placebo as measured by the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS), National Institute of Mental Health Obsessive-Compulsive scale, and the Global Improvement item in the Clinical Global Impression scale. From week 6 onward, 33.3 percent of the fluvoxamine group but only 9.0 percent of the placebo group were rated as much or very much improved on the Global Improvement item.

Fluvoxamine (100 to 250 mg a day) was compared with clomipramine (Anafranil) (100 to 250 mg a day) in 66 outpatients aged 18 to 65 with obsessive-compulsive disorder in a multicenter study. Mean reductions in the Y-BOCS of 33 and 31 percent were observed for the fluvoxamine and clomipramine groups, respectively.

A second comparison study with fluvoxamine (100 to 300 mg a day) and clomipramine (100 to 250 mg a day) was conducted in 79 patients with obsessive-compulsive disorder in a 10-week active treatment trial. Study completion rates were 78 percent for fluvoxamine and 64 percent for clomipramine-treated patients. At the conclusion of the treatment trial, the response rate (defined by a 25 percent or greater decrease in the Y-BOCS) was identical between the two groups, 56 and 54 percent for the fluvoxamine and clomipramine groups, respectively.

Fluvoxamine demonstrates efficacy in the treatment of obsessive-compulsive disorder in children and adolescents. A 10-week, parallel-design, multicenter study of children and adolescents, 8 to 17 years of age, compared fluvoxamine with placebo treatment. The fluvoxamine-treated group showed a mean drop of about 6 units on the Children's Y-BOCS, compared with a drop of about 3 points in the placebo-treated group. Post hoc analysis suggested a more robust response in the

8-to-11-years-of-age group than in the 11-to-17-year-old group. The significance of this is not clear, though it may suggest the need for earlier treatment.

Major Depressive Disorder Fluvoxamine has been compared with a number of different antidepressants. In a comparison with imipramine (Tofranil) (up to 240 mg a day) or placebo, fluvoxamine (up to 150 mg a day) was more efficacious than placebo in this study of 150 outpatients with major depressive disorder. Among those with more-severe depression (Hamilton Rating Scale for Depression is ³30), there was a suggestion of better efficacy for fluvoxamine. Fluvoxamine has been compared with fluoxetine in a trial of 100 outpatients with major depressive disorder. Fluvoxamine (100 to 150 mg a day) and fluoxetine (20 to 80 mg a day) produced comparable improvement in the 21-item Hamilton Rating Scale for Depression score of close to 60 percent over the course of 7 weeks of study medication.

Panic Disorder In addition to demonstrating efficacy in panic disorder, fluvoxamine decreased anxiety following intravenous yohimbine (Yocon) as a pharmacological challenge in patients with panic disorder after 8 weeks of treatment with fluvoxamine. The posttreatment challenge with yohimbine produced significantly less anxiety than the initial pretreatment exposure.

Social Phobia Fluvoxamine was studied in a small double-blind, placebo-controlled trial of patients with revised third edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-III-R) social phobia. Fluvoxamine (150 mg a day) produced improvement in 7 of 15 patients, compared with 1 of 15 in the placebo group, following 12 weeks of treatment.

Other Conditions Fluvoxamine has been used in patients with binge-eating disorders and psychogenic excoriation disorders.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Respiratory System Fluvoxamine appears to have no clinically significant effect on respiratory function. However a drug-drug interaction between fluvoxamine and theophylline (Theo-Dur) is of potential importance in treating the patient with asthma or chronic obstructive pulmonary disease. Increased cough and sinusitis were reported to be frequent (incidence of 1 in 100 or greater) adverse effects in premarketing trials.

Cardiovascular Blood pressure, pulse rate, and electrocardiographs showed no significant difference between patients receiving fluvoxamine and those receiving placebo in premarketing trials. Hypertension, hypotension, syncope, and tachycardia were observed at a frequency of 1 in 100 or greater in premarketing trials.

Blood Administration of fluvoxamine does not appear to have any clinically significant effect on hematological indexes or blood chemistries.

Gastrointestinal Fluvoxamine, like the other SSRIs, may produce nausea, especially with initial dosages. The incidence of nausea in premarketing trials was 40 percent in the fluvoxamine-treated group and 14 percent in the placebo-treated group. Diarrhea or constipation may be less frequent with fluvoxamine than with some of the other selective serotonin reuptake inhibitors (SSRIs). Elevated liver transaminases were reported at a frequency of 1 in 100 or greater in premarketing trials.

Skin Increased sweating is seen in a small percentage of patients receiving fluvoxamine.

Central Nervous System The most commonly observed central nervous system reactions occuring with fluvoxamine treatment are somnolence, insomnia, dry mouth, nervousness, dizziness, tremor, and anxiety. Headache occured at a rate of 22 percent in the fluvoxamine-treated group in premarketing trials which is comparable to an incidence of 20 percent in the placebo-treated groups. Mania or hypomania was observed at a rate of 1 percent, and seizures at a rate of 0.2 percent in premarketing studies among patients receiving fluvoxamine. Both rates are comparable to those seen in trials of other SSRIs.

THERAPEUTIC INDICATIONS

Fluvoxamine is currently indicated for use in the treatment of obsessive-compulsive disorder in both children and adults. Fluvoxamine is approved for use in the treatment of major depression in numerous other countries.

PRECAUTIONS AND ADVERSE REACTIONS

The most common adverse effects observed in premarketing trials were nausea, somnolence, headache (though not at a significantly greater rate than in the placebo-treated group), insomnia, dry mouth, asthenia, nervousness, dizziness, and constipation. In general, the side-effect profile of fluvoxamine is comparable to those of other SSRIs. Patients should be warned of potential medication interactions with fluvoxamine and drugs that are metabolized by the cytochrome P450 (CYP) 1A2 and 3A4 isoenzyme system.

DRUG INTERACTIONS

Fluvoxamine is less tightly protein bound than the other SSRI medications and may be less likely to displace tightly protein-bound drugs. In the P450 system, fluvoxamine is an inhibitor of CYP 1A2, CYP 2C9, and CYP 3A4 isoenzymes. Theophylline, aminophylline (Mudrane) warfarin (Coumadin), propranolol (Inderal), and caffeine metabolism is inhibited through CYP 1A2 by fluvoxamine. Warfarin metabolism is also inhibited through the CYP 2C9 system. Benzodiazepines including alprazolam (Xanax), clonazepam (Klonopin), triazolam (Halcion), and midazolam (Versed) are metabolized by hepatic oxidation via the CYP 3A4 isoenzyme. Approximately twofold elevations in serum concentrations of alprazolam have been observed following coadministration of fluvoxamine. Increased serum concentrations of carbamazepine (Tegretol), clozapine (Clozaril), methadone, propranolol, amitriptyline (Elavil), clomipramine, and imipramine have been reported following the addition of fluvoxamine treatment. The manufacturer warns against the coadministration of fluvoxamine with terfenidine (Seldane), astemizole (Hismanal), or cisapride (Propulsid). All three compounds are inactive prodrugs upon ingestion and require CYP 3A4 activity to produce the active metabolite. Inhibition of metabolism produces an accumulation of precursor that may produce cardiac conduction delay and a potentially fatal torsades de pointes type arrhythmia.

In premarketing trials, smokers demonstrated a 25 percent higher fluvoxamine metabolism than nonsmokers, presumably through induction of CYP 1A2.

Fluvoxamine should not be administered within 14 days of discontinuing treatment with a monoamine oxidase inhibitor (MAOI). A 2-week period should elapse after stopping fluvoxamine treatment and initiating treatment with an MAOI. Caution should be used when combining fluvoxamine, like other SSRIs, with other serotonergic drugs such as sumatriptan (Imitrex) or l-tryptophan.

DOSING AND ADMINISTRATION

Fluvoxamine is available in 25-, 50-, and 100-mg scored tablets. Treatment of adults with obsessive-compulsive disorder is often started at a dose of 50 mg administered at bedtime. The dosage may be increased by 50 mg every 4 to 7 days. The effective dosing range in premarketing trials with obsessive compulsive disorder was 100 to 300 mg a day. The manufacturer suggests dividing the daily dosage if it exceeds 100 mg a day, though many patients tolerate doses larger than 100 mg. Initial upward titration often is easier with a split-dose regimen.

Treatment of obsessive-compulsive disorder in adolescents and children should start at 25 mg a day with 25 mg a day increases every 4 to 7 days. The effective dosing range in premarketing studies was 50 to 200 mg a day. The manufacterer recommends administering 50 mg as a total daily dose and no more than 50 mg as a split dose.

Treatment of elderly patients and those with comorbid medical or hepatic disease should be started at a lower dosage than typically used for adults.

Dosing for disorders other than those for which fluvoxamine is approved in the United States is usually initiated at 25 to 50 mg a day. Patients with major depression typically respond over the range of 100 to 300 mg a day, though often in the range of 100 to 200 mg a day. Treatment of panic disorder is best started at a dosage of 25 mg a day for most patients because of increased sensitivity to adverse effects in this population. The typical effective dosage is usually 100 to 200 mg a day.

SUGGESTED CROSS-REFERENCES

Section 31.25b discusses citalopram, Section 31.25c discusses fluoxetine, Section 31.25e discusses paroxetine, and Section 31.25f discusses sertraline. Complete coverage of all aspects of biological therapies can be found in the rest of Chapter 31. The biological treatment of depression is discussed in Section 14.4, the biological treatment of obsessive-compulsive disorder in Section 15.7, and the biological treatment of eating disorders in Chapter 20.

SECTION REFERENCES

*Arnold LM, Mutasim DF, Dwight MM, Lamerson CL, Morris EM, McElroy SL: An open clinical trial of fluvoxamine treatment of psychogenic excoriation. J Clin Psychopharmacol *19:*15, 1999.

Black DW, Monahan P, Gabel J: Fluvoxamine in the treatment of compulsive buying. J Clin Psychiatry *58:*159, 1997.

Claghorn JL, Earl CQ, Walczak DD, Stoner KA, Wong LF, Kanter D, Houser VP: Fluvoxamine maleate in the treatment of depression: A single-center, double-blind, placebo-controlled comparison with imipramine in outpatients. J Clin Psychopharmacol *16:*113, 1996.

Cottraux J, Mollard E, Bouvard M, Marks I, Sluys M, Nury AM, Douge R, Cialdella P: A controlled study of fluvoxamine and exposure in obsessive-compulsive disorder. Int Clin Psychopharmacol *5:*17, 1990.

*de Beurs E, van Balkom AJ, Lange A, Koele P, van Dyck R: Treatment of panic disorder with agoraphobia: Comparison of fluvoxamine, placebo, and psychological panic management combined with exposure and of exposure in vivo alone. Am J Psychiatry *152:*683, 1995.

den Boer JA, Westenberg HG: Effect of a serotonin and noradrenaline uptake inhibitor in panic disorder: A double-blind comparative study with fluvoxamine and maprotiline. Int Clin Psychopharmacol *3:*59, 1988.

Dominguez RA, Goldstein BJ, Jacobson AF, Steinbook RM: A double-blind placebo-controlled study of fluvoxamine and imipramine in depression. J Clin Psychiatry *46:*84, 1985.

Fichter MM, Kruger R, Rief W, Holland R, Dohne J: Fluvoxamine in prevention of relapse in bulimia nervosa: Effects on eating-specific psychopathology. J Clin Psychopharmacol *16:*9, 1996.

Fleishaker JC, Hulst LK: A pharmacokinetic and pharmacodynamic evaluation of the combined administration of alprazolam and fluvoxamine. Eur J Clin Pharmacol *46:*35, 1994.

*Freeman CP, Trimble MR, Deakin JF, Stokes TM, Ashford JJ: Fluvoxamine versus clomipramine in the treatment of obsessive compulsive disorder: A multicenter, randomized, double-blind, parallel group comparison. J Clin Psychiatry *55:*301, 1994.

Gonella G, Baignoli G, Ecari U: Fluvoxamine and imipramine in the treatment of depressive patients: A double-blind controlled study. Curr Med Res Opinion *12:*177, 1990.

Hudson JI, McElroy SL, Raymond NC, Crow S, Keck PE Jr, Carter WP, Mitchell JE, Strakowski SM, Pope HG Jr, Coleman BS, Jonas JM. Fluvoxamine in the treatment of binge-eating disorder: A multicenter placebo-controlled, double-blind trial. Am J Psychiatry *155:*1756, 1998.

Jeppesen U, Loft S, Poulsen HE, Brsen K: A fluvoxamine-caffeine interaction study. Pharmacogenetics *6:*213, 1996.

*Kiev A, Feiger A: A double-blind comparison of fluvoxamine and paroxetine in the treatment of depressed outpatients. J Clin Psychiatry *58:*146, 1997.

Koran LM, McElroy SL, Davidson JR, Rasmussen SA, Hollander E, Jenike MA: Fluvoxamine versus clomipramine for obsessive-compulsive disorder: A double-blind comparison. J Clin Psychopharmacol *16:*121, 1996.

Linnoila M, Stapleton JM, George DT, Lane E, Eckardt MJ: Effects of fluvoxamine, alone and in combination with ethanol, on psychomotor and cognitive performance and on autonomic nervous system reactivity in healthy volunteers. J Clin Psychopharmacol *13:*175, 1993.

McDougle CJ, Naylor ST, Cohen DJ, Volkmar FR, Heninger GR, Price LH: A double-blind, placebo-controlled study of fluvoxamine in adults with autistic disorder [see comments]. Arch Gen Psychiatry *53:*1001, 1996.

Mendlewicz J: Efficacy of fluvoxamine in severe depression. Drugs *43* (Suppl 2):32, 1992.

Mullin JM, Pandita-Gunawardena VR, Whitehead AM: A double-blind comparison of fluvoxamine and dothiepin in the treatment of major affective disorder. Br J Clin Pract *42:*51, 1988.

Mundo E, Bareggi SR, Pirola R, Bellodi L, Smeraldi E: Long-term pharmacotherapy of obsessive-compulsive disorder: A double-blind controlled study. J Clin Psychopharmacol *17:*4, 1997.

Ottevanger EA: Fluvoxamine and clomipramine in depressed hospitalised patients: Results from a randomised, double-blind study. Encephale *21:*317, 1995.

Phanjoo AL, Wonnacott S, Hodgson A: Double-blind comparative multicentre study of fluvoxamine and mianserin in the treatment of major depressive episode in elderly people. Acta Psychiatr Scand *83:*476, 1991.

Phillips KA, Dwight MM, McElroy SL: Efficacy and safety of fluvoxamine in body dysmorphic disorder. J Clin Psychiatry *59:*165, 1998.

Rahman MK, Akhtar MJ, Savla NC, Sharma RR, Kellett JM, Ashford JJ: A double-blind, randomised comparison of fluvoxamine with dothiepin in the treatment of depression in elderly patients. Br J Clin Pract *45:*255, 1991.

Roth D, Mattes J, Sheehan KH, Sheehan DV: A double-blind comparison of fluvoxamine, desipramine and placebo in outpatients with depression. Prog Neuropsychopharmacol Biol Psychiatry *14:*929, 1990.

Silver H, Nassar A: Fluvoxamine improves negative symptoms in treated chronic schizophrenia: An add-on double-blind, placebo-controlled study. Biol Psychiatry *31:*698, 1992.

Stanley MA, Breckenridge JK, Swann AC, Freeman EB, Reich L: Fluvoxamine treatment of trichotillomania. J Clin Psychopharmacol *17:*278, 1997.

*Terra JL, Montgomery SA: Fluvoxamine prevents recurrence of depression: Results of a long-term, double-blind, placebo-controlled study. Int Clin Psychopharmacol *13:*55, 1998.

van Megen HJ, Westenberg HG, den Boer JA, Slaap B, Scheepmakers A: Effect of the selective serotonin reuptake inhibitor fluvoxamine on CCK-4 induced panic attacks. Psychopharmacology *129:*357, 1997.

Wagner W, Plekkenpol B, Gray TE, Vlaskamp H, Essers H: Review of fluvoxamine safety database. Drugs *43*(Suppl):48, 1992.

Wakelin JS: Fluvoxamine in the treatment of the older depressed patient: Double-blind, placebo-controlled data. Int Clin Psychopharmacol *1:*221, 1986.

White K, Wykoff W, Tynes LL, Schneider L, Zemansky M: Fluvoxamine in the treatment of tricyclic-resistant depression. Psychiatr J Univ Ottawa *15:*156, 1990.

Wilde MI, Plosker GL, Benfield P: Fluvoxamine. An updated review of its pharmacology, and therapeutic use in depressive illness. Drugs *46:*895, 1993.

*Zanardi R, Franchini L, Gasperini M, Smeraldi E, Perez J: Long-term treatment of psychotic (delusional) depression with fluvoxamine: An open pilot study. Int Clin Psychopharmacol *12:*195, 1997.

CHAPTER 31. BIOLOGICAL THERAPIES

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25E PAROXETINE

JEFFREY E. KELSEY, M.D., FH.D. AND CHARLES B. NEMEROFF, M.D., FH.D.

HISTORY

Paroxetine (Paxil), the third member of the selective serotonin reuptake inhibitor (SSRI) family approved for use in the United States, became available in 1993. Although initially approved for the treatment of major depression, paroxetine is currently also approved for the treatment of obsessive-compulsive disorder and panic disorder. It is also used for the treatment of social phobia.

The therapeutic efficacy of paroxetine is thought to be due to the reuptake inhibition of serotonin into the presynaptic serotonergic nerve terminal.

CHEMISTRY

Paroxetine is the hydrochloride salt of a phenylpiperidine compound and is structurally distinct from other antidepressants. Paroxetine is designated as (–)-*trans*-4*R*-(4'-flourophenyl)-3*S*-[(3', 4'-methylenedioxyphenoxy) methyl] piperidine hydrochloride hemihydrate. The empirical formula is $C_{19}H_{20}FNO_3 \cdot HCl \cdot 1/2H_2O$, and the molecular weight of the free base is 374.8. The structure of paroxetine is shown in Figure 31.25e-1.

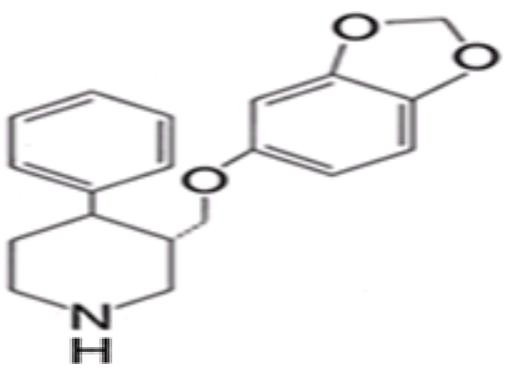

FIGURE 31.25e-1 Molecular structure of paroxetine.

PHARMACOLOGICAL ACTIONS

Absorption Paroxetine is well absorbed orally, and food does not alter absorption. Maximum systemic concentration is achieved approximately 5.2 hours after ingestion.

Distribution Paroxetine, like sertraline (Zoloft) and fluoxetine (Prozac) is tightly bound to serum proteins in vitro, with binding of 95 and 93 percent at plasma concentrations of 100 and 400 ng/mL, respectively. Paroxetine has not been reported to alter the in vitro protein binding of warfarin (Coumadin) or phenytoin (Dilantin). The volume of distribution is about 15 L/kg, and the rate of plasma clearance is close to 79 liters per hour. Most individuals achieve steady state plasma concentrations within 5 to 10 days of dosing.

Metabolism and Elimination The principle path of metabolism for paroxetine is oxidation and methylation. A number of major metabolites have been identified, though none appears to be any more than 1/50th as potent as the parent compound in the in vitro reuptake inhibition of serotonin. A test dose of paroxetine (30 mg) was recovered as 64 percent in the urine (2 percent as parent compound and 62 percent as metabolites over 10 days). In the feces, 36 percent was excreted, with less than 1 percent as the parent compound.

Paroxetine saturates the cytochrome P450 (CYP) system 2D6 isoenzyme and thus exhibits nonlinear pharmacokinetics. Steady-state plasma maximal and minimal concentrations were elevated 6 and 14 times over single-dose values after reaching steady state. This translates into a 24-hour area-under-the-curve value close to eight-fold higher at steady state than would be predicted from a single dose.

In elderly patients, dosing of paroxetine at 20, 30, or 40 mg a day produced plasma concentrations approximately 70 to 80 percent greater than those observed in younger patients. Patients with a creatinine clearance below 30 mL per minute demonstrated a fourfold elevation in plasma concentrations compared with healthy controls. More-moderate renal impairment (creatinine clearance of 30 to 60 mL per minute), as well as patients with moderate hepatic impairment, had a two-fold increase in plasma concentrations. All of the above groups should receive reduced starting dosages of paroxetine (e.g. 10 mg a day).

PHARMACODYNAMICS

Mechanism of Action Paroxetine is believed to exert therapeutic activity by means of serotonin reuptake inhibition within the central nervous system (CNS). The reuptake inhibition of norepinephrine into rat hypothalamic synaptosomal membranes requires a concentration of paroxetine 320 times that necessary for serotonin reuptake inhibition. However, paroxetine is the most potent of the SSRIs in inhibiting norepinephrine, and is in fact more potent than venlafaxine (Effexor), a serotonin-noradrenaline reuptake inhibitor, in this regard. Evidence indicates that at clinically relevant dosages, paroxetine inhibits norepinephrine reuptake. Paroxetine, like other SSRIs also depletes serotonin from blood platelets. In contrast to many tricyclic drugs that downregulate of b-adrenergic receptors in the CNS, paroxetine, like the other SSRIs does not. Terminal serotonin autoreceptors are downregulated following long-term administration of paroxetine, which reflects paroxetine facilitating serotonergic neurotransmission by increasing extracellular fluid serotonin concentrations. This has been posited to play a role in its antidepressant effect and be a critical step in the signal transduction to cellular events that result in altered patterns of gene expression, messenger ribonucleic acid (mRNA) translation, or protein modification.

Paroxetine is essentially devoid of clinically significant affinity for the a_1-, a_2-, or b-adrenergic, dopamine type 2 (D_2), histamine type 1 (H_1), serotonin

(5-hydroxytryptamine [5-HT]) type 1 (5-HT_1) and 5-HT_2 receptors. Paroxetine does have the greatest affinity for muscarinic cholinergic receptors of the available SSRIs, and produces dry mouth, constipation, blurred vision, or urinary hesitancy in a small percentage of patients. However the affinity of paroxetine for muscarinic receptors as measured by inhibition of [^{3}H] quinuclidinylbenzilate binding is 15-fold less potent than that of amitriptyline (Elavil), and its anticholinergic properties may underlie its lack of the diarrhea and loose stools observed with other SSRIs and its occasional sedating properties. Paroxetine is inactive at the cardiac fast sodium channels and has no effect on cardiac conduction. Paroxetine does not inhibit monoamine oxidase activity.

Blood Concentrations and Relation to Action No data suggest the presence of therapeutic concentrations of paroxetine or any correlation between plasma concentrations and subsequent clinical response. Therefore paroxetine plasma concentration determinations are useful clinically only to confirm compliance.

DESIGN AND INTERPRETATION OF CLINICAL TRIALS

Major Depressive Disorder

Clinical Trials Paroxetine has been studied in a large number of subjects with major depression, both in placebo-comparison studies and in those with established antidepressants as the comparison group. Most of these studies have involved healthy outpatients between the ages of 18 and 65. Paroxetine has demonstrated superiority to placebo in several controlled trials in patients with depression. In a pooled analysis of 4 trials involving 273 patients with major depression, paroxetine was superior to placebo in measures of both anxiety and depression. Paroxetine separated from placebo by week 2 and sustained this difference throughout the 6-week trial.

Comparison Trials A number of different trials have compared paroxetine with imipramine (Tofranil). Pooling six studies of similar design in which outpatients with major depression were randomized to treatment with paroxetine (10 to 50 mg a day), imipramine (65 to 275 mg a day), or placebo, both active treatments showed superiority over placebo. Both drugs were equal in efficacy and superior to placebo by week 2, but paroxetine was much better tolerated than imipramine.

Paroxetine has been compared with clomipramine (Anafranil) in a large, international, multicenter study involving 1019 patients with major depression (Montgomery-Asberg Depression Rating Scale [MADRS] score ³20) and associated anxiety (Clinical Anxiety score ³11) randomized to paroxetine 20 to 40 mg a day or clomipramine 75 to 150 mg a day. Only 41 percent of the patients receiving paroxetine and 33 percent of those receiving clomipramine had dosage increases above 20 or 75 mg a day, respectively. After 12 weeks of treatment, assessment with the MADRS and Clinical Anxiety scores demonstrated equal efficacy in the two groups. This contrasts with a Danish study of inpatients with major depression randomized to receive 40 mg a day of paroxetine or 150 mg a day of clomipramine. The clomipramine group demonstrated greater efficacy than the paroxetine group as measured by the Hamilton Rating Scale for Depression (HAM-D). Another comparison study with paroxetine and amitriptyline was performed in 153 hospitalized patients with major depression in Austria and Germany. The average duration of the current episode for most patients was 1 to 3 months. The baseline HAM-D score was 28.6 for the paroxetine group and 28.9 for the amitriptyline group. The mean doses of paroxetine and amitriptyline were 33.3 and 166 mg a day, respectively. Overall improvement was equivalent in the two groups, with response rates (HAM-D reduced to £14) of 66 and 64 percent for the paroxetine and amitriptyline groups, respectively, after 6 weeks. Paroxetine also showed efficacy comparable to that of fluoxetine in a study investigating the treatment of revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) major depression in 178 inpatients randomized to receive paroxetine or fluoxetine, both dosed at 20 mg a day. Subjects were from 18 to 65 years of age, with a baseline MADRS score of 24 or above. At the end of 6 weeks the response rates (MADRS score £11) were 67 percent for the paroxetine group and 64 percent for the fluoxetine group.

Paroxetine and nefazodone (Serdone) were compared in the treatment of DSM-III-R major depression in 206 outpatients with moderate severity (18-item HAM-D score ³17). The average age of patients was approximately 38. Paroxetine (mean dosage, 32.7 mg a day) and nefazodone (mean dosage, 472.0 mg a day) showed equal antidepressant efficacy.

The treatment of geriatric depression with paroxetine was compared with treatment with fluoxetine in a 6-week double-blind study of 106 depressed geriatric (age 65 years or above) outpatients. Patients were randomly assigned to paroxetine (20 to 40 mg a day) or fluoxetine (20 to 60 mg a day) following a 3- to 7-day washout period. At week 3, the paroxetine group showed greater improvement in the HAM-D than the fluoxetine group; however, the two groups showed no difference in efficacy by week 6. HAM-D scores at baseline were approximately 29 for the paroxetine group and 28 for the fluoxetine group. Following 6 weeks of treatment, scores of the the paroxetine group had decreased to approximately 20 and of the fluoxetine group to about 23 in the last observation carried forward analysis. Six weeks is generally believed to be too short a time to obtain optimal treatment results in a geriatric population.

Most antidepressant treatment trials involve 4 to 8 weeks of active medication, often following a variable placebo run-in period. Data indicate that paroxetine is effective in sustaining improvement for at least 12 months with efficacy comparable to that of imipramine but with a considerably more favorable side-effect profile. In this continuation study, patients who had responded to paroxetine, imipramine, or placebo were eligible to continue receiving the same drug for 12 additional months. After 1 year, both antidepressants were more effective than placebo in maintaining euthymia, but the imipramine group had twice the rate of dropout due to adverse experiences of the paroxetine group. Recognizing that full improvement often does not occur with 6 weeks of pharmacotherapy, another long-term extension analysis with paroxetine demonstrated continued improvement after the initial 6-week treatment phase. The baseline HAM-D score of 27.9 declined to 13.5 after 6 weeks of paroxetine treatment. After 12 months, the HAM-D score was 6.9 with no new emergence of adverse effects.

Obsessive-Compulsive Disorder Paroxetine has been compared with clomipramine in the treatment of obsessive-compulsive disorder. A randomized, placebo-controlled trial of 406 subjects with obsessive-compulsive disorder of at least 6 months duration found paroxetine (mean dosage 37.5 mg a day) equal in efficacy to clomipramine (mean dosage, 113.1 mg a day) but better tolerated in terms of adverse effects. In this study 51 percent of the paroxetine group received 60 mg a day, and 51 percent of the clomipramine group received 150 mg a day or more. Dosage-finding studies of paroxetine found dosages of 40 and 60 mg a day (but not 20 mg a day) more effective than placebo, as determined by decreases in Yale Brown Obsessive Compulsive Scale (Y-BOCS) scores. Patients with moderate-to-severe obsessive-compulsive disorder with baseline Y-BOCS scores of 23 to 26 demonstrated a mean decrease of 6 and 7 points for the 40 and 60 mg a day dosage, respectively, following 12 weeks of treatment. This was significantly better than the 4-point decrease in the 20 mg a day group or the 3-point reduction observed in the group receiving placebo.

Panic Disorder Three large-scale, placebo-controlled studies demonstrated the efficacy of paroxetine in DSM-III-R panic disorder with or without agoraphobia. The efficacy of paroxetine in reducing the mean number of panic attacks is shown in Figure 31.25e-2. In a 10-week dosage-finding study comparing placebo with 10, 20, or 40 mg a day, only the 40 mg a day dosage demonstrated superiority to placebo. Endpoint analysis showed 76 percent of patients receiving 40 mg a day of paroxetine to be free of panic attacks, compared with 44 percent of the group receiving placebo. In an extension of this study responders continued for 3 months in a double-blind arm of their original treatment and were then randomized to paroxetine 10, 20, or 40 mg a day or placebo for an additional 3 months. Those randomized to paroxetine were significantly more likely to remain in remission than those in the placebo group. A 12-week flexible-dosing study found that paroxetine (10 to 60 mg a day) eliminated panic attacks in 51 percent of participants, compared with 32 percent of the placebo group. The mean paroxetine dosage for study completers was 40 mg a day. A third 12-week flexible-dosing study compared the efficacy of paroxetine (10 to 60 mg a day) plus cognitive behavioral therapy with placebo plus cognitive behavioral therapy in the treatment of panic disorder. At the endpoint of the study, 33 percent of the paroxetine group had 0 to 1 panic attacks per week, compared with only 14 percent of the placebo group. The mean paroxetine dosage for study completers was 40 mg a day.

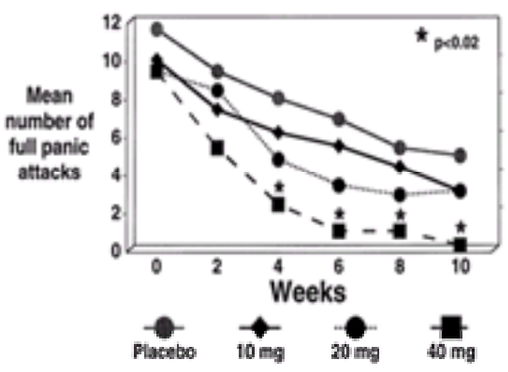

FIGURE 31.25e-2 Paroxetine in the treatment of panic disorder and the mean decrease in number of full panic attacks following treatment with placebo, or paroxetine 10, 20, or 40 mg a day.

Social Phobia An open-label trial of paroxetine (mean final dosage, 36.6 ± 15.0 mg a day) showed efficacy of the SSRI in 15 of 18 subjects rated moderate or marked responders. In another open-label study of paroxetine over 11 weeks, 23 of 30 subjects were considered responders on the basis of a Clinical Global Impressions rating of “very much improved” or “improved.” The mean final dosage of paroxetine was 47.9 ± 6.2 mg a day. Unpublished data derived from placebo-controlled, randomized, double-blind trials reveal the efficacy of paroxetine in the treatment of social phobia. These data are currently being evaluated for submission to the Food and Drug Administration (FDA) to seek approval for the treatment of social phobia.

Premenstrual Dysphoric Disorder Paroxetine has been studied as a treatment for premenstrual dysphoric disorder. In a small, open-label trial, 14 women with moderate-to-severe fourth edition of the DSM (DSM-IV) premenstrual dysphoric disorder with anger and irritability as a central feature were treated with paroxetine. After three active treatment cycles, 7 of 14 patients were given “very much improved” or “much improved” Clinical Global Impressions ratings. The average paroxetine dosage was 22 ± 10 mg a day. In a placebo-controlled comparison study with maprotiline (Ludiomil), a selective norepinephrine reuptake inhibitor and tetracyclic antidepressant, paroxetine (*N* = 22) at a median dosage of 20 mg a day was superior to placebo (*N* = 22) in symptom reduction. Paroxetine was superior to maprotiline (*N* = 22) in reducing symptoms of irritability, increased appetite and carbohydrate craving, bloating, and breast tenderness.

Premature Ejaculation Sexual dysfunction, a common adverse effect of all the SSRIs, takes the form of delayed ejaculation in men and delayed orgasm or anorgasmia in women. In a placebo-controlled trial, paroxetine (40 mg a day) was shown effective in treating premature ejaculation in men.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Respiratory System In placebo-controlled clinical trials, patients receiving paroxetine had rates of yawning and rhinitis of 4 and 3 percent, respectively, compared with 0 percent in the placebo treated group.

Heart and Blood Pressure Paroxetine has not shown any significant effects on heart rate or blood pressure. In a comparison of cardiac effects of paroxetine (20 mg a day) and amitriptyline (150 mg a day), 24 patients with major depression and 24 age and sex matched normal controls were evaluated across a number of cardiac measures. No significant differences were found between the depressed patients and healthy controls in heart rate variability while resting or during deep breathing, spectral analysis of heart rate, the Valsalva test or Posture index. The 24 patients with depression were then randomized to treatment with paroxetine or amitriptyline. Paroxetine elicited no change in any of the indexes in contrast to amitriptyline, which elicited a decrease in all measures but increased heart rate. In a study of patients with comorbid ischemic heart disease and major depression, paroxetine in contrast to nortriptyline (Aventyl, Pamelor) produced no significant alterations in blood pressure, heart rate, cardiac rhythm, or other adverse cardiovascular events.

Blood Paroxetine does not appear to produce any adverse hematological effects.

Gastrointestinal System Nausea is the most common gastrointestinal adverse effect of paroxetine, as it is for the other SSRIs. The incidence of nausea in placebo-controlled trials of paroxetine was usually about 25 percent in the SSRI-treated group. The placebo-treated groups had an incidence of nausea ranging from 9 to 17 percent. Nausea seems to be dose related, and slower escalation of the dosage and ingesting the medication with meals often reduce its incidence. In contrast to other SSRIs, paroxetine seems more likely to produce constipation than diarrhea.

Central Nervous System Like other SSRIs, paroxetine displays CNS adverse effects. Sedation, insomnia, dizziness, and tremor are the most common CNS side effects seen with paroxetine treatment. Both sedation and insomnia can often be easily managed by administering paroxetine at bedtime or in the morning, respectively. Examination of cognitive and motor function have shown no deleterious effects of paroxetine. Paroxetine does not alter the metabolism of alcohol, but as with any CNS active agent, patients should be cautioned about combining the antidepressant with alcoholic beverages.

THERAPEUTIC INDICATIONS

Paroxetine is approved by the FDA for the treatment of major depressive disorder, obsessive-compulsive disorder, and panic disorder. Like many of the other SSRIs it is used in other anxiety and mood disorders such as dysthymic disorder, depression in bipolar disorders, premenstrual dysphoric disorder, social phobia, generalized anxiety disorder, and posttraumatic stress disorder.

PRECAUTIONS AND ADVERSE REACTIONS

The most prevalent CNS effects include headache (although not more frequently than in the placebo group), somnolence, insomnia, and asthenia. In premarket testing, the seizure incidence was 0.1 percent, which is similar to the rate seen with other SSRI antidepressants. Common gastrointestinal adverse effects include nausea, dry mouth, and constipation. Sexual dysfunction was reported to occur at a low rate in the premarketing trials. The actual incidence in clinical practice (as with the other SSRIs) is much higher and consists of delayed ejaculation in men and delayed orgasm or anorgasmia in women as well as occasionally decreased libido. A number of strategies have been suggested to deal with this adverse effect including decreased dosage, drug holidays for shorter-half-life SSRIs, or addition of buspirone (BuSpar) yohimbine (Yocon), amantadine (Symmetrel), cyproheptadine (Periactin), or bupropion (Wellbutrin).

DRUG INTERACTIONS

Potential medication interactions with paroxetine include protein-binding interactions or enzymatic inhibition mediated by the cytochrome P450 isoenzyme system. Paroxetine, like most other SSRIs, is tightly protein bound and may potentially displace other highly protein bound drugs such as warfarin (Coumadin) or digitoxin (Crystodigin). The manufacturer warns that there may be a pharmacokinetic interaction between paroxetine and warfarin that prolongs bleeding, though the prothrombin time is unchanged. Addition of paroxetine (like other SSRIs) to a drug regimen in which tightly protein-bound medications are concurrently prescribed warrants careful observation during dosage adjustment.

Paroxetine is a relatively potent inhibitor of CYP 2D6 but appears to lack clinically significant effects at CYP 1A2, 2C, or 3A4. Coadministration of paroxetine with other medications that are substrates of CYP 2D6 should be approached cautiously, with an awareness of potential increases in serum concentrations of drugs that primarily use CYP 2D6 for metabolism. Examples of CYP 2D6 substrates (see Table 31.25a-3) include the type 1c antiarrhythmics encainide, flecainide (Tambocor), propafenone (Rythmol), and mexiletine (Mexitil) and b-adrenergic receptor antagonists labetalol (Normodyne) metropolol (Lopressor), propranolol (Inderal), and timolol (Blocadren). Atenolol (Tenormin), a cardioselective b-adrenergic receptor antagonist, is renally excreted and exhibits no pharmacokinetic interactions. Codeine is inactive as an analgesic and requires metabolism to morphine via CYP 2D6 to be effective in reducing pain. Inhibition of CYP 2D6 can result in loss of analgesic properties. Paroxetine is both a substrate and an inhibitor of CYP 2D6, but an alternative pathway for its metabolism exists that is not saturated.

Paroxetine, like other SSRIs, should not be administered within 14 days of discontinuing treatment with a monoamine oxidase inhibitor (MAOI). At least 14 days should pass before paroxetine treatment is started after treatment with an MAOI is stopped.

LABORATORY INTERFERENCES

Paroxetine is not known to interfere with any commonly performed laboratory assays.

DOSING AND ADMINISTRATION

Paroxetine is available as 10-, 20-, 30-, and 40-mg tablets. The 20-mg tablets are scored. Treatment for major depressive disorder or obsessive-compulsive disorder is usually initiated at a dosage of 20 mg a day. Taking paroxetine with meals, at least initially, reduces the incidence of nausea, and most patients find this adverse effect transient. Patients with major depressive disorder who do not respond to 20 mg a day can have their dosage increased by 10 mg intervals up to 50 mg a day. In studies that established paroxetine to have efficacy in maintenance treatment, patients received a mean dosage of about 30 mg a day.

The target dosage of paroxetine in the treatment of obsessive-compulsive disorder is 40 mg a day, with most responders receiving 40 to 60 mg a day. The patient's response to treatment of obsessive-compulsive disorder is often slower than that seen in major depressive disorder.

Patients with panic disorder usually tolerate the adverse effects of paroxetine, like antidepressants in general, if paroxetine treatment is started at a lower dosage than is typically used in treating depression. Most panic disorder patients do better if started at an initial dosage of 10 mg a day with gradual titration to a target dosage of 40 mg a day. Dosage increases can be made at the rate of 10 mg a week. The maximal dosage studied in premarketing clinical trials of paroxetine was 60 mg a day.

Other disorders are usually treated with an initial dosage of 10 to 20 mg a day gradually titrated upward. Paroxetine, like the other SSRIs, is well tolerated over a wide dosing range.

SUGGESTED CROSS-REFERENCES

Information relevant to understanding the biochemistry of paroxetine can be found in Section 1.4, on mononamine neurotransmitters, and Section 1.10, on basic molecular neurology. A more general discussion of the biochemistry of mood disorders is found in Section 14.3. Section 14.7 provides an overview of the many treatments for depressive disorders. Section 31.25c and Section 31.25f discuss two other SSRIs, fluoxetine and sertraline. Finally, the use of SSRIs in the treatment of psychiatric disorders is discussed in the relevant chapters (e.g., Chapter 14 on mood disorders, Chapter 15 on anxiety disorders, and Section 51.3d on geriatric mood disorders).

SECTION REFERENCES

*Agid O, Lerer B: Risperidone augmentation of paroxetine in a case of severe, treatment-refractory obsessive-compulsive disorder without comorbid psychopathology. J Clin Psychiatry *60:*55, 1999.

Anonymous: Paroxetine: A selective serotonin reuptake inhibitor showing better tolerance, but weaker antidepressant effect than clomipramine in a controlled multicenter study. Danish University Antidepressant Group. J Affect Disord *18:*289, 1990.

Arminen SL, Ikonen U, Pulkkinen P, Leinonen E, Mahlanen A, Koponen H, Kourula K, Ryyppo J, Korpela V, Lehtonen ML, Vartiainen H, Lehtinen V, Tamminen T, Manniche PM: A 12-week double-blind multi-centre study of paroxetine and imipramine in hospitalized depressed patients. Acta Psychiatr Scand *89:*382, 1994.

Baldwin DS, Hawley CJ, Abed RT, Maragakis BP, Cox J, Buckingham SA, Pover GH, Ascher A: A multicenter double-blind comparison of nefazodone and paroxetine in the treatment of outpatients with moderate-to-severe depression. J Clin Psychiatry *57*(Suppl):46, 1996.

*Ballenger JC, Wheadon DE, Steiner M, Bushnell W, Gergel IP: Double-blind, fixed- dose, placebo-controlled study of paroxetine in the treatment of panic disorder. Am J Psychiatry *155:*36, 1998.

Benkert O, Szegedi A, Wetzel H, Staab HJ, Meister W, Philipp M: Dose escalation vs. continued doses of paroxetine and maprotiline: A prospective study in depressed out-patients with inadequate treatment response. Acta Psychiatr Scand *95:*288, 1997.

Claghorn J: A double-blind comparison of paroxetine and placebo in the treatment of depressed outpatients. Int Clin Psychopharmacol *6*(Suppl):25, 1992.

Cohn JB, Wilcox CS: Paroxetine in major depression: A double-blind trial with imipramine and placebo. J Clin Psychiatry *53*(Suppl):52, 1992.

De Wilde J, Spiers R, Mertens C, Bartholome F, Schotte G, Leyman S: A double-blind, comparative, multicentre study comparing paroxetine with fluoxetine in depressed patients. Acta Psychiatr Scand *87:*141, 1993.

Edwards JG, Goldie A, Papayanni-Papasthatis S, Punter J, Sedgwick EM: The effect of paroxetine on the electroencephalogram, electrocardiogram, and blood pressure. Acta Psychiatr Scand Suppl *350:*124, 1989.

Feighner JP: A double-blind comparison of paroxetine, imipramine and placebo in depressed outpatients. Int Clin Psychopharmacol *6*(Suppl):31, 1992.

Feighner JP, Cohn JB, Fabre LF Jr, Fieve RR, Mendels J, Shrivastava RK: A study comparing paroxetine placebo and imipramine in depressed patients. J Affect Disord *28:*71, 1993.

*Foster CA, Bafaloukos J: Paroxetine in the treatment of chronic daily headache. Headache *34:*587, 1994.

*Franchini L, Gasperini M, Perez J, Smeraldi E, Zanardi R: Dose-response efficacy of paroxetine in preventing depressive recurrences: A randomized, double-blind study. J Clin Psychiatry. *59:*229, 1998.

Geretsegger C, Stuppaeck CH, Mair M, Platz T, Fartacek R, Heim M: Multicenter double blind study of paroxetine and amitriptyline in elderly depressed inpatients. Psychopharmacology *119:*277, 1995.

Guillibert E, Pelicier Y, Archambault JC, Chabannes JP, Clerc G, Desvilles M, Guibert M, Pagot R, Poisat JL, Thobie Y: A double-blind, multicentre study of paroxetine versus clomipramine in depressed elderly patients. Acta Psychiatr Scand Suppl *350:*132, 1989.

Kuhs H, Rudolf GA: A double-blind study of the comparative antidepressant effect of paroxetine and amitriptyline. Acta Psychiatr Scand Suppl *350:*145, 1989.

Kuhs H, Schlake HP, Rolf LH, Rudolf GA: Relationship between parameters of serotonin transport and antidepressant plasma levels or therapeutic response in depressive patients treated with paroxetine and amitriptyline. Acta Psychiatr Scand *85:*364, 1992.

Lauritzen L, Odgaard K, Clemmesen L, Lunde M, Ohrstrom J, Black C, Bech P: Relapse prevention by means of paroxetine in ECT-treated patients with major depression: A comparison with imipramine and placebo in medium-term continuation therapy. Acta Psychiatr Scand *94:*241, 1996.

Lecrubier Y, Bakker A, Dunbar G, Judge R: A comparison of paroxetine, clomipramine and placebo in the treatment of panic disorder. Collaborative Paroxetine Panic Study Investigators. Acta Psychiatr Scand *95:*145, 1997.

Lecrubier Y, Judge R: Long-term evaluation of paroxetine, clomipramine and placebo in panic disorder. Collaborative Paroxetine Panic Study Investigators. Acta Psychiatr Scand *95:*153, 1997.

Lund J, Thayssen P, Mengel H, Pedersen OL, Kristensen CB, Gram LF: Paroxetine: Pharmacokinetics and cardiovascular effects after oral and intravenous single doses in man. Acta Pharmacol Toxicol *51:*351, 1982.

*Lydiard RB, Steiner M, Burnham D, Gergel I: Efficacy studies of paroxetine in panic disorder. Psychopharmacol Bull *34:*175, 1998.

Marshall RD, Schneier FR, Fallon BA, Knight CB, Abbate LA, Goetz D, Campeas R, Liebowitz MR: An open trial of paroxetine in patients with noncombat-related, chronic posttraumatic stress disorder. J Clin Psychopharmacol *18:*10, 1998.

Mertens C, Pintens H: A double-blind, multicentre study of paroxetine and mianserin in depression. Acta Psychiatr Scand Suppl *350:*140, 1989.

Montgomery SA, Dunbar G: Paroxetine is better than placebo in relapse prevention and the prophylaxis of recurrent depression. Int Clin Psychopharmacol *8:*189, 1993.

Nemeroff CB: Paroxetine: An overview of the efficacy and safety of a new selective serotonin reuptake inhibitor in the treatment of depression. J Clin Psychopharmacol *13*(Suppl):10S, 1993.

Oehrberg S, Christiansen PE, Behnke K, Borup AL, Severin B, Soegaard J, Calberg H, Judge R, Ohrstrom JK, Manniche PM: Paroxetine in the treatment of panic disorder. A randomised, double-blind, placebo-controlled study. Br J Psychiatry *167:*374, 1995.

Ozdemir V, Naranjo CA, Herrmann N, Reed K, Sellers EM, Kalow W: Paroxetine potentiates the central nervous system side effects of perphenazine: Contribution of cytochrome P4502D6 inhibition in vivo. Clin Pharmacol Ther *62:*334, 1997.

Raptopoulos P, McClelland GR, Jackson D: The clinical pharmacology of paroxetine in healthy subjects. Acta Psychiatr Scand Suppl *350:*46, 1989.

Rechlin T: The effect of amitriptyline, doxepin, fluvoxamine, and paroxetine treatment on heart rate variability. J Clin Psychopharmacol *14:*392, 1994.

Rey-Sanchez F, Gutierrez-Casares JR: Paroxetine in children with major depressive disorder: An open trial. J Am Acad Child Adolesc Psychiatry *36:*1443, 1997.

Robbe HW, O'Hanlon JF: Acute and subchronic effects of paroxetine 20 and 40 mg on actual driving, psychomotor performance and subjective assessments in healthy volunteers. Eur

Neuropsychopharmacol *5:*35, 1995.

Rocca P, Fonzo V, Scotta M, Zanalda E, Ravizza L: Paroxetine efficacy in the treatment of generalized anxiety disorder. Acta Psychiatr Scand *95:*444, 1997.

Schnyder U, Koller-Leiser A: A double-blind, multicentre study of paroxetine and maprotiline in major depression. Can J Psychiatry *41:*239, 1996.

Schone W, Ludwig M: A double-blind study of paroxetine compared with fluoxetine in geriatric patients with major depression. J Clin Psychopharmacol *13:*34S, 1993.

*Shrivastava RK, Shrivastava SH, Overweg N, Blumhardt CL: A double-blind comparison of paroxetine, imipramine, and placebo in major depression. J Clin Psychiatry *53*(Suppl):48, 1992.

Stein MB, Liebowitz MR, Lydiard RB, Pitts CD, Bushnell W, Gergel I: Paroxetine treatment of generalized social phobia (social anxiety disorder): A randomized controlled trial. JAMA *280:*708, 1998.

Stuppaeck CH, Geretsegger C, Whitworth AB, Schubert H, Platz T, Konig P, Hinterhuber H, Fleischhacker WW: A multicenter double-blind trial of paroxetine versus amitriptyline in depressed inpatients. J Clin Psychopharmacol *14:*241, 1994.

Szegedi A, Wetzel H, Angersbach D, Dunbar GC, Schwarze H, Philipp M: A double-blind study comparing paroxetine and maprotiline in depressed outpatients. Pharmacopsychiatry *30:*97, 1997.

Waldinger MD, Hengeveld MW, Zwinderman AH: Ejaculation-retarding properties of paroxetine in patients with primary premature ejaculation: A double-blind, randomized, dose-response study. Br J Urol *79:*592, 1997.

Yonkers KA, Gullion C, Williams A, Novak K, Rush AJ: Paroxetine as a treatment for premenstrual dysphoric disorder. J Clin Psychopharmacol *16:*3, 1996.

Zohar J, Judge R: Paroxetine versus clomipramine in the treatment of obsessive-compulsive disorder. OCD Paroxetine Study Investigators. Br J Psychiatry *169:*468, 1996.

CHAPTER 31. BIOLOGICAL THERAPIES

31.25 SELECTIVE SEROTONIN REUPTAKE INHIBITORS

31.25F SERTRALINE

JEFFREY E. KELSEY, M.D., FH.D. AND CHARLES B. NEMEROFF, M.D., FH.D.

HISTORY

In 1992 sertraline (Zoloft) was the second selective serotonin reuptake inhibitor (SSRI) approved for use in the United States. Initially approved for the treatment of major depression, sertraline is currently also approved for the treatment of obsessive-compulsive disorder and panic disorder.

Sertraline is a member of the SSRI class of compounds that were developed on the basis of the theory that dysregulation of serotonergic circuits within the central nervous system (CNS) occur in depression and that alterations of serotonergic activity by antidepressant compounds mediate their therapeutic responses.

CHEMISTRY

Sertraline's empirical formula is $C_{17}H_{17}NCl_2 \cdot HCl$; it has a molecular weight of 342.7. The structure of sertraline, distinct from that of other SSRIs, is a napthylamine derivative. The chemical designation is (1*S*-*cis*)-4-(3,4-dichlorophenyl)-1,2,3,4-tetrahydro-*N*-methyl-1naphthalenamine hydrochloride and is shown in Figure 31.25f-1.

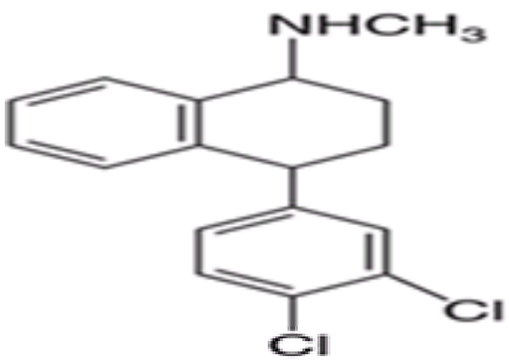

FIGURE 31.25f-1 Molecular structure of sertraline.

PHARMACOLOGICAL ACTIONS

Absorption Sertraline is well absorbed orally, with equivalent absorption from tablets or liquid formulation. Maximum systemic concentration is achieved within 4.5 to 8.4 hours of ingestion following 14 days of dosing in the range of 50 to 200 mg daily. When administered with food, the mean peak plasma concentration is increased by 25 percent, and the peak plasma concentration is achieved 5.5 hours after ingestion rather than 8 hours. Whether this food-related increase in sertraline plasma concentrations is clinically significant in terms of adverse events is unclear.

Distribution In vivo, sertraline is approximately 98 percent bound to serum proteins at concentrations of 20 to 500 ng/mL. The volume of distribution for sertraline is approximately 20 liters per kilogram, and the rate of plasma clearance is 96 liters per hour. The distribution of sertraline into breast milk has been studied in detail. The data suggest that sertraline is present in breast milk and in infant serum, albeit in low concentrations.

Metabolism and Elimination Sertraline undergoes demethylation in the liver to *N*-desmethylsertraline. The half-life of sertraline is 26 hours, and that of *N*-desmethyl sertraline is 62 to 104 hours. However the metabolite is 5 to 10 times less potent in inhibiting serotonin reuptake than the parent compound. Desmethylsertraline displays a five- to nine-fold increase in area under the curve and peak and minimum plasma concentrations following 14 days of sertraline administration. In healthy volunteers, sertraline demonstrates linear pharmacokinetics over the range of 50 to 200 mg a day. Using radiolabeled sertraline in two healthy male subjects, unchanged sertraline was found to constitute less than 5 percent of the plasma radioactivity. After 9 days, 40 to 45 percent of the radioactivity was recovered in urine, but no unchanged sertraline was detected. Another 40 to 45 percent was recovered in feces, of which 12 to 14 percent was unchanged sertraline.

Liver disease is associated with impaired metabolism of sertraline. In a series of subjects with mild, stable cirrhosis of the liver, the mean half-life of sertraline was increased to 52 hours, compared with 22 hours for normal subjects. This suggests the necessity of dosage reduction in individuals with hepatic disease.

Insufficient data are available to assess the effect of renal impairment on sertraline metabolism.

Aging is associated with decreased sertraline clearance. In 16 elderly patients treated with sertraline (100 mg a day for 14 days), the clearance of sertraline was 40 percent lower than that observed in younger (25 to 32-year-old) subjects. Desmethylsertraline clearance was decreased in older males but not females.

PHARMACODYNAMICS

Mechanism of Action The antidepressant effect of sertraline is believed to be due to serotonin reuptake inhibition at the presynaptic nerve terminal. Sertraline is a weak inhibitor of dopamine reuptake but has no effect on norepinephrine reuptake. Sertraline inhibits platelet serotonin reuptake in normal male volunteers in a dosage-dependent fashion. Sertraline also inhibits *p*-chloramphetamine-induced serotonin depletion in the rat brain, which further supports the mechanism of serotonin reuptake inhibition as a mode of action. Short-term administration (hours) of sertraline reduces serotonin turnover in rats, but this effect is not seen following long-term administration. Long-term administration in rats reduces hydrolysis of phosphatidylinositol 4,5-biphosphate, the second messenger that mediates serotonin (5-hydroxytryptamine [5-HT] type 2) ($5\text{-}HT_2$) signal transduction, but without downregulating the receptor. In humans, sertraline at therapeutic dosages (50 to 200 mg a day) does not seem to produce stimulation, sedation, or changes in psychomotor performance.

Like many other drugs in the class, sertraline is essentially devoid of any clinically significant receptor antagonistic properties at the muscarinic cholinergic, histamine type 1 (H_1), a_1-, a_2-, or b-adrenergic, and $5\text{-}HT_1$ or $5\text{-}HT_2$ receptors. Sertraline is about 4 times less potent than paroxetine and 60 times less potent than amitriptyline at inhibiting [3H]quinuclidinylbenzilate binding, which is a measure of muscarinic cholinergic receptor antagonism. Sertraline tends to be significantly better tolerated

than tricyclic drugs. Sertraline also lacks the cardiac fast sodium channel blockade seen with tricyclic drugs and thus demonstrates a superior safety profile with regard to cardiac toxicity than these older agents. Sertraline has no effect on monoamine oxidase activity.

Blood Levels and Relation to Action There does not appear to be a clear relationship between plasma concentrations of sertraline and subsequent therapeutic response.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Major Depressive Disorder Sertraline has been extensively studied in subjects with depression in comparison to placebo and to other antidepressants. Most of these studies have involved healthy adult outpatients 18 to 65 years of age. In a dose-ranging study comparing sertraline at 50, 100, or 200 mg a day with placebo in 369 outpatients with third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) depression, all three dosages in the evaluable patients group (taking study medication for 11 days or more) showed efficacy superior to that of placebo. Females constituted 53 percent of the study population and the mean age of subjects was 37.6 years. Among patients who completed the study, there was a trend for greater improvement in the 100 or 200 mg a day groups, though this did not attain statistical significance. When the data for all patients (as contrasted to study completers) were analyzed, all three sertraline dosage groups had equal efficacy. This may partly be an artifact, since the higher-dosage group had fewer completers (43 percent) than the 50-mg group (62 percent).

Comparison With Other Antidepressants Some 186 outpatients with DSM-III major depression were studied to compare sertraline with the tricycle drug amitriptyline (Elavil). Sertraline (50 to 200 mg a day) was comparable to amitriptyline (50 to 150 mg a day) and had a more benign adverse-effect profile in an 8-week trial. Sertraline was comparable to dothiepin and amitriptyline in two other studies of outpatients with major depression and showed somewhat greater efficacy than imipramine in one study of outpatients with major depression. Sertraline was also demonstrated to possess efficacy comparable to that of moclobemide (Aurorix) in depressed outpatients. In a double-blind, 6-week study of 286 outpatients with revised third edition of DSM (DSM-III-R) major depression (unipolar or bipolar), sertraline (50 to 100 mg a day) and fluoxetine (20 to 40 mg a day) showed equal efficacy.

Depression is a recurrent disorder for many patients. The efficacy of sertraline in preventing the recurrence of a major depressive episode was studied in a comparative study with fluvoxamine (Luvox). Sixty-four patients with a mean age of approximately 50 years and an average of seven lifetime episodes of depression were randomized to receive sertraline 100 mg a day or fluvoxamine 200 mg a day. If a patient's depression worsened (functional impairment and a Hamilton Rating Scale for Depression [HAM-D] score above 15) sertraline was increased to 200 mg a day or fluovoxamine increased to 300 mg a day. The agents were equally effective in this population, with only 7 of 32 sertraline- and 6 of 32 fluvoxamine-treated patients experiencing a recurrence during the 24 months of the study.

Dysthymia Sertraline was shown effective in the treatment of dysthymia. Some 416 patients with revised third edition of DSM (DSM-III-R) dysthymia (271 females, 145 males, aged 25 to 65 years) of at least 5 years duration and without current major depression were randomized to 12 weeks of treatment with sertraline, imipramine (Tofranil), or placebo. The mean duration of illness for all groups was approximately 30 years, with an average age of onset of 12 years of age. The baseline HAM-D scores (17 item) were 12.7 ± 4.0, 13.4 ± 3.8, and 12.7 ± 3.9 for the sertraline, imipramine, and placebo groups, respectively. Both sertraline and imipramine were superior to placebo in improving depression severity ratings, psychosocial function, and quality of life. The mean final dosage of sertraline was 139.6 ± 58.5 mg a day, and of imipramine, 198.8 ± 91.2 mg a day. Sertraline appeared to be better tolerated, with discontinuation rates due to adverse events of 6 percent for sertraline and 18.4 percent for imipramine.

Obsessive-Compulsive Disorder Sertraline is approved by the Food and Drug Administration (FDA) for the treatment of obsessive-compulsive disorder. A placebo-controlled, randomized trial with 50, 100, or 200 mg a day of sertraline in 324 patients with obsessive compulsive disorder without comorbid depression demonstrated efficacy of the 50 and 200 mg a day dosages in a 12-week treatment trial following 1 week of placebo lead-in. Individuals who met entry criteria (National Institute of Mental Health Global Obsessive Compulsive Scale Score ³7 and HAM-D, excluding item 21-obsessions, score £17, and a score of £1 for the depression item) after 1 week of placebo were randomized to 50, 100, or 200 mg a day of sertraline or placebo. Subjects in the two higher dosage groups were titrated upward in dosage to reach 100 mg a day by day 5 and 200 mg a day by day 14. The 100 mg a day dosage differed from placebo in efficacy measures only for the National Institute of Mental Health Global Obsessive Compulsive Scale. A higher dropout rate obtained in the 100 mg a day group (27 of 81) than in the 50 mg a day (17 of 80), 200 mg a day (21 of 80), or placebo groups (24 of 84), which may account for the lack of efficacy in the 100 mg a day group.

Panic Disorder The efficacy of sertraline in panic disorder with or without agoraphobia was assessed in three double-blind, placebo-controlled trials. In two studies, sertraline at mean dosages of 131 and 144 mg a day, was superior to placebo in change from baseline for number of full panic attacks. Sertraline was more effective than placebo in improvement and severity as assessed by the Clinical Global Impression Scale. A third, fixed-dosage study also demonstrated superiority over placebo in panic attack frequency.

Premenstrual Dysphoric Disorder SSRIs are generally considered effective in the treatment of premenstrual dysphoric disorder. In a multicenter, placebo-controlled, double-blind trial with 162 subjects, sertraline produced a positive response (“very much” or “much improved” with the Clinical Global Impression Scale) in 68 percent of patients, compared with 40 percent of those in the placebo group. Dosing of sertraline was 50 mg a day in cycle 1, and the mean dosage were 82 mg a day in cycle 2 and 108 mg a day in cycle 3. By the end of the study 18 percent were receiving 50 mg a day, 48 percent received 100 mg a day, and 34 percent received 150 mg a day of sertraline. In an open-label comparison trial with desipramine (Norpramine) in 32 women with premenstrual syndrome, desipramine (mean dosage, 110 mg a day in the second cycle studied) and sertraline (mean dosage 87 mg a day in the second cycle studied) both improved symptoms. There was a trend (not statistically significant) in favor of sertraline, and the adverse-effect profile favored the SSRI.

Postpartum Depression Sertraline has efficacy in the treatment of postpartum depression. In an open-label trial, 20 of 21 study completers had a decrease of 50 percent or more from baseline in the structured interview for the 21-item HAM-D. Fourteen women obtained remission (HAM-D < 7, CGI = 1, and Global Assessment of Function > 80). The mean daily dosage for the 21 study completers was 109 ± 37 mg a day of sertraline after 8 weeks of treatment.

Posttraumatic Stress Disorder Sertraline had reported efficacy in posttraumatic stress disorder in a pilot, 12-week, open-label study. The mean dosage was 105 mg a day, and 4 of 5 study completers responded. Patients were adult female rape victims with chronic posttraumatic stress disorder, and the completers were on average 41.6 years old and 15.6 years postassault. A large-scale, double-blind trial is currently under way.

Premature Ejaculation Sexual dysfunction, a common adverse effect of all SSRIs, is characterized by delayed ejaculation in males and delayed orgasm or anorgasmia in females. Like fluoxetine (Prozac), sertraline has been studied as a possible treatment for premature ejaculation. In a double-blind, placebo-controlled trial 52 males with self-reported premature ejaculation were randomized to either placebo or sertraline (50 to 200 mg a day) for 8 weeks. Sertraline was more effective than placebo in prolonging latency to ejaculation (self-report), number of successful attempts at intercourse, and overall self-rated improvement.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Sertraline appears to be quite well tolerated in general, with only 15 percent of over 2500 patients in premarketing trials discontinuing because of adverse events. The higher dosages had more dropouts than the lower doses, which may partly reflect forced titration schedules, which are often more aggressive than those in clinical practice.

Heart and Blood Pressure Sertraline has not been found to alter cardiac rhythm or blood pressure significantly. Individuals with existing cardiac disease were excluded from premarketing trials, but no changes on electrocardiograms were noted with treatment. Treatment with sertraline in premarketing trials was associated with a 3 percent increase in total cholesterol and a 5 percent increase in triglycerides.

Blood Abnormal bleeding or bruising has been reported in patients taking SSRIs, including sertraline, but appears to be relatively rare. In premarketing trials sertraline possessed a weak uricosuric effect, with a mean decrease in uric acid of 7 percent. The clinical significance of this, if any, is unclear. Hyponatremia has been re ported after treatment with sertraline and other SSRIs; this effect is reversed upon discontinuation of sertraline and may be secondary to the syndrome of inappropriate antidiuretic hormone secretion. Groups at risk include elderly patients, patients concomitantly treated with diuretics, and patients who were volume depleted.

Gastrointestinal System The incidence of nausea in controlled clinical trials is higher for sertraline than for placebo (26.1 vs. 11.8 percent). In a pooled analysis of placebo-controlled trials about 4 percent of patients taking sertraline experienced an increase or decrease of 7 percent of body weight or more; however, the difference from the placebo groups was not statistically significant. Sertraline also was associated with a higher (17.7 percent) incidence of diarrhea than placebo (9.3

percent) in premarketing trials and likely higher than that observed with fluoxetine or paroxetine.

Skin Approximately 2.1 percent of participants and 1.5 percent of the placebo group in premarketing trials experienced rash. Excessive sweating was reported with sertraline at a rate three time (8.4 versus 2.9 percent) that seen in the placebo treatment groups.

Central Nervous System In premarketing trials, sertraline produced no greater incidence of headache (20.3 vs. 19.0 percent) than placebo. Insomnia, somnolence, agitation, nervousness, anxiety, yawning, and impaired concentration were all reported more often in the sertraline group than the placebo group. The incidence of dizziness, tremor, twitching, hypoesthesia, and hypertonia were also all greater in the sertraline treated group than in the placebo group.

Sexual Dysfunction The reported incidence of sexual dysfunction in sertraline-treated individuals was 15.5 percent for males (2.2 percent in placebo group) and 1.7 percent for females (0.5 percent), though in clinical practice, it is now clear that sertraline and all other SSRIs produce significantly higher rates of sexual dysfunction.

THERAPEUTIC INDICATIONS

Sertraline is currently approved by the FDA to be marketed for use in major depressive disorder, obsessive-compulsive disorder, and panic disorder. Like other newer-generation antidepressants, sertraline is also used in other affective and anxiety disorders including dysthymia, premenstrual syndrome, social phobia, posttraumatic stress disorder, and generalized anxiety disorder.

PRECAUTIONS AND ADVERSE REACTIONS

The most commonly experienced CNS effects include headache (but not greater than placebo), insomnia, somnolence, dizziness, and tremor. Common gastrointestinal system adverse effects include nausea, diarrhea, and dyspepsia. Nausea is generally tolerated over time and can be minimized by administration with food, even though this strategy increases maximal serum concentrations of sertraline. Sexual dysfunction was reported at a low incidence in the premarketing trials. The actual incidence in clinical practice is much higher and consists of delayed orgasm or anorgasmia as well as decreased libido. A number of strategies have been suggested to deal with this important adverse effect including decreased dosage, drug holidays for shorter half-life SSRIs or addition of buspirone (BuSpar) amantidine (Symmetrel), yohimbine (Yocon), cyproheptadine (Periactin), or bupropion (Wellbutrin).

Drug Interactions Sertraline, like fluoxetine and paroxetine (Paxil), is tightly bound to serum proteins, and the manufacturer warns about a potential interaction between sertraline and other tightly protein bound drugs such as warfarin (Coumadin) and digitoxin (Crystodigin). Good clinical practice dictates checking serum concentrations or activity when an antidepressant with high protein binding is added to a tightly protein bound drug with a low therapeutic index. Most assays do not differentiate between bound and free drug, so functional assays such as prothrombin time with warfarin are a better indicator.

Sertraline should not be combined with a monoamine oxidase inhibitor (MAOI). At least 14 days should pass after discontinuation of treatment with an MAOI before sertraline is taken. Similarily, MAOI treatment should not be started in someone taking sertraline until the SSRI has been discontinued for at least 14 days.

Sertraline inhibits the cytochrome P450 (CYP) 2D6 isoenzyme. Inhibition of CYP 2C has potential clinical significance when sertraline is coadministered with tolbutamide (Orinase), an oral hypoglycemic drug. It does not inhibit either CYP 1A2 or CYP 3A4.

LABORATORY INTERFERENCES

Sertraline does not interfere with any laboratory measures.

DOSING AND ADMINISTRATION

Sertraline is available in scored tablets of 25-, 50-, and 100-mg strengths. Initial dosing for major depression and obsessive-compulsive disorder is typically 50 mg a day, with escalation to 100 mg a day after 4 to 7 days of treatment. Patients may require higher dosages. In a 6-week trial of sertraline in major depression the mean dosage for study completers was 145 mg a day. However, in a double-blind, placebo-controlled, continuation trial after 8 weeks of open-label treatment, the mean dosage for study completers after 44 weeks was 70 mg a day. Studies of sertraline in double-blind, placebo-controlled trials for obsessive-compulsive disorder have had mean dosages of approximately 185 mg a day.

Patients with panic disorder often require lower starting dosages of antidepressants than do others. A starting dosage of 25 mg a day with sertraline for 4 to 7 days followed by upward dose titration is a reasonable approach. The mean dosage of sertraline in double-blind, placebo-controlled trials for panic disorder was approximately 140 mg a day.

The dosage range of sertraline in most patients who have affective or anxiety disorders is 100 to 200 mg a day, though some patients respond to 50 mg a day. More-gentle titration is required for the very anxious patient or one who is very sensitive to adverse effects. Initial administration with meals often decreases initial and transient adverse effects.

SUGGESTED CROSS-REFERENCES

Additional information regarding monoamines and their receptors appears in Section 1.4. The basic electrophysiology of neurons is discussed in Section 1.9. The classification of mental disorders is discussed in Chapter 9, and mood disorders in geriatric patients in Section 51.3d. The biochemical basis of mood disorders is discussed in Section 14.4. Lithium therapy is discussed in Section 31.18, and the biological treatment of mental illness in elderly adults in Section 51.4.

SECTION REFERENCES

Aguglia E, Casacchia M, Cassano GB, Faravelli C, Ferrari G, Giordano P, Pancheri P, Ravizza L, Trabucchi M, Bolino F: Double-blind study of the efficacy and safety of sertraline versus fluoxetine in major depression. Int Clin Psychopharmacol *8:*197, 1993.

*Alfaro CL, Lam YMF, Simpson J, Ereshefsky L: CYP2D6 status of extensive metabolizers after multiple-dose fluoxetine, fluvoxamine, paroxetine, or sertraline. J Clin Psychopharmacol *19:*155, 1999.

Chouinard G: Sertraline in the treatment of obsessive compulsive disorder: Two double-blind, placebo-controlled studies. Int Clin Psychopharmacol *7*(Suppl):37, 1992.

*Fabre LF, Abuzzahab FS, Amin M, Claghorn JL, Mendels J, Petrie WM, Dube S, Small JG: Sertraline safety and efficacy in major depression: A double-blind fixed-dose comparison with placebo. Biol Psychiatry *38:*592, 1995.

*Franchini L, Gasperini M, Perez J, Smeraldi E, Zanardi R: A double-blind study of long-term treatment with sertraline or fluvoxamine for prevention of highly recurrent unipolar depression. J Clin Psychiatry *58:*104, 1997.

Greist J, Chouinard G, DuBoff E, Halaris A, Kim SW, Koran L, Liebowitz M, Lydiard RB, Rasmussen S, White K: Double-blind parallel comparison of three dosages of sertraline and placebo in outpatients with obsessive-compulsive disorder. Arch Gen Psychiatry *52:*289, 1995.

*Greist JH, Jefferson JW, Kobak KA, Chouinard G, DuBoff E, Halaris A, Kim SW, Koran L, Liebowtiz MR, Lydiard B: A 1-year double-blind placebo-controlled fixed dose study of sertraline in the treatment of obsessive-compulsive disorder. Int Clin Psychopharmacol *10:*57, 1995.

Halbreich U, Smoller JW: Intermittent luteal phase sertraline treatment of dysphoric premenstrual syndrome. J Clin Psychiatry *58:*399, 1997.

Hindmarch I, Bhatti JZ: Psychopharmacological effects of sertraline in normal, healthy volunteers. Eur J Clin Pharmacol *35:*221, 1988.

Keller MB, Kocsis JH, Thase ME, Gelenberg AJ, Rush AJ, Koran L, Schatzberg A, Russell J, Hirschfeld R, Klein D, McCullough JP, Fawcett JA, Kornstein S, LaVange L, Harrison W: Maintenance phase efficacy of sertraline for chronic depression: A randomized controlled trial. JAMA *280:*1665, 1998.

Londborg PD, Wolkow R, Smith WT, DuBoff E, England D, Ferguson J, Rosenthal M, Weise C: Sertraline in the treatment of panic disorder. A multi-site, double-blind, placebo-controlled, fixed-dose

investigation. Br J Psychiatry *173:*54, 1998.

Mendels J, Camera A, Sikes C: Sertraline treatment for premature ejaculation. J Clin Psychopharmacol *15:*341, 1995.

Montejo-Gonzalez AL, Llorca G, Izquierdo JA: SSRI-induced sexual dysfunction: Fluoxetine, paroxetine, sertraline, and fluvoxamine in a prospective, multicenter, and descriptive clinical study of 344 patients. J Sex Marital Ther *23:*176, 1997.

Murdoch D, McTavish D: Sertraline. A review of its pharmacodynamic and pharmacokinetic properties, and therapeutic potential in depression and obsessive-compulsive disorder. Drugs *44:*604, 1992.

*Pohl RB, Wolkow RM, Clary CM: Sertraline in the treatment of panic disorder: A double-blind multicenter trial. Am J Psychiatry *155:*1189, 1998.

*Reimherr FW, Chouinard G, Cohn CK, Cole JO, Itil TM, LaPierre YD, Masco HL, Mendels J: Antidepressant efficacy of sertraline: A double-blind, placebo- and amitriptyline-controlled, multicenter comparison study in outpatients with major depression. J Clin Psychiatry *51*(Suppl):18, 1990.

Rothbaum BO, Ninan PT, Thomas L: Sertraline in the treatment of rape victims with posttraumatic stress disorder. J Trauma Stress *9:*865, 1996.

Stowe ZN, Owens MJ, Landry JC, Kilts CD, Ely T, Nemeroff CB: Sertraline and desmethylsertraline in human breast milk and nursing infants. Am J Psychiatry *154:*1255, 1997.

Stowe ZN, Casarella J, Landry J, Nemeroff CB: Sertraline in the treatment of women with postpartum major depression. Depression *3:*49, 1995.

Thase ME, Fava M, Halbreich U, Kocsis JH, Koran L, Davidson J, Rosenbaum J, Harrison W: A placebo-controlled, randomized clinical trial comparing sertraline and imipramine for the treatment of dysthymia. Arch Gen Psychiatry *53:*777, 1996.

Yonkers KA, Halbreich U, Freeman E, Brown C, Endicott J, Frank E, Parry B, Pearlstein T, Severino S, Stout A, Stone A, Harrison W: Symptomatic improvement of premenstrual dysphoric disorder with sertraline treatment. A randomized controlled trial. Sertraline Premenstrual Dysphoric Collaborative Study Group. JAMA *278:*983, 1997.

Zanardi R, Franchini L, Gasperini M, Perez J, Smeraldi E: Double-blind controlled trial of sertraline versus paroxetine in the treatment of delusional depression. Am J Psychiatry *153:*1631, 1996.

31.26 SEROTONIN-DOPAMINE ANTAGONISTS

DANIEL P. VAN KAMMEN, M.D., FH.D., AND STEPHEN R. MARDER, M.D.

After more than 40 years of dopamine receptor antagonists with often unavoidable extrapyramidal side effects, newer antipsychotic drugs have become available that have no or minimal extrapyramidal adverse effects, and greater efficacy. These are the serotonin-dopamine antagonists (SDAs), named for their alleged mechanism of action. These agents will have a wider utility in psychiatry than the traditional antipsychotic drugs and could be used as mood stabilizers.

Clozapine (Clozaril) was the first antipsychotic agent with negligible extrapyramidal adverse effects and enhanced efficacy to be available in the United States. Introduced in 1990, it heralded a new era in the treatment of schizophrenia, even though clozapine had been available in Europe from the early 1970s. Fortunately for our patients, the animal models for antipsychotic efficacy based on dopamine receptor blockade were not fully established at that time. If cataleptogenic, antiamphetamine and antiapomorphine animal models had been used to test the agent for potential antipsychotic effects, the test results could have led to the rejection of clozapine for clinical trials.

The success of clozapine has inspired development of new and safer antipsychotic agents. The many unpleasant adverse effects, weekly blood sampling, and high cost of clozapine treatment made alternative drugs attractive. Several other drugs are now available, such as risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine (Seroquel), and others will be approved soon (e.g., aripiprazole). Sertindole (Serlect, Serdolect) is available in Europe but not in the United States. Similarly, ziprasidone was not approved in the United States. Other agents with different mechanisms of action are currently being developed. The terms *typical* and *atypical* will lose their meaning, as these new agents become the treatment of choice. The terms *traditional, conventional*, or *first-generation drugs* for the preclozapine dopamine antagonists and *novel* or *second-generation* antipsychotic agents for SDAs may be preferable to the *typical* and *atypical* agents.

Compared with the traditional agents, all second-generation agents have a higher ratio of serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT $_2$) to dopamine type 2 (D_2) receptor blockade and a greater specificity for the mesolimbic than the striatal dopamine system. These new agents showed selectivity for the limbic system in electrophysiological studies (percentage of silent dopamine neurons with chronic treatment) and animal tests (low catalepsy in rodents, and low extrapyramidal adverse effect liability in nonhuman primates). In contrast, traditional antipsychotic agents affect both the limbic and striatal dopamine neurons, which frequently results in extrapyramidal effects at therapeutic doses.

All the new agents are first-line drugs, except clozapine because of its adverse effects and the required weekly blood sampling. Preliminary data show that negative symptoms may respond better to the newer agents (at least risperidone, olanzapine, ziprasidone, sertindole, and quetiapine). Because negative symptoms are a critical determinant of the social and vocational outcome of schizophrenia, these newer agents should improve quality of life. However, whether the difference results from some intrinsic effects of the drug or can be accounted for by the lack of striatal dopamine type 2 (D $_2$) blockade is unclear. The SDAs are more expensive than conventional agents, which has lead to extensive campaigns to ensure that these agents are available for patients who need them. The high costs of the new agents require justification for their usage. Initial pharmacoeconomic studies of these agents suggest that increased efficacy (e.g., decreased rehospitalization rates and bed days of care) may offset the higher costs, particularly in the second year of treatment (e.g., with clozapine and risperidone). Other studies question this decrease in other costs.

Some intriguing observations, such as lower suicidality (during clozapine treatment), less smoking, lower water intoxication risks; continued improvement potential up to 2 years; and decreased or absent risk of tardive dyskinesia require careful evaluation. Additional benefits include the possibility of better long-term outcomes than with conventional agents, diminished relapse risk, and improved quality of life. So far weight gain has been the major adverse effect of the SDAs. This chapter considers clozapine and five other SDAs in order of their historical appearance on the market, not their potential importance. Because clozapine and risperidone have been on the market for the longest time, more is known about these drugs than about the others.

CLOZAPINE

Clozapine was the first SDA to be approved. Its greater efficacy in schizophrenia, effect in treatment-resistant patients, and low risk for extrapyramidol effects changed the thinking about what antipsychotic drugs could do in schizophrenia. However, the adverse effects of clozapine, ranging from dangerous (agranulocytosis and seizures) to clearly uncomfortable (e.g., constipation, sedation, hypersalivation), created the need to search for similar agents. Its adverse-effect profile keeps it from being prescribed for first-episode patients. At this time clozapine remains the gold standard for treatment of treatment-resistant patients. Studies will soon appear that will show whether clozapine nonresponders respond to other SDAs.

History Clozapine was discovered in 1958 in Bern, Switzerland. Clozapine was first studied in animal experiments in May 1960. In 1972 clozapine was put on the market in Switzerland and Austria. Soon after clinical trials with clozapine began, the drug appeared to be an effective antipsychotic agent that did not cause extrapyramidal effects. The first studies showed that patients became calmer in the first 5 days, and psychotic excitement decreased from day 5 on. Day 20 effects on psychosis were substantial, and negative symptoms improved at 40 days. However, the drug had significant hematological toxicity. In 1975, clozapine was introduced in Finland, where 16 of 1600 treated patients developed granulocytopenia (£1600 cells per mm ^{3}). Eight of the 13 patients in whom granulocytopenia progressed to agranulocytosis died of infectious diseases. After 50 patients around the world had died, clozapine was withdrawn from most European markets, and research with this drug came to a virtual halt. Some patients who were receiving clozapine before its withdrawal from the market continued to receive the drug because of its clear advantages over prior treatment. Regular monitoring of white blood counts in these individuals showed that clozapine-induced agranulocytosis was reversible when drug use stopped. In addition, controlled trials in Europe showed that clozapine was more effective than conventional agents for treating schizophrenic patients who were either severely ill or resistant to other antipsychotic agents.

Discussions with the Food and Drug Administration (FDA) led to a US multicenter trial to determine whether clozapine was effective for treatment-refractory patients. The effectiveness of clozapine was compared with that of chlorpromazine (Thorazine) in 268 schizophrenia patients with thoroughly documented failure to respond to conventional antipsychotic agents. Because of this trial, the FDA approved clozapine in 1990 for schizophrenia patients who are resistant to treatment with other antipsychotic drugs or who are unable to tolerate conventional drugs because of extrapyramidal effects or severe tardive dyskinesia. Treatment resistance is not a well-defined concept and may include patients with distressing residual symptoms. This is important because patients with schizophrenia who improve substantially with traditional antipsychotics frequently retain positive and negative symptoms that impair their adjustment in the community. Such residual symptoms make these patients eligible for clozapine. The clinician and patient are left with the difficult task of weighing clozapine's substantial advantages against its problems, which include the risk of agranulocytosis and sudden death, seizure, adverse effects such as sedation and weight gain and weekly blood monitoring. The newer agents should probably be tried first.

Chemistry Clozapine is a five-membered heterocyclic compound, 8-chloro-11-(4-methyl-1-piperazinyl)-5 *H*-dibenzo-[*b,e*][1,4]diazepine (Fig. 31.26-1). Its chemical formula is $C_{18}H_{19}N_4Cl$, and molecular weight is 326.8. Substitutions for the A-ring have led to other clinically effective clozapine-like agents; however, loxapine (Loxitane), an antipsychotic agent with a clozapine-related molecular structure, has a clinical and adverse-effect profile similar to that of other traditional antipsychotic drugs. Recent studies suggest that loxapine may be effective in lower-than-recommended dosages and may actually be an SDA. Other clozapine-derived antipsychotic agents are olanzapine and quetiapine, but being a derivative does not mean similar efficacy.

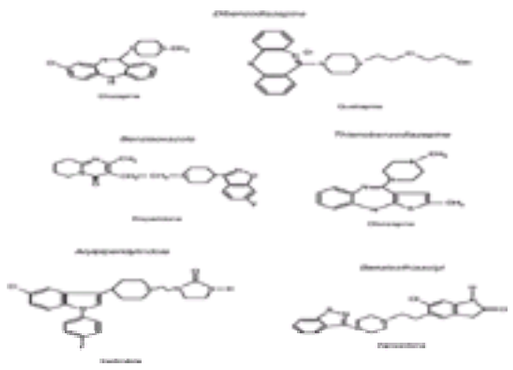

FIGURE 31.26-1 Molecular structures of serotonin-dopamine antagonists.

Pharmacological Actions

Pharmacokinetics The pharmacokinetic profiles of clozapine and other SDAs in special populations are summarized in Table 31.26-1.

Drug	Elderly	Renal Impairment	Hepatic Impairment	Ethnic
Clozapine	↓ dosage	No ↓ dosage	↓ dosage	↓ dosage women
Risperidone	↓ dosage	↓ dosage	↓ dosage	↓ dosage Asians/ CYP 2D6
Olanzapine	↓ dosage (33%)	No ↓ dosage	No ↓ dosage (preliminary)	↓ dosage women ↓ dosage Asians/
Sertindole	No ↓ dosage PD concerns	No ↓ dosage	↓ dosage (1/2)	↓ dosage women CYP 2D6 polymorphism
Quetiapine	NO PK ↓* ↓ dosage ↓* PD	No ↓ dosage	Slight ↓ dosage	No ↓ dosage
Ziprasidone	No ↓ dosage	No ↓ dosage	No ↓ dosage	?

Abbreviations: PD, pharmacodynamic; PK, pharmacokinetic; ↓, change in; CYP2D6, cytochrome P450 system 2D6 polymorphism (three drugs may delay the therapeutic response in CYP2D6, poor metabolizers. Poor metabolizers indicated by this genetic polymorphism include 4 to 10% of whites and smaller proportion of Asians and African-Americans.
*Increased adverse reactions (as thioridazine)
Modified from Hendlin Y: Pharmacokinetics and drug interactions: Update on new antipsychotics. J Clin Psychiatry 57(Suppl 11), 1996.

Table 31.26-1 Changes in Pharmacokinetics: Special Populations

ABSORPTION Clozapine is only available as an oral preparation. As with other antipsychotic agents, peak plasma concentrations are reached approximately 2 hours after oral administration (range, 1.1 ± 0.5 to 3.0 ± 1.5 hours). The elimination half-life is about 12 hours (between 10 and 16 hours; range, 6 to 33 hours). As a result, twice-daily dosing results in steady-state plasma concentrations in less than 1 week. Based on animal studies, coadministration of highly protein bound drugs may increase free clozapine concentrations, although the total concentration (free plus bound) may be unchanged.

DISTRIBUTION Clozapine has a lower volume of distribution than other antipsychotic drugs, but it is large, with a mean of 2.0 to 5.1 L/Kg (range, 1.0 to 10.2 L/Kg).

METABOLISM AND ELIMINATION Clozapine undergoes extensive first-pass metabolism in the liver and gut. Absolute bioavailability (the percentage of clozapine reaching the systemic circulation unchanged) after oral administration varies from 27 to 47 percent. Demethylation and oxidation of the terminal nitrogen of the piperazine side chain form the two main metabolites. These *N*-demethyl and *N*-oxide metabolites are cleared as fast as or faster than clozapine. Reportedly, 80 percent of administered clozapine appears in the urine or feces as metabolites. Less than 5 percent of the parent compound is found unchanged in the urine.

Pharmacodynamics

MECHANISM OF ACTION Clozapine is considered a novel, or atypical, antipsychotic drug because of its biological and clinical profile. It has a relatively low affinity for D_2 receptors and a high affinity for 5-HT_2 receptors, with a low propensity for extrapyramidal adverse effects. Preclinical studies in rats showed that clozapine blocks the conditioned avoidance response, indicating that the drug has potential antipsychotic activity. On the other hand, it did not produce catalepsy in animals, suggesting that it is unlikely to produce extrapyramidal symptoms in man. This made it different from conventional antipsychotic medications, which all cause extrapyramidal effects as a direct result of their D_2 dopamine receptor blockade in the forebrain. The relatively low affinity for D_2 receptors (Fig. 31.26-2) may explain clozapine's lack of extrapyramidal effects. Positron emission tomography (PET) scan studies have shown that with doses comparable to 10 mg of haloperidol (Haldol), more than 80 percent of striatal D_2 receptors are occupied. Higher doses do not increase receptor occupancy further. Clozapine in clinically effective dosages occupies only 40 to 50 percent of D_2 receptors. This weak effect at the D_2 receptor site may be related to clozapine's unusual biological profile that combines a low D_2 affinity with activity at a number of other receptor sites including serotonergic (e.g., 5-hydroxytryptamine [5-HT]) type 2 (5-HT_2), 5-HT_{1c}, adrenergic, and cholinergic sites.

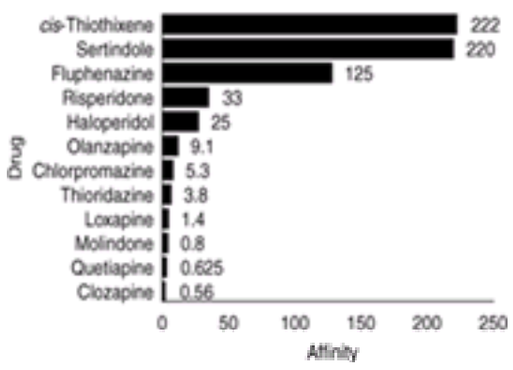

FIGURE 31.26-2 Neuroleptic and dopamine D_2 receptor blockade. Affinity is expressed as 10^7 $1/K_d$, where K_d = equilibrium dissociation in molarity. (Reprinted with permission from preclinical pharmacology of neuroleptics. J Clin Psychiatry *57*(Suppl):5, 1996.)

Several hypotheses address the mechanism of action of clozapine. Because its D_2 receptor activity is lower than that of traditional antipsychotic agents and its 5-HT_2 receptor blockade is stronger than that of haloperidol, Herbert Meltzer has proposed that the differential antipsychotic effect of clozapine is related to its low D_2 to 5-HT_2 ratio. Clozapine's affinity for the 5-HT_2 receptor is among the highest for antipsychotic agents. An alternative hypothesis proposes that clozapine is effective because it has a selective preference for the mesolimbic dopamine system. D_2-receptor blockade in the striatum by conventional agents may interfere with their antipsychotic effects and induce negative symptoms. Others have focused on clozapine's stronger 5-HT_3 blockade and its stronger a_1-adrenergic receptor and a_2-receptor blockade. Other differences include its stronger affinities for histamine type 1 H_1 receptors and for acetylcholine muscarinic receptors but lower affinity for sigma receptors than haloperidol (Table 31.26-2). All values are expressed as nanomoles of the drug needed to replace 50 percent of the radio-labeled receptor ligand (i.e., the inhibitory constant [IC_{50}]). IC_{50} values may vary between laboratories and are sometimes conflicting. Clozapine also has a 10-fold higher affinity for the

D_4 dopamine receptor than other antipsychotic agents. The D_4 receptor is present in the cortex and less so in the striatal areas. Which receptor system is responsible for the clinical difference between conventional and novel antipsychotic agents remains unknown.

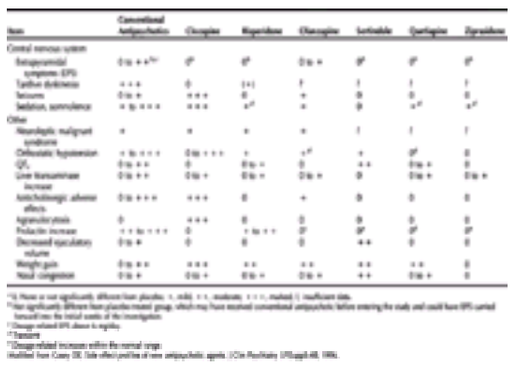

Table 31.26-2 Adverse Effects of Antipsychotic Agents[a]

BLOOD CONCENTRATIONS AND RELATION TO ACTION As with most antipsychotic drugs, patients display considerable variability in their clozapine plasma concentrations—about 45-fold among patients receiving the same daily dose. This probably results from differences in both absorption and metabolism. Variability is also observed within patients, which is presumably related to variable absorption rates rather than changes in hepatic metabolism and clearance. Some reports show that women have slightly higher plasma concentrations, and smokers slightly lower (20 to 30 percent). Older adults may have twofold higher levels than young adults.

Several investigators studied the relation between plasma concentrations of clozapine and clinical response. Their findings suggest that monitoring the clozapine concentration in plasma may be useful under certain conditions. A wide range of clozapine concentrations appears to be associated with clinically meaningful response. Plasma concentrations of clozapine average about 10 to 80 ng/mL per mg of drug given per kilogram of weight. Thus, a typical daily dosage of 300 to 400 mg (about 5 mg/kg) is associated with plasma concentrations ranging from 200 to 400 ng/mL. Studies indicate that patients are more likely to respond when their clozapine plasma concentrations exceed 350 ng/mL. If patients with a plasma concentration of 250 ng/mL have not responded by 6 weeks, the clinician should adjust the dosage to raise the concentration to approximately 350 ng/mL. The likelihood of response to concentrations above 350 ng/mL is small; therefore, patients with high concentrations who exhibit adverse effects may benefit from having their dosage reduced. The correlations between plasma concentration and clinical improvement or adverse effects are weak, as is observed with other antipsychotic agents.

Design and Interpretation of Clinical Drug Studies

Design Open clinical trials conducted in the early 1970s suggested that clozapine was an effective antipsychotic drug that lacked extrapyramidal effects. Double-blind trials then provided further evidence of this drug's unique properties; however, to date no double- blind trials comparing clozapine with a placebo have been reported. In 79 percent of the controlled trials that compared clozapine with another antipsychotic drug, clozapine was superior. In reviewing these studies Ross Baldessarini cautions that the differences between clozapine and the comparison drug were relatively small. Nevertheless, certain patterns emerge. One comparison with chlorpromazine found a small advantage favoring clozapine when a heterogeneous group of patients was studied, but differences favoring clozapine were greater when the most severely ill subgroup was studied. Another study compared haloperidol use with clozapine use in more severely ill and treatment-refractory patients and found a clear advantage of clozapine over haloperidol. One study comparing clozapine with risperidone found no differences in clinical response at the end of 6 weeks. Response to risperidone came sooner, but some extrapyramidal effects were observed with risperidone. Few treatment-refractory patients were included. In summary, these studies suggest that clozapine may be particularly effective for a subgroup of schizophrenic patients who are both severely ill and refractory to treatment with conventional drugs. Whether or not clozapine is more effective in acutely psychotic patients who are likely to respond to conventional antipsychotics is unclear. A study from New York suggests, however, that clozapine also has advantages in these patients. In a US multicenter study of patients who had tardive dyskinesia or were sensitive to extrapyramidal effects, clozapine was superior to chlorpromazine. Double-blind comparison studies with clozapine are difficult to do. The weekly blood sampling and the adverse-effect patterns make it difficult to maintain the double blind with drugs that have so few side effects. Clozapine and other SDAs tend to show their greatest advantages over traditional antipsychotic agents during the second year, when lower hospitalization rates and fewer cognitive effects are noted; however, compliance may improve when patients experience fewer adverse effects.

Efficacy in Treatment-Refractory Patients Clinical experience with clozapine suggested that it was more effective than conventional drugs for patients who were unresponsive to conventional antipsychotic agents. Before marketing clozapine for these patients, the FDA and the manufacturer of clozapine developed a study comparing clozapine with chlorpromazine plus benztropine (Cogentin). The 268 patients in this trial had previously responded poorly in at least three prior antipsychotic agent trials and a prospective single-blind trial of haloperidol. This study compared the effectiveness of chlorpromazine and clozapine in 248 patients who had thorough documentation of poor responsiveness to conventional antipsychotics. During this 6-week trial, the clozapine-treated patients did better on a range of measures of both positive and negative symptoms. Thirty percent of clozapine-treated patients met a priori criteria for significant improvement in contrast to only 4 percent of chlorpromazine-treated patients. Other studies show that this 6-week trial may have significantly underestimated the proportion of patients who respond to clozapine and that up to 60 percent of patients may display substantial improvement following a 6-month trial. Clozapine in this trial was superior to chlorpromazine on 14 of 18 Brief Psychiatric Rating scale items, as well as on the Clinical Global Impression and such nursing ratings as Social Competence, Social Interest, Personal Neatness, Irritability, Manifest Psychosis, and Motor Retardation. When patients were classified as having improved (by prospective criteria) to a clinically significant extent, 30 percent of clozapine patients improved within 6 weeks versus only 4 percent of those receiving chlorpromazine and benztropine ($p < .001$). Based on this trial, the FDA approved clozapine for limited use in the United States.

Long-Term Effectiveness All of the long-term studies with clozapine have been uncontrolled or retrospective. Nevertheless, they provide further support for the unique properties of this drug. For example, patients who received clozapine for up to 13 years were more likely to be discharged and remain out of the hospital. In a European study, patients who received clozapine for at least 2 years increased their rate of employment from 3 to 40 percent. However, none of the reported long-term studies included random assignment and controlled comparisons with conventional treatment or other SDA treatment. Thus the impact of clozapine on the long-term outcome of patients—one of the most compelling questions in the therapeutics of schizophrenia—requires more study. In the meantime Novartis has started a 2-year evaluation of clozapine versus olanzapine in effectiveness against suicidality.

Therapeutic Indications The second author (Steven Marder) and Ted Van Putten suggested before the arrival of other SDAs that three populations of patients with schizophrenia were obvious candidates for clozapine: (1) patients with severe schizophrenic symptoms that respond poorly to conventional antipsychotic medications; (2) patients with severe tardive dyskinesias; and (3) patients who suffer from severe extrapyramidal effects at doses of conventional antipsychotic drugs that are necessary for treating their symptoms. Because the concept of treatment resistance will change with the newer SDAs, the authors suggest that groups 1 and 2 should be replaced by patients who have not responded to other SDAs. An off-label indication may be severe suicidality.

Treatment-Resistant Patients Patients with severe positive symptoms despite treatment with adequate dosages of antipsychotic agents are probably the largest group of candidates for clozapine treatment. These patients may be inpatients or outpatients who are suffering from continuous psychotic symptoms. The findings from John Kane and coworkers' multicenter trial indicate that about one-third of these patients will show significant improvement at 6 weeks, and 60 percent will at 5 to 6 months. In the authors' experience these patients show a range of improvements. Some recover from all or nearly all of their positive symptoms. These patients are then challenging candidates for psychosocial rehabilitation programs. Others demonstrate substantial improvement in psychotic symptoms but still suffer lingering schizophrenic experiences that resist even clozapine treatment. For these patients too, clozapine is a better antipsychotic agent, with fewer adverse effects. Others fail to show substantial improvement with clozapine and may prefer the adverse effects of conventional drugs to those of clozapine and its required participation in the blood-monitoring system.

Severe Tardive Dyskinesia Patients with severe tardive dyskinesias should also be considered for a clozapine trial. Most clinical reports indicate that clozapine can suppress abnormal movements in tardive dyskinesia. This is not surprising since conventional antipsychotic agents are effective in short-term suppression of tardive dyskinesia. However, clozapine is probably less likely to cause the disorder and therefore is promising for the long-term management of patients, particularly those who already have tardive dyskinesia. Although rare cases of tardive dyskinesia have been found in patients who receive clozapine, the risk of developing tardive

dyskinesia is substantially reduced. Studies with animal models indicate that clozapine does not evoke the increased sensitivity of striatal D_2 receptors that is associated with typical antipsychotic agents, which has been theorized to result in tardive dyskinesia. It remains to be established whether a positive effect on tardive dyskinesia results from a direct effect of clozapine on certain transmitter systems or whether clozapine spares the striatal D_2 receptors, which allows spontaneous remission of tardive dyskinesia. The former interpretation is supported by the reports of a return of tardive dyskinetic movements following clozapine withdrawal. None of the SDAs has shown evidence of inducing tardive dyskinesia.

Patients With a Low Extrapyramidal Adverse Effect Threshold Another group of schizophrenic patients who may be reasonable candidates for clozapine are those who develop severe extrapyramidal adverse effects—particularly, severe akathisia—when they receive relatively low doses of conventional antipsychotic drugs. The benefits of antipsychotic treatment are often only obtained at the price of severe discomfort and dysphoria that is not adequately responsive to antiparkinson medications or propranolol (Inderal). The incidence of akathisia is considerably lower with all SDAs.

Schizoaffective Disorder, Treatment-Resistant Mania, and Severe Psychotic Depression A number of early clinical trials of clozapine that included patients with schizoaffective disorder, treatment resistant mania, and severe psychotic depression suggested that clozapine is effective. Several reports suggest that clozapine is effective in treatment-resistant mania, and a consensus is developing that mania responds well to clozapine. A recent trial in patients with psychotic depression indicates that clozapine is effective with those patients as well.

Neurological Illnesses Clozapine has also found uses in patients with neurological disorders. It has been used in idiopathic Parkinson's disease to treat psychotic symptoms that are secondary to levodopa (Larodopa). These patients may respond to relatively low daily dosages of clozapine (25 to 75 mg). Higher dosages (e.g., 100 to 250 mg a day) may exacerbate the parkinsonian symptoms. A case report described a patient with both schizophrenia and Parkinson's disease who was successfully treated with levodopa and 300 mg daily of clozapine. Other reports suggest that clozapine may be useful for tremors in Parkinson's disease, alcohol-related disorders, and benign essential tremor. Patients with Huntington's disease have also responded positively to clozapine. More studies are needed to clarify the role of clozapine in the treatment of neurological disorders. When experience with risperidone, olanzapine, quetiapine, sertindole, and ziprasidone increases, those drugs might become the preferred agents because of their more benign adverse-effect profile.

Treatment Resistance in Other Disorders Treatment-resistant patients with a pervasive developmental disorder, autistic disorder, or obsessive-compulsive disorder have responded well to clozapine, either by itself or in addition to their regular treatment. However, the data on obsessive-compulsive symptoms are somewhat in doubt because some patients may worsen, particularly with concomitant use of selective serotonin reuptake inhibitors (SSRIs). Children with these severe psychiatric illnesses may respond well also.

Severely Suicidal Patients With Schizophrenia or Schizoaffective Disorder Patients who are severely suicidal or agitated with schizophrenia, affective disorders, mental retardation, or dementia become candidates for clozapine treatment. Schizophrenia carries a 9 to 13 percent lifetime risk of suicide, with a projected 0.4 to 1 percent per year. In a 1995 study, the authors reported that suicidality decreased markedly with reduction in depression and hopelessness. Treatment resistance was unrelated to the decrease in suicides or suicidality. Novartis is presently conducting a study in schizophrenia patients with suicidal ideation.

Precautions and Adverse Reactions Adverse effects of clozapine and the other SDAs are compared in Table 31.26-2. The only contraindications to the use of clozapine are a white cell count below 3500, a previous bone marrow disorder, a history of agranulocytosis during clozapine treatment, or the concomitant use of another bone marrow suppressant drug such as carbamazepine (Tegretol). When counts are low to begin with, it will be difficult to determine whether leukocytopenia or agranulocytosis has been induced. Although the mechanism of action is unknown, a slight risk of respiratory depression or collapse may occur if treatment is initiated while patients are taking benzodiazepines.

Leukocytopenia and Agranulocytosis One to 2 percent of patients receiving clozapine develop granulocytopenia or agranulocytosis; if not for this adverse effect, clozapine could be a first- or second-line antipsychotic. These patients may have a sudden or a gradual drop in the white blood count. The risk for this adverse effect is greatest during the first 3 months of treatment, but a significant risk remains during the first year of treatment and perhaps indefinitely. Agranulocytosis due to clozapine is a potentially fatal condition that requires immediate medical attention. The mortality rate is 1 less than in 10,000. The risk for clozapine-induced agranulocytosis is 0.73 percent for the first year of treatment and 0.07 percent in the second year. Review of the clozapine register indicates that the incidence may be as low as 0.038 percent. Peak incidence occurs between 6 and 18 weeks. For neutropenia the risk is 2.32 percent in the first year and 0.69 percent in the second year. Risk of agranulocytosis increases with age and is higher in females. Therefore, before treating patients with clozapine, clinicians were required to register patients with the Sandoz Company (Novartis). Recently the FDA approved a generic form of clozapine for marketing in the United States and registration is no longer required. Until recently, clozapine-treated patients had to receive weekly monitoring of their white blood cell (WBC) counts for as long as they received the drug and for at least a month after discontinuation. Since early 1998 the weekly blood sampling requirement has been every other week after 6 months.

Current guidelines specify that any fever or sign of infection (e.g., pharyngitis) is an immediate indication for a WBC count, particularly in the first 18 weeks of treatment. If the patient has a WBC count below 2000 or a granulocyte count below 1000, clozapine must be discontinued. Reports of clozapine reexposure following hematological recovery have shown that all such patients reexperience agranulocytosis but sooner and at lower doses than before. Thus patients who develop agranulocytosis should not receive clozapine again. Although WBC count below 3500 cells per mm^3 is not a proved risk factor, clozapine should be administered cautiously to patients with such low counts because of the difficulty in early identification of agranulocytosis. Similarly, carbamazepine should not be given concomitantly with clozapine. Clozapine therapy should be discontinued whenever the total WBC count falls below 3000 cells per mm^3 or the granulocyte count below 1500 cells per mm^3. Other hematological changes have been reported with clozapine, including leukocytosis (0.6 percent), eosinophilia (1 percent), leukopenia, neutropenia, and decreased WBC count (3 percent), and rarely, thrombocytopenia.

Several hypotheses about this toxicity have been proposed. A metabolic change might lead to the production of toxic metabolites, possibly higher than usual *N*-demethylclozapine concentrations. An alternative hypothesis proposes that neutrophils and their committed stem cell precursors can metabolize clozapine into free radicals, which are cytotoxic. This has led to suggestions that vitamins E or C and other antioxidants or trace elements that are cofactors of free-radical scavenger enzymes (e.g., selenium, zinc, and copper) may be preventive.

Sialorrhea Sialorrhea is an uncomfortable but tolerable adverse effect that develops early in treatment. Hypersalivation is most profound during sleep, and patients may complain that their pillow is wet in the morning. Placing a towel over the pillow may be of some help. Although the symptom can be treated with anticholinergics, this is not recommended because of the risk of anticholinergic toxicity. It has been proposed that changes in the peripheral adrenergic tone and substance P may override the muscarinic effects of clozapine on the salivary glands. Clonidine (Catapres) (0.1- or 0.2-mg patch once a week) and amitriptyline (Elavil, Endep) have been used at bedtime to treat sialorrhea. Recent studies suggest that clozapine does not increase saliva flow, it decreases swallowing, which becomes more obvious during sleep.

Cardiovascular System The cardiovascular adverse effects most frequently observed are tachycardia and postural hypotension. The tachycardia is probably a direct effect of the vagolytic properties of the drug. It can be present in the supine position and therefore does not result from orthostatic changes. Increased pulse rates of 20 to 25 beats a minute are encountered when clozapine reaches 300 mg a day or more over a 7-day period (e.g., sinus tachycardia). Reversible nonspecific ST-T wave changes, T wave flattening, or inversions (repolarization effects) are seen infrequently but are usually of no clinical significance. These changes are similar to those with other antipsychotics and are dose dependent. Although some tolerance occurs, it will persist unless the dose is lowered. A b adrenergic receptor antagonist (beta-blocker) such as atenolol (Tenormin) but not propranolol (which may increase the risk of agranulocytosis), may be given if the blood pressure allows it.

The risk of postural hypotension progressing to orthostatic collapse is high, if the initial dose exceeds 75 mg a day. Tolerance to the hypotensive effects of clozapine often develops over time. It can usually be managed if the initial dosage is low (e.g., 25 mg a day), and the dose is increased gradually. When it occurs, lowering the dosage to the previous level and slowing titration will limit the problems. Support stockings, increased sodium intake, or fludrocortisone (Florinef) may be helpful. However, orthostatic hypotension with or without syncope can occur with clozapine treatment and may represent a continuing risk in some patients. Rarely (approximately 1 case per 3000 or 4000 patients), collapse can be profound and be accompanied by respiratory arrest, cardiac arrest, or both. Orthostatic hypotension is more likely to occur during initial titration in association with rapid dose escalation and may even occur on the first dose. In one report, initial doses as low as 12.5 mg were associated with collapse and respiratory arrest. When resuming administration to patients who have had even a brief interval within clozapine (i.e., 36 hours or more since the last dose), the authors recommend that treatment be reinitiated at 12.5 mg once or twice daily.

Paradoxically, hypertension has been observed as well (<4 percent). Clozapine induces significant changes in hemodynamic mechanisms; it should be used with caution in patients with preexisting cardiovascular disease, such as a history of myocardial infarction or arrhythmias. The authors recommend careful monitoring of

vital signs during the first few weeks of treatment.

Peripheral Anticholinergic Effects Dry mouth, blurred vision, constipation, and urinary retention are commonly observed with clozapine. Other adverse effects on the autonomic nervous system, such as increased sweating, have also been reported.

Disturbances in Temperature Regulation Occasionally, benign hyperthermia is observed during the first 3 weeks of treatment, with a peak incidence on the 10th day. The increase in temperature is usually not more than 1 to 2°F, which spontaneously resolves over a few days with continued treatment without any clinical significance. Occasionally, however, temperature readings above 101°F have been seen, which require temporary cessation of clozapine treatment and frequent hematological monitoring. Differential diagnoses include drug fever, intercurrent infection, infection secondary to agranulocytosis, dehydration, heat stroke, lethal catatonia, and possibly neuroleptic malignant syndrome. Patients who develop a high temperature can be given the drug again but with a more gradual increase in dosage. If this strategy does not work, clozapine treatment may need to be discontinued. Mild hypothermia is frequently observed (87 percent). Because this effect is observed as frequently as with chlorpromazine, it may indicate the poikilothermic effect of antipsychotic agents (see chapter on antipsychotic agents). It is presumably mediated by the hypothalamus and it might be treated with benzodiazepines or 500 µg of thyrotropin-releasing hormone, according to Garth Bissette.

Weight Changes Weight gain has been reported with clozapine as with most SDAs. The incidence may be higher because of its strong 5-HT $_1$ and histamine type 1 (H_1) receptor affinity. With the improved quality of life often associated with clozapine treatment, weight gain becomes more of a rehabilitation issue.

Gastrointestinal Disturbances Constipation, most probably due to the antimuscarinic effects of the drug, is the most common gastrointestinal adverse effect (up to 33 percent). It should be treated with stool softeners, laxatives, fiber supplements, and adequate fluid intake. The constipation can progress into intestinal obstruction if not appropriately treated. Nausea and less frequent vomiting might be managed by dosage reduction.

Liver Function Disturbances Liver function disturbances are mild and transient. Only one case of cholestatic jaundice induced by clozapine has been reported.

Urogenital Effects Urogenital effects include enuresis, frequency and urgency, hesitancy, urinary retention, and impotence. All of these are independent of dose or blood concentrations. Urinary incontinence or enuresis can be managed by setting an alarm clock to ring at night, intranasal desmopressin (DDAVP) or oxybutynin (Ditrapon) (5 to 15 mg a day). So far, a few cases of clozapine treatment–associated priapism have been noted. Because patients may not volunteer this complaint, clinicians should ask about it because it may lead to noncompliance or medical complications and affect the patients' quality of life. Prolactin concentrations are less likely to increase with clozapine than with other antipsychotic agents, and effects of hyperprolactinemia such as amenorrhea, galactorrhea, or gynecomastia have not been encountered.

Extrapyramidal Adverse Effects Clozapine has a much lower incidence of extrapyramidal adverse effects than other antipsychotic agents. Akathisia (6 percent), tremors (6 percent), and rigidity (3 percent) are considerably less frequent than with traditional antipsychotic agents. The authors are not aware of reports of acute dystonia with clozapine. Currently, there are no confirmed cases of tardive dyskinesia with clozapine; moreover, tardive dyskinesia improved with clozapine. Patients with tardive dyskinesia may benefit from a treatment trial.

Neuroleptic Malignant Syndrome Consistent with clozapine's low affinity for D_2 receptors, there have been few reports of typical neuroleptic malignant syndrome developing in patients treated solely with clozapine. A few case reports exist of patients developing this syndrome who were concomitantly treated with lithium (Eskalith). Although two patients had experienced neuroleptic malignant syndrome with conventional antipsychotic agents before, one case has been reported of a patient with a history of neuroleptic malignant syndrome who was successfully treated with clozapine. The authors recommend, therefore, that lithium not be used as adjunctive treatment with clozapine in patients who have experienced neuroleptic malignant syndrome. Because clozapine treatment is accompanied by adverse effects that resemble several signs and symptoms of neuroleptic malignant syndrome (e.g., hyperthermia, cardiovascular effects, delirium, increased sweating, elevations in creatine phosphokinase, and leukocytosis), clinicians should be alert for the development of a potential neuroleptic malignant syndrome during clozapine treatment. Some case reports have described neuroleptic malignant syndrome without rigidity in patients treated with clozapine, although the lack of rigidity could preclude the diagnosis of this syndrome.

Lowering of Seizure Threshold Like treatment with other antipsychotics, clozapine treatment carries the risk of encephalographic (EEG) changes and grand mal seizures that are dose dependent. The initial data showed that the risk of seizures increases considerably with doses exceeding 500 mg. The cumulative incidence at 1 year for dosages up to 300 mg a day is 1 to 2 percent; 300 to 600 mg a day, 3 to 4 percent; and 600 to 900 mg a day, 5 percent. A preexisting seizure disorder or head trauma places the patient at even greater risk. More-recent data from a 6-month period suggest that the crude rate of seizures is 1 to 3 percent. The real risk is probably much lower when the drug concentration in blood is kept within the recommended range and no preexisting seizure disorder exists.

Several recommendations have been made to lower the risk by 50 percent; they include monitoring clozapine concentrations in blood, an EEG before raising the dosage over 600 mg a day; combining clozapine with anticonvulsants at dosages where a seizure occurred before, and lowering the dosage after a seizure has occurred, followed by a neurological consultation. Looking for other etiologies than clozapine is important as well as avoiding combination treatment with other drugs that lower the seizure threshold. Patients who are on antiepileptic agents may need dosage adjustment according to metabolic changes induced by clozapine administration. Carbamazepine should not be combined with clozapine because of its risk of granulocytopenia. Patients who are taking carbamazepine and are candidates for clozapine should preferably be given another anticonvulsant. If anticonvulsants are required, the effects on metabolism should be included in the decision about the dosage of both clozapine and the anticonvulsant.

Other CNS Effects Sedation is the most common adverse effect of clozapine treatment. It occurs early in treatment and tolerance usually develops after the first few days and weeks. However, if the patient does not take other CNS depressants and takes most of the daily dosage at bedtime, daytime sedation may be avoided or minimized. The a_1- and H_1-receptor blocking effects of clozapine are believed to be the cause. Other CNS effects include dizziness and syncope, which may be related to orthostatic hypotension. Confusion and delirium may occur because of muscarinic toxicity.

Subtle myoclonic movements and (rarely) cataleptic-like events in which the patient experiences a sudden loss of muscle tone in a part of the body without loss of consciousness have been reported and may occur more frequently than is reported in the literature. Lowering the dosage may deal with this. Whether clozapine induces or worsens a restless limb syndrome is uncertain.

Pregnancy and Lactation No data on risks have been reported in pregnant or lactating women to date.

Drug Interactions Because the cytochrome P450 mitochondrial system is involved in the metabolism of clozapine, ethnic differences and interactions with drugs that inhibit or stimulate this system will affect clozapine concentrations in blood. For instance, caffeine N3-demethylation, which is a measure of cytochrome P450 (CYP) isoenzyme 1A2 (CYP 1A2) activity, accounts for 70 percent of the variance in clozapine clearance. This suggests that fluvoxamine (Luvox), a potent inhibitor of CYP 1A2, will elevate clozapine concentrations, while fluoxetine (Prozac) may only lead to changes in higher dosage ranges. Cimetidine (Tagamet), SSRIs, tricyclic drugs, and valproate (Depakene) decrease clearance, presumably through inhibition of CYP 1A2 and 2D6, while phenytoin (Dilantin) and carbamazepine induce CYP 2C and 3A4 hepatic isoenzymes and may decrease concentrations of clozapine (£50 percent). Furthermore, when risperidone, which weakly inhibits CYP 2D6, is added, clozapine levels may rise also. Benzodiazepines do not affect kinetics but interact rarely with clozapine; a few case reports of delirium, increased somnolence, and acute respiratory suppression have appeared in the literature. Erythromycin and ketoconazole (Nizoral) may increase clozapine concentrations.

Dosage and Administration Clozapine treatment should begin with 12.5 to 25 mg during the day and be increased by 25 mg a day every 2 to 3 days. The rate of increase is limited by sedation and orthostatic hypotension. No dosage-response studies have been reported, but daily dosages between 250 and 450 mg are usually considered adequate; daily dosages above 600 mg are seldom indicated. In spite of its low relative potency in many animal studies, clozapine is 1 1/2 to 2 times more potent than chlorpromazine. Except for seizure risk, and possibly early sedation and hypotension, neither clinical nor adverse effects are likely to increase in severity with the dosage. Traditional antipsychotic agents can be discontinued overnight and replaced by clozapine. The authors recommend, however, that the traditional antipsychotic agent be discontinued gradually, until the clozapine dosage is increased to the desired level.

Once the appropriate dosage has been reached and the patient improves, the medication should be continued, probably indefinitely. When drug administration is abruptly discontinued in a successfully treated patient, a rebound worsening is sometimes observed, requiring dosage tapering and an early addition of the next antipsychotic agent. The rebound psychosis may be due to muscarinic supersensitivity in the changeover. An antipsychotic agent that has lower muscarinic blockade than clozapine may be perceived as less effective. The authors recommend that clozapine dosage be tapered gradually and anticholinergic agents added while

administration of the other drug is begun. Available preparations of clozapine and other SDAs are listed in Table 31.26-3.

	Tablets	Solution
Clozapine	25, 100 mg	
Olanzapine	2.5, 5, 7.5, 10 mg	
Quetiapine	25, 100, 200 mg	
Risperidone	1, 2, 3, 4 mg	1 mg/1 mL

Table 31.26-3 SDA Drug Preparations

RISPERIDONE

Results from clinical studies indicate that risperidone is an effective antipsychotic medication with a relatively mild adverse effect profile. In contrast to high-potency conventional antipsychotic agents, risperidone's therapeutic efficacy appears at lower dosages than the extrapyramidal effects, which is dose dependent. Hematological effects have not been observed. Risperidone appears to have a better antipsychotic effect than haloperidol and in all likelihood is also better for the treatment of negative symptoms. Risperidone has not been compared with low-potency antipsychotic agents such as chlorpromazine or thioridazine (Mellaril). Nevertheless, it is less sedating and does not have adverse anticholinergic effects. These particular characteristics indicate that risperidone is an appropriate first-line drug for patients with schizophrenia and other psychotic disorders when an antipsychotic agent is indicated. An early study by Eli Lilly comparing olanzapine with risperidone suggested that olanzapine was more effective; however, the risperidone dosage used was most likely too high for a good evaluation. A recent study by Janssen Research Foundation reported the opposite: risperidone was more effective than olanzapine. Many patients respond optimally to a lower dosage than the recommended 6 mg a day. Risperidone is available as a liquid. Depot formulations are expected to be developed (Phase III). Patients who are likely to need short-acting, injectable drugs may require a drug that is available for other routes of administration.

History Risperidone was the first antipsychotic agent to gain FDA approval after clozapine. Risperidone was discovered, patented, and marketed by the same company that developed haloperidol, pimozide (Orap), penfluridol (Semap), and other antipsychotics.

Chemistry Risperidone is a benzisoxizole derivative (see Fig. 31.26-1).

Pharmacological Action

Pharmacokinetics

ABSORPTION Risperidone is metabolized in the liver to 9-hydroxyrisperidone, which has the same pharmacological profile as the parent compound. After ingestion, peak plasma levels of the parent compound occur within 1 hour and within 3 hours for 9-OH-risperidone. Food does not affect the rate or extent of absorption in the gut.

METABOLISM AND ELIMINATION Risperidone has a bioavailability of 70 percent. The hepatic enzyme system that metabolizes risperidone to the hydroxy metabolite is CYP2D6, or debrisoquin hydroxylase. A genetic polymorphism is present in 7 percent of white people, leading to almost no activity in this enzyme. A much lower incidence has been seen in Asians (Table 31.26-1). Because the metabolite has an activity similar to that of the parent compound, this variance is not clinically significant.

Pharmacodynamics

MECHANISM OF ACTION Potent central antagonism of both serotonin (particularly 5-HT_{2A}) and D_2 receptors characterizes risperidone's presumed mechanism of action. Risperidone also demonstrates high affinity for a_1- and a_2-receptors but low affinity for b-receptors or muscarinic receptors. Preclinical studies indicate that while it is as potent as haloperidol in D_2 antagonism, risperidone is several times less potent than haloperidol at inducing catalepsy. This is consistent with risperidone having substantially less extrapyramidal adverse effect than conventional drugs at the recommended doses. In addition, the activity at 5-HT_2 receptors may result in greater activity against negative schizophrenic symptoms. PET studies indicated that dosages of 1 to 4 mg a day provide the required D_2 blockade for a therapeutic effect.

BLOOD CONCENTRATIONS AND CLINICAL ACTIVITY No clear evidence exists of a relation between clinical efficacy and concentration of risperidone in blood. This may well be because the dose-response curve is an inverted U.

Design and Interpretation of Clinical Drug Studies

Early Clinical Trials Several double-blind trials have compared risperidone with other drugs and placebo. In addition to providing evidence about the effectiveness of risperidone compared with that of other drugs, a number of studies have also focused on the dosage of risperidone by comparing several fixed dosages.

VARIABLE-DOSAGE STUDIES Variable-dosage studies most closely duplicate a real-life clinical setting in which clinicians individualize treatment. The first published double-blind study of risperidone was a multicenter investigation from Belgium. Forty-four patients with relatively refractory forms of schizophrenia were assigned to risperidone (mean daily dosage, 12 mg) or haloperidol (mean daily dosage, 10 mg). Although the two drugs did not differ significantly in their effectiveness for positive or negative symptoms, risperidone tended to be more effective. Risperidone patients were less likely to require antiparkinson medication.

A double-blind study that lasted 8 weeks (N = 107 schizophrenia patients) compared risperidone (mean dosage, 8.5 mg daily) and perphenazine (Trilafon) (mean, 28 mg). Although a higher proportion of patients assigned to risperidone (74 percent) than perphenazine (59 percent) met improvement criteria for total score on the Positive and Negative Symptom Scale (PANSS), the difference was not statistically significant. For positive symptoms, 69 percent of risperidone and 73 percent of perphenazine patients met improvement criteria (not significant); for negative symptoms, 76 percent of risperidone and 53 percent of perphenazine patients met criteria ($P < .05$). Extrapyramidal adverse effects did not differ between the two groups.

FIXED-DOSAGE COMPARISONS Fixed-dosage studies permit comparison of different dosages and they eliminate the effect of differences in prescribing habits among clinicians. The largest risperidone study was a multinational study that included 1362 patients who were randomized to a double-blind comparison of 1, 4, 8, 12, or 16 mg daily of risperidone or 10 mg of haloperidol. The 1-mg risperidone daily dosage was considered subtherapeutic and was therefore used as a control. All of the other risperidone doses and haloperidol resulted in greater improvement on the total PANSS than 1 mg of risperidone. The 4- and 8-mg risperidone daily dosages were the most effective, although their advantages were not statistically significant. A clear dosage-related increase in extrapyramidal adverse effects occurred with risperidone. However, all risperidone dosages were associated with less extrapyramidal effect than haloperidol. The differences in extrapyramidal adverse effects were statistically significant for all risperidone dosages except 16 mg a day.

The most important placebo-controlled trial was carried out in Canada and the United States. Acutely ill patients with schizophrenia were randomly assigned to groups that received 2, 6, 12, or 16 mg of risperidone, 20 mg of haloperidol, or placebo. The results from the United States and Canadian studies were very similar. Risperidone at daily dosages above 2 mg and haloperidol at daily dosages of 20 mg were consistently more effective than placebo for total PANSS and positive symptoms. The most effective daily dosage of risperidone was 6 mg, which was significantly more effective than haloperidol in both total PANSS and positive

symptoms. Risperidone (6 and 16 mg) but not haloperidol resulted in significant improvements in negative symptoms as measured by the PANSS negative symptoms scale. Similar results were obtained when the proportion of patients meeting criteria for improvement (20 percent or greater improvement in total PANSS or Brief Psychiatric Rating Scale [BPRS] were considered.

Risperidone resulted in a dose-related increase in extrapyramidal adverse effects. However, the proportion of patients who required antiparkinson medications was no greater at the 6-mg daily dosage—the dose at which risperidone was most effective—than on placebo. This observation suggests that patients who are treated with an effective dosage of risperidone are less likely to suffer from extrapyramidal effects than patients on high-potency drugs such as haloperidol if they are treated with dosages of 2 to 4 mg a day. In spite of early reported studies, 2 to 4 mg a day is the prefered effective dosage range.

Efficacy in Treatment-Refractory Patients Risperidone is an appropriate choice for individuals who have failed to respond to a conventional antipsychotic. In the U.S. study patients who had been in the hospital for at least 6 months prior to the study tended to show the greatest advantage for risperidone over haloperidol. This suggests that risperidone may be effective for some patients with treatment-refractory illnesses or who respond poorly to traditional agents.

COMPARISONS WITH CLOZAPINE Two published studies have compared risperidone with clozapine. In the first study, carried out in Germany, 59 acutely ill patients received 4 mg of risperidone, 8 mg of risperidone, or 400 mg of clozapine daily for 28 days. All three treatment groups improved, and there were no significant differences among the groups. The 4-mg risperidone daily dosage was the best tolerated when adverse effects were measured. However, subjects in this study were not selected for being resistant to antipsychotic medications.

Another European study compared risperidone and clozapine in treatment-refractory schizophrenic patients with a mean of six previous hospitalizations. Overall no significant difference was observed at the end of the trial. Daily dosages of 300 and 400 mg of clozapine were used compared with 4, 6, or 8 mg a day of risperidone, and at least for these dosages of clozapine, risperidone was as effective as clozapine. Patients receiving clozapine experienced more asthenia and lethargy and increased fatigability.

Replacing clozapine with risperidone does not always prevent a relapse, even when dosages are slowly tapered and given concomitantly. The authors recommend that anticholinergic agents accompany the replacement of clozapine with risperidone. Because of risperidone's lack of associated anticholinergic effects, this treatment should continue for some time after clozapine treatment has been stopped, to prevent symptoms associated with a cholinergic rebound.

Long-Term Effectiveness Evidence indicates that the effectiveness of risperidone does not decrease over time. Initial pharmacoeconomic studies indicate that risperidone may have a better record of preventing relapse and decreasing hospitalizations. Whether this is due to better compliance has not been determined.

Overview of Efficacy Taken together, these studies indicate that risperidone is an effective antipsychotic that may have advantages over conventional antipsychotics in its adverse-effect profile and in its efficacy. Large multicenter studies are most likely to detect differences in efficacy between risperidone and a comparison drug. However, both the 15-nation study and the North American trial compared multiple daily doses of risperidone with a single daily dose of a comparison drug. This type of comparison is biased for obvious reasons. However, John Davis recently analyzed pooled data from five controlled trials. The results of the meta-analysis indicated that 53 percent of risperidone patients met a priori criteria for improvement when their daily dosage exceeded 6 mg, compared with 40 percent of individuals who received a conventional antipsychotic ($P < .001$). Davis also reported that risperidone was 25 percent better for positive symptoms and 60 percent better for negative symptoms than traditional agents.

Treatment Indications

Acute Psychosis Risperidone is effective for both first and later episodes of psychosis in schizophrenia and schizoaffective disorder. There are some suggestions that risperidone is more effective than conventional antipsychotic agents, and it is certainly associated with fewer adverse effects. Other off-label use includes psychotic depression and perhaps autistic disorder. Risperidone may also have antimanic efficacy in bipolar disorder patients in add-on and monotherapy.

Maintenance Treatment in Schizophrenia and Schizoaffective Disorder Once the patient is stabilized, the dosage may be lowered. Although no long-term controlled studies of maintenance risperidone in schizophrenia or schizoaffective disorders have appeared, clinical experience indicates that it is effective in preventing relapse. Depot preparations are being studied.

Tardive Dyskinesia Case reports suggest that tardive dyskinesia may improve with risperidone. Because the disorder waxes and wanes, systematic study is needed before any conclusions can be drawn. Reports from Janssen Pharmaceutica document two cases of tardive dyskinesia in 1156 patients (0.2 percent) and a 0.4 percent risk in those treated for over 1 year. Patients treated with traditional agents have a 5 percent risk per year. A recent study in elderly agitated patients showed that risperidone was safe and efficacious, and there was a lower incidence of tardive dyskinesia than was expected in this age group. Dosages were 0.25 to 1 mg a day.

Patients With a Low Extrapyramidal Adverse Effect Threshold Patients with a low extrapyramidal effect threshold may respond to low dosages of risperidone without extrapyramidal effects.

Other Diagnoses The exploration of the clinical efficacy of risperidone has focused on schizophrenia. Controlled studies for other indications as antimanic and mood-stabilizing effects are awaited. In open studies, risperidone has been successful in patients with schizoaffective disorder, mania, psychotic depression, treatment-resistant depression, autistic disorder, and psychotic disorder due to brain trauma (i.e., hemorrhage), as well as elderly schizophrenic patients, those with acquired immune deficiency syndrome (AIDS)-related dementia and psychosis or with agitation, and psychotic symptoms in patients with dementia or substance-induced psychotic disorder. Some bipolar disorder patients with resistant mania or hypomania have responded well to low-dosage risperidone. In general these patients tend to respond to dosages below 6 mg a day (e.g., 1 to 3 mg). However, dosage increases lead to increased activation in some. Open studies indicate that elderly patients may require lower dosages (0.25 to 3 mg a day). Other case reports indicate that tics in Tourette's disorder, behavior disturbances in Huntington's disease, and levodopa–induced hallucinations may respond favorably as well. One patient with Lesch-Nyhan syndrome showed fewer episodes of self-mutilation during treatment with risperidone. Early reports of effects in children and adults with conduct disorders are encouraging in dosages between 0.02 and 0.06 mg/kg a day.

Precautions and Adverse Reactions Double-blind trials indicate that risperidone treatment is associated with a dose-dependent increase in extrapyramidal adverse effects. Although most adult patients may not experience these effects at the recommended daily dosage of 6 mg, dystonic reactions have occurred with dosages between 4 and 16 mg a day. Akathisia and occasional occurrences of tardive dyskinesia have been described. Most patients require only 2 to 4 mg a day. Risperidone induces a substantial rise in plasma prolactin concentration, a direct effect of all drugs with D_2 activity. This increase could lead to galactorrhea and menstrual disturbances in some women and sexual dysfunction in some men, but no such effects have been reported. Other common adverse effects include sedation, dizziness, constipation, tachycardia, and weight gain. A few cases of neuroleptic malignant syndrome have been reported on risperidone, which is not surprising given this drug's antidopaminergic activity. Other reported cases indicate that risperidone can produce obsessive-compulsive symptoms in some patients, an observation that has been made with clozapine and olanzapine, other drugs with substantial activity at 5-HT_{2A} receptors. One case of risperidone-induced angioedema has been reported. However, many incidents of adverse effects can be prevented if the daily dosage is kept below 5 mg.

Overdose Although antipsychotic drugs are relatively safe as a class, one fatal overdose with risperidone in a 45-year old man with schizophrenia has been reported. He ingested several hundred 1-mg tablets. At autopsy he had blood concentration of risperidone 500 times the normal therapeutic range. A 29-year-old man was brought to the emergency room 45 minutes after ingestion of 245 mg of risperidone. On examination he exhibited a cardiac conduction disturbance and an electrolyte imbalance, but he recovered fully.

Dosage and Administration

Initiation of Treatment The authors suggest a 1-mg starting dose, with 1-mg increments in twice-daily doses for the first few days. Most patients are then evaluated on 4 to 6 mg for the next 2 to 3 weeks. Although 6 mg a day may be the optimal dosage as determined by the phase III studies, new data indicate that many patients may be optimally treated with a 2- to 4-mg daily dosage. Lorazepam (Ativan) may be added during dose titration if the patient shows agitation in the early phases of treatment. Most of the improvement will occur in the first 8 weeks. If the response is not adequate, an increase to 8 mg may be indicated. The response is faster than with haloperidol; however, an 8-week trial should give an adequate assessment. Risperidone can be given once daily, which is as effective as twice-daily dosing.

Elderly and Parkinson's disease patients may need doses of 1 mg or less to prevent other side effects.

Duration of Treatment Until studies show otherwise, the rules used for traditional antipsychotic agents apply.

Drug Interactions At least two types of interactions are possible. Inhibition of CYP 2D6 by fluoxetine and paroxetine (Paxil) should block conversion to the hydroxy metabolite. Inversely, inducers of this enzyme such as carbamazepine should enhance the conversion and may require higher dosages. Raising the concentration of risperidone in blood may increase the risk of extrapyramidal effects and decrease the drug's efficacy. Because the metabolite and parent compound have the same pharmacological effects, such interactions may not be important. Risperidone is only a weak inhibitor of the enzyme and should not significantly decrease the clearance of other drugs. Clinical studies are needed to confirm the in vitro findings. Elderly persons, who are often phenotypically poor metabolizers, may require 50 to 66 percent lower dosages.

OLANZAPINE

Olanzapine is a safe and effective agent in the treatment of schizophrenic symptoms, including positive and negative symptoms, with a very favorable adverse-effect profile. It can be given in a single starting daily dose of 10 mg, with a dosage range of 10 to 30 mg a day. The benign side-effect profile includes weight gain, somnolence, some orthostatic hypotension, and constipation. Liability for extrapyramidal effects and seizures is low. No hematological effects have been observed. Preliminary studies indicate greater improvement in quality of life than with haloperidol as well as a lower rehospitalization rate than with traditional antipsychotic agents. Olanzapine is a first-line antipsychotic agent. Its efficacy in treatment-resistant patients has not been fully established; however, studies are under way to establish olanzapine's mood-stabilizing efficacy in bipolar illness.

History Eli Lilly and Company discovered olanzapine in Great Britain in 1982. It is derived from clozapine, as an agent without strong D_2 receptor blockade. In 1986 renewed interest in the compound emerged. Phase III clinical trials were completed in July 1995, and became available on the American and European markets in the fall of 1996.

Chemistry Olanzapine is a thienobenzodiazepine (Fig. 31.26-1). It can be considered a derivative of clozapine with substitution of a thieno ring for clozapine's carbonyl ring.

Pharmacological Action

Pharmacokinetics

ABSORPTION Food does not affect absorption of olanzapine.

DISTRIBUTION Peak plasma levels of olanzapine are reached in 5 hours. The half-life is 31 hours (range, 21 to 54 hours), which allows once-a-day dosing. The drug is 93 percent protein-bound plasma. Age, gender, or ethnicity effects on olanzapine concentrations are small.

METABOLISM AND ELIMINATION Olanzapine has a very weak affinity for hepatic P450 cytochromes (e.g., CYP 2D6, CYP 1A2, CYP 3A4, CYP 2C19) suggesting that it has little effect on the metabolism of other drugs, and that other drugs minimally affect its concentration in blood.

Pharmacodynamics

MECHANISM OF ACTION Olanzapine specifically blocks 5-HT_{2a}, and D_2 receptors and additionally blocks muscarinic (M_1), H_1, 5-HT_{2c}, 5-HT_3, 5-HT_6, a_1, D_1, and D_4 receptors. Its 5-HT blockade is about eight times as strong as its dopamine receptor blockade. The biochemical profile of olanzapine resembles that of clozapine more closely that those of the SDAs that are available or soon will be, without the typical clozapine adverse effects. The ratio of mesolimbic to striatal dopamine D_2 receptor blockade by olanzapine further indicates that extrapyramidal effects may occur only in some very susceptible patients. In fact, few extrapyramidal effects have been reported in clinical studies to date. Compared with risperidone, sertindole, quetiapine and ziprasidone, olanzapine has considerable M and H receptor antagonism, but less a_2-receptor antagonism. A recent PET study showed that the D_2 effect of olanzapine is greater than that of clozapine but similar to risperidone, which suggests that higher doses will be associated with a rise in prolactin and with extrapyramidal effects. Clozapine has also 5-HT_{1A} agonist properties, which may explain its antianxiety and antidepressant effects.

Olanzapine blocks phencyclidine (PCP) effects in animal studies. PCP is an *N*-methyl-D-aspartate (NMDA) receptor antagonist, that induces a syndrome reminiscent of many aspects of the positive, negative, and related cognitive symptoms of schizophrenia in humans. PCP's biochemical effects have led to the hypoglutamatergic hypothesis of schizophrenia.

BLOOD CONCENTRATIONS AND CLINICAL ACTIVITY Blood concentration studies have not yet appeared; however, side effects do not appear to be dose dependent. Women tend to have higher olanzapine concentrations than men.

Design and Interpretation of Clinical Drug Studies

Early Clinical Trials Double-blind studies compared olanzapine at once-a-day dosages (2.5 to 17.5 mg a day) with placebo and haloperidol (10 to 20 mg a day). Olanzapine was found superior to placebo and comparable to haloperidol in treatment of positive symptoms at intermediate (7.5 to 12.5 mg a day) and high (12.5 to 17.5 mg a day) dosages. Olanzapine was also superior to haloperidol in negative symptom response.

Efficacy in Treatment-Refractory Patients Olanzapine is potentially of interest for treatment-refractory patients because it has a biochemical profile similar to that of clozapine and it is more effective on negative symptoms than haloperidol. One study reported that olanzapine was as effective as chlorpromazine in treatment-resistant patients, while another found that olanzapine was superior to haloperidol in such patients. However, further studies are needed, and such patients may require higher dosages than the 25 mg a day dosages used. A patient who has not responded to one SDA may respond to another.

Long-Term Effectiveness A year-long maintenance study found less relapse with olanzapine than with placebo or haloperidol. In fact, at higher dosages of olanzapine, fewer patients dropped out of the study than with haloperidol.

Treatment Indications

Acute Psychosis Its once-a-day dosage and low side-effect profile make olanzapine an attractive agent for both first-break psychosis patients and relapsed patients who do not like antipsychotic agents because of the extrapyramidal effects. First-break patients tend to improve more with olanzapine than chronic psychotic disorder patients (65 percent vs 40 percent).

Maintenance Treatment Initial studies indicate that efficacy remains high over time and that olanzapine is well tolerated. However, weight gain (>7 percent) can be a serious problem in some patients.

Tardive Dyskinesia No cases of tardive dyskinesia have been observed in the limited time that patients have been studied. Studies of c-*fos* suggest that olanzapine has a low risk of inducing tardive dyskinesia. In all likelihood, olanzapine will be effective in patients with tardive dyskinesia.

Patients With Low Extrapyramidal Effect Threshold So far, akathisia and acute dystonias have been reported only occasionally in the large patient base studied. Because of the low risk of extrapyramidal effects, olanzapine should be a good drug for patients with a low threshold for these effects. However, c-*fos* studies in animals suggest that higher doses of olanzapine may cause extrapyramidal effects.

Other Diagnoses Olanzapine, SDAs, and typical antipsychotics are effective in patients with the same diagnoses.

Precautions and Adverse Reactions Olanzapine's adverse effects (Table 31.26-2) are transient sedation and orthostatic hypotension. The risk of extrapyramidal effects is low, but acute dystonic reactions and akathisia or neuroleptic malignant syndrome have been (rarely) observed. Moderate weight gain, which can be serious in some patients, has been observed. Fortunately, alternatives are now available. Prolactin concentrations increase only transiently and may be dose related. No hematological changes have been observed.

Dosage and Administration

Initiation of Treatment Olanzapine is effective at dosages between 7.5 and 30 mg a day. A single starting dose of 10 mg is well tolerated. The company recommends 10 mg a day for initial dosage. However, patients may need higher or lower dosages to obtain optimal results. Treatment-resistant patients or chronic patients with poor response may need higher doses (30 to 40 mg). This higher dosage range may raise prolactin concentrations and increase the risk of extrapyramidal effects. The half-life suggests that a single daily dose is sufficient.

Duration of Treatment Initial studies suggest that the efficacy increases over time. Early response predicts later response. Guidelines are the same as those for other agents.

Drug Interactions The small effects on hepatic metabolism suggest that interactions with other drugs will not be of much concern. Ethanol increases olanzapine absorption (>25 percent), which may lead to increased somnolence and orthostatic hypotension. Patients who smoke may require higher dosages. Poor extensive CYP 2D6 metabolizers show similar concentrations; thus drugs that inhibit this enzyme should have little effect on olanzapine's efficacy. Carbamazepine and phenytoin decrease olanzapine concentration moderately (£50 percent) by inducing CYP 3A. Cimetidine may increase olanzapine concentration.

SERTINDOLE

Sertindole is an effective, well-tolerated antipsychotic agent, without extrapyramidal adverse effects, and with other, mild adverse effects. It is effective in the treatment of negative symptoms. Its a_1-receptor blockade may cause initial hypotension and sinus tachycardia, nasal congestion, and decreased ejaculatory volume, which usually resolve. Weight gain is about the same or less than that with other SDAs. Prolactin concentrations are not raised to a clinically significant extent. The lack of H_1 and $M_{1–4}$ receptor blockade may explain its lack of sedation, cognitive adverse effects, or constipation. Hematological effects have not been observed. Sertindole has anti–negative symptom effects, but its efficacy in treatment-refractory patients is not yet fully established. Although the initial titration time may be a problem for some patients, the long half-life (3 days) will be helpful in partially compliant patients. Plasma concentrations do not predict therapeutic effect. Initial studies indicate a lower relapse rate than with traditional antipsychotics. Sertindole could be a first-line agent.

History Sertindole was discovered in Denmark in 1985. Abbott Laboratories licensed and brought the drug to the market. Although the New Drug Application (NDA) was filed in 1996, Abbott withdrew its application in the spring of 1998 when the FDA was not forthcoming with approval. The drug is thus only available outside the United States.

Chemistry Sertindole is an arylpiperidylindole derivative (Fig. 31.26-1).

Pharmacological Action

Pharmacokinetics The serum half-life of sertindole is 1 to 4 days, the longest of the drugs discussed in this chapter. Fifty-two (38 to 63) percent of sertindole is excreted in the urine, and 36 (5 to 40) percent in the feces. Plasma half-life is 25 to 29 hours, allowing a once-a-day dosage. The long plasma half-life may lead to a 2-week delay before steady-state concentrations in plasma are reached, which suggests that 3 to 4 weeks must pass before a lack of response is indicated. One circulating metabolite of sertindole is LU 28-092.

Pharmacodynamics

MECHANISM OF ACTION Sertindole blocks $5HT_{2c}$, $5HT_{2a}$, D_2, and a_1 receptors. In clinically effective doses it binds preferential to the mesolimbic dopamine system. Extrapyramidal effects probably require a dosage 10 to 20 times the clinically effective dosage.

BLOOD CONCENTRATIONS AND CLINICAL ACTIVITY No evidence links clinical responsiveness to sertindole to its concentration in plasma.

Design and Interpretation of Clinical Drug Studies

Early Clinical Trials Among clinical trials assessing the clinical efficacy of potential antipsychotic effects, one study stands out, the so-called landmark study with seven treatment arms. Not only did it compare three doses of sertindole with placebo, it also included three arms of haloperidol. Until this study only one dose of the comparison active agent (usually haloperidol, between 12 and 20 mg a day) had been used. All dosages of sertindole (12, 20, and 24 mg a day) and haloperidol (4, 8, and 16 mg a day) were significantly more effective than placebo after 8 weeks of treatment. While all dosages of sertindole and haloperidol were effective on positive symptoms, only sertindole (20 mg a day) was statistically more effective than placebo on negative symptoms. Sertindole and placebo did not differ statistically in extrapyramidal effects; haloperidol (as expected) did. Sertindole was well tolerated; adverse effects were mild and related to its a_1 activity. Improvement in positive symptoms with 4 mg of haloperidol leveled off after 2 weeks, and the 8-mg dosage reached maximum effect after 4 weeks; 16 mg of haloperidol reached its maximum effects in the eighth week but was less effective than the 8-mg dosage. Sertindole (20 mg a day) reached its maximum effect at the eighth week, comparable to 8 mg a day of haloperidol, which was the most effective dosage of that drug. Preliminary long-term studies suggest that patients continue to improve with this dose of sertindole after 8 weeks of treatment.

Efficacy in Treatment-Refractory Patients Seven patients partially responsive to haloperidol received clozapine, risperidone, and sertindole sequentially. Compared with haloperidol, symptoms improved on clozapine (33 percent), on sertindole (35 percent), and on risperidone (45 percent). Results from studies in treatment-resistant patients will soon be available.

Long-Term Effectiveness Studies are under way showing that the efficacy of sertindole continues over time.

Treatment Indications

Acute Psychosis The low adverse-effect profile makes sertindole a good first-line agent, especially for first-episode psychotic patients.

Maintenance Treatment The low adverse-effect profile and preliminary data suggest that maintenance treatment will be well tolerated and effective.

Tardive Dyskinesia The lack of extrapyramidal effects and the actual improvement in involuntary movements (compared with placebo) observed in the studies suggest that long-term risk of tardive dyskinesia with sertindole should be extremely low.

Patients With a Low Extrapyramadal Effect Threshold Because of the lack of extrapyramidal effects at the recommended dosages, patients who cannot tolerate traditional antipsychotic agents are obvious candidates for sertindole.

Other Diagnoses Double-blind studies with patients carrying other diagnoses (e.g., mania, psychotic depression, substance-induced psychotic disorders, dopamine stimulation–induced psychosis in Parkinson's disease, dementia with psychotic features, or substance-intoxication delirium) have not been published. Preliminary evidence from studies with risperidone in patients with obsessive-compulsive disorder suggests that these patients may be helped by sertindole.

Precautions and Adverse Reactions The most frequently observed adverse effects of sertindole are tachycardia, mild nasal congestion, decreased ejaculatory volume (20 percent of patients), mild prolongation of the QT interval, weight gain (0.5 to 1.4 kg), and nausea. In a small number of patients the decreased ejaculatory volume led to discontinuation; in others it resolved spontaneously with continued treatment. This sexual effect and the nasal congestion are attributed to its

a_1-receptor blockade. Other sexual effects observed with the traditional antipsychotic agents such as retrograde ejaculation and diminished libido and ability to experience orgasm (potency) have not been reported. Animal and human studies showed that sertindole prolongs the QT interval, but increasing the dosage does not increase the interval further. The clinical significance of this prolongation has not been established, and this increase may not differ from that observed with some conventional antipsychotic or antidepressants. No incidents of torsade de points have been observed in over 100 patient-years; continued monitoring will clarify the actual risk. However, sudden cardiac death has occurred in patients treated with sertindole. Sertindole has no antihistaminic or antimuscarinic adverse effects. The weight gain may be due to its 5-HT blockade and is similar to that observed with clozapine, olanzapine, risperidone, or quetiapine. Transient elevation of liver enzyme activity (aspartate aminotransaminase [AST], alanine aminotransamine [ALT]) has been reported (as with other agents), without any clinical consequences. Changes in WBC count or other hematological index were not observed. Prolactin concentrations increased only transiently.

Dosage and Administration

Initiation of Treatment The recommended starting dosage of sertindole is 4 mg a day, with increases of 4 mg every other day up to 20 mg a day. Some patients may require 24 mg. If initial agitation is not fully treated before the optimal dose is reached, temporary addition of benzodiazepines or a traditional antipsychotic agent may be helpful. The recommended dosage range is 12 to 20 mg a day. The pharmacokinetics are nonlinear, so that 24 mg a day produces twice the concentration in blood that 20 mg produces. Elderly persons may require much lower dosages. Intramuscular injectable forms are not available. If sertindole replaces clozapine, the authors suggest using the transient anticholinergic strategy used with risperidone.

Duration of Treatment The coming years will provide the needed data about effectiveness and relapse prevention in multiyear studies.

Drug Interactions Fluoxetine and paroxetine decrease sertindole clearance by 50 percent. The authors recommend lowering the dosage by 50 percent in patients also receiving these agents. Carbamazepine and phenytoin increase sertindole clearance by 50 percent, which may require raising the dosage of sertindole. Area under the concentration-time curve for terfendine (Seldane) blood concentrations increased nonsignificantly by 28 percent.

QUETIAPINE

Two double-blind trials with quetiapine indicate that it is a promising SDA with efficacy at least equal to that of haloperidol and chlorpromazine. Doses must be titrated to 150 mg a day and up to 750 mg a day as clinically indicated. Maximum therapeutic effects are seen with 300 mg a day. It is superior to placebo and its extrapyramidal effects do not differ from those of placebo. Adverse effects include drowsiness, increased heart rate, weight gain, and agitation. No blood dyscrasias or effects on prolactin concentrations have been observed in long-term trials. Trials in drug-resistant patients and for pharmacoeconomic comparisons are under way. Quetiapine is a first-line antipsychotic agent.

History Quetiapine was developed by Zeneca Laboratories. Its NDA application was filed at the FDA in 1996 and approved by the FDA in 1998.

Chemistry Quetiapine is a dibenzothiazepine with more potent 5-HT_2 than D_2 receptor-blocking properties. Quetiapine is related to the clozapine (Fig. 31.26-1) and fluperlapine molecules.

Pharmacological Action

PHARMACOKINETICS No difference was found between twice-daily or thrice-daily dosing, in spite of a steady-state half-life of 6.9 hours. The time to maximum concentration (T_{max}) after oral intake is less than 2 hours, with an estimated half-life of 3.0 to 5.0 hours. Steady-state concentrations are reached in 48 hours. PET studies indicated that a single clinical dose of quetiapine occupied 42 percent of D_2 receptors and 72 percent of 5-HT_2 receptors. After 8 and 12 hours, considerable receptor occupancy remained, although concentrations in blood had dropped off, supporting twice a day dosing. Quetiapine has many metabolites; and the only two active ones make up only a small percentage; 95 percent of [^{3}H]quetiapine is recovered as metabolites in urine and feces, and only 1 percent as quetiapine itself.

Pharmacodynamics

MECHANISM OF ACTION Quetiapine has a high affinity for 5-HT_2, histamine H_1, $5HT_6$, a_1, and a_2 receptors, a moderate affinity for D_2 and sigma receptors, and low affinity for D_1 receptors. It has very low affinity for M_1 and D_4 receptors. However, in vitro studies suggest that quetiapine has a much lower receptor blockade profile than other effective antipsychotic agents. An extensive examination in different animal models for schizophrenia (e.g., dopaminergic and nondopaminergic behavior models) indicated that quetiapine has potential antipsychotic and anti–negative symptom effects without significant liability for extrapyramidal adverse effects. Studies in humans suggest that quetiapine has essentially no extrapyramidal adverse effects and that it results in only transient elevations in prolactin concentration.

BLOOD CONCENTRATIONS AND CLINICAL ACTIVITY Quetiapine response is independent of dosage. Because many elderly persons have reduced oral clearance, the dosage may need to be reduced, particularly the starting dosage. Patients with impaired renal or hepatic clearance need to receive 30 to 50 percent lower dosages. No gender or ethnic differences were found in clinical activity or concentrations in blood.

Design and Interpretation of Clinical Drug Studies

Early Clinical Trials Phase II and III trials compared quetiapine at high and low doses. Effectiveness occurs over a wide dosage range, with 300 mg a day generally the optimal dosage. Quetiapine was at least equal to haloperidol in treatment of positive and negative symptoms.

Efficacy in Treatment-Refractory Patients Current trials are ongoing with no results currently available. Anecdotal evidence suggests that some treatment-resistant patients may respond to quetiapine.

Long-Term Effectiveness Although they are expected soon, no results from quetiapine long-term studies have been reported. The low adverse-effect profile and the preliminary clinical data indicate that maintenance treatment will be well tolerated and that therapeutic efficacy will be maintained. A quality of life and pharmacoeconomic study is under way.

Patients With a Low-Extrapyramidal Adverse Effect Threshold Parkinson's disease patients with dopamine agonist-induced psychosis have responded to low doses of quetiapine in open studies.

Treatment Indications Quetiapine is an effective antipsychotic agent in acutely exacerbated schizophrenic and schizoaffective patients. Preliminary data indicate that quetiapine is effective in functional and organic psychoses in elderly persons and very effective in levodopa–induced psychosis in Parkinson's disease patients. It has some effect on negative symptoms. Although not yet reported, quetiapine may be useful in the same nonschizophrenic diagnoses as other SDAs, with the advantage of not being associated with extrapyramidal adverse effects. In patients with a variety of dementias, psychotic symptoms improved with quetiapine. These potential indications require careful examination. Relapse prevention is expected but has not been reported. Response in treatment-refractory patients is being examined.

Precautions and Adverse Reactions The most frequently observed adverse effects (Table 31.26-2) are somnolence, postural hypotension, and dizziness, and tolerance quickly develops to them. Extrapyramidal adverse effects were not observed. Some constipation and dry mouth were observed despite the lack of anticholinergic effects. Prolactin concentrations were not raised after the initial doses. Small increases in heart rate were observed, with supine pulse rate increasing 3.3 to 4.9 beats a minute. Small decreases in thyroid hormone concentration occurred without increases in thyroid-stimulating hormone concentrations, but these changes were not clinically significant. An average weight gain of 1.7 to 2.3 kg (vs. 0.1 kg for placebo) occurred during acute studies. No significant continuing increase in weight was observed in long-term open-label studies. ALT activity increased transiently during the first 2 weeks. No changes in WBC count or other hematological index were observed. QT_c changes were no more likely than with standard antipsychotic agents. Adverse events were as likely on quetiapine as with placebo but less so than with haloperidol or chlorpromazine. Seven overdose incidents (more than 900 mg) were reported (1200 to 9300 mg); none was lethal. One patient developed a first-degree heart block. Occasionally inclusion bodies in the eye are observed. Quetiapine is a safe and well-tolerated antipsychotic agent.

The major risk with quetiapine is the development of cataracts, although a causal relation has not been established. Slitlamp or other sensitive methods to detect lens

abnormalities should be conducted early in treatment and every 6 months thereafter.

Dosage and Administration

Initiation of Treatment Quetiapine doses must be titrated to avoid postural hypotension and syncope. Studies are under way to show how aggressively the dose can be raised. The drug is very well tolerated even in the range of 800 mg a day. Usual titration typically reaches the effective dosage of 300 mg a day by day 4.

Duration of Treatment Most trials had a duration of 6 weeks. Longer-duration studies are under way.

Drug Interactions Interactions between quetiapine and other drugs have been studied to a greater extent than those of most of the new antipsychotic agents. No synergistic effects are noted with alcohol and lorazepam. CYP 3A4 is the major metabolic pathway, and CYP 2D6 is a minor pathway. Quetiapine does not interact with CYT 1A2, 3C9, 2C19, 2D6, or 3A4. Neither cimetidine nor antipyrine (Auralgan) induced quetiapine metabolism or vice versa. Only phenytoin caused a fivefold increase in quetiapine clearance through CYP 3A4 induction. Lithium, lorazepam (or other benzodiazepines), cimetidine, risperidone, haloperidol, fluoxetine, and imipramine (Tofranil) had no effects on the pharmacokinetics of quetiapine, and vice versa (i.e., no required dosage adjustments). Concomitant administration of thioridazine, however, increased oral clearance of quetiapine by 60 percent, which requires further study.

ZIPRASIDONE

Ziprasidone combines antagonist effects at 5-HT_{2A} and D_2 receptors without extrapyramidal, antimuscarinic, anti-a_1 or antihistaminic adverse effects. Ziprasidone is an effective drug for positive symptoms. Negative symptoms improve as well, but no information is available on the deficit state and treatment-refractory patients. Ziprasidone has a low risk of extrapyramidal adverse effects. Food increases absorption but age, gender and renal or hepatic impairments have little effect on the pharmacokinetics. Its norepinephrine and serotonin uptake inhibition may make it of interest in depressed and chronically withdrawn schizophrenic patients. Its potent $5HT_{1A}$ agonist activity suggests that it may also be useful for schizophrenic patients with anxiety symptoms and depression. It is the only SDA that is not associated with weight increase. Hematological effects have not been noted. Drug interactions are expected to be low, suggesting that the drug can safely be used in people with medical treatments and in the elderly. Besides the oral form, an intramuscular preparation would have been available for initiation of treatment. Ziprasidone could have been a first-line drug in schizophrenia or schizoaffective disorder if not for the QT_c prolongation.

History Ziprasidone was synthesized in 1987 by Harry Howard of Pfizer Pharmaceuticals. Tom Seeger screened the compound and identified the activity at D_2 and $5HT_2$ receptors. The FDA did not approve ziprasidone because of QT_c prolongation, and the company withdrew its NDA in the spring of 1998.

Chemistry Ziprasidone is a benzisothiazolyl piperazine (Fig. 31.26-1). Its chemical formula is $C_{21}H_{21}ClN_4OS{\cdot}HCl{\cdot}H_{2O}$ with a molecular weight of 467.41. The compound is not related to any now available or in late-stage development.

Pharmacological Action

Pharmacokinetics

ABSORPTION The peak levels concentrations after single and multiple dosing of ziprasidone were similar. Peak levels were reached 2 to 6 hours after dosing. T_{max} ranged from 4 to 5 hours. Absorption rates did not change from day 1 to 18. Bioavailability doubles when ziprasidone is administered with food; following administration of multiple doses of ziprasidone under fed conditions, peak serum concentrations occur in 6 to 8 hours.

METABOLISM AND ELIMINATION Steady state is attained after 1 to 3 days of dosing. Overall exposure at steady state is relative to the dose over the range of 20 to 80 mg twice a day. The mean terminal plasma half-life for ziprasidone at steady state ranges from 5 to 10 hours in volunteers and schizophrenic patients. The 0.5- to 1-hour lag in reaching peak level concentrations was attributed to delayed gastric emptying.

Age and gender or mild-to-moderate renal or hepatic impairment have no significant effects on the pharmacokinetics of ziprasidone, making dosage adjustments unnecessary. Ziprasidone is extensively metabolized, with less than 1 percent excreted unchanged in the urine or feces. In vitro studies indicate that the major metabolites are produced by CYP 3A4. The major circulating metabolites possess less than 1 percent of the binding affinity of the parent compound for D_2 and 5-HT_{2A} receptors.

Pharmacodynamics

MECHANISM OF ACTION In vitro studies show ziprasidone to be a very potent antagonist at the 5-HT_{2A} receptor, with 5-HT_{2A} to D_2 ratio of 11. PET studies in healthy volunteers confirm that 5-HT_2 occupancy substantially exceeds D_2 occupancy. Ziprasidone has potent affinity for the D_3 receptor and moderate affinity for the D_4 receptor, like haloperidol; the affinity for the D_1 receptor is 100-fold lower than that for other dopamine receptors. Ziprasidone is an agonist at the 5-HT_{1A} receptor and a potent antagonist at the 5-HT_{2C} (5 times greater than clozapine and 19 times greater than risperidone) and 5-HT_{1D} receptors. It also moderately affects norepinephrine and serotonin (and dopamine) uptake sites. Ziprasidone has only moderate affinity for the a_1 receptor and low affinity for the H_1 receptor. The biochemical profile suggests an antipsychotic drug with low risk for extrapyramidal adverse effects, efficacy in treatment of negative symptoms, and additional effects in patients with anxiety and mood disorders. Ziprasidone is the only agent with uptake inhibition of norepinephrine and serotonin. Stimulation of 5-HT_{1A} receptors is shared with clozapine.

BLOOD CONCENTRATIONS AND CLINICAL ACTIVITY With twice-daily administration of ziprasidone, estimates based on systemic exposure and PET data predict that 5-HT_2 and D_2 receptor occupancy ranges from 80 to 90 percent and 45 to 75 percent, respectively. Ziprasidone at serum concentrations of 30 to 40 ng/mL is likely to be associated with about 75 percent 5-HT_2 and more than 65 percent D_2 receptor occupancy as determined from PET scan data.

Design and Interpretation of Clinical Drug Studies

Early Clinical Trials Results from 4- to 6-week fixed-dosage, double-blind clinical trials indicate that ziprasidone is effective in the treatment of positive, negative, and affective symptoms in patients with schizophrenia and schizoaffective disorder. In two separate 4- to 6-week clinical trials in patients with acute psychotic exacerbation, ziprasidone (80 to 160 mg a day) was significantly superior to placebo, and in a 4-week trial, ziprasidone (160 mg a day) was comparable to haloperidol (15 mg a day) in reducing psychotic symptoms. Statistically significant, dosage-dependent improvements in negative symptoms as measured by the PANSS were observed with ziprasidone (80 and 160 mg a day) compared with placebo in a 6-week trial. Further evidence for efficacy in negative symptoms comes from a 4-week, placebo-controlled trial in which a significant improvement in the BPRS anergia factor was observed with ziprasidone (120 mg a day). A dosage-dependent effect was not observed, which suggests that the negative symptom reduction may be related to the lack of extrapyramidal effects. Ziprasidone (160 mg a day) also significantly improved depressive symptoms in patients with clinically significant symptoms at baseline (Montgomery Åsberg Depression Rating Scale [MADRAS] $^{3}14$; baseline mean, 23.5) compared with placebo at 6 weeks. Statistically significant improvement was also observed in affective symptoms at 4 weeks in patients with a BPRS anxiety-depression cluster baseline score of 18 or above.

Long-Term Effectiveness Recent results from a prospective, 1-year, double-blind, fixed-dose ziprasidone study in patients with chronic or subchronic schizophrenia showed that ziprasidone (40 to 160 mg a day) is associated with significantly lower relapse rates than placebo. This group of stable patients experienced a significant reduction in negative symptoms from baseline compared with placebo, as measured with the negative symptom subscale of the PANSS. This improvement in negative symptoms was maintained throughout 12 months of treatment, with a tendency toward continuing improvement in patients who completed the study. Further long-term clinical trials of ziprasidone are ongoing.

Efficacy in Treatment-Refractory Patients No data are available on the efficacy of ziprasidone in treatment-refractory patients.

Treatment Indications The available data from clinical trials show the efficacy of ziprasidone in dosages of 80 to 160 mg a day in the treatment of positive, negative, and depressive symptoms in patients with schizophrenia and schizoaffective disorder. Clinical trial data also indicate that ziprasidone maintenance therapy with

dosages of 40 to 160 mg a day is effective in prevention of psychotic relapse in patients with chronic or subchronic schizophrenia. Most studies so far have focused on chronic and subchronic schizophrenia and schizoaffective disorder. Its biochemical and clinical profile makes ziprasidone a potentially interesting agent in affective disorders with psychosis and in alcoholism and anxiety disorders. The availability of intramuscular administration will make this drug attractive for acutely agitated and psychotic patients.

Precautions and Adverse Reactions Ziprasidone has been well tolerated in short- and long-term clinical trials. Treatment-emergent extrapyramidal adverse effects have been infrequent and mild during the clinical trials. Akathisia has been notably absent. Concomitant benztropine use was significantly lower with ziprasidone than with haloperidol. The major adverse effects encountered with ziprasidone treatment include somnolence, dizziness, nausea, and lightheadedness. Orthostatic hypotension was rare. Unlike other SDAs, ziprasidone evoked a low incidence of clinically significant weight gain (>7 percent), which may be an attractive feature of this agent. Serum prolactin concentrations were not elevated during the chronic phase of the trials (day 18). AST and ALT activities were occasionally elevated but rarely of significance. Treatment-related blood dyscrasias have not been reported with ziprasidone. Prolactin concentrations returned to baseline during the dosing interval in a 4-week study.

Dosage and Administration

Initiation of Treatment Ziprasidone has been administered safely at 40 mg a day in two doses without the need for dose titration. Clinical trial data suggest that 80 to 160 mg a day given as a twice-daily dose with food is effective in acutely exacerbated patients. An intramuscular form of ziprasidone is undergoing testing; this form induces a somnolence from which patients are easily aroused.

Duration of Treatment The preferred dosage for maintenance treatment and prevention of relapse is 40 to 60 mg twice a day. Ziprasidone has been administered safely for up to 2 years, and data from a 1-year clinical trial show that ziprasidone is effective in preventing psychotic relapse.

Drug Interactions Ziprasidone has low potential for clinically significant drug interactions. In vitro studies indicate that ziprasidone is not a substrate for CYP 2D6, CYP 2C9, CYP 2C19 or CYP 1A2 and has low potential to inhibit CYP 1A2, CYP 2C9, CYP 2C19, CYP 2D6, and CYP 3A4. The lack of clinically relevant CYP 2D6 inhibition was confirmed in healthy volunteers; ziprasidone did not inhibit metabolism of dextromethorphan (a CYP 2D6 model substrate) to its major metabolite, dextrorphan. While ziprasidone is a substrate for CYP 3A4, cimetidine (a CYP 3A4 inhibitor) does not inhibit its metabolism, suggesting that significant metabolism occurs by an alternative pathway. Clinically significant interactions have not been reported with lithium, antacids (aluminum and magnesium hydroxides), or oral contraceptives in volunteers. Ziprasidone did not affect steady-state lithium concentrations or renal clearance.

OTHER INVESTIGATIONAL DRUGS

Several other SDAs are being tested in the clinic. Iloperidone is an SDA with high affinity for the 5-HT $_{2A}$ receptor and moderate-to-high affinity for D_{2s}, D_{2L}, D_3, D_4, and $5HT_{1A}$ receptors. This profile suggests that iloperidone will be a second-generation antipsychotic agent with a low liability for extrapyramidal adverse effects and efficacy against negative symptoms and anxiety. The drug is well tolerated, with extrapyramidal adverse effects similar to those of placebo and minimal elevations of prolactin concentration. The preliminary dosage is 8 mg a day.

Aripiprazole (OPC-14597) (Otzuka) is a highly lipid soluble quinolinone derivative that can be given in a once-a-day dosage (2 to 30 mg). Extrapyramidal adverse effects are similar to those with placebo. Prolactin concentrations did not increase with aripiprazole at 4 weeks. The drug stimulates dopamine autoreceptors, with a D_2 antagonism at postsynaptic D_2 receptors in the same dose range. It also blocks 5-HT $_2$ and a_1 receptors. At 4 weeks the 30-mg daily dosage was superior to haloperidol in positive and negative symptoms. Whether aripiprazole increases body weight is uncertain. No evidence of sedation, somnolence, or antimuscarinic effects has been observed.

Several other approaches (e.g., D_4, D_3 and 5-HT_3 antagonists) are also under study. The authors expect the next generation of antipsychotic drugs to have biochemical profiles that are not as solidly based in the monoamines.

IS INDUCTION OF c-*fos* RELATED TO ANTIPSYCHOTIC EFFECTS?

Induction of immediate early genes such as c-*fos* by antipsychotic drugs may assist in understanding the different profiles of the traditional and atypical agents. c-*fos* induction in the dorsolateral striatum suggests extrapyramidal adverse effect liability, and induction of c-*fos* in the prefrontal cortex and nucleus accumbens suggests antipsychotic and potentially anti–negative symptom effects. Haloperidol induces c-*fos* in the dorsolateral striatum and the nucleus accumbens, while clozapine induces c-*fos* in the medial prefrontal cortex, lateral septal nucleus, the major island of Calleja (D_3 receptor blockade), and the nucleus accumbens but not in the dorsolateral striatum. Risperidone strongly induces c-*fos* in the dorsolateral prefrontal cortex, the nucleus accumbens, and the medial prefrontal cortex. Olanzapine treatment shows a pattern like that of clozapine but with some c-*fos* induction in the dorsolateral striatum, although this drug has very low extrapyramidal adverse effect liability at the recommended dosages. Clozapine, risperidone, and olanzapine induce c-*fos* in brain areas that do not show induction with haloperidol, which suggests that these areas may be involved in the treatment of negative symptoms. Similarly, sertindole does not induce c-*fos* in the dorsolateral striatum, and induces very little in the nucleus accumbens and weakly in the medial prefrontal cortex, in spite of its clinical efficacy. Quetiapine also shows c-*fos* induction in the limbic system and ventral striatum but not in the motor areas of the brain. These data indicate that the location of the expression of the immediate early gene c-*fos* differentiates SDAs from conventional agents. The neuronal location also differs among the SDAs as well, even though the anatomical location of its induction is in principle the same. The coming years will include immediate early genes in the screening for potential antipsychotic and anti–negative symptom effects and extrapyramidal adverse effects, showing the individual characteristics of these drugs and their treatment potential.

FUTURE DIRECTIONS

The greater therapeutic efficacy and the low extrapyramidal adverse effect liability of these new agents compared with the traditional agents may be due to the relative blockade of 5-HT $_2$ over D_2 receptors. The differences between the traditional and second-generation antipsychotic agents have also altered our thinking about the biology of psychosis and the clinical indications of these agents. These differences point to neuronal circuitry changes in schizophrenia, rather than an absolute single cause of the changes in behavior. Such changes may involve dopamine, but also serotonin, norepinephrine, GABA, glutamate, and acetylcholine (including muscarinic antagonism and nicotinic receptor stimulation). All these neurotransmitters may play a role in the antipsychotic effects, as they are involved in the cortico-limbic-thalamic pathways in the brain. Conceivably, other, unknown mechanisms may be responsible for the antipsychotic effects when only moderate binding to D_2 receptors provides the optimal response.

These agents differ in their specific receptor affinities, which makes these drugs—including the ones under development—potentially more interesting than just another drug with a low D_2 to 5-HT_2 receptor blockade ratio. For example, risperidone, olanzapine, and quetiapine share a $_2$-receptor blockade with clozapine. 5-HT $_{2A}$ receptor blockade varies widely; risperidone, olanzapine, sertindole, and ziprasidone show high affinities, and quetiapine shows lower affinity. In D_2 receptor blockade, risperidone and olanzapine are similar to haloperidol. The D_4 receptor affinity of haloperidol is similar to that of clozapine, although the implication of this has been disputed. Sertindole and ziprasidone have D_2 affinities somewhat lower than that of haloperidol. Quetiapine is remarkably like clozapine in its low D_2 receptor affinity; while olanzapine has a higher D_2 affinity than clozapine or quetiapine but lower than sertindole and ziprasidone. In addition, ziprasidone is unusual in that it blocks norepinephrine and serotonin uptake, and it shares 5-HT $_{1A}$ receptor stimulation with clozapine. This makes ziprasidone potentially effective for emotional withdrawal and amotivational syndromes and in depressed patients who are psychotic or anxious. Quetiapine has low receptor affinities, with stronger 5-HT $_2$ than D_2 receptor blockade and strong antihistaminic effects. Clozapine, olanzapine, quetiapine, and ziprasidone are considered broad-spectrum, or multireceptor, antagonists, while risperidone, sertindole, and iloperidone share mainly 5-HT $_2$ and D_2 antagonism. Ziprasidone has some specific, additional effects. Sertindole and iloperidone share strong a_1 blockade with different adverse effect profiles. a_1- and a_2-receptor blockade is related to antipsychotic potency of traditional agents.

Improvement in Quality of Life Recent innovations in prescribing antipsychotic drugs have led to the hope that the outcome of schizophrenia can be substantially improved. The second-generation compounds may allow rehabilitation without cognitive impairment as observed with the low-potency agents and clozapine. Research on the effect of the SDAs on the cognitive deficits in schizophrenia has high clinical relevance, and will revolutionize how clinicians and patients will view these drugs. All the new agents are remarkably free of unpleasant side affects. Akathisia is frequently the same as with placebo. Occasionally, initial orthostatic hypotension and somnolence, dizziness (nonorthostatic), stuffed nose, and headaches are observed, while prolactin concentrations are minimally affected or not at

all. Because of their lack of bothersome extrapyramidal effects and drug-induced negative symptoms, all of the new agents may provide a chance for more successful rehabilitation and better compliance. The greater anti–negative symptom response to at least risperidone, olanzapine, and sertindole is a hopeful development. Negative symptoms may fluctuate with relapse and be secondary to positive symptoms or to striatal D_2 blockade (drug induced). Whether the deficit symptoms are responsive to SDAs is unresolved. The lack of antimuscarinic effects (e.g., cognitive interference and sedation) of the newer agents, as with clozapine, chlorpromazine or extrapyramidal adverse effect treatment, may allow better cognitive rehabilitation. Preliminary evidence suggests that cognition improves beyond what is to be expected from the lack of adverse effects. However, well-controlled and well-executed cognitive studies with the newer agents are expected. Other qualitative advantages that are not well addressed with the usual ratings scales seem to be present and require a more focused approach.

Patients with an early age of onset who did not undergo the normal adolescent peer experience and developmental tasks may be unable to deal with the demands of adult life despite absence of schizophrenic symptoms. How this deficiency should be treated is still unknown, but it suggests that therapists and case managers must deal with improvements in a realistic and supportive manner.

How Different Are These Drugs? How much these drugs differ in their clinical profile is unknown. The low adverse-effect profile of the new SDAs makes them attractive first-line agents in first-break patients. Preliminary data suggest that these drugs (e.g., olanzapine and ziprasidone) have antidepressant effects in schizophrenia patients. However, published studies have not addressed the same questions across all SDAs, and others may have similar effects. Anecdotal evidence suggests that an unsatisfactory response to one SDA may not preclude a positive response to any other and that addition of a traditional agent may increase effectiveness as well. Until more studies in treatment-resistant schizophrenic patients have been published with these newer SDAs, clozapine will remain the drug of choice in these patients. However, clinicians should try the newer agents first in chronic patients with serious residual symptoms, as they may be more effective than the traditional agents, with fewer adverse effects. They may replace clozapine for many patients for whom clozapine is considered appropriate. Comparison studies among the SDAs are needed to ascertain specific differences in clinical efficacy and adverse effects.

The authors have spoken out over the years against concurrent use of multiple antipsychotic drugs (polypharmacy). There is a clinical impression from case reports that poor responders to clozapine may do well with addition of a high-potency D_2 blocker (e.g., pimozide). Double-blind trials are obviously needed to address this question. Such additions need to be monitored closely and discontinued as soon as it is clear that they did not induce any change. Similarly, when a patient is gradually switched over to SDA treatment, the original medication can be fully withdrawn within 1 to 2 weeks of having reached the desired dose of the SDA; this is to prevent confusion and to decrease the risk of adverse events. Whether treatment response is associated with specific patient-related factors is not known. With each new agent that reaches the market, the need to identify the responder group becomes more important. The authors expect that genetic approaches to identifying responders and nonresponders to the SDAs will soon become acceptable. At the same time, the fewer adverse effects of these new drugs promise better compliance rates than before. Patients report that they like the new drugs better than conventional agents, but many patients will likely continue to be noncompliant. Initial studies show increased quality of life and increased time in the community, which make these new agents quite attractive.

Therapists' enthusiasm about improvement and unrealistic expectations of a cure must be checked. Patients may be forced into unacceptable situations and stress by the denial of how impaired many patients still are. Clozapine does improve about 30 to 50 percent of patients who did not respond to older drugs, but many patients do not respond, remain severely impaired, or cannot tolerate the drug. Unrealistic expectations of the newer drugs may lead to disappointments. Although perhaps at a lower rate, patients will continue to relapse through noncompliance, stress, or spontaneous fluctuations of the illness, regardless of the level of improvement.

The traditional agents may have some protective effects at the onset of the illness, as shown by a better outcome with a shorter psychotic pretreatment phase during the first episode. If treatment of first-break patients prevents early deterioration, the SDAs may be particularly promising. Such effects may be neuroprotective on the basis of their anti-PCP or sigma receptor binding properties. The ultimate recovery potential may remain limited in some because the major lesions in schizophrenia may have preceded the actual appearance of psychosis. On the other hand if the major lesion in schizophrenia is altered neuronal circuitry, our hope may be justified. The next 5 to 10 years will show whether the course of schizophrenia can be improved.

Some patients may not calm quickly enough with SDAs in the beginning of treatment, and addition of benzodiazepines or a traditional high-potency antipsychotic may be needed temporarily. SDAs are unavailable in short- and long-acting injectable or liquid forms, which limits their use as first-line drugs and in maintenance treatment. Several pharmaceutical companies are exploring injectable forms (e.g., ziprasidone), skin patches, liquid forms (e.g., quetiapine) or easily dissolvable wafers (e.g., olanzapine). Long-acting injectables are not expected within the next few years. The increased activity in antipsychotic drug development and marketing (e.g., treatment guidelines, patient support services, public education) means that clinicians and families will become better educated about treating patients with schizophrenia. The SDAs are being tried in a much wider diagnostic range than the traditional indication of acute and chronic psychosis. Risperidone and olanzapine are presently being developed as mood stabilizers. Preliminary reports of studies in acute mania and chronic affective disorders are encouraging.

Preliminary pharmacoeconomic studies suggest that the extra costs are sometimes offset by the greater efficacy. Even if the newer agents are more expensive, the decrease in hospitalizations, the better quality of life, the fewer adverse effects, and the great improvements in some schizophrenic patients will make treatment with these agents worthwhile. Pharmaceutical companies will continue to search for better antipsychotic compounds, which may not affect dopamine or serotonin activity. At this time, though, the new drugs promise to be real alternatives to traditional agents in first-break patients and in those who respond poorly to the traditional agents or cannot tolerate their adverse effects. The new agents have raised hope for patients and their families.

SUGGESTED CROSS-REFERENCES

Section 12.8 discusses the somatic treatment of schizophrenia. Section 31.17 provides a discussion of the dopamine receptor antagonists. Amino acid and monoamine neurotransmitters are covered in Section 1.5 and Section 1.4 respectively.

SECTION REFERENCES

*Beasley CM Jr, Tollefson G, Tra P, Satterlee W, Sanger T, Hamilton S, the Olanzapine HGAD Study Group: Olanzapine versus placebo and haloperidol: Acute phase results of the North American double-blind olanzapine trial. Neuropsychopharmacology *14:*111, 1996.

*Breier AF, Malhotra AK, Su T-P, Pinals DA, Elman I, Adler CM, Lafargue RT, Clifton A, Pickar D: Clozapine and risperidone in chronic schizophrenia: Effects on symptoms, parkinsonian side effects, and neurendocrine response. Am J Psychiatry *156:*294, 1999.

Casey DE: “Seroquel” (quetiapine): Preclinical and clinical findings of a new atypical antipsychotic. Exp Opin Invest Drugs *5:*939, 1996.

Fink-Jensen A, Kristensen P: Effects of typical and atypical neuroleptics on Fos protein expression in the rat forebrain. Neurosci Lett *182:*115, 1994.

*Galletly CA, Clark CR, McFarlane AC, Weber DL: Effects of clozapine for non–treatment-resistant patients with schizophrenia. Psychiatr Serv *50:*101, 1999.

*Kane JM, Honigfeld G, Singer J, Meltzer H, and the Clozaril Collaborative Study Group: Clozapine for the treatment-resistant schizophrenic: A double-blind comparison versus chlorpromazine/benztropine. Arch Gen Psychiatry *45:*789, 1988.

*Kapur S, Zipursky RB, Remington G: Clinical and theoretical implications of 5-HT_2 and D_2 receptor occupancy of clozapine, risperidone, and olanzapine in schizophrenia. Am J Psychiatry *156:*286, 1999.

Marder SR, Meibach RC: Risperidone in the treatment of schizophrenia. Am J Psychiatry *151:*825, 1994.

Marder SR, Van Putten T: Who should receive clozapine? Arch Gen Psychiatry *45:*865, 1988.

Meltzer HY, Cola PA: The pharmacoeconomics of clozapine: A review. J Clin Psychiatry *55*(Suppl):161, 1994.

Reitz AB, Scott MK: Novel antipsychotics with unique D_2 5-HT_{1A} affinity and minimal extrapyramidal side effect liability. Adv Med Chem *3:*1, 1995.

Rosenheck RA, Cramer J, Xu W, Thomas J, Henderson W, Frisman LK, Fye C, Charney D: A comparison of clozapine and haloperidol in the treatment of hospitalized patients with refractory schizophrenia. N Engl J Med *337:*809, 1997.

Safferman A, Lieberman JA, Kane JM, Szymanski S, Kinon B: Update on the clinical efficacy and side effects of clozapine. Schizophr Bull *17:*247, 1991.

*Seeger TF, Seymour PA, Schmidt AW, Zom SH, Schulz DW, Lebel L, McLean S, Guanowsky V, Howard HR, Lowe JA, Heym J: Ziprasidone (CP-88,059): A novel antipsychotic with combined

dopamine and serotonin receptor antagonist activity. J Pharmacol Exp Ther *275:*101, 1995.

Seeman P: Atypical neuroleptics: Role of multiple receptors, endogenous dopamine, and receptor linkage. Acta Psychiatr Scand *82* (Suppl 358):14, 1990.

Tandon R, Harrigan E, Zorn SH: Ziprasidone: A novel antipsychotic with unique pharmacology and therapeutic potential. J Serotonin Res *4:*159, 1997.

van Kammen DP, McEvoy JP, Targum SD, Kardatzke D, Sebree TB, and the Sertindole Study Group: A randomized, controlled, dose-ranging trial of sertindole in patients with schizophrenia. Psychopharmacology *124:*168, 1996.

31.27 SYMPATHOMIMETICS

JAN FAWCETT, M.D.

Sympathomimetic amines are defined as naturally occurring catecholamines and drugs that mimic their actions. They are currently referred to as stimulant drugs, which can increase motivation, mood, energy, and wakefulness. While these drugs act specifically on symptoms of poor concentration and hyperactivity in children and adults, as well as being approved for use in increasing alertness in narcolepsy, they were also used to maintain wakefulness, alertness, energy, and confidence in combatants in the air and on the ground in several wars, ranging from Bavarian soldiers in the mid 1880s to extensive use in World War II and in fliers in the Gulf war, and most recently in sleep-deprived combat helicopter pilots. Because of their rapid onset, immediate behavioral effects, and propensity to develop tolerance, which leads to the risk of abuse and dependence in vulnerable individuals, they have been classified as controlled drugs and are stigmatized, and their manufacture, distribution, and use are regulated by state and federal agencies.

Despite these caveats, their use persists and may be increasing in medicine and psychiatry in specific clinical situations. Stimulants can be of great help, if appropriately prescribed and monitored, because of their effectiveness in specific clinical situations in which no other drug has been helpful. Sympathomimetics have been widely used in attention-deficit/hyperactivity disorder and narcolepsy because no other equally effective agents exist and have been found effective in certain cognitive disorders that result in secondary depression or profound apathy (e.g., acquired immune deficiency syndrome [AIDS] depression and dementia, closed head injury), as well as in the pharmacological augmentation of antidepressant medications in specific treatment-refractory depressions.

The compound *d,l*-phenylisopropylamine or amphetamine, was first evaluated in the early 1930s when the dextrorotatory isomer, dextroamphetamine (Dexedrine), was found to increase wakefulness and alertness, to promote improved physical and mental performance, and to suppress appetite. Amphetamine was used to treat narcolepsy (in divided doses of 30 to 50 mg a day) then postencephalitic parkinsonism, epilepsy, and for a while barbiturate poisoning. In the early 1940s amphetamine was thought to be useful for treating “psychopathic states and delinquency” with mixed results, which preceded its use in children with attention-deficit/hyperactivity disorder. Controlled studies in depression did not support routine use of amphetamine especially in the face of the risk of dependence and abuse in the late 1960s. In the 1960s amphetamines were used with mixed results in the treatment of obesity. In 1970, amphetamine and its derivatives were scheduled and regulated under the Controlled Substances Act. While the indications for the use of sympathomimetic stimulants have narrowed over the years and their risks have been recognized, they continue to play a growing role in behavioral medicine and psychiatry. Recently, a product that is a combination of dextroamphetamine and amphetamine (Adderall) has been marketed for the treatment of attention-deficit/hyperactivity disorder.

CHEMISTRY

Amphetamine, or racemic phenylisopropylamine, is similar in structure to the catecholamines such as dopamine and norepinephrine, but unlike catecholamines, crosses the blood-brain barrier (Fig. 31.27-1) The dextrorotatary isomer was found to be 3 to 4 times more potent in behavioral effects than the levorotatory isomer. An a-methyl group protects amphetamine from rapid breakdown by the monoamine oxidase (MAO) type A enzyme that breaks down catecholamines (Fig. 31.27-1).

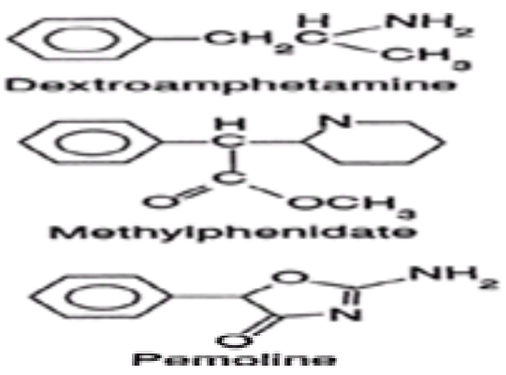

FIGURE 31.27-1 Molecular structures of selected sympathomimetics.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics Dextroamphetamine reaches peak blood levels in 2 to 3 hours, with a variable effect of several hours after oral administration. Though partially metabolized hepatically, almost half is excreted unchanged in the urine, constituting the basis for using ammonium chloride to acidify the urine, which can markedly increase excretion in the management of amphetamine overdose.

Methylphenidate (Ritalin) is a piperidine derivative, structurally related to amphetamine (Fig. 31.27-1), a mild stimulant, with some of its effects. It reaches peak blood concentrations in 2 hours and has a short half-life of 1 to 2 hours. Concentrations of methylphenidate in the brain exceed blood concentrations. Ritanilic acid is the major metabolite excreted in urine.

Pemoline (Cylert) is structurally dissimilar to methylphenidate (Fig. 31.27-1) but exerts similar mild stimulant effects. It peaks in 2 to 4 hours in blood but has a longer half-life of 5 to 7 hours in children. It is partially (about 50 percent) bound to proteins, with about half excreted unchanged in urine.

Pharmacodynamics All of the stimulants are considered indirect-acting sympathomimetics because they do not act directly on receptors as do natural cathecholamines but increase their effect by increasing release at the synapse while blocking reuptake (therefore slowing metabolism and increasing effect). Dextroamphetamine is the most potent by virtue of its ability to release both dopamine and norepinephrine and its capacity to release both newly formed soluble pools and granule-stored pools of catecholamines in nerve endings. Methylphenidate and pemoline act mainly on dopamine (not norepinephrine as well) through release of dopamine from granular pools of dopamine and blockage of the reuptake of dopamine. This may explain why amphetamine is a more powerful stimulant and is more likely to activate latent hypertension than methylphenidate.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Attention-Deficit/Hyperactivity Disorder The treatment of symptoms of narcolepsy, the treatment of attention-deficit/hyperactivity disorder, and more recently residual attention-deficit/hyperactivity disorder in adults are clinical uses that have continued to be dominant treatments since early observations of these effects with the discovery of amphetamine's actions. R.A. Barkley's 1976 review of studies involving a total of 915 child and adolescent patients (mainly treated with methylphenidate) showed that approximately three-quarters showed improvement, though he concluded that long-term psychosocial adjustment was unchanged. A

more recent review of 11 randomized, controlled, longer-term studies (mainly methylphenidate) indicated clear benefit of sustained treatment on core behavioral features of attention-deficit/hyperactivity disorder (poor sustained attention, impulsiveness, and excessive motor activity). Two most recent comprehensive reviews by different groups conclude that given adequate doses of stimulants, approximately 70 percent of those treated achieve significant improvement in cognitive function and behavioral problems; improved social interactions with family, teachers, and peers; and improved self-esteem. After reviewing the available data regarding the use of newer antidepressants, such as buproprion and ventralaxine, as well as other pharmacotherapies for attention-deficit/hyperactivity disorder, in 1997 C.W. Popper concluded that “psychostimulants remain the treatment of choice because of their unique effect on attention.”

Treatment-Resistant Depression and Medical Depression Case reports of the effectiveness of adding psychostimulants such as methylphenidate and dextroamphetamine to tricyclic drugs and serotonin reuptake inhibitors in treatment-resistant depression have been published recently. The effectiveness of these psychostimulants in modest doses in the treatment of depression in the medically ill has been documented as for the use of stimulants in the treatment of depression in the elderly. Methylphenidate has been reported to be beneficial with depressed hospice patients. Early controlled trials of amphetamine use in clinical depression failed to show significant therapeutic effect as monotherapy, but the methodology of these studies has been questioned. This, in the context of the risk of dependence and abuse, resulted in the exile of stimulants from conventional psychiatry. More recently case report data and open clinical trials have again raised the possibility of the effectiveness and usefulness of stimulants as an augmentation medication in treatment-resistant depression and as monotherapy in medical depression (patients with depression associated with comorbid medical conditions such as stroke, AIDS, and cancer pain treated with high-dose opioids, with consequent mood-obtunding effects). Several case reports of the effectiveness of adding stimulants to tricyclic drugs and serotonin-reuptake inhibitors in treatment-resistant depression have been published and have been recently reviewed by A. Nierenberg, and colleagues. The use of stimulants to augment the effects of MAO inhibitors (MAOIs) in treatment-refractory patients (many who had benefited from electroconvulsive therapy [ECT]) who constituted high suicide risks in their untreated state has also been reported. This report emphasized the importance of documenting a risk-benefit analysis, informed consent including the reports of three deaths associated with this combination in years past, as well as obtaining double consent (patient and significant other) and recognition that this combination is listed as contraindicated in the labeling of available monoamine oxidase inhibitors.

Central Nervous System Damage Increasing reports are appearing of the successful use of sympathomimetic stimulants in patients with central nervous system (CNS) damage stemming from stroke, closed head injuries, AIDS dementia, and related disorders characterized by mental dulling, apathy, cognitive decline, lack of motivation, lack of initiative, and lack of energy as well as depression. All of these reports are relatively unsupported by controlled studies, except for comparison studies of stroke patients, but three recent reports of controlled studies in 88 total patients support positive outcomes previously reported. One report of benefit in a patient improving from severe apathy following subcortical infarcts demonstrated improved frontal lobe functions by single photon emission computed tomography (SPECT) scan and reaction times. Five of six placebo-controlled studies of methylphenidate in geriatric patients with early dementia and symptoms of apathy have demonstrated effectiveness. J. Whyte and colleagues recently published positive effects of methylphenidate in traumatic brain injury patients from a randomized placebo-controlled trial. Most recently, M.A. Glen has reviewed the use of methylphenidate for cognitive and behavioral dysfunction after traumatic brain injury, C. Grabe and colleagues have recently published a placebo-controlled study of methylphenidate in post stroke recovery. Of particular interest is the recent report of C.A. Meyers and colleagues showing significant improvement in cognitive function in a study of thirty patients with brain tumor, with very few adverse effects and no seizures noted, with dosages of methylphenidate ranging from 10 to 30 mg twice daily.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Central Nervous System Amphetamine stimulates the medullary respiratory center and has excitatory effects on cortical function. Depending on personality and contextual factors, amphetamine in adults can increase wakefulness, energy, alertness, initiative, self confidence, and physical and mental performance, lessen fatigue, and produce euphoria. These effects occur shortly after dosing. Because of these sometimes dramatic, rapid-onset effects as well as the tendency for tolerance of these effects to develop, the danger of drug dependence and potential for abuse must be kept in mind. After withdrawal of the drug, patients who have been taking excessive doses may note excessive tiredness and sleepiness. More important, the patient may experience severe depression with suicidal im pulses. In children and adults with attention-deficit/hyperactivity disorder, stimulants do not produce euphoria but tend to reduce hyperactivity and help to focus and sustain attention.

Cardiovascular System Amphetamines can cause increased blood pressure (particularly in patients with hypertension), and with high doses, cardiac arrhythmias can result (particularly in patients with cardiovascular disease). Such effects are not likely at usual clinical doses in a patient without cardiovascular disease or hypertension. Amphetamine is more potent in producing cardiovascular effects than dextroamphetamine because of more potent effects on norepinephrine.

More recently, the mild stimulant phentermine (Ionamin) was frequently administered with the serotonergic drug fenfluramine in combination for weight loss; the combination was popularly known as “Fen-Phen.” Because of evidence suggesting the possibility that fenfluramine used alone or in combination may in some cases lead to pulmonary hypertension and cardiac valvular disease, fenfluramine was voluntarily withdrawn from the market. It was not determined whether phentermine increased the likelihood of these adverse effects when used in combination with fenfluramine.

Endocrine Effects Early reports suggested that both dextroamphetamine and methylphenidate might suppress growth in children. A recent controlled study of children and young adolescents found small but significant height differences evident in early (but not late) adolescent children with attention-deficit/hyperactivity disorder, unrelated to the use of psychotropic medications. This study concluded that the effects on growth seemed to be related to the disorder, not its treatment.

Hepatic Effects In the past year attention has been drawn to possible liver toxicity associated with the use of pemoline. The frequency of this risk is still debated, but it has led to recommendations for liver function monitoring and a labeling warning that pemoline should not ordinarily be considered first-line drug therapy for attention-deficit/hyperactivity disorder.

THERAPEUTIC INDICATIONS

Amphetamine is approved for use in attention-deficit/hyperactivity disorder, narcolepsy, and exogenous obesity, though tolerance develops to its anorectic effects. Food and Drug Administration (FDA)-approved uses of methylphenidate include attention-deficit/hyperactivity disorder and narcolepsy. Pemoline is also approved by the FDA for attention-deficit/hyperactivity disorder. As reviewed above, open studies and comparative studies support the usefulness of amphetamine and methylphenidate for depression secondary to medical conditions, apathy and depression secondary to CNS disease or trauma, relief of obtunding effects of opioids in cancer patients, and augmentation of antidepressant medication in some cases of treatment-resistant depression.

PRECAUTIONS AND ADVERSE REACTIONS

Adverse effects (Table 31.27-1) noted with therapeutic doses of amphetamines include mild gastrointestinal disturbance, dryness of mouth, tachycardias, cardiac arrhythmias, insomnia, and restlessness. Headache, palpitations, agitation, confusion, dysphoria, apprehension, and delirium have also been reported. Adults and children may note depressed appetite, which develops tolerance over time, limiting these stimulants' effectiveness as anorectics with long-term use. Adverse effects can include overstimulation and precipitation of mania or psychosis (especially in vulnerable individuals). Medically unsupervised use of amphetamine can involve excessive doses, which can result in such effects as loquaciousness, overcheerfulness, restlessness, rapid and slurred speech, tension, anxiety, and irritability. Excessively dry mouth, with a tendency to rub the tongue along the inside of the lower lip, tachycardia, cardiac arrhythmias, brisk reflexes, dilation of the pupils, and a fine tremor of the limbs may occur. A paranoid psychosis in the setting of clear consciousness may occur. A rare, short-lasting confusional state may occur. There have been an increasing number of case reports of myocardial infarction, cerebral hemorrhage, and cerebral angiitis in individuals who were either abusing the drug intravenously or ingesting toxic doses. While the potential for abuse and dependence should be of concern in adults, especially those with past or family history of substance abuse or antisocial personality, there is no evidence of abuse liability in children treated for attention-deficit/hyperactivity disorder. Because this diagnosis in adults is sometimes problematic and often confounded by comorbid diagnoses of substance abuse or antisocial behavior or both, care must be taken in the use of stimulants in this instance. The possible effectiveness of noradrenergic antidepressants such as desipramine (Norpramine), venlafaxine (Effexor), and buproprion (Wellbutrin) provides alternative choices. Stimulants are relatively well tolerated by elderly patients.

Table 31.27-1 Management of Side Effects of Sympathomimetic Drugs

DRUG INTERACTIONS

Amphetamine may interfere with therapeutic effects of antihypertensive medications. In small doses, amphetamine may augment the analgesic effects of opiates. Methylphenidate may partially inhibit the metabolism of tricyclic drugs, warfarin (Coumadin), primidone (Mysoline), phenobarbital phenytoin (Dilantin), and phenylbutazone (Butazolidin). Because of three deaths from two fulminating hypertensive reactions and one case of hyperthermia reported when amphetamine was added to MAOI, this combination has been contraindicated in the labeling of MAOIs. Combination of any stimulants with MAOIs has been contraindicated by extension. The combination was reported to be effective in treatment-refractory patients in two small open series. Therefore, this combination should be used only when justified by a risk-benefit analysis and with informed consent, preferably of the patient and significant others.

LABORATORY INTERFERENCE

Amphetamines may elevate plasma corticosteroid concentrations.

DOSAGE AND ADMINISTRATION

Sympathomimetic therapy should be started at low dosages and increased in small increments every 3 to 7 days to determine the most-effective and best-tolerated dose for the individual patient. Methyphenidate is the most used drug in the treatment of attention-deficit/hyperactivity disorder, with the most studies reported. It has a lower likelihood of abuse than amphetamine. Dextroamphetamine is about twice as potent as methyphenidale. Dextroamphetamine and methylphenidate are given in divided doses, with the second dose in the early afternoon to avoid insomnia, while pemoline is given in one morning dose. The usual dosage of dextroamphetamine is 20 to 40 mg a day in doses of 5 to 10 mg two or three times daily, and 20 to 60 mg of methylphenidate daily is given in doses of 10 to 20 mg two or three times daily (Table 31.27-2).

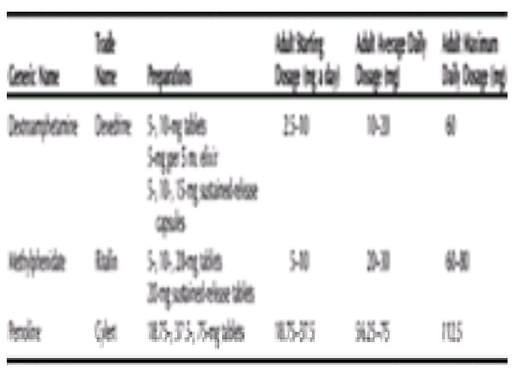

Table 31.27-2 Sympathomimetics

SUGGESTED CROSS-REFERENCES

Attention-deficit/hyperactivity disorder is discussed in Chapter 39. Amphetamine abuse is discussed in Section 11.3 and the treatment of mood disorders is discussed in Section 14.7.

SECTION REFERENCES

Barkley RA: Predicting the response of hyperkinetic children to stimulant drugs: A review. J Abnorm Child Psychol *4:*327, 1976.

Barkley RA: A review of stimulant drug research with hyperactive children. J Child Psychol Psychiatry *18:*137, 1976.

Barkley RA: *Attention Deficit Hyperactivity Disorder*. Guilford, New York, 1990.

Caldwell JA, Caldwell JL: An in-flight investigation of the efficacy of dextroamphetamine for sustaining helicopter pilot performance. Aviat Space Environ Med *68:*1073, 1997.

*Castellanos FX: Stimulants and tic disorders: From dogma to data. Arch Gen Psychiatry *56:*337, 1999.

*Chiarello RJ, Cole JO: The use of psychostimulants in general psychiatry. Arch Gen Psychiatry *44:*286, 1987.

*Das Gupta K: Treatment of depression in elderly patients: Recent advances. Arch Fam Med *7:*271, 1998.

Dulcan MK: Using psychostimulants to treat behavioral disorders of children and adolescents. J Child Adolesc Psychopharmacol *1:*7, 1990.

DuPaul GJ, Rapport MD: Does methylphenidate normalize the classroom performance of children with attention deficit disorder? J Am Acad Child Adolesc Psychiatry *32:*190, 1993.

*Elia J: Drug treatment for hyperactive children: Therapeutic guidelines. Drugs *46:*863, 1993.

*Elia J, Ambrosini PJ, Rapoport JL: Treatment of attention-deficit-hyperactivity disorder. N Engl J Med *340:*780, 1999.

Elia J, Borcherding BG, Potter WZ, Mefford IN, Rapoport JL, Keysor CS: Stimulant drug treatment of hyperactivity: Biochemical correlates. Clin Pharmacol Ther *48:*57, 1990.

Elia J, Welsh PA, Gullotta CS, Rapoport JL: Classroom academic performance: Improvement with both methylphenidate and dextroamphetamine in ADHD boys. J Child Psychol Psychiatry *34:*785, 1993.

Fawcett J, Kravitz HM, Zajecka JM, Schaff MR: CNS stimulant potentiation of monoamine oxidase inhibitors in treatment-refractory depression. J Clin Psychopharmacol *11:*127, 1991.

Glen MB: Methylphenidate for cognitive and behavioral dysfunction after traumatic brain injury. J Head Trauma Rehabil *13:*81, 1998.

Grade E: Methylphenidate in early poststroke recovery: A double-blind, placebo-controlled study. Arch Phys Med Rehabil *79:*1047, 1998.

*Halpern JH: Treatment of attention-deficit/hyperactivity disorder. JAMA *281:*1491, 1999.

Jacobvitz D, Sroufe, LA, Stewart M, Leffert N: Treatment of attentional and hyperactivity problems in children with sympathomimetic drugs: A comprehensive review. J Am Acad Chil Adolesc Psychiatry *29:*677, 1990.

Klein RG, Mannuzza S: Hyperactive boys almost grown up. Arch Gen Psychiatry *45:*1131, 1998.

Kupietz SS, Winsberg BG, Richardson E, Maitinsky S, Mendell N: Effects of methylphenidate dosage in hyperactive reading-disabled children: Behavior and cognitive performance effects. J Am Acad Child Adolesc Psychiatry *27:*70, 1988

Macleod AP: Methylphenidate in terminal depression. J Pain Symptom Manage *16:*193, 1998.

Masand PS, Pierett P, Murray G: Psychostimulants for secondary depression in medical illness. Psychosomatics *32:*203, 1991.

Masand PS, Tesar G: Use of psychostimulants in the medically ill. Psychiatr Clin North Am *19:*515, 1996.

Matochik JA, Liebenauer LL, King AC, Szymanski HV, Cohen RM, Zametkin AJ: Cerebral glucose metabolism in adults with attention deficit hyperactivity disorder after chronic stimulant treatment. Am J Psychiatry *151:*658, 1994.

Mattes JA, Boswell L, Oliver H: Methylphenidate effects on symptoms of attention deficit disorder in adults. Arch Gen Psychiatry *41:*1059, 1984.

Meyers CA, Weitzher MA, Valentine AD, Levin VA: Methylphenidate improves cognition mood and function of brain tumor patients. J Clin Oncol *16:*2522, 1998.

Nierenberg AA, Dougherty D, Rosenbaum JF: Dopaminergic agents and stimulants as antidepressant augmentation strategies. J Clin Psychiatry *59*(Suppl):60, 1998.

Pelham WE, Carlson C, Sams SE, Vallano G, Dixon MH, Hoza B: Separate and combined effects of methylphenidate and behavior modification in boys with attention deficit-hyperactivity disorder in the classroom. J Consult Clin Psychol *61:*506, 1993.

Plutchik L, Synder S, Drooker M, Chodoff L, Sheiner P: Methylphenidate in post liver transplant patients. Psychosomatics *39:*118, 1998.

Popper CW: Antidepressant in the treatment of attention-deficit hyperactivity disorder. J Clin Psychiatry *58*(Suppl):14, 1997.

Prakash M, Rickett P, Murray GB: Psychostimulants for secondary depression in medical illness. Psychosomatics *32:*203, 1991.

*Rapport MD, Kelly KL: Psychostimulant effects on learning and cognitive functioning: Findings and implications for children with attention-deficit hyperactivity disorder. Clin Psychol Rev *11:*61, 1991.

Satel SL, Nelson JC: Stimulants in the treatment of depression: A critical overview. J Clin Psychiatry *50:*241, 1989.

Solanto MV: Neuropharmacological basis of stimulant drug action in attention deficit disorder with hyperactivity: A review and synthesis. Psychol Bull *95:*387, 1984.

Swanson JM, Cantwell D, Lerner M, McBurnett HG: Effects of stimulant medication on learning in children with ADHD. J Learn Disabil *24:*219, 1991.

Thase ME, Nowland RH, Friedman ES: Treating antidepressant nonresponders with augmentation strategies: An overview J Clin Psychiatry *59*(Suppl):5, 1998.

Wagner GJ, Rabkin JG, Rabkin R: Dextroamphetamine as a treatment for depression and low energy in AIDS patients: A pilot study. J Psychosom Res *42:*407, 1997.

Wender PH, Reimherr FW, Wood D: Attention deficit disorder (“minimal brain dysfunction”) in adults. Arch Gen Psychiatry *38:*449, 1981.

Wender PH, Reimherr FW, Wood D, Ward M: A controlled study of methylphenidate in the treatment of attention deficit disorder, residual type, in adults. Am J Psychiatry *142:*547, 1985.

Whyte J, Hart T, Schuster K, Fleming M, Polansky M, Coslett HB: Effects of methylphenidate on attentional function after traumatic brain injury. A randomized, placebo-controlled trial. Am J Phys Med Rehabil *76:*440, 1997.

*Wilens TE, Biederman J: The stimulants. Psychiatr Clin North Am *15:*191, 1992.

Woods SW, Tesar GE, Murray GB, Cassem NH: Psychostimulant treatment of depressive disorders secondary to medical illness. J Clin Psychiatry *47:*12, 1986.

Zametkin AJ, Rapoport JL: Neurobiology of attention deficit disorder with hyperactivity. Where have we come in 50 years? J Am Acad Child Adolesc Psychiatry *26:*676, 1987.

CHAPTER 31. BIOLOGICAL THERAPIES

31.28 THYROID HORMONES

RUSSELL T. JOFFE, M.D.

HISTORY

The link between thyroid disease and psychiatric symptomatology, particularly depression, was well established more than 100 years ago. The clinical literature of the late nineteenth century has rich description documenting a relation between melancholic symptoms and myxedema. These observations in endocrine patients led to the logical assumption that patients with depressive illness might have abnormalities of their thyroid axis that were of etiological importance to their psychiatric illness. A concerted research effort over the last 40 years, however, has failed to demonstrate a specific abnormality of thyroid function test result in patients with primary depressive illness. Over these years, tests of thyroid function have become greatly refined and more accurate, yet they find no direct etiological link between even subtle forms of hypothyroidism and major depressive disorder. Some patients with major depression have evidence of clinical or subclinical hypothyroidism, but they are a relatively small number, and most patients presenting with depression to a psychiatric clinic are euthyroid by both clinical examination and laboratory tests. Although up to one-third of depressed patients have a blunted thyrotropin (TSH) response to thyrotropin-releasing hormone (TRH), the pathophysiological significance of this finding remains unresolved and it has limited clinical utility.

Prompted by the link established between thyroid disease and psychiatric symptomatology in the endocrine clinics and by the lack of effective psychotropic treatments, hormones in general and thyroid hormones in particular were used in the earlier part of the century to treat psychiatric illness. The results were mixed, and the practice fell into disrepute. In the 1960s, levothyroxine (Levoxyl, Levothroid, Synthroid), synthetic forms of thyroxine (T_4) and liothyronine (Cytomel), the synthetic levorotary isomer of triiodothyronine (T_3), were developed and soon replaced desiccated thyroid as the primary agents in replacement therapy in patients with thyroid disease. As a result of seminal work carried out by Arthur Prange and collaborators in the United States, beginning in the late 1960s, the modern era of therapeutic use of thyroid hormones in the treatment of depression began and became established. This effort to generate systematic data on the efficacy of thyroid hormones as psychotropic treatments has been largely confined to the mood disorders, primarily unipolar depression but also some subtypes of bipolar mood disorders.

CHEMISTRY

Two main hormones are produced by the thyroid gland, T4 and T3. The main source of circulating T_4 is the thyroid gland; T_3 is derived largely by monodeiodination of T_4 in extrathyroidal tissues. Both these hormones have been used in the treatment of mood disorders. Their structure is shown in Figure 31.28-1.

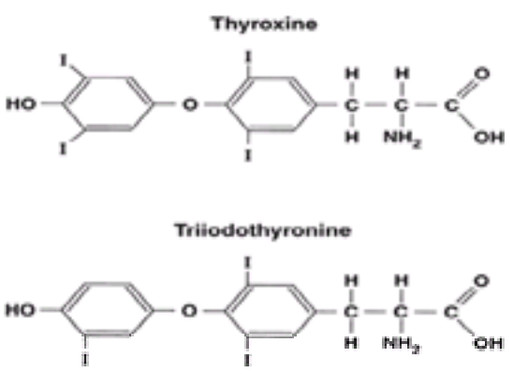

FIGURE 31.28-1 Molecular structures of thyroxine and triiodothyronine.

In addition, several other thyroid hormones are produced including reverse T_3, 3-diiodothyronines, 1-monoiodothyronine, and the acetate metabolites of T_4 and T_3, respectively, tetrac and triac. All of these hormones are produced in substantially lower amounts than either T_4 or T_3 and have uncertain physiological significance.

Secretion of thyroid hormones by the thyroid gland is regulated by the hypothalamic-pituitary-thyroid axis. Regulation of the thyroid axis is illustrated in Figure 31.28-2.

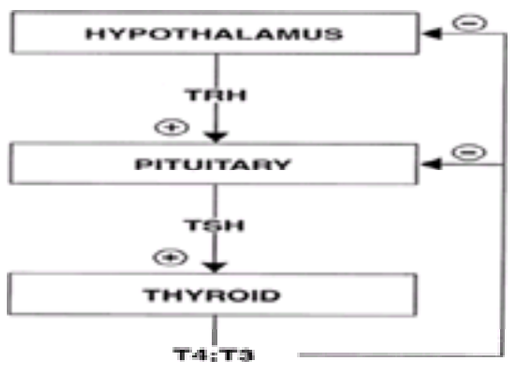

FIGURE 31.28-2 Schema of the feedback control of the hypothalamic-pituitary-thyroid axis.

Cells of the hypothalamus secrete the tripeptide TRH, which regulates the release of TSH from anterior pituitary cells via the portal venous blood system connecting the hypothalamus to the pituitary gland. The TSH is then released from the pituitary where it acts on the thyroid gland to stimulate production of T_4 and T_3. Through a negative-feedback system operating at both the level of the pituitary and of the hypothalamus, circulating levels of T_4 and T_3 regulate the activity of the thyroid axis.

In many respects, T_4 is regarded as a prohormone since its action seems to depend largely on its conversion to T_3, which then acts at the T_3 receptor to initiate the metabolic action of thyroid hormones. For the most part this conversion of T_4 to T_3 occurs in the circulation: thus the thyroid hormone available to peripheral tissues is largely T_3. The major exception is brain and to a lesser extent pituitary where the main source of thyroid hormone for action at the receptor is T_4 that gets converted

locally within the neuron to T_3 before exerting its physiological action.

PHARMACOLOGICAL ACTIONS

Pharmacokinetics Levothyroxine and liothyronine are relatively small molecules that are easily and completely absorbed across the gastrointestinal tract. Both of these hormones are strongly bound to binding proteins and are widely distributed in tissues. There is active transport of thyroid hormones, particularly levothyroxine, across the blood-brain barrier. In addition transport mechanisms exist at the choroid plexus for movement of thyroid hormones within the central nervous system. A series of deiodinase enzymes metabolize thyroid hormones to smaller, less-iodinated molecules that are then excreted both in urine and feces. Levothyroxine has a relatively long half-life of approximately 1 week; liothyronine has a half-life of approximately 8 to 12 hours.

Pharmacodynamics Thyroid hormones have widespread metabolic effects. They are essential for brain development in the maturing infant and, in adults, play key roles in both protein synthesis and calorigenesis.

Until about 20 years ago, it was generally accepted that thyroid hormones had little if any effect on mature brain function. Considerable evidence now indicates that this is not the case and that thyroid hormones may have key roles in mature brain function. Important interactions exist between thyroid hormones and various neurotransmitter and neuropeptide systems in the brain, which may account for the putative effect of thyroid hormones on mood, behavior, and cognition in the adult. Thyroid hormones may also be important in maintaining neuronal cytoarchitecture.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

All of the hormones of the thyroid axis have been used for the treatment of depression, either as monotherapy or in combination with antidepressants. The best evidence for therapeutic efficacy is the use of liothyronine to augment therapeutic response in antidepressant nonresponders. Preliminary evidence suggests an antidepressant acceleration effect of liothyronine and a mood-stabilizing effect of levothyroxine in bipolar I disorder patients.

Various hormones of the hypothalamic pituitary thyroid axis have been used in the treatment of mood disorders. These include the highly purified TSH preparation thyrotropin (Thytropar), TRH, liothyronine, and levothyroxine. Thyroid hormones have been used in several ways including as monotherapy for depression, combined therapy with an antidepressant to accelerate the therapeutic effect of the antidepressant and augmentation therapy to enhance the effects of treatment after patients have failed to respond to a standard antidepressant.

Thyrotropin-Releasing Hormone (TRH) TRH, in addition to being released from the hypothalamus to regulate the thyroid axis, has behavioral effects independent of its effect on thyroid function. TRH affects various aspects of cerebral function including reversal of drug-induced sedation or anesthesia and stimulation of locomotor activity and also affects cardiorespiratory, gastrointestinal, and neurological functions. Because this peptide may stimulate the thyroid axis and may also have widespread effects on the central nervous system, it has been tested as an antidepressant. It has been used both as monotherapy and in combination with electroconvulsive treatment.

Several studies have examined the therapeutic effect of either intravenous or oral administration of TRH in patients with depression (Table 31.28-1). In one of the first placebo-controlled trials A. J. Kastin and collaborators reported a transient antidepressant effect in four of five patients who were administered 500 µg of TRH intravenously as a single dose. Prange and collaborators used a similar design to show that the transient antidepressant effect of TRH was superior to that of placebo in a small number of patients. Later studies have reported no effect or a very minimal therapeutic response to administration of either oral or intravenous TRH as monotherapy for times varying from one dose to several days or weeks. Results of oral and intravenous administration were not substantially different. These preliminary data suggest that TRH has no substantial antide pressant effect. The transient effects noted in some of these studies may be a nonspecific activating effect of the peptide rather than a specific antidepressant response. In another type of study, researchers used a randomized, double-blind, placebo-controlled, crossover design and administered 500 µg of TRH to eight depressed patients who were going to receive electroconvulsive therapy. TRH infusions before the electroconvulsive treatments yielded greater arousal and better cognitive functioning than placebo. TRH did not appear to have any direct effect on any seizure variables induced by the treatment. Although this study is of some interest, the data are preliminary and require replication. In particular, conclusions from this study are limited by the small number of patients and the question of whether this was a nonspecific effect of reducing the sedation associated with anesthesia rather than a true antidepressant response.

Number of Studies	Total N	Route of Administration	Result
3	36	Single intravenous dose	1 of 3 shows TRH > placebo
6	86	Intravenous doses over 3–4 days	1 of 6 shows transient effect
3	52	Oral doses for 7–30 days	1 of 3 shows no difference from amitriptyline

Table 31.28-1 Antidepressant Effect of Thyrotropin-Releasing Hormone (TSH)

Thyrotropin The rationale for investigating the use of thyrotropin was that it would stimulate thyroid function and thus have antidepressant action. Only one study to date has examined the antidepressant effect of thyrotropin. Prange and collaborators administered 10 IU of thyrotropin intravenously to 20 depressed women 1 day before beginning an imipramine (Tofranil) antidepressant trial. The thyrotropin-treated group had a more rapid antidepressant response than the saline-treated controls. Although these data are of considerable interest they require replication with larger samples and modern approaches to the diagnosis of depression, and the definition of response to treatment before one can conclude that thyrotropin accelerates response to antidepressants.

Liothyronine Liothyronine has been used in three ways to treat depressed patients. First, two early studies examined its use as monotherapy for the treatment of depressed patients. Although the data from these studies suggested some antidepressant activity, their findings are limited by methodological flaws including heterogenous patient samples; the lack of operational criteria to define the diagnoses of subjects in the studies; the open, uncontrolled nature of the trials; and the failure to provide systematic criteria for evaluating treatment response. As a result of these methodological limitations, liothyronine as monotherapy has not gained any clinical utility in the treatment of depression.

Second, Prange and collaborators used liothyronine to accelerate the response to tricyclic antidepressant treatment. In several studies they demonstrated that coadministration of liothyronine at the outset of the antidepressant trial with imipramine gave a shorter lag in onset of antidepressant response than that with placebo controls. This acceleration effect was seen in women but not in men. Several other groups replicated these findings, although some studies also gave negative results. All these studies, both positive and negative, suffer from methodological limitations including poorly defined patient samples, suboptimal antidepressant trials by current standards, and small sample sizes. Since the lag in onset of antidepressant response remains one of the great therapeutic challenges, it is puzzling that these early studies have not been followed up to determine whether liothyronine has a place in accelerating antidepressant response.

Finally, liothyronine has been used to augment the response to tricyclic drugs in patients who fail to respond to antidepressant treatment (Table 31.28-2). Both open and controlled studies using a variety of antidepressants have consistently shown that about 55 to 60 percent of patients who are nonresponsive to tricyclic drugs respond within 2 to 3 weeks after the addition of 25 to 50 µg of liothyronine. A recent meta-analysis of these studies suggest that liothyronine may increase response rates and decrease severity of depression scores in patients refractory to tricyclic antidepressant therapy. Patients treated with liothyronine augmentation were twice as likely to respond as controls. Although the meta-analysis demonstrated moderately large improvements in depression scores, the authors concluded that the generally uneven quality of the studies limited conclusions from the meta-analysis.

Study	N	Liothyronine Dosage (mg/day)	Design	Result
Earle, 1970	25	25	Open	14/25
Ogura et al., 1974	44	20-30	Open	29/44
Banki, 1975	52	20-40	Open	39/52
Banki, 1977	10	20	Open	2/10
Tsutsui et al., 1979	11	1-25	Open	10/11
Goodwin et al., 1982	12	25-50	Double-blind, placebo-controlled	8/12
Schwarz et al., 1984	8	25-50	Open	4/8
Gitlin et al., 1987	16	25	Randomized, placebo-controlled, double-blind, crossover	Liothyronine = placebo
Thase et al., 1989	20	25	Open	5/20
Joffe and Singer, 1990	38	37.5	Randomized, double-blind, TA-controlled	9/17
Joffe et al., 1993	51	37.5	Randomized, double-blind, placebo-controlled	10/17 Liothyronine = lithium > placebo

Table 31.28-2 Triiodothyronine Augmentation of Antidepressants

The findings from the open and the controlled studies of liothyronine augmentation show generally good agreement. One study, by Michael Gitlin and collaborators, failed to find a significant difference between liothyronine and placebo in the potentiation of imipramine (Tofranil). The authors, however, used a 2-week, double-blind, crossover design that is problematic in evaluating antidepressant treatment response. However, a later study found liothyronine to be more effective than placebo and comparable to lithium. Despite this study, lithium is still widely accepted as a useful clinical strategy for augmentation of antidepressants in nonresponders, whereas use of liothyronine has been met with scepticism and is less commonly used. The studies to date (Table 31.28-2) as well as the meta-analysis suggest that at least for some patients, liothyronine may be an effective means of treating antidepressant nonresponders.

Liothyronine has also been used to augment the response to electroconvulsive treatment (ECT). In a preliminary study using a double-blind, placebo-controlled design, Robert Stern and collaborators observed that fewer ECT treatments were required for antidepressant response with liothyronine augmentation than with placebo. Furthermore, patients who received liothyronine had less cognitive disturbance associated with ECT. It is unclear whether this preservation of cognitive function was a direct effect of the thyroid hormone or was related to the reduced number of ECT treatments required for response with liothyronine augmentation.

Levothyroxine Most studies of thyroid hormone augmentation of antidepressants have involved liothyronine. It has generally been assumed that because T_3 is largely derived from the extrathyroidal conversion of levothyroxine and that liothyronine would have comparable efficacy in augmenting antidepressant response. Only one study to date has directly compared the two thyroid hormones. This study found liothyronine superior to levothyroxine in augmenting antidepressant response. Clearly, this finding requires replication. Furthermore, the absence of a placebo control in this study limits the conclusions that can be drawn. Another study reported that both liothyronine and levothyroxine were effective augmentation agents in 21 patients who had failed treatment with a tricyclic drug. Seven patients responded to augmentation with levothyroxine. However, five of these seven had evidence of subclinical hypothyroidism as determined by a maximum thyrotropine response to TRH exceeding 25 µ/mL at baseline prior to thyroid hormone augmentation. Therefore it is unclear whether in this study, levothyroxine had an antide pressant augmentation effect or whether the therapeutic response was the result of T_4 replacement therapy for subclinical hypothyroidism.

A number of studies have evaluated whether levothyroxine treatment is effective in patients with bipolar I disorder. The rationale for the use of levothyroxine in this group of patients came largely from early data that suggested a higher prevalence of clinical and subclinical hypothyroidism in patients with bipolar disorders, particularly rapid-cycling. Although the specificity of hypothyroidism in rapid cyclers remains unclear, several open studies document the efficacy of high dosages of levothyroxine in patients with this disorder. Although the earlier studies used very large dosages of levothyroxine (300 to 500 µg a day), later studies used dosages that yielded serum levothyroxine concentrations slightly above the upper range of normal. Although these studies conclude that levothyroxine may have mood-stabilizing effects, particularly when used as an adjunct to other mood stabilizers such as lithium, the study findings are limited by their small sample sizes and open, uncontrolled designs. It has been suggested that the response of patients with bipolar I disorder to high dosages of levothyroxine does not appear related to their pretreatment serum levothyroxine levels, implying that the efficacy of levothyroxine may not be related to correction of underlying subclinical hypothyroidism. Preliminary data suggest that levothyroxine may also be effective in non–rapid-cycling bipolar disorders.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Thyroid hormones have ubiquitous effects on all organ systems. Until approximately 20 years ago mature brain was considered largely impervious to effects of thyroid hormone. Evidence now suggests that not only does the adult brain respond to thyroid hormones but it may be particularly sensitive to changes in concentrations of circulating T_4 and T_3. Preliminary evidence suggests that carefully regulated homeostatic mechanisms maintain brain thyroid hormone concentrations within very narrow limits.

THERAPEUTIC INDICATIONS

None of the thyroid hormones have received formal approval by regulatory authorities for use in the treatment of depression. However, a limited body of empirical data and widespread clinical experience suggest that these hormones may be useful in the treatment of patients with refractory depression and possibly refractory, rapid-cycling bipolar disorders. A substantial number of patients who receive optimum antidepressant treatment fail to respond or have only a partial response to therapy. Liothyronine remains a viable option for such patients along with other augmentation strategies such as lithium, the serotonergic agents such as buspirone (BuSpar) and pindolol (Visken) as well as the psychostimulants, among others. At present, no specific clinical or biochemical features are of use in determining which patients are more likely to respond to any one augmentation strategy. This applies as much to liothyronine as to any of the other augmentation treatments. Substantially more clinical experience and, particularly, empirical data are required to determine which treatment option should be used at what stage of treatment in individual patients. Such data will eventually lead to the rational development of algorithms for the treatment of refractory depression. Until this occurs, liothyronine remains one of many possibilities to consider for use in patients who fail to respond to their first antidepressant trial. Although its efficacy has been best demonstrated with tricyclic drugs, evidence suggests that liothyronine augments the effect of the monoamine oxidase inhibitors, the serotonin reuptake inhibitors, and several of the newer classes of antidepressant.

PRECAUTIONS AND ADVERSE REACTIONS

When liothyronine is used as an augmentation treatment, symptoms of thyroid excess or thyrotoxicosis are very unlikely. A dosage of 25 µg a day is below the replacement dosage for this hormone; thus symptoms of thyroid excess are extremely uncommon. This also applies to dosage of 50 µg a day, although this dosage is closer to the daily T_3 production rate and thus symptoms of thyroid excess could occur.

For the most part liothyronine is very well tolerated. Transient headache is the most commonly reported adverse effect, and this usually subsides spontaneously after 1 or 2 days. Allergic reactions are extremely rare. Liothyronine should be used with great caution in patients with certain medical disorders including endocrine disease and cardiovascular illness and should be avoided in pregnant patients.

If patients respond to liothyronine augmentation and it is used for a prolonged period of time, added caution should be observed in its administration. In particular, long-term thyroid hormone treatment may be associated with bone demineralization resulting in osteoporosis in women. Although this association is still controversial, one should approach the long-term use of liothyronine for the treatment of depression with caution.

DOSAGE AND ADMINISTRATION

Liothyronine is available in 5- and 25-µg tablets. It is usually administered as a single dose in the morning to avoid insomnia. There is a rationale for divided administration of liothyronine because of its short half-life, but a single daily dose is commonly used to improve compliance. Levothyroxine is available in a variety of strengths. It is administered in dosages that produce a circulating levothyroxine concentration exceeding the upper range of normal for the patient with bipolar I disorder. These dosage recommendations apply to healthy adults and should be modified for older patients. Levothyroxine should be used cautiously in elderly patients, particularly if there are concomitant medical disorders. There is no evidence for the efficacy of thyroid hormones in the treatment of mood disorders in

children.

SUGGESTED CROSS-REFERENCES

Psychiatric aspects of endocrine and metabolic disorders are discussed in Section 25.6. Mood disorders are covered in Chapter 14.

SECTION REFERENCES

*Aronson R, Offman HJ, Joffe RT, Naylor CD: Triiodothyronine augmentation in the treatment of refractory depression: A meta-analysis. Arch Gen Psychiatry *53:*842, 1996.

Banki CM: Triiodothyronine in the treatment of depression. Orv Hetil *116:*2543, 1975.

Banki CM: Cerebrospinal fluid amine metabolites after combined amitriptyline-triiodothyronine treatment of depressed women. Eur J Clin Pharmacol *11:*311, 1977.

Bauer MS, Whybrow PC: Rapid-cycling bipolar affective disorders, II: Treatment of refractory rapid cycling with high-dose levothyroxine: A preliminary study. Arch Gen Psychiatry *47:*435, 1990.

*Bauer M, Hellweg R, Graf KJ, Baumgartner A: Treatment of refractory depression with high-dose thyroxine. Neuropsychopharmacology *18:*444, 1998.

*Baumgartner A, Bauer M, Hellweg R: Treatment of intractable non-rapid cyclic bipolar affective disorder with high-dose thyroxine: An open clinical trial. Neuropsychopharmacology *10:*183, 1994.

Coppen A, Whybrow PC, Noguera R: Comparative antidepressant value of carpal L-tryptophan and imipramine with and without attempted potentiation by liothyronine. Arch Gen Psychiatry *26:*234, 1972.

Cowdlry RW, Wehr TA, Zis AP, Goodwin FK: Thyroid abnormalities associated with rapid-cycling bipolar illness. Arch Gen Psychiatry *40:*414, 1983.

Crantz FR, Silver JE, Larsen PR: An analysis of the sources and quantity of 3,5,3-triiodothyronines specifically bound to nuclear receptors in rat cerebral cortex and cerebellum. Endocrinology *110:*367, 1982.

Earle BV: Thyroid hormone and tricyclic antidepressants in resistant depression. Am J Psychiatry *126:*1667, 1970.

Feldmesser-Reiss EE: The application of triiodothyronine in the treatment of mental disorders. J Nerv Ment Dis *127:*540, 1958.

Flach FF, Celian CI, Rawson RW: Treatment of psychiatric disorders with triiodothyronine. Am J Psychiatry *114:*841, 1958.

Gitlin MJ, Winer H, Fairbanks L, Hershman JM, Friedfeld N: Failure of T_3 to potentiate tricyclic antidepressant response. J Affect Disord *13:*267, 1987.

Goodwin FK, Prange AJ, Post RM, Muscettola J, Lipton MA: Potentiation of antidepressant effects by L-triiodothyronine in tricyclic non-responders. Am J Psychiatry *139:*34, 1982.

*Jackson IM: The thyroid axis and depression. Thyroid *8:*951, 1998.

Joffe RT, Singer W: A comparison of triidothyronine and thyroxine in the potentiation of tricyclic antidepressants. Psychiatry Res *32:*241, 1990.

Joffe RT, Singer W, Levitt AJ, McDonald C: A placebo-controlled comparison of lithium and triiodothyronine augmentation of tricyclic antidepressants in unipolar refractory depression. Arch Gen Psychiatry *50:*387, 1993.

Kastin AJ, Ehrensing RH, Schalch DS: Improvement in mental depression with decreased thyrotropin response after administration of thyrotropin releasing hormone. Lancet *2:*740, 1972.

Kirkegaard C: The thyrotropin response to thyrotropin-releasing in endogenous depression. Psychoneuroendocrinology *6:*189, 1981.

Kirkegaard C, Faber J: The role of thyroid hormones in depression. Eur J Endocrinol *138:*1, 1998.

*Laseer RA, Baldessarini RJ: Thyroid hormones in depressive disorders: A reappraisal of clinical utility. Harv Rev Psychiatry *4:*291, 1998.

Loosen PT: Hormones of the hypothalamic-pituitary-thyroid axis: A psychoneuroendocrine perspective. Pharmacopsychiatry *19:*401, 1986.

Ogura C, Okuma T, Uchida Y, Imi S, Yogi H: Combined thyroid (triiodothyronine)-tricyclic antidepressive treatment in depressive states. Folia Psychiatr Neurol Jpn *28:*179, 1974.

Prange AJ Jr, Wilson IC, Lara PP, Lipton M: Effects of thyrotropin-releasing hormone and depression. Lancet *2:*999, 1972.

Prange AJ Jr, Wilson IC, Raybron AM, Lipton M: Enhancement of the imipramine antidepressant activity by thyroid hormone. Am J Psychiatry *126:*457, 1969.

*Rudas S, Schmitz M, Pichler P, Baumgartner A: Treatment of refractory chronic depression and dysthymia with high-dose thyroxine. Biol Psychiatry *45:*229, 1999.

Schwarcz G, Halaris A, Baxter L, Escobar J, Thompson M, Young M: Normal thyroid function in desipramine non-responders converted to responders by the addition of L-triiodothyronine. Am J Psychiatry *141:*1614, 1984.

*Shelton RC: Treatment options for refractory depression. J Clin Psychiatry *60*(Suppl):57, 1999.

Spoov J, Lahdelma L: Should thyroid augmentation precede lithium augmentation—A pilot study. J Affect Disord *49:*235, 1998.

Stancer HC, Persad E: Treatment of intractable rapid-cycling manic-depressive disorder with levothyroxine. Arch Gen Psychiatry *39:*311, 1982.

Stern RA, Nevels CT, Shelhorse ME, Proajska ML, Mason GA, Prange AJ Jr: Antidepressant memory effects of combined thyroid hormone treatment and convulsive therapy: Preliminary findings. Biol Psychiatry *30:*623, 1991.

Surks MI, Oppenheimer JH: Concentration of L-thyroxine and L-triiodothyronine specifically bound to nuclear receptors in rat live in kidney: Quantitative evidence favouring a major role of T_3 and thyroid hormone action. J Clin Invest *60:*555, 1977.

Targun SD, Greenberg RD, Harmon RL, Kessler K, Salerian AJ, Fram DH: Thyroid hormone and the TSH stimulation test in refractory depression. J Clin Psychiatry *45:*345, 1984.

Thase ME, Kupfer TG, Jarrett DB: Treatment of imipramine-resistant recurrent depression: I: An open clinical trial of adjunctive L-triiodothyronine. J Clin Psychiatry *50:*385, 1989.

Tsutsui S, Yamazaki Y, Namba T, Tsushima M: Combined therapy of T_3 and antidepressants in depression. J Int Med Res *7:*138, 1979.

Wheatley D: Potentiation of amitriptyline by thyroid hormone. Arch Gen Psychiatry *26:*229, 1972.

Wilson IC, Prange AJ Jr, McLean TK, Rabon AM, Lipton AM: Thyroid hormone enhancement of imipramine in non-retarded depressions. N Engl J Med *282:*1063, 1979.

31.29 TRAZODONE

STEVEN J. GARLOW, M.D., PH.D. AND CHARLES B. NEMEROFF, M.D., PH.D.

Trazodone (Desyrel) was the first of the second-generation (“atypical”) antidepressants to be developed and marketed. Trazodone was atypical because of its lack of structural similarity to heterocyclic antidepressants. Trazodone was developed in response to the theory that major depressive disorder was caused by a paradoxical response to unpleasant and painful experiences. The compound has antinociceptive properties, and in the course of study of its preclinical pharmacology, it was found effective in animal models specifically developed to test this theory of depression. Trazodone is not active in conventional animal models of major depression where classic tricyclic drugs are effective, and tricyclic drugs are much less active in the animal model in which trazodone displays activity.

The compound was synthesized in Italy in the mid-1960s and was approved for clinical use in Europe in the early 1970s. Trazodone entered clinical trials in the United States in the late 1970s and received Food and Drug Administration (FDA) approval for use in major depression in 1981. Because of its unique pharmacological profile and relative safety, it rapidly captured a large share of the American antidepressant market. With the advent of additional sec ond- and third-generation antidepressants (e.g., selective serotonin reuptake inhibitors [SSRIs], selective-noradrenaline reuptake inhibitors [SNRIs]), the total number of trazodone prescriptions has declined.

Trazodone is a triazalopyridine derivative that had a unique chemical structure until the development of nefazodone (Serzone), a structurally related compound. Trazodone has complex and diffuse pharmacological actions in the central nervous system (CNS), with antagonist and agonist activity in the serotonergic system and antagonist activity in the adrenergic system. Trazodone is a weak inhibitor of serotonin reuptake, but it is a potent antagonist at serotonin (5-hydroxytryptamine [5-HT]) type 2A (5-HT_{2A}) and 5-HT_{2C} receptors. There is at least one active metabolite of trazodone, *m*-chlorophenylpiperazine (mCPP), which acts as an agonist at 5-HT_{2C} receptors and may act as an antagonist at 5-HT_{2A} receptors. The antidepressant property of trazodone is hypothesized to be due to both its 5-HT_2 antagonist and serotonin reuptake blocking activities. Trazodone has little or no effect on norepinephrine or dopamine reuptake. Trazodone has low affinity for muscarinic cholinergic receptors and therefore causes little in the way of anticholinergic adverse effects.

CHEMISTRY

Trazodone is a triazalopyridine derivative with some structural similarity to alprazolam (Xanax) (Fig. 31.29-1). The antidepressant activity may be due in part to the triazalo ring structure on the molecule. The chemical designation of trazodone is 2-[3-[4-(3-chlorophenyl) - 1 - piperazinyl]propyl] - 1,2,4 - triazalo[4,3 - a]pyridin-3(2H)one hydrochloride, with an empirical formula of $C_{19}H_{22}ClN_5 \cdot HCl$, and a molecular weight of 408.3. The compound is soluble in water and has a pKa of 6.7.

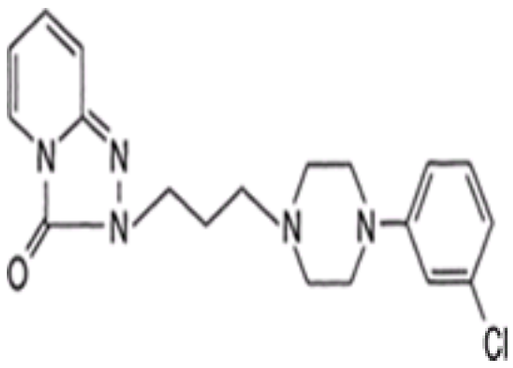

FIGURE 31.29-1 Molecular structure of trazodone.

PHARMACOLOGY

Pharmacokinetics Trazodone is rapidly and efficiently absorbed after oral administration. Peak plasma concentrations in fasting subjects are reached 1 hour after administration. Food slows the rate of absorbance but increases the total amount of drug absorbed, with peak plasma concentrations occurring 2 hours after administration and lower peak concentration values. Taking trazodone with food may minimize dose-dependent adverse effects because of the slower rate of absorbance and the lower peak serum concentration. Steady-state serum concentrations are achieved at 36 hours on a 50-mg twice a day regimen, with new steady-state concentrations being achieved in 12 hours when the dose is increased from that initial regimen. Within the dosage range of 50 to 150 mg twice a day, trazodone displays linear pharmacokinetic properties. There may be significant interindividual variation in the serum concentration of trazodone in subjects receiving the same dose. It is not clear whether there is a relationship between the serum concentration of trazodone and therapeutic response.

Trazodone is 85 to 95 percent protein bound in the blood and does not preferentially distribute to any one tissue. Both the parent compound and mCPP cross the blood-brain barrier. The drug is metabolized in the liver by the cytochrome (CYP) P450 2D6 isoenzyme. Trazodone does enter breast milk of lactating women at a low level, with the concentration in milk being approximately 10 percent of that in serum.

The half-life of the drug is from 5 to 9 hours, depending on the age and hepatic function of the patient, typically with a longer half-life in elderly persons. The elimination of trazodone is biphasic, with the initial phase having a half-life of 3 to 6 hours and the second phase having a half-life of 5 to 9 hours. Trazodone is metabolized in the liver by hydroxylation, oxidation and N-oxidation, and the hepatic metabolites are excreted by the kidneys. The metabolite mCPP has a longer half-life than the parent compound (14 hours vs. 5 to 9 hours) and is cleared more slowly.

Pharmacodynamics Trazodone has a variety of pharmacological actions measured with in vivo and ex vivo assays. Trazodone does block transport of serotonin into synaptosomal membrane preparations but has little impact on the platelet serotonin transporter of patients treated with the drug. There are minimal changes in the concentration of serotonin in the platelets of patients treated with trazodone. Trazodone has significant antagonist activity at the 5-HT_{2A} and receptor sites assayed with radioligand binding techniques on synaptosomal membrane preparations. Trazodone also inhibits a variety of behaviors in rodents produced by agonist action on the 5-HT_{2A} and 5-HT_{2C} receptor sites. Chronic administration of trazodone downregulates the 5-HT_2 receptors, causing a decrease in the number of receptor binding sites in the CNS. This effect has been hypothesized as central to the antidepressant action of the drug, because a number of other antidepressants cause downregulation of 5-HT_2 receptors.

The metabolite mCPP has both antagonist and agonist activities at a number of serotonin receptors. This compound is an antagonist at the 5-HT_{2A} receptor site but

an agonist at the 5-HT_{2C} receptor. mCPP may also be an agonist at other serotonin receptor sites. Administration of mCPP to experimental animals results in behavioral activation and enhancement of anxiety-provoking stimuli. Administration of mCPP to humans also has anxiogenic effects.

Trazodone acts as an antagonist at peripheral a-adrenergic receptors. This adrenergic receptor blockade probably accounts for the adverse effects of orthostatic hypotension, dry mouth, and priapism. Chronic administration of trazodone also downregulates b-adrenergic receptors in the CNS of experimental animals. Treatment of patients with the drug causes downregulation of b-receptors on platelets and lymphocytes. This drug has little or no activity at the noradrenergic or dopamine transporter sites in membrane preparations and behavioral assays. Trazodone has little affinity for muscarinic cholinergic receptors, histamine type 1 (H_1) receptors, and dopamine receptors.

Several studies have attempted to correlate steady-state plasma concentrations of trazodone to clinical response. The results of these studies are variable. One study in geriatric patients reported a threshold of 650 ng/mL for clinical response. A second study of geriatric patients reported an average serum concentration of 1474 ng/mL but indicated that subjects in the study who responded to trazodone had lower levels than the nonresponders. The authors concluded that this concentration may represent the top of the therapeutic window. A study of depressed patients in a primary care setting reported a threshold for response of 250 ng/mL, significantly lower than the thresholds reported in the geriatric studies. This may be due to demographic differences between the geriatric and primary care populations. Currently there has been no consistent report of a therapeutic blood level for trazodone or correlation between concentration in blood and clinical response, and there seems to be little utility in measuring trazodone blood concentrations as part of the general treatment of depression with this drug, with the exception of monitoring patient compliance.

DESIGN AND INTERPRETATIONOF CLINICAL DRUG STUDIES

Trazodone has been systematically studied for efficacy in the treatment of major depression, depression in the elderly, dysthymia, insomnia associated with depression and antidepressant treatment, schizophrenia, obsessive-compulsive disorder, anxiety disorders, eating disorders, chronic pain, geriatric cognitive disorders, developmental disorders, and male impotence. There are case reports of the use of trazodone in a large number of different psychiatric and neurological disorders. Trazodone has been studied extensively as a treatment for the depressive disorders, in which it is clearly as effective as other traditional antidepressants. The use of trazodone in the other neuropsychiatric disorders is reviewed below, though its use in these disorders has not been studied as extensively as its use in the treatment of depression.

Major Depressive Disorder The use of trazodone as an antidepressant has been studied in both inpatient and outpatient settings as well as in the primary care setting. Trazodone as a treatment for depression has been compared with imipramine (Tofranil), amitriptyline (Elavil), mianserin fluoxetine (Prozac), venlafaxine (Effexor), doxepin (Adapin), mirtazapine (Remeron), bupropion (Wellbutrin), and placebo in a variety of clinical settings and with a variety of experimental designs. In general, trazodone showed efficacy equal to the antidepressants with which it was compared, with the most significant differences between the agents being differences in their respective adverse-effect profiles and hence their tolerability. The particular subtype of depressive symptoms suffered by the subjects in the various studies also had an impact on the tolerability of trazodone. Depressed patients who experienced a great deal of anxiety and insomnia seem to benefit more from trazodone than patients suffering from significant anergia and psychomotor retardation.

Several published reviews have examined studies in which trazodone was compared with tricyclic drugs in the treatment of depression. These reviews concluded that trazodone is as effective as traditional tricyclic drugs (generally imipramine or amitriptyline) for the treatment of mild-to-moderate depression. In a meta-analysis that specifically examined those studies in which trazodone was compared with imipramine, a total of six independent reports met criteria for analysis: double-blind design, Hamilton Rating Scale for Depression (HAM-D) scores of 18 or more, and a 50 percent reduction in HAM-D score as a definition of treatment response. The meta-analysis revealed that trazodone and imipramine were equally efficacious in the treatment of depression across all six studies. Trazodone has also been compared with imipramine in the long-term treatment of moderately to severely depressed outpatients; in a double-blind study, with a 12-month blinded treatment phase, trazodone possessed equal or greater efficacy than imipramine. The authors also report continued benefit of trazodone up to 3 years after the 12-month blinded period in an open-label continuation phase.

There has been some question about the effectiveness of trazodone in patients suffering from more severe depressive disorders, as well as speculation that the anxiolytic and sedative properties of trazodone account for some or all of its antidepressant effect. In one case series in which trazodone was given to 13 endogenously depressed patients, only 3 had a positive response as defined by a 50 percent reduction in HAM-D score. This patient cohort was reported to be responsive to other antidepressant agents, and the authors concluded that the main indication for trazodone may be in treating nonendogenously depressed patients. The idea of trazodone's lack of effectiveness has been refuted in at least two different studies. In a retrospective chart review comparing 138 inpatients treated with amitriptyline with 42 inpatients treated with trazodone, the authors conclude that the trazodone-treated patients were started at dosages below the recommended dosage of 150 mg a day, with a mean starting dosage of 113 mg a day (±42 mg) and were at final dosages significantly below those reported optimal for the treatment of depression, with final mean dosage of 217 mg a day (±87.5 mg) compared with an optimal dose of 250 to 350 mg a day. The authors conclude that the apparent lack of response to trazodone may be due in part to inadequate dosing at a subtherapeutic level. In a meta-analysis that compared the effectiveness of imipramine with that of trazodone, bupropion, and fluoxetine in the treatment of depression, all four agents were deemed more effective than placebo. This analysis had very strict criteria including placebo control, double-blind design, HAM-D as a dependent measure, nongeriatric adults meeting DSM-III or Research Diagnostic Criteria (RDC) criteria for major depression, and reported means and standard deviations for responses. The analysis did not support the hypothesis that imipramine was more effective than the newer agents. The authors concluded that there are no differences in the efficacy of all four agents in the treatment of depression.

Review of all available studies of trazodone as a treatment for major depression, as well as the published meta-analyses, leads to several conclusions about the efficacy and utility of trazodone in the treatment of depression. First, trazodone is as efficacious as all other antidepressants with which it has been compared for the treatment of major depression. Second, reports of lack of effectiveness of trazodone are in part due to subtherapeutic dosing of the agent, which may be caused by intolerability of the agent in the patients deemed unresponsive. Third, the adverse-effect profile of trazodone affects its perceived tolerability and therefore its perceived utility for treating depression. Specifically, patients experiencing a great deal of anxiety, agitation, and insomnia will better tolerate the sedating effects of trazodone and in fact may benefit from these effects, whereas patients suffering from a great deal of anergia, somnolence, and psychomotor retardation may find these side effects intolerable and therefore not be amenable to treatment with trazodone.

Major Depressive Disorder in Geriatric Patients Trazodone was initially embraced with a great deal of enthusiasm for the treatment of major depression in elderly patients because of its atypical pharmacological profile, with low anticholinergic activity and negligible cardiotoxicity. As in the case of nongeriatric depression, trazodone has been compared with a number of conventional, tricyclic drugs and a number of nontricyclic agents for the treatment of major depressive disorder in elderly patients. Generally speaking, in geriatric populations as in nongeriatric populations, trazodone is an effective antidepressant agent. Trazodone is as effective as the agents with which it was compared, and the symptom complex of the patients determined the tolerability of the agent. Patients treated with trazodone reported few side effects, and in some studies reported significantly fewer side effects than patients treated with conventional agents. In two different studies comparing trazodone with amitriptyline and mianserin for the treatment of major depression in elderly patients, all three agents were found equally efficacious. Pa tients treated with trazodone had significantly fewer anticholinergic side effects than patients treated with amitriptyline and significantly fewer cardiovascular adverse effects than patients treated with mianserin.

The most frequently reported side effects of trazodone in the various studies were dizziness, orthostatic hypotension, and drowsiness or sedation. It is unclear whether the reported dizziness is in fact a manifestation of orthostatic hypotension, but this hypotensive effect can be worsened by the concomitant administration of antihypertensive agents. The hypotensive effect of trazodone is related to peak serum concentration and can be worsened by taking the drug on an empty stomach and lessened by taking the drug with food. Adjustment of the antihypertensive regimen may also decrease the occurrence of orthostatic hypotension.

As with nongeriatric depression, matching symptoms and adverse effects can improve tolerability of the antidepressant and thus improve compliance and therapeutic response. Geriatric patients who experience a great deal of anxiety and insomnia as part of their depression tolerate the sedating qualities of trazodone better than patients who experience psychomotor retardation. In fact, the sedating and anxiolytic properties of trazodone may contribute significantly to its utility in the treatment of depression in the elderly population because anxiety (often reported as “nerves”) and insomnia are common complaints of elderly depressed patients.

Insomnia Trazodone has been studied as a hypnotic agent both to treat the insomnia associated with other antidepressants and to treat the insomnia that accompanies major depressive disorder in many patients. In an open-label study, patients with major depression who had been treated with monamine oxidase inhibitors (MAOIs) and who had developed insomnia after MAOI treatment was initiated were treated with trazodone in doses of 50 to 75 mg a night. In this study, 20 of 21 patients treated with trazodone had a positive response and reported significantly improved sleep and prolonged duration of sleep. In this same study, depressed patients who had insomnia preceding treatment with the MAOI and thus unrelated to the MAOI were also treated with trazodone. Of those suffering

insomnia principally as a symptom of their depression, 26 of 27 patients treated with trazodone reported significant improvement in sleep quality and duration. Trazodone as a treatment for antidepressant-induced insomnia was examined in double-blind type studies. In one study, depressed patients treated either with fluoxetine or bupropion who were experiencing insomnia were given either trazodone or placebo; 67 percent of the patients treated with trazodone reported a significant improvement in sleep quality and duration, while only 13 percent of the placebo-treated cohort reported positive effects on sleep. This study clearly demonstrates efficacy of trazodone in the treatment of antidepressant-induced insomnia in patients with major depressive disorder.

The effects of trazodone on sleep architecture in depressed patients has been studied with polysomnographic techniques with somewhat contradictory results. In one single-blind study of depressed patients treated with trazodone at antidepressant doses, 150 to 400 mg a day, there was a significant improvement in sleep quality and duration including decreased persistent sleep latency, increased total sleep time, increased rapid eye movement (REM) latency, and improved sleep efficiency. This study reported no change in total REM duration. A second single-blind study of similar design had essentially opposite results, with significant REM suppression but no improvement in any of the other sleep parameters with the exception of increased REM latency.

Two main conclusions can be drawn from the data on the effects of trazodone on sleep. First, trazodone is effective as a sedative-hypnotic agent in patients suffering from the “primary” insomnia of major depression, as well as in those suffering from the “secondary” insomnia resulting from treatment with activating antidepressants. Second, the specific effects of trazodone on sleep architecture remain obscure because contradictory results have been reported from similar studies. The one effect that trazodone does seem to exert on sleep architecture is prolongation of REM latency.

Schizophrenia Trazodone may be expected to have both antipsychotic and antidepressant actions on the basis of its known pharmacological actions. The new generation of antipsychotic agents (e.g., clozapine [Clozaril], olanzapine [Zyprexa], risperidone [Risperdal]) all have significant antagonist action at the 5-HT_{2A} receptor. This is thought to be an essential component of their antipsychotic action and is part of the basis for their designation as atypical. Trazodone has significant antagonist action at the 5-HT_{2A} receptor and therefore may be expected to behave similarly to the atypical antipsychotic agents.

The efficacy of trazodone as a treatment for schizophrenia was examined in a double-blind, multiple crossover study that compared the efficacy of trazodone, haloperidol (Haldol), and amitriptyline in patients suffering from schizophrenia or major depression. Trazodone displayed no antipsychotic action in this study. Trazodone was found to be an effective antidepressant agent and to pose less risk to schizophrenic patients; schizophrenic patients treated with trazodone had significantly fewer instances of antidepressant-provoked psychotic symptoms than did patients treated with amitriptyline. The depressive symptoms that many of the schizophrenic patients displayed responded to treatment with trazodone, but the core psychotic symptoms did not.

In a separate study, trazodone was examined as an adjunctive therapy for treating the negative (deficit) symptoms of schizophrenia. In this study, 47 schizophrenic patients who were in a stable pattern of typical antipsychotic treatment were randomly assigned to treatment with trazodone or placebo. The patients were judged neither floridly psychotic nor depressed at entry into the randomized phase of the study. Treatment with trazodone significantly reduced the severity of two out of three measures of negative symptoms while not increasing the severity of the positive symptoms of psychosis. The improvements in negative symptoms while significant, were modest, with a 10 to 15 percent reduction in severity. Schizophrenic patients treated with typical antipsychotic agents do have some modest improvement in negative symptoms when trazodone is added to their treatment regimen. This seems to be a low-risk approach, but with the emergence of newer serotonin-dopamine antagonists such as olanzapine, the use of dopamine receptor antagonists in conjunction with trazodone would appear to be a secondary or backup strategy in the treatment of schizophrenia.

Obsessive-Compulsive Disorder The use of trazodone as a first-line treatment for obsessive-compulsive disorder has not been established. Trazodone has been reported to be efficacious in the treatment of obsessive-compulsive disorder in a number of open-label studies and case reports. In one study of 9 patients who had failed clomipramine or clomipramine plus lithium (Eskalith) treatment who were subsequently treated with trazodone, a slight but significant improvement was reported. Three of these patients were reported to have a very favorable response. In a separate open-label study, eight patients with obsessive-compulsive disorder were treated with trazodone and six were reported with significant reduction of obsessional symptoms at 4 and 10 weeks of treatment. In contrast to these studies, an open label trial in which trazodone in combination with tryptophan was administered to 11 patients suffering from obsessive-compulsive disorder showed “not encouraging” results: several of the patients were unable to tolerate the regimen, and none of those who could showed any improvement.

There has been one published double-blind, placebo-controlled study of trazodone as a treatment for obsessive-compulsive disorder. In this study, 21 patients were entered into a double-blind, parallel-design study in which trazodone was compared with placebo. There were no baseline differences in obsessive-compulsive disorder or depressive symptoms between the trazodone and placebo groups. Seventeen patients completed 10 weeks of treatment; there were no differences between the trazodone and placebo groups in terms of obsessive symptomatology after the 10 weeks. The obvious conclusion from this study is that trazodone possesses no efficacy in the treatment of obsessive-compulsive disorder. Unfortunately this was a rather small study and was not controlled with a known treatment arm (e.g., clomipramine (Anafranil), fluoxetine, paroxetine (Paxil)). Within the limitations of this study, there was no difference between trazodone and inactive placebo in the treatment of obsessive-compulsive disorder. Whether trazodone is an effective agent in the treatment of this disorder remains unresolved.

Anxiety Disorders Trazodone has been examined as a potential treatment for both generalized anxiety disorder and panic disorders. In these studies, investigators tried to separate patients who were suffering from these conditions from patients suffering from anxiety and panic comorbid with major depression, to assess directly the efficacy of trazodone as a treatment for generalized anxiety disorder and panic. The results of these studies have been somewhat mixed and contradictory.

In one report, 230 patients diagnosed with generalized anxiety disorder that was not complicated by depression or panic disorder were randomized to treatment with either imipramine (143 mg a day mean dosage), trazodone (255 mg a day mean dosage), or diazepam (Valium) (26 mg a day mean dosage) in a placebo-controlled, double-blind study. The patients fulfilled diagnostic criteria for generalized anxiety disorder by DSM-III standards, were excluded if they had a diagnosis of depression or panic disorder, and had a Hamilton Anxiety score of at least 18. The patients were treated for 8 weeks in family practice, private, and academic psychiatric settings. A bimodal treatment response was noted. Patients treated with diazepam showed significant improvement in the first 2 weeks of treatment. The somatic symptoms of anxiety were the most responsive to diazepam during the first 2 weeks of treatment. From week 3 through 8, the patients treated with imipramine and trazodone showed significant improvement, with those treated with trazodone having responses equal to those of patients treated with diazepam, while those treated with imipramine had slightly better responses than those treated with diazepam. The symptoms noted to be more responsive to antidepressant treatment were the psychic symptoms of tension, apprehension, and worry. Among those who completed the study, moderate to marked improvement was registered for 73 percent of patients treated with imipramine, 69 percent of those treated with trazodone, and 66 percent of those treated with diazepam. Clearly, this study indicates that trazodone and imipramine have important roles to play in the treatment of generalized anxiety disorder.

Somewhat contradictory results have been reported for trazodone in the treatment of panic disorder. In a single-blind trial in 11 patients with either panic disorder or panic disorder with agoraphobia, treatment with trazodone resulted in significant improvement at both 4 and 8 weeks of treatment. A different result was reported in a larger double-blind study in which 74 patients with panic disorder were randomized to treatment with either imipramine, alprazolam, or trazodone. Patients were blindly begun on active treatment after 3 weeks of placebo washout. The patients were treated for 8 weeks with active compound. A bimodal treatment response was reported in this study. Patients treated with alprazolam had significant improvement during the first week of treatment, while patients treated with imipramine exhibited a positive therapeutic response by the fourth week of treatment. Both imipramine and alprazolam were reported effective in treating symptoms of generalized anxiety, panic attacks, and phobic avoidance. Trazodone, in contrast, was reported ineffective for panic disorder in addition to being poorly tolerated. Only 17 of the trazodone-treated patients completed 4 weeks of treatment, and of these only 2 were considered to have had a good response. Thus it is unclear whether trazodone is effective for the treatment of panic disorder with or without agoraphobia.

Eating Disorders Treatment of eating disorders is one of the most challenging fields of psychiatry. Many different agents, including a variety of antidepressants, have been studied as potential treatments for both anorexia nervosa and bulimia nervosa. Patients suffering from bulimia nervosa have been reported to respond to a number of antidepressant agents including tricyclic drugs, fluoxetine, and MAOIs. While trazodone has some efficacy in the treatment of bulimia nervosa, it has not been demonstrated to have efficacy equal to the aforementioned agents. In an open-label trial, 13 patients who fulfilled DSM-III criteria for bulimia nervosa were treated with trazodone. Three of these patients dropped out of the study before receiving 4 weeks of active treatment. The remaining 10 subjects were treated for an average of 6.9 weeks with a mean maximum dosage of 410 mg a day (range, 250 to 600 mg). Six of the 10 evaluable patients had significant improvement in bulimic symptoms, with the rate of binging and purging in 4 of the subjects reduced to 0 and in the other 2 improved patients, reduced 50 to 90 percent. The only side effects reported were morning drowsiness and headache. Interestingly, the trazodone-treated patients did not experience significant weight gain, which may prove advantageous in the long-term management of these patients.

In a double-blind, placebo-controlled study, 42 women with revised third edition of DSM (DSM-III-R) bulimia nervosa, were randomized to 6 weeks of treatment with either trazodone or placebo. Trazodone was superior to placebo the frequency of binge eating and purging decreased in the trazodone-treated group. The patients

who received trazodone also reported a subjective sense of improvement. Overall, 40 percent of the patients who received trazodone had a partial or complete remission of bulimic symptoms. This is a rather modest response rate compared with that of other antidepressants, whose rate of response has been reported to be as high as 80 percent, though the “complete remission” rate for all of the different antidepressants in the treatment of bulimia nervosa is disappointingly low, generally less than 25 percent. There are case reports of bulimic patients treated with trazodone who developed delirium, but this adverse effect was not reported in the controlled trials. Thus trazodone may have a role in the treatment of bulimia nervosa, but based on the available data, it would appear to be a secondary or adjunctive role.

Chronic Pain Trazodone was initially identified as having antinociceptive properties in an experimental system designed to test a specific hypothesis of the genesis of depression. Agents that were identified with this screen were expected to possess clinically relevant antinocioceptive activity. Subsequent study of trazodone in other animal models of acute and chronic pain has yielded conflicting results. Use of trazodone in patient populations suffering from chronic pain have likewise yielded equivocal results, partially depending on the type of pain syndrome studied. There are case reports endorsing the use of trazodone to treat intractable headache, migraine, diabetic neuropathy, and other pain syndromes. Trazodone was tested as a prophylactic treatment for pediatric migraine in a double-blind, placebo-controlled, crossover study of children. This 8-month study of 40 children aged 7 through 18 had a crossover design so that patients initially treated with placebo were blindly switched to trazodone and patients initially treated with trazodone were switched to placebo; 35 patients completed the entire trial. The patients who received active compound had a significant reduction in the frequency and duration of migraine headaches. This study indicates that trazodone is an effective prophylactic treatment for pediatric migraine. There are multiple case reports documenting the efficacy of trazodone in adult migraine, but several case reports claim that trazodone treatment induced or provoked adult migraine, presumably through its mCPP metabolite.

Trazodone has been studied as a treatment for a number of peripheral pain syndromes, including chronic low back pain. In a double-blind study, 42 subjects with longstanding chronic low back pain were randomized to trazodone or placebo. The average dosage of trazodone was 201 mg a day. There were no statistically significant differences between the trazodone and placebo-treated groups in a number of self-rating and evaluator rating scales of pain. A similar result was reported for trazodone in the treatment of the dyesthetic pain syndrome resulting from traumatic spinal myelopathy. In this double-blind, placebo-controlled study, patients received either 150 mg a day of trazodone or inactive placebo during a 6-week treatment phase. There were no statistically significant differences between the trazodone- and placebo-treated groups in a number of pain measures. Significantly more patients randomized to the trazodone group dropped out of the study early because of intolerable adverse effects. The opposite result was reported in a separate study that compared trazodone with amitriptyline in the treatment of deafferentation pain, mostly due to cancer. In this double-blind study, 45 patients were randomized to receive either trazodone or amitriptyline, but there was no placebo control group. The therapeutic efficacy of the two treatments was reported to be similar. Unfortunately, because there was no placebo control group in this study, it was difficult to judge the magnitude of the therapeutic response for either treatment and whether either was better than inactive placebo. On the basis of all the available data, trazodone would appear to be ineffective in the treatment of peripheral pain syndromes.

Alzheimer's Disease and Geriatric Cognitive Disorders The use of trazodone to treat the behavioral manifestations of Alzheimer's disease and cognitive disorders has been reported in several case series. In an open-label, pilot study, 13 patients with Alzheimer's disease were treated for 10 weeks with trazodone at a dosage of 25 mg three times a day. Treatment with trazodone resulted in improvement in irritability, anxiety, restlessness, and affective disturbance. Unfortunately there was no change in Mini Mental State Examination scores nor any apparent improvement in cognitive functioning of the treated patients. This study indicates that trazodone may be helpful in alleviating some of the behavioral manifestations of Alzheimer's disease but does not treat the core cognitive deficit. Another case series on four elderly patients with severe cognitive disorders has been published. These patients had failed antipsychotic drug therapy and displayed a number of aggressive and inappropriate behaviors but did not appear to have an affective disorder. Treatment with trazodone was reported to substantially improve the behavior of all four patients. The authors relate this behavioral improvement to the “taming” effect exerted by trazodone in animal models of aggression. Less encouraging results were reported in a different case series in which seven patients with organic mental disorders were treated with trazodone. These patients also displayed a number of aggressive and inappropriate behaviors. Three of them completely ceased aggressive behavior after 4 to 6 weeks of treatment, three had no discernible change in behavior, and one patient dropped out of the study. Given trazodone's overall lack of cardiovascular toxicity and minimal anticholinergic activity, its use in behaviorally disturbed elderly patients would seem to be a reasonable option with little risk and low adverse-effect potential. A trial of trazodone may prove beneficial in achieving behavioral control in patients with dementia and other cognitive disorders.

Autistic Disorders and Other Developmental Disorders A number of case reports document the use of trazodone in the behavioral control of children and adults suffering from a number of developmental disorders. The use of trazodone has been reported in the treatment of the behavioral manifestations of pediatric and adult Down syndrome, autism, and other developmental disorders. In one open-label case series, trazodone was given to 22 hospitalized children with a variety of severe behavioral disturbances. These patients had failed other treatment regimens prior to receiving trazodone; 13 of them were judged to have benefited from trazodone treatment. There was a marked reduction in aggressive, impulsive behaviors in the responsive children. Three of the children were reported to have worsened on trazodone. Follow-up interviews with the parents of the children who had responded to trazodone treatment revealed that the beneficial activity persisted from 3 to 14 months after discharge from the hospital. Again, given the low risk of trazodone treatment, the low adverse-effect potential, and the overall safety of trazodone, a trial of its use in this type of patient seems warranted.

Sexual Dysfunction Trazodone has been systematically studied as a treatment for male impotence. There are several studies of trazodone in the treatment of primary erectile dysfunction. Between 1989 and 1994, 182 patients in a urology clinic were treated with trazodone as empirical treatment for erectile disorder, 127 of these patients were available for follow-up and retrospective analysis. The patients ranged in age from 25 to 85 years, had a number of causes for their sexual dysfunction, and received trazodone for at least 2 consecutive months. Of the patients younger than 60 years of age, 78 percent reported significant improvement in erectile ability. Patients who were older than 60 or were smokers had a poor response to treatment. Duration of erectile disturbance was also a relevant factor: those experiencing 12 months or less of erectile dysfunction had higher response rates, and those experiencing more than 60 months of dysfunction had much lower response rates. In an open-label study of 14 patients with erectile dysfunction, trazodone treatment significantly increased nocturnal erectile time, percentage of sleep with an erection, and maximal erectile rigidity. In a placebo-controlled, single-blind study of 35 patients with organic impotence, 5 weeks of treatment with trazodone (150 to 200 mg a day) resulted in significant improvement over placebo. In a self-report questionnaire, 68.6 percent of the patients reported improvement, compared with 11.4 percent of placebo-treated patients.

In a placebo-controlled, double-blind study of 79 patients who suffered from nonorganic impotence, patients were randomized to receive testosterone, trazodone, hypnosis, or placebo. The response rates were 60 percent for testosterone, 67 percent for trazodone, and 80 percent for hypnosis. The combination of trazodone with yohimbine (Yocon) as a treatment for psychogenic impotence was examined in a double-blind, placebo-controlled, crossover study of 63 patients suffering from primary sexual dysfunction. The patients were assessed for erectile function, ejaculation, interest in sex, and sexual thoughts at the end of 8 weeks of treatment and at 3- and 6-month follow-up; 55 patients completed the entire study and of these, 71 percent had a positive response. Positive responses were maintained in 58 percent of the patients at 3-month and in 56 percent at 6-month follow-up. Minor side effects were reported in 11 percent of the treated group. Trazodone was compared with ketanserin and mianserin as a treatment for erectile dysfunction in a placebo-controlled, double-blind study in which patients received 30 days of treatment. At 30 days, 62.5 percent of the patients treated with trazodone reported a positive result, compared with 19 percent of those treated with ketanserin, 31.6 percent of those treated with mianserin, and 13.6 percent of those treated with placebo. On the basis of these studies, trazodone appears to be a safe and effective agent for the treatment of psychogenic impotence, nonorganic erectile dysfunction, and possibly some types of organic erectile dysfunctions. In the absence of medical contraindications for its use, a 2- to 3-month trial of trazodone would seem to be a reasonable approach to managing patients with erectile disturbances.

Other Disorders A large number of case reports document the therapeutic use of trazodone in an assortment of psychiatric disorders. Diagnoses included in these reports are fibromyalgia; posttraumatic stress disorder; paranoid disorder; sexual paraphilias; climacteric (menopausal) symptoms; drug-seeking behavior, drug use; withdrawal symptoms of alcohol, opioid, benzodiazepines, and cocaine; and a number of anxiety of hysteric conditions. In general, none of these reports have been followed up with placebo-controlled, double-blind studies, and thus the efficacy of trazodone as treatment of any of these disorders is not known.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

Few significant adverse effects were reported for trazodone in premarketing controlled trials and in postmarketing surveillance. Many of the adverse reactions have been reported in case reports or small case series. Most of the adverse effects reported are generally benign, and no lethal reactions have been directly attributable to trazodone. The adverse effects commonly reported include drowsiness, headache, nausea, and orthostatic hypotension. Trazodone is virtually without anticholinergic effects, and one multicenter analysis found no statistically significant difference between trazodone and placebo in the incidence of anticholinergic effects. Allergic reactions to trazodone have been reported. There have been few trazodone-induced hematological reactions: limited reports of mild anemia and slightly decreased neutrophil counts. There are several case reports of agranulocytosis in patients treated with trazodone. These particular reactions appear to be idiosyncratic or allergic and are rare.

A series of 10 patients was reported to suffer from trazodone-induced peripheral edema. This was a limited reaction that disappeared with discontinuation of the drug.

The cause of this effect remains obscure, but it appears to be a rare event, occurring in substantially less than 1 percent of treated patients.

Cardiovascular System Trazodone has little or no direct action on the cardiac contraction or conduction systems in experimental animals and in patients. There are individual case reports of trazodone causing significant cardiac conduction problems including ventricular tachyarrhythmias, bradyarrhythmias, and complete heart block. These reports are mostly from patients with cardiac disease preexisting initiation of trazodone therapy. In a specific double-blind, placebo-controlled study of trazodone in depressed cardiac patients, there were no arrhythmias reported in the trazodone-treated group. These results were confirmed in at least one additional study. Trazodone use in patients without preexisting ventricular arrhythmias appears to be safe, with little or no risk of inducing cardiac arrhythmias. Trazodone is not absolutely contraindicated in patients with known cardiac disease, but its use would seem unwarranted in patients with ventricular arrhythmias, given the vast array of other available antidepressants.

The most commonly reported cardiovascular adverse effect of trazodone is orthostatic hypotension. The incidence of orthostatic hypotension is directly related to the serum concentration of the agent and generally follows oral ingestion by 4 to 6 hours. This hypotensive effect can be blunted by taking the drug with food, which slows the rate of absorption and lowers the peak serum concentration. Orthostatic hypotension can be severe enough to cause syncope. This usually occurs when a large dose is taken without food or with the concurrent use of antihypertensive agents. Trazodone can potentiate the action of antihypertensive agents. Adjustment of the dosage of antihypertensive agents should be considered when coprescribing trazodone.

Genitourinary Effects Trazodone has been well documented to cause priapism, a painful unremitting erection, in a small percentage of patients. This is a true medical emergency that requires immediate medical intervention or it can lead to permanent impotence. The occurrence of priapism was first reported several months after the introduction of trazodone in the U.S. market. In the first year that trazodone was available in the United States, 11 cases of priapism were reported, 5 requiring surgical intervention. Now, with prolonged use of the drug, the rate of priapism appears to be approximately 1 in 10,000 patients treated. The exact cause of trazodone-induced priapism is not known, but it has been postulated to be due to peripheral a $_1$-adrenergic antagonist action within the circulatory system of the penis. Trazodone has minimal effects on CNS dopaminergic and cholinergic systems implicated in other drug-induced cases of priapism. In experimental animal systems, direct infusion of trazodone into the intracorporeal circulatory system causes erection.

Priapism has been reported to occur across the broad spectrum of the therapeutic dosage range of trazodone, from 50 to 400 mg. This effect generally occurs within the first 28 days of treatment but has been reported to occur after 18 months of therapy. There is no particular test or other method for predicting the occurrence of priapism.

Priapism needs to be treated on an emergent basis. Immediate discontinuation of the medication is the first intervention, and patients should be advised to discontinue taking trazodone and seek emergency treatment if they experience an erection that lasts for 1 hour or is excessively painful. Irrigation of the corpora cavernosa with an a-adrenergic receptor agonist agent such as metraminol (Aramine) or epinephrine is the first intervention. If the a-agonist agent is ineffective, then a vascular shunt must be placed, typically between the corpora cavernosa and the corpora spongiosum. If surgical shunting is required, there is a 40 to 50 percent incidence of permanent impotence. At least one case of clitoral priapism in a female patient treated with trazodone was reported in the literature. This patient required a-agonist treatment but did respond to this treatment. Trazodone has been reported to cause a number of other sexual adverse effects including increased or decreased libido in male and female patients, anorgasmia, impotence, and ejaculatory dysfunction.

Central Nervous System Trazodone's principal pharmacological actions are as a 5-HT $_{2A}$ receptor antagonist and as an inhibitor of serotonin transport. The most common neurological adverse effects associated with trazodone treatment are drowsiness, dizziness, lightheadedness, headache, and confusion or mental slowing. As many as 50 percent of patients treated with trazodone report significant drowsiness, which is the main side effect that causes patients to discontinue using this drug. On the other hand, patients with significant psychomotor agitation and insomnia do not find this particularly troublesome.

Trazodone is not known to lower the seizure threshold in patient populations or in experimental animals. There are isolated case reports of patients treated with trazodone having seizures, but this has not been observed in the large-scale clinical trials. Trazodone is not contraindicated in patients with known seizure disorders, though caution should be exercised in the dosing of these patients. Lower starting doses and slower upward titration of the dosage would be prudent when using trazodone in a patient with a known seizure disorder.

Trazodone has been reported to change sleep patterns and electroencephalogram (EEG) tracings during sleep. Generally, slow-wave sleep activity increases in patients taking trazodone. Whether these changes are meaningful is unclear.

Trazodone has been reported to precipitate mania and hypomania in vulnerable patients. The average time from initiating trazodone therapy to onset of manic symptoms is 16 days. The incidence of mania and hypomania in patients receiving trazodone appears to be lower than that reported for other antidepressant agents.

Overdose Trazodone has a favorable therapeutic index and is generally fairly benign in overdose. In the first 4 years that trazodone was available in the United States, 88 cases of overdose were reported to the manufacturer. In 73 of those patients recovery was reported as uneventful. Nine deaths were reported in this series, but all of them involved multiple pharmacological agents or alcohol. There are two case reports of patients ingesting 7.5 and 9 grams, respectively, of trazodone in overdose attempts; both recovered without incident. The symptoms of trazodone overdose are drowsiness, weakness, and prolonged unconsciousness, and general monitoring and supportive measures are the indicated treatment. Treatment with trazodone does not increase suicidal behavior or risk. There is no difference in the rate of suicidal ideation, attempted suicide, or aggressive behavior in patients receiving trazodone and those receiving imipramine, lofepramine, or mianserin.

THERAPEUTIC INDICATIONS

The FDA indication for the use of trazodone stated on the package insert is the treatment of major depressive disorder. Given its lack of anticholinergic and cardiac effects, trazodone seems particularly well suited for the treatment of depression in geriatric patients, although orthostatic hypotension is problematic in this age group. There does seem to be evidence to support the use of trazodone as a hypnotic agent both to treat the insomnia related to a depressive disorder and the insomnia associated with the use of other antidepressant agents. Evidence also supports the use of trazodone to treat generalized anxiety disorder but not panic disorder or obsessive-compulsive disorder. Evidence indicates trazodone may be an effective treatment of psychogenic impotence and organically caused erectile problems. Trazodone may have efficacy in the treatment of bulimia nervosa. Trazodone may also be useful in treating the adverse behavioral manifestations of autistic and other developmental disorders and cognitive disorders. It may be useful in the treatment of some chronic pain syndromes (pediatric migraine) but not others (neuropathic pain, low back pain). Trazodone may have an adjunctive role in the treatment of schizophrenia and is safe to use in schizophrenic patients because it does not worsen positive psychotic symptoms.

PRECAUTIONS AND ADVERSE REACTIONS

Patients should receive adequate information on the side effects and adverse reactions that can be caused by trazodone. Drowsiness is the most common side effect reported. Patients beginning trazodone therapy should be advised that the concomitant use of alcohol or other sedative agents may worsen its sedating effects. Patients, particularly the elderly and those receiving antihypertensive agents should be advised of the occurrence of orthostatic hypotension that can be severe enough to cause syncope. Ingesting trazodone with food reduces the peak serum concentration, thereby reducing the incidence of orthostatic hypotension.

The most significant side effect of trazodone in terms of excess morbidity is priapism. Patients should be well informed on the risks of developing this condition. They should be advised that an erection lasting longer than an hour or one that is excessively painful should be treated urgently and trazodone use immediately discontinued. Patients should be advised to seek immediate medical treatment if priapism develops.

The use of trazodone in patients with cardiac disease is not contraindicated. Patients with known arrhythmias should be monitored closely. Trazodone should not be a first-line agent in patients with known ventricular arrhythmias. The use of trazodone in patients with known seizure disorders is not contraindicated; trazodone dosage should be titrated slowly in patients with known seizure histories.

There is no evidence that trazodone is mutagenic in standard toxicological assays. Trazodone is teratogenic in experimental animals at high doses, 15 to 50 times the maximal human dose. There are no data on the teratogenic potential of trazodone in humans, nor are there any published data on patients exposed to trazodone in utero. Trazodone does enter breast milk. It has no affect on fertility in standard animal models, and there are no reports of alterations in human fertility in response to

trazodone treatment.

DRUG INTERACTIONS

There are case reports documenting a variety of drug-drug interactions in patients receiving trazodone. One patient who was treated with trazodone and fluoxetine had a 30 percent increase in circulating trazodone concentration. This patient experienced increased sedation and an unsteady gait, but the symptoms resolved with dosage adjustment. There are case reports documenting increased concentrations of phenytoin (Dilantin) and digoxin (Crystodigin) in patients receiving trazodone, but this has not been observed in large-scale studies.

There are no known interactions between trazodone and anesthetic agents, but given trazodone's sedating properties, discontinuation of trazodone prior to surgery would seem appropriate. While there is no evidence that trazodone affects seizure threshold, there are case reports claiming that patients on trazodone had prolonged seizures in response to electroconvulsive therapy (ECT). This has not been systematically studied. Trazodone has minimal impact on seizure threshold.

Trazodone can cause orthostatic hypotension, and this effect can be potentiated by the concomitant use of antihypertensive agents. Although this is not a true pharmacokinetic interaction, appropriate dosage adjustment of trazodone and the antihypertensive agent must be considered in patients receiving both agents.

There is evidence that trazodone is safe to use in conjunction with MAOIs. In fact there is evidence documenting the use of trazodone as a sedative hypnotic agent in patients receiving MAOI therapy. While isolated case reports of trazodone causing the serotonin syndrome exist, this reaction seems quite rare in patients being treated with trazodone in addition to other serotonergically active agents.

LABORATORY INTERFERENCES

Trazodone is not known to interfere with routine laboratory tests, including urinalysis, hematological analysis, and chemical analysis. Trazodone blood levels need not be monitored in the routine treatment of depression. There are no reliable data relating concentration in blood to therapeutic response, nor are there any data on an optimal blood concentration for therapeutic response. Monitoring blood concentrations may be a useful method of assessing patient compliance.

DOSAGE AND ADMINISTRATION

Optimal use of antidepressants necessitates matching the side-effect profile of the particular agent to the patient's symptoms. This is known to increase patient compliance, which in turn improves therapeutic response. In the case of trazodone, which is sedating and anxiolytic, patients who have significant psychomotor agitation, anxiety, and insomnia are particularly well suited for treatment with this agent. Patients who have significant neurovegetative symptomatology, anergia, and hypersomnolence would not be expected to be as responsive to trazodone, and other, more-stimulating agents should be considered.

Four controlled studies have examined the optimal dosing regimen for trazodone. They found that once-daily dosing (usually at night) is as effective as multiple daily dosing. This is somewhat unexpected given the agent's short half-life.

Trazodone therapy should be initiated at a dosage of 50 to 100 mg a day. The dosage should be increased at a minimum of 3-day intervals. Rapid dose escalation generally results in significant sedation and discontinuation of treatment. A dosage of 300 mg a day should be achieved within 2 weeks of beginning therapy. There is some controversy about the optimal dosage, with some practitioners advocating a maximum daily dosage of 300 mg, while others recommend 400 to 600 mg a day. Some have postulated a therapeutic window for this agent, but none has been definitively documented. Certainly a lack of therapeutic response at doses below 300 mg a day cannot be judged a treatment failure. Escalation of the daily dosage above 300 mg should be considered in a patient who is responding partially to the 300 mg daily dosage, but escalation of the dosage above 600 mg a day would seem unnecessary. If the patient has experienced no appreciable response at a dosage of 300 mg a day after 4 weeks, then reconsideration of the treatment regimen is in order. Most patients display a positive therapeutic response by 4 weeks of treatment, and some are appreciably improved after 2 weeks. Once-a-day dosing, in the evening, is the preferred regimen because it causes the fewest side effects and increases overall treatment compliance. Trazodone is safe and continues to be effective in the treatment of depression in long-term administration. Generally, 50 to 100 mg at night is a sufficient dose to treat insomnia.

Trazodone is available in scored 50-mg and 100-mg tablets and in 150-mg and 300-mg tablets. The larger-dose tablets can be broken into thirds or halves to facilitate dosage adjustment. There is no difference in treatment response between the regular form of trazodone and a controlled-release form available in Europe.

SUGGESTED CROSS-REFERENCES

Section 1.4 discusses monamine neurotransmitters; Section 12.8 discusses somatic treatments of schizophrenia; Chapter 14 discusses mood disorders. Chapter 15 discusses anxiety disorder. Section 19.1a discusses sexual dysfunctions. Chapter 20 discusses eating disorders. Chapter 21 discusses sleep disorders. Section 31.11 discusses bupropion; Section 31.19 discusses mirtazapine; Section 31.20 discusses MAOIs; Section 31.25, SSRIs; and Section 31.22 discusses nefazodone.

SECTION REFERENCES

Altamura AC, Mauri MC, Colacurcio F, Scapicchio PL, Hadjchristos C, Carucci M, Minervi M, Montanini R, Perini M, Rudas N: Trazodone in late life depressive states: A double-blind multicenter study versus amitriptyline and mianserin. Psychopharmacology *95*(Suppl):34, 1988.

Battistella PA, Ruffilli R, Cernetti R, Pettenazzo A, Baldin L, Zacchello F: A placebo-controlled crossover trial using trazodone in pediatric migraine. Headache *33:*36, 1993.

Beasley CM, Dornseif BE, Pultz JA, Bosomworth JC, Sayler ME: Fluoxetine versus trazodone: Efficacy and activating-sedating effects. J Clin Psychiatry *52:*294, 1991.

*Brogden RN, Heel RC, Speight TM, Avery GS: Trazodone: A review of its pharmacological properties and therapeutic use in depression and anxiety. Drugs *21:*401, 1981.

Brooks D, Prothero W, Bouras N, Bridges PK, Jarman CM, Ankier SI: Trazodone—a comparison of single night-time and divided daily dosage regimens. Psychopharmacology *84:*1, 1984.

Bryant SG, Ereshefsky L: Antidepressant profile of trazodone. Clin Pharmracol *1:*406, 1982.

*Bryant SG, Hokanson JA, Brown CS: A drug utilization review of prescribing patterns for trazodone versus amitriptyline. J Clin Psychiatry *51*(Suppl):27, 1990.

Charney DS, Woods SW, Goodman WK, Rifkin B, Kinch M, Aiken B, Quadrino LM, Heninger GR: Drug treatment of panic disorder: The comparative efficacy of imipramine, alprazolam, and trazodone. J Clin Psychiatry *47:*580, 1986.

Decina P, Mukherjee S, Bocola V, Saraceni F, Hadjichristos C, Scapicchio P: Adjunctive trazodone in the treatment of negative symptoms of schizophrenia. Hosp Community Psychiatry *45:*1220, 1994.

*Feighner JP: Mechanism of action of antidepressant medications. J Clin Psychiatry *60:*4, 1999.

*Feighner JP, Boyer WF: Overview of U.S.A. controlled trials of trazodone in clinical depression. Psychopharmacology *95*(Suppl):50, 1988.

*Gamble DE, Peterson LG: Trazodone overdose: Four years of experience from voluntary reports. J Clin Psychiatry *47:*544, 1986.

Goodkin K, Gullion CM, Agras WS: A randomized, double-blind, placebo-controlled trial of trazodone hydrochloride in chronic low-back-pain syndrome. J Clin Psychopharmacol *10:*269, 1990.

Jacobsen FM: Low-dose trazodone as a hypnotic in patients treated with MAOIs and other psychotropics: A pilot study. J Clin Psychiatry *51:*298, 1990.

Kurt U, Ozkardes H, Altug U, Germiyanoglu C, Gurdal M, Erol D: The efficacy of anti-serotonergic agents in the treatment of erectile dysfunction. J Urol *152:*407, 1994.

Lance R, Albo M, Costabile RA, Steers WD: Oral trazodone as empirical therapy for erectile dysfunction: A retrospective review. Urology *46:*117, 1995.

Lebert F, Pasquier F, Petit H: Behavioral effects of trazodone in Alzheimer's disease. J Clin Psychiatry *55:*536, 1994.

Mann JJ, Georgotas A, Newton R, Gershon S: A controlled study of trazodone, imipramine, and placebo in outpatients with endogenous depression. J Clin Psychopharmacol *1:*75, 1981.

Moises HW, Kasper S, Beckman H: Trazodone and amitriptyline in treatment of depressed inpatients. A double-blind study. Pharmacopsychiatria *14:*167, 1981.

Nierenberg AA, Adler LA, Peselow E, Zornberg G, Rosenthal M: Trazodone for antidepressant-associated insomnia. Am J Psychiatry *151:*1069, 1994.

*Patten SB: The comparative efficacy of trazodone and imipramine in the treatment of depression. Can Med Assoc J *146:*1177, 1992.

Pigott TA, L'Heureux F, Rubenstein CS, Bernstein SE, Hill JL, Murphy DL: A double-blind, placebo controlled study of trazodone in patients with obsessive-compulsive disorder. J Clin Psychopharmacol *12:*156, 1992.

Pope HG, Keck PE, McElroy SL, Hudson JI: A placebo-controlled study of trazodone in bulimia nervosa. J Clin Psychopharmacol *9:*254, 1989.

Rickels K, Downing R, Schweizer E, Hassman H: Antidepressants for the treatment of generalized anxiety disorder. A placebo-controlled comparison of imipramine, trazodone, and diazepam. Arch Gen Psychiatry *50:*884, 1993.

Silvestrini B: Trazodone and the mental pain hypothesis of depression. Neuropsychobiology *15*(Suppl):2, 1986.

Silvestrini B: The paradoxical stress response: A possible common basis for depression and other conditions. J Clin Psychiatry *51*(Suppl):6, 1990.

Weisler RH, Johnston JA, Lineberry CG, Samara B, Branconnier RJ, Billow AA: Comparison of bupropion and trazodone for the treatment of major depression. J Clin Psychopharmacol *14:*170, 1994.

Wheatley D: Trazodone: Alternative dose regimens and sleep. Pharmatherapeutica *3:*607, 1984.

Workman EA, Short DD: Atypical antidepressants versus imipramine in the treatment of major depression: A meta-analysis. J Clin Psychiatry *54:*5, 1993.

Zubieta JK, Alessi NE: Acute and chronic administration of trazodone in the treatment of disruptive behavior disorders in children. J Clin Psychopharmacol *12:*346, 1992.

31.30 TRICYCLICS AND TETRACYCLICS

J. CRAIG NELSON, M.D.

HISTORY

In 1957, Roland Kuhn, a Swiss psychiatrist, reported that imipramine (Tofranil) appeared to have antidepressant effects in 40 patients with endogenous depression. This serendipitous observation was made while testing this chlorpromazine-like antihistamine for its psychoactive properties. This discovery, like those for lithium and chlorpromazine, (Thorazine) was an important landmark in modern psychopharmacology. The knowledge that this compound appeared to alleviate depression, coupled with Julius Axelrod's report of its role in blocking reuptake of norepinephrine, laid the foundation for the norepinephrine hypothesis of depression.

It is with the tricyclic compounds that the field of psychiatry explored a variety of issues involved in the drug treatment of depression. Rating scales for depression were developed. Diagnostic criteria emerged. A better understanding of drug-placebo differences evolved. The necessity for administering an adequate dose for an adequate duration was described. The physiological basis of the secondary effects of tricyclic drugs became better understood. The relationship of plasma concentrations and clinical phenomenon was explored, and the pharmacology of these compounds was investigated as it provided a “psychopharmacologic bridge” to an understanding of depression.

During the next two decades, several other compounds were developed that also had antidepressant properties. These compounds were similar in sharing a basic tricyclic structure and in terms of many of the secondary effects for which the tricyclic drugs came to be known. Later, tetracyclic compounds were also marketed which were somewhat similar in structure and had relatively comparable secondary properties.

CHEMISTRY

Tricyclic and tetracyclic compounds are categorized primarily on the basis of their chemical structure (Fig. 31.30-1). The tricyclic compounds have a three-ring structure, hence the name. The tertiary amine tricyclic compounds, such as amitriptyline (Elavil) and imipramine, have two methyl groups at the end of the side chain. These compounds can be demethylated to secondary amines, such as desipramine (Norpramine, Pertofrane) and nortriptyline (Aventyl, Pamelor). The tetracyclic compounds, maprotiline (Ludiomil) and amoxapine (Asendin) have a four-ring central structure. Five tertiary amines have been marketed in the United States–amitriptyline, clomipramine (Anafranil), doxepin (Sinequan), imipramine, and trimipramine (Surmontil). The three secondary amine compounds are desipramine, nortriptyline, and protriptyline (Vivactil).

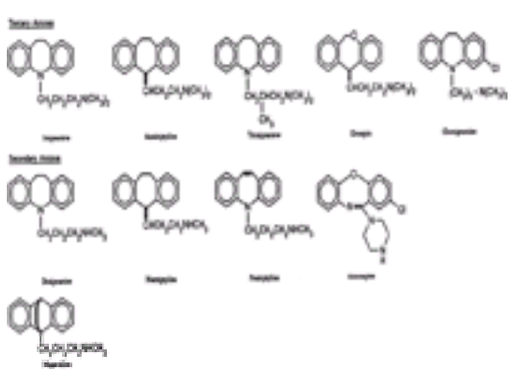

FIGURE 31.30-1 Molecular structures of tricyclics and tetracyclics.

PHARMACOLOGICAL ACTIONS

Pharmacodynamics

Reuptake Blockade The tricyclic and tetracyclic antidepressants block the transporter sites for serotonin and norepinephrine and thus reduce uptake of these amines into the presynaptic neuron (Table 31.30-1). The tertiary amines have greater affinity for the serotonin transporter. The secondary amines primarily block the uptake of norepinephrine. Because the tertiary amines are demethylated to secondary amines, during the administration of amitriptyline, imipramine, or clomipramine, secondary amines are also present, thus both serotonergic and noradrenergic effects occur.

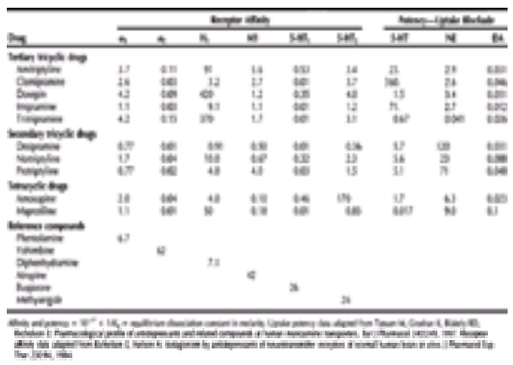

Table 31.30-1 Receptor Affinity or Potency of Cyclic Antidepressants in Human Brain

Receptor Sensitivity Changes Following the initial reuptake blockade, a sequence of events occurs. The tertiary tricyclic compounds, particularly clomipramine, inhibit the uptake of serotonin. Following reuptake blockade, the presynaptic serotonin (5-hydroxytryptamine [5-HT]) type 1A (5-HT $_{1A}$) autoreceptor reduces the firing rate of the presynaptic serotonin neuron, and concentrations of 5-hydroxyindoleacetic acid (5-HIAA), the major metabolite of serotonin, fall rapidly. Over a 2-week

period, the presynaptic autoreceptor is desensitized, which normalizes the firing rate. At this point serotonin transmission is enhanced. With chronic treatment, the tricyclic agents as a group also sensitize, or upregulate, postsynaptic 5-HT $_{1A}$ receptors. The changes in sensitivity of these presynaptic and postsynaptic receptors occur over a 2-week period and appears to be more consistent with the timing of antidepressant response than with the initial uptake blockade.

Depression is also associated with an increase in postsynaptic 5-HT $_{2}$ receptor density. A variety of antidepressants, including the tricyclics, have been shown to downregulate the 5-HT$_2$ receptors. The 5-HT$_2$ receptors principally mediate excitatory effects, while the 5HT $_{1A}$ receptors have an inhibitory effect in the brain; thus these two systems act in opposition. In preclinical experiments, administration of a 5-HT $_{2}$ antagonist enhances the effects of serotonin. Thus, the downregulation of 5-HT$_2$ receptors with tricyclic treatment appears to facilitate the effects of serotonin at postsynaptic 5-HT $_{1A}$ receptors.

In the noradrenergic system, as a consequence of blockade of the norepinephrine transporter, turnover of norepinephrine falls and the concentration of 3-methoxy-4-hydroxyphenylglycol (MHPG), a metabolite of norepinephrine, drops rapidly, within 24 hours. Presumably, this effect results from the presynaptic a$_2$-adrenergic autoreceptor; however, the role of the presynaptic autoreceptor and its possible subsequent desensitization is not as well described for the noradrenergic system as it is for the serotonin system. During this period, the postsynaptic, b-adrenergic receptor, is downregulated, or decreased in density. This downregulation is associated with a corresponding reduction in activity of cyclic adenosine monophosphate (cAMP), the second messenger for the b-receptor.

Early theories of depression posited depletion of serotonin, norepinephrine, or both of these neurotransmitters. Administration of the tricyclic agents was thought to enhance the availability of both neurotransmitters. Subsequent theories suggested that the changes in pre- and postsynaptic receptors were a likely explanation for the delayed effects of these drugs in depression. In the serotonin system, desensitization of presynaptic 5-HT $_{1A}$ autoreceptors and sensitization of postsynaptic serotonin receptors are both changes consistent with enhanced serotonin transmission. Theories of norepinephrine action, however, have become more complicated with the knowledge that postsynaptic b-adrenergic receptors are, in fact, downregulated. It has been suggested that even with downregulation of postsynaptic receptors, the net effect of reuptake blockade is still enhanced formation of the second messenger cAMP. These changes are part of the postreceptor signal transduction cascade, which is a current focus of research on the mechanism of antidepressant action.

An understanding of how the antidepressant drugs act does not necessarily implicate a particular neurotransmitter in the pathophysiology of depression. Recent studies with tryptophan depletion illustrate this point. Administration of a tryptophan-free diet rapidly depletes serotonin. Depressed patients who have been successfully treated, relapse following tryptophan depletion. Further, the subjects who relapse are those treated with serotonergic agents; those treated with norepinephrine agents are relatively unaffected. Alternatively, administration of a-methyl-para-tyrosine, which interrupts the synthesis of catecholamines, causes relapse in patients successfully treated with noradrenergic agents, while those receiving serotonergic drugs are relatively unaffected. Tryptophan depletion in untreated depressed patients, however, has no effect on the patient's depression. These studies support the view that the action of these two types of drugs is, in fact, mediated by serotonin and norepinephrine, but they do not necessarily imply that alterations in these neurotransmitter systems are central to the pathophysiology of depression.

Early theories of antidepressant action suggested that it might be possible to identify serotonergic and noradrenergic depressions on the basis of the relative deficiencies of metabolites associated with each of these neurotransmitters and thus predict which type of antidepressant would be most effective. A number of studies investigated the predictive value of MHPG pretreatment, but their findings were relatively disappointing. These studies were, in part, hampered by the use of agents such as amitriptyline and imipramine, which are not selective.

Secondary Effects The tricyclic and tetracyclic compounds have effects on a variety of other receptors (Table 31.30-1). In particular, these compounds block muscarinic receptors, producing anticholinergic effects. They block histamine type 1 (H $_{1}$) receptors and they block a$_1$- and a$_2$-adrenergic receptors. They affect fast sodium channels, which explains their adverse cardiac effects. These compounds vary considerably in the potency of these effects. Amitriptyline is the most anticholinergic. Doxepin is one of the most potent H $_{1}$ antagonists marketed.

Pharmacokinetics

Absorption The tricyclic and tetracyclic drugs are absorbed in the small intestine, and absorption is reasonably complete and rapid. Peak concentrations following ingestion occur within 2 to 8 hours. One exception is protriptyline, which is absorbed more slowly, with peak concentrations obtained between 6 and 12 hours. Maprotiline also has more-delayed absorption, with peak concentrations occurring in 8 hours or more. Because the primary action of these drugs occurs over several weeks, the timing of peak concentrations is relatively unimportant for treatment of depression. It may have implications for side effects, especially during initial dosing.

Volume of Distribution The tricyclic and tetracyclic compounds are basic lipophilic amines. As a result, they are concentrated in a variety of tissues throughout the body and have a high volume of distribution. In cardiac tissue, for example, drug concentrations exceed those in plasma.

Plasma-Protein Binding Because of their lipid solubility, these compounds are extensively bound to plasma proteins. Binding of the tricyclic and tetracyclic drugs is generally 90 percent or greater. One exception are the hydroxy metabolites of the tricyclics which have somewhat lower plasma protein binding.

First-Pass Metabolism Following absorption, the drug is taken up in the circulation and passes first through the liver, where metabolism of the drug begins (the first-pass effect). As a result, a reduced amount of the compound actually enters the systemic circulation.

Hepatic Metabolism Clearance of the tricyclic and tetracyclic compounds is principally the result of hepatic metabolism. Renal clearance accounts for only a small portion of drug elimination. Rates of hepatic metabolism vary widely from person to person. Elimination half-lives for most of the tricyclic and tetracyclic compounds are about 24 hours or longer, thus the drugs can be given once a day (Table 31.30-2). Amoxapine, which has a shorter half-life, is an exception.

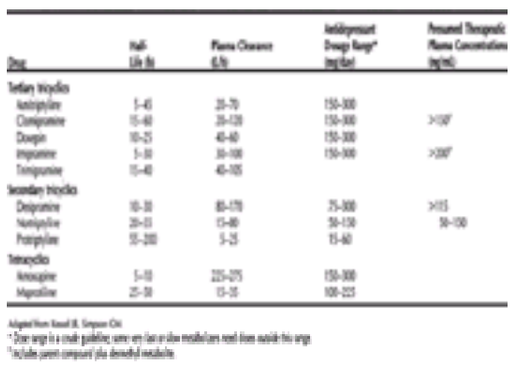

Table 31.30-2 Pharmacokinetic Parameters and Clinical Information for Tricyclic and Tetracyclic Drugs

There are two principal metabolic pathways: demethylation of the side chain and hydroxylation of the ring structure. Demethylation converts the tertiary amine to a secondary amine; examples are the conversion of amitriptyline to nortriptyline and imipramine to desipramine. This process alters the activity of the compound. The tertiary amines are relatively more serotonergic; the demethylated amines are relatively more noradrenergic.

Hydroxylation of the compounds results in hydroxy metabolites. The concentration of the metabolite of nortriptyline, 10-hydroxynortriptyline usually exceeds that of the parent. For desipramine, 2-hydroxydesipramine is present at concentrations approximately 45 percent of those of the parent, but these ratios are quite variable. The ratio of hydroxy metabolite to parent increases in rapid metabolizers. Hydroxyimipramine and hydroxyamitriptyline are present at quite low levels that appear to be clinically unimportant. The hydroxy metabolites are then conjugated and excreted. The conjugated metabolites are not active.

Hydroxynortriptyline and hydroxydesipramine are active compounds. Both have antidepressant activity. Hydroxydesipramine is comparable to the parent in terms of

norepinephrine reuptake blockade. There are two isomers of hydroxynortriptyline, E- and Z-10-hydroxynortriptyline. E-10-hydroxynortriptyline concentrations are four times those of the Z isomer. E-10-hydroxynortriptyline also blocks uptake of norepinephrine but is about 50 percent as potent as nortriptyline. The clinical significance of high concentrations of less potent hydroxynortriptyline is not entirely clear, but if high concentrations of hydroxynortriptyline interfere with the action of nortriptyline, this might explain the therapeutic window described for this drug. In both cases, the hydroxy compounds are less anticholinergic than their parent compounds. The hydroxy metabolites have other effects as well. They appear to contribute disproportionately to the cardiac conduction abnormalities associated with the tricyclic drugs.

The principle metabolic pathway for amoxapine is hydroxylation. 7-hydroxyamoxapine and 8-hydroxyamoxapine are produced, which differ in their clinical activity. 7-hydroxyamoxapine has high-potency neuroleptic properties but a short half-life. 8-hydroxyamoxapine is metabolized more slowly and appears to contribute to the drug's antidepressant action.

An area of great interest in recent years has been the identification of the specific isoenzyme pathways involved in the metabolism of a variety of drugs, including the antidepressants. The cytochrome P450 (CYP) isozyme 2D6 (CYP 2D6) pathway appears to be responsible for hydroxylation of desipramine and nortriptyline. Demethylation of the tertiary amine compounds appears to involve a number of cytochrome P450 isoenzymes, including CYP 1A2, 3A4, and 2C19. These hepatic isoenzymes are under the control of specific genes, and the gene loci have been identified for several of these isoenzymes, including CYP 2D6. Approximately 5 to 10 percent of whites are homozygous for the recessive autosomal CYP 2D6 trait, resulting in deficient hydroxylation of desipramine and nortriptyline. These individuals are known as poor metabolizers, while those with adequate 2D6 isoenzyme are extensive metabolizers. The resulting variability in plasma concentrations is substantial, often described as a 40-fold variation among individuals (Fig. 31.30-2). Even among the extensive metabolizers, there can be moderate variability from individual to individual in the rates of metabolism. Various methods have been used to phenotype the individuals who are slow or fast metabolizers. One technique is debrisoquine phenotyping. Rates of formation of the debrisoquine metabolite in the urine are used to characterize the metabolic rate of CYP 2D6. In clinical practice, blood monitoring for desipramine and nortriptyline concentrations is more readily available than debrisoquine assays, and the concentrations in blood of the compounds themselves are more often used as an index of the rate of metabolism.

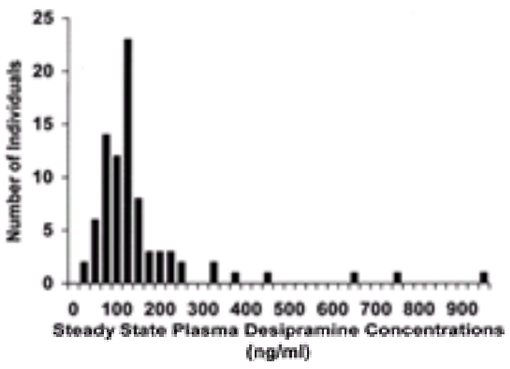

FIGURE 31.30-2 Distribution of steady-state desipramine concentrations in 83 inpatients receiving desipramine (2.5 mg/kg a day). (Adapted from Nelson JC: J Clin Psychiatry *45:*10, 1984.)

Steady-State Concentrations Clinicians must understand the concept of steady state if plasma concentration monitoring is used. Steady state is the point at which plasma concentrations plateau for a given fixed dose. The usual convention in pharmacology is that steady state is achieved after five times the half-life. After five half-lives, the concentration of the drug should be 97 percent of the maximal concentration achieved for that dose. In fact, for the clinician, it is also worth remembering that after three half-lives, the drug will have achieved about 87 percent of the steady-state concentration. The latter is important when dose adjustments are urgently required. If blood concentration monitoring is used, a trough sample (before the next dose) is drawn when the patient has reached steady state. At steady state, drug concentrations are relatively stable as long as the dose is constant, the patient is compliant, and no drugs with interactive effects are added.

When the drug concentration is measured, the total of both the free and bound tricyclic fractions are reported. Few laboratories are prepared to measure free levels, yet free levels are directly related to the concentration of drug in the cerebrospinal fluid. The free concentration depends on dose and hepatic clearance and is not affected by absorption, the volume of distribution, or plasma-protein binding. The latter is one of the most misunderstood aspects of pharmacokinetics. If the concentration of binding protein is low, the total (free and bound) concentration may be low, but the absolute free concentration is unaffected. If another drug affects binding, the absolute free concentration remains unaffected. In these instances, the free fraction may increase because the bound portion decreases, but the absolute free concentration is unchanged.

The biological variability of drug concentrations is not frequently described. In outpatients, the day-to-day variability of nortriptyline concentrations (the coefficient of variation equals the standard deviation divided by the mean plasma concentration) has been reported to be 25 percent. In inpatients at steady state, day-to-day variability of desipramine is approximately 10 to 15 percent. This means that if the average plasma concentration is 150 ng/mL, two thirds of samples obtained will be ±10 percent or within the range of 135 and 165 ng/mL. As a result of this variability, research studies of drug concentrations frequently use the average of two or three plasma samples drawn at steady state to estimate the concentration. For clinical purposes, usually only one sample is drawn, and the clinician needs to remember that even if the laboratory error is low (less than 5 percent), the extent of biological variability means that blood concentrations are better viewed as estimates rather than as precise measures.

Linear Kinetics Drugs for which concentration increases in proportion to dosage within the therapeutic range are said to have linear pharmacokinetics. For most of the tricyclic drugs, kinetics appear to be relatively linear, although there are some exceptions. Desipramine, for example, appears to have nonlinear kinetics in the usual dosage range. The patients most vulnerable to nonlinear changes during increasing dosages are the rapid metabolizers—in other words, patients who at usual doses have very low blood concentrations. Increasing dosages in these patients result in a disproportionate rise in the drug concentration. Nortriptyline appears to have relatively linear kinetics within the usual dosage range. In overdose, nonlinear changes are more likely to occur, and the clinician cannot assume that usual rates of drug elimination will be maintained.

Effects of Aging The pharmacodynamics and pharmacokinetics of drug treatment in the elderly are complicated. A number of changes occur. For example, the ratio of fat to lean body mass increases and cardiac output and hepatic blood flow decrease. There may be further changes associated with medical illness. Yet, the magnitude of these changes are relatively minor in relation to the dramatic variability of hepatic metabolism. Metabolic changes with age appear to vary with the isoenzymes involved. The CYP 3A4 pathway, for example, does appear to decrease or slow with age. Most studies of the tertiary amines (e.g., imipramine) suggest that concentrations of these drugs are increased somewhat in the elderly. The CYP 2D6 isoenzyme pathway, however, is not similarly affected by aging. Concentrations of nortriptyline and desipramine appear to be relatively unaffected by aging. Similarly, the relation of drug concentrations to therapeutic effects appears to be relatively similar in younger and older adults for these agents. Renal clearance of the hydroxy metabolites does decrease with age. Concentrations of hydroxynortriptyline may be substantially elevated in older patients.

Alternatively, the clearance of tricyclic compounds in children is increased. Half-lives of imipramine are shorter and ratios of desmethylimipramine to imipramine are higher, consistent with more-rapid metabolism.

DESIGN AND INTERPRETATION OF CLINICAL DRUG STUDIES

Efficacy in Depressive Disorders The clinical efficacy of the tricyclic and tetracyclic compounds in major depression is well established and was extensively reviewed by John Davis and Alexander Glassman in the fifth edition of this text. Imipramine is the best-studied compound. In 44 placebo-controlled studies, imipramine was more effective than placebo in 30. Overall, in studies reporting response rates (*N* = 1334), 65 percent of the patients completing treatment with imipramine were substantially improved, while 30 percent of those on placebo were similarly improved. In a meta-analysis of intent-to-treat response rates for placebo-controlled studies of tricyclic drugs in outpatients, reported in the Agency for Health Care Policy and Research (AHCPR) guidelines, response rates were 51 percent for imipramine and 30 percent for placebo. In most comparison studies, the other tricyclic and tetracyclic antidepressants have been found to be comparable

to imipramine in efficacy.

In addition to efficacy in acute depression, the tricyclic compounds are also effective for maintenance treatment. Several early studies demonstrated that maintenance treatment with a tricyclic reduced the relapse rate associated with placebo by about 50 percent. These studies, however, often used low doses. Subsequently, an important study by the Pittsburgh group found that if the full dose of imipramine used for acute treatment was continued, nearly 80 percent of the depressed patients maintained their improvement, while only 10 percent of those on placebo remained well at the end of a 3-year period. In this study, maintenance psychotherapy had an intermediate effect, with about 30 percent of the patients remaining well. This study again confirmed the value of maintenance treatment and underscored the importance of continuing the full dose of imipramine. Yet, the magnitude of the findings may in part relate to the sample selected, namely, patients with recurrent depression who had been symptom free between prior episodes. In clinical practice, if patients with chronic depression, prior residual symptoms, or comorbid psychiatric disorders are included, the number of patients who remain in remission may be lower.

Subtypes of Depression The efficacy of the tricyclic compounds varies in different subtypes of depression. Kuhn thought it was essential to establish the efficacy of imipramine in endogenous inpatients. The tricyclic compounds have been frequently studied in severely depressed inpatients and found to be effective. It has been suggested that the tricyclics may be more effective than the selective serotonin-reuptake inhibitors (SSRIs) in severe major depressive disorders with melancholic features. The issue remains unsettled, in part, because of the paucity of studies of the SSRIs in severe melancholic depression.

The tricyclic and tetracyclic compounds have often been studied in anxious depression. Three of the compounds—doxepin, amoxapine, and maprotiline—have Food and Drug Administration (FDA) approval for use in patients with depression and symptoms of anxiety. For many years, clinical lore suggested that amitriptyline was most effective for anxious depression; however, there is little evidence from direct comparison studies that these compounds differ in their efficacy for treatment of anxiety associated with depression. Three studies found that among depressed patients, those with high levels of anxiety respond less well to amitriptyline, imipramine, or desipramine than those with lower levels of anxiety. Yet, these drugs are still more effective than placebo in anxious depressed patients, and it is not established that other classes of antidepressants are more effective in these patients.

In patients with atypical features (defined by the Columbia University or the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV] criteria), imipramine is more effective than placebo but significantly less effective than the monoamine oxidase inhibitors (MAOI). Little information is available regarding comparisons of tricyclic drugs and SSRIs in atypical depression.

Major depressive disorder with psychotic features is less responsive than is nonpsychotic depression to tricyclics used alone. However, several open trials and one prospective study have found that tricyclic drugs combined with an antipsychotic agent are effective in psychotic depression. One double-blind study compared the efficacy of amoxapine with that of the combination of perphenazine (Trilafon) and amitriptyline and found amoxapine to have comparable efficacy in psychotic depression.

In controlled studies of patients with bipolar disorder depression, imipramine was found less effective than tranylcypromine (Parnate) or fluoxetine (Prozac). These studies and others suggest that tricyclic drugs are more likely to induce mania than the newer serotonergic agents or bupropion (Wellbutrin).

Tricyclic antidepressants appear to be effective in elderly patients, but overall drug and placebo response rates appear to be lower than they are in nonelderly patients, although the same appears to be true for other drug classes such as the SSRIs. In children and adolescents, the tricyclics have not demonstrated superiority over placebo.

The tricyclic drugs are effective in dysthymia. Imipramine and desipramine have both been studied in controlled trials and are more effective than placebo for acute treatment of dysthymia and for maintenance treatment.

Relations of Dosage, Plasma Concentrations, and Response Early studies of the tricyclic antidepressants (particularly imipramine) demonstrated that in patients with endogenous major depressive disorder, aggressive dosing, in the 200 to 300 mg a day range, was more effective than lower doses. Because of the adverse effects associated with tricyclic drugs, therapy was usually begun at lower dosages and then titrated upward. The rationale was that patients would accomodate to these side effects and, over time, be able to tolerate higher doses. The need for dose titration and the time required appeared to be obstacles to the use of these drugs by nonpsychiatrists. Studies of these agents in nonpsychiatric settings consistently showed evidence of underdosing.

Knowledge of the wide interindividual variability of tricyclic drug plasma concentrations led to the hope that drug concentration monitoring might ensure that therapeutic concentrations were achieved, and might help to avoid toxic levels. In carefully selected endogenous or melancholic inpatients with major depression, treatment with adequate concentrations of imipramine or desipramine resulted in response rates of about 85 percent. These studies suggested that adequate concentrations in blood would help to maximize response. Alternatively, the careful selection of patients in these studies may have resulted in samples that are not representative of typical clinical practice.

An American Psychiatric Association (APA) task force reviewing the information on blood concentrations and response concluded that a relation had been demonstrated for nortriptyline, for which plasma concentrations between 50 and 150 ng/mL were more effective than lower or higher levels (Table 31.30-2). For imipramine, concentrations above 200 ng/mL were more effective than lower levels. For desipramine, the findings of several individual studies have been variable; however, there was general agreement that concentrations below 110 to 115 ng/mL were not effective. Most studies with an appreciable number of patients at higher plasma concentrations found that response did not decline with higher concentrations, although in one study, patients with concentrations above 155 ng/mL did less well. For both desipramine and imipramine, it appeared that blood concentrations above 300 ng/mL were more likely to be associated with serious adverse effects.

For amitriptyline, it has been more difficult to establish a therapeutic relation of concentration and response. In part, this may be related to the fact that during amitriptyline administration, three active compounds are present (amitriptyline, nortriptyline, and hydroxynortriptyline), and it is unclear whether concentrations of the three compounds should be simply summed or whether a more complicated relationship exists, given that nortriptyline and hydroxynortriptyline have a therapeutic window. During amitriptyline administration, responders usually have total amitriptyline and nortriptyline concentrations in the neighborhood of 150 to 250 ng/mL, but as stated, there is not good agreement between studies. For clomipramine, blood concentrations of 150 to 300 ng/mL (total of clomipramine and desmethylclomipramine) have been suggested for antidepressant effectiveness. Higher doses are usually employed in the treatment of obsessive-compulsive disorder. The data relating blood levels and response are limited for the other tricyclic and tetracyclic compounds.

The clinical value of blood concentration monitoring has been the subject of considerable controversy. It has been suggested that blood concentration is not related to response in depressed outpatients. Indeed, the relation of blood concentration and response has been best demonstrated in inpatients, probably because placebo response rates are low in melancholic inpatients; and drug-placebo differences are reasonably substantial, so the drug effect is relatively easy to detect. In outpatients, placebo rates are higher, the difference between drug and placebo is substantially reduced, and the effect of drug treatment is harder to detect. Moderately depressed outpatients may be a heterogeneous group and include patients who are less responsive to drug treatment. Finally, many studies have not been designed to investigate blood concentration relationships. Fixed dosing is required, and a significant portion of patients in the sample must achieve plasma concentrations below the suspected threshold. If all patients achieve plasma drug concentrations above a lower threshold, no relation with response will be found. It is logical to conclude that blood concentration relationships determined in severely depressed inpatients might be used to guide treatment of outpatients, particularly treatment-resistant patients for whom the clinician wants to ensure that adequate blood concentrations were obtained.

Plasma Concentrations and Toxicity The alternative question is whether blood concentration monitoring might help to avoid toxicity. A variety of data supports this view. The risk of delirium is substantially increased at tricyclic plasma concentrations above 450 ng/mL and moderately increased at levels above 300 ng/mL. However, most of the patients included in these studies were receiving amitriptyline, the tricyclic drug most likely to be associated with delirium. The risk of cardiac arrhythmia is also substantially increased at higher blood concentrations. First-degree atrioventricular block increases with plasma concentrations of imipramine above 350 ng/mL and increases substantially when concentrations exceed 450 ng/mL. The risk of seizures also increases in patients at higher doses and (presumably), higher blood concentrations, although a plasma concentration threshold has not been demonstrated for seizures. The risk of seizures for clomipramine is 1.66 percent with doses above 250 mg a day but is about 0.50 percent for doses up to 250 mg a day. On average, a dosage of 250 mg a day corresponds to a plasma concentration of about 500 ng/mL. In one large sample of patients with obsessive-compulsive disorder at steady state on clomipramine 250 mg a day, the average blood concentration was 155 ng/mL for clomipramine and 335 ng/mL for desmethylclomipramine, or a total of 490 ng/mL. For the older drugs amitriptyline and imipramine, the risk of seizures also appears to be increased at dosages exceeding 250 mg a day. Following overdose, tricyclic drug concentrations in blood can exceed 1000 ng/mL, and the risk of delirium, stupor, cardiac abnormalities, and seizures increase substantially.

One problem in appreciating the relation of toxicity to increased blood concentrations is that although the risk may increase, actual rates are still low, and large samples are required to demonstrate the difference in risk. Certain types of toxicity (e.g. delirium) may have early warning signs, and dosage can be reduced unless the symptoms of delirium are mistaken for symptoms of the underlying psychiatric disorder. Alternatively, there may be no early signs of seizures and cardiac arrhythmia to alert the clinician to reduce dosage. Blood concentration monitoring might be most useful to reduce the risk of these particular adverse effects.

EFFECTS ON SPECIFIC ORGANS AND SYSTEMS

The clinical manifestation of so-called side effects results from the interaction of direct effects of the medication on specific organs and indirect effects of the medication on depression and its associated somatic symptoms. Depression is accompanied by a variety of somatic symptoms; for example, headache, constipation, poor memory, drowsiness, and difficulty sitting still are observed in more than 50 percent of untreated inpatients with major depression. Other symptoms, such as palpitations, dry mouth, tremors, lightheadedness, and poor coordination are reported by more than 40 percent of depressed patients prior to treatment. The determination of whether a physical symptom is an adverse effect of a drug or a symptom of depression involves a judgment about whether the symptom is new or has worsened during drug treatment. Unfortunately, severely depressed patients may not remember accurately that the somatic symptom of concern was more severe prior to the introduction of medication. During treatment, one of the strongest predictors of overall adverse effects reported is the level of depression at the time of assessment.

Antidepressant drugs do, of course, produce new somatic symptoms. The in vitro potency or affinity of antidepressant compounds for various receptor sites (Table 31.30-2) is one method for comparing the likelihood that various agents will produce side effects. The other pharmacological factor that needs to be considered is the bio-availability of the drug.

Central Nervous System In the central nervous system, the principle action of the tricyclic and tetracyclic agents is to alleviate depression. These agents do not have a direct mood-elevating effect. They do not produce euphoria. Rather, they reduce the symptoms of depression. Apparently related to their effects on brain amines, these compounds can cause or aggravate psychotic symptoms and induce mania.

The central anticholinergic and antihistaminic effects of the tricyclic and tetracyclic drugs can produce confusion or delirium. The incidence of delirium is dose dependent; it increases at blood concentrations above 300 ng/mL and may rise to 67 percent when blood concentrations exceed 450 ng/mL in patients receiving the tertiary amines particularly amitriptyline.

Seizures can occur with all of the tricyclic and tetracyclic agents. Seizures are dose and blood concentration related. For clomipramine, the risk for seizures is reported to be 0.5 percent in doses up to 250 mg a day; at higher doses, the seizure risk increases to 1.67 percent. For maprotiline, the overall risk of seizures is reported to be 0.4 percent, but, again, the risk of seizures increases at dosages above the maximum recommended dosage of 225 mg a day. Determination of seizure risk for some of the older compounds was hampered by less rigorous data collection at the time of marketing. In one meta-analysis of data for imipramine, patients receiving less than 200 mg a day had a seizure rate of 1 in 1000. In patients receiving more than 200 mg a day, the rate was 0.6 percent. In patients receiving 250 to 450 mg a day, reported rates vary from 1 to 4 percent, but since several of these samples were small, these rates had wide confidence intervals. Seizure rates for the secondary amines have not been well described. The mechanism by which tricyclic drugs produce seizures is not well understood, although it is proposed that antidepressant drugs may act at the g-aminobuteric (GABA)-receptor chloride-ionophore complex, inhibiting chloride conductance and consequently inducing convulsions. The risk of convulsions is clearly increased in patients with predisposing factors including a prior history of seizures, brain injury, and presence of neuroleptic agents. The risk of seizures increases substantially following overdose.

A fine, rapid tremor can occur with the tricyclic agents and may be a clinical indication of an elevated blood concentration. Because the 7-hydroxy metabolite of amoxapine has neuroleptic properties, administration of amoxapine carries the potential risk of neuroleptic malignant syndrome (which has been reported) and tardive dyskinesia.

Several of the tricyclic compounds and maprotiline have clinically significant antihistaminic effects. Doxepin is one of the most potent H_1 receptor blockers known. Central H_1 receptor blockade can contribute to sedation and delirium. Antihistaminic effects also appear to be related to the increased appetite and associated weight gain that patients develop with long-term treatment.

Autonomic Nervous System The tricyclic drugs block muscarinic receptors and can cause a variety of anticholinergic adverse effects including dry mouth, constipation, blurred vision, urinary hesitancy, and an ocular crisis in patients with narrow-angle glaucoma. The tricyclic and tetracyclic compounds vary substantially in their muscarinic potency (Table 31.30-2). Amitriptyline is the most potent, followed by clomipramine. Of the tricyclic drugs, desipramine has the least severe anticholinergic effects. Amoxapine and maprotiline also have minimal anticholinergic effects. Anticholinergic effects can contribute to tachycardia, but tachycardia commonly occurs in patients receiving desipramine, which has minimal anticholinergic effects. It seems likely that noradrenergic effects also contribute to the increased pulse. Increased sweating is another secondary effect that can occur with use of the tricyclic compounds. The mechanism for this symptom is unclear, but it may be associated with noradrenergic effects.

Cardiovascular Effects Orthostatic hypotension is one of the most common reasons for discontinuation of tricyclic antidepressant treatment. Orthostatic hypotension can occur with all of the tricyclic drugs, but it is less pronounced with nortriptyline. It has been suggested that orthostatic hypotension is caused by a $_1$-adrenergic blockade; however, the postural reflex is primarily affected. Resting supine blood pressure may be unaffected or may be raised. Orthostatic hypotension is most likely to occur or is most severe in patients who have preexisting orthostatic hypotension.

Desipramine has been reported to raise supine blood pressure in younger patients, although it is not clear this effect is limited to that age group. This effect may be similar to that reported for venlafaxine (Effexor).

Tachycardia occurs with all tricyclic drugs, not just the more anticholinergic agents. Both supine and postural pulse changes can occur. Patients do not accommodate to the pulse rise, which can persist for months. Tachycardia is more prominent in younger pa tients who appear more sensitive to the sympathomimetic effects of these agents and is one of the most common reasons for drug discontinuation in this age group.

The effect of tricyclic antidepressants on cardiac conduction has been a subject of great interest. Cardiac arrhythmia is the principle cause of death following overdose. As a result, for many years there was great concern about the use of tricyclic antidepressants in patients with heart disease. The effect of these agents has now been well described. These agents have type I antiarrhythmic qualities, or quinidine-like effects. Apparently, through inhibition of Na^+, K^+-adenosine triphosphatase (ATPase), the tricyclic compounds stabilize electrically excitable membranes and delay conduction, particularly His-ventricular conduction.

At therapeutic blood concentrations, tricyclic drugs can have beneficial effects on ventricular excitability. Alternatively, in patients with preexisting conduction delay, the tricyclic antidepressants can further delay conduction and cause heart block. The QTc interval is regarded as the best index of who might be vulnerable to this effect. Patients with a QTc interval of 450 milliseconds or more are at increased risk and are not candidates for tricyclic antidepressant treatment. The risk of cardiac toxicity is further increased by high blood concentrations. For example, first-degree atrioventricular heart block is increased with imipramine plasma concentrations above 350 ng/mL and is increased more than 30-fold in patients with plasma concentrations above 450 ng/mL.

Tricyclic drugs do not appear to adversely affect cardiac contractility or cardiac output. Studies using radionuclide angiography indicate no adverse effect of imipramine or doxepin on cardiac output, even in patients with diminished left ventricular ejection fractions. Severe orthostatic hypotension was common in these patients.

Recently, another possible hazard of tricyclic drugs was suggested. Based on the report of the Cardiac Arrhythmia Suppression Trial (CAST) studies, which found that type I antiarrhythmic drugs given following myocardial infarction actually increased the risk of sudden death, Alexander Glassman, Steven Roose, and Thomas Bigger suggested that the tricyclic drugs may pose similar risks. Thus the clinician considering long-term use of tricyclic drugs in patients with ischemic heart disease needs to weigh this possible risk against possible benefits.

Hepatic Effects Drug-induced hepatitis can occur with these agents. Mild increases of liver enzymes are not uncommon and usually can be monitored safely over a period of days or weeks without apparent harmful consequences. However, acute hepatitis can occur quickly and is associated with very high enzyme levels (e.g.,

aspartate aminotransferase [AST] levels >800) within days. The rise in enzymes can be either hepatocellular or cholestatic and may precede clinical symptoms, especially in the hepatocellular form. This is a dangerous and potentially fatal condition. The antidepressant must be discontinued. The etiology is not well established but, in some cases, appears to be a hypersensitivity reaction. The risk of severe drug-induced hepatitis is not well established but appears to be between 1 in 1000 and 1 in 100.

THERAPEUTIC INDICATIONS

Indications The tricyclic and tetracyclic compounds have a variety of clinical uses. In fact, it is interesting to consider what these agents might have been called originally had their effectiveness in panic disorder, obsessive-compulsive disorder, pain, or other syndromes been described before their antidepressant action.

All of the tricyclic and tetracyclic compounds discussed in this chapter are approved by the Food and Drug Administration (FDA) for use in the United States for the treatment of depression, with the exception of clomipramine. Clomipramine is also used for depression and in Europe is regarded by many as the most potent antidepressant. This view is based on the dual action of the compound on both serotonin and norepinephrine and empirical data suggesting that this agent is more effective than SSRIs in severely depressed inpatients.

While a variety of antidepressants appear to be effective in depressive disorders, this is not the case for obsessive-compulsive disorder. In this instance the serotonergic antidepressants (e.g., clomipramine) appear to be substantially more effective than noradrenergic tricyclic drugs (e.g., desipramine). Among the tricyclic drugs, clomipramine is the only agent approved by the FDA for use in obsessive-compulsive disorder.

Other Indications The tricyclic agents have been used for a variety of other clinical indications without FDA approval. They are effective in panic disorder; in fact, imipramine was the first drug found effective in these patients.

Amitriptyline has been popular among nonpsychiatrists for use in chronic pain syndromes and migraine headache prophylaxis. During treatment of pain, the usual dosage is lower (up to 75 mg a day) and effects occur more quickly (usually within 3 to 4 days) than in depression. These differences suggest a different mode of action of amitriptyline in pain syndromes. Tricyclic drugs appear to be of some benefit in chronic fatigue syndrome.

For many years, desipramine was commonly prescribed for attention-deficit/hyperactivity disorder in children. Reports of sudden death, apparently of cardiac origin, in children under 12 years, substantially reduced the use of this drug in younger patients. Imipramine has been used for treatment of enuresis in children, with FDA approval. It is doubtful, however, that the risk of cardiac problems would be substantially lower with imipramine than with desipramine in young children.

These drugs have been extensively studied in schizophrenic patients. However, in the absence of a major depressive syndrome, they appear to have limited value, and their use for treatment of negative symptoms has been eclipsed by the introduction of the atypical neuroleptic agents.

PRECAUTIONS AND ADVERSE REACTIONS

Central Nervous System Effects The central anticholinergic effects of tricyclic drugs can produce delirium and seizures. Both of these adverse effects are dose dependent and become more frequent at elevated blood concentrations. The incidence of delirium increases at blood concentrations above 300 ng/mL and becomes common in levels above 450 ng/mL, especially with amitriptyline. The clinician should be aware of this possibility because the signs of delirium may be mistaken for a worsening of the depression. This diagnostic dilemma can be especially problematic in patients with psychotic depression. Patients with concurrent dementia are particularly vulnerable to the development of delirium. Intramuscular or intravenous physostigmine (Antilirium, Eserine) can be used to reverse or reduce the symptoms of delirium. However, the short duration of action of this agent makes its continued use difficult.

Because delirium and seizures depend on dosage and blood concentration, the first goal is to reduce the risk by avoiding unnecessarily high serum concentrations. Some experts advocate routine blood concentration monitoring. This question continues to be debated. A strong case can be made for monitoring in patients at increased risk (e.g., elderly patients and patients with dementia or a history of seizures). Blood monitoring can also be helpful in patients who fail conventional doses, to determine what dosages are required to reach therapeutic levels.

Autonomic Effects As mentioned, tricyclic drugs can cause a variety of anticholinergic adverse effects, including dry mouth, constipation, visual changes, urinary hesitancy, delirium, and an ocular crisis in patients with narrow-angle glaucoma. Use of nortriptyline or desipramine, which are less anticholinergic, can help reduce the likelihood of these problems.

Anticholinergic effects can be annoying but usually are not serious. In some circumstances, however, these effects can be severe. An ocular crisis in patients with narrow-angle glaucoma is an acute condition associated with severe pain. Urinary retention can be associated with stretch injuries to the bladder. Constipation can progress to severe obstipation. (Although paralytic ileus has been described, seldom are bowel sounds absent.) Patients with obstipation have considerable distress. Medication must be discontinued, and appropriate supportive measures instituted. The frequency of severe anticholinergic adverse reactions is increased in elderly persons and in patients receiving concomitant neuroleptic drugs.

Some anticholinergic effects may benefit from other interventions. Patients with urinary hesitancy may benefit from the use of bethanechol (Urecholine) at a dose of 25 mg three or four times a day. Patients with constipation often benefit from regular use of stool softeners. Patients with narrow-angle glaucoma who are receiving pilocarpine eye drops regularly can be treated with a tricyclic drug, as can those who have had an iridectomy. Patients with chronic open-angle glaucoma are not affected by tricyclic agents.

Cardiovascular Effects Orthostatic hypotension is one of the most common reasons for discontinuation of tricyclic antidepressant treatment. Orthostatic hypotension is less common with nortriptyline and is most likely to occur in patients who have preexisting orthostatic hypotension. Elderly patients are not only more likely to have preexisting hypotension but are also more vulnerable to the consequences of orthostatic hypotension—falls and hip fracture. Often, orthostatic hypotension occurs at relatively low blood concentrations. Further, while the subjective experience of lightheadedness may improve with time, the actual orthostatic blood pressure changes do not accommodate within a reasonable period of time (e.g. 4 weeks). Thus, unless the plasma concentration is elevated and the dosage can be reduced, patients who are experiencing serious symptomatic orthostatic hypotension may not be treatable with a tricyclic antidepressant. Fludrocortisone (Florinef) has been used to raise blood pressure, but it is not very effective. If patients are receiving antihypertensives, it may be possible and helpful to reduce these agents.

Tachycardia occurs with all the tricyclic drugs, not just the more anticholinergic agents. Both supine and postural pulse changes can occur. Patients do not accommodate to the pulse rise, which can persist for months. While an eight- to ten-beat increase in heart rate may be subjectively tolerated by the patient, a sustained increase in heart rate may have adverse consequences for the elderly patient because of increased cardiac work. Tachycardia is more prominent in younger patients who appear to be more sensitive to the sympathomimetic effects of these agents and is one of the most common reasons for drug discontinuation in this age group. Propranolol (Inderal) at low dosages may help to reduce the tachycardia unless it is associated with significant hypotension that propranolol might aggravate.

Cardiac Conduction Effects The tricyclic antidepressants have type I antiarrhythmic effects. Patients without conduction delay tolerate the tricyclics well. They may be helpful in patients with premature ventricular beats. However, the tricyclic antidepressants can further delay conduction in patients with preexisting conduction delay and cause heart block. As mentioned above, patients with a QTc interval of 450 milliseconds or more are at increased risk and are not candidates for tricyclic antidepressant treatment.

Hepatic Effects Mild increases of liver enzymes usually can be monitored safely over a period of days or weeks without apparent harmful consequences. However, as discussed above, acute hepatitis develops quickly and is associated with very high enzyme levels. In this situation, the antidepressant must be discontinued. If it is clear which drug caused the hepatitis, that drug should not be introduced again because the next reaction may be more severe. Unfortunately it is not uncommon for the patients to be receiving several medications, making the offending agent hard to identify. If a random blood test indicates mildly elevated liver enzymes, enzyme levels can be followed for a few days to distinguish the more severe form of hepatitis from mild enzyme elevations.

Other Adverse Effects Sexual dysfunction is a potential problem with a variety of antidepressant drugs but appears to be more frequent with clomipramine. Unlike many of the adverse effects, which are more likely to be reported early in treatment, this effect may be more bothersome to the patient in the second or third month of

treatment, after the depression has improved. Numerous pharmacological antidotes have been described but without controlled trials.

Tricyclic drugs can cause allergic rashes. They are sometimes associated with photosensitivity reactions. Various blood dyscrasias also have been reported, but fortunately, these are very rare.

Overdose Because the risk of overdose is increased in depressed patients, the target population for antidepressants use, the lethality of antidepressant drugs in overdose is of great concern. It is suggested that a tricyclic overdose of 10 times the daily dose can be fatal. Death most commonly occurs as a result of cardiac toxicity. However, seizures and central nervous system depression can occur. In 1996 amitriptyline was the second leading cause of death by overdose in the United States (after acetaminophen). The total number of fatalities for amitriptyline was comparable to those for all other tricyclic and tetracyclic drugs combined; yet, the other tricyclic and tetracyclic compounds are also dangerous in overdose. Desipramine has the highest fatality rate (deaths per ingestions). Amoxapine has been reported to produce high rates of seizure in overdose. Because the newer selective serotonergic antidepressants are generally safe in overdose, subtle differences in lethality between the tricyclic and heterocyclic compounds have become relatively unimportant.

Teratogenicity A definitive link between the tricyclic and tetracyclic compounds and teratogenic effects has not been established, although isolated reports of morphogenesis have been reported. Neonatal drug withdrawal can occur. This syndrome includes tachypnea, cyanosis, irritability, and poor sucking reflex. If possible, tricyclic and tetracyclic medications should be discontinued 1 week prior to delivery. The tricyclics are excreted in breast milk at concentrations similar to those in plasma. The actual quantity delivered, however, is small (e.g., if the drug concentration is 300 ng/mL and the amount of milk ingested is 120 mL, the drug dose is only 36 µg, or less than 0.1 mg). These low amounts result in undetectable drug levels in the infant. Because the risk of relapse is a serious concern in patients with recurrent depression and these risks may be increased in some patients during pregnancy or the postpartum period, the risks and benefits of continuing or withdrawing treatment need to be discussed with the patient and weighed carefully.

DRUG INTERACTIONS

Pharmacodynamic Interactions There are two principle types of drug interactions—pharmacodynamic and pharmacokinetic. Pharmacodynamic interactions are those in which the actions of two drugs interact, usually in a deleterious manner. Because tricyclic drugs have acute effects on amine uptake, they are potentially dangerous when given with the MAOIs. The most dangerous sequence is to give a large dose of a tricyclic drug to a patient already on an MAOI. This can result in a sudden increase in catecholamine concentration and a potentially fatal hypertensive reaction. If used cautiously, these two compounds can be used together to treat patients with refractory depression. Treatment is begun with lower doses and the two compounds are started together, or the tricyclic is administered first. Once begun, coadministration may actually reduce the risk of tyramine reactions; however, because this effect is variable, the usual MAOI diet is maintained.

Other pharmacodynamic interactions can occur. By blocking the transporters, tricyclic drugs block the uptake and thus interfere with the action of quanethidine. Quinidine poses two problems. First, the tricyclics have quinidine-like effects, so their effects on cardiac conduction are potentially additive with those of quinidine. In addition, quinidine is a potent enzyme inhibitor that can raise tricyclic drug levels, further adding to the problem.

Pharmacokinetic Interactions Most of the recent emphasis on drug interactions has focused on pharmacokinetic interactions, especially enzyme inhibition. A number of drugs can block the metabolic pathways of tricyclic drugs, resulting in higher and potentially toxic levels. Desipramine has been of particular interest because its metabolism is fairly simple, via the CYP 2D6 isoenzyme (Fig. 31.30-2). Because there are no major alternative pathways, inhibition of CYP 2D6 can result in high desipramine plasma concentrations. A number of drugs inhibit CYP 2D6. Quinidine, mentioned above, is a potent CYP 2D6 inhibitor. Drugs of special concern in psychiatry are fluoxetine and paroxetine (Paxil). At usual doses (e.g., 20 mg a day) fluoxetine, on average, raises desipramine levels three- to fourfold in extensive metabolizers. In slow metabolizers, enzyme inhibitors have less effect because the patient is already deficient in this enzyme. Some individuals on desipramine may experience a greater effect, but they are likely to be rapid metabolizers who would have low levels on a usual dose. Paroxetine would be expected to have effects similar to those of fluoxetine. Sertraline (Zoloft) has less effect on CYP 2D6. At 50 mg a day, sertraline, on average, increases desipramine levels about 30 to 40 percent. Higher doses evoke proportionally greater inhibition, but it is still substantially less than the 300 to 400 percent increase seen with fluoxetine or paroxetine. Venlafaxine, nefazodone (Serzone), mirtazapine (Remeron), and citalopram (Celexa), appear to have minimal effects on CYP 2D6. Nortriptyline metabolism would also be expected to be blocked by CYP 2D6 inhibitors, but the magnitude of this interaction has not been well studied.

The other group of drugs commonly used in psychiatry that interact with tricyclic drugs are the antipsychotic drugs. Chlorpromazine and perphenazine, at antipsychotic doses, inhibit CYP 2D6. The magnitude of the effect appears to be somewhat less than that with the SSRIs (e.g., perphenazine on average doubles desipramine levels), but this effect varies with dosage and the neuroleptic drug used.

Because tertiary tricyclic drugs are metabolized by several pathways (CYP 1A2, 3A4, and 2C19), an inhibitor of one pathway would be expected to have less effect on these compounds. Drug interactions with the tertiary amines do occur but appear to be less robust and are not as well described as those for desipramine.

The other type of pharmacokinetic drug interaction is enzyme induction. Unlike enzyme inhibition, which occurs quickly, enzyme induction involves synthesis of new enzyme. As a result, the full effect of an enzyme inducer may take 2 to 3 weeks to develop and similarly, if the inducer is discontinued, takes 2 to 3 weeks to dissipate. Barbiturates and carbamazepine (Tegretol) are the two potent inducers most commonly encountered. Although CYP 2D6 is thought to be noninducible, phenobarbital reduces the availability of desipramine by about 40 percent. Apparently, other enzymes such as CYP 3A4 are induced, which speed up the metabolism of the tricyclic drugs. Attaining an effective blood concentration of a tricyclic drug in the presence of barbiturates or carbamazepine can be difficult. Phenytoin (Dilantin) is also an enzyme inducer, but its effects on the tricyclic drugs appear to be less dramatic. Nicotine induces the CYP 1A2 pathway and may affect tertiary tricyclic drugs; secondary tricyclic drugs such as desipramine appear to be less affected by nicotine.

Alcohol has a complicated interaction. Acute ingestion of alcohol can reduce first-pass metabolism resulting in higher tricyclic drug concentrations. Because of the association of alcohol use and tricyclic overdose, this may be an important factor resulting in higher tricyclic levels following overdose. Alternatively, chronic use of alcohol induces hepatic isoenzymes and may lower tricyclic drug levels.

LABORATORY INTERFERENCES

The tricyclic compounds are present at low concentrations (nanograms per milliliter) and are not likely or known to interfere with other laboratory assays. It is possible that they may interfere with the determination of neuroleptic blood concentrations because of their structural similarity and the low concentrations of some neuroleptic drugs.

Two types of assays are performed for the tricyclic and tetracyclic antidepressants. Following overdose, an immunoassay is commonly performed to determine whether a tricyclic is present and estimate the concentration. This assay is sensitive, simple to perform, and rapid. However, the immune assays are not specific and can cross-react with metabolites and other drugs; thus they are not precise quantitative measures. A quantitative measure is needed for therapeutic drug monitoring, and a slower turnaround time for the laboratory is usually acceptable. A high-performance liquid chromatography (HPLC) assay is sufficiently precise for this purpose and is usually employed. In a competent laboratory the coefficient of variation is usually less than 10 percent. This assay is relatively specific; but other drugs can interfere. Because the HPLC technique has many modifications (the extraction method, the column used), and because these modifications determine which drugs will interfere, the interfering drugs vary by site. A clinician who frequently obtains tricyclic plasma concentrations is best advised to use one laboratory and find out from that laboratory what agents interfere with their assay.

DOSAGE AND ADMINISTRATION

At one time, amitriptyline and imipramine were the two most commonly prescribed antidepressants, but because of their anticholinergic and antihistaminic adverse effects, they are no longer the tricyclics of choice. Nortriptyline and desipramine are now preferred because of better tolerability. Nortriptyline has the least effect on orthostatic hypotension and the well-established therapeutic window for nortriptyline can be viewed as an advantage because blood concentration monitoring can be used to ensure a therapeutic level. Desipramine is the least anticholinergic and can also be monitored to achieve adequate concentrations. In theory, desipramine should be slightly safer in a patient at risk for cardiac conduction problems, because the hydroxymetabolite concentrations are lower than those for nortriptyline; however, this theoretical advantage has not been empirically demonstrated and in overdose, desipramine appears to be the most lethal tricyclic drug.

Most studies of nortriptyline, find that an average dose needed to reach a blood concentration of 100 ng/mL is about 75 mg a day; however, as indicated, there is considerable variability. For desipramine, the average dose required to reach a target level of 160 ng/mL is 3.0 mg/kg, or 180 mg a day for a 60-kg patient. Both

nortriptyline and desipramine can be adjusted to full dosage quickly, but the clinician needs to assess such factors as medical health, age, concurrent medications, and the severity of the depression before deciding how quickly to adjust dosage. In older patients, medically ill patients, or patients with less-severe depression, treatment can be started by giving a low dose once a day at bedtime. If tolerated well, the dosage can be increased every 3 days. Once the patient is receiving a lowmoderate dose (e.g., 50 mg of nortriptyline) a blood sample can be drawn after 5 to 7 days to determine the plasma concentration. Because the kinetics of nortriptyline are linear, a proportional adjustment in dosage can be made to reach a level of 100 ng/mL. With desipramine, dosage can also be adjusted relatively proportionately, but patients with very low initial plasma levels (very rapid metabolizers) are likely to have nonlinear changes, thus the dosage increase is reduced. Gradual dosage increases are safer but can take weeks. The clinician needs to weigh the need for safety against the need for rapid treatment. This need is usually obvious in severe depression, but rapid treatment can also help reduce the likelihood that a patient who is already discouraged will drop out of treatment. Once the patient is at an adequate dosage, treatment should be continued for 4 to 6 weeks. If there is no change after 3 to 4 weeks, other strategies might be considered. If there is partial response after 3 to 4 weeks, the drug should be given more time. There is some controversy about the value of increasing the dosage if the patient fails this trial. If the plasma concentration is just within the effective range (e.g., 55 ng/mL for nortriptyline or 125 ng/mL for desipramine), then increasing the dosage may be helpful. However, if the drug plasma concentration is already well within the therapeutic range (e.g., 100 ng/mL for nortriptyline or 160 ng/mL for desipramine), the likelihood of response with further dosage increase is minimal, and substantially below what would be achieved with augmentation strategies or switching to a new compound. Further, for nortriptyline, if the plasma level is near the upper limit (e.g., 140 ng/mL) a dosage reduction could be considered. One of the values of a tricyclic trial using drug concentration monitoring is that the efficacy of the drug can be determined after a single trial at an adequate concentration.

If patients experience adverse effects this approach may be altered. Patients on desipramine who report insomnia can be dosed early in the day. For both nortriptyline and desipramine, sedation and anticholinergic effects may diminish with time, and gradual dosage adjustment may help to manage these effects. Blood pressure and pulse changes do not usually accommodate during the first 4 weeks, and gradual dosing does not help to avoid these effects. In patients at high risk for adverse effects, the dosage can be increased in gradual steps every 2 to 3 days to assess tolerance. If anticholinergic adverse effects occur, allowing more time between dose increases will allow the patient to accommodate to these side effects.

Prospective Dosing Techniques Conventional dosing techniques involve administration of a given dose for a long enough time (usually 4 to 6 weeks) to determine whether that dose is effective. Sometimes two or even three drug trials are needed before the best or necessary dosage is determined. Once therapeutic blood concentrations were established, it was tempting to think that dosage could be rapidly adjusted to achieve an effective drug concentration during the first drug trial. Various methods were devised for rapidly adjusting dosage. In one method, a single test dose was administered, two or more samples were drawn to determine clearance of the drug, and the appropriate dosage was calculated. A simpler method used a similar test dose but only one sample was drawn, usually 24 hours later; then dosage was adjusted on the basis of previously established data or a nomogram. This method appeared to be most applicable for nortriptyline, which has linear kinetics, and was also effective for desipramine, because while some patients on desipramine had nonlinear changes, the targeted level was within a broad range. Prospective dosing methods were also reported for amitriptyline and imipramine; however, because of sedation and anticholinergic adverse effects, rapid dosing of these drugs was not feasible.

Several factors have limited the popularity of prospective dosing methods. First, the drug concentration after a single dose is very low, often between 10 and 20 ng/mL. The laboratory performing the assay must be prepared to determine drug concentrations accurately in this range. However, most laboratories do not use low internal standards that permit precise determination of low values and the error is magnified during dosage adjustment. Second, the utility of a blood concentration value for rapid dosage adjustment requires rapid turnaround time from the laboratory. Tricyclic concentration determinations are complicated and tedious and most laboratories are not prepared to provide these values quickly. Finally, even if a precise initial blood concentration value were available quickly, both the clinician and patient may be reluctant to increase the dose quickly to its full level, particularly in an outpatient setting. A more practical and clinically feasible method is to start treatment with a low-moderate fixed dosage, obtain a blood sample after 5 to 7 days on that dosage, and make further adjustments on the basis of that result.

Exceptions Elderly depressed patients may require more gradual dosing to assess tolerance, but most studies of desipramine and nortriptyline indicate that the final dosages and blood concentrations required for response will be similar to those needed in younger patients.

Lower starting dosages are used in panic patients, and dosage is increased slowly to avoid exacerbating panic attacks. In obsessive-compulsive disorder, higher doses of clomipramine are used than are used in depression. In part, this is because obsessive-compulsive disorder responds more slowly and less completely than depression; thus, dosage is raised to ensure the most effective treatment.

If amitriptyline is used for pain, dosages up to 75 mg a day are used. When effective, the action of the drug is observed within the first week.

At this time, it is not clear when the tricyclic antidepressants would be used in children. It is recommended that desipramine not be used in children under 12 because of possible adverse cardiac effects, but the same logic might apply to other tricyclic drugs. The tricyclic drugs have not been shown to be more effective than placebo in adolescents with depression, although this probably relates in part to diagnostic problems. Conceivably, a tricyclic drug might be given to a depressed adolescent whose parent has had a good response to a tricyclic drug. The lethality of tricyclic drugs in overdose should be kept in mind. Younger patients are likely to require higher doses than adults (on a milligram per kilogram basis) to reach the same plasma concentration, but effective plasma ranges have not been established for children.

P>Tricyclic and heterocyclic preparations are shown in Table 31.30-3.

Drug	Tablets	Capsules	Parenteral	Solution
Imipramine	10, 25, 50 mg	75, 100, 125, 150 mg	12.5 mg/mL	—
Desipramine	10, 25, 50, 75, 100, 150 mg	—	—	—
Trimipramine	—	25, 50, 100 mg	—	—
Amitriptyline	10, 25, 50, 75, 100, 150 mg	—	10 mg/mL	—
Nortriptyline	—	10, 25, 50, 75 mg	—	10 mg/5 mL
Protriptyline	5, 10 mg	—	—	—
Amoxapine	25, 50, 100, 150 mg	—	—	—
Doxepin	—	10, 25, 50, 75, 100, 150 mg	—	10 mg/mL
Maprotiline	25, 50, 75 mg	—	—	—
Clomipramine	—	25, 50, 75 mg	—	—

Table 31.30-3 Tricyclic and Tetracyclic Drug Preparations

SUGGESTED CROSS-REFERENCES

More information on the basic chemistry and pharmacological activity of neurotransmitter systems may be found in Section 1.4 on monoamine neurotransmitters, Section 1.5 on amino acid transmitters, and Section 1.10 on basic molecular neurobiology. Section 5.4 describes animal research and its relevance to psychiatry, including the development of psychopharmacological drugs. The general principles of psychopharmacology are detailed in Section 31.1. Specific pharmacotherapeutic considerations for children are discussed in Section 48.6, and for geriatric patients, in Section 51.4.

SECTION REFERENCES

*APA Task Force: Tricyclic antidepressants—blood level measurements and clinical outcome. Am J Psychiatry *142:*155, 1985.

Asberg M, Cronholm B, Sjoqvist F, Tuek D: Relationship between plasma level and therapeutic effect of nortriptyline. Br Med J *3:*331, 1971.

Chan CH, Janicak PG, Davis JM: Response of psychotic and nonpsychotic depressed patients to tricyclic antidepressants. J Clin Psychiatry *48:*197, 1987.

Charney DS, Delgado PL, Price LH, Heninger GR: The receptor sensitivity hypothesis of antidepressant action: A review of antidepressant effects on serotonin function. In *The Role of Serotonin in Psychiatric Disorders*, SL Brown, HM van Praag, editors. Brunner/Mazel, New York, 1991.

Davis JM: Overview: Maintenance therapy in psychiatry: II. Affective disorders. Am J Psychiatry *133:*1, 1976.

*Davis JM, Glassman AH: Antidepressant drugs. In *Comprehensive Textbook of Psychiatry*, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1989.

Duman RS, Heninger GR, Nestler EJ: A molecular and cellular theory of depression. Arch Gen Psychiatry *54:*597, 1997.

*Frank E, Kupfer DJ, Perel JM, Cornes C, Jarrett DB, Mallinger AG, Thase ME, McEachran AB, Grochocinski VJ: Three-year outcomes for maintenance therapies in recurrent depression. Arch Gen Psychiatry *47:*1093, 1990.

Glassman AH, Bigger JT: Cardiovascular effects of therapeutic doses of tricyclic antidepressants. Arch Gen Psychiatry *38:*815, 1981.

Glassman AH, Perel JM, Shostak M, Kantor SJ, Fleiss JL: Clinical implications of imipramine plasma levels for depressive illness. Arch Gen Psychiatry *34:*197, 1977.

Glassman AH, Roose SP, Bigger JT Jr: The safety of tricyclic antidepressants in cardiac patients: Risk/benefit reconsidered. JAMA *269:*2673, 1993.

Joyce PR, Paykel ES: Predictors of drug response in depression. Arch Gen Psychiatry *46:*89, 1989.

Katz IR, Simpson GM, Jethanandani V, Cooper T, Muhly C: Steady state pharmacokinetics of nortriptyline in the frail elderly. Neuropsychopharmacology *2:*229, 1989.

Kocsis JH, Friedman RA, Markowitz JC, Leon AC, Miller NL, Gniwesch L, Parides M: Maintenance therapy for chronic depression. Arch Gen Psychiatry *53:*769, 1996.

Kuhn R: The imipramine story. In *Discoveries in Biological Psychiatry*, FJ Ayd, B Blackwell, editors. Lippincott, Philadelphia, 1970, p 205.

Lapierre YD: A review of trimipramine: 30 years of clinical use. Drugs *38:*17, 1989.

Liebowitz MR, Quitkin FM, Stewart JW, McGrath PJ, Harrison WM, Markowitz JS, Rabkin JG, Tricamo E, Goets DM, Klein DF: Antidepressant specificity in atypical depression. Arch Gen Psychiatry *45:*129, 1988.

Litovitz TL, Smilkstein M, Felberg L, Klein-Schwqartz W, Berlin R, Morgan JL: 1996 annual report of the American Association of Poison Control Centers toxic exposure surveillance system. Am J Emerg Med *15:*447, 1997.

Nelson JC, Jatlow PI, Quinlan DM: Subjective complaints during desipramine treatment: Relative importance of plasma drug concentrations and the severity of depression. Arch Gen Psychiatry *41:*55, 1984.

Nelson JC, Jatlow PI, Quinlan DM, Bowers MB Jr: Desipramine plasma concentrations and antidepressant response. Arch Gen Psychiatry *39:*1419, 1982.

*Nelson JC, Kennedy JS, Pollock BG, Laghrissi-Thode F, Narayan M, Nobler MS, Robin DW, Gergel I, McCafferty J, Roose S: Treatment of major depression with nortriptyline and paroxetine in patients with ischemic heart disease. Am J Psychiatry, *156:*322, 1999.

Nelson JC, Mazure CM, Jatlow PI: Desipramine treatment of major depression in patients over 75 years of age. J Clin Psychopharmacol *15:*99, 1995.

Onghena P, van Houdenhove B: Antidepressant-induced analgesia in chronic non-malignant pain: A meta-analysis of 39 placebo-controlled studies. Pain *49:*205, 1992.

Pimentel L, Trommer L: Cyclic antidepressant overdoses: A review. Emerg Med Clin North Am *12:*533, 1994.

*Pollock BG: Drug interactions. In *Geriatric Psychopharmacology*, JC Nelson, editor. Marcel Dekker, New York, 1997.

Potter WZ, Calil HM, Sutfin TA, Zavadil AP, Jusko WJ, Rapoport J, Goodwin FK: Active metabolites of imipramine and desipramine in man. Clin Pharmacol Ther *31:*393, 1982.

Preskorn SH, Fast GA: Therapeutic drug monitoring for antidepressants: Efficacy, safety, and cost effectiveness. J Clin Psychiatry *52* (Suppl 6):23, 1991.

*Reynolds CF, Miller MD, Pasternak RE, Frank E, Perel JM, Cornes C, Houck PR, Mazumdar S, Dew MA, Kupfer DJ: Treatment of bereavement-related major depressive episodes in later life: A controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry *156:*202, 1999.

Roose SP, Glassman AH, Giardina EGV, Walsh BT, Woodring S, Bigger JT: Tricyclic antidepressants in depressed patients with cardiac conduction disease. Arch Gen Psychiatry *44:*273, 1987.

Roose SP, Laghrissi-Thode F, Kennedy JS, Nelson JC, Bigger JT Jr, Pollock BG, Gaffney A, Narayan M, Finkel MS, McCafferty J, Gergel I: Comparison of paroxetine and nortriptyline in depressed patients with ischemic heart disease. JAMA *279:*287, 1998.

Rosenstein DL, Nelson JC, Jacobs JC: Seizures associated with antidepressants: A review. J Clin Psychiatry *54:*289, 1993.

Rudorfer MV, Potter WZ: Pharmacokinetics of antidepressants. In *Psychopharmacology: The Third Generation of Progress*, HY Meltzer, editor. Raven, New York, 1987.

Ryan ND: The pharmacologic treatment of child and adolescent depression. Psychiatr Clin North Am *15:*29, 1992.

Salzman C, Schneider L, Lebowitz B: Antidepressant treatment of very old patients. Am J Geriatr Psychiatry *1:*21, 1993.

Spiker DG, Weiss JC, Dealy RS, Griffin SJ, Manin I, Neil JF, Perel JM, Rossi AJ, Soloff PH: The pharmacologic treatment of delusional depression. Am J Psychiatry *142:*430, 1985.

Thase ME, Fava M, Halbreich U, Kocsis JH, Koran L, Davidson J, Rosenbaum J, Harrison W: A placebo-controlled, randomized clinical trial comparing sertraline and imipramine for the treatment of dysthymia. Arch Gen Psychiatry *53:*777, 1996.

31.31 ELECTROCONVULSIVE THERAPY

KEITH E. ISENBERG, M.D. AND CHARLES F. ZORUMSKI, M.D.

Electroconvulsive therapy (ECT) is an important nonpharmacological intervention that is effective treatment for patients suffering from certain severe neuropsychiatric disorders. Psychotropic medications produce significant improvement in most patients with major affective disorders and schizophrenia. However, a substantial minority of patients either do not respond to pharmacological treatment, suffer severe adverse effects that make medications intolerable, or have severe symptoms that benefit from urgent intervention and rapid response. For such patients, ECT offers a useful, safe, and, in some cases, lifesaving intervention. ECT involves the passage of a brief electrical current through the brain to induce a generalized central nervous system (CNS) seizure under general anesthesia and muscle relaxation. Although favorable response to ECT can occur quickly, the clinical benefits typically require multiple treatments administered over a period of several weeks. The procedure is generally well tolerated by patients and can be used prophylactically to sustain a partial or complete remission of symptoms. Adverse effects such as memory impairment are minimized by recent advances in the administration of ECT. Presently, ECT is the most effective treatment available for hospitalized patients with major depression, and it is useful in the management of patients with mania, catatonia, schizophrenia, and certain neurological disorders.

HISTORICAL OVERVIEW

The first use of convulsive therapy for the treatment of a psychiatric disorder in modern times dates to Ladislaus von Meduna in 1934. The decision to use this form of treatment was based on the belief that patients with psychosis and epilepsy exhibited improvement in psychotic symptoms following spontaneous seizures. Hypothesizing a biological antagonism between seizures and psychosis, von Meduna used camphor injections to induce seizures in psychotic patients. Of the first 26 patients treated by von Meduna, 13 exhibited partial or complete remission of symptoms. The long action of the camphor injections was associated with numerous clinical problems, which prompted a search for a less traumatic way to induce seizures. von Meduna found that pentylenetetrazol, an agent now known to inhibit g-aminobutyric acid (GABA) type A ($GABA_A$) receptors, was an improvement over camphor injections but was still associated with a number of problems.

In 1938, Luigi Bini and Ugo Cerletti documented the first therapeutic use of electrically induced seizures in humans. Bini and Cerletti were intrigued by the idea of using controlled doses of electrical current for therapeutic purposes but had initially found the treatment dangerous in animals. Although a number of animals stimulated electrically died, Bini and Cerletti observed that the deaths resulted from direct current flow through the heart and that current applied across the head did not carry severe cardiac risks. Following a series of successful experiments in animals, these investigators administered the first ECT to a 39-year-old man who suffered from an acute psychosis. The patient was treated with a series of 11 electrically induced seizures and experienced a dramatic improvement in his psychotic disorder. Electrical induction of seizures was more reliable than the pharmacological approaches used by von Meduna. Following the success of Cerletti and Bini, Lothar Kalinowsky introduced ECT to the United States in 1939.

These initial efforts opened the door to more general use of ECT in an era when most of the currently available psychotropic agents were unknown. Early treatments were fraught with numerous difficulties, including a high incidence of fractured and dislocated long bones as well as compression fractures of the spine. The fractures resulted from severe muscle contractions produced by the direct electrical current flow through the body and the tonic phase of the generalized seizures. In 1940 A. E. Bennett introduced the use of curare as a muscle relaxant to avoid these muscle contractions and minimize the risk of fractures. Equally notable improvements in ECT anesthetic technique included routine use of general anesthesia and vigorous oxygenation throughout the procedure coupled with cardiac and oxygen saturation monitoring to prevent cardiopulmonary complications and minimize hypoxia.

Other innovations in ECT technique (Table 31.31-1) have been developed in an effort to maintain the therapeutic efficacy of the treatment, prevent major adverse effects, and improve the acceptance of the procedure. The therapeutic benefits of chemically induced seizures and the failure of subconvulsive electrical impulses to induce an effective clinical response suggested the necessity of a seizure for ECT to be efficacious. Brief-pulse electrical stimulators and nondominant hemisphere electrode placements diminish the cognitive adverse effects of the treatment. As currently practiced, ECT is a relatively safe and effective means of treating patients with major psychiatric disorders.

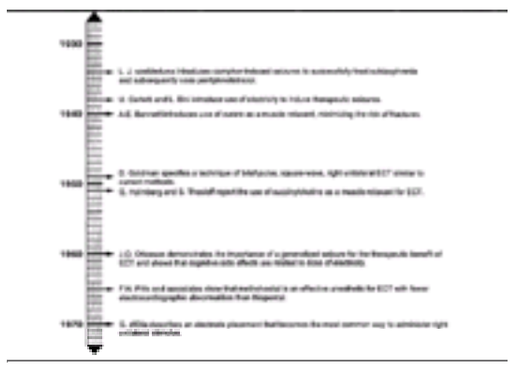

Table 31.31-1 Historical Developments in the Administration of Electroconvulsive Therapy

The use of ECT in psychiatry has waxed and waned over the years. Initial enthusiasm for ECT in the treatment of schizophrenia was gradually replaced by awareness that the procedure was particularly helpful for depression. Innovations described above may result in greater acceptance but usually add to the expense of the procedure as well as enhancing the benefits of ECT. The treatments, largely given at public institutions in the years after inception, are currently more frequently administered at private institutions. It is presently estimated that as many as 100,000 patients per year are treated with ECT.

ELECTRICAL FACTORS IN ECT

ECT Stimulators Considerable evidence indicates that the repeated production of generalized CNS seizures is required to produce the clinical benefits of ECT. Thus, the goal of an ECT treatment session is to induce a generalized seizure of “adequate” duration in the CNS. Subconvulsive electrical stimuli or those inducing only partial (focal) seizures have no therapeutic benefit. Similarly, treatments in which seizures are terminated immediately following stimulation are ineffective. Understanding how seizures are produced by electrical stimulation and how the electrical stimuli used for ECT are quantified requires knowing several basic electrical principles.

Excitable tissues are stimulated by the flow of current (or more properly by the movement of ions across the cell membrane). Current (I) is measured in units of amperes and is defined as the amount of charge (Q) measured in coulombs flowing per unit time (t). Thus, $I = Q/t$. The force that drives current flow is the applied

electrical field (measured in volts [V]). The relationship between current and voltage is given by Ohm's law: $V = IR$, where R is the resistance to current flow measured in ohms. Ohm's law holds for direct current (DC) circuits. An ECT treatment involves the application of electricity as an alternating current (AC); the resistance term of the circuit is more properly described as impedance (Z). *Impedance* includes the DC resistance as well as terms for *capacitance* (the ability to store charge on conductors that are separated by an insulator) and *inductance* (the ability to induce a voltage across the tissue). Since the brain usually contains no metal, the inductance term is not relevant. However, the ability to store charge on either side of the lipid bilayer is a fundamental biophysical property of cell membranes; thus capacitance is important to the ECT circuit. Furthermore, impedance has both static and dynamic components; unfortunately, little is presently known about the dynamic impedance changes associated with ECT. In the ECT circuit, most of the static impedance to current flow is across the skull (approximately 18,000 ohms/cm), while the impedance to current flow across the skin and through brain tissue is only about 200 ohms/cm. When an electrical stimulus is applied via electrodes placed on the patient's head, the low-impedance pathway is along the skin between the electrodes. Thus, most of the stimulus is shunted between the stimulating electrodes and little (<20 percent) enters the cranial cavity to stimulate the nervous system.

Based on these electrical principles, it is possible to understand how ECT stimulators work. These devices typically have only one of two outputs: they are either current generators or voltage generators. In a constant-current stimulator, the applied current is independent of the impedance between the electrodes. According to Ohm's law, the applied voltage varies directly with the impedance. Thus when the impedance between the electrodes is high, the applied voltage from a constant-current stimulator can become very high (sometimes exceeding 500 V, depending on the maximal output of the stimulator). Because the power (measured in watts) dissipated between the electrodes is a product of current and voltage ($P = IV$, or $P = I2R$), significant risk of local tissue damage exists if the impedance between the electrodes is too great. With most constant-current stimulators, the operator is required to perform a self-test prior to stimulating the patient. This test administers a low-amplitude current to test the interelectrode impedance. If the impedance exceeds a limit deemed safe by the manufacturer, the test fails. In the case of a failed self-test, improved contact between the electrodes and the skin usually lowers impedance. Somewhat counterintuitively, when impedance between the ECT electrodes is too low, induction of convulsions with a constant-current stimulator can be more difficult. The difficulty with low-impedance seizure induction develops because the applied voltage becomes too low to drive significant current flow through the high resistance of the skull.

Most ECT devices presently used in the United States are constant-current stimulators. A second type of ECT stimulator uses a constant voltage source. In these devices, the current varies with the resistance between the electrodes; high impedance can cause difficulty inducing seizures because the current flow is too low. Constant-current stimulators offer the advantage of easier quantification of the electrical stimulus. Because current is fixed, the amount of charge (coulombs) administered is simply a product of the current and the time during which current flows (Charge [Q] = It). With a constant-voltage stimulator, calculations of administered charge require information about impedance.

The waveform of the output is another important characteristic of ECT stimulators. Older ECT devices were sine-wave generators. At a frequency of 60 cycles per second (60 Hz), each half sine wave lasts 8.3 ms with a significant stimulus flowing about 75 percent of this time. A basic property of neuronal action potentials, the cellular activity driving generalized seizures, is a duration of a few milliseconds. Furthermore, following an action potential there is a period of several milliseconds during which it is either impossible or relatively difficult to fire a second action potential (termed the *absolute and relative refractory periods*). During a sine-wave stimulus, much of the current flow occurs during inexcitable periods. Thus, sine waves tend to drive neuronal firing rather inefficiently. A constant-step stimulus applied for a long period of time is even more inefficient. A device that administers repeated brief pulses (0.5 to 2.0 milliseconds) of current to trigger action potential firing at rates similar to the intrinsic firing patterns of neurons in critical regions of the CNS is preferred. The benefits of brief-pulse stimuli compared with step pulses or sine waves have been documented in experimental preparations. Evidence suggests that pulses less than 0.5 millisecond in duration (referred to as “ultrabrief” pulses) are likely to be ineffective ECT stimuli. However, this issue may need reexamination in light of recent studies demonstrating the importance of seizure threshold in determining clinical outcome from ECT.

Most ECT stimulators use brief-pulse outputs, typically at frequencies of 30 to 100 Hz. When brief-pulse outputs are given with a constant-current generator, it is relatively easy to quantify the electrical stimulus. The administered charge is calculated by adding the total time that brief pulses are applied and multiplying this duration by the pulse amplitude. In most constant-current stimulators each cycle consists of one positive and one negative pulse. Thus, the calculation of stimulus duration (D) is given by: D = pulse width × pulse frequency × 2 × train duration. Most constant-current stimulators used in the United States have maximal charge outputs of 500 to 600 mC. Assuming an interelectrode impedance of 200 ohms, this output translates into a stimulus energy of less than 100 joules (watt-seconds). Because stimulus energy requires information about the interelectrode impedance and impedance measures must take into account both static and dynamic factors, it has been preferable to quantitate ECT stimuli in units of charge rather than units of energy.

Seizure Threshold The work of Harold Sackeim and colleagues demonstrated that features of the electrical stimulus interact with the mode of stimulus to play a role in the therapeutic benefits of ECT. When treatments are administered with electrodes placed bitemporally (Fig. 31.31-1), minimally suprathreshold electrical doses produce significant clinical benefits. However, treatment with nondominant hemisphere unilateral electrode placement at stimuli minimally above the seizure threshold produces only marginal clinical improvement, despite inducing what appear to be generalized seizures of adequate duration. Sackeim and colleagues extended this work to demonstrate that benefits of unilateral ECT increase significantly when electrical doses at least 2.5 times the seizure threshold are used. These studies suggest that the degree to which the electrical stimulus exceeds the seizure threshold is critical in determining therapeutic effects of unilateral treatments. Sackeim and colleagues confirmed earlier work suggesting that electrical dose plays a role in the cognitive adverse effects of ECT. Higher stimulus intensities are associated with greater memory impairment. Thus, attention to seizure threshold is becoming a major concern in centers that use ECT.

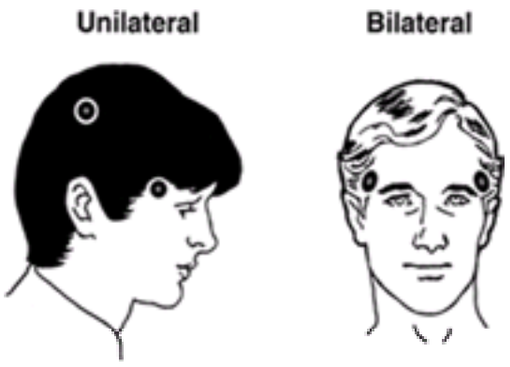

FIGURE 31.31-1. Types of electrode placement used with ECT—bilateral (right) and right unilateral nondominant (left). In the latter case, a wide centroparietal-frontotemporal derivation (d'Elia placement) is depicted.

Seizure threshold is defined empirically as the minimum amount of electrical charge that induces a generalized CNS seizure. There is some debate concerning the proper length of a threshold seizure and whether duration should be measured by electroencephalogram (EEG) or by motor seizure in an isolated limb. Some centers use a cutoff of 25 seconds, but this limit is arbitrary. Sackeim and colleagues found a great deal of patient-to-patient variability in seizure threshold, with estimated values exhibiting about a 40-fold difference. A number of factors appear to influence seizure threshold, including age, sex, and electrode placement. Seizure thresholds tend to be higher in men than in women and higher in older patients than in younger patients. Age-related differences may reflect differences in skull density as well as plasticity of an aging nervous system. Electrode placement also plays a major role, with bilateral (bitemporal) placements having a higher threshold than non–dominant hemisphere placements. Other variables include the patient's electrolyte and hydration status (e.g., hyponatremia and hypocalcemia are associated with lower seizure thresholds) as well as concomitant use of CNS-active medications. Anticonvulsants, benzodiazepines, barbiturates, sedative-hypnotics, and certain antiarrhythmic drugs are associated with higher seizure thresholds, whereas theophylline derivatives, lithium, and sedative-hypnotic and alcohol withdrawal are associated with lower thresholds. ECT has anticonvulsant effects, so recent treatment with ECT can influence threshold measurements. Recent work suggests that perhaps the most important determinant of seizure threshold with current stimulators is pulse duration and frequency.

The nature of the electrical stimulus and the anesthetic drugs used for the ECT procedure may also influence seizure threshold. For example, sine wave stimuli are inefficient and tend to yield higher thresholds. With brief-pulse treatments, pulses of shorter duration (0.5 ms) and lower frequency (30 to 40 Hz) produce lower thresholds than longer-duration pulses (2.0 ms) administered at high frequencies (70 to 100 Hz) for a given amplitude of current pulses. Among anesthetic agents, barbiturates and propofol (Diprivan) clearly raise seizure threshold. Anesthetics that may have less effect on seizure threshold include etomidate (Amidate), ketamine (Ketolar), and alfentanil (Alfenta), but this issue has not received adequate systematic study. Prior to delivering the electrical stimulus, vigorous hyperventilation and

decreased carbon dioxide concentration may also help to lower threshold.

Several electrical dosing schedules have been described for estimating seizure threshold. Typically these dosing regimens begin with a low electrical charge (e.g., 25 mC); increases in the charge are delivered according to a predetermined plan until a generalized seizure is induced. In clinical settings, threshold titrations involving a minimal number of stimulations (four or five) are preferred to diminish the risks associated with titration. The last stimulation in the titration series is given at maximal charge. About 30 seconds are allowed between stimulations to ensure that the prior stimulus has not produced a seizure. When the stimulus is near threshold, onset of a generalized seizure may be delayed for several seconds.

ECT MECHANISMS

Possibly no treatment in psychiatry is more misunderstood or more maligned than ECT. In part, this stems from the fact that ECT is one of the most invasive procedures used by psychiatrists. Furthermore, critics of ECT are often bothered because the mechanisms responsible for the therapeutic and adverse effects are not well understood. The lack of mechanistic understanding should come as no surprise, since unequivocal information about the pathophysiology of most major psychiatric disorders is lacking. The lack of understanding of ECT mechanisms does not result from a lack of hypotheses. Systematic reviews reveal more than a hundred theories in the literature dealing with this topic, ranging from psychological and psychodynamic explanations to molecular, biochemical, neuroendocrine, and structural theories. Various theories are not mutually exclusive but rather indicate the complexity of the CNS and the multitude of academic disciplines interested in providing mechanistic explanations.

Certain hypotheses can be dismissed fairly easily. For example, there is no evidence that fear, psychological regression, increased medical attention, anesthetic medications, or brain hypoxia result in therapeutic effects. Similarly, memory loss does not contribute to the benefits of ECT. Most of the major innovations in ECT technique, including the use of brief-pulse generators, titrated electrical doses, and nondominant hemisphere stimulation, have been directed toward minimizing this adverse effect while maintaining treatment efficacy. The latter point is particularly important since memory loss is a popular lay theory about ECT. Opponents of ECT have argued that the treatment works by damaging the CNS. This topic has been dealt with in detail in two excellent reviews by Richard Weiner and D. P. Devanand and colleagues. Based on careful and thorough reviews of the literature, including both human and animal studies, these authors conclude strongly that there is no evidence that ECT produces structural damage to the brain.

Currently, it is believed that the therapeutic and adverse effects of ECT result from changes in CNS biochemistry and physiology. Furthermore, the beneficial effects of ECT require several treatments over a period of several days, which has spurred considerable interest in understanding the effects of repeated brief seizures on CNS function. One major difficulty in understanding critical effects of ECT is that the treatment is administered to human beings with severe psychiatric disorders, while most detailed biochemical studies are conducted in presumably normal animals, particularly rodents. While studies in animals must be viewed with some skepticism, insights generated in animal studies have often been useful in understanding human pathophysiology. Improvements in CNS functional imaging methods lend hope that sophisticated biochemical studies in humans will be possible in the future. An additional problem with understanding the effects of ECT is that the treatment affects multiple CNS systems. Studies of the effects of repeated seizures are unlikely to produce insights into disease mechanisms because it will be difficult to determine causal relations.

The best hope for understanding ECT mechanisms may exist when considering disorders whose underlying biology is better understood; Parkinson's disease is one such example. In this disorder there are clear disruptions of dopaminergic transmission in the striatonigral pathway. Drugs that increase dopamine function or diminish muscarinic cholinergic function are useful in treating motor symptoms of the disorder. Thus, it might be expected that ECT would alter dopamine or muscarinic cholinergic responses in the CNS. Indeed, electroshock (ECS) in animals rapidly increases dopamine concentrations in the frontal cortex and striatum while having variable effects on resting dopamine concentrations. Dopamine autoreceptor sensitivity is diminished following ECS and D_1 dopamine receptor agonists enhance stimulation of adenylate cyclase. These effects suggest enhanced dopamine transmission following ECS. Effects on D_1 receptor binding have been variable, with no change in the striatum and increases in the substantia nigra. Although less consistently found, evidence suggests that ECS diminishes muscarinic function in some CNS regions. Together, these effects of ECS suggest plausible mechanisms for the antiparkisonian effects of ECT in humans.

P>Over a course of treatment, ECT has anticonvulsant effects that raise seizure threshold and decrease seizure duration. There has been considerable work on the cellular correlates of seizure generation in the CNS; this is also an area in which the effects of ECT can be potentially understood. At a fundamental level, anticonvulsant drugs do one of two things: enhance inhibition or diminish excitation. Since GABA is the principal fast inhibitory transmitter in the mammalian CNS and drugs that enhance GABAergic transmission via GABA, receptors (e.g., barbiturates, benzodiazepines, loreclezole) have anticonvulsant effects, it might be expected that ECS would alter this system. Some evidence suggests that GABA levels increase in certain CNS regions following ECS, suggesting a possible increase in tonic inhibition. There is also evidence that ECS causes increases in $GABA_B$ (baclofen [Lioresal]-sensitive) receptors. These sites mediate both presynaptic and postsynaptic inhibition at certain synapses. It is less clear whether ECS produces consistent changes in $GABA_A$ receptors. Interestingly, ECS produces changes in endogenous opiate systems that could have anticonvulsant effects, and some evidence suggests that the anticonvulsant effects of ECS in animals are blocked by naloxone (Narcan), a broad-spectrum opiate receptor antagonist. ECS also upregulates certain opiate receptors (e.g., those for D-alanine-D-leucine-enkephalin [DADLE]. A third potential anticonvulsant mechanism is via release of endogenous adenosine. Periods of increased electrical activity in the CNS, including seizures, promote extracellular release of adenosine, which in turn acts on several receptor types to produce inhibition. Adenosine A_1 receptors are upregulated in neocortex but not hippocampus or striatum following ECS, suggesting regional increases in inhibition. Caffeine, a drug used to increase ECT seizure durations clinically, antagonizes adenosine A_1 receptors, suggesting a possible link to seizure termination mechanisms. However, caffeine has other effects (e.g., phosphodiesterase inhibition and release of calcium from intracellular stores) that could also contribute to increased duration of seizure. Effects of ECS on the glutamate system that is the prime mediator of fast excitatory synaptic transmission could also be important but have received less systematic attention.

Understanding how ECT produces memory impairment would be helpful, since it could potentially lead to strategies to diminish this adverse effect. Although cellular mechanisms involved in memory are not well understood, changes in several transmitter systems could contribute. For example, the antimuscarinic effects of ECS could be involved, because drugs that disrupt this system produce cognitive impairment and delirium in humans. Similarly, the increased inhibition that possibly contributes to the anticonvulsant effects of ECS could be involved, and drugs that alter inhibitory systems (e.g., barbiturates, benzodiazepines, alcohol) are associated with amnesia (blackouts). Some evidence suggests that the phenomenon of long-term potentiation (LTP) could be a cellular correlate of memory formation in the CNS. This process occurs when glutamatergic synapses are used repeatedly under conditions in which receptors of the *N*-methyl-d-aspartate (NMDA) class are activated. Although the mechanisms are not well understood, it appears that ECS and generalized seizures disrupt LTP generation in animals. An interesting possibility is that ECT-induced seizures produce a form of NMDA receptor–mediated LTP inhibition that has been described in rodent hippocampal slices. Since there is presently intense interest in mechanisms involved in LTP, these insights may shed some light on ECT-induced cognitive impairment, particularly anterograde amnesia.

Because the mechanisms involved in psychiatric disorders are so poorly understood, several indirect strategies have been used to assess ECT mechanisms. One of the favored strategies has been to compare the effects of ECS in animals with the effects of psychotropic medications. This approach assumes that ECT and psychoactive drugs exert their effects by similar means. Certainly some of the effects described above could contribute to beneficial effects of ECT. For example, the usefulness of anticonvulsants (valproic acid [Depakene], carbamazepine [Tegretol]) in management of affective disorders makes it possible that the anticonvulsant effects of ECT contribute to therapeutic benefits. One interesting finding is that ECS and a variety of antidepressant medications produce b-adrenergic receptor subsensitivity, suggesting some commonality in mechanism. Additionally, ECS increases norepinephrine turnover and a_1-adrenergic receptor sensitivity and decreases presynaptic a_2-adrenergic receptors. ECS also appears to enhance the effects of the serotonergic system but differs from chronic antidepressant medication treatment in producing increases in serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptor binding in the cerebral cortex. These findings suggest that ECT may have important actions on monoaminergic transmission that contribute to therapeutic effects.

Advances in understanding receptor biology and gene function in the CNS over the past decade offer hope of eventually achieving a better understanding of the CNS effects of ECT. Beyond transmitter receptors, considerable interest now exists in defining effects of ECT on second messenger systems and gene transcription. Effects on gene transcription and protein synthesis could be particularly important, since effects of the treatment tend to develop slowly and persist for some period of time. Similar approaches in the study of CNS plasticity during chronic exposure to psychotropic medications and drugs of abuse are already paying dividends in understanding long-term changes associated with drug exposure.

INDICATIONS FOR ECT

Major Depression For patients with severe major depression, ECT is the most effective treatment available and the standard against which other treatments can be

judged. A substantial body of experimental evidence supports the use of ECT in the treatment of depression (Table 31.31-2). ECT was developed prior to general acceptance of the prospective, randomized, double-blind clinical trial as the most rigorous method of evaluating benefits of treatment, a circumstance that applies to numerous other procedures and many psychotherapies. Evaluation of the efficacy of ECT was further facilitated by development of standardized assessment instruments and acceptance of well-defined diagnostic criteria. Results of studies shown in Table 31.31-2 deserve further comment. Substantial numbers of patients participated in each trial; the number in parentheses indicates the number of individuals enrolled in a given trial, when this information is available. Five of the six studies clearly show ECT to be superior to sham procedures; benefits of ECT occur despite the weak but consistent pattern of improvement observed in the placebo-treated population, similar to that seen in other trials of antidepressant treatment. Most of the studies used sine-wave stimulators to induce a seizure. Studies reported at later dates may have benefited from flaws noted in the design and execution of the first two studies reported. In aggregate these studies demonstrate the overwhelming superiority of ECT to sham, a result that is sustained in clinical practice.

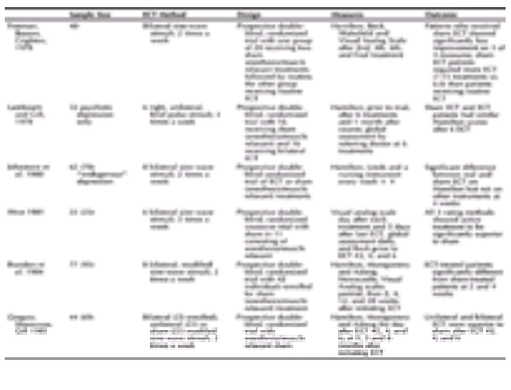

Table 31.31-2 Effectiveness of ECT in the Treatment of Depression

ECT is effective in producing significant-to-complete symptom remission in more than 70 percent of patients with previously untreated major depression. Despite this benefit, ECT rarely is the treatment of first choice for depression. In part, this stems from the usefulness of pharmacological and psychotherapeutic interventions for depression and from the significant adverse effects of ECT, particularly memory loss and confusion. Furthermore, misunderstanding and misinformation among psychiatrists and the lay populace about the role of ECT compound the problem. ECT is most often used for depressed patients who have failed to respond to one or more trials of antidepressant medications, including individuals who cannot toler ate the adverse effects of medications as well as those who do not respond to therapeutic doses of medications. Thus the effectiveness of ECT in treating patients who are known medication nonresponders is of interest, since these patients could be unresponsive to any biological treatment. Depending on how ECT response is defined, it appears that more than 50 percent and perhaps as many as 70 percent of medication nonresponders exhibit significant improvement with ECT. A major difficulty in this population is not the initial response to ECT but rather the ability to maintain remission following clinical improvement. Evidence now indicates that patients who have failed particular medications should not be treated again with those medications after they respond to ECT. Rather, other antidepressant medications, particularly those to which the patient has previously shown a good response or to which the patient has not been previously exposed, should be used for maintenance therapy. Alternatively, following a successful course of ECT, continuation or maintenance ECT can be implemented with good clinical results. The difficulty with maintenance ECT lies in the need to administer treatments on an ongoing, outpatient basis and problems with treatment adverse effects, particularly memory impairment.

ECT should be strongly considered as the initial choice for some depressed patients, including individuals with particularly severe symptoms, including psychosis, marked suicidal intent, and refusal to eat. Because response to medication is often delayed for days to weeks, ECT offers the prospect of more rapid, perhaps lifesaving response. Other patients for whom ECT is considered first-line therapy include those with a prior history of nonresponsiveness to medications who have had a prior good response to ECT. Finally, consideration should be given to patient preference; psychiatrists should always present ECT as a treatment option to patients suffering from depression, an important alternative to psychotherapy and pharmacotherapy.

No clinical symptoms of depression or demographic variables uniquely predict a favorable response to ECT; it has been used to treat depression in a variety of clinical populations, including elderly adults, medically ill patients, pregnant women, and adolescents. No evidence suggests undue risk in any of these populations or lack of ECT responsivity. When considering ECT use in these populations, the risks and benefits of treatment must be weighed carefully, particularly in patients with severe medical illnesses for whom anesthesia alone carries increased risk. In most regions, ECT use in adolescents is restricted legally but may be administered following consultation with parents or guardians and appropriate medicolegal review.

Some depressed patients may not respond to ECT. For example, some evidence suggests that depression in patients with preexisting somatization disorder may be less responsive to the treatment. Additionally, ECT appears to be relatively ineffective for some psychiatric disorders, including somatization disorder, personality disorders, obsessive-compulsive disorder, and primary anxiety disorders. In some of these patients, however, ECT can be effective in treating concomitant depressions that do not respond to antidepressant therapy.

Mania In general, acute mania responds well to treatment with mood stabilizers and antipsychotic medications, the first-line treatments for the disorder. Some patients with mania, however, either cannot tolerate medications or fail to respond to medications. Additionally, some patients with mania have severe symptoms, including exhaustion, extreme agitation, and delirium. ECT should be considered as an effective therapeutic alternative for these patients. The prospectively collected evidence that ECT is effective treatment for mania is presented in Table 31.31-3. The number in parentheses indicates the total number of patients enrolled in the larger trial; over 75 percent completed the initial phase of this trial. Although neither of the studies is as well designed as the studies considering the benefits of ECT in the treatment of depression, both used constant-current square-wave generators to induce seizure and are thus similar to current ECT practice. A number of retrospective studies further support the observation that perhaps more than 80 percent of individuals suffering from mania exhibit a partial or complete remission of symptoms. Electrode placement in treating mania is controversial, with some studies indicating a better response to bilateral treatments. As in depression, manic patients should be treated with ECT until symptomatic improvement plateaus or adverse effects preclude further treatments. When ECT is stopped, pharmacological management with medications such as mood stabilizers should be instituted. Maintenance ECT can also be an option for some patients with bipolar disorder. Because of its effectiveness against both mania and depression, ECT should be considered early in the treatment of depressed patients with bipolar disorder.

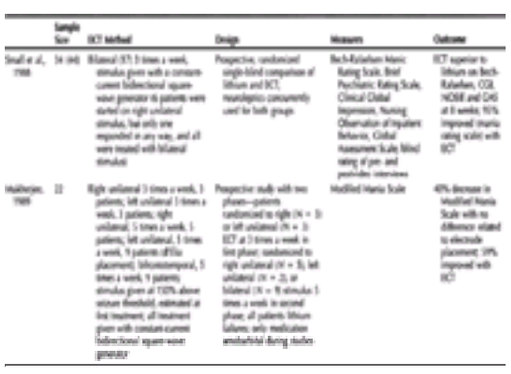

Table 31.31-3 Effectiveness of ECT in the Treatment of Mania

Schizophrenia There is little evidence that ECT alters the chronic symptoms of schizophrenia, including either chronic positive (delusions, hallucinations) or negative (flat affect, poverty of thought content) features of the disorder. Nevertheless, ECT can be effective in management of acute psychotic exacerbations, with as many as 80 percent of patients responding favorably. In schizophrenic patients with acute exacerbations, antipsychotic medications are the mainstays of treatment. ECT should be reserved for patients who either have very severe symptoms or who cannot tolerate or fail treatment with psychotropic medications. Acute psychotic symptoms typically respond to 8 to 12 treatments, although some patients may require additional treatments depending on their clinical response and, ability to

tolerate adverse effects. Treating schizophrenic patients with prolonged courses of ECT appears to produce no substantial clinical benefit and is associated with severe memory loss and confusion. Some schizophrenic patients who respond well to a short-term course of ECT may benefit from maintenance treatments, although most of these patients will also be treated with antipsychotic medications.

Catatonia Catatonia is a syndrome characterized by stupor, mutism, and motor symptoms and is associated with a variety of psychiatric, neurological, and medical disorders. Most often the syndrome is observed in patients with schizophrenia or bipolar disorder. Regardless of the underlying disorder, catatonia responds relatively well to ECT, with improvement in motor symptoms (posturing, rigidity, catalepsy) usually occurring after only a few (three to five) treatments. Some patients, however, require longer courses of treatment for full remission of symptoms. ECT can be particularly effective in patients who develop extremely severe symptoms including exhaustion (lethal catatonia) and delirium. The beneficial effect of ECT upon motor symptoms may explain its usefulness in management of patients with severe forms of neuroleptic malignant syndrome, a syndrome that shares a number of features with severe catatonia, including marked muscular rigidity and autonomic instability.

Neurological Disorders In addition to its beneficial effects in catatonia and neuroleptic malignant syndrome, ECT also improves motor symptoms of Parkinson's disease, particularly rigidity and bradykinesia, independent of effects on the depressive symptoms that often accompany this disorder. Interestingly, ECT may prolong the “on” periods in patients who experience the “on-off” phenomenon associated with antiparkinsonian drug use. In general, the beneficial effects of ECT in parkinsonism are time-limited, with relapse occurring over 4 to 6 weeks. Some evidence suggests that maintenance ECT can benefit these patients, but systematic studies examining this issue are needed.

Over a course of treatment, ECT has anticonvulsant effects, with increased seizure threshold and decreased seizure duration. In the past, this prompted the use of ECT to treat patients with intractable epilepsy. Presently, no evidence indicates that ECT is more effective for these patients than anticonvulsant medications, and the treatment is rarely, if ever, used for this purpose. The anticonvulsant effects of ECT are time limited, dissipating over several weeks. Epileptic patients referred for ECT do not appear to be at substantially increased risk for worsening of their seizures either during or following a course of ECT. Patients with epilepsy who are referred for ECT should remain on anticonvulsant medications through the course of treatment because significant risks are associated with discontinuing these medications. There is no need to increase the dosage of anticonvulsants during a course of ECT unless increases in medication are warranted on the basis of clinical history.

Since memory loss and confusion are major adverse effects of ECT, it is surprising that it has been used in the management of severe delirium. As with the catatonic syndrome, ECT improves delirium associated with a variety of psychiatric, medical, and neurological disorders. In particular, ECT appears to be effective in managing the severe agitation and exhaustion that some of these patients experience. Most of the evidence for beneficial effects of ECT in delirium is from anecdotal reports, and ECT is usually considered only in patients who have failed more conventional management.

GENERAL PATIENT MANAGEMENT

Informed Consent ECT is typically an elective medical procedure involving general anesthesia, muscle relaxation, and electrical stimulation. The procedure should be performed after obtaining in formed patient consent, documented in writing. Consent requires a thorough discussion with the patient and when possible, with the patient's family, of the details of the procedure, indications for treatment, risks from treatment, adverse effects, alternative forms of treatment, and risks of no treatment. A number of excellent pamphlets and videotapes are available to aid psychiatrists in presenting this information. While these can be very helpful, they do not replace direct dialogue between the treating physician and the patient.

Some patients with severe psychiatric disorders are unable or unwilling to give informed consent for ECT. Yet, ECT may be indicated and deemed necessary to prevent serious risks to the patient. Although restrictions on the use of ECT in such cases vary, most jurisdictions provide for use of the treatment involuntarily following proper legal review. Whenever ECT is used, but particularly when it is used without the patient's consent, one must weigh the risk-benefit ratio carefully.

Pre-ECT Evaluation All patients treated with ECT require careful psychiatric and medical histories and a thorough physical examination. The psychiatric history should include sufficient information to make a clear DSM-IV diagnosis, document prior response to pharmacological and psychotherapeutic interventions, describe any prior treatment with ECT (including response to treatment and adverse effects), and discuss the indications for ECT. The medical history and physical examination should focus on any documented medical problems. Of particular importance is any prior history of neurological, cardiorespiratory, or gastrointestinal disorders that might impinge on the patient's ability to tolerate general anesthesia or the ECT procedure. The past medical history should also include a description of prior surgeries, including any difficulties with anesthetic medications in the patient or the patient's family. A careful oral examination is also needed, since severe dental problems may necessitate adjustments in the procedure to protect the patient's teeth.

The pre-ECT laboratory evaluation need not be extensive but should follow guidelines for outpatient surgical procedures. Emphasis in the laboratory evaluation should be placed on any disorders identified in the initial history and physical examination. Minimal laboratory tests include a complete blood count, serum electrolytes, and electrocardiogram (ECG). Older patients and those with a history of cardiorespiratory disorders, should have a chest x-ray. Spinal xrays are not required for ECT but should be considered for patients with a history of significant back problems. Neuroimaging studies are not routine prior to ECT. Rather, these tests should be reserved for patients with suspicious neurological histories and physical examinations, particularly those with focal neurological findings on examination. Other laboratory studies should be based on the history and physical examination.

ECT Anesthesia. In most settings, a team of at least three individuals is involved in the administration of ECT: a psychiatrist, an anesthesiologist or nurse anesthetist, and a nursing assistant. Depending on the size of the ECT service, other personnel involved might include staff performing pre-ECT patient evaluation and staff assisting with recovery from treatment. Because ECT is administered under general anesthesia and involves a generalized seizure, patients should have nothing to eat or drink for at least 6 hours prior to the procedure. This precaution greatly reduces the risk of aspiration during treatment. Minor exceptions to this rule can be made when, for example, the benefits of administering cardiovascular medications or agents that diminish gastric acid secretion prior to treatment are greater than the extremely low risk of aspiration associated with a sip of water necessary to give medication.

Standards developed for administration of anesthesia in outpatient surgical settings are generally appropriate for ECT. These standards include regimens for pre-ECT evaluation, anesthetic care, and postoperative recovery. ECT is typically administered in either a specialized treatment area or in an operating room setting. In either case, facilities must be available to manage medical emergencies (particularly cardiac events) and allow the patient to recover from the procedure. Prior to treatment, all patients must be evaluated by the person performing anesthesia to assess independently the risks of anesthesia and potential problems in managing the patient's airway. Consultation with the anesthesia provider may also suggest ways to maximize therapeutic benefits of ECT (e.g., choice of anesthetic drug) and to diminish adverse effects of treatment (e.g., use of analgesics, antiemetics, or both).

At the time of treatment, an intravenous catheter is inserted into a peripheral vein to provide access for medications and (if required) fluids. Patients are typically connected to several monitors including an ECG, two-lead EEG, and pulse oximeter. Respiratory support during ECT is almost always provided via ventilation mask. Prior to administering the anesthetic agent, the patient's mouth is inspected to remove any foreign objects and dentures. Patients are usually preoxygenated prior to anesthetic induction using 100 percent oxygen by mask; exceptions can be made if the patient suffers intolerable anxiety from having a ventilation mask over the face. Oxygen is administered throughout the procedure until recovery of sustained respiratory effort occurs. The routine use of oxygen prevents problems with hypoxia and diminishes cognitive impairment from treatment and risk of cardiac complications.

Two or three medications are typically used for the ECT procedure (Table 31.31-4). The information provided in the table offers guidance regarding medication administration, but each ECT practice should individualize medication usage to reflect the experience of providers. Because the passage of the electrical stimulus can produce asystole lasting 6 to 12 seconds, many patients are treated with an anticholinergic drug. Atropine is a typical agent, administered at doses of 0.4 to 1.0 mg intravenously immediately before stimulation. Individuals with substantial oral secretions (e.g., tobacco smokers) can be given atropine intramuscularly 30 to 60 minutes before the procedure. Advantages of atropine include its effectiveness as an anticholinergic and low cost. Disadvantages include increased heart rate, which can compound the tachycardia that typically accompanies an ECT-induced seizure. Additionally, atropine has CNS effects and is associated with memory impairment and delirium. The latter problems can be prevented by using a peripherally acting anticholinergic agent such as glycopyrrolate (Robinul; 0.2 to 0.4 mg intravenously), although patients may not tolerate the prolonged drying effect of this agent. Because anticholinergic-induced tachycardia can increase cardiovascular risks of ECT, the decision to use these agents is best made on a case-by-case basis. In our practice, patients with resting heart rates above about 90 beats per minute are unlikely to be treated with anticholinergics unless, for example, they are also taking medications that promote bradyarrhythmias (e.g., b-adrenergic receptor antagonists, verapamil [Isoptin], digoxin [Lanoxin]).

Drug	Dose
Anticholinergics	
Atropine	0.4–1.0 mg IV or IM
Glycopyrrolate	0.2–0.4 mg IV or IM
Anesthetics	
Methohexital	0.5–1.0 mg/kg IV
Thiopental	1.5–2.5 mg/kg IV
Etomidate	0.1–0.3 mg/kg IV
Alfentanil	0.2–0.3 μg/kg IV
Ketamine	0.5–1.0 mg/kg IV
Propofol	0.75–1.5 mg/kg IV
Midazolam	0.15–0.3 mg/kg IV
Muscle Relaxants	
Depolarizing	
Succinylcholine	0.75–1.5 mg/kg IV
Nondepolarizing	
Mivacurium	0.1–0.2 mg/kg IV
Atracurium	0.3–0.4 mg/kg IV
Antihypertensives	
Esmolol	0.05–0.1 mg/kg IV
Labetalol	0.04–0.2 mg/kg IV
Nifedipine	10–30 mg PO

Table 31.31-4 Medications Used in the Administration of ECT

Anesthetic agents are given because they tend to obliterate memory of the procedure. Intravenous anesthetics with rapid onset and brief duration of action are the agents of choice for ECT; inhalational anesthetics play little role in administration of the treatments. Methohexital (Brevital; about 1 mg/kg intravenously), a short-acting barbiturate, is the most commonly used ECT anesthetic; it has a rapid onset and short (about 10 minutes) duration of action. Advantages include ease of use, low incidence of adverse effects, and low cost. Potential problems include respiratory depression, risk in patients with porphyria, and laryngospasm at low doses. Thiopental (Pentothal), a barbiturate anesthetic commonly used for induction prior to surgical procedures, appears to carry significantly greater risk of ECG abnormalities during ECT than methohexital. Since other adverse effects of thiopental are similar to those seen with methohexital, methohexital is the agent of choice. Other intravenous anesthetics that can be used for ECT include etomidate, alfentanil, ketamine, and propofol. Among these, etomidate is particularly interesting because it has few cardiovascular adverse effects and perhaps less effect on seizure induction than barbiturates. Alfentanil and other short-acting opioids may prolong seizures but require concurrent administration of an amnestic agent (e.g., barbiturates such as methohexital) to perform effectively as anesthetics. Furthermore, opioids are notorious suppressants of respiratory effort, an effect that can be reversed with antagonists such as naloxone. Ketamine also has less effect on seizure induction but tends to worsen ECT-induced blood pressure and heart rate changes and, in adults, is associated with psychotomimetic symptoms that would be particularly detrimental in psychiatric patients. Propofol has strong anticonvulsant properties that may diminish its usefulness for ECT anesthesia. The benzodiazepines (e.g., midazolam [Versed]) are to be avoided because of their deleterious effects on seizure induction. Etomidate is probably the anesthetic agent to choose if a barbiturate such as methohexital is undesirable.

Routine use of muscle relaxants during ECT has virtually eliminated fractures of long bones and vertebrae. After the patient has been anesthetized, a blood pressure cuff or tourniquet is placed on an extremity (usually the right ankle) and inflated to exceed the systolic blood pressure. A short-acting muscle relaxant is injected intravenously. The blood pressure cuff isolates the limb from the muscle relaxant and allows monitoring of motor seizure activity. Succinylcholine (Anectine; about 1 mg/kg, intravenously), a depolarizing muscle relaxant, is the drug of choice for ECT. This drug has a rapid onset (about 1 to 2 minutes) and short (less than 10 minutes) duration of action that is almost ideal for ECT. Advantages include ease of use and low cost. Adverse effects of succinylcholine include muscle soreness (due to fasciculations), increased extracellular potassium (due to depolarization of the neuromuscular junction), and risk of prolonged apnea in patients with pseudocholinesterase deficiency. Risk of hyperkalemia is greatest in patients with recent trauma (fractures and burns) and in those with neuromuscular disorders. Patients with neuromuscular disorders are also at risk for developing malignant hyperthermia following succinylcholine use. For these patients, a nondepolarizing muscle relaxant is a more appropriate choice. Alternatives include atracurium (Tracrium) and mivacurium (Mivacron). For patients who experience recurrent intolerable muscle soreness from succinylcholine, pretreatment with a low dose of a nondepolarizing agent (e.g., mivacurium, 1 mg intravenously) can be beneficial.

Following injection of the muscle relaxant, a bite block is placed in the patient's mouth to protect teeth and tongue. The ECT stimulus is then applied via electrodes placed either bitemporally (bilateral ECT) or over the nondominant hemisphere (unilateral ECT). A bite block is required because passage of electrical current directly stimulates jaw muscles and causes teeth clenching. Since this problem cannot be prevented by use of muscle relaxants, patients should be warned before treatment that some risk of damage to the teeth is associated with the procedure, particularly in patients with poor baseline dentition. Following the electrical stimulus and seizure induction, the bite block is removed, and oxygenation via mask is continued until the patient resumes spontaneous and sustained respirations.

An ECT seizure usually lasts between 30 and 90 seconds; the observed motor seizure is typically shorter than the EEG seizure. Brief seizures (less than 25 seconds) can be followed by restimulation at a higher charge during the same session if anesthetic conditions permit. Prolonged-ECT seizures are defined as those lasting more than 180 seconds on EEG. Prolonged seizures are terminated by an intravenous anticonvulsant such as a barbiturate (methohexital, 30 to 50 mg) or a benzodiazepine (diazepam [Valium], 5 to 10; lorazepam [Ativan], 1 to 2 mg). Following the end of the seizure and full recovery of respirations, the patient is taken to the recovery area where blood pressure, pulse, ECG, and oxygen saturation are monitored until the patient is fully alert and able to ambulate. This typically takes about 30 minutes.

During each ECT treatment there are marked increases in blood pressure and pulse rate. Most patients tolerated these changes well, but hemodynamic changes may predispose patients with underlying cardiovascular illnesses to episodes of myocardial ischemia or rate-dependent arrhythmias. In the ECT area, marked increases in pulse and blood pressure can be dampened by judicious use of esmolol (Brevibloc; 0.2 mg/kg, intravenously), a short-acting b-adrenergic receptor antagonist, or labetalol (Trandate; 5 to 20 mg intravenously), a combined b- and a-adrenergic antagonist. These agents can be administered either before or after electrical stimulation. At higher doses, both agents have been associated with increases in seizure threshold, and they may increase poststimulus asystole in patients not treated with an anticholinergic drug. Thus, poststimulus administration of esmolol and labetalol may be preferred for ECT. Cardiovascular changes induced by ECT are brief and time limited, so the briefer duration of action may make esmolol the preferred agent. However, both esmolol and labetalol may produce bradycardia that persists beyond the treatment period. Esmolol and labetalol tend to be effective for ECT-induced increased heart rate, but in some patients are less effective for increased blood pressure. For patients who exhibit marked changes in blood pressure, premedication with nifedipine (Procardia; 10 to 30 mg orally about 30 minutes prior to treatment) can be useful. Other medications (e.g., nitroglycerin) may also be useful.

Electrode Placement Typically, one of two electrode placements is used for ECT: unilateral, nondominant hemisphere placement or bilateral placement. The electrodes are usually either hand-held devices or metal plates contained in a band that is affixed to the patient's head. Hand-held electrodes are easier to position, but the position may change as pressure is applied to form a good electrical contact. Several varieties of unilateral placement have been described, but most practitioners prefer, and most research supports, a temporoparietal (d'Elia) position. For all right-handed patients, the electrodes are placed over the right cerebral hemisphere. Most (more than 60 percent) of left-handed patients are either left hemisphere dominant for language or have mixed dominance, so a right hemisphere placement is appropriate for these patients as well. For bilateral treatments, electrodes are usually positioned bitemporally; other placements are experimental. Prior to positioning the electrodes, the skin must be carefully prepared to improve electrical contact and diminish interelectrode impedance. This is accomplished by cleansing the electrode area with a saline-soaked pad and coating the electrodes with a conducting gel.

When the electrical dose is taken into consideration and electrical charges that are more than 2.5 times the seizure threshold are used for unilateral treatments, both unilateral and bilateral ECT are effective in treating psychiatric disorders. Nondominant hemisphere electrode placements are the preferred initial mode of treatment for most patients, because this placement tends to produce less memory impairment than bilateral stimulus. Exceptions to this approach may include patients who have a history of poor response to prior courses of nondominant hemisphere treatment, a history of positive prior response to bilateral ECT, or extremely severe symptoms. When a patient is to be treated with bilateral ECT, the reasons for this selection should be clearly documented in the medical record.

For reasons that are not clear, some patients do not respond to unilateral treatments. If a patient fails to respond after four to eight unilateral treatments using an electrical dose that is 2.5 to 5.0 times the seizure threshold, several options are open to the treating psychiatrist. If possible, the electrical dose can be increased to more than five times threshold. If the patient fails to respond to this form of stimulation, electrode placement should be changed to bilateral using a minimally suprathreshold electrical dose. Failure to respond to low-charge bilateral stimulation should prompt a change to high-charge bilateral treatments. In all cases, a decision to increase electrical charge, switch to bilateral treatments, or both should be made after carefully considering the patient's cognitive status and after discussion with the patient of the risks and possible benefits of the change in treatment. A patient should not be considered an ECT failure until a course of high-charge bilateral ECT has been administered. Additionally, failure to respond to a prior course of ECT does not necessarily predict failure to respond to ECT at later times, particularly during different episodes of the disorder.

ECT Course An average ECT seizures lasts from 30 to 90 seconds. ECT has anticonvulsant properties, and over a course of treatments, seizure threshold increases and seizure duration decreases. Seizures lasting less than 25 seconds are considered less therapeutic than longer seizures yet are associated with the risks and

adverse effects of longer seizures. When seizures routinely last less than 25 seconds, several approaches can be use to lengthen them. First, vigorous hyperventilation prior to and during the seizure can lengthen seizures in some patients by diminishing carbon dioxide levels. Second, any medications that raise seizure threshold and that can be withheld safely should be discontinued; these include benzodiazepines, antidepressants, and anticonvulsants given for psychiatric indications. Third, consideration should be given to the dose and type of anesthetic drug. High doses of barbiturates clearly have anticonvulsant effects. Thus, either lowering the dose of barbiturate or changing the anesthetic to etomidate or ketamine can lengthen seizures in some patients. Alternatively, the dose of barbiturate can be significantly lowered (to 20 to 30 mg) and alfentanil (0.25 µg/kg) added to the regimen. Fourth, intravenous administration of caffeine (250 to 1000 mg) significantly prolongs seizure activity in most patients. Theophylline has effects similar to those of caffeine but has been associated with status epilepticus during ECT.

Sustained improvement in psychiatric symptoms rarely occurs with a single ECT treatment. Most, if not all, patients require a course of repeated treatments. A typical course of ECT consists of 6 to 12 treatments administered two or three times per week over a period of several weeks until improvement in target clinical symptoms reaches a plateau. There have been several attempts to speed this course of treatment by inducing multiple seizures in succession at a single treatment session (referred to as “multiple-monitored ECT”). There is little evidence that these multiple treatments enhance the benefits of ECT, but they seem to be associated with significantly greater memory impairment and confusion, particularly in elderly patients. In an effort to prolong the duration of improvement following ECT, some practitioners have administered prolonged courses of treatment (20 to 40 treatments). There is no evidence that prolonged courses of treatment enhance clinical benefit. In the past, prolonged courses of treatment were used to induce a state of cognitive impairment (so-called regressive ECT). This form of treatment has no place in modern ECT practice, since it has no clear therapeutic advantage and is associated with severe cognitive impairment. Fixed courses of treatment for all patients are inappropriate, and decisions about discontinuing ECT should be made on the bases of clinical response and adverse effects of treatment.

The benefits of ECT are time limited. Following a successful course of ECT, an effective maintenance strategy must be implemented for each patient. This typically involves aggressive pharmacological treatment using medications that the patient has not failed in the past and psychotherapy to facilitate psychosocial rehabilitation. ECT can also be used for maintenance treatment. Because risk of relapse following ECT is highest in the first 6 to 8 weeks following a successful course, maintenance treatments are usually administered once a week for the first 2 months. Treatment frequency is then gradually decreased, aiming at one treatment a month. After 6 to 12 months of maintenance ECT, treatments are usually discontinued, and the patient continues taking psychotropic medications. During the maintenance ECT period, patients are commonly treated with appropriate psychotropic medications.

CONTRAINDICATIONS TO ECT

There are no absolute contraindications to ECT. Previously, the presence of a brain tumor was viewed as an absolute contraindication to treatment because increases in cerebral blood flow and intracranial pressure that accompany seizures increase the risk of brain herniation, neurological decompensation, and death. However, numerous reports have described successful use of ECT to treat psychiatric disorders in patients with CNS tumors. At present, the presence of any CNS tumor must be viewed as significantly increasing the risk of ECT. Patients at greatest risk are those with large CNS lesions, evidence of increased intracranial pressure, evidence of midline shift on neuroimaging studies, or neurological examinations that show clear focal deficits. Patients with small lesions in silent regions of the CNS who lack the findings described above are at lower risk. Because changes in intracerebral pressure are related to changes in cerebral blood flow, pressor responses to ECT must be managed carefully in patients with CNS tumors. Use of steroids may diminish local swelling in the tumor area and decrease risks associated with ECT. In all patients with CNS tumors, the decision to use ECT should be made carefully in collaboration with neurologists and neurosurgeons.

Patients who have had neurosurgical procedures or who have had skull fractures may also be at increased risk for complications from ECT because of bony defects in the region of the surgery or fracture. These defects provide a low-resistance pathway for current to enter the CNS more directly during electrical stimulation, thus increasing risk of electrolytic brain injury. In these patients, ECT electrodes should not be placed directly over regions where the skull has been significantly compromised.

Other factors that increase the risk of complications from ECT include a recent myocardial infarction or stroke, particularly when ECT is administered within 4 to 6 weeks of the event. These patients are at increased risk from general anesthesia alone. Similarly, patients with uncompensated congestive heart failure are at increased risk of complications from the treatment. In all of these cases, close interaction between the treating psychiatrist and the patient's cardiologist or neurologist may suggest management that decreases risk.

ADVERSE EFFECTS AND RISKS

Cardiovascular Effects of ECT Significant changes in hemodynamic function occur during each ECT treatment. Immediately following the electrical stimulus, a period of increased parasympathetic tone produces profound bradycardia and asystole that can last several seconds. This initial asystole may be more pronounced with subconvulsive stimuli, particularly in patients treated with medica tions that slow the heart rate (e.g., b-adrenergic receptor antagonists). The initial bradycardia can be diminished by pretreatment with anticholinergic agents (atropine, glycopyrrolate). During the ECT-induced seizure, there are marked increases in heart rate (up to 140 to 160 beats per minute) and blood pressure (often up to 220 to 260/120 to 140 mm Hg). These increases in heart rate and blood pressure can last for 20 minutes or longer following the seizure and are believed to result from the release of catecholamines during the generalized seizure. The increase in heart rate may be enhanced by pretreatment with anticholinergic agents. For this reason, many ECT practitioners determine the need for anticholinergics on a treatment-by-treatment and case-by-case basis. Hemodynamic changes can be dampened by judicious use of b-adrenergic receptor antagonists (esmolol, labetalol) at the time of treatment and by premedication with short-acting calcium channel blockers such as nifedipine. Following the ECT seizure, some patients develop bradycardia, which can last for several minutes. The mechanisms responsible for the postictal bradycardia are uncertain, but the slowing can be overcome by anticholinergic drugs.

Changes in hemodynamic function during ECT render patients with cardiovascular illnesses susceptible to episodes of myocardial ischemia and arrhythmias. Although in most cases, ECT-induced ECG changes are time limited and do not require specific interventions, some patients may develop more-malignant changes that require immediate attention. Thus equipment and medications for cardiac emergencies must be readily available in the ECT treatment room, and the ECT staff must be familiar with appropriate interventions for cardiac emergencies.

Death ECT is a medical procedure that involves general anesthesia. As a consequence it carries a small but definite risk of death. Furthermore, ECT is often used to treat psychiatric disorders in patients with severe medical illnesses. Although definitions of ECT-related deaths vary, the incidence of a death occurring within 24 hours of a treatment appears to be about 1 in every 25,000 treatments. This incidence does not different significantly from the incidence of death in patients exposed to general anesthesia alone. Most often, cause of death is a cardiovascular complication, including myocardial infarction, myocardial rupture, or ventricular arrhythmia. Other causes include brain herniation (particularly in patients with large CNS tumors) and respiratory complications, including aspiration. Given the cardiovascular changes that occur with ECT, it is perhaps surprising that ruptured aneurysms are rare following ECT. Despite risk of death and cardiovascular changes associated with the treatment, ECT has been used as a treatment for more than 50 years and is generally a very safe procedure. Furthermore, untreated severe psychiatric disorders carry significant morbidity and mortality risks from suicide and deterioration in physical health. All of these factors must be weighed when considering the use of ECT in severely ill individuals.

Cognitive Impairment Cognitive adverse effects of ECT show great individual variability. For example, some patients have little recollection of the ECT procedure while others can describe in detail all events up to the time they lose consciousness. The reasons for this variability are not certain.

Most patients experience a period of postictal and postanesthetic confusion that lasts about 30 minutes, although the duration can (rarely) extend to hours. During this time, some patients (roughly 5 percent) may become severely agitated and require restraint and sedation. Preferred agents for this purpose include benzodiazepines (e.g., lorazepam, 1 to 2 mg intravenously, or diazepam, 5 to 10 mg intravenously) or antipsychotic medications (e.g., haloperidol [Haldol], 5–10 mg intramuscularly). Factors that contribute to postictal confusion include frequency and number of ECT treatments, electrical dose, anesthetic agents used, and concomitant medications, including anticholinergic drugs and other CNS-active agents. Age and medical status may also contribute, but severe bouts of postictal confusion can occur in young, healthy persons. Patients who have a documented history of significant difficulties during the recovery period may benefit from prophylactic treatment with a benzodiazepine (e.g., 5 to 10 mg of diazepam intravenously) following termination of the seizure. These patients should also receive intensive nursing intervention in the recovery room.

Memory loss, a major adverse effect of ECT, has both retrograde and anterograde components. Because of the repeated treatments, memory is characteristically worse for events occurring during the ECT course (anterograde amnesia). Most patients also experience retrograde amnesia that is usually worse for events occurring in the weeks prior to treatment. Typically, severity and duration of amnesia diminish as ECT administration becomes more remote. Some patients report difficulties with memory for more-distant events, including specific problems with autobiographical memories. These problems are often confounded by the fact that memory can be impaired by episodes of depression and other treatments used for depression. ECT-induced memory problems usually improve within 6 to 8 weeks following a

course of treatment and coincide with the period during which the EEG shows significant slowing. Some patients report more-sustained difficulties with memory, lasting months, but persistent problems with memory formation are often difficult to demonstrate systematically, and interpretation can be confounded by recurrence of psychiatric symptoms.

As with postictal confusion, several variables contribute to memory impairment, including frequency and number of treatments, electrical charge used to induce seizures, and perhaps the drugs used for anesthesia. Electrode placement is possibly the greatest contributor to ECT-induced memory problems. Systematic studies clearly demonstrate that bilateral treatments are associated with significantly greater verbal memory impairment than nondominant hemisphere unilateral treatments. For this reason alone, unilateral electrode placement is considered the treatment of first choice for most patients referred for ECT.

Memory impairment is an adverse effect of ECT and has nothing to do with clinical benefits of the treatment. A number of attempts to prevent memory deficits have used pharmacological approaches, but none of these has had significant success or impact on clinical practice. Consequently, cognitive impairment remains a major concern in patients treated with ECT.

Brain Damage The concern that ECT may induce structural damage to the CNS is motivated by ECT-associated cognitive impairment. Furthermore, prolonged seizures in animals and humans can produce neuronal injury. Presently no convincing evidence indicates that either ECT or intermittent brief seizures, such as those induced during ECT, produce structural damage to the brain. The brain damage that occurs during status epilepticus is particularly irrelevant because this form of damage takes more than 20 minutes of continuous seizure activity to become manifest and often occurs in the context of hypoxia. ECT seizures rarely exceed 3 minutes and occur under conditions in which patients are well oxygenated. Perhaps the best clinical evidence against ECT-induced CNS damage is derived from magnetic resonance imaging (MRI) studies of the brain performed by C. Edward Coffey and colleagues.

Spontaneous Seizures The repeated induction of generalized seizures during a course of ECT led to concerns that the treatment may increase the risk of spontaneous seizures, perhaps via a kindling effect. The incidence of spontaneous seizures in patients treated with ECT appears to be about 0.1 to 0.5 percent, similar to the incidence of spontaneous seizures in the general population. Some of these patients may have had underlying seizure disorders, but it remains possible that rare patients have a spontaneous seizure following ECT. Whether this becomes a recurrent seizure disorder is less certain.

Other Adverse Effects of ECT Fractures often accompanied treatments in the early days of ECT. With routine use of muscle relaxants, fractures of long bones or vertebrae should not occur. However, some patients may break teeth or experience back pain because of contractions during the procedure. Muscle soreness may occur in some individuals but often results from the effects of muscle depolarization by succinylcholine and is most likely to be particularly troublesome after the first session in a series. This soreness can be treated with mild analgesics, including nonsteroidal anti-inflammatory drugs (NSAIDs). A significant minority of patients experience nausea, vomiting, and headaches following an ECT treatment. Nausea and vomiting can be prevented by treatment with antiemetics at the time of ECT (e.g., metoclopramide [Reglan], 10 mg intravenously, or prochlorperazine [Compazine], 10 mg intravenously; ondansetron [Zofran] is an acceptable alternative if adverse effects preclude use of dopamine receptor antagonists).

ECT can be associated with headaches, although this effect is usually readily manageable. Headaches often respond to NSAIDs given in the ECT recovery period. In patients with severe headaches, pretreatment with ketorolac (Toradol) (30 to 60 mg intravenously), NSAID approved for brief parenteral use, can be helpful. Acetaminophen (Tylenol), tramadol (Ultram), propoxyphene (Darvon) and more potent analgesia provided by opioids can be used individually or in various combinations (e.g., pretreatment with ketorolac and postseizure management with acetaminophen-propoxyphene) to manage more intractable headache. ECT can induce migrainous headache and related symptoms; sumatriptan (Imitrex) (6 mg subcutaneously or 25 mg orally) may be a useful addition to the agents described above. Ergot compounds can exacerbate cardiovascular changes observed during ECT and probably should not be a component of ECT pretreatment.

ECT AND PSYCHOTROPIC MEDICATIONS

The combined use of psychotropic medications and ECT does not appear to enhance the clinical benefits of ECT. However, in some settings, particularly outpatients undergoing maintenance ECT, it is likely that patients will be treated with a combination of medications and ECT. One concern is the risk of adverse interactions between psychotropic medications, adverse effects of ECT, or vice versa. Clear evidence of drug/ECT interactions is lacking for most medications, but it may be prudent to avoid combined use. For example, some evidence suggests that lithium may have adverse effects when combined with ECT, leading to greater confusion and possibly increasing the risk of spontaneous seizures. These findings have been disputed in the literature, but careful consideration of the clinical situation will usually encourage discontinuing lithium use in the period preceding ECT. Additionally, some evidence suggests that lithium may interfere with pseudocholinesterase, the enzyme that degrades succinylcholine; thus, lithium may be associated with prolonged muscle paralysis during ECT. There may be increased risks of combining ECT with irreversible monoamine oxidase inhibitors (MAOIs), particularly problems with blood pressure management and perhaps with pseudocholinesterase inhibition, but this is also controversial. Some studies indicate that benzodiazepines diminish the effectiveness of unilateral ECT, but this may result from effects on seizure threshold that can be corrected by titration of the electrical dose. On theoretical grounds, it may be wise to avoid agents with significant anticholinergic effects (e.g., tricyclic drugs and phenothiazines) because of potential problems with cognitive impairment, but data strongly supporting this contention are lacking. Certain medications (bupropion [Wellbutrin], clozapine [Clozaril], amoxapine [Asendin], and maprotiline [Ludiomil] may prolong ECT seizures, but this is not well documented. Tricyclic drugs and low-potency phenothiazines might be expected to produce similar problems, but no evidence supports this. Problems will other medications are mostly recounted in case reports and anecdotal experiences.

Despite the unclear literature, avoiding the combined use of ECT with psychotropic medications to the extent possible seems most prudent. Patients with marked anxiety may require treatment with low doses of shorter-acting benzodiazepines (e.g., lorazepam, alprazolam [Xanax]) may be necessary. If patients require a sedative-hypnotic, zolpidem (Ambien) or chloral hydrate (Noctec) would be reasonable choices. For severely agitated psychotic patients, high-potency neuroleptics antipsychotics (e.g., haloperidol, thiothixene [Navane]) are useful.

TRANSCRANIAL MAGNETIC STIMULATION

ECT has multiple effects on brain function that are responsible for both its therapeutic and adverse actions. If changes in only certain regions of the CNS are required for therapeutic benefits, it may be possible to develop stimulation paradigms that target these areas. Such treatments could have great advantages in avoiding many of the unwanted effects of ECT, perhaps including cognitive impairment. Transcranial magnetic stimulation (TMS) is one such treatment. In neurology, TMS has been developed as a way to stimulate the CNS noninvasively by application of a focal magnetic field over regions of the cortex. Refinements of magnetic stimulators, including the development of stimulators capable of discharging at frequencies up to 60 Hz (referred to as rapid-rate TMS [rTMS]) have allowed focal stimulation of the CNS to estimate motor thresholds and determine hemispheric language dominance. Interestingly, rTMS was found to benefit some patients with Parkinson's disease, and some Parkinson's disease patients exhibited improved mood following rTMS. Additionally, subjects exposed to rTMS for purposes of determining hemispheric language dominance exhibited affective responses following stimulation of the left frontal cortex.

These observations suggest that rTMS may have therapeutic potential in psychiatry and may allow focal stimulation of areas most involved in affective states. While experience with rTMS in psychiatry is limited, some evidence suggests that depending on the placement of the magnetic coil, rTMS can improve or worsen affective state. In one of the best studies to date, left dorsolateral prefrontal cortex stimulation significantly improved depression ratings in 11 of 17 patients with psychotic major depression. rTMS appears to be well tolerated and does not require general anesthesia. Seizures may be a side effect in some patients but do not appear to be required for therapeutic effects.

TMS is performed using a high-speed magnetic stimulator that generates a 1.5 to 2.5 Tesla field for brief periods. This field is similar to that used for nuclear magnetic resonance imaging. rTMS stimulus is delivered at frequencies of 10 to 60 Hz using a figure-8 shaped coil that is placed over the desired region of the skull and continuously cooled with water to prevent overheating. Patients and staff typically wear earplugs because of the noise generated by the stimulator. Stimulation is typically given several times per session and is repeated over several days to weeks. At present, optimal stimulation patterns for rTMS in psychiatric disorders are not known.

ROLE OF ECT

More than 50 years after it was developed, ECT remains an important, effective, and safe treatment for a variety of neuropsychiatric disorders. Major depression is presently the most common indication for the treatment. Despite the fact that the treatment is usually reserved for patients who fail to respond to psychotropic medications, ECT remains effective in this population. ECT may be a life-saving intervention for some patients. The invasiveness of the procedure and the major adverse effects of memory loss and confusion are limiting variables in the use of ECT. However, cognitive adverse effects of ECT clearly have nothing to do with

therapeutic benefits. Most major innovations in ECT technique over the past 20 years have sought to diminish cognitive effects while maintaining benefits. New developments in ECT technique and perhaps the development of rTMS as a therapeutic tool offer the hope that this form of treatment will find better acceptance among psychiatrists and patients. Further developments in understanding the neurobiology of major psychiatric disorders will only help in this regard.

SUGGESTED CROSS-REFERENCES

Basic electrophysiology is discussed in Section 1.9, and applied electrophysiology in Section 1.14. Biology of memory is covered in Section 3.4 Medical assessment and laboratory testing in psychiatry are covered in Section 7.7. Chapter 12 discusses schizophrenia, and Chapter 13 discusses mood disorders. Section 51.3d covers mood disorders in the elderly, and Section 51.4g covers the use of ECT in older persons. Transcranial magnetic stimulation also is discussed in Section 31.33.

SECTION REFERENCES

*Abrams R: *Electroconvulsive Therapy*, ed 3. Oxford University Press, Oxford, England, 1997.

Abrams R: The mortality rate with ECT. Convuls Ther *13:*125, 1997.

American Psychiatric Association: Practice guideline for major depressive disorder in adults. Am J Psychiatry *1500*(Suppl):4, 1993.

American Psychiatric Association: Practice guideline for the treatment of patients with bipolar disorder. Am J Psychiatry *151*(Suppl):12, 1994.

*American Psychiatric Association Task Force on Electroconvulsive Therapy: *The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging: a Task Force Report of the American Psychiatric Association*. American Psychiatric Association, Washington, DC, 1990.

Applegate RJ: Diagnosis and management of ischemic heart disease in the patient scheduled to undergo electroconvulsive therapy. Convuls Ther *13:*128, 1997.

Bertagnoli MS, Borchardt CM: A review of ECT for children and adolescents. J Am Acad Child Adolesc Psychiatry *29:*302, 1990.

Beyer JL, Weiner RD, Glenn MD: *Electroconvulsive Therapy: A Programmed Text*, ed 2. American Psychiatric Press, Washington, DC, 1998.

Brandon S, Cowley P, McDonald C, Neville P, Palmer R, Wellstood-Eason S: Electroconvulsive therapy: Results in depressive illness from the Leicestershire trial. Br Med J *288:*22, 1984.

Coffey CD, Weiner RD, Djang WT, Figiel GS, Soady SAR, Patterson LJ, Holt PD, Spritzer CE, Wilkinson WE: Brain anatomic effects of electroconvulsive therapy: A prospective magnetic resonance imaging study. *48:*1013, 1991.

Devanand DP, Dwork AJ, Hutchinson ER, Bolwig TG, Sackeim HA: Does ECT alter brain structure? Am J Psychiatry *151:*957, 1994.

Freeman CPL, Basson JV, Crighton A: Double-blind controlled trial of electroconvulsive therapy (ECT) and simulated ECT in depressive illness. Lancet *I:*738, 1978.

*George MS, Lisanby SH, Sackeim HA: Transcranial magnetic stimulation. Arch Gen Psychiatry *56:*300, 1999.

Gregory S, Shawcross CR, Gill D: The Nottingham ECT study: A double-blind comparison of bilateral, unilateral and simulated ECT in depressive illness. Br J Psychiatry *146:*520, 1985.

Grisaru N, Chudakov B, Yaroslavsky Y, Belmaker RH: Transcranial magnetic stimulation in mania: A controlled study. Am J Psychiatry *155* :1608, 1998.

*Helsley S, Sheikh T, Kim KY, Park SK: ECT therapy in PTSD. Am J Psychiatry *156:*94, 1999.

Johnstone EC, Deakin JFW, Lawler P, Frith CD, Stevens M, McPherson K, Crow TJ: The Northwick Park electroconvulsive therapy trial. Lancet *ii:*1317, 1980.

Kellner CH, Pritchett JT, Beale MD, Coffey CE: *Handbook of ECT*. American Psychiatric Press, Washington, DC, 1997.

*Klein E, Kreinin I, Chistyakov A, Koren D, Mecz L, Marmur S, Ben-Shachar D, Feinsod M: Therapeutic efficacy of right prefrontal slow repetitive transcranial magnetic stimulation in major depression: A double-blind controlled study. Arch Gen Psychiatry *56:*315, 1999.

Lambourn J, Gill D: A controlled comparison of simulated and real ECT. Br J Psychiatry *133:*514, 1978.

Miller LJ: Use of electroconvulsive therapy during pregnancy. Hosp Community Psychiatry *45:*444, 1994.

Moise FN, Petrides G: Case study: Electroconvulsive therapy in adolescents. J Am Acad Child Adolesc Psychiatry *35:*312, 1996.

Mukherjee S: Mechanisms of the antimanic effect of electroconvulsive therapy. Convulsive Ther *5:*227, 1989.

*Mukherjee S, Sackeim HA, Schnur DB: Electroconvulsive therapy of acute manic episodes: A review of 50 years' experience. Am J Psychiatry *151:*169, 1994.

Olfson M, Marcus S, Sackeim HA, Thompson J, Pincus HA: Use of ECT for the inpatient treatment of recurrent major depression. Am J Psychiatry *155:*22, 1998.

Pascual-Leone A, Rubio B, Pallardó F, Catalá MD: Rapid rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. Lancet *348:*233, 1996.

Prudic J, Haskett RF, Mulsant B, Malone KM, Pettinatti HM, Stephens S, Greenberg R, Rifas SL, Sackeim HA: Resistance to antidepressant medications and short-term clinical response to ECT. Am J Psychiatry *153:*985, 1996.

*Prudic J, Sackeim HA: Electroconvulsive therapy and suicide risk. J Clin Psychiatry *60:*104, 1999.

Rey JM, Walter G: Half a century of ECT use in young people. Am J Psychiatry *154:*595, 1997.

Rice EH, Sombrotto LB, Markowitz JC, Leon AC: Cardiovascular morbidity in high-risk patients during ECT. Am J Psychiatry *151:*1637, 1994.

*Sackeim HA, Long J, Luber B, Moeller JR, Prohovnik I, Devanand DP, Nobler MS: Physical properties and quantifications of the ECT stimulus: I Basic principles. Convulsive Ther *10:*93, 1994.

Sackeim HA, Prudic J, Devanand DP, Decina P, Kerr B, Malitz S: The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. J Clin Psychopharm *10:*96, 1990.

*Sackeim HA, Prudic J, Devanand DP, Kiersky JE, Fitzsimmons L, Mood BJ, McElhiney MC, Coleman EA, Settembrino JM: Effects of stimulus intensity and electrode placement on the efficacy and cognitive effects of electroconvulsive therapy. N Engl J Med *328:*839, 1993.

Salzman C: The use of ECT in the treatment of schizophrenia. Am J Psychiatry *137:*1032, 1980.

Schwarz T, Loewenstein J, Isenberg KE: Maintenance ECT: Indications and outcome. Convulsive Ther *11:*14, 1995.

Small JG, Klopper MH, Kellams JJ, Miller MJ, Milstein V, Sharpley PH, Small IF: Electroconvulsive treatment compared with lithium in the management of manic states. Arch Gen Psychiatry *45:*727, 1988.

Weiner RD: Does electroconvulsive therapy cause brain damage? Behav Brain Sci *7:*1, 1984.

West ED: Electric convulsion therapy in depression, a double-blind controlled trial. Br Med J *282:*355, 1981.

Zielinski RJ, Roose SP, Devanand DP, Woodring S, Sackeim HA: Cardiovascular complications of ECT in depressed patients with cardiac disease. Am J Psychiatry *150:*904, 1993.

Zorumski CF, Rubin EH, Burke WJ: Electroconvulsive therapy for the elderly: A review. Hosp Community Psychiatry *39:*643, 1988.

Zorumski CF, Rutherford JL, Burke WJ, Reich T: ECT in primary and secondary depression. J Clin Psychiatry *47:*298, 1986.

31.32 NEUROSURGICAL TREATMENTS

SCOTT L. RAUCH, M.D. AND G. REES COSGROVE, M.D., F.R.C.S.(C)

Neurosurgical treatment of psychiatric disease has a long and checkered history. Crude freehand “psychosurgery,” such as prefrontal lobotomy, was widely and indiscriminately used to treat severe psychiatric disease during an era that predated modern psycho-pharmacology. Those primitive operations yielded modest reductions in symptoms but were accompanied by unacceptable adverse effects. Unfortunately, many people still associate neurosurgical treatment of psychiatric disease with the techniques and practices of that bygone era. Over the past 35 years neurosurgical techniques have evolved that differ substantially from early psychosurgical interventions. First, accurate, precise, and reproducible placement of lesions in specific target areas is performed using advanced magnetic resonance imaging (MRI)-guided stereotactic methods. Second, strict criteria for patient selection are observed, and the process of determining appropriate candidacy has been formalized. Currently, surgical intervention is predominantly reserved for patients with severe, incapacitating major depressive disorder or obsessive-compulsive disorder who have failed an exhaustive array of other available treatments. Surgery is not approved unless a multidisciplinary committee reaches consensus regarding its appropriateness for a given candidate and the patient renders informed consent. Third, although a large body of clinical data indicates the efficacy and safety of modern neurosurgical interventions, major centers providing these treatments continue to gather information prospectively. In the context of smaller, more precisely placed lesions and better-established selection criteria, modern psychiatric neurosurgery yields substantial improvement in approximately 25 to 70 percent of cases, with drastically reduced morbidity and mortality.

HISTORY

Trephination, as performed by ancient civilizations, probably represents the earliest form of surgical intervention for psychiatric disease. In 1891 the first formal report of neurosurgical treatment in psychiatry was published, describing bilateral cortical excisions in demented and depressed patients, which yielded mixed results. During the ensuing four decades, little progress was made in this arena. In 1935 John Fulton and Charles Jacobsen presented their research on primate behavior following frontal cortical ablation. They observed that lobectomized chimpanzees showed reduced “experimental neurosis” and were less fearful, while retaining an ability to perform complex tasks. Egaz Moniz (Fig. 31.32-1), a renowned Portuguese neurologist, was in attendance at Fulton's lecture. Prompted by these findings in primates, Moniz pioneered prefrontal leukotomy in collaboration with his neurosurgical colleague Almeida Lima. First by using pure alcohol injections and subsequently by mechanical means with a leucotome, Moniz and Lima performed “psychosurgery” on 20 severely ill institutionalized patients; 14 were said to have exhibited worthwhile improvement. In an era of overflowing asylums and few effective treatments for chronic debilitating psychiatric illness, this mode of therapy was initially enthusiastically embraced. In fact, Moniz was awarded the 1949 Nobel Prize in Medicine and Physiology for this contribution. From the mid-1930s until the emergence of the phenothiazines in the mid-1950s, there was a global proliferation of psychosurgical techniques and practices. Walter Freeman was perhaps the most zealous promoter of psychosurgery in the United States. Pioneering a series of freehand procedures to achieve prefrontal lobotomy (i.e., severing the white matter connections between the prefrontal cortex and the rest of the brain), Freeman and neurosurgeon James Watts reported on their first 200 cases by 1942. Although the benefits of the surgery were highlighted, others acknowledged a significant complication rate, including frontal lobe syndrome, seizures, and even deaths. At its peak, this brand of psychosurgery was being performed on approximately 5000 patients per year in the United States alone. A review of the results of 10,365 prefrontal lobotomies performed from 1942 to 1954 in Britain concluded that while 70 percent showed improvement, adverse events included 6 percent mortality, plus 1 percent with seizures and 1.5 percent with disinhibition syndromes. Widespread reports of blunted personality and socially inappropriate behavior also surfaced. In the late 1940s and early 1950s, recognition of these risks prompted attempts to develop modified surgical procedures that might yield better results. However, with the introduction of chlorpromazine (Thorazine) in 1954, satisfactory medical management of psychiatric illness became possible for the first time. Thus, despite the advent of stereotactic neurosurgical techniques in the 1950s and a continued high prevalence of severe treatment-refractory psychiatric disease, psychosurgery was all but abandoned in favor of nonsurgical therapies. Later, as the limitations of antipsychotic drugs and electroconvulsive treatment became known, there was a resurgence of interest in psychiatric neurosurgery at a few academic medical centers in the 1960s. Hence, refinement of stereotactic methods and emerging theories regarding the role of the limbic system in emotion shaped a new age of neurosurgical treatments for psychiatric disease. The pioneers of these contemporary procedures faced harsh opposition from political groups, as well as the populace, who equated psychosurgery with crude and indiscriminately applied operations of an earlier era. Political and social pressure dissuaded all but a handful of neurosurgeons from offering these treatments in the 1960s and 1970s. Gradually, over the past two decades, research and education regarding modern neurosurgery has led to its legitimacy as an acceptable treatment modality for severe, intractable psychiatric disease—now in the context of more-refined procedures as well as better-defined indications, risks, and potential benefits.

FIGURE 31.32-1 Egas Moniz. (Courtesy of New York Academy of Medicine.)

SURGICAL APPROACHES

Although numerous approaches have been tried, four neurosurgical procedures have evolved as the safest and most effective for treating psychiatric disorders. All four entail bilateral lesions and are currently performed using modern stereotactic methods (Fig. 31.32-2).

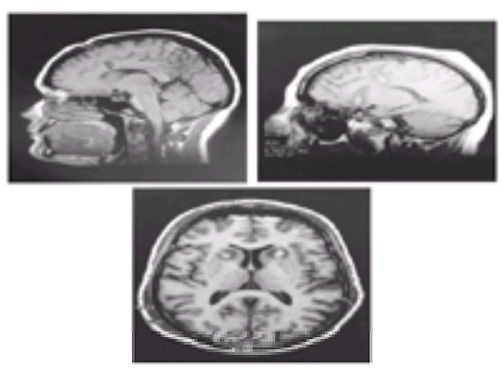

FIGURE 31.32-2 Appearance of psychiatric neurosurgical lesions by postoperative T_1-weighted MRI. **A**, Sagittal view depicting an anterior cingulotomy lesion. **B**, Sagittal view depicting the lesions of limbic leukotomy (i.e., anterior cingulotomy plus subcaudate tractotomy). **C**, Transaxial view depicting anterior capsulotomy lesions.

Subcaudate Tractotomy Subcaudate tractotomy was introduced by Geoffrey Knight in Great Britain in 1964 as one of the first attempts to limit adverse effects by restricting lesion size. By targeting the substantia innominata (just inferior to the head of the caudate nucleus), the goal was to interrupt white matter tracts connecting orbitofrontal cortex and subcortical structures. The surgery involved placement of radioactive yttrium-90 seeds at the desired centroid, yielding lesion volumes of approximately 2 cubic centimeters on each side. Indications for subcaudate tractotomy are major depressive disorder, obsessive-compulsive disorder, and other severe anxiety disorders.

Anterior Cingulotomy Fulton was the first to propose that the anterior cingulum might represent an appropriate target for treatment of psychiatric disease. While the first cingulotomies were performed for intractable pain, it was noted that the best results were in patients with comorbid anxiety or depressive conditions. H. Thomas Ballantine, Jr., and colleagues subsequently demonstrated the safety of anterior cingulotomy and studied its efficacy for a broad range of psychiatric indications. Since 1962 the group at Massachusetts General Hospital has performed nearly 1000 of these procedures, and it remains the most commonly used neurosurgical treatment for psychiatric disease in North America. The surgery is conducted under local anesthesia and two or three approximately 1.0 cubic centimeter lesions are made on each side by thermocoagulation through bilateral burr holes. The target is within the anterior cingulate cortex (Brodmann areas 24 and 32), at the margin of the white matter bundle known as the cingulum. Originally, lesion placement was determined by ventriculography; however, since 1991 anterior cingulotomy has been conducted via MRI-guidance. This group has always emphasized making the smallest possible lesion; consequently, up to 40 percent of patients return several months following the first operation for a second procedure to extend the first set of lesions. The indications for anterior cingulotomy include major depressive disorder, and obsessive-compulsive disorder.

Limbic Leukotomy Limbic leukotomy was introduced by Desmond Kelly and colleagues in England in 1973. The procedure combines the targets of subcaudate tractotomy and anterior cingulotomy. The lesions have typically been made via thermocoagulation or with a cryoprobe. Historically, the precise placement of the lesions was guided by intraoperative stimulation; pronounced autonomic responses were believed to designate the optimal lesion site. The indications for limbic leukotomy include major depression, obsessive-compulsive disorder, and other severe anxiety disorders.

Anterior Capsulotomy Anterior capsulotomy was first performed in France; however, Lars Leksell and colleagues popularized the procedure in Sweden. Anterior capsulotomy remains the most commonly used surgical treatment for psychiatric disease in Scandinavia. The procedure involves small lesions placed within the anterior one-third of the anterior limb of the internal capsule at the level of the intercommissural plane, thereby interrupting fibers of passage from prefrontal cortex to subcortical nuclei. Anterior capsulotomy was originally performed via thermocoagulation; more recently the lesions have been made noninvasively using a radiosurgical device known as the *gamma knife*. The relative advantages and disadvantages of this stereotactic radiosurgical approach remain undetermined. Indications for anterior capsulotomy include major depressive disorder, obsessive-compulsive disorder, and other severe anxiety disorders.

PATIENT SELECTION

Indications and Contraindications Although limited reports have suggested efficacy across a broad range of psychiatric conditions, the best-established indications for contemporary psychiatric neurosurgery are major depressive disorder and obsessive-compulsive disorder. In evaluating candidates several factors are considered:

1. **Primary diagnosis.** The patient must meet clinical criteria for the diagnostic indication, and this disorder should be a primary cause of the patient's illness.
2. **Severity.** The patient must have chronic, severe, and debilitating illness; duration of the primary illness must exceed 1 year and typically exceeds 5 years. Severity is gauged on standardized instruments (e.g., patients with obsessive-compulsive disorder typically have Yale-Brown Obsessive-Compulsive Scale scores of 25 or above; patients with depression typically have Beck Depression Inventory scores of 30 or above), while debility should be indicated by a low level of functioning (e.g., a Global Assessment of Functioning score £ 50) and a poor quality of life.
3. **Adequacy of previous treatment.** Patients must have already undergone an exhaustive array of other available treatments.
4. **Psychiatric comorbidity.** Appropriate treatment must have been rendered for any comorbid psychiatric disorder; the presence of psychoactive substance use or personality disorders are considered strong relative contraindications.
5. **Medical comorbidity and surgical fitness.** Structural brain lesions or significant central nervous system injuries are strong contraindications. Medical conditions that increase neurosurgical risks (e.g., cardiopulmonary disease) and age above 65 years are relative contraindications. A history of past seizures is a risk factor for perioperative seizures and must be weighed in the overall risk-benefit assessment.
6. **Access to postoperative care.** Patients must have access to adequate postoperative care, including a psychiatrist (typically the referring physician) who will accept responsibility for managing them after discharge. Arrangements for postoperative care (e.g., intensive behavior therapy) should be confirmed ahead of time.
7. **Informed consent.** Under no circumstances should psychiatric neurosurgery be performed on patients against their will. The patient must be able and willing to render an informed consent. In rare instances these procedures are performed with assent of the patient and formal consent from a legal guardian. In this context, age less than 18 years also represents a strong contraindication.

Multidisciplinary Case Review The process of assessing candidacy differs somewhat from center to center and between nations as governed by law. Across centers, however, there is consensus that decisions regarding candidacy must be made after thorough review of the patient's case, taking into account the written record, current clinical examination, and formal referral from the patient's treating psychiatrist. In Boston, Stockholm, and London, the decision is typically made by an interdisciplinary committee; in Britain, psychosurgery requires additional formal approval from the Mental Health Act Commission.

PREOPERATIVE WORKUP

In addition to a multidisciplinary clinical evaluation and formalized case review, the preoperative workup for psychiatric neurosurgery includes standard blood and urine laboratory tests, an electrocardiogram, and specialized tests of brain structure and function. A comprehensive preoperative assessment should include brain MRI, an electroencephalogram, and psychometric testing.

POSTOPERATIVE CARE

Immediate postoperative care includes the standard medical and surgical considerations following any stereotactic neurosurgical procedure. Special attention is paid to signs or symptoms of potential surgical complications, including infection, hemorrhage, seizures, or altered mental status. A postoperative MRI should be obtained to document the placement and extent of lesions. Intensive postoperative psychiatric treatment is recommended, since the efficacy of the surgery may rely on some synergy between the neurosurgical intervention itself and enhanced response to pharmacological or behavioral therapies. Although dosages of psychotropic medications may be reduced during the immediate perioperative period, the medication regimen should be readjusted as tolerated postoperatively. In the case of obsessive-compulsive disorder, intensive behavior therapy should be initiated as soon as possible, preferably within the first month postoperatively.

DESIGN AND INTERPRETATION OF CLINICAL STUDIES

One must appreciate the inherent challenges surrounding research in this domain. First, the population in question is very heterogeneous, has high rates of comorbidity, and is relatively small (the major psychiatric neurosurgical centers now treat 10 to 30 new patients per year). These factors limit statistical power and complicate the quantification of efficacy and adverse-effect rates. Second, the major academic centers each have their own surgery of preference and use somewhat different measures to assess treatment response. Consequently, it is difficult to contrast or combine data across these different groups. Third, determining the cause of postoperative improvement is often confounded by exposure to other new therapies as they emerge. Fourth, the surgical equivalent of placebo-controlled trials has not been possible, because the inherent risks associated with surgical incisions and craniotomy make sham neurosurgical procedures unethical. With the advent of new radiosurgical techniques, sham-controlled studies of neurosurgical treatments may now be feasible.

The initial literature on modern psychiatric neurosurgery suffered several additional limitations. Some of the first case series entailed retrospective designs and did not include rigorous clinical evaluations with valid or reliable standardized instruments for assessing diagnosis and outcome. More recently, the major centers have instituted protocols under which high-quality data are gathered prospectively. Further progress is still needed regarding blind evaluation, identification of optimal controls, and standardization of dependent measures across groups. Nonetheless, the best available data indicate that contemporary psychiatric neurosurgery is effective in 25 to 70 percent of patients, with a modest rate of adverse effects. Current data do not, however, provide compelling evidence that any one of the four modern operations outlined is clearly superior to the others.

TREATMENT OUTCOME

For all four contemporary procedures, outcome cannot be fairly assessed in less than 3 to 6 postoperative months. Most clinical reports have used the Pippard Postoperative Rating Scale or a similar measure of global improvement. The Pippard Scale rates outcomes as follows: (1) symptom free, (2) much improved, (3) slightly improved, (4) unchanged, and (5) worse. Most studies consider categories 1 and 2 as significant improvement. Since most studies focus on one or another of the procedures, they are best reviewed according to surgical approach.

Outcome With Subcaudate Tractotomy In 1975 investigators reviewed results of subcaudate tractotomy in 208 patients with depression and obsessive-compulsive disorder over a mean 2.5-year follow-up period. Optimal data, including structured interviews, were available from a subset of 134 patients. Significant improvement was seen in 68 percent of patients with major depressive disorder, 50 percent of patients with obsessive-compulsive disorder, and 62.5 percent of patients with other anxiety disorders. Patients with schizophrenia, substance abuse, or personality disorders did poorly. Short-term adverse effects include transient headache and confusion or somnolence, which typically resolve in less than 1 week. Patients are usually ambulatory by the third postoperative day. Adverse effects include mild undesirable personality changes in 6.7 percent and seizures in 2.2 percent. Transient disinhibition syndromes were common. One patient died because of migration of an yttrium seed, and three others committed suicide. In 1994 a large-scale review of 1300 cases concluded that the procedure enables 40 to 60 percent of patients to lead normal or near normal lives, with a reduction in suicide rate to 1 percent versus 15 percent in a similarly affected control group with major mood disorders. There were no additional deaths in this series, and although the seizure rate was comparable to that in previous reports, the 1994 review did not find enduring personality changes. In addition, the investigators proposed predictors of positive response, including major depressive disorder, sudden onset of symptoms, emergence in the second half of life or peripartum, positive family history, and previous positive response to electroconvulsive therapy.

Outcome With Anterior Cingulotomy Ballantine and coworkers retrospectively reviewed 198 cases with mean follow-up of 8.6 years. They noted significant improvement in 62 percent of patients with affective disorders, 56 percent with obsessive-compulsive disorder, and 79 percent with other anxiety disorders. In 1995 a prospective report on 18 patients with obsessive-compulsive disorder was published. The investigators, using more-stringent criteria, found that 28 to 44 percent had responded favorably, with no serious long-term adverse effects. A subsequent report reviewed a series of 34 patients who underwent MRI-guided cingulotomy. Among patients with major depressive disorder, 60 percent responded favorably; among patients with bipolar I disorder, 40 percent responded favorably; and among patients with obsessive-compulsive disorder, 27 percent were classified as responders with another 27 percent categorized as possible responders. Short-term adverse effects include headache, nausea, and difficulty with urination; however, these typically resolve within 3 days. Patients are usually ambulatory within 12 hours following the operation and discharged on the third to fifth postoperative day. In the Massachusetts General Hospital experience of nearly 1000 anterior cingulotomies there have been no deaths and no infections; since the advent of MRI guidance, there have been no cases of stroke or hemorrhage. The incidence of seizure is from 1 to 5 percent—most often occurring in patients with a preexisting seizure history. Over the past 5 years the practice of treating patients who experience perioperative seizures with long-term anticonvulsant therapy has been discontinued, and no cases of multiple recurrent seizures have been seen. An independent analysis of 34 patients demonstrated no significant intellectual or behavioral impairments attributable to anterior cingulotomy; a subsequent study of 57 patients likewise found no evidence for lasting neurological or behavioral adverse effects.

Outcome With Limbic Leukotomy Kelly's group has published several reports, including an initial series of 66 patients studied prospectively with a mean follow-up of 16 months. They found significant improvement in 89 percent of patients with obsessive-compulsive disorder, 78 percent with major depressive disorder, and 66 percent with other anxiety disorders. In a separate report regarding 49 patients with obsessive-compulsive disorder, they noted that 84 percent were improved at 20 months follow-up. Short-term adverse effects include headache, lethargy, confusion, and lack of sphincter control, which may last from a few days to a few weeks. In particular, postoperative confusion commonly lasts at least several days, and patients are rarely discharged in less than 1 week. There were no seizures and no deaths; however, one patient suffered severe memory loss because of improper lesion placement, and enduring lethargy was present in 12 percent of patients.

Outcome With Anterior Capsulotomy Leksell and colleagues reported on their first 116 cases, noting a favorable response in 50 percent of those with obsessive-compulsive disorder and 48 percent of those with major depression. Other investigators conducted a prospective study of 35 patients with obsessive-compulsive disorder and observed that 70 percent had satisfactory outcomes. One review of the available literature on capsulotomy uncovered adequate data on 213 of 362 cases; of these, 64 percent were deemed to have shown significant benefit. Short-term adverse effects can include transient headache or incontinence. Postoperative confusion often lasts for up to 1 week. Recovery from gamma capsulotomy may be swifter and cause less discomfort, but adverse effects from radiation necrosis may be delayed for up to 8 to 12 months. Although patients are typically ambulatory in a matter of hours to days following the operation, length of hospital stay may be influenced by the duration of confusion. Investigators followed a cohort of 24 patients prospectively and observed one intracranial hemorrhage and one episode of seizure. Fatigue was present in 29 percent, 17 percent had poor memory, and 8 percent exhibited slovenliness. Weight gain is a common, enduring adverse effect, with a mean increase in mass of 10 percent. In other reports, up to 200 capsulotomy patients have been assessed via a variety of psychometric tests with no evidence of significant cognitive dysfunction or adverse personality changes. Although it can be difficult to discern from the published literature, apparent discrepancies in terms of efficacy and adverse effects in part reflect disparities in lesion size across different study cohorts. In some case series with larger capsulotomy lesions, reductions in symptoms were greater but came at the expense of more common and more-severe adverse effects; conversely, series involving smaller lesions yielded more-modest efficacy with a superior adverse effect profile.

Outcome Across Contemporary Neurosurgical Procedures Carefully selected psychiatric patients should significantly benefit, from contemporary neurosurgical treatment in 25 to 70 percent of cases, with at least 25 percent showing outstanding improvement. Response rates are marginally better for major affective disorders than for obsessive-compulsive disorder. While minor short-term adverse effects are common, severe or enduring adverse consequences are rare but include seizures in approximately 1 to 3 percent of cases. Although frontal syndromes, confusion, or subtle cognitive deficits can still be seen, they are influenced by lesion size as well as choice of surgical approach, and they are minimal in comparison with past procedures. Notably, overall cognitive function, as indicated by the standard intelligence quotient, is generally enhanced, which is attributed to the overriding beneficial effects of symptomatic improvement. Psychiatric neurosurgery likely reduces mortality as evidenced by data on comparative suicide rates. Nonetheless, patients who undergo and fail to benefit from these procedures are at particularly high risk for completed suicide. Therefore, as with any therapy, the potential risks and benefits of psychiatric neurosurgery must be weighed against the potential risks and benefits of foregoing this brand of treatment. Future research should help to identify predictors of response, enable further refinement of the procedures, and elucidate the mechanisms by which these surgeries produce their observed effects.

SUGGESTED CROSS-REFERENCES

Depressive disorders are discussed in Chapter 14 on mood disorders. Obsessive-compulsive disorder is discussed in Chapter 15 on anxiety disorders. The history of psychiatry is covered in Section 55.1.

SECTION REFERENCES

Baer L, Rauch SL, Ballantine HT, Martuza R, Cosgrove R, Cassem E, Giriunas I, Manzo P, Jenike MA: Cingulotomy for intractable obsessive-compulsive disorder: Prospective long-term follow-up of 18 patients. Arch Gen Psychiatry *52:*384, 1995.

Ballantine HT: Historical overview of psychosurgery and its problematic. Acta Neurochir *44:*125, 1988.

Ballantine HT, Bouckoms AJ, Thomas EK, Giriunas IE: Treatment of psychiatric illness by stereotactic cingulotomy. Biol Psychiatry *22:*807, 1987.

Bartlett JR, Bridges PK: The extended subcaudate tractotomy lesion. In *Neurosurgical Treatment in Psychiatry, Pain and Epilepsy*, WH Sweet, S Obrador, JG Martin-Rodriguez, editors. University Park Press, Baltimore, 1977.

Bingley T, Leksell L, Meyerson BA, Rylander G: Long-term results of stereotactic capsulotomy in chronic obsessive compulsive neurosis. In *Neurosurgical Treatment in Psychiatry, Pain and Epilepsy*, WH Sweet, S Obrador, JG Martin-Rodriguez, editors. University Park Press, Baltimore, 1977.

*Bridges PK, Bartlett JR, Hale AS, Poynton AM, Malizia AL, Hodgkiss AD: Psychosurgery: Stereotactic subcaudate tractotomy—an indispensible procedure. Br J Psychiatry *165:*599, 1994.

Corkin S, Twitchell TE, Sullivan EV: Safety and efficacy of cingulotomy for pain and psychiatric disorders. In *Modern Concepts in Psychiatric Surgery*, ER Hitchcock, HT Ballantine, BA Myerson, editors. Elsevier, Amsterdam, 1979.

*Cosgrove GR, Rauch SL: Psychosurgery. Neurosurg Clin North Am *6:*167, 1995.

Fodstad H, Strandman E, Karlsson B, West KA: Treatment of obsessive-compulsive states with stereotactic anterior capsulotomy or cingulotomy. Acta Neurochir *62:*1, 1982.

Freeman W, Watts JW: Prefrontal lobotomy in the treatment of mental disorders. South Med J *30:*23, 1937.

Freeman W, Watts JW: Psychosurgery: Intelligence, emotion and social behavior following prefrontal lobotomy for mental disorders. Thomas, Springfield, IL, 1942.

Goktepe EO, Young LB, Bridges PK: A further review of the results of stereotactic subcaudate tractotomy. Br J Psychiatry *126:*270, 1975.

Kelly D, Richardson A, Mitchell-Heggs N: Stereotactic limbic leucotomy: Neurophysiologic aspects and operative technique. Br J Psychiatry *123:*133, 1973.

Knight GC: Bifrontal stereotactic tractotomy: An atraumatic operation of value in the treatment of psychoneurosis. Br J Psychiatry *115:*257, 1969.

Kullberg G: Differences in effect of capsulotomy and cingulotomy. In *Neurosurgical Treatment in Psychiatry, Pain and Epilepsy*, WH Sweet, S Obrador, JG Martin-Rodriguez, editors. University Park Press, Baltimore, 1977.

Mayberg HS: Limbic-cortical dysregulation: A proposed model of depression. J Neuropsychiatry *9:*471, 1997.

Mindus P, Nyman H: Normalization of personality characteristics in patients with incapacitating anxiety disorders after capsulotomy. Acta Psychiatr Scand *83:*283, 1991.

*Mindus P, Rauch SL, Nyman H, Baer L, Edman G, Jenike MA: Capsulotomy and cingulotomy as treatments for malignant obsessive-compulsive disorder: An update. In *Current Insights in Obsessive-Compulsive Disorder*, B Berend, E Hollander, D Marazitti, J Zohar, editors. Wiley, Chichester, 1994.

*Mitchell-Heggs N, Kelly D, Richardson A: Stereotactic limbic leucotomy—a follow-up at 16 months. Br J Psychiatry *128:*226, 1976.

Moniz E: Prefrontal leucotomy in the treatment of mental disorders. Am J Psychiatry *93:*1379, 1937.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research: Report and recommendations: Psychosurgery. Department of Health and Human Services, publ no. (OS) 77-002. U.S. Government Printing Office, Washington, DC, 1979.

Pippard J: Rostral leucotomy: A report on 240 cases personally followed up after one and one-half to five years. J Ment Sci *101:*756, 1955.

Rauch SL, Baer L, Cosgrove GR, Jenike MA: Neurosurgical treatment of Tourette's disorder: A critical review. Compr Psychiatry *36:*141, 1995.

Rauch SL, Whalen PJ, Dougherty DD, Jenike MA: Neurobiological models of obsessive compulsive disorders. In *Obsessive-Compulsive Disorders: Theory and Management*, MA Jenike, L Baer, WE Minichiello, editors. Mosby, Philadelphia, 1998.

Rylander G: Stereotactic radiosurgery in anxiety and obsessive-compulsive states: Psychiatric aspects. In *Modern Concepts in Psychiatric Surgery*, ER Hitchcock, HT Ballantine, BA Myerson, editors. Elsevier, Amsterdam, 1979.

*Spangler WJ, Cosgrove GR, Ballantine HT, Cassem EH, Rauch SL, Nierenberg A, Price BH: MRI-guided stereotactic cingulotomy for intractable psychiatric disease. Neurosurgery *38:*1071, 1996.

Swayze VW: Frontal leukotomy and related psychosurgical procedures in the era before antipsychotics (1935–1954): A historical overview. Am J Psychiatry *152:*505, 1995.

Tooth JC, Newton MP: Leucotomy in England and Wales 1942–1954. Reports on public health and medical subjects no. 104. Her Majesty's Stationary Office, London, 1961.

Valenstein ES: *Great and Desperate Cures. The Rise and Fall of Psychosurgery and Other Radical Treatments for Mental Illness*. Basic Books, New York, 1986.

*Weingarten SM: Psychosurgery. In *The Human Frontal Lobes: Functions and Disorders*, BL Miller, JL Cummings, editors. Guilford, New York, 1999.

31.33 OTHER BIOLOGICAL AND PHARMACOLOGICAL THERAPIES

CHARLES DEBATTISTA, M.D., D.M.H. AND ALAN F. SCHATZBERG, M.D.

Biological therapies for the treatment of mental disorders have been available since the dawn of civilization. Herbs, potions, and other treatments for emotional disturbance date back thousands of years. In the twentieth century, many new pharmacological and other biological therapies have been developed to treat psychiatric disorders. Some treatments such as insulin coma therapy have not survived. It remains to be seen whether newer interventions, such as rapid transcranial magnetic stimulation, will become important tools in modern psychiatric practice or be quickly dispatched to the vaults of history.

RAPID TRANSCRANIAL MAGNETIC STIMULATION

Rapid transcranial magnetic stimulation (rTMS) is a noninvasive technique for stimulating cells of the cerebral cortex. rTMS uses a magnet to allow focused electrical stimulation across the scalp and cranium without the pain associated with percutaneous electrical stimulation. rTMS was originally used to map cortical motor control and hemisphere dominance. Stimulating the motor cortex with rTMS results in a contralateral motor response. Likewise, stimulating Broca's area with rTMS has resulted in speech blockage. Currently, the potential use of rTMS for the treatment of neurological and psychiatric disorders is being explored actively.

In rTMS a powerful electrical current is passed through a small coil applied to the scalp. This current generates a focused magnetic field of 1.5 to 2 teslas that passes through the scalp and is largely unimpeded by bone or tissue. The magnetic field in turn depolarizes brain cells to a depth of 2 cm from the coil. Cortical interneurons are more likely to be stimulated than are cortical output cells because the interneurons tend to lie parallel to the brain surface. rTMS uses magnetic stimulators with multiple capacitors capable of generating very rapid pulses up to 60 Hz. Low-frequency pulses in the range of 1 Hz may have an inhibitory effect on cortical cells, while higher frequencies have an excitatory effect.

Some psychiatric conditions, such as major depression, may be characterized by hypoactive cortical areas. Functional imaging, including positron emission tomography (PET), has revealed a relative hypofrontality in some patients with major depression. It has been proposed that rTMS stimulation of these frontal areas would relieve symptoms.

rTMS is an experimental technique, and it currently has no approved psychiatric indications. At this time, the most studied psychiatric disorder in rTMS research is major depression. Several open-label and controlled studies have suggested that rTMS may be at least temporarily effective in both animal models of depression and patients with major depression. For example, in the forced swim test, in which rats quickly learn to give up swimming after repeated failures at escaping, they will swim longer after exposure to rTMS over the whole cortex. Other animal models of depression, such as apomorphine-induced stereotypy, are also effectively altered with rTMS.

A number of small, open-labeled studies have suggested that rTMS may be effective in some patients with treatment-resistant major depressive disorder as well as in those with milder major depressive disorder. More recently some small, randomized, placebo-controlled (sham) rTMS studies have suggested that the procedure may also be effective in psychotic depression. In addition, recent functional imaging studies have suggested that the baseline hypofrontality associated with major depressive disorder can be reversed with active treatment but not with a sham rTMS procedure. In addition to major depressive disorder, rTMS has shown some preliminary efficacy in obsessive-compulsive disorder and in posttraumatic stress disorder.

The application of rTMS to psychiatric conditions has lagged behind its neurological applications. rTMS has been used to map the motor cortex, help determine hemispheric dominance, and probe short-term memory. In some symptoms of Parkinson's disease, including bradykinesia, diminished reaction time has improved transiently with rTMS. Finally, rTMS has been used to help elucidate the pathophysiology of migraine headache, and some patients have had temporary symptom relief with rTMS.

PHOTOTHERAPY

Phototherapy (light therapy) was introduced in 1984 as a treatment for seasonal affective disorder (mood disorder with seasonal pattern). In this disorder, patients typically experience depression as the photo period of the day decreases with advancing winter. Women represent at least 75 percent of all patients with seasonal depression, and the mean age of presentation is 40. Patients rarely present over the age of 55 with seasonal affective disorder.

Phototherapy typically involves exposing the afflicted patient to bright light in the range of 1,500 to 10,000 lux or more, typically with a light box that sits on a table or desk. Patients sit in front of the box for approximately 1 to 2 hours before dawn each day, although some patients may also benefit from exposure after dusk. Alternatively, some manufacturers have developed light visors, with a light source built into the brim of the hat. These light visors allow mobility, but recent controlled studies have questioned the utility of this type of light exposure. Trials have typically lasted 1 week, but longer treatment durations may be associated with greater response.

Phototherapy tends to be well tolerated. Some patients have complained of irritability and headache. Ophthalmic damage was reported with earlier light boxes that emitted high-intensity light over long periods of exposure. Some investigators have implicated the concurrent use of tricyclic drugs or neuroleptic agents in ophthalmic damage associated with phototherapy. Newer light sources tend to use lower light intensities and come equipped with filters; patients are instructed not to look directly at the light source. As with any effective antidepressant, phototherapy has on rare occasions been implicated in switching some depressed patients into mania or hypomania.

Phototherapy is primarily indicated in the treatment of so-called seasonal depressions. Patients with more prominent hypersomnia as a feature of their seasonal depression may show more robust response to phototherapy. However, recent data suggest that phototherapy may benefit both other types of depression and other psychiatric disorders. For example, exposure to bright light has been suggested to reduce hospital stays in depressed patients, whether or not a seasonal component is evident. Also, case reports suggest that phototherapy may alleviate depression in immobile, terminally ill patients. Some preliminary data indicate phototherapy may benefit some patients with obsessive-compulsive disorder that has a seasonal variation.

In addition to seasonal depression, the other major indication for phototherapy may be in sleep disorders. Phototherapy has been used to decrease the irritability and diminished functioning associated with shift work. Sleep disorders in geriatric patients have reportedly improved with exposure to bright light during the day. Likewise, some evidence suggests that jet lag might be responsive to light therapy.

In mammals, the photo period of the day may significantly influence circadian cycles. For example, in some mammals a decreasing photo period stimulates hormonal events that precipitate hibernation.

Circadian rhythms are frequently disrupted in major depression, resulting in sleep disturbance and other symptoms. Exposure to light results in a phase advance that shifts the phase response curve earlier in the day. In addition, light suppresses the production of melatonin from the pineal gland at night. This hormone may affect

both mood and sleep cycles. The interplay of melatonin and the phase response curve in phototherapy is the focus of current investigation.

A number of controlled studies suggest that phototherapy is effective as monotherapy and as an adjunctive agent in the treatment of winter depressions. Studies have generally compared low-intensity light (<100 lux) with high-intensity light (>1500 lux). In general, patients with milder seasonal depressions appear more likely to respond to phototherapy treatment alone than do patients with more serious depression. Bipolar depression with a seasonal component frequently appears responsive to phototherapy, and these patients also appear to be more susceptible to light-induced mania.

Seasonal depressions often are characterized by the reverse neurovegetative symptoms of atypical depressions. However, studies of phototherapy in atypical depression without a seasonal component have failed to show benefit.

Research into the efficacy of phototherapy has been hindered by the lack of a placebo for bright light. Patients can differentiate low-intensity from high-intensity exposures.

ACUPUNCTURE AND ACUPRESSURE

Acupuncture originated in China about 5000 years ago and continues to be an important medical intervention in the East. America was first exposed to acupuncture by Chinese immigrants in the 1800s. The practice involves inserting fine needles at specific points in the body to relieve pain and other ailments. In the East, acupuncture is used to treat mental disturbances, including agitation, depression, insomnia, and anxiety states. Unfortunately, these indications have not been fully investigated in the West. The two indications for acupuncture that have been best studied in the West are pain management and the treatment of substance abuse.

Abundant data from both animal and human studies suggest that acupuncture can modulate pain responses. Clinically, acupuncture has been used to treat a variety of pain conditions including trigeminal neuralgia, neuropathies, chronic back pain, and headache. The analgesic effects of acupuncture appear to be at least partly mediated by endorphins and enkephalins. Multiple studies have reported that naloxone (Narcan) appears to block the analgesic effects of acupuncture.

A number of studies in the past 25 years suggest that acupuncture may be beneficial in the treatment of alcohol, cocaine, and opioid dependence. Studies in rats and humans suggest that acupuncture may limit narcotic withdrawal symptoms. Among alcoholic patients, acupuncture appears to improve the rate at which they complete detoxification programs and may reduce the frequency of alcohol cravings when used as a maintenance treatment.

Acupressure involves the application of tactile pressure (rather than the insertion of needles) in the same points used in acupuncture. Acupressure and acupuncture share a reported utility in the treatment of headaches, pain states, and possibly addiction. A number of recent studies have reported on the utility of acupressure in the treatment of nausea including postoperative nausea, nausea associated with chemotherapy, and hyperemesis gravidarum. Compared with pharmacological interventions, many patients reported equal benefit from acupressure.

SLEEP DEPRIVATION

Affective illness is characterized by sleep disturbance. Mania tends to be characterized by a decreased need for sleep, while depression may be associated with either hypersomnia or insomnia. Sleep deprivation may precipitate mania in bipolar I patients and temporarily relieve depression in unipolar patients. Approximately 60 percent of depressive disorder patients exhibit significant but transient benefit from total sleep deprivation. The positive results are typically reversed by the next night of sleep. The pathophysiology of sleep deprivation is not well understood; however, monaminergic transmission is reportedly enhanced by sleep deprivation. In animal studies, sleep deprivation resulted in substantial increases in both triiodothyronine (T_3) and thyroxine (T_4). Beta endorphin concentrations appear to be increased in patients who respond to sleep deprivation, but cortisol concentrations tend not to differ from those in nonresponders. Finally, sleep deprivation is associated in responders with increases in auditory evoked potentials (P300).

Several strategies have been used in an attempt to achieve a more sustained response to sleep deprivation. One method used serial total sleep deprivation with a day or two of normal sleep in between. This method does not achieve a sustained antidepressant response since the depression tends to return with normal sleep cycles. Another approach used phase delay in the time patients go to sleep each night, or partial sleep deprivation. In this method, patients may stay awake from 2 AM to 10 PM daily. Up to 50 percent of patients get same-day antidepressant effects from partial sleep deprivation, but this benefit also tends to wear off in time. However, in some reports serial partial sleep deprivation has successfully treated insomnia associated with depression. The third and probably most effective strategy combines sleep deprivation with pharmacological treatment of depression. A number of studies have suggested that total and partial sleep deprivation followed by immediate treatment with an antidepressant or lithium (Eskalith) sustains the antidepressant effects of sleep deprivation. Likewise, several recent reports have suggested that sleep deprivation accelerates the response to antidepressants including fluoxetine (Prozac) and nortriptyline (Aventyl, Pamelor). Sleep deprivation has also been noted to improve premenstrual dysphoria.

HERBS, VITAMINS, AND AMINO ACIDS

Nutritional status has long been deemed important in mental health, and certain, now rare, vitamin deficiencies can produce psychiatric symptoms. For example, severe niacin deficiency results in pellagra with its characteristic triad of skin lesions, gastrointestinal disorder, and psychiatric symptoms. The psychiatric symptoms include irritability and emotional instability progressing to severe depression progressing to disorientation, memory impairment, hallucinations, and paranoia. Folic acid deficiency is associated with depression and dementia, while B_{12} deficiency is associated with cognitive impairment, depression, and other affective symptoms. Severe malnutrition can result in apathy and emotional instability. Herbs of undefined safety and unsafe herbs are shown in Table 31.33-1.

Table 31.33-1 Safety Profiles of Herbs Sold for Brewing Teas

In 1968 the eminent chemist Linus Pauling coined the term “orthomolecular” to refer to the connection between the mind and nutrition. In his book *Orthomolecular Psychiatry* research articles were compiled supporting the notion that taking many times the recommended minimum daily dose of vitamins is useful in the treatment of schizophrenia and other psychiatric disorders. While some severe vitamin deficiencies may result in syndromes with a psychiatric component (i.e., niacin deficiency resulting in pellagra), empirical data and an American Psychiatric Association (APA) task force have failed to find evidence supporting the notion that schizophrenia and other disorders respond to vitamin therapies. However, that is not to say that vitamins and amino acids are of no importance in preserving mental health. Evidence indicates that severe vitamin deficiencies can result in psychiatric symptoms and that amino acid supplements may be pharmacologically useful in the treatment of some disorders. These are briefly reviewed below.

Thiamine, Vitamin B-12, and Folate In industrialized societies, severe vitamin deficiencies are rarely encountered except in certain populations. Those who are elderly, alcohol dependent, chronically ill, or who have certain types of gastrointestinal surgery are at greatest risk. Among the forms of vitamin deficiency most commonly encountered in the emergency room is acute thiamine depletion from alcohol dependence. While the chronic forms of thiamine deficiency that lead to

beriberi are rarely seen in the Western world, the fulminant depletion of already low stores of thiamine result in Wernicke's encephalopathy and Korsakoff's syndrome.

Wernicke's encephalopathy classically presents with the triad of ataxia, ophthalmoplegia, and mental confusion, but confusion and a staggering gait are perhaps most common. While Wernicke's encephalopathy is an acute process, Korsakoff's syndrome may be the permanent residue of this encephalopathy. Patients with Korsakoff's syndrome exhibit a well-circumscribed retrograde and anterograde amnesia that results from destruction of the mamillary bodies, and psychotic symptoms are also reported. Wernicke's encephalopathy is a medical emergency that responds to acute treatment with 50 mg of thiamine intravenously followed by 250-mg intramuscular injections daily until a normal diet is attained. The treatment of uncomplicated acute thiamine deficiencies usually involves 100 mg given orally 1 to 3 times per day.

Vitamin B_{12} deficiency or pernicious anemia is often seen in elderly adults, patients with gastric surgery, and malnourished depressed patients. The most typical psychiatric presentations include apathy, malaise, depressed mood, confusion, and memory deficits. B_{12} concentrations 150 mg/mL of serum are sometimes associated with these symptoms. B_{12} deficiency is a more common cause of reversible dementia and is typically assessed in dementia evaluations. The treatment of pernicious anemia usually involves daily intramuscular injections of 1000 mg of B_{12} for approximately 1 week followed by maintenance doses of 1000 mg every 1 to 2 months.

Folate deficiency has been associated with depression and dementia. Other psychiatric symptoms occasionally associated with depression include paranoia, psychosis, agitation, and confusion. The relationship of folate to depression has been debated over the years. Folate deficiency may be the consequence of anorexia in depressed patients and may also contribute to depression by interfering with the synthesis of norepinephrine and serotonin. Folate deficiency has been associated with anticonvulsant use (particularly phenytoin (Dilantin), primidone (Mysoline), and phenobarbital (Solfoton) and the sex steroids, including oral contraceptives and estrogen replacement. Perhaps the most common cause of folate deficiency is the malnourishment associated with alcoholism. Many folate deficiencies respond to 1 mg of folate orally per day; however, some more severe forms may require dosages of 5 mg up to three times a day.

Amino Acids Amino acids provide the substrate for neurotransmitters and have been used as adjunctive agents in the treatment of depression and sleep. L-Tryptophan (Fig. 31.33-1) was used for many years in this country and elsewhere to treat insomnia and augment standard antidepressants. L-Tryptophan, the precursor to serotonin, must be obtained from the diet. Patients who respond to serotonergic antidepressants may rapidly relapse into depression on a diet that is deficient in L-tryptophan. Interestingly, patients who respond to more-noradrenergic antidepressants appear less vulnerable to relapse with an L-tryptophan-free diet. In the 1970s and 1980s, l-tryptophan was used to augment standard antidepressants. It was part of the Newcastle cocktail for refractory depression, which included clomipramine (Anafranil) or phenelzine (Nardil), lithium, and L-tryptophan. A number of small studies have suggested that it did have some effect on augmenting antidepressants; however, it has little antidepressant effect of its own.

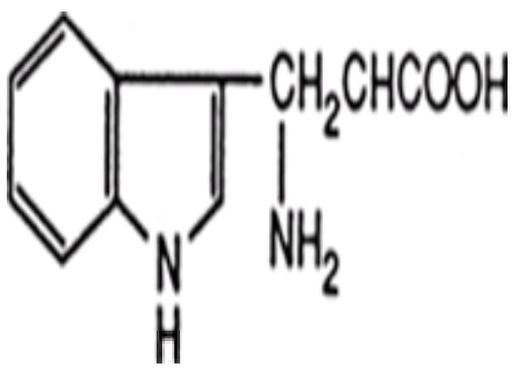

FIGURE 31.33-1 Molecular structure of L-tryptophan.

l-Tryptophan was also used as an over-the-counter treatment of insomnia in the United States. A number of studies suggested that l-tryptophan in doses of 1 to 4 grams before bedtime decreased sleep latency. l-Tryptophan has been unavailable in the United States since 1989, because of its association with the eosinophilia-myalgia syndrome, which may have been secondary to an impurity resulting from the processing of the compound.

Another amino acid that has been examined as an augmentor to antidepressants is phenylalanine. Phenylalanine is converted to tyrosine as a catecholamine precursor. Phenylalanine has been added to selegiline (Eldepryl) successfully in the treatment of some patients with refractory major depression. Tyrosine has been investigated as an augmentor to tricyclic drugs and may also have some mild antidepressant activity itself.

Herbs Over the years many herbs have been purported to affect mood and mental health. Homeopathic practitioners in many parts of the world, particularly Europe, use plant derivatives including belladonna, *Aconitum napellus*, and *argentum nitricum* (silver nitrate) to treat a variety of emotional maladies. Herbalists continue to recommend *Valeriana officinalis* (valerian) for insomnia, *Passiflora incarnata* (passion flower) for nervous unrest, and *Ginkgo biloba* (ginkgo) for memory problems and anxiety. Ginkgo is currently being investigated in the treatment of selective serotonin reuptake inhibitor (SSRI)-induced sexual dysfunction. However, few herbs have been subjected to any systematic investigation. One herb that has caught the attention of Western psychiatry recently is *Hypericum perforatum* (St. John's wort) for the treatment of major depressive disorders. St. John's wort has been used in folk medicine for hundreds of years and is still commonly used in Europe. In Germany, it is estimated that several million prescriptions for hypericum are obtained annually for the treatment of depression, anxiety, and sleep problems. Over 20 studies in the past 15 years have compared hypericum extracts with placebo and tricyclic drugs. A recent meta-analysis of these studies found that hypericum extracts were more effective than placebo and as effective as standard antidepressants in the treatment of mild-to-moderate depression. Furthermore, the hypericum-treated patients exhibited far fewer adverse effects than those taking standard antidepressants. The major problem with these studies has been the lack of rigor in the diagnosis of depression and the assessment of efficacy. Most of the trials were performed with private patients of internists, psychiatrists, and general practitioners.

ENDOCRINE THERAPIES

Hormones may act directly as neurotransmitters and also indirectly influence the activity of some neurotransmitters. For example, T_3 has numerous cortical receptors, and this thyroid hormone may also modulate central and peripheral noradrenergic activity. These actions may be the basis for liothyronine's (Cytomel) utility as an augmentor to antidepressants. In addition to thyroid hormone, the psychotropic effects of a number of other hormones remain under investigation.

Estrogen Estrogen has been used for many years to relieve menopausal symptoms and as hormone replacement therapy after menopause. Some evidence suggests that estrogens may have some antidepressant effects in postmenopausal women and that progesterone may be depressogenic. Estrogens appear to lower monoamine oxidase (MAO) concentrations and increase synaptic availability of monamines. In addition, estrogen displaces tryptophan from its albumin binding site and enhances the availability of this precursor to serotonin synthesis. Attempts to augment tricyclic drugs with estrogen have been mixed. Some studies have indicated that estradiol in dosages of 15 to 25 mg per day does augment tricyclic drug response in some women; other studies have failed to find any benefit. More-recent data suggest that estrogens modulate the effect of serotonin agonists. For example, estrogen appears to enhance the hormonal response of the serotonin agonist methylchlorophenyl piperazine (mCPP). In addition, preliminary data indicate that postmenopausal women who are on hormone replacement therapy are more likely to have a robust response to fluoxetine treatment for major depressive disorder than those who are not.

Dehydroepiandrosterone Dehydroepiandrosterone (DHEA), a precursor hormone for both estrogens and androgens, is available over the counter. Recent years have seen an interest in DHEA for improving cognition, depression, sex drive, and general well-being in elderly adults. Some reports suggest that DHEA in dosages of 50 to 100 mg per day increases the sense of physical and social well-being in women aged 40 to 70 years. Reports also exist of androgenic effects, including irreversible hirsutism, hair loss, voice deepening, and other undesirable sequelae. In addition, DHEA has at least a theoretical potential of enhancing tumor growth in people with latent, hormone-sensitive malignancies such as prostate, cervical, and breast cancer. Despite its significant popularity, there is a dearth of controlled data on the safety or efficacy of DHEA.

Melatonin Melatonin is another popular over-the-counter hormone used by many Americans on a regular basis for insomnia and jet lag. Melatonin is produced by the

pineal gland, and commercially available supplies are derived synthetically or from hog pineal glands. The hormone is released naturally by the pineal gland early in the sleep cycle and appears to contribute to natural sleep cycles. A number of small, brief studies have suggested that melatonin can act as a hypnotic in doses of 0.2 and 5 mg at night. Mostly anecdotal reports suggest that melatonin can also reduce the insomnia associated with jet lag. Some uncontrolled reports suggest that melatonin has mild antidepressant effects. However, because of its reciprocal relationship to b-adrenergic receptor activity, it may worsen depression in some patients. The long-term effects of melatonin use are unknown, and the efficacy of melatonin has been poorly studied at this time, given the widespread use of the drug.

Testosterone Testosterone derivatives are anabolic-androgenic steroids that appear to have psychotropic effects in both men and women. Testosterone is now given quite commonly to postmenopausal women to enhance libido, since this hormone appears to have significant effects on libido in both men and women. In hypogonadal men, testosterone may improve sexual dysfunction, mood, energy, muscle to fat ratio, and sex drive. High doses of testosterone derivatives are sometimes abused by bodybuilders, football players, and even teenagers trying to increase muscle bulk. Anabolic steroids have sometimes been associated with rage attacks, aggressive behavior, and untoward physical effects including accelerated atherosclerosis, testicular atrophy, alopecia, and enhanced tumor growth. Women who use testosterone at high doses also may experience irreversible virilization. Testosterone is now available in a transdermal patch and has been marketed fairly aggressively, resulting in a recent increase in prescriptions for both men and women.

OTHER PHARMACOLOGICAL STRATEGIES

Reboxetine In 1997, the first truly selective noradrenaline (norepinephrine) reuptake inhibitor (NRI), reboxetine, was introduced in Europe for the treatment of depression. Pharmacologically and chemically unrelated to tricyclics or SSRIs, reboxetine has significant affinity for the norepinephrine transporter, and little affinity for other neuroreceptors including serotonin, dopamine, histamine, muscarinic, and alpha adrenergic sites. Norepinephrine depletion studies suggest that while norepinephrine reuptake inhibition may improve all core symptoms of depression, norepinephrine regulation may be most closely correlated with patient improvements in energy, interest, concentration, agitation, helplessness, and hopelessness. Reboxetine, therefore, through its mechanism of action, may be useful for those depressed patients with either anergic symptoms or for those patients with comorbid anxiety.

Reboxetine is rapidly and extensively absorbed following oral administration. Maximal concentrations of 111 ng/mL are achieved within a mean of 2.4 hours after administration of a 4-mg tablet. The half-life of reboxetine is approximately 13 hours, and steady state levels are achieved within 5 days of dosing. Reboxetine is extensively metabolized after oral administration, primarily via cytochrome P450 (CYP) isoenzyme 3A4. Reboxetine does not inhibit the activity of cytochrome P450: CYP 1A2, CYP 2C9, CYP 2C19, and CYP 2E1. Reboxetine is a very weak inhibitor of CYP 2D6 and CYP 3A4, and based upon its low affinities for these enzymes, the potential for significant drug-drug interactions is minimal.

Area under the curve (AUC) and serum half-life values for reboxetine increase in elderly patients and in those with hepatic or renal insufficiency. A reduction in dose is warranted in these patients.

Reboxetine is an effective treatment for depression and is demonstrated effective long-term (one year) in preventing symptom relapse. Clinical studies have demonstrated efficacy for reboxetine as compared to imipramine, desipramine, and fluoxetine. In one study reboxetine demonstrated greater impact on patient-rated social functioning compared with fluoxetine. In a study of inpatients, reboxetine was also found to be significantly more effective than fluoxetine for patients with severe depression. A one-year maintenance study demonstrated the long-term effectiveness of reboxetine in preventing symptom relapse. One, double blind, placebo-controlled trial has demonstrated efficacy for reboxetine in the treatment of panic disorder. In this study, reboxetine-treated patients demonstrated significant reductions in weekly panic attacks, phobic sympton scores, and level of anticipatory anxiety.

Reboxetine is available in 4-mg functionally scored tablets. The initial dose for adults is 4 mg twice daily. For most patients, an increase in dose will not be necessary. However, if needed, the dose may be increased to a total of 10 mg/day in two doses after 3 weeks. In elderly patients or those with renal or hepatic impairment, therapy may be initiated at 2 mg (one-half tablet) one-half tablet) twice daily, and increased to a maximum of 6 mg/day in two doses after 3 weeks.

A variety of studies have demonstrated greater efficacy for norepinephrine-active compounds in the treatment of severe depression in comparison with SSRIs. While data is currently limited to one study, reboxetine may prove useful for patients with severe or treatment resistant depression. Depression in elderly patients is often characterized by lack of energy, impaired cognition, apathy, and agitation. Reboxetine appears to be effective and well tolerated in elderly patients at prescribed doses, offers minimal risk of drug interactions, and causes less agitation than placebo. In clinical tests, reboxetine does not appear to have a negative impact upon patient cognition.

As a result of the selectivity of reboxetine for norepinephrine, the product is generally well tolerated with a benign side effect profile. As compared with tricyclics, reboxetine has been associated with fewer reports of dry mouth, constipation, and other anticholinergic side effects. In addition, typical CNS side effects such as, agitation; anxiety; nervousness; and daytime somnolence; are reported less frequently by reboxetine-treated patients as compared to placebo-treated patients. A small percentage of patients reported sexual dysfunction on reboxetine therapy. The safety of reboxetine in overdose has been studied in animal models that suggest a very wide therapeutic margin for this product. Pharmacovigilance reports suggest safety in overdose for reboxetine, without a single death attributed solely to reboxetine overdosage. Unlike tricyclics, clinical pharmacology studies suggest that reboxetine does not prolong the QTc interval or impact significantly on cardiac conduction.

Bromocriptine and Levodopa Bromocriptine (Parlodel) (Fig. 31.33-2) is an ergot-derived dopamine agonist that has been used in medicine for many years as an adjunctive agent in the treatment of Parkinson's disease. Similarly, levodopa (Larodopa) (Fig. 31.33-3) is a dopamine agonist that has been a central treatment for Parkinson's for many years. Other uses for bromocriptine have included the treatment of hyperprolactin states including those associated with pregnancy and prolactinomas. Both levodopa and bromocriptine have proved useful in the treatment of restless leg syndrome and periodic leg movements. All dopamine agonists, including levodopa and bromocriptine, have been associated with producing psychotic symptoms, particularly in high doses and in elderly adults.

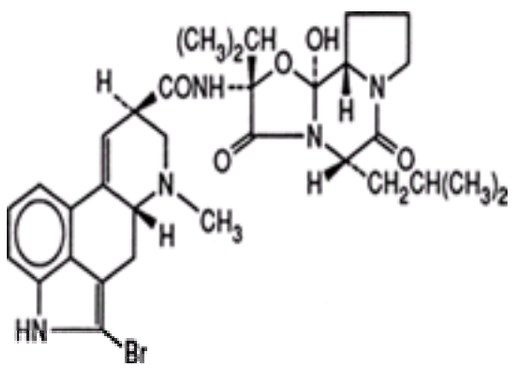

FIGURE 31.33-2 Molecular structure of bromocriptine.

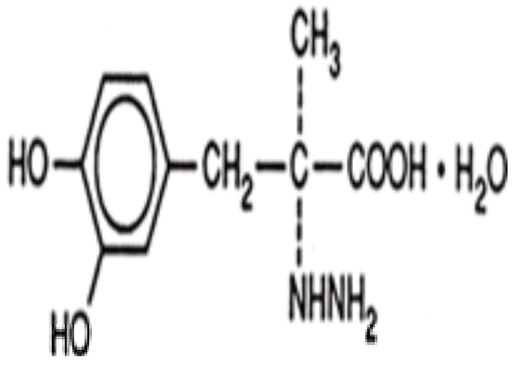

FIGURE 31.33-3 Molecular structure of levodopa.

In general psychiatry, bromocriptine has been used in the treatment of a variety of disorders. Bromocriptine has proved useful in treating neuroleptic malignant syndrome in many case studies. Antipsychotic- and head injury–induced akathisia has also responded to bromocriptine, although psychosis has also sometimes been induced in these populations. Some reports exist of bromocriptine helping with the negative symptoms of schizophrenia. Bromocriptine may also reverse the galactorrhea and amenorrhea associated with neuroleptic drug use. Recently, there has been an interest in the use of bromocriptine in addictive states, particularly cocaine dependence. The dopamine agonist properties of bromocriptine would appear to make it a rational choice for treating dependence to a dopamine agonist illicit drug. While some case reports have suggested that bromocriptine is efficacious in treating cocaine dependence, results of double-blind studies have generally been negative. Likewise, bromocriptine has not proved particularly useful in the treatment of alcoholism.

Fenfluramine and Dexfenfluramine Fenfluramine (Fig. 31.33-4) and dexfenfluramine are serotonin agonists that had been approved by the Food and Drug Administration as anorexiants. While fenfluramine has been available for many years, its popularity did not soar until reports of its success in the treatment of obesity when combined with the stimulant phentermine (Adipex, Zantryl) were widely reported in the popular media. In double-blind studies, both fenfluramine and dexfenfluramine were efficacious in the treatment of obesity. The drugs were approved for the treatment of significant or morbid obesity but quickly began to be used by many people with less-substantial weight difficulties. By 1996 16 million Americans were using one of these drugs. There were early reports of pulmonary hypertension associated with fenfluramine but the incidence was extremely low and did not significantly influence prescription practices. In 1997 a strong link was noted between the use of fenfluramine and valvular heart disease in young patients. All heart valves potentially involved in these patients were examined; the histopathology included plaque formation and fibrosis of the chordae. A number of patients required valve replacement surgery, and it was estimated that up to 25 percent of patients on fenfluramine or dexfenfluramine might have sustained some (although generally mild) valve damage. Thus these drugs were pulled from the American market in 1997, and many patients who had been on fenfluramine or dexfenfluramine were advised to obtain an echocardiogram.

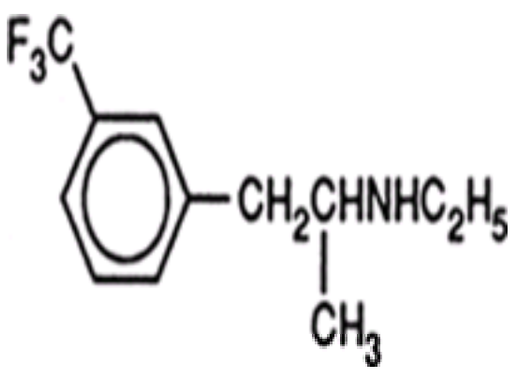

FIGURE 31.33-4 Molecular structure of fenfluramine.

In addition to their ill-fated use as anorexiants, fenfluramine and dexfenfluramine have had other uses in psychiatry. Perhaps most importantly, they have been used as chemical probes to evaluate the serotonin system and its relationship to the endocrine system. Case reports have suggested that fenfluramine may augment the anti–obsessive-compulsive and antidepressant effects of SSRIs. There have also been rare reports of serotonin syndrome associated with SSRI fenfluramine or dexfenfluramine combinations. Fenfluramine has been evaluated in the treatment of eating disorders such as bulimia nervosa and anorexia nervosa with inconsistent results. Results of recent studies of fenfluramine in the treatment of autistic disorder were negative.

Yohimbine Yohimbine (Yocon) (Fig. 31.33-5), an a_2-adrenergic receptor antagonist, has proved useful in the study of various anxiety disorders. In most animal models of anxiety, yohimbine is anxiogenic. In human studies, yohimbine has been used as a chemical probe to examine the biology of panic disorder and posttraumatic stress disorder. For example, intravenous yohimbine produces panic attacks in patients with a history of panic disorder at a significantly higher rate than in patients without such a history. Likewise, combat veterans with posttraumatic stress disorder who are exposed to oral or intravenous have a greater acoustic startle response than patients without posttraumatic stress disorder. In addition, some studies of combat veterans with posttraumatic stress disorder have demonstrated that yohimbine can precipitate an increase in core symptoms, including intrusive traumatic thoughts, emotional numbing, and grief.

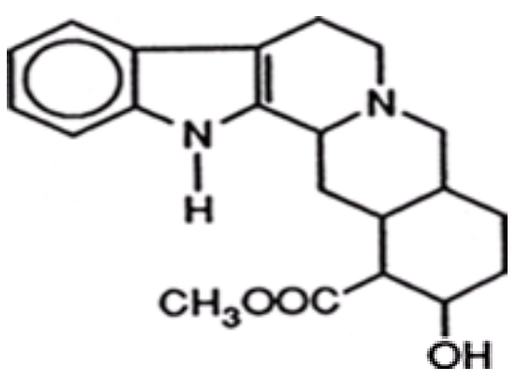

FIGURE 31.33-5 Molecular structure of yohimbine.

Yohimbine may also have therapeutic applications. For example, yohimbine appears efficacious in the treatment of some types of sexual dysfunction. Several case reports suggest that yohimbine can reverse the sexual dysfunction associated with the use of a serotonergic antidepressant. A number of placebo-controlled studies have examined the effects of yohimbine on various types of impotence. Yohimbine in combination with trazodone (Desyrel) has been effective in the treatment of psychogenic impotence. Some studies have also suggested that yohimbine may help mixed psychogenic and organic impotence while other studies have been negative.

The utility of yohimbine in other disorders remains to be seen. A few reports have suggested that yohimbine may be a useful strategy in the treatment of refractory depression but these reports have not been conclusive.

Dantrolene Dantrolene (Dantrium) is a direct-acting skeletal muscle relaxant. In contemporary clinical psychiatry, dantrolene is one of the potentially effective treatments for neuroleptic malignant syndrome, catatonia, and serotonin syndrome. Dantrolene is derived from hydantoin, as indicated in its molecular structure (Fig. 31.33-6). Dantrolene is structurally and pharmacologically unrelated to other skeletal muscle relaxants.

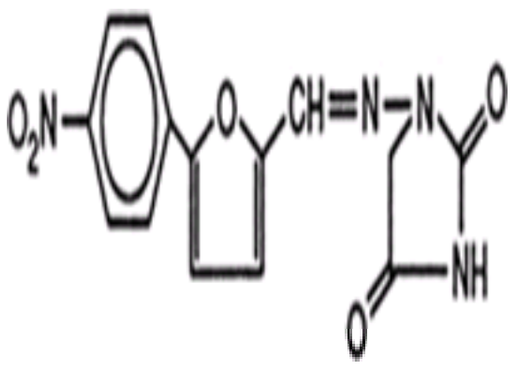

FIGURE 31.33-6 Molecular structure of dantrolene.

The skeletal muscle relaxant effect of dantrolene can cause muscle weakness and such symptoms as slurring of speech and drooling. Dantrolene also affects the gastrointestinal system (e.g., diarrhea) and the nervous system (e.g., headache and depression) and possibly toxic effects on hepatocytes, as indicated by an association with elevated liver function test results.

The primary psychiatric indication for intravenous dantrolene is muscle rigidity in neuroleptic malignant syndrome. Dantrolene is almost always used in conjunction with appropriate supportive measures and a dopamine receptor agonist (e.g., bromocriptine). Dantrolene has been used in efforts to treat other psychiatric conditions characterized by life-threatening muscle rigidity, such as catatonia and serotonin syndrome.

Because of its potential for severe adverse effects, dantrolene should not be used by psychiatric patients for any long-term treatment. Dantrolene should be used with caution by patients with hepatic, renal, and chronic lung diseases.

Disulfiram Disulfiram (Antabuse) is used to ensure abstinence in the treatment of alcohol dependence. Its main effect is to produce a rapid and violently unpleasant reaction in a person who ingests even a small amount of alcohol while taking disulfiram. Because of the risk of severe and even fatal disulfiram-alcohol reactions, disulfiram therapy is used less often today than previously.

The molecular structure of disulfiram is presented in Figure 31.33-7.

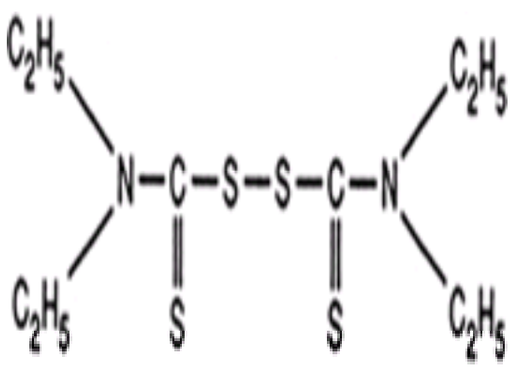

FIGURE 31.33-7 Molecular structure of disulfiram.

Disulfiram is an aldehyde dehydrogenase inhibitor that interferes with the metabolism of alcohol and produces a marked increase in blood acetaldehyde levels.

The treatment of alcohol dependence requires careful monitoring, since a patient can simply decide not to take the disulfiram; compliance with the medication should be checked if possible.

The intensity of the disulfiram-alcohol reaction varies with each patient. In extreme cases it is marked by respiratory depression, cardiovascular collapse, myocardial infarction, convulsions, and death. Most fatal reactions occur in patients who are taking more than 500 mg a day of disulfiram and who consume more than 3 ounces of alcohol.

The adverse effects of disulfiram in the absence of alcohol consumption include fatigue, dermatitis, impotence, optic neuritis, a variety of mental changes, acute polyneuropathy, and hepatic damage. A metabolite of disulfiram inhibits dopamine hydroxylase and thus potentially exacerbates psychosis in patients with psychotic disorders.

Drug-Assisted Interviewing The use of amobarbital (Amytal) and other sedative-hypnotic agents to facilitate interviewing in some patients has become uncommon but is still used in some settings. The primary indication for drug-assisted interviewing is ostensibly to recover repressed material in amnesiac patients and as a diagnostic probe in assessing conversion disorder, catatonia, confusion, and mutism. For example, if a patient is mute because of severe anxiety, the use of a sedative-hypnotic agent may facilitate gaining access to traumatic material. Catatonic patients anecdotally may become more activated when a hypnotic agent is administered. Confusion due to functional causes such as schizophrenia may improve with drug-assisted interviewing, while confusion due to organic causes may worsen. Despite the use of drug-assisted interviewing for over 60 years, there is very little evidence that interviewing with a sedative-hypnotic agent is any better than skillful interviewing without drugs. In some instances, memories are inhibited by the concurrent use of a sedative-hypnotic agent.

The amobarbital interview involves administering a 10 percent solution of sodium amobarbital at 0.5 to 1.0 mL per minute until mild sedation is achieved. Alternatively, intravenous diazepam (Valium) or lorazepam (Ativan) have been used as safer alternatives to barbiturates. Respiratory suppression is the most serious complication of drug-assisted interviewing, and the clinician must be prepared to take emergency measures as needed.

Given the risks of amytal interviews, drug-assisted interviewing using safer agents, such as benzodiazepines, has emerged as an alternative. Diazepam, lorazepam and other benzodiazepines have been used for this purpose. One of the most recent studies on drug-assisted interviewing used the very short-acting benzodiazepine midazolam (Versed) and a pulse oximeter to decrease the risk of respiratory suppression. Patients are treated with midazolam, 1 mg/mL given up to every minute depending on the levels of sedation observed and oxygen saturation. In addition to having a wider therapeutic index than barbiturates, benzodiazepines have the added advantage of having an effective antagonist available, flumazenil (Romazicon), which can quickly reverse the effects if needed.

Placebos Pharmacologically inactive substances have long been known to sometimes produce significant clinical benefits. The belief that a compound is helpful may often result in a patient's deriving considerable benefit from taking a substance, whether it is known to be pharmacologically active or not. For many psychiatric disorders, including mild-to-moderate depression and some anxiety disorders, well over 30 percent of patients can have significant improvement or a remission of symptoms on a placebo. For other conditions such as schizophrenia, manic episodes, and psychotic depression, the placebo response rate is very low. While suggestion is undoubtedly important in the efficacy of placebos (and active drugs), placebos may produce biological effects. For example, placebo-induced analgesia may sometimes be blocked by naloxone, which suggests that endorphins may mediate the analgesia derived from taking a placebo. It is conceivable that placebos may also stimulate endogenous anxiolytic and antidepressant factors, resulting in clinical improvement in patients with depression and anxiety disorders.

Just as placebos may produce benefit, they may also have adverse effects. In many studies, some adverse effects are likely to be more common with placebos than with the active drug. Some patients will not tolerate placebos despite the fact that they are supposedly inert. It is easy to discount such patients as being overly

suggestible. However, if beneficial endogenous factors may be stimulated by placebos, perhaps toxic endogenous factors may also be produced.

There are relatively few instances in modern medicine in which placebos should be routinely used. Clinical drug investigations are the clearest indication for the use of a placebo. However, even in pharmacological studies, the use of placebos sometimes presents an ethical dilemma. While placebo controls represent the gold standard in determining the safety and pharmacological efficacy of a given compound, the health of some patients may be jeopardized by being given placebo when there are known effective interventions. Psychotic patients sometimes decompensate on placebos, and placebo-treated depressed patients sometimes need to be hospitalized. However, without a placebo control in the early investigation of a new psychotropic agent, questions cannot be satisfactorily answered about the safety and efficacy of the drug.

Placebos are also routinely used in the detoxification of patients dependent on pain medications. This is often done in the form of pain cocktails in which the percentage of active drug is reduced with the patient being blind to the rate or extent of the reduction. The patient receives the same amount and type of liquid each time, although the composition of the cocktail may change daily. Patients are informed ahead of time, however, that the cocktail may be altered to contain the lowest amount of active drug they can tolerate. Some pain patients have been able to discontinue taking their narcotics and muscle relaxants altogether by being treated with a cocktail that no longer contains active drug.

Prudence is needed in contemplating the use of a placebo in clinical practice. Treating a patient with a placebo without consent can seriously undermine a patient's confidence in the physician if, and when, it is discovered.

TREATMENTS OF HISTORICAL SIGNIFICANCE

Chemical Convulsive Therapies Convulsive therapies for the treatment of serious psychiatric disorders date back hundreds of years, with the Swiss physician Paracelsus reportedly giving camphor by mouth to induce seizures and treat lunacy in the sixteenth century. Several European manuscripts from the 1700s describe the benefits of camphor-induced seizures for the treatment of mania and other forms of insanity. These manuscripts were largely forgotten until the work of Ladislas von Meduna in the 1930s. Von Meduna had experimented with intramuscular camphor monobromide, caffeine, strychnine, brucine, and other compounds before settling on pentylenetetrazol as a more reliable convulsant for the treatment of dementia precox. Pentylenetetrazol was more soluble than many other compounds and also had a quicker onset of action. Meduna typically used an initial dose of 5 mL of a 10 percent solution of pentylenetetrazol, followed by additional doses every minute if convulsions were not achieved. The major drawbacks of these chemical convulsions is that seizures sometimes did not occur, and patients would experience significant preictal discomfort including nausea and anxiety and thus, tend to decline further treatment. In the late 1930s and early 1940s, chemical convulsive therapy was replaced by the considerably more reliable electroconvulsive therapy (ECT).

Coma-Inducing Therapies. Insulin coma therapy emerged approximately the same time in the 1930s as electroconvulsive therapy. As von Meduna observed that epilepsy and dementia praecox were incompatible, Manfred Sakel observed that dementia praecox patients who went into a coma tended to come out less symptomatic. The treatment involved using incrementally higher doses of intramuscular insulin until the patient became comatose. Comas were initially terminated with glucagon after approximately 15 minutes, but an attempt was made to increase subsequent comas to a maximum of 60 minutes. Patients often required 60 or more treatments before results were seen. Complications including arrythmias and laryngeal spasms were not uncommon, and insulin coma therapy had a fatality rate of at least 1 percent and in some samples was considerably higher. The danger of the procedure and a controlled study in 1962 that suggested that it was no more effective than a similar period of unconsciousness induced by barbiturates hastened the demise of the procedure. However, some patients clearly appeared to respond to insulin coma therapy who did not respond to other available treatments.

Variations of insulin coma therapy included atropine coma therapy used briefly in the 1950s. Atropine in dosages of up to 200 mg a day was given to induce comas lasting 6 to 8 hours. If the patient did not wake up spontaneously, the coma was aborted by intramuscular physostigmine (Antilirium). As with insulin coma therapy, atropine coma therapy was said to be effective for the treatment of schizophrenia and mania. The most serious complications were hyperthermia and rhabdomyolysis. By the late 1950s, coma therapies had been all but abandoned for safer treatments, including ECT and effective antipsychotic drugs.

Continuous Sleep Therapy In the 1930s, therapies that altered consciousness for extended periods by seizure or coma were thought to be effective in the treatment of psychosis. Even earlier, psychosis was treated by inducing a state of continuous sleep for 10 days or more. Continuous sleep therapy was introduced for the treatment of psychosis into the early 1920s. This intervention originally involved the use of barbiturates to keep patients asleep 20 hours or more a day with brief interruptions for the patient to eat and use the bathroom. Complications of barbiturate-induced continuous sleep included allergic reactions, seizures on withdrawal, and respiratory depression ending in death. Later the combination of chlorpromazine (Thorazine) with benzodiazepines and other hypnotics was used to keep patients asleep for therapeutic purposes. Although there are some reports of improvement in anxiety states, obsessive-compulsive disorder, and schizophrenia, no controlled data are available to support these claims. Given the significant morbidity and clear lack of efficacy of this method, it was largely abandoned in the United States by the 1960s.

Hallucinogen Therapy. Many cultures have used hallucinogens, including mescaline, psilocybin, and ergots, for thousands of years to gain spiritual and personal insight. Lysergic acid diethylamide (LSD) was synthesized in the 1930s and was marketed to psychiatrists and other practitioners in the late 1940s under the trade name Desylid as a tool for understanding psychosis and for facilitating psychotherapy. Using LSD reportedly helped patients capture repressed memories and deal with anxiety, and it allowed patients to gain insight through an analysis of the primary process induced by the hallucinogen. Oral doses of 150 to 250 µg were administered occasionally by psychiatrists throughout the 1950s and early 1960s to facilitate psychotherapy with some patients. In the 1960s Timothy Leary advocated the wide spread use of hallucinogens, but the drugs were outlawed as class I controlled substances in 1965.

While no longer used for therapeutic purposes in this country, LSD has fulfilled part of its early promise as a probe for psychosis. Our more recent understanding of the pharmacology of LSD and its affinity to serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT $_2$) receptors has supported the interest in developing serotonin-dopamine antagonists (atypical antipsychotics) with the 5-HT $_2$-receptor blocking properties.

Detoxification Therapies. The notion that some mental disorders may be related to a toxin of some sort is very old. Various methods have been used to combat potential toxins suspected to be included in the etiology of psychosis. More recent attempts to deal with suspected toxins include the use of blood transfusions in the 1940s and 1950s and hemodialysis in the 1970s. A few case reports in the late 1970s suggested that hemodialysis was an effective short-term and maintenance treatment in some schizophrenic patients. Patients were dialyzed daily until improvement was seen and were then maintained with dialysis every 2 to 8 weeks. Several patients were said to recover with hemodialysis and to relapse when the treatments were stopped. The investigators presumed a leucine-containing endorphin was the responsible toxin, but they (and other investigators) were unable to replicate their initial findings. Thus, hemodialysis joined blood transfusions and other detoxification therapies in the annals of psychiatric history.

SUGGESTED CROSS-REFERENCES

Of interest to readers of this section are Section 55.1 on the history of psychiatry, Section 31.31 on electroconvulsive therapy, Section 51.3d on the somatic treatment of mood disorders, Section 12.8 on the somatic treatment of schizophrenia, Section 1.11 on psychoendocrinology, Section 11.7 on hallucinogens, and Section 31.1 on the general principles of psychopharmacology. Section 11.13 deals with DHEA and steroid abuse, Section 28.6 with chronic pain and placebo. Alternative medicine is covered in Section 28.8.

SECTION REFERENCES

al-Semaan Y: Bromocriptine as adjunctive therapy to clozapine in treatment-resistant schizophrenia (Letter). Can J Psychiatry *41:*484, 1996.

Barbini B, Bertelli S, Colombo C, Smeraldi E: Sleep loss, a possible factor in augmenting manic episode. Psychiatry Res *65:*121, 1996.

Belluomini J, Litt RC, Lee KA, Katz M: Acupressure for nausea and vomiting of pregnancy: A randomized, blinded study. Obstet Gynecol *84:*245, 1994.

Benedetti F, Barbini B, Lucca A, Campori E, Colombo C, Smeraldi E: Sleep deprivation hastens the antidepressant action of fluoxetine. Eur Arch Psychiatry Clin Neurosci *247:*100, 1997.

Brill JR: Acupressure for nausea and vomiting of pregnancy: A randomized, blinded study (Letter; comment). Obstet Gynecol *85:*159, 1995.

Brenot F, Herve P, Petitpretz P, Parent F, Duroux P, Simonneau G: Primary pulmonary hypertension and fenfluramine use. Br Heart J *70:*537, 1993.

Cappiello A, McDougle CJ, Malison RT, Heninger GR, Price LH: Yohimbine augmentation of fluvoxamine in refractory depression: A single-blind study. Biol Psychiatry *38:*765, 1995.

*Connolly HM, Crary JL, McGoon MD, Hensrud DD, Edwards BS, Edwards WD, Schaff HV: Valvular heart disease associated with fenfluramine-phentermine (see comments). N Engl J Med *337:*581, 1997.

Ernst E: Acupuncture/acupressure for weight reduction? A systematic review. Wien Klin Wochenschr *109:*60, 1997.

Fahy TA, Eisler I, Russell GF: A placebo-controlled trial of d-fenfluramine in bulimia nervosa (see comments). Br J Psychiatry *162:*597, 1993.

*Fan CF, Tanhui E, Joshi S, Trivedi S, Hong Y, Shevde K: Acupressure treatment for prevention of postoperative nausea and vomiting (see comments). Anesth Analg *84:*821, 1997.

Feinsod M, Kreinin B, Chistyakov A, Klein E: Preliminary evidence for a beneficial effect of low-frequency, repetitive transcranial magnetic stimulation in patients with major depression and schizophrenia. Depress Anxiety *7:*65, 1998.

Figiel GS, Epstein C, McDonald WM, Amazon-Leece J, Figiel L, Saldivia A, Glover S: The use of rapid-rate transcranial magnetic stimulation (rTMS) in refractory depressed patients. Neuropsychiatry Clin Neurosci *10:*20, 1998.

Galletly C, Clark A, Tomlinson L: Evaluation of dexfenfluramine in a weight loss program for obese infertile women. Int J Eating Disord *19:*209, 1996.

George MS, Wassermann EM, Williams WA, Callahan A, Ketter TA, Basser P, Hallett M, Post RM: Daily repetitive transcranial magnetic stimulation (rTMS) improves mood in depression. Neuroreport *6:*1853, 1995.

*Gross F, Gysin F: Phototherapy in psychiatry: Clinical update and review of indications. Encephale *22:*143, 1996.

Handelsman L, Rosenblum A, Palij M, Magura S, Foote J, Lovejoy M, Stimmel B: Bromocriptine for cocaine dependence. A controlled clinical trial. Am J Addict *6:*54, 1997.

Hemmeter U, Seifritz E, Hatzinger M, Muller MJ, Holsboer-Trachsler E: Serial partial sleep deprivation as adjuvant treatment of depressive insomnia. Prog Neuropsychopharmacol Biol Psychiatry *19:*593, 1995.

Labbate LA, Lafer B, Thibeu HA, Rosenbaum JF, Sechs GS: Influence of phototherapy treatment duration for seasonal affective disorder: Outcome at one vs. two weeks. Biol Psychiatry *38:*747, 1995.

Lafer B, Sachs GS, Thibau HA, Rosenbaum JF: Phototherapy for seasonal affective disorder: A blind comparison of three different schedules. Am J Psychiatry *151:*1081, 1994.

*Linde K, Ramirez G, Mulraw CD, Pauls A, Weidenhammer W, Melchart D: St. John's wort for depression—an overview and meta-analysis of randomized clinical trials (see comments). Br Med J *313:*253, 1996.

Ma J: Periomarthritis treated with pain point pressure in combination with local exercises. J Tradit Chin Med *15:*289, 1995.

Maa SH, Gauthier D, Turner M: Acupressure as an adjunct to a pulmonary rehabilitation program. J Cardiopulm Rehabil *17:*268, 1997.

Mann K, Klingler T, Noe S, Roschke J, Muller S, Benkert O: Effects of yohimbine on sexual experiences and nocturnal penile tumescence and rigidity in erectile dysfunction. Arch Sex Behav *25:*1, 1996.

*Nahas Z, Bohning DE, Molloy MA, Ovotz JA, Risch SC, George MS: Safety and feasibility of repetitive transcranial magnetic stimulation in the treatment of anxious depression in pregnancy: A case report. J Clin Psychiatry *60:*50, 1999.

*Neary JT, Bu Y: Hypericum LI 160 inhibits uptake of serotonin and norepinephrine in astrocytes. Brain Res *816:*358, 1999.

Oren DA, Jacobsen FM, Wehr TA, Cameron CL, Rosenthal NE: Predictors of response to phototherapy in seasonal affective disorder (published erratum appears in Compr Psychiatry *33:*419, 1992). Compr Psychiatry *33:*111, 1992.

*Palmer ME, Rao RB: Problems evaluating contamination of dietary supplements. New Engl J Med *340:*568, 1999.

Pascual-Leone A, Rubio B, Pellardâo F, Catalâa MIJ: Rapid-rate transcranial magnetic stimulation of left dorsolateral prefrontal cortex in drug-resistant depression. Lancet *348:*233, 1996.

Powell BJ, Campbell JL, Landon JF, Liskow BI, Thomas HM, Nickel EJ, Dale TM, Penick EC, Samuelson SD, Lacoursiere RB: A double-blind, placebo-controlled study of nortriptyline and bromocriptine in male alcoholics subtyped by comorbid psychiatric disorders. Alcohol Clin Exp Res *19:*462, 1995.

*Saitz R, O'Malley SS: Pharmacotherapies for alcohol abuse. Withdrawal and treatment. Med Clin North Am *81:*881, 1997.

Sartori S, Poirrier R: Seasonal affective syndrome and phototherapy: Theoretical concepts and clinical applications. Encephale *22:*7, 1996.

Shelton RC, Loosen PT: Sleep deprivation accelerates the response to nortriptyline. Prog Neuropsychopharmacol Biol Psychiatry *17:*113, 1993.

Teicher MH, Glod CA, Oren DA, Schwartz RJ, Luetke C, Brown C, Rosenthal NE: The phototherapy light visor: More to it than meets the eye. Am J Psychiatry *152:*1197, 1995.

CHAPTER 32. CHILD PSYCHIATRY

32.1 INTRODUCTION AND OVERVIEW

CAROLY S. PATAKI, M.D.

HISTORY

Over the last few decades, child and adolescent psychiatry has emerged as an academic discipline that covers a wide range of related topics from the development of the cortex, to the public health implications of childhood disorders. Child psychiatry began with the child guidance service delivery system. The goals of this system were to evaluate and provide psychiatric therapeutic interventions to children with psychiatric disorders. Although no psychiatrists were formally trained in child and adolescent psychiatry, a multidisciplinary team performed the evaluations and provided recommendations. That system developed from the historic Commonwealth Child Guidance Clinics of the 1920s, which endeavored to provide services to emotionally disturbed children and their families. The services rendered typically included family interventions by a social worker, some form of psychological testing administered by a psychologist, and psychodynamically based play therapy by a psychiatrist. Scientific study of childhood psychiatric disorders was not a part of the child guidance mission.

In 1946 the American Association of Psychiatric Clinics for Children was established, using the Commonwealth Child Guidance Clinics as a base for training. In 1953 an organization of medical practitioners was formed, the American Academy of Child Psychiatry. In 1959 the academy was legitimized as a medical subspecialty group with the establishment of board certification under the American Board of Psychiatry and Neurology.

EMPIRICAL MODEL FOR THE STUDY OF CHILD AND ADOLESCENT PSYCHOPATHOLOGY

Dennis Cantwell (1939–1997), was instrumental in the field of child and adolescent psychiatry by promoting an empirical model for the study of psychopathological disorders of childhood and adolescence. This was a novel direction for the field to take, given its history and the tradition of a psychoanalytic framework. The starting point for the empirical method was the recognition and classification of clinical phenomenology of child and adolescent psychiatric disorders.

PHENOMENOLOGY AND CLASSIFICATION

The clinical phenomenology of a psychiatric disorder comprises the essential core features, associated features, and relevant subtypes of the disorder. Once the essential features of a psychiatric disorder are identified, its natural history can be followed over time. Given the overlap of behavioral and mood symptoms in a variety of different childhood and adolescent disorders, defining the boundary between one disorder and another can be challenging. Additionally, within the more severely psychiatrically impaired child and adolescent populations, comorbidity is the rule rather than the exception.

Both categorical and dimensional approaches are useful in describing psychiatric syndromes in childhood and adolescence. The *Diagnostic and Statistical Manual of Mental Disorders* (DSM) uses categorical criteria to identify the clusters of symptoms that present together in multiple cases. This basis of identifying the clinical phenomenology of psychiatric syndromes and classifying them into reliable disorders facilitates meaningful investigation of many areas of child and adolescent psychiatry.

DEVELOPMENTAL PERSPECTIVE

Onset of a psychiatric disorder is a challenging concept in child and adolescent psychiatry, since it is difficult to pinpoint reliable landmarks for many of the disorders. Defining onset of a disorder by a single symptom might lead to inclusion of a heterogeneous group of disorders, whereas inclusion of only a rigid set of symptoms may exclude important subtypes. Age at onset of a disorder is an important scientific variable because it can give important etiological clues. Thus, age of onset can differentiate genetically determined disorders from those occurring sporadically and can separate cases of the same disorder that result from different genes. Earlier age of onset of some disorders (e.g., depressive disorders) have been associated with increased probability of depressive disorders in relatives.

Clinical Markers of Psychiatric Disorders Investigators are faced with the problem of finding reliable clinical indicators of disorders, while keeping in mind that the clinical emergence of symptoms is not necessarily correlated with the beginning of the pathological process.

Onset The investigation of age at onset of a disorder leads into studies of models of delayed onset of illnesses. Both developmental and genetic mechanisms of delayed onset of disease processes within the human body are applicable to psychiatric diseases. The following examples of delayed onset of disease shed light on potential future studies of the psychiatric disorders. Degeneration is a model of delayed onset of disease that characterizes diseases such as Alzheimer's disease and Parkinson's disease; that is, gradual loss of target neurons over time results in a predictable behavioral and neurological deficit. Developmental failure (i.e., a lack of physiological development within the nervous system at the appropriate time), is another model of disease process. Although most of the pervasive developmental disorders of childhood do not appear to have a delayed onset, a deficit of brain function and aberrant brain function seem to characterize these disorders.

Vulnerability Another theory of delayed onset of a disorder, the two-hit model, was first proposed in relation to the genetics of retinoblastoma. Essentially, the emergence disease is related to a genetic vulnerability triggered by exposure to a second event that might be environmental. For example, the rate of bipolar disorders increases markedly during adolescence, is maximal in young adulthood, and then gradually diminishes. One model to explain the emergence of a disorder with this distribution is a genetic predisposition (first hit) for the disorder and a second exposure, such as an adverse life event (second hit) that results in the emergence of the disorder.

Genetic anticipation is defined as the emergence of more-severe forms of a genetic disorder in successive generations. This mechanism has been shown to contribute to the severity of pathology in a number of genetic disorders that are caused by mutations of deoxyribonucleic acid (DNA) triplet repeat nucleotide sequences. Increases in repeat patterns of three nucleotides, (e.g., GACGACGAC) increase the length of DNA sequences and result in pathological genes. Triplet repeat expansion has been shown to be the mechanism of a number of genetic disorders including Huntington's disease and fragile X syndrome. Anticipation has been suggested as a possible explanation for increasingly severe psychopathology in successive generations in psychiatric disorders. Future investigation of possible molecular bases of anticipation in psychiatric disorders, such as the increased density of genetic triplet repeat expansions shown in Huntington's disease and fragile X mental retardation, is on the horizon.

NEUROBIOLOGY

The early development of the brain is an area of research that is of increasing interest to child psychiatrists. Developmental biologists have identified a critical period for the acquisition of various cognitive and motor skills. Environmental factors including exposure to alcohol, drugs, or infections may alter brain development during critical periods of nervous system growth and development and thus change later skills potentials. For example, differentiation of neurons and their migration to their proper places are critical for making necessary synaptic connections. Recent research on the acquisition of language has shown that early neuronal activity can influence the organization of the brain. Functional magnetic resonance imaging scans have determined the spatial relationship of language centers in children who have first and second languages. Children who learn first and second languages early in life have evidence of changes in both language centers in the same cortical region, whereas when a second language is acquired in adult life, the new language center is not represented in the same cortical region as the first language. Studies such as this are crucial for ultimate understanding of both cognitive and behavioral function and dysfunction. Moreover, an understanding of the critical

periods of brain development may drastically alter our understanding of early childhood behavior and its influence on later function.

In addition to the timely birth and migration of neurons to their final cortical destinations, interneuronal synaptic connections must develop for the acquisition of important skills. Finally, during development multiple neurons will die, and the brain will go on to function quite well. These various maturational steps are believed to be mediated by growth factors that send signals from the surface of a neuron into a cell.

Advances in genetics and neuroscience have opened up doors within child and adolescent psychiatry to the understanding of single-gene disorders, such as fragile X syndrome and more complex neuropsychiatric phenotypes. Disruptions of brain development may account for some of the more common disorders seen by child and adolescent psychiatrists, such as dyslexia. Overall, the broader our understanding of neurobiology and neuroscience of the brain, the greater our chance of comprehending the complex behavior of children and adolescents.

RISK AND PROTECTIVE FACTORS

Risk factors include all variables that increase the probability that a given child or adolescent will develop psychopathology. Protective factors decrease the risk of developing psychopathology. Demographics, psychosocial factors, biological factors, genetics, family environment, and external environment interact to produce risk and protection from psychiatric disorders. Rarely can a single risk factor account for the entire variance between the emergence or inhibition of a psychiatric disorder. Thus, the study of psychiatric risk is complicated and multifactorial. Risk factors often studied for influence on the emergence of psychiatric disorders among children and adolescents include family history of psychiatric disorders, social class, intellectual function, adverse life events, temperamental factors, and the quality of family and peer relationships. Developmental delays, child maltreatment, and high levels of family conflict increase the risk of psychiatric disorder. Protective factors that have been identified and studied include individual temperamental predisposition, positive family relationship, and other attributes of the external environment.

TYPES OF INTERVENTIONS

Interventions that can be investigated for efficacy include clinical interventions, targeted interventions, and universal interventions. Intervention studies may be the most important component of an empirical model of study within child and adolescent psychiatry.

Clinical interventions are those in which a family with a child or adolescent who has been identified as having a psychiatric problem seeks treatment. Psychosocial, psychopharmacological, and other environmental interventions can be compared with a placebo condition and with each other. Such efficacy studies are necessary to substantiate the benefit of the treatments. Because close to one half of families who begin psychiatric treatment terminate prematurely, the nature and delivery of the treatments also merit investigation.

Targeted interventions are those designed for children who have been identified as having an increased risk of a psychiatric disorder but whose families are not seeking treatment. Children can be identified either by an external factor such as a family characteristic (e.g., drug-addicted parent, family receiving state assistance) or through a behavior (e.g., defiance or aggression in the classroom). Thus, a child can be monitored or receive some psychosocial intervention on the basis of the risk factor, in some cases correlated with an already existing psychiatric disorder and in other cases preempting the emergence of a psychiatric disorder. The difficulty with these interventions is that sometimes interventions may not be wanted by the families or the threshold for identification is not correlated with a clinically relevant behavior.

Universal interventions are received by all children and families within a particular geographical distribution. They may occur throughout a targeted school or community or on a citywide, statewide, or national basis. This prevention strategy, also termed *primary prevention*, obviously reaches more children than any other method of intervention, but questions arise about the benefit and need of the intervention for the population at large as well as the cost-benefit ratio.

Much study of the risk factors and etiology of psychopathology is needed to develop the most effective combination of prevention strategies for children and adolescents. Interventions in child and adolescent psychiatric disorders will continue to be influenced by genetics, family environment, adverse life events, biological factors, and behavioral genetic studies. The goals within child and adolescent psychiatry are to diminish risk factors and enhance protective factors to prevent the emergence of psychiatric disorders in this population.

SUGGESTED CROSS-REFERENCES

In-depth reviews of the following areas of child and adolescent psychiatry can be found in Chapter 32, Chapter 33, Chapter 34, Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39, Chapter 40, Chapter 41, Chapter 42, Chapter 43, Chapter 44, Chapter 45, Chapter 46, Chapter 47, Chapter 48 and Chapter 49. Chapter 32 details normal child and adolescent development; Chapter 33 discusses the psychiatric examination of the infant, child, and adolescent. Chapter 34 covers mental retardation; Chapter 35 ,Chapter 36 and Chapter 37 detail learning disorders, motor skills disorders, and communication disorders. Chapter 38 reviews the pervasive developmental disorders; Chapter 39, the attention-deficit disorders; and Chapter 40, the disruptive behavior disorders. Chapter 41 covers feeding and eating disorders of infancy and early childhood. Chapter 42 covers tic disorders; Chapter 43, elimination disorders; Chapter 44 details stereotypic movement disorders and other disorders of infancy, childhood, and adolescence and reactive attachment disorder of infancy or early childhood. Chapter 45 discusses mood disorders; Chapter 46 reviews anxiety disorders of childhood, including separation anxiety disorder, obsessive-compulsive disorder, and selective mutism. Chapter 47 discusses schizophrenia with childhood onset. Chapter 48 details psychiatric therapies, including individual psychotherapy, short-term pshildhood, including separation anxiety disorder, obsessive-compulsive disorder, and selective mutism. Chapter 47 discusses schizophrenia with childhood onset. Chapter 48 details psychiatric therapies, including individual psychotherapy, short-term psychotherapy, cognitive and behavioral therapies, group psychotherapy, pharmacotherapy, residential and inpatient treatment, community-based treatments, partial hospital and ambulatory services, and psychiatric treatment of adolescents. Chapter 49 includes special areas of interest within child and adolescent psychiatry including psychiatric aspects of day care; adoption; foster care; physical abuse, sexual abuse and neglect of children; children's reaction to illness, hospitalization, and surgery; psychiatric sequelae of human immunodeficiency virus (HIV) and acquired immune deficiency syndrome (AIDS); childhood or antisocial behavior; borderline intellectual function and academic problems; dissociative disorders; posttraumatic stress disorder; gender identity and sexual issues; identity problem and borderline disorders; adolescent substance abuse; forensic child psychiatry; ethical issues in child psychiatry; school consultation; and psychiatric prevention in children.

SECTION REFERENCES

*Harmon RJ, Morgan GA: Clinicians' guide to research methods and statistics: Series preface. J Am Acad Child Adolesc Psychiatry *38:*99, 1999.

*Hockfield S, Lombroso P: Development of the cerebral cortex, IX. Cortical development and experience, I. J Am Acad Child Adolesc Psychiatry *37:*992, 1998.

*Husain A, Cantwell DP: *Fundamentals of Child and Adolescent Psychopathology*. American Psychiatric Press, Washington, DC, 1991.

*McMahon FI, DePaulo JR: Genetics and age of onset. In *Mood Disorders Across the Life Span*, KI Shulman, M Tohen, SP Kutcher, editors. Wiley, New York, 1996.

*Offord DR, Kraemer HC, Kazdin A, Jensen PS, Harrington R: Lowering the burden of suffering from child psychiatric disorder: Trade-offs among clinical, targeted, and universal interventions. J Am Acad Child Adolesc Psychiatry *37:*686, 1998.

*Vaccarino F, Lombroso P: Development of the cerebral cortex, VII. Growth factors, II. J Am Acad Child Adolesc Psychiatry *37:*789, 1998.

32.2 NORMAL CHILD DEVELOPMENT

MAUREEN FULCHIERO GORDON, M.D.

A growing, multidisciplinary body of developmental information encourages contemporary psychiatry to evaluate child development and its principles from an integrative perspective. Such an approach emphasizes certain essential elements of development: its inherent genetic predispositions, its interactive qualities, its repetitive patterns, and its transactional processes. An integrative approach considers such concepts as the “tutoring” of development. It focuses upon newer areas of research, particularly those involving temperamental and moral development, and stresses the independence and interdependence of separate areas of growth. The concept of normality, as it relates to development, frames the dynamic process.

Just as the developmental process necessarily incorporates change and continuity, an integrative approach to the study of development does not diminish any prior knowledge of child development. It allows instead an opportunity for the integration of the various biological, psychoanalytic, learning, psychosocial, and interactive theories into a coherent view of a growing and developing child. This approach is especially important for psychiatrists who are trying to understand the period from conception to the end of middle childhood, when development is the most substantial, dramatic, and fundamentally important. An integrative approach perceives infants and children as active participants in their move toward competence. It is a dynamic model of growth that is multidimensional and constantly evolving (Fig. 32.2-1). It incorporates the most recent biotechnical evidence to offer specific mechanisms that confirm development's essential interactive nature.

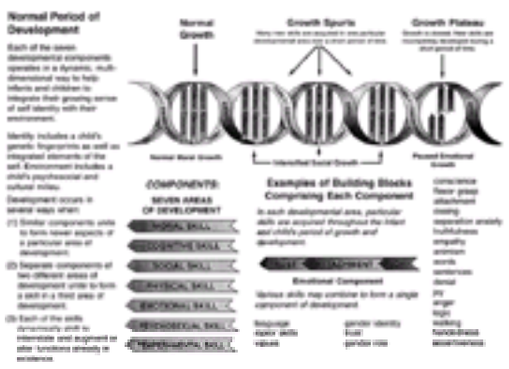

FIGURE 32.2-1 A transactional model of normal child development. The transactional model is aimed at helping to demonstrate the dynamic nature of the developmental process. This model, inspired by DNA's structure, describes an integrated approach to the normal development of infants and children. It represents the essential way in which a child's identity interacts with the child's environment to support the developmental process. In this view, identity represents a child's genetic prints as well as components that compose a child's sense of identity. Family environment represents the psychosocial milieu that surrounds the child. Development occurs when various components of behavior, supported by the child's identity and milieu, come together to form a completed set of maturational skills in each of the seven developmental areas. In contrast to the more traditional linear or branching model, this interactive schematic helps one see how development in one area may continually influence past or future growth in each of the other separate developmental lines. It demonstrates the active, dynamic independence and interdependence of development as well as its interactive nature. (© 1997, M. Gordon, M.D.)

SIGNIFICANT THEMES UNDERLYING CHILD DEVELOPMENT

The various theories contain certain important themes that help psychiatrists understand the complexity and consistency of children's growing experiences. These principles offer explanations for the transformation of simple collections of cells into 11-year-old children who can respond and interact with their world. There are at least four recurring and significant themes.

Development as Inherent in Nature Genetic predispositions are the foundation for any consideration of development's inherent nature. Developmentalists refer to this internally generated, exquisitely programed part of a child's development as organismic growth. Genetics dictates species-specific developmental potentials. Children, however, have their own special, genetically motivated patterns and ranges that shape their individual differences.

Only recently have developmentalists identified genes that are associated with behavioral traits and syndromes. Replacing the imprecision of genetic inferences—as obtained from twin studies—with the clarity of information derived from specifically identified genes is just one example of the enormous potential contribution of this new information to the study of child development. It has been suggested that contemporary developmentalists will have to learn to use newly found deoxyribonucleic acid (DNA) markers to answer questions about developmental interactions between nature and nurture, including psychosocial and physical environmental risks and influences, comorbidities, and genetically influenced, revised concepts of normality.

Development as Transactional in Nature The transactional quality of development, which is an interplay of nature and nurture, forms the basis of most current thinking and research. Mechanistic factors from conception on are quickly interwoven with the child's environmental and behavioral circumstances to alter an infant's growth patterns. The practice of tutoring is a transactionally derived principle. Developmental tutoring recognizes that children interact with their environment even before birth in ways that significantly stimulate or inhibit potential. A mother's ingested folic acid helps her fetus avoid the risk of spina bifida. A carefully designed mobile lets an infant engage in activities that are visually and auditorily stimulating. Jungle gyms stimulate physical development and coordination. Play dates allow an 8-year-old child to advance social and moral skills.

Transactional research and views have opened up new considerations for child developmentalists. The developmental entity, temperament, was differentiated as the result of observations made during transactionally oriented studies called longitudinal studies. The practice of interconnecting various theories has promoted a new emphasis to be placed upon other developmental areas, such as moral development, which is significantly influenced by cognitive, social, and gender issues.

Normality in Development The concept of normalcy is often implied in development; yet, as used in this arena, it has different connotations. *Normality* may mean freedom from pathology. In this context (but not without opposition), labeling a child as normal with respect to developmental maturity is a common practice. Normality may simply refer to statistics or it may reflect what is acceptable culturally. In the near future, normality may involve a comparison of genetic profiles. In transactional theory, the cultural concept, once having limited validity because of its changing norms, has taken on new importance. In child psychiatry *normality* conveys a continuum with various periods of rapid change. This concept of normalcy is important for understanding a critical, transactional principle: a child's development operates as a two-way street of regression and progression (Fig. 32.2-2). Typically occurring before growth spurts, environmental or internal changes can create a stress that causes developmental skills in a particular area to be lost. This phenomenon is called normal regression. Just before an infant makes an exciting quantum leap in growth, for example, it is common for parents to describe their child as particularly irritable or cranky. From a transactional viewpoint normal regression does not impede progression in other areas. A child can continue working socially on peer relations, for example, while regressing cognitively by resisting school assignments. This is in contradistinction to abnormal, or neurotic, regression. Neurotic regression creates a self-perpetuating psychosocial isolation—a break between individuals and their environments—that impedes further development in other areas. A neglected 5-year-old who is emotionally regressed will have trouble developing social empathy, applying temperamentally assertive skills, and cognitively refining problem-solving tools.

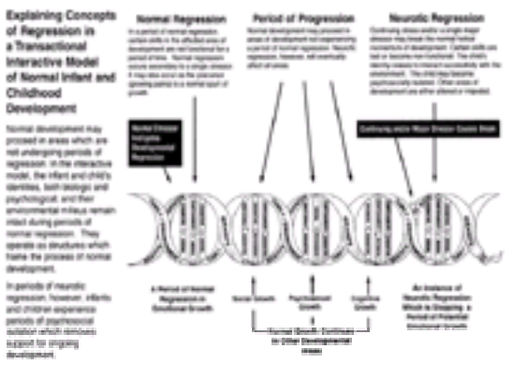

FIGURE 32.2-2 A transactional model of regression in normal infant and child development. Diagrammatic demonstration of how a transactional view of normal development can explain the concepts of normal regression, progression, and neurotic regression in normal child development. A transactional, dynamic model is used to explain how normal regression may occur yet not inhibit further development. It shows the sequence of normal development, called progression, and demonstrates how continued stress may actually separate the two supports a child must maintain to sustain and continue development. When enough stress occurs, psychosocial isolation results. In the process, children (identity) become separated from their psychosocial milieu (family-environment), *neurotic regression*. The model shows how disruptive this type of regression (as opposed to normal regression) can be to the entire developmental process.

Distinct Areas of Development To study child development, growth is often differentiated into distinct developmental areas. Each area is considered to mature independently, yet they interact in ways that ultimately change the outcome of each separate event. Developmentalists term this the *interdependence and independence of development*. Normal healthy development exists when there is significant harmony between both the level attained and the separate areas or lines. When infants and children do not mature in quantitative ways in each or all of these areas they are said to have *developmental delays. Developmental disorders*, on the other hand, involve behavior that differs qualitatively from the norm. These disorders are differentiated from *discontinuous development*, which refers to new ways of reacting at particular periods called *stages*. Child development has seven significant areas to be considered: physical, temperamental, cognitive, social, emotional, moral, and psychosexual.

THEORIES OF DEVELOPMENT

Theories help psychiatrists understand how a single cell grows into a complex child who has all the skills necessary to enter adolescence. By incorporating major developmental themes, theories help explain principles of development and create guidelines for translating concepts into practical applications. Theories help psychiatrists predict behavior in children and often explain the significance of their observations. Psychiatrists can test hypotheses that pinpoint factors that influence development. Understanding the significant theories in each of the seven major developmental areas is a valuable tool for the psychiatrist attempting to integrate the wealth of developmental knowledge of infants and children.

Physical Development of Infants and Children Human physical development is directed at helping infants and children perceive and negotiate with their external world. During infancy and early childhood there is enormous physical growth, including changes in overall body growth, size, shape, body composition, and structures. By the time children end the period of infancy, at about 2 years of age, their heights have reached half of their adult height potential; their weight has quadrupled; and they have acquired skills that help them to function in their external worlds. Especially important, however, is the astounding maturation that occurs in the central nervous system (CNS). This development allows children to acquire several paramount skills: motor abilities, perceptual abilities and pathways, and language (Table 32.2-1). This particular part of physical development uses much of the developmental energy of infants and smaller children. Most often the play of infants and young children involves practicing new physical skills or demonstrating competence with others.

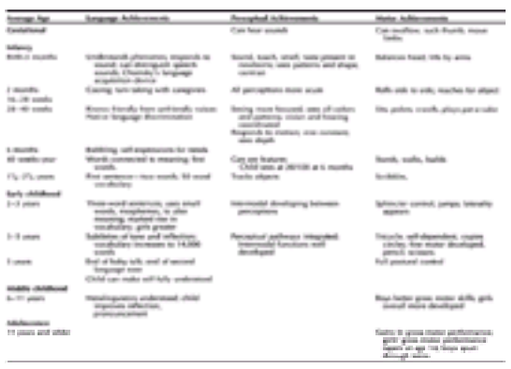

Table 32.2-1 Physical Development—Maturation of the Central Nervous System: Comparing Language, Perception, and Motor Skills

As with other areas of development, physical development is motivated by multiple pushes toward growth. Biological influences, for example, are mediated by the body's hormones: the pituitary monitors hormonal effects upon all tissues; sex hormones regulate sexual maturational changes; thyroxine influences brain and overall body development; adrenaline and neurotransmitters alter CNS maturation in particular and may influence overall development in general. This significant process begins long before birth. By the end of the first intrauterine month, a fetus has grown up to 10,000 times its size at conception and has a heartbeat and the beginnings of several organs. By the second month, a fetus has all its organs. By the fourth month it has human qualities, including the ability to breathe, suck, squint, hear, swallow, urinate, and move.

Even at this earliest of human stages, environmental factors interact with the organic to serve as tutors of growth. Maternal drug and alcohol use and maternal emotional issues, including stress, significantly affect outcome for the developing fetus. Cocaine, for example, increases norepinephrine and serotonin concentrations at the cellular level. In recent survey, newborns whose mothers had untreated prenatal depression showed lower birth weights and less alert behavior than newborns whose mothers had been treated with selective serotonin reuptake inhibitors (SSRIs). Proper nutrition, appropriate caregiving, and adequate stimulation allow maximal physical growth; illness, malnutrition, inadequate stimulation, and emotional deprivation can inhibit growth potential.

Sociological trends also alter physical growth patterns between generations and geographical regions. In economically deprived nations children often grow more slowly and reach maturity at much lower heights than their predecessors. In the United States, on the other hand, California children are rated on height scales that are calibrated 2 inches above the norm.

The enormous physical growth in infancy and childhood helps to magnify growth patterns, some of which are not exclusive to this area. Development is not a simple linear process. Growth in all areas occurs in spurts, which are followed by plateaus of much slower development. In one study, infants grew up to ½ inch a day and then did not grow taller for as long as 5 weeks. In older children, the plateau between spurts widens, but the pattern remains the same. Growth spurts are characteristically preceded by a regressive phenomenon, colloquially termed *growing pains*.

Unique to physical growth, however, is a pattern of universal direction, both a *cephalocaudad direction*—an infant's head is largest at birth and gradually becomes smaller in relation to the extremities and trunk and a *proximodistal direction*—growth extends from the center of the body and moves outward to the extremities, the lower extremities, and then the hands. These directions continue and reverse only when adolescence has begun. Development is also affected by sex. The growth rate for males lags behind that of females until both reach maturity. Girls between the ages of 3 and 10 typically spurt physically 6 months to 1 year ahead of boys. Even in the fetal stage, gender influences the rate of growth. By 20 weeks following conception, male fetuses are already 2 weeks behind female fetuses in development. Growth in children physically is *asynchronous*—various parts of the body spurt at different rates and at different times. In infants, for example, heads

and chests mature much more rapidly than the extremities. The brain, sexual organs, bladder, lymphatic organs, and individual tissues differ in both times of peak growth as well as appearance of growth spurts. Body composition changes as well. The fat composition of the body, for example, increases until the tenth month and then diminishes until adolescence.

Recording Physical Growth—Growth Curves Versus Milestones The various ways in which growth is recorded reflects physical development's patterns and principles. *Distance curves* are used to monitor an individual child's annual physical progress by charting a child's specific height and weight and comparing them with those of a sample of children at each chronological age. Results on such curves demonstrate, for example, the difference between growth patterns of boys and girls. *Velocity curves*, on the other hand, show the average annual growth for an infant or child. Data from velocity curves show growth spurts for an individual child. Milestones are based upon comparisons between children and convey a normal range of expectancy for acquiring various skills. One commonly used milestone-based tool is the Denver Developmental Screening Test devised to screen children from 1 month to 6 years of age in selected areas of development, such as language. Another measuring instrument popular with child psychiatrists and pediatricians is Berry Brazelton's Neonatal Behavioral Assessment Scale.

Physical Development of the Central Nervous System In infants and children, CNS maturation allows the development of a child's motor skills, perceptual patterns, and language. Thus, children eventually master their physical environment; take in essential information; and communicate their needs, wants, and feelings to those with whom they relate. The importance of CNS development makes it worthwhile to look closely at various theories that describe its principles.

Biological theorists describe the development of the CNS as a gene-regulated, orderly appearance of certain morphological structures, brain development, and synaptic maturation. They emphasize milestones for rating the timing of cell proliferation, differentiation, and maturation. One of their major contributions to understanding children's growth is the introduction of important terminology. They describe *critical periods*, which refer to increased activity in certain parts of the system within a short, specified period of time. Such periods suggest windows of opportunity for maximal growth in that area as well as times of particular vulnerability. The latter developmental phenomena, called selective vulnerability, explain why certain areas are damaged by harmful agents at particular times. This principle is used to explain the appearance of certain pathology, such as perceptual problems. Rubella, for example, causes the greatest damage to the fetus' auditory and cardiac system if it is contracted by the mother prior to the eleventh week of pregnancy, when the most fundamental development in those particular organs is occurring. Critical periods are sometimes used in reference to other areas of development as well. Biologists also refer to *plasticity*, the concept that the brain can compensate, within limits, by forming alternative pathways when toxic events damage part of its functions.

Environmentally oriented developmentalists emphasize the importance of a second category of regulators that include externally influenced and dependent factors. Illness, toxins, X-rays, drugs, and even therapeutic interventions can alter brain structure and chemistry. This schema can explain normal alterations in development as well as neuropathology. External assaults are most damaging during critical vulnerable periods and are most effective in restoring function at those same times.

New and exciting biotechnical research, however, has discovered processes that integrate much of the information previously used to describe development from gestation through middle childhood. This work demonstrates that although the brain's full supply of neurons is present at birth, critical neural pathways that depend upon environmental stimulation for their maintenance and development—shape and reform the original morphology. From the third gestational week, when neural tube development enters a rapid phase of proliferation, an infant must interact with the environment to generate the forces that will cause hundreds of thousand of proliferating neurons to shape structures and pathways. Identification of the genes and the mechanisms that accomplish this bridge between nature and nurture is the first solid evidence for the validity of a transactional theory of development. Biotechnical sophistication has resulted in identifying over 50,000 genes that use environmental stimuli to reshape the nervous system. One group includes morphology-inducing genes, such as the “hedgehog gene,” which encode proteins that create interneurons responsible for neural structural changes. A second group of genes, which includes the cyclic adenosine monophosphate (cAMP) response-element binding protein (CREB), encode electrical stimuli that translate environmental responses into new synaptic pathways. Together, these groups of genes actually determine whether there is proliferation or destruction of neural synapses and pathways.

Practical implications of this mechanism for explaining child development are enormous. During the first 10 years of life a child's nervous system most actively interrelates with environmental stimuli through gene-induced neural synapses. The first 3 years, in particular, when the bulk of neural synapses are formed, are crucial. At age 2, a child has twice the number of adult synapses. A child's brain, which begins to develop during the first gestational month, becomes proportionately the largest organ in the child's body, as it reaches 95 percent of its adult size by age 5. After age 10 the brain becomes actively involved in destroying weakly formed synapses. Such findings reinforce the critical importance of early childhood development. Peak times of synapse formation for each skill help explain the mechanism of critical periods. Biotechnically explained mechanisms shed new light on the ways in which environmental factors, including “tutoring,” can affect development of the CNS. During the period of most active synapse formation, a child's repeated experiences stimulate neural pathways. Such pathways interact with genetically controlled enzymes and other proteins to create neural synapses. The synapses in turn actually form parts of the brain responsible for special functions: motor development, perceptual development, and language. Understimulated pathways atrophy the parts of the brain that correspond to unused pathways. Newer and safer techniques for evaluating brain morphology, such as magnetic resonance imagery (MRI), delineate gross morphological changes that may parallel the synaptic changes. For example, MRI studies verify that between the ages of 3 and 18 years, the corpus callosum—composed of white matter or myelinated matter—increases approximately 1.8 percent in size while the brain's gray matter decreases.

CNS MATURATION AND MOTORIC GROWTH Motor activity, which allows children to begin exerting control over their environment, results from similar, specific patterns of brain synaptic maturation. By the time children reach age 2, they have trillions of new synapses available for their motor work. Children engaged in physical play are actually stimulating structures in the brain that are responsible for motor balance and coordination. Simple actions, such as a baby's reaching for a rattle, activate the vast motor-neural network, and the brain in turn responds to such activity. The cerebellum, a major part of the motoric brain, reacts by myelinating to its full maturity by the time a child is 4 years old. It is no coincidence that gross motor skills peak in 4-year-old children.

In general, motor development in children is extremely important because it allows them to acquire two of the most fundamentally important functional human skills: locomotion and handedness. *Locomotion*, the ability to navigate the environment, progresses as synaptic connections grow through stimulation initiated by children's repeated motor behaviors. In this way, infants progress from their struggle with simple movements (e.g., chin lifting at 20 weeks) to full postural control at 5 years (Table 32.2-1). In the same way, *handedness*, the ability to reach and grasp objects, begins as a simple reflex at age 1 month, progresses to a full grasp response at 1 year, and at 3 years shifts from gross to fine motor control that shows high dexterity. As part of this development, laterality also makes its appearance. *Laterality*, the asymmetric functioning of a paired faculty, is observable in a child's preference in using each of several organs: eyes, feet, visual fields, and hands.

CNS MATURATION AND PERCEPTUAL DEVELOPMENT To process their world, infants and children must develop their perceptions. Theorists who endorse nature over nurture are encouraged by the presence of well-developed perceptions, except for vision, in newborns. Even in utero, a fetus can hear sounds, and 3-day-old infants can recognize their mother's voice. By age 2, auditory perception has reached adult levels. Taste and odor are well perceived even by newborns, who show a preference for sweet milk and can identify their mothers by odor. Newborns have well-developed sense of touch and can discriminate between reassuring and threatening sensations.

Visual perception, however, is not well developed at birth and confirms the developmental principle that nature and nurture must operate together for optimal maturation. At birth an infant's visual acuity is only 20/800, but at 8 months it can range from 20/200 to 20/70. In the 6 months after birth, an infant's visual cortex changes enormously. As many as 18,000 new visual synapses can potentially form in that time. The number of synapses is predicated not only on genetic priming of electrical impulses, but also on environmental visual feedback that stimulates certain synapses and prevents the atrophying of others. Functionally, the new synapses are responsible for the appearance of depth perception, better visual acuity, perceptual organization, and eye-hand coordination. In addition, visual acuity is significantly affected by environmental factors such as diet. In one recent study prompted by current cultural concerns with fat content of foods, researchers determined that infants who were breast-fed milk, which contains higher levels of certain fatty acids, had acquired better visual acuity at both 3 months and 6 months of age than those given milk with lower fatty acid compositions (i.e., commercial formula milk or evaporated milk).

As a child matures, perceptions in all areas become intermodal or interrelated. By age 6, perceptions must continue to be differentiated and interrelated so that a child can perform tasks expected in the academic environment. Perceptual difficulty is often diagnosed when tasks require interrelated perceptions. According to biotechnical theories of synaptic peak formation, however, it would be valuable if such perceptual intermodal difficulties, including dyslexias, could be identified much earlier—at the peak of perceptual formation at age 3. For example, children treated for strabismus before the age of 2 have no permanent binocular visual damage (Table 32.2-1).

CNS MATURATION AND THE ACQUISITION OF LANGUAGE Developmentally, *language* refers to the ability to communicate as well as the capacity to use symbolic thought in the speech process. *Speech* refers to the ability to produce sounds and words. In addition to sound, language involves syntax (the rules of language) and semantics (the vocabulary). Developmental theorists generally concede that language maturation proceeds along a schema with a particular set of rules. Infants' prelinguistic development starts as early as a few hours after birth. An infant's language at age 18 months reflects a tendency to rely upon sound, a

telegraphic style of omitting insignificant words, and a significant development of receptive, as opposed to expressive, language. As children mature into early school age and middle childhood, language develops with less reliance upon sound and more emphasis upon semantics, grammar, and an understanding of metalinguistics, that is, the ability to perceive language as a system apart from themselves. The importance of this area of maturation to a child's overall development cannot be overstressed. In many studies, infants who are blind, for example, have demonstrated the auditory ability to recognize their mothers' voices, but suffer from lacking eye-to-eye contact. On follow-up, severe language difficulties, and serious social and emotional delays were found in the same infants (Table 32.2-1).

In language acquisition, as in much of development, widely divergent theories have made valuable contributions to understanding the process of growth. Noam Chomsky, for example, advocates a nativist language theory. He uses his observation that children create self-coined words to prove the presence of a *language acquisition device* that innately permits children to speak a universally grammatically correct language. Behaviorists, on the other hand, suggest that language is learned by association. They note that 2- to 3-month-old infants shape their localizations in response to their parents' special ways of speaking to them, a language termed *parentese*.

A biotechnically proven, transactional approach to language development, however, seems to be the only explanation that accounts for both overall patterns and individual differences. The complex neural synaptic connections between genetically regulated brain formations controlling language, and neural connections peripherally, explain language acquisition as well as its sensitive periods of acquisition. Infants and children have the most-sensitive, receptive language skills and ability to recognize speech until neural language synapses slow their growth markedly at ages 5 to 6. Between the ages of 1½ to 6 years, children learn approximately five to nine new words daily. After age 6, as synaptic formation rate slows, children have more difficulty learning a new language. Although children older than 6 years of age have more difficulty learning a new language, they remain equally responsive to the acquisition of new vocabulary at any age. As with other areas of development, children play active roles in language acquisition, especially during this peak period of neuronal stimulation and synapse formation. Children stimulate new language pathways by several widely used vocabulary-building strategies: inferring the meaning of a new word by its context, deriving the meaning of a novel word by comparing it with a familiar verb in the same sentence, and retrieving the meaning when presented with the same word a second time.

Temperamental Development Temperamental development involves the maturation of traits that allow children to respond to new situations, tasks, and opportunities. Parents frequently comment colloquially upon a child's temperament: “Where did this child of mine come from?” Usually such comments stem from parents' acknowledgments that they do not recognize a particular child's traits as familiar to their own experiences.

Principles of Temperamental Development Each child is born with specific temperamental patterns of behavioral response. These traits can be discerned throughout the child's life span. As with other developmental areas, temperamental traits can be stimulated or inhibited by external factors, including family interactions, cultural and environmental influences, and social interactions. Using inhibition as a focus for distinguishing temperament, Jerome Kagan shows that certain children are irritable and fearful in new situations regardless of a parent's sensitivity and support. Kagan's work also demonstrates the internal consistency and continuity of temperament up to a point. On follow-up, temperament remained the same until a child reached 5½ years. Sometime near the 7½ year mark however, environment begins to shape such traits. Unlike other areas of development, temperamental traits have especially strong negative and positive connotations, which magnify the influence of sociocultural influences upon a child's growing experience. Temperamental development can significantly affect other areas of growth. In the parent-child relationship, temperamental traits are particularly relevant for emotional growth. The compatibility of traits between children and their parents is termed a *goodness of fit*. A good fit results in parents behaving with more sensitivity and positive interactions with their children.

Unlike other areas of development, temperament has been considered a separate area of study only since the work of Stella Chess and Alexander Thomas in the 1970s. In the New York Longitudinal Study, Chess and Thomas found that temperamental traits in children could be identified from infancy and at least through adolescence. The results strongly indicate that temperament has a genetic base but that such traits are significantly altered by environmental factors, including family and sociocultural situations. Successful attempts to identify specific genetic markers involving temperament have bolstered both the genetic foundations as well as transactional aspects of temperamental development. In two separate studies, certain alleles of *DRD4*, a dopamine receptor gene, and the temperamental trait labeled novelty seeking, have shown strong association. In addition, the same *DRD4* alleles frequently occur in those with attention-deficit/hyperactivity disorder, which reinforces the transaction concept of comorbidity.

Assessment of Temperament The relative newness of this area of study results in the lack of uniformity among those who describe temperament. Chess and Thomas for example, categorize temperament into three general clusters based upon a continuum of ease with which a child approaches the environment. The “easy child” typifies 40 percent of the population; the “difficult child,” 10 percent of the population; and the “slow-to-warm-up child,” 15 percent of the population. The remaining 35 percent of children do not fall into any particular grouping. Within these broad categories are nine specifically described traits: activity level, rhythmicity, approach/withdrawal, adaptability to change, threshold of sensitivity response, mood, emotional expression, attention span, and distractibility. The Carey Infant Temperament Questionnaire, based upon the Chess and Thomas model, uses family interviews to rate temperament for 4- to 8-month-olds and for 1- to 3-year olds.

The Clinical Assessment of Temperamental Development focuses on a task-orientation approach that incorporates the negative and positive aspects of temperamental traits. In this system, eight temperament traits and their opposites are listed and categorized by their presentation at significant chronological ages and in each gender: goal-directed and non-goal-directed temperament, deliberate or impulsive; active or passive, assertive or nonassertive, verbal or nonverbal, and receptive or rejecting. Information for assessing such traits comes from clinical observation of each child. A 3-year-old child who is observed to move into another child's space without permission, for example, is given a highly aggressive rating. A 5-year-old child who is easily distracted yet moves consistently toward a goal is rated impulsive but goal-directed (Table 32.2-2).

Thomas and Chess		Carey	
Dimension	Description	Trait	Description
Activity level	% of time spent in activities	Verbal/Nonverbal	Utterances, including words but not cognitive quality
Distractibility	Degree to which stimuli are allowed to alter behavior	Goal-directed/nongoals	Degree to which a child thinks in terms of goals
Adaptability	Ease moving into change	Assertive/nonassertive	Expresses effectively needs
Attention span	Amount of time spent on attending	Active/passive	Degree of movement
Intensity	Energy level	Aggressive/Nonaggressive	Ease of crossing into another's space without permission
Threshold of responsiveness	Intensity required for response	Deliberate/impulsive	Time spent doing tasks, not necessarily goal-directed
Quality of mood	Positive compared to amount negative behavior	Resistance/acceptance	Attitude toward change
Rhythmicity	Regulation of functions		
Approach/withdrawal	Response to new situations		

Table 32.2-2 Two Models of Temperament—Newborn to 6 Years

In addition, measurements of temperament have been attempted by drawing correlations between observable traits and laboratory measurements of physiological criteria. For example, 12-week-old infants can be assessed for positive and negative traits on the basis of cardiac vagal tone, which is measured as the amplitude of respiratory sinus arrhythmias. Infants who show higher baseline cardiac vagal tone rate higher on assessments of temperamentally negative traits; infants with low cardiac tone show a higher incidence of positive temperamental traits such as goal-directedness and ease in being soothed.

Practical Implication of Temperamental Development In general, traits that coincide with a child's chronologically determined developmental tasks are the most significant because they frequently determine how children are received and perceived. Even before birth, temperament influences a child's reception. Expectant mothers often relate to their unborn child's trait unwittingly. One mother describes a temperamentally active fetus: “He's going to be a football player! His kicks are so hard.” Another mother exclaims: “She's never going to let me rest.” Newborns who are assertive cry with energy and receive attention much more emotionally. Parents respond according to their own personalities (“My child is feisty,” “My child is never satisfied”). In early childhood, other traits come into play. A temperamentally verbal child often appears “bright” (“My child asks a lot of questions”). The cognitive maturity of a less-verbal child is frequently underestimated. As childhood progresses, the traits that determine school success predominate. A child who is temperamentally deliberate and goal-directed functions well in the traditional school setting. An impulsive, non-goal-directed child may be mislabeled as having an attention problem or characterized as unmotivated. Temperamental qualities may alter behavior in critical ways. Verbal temperaments in children are a major factor in recognizing the need for psychiatric treatment. As a child matures, temperamental qualities can have critical long-lasting consequences. Preadolescent suicidal children who have verbal temperaments, for example, have been shown to receive timely and more adequate psychiatric treatment for their serious problems than their temperamentally nonverbal counterparts, whose verbal reticence makes

recognition of their suicidality much less likely.

Cognitive Development Children must find ways to understand, to remember, to solve problems, and to organize their environment. Cognitive development in children includes all the processes that lead to maturation of their mental activity: a fetus associating sound with its mother; a newborn learning to suck with the help of a guiding hand; a toddler tinkering with a toy; a 5-year-old putting together a six-piece puzzle; a 10-year-old's questioning the reason for an illness. Explanations for the enormous range of cognitive maturation that can occur within a child's first decade are widely divergent. Two theories—those of Jean Piaget and Leon Vygotsky—present opposing views of cognitive maturation. A third theoretical framework is favored by many contemporary researchers, an information-processing framework. An adequate explanation for cognition incorporates all three theories (Table 32.2-3).

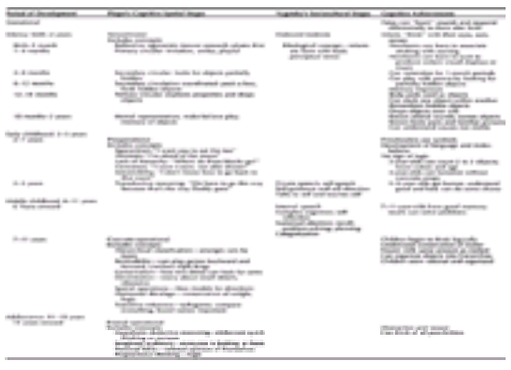

Table 32.2-3 Theories and Skills of Cognitive Development in Infants and Children

Biological Approach to Cognitive Development Piaget explains cognitive maturation by use of a biological model that focuses upon *schemes* (i.e., organized ways for children's developing brains to make sense of their experiences). To structure such schemes, children use two complementary processes: assimilation and accommodation. According to Piaget, cognitive growth occurs in stages, which are always in fixed order (invariant). Overall, cognitive maturity is defined by a child's increasingly refined ability to conceptualize space, both internally and externally. In the first stage, the *sensorimotor stage*, from birth to 2 years, infants use body senses and activity to explore their environment. In the process of exploration, infants learn to anticipate an experience by internally constructing a model of each of their own experiences. Cognitive growth results in forming new skills, such as object permanence (i.e., the ability to remember an object once it is out of sight). Children discover that objects can be assimilated into an experience they have already encountered: a top spins, a mobile makes music, apple juice tastes sweet, hot food burns. Infants begin to play or practice cognitive skills with games like peek-a-boo. Eventually, infants discover that each experience does not necessarily assimilate with the same function. By adjusting to this newer function, accommodating, infants broaden their cognitive development. Piaget asserts that infants and children have the cognitive challenge of internally constructing new schemes to fit old and new experiences into a universally uniform reality. Piaget's schemes are thought of in absolute terms. During the second stage, the preoperational phase, from ages 2 to 7, children become intuitive, anticipating experiences with consequences. Children in the preoperational phase think symbolically but illogically and with egocentricity and a distinct inability to perceive self from others in their environment. Preoperational processes combine to create a child's inner world populated with the creations of magical thinking. A stuffed kitty becomes a ferocious tiger, a darkened room is transformed into a dungeon, and the shadow of a swaying tree's bough conjures up monsters with unlimited powers. From ages 7 to 11, children enter a *concrete operational phase* that involves an ability to think logically and in an organized fashion. The magical woes of the younger child are replaced by a more realistic set of concerns that are stimulated by a new-found conception of cause and effect. Between the ages of 8 and 13, children normally spend some of their time worrying about such issues as school success, health and dying, and social relationships. At the end of this stage, at age 11, children develop the capacity to think abstractly. By adolescence, formal operations appear as abstract thought and the interesting imaginary-audience phenomenon, which joins egocentrism and the capacity to abstract to torment adolescents into thinking that the entire world can watch whatever they are doing in public. In the Piagetian schema, language becomes important as a tool only after a child is capable of hypothetical thought. However, some criticism of this theory of cognition exists. Many researchers who feel that certain social contexts obscure valid cognitive assessments argue that much younger children can display logic.

Sociocultural Schema for Cognitive Development In direct opposition to Piaget, Vygotsky uses a sociocultural framework to describe cognitive maturation. According to Vygotsky, an infant has a basic instinctive cognition that includes some basic memory and perceptual capacity. Until the age of 2, children simply develop this primitive cognition by interacting with their environment. After age 2, however, Vygotsky conceives cognitive maturity as children's acquisition of skills, which in turn mold their mental representations. Both tangible and intangible tools, from computers to social strategies, help children make the translation. For Vygotsky, however, language is the essential tool for cognitive maturation. Cognition is developed by children's social dialogues, (cooperative conversation with others), which are incorporated internally into private speech. Private speech, once internalized, is then used by children as a support, or scaffolding, to guide them through new learning. The motivation for cognitive development is created by a *zone of proximal development*, which is the distance between children's actual developmental levels and their potentials for tasks that can only be accomplished with adult help. Critics of Vygotsky's explanation for cognition argue that he eliminates the biological line of cognition. Others assert that language cannot be primary for cognition. Cognition develops in cultures and families that do not emphasize language. Children who can learn primarily by observation raise questions about Vygotsky's claims as well.

Information-Processing Model Explaining Cognitive Development A third model for understanding cognition relies upon a computer-like metaphor to explain cognitive maturation. According to this model, children use four strategies to develop cognition: taking in information from their environment, encoding the information into a symbolic mental system, recording it into a meaningful effective model for knowing their environment, and finally decoding it so that task solutions are possible. The processing strategies have been compared with software for the brain's hard drive system. This model is often used as a research tool for understanding cognition, memory, and perception. Short-term memory, which holds a certain amount of information until it is processed, is the center of this system. Detractors of this theory have argued that selective attention must be present and that it drives the cognitive process. They reason that prelinguistic children could not use this process for cognitive development because they do not have the required attention. Recent work, however, has confirmed that prelinguistic children do, indeed, have the capacity for selective attention as well as explicit memory. By using display training, researchers have demonstrated that not only can 3-month-old infants have their attention selectively directed toward objects, but such selection also increases their memories of these objects.

Possible Mechanism for an Integrated Cognitive Theory Over the last decade, many researchers have tended to agree that genetic influence is a major factor in dictating cognitive development; in fact, most researchers assert that over half of the cognitive ability within an individual may be determined by heredity. From birth, heredity exerts its influence upon cognitive development until it reaches its maximum effect during the midteens. Recent identification of the mechanism of operation of the *CREB* amplifier gene, however, has reinforced the concept that acquisition of cognition requires the interrelation of genetic, biological, and environmental factors. To assist in learning, the gene depends upon several elements: a child's repetition of cognitive tasks in the environment; interaction with certain neuronal synapses; and the individual child's DNA decoding. In particular, the *CREB* gene appears to be essential to develop pathways that allow learning and long-term memory—an element important to all the cognitive theories. In addition to the *CREB* gene, a single gene more closely connected to general cognitive ability has been identified; thus researchers can trace more directly the cognitive maturation process and observe the interplay between this specific gene and environmental factors.

Measurement of Cognitive Development Measuring cognition and intelligence have in the past been viewed as synonymous. Several contemporary developmentalists, however, have attempted to redefine the concept of intelligence in transactional terms—to include environmental factors and such other aspects of development as emotional and social maturation as part of a broader intelligence quotient. As the concepts are questioned, so are the measuring tools, especially standardized testing. Many clinicians have begun to rely upon clinical assessments based upon Piaget's work to assess cognition. Advantages include their ability to remove cultural influences; they can also adapt cognitive assessments to everyday situations that a child faces.

Social Development Once children have formed a sense of themselves, they can think about and interpret their experiences in other situations. The process by which children develop a sense of themselves and then relate that knowledge to their experience of others constitutes social development. Social maturation begins from birth when infants' cries tell others: “I'm hungry” or “I'm lonely.” It operates meaningfully in the small child's world; for example, when a 3-year-old gleefully shouts: “Throw the ball again, Daddy” or “I'm a great kicker!” Social maturation involves a shift away from adult orientation, when children depend upon adults for their needs, to a peer orientation in which children make friends their focus (e.g., “I don't want to go to the store with you, Mom. Can I take my friend Sarah with us at least?”).

There are four distinct phases in the social maturation process: a period of attachment, during which children develop a sense of self; a period of self-understanding

that allows children to analyze their own makeups; a period of perspective taking, during which children apply their self-knowledge to adults and peers; and a period of true friendships and peer orientation. Certain theories better explain social development at particular chronological stages. Attachment theories best describe the first phase of socialization—the development of self (Table 32.2-4).

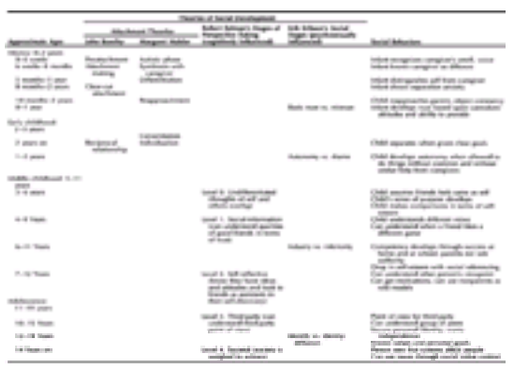

Table 32.2-4 Theories of Social Development and Skills in Infants and Children

Theories of Infant Socialization—Attachment Theories To begin to feel a sense of themselves, infants must first attach over a period of time to a parenting person who feeds, bathes, holds, and plays with them. Without this attachment, children have literally failed to thrive. More commonly, children with inadequate or faulty forms of attachment fail to develop the whole sense of themselves that is prerequisite to having good social relationships. Children of antisocial mothers who cannot empathize with their children's autonomy tend to form avoidant styles that persist as poor adult relationships in adulthood. John Bowlby, who coined the word “attachment,” describes the process in four phases. A baby experiences a preattachment phase during the first 6 weeks of life. With appropriate caregiving, infants are able to prefer one particular person and begin the attachment-making phase, which ends at approximately 8 months. A phase of clear-cut attachment follows. After age 2 years, children who have securely attached have learned three things about themselves and others: trust (i.e., people can be counted on to be there for them), reciprocity (i.e., people take turns interacting), and effectance (i.e., their behavior produces some kind of result in the parenting person).

Using psychoanalytic theory, Margaret Mahler proposes a more complex sequence whereby an infant arrives at some basic sense of self. According to her theory a child's ability to separate and individuate as a person is important. From ages 0 to 2 months, the infant exists in an autistic phase. Through a process of symbiosis, occurring during ages 2 to 5 months, an infant first becomes attached through an intense interaction with a mothering person. From 5 to 12 months, a child begins *differentiation* that is, having a sense of self as different from the parent. From 12 to 18 months, the differentiated infant practices separation. At 18 to 24 months a socially successful infant is free to leave and reapproach the parental person. Individuality and the beginning of object constancy follow. Indications of successful, normal attachment are the infant's ability to use the mothering figure for security and trust.

The father's role (along with the culturally changing maternal role) raises questions about a child's ability to attach to the father as primary parent. Most researchers recognize that attachment of a primary caregiver—whether mother or father—is essential for developing an internal working model for future relationships. *Engrossment* is the term used to describe the father's initial attachment to his child. Recent research describes the critical role that a father plays, not only in the social development of his children, but also in other areas, including cognition and moral development. Specifically, children reared in single-parent households with the mothers as heads have dramatically different social relationships with peers and with other adults, including teachers, when warm and positive social interactions are maintained with their nonresidential fathers. Results of this study show that peers perceive such children as making better friends, and teachers perceive them as having fewer behavioral problems and showing better academic performance than children who have inadequate relationships with their fathers. Daughters show an even higher correlation between adequate father-child relationships and acquiring socially positive developmental skills. Daughters with close attachments to their fathers, whether living in the same household or apart, develop higher levels of critical thinking and cause-and-effect-processing cognitive skills than daughters who are not involved with their fathers.

Behaviorists attempt to describe attachment as a series of learned behaviors. Drive-reduction models stress feeding habits between caregiver and child as stimulating and shaping the attachment. Operant conditioning models stress the need for a reciprocal responsiveness between the child and caregiver for attachment and separation to occur. As in others areas of development, incorporating elements of each theory gives a more complete explanation for the universal phenomenon of social maturation, which requires bonding before social maturation can proceed.

Measuring Social Interaction In addition to information derived from clinical observations, several tools are available to assess basic social interaction skills as a measure of social maturation. *The Ghuman-Folstein Screen for Social Interaction* (SSI) is one such tool. It is a questionnaire designed to measure preschool children's capacity for basic social interaction skills in a variety of contexts. It is most valid and reliable for assessing children from 24 months to 5 years of age and is especially helpful when a relatively quick assessment of children is required (e.g., school consultations, medical-clinical situations, and social intervention programs).

CONSEQUENCES OF ABNORMAL ATTACHMENTS Many studies have attempted to correlate attachment, a child's sense of self, and future behavior. In toddlers between the ages of 12 and 37 months, for example, feeding problems such as infantile anorexia and picky eating are significantly more frequent in children with faulty attachments than in those with secure attachments. Further cognitive and emotional development suffers when attachment is inadequate or faulty. Disruptive school behavior, poor social interactions, and a lack of appropriate long-term relationships correlate with inadequate attachment. On the other hand, children who have experienced secure attachments show high cooperation, enthusiasm, self-esteem, problem-solving ability, and normal relationships.

Socialization in the Childhood Years A child who adequately attaches and differentiates develops a strong sense of self. This rudimentary sense of self-discrimination continues to grow, and in midchildhood, children who have developed a strong self-sense can move ahead into successful peer relationships. Three skills are generally considered essential for children to take the next social step with peers: people perception (i.e., the ability to size up others), intention taking (i.e., the ability to understand motives), and perspective taking (i.e., the capacity to understand what others think and feel). Perspective taking stems from Piagetian concepts and is considered to be motivated by a child's active need to interpret experience.

Robert Selman has devised a system of perspective taking that incorporates all three elements in a child's movement toward a peer orientation. In this system children aged 3 to 6 years are placed at *level 0*, a level of undifferentiated perspective, during which they cannot separate their own needs from those of their peers. At ages 4 to 9 years, children advance to *level 1*, a social-informational stage; following which they can understand that friends have qualities that make them trustworthy or not. At *level 3*, from 7 to 12 years, children become self-reflective. They know that they have ideas and attitudes and consider friends as assistants in the process of self-discovery. By this stage, peers have replaced adults as the source of children's emotional intimacy (Table 32.2-5).

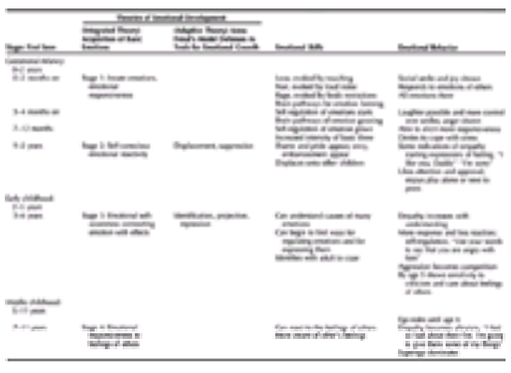

Table 32.2-5 Emotional Development: Theories of Emotional Development and Emotional Skills

Social maturation has been described by other theorists who share a child's sense of self in relation to their environment as their chief focus. Social cognition theory offers a studied approach to socialization on the basis of a recognition of self. From birth to 2 years of age, social cognitionists believe that children engage in learning self-recognition. Infants as young as 5 months can discriminate their images and sounds from those of peers, dolls, and adults. The work of cognitionists shows that from 2 to 4 years of age, children learn how to translate their sense of self and their desires into actions. From ages 3 to 5 years, children discover how to use their self-concept to relate to their environment. From age 5 onward, social comparisons are made in terms of the child's self-esteem. At age 10, as the result of social referencing, children actually drop in self-esteem. From age 8 onward, children learn about their attributes, which lead them to have a sense of master-orientation versus learned helplessness. Finally, from age 11 onward, a true identity is formed. Erik Erikson uses identity formation as the central focus for his own social model as well. According to his schema, children are driven toward the formation of identity in adolescence by first developing a sense of self based upon the perceptions of their ability to function independently and successfully in their social environment. James Youniss offers a model of social interaction that emphasizes social development based upon peer structure and status. He perceives social maturation as proceeding along two tracks. Peer relations develop through a sharing of mutual goals and behavior. Peers enable a child to develop a sense of equality, awareness, cooperation, intimacy, and of being able to cope with differences. Child-adult relationships, which allow internal controls to develop, are based upon reciprocity and authority conveyed through nurturing, modeling, setting rules, explanations, and discipline.

Emotional Development Emotional development can be seen as the literal acquisition of emotions. Children must develop the ability to recognize and use their emotions appropriately. They must also become successful in a complex maturation process that entails learning to become emotionally responsive rather than emotionally reactive to internal experiences of emotion. In addition, children must learn to use their emotional repertoire to handle the inherent anxiety and stresses that are universal to the human condition. Emotional maturation can be understood as the acquisition of coping defenses in infancy and childhood. In infants and children, emotional maturation is best understood by theories that incorporate all of these elements.

Theories of Emotional Development Several approaches help explain how infants and children acquire emotions. Behaviorists and social learning theorists explain emotions as developing secondary to either stimulation or classical conditioning. John Watson postulates a theory of three innate emotions that states that emotions are available and are easily stimulated from birth: fear is evoked by loud noises or loss of support, rage is stimulated by body restrictions, and love is aroused by touching. Others postulate that operant conditioning stimulates emotions. However, none of these theories adequately explains the appearance of spontaneous emotion in older children.

A second group of theorists, cognitive developmentalists such as Donald Hebb, postulate that emotions are simply the byproduct of the cognitive process. In the discrepancy theory of emotional development, Hebb postulates that emotional learning results when children attempt to rectify the discrepancy between a previously learned experience and a new one. This theory does not explain why old situations can be interesting and new ones boring or why some children react differently than others to the same situation. Three-year-old boys, for example, cry more spontaneously in new situations than girls of the same age. A group of children who experience the same event display a wide range of reactions. For example, children can either laugh or cry at a circus clown's performance, scream or smile at a tiger's growling in its home at the zoo, and giggle or moan at an animated elephant's escape from its captor.

The functionalist theories of emotional development—including Anna Freud's defense model and the adaptation of emotional acquisition—transactionally combine previous theories of emotional development with a new, more pragmatic formulation. The functionalist's premise is that emotions are acquired inherently to help children adapt to their world. They are applied through a learned practice that derives from cognitive and social maturation, called *social referencing* (infant's use of parents for cues as to how to react). Functionalists believe that emotional maturity is contingent upon acquiring defense mechanisms for coping with the inevitable anxiety and stress of life. Emotional maturity is quantitatively based upon the number of available emotional defenses acquired successfully by the child. Such defenses include denial, repression, sublimation, regression, projection, reaction formation, substitution, rationalization, compensation, and escape.

Recently, researchers used the most advanced biotechnical equipment to document a series of complex interactions for explaining emotional maturation. Their work appears to integrate the previous theories into a transactional, biological, sociological, interactive formulation. They suggest that emotional maturation—predicated upon genetic predispositions, environmental exposure, and behavioral repetition—occurs through the development of complex interactive pathways within a child's brain. They postulate, for example, that the persistent and pervasive anxiety of children who are neglected or abused as infants stems from the enormous stimulation and subsequent high degree of development of those particular brain synapses.

Proposed Model of Emotional Behavioral Milestones Incorporating aspects of all three theoretical models, a four-stage model for the emotional development of infants and children can be proposed. First in the developmental process must come *acquisition of basic emotions*—those learned directly or inferred from facial expressions, including happiness, anger, sadness, and fear. Throughout the first year, an infant's emotions grow as they are stimulated and rewarded. Smiles, for example, are simple reflexes at birth. By 5 to 6 months, however, smiles or laughs can be elicited when a parent makes a familiar motion that the child anticipates as funny. Appearing at the end of the second year are the second-stage emotions, the *self-conscious emotions*, which arise from injury to, or enhancement of, the sense of self: shame, envy, pride, embarrassment, and guilt. The third stage involves the development of *emotional self-regulation*—strategies to adjust emotion to a comfortable level of intensity. This phase begins as early as 2 to 4 months but increases markedly in middle childhood. Certain patterns corroborate this view of emotional maturation. For example, irritability and temper tantrums diminish progressively with age as children learn to respond rather than react to situations. Crying diminishes in childhood, and worry becomes a common weekly experience for 70 percent of middle-school-age children. Fears diminish with age, but different fears arise at different times in relation to varying life circumstances. Moodiness and unhappiness normally increase toward the end of childhood and into adolescence. Finally, emotional maturity is reached at a fourth stage, when true *empathy* (i.e., the capacity to respond to the emotions of others) is attained. Empathy is a major underlying theme that supports each of the stages of maturation. Rudimentary empathy occurs at age 1 with the development of some self-awareness. Children with a basic level of empathy understand the expression “I'm sorry.” Empathic capacities of children continue to mature until the age of 5, when children can apply their own experiences with feeling to something that they see happening to another. An empathic child can show support: “I'm sorry you didn't get invited to the party.” Finally, by late childhood, empathy develops into *altruism*—empathy directed toward strangers and people with different life experiences (“I'm going to volunteer in a nursing home. So many older people have no one to visit them”).

Psychosexual Development Psychosexual development involves the process of infants and children learning to view themselves and others in terms of gender. It includes aspects of sexual, physical maturation. Three stages describe psychosexual maturation in children: the development of *gender identity*, the formation of a concept of behavior related to their own gender identity (*gender roles*), and the formation of children's attractions to a particular gender in others (*gender relationships*). Psychosexual developmental principles are best explained by integrating several distinctly different theories: a biological explanation, a cognitive-developmental system, a social learning premise, and traditional psychoanalytic theories.

In infants and children physical sexual maturation proceeds in stages. Even before birth there are distinct biological, genetic, and hormonally driven physical changes. In the first 3 months of fetal development, the presence of a Y chromosome and fetal androgens produces masculinization and inhibits feminization. Without a Y chromosome, ovaries form at 12 weeks. Very little sexual physical development occurs until late childhood. In boys, puberty occurs between the ages of 11 and 16 1/2 years; in girls, it generally occurs between the ages of 8 and 16 1/2 years. The current trend is toward earlier sexual physical maturation. It has been speculated that environmental factors, such as hormones in food and better nutrition, may be responsible for such changes.

From infancy onward, however, most psychosexual maturation focuses upon the psychological stages of sexual development. *Gender identity*—a child's perception of the self as either male or female—begins at 3 to 4 years. Two thirds of 3- to 7-year-olds know their own sex, based upon cues such as clothing and hair. There is a large range of individual difference in the timing of gender identity. Gender identity generally appears earlier in girls than in boys. Sociocultural factors can influence the timing as well. Changing gender identity, however, is very difficult after age 4. Gender roles tend to appear at approximately the same ages. By ages 3 to 4 years, most boys show masculine preferences in choices of activities, toys, and peers. Studies show that gender role appears, especially in boys, as early as age 2, even before gender identity, with gender-labeling. A 2-year-old can say, “He is a boy.” A 3-year-old can identify herself: “I am a girl.” Gender role is refined through early childhood so that by early school years a child has a sense of gender stability as well: “I will always be a girl.” From the early school years to adolescence, gender role is further developed to include gender consistency: “I know I'm a girl even if my haircut looks like my brother's.” Behaviors that are labeled clearly masculine or feminine are relatively fixed by ages 4 to 5. Gender stereotyping is well set by age 5. Gender relationships have certain patterns. Infant boys and girls can engage in genital manipulation, but between ages 2 to 5 years, genital interest and sex play increases markedly. In spite of the use of the term *latency*, children between the ages of 5 and puberty have an ongoing interest in sex. Masturbation normally increases during this period. Heterosexual play rises from 5 percent in 5-year-olds to 33 percent in 8-year-olds and 66 percent in 13-year-olds. The cause of gender preference in relationships is currently quite controversial. Homosexual play in early childhood is considered transient and generally does not indicate adult sexual orientation.

Four significant areas of theory explain psychosexual development in infants and children. No single school of thought accounts for the complexity of sexual

development; each theory offers a piece of the psychosexual puzzle. The biological theory of psychosexual development argues that sex differences are established in the brain early in life and that sex-typed behaviors, gender identity, general sex object choice, and sexual mechanisms are all neurobiologically determined. However, this approach fails to account for any significant environmental influences. The cognitive-development theory of psychosexual development links such development to cognition, concluding that gender identity is constructed from an awareness of physical reality and that sexual behavior is determined by the child's self-image. A failure to explain individual difference, sexual preference, or the existence of sex-typing behavior prior to a development of gender constancy makes this theory incomplete. Social learning theories of psychosexual development suggest that sexuality is determined by identification with the same-sex parent. Social theorists emphasize the importance of differential reinforcing of sex-type behavior as the basis for psychosexual development. This theory is inconsistent when children's behavior fails to resemble that of their same-sex parent. The psychoanalytic theory of psychosexual development has changed significantly since Freud's initial schema. His theory, however, has both historical significance and relevance for social, emotional, and moral development. The basis of Freud's schema is that children move through a series of stages during which a conflict between their biological drives and social expectations is confronted. Inherent in his theory is the concept of the critical role that ages 0 to 5 play in personality development. Freud suggests that psychosexual development occurs in five stages. Birth to 1 year is a period of oral gratification, used by the infant's *ego*—the conscious part of the personality—to direct the *id*, which is the inherent biological instinct. From ages 1 to 3 children's egos redirect their ids in anal terms. From ages 3 to 6, the id impulses of children are transferred to the genitals. From ages 6 to 11, sexual instincts cool down or become latent while the superego develops in relation to peers. Arguments against the Freudian explanation include a progressive increase of sexual activity during the latent period as well as difficulties with incomplete explanations for gender behavior, especially in terms of roles (Table 32.2-6).

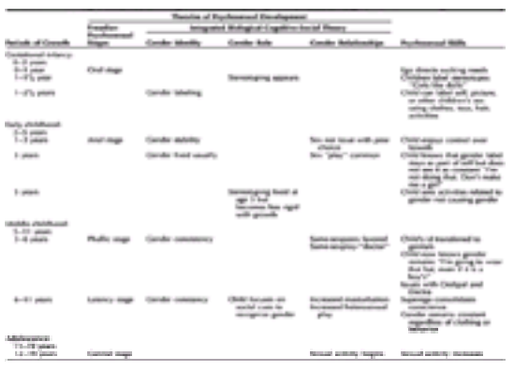

Table 32.2-6 Psychosexual Development: Theories of Psychosexual Development and Psychosexual Skills

Moral Development In developmental terms, moral maturation refers to children's acquisition of internal standards that guide their observable actions or behavior. Infants have only a rudimentary awareness of the effects of their responses on others. They know that smiling makes their parents hug them. Small children can understand that sometimes they do things that make their parents concerned. A 3-year-old can say: “My daddy gets upset when I say yucky words.” However, by the time children enter adolescence, they should have acquired enough moral skills to be able to make decisions on the basis of something other than fear of punishment. They are already able to understand rules in terms of a sense of fairness. At this point, external restrictions are replaced by *conscience*, a person's inner ability to make decisions regarding concepts of right and wrong. In certain cultures elaborate rituals mark the acquisition of conscience. Transactional thinking has catapulted this area of development to its important position in the developmental schema. As with other areas, various explanations are offered to help understand the moral developmental process.

Ethologists suggest that moral principles are innate and universal in origin. They suggest that moral principles are inherited by humans as a species and are developed through social experiences. Such theorists see moral development as adaptive.

Cognitive developmentalists, such as Piaget and Lawrence Kohlberg, suggest that moral maturation is an active process of acquiring a moral sense through a set of cognitively derived constructions that operate in conflictual situations. In Piaget's schema, morality is formed in two stages. Heteronomous morality develops as children are able to see rules as being fixed and dictated by authority, a period also called “the morality of constraint.” A child's behavior is determined by physical consequences and rules are considered unchangeable; motives are not related to intention, actual consequences rule intention. Justice is a concept that is meted out as punishment by some supernatural agent in response to the child's misbehavior. Autonomous morality, or the morality of cooperation, forms as children see rules as flexible and people-dictated rather than absolute. In this second stage, children are able to see morality as linked to intentions, not consequences, and are able to see that moral punishment can be modified by circumstances. In the latter stage, incidents such as illness are no longer connected to punishment for bad behavior in the child's mind. Justice becomes a more abstract consequence and goal. Kohlberg elaborates upon Piagetian views by dividing the moral maturation of children into three levels. At the preschool and early childhood level, morality is based on a system of rewards and punishment meted out by authority figures. In a later phase, beginning after age 6, relativism and social approval are important enforcers. In the middle childhood to preteen years, morality is based on a system of laws that overrides simple approval. The framework for both Piagetian moral development and Kohlberg's form lies in cognitive growth and understanding.

Social theorists, on the other hand, stress that a child's morality is learned when adult models offer instructive examples in their interactions with children. Helpfulness, generosity, and altruism are learned and reinforced by positive feedback. Punishment tends to preclude the development of a conscience. Social theorists explain that negative experiences make the child more stressed and less amenable to adopting any part of the parental directive.

Psychoanalytic theories also stress the importance of the environment on the child. Classical psychoanalytic theory suggests that the conscience is actually a superego that arises from repressed hostility held toward a parent following the resolution of the Oedipal and Electra issues. Guilt and self-punishment occur when a child behaves in a way that is contradictory to internalized parental values. More-modern psychoanalytic theory stresses conscience development as the growth of a superego—arising out of a positive identification with parental values rather than over guilt.

More recent approaches to explain moral development are the gender-based theories. Carol Gilligan asserts that male and female morality develop along different tracks. Women's moral development ends in compassion and an ethics of caring, whereas men's morality culminates in a moral system dominated by the ethics of justice and the assertion of rights (Table 32.2-7).

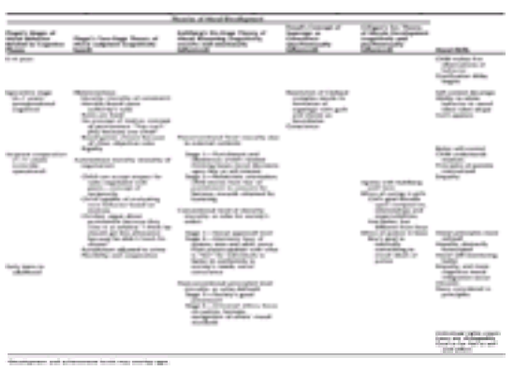

Table 32.2-7 Moral Developmental Theories and Achievements*

Integrated Approach to Moral Development Attempts have been made to formulate a system of moral development that integrates aspects of cognitive, social, and moral maturation. One approach uses the concept of moral volition to frame the moral developmental process. In this system, children mature morally by applying their growing cognitive understanding and social ability to a process of integrating intent and autonomy with the development of conscience—a process that involves five stages of conceptualization: the morality of restraint, morality of mastery, morality of virtuous striving, idealization, and individual responsibility. According to this

view, children can better see their own behavior as independent from their conscience as they progress through the five stages; consequently, they can take full responsibility for their behavior by the time they reach the last two levels of moral development.

DSM-IV CHANGES IN DEVELOPMENT AND THEIR IMPLICATIONS FOR THE FUTURE

Recent significant changes in developmental thinking, as reflected in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), hint at the direction in which developmental research and clinical observation might be leading psychiatry. Atypical development, maturation that quantitatively or qualitatively differs from normal development, is defined with more emphasis upon process, communication, and the integration of various elements of development. Certain language and motor developmental disorders, for example, are differentiated in a more detailed way in DSM-IV and have been moved from Axis II to Axis I categories. DSM-IV models rely upon developmental process concepts. Learning disorders, for example, are regrouped with principles of developmental theory in evidence.

SUGGESTED CROSS-REFERENCES

The contributions of the psychological sciences are discussed in Chapter 3. Classic psychoanalytic theory and Erikson's developmental theory are discussed in Section 6.1 on psychoanalysis and Section 6.2 on Erikson, respectively. Normal adolescent development is reviewed in Section 32.3, and normal adult development in Chapter 50. Principles of examination are discussed in Chapter 33, and clinical syndromes observed in children are reviewed in Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39, Chapter 40, Chapter 41, Chapter 42, Chapter 43, Chapter 44 and Chapter 45. Treatment modalities specific to children are discussed in Chapter 48, and special areas of interest in child psychiatry are discussed in Chapter 49.

SECTION REFERENCES

Adler SA, Gerhardstein P, Rovee-Collier C: Levels-of-processing effects in infant memory. Child Dev *69:*280, 1998.

Astor RA: Children's moral reasoning. Child Dev *65:*1054, 1994.

Bandura A: *A Social Learning Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

*Berk LE: *Child Development.* Allyn & Bacon, Boston, 1994.

Bowlby J: *Attachment and Loss: Attachment*, vol 1. Basic Books, New York, 1969.

Brainard MS, Knudsen EI: Experience affects brain development. Am J Psychiatry *155:*1000, 1998.

Brazelton TB: *Neonatal Behavioral Assessment Scale*. Lippincott, Philadelphia, 1984.

Campos JJ: *Handbook of Child Psychology*. Wiley, New York, 1983.

Chatoor I, Ganiban J, Colin V, Plummer N, Harmon RJ: Attachment and feeding problems: A reexamination of nonorganic failure to thrive and attachment insecurity. J Am Acad Child Adolesc Psychiatry *37:*1217, 1998.

*Chess S, Thomas A: The process of development: Interaction and goodness of fit. In *The Dynamics of Psychological Development*. Bruner, Mazel, New York, 1980.

Chomsky N: *Reflections on Language*. Temple Smith, London, 1976.

Coley RL: Children's socialization experiences and functioning in single-mother households: The importance of fathers and other men. Child Dev *69:*219, 1998.

Courage ML, McCloy UR, Herzberg GR, Andrews W, Simmons BS, McDonald AC, Mercer CN, Friel JK: Visual acuity development and fatty acid composition of erythrocytes in full-term infants fed breast milk, commercial formula, or evaporated milk. J Dev Behav Pediatr *19:*9, 1998.

*Eisenberg N, Fabes RA, Shephard S, Guthrie IK, Murphy BC, Riser M: Parental reactions to children's negative emotions: Longitudinal relation to quality of children's social functioning. Child Dev *70:*2, 1999.

Erikson EH: *Identity, Youth, and Crisis*. Norton, New York, 1968.

Freud S: Three essays on the theory of sexuality. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 7. Hogarth Press, London, 1966.

Gerwitz JL, Boyd EF: In reply to the rejoinder to our critique of the 1972 Bell and Ainsworth report. Child Dev *48:*1217, 1977.

Gesell A, Amatruda CS: *Developmental Diagnosis*. Harper & Row, New York, 1974.

Ghuman JK, Freund LRA, Serwinet J, Folstein S: Early detection of social interactions problems: Development of a social interaction instrument in young children. J Dev Behav Pediatr *19:*411, 1998.

*Giedd J: Human brain growth. Am J Psychiatry *156:*4, 1999.

Gilligan C: New maps of development: New visions of maturity. Am J Orthopsychiatry *52:*199, 1982.

Goodman JC, McDonough L, Brown NB: The role of semantic context and memory in the acquisition of novel nouns. Child Dev *69:*1330, 1998.

*Gorski RA: Development of the cerebral cortex XV. J Am Acad Child Adolesc Psychiatry *38:*344, 1999.

Huffman LC, Bryan YE, del Carmen R, Pedersen FA, Doussard-Roosevelt JA, Porges SW: Infant temperament and cardiac vagal tone: Assessments at twelve weeks of age. Child Dev *69:*624, 1998.

Kagan J, Resnick J, Gibbons J: Inhibited and uninhibited children. Child Dev *60:*838, 1989.

*Kaufman AS: Genetics of childhood disorders II: Genetics and intelligence. J Am Acad Child Adolesc Psychiatry *38:*487, 1999.

*Kohlberg L: Moral stages and moralization: The cognitive-developmental approach. In *Moral Development and Behavior*, T Lickona, editor. Rand McNally, Chicago, 1969.

Kohlberg L: Essays on moral development. In *The Psychology of Moral Development*. Harper Row, San Francisco, 1984.

Milling L, Giddan JJ, Bush E, Laughlin A: Preadolescent suicidal behavior: The role of cognitive functioning. Child Psychiatry Hum Dev *28:*103, 1997.

Muris P, Cor M, Merckelbach H, Sermon A, Zwakhalen S: Worry in normal children. J Am Acad Child Adolesc Psychiatry *37:*703, 1998.

Nash JM: Fertile minds. Time Special Report *149:*48, 1997.

Piaget J: *Origins of Intelligence in Children*. International University Press, New York, 1953.

Plomin R, DeVries JC: The genetics of cognitive abilities and disabilities. Sci Am *278* 62, 1998.

Plomin R, Rutter M: Child development, molecular genetics, and what to do with genes once they are found. Child Dev *69:*1223, 1998.

Richards LK, Tuttle SE, O'Cleary DM: Directed growth of early cortical axons is influenced by a chemoattractant released from an interneuron target. J Neurosci *17:*2445, 1997.

*Rutter M: *Developmental Psychiatry*. American Psychiatric Press, Washington, DC, 1987.

*Rutter M: Psychosexual development. In *Developmental Psychiatry*, M Rutter, editor. American Psychiatric Press, Washington, DC, 1987.

Rutter JJ, Baumann MH, Waterhouse BD: Systemically administered cocaine alters stimulus-evoked responses of thalamic somatosensory neurons to perithreshold vibrissae stimulation. Brain Res *798:*7, 1998.

Saudino K, Eaton WO: Infant temperament and genetics. Child Dev *62:*1167, 1991.

Sellman RL: *The Growth of Understanding*. Academic Press, New York, 1980.

Stewart BA, Schuster C, Goodman CS: Homeostasis of synaptic transmission in Drosophila: Genetically altered transmitter nerve morphology. J Neurosci *16:*3877, 1996.

Stillwell BM, Galvin MR, Kopta SM, Padgett RJ: Moral volition: The fifth and final domain leading to an integrated theory of conscience understanding. J Am Acad Child Adolesc Psychiatry *37:*202, 1998.

Vallee M, Mayo W, Delli FD, LeMoal M: Prenatal stress induces high anxiety and postnatal handling induces low anxiety in adult offspring: Correlation with stress-induced corticosterone secretions. J Neurosci *17:*2334, 1997.

Vygotsky L: *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA, 1978.

Youniss J, Smoller T: *Adolescent Relations with Mothers, Fathers, and Friends*. University of Chicago Press, Chicago, 1986.

32.3 NORMAL ADOLESCENCE

NANCY S. COTTON, PH.D.

Adolescence is a developmental period of dramatic change. Healthy integration of these changes stems from facilitating dialogues between teenagers and the people involved, in the contexts in which they develop. Adults are always on the outside of childhood development; however in this developmental stage, being on the outside can lead to more misunderstandings, distortions, inaccurate readings, and just plain ignorance of what a teenager is trying to understand, do, feel, or be.

What is a teenager trying to do with the behaviors, feelings, and interests adults observe? What is the response from individual adults and peers in the child's historical, cultural, socioeconomic, and geographical context? The adult response to the outside is largely shaped by adult expectations and understanding of what should happen in adolescence and adult feelings about what is happening. These feelings can be intense and range from delight, admiration, and interest to disapproval, dislike, disgust, and fear. Accurate understanding of what should, will, and may happen in adolescence shapes the nature and quality of the dialogue that occurs between an adolescent and family, peers, and society.

DEFINITION

Adolescence refers to the long transitional developmental period between childhood and adulthood and to a maturational developmental process involving major physical, psychological, cognitive, and social transformations. The onset of adolescence is marked by puberty, which is primarily a physical, maturational, hormonal, and growth process. Physical changes and expanding cognitive abilities initiate broader psychological and social changes shaped by the cultural, socioeconomic, and historical context in which adolescence occurs. The concept of a transitional period between childhood and young adulthood is a modern invention; earlier, childhood was directly followed by adult status, marked by puberty.

Earlier views of adolescence described normal adolescence as a time of universal and extreme internal and external turmoil and stress marked by marked emotional reactivity, volatile behavioral unpredictability, dramatic separations from family ties, and severe disruptions in the personal sense of self, leading to separation and autonomy. These extremes were expected and considered normal and necessary for subsequent healthy adult development. However, empirical studies in the last 20 years have described normal adolescence as a time of reordering, realignment, and transformation leading to physical and cognitive development, relational transformations, and psychological differentiation, interdependence, and greater selfcoherence. Daniel Offer found that only about 10 to 20 percent of adolescents experienced severe emotional disturbance or dramatic disruptions in family relationships and self-concepts.

Studies of adolescent development reveal both continuity and metamorphosis—moving forward toward maturity and reaching backward toward childhood for familiarity and grounding. Adolescent development is shaped differently by gender, including sexuality, sources of self-esteem, relationship needs, paths of separating from parents, and the psychological impact of timing of physical development. Rigid schedules of phases of development and categorizations of typical behaviors have been deemphasized and viewed as generic blueprints of adolescence. Normal adolescence can only be seen and understood by studying individual paths of adjustment, with particular attention to their contexts.

Adolescent development is best understood as a dialogue between teenagers and their total environment. Descriptions of adolescent development that ignore the nature of the participants and the characteristics of this dialogue can be inaccurate and deceptive. Perhaps the older views emphasizing the inevitability of adolescent turmoil and psychopathology resulted from a failure to recognize the power of this dialogue to steer development in positive or negative directions.

STAGES OF ADOLESCENCE

Joseph Noshpitz, Robert King, John Schowalter, and David Elkind provide “outsider” reviews of the adolescence. Adolescence is usually divided into phases of early (12 to 14 years), middle (14 to 17 years), and late adolescence (17 to 19 years) described in terms of physical, cognitive, psychological, and social changes; developmental tasks; typical and pathological behaviors and feelings; and healthy outcome measures for each task. These summaries also list parenting tasks, typical parental behavior, and acceptable and unhelpful or pathological responses by parents to adolescent behavior and feelings.

Research describes late childhood and early adolescence as the period of major developmental shifts in all domains, with smaller changes occurring in middle and late adolescence. Although separate developmental lines are usually constructed for each area, actual development involves the mutually influential interplay of all developmental changes. Thus, teenagers who exhibit low-level decision-making in real life situations because of constraints derived from a range of personal and social forces can exhibit higher-order thinking during discussions of hypothetical situations. Physical, cognitive, and psychological changes combine to form a complex web that becomes increasingly difficult to unravel.

Physical Changes Brain maturation and hormonal growth set off major physical changes in early adolescence with a growth spurt and the appearance of secondary sex characteristics. At about 10 years of age, girls show dramatic increases in their height and weight. Boys begin their growth spurt about 2 years later. This is why the average 12- or 13-year-old girl is likely to be heavier and taller than the average boy of this age. Boys grow most rapidly during the thirteenth year, when they may grow as much as 4 or 6 inches in height.

Secondary sex characteristics begin to appear about a year after the onset of the growth spurt. These changes are charted by J.M. Tanner's five-stage scale of pubertal maturity, which traces changes in pubic hair and breast development in girls and pubic hair and genital development in boys. Menarche, beginning around the age of 12 or 13 years, occurs after the growth spurt has peaked. A wide range of factors influence the great variability in the onset and rate of physical changes, including gender, nutrition, body type, stress, ethnicity, geography, and climate.

The body changes in almost every area. The nose and mouth widen, the nose becomes longer and more prominent, and the jaw juts out. The chin is the last facial feature to increase in size. Body hair darkens and lengthens for both girls and boys. Sweat glands, inoperative in childhood, become active, leading to secretions that cause body odor. Voice quality changes. Complexion and acne problems begin to appear. Endurance and strength increases.

Cognitive Changes Cognitive development in adolescence builds the foundation for a more differentiated, multileveled, abstract, subtle, and complex view of self, others, and society. From early adolescence on, adolescents acquire formal or abstract thought, information-processing skills, a knowledge base, and decision-making strategies. Emerging adolescent thinking tends to involve abstract rather than concrete descriptions; to entertain future or ideal possibilities; to become self-reflective and self-aware; to become multidimensional rather limited to a single issue; and to become relative rather than absolute in the conception of knowledge.

These developing abilities have enormous practical impact. Teenagers can integrate concrete experiences, inferences, possible scenarios, historical facts, and remembered experiences into an abstract generalization (e.g. a new perception of self, parents, or a peer as reliable or trustworthy). Self-reflection can create new dimensions of the self involving higher-order personal or moral categorizations. Abstractions can represent possible situations or ideal situations that stimulate new feelings of hope or despair. A teenager can be inspired by the “what ifs” if they suggest a better life and despair if they suggest a dismal future.

Thoughts about ideal adults, parents, friends, or worlds can lead to complex emotional reactions. The ability to separate one's thinking from that of others enhances the capacity for perspective taking and interpersonal empathy. These new capacities lead to preoccupation with the self, extended periods of private introspection,

self-consciousness, and (in some cases) inaccurate application. Elkind describes this as “the imaginary audience” problem in which adolescents falsely assume that others are as preoccupied with their behavior and appearance as they are themselves.

Teenagers learn to process information more efficiently and develop selective attention, better short- and long-term memory, and organizational strategies. Appreciation of multileveled realities leads to more complex understanding of causality in academic, personal, and interpersonal matters. Investigators of adolescent thought are careful to point out that many of these new thinking capacities benefit academic pursuits and theoretical discussions but are not typical of everyday thinking, in which issues are more influenced by psychological and affective pressures. The development and use of new thinking is influenced by personality, gender, culture, and social class to create individual pathways, as in all areas of development.

INDIVIDUAL PATHWAYS THROUGH PHASES

Empirical studies of adolescence reveal multiple adaptive pathways for negotiating the biological, cognitive, psychological, and social transitions of this developmental period. This research leads to the study of individual differences and deemphasizes a unitary picture of adolescence with predictable sequences and phases and typical profiles. Adolescents themselves would certainly agree that “teenagers don't come in neatly wrapped boxes.”

Individual adaptations are shaped by variables within the individual (individual differences), variables in the individual's environment (context), transactions between individual differences and context (the dialogue), and the fit of individual milestones to normative or expected growth curves (timing). Longitudinal research and empirically derived profiles of adolescents show more variability than conformity in development. In their longitudinal study of psychiatrically hospitalized adolescents Stuart Hauser and colleagues reported that some of these adolescents went on to function exceptionally well because of adaptive ego functioning marked by capacities for self-reflection, self-efficacy or agency, self-complexity, persistence and ambition, and positive self-esteem. Their research identifies multiple paths of adolescent development characterized by arrested ego development, steady conformity, progressive ego development, and accelerated ego development. Ego development traces the evolution of meanings that the teenager imposes upon inner experiences and perceptions of people and events. This research concludes that healthy adolescence results from healthy ego development that generates an extensive and flexible repertoire of responses to the normal vicissitudes of living and facilitates a healthy balance of autonomy and involvement with others.

In adolescence, a person is male or female first and an adolescent second. Timing, gender, and match are the crucial variables in whether early puberty is a positive or negative experience. Early maturation appears to benefit boys who are taller and stronger and perceived as more sexually attractive than their late-maturing peers. Since boys measure positive self-esteem by athletic and academic prowess, early-maturing boys also benefit psychologically. Early-maturing girls must cope with sexual advances from boys at an earlier age and a longer awkward period when they are taller and heavier than boys as well as hair growth, body odor, and changes in facial bones and features before their peers. Early-maturing girls have poorer body images and engage in smoking, drinking, and sexual activity at earlier ages than their late-maturing peers. The differences in gender reactions to puberty are in part influenced by adult and societal attitudes that increase expectations for early maturers and decrease expectations for late maturers. Mary Pipher's popular book on adolescent girls and William Pollack's book on boys describe the gender-specific struggles and socialization challenges of adolescents and their parents.

CONTEXT

Individual pathways are constructed and shaped by sex and gender, socioeconomic status, historical context, parental education, culture, race, ethnicity, religious affiliation, sexual orientation, and geographic location. These variables provide positive or negative possibilities, guidelines, and resources to shape the consequences and meanings of change. Socioeconomic class mediates exposure to the entire catalogue of ills in society. Poor, urban youths measure significantly lower on perceived life chances than privileged youths, which places them at greater risk for troubles in adolescence.

The predominant value system of the culture or community can also shape development regardless of socioeconomic status, as noted in adolescent adaptations in privileged boarding-school and poor, urban gang cultures. Immigrant parents from some cultures frequently influence children with their high expectations for educational achievement. Children from financially privileged backgrounds can experience unreachable expectations or “benign” neglect.

Adolescent development is affected by the historical era. Today, children living in isolated, culturally homogeneous communities (e.g., rural or elite) are exposed to broader worlds via television, newspapers and computers and can interact with this world through the Internet. Guiding and supportive social structures such as the traditional family or religious institutions are currently in transition. American adolescents are growing up in a world with fewer rules—a much less predictable, unstable, larger, more unfamiliar world. The sexual practices of the 1990s adolescent have been shaped by changes in adult practices, the individual choice of partner, and premarital sexual relations. Modern teenagers are most likely to die in a substance-abuse-related automobile accident, by suicide, or by violence.

Membership in a marginalized community has always been a risk factor in development, but the interaction of culture and socioeconomic factors can change the context for the developing teen. Teenagers from diverse cultures or with alternative sexual orientations whose families have sufficient economic resources to choose tolerant living and educational communities can be shielded from the destructiveness of hostile and judgmental communities. Relatively normal adolescent development is possible for teenagers with major learning and language disabilities, mental illnesses, or retardation or physical disabilities given sufficient social or familial economic resources, sophisticated treatment, educated and understanding parents, effective school systems, and safe and attractive living spaces and communities.

Accounts of adolescents have been based on narrow samples of white, middle-class, urban, Western-European adolescents growing up in conventional family structures consisting of one or two heterosexual parents raising their biological children, born to them during their twenties and thirties. Barbara Okun has described successful variants on conventional family formats created by older, adoptive, homosexual, and multiracial families.

PARENT-CHILD DIALOGUE

When the point of adolescent development was separation and autonomy, there was relatively little research on normal parent-child relationships. Lawrence Steinberg describes more-complex parent-child roles in which parents and teenagers actively participate in the mutual and reciprocal process of redefining the relationship. The result is a realignment in which parents and adolescents maintain close ties and resolve differences through mature cooperative negotiation and mutual respect. Reaching this point involves inevitable perturbations (rather than storms) along the way.

The quality of the conflict-resolution strategies shapes the nature and outcome of the dialogue. Conflict starts with behaviors that deviate from expectations. Children and parents argue more at the onset of puberty, continue to squabble throughout adolescence, and begin to resolve differences more easily in late adolescence. Parents and children argue more often about such issues as curfews, chores, schoolwork, and personal hair and clothing styles than about conflicting value and political systems. Parent-child conflicts yield positive outcomes when they are temporary, take place in the context of warm close relationships, and proceed with mature conflict resolution strategies that end in negotiated compromises.

Jenny was a responsible 16-year-old who remained home for the weekend to attend practice for the Latin club's annual theater night. Her parents and younger sister went on an annual ski trip with old friends in another state. When her parents returned home it was obvious that Jenny had given a party. Her parents were furious with their daughter's betrayal of their trust in her and horrified by the condition of their house. Several months of strained relationships ensued, marked by initial parental blowups; constant parental monitoring (including picking up and dropping off Jenny at school); calling the parents of other children known to have attended the party; indefinite weekend grounding; and frequent, intense, emotional monologues.

Jenny was appropriately ashamed of her actions and shocked in retrospect about her poor judgment and inconsiderate behavior. Throughout the punishment period she began to wonder how she could have thought the evening was worth it. She mourned the loss of her parents' trust and respect. Although she was humiliated by the supervision and rejected by some of her friends who were furious about her parents contacting their parents, Jenny “took her punishment” because she felt that she deserved it.

The third Saturday night following the party, Jenny was watching television with her mother. They were laughing and Jenny found herself crying and then weeping uncontrollably. Her mother spontaneously hugged her for the first time since the incident. Finally, Jenny could say, “Will it ever be the same? I can't stand this feeling. I'm such a failure.” Her mother was shocked by the intensity and depth of her feelings and her words. Both parents wanted Jenny to experience guilt and regret over her actions—“to learn from her mistake”—but something was going wrong. Her mother told Jenny, “Things will never be the same. But things do not always need to be this bad. We need to figure this out together. Your Dad and I do not think that you're a failure. In fact we felt we were the failures. But we all need to learn from this and rebuild our relationship. It will take some time. I'm still pretty mad and frightened by all this. But I know I love you so much, Jenny. I always have and I always will. That's going to get us through this.”

This began a new dialogue and realigned the relationship between this basically healthy teenager and her parents who had a history of love and caring and being able to speak sincerely to each other. The new dialogue was based on a new-found respect for the unpredictability of feelings, thoughts, and actions in adolescence that necessitates less emphasis on global parental trust and global adolescent “goodness and reliability” and more emphasis on dialogue and anticipation of challenges to existing capacities.

The inevitable bumps along the road may result in negative outcomes when they become pervasive, occur with a backdrop of shaky emotional relationships, involve parents or children who may be temperamentally or psychiatrically more vulnerable, and escalate regularly to devaluing and hostile interactions that end in withdrawal or disengagement. Constructive conflict resolution contributes to adolescent identity formation and the development of social-cognitive skills, relatively advanced reasoning, and ego. Ineffectively managed and continuously high levels of unresolved conflict are associated with psychosocial problems during adolescence.

From the Adult Side Constructive parent-adolescent engagement is shaped to a large extent by parental responses to teenagers' legitimate, but often annoying or anxiety-provoking, efforts to make sense of the internal and external changes that confront them. Studies have identified certain parental and familial factors associated with effective management of parent-adolescent tensions: the affective quality of the preadolescent relationship, parental expectations and assumptions about adolescence and their adolescent in particular, parenting style, and parental ego functioning.

Adolescent development is better in a parent-child relationship characterized by secure attachments, well-worn styles of discussion, supportive communication, and adaptive strategies for conflict resolution. Steinberg points out that stepparent-adolescent relationships are particularly vulnerable because they lack the relational context to prevent low-level bickering from escalating into dysfunctional discord or abuse.

Hauser and colleagues describe parents with more-mature ego development as more prepared to engage in, and respond to, the adolescent dialogue. These parents have greater self-awareness and appreciation of individual differences, leading to more curiosity, reasoning, and joint searches for solutions in their family discussions. A more reasoned approach to adolescent experimentation with style, rule breaking, and questioning of society can help clarify values, prioritize concerns, and effect compromises.

More-mature ego functioning may buffer increased feelings of inadequacy around parenting, greater anxiety, negative midlife self-evaluations, and greater marital tension reported in parents of early adolescents. These parents also start the dialogue with more-adaptive responses to the adolescent's emotional distancing, body changes, increased questioning of family rules and conventions, different tastes in clothes and music, and search for privacy and extrafamily relationships. Parental developmental also shapes their response to the dialogue. Middle-age parents may be confronting diminished physical attractiveness and life opportunities while their adolescent's physical development is in bloom and life opportunities are opening up.

Parents with mature ego functioning are more able to use the authoritative parenting style described by Diana Baumrind and her colleagues and associated with more-successful adolescent psychosocial development as measured by social competence, self-reliance, school performance, and self-esteem. Authoritative parents provide a supportive framework in which adolescents can stretch and grow. Authoritative parenting balances limit setting and negotiating, values autonomous expression and disciplined conformity, and provides emotional responsiveness through love, acceptance, empathy and mature conflict-resolution strategies. These qualities create a family environment that encourages and uses mature conflict-resolution strategies grounded in warmth and caring, including discussion, negotiation, self-expression, and respectful interactions during the most heated debates. The dialogue needed in adolescence was nurtured and modeled throughout childhood and can now be used.

As their children mature, parents recall and reexperience their own childhoods. This can be deeply troubling when they have had painful childhoods, even for parents with mature ego functioning.

Susan arrived for her first appointment red-eyed and anxious. She had called about her 13-year-old son. She described the escalation in fighting over the past 3 months. Sean and she had always been close. This had been important because he had always been “uptight” in spite of his general success with friends, school, and athletics. In addition to starting a psychological evaluation of Sean, an additional source of stress in this relationship emerged. Susan's mother had died when she was 12 years old. As the oldest of seven children, she assumed her mother's role in housekeeping and childcare. Her devastated father worked diligently to provide financially for the family but remained emotionally unavailable. Susan realized that she was working hard to give Sean the kind of adolescence that she had missed and yearned for. She resented the fact that Sean was so ungrateful for his opportunity to participate in athletics and after-school activities while her high school years had been filled with household chores and childcare. Her resentment had been largely unconscious before her treatment. In addition she realized how hurt she was when Sean blocked her vicarious enjoyment of his adolescent experience in his efforts to lessen his dependence on his relationship with his mother. Her broader awareness of her multiple feelings about Sean's development allowed her to tolerate and then transform their frequent battles. They agreed on Sean's legitimate need to be more independent of her and they were able to recover some of the pleasure that they had always had in each others company.

Stereotypic parental expectations and assumptions can severely hamper the dialogue. American culture stereotypes teenagers as victims, rebels, victimizers, or sophisticated young adults, and stereotypes inevitably distort interpersonal interactions. Parents can become too controlling after the first missed curfew; too judgmental after a drunken episode or too devaluing after a drop in school grades if they fail to put these behaviors in the context of their child's ongoing development and see them as signs of the child fitting their most feared stereotype.

If the interpersonal dialogue between a developing adolescent and caring adults deteriorates, negative stereotypes are more likely to dominate the person-to-person interactions. Adults plead with adolescents to change their style to avoid the negative images, and adolescents respond by requesting that adults stop oversimplifying and being emotional about their harmless experimentation with their “surfaces.” In the 1960s reasonable adults could panic over long hair on a boy; in the 1990s adults are undone by tattoos and body piercings. Negative stereotypes also function as mirrors held up to adolescents by their parents, and the teenager may accept the adult assumptions and interpretations about their behaviors, ideas, or dress.

Table 32.3-1 lists guidelines for the parent side of the dialogue.

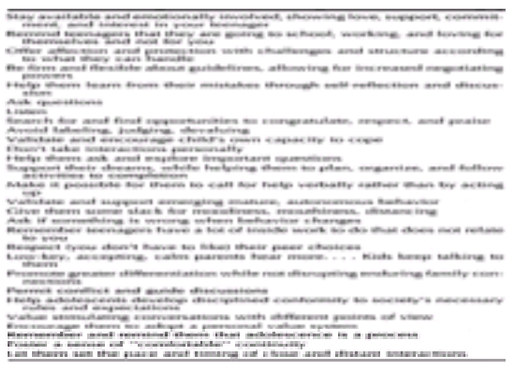

Table 32.3-1 Guidelines for the Parent Side of the Dialogue

From the Adolescent Side Table 32.3-2 summarizes the major life questions posed by a focus group of primarily white, middle-class adolescents. The list mirrors the process of self development, moving from “Who I am” to questions about measuring up, likability, and attractiveness.

Table 32.3-2 The Big Questions: Working Through What?*

Journey to Self Versus Rebellion Adults can miss the teenager's absorption in these bigger questions if they overlook the arenas in which they ask them or if they interpret the teenagers' lifestyle choices only in terms of the parent-child relationship. Adolescent fashion, behaviors, and tastes are not merely “rebellious” alternatives to parental values and behaviors. Teenagers are learning and making serious life choices when they do the varied and sometimes extreme things they do. Girls and boys do these things differently as do daughters and sons, sisters and brothers, firstborns and later borns. Individuals create different pathways.

Adults see the outside (e.g., struggles over schoolwork despair over a bad haircut, a rebuff from a peer). What appears to be a banal concern to an adult can be a crucial life experience for a developing teen. For the teenager they are explorations into who I have been, am, or could be. They are also badges of membership in some peer groups and nonmembership in others.

Carol had always been able to depend upon her 13-year-old son, Darien to pick up his younger sister after school on Tuesdays and take her to gymnastics. She received a hysterical call from Cheryl who was left at school. In the evening, Darien was initially angry when confronted by his mother. He told her, “Anyway, I needed to go with the kids. I can't be taking care of my little sister all the time.” As the evening wore on and Carol and Darien became less emotional, he apologized and told her, “Josh had never asked me over before. I couldn't not go. What if he never asked me again?” Carol could finally see the situation the way that Darien did and was able to help him think of other ways to juggle his responsibilities in the future.

Cognitive Development and Social Context The questions asked are partially determined by maturing cognitive skills and social context in early, middle, and late adolescence. Questions and order of importance of topics change drastically in focus groups of socioeconomically or culturally diverse youth. Nothing separates urban, poor teenagers from their economically privileged peers as much as their proximity to daily violence and the smells, sights, and sounds of poverty, which make the question not who you are or what you will be but whether you will be.

In less dramatic and compromising contexts, teenagers are primarily focused on self-development, which often leads to periods of intense self-absorption. The famous emotional volatility of adolescence is the external evidence of the challenging work within. The stable self-images created and supported in the middle-school years by parents and other nurturing adults are reexamined by the teenager and put to a test: Is this still me? Will this be me? Do I want this to be me? Is it likeable, worthwhile, attractive to me, peers, adults?

Psychological Milestones of Self-Development Susan Harter has mapped out the milestones of the self-development process: generation of components of the self; differentiation and exploration of these components; valuing of some parts and devaluing of others (self-esteem regulation); dropping of undesirable and development of valued parts; reconciling of contradictory self images; and integration of self components. During this process the adolescent moves from creating a self portrait to exploring broader societal roles in terms of gender, occupation, religion, and political identities.

Identity Formation Erik Erikson has called adolescence the developmental period of identity versus identity confusion, as the teenager works through being or not being oneself and issues of sharing oneself. Even in normal adolescence, self-discovery involves discomfort, confusion, anxiety, and unhappiness in the short run. Early adolescence is particularly marked by the negative emotions caused by the inner reworking of the self. The positive outcome of this process, in later adolescence, is the development of “the real me” or a clear, integrated, realistic, internalized sense of an acceptable self (positive self-esteem) perceived as acceptable to society. The healthy adolescent identity is not the final self, but the basis for the continued search for meaningful identity in the future.

Erikson describes the normative process of identity formation as moving through a period of exploration of alternative selves (identity diffusion) toward a period of choice, commitment, and consolidation. Maladaptive attempts at identity formation can avoid or get stuck in the process by premature commitment (foreclosure) or failure to commit (identity diffusion). Teenagers who foreclose the process adopt identities prescribed by parents or other authority figures without ever experiencing tension or exploring options. Although diffusion is developmentally appropriate in early adolescence, in later adolescence and young adulthood it represents a maladaptive inability to seek decisions and make commitments. Pathological outcomes in identity formation lead to turning against the self (e.g., depression, eating disorders, and suicidality) or turning against others or society (e.g., delinquent behaviors).

Components of the Self The components of the self come from all domains including physical, academic, athletic, social, artistic, and moral, which are gradually sorted into the more important and less important parts of the self. In early adolescence the self is defined by relatively concrete self-descriptions of the social and behavioral exterior, physical attractiveness, athletic ability, popularity. The late adolescent self-portrait is defined by the psychological interior, a composite of beliefs, emotions, motives, wishes, and attitudes. Developing abstract capacities for self-reflection and reflection on the thoughts of others, teenagers create personal narratives about themselves and use the imaginary audience of others to validate and evaluate themselves.

Self-Esteem Development Self-esteem is positively regulated when teenagers approve their competence in domains that they value. The search for self also incorporates evaluations from others (“the looking glass–self”) such as parents, other adults in authority, individual peers, and peer groups. Thus self-esteem regulation consists of a general sense of liking oneself as a person, global self-esteem, and domain-specific self-esteem, based on how adequate adolescents feel in the areas they consider important.

Researchers have found that global self-esteem depends heavily on what others think—individual adults and peers as well as society as a whole. Studies of African-American youths with positive self-esteem found that the positive support of their family and members of the African-American community filtered out destructive racist messages from the white community. Pipher's discussion of developing adolescent girls emphasizes the destructive role of sociocultural messages to developing girls. Physical appearance consistently correlates most highly with global self-esteem, followed closely by peer and parental acceptance, and then scholastic, athletic, and behavioral competence. Early teenage girls are more vulnerable to comparisons with cultural standards of the ideal female as an ultrathin person, subsuming her personal assertiveness, competence, and individuality under the desire to please others. Adolescent girls' self-development also needs supportive, affirming subcultures to buffer the "girl-hurting isms," such as sexism, consumerism, and "lookism."

Although parent and adult feedback remain essential to the process of self-understanding, adolescents begin to look more and more to their peers for validation and approval. Peer pressure is probably better understood as "self-pressure" to meet the expectations and win the approval of the selected peer group. When teenagers pick a peer group with behaviors and values more similar to those of their parents, the teenagers and parents have a more harmonious journey. Selection of the peer group depends upon who the teenagers are trying to be or think they can be. Teenagers with ability in academics, athletics, or the arts who can function well in those arenas are accepted by these peer groups. If teenagers are rejected by more-competent peer groups, they seek out groups who will accept and validate their level of competence. Self-development and self-esteem regulation depend upon feedback from peer groups and close friends. The feedback from close friends is often discounted in favor of the perceived objectivity of the greater peer group.

Early, Middle, and Late Adolescent Self-Development The course of self-development is most tortuous in early and middle adolescence when self-esteem is at a low point. Teenagers are exploring multiple, often contradictory, selves based upon fragments of childhood selves and the diversity of possible selves generated by their capacity to imagine ideal or future selves. The comforts of a more cohesive and continuous self need to be delayed for a genuine attempt at integrating new selves. The "quality of the dialogue" is seriously threatened. From the outside, parents watch their athletic child give up sports, their kind daughter become critical, their trustworthy child throw a party while they are away for the weekend, or their studious child fall off the honor roll. Parents' responses are less helpful if they take these exploratory choices as personal affronts, treat them as final decisions on identity, cannot explore the choices with the teenager, refuse to set firm boundaries and flexible guidelines around dangerous and safe choices, or fail to validate and respect the power of this process. Everyone must remember that a process is going on and that learning takes place through mistake making.

Ken and Stacey called the clinician about their 14-year-old daughter, Kim. They described escalating fighting, Kim's apparent abandonment of her previous friends and field hockey, and her demands to quit band. Ken and Stacey were well-respected professionals in their community, married happily for 20 years after meeting in college, and articulate and caring parents. They described Kim as a challenge since early childhood, with a quick temper, bossy with friends, irritable doing family activities and routines, academically bright, and testy with some (not all) teachers. "If a teacher gets Kim, she or he likes her and Kim likes the teacher and does well in that class. When Kim feels unliked by a teacher, she becomes sullen, talks back, and doesn't perform academically as well as she could." Ken and Stacey feared that they were losing control over Kim. She had run away from home the previous weekend when they refused to let her stay out later than her curfew. Ken had followed her along a dark country road until she had finally agreed to walk home.

The dialogue in this family had seriously broken down. Kim's angry declaration of her rights to be independent were met by parental rules and standards that they believed to be reasonable for a child her age. Kim was escalating her position from words to actions.

Kim refused to come to see the clinician. She agreed to come only when it was suggested that the psychologist needed to meet her because her parents wanted to work on being better parents for her. She could provide her view of how they functioned as parents. The most striking moment in her first meeting was when Kim said, "So you don't think I'm the monster."

The clinical work with this child and her parents took the form of shuttle diplomacy. The clinician suggested that the problem was in the quality of the dialogue not in "problem parents" or a "problem child." They were all reaching out to each other and frightening each other. Kim had given her mother a book on adolescent girls to help her mother understand her. The book contained examples of severe eating disorders, drug addiction, and identity breakdowns. When Kim asked her mother theoretically what she would do if a child of hers tried alcohol or marijuana, Stacey said she would send her to drug and alcohol treatment. It was obvious that both parents and child were trying desperately to connect with each other. In separate clinical sessions, the clinician tried to help them identify ineffective ways of responding to each other or of bringing up situations. Kim learned that her provocative and combative style was defeating her sincere wish to be understood and respected. Stacey learned that her inability to discuss risky behaviors or more flexible rules was closing off conversation and encouraging deception and an overreliance on Kim's new and older friends, who were experimenting with parental limits and drugs and alcohol. Ken learned that he intimidated his daughter with his definitive stances, articulated clearly with suppressed emotion. When he felt most hurt by his daughter, she was usually feeling most attacked by her father.

A new emphasis on building respect and trust, rather than assuming that these qualities were automatically present, helped to refocus the dialogue. Kim was able to refine her asking skills and take responsibility for her deceptions. Stacey was able to listen to her daughter's concerns and try to understand her, rather than try to fix or protect. Kim's father became less reactive to her provocative remarks, which were often pointed personality critiques of his shortcomings. In the space provided by some peaceful dialogue, parents and adolescent noticed small ways in which they cared for each other. Diplomacy and restraint, coupled with a willingness to compromise (more compliance by Kim and more flexibility by her parents) brought more harmony to their relationship. This, in turn, facilitated Kim's reengagement in more-positive developmental pursuits (e.g. academics and athletics) and her parents' return to their more-mature conflict resolution strategies.

Kim's parents were amazed when they received a call from her on a Friday night, before her curfew. She asked if they could pick her up from a party. She instructed them to wait three houses down from the party. Without asking any questions, Ken drove to pick her up. Kim did not want to give details. She simply told her father, "That's not me anymore."

Discovering the true self in middle adolescence involves choosing between the differentiated and valued parts of the self, adopting some and rejecting others. The feedback from self-evaluations, based upon social comparisons and the feedback from adults and peers, partially determines what is kept and what is shed. Later stages of self-development involve the integration of parts of self, based upon what feels compatible and feasible. More-mature ego development allows higher-order integration of more-complex portraits of the self, just as less developed ego skills simplify self-knowledge. Healthier teenagers can see how they can be both smart and dumb, kind and mean, capable and incompetent, lovable or unpopular depending upon the context.

In older adolescence, integration of past, current, and future selves involves reconciling and coordinating conflicting insights into self-knowledge and self-acceptance. Older adolescents will refuse to participate in an activity with reasons such as "That's not me."

The late adolescent version of the real me is a more stable, comfortable sense of self with continuity to childhood identities and confidence in newer competencies and roles. Although life choices for career and religious and political identities may not be made, the process of identity formation up to this point has developed the skills to make those choices and the self-confidence to face those decisions with less anxiety and confusion than the junior high school student brought to the challenges of early adolescence.

Views From Inside Young-adult literature provides a unique opportunity to listen to the inner voice of adolescence. These stories can show adults how it looks and feels from within.

Many of these stories are written as journals or narratives: *Dear Nobody* is about a pregnant teen; *Ironman* is about an angry teen learning how to be heard and how to listen; *Ruby* and *Sees Behind Trees* place adolescent development in other cultures; *Deliver Us from Evie, Bad Boy*, or *A Boy's Own Story* reveal different experiences growing up gay; *Imitate the Tiger* explores the development of alcoholism in a football player. Gang life with conflicting loyalties to self, friends, and society is the subject of *The Outsiders* and *That Was Then, This is Now* and *A Way Out of No Way* collects stories of black childhoods. Holden Caufield in *Catcher in the Rye* became the archetypal angst-filled teenager of the 1950s and is the hero for a 1996 teenager in *Bottom Drawer* who finds relief from his troubled relationship with his stepfather via the Internet.

SUGGESTED CROSS-REFERENCES

The ideas of Jean Piaget are explained in Section 3.2 and of Erik Erikson in Section 6.2. Normal human sexuality is discussed in Section 19.1, eating disorders are

discussed in Chapter 20, normal child development is outlined in 32.2, family therapy in Section 30.5 and Section 48.5, mood disorders and suicide in Chapter 45, and psychiatric treatment of adolescents in Section 48.10.

SECTION REFERENCES

Allen JP, Hauser ST, O'Connor TG, Bell KL, Eickholt C: The connection of observed hostile family conflict to adolescents' developing autonomy and relatedness with parents. Dev Psychopathol *8:*425, 1996.

Baumrind D: Rearing competent children. In *Child Development Today and Tomorrow*, W Damon, editor. Jossey-Bass, San Francisco, 1989.

Brown LM, Gilligan C: *Meeting at the Crossroads*. Harvard University Press, Cambridge, MA, 1992.

Bogenschneider K, Wu M, Raffaelli M, Tsay JC: Parent influences on adolescent peer orientation and substance use: The interface of parenting practices and values. Child Development *69:*1672, 1998.

*Boyd D: *Bottom Drawer*. Rubicon, Canada, 1996.

Cheripko J: *Imitate the Tiger*. Boyds Mills, Honesdale, PA, 1996.

Crosier LM, editor: *Casualties of Privilege: Essays on Prep Schools' Hidden Culture*. Avocus, Washington, DC, 1991.

Damon W: *Greater Expectations Overcoming the Culture of Indulgence in America's Homes and Schools*. Free Press, New York, 1995.

Doherty B: *Dear Nobody*. Beech Tree Edition, New York, 1994.

Dorris M: *Sees Behind Trees*. Hyperion Books for Children, New York, 1996.

Elkind D: *Understanding Your Child from Birth to Sixteen*. Allyn & Bacon, Boston, 1994.

Erickson SJ, Feldman SS, Steiner H: Defense mechanisms and adjustment in normal adolescents. Am J Psychiatry *153:*826, 1996.

*Fabes RA, Carlo G, Kupanoff K, Laible D: Early adolescence and prosocial/moral behavior I: The role of individual processes. J Early Adolesc *19:*5, 1999.

Guy R: *Ruby*. Bantam Doubleday Dell, New York, 1976.

*Harter S: Self and identity development. In *At the Threshold The Developing Adolescent*, SS Feldman, GR Elliott, editors. Harvard University Press, Cambridge, MA, 1990.

*Hauser ST, Powers SI, Noam GG: *Adolescents and their families: Paths of Ego Development*. Free Press, New York, 1991.

Hinton SE: *The Outsiders*. Bantam Doubleday Dell, New York, 1995.

Hinton SE: *That Was Then, This is Now*. Bantam Doubleday Dell, New York, 1971.

Jessor R: Risk behavior in adolescence: A psychosocial framework for understanding and action. J Adolesc Health Care *12:*597, 1991.

Kastner LS, Wyatt JF: *The Seven-Year Stretch: How Families Work Together to Grow Through Adolescence*. Houghton Mifflin, Boston, 1997.

Kerr ME: *Deliver Us From Evie*. Harper Collins Children's Book, New York, 1994.

Noshpitz JD, King RA: Adolescence. In *Pathways of Growth Essentials of Child Psychiatry*, JD Noshpitz, editor. Wiley, New York, 1991.

Offer D, Schonert-Reichl KA: Debunking the myths of adolescence: Findings from recent research. J Am Acad Child Adolesc Psychiatry *31:*1003, 1992.

*Okun BF: *Understanding Diverse Families*. Guilford, New York, 1996.

Pollack W: *Real Boys: Rescuing Our Sons from the Myths of Boyhood*. Random House, New York, 1998.

Pipher M: *Reviving Ophelia: Saving the Selves of Adolescent Girls*. Ballantine, New York, 1995.

Powers SI, Hauser ST, Kilner LA: Adolescent mental health. Am Psychol *44:*200, 1989.

Salinger JD: *The Catcher in the Rye*. Viking Penguin, New York, 1951.

Schowalter JE: Normal adolescent development. In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Shantz CU, Hartup WW: *Conflict in Child and Adolescent Development*. Cambridge University Press, New York, 1992.

*Steinberg L: Autonomy, conflict, and harmony in the family relationship. In *At the Threshold: The Developing Adolescent*, SS Feldman, GR Elliot, editors. Harvard University Press, Cambridge, MA, 1990.

Steinberg L: *Crossing Paths: How Your Child's Adolescence Triggers Your Own Crisis*. Simon & Schuster, New York, 1994.

Steiner H, Lock J: Anorexia nervosa and bulimia nervosa in children and adolescents: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*352, 1998.

Weinberg NZ, Rahdert E, Colliver JD, Glantz MD: Adolescent substance abuse: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*252, 1998.

White E: *A Boy's Own Story*. EP Dutton, New York, 1982.

Wieler D: *Bad Boy*. A Groundwood Book, Douglas & McIntyre, Toronto, Canada, 1989.

Woodson J: *A Way Out of No Way*. Henry Holt, New York, 1996.

CHAPTER 33. PSYCHIATRIC EXAMINATION OF THE INFANT, CHILD, AND ADOLESCENT

ROBERT A. KING, M.D., MARY E. SCHWAB-STONE, M.D. , BRADLEY S. PETERSON, M.D. , AND ARMIN PAUL THIES, H.D.

Psychiatric assessment of the child or adolescent is undertaken for a variety of purposes including outpatient office or clinic evaluation to determine the need for treatment; emergency room assessment of safety issues and need for hospitalization; consultation to pediatric, school, or legal services; as a component of family or other mode of treatment; and for research purposes. While the structure and style of the assessment will vary to some extent depending on the purpose and setting, certain elements are common to all.

Core features of the clinical diagnostic assessment of the child are identification of the reasons for referral; evaluation of the nature and extent of the child's emotional or behavioral difficulties or both; and determination of factors in the child, family, and environment that cause, exacerbate, or potentially ameliorate those difficulties. While information gathering and differential diagnosis are the primary aims of such assessment, the evaluation process should also facilitate the development of an alliance with the clinician around understanding the child's difficulties and pursuing the treatment recommendations that emerge from the evaluation.

Since decisions about intervention follow from evaluation, clinical assessment may rightly be regarded as a cornerstone of child and adolescent psychiatry. Clinical evaluation of the child or adolescent is both challenging and complex in that the clinician must draw upon multiple sources and a range of techniques for eliciting information on various aspects of the child's functioning. This information must then be integrated from multiple theoretical perspectives using the clinician's knowledge of both development and psychopathology to arrive at a diagnostic formulation that can guide treatment planning effectively.

The initial portion of this chapter focuses on the comprehensive clinical assessment of the child and adolescent; subsequent sections discuss more-specialized types of assessment, such as the assessment of infants and adolescents, psychological testing, laboratory assessment, and clinical research assessment instruments.

CLINICAL INTERVIEW OF THE CHILD

The psychiatric assessment of the child requires a comprehensive approach that evaluates the child's developmental progress in various domains and positive adaptive capacities, as well as the presence of the pathognomonic symptoms of specific disorders. A developmental approach to the assessment of the child is essential because children differ from adults in certain key respects.

Distinctive Features of Childhood Psychopathology First, psychiatric disturbances in children often consist of a lack of developmental progress in one or more domains, rather than the presence of specific symptoms that are pathognomonic of adult disorders. For example, a nursery school child's failure to develop useful social language or interactions or a school-aged child's inability to meet the developmental expectation of separating from parents and settling into the school day may prompt the parent, school, or both to request an assessment.

Second, the child's developmental status may affect the clinical presentation of various syndromes. For example, in children, depression often presents with irritability and somatic complaints, while excessive guilt or depressive delusions are rare.

Third, development brings expectable periods in which distressing emotions or impairing behavior may occur as part of a normal transition, for example, the separation anxiety of a child starting preschool or the oppositionality of the adolescent.

In many cases clinical conditions represent severe forms of symptoms found in milder form in nonreferred children. Fears, tantrums, moodiness, or restlessness are relatively common in childhood and occur transiently at different stages. Assessment may be sought by concerned parents needing guidance on how to understand and manage these developmental manifestations. Thus, the clinician must judge whether the behavior is likely to resolve with time and without substantial deleterious impact on the child or family or whether, instead, the level of distress, compromised functioning, or symptom persistence indicates the need for clinical intervention.

To distinguish transient or normative difficulties from those that are more clinically worrisome, the evaluator must possess a solid knowledge of both normal and abnormal child development. This developmental frame of reference includes an understanding of what behaviors can be expected normally in children of different ages, the time frame within which various behaviors normally wax and wane, and the natural history of psychiatric disturbances at different stages in development, including knowledge of the ages at which particular syndromes are more or less likely to present.

Another difference in the adult and child psychiatric assessment is that many children coming to clinical attention have difficulties that cannot be neatly subsumed under the rubric of a single diagnostic label. Thus, comorbidity is usually not the exception but the rule in childhood disorders. Even in epidemiological studies of children and adolescents, as many as half of those who meet diagnostic criteria for one disorder also meet criteria for at least another disorder. This high rate of comorbidity, found even in nonreferred populations, may have several sources. Many traditional nosological entities draw their definition from clinical experience with adults; although childhood analogues clearly exist, the appropriate descriptive boundaries may not be the same. Furthermore, impairing symptoms in childhood tend to have a ramifying effect by interfering with the ongoing acquisition of key developmental skills in multiple areas. Finally, regardless of whether the origins of a given form of childhood psychopathology lie in biological factors, family or social environment, or an interaction between these realms, these pathogenic forces often produce symptomatic difficulties that cut across many diagnostic lines.

Distinctive Aspects of Child Assessment Save for patients brought involuntarily for evaluation, most adults coming for psychiatric assessment acknowledge, at least implicitly, some degree of self-perceived distress and wish for help, no matter how ambivalent the wish may be. In contrast, for most children, the responsibility for initiating and following through with a psychiatric evaluation lies with the parents. In many cases, children's behavior is a greater source of distress to others—parents, teachers, peers—than to the children themselves. Other children, even while acknowledging distress, view the locus of the troubles as purely external, and hence see little need for evaluation, even if they comprehend the process. In any event, the lack of any stated wish for help or active role in seeking the evaluation colors the child assessment process from the start. As a result, explicit attention must be paid to assessing the child's view of the problem and to cultivating the child's engagement in the process.

A related aspect of the child psychiatric assessment involves the need for the clinician to assess and accommodate to limitations in the child's ability to reflect, conceptualize, and report experiences and internal states. These capacities emerge only gradually with maturation and are influenced by both normal and pathological factors. Thus, the clinician's communication must be attuned to the child's developmental level and may require play, stories, drawing, and other alternative modes of interaction in addition to direct discourse.

Context and the Need for Multiple Informants Even more so than for an adult, the child's functioning and psychological well-being are strongly related to the social context—the family, school, and community settings—in which the child lives and develops. Thus, to evaluate the nature and severity of the child's psychiatric status, information is needed about the child's relationship to, and functioning in, these diverse settings. The fit of the child and social environment can be appraised from what parents say about their relationships with the child and from the child's direct report, as well as from play and drawings; however, direct observation of the content and tone of parent-child interactions in the office setting also provides much useful additional data. In some cases, a home visit may provide invaluable information.

It is also important to gather information on the child's functioning in school, a task that often requires speaking with the child's teacher or guidance counselor, in addition to reviewing school reports and educational assessments. Here too, direct classroom observation can provide invaluable in vivo information. For children

involved with child welfare, juvenile justice, or other institutional systems, information from caseworkers, probation officers, and institutional caretakers is essential.

Special Types of Assessment The main emphasis of this chapter is on the assessment of children and adolescents, especially those capable of participating verbally in the assessment process. Because developmental considerations shape the evaluation process, infant and toddlers on one hand and adolescents on the other require specific modifications of the general approach outlined in this chapter (see sections below on these age groups). Nonverbal or language-impaired children, such as those with pervasive developmental disorders or mental retardation, also require specialized assessment techniques.

Although the prototype for clinical assessment of the child is the office consultation or clinic visit, evaluations occur in a variety of settings including schools, pediatric wards, residential treatment centers, detention centers, and hospital emergency rooms. Procedures necessarily vary with the setting and reasons for carrying out the assessment. The following general model for the outpatient evaluation can be adapted to fit more-specific clinical situations. More-specialized types of evaluation are discussed in a separate section below.

Preliminary Contact In the initial phone call to the clinician or clinic intake office it is helpful to gather some basic information about the nature of the chief complaint; to set the time, date, and place of the first appointment; and, if relevant, to discuss issues of fee or insurance coverage.

During the evaluation the clinician will want to see the parents alone, the child alone, and the parents and child together. The order in which these should be done and the corollary issue of who will attend the first meeting should be discussed at the initial contact. This decision is usually made with reference to the child's age. When evaluating a young child, the parents are often seen first, without the child. This allows the clinician to obtain a history and to discuss how to prepare the child for the evaluation. On the other hand, the adolescent should usually be included in the initial interview so that the youngster does not identify the clinician as an agent of the parents, which may interfere with the adolescent's willingness to talk with the clinician and impair the formation of a treatment alliance. If children or adolescents are to attend the initial appointment, the clinician should discuss in the initial call how they are to be prepared for the appointment.

In the initial phone contact the issue of which adults will attend should also be discussed. When the family is intact, the presence of both parents is optimal. When this is not the case, the clinician and parent should discuss whether a noncustodial parent, stepparent, or other adult should be present.

Parent Interview In the initial parent interview the clinician seeks to understand the parents' view of the problems that have led to the referral, including the nature of the child's current difficulties; the explicit and implicit reasons for the timing of the referral; and the impact of the child's difficulties on the parents, individually and as a couple, as well as on the well-being and functioning of the family as a whole. A detailed history of the child's development and a review of family medical and psychiatric disorders are obtained. Finally, the clinician seeks to understand how the family is functioning in their community and cultural setting. This includes a careful appraisal of the strengths of both the child and the family, in addition to the history of the presenting symptoms and areas of impairment.

Child's Current Difficulties and Reasons for Referral In child psychiatry, referrals are usually initiated by the parent or parents rather than the child. Similarly, parents are key in determining whether a recommendation for treatment is followed. Over and above the need to gather information about the child's symptoms and functioning, the parent interview provides an opportunity to develop an alliance with the parents around the goals of identifying and helping the child's difficulties. Success of this process requires understanding the parents' expectations and concerns regarding the evaluation and its outcome. How these issues are addressed will depend on the nature of the child's problems and the parents' knowledge of and previous experience with psychiatric evaluations and treatment. Attention to the parents' view of the evaluation process allows the clinician to address unrealistic fears or excessive expectations about what can be accomplished in a given time and to reduce family stress when parental anxiety is too high.

In taking a history of the current problem, the clinician obtains a detailed picture of the parents' current emotional and behavioral concerns about the child. Parents should be asked to describe specific instances of the child's problematic behavior, frequency, intensity, duration, circumstances in which the behavior occurs, their responses, and the child's responses to it. The extent to which the symptoms cause functional impairment is ascertained by asking about the level of the child's distress, interference with social and academic activities, impact on ongoing development, and the effect of the child's behavior on others. The clinician should also inquire directly about behaviors or symptoms not reported by the parents but known to be commonly associated with the presenting problem or with disorders that the presenting problem may suggest. Thus one goal of the initial parent interview is to develop a detailed picture of the presenting problems, their course, their effect on functioning and on the family, and what has been tried to help alleviate them.

The clinician is also interested in the implicit and explicit reasons why the child has been brought for help at this particular moment. Often the reason lies in an escalation of troublesome behavior or in the deleterious impact of a deficit emerging into sharper relief (e.g., a learning disability or attention deficit becoming more apparent as the child faces the increased demands of a new grade). In other cases, the child's behavior may not have changed, but some change within the family or loss of compensatory support may have made it impossible for the parents to tolerate or ignore previously bearable behavior. These iatropic factors (factors propelling families to seek clinical attention) are often complex. In surveys of the general child population, rates of diagnosable psychiatric disorder (defined as meeting both diagnostic and impairment criteria) approach 20 percent, yet the number of children actively seeking or receiving clinical services is far lower. Perceived barriers to services alone do not account for this low rate of seeking help.

Throughout the history taking, the clinician strives to understand the meaning and function of the symptoms in relation to the factors in the child and environment that influence them. A given symptom (e.g., a period of anxiety, theft, or hallucinations) may have quite different meanings, functions, and clinical implications in different children. Thus, detailed information on precipitating circumstances, behavioral concomitants, and consequences and alleviating factors is needed, much as is the case in taking a medical history. However, this information must be interpreted within the broader developmental and family contexts in which the symptoms occur; thus while gathering history, the clinician simultaneously takes in information about these contextual elements. History taking and diagnostic formulation are not distinct processes; during the interview process clinicians continually formulate and test tentative hypotheses that guide further questions and the diagnostic possibilities under consideration.

To understand the child's presenting problems, developmental status, and family and social context, the clinician must learn about the child's strengths, interests, talents, and areas of adequate or superior adjustment. Such information is essential for a full diagnostic understanding of the child and helps the clinician appraise the child's overall adjustment and functioning. Moreover, in planning for treatment, factors that may help counterbalance or remediate the child's vulnerabilities must be identified. Finally, the clinician's appreciation and discussion of the child's strengths and areas of good functioning provides support and reassurance at a time when the child's problems and limitations have been the parents' major focus of attention.

The parent interview should also cover the practical and administrative aspects of the diagnostic assessment, including discussion of fees, scheduling, confidentiality, and permission to obtain information from school personnel and other clinicians. If the child was not included in the initial interview, the clinician should discuss appropriate preparation of the child for a meeting.

Developmental History The developmental history is a detailed accounting of the child's development across physical, cognitive, linguistic, social, and emotional domains. One task of this part of the parent interview is to ascertain the timing and sequence of developmental events from pregnancy to the present. Parents vary in their ability to provide precise information on the timing of specific milestones; however, an adequate sense of the course of development can often be obtained by having the parents compare the patient's progress in a particular domain with that of siblings. Asking parents about the timing of changes in the family, such as moves or other memorable life events, may help provide a chronological framework for remembering features of the child's development. Besides gathering descriptive information about the child's development, this part of the interview often allows one to tap the emotional aspects of the parent-child relationship by querying the parents' hopes, worries, expectations, and life circumstances in relation to different developmental events.

It is useful to conceptualize the child's past and current development, in the following domains.

COGNITIVE AND ACADEMIC DEVELOPMENT The child's pattern of cognitive strengths and weakness should be traced from early childhood, including verbal and attentional skills. While academic performance is important (including areas of particular strength and weakness), important information about aspects of emotional and social development also flows naturally from the discussion of how the child progressed from one school year to the next. Thus, in tracking cognitive development and academic success, ask about the child's ability to separate from parents and to attend school regularly, interpersonal relationships with peers and teachers, motivation to learn, ability to function independently, tolerance for frustration and delay of gratification, attitudes toward authority, ability to accept criticism, etc. A grade-by-grade history of the schools attended can be obtained during this part of the interview, as well as the reasons for any changes. Retentions, the reasons for them, and the child's reaction at the time and subsequently should be noted. When the child's behavior or progress at school is among the reasons for the psychiatric evaluation, one should obtain permission to communicate with the child's teachers, counselors, or other school personnel and to review the school records, including

results of standardized tests.

FAMILY RELATIONSHIPS The parent interview should include assessment of how the child relates to each family member and how the child fits into the overall family system. The child's reactions to major family changes should be noted (e.g., deaths; birth of siblings; marital separation, divorce, or remarriage; and changes in caretaking arrangements, custody, or visitation), as well as parental responses to those reactions. Ask about compliance with family rules and standards as well as consequences when the child does not comply (i.e., usual mode of discipline or limit setting) and the child's response to such interventions.

PEER RELATIONSHIPS The clinician should gather information about how the child relates to peers, including the number of friends, preferences regarding age and gender of friends, and any major changes in peer group; the child's satisfaction with these relationships; their relative stability; activities and interests shared with peers; and parents' feelings about the child's close peers or lack thereof. The parents' perspective on the child's social skills and deficits should be assessed, including their sense about any difficulties the child has in this domain. For adolescents, this part of the history includes issues such as the capacity for intimate relationships, romantic interests, sexual activity, and concerns over sexual orientation.

PHYSICAL DEVELOPMENT AND MEDICAL HISTORY Assessment includes fine and gross motor development, toilet training and its lapses, eating behavior and attitudes, and sleep patterns. Precocious development or delays in physical growth or pubertal maturation should be noted. In this context, parents should be asked systematically about medication, illnesses, hospitalizations, operations, episodes of loss of consciousness, seizures, head injury, or other serious injuries, as well as the child's reactions to these and their impact on subsequent health and activities. Inquiry about tics, difficulties with hearing or vision, lead exposure, and other specific conditions may be suggested by the clinical presentation. When necessary, medical reports should be obtained and reviewed.

EMOTIONAL DEVELOPMENT AND TEMPERAMENT While some of these topics will have been covered in taking the history of the presenting problems, the clinician should fill in gaps with a systematic review. This domain of assessment includes the child's present and past mood and capacity for affect regulation, anxieties, and ability to adapt to changes and situations that are new, challenging, or frustrating. Prevailing mood as perceived by the parents should be assessed, and if the child has shown depressive symptoms, including irritability, the clinician should ask specifically about suicidal ideation, gestures, or attempts. Worries, fears, and other manifestations of anxiety should be covered and the level of any distress and impairment explored. The generalized or specific nature of symptoms, situational evokers, historical precipitant, and impact on the child are relevant. Unusual fears, excessive shyness or withdrawal, and obsessive or compulsive symptoms are also important.

The child's capacity to regulate aggression falls within this domain and should be reviewed. Ask about circumstances in which the child becomes angry or aggressive, the mode of expressing anger (i.e., verbal, physical, or both), the impact on others, and the child's reaction to and processing of such feelings and behavior. In this context the clinician seeks to evaluate how the child manages aggression, and whether the child is too aggressive, or too fearful of his or her own anger or that of others. Specific aggressive symptoms, (e.g., bullying, firesetting, vandalism, or cruelty to animals) should also be sought.

DEVELOPMENT OF CONSCIENCE AND VALUES The clinician should assess the development of conscience to gauge whether it is too harsh, lax, or overly focused on particular issues. The effectiveness of conscience in helping the child conform to family and societal expectations is important. Religious or ethical concerns and their concordance with those of the family can be addressed in this phase of history taking. The family's expectations with regard to the child's values and future life choices should be ascertained from the parents' perspective, and areas of potential discord noted.

INTERESTS, HOBBIES, TALENTS, AND AVOCATIONS Although this line of inquiry is also pursued with the child, it is useful to obtain the parents' perspectives on the child's interests and activities and to assess the parents' approval, involvement, and support for them. Indications of parent-child conflict should be probed and may provide useful information about the relationship. The presenting difficulties may have affected the child's ability to focus on or engage in previously enjoyable areas of interest or activity such as sports or music.

The quantity and type of TV programs, movies, and videos a child is permitted to watch provides information on both the child's interests and the quality of parental limit setting.

UNUSUAL CIRCUMSTANCES The clinician should assess the child's exposure to unusual or traumatic circumstances such as sexual or physical abuse, family or community violence, natural disaster, or armed conflict. If a history of such exposure exists, the child's immediate and subsequent reactions and the nature of the response from parents or other adults should be assessed.

Family and Community Background Children develop in the contexts of the family and community, and these provide opportunities and challenges that influence and shape the course of that development. The clinician should gain a picture of the child's functioning in these contexts and their influences on the child. While the clinician will learn a great deal about these areas from taking the history of the presenting problem and the developmental history, direct questioning may also be needed.

PARENTS The clinician should assess the parents as individuals and as a couple, including strengths, weaknesses, and areas of conflict or difficulty. If the child resides with someone other than the parents (e.g., a relative or foster parent) obtain information on the history and circumstances of that relationship, including how the child came to be in that adult's care. Determine what the child's previous experiences were with the parents and the nature of the bond with them. The clinician should similarly assess the current caregivers and their parenting relationship with the child.

If the child is adopted, obtain information on the circumstances, the parents' feelings and expectations about the adoption, the age at adoption, the child's adjustment at the time and subsequently, and how the child and family members currently view the fact of the adoption. Review how the adoption has been discussed with the child and any concerns or questions about the biological parents.

When taking the developmental history it is often natural to inquire about the parents' feelings and involvement with the child at different stages. Parents' past attitudes and behavior are likely to reflect on their responses in the current situation, hence, one wants to know how they felt about past difficulties, and what they have done to help and support the child. The parents' individual views on the causes of the child's difficulties, and the extent to which they agree or disagree on this and other aspects of the child's care and management affect how they will respond to the clinician's recommendations. Information on the parents' own backgrounds and histories with their families of origin will help the clinician to understand and work with their responses to their child.

Ask about the parents' ethnic and religious backgrounds. They may be a source of shared identification and values to be imparted to the child or a source of conflict between the parents or between the adolescent and the parents. A clinician should be sufficiently familiar with the culture and language of the family to understand the child's and family's situation in that context.

The parents' education and occupation may bear on how they understand and respond to the evaluation. How knowledgeable and worried they are may depend on their past experiences and familiarity with psychiatric issues. Financial resources and insurance coverage should be ascertained because this may be a source of stress or worry and may pose practical limitations to the treatment options available to the family.

The clinician should gain an understanding of how the family system functions, with whom the child lives, and the nature of relationships between household members and the child. Explore boundaries and alliances within the family and the child's affinities and conflictual relationships with parents, siblings, and other household members. Assess family communication and problem solving, including how issues of separation or disagreement are handled. The emotional tone of the family (e.g., tense, anxious, critical, warm, supportive) can be ascertained by observing interactions and by direct inquiry. Stresses on the family as a whole or individual members (e.g., moves, migration, illness, accidents, job loss, legal difficulties) should also be addressed.

COMMUNITY The clinician should understand the community or neighborhood in which the family lives and their feelings and identification with it. Information on the family's involvement with civic, community, and religious activities (and the extent to which the child participates) should be ascertained. The clinician will also want to know about issues of safety and adversity (e.g., poverty, poor housing, high rates of crime or urban violence) in the neighborhood. In some communities, mental health resources are viewed positively; in others their use is stigmatized, which may adversely affect the child. These issues should be assessed and understood because they may affect how and where the child receives any indicated treatment.

Family Medical and Psychiatric History A careful family medical and psychiatric history is needed to identify any medical or psychiatric disorders with potential genetic or environmental implications for the child. Among those to be routinely considered are psychotic disorders, mood disorders, anxiety disorders, tic and

obsessive-compulsive spectrum disorders, alcohol or substance use disorders, attention-deficit disorders, learning disorders, and antisocial personality disorder as well as metabolic and neurological disorders. The impact of any family members' symptoms, hospitalizations, or incarcerations should be probed, with particular reference to the child.

Special Issues in the Parent Interview

INFORMANT DIFFERENCES While parents' concerns usually play a major role in initiating the evaluation, parents may perceive the child's problems differently and may disagree about their severity and the need for treatment. Parents' accounts of the child's difficulties and of various aspects of the developmental history may not agree completely with each other or with those of the child, teachers, or records of past events. To note these discrepancies is not to denigrate the reliability of parents as informants, but rather to emphasize the necessity of multiple informants.

Informants provide differing accounts of children's problems for a number of reasons. Thus, discrepancies should be noted and regarded as potentially useful data regarding what they reflect about the clinical situation. Informants may interact with the child in different settings that vary in the extent to which the child expresses disturbing feelings and problematic behavior. This is commonly true for behaviors expressed most dramatically in settings that challenge the child in vulnerable areas (e.g., the school's demand that a child with attention-deficit/hyperactivity disorder sit still and pay attention). Information about such situation-specific symptoms is best provided by those who spend the most time with the child in that setting. These different perspectives on the child's behavior may also have led to tension between informants who see the child discrepantly. (This is often true with divorced or separated parents whose child may show quite different behavior in the context of each parent's home). The clinician should be aware of these different perspectives and the feelings they generate so that the focus is maintained on evaluating and providing services for the child, with adults working together rather than at cross purposes.

Another reason for discrepant reporting involves differences in how the informants evaluate the child's behavior. Observers may judge the same behavior by differing standards or expectations for many reasons. The clinician's appreciation of the informant's psychology or cultural background will be useful for understanding and working with such differences. For example, when one parent is overly worried about the child having a serious disorder and the other is too ready to chalk up symptoms as something the child will “just grow out of,” the clinician must navigate carefully to allow both parents to move to a reasonable middle ground that will best serve the child's needs. In such instances, understanding each parent's defensive style will help the clinician avoid appearing to collude with one parent and alienating the other.

Discrepant reporting also results from differences between informants in their willingness or ability to report their views to the clinician. For example, parents may be quick to report a behavior that they find disturbing or annoying, while a child may be unable to describe the problematic behavior verbally, or may not appreciate the level of disturbance it creates for others, or may not report it because of feelings of shame or fear of being blamed or punished.

Generally speaking, both methodologic studies and clinical experience suggest that parents are more likely than children to report disruptive or externalizing behaviors such as restlessness, impulsiveness, oppositionality, or aggression. Conversely, children may be more likely to report anxious or depressive feelings and symptoms (including suicidal thoughts and acts) of which the parents may be unaware. Because children may be immature in their capacity to report information, such as frequency and duration of symptoms, parents are usually more accurate informants about chronological details. The child, however, may be the only source of information on events such as sexual abuse, exposure to violence, or behavior that occurred in peer contexts. Generally children become more reliable reporters of specific symptoms with age, with prepubertal children tending to be less reliable reporters of symptoms than their parents.

Mothers and fathers also differ in their experiences with their children, depending on the child's age and gender, shared interests, parental availability, and the respective temperaments and personalities of the parents and child. Although mothers have been reported to be more reliable informants than fathers, in practice it is wisest, both for gathering the fullest information possible and for fostering an alliance with the parents, to involve both parents from the outset. When appropriate, other informants such as siblings, grandparents, or child caregivers, may be called on to contribute information, but only after the implications for the family system and the child's specific relationships with household members are carefully thought out.

Parents and teachers may view the nature and severity of the child's difficulties differently because of the situation-specific nature of some presentations. When present both at home and at school, symptoms such as hyperactivity and inattention may carry a different clinical significance and prognosis than those noted in only one (but not both) setting. Parents and teachers may also differ in their appraisals of the child's behaviors. For example, a stressed and preoccupied mother may be oblivious to her child's behavior while an observant teacher with years of experience with similar-aged children may readily identify the child as distressed or disruptive.

For the clinician conducting the assessment, the issue is not which informant is “right” but how to understand the differences in viewpoint most effectively for the patient's benefit. Thus, eliciting and synthesizing information from multiple informants is a critical aspect of the clinical evaluation of children and adolescents.

INTERVIEW TECHNIQUES WITH PARENTS The clinician should use a variety of interview techniques to elicit information from parents. The type and quality of information obtained is in part determined by the methods used and the clinician's skill and comfort in using them. A series of empirical studies confirm that systematic factual information and details about onset and timing of symptoms are generally most easily elicited by specific and direct questions, while an understanding of feelings and relationships in the family is usually gained most effectively by means of more open-ended, indirect questioning strategies.

In the initial interview, parents should have the opportunity to tell their story in their own way. For rapport to develop they must feel that they have been able to express their concerns in their own words and were heard and empathically understood. An open-ended beginning also avoids narrowing down too quickly on a single aspect of the child's situation and conveys the clinician's interest in many different aspects of the child's situation. Subsequent probing can clarify details and help shape the clinician's understanding of diagnostic and potential treatment issues.

Parents are most likely to be nondefensive and comfortable about sharing personal information when they feel that the clinician is trustworthy, interested, understanding, and nonjudgmental. Clinicians must thus convey respect and empathy and avoid seeming aloof or excessively familiar. They should discuss the child's and family's situation in a natural way, using words and concepts that the parents can understand. The range of questions and the tone of the inquiry will convey the clinician's interest in the whole child (including strengths, interests and talents) and an appreciation that the child is not merely the repository of a set of symptoms.

EXPECTATIONS AND ANXIETIES Parents may hold strong preconceived ideas about the nature of their child's problem and the kind of help needed. Because of the extensive coverage of issues relating to psychotropic medications in the popular press, many parents have drawn their own conclusions about the appropriateness of medications both in general and in relation to their child's needs. For example, parents may arrive convinced that their child needs fluoxetine (Prozac) or that methylphenidate (Ritalin) is an addicting and inherently harmful drug.

Preconceptions or parental anxiety or conflict may also be played out around the issue of blame for the child's problems. In such situations, a parent may maneuver to have the clinician arbitrate or to agree that the teacher, school, or other parent is responsible for the child's behavior and distress. The clinician should be alert to such situations and handle them without making either parent feel defensive or discounted. If ignored, such hidden or overt agendas can have a deleterious effect on the assessment process and the likelihood that treatment recommendations will be accepted.

Parents may bring other entrenched, preconceived agendas to the clinical evaluation process. A parent may insist that the evaluation will confirm that the child has been abused (e.g., by an ex-spouse) or does not have a psychiatric problem but a medical one. Alternatively, the parent may demand that the clinician agree that there is nothing wrong with the child. In such instances, clinicians must explain carefully and clearly that they will conduct a thorough assessment and provide thoughtful professional advice, adding that the assessment must be done openly, without preordained conclusions. This may not satisfy some parents, and some situations may simply not be workable.

Bringing a child for psychiatric evaluation is a charged and emotionally difficult step for nearly all parents, and it is only taken after much worry and disappointment that home remedies or school-based adjustments have not sufficed. Our children are the repositories of our dearest hopes (and often our darkest fears) for ourselves. Although a parent may rationally conclude that seeking consultation is reasonable and necessary, admitting that one's child (or oneself) needs help is always painful. Submitting one's child and one's parenting to professional scrutiny stirs many anxieties, as well as positive expectations of help, even in the most enlightened parents.

Parents often bring unspoken anxieties about the child's possible diagnosis or prognosis to the evaluation. These anxieties may be based in the parent's own personal or family experience with psychiatric disturbance or may derive from accounts in the popular media. For example, is the child's uneven development a sign of

autism or retardation? Do the facial grimaces mean the child is developing a disabling tic disorder? Is the child's moodiness a harbinger of a serious depression or suicide? Other parental anxieties may be guilty ones. Is failed parental rearing or transmitted heredity somehow to blame for the child's difficulties? Is the child fated to share the parent's (or other relative's) own painful struggle with depression, anxiety, or psychosis? Does the parent' relationship to the child seem to recapitulate some distressing aspect of the parent's own early family experience?

The clinician must be aware of such parental anxieties and remain alert for hints of them as the evaluation progresses; some parents are too anxious or ashamed to be forthcoming about them. By and large, it is best to have these concerns aired, even if they cannot be settled at the initial stage of the evaluation and perhaps will not be settled for some time. When such anxieties are revealed or suggested indirectly, the clinician must express an interest in them with sensitivity. Respect and concern for these anxieties can be conveyed without dismissing them or providing premature or unfounded reassurance. The clinician may need to acknowledge the limits of knowledge about cause or prognosis in individual cases. Parents often seek psychiatric assessment after years of worry, having been given wellmeaning (but inappropriate) reassurances that the child would grow out of it. Frank discussion of these issues may provide some relief to the parents, albeit perhaps not definitive answers to their questions.

PREPARATION OF THE CHILD At some point in the initial interview, the clinician and parents should discuss the preparation of the child for his or her first appointment. Children have many reactions to a first visit to the psychiatric clinician, and parents must anticipate their child's possible reaction and prepare their own response. Children may feel anxious because they cannot envision what will transpire or because the evaluation may feel like a punishment. Some children fear that the assessment will reveal some profound or irremediable problem (this may be projected on to the clinician or parent with the reproach “You think I'm crazy”). Many younger children fear shots or having to get undressed for a physical examination. In most cases, the child is probably aware of many of the parental or school concerns that have led to the evaluation and may feel a sense of relief (which can nevertheless coexist with anxiety) about the possibility of getting some help. The child's anxiety or defensiveness may be reduced if the purpose of the visit is put in supportive (rather than pejorative or accusatory) terms. The parents might be encouraged to tell the child something like, “We know you've been having a hard time at school (or feeling upset and sad, or having trouble with friends), so we are going to see Dr.——, who is used to helping kids with that sort of problem and can maybe figure out what would help.” For younger children, explanatory terms such as a “feelings doctor” or a “talking doctor” who helps children with problems and worries may be useful. The parents must give permission and encourage children to tell the clinician about their concerns, even those that might seem to involve private family matters not otherwise spoken about.

Child Interview In the child interview the clinician directly assesses the child's psychological functioning, developmental status across different domains, and perceptions and experiences of the presenting problems. Like the parent interview, the child interview also provides an important opportunity to establish rapport. Such rapport will reduce the level of stress experienced by the child during the assessment process, enhance the child's engagement and willingness to share information, and facilitate the child's cooperation with subsequent recommendations for intervention. Thus, from the outset, the clinician should convey empathy and interest in understanding and helping the child.

There has been debate regarding the reliability, validity, and clinical utility of various approaches to the child interview, but it is generally agreed that the direct child interview is an essential component of a comprehensive clinical child assessment. It provides both observational data and diagnostic and historical information that is not accessible from other sources. In particular, information about feelings and moods, level of distress, and the child's view of significant events is essential for a full diagnostic understanding of the child's problems. For their own reasons, parents may not understand or fully appreciate the importance of these issues, or the child or adolescent may not have divulged them. The child may be guarding important information, such as secrets about abuse, sexual or antisocial activities, suicide attempts, pathological patterns of eating, or collusive family dynamics. Certain symptoms such as compulsions, premonitory urges, obsessive or suicidal thoughts, or hallucinations may be apparent only to the child. Other symptoms may escape or mislead the untrained eye; detection of tics or a thought disorder may require direct clinical observation.

There are many varieties of child interviews, depending on the nature of the presenting problem, the clinical setting, and the purpose of the assessment. Thus the scope, focus, duration, setting, and predominant techniques used will vary, depending on whether the assessment is an emergency department evaluation of a suicidal threat, a pediatric ward consultation of a regressed or uncooperative child, a forensic custody evaluation, an investigation of alleged child abuse, or the evaluation of a child who is academically underachieving, anxious, or manifesting tics.

The clinician approaches the child interview with a developmentally based understanding of child and adolescent functioning and psychopathology. Within this framework, the clinician expects to obtain information relevant to the diagnostic questions raised by the presenting problems, and so will systematically review the major domains of symptomatology. The order and manner of conducting the child interview vary with the child's age and developmental status, the nature of the presenting problems, the setting and context, and the clinician's style. The constant elements in the interview process are the orienting framework chosen and the clinician's clinical skill in engaging and eliciting relevant data from children of varying ages and capacities, with diverse problems and concerns. Some data may emerge spontaneously; some require questioning or other deliberate means of eliciting information.

Components of the Child Interview The components of the child interview are often conceptualized under two broad headings: history taking and the mental status examination. *History taking* involves inquiry into the significant areas of the child's life and functioning, past and present, including the presenting problem. The *mental status examination* consists of an assessment and description of the child's appearance and functioning as manifested in the interview. While these aspects of the child interview can be conceptually distinguished, they usually proceed concurrently. For example, the interviewer may ask a specific question such as who lives at home with the child or how the child gets along with a sibling or a teacher. From a history-taking point of view, the child's response provides some information about these facts, as well as how the child feels about them. While the child's response is thus of interest for the historical information it reveals, the child is also providing the clinician with data relevant to the mental status examination (e.g., how the child conceptualizes the response; how fluent the child is; how readily the child engages; how confiding or suspicious the child is toward the interviewer). Similarly, while engaging a young child in play with figures or puppets, the clinician can simultaneously gather information for the mental status examination (e.g., speech, relatedness, conceptual skills) and inferential data about the child's feelings, fantasies, and conflicts.

Several distinctive characteristics of clinical work with children bear on the manner of conducting the child interview and mental status examination and on interpreting the data they yield. The focus of both history taking and the mental status examination of the child is developmental, that is, it seeks to describe the child's current presentation in various domains and to compare the child's functioning in these areas with that expected for the child's age and phase of development. Another distinctive aspect of the child interview involves the child's potential lability and propensity to react in more immature ways when tired, sick, anxious, or stressed by an unfamiliar situation. While one can sometimes gather sufficient information for an assessment from a single interview, usually more than one interview is needed to assess the child's optimal or characteristic level of functioning.

Mental Status Examination For the mental status examination, the child psychiatrist observes and assesses the following areas.

Physical appearance—including physical development, presence of minor congenital anomalies, style of hair and dress, cleanliness, and other indicators of the quality of self-care or parental attentiveness (or its lack) to the child's physical appearance and hygiene.

Manner of relating to examiner and parents—including ease of separation, how confiding or guarded the child seems toward the clinician, eagerness to please or impress, flirtatious behavior, readiness to make excuses or minimize, passive withholding, or open defiance. The child's behavior toward the examiner is compared with that observed or reported toward the parents or other adults. Reactions to the clinician's physical appearance, gender, or ethnicity (especially if different from that of the child) should be noted.

Affect—the child's predominant mood and range of emotion during the interview and appropriateness of mood.

Coping mechanisms—as manifested toward the clinician, expressed in play, or described by the child in relation to important others; age-appropriateness of the child's dependent longings, sexual interests and impulses, and aggressive feelings with respect to intensity, object, and mode of expression; child's control or modulation of such urges (e.g., finding alternative or socially permissible means of satisfying them); strategies used for coping when anxious or frustrated.

Orientation to time, place, person:—whether children are accurately aware of the date, where they are, and who they are. This can be evaluated by asking the child “What is the year? season? date? day? month?” or “Where are we?” (Name of building, town, state, country).

Motor behavior including activity level—coordination, presence of unusual postures or motor patterns (e.g., tics, compulsions, or stereotypies, such as hand flapping or

twirling).

Quality of thinking and perception—including presence of hallucinations, delusions, thought disorder, flight of ideas, and adequacy of hearing and vision.

Speech and language—including reading and writing. Articulation, inflection, pitch, rhythm and fluency of the child's speech; richness, limitation, or idiosyncratic aspects of vocabulary and syntax; presence of echolalia or persistent neologisms; misuse of pronouns and gender; delayed or deviant social use of language; and lack of accompanying nonverbal communicative behavior (e.g., lack of facial expression or eye contact with interlocutor). When evidence of delays or problems in these realms exists, formal speech and language evaluation by a qualified specialist may be warranted. Some of the instruments used are shown in Table 33-1.

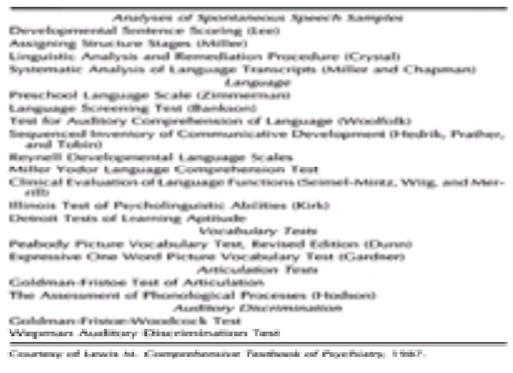

Analyses of Spontaneous Speech Samples
Developmental Sentence Scoring (Lee)
Assigning Structure Stages (Miller)
Linguistic Analysis and Remediation Procedure (Crystal)
Systematic Analysis of Language Transcripts (Miller and Chapman)
Language
Preschool Language Scale (Zimmerman)
Language Screening Test (Bankson)
Test for Auditory Comprehension of Language (Woolfolk)
Sequenced Inventory of Communicative Development (Hedrik, Prather, and Tobin)
Reynell Developmental Language Scales
Miller Yodor Language Comprehension Test
Clinical Evaluation of Language Functions (Seinsel-Mintz, Wiig, and Merrill)
Illinois Test of Psycholinguistic Abilities (Kirk)
Detroit Tests of Learning Aptitude
Vocabulary Tests
Peabody Picture Vocabulary Test, Revised Edition (Dunn)
Expressive One Word Picture Vocabulary Test (Gardner)
Articulation Tests
Goldman-Fristoe Test of Articulation
The Assessment of Phonological Processes (Hodson)
Auditory Discrimination
Goldman-Fristoe/Woodcock Test
Wepman's Auditory Discrimination Test

Courtesy of Lewis Ist. Comprehensive Textbook of Psychiatry, 1985.

Table 33-1 Language and Speech Tests

Overall intelligence and fund of knowledge—capacity for classification, abstraction, and inference appropriate to age; counting, alphabet, days of the week, months of year, reading, writing, computational abilities; geographical and historical facts.

Attention, concentration, impulsivity—level of attention to an activity or topic of discussion, degree to which child jumps from activity to activity; need for physical limits; distractability (e.g., by outside noises).

Memory—immediate short- and long-term memory (e.g., repeat three items and recall at 5 and 30 minutes).

Neurologic Functioning—fluctuating alertness, tremor, nystagmus, choreiform movements, abnormal gait, neurological soft signs, cerebral dominance (preferred hand, foot, eye, ear); testing for neurological soft signs can be incorporated in a game-like fashion (e.g., ball throwing, walking on a line, heel-to-toe walking, evidence of hand overflow when walking on outside of feet, and timing 20 repetitions of finger tapping, successive finger tapping, alternating pronation-supination of the hand are reliable components of the revised Neurological Examination for Subtle Signs [NESS]).

Judgment and insight—especially concerning the presenting problem; this is most usefully judged after the child has had the opportunity to develop some rapport with the clinician, as the child's initial impulse may be deny or minimize the parents' presenting concerns.

Preferred modes of communication—Is the child open to talking directly about the presenting problems and the significant aspects of his or her life or is the child more comfortable with indirect modes of expression such as play or drawing?

Reliable detailed assessment of a child's speech and language capacities, intelligence, academic attainment, attentional and executive functioning, and memory requires standardized psychometric testing (see section on psychological testing below). Some simple tasks, such as those used in the Mini-Mental State Examination (as adapted for children by R.A. Ouvrier and colleagues (Table 33-2) can be incorporated in a game-like fashion into the interview of the child to provide a rough screen of higher mental functions such as orientation, attention, memory, language, and constructional ability.

Orientation		Score	Points
1. What is the	year?	_____	1
	[season]?	_____	1
	date?	_____	1
	day?	_____	1
	month?	_____	1
2. Where are we?	Country	_____	1
	State or territory	_____	1
	Town or city	_____	1
	Hospital or school	_____	1
	Floor or address	_____	1
Registration			
3. Name three objects, taking one second to say each. Then ask the patient all three after you have said them. Give 1 point for each correct answer. Repeat the answers until patient learns all three.			3
Attention and Calculation			
4. Serial sevens. Give one point for each correct answer. Stop after five answers.			5
5. Spell WORLD backwards.			5
Recall			
6. Ask for names of three objects learned in Q.3. Give one point for each correct answer.			3
Language			
7. Point to a pencil and a watch. Have the patient name them as you point.			2
8. Have the patient repeat "No ifs, ands or buts."			1
9. Have the patient follow a three-stage command: "Take a piece of paper in your right hand. Fold the paper in half. Put the paper on the floor."			3
10. Have the patient read and obey the following: "CLOSE YOUR EYES." (Write in in large letters.)			1
11. Have the patient write a sentence of his or her choice. (That sentence should contain a subject and an object, and should make sense. Ignore spelling errors when scoring.)			1
12. Have the patient copy the design printed below. (Give one point if all sides and angles are present and if the intersecting sides form a four-sided figure.)			1
	Total		30

Reprinted with permission from R.A. Ouvrier, M.D., Professor of Paediatrics, The Children's Hospital at Westmead, Sydney, Australia.

Table 33-2 Mini-Mental State Examination for Children

Although the clinician should usually evaluate all areas of the mental status assessment, the emphasis and detail given to the different elements varies with the presenting problem and the type of assessment. For example, in an agitated, potentially delirious child on a pediatric ward, particular attention is paid to the child's orientation, alertness, memory, verbal coherence, neurological functioning, and the possible presence of hallucinations.

Child Interview Techniques A range of techniques is used in the psychiatric interview of the child or adolescent. The choice and timing of techniques depends on the child's developmental, cognitive, and linguistic level; the emotional difficulty of the issue being addressed; and the degree of rapport between child and clinician. Being able to gauge the potential utility of a technique at a particular moment and feeling comfortable with shifting from one technique to another as the occasion demands are important skills for those conducting child and adolescent clinical assessments.

PLAY TECHNIQUES Children under 7 years of age have limited capacities to verbally recount their feelings or interpersonal interactions. For these younger children as well as a number of older ones, play is a useful adjunct to direct questioning and discussion and is often a less challenging mode for the child. Some children find it easier to communicate in displacement; thus, imaginative play with puppets, small figures, or the interviewer can provide useful inferential material about the child's concerns, perceptions, and characteristic modes of regulating affects and impulses.

The skilled interviewer can facilitate the child's engagement in play, without prematurely introducing speculations or reactions that might distort or cut short the presentation of certain types of material. During the course of play, the clinician follows the sequences of play content, noting themes that emerge, points at which a child backs away from the story line or shifts to a new sequence or activity, and situations in which the child gets stuck or falls into a repetitive loop. To facilitate the play component of an interview, the interview room should have a supply of human and animal figures or dolls and appropriate props. These should be relatively simple, since elaborate toys can become distractions rather than serve as vehicles for the expression of the child's fantasies. Stock characters (such as Barbie or Disney characters) may impose their own specific story lines and thus limit access to the child's own concerns.

Not only is the content of the child's play a rich source of information, careful observation of the form of play also provides important information for the mental status examination. During imaginative play, the clinician can observe the child's coordination and motor capacities, speech and language development, attention span,

readiness to engage the interviewer, capacity for complex thought, and affective state. Absence of imaginative play or limited, concrete, noninteractive play may indicate a pervasive developmental disorder.

Games such as cards or board games are also useful for putting the child at ease and developing rapport. These provide opportunities to observe the degree of the child's engagement in and enjoyment of the shared activity; how the child manages competition, including reactions to winning or losing; or whether the child is prone to cheating. Some play activities (e.g., throwing a ball back and forth or easy card games such as war) are simple enough to permit ongoing conversation, while helping to discharge tension and diminish the pressures of the interview situation. As with toys, elaborate games (e.g., chess) should be avoided; games that demand much cognitive energy and concentration usually preclude discussion of issues relevant to the assessment and may become a means of avoiding issues involved with the child's difficulties.

PROJECTIVE TECHNIQUES In addition to imaginative play, projective techniques often help provide an indirect picture of concerns that the child may be reluctant or unable to report directly. These techniques can help the child feel more comfortable with the clinician, are often experienced as fun, and may provide access to concerns that are important to the diagnostic formulation.

One commonly used technique is picture drawing. The child can be asked to draw a picture, leaving the choice of subject up to the child or, alternatively be given a specific request (e.g., draw a person or the child's family doing something). When the picture is finished or nearly so, it is helpful to compliment the child's effort and express interest in what is happening in the drawing. The child's elaboration provides information that may not be readily apparent from the drawing itself. Both the content and form of the drawing offer a window on the child's emotional concerns and aspects of intellectual and visuomotor development. For example, the relative size and placement or omission of family members in a family drawing may be important nonverbal indicators of the child's perceptions or feelings about the family. Aggressive or sexual themes may be reflected more readily in drawings than in words. Self-image may be indicated through depictions of the child as nonhuman, grotesque, inconsequential, or of the opposite gender. The clinician should become familiar with the developmental progression and norms for human figure details, such as limbs, joints, facial features, and clothing, which can provide a useful rough estimate of intellectual maturity. Various systems have been developed for systematically assessing the cognitive and emotional aspects of children's drawings. Also, the child's behavior and speech while drawing may yield useful information (e.g., throwing the picture away unfinished and saying that it is no good).

Frequently used verbal projective techniques are asking what animal the child would most or least like to be or whom the child would pick to take along to a desert island. It is useful to ask what the child would do with three magic wishes; if elaboration is needed, the clinician can explain that the wishes could be to have anything, to have the world be any way, or to change oneself in any way. Children's responses are often revealing. Some may needily or impulsively wish for material possessions, such as a videogame or a million dollars. Other responses may reveal longings to change distressing circumstances, such as “for my mom and dad to get back together again,” “not to have tics anymore or get teased about them,” or “to have a dad who doesn't yell at me all the time.” Still other children appear uncomfortable wishing for something for themselves, preferring instead seemingly altruistic responses, such as “no more poverty or wars.” Children's responses can be used as the starting point for further exploration. For example, the child who wishes for “a big house and lots of money,” may be asked who else would live there and what they would do. Children who wish for “no more fighting in the world” can be asked if there are some particular fights that they would especially like to stop.

Some interactive imaginative techniques useful to interviewers comfortable with them provide elements of playing a game, which may appeal to the child. The squiggle drawing game developed by Donald Winnicott consists of the clinician drawing a curvy line and asking the child to turn it into a picture of something; the child then draws a curvy line that the therapist elaborates, and so on, taking turns. The Despert fables are a series of incomplete affectively evocative stories that the child is asked to complete.

Asking the child to tell about a dream or a book, movie, or television program can provide information about the child's interests and preoccupations. (If the clinician is familiar with the plot, the distortions the child may introduce can be informative regarding the child's cognitive and emotional style.) Inquiring about what the child would like to do for a living when grown up can provide insight into the child's aspirations, values, and concerns as well as those of the family.

DIRECT QUESTIONING Direct questioning about the current problems and other aspects of the child's life requires tact, attention to the child's level of cognitive and linguistic development, and regard for the impact of the questions on the child's self-esteem. The goal of such discussion is to determine how children see their world and functioning in it, including attachments and antipathies, pleasures and anxieties, and strengths and weakness, which may include difficulties that led to the evaluation.

It is important to phrase questions in language that is appropriate to the child's developmental level and verbal capacities. Children may become confused or made more anxious by technical words or abstract or complicated questions. Closed questions often yield little information. For example, “Do you get along OK with your brother?” may bring forth a simple “yes” or “no,” while open-ended or descriptive questions are more likely to elicit a richer picture of how a child thinks and feels (e.g., “What sort of guy is your brother?” or “Tell me about your last play-date with Danny.”) Some usefully evocative questions include “What sorts of things make you mad (or make you feel afraid)?” “What is your saddest (or happiest) memory?” “What do you like to daydream about?” If it is important to establish the order or duration of events, school-aged children will do better if the questions are anchored to markers in their own lives (e.g., “Did that happen before or after your birthday?” or “Have you had trouble with that since school got out?”) Several empirical studies by Michael Rutter and colleagues have examined the effects of various interview styles on the information elicited from children and parents.

Although direct questions are often needed to initiate or focus discussion, settling into a direct question format may constrict the interview process and give children the impression that the clinician just wants to hear the facts and is not so interested in their own feelings and experiences. Young children frequently respond in a manner they believe to be socially desirable, while older children are often reluctant to acknowledge, even to themselves, feelings of sadness or vulnerability. Clinicians should convey, by demeanor and tone of voice, their interest in the child's real feelings.

When discussion of feelings seems difficult for the child, it may be useful for the clinician to acknowledge the link between a situation and the child's affective response (“Some kids might have felt mad if that happened to them; how about you?”). However, the clinician must avoid leading a child who is not yet well known by assuming experiences or feelings. Probing a bland, equivocal response may prove useful; for example, in response to “Okay, I guess,” the clinician may say, “Sounds like you're not quite sure.”

The assessment should convey respect for the child and protect the child's self-esteem. Many children feel ashamed or guilty about the difficulties that have brought them to the evaluation. When the clinician conveys a nonjudgmental understanding, using language that suggests that the difficulties are problems to be helped rather than transgressions, the child is more likely to be forthcoming. The child has come to the evaluation because of problems, but one should also talk with the child about strengths. For example, in inquiring about school, it may be useful to say, “Tell me the parts of school you like best” before asking “Are there subjects (classmates, teachers) that give you trouble?” or “What are the rough spots at school?” It may be better to ask the socially isolated child “Do you sometimes have trouble making or keeping friends as much as you'd like?” then “What do you think the trouble is?,” rather than “Do you have any friends?” (especially if the truthful answer to this question is “no”).

Structure and Sequence of the Child Interview When interviewing the child, the clinician has in mind an overall map of the functional and historical areas to be covered; however, the format and sequence of any given interview is responsive to the individual child, the presenting problems, the unfolding pattern of the interview, and the clinician's personal style. For example, in every assessment, the clinician should ask about the cardinal symptoms of depression, anxiety, and disruptive behavior disorders, but when, how, and in what order this is done, varies from one interview to another.

BEGINNING THE INTERVIEW There are various ways to start the interview. The one chosen should be comfortable for both clinician and child and set the stage for a productive interchange. Many clinicians prefer not to address the presenting problems at the very outset. Some clinicians, for example, remain quiet and let the child look about the room and begin playing or talking without prompting or guidance. This approach imposes a minimum amount of structure; the clinician can see how the child reacts to both the materials present in the playroom and to the clinician. It also provides information about how the child handles the situational anxiety that is often evoked by the first meeting with the clinician. However, this unstructured approach may leave many children disoriented and uncomfortably uncertain as to the clinician's goals and intentions. Other clinicians are more direct in attempting to place the child at ease and begin by inviting the child to explore the available play materials (“What looks interesting?”). Others open the interview by inquiring about neutral or pleasurable topics (“Tell me about what you like to do for fun?”). The clinician must read the child's cues regarding level of comfort and urgency in deciding when to address the problems that led to the evaluation. Some children will prefer to first talk about nonthreatening topics, while others expect to get down to business and talk about their problems right away.

Although it may be useful to establish some initial rapport around neutral or interesting topics, most children understand that they have not come for a conventional

social encounter. Hence, relatively early in the interview it is useful to ask what the child's ideas are about the purpose of the meeting. It is usually best to do this in a matter-of-fact manner by asking, “What did your mom and dad tell you [or what do you know] about why they wanted you to come and see me today?” If the child has already met with the clinician together with the parents as part of an earlier contact, the clinician will have asked something similar at that time and may have heard about the child's view of the presenting problem and its history. If so, the clinician can follow up by saying, “Last time, mostly your mom and dad said what they thought about things. Today, we have a chance to talk more about how things look to you. Are there any things we talked about then that we should talk about today?” Asking what the child knows or thinks about the purpose of the evaluation shows the clinician's interest in the child's view of the situation and allows the clinician to hear how the child describes or assumes responsibility for the problem (or repudiates it as someone else's concern). Asking for the child's view up front, before the child has become deeply engaged or focused on some aspect of the interview process, allows the clinician to address misapprehensions. It also provides an opportunity for the clinician to explain, in terms that can be understood and potentially acknowledged by the child, the issues (at least as the clinician understands them) that have prompted the assessment.

ASSESSMENT OF MAJOR REALMS OF FUNCTIONING Throughout the assessment, the clinician elicits and assimilates information to develop a picture of the child's functioning in the realms of family, school, peers, and recreation. Discussion of the presenting problem will reveal information, particularly with respect to impairments, in these domains; however, additional discussion of the child's experiences in these areas enhances the picture of the child's life and level of adaptation. It is often useful to begin exploration of each of these realms by asking for objective facts (e.g., “Who lives at home with you?” “Whom do you like to play with [or hang out with?]” “What school do you go to? What grade are you in?” or “What do you like to do for fun?”). This can be followed up with probes for more subjective views: “How do people in your family get along? Are there any problems at home? things people fight about? or things you get in trouble for?, or things people at home do that you don't like?).” This approach is intended to open up discussion of the major realms of functioning, not to simply elicit rote responses.

It is often helpful to begin by inquiring about the child's recreational activities (“What do you like to do for fun?”). This allows the clinician to learn about the child's interests, preoccupations, and talents and may provide a less threatening or problematic starting point than the areas of peer and academic functioning. Family functioning can be approached by asking who lives at home, how the child gets along with each person, and how family members get along with each other. One can also inquire more specifically “What do you like to do with your dad [or other relative]?” or “What is your brother like?” In the case of nontraditional family structures or families of divorce or separation, the clinician may inquire about the history of the current arrangement and what sort of relationship is maintained with the noncustodial parent.

INQUIRY ABOUT PSYCHOPATHOLOGICAL SYMPTOMS Just as a medical evaluation includes a review of key symptoms in the major systems, the child psychiatric evaluation should include a set of questions that screen for symptoms of common or major disorders. The family or child may not see a link between one behavior or area of distress and another and may not volunteer symptoms that seem unrelated to the presenting problem. Comorbidity is relatively common in children with psychiatric disorders; thus, the child psychiatric clinician must routinely question across the major areas of psychopathology. Using developmentally appropriate language, this inquiry should cover depressive symptoms (e.g., low self-esteem, anhedonia, and suicidal ideation or behavior); excessive anxiety or fears; obsessions and compulsions; hallucinations and delusions; and antisocial behaviors. For example, one should ask “Do you feel sad or grumpy a lot of the time? Do you often feel like crying?” If the answer is positive, one can follow up concerning persistence (“Does that last for more than a few minutes?”), frequency (“Does that usually happen every week [or most days?”]) and context (“What sorts of things make you feel that way?”). To inquire about suicidal ideation, one can begin by asking, “Do you ever feel that life is so hard, you wish you weren't alive (or would be better off dead)?” following up with the more explicit “Have you ever tried to hurt yourself (or done anything to try to make yourself dead?)” Exposure to potentially traumatic experiences should be routinely covered and developmentally appropriate language should be used (e.g., “Have any really scary things ever happened to you?”).

MENTAL STATUS EXAMINATION Throughout the clinical interview the clinician observes and notes information relevant to the domains of the mental status examination. Some abnormalities or deficits may be evident through this process; however, other areas will require direct examination, which can usually be incorporated in a playful way into the clinical interview. While intensive developmental, psychometric, language, and neurological assessment usually cannot be subsumed within the clinical interview, a brief semistructured evaluation requires only about 5 to 10 minutes and provides useful information. This generally involves drawing; copying a design; simple tests of orientation, recall, general knowledge, and language (e.g., naming three objects; repeating a phrase such as “No ifs, ands, or buts”; a three-stage command such as “Take a piece of paper in your right hand, fold the paper in half, put the paper on the floor”; a writing sample [such as asking the child to write a sentence of their choice]; and a brief screen of neurological soft signs).

CLOSING THE INTERVIEW In winding up the interview, the clinician can check for concerns that have not been covered and also signal that the interview is coming to a close by asking, “We've talked about a lot of things; are there any other things that would be important for me to know about you?” Similarly, one may ask, “I've asked you a lot of questions; do you have any questions you want to ask me?” The clinician should also give the child a general idea of what will be done with all the information the child has shared and some sense of what will happen next. For example, one may explain “I'm going to think over what I've heard from you and your folks and meet again with them [and you] to talk about what might help with .... Are there any particular things I should be sure to mention?” If the clinician already has a sense of what interventions might be helpful, this may be tentatively broached with the child, keeping in mind the parents' right to hear and decide on those recommendations.

SPECIAL ISSUES IN THE ASSESSMENT OF ADOLESCENTS

The clinical assessment of adolescents requires familiarity with the distinctive dynamics of adolescents and their characteristic patterns of managing conflicts, anxiety, and interactions with adults.

The adolescent's attitude towards the evaluation has important implications for how the assessment process unfolds. In contrast to the adult patient (who usually takes the responsibility and initiative for seeking help) and to the child patient (who is simply brought by parents), many adolescents are cautiously ambiguous as to whether there is a problem with which they want help or whether they are merely passively complying with their parents' initiative.

One important initial decision concerns who is to be seen in the initial interview: the parents, the adolescent, or all together. Seeing adolescents alone first has the advantage of emphasizing their active participation in the process and underlines the clinician's interest in their perspective; on the other hand, it may mean that crucial elements of the history are not available to the assessor. Furthermore, some embattled youngsters may protest that they are being identified as the patient or bearer of a problem that they feel resides in their parents or in the family as a whole. The clinician may to need to explain explicitly that the evaluation is not a judicial proceeding to assess blame, but rather an attempt to understand what the problems are.

In some cases, when the adolescent balks at being seen at all the clinician may need to meet with the parents alone to facilitate the youngster's participation in the evaluation. Regardless of when the adolescent is seen, the parents should at some point be seen alone to explain their concerns about the child, to obtain a full developmental and family history, and to hear about aspects of their spousal and parenting relationship that may be clinically significant but that may not be appropriate to discuss initially in the adolescent's presence.

Meeting with both the adolescent and parents together is usually helpful. This provides an opportunity to explore how family members interact, their convergences or divergences of perspective, them, and their capacity to work together therapeutically. Such conjoint interviews usually require active structuring by the clinician so that the purpose is framed as an occasion to talk together. Without such active intervention, the family session can easily deteriorate into an adversarial proceeding in which the parents complain about the adolescent, who in turn retreats into hostile or guarded silence.

Confidentiality The assessment of adolescents often raises complex issues of confidentiality. Although the parents and the adolescent must know that the adolescent's communications will be kept confidential, they must all understand from the onset that this does not extend to situations that pose a clear danger to the patient or others. For example, if the adolescent reveals significant suicidal ideation or a recent suicide attempt of which the parents are unaware, parents must be informed for the youngster's safety. Having made clear that the parents need to know, the task is then to negotiate whether the adolescent prefers to tell the parents directly or to have the clinician inform them. It is usually best to have the adolescent undertake this responsibility, but sometimes after discussion with the youngster, the clinician must be the one to speak with the parents, preferably in the adolescent's presence. Less clear-cut situations in which the adolescent reveals problematic, but not imminently dangerous, antisocial or sexual behavior or substance use must be decided on an individual basis. Treatment may not be possible without some confidentiality; on the other hand, confidentiality must not become collusion.

Developmental Issues Adolescents' usual strong push toward autonomy is often linked to a wariness of feeling vulnerable, dependent, or controlled. Tolerating feelings of ambivalence, guilt, or other painful affects or internal conflicts does not come easily to adolescents. In addition, adolescents are prone to externalize their concern or embarrassment over self-perceived shortcomings or difficulties. As a result, even adolescents who consciously desire help often approach a clinical

assessment with the fear that revealing what they perceive as shameful weaknesses will leave them vulnerable to criticism, control, or regressive dependency. Fear of acknowledging vulnerability or dependency may result in bland denials of any problems or insistence that “I can handle it by myself” or “I don't need a shrink; I can talk about my problems with my friends.” Externalizing their conflicted feelings about autonomy versus dependence, adolescents may fluctuate between grievances about parental overprotectiveness and complaints about perceived lack of care. Reversal of affect and counterphobic defenses may mask anxiety, shame, or other painful affects. The skilled clinician learns to look beyond surface behavior and to notice when an adolescent “doth protest too much.”

Although many adolescents seem intolerant of self-reflection, an unrealistic belief in the omnipotence of thought often helps support denial and avoidance. Even when confronted with a pattern of long-standing maladaptive behaviors, the adolescent may insist “I can stop any time I really want to” or “I don't do that any more” (i.e., not since Saturday).

The lability of adolescents' moods and their truncated time perspective often make it useful to extend the evaluation over time to distinguish the transient issues from those that are more enduring.

Technical Issues in the Adolescent Interview Interviewing adolescents requires tact, flexibility, a sense of humor, and clinicians who have come to terms with their own adolescent issues. Condescension, sarcasm, competitiveness, or excessive passivity on the part of the interviewer are likely to result in failure. On the other hand, although conveying a genuine interest in the youngster is essential, seductiveness, overfamiliarity, or the failure to maintain an adult role are likely to frighten the adolescent. The interviewer should resist the temptation to be too “with-it” or too “cool”— nothing is more dated than last year's slang. Because adolescents can so easily feel defensive, criticized, or embarrassed, the interviewer must learn to talk frankly, yet tactfully, about areas of apparent difficulty. Many embattled adolescents see adults as either for them or against them; hence, the clinician needs to convey a genuine empathic interest in the adolescent's view of a situation without implying collusive, uncritical acceptance of it or entering into an adversarial dispute.

The general observation that it is important to learn about youngsters' strengths and interests as well as their problems and vulnerabilities is especially true of adolescents. Without appearing squeamish or avoidant of the problem areas, the clinician should inquire about the adolescent's interests, hobbies, recreations, and friendships. Knowing about the areas of accomplishment, talent, and gratification is required for a full picture that includes the youngster's adaptive and compensatory strengths and assets. Furthermore, such an inquiry helps to facilitate rapport and conveys an interest in the adolescent as a whole person, rather than simply a bearer of problems. A too-exclusive focus on pathology puts many youngsters on the defensive, and an opportunity to talk about more-neutral areas or those of greater mastery or confidence helps the adolescent feel less on the spot. Instead of the interviewer straining to demonstrate familiarity with the latest band or winning team, letting adolescents teach one about the details of their interests lets the youngsters enjoy a sense of mastery, control, and a degree of parity with the adult interviewer that helps put them at ease. Indeed, what has been called the constructive use of ignorance is often a useful technical element of the adolescent interview. This consists of being able to ask about some missing piece in the story or about some puzzling impasse with an ingenuous curiosity that invites the adolescent's self-reflection.

Although interviewing and taking the history of the adolescent cover the same general areas outlined above, certain domains are of particular salience with the adolescent. These relate to the core developmental tasks of adolescence: developing increased autonomy and individuation from parents, taking over the tasks of physical self-care; developing a satisfying, realistic body image in the face of pubertal changes; mastering burgeoning sexual and aggressive impulses; consolidating an identity including ethical values and vocational goals, developing a network of satisfying peer relations; and, in later adolescence, a capacity for romantic intimacy with an appropriate partner (Table 33-3 and Table 33-4).

1. Gradual development as an independent individual
2. Mental evolvement of a satisfying, realistic body image
3. Appropriate control and expression of sexual drives
4. Expansion of relationships outside the home
5. Implementation of a realistic plan to achieve social and economic stability
6. Transition from concrete to abstract conceptualization
7. Integration of a value system applicable to life events

Reprinted with permission from Cross RT, Duke PM: Adolescence. In *Developmental Behavioral Pediatrics*, MD Levine, WB Carey, AC Crocker, RT Cross, editors. Saunders, Philadelphia, 1983.

Table 33-3 Developmental Growth Tasks of Adolescence

Tasks	Early: 10 to 13 Years	Middle: 14 to 16 Years	Late: 17 Years and Older
1. Independence	Emotionally breaks from parents and prefers friends to family	Ambivalence about separation	Integration of independence issues
2. Body image	Adjustment to pubescent changes	"Trying on" different images to find real self	Integration of a satisfying body image with personality
3. Sexual drives	Sexual curiosity; occasional masturbation	Sexual experimentation; opposite sex viewed as sex object	Beginning of intimacy and caring
4. Relationships	Unisexual peer group; adult crushes	Begin heterosexual peer group; multiple adult role models	Individual relationships more important than peer group
5. Career plans	Vague and even unrealistic plans		Specific goals and specific steps to implement them
6. Conceptualization	Concrete thinking	Fascinated by new capacity for thinking	Ability to abstract
7. Value system	Drop in superego; testing of moral system of parents	Self-centered	Idealism; rigid concepts of right and wrong; other oriented; asceticism

Reprinted with permission from Cross RT, Duke PM: Adolescence. In *Developmental Behavioral Pediatrics*, MD Levine, WB Carey, AC Crocker, RT Cross, editors. Saunders, Philadelphia, 1983.

Table 33-4 Growth Tasks by Developmental Phase

In the realm of values, ideals, and aspirations, the clinician should explore the adolescent's values and the models whom the adolescent wishes to emulate or reject. To what extent do these values complement or conflict with those of the parents? What expectations and aspirations (realistic or not) does the adolescent have concerning the future?

The world of peers and friends should also be explored with adolescents. With whom do they “hang out?” What crowd do they see themselves part of? How do they get along with peers, what do they do for fun? Asking teenagers to describe a close friend provides a chance to learn how they think about people and relationships and to assess their capacity for empathy.

The topic of friends leads naturally to the topic of romantic and sexual interests and relationships. Exploration of this area requires tact, and the clinician cannot always be sure of receiving a candid response. Nonetheless, as with questions concerning depression or substance use, the clinician's tone can convey the sense that the inquiry into these sensitive areas is motivated by a genuine and legitimate wish to understand the adolescent. One can begin by asking whether the teenager dates and whether there is anyone of either sex with whom the youngster is close? If so, who are they, and what attracted the adolescent to them? Are the adolescent's interests reciprocated? How has the relationship gone? Does the teenager note any recurrent patterns in terms of the type of person to whom he or she is attracted or how such relationships have unfolded in the past? Have any of the teenager's relationships developed into sexual ones? Has the adolescent had other sexual experiences? (Here one must be open to hearing about possible episodes of sexual abuse or concerns about sexual orientation.)

Because of the high prevalence of depression, suicidal ideation, and even suicide attempts in the general population, specific inquiry about these areas as outlined above is particularly important. The clinician should also inquire matter-of-factly about whether the adolescent smokes, drinks, or uses drugs illicitly, and, if so, with what frequency, in what amounts, in what contexts, and with what consequences.

ASSESSMENT OF INFANTS AND TODDLERS

The assessment of infants and toddlers poses special challenges for the clinician. First, the evaluation of young children requires specific and detailed familiarity with the cognitive, linguistic, motor, and adaptive developmental landmarks of 1 month to 6 years of age, as well as the specialized techniques needed to assess these competencies. Second, although direct observation of the child is an important source of data, the preverbal or limited verbal capacities of the child severely restrict the historical material that can be obtained from the direct interview of the child. Third (and perhaps most important), infants and young children cannot be assessed, diagnosed, or treated apart from their caregiving environment, which includes not only the parents, but also other members of the extended family and day-care and other care providers.

The developmental tasks of infancy include achieving physiological homeostasis, developing stability of state and learning to negotiate transitions between states, organizing basic physiological rhythms of wakefulness and feeding; coping with novel stimuli; perceiving coherent patterns of experience; mastering simple sensorimotor schema; establishing a specific and secure relationship with a caretaker; maintaining a positive affective tone; developing expressive and receptive channels of communication (including vocalizations and facial and gestural cues); and forming schemas of causality, intentionality, object permanence, and affect. The older infant or toddler must grapple with the regulation and socialization of bodily functions (e.g., eating, toileting, sleep), learning to concentrate and regulate shared attention, more-complex forms of imitation and acquiring the capacity for imaginative play, the challenge of socialization (including development of more-complex forms of social perception, communication, and empathy and negotiating conflicts of will with caregivers and peers), control of negative mood and aggression, and formation of core gender identity. Development can proceed at a different pace in each area and may go awry in any or all of them.

The goal of the developmental assessment of infants and toddlers is to depict functional capacities in each of these areas, the relationship between the various domains of development, their ability to adapt, and the range of coping strategies they use in doing so, as well as the caregiving environment as it impinges on them, the quality of caregiving interactions, and the impact of their developmental status on the caregiver. As with older children, the purpose of the assessment is to define what, if any, interventions are necessary, while at the same time helping to form an alliance with the parents that will facilitate implementation of any recommendations that may come out of the assessment.

The most common referral problems for infants and toddlers involve:

1. Developmental delays (e.g., specific delays such as delayed, absent, or deviant motor or speech development or general delays in multiple areas of development).
2. Psychophysiological dysregulation (fussy behavior, disturbances of sleep or feeding, rumination, eczema, self-stimulatory behaviors [head-banging, rocking], failure to thrive).
3. Behavioral disturbances (excessive tantrums, excessive negativism, aggression, hyperactivity).
4. Disturbed social development (lack of apparent awareness of or interest in others, response to abuse, neglect or multiple placements or repeated or prolonged separations).
5. Medical or genetic problems (e.g., prematurity, genetic syndromes).
6. Exposure to domestic or urban violence, sequelae of trauma.

Because so many aspects of early development are interdependent, the significance of difficulties in one functional domain cannot be understood without reference to the child's developmental status in other areas. For example, speech, language, and hearing difficulties profoundly affect personal and social development. Assessment of motor or adaptive competency that involves interaction with the caregiver or assessor will be affected by disturbances in relatedness and hence yield findings that appear erratic or uneven. Infants and toddlers are particularly sensitive to environmental stresses, which may delay the appearance of phase-appropriate skills or lead to transient regressions.

Technical Issues in the Development of the Infant and Young Child The assessment of the young child usually requires several sessions. In addition to sessions spent with the parents alone to gather history and to present the results of the assessment, familiar adults must be present during the assessment of the child; this is necessary both to place the child at ease in the assessor's presence and to permit direct observation of child and caregiver interaction. Furthermore, the behavior of infants and young children is highly state dependent (and sensitive to fatigue or illness), and one usually needs to see the child on more than one occasion to have a sufficient sample of characteristic behavior.

The general principles of history taking from the parents and assessment of the caregiving environment follow those outlined above. However, particular emphasis is given to the parental perception of the pregnancy—whether the conception was planned or came at a good time in the life of the family, and what the impact of the child's arrival was on the parents individually and on the family as a whole. The history should explore parental (and grandparental) expectations for the child, the meaning the child has for them, whom the infant reminds them of, and what they perceive as the child's best and worst traits.

The developmental history for the young child includes prenatal, perinatal, and postnatal complications, including maternal illness, medication and substance use, and neonatal status. In addition to the areas outlined above, assessment of the child's behavioral organization should include the regulation of physiological state, alertness, and activity patterns. Children's ability to console themselves and to adapt to new situations and their response to separations are also key areas of evaluation. The history should also include the history of child-care arrangements (nannies, relatives, out of home day-care) and the child's reaction to any changes in these arrangements.

The child's mastery of various gross motor, fine motor, language, adaptive, and personal-social milestones should be systematically assessed, both by history and in the direct evaluation of the child. In taking such a history, the assessor may be aided by a variety of screening instruments, such as Vineland Adaptive Behavior Scales. It is useful to inquire about the timing of the child's mastery of these milestones relative to any siblings.

Within the first few months of life, children begin to manifest distinctive and often enduring behavioral styles or temperament. Table 33-5 shows the nine temperamental categories and three distinctive constellations of temperament as defined by Stella Chess and Alexander Thomas. The goodness of fit, or degree of consonance between the child's temperament and the parents' expectations or demands of the child, is an important predictor of future development.

Table 33-5 Temperament Categories

Particularly salient aspects of the parental interview include a detailed history of the parents' own childhoods and history of care, out-of-home placement, separations, and experience in their families of origin. The impact of the child on each parent and their marital relationship should be assessed as well as the appropriateness and degree of congruity in their respective expectations of the child.

Adequate caregiving for the young child requires a delicate balance between gratification, comfort, support, and frustration appropriate to the child's developmental phase. It is important to hear (and observe) which aspects of parenting and which features of the child's behavior the parent feels most comfortable with or most

challenged by. The clinician should assess the parents' overall level of psychological functioning; the parental difficulties that adversely affect a young child's development do not always involve such gross failures as serious neglect or abuse, but may involve problems more difficult to identify such as persistent maternal depression. To help obtain a concrete picture of the parent, child, and their interaction, it is useful to ask for a detailed description of a day in the life of the child. In addition to providing data regarding the child's capacities, responses, and temporal organization, such a description provides clues to the parents' attunement and affective responses to the various elements of the child's behavior.

Although much of the clinical observation is focused on the child's behavior and capacities, direct observation of the parent-child relationship is essential. Inviting the family to interact in the play room as they usually would at home allows observation of their interaction in unstructured parent-child or family play; assigning various structured tasks may also be useful. Observing the child in interaction with each parent separately helps reveal relationship-specific patterns of behavior. Among the dimensions of parental behavior to be assessed are the ability to engage the child (both verbally and nonverbally); the quality and affective tone of this engagement; the parent's attunement to the child's cues; parental vigilance, protectiveness, and limit setting; the parent's facilitation or restriction of autonomous play by the child; thematic content; and the degree of mutual pleasure during play. Although instruments for formally assessing parental responsivity and parent–young child interaction, such as the Home Observation for Measurement of Environment (HOME) and the Parent-Child Early Relational Assessment (PCERA) are used primarily for research purposes, familiarity with the various facets of such instruments is clinically useful.

Observation and Examination of the Infant and Toddler The direct examination of the young child draws on three levels of observation: the child's response to structured assessment items during formal testing, the child's response to other interactions with the assessor and the assessment environment, and the interactions of the child and parent.

Structured Assessment Instruments A variety of instruments exists for the structured assessment of infants and young children, and each has somewhat different goals, theoretical orientation, and psychometric properties. These instruments do not yield diagnoses; they detail the child's development in various areas relative to a normative population. For example, the Denver Developmental Screening Test (Denver II) is suitable for screening use by pediatricians or trained paraprofessionals to help identify children with significant motor, social, or language delays who need fuller evaluation. Population-specific norms are also available for assessing children from families of various ethnic or educational backgrounds. The Bayley Scales of Infant Development II, which are administered by a trained assessor, can be used to assess children 1 to 42 months of age; they include a mental scale (assessing information processing, habituation, memory, language, social skills, and cognitive strategies); a motor scale assessing gross and fine motor skills; and a Behavior Rating Scale for assessing qualitative aspects of the child's behavior during the assessment. This well-standardized instrument yields standard scores for a Mental Development Index and Psychomotor Development Index.

Although tests such as the Bayley Scales may show good reliability and concurrent validity, their ability to predict later performance on Intelligence Quotient (I.Q.) assessments or later adaptive functioning is highly variable. Among the reasons for this weakness of prediction are the intervening effects of social and family environment and the heavy emphasis infant tests place on perceptual and motor skills that may have relatively little to do with information-processing capacities.

The mental status examination of the infant and young child may be organized using a schema such that shown in Table 33-6.

Table 33-6 Infant and Toddler Mental Status Exam by Anne L. Benham, M.D.

Diagnostic Formulation of the Assessment of Infants and Young Children Although mental retardation, autism, and other pervasive developmental disorders can be at least tentatively diagnosed in infants and toddlers, the most widely used official diagnostic systems—the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10)—generally lack a developmental perspective and diagnostic criteria or categories suitable for infants and young children. The Zero to Three/National Center for Clinical Infant Programs's *Diagnostic Classification of Mental Health and Development Disorder of Infancy and Early Childhood* provides a diagnostic schema for this age group that is intended to complement DSM-IV. This new schema is still under development, and the nosological approach to this developmental epoch remains an area of ongoing research and debate.

The appropriate outcome of a thorough clinical assessment of the young child thus extends far beyond an attempt at a specific categorical diagnosis. The diagnostic formulation should provide a detailed picture of the strengths and weaknesses in the child's social, emotional, cognitive, linguistic, motor, and perceptual development, the quality of the child's temperament, attachment, and regulatory capacities; hypotheses regarding the etiology of the child's difficulties; and potential exacerbating or risk factors and protective or ameliorative factors in the child and family. This formulation, which should be shared with the family in a comprehensible and usable form, ought to provide the basis for recommendations regarding intervention or further assessment.

SPECIAL TYPES OF ASSESSMENT

This chapter describes the comprehensive clinical assessment of the child. Although the clinician must be aware of the many facets of history taking, interview, and observation described herein, certain clinical settings or tasks may call for a narrower focus of inquiry or distinctive emphasis for the evaluation. For example, an urgent situation may require therapeutic intervention before a comprehensive assessment can be completed; in such cases the immediate evaluation focuses on defining the crisis and selecting the most suitable immediate intervention. Thus, the emergency room or crisis center evaluation focuses on precipitants and extent of the current decompensation, predisposing vulnerabilities, the exacerbating factors impinging on the child and family and the ameliorative resources available to them, the risk of danger to self or others, and other issues germane to making an immediate disposition (including assessing the potential need for hospitalization). Evaluation of suicidal ideation, threats, or attempts or aggressive outbursts are perhaps the most common clinical context for such urgent evaluations (Table 33-7). Children exposed to acute trauma (e.g., natural catastrophe or violence) are often screened outside the traditional clinic setting, in a combination of emergency assessment and crisis intervention. The noncompliant, acutely agitated, or delirious child in the pediatric setting raises yet a different set of assessment issues (Table 33-8).

Table 33-7 Questions to Ask in the Evaluation of Suicidal Risk in Children

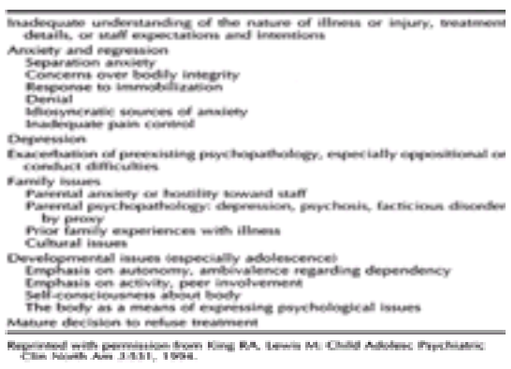

Inadequate understanding of the nature of illness or injury, treatment details, or staff expectations and intentions
Anxiety and regression
Separation anxiety
Concerns over bodily integrity
Response to immobilization
Denial
Idiosyncratic sources of anxiety
Inadequate pain control
Depression
Exacerbation of preexisting psychopathology, especially oppositional or conduct difficulties
Family issues
Parental anxiety or hostility toward staff
Parental psychopathology: depression, psychosis, factitious disorder by proxy
Prior family experiences with illness
Cultural issues
Developmental issues (especially adolescence)
Emphasis on autonomy, ambivalence regarding dependency
Emphasis on activity, peer involvement
Self-consciousness about body
The body as a means of expressing psychological issues
Mature decision to refuse treatment

Reprinted with permission from King RA, Lewis M: Child Adolesc Psychiatric Clin North Am 3:551, 1994.

Table 33-8 Differential Psychosocial Diagnosis of the Uncooperative Patient

Various specialized consultative purposes also require distinctive techniques or procedures or specific areas of inquiry. For example, forensic evaluations raise special issues of role definition, confidentiality, and privilege and require familiarity with the legal issues involved. Clinicians must be clear about whether they are serving as neutral impartial advocates for the best interests of the child at the behest of the court or as consultants to one party or another. Although the clinician treating or assessing the child for clinical purposes may have information valuable to a forensic evaluation, it is usually in the best interests of both the therapy and the legal proceedings for the evaluator to be someone other than the therapist.

The forensic evaluation of children or adolescents who may have been physically or sexually abused is a common form of forensic evaluation. In addition to assessing the child's competency and credibility, the examiner may also be asked for recommendations regarding whether the child should testify in court (and if so, how) and whether removal of the child or alleged perpetrator from the home is needed, as well as any therapeutic interventions that may be required. A forensic evaluation requires careful documentation of parent and child interviews. The suggestibility of younger children raises the danger of both false accusations and false denials and requires meticulous safeguards against repetitive, suggestive questioning.

Child custody evaluations also require careful definition of the assessor's role with all concerned; specific attention must be given to the best interests of the child in terms of continuity and stability of care, the quality of parent-child attachment and attunement, the child's stated preferences, the mental health of each parent, and the level of parental conflict and its impact on the child. Consultations for the juvenile court regarding youthful offenders or children in need of services and probate or commitment hearings represent other forms of forensic evaluation. Other specialized forms of evaluations include medication consultations and consultation to day-care centers, schools, or pediatric specialty clinics (e.g., pain or oncology clinics).

STANDARDIZED ASSESSMENT INSTRUMENTS FOR CHILDREN AND ADOLESCENTS

The past two decades have seen the development of a number of standardized interviews and rating scales that provide systematic methods for eliciting information about child psychiatric symptoms and for combining that information to yield categorical diagnoses or dimensional scores on statistically derived scales. The wave of interest in developing and testing methodologies was stimulated in large part by the rapid expansion of research activity in child and adolescent psychiatric disorders. This burgeoning research interest has confronted researchers with the need for assessment techniques that yield reliable and valid data.

There are a number of reasons why the development of measures to assess psychopathology in children is more challenging than the comparable task with adults. These challenges include (1) the need to incorporate an understanding of various aspects of development into the framing and wording of questions about symptoms and into the process of discriminating normative behavior from pathological behavior; (2) the necessity of conducting the assessment with both child and parent (and at times with the teacher), since often some kinds of information are uniquely reported by one or the other informant; (3) the importance and difficulty of integrating concepts of social context and adaptive functioning into a symptom-based assessment; (4) the lack of a readily available gold standard with which to compare the results from assessment methods in order to modify and improve them. Despite—or perhaps because of—these challenges, extensive research effort has been directed toward diagnostic assessment in child psychiatry; this effort has yielded a number of measures for use in clinical and epidemiological research. Some of the methods described below have found appropriate adjunctive roles in certain clinical settings.

Structured and Semistructured Interviews Available diagnostic interviews differ from each other in several respects. A fundamental one is the degree of structure that is built into the interview. Interviews range from highly structured instruments that specify the exact order and wording of all components to semistructured interviews that delineate the symptoms to be covered and suggest phrases that may be used but permit much latitude in the order and phrasing of questions. The degree of structure reflects the intended use of the interview, the type of interviewer who will administer it, and the cost of using it in a study.

On a continuum of structure, semistructured interviews resemble clinical interviews more closely than do structured interviews, but they nevertheless differ substantially in style and content from a true clinical assessment. Semistructured interviews are designed to be administered by interviewers with clinical training or with very intensive training in the assessment of symptoms covered by the measure (e.g., the Child and Adolescent Psychiatric Assessment). The flexible structure of the interview allows the clinically informed interviewer some freedom in the manner of inquiry and permits judgments about whether a reported behavior or expression of distress is of the quality and severity required to be considered a symptom. Some structure is needed to ensure that symptom inquiry is complete, and the interviewer is also required to make specified ratings that can be compared and aggregated across the sample. Interviews of this type include the many versions of the Child Schedule for Affective Disorders and Schizophrenia (K-SADS), the Child Assessment Schedule, (CAS), the Interview Schedule for Children (ISC), the Child and Adolescent Psychiatric Assessment (CAPA), and the recent forms of the Diagnostic Interview for Children and Adolescents (DICA). Semistructured interviews are more costly to administer than highly structured ones, but they are particularly appropriate for clinically based research in which subtle diagnostic distinctions may be critical for defining samples and where sample sizes are small enough to justify the expense of hiring interviewers trained to make these judgments.

Structured interviews, on the other hand, provide highly specified protocols that are particularly useful for epidemiological studies in which large sample sizes (i.e., in the thousands) require using many nonclinician interviewers. These studies investigate issues of prevalence of disorders, developmental patterns of psychopathology, psychosocial correlates, and risk and protective factors, topics that often require large samples to provide enough statistical power to examine study hypotheses. Because it is obviously unwise to have untrained interviewers make clinical distinctions and because using more highly trained interviewers becomes prohibitively expensive, considerable effort has been devoted to the development of highly structured techniques. In recent years, the bulk of this work has focused on writing and testing the Diagnostic Interview Schedule for Children (DISC), which has proceeded through several rounds of methodological testing in clinical and community samples.

Choosing the appropriate instrument for a particular purpose and interpreting results from child psychiatric studies that have used one or another diagnostic method require some understanding of a few key issues concerning the quality of the assessment method. Two critical properties in the evaluation of such instruments are validity and reliability. *Validity* is the extent to which an instrument measures what it purports to measure. There are various ways to conceptualize and assess validity. Child psychiatric diagnoses generated by a particular instrument are usually compared with those generated by an expert, or criterion, assessment. Statistical methods are then applied to quantify the degree of agreement on whether diagnoses are present or absent. In the methodological evaluation of the validity of an instrument, agreement is usually determined diagnosis by diagnosis, so that the instrument can be revised in areas of poor agreement with the criterion.

The *reliability* of an assessment method is often conceptualized as the extent to which the results of an assessment are repeatable. There are various types of reliability. *Interrater reliability* involves the stability of an assessment when it is administered by different raters (in this case, interviewers). *Test-retest reliability* reflects the stability of result over time (i.e. when a test is administered on two occasions separated by a period of time in which the degree of real change in the condition being measured is expected to be trivial). In child psychiatric assessment, the test is often administered twice, about 2 or 3 weeks apart. This interval is chosen to

minimize memory effects (i.e., the child or parent is not likely to remember, and thus repeat, precise responses to a great number of the questions previously asked). On the other hand, the relatively long duration of many child psychiatric conditions makes it more likely that changes between the first and second administration reflect unstable aspects of the assessment instrument than real changes in the disorder.

In addition to writing and testing the interview itself, developers of structured and semistructured diagnostic methods must determine how to use symptom information to yield diagnoses. While DSM-IV criteria determine the overall rules for defining a diagnosis, much detailed work goes into the decisions about the number of positive responses required to meet a DSM-IV symptom criterion. The specification of frequency and duration for behaviors also requires much fine tuning (e.g., contrast the number of times truant versus episodes of fire-setting for meeting DSM-IV symptom criteria). For some instruments (e.g., particularly highly structured interviews such as DISC), diagnoses are generated by having a computer count and combine the interview information according to a specified algorithm. For the semistructured instruments (e.g., K-SADS), the clinician performs this process. Depending on the instrument, information elicited from the child and parent may be combined to generate diagnoses, simulating clinical assessment. Alternatively, diagnoses may be generated for each informant separately, so that the investigator can explore issues of differential reporting (and perhaps awareness of) types of psychopathology, as well as different correlates for parent- and child-reported psychopathology. Recent research has placed increased emphasis on incorporating indicators of associated impairment, over and above the mere presence of symptomatology.

Used in a clinical setting, these interviews may be useful in prompting the clinician to inquire systematically about a broad array of symptoms and disorders, including those that may be clinically significant but are not part of the initial chief complaint.

Rating Scales A variety of checklist rating scales have been developed over recent years to assess symptomatology and various aspects of behavior and adaptive functioning. Different measures have been designed to be completed by parents, teachers, or children. These may focus on either a broad range of symptoms or a specific area (e.g., depression, anxiety). The development of each scale involves testing its reliability and validity and administering it to large and diverse samples to determine norms and clinically relevant cutoffs.

Probably the most widely used comprehensive symptom rating scale is the Child Behavior Checklist developed and extensively tested by Thomas Achenbach and Craig Edelbrock. It is a 118-item inventory of behavior and functioning that is filled out by the parent or primary caregiver. The component behavior and competency scales have age and gender group norms and clinically meaningful cutoffs. Companion forms exist for the teacher to fill out (Teacher Report Form [TRF]) and for the youth (Youth Self Report [YSR]). These measures provide a convenient way to screen for overall level of symptomatology, and when scored, they provide a profile of problem areas and competencies. The Child Behavior Checklist is stable over relatively brief intervals (i.e., good test-retest reliability) but also reflects changes resulting from treatment. The Child Behavior Checklist TRF and YSR have been widely used in epidemiological and clinical research and in many clinical settings and countries. These measures can be a useful clinical screening tool and an efficient way for clinics to gather uniform information for characterizing the populations served and changes in those groups over time.

The Behavior Assessment System for Children (BASC), developed by Cecil Reynolds and Randy Kamphaus, is a relatively new multimethod assessment for evaluating aspects of behavior and personality in children ages 4 to 18 years. It has five components including parent, teacher, and child self-reports, a structured developmental history, and a system for systematic collection of observed classroom behavior. A distinctive feature of the BASC is its emphasis on positive aspects of the child's adjustment in addition to rating symptoms and problem behaviors. An additional feature of this system is its coverage of problems related to school functioning, such as difficulties in social or study skills. Norms have been established for children by age, gender, and clinical status. This system has been designed for application in clinical, educational, research, and program evaluation contexts.

The Conners' Parent Rating Scale (CPRS) is a commonly used rating scale that collects systematic information from a parent on the child's conduct, activity level, attention, impulsiveness, and somatic complaints. The scale can be scored to reflect a child's symptom level on six component subclasses, including a general index of hyperactivity. Norms have been determined for boys and girls ages 3 to 17 years. The Conners' Teacher Rating Scale (CTRS) is a companion instrument completed by the child's teacher that can be a convenient means of collecting information on a child's behavior in the classroom. It is often a useful adjunct in evaluating and (particularly) monitoring children being treated with medication for attention-deficit/hyperactivity disorder. The CPRS and CTRS are well known and widely used. They have been revised and organized into a family of six instruments developed by Kieth Conners and colleagues, known as the Conners' Rating Scales–Revised (CRS-R). This family of instruments has scales of varying lengths for parent, teacher, and adolescent informants.

The Child Behavior Checklist and BASC families of rating scales are examples of assessment tools designed to collect information across a broad range of symptoms and competencies. The CRS-R focuses especially on disruptive behavior and symptoms of attention-deficit/hyperactivity disorder. More narrowly focused symptom rating scales have been developed to permit the quantitative assessment of specific symptom realms. Such symptom scales may be useful in quantifying the presenting severity of a symptom to establish a baseline against which response to a therapeutic intervention, such as medication, can be compared. A number of these scales are self-report measures. These have the advantages of providing an opportunity for the child to report on symptoms that might be difficult to conceptualize or verbalize. Sometimes parents or children (especially adolescents) are initially more willing to report an area of difficulty in the seemingly more anonymous context of a questionnaire than they are in a face-to-face interview. On the other hand, self-reports also carry the risk of the child misunderstanding some questions or feeling compelled to provide socially desirable responses rather than acknowledging behaviors and feelings that the child believes would be met with disapproval. With these potential limitations in mind, child self-report measures can provide useful adjunctive information in clinical settings or research studies.

Among the rating scales useful in clinical practice are a variety of scales for assessing the severity of depressive symptoms. One of the most widely used, the Children's Depression Inventory (CDI), developed by Maria Kovacs, is a self-report measure for assessing symptoms of depression in children aged 7 to 17 years. It was adapted from the well-known adult scale, the Beck Depression Inventory, and has seen considerable use in clinical and research settings. The Reynolds Adolescent Depression Scale (RADS) was designed specifically to assess depressive symptoms in adolescents and has been used in a range of populations. In addition to research applications, it may be particularly appropriate to screen for depression in schools or other such settings. The Children's Depression Rating Scale (Revised) of Elva Poznanski and Hartmut Mokros, modeled after the adult Hamilton Rating Scale for Depression, assesses 17 depressive symptom areas, based on a 20- to 30-minute, focused, semistructured interview; well standardized, it is useful as a sensitive and reliable screening instrument and severity measure of depression for children 6 years old and above.

Various rating scales for different types of anxiety symptoms, traits, and disorders are also available. For example, the Revised Children's Manifest Anxiety Scale (R-CMAS), developed by Cecil Reynolds and Bert Richmond, is a child self-report measure that assesses symptoms of anxiety (including physiological anxiety), worry, and oversensitivity, and social concerns and concentration. Children at the third-grade reading level or above can fill it out individually or by group administration, and it has been used with younger children by reading the items aloud. The Multidimensional Anxiety Scale for Children (MASC) is a self-report measure for children and adolescents aged 8 through 18 years that assesses symptoms of anxiety in the areas of physical symptomatology, social anxiety, harm/avoidance, and separation anxiety, as well as seven more-refined subdomains. Although relatively new, it is being used in a number of treatment studies to monitor response.

Instruments are available for assessing many different types of child and adolescent symptomatology, including symptoms of obsessive-compulsive disorder (the child version of the Leyton Obsessional Inventory by Carol Berg and coworkers and the Child Yale-Brown Obsessive-Compulsive Scale by Lawrence Scahill and coworkers); tic severity (Yale Global Tic Severity Scale by James Leckman and coworkers); eating disorders and related attitudes (Eating Disorder Inventory by David Garner and coworkers and Eating Attitudes Test by David Garner and Paul Garfinkel). For assessment of autism and related disorders there is the Autism Diagnostic Observation Schedule (ADOS), a play-based interview with standardized probes and ratings for various facets of social relatedness, communication, and repetitive behaviors; the revised Autism Diagnostic Interview (ADI-R) is a semistructured interview administered to parents regarding developmental milestones and assessment of these same facets.

Instruments also exist for assessing level of adaptive functioning and impairment, either globally or with respect to specific domains. The Children's Global Assessment Scale (C-GAS) by David Shaffer and coworkers provides a quantitative scale for the clinician's rating of overall impairment. The Columbia Impairment Scale by Hector and coworker Bird is a rating scale of impairment that taps four major areas of functioning: interpersonal relations, broad psychopathological domains, functioning in school or at a job, and use of leisure time. The Vineland Adaptive Behavior Scales provide a comprehensive framework for systematically recording and assessing the development of children with respect to various realms of adaptive functioning. The Social Adjustment Inventory for Children and Adolescents (SAICA) provides a semistructured interview format for assessing social functioning.

Various structured and semistructured formats for performing or recording portions of the child mental status examination have been developed. Some such as the child version of the Mini-Mental State Examination by R.A. Ouvrier and coworkers contain specific standardized items for screening cognitive functions. Other instruments, such as the Mental Health Assessment Form by Clarice Kestenbaum and Hector Bird, provide a format for organizing mental status data derived from a

semistructured clinical interview.

Limitations of Structured Measures in Clinical Assessment Interviews and scales such as those described above can at times be useful adjuncts in clinical assessment, if thoughtfully chosen and administered with sensitivity to their meaning in the overall process of the evaluation. For example, they can help ensure a comprehensive review of symptoms or provide a standardized baseline measure of severity for later determination of intervention efficacy. They cannot, however, replace an individualized child psychiatric interview nor be relied upon as the sole basis for establishing diagnoses or planning treatment. Most of the existing instruments were developed for use in research studies where data must be gathered from all subjects in a specified, uniform manner. Eliciting information for a clinical child assessment, on the other hand, requires comprehensive, detailed, and flexible inquiry and an empathic rapport with the parent and child. These elements cannot be provided by a standardized interview format alone. Furthermore, most standardized interview schedules are designed as symptom inventories and not the comprehensive assessment of feelings, personality style, coping mechanisms, situational context, and adaptive strengths that the clinical interview affords. Such factors may be as crucial to the clinical assessment and treatment planning as the presence or absence of a given pathognomonic symptom or categorical diagnosis.

The researcher interested in selecting measures for a study protocol must review carefully the existing measures in the area of interest and consider the psychometric characteristics of potential measures and the appropriateness of the measures for the study sample and the goals of the study. It is useful to begin by examining some up-to-date critical reviews of instruments.

PSYCHOLOGICAL AND NEUROPSYCHOLOGICAL ASSESSMENT

Psychological assessment using tests of personality and mental functioning can be a valuable part of the child psychiatric evaluation. Psychological evaluations provide a unique perspective on the child's condition by generating information complementary to that obtained from history, observation, and interview. With history and observation, such psychological data may clarify relationships among various etiological factors or help direct the selection of an appropriate intervention. Although partly inferential, the clinical information derived from projective tests of personality is less influenced by certain reporting biases (e.g., social desirability) than self-report. Psychological testing also provides data that are more quantitative than the qualitative information provided by the child's history, presenting complaints, and clinical presentation.

The quantitative character of test findings is more important in the assessment of cognitive abilities than in personality assessment, because the child's relative strengths and weaknesses in mental processing can be estimated only grossly from history taking and clinical observation. Such clinical estimates are also, necessarily, less accurate. The ability to discriminate meaningful differences among various domains of functioning helps to clarify the particular pattern or syndrome evidenced by the child. The assessment of cognitive functioning also reveals cognitive impairments that may underlie or influence the psychological condition under investigation (e.g., whether poor social judgment in a child with conduct disorder also results from a nonverbal learning disability). Furthermore, the manner in which any cognitive impairments or strengths interact with the child's psychological disorder can be delineated. Findings from tests of personality also identify important dynamics that affect the child's clinical condition.

Referral for Psychometric Assessment A referral for a psychometric assessment of the child is a request for a professional consultation from which an opinion regarding the referral questions can be expected. The referral for psychological examination is analogous to that made to a neurologist or radiologist, who use their assessment methods and particular expertise to contribute potentially valuable information that can then be considered, evaluated, and integrated into the psychiatric evaluation. Referrals are, therefore, best worded in terms of the information that is needed, the issue to be resolved, or the decisions that must be made, rather than requesting that specific tests be administered.

Referral questions can be broad in scope, such as requesting a second opinion regarding the child's condition from the unique perspective of psychological test findings. Requests for assistance in delineating differential diagnoses or in identifying idiosyncratic attributes of the child that may affect treatment are narrower referral questions that, nonetheless, still require relatively broader psychological examinations. In contrast, more narrowly focused questions, such as the adequacy of the child's reality testing, may permit more focused examinations. However, the extent to which psychological evaluations can be restricted to a narrowly defined referral question has a lower limit. Usually a critical mass of data is required to interpret test findings properly.

Referrals for psychological examination should be made to professionals who will use the appropriate instruments and can interpret the findings. Just as psychiatrists may specialize in the treatment of adult, child, or geriatric patients or develop a special expertise with certain disorders, psychologists develop specialized skills in assessing certain groups of children or syndromes.

While it is still best to state a referral question, referrals are frequently made for a particular kind of evaluation (e.g., psychological versus neuropsychological). A request for a psychological evaluation is commonly understood by the psychologist to mean assessment of personality and intelligence, with only a screening of achievement or other cognitive functioning unless specifically requested. In contrast, a request for a neuropsychological evaluation by a neuropsychologist is assumed to mean comprehensive assessment of cognitive functioning, including intelligence, complex mental processing, attention, memory, language, perception, and motor functioning, with only a screening of personality.

The primary purpose of neuropsychological evaluations is not to detect brain damage. Rather, comprehensive assessment of cognitive functioning, attention, memory, language, perception, and motor functioning permits the identification of deficits that cause academic, social, or developmental disabilities or that impede psychiatric treatment. Such assessments also identify the child's areas of strengths and weaknesses and help provide the basis for prescribing rehabilitative interventions that may be delivered as special education, occupational therapy, physical therapy, or vocational or cognitive rehabilitation. By parsing the psychological and cognitive impairments that contribute to the child's functional disability, such assessments also help to delineate the interaction between the child's psychological disorder and cognitive impairments. While neuropsychologists are familiar with the cognitive sequelae of various neurological conditions and can be particularly helpful in the assessment of children with neurological disorders, any child requiring a detailed assessment of cognitive functioning is appropriate for referral.

Psychological Test Batteries Table 33-9 contains a list of commonly used psychological measures for various domains of functioning. A comprehensive examination samples all of the domains, whereas a more focused evaluation selects a subset of domains. In the most general sense, a battery of tests consists merely of a collection of tests to measure a variety of mental functions. There are loosely defined batteries that are routinely administered in psychological examinations. For example, the psychological battery for a child often includes the Wechsler Intelligence Scale for Children–III (WISC-III) as the measure of intelligence. In addition, there may be a measure of achievement in basic linguistic skills and math, such as the Wide Range Achievement Test–3 (WRAT-3) or the Kaufman Test of Educational Achievement (K-TEA). A test of visuomotor integration or visual construction, such as the Beery-Buktenica Test of Visual Motor Integration (VMI) or the Bender Visual-Motor Gestalt Test, is also usually administered. Finally, a variety of personality tests and behavior rating scales is used.

Table 33-9 Commonly Used Child and Adolescent Psychological Assessment Instruments

Some published test batteries consist of measures of various mental functions that have been developed together and can usually be combined into composite or factor scores. For infants and preschool children, psychometric instruments primarily focus upon development, including motor, sensory, perceptual, and language development and the development of conceptual reasoning. Comprehensive developmental quotients are derived from many of these instruments. The revision of the

Bayley Scales of Infant Development and the Mullen Scales of Development are two notable examples. The Woodcock-Johnson Psychoeducational Battery and the Halstead-Reitan Neuropsychological Batteries are appropriate for school-aged children and examine a variety of smaller domains of functioning.

Personality Measures Personality tests for the young school age child consist primarily of projective tests of personality. Projective tests contrast with objective tests of personality in that projective tests use limited structure and ambiguous stimuli to which the child can respond in a virtually unlimited variety of ways. Perhaps the best known example of a projective technique is the Rorschach Inkblot test. Other projective tests commonly used for the young, school-age child include the Thematic Apperception Test (TAT) and the Children's Apperception Test (CAT), which both require that the child compose a story in response to a picture depicting a scene. A variety of tests administered to this age group require the child to draw, such as the Draw-A-Person Test, the House-Tree-Person Test (HTP), and the Kinetic Family Drawing (KFD). (Methods exist for scoring children's drawings to assess cognitive and visual motor development as well as affective content). For the most part, the interpretation of responses to projective tests of personality is inferential, that is, the meaning is not self-evident in the response. The inferences are based upon accumulated clinical experience reported in the literature, symbolic interpretation or analogous reasoning based upon a theory (e.g., psychoanalytic theory) and, to a much lesser extent, empirically established norms. The notable exception is the interpretation scores derived from the Exner Scoring System for the Rorschach, which have norms and have been extensively studied in empirical research.

Objective personality tests are usually more structured than are projective tests, in that the stimuli have limited ambiguity and only a select number of responses are permitted. In most objective measures, individuals are asked whether various statements are true or false for them. Perhaps the best known example of an objective test of personality is the Minnesota Multiphasic Personality Inventory, which is now in its second revision (MMPI-2) and is appropriate only for adults. Objective measures of personality for younger children are usually limited to behavior rating scales or structured reports by parents or caregivers. The Personality Inventory for Children (PIC) and the Behavior Assessment Scales for Children (BASC) are good examples of these measures. There are objective tests of personality available for adolescents in the format of the MMPI-2. The Minnesota Multiphasic Personality Inventory–Adolescent (MMPI-A), and the Million Adolescent Personality Inventory (MAPI) are two well-known examples. Both of these tests yield personality profiles based upon the relative deviations of scales that measure various psychological attributes.

Validity and Reliability While the validity and reliability of individual tests as measures of specific mental functions are established with group data, the reliability and validity of findings for the individual child is also a concern for the examining psychologist. Test batteries provide some advantages in determining the reliability and validity of individual test findings. An integrated battery provides some overlap in the testing of mental functions, permitting comparative assessment of the consistency of the child's performance across measures. In addition, if there is a theoretical or empirical basis for knowing that some mental abilities are associated with each other, then consistent findings across measures of those abilities strengthens the confidence in the findings for the individual child. Examinations that focus on a few mental functions by administering a limited number of tests do not permit detection of transient performance factors that may cause spurious scores. In addition, no test is a pure measure of the target variable; therefore, performance on a single test may be impeded by impairment of a mental function necessary for performing the task but which is not the function that the test purports to measure. If previous psychometric evaluations have been performed, the reliability of findings for the particular child can be estimated from the consistency of findings across examinations.

Finally, the assessment of children with motor, language, hearing, or visual impairments presents a special challenge. Every psychological measure requires a modality of perception, mental processing, and modality of expression. To assess children with impairments in any of these functions, tests are selected with forms of input and expression that avoid the obviously impaired functions. Children with pervasive disorders of mental processing require assessment methods appropriate for their level of functioning. In such cases, it is frequently more helpful to characterize the impairments in terms of criterion-referenced measures, which list abilities mastered, or in terms of the child's standing among children of the same diagnostic group.

Interpretation of Psychometric Findings The interpretation of psychometric findings is predicated upon standardized administration of the tests. In addition, the proper interpretation of psychometric findings generally requires knowledge of parametric and nonparametric statistics, mastery of the literature on test performance and test construction, familiarity with the construction and psychometric properties of the specific tests used, and knowledge of the literature on those instruments.

The results of psychological examination are often best integrated in a concluding discussion of the findings as they pertain to the referral question. The results of a neuropsychological examination are organized around a discussion of various functions with a concluding integration. The interpretation of test results includes consideration of the immediate context in which the tests were administered (e.g., performance factors such as fatigue, anxiety, rapport, level of cooperation, and motivation of the child). There are generally three levels of analysis for the test scores themselves: (1) symptoms, (2) levels of performance, and (3) syndromes or patterns. Syndrome analysis interprets scores within the context of the results of other measures and is the most powerful form of analysis in terms of the ability to extrapolate beyond the immediate test data. This capacity to extrapolate is based upon the known characteristics of children evidencing the syndrome. Finally, in testing children, one must bring a developmental perspective to the findings, so that future difficulties can be anticipated with the changes that occur as the child grows older.

Psychometric findings are communicated more easily if the clinician receiving them understands some basic principles of psychometrics. Most people cluster around some average score for most measures of mental abilities, and fewer persons score either very high or very low. The distribution approximates a binomial distribution in the form of an inverted U, known as the normal curve. Raw scores are frequently the number of correct responses during a test and are not usually reported. Percentile scores indicate the percentage of persons obtaining raw scores equal to or less than the raw score obtained by the individual. In the normal distribution, small changes in raw scores close to the average (where many people are grouped) result in relatively large shifts in the percentile rank, whereas large shifts in raw scores at either the high or low ends of the distribution may change the person's percentile rank very little.

To compensate for the variations in the rate of change of percentile scores and to make performance levels comparable across tests, raw scores are converted to standard, or deviation, scores that indicate the extent to which an individual deviates from the mean or average. By convention, the average standard score for most psychological tests is 100, and the standard deviation is set at 15. I.Q. scores are standard scores with these psychometric properties. By the use of standard scores, performance levels on measures of specific mental abilities can be compared with I.Q. Standard scores also give the best indication of significant differences among scores, because the ntervals between scores of the same magnitude are equal anywhere in the distribution. Statistical significance occurs when the probability is very low that a difference between scores could have occurred by chance, as the result of inherent error variance in the test. By convention, a level of .05 probability is set as the standard for statistical significance. Although two scores may differ, that does not mean that the difference is meaningful or has clinical significance. Many statistically significant differences occur frequently in the normal population. For example, a difference of 15 points between the Verbal and Performance I.Q. scores of the WISC-III is statistically significant, but a difference of 15 points is not necessarily clinically significant, because it occurs in 24.3 percent of the normal subjects used to standardize test. Of course, a difference of this magnitude could have clinical significance in a particular case, if the adverse effect of the difference was exacerbated by other factors.

Some tests for children allow calculating standard scores from either age norms or grade norms, for example, the Woodcock-Johnson Psychoeducational Battery and the Kaufman Tests of Educational Achievement. The difference between these two methods of calculation illustrates the importance of understanding test scores in relation to the normative reference group. A standard score based upon age norms compares the performance of the child with that of other children of the same age. A standard score based upon grade norms compares the child's performance with that of other children in the same grade. If the child has been retained in grade to compensate for learning problems, the standard score based upon grade norms will be artificially inflated, masking the extent of the child's impairment. On the other hand, an achievement score based upon grade norms will indicate how well the child has mastered the curriculum at that grade level. In general, grade scores and age scores should be interpreted cautiously, because their psychometric properties are poor and the implications of the scores vary across levels. Clinical decisions should not be based solely on these types of scores.

There has traditionally been a concern about reporting I.Q. scores, because people do not understand that the scores are not absolute and immutable. Reporting the standard error of measurement or confidence intervals helps others to realize that scores have inherent error and will vary to some extent over successive administrations of the test, even when the test is valid and reliable by present standards. The standard error of measurement is the standard deviation of the hypothetical distribution of scores obtained from repeated administrations of the test. For example, the Woodcock/Johnson Psychoeducational Battery provides standard error of measurement for its subtest scores. Confidence intervals are the range of scores believed to contain the true score at a specified level of probability. The WISC-III and K-TEA test manuals, for example, provide tables that specify confidence intervals for standard scores, at both 90 and 95 percent probability. Increasing levels of confidence require larger intervals.

Report of the Psychometric Findings Ordinarily, the child's parents are informed of the results of a psychological or neuropsychological evaluation. Because test scores are easily misinterpreted, reports of psychological test findings should not to be given to the child or parents unless a qualified professional is present to explain the findings and answer questions. It is helpful if the referring clinician indicates how such feedback is to be given. In many situations, especially if the findings could create distress for the child or family, the referring clinician may wish to provide the feedback. If a written report is desired, then a report can be written for the

family that does not include confusing technical information and is designed to minimize adverse effects of sensitive findings. In other cases, the referring clinician may choose to have the psychologist provide feedback. Alternatively, the referring clinician may decide to meet with the psychologist and the family so that the psychologist can explain any technical aspects of the evaluation and its findings, while the referring clinician can place the findings within the context and perspective of the overall assessment and treatment recommendations.

LABORATORY MEASURES

A burgeoning number of research studies applying new techniques of molecular genetics, neurobiology, functional and structural imaging, and neuroendocrinology are rapidly expanding our understanding of the pathogenesis of various childhood psychiatric disorders. This section, however, focuses on the practical use of various laboratory studies in the clinical assessment of the child presenting for psychiatric evaluation.

PEDIATRIC AND OTHER ADJUNCTIVE EVALUATIONS

Pediatric Evaluation Close collaboration with the child's pediatrician is usually an important component of any evaluation. The pediatrician is a valuable source of information about the child's developmental and medical history and can contribute a unique long-term perspective on the child and family.

A general pediatric examination is a useful adjunct to the psychiatric evaluation of a child. In addition to a general review of systems, the pediatrician should assess the child's current and past physical growth (head circumference, height, weight, pubertal status). Screening of hearing and vision, with more detailed testing as indicated, is important because deficits in these areas can manifest themselves in learning, language, and social delays or seeming inattention or oppositionality. For children with developmental difficulties, a careful history of perinatal risk factors is taken; the pediatric examination includes careful evaluation for major or minor congenital anomalies, dysmorphic features, dermatological abnormalities (hyperpigmentations, tubers, fibromas), or ophthamological findings (lens abnormalities, retinal pigmentation) (Table 33-10). History taking includes abnormal movements or absence spells suggesting seizures, abnormal postures, and tics. The physical examination notes difficulties in muscle tone and strength; asymmetrical, abnormal, or primitive reflexes; tremor; and gait or postural disturbances

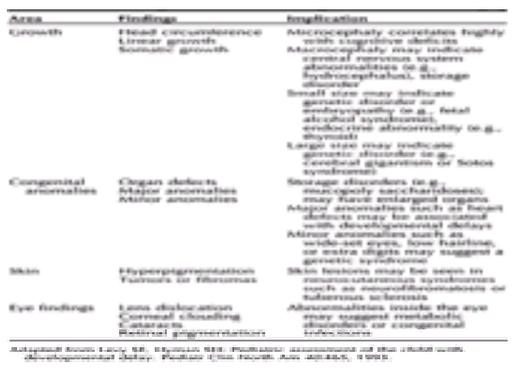

Area	Findings	Implication
Growth	Head circumference Linear growth Somatic growth	Microcephaly correlates highly with cognitive deficits Microcephaly may indicate central nervous system abnormalities (e.g., hydrocephalus), storage disorder Small size may indicate genetic disorder or embryopathy (e.g., fetal alcohol syndrome), endocrine abnormality (e.g., thyroid) Large size may indicate genetic disorder (e.g., cerebral gigantism or Sotos syndrome)
Congenital anomalies	Organ defects Major anomalies Minor anomalies	Storage disorders (e.g., mucopolysaccharidoses) may have enlarged organs Major anomalies such as heart defects may be associated with developmental delays Minor anomalies such as wide-set eyes, low hairline, or extra digits may suggest a genetic syndrome
Skin	Hyperpigmentation Tumors or fibromas	Skin lesions may be seen in neurocutaneous syndromes such as neurofibromatosis or tuberous sclerosis
Eye findings	Lens dislocation Corneal clouding Cataracts Retinal pigmentation	Abnormalities inside the eye may suggest metabolic disorders or congenital infections

Adapted from Levy SE, Hyman SL. Pediatric assessment of the child with developmental delay. Pediatr Clin North Am 2003; 50:843.

Table 33-10 Clues to Developmental Disability From the Physical Examination

The pediatric history and examination, in turn, help guide decisions about what further medical consultations (e.g., neurological, orthopaedic, audiometric, genetic) and which diagnostic tests (e.g., neuroimaging, laboratory tests, electroencephalogram) are likely to be informative.

Diagnostic Laboratory Tests The clinical utility and cost effectiveness of routine laboratory tests for children who present with psychiatric problems has not been thoroughly studied. Most guidelines for performing these tests for children have historically been developed using data from studies of adults. Adult studies generally suggest that routine laboratory tests are not clinically useful in such typical psychiatric settings as outpatient clinics and most inpatient units. Laboratory screening tests are of some use in psychiatric settings where patients are at high risk for medical illness, such as the emergency room, substance abuse treatment centers, acquired immune deficiency syndrome (AIDS) clinics, and geriatric clinics, as well as for patients who have new-onset psychosis, depression, or dementia. Similarly, routine screening laboratory measures are more likely to yield clinically significant information when clinical symptoms of physical illness are present.

The few studies of the use of routine laboratory tests in child psychiatry patient populations have yielded similar conclusions. One review of routine laboratory screenings (thyroid function tests, electroencephalograms [EEG], chest X-ray, chemistry panel, urinalysis, complete blood count, electrocardiogram, and rapid plasma reagin) in 100 consecutive adolescent inpatient admissions reported variable rates of abnormal values, depending on the specific test, but in only 1 of these 100 patients did the tests produce a change in diagnosis from functional to organic—and even then the diagnostic information proved to have little clinical relevance. Most laboratory abnormalities in this study were regarded as minor and did not indicate a need for clinical follow-up.

More-specialized diagnostic laboratory evaluations (EEG, computed tomography [CT], magnetic resonance imaging [MRI], or chromosomal analyses) also provide a relatively low yield of clinically useful information. In a study of 200 consecutive child psychiatric inpatients, these evaluations were done only when “clinically indicated”; despite their judicious use, the tests provided clinically relevant information in only 7 patients (3.5 percent of the total patient sample), or in 7 of 136 tests (5.1 percent of all tests performed). Chromosomal analyses proved to be the most informative of all these tests, yielding new medical diagnoses in 5 of 32 (15.6 percent selected children for whom they were performed. A study of 111 putatively high-risk inpatients with new-onset adolescent psychoses produced similar conclusions. In this population, routine endocrine and neuroimaging screening evaluations failed to provide any information of diagnostic utility in any patient (although inconsequential laboratory abnormalities were present in 15.4 percent of the neuroendocrine screens and 11 percent of the neuroimaging screening tests). More-specialized neuroimaging technologies, such as positron emission tomography (PET), single photon emission computerized tomography (SPECT), functional MRI (fMRI), and brain electrical activity mapping (BEAM) currently have no routine clinical or diagnostic utility in child and adolescent psychiatric populations.

Testing in Specific Childhood Disorders Certain child psychiatric patient populations may warrant specific diagnostic tests. Generally the pediatric history and physical examination guide the appropriate use of laboratory tests.

Mental Retardation and Pervasive Developmental Disorders In general, the more severe the child's disability, the greater the likelihood of an identifiable (albeit not necessarily treatable) cause. For example, one study found that a medical cause could be identified in 60 to 70 percent of children with severe mental retardation, versus only 35 to 55 percent of children with mild mental retardation. Chromosomal abnormalities, especially, trisomy 21 (Down syndrome) or fragile X syndrome, account for over half of cases of mental retardation.

The pediatric history and examination provide important clues. Loss of previously acquired developmental milestones or the presence of congenital anomalies, dysmorphic features, microcephaly or macrocephaly and other growth problems, abnormal skin pigmentation or lesions, and symptoms suggesting seizure activity or other neurological abnormality warrant more-detailed genetic, metabolic, or neurological evaluation (Table 33-10).

Autistic children have relatively high rates of mental retardation, seizure disorder (20 to 40 percent), and tuberous sclerosis (1 to 5 percent). Most experts would therefore still advocate a Wood's lamp examination to search for tubers, an EEG to exclude seizures, and chromosome analysis to exclude fragile X in children who display features of autism, retardation, or pervasive developmental disorder. Lead testing is also indicated in cases of mental retardation, since chronic lead ingestion due to pica can cause substantial retardation; conversely, many children with retardation or pervasive developmental disabilities from other causes indulge in pica and may accumulate a substantial lead burden.

Mood Disorders Common medical causes of mood disorders in child and adolescent psychiatric populations include substance abuse, infectious diseases (e.g., mononucleosis, Lyme disease, human immunodeficiency virus [HIV], and thyroid abnormalities). Rapid onset, severe fatigue, and a change in cognition suggest an organic basis for the mood disturbance. Routine thyroid function, infectious disease, and toxicology testing is probably indicated in this population.

Psychotic Disorders The patient's history and physical examination should guide the use of diagnostic tests beyond a general health screening battery. Cognitive decline, an altered level of consciousness, headache, an abnormal neurological examination, altered vital signs, leukocytosis, or new-onset seizure disorder are all

indications for further clinical and laboratory workup that may include examination of the cerebrospinal fluid.

Attention-Deficit/Hyperactivity Disorder Thyroid function abnormalities (particularly hypothyroidism) in attention-deficit/hyperactivity disorder were first reported in 5.4 percent of a group of 277 children with this disorder compared with reported rates in the general population below 1 percent. Matched control children were not included in this study, however, so the results were considered preliminary. In a subsequent study of 196 inpatient adolescents with attention-deficit/hyperactivity disorder who were routinely tested for thyroid function abnormalities, most of the observed abnormalities had normalized by the time of repeat testing 1 week later, and none of the adolescents actually required treatment for the abnormalities. Therefore routine thyroid testing is not currently indicated in children newly diagnosed with this disorder who do not have other signs or symptoms of thyroid dysfunction.

Tic and Obsessive-Compulsive Disorders History and physical examination differentiate tics and compulsive behaviors from other stereotypies and movement disorders such as choreoathetosis, myoclonus, ballismus, akathisia, and tremor. Once the diagnosis of a tic disorder or obsessive-compulsive disorder is made, further laboratory testing is generally not indicated. However, with an acute onset or acute exacerbation of symptoms or with chronic or recurrent pharyngitis, a throat culture and serological studies for group A b-hemolytic streptococcal (GABHS) infection, which include anti-deoxyribonuclease B and antistreptolysin O antibody titers, are probably warranted. Recently, GABHS has been postulated to initiate and sustain an autoimmune disorder due to cross-reactivity of GABHS and neuronal antigens, similar to the presumed cause of another movement disorder, Sydenham's chorea. The evidence for the role of GABHS in the cause of tic disorders and obsessive-compulsive disorder, however, is much less clear than it is in Sydenham's chorea.

Substance Use Disorders Substance abuse and dependence can cause a wide range of neuropsychiatric symptoms. The high prevalence of substance abuse in adolescent populations has prompted recommendations that toxicology screens be obtained for (1) all adolescents who have psychiatric symptoms or who exhibit acute behavioral changes; (2) high-risk adolescents, such as delinquents and runaways; and (3) adolescents who have recurrent accidents or unexplained somatic symptoms.

Sexually Transmitted Diseases Children and adolescents with a history of either sexual activity or sexual abuse who are being evaluated for depression or a change in cognitive function should be evaluated for sexually transmitted diseases, including HIV infection and syphilis.

DIAGNOSTIC FORMULATION AND RECOMMENDATIONS

In the diagnostic formulation the clinician attempts to integrate the data gathered through the assessment process into a coherent account of the child's difficulties, the factors that appear to have predisposed the child to develop the problem, the concomitants and consequences of the problem, and the factors that maintain the problem or might ameliorate it. It is not always possible to provide a complete formulation at the conclusion of the initial assessment; under such conditions, the best that the clinician may be able to provide is a differential diagnosis that can point toward the appropriate treatment options and the subsequent steps needed to establish a diagnosis.

One product of the assessment may be the assignment of a diagnosis of one or more categorical psychiatric disorders using the multiaxial diagnostic schema of DSM-IV or ICD-10. While these nosological systems have done much to standardize diagnostic practices and improve the quality of research, one must be aware of their clinical limitations as applied to children. Many children experience sufficient symptoms, distress, or impairment to be regarded as psychiatrically disturbed and to warrant therapeutic intervention without meeting the full criteria of a categorical psychiatric disorder. Furthermore, as currently conceptualized, the DSM-IV classification is descriptive and atheoretical; it specifies neither presumed etiology nor treatment for the various diagnostic categories. If the assessment process is to provide a picture of the child that suffices to guide intervention, the resulting diagnostic formulation must go beyond simply assigning a categorical psychiatric diagnosis (if one is warranted). As noted in the introduction, comorbidity is often the rule rather than the exception in childhood, and many children have complex patterns of developmental and behavioral difficulties that are not adequately captured by a list of multiple categorical diagnoses. Children who share a given categorical diagnosis may differ in crucial respects that influence current severity, natural history, and treatment response. Examples of potential risk (or exacerbating) factors and protective (or ameliorative factors) include intelligence, compensatory skills, family resources and supports, and vulnerabilities such as comorbid pathological personality traits or neurological impairments. When the DSM-IV or ICD-10 is being used, the full multiaxial diagnostic system must be used. In addition to Axis I (categorical psychiatric disorders), this includes specifying comorbid personality disorders and mental retardation (Axis II), general medical conditions (Axis III), level of psychosocial and environmental problems (Axis IV), and a global assessment of current functioning (Axis V).

The clinician's diagnostic formulation should thus supplement the assignment of any formal categorical diagnoses by attempting to identify, as far as possible, the potential causes, predisposing factors, and current determinants of the child's difficulties. On the basis of this information and the clinician's expertise regarding the various interventions available for the diverse forms of childhood psychopathology and developmental difficulties, the clinician formulates appropriate recommendations for treatment of the child's problems.

Communicating Findings and Recommendations Communicating the findings and recommendations that result from the assessment to the parents and child is an essential part of the evaluation process and may require one or more sessions. Depending on the problem and the child's age and level of comprehension, presenting the findings usually involves meeting with the child and parents, either separately or together.

Several principles must be observed to maximize the likelihood that the parents and child will hear, understand, and experience the clinician's findings and recommendations as helpful. First, the clinician must communicate a sense of the child as a whole person, with strengths and abilities as well as problems and vulnerabilities. This perspective conveys a sense of the clinician's appreciation and empathic understanding of the child, reduces defensiveness, and helps mitigate the fear that there will be only bad news.

The findings and recommendations will be most useful to the family and likely to be implemented if the interpretive session is a dialogue rather than a lecture. The clinician must use language comprehensible to the parents and child, avoid jargon, and keep technical terms to a minimum. When technical terms are necessary, they should be explained clearly so the parents and child understand them; terms that are merely descriptive to the clinician may hold frightening or negative connotations for the child or family. Allow ample opportunity for the parents and child to discuss the recommendations; this provides a chance to judge whether the clinician's formulation made sense to the family, to address any differences of opinion, and to explore the feasibility and acceptability of the recommendations offered. Remaining areas of uncertainty or ambiguity about the assessment or recommendations should be clearly noted and discussed.

If the assessment is a consultation requested by an agency, a school, or another clinician, the findings and recommendations should be communicated to the referring party in appropriate terms, after being discussed with the parents and child and with their consent. If further evaluation or treatment is indicated but is best done by someone else, the clinician should offer to assist in appropriate referral.

SUGGESTED CROSS-REFERENCES

General principles of the psychiatric interview, history, and mental status examination are presented in Section 7.1. The psychiatric report is discussed in Section 7.2, psychiatric rating scales in Section 7.8, and medical assessment and laboratory testing in Section 7.7. Theories of personality and psychopathology are discussed in Chapter 6. The normal development of children is in Section 32.2, and of adolescents in Section 32.3. Details regarding the personality assessment of children are found in Section 7.5. Treatment in child psychiatry is discussed in Chapter 48 and special areas of interest in child psychiatry in Chapter 49.

CHAPTER REFERENCES

*American Academy of Child and Adolescent Psychiatry: Practice parameters for the psychiatric assessment of children and adolescents. J Am Acad Child Adolesc Psychiatry *34:*1386, 1995.

*American Academy of Child and Adolescent Psychiatry: Practice parameters for the psychiatric assessment of infants and toddlers. J Am Acad Child Adolesc Psychiatry *36*(Suppl):21S, 1997.

American Academy of Child and Adolescent Psychiatry: Practice parameters for the forensic evaluation of children and adolescents who may have been physically or sexually abused. J Am Acad Child Adolesc Psychiatry *36*(Suppl):37S, 1997.

American Academy of Child and Adolescent Psychiatry: Practice parameters for child custody evaluation. J Am Acad Child Adolesc Psychiatry *36*(Suppl):57S, 1997.

Angold A: Clinical interviewing with children and adolescents. In *Child and Adolescent Psychiatry: Modern Approaches*, ed 3, M Rutter, E Taylor, L Hersov, editors. Blackwell Scientific, Oxford, 1994.

Angold A, Messer SC, Stangl D, Farmer EM: Costello EJ, Burns BJ. Perceived parental burden and service use for child and adolescent psychiatric disorders. Am J Public Health. *88:*75, 1998.

Bird HR, Kestenbaum CJ: A semistructured approach to clinical assessment. In *Handbook of Clinical Assessment of Children and Adolescents*, vol 1, CJ Kestenbaum, DT Williams, editors. New York University Press, New York, 1988.

Chess S, Thomas A: Temperament. In *Child and Adolescent Psychiatry, ed 2*, M Lewis, editor. Baltimore, Williams & Wilkins, 1996.

Costello EJ, Angold A: Scales to assess child and adolescent depression: Checklists, screens, and tests. J Am Acad Child Adolesc Psychiatry *27:*726, 1988.

*Costello EJ, Angold A, Keeler GP: Adolescent outcomes of childhood disorders: The consequences of severity and impairment. J Am Acad Child Adolesc Psychiatry *38:* 121, 1999.

Cox AD: Interviews with parents. In *Child and Adolescent Psychiatry: Modern Approaches*, ed 3, M Rutter, E Taylor, L Hersov, editors. Blackwell Scientific, Oxford 1994.

Freud A: *Normality and Pathology in Children.* International Universities Press, New York, 1965.

*Gold A, Costello EJ, Farmer MZ, Burns BJ, Erlanki A: Impaired but undiagnosed. J Am Acad Child Adolesc Psychiatry *38:*129, 1999.

Granero Perez R, Expeleta Ascaso L, Domenech Massons JM, de la Osa Chaparro N: Characteristics of the subject and interview influencing the test-retest reliability of the Diagnostic Interview for Children and Adolescents—Revised. J Child Psychol Psychiatry *39:*963, 1998.

Greenspan SI, Greenspan NT: *The Clinical Interview of the Child*, ed 2. American Psychiatric Press, Washington, DC, 1991.

Herjanic B, Reich W: Development of a structured psychiatric interview for children: An agreement between child and parent on individual symptoms. J Abnorm Child Psychol *10:*307, 1982.

Hodges K: Structured interviews for assessing children. J Child Psychol Psychiatry *34:*49, 1993.

*Jensen PS, Watanabe H: Sherlock Holmes and psychopathology assessment approaches: The case of the false positive. J Am Acad Child Adolesc Psychiatry *38:*138, 1999.

Kestenbaum CJ: The clinical interview of the child. In *Textbook of Child and Adolescent Psychiatry*, ed 2, JM Weiner, editor. American Psychiatric Press, Washington, DC, 1997.

*King RA, Noshpitz JD: *Pathways of Growth: Essentials of Child Psychiatry*, vol 2, *Psychopathology.* Wiley, New York, 1991.

Levy SE: Pediatric evaluation of the child with developmental delay. Child Adolesc Psychiatr Clin North Am *5:*809, 1996.

Lewis M: Psychiatric assessment of infants, children, and adolescents. In *Child and Adolescent Psychiatry*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Naglieri JA, McNeish TJ, Bardos AN: *DAP: SPED, Draw a Person: Screening Procedure for Emotional Disturbance, Examiner's Manual.* Pro-ed, Austin, TX, 1988.

*Noshpitz JD, editor: *Handbook of Child and Adolescent Psychiatry, Volume 5: Clinical Assessment and Intervention Planning.* Wiley, New York, 1998.

Ouvrier RA, Goldsmith RF, Ouvrier S, Williams IC: The value of the Mini-Mental State Examination in Childhood: A preliminary study. J Child Neurol *8:*145, 1993.

Puura K, Almqvist F, Tamminen T, Piha J, Kumpulainen K, Rasanen E, Moilanen I, Koivisto AM. Children with symptoms of depression—What do the adults see? J Child Psychol Psychiatry *39:*577, 1998.

Rabin AI, Haworth MR, editors: *Projective Techniques with Children.* Grune & Stratton, New York, 1960.

Racusin GR, Moss NE: Psychological assessment of children and adolescents. In *Child and Adolescent Psychiatry*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Rutter M: Routes from research to clinical practice in child psychiatry: Retrospect and prospect. J Child Psychology Psychiatry *39:*805, 1998.

Simmons JE: *Psychiatric Examination of Children*, ed 4. Lea & Febiger, Philadelphia, 1987.

Solnit AJ, Cohen DJ, Neubauer PB: *The Many Meanings of Play: A Psychoanalytic Perspective.* Yale University Press, New Haven, CT, 1993.

*Sparrow SS, Carter AS, Racusin G, Morris R: Comprehensive psychological assessment throughout the lifespan: A developmental approach. In *Manual of Developmental Psychopathology*, vol 1, D Cicchetti, DJ Cohen, editors. Wiley, New York, 1995.

Sparrow SS, Balla DA, Cicchetti DV: *The Vineland Adaptive Behavior Scales: Interview Editions, Survey Form.* American Guidance Service, Circle Pines, MN, 1984.

Thomas GV, Silk AM: *An Introduction to the Psychology of Children's Drawings.* New York University Press, New York, 1990.

Woolston JL, Riddle MA: The role of advanced technology in inpatient child psychiatry: Leading edge or useful aid? J Am Acad Child and Adolesc Psychiatry *29:*905, 1990.

Zametkin AJ, Ernst M, Silver R: Laboratory and diagnostic testing in child and adolescent psychiatry: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*464, 1998.

CHAPTER 34. MENTAL RETARDATION

BRYAN H. KING, M.D., ROBERT M. HODAPP, PH.D., AND ELISABETH M. DYKENS, PH.D.

HISTORY

Persons with mental retardation have long been of interest to Western societies. This interest has ranged from the overly positive to the overly negative, from the French expression *les enfants du bon Dieu* (God's children) to Martin Luther's exclamation that “The Devil sits where their souls should be.”

The field's modern history begins in the late eighteenth and early nineteenth centuries. At that time, Jean-Marc Itard attempted a natural experiment to educate Victor, a “wild child” discovered in the forests of Aveyron, France. Although Itard himself judged his work with Victor a failure, this renowned experiment marked the first time that anyone had considered the possibility that persons with disabilities could be educated. An overview of the history of this complex field should mention three interrelated topics: service delivery, research, and organizations.

Service Delivery Spurred by Itard's work, the French educator Edouard Seguin developed his physiological view of education in the mid-1800s. Seguin's education focused on children's development in three areas: activity, intelligence, and will. *Activity* involved muscular or physical education, including exercises and movement to “awaken” the child's body. *Intelligence* emphasized educating the senses. Children with mental retardation were taught to handle objects, discriminate musical and environmental sounds, taste and perform tongue movements for speaking, and visually discriminate forms, colors, and sizes. Such visual discriminations eventually led to drawing and writing. The final component of Seguin's educational program involved educating the *will*, a “moral education” akin to what one might call socialization. Throughout, Seguin wanted children with mental retardation to take their rightful place in the societies of their day. His views that children with mental retardation could be trained successfully fit well with the progressive, reform-minded spirit of the times.

Such optimism helped create the early training schools, the residential component of service delivery in mental retardation. First begun in 1848 in Massachusetts by Samuel Gridley Howe (the first public facility) and Hervey Wilbur (the first private facility), the late 1800s saw the creation of many such training schools. These schools were originally small and homelike, often housing only 8 to 10 residents. Residential schools had as their original goal the return of children to their families after a period of intervention. By 1890, 20 such residential schools had sprung up in 15 states.

Gradually, however, these residential schools became larger and less focused on education than on custodial care. To many superintendents, it was becoming clear that many residents could not return home. In addition, the isolated placement of most training schools allowed for the segregation, overcrowding, and abuses seen in later years. Such isolated placements also fostered a change in goals: from an original focus on a warm, homelike residence that might help children return to their communities, residential schools began to be seen as custodial institutions designed to keep persons with mental retardation away from society. During their peak usage, from 1950 until the late 1960s, institutions were home to approximately 1 per 1000 Americans.

Since the mid-1960s, many fewer persons with mental retardation have been institutionalized in the United States. This move toward deinstitutionalization has arisen from many sources. First, the overcrowding and neglect common in many large institutions came to light in various exposés during the late 1960s. Burton Blatt and Fred Kaplan's photographic exposé *Christmas in Purgatory* (reproduced in *Look* magazine), Geraldo Rivera's television reports, and Robert F. Kennedy's attacks on the Rome and Willowbrook institutions in New York all shocked the nation.

Normalization Other forces also led to deinstitutionalization. Probably the most important of these forces was the philosophy of “normalization,” the idea that individuals with mental retardation were entitled to a more normal lifestyle, including a culturally normal rhythm to the day (school or work, leisure-time, sleep), week (weekdays and weekends), and year (vacations, holidays). In 1972, Wolf Wolfensberger extended the idea of normalization to the service delivery system itself, calling on all residences, schools, and other services for persons with retardation to be as normative as possible. Parent and professional advocacy groups also fought hard for legislative and legal victories to decrease the size of large institutions.

Compared with institutions even 30 years ago, changes in U.S. residential institutions have been dramatic. Several states have closed their institutions, and remaining institutions have become much smaller. Typically, patients remaining in large or even medium-sized institutions are the most severely and profoundly impaired; many of these residents also have severe behavior problems or motor or sensory disabilities in addition to their severe or profound mental retardation. Overall, from 1967 to 1994, the total institutional population (defined as those in state-operated institutions with 16 or more residents) has decreased by almost two thirds, from 194,650 to 65,818.

Schooling is the other major service for persons with mental retardation. The move to school all American children began in the mid-1800s, and teachers and administrators soon discovered that a subset of children were having difficulty performing school lessons. As a result, classes for problem children began in many cities and towns in the late 1800s. In 1894 the first classes specifically for children with mental retardation were formed in Providence, Rhode Island. In rapid succession, special education classes were begun in Springfield, Massachusetts (1894), Chicago (1898), Boston (1899), New York (1901), Philadelphia (1901), and Los Angeles (1902).

Local school districts have historically held an ambivalent attitude toward special education classes. Often housed within the least desirable buildings and rooms, these classes were frequently furnished with sparse equipment and materials, and had few specially trained teachers. By 1968, Lloyd Dunn declared that most children with mental retardation could be “mainstreamed” in classes with nonretarded age-mates; he questioned the need for segregated special education classes for most children with mental retardation. Dunn's article took note of the Supreme Court's ruling that “separate could not be equal” in educational settings, that research had not shown that children with mild mental retardation learn better in special education classes than in regular classes, and that educational techniques had advanced sufficiently to allow the effective schooling of most children with retardation alongside other, typical children.

In both education and living settings then, the post-1970 era has witnessed the strong influence of mainstreaming, community living, and normalization. On the whole, such movements have proven beneficial, as persons with mental retardation increasingly take their rightful place within modern society. Many professionals do, however, question whether normalization has sometimes gone too far. Not every child with mental retardation may be able to be schooled optimally with typically developing age-mates, nor might every adult be able to live independently in the community. These issues continue to be hotly debated among professionals, families, and organizations associated with persons with mental retardation.

Research Like service delivery, the modern history of research also begins in the mid-nineteenth century. In 1838, Jean Esquirol differentiated mental retardation from mental illness and proposed several levels of mental retardation. Later in the century, various workers proposed different classification systems. Probably the best known was a system of “ethnic classifications” proposed by J. Langdon Down in 1866; hence the term “mongolism” for persons with Down syndrome, the type of retardation he first identified.

In addition to classifying persons with mental retardation, genetics plays an important historic role in retardation research. Early in this century, that role was mainly negative. Spurred by the rediscovery of Gregor Mendel's genetic findings, various workers attempted to show that mental retardation was inherited. Most notorious were Dugdale's (1877) study of the Jukes, and Henry Goddard's (1913) study of the Kallikaks. In both cases, families with multigenerational retardation were used to argue that mental retardation was inherited and that sterilization was necessary for the “preservation of the race.” Such eugenic scares led to court decrees and

sterilization laws in California and other states during the 1920s.

Also playing an important role in mental retardation research was the new science of psychological testing. Goddard, the research director at the Vineland (New Jersey) Training School, was the first to use Binet and Simon's intelligence quotient (I.Q.) tests in the United States. After testing Vineland residents over several years, Goddard concluded that “the vast majority of feeble-minded children are not changing and are not improving in their intelligence levels,” a finding that another leading worker, Walter Fernald, called “the most significant ... and the most discouraging that we have ever known.”

Since early in this century, research has progressed in both the behavioral and biomedical fields. Behaviorally, early and more recent work has allowed better diagnoses and classification. Beginning as early as the 1910s, psychologists invented tests of motor skills, nonverbal intelligence, achievement, adaptive behavior, and other skills. Following similar work with typically developing children, psychologists have recently learned much about the development of cognitive, linguistic, social, and adaptive skills in persons with mental retardation. In addition, many studies now examine the presence of psychiatric disorders in children and adults with mental retardation, and families, schools, group homes, and workshops have all received research attention. Much of this work has occurred from the 1960s to today, a time of strong, often federally supported, progress in mental retardation behavioral research.

Arguably the most important progress has occurred in the biomedical and genetic fields. As early as 1934, the Norwegian physician Asbjorn Folling hypothesized that several of his patients with retardation were unable to metabolize phenylalanine. Such insights later led to a dietary treatment of this inborn error of metabolism; the phenylalanine-free diet, though not perfect, has significantly limited phenylketonuria (PKU) as a cause of mental retardation is most industrialized countries. In addition, PKU can now be detected at birth by the Guthrie test, after which dietary treatments can be instituted. PKU remains one of the major success stories in mental retardation.

Modern genetics constitutes another success story in mental retardation research. In 1959 J. Lejeune, M. Gautier, and R. Turpin discovered that most cases of Down syndrome were due to a trisomy of chromosome 21. Fragile X syndrome, the second most common genetic cause of mental retardation, was discovered in 1969, and its cytogenetic diagnosis was refined in the mid-1970s. With the banding techniques of the 1970s and the molecular genetic techniques of the 1980s, exact causes are being described for many of the 750 different genetic disorders associated with mental retardation. The genetic—and, increasingly, the behavioral—characteristics of many of these disorders are now known.

Organizations The modern field of mental retardation is also a history of organizations. The first such organization was the Association of Medical Officers of American Institutions for Idiotic and Feeble-Minded Persons, begun in 1876. As the first professional group devoted exclusively to persons with mental retardation, the Association of Medical Officers was the forerunner to the American Association on Mental Retardation (AAMR), the field's most influential organization. Through its various journals and committees, AAMR has been at the forefront of research, policy, and legislative advances for persons with mental retardation.

In addition to AAMR, other groups have also played important roles, particularly concerning changes in federal policy toward persons with disabilities. The Council for Exceptional Children (CEC), begun in 1922, continues to champion the profession of special education and the education of all children with disabilities. The National Association for Retarded Citizens (NARC) is the main parental lobbying group. NARC was particularly influential in the passage of the Education for All Handicapped Children Act of 1975 (PL 94&142), the federal law that for the first time mandated the “free, appropriate public education” of all children with disabilities throughout the United States. Each of these groups was also influential in the passage of the Americans with Disabilities Act, federal legislation that took effect in 1992 and outlawed discrimination of persons with disabilities.

Federal agencies have also been influential. Begun during the presidency of Kennedy (who had a sister with mental retardation), the National Institutes of Child Health and Human Development (NICHD) have long supported disability research. Other federal programs, including the University Affiliated Facilities (UAFs), University Affiliated Programs (UAPs), and Mental Retardation Research Centers (MRRCs), have also played important roles in research and policy. Taken together, professional, parental, and federal organizations provide an important infrastructure to the mental retardation field.

DEFINITION

The struggle to define and classify mental retardation is long-lived. Esquirol (1843) is considered the first medical writer to have penned a definition, and his seminal characterization of mental retardation as a disorder of development instead of a disease is maintained in all modern definitions (which require an onset during childhood or adolescence). Writing over a century ago, Wilbur (1852) reasoned that mental retardation was defined primarily by deficits in social or moral reasoning. This line of thinking has also persisted over time and is now woven into contemporary controversies about the respective roles of social adaptation and intelligence in defining mental retardation. Such controversy is not entirely recent however. William White (1919) observed that “feeble-mindedness, even from the standpoint of an intelligence measuring scale, is a relative affair when expressed in the behaviour of an individual, and conduct which would be considered normal under certain conditions might well be open to inquiry, as possibly defective, under others.” Alfred Tredgold's (1922) presentation of the case of Gottfried Mind is also illustrative:

> Gottfried Mind was a cretin imbecile who was born at Berne in 1768, and died in the same city at the age of forty-six years. At an early age he showed considerable talent for drawing, and as it was obvious that he would never be able to earn his living in any ordinary occupation, his father's employer interested himself in providing young Gottfried with some training. He could neither read nor write, he had no idea of the value of money, his hands were remarkable for their large size and roughness, and his general appearance was so obviously indicative of mental defect that his walks through the city were usually to the accompaniment of a crowd of jeering children. In spite of all this his drawings and water-color sketches of not only cats, but of deer, rabbits, bears, and groups of children were so marvelously lifelike and so skillfully executed that he acquired a European fame. One of his pictures, indeed, of a cat and kittens was purchased by King George IV.

Thus, central to the struggle with how to conceptualize and understand mental retardation is the idea that something more than cognitive deficits or low I.Q. scores on intelligence tests is involved. Over the years, the American Psychiatric Association (APA) adopted definitions of mental retardation from the AAMR. Looking back to the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) for example, the diagnosis of mental retardation is virtually identical to the previous AAMR definition. Most recently, however, the AAMR definition has diverged somewhat from that of the fourth edition of DSM (DSM-IV), thus generating some controversy. The cognitive and adaptive components of these definitions must be examined to understand the controversy.

Cognitive and Adaptive Functioning

Cognitive Functioning Both DSM-IV and the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) specify an I.Q. of 70 or less in their diagnostic criteria for mental retardation. I.Q. scores are presumably derived from standardized intelligence tests that meet appropriate psychometric criteria for reliability and validity. I.Q. tests that are commonly used to identify mental retardation are summarized in Table 34-1, along with each test's age range and cognitive domains. Though all are generally acceptable, diagnosticians should shy away from tests that tap a single domain (e.g., receptive vocabulary) in favor of more extensive batteries such as the Kaufman or Wechsler tests, as these rely on performance across multiple cognitive domains (Table 34-1).

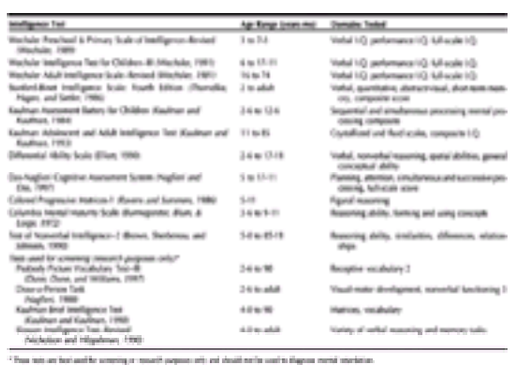

Table 34-1 I.Q. Tests for Diagnosing Mental Retardation

In many ways, administering I.Q. tests to persons with mental retardation is quite challenging, both in terms of the testing situation itself and in the choice of an appropriate I.Q. test. Since persons with mental retardation have increased risks of comorbid psychiatric or behavioral dysfunction, examiners need to ensure that difficulties such as hyperactivity or poor frustration tolerance do not impede optimal test performance. Even in persons without behavioral dysfunction, certain personality characteristics may interfere with testing. Many persons with mental retardation look to others for solutions to difficult problems and are quick to acquiesce or become easily discouraged by failure. Certain techniques minimize these problems, for example, ensuring that the examinee succeeds with easy tasks before administering hard tasks, or providing positive incentives for effort.

Challenges also emerge with individuals who are nonverbal, and several I.Q. tests now exist that do not rely on expressive language (Table 34-1). Persons with certain genetic syndromes (e.g., Williams syndrome, Down syndrome) also present unique test challenges as many display syndrome-specific profiles of cognitive strength and weakness that are not easily described by a global I.Q. score. Persons with mental retardation and co-occurring sensory or motor deficits also require adaptation to routine test procedures. Finally, examiners need to be extra cautious when testing individuals from minority groups. Sparked by a series of lawsuits in the 1970s (e.g., *Larry P. v W. Riles*), controversies still rage about test bias and the appropriateness of I.Q. testing in children from minority groups. Court rulings have come out both against and in favor of I.Q. tests. Most psychologists readily agree that testing persons from minority groups requires extra sensitivity to language and cultural issues as well as a multimethod test approach, especially when diagnosing persons with mild mental retardation.

Adaptive Functioning Based on his pioneering work at the Vineland Training School in Vineland, New Jersey, Edgar Doll (1935) was the first to develop a formal definition and measure of adaptive behavior. Two decades later, the AAMR officially included deficits in adaptive behavior in its definition of mental retardation. Since then, deficits in adaptive behavior have been formally included in all definitions of mental retardation. Though meanings vary, adaptive behavior is typically viewed as the performance of behaviors required for social and personal sufficiency.

Further, adaptive behavior is an inherently developmental and social construct. Adaptive behavior changes as children grow into adolescence and adulthood, and demands for social adaptation are also defined by expectations from others—from one's family, society, and culture. Adaptive skills typically change across various settings; one's adaptive performance on the job or at school may differ from one's performance with friends or at home. Measurements of adaptive behavior, then, need to have a developmental orientation, to be socially and culturally sensitive, and to represent the many settings in which people live, work, and play.

In contrast to Doll's time, many instruments now exist that measure adaptive behavior across multiple domains (e.g., community skills, personal grooming). Table 34-2 summarizes commonly used measures of adaptive behavior, their age ranges, and various domains. All are acceptable measures of adaptive behavior in persons with mental retardation, yet the Vineland probably enjoys the most widespread use. Although most of these measures are administered as interviews with parents or other care providers, some of the newer measures are given directly to persons with mental retardation. Central to all measures is the idea that adaptive behavior is measured by typical, everyday performance, not ability. If individuals with mental retardation can perform certain behaviors but for any reason do not routinely do so, then they necessarily have compromised adaptive functioning.

Measure	Age Range	Domains Tested
Vineland Adaptive Behavior Scales (Sparrow, Balla, and Cicchetti, 1984)	Birth to 18	Communication: receptive, expressive, written; daily living skills: personal, domestic; community; socialization: interpersonal relations, play and leisure; coping skills; motor skills: fine, gross
Scales of Independent Behavior–Revised (Bruininks, Woodcock, Weatherman, & Hill, 1996)	Birth to 80+	Motor: fine, gross; social interaction and communication: social interaction, language comprehension, language expression; personal living skills: eating and meal preparation, toileting, dressing; community, living skills: time and punctuality, money and value, work
AAMR Adaptive Behavior Scales (Lambert, Nihira, and Leland, 1993)	3 to 18–21	Independent functioning, physical development, economic activity, school
Kaufman Functional Academic Skills (Kaufman and Kaufman, 1994)	15 to 75+	Functional math and reading skills, administered directly to person
Street Survival Skills Questionnaire (Grieve-Volker and McCarron, 1993)	9 to 45+	Administered to person; basic concepts, functional signs, tools, health

Table 34-2 Common Measures of Adaptive Behavior

Definition Controversies

Cognition Versus Adaptation What, then, is so controversial about cognitive and adaptive functioning? One controversy lies in the relative importance of these domains in defining mental retardation. Some workers object to the equal role afforded to cognitive and adaptive deficits and argue that intelligence is more important to the definition of mental retardation than adaptive behavior. These workers consider low I.Q. the benchmark of mental retardation and view adaptive deficits as correlates or sequelae of low I.Q. Further, unlike adaptive behavior, intelligence has firm theoretical and empirical roots, evolving from Francis Galton's pioneering work on individual differences to a host of modern views of intelligence, ranging from Howard Gardner's “multiple intelligences” to Robert Sternberg's triarchic theory of intelligence. Critics of adaptive behavior, then, embrace the strong theoretical and empirical history of the intelligence field and question whether adaptive behavior has equally robust psychometric properties.

Yet persons who have low I.Q. scores do not necessarily show deficits in everyday behavior. Proponents of adaptive behavior argue that cognitive and adaptive functioning are generally unrelated, with dramatically different theoretical underpinnings and measurement techniques. Indeed, some studies of persons with mixed etiologies find nonsignificant correlations between cognitive and adaptive functioning. A disjuncture between I.Q. and adaptive behavior may be particularly apparent in those with mild mental retardation. Variable outcomes were seen, for example, in Robert Ross and colleagues' 40-year study of students with mild mental retardation followed into adulthood. As adults, some of these subjects were completely to partially independent and adapted quite well, whereas others were highly dependent and showed poor social adaptation. Robert Edgerton, who followed up 48 residents of a large institution after their discharge and integration into the community, observed that the I.Q. of any given individual said little about what to expect in the outside world, and even very low I.Q. was no sure indicator. His case of Fred Barnett (pseudonym) is illustrative.

> Fred was the product of an uncomplicated pregnancy and delivery. Developmental milestones were remarkable for walking at 17 months of age, but slowness to talk and stammering, unclear speech. He was described as a little slow, but at 5 years of age when verbal and learning delays became more obvious, the family became concerned. At age 6, Fred was injured in an automobile accident with resultant coma for 5 days. Hospital records at the time suggested the possibility of slight neurological injury, but his mother attributed all of Fred's subsequent problems to this accident. He was unable to keep up in school and fell into greater conflict with his younger siblings. At 10 years of age he dropped out of school, and 4 years later was admitted to the Developmental Center. I.Q. testing over the years consistently placed Fred in the moderate range (middle to lower 50s). While in the institution Fred was seldom in trouble and worked on the wards, in the kitchen, and occasionally as a messenger. After 3 years he was discharged to the community, where despite nearly continuous supervision from employers and social workers, he was said to be in frequent trouble. In virtually all of his multiple places of employment, primarily as a dishwasher, his reputation was highlighted by temper, aggression, profanity, and boorishness. He was also involved in several minor difficulties with police including several charges of vagrancy. On one occasion he took an employer's car without permission and drove it recklessly. Over time, his conduct became more competent, and at the age of 32 he presented as a robust-appearing man perpetually smoking a cigar who gave every appearance of “being a happy-go-lucky, easygoing, altogether happy man. There is nothing in his appearance to suggest that he is anything less than a normal man, and his speech is likewise unexceptionable... It is only upon much closer inspection that Fred's intellectual deficit becomes apparent. To the casual observer, he is an ordinary man, competent to live within the not too demanding constraints of his life circumstances.”

Other studies, however, find strong correlations between cognitive and adaptive behavior, especially in persons with moderate-to-profound mental retardation. Significant I.Q.-adaptive behavior correlations are seen in studies of persons with moderate mental retardation of mixed etiologies, as well as in studies of individuals with distinctive syndromes such as fragile X syndrome, Down syndrome, Prader-Willi syndrome and 5p- (or cri du chat) syndrome.

Thus, I.Q. may set an upper limit or ceiling to adaptive accomplishments, which resolves at least some of the controversy about the relative importance of these two constructs. If so, this limit may be more pronounced at the lower levels of intelligence, with increased variability in adaptive outcome found in persons with mild mental retardation. Though many still feel that adaptive behavior does not belong in the definition of mental retardation, virtually all workers agree that adaptive skills are

critical to the long-term adjustment and success of people with mental retardation.

AAMR's New Definition Controversies about how best to conceptualize mental retardation have recently been piqued by the AAMR's so-called new definition of mental retardation. The new definition does not view mental retardation as an inherent characteristic of people, but as an interaction between individuals and their environments. With this assumption, the new definition eliminates traditional nosology based on level of cognitive impairment (i.e., mild, moderate, severe, profound) and instead proposes four levels of environmental supports (intermittent, limited, extensive, pervasive) across 10 different adaptive domains (e.g., health and safety, self-care, leisure). Thus, instead of giving a person a diagnosis of “moderate mental retardation,” the new definition specifies that a person has intermittent needs for supports in health and safety, limited needs for supports in self-care, and so on, across 10 domains. Critics argue that this system is unwieldy, is more pertinent for practice than diagnosis, and represents a giant step backward for research as it leaves researchers without a meaningful way to classify subjects.

Two other features of the new definition have been hotly debated. The new definition extends the I.Q. criterion from “70 and below” to “70 or 75 and below,” opening up the possibility for a more lenient I.Q. cutoff point of 75. Although a 5-point increase seems small, the Gaussian, or bell-curve, I.Q. distribution makes this high-end change in the I.Q. cutoff point particularly important. Donald MacMillian and colleagues note that “Small shifts in the upper limit have substantial consequences for the percentage of the population eligible to be diagnosed with mental retardation. *Twice as many people are eligible when the cut-off is 75 and below than when it is 70 and below*.” Many critics thus predict that if followed, the new definition will increase the size of the population with mental retardation, including increases in the overrepresentation of several minority groups.

Many critics also deride the new definition's adaptive behavior criteria, which now specify that documented deficits must be seen in at least 2 of 10 adaptive domains. Factor-analytical studies of adaptive behavior provide no empirical support for 10 domains and instead reveal from 2 to 7 factors of adaptive behavior. Further, no single measure of adaptive behavior taps these 10 particular domains, which forces workers to mix domains across two or more different tests or to simply rely on their clinical judgment.

For all these reasons, then, the new 1992 AAMR definition has generated much controversy in the mental retardation field. The controversy has spread to the field of psychiatry as well. In particular, the new definition was published just in time to create a dilemma for the APA and the DSM-IV work group on childhood disorders. The crux of the dilemma was whether DSM-IV should adopt the new AAMR definition—if the work group followed this organizational tradition, DSM-IV would depart radically from DSM-III-R. The solution was in large part a compromise. DSM-IV and ICD-10 criteria retained the I.Q. cutoff point of 70 and the traditional level-of-impairment nosology yet adopted the AAMR's adaptive behavior criteria. While some practitioners have embraced the new AAMR definition, others (both practitioners and researchers) have rejected the new definition in favor of DSM-IV or other more empirically based models. Given the many criticisms of the new definition, its fate remains in doubt.

EPIDEMIOLOGY

According to some estimates, approximately 1 percent of the population has mental retardation. This 1 percent figure is cited by DSM-IV and is roughly the percentage found in most prevalence studies. Yet the widely cited 1 percent figure hides a variety of controversies within mental retardation. In particular, many have reasoned that mental retardation is more frequent—nearer to a 3 percent prevalence rate—while others question whether some categories of mental retardation are disappearing altogether. These controversies have enormous societal implications. Based on the current U.S. population, a difference of even 1 percent means that an extra 2.6 million Americans have mental retardation and may require services.

Two Groups of Individuals With Mental Retardation To understand the prevalence issue, one must consider who, historically, has been thought to make up the retarded population. Given a Gaussian, bell-curve distribution of intelligence, 2.28 percent of individuals should fall two or more standard deviations below the general population average of I.Q.-100, which translates into scores below 70 on most psychometric tests. Following the Gaussian distribution, most of these individuals should have an I.Q. just below 70, and increasingly fewer individuals should have an I.Q. below 55 or 40 (three or four standard deviations below the mean). The percentage of persons whose I.Q. is below 40 (4 standard deviations below the mean) should equal 0.00003, or roughly 7800 persons in the entire United States (0.00003 × 260 million).

As early as 1960, Harvey Dingman and George Tarjan realized that many more persons with severe and profound mental retardation exist than would be expected by a Gaussian distribution. The relatively large number of persons at the lowest I.Q. levels seems to be due to the many persons with clear organic causes for retardation. Most prevalence studies find that as one goes lower down the I.Q. spectrum, increasingly high percentages of individuals show one or more clear organic causes for their mental retardation. Earlier, Leta Hollingworth observed that “a small percentage of mental deficiency is due to disease of the nervous system ... research points to the conclusion that approximately 90 percent of the mentally subnormal are the products of inferior germ-plasm. The remainder are the victims of organic causes, and are in a true sense pathological cases.”

In 1967 Edward Zigler proposed that there are two groups of persons with mental retardation. Retardation in the first group was caused by the usual (and so far unidentified) factors that distinguish individuals across the normal range of intelligence. Researchers have variously referred to this group as having cultural-familial, familial, or nonspecific mental retardation. Many of these individuals come from poor, minority, and low educational backgrounds, and low I.Q. scores are also common in parents or siblings. Like variations across the entire I.Q. spectrum, some interplay of environmental and biological factors seems to be involved in this type of retardation.

Zigler referred to the second group as having “organic” mental retardation. The retardation of this group was caused by one of many different pre-, peri-, or postnatal causes. Over the years, increasing numbers of persons with organic mental retardation have been identified. Mental retardation has now been linked to approximately 750 different genetic causes, and other causes also exist. In all cases, however, these individuals have one or more organic insults associated with their mental retardation.

The 1 Percent Versus 3 Percent Issue By summing percentages of the two groups, most workers of the 1960's and 1970's concluded that approximately 3 percent of individuals have mental retardation. In the early 1970s, however, Jane Mercer criticized the 3 percent figure and concluded that only 1 percent of the U.S. population has mental retardation. Although the 1 percent versus 3 percent debate may ultimately be unresolvable, its details illustrate some of mental retardation's most complicated issues.

I.Q. as the Sole Criterion of Mental Retardation The 3 percent prevalence rate considers I.Q. the sole criterion for mental retardation. That is, persons with an I.Q. below 70 are considered to have mental retardation, those with an I.Q. of 70 or above do not. Even considerations of the excess of individuals at the lowest I.Q. levels are based solely on the Gaussian distribution of I.Q.

But I.Q. is not the sole criterion of mental retardation. To be considered to have mental retardation, a person must have an I.Q. below 70 as well as deficits in adaptive behavior. If the two are perfectly correlated—if every person with a below-70 I.Q. also has deficits in adaptive behavior—then a 3 percent figure becomes tenable. If the two correlate weakly or not at all, much lower prevalence rates hold. Although strong I.Q.-adaptive correlations have been found for some organic groups and for individuals with severe and profound mental retardation, lower correlations probably hold for children with mild mental retardation. Throughout the entire population of persons with mental retardation, a moderate, but by no means perfect, correlation probably exists between I.Q. and adaptive behavior. To the extent that the two are uncorrelated for persons with an I.Q. fall below 70, less than 3 percent of individuals have mental retardation.

In recent years two new twists to the I.Q.-adaptive issue have arisen. The first concerns the possible change in the I.Q. cut-off criterion itself, from an I.Q. of 70 to one of 75. Given the Gaussian distribution of I.Q., as many people have an I.Q. between 70 and 74 as have an I.Q. between 0 and 69. A second issue concerns the change in the AAMR definition's adaptive criteria, from “deficits in adaptive behavior” to inadequate functioning in 2 of 10 adaptive domains. In both instances, more persons become eligible for a diagnosis of mental retardation.

I.Q. Remains Constant The 3 percent prevalence rate also assumes that an individual's I.Q. remains relatively stable over time. This assumption seems justified in some ways, unjustified in others. Once children reach the late-preschool years, correlations between separate testings many years apart become fairly high, particularly in samples of children with disabilities. Indeed, across the entire I.Q. distribution, a median correlation of .77 has been shown between testings at age 4 and age 12. However, correlations between I.Q. scores during infancy and later I.Q. scores are essentially zero. That is, with the exception of infants with whose I.Q. scores are very-low, any baby's I.Q. score (e.g., “developmental quotient”—or D.Q.—on the Bayley) has little relation to that child's I.Q. during later childhood or adulthood.

Further, different groups may show different trajectories of I.Q. with development. Children with Down syndrome show their highest I.Q. (or D.Q.) scores during the first year of life, then decline in I.Q. over the early and middle childhood years. Boys with fragile X syndrome also decline in I.Q., but their declines seem to begin at approximately 10 to 15 years. Conversely, children with cerebral palsy (half of whom have mental retardation) remain remarkably stable in their I.Q. scores over time, much like groups with mixed or nonspecific etiologies of mental retardation. Thus, considering I.Q. alone, the age of the individual must be kept in mind when considering the prevalence rates of different degrees of mental retardation.

Identification Rates Are Equal Across a Range of Variables A related issue concerns identification rates. To receive a diagnosis of mental retardation, individuals must be brought to the attention of professionals, tested, and found to meet definitional criteria. But the workings of this process vary widely for a host of reasons.

The first concerns age. Nearly every study finds that prevalence rates are low during the years from birth to age 5 or 6, increase rapidly until the adolescent years, and decrease thereafter. As might be expected, the increased intellectual demands of school seem to be responsible for such discrepancies.

But increased intellectual demands are not the only school variable affecting prevalence rates. Schools also vary in their formal procedures for identifying children with problems such as mental retardation. In many schools, these processes begin when the classroom teacher calls the child to the attention of a student study team, a multidisciplinary group designed to help teachers handle difficult children. The team then decides on within-class changes that the teacher might try or refers the child for further evaluation. Unfortunately, researchers currently have only a general sense of why teachers refer children to student study teams, why rates vary so much from one school or district to another, or what percentage of children who meet criteria for referral to student study teams ultimately receive a diagnosis of mental retardation. Such variations undoubtedly affect mental retardation prevalence rates.

These diagnostic issues are further complicated by regional and local variations. Many more children receive mental retardation diagnoses in certain areas of the country than in others. Using percentages of children enrolled in special education services across the 50 states during 1993, only 3.2 per 1000 schoolchildren were identified as having mental retardation in New Jersey and 4.5 in California versus 31.4 in Alabama and 25.4 in Kentucky. Although many factors probably relate to these interstate differences, Philip Massey and Richard McDermott note that three account for most of the variance: the percentage of a state's population with less than a ninth-grade education, the median household income, and the percentage of births to teen mothers.

In addition to these demographic factors, policy variables are also important. Over the past 20 years, the number of children given a diagnosis of mental retardation has decreased markedly. This decrease arises from many sources. Particularly in California, the *Larry P. v. W. Riles* case lessened reliance on I.Q. tests for diagnosis, particularly with minority children. Many even consider the very term *mental retardation* stigmatizing. Partly as a result, fewer children who might otherwise be considered to have mild mental retardation (i.e., I.Q. scores from 55 to 69) are receiving a mental retardation diagnosis.

Where are all the children with mild mental retardation? Partly, it seems, these children now receive diagnoses of learning disabilities. According to U.S. Department of Education statistics, over the school years 1976–77 to 1992–93, there was a 41 percent decline in the percentage of children considered to have mentally retardation by U.S. schools, while the percentage of children considered to be learning disabled jumped 198 percent.

But the move from mental retardation to learning disabilities is not the total answer. Indeed, many children with an I.Q. score below 70 do not have a diagnosis or either mental retardation or learning disabilities. They do not receive the diagnosis of mental retardation because of fear of stigma and decreased reliance on I.Q. tests, and they do not receive a diagnosis of learning disabilities because they do not function two or more standard deviations worse on achievement tests than on I.Q. tests (the usual definition of learning disorders). To stop the underdiagnosis of these children, some workers have suggested abandoning the category of mild mental retardation in favor of a new diagnostic category that highlights this group of low-intelligence, low-achieving, and less-often-served children.

Equal Mortality Rates A final assumption of the 3 percent figure is that persons with mental retardation and those without have equal life spans. This assumption seems obviously incorrect, particularly for persons with more severe and profound mental retardation.

Several variables lower life expectancies. For example, many early deaths—sometimes for persons in their 20s and 30s—are related to ambulatory and respiratory problems. Etiology may also play a role. Specifically, children with Down syndrome continue to be prone to heart problems, leukemia, and (by age 35) the plaques and tangles of Alzheimer's disease. Similarly, children with Prader-Willi syndrome are prone to hyperphagia and obesity, and complications of obesity (e.g., diabetes, heart disease) remain the main cause of higher death rates for persons with Prader-Willi syndrome.

How many persons with mental retardation are there? No one can say for certain. Given a definition that features an I.Q. below 70 and deficits in adaptive behavior, the prevalence rate is probably below 3% but above 1%. If one adopts a cutoff below I.Q. 75, higher percentages might be obtained. At present, it seems best to conclude that although the 1% figure is cited by most studies and professional-organizations, several difficult (possibly unresolvable) issues complicate even the simple question of how many persons have mental retardation.

ETIOLOGY

As noted by Esquirol in 1845, intellectual disability is not a disease in and of itself, but the developmental consequence of some pathogenic process. In 1898 William Ireland could classify idiocy into ten categories on the basis of etiology including “genetous, microcephalic, eclampsic, epileptic, hydrocephalic, paralytic, cretinism, traumatic, inflammatory, and idiocy by deprivation.” More recently, the AAMR offered an admittedly partial listing that enumerated over 350 causes of mental retardation.

With advances in medicine generally and in molecular genetics in particular, new causes of mental retardation or the genetic causes of formerly unspecified syndromes are identified each year. John Opitz counts over 750 genetic causes of intellectual disability alone. Eleanor Feldman notes that some 95 mental retardation syndromes have been linked to the X chromosome. The most common causes of mental retardation are Down syndrome, fragile X syndrome (accounting for 40% of all X-linked retardation), and fetal alcohol syndrome. Together, these three conditions are responsible for about 30 percent of all identified cases of mental retardation. Recently, early childhood anemia has been identified as a risk for mild to moderate mental retardation.

In addition to new causes being identified, additional knowledge is accumulating about underlying mechanisms in causes previously known. In Down syndrome, for example, a critical region on chromosome 21 (*DSCR1*) has been described at the 21q22.1-q22.2 locus. The availability of molecular markers for this and flanking regions has enabled clinicians to confirm cases of Down syndrome in the context of subtle translocations and chromosomal abnormalities other than trisomy. At least 170 different mutations leading to phenylketonuria have been reported in the phenylalanine hydroxylase gene.

As clinicians approach the cause of mental retardation in a particular patient, it is helpful to work from a broad framework initially. For example, an initial distinction might be drawn between congenital and acquired causes. For the latter, the timing of the insult that led to retardation may be further broken down into perinatal or postnatal causes. Congenital causes might be divided into genetic disorders or developmental disorders of brain formation, or more specifically still into inborn errors of metabolism and so on.

The frequency with which mental retardation associates with a particular cause depends highly on the population surveyed. For example, persons drawn from institutional settings, where severe and profound mental retardation and visible congenital anomalies are seen more commonly, are more likely to have an identifiable chromosomal abnormality than are persons with mild disability residing in community settings. In other words, more severe presentations of mental retardation are likely to have an organic cause. Table 34-3 presents a classification scheme with associated frequency estimates for mental retardation etiopathogenesis.

Cause	Examples	Estimated Frequency (%)
Prenatal		
Genetic disorders		4–28
Chromosomal aberrations	Down syndrome	
Single gene mutations	Tuberous sclerosis, phenylketonuria and other metabolic disorders, fragile X syndrome	
Multifactorial	"Familial" mental retardation	
Malformation syndromes due to microdeletions	Prader-Willi, Williams, and Angelman syndromes	
Congenital malformations		7–17
Malformations of the central nervous system	Neural tube defects	
Multiple malformation syndromes	Cornelia de Lange syndrome	
Exposure		5–13
Maternal infections	Congenital rubella, HIV	
Teratogens	Fetal alcohol syndrome	
Toxemia, placental insufficiency	Prematurity	
Other	Radiation, trauma	
Perinatal		2–10
Infections	Meningitis	
Delivery problems	Asphyxia	
Other	Hyperbilirubinemia	
Postnatal		3–12
Infections	Encephalitis	
Toxins	Lead poisoning	
Other postnatal causes	Traumas, brain tumors	
Psychosocial problems	Poverty, psychotic illness	
Unknown		30–50

Table 34-3 Mental Retardation Causes

Animal Models Over the past 100 years, increased understanding of basic brain organization and function and some of the causes of intellectual disability has been derived from animal models. It is hoped that such modeling will help clarify underlying mechanisms and suggest potential treatments for mental retardation.

Rodent models have been described for Down syndrome, phenylketonuria, fragile X syndrome, Lesch-Nyhan syndrome, and prenatal exposure to alcohol or cocaine. Trisomy 16 mice, a potential Down syndrome model, demonstrate visual, spatial, and attendant learning deficits. Homologues of the single-minded gene in the fruit fly *Drosophila* have been identified in both man and mouse, and the human gene maps to the Down syndrome critical region. Likewise, the *Drosophila* minibrain gene appears to be associated with the Down syndrome critical region. Both of the genes in *Drosophila* are involved in brain development, and their study will suggest possible mechanisms underlying mental retardation in humans. The curly tail mouse has been a valuable model for the study of neural tube defects, and a number of agents derived from this paradigm will likely be evaluated for their potential to prevent neural tube defects in humans. A mouse model for the fragile X syndrome, in which the *FMR*-1 gene is molecularly knocked out, is remarkable in that like humans with the disorder, the knockout mice show macroorchidism, learning deficits, and hyperactivity. These models may help to unravel the physiological role of the FMR1 protein in humans.

Genetic Mental Retardation Syndromes and Behavioral Phenotypes New genetic technologies have played a key part in elucidating causes of mental retardation, and the study of intellectual disability has contributed to advancing the understanding of the genetics of disease more broadly. Largely as a result of research into mental retardation syndromes, primarily fragile X and Prader-Willi syndromes, novel genetic mechanisms have been identified that have led to collective rethinking of classic Mendelian principles.

These remarkable genetic advances have sparked renewed research interest in the behavioral features of persons with fragile X, Prader-Willi, and other genetic syndromes. In their review of behavioral aspects of chromosomal disorders, Seymour Kessler and Rudolf Moos, for example, discuss only five syndromes. Since then, the number of disorders for which a “behavioral phenotype” has been suggested has grown considerably. A good working definition of a *behavioral phenotype* was proposed by Elisabeth Dykens, namely, a syndrome in which the probability of the expression of certain behaviors or constellations of behaviors is greater than expected. Stereotyped hand movements, for example, have been described in many contexts, but a particular form of stereotyped hand wringing is almost invariably seen in Rett's syndrome. Hyperphagia and obesity are nonspecific symptoms, but without aggressive dietary intervention, the probability of an individual with Prader-Willi syndrome becoming obese is nearly 100 percent.

Table 34-4 provides a representative sample of syndromes for which behavioral phenotypes have been described, some only anecdotally, others with controlled studies. Though the study of behavioral phenotypes is in its beginning stages, this line of work holds much promise for both clinicians and researchers. Identified syndrome-specific behavioral patterns guide clinicians toward some diagnostic possibilities over others, and these behavioral profiles also guide intervention and treatment. From a research standpoint, phenotypic work advances the understanding of gene-brain-behavior relation. To illustrate these advantages, behavioral and genetic features are reviewed below in fragile X, Prader-Willi, and Down syndromes.

[illegible]	[illegible]	[illegible]

Table 34-4 Representative Sample of Mental Retardation Syndromes and Behavioral Phenotypes

Examples of Behavioral Phenotypes in Mental Retardation Syndromes

FRAGILE X SYNDROME Fragile X syndrome, the most common inherited cause of mental retardation, results in a wide range of learning and behavioral problems, with males being more often and severely affected than females. The recently discovered fragile X gene (*FMR*-1) represents a newly identified type of human disease, caused by amplification (or excessive repetition) of a three-nucleotide sequence (CGG) in deoxyribonucleic acid (DNA). Above a certain threshold of these triplet repeats (about 200), people are fully affected with the syndrome. Numbers of repeats between the normal threshold of 50, and below 200, are termed “premutations.” As many as 1 in 259 women in the general population may carry the premutation, with 1 in 4000 males and 1 in 8000 females being fully affected with the syndrome.

Many of the behavioral features in people with fragile X syndrome are a function of the number of CGG repeats and a cascade of subsequent effects. The number of repeats is related to the extent to which the *FMR*-1 gene is methylated and subsequently silenced, or blocked from transcription and translation. This, in turn, affects the amount of FMR-1 protein (FMR1P) produced, and the range of clinical expression in fragile X syndrome is associated with the amount of FMR1P. Fully affected, fully methylated males have no FMR1P, whereas high-functioning males, though rare, have some FMR1P. In females, only one X chromosome carries the *FMR*-1 gene; the other allows production of some FMR1P. Because all females have one X chromosome randomly inactivated in all cells, the degree of involvement in females with fragile X is related to the ratio of normally active X chromosomes to the total number of X chromosomes (active plus inactive). About half of fully affected females show mild-to-moderate mental retardation and half have average I.Q. scores. But even among those with average intelligence, up to 80 percent may show specific problems in planning, memory, and attention. Neuroimaging studies link performance on some of these cognitive tasks to the size of the posterior cerbellar vermis, which is smaller in males and females with fragile X syndrome than in controls.

FRAGILE X SYNDROME AND AUTISTIC DISORDER Reportedly, many fully affected males with fragile X syndrome show autistic-like symptoms such as language delay, echolalia, stereotypies, self-injurious behavior, perseveration, poor eye contact, and tactile defensiveness. Not surprisingly, then, early workers tried to associate fragile X syndrome with autism, spurred on by the possibility of a common genetic cause of autism. Yet highly variable prevalence rates were found in this work, primarily because of discrepancies in the diagnostic criteria for autistic disorder.

This flurry of research faded as new studies suggested that instead of autism per se, many males showed a willingness to interact with others that was coupled with social and performance anxiety and mutual gaze aversion. Controlled studies and meta-analyses now suggest that only about 5 percent of males with fragile X have autistic disorder. Most affected males can be placed on a spectrum of social anxiety, shyness, avoidance, and gaze aversion. While some of these boys have anxiety

disorders or pervasive developmental disorder not otherwise specified, others may simply display slow-to-warm temperament styles, including shyness or social withdrawal.

Similarly, many females with fragile X syndrome exhibit variable levels of social dysfunction, primarily shyness, gaze aversion, and social anxiety. Many of these women meet clinical criteria for schizotypal disorder, showing interpersonal discomfort and difficulties in communication and social relationships. Fully affected women (with more than 200 CGG repeats) are more likely to have schizotypal disorder or schizotypal features than women with a premutation or appropriately matched control women without a fragile X. Although shyness is thus a central feature of the fragile X behavioral phenotype, affected females may also have a higher risk of depression than non–fragile X mothers of developmentally delayed children.

In addition to these difficulties, hyperactivity and attention deficits are seen in the vast majority of boys and girls with fragile X syndrome. Attention-deficit/hyperactivity disorder symptoms are higher among fragile X boys than among control subjects, and girls with a fragile X have lower prevalence rates of attention-deficit/hyperactivity disorder than fragile X boys. Among adults, problems in attending and in sustaining effort have been found in the neuropsychological profiles of women who carry the *FMR-1* gene, which may contribute to problems in mathematics, abstract reasoning, and planning.

Fragile X syndrome thus involves vulnerabilities toward shyness, gaze aversion, social anxiety, avoidant disorders, schizotypal disorder, attention/deficit/hyperactivity disorder, pervasive personality disorder not otherwise specified and, more rarely, autistic disorder. These difficulties vary in severity but are typically found in persons across the I.Q. spectrum, from those with moderate mental retardation to those with mild learning disabilities.

PRADER-WILLI SYNDROME First identified 41 years ago, Prader-Willi syndrome affects about 1 in 15,000 births and is best known for its food-related characteristics. Babies invariably show hypotonia and pronounced feeding-sucking difficulties, while young children between 2 and 6 years of age develop hyperphagia and food-seeking behavior such as foraging and hoarding. Hyperphagia is probably associated with a hypothalamic abnormality resulting in a lack of satiety. Food preoccupations are lifelong, and without prolonged dietary management, affected individuals invariably become obese. Complications of obesity remain the leading cause of death in persons with this syndrome.

Prader-Willi syndrome is the first known human disease to show the effects of genomic imprinting, that is, genes being modified and expressed differently depending upon whether they are inherited from the mother or the father. About 70 percent of Prader-Willi syndrome cases are caused by a paternally derived deletion on the long arm of chromosome 15. The remaining cases are attributed to maternal uniparental disomy of chromosome 15, in which both members of the chromosome 15 pair come from the mother. In either case, the paternally derived contribution to this specific region of the genome is missing. When missing information in this same region of chromosome 15 is maternally derived, it results in a completely different and more severe developmental disorder, Angelman's syndrome.

Although people with Prader-Willi syndrome invariably obsess about food, a remarkably high proportion also show nonfood obsessions and compulsive behaviors. These nonfood symptoms include skin picking; hoarding; needing to tell, ask, or say things; and having concerns with symmetry, exactness, ordering, arranging, cleanliness, and sameness in daily routine. Often these symptoms are associated with distress or adaptive impairment, suggesting a marked risk of obsessive-compulsive disorder in this population. Indeed, the authors estimate that obsessive-compulsive disorder is many times more likely in persons with Prader-Willi syndrome than in the general population of persons with mental retardation.

In addition, even compared with others with mental retardation, children and adults with Prader-Willi syndrome display high rates of temper tantrums, aggression, stubbornness, underactivity, excessive daytime sleepiness, and emotional lability. Coupled with food seeking, these impulsive behaviors often lead those with Prader-Willi syndrome to need more-restrictive care than would be predicted by their mild levels of mental retardation.

Psychopathology may be associated with the derivation of the individual's Prader-Willi syndrome. Preliminary findings from ongoing work suggest some behavioral differences between cases due to paternal deletion and those due to maternal uniparental disomy. Patients with deletions may have a lower I.Q., especially verbal I.Q., and more frequent or severe problem behaviors such as skin picking, hoarding, temper tantrums, overeating, and social withdrawal. Although a dampening of symptom severity is suggested in many uniparental disomy patients, the authors also observe occasional patients with more-severe problems in uniparental disomy, primarily autistic-like features and relatively low I.Q. scores.

Many people with Prader-Willi syndrome then, are at increased risk for obsessive-compulsive, impulse control, and affective disorders. Yet even those who do not meet diagnostic criteria for these psychiatric disorders often show significant maladaptive behaviors that interfere with optimal adaptive functioning.

DOWN SYNDROME Occurring in approximately 1.2 of 1000 live births, Down syndrome is the most common chromosomal abnormality leading to mental retardation. It usually results from nondisjunction of chromosome 21. The syndrome has been the subject of considerable investigation.

Studies of children have revealed distinctive cognitive and linguistic profiles. Particular strengths have been noted in visual (versus auditory) processing. Language impairment can be extensive, with particular difficulty in expressive language, grammar, and pronunciation.

In contrast to the relative overall weakness in the cognitive arena, those with Down syndrome are often described as being particularly socially adept. However, studies of adaptive functioning suggest that these skills are not uniform, with communication scores trailing behind measures of daily living skills and socialization. The stereotype of a Down personality—happy, good tempered, affectionate, placid, and stubborn—has been difficult to verify.

Compared with others with mental retardation, persons with Down syndrome appear to suffer less often and less seriously from psychopathology. Adults with Down syndrome appear somewhat less prone to psychiatric disturbance than controls. This trend seems to extend to children and adolescents as well; rates of psychiatric and behavioral problems exceed those in the general population but are appreciably lower than those in other groups with mental retardation. Commonly noted problems include attention difficulties, impulsivity, hyperactivity, and aggression. In contrast to these problems, depression seems to be less common among children and adolescents than expected norms. Autism and pervasive developmental disorders appear to be relatively rare.

DIAGNOSIS AND CLINICAL FEATURES

Table 34-5 presents the DSM-IV diagnostic criteria for mental retardation, and Table 34-6 lists the ICD-10 criteria.

A. Significantly subaverage intellectual functioning: an IQ of approximately 70 or below on an individually administered IQ test (for infants, a clinical judgment of significantly subaverage intellectual functioning).
B. Concurrent deficits or impairments in present adaptive functioning (i.e., the person's effectiveness in meeting the standards expected for his or her age by his or her cultural group) in at least two of the following areas: communication, self-care, home living, social/interpersonal skills, use of community resources, self-direction, functional academic skills, work, leisure, health and safety.
C. The onset is before age 18 years.

Code based on degree of severity reflecting level of intellectual impairment:

Mild mental retardation: IQ level 50–55 to approximately 70
Moderate retardation: IQ level 35–40 to 50–55
Severe mental retardation: IQ level 20–25 to 35–40
Profound mental retardation: IQ level below 20 or 25
Mental retardation, severity unspecified: when there is a strong presumption of mental retardation but the person's intelligence is untestable by standard tests

Reprinted with permission from American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorder*, ed 4, © American Psychiatric Association, Washington, DC, 1994.

Table 34-5 DSM-IV Diagnostic Criteria for Mental Retardation

Table 34-6 ICD-10 Diagnostic Criteria for Mental Retardation

Levels of Mental Retardation At present, the field lacks a classification system that reflects the remarkable diversity, strengths, and competencies of people with mental retardation. The new AAMR definition rejected traditional nosology because of its emphasis on deficits in people, yet it fell short of developing a successful alternative based on strengths or competencies. Until a viable new scheme is developed, it makes sense to use traditional nosology to describe persons with mental retardation.

Mild Mental Retardation Mild mental retardation (I.Q., 55 to 70) characterizes the largest group of persons with mental retardation, possibly as many as 85 percent of the total. These individuals appear similar to nonretarded individuals and often blend into the general population in the years before and after formal schooling. Many achieve academic skills at the sixth grade level or higher, and some graduate from high school. As adults, many of these individuals hold jobs, marry, and raise families—yet at times they may appear slow or need extra help negotiating life's problems and tasks.

Johnny is a 10-year-old boy with mild mental retardation. Although from birth his parents considered him somewhat “slow,” Johnny was not diagnosed with mild mental retardation until his early grade-school years. To this day, no clear diagnosis has been provided for Johnny's mild mental retardation. As measured by the Stanford-Binet IV, his I.Q. is 67, with no significant differences between his verbal and perceptual processing scores. Johnny does, however, show impulsivity and problems in attending. These attentional problems are apparent on his Stanford-Binet IV auditory short-term memory subtests and in his adaptive behavior. For example, he has difficulty waiting his turn at school and needs many prompts to follow through with directions from his teachers and parents.

Johnny's mental retardation first became apparent at the end of first grade. At that time, a student study team at his school worked with Johnny's classroom teacher and the resource room teacher to help Johnny improve his basic work organization skills and increase his attention span. At his parents' request, Johnny was also evaluated for attention-deficit/hyperactivity disorder by a local psychiatrist, who subsequently prescribed stimulant medication that seemed to help.

Traditionally, individuals with mild mental retardation were thought to show relatively few clear-cut organic causes for their delay. While this may still be the case, recent years have seen an increase in the number of people with genetic syndromes who function in the mild range. Examples include most people with Prader-Willi syndrome and some males and most females with fragile X syndrome.

A more striking characteristic, however, is that more people with mild mental retardation come from minority groups and low socioeconomic backgrounds than would be expected from their percentages in the general population. This overrepresentation of minority groups has been used to criticize I.Q. tests and to highlight the importance of both environmental-cultural and genetic influences on mental retardation.

Moderate Mental Retardation Moderate mental retardation (I.Q., 40 to 55) is seen in approximately 10 percent of those with mental retardation, including persons with more-impaired cognitive and adaptive functioning. Individuals with moderate mental retardation typically receive their diagnosis in their preschool years, and some show a clear organic cause for their delay. Persons with Down syndrome often function in this range, as do many adolescents and adults with fragile X syndrome. Most children with moderate mental retardation require special education services and achieve academic skills at the second to third grade level. Supportive services are needed throughout life. With proper supports, many live, work, and thrive in their local communities. A study by Ross and colleagues found that 20 percent of persons with an I.Q. from 40 to 49 lived independently, 60 percent were considered partially dependent, and 20 percent were totally dependent on others. Similarly, some individuals in this range are employed in the competitive job market and need minimal job supervision, whereas others require more extensive supervision on the job and may work in sheltered workshops or other, more segregated settings.

Harold is a 16-year old boy with moderate mental retardation. Diagnosed at birth with Down syndrome, Harold has been mainstreamed with same-aged peers throughout his schooling. Harold, whose I.Q. is 53, has some difficulties with articulation, and his language consists mainly of two- to four-word sentences. Recently, Harold was introduced to reading; he now can read 50 sight words. Although he can occasionally be stubborn and argumentative, Harold is well liked by his teachers and classmates, and his parents boast that Harold has been a charming, pleasant child from infancy on. In addition to attending his classes at the local high school, Harold serves as the assistant manager for his high school's cross-country team and actively participates in Special Olympics.

Severe Mental Retardation Severe mental retardation (I.Q. 25 to 40) occurs in about 3 to 4 percent of persons with mental retardation. Individuals at this level often have one or more organic causes for their delay, and many show concurrent motor, ambulatory, and neurological problems as well as poorly developed communication skills. Most persons with severe mental retardation require close supervision and specialized care throughout their lives. Some individuals learn to perform simple tasks or routines that facilitate their self-care or their ability to perform in a sheltered workshop or preworkshop-type setting.

Robert was born prematurely at 8 months, weighing 4 pounds 7 ounces. Because of hyperbilirubinemia and respiratory problems, Robert remained in the hospital for several weeks before going home. For the first 2 years of his life, his mother recalls that Robert was quite a “passive” baby who would seldom even make his needs known; he did not even cry when he was hungry or wet. Similarly, while Robert gave no indication that he enjoyed cuddling early on, neither did he indicate that he disliked it. At approximately 2 years of age, Robert was hospitalized with a very significant episode of chicken pox that apparently included pneumonia. During that illness, he also had at least one seizure, which prompted the initiation of phenobarbital (Solfoton) treatment. During that hospitalization, physicians observed Robert's developmental delays: he was not bowel or bladder trained and did not talk or walk independently. A subsequent evaluation yielded a diagnosis of severe-to-profound mental retardation. His mother recalls receiving a recommendation at this time to consider institutional placement for Robert, given the severity of his disabilities. She recalls that Robert continued to acquire skills, including bowel and bladder training at approximately 4 years of age, and that he began to speak some years later. He also began to change from a relatively placid child to one with more motor activity, and mother describes him as having been quite hyperactive and also as having exhibited insomnia. Robert was placed in a residential school setting at age 6. During this placement, Robert's medications were tapered and discontinued without incident, and through a behavioral modification program, his behaviors were significantly improved, which lead to a recommendation for discharge when Robert was ten. In the discharge summary, Robert was characterized as a young man who was very pleasant and cooperative, who enjoyed working with others and even enjoyed hugs. There were no reports of symptoms that might be regarded as autistic, either stereotypies or unusual preoccupations. Following his discharge from the institutional setting, Robert returned home and lived with his family until the age of 21, when he moved into a sponsored living situation with one other peer. Robert, now 37, is a handsome man who is quite particular about his appearance. He speaks in one- and two-word phrases and follows simple requests without difficulty. He does not consistently demonstrate good safety awareness and requires supervision when crossing streets. For approximately the last 5 years, Robert has been treated with carbamazepine (Tegretol) for staring spells. When carbamazepine treatment started, Robert was described as having become somewhat irritable. This adverse effect has since resolved, and he is doing well at the moment in all settings, both at his structured workshop and at home, with no problems with his sleep, appetite, or behavior.

Profound Mental Retardation Profound mental retardation (I.Q. of 25 or below) affects relatively few individuals (1 to 2 percent) and involves pervasive deficits in cognitive, motor, and communicative functioning. Impaired sensory-motor functioning is often seen from early childhood on, and most individuals require extensive

training to complete even the most rudimentary aspects of self-care such as eating and toileting. The vast majority of people with profound mental retardation have organic causes for their delay, and most require total supervision and care throughout life.

> Martha is a 24-year-old woman with profound mental retardation. Diagnosed from birth as having 5p syndrome, Martha's I.Q. is below testable limits, she walks only with difficulty, and she shows limited toileting and grooming skills. Martha is nonverbal but seems to understand simple commands and phrases. She is also extremely hyperactive and prone to irritability and temper tantrums, especially when going from one activity to another. Despite her difficulties, Martha shows an interest in others, has good eye contact, and is well liked by the residents and staff of her community-based group home. Martha attends a day-training program, where she is learning grooming and toileting, as well as basic visual-motor tasks.

Metabolic Testing The presence of certain clinical characteristics may prompt the focused use of metabolic studies to identify inborn errors of metabolism. A partial listing of signs and symptoms that might raise the suspicion of a metabolic disorder appears in Table 34-7. Testing in the presence of clinical suspicion might include plasma and urine amino and organic acid studies, acid-base balance determination, thyroid function tests, lysosomal enzyme analysis, plasma and urine carnitine analyses, and plasma very long chain fatty acid assays, among others. Because neonatal screening picks up phenylketonuria, hypothyroidism, and galactosemia, the yield for subsequent unselected screening for inborn metabolic disorders is extremely low. In general, in the absence of specific indications other than mere mental retardation, metabolic testing is likely to be of little diagnostic value.

Growth abnormality
Recurrent, unexplained illness
Seizures
Ataxia
Loss of psychomotor skills
Hypotonia
"Coarse" appearance
Eye abnormalities (cataracts, ophthalmoplegia, corneal clouding, retinal abnormality)
Recurrent somnolence/coma
Abnormal sexual differentiation
Arachnodactyly
Hepatosplenomegaly
Metabolic/lactic acidosis
Hyperuricemia
Hyperammonemia
Low cholesterol
Structural hair abnormalities
Unexplained deafness
Bone abnormalities (dysostosis, occipital horns, punctate calcifications)
Skin abnormalities (angiokeratoma, "orange-peel" skin, ichthyosis)

Reprinted with permission from Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, Cunniff C, Graham JM Jr, Jones MC, Kaback MM, Moeschler JL, Schaefer GB, Schrander JJ, Tarleton J, Opitz J: Evaluation of mental retardation: Recommendations of a consensus conference: American College of Medical Genetics. Am J Med Genet 72:468, 1997.

Table 34-7 Selected Clinical Findings and Laboratory Abnormalities That Increase Suspicion for Underlying Metabolic Disorder

DEVELOPMENTAL CONSIDERATIONS IN CHILDREN WITH MENTAL RETARDATION

Throughout the twentieth century children with retardation have been an ongoing topic of developmental research. Small groups of researchers have examined the development of various Piagetian cognitive domains, language, social skills, and even morality. By now, much is known about how children with mental retardation develop.

Besides providing basic milestones or natural histories, these studies help reveal the underlying processes of development. Over the past decade, researchers have begun to specify the ordering of development (sequences), the ways in which levels in certain areas relate to levels in other areas (cross-domain relations), and the points at which the speed of development (rate) changes. Such analyses have begun to be applied to children with different causes of mental retardation.

In addition, developmental analyses have been extended beyond children themselves. Many developmental psychologists examine mother-child interactions, family behaviors in families of different-aged children, and the child's ongoing interactions with peers, neighbors, and schools.

Child-Related Aspects of Development Following developmental psychology's study of normal children, researchers throughout this century have examined children with mental retardation. Such studies have focused in two main areas: development of children themselves and children's interactions with their surrounding environments.

Children with mental retardation are now known to traverse the same sequences of development as nonretarded children, but children with certain forms of mental retardation show intellectual strengths and weaknesses not generally found in those who are nonretarded. Children with mental retardation also show certain ages and tasks with particularly slowed development that seem to be related to specific types of mental retardation; the reasons for such slowing are little understood.

Mothers of children with mental retardation resemble mothers of typical children in shortening their speech, focusing on key words, and in other ways structuring the environment for their child's development. However, they seem much more didactic and intrusive. Families of children with mental retardation differ widely one from another. When the child with retardation shows less maladaptive behavior and when the mother's style of coping is more problem focused, parental and family stress seems to be lowered.

Sequences Normal children have a specific, possibly universal, order to their development. For example, in Piagetian cognitive development, children proceed from sensorimotor, to preoperational, to concrete operational, to formal operational thought. Even within these four larger stages, smaller orderings hold; within sensorimotor development, normal infants proceed in order through Piaget's six substages in each of several subdomains.

Do children with mental retardation also follow a so-called similar sequence in development? For almost 30 years, similar sequences have been found for a variety of children over many tasks. Similar sequences even hold for children with genetic or other organic causes for their retardation. The only possible exceptions include some children with uncontrollable seizures (which make accurate testing difficult) and some autistic children, who may show different orderings because of their particular disabilities on certain social tasks. Such sequences have been noted in many areas: in almost twenty Piagetian domains, in symbolic play, and in linguistic grammar and pragmatics. For almost all children and for a wide variety of behaviors, children with mental retardation and other children develop along similar sequences.

Cross-Domain Relations If children with retardation proceed in the same developmental ordering as other children, do they also show the even, or flat, profiles that groups of children without mental retardation show from one domain to another? Such a “similar structure” to development was proposed by Zigler in the late 1960s.

Unlike similar sequences—which hold for all children with mental retardation—cross-domain structures may differ, depending on which retarded group one considers. With a few exceptions, children with cultural-familial mental retardation do show even or near-even performance across various intellectual domains.

In contrast, children with different organic forms of mental retardation show specific intellectual strengths and weaknesses. For example, three separate groups are weaker in sequential (i.e., bit-by-bit, serial) processing than in simultaneous (i.e., Gestalt, holistic) processing or achievement abilities. Such sequential deficits are found in boys with fragile X syndrome, children with Prader-Willi syndrome, and children with Smith-Magenis syndrome. Children with Down syndrome may have particular weaknesses in expressive (versus receptive) communication and special problems in grammatical abilities. Children with Williams syndrome show extra deficits in visuospatial processing skills, and some subsets of these children show heightened abilities in language. Different causes of mental retardation thus result in different characteristic intellectual strengths and weaknesses.

Rates A further issue concerns children's rates of development. By definition, children with mental retardation develop at slower rates than other children. Nonspecific or mixed groups show reasonably steady rates of development, evinced by stable a I.Q. after the early childhood years. Children with other types of mental retardation, however, may have periods of speeded or slowed development.

Times of speeded or slowed development seem to be related to either age or task difficulties. In age-related slowing, most children with a specific type of mental retardation slow in their development during a certain age span. For example, boys with fragile X syndrome have stable developmental rates until about age 9 or 10, when development slows. Similarly, both cross-sectional and small-scale longitudinal studies show that children with Down syndrome slow from ages 6 to 11;

throughout this age span, these children slow in their development of linguistic grammar, intellectual skills, and adaptive behavior. It remains unknown why such age-related slowings occur in both fragile X and Down syndromes, whether all children with the disorder show such slowings, or if slowings are seen across all domains.

A second type of slowing relates to pronounced difficulties mastering a specific developmental task. Unlike age-related slowings, these slowings occur at a variety of ages, whenever children with a particular type of retardation reach a specific developmental task. Consider young children with Down syndrome (who also show age-related slowings from 6 to 11 years). These children have difficulty with certain tasks of infant intelligence, even after one accounts for their already-slower rates of development, and toddlers are slower to develop expressive skills than receptive language skills. As a result of both age- and task-related slowings, children with Down syndrome generally exhibit their highest I.Q. scores during the first year of life.

An overview of development in children with mental retardation shows that these children develop along the usual sequences. At the same time, different subgroups show exceptional strengths or weaknesses in one or another area and have particularly slowed development in mastering different tasks or during differing age spans. Why different mental retardation groups show such developmental patterns remains generally unknown.

Contextual Aspects of Development

Mother-Child Interactions Following from research on interactions between typically developing children and their mothers, mother-child interaction studies in mental retardation began during the mid-1970s. Some similarities and some differences were found in mother-child interactions with children with mental retardation and those with other children of the same mental age (M.A.). The two groups of mothers behave similarly on what might be called the structural aspects of behavior. In language, for example, they both shorten their speech, simplify their vocabularies, raise their pitch, repeat key words, and generally provide typical “motherese.” Thus mothers of typically developing 2-year-olds and mothers of children with retardation who are at 2 years M.A. act almost identically.

At the same time, mothers of children with mental retardation show very different styles in their interactions. They are often more didactic and intrusive than mothers of typically developing, M.A.-matched children. They more often initiate the topic of interaction, take longer speaking turns, talk at the same time as their child, and in other ways control the interaction. As a result, the child has fewer opportunities to initiate and control the conversation.

Why such stylistic differences occur remains unclear. The prevailing view holds that mothers of children with mental retardation are worried about their child's development, and in their zeal to teach and instruct their children, they may go overboard. As one mother of a child with Down syndrome explained: “It's put him on your knee and talk to him, that's the main object. Play with him, speak to the child, teach him something.”

A more didactic interactive style may not, in fact, be most beneficial for the child's development. Although few studies have been performed, young children with retardation seem to develop faster when their mothers are less intrusive. One study examined mother-child interactions at 23 months and child outcomes at 36 months. Children whose mothers constantly redirected the child's attention had lower receptive language ages at 36 months than those whose mothers commented on the object at which the children was looking. These studies suggest that mothers might better respond to the child's interests and initiatives than be didactic or intrusive.

Family Reactions Over the past two decades, family studies in mental retardation have changed from a predominantly negative to a more balanced perspective. Before the early 1980s, these families were perceived to be families in crisis. Mothers were examined for their mourning reactions, couples for divorce, and mothers, fathers, and unaffected siblings for the presence of depression and other forms of psychopathology. The basic findings were that families of children with disabilities may suffer more divorces, parents and siblings are somewhat more prone to depression, and families and individual family members have more difficulty when there is only one parent, when the mother receives little support from the husband, or when the family is of low socioeconomic status.

In the early 1980s family researchers began to change their conception from the earlier focus on family pathology to families facing increased stress. Children with mental retardation might add stress to the family system, but this stress could result in negative or positive adaptation. This change implies a more normative view of these families: every family faces a variety of stressors (e.g., illness, loss of job), and having children with retardation can be thought of as a stressful situation.

The move from a negative to a stress-coping perspective has other implications. First, it permits change from a predominantly “between-group” research strategy. Previously, most studies compared families of children with disabilities to families of typical, nondisabled children of the same chronological age. Recently, the between-group focus has been complemented by within-group studies. Since families of children with mental retardation vary in their ability to cope, what causes better adjustment in one family and worse adjustment in another? What child, parent, or family characteristics foster better adaptation?

To date, most work has examined such child characteristics as age and degree of impairment. Findings have been inconsistent for both variables. Some studies find it more stressful to parent older children with retardation; others suggest that families experience more stress when the child begins puberty (11 to 15 years) and again when the child reaches early adulthood (20 to 21 years). Still other studies find no relation between increased family stress and the child's age. Similarly, some studies find that it is more difficult to parent children with more severe levels of retardation, and others do not.

Another, less-often examined characteristic involves the extent of the child's maladaptive behavior. So far, strong connections have been found between child maladaptive behavior and family stress in families of children with retardation of mixed etiology, with Prader-Willi syndrome, and with 5p (*cri du chat*) syndrome. Why such correlations occur is unclear. Although most researchers feel that child maladaptive behavior increases familial stress, increased familial stress may also elicit child behavior problems.

Another important variable concerns the coping style of the parents. Across several studies, parents who more actively, constructively attempt to deal with their child did better than parents who adopted a palliative coping style, one that either dwells on or ignores parental emotions. This difference in personality style may help buffer parents and families from the increased stresses of parenting a child with mental retardation.

COMORBIDITY

Attention-Deficit/Hyperactivity Disorder The rates of attention-deficit/hyperactivity disorder in mental retardation are estimated to be between 9 and 18 percent. Although considerable research has been done on the treatment of attention-deficit/hyperactivity disorder, virtually all of these studies have specifically excluded children with cognitive disability. Moreover, the diagnosis of attention-deficit/hyperactivity disorder is based upon developmental considerations, namely, significantly greater motoric hyperactivity, impulsivity, and inattention than is expected for a given developmental age; thus the threshold for diagnosis in persons with severe-to-profound mental retardation should be elevated.

For persons with mental retardation, the diagnosis of attention-deficit/hyperactivity disorder is qualified as being excessive for an individual's mental age. In the context of profound mental retardation, attention span, distractibility, or on-task behavior are predictably quite variable. Individuals given the diagnosis of attention-deficit/hyperactivity disorder in this context should exhibit shorter attention spans, greater psychomotor activity, and more remarkable impulsivity than their peers with similar levels of retardation. Often clinicians encounter situations in which an individual does not evidence remarkable psychomotor activity or attentional difficulties but may be unusually impulsive. In these situations one should entertain the diagnosis of an impulse control disorder not otherwise specified. Such a diagnosis, for example, might be appropriate for an individual who inexplicably strikes out at a peer in the absence of any identifiable environmental stressor.

Impulse-Control Disorders: Self-Injury and Aggression Every effort should be made to treat the underlying cause of self-injurious behavior or aggression, not merely suppress the behavior. However in some cases, for example, stereotyped movement disorder with self-injurious behavior, the symptom is essentially the diagnosis. Aggression and self-injurious behavior are common in mental retardation and increase as cognitive disability becomes more severe. Self-injurious behavior typically is a chronic, repetitive, and frequently stereotyped behavior causing trauma. It occurs in the context of specific genetic syndromes (e.g., Lesch-Nyhan syndrome and Smith-Magenis syndrome) but more commonly in persons with unknown or nonspecific causes for their mental retardation.

A review of psychiatric consultations in an institutional setting reported that self-injurious behavior was specifically cited as a reason for referral in 36 percent of the 251 cases examined. However, self-injurious behavior as a presenting symptom generally does not help to predict the ultimate psychiatric diagnosis. The single exception perhaps is the diagnosis of stereotyped movement disorder, in which the designation “with self-injurious behavior” was specifically created for persons who engage in this behavior in the absence of other diagnosable psychopathology.

Since self-injurious behavior and aggression are nonspecific symptoms, one must consider the presence or absence of a variety of factors to reach a presumptive diagnosis: the chronicity of the behavior, whether it may serve a communicative function, whether it is invariant in topography (e.g., hitting only the right ear, suggesting an ear infection), whether it is situational, whether it occurs in concert with regression from a previous level of function, and whether any associated neurovegetative signs correlate with its onset.

Oppositional Defiant Disorder and Conduct Disorder The DSM-IV diagnosis of oppositional defiant disorder or conduct disorder also requires comparisons with others of similar mental age. Further, both diagnoses assume some deliberateness on the part of patients (e.g., disobedience motivated by spite or resentment), which can be very difficult to discern in nonverbal subjects with profound cognitive deficits.

Anxiety Disorders Specific anxiety disorders (e.g., separation anxiety, overanxious disorder, obsessive-compulsive disorder, panic disorder, and generalized anxiety disorder) rely heavily on an individual's ability to describe the subjective symptoms of anxiety. According to DSM-IV, concurrent pervasive developmental disorder specifically precludes the diagnosis of most of these disorders. Yet, some individuals have constellations of signs and symptoms that are best captured in the anxiety disorder spectrum. Patients who are clearly avoidant, who exhibit autonomic arousal in the face of stimuli that most of their peers would not find aversive, and who evince other features of anxiety but cannot articulate their subjective states might be given a diagnosis of anxiety disorder not otherwise specified.

When individuals engage in behavior that appears compulsive or driven and seems ego-alien, the diagnosis of obsessive-compulsive disorder not otherwise specified might be considered. Often these patients engage in self-restraint (securing their extremities in their clothing) or cling to their parents or careproviders—seemingly to prevent self-injurious behaviors. Simple repetitive behaviors or an insistence on sameness that can accompany pervasive developmental disorder should not be given a separate obsessive-compulsive spectrum diagnosis. It may be useful to organize treatment with an additional diagnosis beyond “autism” if, for example, the focus of treatment is a sleep disturbance, impulsive aggression, or the like.

Although common, anxiety disorders appear to be underdiagnosed in persons with mental retardation. Variability in prevalence rates, from 1 to 25 percent, is attributed to difficulty in making a diagnosis. Moreover, an individual with mental retardation may not be able to identify subjective anxiety as an underlying cause of distress, and a patient's aggression or agitation may suggest a disorder of impulse control rather than reflecting underlying anxiety. Common symptoms of anxiety in persons with mental retardation include aggression, agitation, compulsive or repetitive behaviors, self-injury, and insomnia. Panic may be expressed as agitation, screaming, crying, or clinging, which might even pass for delusional or paranoid behavior. Phobias also occur in this population and may even be more common in persons with developmental disabilities. Ruth Ryan has noted that persons with developmental disabilities are at high risk for abuse, which puts them at a greater risk for posttraumatic stress disorder. With data suggesting that it may be seen in nearly 8 percent of the general population, posttraumatic stress disorder is an important diagnosis to consider in individuals with mental retardation.

Eating Disorders Because the diagnostic criteria rely upon subjective experiences, the diagnoses of anorexia nervosa and bulimia are effectively precluded for individuals with severe or profound mental retardation. For example, one would be hard pressed to identify such things as classic distortions in body image or guilt feelings associated with bingeing in nonverbal patients. Food refusal or self-induced vomiting would have to be considered atypical eating disorders if they occurred in the absence of other diagnosable disorders (e.g., depression or rumination). Pica is perhaps the most common eating disorder among persons with mental retardation.

Mental Disorders Due to a General Medical Condition One could argue that by definition everyone with mental retardation has some organic cerebral dysfunction, and thus any psychiatric illness should be regarded as organic or due to a general medical condition. In his study of psychiatric illness in a sample of institutionalized patients with Down syndrome, Frank Menolascino (1984) appropriately argued that psychiatric nosology did not have to be reinvented to accommodate individuals with a “tissue diagnosis.” Moreover, he reasoned that patients with so-called dual diagnoses of mental illness and Down syndrome need to be distinguished from those with Down syndrome alone. Thus, the application of the diagnoses of organic mental syndromes and disorders is best approached as if patients do not have mental retardation. The same principle should apply to Axis II personality disorders. The diagnosis of a personality disorder due to a general medical condition is best reserved for individuals whose preexisting personality was altered in a pathological way by some additional cerebral insult. In essence, this category was reserved for patients whose mental retardation is acquired, usually secondary to trauma experienced in childhood or early adolescence.

Psychosis Patients with developmental disorders are at increased risk for schizophrenia, bipolar disorder, and other mental illnesses that may include symptoms of thought disorder and hallucinations. The diagnosis of schizophrenia essentially requires that a patient relate the experience of delusions or hallucinations. As has been suggested by others, the diagnosis of classic schizophrenia is arguably impossible for individuals with profound mental retardation and limited communicative ability. Nonetheless, some individuals display presumptive evidence of response to hallucinations (e.g., striking or shouting at empty space, throwing imaginary peers from furniture) or adopt catatonic postures that can appear to be of psychotic origin. In these cases the diagnosis of psychotic disorder not otherwise specified should be considered if these signs exist in the absence of sufficient evidence to warrant the diagnosis of a supervening mood disorder.

Mood Disorders Even in profound mental retardation, the diagnosis of mood disorders is fairly straightforward. Generally, a change in mood from baseline is obvious (recent-onset lability, tearfulness, mood elevation, irritability). If it is coupled with changes in interests, activity level, sleep, appetite, or sexual behavior of sufficient duration and causing sufficient impairment in habilitative function, the diagnoses of mania or of depression can be made in nonverbal patients.

Mood disorders are not uncommon in persons with mental retardation. Learning problems, social skills deficits, and low self-esteem are often associated with developmental disabilities and represent risk factors for the development of mood disorders. No striking differences exist between the expression of mood disorders in persons functioning in the mild and moderate ranges of mental retardation and their normally developing peers. Differences may emerge among persons with severe-to-profound disability, but equivalents of mood disturbance are easily recognizable, including irritability, crying, problems in sleep or appetite regulation, agitation, mood lability, social withdrawal, and isolation. Aggression or self-injurious behavior may be seen as behavioral manifestations of dysphoria in persons regardless of developmental level.

Other Disorders The diagnosis of Tourette's disorder is difficult in persons with profound mental retardation. These individuals frequently also display stereotyped or other movements, and it is difficult to distinguish intentional from unintentional movements or sounds or vocal tics from spontaneous, stereotyped, or echolalic vocalizations in individuals frequently incapable of functional speech. The diagnosis of stereotyped movement disorder might be considered in such circumstances.

Since mental age of 4 years is required for elimination disorders, the diagnoses of functional encopresis or functional enuresis are seldom made in the context of severe intellectual disability. In some instances, individuals appear to lose previously acquired skills, (e.g., urinary continence), but such losses typically do not occur in isolation, suggesting alternate diagnoses (e.g., delirium or depression).

Somatoform disorders, depersonalization disorders, and sexual disorders are less frequently diagnosed in the context of mental retardation, though they are certainly not precluded. Sleep disorders ultimately require subjective input by the patient regarding the adequacy of rest, occurrence of nightmares, and so on. Given the frequent history of abuse reported for people with mental retardation as a group, one should not overlook the possibility of posttraumatic stress disorder when sleep disturbance is a presenting problem.

Provisional Diagnoses Individuals often do not clearly fall into a single diagnostic category. Comorbidity is common. Additionally, some individuals have psychiatric symptoms that significantly interfere with habilitative function but do not allow a clear distinction between certain diagnoses. It may be very difficult to distinguish between an impulse control disorder not otherwise specified (perhaps characterized by an individual who engages in impulsive aggressive acts) and an anxiety disorder not otherwise specified (perhaps suggested by an individual who strikes out in the context of a stressor that would go unnoticed by most people). The clinician should always make a best effort to generate working diagnosis and be prepared to modify it as indicated by data gathered through collateral sources and from increasing familiarity with a particular patient.

APPROACH TO MALADAPTIVE BEHAVIOR

As with child psychiatry in general, little specificity can be attached to a given symptom. Persons with mental retardation typically are referred for evaluation because of self-injurious, aggressive, impulsive, or hyperactive behavior. These symptoms lack diagnostic specificity and no diagnostic decision tree can be constructed. It may be more useful to ask a series of questions about the expression of a particular behavior. If the behavior is of recent onset, one is more likely to consider an acute medical or psychiatric cause. If the behavior is highly situational, occurring primarily in the context of the stress of task demands, the likelihood of a psychosis or mood disorder is probably reduced. If attempts are made to avoid the behavior by self-restraint, the inference of some ego-dystonic features may be tenable.

Assessing the sum of these and collateral data will lead the clinician to a presumptive diagnosis that will form the basis for a treatment plan.

PATHOLOGY AND LABORATORY EXAMINATION

Developmental disability is a significant risk factor for psychopathology in general, and this increased risk may derive from both biological vulnerabilities and the environment. Moreover, individuals with an I.Q. below 70 have a two- to five-fold higher rate of psychiatric disorders than normally developing persons.

A host of explanations, highlighted in Table 34-8, have been put forward to account for this added risk, including developmental experiences with which these individuals must contend such as perceived rejection from peers and frustration from parents. Repeated failure or difficulty achieving what appears to come naturally to normally developing peers, the recapitulation failure with each developmental stage, and less well developed or perhaps less-supportive peer groups must all take a toll on ego development. In addition to specific biological vulnerabilities, the common comorbidity of mental retardation with physical illness (e.g. epilepsy) may also increase the risk of mental disturbance.

Neuropathological process responsible for mental retardation may also cause or increase risk for mental illness
Increased likelihood of loss or separation, particularly in out-of-home placements
Communication deficits may predispose to emotional or behavioral disturbance
Vulnerability to exploitation or abuse by others
Inadequate coping skills
Family stress may be heightened by presence of child with developmental disability
Risk of limited network of social relationships and repertoire of social skills
Risk of reduced opportunities for development and exercise of recreational and occupational skills
Adverse effect on self-esteem of disability, possible dysmorphology

Table 34-8 Possible Contributions to Increased Vulnerability to Mental Disorders in Persons With Mental Retardation

The treatments for epilepsy and other medical conditions may carry some behavioral toxicity that can increase the likelihood of diagnosed mental illness. Phenobarbital has been widely reported to increase the risk of motoric hyperactivity and disinhibition in children and in individuals with developmental disorders, and phenytoin (Dilantin) may cause cognitive toxicity (as can essentially any of the medications used to manage epilepsy). Studies of the incidence of specific mental disorders reveal that individuals with developmental disorders may experience the same range of mental illnesses that occur in the population in general.

Medical Diagnostic Approach to the Patient With Mental Retardation

Clinical History The elements of a comprehensive clinical evaluation of a patient with mental retardation are summarized in Table 34-9. A good history is essential, and in many cases collateral informants are extremely useful. Family members may be able to fill in important aspects of the family pedigree by providing information about relatives ranging from learning problems to physical anomalies and developmental disabilities.

Clinical history
Prenatal and birth history
Family pedigree (three generations)
Relatives with learning problems, psychiatric disorders, mental retardation, neurological, or degenerative disorders
Family resemblance (reduced in aneuploidy)
Physical examination
Assessment of minor physical anomalies
Growth and physical development
Head circumference compared with norms
Growth trajectory (comparison with earlier measures)
Description of facial features (micrognathia, hypertelorism, thin upper lip); use of photographs or video, to document minor morphological variants, gait
Complete neurological examination
Documentation of behavioral phenotype
Wood's light, dermatoglyphic examinations as indicated
Adjunct diagnostics
Audiologic, ophthalmological, psychometric assessments
Diagnostic tests (for selective use as indicated)
Skeletal radiographs
Metabolic studies for lysosomal, peroxisomal, and mitochondrial disorders
Muscle biopsies
DNA molecular studies
Chromosome analysis, fluorescent in situ hybridization (FISH)
Fragile X testing
Organic and amino acid assays
Imaging studies (MRI, CT)

Table 34-9 Clinical Evaluation of the Patient With Mental Retardation

The clinical examination should give particular attention to minor anomalies and aberrant growth and physical development. Behavioral symptoms that may comport with a known behavioral phenotype and help delineate a diagnosis should be sought. In some disorders, the clinical course evolves over time, and ongoing attention to the unfolding behavioral phenotype may help establish a diagnosis. A representative list of such disorders appears in Table 34-10. Thus, a history of significant early feeding difficulties associated with hypotonia (and the tell-tale muscle biopsy scar) that later gives way to hyperphagia and weight gain is perhaps more useful for the diagnosis of Prader-Willi syndrome than the observation of hyperphagia alone.

Rett syndrome
Prader-Willi syndrome
Angelman syndrome
Kabuki syndrome
Velocardiofacial syndrome
Williams syndrome
Tuberous sclerosis
Hurler syndrome
Hunter syndrome
Neurofibromatosis I
Noonan syndrome
Smith-Magenis syndrome
Fragile X syndrome
Lesch-Nyhan syndrome
Hallervorden-Spatz syndrome
Metachromatic leukodystrophy
XL adrenoleukodystrophy

Reprinted with permission from Curry CJ, Stevenson RE, Aughton D, Byrne J, Carey JC, Cassidy S, Cunniff C, Graham JM Jr, Jones MC, Kaback MM, Moeschler J, Schaefer GB, Schwartz S, Tarleton J, Opitz J: Evaluation of mental retardation: Recommendations of a consensus conference: American College of Medical Genetics. Am J Med Genet 72:468, 1997.

Table 34-10 Selected Syndromes and Conditions in Which a Recognizable Phenotype May Evolve Over Time

Chromosome Studies Since chromosome abnormalities are the single most common known cause of mental retardation found in series of unselected patients with mental retardation, a chromosome analysis is usually obtained. The presence of multiple physical anomalies and the severity of the mental retardation will influence whether chromosomal study is obtained. It is not uncommon to be presented with a history of birth complications, perinatal hypoxia, falls with suspected injury, and so on, and later discover an underlying genetic cause for mental retardation. Moreover, in some cases the genetic syndrome may set the stage for birth complications, with the latter being credited with the cause of disability until an appropriate diagnostic procedure is done. Thus, in some settings, the mere presence of developmental delay may prompt a chromosomal study.

Given advances in diagnostic techniques, there may be situations in which multiple physical anomalies in an individual suggest aneuploidy despite normal results

from a chromosome study. In these situations the clinical geneticist may suggest multiple skin biopsies, or special karyotyping, or molecular cytogenetic studies.

Neuroimaging At the turn of the century, William Thomson observed that “the brains of most idiots and of half-witted persons are usually smaller and weigh less than the average of normal brains, while many men distinguished for their mental powers have had large and heavy brains. But the exceptions are very numerous both ways.... No man's intellect can be judged by the size of his hat.” Thomson went on to argue that the material organization of the brain determines thought, feeling, and volition. Advances in brain-imaging technology now permit a much more refined look at the relation between neuroanatomy and mental retardation.

Table 34-11 summarizes the findings from computerized tomography (CT) and magnetic resonance imaging (MRI) in groups of patients with mental retardation. Since mental retardation should suggest obligate brain abnormality, it is surprising that the percentage of individuals in imaging studies who exhibit an abnormality ranges from only 9 to 96 percent. Clearly, the selection criteria and type of study significantly affect the resultant yield.

Study	Population	Number of Subjects	Type of Study	Findings
Hunt, Miller et al. (1981)	Moderate-to-severe mental retardation (included micro- and macrocephaly and significantly dysmorphic)	23	CT	1 case atrophy, 1 case prominent preposition cistern
Lingam et al. (1982)	Nonspecific mental retardation	76	CT	20% atrophy, 8% other findings
Sugimoto et al. (1983)	Microcephaly	15	CT	80% abnormal
Curry et al. (1991)	Microcephaly	31	CT/MR	68% abnormal
	Normocephaly	46	CT, 36; MR, 10	39% abnormal
Hajnamer and Shevell (1995)	Significant delay	60	85% CT; 19% MR	16.7% cerebral dysgenesis; 10% hypoxic ischemic encephalopathy
Schaefer (1995)	Mental retardation or developmental disorder disability	681	MR	62% variety of abnormalities
Curry et al. (1996)	CA cerebral palsy project	76/192	CT, 48; MR, 14; both, 14	96% abnormal
Root and Curry (1996)	Profound mental retardation	40	CT/MR	85% abnormal
	Genetic clinic patients	32	CT/MR	28% abnormal

Table 34-11 CT and MRI Findings in Persons with Mental Retardation

In general, MRI studies are considered superior to CT because the former provides better anatomical resolution. Myelination and more subtle anatomical abnormalities such as heterotopias are more easily visualized with MRI. On the other hand, CT may be more useful when the intent is to look for possible intracranial calcification (e.g., in tuberous sclerosis complex) or abnormalities of the skull itself (e.g., craniosynostosis syndromes).

When little doubt exists that structural brain abnormalities are present, an imaging study may serve as a baseline against which to measure progression. An individual with tuberous sclerosis may be followed with serial imaging studies to chronicle the emergence of new lesions, or an individual who engages in severe head banging behavior may warrant a baseline imaging study. Neuroimaging is also indicated for patients with seizures, microcephaly, or macrocephaly; the loss of previously acquired skills; or neurological signs such as spasticity, dystonia, or altered reflexes.

When none of these conditions are met, an imaging study may not be indicated. For example, for a normocephalic patient with no localizing neurological signs, the yield of an imaging study is remote, and its interpretation is confounded by a paucity of data concerning the incidence of minor anomalies in the population in general. Similarly, for patients with known genetic syndromes but without neurological signs (e.g., Down syndrome or Prader-Willi syndrome), there is little to be gained by subjecting the patient to an imaging study outside of research interest.

Diagnostic Instruments and Rating Scales A number of diagnostic and behavioral rating scales have been developed for the population with mental retardation. While neither diagnostic nor treatment decisions should be based entirely upon the results of a given instrument, systematic data collection with such a scale is a useful adjunct to clinical diagnostics or monitoring. Among the most common general behavior rating scales are the Aberrant Behavior Checklist and the Developmental Behaviour Checklist. The Behavior Problem Inventory is particularly useful for self-injurious, aggressive, and stereotyped behaviors. The Psychopathology Inventory for Mentally Retarded Adults, the Diagnostic Assessment of the Severely Handicapped, the Reiss Screen for Maladaptive Behavior, and more recently the Reiss Scale for Children's Dual Diagnosis are all instruments developed to screen for the presence of mental disorders.

TREATMENT

Prevention Treatment strategies largely focus on preventing intellectual disability and mitigating associated complications (e.g., treating associated mental disorders). The merits of primary prevention are obvious, and the successes enjoyed with PKU should continue to provide powerful incentive for the ongoing collaborations of basic scientists and clinicians. The impact of more-recent programs is less clear. For example, although folic acid supplementation appears to reduce the risk of neural tube defects significantly, compliance with recommendations to increase dietary folate appears disturbingly negligible. It also appears that the prevalence of trisomy 21 is likely to remain unchanged or increase despite the availability of prenatal diagnostic programs.

Influence of Etiological Factors on Treatment The approach to treatment begins with diagnosis. In some cases the underlying cause of mental retardation may be particularly important in considering treatments. For example, in mental retardation associated with phenylketonuria, a number of attempts have been made to minimize or to attenuate hyperactivity and impulsivity by dietary modification. More recently, as animal models of this disorder are developed and explored, it appears that dopaminergic abnormalities may play a particularly salient role in the expression of maladaptive behaviors and perhaps in the design of specific treatments. Thus, early intervention with dopamine agonists may ultimately play a unique role in the treatment of behavioral and cognitive disturbance associated with PKU.

Abnormalities in serotonergic function have been reported in the context of Down syndrome. This hyposerotonemia led to trials of serotonin replacement with mixed results, but serotonergic drugs may yet hold particular relevance for persons with Down syndrome. In Prader-Willi syndrome, both serotonergic abnormalities and responses to serotonergic agents have been reported. Identification of the underlying cause of mental retardation has become increasingly important in considering biological treatments. Likewise, the diagnosis of mental disorders or syndromes in persons with mental retardation will guide and influence treatment strategies.

Experience over the past decade clearly shows that mental retardation is a multidisciplinary problem and optimal treatment is multimodal. Typically, a treatment plan includes attention to psychoeducational, psychotherapeutic, and psychopharmacological interventions.

Psychotherapy One should never assume that persons with mental retardation cannot benefit from psychotherapeutic intervention simply because of their impaired intellectual functioning. For example, Anton Dosen has highlighted the use of psychoanalytic approaches that focus on developmental theories to improve emotional expression, enhance self-esteem, increase personal independence, and broaden social interactions. Christian Gaedt has similarly advanced the usefulness of ego psychology (in particular, object-relation theory) in the approach to individuals with mental retardation.

In addition to psychoanalytic or developmentally based approaches, cognitive therapy may benefit the treatment of depression, and brief relaxation therapy may help reduce anxiety even in the context of moderate-to-severe mental retardation.

However, all types of individual therapies in this population benefit from certain modifications in approach. For example, an active therapeutic stance should be used with concrete, supportive interventions and careful attention to the language abilities and developmental level of the patient. When these types of alterations are made, many patients with mental retardation clearly can benefit.

Behavioral therapies are also demonstrably effective in managing many difficulties in persons with mental retardation. Typically, a behavioral assessment begins with a functional analysis of behavior, that is, a detailed examination of the variables that reinforce or maintain particular behaviors. One considers the antecedent events and consequences of a behavior in question and typically tests hypotheses to confirm the results of the behavioral analysis. Some instruments (e.g., the Functional Analysis of Stereotyped Behavior [FAST]) may be useful in beginning to identify reinforcing variables, but a behavioral psychologist can best generate and implement a behavior program based upon a functional analysis of behavior.

Group therapy can be an important part of the treatment program for persons with mental retardation, particularly in the area of social skills building. Supportive groups for parents and siblings may also be of particular benefit. Family therapy can be particularly useful and provides a setting for education, support, and

consolidation of behavioral treatment and other interventions.

Pharmacotherapy Relatively few controlled clinical drug trials include persons with mental retardation, particularly those with the most severe cognitive impairment. Thus, clinicians must generally extrapolate from pharmacotherapy in the general population for persons with mental retardation. Fortunately, there is little reason to suspect that mechanisms of action of drugs change on the basis of retardation.

However, because of drug-drug interactions that may affect the availability or effectiveness of concurrent medications, individuals with mental retardation may require different dosing strategies. In institutionalized populations, for example, where 30 to 40 percent of persons have epilepsy and as many as 70 percent may have some other significant medical condition, drug interactions become an increasingly important consideration.

Antidepressants The use of antidepressant medications in persons with mental retardation appears to remain relatively low. Special considerations in the use of antidepressant medication include the common medical comorbidities. Use of tricyclic drugs in particular must acknowledge the risk of lowering seizure threshold. This risk in the general population is on the order of 1 in 1000, and in individuals with mental retardation it may increase to nearly 1 in 5. Cardiac anomalies are common in some mental retardation syndromes, and the anticholinergic adverse effects of some medications may be particularly significant in persons with Down syndrome. Individuals with mental retardation may require lower concentrations of antidepressant drug than their normally developing peers, and disinhibition has been described with typical antidepressant doses of selective serotonin reuptake inhibitors (SSRIs).

Some individuals (particularly a subgroup within the pervasive developmental disorder spectrum) appear to exhibit extraordinary sensitivity to drugs of this and other classes. This sensitivity is manifested by disinhibition and may present as an increase in target symptoms at doses of medication that otherwise appear quite standard. These individuals may be converted to responders with significant dosage reductions, for example, reducing the starting dosage of fluoxetine (Prozac) from 10 to 1 mg daily.

A subgroup of persons with self-injurious behavior also exhibit self-restraining behaviors; for example, binding their extremities in their clothing or seeking out helmets to wear for self-protection. Such behavior might indicate an ego-dystonic quality of self-injury and suggests the possibility of self-injurious behavior as a compulsion. Obsessive-compulsive disorder occurs more commonly in the Prader-Willi syndrome than in the general population, and treatment with an SSRI may improve compulsive skin picking along with other compulsive behaviors. Given these findings, trials of SSRIs are increasingly common among patients with self-injurious behavior. Favorable results have been reported for fluoxetine, paroxetine (Paxil), sertraline (Zoloft), trazodone (Desynel), and clomipramine (Anafranil). However, of these agents, only clomipramine has been shown effective in well-controlled studies. Because it lowers seizure threshold, clomipramine is generally not a first-line treatment for compulsive self-injurious behavior in individuals frequently comorbid for epilepsy.

Anticonvulsants Data on the use of anticonvulsant medications for indications other than epilepsy are limited. However, considerable experience suggests that as in the population in general, some anticonvulsant drugs may improve cyclical mood disorders and impulsive aggression.

Self-injurious behavior and aggression can also dramatically improve when phenobarbital administration is stopped. The potential behavioral toxicity of this or any drug should not be overlooked in designing treatment strategies for persons with mental retardation. Moreover, medical comorbidity is the rule in individuals with developmental disabilities, and the importance of identifying and treating underlying medical problems (or refining that treatment) cannot be overstated.

Anxiolytics Although benzodiazepines are commonly prescribed to treat anxiety in the general population, unique concerns arise in the context of developmental disorders, particularly regarding the possibility of increased confusion, cognitive impairment, unsteadiness, and paradoxical excitement. Jennifer Barron and Curt Sandman reviewed disinhibition associated with benzodiazepines, which occurred in 35 to 68 percent of an institutionalized mentally retarded population compared with controls. Nevertheless, alprazolam (Xanax), clonazepam (Klonopin), and lorazepam (Ativan) are widely used in the treatment of acute anxiety, particularly anxiety associated with medical or surgical procedures. Their use should be considered in the absence of evidence of previous abnormal responses to these agents.

Buspirone (BuSpar) is another serotonergic agent that has been reported to benefit some persons with developmental disorders. John Ratey and colleagues reported the use of buspirone in an open trial in persons with diagnosed anxiety disorders manifested by aggressive and self-injurious behaviors. Typical dosages at which patients responded were about 15 to 45 mg a day. In an open trial by Bryan King and Pablo Davanzo, nonresponders to buspirone remained so even when the medication dosage was increased to 60 mg daily. Willem Verhoeven and Siegfried Tuinier also studied buspirone in the treatment of anxiety in persons with mental retardation and suggested that this agent should be among those considered for treating this population. Its advantages include a relatively benign side-effect profile, specifically the absence of common motor or cognitive adverse effects at dosages used to treat anxiety.

Antipsychotics Antipsychotic medications have long been used, arguably indiscriminately, in persons with mental retardation. Antipsychotics continue to be the most widely prescribed class of psychotropic medication and are even more commonly prescribed than anticonvulsant drugs for persons with mental retardation. Given this ample experience with antipsychotics in persons with mental retardation, where in residential or institutional settings as many as 50 percent of individuals may be treated with this class of drug, the adverse effects are well-known.

Individuals with mental retardation appear to be at greater risk of developing tardive dyskinesia than the general population; recorded rates range from 18 to over 30 percent. On the other hand, spontaneous abnormal involuntary movements are not uncommon in this population, which may confound interpretation of rates of neuroleptic-induced tardive dyskinesia. The growing availability of serotonin-dopamine antagonists (atypical antipsychotics) and their use in persons with mental retardation suggest that these drugs may be very helpful considering their apparently reduced risk for tardive dyskinesia and extrapyramidal symptoms as well as cognitive toxicity. Studies reporting the effectiveness of clozapine (Clozaril), risperidone, and sulpiride through open trials now exist to support the use of these agents in individuals with mental retardation.

Evidence supports the use of dopamine antagonists in self-injurious behavior and aggression, both in theory and in practice. For thioridazine (Mellaril) in particular, ample experience indicates that both self-injurious behavior and aggression may improve. Typical dosages average less than 300 mg a day, with dosages as low as 15 mg a day reported for some children. Positive reports also exist for most other neuroleptics in this population. Although suggested by some, no convincing evidence suggests that these merely suppress behavior generally through a nonspecific sedating effect. Such an outcome is clearly undesirable in individuals with preexisting cognitive impairment, and many of the earliest reports specifically note the absence of sedation. Nonetheless, enthusiasm for neuroleptic use in self-injurious behavior should be dampened by significant adverse effect liability. Interest in atypical neuroleptics for self-injurious behavior and aggression has grown in recent years. Based upon preclinical data, drugs that selectively antagonize dopamine type 1 (D_1) receptor function may have a unique potential for attenuating repetitive self-injurious behavior.

Psychostimulants Despite reports of paradoxical responses to stimulant medications in persons with mental retardation, with higher than expected rates of emergent motor tics and emotional lability, a growing body of literature supports the use of stimulant drugs for the treatment of attention-deficit/hyperactivity disorder in the context of mental retardation. Additional study is needed concerning the likelihood of a response to stimulants in relation to the severity of cognitive disability. Data are mixed on whether I.Q. negatively or positively predicts treatment response.

Opioid Antagonists Some individuals who self-injure appear to have altered pain sensitivity, as Ireland suggested nearly a century ago. This observation, coupled with data indicating that opioid antagonists can attenuate stereotypies and self-injury in animal models and data indicating that opioids may modify the function of dopaminergic systems has fueled interest in the opioids. Naltrexone (ReVia) is the opioid antagonist most widely used for self-injurious behavior, but the literature is mixed. Typical dosages range from 0.5 to 2.0 mg/kg a day in children and up to 200 mg a day in adults. The relatively long half-life of naltrexone (72 hours in brain) must be kept in mind in designing titration strategies for this drug. On balance, naltrexone appears to be well tolerated in persons with developmental disorders; sedation is the side effect most likely to be observed.

Other Drugs b-Adrenergic receptor antagonists have also been reported to be of use in the population with developmental disorders, it is not clear whether the mechanism is central or peripheral, however Ratey and colleagues explored this class of drugs on the basis of their perception that many individuals with developmental disabilities appear to have a low frustration tolerance for ambient stimuli. At dosages of propranolol (Inderal) above 1000 mg a day, the precise mechanism of action of this drug arguably becomes nonspecific, but lower dosages of this and other b-adrenergic receptor antagonists are also reportedly beneficial.

A growing appreciation exists for the potential therapeutic utility of drugs acting at glutamate receptor subtypes. Glutamatergic and dopaminergic interactions in the neostriatum are the focus of research on the pathogenesis of a host of neuropsychiatric illnesses including schizophrenia, obsessive-compulsive disorder, self-injurious behavior, and aggression. Dextromethorphan (Dimetane, Sudafed) an antitussive agent which also binds to the NMDA receptor, was reported by Leisa

Welch and Robert Sovner to have markedly attenuated self-injurious behavior in a 25-year-old individual with congenital rubella syndrome. The patient was maintained on dextromethorphan for 16 months with sustained benefit. Surprisingly (and perhaps ominously) no follow-up studies have been reported with this drug in self-injurious behavior. Lamotrigine (Lamietal) is an anticonvulsant drug that also appears to antagonize glutamate by reducing its release. A recent case report suggests that lamotrigine may be useful in reducing self-injury in the context of a stereotyped movement disorder. A larger study in which lamotrigine was added to the anticonvulsant regimen of children with mental retardation suggested that those with autism experienced gains that could not be readily attributable to better seizure control alone. However, a recent double-blind, placebo-controlled trial of lamotrigine in autism had negative results.

Services and Supports In contrast to even 20 years ago, individuals with mental retardation now enjoy a wide variety of services that can be found throughout the United States and that last throughout the lifespan; Table 34-12 lists some of these services.

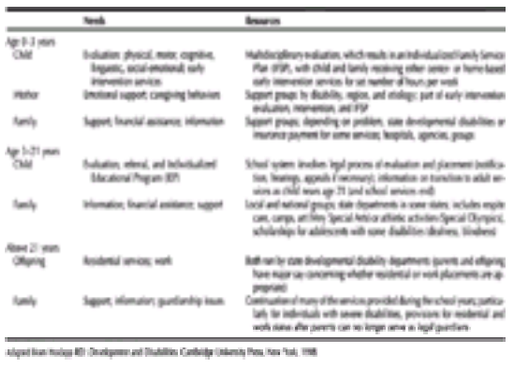

Table 34-12 Service Needs and Resources for Families of Children at Different Ages

Services

Early Intervention Early intervention serves individuals from birth until 3 years. Such services are provided by the state, with a different lead agency coordinating services in each state. These services begin with early intervention specialists visiting the child's home for a few hours a week; when the child is slightly older, center-based care is also often provided.

Since the passage of part H of PL 99-457—the Education of the Handicapped Amendments of 1986—early intervention services have increasingly emphasized the family. Agencies are required to develop an Individualized Family Service Plan (IFSP) for each family, focusing on family strengths and needs in caring for their young child with mental retardation.

School From ages 3 until 21, school takes over. These educational services became widespread following the passage of PL 94-142 (the Education for all Handicapped Children Act of 1975); this legislation has recently been extended and modified by the Individuals with Disabilities Act of 1990.

These laws require that public school systems serve all children with disabilities and institute what are essentially legal provisions to do so. This process provides notice that the child is being considered for special services; a formal, scheduled hearing at which the parents and school personnel discuss the most appropriate school placement and specialized services; the right to present a defense if parents disagree with school personnel; a written decision given to the parents notifying them of the school's decision; and the parents' right to appeal this decision, first to the school itself, then to district and state educational authorities, eventually to the courts. This entire process results in an Individualized Educational Program or Plan (IEP), which assesses the child's academic performance, provides annual goals, provides instructional objectives, describes services the child will receive, and includes starting dates of services and when and how the child's progress will be reevaluated. Although receiving services for one's child can be complicated, educational services have become a right—not a privilege—for parents of children with mental retardation.

A further element of education law involves the right to a public education “within the least restrictive environment.” This phrase has engendered many debates within the special education community about inclusion. Essentially, the debate centers on whether all children should be educated in classes composed primarily of typically developing children. Inclusion advocates consider education alongside typically developing children to be every child's right, while proponents of a continuum of care argue that the child's individual educational needs should be paramount.

Adult Services When an individual with mental retardation turns 21, service delivery comes under the auspices of the state department of developmental disabilities, but in contrast to educational services, adult services are not federally mandated. States thus vary widely in their services and how they offer them. Services are often spread out across various locations, and qualifying for and getting services can sometimes be difficult.

During the adult years, residential and vocational services predominate. Residential services run the gamut from institutional to community-based, and from constant to no supervision. More-restrictive placements involve large or small institutional residences, which mainly serve the lowest functioning, often multiply-impaired individuals. Group homes for from four to eight residents are probably the most common residential option; these homes typically exist within residential neighborhoods and provide round-the-clock staff to ensure safety and programming. Supervised living arrangements usually consist of apartments that are shared by a few persons with mental retardation; the residents are visited by program staff, who often instruct them in cooking, cleaning, and shopping. Unsupervised apartments, where individuals live by themselves or with a roommate under little supervision, mainly succeed with clients who have milder intellectual impairments and fewer behavior problems.

Vocational services also show a continuum from most to least restrictive. The most restrictive are sheltered workshops in which individuals with retardation assemble or test products next to others with disabilities. Supported employment provides a job coach or other specially trained individual to help the client find a job, learn how to take public transportation to get to work, and acquire the technical and social skills needed to handle the job. Competitive employment is the least restrictive job option for adults with mental retardation. Competitive employment involves a job that a nondisabled person might have, with the same pay levels and benefits and the same pressures and responsibilities for work performance. As in unsupervised apartment living, competitive employment operates best for the highest functioning and least behaviorally disordered individuals.

In considering working and living arrangements for any young adult with mental retardation, one should also consider transitional services. Schools offer these services to children with disabilities who have reached age 14. They teach vocational and adaptive skills that will be necessary later on. Parents need to work with schools to determine the most appropriate residential and vocational goals for each young adult.

Supports In addition to services provided by early intervention, school, and adult services, individuals with mental retardation and their families can also benefit from a variety of other supports. These range from short- or longer-term respite care (allowing families a break from full-time care), to summer camps, to sports activity programs such as Special Olympics (Fig. 34-1). In addition, several major universities have specialty clinics for behavioral and medical management of certain types of mental retardation.

FIGURE 34-1 Athletes competing in activities sponsored by Special Olympics, whose athletes take the oath: “Let me win. But if I cannot win, let me be brave in the attempt.” Special Olympics was founded in 1968 by Eunice Kennedy Shriver. The mission of the organization is to provide year-round sports training and athletic competition in a variety of Olympic-type sports for individuals with mental retardation by giving them continuing opportunities to develop physical fitness, demonstrate courage, experience joy, and participate in sharing gifts, skills, and friendship with their families, other Special Olympics athletes, and the community.

A wide variety of organized groups are also helpful to affected individuals and their families. Parent groups are available for parents of children with disabilities and for parents of children with specific types of retardation. Sibling groups and networks are also available, and the federal government fosters service and advocacy through University Affiliated Programs in almost every state. See Table 34-13 for some of the parent and professional groups that help persons with disabilities and their families.

Table 34-13 Prominent Organizations in Mental Retardation

LEGAL ISSUES IN MENTAL RETARDATION

The Americans with Disabilities Act of 1990 spells out four aims: (1) to provide a clear and comprehensive national mandate for the elimination of discrimination against individuals with disabilities; (2) to provide clear, strong, consistent enforceable standards addressing discrimination against individuals with disabilities; (3) to ensure that the federal government plays a central role in enforcing the standards established in the act on behalf of individuals with disabilities; and (4) to invoke the sweep of congressional authority, including the power to enforce the fourteenth amendment and to regulate commerce, to address the major areas of discrimination faced day-to-day by people with disabilities. This act represents a major turning point in ensuring the rights of persons with disabilities who, even 50 years ago, were frequently institutionalized and given compulsory sterilization in misguided efforts to protect society. While attitudes have shifted repeatedly over the centuries from asylum to punishment of persons with disabilities, the prevailing paradigms in developed countries today are community integration and support. Disabilities law over the past three decades reflects these values—section 504 of the Rehabilitation Act of 1973, the Individuals with Disabilities Education Act of 1990 (originally Public Law 94-142 of 1975), and the Americans with Disabilities Act Collectively, these laws ensure that all children with disabilities receive a free appropriate public education in the least restrictive environment and that reasonable accomodation is made for the employment of adults. Clinicians working with this population should be aware of their personal obligation to provide an accomodating environment but should also familiarize themselves with local guardianship, informed consent, and abuse-reporting laws to serve this population best.

SUGGESTED CROSS-REFERENCES

The neural sciences are discussed in Chapter 1, functional neuroanatomy in Section 1.2, principles of neuroimaging in Section 1.15 and Section 1.16, basic molecular neurobiology in Section 1.10. Neuropsychiatry and behavioral neurology are discussed in Chapter 2, including approach to diagnosis in Section 2.1, neuropsychiatric aspects of epilepsy in Section 2.4, brain trauma in Section 2.3, movement disorders in Section 2.6, and neuroimaging in clinical practice in Section 2.13. Contributions of the psychological sciences appear in Chapter 3, including perception and cognition in Section 3.1, Piaget's approach in Section 3.2, learning theory in Section 3.3, and biology of memory in Section 3.4. Neuropsychological and intellectual assessment is discussed in Section 9.5 and Section 9.6, medical assessment and laboratory testing in Section 9.7, and psychiatric rating scales in Section 9.8. Borderline intellectual functioning and academic problems are covered in Section 27.4. Abnormal and atypical patterns of behavior are also discussed in a Chapter 35, Chapter 36 and Chapter 37 on learning disorders, motor skills disorder, and communication disorders. Chapter 38 covers pervasive developmental disorders, and Section 44.2 provides an in-depth review of stereotypic movement disorder. Forensic psychiatry is covered in Section 49.12 and Section 54.1.

CHAPTER REFERENCES

American Association on Mental Retardation: *Mental Retardation: Definition, Classification, and Systems of Supports*. American Association on Mental Retardation Washington, DC, 1992.

*Beutler E: Gaucher disease. Arch Int Med *159:*881, 1999.

Borthwick-Duffy SA, Eyman RK: Who are the dually diagnosed? Am J Ment Retard *94:*586, 1990.

Dingman HF, Tarjan G: Mental retardation and the normal distribution curve. Am J Ment Defic *64:*991, 1960.

Down JL: Observations on an ethnic classification of idiots. London Hosp Rep *3:*259, 1866.

Dykens EM, Hodapp RM, Evans DW: Profiles and development of adaptive behavior in children with Down syndrome. Am J Ment Retard *98:*580, 1994.

Dykens EM, Hodapp RM, Walsh K, Nash LJ: Profiles, correlates and trajectories of intelligence in Prader-Willi syndrome. J Am Acad Child Adolesc Psychiatry *31:*1125, 1992.

Feinstein C, Reiss AL. Psychiatric disorders in mentally retarded children and adolescents. Child Adolesc Psychiatr Clin North Am *5:*827, 1996.

Grossman H: *Classification in Mental Retardation*, rev 3. American Association of Mental Deficiency, Washington, DC, 1983.

*Harris JC: *Developmental Neuropsychiatry*, vols 1 and 2. Oxford University Press, New York, 1995.

Haskell PH: Mental deficiency over a hundred years. Am J Psychiatry *100:*107, 1944.

*Hodapp RM: *Development and Disabilities: Mental, Motor, and Sensory Impairments*. Cambridge University Press, New York, 1998.

*Hodapp RM, Zigler E: Past, present, and future issues in the developmental approach to mental retardation. In *Manual of Developmental Psychopathology*, vol 2, D Cicchetti, D Cohen, editors, *Risk, Disorder, and Adaptation*. Wiley New York, 1995.

*Hurtado EK, Claussen AH, Scott KG: Early childhood anemia and mild or moderate mental retardation. Am J Clin Nutr *69:*115, 1999.

*Inan C, Marcus J: Sturge-Weber syndrome: Report of an unusual cutaneous distribution. Brain Dev *21:*68, 1999.

*Kaufmann WE, Reiss AL: Molecular and cellular genetics of fragile X syndrome. Am J Med Genet *88:*11, 1999.

Lambert N, Nihira K, Leland H: *AAMR Adaptive Behavior Scales—School*. Pro-ed, Austin, TX, 1993.

Luckason R, Coulter DL, Polloway EA, Reiss S, Schalock RL, Snell M, Spitalnick D, Stark J: *Mental Retardation: Definition, Classification, and System of Supports*. American Association on Mental Retardation, Washington, DC, 1992.

MacLean WE, editor: *Ellis' Handbook of Mental Deficiency, Psychological Theory, and Research*, ed 3. Erlbaum, Mahwah, NJ, 1997.

MacMillan DL, Gresham FM, Siperstein GN: Conceptual and psychometric concerns about the 1992 AAMR definition of mental retardation. Am J Ment Retard *98:*325, 1993.

Massey PS, McDermott S: State-specific rates of mental retardation—United States, 1993. MMWR *45:*61065, 1995.

Minnes PM: Family stress associated with a developmentally handicapped child. Int Rev Res Ment Retard *15:*195, 1988.

Nirje B: The normalization principle and its human management implications. In *A History of Mental Retardation. Collected Papers*, vol 1, M Rosen, GR Clark, M Kivitz, editors. University Park Press, Baltimore, 1976.

Ogbu J: Culture and intelligence. In *Encyclopedia of Intelligence*, R Sternberg, editor. Macmillan, New York, 1994.

Pober BR, Dykens EM: Williams syndrome: An overview of medical, cognitive, and behavioral features. Child Adolesc Psychiatry Clin North Am *5:*929, 1996.

Reiss S: *Handbook of Challenging Behavior: Mental Health Aspects of Mental Retardation*. IDS Publishing, Worthington, OH, 1994.

*Reiss S, Aman MG, editors: *Psychotropic Medication and Developmental Disabilities: The International Consensus Handbook*. The OSU Nisonger Center, Columbus, OH, 1998.

Roeleveld N, Zielhuis GA, Gabreels F: The prevalence of mental retardation: A critical review of the literature. Dev Med Child Neurol *39:*125, 1997.

Scheerenberger R: *A history of Mental Retardation*. Brookes, Baltimore, 1983.

Sparrow SS, Balla D, Cicchetti DV: *Vineland Adaptive Behavior Scales*. American Guidance Service, Circle Pines, MN, 1984.

*Szymanski L, King BH: *Practice Parameters for the Assessment and Treatment of Children, Adolescents, and Adults with Mental Retardation and Comorbid Mental Disorders* . American Academy of Child and Adolescent Psychiatry, Washington, DC, in press.

Szymanski LS, King BH, Goldberg B, Reid A, Tonge B, Cain N: Diagnosis of mental disorders. In *Psychotropic Medication and Developmental Disabilities. The International Consensus Handbook*, Reiss S, Aman MG, editors. OSU Nisonger Center, Columbus, OH, 1998.

Trent JW: *Inventing the Feeble Mind: A History of Mental Retardation in the United States*. University of California Press, Berkeley, 1994.

United States Department of Education: Sixteenth annual report to Congress on the implementation of the Individuals with Disabilities Education Act. United States Department of Education, Washington, DC, 1994.

Wodrich DL: *Children's Psychological Testing: A Guide for Nonpsychologists*. Brookes, Baltimore, 1997.

Zigler E: Developmental versus difference theories of mental retardation and the problem of motivation. Am J Ment Defic *73:*536, 1969.

Zigler E, Hodapp RM: *Understanding Mental Retardation*. Cambridge University Press, New York, 1986.

CHAPTER 35. LEARNING DISORDERS

35.1 READING DISORDERS

MICHAEL E. SPAGNA, FH.D., DENNIS P. CANTWELL, M.D. AND LORIAN BAKER, FH.D.

DEFINITION AND COMPARATIVE NOSOLOGY

Reading disorder is characterized by significant impairment in reading acquisition that does not have any demonstrable cause in visual, hearing, or physical disorders; mental retardation; emotional disturbance; or environmental, cultural, or economic disadvantage. Specific symptoms of reading disorder include difficulties in single-word decoding, slow oral reading, and poor comprehension of written text.

The study of children with reading disorders dates back to the late nineteenth century as recorded in case studies collected primarily by physicians. The terms “pure word blindness” and “congenital word blindness” were used to discriminate between adults and children, respectively, who were diagnosed as experiencing difficulty reading words in spite of normal visual acuity and adequate exposure to formal education. These reading disorders were recognized as heterogeneous; some individuals experienced difficulty labeling letters, some could only read letters and single syllables, some manifested word reversals in reading, and some had obtained an adequate sight-word vocabulary but could not read uncommon monosyllabic words. Als at this time several investigators postulated that genetic links might exist within families in which at least one member was diagnosed as having a reading disorder. Attempts were made in the early 1900s to refine the conceptualizations of word blindness, and one researcher proposed the *developmental alexia*, defined as a developmental difficulty in recognizing printed symbols. This term received wide acceptance in professional circles and was eventually shortened in the 1960s to *dyslexia*.

By the early 1900s, a significant knowledge base was established. Case study research suggested that reading disorders (1) were heterogeneous, affecting different areas of reading ability; (2) existed in children, youth, and adults with otherwise intact cognitive functioning; (3) occurred in more males than females; (4) appeared to have a familial component; and (5) were for the most part resistant to traditional educational intervention programs.

Despite over a century of research, however, the definition of a reading disorder remains controversial. A review of various definitions yields five common elements, each of which has generated considerable controversy in the field. The first common element is that the underlying cause for reading disorder is central nervous system dysfunction. Although, neurological conditions often cannot be determined in individuals via external tests and medical examinations, it is presumed that dysfunction can be inferred from observation of behaviors.

The second common element is an uneven pattern of cognitive functioning. In other words, although overall cognitive functioning is intact, specific areas (what one researcher calls informationally encapsulated “vertical modules”) are significantly degraded—in some instances to such an extent that the acquisition of basic decoding skills is extremely difficult. This uneven pattern of cognitive functioning has been referred to in the literature as the presence of a “psychological processing deficit” in spite of overall adequate cognitive functioning.

Most reading disorder definitions contain the third common element, difficulty in single-word decoding. All 50 states have adopted definitions of “specific learning disabilities” that identify decoding skills as an area of academic achievement affected by reading disorder.

The fourth element common to most definitions of reading disorder is a discrepancy between learning potential and actual reading achievement. This element has been the target of considerable recent debate. It assumes that individuals are underachievers in the area of reading achievement, failing to exhibit decoding or comprehension skills or both commensurate with their ability level.

The final common definitional element is the exclusion of other causes of reading difficulty. Similar to the first, second, and fourth elements, considerable disagreement exists about whether or not reading disorders can exist in combination with other disorders. However, 48 of the 50 states have adopted this element in their definitions of specific learning disabilities. Inclusion of this fifth element reflects the need in the mid-1970s to establish specific learning disabilities and thereby reading disorders as a discrete and separate entity for federal legislation.

Lacking an agreed-upon definition, the terms most commonly used in the educational and psychiatric fields today include “developmental reading disorder,” used in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and the revised third edition of DSM (DSM-III-R), “reading disorder,” used in the fourth edition of DSM (DSM-IV), and “specific reading disorder,” used in the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). The conceptualization of reading disorder has changed little across DSM editions. However, the definition in ICD-10 differs somewhat in requiring a history of impairment in speech, language, sound categorization, motor coordination, visual processing, or control of attention or activity.

Recently domain-specific definitions of learning disorders have been developed. One professional organization has offered the following definition of dyslexia:

> Dyslexia is one of several distinct learning disabilities. It is a specific language-based disorder of constitutional origin characterized by difficulties in single word decoding, usually reflecting insufficient phonological processing. These difficulties in single word decoding are often unexpected in relation to age and other cognitive and academic abilities; they are not the result of generalized developmental disability or sensory impairment. Dyslexia is manifest by variable difficulty with different forms of language, often including, in addition to problems with reading, a conspicuous problem with acquiring proficiency in writing and spelling.

Although this definition contains several of the aforementioned common definitional elements, it poses two important advances. First, instead of defining reading disorders generically, it focuses on one type of reading disorder, dyslexia. Second, this definition of dyslexia, for the first time, localizes the difficulty associated with dyslexia at the single-word level and pinpoints the cause as insufficient phonological processing. Although this definition has not gained wide acceptance, it represents a significant first step in addressing some of the previous confusion and disagreement surrounding definitions of learning disabilities.

EPIDEMIOLOGY

Few studies have examined the incidence of reading disorder, and those that have used different definitions of the disorder. More incidence studies have focused on learning disorder without specifically differentiating the combinations of reading, spelling, mathematics, and other problems involved. Generally, reading disorder is estimated to affect 80 percent of children identified as having learning disabilities; this represents approximately 4 percent of the school-age population. Similar incidence estimates have been reported worldwide.

Clinical and school-based studies have typically reported that three to four males are affected with reading disorder for every female. However, these data may reflect a referral bias in which symptoms in females are overlooked. One recent epidemiological study found no significant differences between the sexes in rates of reading disorder among research-identified children. However, the same study found that school-referred children with reading disorder were predominantly male. Boys with reading disorder appear to be more noticeable in the educational setting than girls with reading disorder because of associated behavioral and activity symptoms and disorders.

ETIOLOGY

The etiology of reading disorder is not known. The traditional view is that the disorder has multiple causes, most likely involving some biologically based dysfunctions that result in a lag or impairment in the development of cognitive skills needed for learning to read. Numerous hypotheses exist regarding the types and sources of biological factors and cognitive impairments that may be involved.

Some of the conflicting results and part of the overall confusion exist because many studies failed to distinguish reading disorder from general learning disability or low academic achievement or to distinguish among different subtypes of reading disorder. The co-occurrence of reading disorder with language disorders, attention-deficit/hyperactivity disorder, and other learning disorders is also of concern. The often-associated disorders may reflect distinct subtypes of reading disorder, each with a different cause.

Theories of Brain Function Many of the theories of brain function in children with reading disorder are based on observations of associated symptoms. For example, because language and speech problems are often associated, abnormalities in the left hemisphere and in the frontal speech regions in both hemispheres have been postulated. Similarly, because motor problems with balance and equilibrium are often associated, abnormalities in cerebellar-vestibular function have been postulated. Thus, brain anomalies are inferred but not demonstrated. Unfortunately, the type of neuroanatomical studies needed to demonstrate brain anomalies (e.g., computed tomography, electrophysiological studies, or postmortem studies) are very limited and have produced contradictory results.

The notion that reading disorder results from a delay or abnormality in lateralization or cerebral dominance is popular. It was first proposed by Samuel Orton in the 1930s as a cause of two common symptoms of reading disorder—poorly established or unstable handedness and the tendency to reverse letters in reading and writing.

Lateralization Although controversial, some evidence suggests abnormal or reversed lateralization or cerebral asymmetry in some children with reading disorder. The data come from handedness reports, dichotic listening studies, an evoked-potential study, a cerebral axial tomography study, and one postmortem study. Much of the data, particularly that on handedness and dichotic listening, can be explained by alternative hypotheses; for example, task-dependent attentional dysfunction interferes with left hemispheric language processes by overengaging either hemisphere.

The postmortem data have led to an intriguing hypothesis about dominance that also considers clinical and familial data. In postmortem examinations of the brains of five males with a familial reading disorder, Albert Galaburda, Norman Geschwind, and their colleagues found a consistent lack of hemispheric asymmetry along with some developmental cortical abnormalities and some alterations in subcortical structures. The affected areas of the left hemisphere were those whose fetal development is regulated by testosterone. Consequently, it was hypothesized that abnormal testosterone levels in fetal blood during critical periods of development, possibly genetically controlled, could be responsible for producing the brain abnormalities and the reading disorder syndrome.

Neuroanatomical Data Some neuroanatomical data (from regional cerebral blood flow studies and positron emission tomography studies) indicate bihemispheric activation during reading. Those findings are relevant to the theories that suggest that different subtypes of disability are associated with different dysfunctional regions in both hemispheres of the brain. For example, it has been hypothesized that disorders involving basic phonological processing may originate in the left temporal lobes, disorders involving phoneme-grapheme matching may originate in the parietotemporal region of the left hemisphere, and disorders involving difficulties with outputs in all modalities may originate in the prefrontal regions of both hemispheres.

A study that used sophisticated computer graphics technologies to map brain electrical activity delineated specific areas of the brain that, in combination, could constitute a physiological signature of dyslexia. In addition to aberrant brain function in the left posterior language region, differences in medial frontal lobe function were also found.

Genetic Factors Perinatal and postnatal factors associated with reading disorder include prematurity, low birth weight, toxemia of pregnancy, hyperbilirubinemia, recurrent otitis, meningitis, encephalitis, and anemia. Some studies of groups of children with reading and learning problems have reported that as many as 80 percent had at least one perinatal factor. However, several longitudinal studies, including the Kauai study and the Collaborative Perinatal Project, found that environmental factors (e.g., family size and socioeconomic status) were stronger predictors of learning problems than biomedical factors.

In the 1950s a landmark study of adults with reading disorder found that almost half of their offspring also had reading disorder. Subsequent family studies have all found high rates of reading disorder in the families, and especially in the first-degree relatives of persons with reading disorder. However, family studies have been unable to establish a specific mode of genetic transmission.

Twin studies have consistently found higher concordance rates for dyslexia in monozygotic twin pairs than in dizygotic pairs. An elevated concordance rate in monozygotic twins is generally accepted as proof of a genetic component. However, one investigator has argued that, owing to more cognitive or emotional similarity, the monozygotic twins could have created more similar environments for themselves, which in turn could have affected their reading acquisition. Nonetheless, the various twin studies that have looked at the heritability of reading disorder have attributed 30 to 60 percent of the similarity between monozygotic twins to genetic factors and the balance to environmental factors. The observation that some monozygotic twin pairs are discordant for reading disorder proves that genetics is not the only cause.

Most studies that have used segregation analyses to examine family pedigrees have failed to find any single dominant or recessive gene model to explain the data. The heterogeneity of reading disorder may be responsible for these negative findings, and once valid subtypes are identified, genetic research may be more fruitful.

In fact, a recent review pointed out that when separate subtypes of reading disorder are considered, patterns of possible genetic influence do become apparent. Limited evidence of inheritance exists for reading disorder with associated visual-spatial deficits, and a fairly strong literature suggests a genetic influence for reading disorder with associated verbal or language deficiencies. One study that looked at that type of reading disorder isolated a gene on chromosome 15 that may be responsible for transmission of the disorder, which seemed to follow an autosomal dominant pattern.

Cognitive Factors Children with reading disorder perform poorly on so many tasks that there is almost no limit to the hypothesized underlying cognitive difficulties. Difficulties have been reported with tasks spanning visual, motor, and auditory modalities and including such abilities as discrimination, integration, attention, and memory. In general, cognitive theories of reading disorder tend to fall into three major groups: models of visual-spatial-motor deficits; models of verbal, auditory, or linguistic deficits; and models of cross-modality integration deficits. As with all the research on reading disorder, definitions and subclassifications are problematic. In addition, many proposed cognitive-psychological deficiencies have problems of construct validity. It is unclear, for example, whether some of the constructs represent real ability differences or are simply measurement artifacts. Similarly, some of the constructs may not actually underlie or even be related to reading ability.

Perceptual Factors The notion that visual-perceptual, visual-spatial, or visual-motor deficits underlie reading disorder dates back to Orton's work in the 1930s, which focused on two types of letter reversals, kinetic or sequencing reversals (e.g., reading *was* for *saw* or *on* for *no*) and static or letter-form orientation reversals (e.g., confusing *b* and *d*). Such reversals led to the hypotheses that children with reading disorder experienced spatial disorientation or left-right discrimination difficulties. Recent attention to the reversals made by children with reading disorder has shown that these errors constitute a small fraction of the reading errors made. Some children with reading disorder do have deficits in visual perception. However, they have not been shown to be major deficits for the children, and no direct correlations have been found between these deficits and reading levels. Further, much of the evidence for a visual perceptual deficit concerns aspects of visual processing that are peculiar to written language. For example, a group of poor readers had difficulties matching, recognizing, and recalling visual stimuli when the stimuli were printed nonsense words. However, they did not have difficulties when the visual stimuli were meaningless shapes. Similarly, poor readers demonstrated impaired visual recall for pictures of objects that could be named and for printed nonsense syllables but not for nonsense doodles or faces of strangers.

The ability to integrate different modalities, known variously in the literature as intersensory integration, cross-modal perception, or intermodal transfer, has been suggested as another underlying dysfunction in children with reading disorder. There is very little support for this hypothesis.

Language The most compelling evidence for deficits underlying reading disorder is evidence of deficits in auditory, verbal, or linguistic functioning. An extremely large literature, including both experimental and clinical studies, demonstrates that children with difficulties in reading also have pervasive and enduring language problems. A recent screening of more than 200 children in learning-disorders classes, using standardized testing data, found that 90 percent had a language deficit and 96 percent had a language, speech, or hearing problem. This is not entirely surprising because factor analyses of neuropsychological testing results in children

with reading disorder have shown that the largest subgroups have disturbances in language functioning.

Virtually all areas of language development are deficient in at least some groups of children with reading disorder. Studies using formal testing procedures have revealed deficits in phonology or sound system processing, grammar or syntax, vocabulary or semantics, and pragmatics or conversational strategies. Similarly, analyses of the spontaneous speech and interactions of children with reading disorder have found linguistic problems, including limited verbalizations, nonfluency of speech, limited vocabulary, word-finding difficulties, poor comprehension of word derivations, lack of grammatical rules for pronoun use, and inability to abstract meanings from interactions.

It has been suggested that some of the language deficiencies found in poor readers (e.g., lags in higher-level semantic and grammatic skills) may actually be a result of limited exposure to certain vocabulary and grammatical forms. However, language deficits occur in very young children with reading disorder, before the lack of reading exposure could exert a large impact. Moreover, most follow-up studies have found that children with early speech or language problems have reading or other learning disorders once they reach school age.

Furthermore, it is difficult to see how weak phonological skills could result from reading disorder rather than vice versa. In fact, several studies that looked at various aspects of language development in children with reading disorder find no deficits in syntax and semantics, although deficits in phonological development were found.

Phonological processing deficits are deficits in skills such as segmenting sound sequences and forming and accessing representations of the sound sequences in words—skills that assist in the storage and retrieval of linguistic information. Thus, an underlying phonological processing deficit could explain the difficulties that children with reading disorder have with perceiving speech through noise, with repeating sentences and sequences of words, with comprehending certain spoken sentences, and with remembering linguistic material. The fact that children with reading disorder seem able to manage print when they do not have to translate print to speech is further evidence that their word-recognition difficulties may be related to underlying phonological processing deficits.

DIAGNOSIS AND CLINICAL FEATURES

The essential feature of reading disorder is a significant impairment in the development of reading ability that is not due to low intelligence, a visual or hearing deficit, a neurological disorder, or lack of opportunity for instruction. Although the severity of impairment and the type of associated symptoms vary widely, all children with the disorder exhibit three key symptoms: inaccurate reading, slow reading, and poor reading comprehension. The DSM-IV diagnostic criteria for reading disorder are given in Table 35.1-1. The ICD-10 diagnostic criteria for specific developmental disorders of scholastic skills, including specific reading disorder, are listed in Table 35.1-2.

A. Reading achievement, as measured by individually administered standardized tests of reading accuracy or comprehension, is substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education.

B. The disturbance in criterion A significantly interferes with academic achievement or activities of daily living that require reading skills.

C. If a sensory deficit is present, the reading difficulties are in excess of those usually associated with it.

Coding note: if a general medical (e.g., neurological) condition or sensory deficit is present, code the condition on Axis III.

Table 35.1-1 DSM-IV Diagnostic Criteria for Reading Disorder

Table 35.1-2 ICD-10 Diagnostic Criteria for Specific Developmental Disorders of Scholastic Skills

Word recognition (ability to read single words aloud accurately) is poor. Children with reading disorder demonstrate word recognition problems (compared with normal readers of the same chronological age and with normal readers at the same reading level). Nevertheless, a recent study indicates that the general approaches to, and patterns of, word recognition are the same for normal readers and children with reading disorder. Thus, for example, both normal readers and those with reading disorder rely primarily on phonological information for word recognition, both use contextual information to assist in word recognition, and both find recognition of low-frequency words more difficult.

The misreadings of a child with reading disorder may be distortions, substitutions, or omissions of words or morphemes (meaningful parts of words). All types of words can be misread, both important content words (usually nouns or verbs) and the less important function words (e.g., prepositions or articles). Substitutions often, but not always, involve words belonging to the same grammatical category (e.g., nouns are usually substituted for nouns). Substitutions may involve phonetically or graphemically similar words (*read* for *red*), semantically similar words (*blue* for *violet*), or unrelated words. Errors occur just as often in reading single words as in reading paragraphs or stories.

Both oral and silent reading are characterized by slowness and comprehension difficulties. Reading is slower with unfamiliar words. Part of the slowness results from the child's need to go back and revise or recheck previous words that were guessed at or read inaccurately, and part is likely due to the word recognition deficit. An artifact of this rechecking of previously read words can be observed in “kinetic reversals”; for example, reading *was* as *saw*.

A recent study examined the reading comprehension performance of reading-disordered and normal-reading college students on timed and untimed tests. Poor readers did not differ significantly from normal readers in the untimed condition but differed significantly in the timed condition. The authors of the study interpreted this as suggesting that problems with attention span and concentration among the poor readers necessitated rereading passages.

The associated impairments and disorders found with reading disorder vary more than the specific symptoms of the disorder. The disorders most often associated with reading disorder are mathematics disorder, disorder of written expression, specific expressive language disorder, and mixed receptive-expressive language disorder. The vast majority of children with reading disorder, perhaps as many as 95 percent, have at least one of those disorders concurrently. In addition, various cognitive processing difficulties in perception, language, attention, and memory as well as motor coordination difficulties are frequently associated with reading disorder.

Another important associated symptom is problems with social skills. Many children with reading disorder experience impaired social skills. It is not clear whether

those impairments constitute a separate underlying deficit or are among the social consequences of having a learning difficulty.

The psychiatric diagnoses reported most frequently in children with reading disorder are attention-deficit/hyperactivity disorder, conduct disorder, oppositional defiant disorder, and depressive disorder. A large literature exists on the association between reading disorder and attention-deficit/hyperactivity disorder, conduct disorder, and juvenile delinquency. It is not clear if reading disorder predisposes to the development of those psychiatric disorders, if the psychiatric disorders lead to underachievement and thence to reading disorder, or if a common risk factor leads to the simultaneous development of both.

An important issue in the learning disabilities field is subtyping. Some of the different associated symptoms and disorders may be markers for distinct subtypes of reading disorder, each with a different cause and prognosis. A number of different subtypes of reading disorder have been postulated (e.g., subtypes with specific types of intelligence quotient [I.Q.] profiles on the Wechsler Intelligence Scale for Children, subtypes with particular patterns of reading or spelling performance, and subtypes with or without motor impairment, language disorder, or attentional deficits).

Most research has concentrated on defining and describing subtypes; attempts to validate the subtypes with follow-up data, treatment data, and laboratory measures are more limited. Factor analysis studies of children with reading disorder has identified two major subtypes of reading disorder: a verbal-linguistic subtype and a nonverbal subtype. Children with the two subtypes of reading disorder can be distinguished by their performances on batteries of neuropsychological and educational achievement tests. In addition, some evidence suggests that they may have different associated behavioral or emotional problems and different prognoses.

N. was a 13-year-old girl who was referred for evaluation because of poor academic performance noted by several of her middle school teachers. An initial conference conducted by the middle school's student study team revealed that although no complications were reported during the pregnancy, the mother indicated that N. had a history of frequent ear infections in childhood. Those infections eventually led to myringotomes performed in both ears at the ages of 3 and 4 years. N. attended a nonacademic preschool and kindergarten. Her teachers described her as quiet, with no behavioral problems in the classroom. She attended a regular public school for first through fourth grades and was initially referred for evaluation of a reading problem in fifth grade. Testing revealed a normal pattern of cognitive functioning, language development, and adaptive behaviors; academic achievement testing revealed age-level performance in mathematics computation and reasoning, and spelling; however, although her sight-word reading skills were adequate, word attack and reading comprehension performance was below average. It was recommended at the time that N. receive after-school tutoring with an educational therapist to work on word attack skills and reading comprehension strategies. In spite of this intervention, N.'s academic performance did not improve markedly, and a decision was made to reconsider her case for possible referral for special education and related services.

As a result of student study team deliberations, a multidisciplinary team conducted a comprehensive assessment with the following results: intellectual testing revealed a nonverbal I.Q. in the high normal range and a verbal I.Q. in the normal range. However, closer inspection revealed that N.'s performance on the Digit Span, Information, Arithmetic, and Coding subtests was significantly below average. Achievement testing revealed slightly below average performance in mathematics and written language but significantly below average performance on measures of phonics, sight-word, and reading comprehension skills. Findings obtained from outside testing conducted at a university hospital revealed that N. has an attention-deficit disorder. This information combined with previously reported findings led the school-based Individualized Education Program committee to classify N. as having a specific learning disability (dyslexia) occurring along with an attention-deficit disorder. In an individualized education program the committee recommended that the family establish contact with the university hospital for possible treatment of the attention-deficit disorder. The committee also recommended intensive remedial reading instruction for one period a day through a resource specialist program.

Over the next 6 months N.'s reading performance improved dramatically. When she was retested at the end of the year, she had made significant progress in acquiring decoding skills and a decision was made to continue medical treatment for attention-deficit disorder and discontinue placement in the resource specialist program. The committee recommended bimonthly monitoring N.'s progress in the eighth grade general education program.

DIFFERENTIAL DIAGNOSIS

The essential feature of reading disorder is a significant impairment in development of reading accuracy, speed, or comprehension that is not due to low intelligence, visual or hearing deficit, neurological disorder, or lack of opportunity for instruction. However, distinguishing children with reading disorder from normal children with poor reading requires defining a *significant impairment*. This issue has generated extensive debate in the learning disabilities field.

Various methods and formulas have been applied to the problem. The approaches include identifying discrepancies between mental age and reading age scores, between reading age scores and chronological age, between reading age scores and standard scores, and between reading age scores and expected reading scores.

Expected reading scores are calculated from a regression equation based on population achievement, age, and intelligence. The various approaches identify different children with reading disorder. Although the regression equation method appears to be the most consistent across the I.Q. range, it is the most difficult method for clinicians to use. DSM-IV permits clinicians to compare a child's reading test performance with age and I.Q. norms. Generally, a discrepancy of 2 standard deviations is required for the diagnosis, but some flexibility is permitted in the DSM system. School districts may set more rigorous standards.

DSM-IV also permits concurrent diagnoses of reading disorder and mental retardation. The diagnosis of reading disorder is made in exactly the same way for a child with mental retardation as for a child with normal intelligence. Mentally retarded children who read poorly, but no more so than would be expected from their other cognitive functioning, are not given the concurrent diagnosis of reading disorder.

Inadequate schooling and impaired vision or hearing are other possibilities to consider in differential diagnosis. Inadequate schooling can usually be determined from the history and interviews; frequent school changes, frequent absences, or classmates with similar impairments in reading are clues. Impaired vision or hearing can be ruled out through screening tests.

COURSE AND PROGNOSIS

Although symptoms of reading disorder may appear as early as age 5 years (e.g., inability to distinguish among common letters or to associate phonemes with letter symbols), referral and formal diagnosis may not occur until later. Many children are first diagnosed when they fail to respond to formal instruction in the second grade. Some, especially those with high intelligence, may not be diagnosed until the fourth or fifth grade.

Over time, reading disorder tends to improve, with or without treatment. However, even with optimal remediation, short-term improvement is slow, and long-term improvement may not be complete. Complications include conduct or oppositional defiant disorder, low self-esteem and demoralization, and school dropout.

Adults with a childhood diagnosis of reading disorder show a wide range in reading skills. Typically, some residual difficulties exist, but complete inability to read is rare. Despite residual reading problems, adult educational and occupational attainments are typically those of the general population. Predictors of better outcome include less-severe impairment, higher intelligence and socioeconomic level, good psychiatric status, and earlier age at diagnosis and treatment.

TREATMENT

Attempted therapies for reading disorder include educational remediation, medications, and psychosocial interventions. The most effective therapy appears to be an educational approach that combines “meaning-emphasis” and “code-emphasis” programs. In a review of research studying beginning reading instruction, one educator found that code-emphasis programs that emphasize systematic instruction in phonics are most effective in the early grades, especially for students with reading disorders. Once automaticity is achieved in decoding, meaning-emphasis programs that stress the use of common sight words are recommended to increase fluency and develop comprehension.

In certain cases, specific remedial reading programs and approaches have proved particularly effective in educating students otherwise resistant to such a combination approach. Examples of these remedial programs and approaches are as follows: Reading Mastery, an intensive, highly structured program designed to be implemented up through third grade; the Corrective Reading Program, an advanced reading program, also highly structured and designed for students in grades four through twelve; the Fernald Method, a multisensory approach that teaches words as whole units using language experience and tracing (kinesthetic) techniques

to teach vocabulary selected from student-dictated stories; and the Gillingham Method, another highly structured approach that emphasizes consonants and vowels with only one sound as initially presented on drill cards, followed by learning of consonant-vowel-consonant words through blending exercises, and spelling and writing skills as a means of practicing words. Approaches have also been designed to increase fluency (e.g., Neurological Impress method; Repeated Reading method) and assist older students with comprehension skills (e.g., High-Interest Low-Vocabulary method).

Other instructional approaches that have proved highly successful with children with reading problems are Reciprocal Teaching (an interactive approach that promotes comprehension by engaging teachers and students in an active dialogue about the text), Story-Mapping (an organizational map used by readers to keep track of story components such as setting, problem, goal, action, and outcome), and Multi-Pass (an approach that explicitly trains students to make several structured passes through content-area texts, picking out important textual cues such as italicized words and captions to improve comprehension).

Various types of cognitive-perceptual skills training (e.g., sensory integration training, perceptual-motor training, occupational therapy, auditory memory training, vestibular stimulation, hemispheric stimulation, optometric training) have been used with reading-disabled children. The effectiveness of these methods in improving reading disorder in most children has not been convincingly demonstrated.

Medical approaches to treating reading disorder include stimulant medications, antianxiety medications, motion sickness stimulants, and piracetam. Both methylphenidate (Ritalin) and piracetam have been studied in double-blind trials. Although methylphenidate does not appear to be specifically beneficial to reading skills, there is some indication that piracetam may facilitate reading performance.

Psychosocial approaches to reading disorders include supportive psychotherapy, parent guidance and training, social skills training, relaxation training, and behavioral modification approaches. The effectiveness of such approaches for reading disorder has not been proved, although they are likely to be helpful for associated behavioral and emotional problems. Similarly, the benefits of speech or language therapy for reading are not known, although it may prove helpful for associated speech and language deficits.

An overall treatment program should include attention to associated disorders such as language disorders or psychiatric disorders. In addition, individual characteristics of the child—motivation, learning style, attentional abilities, and responses to instructional strategies or to curriculum materials may be significant to treatment selection and outcome.

SUGGESTED CROSS-REFERENCES

Commonly associated developmental or learning disorders are discussed in Section 35.2 on mathematics disorder, Section 35.3 on disorder of written expression, Section 37.1 on expressive language disorder, Section 37.2 on mixed receptive-expressive language disorder, Chapter 38 on developmental coordination disorder, and Chapter 34 on mental retardation. Commonly associated psychiatric disorders are discussed in Chapter 39 on attention-deficit disorders, Chapter 40 on disruptive behavior disorders, and Chapter 45 on depressive disorders.

SECTION REFERENCES

Ackerman PT, Dykman RA: Phonological processes, confrontational naming, and immediate memory in dyslexia. J Learn Disabil *26:*9, 1993.

Ackerman PT, McPherson WB, Oglesby DM, Dykman RA: EEG power spectra of adolescent poor readers. J Learn Disabil *31:*83, 1998.

*Baker L, Cantwell DP: Specific language and learning disorders. In *Handbook of Child Psychopathology*, ed 2, TH Ollendick, M Herson, editors. Plenum, New York, 1989.

Bryan T: The social competence of students with learning disabilities over time: A response to Vaughn and Hogan. J Learn Disabil *27:*5, 1994.

Carr SC, Thompson B: The effects of prior knowledge and schema activation strategies on the inferential reading comprehension of children with and without learning disabilities. Learn Disabil Q *19:*48, 1996.

*Castles A, Datta H, Gayan J, Olson RK: Varieties of developmental reading disorder: Genetic and environmental influences. J Exp Child Psychol *72:*73, 1999.

Das JP, Mishra RK, Kirby JR: Cognitive patterns of children with dyslexia: A comparison between groups with high and average nonverbal intelligence. J Learn Disabil *27:*4, 1994.

Dixon ME, Rossi JC: Directors of their own learning: A reading strategy for students with learning disabilities. Teach Except Child *27:*10, 1995.

Duffy FH, Geschwind N, editors: *Dyslexia: A Neuroscientific Approach to Clinical Evaluation*. Little, Brown, Boston, 1985.

Felton RH: Effects of instruction on the decoding skills of children with phonological-processing problems. J Learn Disabil *26:*9, 1993.

Flowers DL: Brain basis for dyslexia: A summary of work in progress. J Learn Disabil *26:*9, 1993.

Geschwind N, Galaburda A: *Cerebral Lateralization*. MIT Press, Cambridge, MA, 1987.

Hammill DD: On defining learning disabilities: An emerging consensus. J Learn Disabil *23:*74, 1990.

Hurford DP, Schauf JD, Bunce L, Blaich T, Moore K: Early identification of children at risk for reading disabilities. J Learn Disabil *27:*6, 1994.

*Hynd GW, Clikeman-Semrud M: Dyslexia and neurodevelopmental pathology: Relationships to cognition, intelligence, and reading skill acquisition. J Learn Disabil *22:*204, 1989.

Jimenez Glez JE, Lopez MR: Is it true that the differences in reading performance between students with and without LD cannot be explained by IQ? J Learn Disabil *27:*3, 1994.

Kavale KA, Forness SR: Learning disability and the history of science: Paradigm or paradox. Remedial Spec Educ *6:*12, 1985.

Keogh BK, Wiesner T: An ecocultural perspective on risk and protective factors in children's development: Implications for learning disabilities. Learn Disabil Res Pract *8:*1, 1993.

Korhonen TT: The persistence of rapid naming problems in children with reading disabilities: A nine-year follow-up. J Learn Disabil *28:*232, 1995.

Korkman M, Pesonen AE: A comparison of neuropsychological test profiles of children with attention deficit-hyperactivity disorder and/or learning disorder. J Learn Disabil *27:*6, 1994.

*Lyon GR: Toward a definition of dyslexia. Ann Dyslexia *45:*3, 1995.

Mauer DM, Kamhi AG: Factors that influence phoneme-grapheme correspondence. J Learn Disabil *29:*259, 1996.

McKinney JD, Osborne SS, Schulte AC: Academic consequences of learning disability: Longitudinal prediction of outcomes at 11 years of age. Learn Disabil Res Pract *8:*1, 1993.

Orton ST: *Reading, Writing, and Speech Problems in Children*. Norton, New York, 1937.

Riley RW: Improving the reading and writing skills of America's students. Learn Disabil Q *19:*67, 1996.

Share DL, Silva PA: Language deficits and specific reading retardation: Cause or effect. Br J Disord Commun *22:*219, 1988.

*Snowling M: *Dyslexia: A Cognitive Developmental Perspective*. Blackwell Scientific, Oxford, 1987.

*Swank LK: Specific developmental disorders. The language-learning continuum. Child Adolesc Psychiatr Clin North Am *8:*89, 1999.

Swanson HL, Berninger V: The role of working memory in skilled and less skilled readers' comprehension. Intelligence *21:*83, 1995.

Swanson PN, De La Paz S: Teaching effective comprehension strategies to students with learning and reading disabilities. Interv School Clin *33:*209, 1998.

Torgesen JK, Barker TA: Computers as aids in the prevention and remediation of reading disabilities. Learn Disabil Q *18:*76, 1995.

Torgesen JK, Wagner RK, Rashotte CA: Longitudinal studies of phonological processing and reading. J Learn Disabil *27:*5, 1994.

Vaughn S, Hogan A: The social competence of students with learning disabilities over time: A within-individual examination. J Learn Disabil *27:*5, 1994.

*Vellutino FR: Recent advances in the study of developmental dyslexia. Sci Am *256:*34, 1987.

35.2 MATHEMATICS DISORDER

MICHAEL E. SPAGNA, FH.D., DENNIS P. CANTWELL, M.D. AND LORIAN BAKER, FH.D.

DEFINITION

Mathematics disorder is an impairment in the development of arithmetical or mathematical skills that is sufficiently serious to interfere with academic achievement or daily living. This impairment in mathematical skills cannot be explained by the person's measured intelligence; educational background; visual, hearing, or physical disorders; mental retardation; emotional disturbance; or environmental, cultural, or economic disadvantage.

HISTORY AND COMPARATIVE NOSOLOGY

Compared with reading disorder, the study of learning disorders in arithmetic have been largely ignored, with empirical investigations of mathematics disorder few and narrowly focused. Early study of mathematics disorder focused on acquired calculation disturbances (acalculia) in adults that resulted from cerebral trauma. From this work, three broad categorizations of acalculia were accepted in the professional community: (1) number alexia or graphia (also termed *aphasic acalculia*), difficulties related to reading and writing numbers correctly or handling numbers as words; (2) *visual-spatial acalculia*, difficulties associated with the inability to align problems or sustain place-holding values; and (3) *anarithmetia*, a primary inability to perform calculations with errors particularly noted in carrying out arithmetic operations. A developmental disorder specific for mathematical skills was first described in 1937. Various terms have been used to describe the disorder, including *dyscalculia, Gerstmann syndrome, developmental acalculia, developmental dyscalculia*, and *developmental arithmetic disorder*.

The disorder appeared in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), the revised third edition of DSM (DSM-III-R) (under the name “developmental arithmetic disorder”), and the fourth edition of DSM (DSM-IV), with only minor changes in the wording of the diagnostic criteria. Specific disorder of arithmetical skills is included in the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) with largely similar diagnostic criteria (see Table 35.1-2). A key difference between the DSM definition and the ICD definition is that the ICD definition specifically excludes reading and spelling difficulties and disorders.

EPIDEMIOLOGY

Although the prevalence of mathematics disorder has not been studied in detail, several studies agree that approximately 6 percent of the school-age population has serious arithmetic difficulties unexplained by cognitive or sensory functioning. The study of mathematics disorder has been neglected even though it appears to occur with the same frequency as learning disorders in other academic areas. Several researchers have asserted that this situation reflects cultural values that afford literacy skills (i.e., reading and writing) a higher status than mathematics skills.

The sex ratio of the disorder is also unclear, and data are limited. However, girls are thought to have a greater predisposition to the disorder than boys.

ETIOLOGY

The cause of developmental mathematics disorder is unknown. The earliest view of mathematics disorder was derived from neurological studies of adults with acquired brain lesions that were accompanied by loss of arithmetical-mathematical skills. That early work implicated abnormalities in the cerebral right hemisphere (occipital) processing of visual-spatial stimuli. Subsequent work indicated that such lesions could not account for all cases of acalculia, in either children or adults. However, when children with developmental language disorders are excluded from consideration, symptom clusters of adults with acquired acalculia and children with mathematics disorder are markedly similar.

Some persons may have a genetic predisposition to mathematics disorder. Anecdotal and case history reports describe families with consistently high or low mathematical abilities. Furthermore, in the general population monozygotic twins show a higher than chance correlation in mathematics test scores.

More-recent models suggest that difficulties with mathematics are produced by differences in cognitive development that result in deficient learning strategies for arithmetic. In one model, internal developmental-cognitive factors and external environmental factors interact to produce the deficient learning strategies. Internal developmental-cognitive factors include (1) a general intelligence factor, (2) a verbal ability factor, (3) a visual-spatial ability factor, and (4) a numeric factor (specific to numerical test performance). External environmental factors identified as contributing to poor performance in mathematics include (1) complexity of subject matter, (2) insufficient mastery of prerequisite skills, (3) motivational disposition, and (4) inadequate instruction. Although educators have asserted that the most significant factors responsible for poor mathematical performance are such nonpersonal factors as confusing textbooks and inadequate teaching, these factors are generally discounted in individuals identified as having mathematics disorder.

DIAGNOSIS AND CLINICAL FEATURES

Common signs of mathematics disorder are listed in Table 35.2-1. The age of onset is unclear. Diagnosis may be made as early as first grade, but it is usually made after several years of schooling. Some children with mathematics disorder can make normal progress by using rote memory for the first several years at school. Those children begin to exhibit problems only in the later grades, when the demands in mathematics increase and rote memory is insufficient for problem solving.

Difficulty learning number names
Difficulty establishing rote count sequences
Difficulty learning to print numerals
Difficulty with concepts of combining and separating
Difficulty learning the meaning of operational signs or accurately acting on those signs
Difficulty in committing basic facts to memory
Difficulty with concept of place value
Difficulty in aligning numerals when copying problems
Difficulty maintaining alignment during computation
Idiosyncratic error patterns
Inaccurate computations
Slow computation
Difficulty with word problems
Difficulty with graphic representation of information

Courtesy of J.E. Fleischner, K. Garnett, B. Silver.

Table 35.2-1 Common Signs of Mathematics Disorder

The specific symptoms of mathematics disorder vary across individuals, falling into four major groups: linguistic symptoms, perceptual symptoms, mathematical symptoms, and attentional symptoms. Linguistic symptoms include difficulties in understanding or naming mathematical terms, operations, or concepts; decoding written problems into mathematical symbols; and understanding concepts used in mathematical problems (e.g., more and less, first and last, and before and after). Perceptual symptoms include difficulties recognizing or reading numerical symbols or arithmetical signs, clustering objects into groups, aligning strings of numbers during calculations, and ordering strings of numbers.

Mathematical symptoms include difficulties performing basic arithmetical operations, memorizing numerical facts, following sequences of mathematical steps, counting objects, and multiplying. Learning the multiplication tables frequently poses great difficulty for many children with mathematics disorder. Such children frequently develop elaborate adaptive strategies to avoid multiplication. These strategies, however, often involve an increased number of processes (grouping, adding, and counting) and may produce higher rates of errors.

Attentional symptoms include inaccurate copying of numbers; omitting digits, decimals, or symbols when writing answers; forgetting to add in carried numbers during addition; and failing to note arithmetical signs (e.g., adding when subtraction is indicated). Such errors are common among children with learning disorders. Sometimes they represent inattention or carelessness, but at other times they represent a true lack of understanding of what is needed in mathematical operations. One study of such errors showed that they occurred most often in problems on which the students spent the most time.

Other developmental learning disorders and difficulties are often associated with mathematics disorder. Reading disorder, disorder of written expression, expressive language disorder, mixed receptive-expressive language disorder, and developmental coordination disorder have all been reported in association with mathematics disorder. Specific cognitive processing difficulties (e.g., auditory-verbal deficits, visual-spatial deficits, motor deficits, memory deficits, and attention deficits) may be associated features.

Social immaturity, school and peer problems, social skills deficits, anxiety, and depression have also been reported as associated problems. However, the psychiatric syndrome that is most often found is attention-deficit/hyperactivity disorder. The precise nature of the association is not clear; however, three recent studies shed some light. One study compared relative mathematical underachievers with relative reading underachievers and found that the mathematically impaired children had higher rates of inattention but not of hyperactivity or impulsivity. The second study examined children with attention-deficit disorder and found correlations between measures of vigilance and distractibility and arithmetic test performance. The third study found that older children with attention-deficit disorder were significantly more likely than younger children with the disorder to show a discrepancy between intelligence levels and mathematical achievement. The DSM-IV diagnostic criteria for mathematics disorder are listed in Table 35.2-2.

A. Mathematical ability, as measured by individually administered standardized tests, is substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education.

B. The disturbance in criterion A significantly interferes with academic achievement or activities of daily living that require mathematical ability.

C. If a sensory deficit is present, the difficulties in mathematical ability are in excess of those usually associated with it.

Coding note: if a general medical (e.g., neurological) condition or sensory deficit is present, code the condition on Axis III.

Table 35.2-2 DSM-IV Diagnostic Criteria for Mathematics Disorder

Differential Diagnosis The differential diagnosis requires individual standardized testing of both intellectual functioning and achievement in mathematics. Mathematics disorder is indicated by an achievement score significantly (e.g., two standard deviations) below the score expected on the basis of the child's chronological age and intelligence.

Poor development in mathematics may also be associated with mental retardation, poor vision or hearing, and inadequate education. In mental retardation achievement scores in mathematics are generally commensurate with scores for intellectual functioning. However, children with mild mental retardation may receive the concurrent diagnosis of mathematics disorder if mathematical skills are impaired beyond other levels of functioning. Impaired vision or hearing may be ruled out with audiometric or vision screening tests. Inadequate educational exposure can usually be determined by obtaining a detailed educational history from the parents and the school. The history should cover such information as number of schools attended; frequency of school absences; presence of personality conflicts among the child, parents, and teacher; and the curriculum and styles of teaching used.

Rosa's sixth-grade general education mathematics teacher referred Rosa to the school's Student Study Team for evaluation to determine whether she might have a specific learning disability after noting the following behaviors in class: Rosa constantly lost her place on text pages; she appeared unable to copy problems accurately, either from the board or from a page; her written mathematics work was carelessly completed according to the teacher, with numerous mistakes such as copying operation symbols incorrectly and incorporating the number of the problem into the problem itself; and she continually reversed the order of digits, especially in two- and three-digit problems (e.g., writing 52 for 25).

After a series of in-class accommodations (e.g., moving Rosa's seat closer to the front of the classroom) and modifications of materials (e.g., reducing the number of problems required in assignments), the Student Study Team recommended a full assessment of Rosa's difficulties in mathematics.

Testing consisted of administering the Woodcock Johnson Psycho-Educational Battery-Revised, parts 1 and 2; the Enright Diagnostic Test of Mathematics; the Brigance Test of Basic Skills; and various curriculum-based assessments. The testing revealed that Rosa had a specific learning disability in the area of mathematics and thus qualified for special education and related services. More specifically, the multidisciplinary team, assembled for an Individualized Education Program meeting with Rosa's family, reported that Rosa encountered difficulties in mathematics due to visual perceptual deficits, evidenced by figure-ground difficulties, discrimination problems, and reversals. They recommended specific interventions in the general education classroom by her present teacher, with support provided to this teacher by a resource specialist specifically trained to help students encountering these difficulties.

To address Rosa's figure-ground difficulties, a geometrically shaped template was used to box problems in the textbook, effectively isolating one problem at a time. Additionally, problem numbers were color-coded to differentiate the number of the problem from digits in the problem itself. These modifications were incorporated into assignments by the teacher so that Rosa would box out problems and color-code problem numbers for homework and actually solve the problems in the subsequent class meeting. These adaptations were reported to be quite effective by the general education classroom teacher.

To address Rosa's difficulties with copying operation symbols and reversing numbers, lined paper was used to align number columns, and color-coding was used to identify each operation symbol (e.g., addition was identified by a blue “+” symbol) as well as ones, tens, and hundreds place values (e.g., ones identified in green, tens in red, etc.).

At a 6-month follow-up meeting with Rosa's family, the general education mathematics teacher reported that Rosa was making significant progress and had appeared to be using the curricular modifications effectively; for example, fewer reversals occurred and her operational symbol confusion was no longer observed.

Roger, a 15-year-old student, was recently diagnosed as having a nonverbal learning disability known as dyscalculia; in other words, Roger's verbal skills, as measured on the Wechsler Intelligence Scale for Children–Third Edition, were in the average range, but performance skills were significantly below average. Teachers consistently reported that Roger exhibited good-to-excellent reading, spelling, and writing skills but encountered difficulty in mathematic tasks that required abstract reasoning, particularly mathematic word problems. Teachers described Roger's work as sloppy and lacking proper organization. His parents reported that Roger had always experienced these difficulties; additionally, they indicated that Roger had always had difficulty correctly perceiving certain social cues such as tone of voice and facial expression, and that these deficits impeded his ability to build and maintain friendships with peers.

Presently, Roger is receiving special education in a resource room setting. His resource specialist created a highly structured program for Roger that systematically addresses his difficulties with mathematical word problems. In particular, the resource specialist uses flip charts that take Roger through each step of a mathematical word problem sequentially; a separate step is diagramed on each page and illustrations present concepts visually. This technique has had moderate success with Roger.

Recently, the resource teacher taught Roger how to use a highlighter pen to identify essential and extraneous information contained in mathematical word problems: yellow ink indicates important information necessary to solve the problem and blue ink identifies information unnecessary for correct problem solution. This highlighting strategy was combined with a poster chart that identifies key words indicating certain operations in the word problems (e.g., the words “take away” are equivalent to subtraction)

.

At follow-up, participation in this program resulted in marked improvement in Roger's performance on mathematical word problems; through consultation with the general education teacher, the resource specialist recommended that Roger be mainstreamed into a general education ninth-grade mathematics class.

COURSE AND PROGNOSIS

Although difficulties with counting and number concepts may be apparent as early as kindergarten, diagnosis is generally not made until the second or third grade. The prognosis of mathematics disorder is unclear; case reports show a full spectrum of outcomes, from chronic disability through partial recovery to complete remission of mathematical difficulties. Currently it is not known whether those differing outcomes reflect different subtypes of underlying disorder, different severity levels or degrees of impairment, or treatment, environmental, or other factors. School and peer relationship problems are considered possible complications, but how they develop has not been studied.

TREATMENT

The only treatment of mathematics disorder proved effective is systematic instruction guided by eight principles—two general and six specific. The first general guiding principle is making sure to consider the strengths and weaknesses of the individual with mathematics disorder when planning and implementing a course of education. To investigate the student's abilities, one should determine whether the student (1) comprehends number structure and arithmetic operations, (2) exhibits evidence of adequate spatial-relation skills, (3) has language difficulties that might contribute to learning mathematics, (4) has reading difficulties that affect mathematics learning, and (5) encounters attentional or memory problems or both that might affect mathematics performance.

The second general guiding principle involves providing a balanced mathematics program. In other words, the instructional program should include concepts, skills, and problem solving in an appropriate combination. The teaching of concepts emphasizes such basic understanding as the ability to group and classify objects and skills instruction equips students to perform correct operations using number concepts. Concepts and skills are both used in the third element of a balanced mathematics program, problem solving. In problem solving activities, students are given opportunities to apply understanding of concepts and skills acquisition to new and different settings.

In addition to these general principles, six specific principles are used to guide mathematics instruction for those with mathematics disorder. The first involves checking previously acquired learning of numbers to ensure that these individuals have the necessary foundation (prenumber learnings such as matching, recognizing groups of objects, relationships of parts to whole operations, among others) for more-advanced and abstract concepts.

Second specific guiding principle is that teaching of mathematical concepts should progress through concrete, semiconcrete, and abstract instructional phases. For example, at the concrete stage, concepts are demonstrated by allowing students to manipulate real objects; in the semiconcrete stage, graphic representations are substituted for actual, real-world objects to allow further practice in concept learning; and in the abstract stage, numerals are finally represented by graphic symbols (i.e., 4 + 6 = 10).

The third specific instructional principle emphasizes the need for constant review, drill, and practice of concepts. The objective is enabling the student to use mathematical concepts with a high degree of automaticity. Worksheets, flashcards, and instructional and computer games can be incorporated into the educational program to achieve this objective.

The next principle involves purposeful design and implementation of activities that allow individuals to generalize learned concepts to new and novel situations. For example, measurement skills can be applied to such real-world situations as following a recipe.

A fifth specific principle requires the explicit teaching of mathematics vocabulary to individuals with mathematics disorder. Although many students with mathematics disorder understand the operations required to complete calculation problems successfully, at the same time, they lack the vocabulary (e.g., division requires understanding the terms *divisor* and *quotient*) necessary to unerstand operations fully.

The sixth specific guiding principle involves using calculators, not as a substitution for learning numerical concepts but in lessons emphasizing mathematical reasoning. In other words, calculators are used judiciously and primarily for advanced conceptual work, not when the objective is teaching calculation procedures.

In addition, several recent approaches have been developed to address the unique learning needs of individuals with mathematics disorder: Acronym Mnemonics, the Think Aloud strategy, and the Cognitive Assault strategy. Acronym Mnemonics teaches children with mathematics disorder in primary grades to use a systematic procedure to answer mathematic fact questions from memory. For example, students learn to use the acronym SOLVE whenever they cannot quickly remember a math fact. SOLVE stands for the following steps:

1. **S**ee the sign. (Cues students to determine which operation to use on the math fact.)
2. **O**bserve and answer (if unable to answer, keep going). (Cues students to look at the problem and provide an answer. If the answer is unknown, students continue with steps 3 through 5.)
3. **L**ook and draw. (Reminds students to figure out answers by looking at numbers and drawing tallies.)
4. **V**erify your answer. (Reminds students to check their answers to ensure accuracy)
5. **E**nter your answer. (Cues students to record answers in the space provided.)

This approach is reportedly highly effective with children who have difficulty remembering basic math facts.

The second approach was developed for use with middle school students with mathematics disorder. Based on previous observations that students with mathematics disorder make fewer cognitive and metacognitive verbalizations than normal peers when engaged in solving word problems, this approach teaches students with mathematics disorder to use Think Aloud procedures to solve one-, two-, and three-step word problems. In addition to training students to become more strategic in their problem solving, this approach also resulted in increased perseverance by individuals with mathematics disorder, even when faced with tasks of increasing difficulty.

The third approach, the Cognitive Assault Strategy, was developed for secondary students with mathematics disorder. In this approach, students are assigned instructors or mentors who are responsible for detecting and correcting errors and developing verbalization and writing skills to produce a written description of how

students solve problems. More specifically, students receive a mathematics problem and then verbalize and write out all steps associated with solving it; next, the students write down each step in the sequence, indexing each step with its location in the actual problem; finally, students laminate correct solutions and place them in a notebook for future reference when encountering similar problems in the future. Additional steps include providing a title just above the problem (e.g., basic algebra problem) and possibly color-coding similar types of problems (e.g., algebra problems coded green, calculus problems coded orange). Overall, high school students who have used this approach have gained increased understanding of concepts and procedures and encounter similar problems with more confidence.

SUGGESTED CROSS-REFERENCES

Mental retardation is discussed in Chapter 34 and attention-deficit disorders are covered in Chapter 39. Developmental coordination disorder is discussed in Chapter 38, expressive language disorder in Section 37.1, mixed receptive-expressive language disorder in Section 37.2, and reading disorder in Section 35.1.

SECTION REFERENCES

Babbitt BC, Miller SP: Using hypermedia to improve the mathematics problem-solving skills of students with learning disabilities. J Learn Disabil *29:*391, 1996.

Benbow CP: Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects and possible causes. Behav Brain Sci *11:*169, 1988.

*Bley NS, Thornton CA: *Teaching Mathematics to Students with Learning Disabilities*, ed 3. PRO-ED, Austin, TX, 1995.

Bryant BR, Rivera DP: Educational assessment of mathematics skills and abilities. J Learn Disabil *30:*57, 1997.

*Carey WB: Problems in diagnosing attention and activity. Pediatrics *103:*664, 1999.

*Cawley JF, editor: *Cognitive Strategies and Mathematics for the Learning Disabled*. National Institute of Mental Health, Rockville, MD, 1985.

Cocking RR, Mestre JP, editors: *Linguistic and Cultural Influences on Learning Mathematics*. Erlbaum, Hillsdale, NJ, 1988.

*Deloche G, Seron X, editors: *Mathematical Disabilities: A Cognitive Neuropsychological Perspective*. Erlbaum, Hillsdale, NJ, 1993.

Dunn C, Rabren K: Functional mathematics instruction to prepare students for adulthood. LD Forum *21:*34, 1996.

Grant ML, Ilai D, Nussbaum NL, Bigler ED: The relationship between continuous performance tasks and neuropsychological tests in children with attention-deficit hyperactivity disorder. Percept Motor Skills *70:*435, 1990.

Hofmeister A: Elitism and reform in school mathematics. Remed Spec Educ *14:*8, 1993.

Hutchinson N: Effects of cognitive strategy instruction on algebra problem solving of adolescents with learning disabilities. Learn Disabil Q *16:*34, 1993.

Jitendra AK, Hoff K: The effects of schema-based instruction on the mathematical word-problem-solving performance of students with learning disabilities. J Learn Disabil *29:*422, 1996.

Kelly B, Carnine D: Teaching problem-solving strategies for word problems to students with learning disabilities. LD Forum *21:*5, 1996.

Koontz KL, Berch DB: Identifying simple numerical stimuli: Processing inefficiencies exhibited by arithmetic learning disabled children. Math Cogn *2:*1, 1996.

Lewis RB: Assistive technology and learning disabilities: Today's realities and tomorrow's promises. J Learn Disabil *31:*16, 1998.

Marsh LG, Cooke NL: The effects of using manipulatives in teaching math problem solving to students with learning disabilities. Learn Disabil Res Pract *11:*58, 1996.

Mastropieri M, Scruggs T, Whittaker M, Bakken J: Applications of mnemonic strategies with students with mild mental disabilities. Remed Spec Educ *15:*34, 1994.

Mayer R: Understanding individual differences in mathematical problem solving: Towards a research agenda. Learn Disabil Q *16:*2, 1993.

Miles DD, Forcht JP: Mathematics strategies for secondary students with learning disabilities or mathematics deficiencies: A cognitive approach. Intervent School Clin *31:*91, 1995.

Miller SP, Mercer CD: Mnemonics: Enhancing the math performance of students with learning difficulties. Intervent School Clin *29:*78, 1993.

*Montague M, Applegate B: Middle school students' mathematical problem solving: An analysis of think-aloud protocols. Learn Disabil Q *16:*19, 1993.

Nussbaum NL, Grant ML, Roman MJ, Poole JH, Bigler ED: Attention deficit disorder and the mediating effect of age on academic and behavioral variables. J Dev Behav Pediatr *11:*22, 1990.

Rivera DP: Using cooperative learning to teach mathematics to students with learning disabilities. LD Forum *21:*29, 1996.

Rivera DP: Effective mathematics instruction for students with learning disabilities: Part two of the series. LD Forum *21:*4, 1996.

Rosenberg PB: Perceptual-motor and attentional correlates of developmental dyscalculia. Ann Neurol *26:*216, 1989.

Rourke BP, Conway JA: Disabilities of arithmetic and mathematical reasoning: Perspectives from neurology and neuropsychology. J Learn Disabil *30:*34, 1997.

*Rourke BP, Strang JD: Subtypes of reading and mathematical disabilities: A neuropsychological analysis. In *Developmental Neuropsychiatry*, M Rutter, editor. Guilford, New York, 1983.

Shalev RS, Weirtman R, Amir N: Developmental dyscalculia. Cortex *24:*555, 1988.

Sikora D, Plapinger D: Using standardized psychometric tests to identify learning disabilities in students with sensorineural hearing impairments. J Learn Disabil *27:*352, 1994.

Strang JD, Rourke BP: Adaptive behavior of children who exhibit specific arithmetic disabilities and associated neuropsychological abilities and deficits. In *Neuropsychology of Learning Disabilities: Essentials of Subtype Analysis*, BP Rourke, editor. Guilford, New York, 1985.

Woodward J: Procedural knowledge in mathematics: The role of the curriculum. J Learn Disabil *24:*242, 1991.

Zawaiza T, Gerber M: Effects of explicit instruction on math word-problem solving by community college students with learning disabilities. Learn Disabil Q *16:*64, 1993.

Zentall S, Ferkis MA: Mathematical problem solving for youth with ADHD, with and without learning disabilities. Learn Disabil Q *16:*6, 1993.

CHAPTER 35. LEARNING DISORDERS

35.3 DISORDER OF WRITTEN EXPRESSION AND LEARNING DISORDER NOT OTHERWISE SPECIFIED

MICHAEL E. SPAGNA, FH.D., DENNIS P. CANTWELL, M.D. AND LORIAN BAKER, FH.D.

DISORDER OF WRITTEN EXPRESSION

Definition Disorder of written expression is a significant impairment in written communication that is not attributable to low intelligence; visual, hearing, or physical disorders; emotional disturbance; environmental, cultural, or economic disadvantage; or lack of adequate instruction. The written compositions of persons with the disorder are usually short, poorly organized, and lack sufficient ideation. These written products are also replete with errors in handwriting, spelling, punctuation, and capitalization. The disorder is frequently associated with other learning and language disorders.

History and Comparative Nosology The major focus of learning has traditionally been on reading and only recently has any attention been paid to writing. The first attention paid to writing difficulties focused on handwriting and spelling. The earliest view was that writing difficulties were always associated with reading difficulties.

In the late 1960s the first comprehensive work was carried out describing disorders of written language. These early case studies reported learning disorders that only occurred in written form, with other forms of verbal behavior remaining intact. Difficulties in written expression were grouped into three primary categories: (1) disorders in visual-motor integration (persons could speak and read but could not correctly execute the motor operations necessary to print symbols such as letters and numbers), (2) deficits in revisualization (individuals could recognize and read words but could not revisualize letters and words and thus could not write words from dictation or spontaneously), and (3) deficiencies in formulation and syntax (persons could copy printed symbols accurately, could revisualize words, but could not organize thoughts into meaningful written communication).

Since this first set of disorders involved visual-motor integration and in an effort to parallel the use of the term “dyslexia” for reading disorder, the term “dysgraphia” was the first term associated with disorders of written expression. Other terms used to refer to spelling disorder were *spelling dyslexia, specific spelling retardation, developmental spelling retardation*, and simply *spelling disorder*.

The revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) was the first edition of DSM to include a description of a disorder of written language, then called “developmental expressive writing disorder.” The disorder was defined as an impairment in the development of writing skills not explainable by mental retardation or inadequate schooling and not due to visual or hearing defects or to a neurological disorder. Writing skills according to that definition included spelling, grammar, punctuation, sentence construction, and paragraph organization. The disorder is included in the fourth edition of DSM (DSM-IV) as disorder of written expression, with only minor changes in the wording of the definition. The disorder is also included in the 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) with essentially the same diagnostic criteria (see Table 35.1-2) as those in DSM-IV (i.e., including spelling, grammar, and compositional difficulties).

Epidemiology The prevalence of disorder of written expression has not been studied. It is thought that the prevalence and sex ratio of the disorder are roughly similar to those of reading disorder (i.e., prevalence of 4 percent in school-age children; sex ratio of three to four boys to each girl). However, that does not mean that reading disorder and disorder of written expression always co-occur.

Etiology The cause of disorder of written expression is unknown, and the factors that may be associated with the disorder have not been examined. The factors associated with spelling difficulties have been studied, but the studies did not necessarily consider other forms of writing difficulties or associated reading disorder.

Etiological factors are generally assumed to be similar to those for reading disorder or the specific developmental language disorders. Some professionals believe that underlying cognitive processing difficulties are present, but no specific types have been identified.

Difficulties with attention and concentration, visual memory, and eye-hand coordination can all produce writing problems (e.g., careless punctuation, misspellings, and poor letter formation). Similarly, language disorder can be manifest in written language (e.g., word-finding problems, limited knowledge of grammar and syntax, and inability to sequence ideas expressively). How frequently or to what extent linguistic or other deficits underlie disorder of written expression is unknown.

Diagnosis and Clinical Features The essential feature of disorder of written expression is significant impairment in the ability to compose written text. The impairment may be exhibited in a variety of ways, the most common of which are spelling errors, grammatical or syntactic errors, punctuation errors, poor paragraph organization, and excessively poor handwriting. Written compositions may also show poor thematic development.

Spelling errors are among the most common and most noticeable errors in the writings of children with disorder of written expression. Various types of misspellings may occur, but the most common type appears to be phonetic or phonographical errors. In these errors, misspellings reflect the way the word sounds (e.g., spelling “fotograf” for “photograph” or “caim” for “came”). These spelling errors are similar to those made by young children just learning to spell. The misspellings of children with disorder of written expression sometimes resemble a mispronunciation or slurring of the word (e.g., “ineverable” for “inevitable” or “offen” for “often”). Other less common types of spelling errors include morphological errors (e.g., “laughter” for “laughed”), segmentation errors (e.g., “away” for “away”), reversals (e.g., “god” for “dog”), and spelling-rule errors (e.g., “bitting” for “biting”).

Grammatical or syntactic errors include word omissions, incorrect word order, incorrect verb and pronoun usage, and incorrect word endings. Written sentences may be incomplete or run-on sentences. Syntactic analyses of written language (e.g., mean number of words, subordinate clauses, or morphemes per sentence) typically reveal unsophisticated syntax. Among older (i.e., college) students with disorder of written expression, the grammatical and syntactic errors may be seen only with higher-level linguistic structures, such as coordination or subordination.

Punctuation and capitalization errors, the so-called writing mechanics, are also considered a major part of disorder of written expression. These errors are particularly common in the writings of younger children but also occur in the works of older disabled writers. Some researchers feel that this aspect of writing poses greater difficulties than do the more creative areas, such as choice of vocabulary and thematic maturity.

Excessively poor handwriting is another indication of disorder of written expression. Disordered handwriting (as opposed to handwriting that is simply messy) includes letter forms that are not decodable, rotations or inversions of letters, a mixture of printing and cursive writing, and inappropriate mixtures of upper-case and lower-case letters.

Poor story composition and poor paragraph organization are other essential features of disorder of written expression. Two main aspects of poor story composition are poor or missing story components (e.g., settings, characters, themes, conflict, and conflict resolution) and poor cohesion (e.g., unclear referents, rough transitions, and abrupt endings). Thus, the written paragraphs and stories of children with this disorder usually convey a clear message but have an unplanned, simplistic quality.

Learning disorders thought to co-occur frequently include phonological disorder, expressive language disorder, mixed receptive-expressive language disorder, reading disorder, developmental coordination disorder, and mathematics disorder. Deficits in social skills and self-esteem and peer relationship problems may accompany disorder of written expression. Attention-deficit disorder is the psychiatric disorder most likely to co-occur.

The diagnostic criteria for disorder of written expression appear in Table 35.3-1.

A. Writing skills, as measured by individually administered standardized tests or functional assessments of writing skills), are substantially below those expected given the person's chronological age, measured intelligence, and age-appropriate education.
B. The disturbance in criterion A significantly interferes with academic achievement or activities of daily living that require the composition of written texts (e.g., writing grammatically correct sentences and organized paragraphs).
C. If a sensory deficit is present, the difficulties in writing skills are in excess of those usually associated with it.
Coding note: if a general medical (e.g., neurological) condition or sensory deficit is present, code the condition on Axis III.

Table 35.3-1 DSM-IV Diagnostic Criteria for Disorder of Written Expression

B.C. was a 12-year-old male student who presented for evaluation of problems in school. He attended an academic preschool and was presently enrolled in a regular sixth grade class at a public school.

Current evaluation revealed no history of neurological, visual, or hearing problems that could explain B.C.'s school difficulties. Intelligence testing revealed high-average scores in both the verbal and performance subtests of the Wechsler Intelligence Scale for Children-III (WISC-III). Reading and mathematical scores on standardized tests of academic performance were consistent with his intelligence and chronological age; however, spelling scores were significantly below the predicted level of performance. Multiple misspellings occurred. Although the examiner noted adequate handwriting, B.C. appeared unable to express thoughts in complete sentences. His sentences were short and failed to state intended points clearly. Careful study of B.C.'s written paragraphs revealed numerous grammatical and syntactic errors as well as errors in punctuation and capitalization.

The clinical picture of the inability to compose a written text, poor spelling, and grammatical errors in the absence of low intelligence, problems with reading or mathematics, and pervasive attentional problems led to a diagnosis of a disorder of written expression.

Differential Diagnosis The diagnosis of disorder of written expression requires a score on an individually administered standardized test of spelling, handwriting, or written expression that falls significantly below that expected for the child's chronological age and measured intelligence. Numerous educational tests provide reliable standardized assessments of spelling skills. However, few standardized tests assess writing skills. Available tests of written language include the Test of Written Language, first and second editions (TOWL), the Diagnostic Evaluation of Writing Skills (DEWS), and the Test of Early Written Language (TEWL). In addition, some tests of educational achievement and of language development have subtests dedicated to written language (e.g., California Achievement Tests, Sequential Tests of Educational Progress, Test of Adolescent Language, and Woodcock-Johnson Psychoeducational Battery–Revised).

Normative standards differ for the different tests, and clinicians are cautioned to check those standards before relying on a particular test. Also, some tests do not require writing samples but provide examples of poor writing to be identified. In effect, they test editing rather than writing.

Owing to the limited number of tests available, DSM-IV does not require standardized testing in its diagnostic criteria. Instead, samples of a child's written schoolwork may be analyzed and compared with those of peers of similar age and intelligence. The limited studies available suggest that analysis of a spontaneous sample of written composition is as effective in distinguishing children with syntactic difficulties as a formal test for written language.

Reliable judgments of disorder require experience with the levels of writing acquisition appropriate for different age levels. Telegraphic statements, verb-tense switching, and repetitious sentence forms, for example, are normal in the writing of a first-grade child but not of a third-grade child. Similarly, unrelated characters and actions and focus on the ending rather than on the details leading to the ending are normal in the writing of a second-grade child, but not in that of a fourth-grade child.

Mental retardation, impaired vision or hearing, impaired motor coordination, inadequate schooling, language disorders, learning disorder not otherwise specified, and attention-deficit disorder can all produce difficulties in aspects of written language. In a mentally retarded child (as in a child with normal intelligence) disorder of written expression may be diagnosed only if the child's writing skills are significantly below those expected of children with that general intellectual functioning. Impaired vision and hearing can both be ruled out through screening tests. Inadequate schooling can usually be ruled out by a careful history.

Impaired motor coordination, as a result of developmental coordination disorder or neuromotor damage, may produce illegible handwriting. If spelling and expression of thoughts in writing are not affected, the diagnosis of disorder of written expression is not made.

Attention-deficit disorders may be associated with disorder of written expression or may itself result in difficulties with producing written work. Children with attention deficits may not pay attention to the details of writing, so that their written work is marked by careless punctuation errors and messy handwriting. In such cases, careful attention to spelling and syntactic performance is needed to determine whether the diagnosis of disorder of written expression should be made. Language disorders frequently precede disorder of written expression. Concurrent diagnoses of oral language disorders and disorder of written expression are permissible. However, DSM-IV suggests that severe disorder of written expression along with severe disorders in reading and mathematics, should be diagnosed as a learning disorder not otherwise specified (Table 35.3-2) rather than as three separate diagnoses.

This category is for disorders in learning that do not meet criteria for any specific learning disorder. This category might include problems in all three areas (reading, mathematics, written expression) that together significantly interfere with academic achievement even though performance on tests measuring each individual skill is not substantially below that expected given the person's chronological age, measured intelligence, and age-appropriate education.

Table 35.3-2 DSM-IV Diagnostic Criteria for Learning Disorder Not Otherwise Specified

Course and Prognosis Symptoms of writing difficulties, such as letter reversals and the inability to copy printed materials, may appear quite early. However, diagnosis of disorder of written expression should generally not be made until the age of 8 years.

Diagnosis is difficult and often inappropriate at early ages for several reasons. First, sufficient motor development to coordinate hand, arm, and shoulder movements necessary for pencil control is frequently not acquired until the age of 6 years. Second, depending on educational opportunities and socioeconomic background, sufficient exposure to written language may not take place until after the age of 7 years. Third, letter reversals, miscopying, and difficulty in mastering graphemes are normal in the early stages of writing acquisition. Finally, in the first several grades of school only simple narrative writings are expected, which may not reveal milder difficulties. The most common age of diagnosis appears to be 10 years (fifth grade). At this time the writing curriculum shifts from narrative to expository text, a shift that most children with disorder of written expression cannot make.

The course of the disorder has not been systematically studied. It is assumed that it resembles reading disorder in that writing skills eventually will improve, but aspects of the disorder may remain throughout life.

The course and outcome of spelling problems have been somewhat studied. In the Isle of Wight studies, spelling problems were more persistent than reading problems, even with remedial interventions.

Although there are no longitudinal studies of children with disorder of written expression, four publications reported on the writing abilities of learning-disabled children at different age levels. All of these papers found the writings of older children superior to those of younger children.

Just as there have been no studies on the possible etiological factors underlying disorder of written expression, no studies have examined factors that may influence the outcome of the disorder. Both educational factors (e.g., opportunities to practice writing and teaching of mechanics of writing) and personal factors (e.g., lack of self-confidence and negative attitude) are believed relevant.

Treatment Educational intervention is thought to improve disorder of written expression, although specific studies of its effectiveness are lacking. A number of educational approaches to remediation of disorder of written expression exist and no convincing evidence favors any one of them. Both formal instruction in writing and practice with writing are believed to be beneficial. Recent attention has focused on the use of computers to teach writing skills, but there is no empirical evidence that computer use is superior to any other method.

LEARNING DISORDER NOT OTHERWISE SPECIFIED

Learning disorder not otherwise specified is a catch-all category for disorders in learning that do not meet the criteria for a specific learning disorder (Table 35.3-2). The category could be used, for example, for a specific spelling disorder.

ICD-10 includes the diagnosis of specific spelling disorder, defined as spelling problems in the absence of reading and arithmetic difficulties (see Table 35.1-2). A specific spelling disorder is not included in DSM-IV, reflecting the view that spelling difficulties are pervasive in the general population. Consequently, in most of the United States, children who have difficulties only with spelling are not eligible for learning disability programs.

SUGGESTED CROSS-REFERENCES

Reading disorder is discussed in Section 35.1, mathematics disorder in Section 35.2, and developmental coordination disorder in Section 36. Expressive language disorder is discussed in Section 37.1, mixed receptive-expressive language disorder in Section 37.2, and phonological disorder in Section 37.3. Mental retardation is the subject of Chapter 34 and attention-deficit disorders of Chapter 39.

SECTION REFERENCES

*Baroff GS: General learning disorder: A new designation for mental retardation. Ment Retard *37:*68, 1999.

Bereiter C, Scardamalia M: *The Psychology of Written Composition*. Erlbaum, Hillsdale, NJ, 1987.

Bulgren J, Scanlon D: Instructional routines and learning strategies that promote understanding of content area concepts. J Adolesc Adult Literacy *41:*292, 1998.

Crank J, Bulgren J: Visual depictions as information organizers for enhancing achievement of students with learning disabilities. Learn Disabil Pract *8:*140, 1993.

Dobbie L, Askov EN: Progress of handwriting research in the 1980s and future prospects. J Educ Res *88:*339, 1995.

Flowers DL: Brain basis for dyslexia: A summary of work in progress. J Learn Disabil *26:*575, 1993.

Freedman SW, editor: *The Acquisition of Written Language Response and Revision*. Ablex, Norwood, NJ, 1985.

Friedland J: Development and breakdown of written language. J Community Disord *23:*171, 1990.

Fulk BM: The effects of combined strategy and attribution training on LD adolescents' spelling performance. Exceptionality *6:*13, 1996.

Gagar A: A computer analysis of written language variables and a comparison of compositions written by university students with and without learning disabilities. J Learn Disabil *22:*125, 1989.

Gleason MM: Using direct instruction to integrate reading and writing for students with learning disabilities. Read Writ Q *11:*91, 1995.

Gordon J, Vaughn S, Schumm JS: Spelling intervention: A review of literature and implications for instruction for students with learning disabilities. Learn Disabil Res Pract *8:*175, 1993.

Graham S, Harris K, Loynachan C: The spelling for writing list. J Learn Disabil *27:*210, 1994.

Hamstra-Betz L, Blote A: A longitudinal study on dysgraphic handwriting in primary school. J Learn Disabil *26:*689, 1993.

*Houck C, Billingsley B: Written expression of students with and without learning disabilities: Differences across the grades. J Learn Disabil *22:*561, 1989.

Laughton J, Morris N: Story grammar knowledge of learning disabled students. Learn Disabil Res *4:*87, 1989.

*Lynch EM, Jones SD: Process and product: A review of the research on LD children's writing skills. Learn Disabil Q *12:*74, 1989.

*MacArthur CA: Using technology to enhance the writing processes of students with learning disabilities. J Learn Disabil *29:*344, 1996.

MacDonald GW, Cornwall A: The relationship between phonological awareness and reading and spelling achievement eleven years later. J Learn Disabil *28:*523, 1995.

Martin KF, Manno C: Use of a check-off system to improve middle school students' story compositions. J Learn Disabil *28:*139, 1995.

Montague M, Leavell A: Improving the narrative writing of students with learning disabilities. Remed Spec Educ *15:*21, 1994.

*Myklebust H: *Developmental and Disorders of Written Language*. Grune & Stratton, New York, 1973.

Nelson JR, Smith D, Dodd J: The effects of learning strategy instruction on the completion of job applications by students with learning disabilities. J Learn Disabil *27:*104, 1994.

*Newcomer PL, Barenbaum EM: The written composing ability of children with learning disabilities: A review of the literature from 1980 to 1990. J Learn Disabil *24:*578, 1991.

Nodine BF, Barenbaum E, Newcomer P: Story composition by learning disabled, reading disabled, and normal children. Learn Disabil Q *8:*167, 1985.

Outhred L: Word processing: Its impact on children's writing. J Learn Disabil *22:*262, 1989.

*Penney CG, Godsell A: Unusual modality effects in less-skilled readers. J Exp Psychol Learn Mem Cogn *25:*284, 1999.

Rourke BP: *Nonverbal Learning Disabilities: The Syndrome and the Model*. Guilford, New York, 1989.

Sexton M, Harris KR, Graham S: Self-regulated strategy development and the writing process: Effects on essay writing and attributions. Except Child *64:*295, 1998.

Shafrir U, Siegel L: Subtypes of learning disabilities in adolescents and adults. J Learn Disabil *27:*123, 1994.

Singer BD: Written language development and disorders: Selected principles, patterns, and intervention possibilities. Top Lang Disord *16:*83, 1995.

Trammel D, Schloss P, Alper S: Using self-recording, evaluation, and graphing to increase completion of homework assignments. J Learn Disabil *27:*75, 1994.

Vaughn S, Hughes MT, Klingner J, Schumm JS: A collaborative effort to enhance reading and writing instruction in inclusion classrooms. Learn Disabil Q *21:*57, 1998.

Welch M: It's in the bag: An instructional game to promote positive student attitudes toward writing. Except Child *27:*63, 1995.

Zipprich MA: Teaching web making as a guided planning tool to improve student narrative writing. Remedial Special Educ *16:*3, 1995.

Textbook of Psychiatry

CHAPTER 36. MOTOR SKILLS DISORDER: DEVELOPMENTAL COORDINATION DISORDER

MICHAEL E. SPAGNA, PH.D., DENNIS P. CANTWELL, M.D., AND LORIAN BAKER, PH.D.

HISTORY

Movement abnormalities were first classified by the medical profession during the fifteenth century. By the 1700s and 1800s, neurological correlates with motor abnormalities were probed and described in the professional literature. This early history is best described as a period when researchers were primarily concerned with investigating and explaining dramatic, involuntary motor behaviors observed in adults supposedly at rest. Not until the early twentieth century, however, were motor signs of clumsiness documented, first described in 1911 as “motor deficiency syndrome,” marked by awkward voluntary actions, excessive tendon reflexes, mild hypertonicity, and neurological overflow. A shift occurred here, with research efforts focusing on “hidden” motor coordination behaviors observed in children given reasonably complex tasks to perform.

By the early 1920s, assessment instruments were developed to evaluate motor awkwardness in children; the most influential device was created in 1922 at the Moscow Neurological Institute. The term “clumsy child syndrome” first appeared in the literature in the 1930s to describe a set of symptoms associated with awkward motor behaviors that did not result from overt neurological damage. During the 1940s, remedial programs were established in an attempt to correct and improve academic learning through treatment of coordination difficulties observed in children.

Not until the 1960s was clumsy child syndrome clinically set apart from other categories of learning disorders. From this period to the mid-1980s, significant investigations were conducted into possible perceptual-sensory causes for motor impairments as well as social and emotional problems associated with such impairments. Over the past decade, scholars have begun to investigate the presence of several subsyndromes suggesting that clumsiness might be a multidimensional condition warranting syndrome-specific intervention and remedial programs.

Currently, the *clumsy child syndrome* implies difficulties or delayed development in fine and gross motor coordination skills. The motor skills of clumsy children tend to be imprecise or mildly affected rather than grossly impaired. In addition, clumsy children tend to have problems in peer relationships and social adjustments and frequently suffer from language and other learning difficulties.

DEFINITION

Both the revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) and the fourth edition of DSM (DSM-IV) define developmental coordination disorder as significantly impaired development of motor coordination not directly attributable to either a neurological disorder or general mental retardation. Children with developmental coordination disorder perform motor coordination tasks at levels markedly below those of peers of the same chronological age and intellectual capacity. The motor performance deficits are sufficiently marked to interfere with academic achievement and activities of daily living. Specific manifestations of the disorder are described in Table 36-1. Many are age related, including delays in achieving motor milestones (walking, crawling, sitting), dropping things or clumsiness, poor performance in sports, and poor handwriting.

Gross motor manifestations
Preschool age
Delays in reaching motor milestones such as sitting, crawling, walking
Balance problems: falling, getting bruised frequently, poor toddling
Abnormal gait
Knocking over objects, bumping into things, destructiveness
Primary-school age
Difficulty with riding bikes, skipping, hopping, running, jumping, doing somersaults
Awkward or abnormal gait
Older
Poor at sports, throwing, catching, kicking, hitting a ball
Fine motor manifestations
Preschool age
Difficulty learning dressing skills (tying, fastening, zipping, buttoning)
Difficulty learning feeding skills (handling knife, fork, or spoon)
Primary-school age
Difficulty assembling jigsaw pieces, using scissors, building with blocks, drawing, or tracing
Older
Difficulty with grooming (putting on makeup, blow-drying hair, doing nails)
Messy or illegible writing
Difficulty using hand tools, sewing, playing piano

Table 36-1 Manifestations of Developmental Coordination Disorder

COMPARATIVE NOSOLOGY

Various terms have been used for developmental coordination disorder including *congenital maladroitness, choreiform syndrome, psychomotor syndrome, developmental apraxia, developmental dyspraxia, congenital clumsiness, developmental agnosia,* and *developmental dyspraxia-dysgnosia*. Although the most appropriate label for the disorder is controversial, the terms in widest current usage are *clumsiness, clumsy child syndrome, development coordination disorder,* and *specific developmental disorder of motor function*.

The disorder was not included in the third edition of DSM (DSM-III) as a separate syndrome, although motor coordination problems were noted as common associated features for most of the childhood developmental disorders. Developmental coordination disorder has been included as a separate diagnosis in DSM-III-R and appears in essentially the same form in DSM-IV. The 10th revision of the *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) also includes the disorder, but the slightly more formalized definition requires an intelligence quotient of at least 70 and performance at least two standard deviations below the mean on a standardized test of motor coordination.

Although DSM and ICD view the disorder as a unitary undifferentiated syndrome, several investigators have suggested that there may actually be several distinct subtypes. Supporting that view is the interindividual variation in symptomatology, only some of which is clearly age related. One recent attempt to subclassify developmental coordination disorder proposed subtypes such as ataxic syndrome (marked by unsteadiness and mild tremors), hypotonic syndrome (characterized by below-average resting muscle tone), tension syndrome (manifested by constant and excessively high levels of muscle tone), dyspraxic syndrome (identified by the inability to chain together a series of submovements necessary to execute smooth motor movements), manual-graphic syndromes (two separate manual syndromes, one marked by inadequate dexterity but adequate ability to reproduce written symbols, and the other characterized by adequate dexterity but inability to reproduce written symbols accurately using fine motor movements), visual-perceptual clumsiness (ranging from ocular-motor problems involving inefficient movement of eye muscles to severe retinal deterioration), and mixed syndromes (a combination of symptoms already presented under the aegis of the previous six classifications). DSM-IV classifies developmental coordination disorder as the only motor skills disorder.

EPIDEMIOLOGY

The prevalence of developmental coordination disorder has been estimated at approximately 5 percent of the school-age population. Developmental coordination disorder has been noted in Europe, Asia, Africa, Australia, and North America. Although it has been traditionally assumed that clumsiness is more prevalent among

males than females, two recent studies indicate that the prevalence figures are the same for male and female children. Discrepancies between these latest findings and earlier estimates of gender differences have been attributed to either of two factors: a tendency for schools to refer more males than females for testing and possible special education and related services, and societal pressures for males to excel at physical tasks at earlier ages than their female counterparts. Further research is required to resolve the present gender imbalance debate.

ETIOLOGY

The mechanisms and causes of developmental coordination disorder are unknown. Different conceptualizations of the disorder exist, but research supporting or differentiating those views is very limited.

Evidence indicates both biological-organic and cognitive-developmental correlates of developmental coordination disorder. The biological-organic factors associated with the disorder include various birth and perinatal problems, such as low birth weight, prematurity, hypoxia, and neonatal malnutrition. Some investigators have postulated a continuum of reproductive casualty. There are also suggestions of incomplete cerebral dominance or lateralization, neurochemical abnormalities in the brain, and structural lesions in the parietal lobes.

Cognitive-developmental problems sometimes associated with developmental coordination disorder include speech-language disorders and delays, hyperactivity, distractibility, impulsivity, and academic and learning difficulties. Apparently, the association between speech-language problems and coordination problems is quite strong. Studies of children with motor coordination delays have frequently reported delays in speech and language development; similarly, studies of children with specific language impairment have frequently reported motor delays and difficulties. One recent study of 82 school-age children with specific, severe speech and language disorders revealed that 90 percent were clumsy and 22 percent had walked late (after 18 months of age).

Whether the association between coordination difficulties and speech-language difficulties is global or is limited to specific aspects of coordination and communication is not known. However, the literature suggests that the associations may be rather specific. One longitudinal study of children with language impairments revealed a continuing correlation between peg-moving performance and language performance. Another study found that fine motor skills, response speed, and upper limb speed were associated with language disturbances whereas gross motor coordination and visual motor control were not. Furthermore, evidence indicates articulation problems are related to motor impairment, even though the motor impairment is not necessarily dyspraxia.

The evidence for any one of those postulated factors is inconclusive. Most likely, developmental coordination disorder has multiple causes. One recent study of motor coordination, speed, and inhibition in poor achievers suggested two underlying processes: specific motor deficits and developmental motor lags.

DIAGNOSIS AND CLINICAL FEATURES

The essential feature of developmental coordination disorder is poor motor coordination. The manifestations may include difficulties or delayed development in fine and gross motor coordination skills. However, motor skills tend to be imprecise or clumsy rather than globally impaired.

Coordination problems are diagnosed according to movement characteristics, generally observed while the person is engaged in several different types of tasks requiring the use of different muscle groups. They are clinically assigned to one or more of seven diagnostic categories: dyspraxia, synkinesia, hypotonus, hypertonus, tremors, impersistence, and asymmetries. *Dyspraxia* describes the inability of the child to produce correctly sequenced, coordinated motor movements when presented with a demonstration or an oral request. Movements identified as *synkinesia* are best described as unintentional muscle movements, also termed *muscle overflow*. Mirror-movements such as finger twitching on the opposite hand when the child is asked to perform a finger opposition task, or facial grimaces when the child is asked to make hand movements, are examples of synkinetic motor actions.

Two categories clinically describe tonus abnormalities, hypotonus and hypertonus. *Hypotonus* can be observed in all parts of the body and is characterized by a flaccid or sleepy quality in the child's facial expression and general muscle tone. These children appear lazy and unfit; the condition is often accompanied by poor posture and obesity. *Hypertonus* describes generally high levels of muscle tone in a child. For example, a child with hypertonus movement cannot finish drawing a line at a given point, instead overshooting the intended stopping point. Another example is a child who tends to throw a ball too hard and inaccurately at short distances.

The movement category of *tremors* is marked by irregular unsteadiness in motor movements observed in tasks requiring walking and/or drawing. For example, legibly formed letters might reveal on closer inspection, rapid tremors in the actual lines themselves. Unsteadiness might be exhibited in the child's gait as high-amplitude fluctuations in leg muscle movement.

The category of *impersistence* refers to the child's inability to sustain and maintain various body postures for reasonable periods of time. Asking a child to “stick out your tongue” is an example of a task used often to investigate impersistent motor movements. In this example, the clinician would keep track of how long the child could maintain this tongue position.

The final movement category, *asymmetries,* describes motor behaviors that affect only one side of the body. This category includes unusual flexions, muscle weakness, or both on one side of the body or general muscle weakness in the limbs on one side. These asymmetric muscle movements can be observed by requiring a child to move laterally in a rapid manner (“like a basketball player on defense, not crossing your feet”). If the child executes the movement more fluidly on one side than on the other, it might indicate an asymmetry.

Specific manifestations vary across individuals; thus there is no typical clumsy child. In some instances, manifestations may be quite specific; one case report describes a child who could sew and do jigsaw puzzles without difficulty but could not write neatly. In fact, total disability in any one area of motor performance is rare.

A common symptom is a history of delays in achieving motor milestones (e.g., sitting, standing, crawling, and walking). Gross motor coordination manifestations may be seen in difficulties with such activities as sitting, walking, running, balancing, throwing, and kicking. Fine motor coordination manifestations may be seen in difficulties in performing such tasks as tying laces, fastening fasteners, drawing, tracing, sewing, and printing or writing. During coordination tasks, children with developmental coordination disorder often (but not always) exhibit fine choreiform movements in unsupported limbs, mirror movements or persistent associated movements, slight abnormalities of reflex, rhythmic disturbances, and other neurological signs.

Some manifestations of developmental coordination disorder are age dependent. Young children show such symptoms as delayed development in early motor milestones and delayed acquisition of such self-help skills as self-dressing (i.e., tying ties, fastening snaps, zipping zippers, buttoning, and unbuttoning) or self-feeding (handling a knife, fork, or spoon). Preschool-age children may also have balance problems, awkward gait, and a history of frequently bumping into objects, falling down, and getting bruised. Problems with balance, hopping, jumping, catching and bouncing a ball, and tracing appear to be among the more common symptoms of school-age children with developmental coordination disorder. Another common sign among school-age children is the very strong motivation (often to the point of truancy) to avoid physical education classes.

In addition, children with developmental coordination disorder frequently suffer from other developmental-learning difficulties. Among the more common of these difficulties are expressive language disorder, mixed receptive-expressive language disorder, phonological disorder, reading disorder, disorder of written expression, and mathematics disorder.

Secondary complications primarily involving peer group problems are frequently reported in developmental coordination disorder. The repeated failures that the children experience in sports and games often engender peer rejections and subsequent relationship and social adjustment problems. The DSM-IV diagnostic criteria for developmental coordination disorder are listed in Table 36-2. The ICD-10 diagnostic criteria for specific developmental disorder of motor function are presented in Table 36-3.

A. Performance in daily activities that require motor coordination is substantially below that expected given the person's chronological age and measured intelligence. This may be manifested by marked delays in achieving motor milestones (e.g., walking, crawling, sitting), dropping things, clumsiness, poor performance in sports, or poor handwriting.
B. The disturbance in criterion A significantly interferes with academic achievement or activities of daily living.
C. The disturbance is not due to a general medical condition (e.g., cerebral palsy, hemiplegia, or muscular dystrophy) and does not meet criteria for a pervasive developmental disorder
D. If mental retardation is present, the motor difficulties are in excess of those usually associated with it.

Coding note: If a general medical (e.g., neurological) condition or sensory deficit is present, code the condition on Axis III.

Table 36-2 DSM-IV Diagnostic Criteria for Developmental Coordination Disorder

A. The score on a standardized test of fine or gross motor coordination is at least 2 standard deviations below the level expected for the child's chronological age.
B. The disturbance described in criterion A significantly interferes with academic achievement or with activities of daily living.
C. There is no diagnosable neurological disorder.
D. *Most commonly used exclusion clause.* I.Q. is below 70 on an individually administered standardized test.

Table 36-3 ICD-10 Diagnostic Criteria for Specific Developmental Disorder of Motor Function

Differential Diagnosis Informal screening for gross motor coordination (e.g., hopping, jumping, and standing on one foot), fine motor coordination (e.g., finger tapping, shoelace tying), and hand-eye coordination (e.g., catching a ball and copying letters) is helpful in identifying difficulties. However, the cutoff between the normally clumsy child and the child with developmental coordination disorder is not always clear. The general guideline is that coordination difficulties interfere with academic achievement or activities of daily living. Formal tests of motor coordination (e.g., the Bruininks-Oseretsky Test of Motor Development, the Frostig Movement Skills Test Battery, the Slosson Drawing Coordination Test, the Southern California Sensory Integration Test, and the Stott Test of Motor Impairment) provide a comparative measure of a child's abilities.

Differential diagnosis must rule out other disorders that may lead to coordination problems, such as specific neurological disorders (cerebral palsy or lesions), pervasive developmental disorders, and generalized mental retardation. Children with physical-neurological disorders tend to exhibit more-global coordination difficulties. Children with mental retardation and pervasive developmental disorders do not show a large discrepancy between coordination skills and other skills.

Clumsiness due to excessive motor activity (in attention-deficit/hyperactivity disorder) must also be ruled out. The tendency for children with developmental coordination disorder to fall down, hurt themselves, and break objects may evoke presenting complaints of behavioral problems such as destructiveness. In the school setting, a child's motor difficulty with manipulating a pencil may result in discouragement and refusal to continue trying, which are mislabeled as laziness. Children with developmental coordination disorder have reportedly presented medically under the guise of various psychosomatic aches and pains.

Kevin (benign hypotonia) was brought to a university clinic for evaluation for suspected muscular dystrophy by his parents when he turned 7 years of age. In an initial interview, parents reported that Kevin had difficulty sitting up without assistance through 23 months of age and, over the past year and a half, had begun to put on excessive weight and had no interest in physical activities.

A comprehensive assessment of fine and gross motor skills yielded the following results: Kevin attempted to complete a skipping task with only partial success; he could initiate a skipping motion but could only succeed with one foot at a time. Kevin could stand with both feet together but could not rise on the balls of his feet without excessive wobbling. Assessment of dynamic balance showed that Kevin could walk on a balance beam 4 inches wide but not one less than 2 inches wide. Kevin's measured throwing motion resembled that of a 5-year-old: he could not step forward simultaneously on the leg opposite his throwing arm. Although Kevin could catch a ball, he cradled a ball bounced to him at chest-high level with stiff arms; he could not catch a tennis ball bounced to him on the ground from a distance of 15 feet. Kevin's agility and coordination were measured with the Bruinink's Oseretsky battery, which revealed functioning levels commensurate with those of an average 5-year-old child.

Based on these assessment results, Kevin was given the diagnosis of benign hypotonia, a nonprogressive disorder believed to be connected to central nervous system deficits. A rigorous physical conditioning program was recommended that included work with physical and occupational therapists as well as a coordinated adapted physical education program to be implemented at Kevin's elementary school. This intervention program was based on several accepted principles of motor learning and acquisition. First, all practice of motor skills took place 3 days a week, with rest days in between for more rapid improvement. Second, it was suggested that parents provide extrinsic rewards to Kevin for his participation in the program; in this case, he would be rewarded with trips to the zoo, park, and so forth by members of his family. Third, all specific motor performance objectives were clearly explained to Kevin as he participated in weekly tasks; for example, skipping activities were demonstrated for Kevin, and his success was graphed and placed on a bulletin board at the physical therapist's office. Finally, Kevin's improved performance on specific tasks was generalized to new environments and activities; for example, catching skills developed through a training program with the physical therapist were later applied to playground activities in adapted physical education at Kevin's school.

As a direct result of 10 months of this comprehensive training program, Kevin mastered many of his motor training performance objectives, and his parents reported that Kevin's weight had stabilized. Additionally, his family reported that Kevin had indicated that he now enjoyed participating in physical education activities at school and wanted to play games with neighborhood friends. At follow-up, the family was told that Kevin should be encouraged to maintain a systematic exercise program to prevent further weight gain and loss of muscle tone.

Maureen (dysgraphia), a 14-year-old student, had been referred to her school's Student Study Team for evaluation to determine whether she had a specific learning disability. The referring teacher had noticed that Maureen's handwriting was barely legible, with poorly formed letters, lack of appropriate spacing between words, and generally poor line quality (i.e., words appearing below and above page lines). One of the team members observed Maureen in class and also noticed that Maureen exhibited a high degree of tension in her face while writing, grimacing periodically as she attempted to form letters and words in handwriting exercises.

A comprehensive battery of tests was administered, including the Test of Written Language, third edition (TOWL-3) and the Zaner-Bloser. Results showed that despite average intellectual functioning and exposure to formal education in handwriting, Maureen had a specific learning disability in the area of handwriting and was thus eligible for special education and related services. More specifically, Maureen was identified as having the specific learning disability known as dysgraphia. Further informal assessment showed that Maureen could print block letters containing right angles (e.g., F, E) but could not execute the smooth fine-motor movements necessary to produce letters formed from circles (e.g., O, Q).

Based on these assessment results, the team decided that Maureen would receive assistance from a resource specialist for one period during the school day. This assistance consisted of using multisensory drill techniques accompanied by fading exercises to overlearn circular letter formation. Additionally, gross-motor skill strengths were used to teach letter formation by using sky-writing tasks. Generalization of these handwriting skills was ensured by allowing Maureen to use letter-formation cards at her desk to reinforce her handwriting when completing independent seatwork.

Maureen made good progress with this program, mastered the formation of all manuscript letters, and progressed to working on cursive letters. Assistance by the resource specialist was discontinued and in a follow-up meeting with the team Maureen's parents reported that they were satisfied with Maureen's academic progress. The family was told that learning typewriting skills with computer word processing software might facilitate development of Maureen's written composition skills.

COURSE AND PROGNOSIS

Little is known about the outcome of developmental coordination disorder; however, impaired coordination appears to be relatively stable. A 2-year follow-up study of children of primary school age who had failed a motor test battery revealed motor performance that, although improved, was still inferior to that of controls. The children with motor impairments also showed poorer educational performance than controls.

A more recent study reported on a 10-year follow-up of clumsy children and their age-matched, sex-matched peers. When first identified, the clumsy children were approximately 6 years of age; at follow-up they ranged from 15 to 17 years of age. Follow-up testing indicated that the clumsy children were still significantly below the controls in most motor skills (e.g., dexterity, ball skills, balance, cutting with scissors, catching, walking backward) and in global ratings by physical education teachers. The clumsy adolescents showed higher rates of dysgraphesthesia, dysdiadokinesis, and motor slowness. In addition the educational achievements of the clumsy adolescents were significantly poorer, and they had significantly lower perceived social competence than their nonclumsy peers. That finding substantiates reports of continuing secondary problems, including peer problems, low self-esteem, and academic difficulties.

Different aspects of motor coordination may progress at different rates. One study of 50 kindergarten and first-grade children typically found some improvement but overall continued poor performance in motor skills. Improvement was greatest in motor speed and poorest for motor inhibition.

TREATMENT

Although motor training programs gained wide acceptance in remediating various types of learning disorders in the late 1960s, 1970s, and early 1980s, no research supports the notion that motor skills training in isolation results in gains in academic achievement. However, two particular programs are widely accepted by schools as effectively addressing the motor needs of individuals with developmental coordination disorder: adaptive physical education and sensory integration programs. Adaptive physical education programs are basically physical education programs modified to help students with developmental coordination disorder enjoy exercise, recreation, and leisure activities. Such programs usually emphasize inclusion of movement games into the physical education regimen (e.g., serving a volleyball or kicking a football). For this approach to be effective, it is recommended that adaptive physical education programs be available at least two to five times a week, in sessions lasting from 40 minutes to an hour.

Sensory integration programs are usually delivered as a related service to children with developmental coordination disorder under the direction of occupational therapists. These programs prescribe specific physical therapies that attempt to modify the motor and sensory functioning of these children. The entire set of therapies are based on the notion that sensory integration disorders interfere with both body movements and body awareness. Activities are designed to target the tactile, vestibular, and proprioceptive systems. Some children with developmental coordination disorder, for instance, experience “tactile defensiveness” (discomfort when touched by another person). In addressing a disorder affecting the tactile system, occupational therapists might use a sensory integration program that includes touching and rubbing skin surfaces, using creams, and brushing skin surfaces (if working with infants). A sensory integration program that addresses disorders affecting the proprioceptive system (bumping into walls, inability to button, skip, or write) might involve using scooter boards to improve balance and body awareness.

SUGGESTED CROSS-REFERENCES

Mental retardation is discussed in Chapter 34, pervasive developmental disorder in Chapter 38, and attention-deficit disorders in Chapter 39. Reading disorder is covered in Section 35.1, mathematics disorder in Section 35.2, disorder of written expression in Section 35.3, expressive language disorder in Section 37.1, mixed receptive-expressive language disorder in Section 37.2, and phonological disorder in Section 37.3.

CHAPTER REFERENCES

*Ayres A: *Sensory Integration and the Child*. Western Psychological Services, Los Angeles, 1981.

Barnett BE, Merriman WJ: Misconceptions in motor development. Strategies *5:*5, 1991.

Bendersky M, Lewis M: Environmental risk, biological risk, and developmental outcome. Dev Psychol *30:*484, 1994.

Benelli C, Yongue B: Supporting young children's motor skill development. Child Educ *71:*217, 1995.

Blondis TA, Snow JH, Accardo PJ: Integration of soft signs in academically normal and academically at risk children. Pediatrics *85:*421, 1990.

Blumsack J: Neurodevelopmental precursors to learning disabilities: A preliminary report from a parent survey. J Learn Disabil *30:*228, 1997.

Bushnell E, Boudreau JP: Motor development and the mind: The potential role of motor abilities as a determinant of aspects of perceptual development. Child Dev *64:*1005, 1993.

Campbell L: Perceptual-motor programs, movement and young children's needs: Some challenges for teachers. Aust J Early Child *22:*37, 1997.

Corrie L, Barratt-Pugh, C: Perceptual-motor programs do not facilitate development: Why not play? Aust J Early Child *22:*30, 1997.

*Cratty BJ: *Clumsy Child Syndromes: Description, Evaluation, and Remediation*. Harwood Academic Publishers, Langhorne, PA, 1994.

*Dunn JM: *Special Physical Education: Adapted, Individualized, Developmenta*l, ed 7. Brown and Benchmark, Madison, WI, 1997.

Dussart G: Identifying the clumsy child in school: An exploratory study. Br J Special Educ *21:*81, 1994.

Everhart B: Assessing motor and sport skill performance: Two practical procedures. J Phys Educ Recreation Dance *67:*49, 1996.

*Gordon N, McKinlay I: *Helping Clumsy Children*. Churchill Livingstone, New York, 1980.

Jarus T, Loiter Y: The effect of kinesthetic stimulation on acquisition and retention of a gross motor skill. Can J Occup Ther *62:*23, 1995.

*Losse A, Henderson SE, Elliman D, Hall D, Knight E, Jongmans M: Clumsiness in children—do they grow out of it? A ten year follow up study. Dev Med Child Neurol *33:*55, 1991.

Pinheiro VED, Simon HA: An operational model of motor skill diagnosis. J Teach Phys Educ *11:*288, 1992.

Roussounis SH, Gaussen TH, Stratton P: A 2-year follow-up study of children with motor coordination problems identified at school entry age. Child Care Health Dev *13:*377, 1987.

Smits-Engelsman BCM, Van Galen GP: Dysgraphia in children: Lasting psychomotor deficiency or transient developmental delay? J Experimental Child Psychol *67:*164, 1997.

Smyth TR: Clumsiness: Kinesthetic perception and translation. Child Care Health Dev *22:*1, 1996.

*Stagg V, Burns MS: Specific developmental disorders. In *Handbook of Prescriptive Treatments for Children and Adolescents*, ed 2, RT Ammerman, M Hersen, editors. Allyn & Bacon, Boston, 1999.

Taft LT, Barowsky EI: Clumsy child. Pediatr Rev *10:*247, 1989.

Thelen E: Motor development: A new synthesis. Am Psychol *50:*79, 1995.

Wilson PH, McKenzie BE: Information processing deficits associated with developmental coordination disorder: A meta-analysis of research findings. J Child Psychol Psychiatry *39:*829, 1998.

CHAPTER 37. COMMUNICATION DISORDERS

37.1 EXPRESSIVE LANGUAGE DISORDER

CARLA J. JOHNSON PH.D., AND JOSEPH H. BEITCHMAN, M.D.

Spoken communication is fundamental to most human endeavours. A child who is not proficient in communication may be unable to participate fully and competently in many personal, social, academic, and recreational situations. In turn, limited participation in these key life events may jeopardize further growth in communication and related abilities. Effective communication requires mastery of both language, the conventional code used to express ideas about the world, and speech, the complex and rapid motor movements that translate ideas into spoken words. Communication disorders affect language, speech, or both of these key abilities. The fourth edition of *Diagnostic and Statistical Manual of Mental Health Disorders* (DSM-IV) includes two diagnostic categories for language disorders (expressive and mixed receptive-expressive language disorders), two for speech disorders (phonological disorder and stuttering), and one for communication disorders not otherwise specified.

DEFINITION, AND COMPARATIVE NOSOLOGY

Children with expressive language disorders have difficulty communicating their needs, thoughts, and intentions via spoken language. They frequently (1) have speaking vocabularies that are limited in size and variety; (2) use sentences that are short, incomplete, or ungrammatical; and (3) relate stories and events in a disorganized, confusing, or unsophisticated manner. These communication problems are evident despite performance within the normal range on measures of hearing acuity, nonverbal intelligence, and understanding (comprehension, reception) of spoken language. As with many other childhood mental disorders, the definitional and diagnostic issues concerning expressive language disorders are complex and often controversial.

DSM-IV makes a categorical distinction between expressive language disorder, which affects only production of spoken language, and mixed receptive-expressive disorder, which also affects comprehension of spoken language. The distinction accords with common practice and with evidence that the two disorders vary in prognosis and comorbid conditions.

Many experts, however, question whether separate categories are warranted. An alternative view is that expressive language disorder represents a less severe manifestation of the same underlying problems of mixed receptive-expressive disorder. In this view, differences in severity account for the differences in prognosis and comorbidity. The view also accords with evidence that children with expressive language disorder frequently exhibit mild concomitant deficits in language comprehension and typically exhibit expressive language features similar to those shown in receptive-expressive language mixed disorder.

DSM-IV does not include a diagnostic category for receptive disorders in the presence of intact expressive language. The exclusion can be defended on the basis that in normal development, language comprehension usually precedes production. In some cases of pervasive developmental disorder or spina bifida, however, language comprehension skills appear to be weak relative to productive abilities. Further study of these children with possible receptive deficits may help to refine the understanding and diagnosis of language disorders.

Within the category of expressive language disorders, DSM-IV recognizes two types, which are also generally acknowledged in the literature on communication disorders. The developmental type, by far the most common, has no known cause and manifests gradually in delayed onset and below normal rates of expressive language growth. The acquired type, which is quite rare, arises suddenly from a known neurological cause, such as disease or head trauma. For the practitioner, the inclusion of developmental and acquired types under the same DSM-IV diagnostic category may be misleading. Although the two types are similar in certain expressive language manifestations, they vary substantially in etiological factors, prevalence, course, prognosis, treatment, and comorbid conditions.

Another controversial aspect of the DSM-IV scheme is its use of cognitive referencing, the requirement that nonverbal cognitive performance must be substantially higher than language performance for diagnosis of a language disorder (expressive or mixed). For years, cognitive referencing was a widely accepted practice in defining language disorders because it presumably identified children whose language growth was not keeping pace with their nonverbal cognitive development. Recently, cognitive referencing has been challenged on various theoretical and psychometric grounds. A key theoretical concern is that nonverbal cognitive functioning does not necessarily limit a child's potential for language growth, as is implied by cognitive referencing. For instance, children with little or no discrepancy between nonverbal and verbal scores may benefit from language intervention as much or more than those with a large discrepancy between the two. A related psychometric concern is that a discrepancy in performance on language and cognitive tests may be spurious if the effects of measurement error are not adequately considered, as is often the case in diagnostic practice. Further theoretical and empirical efforts are needed to resolve these issues.

The DSM-IV classification scheme is similar, but not identical, to other well-known diagnostic schemes, such as that of its predecessor, the revised third edition of DSM (DSM-III-R) and the 10th revision of the *International Statistical Classification of Disease and Related Health Problems* (ICD-10). DSM-III-R included separate expressive and receptive categories, but the expressive category could be coded either alone or as an associated feature of the receptive disorder. In DSM-IV and ICD-10, the categories of expressive and mixed (receptive) disorders are mutually exclusive. DSM-III-R and ICD-10 consider only developmental language disorders. The mention of acquired types is unique to DSM-IV. A number of other schemes for classification or subtyping of childhood language disorders are also elaborated in the scientific literature. The result is a diverse set of terms that refer to types and subtypes of childhood language disorders (developmental dysphasia, auditory verbal agnosia, semantic-pragmatic disorder), many of which do not correspond in a straightforward manner to the DSM-IV categories.

EPIDEMIOLOGY

Developmental Type Prevalence estimates for developmental language disorders (combined expressive and mixed receptive-expressive) range from 3 to 13 percent of children, depending on the age of the children surveyed, the nature of the sample tested (community or clinic-referred), and the specific language measures and definitional criteria used. DSM-IV suggests that the expressive form of developmental language disorder has a prevalence rate of 3 to 5 percent in school age children, which is generally consistent with results from the few studies that have separately identified expressive disorders.

Boys are more likely than girls to have developmental language disorders, including expressive language disorder. In some studies, boys outnumbered girls by ratios as high as 5 to 1. In others, the sex imbalance was less pronounced, with ratios of 2 to 1 or even less observed. The discrepancies likely reflect differences in sampling practices, testing procedures, and definitional criteria.

Evidence is mounting in support of familial aggregation of developmental language disorders, but there are few relevant data specific to expressive language disorder. One research group found no evidence of familial aggregation in a sample of 2-year-olds with expressive delays (but not necessarily disorders). A study of older children yielded higher rates of familial aggregation for children with expressive disorders than for those with mixed receptive-expressive language disorder. Further work is needed to establish whether expressive language disorder aggregates in families and, if so, to identify specific genetic and environmental contributions to such familial effects.

Acquired Type Acquired language disorders are rare in children, with those specific to expressive language rarer still. Prevalence rates, sex ratios, and familial aggregation data are available for certain specific causes (e.g., stroke, head injury) but not for the heterogeneous category of acquired expressive language disorder.

ETIOLOGY

Developmental Type The causes of developmental language disorders are unknown and most probably heterogeneous. Theoretical and empirical etiological efforts typically have focused on developmental language disorders in general, not on expressive language disorder in particular. A variety of biological and environmental risk factors have been identified, but none have been confirmed as causal agents.

Acquired Type Acquired language disorders arise suddenly, at any age, as a result of known neurological conditions such as stroke, head injury, disease, poisoning, or seizure disorder. At their onset, many acquired disorders affect both receptive and expressive language. Following a period of spontaneous recovery, language problems may partially resolve to be primarily expressive in nature. Selective expressive deficits also result on occasion from a focal injury to the left hemisphere, for example, from a gun shot wound or stroke.

DIAGNOSIS AND CLINICAL FEATURES

Expressive language disorder is diagnosed when a selective deficit in expressive language development occurs in the presence of intact nonverbal intelligence and receptive language skills (Table 37.1-1). Children with expressive disorders do acquire language, albeit at a slower rate than their peers. Accordingly, the specific manifestations of expressive language disorder change with maturation. Affected children often show expressive language characteristics that resemble those of younger children who are developing language at a normal rate. Table 37.1-2 provides an overview of typical milestones in language and nonverbal development. Although milestones are listed only up to age 8, it is commonly recognized that substantial growth in vocabulary, grammar, and language use continues through adolescence.

A. The scores obtained from standardized individually administered measures of expressive language development are substantially below those obtained from standardized measures of both nonverbal intellectual capacity and receptive language development. The disturbance may be manifest clinically by symptoms that include having a markedly limited vocabulary, making errors in tense, or having difficulty recalling words or producing sentences with developmentally appropriate length or complexity.
B. The difficulties with expressive language interfere with academic or occupational achievement or with social communication.
C. Criteria are not met for mixed receptive-expressive language disorder or a pervasive developmental disorder.
D. If mental retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

Coding note: if a speech—motor or sensory deficit or a neurological condition is present, code the condition on Axis III.

Table 37.1-1 DSM-IV Diagnostic Criteria for Expressive Language Disorder

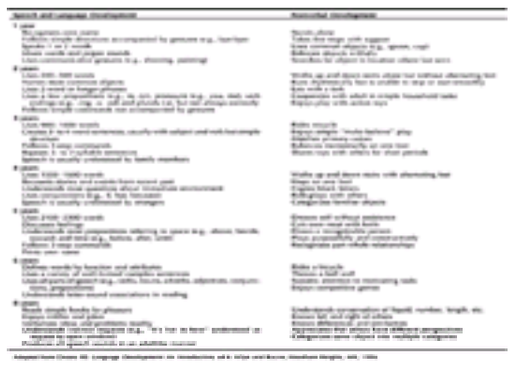

Table 37.1-2 Normal Development of Speech, Language, and Nonverbal Skills in Children

Developmental Type Expressive disorders may be observed to varying degrees in one or more aspects of the content, form, or use of language. Difficulties with content may be evident in a small vocabulary and limited ability to express complex or abstract ideas. Problems with the form of language may be manifest in short, incomplete, and ungrammatical sentences or in marked delays in learning the correct use of certain grammatical forms, such as pronouns (e.g., *he, she, they*), auxiliary or modal verbs (e.g., *is, were, could*), plurals (e.g., *cats, geese*), and verb tense endings (e.g., *-ing, -s, -ed*). Restrictions in the use of language may limit the expression of a range of communicative purposes (e.g., question, explain, comment) or the ability to maintain and extend conversational topics. Particular difficulties may arise in connected discourse, for instance, in telling stories, relating events, or giving explanations. The discourse may be uninformative, poorly organized, and difficult to follow.

Miguel was an alert, energetic 2-year-old with an expressive vocabulary that consisted of only five words (mama, daddy, bottle, up, bye-bye). He communicated by using these words one at a time or by pointing or using other simple gestures. He typically used his words and gestures to request desired objects or actions, but rarely used them for other communicative purposes (e.g., commenting, protesting). Except for his limited language skills, he appeared to be developing normally in other areas. He sat, stood, and walked at the expected times. He enjoyed the company of other children. His play skills and preferences were comparable to those of other 2-year-olds. He had a history of frequent ear infections, but a recent hearing test revealed normal hearing. Comprehension testing, carefully conducted to ensure his attention and motivation, indicated that he understood the names for most familiar objects and actions. He also had no difficulty following the types of simple verbal instructions that are understood by most children his age (e.g., Show me your eyes. Get your shoes. Say “bye-bye.”).

Although Miguel was clearly developing language more slowly than most children his age, it was not yet clear that he had an expressive language disorder. Recent prospective research on the development of children like Miguel, often termed *late talkers*, suggests that a significant percentage of them will spontaneously overcome their initial slow start in language development.

Jennifer was a sociable, active 5-year-old who was diagnosed with expressive language disorder. She often played with her best friend, Sarah. One day, in the course of pretend play, each girl told the story of Little Red Riding Hood to her doll. Sarah's story began: “Little Red Riding Hood was taking a basket of food to her grandmother who was sick. A bad wolf stopped Riding Hood in the forest. He tried to get the basket away from her but she wouldn't give it to him.”

By contrast, Jennifer's story illustrated her marked difficulties in verbal expression: “Riding Hood going to grandma house. Her taking food. Bad wolf in a bed. Riding Hood say, what big ears, grandma? Hear you, dear. What big eyes, grandma? See you, dear. What big mouth, grandma? Eat you all up!”

Many features of Jennifer's story were characteristic of children with expressive language disorder, including the short, incomplete sentences, the simple sentence structures, the omission of grammatical function words (e.g., *is, the*) and endings (e.g., possessive -'*s*, present tense verb -*s*), the problems in question formation, and the incorrect use of pronouns (e.g., “her” for *she*). Nonetheless, when tested by methods that did not require verbal responding, it was clear that Jennifer understood the details and plot of the Riding Hood tale as well as Sarah did. Jennifer also demonstrated adequate comprehension skills in her kindergarten classroom, where she readily followed the teacher's complex, multistep verbal instructions (e.g., “Before you get ready for recess, make sure that you draw a green circle around all the animals, put your library books under your chair, and line up at the back of the room.”).

Sean was a quiet, sullen 8-year-old whose expressive language problems were no longer obvious in casual, social conversations. He now rarely omitted words or made grammatical errors as he often did when he was younger. His expressive difficulties, however, were still apparent in tasks that require more elaborate or reflective use of language, including many that were part of his third-grade academic work. Take, for example, his explanation of the outcome of a simple science demonstration: “The teacher had stuff in some jars. He poured it and it got pink. The other thing made it white.” Although each sentence was grammatical by itself, the entire explanation was difficult to follow, as important ideas and details were omitted or poorly explained. Sean also showed problems in word finding, that is, in using specific words for the concepts and actions he was describing. Instead, he relied on vague and nonspecific terms such as *thing, stuff*, and *got*. Similar problems were evident in his written work. In earlier grades, Sean kept up with his classmates fairly well in reading and other activities, but the increased demands for written work in third grade negatively affected his overall academic standing. He also began to experience some teasing from classmates, to which he reacted quite aggressively, sometimes to the point of fighting. Nonetheless, he continued to show relatively good comprehension of spoken language, as evidenced by his ability to follow classroom lectures and to understand sentences that were grammatically and conceptually complex (e.g., “The boy the girl hit had a shirt with an abstract design; Had it been possible, they would have traveled in a van or in a bus.”).

It is relatively common for other disorders of speech, learning, development, and behavior to occur with expressive disorders. In preschoolers, the most common concomitant disorder is phonological disorder, which, if severe, may persist into the early school years. Problems in acquisition of written language may also be diagnosed at school age. Reading disorder is the most common of these, but disorder of written expression may also occur. Some children also show developmental coordination disorder.

Concomitant psychiatric disorders may also occur in as many as half of the children with persistent expressive language disorders. By far the most common associated psychiatric diagnoses are attention-deficit disorders, but externalizing or internalizing disorders may also be observed.

Most children with developmental expressive disorders have no overt neurological symptoms (e.g., electroencephalographic [EEG] abnormalities, seizures, dysarthria, or dyspraxia). The presence of neurological symptoms does not, however, preclude diagnosis of developmental expressive disorder, unless their onset is attributable to a known, acquired brain insult.

Acquired Type Acquired expressive language disorder may manifest somewhat differently than developmental disorders. The most common difference (after initial recovery stages) is that expressive grammar skills (language form) may be relatively preserved, resulting in sentences that resemble those of age-mates in length and grammatical complexity. Problems in word finding and discourse organization are common in acquired disorders but may be qualitatively different than the word-finding and discourse difficulties shown in developmental disorders. Qualitative differences in symptomatology may also be observed across different etiological factors for acquired language disorders.

Louise developed normally until age 10, when she had a stroke that caused focal left hemisphere damage. Immediately following the stroke, Louise could not speak and showed some comprehension difficulties. She also experienced hemiparesis, a partial paralysis on the right side of her body. In the weeks following the stroke, Louise quickly regained her language comprehension skills. Her expressive language skills recovered more slowly. When she began to speak again, she used short, telegraphic sentences with few grammatical function words or endings. Her speech, mildly dysarthric, was slow, labored, and slurred. Gradually the complexity of her sentences and the rate and precision of her speech returned to normal. Six months after her stroke, the hemiparesis had also resolved. Louise continued, however, to show subtle residual expressive problems, particularly in word finding during discourse. She often inadvertently replaced words that she intended to say with others that were closely related in meaning (e.g., “broccoli” for *cauliflower*) or sound (e.g., *symphony/sympathy*). Despite relatively complete recovery of her language skills, Louise experienced residual memory and attention problems that interfered with her academic performance in fifth grade. Louise was a competent student prior to her stroke but now became easily frustrated with her changed and inconsistent performance.

As in Louise's case, neurological insults may also affect cognitive, affective, gross motor, and speech motor skills, either temporarily or permanently. Many children with acquired disorders recover sufficiently to score within the normal range on structured intelligence tests but may not function adequately in less structured settings, such as school or home. Not surprisingly, overt neurological symptoms are more frequently associated with acquired disorders than with developmental ones. Concomitant conditions may include speech disorders, learning disorders, and (perhaps) behavioral and emotional difficulties severe enough to warrant psychiatric diagnosis.

DIFFERENTIAL DIAGNOSIS

Differential diagnosis of developmental expressive language disorder requires standardized evaluations of expressive language, receptive language, and nonverbal intellectual functioning (Table 37.1-1 and Table 37.1-2). Expressive language development must fall substantially below (1) the range of normal expressive performance expected for a child's age, (2) receptive language performance, and (3) nonverbal intellectual performance. Further, the expressive language difficulties must be severe enough to impair academic performance or social communication. To help determine whether this clinical severity criterion is met, standardized testing is often supplemented with observational techniques and analysis of spontaneous language use.

Substantial deficits in either receptive language or nonverbal intelligence preclude the diagnosis of expressive language disorder but may contribute to the diagnosis of other disorders that affect language growth. If both expressive and receptive deficits occur in the absence of nonverbal deficits, the diagnosis of mixed receptive-expressive language disorder is appropriate. If language and nonverbal functioning are both substantially below age level expectations, the diagnosis of mental retardation should be made.

A hearing test should also be a regular part of the diagnostic procedure when expressive language disorder is suspected. Many children with expressive problems experience regular ear infections or allergies, which may be accompanied by mild hearing loss of a transient or persistent nature. Medical treatment and monitoring of such conditions is warranted. A hearing test may also identify cases of language delay associated with more severe hearing impairments, in which hearing aids or other amplification devices may assist in language learning. If expressive language deficits exceed those expected for a given degree of hearing impairment or mental retardation, a concurrent diagnosis of expressive language disorder may be made.

Language delay is also associated with environmental deprivation, pervasive developmental disorders, and selective mutism. A careful history will help identify environmental deprivation. Observations of disturbed social interaction and communicative intent, as well as unusual repetitive behaviors, will suggest pervasive developmental disorder. Documentation of relatively normal expressive language use in some settings (e.g., home) but refusal to speak in others (e.g., school) indicates selective mutism.

Children from diverse linguistic and cultural backgrounds may pose a particular diagnostic challenge because it may be difficult to decide whether particular language characteristics are normal in the acquisition of a new language or reflect a disorder. Cultural and linguistic barriers may also make it difficult to communicate, not only with the child, but also with other family members for the purposes of collecting case history data and conveying assessment results. Nonetheless, it is usually possible to conduct at least a preliminary language assessment. If inconclusive results are obtained, the child can be monitored periodically to assess language progress.

The hallmark of the acquired type of expressive language disorder is a history of sudden onset of language symptoms associated with known neurological insult. Brain scans are often used to verify the existence and extent of brain lesions.

COURSE AND PROGNOSIS

Developmental Type Developmental expressive language disorder is characterized by considerable variability in severity, course, and outcome. Recently several research teams conducted prospective studies of late talkers (i.e., children with normal cognitive functioning who use fewer than 50 words and no word combinations at age 2). Experts disagree on whether late talkers actually meet criteria for the diagnosis of expressive language disorder. Nonetheless, their language delays often provoke parental concern and professional referral. The various studies of late talkers generally agree that 50 to 80 percent of these children achieve language skills within the normal range during the preschool years, with vocabulary skills perhaps recovering more fully than grammatical abilities. Most late talkers who recover during preschool appear to be at little risk for severe learning and behavioral problems, at least through the early school years. Nonverbal intelligence scores, receptive language levels, the extent of initial expressive delay, and the amount and variety of early communicative intentions have been suggested as possible

predictors of outcome in late talkers, but these suggestions require further empirical validation.

Prognosis is generally less favorable for children whose expressive language disorders persist into the late preschool or early school-age years. For these children, language growth proceeds but at a slower rate than for children who are developing typically. Children with persistent expressive disorders may also experience associated problems, particularly reading disorder and attention-deficit disorders. By adolescence, most children with expressive language disorder acquire sufficient language skill to function reasonably well in most daily communication activities. Subtle residual deficits may still be apparent in more demanding speaking tasks, such as those that require precise vocabulary and complex explanations. As a group, however, children with persistent expressive disorders show more favorable long-term outcomes than their counterparts with mixed receptive-expressive disorder.

Acquired Type Approximately two thirds of children with acquired language disorders recover to have language functioning that is broadly within the normal range, although subtle deficits may persist. Recovery may occur rapidly within days or weeks of the neurological insult or may be more protracted, occurring over months or even years. Persistent language and cognitive deficits of varying severity are seen in those who do not recover normal language functioning. Prognosis depends on a variety of factors, including the specific etiology for the disorder, the severity, extent, and location of brain damage, the age at injury, the extent of preinjury language development, and the existence of other preinjury disorders.

TREATMENT

Developmental Type Experts disagree on when intervention is warranted for expressive language disorder. Some suggest that a watch-and-see attitude is appropriate for young children with early expressive delays (late talkers) because most will acquire language functioning within the normal range during preschool. Intervention to improve expressive language would only be provided for children whose problems persist to age 4 or 5. Others argue that this policy would deny children early help that might serve to prevent or minimize later language, academic, and behavioral difficulties. Unfortunately, no relevant data are available on the efficacy of early versus late language intervention.

Research on the efficacy of language intervention is, in fact, rather sparse. Most available efficacy studies have targeted children with developmental language disorders, not just those with expressive disorders. Nonetheless, a variety of techniques have been shown experimentally to result in improved use of specific expressive language features (e.g., pronouns, question forms, verb tense markers, complex sentence forms, narrative coherence) during intervention. What is not yet clear is whether such intervention can close or narrow the gap in language growth on a long-term basis and so prevent or ameliorate the adverse academic and psychosocial consequences sometimes associated with persistent language disorders.

Another issue concerns whether intervention should be direct or mediated. In direct interventions, a professional (usually a speech-language pathologist) serves as the primary intervention agent and works directly with the child. In mediated interventions, the professional teaches others (e.g., parents, teachers, paraprofessionals) to be the primary intervention agents. Mediated interventions are thought to make efficient use of professional time and to allow mediators to incorporate language facilitation techniques into daily activities and routines, thereby providing regular, naturalistic opportunities for children to use newly learned language features. Several recent studies indicated that both direct and mediated (parent) interventions can be effective in promoting language growth in children with developmental language disorders (expressive and mixed), at least in the short term. Late talkers were the target population in another recent randomized control study that also demonstrated that parents could be taught to use specific techniques to effectively promote short-term language growth.

Additional systematic, well-controlled research on treatment efficacy is clearly needed. Although the limited literature to date supports the possible benefits of intervention, it provides little guidance on whether or how interventions should be tailored to address the specific language needs of individual children.

Acquired Type Little if any research has systematically evaluated the efficacy of language intervention for acquired expressive disorders in children. Recommendations are available in the literature, however, concerning how to facilitate the integration of children who have suffered brain insults back into their home and school environments. Many of these suggestions address cognitive, motor, and emotional sequelae of brain damage, but others deal with common expressive language symptoms, such as word-finding and discourse organization problems. Recommendations often incorporate techniques devised for adults with acquired language disorders, some of which have been shown to be effective with that population.

SUGGESTED CROSS-REFERENCES

Section 37.2 describes mixed receptive-expressive disorder, which is similar in many respects to expressive language disorder. Another closely related disorder is reading disorder, which is discussed in Section 35.1. Several other disorders are relevant to the differential diagnosis of expressive language disorder, including mental retardation (Chapter 34), pervasive developmental disorder (Chapter 38), selective mutism (Section 46.4). Other disorders may occur with expressive language disorder, including mathematics disorder (Section 35.2), disorder of written expression (Section 35.3), developmental coordination disorder (Chapter 36), phonological disorder (Section 37.3), stuttering (Section 37.4), and attention-deficit disorders (Chapter 39).

SECTION REFERENCES

Beitchman JH, Nair R, Clegg M, Ferguson B, Patel PG: Prevalence of psychiatric disorders in children with speech and language disorders. J Am Acad Child Psychiatry *25:*528, 1986.

Beitchman JH, Nair R, Clegg M, Patel PG: Prevalence of speech and language disorders in 5-year-old children in the Ottawa-Carleton region. J Speech Hear Disord *51:*98, 1986.

*Beitchman JH, Wild J, Kroll R: An overview of childhood speech and language disorder. In *Handbook of Child and Adolescent Psychiatry*, vol 4, JD Noshpitz, NE Alessi, editors. Wiley, New York, 1997.

Bishop DVM: Language development after focal brain damage. In *Language Development in Exceptional Circumstances*, D Bishop, K Mogford, editors. Erlbaum, Hillsdale, NJ, 1992.

Bishop DVM, Adams C: A prospective study of the relationship between specific language impairments, phonological disorders, and reading retardations. J Child Psychol Psychiatry *31:*1027, 1990.

Cantwell DP, Baker L: *Psychiatric and Developmental Disorders in Children with Communication Disorder*. American Psychiatric Press, Washington, DC, 1991.

*Cole KN, Coggins TE, Vanderstoep C: The influence of language/cognitive profile on discourse intervention outcome. Lang Speech Hear Serv Schools *30:*61, 1999.

Dennis M: Word finding in children and adolescents with a history of brain injury. Top Lang Disord *13:*66, 1992.

Eiserman W, Weber C, McCoun M: Parent and professional roles in early intervention: A longitudinal comparison of the effects of two intervention configurations. J Spec Educ *29:*20, 1995.

Enderby P, Emerson J: Children with speech and language disorders. In *Does Speech and Language Therapy Work? A Review of the Literature*, P Enderby, J Emerson, editors. Whurr, London, 1995.

Fey M, Cleave P, Long S: Two models of grammar facilitation in children with language impairments: Phase 2. J Speech Lang Hear Res *40:*5, 1997.

Fey ME, Cleave PL, Long SH, Hughes DL: Two approaches to the facilitation of grammar in language-impaired children: An experimental evaluation. J Speech Hear Res *36:*141, 1993.

Francis D, Fletcher J, Shaywitz B, Shaywitz S, Rourke B: Defining learning and language disabilities: Conceptual and psychometric issues. Lang Speech Hear Serv Sch *27:*132, 1996.

Gilger JW: Behavioral genetics: Concepts for research and practice in language development and disorders. J Speech Hear Res *38:*1126, 1995.

Girolametto L, Pearce P, Weitzman E: Interactive focused stimulation for toddlers with expressive vocabulary delays. J Speech Hear Res *39:*1274, 1996.

Lahey M, Edwards J: Specific language impairment: Preliminary investigation of factors associated with family history and with patterns of language performance. J Speech Hear Res *38:*643, 1995.

*Lahey M, Edwards J: Naming errors of children with specific language impairment. J Speech Lang Hear Res *42:*195, 1999.

*Leonard LB: *Children with Specific Language Impairment*. Bradford, Cambridge, MA, 1998.

Leonard LB: Children with specific language impairment (developmental dysphasia): Treatment. In *Linguistic Disorders and Pathologies: An International Handbook*, G Blanken, J Dittmann, H Grimm, J Marshall, C Wallesch, editors. DeGruyter, New York, 1993.

*Nelson NW: *Childhood Language Disorders in Context: Infancy through Adolescence*, ed 2. Allyn and Bacon, Boston, MA, 1998.

*Owens RE: *Language Development: An Introduction*, ed 4. Allyn and Bacon, Needham Heights, MA, 1996.

Paul R: Clinical implications of the natural history of slow expressive language development. Am J Speech Lang Pathol *5:*5, 1996.

Paul R, Murray C, Clancy K, Andrews, D: Reading and metaphonological outcomes in late talkers. J Speech Lang Hear Res *40:*1037, 1997.

Rescorla L, Ratner N: Phonetic profiles of toddlers with specific expressive language impairment (SLI-E). J Speech Hear Res *39:*153, 1996.

Rescorla L, Roberts J, Dahlsgaard K: Late-talkers at 2: Outcome at age 3. J Speech Lang Hear Res *40:*556, 1997.

Roberts J, Rescorla L, Giroux J, Stevens L: Phonological skills of children with Specific Expressive Language Impairment (SLI-E): Outcome at age 3. J Speech Lang Hear Res *41:*374, 1998.

Russell N: Educational considerations in traumatic brain injury: The role of the speech-language pathologist. Lang Speech Hear Services Schools *24:*67, 1993.

Stothard S, Snowling M, Bishop DVM, Chipchase B, Kaplan C: Language-impaired preschoolers: A follow-up into adolescence. J Speech Lang Hear Res, *41:*407, 1998.

Thal D, Tobias S: Relationships between language and gesture in normally developing and late-talking toddlers. J Speech Hear Res *37:*157, 1994.

Tomblin JB, Records N, Buckwalter P, Zhang X, Smith E, O'Brien M: Prevalence of specific language impairment in kindergarten children. J Speech Lang Hear Res *40:*1245, 1997.

*Turkstra L: Language testing in adolescents with brain injury: A consideration of the CELF-3. Lang Speech Hear Serv Schools *30:*132, 1999.

Whitehurst G, Fischel J: Practitioner review: Early developmental language delay: What, if anything, should the clinician do about it? J Child Psychol Psychiatry *35:*613, 1994.

37.2 MIXED RECEPTIVE-EXPRESSIVE LANGUAGE DISORDER

CARLA J. JOHNSON PH.D., AND JOSEPH H. BEITCHMAN, M.D.

DEFINITION, AND COMPARATIVE NOSOLOGY

Children with mixed receptive-expressive language disorders, like their peers with expressive language disorders, characteristically have small speaking vocabularies, use unsophisticated sentences, and produce disorganized stories and descriptions. More importantly, however, they also have problems in understanding (reception, comprehension) of spoken language. Children with receptive problems may have difficulty learning new words, particularly those with uncommon or abstract meanings. They may misinterpret conversational questions or comments, leading them to make inappropriate or nonsensical remarks. They may fail to understand lengthy or grammatically complex directions and explanations. Such comprehension failures may not be recognized and may instead be attributed incorrectly to inattention, stubbornness, or other behavioral problems.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) definition of mixed receptive-expressive language disorder is similar in many ways to that for expressive language disorder. Both mixed receptive-expressive and expressive disorders are defined by language difficulties in the presence of normal levels of nonverbal intelligence. Both mixed receptive-expressive and expressive language disorders have developmental or acquired subtypes. The developmental forms have no known causes and are more prevalent; the acquired forms result from known neurological conditions and are less common. Not surprisingly, many of the problematic definitional and diagnostic issues that arise with respect to expressive language disorder also apply to mixed receptive-expressive language disorder.

EPIDEMIOLOGY

Developmental Type The prevalence of mixed receptive-expressive disorder has not been established definitively. Study methodologies and definitions have varied greatly, resulting in a wide range (3 to 13 percent) of prevalence estimates for developmental language disorder (all types). Moreover, some studies have counted children with mental retardation, hearing impairment, and pervasive developmental disorder as language disordered; others have excluded such children. DSM-IV suggests that the mixed receptive-expressive form of developmental language disorder has a prevalence rate of 3 percent in school-age children and may be less common than expressive language disorder. These suggestions, however, await empirical verification, as no existing prevalence study has specifically used the DSM-IV criteria for expressive and mixed receptive-expressive language disorders.

Boys are more likely than girls to have developmental language disorders, including mixed receptive-expressive language disorders. Sex ratios as high as 4 to 1 and as low as 2 to 1 have been reported across various studies.

Familial aggregation is also characteristic of developmental language disorders, but there are few available data specific to mixed receptive-expressive language disorders. Work continues to identify relevant disorder subtypes and to discover the possible genetic and environmental factors that underlie familial aggregation.

Acquired Type Acquired language disorders are infrequent in children. Both receptive and expressive language are usually affected initially. Spontaneous recovery of function often occurs to varying degrees, but residual deficits in both receptive and expressive language may persist. Prevalence rates, sex ratios, and familial aggregation data are not available for the heterogenous category of acquired mixed receptive-expressive language disorders.

ETIOLOGY

Developmental Type The cause of mixed receptive-expressive disorder is unknown. Several correlates of developmental language disorder have been identified, but it is not yet clear whether these factors are causative and, if so, how they operate. A plausible working hypothesis, given current evidence, is that language disorders are multiply determined, with both biological and environmental factors playing a role. The heterogeneity of the population is also consistent with the possibility that certain factors or combinations of factors may be critical in some cases, while other factors or combinations may be important in others. Attempts to understand causation have focused on seven general possibilities: (1) brain abnormalities, (2) genetic factors, (3) environmental influences, (4) neurodevelopmental immaturity, (5) general cognitive factors, (6) specific linguistic factors, and (7) auditory processes. The possibilities are not mutually exclusive, and each enjoys some empirical support. None, however, currently accounts for all cases of language disorder.

Structural or functional brain abnormalities have long been considered possible causative factors because the symptoms of developmental language disorders resemble those of the acquired language disorders that result from known neurological insults in both children and adults. Children with developmental language disorders typically do not show hard neurological signs indicating brain damage (e.g., paralysis, seizures). As a group, however, they often show elevated rates of various soft neurological signs (e.g., mixed cerebral dominance, attention problems, subtle motor difficulties). Neuroimaging studies also indicate that some children with language and learning disorders have abnormal neuroanatomical characteristics, particularly in areas of the left hemisphere known to be crucial for language processing. Nonetheless, many questions remain. The brain abnormalities identified thus far do not occur in all children with developmental language disorders, nor has their functional significance been established.

Genetic factors are also implicated in some (but not all) cases of developmental language disorder. Familial aggregation of language disorders is commonly reported, but possible modes of genetic transmission have not yet been ascertained. The fact that boys are affected more often than girls may also reflect genetic factors, although other influences may contribute.

Environmental determinants may be critical in certain cases and of limited importance in others. Developmental language disorders are more common in children from families with low socioeconomic status, as indexed by family income, parental occupational status, or parental education level. In turn, low socioeconomic status is correlated with other known risks for developmental language disorders, such as single-parent family status, large family size, pregnancy and birth complications, and abuse and neglect. Children from low socioeconomic families are also at risk for malnutrition, health problems, increased exposure to known toxins (e.g., lead), and limited educational and social opportunities. It is not yet clear whether or how these various factors may interact with each other and perhaps with biological factors to cause developmental language disorders.

Another view is that neurodevelopmental immaturity underlies developmental language disorders. Proponents point to evidence that language development is delayed, not deviant, and generally follows a normal sequence and course but at a slowed rate. Slow myelinization of neural pathways or delayed maturation of other neural structures might also account for other characteristics often seen in children with language disorders, including subtle fine motor and oral motor problems, behavioral difficulties, and soft neurological signs. Boys may be more likely to evidence neurodevelopmental immaturities than girls. A related view holds that children with developmental language disorder are not disordered at all, but simply fall at the lower end of the normal distribution of language-related skills. Because language skills are so crucial for success in our literate society, however, these children are perceived as handicapped. Children who fall in the lower end of the normal distribution on other abilities, for example, musical talent, are not perceived as handicapped because musical proficiency is not expected of every member of society.

General cognitive factors have also been suggested as candidates for the core deficit in developmental language disorders. One such factor is speed of cognitive processing. Children with language disorders often perform more slowly than children without disorders on a variety of linguistic (e.g., naming) and nonlinguistic (e.g.,

fine motor) tasks. Problems in speed of processing may disrupt the language learning process. Slow processing could also reflect neurodevelopmental immaturity. Another cognitive factor implicated in language disorder is phonological working memory, the skill required for immediate processing and maintenance of verbal information. Deficits in phonological working memory have been linked to problems in learning new vocabulary items and in developing proficient reading skills, two areas of deficit for many children with language disorder. It is not clear, however, that phonological memory deficits can account for all characteristics of language disorder.

Specific linguistic factors have also been investigated in relation to developmental language disorders. Some theorists argue that children with language disorders cannot fully exploit the regularities in the language that they hear to induce its underlying grammatical rules. Accordingly, children may have difficulty in acquiring sets of grammatical features that are thought to share an underlying linguistic relationship (e.g., articles *a, an, the*; determiners *this, that, these, those*; possessives, both noun and pronoun forms). Certain parameters of grammar are thought to be especially problematic for children with language disorders to induce. Theories of this sort, however, are primarily descriptive rather than explanatory at this stage in their development.

Finally, auditory problems have been studied as possible causes of developmental language disorders. One type of auditory deficit is the temporary, mild hearing loss that accompanies inflammation of the middle ear (otitis media), which may result from ear infections or allergies. Many young children suffer repeatedly from otitis media and associated hearing loss during their early years when language skills are developing. Current research suggests that otitis media alone does not suffice to cause language disorder but that it may interact, in as yet unknown ways, with other risk factors. A second type of auditory problem involves perception of sounds. Children with language disorders often show deficits in perceiving small sound differences and rapid auditory transitions, aspects of the acoustic signal that are critical to understanding speech. As with other proposed causal factors, however, neither type of auditory difficulty appears to occur in all children with developmental language disorder.

Acquired Type Acquired mixed receptive-expressive language disorder may result from a variety of neurological insults, including (1) closed head injuries resulting from motor vehicle accidents, falls, or other blows to the head; (2) open head injuries such as those from gunshot wounds; (3) cerebral vascular accidents such as strokes; (4) infections such as encephalitis or meningitis; (5) severe respiratory problems that may lead to intraventricular hemorrhage or anoxia; (6) drownings or near-drownings; (7) toxins such as lead or household poisons; (8) damage from tumors or radiation treatment; or (9) seizure activity. Landau-Kleffner syndrome (acquired aphasia with epilepsy) is a rare condition characterized by a sudden, marked deterioration of language skills accompanied by the onset of seizures or abnormal electroencephalographic (EEG) activity.

DIAGNOSIS AND CLINICAL FEATURES

Mixed receptive-expressive language disorder is diagnosed when selective deficits in receptive and expressive language development occur in the presence of intact nonverbal intelligence (Table 37.2-1). Children with the disorder do acquire language, but they do so at a slower rate than their peers. Thus, the specific symptoms of mixed receptive-expressive language disorder change with age. Affected children often show receptive and expressive language characteristics that resemble those of younger children who are developing language at a normal rate. The 10th revision of *International Classifications of Diseases and Related Health Problems* (ICD-10) provides diagnostic criteria for receptive language disorder and acquired aphasia with epilepsy (see Table 37.5-2).

A. The scores obtained from a battery of standardized individually administered measures of both receptive and expressive language development are substantially below those obtained from standardized measures of nonverbal intellectual capacity. Symptoms include those for expressive language disorder as well as difficulty understanding words, sentences, or specific types of words, such as spatial terms.
B. The difficulties with receptive and expressive language significantly interfere with academic or occupational achievement or with social communication.
C. Criteria are not met for a pervasive developmental disorder.
D. If mental retardation, a speech-motor or sensory deficit, or environmental deprivation is present, the language difficulties are in excess of those usually associated with these problems.

Coding note: If a speech—motor or sensory deficit or a neurological condition is present, code the condition on Axis III.

Table 37.2-1 DSM-IV Diagnostic Criteria for Mixed Receptive-Expressive Language Disorder

Developmental Type Mixed receptive-expressive language disorder may be observed to varying degrees in one or more aspects of the content, form, or use of language. The expressive manifestations of the disorder are usually similar to those described for expressive language disorder. The receptive difficulties typically have adverse effects on social interaction and academic performance. Receptive problems may be evident in tasks that tap comprehension of single words, sentences, and larger units of discourse. Single words that refer to common observable objects, actions, or relations may be adequately understood (e.g., bowl, key, pull, sleep, on, under), but those that have less-common, less-observable meanings may not be well known (e.g., liquid, creation, theorize, obtain, among, if). Likewise, sentences with simple grammatical structures may be comprehended (e.g., The rabbit is hopping, The car stopped), while those with more complex syntax may be misinterpreted (e.g., The boy who chased the girl was no longer my friend; If I were to take this trip, I would need the approval of my manager). Lengthy and complex instructions, questions, and explanations may be misunderstood. Children with mixed disorders may make inappropriate comments or contributions to the ongoing conversation because they fail to correctly interpret the language of others. Partial or incomplete understanding of the language of others may also contribute to confusion or inaccuracy in telling stories, relating events, and giving explanations. Expressive language limitations may further exacerbate these problems. Children with mixed receptive-expressive language disorders are sometimes thought to be stubborn or uncooperative because their comprehension difficulties are not recognized.

Maria was a quiet, even-tempered 2-year-old, who had not yet begun to use any recognizable words. She communicated with vocalizations and simple gestures (e.g., pointing, showing) similar to those typically used by younger children. She pointed, on request, to a few familiar people and objects in the environment (e.g., mommy, daddy, dog, shoe, cup). Compared with other children her age, however, she understood few words and rarely responded correctly to simple verbal commands (e.g., Get your book, Throw the shirt). Nonetheless, she had normal hearing acuity, had reached early motor milestones at the expected ages, and enjoyed simple symbolic play activities commensurate with those expected for her age (e.g., pretending to fall asleep, pretending to feed her doll).

Trang was a shy, reserved 5-year-old who grew up in a bilingual home, where his parents and older siblings spoke both English and Vietnamese proficiently. Trang began to understand and speak both languages much later than his siblings had, and he developed slowly in both. When he entered kindergarten, assessment revealed difficulties in both comprehension and production of English. Compared with other children his age, Trang understood a limited number of words for objects, actions, and relations. He often failed to follow classroom instructions, particularly those that involved words for concepts of time (e.g., yesterday, after, week) and space (e.g., beneath, in front of, around). Trang also had difficulty identifying which of several pictures corresponded to a syntactically complex sentence he had heard (e.g., It was not the bus she was waiting for; Because he had already eaten his lunch, he was not kept in the cafeteria). His conversations with other children often broke down because he did not understand fully what they were saying, nor could he express his own ideas clearly. As a result he was not a favored playmate, and most of the children in his class ignored him. His limited interactions further reduced Trang's opportunities for improving and practicing his already weak language skills. Additional assessment, conducted with the assistance of a Vietnamese interpreter, revealed that Trang showed similar receptive and expressive language deficits in Vietnamese. His nonverbal skills, however, were generally appropriate for his age. He readily constructed intricate buildings and vehicles with small, plastic building blocks; he easily completed complex jigsaw puzzles; and he successfully solved numerical, conceptual, or analogical problems, as long as they were presented nonverbally.

As a preschooler, Jerry was diagnosed with mixed receptive-expressive language disorder. By age 8, he had also received the diagnoses of comorbid attention-deficit/hyperactivity disorder and reading disorder. Jerry's language, attention, and reading problems contributed to a negative spiral of academic failure. His receptive language difficulties and attention problems resulted in frequent failures to understand and learn important information from classroom instructions, discussions, and lectures. Because information critical to the understanding of future lessons was not learned, Jerry continued to fall further and further behind his classmates. Moreover, because he could read only a few single words, Jerry could not acquire other key information as his classmates did by reading textbooks, library books, newspapers, and other reading materials. His inability to read proficiently further limited his opportunities to acquire new words, learn complex sentence forms, and appreciate sophisticated ideas. However, Jerry continued to score within the low normal range on nonverbal intellectual tasks, albeit somewhat lower than he had scored as a preschooler.

Children with mixed receptive-expressive language disorder are also at considerable risk for other disorders of speech, learning, and development. Approximately 40 to 60 percent of preschoolers with mixed receptive-expressive language disorders also have difficulties in producing speech sounds, indicating phonological disorder. In a few cases, stuttering may accompany mixed receptive-expressive language disorder. Reading disorder occurs in as many as 50 percent of school-aged children with mixed receptive-expressive language disorder, a much higher rate than in children with expressive language disorders. Reading problems may be evident at the earliest stages of learning to read or may not become obvious until later stages when successful reading comprehension demands knowledge of sophisticated vocabulary and sentence structures. A few children with mixed receptive-expressive language disorder also experience disorder of written expression or developmental coordination disorder.

Children with persistent mixed receptive-expressive language disorder are also at greater risk for concomitant psychiatric disorders than children with expressive language disorder. As many as 70 percent of school-aged children with mixed receptive-expressive language disorder may also have comorbid psychiatric diagnoses. Attention-deficit disorders are the most common comorbid conditions, occurring in 30 to 60 percent of cases. Externalizing or internalizing disorders may also be observed, but they are less frequent than attention disorders.

Children with developmental mixed receptive-expressive language disorder usually do not show overt neurological symptoms (e.g., electroencephalographic abnormalities, seizures, dysarthria, or dyspraxia). However, neurological symptoms do not preclude diagnosis of mixed receptive-expressive language disorder, unless they arise from a known acquired brain insult.

Acquired Type The language symptoms of acquired mixed receptive-expressive language disorder vary considerably, depending on the nature and severity of the underlying injury or disease. Although acquired disorders may have certain language characteristics in common with developmental disorders, they also show distinctive patterns. One of the more common patterns of outcome follows closed head injury, which often occurs in motor vehicle accidents or falls. Basic skills in vocabulary and grammar are relatively spared (after initial recovery stages), but more-complex skills remain significantly impaired. Comprehension of extended or sophisticated discourse may be compromised. Likewise, expression in discourse may be characterized by inappropriate, illogical, disorganized, or tangential comments, despite relatively proficient use of basic vocabulary and sentence structure. ccompanying deficits in attention, memory, affect, behavior, and motor skills may further impair overall communication effectiveness. Landau-Kleffner syndrome also has a distinct profile.

Patrick, age 4, was a cheerful, talkative preschooler who had good language skills until he experienced a series of seizures for no apparent reason. His language skills then began to deteriorate noticeably. He no longer understood many words, commands, and sentences that he had readily understood before the seizures. His expressive skills also deteriorated. His sentences became shorter and less sophisticated. It became increasingly difficult to decipher what he intended to say, as important words were omitted or incorrectly used. He also experienced significant attention and memory problems, although he continued to show nonverbal intellectual skills that were grossly within the normal range. Eventually, Patrick's seizures were controlled with medication, but he continued to exhibit substantial difficulties in both comprehension and production of language that never fully resolved.

DIFFERENTIAL DIAGNOSIS

Differential diagnosis of mixed receptive-expressive language disorder requires standardized evaluations of expressive language, receptive language, and nonverbal intellectual functioning. Standardized testing of receptive language is particularly critical, as deficits in this area may be less obvious than those in expressive language. The diagnosis requires that both receptive and expressive language development fall substantially below (1) the range of normal language performance expected for a child's age and (2) nonverbal intellectual performance. Moreover, the receptive and expressive language difficulties must be severe enough to impair academic performance or social communication. To make this determination, speech-language pathologists often use observational techniques and analysis of spontaneous language in conjunction with standardized assessments.

The related diagnoses of expressive language disorder, selective mutism, and mental retardation are ruled out by the presence of a significant discrepancy between receptive language and nonverbal abilities. In expressive language disorder and selective mutism, receptive language and nonverbal abilities are generally commensurate and are notably better than expressive language skills. In mental retardation, performance is similar across receptive language, expressive language, and nonverbal abilities, but all three are depressed relative to age-level expectations.

Language comprehension may also be affected in children with hearing impairment, environmental deprivation, or a pervasive developmental disorder. A hearing test is an important diagnostic procedure whenever a language disorder is suspected. Environmental deprivation can usually be identified through careful history taking. Pervasive developmental disorder may be indicated by disturbed social interaction and communicative intent, as well as unusual repetitive behaviors. If receptive and expressive language deficits exceed those expected for a particular degree of hearing impairment or mental retardation, a concurrent diagnosis of mixed receptive-expressive language disorder may be made.

Other comorbidity issues require attention in differential diagnosis. Difficulties in speech sound production may accompany language disorders (either expressive or mixed) and may suffice to warrant a comorbid diagnosis of phonological disorder. Because of the high concurrence of language and reading disorder, school-aged children diagnosed with either disorder should be screened for possible problems in the other area. Likewise, a high proportion of children referred for psychiatric assessment may exhibit previously undetected comorbid language disorders. Thus, children with psychiatric diagnoses, particularly attention-deficit/hyperactivity disorder, should also be screened for language disorders.

Occasionally a child may show expressive language so difficult to follow and so disorganized that it may be considered thought disordered and symptomatic of schizophrenia.

The following is a brief discussion with a highly anxious 9-year-old youngster with mixed receptive-expressive disorder who earlier that day had visited the dentist.

What do you suppose you'd like to do today?

Taking a dentist appointment ...

[You're taking a dentist appointment]

Yeah, he has a little cavity ...

[He has a little cavity]

... and I'm going to take it from him cause it's stuck there he sure doesn't like it but he knows there is too far (mumbled)

... too far and so he was losing a lot of hair so he needs a new way... sounds like he's bald... so he has everything (mumbling) a guy whos not recep... teaching he will have to it's important like he will or he won't

Discussion The stream of language from this child raises concerns about the possibility of schizophrenia. However, language this disorganized can occur in children with a mixed receptive-expressive disorder, especially in situations of heightened anxiety. Reassessing the child when less anxious and careful attention to the presence or absence of comorbid symptoms usually help clarify the diagnosis.

Children from diverse linguistic and cultural backgrounds may also have developmental language disorders. The diagnostic challenge is to differentiate cases of disorder from cases in which children are making language errors that are a normal part of acquiring a new language. A key factor in a successful assessment is often the cooperation of a well-trained informant from the same cultural and linguistic background as the child in question. The informant assists in obtaining history information, evaluating the child's proficiency in the first language, and conveying assessment results to the family.

Mixed receptive-expressive language disorder of the acquired type is usually identifiable by its sudden onset and association with known neurological conditions. A difficulty in using standardized tests for diagnosis is that the structure of a one-to-one testing situation may help to optimize the performance of children with acquired disorders, who often have associated attention and behavioral problems. These children may experience greater difficulties in language comprehension and production when they must function in the more distracting and less structured environments of home, school, and community. For this reason, observational techniques may also be important in diagnosis. In cases where Landau-Kleffner syndrome is suspected, a sleep EEG is often used to document abnormal brain wave activity and to help differentiate the disorder from other conditions in which previously learned skills deteriorate (e.g., childhood disintegrative disorder).

COURSE AND PROGNOSIS

Developmental Type Wide variations in severity, course, and outcome are observed for the developmental type of mixed receptive-expressive disorder. Recent studies of late talkers (i.e., children with normal cognitive functioning who use fewer than 50 words and no word combinations at age 2) suggest that youngsters with early mixed receptive-expressive delays have less favorable long-term prognoses than those with early expressive-only delays. Some preliminary evidence suggests that higher levels of nonverbal intelligence, receptive language, and early communicative intentions may be associated with better prognoses. However, it is not possible to predict the outcome for an individual child with any degree of certainty.

Throughout the preschool and school-age years, children with persistent mixed receptive-expressive language disorder continue to have poorer prognoses than those with expressive language disorder. Both language comprehension and production skills typically continue to develop, but often quite slowly. Thus, these children may become increasingly disadvantaged relative to their peers who are acquiring language skills at a typical rate. Children with mixed receptive-expressive language disorder also have a high probability of developing other associated problems during the school years, particularly reading disorders. Comprehension deficits also appear to place a child at high risk for attention-deficit disorders. Language problems, particularly when coupled with reading or attention problems, may lead into a vicious cycle of poor academic performance, low self-esteem, reduced motivation, and social isolation. Some theorists also posit that language plays a central role in moral development. The possibility that language deficits may prevent children from attaining age-appropriate levels of moral development is an intriguing one because of its potential link to the development of comorbid psychiatric and behavioral difficulties. Eventually, some children with mixed receptive-expressive language disorder learn to communicate well enough to conduct their daily affairs as adults, but many continue to experience subtle (or not so subtle) language deficits that limit their ability to participate effectively in a full range of communication and life situations. Indeed, detectable language problems are evident in a substantial portion of adults who live in poverty, commit crimes, or suffer psychiatric disturbances.

Acquired Type Acquired mixed receptive-expressive language disorders may have their onset suddenly at any age. Children generally recover more completely than adults from acquired language disorders of similar severity. Spontaneous recovery of function may occur rapidly or may have an extended course over months or years. Approximately two thirds of children with acquired language disorders regain language functioning that is grossly within the normal range, but some subtle deficits, particularly in high-level receptive or expressive tasks, may persist. When language deficits persist, they may have an adverse impact on academic performance and social communication, particularly when they are also accompanied by associated deficits in other areas of functioning (e.g., attention, memory, emotion regulation, behavior, motor skills). The prognosis for recovery from mixed acquired language disorders depends in a complex fashion on a number of factors, including the severity, location, and extent of brain damage; the age of the child at injury; the specific cause of the acquired disorder; the extent of other associated deficits; and the child's level of functioning prior to the brain insult.

Landau-Kleffner syndrome has a less favorable prognosis than acquired disorders attributable to other etiologies. The typical age of onset for the syndrome is from 4 to 7 years, but it ranges from 1 and 13 years. The observed language degeneration is usually sudden and pronounced. The comprehension impairment may be so marked that the child appears to have suddenly gone deaf, although hearing remains normal. Language deterioration may precede, follow, or accompany the onset of seizures or abnormal brain wave activity. In a few cases, overt seizures do not occur, but abnormal EEG readings are observed. The seizure activity often responds well to medication, but language functioning may not recover. Fluctuations in language performance are characteristic. Language recovery, when it occurs, has a gradual, inconsistent course usually over a period of months or years. In general, the younger the age at onset, the poorer the prognosis for language recovery in this syndrome.

TREATMENT

Developmental Type Intervention is generally recommended for children with mixed receptive-expressive language disorder, because they are at substantial risk for persistent linguistic, academic, and behavioral difficulties. The goals of intervention for these children usually encompass both receptive and expressive language needs. A variety of intervention techniques, styles, and programs have been recommended in the literature, but many have not been subjected to rigorous scientific evaluation for efficacy. Reviewers of the available efficacy literature, however, generally conclude that language intervention can be effective in fostering short-term progress toward specific language goals, a conclusion that appears to hold across a variety of intervention techniques, styles, and programs. To date, however, the possible long-term benefits of language intervention have received limited study. Future research must determine whether and under what circumstances language intervention can limit or reverse the negative impact of language disorder on long-term linguistic, academic, and psychosocial outcomes.

Interventionists now appear to be recognizing the need for a focus on long-term, broad-based outcomes. Language interventions for preschoolers are being designed to focus not only on oral language outcomes, but also on promoting social communication and early literacy skills. Kindergarteners with language disorders are being specifically taught certain key prereading skills in an effort to prevent the early reading failure that is often associated with language disorder and that may lead to subsequent academic and psychosocial problems. Peers, parents, and teachers are being taught methods for fostering language growth, which can be used on a regular basis in natural communicative interactions with children who have language disorders. Preliminary results of these efforts appear to be positive, but many key questions remain concerning optimal goals, techniques, and programs for language intervention.

Other interventions that rely heavily on language as a medium (e.g., psychotherapy, cognitive behavioral modification) may need to be modified for children with language disorders. To compensate for comprehension problems, vocabulary, sentence structure, questions, and verbal instructions should be simplified, and (if

necessary) repeated. To minimize the impact of expressive difficulties, nonverbal response modes may be used to supplement verbal responses. Clarification of ambiguous or unclear responses should also be sought. Similar modifications may be required for assessment procedures that depend heavily on language (e.g., projective tests, diagnostic interviews).

Acquired Type Language intervention for acquired language disorders often has three distinct phases. In the first phase, intervention is intended to stimulate and complement the spontaneous recovery of function that typically follows the neurological insult. In the second phase, the goal is to reteach functional and adaptive language skills, usually in a one-to-one setting that is relatively free of distractions. In the final phase, the focus is on helping the child reintegrate into home, community, and school environments. At this stage, a child may be taught strategies intended to compensate for persistent deficits. Throughout, language goals are usually part of a larger program intended to foster improvement in other affected areas as well (e.g., cognitive, behavioral, motor). Intervention techniques frequently are a blend of those typically used for developmental language disorders and those used for acquired disorders in adults. Unfortunately, there has been little systematic evaluation of the efficacies of various interventional approaches for children with acquired language disorders.

SUGGESTED CROSS-REFERENCES

Two disorders closely related to mixed receptive-expressive disorder are expressive language disorder (Section 37.1) and reading disorder (Section 35.1). Disorders that are relevant to the differential diagnosis of mixed receptive-expressive language disorder include mental retardation (Chapter 34), pervasive developmental disorder (Chapter 38) and selective mutism (Section 46.4). Disorders that may occur with mixed receptive-expressive language disorder include reading disorder Section 37.2), mathematics disorder (Section 35.2), disorder of written expression (Section 35.3), developmental coordination disorder (Chapter 36), phonological disorder (Section 37.3), stuttering (Section 37.4), and attention-deficit disorders (Chapter 39).

SECTION REFERENCES

*Baltaxe C: Psychiatric disorders and communication. In *Introduction to Organic and Neurogenic Disorders of Communication*, C Ferrand, R Bloom, editors. Allyn & Bacon, Needham Heights, MA, 1997.

Bashir AS, Scavuzzo A: Children with language disorders: Natural history and academic success. J Learn Disabil *25:*53, 1992.

Beitchman JH, Wild J, Kroll R: An overview of childhood speech and language disorder. In *Handbook of Child and Adolescent Psychiatry*, vol 4, JD Noshpitz, NE Alessi, editors. Wiley, New York, 1997.

*Beitchman JH, Wilson B, Brownlie EB, Walters H, Inglis A, Lancee W: Long term consistency of speech/language profiles, II. Behavioral, emotional, and social outcomes. J Am Acad Child Adolesc Psychiatry *35:*815, 1996.

*Beitchman JH, Wilson B, Brownlie EB, Walters H, Lancee W: Long term consistency of speech/language profiles, I. Developmental and academic outcomes. J Am Acad Child Adolesc Psychiatry *35:*804, 1996.

Benasich A, Curtiss S, Tallal P: Language, learning, and behavioural disturbances in childhood: A longitudinal perspective. J Am Acad Child Adolesc Psychiatry *32:*585, 1993.

*Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P: Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. J Speech Lang Hear Res *42:*155, 1999.

Blosser J, DePompei R: *Pediatric Traumatic Brain Injury*. Singular, San Diego, 1994.

Brinton B, Fujiki M, Spencer J, Robinson L: The ability of children with specific language impairment to access and participate in an ongoing interaction. J Speech Lang Hear Res *40:*1011, 1997.

Cantwell DP, Baker L: *Psychiatric and Developmental Disorders in Children with Communication Disorder*. American Psychiatric Press, Washington, DC, 1991.

Chapman SB, Watkins R, Gustafson C, Moore S, Levin H, Kufera JA: Narrative discourse in children with closed head injury, children with language impairment, and typically developing children. Am J Speech Lang Pathol *6:*66, 1997.

Cohen NJ, Davine M, Horodezky N, Lipsett L, Isaacson L: Unsuspected language impairment in psychiatrically disturbed children: Prevalence and language and behavioral characteristics. J Am Acad Child Adolesc Psychiatry *32:*595, 1993.

Enderby P, Emerson J: Children with speech and language disorders. In *Does Speech and Language Therapy Work? A Review of the Literature*, P Enderby, J Emerson, editors. Whurr, London, 1995.

*Fazio B: Arithmetic calculation, short-term memory, and language performance in children with specific language impairment. J Speech Lang Hear Res *42:*420, 1999.

Gathercole SE, Baddeley A: *Working Memory and Language*. Erlbaum, Hillsdale, NJ, 1993.

Gauger L, Lombardino L, Leonard C: Brain morphology in children with specific language impairment. J Speech Lang Hear Res *40:*1272, 1997.

Goldstein H, English K, Shafer K, Kaczmarek L: Interaction among preschoolers with and without disabilities: Effects of across-the-day peer intervention. J Speech Lang Hear Res *40:*33, 1997.

*Johnson CJ, Beitchman JH, Young A, Escobar M, Atkinson L, Wilson B, Brownlie EB, Douglas L, Taback N, Lam I, Wang M: Fourteen year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. J Speech Lang Hear Res *42:*132, 1999.

Leonard LB: *Children with Specific Language Impairment*. Bradford, Cambridge, MA, 1998.

Leonard L: Children with specific language impairment (developmental dysphasia): Treatment. In *Linguistic Disorders and Pathologies: An International Handbook*, G Blanken, J Dittmann, H Grimm, J Marshall, C Wallesch, editors. DeGruyter, New York, 1993.

Merzenich M, Jenkins W, Johnston P, Schreiner C, Miller S, Tallal P: Temporal processing deficits of language-learning impaired children ameliorated by training. Science *271:*77, 1996.

Nye C, Foster SH, Seaman D: Effectiveness of language intervention with the language/learning disabled. J Speech Hear Disord *52:*348, 1987.

Owens RE: *Language Disorders: A Functional Approach to Assessment and Intervention*, ed 2. Allyn & Bacon, Boston, 1995.

*Paul R: *Language Disorders from Infancy through Adolescence: Assessment & Intervention*. Mosby, St. Louis, MO, 1995.

*Rapin I: Practitioner review: Developmental language disorders: A clinical update. J Child Psychol Psychiatry *37:*643, 1996.

Redmond S, Rice M: The socioemotional behaviors of children with SLI: Social adaptation or social deviance? J Speech Lang Hear Res *41:*688, 1998.

Riccio C, Hynd G: Developmental language disorders in children: Relationship with learning disability and attention deficit disorder. School Psychol Rev *22:*696, 1993.

Rossetti L: Epidemiology of risk and socio-communicative development in medically fragile children. In *Introduction to Organic and Neurogenic Disorders of Communication*, C Ferrand, R Bloom, editors. Allyn & Bacon, Needham Heights, MA, 1997.

Spitz R, Tallal P, Flax J, Benasich A: Look who's talking: A prospective study of familial transmission of language. J Speech Lang Hear Res *40:*990, 1997.

Tallal P, Miller S, Bedi G, Byma G, Wang X, Nagarajan S, Schreiner C, Jenkins W, Merzenich M: Language comprehension in language-learning impaired children improved with acoustically modified speech. Science *271:*81, 1996.

Tappan M: Language, culture, and moral development: A Vygotskian perspective. Dev Rev *17:*78, 1997.

Tomblin JB, Buckwalter P: Heritability of poor language achievement among twins. J Speech Lang Hear Res *41:*188, 1998.

Tomblin JB, Records N, Zhang X: A system for the diagnosis of specific language impairment in kindergarten children. J Speech Hear Res *39:*1284, 1996.

37.3 PHONOLOGICAL DISORDER

CARLA J. JOHNSON PH.D., AND JOSEPH H. BEITCHMAN, M.D.

DEFINITION AND COMPARATIVE NOSOLOGY

Children with phonological disorders have difficulties in correctly producing the speech sounds appropriate for their age and dialect. They may omit sounds (e.g., saying “at” for *hat* or “bup” for *bump*), substitute sounds for other sounds (e.g., saying “tum” for *come* or “hewo” for *hello*), or distort sounds by producing them in an unusual manner (e.g., producing /sh/ while allowing too much air to flow noisily over the sides of the tongue or producing /*s*/ or /*z*/ while protruding the tongue). In many cases, the speech sound differences occur in systematic patterns. For example, consonants in the final position of words may be omitted (e.g., “fee” for *feet*, “no” for *nose*) or consonants that require a steady airflow may be produced with an interrupted airflow (e.g., “do” for *zoo*, “pan” for *fan*). Listeners may be unable to understand the speech of children with severe phonological disorders.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) uses somewhat different terminology for speech sound production disorders than that commonly used in the clinical and scientific literature, which may cause a problem for diagnosticians. In DSM-IV the category of phonological disorder is an inclusive one, encompassing difficulties in speech sound production that have no known cause as well as those that arise from hearing impairment, structural abnormalities of the speech mechanism (e.g., cleft lip or palate), or neurological conditions (e.g., cerebral palsy, head injury). In the literature, the terms *articulation disorder* or *speech sound production disorder* are more often used to label such an inclusive category. The terms *dysarthria* and *dyspraxia* are typically used to refer to speech sound production difficulties that have a neurological origin. The term *phonological disorder* is usually reserved for speech sound problems that have no known cause and presumably reflect developmental difficulties in acquiring the regularities of phonology (the sound system of a language). In this section this more restrictive category is referred to as *developmental phonological disorder* to distinguish it from the larger DSM-IV category of phonological disorder.

The inclusive nature of the DSM-IV category of phonological disorder obscures some potentially important differences between speech sound difficulties of developmental (unknown) origin and those with known structural or neurological causes. In addition to etiology, there may be marked differences in characteristics, prevalence, course, prognosis, and treatment. Moreover, the inclusive category used in DSM-IV differs from the more restrictive category of developmental articulation disorder used in its predecessor, the revised third edition of DSM (DSM-III-R). The earlier category included only developmental speech disorders of unknown origin and excluded those with structural or neurological causes. Similarly, the category of specific speech articulation disorder in the 10th revision of *International Classification of Diseases and Related Health Problems* (ICD-10) encompasses only developmental speech difficulties. Structural or neurological conditions are excluded (see Table 37.5-2).

Developmental phonological disorders are far more common than speech production problems from structural or neurological causes. Phonology (the sound system) is an aspect of language that children must learn. Thus, it is not surprising that developmental phonological disorder shares many similarities with the developmental language disorders discussed in Section 37.1 and Section 37.2. All are relatively common disorders in young children. All have unknown etiologies. All are characterized by a delayed onset and slow rate of growth but a relatively normal sequence of development in the affected communication skills.

Speech sound development is thought to require both linguistic and motor learning, as well as integration and coordination of the two. Linguistic aspects of the phonological system that need to be acquired include information on the specific sounds that are used in a particular language and how those sounds can be contrasted and combined with each other. Permissible patterns of syllable and word stress and patterns of speech melody (intonation) also need to be internalized. Thus, speech sound development is often thought to involve the gradual implicit development of a complex system of linguistic rules that underlie the representation, perception, and production of speech. Whether these rules explain or merely describe phonological knowledge is a point of controversy.

Considerable motor learning is required for speech sound development. Speech is a complex fine-motor skill that requires rapid, precise, and coordinated action of the large number of muscles involved in breathing, voicing, and articulating speech sounds. The coordination of these numerous muscle actions must be accurate and rapid enough to permit rates of speech as high as 250 words a minute for adult speakers.

The normal course of speech sound development is systematic. Children's first words generally comprise consonant (C) and vowel (V) sequences that form simple syllables of the forms CV (e.g., “hi”), VC (e.g., “up”), or CVC (e.g., “pop”) and short repetitive sequences of the form CVCV (e.g., “mama”). Gradually, more-complex syllable and word shapes are mastered, including those with initial or final consonant cluster sequences, such as /*thr*-/, /*k*l-/, -/rst/, or /-nt/. The consonants typically used in early words are those designated the *early 8* (/*m*/, /*b*/, /*y*/, /*n*/, /*w*/, /*d*/, /*p*/, /*h*/). As development proceeds, consonants termed the *middle 8* (/t/, /*ng*/, /*k*/, /*g*/, /*l*/, /*v*/, /*ch*/, /*l*/) are gradually mastered; followed by those known as the *late 8* (/*sh*/, /*th*/ [two variants], /*s*/, /*z*/, /*l*/, /*r*/, /*zh*/). Mastery of the late-8 consonants may not be fully completed until the age of 7 to 8 years. In the course of normal development, children often produce speech sounds inconsistently and incorrectly. Their errors can be described as omissions (e.g., “dah” for *dog*), substitutions (e.g., “fink” for *think*), or distortions (e.g., a slurpy-sounding /*sh*/). Sound errors may also be described in terms of phonological processes or rules that seem to capture a pattern of errors. For example, many young children appear to be using a phonological process of final consonant deletion; that is, they consistently omit consonants that occur at the end of words. The process occurs normally in the speech of very young children but is typically overcome by the age of 3. Other common processes that are gradually overcome during development include reduplication (i.e., repeating a syllable to form a multisyllabic word—e.g., “bahbah” for *bottle*), unstressed syllable deletion (i.e., omitting a syllable that is not stressed—e.g., “tefon” for *telephone*), stopping (i.e., using an interrupted airflow to produce sounds that should have a continuous airflow—e.g., “tu” for *shoe*), *consonant harmony* (i.e., using a sound similar to one that occurs elsewhere in a word—e.g., “tat” for *cat*), and consonant cluster reduction (i.e., omitting or simplifying sounds in a consonant cluster—e.g., “moke” for *smoke*). All of these processes are common in the speech of young children, but all are usually eliminated by the age of 5 years.

EPIDEMIOLOGY

The prevalence of phonological disorder has not been definitively established, in part because definitions and test procedures have varied considerably across studies. One study of a large community sample found that 7 to 8 percent of 5-year-old children had speech sound production difficulties of developmental, structural, or neurological origins. Approximately 55 percent of them had problems only in speech sound production; the others experienced additional language difficulties. Another report suggested that 7.5 percent of children from 3 to 11 years of age have developmental phonological disorders: 2.5 percent with delayed speech (deletion and substitution errors persisting past 4 years of age) and 5 percent with residual articulation errors (distortions of a few sounds, often /*s*/, /*l*/, or /*r*/, persisting beyond age 8). Developmental phonological disorders are far more common than phonological disorders with known structural or neurological causes.

Boys are more likely to have speech sound production difficulties than girls. Sex ratios as high as 4 to 1 and as low as 2 to 1 have been reported in the literature. Familial aggregation of speech disorders is also common, with 30 to 60 percent of children with developmental phonological disorders having one or more immediate family members with a history of a similar disorder. Further work is required to determine the genetic and environmental factors that may underlie familial aggregation.

ETIOLOGY

The cause of developmental phonological disorder is not known. Many of the factors proposed as possible causes are the same as those considered for developmental language disorders, including (1) subtle brain abnormalities, (2) genetic factors, (3) environmental influences, (4) neurodevelopmental immaturity, (5)

linguistic factors, and (6) auditory processes. Speech mechanism and speech motor factors have also been considered as possible causes.

Each proposed cause has some support in the literature, but no single factor seems to account for all cases of developmental phonological disorder. The evidence suggesting brain abnormalities includes a higher-than-normal incidence of soft neurological signs as well as findings of abnormalities in children with related disorders, such as mixed receptive-expressive language disorder and reading disorder. Genetic factors are implicated by the observed patterns of familial aggregation and results from twin and adoption studies and linkage studies involving phonological awareness, a component skill in reading. Environmental influences may also contribute to familial aggregation, but no conclusive evidence links specific environmental factors to developmental phonological disorders. Neurodevelopmental immaturity may be reflected in the fact that children with developmental phonological disorder typically develop speech skills more slowly than (but in roughly the same sequence as) children without the disorder. Other characteristics such as soft neurological signs, subtle deficits in motor skills, and associated behavioral difficulties could also reflect neurodevelopmental immaturity. Linguistic deficits may limit the ability to extract the regularities of phonology (e.g., permissible sounds, combinations, and positions of occurrence) from the available linguistic input. Temporary hearing losses or more permanent problems in auditory perception or discrimination may also contribute to developmental phonological disorder in some cases. Subtle abnormalities in the structure and function of the speech mechanism also are relatively common in children with developmental phonological disorder, but it is not clear whether these abnormalities are causal factors or simply correlated deficits. Thus, no single causal hypothesis enjoys convincing support, which suggests that multiple factors may be responsible for developmental phonological disorder.

Abnormalities in the structure of the articulators (e.g., lips, tongue, jaw, soft palate) occasionally cause speech sound production problems. The most common of these structural abnormalities is cleft palate, a congenital condition in which a cleft (opening) in the palate (roof of the mouth) fails to close properly during fetal development. The condition is sometimes accompanied by cleft lip. Other structural abnormalities in the articulators may occur following tumor removal or injury to the mouth or face. Tongue tie and dental abnormalities are rarely responsible for speech sound production disorders.

Neurological conditions may also cause speech sound production problems. One type of speech difficulty attributable to neurological damage is dysarthria, which is characterized by paralysis, weakness, or incoordination of the speech musculature. Dysarthria may result from congenital neurological conditions such as cerebral palsy, from progressive degenerative conditions such as muscular dystrophy, or from acquired brain insults such as head injury, stroke, tumors, or infections. A second type of speech problem arising from neurological damage is dyspraxia or apraxia, which is characterized by difficulty in planning and executing speech movements in the absence of paralysis or weakness of the speech musculature. Apraxia may follow acquired brain insults. A developmental form of apraxia with no known cause has also been posited, but its status as an identifiable disorder is controversial.

DIAGNOSIS AND CLINICAL FEATURES

Phonological disorder is diagnosed when speech sound production is inappropriate for the age and dialect of the speaker (Table 37.3-1). The severity of the disorder may range from mild, in which speech intelligibility is barely affected, to severe, in which speech is totally unintelligible. Errors may occur with only one or two sounds or with virtually all sounds.

A. Failure to use developmentally expected speech sounds that are appropriate for age and dialect (e.g., errors in sound production, use representation or organization, such as, but not limited to, substitutions of one sound for another [use of /t/ for target /k/ sound] or omissions of sounds such as final consonants).

B. The difficulties in speech sound production interfere with academic or occupational achievement or with social communication.

C. If mental retardation, a speech—motor or sensory deficit, environmental deprivation is present, the speech difficult are in excess of those usually associated with these problems.

Coding note: If a speech—motor or sensory deficit or a neurological condition is present, code the condition on Axis III.

Table 37.3-1 DSM-IV Diagnostic Criteria for Phonological Disorder

Children with developmental phonological disorders typically have a delayed onset and slow rate of speech sound development, but the sequence of development is generally quite similar to that found in children who are developing normally. A repertoire of simple syllable and word shapes is slowly expanded to include more complex shapes. The early-eight consonant sounds are generally mastered first, followed by the middle eight and late eight, but acquisition may be very slow. Speech sound errors are usually similar to those that occur in normal development, but the relative distribution of omissions, substitutions, or distortions may be dissimilar. For example, one study of 3- to 6-year-olds found that omissions and substitutions accounted for 25 and 47 percent, respectively, of errors in the speech of children with developmental phonological disorders but only 5 and 14 percent of errors in the speech of children without disorders. Phonological processes common in normal development also characterize the speech of children with developmental phonological disorder, but these processes may persist well beyond the age when they are normally overcome. Occasionally, unusual or idiosyncratic phonological processes may be used by children with developmental phonological disorder.

Sasha was a talkative, likable 3-year-old whose speech was virtually unintelligible. Sasha had normal hearing and language comprehension skills. No firm conclusion about his level of expressive language development could be made because he was so difficult to understand. He did, however, seem to be producing multiword utterances. Sasha produced only a small number of early-developing consonants (/ *m*/, /*n*/, /*d*/, /*t*/, /*p*/, /*b*/, /*h*/, /*w*/), vowels (/*ee*/, /*ah*/, /*oo*/), and syllable shapes (V, CV, CVCV). As a result, many of his spoken words were indistinguishable from each other (e.g., he said “bahbah” for *baby, bottle*, and *bubble;* he used “nee” for *knee, need*, and *Anita* [his sister]). Moreover, he never produced consonant sounds at the end of words or used consonant cluster sequences (e.g., /*tr-*/, /*st-*/, /*-nt*/, /*-mp*/). On occasion, Sasha reacted with frustration and tantrums to his difficulties in making himself understood.

Don was a pleasant, cooperative 5-year-old diagnosed with a developmental phonological disorder. He had normal hearing and language comprehension skills. He showed some mild deficits in certain aspects of expressive language, particularly in using grammatical features (e.g., pronouns, auxiliary verbs, past tense word endings) and in formulating complex sentences. He produced all vowels and most early-developing consonants correctly, but his productions of later-developing consonants (e.g., /*r*/, /*l*/, /*s*/, /*z*/, /*sh*/, /*th*/, /*ch*/) were much less stable. Sometimes he used them correctly, sometimes he omitted them altogether, sometimes he distorted them or substituted other sounds for them (e.g., “w” for /*r*/, “f” for /*th*/). Don's phonological difficulties were most apparent in his productions of consonant cluster sequences and multisyllabic words. Clusters rarely contained the correct sounds or the required number of sounds (e.g., blue might be produced as “bue” or “bwue,” *hearts* might be said as “hots” or “hars”). Multisyllabic words also included incorrectly produced sounds and sometimes had syllables omitted (e.g., “efant” for *elephant*, “getti” for *spaghetti*) or sounds transposed (e.g., “aminal” for *animal*, “lemon” for *melon*). Strangers usually understood about 80 percent of what Don said. Fortunately, if asked to repeat something, he sometimes spoke more slowly and clearly than usual.

Barb was a shy, reserved 8-year-old whose speech development had been delayed as a child. Most of her earlier speech problems had resolved gradually during her preschool and early elementary school years. She continued, however, to have problems in producing a few late-developing sounds (/ *r*/, /*l*/, /*th*/). She often substituted /*w*/ for /*r*/ and /*l*/ and used /*t*/ or /*d*/ for /*th*/. Despite these minor errors, her speech was readily understandable. Nonetheless, her classmates teased her about her speech, which made her reluctant to speak in the presence of others.

Developmental phonological disorder may accompany other language, learning, and developmental disorders. Approximately 50 to 75 percent of children with developmental phonological disorders have mild-to-severe deficits in language expression (see Section 37.1 on expressive language disorder); 10 to 40 percent also show mild-to-severe deficits in language comprehension. Reading disorder is another common comorbid diagnosis, especially in children with associated language involvement, who are also at increased risk for other problems such as developmental coordination disorder, learning disorders, and attention-deficit disorders. Children with isolated developmental phonological disorders are at relatively little risk for behavioral or psychiatric disorders.

In contrast to children with developmental phonological disorders, children with structural or neurological disorders often exhibit patterns of speech sound errors that

are unlike those found in normal development. Children with structural and neurological deficits may also experience developmental phonological problems, which complicates diagnosis.

The speech of children with cleft palate may be characterized by excessive nasality and by inability to produce the many consonants that require the oral cavity to be closed off from the nasal passage by the palate. Children may develop unusual ways of articulating certain sounds in an attempt to compensate for structural deficiencies. Even after structural problems are corrected by surgical or prosthetic means, these unusual patterns may persist. Children with cleft palate are also at increased risk for upper respiratory infections and accompanying hearing losses, as well as developmental language and learning difficulties.

> Grace was a cheerful 6-year-old who was born with a severe cleft lip and palate. During infancy, she required surgery to effect initial repairs on her lip and palate. Several subsequent surgeries were needed to adjust those initial repairs to the growth of her face. Grace did not have sufficient palatal tissue for surgeons to achieve a completely adequate repair of the soft palate, so she was fitted with a prosthetic device for the roof of her mouth, which provided the separation of the oral and nasal passages required for speech. The prosthesis distorted her production of a few speech sounds that normally are produced with no separation between the oral and nasal passages (*m, n*, and *ng* sounded more like /*b*/, /*d*/, and /*g*/). These distortions, however, were minor relative to the severe speech difficulties Grace would have experienced without the prosthesis. Her speech likely would have been unintelligible because of excessive nasality and her inability to produce the oral consonants. Grace's upper lip was also a bit tight as a result of the corrective surgery. This limited her ability to round her lips but had no audible effects on her speech production. Grace experienced frequent upper respiratory and ear infections as a child. She also showed some mild expressive language delays, but her academic skills developed as expected.

Children with neurological conditions may have apraxia, a difficulty in planning and executing speech movements despite normal muscle strength and range of movement. Groping and hesitant articulation is characteristic of apraxia, which generally arises suddenly as a result of acquired neurological insults, such as head injury, stroke, or infections.

Neurological conditions may also result in dysarthria, a speech difficulty caused by weakness or paralysis in the speech musculature. There are several types of dysarthrias, each characterized by particular neuromuscular deficits and speech features. In one common type of dysarthria, speech is slow, labored, slurred, and imprecise, sounding much like the speech of someone who is drugged or intoxicated. The underlying muscular problems in dysarthria may affect not only production of speech sounds, but also breathing, voice quality, pitch, loudness, stress, and speech melody. Feeding, swallowing, and drooling problems may also result from impaired functioning of the oral musculature. Gross motor functioning of the limbs may also be affected by the underlying neurological disorder. Dysarthrias range from mild to severe. In severe cases intelligible speech may not be possible.

> Shane was an intelligent, likeable 12-year-old who was born with severe cerebral palsy. His quadriplegia made it necessary for him to use an electric wheelchair. Shane had a severe dysarthria and was able to produce only a few vocalizations consistently enough for them to be used for communication purposes. He was unable to use sign language as a method of communication because he could not execute the necessary hand movements. Nonetheless, he was able to communicate quite effectively, albeit somewhat slowly, by using a sophisticated computerized device as a communication aid. He operated this computerized system by selecting words or messages via an electronic switch that he activated with head movements, which he could control reasonably well. The message he composed could then be produced for his listener by a computerized voice output system. Using his computerized system, Shane also pursued many of his favorite leisure activities, such as playing chess and participating in chat groups on the Internet.

DIFFERENTIAL DIAGNOSIS

The key element in diagnosing phonological disorder is determining whether speech sound production is inappropriate for the age and dialect of the speaker. The decision may not be straightforward because the speech characteristics of children with developmental phonological disorder often resemble those of younger children who are developing normally. Speech errors indicating disorder in a 6-year-old may be perfectly normal in a 4-year-old. Thus, a particular pattern of speech sound errors must be carefully compared with the normal patterns of speech sound development in children.

Consideration must also be given to the dialect spoken by the child. Some sound differences that are normal in one dialect may indicate disorder in another. For example, the production of “hot” for *heart* is normal for speakers from Boston, but not for those from most other parts of North America. Children who are learning English as a second language may also present a difficult diagnostic challenge. Some difficulties in speech sound production are a normal part of acquiring a second language; others may suggest disorder. A careful history and an assessment of speech sound production skills, ideally in both English and the child's first language, may help to determine the correct diagnosis.

A hearing test and an oral examination are also important parts of the diagnostic assessment. The hearing test may identify temporary or permanent hearing problems that require medical or other interventions to improve the prognosis for development of intelligible speech. The examination of oral structures may reveal previously undiagnosed structural or functional problems that may be contributing to the speech difficulties. For example, children with neurologically based disorders may have difficulties in accurately and rapidly performing various speech and nonspeech movements (e.g., sequencing syllables, as in “putuku”; protruding the tongue; rounding the lips). Findings from the oral examination may, therefore, contribute to differentiating developmental and neurologically based disorders. Dysarthria involves muscle weakness and paralysis, signs of which may be observable during the oral examination. Drooling, feeding, or swallowing problems, either concurrently or in the history, may also suggest dysarthria.

Children diagnosed with developmental phonological disorders should also be screened for possible comorbid expressive or mixed receptive-expressive language disorders. When the history warrants, cognitive assessments may also help to rule out or confirm developmental delay or mental retardation. In these cases, a concurrent diagnosis of phonological disorder may be made when the speech difficulties exceed those expected for the degree of retardation or delay.

Referral to a speech-language pathologist for assessment of speech sound production skills is indicated when (1) a 2-year-old uses little or no speech, (2) a 3-year-old is poorly understood by family members familiar with his speech, (3) a 4-year-old is poorly understood by peers and strangers, (4) a child of any age appears to be withdrawing from social interactions because of teasing or frustration associated with speech difficulties, (5) a parent or teacher is concerned about a child's speech, or (6) a child (or adult) experiences a sudden deterioration in speech production skills. Deterioration in speech skills may be the first warning sign of a developing neurological condition and, therefore, warrants referral for neurological examination.

COURSE AND PROGNOSIS

Severe cases of developmental phonological disorder with comorbid language problems may be identified as early as 2 years of age, but diagnosis at about age 4 is typical. Variability in severity, course, and outcome is characteristic of developmental phonological disorders. In most children with the disorder, speech sound development proceeds gradually, at a slower than normal rate. Speech sound production skills often normalize eventually. One recent study investigated sound normalization over a 1-year period in a large group of children with developmental phonological disorders, whose average age was slightly more than 4 years at the start of the study. Approximately 20 percent of the children achieved normalized speech production skills during the year, but investigators were unable to find any significant speech, hearing, oral, cognitive, linguistic, or psychosocial variables that predicted normalization. A second study of long-term normalization suggested that speech sound development proceeds in spurts and plateaus until approximately 8 to 9 years. If speech has not fully normalized by then, residual articulation problems may persist into adulthood. Residual articulation errors are usually distortions of later-developing sounds, such as /*s*/, /*z*/, /*sh*/, /*ch*/, /*l*/, /*r*/, and /*th*/. Distortions that affect the sounds /*s*/, /*z*/, /*sh*/, and /*ch*/ are often called *lisps*. Residual errors are often mild, with little effect on speech intelligibility. Research suggests, however, that others may incorrectly consider a speaker with residual speech errors to be immature, lazy, or of low intelligence.

Children with mild-to-moderate isolated developmental phonological disorders are at relatively little risk for adverse long-term outcomes. Speech usually normalizes by 8 to 9 years of age, although subtle deficits in the speed and accuracy of articulation during complex speech tasks may remain detectable into adulthood. Adults with a history of mild-to-moderate developmental phonological disorder generally attain language, cognitive, academic, psychosocial, and psychiatric outcomes within the normal range.

Children with severe developmental phonological disorders that persist to school age are at greater risk for poor outcomes. They may fail to develop adequate phonological awareness (the ability to perceive, identify, and mentally manipulate speech sounds), which may contribute to marked difficulties in learning to read and spell. Thus, parents and teachers should carefully monitor the reading and spelling progress of school-age children with severe, persistent phonological disorders. Explicit training in phonological awareness may help to improve reading and spelling performance. There has been little systematic study of long-term psychosocial

and psychiatric outcomes for children with severe developmental phonological disorders.

The risk of adverse outcomes in language, reading, academic, psychosocial, or psychiatric functioning also increases for children who have comorbid mixed receptive-expressive language disorder. Those with significant receptive involvement are particularly at risk.

Children whose speech difficulties arise from structural abnormalities also have variable outcomes. In children with cleft palate, the adequacy of the speech outcomes is often determined by the success of surgical or prosthetic management. Children with cleft palate are, however, vulnerable to hearing and language difficulties, which may have a negative impact on overall prognosis for academic and social success. In addition, cleft palate is often part of a syndrome of abnormalities that may include mental retardation. Thus, prognosis for individual children will vary, depending on the extent of speech and associated involvements.

Children with speech difficulties of neurological origin also have varied prognoses. Those with mild difficulties may achieve intelligible speech, whereas those with more-severe problems may be unable to do so. Children with neurologically based speech problems also often have other involvements in the motor, language, and cognitive domains, which will influence overall outcome.

TREATMENT

Intervention is usually recommended for children of any age who are diagnosed with moderate-to-severe developmental phonological disorders. A speech-language pathologist may deliver the intervention directly or may teach parents, teachers, volunteers, or paraprofessionals to deliver the intervention under supervision. A variety of intervention techniques, styles, and programs have been recommended, but most can be grouped into two main approaches to treatment.

A *phonological approach* is usually implemented with children whose multiple speech sound errors can be described by the phonological processes (e.g., final consonant deletion, consonant cluster reduction). The goal in this approach is to help the child eliminate the use of these processes by emphasizing the communicative value of correctly using certain classes of sounds, such as final consonants or consonant clusters. Practice in producing a variety of final consonants and clusters then helps to consolidate these classes of sounds within the child's phonological system. Treatment generally continues until the child is using the new classes of sounds consistently in everyday communication situations.

A second approach, called the *traditional approach* because it was developed first, is more often used with children who have only a few error sounds, usually substitutions or distortions. The goal of this approach is to teach correct production of a sound by emphasizing its auditory and motor characteristics. Practice in producing the sound is provided, with cues about correct articulator placement or movement given by the clinician. Treatment generally continues until the child is using the new sound correctly in spontaneous speech.

One pattern of residual errors is often treated somewhat differently. Some children who lisp by protruding their tongue may also have an abnormal (for older children) swallowing pattern known as tongue thrust. These children may benefit from simultaneous treatment of the swallowing pattern and the lisp.

The research on treatment efficacy, while not extensive or comprehensive, does suggest that both phonological and traditional approaches can improve speech sound production skills. Long-term success appears to depend on the extent to which the new speech skills were established as automatic patterns of production. Not surprisingly, the extent to which children are motivated to change their speech patterns also appears to be a factor in the short- and long-term success of intervention.

Children with comorbid communication disorders may require language or fluency interventions in addition to speech sound intervention. There has been little systematic evaluation of the efficacy of such joint interventions. A few studies have examined whether speech production gains occur as an additional benefit of language or fluency interventions that do not focus specifically on speech sound production. The limited results to date have been mixed. Further systematic work is required on these and other questions concerning the efficacy of speech interventions.

For children with structural abnormalities, surgical or prosthetic management is often the first, and sometimes the only, step in intervention. Speech sound difficulties that remain after surgical or prosthetic management may require speech treatment, often with techniques similar to those used for developmental phonological disorders. On occasion, persisting structural deficits may not permit a sound or sounds to be produced in the usual way. Compensatory articulation strategies are then included in the intervention program.

Children with apraxia following acquired brain insults are often treated with techniques similar to those used in the traditional approaches. Production practice emphasizes rhythmic and efficient sequencing of difficult articulatory movements. Some children are also thought to have a developmental apraxia of speech (with no known cause), but the diagnostic category is controversial, and its characteristic features are poorly understood. Treatment approaches similar to those for acquired apraxia are generally recommended.

Interventions for children with dysarthria cannot eliminate the basic problems in muscle strength and control that underlie the disorder but can promote optimal use of available strength and control for speech purposes. Compensatory techniques for speech production are often a key part of this process. Prosthetic devices may also help to optimize speech intelligibility. Early intervention to limit abnormal muscular reflexes and promote optimal feeding and swallowing patterns may also be required. For children who are unable to achieve intelligible speech, a variety of augmentative and alternative communication systems may be recommended. Sign language may be used to supplement speech if hand movements can be well controlled. Other communication systems involve picture or symbol boards or sophisticated computerized devices, such as that described in the case example above. Research has shown that augmentative and alternative communication systems do not reduce a child's motivation to improve speech skills. Indeed, the opposite effect is often found.

Research that evaluates the efficacy of speech-sound interventions for children with structural and neurological problems is sparse, in part because such interventions are often tailored specifically to the particular needs of an individual child. Accordingly, the available studies often focus on outcomes for individual children.

SUGGESTED CROSS-REFERENCES

Other communication disorders that may accompany phonological disorder are discussed in Section 37.1 on expressive language disorder, Section 37.2 on mixed receptive-expressive language disorder, Section 37.4 on stuttering, and Section 37.5 on communication disorder not otherwise specified. Other possible concurrent disorders are described in Section 35.1 on reading disorder, Section 35.2 on mathematics disorder, Section 35.3 on disorder of written expression, and Chapter 36 on developmental coordination disorder. Disorders that may require attention in differential diagnosis are discussed in Chapter 34 on mental retardation and Chapter 38 on pervasive developmental disorders.

SECTION REFERENCES

Beitchman JH, Brownlie EB, Inglis A, Wild J, Mathews R, Schachter D, Kroll R, Martin S, Ferguson B, Lancee W: Seven-year follow-up of speech/language-impaired and control children: Speech/language stability and outcome. J Am Acad Child Adolesc Psychiatry *33:*1322, 1994.

Beitchman JH, Nair R, Clegg M, Patel PG: Prevalence of speech and language disorders in 5-year-old children in the Ottawa-Carleton region. J Speech Hear Disord *51:*98, 1986.

*Bernthal J, Bankson N: *Articulation and Phonological Disorders*, ed 4. Prentice Hall, Englewood Cliffs, NJ, 1998.

Bird J, Bishop DVM, Freeman NH: Phonological awareness and literacy development in children with expressive phonological impairments. J Speech Hear Res *38:*446, 1995.

Camarata S: The application of naturalistic conversation training to speech production in children with speech disabilities. J Appl Behav Anal *26:*173, 1993.

Dodd B, Barker R: The efficacy of utilizing parents and teachers as agents of therapy for children with phonological disorders. Aust J Commun Disord *18:*29, 1990.

*Edwards J, Fourakis M, Beckman M, Fox R: Characterizing knowledge deficits in phonological disorders. J Speech Lang Hear Res *42:*169, 1999.

*Felsenfeld S, Broen P, McGue M: A 28-year follow-up of adults with a history of moderate phonological disorder: Linguistic and personality results. J Speech Hear Res *35:*1114, 1992.

Felsenfeld S, McGue M, Broen PA: Familial aggregation of phonological disorders: Results from a 28-year follow-up. J Speech Hear Res *38:*1091, 1995.

Felsenfeld S, Plomin R: Epidemiological and offspring analyses of developmental speech disorders using data from the Colorado Adoption Project. J Speech Lang Hear Res *40:*778, 1997.

Gierut J: Treatment efficacy: Functional phonological disorders in children. J Speech Lang Hear Res *41:*S85, 1998.

Grigorenko EL, Wood FB, Myers MS, Hart LA, Speed WC, Shuster A: Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet *60:*27, 1997.

*Harbers H, Paden E, Halle J: Phonological awareness and production: Changes during intervention. Lang Speech Hear Serv Schools *30:*50, 1999.

Klein E: Phonological/traditional approaches to articulation therapy: A retrospective group comparison. Lang Speech Hear Serv Schools *27:*314, 1996.

Lewis BA, Freebairn L: Residual effects of preschool phonology disorders in grade school, adolescence, and adulthood. J Speech Hear Res *35:*819, 1992.

Masterson JJ: Classroom-based phonological intervention. Am J Speech Lang Pathol *2:*5, 1993.

Peterson-Falzone SJ: Speech outcomes in adolescents with cleft lip and palate. Cleft Palate Craniofac J *32:*125, 1995.

Ruscello D: Visual feedback in treatment of residual phonological disorders. J Commun Disord *28:*279, 1995.

Shriberg L, Aram DM, Kwiatkowski J: Developmental apraxia of speech I: Descriptive and theoretical perspectives. J Speech Lang Hear Res *40:*273, 1997.

Shriberg L, Aram DM, Kwiatkowski J: Developmental apraxia of speech II: Toward a diagnostic marker. J Speech Lang Hear Res *40:*286, 1997.

Shriberg L, Aram DM, Kwiatkowski J: Developmental apraxia of speech III: A subtype marked by inappropriate stress. J Speech Lang Hear Res *40:*313, 1997.

*Shriberg L, Austin D, Lewis B, McSweeny J, Wilson D: The speech disorders classification system (SDCS): Extensions and lifespan reference data. J Speech Lang Hear Res, *40:*723, 1997.

Shriberg L, Gruber F, Kwiatkowski J: Developmental phonological disorders III: Long-term speech-sound normalization. J Speech Hear Res *37:*1151, 1994.

*Shriberg L, Kwiatkowski J: Developmental phonological disorders I: A clinical profile. J Speech Hear Res *37:*1100, 1994.

Shriberg L, Kwiatkowski J, Gruber F: Developmental phonological disorders II: Short-term speech-sound normalization. J Speech Hear Res *37:*1127, 1994.

Smit AB, Hand L, Freilinger JJ, Bernthal JE, Bird A: The Iowa articulation norms project and its Nebraska replication. J Speech Hear Disord *55:*779, 1990.

Sommers R, Logsdon B, Wright J: A review and critical analysis of treatment research related to articulation and phonological disorders. J Commun Disord *25:*3, 1992.

Stoel-Gammon C: Evaluation of phonological skills in preschool children. Semin Speech Lang *9:*15, 1988.

Weston A: The influence of sentence elicitation variables on children's speech production. J Speech Lang Hear Res *40:*975, 1997.

Yorkston KM: Treatment efficacy: Dysarthria. J Speech Hear Res *39:*S46, 1996.

37.4 STUTTERING

ROBERT KROLL PH.D., AND JOSEPH H. BEITCHMAN, M.D.

DEFINITION

Stuttering is a condition in which the flow of speech or fluency is disrupted by involuntary speech motor events. Stuttering may include a variety of specific disfluency types such as sound and syllable repetitions, sound prolongations, dysthymic phonations, and complete blocking or unusual pauses, or both between sounds and syllables of words. More-severe forms of stuttering are frequently accompanied by a variety of struggle behaviors, commonly referred to as accessory, or secondary, features of the disorder. These struggle behaviors may be observed at the respiratory, phonatory, or articulatory levels of the speech mechanism and may manifest themselves in disordered breathing, glottal fry, lip pursing, and tongue clicking. Additional concomitant behaviors may be observed such as eye blinks, facial grimacing, head jerks, and abnormal body movements prior to or during instances of disrupted speech. Linguistic avoidance is a commonly observed feature of stuttering, which results in word substitution and circumlocution as the individual approaches the feared sound or word during the communicative attempt. Similarly, situational avoidance is often characteristic of chronic stuttering, and the type of communicative situation often influences the amount and severity of stuttering. Stuttering can vary in severity, ranging from mild, occasional disfluencies in speech to severe stuttering that renders communication all but impossible for the affected individual and often results in serious psychological difficulty and social withdrawal.

HISTORY AND COMPARATIVE NOSOLOGY

Psychiatric syndromes can be defined from the point of view of the observer or the patient. Depressed individuals may not look depressed, but if they report subjective symptoms sufficient to meet criteria, a diagnosis of a depressive disorder can be made. Likewise, stuttering can be defined from the point of view of the listener or the speaker. The speaker's sense of subjective distress or loss of control of the speech apparatus reflects an approach to nosology that focuses on the speaker's experience. The *Diagnostic and Statistical Manual of Mental Disorder* (DSM), on the other hand, focuses on the observer's point of view, that is, on the observable behaviors.

In both the third edition of DSM (DSM-III) and the revised third edition (DSM-III-R) stuttering was placed in the category of disorders usually first evident in infancy, childhood, or adolescence. In both these systems stuttering was coded on Axis I, along with clinical psychiatric syndromes. This remains controversial, as stuttering is not commonly viewed as a psychiatric disorder among speech-language professionals, and speech-language pathologists have argued that it should not be included in DSM at all.

In the fourth edition of DSM (DSM-IV), stuttering is categorized as a communication disorder. This is consistent with the speech-language pathology classification in which all speech and language disorders are kept within the same category. However, it remains to be shown that it belongs in this category; aside from the similarity of causing communicative difficulties, stuttering may differ fundamentally from the other communication disorders in this category. The 10th revision of *International Statistical Classification of Disease and Related Health Problems* (ICD-10) separates stuttering (stammering) from the speech and language disorders and places it in the category of other behavioral and emotional disorders with onset usually occurring in childhood and adolescence. The diagnostic criteria in ICD-10 are similar to the DSM-IV diagnostic criteria, but a few differences are notable.

DSM-IV requires that the disorder interfere with academic or occupational achievement or social communication, whereas the ICD-10 diagnosis (Table 37.4-1) applies only if it is severe enough to cause marked disturbances in fluency. ICD-10 requires a minimum duration of 3 months; DSM-IV has no duration criterion. DSM-IV further specifies that the disturbance must exceed what is appropriate for the individual's age, and if a speech-motor or sensory deficit is present, the speech difficulties must exceed those usually associated with these problems. ICD-10 has an additional category for cluttering, which is not accorded separate status DSM-IV (Table 37.4-2).

A. Stuttering (i.e., speech characterized by frequent repetition or prolongation of sounds or syllables or words, or by frequent hesitations or pauses) is persistent or recurrent and sufficiently severe to cause marked disruption of the fluency of speech.
B. Duration of the disorder is at least 3 months.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 37.4-1 ICD-10 Diagnostic Criteria for Stuttering

A. Cluttering (i.e., rapid speech with breakdown in fluency, but no repetitions or hesitations) is persistent or recurrent and sufficiently severe to reduce significantly the intelligibility of speech.
B. Duration of the disorder is at least 3 months.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 37.4-2 ICD-10 Diagnostic Criteria for Cluttering

The differences between the ICD-10 and DSM-IV classifications of stuttering reflect the ongoing controversy about the true nature of stuttering, whether it should be regarded as part of the broader speech and language diathesis or as a separate entity. Similarly, whether stuttering should be considered a psychiatric disorder at all

remains controversial. Clinically it appears to differ qualitatively from more prototypical psychiatric syndromes (e.g., schizophrenia, autistic disorder, depressive disorders, and attention-deficit/hyperactivity disorder). However, those who suffer from stuttering show psychological distress that interferes with function and impairs enjoyment of life. The cause is unknown, but psychological factors may exacerbate or ameliorate the condition. Until the etiology is firmly rooted, the precise classification of this condition is likely to remain controversial.

EPIDEMIOLOGY

Surveys conducted chiefly in the United States and Europe indicate that the prevalence of stuttering in the general population is approximately 1 percent. Although data regarding incidence vary widely, most specialists would agree that the likelihood of an individual ever exhibiting some form of stuttering is between 3 and 4 percent. Of those who begin to stutter, a large portion stop by adulthood. The estimated percentages of spontaneous recovery in stuttering again vary widely because of differing definitions of stuttering and normal nonfluency, age groups studied, and data collection techniques. Several studies have estimated spontaneous recovery to be as high as 80 percent. Stuttering is more prevalent among males than females, with the sex ratio among American school children being approximately 3 to 4. A high familial incidence is indicated by the presence of stuttering among the relatives of about 50 percent of individuals who stutter. Additionally, a high degree of concordance of stuttering exist in identical twins. Individuals who stutter appear to have normal developmental histories except for speech and language. Numerous studies have demonstrated that young stuttering children are somewhat delayed in early language and articulation development; however, studies of older stuttering children appear to demonstrate that language skills are once again age appropriate. With regard to personality and social and emotional adjustment, children who stutter are often regarded by their parents or primary caregivers as nervous, and prone to specific fears. They are often described as overly sensitive to their environment. Adults who stutter tend to be somewhat more anxious and emotionally reactive than their nonstuttering peers. For the most part studies relating intellectual ability to stuttering are inconclusive; however, there is a significantly higher prevalence of stuttering in the developmentally delayed population.

ETIOLOGY

Researchers and clinicians working in the area of stuttering recognize that the causes of stuttering are complex. Understanding the etiology of stuttering can be facilitated by specifying the level of causation.

The first of these levels pertains to underlying causation. This refers to the genetic, neurophysiological, and psychological factors that may predispose a child to fluency disruption. Research in these areas has been rather cyclic since the 1940s, and various models implicating psychological, learning, and biological variables in stuttering have been postulated.

One early psychoanalytic theory of stuttering postulated that disrupted speech represents the individual's attempt to fulfill some type of unconscious neurotic need and usually results from disturbed early parent-child interaction. Empirical support for this psychoanalytic model is largely lacking, and studies investigating personality factors associated with stuttering have not revealed anything suggesting emotional disturbance or neurosis as a possible cause.

Later psychological models of stuttering have drawn upon the principles of classical and operant conditioning to explain stuttering. Several behaviorally oriented researchers offered a model of operant conditioning in which the early disfluencies of the child were subject to an intermittent schedule of reinforcement, resulting in incipient stuttering. The two-factor theory of stuttering posited that fluency failure resulting from negative emotion becomes associated with speech-related stimuli through a process of classical conditioning. The individual's response to the fluency failure, the repetitions, sound prolongations, and attempts to terminate the instance of disruption, are instrumentally conditioned. Perhaps the best known of modern psychological theories of stuttering is the diagnosogenic (semantogenic) model. This theory postulates that stuttering is caused by parents' diagnoses of the developing child's normal disfluencies as stuttering.

The large proportion of stuttering individuals who have relatives with a similar speech disorder and the high concordance for stuttering in identical twins have long suggested that biological inheritance plays some part in stuttering. Current models of genetic transmission that can account for familial patterns of stuttering provide strong evidence for the biological inheritance of stuttering. However, these genetic models point to a complex interaction between heredity and environment. Studies of identical twins reared apart generally show reduced concordance for stuttering, implicating the environment.

The theory of cerebral dominance, proposed in the late 1920s, was one of the earliest neurophysiological models used to explain the cause of stuttering. It arose from the observation that an unusually high number of individuals who stuttered were either left-handed, ambidextrous or had been shifted to right-handedness in childhood. It was further reported that the onset of stuttering occurred shortly after the shift of handedness. The theory of incomplete or atypical cerebral lateralization for speech has received varying amounts of support, but not until relatively recently did advances in electrophysiological techniques and brain imaging began to yield more definite data on cortical and subcortical processing of speech in stuttering. Several investigators have reported evidence of atypical cortical activation patterns in stutterers, including abnormal electroencephalogram measures and patterns of alpha wave suppression. Most recently, several brain-imaging techniques used to measure regional cerebral blood flow have been applied to the investigation of stuttering. Data from these initial studies using single photon emission computerized tomography and positron emission tomography point to differential activation patterns in stuttering during both oral and silent reading and even immediately following treatment for stuttering. Findings from these studies show increased right hemisphere activation during simple oral reading tasks conducted by stuttering adults. These subjects also show significant activation in the left anterior cingulate cortex during silent reading, a finding interpreted in the framework of anticipatory anxiety. A recent study investigating pretreatment and posttreatment brain activation patterns in stuttering suggests that increased fluency following intensive speech therapy results from greater sensorimotor control over articulatory movements, without eliminating a right hemisphere bias for language processing.

A second etiological level of stuttering involves the precipitating causes, that is, when, where, and under what conditions stuttering is most likely to occur. At this level concern is with the loci of stuttering and the specific speaking situations affecting fluency. Linguistic factors associated with stuttering have led some theorists to view it as a higher-level cognitive dysfunction. The vast majority of stuttering occurs mostly on the initial sounds of words or syllables. Fluency-facilitating conditions include choral reading, repeated reading of the same material, whispering, and speaking under conditions of altered auditory feedback or masking noise. Based upon these and other phenomenological features of stuttering, new theories have implicated higher-level linguistic processes as the fundamental causal elements for speech disruption in stuttering. These processes involve the planning or adjustments of linguistic plans via internal monitoring loops. One such theory, the covert-repair hypothesis, emphasizes the role of internal monitoring in detecting and correcting phonological errors in speech planning and production. The theory uses elaborate models of internal monitoring, self-repair, and priming in attempts to explain both the type of disfluency and the conditions that facilitate speaking.

A third level of causation and an area of active investigation concerns the motor speech and vocal tract dynamics characteristics of stuttering. The vocal tract events generating disrupted speech patterns have been the focal point of much of the physiological research in stuttering for over two decades. It is reasoned that increased knowledge of vocal tract events and the component behaviors of stuttering will ultimately have a direct impact on the assessment, measurement, and modification of these behaviors. Much information is available on the respiratory, laryngeal, and articulatory dynamics of stuttering. Physiological studies indicating differences between stutterers and nonstutterers are closely scrutinized because they may in fact be dealing only with peripheral events surrounding stuttering, not with underlying causation. The above argument notwithstanding, individuals who stutter, as a group, exhibit longer voice reaction times and phonatory onset times than do nonstutterers. This has led to a number of hypotheses implicating the larynx as the primary dysfunctional element in stuttering.

The final level of causation involves the environmental factors associated with the onset and development of stuttering. Much information has been collected on personality traits of young children who stutter and their parents, and psychodynamic, sociological and conditioning factors influencing stuttering and their interaction with linguistic development. Considerable inconsistency exists in data obtained from research on the personality of stutterers. There is little (if any) conclusive evidence that stutterers as a group exhibit a definable character structure or predictable set of personality traits. Moreover, there appears to be much overlap in overall adjustment profiles between stutterers and nonstutterers, with lower social adjustment and self-esteem scores characterizing individuals who stutter. These traits are most often explained by the impact of the stuttering itself on these individuals and their environment.

DIAGNOSIS AND CLINICAL FEATURES

Stuttering is typically characterized by involuntary disruptions in the fluency of verbal expression. These disruptions may be audible or silent and are usually characterized by repetitions or prolongations of sounds or syllables or by tense pauses or the complete cessation or blocking of speech sounds.

Instances of stuttering are frequently accompanied by accessory or secondary features. These secondary behaviors involve both the speech mechanism and related or unrelated body structures. Such behaviors may include eye blinks, facial grimacing, lip and tongue tremors, tics, vocal fry, head jerks, abnormal movements of the limbs, and breathing irregularities. Linguistic and situational avoidance behavior is often observable with confirmed stuttering. The affected speaker will often scan the intended utterance to predict stuttering and then attempt to avoid such disruption by substituting words, revising phrases, circumlocuting, or stopping the entire

communicative process. Situational avoidance is often reported, such as speaking in front of a group or on the telephone. Most adult stutterers report specific sounds, words, or situations that increase stuttering. Most of this information comes from studies carried out on chronic adult stutterers. Children who stutter have not been studied extensively, but it is generally accepted by clinicians that stuttering in its earlier forms is more often restricted to simpler types of disfluency such as repetitions and prolongations and less often involves the more severe struggle and avoidance behaviors. Definitions of stuttering vary widely and are usually based on listeners' observations of the disrupted speech pattern. However, some writers emphasize that the affected individual's sensations of the stuttering moment are crucial to any complete definition. Many forms of stuttering are so subtle that they are undetectable by the listener and are only perceived by the speaker.

While stuttering varies markedly in terms of frequency and severity both within and among individuals who stutter, some fairly predictable events or situations are commonly observed that either facilitate fluency or result in a deterioration of the stuttering (e.g., choral reading, singing, whispering, speaking in the presence of masking noise or under conditions of delayed auditory feedback, speaking when alone). Conditions commonly observed to increase stuttering include talking on the telephone, talking to groups, making introductions and saying one's own name, talking to authority figures, and attempting to communicate during stressful situations such as job interviews or formal presentations.

Although severe forms of stuttering are most often observed in adults, children and adolescents can be similarly affected. Stuttering can have a significant impact on the individual's quality of life, causing significant social, emotional, educational, and vocational problems. The preschool and young school-aged stutterer frequently develops social and school-related anxiety. The adolescent who stutters tends to withdraw from normal social interaction. The young adult stutterer often reports feelings of restricted academic and career choices and often foregoes the opportunity to develop specific talents and career paths.

The DSM-IV diagnostic criteria for stuttering are shown in Table 37.4-3.

A. Disturbance in the normal fluency and time patterning of speech (inappropriate for the individual's age), characterized by frequent occurrences of one or more of the following:
   (1) sound and syllable repetitions
   (2) sound prolongations
   (3) interjections
   (4) broken words (e.g., pauses within a word)
   (5) audible or silent blocking (filled or unfilled pauses in speech)
   (6) circumlocutions (word substitutions to avoid problematic words)
   (7) words produced with an excess of physical tension
   (8) monosyllabic whole-word repetitions (e.g., "I-I-I-I see him.")
B. The disturbance in fluency interferes with academic or occupational achievement or with social communication.
C. If a speech-motor or sensory deficit is present, the speech difficulties are in excess of those usually associated with these problems.

Coding note: If a speech-motor or sensory deficit or a neurological condition is present, code the condition on Axis III.

Table 37.4-3 DSM-IV Diagnostic Criteria for Stuttering

DIFFERENTIAL DIAGNOSIS

The ability to produce fluent speech is a skill that develops as the young child grows and it depends on such factors as motor coordination and timing and linguistic and cognitive development as well as emotional development and maturity. Preschool children quite commonly exhibit disfluency in speech during their communication efforts. The factors that distinguish normal nonfluency from incipient stuttering involve both the type and frequency of the observed disrupted speech patterns. Children who stutter exhibit predominantly part-word repetitions, voiced or unvoiced sound prolongations, and broken or fragmented words, whereas the normally nonfluent child is more likely to demonstrate whole word and phrase repetitions and revisions. Several researchers have emphasized the degree of speech fragmentation as the distinguishing factor between stuttering and normal nonfluency. The child who is developing normal speech patterns usually does not evidence more than 10 percent of any type of disfluency in speech, whereas the stuttering child often exhibits higher amounts of speech disruption. Another distinguishing factor is the presence or absence of struggle during speech attempts. The child who is stuttering often exhibits difficulty in initiating or maintaining breath flow or voicing during speech attempts. The normally developing child is more likely to evidence easy, effortless speech, even if it is intermittently marked by repetitions of whole words or phrases. Unlike the stuttering child, the normally developing child produces speech, voice, and airflow in a continuous, uninterrupted fashion.

Cluttering is another disorder of fluency characterized by rapid speech, reduced intelligibility, and frequent telescoping of words. Cluttering is often marked by grammatical and syntactic errors and is not typically associated with any specific sound or word fears or even limited awareness of the speech disorder. Cluttering has often been associated with a central language imbalance. Prognosis for improvement with treatment for cluttering is usually guarded, in large part because of the individual's general lack of awareness of the speech pattern.

COURSE AND PROGNOSIS

The onset of stuttering is usually between the ages of 2 and 7, peaking at about age 4. The stuttering fluctuates in severity initially, so that the course of development is not steady. Initially, days or weeks may go by with little or no speech difficulty only to be followed by a sharp increase in the amount and type of stuttering behavior. The stuttering pattern changes with time. Despite fluctuations in severity, young children beginning to demonstrate signs of true stuttering do not continue to be disfluent in the same manner in which they began. As children begin to demonstrate more struggle and fragmentation of words and syllables, the stuttering pattern becomes more complex. Moreover, while the stuttering is usually episodic in the beginning, it becomes more chronic as it progresses. In its most advanced form, speech disruptions are accompanied by secondary behaviors such as those described above. Complete recovery from stuttering is rare after the teenage years. Up to 80 percent of stuttering children experience spontaneous recovery, with girls more likely to become spontaneously fluent than boys. Investigations of prognosis in therapy for the older, confirmed stutterer have focused primarily on the influence of personality factors in treatment; subjects with less favorable speech attitudes were less likely to obtain long-term benefit from behaviorally oriented speech therapy for stuttering.

TREATMENT

The vast majority of treatment techniques for stuttering have been conducted on adults. Speech-language pathologists have traditionally practiced two distinct forms of therapy. The first deals directly with the psychological aspects of the problem, that is, the attitudes, feelings, and emotions displayed by the individual who stutters. Treatment procedures use various types of psychotherapy directed at self-acceptance, attitude change, and avoidance and anxiety reduction. The second major speech-therapy approach involves modification of the stuttering response to fluent-sounding speech by the application of systematic steps and rules of speech mechanics. In this framework, therapy procedures attempt to reconstruct the respiratory, phonatory, and articulatory gestures used in speech production, either by stuttering modification or fluency-shaping techniques. Therapy formats for adults who stutter have been intensified, and several major stuttering treatment programs now offer full-day treatment approaches administered in blocks of several weeks. Current research and clinical efforts are directed at maintenance of fluent speech in the posttreatment environment. A significant number of individuals have been reported to exhibit either partial regression or even total relapse of stuttering without continued practice of speech skills following treatment.

Measures of treatment efficacy have been largely hindered by varying definitions of stuttering and successful treatment. Current therapies report success rates ranging from 60 to over 90 percent when the sole dependent variable is perceived fluent-sounding speech immediately following treatment. However, thorough treatment outcome measures for stuttering should include three elements: (1) a significant positive change in speech output is produced (2) the change is generalized across speaking situations, and (3) the change is maintained over time.

Far fewer studies have been done on the effects of speech therapy for young stuttering children. Clinicians tend to be somewhat divided on the issue of direct therapeutic intervention for the preschooler who stutters. Many advocate an indirect approach, incorporating parent counseling and environmental manipulation to treat the disfluent child. This approach incorporates such techniques as parental modeling of relaxed, unhurried speech and reduction of communicative stressors in the home environment. The second approach introduces direct techniques of speech modification or fluency shaping, such as respiratory control and gentle phonatory onsets, as soon as the child's stuttering behavior has been identified.

SUGGESTED CROSS-REFERENCES

Chapter 37, especially Section 37.3 and Section 37.5, contains further information on speech and language disorders and speech sound production difficulties.

SECTION REFERENCES

Adams M: A clinical strategy for differentiating the normally nonfluent child and the incipient stutterer. J Fluency Dis *2:*141, 1977.

*Andrews G: Epidemiology of stuttering. In *Nature and Treatment of Stuttering: New Directions*. RF Curlee, WH Perkins, editors. College-Hill Press, San Diego, CA, 1984.

Andrews G, Harris M: *The Syndrome of Stuttering*. The Spastics Society Medical Education and Information Unit in association with William Heinemann. Medical Books, London, 1964.

*Bloodstein O: *A Handbook on Stuttering*, ed 4. Singular, San Diego, 1995.

Boberg E: Maintenance of fluency. Proceedings of the Banff Conference. Elsevier, New York, 1981.

Brutten EJ, Shoemaker DJ: *The Modification of Stuttering*. Prentice-Hall, Englewood Cliffs, NJ, 1967.

*Cooper E, De Nil LF: Is stuttering a speech disorder? ASHA *41:*10, 1999.

Cox NJ, Seider RA, Kidd KK: Some environmental factors and hypotheses for stuttering in families with several stutterers. J Speech Hear Res *27:*543, 1984.

*Curlee RF, Siegel GM: *Nature and Treatment of Stuttering: New Directions*, ed 2. Allyn & Bacon, Boston, 1997.

De Nil LF, Kroll RM, Houle S, Ludlow CL, Braun A, Ingham RJ: Advances in stuttering research using positron emission tomography brain imaging. ASHA *37:*89, 1995.

Devinsky O, Morrell MJ, Vogt BA: Contributions of anterior cingulate cortex to behavior. Brain *118:*279, 1995.

Drayna DT: Genetic linkage studies of stuttering: Ready for prime time? J Fluency Dis *22:*237, 1997.

Fox PT, Ingham RJ, Ingham JC, Hirsch TB, Downs JH, Martin C, Jerabek P, Glass T, Lancaster JL: A PET study of the neural systems of stuttering. Nature *382:*158, 1996.

Freeman F. Stuttering. In *Speech, Language and Hearing*, vol II, N Lass, L McReynolds, J Northern, D Yoder, editors. Saunders, Philadelphia, 1982.

*Howell P, Au-Yeung J, Pilgrim L: Utterance rate and linguistic properties as determinants of lexical dysfluencies in children who stutter. J Acoust Soc Am *105:*481, 1999.

Ingham RJ, Cordes AK: Self-measurement and evaluating stuttering treatment efficacy. In *Nature and Treatment of Stuttering: New Directions*, ed 2, RF Curlee, GM Siegel, editors. Allyn & Bacon, Boston, 1997.

Johnson W: *The Onset of Stuttering*. University of Minneapolis Press, Minneapolis, 1959.

Kidd KK: Stuttering as a genetic disorder. In *Nature and Treatment of Stuttering: New Directions*, RF Curlee, WH Perkins, editors. College-Hill Press, San Diego, CA, 1984.

Kolk H, Postma A: Stuttering as a covert repair phenomenon. In *Nature and Treatment of Stuttering: New Directions*, ed 2, RF Curlee, GM Siegel, editors. Allyn & Bacon, Boston, 1997.

*Kroll RM, De Nil LF, Kapur S, Houle S: A positron emission tomography investigation of post-treatment brain activation in stutterers. In *Speech Motor Production and Fluency Disorders*, W Hulstign, HMF Peters, editors. Proceedings of the third international conference on speech motor production and fluency disorders. Elsevier, Amsterdam, 1997.

Ludlolw HG, Braun HG: Advances in stuttering research using positron emission tomography brain imaging. ASHA *37:*89, 1995.

Pardo JV, Pardo PJ, Janer KW, Raichle ME: The anterior cingulate cortex mediates processing selection in the Stroop attentional conflict paradigm. Proc Natl Acad Sci USA *87:*256, 1990.

Perkins WH: Stuttering: Why science hasn't solved it. In *Nature and Treatment of Stuttering: New Directions*, ed 2, RF Curlee, GM Siegel, editors. Allyn & Bacon, Boston, 1997.

*Peters TJ, Guitar B: *Stuttering: An Integrated Approach to Its Nature and Treatment*. Williams & Wilkins, Baltimore, 1991.

Petersen SE, Fox PT, Posner MI, Mintun M, Raichle ME: Positron emission tomographic studies of the cortical anatony of single-word processing. Nature *331:*585, 1988.

Pool KD, Devous MD, Freeman FJ, Watson BC, Finitzo T: Regional cerebral blood flow in developmental stutterers. Arch Neurol *48:*509, 1991.

St Louis KO, Myers FL: Management of cluttering and related fluency disorders. In *Nature and Treatment of Stuttering: New Directions*, ed 2, RF Curlee, GM Siegel, editors. Allyn & Bacon, Boston, 1997.

Shames GH, Sherrick CE Jr: A discussion of nonfluency and stuttering as operant behavior. J Speech Hear Dis *28:*3, 1963.

Sheehan JG, Martyn MM: Spontaneous recovery from stuttering. J Speech Hear Res *9:*121, 1966.

Sheehan JG, Martyn MM: Stuttering and its disappearance. Hear Res *13:*279, 1970.

Starkweather CW: Learning and its role in stuttering development. In *Nature and Treatment of Stuttering: New Directions*, RF Curlee, GM Siegel, editors. Allyn & Bacon, Boston, 1997.

Webster RL: *Precision Fluency Shaping Program: Speech Reconstruction for Stutterers*. Communications Development Corporation, Roanoke, VA, 1974.

Wischner GJ: Stuttering behavior and learning: A preliminary theoretical formulation. J Speech Hear Dis *15:*324, 1950.

Webster WG: Principles of human brain organization related to lateralization of language and speech motor functions in normal speakers and stutters. In *Speech Motor Production: Motor Control, Brain Research and Fluency Disorders*, W Hulstijn, FM Peters, PHHM Van Lieshout, editors. Elsevier, Amsterdam, 1997.

Wu JC, Maguire G, Riley G, Fallon J, Lacasse L, Chin S: A positron emission tomography [18F] deoxyglucose study of developmental stuttering. Neuroreport *6:*501, 1995.

CHAPTER 37. COMMUNICATION DISORDERS

37.5 COMMUNICATION DISORDER NOT OTHERWISE SPECIFIED

REBECCA F. DETWEILER PH.D., AND JOSEPH H. BEITCHMAN, M.D.

DEFINITION AND COMPARATIVE NOSOLOGY

Any disorder of communication that does not meet criteria for inclusion as one of the defined communication disorders is coded as a communication disorder not otherwise specified. Examples of disorders subsumed under this inclusive heading are voice disorders and communication impairment secondary to structural anomaly (e.g., glossectomy, tracheostomy). The latter type of disorder is relatively self-evident; therefore, the balance of this section is focused on voice disorders in children and adolescents.

Voice production is a highly sophisticated physiological activity, involving minute muscular movement precisely coordinated with aerodynamic events, all subserved by multiple organ systems. It reflects physical and emotional state and plays a central role in human communication. A voice disorder, as described in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), is any “abnormality of vocal pitch, loudness, quality, tone or resonance.” Given the complexity and sensitivity of the process of voice production, it is not remarkable that these disorders may arise from a multiplicity of factors, including medical disease, emotional stress, psychological conflict, and environmental toxicity.

EPIDEMIOLOGY

Reliable data on the prevalence of voice disorders in the general population are lacking. Estimates of prevalence range from 5 to 25 percent in school-aged children and adolescents. Available data suggest that in younger children, voice disorders are more common among boys than girls; however, in adolescence, prevalence is roughly equal among males and females. In addition, the prevalence of voice disorder is higher in populations with mental retardation, hearing impairment, or disruptive behavior disorders.

Efforts to accumulate data on the incidence of voice disorders have been confounded by discrepancies in the definition of the disorder among studies. Cultural variation in the social acceptability of certain voice characteristics (i.e., low, hoarse voice in females; high-pitched voice in males; strident or hoarse voice in children; hypernasality) has made development of a precise, generally accepted definition problematic.

ETIOLOGY

Organic, behavioral, psychogenic, and environmental factors may all contribute, singly or in combination, to the development of a voice disorder. In patients referred for evaluation of voice symptoms, laryngeal pathology, or structural change to the laryngeal system was observed in 50 to 80 percent. The discrepancy in reported rates of organic pathology is due not only to differences in referral criteria among studies, but primarily to variation in the method of laryngeal visualization used to examine the patient. The application of stroboscopic laryngeal endoscopy in laryngological evaluation has significantly improved the identification of earlier and more subtle forms of organic disease. Stroboscopic illumination provides apparent slowing of the motion of the vocal folds, allowing more-detailed examination of their structure and oscillation patterns. In addition, stroboscopic endoscopy of the larynx is generally performed in the clinic with minimal topical anesthesia on an alert patient; thus the clinician has visual access that previously was obtainable only by direct laryngoscopic examination of patients under general anesthesia.

Voice disorders of organic cause among children and adolescents include laryngeal papilloma, vocal fold nodules, vocal hemorrhage, laryngomalacia, laryngeal webbing (congenital or traumatic), vocal polyps, vocal fold paralysis or paresis, endocrine dysfunction, and laryngeal cancer. Premature infants and children with a history of extended endotracheal intubation are at high risk for development of a number of laryngeal pathologies, including subglottic stenosis, arytenoid dislocation, vocal fold paralysis, and granuloma formation or loss of tissue of the vocal folds. Reflux esophagitis and irritation of the larynx are observed in anorexia nervosa and bulimia nervosa patients and are commonly accompanied by hoarseness and low-pitched voice. Because persistent hoarseness in a child or adolescent, as in an adult, may be a sign of malignancy or another serious medical condition, otolaryngological evaluation is warranted.

Behavioral factors are a major cause of voice disorders in children and adolescents. Habitual, maladaptive vocal behaviors such as throat clearing, use of excessive effort and tension in speaking, or screaming, are often the primary cause of vocal dysfunction in this population. Maladaptive behaviors may give rise to organic lesions of the larynx (e.g., vocal nodules, contact ulcers, chronic laryngitis) or may result in functional dysphonia. Cheerleaders, athletes, and young vocal performers are commonly associated with these types of voice disorders. Less obviously, children with histories of allergies or respiratory disease or who have suffered abuse or trauma, have a history of a learning disorder, disruptive behavior disorder, low academic performance, or substance abuse are also at risk. Research has yet to determine the precise relation of the latter factors to the development of voice disorders, but these children have a significantly higher incidence of voice pathology than the general population. Vocally abusive behaviors may be evidence of underlying psychological problems or may simply be detrimental habits developed spontaneously or imitated from another individual in their environment.

A significant number of voice disorders are primarily or secondarily psychogenic; conversion aphonia or dysphonia and persistent falsetto are classic examples. Conversion voice disorders include total (aphonic) or partial (dysphonic) loss of phonatory function in the absence of identified laryngeal pathology, associated with psychological factors that are judged to be etiologically involved (consistent with diagnostic criteria for conversion disorder as defined in DSM-IV). Persistent falsetto, also referred to as mutational falsetto or puberphonia, is the persistence of high-pitched voice production in adolescents following pubertal maturation of the larynx. In this disorder, the voice fails to change from the higher-pitched, harmonically thinner quality of childhood to the lower, harmonically richer quality of adolescence and adulthood. Though associated primarily with adolescent males, a similar disorder is observed in females. Psychosocial factors observed by clinicians to be associated with this disorder include predominant feminine emotional attachment and self-identification, self-consciousness about physical maturation, or resistance to or rejection of the roles and responsibilities of adulthood.

Functional or organic dysphonias associated with maladaptive vocal behaviors resulting from underlying emotional problems also fall in this category. Thus, a significant proportion of cases of muscle tension dysphonia, vocal nodules, and chronic laryngitis, polypoid degeneration may have a primary or contributory psychogenic cause. Studies indicate that approximately one third of patients evaluated for functional dysphonia received psychiatric diagnoses. Diagnoses in these cases included major depressive disorder and dysthymic disorder, generalized anxiety disorder, conversion disorder, phobias, adjustment disorders with depressed mood or anxiety, posttraumatic stress disorder, personality disorders, and (very rarely) schizophrenia.

In addition, reactive emotional stress is frequently concomitant to vocal dysfunction in cases arising from nonpsychogenic factors. The child or adolescent whose voice differs from those of peers is often subject to teasing; cheerleaders or young performers who lose their voice may suffer anxiety about continued success or popularity. Believing that one's voice is different can contribute to the development of selective mutism. Adults, similarly, have become anxious or depressed when a voice disorder jeopardized their established vocational or interpersonal patterns.

Environmental factors may be significant in the development or perpetuation of a voice disorder. Chronic laryngeal irritation and edema is observed among those exposed to chemical fumes and airborne particles. Second-hand smoke is a frequent cause of laryngeal and respiratory irritation among children, producing wheezing and a hoarse voice in those affected. Inhaled corticosteroid preparations used in the treatment of asthma are implicated in voice changes associated with laryngeal edema and mucosal thickening, as are the propellants used to deliver those medications. Another medication effect with implications for voice dysfunction is mucosal

dehydration produced by several classes of drugs, including antihistamines, diuretic agents, and opioids. Many psychoactive medications have this adverse effect, including several tricyclic and tetracyclic drugs, antipsychotic agents, anxiolytic agents, and particularly the older, high-dosage, low-potency antipsychotic drugs. Drugs with anticholinergic adverse effects (e.g., amitriptyline [Elavil, Endep], chlorpromazine [Thorazine], thioridazine [Mellaril]) and those with sympathomimetic actions (e.g., dextroamphetamine [Dexedrine], methylphenidate [Ritalin]) are known causes of mucosal dryness. Additionally, both intrinsic hormonal cyclic fluctuation (i.e., menstrual cycle) and ingested hormonal preparations may affect vocal function. Even drugs as seemingly innocuous as aspirin may have profound impact on laryngeal health, for example, a premenstrual adolescent who takes aspirin for discomfort and suffers acute vocal hemorrhage during her subsequent cheerleading activity.

The clinician should be alerted to the significance of the drying effect that many psychoactive drugs exert on the laryngeal and oropharyngeal mucosa. This mucosal dehydration reduces lubrication of the vibrating structures of the larynx, potentially compromising a critical protective mechanism against frictional damage to these structures. Use of these medications may exacerbate existing vocal symptoms or trigger the development of voice pathology in predisposed patients. In such cases, the patient should be counseled to maintain adequate oral hydration; consultation with a speech-language pathologist may be warranted if the patient is not currently being seen by one.

DIAGNOSIS AND CLINICAL FEATURES

The essential feature of voice disorder is that one or more parameters of voice production fall outside the normal range for the individual's sex, age, and developmental status. Parameters of voice production include fundamental frequency (pitch), intensity (loudness), voice quality, resonance, and prosody. The diagnosis is primarily perceptual; however, objective measures are increasingly valuable in differential diagnosis, determination of prognosis, and treatment planning.

Fundamental frequency is impaired when the individual's speaking pitch is inappropriately high or low for her or his sex, age, and physical maturation. *Disordered intensity* includes a voice that is inappropriately loud or soft or is inadequate in volume for functional communication in the individual's everyday speaking environment. *Voice quality disturbance* includes voices that are hoarse, breathy, harsh, strident, distracting, or effortful or painful to produce. *Resonance* refers to the quality of nasality in the voice; disorders of this parameter include hypernasality and hyponasality. Finally, *prosodic disturbances* include speech patterns that are monotonic in pitch or loudness, that is, lack the normal inflection of speech, exhibit an abnormally fast or slow rate of production, or include inappropriately spaced or distracting pauses for breath.

The DSM-IV diagnostic criteria for communication disorder not otherwise specified, which subsumes voice disorder, are presented in Table 37.5-1. The criteria for specific disorders of speech and language from the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) are listed in Table 37.5-2.

This category is for disorders in communication that do not meet criteria for any specific communication disorder; for example, a voice disorder (i.e., an abnormality of vocal pitch loudness, quality, tone, or resonance).

Table 37.5-1 DSM-IV Diagnostic Criteria for Communication Disorder Not Otherwise Specified

Table 37.5-2 ICD-10 Diagnostic Criteria for Specific Developmental Disorders of Speech and Language

DIFFERENTIAL DIAGNOSIS

A thorough voice evaluation is the foundation of accurate differential diagnosis of voice disorders. In addition to perceptual assessment, the voice evaluation includes investigation of the patient's history, instrumental measures of the respiratory function for speech, and objective analysis of acoustic output over the entire vocal dynamic range. Comprehensive examination of the patient's personal, medical, social, educational, and vocational history provides critical data relative to factors that may contribute to the development or maintenance of the disorder. Aerodynamic measures and acoustic analysis document deficits in the subsystems serving the perceptual parameters of voice production; identified deficits should be consistent with noted perceptual abnormalities.

Laryngological evaluation includes endoscopic visualization of the larynx, preferably performed under stroboscopic illumination to facilitate minute examination of vocal fold mobility and oscillation patterns. Small lesions and early tissue changes may be difficult to detect by standard indirect laryngoscopic technique. Accurate differentiation of organic and functional pathologies, which may require entirely different treatment approaches, depends upon the quality of laryngeal visualization. In addition, the otolaryngologist may refer the patient for complete medical examination to rule out underlying systemic disease.

Several other pathologies may disturb vocal function and must be differentiated by the clinician from primary disorder of voice. These include stuttering, verbal apraxia, and dysarthria.

Certain symptoms of voice disorder may resemble those of stuttering, including excess physical tension in speech production, laryngeal blocking, and evidence of struggle in production. These symptoms are among the most notable presenting features of spasmodic dysphonia. In general, the two pathologies may be differentiated by the clinician by the time of onset and clinical course of the disease.

Symptoms of spasmodic dysphonia (spastic dysphonia) first appear in early to midadulthood, generally between the ages of 25 and 40. Onset in adolescence or childhood is rare. Stuttering, by contrast, clearly manifests in childhood and is more common in boys than girls. In addition, stuttering shows a typical course of gradual recovery through adolescence, often resolving by age 16. Spasmodic dysphonia demonstrates a progression in which symptoms gradually or suddenly worsen until many patients are left with severe communication impairment. In many cases, symptoms have been unresponsive to voice therapy. Injection of botulinum toxin to paralyze the adductor musculature of the larynx appears to offer these patients the greatest promise of relief at the present time.

Verbal apraxia (also referred to as apraxia of speech) is a disturbance in the programing of speech movements. It is differentiated from voice disorder by its

association with a primary insult to the left cerebral hemisphere. Also, verbal output in apraxic patients is often easier or normal in automatic or overlearned speech output (e.g., automatic social greetings, the song “Happy Birthday”). In most voice disorders, again particularly spasmodic dysphonia, struggle and the strained-strangled quality of production is relatively constant across communication contexts.

When verbal apraxia is observed during childhood speech and language development, it is referred to as *developmental verbal apraxia* (or developmental apraxia of speech). Though the diagnosis is not universally accepted, with some arguing that these children exhibit severe phonological disorder, the literature suggests that this disorder may be differentiated from other developmental articulation disorders by several characteristics, including the coexistence of an oral apraxia or oral sensory deficit, imitative responses that are marked by articulatory groping and struggle, unusual substitutions or distortions, and much more difficulty producing multisyllabic words than monosyllabic ones. In addition, children diagnosed with developmental verbal apraxia often demonstrate associated problems, such as feeding and swallowing problems and difficulty with fine motor coordination or coordination of the extremities. In contrast to patients with acquired verbal apraxia, children with this diagnosis generally do not evince localized neurological insult in neuroimaging studies.

Finally, dysarthria typically affects the laryngeal and resonance musculature as well as that of the respiratory and articulatory systems, resulting in disturbed vocal function. It is differentiated from primary voice disorder by its association with specific neurological insult (e.g., stroke, cerebral palsy, traumatic brain injury) and by the multiplicity of systems involved.

COURSE AND PROGNOSIS

Clinical course and prognosis vary, depending on the nature of the voice disorder, the characteristics and motivation of the patient, and the number of factors functioning to reinforce and maintain the disorder. Straightforward medical disorders that are not complicated by behavioral factors tend to respond well to surgical management and a short course of subsequent voice therapy. Pathologies of this type include vocal polyp and submucous cyst, particularly if the latter is lateral to the vibrating margin of the fold. Other causes, particularly those with habitual maladaptive or abusive vocal behaviors, carry a poorer prognosis. Vocal nodules may be surgically excised but tend to recur unless the underlying vocally abusive behaviors are remediated by an adequate, occasionally extensive, course of voice therapy. Vocal nodules in children are typically not removed, because they may improve spontaneously with the structural change inherent in the physiological maturation of the vocal folds. However, unless the perpetuating behaviors are modified, the pathology tends to persist.

Psychogenic voice disorders tend to show prompt improvement once the underlying emotional problems are resolved. Interdisciplinary management by the speech-language pathologist and psychiatrist or psychologist appears to be most efficacious for these patients.

TREATMENT

Intervention in voice disorders is etiologically and symptomatically determined. Pathologies with an organic basis are typically managed medically or surgically, with adjunctive voice therapy to aid rehabilitation. Behavioral and psychogenic voice disorders are treated by the speech-language pathologist with a course of education, counseling, and behavior modification therapy aimed at reducing, modifying, or eliminating maladaptive vocal behavior. Psychological intervention to address underlying emotional problems is recommended when these issues are identified.

Referral to otolaryngology and speech-language pathology for voice evaluation should be made if any of the following are observed: (1) hoarseness in a child or adult that persists longer than 2 weeks, (2) pain in voice or speech production that persists longer than 2 weeks, (3) a voice quality or difference that either distresses the speaker or distracts the listener, or (4) any unexplained change in the voice. Thorough evaluation is requisite to effective treatment of voice disorders and to resolution of the complex of physiological and emotional factors often associated.

SUGGESTED CROSS-REFERENCES

Section 37.1 and Section 37.2 contain information relative to language disorders in childhood. Section 37.5 provides information on stuttering, which may have symptoms apparently similar to those of certain voice disorders. Information relative to hearing loss and mental retardation may be found in Chapter 34. Information on dysarthria and apraxia is contained in Chapter 35.

SECTION REFERENCES

Andersson K, Schalen L: Etiology and treatment of psychogenic voice disorder: Results of a follow-up study of thirty patients. J Voice *12:*96, 1998.

Andrews ML: *Voice Therapy for Children*. Singular Publishing, San Diego, CA, 1991.

Baker L, Cantwell DP: Factors associated with the development of psychiatric illness in children with early speech/language problems. J Autism Dev Disord *17:*499, 1987.

Benninger MS, Jacobson BH, Johnson AF: *Vocal Arts Medicine: The Care and Prevention of Professional Voice Disorders*. Thieme Medical Publishers, New York, 1994.

Blitzer A, editor: *Neurologic Disorders of the Larynx*. Thieme, New York, 1992.

Cantwell DP, Baker L: Clinical significance of childhood communication disorders: Perspectives from a longitudinal study. J Child Neurol *2:*257, 1987.

Cantwell DP, Baker L: Prevalence, type and correlates of psychiatric diagnoses in 200 children with communication disorder. J Dev Behav Pediatr, *2:*131, 1981.

Gerritsma EJ: An investigation into some personality characteristics of patients with psychogenic aphonia and dysphonia. Folia Phoniatr *43:*13, 1991.

*Giddan JJ, Milling L: Comorbidity of psychiatric and communication disorders in children. Child Adolesc Psychiatr Clin North Am *8:*19, 1999.

Gray SD, Smith ME, Schneider H: Voice disorders in children. Pediatr Clin North Am *43:*1357, 1996.

Johns DF, editor: *Clinical Management of Neurogenic Communicative Disorders*. Little, Brown, Boston, 1990.

Kahane JC: Growth and development of the human prepubertal and pubertal larynx. J Speech Hear Res *25:*446, 1982.

Kaslon KW, Stein RE: Chronic pediatric tracheotomy: Assessment and implications for habilitation of voice, speech and language in young children. Int J Pediatr Otorhinolaryngol *9:*165, 1985.

Ko ML, McConachie H, Jolleff N: Outcome of recommendations for augmentative communication in children. Child Care Health Dev *24:*195, 1998.

Korstanje MJ: Drug-induced mouth disorders. Clin Exp Dermatol *20:*10, 1995.

Lewis JR, Andreassen ML, Leeper HA, Macrae DL, Thomas J: Vocal characteristics of children with cleft lip/palate and associated velopharyngeal incompetence. J Otolaryngol *22:*113, 1993.

*Maddern BR, Campbell TF, Stool S: Pediatric voice disorders. Otolaryngol Clin North Am *24:*1125, 1991.

*Martin FG: Tutorial: Drugs and vocal function. J Voice *2:*338, 1988.

*Matas M: Psychogenic voice disorders: Literature review and case report. Can J Psychiatry *36:*363, 1991.

Moller KT, Starr CD, editors: *Cleft Palate: Interdisciplinary Issues and Treatment*. Decker, Philadelphia, 1993.

Morrison M, Rammage L: *The Management of Voice Disorders*. Singular Publishing, San Diego, CA, 1994.

*Nichol H, Morrison MD, Rammage LA: Interdisciplinary approach to functional voice disorders: The psychiatrist's role. Otolaryngol Head Neck Surg *108:*643, 1993.

Pope CE: Acid reflux disorders. N Engl J Med *331:*656, 1994.

Putnam CE, Orenstein SR: Hoarseness in a child with gastroesophageal reflux. Acta Pediatr *81:*635, 1992.

Ramig LO, Verdolini K: Treatment efficacy: Voice disorders. J Speech Lang Hear Res *41:*S101, 1998.

Sataloff RT: *Professional Voice: The Science and Art of Clinical Care*. Raven, New York, 1991.

Simon BM, McGowan JS: Tracheostomy in young children: Implications for assessment of communication and feeding disorders. Infants Young Child *1:*1, 1989.

*Square PA, Roy EA, Martin R: Apraxia of speech: Another form of praxis disruption. In *Apraxia*, L Gonzalez Rothi, K Heilman, editors. Erlbaum, London, 1996.

Stemple JC: *Voice Therapy: Clinical Studies.* Mosby Year Book, St. Louis, MO, 1993.

Subtelney JD, Orlando NA, Whitehead RL: *Speech and Voice Characteristics of the Deaf*. Alexander Graham Bell Association for the Deaf, Washington, DC, 1981.

Titze IR, Lemke J, Montequin D: Populations in the U.S. workforce who rely on voice as a primary tool of trade: A preliminary report. J Voice *11:*254, 1997.

Zalzal GH, Loomis SR, Fischer M: Laryngeal reconstruction in children: Assessment of vocal quality. Arch Otolaryngol Head Neck Surg *119:*504, 1993.

CHAPTER 38. PERVASIVE DEVELOPMENTAL DISORDERS

FRED R. VOLKMAR, M.D., AND AMI KLIN, PH.D.

DEFINITION

The pervasive developmental disorders are early-onset conditions characterized by delay and deviance in the development of social, communicative, and other skills. The individual lacks interest in the social environment, but unusual behavioral responses to the inanimate environment are typical, including various motor mannerisms (stereotypies), resistance to change, and idiosyncratic interests and preoccupations. In the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) this category includes autistic disorder, Rett's disorder, childhood disintegrative disorder, Asperger's disorder, and pervasive developmental disorder not otherwise specified.

AUTISTIC DISORDER

Definition Autistic disorder, also known as childhood autism, infantile autism, or early infantile autism, is by far the best known of the pervasive developmental disorders. In this condition there is marked and sustained impairment in social interaction, deviance in communication, and restricted or stereotyped patterns of behavior and interest. Abnormalities in functioning in each of these areas must be present by age 3 years. Approximately 70 percent of individuals with autistic disorder function at the mentally retarded level, and mental retardation is the most common comorbid diagnosis.

History In 1943 Leo Kanner first described 11 cases of what he termed “autistic disturbances of affective contact.” These 11 children had an “inability to relate” in usual ways to people, from the beginning of life. Kanner also noted unusual responses to the environment, which could include both stereotyped motor mannerisms and resistance to change (or insistence on sameness) as well as communication idiosyncrasies such as pronoun reversal and tendency to echo language (echolalia). Kanner's phenomenologically based description of this condition remains worthy of review. At the same time he was mistaken about some aspects of the condition, and his early misconceptions took some years to clarify. For example, Kanner believed that the condition was not associated with mental retardation because the children looked “intelligent” and did well on some parts of intelligence quotient (I.Q.) tests. As time went on it became clear that most of these children function in the mentally retarded range. However, consistent with Kanner's original impression, marked scatter in skills—with occasional splinter skills—is often observed. Kanner also mentioned that the condition was congenital but that in his original cases parents were unusually well educated, successful in their occupations, or both. This led to the notion that autistic disorder might somehow result from pathological patterns of care. Considerable evidence suggests that this is not the case. No particular bias in terms of social class distribution exists if factors that determine case ascertainment are controlled. Also, parents of autistic children do not exhibit specific deficits in parenting, do not exhibit unusual personality characteristics, and are not likely to suffer from other psychiatric disabilities at an increased rate (with the possible exception of depression). Finally, Kanner's use of the term “autism” was reminiscent of Leo Bleuler's earlier use of the term to describe the quality of self-centered thinking typically seen in schizophrenia. Although Kanner originally believed that autism and schizophrenia were unrelated, his use of this word suggested a relationship, which caused much confusion for many years. One part of this assumption rested on a presumption of continuity based on severity; that is, the presumption was that since autism was so severe, it must somehow be on a continuum with schizophrenia.

In his report Kanner carefully provided a developmental context for his observations. He emphasized both the centrality of deficits in social relatedness and unusual behaviors in the definition of the condition. During the 1960s, there was much confusion about the nature of autism and its etiology. In the early 1960s a growing body of evidence began to suggest that the condition resulted from a neuropathological process. Difficulties in consensual definitions and confusion about the similarities and differences of autism and childhood schizophrenia were complications. By the 1970s a considerable body of evidence began to suggest the neurobiological basis of the disorder. This included high rates of seizures in children followed over time, persistence of unusual primitive reflexes, and other neurological signs. Follow-up studies of large groups of psychotic children revealed a bimodal distribution of onset of this poorly defined condition. The early-onset group, noted by Israel Kolvin, M. Rutter, and others, appeared to have many of the characteristics described by Kanner. Children with later-onset psychosis (after age 5 or 6) seemed to resemble those with schizophrenia in having delusions and hallucinations. Family history also discriminated the groups; the family members of those with the late-onset-type had higher rates of schizophrenia.

A landmark in classification occurred in 1978 when Rutter proposed a definition of autism based on (1) social delay and deviance that were not just a function of mental retardation; (2) communication problems, again, not just a function of associated mental retardation; (3) unusual behaviors such as stereotyped movements and mannerisms; and (4) onset before age 30 months.

Rutter's definition and the growing body of work on autism were influential in the definition of the condition in the third edition of DSM (DSM-III), in which the condition was first recognized and placed in a new class of disorder—the pervasive developmental disorders.

Comparative Nosology The first and second editions of DSM (DSM-I and DSM-II) did not officially recognize autistic disorder; rather, it was viewed as being on some continuum with schizophrenia. Childhood schizophrenia, particularly very-early-onset childhood schizophrenia, is now known to be quite rare. However, in the 1950s and 1960s a broad view of schizophrenia predominated.

The definition of autism in DSM-III was based largely on Rutter's synthesis of Kanner's original description and subsequent research. The name of the disorder “infantile autism” emphasized its early onset but also reflected a certain lack of developmental orientation in the diagnosis and criteria for it. The criteria were monothetic and most appropriate to younger children with autism; a category “residual infantile autism” was available for children, adolescents, and adults who once had met criteria for the condition, but no longer did so. This concept was problematic in several respects and reflected the lack of a developmental orientation.

The revised third edition of DSM (DSM-III-R) addressed the concern about the lack of developmental emphasis. Age of onset could be specified before or after 36 months. The condition was diagnosed on the basis of a set of 16 very detailed criteria grouped in the traditional 3 categories of social disturbance, communicative disturbance, and restricted and repetitive behaviors. The greater developmental orientation of DSM-III-R also engendered a broader diagnostic concept, which was probably overinclusive. An additional change in DSM-III-R was the multiaxial placement of autism and other pervasive developmental disorders; the condition was moved to Axis II, and only the two diagnoses of autistic disorder and pervasive developmental disorder not otherwise specified were included.

The term “pervasive developmental disorder” has been controversial. Coined originally in 1980 as a new term for the class of disorders to which autism was assigned, the term was chosen to reflect the fact that multiple areas of functioning are affected in autism and related conditions. However, some investigators and clinicians have objected to the term since it seems to imply a greater severity than always occurs in individual cases. However, by the time of DSM-III-R the term had come into general usage, and in the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), the term was used as a category as well.

In the ninth revision of *International Statistical Classification of Diseases* (ICD-9) the category of psychoses with onset in childhood included the subgroup of infantile autism among others. The use of the term “psychosis” in ICD-9 perpetuated the unfortunate impression that autism was related to adult schizophrenia. In ICD-10 the pervasive developmental disorder category includes childhood autism, atypical autism (in which there is a failure to meet the behavioral or onset criteria), Rett's syndrome, other childhood disintegrative disorder, overactive disorder with mental retardation and stereotyped movements, Asperger's syndrome, other pervasive developmental disorder, and pervasive developmental disorder unspecified (Table 38-1). In contrast to DSM-IV there is an even greater attempt to differentiate subgroups of pervasive developmental disorders. However, the validity of these conditions has been controversial, and definitive data are lacking. Despite the concerns about such distinctions, the working group responsible for DSM-IV decided, partly on the basis of a large international field trial, to try to achieve as much consistency with ICD-10 as possible. In the end they largely succeeded, although some differences in emphasis exist.

Table 38-1 ICD-10 Diagnostic Criteria for Pervasive Developmental Disorders

In DSM-IV, autistic disorder is defined on the basis of behavioral features and age of onset, which must be before 3 years. Behavioral difficulties must include some feature of social disturbance, communicative disturbance, and restricted interests or repetitive behaviors. The definitions of autism in both ICD-10 and DSM-IV are conceptually identical. There was an explicit attempt to avoid the overinclusiveness of DSM-III-R criteria for autistic disorder. In addition, the disorder was returned to Axis I. Criteria number and detail were also reduced.

The field trial for autistic disorder and related conditions in DSM-IV was an international, multisite study that included 977 subjects, 454 with autism and 283 with conditions other than pervasive developmental disorders. Characteristics of the field trial sample and results are summarized in Table 38-2 and Table 38-3. Interrater reliability was assessed for autism and related conditions and was generally good to excellent, particularly for experienced clinicians.

	Clinically Autistic (N = 454)	Other PDDs (N = 240)	Non-PDD (N = 283)
Sex ratio (M:F)	4.49:1	3.71:1	2.29:1
% Mute	54%	35%	33%
$\bar{X}$ Age (SD)	8.99 (7.18)	9.68 (6.57)	9.72 (8.26)
$\bar{X}$ I.Q. (SD)	58.1 (25.6)	77.2 (29.6)	66.9 (23.4)

PPD, pervasive developmental disorder.
Data from Volkmar FR, Klin A, Siegel B, et al: Field trial for autistic disorder in DSM-IV. Am J Psychiatry 151:1361, 1994.

Table 38-2 DSM-IV Field Trial Group Characteristics by Clinical Diagnosis

	Se	Sp	PPV	NPV	TPV	K
DSM-III (current/residual IA)	.82	.80	.78	.84	.81	.62
DSM-III (current IA)	.71	.88	.84	.80	.80	.60
DSM-III-R	.86	.73	.73	.86	.79	.58
DSM-IV/ICD-10	.79	.87	.87	.83	.85	.70

Se, sensitivity; Sp, specificity; PPV, positive predictive validity; TPV, total predictive validity; K, agreement with clinician's diagnosis of autism. Data from Volkmar FR, Klin A, Siegel B et al: Field Trial for Autistic Disorder in DSM-IV. Am J Psychiatry 151:1361, 1994.

Table 38-3 Results of DSM-IV Field Trial: Sensitivity/Specificity for DSM-III, DSM-III-R, and ICD-10

Epidemiology

Prevalence The first epidemiological study of autism was conducted by Victor Lotter in 1966, who reported a prevalence rate of 4.5 in 10,000 children among the entire 8- to 10-year-old population of Middlesex, a county northwest of London. Since then 23 epidemiological studies worldwide reported in the literature have surveyed over 4,000,000 children and identified 1545 individuals thought to have autism. Prevalence rates ranged from 0.7 per 10,000 to 21.1 per 10,000 (Table 38-4). The median prevalence estimate of 4 to 5 per 10,000 is consistent with Lotter's initial value. Variability among studies reflect methodological issues such as sample size, syndrome definition, and aspects of screening and ascertainment. Recent studies have reported higher prevalence rates. Possible reasons for the increased rates involve (1) broader definitions of autism; (2) smaller target populations, since in general, smaller studies have yielded the higher rates, and (3) better detection of cases in the extreme ranges (i.e., severely mentally retarded children and nonretarded individuals with autism). Some studies have also included estimates of the broader spectrum of autistic conditions. Although the reliability for diagnostic ascertainment of these autistic-like conditions remains questionable, compelling evidence indicates that perhaps 1 in every 1000 children may exhibit social disabilities consistent with the autistic spectrum of disorders.

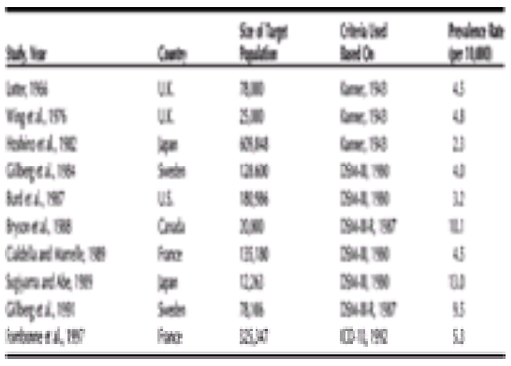

Table 38-4 Representative Epidemiological Studies of Autism

Sex Ratio Studies based on both clinical and epidemiological samples have suggested a higher incidence of autistic disorder in boys than in girls, with reported ratios averaging about 3.5 or 4.0 to 1. This ratio varies, however, as a function of intellectual functioning. Some studies have reported ratios of up to 6.0 or more to 1 in individuals with autism without mental retardation, whereas reported ratios within the moderately to severely mentally retarded range have been as low as 1.5 to 1.

Why females are underrepresented in the nonretarded range remains unclear. One possibility is that males have a lower threshold for brain dysfunction than females or, conversely, that more-severe brain damage is required to cause autism in a girl. According to this hypothesis, a girl with autism is more likely to be severely cognitively impaired.

Social Class Although a few early studies supported Kanner's impression of an association between autism and upper socioeconomic status, epidemiological studies by L. Wing, Eric Schopler, and others have failed to reveal such association. In addition to the bias for more-educated and successful parents to seek referral, families from disadvantaged backgrounds still seem to be underrepresented in clinically referred samples. Outreach initiatives are needed to give children from all socioeconomic backgrounds equal access to diagnostic and intervention services.

Etiology

Psychosocial Theories Kanner's original speculation that emotional factors might be involved in the pathogenesis of autism led others to conclude that the condition was always caused by a “refrigerator” mother who was not responsive to the child's emotional needs. This view was particularly expounded in this country by Bruno Bettelheim, who recommended intensive psychotherapy for mother and child or sometimes removal of the child from the family, in an attempt to remediate the basic deficit. Unfortunately, no evidence showed that such efforts were efficacious. A generation of parents was traumatized by the experience of being blamed for their child's condition.

Biological Theories Factors that suggested a biological basis for the condition included the high rate of mental retardation and seizure disorders and the recognition that various medical or genetic conditions are sometimes associated with the syndrome. The present consensus is that autistic disorder is a behavioral syndrome caused by one or more factors acting on the central nervous system (CNS). While the underlying biological abnormalities of autistic disorder are unknown, efforts are now under way to develop precisely testable neuropathological mechanisms.

Genetic Factors The early impression was that genetic factors had no role in the pathogenesis of autism. The condition is relatively rare and patients did not seem to reproduce. Studies of twins indicated high concordance, especially for monozygotic twin pairs, with reduced concordance for fraternal, or dizygotic, same-sex twin pairs. Evidence also suggested that the high rates of cognitive difficulties in the unaffected monozygotic twin were associated with perinatal complications in the autistic cotwin, suggesting a perinatal insult related to autism in the face of some inherited liability for the disorder. In general, family studies have shown a rate of recurrence in families of approximately 2 to 3 percent of autism among siblings. However, this is 50 to 100 times the rate of autism in the general population. Moreover, parents who are given the early diagnosis and presentation of autism might consciously or unconsciously decide against having additional children. If this phenomenon, “stoppage,” is taken into account, the risk to siblings is even higher. Even when not affected, siblings are at increased risk for various developmental difficulties, including problems in language and cognitive development. It remains unclear whether what is inherited is a specific predisposition to autistic disorder or a more general predisposition to developmental difficulties. Recent work on the family members of autistic persons finds higher rates of mood and anxiety problems and increased frequency of social difficulties. Although the role of genetic factors in autistic disorder is now well established, specific modes of inheritance remain unclear, and efforts are under way to identify potential genetic mechanisms. It is possible and indeed even likely that some genetic forms of autism will be identified over the next few years.

Other Medical Conditions and Autistic Disorder Autistic disorder has also been associated with other conditions with a strong genetic component, most notably fragile X syndrome and tuberous sclerosis. In fragile X syndrome, a marker X chromosome in cells grown in a special medium deficient in folic acid has a site that commonly fractures. The fragile X mutation includes a triplet repeat of cytosine-guanine-guanine (CGG) that may be amplified in succeeding generations. Physical signs of the condition include characteristic facies, enlarged testicles, associated mental retardation, and some autistic features. Behavioral difficulties include attention problems, impulsivity, and anxiety. Initially there was great enthusiasm for the notion that a fragile X chromosome might account for most cases of autism in males. In fact, only about 1 percent of individuals with autism are affected. This condition remains the second most important known chromosomal cause of mental retardation, after Down syndrome.

Tuberous sclerosis is characterized by abnormal tissue growth, or benign tumors (hamartomas), that affect various organ systems. This autosomal dominant disorder is associated with a range of phenotypes including mental retardation and seizure disorder. Studies find tuberous sclerosis in 0.4 to 2.8 percent of autistic individuals, a significantly higher rate than that in the general population. Rates of autistic disorder in individuals with this disorder are high.

Perinatal Factors Several studies have shown increased rates of prenatal, perinatal, and neonatal complications in autistic disorder children. Much of the difference relates to observations that something unusual is noted about the child at birth, which may reflect the operation of both genetic and perinatal factors. The genetic predisposition to autistic disorder may interact with perinatal factors in producing the syndrome.

Other Causes Various reports have associated autistic disorder with a host of other conditions. However, these reports usually describe single cases rather than controlled studies. Several associations have been of interest, for example, autism associated with phenylketonuria, neurofibromatosis, and congenital rubella. However, when children with congenital rubella who were initially thought to have autistic disorder were followed over time, their autistic-like features tended to diminish; also these children exhibit a range of sensory deficits and mental retardation, both of which complicate diagnosis.

Neuroanatomical Models Studies have focused on the cortical and subcortical systems related to language and cognitive processing, that is, on areas of the frontal and temporal lobes, as well as the neostriatum, sensory processing systems, and the cerebellum. A role for the mesial temporal lobe was suggested by dilatation of the temporal horn in the left lateral ventricle observed in early studies using pneumoencephalogram. Subsequent findings by computed tomography (CT) and magnetic resonance imaging (MRI) have been somewhat less consistent. Some autistic individuals have enlarged brains and heads, whereas others (particularly those more retarded) have smaller heads. Neuropathological studies have suggested cellular changes in the hippocampus and the amygdala; increased cell packing has been seen in the amygdala. The cerebellum was the focus of some interest after reduced cerebellar size in the neocerebellar vermal lobules VI and VII was reported; however, other investigators have failed to replicate this finding. Some neuropathological studies have suggested decreased numbers of Purkinje's cells in the cerebellar vermis and hemispheres.

Although animal models of autistic disorder have been attempted, young animals with marked social deficits would be much less likely to be cared for by their parents and thus are at greatly increased risk of mortality. Models of the condition have been attempted by administration of drugs (e.g., amphetamine) to induce motor stereotypy as well as by lesions of certain brain structures.

The severe deficits in language and communication that characterize autistic disorder have suggested the possibility of left cortical involvement to many investigators. Results of studies have, however, been equivocal. Since at least some functions affected in autistic disorder (prosody and language pragmatics) are more likely to be right-hemisphere related, a left hemisphere hypothesis cannot account for all deficits.

Neurochemistry Beginning in 1961 a number of studies have reported that approximately one third of children with autistic disorder have increased peripheral concentrations of the neurotransmitter serotonin. Despite much research the significance of this finding remains unclear since it is not specific to autism and the relation of peripheral concentrations to central concentrations of serotonin is unclear.

Other work has focused on other neurotransmitters, such as dopamine. Hyperdopaminergic functioning of the brain might explain the overactivity and stereotyped movements seen in autism. Administration of stimulants that increase dopamine concentration typically worsens behavioral functioning in autistic disorder. Studies of dopamine metabolites and catecholamine metabolites in cerebrospinal fluid (CSF) have yielded inconsistent results; however, agents that block dopamine receptors are effective in reducing the stereotyped and hyperactive behaviors of many autistic children.

The endogenous opioids were investigated given the possibility that these compounds, enkephalins and endorphins, might lead to social withdrawal and unusual sensitivities to the environment. This was the rationale for using opioid antagonists such as naltrexone (ReVia) to treat children with autistic disorder. Although these agents may have a modest effect on the high levels of activity and agitation, overall results have been disappointing.

Immune Theories Some work has suggested a possible role of immunological factors in autistic disorder. There has been a suggestion that maternal antibodies directed against the fetus may be produced in utero. There also have been reports of autism associated with viral infections.

Diagnosis and Clinical Features A diagnosis of autistic disorder requires at least six behavioral criteria, one from each of the three areas of disturbance in social interaction, communication, and restricted patterns of behavior and interest (Table 38-5). There is a marked range of syndrome expression in autistic disorder. The

lowest-functioning children typically are largely or entirely mute, are isolated from social interaction, and make few social overtures. At the next level children may accept social interaction passively but do not seek it out. At this level, some spontaneous language may be observed. The higher-functioning and somewhat older children have a different social style; they may be interested in social interaction but cannot sustain it in typical ways. The social style of such individuals has been termed “active but odd” in that they often have difficulty regulating social interaction once it has begun. Behavioral features of autistic disorder change over the course of development. Considerable potential for misdiagnosis exists, especially at the extremes of intellectual functioning. Evaluation of the autistic child should include a detailed history, comprehensive medical and developmental examination, and psychological testing.

Table 38-5 Diagnostic Criteria for Autistic Disorder

Age at Onset The onset of autism is almost always before age 3 years (Fig. 38-1); parents typically become concerned between the ages of 12 and 18 months as language fails to develop. Although they may be concerned that the child is deaf, the parents also note the child may respond quite dramatically to sounds in the inanimate environment; occasionally parents report in retrospect that the child was “too good,” made few demands, and had little interest in social interaction. This is in stark contrast to normally developing infants for whom the human face and social interaction are among the most interesting and salient features of the world. In 20 to 25 percent of cases parents report that children seemed to develop some language, and then their language either plateaued or was lost. Parents almost always report being worried by age 2 and inevitably by age 3 (onset of the condition after age 3 would result in a diagnosis of atypical autism). Occasionally parents of very high-functioning children with autistic disorder may be less concerned in the first year or two of life, but even these parents become concerned before age 3 as the severe deficits in social interaction become more apparent.

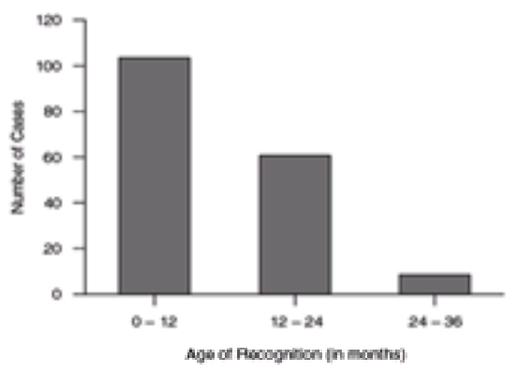

FIGURE 38-1 Age of recognition of autism in a series of 174 consecutive cases. (Reprinted with permission from Lewis M, Volkmar F: *Clinical Aspects of Child Development*, ed 4. Williams & Wilkins, Baltimore, 1996.

Qualitative Impairments in Social Interaction Normally developing infants have a marked interest in social interaction and the social environment from birth. This predisposition is an important foundation for development of other skills. In autistic infants and young children the human face holds little interest, and disturbances are seen in the development of joint attention, attachment, and other aspects of social interaction. For example, the child may not engage in the usual games of infancy, may have difficulties with imitation, and may lack usual play skills. These deficits are highly distinctive and do not just reflect associated developmental delay.

Social interest may increase over time. There is often a developmental progression with younger and more impaired individuals being avoidant or aloof from interaction while somewhat older or more advanced individuals are willing to accept interaction passively but do not seek it out. The most able persons with autistic disorder often display social interest but have difficulty managing the complexities of social interaction, which often leads to an unusual or apparently eccentric social style.

Qualitative Impairment in Verbal and Nonverbal Communication and Play As many as 50 percent of individuals with autistic disorder never speak. Delays in the acquisition of language are the most frequent presenting complaint of parents. Usual patterns of language acquisition (e.g., playing with sounds and babbling) may be absent or infrequent. Infants and young children with autistic disorder may take the parent's hand to obtain a desired object without making eye contact (i.e., as if the hand, rather than the person, is obtaining the item). In contrast to children with language disorder, these children have no apparent motivation to engage in communication or attempt to communicate via nonverbal means.

When individuals with autistic disorder do speak, their language is remarkable in various ways. They may echo what they have heard (echolalia). Speech tends to be less flexible so that for example, there is no appreciation that change in perspective or speaker requires pronoun change, which leads to pronoun reversal. Speech may be nonreciprocal (e.g., the child produces language that is not meant as communication). While the syntax and morphology of language are relatively spared, vocabulary and semantic skills may be slow to develop, and aspects of the social uses of language (pragmatics) are particularly difficult for individuals with autistic disorder. Thus humor and sarcasm may confuse a person with autistic disorder who fails to appreciate the speaker's communicative intent—resulting in an overly literal interpretation of the utterance. Often intonation is monotonic and robotlike.

Deficits in play may include a failure to develop usual patterns of symbolic-imaginative play. Children with autistic disorder may explore nonfunctional aspects of play materials (e.g., taste or smell) or use aspects of materials for self-stimulation (spinning the tires on a toy truck).

Markedly Restricted Repertoire of Activities and Interests Children with autistic disorder often have difficulty tolerating change and variation in routine. For example, an attempt to alter the sequence of some activity may be met with what appears to be catastrophic distress on the part of the child. Parents may report that the child insists that they engage in activities in very particular ways. Changes in routine or in the environment may elicit great opposition or upset. The child may develop an interest in a repetitive activity such as collecting strings and using them for self-stimulation, memorizing numbers, or repeating certain words or phrases. In younger children attachments to objects, when they occur, differ from usual transitional objects in that the objects chosen tend to be hard rather than soft, and often it is the class of object, rather than the particular object, which is important (e.g., the child may insist on carrying a certain kind of magazine around). Stereotyped movements may include toe walking, finger flicking, body rocking, and other mannerisms, which are engaged in as a source of pleasure, or self-soothing. The child may be preoccupied with spinning objects, for example, spending long periods of time watching a ceiling fan rotate.

Associated Features In contrast to Kanner's early belief that children with autism had good cognitive potential, approximately 75 to 80 percent are mentally retarded, with about 30 percent falling within the mild-to-moderate range and about 45 percent being severely to profoundly mentally retarded (Fig. 38-2). The mental retardation is not simply a consequence of negativism or lack of motivation. The typical profile on psychological testing is marked by significant deficits in abstract reasoning, verbal concept formation, and integration skills and in tasks requiring social understanding. Therefore, on the Intelligence Scale for Children, for example, weaknesses are usually obtained on the Similarities and Comprehension subtests. In contrast, relative strengths are usually observed in the areas of rote learning

and memory skills and visual-spatial problem solving, particularly if the task can be completed piecemeal, that is, without having to infer the context, or gestalt, of the task. Therefore, performance on the Block Design and Digit Recall subtests of the Wechsler scales usually correspond to peak performances. The typical preference for rote and sequential tasks rather than reasoning and integrative tasks usually carries the implication that individuals with autistic disorder fail to see “the trees from the leaves,” a difficulty that cuts across functioning modalities, from cognitive testing to communication and social interaction. Given the ubiquity of verbal deficits in autistic disorder, individuals usually have higher performance than verbal scores, particularly in the individuals scoring in the mentally retarded range. Interestingly, some studies have suggested the opposite pattern in individuals with Asperger's disorder.

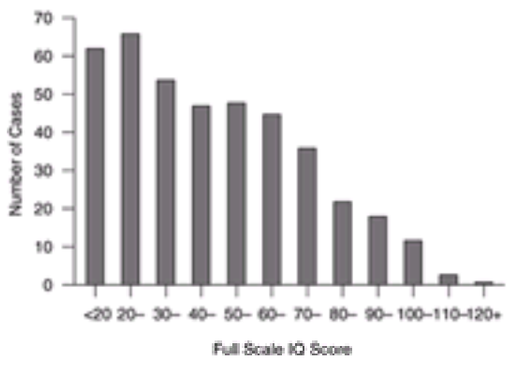

FIGURE 38-2 Full-scale I.Q. in a series of individuals with autism.

Several influential cognitive theories of the social dysfunction in autistic disorder have been proposed. One hypothesis posits a lack of a central drive for coherence, with the consequent focus on dissociated fragments rather than integrated wholes, leading to a fragmentary and overly concrete experience of the world. Another hypothesis posits deficits in executive functioning, that is, in the ability to abstract rules, inhibit irrelevant responses, shift attention and profit from feedback, and maintain a focus on multiple aspects of information in decision making. Since executive functions are thought to be mediated by frontal areas, this hypothesis highlights the similarities between autistic disorder and conditions resulting from frontal lobe lesions. However, this hypothesis lacks specificity for autistic disorder, since a range of developmental psychiatric conditions share such deficits. Another hypothesis, probably the most influential, posits that autistic disorder is caused by the child's inability to attribute mental states (e.g., beliefs and intentions) to others. Devoid of such a “theory of mind,” autistic individuals are presumed unable to infer the thoughts and motivations of others; and thus unable to predict their behavior and adjust accordingly, which results in a lack of reciprocity in communication and social contact. Although the topic of much research, this hypothesis does not account for the very early social abnormalities in autism, and it is not specific to autism.

One of the most fascinating cognitive phenomena in autistic disorder is the presence of so-called islets of special abilities, or splinter skills (i.e., preserved or very highly developed skills in specific areas, which contrast with the child's overall deficits in cognitive functioning). For example, autistic children frequently have great facility in decoding letters and numbers, at times precociously (hyperlexia), even though comprehension of what is read is much impaired. Perhaps 10 percent of individuals with autistic disorder exhibit a form of savant skills—high, sometimes prodigious, performance on a specific skill in the presence of mild or moderate mental retardation. This fascinating phenomenon usually relates to a narrow range of capacities, for example, memorizing lists or trivial information, calendar calculation, visual-spatial skills such as drawing, or musical skills involving perfect pitch or playing a piece of music after hearing it only once. Individuals with autistic disorder represent a disproportionate majority of all savants.

ABNORMALITIES OF MOTOR BEHAVIOR One of the prototypical aspects of autistic disorder is the display of motor stereotypies such as hand flapping, body rocking, and finger twiddling or waving in front of the eyes, as well as other repetitive and purposeless motor mannerisms such as tiptoe walking and assuming odd postures. Stereotypies are also seen in other conditions, particularly severe mental retardation and sensory impairments such as blindness and deafness, but they are clearly more common in autistic disorder, often emerging about ages 3 to 4. Often they are less frequent in adolescence and adulthood and in higher intellectually functioning individuals. Some children with autistic disorder may copy other people's motor movements without necessarily learning the purpose of that movement. This phenomenon, echopraxia, is to some extent the motor equivalent of echolalia in language. Normative motor skills are usually preserved relative to the child's intellectual level, although higher-functioning children can be clumsy and ill coordinated. Developmental course is an important consideration, since a large percentage of children with autistic disorder appear quite graceful and agile in preschool years but awkward and unnatural in adolescence and adulthood. Recent research on motor development has highlighted the ever-increasing control that higher cognitive and social functions exercise on motor skills and presentation, potentially accounting for this developmental course observed in autistic disorder. Overactivity is common in autistic disorder, particularly in preschool years. At the same time a lack of curiosity in the environment and passivity (i.e., hypoactivity) may also be observed in these children. The two may alternate in the same child in different settings or in relation to different activities. Task engagement and attention are important considerations in educational programing.

ABNORMAL RESPONSES TO SENSORY STIMULI Both hypersensitivity and hyposensitivity to sensory stimuli are typical of children with autistic disorder. Theoretical models have proposed that aberrant sensory modulation arises from impaired brainstem attentional mechanisms or from excessive amounts of striatal beta endorphin. Children with autistic disorder may be acutely sensitive to sounds (hyperacusis), for example, covering their ears when hearing a dog bark or the noise of a vacuum cleaner. Other children may appear oblivious to loud noise or people calling them but be fascinated by the faint ticking of a wristwatch or the sound of crumpling paper. Bright lights may be distressing, although some children are fascinated with light stimulation, for example, moving an object back and forth in front of their eyes. There may be extreme sensitivity to touch (tactile defensiveness), including major reactions to specific fabrics and social or affectional touch, but many children appear to be insensitive to pain and may not cry after a severe injury. Many children are fascinated by certain sensory stimuli, such as spinning objects or toy parts that can spin, while some enjoy vestibular sensations such as twirling and engage in this action apparently without becoming dizzy.

SLEEPING AND EATING DISTURBANCES Sleeping and eating disturbances can greatly disturb family life, particularly during childhood. Children with autistic disorder may display erratic sleep patterns with recurrent awakening at night for long periods. Eating disturbances may involve aversion to certain foods because of their texture, color, or smell or insistence on eating a limited choice of foods and refusal to try new foods. In the more severely cognitively impaired children, pica may pose a range of safety issues including the risk of lead poisoning.

MOOD AND AFFECT DISTURBANCES Poor affect modulation and the display of emotions inappropriate to a given social situation are common. Some individuals may show sudden mood changes, and laugh, cry, or giggle to themselves for no apparent reason. Higher-functioning individuals may display intense anxiety in social situations; they may also develop depression in adolescence, usually as a result of their negative social experiences over the years and their partial insight into their condition—knowing that they are different from others without fully understanding their own contribution to the rejecting or otherwise isolating reactions of their peers.

SELF-INJURIOUS BEHAVIOR AND AGGRESSION AGAINST OTHERS Lower-functioning children with autistic disorder may bite their hands or wrists, often causing bleeding and callus formation. Head banging, particularly by severely or profoundly mentally retarded children, may necessitate the use of helmets or other protective devices. Children may also pick their skin, pull their hair, bang their chests, or hit themselves. They have a decreased sense of danger that, along with impulsivity, may lead to injuries. Temper tantrums are common, particularly in reaction to demands (e.g., to comply with a task), changes in routine, or otherwise unexpected events. Lack of understanding or inability to communicate, or sheer frustration may occasionally prompt aggressive outbursts. Although some higher-functioning individuals (e.g., those with Asperger's disorder) have been described as particularly apt to exhibit antisocial behaviors, these individuals are in fact more likely to be victims of practical jokes or other forms of aggression; more commonly, these individuals tend to gravitate toward the periphery of social settings.

SEIZURE DISORDERS Epilepsy develops in approximately 10 to 35 percent of the autistic disorder population by young adulthood, based on the occurrence of readily identifiable major motor seizures (Fig. 38-3). Although there is evidence of seizure onset at all ages, early childhood and adolescence have been reported to be the peak periods. Lower-functioning individuals are at increased risk. The onset of seizures may be associated with deterioration.

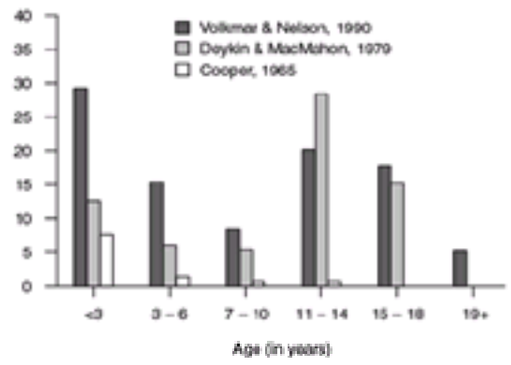

FIGURE 38-3 Rates of first seizure (incidence rate per 1000) in samples of individuals with autism. Data from patients with autism (Volkmar F, Nelson D: Seizure disorders in autism. J Am Acad Child Adolesc Psychiatry *29:*127, 1990) and patients with autism/PDD Deykin EY, MacMahon B: The incidence of seizures among children with autistic symptoms. Am J Psychiatry *126:*1310, 1979) compared with rates in a normative sample Cooper JE: Epilepsy in longitudinal survey of 5000 children. Br Med J *1:*1020, 1965).

PHYSICAL CHARACTERISTICS Children with autistic disorder exhibit a higher incidence of minor physical anomalies such as ear malformations. These anomalies may reflect the embryological period in which the factors responsible for autism act (e.g., ears are formed at about the same time as older regions of the brain). Young children with autistic disorder are often described as very attractive and fail to exhibit any forms of stigmata. Such descriptions are less common with adolescents and adults when the social dysfunction has a cumulative effect on physical appearance.

Pathology and Laboratory Examination

Electroencephalography A variety of electroencephalographic (EEG) abnormalities may be seen in autistic disorder, including diffuse and focal spikes, paroxysmal spike and wave patterns, multifocal spike activity, and a mixed discharge. The prevalance of EEG abnormalities in autistic disorder (in the absence of a clinical seizure disorder) ranges from 10 to 83 percent and depends on the number of recordings and the nature of the sample obtained. The prevelance of abnormal EEGs is significantly higher in individuals with mental retardation and autistic disorder. The high prevalance of EEG abnormalities and seizure disorders in autism were among the first compelling pieces of evidence supporting a biological basis for the condition.

Evoked Potential Recording Both early and more-recent studies of auditory brainstem evoked potentials in autistic disorder indicate that if children with underlying neurological conditions are excluded and age and gender are controlled, no evidence exists for abnormalities in the auditory brainstem pathways. However, abnormalities of cognitive potentials, particularly the auditory P300 (which represents the brain's processing of sensory stimuli) have been demonstrated in autistic disorder. This presumably reflects abnormalities in higher auditory processing and neural pathways.

Oculomotor and Postural Physiology Studies that focus on oculomotor and postural physiology of motor pathways have found bilaterally symmetric abnormalities in functions that depend on neocortical circuitry, with subtle or no abnormalities in parameters that depend on widely distributed pathways between posterior fossa and cortex and intact oculovestibular reflexes and posterior fossa circuitry.

Neuroimaging Studies

Computed Tomography Although some CT studies have shown enlargement of the lateral and third ventricles in approximately 15 to 45 percent of autistic individuals, several subsequent studies failed to corroborate this finding. Additionally, with the exception of the ventricular finding, the CT scans of subjects participating in these studies were otherwise normal, and since ventricular size was unrelated to all clinical indices examined in these studies, the implications of ventricular enlargement for the pathophysiology of autism are unknown.

Magnetic Resonance Imaging Two MRI studies of total brain area and volume found increased total brain volume above the lower boundary of the brainstem, reflecting increased tissue volume and lateral ventricular volume. A follow-up study reported that the enlargement of the cerebral hemisphere was regional, involving occipital, parietal, and temporal regions but not the frontal lobe. A series of MRI studies focusing on the cerebellar vermis revealed decrease in the midsagittal area of vermal lobules VI and VII, but these findings have not been independently replicated in studies controlling for age and I.Q. A small number of MRI studies of the brainstem revealed a reduction in area, although most studies found no differences from controls; similarly, volumetric studies of hippocampus revealed no abnormalities. While an early MRI study of corpus callosum found no abnormalities in the midsagittal area, a recent study reported decreases in the middle and posterior regions when measurements were adjusted for total brain volume. The latter study involved the same subjects in whom increased volumes of the parietal, temporal, and occipital lobes but not the frontal lobes were found. The dissociation between the sizes of the cerebral cortex and corpus callosum was interpreted as evidence of abnormal development of neural connectivity between the hemispheres.

Neuropathological Studies Autopsy studies of a small number of autistic individuals have failed to reveal gross brain abnormalities. However, microscopic analysis has revealed reduced neuronal size and increased cell packing density in the hippocampus, amygdala, mammillary body, anterior cingulate cortex, and septum. These areas are known to be related to each other by interconnecting circuits making up a major portion of the limbic system of the brain. There is evidence of decreased numbers of Purkinje's cells and granule cells in the neocerebellar cortex.

Differential Diagnosis Autistic disorder must be differentiated from other pervasive developmental disorders and other developmental disorders (e.g., mental retardation and communication disorders) as well as from sensory impairments, particularly deafness. Both history and current examination help in differentiating autistic disorder from other pervasive developmental disorders (Table 38-6). The onset of autistic disorder is usually well before the third birthday. The unusual patterns of onset noted in Rett's disorder and childhood disintegrative disorder are not observed. Some potential for confusion with Asperger's disorder exists, since a diagnosis of autistic disorder takes precedence over Asperger's disorder. The early preservation of language skills in Asperger's disorder and the later age of parental concern are helpful diagnostic features. Sometimes, extensive questioning elicits the observation that parents noted some minimal abnormality or oddity in the child with Asperger's disorder in the first 3 years of life; usually this should not be taken to mean that the child meets criteria for autistic disorder.

Feature	Autistic Disorder	Asperger's Disorder	Rett's Disorder	Childhood Disintegrative Disorder	Pervasive Developmental Disorder NOS
Age at recognition (months)	0–36	Usually >36	5–30	>24	Variable
Sex ratio	M>F	M>F	F	M>F	M>F
Loss of skills	Variable	Usually not	Marked	Marked	Usually not
Social skills	Very poor	Poor	Varies with age	Very poor	Variable
Communication skills	Usually poor	Fair	Very poor	Very poor	Fair to good
Circumscribed interests	Variable (mechanical)	Marked (facts)	NA	NA	Variable
Family history of similar problems	Sometimes	Frequent	Not usually	No	Unknown
Seizure disorder	Common	Uncommon	Frequent	Common	Uncommon
Head growth decelerates	No	No	Yes	No	No
I.Q. range	Severe MR to normal	Mild MR to normal	Severe MR	Severe MR	Severe MR to normal
Outcome	Poor to fair	Fair to good	Very poor	Very poor	Fair to good

Adapted with permission from Volkmar FR, Cohen D: Nonautistic pervasive developmental disorders. In Psychiatry, R Michels, editor. Lippincott-Raven, Philadelphia, 1998.

Table 38-6 Differential Diagnostic Features of Autism and Nonautistic Pervasive Developmental Disorders

In mild-to-moderate mental retardation social skills are usually consistent with abilities in cognition and communication. In persons with severe and profound retardation the frequency of both autistic disorder and autistic-like behaviors increases. A common source of diagnostic error is the confusion of stereotypies in such

individuals. In higher-functioning individuals there is potential confusion with personality disorders that involve social isolation (e.g., schizoid personality disorder).

Videotapes or other aids to memory may facilitate in obtaining a history. It also may help to ask the parent to recall the child at particular dates (e.g., third birthday). In autistic disorder usually there is no history of normal development, although about 20 to 25 percent of cases there may be a question of an apparent period of normal development prior to either developmental stagnation or developmental regression. A history of clearly normal development raises the possibility of selective mutism, Rett's disorder, childhood disintegrative disorder, language disorder, schizophrenia, and degenerative CNS disorders. Typically children who have experienced severe neglect exhibit delayed or deviant social skills, but other features of autistic disorder are not present, and the social deficits remit with appropriate care. In schizophrenia of childhood onset there is usually a long period (many years) of normal or near normal development prior to the onset of characteristic delusions, hallucinations, and so forth. In some instances the diagnosis can be made with certainty only over time.

John was the second of two children born to middle-class parents after normal pregnancy, labor, and delivery. As an infant, John appeared undemanding and relatively placid; motor development proceeded appropriately, but language development was delayed. Although his parents indicated that they were first concerned about his development when he was 18 months of age and still not speaking, in retrospect they noted that in comparison to their previous child, he had seemed relatively uninterested in social interaction and the social games of infancy. Stranger anxiety had never really developed, and John did not exhibit differential attachment behaviors toward his parents. Their pediatrician initially reassured John's parents that he was a “late talker,” but they continued to be concerned. Although John seemed to respond to some unusual sounds, the pediatrician obtained a hearing test when John was 24 months old. His hearing appeared adequate for development of speech, and John was referred for developmental evaluation. At 24 months, motor skills were age appropriate, and John exhibited some nonverbal problem-solving skills close to age level. His language and social development, however, were severely delayed, and he was noted to be resistant to changes in routine and unusually sensitive to aspects of the inanimate environment. His play skills were quite limited, and he used play materials in unusual and idiosyncratic ways. His older sister had a history of some learning difficulties, but the family history was otherwise negative. A comprehensive medical evaluation revealed no EEG and CT abnormalities; results of genetic screening and chromosome analysis were normal as well.

John was enrolled in a special education program, where he gradually began to speak. His speech was characterized by echolalia, extreme literalness, a monotonic voice quality, and pronoun reversal. He rarely used language in interaction and remained quite isolated. By school age, John had developed some evidence of differential attachments to family members; he also had developed a number of self-stimulatory behaviors and engaged in occasional periods of head banging. Extreme sensitivity to change continued. Intelligence testing revealed marked scatter, with a full-scale I.Q. in the moderately retarded range. As an adolescent, John's behavioral functioning deteriorated, and he developed a seizure disorder. Now an adult, he lives in a group home and attends a sheltered workshop. He has a rather passive interactional style but exhibits occasional outbursts of aggression and self-abuse. (Reprinted with permission from Volkmar F: Autism and the pervasive developmental disorders. In *Child and Adolescent Psychiatry: A Comprehensive Approach*, ed 2, M Lewis, editor. Williams & Wilkins Baltimore, 1996.

Course and Prognosis Autistic disorder is a lifelong disability, and most individuals affected with this condition remain unable to live independently and require family or community support or institutionalization. However, most children with autistic disorder show improvement in social relatedness, communication, and self-help skills with increasing age. Several factors are thought to predict course and long-term outcome, particularly the presence of some communicative speech by the age of 5 or 6, nonverbal intellectual level, severity of the condition, and response to educational intervention. Younger children more typically display the “pervasive” unrelatedness typically included in earlier diagnostic systems. Although some evidence of differentiated responsiveness to parents may be observed as the child reaches elementary school, patterns of social interaction remain quite deviant. For example, enjoyable social interaction may be restricted to circumscribed activities, such as “rough-and-tumble” or “tickling” games, whereas left to their own devices, children with autistic disorder may be withdrawn or absorbed with trivial aspects of the environment (e.g., light switches) or isolating activities (e.g., puzzles). That notwithstanding, gains in social compliance and communication are often made during elementary school years, particularly if structured, individualized, intensive interventions are in place. During adolescence, some autistic disorder children may exhibit behavioral deterioration; in a minority of these, the decline in language and social skills may be associated with the onset of a seizure disorder. Various interactional styles can be observed, ranging from aloof to passive to eccentric (e.g., children attempt to initiate contact with others, but do so in a very awkward or rigid fashion); these styles are related to developmental level. Depressive and anxiety symptoms may appear in higher-functioning adolescents, who become painfully aware of their inability to form friendships despite a desire to do so, and who begin to suffer the cumulative effect of years of failed contact with others and teasing by peers.

Several long-term outcome studies suggest that approximately two thirds of autistic disorder children have a poor outcome (unable to live independently), with perhaps only one third able to achieve some personal independence and self-sufficiency as adults. Among the latter, most may have a fair outcome (social, educational, or vocational gains despite behavioral and other difficulties), and a minority (about a tenth of all individuals with autistic disorder) may have a good outcome (successful work placement and some social life). However, the autistic person is rarely capable of being self-supporting as an adult, having reciprocal friendships, and getting married. Nevertheless, several factors must be considered when evaluating outcome data. For example, much of the available outcome information is based on data collected at a time when institutionalization of autistic disorder patients was commonplace, schooling was not mandatory, early diagnostic services were not accessible, and early intervention services did not exist. There is some reason to hope that over the past decade, the increased interest in early screening techniques and early diagnostic services, more intensive and early interventions, and the current focus on realistic life skills such as self-sufficiency and vocational training have improved the long-term outcome for those with the disorder.

Treatment The goals of treatment for children with autistic disorder are to reduce disruptive behaviors and to promote learning, particularly language acquisition and communication and self-help skills. These goals are best achieved after a comprehensive assessment has determined a profile of strengths and needs and a highly structured and individualized intervention program is put in place that addresses the child's weaknesses and maximizes the child's assets. This program should usually take place in a special education setting and should be carried out by professionals experienced in working with children with autistic disorder. A recent review of several treatment studies highlighted the importance of intensive, early intervention to individual outcome. Treatment goals should be updated regularly, because different issues should be prioritized as a function of the child's rate and profile of progress and age (e.g., vocational training and independent living skills should be aggressively promoted in adolescence). Additionally, psychopharmacological treatments may be beneficial when specific symptoms are not amenable to other treatment modalities.

Educational Approaches Children with autistic disorder require intensive, highly structured special education, starting as early as the child can tolerate a school routine. Given the challenges involved in teaching autistic disorder children, usually a classroom setting with a low student to teacher ratio is essential. For the more impaired children, a typical hierarchy of priorities should include the ability (1) to tolerate individual adult guidance in performing tasks, (2) to follow a daily routine consistently, (3) to develop communication intent and communication means, and (4) to move from associative to conceptual learning. Learning should take place in an environment that minimizes such distractions as nearby windows or richly decorated walls; the more-challenged children may require individual work stations and forceful adult intrusion. Highly predictable, consistent routines are necessary to promote the child's own internal sense of order, scheduling, and organization of experiences and thus more-systematic learning. Children with autistic disorder often learn skills in a highly circumscribed fashion, exhibiting a capacity only in a very concrete and specific setting. Therefore, one must ensure that the child both begins to use the skill spontaneously (e.g., uses new words for requesting in an unpromoted fashion) and generalizes from the setting in which the skill was acquired to a different setting (e.g., uses the new words at home and in other settings where they would be appropriate). An improved ability to predict impending activities and to express one's wishes and protest usually help reduce frustration and associated disruptive behaviors.

Speech and language therapy focuses on the use of words for meaningful communication. Children with autistic disorder may acquire a considerable vocabulary that is dissociated from the act of communication. Therefore, language acquisition should go hand in hand with promotion of the child's intent to communicate with others. For this purpose, vocabulary expansion should focus on words that are relevant to the child's attempt to negotiate the demands of everyday life. Children who do not vocalize should be engaged in programs that use alternative forms of communication, including signs, communication boards, or other forms of augmentative communication. The use of nonvocal forms of communication should not preclude the simultaneous use of words by children whose vocalizations are just emerging and for whom vocal communication is a realistic goal. For older or higher-functioning children, the educational program should focus intensively on social and communication skills training. Positive actions in frequently troublesome situations may have to be rehearsed and scripted; concrete social and communication skills—eye gaze, voice modulation, gestural communication, posture, proximity, greeting behaviors, rules of conversation, and social expectations—may have to taught in a very explicit fashion. Social and communication skills therapy may have to alternate between small group instruction (where appropriate behaviors can be practiced and supportive feedback can be gained) and naturalistic settings (where the newly acquired skill can be practiced or where additional target problematic behaviors can be identified for practice in the small setting). Successful techniques include modeling of behaviors by an instructor, self-observation, role-playing, and the use of individualized social stories.

Behavior Therapy Behavior therapy, which is based on learning theory principles, uses behavior modification technique to establish desired behaviors and eliminate

problem behaviors. Most educational programs for children with austistic disorder use behavioral management technique, although they vary in how integrated these procedures are in the comprehensive educational program. Behavior therapy is particularly useful in managing disruptive behaviors, which range from difficulties in attention and compliance to tantruming and self-injurious behaviors. After a functional analysis of the target behavior is performed and patterns of reinforcement are identified, techniques such as shaping, prompting, and extinction are used to promote the desired alternative behavior, which is then reinforced by increasingly mature rewards. Behavior therapy is also used to facilitate learning, including the promotion of early cognitive skills such as categorization and elicitation of vocalization and speech. Given the autistic child's tendency to learn things in isolation, skills learned through behavior interventions must be relevant to the child's adaptation and must be increasingly used spontaneously and in different settings. Thus, an appropriate behavioral curriculum should place special emphasis on generalization and self-initiated skills. Although most agree that children with autistic disorder profit from behavior therapy, it is still unclear whether some claims of extraordinarily good outcome for autistic children receiving intensive behavior treatment are justified. One must also ensure that higher-level forms of teaching are not more appropriate to a given child (e.g., a more intellectually able child) before most of the educational resources are focused on behavior therapy.

Psychotherapy With the shift from a psychogenic to a biological understanding of autistic disorder, it became clear that psychodynamic psychotherapy and unstructured play therapies in general were not appropriate for the treatment of young children. Individual psychotherapy may be appropriate for higher-functioning individuals who may present with anxiety and depressive symptoms as they grow older and become more aware of their differences and difficulties relating to others. In these instances, psychotherapy should focus on rather explicit problem-solving skills rather than being insight oriented, with the goal of promoting better adjustment and self-satisfaction.

Psychopharmacology No pharmacological agent has proved curative, but certain medications may be of benefit for specific symptoms such as self-injury, aggression, stereotyped movements, and overactivity. Because autistic individuals are often enrolled in educational and behavioral programs, it may be possible to have staff participate, for example, by collecting behavioral data that can be used to monitor the effectiveness of the drug. As with all medications the potential benefits and adverse effects should be considered; for example, the major tranquilizers carry a potential for oversedation, which can be mistaken for a positive therapeutic response. Prior to beginning drug treatment, baseline laboratory studies (tests of liver and renal function, electrocardiogram [ECG], urinalysis, blood count, weight, blood pressure) should be conducted. An examination for abnormal movements is usually performed (typically with the Abnormal Involuntary Movements Scales [AIMS]).

ANTIPSYCHOTICS The dopamine receptor antagonists are the most extensively studied agents in autistic disorder. At relatively low dosages they may decrease stereotyped behaviors and agitation and may help the individual profit from remedial programming. Some data suggest that the combination of these agents with behavior therapy is more effective than either treatment alone. Adverse effects may limit the usefulness of these agents (e.g., sedation, withdrawal, and tardive dyskinesia). The higher-potency neuroleptics have often been used because of their lower likelihood of sedation, although these agents do carry increased risk for dystonias. Recent interest has centered on the serotonin-dopamine antagonists, which appear to offer considerable potential promise.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS Several reports suggested the potential usefulness of the selective serotonin reuptake inhibitors (SSRIs) in autistic disorder and related conditions with the observation of higher peripheral serotonin concentrations in autistic persons. In several controlled studies these medications have proved superior to placebo in treating such symptoms as repetitive behaviors, impulsive aggression, or both. There is a suggestion that seizure disorders may be exacerbated in some cases.

CLONIDINE The use of clonidine (Catapres) was suggested because of the apparently high levels of arousal in autistic disorder. This agent, an a $_2$-adrenergic receptor agonist, reduced noradrenergic activity. It may exert a modest effect on activity levels, but sedation and hypotension can be serious adverse effects.

NALTREXONE Interest in the endogenous opioid system was spurred by early studies that reported high endogenous opioid concentrations. The opioid receptor antagonist naltrexone has been evaluated in several studies. The major effect of this agent appears to be a mild decrease in activity levels. Neither increased social engagement nor facilitation of learning has been observed.

OTHER AGENTS Various other agents have been reported to sometimes be effective in treating at least some of the symptoms of autistic disorder. Such evidence is usually based on single case reports, and controlled studies are often lacking. Lithium (Eskalith) is generally not clearly helpful unless a personal or family history of bipolar disorder exists. Although the available data are limited, in general, psychostimulants appear to exacerbate behavioral difficulties in autistic disorder, probably because of their effect on dopamine. b-Adrenergic receptor antagonists (beta-blockers) have been used in several open trials concerned with reducing aggressive, self-injurious, and impulsive behaviors. These agents may decrease arousal levels, and positive responses have been noted, although the potential for serious adverse effects limits their more general use.

Parents are often interested in so-called alternative treatments, including diet and vitamin therapy. Although they are the focus of much interest, substantive data are lacking. Various somatic treatments have not proven clinically useful. In general, unproved treatments should be avoided, particularly if they are pursued at the expense of educational and behavioral interventions that are known to be efficacious. Treatments that pose actual danger to the child are sometimes proposed and clearly should be avoided. Given the relatively poor prognosis associated with autistic disorder, parents are easily attracted to treatments that propose a rapid cure. Parents should be encouraged to pursue treatments with known efficacy.

RETT'S DISORDER

Definition Rett's disorder is a progressive condition that develops after some months of apparently normal development. Head circumference at birth is normal, and early developmental milestones are unremarkable. Between 5 months and 48 months (usually between 6 months and 1 year), head growth begins to decelerate. Purposeful hand movements are lost, and characteristic midline hand-wringing or hand-washing stereotypies develop. Expressive and receptive language skills become severely impaired and are associated with marked mental retardation. Gait apraxia and truncal apraxia and ataxia develop in the preschool years. A loss of social interactional skills is frequently observed during the preschool years, but social interest often increases later. To date the condition has been convincingly demonstrated only in girls.

History Cases of Rett's disorder were first reported in 1966 by Andreas Rett. His initial report noted the characteristic history and such clinical findings as gait and truncal ataxia and apraxia, breathing difficulties, EEG abnormalities, and occasional seizures. He also observed some features suggesting autism. Subsequently, Bengt Hagberg and colleagues reported similar findings in a new series of patients. The presence of some symptoms suggesting autism, particularly in the preschool years, is the major rationale for including this condition in the pervasive developmental disorder class of disorders in both DSM-IV and ICD-10.

Comparative Nosology Rett's original report provided guidelines for the diagnosis, which subsequently were revised at a 1984 conference on the condition. These Vienna criteria included features necessary for the diagnosis, those that supported it, and those that were exclusionary. Features that support the diagnosis but are not necessarily present include breath-holding spells, periodic hyperventilation, periodic apnea, growth retardation, EEG abnormalities and seizures, dystonia, spasticity, scoliosis, and peripheral vasomotor problems. Features that mitigate against the diagnosis include demonstrable prenatal growth retardation or postnatal CNS trauma. The definitions in DSM-IV and ICD-10 are essentially the same and focus on the features necessary for diagnosis.

Epidemiology Estimates of the prevalence of the condition range from 1 in 15,000 to 1 in 22,000 females. Several thousand cases are now registered with the International Rett's Syndrome Association. Although several males have been reported to have some features of the condition, to date no male has clearly met all criteria.

Etiology Although it is generally agreed that Rett's disorder has a neurobiological basis, the exact cause remains unknown. Rett originally speculated that the condition was associated with high peripheral ammonia concentrations, but this did not prove to be the case. There are no specific biological markers or laboratory tests for the condition. Genetic factors have been implicated by reports of the condition in monozygotic twins and in extended family members, but most cases appear to be sporadic.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for the condition are presented in Table 38-7. Early development of the child is normal. The onset of the condition may be insidious and follow a period of developmental stagnation; it may delay recognition slightly. Over time the developmental delay, decelerated head and body growth, and diminished interest in the environment become quite striking. Previously acquired abilities are lost, including purposeful hand movements. The lack of social interest and potential for misdiagnosis of autistic disorder are greatest in the preschool years, since usually by the time the child reaches school age the autistic-like features are less prominent and development plateaus for a time. At this point severe mental retardation, seizures, and motor problems are of major concern. During this plateau, or pseudostationary phase, breathing difficulties, bruxism, motor problems, and early scoliosis may be noted. Apneic episodes may

alternate with hyperventilation. Most children remain ambulatory until a final period of motor deterioration. The EEG is frequently abnormal, and seizures are common.

A. All of the following: (1) apparently normal prenatal and perinatal development (2) apparently normal psychomotor development through the first 5 months after birth (3) normal head circumference at birth
B. Onset of all of the following after the period of normal development: (1) deceleration of head growth between ages 5 and 48 months (2) loss of previously acquired purposeful hand skills between ages 5 and 30 months with the subsequent development of stereotyped hand movements (e.g., hand wringing or hand-washing) (3) loss of social engagement early in the course (although often social interaction develops later) (4) appearance of poorly coordinated gait or trunk movements (5) severely impaired expressive and receptive language development with severe psychomotor retardation

Table 38-7 DSM-IV Diagnostic Criteria for Rett's Disorder

Darla was born at term after an uncomplicated pregnancy. An amniocentesis obtained because of maternal age was normal. At birth, Darla was in good condition; weight, height, and head circumference were all near the 50th percentile. Her development during the first months of life was within normal limits. At about 8 months of age, her development seemed to stagnate, and her interest in the environment, including the social environment, waned. Her developmental milestones then became markedly delayed; she was just starting to walk at her second birthday and had no spoken language. Evaluation at that time revealed that head growth had decelerated. Some self-stimulatory behaviors were present. Marked cognitive and communicative delays were noted on formal testing. Darla began to lose purposeful hand movements and developed unusual hand-washing stereotyped behaviors. By age 6, her EEG was abnormal, and purposeful hand movements were markedly impaired. Subsequently, she developed truncal ataxia and breath-holding spells, and motor skills deteriorated further. (Reprinted with permission from Volkmar F: Autism and the pervasive developmental disorders. In *Child and Adolescent Psychiatry: A Comprehensive Approach*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.)

Pathology and Laboratory Examination Various nonspecific abnormalities reported in Rett's disorder include elevated copper and ammonia concentrations in the blood, cortical atrophy on brain scan, and EEG abnormalities. Some neuropathological studies have noted decreased brain weight and loss of neurons, with changes in the substantia nigra and caudate nucleus. The results of neurochemical analyses (e.g., endorphins, cortisol, and dopamine) have been contradictory.

Differential Diagnosis Diagnosis of the condition is most straightforward in somewhat older individuals (i.e., school age), but it can be made reliably early in life. The condition differs from autism in its unusual pattern of developmental loss and characteristic clinical features. The potential for incorrect diagnosis of autistic disorder is greatest in the preschool years when many patients with Rett's disorder may exhibit some deterioration of social skills. The onset of childhood disintegrative disorder is typically later than that of Rett's disorder, and the period of normal development is usually much longer in that condition, which also usually affects males. In Asperger's disorder cognitive and language skills are preserved and there is not a marked loss of abilities.

Course and Prognosis Rett's disorder is a progressive neurodegenerative condition. Adult patients may be nonambulatory because of motor problems and scoliosis. There is increased risk of sudden death.

Treatment There is no specific treatment for Rett's disorder. Special education, behavior modification, and physical and respiratory therapies may be useful. No specific pharmacological treatment is available. Given the high rate of seizure disorder, caution is needed in the use of medications that lower the seizure threshold. As with other pervasive developmental disorders, consideration should also be given to support for the parents and siblings of affected individuals.

CHILDHOOD DISINTEGRATIVE DISORDER

Definition Childhood disintegrative disorder is a rare condition characterized by a marked regression in multiple areas of development after several years of normal development.

History Childhood disintegrative disorder was first described by an educator, Theodore Heller, in 1908. He reported a series of patients who displayed a marked and persisting developmental regression after 3 or 4 years of normal development. He originally termed the condition *dementia infantilis;* subsequently it has also been termed *disintegrative psychosis* or *Heller's syndrome*. Over 100 cases have been reported in the years following Heller's report. While the condition is certainly quite rare, it also is probably underrecognized.

Comparative Nosology ICD-9 included a category for disintegrative psychosis defined on the basis of “normal or near normal development in the first years of life, followed by a loss of social skills and of speech together with a severe disorder of emotion, behavior, and relationships.” The condition was not included in DSM-III or DSM-III-R on the presumption that the condition was almost invariably associated with some identifiable general medical condition or progressive neuropathological process; however, review of cases suggests that is not usually the case. Review of cases also suggested important potential differences from autistic disorder, for example in course and outcome.

The definitions of the disorder in DSM-IV and ICD-10 (Table 38-1) are very similar. The ICD-10 definition notes that a general loss of interest in the environment may be observed.

Epidemiology Although available data are limited the disorder appears to be quite rare. Prevalence estimates of 1 in 100,000 children have been suggested. More-recent case series suggest a preponderance of the condition in males. It is likely that some early cases reported in girls were actually cases of Rett's disorder.

Etiology Several lines of evidence suggest the importance of neurobiological factors in the pathogenesis of childhood disintegrative disorder. In about half of cases the EEG is reported to be abnormal, and seizures are sometimes observed. The condition has been associated with various general medical conditions (e.g., the neurolipidoses, metachromatic leukodystrophy, Addison-Schilder's disease, and subacute sclerosing panecephalitis). Although an intensive search for such conditions is always indicated, they are usually not found. Such conditions are more likely if the onset is later (after age 6). Data on other aspects of the neurobiology of the condition are very limited.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for childhood disintegrative disorder are presented in Table 38-8. Early development must be normal for at least 2 years; this should include normal communication and social skills. Before age 10 years there is a regression in at least two (usually many) different areas (e.g., loss of previous skills in communication, social interaction, toileting, or motor abilities) and the development of symptoms similar to those seen in autistic disorder.

A. Apparently normal development for at least the first two years after birth as manifested by the presence of age-appropriate verbal and nonverbal communication, social relationships, play, and adaptive behavior.
B. Clinically significant loss of previously acquired skills before age 10 years) in at least two of the following areas: (1) expressive or receptive language (2) social skills or adaptive behavior (3) bowel or bladder control (4) play (5) motor skills
C. Abnormalities of functioning in at least two of the following areas: (1) qualitative impairment in social interaction (e.g., impairment in nonverbal behaviors, failure to develop peer relationships, lack of social or emotional reciprocity) (2) qualitative impairments in communication (e.g., delay or lack of spoken language, inability to initiate or sustain a conversation, stereotyped and repetitive use of language, lack of varied make-believe play) (3) restricted, repetitive, and stereotyped patterns of behavior, interests, and activities, including motor stereotypies and mannerisms
D. The disturbance is not better accounted for by another specific pervasive developmental disorder or by schizophrenia.

Table 38-8 DSM-IV Diagnostic Criteria for Childhood Disintegrative Disorder

Onset is usually between the ages of 3 and 4 years and may be either abrupt or gradual. There may be nonspecific agitation or anxiety prior to developmental deterioration. Figure 38-4 provides a summary of clinical features. The loss of social and communicative skills is (understandably) of great concern to parents. Stereotyped behaviors, problems with transitions and change, and nonspecific overactivity often develop. Deterioration in self-help skills can be striking and is in contrast to autistic disorder, in which such skills are acquired somewhat later than usual but typically are not lost.

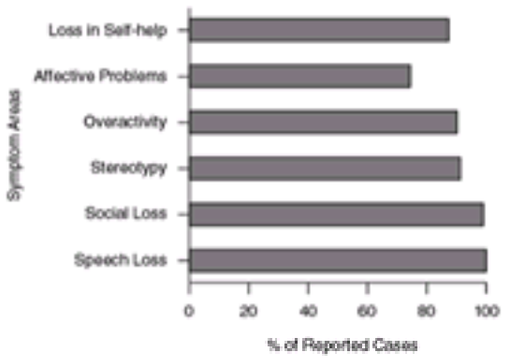

FIGURE 38-4 Clinical features in childhood disintegrative disorder. (Data from Volkmar F, Klin A, Marons W, Cohen DJ: Childhood disintegrative disorder. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.)

Bob's early history was within normal limits. By age 2, he was speaking in sentences, and his development appeared to be proceeding appropriately. At age 40 months he abruptly exhibited a period of marked behavioral regression shortly after the birth of a sibling. He lost previously acquired skills in communication and was no longer toilet trained. He became uninterested in social interaction, and various unusual self-stimulatory behaviors became evident. Comprehensive medical examination failed to reveal any conditions that might account for this developmental regression. Behaviorally, he exhibited features of autistic disorder. At follow-up at age 12 he spoke only an occasional single word and was severely retarded. (Reprinted with permission from Volkmar F: Autism and the pervasive developmental disorders. In *Child and Adolescent Psychiatry: A Comprehensive Approach*, ed 2, M Lewis, editor. Baltimore, Williams & Wilkins, 1996.)

Differential Diagnosis Children with autistic disorder typically exhibit difficulties before age 2 and almost always before age 3 years. Early development is usually not unequivocally normal in autistic disorder, although occasional children are reported to either stagnate in development or experience developmental regression. Baby books or early home movies or videos may help establish that early development was normal. In Rett's disorder there is characteristic head growth deceleration as well as clinical features such as unusual hand movements. In rare instances childhood disintegrative disorder may be confused with schizophrenia, but usually the characteristic findings of schizophrenia on clinical examination will clarify the diagnosis. In the syndrome of acquired aphasia with epilepsy (Landau-Kleffner syndrome) social interest is preserved, and nonverbal communicative skills may be extensive. Specific findings on examination or in the history may help guide diagnostic evaluation.

Course In about three fourths of cases the child's behavior and development deteriorate to a much lower functional level and then stabilize with no further deterioration but only minimal subsequent gains. Some children show more recovery of previous developmental skills; a few have made very good recovery. Deterioration is progressive in some children, particularly in cases associated with a progressive neuropathological process, and death may result. Otherwise life expectancy is normal. In general the outcome appears to be worse than that in autistic disorder.

Treatment As with autistic disorder, special education and behavioral treatments are indicated to help encourage reacquisition of skills. No specific pharmacological treatments exist.

ASPERGER'S DISORDER

Definition Asperger's disorder is characterized by impairments in social interaction and restricted interests and behaviors as seen in autism, but its early developmental course is marked by a lack of any clinically significant delay in spoken or receptive language, cognitive development, self-help skills, or curiosity about the environment. All-absorbing and intense circumscribed interests and motor clumsiness are typical of the condition but are not required for diagnosis.

History In 1944 Hans Asperger, an Austrian pediatrician with interest in special education, described four children who had difficulty integrating socially into groups. Unaware of Kanner's description of early infantile autism published just the year before, Asperger called the condition he described “autistic psychopathy,” indicating a stable personality disorder marked by social isolation. Despite preserved intellectual skills, the children showed marked paucity of nonverbal communication involving both gestures and affective tone of voice; poor empathy and a tendency to intellectualize emotions; an inclination to engage in long-winded, one-sided, and sometimes incoherent speech; rather formalistic speech (Asperger called them “little professors”); all-absorbing interests involving unusual topics, which dominated their conversation; and motoric clumsiness. These children were not less withdrawn or aloof than Kanner's patients; they also developed (sometimes precociously) highly grammatical speech and could not in fact be diagnosed in the first years of life. Discarding the possibility of a psychogenetic origin, Asperger highlighted the familial nature of the condition and even hypothesized that the personality traits were primarily male transmitted. Asperger's work, originally published in German, became widely known in the English-speaking world in 1981, when Lorna Wing published a series of cases with similar symptoms. Her codification of the syndrome, however, blurred somewhat the differences between Kanner's and Asperger's descriptions, as she included a small number of girls and mildly mentally retarded children, as well as some children who had presented with some language delays in their first years of life. Since then, several studies have attempted to validate Asperger's disorder as distinct from autism without mental retardation, although comparing findings has been hindered by the lack of consensual diagnostic criteria for the condition. Although Asperger's syndrome was first granted official recognition in ICD-10 and appears as Asperger's disorder in DSM-IV, its nosological status is still uncertain.

Comparative Nosology Asperger's disorder was not accorded official recognition until publication of ICD-10 and DSM-IV, although it was first reported in the German literature in 1944. Asperger's work was known primarily in German-speaking countries, and only in the 1970s were the first comparisons with Kanner's work made, primarily by Dutch researchers such as D. A. Van Krevelen who were familiar with both the English and German literature. The initial attempts at comparing the two conditions were difficult because of major differences in the patients described—Kanner's patients were both younger and more cognitively impaired. Also, Asperger's conceptualization was influenced by accounts of schizophrenia and personality disorders, whereas Kanner had been influenced by the work of Arnold Gesell and his developmental approach. Several influential researchers in Europe and North America attempted to codify Asperger's prose into a categorical definition for the condition, but no consensual definition emerged until the advent of ICD-10. Given the reduced empirical validation of the ICD-10 and DSM-IV criteria, the definition of the condition is likely to change with new and more rigorous studies in the near future.

Epidemiology Given the lack of consensual definitions of the conditions until recently, it is not surprising that the prevalence of the condition is unknown, although a rate of 1 in 10,000 has been reported. The condition is more prevalent in males than females, with a reported ratio of 9 to 1. In the past few years, parent support organizations for Asperger's disorder have proliferated, and clinicians appear to be giving this diagnosis much more frequently than they did just a few years ago. There are also indications that Asperger's disorder is currently functioning as a residual diagnosis given to normal-intelligence children with social disabilities who do not fulfill criteria for autistic disorder, overlapping in this way with the DSM-IV diagnosis pervasive developmental disorder not otherwise specified. This pattern has diluted the concept and reduced its clinical utility. Empirical validation of specific diagnostic criteria is badly needed, although this must await reports of rigorous studies using standard diagnostic procedures, and validators truly independent of the diagnostic definition such as neuropsychological, neurobiological, and genetic data.

Etiology As with autistic disorder, the cause of Asperger's disorder is unknown. A few case studies reporting the presence of autism and Asperger's disorder in different family members and some recent studies suggesting that Asperger's disorder may be one of the conditions along the autistic spectrum suggest that autistic disorder and Asperger's disorder may be related genetically. However, some reports have described a stronger genetic contribution in Asperger's disorder than in autistic disorder, with more family members and sometimes the child's father either fulfilling criteria for the condition or having similar social difficulties. Asperger himself had alluded to some children who were much more severely involved than those he described and who suffered from varying degrees of mental retardation. His notion that these were children with “autistic psychopathy” who had also suffered brain damage has been taken up by some authors, although no neuroanatomical or neurofunctional data substantiate this view. In fact, there are a few reports of brain anomalies in individuals with Asperger's disorder with otherwise normal or superior I.Q.s.

Diagnosis and Clinical Features Diagnosis of Asperger's disorder (Table 38-9) requires the demonstration of qualitative impairments in social interaction and restricted patterns of interest, criteria identical to those for autistic disorder. In contrast to autistic disorder, there are no criteria in the cluster of language and communication symptoms, and onset criteria differ in that there should be no clinically significant delay in language acquisition, cognitive and self-help skills, symptoms that result in significant impairment in social and occupational functioning.

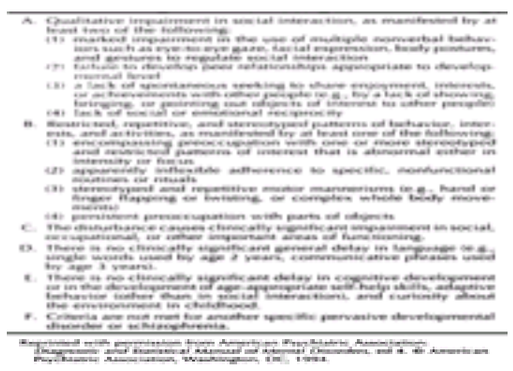
A. Qualitative impairment in social interaction, as manifested by at least two of the following:
(1) marked impairment in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
(2) failure to develop peer relationships appropriate to developmental level
(3) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest to other people)
(4) lack of social or emotional reciprocity
B. Restricted, repetitive, and stereotyped patterns of behavior, interests, and activities, as manifested by at least one of the following:
(1) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
(2) apparently inflexible adherence to specific, nonfunctional routines or rituals
(3) stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting, or complex whole body movements)
(4) persistent preoccupation with parts of objects
C. The disturbance causes clinically significant impairment in social, occupational, or other important areas of functioning.
D. There is no clinically significant general delay in language (e.g., single words used by age 2 years, communicative phrases used by age 3 years).
E. There is no clinically significant delay in cognitive development or in the development of age-appropriate self-help skills, adaptive behavior (other than in social interaction), and curiosity about the environment in childhood.
F. Criteria are not met for another specific pervasive developmental disorder or schizophrenia.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. Washington, DC, American Psychiatric Association, 1994.

Table 38-9 DSM-IV Diagnostic Criteria for Asperger's Disorder

In some contrast to the social presentation in autistic disorder, individuals with Asperger's disorder find themselves socially isolated but are not usually withdrawn in the presence of others; typically they approach others but in an inappropriate or eccentric fashion. For example, they may engage the interlocutor, usually an adult, in one-sided conversation characterized by long-winded, pedantic speech about a favorite and often unusual and narrow topic. They may express interest in friendships and in meeting people, but their wishes are invariably thwarted by their awkward approaches and insensitivity to the other person's feelings, intentions, and nonliteral implied communications (e.g., signs of boredom, haste to leave, and need for privacy). Chronically frustrated by their repeated failures to engage others and form friendships, some individuals with Asperger's disorder develop symptoms of a mood disorder that may require treatment, including medication. They also may react inappropriately to, or fail to interpret, the valence of the context of the affective interaction, often conveying a sense of insensitivity, formality, or disregard for the other person's emotional expressions. They may be able to describe correctly (in a cognitive and often formalistic fashion) other people's emotions, expected intentions, and social conventions; yet they cannot act upon this knowledge in an intuitive and spontaneous fashion and thus lose the tempo of the interaction. Their poor intuition and lack of spontaneous adaptation are accompanied by marked reliance on formalistic rules of behavior and rigid social conventions. This presentation is largely responsible for the impression of social naiveté and behavioral rigidity that these individuals so forcefully convey.

Although significant abnormalities of speech are not typical of individuals with Asperger's disorder, at least three aspects of their communication patterns are of clinical interest. First, speech may be marked by poor prosody, although inflection and intonation may not be as rigid and monotonic as in autistic disorder. They often exhibit a constricted range of intonation patterns used with little regard to the communicative functioning of the utterance (e.g., assertions of fact, humorous remarks). Rate of speech may be unusual (e.g., too fast) or may lack in fluency (e.g., jerky speech), and volume is often poorly modulated (e.g., voice is loud despite physical proximity to the conversational partner). The latter feature may be particularly noticeable as a lack of adjustment to a given social setting (e.g., a library or a noisy crowd). Second, speech is often tangential and circumstantial, conveying a sense of looseness of associations and incoherence. Even though in a very few patients this symptom may indicate a possible thought disorder, the lack of contingency in speech is a result of the one-sided, egocentric conversational style (e.g., unrelenting monologues about the names, codes, and attributes of innumerable TV stations in the country), failure to provide the background for comments and to clearly demarcate changes in topic, and failure to suppress the vocal output accompanying internal thoughts. Third, the communication style of individuals with Asperger's disorder is often markedly verbose. The child or adult may talk incessantly, usually about a favorite subject, often completely disregarding whether the listener is interested, engaged, or attempting to interject a comment or change the subject. Despite such long-winded monologues, the individual may never come to a point or conclusion. Attempts by the interlocutor to elaborate on issues of content or logic or shift the interchange to related topics are often unsuccessful.

Individuals with Asperger's disorder typically amass a large amount of factual information about a topic in a very intense fashion. The actual topic may change from time to time, but it often dominates the content of social interchange. Frequently the entire family may be immersed in the subject for long periods of time. This behavior is peculiar in the sense that often extraordinary amounts of factual information are learned about very circumscribed topics (e.g., snakes, names of stars, television guides, deep fat fryers, weather, personal information on members of Congress) without a genuine understanding of the broader phenomena involved. This symptom is not always easily recognized in childhood since strong interests in certain topics (e.g., dinosaurs or fashionable fictional characters) are so ubiquitous. However, in both younger and older children the special interests typically become more unusual and narrowly focused.

Individuals with Asperger's disorder may have a history of delayed acquisition of motor skills such as pedaling a bike, catching a ball, opening jars, and climbing outdoor play equipment. They are often visibly awkward and poorly coordinated and may exhibit stilted or bouncy gait patterns and odd posture. Neuropsychologically, there is often a pattern of relative strengths in auditory and verbal skills and rote learning and significant deficits in visual-motor and visual-perceptual skills and conceptual learning. Many children exhibit high levels of activity in early childhood, and the most commonly reported comorbid symptoms in adolescence and young adulthood are anxiety and (particularly) depression.

Tom was an only child. Birth, medical, and family histories were unremarkable. His motor development was somewhat delayed, but communicative milestones were within normal limits. His parents became concerned about him at age 4 when he was enrolled in a nursery school and displayed marked difficulties in peer interaction that were so pronounced that he could not continue in the program. In grade school, he was enrolled in special education classes and was noted to have some learning problems. His greatest difficulties arose in peer interaction; he was viewed as markedly eccentric and had no friends. His preferred activity, watching the weather channel on television, was pursued with great interest and intensity. On examination at age 13, he had markedly circumscribed interests and exhibited pedantic and odd patterns of communication with a monotonic voice quality. Psychological testing revealed an I.Q. within the normal range, with marked scatter evident. Formal communication examination revealed age-appropriate skills in receptive and expressive language but marked impairment in pragmatic language skills. (Reprinted with permission from Volkmar F: Autism and the pervasive developmental disorders. In *Child and Adolescent Psychiatry: A Comprehensive Approach*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.)

Differential Diagnosis Unlike children with Asperger's disorder, the great majority of autistic children experience early delays and deviance in language acquisition and cognitive impairment. The differential diagnosis is more difficult when the comparison is made with children with autistic disorder without mental retardation. Asperger's disorder differs from the latter in that the onset is usually later and the outcome more positive. In addition, social and communication deficits are less severe and motor mannerisms are usually absent, whereas circumscribed interest is more conspicuous, motor clumsiness is more common, and a family history of similar problems is more frequent. In both ICD-10 and DSM-IV, if the child meets criteria for autistic disorder, this should take precedence over a diagnosis of Asperger's disorder. The distinction between Asperger's disorder and atypical autism (ICD-10) or pervasive developmental disorder not otherwise specified (DSM-IV) is more difficult, because the latter are basically subthreshold or residual categories without specific defining criteria. Reports attempting to distinguish between them have indicated that the social impairment is more severe and the communication deficits and deviance are less pronounced in Asperger's disorder than in atypical autism or pervasive developmental disorder not otherwise specified. Clinical reports have also emphasized the associated features of all-absorbing and interfering circumscribed interests, verbosity, and motor clumsiness in Asperger's disorder. Individuals with schizoid personality disorder show neither the level of severity of social impairment nor the early developmental patterns seen in Asperger's disorder. The combination of verbosity and severe pragmatic deficits (involving

long-winded, one-sided, and incoherent conversational patterns marked by failure to demarcate changes of topic, to provide background, or to adhere to a communicative message) may lead to the erroneous diagnosis of schizophrenia. However, in the vast majority of cases, these symptoms reflect a communication dysfunction rather than a thought disorder. Again, a detailed developmental history documenting early onset and continuity of symptoms may be important in the differential diagnosis.

Course and Prognosis There are no systematic long-term follow-up studies of children with Asperger's disorder as yet, partially because of nosologic issues. Many children can attend regular education classes with additional support services, although these children are especially vulnerable to being seen as eccentric and being teased or victimized; others require special education services, usually not because of academic deficits but because of their social and behavioral difficulties. Asperger's initial description predicted a positive outcome for many of his patients, who could often use their special talents to obtain employment and so lead self-supporting lives. His observation of similar traits in family members (i.e., fathers) may also have made him more optimistic about ultimate outcome. Although his account was tempered somewhat by the time he had seen 200 patients with the syndrome (25 years after his original paper), Asperger continued to believe that a more positive outcome was a central criterion differentiating individuals with his syndrome from those with Kanner's autism. Although some clinicians have informally concurred with this statement, particularly in regard to gainful employment, independence, and establishing a family, no studies specially addressing the long-term outcome of individuals with Asperger's disorder are currently available. The social impairment (particularly the eccentricities and social insensitivity) is thought to be lifelong.

Treatment As in autistic disorder, treatment of Asperger's disorder is essentially supportive and symptomatic and to a great extent overlaps with the treatment guidelines applicable to individuals with autistic disorder unaccompanied by mental retardation. One initial difficulty encountered by families is proving eligibility for special services. Since these children are often very verbal and many of them do well academically, educational authorities might judge that the deficits—primarily social and communicative—are not within the scope of educational intervention. In fact, these two aspects should be the core of any educational intervention and curriculum for individuals with this condition. Skills, concepts, appropriate procedures, cognitive strategies, and behavioral norms may be more effectively taught in an explicit and rote fashion using a parts-to-whole verbal instruction approach, in which the verbal steps are in the correct sequence for effective behavior. Additional guidelines should be derived from the individual's neuropsychological profile of assets and deficits. Acquisition of self-sufficiency skills in all areas of functioning should be a priority. The tendency of individuals with Asperger's disorder to rely on rigid rules and routines can be used to foster positive habits and enhance the person's quality of life and that of family members. Specific problem-solving strategies (usually following a verbal algorithm) may be taught for handling the requirements of frequently occurring, troublesome situations (e.g., involving novelty, intense social demands, or frustration). Training is usually necessary in recognizing troublesome situations and selecting the best available learned strategy to use in such situations. Social and communication skills are best taught by a communication specialist with an interest in pragmatics in speech, in the context of both individual and small group therapy. Communication therapy should include appropriate nonverbal behavior (e.g., the use of gaze for social interaction, monitoring and patterning of voice inflection), verbal decoding of nonverbal behaviors of others, social awareness, perspective-taking skills, and correct interpretation of ambiguous communications (e.g., nonliteral language). Often, adults with Asperger's disorder fail to meet entry requirements for jobs in their area of training (e.g., college degree) or fail to maintain a job because of their poor interviewing skills, social disabilities, eccentricities, or anxiety attacks. Thus they must be trained for and placed in jobs for which they are not neuropsychologically impaired and in which they will enjoy some support and shelter. Preferably the job should not involve intensive social demands, time pressure, or the need to improvise quickly or generate solutions to novel situations. The little experience available with self-support groups suggests that individuals with Asperger's disorder enjoy the opportunity to meet others with similar problems and may develop relationships around an activity or subject of shared interest. Special interests may be used to create social opportunities through hobby groups. Supportive psychotherapy as well as pharmacological interventions may be helpful in dealing with feelings of despondency, frustration, and anxiety, although a more direct, problem-solving focus is thought to be more beneficial than an insight-oriented approach.

PERVASIVE DEVELOPMENTAL DISORDER NOT OTHERWISE SPECIFIED

Atypical autism in ICD-10 and the equivalent pervasive developmental disorder not otherwise specified in DSM-IV (Table 38-10) essentially refer to a residual category with minimal defining criteria. The diagnosis is used to denote a subthreshold form of autism or an autistic manifestation that is atypical in terms of onset patterns or symptomatology. *Atypical autism* and *pervasive developmental disorder not otherwise specified* are used to describe a rather large and heterogeneous group of children who do not meet strict criteria for autistic disorder or another pervasive development disorder but do exhibit a pattern of developmental and behavioral dysfunction similar to that observed in autistic disorder. Such children typically exhibit unusual sensitivities and affective responses in the presence of more differentiated social relatedness and better cognitive and communicative skills than do most autistic children. It is likely that regularity patterns or new definable syndromes will be identified in this group in the future, thus reducing the number of children with social disabilities who are given what is essentially an undefined diagnosis. Some studies have attempted to delineate reliable and specific criteria for this group, whereas others have proposed new syndromes. For example, several studies have examined the validity of a condition named multiple complex developmental disorder, in which there are basic and early emerging deficits in affective modulation, capacity for relating, and stability of thinking, compounded by anxiety and learning difficulties. Validation studies of this and other proposed syndromes are needed before they can be incorporated in current nosology.

This category should be used when there is a severe and pervasive impairment in the development of reciprocal social interaction or verbal and nonverbal communication skills, or when stereotyped behavior, interests, and activities are present, but the criteria are not met for a specific pervasive developmental disorder, schizophrenia, schizotypal personality disorder, or avoidant personality disorder. For example, this category includes "atypical autism"—presentations that do not meet the criteria for autistic disorder because of late age at onset, atypical symptomatology, or subthreshold symptomatology, or all of these.

Table 38-10 DSM-IV Diagnostic Criteria for Pervasive Developmental Disorder Not Otherwise Specified

Leslie was the oldest of two children. She had been a difficult baby who was not easy to console but whose motor and communicative development seemed appropriate. She was socially related and sometimes enjoyed social interaction, but was easily overstimulated. She exhibited some unusual sensitivities to aspects of the environment and, at times of excitement, exhibited some hand flapping. Her parents sought evaluation when she was 4 years of age because of difficulties in nursery school. Leslie had problems with peer interaction. She was often preoccupied with possible adverse events. At evaluation she displayed both communicative and cognitive functions within the normal range. Although differential social relatedness was present, Leslie had difficulty using her parents as sources of support and comfort. She displayed behavioral rigidity and a tendency to impose routines on social interaction. Leslie was enrolled in a therapeutic nursery school where she made significant gains in social skills. Subsequently, she was placed in a transitional kindergarten and did well academically, although problems in peer interaction and unusual affective responses persisted. As an adolescent, she describes herself as a "loner" who has difficulties with social interaction and tends to enjoy solitary activities. (Reprinted with permission from Volkmar F: Autism and the pervasive developmental disorders. In *Child and Adolescent Psychiatry: A Comprehensive Approach*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.)

SUGGESTED CROSS-REFERENCES

Mental retardation is discussed in Chapter 34, learning disabilities in Chapter 35. Normal child development is discussed in Section 32.2. Detailed information on behavior therapy is presented in Section 30.2, on general principles of psychopharmacology in Section 31.1, on medication-induced movement disorders in Section 31.4. The psychiatric treatment of children is discussed in Chapter 48.

CHAPTER REFERENCES

Asperger H: 'Autistic psychopathy' in childhood (U Frith, translator). In *Autism and Asperger Syndrome*, U Frith, editor. Cambridge University Press, Cambridge, 1944/1992.

Bregman J: Behavioral interventions. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

*Bristol MM, Cohen DJ, Costello EJ, Denckla M, Eckberg TJ, Kallen R, Kraemer HC, Lord C, Maurer R, McIlvane WJ, Minshew N, Sigman M, Spence MA: State of the science in autism: Report to the

National Institutes of Health. J Autism Dev Disord *26:*121, 1996.

Campbell M, Schopler E, Cueva JE, Hallin A: Treatment of autistic disorder. J Am Acad Child Adolesc Psychiatry *35:*134, 1996.

*Fombonne E: Epidemiology of autism and related conditions. In *Autism and Pervasive Developmental Disorders*, FR Volkmar, editor. Cambridge University Press, Cambridge, 1998.

Harris SL, Handleman JS: Helping children with autism enter the mainstream. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

Heller T: Dementia infantilis. Z Erforsch Beh Jugenli Schwachsinns *2:*141, 1908.

Howlin P: Outcome in autism and related conditions. In *Autism and Pervasive Developmental Disorders*, FR Volkmar, editor. Cambridge University Press, Cambridge, 1998.

Kanner L: Autistic disturbances of affective contact. Nerv Child *2:*217, 1943.

Klin A, Volkmar FR, Sparrow SS, Cicchetti DV, Rourke BP: Validity and neuropsychological characterization of Asperger syndrome: Convergence with nonverbal learning disabilities syndrome. J Child Psychol Psychiatry *36:*1127, 1995.

*Lord C: Diagnostic instruments in autism spectrum disorders. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York 1997.

*McDougle CJ: Psychopharmacology. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York. 1997.

Minshew NJ, Sweeney JA, Bauman ML: Neurological aspects of autism. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

Olsson BO, Rett A: Autism and Rett syndrome: Behavioural investigations and differential. Dev Med Child Neurol *29:*429, 1987.

*Piven J, Palmer P: Psychiatric disorder and the broad autism phenotype: Evidence from a family study of multiple-incidence autism families. Am J Psychiatry *156:*557, 1999.

Rapin I: Autistic children: Diagnosis and clinical features. Pediatrics *87*(5 pt 2):751, 1991.

Rett A: Uber ein eigenartiges hirntophisces Syndroem bei hyperammonie im Kindersalter. Wein Med Wochensch *118:*723, 1966.

Rutter M: Autistic children: Infancy to adulthood. Semin Psychiatry *2:*435, 1970.

Rutter M: Diagnosis and definitions of childhood autism. J Autism Dev Dis *8:*139, 1978.

Rutter M: Debate and argument: There are connections between brain and mind and it is important that Rett syndrome be classified somewhere. J Child Psychol Psychiatry *35:*379, 1994.

*Rutter M: Autism research: Prospects and priorities. J Autism Dev Disord *26:*257, 1996.

Rutter M, Bailey A, Bolton P, Le Couter A: Autism and known medical conditions: Myth and substance. J Child Psychol Psychiatry *35:*311, 1994.

Rutter M, Bailey A, Simonoff E, Pickles A: Genetic influences in autism. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

Szatmari P: Asperger's syndrome: Diagnosis, treatment, and outcome. Psychiatr Clin North Am *14:*81, 1991.

Towbin KE: Pervasive developmental disorder not otherwise specified. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

Tsai LY: Rett's syndrome: A subtype of pervasive developmental disorder? In *DSM-IV Source Book*, vol 3, AJFTA Widiger, HA Pincus, R Ross, MB First, W Davis, editors. American Psychiatric Association, Washington, DC, 1997.

*Volkmar FR, Klin A, Marans W, Cohen DJ: Childhood disintegrative disorder. In *Handbook of Autism and Pervasive Developmental Disorders*, ed 2, DJ Cohen, FR Volkmar, editors. Wiley, New York, 1997.

Volkmar FR, Klin A, Siegel B, Szatmari P: Field trial for autistic disorder in DSM-IV. Am J Psychiatry *151:*1361, 1994.

Wing L: Asperger's syndrome: A clinical account. Psychol Med *11:*115, 1981.

CHAPTER 39. ATTENTION-DEFICIT DISORDERS

39.1 ATTENTION-DEFICIT DISORDERS

JAMES T. McCRACKEN, M.D.

ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

Attention-deficit/hyperactivity disorder (ADHD) is the most common psychiatric disorder among school-age children and the best understood. Children with ADHD display the early onset of symptoms consisting of developmentally inappropriate overactivity, inattention, academic underachievement, and impulsive behavior. The need for treatment of ADHD is highlighted by the increased risk of ADHD children for delinquency, accidents, and substance abuse. Although the disruptive behaviors of ADHD are usually the reason for referral for treatment, ADHD in childhood and adolescence is often associated with other psychopathology. Considerable knowledge supports the premise that ADHD is a familial disorder associated with differences in central nervous system structure, metabolism, and processing. Treatment of patients with ADHD involves multiple interventions and should be guided by a complete assessment of current functioning in multiple domains of school, family, and peer relationships and comorbid symptoms.

Definition and Comparative Nosology The diagnostic criteria for ADHD have undergone frequent, and at times significant, changes. The definition in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) is based upon extensive field testing that led to grouping symptoms into two primary categories: inattentive and hyperactive-impulsive symptoms. A minimum of six symptoms from at least one of the symptom categories is required for the diagnosis. Three different subtypes of ADHD may be diagnosed on the basis of the number and type of symptoms from each category, with subtypes referred to as either ADHD, inattentive type; ADHD, hyperactive-impulsive type; or ADHD, combined type. According to DSM-IV, symptoms of ADHD must be evident by age 7 years and have a minimum duration of 6 months. Although developmental shifts in symptom frequency and severity are well known to occur, the current criteria make no modification for age. While the diagnosis of ADHD has been frequently criticized as a form of inappropriate labeling of natural variants of child behavior, its validity is among the strongest of any psychiatric disorder.

In the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), the category hyperkinetic conduct disorder overlaps considerably with ADHD. Besides generating overall lower prevalence rates for disorder in community samples, the ICD-10 criteria (Table 39.1-1) tend to identify children with higher rates of soft neurological signs, developmental delays, and learning problems. Recent data from family-genetic studies provide some support for the existence of a subtype of ADHD comorbid with conduct disorder, quite similar to the ICD-10 definition, which may possess additional and independent family-genetic and environmental risk factors. In general, hyperkinetic disorders are considered a subgroup within ADHD as currently defined.

Table 39.1-1 ICD-10 Diagnostic Criteria for Hyperkinetic Disorders

History The history of the recognition of ADHD as a child psychiatric disorder spans more than a century. Early conceptualizations included both neurological and moral notions of causation, which have evolved over time into models of the disorder that attempt to integrate knowledge of brain anatomy and function with behavior. Classic early case histories such as Fidgety Phil embody all of the core features of the disorder as currently defined. Later descriptions of hyperactivity in various neurological samples of children with coarse brain injury or disorder, including children with postencephalitic syndromes, mental retardation, and epilepsy, led to the suggestion of a possible syndrome of overactivity. Particularly notable were the common sequelae in children of overactivity and impulsive and disruptive behavior following the pandemic of encephalitis lethargica seen during, and for the 10 years following, World War I. The prominent increase in psychomotor activity in children was in stark contrast to the parkinsonism observed in many adults after recovery from the infection. Though the importance of brain dysfunction presumably resulting from injury or maldevelopment stimulated more interest in understanding the pathophysiology of ADHD, the uncritical application of the assumption of brain damage to the behavior of children with varying degrees of disruptive behavior (as seen in earlier diagnostic labels such as minimal brain damage or minimal brain dysfunction) led to wide differences in syndrome definition and therapeutic approaches between nations.

Over the past 25 years, much research has attempted to identify core deficits associated with the disorder, develop empirically tested diagnostic criteria, and further validate the diagnosis of ADHD by demonstrating the disorder's characteristic symptom profiles, natural history, family-genetic aspects, and response to treatment. Although emphasis has shifted somewhat from the importance of attention deficits to the current view of two primary dimensions of inattentive and hyperactive-impulsive behaviors, along the way ADHD has become one of the best-validated disorders in psychiatry. Furthermore, public acceptance of the need for treatment of the disorder indicated by the increased number of children receiving stimulant treatment.

Epidemiology In general, a variety of epidemiological data consistently find ADHD to be a common disorder in community samples of children and adolescents (on average identified in 3 to 5 percent) and one of the most common disorders among children referred to child mental health services. However, as with most mental disorders, changes in definitions also influence estimates of prevalence, and some variability in rates of ADHD in nonreferred samples have been noted. Available large-scale epidemiological surveys using contemporary diagnostic criteria are summarized in Table 39.1-2.

Author	ADD (%)	ADHD (%)
Anderson et al., 1987	6.7	5.7
Costello et al., 1988	2.2	2.0
Szatmari et al., 1989	—	6.3
Velez et al., 1989	12.6	—
Jensen et al., 1995	—	7.4
Costello et al., 1996	—	1.9
Shaffer et al., 1996	—	4.1
Wolraich, 1996	—	11.4

ADD, attention-deficit disorder; ADHD, attention-deficit/hyperactivity disorder; —, not reported.

Table 39.1-2 Prevalence Estimates for ADHD in Community Samples

These studies have found that prevalence rates of ADHD in nonreferred samples range between 2.0 and 9.5 percent. Prevalence rates of ADHD under DSM-IV definitions are 15 to 57 percent higher than those determined by use of revised third edition of DSM (DSM-III-R) criteria. Despite these differences, some interesting consistencies are seen across studies. A strong male predominance of approximately 2–3 to 1 has been seen for both clinic samples of children with ADHD and school-age children. Rate estimates for ADHD in girls have been generally consistent at approximately 3 percent, in spite of study differences. The field awaits a large-scale epidemiological study applying DSM-IV criteria to assess the full impact of this most recent criteria change on estimates of the prevalence of disorder, correlates, and comorbid difficulties.

Etiology Considerable progress has been made toward revealing the pathophysiological basis of ADHD. Studies of ADHD neurochemistry, neuroimaging, epidemiological risk factors, and genetics have supported the notion that ADHD is a familial disorder involving differences in monoamine regulation and frontal-striatal neural circuitry. Additional studies should soon add detail to early findings and improve early detection of risk and intervention.

Neurochemistry The incontrovertible benefit of psychostimulants on hyperactive and impulsive behaviors has resulted in research into neurochemical differences in children with ADHD, principally among the monoamines dopamine and norepinephrine. In general, earlier studies of peripheral and cerebrospinal fluid monoamine measures in ADHD were considered inconclusive; similarly, studies on the effects of medications on monoamine metabolism in children with ADHD initially yielded variable results. However, more-recent studies have yielded significant evidence supporting the involvement of both the dopaminergic and catecholaminergic systems in ADHD. For example, one study showed a significant correlation between the pretreatment cerebrospinal fluid concentrations of the major dopamine metabolite homovanillic acid (HVA) and response to psychostimulant administration. Alternatively, recent studies of changes in peripheral monoamine regulation reveal differences in catecholamines in children with ADHD and normal controls.

Although difficult to extrapolate to the human disorder, animal models incorporating neonatal lesions of dopamine tracts, rodent strains with alellelic variants at the dopamine transporter (DAT1) and SNAP-25 proteins, and inbred strains such as the spontaneously hypertensive rat (SHR) also support the prominent role of monoamines in the regulation of activity. Taken together, these clinical and basic studies support the importance in ADHD of differences in the dopaminergic ennervation of mesolimbic and cortical areas as modulators of persistence, distractibility, motivation, and motor control. Overactivity or poorly modulated responses of the noradrenergic locus coeruleus to novelty or challenge may elicit the poorly regulated behavioral responses seen in children with ADHD. Increasing dopamine availability and inhibition of locus coerulus activity via adrenergic agonism are hypothesized to be required for maximal therapeutic efficacy on ADHD symptoms, as is found with medications such as the rapidly acting stimulants.

Neuroimaging Both structural and functional neuroimaging studies have contributed to elucidating the etiology of ADHD. Early studies using xenon-133 regional cerebral blood flow pointed to brain areas that form the crux of current neuroanatomical models of pathology in ADHD. Although complicated by the co-occurrence of specific developmental problems in some samples, reduced perfusion in bilateral frontal areas, the caudate nuclei, and additional basal ganglia areas, partially increased by methylphenidate administration, have been reported. The apparent reductions in blood flow in frontal and basal ganglia regions were consistent with the emerging view that many of the self-regulation difficulties seen in children with ADHD resembled the behavior seen in classic frontal lobe damage syndromes. The first positron emission tomography (PET) study of adults with familial ADHD also found both global and specific patterns of reduced metabolism in the brains of patients compared with controls. Compared with normal control adults, those with ADHD showed global reduced metabolism bilaterally, as assessed by a [^{18}F] fluorodeoxyglucose (FDG) tracer. In addition, a survey of specific regions found significant metabolic reductions in superior prefrontal and premotor cortices. Attempts to replicate these findings in samples of adolescents with ADHD have found fewer differences; global metabolism did not differ, although metabolism in six specific regions (an equal number from right and left brain areas) was significantly lower than that in controls in one study. Gender apparently strongly influences regional cerebral blood flow, as only females with ADHD differed in group effects from controls in some studies. In general, the initial functional brain imaging studies indicate differences in brain activity associated with ADHD, which may vary with age. Reduced activity in brain areas associated with executive functions is consistent with the types of behavioral problems seen in the disorder.

Structural brain imaging studies complement the functional imaging studies cited above, and together they underscore the involvement of striatal and frontal brain areas in ADHD, although possible involvement of other brain regions has not been excluded and great variability in brain size is evident in both patient and controls. Initial small-scale magnetic resonance imaging (MRI) studies of ADHD reported lower right anterior frontal width in ADHD than in normal controls and loss of the normal right greater than left asymmetry of the caudate nucleus. The largest structural imaging study of ADHD identified a number of global and specific brain areas that differed between a sample of 57 boys with ADHD and 55 normal controls. Total cerebral volume was 4.7 percent smaller, and total cerebellar volumes were also significantly lower in the children with ADHD. In the striatum, the right globus pallidus was smaller in children with ADHD, and the normal right greater than left asymmetry of the caudate nucleus was lost. Elsewhere, a gross measure of the right anterior frontal lobe volume was also reduced in the ADHD sample. Other reports have generally confirmed these findings, although with some differences. In one report, subjects with ADHD did not differ in global or hemispheric volumes but displayed bilateral reduction in frontal lobe and parietal lobe measures. Other reports have commented on reductions in corpus callosal regions in ADHD.

The relation between these reported brain volume decreases associated with ADHD and their functional significance is largely unclear. One study examined the relation between regional volume and performance on cognitive measures of response inhibition. Significant correlations were observed between prefrontal cortex, caudate nucleus, and globus pallidus volume and task performance, greatest with right-sided structures. Such studies need replication. Additional experiments using noninvasive functional measures, such as functional MRI (fMRI), could significantly increase knowledge of areas and circuits that are involved in producing ADHD symptoms.

In summary, the available neuroimaging data strongly support the existence of differences in brain structure and, to a lesser extent, brain function in individuals with ADHD and normal controls. Rather than finding global brain involvement in ADHD, most studies identify relatively specific areas of difference associated with the disorder, mainly the richly interconnected areas of the frontal lobes and the striatum. However, additional research is needed before these findings can be fully integrated with other knowledge about the disorder.

Genetics Compelling evidence supports the view that ADHD is familial and in large part genetic. In spite of differences in diagnostic criteria, most published controlled family studies report a significantly higher relative risk of ADHD in first-degree and second-degree relatives of probands with ADHD than in normal controls. For example, the risk of ADHD for a sibling of a child proband with ADHD increases from 1.8 to 5 times, depending upon the study, the sample enrolled, and the gender of the probands. Comparisons of concordance rates for ADHD in monozygotic and dizygotic twins also strongly support genetic influence in ADHD and the core symptoms of hyperactivity and inattention. Concordance rates for ADHD have ranged from 51 to 80 percent for monozygotic twins versus 29 to 33 percent for dizygotic twins. Heritability estimates for individual symptom domains also strongly support a genetic influence in producting hyperactive and inattentive behavior. The heritability of hyperactivity has been calculated in twin studies to be between 64 and 77 percent, and that of inattention-related behaviors, between 76 and 98 percent. Therefore, from a dimensional perspective, these extremes along behavioral traits of activity and persistence also appear to be highly influenced by genes.

Analysis of patterns of illness within well-studied pedigrees of families with a member with ADHD may be consistent with the effect of a single major gene; other modes of mendelian transmission have been rejected. However, the cause of ADHD likely involves a complex interaction of multiple genetic factors (perhaps several genes) with effects of gender and environmental influences. This conclusion is supported by initial molecular genetic studies of ADHD, which have indicated a possible association of two specific genes with ADHD. Several studies have replicated an association between a specific allele for the dopamine 4 receptor subtype (the *DRD*4.7 allelle) with childhood ADHD. The association with ADHD follows other work suggesting an association between the *DRD*4.7 allele and novelty-seeking in adults from community samples. Therefore, this allele may relate to specific behavioral dimensions within ADHD, although this is speculative. Another allele (the 480 bp allele) of DAT1 was reported to be associated with transmission of ADHD, using data obtained from affected children and their parents (the haplotype relative risk method). However, some studies have been unable to replicate the association with the *DAT*1 allele. These findings underscore the likely genetic heterogeneity of ADHD (and its subgroups) and the need for additional research using these techniques within large samples.

Acquired Etiological Influences A variety of acquired influences have received support as etiologic factors for some children and adolescents with ADHD, including pregnancy and delivery complications, low birth weight, traumatic brain injury, and prenatal substance exposure. Although the strength and specificity of these influences appear to be more limited than familial factors within groups of subjects with ADHD, the risk factors below may be critical forces in the genesis of nonfamilial forms of ADHD for individual children and adolescents.

Several studies have reported associations between ADHD, pregnancy, and delivery complications. Some but not all reports have noted exposure to maternal toxemia in children later diagnosed with ADHD. Other studies suggest that problems during labor, fetal distress, and other birth complications are associated with later disruptive behavior problems. Such experiences may be particularly germane to the etiology of ADHD in children who lack a family history of the disorder. Similarly,

these adverse neonatal events may be associated with particular forms or subgroups of ADHD such as ADHD with comorbid conduct disorder.

Low birth weight has been identified as a risk factor for disruptive behavior disorders, including ADHD in particular, in several studies. The relation between low birth weight and risk for ADHD does not appear to be due to other perinatal risk factors commonly associated with prematurity. However, not all studies of pre- or perinatal adversity have found associations with ADHD per se: therefore, these events appear to function usually as nonspecific risk factors for psychopathology in children.

While traumatic brain injury has been thought to represent a risk factor for later hyperactivity, support for this association has been mixed. Nevertheless, more-recent studies of children and adolescents with carefully ascertained brain injury have uncovered a complex connection between severe brain injury and attentional and behavioral problems. The relation between brain injury and ADHD symptoms is defined by several aspects including severity of the injury, location of damage to brain tissue, and age at time of injury. The strongest relation has been found for more severe injuries and younger age at the time of injury. In one prospective sample of children with severe head injury, 19 percent had developed new-onset ADHD at follow-up. An association between mesial frontal lobe lesions and postinjury ADHD has also been described. These findings have been interpreted to suggest that earlier injuries may compromise ongoing brain development and disrupt executive functions and that individuals who have achieved greater cognitive development prior to injury may adapt better to loss of skills.

Substance exposure in utero has long been believed to lead to problems of behavioral control and emotional regulation. Prenatal exposure to alcohol and nicotine has been strongly associated with later assessment of ADHD symptomatology. While these effects may be moderated in part by socioeconomic variables and other environmental features, for many children, alcohol exposure represents a major etiologic factor associated with cognitive and behavioral problems. Similarly, prenatal nicotine exposure is also significantly associations with later development of disruptive behavior problems. Though the connection between prenatal nicotine exposure and ADHD has not always been consistent, there is general support for a relation between nicotine exposure and later disruptive behavior problems, including overactivity, in some individuals.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for ADHD are listed in Table 39.1-3. The symptom heterogeneity exhibited by children diagnosed with ADHD is considerable. The variability in behavioral profile within groups of children with ADHD is colored by many factors—age, symptom subtype, cognitive deficit, and comorbidity. Inconsistent behavior and poor academic performance are hallmarks of ADHD. Frequently, parents note that many child problems depend on context—with greater supervision, reduced stimulation, and support, their child's behavior and functioning can improve, at least momentarily. However, a careful history generally clearly shows an overall pattern of symptoms fitting the definition of ADHD. Besides a detailed clinical history of symptoms of ADHD, a general review of other possible psychopathology is indicated. Although children may report few indicators of their own ADHD symptoms, frequently their report can confirm impairment due to ADHD. Detailed descriptions of the individual's school history and performance is important. Additional direct observations of the child in the school setting and with peers can be revealing, particularly in less severe cases. A collection of observations and ratings on standardized behavioral rating forms from parents and teachers greatly aids diagnosis of ADHD. When available, the report of a spouse or parent is a key element in confirming the report of an adult being evaluated for ADHD. Additional self-report measures of ADHD symptoms are available for adults. At present, the specificity of computerized attentional tests for the diagnosis of ADHD is uncertain, and such results should be interpreted with caution.

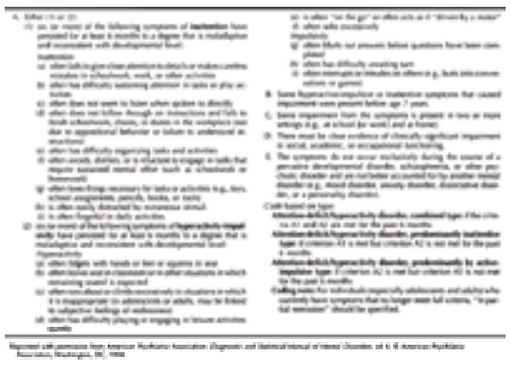

Table 39.1-3 DSM-IV Diagnostic Criteria for Attention-Deficit/Hyperactivity Disorder

Subtypes The DSM-IV subtypes of ADHD describe some of the symptom variability in ADHD: ADHD, inattentive type; ADHD, hyperactive-impulsive type; or ADHD, combined type. Estimates of the frequency of subtypes in clinic-referred children suggest that the combined type is most common, followed by the inattentive type, and finally by the hyperactive-impulsive type. The average age at diagnosis of ADHD is approximately 7 to 9 years. In the field testing of the DSM-IV ADHD categories, most children identified as having ADHD, hyperactive-impulsive type were 3 to 4 years younger than children diagnosed with other subtypes. ADHD children with the inattentive subtype tend to have a later age of recognition, perhaps because they exhibit fewer of the oppositional, defiant, and aggressive behaviors that often spark referrals for evaluation. Although symptoms often overlap across subtypes of ADHD and hyperactivity-impulsivity and inattentive behaviors are significantly correlated, some prominent differences can be observed in individual cases. For example, children with inattentive subtype are often described as sluggish, anxious, subject to daydreaming, and sleepy—if anything, hypoactive.

Nature of ADHD Cognitive Deficits Surprisingly, the nature of the attention deficits encompassed by the syndrome of ADHD has eluded precise identification. Initial conceptualizations of performance differences between ADHD children and normal controls on laboratory tests of cognition emphasized difficulties in sustained attention in ADHD. However, these differences were inconsistent across studies. The application of newer cognitive test paradigms has revealed more-specific patterns of weakness in samples of children with ADHD. Difficulties in inhibiting a behavioral response to a stop signal have been noted in samples of children with ADHD as well as children with other patterns of disruptive or aggressive behavior. However, the deficit in response inhibition is typically not seen in children with other types of psychopathology and thus may be an important correlate of externalizing behavior problems. Other cognitive paradigms applied to the study of ADHD have also observed asymmetries in response patterns in ADHD and normal control children. Tests of covert visual spatial orienting have revealed slower reaction times and unique types of hemispheric asymmetries involving poorer performance of both right and left hemispheres in children with ADHD, depending on cue type. An important area of current research focuses on whether types of cognitive deficits may be familial and serve as possible risk markers for the transmission of ADHD.

Comorbidity Studies of epidemiological and clinical samples of children with ADHD find a high frequency of other overlapping symptoms and diagnoseable disorders, including significant rates of both mood and anxiety disorders as well as other disruptive behavior disorders. Recognition of these associated conditions carries important implications for assessment, prognosis, treatment approaches, and research. For example, the association of ADHD and conduct disorder is reportedly related to later development of substance abuse. While continued debate exists over whether these apparent comorbidities are due to artifactual comorbidity produced by current criteria, other research finds cause for comorbidity in shared deficits and other possible mechanisms of association. In general, the high frequency of associated psychopathology has been argued to reflect the heterogeneity of ADHD itself.

The most common comorbid disorders found in both clinical and epidemiological samples of children with ADHD are oppositional defiant disorder and conduct disorder. These children display argumentativeness, temper outbursts, defiance of authority and rules, and aggressive, antisocial behavior in addition to symptoms of ADHD. Rates of oppositional defiant disorder in samples of children with ADHD reportedly average 35 percent; when oppositional defiant disorder and conduct disorder are combined, rates of comorbidity with ADHD rise to about 50 to 60 percent. A variety of influences may shape this association. For example, school-age children with oppositional defiant disorder or conduct disorder almost invariably meet criteria for ADHD. Conversely, adolescents with conduct disorder without ADHD are more common. Other influences include greater symptom severity, reading disorder, school impairment, and lower socioeconomic status in comorbid ADHD–conduct disorder children. Family and genetic studies support the premise that children with comorbid ADHD and conduct disorder may represent an etiologically distinct subtype. Children with ADHD associated with other disruptive behaviors are particularly challenging and require intensive intervention to prevent greater morbidity and impairment.

Significant comorbidity exists with ADHD and mood and anxiety disorders. Estimates of the co-occurrence of ADHD and mood disorders have ranged from 15 to 75 percent, with an average comorbidity of about 25 to 30 percent; association with anxiety disorders has been reported in up to 25 percent. Studies of comorbidity of ADHD with mood disorders have found common co-occurrence of ADHD with both dysthymic disorder and major depressive disorder. In general, comorbid depression and ADHD does not appear to affect the manifestations of either disorder. With the early onset of ADHD, most mood disorder diagnoses are made following the emergence of ADHD, which suggests that some instances of ADHD–mood disorder comorbidity are secondary to the experience of a chronic impairing disorder such as ADHD. However, most studies of the first-degree relatives of children with ADHD and mood disorder find an increased risk for mood disorders, suggesting that both

disorders breed true in families. Whether ADHD and mood disorders share vulnerability factors is under active investigation. In contrast, some differences in patterns of performance on cognitive tasks and reduced cognitive improvement with stimulants distinguish children with ADHD plus anxiety and those with ADHD without anxiety. These findings may suggest dysfunction in different or additional neural systems in children with ADHD plus anxiety.

A recent controversy involves a suggested association of ADHD with prepubertal bipolar disorder, with over 90 percent of those meeting criteria for bipolar disorder also meeting criteria for ADHD. In one clinical sample of children with ADHD, 16 percent reportedly met criteria for bipolar disorder. Children believed to display co-occurring ADHD and bipolar disorder have also been described as showing more severe ADHD and aggressive symptoms and higher internalizing symptom scores. However, the absence of patterns of classic cyclical mood and the lack of increased rates of bipolar disorder in all of the extant follow-up studies of ADHD suggest that ADHD plus bipolar disorder is infrequent.

Distinct from the academic underachievement and slightly lower (intelligence quotient) I.Q. test scores found in groups of ADHD children, learning disorders and reading disabilities also commonly occur in children with ADHD. Use of rigorous criteria has shown that learning disorders are more frequent in children with ADHD than in either psychiatric controls of normal children, but this comorbidity is modest (15 to 30 percent). Compared with conduct disorder, several studies find ADHD to be more strongly associated with reading disorders than with mathematics disorder. The co-occurrence of learning disability and ADHD has been suggested to result in poorer outcomes. This overlap may also relate to the association of ADHD with other disruptive behavior disorders, such as conduct disorder.

Other types of comorbidity can be found in individuals with ADHD. The overlap of ADHD and chronic tic disorders, including Tourette's disorder, is widely acknowledged, and rates of ADHD in children and adolescents with Tourette's disorder range from 30 to 50 percent. Individuals with mental retardation also often display attentional difficulties, hyperactivity, and impulsivity. For those who meet criteria for ADHD after developmental level is considered, treatment can be highly effective. In adults with ADHD, alcohol and substance abuse is a common comorbidity that poses a definite diagnostic and treatment challenge. Frequently, a period of abstinence is necessary before the adverse effects of substance use can be separated from a possible overlap with symptoms of ADHD. Anger outbursts, other patterns of mood lability, and relationship conflicts are frequent complaints of adults with possible ADHD.

Bill was an 8-year-old boy who lives with his adoptive parents and younger sister. After his school teacher raised concerns over his poor adjustment to his second grade classroom, his parents sought out a consultation. Bill was the product of an apparently normal pregnancy and delivery. He was adopted at age 3 months by his current adoptive parents. Early development was normal except for the description of Bill's temperament as hard to soothe and irregular sleep patterns in the first year of life. After he began preschool at 30 months, staff noted that he had difficulty sitting still for group activities and needed frequent reminders to resist hitting peers over disagreements during play. Later, a supportive first grade teacher described Bill as active but learning normally without concerns. However, soon after Bill entered second grade, his teacher called parents to express concerns that Bill showed problems listening in class and completing assignments without multiple reminders. He also frequently rose from his chair in class during quiet study activities. Overall, he was performing below expectations, and multiple reprimands had not led to sustained improvement.

Bill's parents noted that for about 9 months prior to evaluation, he had become more challenging to his parents at home; in particular, he was frequently argumentative and resistant to requests. Daily struggles over simple routines and limits ensued. Bill's mother had attempted to restrict his intake of sugar but without significant change in behavior. On examination, Bill appeared restless and shifted frequently from one item to another. He acknowledged difficulty listening to his teacher “sometimes.” Psychomotor activity was increased. Mood and affect were normal. Teacher ratings on behavior rating scales were above the 95th percentile, with no indication of a learning disorder. Parent ratings on the Child Behavior Checklist confirmed attention problems and aggresion. A physical examination performed 5 months earlier by Bill's pediatrician showed no abnormalities.

Results of the consultation were consistent with attention-deficit hyperactivity disorder and oppositional defiant disorder. Recommendations included a trial of methylphenidate (Ritalin) at 5, 10, and 15 mg given three times a day over 3 weeks and referral to a parent training program for behavioral management.

Pathology and Laboratory Examination ADHD remains a clinical diagnosis, based upon clinical history and examination. No specific diagnostic tests exist at present. Some data obtained in psychoeducation testing may be confirmatory, including behavioral observations of inattention or poor task persistence during test situations, and relative test performance scores across domains, including overall intelligence, presence or absence of specific learning disorders, and items such as the Freedom from Distractibility factor of the Wechsler intelligence test. Brain imaging, electroencephalography (EEG), and specific neuropsychological testing may be indicated under specific clinical situations to rule out other diagnostic possibilities such as mental retardation, epilepsy, or structural lesions or to evaluate for suspected uncommon comorbidities but should not be considered on a routine basis.

Differential Diagnosis The differential diagnosis of ADHD versus other types of disruptive behavior disorders is generally straightforward by clinical history and examination of specific symptoms and behavior patterns. Descriptions of inattention, overactivity, and impulsivity should be evident across situations and to multiple observers. Other sources of inattention and frustration in school settings, such as learning disorders, can be discerned by differences in behavior across settings, school grades, and results of psychoeducational evaluations. At times, these problems may co-occur, making accurate diagnosis more complex.

Another diagnostic challenge lies in separating interference from other psychopathology such as mood or anxiety disorders on concentration, school adjustment, and performance. In general, ADHD is distinguished by its chronicity and early onset rather than an episodic pattern, although school impairment from ADHD may be progressive, with increasing academic demands over many years. The temporal relationship between the appearance of internalizing symptoms with greater school difficulty usually is evident in instances of mood or anxiety disorders uncomplicated by ADHD.

A variety of other psychiatric and developmental disorders are commonly associated with attention deficits, including autistic disorder and other pervasive developmental disorders, schizophrenia, and mental retardation. In most instances, these disorders, if diagnosed, are considered primary, and a separate diagnosis of ADHD is not made, although interventions to enhance concentration, reduce interfering symptoms, and improve academic performance may be an important part of treatment efforts.

Course and Prognosis Although complicated by differing diagnostic criteria, sample characteristics, and methodology, studies have yielded much about the course of ADHD. These findings include reports from a variety of well-controlled prospective follow-up studies with follow-up as long as 15 to 17 years after initial assessment. In general, both longitudinal studies and descriptions of older samples of individuals with ADHD consistently reveal changes in the core symptoms of ADHD with the development and the appearance of new problem behaviors. The most striking observation is the strong persistence of the diagnosis over time, supporting the prognostic validity of the diagnosis. Recognition of the variety of adverse outcomes of individuals with ADHD shows the need for aggressive, multimodal intervention approaches.

P>Studies following school-age children with ADHD into adolescence find strong persistence of the syndrome over time, usually in spite of treatment. These longitudinal studies find that the symptoms of inattention are more persistent than those of hyperactivity-impulsivity, which tend to decline as a group over time. Besides continuation of most core ADHD symptoms, early studies found worse outcomes for ADHD adolescents: high rates of deliquent behaviors in 25 to 50 percent, poorer self-esteem, lower academic achievement, new diagnoses of conduct disorder, and greater substance abuse. However, subsequent research revealed that early conduct problems, particularly aggression and defiance, rather than severity of ADHD symptoms per se, tend to be more strongly predictive of poor outcomes at later ages. Other predictors, such as I.Q. and parental psychopathology have been identified in some, but not all, follow-up studies. However, given the high comorbidity of ADHD with oppositional defiant disorder and conduct disorder and the relatively modest strength of individual variables as predictors, children with ADHD should be considered at increased risk for adverse outcomes. On average, about 60 to 80 percent of children with ADHD continue to manifest the full syndrome well into adolescence.

In many respects, a slightly more optimistic picture is seen in the young-adult follow-up studies of individuals with ADHD. ADHD persists in 10 to 50 percent of young adults at follow-up, although many more individuals report some interference because of at least one remaining core symptom. Controlled follow-up studies consistently document increased incidence of the development of antisocial personality disorder and substance abuse problems at follow-up. Other characteristics seem to distinguish ADHD groups at follow-up, including high numbers of subjects with histories of incarceration, low educational achievement, and poor self-esteem. The worst outcomes of antisocial personality disorder and substance abuse occurred almost exclusively in subjects who had experienced persistent ADHD over follow-up. Interpretation of these follow-up studies is complicated and controversial. Nevertheless, they highlight the importance of intervention in an effort to prevent these adverse outcomes.

Treatment Practice parameter guidelines for the assessment and treatment of ADHD have been published by the American Academy of Child and Adolescent Psychiatry. While a wide variety of treatments have short-term efficacy for some symptoms of ADHD, few demonstrations of long-term efficacy of comprehensive and

clinically relevant treatment programs exist. Treatments with strong empirical validation of efficacy include medications (particularly the psychostimulants) and certain psychosocial therapies.

The short-term benefits of medication treatment of ADHD has been confirmed in well over 100 studies that included thousands of children with ADHD. The short-acting psychostimulants remain the first-line choice for the pharmacotherapy of ADHD, principally because of their ability to improve both behavioral and cognitive aspects of the disorder in 70 to 80 percent of children. The safety profile for the stimulants is also excellent. However, interest continues in identifying alternatives to stimulants, for such reasons as the significant comorbidity of ADHD with other disorders, occasional intolerance of stimulant adverse effects, the inconvenience of multiple dosing, and infrequent lack of efficacy.

The behavioral and cognitive effects of stimulants on ADHD are robust and extensive. These positive effects include a reduction in psychomotor activity in ADHD patients and normal subjects (not a paradoxical response) and decreased disruptive behavior, aggression and deviance from classmates as measured by peer ratings. The beneficial cognitive effects of stimulants in ADHD children and adolescents are also well documented, including enhanced attention, short-term increases in academic acquisition, and facilitated working memory. However, behavioral improvement clearly falls short of normalization, and significant residual ADHD symptoms, peer problems, and academic challenges remain in most children.

Four rapid-acting stimulants are currently available: methylphenidate, dextroamphetamine (Dexedrine), magnesium pemoline (Cylert), and a mixed preparation of amphetamine salts (Adderall). Table 39.1-4 describes commonly prescribed medications for ADHD and their approximate effective dosage ranges. While methylphenidate is the most commonly prescribed stimulant, there is no indication of differential efficacy among the currently available psychostimulants. In the few studies that have directly compared the responses of individual ADHD children to more than one stimulant, occasional preferential improvement with one agent has been observed.

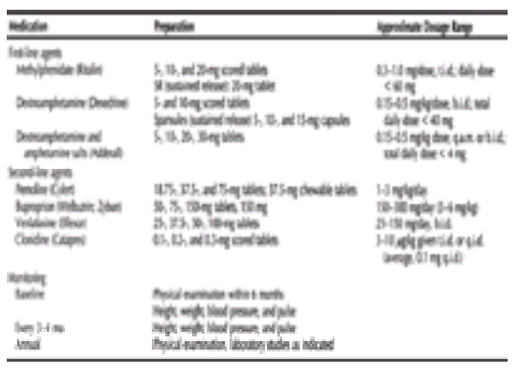

Table 39.1-4 Medications for the Treatment of ADHD and Suggested Monitoring

The psychostimulants have an excellent track record of safety. Adverse effects of the stimulants are fortunately minor or uncommon, although they can pose dilemmas in management. In general stimulant treatment–emergent effects are minor, dose related, and transient. The most common adverse effects are initial insomnia and appetite reduction, followed by headache, irritability, and dysphoria. Less commonly, reduced weight gain is problematic. In some patients, a rebound in ADHD symptoms can be difficult to manage late in the day. Although the convention of not prescribing stimulants to patients with comorbid tic disorders has been largely discounted, some individuals cannot tolerate the increase in tic symptoms produced by stimulants. Pemoline has been associated with a rare hepatotoxic reaction that has not been observed with the other stimulants. Although typical stimulant adverse effects can usually be managed by dosage reduction, or occasional change in agent, the continued presence of even minor adverse effects may strongly influence compliance. Causes for consideration of other medications than the stimulants include impairing comorbid psychopathology, the presence of a chronic tic disorder, comorbid substance abuse, and marked anorexia.

A number of nonstimulant alternatives have been proposed for the pharmacological management of ADHD (Table 39.1-4), however, their empirical support varies considerably, and they should be reserved for stimulant refractory or intolerant patients. These include bupropion (Wellbutrin, Zyban) tricyclic drugs, a-adrenergic receptor agonists such as clonidine (Catapres) and guanfacine (Tenex), and other atypical antidepressants. In general, these alternatives fall short of equaling the improvements on cognition produced by the stimulants, although modest positive effects of bupropion on cognition have been suggested. Safety concerns also limit the indications for wider use of these medications for ADHD.

In general, the approach to pharmacotherapy involves initial titration of a stimulant, usually methylphenidate or dextroamphetamine given in multiple daily doses. Intolerance or lack of efficacy should prompt a trial of a second stimulant. Thereafter, additional medication trials should be guided by comorbidity, acknowledging the substantial empirical support for bupropion and tricyclic drugs as effective second-line agents.

Despite the broad base of empirical support for medication treatment for ADHD, a number of gaps still exist in the literature. Most studies have primarily included male school-age children, and used short-term treatment periods. Emerging data does support the continued efficacy of treatment periods as long as 12 to 24 months, but additional data are needed.

Behavior therapy has clear support as a component of treatment for ADHD. In particular, behavioral parent training and classroom behavioral modification approaches are well documented to be probably effective interventions for ADHD, although the magnitude of behavior change typically falls short of that produced by stimulant treatment. Effective ingredients of these approaches include providing consistent contigencies for adaptive and maladaptive behavior, consistent praise, use of time out, and verbal reprimands. Additional elements that appear to be effective are daily report cards for behavior and point systems for positive and negative behaviors. Other, more-intensive behavior modification interventions may be helpful.

A variety of other traditional and nontraditional therapies have been touted as useful complementary treatments for ADHD but lack supporting studies. Attempts to show the benefits of cognitive therapy and other psychological therapies for ADHD have not proved fruitful. Dietary manipulations and restrictions, herbal preparations, biofeedback, and perceptual training cannot be endorsed for ADHD because of lack of supporting evidence. Because of the well-known academic underachievement problems of ADHD children and the significant comorbidity of ADHD with learning disabilities, educational tutoring and remediation are widely indicated.

Combined Approaches Few studies have provided true tests of the widely held belief that a combination of treatments may be most efficacious for children with ADHD. Initial findings from the recent Multisite Treatment Study of ADHD (MTA study) suggest that although psychostimulants exhibit the most robust short and longer-term benefits for ADHD symptoms, some additive benefits in peer relationships and parent satisfaction occurred in children receiving medication treatment combined with home and school behavioral interventions. Future research should help clarify better indicators for the use of combined treatments and the benefits of multimodal treatment approaches in other populations of youngsters with ADHD.

ATTENTION-DEFICIT DISORDER NOT OTHERWISE SPECIFIED

DSM-IV includes attention-deficit disorder not otherwise specified as a residual category for disturbances with prominent symptoms of inattention or hyperactivity that do not meet the criteria for ADHD (Table 39.1-5).

Category is for disorders with prominent symptoms of inattention or hyperactivity-impulsivity that do not meet criteria for attention-deficit/hyperactivity disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 39.1-5 DSM-IV Diagnostic Criteria for Attention-Deficit Disorder Not Otherwise Specified

The incidence of adult manifestations of ADHD is unknown, but there are many more cases than were previously thought or diagnosed. This category of illness is being more frequently diagnosed and requires much greater attention and study. In adults, residual signs of the disorder include impulsivity and attention deficit (e.g., difficulty in organizing and completing work, inability to concentrate, increased distractibility, and sudden decision making without a thought of the consequences).

SUGGESTED CROSS-REFERENCES

Adult manifestations of attention-deficit disorders are discussed in Section 39.2. Conduct disorder and oppositional defiant disorder are discussed in Chapter 39, and Tourette's disorder is discussed in Chapter 41. Pharmacotherapy for children is discussed in Section 46.3, individual psychotherapy for children in Section 46.1, and family therapy in Section 31.5 and Section 48.5.

SECTION REFERENCES

American Academy of Child and Adolescent Psychiatry Work Group on Quality Issue: Practice parameters for the assessment and treatment of attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry *30:*1, 1991.

*Barkley RA: Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of ADHD. Psychol Bull *121:*65, 1997.

Barkley RA, Grodzinsky G, DuPaul GJ: Frontal lobe functions in attention deficit disorder with and without hyperactivity: A review and research report. J Abnorm Child Psychol *20:*163, 1992.

Barrickman LL, Perry PJ, Allen AJ, Kuperman S, Arndt SV, Herrmann KJ, Schumacher E: Bupropion versus methylphenidate in the treatment of attention-deficit hyperactivity disorder. J Am Acad Child Adolesc Psychiatry *34:*649, 1995.

*Baughman FA Jr: Treatment of attention-deficit/hyperactivity disorder. JAMA *281:*1490, 1999.

Biederman J, Faraone S, Mick E, Wozniak J, Chen L, Ouellette C, Marrs A, Moore P, Garcia J, Mennin D, Lelon E: Attention-deficit hyperactivity disorder and juvenile mania: An overlooked comorbidity? J Am Acad Child Adolesc Psychiatry *35:*997, 1996.

*Biederman J, Munir K, Knee D: Conduct and oppositional disorder in clinically referred children with attention deficit disorder: A controlled family study. J Am Acad Child Adolesc Psychiatry *26:*724, 1987.

Biedermann J, Newcorn J, Sprich S: Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety and other disorders. Am J Psychiatry *148:*564, 1991.

Casey BJ, Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Schubert AB, Vauss YC, Vaituzis AC, Dickstein DP, Sarfatti SE, Rapoport JL: Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. J Am Acad Child Adolesc Psychiatry *36:*374, 1997.

Castellanos FX, Elia J, Kruesi M, Gulotta C, Mefford I, Potter W, Ritchie G, Rapoport J: Cerebrospinal fluid monoamine metabolites in boys with attention-deficit hyperactivity disorder. Psychiatry Res *52:*305, 1994.

Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Vaituzis AC, Dickstein DP, Sarfatti SE, Vauss YC, Vaituzis AC, Snell JW, Lange N, Kaysen D, Krain AL, Ricthie GF, Rajapakse JC, Rapoport JL: Quantitative brain magnetic resonance imaging in attention-deficit/hyperactivity disorder. Arch Gen Psychiatry *53:*607, 1996.

Cheyette SR, Cummings JL: Encephalitis lethargica: Lessons for contemporary neuropsychiatry. J Neuropsychiatry Clin Neurosci *7:*125, 1995.

Clark E: Children and adolescents with traumatic brain injury: Reintegration challenges in educational settings. J Learn Disabil *29:*549, 1996.

Conners CK, Casat CD, Gualtieri CT, Weller E, Reader M, Reiss A, Weller RA, Khayrallah M, Ascher J: Bupropion hydrochloride in attention deficit disorder with hyperactivity. J Am Acad Child Adolesc Psychiatry *35:*1314, 1996.

Cook E, Stein MA, Krasowski MD, Cox N, Olkon D, Kieffer J, Leventhal B: Association of attention-deficit disorder and the dopamine transporter gene. Am J Hum Genet *56:*993, 1995.

*Costello EJ, Angold A, Burns B, Stangl DK, Tweed D, Erkanli A, Worthman C: The Great Smoky Mountains study of youth. Arch Gen Psychiatry *53:*1129, 1996.

Douglas VI: Stop, look, and listen: The problem of sustained attention and impulse control in hyperactive and normal children. Can J Behav Sci *4:*259, 1972.

Elia J, Borcherding B, Rapoport J, Keysor C: Methylphenidate and dextroamphetamine treatments of hyperactivity: Are there true nonresponders? Psychiatry Res *36:*141, 1991.

Ernst M, Liebenauer LL, King AC, Fitzgerald GA, Cohen RM, Zametkin AJ: Reduced brain metabolism in hyperactive girls. J Am Acad Child Adolesc Psychiatry *33:*858, 1994.

Faraone SV, Biederman J: Genetics of attention-deficit hyperactivity disorder. Child Adolesc Clin North Am *3:*285, 1994.

Forness SR, Cantwell DP, Swanson JM, Hanna GL, Youpa D: Differential effects of stimulant medication on reading performance in hyperactive boys with and without conduct disorder. J Learn Disabil *30:*173, 1991.

Goodman R, Stevenson J: A twin study of hyperactivity. II: The aetiological role of genes, family relationships and perinatal adversity. J Child Psychol Psychiatry *30:*691, 1989.

Greenhill LL, Abikoff HB, Arnold LE, Cantwell DP, Conners CK, Elliott G, Hechtman L, Hinshaw SP, Hoza B, Jensen PS: Medication treatment strategies in the MTA study: Relevance to clinicians and researchers. J Am Acad Child Adolesc Psychiatry *35:*1304, 1996.

Hart EL, Lahey BB, Loeber R, Applegate B, Frick PJ: Developmental change in attention-deficit hyperactivity disorder in boys: A four-year longitudinal study. J Abnorm Child Psychol *23:*729, 1995.

Heilman KM, Voeller KK, Nadeau S: A possible pathophysiologic substrate of attention deficit hyperactivity disorder. J Child Neurol *6:*S76, 1991.

Hinshaw SP: On the distinction between attentional deficits/hyperactivity and conduct problems/aggression in child psychopathology. Psycho Bull *101:*443, 1987.

Hinshaw SP: Intervention for social competence and social skill. Child Adolesc Psychiat Clin North Am *1:*539, 1992.

Hynd GW, Semrud-Clikeman M, Lorys AR, Novey ES, Eliopulos D, Lyytinen H: Corpus callosum morphology in attention-deficit hyperactivity disorder: Morphometric analysis of MRI. J Learn Disabil *24:*141, 1991.

Lahey B, Applegate B, McBurnett K, Biederman J, Greenhill L,: DSM-IV field trials for attention deficit hyperactivity disorder in children and adolescents. Am J Psychiatry *151:*1673, 1994.

LaHoste GJ, Swanson JM, Wigal, SB, Glabe C, Wigal T, King N, Kennedy JL: Dopamine D4 receptor gene polymorphism is associated with attention deficit hyperactivity disorder. Mol Psychiatry *1:*121, 1996.

Mannuzza S, Klein RG, Bessler A, Malloy P, LaPadula M: Adult outcome of hyperactive boys. Arch Gen Psychiatry *50:*565, 1993.

McCracken JT: A two-part model of stimulant activity. J Neuropsychiatry Clin Neurosci *3:*201, 1991.

McGee R, Williams S, Silva PA: Behavioral and developmental characteristics of aggressive, hyperactive, and aggressive-hyperactive children. J Am Acad Child Adolesc Psychiatry *23:*270, 1984.

Milberger S, Biederman J, Faraone SV, Guite J, Tsuang MT: Pregnancy, delivery and infancy complications and attention deficit hyperactivity disorder. Issues of gene-environment interaction. Biol Psychiatry *41:*65, 1997.

Nigg JT, Swanson JM, Hinshaw SP: Covert spatial attention in boys with attention deficit hyperactivity disorder: Lateral effects, methylphenidate response and results for parents. Neuropsychologia *35:*165, 1997.

Pelham WE, Wheeler T, Chronis A: Empirically supported psychosocial treatments for attention deficit hyperactivity disorder. J Consult Clin Psychol *27:*190, 1998.

Pliszka S: Comorbidity of attention-deficit hyperactivity disorder and overanxious disorder. J Am Acad Child Adolesc Psychiatry *31:*197, 1992.

Porrino L, Rapoport J, Behar D, Sceery W, Ismond D, Bunney W: A naturalistic assessment of the motor activity of hyperactive boys. Arch Gen Psychiatry *40:*681, 1983.

Richters JE, Arnold LE, Jensen PS, Abikoff H, Conners CK, Greenhill LL, Hechtman L, Hinshaw SP, Pelham WE, Swanson JM: NIMH collaborative multisite multimodal treatment study of children with ADHD: I. Background and rationale. J Am Acad Child Adolesc Psychiatry *34:*987, 1995.

Satterfield JH, Satterfield BT, Cantwell DP: Three-year multimodality treatment of hyperactive boys. Pediatrics *98:*650, 1981.

Schachar R, Tannock R: Test of four hypotheses for the comorbidity of attention-deficit hyperactivity disorder and conduct disorder. J Am Acad Child Adolesc Psychiatry *34:*639, 1995.

Seeman P, Madras BK: Anti-hyperactivity medication: Methylphenidate and amphetamine. Mol Psychiatry *3:*386, 1998.

Seidman LJ, Biederman J, Faraone SV, Weber W, Ouellette C: Toward defining a neuropsychology of attention deficit-hyeractivity disorder: Performance of children and adolescents from a large clinically referred sample. J Consult Clin Psychol *65:*150, 1997.

Sherman DK, McGue MK, Iacono WG: Twin concordance for attention deficit hyperactivity disorder: A comparison of teachers and mothers reports. Am J Psychiatry *154:*532, 1997.

*Spencer T, Biederman J, Wilens T, Harding M, O'Donnell D, Griffin S: Pharmacotherapy of attention-deficit hyperactivity disorder across the life cycle. J Am Acad Child Adolesc Psychiatry *35:*409, 1996.

Sprich-Buckminster S, Biederman J, Milberger S, Faraone S, Krifcher Lehman B: Are perinatal complications relevant to the manifestations of ADD? Issues of comorbidity and familiality. J Am Acad Child Adolesc Psychiatry *32:*1032, 1993.

Swanson JM, Shea C, McBurnett K, Potkin SG, Fiore C, Crinella F: Attention and hyperactivity. In *The Development of Attention: Research and Theory*, editor. JT Enns, Elsevier Science, New York, 1990.

Tannock R, Ickowicz A, Schachar R: Differential effects of methylphenidate on working memory in ADHD children with and without comorbid anxiety. J Am Acad Child Adolesc Psychiatry *34:*886, 1995.

Velez CN, Johnson J, Cohen P: A longitudinal analysis of selected risk factors for childhood psychopathology. J Am Acad Child Adolesc Psychiatry *28:*861, 1989.

Weiss G, Hechtman LT: *Hyperactive Children Grown Up*, ed 2. Guilford, New York, 1993.

*Wolraich M, Hannah J, Pinnock T, Baumgaertel A, Brown J: Comparison of diagnostic criteria for attention-deficit hyperactivity disorder in a county-wide sample. J Am Acad Child Adolesc Psychiatry *35:*319, 1996.

Zametkin A, Nordahl T, Gross M, King C, Semple W, Rumsey J, Hamburger S, Cohen R: Cerebral glucose metabolism in adults with hyperactivity of childhood onset. N Engl J Med *323:*1361, 1990.

CHAPTER 39. ATTENTION-DEFICIT DISORDERS

39.2 ADULT MANIFESTATIONS OF ATTENTION-DEFICIT/HYPERACTIVITY DISORDER

PAUL H. WENDER, M.D.

Attention-deficit/hyperactivity disorder (ADHD) is a common psychiatric disorder in adults that was seldom diagnosed until recently, despite the publication of systematic studies since the mid-1970s. It was denoted in the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) as attention-deficit disorder, residual type. The revised third edition of DSM (DSM-III-R) stated that approximately one third of children with ADHD showed continuing signs in adulthood but did not code this as a separate category. The fourth edition of DSM (DSM-IV) states that "symptoms attenuate during late adolescence and adulthood, although a minority experience the full complement of systems of ADHD into mid-adulthood," and thus continues to use the attention-deficit/hyperactivity disorder diagnosis with adults.

HISTORY

The systematic exploration of adults with ADHD, using operational definitions of the syndrome and evaluating treatment in a placebo-controlled fashion, was started in 1976 by the Utah group of the author, David Wood, and Paul Reimherr, and they continued it through the early 90s. During the past few years awareness of the syndrome has spread widely and will undoubtedly increase further.

EPIDEMIOLOGY

No data exist concerning the prevalence of ADHD in the adult population. However, an estimate can be made on the bases of its prevalence in childhood and longitudinal studies that have determined the persistence of the disorder into adulthood. Studies have variously estimated the prevalence at 3 to 10 percent of children. The variability in rates reflects the lack of validating gold-standard methods for ascertaining psychiatric diagnosis, hence diagnosis is based on a quantitative measure—an arbitrary cutoff in the number of symptoms present or an equally arbitrary cutoff in rating scale scores. Two large longitudinal studies have addressed the prevalence and persistence issues. Salvatore Mannuzza and Rachel Klein followed up a population of 103 boys, ages 6 to 12 years, with "hyperactive reaction of childhood" and 100 controls. They found persistence in 40 percent of patients at a mean age of 18 (versus 3 percent of the controls). At a mean age of 26, only 11 percent of the subjects manifested ADHD symptoms, while 18 percent had antisocial personality disorder and 16 percent had nonalcohol substance abuse disorder (versus 2 to 4 percent in matched controls). Gabrielle Weiss and Lily Hechtman followed up 6- to 12-year-old hyperactive patients to the age of 25 and could evaluate only 60 percent of the now-adult subjects. Two thirds of them exhibited at least one symptom of ADHD such as restlessness, distractibility, or impulsivity compared with 7 percent of the controls. Approximately one half continued to have moderate-to-severe problems, and one fourth manifested symptoms of antisocial personality disorder. The difference in outcome cannot be readily explained, but one factor may have been the absence of a reporting adult in the Mannuzza-Klein studies. Evaluation of initial psychopathology and changes produced by treatment are extremely difficult to evaluate without the help of an adult reporter. Based on these studies the estimate of the prevalence of ADHD in adults varies from 0.3 to 6 percent. Other studies have found an increased frequency of a history of ADHD in young male alcoholics and in populations of substance abusers.

An additional difficulty in determining the prevalence of ADHD in adults is the frequency with which it is mistaken for one of several, often common, disorders in adulthood or is hidden from view altogether. For example, not only is it found with increased prevalence in young alcohol abusers and in college students with poor study habits (generally due to an inability to concentrate), but it often mimics minor depressions. Typically it presents as a mild, periodic depression of short duration and marked lability that responds rapidly to improved environmental conditions. Other substantial groups of ADHD individuals go completely unrecognized because they belong to populations that generate little diagnostic attention such as the lowest classes in society and those embedded in the criminal and penal systems, which both appear to contain a high prevalence of individuals with this disorder.

ETIOLOGY

The etiology of the disorder in adulthood is obviously the same as that in childhood. The author hypothesized in 1971 that the disorder was a genetically transmitted disorder of catecholaminergic functioning. A number of studies have documented the genetic basis of ADHD. The author's inferences on genetic transmission were based on two crucial findings: the higher instance of ADHD in siblings of children and the increased incidence of other psychiatric conditions, particularly alcohol abuse and antisocial personality disorder in their parents. The latter relationship may have been mediated by the well-known comorbidity of ADHD and conduct disorder in child probands. Since one half of conduct-disordered children go on to develop antisocial personality disorder in adulthood, an increased frequency of alcoholism and antisocial personality disorder would be predicted in adults who had ADHD in childhood. Moreover, the clustering of alcohol abuse, antisocial personality disorder, somatization disorder agrees with results of the family studies of Samuel Guze at Washington University. Familial clustering can obviously result from genetic or psychosocial transmission. Studies of the concordance of ADHD in monozygotic and dizygotic twins found greater concordance in the former and no evidence for psychosocial transmission. Finally, adoption studies support the genetic hypothesis. Daniel Safer found that separately placed full siblings of ADHD children were twice as likely to have ADHD as their half-siblings, while Helene Deutsch found a higher frequency of ADHD symptoms in the biological relatives of ADHD children than in the adopting parents of these children.

The catecholaminergic hypothesis also drew early support from the development of ADHD-like symptoms in children who recovered from von Economo's disease. Pathological studies by Constance von Economo of children who died during the course of the encephalitis showed that they had damage to a variety of structures including the substantia nigra, the major site of origin of dopaminergic cells. The biochemical and genetic basis for ADHD is also supported by the fact that such children and adults often show dramatic improvement when treated with two indirect dopamine agonists, the amphetamines and methylphenidate (Ritalin). The availability of adults with ADHD allowed studies ethically precluded in children. The only study of cerebrospinal fluid (CSF) in adults with ADHD demonstrated a lower concentration of homovanillic acid (the principal metabolite of dopamine) in patients than in controls. Further, administration of amino acid precursors of dopamine revealed that tyrosine, the immediate precursor of dopamine, decreased ADHD behavior in adult patients. Finally, drugs with a specific effect on dopamine metabolism were used. Dopamine is metabolized by monoamine oxidase (MAO) type B (MAO_B). Administration of two inhibitors of this enzyme, pargyline and selegiline (Eldepryl), improved the behavior of ADHD subjects, these studies were not placebo controlled and must be replicated. Finally, there are an increasing number of imaging studies of children with ADHD and presumably such studies will soon be conducted in adults.

DIAGNOSIS AND CLINICAL FEATURES

DSM-IV defines subsyndromes of ADHD in children as predominantly inattentive type, predominantly hyperactive-impulsive type, and combined type. No definition is provided for the syndrome in adulthood. As might be expected, many items are age limited; for example, features such as "often runs about or climbs excessively" or "has difficulty playing or engaging in leisure activities quietly" clearly are designed for children. To study the syndrome in adults, the Utah group constructed operational diagnostic criteria. The first prerequisite specified in DSM-IV is that the specific symptoms have been present since before the age of 7. Since no evidence indicates that adults can accurately remember the presence or absence of such symptoms in that age span and parents or former caretakers were infrequently available, alternate diagnostic techniques were needed. To address the childhood requirements, the Utah adult ADHD research group demanded that either the narrow or the broad criterion below be met.

I. Childhood characteristics
   A. Narrow criterion: the individual met DSM-IV criteria for ADHD in childhood, determined from a parent interview.

B. Broad criterion: both characteristics 1 and 2 below are met, as judged by the patient.
   1. *Hyperactivity:* more active than other children, unable to sit still, fidgetiness, restlessness, always on the go, talking excessively
   2. *Attention deficits:* sometimes described as having a “short attention span,” distractibility, unable to finish school work

Even with best efforts, it was sometimes difficult to obtain this early history with certainty. However, the careful schedule of symptoms reported by patients about their current, adult functioning had greater reliability and importance. This information was gathered primarily from the patient but also, when possible, from a significant other as well. Thus, a second diagnostic criterion was (typically) the presence of motor hyperactivity in adulthood as well as at least three of the following descriptive groups of symptoms. As further measures of childhood symptoms, two rating scales were employed. The first, the Parents' Rating Scale is a 10-item scale on which the subject's mother rates him as he was between the ages of 6 and 10. The second, the Wender Utah Rating Scale (WURS) is a 61-item questionnaire on which the subject rates his behavior. Both have been standardized and permit the determination of the percentile of childhood ADHD into which the subject fell.

II. Adult characteristics
   A. *Motor hyperactivity:* restlessness; inability to relax; “nervousness” (meaning inability to settle down, not anticipatory anxiety); inability to persist in sedentary activities (e.g., watching movies or TV, reading the newspaper); always on the go, dysphoric when inactive
   B. *Attention deficits:* an inability to keep mind on conversations; distractibility (incapacity to filter out extraneous stimuli); difficulty keeping mind on reading materials or tasks (“mind frequently somewhere else”); frequent “forgetfulness”; often losing or misplacing things; forgetting appointments, plans, car keys, purse, and so on
   C. *Affective lability:* usually described as antedating adolescence and in some instances as far back as the patient can remember; manifested by definite shifts from a normal mood to depression or mild euphoria or (more often) excitement; depression described as being “down,” “bored,” or “discontented”; anhedonia not present; mood shifts usually last hours to at most a few days and are present without significant physiological concomitants; mood shifts may occur spontaneously or be reactive
   D. *Hot temper, explosive, short-lived outbursts:* outbursts of hot temper, “short fuse,” or “low boiling point” usually followed by quick calming down. Subjects may report having transient loss of control and being frightened by their own behavior; easily provoked or constant irritability; temper problems interfere with personal relationships
   E. *Emotional overreactivity:* cannot take ordinary stress in stride and reacts excessively or inappropriately with depression, confusion, uncertainty, anxiety, or anger; emotional responses interfere with appropriate problem solving; experience repeated crises in dealing with routine life stresses; describe themselves as easily “hassled” or “stressed out”
   F. *Disorganization, inability to complete tasks:* a lack of organization in performing on the job, running a household, or performing schoolwork; tasks are frequently not completed; the subject switches from one task to another in haphazard fashion; disorganization in activities, problem solving, organizing time; lack of “stick-to-it-tiveness”
   G. *Impulsivity:* minor manifestations include talking before thinking things through; interrupting others' conversations; impatience (e.g., while driving); impulse buying. Major manifestations may be similar to those seen in mania and antisocial personality disorder and include poor occupational performance; abrupt initiation or termination of relationships (multiple marriages, separations, divorces); excessive involvement in pleasurable activities without recognizing risks of painful consequences (buying sprees, foolish business investments, reckless driving). Subjects make quick and easy decisions without reflection, often based on insufficient information and to their disadvantage; inability to delay acting out without experiencing discomfort
   H. *Associated features:* marital instability; academic and vocational success below that expected on the basis of intelligence and education; alcohol or drug abuse; atypical responses to psychotropics; family histories of ADHD in childhood as well as antisocial personality disorder in male relatives and Briquet's syndrome in female relatives

These required characteristics are of a combined type and are based on disjunctive categories; hyperactivity is prerequisite both in childhood and adulthood, and a sufficient number of other related symptoms are required as well. However, it was possible to study (as the Utah group did) individuals who possessed a sufficient number of symptom groups but not hyperactivity.

These adult criteria are preliminary but have proved their usefulness. They are based on symptoms about which an adult can inform the examiner; the childhood psychiatric diagnosis is based on children's behavior or signs. Further delineation of the adults' symptoms benefit from direct subjective impressions of the adults rather than their presumably unreliable extrapolation of childhood behavior.

DIFFERENTIAL DIAGNOSIS

ADHD patients are often identified by prominent affective lability, irritability, and impulsivity. Thus, ADHD may be confused with chronic mood disorders such as dysthymic and cyclothymic disorders. Unlike these classic mood disorders, the mood symptoms in adults with ADHD are relatively brief, are not characterized by anhedonia, and have no physiological concomitants. The emotion most frequently described by these patients is “boredom.” ADHD “cyclothymic symptoms” differ from symptoms of cyclothymic disorder in that the up periods rarely last more than hours, are characterized by excitement rather than euphoria, and are usually reactive. Some symptoms of ADHD overlap those of borderline personality disorder such as impulsivity, affective inability, angry outbursts, and feelings of boredom. However, patients in the two groups show quantitative and qualitative differences in these symptoms. Impulsivity in ADHD is typically abrupt and thoughtless rather than driven, and the anger displayed is usually quick and short, unlike the brooding seen in patients with borderline personality disorder. Other associated conditions are antisocial personality disorder and substance abuse. The former is to be predicted because many conduct disordered children are comorbid for ADHD and half of them develop antisocial personality disorder in adulthood. Both the Mannuzza and Klein findings and those of David Wood predict an increased frequency of ADHD among alcohol and other substance abusers. The therapeutic consequences of this observation remain to be explored.

COURSE AND PROGNOSIS

In spite of the increasing recognition of ADHD among adults, there are no longitudinal studies of its course in adults. The Utah group has treated individuals in their 30s through their 50s and has found that they can be maintained for extended periods on the same dose of stimulants. In addition, clinical observation of parents of ADHD adults suggests that many symptoms continue well into the 60s (the seventh decade). No controlled studies of these issues exist.

TREATMENT

Sophisticated medication trials in ADHD adults are limited. The Utah group has conducted placebo-controlled trials and open-label studies in over 300 patients. These include four double-blind, placebo-controlled trials with 240 patients taking stimulants: three trials of methylphenidate (one of which is in progress) and one with pemoline (Cylert). In addition, 79 patients were included in trials of various medications including pargyline, selegiline, bupropion (Wellbutrin), levodopa (Larodopa), *dl*-phenylalanine, and L-tyrosine. Two additional placebo-controlled trials of methylphenidate have been conducted by Jeffrey Mattes and coworkers (who failed to replicate treatment effects, probably due to major differences in design) and Thomas Spencer and co-workers (who replicated the methylphenidate findings). The Utah group found that about 60 percent of patients receiving stimulant medication experience moderate-to-marked improvement, compared with 10 percent of controls. This is reflected by changes in Global Assessment of Functioning scores from pretreatment levels of 55 (moderate symptoms) to posttreatment scores of about 75 (slight symptoms present only in response to stress).

In general, current conclusions are that the drugs useful for ADHD children are useful in the management of adults with the disorder, with similar qualitative and quantitative effects in drug-responsive individuals. In fully responsive patients, concentration span is increased and distractibility decreased. Organizational ability improves, and patients think ahead and plan their tasks. Motor hyperactivity and feelings of restlessness disappear without the appearance of sedation; affective lability decreases, both highs and lows are eliminated, and patients experience calm without feelings of euphoria; hot temper and loss of control diminish; and many frustrations become tolerable. Impulsivity diminishes, at least in the short term; longer-term evaluation is necessary to observe important changes in vocational adjustment and interpersonal relationships. Stress intolerance abates. Patients report that they become “thicker skinned”; they can “roll with the punches,” and reverses may get them down slightly, but they bounce back.

The medications most widely used are the stimulant drugs. Methylphenidate, dextroamphetamine (Dexedrine), and methamphetamine (Desoxyn) appear to be equally efficacious; pemoline seems least effective overall. However, no predictors exist, and any one patient may do best on any one of these three agents. Dextroamphetamine and methamphetamine are equally potent; methylphenidate is approximately one half and pemoline approximately one quarter as potent as the amphetamines. The pharmacokinetics of the stimulants differ. Methylphenidate lasts 1½ to 3 hours, and many patients require four to six daily doses; amphetamines last 3 to 5 hours, and most patients need two or three doses. Pemoline has the longest duration of all, and one morning dose may suffice. Approximate dosage ranges

are methylphenidate, 30 to 90 mg a day; amphetamines, 15 to 45 mg a day; and pemoline, 37.5 to 150 mg a day.

Although long-acting forms of methylphenidate and amphetamines are marketed, the clinical consensus is that they do not always increase the duration of action; clinical experimentation may be necessary. Laboratory tests are unnecessary except for pemoline, which produces liver abnormalities in 2 to 3 percent of patients. The adverse effects of the stimulant drugs are anorexia (usually temporary) and insomnia (if the drug is given too late in the day); the latter can often be controlled with a small bedtime dose of a sedative phenothiazine such as thioridazine (Mellaril), 10 to 50 mg. The development of paranoid behavior usually indicates an undiagnosed schizophrenia spectrum condition.

The MAO inhibitors (MAOIs) appear to be effective in some patients with ADHD. The only systematic trials have been open ones with pargyline and selegiline. The advantages of MAOIs are their lack of abuse potential and their long duration of action; the latter averts fluctuations during the day and hence destructive temper outbursts before stimulant medication is taken in the morning or after it wears off in the evening. Compared with other MAOIs, selegiline in low (possibly subtherapeutic) dosages has no hypertensive “cheese effect.” Unfortunately it has not been evaluated in a placebo-controlled trial, and the higher dosages likely to be used run the risk of nonselective MAO inhibition and the cheese effect. Bupropion also appears to benefit approximately one half of stimulant-drug responders at a dosage of 300 to 400 mg a day. Bupropion is considerably less effective in stimulant nonresponders. It appears to control affective lability, dysthymia, and temper more effectively than concentration problems. A small, recent, open study suggests that venlafaxine (Effexor) may benefit some patients. Some clinicians have reported that tricyclic drugs are beneficial, but the author has found them of only moderate effectiveness. A therapeutic response occurs in days, but tolerance generally develops in weeks and does not respond to increased dosage. Symptom response appears more limited than with the stimulants, and ADHD patients seem less tolerant of the adverse effects than depressed patients; they are more troubled by the anticholinergic effects, weight gain, and impaired sexual functioning. Selective serotonin reuptake inhibitors (SSRIs), lithium (Eskalith), and antipsychotics are generally ineffective in patients without major mood disorders, although limited clinical experience suggests that the combination of SSRIs and stimulants may be of considerable benefit in individuals with ADHD and major depressive disorder.

Good clinical management requires having another person help evaluate treatment responsiveness. Previously untreated ADHD adults are often as unobservant about their behavior, its effect on others, and its changes with treatment as children with the disorder. A part of effective drug treatment requires teaching patients to observe and monitor their own behavior; therefore, another person is necessary. Failure to have an external judge results in false-negative treatment outcomes—substantial changes that are unobserved by the patient.

Appropriate management also requires assessing target symptoms. Of accurately diagnosed patients whose treatment progress is evaluated with the help of another person, 60 to 70 percent experience moderate-to-marked therapeutic benefit (an increase of 15 to 20 points on the Global Assessment of Functioning Scale), which often improves with time; indeed, the most dramatic benefits are seen over years.

The most effective psychological intervention begins with education. Patients are taught that some of the problems have a biological origin and can be suppressed by the use of medication. Patients learn how to monitor the severity of these symptoms and gradually come to understand that these biological problems have engendered psychological, vocational, and interpersonal problems that are, so to speak, self-sustaining. Psychological intervention is then directed at recognizing and altering these maladaptive patterns. Only individual and couple therapies have been used to date. Group approaches using these principles would likely be even more effective.

A 55-year-old high school teacher entered a long-term study of methylphenidate as a symptomatic volunteer. His Parents' Rating Scale score placed him in the 99th percentile of “hyperactivity” while his WURS placed him in the most severe category. His presenting problem was that his temper and verbal abuse of students had led to his being given the choice of quitting or being fired from his teacher's job. He perceived himself as the “most unlucky and misunderstood” person in the world. He manifested all seven symptoms of ADHD, was close to abusive at home, and had had a serious drinking problem (like most members of this family) which he controlled with the help of Alcoholics Anonymous. At the time he chose to quit his job, his family life (four biological and four adopted children) was chaotic, and several of the children were having emotional problems. His wife was ready to leave but could see no way to support the children financially or psychologically.

He was given a placebo-controlled trial of methylphenidate (15 mg every 2½ hours, 6 times daily). After this his wife described marked changes: he could concentrate throughout an entire telephone conversation (something he had been unable to do before), could communicate with his children and follow through with reasonable discipline, and manifested none of his early disabling temper outbursts. While his wife worked, he decided to return to school and study computers. On the basis of these skills, he obtained a government job that required great attention to detail and more concentration than “I ever could have had produced before this study.” At the job he helped integrate computers throughout his department, and several of his ideas were incorporated as policy. Over a period of 5 years he was advanced several times and became the supervisor of seven employees. At the same time, he returned to school and obtained a 4.0 average in mathematics and a 3.7 average overall. Family life stabilized, his wife reported that the marriage was better than it ever had been, and the children's psychological problems were greatly improved.

SUGGESTED CROSS-REFERENCES

ADHD in children covered in Section 39.1. Sympathommetics (psychostimulants) are covered in Section 31.27, MAOIs, in Section 31.20, and tricyclic drugs in Section 31.30.

SECTION REFERENCES

Deutsch CK, Swanson JM, Bruell JH, Cantwell DP, Weinberg F, Baren M: Over-representation of adoptees in children with the attention deficit disorder. Behav Genet *12:*231, 1982.

Guze SB: *Criminality and Psychiatric Disorders*. Oxford University Press, New York, 1976.

*Heiligenstein E, Guenther G, Levy A, Savino F, Fulwiler J: Psychological and academic functioning in college students with attention-deficit/hyperactivity disorder. J Am Coll Health *47:*181, 1999.

Klein RG, Abikoff H, Klass E, Ganeles D, Seese LM, Pollack S: Clinical efficacy of methylphenidate in conduct disorder with and without attention deficit hyperactivity disorder. Arch Gen Psychiatry *54:*1073, 1997.

*Mannuzza S, Klein RG, Bessler A, Malloy P, La Padula M: Adult outcome of hyperactive boys: Educational achievement, occupational rank, and psychiatric status. Arch Gen Psychiatry *50:*565, 1993.

*Mattes JA, Boswell L, Oliver H: Methylphenidate effects on symptoms of attention deficit disorder in adults. Arch Gen Psychiatry *41:*1059, 1984.

Reimherr FW, Wender PH, Wood DR, Ward M: An open trial of L-tyrosine in the treatment of attention deficit disorder, residual type. Am J Psychiatry *144:*1071, 1987.

Safer DJ: A familial factor in minimal brain dysfunction. Behav Genet *3:*175, 1973.

Spencer T, Wilens T, Biederman J, Faraone SV, Ablon JS, and Lapey K: A double-blind comparison of methylphenidate and placebo in adults with childhood-onset attention-deficit hyperactivity disorder. Arch Gen Psychiatry *52:*434, 1995.

Ward M, Wender P, Reimherr F: The Wender Utah Rating Scale: An aid in the retrospective diagnosis of attention deficit hyperactivity disorder. Am J Psychiatry *150:*885, 1993.

*Weiss G, Hechtman LT: *Hyperactive children grown up: ADHD in children, adolescents, and adults*, ed 2. Guilford, New York, 1993.

Wender PH: *Minimal Brain Dysfunction in Children*. Wiley, New York, 1971.

*Wender PH: *Attention-Deficit Hyperactivity Disorder in Adults*. Oxford University Press, New York, 1995.

Wender PH, Reimherr FW: Bupropion treatment of attention-deficit hyperactivity disorder in adults. Am J Psychiatry *147:*1018, 1990.

Wender PH, Reimherr FW, Wood DR: Attention deficit disorder (‘minimal brain dysfunction') in adults: A replication study of diagnosis and drug treatment. Arch Gen Psychiatry *38:*449, 1981.

Wender PH, Reimherr FW, Wood D, Ward M: A controlled study of methylphenidate in the treatment of attention deficit disorder, residual type, in adults. Am J Psychiatry *142:*547, 1985.

Wender PH, Wood DR, Reimherr FW, Ward M: An open trial of pargyline in the treatment of attention deficit disorder, residual type. Psychiatry Res *9:*329, 1983.

Wood DR, Reimherr FW, Wender PH: Effects of levodopa on attention deficit disorder, residual type. Psychiatry Res *6:*13, 1982.

*Wood DR, Reimherr FW, Wender PH: The use of L-deprenyl in the treatment of attention deficit disorder, residual type (ADD,RT). Psychopharmacol Bull *19:*627, 1983.

Wood DR, Reimherr FW, Wender PH: The treatment of attention deficit disorder with dl-phenylalanine. Psychiatry Res *16:*21, 1985.

Wood DR, Reimherr FW, Wender PH, Johnson GE: Diagnosis and treatment of minimal brain dysfunction in adults. Arch Gen Psychiatry *33:*1453, 1976.

Wood DR, Wender PH, Reimherr FW: The prevalence of attention deficit disorder, residual type, or minimal brain dysfunction, in a population of male alcoholic patients. Am J Psychiatry *40:*95, 1983.

*Yuen KM, Pelayo R: Sleep disorders and attention-deficit/hyperactivity disorder. JAMA *281:*797, 1999.

CHAPTER 40. DISRUPTIVE BEHAVIOR DISORDERS

HANS STEINER, M.D.

The introduction of a psychopathological grouping containing disorders manifesting themselves mainly in antisocial conduct is a recent advance but not one without controversy. Prior to the introduction of these diagnostic categories that imply discontinuity from normal development, there existed voluminous research on aggression, delinquency, and criminality from a dimensional and sociological perspective. The ultimate validity of these categorical diagnoses remains to be established. In addition, there is some concern that by transferring the focus of study from criminogenic environments to psychopathology, the essence of these disorders will be misrepresented. Disruptive behavior disorders contain two subgroups: oppositional defiant disorder and conduct disorder. Both diagnoses have been remarkably robust in several field trials. Both categories are to be distinguished from antisocial behavior, where the kind of temporal, qualitative, and quantitative syndromal clustering is lacking. All three entities are independent of the legal status of the child or adolescent, offering opportunities for study and intervention regardless of legal proceedings. Such an independence is increasingly warranted as evidence accumulates that certain biological and psychological factors (in addition to social ones) define these disorders. These additional factors can be genetic, constitutional, or environmental in origin. They offer new ways of conceptualizing necessary prevention, diagnosis, and treatment independent of criminological management. Such emerging data help in the active debate as to whether it is most accurate and advantageous to address different forms of antisocial behavior as ecological adaptations to psychosocially toxic environments or as psychopathology that is inherently located within an individual. The predominant opinion at the present time is that only a subset of delinquency will ultimately be ascribed to psychopathological conditions, but this subcategory will be a significant one because these cases will have the highest chances of persisting in their criminal activities regardless of environment and developmental phase. Finding accurate positive predictors of disruptive behavior in early development is an important issue. Finally, debate is ongoing as to the best way to conceptualize any antisocial behavior (dimensional versus categorical or a combination of the two). There is some support from retrospective data in the Epidemiological Catchment Area Study that later substance abuse was best predicted in a linear fashion. Other aspects of antisocial behavior have not been comparatively tested. Current criteria of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM) for oppositional defiant disorder and conduct disorder suggest discontinuities after a certain number of criteria are met, but data do not necessarily support this concept. On the other hand, the criteria as they currently stand are purely descriptive and allow only for some contextual interpretation, thus leading to considerable within-category heterogeneity. Different subtypes will be found and there is an active debate on the best way of subdividing these large categories.

OPPOSITIONAL DEFIANT DISORDER

Definition This disorder consists of negativistic, hostile, or defiant behavior, creating disturbances in one of three domains of functioning (academic, occupational, or social), lasting at least six months. The diagnosis also refers to angry and vindictive behavior and problems with control of temper. Most of the behaviors are directed at someone, that is, an authority figure. However, there are no major antisocial violations. The behavior is also not part of a developmental stage (i.e., coercive behavior around ages 2 to 3 years and in early adolescence). The diagnosis is not limited to a particular age group, but most commonly emerges in late-preschool or early–school-age children. Exclusion criteria for other diagnoses (conduct disorder, antisocial personality disorder, psychotic or mood disorder) are added. The diagnosis implies more circumscribed disturbances of lesser severity than in conduct disorder, but represents more troublesome behavior than normative oppositionality. Behaviors of the oppositional defiant disorder type on the average appear 2 to 3 years earlier than those of conduct disorder. The latest factor analysis also suggests that there is significant coherence of the oppositional defiant disorder behaviors as outlined in the diagnostic criteria. However, support for the diagnosis has not been uniform; some authors question its status as a separate diagnosis or even as any kind of taxon at all and there have been some negative public reactions that categorizing oppositional defiant disorder as a disorder is an attempt to characterize normative behavior as pathological.

History The Group for the Advancement of Psychiatry suggested this diagnosis first in 1966 and thus it represents a relatively new addition to the taxonomy.

COMPARATIVE NOSOLOGY

DSM In 1980 the third edition of DSM (DSM-III) included this diagnosis as *oppositional disorder* for the first time. There were five criteria, of which at least two needed to be met. In 1987 the revised third edition of DSM (DSM-III-R) expanded the diagnostic label to include the word “defiant.” The list of criteria was expanded to nine, and at least five had to be met for a diagnosis. In 1994 the fourth edition of DSM (DSM-IV) reduced the criteria to eight and required that at least four be met. The current set of criteria has good reliability when parents are the informants, has high internal consistency, and negative and positive predictive value.

ICD The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) includes this diagnosis for the first time under the category of conduct disorders. The label is distinct from Diagnostic and Statistical Manual of Mental Disorders in two major ways: there is an assumption, partially supported by data, that it is a milder form of conduct disorder, that is, that there are significant continuities between the two diagnoses; and (2) it specifies an age of occurrence, discouraging use in older children. The general criterion for conduct disorder and four of the specific symptoms listed must be met, but no more than two acts from the more severe antisocial acts of twenty-three items (as opposed to six for a conduct disorder diagnosis). There is no additional listing of behaviors specific to oppositional defiant disorder. Comparative studies assessing ICD and DSM diagnoses are not available.

Epidemiology The epidemiological data for oppositional defiant disorder need to be regarded with some caution because of the recent modifications of the diagnostic criteria. The point prevalence of the disorder has been reported to vary between 1.7 to 9.9, with a weighted average of 5.7 percent. The average age of onset is about six years. Boys outnumber girls in the prepubertal age range, after which the two genders are more equal. The disorder occurs mostly in cohorts of lower socioeconomic status.

Etiology The etiological studies of this diagnosis are usually contained in the body of research on conduct disorder. So far there are no separate systematic investigations into the origin of oppositional defiant disorder. The disorder is believed to be multifactorial and developmental in origin and a cumulative risk factor model might help explain its genesis. Such a model leaves open the questions of discontinuity (i.e., it allows for the possibility that qualitative changes may result in a categorical diagnosis). It is also consistent with a dimensional conceptualization where linear accumulation of risks leads to more disorder and impairment.

Biological Factors Most authorities believe that genes, temperament, and other constitutional factors may be responsible in interaction with adverse social factors in the production of oppositional defiant disorder. There is familial clustering of certain disorders that are at least in part related to oppositional defiant disorder: other disruptive behavior disorders, attention-deficit/hyperactivity disorders, substance-use disorders, and mood disorders are common in family members.

Psychological Factors Very little is known regarding the role of psychological factors in this disorder. Attachment theorists have noted the similarities between the behavioral manifestations of insecure attachment and disruptive behavior disorders. Antisocial behavior is seen as a special signal to an unresponsive parent. Findings linking infant attachment status with problematic behavior in preschool have been inconsistent, but interesting. Oppositional defiant disorder has been linked to the presence of anxious-avoidant attachment in particular. Insecure attachment also predicts aggression in elementary school in boys and multiple behavior problems in the classroom. Given the inconsistency in the data, most experts believe that multivariate transactional pathways will be found in this area of research. Another important area of research in a related area is the work of Kenneth Dodge who focused on aggressive children's deficient information processing in regard to social stimuli. Aggressive children have shown deficits at every phase of this multistage process: they underutilize pertinent social clues, misattribute hostile intent to peers, generate fewer solutions to problems, and expect to be rewarded for aggressive responses.

Social Factors Ecological factors such as poverty, crowding, and high-crime neighborhoods are believed to contribute to this disorder, but are mostly mediated by poor familial functioning and deficient parenting.

Developmental Trajectories One of the main explanatory problems in developmental psychopathology is the simultaneous existence of stability in behavior paired with its protean ability for appearance. Nowhere is this as apparent as in the area of antisocial behavior and its syndromal disturbances. Models need to be developed

that do justice to this heterotypic continuity. Underlying process may remain stable, but manifest disturbances can change depending on the context of the situation and the developmental phase of the individual in question. One of the crucial issues for the diagnostic category of oppositional defiant disorder is the demonstration of its continuity to conduct disorder and antisocial personality disorder. One longitudinal study establishes such a link: researchers have demonstrated that about one third of boys with oppositional defiant disorder progress to suffer conduct disorders. Conversely, 90 percent of boys with conduct disorder previously fulfilled criteria for oppositional defiant disorder. Thus, the syndrome has considerable sensitivity for the prediction conduct disorder from oppositional defiant disorder, but the positive predictive power is much less. One half of the sample retained oppositional defiant disorder diagnoses while one quarter desisted from oppositional defiant disorder at 3-year follow-up. It is an open issue whether children with oppositional defiant disorder who do not go on develop conduct disorder develop other psychiatric diagnoses.

Diagnosis and Clinical Features Typically, the patient is brought in by parents for problems in the home where difficulties are usually contained. Only with increasing severity does the child's behavior become out of control outside of the home. The DSM-IV diagnostic criteria for conduct disorder are listed in Table 40-1; Table 40-2 presents the ICD-10 criteria for conduct disorders.

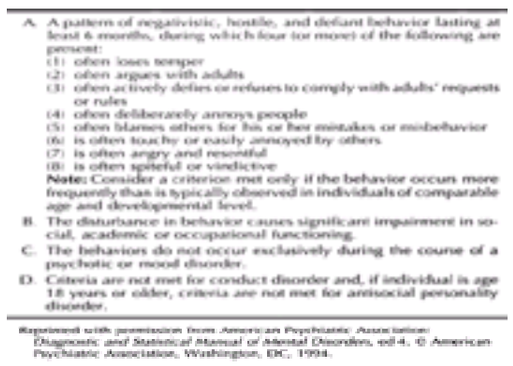

A. A pattern of negativistic, hostile, and defiant behavior lasting at least 6 months, during which four (or more) of the following are present:
   (1) often loses temper
   (2) often argues with adults
   (3) often actively defies or refuses to comply with adults' requests or rules
   (4) often deliberately annoys people
   (5) often blames others for his or her mistakes or misbehavior
   (6) is often touchy or easily annoyed by others
   (7) is often angry and resentful
   (8) is often spiteful or vindictive
   Note: Consider a criterion met only if the behavior occurs more frequently than is typically observed in individuals of comparable age and developmental level.
B. The disturbance in behavior causes significant impairment in social, academic or occupational functioning.
C. The behaviors do not occur exclusively during the course of a psychotic or mood disorder.
D. Criteria are not met for conduct disorder and, if individual is age 18 years or older, criteria are not met for antisocial personality disorder.

Reprinted with permission from American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 40-1 DSM-IV Diagnostic Criteria for Oppositional Defiant Disorder

Table 40-2 ICD-10 Diagnostic Criteria for Conduct Disorders

Robert, age 7, presented to the consultation-liaison team. He suffered from leukemia and was extremely difficult to manage. He would refuse all necessary blood work and repeatedly ran away from the clinic when asked to cooperate with requests for X-rays, blood tests, etc. He was sullen, argumentative, and irritable. The behavior was unchanged when mother was used as a “filter” for the demands. He was a chronically cranky child, although his illness was in remission and he was mostly medication free at the time of consultation. His care was severely compromised. At home his behavior was very similar and had been so for several months. He continuously argued with his single mother about any kind of request, such as cleaning up his room. Prior to his difficulties at the hospital, he began to exhibit similar problems at school. He was suspended for one week for being verbally abusive and out of control with his teacher. His mother had been diagnosed with acquired immune deficiency syndrome (AIDS) 2 years ago, having become infected by her drug-abusing husband who had died about 3 years ago from the effects of AIDS. She seemed dysphoric, passive, and extremely permissive with the child. She had no additional help because her own mother was also gravely ill and she had no current social support network. The patient would often scream at his mother at the top of his voice without her evidencing any kind of reaction or making any attempt to contain him. Testing showed him to be of normal intelligence, without the symptoms of attention-deficit/hyperaction disorder or learning disabilities. He knew of his father's death and his mother's illness. His mood remained dysphoric for many weeks when discussing this. He had never exhibited other reactions to being told about the illnesses of his parents. His mother described his early development as unremarkable except for his tendencies to have irregular sleep and eating patterns and his propensity to be cranky. His mother had intermittently taken drugs while pregnant with him, her only child.

This case illustrates a severe form of oppositional defiant disorder. In part, the severity was iatrogenic because the pediatric team and the school had attempted to manage this child without psychiatric input in order not to add insult to injury. The mother had been dysfunctional for many years. Her inability to regulate the child's behavior was not the effect of AIDS impairment; her illness compounded longstanding problems with mood regulation and self-efficacy. As is usually true with externalizing disorders, there was substantial discrepancy between the medical team, the school team, the parent, and the child as to precisely which behaviors presented problems and when, highlighting the necessity for multiple informants in the assessment of such cases.

Pathology and Laboratory Examination There are no pathognomonic tests and findings. A general chemistry panel may be helpful to identify chronic illness that might contribute to the problem. Often oppositional defiant disorder may be the first manifestation of more subtle specific developmental disabilities, and thus assessment of intelligence and learning might be helpful. Temperamental rating scales such as the Dimensions of Temperament–Revised, normed from preschool to adulthood and available in self and observer rating format can contribute to assessment.

Differential Diagnosis and Comorbidity Delineation of oppositional defiant disorder from normative oppositional behavior, transient antisocial acts, and conduct disorder is of paramount importance. Oppositional defiant disorder is not transient, leads to significant impairment, but does not involve major violations of the law and the rights of others. Attention-deficit/hyperactivity disorder is the most common comorbidity: between 25 to 60 percent of children with oppositional defiant disorder also fulfill criteria for attention-deficit/hyperactivity disorder by parental report, and half of attention-deficit/hyperactivity disorder children have oppositional defiant disorder. As with conduct disorder, the association of oppositional defiant disorder and attention-deficit/hyperactivity disorder confers poor prognosis. Youngsters tend to be more aggressive, show a greater range and persistence of problem behaviors, are rejected at higher rates by peers, and underachieve more severely in the academic domain. Furthermore, attention-deficit/hyperactivity disorder facilitates the early appearance of oppositional defiant disorder and conduct disorder. Antagonistic behavior is commonly found in internalizing disorders in this age group: dysthymic disorder, major depressive disorder, and early-onset bipolar I disorder should be considered. Anxiety disorders, especially separation anxiety disorder, can present with predominant temper control problems. Pervasive developmental disorders also can demonstrate oppositionality, but the underlying bizarre problems with relating to others are usually absent in oppositional defiant disorder children.

Course and Prognosis With the exception of one 3-year prospective study there is no information about the naturalistic progression or response to treatment in these children. Most of these children will not develop conduct disorders or antisocial personality disorder. One half will show stable signs of oppositional defiant disorder after 3 years; 25 percent will have no further diagnosis. Extrapolating from studies on conduct disorder and oppositional defiant disorder about one quarter of the initial total will go onto develop conduct disorders and about 10 percent will progress into antisocial personality disorder.

Treatment Adequately designed studies of controlled treatments are not available. The clinician needs to develop an individualized treatment plan based on the specific situation. First, the setting for the intervention needs to be determined. Issues of the safety of patients and those around them need to be considered. Intensive settings are needed for crisis management only. Long-term solutions are typically necessary to achieve positive outcome. Family therapy, parent management training, and conjoint treatment of parent and child along psychodynamic lines have some empirical support; day treatment may also be effective. In all

cases, there should be close collaboration with schools. Psychopharmacological treatments are mostly indicated to address issues of comorbidity, as there are no data on the pharmacological treatment of oppositionality without comorbidity. Generally, it is believed that early application of treatment is more effective than later.

CONDUCT DISORDER

Definition *Conduct disorder* is a clinical term referring to the clustering of persistent antisocial acts of children and adolescents. The condition is thought to be due to underlying psychopathology leading to significant impairment in one or more domains of functioning. The symptoms are clustered in four areas: aggression to people and animals, destruction of property, deceitfulness and theft, and serious violations of rules. Subtyping is allowed based on the age of onset of symptoms. Severity can be specified as mild, moderate, or severe. The category is currently conceived of as a polythetic diagnosis in that no one specific criterion is necessary for and any combination of criteria will suffice to establish the diagnosis. There is no formal provision for evaluating the context in which these antisocial clusters occur. Both these features contribute to the fact that the category is inherently heterogenous. The current criteria require that at least three of a list of fifteen antisocial behaviors be present over a period of 12 months; one of them has to be present in the past 6 months. Exclusion criteria for antisocial personality disorder are added. In epidemiological studies, this category has been robust, especially in its most recent more stringent versions. Its inherent heterogeneity has made conduct disorder less useful for causal and treatment studies. National practice guidelines have been developed for this disorder by this author.

History Clinical interest in antisocial behavior began with the specialty of forensic pathology. Early investigations focused on the detection of criminal stigmata in physiognomy and anatomy. Subsequently, Richard von Krafft-Ebing's detailed descriptions of deviant sexual behavior began to move the field into the realm of behavior observations. Following the writings of Sigmund Freud, whose speculations regarding aggression as an instinct were of direct relevance, psychiatric clinicians began the exploration of antisocial acts as psychopathology. Soon afterwards, ethologists began their naturalistic observations of animal aggression and explored its relationship to human behavior, providing ontogenetic and phylogenetic anchors. Konrad Lorenz was especially interested in aggression as an instinct. Following the results from several longitudinal studies that demonstrated the stability of antisocial acts within the same individuals over time and in diverse environments, their poor response to criminological management and supportive counseling, interest intensified in the clinical definition of antisocial behavior. Psychometric work in the middle of the twentieth century established the coherence of antisocial and aggressive patterns of behavior. At the same time, John Bowlby was studying subtypes of delinquents. The description of “affectionless characters,” with their histories of prolonged disruptions of early relationships, became one of the sources of inspiration for the study of attachment and also provided a bridge from analytic concepts and ethology to the clinical study of antisocial acts. Most recently, interest in biological substrates has reemerged. Stella Chess and Alexander Thomas introduced the notion of the “difficult temperament” which served as an early childhood antecedent of behavior problems in some boys; Robert Cloninger further specified the risks. High novelty seeking with low harm avoidance, both heritable traits, were instrumental in generating risk of conduct disorder. Recent studies of genetic transmission indicate certain heritability for antisocial behavior.

COMPARATIVE NOSOLOGY

DSM The first edition of DSM (DSM-I) did not delineate children's antisocial acts; the second edition (DSM-II) listed three reactions, suggesting a strong environmental bias; and DSM-III for the first time introduced conduct disorder as a syndrome. One action, displayed over long time periods, was necessary for diagnosis. There were four subtypes of conduct disorder, based on a 2 × 2 matrix on the axes of socialization and aggressivity. The resulting categories, informed by the initial psychometric work and supported in factor analytic studies by Herbert Quay, showed poor reliability and subsequently were dropped in 1987 in the DSM-III-R. DSM-III-R made the diagnostic criteria more stringent by expanding them to 13 and requiring that at least three were present for at least 6 months. This reflected the empirical finding that diversity of antisocial behaviors rather than specificity of a behavior best predicted chronic antisocial problems and recidivism. Subtyping was changed: group or socialized; solitary or aggressive; and undifferentiated subtypes were permitted. This subtyping was dropped in DSM-IV, reflecting its poor reliability. Instead, a subtyping by age of onset is permitted. Once again, criteria were made more stringent: in addition to three of 15 behaviors being present for at least 12 months, one of those three has to have been present in the past 6 months. There is still no differential weighing of the various antisocial behaviors listed, contributing to the poor negative predictive value. Some of the behaviors listed are endorsed by a majority of adolescents (e.g., truancy from school, staying out all night despite parental prohibitions) whereas others occur only exceedingly rarely (carrying out robberies).

ICD There are substantial differences between the two classifications in the definition of this disorder. For one, ICD-10 lists this grouping as a plural to indicate the heterogeneity of the category. ICD-10 only requires one major behavior to have been present in a repetitive and persistent pattern in the past 6 months; isolated dyssocial acts are excluded. A list of 23 behaviors is given and there is some weighing of criteria, as seven behaviors of greater severity (i.e., forcing sexual activity; using weapons; being physically cruel to others) need to have been present only once in order for the criterion to be fulfilled. The instructions exclude behaviors due to other illness and the clinician is encouraged to use other diagnoses as primary if present. Hyperkinetic disorder, mood disorders, pervasive developmental disorder, and schizophrenia are excluded. Subtyping by age of onset is recommended. In addition, describing cases in terms of the following characteristics: hyperkinesis, emotional disturbance and severity (mild, moderate and severe), is also recommended. Other subtypes are: conduct disorder confined to family context; unsocialized conduct disorder; socialized conduct disorder; oppositional defiant disorder; other conduct disorder; conduct disorder unspecified. In addition, there is a separate category provided for mixed disturbances of conduct and emotions, where both criteria for emotional disturbance (anxiety, depression, other emotion), an adult-type neurotic diagnosis, or a mood disorder are met. Subtypes here include depressive conduct disorder and other mixed disorder. ICD thus uses finer stratification of conduct disorder and avoids comorbidity diagnoses. The proposed subtypes await definitive empirical validation.

Epidemiology Conduct disorders are among the most frequently referred, diagnosed, and treated psychiatric problems. A number of large and well-designed epidemiological studies employing a range of assessment methods provide information. However, the diagnostic criteria have been changed substantially since the early 1980s, so results of older studies (i.e., prior to DSM-III-R) may not be fully comparable to current ones. The general population prevalence is estimated somewhere between 1.5 and 3.4 percent of children and adolescents when clinical interviewing is used as a method of detection, but reported prevalences vary widely (between 1 and 16 percent) depending on the samples used, the definitions given to the disorder, and the instruments applied. The male-to-female ratio has been found to range between 5:1 and 3:1, depending on the age range studied, but at all ages boys predominate over girls. It is only in adolescence that the gap between the sexes begins to close because of the increase of the disorder in girls. There is debate as to whether diagnostic criteria should be modified for girls to reflect their greater propensity for covert antisocial acts. Peak age of onset is in late childhood and early adolescence, but onset can range from preschool to late adolescence. The occurrence of conduct disorder appears to have increased in the 1990s, which may be a reflection of the increase in the adolescent cohort or improved case identification and diagnostic methods. At any given time it is estimated that about 1.3 to 4 million children and adolescents are afflicted with conduct disorders.

Etiology Most authorities agree that conduct disorder is a heterogenous disorder. It can best be described as a final common pathway for several initially divergent developmental trajectories. Although there is currently no agreement about a uniform model applying to all forms of conduct disorder, one possible model for the combination of causal factors is that of genetic liability triggered by an environmental adversity, mediated by other factors such as poor coping. Although there is some debate on the relative importance of the factors that have been implicated, there is general support for the developmental nature of the disorder. The following model describes one possible integration of such risk and resiliencies in the genesis of this disorder.

Risk-Resilience Model for Conduct Disorders This model posits that it is the gradual accumulation of risk as well as the absence or weak presence of protective factors and their interactions that ultimately lead to the conduct disorder rather than single risk factors operating in isolation. Rolf Loeber has illustrated the gradual stacking of factors in the genesis of conduct disorder. An expanded model would include a parallel pyramid of resilience or protective factors, balancing the gradual aggregation of risk as it accumulates over time. Figure 40-1 portrays the predominance of risks in ecological (e.g., poverty), constitutional (e.g., difficult temperament), and parenting (e.g., poor response to coercive behaviors, abuse) factors. This results in poor internal self-regulation, which becomes manifest especially during school age. School performance is also affected because these children lack the skills to deal with authority and cannot fulfill their academic potential. Peer relationships also suffer as the child tends to find acceptance only from similarly socially inept peer groups. As there is an increasing aggregation of risk it takes more and stronger protection to offset the risk and more domains may be adversely affected. Empirical data also support the fact that as risks accumulate in number, the greater is the chance that conduct disorder will develop via multiple interactive loops between risks.

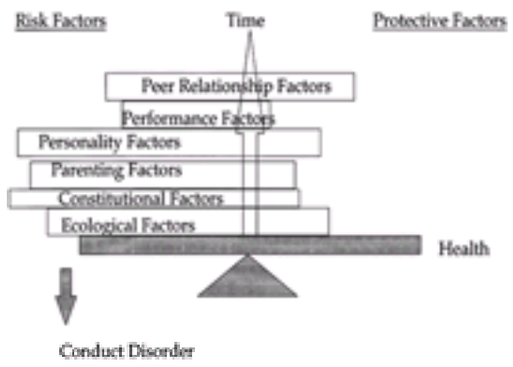

FIGURE 40-1 Developmental pathways to disruptive behavior disorders.

Biological Factors The familial aggregation of the disorder has suggested that there may be a genetic risk for conduct disorder. Data from the comparison of monozygotic and dizygotic twins in terms of concordance rates for criminality and the results of fast adopt studies support such a link, but a cautionary note is in order: these heritability estimates were mostly based on adult criminal populations, which represent the most severely ill individuals. The results also apply mostly to nonviolent offending, which appears paradoxical. The results in juvenile delinquency are much less impressive, most likely because of the high prevalence rates and the heterogeneity of conduct disorder. However, the results are interesting, suggestive, and merit attention. Most authorities feel that it is unlikely that any one gene or even any combination of genes would account for the occurrence of such highly complex behaviors as are expressed in conduct disorder, but that there may be genetic causes for certain risk factors, such as hyperactivity. Recent reviews present new and promising leads. Complex interactions of constitutional risk and environmental adversity will better explain etiological variance than any single factor model. Certain psychophysiological abnormalities have repeatedly been reported as constituting risk for the development of conduct disorder, supporting the claim that biological abnormalities that are under partial genetic control are involved in the problem. The best researched area is the autonomic nervous system, which shows low baseline activity and reactivity on a variety of parameters in the early-onset group, paralleling findings in the adult psychopathy studies. These abnormalities are presumed to relate to diminished avoidance conditioning in response to socialization. Another abnormality is the high base activity and reactivity in this system in the late-onset group, which has been shown to relate to increased desistance from crime on follow-up. Neurotransmitter abnormalities are found with some regularity, although the findings are less consistent when compared to the psychophysiological literature. Of particular interest are the compounds reflecting the activity of sympathetic arousal, which have direct connections to anxiety and sensation-seeking behavior. The latter is inversely related to 3-methoxy-4-hydroxyphenylglycol (MHPG) concentrations in older youth, but not in younger boys with delinquent behavior. Abnormalities in the systems reflecting noradrenergic and dopaminergic activity have been described, although these findings do not hold uniformly across other small samples. More recently, serotonin has been added to the list of neurotransmitters implicated in aggressive youth, but the literature is not methodologically strong. Neurotransmitter findings have been linked with an imbalance favoring the behavioral reward over the behavioral inhibition system. Many of these findings raise issues about the causal relationship between environmental stress and biological factors, because environmental adversity can result in such alterations. Multiple promising leads in terms of the biological implications of child maltreatment are currently identified, including associations between abuse and reduced hippocampal size, and lack of autonomic reactivity in victims of extended and severe abuse.

Sex is a clear risk factor in multiple studies. Much less is known about females in this population, although some recent studies have begun to document details. The findings implicating gender-specific hormones in the genesis of the disorder are very inconsistent and the current understanding is that responses to androgens depend on the biochemical, environmental, and historical context of the individual involved. The clear and consistent findings in animals regarding the role of androgens in aggression, brain development, and social dominance are not paralleled in the human literature. Difficult temperament has been repeatedly implicated in the genesis of the disorder. It could work in at least two ways: it may make children more likely to be the target of parental anger and thus of poor parenting or it may be directly linked to behavior problems later on. Inappropriate aggression at an early age, especially in combination with shyness, predicts later conduct disorder. Chronic illness and disability also have been known to increase the prevalence of conduct disorder, especially if the primary illness affects the central nervous system (CNS). Chronically ill children have three times the incidence of conduct problems than healthy peers; if chronic illness affects the CNS, the risk is about five times as high.

Psychological Factors Conduct disorder is more likely to be paired with diverse and complex disturbances in psychological domains. The origin of these disturbances is not clear, but their presence implies that many risks for conduct disorder are retained and internalized and independent of specific environments. Academic underachievement, learning disabilities, and problems with attention span and hyperactivity are all associated with conduct disorder. Hyperactivity, especially in the presence of poor parental functioning, is a risk; it seems to facilitate rapid development of conduct disorder. Neuropsychological deficits have been documented implicating frontal and temporal lobe dysfunctions. Laterality and language performance are disturbed. Higher personality functions are also affected: in complex social situations, children with conduct disorder have been shown to perceive fewer appropriate responses, lack the skills to negotiate conflict, and lose their ability to restrain themselves when emotionally stressed.

Social Factors Poor family functioning, familial aggregation of drug and alcohol abuse, psychiatric problems, marital discord, and especially poor parenting are all associated with conduct disorder. Abusive, neglectful parenting and child maltreatment is a highly specific risk factor for the development of conduct disorder. The specific parenting patterns that contribute to the development of conduct disorder have been described as training in noncompliance by inconsistent responses to coercive behavior of the child and by capitulating to demands in response to the child's coercion. There is fairly substantial evidence that viewing televised or other media violence and violence in the child's community contributes to the development of conduct disorder problems, especially in children who are at high risk for other reasons. Socioeconomic disadvantage as manifested in poor housing, crowding, and poverty all exert consistently negative influences.

Protective Factors Protection implies that certain factors affect core aspects of functioning positively and pervasively in the presence of considerable risk. Much less is known about these factors at the present time than is known about risk. An easy temperament and high base rates of autonomic nervous system activity offer protection. Ability to relate to others, good work habits at school, areas of competence outside school, and high intelligence are important. Social factors include a good relationship with at least one parent or another important adult, prosocial peers, a good school that fosters success, responsibility and self-discipline, and the selection of a good mate (as demonstrated by a stable interpersonal relationship, a good work history, and capacity for good parenting). Currently, there are no studies available that look at the experimental manipulation of these protective variables.

Diagnosis and Clinical Features The clinical manifestations of conduct disorder can be very diverse and difficult to assess. Research has repeatedly supported a fundamental division among externalizing behaviors: aggressive-antisocial on one hand and impulse-reactive on the other, but this distinction is not always easily made. In addition, the context in which these diverse symptoms occur is quite variable. Although DSM-IV does not provide a contextual subtyping and qualification of the symptoms (Table 40-3), there is a stipulation that symptoms only be considered as indicative of conduct disorder if the behavior in question is indicative of an underlying dysfunction in the individual and not a reactive adaptation to an immediate social context. In reality, it may be quite difficult to separate these two causes. Furthermore, patients and other informants differ widely in their appreciation of some of the problems and their evaluation, and thus it is of great importance to use a range of methods for assessment across different settings, addressing multiple domains, and using multiple informants. In general, there is a tendency for lower-grade delinquent behaviors to appear first, and gradually for more severe ones to appear later in the course of conduct disorder. In taking the patient's history particular attention should be paid to data that allow for further subtyping. Childhood-onset conduct disorder is characterized by more aggression, higher likelihood of neuropsychiatric impairment, and poorer prognosis. Other subtypes are still relevant: the literature supports a differentiation of predatory (i.e., planned and coldly executed) from reactive (i.e., occurring in response to frustrating circumstances) aggression. Rolf Loeber's differentiation of separate pathways for antisocial behavior manifesting itself in conflict with authority, overtly or covertly in high-risk youth can be useful. These pathways occur in the context of different comorbidities and family backgrounds and follow different trajectories, although there is a tendency for them to blend together in most severely disturbed youths. Although the distinction between socialized and unsocialized delinquents has been dropped in the DSM system, it is still retained in ICD (Table 40-2) and some authorities believe that it continues to have considerable support and usefulness. An assessment of the interpersonal context of the patient's symptoms would also give additional help in deciding management and prognosis: A pervasive lack of relationships and the isolated predatory nature of certain acts imply a worse prognosis than the crimes committed in consort or under the influence of dyssocial peers. Finally, there have been recent descriptions of conduct disorder that presents with personality profiles of paradoxical overrestraint. These antisocial acts tend to follow a different path than their more common and typical underrestrained conduct disorder cohort: they exhibit an intermittent loss of impulse control, precipitated by specific personal triggers. These crimes tend to be fewer in number, but more severe in degree. These patients do not present with all the typical risks of most delinquent youths, but follow the trajectories of what has been described as the “neurotic delinquent.” Remarkable is the capacity of these patients to selectively ignore negative information, especially about themselves and those whom they love. Such difficulty seems to relate to a repressor profile that has been amply studied in the adult literature. Recent investigations have also shown the validity of this concept in the juvenile age

range.

Table 40-3 DSM-IV Diagnostic Criteria for Conduct Disorder

In addition, ICD-10 contains a separate diagnostic category for patients that meet the criteria for conduct disorders and another emotional disorder. The ICD-10 criteria for mixed disorders of conduct and emotions are listed in Table 40-4.

Depressive conduct disorder

A. The general criteria for conduct disorders must be met.

B. Criteria for one of the mood (affective) disorders must be met.

Other mixed disorders of conduct and emotions

A. The general criteria for conduct disorders must be met.

B. Criteria for one of the neurotic, stress-related, and somatoform disorders or childhood emotional disorders must be met.

Mixed disorder of conduct and emotions, unspecified

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 40-4 ICD-10 Diagnostic Criteria for Mixed Disorders of Conduct and Emotions

It is quite likely that many subtypes will be found to overlap. The late-onset one–channel-disturbed-repressor-socialized profile can be contrasted to the early-onset two–or three–channel-disturbed reactive-undersocialized conduct disorder, with perhaps a distinct core group of more predatory-psychopathic conduct disorder.

J.O. was raised in an impoverished minority neighborhood adjacent to a major urban center on the west coast. The area was infested by crime and gang activity and relatively devoid of community structure and resources. His mother was a teenage single woman with a history of drug abuse dating back to age 14. From age 16 on she lived outside the parental home having been asked to leave by her mother because she was found in bed with her father. J.O.'s mother took up prostitution as a way to support herself. At age 18 she married the father of J.O. because she was pregnant by him. J.O., the child from this relationship, was born with signs of drug addiction. The child was tracked by Child Protective Services because of the circumstances of his birth. He was repeatedly placed outside of the home because he was found in need of food and adequate care. The couple lived on Aid for Families with Dependent Children funds and the mother continued her prostitution, bringing home customers and exposing her infant to the details of her business. In addition, J.O.'s father started a pattern of bringing home men for either his own pleasure or for the business of prostitution. This led to a permanent breakup when the child was about 3 years old. J.O. was removed and placed in his sixth foster home since birth. His foster parents noted a high degree of activity in the child and very difficult temperamental characteristics, which led to frequent harsh parenting on their part. After 6 months the child was placed in another foster home because the current foster parents declared themselves unable to deal with him. A new placement was sought and a psychiatric assessment led to a diagnosis of oppositional defiant disorder and attention-deficit/hyperactivity disorder. A stimulant was started that had some effect. However, with entry into school J.O.'s status deteriorated and he rapidly developed difficulties with teachers, peers, and school work. He was placed in a special classroom and treatment of attention-deficit/hyperactivity disorder and conduct disorder were coordinated between foster home and teachers. An unremitting pattern of lying, stealing, and deceptiveness began to manifest itself in the home and school. Two years into this regimen his mother reappeared and wanted to be reunited with him. After 1 month of trial in the home J.O. once again had to be removed because he was uncontrollably angry and menacing towards others, was found to be cruel to animals in the neighborhood, and had set fire to a local elementary school. He was placed into a different foster home where after 1 year he was found to have been molested by one of the other children in the home and in turn had begun to victimize younger children sexually. He was placed in a locked unit and released into a residential treatment setting, which in turn sent him to another group home after 2 years. He was now about 13 years old, and shortly before his next birthday, he ran away from his placement, surviving by robbing and stealing while beginning to indulge in drug and alcohol abuse. At 15 he was apprehended while driving a stolen vehicle and put on probation. He ran away again and robbed a store at gun point. This lead to his arrest, conviction, and commitment for 5 years to the Youth Authority. His case shows the gradually escalating course of antisocial activities in a child with extreme environmental and severe temperamental risks, leading to a chronic and unremitting pattern of closely clustered antisocial behavior that becomes increasingly independent of the social context in which the child finds himself.

Evaluation From the outset it must be made clear to all concerned that the clinician is predominantly assessing the situation from the viewpoint of the patient. Issues of compliance in the child and the parents will be a prominent part of the process of evaluation. Even with such clarity, it is highly likely that there will be multiple inconsistencies and contradictions that directly bear on the situation. The clinician is well advised to expect these and not attempt to clear them up in one session, lest this lead to alienating and premature confrontations. The best way to approach them is from a longitudinal perspective: as the evaluation unfolds, it will become increasingly clear what transpired. Interviews with parent and child are necessary, as are auxiliary ones. The child should be interviewed alone for significant parts of the evaluation. Probation and legal information should be presented in a dispassionate and unobtrusive manner, as a point of departure for discussion and not as a standard against which the patient's reports are going to be assessed. The evaluation must include a detailed physical examination and an assessment of neuropsychiatric and neuropsychological functioning. Particular attention must be given during the mental status exam to disturbances of mood and affect, suicidality, impulsivity, and associated comorbidities.

Pathology and Laboratory Examination At the present time there are no pathognomonic test findings for this diagnosis, but as the disorder presents with disturbances in multiple domains, several auxiliary assessments can be of value. As part of the routine assessment of conduct disorder, a careful assessment of the physical health domain needs to be performed to rule out chronic illness that contributes to the onset of disorder or to detect the ill-effects of comorbidities and the effects of abuse and general trauma. Head traumas are common in this group of patients, as are the sequelae of high–risk-taking behaviors, such as drug abuse and indiscriminate sexual conduct. Nutritional deficiencies are common and aggressivity has been linked to low levels of trace elements. Imaging techniques have been used in small-scale studies to document specific frontal and temporal lobe dysfunctions, but they are in the early states of investigation and should not be routinely applied. Referral for psychometric testing is always indicated to ascertain the level of intellectual functioning, specific developmental disabilities, personality functioning, and of indicators of psychopathy.

Differential Diagnosis and Comorbidity Conduct disorder has to be distinguished first and foremost from antisocial behavior without underlying psychopathology, oppositional defiant disorder, antisocial personality disorder, and impulse-control disorders. Substance use disorders are extremely common in conduct disorder and can be the primary diagnostic reason for antisocial conduct. Attention-deficit/hyperactivity disorder is the next most common differential. The association with conduct disorder has been so frequent that there has been a debate over combining the diagnoses, but empirical data finds that the disorders differ in terms of premorbid risk and their respective predictive power for adult criminal outcomes. Divergent validity for conduct disorder from attention-deficit/hyperactive disorders is generally

considered as established. The latter disorder is considered the most virulent comorbid condition: it facilitates the early appearance of conduct disorder, which is a strong predictor of adverse outcome. Psychotic disorders, especially those with paranoid processes, can be mistaken for conduct disorder. Internalizing disorders, such as mood disorders, posttraumatic stress disorder, and dissociative disorder can be confused with conduct disorder, although less commonly. Age under 18 prevents the diagnosis of personality disturbances, but in some cases borderline, narcissistic, and antisocial personality disorders should be considered. DSM-IV allows for extensive comorbidities as no exclusion criteria are provided. Conduct disorder is usually accompanied by a wide range of comorbid conditions. These comorbidities contribute independently and interactively to prognosis and outcome. Internalizing and externalizing comorbidities become increasingly common as the disorder becomes more severe. Repeated studies have shown that complex diagnostic patterns are the rule and not the exception in this population, resulting in compound psychopathology that requires special management and treatment. Loeber's data provide an unique opportunity to examine the developmental unfolding of various comorbid states: prior to adolescence, attention-deficit/hyperactivity disorder is most frequently associated with conduct disorder, especially in boys, declining in importance thereafter. The typical sequence seems to be attention-deficit/hyperactivity disorder-opposition defiant disorder-conduct disorder; alcohol and substance abuse follows. Each additional comorbidity adds to the poor prognosis in boys. The picture is different for girls: they also have a higher chance of getting conduct disorder when they suffer from attention-deficit/hyperactivity disorder, but it is not clear that the prognostic implications are as grave. Internalizing disorders, most commonly mood and anxiety disorders, but also somatization disorders seem to appear during adolescence when conduct disorder is firmly entrenched. Depression, in particular, affects both sexes but especially girls, particularly after they have reached pubertal maturation. The evidence on the prognostic impact of internalizing comorbidity and conduct disorder is mixed.

Course and Prognosis There are several studies demonstrating the stability of conduct disorders over time. Although the precise manifestations of the disorder may change, the pattern of repeated violations of major societal rules and laws continues. Modern naturalistic and long-term results suggest that only approximately 40 percent go on to develop the most pernicious variant of the personality spectrum: antisocial personality disorder. The majority of the remainder, however, also lead problematic lives in which many domains of functioning continue to be affected significantly (e.g., interpersonal relationships, ability to maintain healthful styles of living, ability to support themselves). Whether such a poor outcome is attributable to lack of intervention by the penal system remains an open question.

Treatment There is only modest evidence that treatment of conduct disorder is effective. Several recent reviews of the literature and a meta-analysis of over 500 studies show that a wide variety of treatments have been tried and on the average only show modest effect sizes. There is consensus among experts that early intervention is better; prevention is more effective than treatment, (although the evidence for effective prevention programs is also incomplete); and extensive approaches in naturalistic settings are preferable over those who work intensively in special settings, which bear little or no resemblance to the patient's daily environment. Dramatic interventions such as shock, incarceration, or boot camps are not supported by the evidence and may even have negative outcomes. Realistic programs should be multimodal, addressing deficiencies in the multiple domains of functioning. Finally, treatment packages should reflect the developmental needs of the child because there is no one intervention that is effective across all ages.

Preschool Conduct disorder rarely is manifested in this age group, but there is important evidence that interventions may prevent future conduct disorder. Programs such as Head Start have as one of their byproducts the prevention of future delinquency. Programs consist of special stimulation packages for the child, parent education about normal development and maturation, and parental support in times of crisis. Identification of the temperamental characteristics of the child, the goodness of fit between the child and the parent, and the facilitation of parental efficacy, especially in handling normative coercive behaviors and tantrums are reasonable targets for intervention. Stimulants and other medications are used quite frequently in this age group for symptom control, but there is no convincing support that medication is effective in the short or long term.

School-Age Multiple studies show the effectiveness of parenting training and social skills training aimed at improving peer relationships and the child's ability to comply with demands from authority figures as well as the improvement of academic skills. The primary target for intervention should be the child and the family as well as the school environment, that is, teachers and peers. Behavioral techniques are superior to client-centered therapy in short-term outcome. Suitable targets are prosocial functioning and antisocial behaviors. Social competence can also be induced by other techniques, including television viewing of problem situations and fantasy play. Family therapy is one of the mainstays of intervention in this age group and can be used primarily to explore conflict and problems in parenting, but also secondarily as a means of building and maintaining a working alliance with these families. Individual therapy used in isolation is not considered to be an effective treatment approach. Psychopharmacological approaches are definitely indicated to treat significant comorbidities, but evidence that they are effective in cases of uncomplicated conduct disorder is limited. Medication is best used to control symptoms and crises, while the rest of the intervention package is allowed to take effect. A reasonable algorithm for the use of types of medication is discussed below.

Adolescence Youngsters are increasingly beyond parental control and rely on internal psychological structures and peer opinion to guide themselves in their daily conduct. Treatment needs to reflect this shift: individual treatment should probably be added to the package because there needs to be a specific alliance with the patient in order to maintain some compliance with treatment demands. In addition, most youngsters with conduct disorder in this age group will have charges pending or assessed against them, and there most likely will be some curtailment of their freedom in some rehabilitative setting. Furthermore, substances and alcohol play an increasingly prominent part of the patient's symptoms. Their adverse effects and interactions with prescribed treatment need to be considered at all times. The most promising approach with youth in this age group is multisystemic therapy, which aims at keeping delinquents in their psychosocial environments while coordinating family and multiple systems interventions. Outcome studies have shown the clear superiority of this approach over incarceration and other treatments at substantially reduced cost. Psychoeducational packages targeting social skills, conflict resolution, and anger management are available to augment treatment; some have better empirical support than others. Psychopharmacological management is increasingly being considered for this age group, but it is only very recently that we have evidence from double-blind placebo-controlled trials that lithium (Eskalith), selective serotonin reuptake inhibitors (SSRIs), stimulants, and antikindling agents might be effective against aggression and some forms of antisocial behavior, even in the absence of particular comorbidities. There is no evidence that drugs by themselves are helpful in cases of uncomplicated or comorbid conduct disorder in the long run. In addition, antipsychotics, lithium, and carbamazepine (Tegretol), clonidine (Catapres), and propranolol (Inderal) all have been studied to some extent. As there are no definitive data available, the current treatment guidelines recommend the following algorithm. In the presence of mental retardation or organic brain impairment, propranolol or a brief course of psychiatric drugs might be helpful. In the presence of clear danger to others, antipsychotic drugs become the first choice. In the presence of lability of mood or explosive outbursts, lithium, divalproex (Depakote) or valproate (Depakene), or carbamazepine might be helpful. Clonidine can also be considered for explosive outbursts. If all these drugs do not help, buspirone (Buspar) or trazodone (Desyrel) might be prescribed. If irritability is a major antecedent to violence and aggression, SSRIs should be considered.

DISRUPTIVE BEHAVIOR DISORDER NOT OTHERWISE SPECIFIED

This is a residual diagnostic category, intended to be used in those cases that do not fulfill all the criteria but suffer significant impairment in functioning in one of several domains (Table 40-5). No research data is available on the validity and utility of this category.

This category is for disorders characterized by conduct or oppositional-defiant behaviors that do not meet the criteria for conduct disorder or oppositional defiant disorder. For example, include clinical presentations that do not meet full criteria either for oppositional defiant disorder or conduct disorder, but in which there is clinically significant impairment.

Table 40-5 DSM-IV Diagnostic Criteria for Disruptive Behavior Disorder Not Otherwise Specified

SUGGESTED CROSS-REFERENCES

Antisocial behavior in children and adolescents is discussed in Chapter 40, aggression and specific developmental disabilities are discussed in Chapter 38. Attention-deficit disorders are presented in Section 39.1, antisocial personality disorder in Chapter 24, impulse control disorder in Chapter 22, neuropsychological and intellectual assessment of children in Section 7.6, substance use disorder in Chapter 1, normal development of children and adolescents in Section 32.2, mood

disorders in Chapter 45, trauma-related psychopathology in Section 46.2, forensic assessment of children and adolescents in Section 49.12, psychopharmacology of children and adolescents in Section 48.6, family therapy in Section 48.5, individual therapy in Section 48.1, behavior therapy in Section 48.3, and prevention in Section 49.15.

CHAPTER REFERENCES

*Borduin C, Mann B, Cone L, Henggeler S, Fucci B, Blaske D, Williams R: Multisystemic treatment of serious juvenile offenders: Long-term prevention of criminality and violence. J Cons Clin Psychol *63:*569, 1995.

Campbell M: The pharmacological treatment of conduct disorders and rage outbursts. Psychiatr Clin North Am *15:*69, 1992.

Campbell M, Adams PB, Small AM, Kafantaris V: Lithium in hospitalized aggressive children with conduct disorder: A double-blind and placebo-controlled study. J Am Acad Child Adolesc Psychiatry *34:*445, 1995.

Cauffman E, Feldman SS, Waterman J, Steiner H: Posttraumatic stress disorder among female juvenile offenders. J Am Acad Child Adolesc Psychiatry *37:*1209, 1998.

Dodge K: Social-cognitive mechanisms in the development of conduct disorder and depression. Annu Rev Psychol *44:*559, 1993.

Earls F: Oppositional-defiant and conduct disorders. In *Child and Adolescent Psychiatry: Modern Approaches,* M Rutter, E Taylor, L Hersov, editors. Blackwell Scientific Publications, London, 1994.

Frick P, Lahey B, Loeber R, Tannenbaum L: Oppositional defiant disorder and conduct disorder: A meta-analytic review of factor analyses and cross-validation in a clinic sample. Clin Psychol Rev *13:*319, 1993.

*Guzder J, Paris J, Zelkowitz P, Feldman R: Psychological risk factors for borderline pathology in school-age children. J Am Acad Child Adolesc Psychiatry *38:*206, 1999.

Hechtman L, Weiss G: Controlled prospective fifteen-year follow-up of hyperactives as young adults: Non-medical drug and alcohol use and anti-social behavior. Can J Psychiatry *31:*557, 1996.

Hinshaw SP, Anderson CA: Conduct and oppositional defiant disorders. In *Child Psychopathology,* EJ Mash, RA Barkley, editors. Guilford, New York, 1996.

Kazdin A: Conduct disorders in childhood and adolescence. Sage, Thousand Oaks, CA, 1995.

Kelley BT, Thornberry TP, Smith C: In the wake of childhood maltreatment. Office of Juvenile Justice and Delinquency Prevention Juvenile Justice Bulletin, August, 1997.

Kerr M, Tremblay R, Pagini L, Vitaro F: Boy's behavioral inhibition and the risk of later delinquency. Arch Gen Psychiatry *54:*809, 1997.

Klein RG, Abikoff H, Klass E, Ganeles D, Seese LM, Pollack S: Clinical efficacy of methylphenidate in conduct disorder with and without attention deficit hyperactivity disorder. Arch Gen Psychiatry *54:*1073, 1997.

Kruesi M, Lelio D: Disorders of conduct and behavior. Child Adolesc Psychiatr Clin North Am *25:*2, 1995.

Lahey B, Hart E, Pliszka S, Applegate B, McBurnett K: Neurophysiological correlates of conduct disorder: A rationale and a review of research. J Clin Child Psychol *22:*141, 1993.

*Lewis D, Yeager C, Lovely R, Stein A, Cobham-Portorreal C: A clinical follow-up of delinquent males: Ignored vulnerabilities, unmet needs, and the perpetuation of violence. J Am Acad Child Adolesc Psychiatry *33:*518, 1994.

Loeber R, Wung P, Keenan K, Giroux B, Stouthamer-Loeber M, Van Kammen W, Maughan B: Developmental pathways in disruptive behavior. Dev Psychopathol *5:*101, 1993.

Loeber R, Keenan K: Interaction between conduct disorder and its comorbid conditions: Effects of age and gender. Clin Psychol Rev *14:*497, 1994.

Loeber R, Green S, Keenan K, Lahey B: Which boys will fare worse? Early predictors of onset of conduct disorder in a six-year longitudinal study. J Am Acad Child Adol Psychiatry *34:*499, 1995.

Moffit TE: Life-course persistent and adolescence-limited antisocial behavior: A developmental taxonomy. Psychol Rev *100:*674, 1993.

Offord D, Bennett K: Conduct disorder: Long-term outcomes and intervention effectiveness. J Am Acad Child Adolesc Psychiatry *33:*1069, 1994.

Patterson G, Narrett C: The development of a reliable and valid treatment program for aggressive young children. Int J Ment Health *19:*19, 1990.

Quinton D, Pickles A, Maughan B, Rutter M: Partners, peers, and pathways: Assortative pairing and continuities in conduct disorder. Special Issue: Milestones in the development of resilience. Dev Psychopathol *5(4)*:763, 1993.

Raine A, Venables P, Williams M: High autonomic arousal and electrodermal orienting at age 15 years as protective factors against criminal behavior at age 29 years. Am J Psychiatry *152:*1595, 1995.

Rey J: Oppositional defiant disorder. Am J Psychiatry *150(12)*:1769, 1993.

Richters J: Mark Twain meets DSM-III-R: Conduct disorder, development, and the concept of harmful dysfunction. Dev Psychopathol *5:*5, 1993.

Robins L, Rutter M, editors: *Straight and Devious Pathways from Childhood to Adulthood.* Cambridge University Press, New York, 1990.

Rutter M: Introduction. In *Genetics of Criminal and Antisocial Behavior,* G Bock, J Good, editors. Wiley, Chichester, 1996.

Sholevar G: *Conduct Disorders in Children and Adolescents.* American Psychiatric Press, Washington, DC, 1995.

*Steiner H: Practice parameters for the assessment and treatment of conduct disorders. J Am Acad Child Adolesc Psychiatry *37*(Suppl):122S, 1997.

Steiner H: *Treating Adolescents.* Jossey-Bass, San Francisco, 1996.

Steiner H, Matthews Z, Tanzen G, Duxbury E: Randomized clinical trial of depakote in delinquents. *Scientific Proceedings of the Annual Meeting of the American Academy of Child and Adolescent Psychiatry* October, Toronto, 1997.

*Steiner H, Cauffman E, Duxbury E: Personality traits in juvenile delinquents: Relation to criminal behavior and recidivism. Am Acad Child Adolesc Psychiatry *38:*256, 1999.

*Steiner H, Stone L: Violence and related psychopathology. J Am Acad Child Adolesc Psychiatry *38:*232, 1999.

Vitiello B, Stoff DM: Subtypes of aggression and their relevance to child psychiatry. J Am Acad Child Adol Psychiatry *36:*3, 307, 1997.

Waldman I, Lilienfeld S, Lahey B: Toward construct validity in the childhood disruptive behavior disorders: Classification and diagnosis in DSM-IV and beyond. Adv Clin Child Psychol *17:*323, 1995.

Weinberger D: The construct validity of the repressive coping style. In *Repression and Dissociation,* Singer JL, editor. University of Chicago Press, Chicago, 1990.

Zoccolillo M, Rogers K: Characteristics and outcomes of hospitalized adolescent girls with conduct disorder. J Am Acad Child Adolesc Psychiatry *30:*973, 1991.

CHAPTER 41. FEEDING AND EATING DISORDERS OF INFANCY AND EARLY CHILDHOOD

IRENE CHATOOR, M.D.

Because infants and young children cannot feed themselves independently, disturbances related to eating are frequently referred to as “feeding disorders” to emphasize the dyadic nature of the eating process in this young age group. The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) lists three different diagnoses under feeding and eating disorders of infancy or early childhood: pica, rumination disorder, and feeding disorder of infancy or early childhood.

In the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), feeding disorder of infancy and childhood (which also includes rumination disorder) and pica are included under the category of other emotional and behavioral disorders with onset usually occurring in childhood and adolescence.

PICA

In DSM-IV, pica is described as the persistent eating of nonnutritive substances for at least 1 month. The behavior must be developmentally inappropriate, not culturally sanctioned, and sufficiently severe to merit clinical attention. Pica is diagnosed even when these symptoms occur in the context of another disorder such as autistic disorder, schizophrenia, or Klein-Levin syndrome.

Pica appears much more frequently in young children than in adults; it also occurs more frequently in persons who are mentally retarded. Among adults, certain forms of pica, including geophagia (clay eating) and amylophagia (starch eating), have been reported to occur in pregnant women. In certain regions of the world and among certain cultures, such as the Australian aborigines, rates of pica in pregnant women have been reported to be high. According to DSM-IV, however, if such practices are culturally accepted, the diagnostic criteria for pica are not met.

History The term “pica” was first applied by the French surgeon, Ambroìse Páre. It was derived from the Latin word for magpie, a bird considered to have unusual appetites. Many attributions for pica that have been described over the past 150 years were summarized by Myra Cooper in the 1950s.

Epidemiology Since young infants mouth objects quite commonly, it is difficult to diagnose pica when the child is under the age of 2 years. A survey of a large clinic population with a wide range of ethnic backgrounds revealed that 75 percent of 12-month-old infants and 15 percent of 2- to 3-year-old toddlers were reported by their parents to put nonnutritive substances in their mouths. Among individuals with mental retardation, the prevalence of pica appears to increase with the severity of the retardation. The prevalence of pica among institutionalized mentally retarded individuals has been estimated to range from 10 to 33 percent.

Etiology Organic, psychodynamic, socioeconomic, and cultural factors have been implicated in the etiology of pica. Inadequate intake of iron and calcium induces pica in rats and has been considered to lead to abnormal cravings and pica in humans. Psychosocial factors such as poverty, maternal neglect and abuse, and disorganized family situations with inadequate nurturing and supervision of the children have been implicated in the etiology of pica. The cultural acceptance of pica in certain population groups is also considered to play a role in pica. Some authors have proposed a multifactorial model in which organic, familial, socioeconomic, and cultural factors interact.

Diagnosis and Clinical Features Young children with pica typically eat plaster, paper, paint, cloth, hair, insects, animal droppings, sand, pebbles, and dirt. Many of the children engage in other oral activities (e.g., thumb sucking or nail biting), which they seem to use for relief of tension and self-soothing. Pica may lead to anemia, diarrhea/constipation, worm infestation, toxoplasmosis, lead poisoning, and malnutrition. Intestinal obstruction may develop as the result of hair ball tumors. The DSM-IV diagnostic criteria are listed in Table 41-1, and the ICD-10 criteria are found in Table 41-2.

A. Persistent eating of nonnutritive substances for a period of at least 1 month.
B. The eating of nonnutritive substances is inappropriate to developmental level.
C. The eating behavior is not part of a culturally sanctioned practice.
D. If the eating behavior occurs exclusively during the course of another mental disorder (e.g., mental retardation, pervasive developmental disorder, schizophrenia), it is sufficiently severe to warrant independent clinical attention.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 41-1 DSM-IV Diagnostic Criteria for Pica

A. There is persistent or recurrent eating of nonnutritive substances, at least twice a week.
B. Duration of the disorder is at least 1 month. (For some purposes, researchers may prefer a minimum period of 3 months.)
C. The child exhibits no other mental or behavioral disorder in the ICD-10 classification (other than mental retardation).
D. The child's chronological and mental age is at least 2 years.
E. The eating behavior is not part of a culturally sanctioned practice.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 41-2 ICD-10 Diagnostic Criteria for Pica of Infancy and Childhood

Differential Diagnosis Before the age of 18 to 24 months, mouthing and sometimes eating nonnutritive substances is relatively common and should only be diagnosed as pica if it is persistent and inappropriate for the developmental level of the child.

Course and Prognosis Usually, pica lasts for several months and then remits. It may occasionally continue into adolescence or, less frequently, into adulthood. Several authors have pointed to the serious developmental impact of the disorder. The younger children showed delays in their speech and psychosocial development. Half of the adolescents suffered from depression, personality disorders, or both; engaged in other forms of disturbed oral activities (e.g., thumb sucking and nail biting), and abused tobacco, alcohol, or drugs. More recently, some authors found a strong relation of pica during early childhood and bulimia nervosa during

adolescence.

Treatment A variety of treatments have been proposed, ranging from physical restraints, aversive and nonaversive behavioral therapy, time out and overcorrection, to environmental enrichment, and individual and group therapy for the children and their mothers. One must conduct a comprehensive evaluation of the child, the caregivers, and their environment to develop an individualized treatment plan. This should include helping the adults become aware of the short- and long-term dangers of pica, providing a lead-free and childproof environment, facilitating a more-satisfying parent-child relationship, and applying behavioral techniques to extinguish the maladaptive behavior of pica.

RUMINATION DISORDER

In DSM-IV rumination disorder is described as an infant's or a child's repeated regurgitation and rechewing of food, after a period of normal functioning. The symptoms last for at least 1 month, are not caused by a medical condition, and are severe enough to merit clinical attention. The onset of the disorder generally occurs after 3 months of age. After regurgitation, the food may be swallowed or spit out. Infants who ruminate are observed to strain to bring the food back into their mouths and appear to find the experience pleasurable. The infants are often brought for evaluation because of failure to thrive. The disorder is rare in older children, adolescents, and adults. It varies in its severity, and it is sometimes associated with medical conditions, such as hiatal hernia, that result in esophageal reflux. In its most severe form, the disorder can be fatal.

Rumination has been recognized for hundreds of years. An awareness of the disorder is important for correct diagnosis to avoid unnecessary surgical procedures and inappropriate treatment.

Rumination is derived from the Latin word *ruminare*, meaning “to chew the cud.” The Greek equivalent is *merycism*, the act of regurgitating food from the stomach into the mouth, rechewing the food, and reswallowing it.

History As early as in the seventeenth century, rumination was described in conjunction with hiatus hernia by the Italian anatomist Fabricius ab Aquapendente. Paul Maas first described rumination disorder of infancy in 1907, and in the 1950s Reginald Lourie linked rumination to disturbances in the mother-infant relationship.

Epidemiology Rumination disorder is believed to be rare. It appears to occur more often in males than in females and also in individuals with mental retardation. A few reports describe adults with eating disorders who developed rumination.

Etiology Various causes have been proposed. Several authors have attributed rumination to separation or a disturbed mother-infant relationship (e.g., lack of responsiveness or neglect by the mother because of stressful family relations or life events). Others consider rumination a symptom of gastroesophageal reflux. Several authors have postulated that rumination is a learned behavior that is maintained by special attention to the regurgitation. More recently opiate receptor insensitivity or reduced endorphinergic transmission has been implicated in rumination.

The author has proposed a biopsychosocial model in which rumination is seen along a continuum in which an infant may have gastrointestinal pathology and little psychopathology at one end of the spectrum and no organic pathology and severe psychopathology in the mother-infant relationship at the other. Frequently, vomiting secondary to gastroesophageal reflux or vomiting associated with an acute illness precedes the beginning of rumination. Apparently at some point the infant learns to initiate vomiting and rechew the food to relieve tension, self-soothe, or self-stimulate. When the infant fails to elicit or loses caring attention or tension-relieving responses from the mother, rumination seems to become a means of self-regulation. Once the infant has experienced rumination as a means of self-soothing or self-stimulation, the rumination develops into a habit that is difficult to break.

Diagnosis and Clinical Features Some infants put their thumb or whole hand in the mouth, suck their tongue rhythmically, or arch their back to initiate rumination. Others ruminate after they have been placed in the crib to sleep and are frequently found in a puddle of vomitus, which is then attributed to an organic problem. The latter type of rumination is most commonly observed in infants who have received little emotional stimulation and have learned to stimulate and soothe themselves through rumination. Although initially most of the regurgitated food is vomited, some infants gradually learn to hold more of the food in their mouth and rechew and reswallow it. Experienced ruminators can bring up the food through tongue movements and reswallow it without losing any. Their rumination can only be discovered by observing the movements of their cheeks and the foul oral odor.

The DSM-IV diagnostic criteria for rumination disorder are presented in Table 41-3.

A. Repeated regurgitation and rechewing of food for a period of at least 1 month following a period of normal functioning.
B. The behavior is not due to an associated gastrointestinal or other general medical condition (e.g., esophageal reflux).
C. The behavior does not occur exclusively during the course of anorexia nervosa or bulimia nervosa. If the symptoms occur exclusively during the course of mental retardation or a pervasive developmental disorder, they are sufficiently severe to warrant independent clinical attention.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 41-3 DSM-IV Diagnostic Criteria for Rumination Disorder

Differential Diagnosis Rumination must be differentiated from vomiting secondary to organic conditions. Since reflux and rumination frequently coexist, one must observe the infant in different situations and actually see the initiation and rhythmic movements of the ruminatory activity.

Course and Prognosis The onset of rumination is frequently in infancy, most commonly in the first year of life. However, rumination may start at any point in life and has been reported to start later in childhood in individuals with developmental delays and in adulthood in individuals with eating disorders. In some individuals, the disorder remits spontaneously. However, some infants suffer serious complications such as electrolyte imbalance, dehydration, and weight loss. Earlier reports described mortality rates as high as 25 percent.

Treatment Different treatments have been suggested on the basis of different theories about the cause of rumination. Treatments range from the mechanical restraints described in the early literature and surgical interventions to prevent reflux to behavioral and psychodynamic treatments.

Based on the assumption that rumination is a learned habit that is reinforced by the increased attention given to the regurgitation, unlearning by behavioral techniques has been suggested. These treatments range from the use of aversive taste stimuli (lemon juice or hot sauce), oral hygiene, differential reinforcement of incompatible behaviors, to electric shock. Those who assume that rumination is based on an unsatisfactory mother-infant relationship have suggested a psychodynamic approach. Having a mother substitute feed the infant while the mother is helped through psychotherapy has been proposed.

After a thorough assessment of both infant and mother, the therapist needs to individualize the treatment. In severe situations, if the infant is malnourished and continues to lose most of the food through rumination, a jejunal tube may have to be inserted before any psychosocial treatment can begin. Frequently, a combination of psychodynamic and behavioral interventions to enhance the mother-infant relationship in general and to treat the symptom of rumination, in particular, is most effective.

FEEDING DISORDER OF INFANCY OR EARLY CHILDHOOD

According to DSM-IV, feeding disorder of infancy or early childhood is a persistent failure to eat adequately, reflected in significant failure to gain weight or in significant weight loss over at least 1 month. The symptoms are not better accounted for by a medical condition or by another mental disorder and are not caused by lack of food. The disorder has its onset before the age of 6 years.

The DSM-IV diagnostic criteria (Table 41-4) are broad and do not address the specificity of various feeding disorders not included in DSM-IV or ICD-10. Feeding disorder of infancy or early childhood serves as an umbrella diagnosis for the more specifically defined feeding disorders described in this chapter.

A. Feeding disturbance as manifested by persistent failure to eat adequately with significant failure to gain weight or significant loss of weight over at least 1 month.
B. The disturbance is not due to an associated gastrointestinal or other general medical condition (e.g., esophageal reflux).
C. The disturbance is not better accounted for by another mental disorder (e.g., rumination disorder) or by lack of available food.
D. The onset is before age 6 years.

Table 41-4 DSM-IV Diagnostic Criteria for Feeding Disorder of Infancy or Early Childhood

The ICD-10 diagnostic criteria for the disorder are listed in Table 41-5. Unlike DSM-IV, ICD-10's feeding disorder of infancy or early childhood subsumes rumination disorder. However, similar to DSM-IV, the criteria are broad and do not delineate additional specific feeding disorders.

A. There is persistent failure to eat adequately, or persistent rumination or regurgitation of food.
B. The child fails to gain weight, loses weight, or exhibits some other significant health problem over a period of at least 1 month. (In view of the frequency of transient eating difficulties, researchers may prefer a minimum duration of 3 months for some purposes.)
C. Onset of the disorder is before the age of 6 years.
D. The child exhibits no other mental or behavioral disorder in the ICD-10 classification (other than mental retardation).
E. There is no evidence of organic disease sufficient to account for the failure to eat.

Table 41-5 ICD-10 Diagnostic Criteria for Feeding Disorder of Infancy and Childhood

According to ICD-10, feeding disorder of infancy and childhood is “a feeding disorder of varying manifestations, usually specific to infancy and early childhood. It generally involves refusal of food and extreme faddiness in the presence of an adequate food supply and a competent caregiver, and the absence of organic disease. There may or may not be associated rumination (repeated regurgitation without nausea or gastrointestinal illness).”

History and Comparative Nosology Prior to the inclusion of the diagnostic category of feeding disorder of infancy or early childhood in DSM-IV, no nationally defined criteria existed for the diagnosis of feeding disorders. In the absence of a recognized standard, feeding problems were given various labels by different authors: food refusal, food phobia, food aversion, feeding resistance, picky eaters, problem eaters. The field of feeding disorders is further complicated by the use of *failure to thrive* as a diagnostic label primarily in the pediatric literature. *Failure to thrive* refers to inadequate weight gain on the basis of standard growth charts. Frequently, *failure to thrive* and *feeding disorder* are used interchangeably, although not all infants and children who fail to thrive have feeding problems and not all infants and children with feeding problems fail to thrive.

Because of the multiple causes of feeding problems, which may or may not be associated with growth deficiency, the author has developed clinical operational criteria for three developmental transactional feeding disorders associated with growth deficiency: (1) feeding disorder of homeostasis, (2) feeding disorder of attachment (described in the third edition of DSM [DSM-III] as reactive attachment disorder), and (3) feeding disorder of separation, later defined as infantile anorexia. Posttraumatic feeding disorder usually follows a traumatic experience related to feeding (e.g., choking or severe gagging) and may or may not be associated with growth deficiency.

Epidemiology It is estimated that between 15 and 35 percent of infants and young children have feeding problems. The more common feeding difficulties include eating too little, refusing certain types of food, objectionable mealtime behaviors, and bizarre food habits. Severe feeding problems associated with poor weight gain, such as refusal to eat or vomiting, have been reported in 1 to 2 percent of infants and toddlers. A few studies found that about 70 percent of infants who exhibit food refusal in the first year of life continue to have feeding problems when followed up to school age. Picky eating and gastrointestinal symptoms in early childhood have been linked to anorexia nervosa, and pica and problem behaviors during mealtime have been associated with bulimia nervosa during the adolescent years.

Other Feeding Disorders

Feeding Disorder of Homeostasis

DIAGNOSIS AND CLINICAL FEATURES Feeding disorder of homeostasis begins in the postnatal period and is characterized by irregular, poor feedings and inadequate food intake by the infant. Both infant and parent contribute to the feeding difficulties. The infant demonstrates poor regulation of state during feeding, expressed by irritability, easy fatiguability, and excessive sleepiness. The parent may be anxious or depressed or may exhibit more severe psychopathology. Feedings are characterized by infant irritability, sleepiness, or both; maternal tension; and poor engagement between mother and infant.

Feeding disorder of homeostasis can be seen in infants without any organic problems, but it is frequently associated with prematurity or dysmaturity, functional or structural abnormalities of the oropharynx or gastrointestinal tract, and cardiac or pulmonary disease.

DIFFERENTIAL DIAGNOSIS This feeding disorder needs to be differentiated from pure organic disorders which can interfere with the regulation of state of the infant. As described above, organic problems can trigger the feeding disorder and should always be addressed first before making the diagnosis of feeding disorder of homeostasis.

COURSE AND PROGNOSIS Mother-infant interactional patterns and the foundation for the regulation of feedings are established in the first months of life. Infants with feeding problems during the early months trigger anxiety in their mothers and frequently have difficulty self-regulating their food intake during the transition to self-feeding in the second year of life.

TREATMENT Treatment must be individualized and can be directed toward the infant, the mother, and the mother-infant interaction. If the infant tires easily and cannot take in adequate calories for growth, nasogastric tube feedings might be needed to supplement oral feedings. On the other hand, if the mother is exhausted,

overly anxious, or depressed, the mother's difficulties must be addressed to enable her to be more effective with her infant.

Videotaping the feedings and observing the tape with the mother frequently heightens the mother's awareness of the infants' cues and helps her consider alternative ways to facilitate the infant's regulation of state and ability to feed.

Feeding Disorder of Attachment

DIAGNOSIS AND CLINICAL FEATURES Feeding disorder of attachment is characterized by a lack of engagement between mother and infant that leads to inadequate food intake and growth failure. The onset of growth failure is usually between 2 to 8 months of age. The infant shows lack of age-appropriate social responsibility (e.g., lack of smiling response, lack of vocal reciprocity, lack of anticipatory reaching out when picked up, lack of molding and cuddling when held). The mother frequently presents with acute or chronic depression, drug or alcohol abuse, or high psychosocial stress that interferes with her care of the infant. Mother-infant interactions are characterized by a lack of mutual engagement and lack of pleasure in their relationship. The mother frequently appears detached and noncontingent in her responses to the infant's cues.

This feeding disorder has been referred to in the literature as “maternal deprivation” or “deprivation syndrome.” It has been postulated that lack of emotional nurture and infrequent feedings lead to the growth failure and developmental delays of these infants. Characteristically, infants with feeding disorder of attachment become engaging and gain weight when admitted to the hospital and put in the care of a nurturing nurse.

In DSM-III, this disorder was defined as “reactive attachment disorder of infancy” associated with failure to thrive; in DSM-IV this diagnosis was changed to encompass only problems in relatedness without growth failure. Several studies of infants with the diagnosis of “nonorganic failure to thrive” have demonstrated very high rates of insecure attachment, ranging from 50 to 90 percent of the infants studied. However, not all infants with insecure attachment fail to thrive. Apparently only a severe disturbance in the attachment process of the infant to the caregiver, what some authors describe as “nonattachment,” results in the severe disturbances in relatedness and growth failure seen in infants with this feeding disorder.

DIFFERENTIAL DIAGNOSIS Feeding disorder of attachment must be differentiated from organic conditions that lead to a lack of weight gain and weakness of the infant. However, in organic conditions mother and infant usually show better mutual engagement and the infant responds more readily to the examiner.

COURSE AND PROGNOSIS Children with a feeding disorder of attachment are at high risk for delayed cognitive and disturbed emotional development. When followed up to school age, deficits in behavioral organization, ego control, ego resiliency, and behavioral symptoms have been reported in these children.

TREATMENT Depending on the severity of the growth deficiency and the evidence of neglect, various interventions have been suggested. When no evidence exists of deprivational behavior by the mother and the parents have sought medical care in the past and have some support in the extended family or the community, an outpatient approach may be safe. However, if the growth deficiency is more severe, if there is serious hygiene neglect, if the mother is abusing drugs or alcohol, or if the mother has a chaotic lifestyle, immediate hospitalization is indicated.

During the hospitalization, a warm and nurturing nurse must be assigned to engage the infant during feedings. While the nutritional, emotional, and developmental rehabilitation of the infant takes place, a more in-depth evaluation of the mother and the mother-infant relationship must occur. Many of these mothers experienced unsatisfactory relationships with their own caregivers when they were growing up, and consequently they distrust and avoid professionals. It is helpful to identify and point out any positive behavior the infant shows toward the mother in order to see whether the mother has any potential to respond and to engage in a mutually satisfying relationship with her infant. In addition, the mother's support system and her ability to become engaged with a therapist must be explored before the infant is returned to her care. In some situations of severe neglect and associated abuse, Protective Services must be involved, and the infant must be placed in foster care.

Before discharge from the hospital all services must be in place to ensure adherence to the treatment plan. The treatment plan should be individualized and may include home visits by a health care professional, day care for the infant, mother-infant psychotherapy, and family therapy for the parents.

Infantile Anorexxia (Feeding Disorder of Separation)

DIAGNOSIS AND CLINICAL FEATURES Infantile anorexia is characterized by food refusal by the infant that leads to inadequate food intake and acute or chronic malnutrition. The food refusal causes high anxiety in the parents and leads to intense conflict in the parent-infant relationship over issues of autonomy, dependency, and control, with food being the battleground.

The food refusal by the infant usually begins or worsens considerably at the age of 6 months to 2 years, during the transition to spoon- and self-feeding. The food refusal may vary from meal to meal and with different caregivers. The infant may drink from the bottle or from the breast and primarily refuse solid food. This leads the parents to believe that the infant has a poor appetite and loses interest in food quickly. On the other hand, the infant is seen as very curious and demanding a lot of attention. The food refusal causes intense anxiety in the mother or both parents. Usually the parents resort to coaxing, cajoling, bribing, distracting, and offering different foods at all times, and when desperate, they may resort to force feeding.

Infantile anorexia interferes with the development of somatopsychological differentiation, the ability to differentiate physiological sensations of hunger and satiety from emotional feelings and needs (e.g., feelings of anger and the need for affection). The mothers are usually so anxious that they offer food, the bottle, or breast at any time, regardless of whether the infant is hungry or distressed. Frequently, the infants refuse to eat to control the parent, if they cannot have their way, if they cannot get out of the high chair, or if they are not offered the food they want. Eating or not eating becomes increasingly controlled externally by the interactions between the infant and the caregiver instead of internally by feelings of hunger and fullness.

DIFFERENTIAL DIAGNOSIS Infantile anorexia must be differentiated from another feeding disorder characterized by food refusal. Posttraumatic feeding disorder usually has a more sudden onset of food refusal following a traumatic event to the oropharynx or gastroesophageal tract (e.g., choking, severe gagging or vomiting, or the insertion of feeding or endotracheal tubes). The infant may refuse all food, only solid food, or only the bottle, depending on what mode of feeding the infant associates with the traumatic experience. Picky eaters may show the same behavior patterns as infantile anorectics, but their food intake is adequate for age-appropriate growth. In addition, infantile anorexia must be differentiated from food refusal caused by organic problems (e.g., gastroesophageal reflux, zinc deficiency, and metabolic disorders).

COURSE AND PROGNOSIS No systematic data from prospective studies are available. Studies from Sweden in which infants with food refusal in the first year of life were followed up to school age, showed that 70 percent of the children continued to have feeding problems at 4 years of age, and by 6 years of age the eating problems were noted not only in the home but in the school as well.

A follow-up study by the author of 20 children, 6 months to 2 years after they had received treatment, found that 17 children had changed considerably. The conflict in the parent-child relationship over the child's food intake was completely gone. The children had learned to recognize hunger and had increased their food intake, which had led to weight gain and improved growth. However, the parents reported that many of these children continued to lose their appetite when they were excited, had house guests, had a birthday party, or traveled. Once they calmed themselves, they would regain their appetite and usually make up for the lost meals. Three children continued to struggle with their parents over their lack of eating; in two cases, day care seemed to be too stimulating to allow these children to relax enough to eat, and in the third case, the home situation was very overcrowded and disorganized.

TREATMENT The author has developed a treatment model that is based on a transactional model for development of infantile anorexia. According to this model, the infant's emotional intensity, lower threshold for stimulation, and stubbornness evoke conflicts over issues of control, autonomy, and dependency in a vulnerable parent who is insecure about limit setting. The treatment addresses the infant's temperament and the parent's anxiety about limit setting to facilitate internal regulation of eating in the infant.

1. The infant's temperament is discussed with the parents to help them understand that the infant's curiosity and emotional intensity interfere with the awareness of hunger. Because these infants are strong willed and have learned that food refusal is a powerful tool to control the parents' attention, parental efforts to distract or entertain the infants to get them to eat only perpetuate the problem by getting the infants more excited.
2. The source of the parents' anxiety about limit setting is explored. Some parents have been sensitized by lost pregnancies, infertility problems, or the harshness or emotional unavailability of their own parents. Understanding the parents' sensitivity about limit setting is necessary to help them set limits more effectively with

their infant.

3. After this groundwork has been laid, the parents are trained to change their behavior to facilitate internal regulation of eating by the infants. To boost the infants biological rhythms of hunger and satiety, parents are encouraged to feed them only at regular 3- to 4-hour intervals and to offer nothing but water between these regular feedings. They are helped to understand the importance of encouraging the infants' self-feeding by commenting on their feeding skills (“you get the spoon in your mouth all by yourself”), but they are not to comment on how much or how little the infants eat. Infants must learn to regulate their food intake by their internal signals of hunger and satiety. Eating should not be a performance for the parents. There should be no distractions during feedings, no television or games, so that the infants focus on eating. If infants engage in inappropriate behavior during feeding (climbing out of the highchair, throwing food or feeding utensils) the parents need to learn to use the time-out procedure.

This three-step intervention is best accomplished in three double sessions, lasting 2 hours each, and grouped close together within a few weeks. The intensity of this brief intervention facilitates a close therapeutic alliance between the therapist and the parents and enables many parents to make major shifts in their interactions with their infants. Many of these parents need only a few follow-up sessions to deal with some specific questions that may arise. Frequently, after infants change their eating patterns, parents ask for help concerning the sleeping problems many of these children have as well. However, there are some families who are more deeply entrenched in maladaptive interactional patterns with their infants, and some infants are so severely malnourished that more intensive interventions or hospitalization may be required.

Posttraumatic Feeding Disorder

DIAGNOSIS AND CLINICAL FEATURES Posttraumatic feeding disorder may present with total food refusal, refusal to swallow solid food or refusal to drink from the bottle depending on what kind of feeding the infant associates with the traumatic event. The onset of food refusal is frequently rather sudden and follows a traumatic experience that involves the oropharynx or gastrointestinal tract (e.g., severe gagging, choking, vomiting, insertion of feeding or endotracheal tubes, or force-feeding). Frequently, the parents are not aware that the event was so frightening to the infant that it triggered the food refusal because other infants who undergo the same experience do not necessarily develop the disorder. Infants who develop a posttraumatic feeding disorder appear to be more prone to anxiety or are more sensitive to pain than the average infant. Older children who refused to eat solid food after an incident of gagging or choking reported being afraid that food would get stuck in their throat and choke them to death. Infants and young children express this fear by crying in anticipation of being fed when seeing the highchair, bottle, or spoon. Some infants or children are brave enough to put the food in the mouth but cannot bring themselves to swallow it. They keep the food in their cheeks or spit it out. Some infants and children may become more irritable and cling to their parents, others experience frightening dreams or even nightmares.

The parents frequently react with anxiety to the infant's food refusal. They worry about the infant's nutrition. Frequently, they try to coax and distract the infant, they offer various types of food, and they try to feed the infant day and night. Some infants who are afraid to drink from the bottle may drink when they are asleep and not be aware of what they are doing. However, if they wake up and see the bottle, they push it away and cry.

DIFFERENTIAL DIAGNOSIS Posttraumatic feeding disorder must be differentiated from infantile anorexia and from food refusal because of sensory aversions to certain types of food. Infantile anorexia is characterized by a more inconsistent pattern of food refusal, depending on the mood of the infant. Anorectic infants are not afraid to eat and can eat all types of food if they want to eat. Food aversions involve certain types of food with a certain taste, texture, or smell. Usually the food refusal is more selective and not as global as that in a posttraumatic feeding disorder.

COURSE AND PROGNOSIS No systematic longitudinal data on the course of posttraumatic feeding disorder are available. However, individual case studies indicate that many infants and children get locked into their fears of eating. Some drink milk from the bottle and eat only pureed food until they reach school age, when the social embarrassment of their eating behavior causes the parents to seek help. In extreme cases, when infants refuse all food, gastrostomy feedings have to be implemented. Some of these children depend on gastrotomy feedings to survive for years.

TREATMENT Because of the complexity of many cases, particularly those that result from trauma inflicted by unavoidable medical procedures (e.g., intubation, suctioning), a multidisciplinary team consisting of a pediatrician or gastroenterologist, nutritionist, occupational therapist, and a psychiatrist is best equipped to meet the needs of these infants.

Total food refusal puts the infant in acute medical danger of dehydration, and intravenous fluids must be given. It is important to determine whether the infant can be coaxed into accepting any form of food or whether nasogastric tube feedings need to be instituted. Unfortunately, the insertion of feeding tubes can intensify the fear of feeding, and if this is the case, the infant is better served by a gastrostomy tube. In any case, the infant's nutritional requirements must be addressed prior to the implementation of the psychiatric treatment.

The psychiatric treatment of infants who exhibit total or partial food refusal centers on desensitizing them to the fear of eating. Behavioral techniques of positive reinforcement of food acceptance and negative reinforcement of food refusal are instituted in the initial management of these infants. Encouraging self-feeding frequently helps these infants gain mastery over the anticipatory anxiety of eating. When the infant begins to accept food from the spoon, it is important to proceed slowly in advancing the texture of food, to avoid any incident of gagging that might set the infant back.

The behavioral manipulation of the infant's eating frequently leads to external regulation of eating in response to the reinforcers. Once the infant has overcome the fear of eating, the external reinforcers must be phased out and eating according to internal signals of hunger and satiety encouraged. The principles described above in facilitating internal regulation of eating in infants with anorexia should also be applied in this final stage of the treatment.

SUGGESTED CROSS-REFERENCES

The general topic of eating disorders is covered in Chapter 20. Psychological aspects of gastrointestinal disorders are discussed in Section 25.2. Regarding relevant treatment modalities see Section 3.2 on behavior therapy and Section 31.31 on electroconvulsive therapy. Normal child development is reviewed in Section 32.2. Psychiatric examination of infants and children is discussed in Chapter 33. Various other psychiatric disorders of childhood are the topics of Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39 and Chapter 40 and Chapter 42, Chapter 43, Chapter 44, Chapter 45, Chapter 46, Chapter 47 and Chapter 48.

CHAPTER REFERENCES

Beckner J, Turner S, Sajwaj T: Multiple behavioral effects of the use of lemon juice with a ruminating toddler. Behav Modif *2:*267, 1978.

Benoit D, Coolbear J: Posttraumatic feeding disorders in infancy: Behaviors predicting treatment outcome. Infant Ment Health J *19:*409, 1998.

Benoit D, Green D, Arts-Rodas D: Posttraumatic feeding disorders. J Am Acad Child Adolesc Psychiatry *36:*577, 1997.

Benoit D, Zeanah CH, Barton ML: Maternal attachment disturbances in failure to thrive. Infant Ment Health J *10:*185, 1989.

Bithoney W, Dubowitz H, Egan H: Failure to thrive/growth deficiency. Pediatr Rev *13:*453, 1992.

Black M, Hutcheson J, Dubowitz H, Berenson-Howard J: Parenting style and developmental status among children with non-organic failure to thrive. J Pediatr Psychol *19:*689, 1994.

*Chatoor I: Infantile anorexia nervosa: A developmental disorder of separation and individuation. J Am Acad Psychoanal *17:*43, 1989.

*Chatoor I, Conley C, Dickson L: Food-refusal after an incident of choking. J Am Acad Child Adolesc Psychiatry *27:*105, 1988.

Chatoor I, Dickson L, Einhorn A: Rumination: Etiology and treatment. Pediatr Ann *13:*924, 1984.

Chatoor I, Egan J, Getson P, Menveille E, O'Donnell R: Mother-infant-interactions in infantile anorexia. J Am Acad Child Adolesc Psychiatry *27:*535, 1988.

Chatoor I, Ganiban J, Colin V, Plummer N, Harmon R: Attachment and feeding problems: A reexamination of non-organic failure to thrive and attachment insecurity. J Am Acad Child Adolesc Psychiatry *37:*1217, 1998.

*Chatoor I, Getson P, Menveille E, Brasseaux C, O'Donnell R, Rivera Y, Mrazek D: A feeding scale for research and clinical practice to assess mother-infant interactions in the first three years of life.

Infant Ment Health J *18:*76, 1997.

Chatoor I, Hirsch R, Ganiban J, Persinger M, Hamburger E: Diagnosing infantile anorexia: The observation of mother-infant interactions. J Am Acad Child Adolesc Psychiatry *37:*959, 1998.

Chatoor I, Hirsch R, Persinger M: Facilitating internal regulation of eating: A treatment model for infantile anerexia. Infants Young Child *9:*12, 1997.

Chatoor I, Kerzner B, Zorc I, Persinger M, Simenson R, Mrazek D: Two-year-old twins refuse to eat: A multidisciplinary approach to diagnosis and treatment. Infant Ment Health J *13:*252, 1992.

Dahl M, Sundelin C: Feeding problems in an affluent society: Follow-up at four years of age in children with early refusal to eat. Acta Paediatr Scand *81:*575, 1992.

Drotar D: Behavioral diagnosis in non-organic failure to thrive: A critique and suggested approach to psychological assessment. J Dev Behav Pediatr *10:*48, 1989.

Evans SL, Reinhart JB, Succop RA: Failure to thrive: A study of 45 children and their families. J Am Acad Child Psychiatry *110:*44, 1972.

Flanagan CH: Rumination in infancy-past and present. J Am Acad Child Psychiatry *16:*140, 1977.

Fraiberg S, Anderson E, Shapiro U: Ghosts in the nursery. J Am Acad Child Psychiatry *14:*387, 1975.

Goldbloom R: Growth in infancy. Pediatr Rev *9:*57, 1987.

Haynes C, Cutler C, Gray J, Kempe R: Hospitalized cases of non-organic failure to thrive: The scope of the problem and the short term lay health visitor intervention. Child Abuse Negl *8:*229, 1984.

Homer C, Ludwig S: Categorization of etiology of failure to thrive. Am J Dis Child *135:*848, 1981.

Lindberg L, Bohlin G, Hagekull B, Thunstrom M: Early food refusal: Infant and family characteristics. Infant Ment Health J *15:*262, 1994.

Linscheid TR, Thomas R, Cunningham CE: A controlled demonstration of the effectiveness of electric shock of chronic infant rumination. J Appl Behav Anal *10:*500, 1977.

Lourie RS: Pica and lead poisoning. Am J Orthopsychiatry *41:*697, 1977.

*Marchi M, Cohen P: Early childhood eating behaviors and adolescent eating disorders. J Am Acad Child Adolesc Psychiatry *29:*112, 1990.

Mayes SD, Humphrey FJ, Handford HA, Mitchell JF: Rumination disorder: Differential diagnosis. J Am Acad Child Psychiatry *27:*300, 1988.

Robinson BA, Tolan W, Golding, Beecher-Beecher O: Childhood pica. Some aspects of the clinical profile in Manchester, Jamaica. West Indian Med J *39:*20, 1990.

Sanders MR, Patel RK, LeGrice B, Shephard R: Children with persistent feeding difficulties: An observational analysis of the feeding interaction of problem and non-problem eaters. Health Psychol *12:*64, 1993.

*Satter E: The feeding relationship: Problems and interventions. J Pediatr *12:*115, 1990.

Starin PS, Fugua RW: Rumination and vomiting in the developmentally disabled: A critical review of the behavioral, medical and psychiatric treatment research. Res Dev Disabil *8:*575, 1987.

Sauvage D, Leddet I, Hameury L, Barthelemy C: Infantile rumination: Diagnosis and follow up study of twenty cases. J Am Acad Child Psychiatry *24:*197, 1985.

*Story M, Neumark-Sztainer D: Promoting healthy eating and physical activity in adolescents. Adolesc Med *10:*109, 1999.

*Waugh E, Bolik CM: Offspring of women with eating disorders. Int J Eat Disord *25:*123, 1999.

Wooster D, Brady N, Mitchell A, Grizzle M, Barnes M: Pediatric feeding: A transdisciplinary team's perspective. Top Lang Disord *18:*34, 1998.

CHAPTER 42. TIC DISORDERS

JAMES T. McCRACKEN, M.D.

The tic disorders comprise a group of neuropsychiatric disorders that appear in childhood or adolescence, encompass a wide range of severity, and share the primary symptom of persistent and interfering tics. The most widely recognized and severest form of tic disorder is Gilles de la Tourette syndrome, or Tourette's disorder. The tic disorders are increasingly recognized as highly familial conditions involving disruptions of frontal, striatal, and subcortical brain circuitry and commonly accompanied by other forms of psychopathology. Treatment of patients with tic disorders usually involves multiple approaches and should be guided by a comprehensive assessment of current associated symptoms and impairment.

DEFINITIONS

Tics are defined as rapid, repetitive muscle contractions or sounds that usually are experienced as outside volitional control and which often resemble aspects of normal movement or behavior. Tics can be elicited by stimuli or preceded by an urge or sensation. They are usually beyond attempts at suppression. Tics are classified as simple or complex tics and as motor or phonic or vocal tics. *Simple motor tics* involve one or a small number of muscle groups (e.g., eyeblinking tic, facial grimace, or shoulder shrug). Simple motor tics can be further subdivided into clonic, tonic, or dystonic types. *Clonic tics* are very brisk movements; *tonic and dystonic tics*, on the other hand, can include more-sustained muscle contractions, such as arm extension and muscle tensing. Other dystonic tics are very common among individuals with Tourette's disorder and can involve oculogyric movements or torticollis.

Complex motor tics, on the other hand, may come close to mimicking normal movements of various kinds, through the synchronous contraction of several muscle groups, as in hopping movements, or the simultaneous extension of arms and legs, in knee bends, and rarely, obscene gesturing or copropraxia.

Phonic tics can also take simple or complex forms. Simple phonic tics include sniffing, grunting, or yelping. Examples of complex vocal tics include intelligible syllables or even phrases, such as “hi,” “I love you, I love you,” or (uncommonly) obscene utterances, defined as coprolalia. While simple tics are relatively easy to discern, complex tics may be mistaken for volitional acts.

As diagnoseable conditions, the tic disorders are currently grouped into four categories, transient tic disorder, chronic motor or vocal tic disorder, Tourette's Disorder, and tic disorder not otherwise specified. According to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders*. (DSM-IV), both chronic tic and Tourette's disorder require that tic symptoms emerge prior to age 18 years and have a minimum duration of 12 months. The criteria for transient tic disorder require a duration of more than 4 weeks but less than 12 months. All of the categories require that the tics produce marked distress or impair functioning, which for many children and teenagers often is manifest in the social ridicule that they receive from peers. For other individuals, the distress of experiencing poorly controlled impulses generates considerable internal tension, dysphoria, frustration, and demoralization.

HISTORY

The formal history and discovery of Tourette's disorder and the tic disorders is credited to a nineteenth century French neurologist, Gilles de la Tourette (1857–1904), who in 1885 published a case series of nine patients afflicted with a disorder characterized by abrupt, involuntary movements, hyperexcitability, echolalia, echopraxia, and compulsive swearing, or coprolalia. de la Tourette had been led to examine and describe the characteristics of these patients by his mentor, Jean Charcot, who had earlier described a woman with abnormal movements and coprolalia. At the time, the classification and relation of tic movements to symptoms of several movement disorders, including the “jumping Frenchmen of Maine,” *latah, miryachit*, and Sydenham's chorea, was uncertain. Tourette accurately identified several integral features of Tourette's disorder still accepted today, such as the childhood onset, the likely hereditary nature of the condition, the distinction between the brisk tic movements and other abnormal movements, and the lack of association of Tourette's disorder with cognitive disability. However, the emphasis on coprolalia and echopraxia are now considered less central to the disorder. Likewise, the possibility that some cases of tic disorders may be associated with prior infections qualifies the earlier separation of Tourette's disorder and Sydenham's chorea in some patients. Indeed, other historians have pointed out an extensive older clinical literature on the role of infection and movement disorders that is being rediscovered in current research. Also of note, biographers have suggested that Samuel Johnson, Lev. Nickolayevich Tolstoy's Anna Karenina, and possibly Wolfgang Amadeus Mozart had tic disorders.

COMPARATIVE NOSOLOGY

The criteria from the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) includes the same tic disorder as does DSM-IV and adds one more, other tic disorders. A different classification has been endorsed by the Tourette Syndrome Study Group and the Tourette Syndrome Association. This system divides tic disorders into transient and chronic types. The chronic type is further divided into chronic single tic disorder, chronic multiple motor or phonic tic disorder, and Tourette disorder.

EPIDEMIOLOGY

In general, increased awareness of the wider range of expression and severity of tic disorders has led to their greater recognition and revised estimates of their prevalence and associated features in various populations. However, as with most mental disorders, changes in definitions influence estimates of prevalence. Although Tourette's disorder was previously considered rare, even this most severe tic disorder has recently been estimated to occur at a point prevalence as high as 0.05 to 5.2 per 10,000. Estimates of the other tic disorders are few, but one study, using the less stringent revised third edition of DSM (DSM-III-R) criteria, screened a sample of all 16- to 17-year-old Israeli military inductees and found point prevalences of 1.8 and 1.6 percent for chronic multiple tics and transient tic disorder, respectively. Similarly, chronic motor and vocal tic disorders were noted to occur in less than 1 percent of control families seen in a family history study of Tourette's disorder. Other surveys of the frequency of transient tic disorder find this disorder to be relatively common in community samples of school-age children, appearing in approximately 5 to 15 percent of children. In clinical samples, chronic tic disorders have been reported to be as frequent as 7 to 34 percent among children with attention-deficit/hyperactivity disorder in child psychiatric clinics. Although the overlap of disruptive behavior and tic disorders is viewed as a common comorbidity in children with chronic tic disorders, it is not shared by most other forms of child psychopathology. Tic disorders are much more common in males than females, by ratios of 1.5 to 9 to 1.

ETIOLOGY

Conclusive evidence of the pathophysiologic basis of tic disorders is lacking, but converging data support the notion that the tic disorders are genetic disorders involving abnormal dopaminergic–excitatory amino acid interactions in neural circuits bridging parts of frontal cortex, basal ganglia, and the thalamus. Additional neuropathological, neuroimaging, and molecular genetic studies under way should greatly advance our knowledge of these disorders and likely will also contribute to our understanding of other, more common forms of psychopathology.

Genetics While the familiality of Tourette's disorder was recognized in the earliest descriptions of the condition, it has proved a difficult disorder to fully understand using current techniques and models of genetic disorders. The inability to confirm a clear genetic basis for the condition may reflect some genetic and etiologic heterogeneity among families with members afflicted with Tourette's disorder. Nevertheless, twin and family studies convincingly demonstrate patterns of inheritance

in many families that are consistent with an autosomal dominant disorder with incomplete penetrance. Furthermore, careful screening of relatives with Tourette's disorder has consistently revealed a relation between Tourette's disorder, other tic disorders, and obsessive-compulsive disorder in some families, suggesting that obsessive compulsive disorder and less severe forms of tic disorders may reflect alternative expressions of a common genetic vulnerability. The high rate of comorbidity of tic disorders with disruptive behavior disorders such as attention-deficit/hyperactivity disorder provides some support for inclusion of attention-deficit/hyperactivity disorder as a possible alternative phenotype. Although some have suggested that tic disorders are part of a spectrum of psychiatric disorders that includes many other disorders with repetitive behaviors, preoccupations, and poor impulse control, evidence supporting a broader spectrum concept is lacking. The recognition that less than half of the offspring of individuals with Tourette's disorder will display Tourette's disorder or related phenotypes has led to the conclusion that epigenetic, or environmental, factors may also influence the expression, severity, and manifestations of tic disorders. These interacting epigenetic factors may include intrauterine experience, birth weight, and other pregnancy or delivery complications. Initial attempts at testing the possible association of candidate genes suspected of causal involvement with Tourette's disorder have had little success, although some positive findings have been reported including an association with the “seven-repeat” allelle of the dopamine type 4 (D_4) receptor subtype (*DRD4.7*). Additional progress in understanding the molecular genetics of Tourette's disorder is expected shortly from ongoing multisite genetic studies of affected sib pairs with Tourette's disorder.

Neurochemistry The potent tic-suppressing effects of drugs such as haloperidol (Haldol) that possess high affinity for D_2 receptors have long been inferred to support the involvement of abnormal dopamine neurotransmission in Tourette's disorder, often suspected to involve hypersensitivity of D_2. Similarly, the often more-subtle benefits of a-adrenergic receptor agonists such as clonidine (Catapres) have expanded notions of the role of other monoamines in the etiopathophysiology of Tourette's disorder. The difficulty directly measuring brain neurochemical systems has impeded researchers' ability to confirm or reject these theories. More-recent models of dopamine abnormalities in Tourettes's disorder include possible interactions between excitatory amino acids such as glutamate and dopamine systems.

A small amount of postmortem and neuroimaging research supports suggested monoamine and excitatory amino acid disturbances in Tourette's disorder. For example, differences in D_2 ligand binding with ^{123}I-benzamide in the caudate nucleus correlate with differences in symptom severity among twins with Tourette's disorder. Similarly, comparisons of dopamine reuptake transporter ligand binding with ^{123}I-b-carbomethoxy-3b-(4-iodophenyl)tropane (^{123}I-b-CIT) in Tourette's disorder patients and controls showed greater dopamine reuptake protein expression in the basal ganglia of Tourette's syndrome patients. In addition, postmortem analyses of brains from Tourette's syndrome patients have suggested higher concentrations of glutamate in globus pallidus and other striatal areas in these patients. Other neurotransmitters less central to tic genesis but probably also dysregulated in Tourette's disorder include serotonin, opioids such as dynorphin, and norepinephrine.

Neuroimaging Both structural and functional neuroimaging studies have contributed to understanding the etiology of Tourette's disorder. Studies examining structural brain differences in subjects with Tourette's disorder have found consistent volume reductions in the region of the basal ganglia as well as loss of normal asymmetries in these structures. Other regional volume differences reported include reductions in the corpus callosum. Taken together, these suggest the involvement of abnormalities in the corticostriatothalamocortical network and subtle disturbances of brain lateralization that may be related to both motor inhibition and information-processing deficits in patients with Tourette's disorder.

Functional imaging studies of Tourette's disorder patients have yielded less-conclusive information on patterns of disturbed brain activity but confirm that many of these patients show deviations from normal. Both hypo- and hypermetabolic regions have been identified by positron emission tomography (PET). In part, the greater variability in brain metabolic patterns in Tourette's disorder patients may reflect the extensive comorbidity common to the disorder.

Rediscovery of a possible link between exposure to infectious diseases and the appearance or worsening of tics has led to studies of immune parameters in Tourette's disorder patients. Overall, these studies have reopened the possibility of an autoimmune subtype of tic disorder, although the frequency of this subtype is far from clear. Support for the involvement of immune responses in tic disorders includes new onsets of tics and obsessive-compulsive symptoms following streptoccocal illnesses. Examination of symptoms of obsessive-compulsive disorder in patients with Sydenham's chorea found that up to one third meet criteria for obsessive-compulsive disorder during the height of the illness. Two studies in Tourette's disorder patients and related groups have found high concentrations of a lymphocyte antigen, the D8/17 protein, which earlier research associated with risk for development of rheumatic fever in response to stretococcal exposure. A possible mechanism for symptom production was suggested by apparent autoantibody binding to brain protein antigens in patients with tic disorders, obsessive-compulsive disorder, and other childhood hyperkinetic movement disorders. These intriguing studies await more-conclusive demonstrations of autoimmune abnormalities in patients with Tourette's disorder and related disorders.

DIAGNOSIS AND CLINICAL FEATURES

The clinical presentation of tic disorders in the psychiatric setting typically falls into one of two general categories: individuals with clearcut, often dramatic, hyperkinetic movements who frequently present during an exacerbation of an evolving or known disorder or patients who, during the course of evaluation or treatment for disruptive behavior or obsessions and compulsions, display more-subtle and possibly previously undetected tics. Both clinical situations require a comprehensive assessment of the child's functioning, range of symptoms, type and severity of impairments, current stressors, developmental issues, and family and school issues and supports.

The frequency and severity of tic symptoms typically wax and wane over weeks and months, which can greatly complicate attempts to gauge the response to treatments. Therefore, clinicians should avoid abrupt, multiple changes in treatment approaches and adopt a careful, longer-term view of treatment for the condition. While some increases in tic severity and frequency are often due to identifiable common stressors such as starting school, peer conflicts, academic challenges, fatigue, excitement, and illnessess, random exacerbations can occur. Another sometimes confusing element of Tourette's disorder and other tic disorders is the degree to which some patients can suppress tics for significant periods. Many children can so effectively suppress, hide, or postpone tic symptoms that even teachers are unaware of the child's disorder. Alternatively, some patients describe their tics as intentional but irresistible. Age influences symptom description and report; younger patients can less often describe mental and volitional aspects of their experience of tic phenomena.

Often the clinical picture is dominated by the common comorbid psychiatric disorders seen in high numbers of individuals with tic disorders, especially obsessive-compulsive disorder and attention-deficit/hyperactivity disorder. Epidemiological, family, and clinical samples display co-ocurring obsessive-compulsive disorder in as many as 20 to 40 percent of patients with Tourette's disorder. Many other Tourette's disorder patients complain of obsessive-compulsive behaviors that are below the threshold for diagnosis but cause internal preoccupations and worry. Recent phenomenological data support some distinctions between obsessive-compulsive symptoms associated with Tourette's disorder (tic-related obsessive-compulsive disorder) and obsessive-compulsive disorder without comorbid Tourette's disorder. Tic-related obsessive-compulsive symptoms are more likely to include more aggressive, sexual obsessions as well as more hoarding, ordering, and touching compulsions.

The association of attention-deficit/hyperactivity disorder and other disruptive behaviors such as aggression with Tourette's disorder is robust among clinic-referred children and adolescents, with 40 to 60 percent of children with Tourette's disorder meeting criteria for attention-deficit/hyperactivity disorder, possibly reflecting shared deficits in response inhibition affecting greater portions of corticostriatothalamocortical circuitry. Curiously, epidemiological samples of individuals with Tourette's disorder do not display the extensive comorbidity with attention-deficit/hyperactivity disorder, perhaps reflecting ascertainment differences. In most instances, the burden of symptoms created by the child's attentional problems, academic underacheivement, and impulsivity far outweigh that produced by tic severity. However, treatment of these associated comorbid conditions can be challenging because of the adverse effects of some treatments on tics, the temporal fluctuations in symptoms, and the frequent presence of other common associated problems such as learning disorders.

Estimates of the exact frequency of co-occurring learning problems in patients with Tourette's disorder and other tic disorders have varied widely, in part because of ascertainment differences, screening approaches, subject age range, and small sample sizes. Understanding learning difficulties in children with Tourette's disorder is complicated, since learning interference may be caused by the direct distracting effects of the tics themselves, by comorbid psychopathology, by specific learning disorders, or by the child's demoralization. Some reports note greater deficits in mathematical skills among Tourette's disorder children than among controls. In addition, interactions between comorbid attention-deficit/hyperactivity disorder, Tourette's disorder, and reduced performance in intelligence quotients (I.Q.) measures have also been reported. Assessments of children with tic disorders must include screening for possible learning disorders.

Other vexing problem behaviors are frequently encountered in patients with Tourette's disorder and other tic disorders, including forms of self-injurious behavior. These may range from skin picking to self-hitting and are clearly related to tics in some individuals. Other less common symptoms in individuals with tics include vomiting, freezing behaviors, rage outbursts, and depressive disorders. In assessing the treatment needs of patients with Tourette's disorder and other tic disorders, one must clearly separate the impact of tic severity on functioning from effects of other psychopathology, developmental stage, life goals, and family responses.

Tourette's Disorder To make a diagnosis of Tourette's disorder, clinicians must obtain a history of multiple motor tics and the emergence of at least one vocal tic at

some point in the disorder. According to both DSM-IV and ICD-10 the tics must occur many times a day nearly every day or intermittently for more than 1 year and cause impairment or distress (Table 42-1 and Table 42-2). The average age of onset of tics is 7 years, but tics may occur as early as the age of 2 years. The onset must occur before the age of 18 years.

A. Both multiple motor and one or more vocal tics have been present at some time during the illness, although not necessarily concurrently. (A tic is a sudden, rapid, recurrent, nonrhythmic, stereotyped motor movement or vocalization.)
B. The tics occur many times a day (usually in bouts), nearly every day or intermittently throughout a period of more than 1 year, and during this period there was never a tic-free period of more than 3 consecutive months.
C. The onset is before age 18 years.
D. The disturbance is not due to the direct physiological effects of a substance (e.g., stimulants) or a general medical condition (e.g., Huntington's disease or postviral encephalitis).

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 42-1 DSM-IV Diagnostic Criteria for Tourette's Disorder

Table 42-2 ICD-10 Diagnostic Criteria for Tic Disorders

In Tourette's disorder, the initial tics are in the face and neck. Over time, the tics tend to occur in a downward progression or involve more complex movements of several muscle groups. The most commonly described tics are those affecting the face and head, the arms and hands, the body and lower extremities, and the respiratory and alimentary systems. In these areas, the tics take the form of grimacing; puckering the forehead; raising eyebrows; blinking eyelids; winking; wrinkling the nose; trembling nostrils; twitching mouth; displaying the teeth; biting the lips and other parts; extruding the tongue; protracting the lower jaw; nodding, jerking, or shaking the head; twisting the neck; looking sideways; head rolling; jerking the hands; jerking the arms; plucking fingers; writhing fingers; clenching fists; shrugging the shoulders; shaking a foot, knee, or toe; walking peculiarly; body writhing; jumping; hiccuping; sighing; yawning; snuffing; blowing through the nostrils; whistling inspiration; breathing exaggeratedly; belching; making sucking or smacking sounds; and clearing the throat.

Typically, prodromal behavioral symptoms—such as hyperactivity, attention difficulties, and poor frustration tolerance—are evident before or coincide with the onset of tics. The most frequent initial symptom is an eye-blink tic, followed by a head tic or a facial grimace. Most complex motor and vocal symptoms emerge several years after the initial symptoms. Coprolalia usually begins in early adolescence and occurs in about 15% of all cases. Mental coprolalia—in which a patient thinks a sudden, intrusive, socially unacceptable thought or obscene word—may also occur. In some severe cases, physical injuries, including retinal detachment and orthopedic problems, have resulted from severe tics.

Chronic Motor or Vocal Tic Disorders The onset of chronic motor or vocal tic disorder appears to be in early childhood. The types of tics and their locations are similar to those in transient tic disorder. Chronic vocal tics are considerably rarer than are chronic motor tics. The chronic vocal tics are usually much less conspicuous than those in Tourette's disorder. The vocal tics are usually not loud or intense and are not primarily produced by the vocal cords; they consist of grunts, throat-clearing, or sniffing or other noises caused by thoracic, abdominal, or diaphragmatic contractions. The DSM-IV diagnostic criteria are given in Table 42-3. The ICD-10 criteria are listed in Table 42-2.

A. Single or multiple motor or vocal tics (i.e., sudden, rapid, recurrent, nonrhythmic, stereotyped motor movements or vocalizations), but not both, have been present at some time during the illness.
B. The tics occur many times a day nearly every day or intermittently throughout a period of more than 1 year, and during this period there was never a tic-free period of more than 3 consecutive months.
C. The disturbance causes marked distress or significant impairment in social, occupational, or other important areas of functioning.
D. The onset is before 18 years.
E. The disturbance is not due to the direct physiological effects of a substance (e.g., stimulants) or a general medical condition (e.g., Huntington's disease or postviral encephalitis).
F. Criteria have never been met for Tourette's disorder.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 42-3 DSM-IV Diagnostic Criteria for Chronic Motor or Vocal Tic Disorder

Transient Tic Disorder The DSM-IV criteria for establishing the diagnosis of transient tic disorder (Table 42-4) are as follows: The tics are single or multiple motor or vocal tics. The tics occur many times a day nearly every day for at least 4 weeks but for no longer than 12 consecutive months. The patient has no history of Tourette's disorder or chronic motor or vocal tic disorder. The onset is before age 18. The tics do not occur exclusively during substance intoxication, and they are not caused by a general medical condition. The diagnosis should specify whether a single episode or recurrent episodes are present. Transient tic disorder can be distinguished from chronic motor or vocal tic disorder and Tourette's disorder only by observing the symptoms' progression over time. The ICD-10 criteria are listed in Table 42-2.

A. Single or multiple motor and/or vocal tics (i.e., sudden, rapid, recurrent, nonrhythmic, stereotyped motor movements or vocalizations)
B. The tics occur many times a day, nearly every day for at least 4 weeks but for no longer than 12 consecutive months.
C. The disturbance causes marked distress or significant impairment in social, occupational, or other important areas of functioning.
D. The onset is before age 18 years.
E. The disturbance is not due to the direct physiological effects of a substance (e.g., stimulants) or a general medical condition (e.g., Huntington's disease or postviral encephalitis).
F. Criteria have never been met for Tourette's disorder or chronic motor or vocal tic disorder.

Specify if: Single episode or recurrent.

Table 42-4 DSM-IV Diagnostic Criteria for Transient Tic Disorder

Tic Disorder Not Otherwise Specified According to DSM-IV, tic disorder not otherwise specified refers to disorders characterized by tics but not otherwise meeting the criteria for a specific tic disorder Table 42-5.

This category is for disorders characterized by tics that do not meet criteria for a specific tic disorder. Examples include tics lasting less than 4 weeks or tics with an onset after age 18 years.

Table 42-5 DSM-IV Diagnostic Criteria for Tic Disorder Not Otherwise Specified

Alan is a 11-year-old boy who lives with his parents and younger sister. His parents reported a long previous history of disruptive symptoms, especially noticeable in school, which led to the diagnosis of attention-deficit/hyperactivity disorder at age 7. His parents sought consultation because of the persistence of several ticlike muscle movements.

Alan was the product of a delivery complicated by breech positioning, leading to birth by casearean section associated with brief fetal cardiac decelerations. Early development was normal except for poor fine motor coordination, persisting into the present in poor handwriting. Preschool staff noted that Alan had difficulty sitting still and attending to class instruction. Poor school performance in the first and second grades improved considerably with the institution of psychostimulant treatment and behavior management.

About 12 months prior to evaluation, the parents had first noted Alan display frequent eyeblinking, which they ascribed to anxiety about restarting school. The eyeblinking persisted, although the frequency of the movements varied from week to week. About 2 months earlier, the parents noticed Alan begin to thrust his head forward violently. For the first time, Alan himself acknowledged these behaviors as outside his control and expressed his frustration over his inability to suppress them. When another movement appeared—repeated shoulder shrugging—the parents sought consultation. At the evaluation, Adam described his frustration and discouragement over his trouble controlling the increasingly frequent muscle jerks. The head thrusts often led to neck soreness and occasional headaches. The eyeblinking was frequent enough to interfere with reading and other activities requiring sustained effort. On examination Alan appeared cooperative but sad and displayed several frequent tic movements involving the face, neck, and shoulders. No vocal tics were noted. Psychomotor activity was increased. Results of the consultation were consistent with chronic motor tic disorder and attention-deficit/hyperactivity disorder. Recommendations included a trial of an a-adrenergic receptor agonist.

PATHOLOGY AND LABORATORY EXAMINATION

Tourette's disorder and the other tic disorders remain clinical diagnoses based upon clinical history and examination. No confirmatory or diagnostic tests exist at present. Brain imaging, electroencephalography (EEG), and neuropsychological testing may be indicated under specific clinical situations to rule out other diagnostic possibilities such as epilepsy or structural lesions or to evaluate for suspected comorbidities, but they should not be considered on a routine basis. Nonspecific EEG abnormalities have been reported to occur in up to 20 to 35 percent of patients with Tourette's disorder and have no clinical significance. Individuals with new onset of tics or a significant exacerbation who also have a recent history of infectious illness should receive throat cultures to test for group A b-hemolytic streptococcal infection, and if positive, serological screening for antistreptolysin O. Aggressive treatment of the infection and eradication of carrier status is advised.

DIFFERENTIAL DIAGNOSIS

Differentiating tics from other abnormal movements is generally straightforward by clinical history and examination (Table 42-6). Although more-complex tics, dystonic tics, and compulsive tics lack the rapidity of other tic movements, in general patients exhibit a mixture of simple and more complex tics, enabling discrimination between tic disorders and other movement disorders. Repetitive vocalizations are rare in nontic movement disorders, which also aids accurate classification. Additional features more strongly associated with tics are premonitory sensations, which are described by up to 80 to 90 percent of older patients. These sensations or experiences can take many forms, including an urge, an itch, a feeling of tightness, tingling, or the experience of irritation or worry. Other subjective features of tics include the experience of relief following the tic; conversely some patients note a buildup of an urge or tension prior to the activity. Other aspects associated with compulsive tics involve the need to repeat behaviors until relief is achieved or the individual feels “just right.” Similarly, the description of tic movements as intentional, or purposely performed, in response to an urge or sensation is uncommon among other types of abnormal movements. The complicated mental phenomena seen in some patients with tics often make the precise boundary between compulsive tics and obsessive-compulsive disorder difficult to define.

Table 42-6 Differential Diagnosis of Tic Disorders

Other features that distinguish Tourette's disorder and the tic disorders from other movement disorders include differences in age of onset (younger in autistic disorder, athetoid cerebral palsy, Pelizaeus-Merzbacher disease, Lesch-Nyhan syndrome, and status dysmyelinatus; older in Huntington's disease, Wilson's disease, spastic torticollis) and the lack of a frequent association of tic disorders with mental retardation or a dementing process (in contrast to autistic disorder, cerebral palsy, Lesch-Nyhan syndrome, Wilson's disease, and Huntington's disease). Occasionally, tics are an adverse effect of commonly prescribed psychostimulants. Generally these tics resolve promptly with dosage reduction or drug discontinuation.

Renee was a 7-year-old girl who lived with both parents and her older brother. She was referred by her parents because of worsening ticlike movements over the previous 12 months. Renee had seemed a healthy, normally developing child until about age 6, when her parents first observed the gradual emergence of several unusual muscle jerks affecting her face and neck.

An initial evaluation by her pediatrician noted the presence of multiple motor tics, but since no clear impairment existed, further monitoring was suggested. Her parents reported that since then, the frequency of tics had fluctuated considerably. However, soon after the most recent resumption of school, her movements had definitely worsened. Especially notable was Renee's difficulty in speaking clearly, which was reported by her teacher. Renee had also asked her parents if there was a medicine available for her “moves.” The parents and Renee denied teasing by peers.

On examination, Renee was a frail girl who was shy but able to engage the interviewer. Her behavior was marked by a number of frequent, vigorous movements that appeared to be involuntary. Renee's responses to questions were often difficult to decipher because of frequent forced exhalations that interfered with her vocalizing. Other behaviors included facial grimacing, head tossings, truncal flexing, and frequent squeaking vocalizations. Renee was tearful at one point while discussing her difficulty in controlling her movements.

Results of the consultation were consistent with the presence of Tourette's disorder and the possible co-occurrence of a specific learning disorder. Recommendations included referral to a local Tourette's disorder parent support group, education about tic disorders, educational and speech-language testing, and a trial of risperidone (Risperdal).

COURSE AND PROGNOSIS

Although the onset criterion in DSM-IV for Tourette's disorder allows symptom onset by age 18, prior reports describe tic onset prior to age 11 in 96 percent of patients. Initial symptoms most commonly involve the face, head, or neck, with eyeblinking being the most common first symptom. Phonic tics generally emerge years after the appearance of motor tics, although exceptions occur as well as cases with fulminant emergence of Tourette's disorder in which multiple motor and phonic symptoms appear within days or weeks. More-distal tics involving the trunk or limbs tend to appear later in the course of the condition; it is often suggested that tic symptoms emerge in a rostral-to-caudal fashion. Disruptive behavior symptoms in Tourette's disorder children frequently tend to precede the initial development of tics. Conversely, obsessions and compulsions generally arise later than the tics. The modal onset period of Tourette's disorder occurs during adrenarche in most children, raising questions about a possible relation between androgenic steroid sensitization and tic symptomatology. However, in experimental trials androgen antagonists have not consistently altered tic frequency or severity in patients with Tourette's disorder.

The course of tic symptoms varies from individual to individual. Tourette's original case descriptions included periods of tic remission followed by exacerbations. Recent follow-up data from small cohorts of children with Tourette's disorder tracked over 8 years found that tic severity reached maximum levels in early adolescence, followed by consistent falls in symptoms over the transition from late adolescence to young adulthood, with some patients reporting negligible or no tics. In other samples followed over time only about 25 percent of Tourette's disorder patients continue to experience moderate-to-severe tics in young adulthood. However, other comorbid conditions can persist and continue to produce significant morbidity.

TREATMENT

Children, adolescents, and adults with tic disorders should be healed from a longitudinal perspective guided by the patient's overall functioning, associated psychopathology, developmental challenges, and family and social adjustment rather than a narrow focus on tic suppression. Indeed, the goal of tic suppression is frequently secondary to improving classroom behavior or academic performance or relieving burdensome obsessions and compulsions. Thus clinicians should be prepared to use a variety of treatment modalities, depending on individual needs, including pharmacotherapy, behavioral treatments, psychotherapy, educational interventions, and family therapy. Treatment choice should be guided by a hierarchy of treatment goals for the individual.

Psychoeducation of the patient and family is required. Helping the patient and parents understand and anticipate periods of symptom exacerbation can reduce feelings of helplessness and frustration. Identifying sources of symptom fluctuations can aid efforts to reduce stressors and weather milder increases in tics. Additional information about resources available through support groups and other organizations can help families navigate complicated educational issues (eg., Tourette Syndrome Association).

If tic suppression is one of the primary goals of treatment, pharmacological approaches can be effective, but the risks and benefits of available agents should be carefully reviewed. High-potency neuroleptics with potent D_2 receptor antagonism such as haloperidol, trifluoperazine (Stelazine), and pimozide (Orap) possess robust tic suppressing effects and significantly reduce tic frequency and severity in 70 to 80 percent of patients at daily dosages of 1 to 8 mg. One study of haloperidol and pimozide found more than 70 percent tic reduction in 64 percent of patients receiving active medication. However, adverse effects may limit dosing in some patients. Among the neuroleptics, pimozide may be better tolerated, with less-frequent dysphoric effects, extrapyramidal symptoms, and cognitive dulling.

The more recent availability of atypical neuroletics such as risperidone and olanzapine (Zyprexa) has led to attempts to test the efficacy of these agents for tic control, with the hope of identifying better tolerated treatments. Initial reports indicate risperidone efficacy for tic suppression at dosages of 1 to 6 mg a day. In general, risperidone is well tolerated, with adverse effects of sedation and weight gain and fewer complaints of cognitive dulling or dysphoria. Despite their effectiveness as tic suppressing drugs, concerns over long-term risks of tardive dyskinesia, interference with learning, and dysphoric side effects often prompt use of nonneuroleptic alternative agents.

Although not fully approved for use in the treatment of tic disorders, the a-adrenergic receptor agonists clonidine and guanfacine (Tenex) are commonly prescribed to treat individuals with tic disorders. Some variability in the efficacy of these agents has been noted in controlled studies with clonidine, with only about 30 to 40 percent of patients showing significant improvements in tic severity. Another advantage of the a-adrenergic receptor agonists is their benefit on other symptom domains common to patients with tic disorders, including hyperactivity and disruptive behavior. While additional studies on the safety and efficacy of these agents in the treatment of tic disorders are still needed, these agents are frequently preferred because of better tolerability than the neuroleptics. At usual dosages (0.2 to 0.4 mg a day for clonidine and 1 to 4 mg a day for guanfacine), most-frequent adverse effects include drowsiness, fatigue, headache, irritability, and occasional hypotension.

A variety of other medications have been suggested as possible treatments for tic disorders. The addition of high-potency benzodiazepines such as clonazepam (Klonopin) to ongoing neuroleptic treatment may further improve tic control, especially in individuals with comorbid anxiety or pronounced internal tension associated with tics. However, treatment with benzodiazepines may risk disinhibition, particularly in younger patients with associated neurological symptoms.

Other suggested pharmacological treatments may be indicated by the comorbidity of tics with other disorders. The effects of stimulants on tic frequency and severity are variable, but most patients with comorbid attention-deficit/hyperactivity disorder and tics can benefit from the effects of stimulants on attention-deficit/hyperactivity disorder without experiencing intolerable increases in tic symptoms. Other alternative treatments for attention-deficit/hyperactivity disorder, such as the tricyclic drugs imipramine (Tofranil), desipramine (Norpramine), or nortriptyline (Aventyl, Pamelor) have both decreased disruptive behaviors and produced mild-to-moderate reductions in tic symptoms.

Other, more experimental treatments for tic disorders have been suggested but require additional study of their efficacy, safety, and tolerability in larger samples of tic disorder patients. In one case series the dopamine receptor agonist pergolide (Permax) benefited patients poorly responsive to standard treatments. The selective monoamine oxidase B inhibitor 1-deprenyl (selegiline; Eldepryl) has benefited a group of children with comorbid attention-deficit/hyperactivity disorder and tics, but conflicting reports exist. Other novel approaches under evaluation are nicotine as an adjunct to neuroleptics, as well as one report of the beneficial effects of mecamylamine (Inversine), a nicotinic antagonist. Other atypical neuroleptics may prove useful, with the promise of fewer adverse effects than currently available

medications.

Besides pharmacological treatments, behavioral interventions may have a role in the treatment of some individuals with tic disorders. Useful behavioral approaches for patients with a smaller number of persistent, bothersome tics include the use of habit reversal techniques, awareness training, and relaxation methods. Other behavioral approaches for patients with tic disorders may involve the use of behavioral management for disruptive behaviors, with application of contingencies for prosocial behaviors, clear expectations for behavior, and consistent environmental responses to challenging behaviors.

While traditional psychotherapy is not considered a primary treatment for patients with tic disorders, it can help educate the patient and family about the illness, help them identify and manage life stressors, develop a healthy acceptance of the disorder without loss of self-esteem, and encourage compliance with other interventions. Patients with demoralization or chronic depression due to peer rejection or other disappointments may benefit from focused psychotherapeutic intervention.

Overall, given the favorable course of symptoms for most patients, available and anticipated treatments, and expanding knowledge about these conditions, the outlook for patients diagnosed with Tourette's disorder or other tic disorders is quite positive. The overwhelming majority of patients with tic disorders can benefit from known interventions.

SUGGESTED CROSS-REFERENCES

Section 44.2 discusses stereotypic movement disorder. Obsessive-compulsive disorder is covered in Chapter 15, Section 2.6 discusses the neuropsychiatric aspects of movement disorders. Further discussion of various drugs used to treat tic disorders is given in Section 31.16 on clonidine. Section 31.17 on dopamine receptor antagonists, and Section 31.4 on medication-induced movement disorders.

CHAPTER REFERENCES

Anderson G, Pollak E, Chatterjee D, Leckman JF, Ridelle MH, Cohen DJ: Postmortem analysis of subcortical monoamines and amino acids in Tourette syndrome. Adv Neurol *58:*123, 1992.

Apter A, Pauls D, Bleich A, Zohar A, Kron S, Ratzoni G, Dycian A, Kotler M, Weizman A, Gadot N, Cohen D: An epidemiologic study of Gilles de la Tourette's syndrome in Israel. Arch Gen Psychiatry *50:*734, 1993.

Braun A, Randolph C, Stoetter B, Mohr E, Cox C, Vladar K, Sexton R, Carso RE, Herscovitch P, Chase TN: The functional neuroanatomy of Tourette's syndrome: An FDG-PET study. II: Relationships between regional cerebral metabolism and associated behavioral and cognitive features of the illness. Neuropsychopharmacology *13:*151, 1995.

Bruun R: Gilles de la Tourette's syndrome: J Am Acad Child Psychiatry *23:*126, 1984.

*Chappell P, Leckman J, Riddle M: The pharmacologic treatment of tic disorders. Child Adolesc Psychiatr Clin North Am *4:*197, 1995.

Como P: Neuropsychological tests for obsessive-compulsive disorder and attention deficit hyperactivity disorder. Neurol Clin North Am *15:*255, 1997.

Cruz C, Camarena B, King N, Paez F, Sidenberg D, de la Fuente JR, Nicolini H: Increased prevalence of the seven-repeat variant of the dopamine D4 receptor gene in patients with obsessive-compulsive disorder with tics. Neurosci Lett *231:*1, 1997.

Dykens E, Leckman J, Riddle M, Hardin M, Schwartz S, Cohen D: Intellectual, academic, and adaptive functioning of Tourette syndrome children with and without attention deficit disorder. J Abnorm Child Psychol *18:*607, 1990.

Eapen V, Robertson MM, Alsobrook JP 2nd, Pauls DL: Obsessive compulsive symptoms in Gilles de la Tourette syndrome and obsessive compulsive disorder: Differences by diagnosis and family history. Am J Med Genet *74:*432, 1997.

*Ernst M, Zametkin AJ, Jons PH, Matochik JA, Pascualvaca D, Cohen R: High presynaptic dopaminergic activity in children with Tourette's disorder. J Am Acad Child Adol Psychiatry *38:*86, 1999.

*Jankovic J: Phenomenology and classification of tics. Neurol Clin North Am *15:*267, 1997.

Kiessling L, Marcotte A, Culpepper L: Antineuronal antibodies: Tics and obsessive-compulsive symptoms. Dev Behav Ped *15:*421, 1994.

Kompoliti K, Goetz C: Clinical rating and quantitive assessment of tics. Neurol Clin North Am *15:*239, 1997.

Kurlan R: Tourette's syndrome and 'PANDAS': Will the relation bear out? Pediatric autoimmune neuropsychiatric disorders associated with streptococcal infection [see comments]. Neurology *50:*1530, 1998.

Lajonchere C, Nortz M, Finger S: Gilles de la Tourette and the discovery of Tourette syndrome. Arch Neurol *53:*567, 1996.

Leckman J, Hardin MT, Riddle MA, Stevenson J, Ort SI, Cohen DJ: Clonidine treatment of Gilles de la Tourette syndrome. Arch Gen Psychiatry *48:*324, 1991.

*Leckman J, Peterson B, Anderson G, Arnsten AF, Pauls DL, Cohen DJ: Pathogenesis of Tourette's syndrome. J Child Psychol Psychiatry *38:*119, 1997.

Leckman J, Zhang H, Vitale A: Course of tic severity in Tourette syndrome: The first two decades. Pediatrics *102:*14, 1998.

Mason A, Banerjee S, Eapen V, Zeitlin H, Robertson MM: The prevalence of Tourette syndrome in a mainstream school population. Dev Med Child Neurol *40:*292, 1998.

Miguel EC, Baer L, Coffey BJ, Rauch SL, Savage CR, O'Sullivan RL, Phillips K, Moretti C, Leckman JF, Jenike MA: Phenomenological differences appearing with repetitive behaviours in obsessive-compulsive disorder and Gilles de la Tourette's syndrome. Br J Psychiatry *170:*140, 1997.

Murphy TK, Goodman WK, Fudge MW, Williams RC Jr, Ayoub EM, Dalal M, Lewis MH, Zabriskie JB: B lymphocyte antigen D8/17: A peripheral marker for childhood-onset obsessive-compulsive disorder and Tourette's syndrome? Am J Psychiatry *154:*402, 1997.

Pauls D, Hurst C, Kruger S, Leckman J, Kidd K, Cohen D: Gilles de la Tourette's syndrome and attention deficit disorder with hyperactivity. Arch Gen Psychiatry *43:*1177, 1986.

*Pauls D, Towbin K, Leckman J, Zahner G, Cohen D: Gilles de la Tourette's syndrome and obsessive-compulsive disorder. Arch Gen Psychiatry *43:*1180, 1986.

*Peterson BS, Cohen DJ: The treatment of Tourette's syndrome: Multimodal, developmental intervention. J Clin Psychiatry *59*(Suppl):62, 1998.

Peterson B, Riddle M, Cohen D, Katz LD, Smith JC, Hardin MT, Leckman JF: Reduced basal ganglia volumes in Tourette's syndrome using three-dimensional reconstruction techniques from magnetic resonance images. Neurology *43:*941, 1993.

Peterson BS, Skudlarski P, Anderson AW, Zhang H, Gatenby JC, Lacadie CM, Leckman JF, Gore JC: A functional magnetic reasonance imaging study of tic suppression in Tourette syndrome, Arch Gen Psychiatry *55:*326, 1998.

Sallee FR, Nesbitt L, Jackson C, Sine L, Sethuraman G: Relative efficacy of haloperidol and pimozide in children and adolescents with Tourette's disorder. Am J Psychiatry *154:*1057, 1997.

Schuerholz LJ, Cutting L, Mazzocco MM, Singer HS, Denckla MB: Neuromotor functioning in children with Tourette syndrome with and without attention deficit hyperactivity disorder. J Child Neurol *12:*438, 1997.

Singer HS, Schuerholz LJ, Denckla MB: Learning difficulties in children with Tourette syndrome. J Child Neurol *10*(Suppl):S58, 1995.

Walkup JT, LaBuda MC, Singer HS, Brown J, Riddle MA, Hurko O: Family study and segregation analysis of Tourette syndrome: Evidence for a mixed model of inheritance. Am J Hum Genet *59:*684, 1996.

Wolf S, Jones D, Knable M, Gorey JG, Leek S, Hyde TM, Coppola R, Weihberger DR: Tourette syndrome: Prediction of phenotypic variation in monozygotic twins by caudate nucleus D_2 receptor binding. Science *273:*1225, 1996.

CHAPTER 43. ELIMINATION DISORDERS

EDWIN J. MIKKELSEN, M.D.

ENURESIS

References to bedwetting have occurred almost since the beginning of recorded history. The modern term “enuresis” actually derives from the Greek word *enourein*, which means to void urine. Over the centuries the phenomenon of enuresis spawned a number of etiological theories, which carried with them differing approaches to treatment. Recent decades have seen substantial progress in the etiology of this disorder. The old observation that enuresis tends to run in families has evolved into sophisticated studies pointing to potential genetic loci. Progress in understanding the etiology of the disorder has been paralleled by the development of new treatment options. The particulars of these advances are detailed below. Viewed from a historical vantage, this progress is both impressive and gratifying.

DEFINITION

Enuresis is currently defined as the involuntary or intentional voiding of urine. The severity is determined by frequency of urination; quantity is not a diagnostic consideration per se. Quantity can become a factor in treatment decisions if a child emits only small quantities of urine; in actual practice, quantity usually does not figure heavily in the treatment plan. Frequency, however, can be important in planning a hierarchy of treatment approaches. The major diagnostic qualifier relates to age of onset. Finally, the definition precludes a physical cause for the disorder.

The length of time before continence is considered established varies in the literature between 6 months and 1 year. DSM-IV does not address that specific chronological issue but rather refers to “a ‘primary' type in which the individual has never established urinary continence, and a ‘secondary' type in which the disturbance develops after a period of established urinary continence.” In practical terms children who have never maintained continence for more than 1 year are referred to as having primary enuresis, whereas children who have achieved continence for 1 year or longer and then lost it are labeled as having secondary enuresis.

The disorder is further defined according to the timing of the episodes throughout the day. Episodes occurring only at night are referred to as nocturnal; daytime wetting is labeled diurnal. Most children display only nocturnal enuresis, but some manifest diurnal or nocturnal and diurnal patterns.

History Enuresis has been described throughout most of written history, with the earliest description being dated to the Papyrus Ebers of 1550 BC. Many treatment approaches prescribed in the past appear sadistic in retrospect. While those attempts may have been well intentioned at the time (as was leeching for other disorders), history underscores the anger that the disorder can engender in parents and clinicians who view the enuresis as voluntary or somehow under the control of the individual.

The modern era of research into enuresis dates from the development of polysomnographic recordings during sleep, which have enabled researchers to pinpoint the physiological correlates of nocturnal episodes. The modern era in treatment parallels the development of behavioral psychology and the application of those techniques to the disorder. The history of successful pharmacological treatments can be seen as dating to the initial description of the efficacy of imipramine (Tofranil) for the disorder in 1960.

Comparative Nosology The variety of differing nosological and diagnostic schemas seen with many disorders does not exist for enuresis. The behavior is concrete and definite: the child either wets or does not. That criterion makes it possible to integrate research data from different countries, cultures, and treatment centers with greater reliability than is possible for many other disorders. The efficacy of a given treatment is usually reported in terms of its impact on the frequency of enuretic events. The primary differences with regard to diagnosis concern the frequency required to make a diagnosis of enuresis as a pathological state and the period of continence necessary to separate primary from secondary enuresis, with a duration of 6 months occasionally being used instead of 1 year. Only slight differences exist between the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) and the fourth edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV). ICD-10 states that enuresis “would not ordinarily be diagnosed in a child under the age of 5 years or with a mental age under 4 years.” This is slightly different from the DSM-IV cutoff of 5 years for both chronological and mental age. However, the tables that accompany ICD-10 indicate the same 5-year cutoff for both chronological and mental age as is found in DSM-IV. The ICD-10 also notes a change in required frequency of enuresis after age 7 (at least twice a month before and once a month after).

Epidemiology The most important data concerning the epidemiology of enuresis come from the Isle of Wight study. That study found that 15.2 percent of boys (7 years of age) were enuretic less than once a week, whereas 6.7 percent were wetting at least once a week or more. In 7-year-old girls the frequency was 12.2 percent for those wetting less than once a week and 3.3 percent with a frequency of once a week or more. By ages 9 and 10 years, 6.1 percent of the boys were wetting less than once a week and 2.9 percent were wetting once a week or more. The corresponding figures for 9- and 10-year-old girls were 3.5 percent wetting less than once a week and 2.9 percent once a week or more. The frequency for boys decreased considerably by age 14, at which time 1.9 percent were wetting less than once a week and 1.1 percent at a frequency of once a week or more. The rates also decreased or girls by age 14, with 1.2 percent wetting less than once a week and only 0.5 percent once a week or more.

A New Zealand study involving 8-year-old children found the prevalence of nocturnal enuresis to be 7.4 percent, with 3.3 percent labeled primary and 4.1 percent secondary. A similar large Scandinavian study involving 7-year-old children reported a prevalence of 9.8 percent, with most (6.4 percent) involving children with nocturnal enuresis, 1.8 percent with diurnal enuresis, and 1.6 percent with mixed nocturnal and diurnal enuresis.

The point prevalence figures cited in DSM-IV are 7 percent of boys and 3 percent of girls at age 5 years, dropping to 3 percent of boys and 2 percent of girls by age 10 years. Only 1 percent of boys are still wetting at age 18 years and still fewer girls. DSM-IV also cites a spontaneous remission rate of between 5 and 10 percent per year after age 5 years. Secondary enuresis may occur at any time but most commonly begins between 5 and 8 years of age.

Etiology Inclusion in the DSM-IV definition of enuresis of children who wet intentionally or involuntarily can be problematic. The vast majority of enuretic children do not wet intentionally or even on a subconsciously motivated basis. Increasingly, the research is pointing toward causal factors that may involve irregularities in physiological processes. Clearly, children who wet intentionally are in a different phenomenological grouping than those who do so involuntarily, even if they do meet the other diagnostic requirements. The most likely explanations for voluntary intentional wetting are either an oppositional defiant disorder or a psychotic disorder. A small number of children originally have enuretic events involuntarily and subsequently manifest the behavior on a voluntary learned basis as well.

There is a correlation between enuresis and psychological disturbance that increases with age. Children living in socially disadvantaged situations and experiencing psychosocial stress have a greater frequency of enuresis than those who are not. The type and range of behavioral disturbance seen in children with enuresis are broad, and no marker can reliably differentiate behaviorally disturbed from nondisturbed enuretic children. Thus, the associated behavioral disturbances are nonspecific and may represent a coincidental or secondary relationship rather than a causal correlation. Further supporting a secondary correlation are the repeated findings that children with enuresis have significantly more developmental delays than nonenuretic children, compared with both controls and other children attending a psychiatric clinic. One comprehensive study reviewed the psychodynamic literature concerning enuresis and encopresis to define the concepts conveyed in that literature and then determine empirically how often those concepts were verified and supported by the analysis of clinical material. The literature review yielded 24 generalizations that were consistently applied to enuretic and encopretic children. In subsequent statistical analysis of the clinical material, only 2 of the 24 generalizations reached statistical significance.

The association between behavioral disturbance and enuresis is stronger with secondary enuresis, and there are anecdotal reports of an association between the onset of enuresis and the loss of a father through divorce or death. The prevalence of psychodynamic factors in children with secondary enuresis is intuitively

reasonable as they have, by definition, demonstrated the physiological competence to maintain continence for a prolonged period of time before losing that ability.

There has been one innovative attempt to explore the possible relation between physiological and psychological factors. Specifically, it was postulated that children with enuresis who did not exhibit a concomitant behavioral disturbance would be those with dysfunctional or abnormal bladders, whereas those with a behavioral disturbance would be more apt to have normal bladders. The study found the converse to be true; the group with behavioral disturbance had more-dysfunctional bladders with lower volumes. The behaviorally disturbed group also had more developmental delays.

The observation that enuresis tends to occur in family members has been made for some time. A large Scandinavian study involving over 3000 children found that a child's risk for being enuretic was 5.2 times greater if the mother was enuretic and 7.1 times greater if the father was enuretic, lending further support to a genetic influence. DSM-IV notes that the concordance rate for the disorder is greater in monozygotic than in dizygotic twins, and 75 percent of children with enuresis have a similarly affected first-degree biological relative. Within the last few years, there have been tremendous advances in regard to the genetic transmission of enuresis. A Scandinavian group identified 11 families that manifested primary nocturnal enuresis over 3 generations in a pattern suggesting autosomal dominant inheritance with penetrance above 90 percent. Genetic linkage studies with these families indicated markers on chromosome 13q. Similarly designed studies have also implicated chromosomes 8 and 12, suggesting genetic heterogeneity.

One particularly important etiological hypothesis was that urinary tract obstruction frequently caused enuresis. The significance of that belief evolves around the related implications for surgery to correct the obstruction. Extensive review of the relevant literature found no basis for concluding that bladder neck repair or urethral dilation was a reasonable or effective treatment for enuresis. That generalization does not apply to patients with specific anatomical or pathophysiological findings.

Research into the association between sleep physiology and enuretic events has evolved steadily over the past three decades. The research can be seen as evolving through the following four stages. During the first phase of the research it was postulated that enuretic events were dream equivalents that occurred in deep sleep. The second major theory grew out of the observation that enuretic events originated in delta sleep and followed arousal signals. Thus, enuresis was seen as a disorder of arousal, which led to the hypothesis that enuretic children who were not psychologically disturbed did not produce arousal signals, whereas behaviorally disordered children did, but failed to respond to them.

The third phase of sleep research is represented by the two largest controlled studies, both of which found that when the time of night is also considered, enuretic events occur in all sleep stages in proportion to the amount of time spent sleeping in each. A fourth phase of investigation has attempted to couple cystometry with sleep studies in an attempt to identify subtypes.

The efficacy of desmopressin (DDAVP) as a treatment for enuresis has led to the theory that some children lack the ability to concentrate urine produced during the night and thus cannot reduce urine volume and manifest enuretic episodes as a result. That hypothesis has led to studies investigating the circadian variation of plasma atrial natriuretic peptide (ANP). Children with enuresis have not consistently been found to differ from controls with regard to ANP. In one study, 14 of 55 children with enuresis had lower ANP concentrations than controls, and 9 of the 14 had an excellent response to treatment with desmopressin. An investigation into the concentration of vasopressin in plasma and urine in 18 children with primary nocturnal enuresis and 20 matched controls found the concentration of vasopressin in plasma at 8:00 AM was significantly lower in the enuretic group. Total 24-hour urinary vasopressin excretion was also lower in the enuretic group but not significantly so.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for enuresis are listed in Table 43-1. The concreteness and simplicity of enuresis make the diagnosis relatively easy. As previously indicated, the diagnosis is not made in a child whose chronological or mental age is below 5. The wetting must occur at least twice a week for at least 3 consecutive months, or if less frequent, it must produce significant distress or functional impairment. Physical causes such as a bladder infection must be excluded. Qualifiers to the diagnosis indicate whether it is primary or secondary enuresis. The other qualifiers refer to the timing of the enuretic event. While most children exhibit only nocturnal enuresis, some have daytime (diurnal) patterns or a combined nocturnal and diurnal pattern.

A. Repeated voiding of urine into bed or clothes (whether involuntary or intentional).
B. The behavior is clinically significant as manifested by either a frequency of twice a week for at least 3 consecutive months or the presence of clinically significant distress or impairment in social, academic (occupational), or other important areas of functioning.
C. Chronological age is at least 5 years (or equivalent developmental level).
D. The behavior is not due to the direct physiological effect of a substance (e.g., a diuretic) or a general medical condition (e.g., diabetes, spina bifida, a seizure disorder).

Specify type:
Nocturnal only
Diurnal only
Nocturnal and diurnal

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 43-1 DSM-IV Diagnostic Criteria for Enuresis

Because DSM-IV includes children who wet intentionally and involuntarily under the same heading, the distinction between the two groups is currently not necessary for diagnosis. However, that distinction has significant clinical implications, as children who are wetting intentionally almost certainly are doing so as a manifestation of a psychological disturbance. There is also an association between behavioral disturbance and involuntary enuresis, but the nature of the correlation is less clear. Thus, although the coexistence of other behavioral problems is not a diagnostic issue per se, it is of clinical importance. The ICD-10 criteria for nonorganic enuresis are presented in Table 43-2.

A. The child's chronological and mental age is at least 5 years.
B. Involuntary or intentional voiding of urine into bed or clothes occurs at least twice a month in children aged under 7 years, and at least once a month in children aged 7 years or more.
C. The enuresis is not a consequence of epileptic attacks or of neurological incontinence, and not a direct consequence of structural abnormalities of the urinary tract or any other nonpsychiatric medical condition.
D. There is no evidence of any other psychiatric disorder that meets the criteria for other ICD-10 categories.
E. Duration of the disorder is at least 3 months.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 43-2 ICD-10 Diagnostic Criteria for Nonorganic Enuresis

John was a 10-year-old male who had consistently wet the bed throughout his life. He averaged three to four nocturnal episodes per week, and his longest period of continence was a few weeks. Thus, John warranted a diagnosis of primary enuresis. John's father had been enuretic until age 12, as was his father and a paternal uncle. Accordingly, he had always assumed that this was a genetic hereditary disorder that John could not control. In keeping with this perception, John's parents had never adopted a punitive approach to the bedwetting. The wetting was handled in a matter-of-fact manner, with John washing his sheets every morning and showering before going to school.

The family sought consultation at this time because John was increasingly being asked to sleep overnight at the homes of friends and also wanted to attend an overnight Boy Scout camp. These concerns led the family to meet with their pediatrician, who suggested a trial of the bell-and-pad method of conditioning. This approach was not effective, in part because the family lived in cramped quarters and the alarm that was meant to awaken John also woke his two brothers who shared the bedroom with him.

At this point, the family's pediatrician suggested that they meet with a child psychiatrist that she worked with. The child psychiatrist described the clinical literature on desmopressin. However, when John's father indicated that his health insurance did not cover prescriptions and that the family was having difficulty covering their household expenses on his income, which was only moderately above minimum wage, the psychiatrist felt compelled to discuss the relative costs of desmopressin and imipramine, which was only a fraction of the cost of desmopressin, since it was available in generic form. However, the psychiatrist also indicated that John would require a baseline electrocardiogram (which his insurance could cover) and that imipramine had more potential adverse effects.

Ultimately, the family decided that John would need medication only for brief discrete periods (overnights at friends and 1 to 2 weeks of summer camp), since they had worked out an acceptable plan for responding to the wetting on a day-to-day basis. Further, they reasoned that within approximately 2 years John would experience a spontaneous remission of his enuresis, as this was the family pattern. Thus, they decided that since desmopression has a rapid onset of action, they would pay out of pocket for the relatively small amounts of demopressin that would be needed for these discrete periods over the next 2 years.

Pathology and Laboratory Examination Because urinary tract infections can produce enuresis, a urinalysis should be part of every evaluation. Using radiographic procedures with contrast media to detect an anatomical or physiological cause for the enuresis is more problematic, as the procedures are invasive and painful, and the diagnostic yield is low. A large study carried out in a pediatric primary care setting found a 3.7 percent incidence of obstructive lesions in children with enuresis. Similar findings have been reported by others.

Differential Diagnosis As suggested above, the primary differential diagnosis is a urinary tract infection. This is especially true for girls, who are more prone than boys to urinary tract infections. With a girl who has been continent for a considerable period and has recently begun wetting, a urinary tract infection should be the first consideration. Although the diagnostic yield is smaller with boys, an urinalysis should still be carried out.

Enuresis can result from anatomical malformations or obstructive lesions, but the percentages are relatively low. If the history and interviews with the child suggest that the enuresis is intentional, then it is almost certainly related to an underlying psychological disturbance. The relation between psychological disturbance and involuntary enuresis is less clear. The coexistence of another behavioral disturbance should be noted and attended to clinically.

Course and Prognosis The natural history of enuresis is significant because it figures prominently in treatment decisions, since enuresis is a self-limited disorder that at some time will spontaneously remit. The fact that the diagnosis is not made until age 5 takes into account children who have delayed toilet training that is not outside the accepted range (between 2 and 5 years of age). The prevalence of enuresis is relatively high between the ages of 5 and 7 and then drops off substantially. The vast majority of enuretic children experience spontaneous resolution of the problem at some time, and only a few remain enuretic into adulthood. By age 14 years only 1.1 percent of boys wet once a week or more. The greatest rates of spontaneous remission occur after age 7 years and again after age 12 years.

Treatment The proved methods for the treatment of enuresis are primarily behavioral and pharmacological. Psychotherapy may be useful for ameliorating some of the associated behavioral problems that can be seen with enuresis, especially secondary enuresis. A particularly common clinical scenario for secondary enuresis is the development of wetting in boys following the loss of their father through death or divorce. In these patients, psychotherapy is the primary treatment modality. A review of the efficacy of psychotherapy for primary enuresis found a 20 percent success rate, which is probably not significantly above what would be expected by spontaneous remissions and random chance.

Behavior Therapy A comprehensive review of several studies determined the success rate for behavioral interventions to be 75 percent. Recent studies have yielded comparable response rates. The primary behavioral intervention is the bell-and-pad method of conditioning. A pad is placed on the bed with a wire running to a bell. When the child wets, the moisture completes a circuit in the pad, ringing the bell and waking the child. With repeated use the child learns to awaken before wetting occurs. Buzzer ulcers were a potential adverse effect of the treatment, but the frequency has decreased considerably with improved technology.

Recent behavioral studies have focused on embellishing the standard method and comparisons with different methods. A study that compared the basic bell-and-pad method with methods in which the bell and pad were coupled with more response contingencies for wetting versus not wetting found no statistical difference other than a slightly lower relapse rate with the adjunctive use of reward contingencies.

The relation between bladder capacity and response to behavioral treatment has also been investigated. One study found that bladder capacity did not affect outcome with the bell-and-pad method of conditioning, whereas another found that children with smaller bladder capacities tended to do slightly better when retention-control training was linked to the bell-and-pad method. In general, bladder capacity does not seem to significantly affect response to behavioral treatment. However, concomitant behavioral disturbance does appear to have a negative impact on the response to behavioral conditioning.

Pharmacotherapy Since the efficacy of imipramine for the treatment of enuresis was described in 1960, over 40 double-blind studies have confirmed that observation. Recent studies have focused on the relation between response and concentration of imipramine in the blood. While one study found no connection between positive response and imipramine concentration either alone or in combination with its metabolite desipramine, three studies found a significant relation between positive response and total concentration of imipramine and desipramine. More specifically, optimal response was found above combined concentrations (imipramine plus desipramine) of 60 ng/mL in one study and above 80 ng/mL in another. The most recent study also found that efficacy was related to increasing dosage but noted tremendous variation in serum concentrations among individual children receiving the same dosage. The adverse effect of dry mouth also correlates with concentration in blood, providing clinicians with a crude index of blood concentration in children who are extremely reluctant to have their blood drawn.

Clinically, when one encounters a history of a lack of response to imipramine, one must ascertain the specific dosage used, as nonresponse is frequently related to dosage: primary care physicians often prescribe only 25 to 50 mg a day. However, low-dosage responders do exist, and it is reasonable to start treatment at a daily dosage of 25 mg, titrating up by 25 mg every 4 to 7 days. This enables identification of the low-dose responders, as the response to imipramine occurs within a day or two. Most children exhibit a positive response in the 75- to 125-mg range. The standard maximal limit for dosage is 5 mg/kg body weight, and electrocardiographic (ECG) monitoring is recommended at dosages above 3.5 mg/kg. A pretreatment baseline ECG is also recommended. Blood concentration values may also be clinically useful if higher dosages fail to elicit a response.

There is the risk of overdosing both by the enuretic child and younger siblings. Children can engage in magical thinking, and there have been reports of children who believed that if taking a few pills would stop the wetting for a night then consuming the whole bottle would make it disappear forever. Thus, parents should be advised to control the medication carefully and to store it in a secure place. Severe overdoses may require treatment with physostigmine (Antilibrium), while more moderate overdoses may be treated by the symptomatic management of arrhythmias and seizures.

Because the vast majority of children at some point experience spontaneous remission of the enuresis it does not make sense to prescribe medication for years without checking for a spontaneous remission. One solution to that problem is to taper the imipramine dosage at 3-month intervals. Should enuresis reappear as imipramine dosage is tapered, it can simply be returned to the optimal level for another 3-month period. Clinically, one often finds that more children remain dry after 3 to 6 months of treatment than can be expected from spontaneous remission alone. In general, treatment with imipramine should be viewed as necessary until spontaneous remission occurs.

Three clinical subtypes of enuretic response to imipramine exist: true nonresponders; true responders who maintain a response over time; and transient responders, who respond for a few weeks and then resume wetting. Increasing the dosage for transient responders usually recaptures the response but only for another few

weeks, and ultimately one reaches a point at which the dosage cannot be increased further. However, the initial transient response can be replicated after a medication-free interval; thus one can use the medication for especially important, time-limited social events such as summer camp.

The most recent advance in the pharmacological treatment of enuresis involves the use of desmopressin. A thorough review of the literature on the use of desmopression for the treatment of enuresis revealed 18 randomized, controlled trials, including a total of 689 subjects. Many subjects had been refractory to previous treatment for enuresis. Desmopression reduced enuretic frequency from 10 to 91 percent in these studies. Positive prognostic factors were age above 9 years and lower pretreatment frequency of enuretic events. In most studies wetting resumed after the medication was discontinued. Follow-up studies indicated that 5.7 percent of children remained dry after cessation of medication, but many of those could have been due to spontaneous remission. The most common reported adverse effects were mild abdominal pain, epistaxis, headache, and nasal stuffiness. Several subsequent studies attempted to find factors that might predict or relate to an increased likelihood of a positive response to desmopression. No one factor appears to predict a positive response, but larger functional bladder capacity and a positive family history of enuresis appear to be positively correlated with treatment success. Factors that are not related to treatment outcome are urine osmolality, plasma osmolality, antidiuretic hormone (ADH) secretion, urine production and nocturnal arginine vasopressin concentration.

Studies have also specifically investigated the efficacy of desmopressin for children who were refractory to conditioning treatment and imipramine. One of these involving 52 children (dosage range, 20 to 40 µg intranasally) found complete cessation of wetting in 53 percent, partial response in 19 percent, and minimal or no response in 28 percent. A long-term follow-up study (mean length of follow-up, 13 months) of children treated with desmopressin found no hematological or hormonal adverse effects. However, there have been an increasing number of case reports of hyponatremic seizures with prolonged use. A review of 14 articles that reported data on serum sodium concentration in desmopressin-treated patients and 11 articles that reported individuals who developed altered consciousness or a seizure while taking desmopressin for enuresis concluded that hyponatremia is a potential adverse effect of desmopressin treatment. The authors also noted that excess fluid intake was a contributing factor in 6 of the 11 case reports. They concluded that patients receiving desmopressin for nocturnal enuresis should be advised not to consume more than 8 ounces of fluid on nights when desmopressin is administered. The data on hyponatremia also suggest that periodic monitoring of serum sodium concentration would be prudent.

Treatment with desmopression has been compared with the bell-and-pad method of conditioning. The difference in response rate was not statistically significant; 86 percent improved with the bell and pad and 70 percent with desmopressin. Another investigation comparing the combined use of the bell-and-pad method and desmopressin with placebo indicated that the combination achieved significantly more dry nights. Similar results were found in a study comparing alarm monotherapy with alarm treatment coupled with desmopressin. The combination proved superior, especially for children with severe, almost nightly, wetting.

The newest innovation in desmopressin treatment is the introduction of oral tablets, which were introduced for treatment of diabetes insipidus and are now used for primary nocturnal enuresis as well. A controlled multicenter study found no difference in response between oral desmopressin and the traditional nasal spray. A dosage of 400 µg produced more dry nights than did a 200-µg dose.

General Treatment Considerations The results of the studies reviewed in conjunction with the natural history of enuresis can help the clinician design a decision tree tailored to a given child. The first decision is whether to treat at all. That decision is affected by the frequency of the wetting, the age of the child, and the amount of psychological distress that the wetting causes the child. Enuresis has a negative impact on the self-esteem of affected children and self-esteem improves following successful treatment. One may choose not to treat a child who is wetting infrequently and is at an age in which a spontaneous remission may be expected.

If one does elect to treat, it makes sense to first use the bell-and-pad method of conditioning, because its efficacy has been repeatedly demonstrated, and because a behavioral technique is considered less invasive than pharmacological approaches. A study comparing observation, imipramine, desmopressin, and the bell and pad indicated that while all of the active treatments were significantly superior to observation alone, the positive response was much more likely to be sustained after treatment cessation with the bell and pad than with either pharmacological treatment. Despite these apparent advantages, one large population-based investigation revealed that only a relatively small number of enuretic children (38 percent) actually saw a physician, and of that number, only 3 percent were treated with the bell-and-pad method of conditioning, while over one third received pharmacological treatment. A more geographically restricted survey of primary care physicians yielded significantly different responses. Eighty percent of physicians indicated that they recommended the bell-and-pad method of conditioning. When the bell-and-pad method of treatment is unsuccessful or not feasible, it makes sense to proceed to pharmacological treatment, assuming that the enuresis is severe enough to warrant pharmacological intervention.

For decades imipramine was the first choice for pharmacological treatment. Although precise figures are not available for its relative use in the treatment of primary nocturnal enuresis, desmopressin appears to be emerging as the most popular pharmacological treatment. If studies continue to indicate that the oral tablet preparation is as effective as the nasal spray, this popularity will likely continue to expand.

Only recently, with the impact of managed care, have articles appeared that address the financial cost of the various treatments for enuresis. The bell-and-pad method of conditioning appears to be the most cost-effective treatment; it is as effective as the pharmacological approaches, is more likely to lead to sustained improvement after the cessation of active treatment, and requires only the initial expense of the apparatus. Other expenses to be considered are the time of the professional who counsels the parents on its use and the parental time involved. Imipramine is available generically, so its actual cost is quite modest. Recent wholesale prices of generic imipramine were $1.71 per 100 for 25-mg tablets and $2.38 per 100 for 50-mg tablets. Related costs are the physician's time and the costs of a baseline ECG, periodic monitoring, and blood concentration determinations. Desmopressin is not available generically. The recent average wholesale price for a 5-ml container of the DDAVP nasal spray was $131.16 and the average wholesale cost for 100 0.2-mg tablets was $260.60. Related costs are the physician's time and monitoring for serum sodium and urine osmolality.

ENCOPRESIS

The problem of fecal soiling, encopresis, has not yielded as well to research as has enuresis. Although there has been progress in understanding some of the pathophysiological elements of the disorder, so far these have not evolved into a unifying theory. As discussed below, this most likely represents the heterogenous psychosomatic nature of the disorder.

Definition DSM-IV currently defines encopresis with four related criteria: (1) the repeated inappropriate passage of feces, usually involuntary; (2) occurrence at least once a month for at least 3 months; (3) a chronological or mental age of 4 years; and (4) exclusion of a substance or medical condition as a cause.

For several years the clinical and research literature concerning encopresis has distinguished between retentive and nonretentive encopresis. Accordingly, DSM-IV lists two subtypes: with constipation and overflow incontinence and without constipation and overflow incontinence. The subtype with constipation and overflow incontinence corresponds to retentive encopresis and is described by DSM-IV as follows: “Feces are characteristically (but not invariably) poorly formed and leakage is continuous, occurring both during the day and during sleep. Only small amounts of feces are passed during toiletting and the incontinence resolves after treatment of the constipation.” The subtype without constipation and overflow incontinence corresponds to nonretentive encopresis and, as DSM-IV notes, “feces are likely to be of normal form and consistency and soiling is intermittent. Feces may be deposited in a prominent location.”

Some clinicians refer to a primary type in which the person has never established fecal continence and a secondary type in which the disturbance develops after a period of established fecal continence.

History The history of encopresis is not as richly detailed as is that of enuresis, possibly because encopresis affects fewer children and the events are less frequent. Unless the feces are particularly fluid it may also be easier for children to hide the symptoms. A high frequency of encopresis and enuresis was found in children who were separated from their parents in World War II, thus documenting the impact of environmental factors.

Comparative Nosology DSM-IV has adopted the widely held clinical distinction between retentive and nonretentive encopresis, although slightly different terminology is used. Although the current DSM-IV definition includes both voluntary and intentional encopresis, that is obviously a distinction of major clinical importance. DSM-IV states that nonretentive encopresis “is usually associated with the presence of Oppositional Defiant Disorder or Conduct Disorder or may be the consequence of anal masturbation.” This statement appears to negate the existence of a rather substantial subgroup of children with nonretentive encopresis who are encopretic because they cannot adequately control the anal sphincter or do not recognize the need to defecate in time.

The most comprehensive categorization of encopresis distinguishes three types of patients: (1) intentional—children who have bowel control but intentionally deposit feces in inappropriate places for psychological reasons; (2) involuntary—those who cannot adequately control the sphincter or lack an awareness of the process; and

(3) involuntary—those whose soiling is related to excessive fluid from retentive overflow (75 percent of this category), diarrhea, or anxiety. Thus, this classification schema incorporates both the retentive and nonretentive dichotomy and the distinction between voluntary and involuntary incontinence.

The ICD-10 diagnostic classification of encopresis essentially follows this schema and allows the notation of the three subtypes in DSM-IV. The primary differences with the DSM-IV criteria are the ability to code for three subtypes rather than two and a duration of 6 months at a frequency of at least once a month, as opposed to the DSM-IV duration of 3 months at a frequency of at least once a month. The ICD-10 also states that if encopresis coexists with enuresis "the coding of the encopresis should have precedence."

Epidemiology Two large epidemiological studies have yielded consistent results. The Isle of Wight study reported a prevalence of 1.3 percent for boys and 0.3 percent for girls ages 10 to 12 years soiling at least once a month. A large study involving 8863 children found a 1.5 percent prevalence rate for boys ages 7 and 8 years. In that study the ratio of boys to girls was more than three to one. A significant relation between encopresis and enuresis has also been found. DSM-IV cites a prevalence of 1 percent in 5-year-old children, with boys being more commonly affected than girls.

Etiology The dramatic nature of fecal soiling has contributed to psychodynamic speculation. The term "chronic neurotic encopresis" has been used to describe the small number of children whose encopresis is predominantly psychological. The clinical characteristics described in these children are (1) a distant father and neurotic mother, (2) early and often harsh bowel training, and (3) a history of neurological delay. The one systematic study that attempted to verify the clinical features usually attributed to enuretic and encopretic children in the literature by analyzing actual case material could not substantiate the existence of 20 of the 22 factors developed through the literature review.

As with enuresis, the distinction between primary and secondary encopresis relates to the issue of associated psychopathology. A study involving 63 boys with encopresis found that boys with primary encopresis were more likely to have developmental delays and associated enuresis while those with secondary enuresis were more likely to have experienced higher levels of psychosocial stressors and to be diagnosed with conduct disorder. A large study looked at several psychological and physiological variables in children with chronic constipation and children with encopresis unrelated to constipation. The group with constipation had significantly longer colonic transit times and both groups had relatively high frequencies of abnormal defecation dynamics. Both groups were significantly different from controls but were not significantly different from each other.

No correlation between encopresis and social class was reported in two independent studies. The issue of neurological competence as a contribution to encopresis has been investigated. One study found no abnormalities of anal-rectal motor or sensory function but did find that a significant number of encopretic boys had abnormal anal-rectal expulsion dynamics. A sophisticated physiological study using a triple-lumen catheter attached to a hydraulic manometry infusion system found that age of onset, frequency, and duration of encopresis were all highly correlated with the amount of anal sphincter spasm that occurred when defecation was attempted. The amount of pain reported with bowel movements correlated positively with frequency of encopresis and inversely with maximum squeeze pressure. An investigation involving the concentrations of gastrointestinal hormones in controls and children with retentive encopresis indicated that the latter had higher concentrations of pancreatic polypeptide after a meal and a lower motilin response. However, the authors noted that these abnormalities could be the result of, or the cause of the chronic constipation.

An extensive series of studies concerning physiological factors and the interplay of physiological and psychological factors was carried out. The principal physiological findings were (1) 56 percent of children with retentive encopresis could not defecate rectal balloons (research technique used to assess sphincter competency), and most had abnormal anal sphincter contractions; (2) only 14 percent of those who could not defecate the balloons had responded to treatment at 1 year follow-up compared with a 64 percent success rate for those who could; (3) at 1-year follow-up, 70 percent of children who could relax the anal sphincter at the time of initial evaluation were improved as opposed to 13 percent of those who could not; and (4) none of the children who initially presented with an abdominal fecal mass were improved at 1-year follow-up, regardless of other factors.

One subset of these studies physiologically compared constipated children with controls during the act of bearing down. All of the control children displayed decreased anal sphincter activity when bearing down, compared with 58 percent of the constipated children who could defecate a balloon and 7 percent of constipated children who could not defecate the balloon. The constipated children who could not defecate the balloon were also significantly less likely to respond to laxative treatment.

An attempt was also made to assess the relative impact of physiological and psychological factors. Specifically, anal-rectal manometric and electromyographic assessments were correlated with behavioral and social competence profiles as related to treatment outcome. Psychological variables were not found to predict outcome, but physiological variables were.

Diagnosis and Clinical Features The DSM-IV diagnostic criteria for encopresis are listed in Table 43-3. The rather objective and concrete nature of the encopretic event makes diagnosis relatively easy and uncomplicated from a phenomenological perspective, if the child meets the frequency, duration, and age requirements set forth in DSM-IV. The involuntary versus intentional dichotomy is neither a requirement nor a reason for subclassification according to DSM-IV, but it has important clinical implications. Similarly, the retentive versus nonretentive schema has clinical utility. Differentiating between primary and secondary encopresis is important, as one might expect medical or psychological factors to play a greater role in secondary encopresis. The ICD-10 criteria for inorganic encopresis are presented in Table 43-4.

A. Repeated passage of feces into inappropriate places (e.g., clothing or floor) whether involuntary or intentional.
B. At least one such event a month for at least 3 months.
C. Chronological age of at least 4 years (or equivalent developmental level).
D. The behavior is not due exclusively to the direct physiological effects of a substance (e.g., laxatives) or a general medical condition except through a mechanism involving constipation.

Code as follows:
With constipation and overflow incontinence
Without constipation and overflow incontinence

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 43-3 DSM-IV Diagnostic Criteria for Encopresis

A. The child repeatedly passes feces in places that are inappropriate for the purpose (e.g., clothing, floor), either involuntarily or intentionally. (The disorder may involve overflow incontinence secondary to functional fecal retention.)
B. The child's chronological and mental age is at least 4 years.
C. There is at least one encopretic event per month.
D. Duration of the disorder is at least 6 months.
E. There is no organic condition that constitutes a sufficient cause for the encopretic events.

Reprinted with permission from World Health Organization: *The ICD-10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*. © World Health Organization, Geneva, 1993.

Table 43-4 ICD-10 Diagnostic Criteria for Nonorganic Encopresis

Henry was an 11-year-old male with almost-daily encopresis and a number of associated behaviors including hiding the feces around the house. He resided in a specialized foster care setting, having been removed from his biological parents at age 7, because of physical and sexual abuse. Both parents were involved with substance abuse, and his early history is not well documented. However, the parents did indicate that he had never exhibited sustained bowel continence for several months. Henry had also been enuretic until age 6, but this had resolved to an occasional nocturnal episode every 4 to 6 months.

Henry also qualified for a diagnosis of oppositional-defiant disorder. Although he had experienced physical and sexual abuse, he did not have flashbacks or other symptoms that would meet the criteria for posttraumatic stress disorder. Henry also had an attention-deficit/hyperactivity disorder and was being effectively treated with 10 mg of methlyphenitate (Ritalin) twice a day.

The foster family resided in an urban area that had access to a nationally recognized children's hospital. The Ambulatory Care Department had a specialized behavioral encopresis program that coupled the bowel training method with a psychoeducational component and psychotherapy. The psychiatric consultant to the specialized foster care program doubted that this program would be successful for Henry since he had so much associated psychopathology and the feces were often deposited around the house in a symbolic manner. Also, the encopresis was not of the retentive-overflow type, and the feces were always well formed. However, since no apparent harm could come from the referral, the consulting child psychiatrist agreed to it. Much to the surprise of the consultant, the several-week outpatient bowel training course coupled with the psychoeducational component and psychotherapy resulted in a complete cessation of the encopresis. On one of her visits to the home, Henry proudly showed his case manager a diagram of the functioning of the digestive system that was part of the psychoeducational program. In retrospect, it appeared that although there were symbolic aspects to Henry's encopretic behavior, the soiling was ego-dystonic, and he was highly motivated to change the behavior, although this motivation could not be prospectively detected by the treatment team because of his oppositional-defiant manner of responding to adults.

Pathology and Laboratory Examination The extensive, detailed physiological studies cited above suggest that physiological factors are important in many encopretic children as they correlate with treatment outcome at 6-month and 1-year follow-up. However, that does not imply that a physiological workup this thorough is required for every encopretic child who presents clinically. The clinician needs to rule out definite physiological illnesses, such as Hirschsprung's disease. Beyond that clinicians might consider more detailed physiological investigations, such as those used in the experiments described above, for children who prove refractory to conventional treatment. That type of specialized testing could be difficult to obtain outside an academic research center. A simple plain abdominal roentgenogram may aid in the diagnosis and confirmation of retentive encopresis. In one study 78 percent of children who met the diagnostic criteria for encopresis in the revised third edition of DSM (DSM-III-R) also met roentgenographic criteria, while 22 percent did not.

Psychological testing can be useful for children who exhibit concomitant behavioral problems or who are intentionally encopretic. Finally, clinicians should not overlook the importance of a detailed history obtained from both parents and the child that documents the time course, frequency, and circumstances surrounding the encopresis.

Differential Diagnosis Encopresis can be a symptom of other processes or be a syndrome in itself as described above. The medical illnesses that need to be ruled out include stenosis of the rectum or anus, endocrine abnormalities, smooth muscle disease, and Hirschsprung's disease.

The child's psychological profile should be considered, as children with mental retardation or pervasive developmental delay may have episodes of soiling related to those processes. Children with impulse-control disorders or attention-deficit disorder may at times have episodes of soiling related to lack of attention. A thorough history and psychiatric assessment should make it possible to identify both those children and children who soil on an intentional-oppositional basis. Extreme stress may also provoke encopretic episodes in children who otherwise function well.

Course and Prognosis As with enuresis, continued maturation provides increasing numbers of spontaneous remissions with time. That trend should be factored into the assessment of the efficacy of any long-term intervention. The results from relatively simple behavioral interventions involving educational, behavioral, and physiological components are striking (78 percent), suggesting that the disorder responds to treatment in most cases.

Treatment The first clinical approach to encopresis is primarily behavioral, with educational and physiological components. That method entails initial educational interventions with the child and family around bowel function. The process is meant to correct misconceptions the family may have and dissipate anxiety that may develop around the soiling. The physiological component is instituted next with an initial bowel catharsis followed by daily dosages of laxatives or mineral oil. The behavioral element of the program entails daily timed intervals on the toilet with success being rewarded. Follow-up studies have reported a 78 percent success rate for the method. The physiological research findings concerning abnormal defecation dynamics in children with encopresis led to interest in biofeedback training as an adjunctive treatment. Although this treatment received positive reports from uncontrolled studies, a review of controlled studies indicated that biofeedback training did not provide additional benefit over conventional treatment.

One element of the aforementioned physiological studies, namely the ability to defecate a rectal balloon, correlates with treatment outcome. A study involving 139 children with constipation and encopresis and 20 controls found that all of the controls could defecate the balloon but only 47 percent of the encopretic group could do so.

One year after beginning treatment, 51 percent of the children who could defecate the balloon at baseline were considered recovered but only 34 percent of those who could not. Although the group comparisons were considered significant, the balloon test alone could not predict response to treatment on a individual basis.

Psychotherapy may be useful for children who have concomitant behavioral problems and those with psychodynamic factors that appear to be especially strong and contributing to the disorder. Naturally, psychotherapeutic approaches are of primary importance for children who are intentionally soiling or depositing feces in inappropriate places around the home.

Pharmacological treatment for encopresis has not been as extensively studied as for enuresis. Six case reports (encompassing 15 individual patients) have described imipramine as effective for encopresis. Twelve of the 15 patients were boys. The therapeutic effect had relatively rapid onset (as it does with enuresis), occurring within a few days to 2 weeks, usually at dosages of 25 to 75 mg. Although the subtype of encopresis is not delineated in all of the case reports, one would anticipate that if imipramine were to be helpful, it would be in those with nonretentive encopresis.

The relatively high success rate of the combined educational, physiological, and psychological approach described makes it a reasonable first approach once other diagnoses have been ruled out. Psychotherapeutic and pharmacological approaches could be considered for children who prove refractory to that form of treatment.

SUGGESTED CROSS-REFERENCES

Section 1.19 on the basic science of sleep, Section 3.3 discusses learning theory, Chapter 21 presents the sleep disorders, Section 30.2 discusses behavior therapy, and Section 31.30 covers tricyclic and tetracyclic drugs. Psychotherapy with children is discussed in Section 46.1 and Section 46.2.

CHAPTER REFERENCES

*Arnell H, Hjalmas K, Jagervall M, Lackgren G, Stenberg A, Bengtsson B, Wassen C, Emahazion T, Anneren G, Pettersson U, Sundvall M, Dahl N: The genetics of primary nocturnal enuresis—inheritance and suggestion of a second major gene on chromosome 12Q. J Med Genet *34:*360, 1997.

Bernstein SA, Williford SL: Intranasal desmopressin-associated hyponatremia: A case report and literature review. J Fam Pract *44:*203, 1997.

Bradbury M: Combination therapy for nocturnal enuresis with desmopressin and an alarm device. Scand J Urol Nephrol *31:*61, 1997.

Butler R, Holland P, Devitt H, Hiley E, Roberts G, Redfern E: The effectiveness of desmopressin in the treatment of childhood nocturnal enuresis: Predicting response using pretreatment variables. Br J Urol *81:*29, 1998.

Donoghue MB, Latimer ME, Pillsbury HL, Hertzog JH: Hyponatremic seizure in a child using desmopressin for nocturnal enuresis. Arch Pediatr Adolesc Med *152:*290, 1998.

*Eggert P, Muller-Schluter K, Muller D: Regulation of arginine vasopressin in enuretic children under fluid restriction. Pediatrics *103:*452, 1999.

*Eiberg H, Berendt I, Mohr J: Assignment of dominant inherited nocturnal enuresis (ENUR1) to chromosome 13Q. Nature Genetics *10:*354, 1995.

*El-Anany FG, Maghraby HA, Shaker SED, Abdel-Moneim AM: Primary nocturnal enuresis: A new approach to conditioning treatment. Urology *53:*405, 1999.

Foreman DM, Thambirajah MS: Conduct disorder, enuresis and specific developmental delays in two types of encopresis: A case-note study of 63 boys. Eur Child Adolesc Psychiatry *5:*33, 1996.

Friman PC, Handwerk ML, Swearer SM, Mcginnis JC, Warzak WJ: Do children with primary nocturnal enuresis have clinically significant behavior problems? Arch Pediatr Adolesc Med *152:*537, 1998.

Fritz GK, Rockney RM, Yeung AS: Plasma levels and efficacy of imipramine treatment for enuresis. J Am Acad Child Adolesc Psychiatry *33:*60, 1994.

Hagglof B, Andren O, Bergstrom E, Marklund L, Wendelius M: Self-esteem before and after treatment in children with nocturnal enuresis and urinary incontinence. Scand J Urol Nephrol *31:*79, 1997.

Hersov L: Faecal soiling. In *Child and Adolescent Psychiatry: Modern Approaches*, ed 2, M Rutter, L Hersov, editors. Blackwell Scientific, London, 1985.

Hogg RJ, Husmann D: The role of family history in predicting response to desmopressin in nocturnal enuresis. J Urol *150:*444, 1993.

Ilyas M, Jerkins GR: Management of nocturnal childhood enuresis in managed care: A new challenge. Pediatr Ann *25:*261, 1996.

*Janknegt RA, Zweers HM, Delaere KP, Kloet AG, Khoe SG, Arendsen HJ: Oral desmopressin as a new treatment modality for primary nocturnal enuresis in adolescents and adults: A double-blind, randomized, multicenter study. J Urol *157:*513, 1997.

*Loening-Baucke V: Biofeedback training in children with functional constipation. A critical review. Dig Dis Sci *41:*65, 1996.

Loening-Baucke V: Balloon defecation as a predictor of outcome in children with functional constipation and encopresis. J Pediatr *128:*336, 1996.

Mikkelsen EJ: Enuresis and encopresis. In *Child and Adolescent Psychiatry: A Comprehensive Textbook*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Moffatt ME, Harlos S, Kirshen AJ, Burd L: Desmopressin acetate and nocturnal enuresis: How much do we know? Pediatrics *92:*420, 1993.

Monda JM, Husmann DA: Primary nocturnal enuresis: A comparison among observation, imipramine, desmopressin acetate and bed-wetting alarm systems. J Urol *154:*745, 1995.

*Neveus T, Lackgren G, Tuverno T, Stenberg A: Osmoregulation and desmopressin pharmacokinetics in enuretic children. Pediatrics *103:*65, 1999.

Nolan T, Cattosmith T, Coffey C, Wells J: Randomised controlled trial of biofeedback training in persistent encopresis with anismus. Arch Dis Child *79:*131, 1998.

Oredsson AF, Jorgensen TM: Changes in nocturnal bladder capacity during treatment with the bell and pad for monosymptomatic nocturnal enuresis. J Urol *160:*166, 1998.

Pugner K, Holmes J: Nocturnal enuresis—Economic impacts and self-esteem preliminary research results. J Clin Urol Nephrol *31:*65, 1997.

Rappaport L: Prognostic factors for alarm treatment. Scand J Urol Nephrol *31:*55, 1997.

Rittig S, Knudsen UB, Norgaard JP, Gregersen H, Pedersen EB, Djurhuus JC: Diurnal variation of plasma atrial natriuretic peptide in normals and patients with enuresis nocturna. Scand J Clin Lab Invest *51:*209, 1991.

Robson WL, Norgaard JP, Leung AK: Hyponatremia in patients with nocturnal enuresis treated with DDAVP. Eur J Pediatr *155:*959, 1996.

Rockney RM, McQuade WH, Days AL: The plain abdominal roentgenogram in the management of encopresis. Arch Pediatr Adolesc Med *149:*623, 1995.

*Rockney RM, McQuade WH, Days AL, Linn HE, Alario AJ: Encopresis treatment outcome: Long-term follow-up of 45 cases. J Dev Behav Pediatr *17:*380, 1996.

Rushton HG, Belman AB, Zaontz MR, Skoog SJ, Sihelnik S: The influence of small functional bladder capacity and other predictors on the response to desmopressin in the management of monosymptomatic nocturnal enuresis. J Urol *156:*651, 1996.

Rutter M: Isle of Wight revisited: Twenty-five years of child psychiatric epidemiology. J Am Acad Child Adolesc Psychiatry *28:*633, 1989.

Shaffer D, Gardner A, Hedge B: Behavior and bladder disturbance of enuretic children: A rational classification of a common disorder. Dev Med Child Neurol *26:*781, 1984.

Stark LJ, Opipari LC, Donaldson DL, Danovsky MB, Rasile DA, DelSanto AF: Evaluation of a standard protocol for retentive encopresis: A replication. J Pediatr Psychol *22:*619, 1997.

Steffens J, Netzer M, Isenberg E, Alloussi S, Ziegler M: Vasopressin deficiency in primary nocturnal enuresis. Results of a controlled prospective study. Eur Urol *24:*366, 1993.

Stern HP, Stroh SE, Fiedorek SC, Kelleher K, Mellon MW, Pope SK, Rayford PL: Increased plasma levels of pancreatic polypeptide and decreased plasma levels of motilin in encopretic children. Pediatrics *96:*111, 1995.

Sutphen J, Borowitz S, Ling W, Cox DJ, Kovatchev B: Anorectal manometric examination in encopretic-constipated children. Dis Colon Rectum *40:*1051, 1997.

Vogel W, Young M, Primack W: A survey of physician use of treatment methods for functional enuresis. J Dev Behav Pediatr *17:*90, 1996.

Vongontard A, Eiberg H, Hollmann E, Rittig S, Lehmkuhl G: Molecular genetics of nocturnal enuresis—clinical and genetic heterogeneity. Acta Paediatr *87:*571, 1998.

Vongontard A, Hollmann E, Eiberg H, Benden B, Rittig S, Lehmkuhl G: Clinical enuresis phenotypes in familial nocturnal enuresis. Scand J Urol Nephrol *31:*11, 1997.

Yap HK, Chao SM, Tan AYS, Murugasu B, Ong EK, Low EH: Efficacy and safety of oral desmopressin in the treatment of primary nocturnal enuresis in Asian children. J Paediatr Child Health *34:*151, 1998.

CHAPTER 44. OTHER DISORDERS OF INFANCY, CHILDHOOD, AND ADOLESCENCE

44.1 REACTIVE ATTACHMENT DISORDER OF INFANCY AND EARLY CHILDHOOD

NEIL W. BORIS, M.D., AND CHARLES H. ZEANAH, M.D.

Reactive attachment disorder of infancy and early childhood is one of the few disorders in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) applicable to children under the age of 5. Since this disorder first appeared in the third edition of DSM (DSM-III), the criteria have been substantially revised. The rationale for DSM-IV and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) criteria for the disorder clearly arises from converging lines of research on institutionalized and maltreated infants and young children. However, no systematic research validates these criteria, and reactive attachment disorder of infancy and early childhood is rarely cited in the literature.

The normal development of the infant and young child's attachment system, first elucidated in John Bowlby's theory of attachment, is the marker against which behavior indicating disordered attachment is compared (Table 44.1-1). Two general patterns of deviant social responses have been described in published case studies and reviews of the literature on disturbances of attachment in institutionalized or maltreated infants and young children are reflected in current criteria for the disorder. Other less well studied or more subtle patterns also appear to exist. While the development of primary attachment relationships in early childhood is reciprocal, criteria clearly specify that this disorder occurs “as a result of grossly pathogenic care.”

	Birth to 2 months	3 to 7 months	7 to 12 months	12 to 18 months	Over 18 months
Phase of attachment	Limited discrimination	Discrimination with limited preference	Preferred attachment	Secure base	Goal-corrected partnership
Characteristics	Physical attributes of babies/news attract caregivers, but infant expressions of preference limited to olfactory and auditory realms and require experimental conditions to demonstrate	Infant differentiates among different interactive partners; may seem more comfortable with primary caregiver but willing to interact readily with other social partners	Clearly expressed preference for a small number of caregiving adults; separation protests and stranger wariness are normative	Use of the attachment figures as a secure base from which to venture out confidently and explore the world; and a safe haven to return to in times of danger; proximity to caregiver promotes internal feeling of security in infant	Cooperation with caregiver despite frequent conflicts; goal is to balance autonomous functioning with reliance on caregiver for help when needed; verbal relationships important including attention to emotional integration and fluency of discourse

Table 44.1-1 Developmental Features of Attachment in Early Childhood

A number of important issues regarding reactive attachment disorder remain unresolved. First, can the disorder be reliably diagnosed? Does the disorder exist within the child or within the relationship? Can the disorder exist when grossly pathogenic care cannot be documented? What other abnormal attachment patterns are disturbed enough to be considered disordered? What is the prevalence and natural history of the disorder? How is the disorder related to other disorders, particularly pervasive developmental disorders?

DEFINITION

Reactive attachment disorder of infancy and early childhood as described in DSM-IV is characterized by “markedly disturbed and developmentally inappropriate social relatedness in most contexts.” These findings must occur in the context of “grossly pathogenic care.” The disorder must begin before 5 years of age to meet criteria and cannot be “accounted for solely by developmental delay.” Children who are mentally retarded are thus difficult to diagnose; those who meet criteria for pervasive development disorder are explicitly excluded from consideration for reactive attachment disorder.

Two subtypes are spelled out in the DSM-IV criteria. The first pattern, generally linked in the literature to early childhood maltreatment, is characterized by inhibition of the normal developmental tendency to seek comfort from a select group of caregivers. Responses to social interactions are “excessively inhibited, hypervigilant, or highly ambivalent,” reflecting the overall inhibition of the attachment system in affected children. The second pattern, linked to institutionalization or exposure to multiple caregivers before age 5, is characterized by a relative hyperactivation of the attachment system, resulting in “diffuse” and unselective attachments, and patterned behavior labeled “indiscriminate sociability.”

Attachment and Development Attachment, as described by Bowlby in his influential trilogy, refers to a biobehavioral system whose goal is to coordinate the balance between the need for safety in proximity to a caregiver or set of caregivers with the tendency for exploration and autonomy in infancy and early childhood. Bowlby argued that human infants were motivated by this heritable attachment system to seek an external goal of safety in proximity to a small number of identified caregivers (usually with the infant's mother) and an internal goal of “felt security.” This internal subjective sense of security is influenced heavily by the experience of the caregiver's emotional and physical availability in times of need and modified by the infant's own temperamental makeup. From the usual biobehavioral shift at age 7 to 9 months through the first 3 years, infants are thought to begin to form internal representations of their relationships with important caregivers. These representations (Bowlby called them “internal working models”) form the basis for both the intense emotional bond between infants and their primary caregivers and behavior in later relationships. This link provides evidence that attachment is salient throughout the course of development. Although these representations are relatively stable, they may be modified by significant experiences or relationships later in life.

Bowlby's theoretical framework, refined from the early 1950s to the 1970s, has evoked much research in the last two decades on the development of patterns of secure and insecure attachment in early childhood. The normal developmental progression of attachment, captured in attachment-related behaviors, is summarized in Table 44.1-1. Though little of this research has been directed toward defining clinical disorders of attachment, much of it is germane to clinical medicine; assessment of attachment has become necessary for infant mental health professionals.

Deviant patterns of attachment like those reflected in criteria for reactive attachment disorder have been recognized in populations raised in environments marked by extreme deprivation. The normative tendency to seek comfort actively from a restricted number of familiar caregivers in times of distress, a hallmark of secure attachment, is commonly not apparent in many of these children. Moreover, a series of anomalous behaviors in these relationships are evident and are reflected in the criteria for reactive attachment disorder.

These anomalous behaviors are also the hallmark for a pattern of attachment labeled *disorganized/disoriented*. This pattern of attachment has been identified using standardized laboratory assessments. Recent longitudinal research suggests that disorganized/disoriented attachment in infancy is a risk factor for the development of psychopathology. Further, the antecedents of this pattern of attachment are primarily environmental factors that impact early relationship formation.

These findings suggest that subsyndromal disturbances of attachment such as disorganized/disoriented attachment exist on a continuum. The findings also suggest that clinicians should be familiar with risk factors impacting early caregiver–infant interactions that are related to the development of disorganized/disoriented

attachment. Maternal psychopathology, child maltreatment, family violence, and poor parental sensitivity to infant cues have been linked to the development of disorganized/disoriented attachment.

HISTORY

The importance of early experience on infant development was recognized as far back as the thirteenth century when the Holy Roman Emperor Frederick II ordered that no one should speak to, or interact with, a group of infants. His experiment, designed to see what language the children would learn to speak, ended with the premature death of all its subjects. Notes inscribed at the time by a Franciscan monk document the devastating social and emotional effects this lack of social interaction had on the infants.

This crude experiment was one of the first designed to test the relative importance of nature versus nurture in influencing development. The nature-versus-nurture theme is as evident in the history of philosophy as it is today in the field of experimental developmental psychology. The work of Aristotle and Plato includes reflection on the relative importance of biology and environment in influencing development. The Enlightenment in the seventeenth and eighteenth centuries brought more interest in the origins of personality; the eminent English philosopher John Locke argued that the mind of the infant was a tabula rasa on which layers of experience were imprinted to create a person with a unique personality. This view was countered by Jean-Jacques Rousseau, among others, who viewed infants as being endowed at birth with inherent goodness that would naturally develop except in circumstances marked by parenting deficiency. By the nineteenth century, Charles Darwin had suggested that complex individual traits, including intelligence, could be largely accounted for by hereditary factors.

The continued interest in the relative importance of nature and nurture has shaped much of the research in early childhood development in this century. The relatively common practice of institutionalization of infants in orphanages in the first half of the century unwittingly provided more evidence of the detrimental effects of what has been called “maternal deprivation.” Pediatricians had long argued against the practice of institutionalization, particularly because of the high mortality rates related to failure to thrive, but health care providers did not take notice of the effects of institutionalization on social and emotional development until René Spitz in the 1940s and James and Joyce Robertson in the 1950s conducted more rigorous studies (including control groups). Films that showed the effects of prolonged separations from caregivers and of institutionalization heightened interest in this social problem. These studies also led Spitz to propose a diagnostic entity he called “anaclitic depression,” which foreshadowed the DSM-III criteria for reactive attachment disorder and is consistent with some of the features of the DSM-IV inhibited subtype. Anaclitic depression seemed to follow prolonged separation of infants from their caregivers and was most severe if the infants were old enough to have developed a preference for that caregiver. The eventual recognition of the clinically meaningful consequences of institutionalization led to a marked decrease in state-run orphanages in many industrialized countries. However, the recent influx of young children from Romania and some Russian states has renewed interest in studying this population of children.

Recognition of the scope of the problem of child maltreatment, on the other hand, did not begin until relatively recently, spurred initially by C. Henry Kempe's classic 1962 article on the “battered child syndrome.” Since the early 1960s, there has been a great deal of research on the development effects of maltreatment on children. Though this research is confounded by the number of competing variables influencing observed outcome, disturbed attachment patterns have been consistently documented in a range of samples of young children. On the other hand, it is not yet clear what percentage of these children might meet criteria for an attachment disorder at any given developmental stage.

COMPARATIVE NOSOLOGY

The criteria for reactive attachment disorder have evolved markedly since the diagnosis was first introduced in DSM-III. This early version of the disorder included growth failure and lack of social responsivity as central features. The diagnosis had to start by 8 months of age (the age at which preferred attachment to a restricted set of caregivers is usually just beginning to be evident) and could not result from a diagnosable medical condition. Gross neglect of the infant's physical and emotional needs had to be evident, and, as with later criteria, presence of autism or mental retardation precluded the diagnosis. Pertinent behaviors cited in the criteria included poor tone, weak cry, excessive sleep, lack of interest in the environment, and weak rooting and grasping when feeding.

Though so-called nonorganic failure to thrive remains an important diagnostic entity in infant mental health, the restrictive age range of this set of criteria and the requirement for growth failure made the DSM-III definition difficult to apply in many cases. Further, it is rarely possible to rule out autistic disorder or mental retardation in severely malnourished infants under 1 year of age. The link between failure to thrive and reactive attachment disorder was dropped in the revised third edition of DSM (DSM-III-R), and the age of onset was changed to the first 5 years. The two subtypes of the disorder, inhibited and disinhibited, were introduced with DSM-III-R and have persisted in DSM-IV. The etiological link between the disorder and evidence of “pathogenic care” at the hands of the young child's primary caregivers and the exclusion of children whose symptoms might be accounted for by cognitive delay or one of the pervasive developmental disorders remained an emphasis of the criteria.

Though reactive attachment disorder was not included in the ninth revision of *International Statistical Classification of Diseases* (ICD-9), it does appear in ICD-10—in a form largely consonant with DSM-IV criteria. The ICD-10 criteria do not explicitly link the disorder to pathogenic care, but a warning against making the diagnosis without evidence of abuse or neglect is included in the clinical description attached to the criteria. ICD-10 lists the two subtypes from DSM-IV as separate disorders (Table 44.1-2), and neither explicitly excludes children with mental retardation or pervasive developmental disorders. However, the clinician is required to document that the child shows “elements of normal social relatedness” with “non-deviant adults.”

Table 44.1-2 ICD-10 Diagnostic Criteria for Disorders of Social Functioning With Onset Specific to Childhood or Adolescence

DSM-IV and ICD-10 criteria for attachment disorders have been criticized in the literature on a number of fronts. First, the implication that the disorder is reactive is problematic. A temporal association of factors like maltreatment and inhibition should not imply a causative link. From a practical perspective, one cannot always know what experiences children might have had with their caregivers, and the limits of what is or is not pathogenic care are unclear. Some children might meet criteria for reactive attachment disorder without having experienced extreme deprivation or abuse. Second, the exclusion of infants and children with pervasive developmental disorders and mental retardation ostensibly on grounds that each has a different cause suggests that these disorders cannot coexist. Certainly, children with a mental age below 8 months are unlikely to develop a focused set of attachments with caregivers. However, in children with milder cognitive impairment, the clinical distinction between these disorders is occasionally difficult, particularly when caregiving deficiencies are notable.

The criteria emphasize the general social behavior of affected children across relationships rather than more specific attachment-related behaviors. These latter behaviors might include comfort seeking, compliance with caregiver requests, exploratory behavior, and attempts to control the caregiver either by acting in a caregiving role or by being bossy and punitive. Responses to reunion, used successfully to classify patterns of attachment, may also reveal extreme affective responses (e.g., ignoring, fearful, angry, or affectless reunions) that are clinically meaningful.

A final concern is raised by the clear evidence that children act differently in different relationships. Both DSM-IV and ICD-10 criteria require that the aberrant behavior be evident across relationships, which excludes children whose behavior is clearly compromised in the presence of their primary caregivers but not in the few other relationships they might have. Changing this focus would compel the clinician to identify a relationship disorder rather than a pervasive pattern of abnormal

social behavior, which might identify more children and families requiring intervention. This approach, however, flies in the face of current conceptualizations of psychiatric disorders—namely that they exist within individuals, not between them.

These concerns about DSM-IV and ICD-10 criteria have led a group of researchers and clinicians to propose a series of alternate criteria designed to redress most of these concerns. Criteria have been proposed for six separate attachment disorders, beginning with disorders of nonattachment that can be diagnosed in any child with a mental age of 10 months or above (i.e., the age at which children typically show preferred attachments). These disorders follow DSM-IV criteria in including inhibited and disinhibited subtypes. However, the link to pathogenic care is dropped and evidence of less extreme social behavior in relationships with caregivers other than the primary caregiver does not disqualify the child for the diagnosis. A second general type of disorder is also proposed, disrupted attachment disorder, which applies when a child experiences the sudden loss of the caregiver; it includes behaviorally anchored criteria consistent with the descriptions Bowlby and others originally documented. The third general type of disorder, so-called secure base distortions, arises from clinical work with young children. Three types have been identified: attachment disorder with self-endangerment, attachment disorder with inhibition, and attachment disorder with role reversal. In each of these disorders, the child has a preferred attachment figure, but the relationship with this figure is markedly distorted in one of these three ways.

Unfortunately, none of the available criteria, including those from DSM-IV and ICD-10, have been subjected to research aimed at validation and reactive attachment disorder remains a little-used and poorly studied phenomenon.

EPIDEMIOLOGY

Virtually no data exist on the prevalence and clinical course of reactive attachment disorder. Retrospective chart reviews of consecutive patients suggest that the diagnosis can be made reliably in a clinic-referred population. Because pathogenic care is often associated with broad-risk factors like poverty, family disruption and low social support, in clinical samples, these may be considered frequent conditions associated with the diagnosis. However, reactive attachment disorder appears to exist across socioeconomic strata.

ETIOLOGY

While reactive attachment disorder typically occurs in the context of grossly neglectful or overtly abusive care, the critical elements of this care and their relation to the onset of the disorder are unclear. Certainly children institutionalized in the same facility or siblings raised together in a markedly disturbed family may have divergent outcomes. Individual temperamental or personality factors may combine with corrective experiences to yield a nonpathological outcome in some children who experience extremes of care. The individual fit between neglectful or abusive adults and their infants or children may be critical in determining which children develop symptoms consistent with an attachment disorder.

Though research on what causes caregivers to be abusive or neglectful is limited, caregiver characteristics from severe personality disturbance to psychopathology appear to be important. Parental mental retardation and poor basic parenting skills, particularly in the context of little social support, may also lead to pathogenic care. Infants who are cared for by multiple caregivers in succession, as is common in the foster care system in the United States, are also at increased risk for attachment disorders. The traumatic experience of prolonged separation from a caregiver in early childhood, documented by Spitz, the Robertsons and Bowlby, is itself enough to cause attachment disturbance. Repeated prolonged separations typically cause more severe symptoms, and when these separations are accompanied by abuse, neglect, or both, attachment disorders appear to be particularly likely.

DIAGNOSIS AND CLINICAL FEATURES

The diagnosis of reactive attachment disorder is contingent upon documenting clear evidence of pervasive disturbance in social relatedness that began prior to age 5 (Table 44.1-3). The pattern of behavior should fit one of the two different subtypes described in the criteria. The inhibited subtype is characterized by hypervigilant and fearful behavior (often recognized as a pattern of compulsive compliance with the abusive caregiver) or extreme ambivalence and contradictory behavior in relationships. Mixtures of approach and avoidance may be apparent, and frozen watchfulness may be noted. The disinhibited subtype is marked by a lack of selectivity in choosing social partners, resulting in diffuse attachments and a peculiar overfriendliness that has been labeled indiscriminate sociability. Caregivers may remark on their own subjective sense that the child is not truly attached to them.

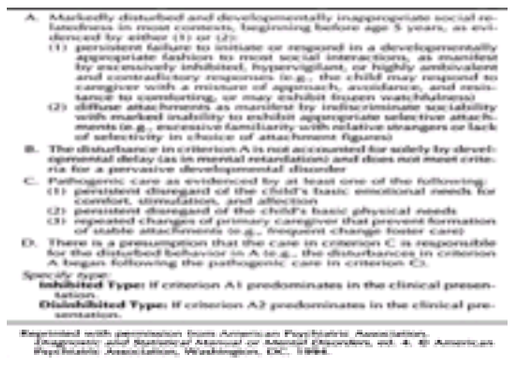

A. Markedly disturbed and developmentally inappropriate social relatedness in most contexts, beginning before age 5 years, as evidenced by either (1) or (2):
   (1) persistent failure to initiate or respond in a developmentally appropriate fashion to most social interactions, as manifest by excessively inhibited, hypervigilant, or highly ambivalent and contradictory responses (e.g., the child may respond to caregiver with a mixture of approach, avoidance, and resistance to comforting, or may exhibit frozen watchfulness)
   (2) diffuse attachments as manifest by indiscriminate sociability with marked inability to exhibit appropriate selective attachments (e.g., excessive familiarity with relative strangers or lack of selectivity in choice of attachment figures)
B. The disturbance in criterion A is not accounted for solely by developmental delay (as in mental retardation) and does not meet criteria for a pervasive developmental disorder
C. Pathogenic care as evidenced by at least one of the following:
   (1) persistent disregard of the child's basic emotional needs for comfort, stimulation, and affection
   (2) persistent disregard of the child's basic physical needs
   (3) repeated changes of primary caregiver that prevent formation of stable attachments (e.g., frequent change foster care)
D. There is a presumption that the care in criterion C is responsible for the disturbed behavior in A (e.g., the disturbances in criterion A began following the pathogenic care in criterion C).

Specify type:
Inhibited Type: If criterion A1 predominates in the clinical presentation.
Disinhibited Type: If criterion A2 predominates in the clinical presentation.

Reprinted with permission from American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, ed. 4. © American Psychiatric Association, Washington, DC, 1994.

Table 44.1-3 DSM-IV Diagnostic Criteria for Reactive Attachment Disorder of Infancy or Early Childhood

It is difficult to make the diagnosis of reactive attachment disorder without a multisession evaluation, particularly when the caregiver's behavior is not known to be abusive or the child's history does not include multiple placements. Ascertaining whether the child has less-deviant behavior in the presence of the clinician or another trusted adult is important and may take time. The child's relationship with the primary caregiver can be adequately assessed with a combination of free play, structured teaching tasks that stress the dyad, and a brief separation and reunion. Videotaping the assessment is a useful way to document social responses, and the introduction of a clinically trained person whom neither partner has met is often informative. Children who meet criteria for this disorder may look quite different depending on their developmental stage.

A 26-month-old girl, recently placed in foster care, was referred by state child protective services with her biological and foster families to assist with long-term case management. Her history included two admissions for failure to thrive in the first year of life and a third admission at 13 months that revealed retinal hemorrhage and a subdural hematoma from suspected shaken baby syndrome. No perpetrator was conclusively identified. When seen with her biological mother in a comfortable, toy-filled room, she stood completely still and maintained little facial expression. She complied completely and in rote fashion with her mother's often angry instructions, maintaining no sustained eye contact with either her mother or the examiner. When briefly separated from her mother, she showed little reaction, looking up briefly with an odd grimace when her mother returned to the room. Her mother confirmed that her behavior had been similar when she had lived in her home; the child spoke infrequently and rarely sought comfort when distressed. When seen with her foster mother of 3 months, she was markedly more animated, though frequently irritable. She engaged in play freely and referenced both her foster mother and the examiner during play. She stopped playing and stared blankly when separated from her foster mother, though she actively reengaged her foster mother upon her return. The biological mother's parental rights were eventually terminated and though the child was placed in two more homes, she showed the capacity to engage with her new caregivers each time. The girl was diagnosed with reactive attachment disorder, inhibited type.

A 6-year-old boy was referred by his adoptive parents because of hyperactivity and disruptive behavior at school. He had been adopted at age 5, after living most of his life in a Romanian orphanage where he received care from a rotating shift of caregivers. Though he had been below the fifth percentile for height and weight upon arrival, he quickly approached the tenth percentile in his new home. However, both his adoptive parents were frustrated by their inability to "reach him." They had initially worried about a hearing disturbance, though testing and his capacity to engage many adults and children verbally suggested otherwise. He showed interest in anyone and would often follow strangers willingly. He showed little empathy when others were hurt and blandly resisted redirection in school. He was frequently injured because of seemingly reckless behavior, though he had an extremely high tolerance for pain. Intensive intervention focused on problem behaviors at home decreased his self-endangering behavior, though he remained oddly overfriendly and unempathic both at home and in school. The boy was diagnosed with reactive attachment disorder, disinhibited type.

DIFFERENTIAL DIAGNOSIS

Children with marked disinterest in social interaction altogether may require cognitive testing for developmental delay or a neurological workup. Evidence of stereotypies, grossly restricted range of interests, and poor response to changes in routines suggest the spectrum of pervasive developmental disorders; cognitive impairment is frequently moderate to severe in these children. Interviews with the caregivers of these children usually yield little evidence of inappropriate care, and consistent blunting of social interactions outside the family unit is common. However, many children with pervasive developmental disorders or mental retardation form secure attachments with their primary caregivers despite an overall restriction in the range of attachment-related behaviors.

Children with severe receptive and expressive language delays may present linically with difficulty in social relatedness. Often the clinical picture includes externalizing behavior that heightens as unsuccessful attempts to communicate are made; rarely does this behavior fit the reactive attachment disorder subtypes, and grossly inappropriate care is uncommon in this population. As communication improves, social interactions improve.

Temporary or permanent loss of a primary caregiver to whom a young child is already attached is associated with significant disturbance. Bowlby described a common progression of reactions in this instance: protest, despair, and detachment. The clinical picture may mimic the inhibited form of reactive attachment disorder, though pathogenic care is not apparent in these cases. These children may quickly form new attachments to sensitive caregivers, particularly if they had some familiarity with the replacement caregiver before the loss. The clinician may diagnose depression in these cases, though the criteria for depression in infancy are not fully consonant with DSM-IV criteria for major depressive disorder.

The relationship between failure to thrive and reactive attachment disorder has not been well studied. If failure to thrive is not caused by a medical condition, then investigation of the child's relationships with the caregivers is necessary. In severe cases of neglect, reactive attachment disorder and failure to thrive may coexist. Prolonged lack of stimulation may lead to psychosocial dwarfism and many of these children would be expected to meet criteria for reactive attachment disorder. Worldwide, the interplay of kwashiorkor, lack of appropriate caregiving, and attachment disorders may be significant.

Comorbid psychiatric conditions appear common in older children whose history is consistent with reactive attachment disorder. Disorganized patterns of attachment behavior in the first 6 years appear to be associated with disruptive behavior disorders in later childhood; this may also be true for children with reactive attachment disorder.

COURSE AND PROGNOSIS

Because a sizable cohort of patients with reactive attachment disorder has never been followed, the typical course and prognosis for this disorder are not clear. It is likely that associated conditions from the nutritional and neurological sequelae of psychosocial deprivation to the stability of later relationships are critical in influencing the outcome. The possible range of outcomes include death in the most severe cases to relatively normal functioning with intervention leading to the establishment of healthy relationships. There have been several recent controlled studies of children adopted from international orphanages to Canada, England, and the United States. Despite the absence of a gold standard for diagnosing reactive attachment disorder and significant variation across samples (from age at adoption to conditions prior to adoption), these studies provide evidence that a subsample of adopted orphans have symptoms of indiscriminate sociability which continue for months or years after adoption. A smaller subsample appears to be at risk for externalizing disorders, and a handful of children from different samples have been reported to have chronic quasiautism. On the other hand, despite their histories of severe privation, it appears that a majority of these international adoptees rapidly assimilate after stable placement and have few ongoing psychiatric symptoms. Factors that are consistently related to these disparate outcomes have yet to be identified.

There are few long-term, controlled follow-up studies of children and adults who were raised primarily in institutional settings. The few available studies also suggest that indiscriminate sociability may persist and that the choice of marital partners in later life strongly influences eventual adult psychological functioning. The link between early disturbances of attachment and antisocial tendencies, particularly a lack of empathy, was first made by Bowlby; longitudinal studies of this association are needed to firmly establish this pathway.

TREATMENT

The first consideration in the assessment of children exposed to grossly inadequate care is the child's safety. Child maltreatment is associated with significant morbidity, and mortality is not uncommon, particularly in children under 24 months of age. Early involvement of child protective services is often warranted, and assessment of parental fitness may be a necessary part of the evaluation. In some cases placement of the child may be necessary, and reunification of parent and child may not be warranted. Children in this situation have often not had appropriate medical care, and medical referral is almost always indicated. Unfortunately, the foster care system and the family court system may lead to multiple placements in the first years of life and thus increase the likelihood of an attachment disorder. The clinician may play a crucial role in staying involved with the child who is placed in foster care, providing expert testimony in court and individual or family treatment.

Once the child is in a relatively stable placement and is medically healthy, full attention can be paid to psychosocial intervention, which must often be tailored to the individual. Possible interventions include individual psychotherapy for the child or caregiver; parent training with emphasis on developmental expectations; family therapy; or caregiver-child dyadic therapy, which is perhaps most specifically directed toward disturbances of attachment and in many instances is the treatment of choice. As described by Alicia Lieberman, this approach weaves together developmental training and guidance with an active attempt to address pressing instrumental issues (e.g., poor housing, inadequate medical care) and insight-oriented psychotherapy with the child present. The complexity of this approach reflects the complexity of the clinical problem. Long-term interventions are necessary in these cases, and psychiatric treatment should be bolstered by early intervention programs and ongoing medical care for the child.

SUGGESTED CROSS-REFERENCES

The role of early experience in development is discussed in different contexts in Section 4.1 on anthropology and psychiatry, Section 5.4 on animal research and its relevance to psychiatry, and Section 6.1 on psychoanalysis. The role of early experience in personality formation is discussed in Chapter 24 on personality disorders. Neuropsychological and intellectual assessment of children is discussed in Section 7.5. Mood disorders are presented in Chapter 14. Feeding and eating disorders of infancy or early childhood are discussed in Chapter 41. Aspects of physical and sexual abuse of children and neglect are reviewed in Section 49.4. Foster care is discussed in Section 49.3.

SECTION REFERENCES

Ainsworth MDS, Blehar MS, Waters E, Wall S: *Patterns of Attachment: A Psychological Study of the Strange Situation*. Erlbaum, Hillsdale, NJ, 1985.

Boris NW, Fueyo MA, Zeanah CH: The clinical assessment of attachment in children less than five. J Am Acad Child Adolesc Psychiatry *36:*295, 1997.

Boris NW, Zeanah CH: Clinical disturbances and disorders of attachment in infancy and early childhood. Curr Opin Pediatr *10:*365, 1998.

*Boris NW, Zeanah CH: Disorders and disturbances of attachment in infancy: An overview. Infant Ment Health J *20:*1, 1999.

Boris NW, Zeanah CH, Larrieu JA, Scheeringa MS, Heller SS: Reactive attachment disorder of infancy and early childhood: A preliminary investigation of diagnostic criteria. Am J Psychiatry *155:*295, 1998.

Bowlby J: *Maternal Care and Child Health*. World Health Organization, Geneva, 1951.

*Bowlby J: *Attachment and Loss*, ed 2, vols 1–3. Basic Books, New York, 1982.

Carlson EA: A prospective longitudinal study of attachment disorganization/disorientation. Child Dev *69:*1107, 1998.

Chapin HD: Are institutions for infants necessary? JAMA *64:*1, 1915.

Chisholm K: A three-year follow-up of attachment and indiscriminate friendliness in children adopted from Romanian orphanages. Child Dev *69:*1092, 1998.

Emde RN, Sameroff AJ: Understanding early relationship disturbances. In *Relationship Disturbances in Early Childhood*, AJ Sameroff, RN Emde, editors. Basic Books, New York, 1989.

Greenspan SI, Lieberman A: A clinical approach to attachment. In *Clinical Implications of Attachment*, J Belsky, T Nezworski, editors. Erlbaum, Hillsdale, NJ, 1988.

Guedeney A: Kwashiorkor, depression and attachment disorders. Lancet *346:*1293, 1995.

*Hinshaw-Fusilier S, Boris NW, Zeanah CH: Reactive attachment disorder in maltreated twins. Infant Ment Health J *20:*42, 1999.

Hodges J, Tizard B: Social and family relationships of ex-institutional adolescents. J Child Psychol Psychiatry *30:*77, 1989.

Karen R: *Becoming Attached*. Warner, New York, 1994.

Kempe CH, Silverman FN, Steele BF: The battered-child syndrome. JAMA *181:*17, 1962.

Lieberman AF, Pawl JH: Disorders of attachment and secure base behavior in the second year of life: Conceptual issues and clinical intervention. In *Attachment in the Preschool Years*, MT Greenberg, D Cicchetti, EM Cummings, editors. University of Chicago Press, Chicago, 1990.

Lieberman AF, Pawl JH: Infant-parent psychotherapy. In *Handbook of Infant Mental Health*, CH Zeanah, editor. Guilford, New York, 1993.

Lieberman AF, Weston D, Pawl JH: Preventive intervention and outcome with anxiously attached dyads. Child Dev *62:*199, 1991.

Lieberman AF, Zeanah CH: Disorders of attachment in infancy. Child Adolesc Psychiatry Clin North Am *4:*571, 1995.

Main M, Kaplan N, Cassidy J: Security in infancy, childhood, and adulthood: A move to the level of representation. Monogr Soc Res Child Dev *50:*66, 1985.

*O'Connor TG, Bredenkamp D, Rutter M, the English and Romanian Adoptees (ERA) Study Team: Attachment disturbances and disorders in children exposed to early severe deprivation. Infant Ment Health J *20:*10, 1999.

*Provence S, Lipton R: *Infants in Institutions*. International Universities Press, New York, 1962.

Richters MM, Volkmar FR: Reactive attachment disorder of infancy or early childhood. J Am Acad Child Adolesc Psychiatry *33:*328, 1994.

Robertson J, Robertson J: *Separation and the Very Young*. Free Association Books, London, 1989.

Rutter M: Maternal deprivation, 1972–1978: New findings, new concepts, new approaches. Child Dev *50:*283, 1979.

Skeels HM: Adult status of children with contrasting early life experiences. Monogr Soc Res Child Dev *31:*1, 1966.

Spitz R: Anaclitic depression. Psychoanal Study Child *2:*313, 1946.

Tizard B, Hodges J: The effect of early institutional rearing on the development of eight-year-old children. J Child Psychol Psychiatry Allied Discip *19:*99, 1978.

Tizard B, Rees J: A comparison of the effects of adoption, restoration to the natural mother, and continued institutionalisation on the cognitive development of four-year-old children. J Child Psychol Psychiatry Allied Discip *15:*61, 1974.

*Tizard B, Rees J: The effect of early institutional rearing on the behaviour problems and affectional relationships of four-year-old children. J Child Psychol Psychiatry Allied Discip *16:*61, 1975.

*Zeanah CH: Beyond insecurity: A reconceptualization of attachment disorders in infancy. J Consult Clin Psychol *64:*42, 1996.

Zeanah CH, Emde RN: Attachment disorders in infancy and childhood. In *Child and Adolescent Psychiatry: Modern Approaches*, M Rutter, L Hersov, E Taylor, editors. Blackwell, Oxford, 1994.

Zeanah CH, Mammen OK, Lieberman AF: Disorders of attachment. In *Handbook of Infant Mental Health*, CH Zeanah, editor. Guilford, New York, 1993.

44.2 STEREOTYPIC MOVEMENT DISORDER OF INFANCY AND DISODERS OF INFANCY AND EARLY CHILDHOOD NOT OTHERWISE SPECIFIED

JOAN L. LUBY, M.D.

STEREOTYPIC MOVEMENT DISORDER OF INFANCY AND EARLY CHILDHOOD

Both the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) contain a category that includes a broad variety of etiologically unrelated stereotyped movements that do not form a part of any other recognized psychiatric or neurological condition.

Definition and Comparative Nosology The stereotypic movement disorder is described in DSM-IV as “repetitive, seemingly driven, and nonfunctional motor behavior” that “interferes with normal activities or results in self-inflicted bodily injury that requires medical treatment.” These movements may include hand waving, head banging, rocking, self-biting, picking at skin or bodily orifices, or hitting various body parts. ICD-10 contains a basically identical category called stereotyped movement disorders. In both systems, the diagnosis is given when the behavior is present for 4 weeks or longer. When the behavior is self-injurious this feature is specified; the ICD further elaborates on non-self-injurious, self-injurious, and mixed types. When the movement occurs as a part of another general neurological, medical, or psychiatric condition (e.g., pervasive developmental disorder not otherwise specified or autism) a comorbid diagnosis of stereotypic movement disorder is not deemed necessary. Lesh-Nyhan syndrome, the genetic deficiency of hypoxanthine-quinine-ribosyltransferase activity associated with choreiform movements and severe self-mutilating behavior, is perhaps the most dramatic example of a stereotyped movement that does not fall into this classification. One exception to this rule is that stereotyped movements that occur in conjunction with mental retardation warrant a separate diagnosis when the stereotypic or self-injurious behavior is severe enough to require treatment.

History Stereotypic movements were first identified as an important psychiatric symptom in the early 1900s by Emil Kraepelin, who described stereotypies as a characteristic of dementia precox. These movements have since become widely recognized as symptoms associated with psychotic disorders in general and with a number of neurodevelopmental disorders.

Stereotypic movement disorder has undergone some modification with the evolution of the DSM system. The third edition of DSM (DSM-III) contained a broader definition of the disorder than DSM-IV contains. The DSM-III criterion describes voluntary, nondistressing, nonspasmodic movements and goes on to specify that the behaviors were more likely to be associated with extreme psychosocial deprivation, mental retardation, or pervasive developmental disorder but could occur independent of these conditions. Stereotypy/habit disorder as it was called in the revised third edition of DSM (DSM-III-R), represented a broadened diagnostic category that included periodic and persistent nonfunctional behaviors such as nail biting. Also new in the DSM-III-R was the inclusion of functional impairment in the diagnostic criteria; the behavior had to either cause physical injury or interfere with normal activities. DSM-III-R also added exclusions; the diagnosis could not be made if the movements occurred in the presence of a tic disorder or a pervasive developmental disorder. In addition to new criteria described above, DSM-IV also excludes specific behaviors such as hair pulling (trichotillomania), tics, and compulsions since they are thought to fit more coherently into other diagnostic categories.

Epidemiology The prevalence of stereotyped movement disorders as a whole in the general population remains unknown; however, some data exist on the prevalence of individual stereotyped behaviors. The prevalence of self-injurious behaviors among individuals with mental retardation has been estimated to be between 2 and 3 percent in community samples of children and 25 percent among mentally retarded adults living in institutions. Head banging has been estimated to have a prevalence of approximately 5 percent in child populations, with males affected three times more frequently than females. The behavior is usually self-limiting, occurring in the first 3 years of life, with typical age of onset between 5 and 11 months. It is relatively rare after age 3 but persists longer in about 5 percent of cases. Similar statistics are reported for head rolling, a behavior that can result in significant hair loss. Similarly, breath holding is also rare after the third birthday and occurs most commonly between 12 and 18 months of age. This behavior is also more common among boys than girls. Bruxism (grinding the teeth) is estimated to occur in as many as 56 percent of the normal population in infancy, after eruption of the teeth.

Etiology Although many theories (behavioral, neurobiological, genetic, and analytic) have been proposed to explain the pathogenesis of various stereotyped movements, no clear understanding of the cause of the disorder has emerged. A number of genetic, neurological, and metabolic disorders are associated with stereotyped movements, and there is a much higher prevalence of these behaviors in individuals with mental retardation. In addition, twin studies have shown that a few stereotypic movement disorders (e.g., nail biting) have a familial inheritance pattern. For example, monozygotic twins had a 66 percent concordance for nail biting while dizygotic twins had a rate of 34 percent.

Stereotypic movements can be induced or exacerbated by a number of sympathomimetic pharmacological agents (e.g., amphetamine, cocaine). Similarly, both stereotypic movements and tics occur in children taking sympathomimetics (e.g., methylphenidate [Ritalin], dextroamphetamine [Dexedrine], pemoline [Cylert]). These movements generally disappear or diminish when the drug is withdrawn. Recent studies have also suggested that these behaviors have a rhythmic component.

Diagnosis and Clinical Features Although stereotypic movements may take a wide variety of forms, the most commonly observed forms include nail biting and onychotillomania (nail or cuticle picking), nose picking, hair pulling, head banging, head rolling, skin picking, poking or picking at other bodily orifices, bruxism, and breath holding. The behavior is repetitive and volitional. In younger children it may be more obvious on clinical examination, since they have less desire to suppress it; in older children it is more likely to be disclosed by a parent in the psychiatric history. In primary care settings, the diagnosis may be suspected when there is evidence of tissue trauma or damage or other physical signs such as anemia from chronic blood loss due to a self-destructive injury in extreme cases. No biological or physiological markers for the disorder are known, and thus no laboratory tests. Perhaps the most challenging aspect of the differential diagnosis is determining whether the movements are simply a part of another underlying disorder such as mental retardation, pervasive developmental disorder not otherwise specified, obsessive-compulsive disorder, or an underlying tic or neurological movement disorder. The DSM-IV diagnostic criteria for stereotypic movement disorder are listed in Table 44.2-1. Table 44.2-2 presents the ICD-10 diagnostic criteria for stereotyped movement disorders.

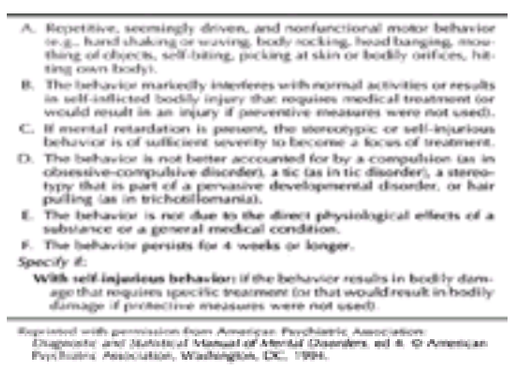

A. Repetitive, seemingly driven, and nonfunctional motor behavior (e.g., hand shaking or waving, body rocking, head banging, mouthing of objects, self-biting, picking at skin or bodily orifices, hitting own body).
B. The behavior markedly interferes with normal activities or results in self-inflicted bodily injury that requires medical treatment (or would result in an injury if preventive measures were not used).
C. If mental retardation is present, the stereotypic or self-injurious behavior is of sufficient severity to become a focus of treatment.
D. The behavior is not better accounted for by a compulsion (as in obsessive-compulsive disorder), a tic (as in tic disorder), a stereotypy that is part of a pervasive developmental disorder, or hair pulling (as in trichotillomania).
E. The behavior is not due to the direct physiological effects of a substance or a general medical condition.
F. The behavior persists for 4 weeks or longer.

Specify if:

With self-injurious behavior: if the behavior results in bodily damage that requires specific treatment (or that would result in bodily damage if protective measures were not used).

Table 44.2-1 DSM-IV Diagnostic Criteria for Stereotypic Movement Disorder

A. The child exhibits stereotyped movements to an extent that either causes physical injury or markedly interferes with normal activities.
B. Duration of the disorder is at least 1 month.
C. The child exhibits no other mental or behavioral disorder in the ICD-10 classification (other than mental retardation).

Reprinted with permission from World Health Organization: *The ICD 10 Classification of Mental and Behavioural Disorders: Diagnostic Criteria for Research*, © World Health Organization, Geneva, 1993.

Table 44.2-2 ICD-10 Diagnostic Criteria for Stereotyped Movement Disorders

A white, 4½-year-old boy was brought to the clinic by his parents because of their concern about his persistent nail and cuticle biting that resulted in bleeding and tissue damage and his picking at scabs and moles on various parts of his body led to poor healing and recurrent infection. Parents described the behavior as ongoing for the past 4 to 6 months. They had hoped it was simply a bad habit and that the child would spontaneously "grow out of it." A comprehensive mental health evaluation revealed that in addition to the chief complaint the child also suffered from mild delays in expressive language as well as delays in gross motor control. On examination, the nails had been bitten down to the quick, and several cuticles appeared scabbed. During the 45-minute play observation, the child was noted to pick at moles and scales on his arms and upper body during his play. No other associated psychiatric or medical symptoms were noted. Family psychiatric history was negative for any major psychiatric disorder; however the mother was both a smoker and a nail biter herself. The diagnosis of stereotypic movement disorder was made, and a behavioral intervention program for the nail and cuticle biting was implemented. The child was also referred for a speech and language and occupational therapy evaluation. Given the tissue damage that had resulted from picking of moles, the primary care physician decided (in consultation with the child psychiatrist) to have the moles removed. The need to use large, impenetrable bandages to facilitate postsurgical healing was emphasized, and the procedure was successful.

Precautions Given the opportunity for early classification that the DSM-IV disorder of infancy and early childhood not otherwise specified category provides and the inherent lack of formal restrictions in using this classification, the responsibility to exercise caution in its application to very young children rests in the hands of the individual clinician. This is particularly important given the potentially detrimental social effects of applying psychiatric diagnoses or classifications to children. The infant-preschool clinician faces the additional unique challenge of distinguishing clinically significant phenomena from normal developmental vicissitudes known to appear so commonly in early childhood. Magda Campbell and others have expressed concern about the difficulty in distinguishing such developmental problems from clinically significant psychopathology in very young children. These investigators emphasize the importance of attention to symptom severity and the co-occurrence of other symptoms in determining clinical significance. A tangible example of the ambiguity in distinguishing developmental and clinical problems is seen when evaluating defiance or oppositionality in the preschool child. Here, the normative developmental move toward autonomy may alone produce these difficult behaviors. In these situations, the clinician must judge both the appropriateness of the giver's perceptions of and tolerance for the child's behaviors and the degree to which the behaviors deviate from developmental norms and occur in conjunction with other symptoms. The availability of the not otherwise specified category, with its inherent lack of constraints, makes this distinction particularly important to prevent over diagnosis in infants and preschoolers with only normative, transient developmental problems.

Course and Prognosis Stereotypic movements that occur in association with other psychiatric or neurological disorders generally follow the course of the core disorder. For example, stereotypic movements associated with a pervasive developmental disorder not otherwise specified are generally more severe when there is greater impairment from the underlying disorder and less severe when the underlying disorder is more controlled. Stereotypic movement disorder can occur at any age, but prevalence peaks in adolescence and declines in older age groups. Although the specific behaviors may change over time, the underlying disorder is generally quite stable.

Treatment No specific treatment has been shown to be effective for stereotypic movement disorder in general; however, a small number of double-blind studies have investigated the efficacy of pharmacological treatments for specific behaviors. For example, investigations have suggested the superiority of clomipramine (Anafranil) over desipramine (Norpramin) for severe nail biting. Although a wide variety of pharmacological agents have been used for stereotyped movements, with varying success, the standard neurolepics are perhaps most frequently used and demonstrate most clinical benefit. Both haloperidol (Haldol) and chlorpromazine (Thorazine) have demonstrated efficacy in placebo-controlled studies of patients with stereotyped movements associated with mental retardation and autistic disorder.

Behavioral modification techniques have been perhaps the most widely used interventions for the treatment of stereotypic movements. These techniques, which include both positive and negative reinforcement, have shown some success in diminishing the severity and frequency of the movements and in some cases extinguishing the disorder.

DISORDERS OF INFANCY AND EARLY CHILDHOOD NOT OTHERWISE SPECIFIED

The ongoing evolution of the DSM and International Classification of Diseases (ICD) is based on the recognized need to modify diagnostic criteria and add new diagnostic categories with advances in the understanding of psychiatric disorders. In keeping with this DSM-IV has a category disorder of infancy, childhood, or adolescence not otherwise specified (Table 44.2-3) so that clinically significant disturbances in children that do not meet criteria for specific, established disorders can be coded on Axis I. This is a residual category, intended for use only when the clinical characteristics cannot be adequately described by a more specific, established diagnosis.

This category is a residual category for disorders with onset in infancy, childhood, or adolescence that do not meet criteria for any specific disorder in the classification.

Reprinted with permission from American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorder*, ed 4. © American Psychiatric Association, Washington, DC, 1994.

Table 44.2-3 DSM-IV Diagnostic Criteria for Disorder of Infancy, Childhood, or Adolescence Not Otherwise Specified

The ICD-10 included two residual categories for childhood mental disorders: (1) other specified behavioral and emotional disorders with onset usually occurring in childhood and adolescence, and (2) unspecified behavioral and emotional disorders with onset usually occurring in childhood and adolescence. However, inclusion of these categories implies that there are likely to be clinical disorders or variations of disorders that are not yet specifically described in the DSM-IV. These categories are intended for use throughout childhood. Unfortunately, they may be most frequently used by clinicians working with the youngest group of children, infants and preschoolers, for whom DSM-IV and ICD-10 include very few specific diagnostic categories. The greater need for not-otherwise-specified diagnoses in the youngest portion of clinical infant and preschool samples, to whom a specific DSM-IV diagnosis could not be applied, has been described by infant-preschool clinicians.

A NEW CLASSIFICATION FOR INFANTS AND PRESCHOOLERS

Over the last decade, advances in research on the socioemotional development and developmental psychopathology of infants and preschoolers and the clinical

application of these findings have contributed increased recognition and treatment of very early onset behavioral and emotional disturbances. This progress along with public health emphasis on early and potentially preventive intervention has catalyzed the growth of infant and preschool mental health into an increasingly practiced clinical subspecialty of child psychiatry. Although infant and preschool patients and their caregivers are seen in general child psychiatric clinics, increasing numbers of infant/preschool specialty clinics have been established across the nation. This led to the establishment of specific practice parameters for the clinical assessment of this age group. Along with this increase in clinical attention has come progress in developing a classification system specifically for these very young patients. The *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood* (DC:0-3), developed by a task force of the ZERO TO THREE/National Center for Clinical Infant Programs (NCCIP), represents the first diagnostic manual designed for children from birth through age 4. The system developed from the concern of a large group of experienced infant-preschool clinicians who agreed that DSM-IV and ICD-10 systems did not provide adequate descriptive criteria for this age group.

The DC:0-3 outlines a number of new diagnostic categories and developmentally modified DSM-IV categories. The manual is modeled after, but not identical to, the DSM-IV multiaxial system (Table 44.2-4). It is intended to augment and complement DSM-IV and ICD-10, not replace them. As in DSM-IV, clinical disorders ("primary diagnoses") are listed on Axis I. Physical, neurological, developmental, and mental disorders or conditions described in other classification systems (e.g., DSM-IV, ICD) are listed on Axis III. However, when a DSM-IV or ICD diagnosis is determined to best describe the clinical picture, these standard diagnoses are listed on Axis I. Also as in the DSM-IV, Axis IV is used to rate psychosocial stressors determined to affect the mental or developmental condition. Axis V rates the functional emotional developmental level. Given the developmental context of DC:0-3, this refers to the child's capacity for organizing "affective, interactive, communicative, cognitive, motor and sensory experience." In this sense it is similar to the DSM-IV Global Assessment of Functioning (GAF) Scale but more specific to emotional competence.

Axis I:	Primary classification
Axis II:	Relationship classification
Axis III:	Physical, neurological, developmental, and mental health disorders or conditions (described in other classification systems)
Axis IV:	Psychosocial stress
Axis V:	Functional emotional developmental level

Table 44.2-4 DC:0-3 Classification System

Relationship Classifications Unique to the DC:0-3 axial system and perhaps its most important contribution is the use of Axis II to describe classify relationships. Given the central importance of the primary caretaking relationship to the emotional development and mental status of the very young child, this axis is used in all cases to describe the quality of this primary relationship and (when indicated) to designate a relationship disorder. The Parent-Infant Relationship Global Assessment Scale (PIR-GAS) was developed to rate the quality of the relationship. The PIR-GAS ranges from 90 (well adapted) to 10 (grossly impaired). The clinical diagnosis of relationship disorders is made only when scores are below 39, while scores between 40 and 79 constitute "trend" or "tendency" levels of impairment similar to "trait" level personality disorders in DSM-IV Axis II. The PIR-GAS also constitutes a variant of the GAF Scale but is specific to the primary relationship rather than the individual. A number of specific relationship disorders described in the DC: 0-3 are outlined in Table 44.2-5.

901.	Overinvolved relationship
902.	Underinvolved relationship
903.	Anxious/tense relationship
904.	Angry-hostile relationship
905.	Mixed relationship
906.	Abusive relationships

Table 44.2-5 DC:0-3 Relationship Disorder Classifications

The use of Axis II for relationship disorders in infants and toddlers reflects the central importance of early caregiving relationships and environments as primary factors in the early manifestation of psychopathology. This theoretical model has been supported by both developmental child research and mental health research over the last two decades. At very young ages and in some circumstances (e.g., parental abuse or extreme neglect) the primary caregiving relationship is thought to be the unit of psychopathology rather than the individual child, particularly with children under 3 years of age, who remain fundamentally dependent on, and as yet undifferentiated from, their caregivers. Thus, early child assessment should involve assessing the child in interaction with the primary caregivers independent of the classification system used. Particular attention has been paid to this issue in the clinical assessment of attachment.

Regulatory Disorders While a comprehensive review of all diagnoses included in the DC:0-3 is beyond the scope of this chapter, several novel and developmentally modified diagnostic categories exist. The DC:0-3 regulatory disorders, specified by Stanley Greenspan in 1992, first described by Klaus Minde and Regina Minde in 1986 as "disorders of behavioral organization," have no corollary in DSM-IV. They represent the first attempt to categorize severe difficulties in sensorimotor organization or self-regulation in infants and preschoolers. Historically, children displaying these symptoms were likely described as extremely fussy or difficult temperamentally. An etiological basis for regulatory disorders is inherent in its definition: hyperreactivity or hyporeactivity to sensory stimuli occurring as a result of a core impairment in sensory motor processing and integration. Behavioral difficulties and problems in adaptation must be coupled with evidence of sensorimotor processing difficulties for the diagnosis to be made. Some physiological differences (i.e., changes in vagal tone) were noted in groups of children meeting criteria for the disorder, and evidence of later difficulty in school adaptation was found among a small group of children followed up after 4 years.

A 3-year and 2-month-old girl was brought to the infant-preschool clinic by her mother who was concerned about her extreme and chronic irritability, fussiness, and difficulty adapting to environmental change. She was described as a very difficult infant, the product of an uncomplicated full-term pregnancy, who was slow to develop a stable eating and sleeping cycle and very difficult to soothe. The mother was finding it increasingly difficult to leave her with alternate caregivers (i.e., babysitters) and was becoming concerned about her ability to make the transition to preschool. Further probing revealed that the child was extremely rigid in a number of areas that the family had unwittingly complied with. The child had a very small repertoire of food that she would eat, which appeared to be largely limited by texture. In addition, she refused to eat anything that was either too warm or too cold; all foods had to be at room temperature. The child was also extremely sensitive to tactile stimuli, tolerating only cotton fabrics and preferring not to wear clothing at all. She often comforted herself by stroking a fur blanket, which she seemed to find very soothing. Strikingly, she was drawn to articles of clothing of only a few colors. Family vacations were nearly impossible as the child would become extremely irritable and have tantrums when faced with the need to sleep in an unfamiliar bed, on unfamiliar sheets. She openly expressed her discomfort with the unusual smell of the new setting and was unable to enjoy herself, appearing distressed and whining throughout the trip. The child's verbal skills were within normal limits for her chronological age, and no core impairment in interpersonal relatedness was noted after a multiple session dyadic evaluation with both parents, at which time the diagnosis of a regulatory disorder was made. Previous clinicians had suggested the diagnosis of Asperger's disorder; however, her parents (who were quite savvy consumers) felt that she was simply a child with an extremely difficult temperament.

Mixed Disorder of Emotional Expressiveness In the area of affective disorders, the DC:0-3 includes another novel diagnosis, mixed disorder of emotional expressiveness. This disorder, unlike any mood disorder currently described in DSM-IV or ICD systems, identifies children who have an “ongoing difficulty expressing developmentally appropriate emotions.” This may be characterized either by the absence of the capacity to express discrete emotional states (e.g., shame, excitement, empathy) that would be expected for chronological age or by impairments in affect as evidenced by excessive inhibition or intensity of emotional reactivity. Although this category is based largely on clinical observations, it appears to be consistent with recent research findings on emotional regulation and affective development in infants and toddlers. This research supports the notion of affective development in infancy and early childhood and has helped identify and operationalize its multiple components. Normative developmental studies suggest that a core group of discrete emotions are present from birth and that others develop along a predictable trajectory. The development of both affective range and repertoire is thought to be a function of both constitutional and environmental factors. Although this new diagnostic category remains controversial (some forms could represent other nonmood disorders already included in the DSM-IV), it may be helpful to describe a subgroup of children who are not otherwise specified in DSM-IV.

In addition to novel diagnostic categories, the DC:0-3 also offers a reorganization that implicitly reconceptualizes some DSM-IV pervasive developmental disorders. DSM-IV includes autistic disorder, childhood disintegrative disorder, Asperger's disorder, Rett's disorder, and pervasive developmental disorder not otherwise specified in the broad category of pervasive developmental disorders. This categorization implies that these disorders are related, characterized by a core impairment in interpersonal relatedness, and associated with multiple developmental delays.

Disorders of Relating and Communicating In contrast to the DSM-IV organization, the DC:0-3 proposes a new general category of disorders of relating and communicating, which contains a new disorder, the multi-system developmental disorder. Multi-system developmental disorder can be described as a milder manifestation of traditional pervasive developmental disorder but is thought to result from a core impairment in sensory and motor processing that gives rise to deficits in social development. This new organization reflects the hypothesis that infants and preschoolers with milder impairments, who demonstrate a stronger capacity for interpersonal relatedness with primary caregivers, do not have a primary deficit in relatedness characteristic of autistic disorder. Rather, these more mildly affected children, who could also be classified in the currently very broad DSM-IV category pervasive developmental disorder not otherwise specified, are thought to have a primary deficit in sensorimotor processing and secondary impairments in relatedness that are not fixed and permanent but responsive to early therapeutic interventions. This new category is controversial as recent studies have demonstrated that early intervention in pervasive developmental disorders not otherwise specified also resulted in significant improvements. This new category and reorganization reflects another use of the DC:0-3, a template for scientific investigations of diagnostic classifications and treatment strategies in very young children.

Traumatic Stress Disorder The DC:0-3 also modifies criteria in established DSM-IV diagnoses to capture more age-appropriate manifestations. DC:0-3 traumatic stress disorder is a prime example of this. A modification of the DSM-IV posttraumatic stress disorder, the diagnostic criteria include several unique developmental manifestations of trauma such as posttraumatic play, interference or regression in development (e.g., temporary loss of previously attained developmental skills), and the onset of fears that were not present before the traumatic event. Outlining these age-adjusted criteria provides clear guidelines in lieu of the previous need to extrapolate from the more adult-based DSM-IV criteria. Support for this presumed need to modify the DSM-IV posttraumatic stress disorder criteria for infants and toddlers comes from empirical data showing that traumatized children under 48 months demonstrate these developmentally unique manifestations not included in the DSM-IV criteria.

Like a number of remaining unvalidated diagnoses in the DSM-IV, the current version of the DC:0-3 is based on the collective clinical observations of expert clinicians who treat infants and preschoolers. The manual provides an initial framework for further specification, definition, and controlled investigations into the validity of specific diagnostic criteria for infants and preschoolers. The DC:0-3 is currently in use in a number of infant and preschool centers worldwide, and preliminary testing in clinical samples has demonstrated its fundamental descriptive utility.

SUGGESTED CROSS-REFERENCES

Chapter 37 covers pervasive developmental disorders, and Chapter 34 covers mental retardation. Tic disorders are discussed in Chapter 42.

SECTION REFERENCES

Abe K, Oda N, Amatomi M: Natural history and predictive significance of head-banging, head-rolling and breath-holding spells. Dev Med Child Neurol *26:*644, 1984.

Boris NW, Fueyo M, Zeanah CH: The clinical assessment of attachment in children under five. J Am Acad Child Adolesc Psychiatry *36:*291, 1997.

*Campbell SB: Behavior problems in preschool children: A review of recent research. J Child Psychol Psychiatry *36:*113, 1995.

Dawson G, Osterling J: Early intervention in autism. In *The Effectiveness of Early Intervention*, MJ Guralnick, editor. Paul Brookes, Baltimore, 1996.

DeGangi GA, DiPietro JA, Greenspan SI, Porges SW: Psychophysiological characteristics of the regulatory disordered infant. Infant Behav Dev *14:*37, 1991.

*Emde RN, Bingham RD, Harmon RJ: Classification and the diagnostic process in infancy. In *Handbook of Infant Mental Health*, CH Zeanah, editor. Guilford, New York, 1993.

*Greenspan SI: Reconsidering the diagnosis and treatment of very young children with autistic spectrum or pervasive developmental disorder. Zero Three National Center for Clinical Infant Program *13:*1, 1992.

*Harmon RJ: Diagnostic thinking about mental health and developmental disorders in infancy and early childhood: A core skill for infant/family professionals. In *Educating and Supporting the Infant/Family Work Force: Models, Methods and Materials*, L Eggbeer, E Fenichel, editors. National Center for Clinical Infant Programs, Arlington, VA, 1995.

Harmon RJ, Frankel KA: The growth and development of an infant mental health program: An integrated perspective. Infant Ment Health J *18:*126, 1997.

Leonard HL, Lenane MC, Swedo SE, Rettew DC, Rapoport JL: A double-blind comparison of clomipramine and desipramine treatment of severe onychophagia (nail biting). Arch Gen Psychiatry *48:*821, 1991.

*Luby JL, Morgan K: Characteristics of an infant/preschool psychiatric clinic sample: Implications for clinical assessment and nosology. Infant Ment Health J *18:*209, 1997.

Minde K, Minde R: *Infant Psychiatry. An Introductory Textbook*. Developmental Clinical Psychology and Psychiatry Series, vol 4, AE Kazdin, editor. Sage, Beverly Hills, CA, 1986.

Ross LL, Yu D, Kropla WC: Stereotyped behavior in developmentally delayed or autistic populations. Behav Modif *22:*321, 1998.

Sameroff AJ, Emde RN, editors: *Relationship Disturbances in Early Childhood. A Developmental Approach*. Basic Books, New York, 1989.

Scheeringa MS, Zeanah CH, Drell MJ, Larrieu JA: Two approaches to the diagnosis of posttraumatic stress disorders in infancy and early childhood. J Am Acad Child Adolesc Psychiatry *34:*191, 1995.

Thomas JM, Benham AL, Gean M, Luby J, Minde K, Turnes S, Wright HH: Practice parameters for the psychiatric assessment of infants and toddlers (0–36 months) [Abstract]. J Am Acad Child Adolesc Psychiatry *36:*21S, 1997.

Winchel RM, Stanley M: Self-injurious behavior: A review of the behavior and biology of self-mutilation. Am J Psychiatry *148:*306, 1991.

CHAPTER 45. MOOD DISORDERS AND SUICIDE IN CHILDREN AND ADOLESCENTS

CAROLY S. PATAKI, M.D.

Mood disorders in children and adolescents and their treatments have received increasing attention and clinical investigation over the last few decades. The core features of mood disorders are essentially the same across the life span. Developmental level, however, appears to influence the expression of certain mood symptoms with greater frequency than others within the framework of depressive disorders. For example, pervasive anhedonia and significant psychomotor retardation, typical of adults with depression, are not often observed as dramatically in children with major depressive disorder. Auditory hallucinations and somatic complaints appear with greater frequency among prepubertal children with major depressive disorder than among adolescents and adults, whereas delusions and psychomotor retardation are endorsed more often by depressed adolescents. Some depressive symptoms including suicidal ideation are equally prevalent at all ages. The manner of communicating depressive symptomatology such as irritability, sadness, and negative attributions is always subject to the child's developmental level.

Reports of suicide by children and adolescents have increased over the last decade. Suicide is the fourth leading cause of death in children between the ages of 10 and 15 years and the third leading cause of death among adolescents and young adults 15 to 25 years. Although the rate of death is significantly lower in children and younger adolescents than in older adolescents, the risk factors and methods of suicide need to be examined for each group. Recent findings indicate that psychiatric disorders have relatively less association with suicidal behavior among children and younger adolescents and that family conflict and peer difficulties may be more significant in the expression of suicidal behavior in this age group. Among older adolescents, although interpersonal conflict and loss are often reported to precipitate suicidal behavior, mood disorders and other psychiatric disturbance appear to be more important in the emergence of suicidal behavior. The task ahead is to identify both the psychosocial and the psychiatric risk factors to develop prevention strategies for suicidal behavior among these young persons.

DEFINITION

Major mood disorders comprised mood episodes including major depressive episode, manic episode, mixed episode, and hypomanic episode. Mood episodes, according to the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), are not coded as separate entities, but they serve as the building blocks for many mood disorder diagnoses. The mood disorders include the depressive disorders (major depressive disorder, dysthymic disorder, and depressive disorder not otherwise specified); the bipolar disorders (bipolar I disorder, bipolar II disorder, cyclothymic disorder, and bipolar disorder not otherwise specified), mood disorder due to a medical condition, substance-induced mood disorder, and mood disorder not otherwise specified.

A *major depressive episode* consists of at least five depressive symptoms that have been present for at least 2 weeks and represent a change from previous functioning. Either a depressed mood or a loss of interest or pleasure must be present. In children or adolescents, an irritable mood may suffice for the depressed mood criterion. Significant weight loss or marked change in appetite may be substituted by a failure to make expected weight gains in children or adolescents. Insomnia or hyposomnia, psychomotor agitation or retardation, loss of energy, feelings of worthlessness or excessive guilt are mood symptoms for children and adolescents as well as for adults. Diminished ability to concentrate and recurrent thoughts of death are also depressive symptoms that apply to all ages. The symptoms must cause significant distress or functional impairment.

A *manic episode* consists of a distinct period of persistently elevated, expansive, or irritable mood for at least 1 week or of any duration if hospitalization is necessary. During the mood disturbance at least three additional symptoms (or four if the mood is only irritable) must emerge. These include grandiose thinking or inflated selfesteem, decreased need for sleep, pressured speech or increased verbalizations, racing thoughts or flight of ideas, distractibility, increase in goal-directed activities, or ideation and hopelessness, which seem to be equally likely at any age. The sepcific manner of excessive involvement in pleasurable activities is one of recklessness and thereby may lead to dire consequences. No modifiactions of manic episodes apply to children or adolescents according to DSM-IV.

A *mixed episode* consists of a combination of all the criteria for both a manic episode and a major depressive episode, except that they need only last for 1 week, not 2 weeks. The mood disturbance must be severe enough to cause functional impairment. The modifications within the major depressive episode for children and adolescents also apply in the mixed episode.

A *hypomanic episode* is defined in the DSM-IV as a distinct period of persistently elevated, expansive, or irritable mood lasting at least 4 days. During the mood disturbance, at least three symptoms of mood disturbance must be present or four symptoms if the mood change is only irritability. These symptoms include grandiosity, decreased need for sleep, talkativeness, flight of ideas or racing thoughts, distractibility, increased goal-directed activity, and excessive involvement in pleasurable activities. The episode is associated with an unequivocal change in functioning, but it is not severe enough to cause a marked deterioration in either social or occupational functioning. Hospitalization is not required, and there are no psychotic features. There are no modifications in the criteria for children or adolescents.

According to DSM-IV *dysthymic disorder* is defined by a depressed mood for most of the day, for most days, for a period of at least 2 years. The duration criterion is modified to 1 year for children or adolescents, and the quality of the mood may be described as irritable rather than depressed for youths. At least two additional depressive symptoms must be present in dysthymic disorder. These include change in appetite, insomnia or hyposomnia, diminished energy, low self-esteem, poor concentration or difficulty making decisions, or hopelessness. To meet the criteria for dysthymic disorder, the mood symptoms cannot be absent for more than 2 months.

Depressive disorder not otherwise specified includes disorders with depressive features that do not meet the criteria for other depressive disorders or adjustment disorders with disturbances of mood.

Bipolar I disorder requires a manic or mixed manic episode or a history of these episodes. The initial manic episode may be specified as mixed; that is, the criteria for major depression are also met. Bipolar II disorder is defined in the DSM-IV as recurrent major depressive episodes with hypomanic episodes. The criteria for bipolar II disorder require the presence or history of one or more major depressive episodes as well as the presence or history of at least one hypomanic episode. In bipolar II disorder, there has never been a manic or a mixed manic episode. The symptoms must cause functional impairment.

Cyclothymic disorder consists of at least 2 years with numerous periods of hypomanic symptoms and numerous periods with depressive symptoms that do not meet the criteria for a major depressive episode. In children and adolescents, the duration must be at least 1 year instead of 2 years. During the duration of symptoms (1 year for children and adolescents), symptoms cannot be absent for more than 2 months at a time. There cannot be a major depressive episode, a manic episode, or a mixed manic episode in the first year of the disturbance. After the initial year of symptoms, a manic or mixed episode may be superimposed in which case both disorders may be diagnosed.

Bipolar disorder not otherwise specified includes disorders with bipolar features that do not meet the criteria for the above-mentioned bipolar illnesses.

HISTORY

Mood disorders were recognized in ancient times. In the late 1800s and early 1900s Emil Kraepelin distinguished the syndromes of depression and mania from the deteriorating course of schizophrenic illness in adult populations. Although there was a rare report of a child or adolescent with serious mood alteration in the

literature, for the most part until recently, enduring mood disorders were not believed to occur in childhood.

One exception to this notion was the accepted syndrome of “anaclitic depression,” described by Rene Spitz in the 1940s among infants under 1 year of age who were separated from their primary attachment figure and placed in institutional settings. Spitz described a pervasive deterioration in the emotional and physical functioning of infants who experienced minimal interaction with the primary caregiver. The syndrome that developed over a period of months was characterized by expressed discontent and sadness in the deprived infants, which progressed to more overt distress, and then evolved into a withdrawn, apathetic, and dejected state. After 5 months of continued deprivation of interaction with an attachment figure, the infants failed to thrive and developed loss of appetite, weight loss, and diminished emotional reactivity. He called this syndrome hospitalism. Although this syndrome in infants had many similarities with depressive disorders among adults, the links between mood disorder in adults and during childhood remained dormant for several more decades.

The interpretations of mood disturbance in children and adolescents were constrained by the notion that children were developmentally too immature to mount true depressive responses. For some, this was supported by the belief that depression developed in a process requiring a harsh superego and the inward directing of rage, a chronological and developmental psychological process that could not have been accomplished by young children. Others held that since children were unable to engage in the adult psychological mechanisms generating depression, behavioral disturbances commonly seen among some troubled children must represent “masked depression,” that is, the childhood equivalents of depression. However, when children were assessed directly and standardized criteria for mood disorders evolved, it became evident that children of all ages met criteria for enduring and severe mood disorders.

The issue of developmental modifications of criteria for mood disorders in children and adolescents is still quite controversial. Early age of onset of a mood disorder introduces questions about whether this predicts a more severe, chronic, or recurrent illness and whether it is distinct from the adult form. Additionally, an illness that begins in childhood may differ from one beginning in adulthood with respect to efficacy of treatments and the subsequent pathophysiology of the disorder itself. Current research efforts are aimed at elucidating the relation of childhood mood disorders with those in adulthood, identifying genetic markers, and clinical treatment trials with pharmacological agents and psychosocial interventions.

COMPARATIVE NOSOLOGY

Although there were a number of alterations in the criteria for major depression and dysthymia between the third edition of the DSM (DSM-III) and the revised DSM-III (DSM-III-R), there were relatively few changes between the DSM-III-R and the DSM-IV. DSM-IV like DSM-III-R requires impairment for a diagnosis of major depressive disorder. There is one modification for youth in the DSM-IV for major depression: the criterion of depressed mood may be replaced by *irritable mood*, and in the dysthymic disorder criteria, the 2-years duration used for adults may be modified to 1 year for youths. In DSM-IV, the criteria for mania returned from a distinct period in DSM-III-R to a distinct period of at least a week or of any duration if hospitalization is necessary. Consistent with all of the diagnostic criteria in DSM-IV, there must be some impairment in social or occupational functioning. In DSM-IV, the diagnostic criteria for hypomania include a duration of 4 days and the same specific menu of symptoms given for a manic episode. In contrast to a manic episode, in which the symptoms must be severe enough to markedly impair occupational or social functioning, a hypomanic episode is not severe enough to cause marked impairment but must be associated with some unequivocal change in functioning that is uncharacteristic of the person when not symptomatic.

The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) divides depressive disorders into mild, moderate, and severe categories for single episodes of depression and for recurrent depression. There are no symptom modifications within the ICD-10 criteria for depression in children or adolescents. According to ICD-10, mild depression requires fewer symptoms than moderate or severe depression. The symptom of ‘loss of confidence and self-esteem,' listed as part of the criteria for depressive disorder in ICD-10, is not part of the criteria for major depressive disorder in the DSM-IV, but low self-esteem is listed as a symptom of dysthymic disorder in DSM-IV. ICD-10 does not list a category of depression that is exactly analogous to dysthymic disorder in so far as no depressive disorder requires persisting symptoms for 2 years (or 1 year for youths). ICD-10 does contain one other depressive disorder, depressive conduct disorder, within the broader diagnostic category of mixed disorders of conduct and emotions. Depressive conduct disorder (not contained within DSM-IV) may be particularly relevant to children and adolescents who often exhibit both a conduct disorder and a depressive disorder concurrently. To meet the ICD-10 criteria for depressive conduct disorder, the full criteria for both a depressive disorder and conduct disorder must be met. The most similar combined conduct disturbance and mood disturbance listed within DSM-IV is the category of adjustment disorders. Adjustment disorder with mixed disturbance of emotions and conduct in DSM-IV differs significantly from depressive conduct disorder in ICD-10, insofar as all adjustment disorders in DSM-IV must occur in response to an identifiable stressor within 3 months of the disorder. Furthermore, the disturbance must not meet the full criteria of another disorder such as a depressive disorder or a conduct disorder.

ICD-10 criteria for mania, hypomania, and bipolar I and bipolar II disorder are equivalent to the criteria in DSM-IV.

EPIDEMIOLOGY

The epidemiology of mood disorders and suicide among children and adolescents is complicated by their increasing rates with increasing age. Also, over the last few decades, the criteria for diagnosing the disorders have changed. Finally, the reported incidence of mood disorders among youths over the last few decades has consistently increased, and the age of onset has decreased. This phenomenon has been called the “cohort effect.” It is most evident in studies of mild-to-moderate depression, but it has not been noted for dysthymic disorder. More-severe melancholic depressions are less affected by the cohort effect. Although the mechanism of the cohort effect is not clear, genetic anticipation, in which a genetic disorder tends to worsen in successive generations, has been suggested to explain it.

Prevalence rates for mood disorders and suicide are often reported separately for prepubertal children and for adolescents. Rates of completed suicide are separated by gender as well as age. Large differences in sampling and measurement instruments make it not surprising that a range of prevalence rates were found by different studies. There is also some variation in rates of depressive illness depending on who the informant was. In most cases, patients reported higher rates of disorder than their parents. See Table 45-1 for prevalence rates.

Major depression	
Preschoolers	0.3%
Children	0.4–3%
Adolescents	0.4–6.4%
Dysthymic disorder	
Children	0.6–1.7%
Adolescents	1.6–8%
Bipolar I disorder	
Children	0.2–0.4%
Adolescents	1%
Attempted suicide	
Children	1%
Adolescents	1.7–5.9%
Completed suicide	
5–9-year-old boys	0.04/100,000
5–9-year-old girls	0
10–14-year-old boys	2.4/100,000
10–14-year-old girls	0.96/100,000
15–19-year-old males	18.25/100,000
15–19-year-old females	3.48/100,000

Table 45-1 Prevalence (Point or 1-year)

The lifetime prevalence rate of major depressive disorder in adolescents is estimated to fall between 15 and 20 percent, similar to that in adult populations. Epidemiological studies of mood disorders in children or adolescents support the notion that pure depression is rare among youths. Between 40 and 70 percent of children and adolescents have comorbid psychiatric disorders, and up to 50 percent have two or more comorbid disorders. Anxiety disorders and dysthymic disorder tend to be the most common comorbid diagnoses, major depressive disorder accompanying from 30 to 80 percent of the time. Disruptive behavior disorders co-occur with major depressive disorder 10 to 80 percent of the time, and substance abuse co-occurs 20 to 30 percent. Major depression often follows the other psychiatric disorder, although in general, substance use disorders occur after the onset of major depression. Approximately 70 percent of patients with early-onset dysthymic disorder have a secondary major depressive episode. A number of factors were found to be associated with depressive disorders among youths, including age. In most studies, adolescents had significantly more depressive disorders than children under 12 years of age. In studies comparing early and late adolescents, the older groups also reported higher rates of depressive disorders. Studies have found depression to be significantly associated with the patient's report of family dysfunction. Low self-esteem, not a core feature of major depressive disorder in DSM-IV (although a core feature of dysthymic disorders in DSM-IV), was associated with major depressive disorder in a number of studies. Stressful life events were associated with depression in at least one study. Results with respect to gender were

inconsistent; in children, depressive disorders were either equal in males and females or higher in males and among adolescents, they were either equal or higher in females. Race, poor school performance, and socioeconomic status were inconsistently associated with depressive disorders among youths.

The most frequent age of onset of bipolar I disorder (i.e., an episode of mania) is between 20 and 30 years, followed by 15 to 19 years. Mixed manic and depressed states in youth are often confounded by other diagnostic entities including attention-deficit/hyperactivity disorder and conduct disorder mixed with depression. Among adolescents, bipolar disorder appears to be equally distributed by gender, and 20 to 40 percent of adolescents with major depressive disorder develop a manic episode, either alone or concurrent with depression, within 5 years after the onset of the depression. In young adults with depression, development of hypomania has been associated with early-onset depression, atypical depression, seasonal affective disorder, protracted depressive episodes, mood liability, substance abuse, and significant psychosocial problems.

There have been reports of particular family pedigrees, for example, certain Old Order Amish families, in which bipolar I and II disorders were extremely prevalent. In these family pedigrees the prevalence of bipolar I and bipolar II disorder in family members 15 years or older was 46 percent.

Suicide in children and young adolescents is extremely rare in the United States and worldwide. Among 15- to 19-year-olds, suicide rates have quadrupled over the last four decades. The suicide rate for adolescents represents approximately 12 percent of the total mortality in this age group. Lifetime prevalence rates of suicide attempts among adolescents are reported to range from 3.0 to 7.1 percent. Male suicide rates in this age group are approximately 5 times higher than female rates. Suicide attempts are reported to occur approximately 3 times more often in female adolescents than in males in community samples. Protective factors for children and young adolescents may include less access to lethal means of suicide (firearms) and less realistic plans for suicide. Additional factors increasing the suicide risk, including substance use and depressive disorders, are more likely to be present in older adolescents than in children or younger adolescents. In the United States, the United Kingdom, and Europe, suicide is more common among males at every age, except the infrequent young childhood death by suicide. In Asia and some Latin American countries, suicide rates are equally distributed by gender or are higher in female adults. Ethnic and cultural factors also play a role in suicidal behavior. In the United States, suicide rates are more common among whites than among other ethnic groups, with the exception of certain Native American groups.

Suicide method is the most significant determinant of whether the attempt results in death. In the United States firearms and hanging are the two most common methods of suicide in males. Female adolescents who complete suicide are more likely to take a lethal ingestion or jump from a height. Asphyxiation by automobile fumes from a parked car (often in a garage) occurs in suburban areas, while jumping from a height is more common in urban areas that have tall buildings.

Suicide imitation and contagion, particular features of adolescent and young adult suicides, influence the suicide rate on a short-term basis but not on a long-term basis. Suicidal behavior increases in adolescents following exposure to well-publicized news stories of suicide or a film involving a teen suicide. The baseline suicide rate may increase for several weeks after such a dramatization. Suicide clusters also occur among adolescents within a community after a publicized suicide of a young person. The cluster consists of a larger-than-expected number of suicides taking place in a short period of time. The deaths are not more likely to be adolescents who were close friends or well known to the initial suicide victim. The adolescents who suicide as part of a cluster are those vulnerable to suicidal behavior, often with past histories of suicide attempts.

The great majority of suicide attempts among adolescents have little lethal potential. While use of a lethal suicide means (firearms, hanging) accounts for the outcome in most cases, severity of a depressive disorder, substance use, and persistent suicidal intent predict future lethality. Many adolescents who have survived a suicide attempt deny persistent suicidal intentions. In addition to mood disorders and their comorbid psychiatric disorders, additional predisposing factors for suicidality among adolescents include a family history of suicidal behavior or mood disorder, exposure to family violence, impulsivity, and availability of lethal methods.

ETIOLOGY

Genetic Factors Evidence from offspring studies, adoption studies, and studies of twins reared together and apart indicates that major depressive disorder and bipolar disorder aggregate in families. So far no twin studies have examined rates of depression among adolescents. In adult twin studies, rates of major depression range from 13 to 28 percent among dizygotic twins and from 53 to 69 percent among monozygotic twins. In one twin study in Denmark, concordance for bipolar illness was 19 percent in dizygotic twins and 69 percent in monozygotic twins.

Studies of rates of mood disorders in the offspring of adults with these disorders show at least a threefold increased risk of developing a mood disorder compared with the general population. The lifetime risk of depression in children of depressed parents has been estimated to range from 15 to 45 percent. Depression is also increased in offspring when both parents have depression. Significantly higher rates of depression emerge in adoptees who are offspring of a depressed parent and are reared outside the biological home. Relatives of adoptees with mood disorder have an eightfold increase in the rate of depression over nondepressed adopted controls. Children of depressed parents are also at higher risk for other psychiatric disorders, such as anxiety and disruptive behavior disorders. Other first-degree relatives of adults with mood disorder also show significantly increased risk of mood disorders. Genetic factors have been estimated to account for approximately 50 percent of the variance in the transmission of mood disorders in adult studies, with the other 50 percent attributable to nonshared familial and extrafamilial environmental factors. Offspring of bipolar I parents tend to appear well adjusted in early life, but they have significantly higher rates of bipolar I and bipolar II disorder.

An early age of onset of major depressive disorder, dysthymic disorder, or bipolar disorders predicts a higher family risk of mood disorders among first-degree relatives of the proband. Recent studies of early-onset depression indicate a higher risk of both depressive and bipolar disorders in family members and potentially the earlier onset of illness in these relatives. Lifetime prevalence rates for depression in first-degree relatives of depressed children and adolescents have been estimated to be from 20 to 46 percent. Adolescents with clear-cut manic episodes during adolescence who had prepubertal histories of complex psychiatric symptoms including mood disorders, attentional problems, and aggressive behavior were found to be less responsive to lithium (Eskalith) treatment and to have greater family loading for bipolar I disorder. Among genetic mechanisms considered for bipolar I disorder are preferential maternal or paternal transmission and mitochondrial inheritance (indicating maternal transmission). Studies of family members of children and adolescents with bipolar and depressive disorders have documented higher-than-expected rates of alcohol use disorders for both types of mood disorders.

There have been recent attempts to examine bipolar disorders or symptoms of bipolar disorders among children and adolescents with chromosomal disorders. These studies are preliminary and have not shown generalizability to most patients with bipolar disorder.

The ramifications of the age of onset of a mood disorder are still under investigation. Age of onset can indicate genetic heterogeneity of an illness, indicate a marker of etiology, herald a particular phenotype of the illness, or represent severity of an illness. Defining the age of onset of an illness is complicated by retrospective reporting, ascertainment bias (i.e., the ages present within the sample), and the cohort effect (i.e., an earlier onset of unipolar and bipolar illness reported over the last four decades). Various models have been proposed to explain the age delay in the emergence of mood disorders. One such model, taken from the observation that certain cancers seem to occur with greater frequency with increasing age, is termed the two-hit model. This model posits that to express a mood disorder, a gene carrier (demonstrating the “first-hit”) would need to be exposed to a second hit, thus explaining why some family members presumed to be gene carriers never manifest a mood disorder. The definition of a *hit* might include nongenetic adverse life events. On the other hand, a mood disorder in some individuals might be attributed to two noninherited hits. Thus, for bipolar disorders the age-specific incidence rises abruptly during midadolescence consistent with a model of an inherited mutation compounded by an environmental hit resulting in emergence of disorder. Since the incidence of major depressive disorder continues to increase over time, this disorder could fit a model in which one or more nongenetic hits occurring over time are responsible for the emergence of illness.

Several genetic mechanisms suggested in the literature could contribute to an age delay in the emergence of mood disorders. These include (1) triplet repeat expansion, (2) damage to mitochondrial deoxyribonucleic acid (DNA), and (3) age-specific changes in gene expression. Triplet repeat expansion is a type of mutation consisting of a repeating pattern of the same three nucleotides. The actual size of the triplet repeat expansion can influence the age of onset of a disease resulting from the mutation and the severity of the illness. This is one example of genetic anticipation, since a parent with a short expansion may exhibit a mild form of a disease, but if a longer mutation is transmitted, the offspring may exhibit a more virulent form of the disease with a younger age at onset.

Damage to mitochondrial DNA may play a role in age-dependent processes. Mitochondrial DNA is outside the cell's nucleus, but when damaged, it may exacerbate a cell's potential for dysfunction. Thus, mitochondrial damage can potentiate an already damaged cell, resulting in an earlier display of a disease process. It is not clear whether this process occurs in the emergence of mood disorders.

Age-specific changes in gene expression are related to the complex processes of transcribing genes into messenger ribonuclear acid (RNA) and then translating them into a protein product. Although all cells carry the same amount of DNA, mosaicism occurs (i.e., DNA varies qualitatively between cells). To be expressed, genes are

first transcribed into messenger RNA molecules, which undergo another process called splicing; then the spliced messenger RNA is translated into protein products. The splicing mechanism allows one messenger RNA precursor to generate different final messenger RNA molecules that will ultimately result in different protein products. Thus the DNA a cell starts with may eventually give rise to a variety of protein products. It is not well understood whether this mechanism can account for variation in the age of onset or severity of a mood disorder.

Adolescents who make suicide attempts have higher rates of psychiatric disorders including major depressive disorder, bipolar I disorder, and substance use disorders than nonsuicidal adolescents. A family history of suicide attempts is associated with suicide in adolescents, particularly when the rates are examined in second-degree relatives. Some studies have found higher rates of suicidality in the family members of suicide completers (23 to 38 percent) than in family members of suicide attempters (5 to 22 percent). In one recent study, a family history of depression was associated with an increased risk for completed suicide in adolescents, even after controlling for depression in the proband. This supports the hypothesis that both genetic and environmental contributions predispose an adolescent toward suicide. Twin studies (mainly in adults, but some in adolescents) have provided evidence of a genetic contribution to suicidal behavior. Studies of monozygotic and dizygotic twins showed concordance for suicide (both twins committed suicide vs. one twin) only in monozygotic pairs. In half of the cases, twins were concordant for the same psychiatric disorder. Another study compared 57 adopted probands who completed suicide with matched controls and found that 12 percent of biological relatives of the suicide victims had committed suicide whereas only 0.7 percent of the controls had relatives who had completed suicide. None of the adoptive relatives had completed suicide. The probands with diagnosed affective disorders had the ighest rates of biological family members who had completed suicide. Biochemical abnormalities found in suicide victims, particularly in the serotonin system (discussed below), may also imply a genetically determined vulnerability to suicide.

Cognitive Factors A number of theories have been developed to explain the relation between mood disorders, especially depression, and cognitive style. Cognitive distortions and negative attributions are commonly observed among depressed children, adolescents, and adults. Depressed children and adolescents exhibit higher levels of cognitive distortion and hopelessness and the tendency to attribute their experiences to uncontrollable external causes. How a negative cognitive style is developed is not clear; suggestions include a learned component (i.e., modeling of these styles at home or in the environment), experience with stressful unpredictable life events, and a temperamental component of perfectionistic traits that can enhance such cognitive patterns. The cognitive distortion theory, developed by Aaron Beck, hypothesizes that persons who seem to focus on negative information perpetually reinforce a negative view of about themselves, the world, and the future, and are at higher risk of becoming depressed. The distortions are negative catastrophic beliefs and overgeneralizations of negative events and beliefs. The notion of learned helplessness, first proposed by Martin Seligman describes the cognitive experience of a person who perceives unpredictable and uncontrollable events as the basis for feeling helpless and powerless. Another form of cognitive distortion observed in depressed persons is the belief that they are responsible for negative events that may be uncontrollable.

It is not well understood whether negative cognitions are more often the cause or the result of depression. A depression seems to exacerbate low self-esteem, a sense of being powerless over negative events, and negative attributions. When a depression resolves, persistent low self-esteem may linger as a scar and predispose a child or adolescent to future episodes. Recent reports indicate that a negative cognitive style predisposes children to more-prolonged mood disturbance in the face of a negative event. The cognitive-diathesis model proposes that the interactions of negative cognitive style with exposure to adverse life events are associated with development of depressive disorders.

Cognitive characteristics of children and adolescents of adults with bipolar disorders include a discrepancy between verbal and performance intelligence quotient (I.Q.). This discrepancy exceeds the variations expressed by normal controls. A recent report of school function among children stated that the children destined for bipolar I disorder showed a relative weakness in mathematics skills.

Cognitive patterns and attributional style in suicide attempters are shared with those in children and adolescents with mood disorders, since these two groups overlap. There is a subgroup of suicide attempters and completers who do not meet criteria for a mood disorder. Suicidal children and adolescents are less flexible in generating solutions for problems than nonsuicidal children. Suicidal children also predict more-negative consequences for their solutions than nonsuicidal psychiatric controls. Suicide attempters seem to be focused on the problem situation and unable to synthesize an effective strategy to deal with it. They tend to have less confidence that a solution will work. The adult literature reports a strong relation between hopelessness and suicidal behavior. Studies of cognitive features of suicidal children and adolescents have yielded contradictory results. Several recent studies of children and adolescents concurred that suicidal children and adolescents report lower self-esteem and more depression. One recent study examined both negative affect and cognitive biases in adolescents with suicidal ideation, suicide attempters, and nonsuicidal controls. The suicide ideators reported significantly greater hopelessness and poorer self-concept than psychiatric controls without suicidal ideation. Suicide attempters did not differ significantly from either group on any measure. The suicide ideators in this study reported the highest levels of hopelessness and the poorest self-concept. Negative affect and cognitive bias variables accounted for 48 percent of the variance in suicidal ideation. Among the measures of negative affect, only anxiety and depression were uniquely related to suicidal ideation and to each other. Several past studies have shown that impulsive adolescent suicide attempers report less depression and hopelessness than nonimpulsive suicide attempters. Some studies have found gender differences, with suicidal adolescent females reporting higher levels of depression and anxiety than suicidal males, but this is not universal.

Environmental Factors Family interactions with regard to the development of depression in a child or adolescent are complicated and difficult to interpret. In families in which one or both parents have a mood disorder, a child may be faced with an irritable or withdrawn parent whose reactions are inconsistent or hostile. These parenting factors are associated with disruptive behavior and (in some cases) aberrant attachment in the children, even in nondepressed populations. Children raised in family environments characterized by high levels of parental depression, conflict, less involved parents, or harsh parenting styles have increased risk for a variety of psychiatric disorders. More specifically, depressed parents may model negative cognitive styles and poor self-esteem, resulting in poor coping strategies in the child. Young children of mothers with unipolar depression exhibit observable anxiety and disruptive behavior, whereas toddlers of parents with bipolar illness do not. At slightly older ages, however, children at risk for bipolar illness showed higher-than-average rates of attentional and behavioral disturbances. Marital discord and lack of adequate support system for the family are additional risks for development of depression in the offspring. Overall, risk for depression in children of depressed parents is highest when parental illness is of earlier onset, is recurrent, and disrupts parental functioning.

Recent studies attempting to determine the relative roles of genetic endowment and family environmental factors in the development of depression and anxiety in childhood indicate that the greatest contribution to depression and anxiety is made by nonshared environmental factors affecting an individual child within the family, rather than genetic transmission or shared family factors. The mechanisms of these unique influences on a vulnerable child remain to be clarified.

Work with adult patients with schizophrenia and bipolar illness has shown that high levels of negative expressed emotion in families predict increased relapse rates. Preliminary evidence in studies of adults with bipolar disorders indicates that psychosocial interventions that educate families about these interactional styles may decrease relapse rates.

Family conflict is the most frequent precipitating event reported by adolescents with suicidal behavior. A recent study found that suicidal behavior in adolescents is independently associated with conflictual family relationships and depression. That is, lack of family warmth, family discord, and disturbed maternal relationships made an independent contribution to the risk of suicidal behavior. Among nondepressed adolescents, females with a history of conduct-disordered behavior were at increased risk of suicidal behavior. Family discord increased the risk for suicidal behavior in depressed adolescents. Disturbed family relationships may interfere with a child's acquisition of social problem-solving skills, creating a deficit in coping with stressful life events. Since family relationship problems make a contribution to the risk of suicidal behavior beyond that of depression, family interventions may play a major role as a strategy for diminishing suicidal behavior among youth.

Adverse Life Events The adult literature and a number of studies with children and adolescents document a correlation between stressful life event and depressive symptoms. Several studies have shown more stressful life events in depressed children and adolescents than in nondepressed controls. Based on life events inventories delineating the chronology of the events, perceived level of stress, and depressive symptoms, Ian Goodyer found that up to 70 percent of children and adolescents with new-onset depression reported an adverse life event in the 12 months preceding the depression, compared with 29 percent of controls.

Stressful life events that have been studied include interpersonal losses, relationship failures, divorce, bereavement, and exposure to a suicide. David Brent showed that exposure to a suicide was associated with a threefold increase in development of a depressive episode. The risk of developing depression is correlated with the relationship of the survivor to the suicide victim (i.e., the closer the relationship, the higher the incidence of a depressive episode).

A recent study by Douglas Williamson and coworkers found that depressed adolescents experienced more stressful life events during the depressive episode than nondepressed adolescents. A recent study by Boris Birmaher found that stressful life events in depressed adolescents were correlated with higher immune activity. It is not surprising that adverse life events are correlated with mood disorder especially in young children. One study of major depression in preschoolers found that 100 percent of the subjects had been victims of neglect or abuse. Finally, a past episode of major depression can produce a scar effect that results in heightened

interpersonal sensitivity and a persistent need for increased support.

A number of studies have investigated life events reported to precipitants of serious suicide attempts among adolescents. One such study concluded that adverse life events among adolescents during the preceding year (especially an interpersonal loss or discord with family, friends, or partners) significantly increases the risk of a suicide attempt. The predominant role of interpersonal life events in the suicide attempt could not be accounted for by antecedent family or life circumstances. Interpersonal adverse life events were experienced by one third to one half of adolescents with suicide attempts studied. A secondary category of adverse life event associated with suicide attempts among adolescents pertained to legal problems and difficulties with the police in the year preceding the attempt. In one recent study, 24 percent of adolescents who had made suicide attempts experienced legal problems within the year prior to the attempt, compared with less than 2 percent of the adolescent control group. These results are consistent with earlier studies that found strong associations between impending disciplinary crises and suicide attempts among adolescents. Other categories of life events associated with suicide attempts include work difficulties and financial problems. Personal illness and family illness were not significantly associated life events, but they may become more of a risk factor as the subjects age.

A recent study investigating the relation of physical abuse among adolescents to the risk of suicide attempts did not find a significantly higher number of suicide attempts among the adolescent group that was abused. The two groups did differ, however; the physically abused group had a cumulative higher number of risk factors for suicide and were believed to be more likely to make a suicide attempt in the future. Abused adolescent suicide attempters were characterized by higher rates of divorced parents and higher rates of comorbid psychiatric disorders, including depressive disorders, disruptive behavior disorder, and substance use disorders. The nonabused suicide attempters had less comorbidity and lower rates of psychiatric disorders altogether. Physically abused adolescents who made suicide attempts were more likely to perceive their mothers as noncaring than those who had never attempted suicide and they described their families as noncohesive and hostile. They were also more likely to carry a diagnosis of a disruptive behavior disorder, a substance use disorder, or both and were more likely to report exposure to a suicide in a peer or family member.

Biological Factors

Hypothalamic-Pituitary Axis Abnormal cortisol secretion in adults with major depressive disorders was demonstrated with overall elevation of 24-hour basal secretion, nonsuppression on a dexamethasone (Decadron) challenge test (DST), blunting of normal diurnal rhythms, and increased nocturnal cortisol secretion. It has been more difficult to document these changes systematically in depressed children and adolescents. Neither studies of 24-hour cortisol secretion nor nocturnal measurements have yielded robust differences between cortisol concentrations in depressed children and adolescents and those in normal controls.

DST studies in children and adolescents have been less robust in eliciting differences in cortisol secretion among children with depression from children with other psychiatric disorders. DST studies have also failed to show consistent differences among children with depression from children without any psychiatric disorders. A review of Connie Casat revealed a greater number of nonsuppressors with all types of psychopathology among hospitalized children and adolescents than among outpatients. DST appears to be a more specific test in adolescence compared to childhood, that is, in adolescents, a positive DST result was correlated with major depressive disorder 85 percent of the time, whereas for prepubetal children, the specificity was only 60 percent. The DST is not hightly specific, since 47 percent of hospitalized adolescents with major depressive disorder were nonsuppressors on the DST, and 20 percent of inpatients with other psychiatric disorders were also nonsuppressors.

Results of studies investigating hypothalamic-pituitary axis in relation to suicide have been inconsistent. In a study of hospitalized adolescents, nonsuppression of cortisol secretion after dexamethasone challenge was associated with patients who had made medically serious suicide attempts; those who exhibited less potentially lethal suicidal behaviors were less likely to have nonsuppression. Higher levels of suicidal intent existed among those who had made serious attempts. These results may be confounded by the potential association of serious depressive disorders with nonsuppression. Another study of prepubertal children showed an association between suicidality and nonsuppression of cortisol during a DST, independent of diagnosis.

Other measures of hypothalamic-pituitary-adrenal axis dysregulation found in adults with major depressive disorder include blunted secretion of adrenocorticotropic hormone (ACTH) after a challenge with corticotropin-releasing hormone (CRH). Cortisol concentrations remain normal. A recent study by Boris Birmaher found no significant differences in cortisol or ACTH concentrations in children with major depression and normal controls.

Sleep Studies Among adults with major depression, sleep abnormalities have included reduced slow-wave (delta) sleep, diminished latency to rapid eye movement (REM) sleep, increased REM density, and increased awakening during the night. Sleep studies with children and adolescents have provided variable results with respect to sleep architecture abnormalities. Prepubertal children with major depression have shown diminished REM latency and decreased slow-wave sleep in some studies; other studies have found no differences between depressed children's sleep and that of normal controls. Joaquin Puig-Antich followed sleep patterns of depressed children after recovery and noted that diminished REM latency persisted after recovery. This finding (not yet duplicated) has raised the possibility that the sleep abnormalities in these children were markers of risk for depression rather than changes based on the episode of depression. The persistence of abnormal sleep after recovery in some depressed children may also reflect a scar from the episode of depression. Among sleep studies in adolescents, Rao followed depressed subjects over time and found that those who went on to have recurrent major depressive episodes had shorter REM latency during the index depression than adolescents who had no recurrence. Adolescents with a single episode of depression or those who go on to have bipolar I disorder may have less consistent sleep-architecture abnormalities than more severe recurrent depressives, compared with normals.

Growth Hormone Children with depression hyposecrete growth hormone after pharmacological challenges with clonidine (Catapres), levodopa (Larodopa), growth hormone–releasing hormone, and desmethylimipramine. Growth hormone secretion is blunted in response to insulin-induced hypoglycemia during the depressive episode as well as after resolution of the depression. Nocturnal secretion of growth hormone has been reported to range from hypersecretion to hyposecretion. Growth hormone, secreted by the anterior pituitary, is secreted in greatest quantity during sleep throughout the life span, and during adolescence an increase in daytime growth hormone secretion results in an overall increase in 24-hour growth hormone secretion. Blunted growth hormone response may be a trait marker for depression (as opposed to a state marker), since it occurs among children who have had depression, even after the depression has remitted. Alternatively, after a depression has occurred, blunted growth hormone response may persist as a scar response.

There have been only a few studies suggesting that suicidality among adolescents is associated with a blunted growth hormone response to a desipramine (Norpramin) challenge. More studies are needed to differentiate whether this response is really just a measure of severity of depressive disorder or if it is associated with suicidality independent of the severity of mood disorder.

Serotonin Some evidence in adult studies indicates that a dysregulation of the serotonin system might contribute to the development of depression. This hypothesis is based on the blunted response of cortisol secretion to serotonergic agents used to challenge the system. One study to date has suggested that depressed children have a blunted cortisol response when challenged with an intravenous dose of L-5-hydroxytryptophan, a precursor of serotonin. Infusion of L-5-hydroxytryptophan increases serotonin turnover. These studies need to be replicated.

Most of the current investigations of biological correlates among suicide attempters and completers have involved adults. These studies have found significantly lower concentrations of the serotonin metabolite 5-hydroxyindoleacetic acid (5-HIAA) in the cerebrospinal fluid (CSF) of both suicide attempters and suicide completers. In postmortem studies, lower concentrations of serotonin itself and its major metabolite (5-HIAA) were found in the brains of suicide completers than in normal postmortem brains. These lower concentrations of 5-HT and 5-HIAA were localized to the brainstem and were not found in the cortex. Other studies have found that brains of suicide completers contain an increased postsynaptic number of serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptors in the prefrontal cortex, suggesting that this increase in receptors may be a mechanism to compensate for diminished serotonin release. Low concentrations of serotonin and its metabolites in CSF also characterize subjects with impulsive, labile, and aggressive behaviors across a variety of diagnoses. The postmortem studies of receptor density and serotonin concentration among suicide completers have led to controversial results; further studies are necessary to confirm some of these results. Several studies have demonstrated a higher number of platelet 5-HT_2 receptors in subjects with suicidal behavior across a range of psychiatric disorders. There is evidence (mainly from adult samples) of abnormal serotonergic function and possibly of changes in platelet 5-HT_2 receptors among suicide attempters.

Imaging Studies One open pilot study of magnetic resonance imaging (MRI) of children and adolescents with bipolar I disorder revealed enlarged ventricles and an increased number of hyperintensities.

Behavioral Genetic Studies Behavioral genetic research on childhood mood disorders investigates the causes of individual differences pertaining to genetic variability, shared environmental influences, and nonshared environmental influences. Environmental influences in this context include all sources of variation not explained by heritable genetic effects, including perinatal factors, psychosocial experiences, illness, and any developmental chromosomal influences that are not

inherited. This kind of investigation attempts to identify factors that are either shared or not shared within a family environment to explain variability in the emergence of depressive symptoms and disorders among siblings who share (1) no genetic material (adoptive siblings), (2) some genetic material (dizygotic twins), and (3) virtually all of their genetic material (monozygotic twins). One recent twin study of self-report of nonclinical depressive symptoms using dizygotic and monozygotic twin pairs found .20 for a heritability estimate, .27 for the shared environmental estimate, and .53 for the nonshared environmental factors relating to the symptoms. When the investigators looked only at the adolescents at the high end of reported depressive symptoms, however, genetics was estimated to account for 59 percent of the variation. Another study of twins who self-reported nonclinical depressive symptoms found a difference in the factors that accounted for the variation in depressive symptoms for children (8 to 11 years) and adolescents (11 to 16 years). For the children, shared environment accounted for 60 percent of the variation, nonshared environment accounted for 22 percent and genetics accounted for 18 percent. In the adolescent group, the variation in depressive symptoms was largely due to genetic factors, 78 percent, with shared environmental factors accounting for 4 percent, and nonshared environmental factors accounting for 18 percent of the variation.

Finally, in a recent large study of nonshared environmental factors in adolescents, shared, nonshared, and genetic factors accounted for 4, 62, and 34 percent of the variation in depressive symptoms for the entire group, respectively. When only the most depressed adolescents were analyzed, the shared environment accounted for 44 percent of the variation, the nonshared environment accounted for 33 percent of the variation, and genetics accounted for the remaining 23 percent. These results conflict with those of the first study mentioned, in which genetic factors accounted for more of the variation when the most depressed adolescent subgroup was analyzed.

Although there is much left to study in this area, several conclusions can be drawn. The most obvious but significant conclusion is that genetics clearly plays at least a moderate role in adolescent depression. Additionally, both shared and nonshared environmental influences appear to be important in the emergence of depression in adolescents. This means that differences experienced by siblings within family systems have a significant impact on the emergence of depressed symptoms in adolescents. Thus, the relative experience, in comparison to the experience of siblings, bears upon the development of depressive symptoms. Experiences outside the family, such as peer and school experiences, are also likely to have significant influence as nonshared environmental factors in the emergence of depressive symptoms. In general, assessment of an entire family is needed to understand the relative influences on an adolescent who presents with depressive symptoms.

DIAGNOSIS AND CLINICAL FEATURES

Major depressive disorder among children and adolescents presents as an enduring syndrome with the same core features manifested in adults. Often the two DSM-IV criteria that are modified for youth—(1) irritable mood in a child or adolescent may fulfill the depressed mood criterion, and (2) failure to attain expected weight gain may replace the weight loss and/or disturbance of appetite criterion for a child or adolescent—are not needed to make the diagnosis of major depressive disorders in children and adolescents.

Children with major depressive disorder may describe their predominant mood symptom as sad, mad, or bored. Children who exhibit an acute change in mood and functioning are most often identified, whereas children with long-standing sadness and irritability are sometimes overlooked. Among adolescents, certain features of depression such as guilt, hopelessness, and beliefs about their future are more likely to be experienced and described than they are among younger children. Currently no specific notations exist for subclusters of symptoms that occur more regularly in children of different ages in the diagnostic system. Clusters of symptoms within the criteria for major depression are more or less likely to occur depending on the age of the patient, from the very young child to adulthood. Table 45-2 summarizes the frequency of symptoms of depression on the basis of age.

Symptom	Adult (*N* = 54)	Adolescent (*N* = 92)	Child (*N* = 95)	Preschool (*N* = 9)
Felt sad	93	99	95	100
Appeared sad	74	47	64	89
Crying	50			89
Irritability	67	83	83	78
Social withdrawal	78	73	64	89
Anxiety, seasonal affective disorder	30	59	78	67
Somatic complaints	74	66	83	100
Fighting/disorder conduct	—	5	38	11
Mood-congruent hallucinations	2	10	22	—

Adapted from Carlson GA, Abbott SF: Mood disorders and suicide. In *Comprehensive Textbook of Psychiatry*, ed 6. HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Table 45-2 Frequency (%) of Depressive Symptoms Among Children and Adolescents

Additional symptoms that appear more frequently as age increases include anhedonia, delusions, psychomotor retardation, hopelessness, hypersomnia, and diurnal variation. Symptoms that decrease in frequency with age include somatic complaints, auditory hallucinations, poor self-esteem, and sad appearance. Some depressive symptoms seem to be represented equally at all ages, including suicidal ideation, poor concentration, depressed mood, and insomnia. While suicidal ideation occurs with equal frequency at all ages, the likelihood of a prepubertal child completing suicide is extremely low (0.04 per 100,000). Factors that protect prepubertal children from completing suicide include little access to lethal methods (firearms), limited ability to plan and carry out a lethal attempt, and the developmentally expected cognitive gap between having a thought (suicidal ideation) and following through behaviorally (a lethal attempt). In adult studies, hopelessness was often correlated with suicidal behavior. Hopelessness occurs more frequently in older children and adolescents with depression.

Specific signs of depression are especially salient in each developmental period. For example, major depression among preschoolers is often evinced by a lack of smiling, apathy toward play, and an overriding lack of involvement in all activities. These children may become tearful or irritable easily, and their activities may be destructive toward themselves, others, and property. Physical complaints are prominent, and physical aggression may be high. It is necessary to assess the family environment, because depressed preschoolers are at high risk for having been victims of some form of maltreatment.

Among school-aged children, deteriorating school performance and avoidance of peers may signal depression. Not infrequently, increased irritability, fighting, or argumentativeness may be associated features of a major depression in a child of this age. Exacerbation of anxiety symptoms and school refusal are not uncommon among children who are depressed. While vegetative symptoms are usually not predominant, mood-congruent auditory hallucinations that frighten and berate the child may emerge.

In adolescents, major depressive episodes occur with increasing frequency. As with children of all ages, the depression may be the primary psychiatric disorder, or it may be superimposed on a number of other behavior and anxiety disorders that generally emerge prior to puberty. Some adolescents recognize their state as depression, while others cannot identify the various components of the depressive state. More-severe depressions among adolescents that are characterized by mood-congruent auditory hallucinations and delusions, psychomotor retardation, and pervasive anhedonia have been recognized as the syndromes that more often are followed in the future by a period of mania and, consequently, a diagnosis of bipolar disorder. Adolescents with incipient schizophrenia may also present with social withdrawal, auditory hallucinations and delusions, and anhedonia. In the past, many adolescents who went on to develop bipolar disorder were misdiagnosed as having schizophrenia.

Several categories of specifications can be coded within the DSM-IV in relation to the current major depressive episode. These include severity, presence or absence of psychosis, presence or absence of melancholic features, presence or absence of catatonic features, and presence or absence of atypical features. *Melancholic features* include loss of pleasure in all activities or lack of reactivity to usually pleasurable stimuli or both in conjunction with three of the following symptoms: a depressed mood qualitatively different from the feeling of grief over loss of a loved one, depression that is worse in the morning, early morning awakening, marked psychomotor retardation, significant weight loss, or excessive guilt. *Atypical features* refer to mood reactivity in conjunction with two or more of the following: hypersomnia, heavy feeling in the extremities, long-standing pattern of interpersonal rejection sensitivity, or significant weight gain.

Another specified symptom within the description of a depressive illness pattern is seasonality. Winter seasonal affective disorder (mood disorder with seasonal pattern) affects between 3 and 4 percent of children. Relatively few studies have been done on children and adolescents with seasonal affective disorder even though more than one third of adults with seasonal mood disorders report that their symptoms began in childhood. Those studies that have been published describe

depressed states characterized by sadness, fatigue, hypersomnia, and carbohydrate craving. This constellation of symptoms has been categorized as atypical depression. Adult seasonal affective disorder studies have shown that disturbance in activity (i.e., circadian rhythm disruptions) were present. A recent study found that children with seasonal affective disorder displayed abnormal patterns of at-rest motor activity that are identical with patterns of depressed non–seasonal affective disorder patients, but different from patterns in adults with seasonal affective disorder.

Dysthymic disorder is described in DSM-IV as a less-severe depressive entity than major depressive disorder in certain respects, yet a more chronic form of depression with a longer recovery period than major depressive disorder. It consists of a depressed mood for most of the day or for more days than not for at least 1 year. In adults, the duration necessary for the diagnosis is 2 years. In addition to depressed mood, the criteria for dysthymic disorder may include low self-esteem, diminished energy level, feelings of hopelessness, insomnia or hypersomnia, poor concentration, and appetite disturbance. There is never a period of 2 months or more within the year of an absence of the dysthymic symptoms.

Mania in children and adolescents is currently controversial because of a lack of consensus among clinicians and researchers as to whether its identification requires a discrete episodic presentation. There are no modifications in the DSM-IV criteria for manic episode in children and adolescents, and yet there is considerable variation in defining a manic or hypomanic episode in a child who exhibits symptoms of the above but not in easily definable discrete episodes. Thus, a developmental viewpoint must be incorporated into future diagnostic criteria for bipolar I disorder in childhood. According to a recent review of bipolar illness in childhood and adolescence, childhood-onset bipolar illness may be more severe; it may have a chronic, nonepisodic course; and it may present in a mixed, or rapidly cycling, form that resembles the form seen in the most severely ill adult bipolar patients, who are relatively treatment resistant.

Children and adolescents with mania may present with infectious elation and interact in an amusing and silly manner. Children or adolescents with mania are likely to disclose age-appropriate grandiose delusions (e.g., they are that on the brink of achieving greatness in academia or in the arts), even in the face of impending failure. They may feel compelled to share solutions to world problems and instruct others in the meaning of life. Grandiose children may believe that they should be exempt from rules because of their more profound understanding of the laws of nature or the will of God. Adolescents with mania may also experience auditory hallucinations or visions of a grandiose nature. Children and adolescents with mania are well known for their extremely high energy levels, and they may manifest both high-risk behaviors with peers and hypersexual episodes far exceeding the comfort level of their friends. In contrast to adolescents with hyperactivity syndromes, manic adolescents are often illogical and driven to interact incessantly with peers and family members. Adolescents are often frightened and put off by a manic adolescent peer; however, when feedback is provided, it is often dismissed by the patient. Many adolescents do exhibit classic adultlike manic episodes. There is often no discrete episode in prepubertal children who exhibit hypomanic or manic symptoms that had an insidious onset and may not be increasing but are now impairing the child. In such cases, it is more difficult to make the diagnosis. There does seem to be a subgroup of children who present with hyperactivity in the preschool years and exhibit manic symptoms in the early grade school years.

Suicidal behavior can take many forms. Adolescent suicidal behavior usually occurs within the context of family conflict or recent perception of an interpersonal loss. Studies have also found that a subgroup of adolescent suicides took place during a disciplinary crisis within the family setting, with the law, or at school. Anticipation of punitive actions or social humiliation has precipitated teenage suicide. There is also evidence that suicidal behavior among adolescents has been exacerbated by exposure to media presentations pertaining to suicide, either in the form of news stories or television movies. Suicidal intentions can be in some patients recognized and identified. Psychological autopsy studies have found that up to 50 percent of adolescents with completed suicide had verbalized suicidal intentions within a day of the event. Thus, all suicidal threats by adolescents should be considered potential serious warnings. Clinical evidence of increased risk of suicide potential in a child or adolescent include previous attempts, a lethal plan, accessibility of lethal means (especially guns), probability of alcohol use, and a plan that includes precautions against being rescued. Additional features of high-risk clinical scenarios in a child or adolescent include suicide pacts with peers, desire to join a deceased friend or relative, inadequate parental supervision and support, and inability of the child or adolescent to agree to an explicit verbal or written no-suicide contract with the clinician.

Taylor, a 16-year-old eleventh grader was admitted to the adolescent inpatient unit from the emergency room where he had been brought by ambulance after he was found hanging in his bedroom, cyanotic, with a rope around his neck. He was found by his 17-year-old sister who had gone to his room looking for a cigarette. Since his mother was out for the evening, his sister panicked and then quickly called 911. According to his sister, Taylor seemed to be more secretive, less trusting, and more isolative since the beginning of the eleventh grade. Since his parents' divorce 4 years earlier, Taylor had taken care of things around the house because his mother was chronically depressed and alcoholic and his estranged father had a long history of being irresponsible and impulsive. Lately, Taylor seemed to forget what needed to be done, and he seemed preoccupied and distraught. Prior to the 11th grade, Taylor had been a good athlete, he had played on the school basketball team, and he had gotten along well with peers. This year, he had quit the team, and had become strangely isolated. Taylor's sister had mentioned to their mother that Taylor seemed to be sleeping all the time and was avoiding his friends in school. Although his mother was aware that his grades had dropped considerably, she thought he was just going through a phase and decided to leave him alone.

Once situated on the adolescent inpatient unit, Taylor admitted that he had been contemplating suicide off and on for several months. He felt extreme guilt for letting his mother and sister down by becoming more irresponsible around the house, just like his father. He didn't have the energy to get up in the morning or to get to classes. He was unable to concentrate when in school, and he began to feel that his friends were rejecting him. He realized that he was not being invited to parties or included in social activities with this friends. He believed that he was possessed by a satanic spirit that would whisper to him at night that he should kill himself. On the night of his hanging attempt, Taylor had decided that his mother and sister would be better off without him.

Taylor was given the diagnosis of major depressive disorders with psychotic features and treated with fluoxetine (Prozac), which was titrated to 40 mg over a 10-day period. On the seventh day the staff began to notice an increase in energy and a marked increase in his verbalizations. He was less and less isolative and appeared to be pressured with regard to conversing with patients. Several patients began to complain that Taylor was intrusive, irritable, and silly. Taylor began to believe that he had come to understand secrets of life that he felt compelled to share with others. He was speaking louder and was unable to sleep. He didn't understand why others were not fascinated by his revelations. He began to mistrust the other patients and staff, believing that he had been chosen by God to share his special knowledge. He was in constant motion, barely sitting down long enough to eat. Although he was overly active, intrusive, and silly, Taylor was emotionally labile, crying often and reporting that he felt depressed. Taylor's mother and sister had never seen him this way, but his mother reported that he reminded her of his father. Taylor continued in this manner for 4 days, and even after his fluoxetine treatment was discontinued, he continued to exhibit manic symptoms. Taylor was given the diagnosis of a mixed episode and was given divalproex (Depakote). The dosage was titrated to dose of 1000 mg a day, with a blood concentration of 80 µmg/mL. Taylor became subdued; he was able to converse more reasonably and reported that he was not suicidal. He still felt depressed and wondered if he would be able to make anything of his life now. A trial of sertraline (Zoloft) was added to his divalproex therapy, and within 2 weeks, he was somewhat less depressed. Taylor was discharged from the hospital and placed in a partial hospital program where he remained for another month before he returned to school. During this month family meetings were held twice a week, during which the family was given information about bipolar illness and the importance of family support. Taylor's mother entered treatment herself and vowed to drink less and be more available to Taylor and her daughter. Taylor's sister, who was exhibiting symptoms of depression and anxiety about caring for her brother, also entered into her own therapy. When he returned to school, Taylor was mildly depressed, nonsuicidal, and willing to participate in outpatient therapy. He remained concerned about how bipolar disorder would influence the rest of his life.

Brian is an 11-year-old boy who was transferred from his residential facility to the children's psychiatric unit after he had broken a window, threatened to run away, and stated that he was going to kill himself. Brian had a history of attention-deficit/hyperactivity disorder, and he had been treated with a variety of stimulants without great success. He had been placed in a residential treatment facility just 6 weeks prior to the admission, after he had tried to strangle his infant baby sister. Although Brian was quite bright, he had never done well in school because of his inability to follow directions, wait his turn, or get along with his peers. At times his mother and his teacher would describe periods in which his activity level seemed even more frenetic, and he had told his teacher that he would like to have sex with her. Although he was chronically hyperkinetic, at times he appeared giddy and grandiose in his thinking. He constantly made up stories, and he was continually rejected by his peers because of his absurd story telling. He also had another side, however, which was perpetually present but seemed to wax and wane in intensity. He was irritable and believed that nobody liked him or cared for him. When in this mood, he became inconsolable and reported that he was not going to listen to any adults and that life was not worth living. He was devastated when his parents put him into placement, and he had been defiant and irritable since his placement. On admission, Brian was tearful and hopeless. He reported that he wanted to die because he could not control himself and nobody would ever love him the way he was. He stated that he simply could not get his behavior under control even if he tried. He believed that there was definitely something wrong with his brain, and although he denied hearing voices, he felt that there was a noise in his head. Brian's mother carried a diagnosis of bipolar I disorder, as did his maternal grandmother. Brian believed that he always got blamed unjustly and that there was no reason to believe that things could ever get better. After a complete screening of Brian's blood chemistry, he was given a trial of lithium. The dosage was titrated to 1200 mg a day, with a blood concentration of 1.0 mmoL/L. On this medication, Brian seemed calmer, less irritable, and less aggressive. He still had poor concentration and had difficulty focusing on his school work. 10 mg a day of dextroamphetamine (Dexedrine) added to the lithium treatment resulted in some increase in vigilance and performance in the classroom. Although Brian did not exhibit clear-cut periods of pure mania or pure depression, his waxing and waning of mood symptoms seemed most consistent with a diagnosis of bipolar I disorder.

PATHOLOGY AND LABORATORY EXAMINATION

Since no specific laboratory tests serve as direct measures of mood disorders or suicide risk, the clinical history, medical history, and interview remain the most important sources of diagnostic information. A number of laboratory measures have been suggested as potential markers of past or present mood disorder or as influential in affective symptoms. The following paragraphs briefly review these tests.

A child or adolescent who complains of a significant change in energy level, appetite, motivation, or sociability may be experiencing both a medical condition and a psychiatric condition. Viral syndromes and other low-grade infections may produce symptoms that overlap with those of a mood disorder. Thus, a physical examination and a screening blood panel for blood count and (in some cases), electrolyte concentrations may be indicated. A pregnancy test or a reliable history of absence of sexual activity is needed before any adolescent female is given a medication that may have teratogenic effects.

The DST has been suggested to be a biological marker of depression. Nonsuppression of cortisol secretion during the DST has been associated with severe depression, more consistently with adults than with children or adolescents. Nonsuppression of the DST is not specific for mood disorders, however, and thus it does not help in making the diagnosis. DST also does not predict treatment choice or probability of response. Therefore, the DST is not a helpful diagnostic tool.

Thyroid function tests that measure serum thyroxine, triodothyronine resin uptake, thyroid-stimulating hormone (TSH), and triodothyronine may be useful with a child or adolescent with a psychiatric disorder because of the high frequency of thyroid dysfunction in this population and the potential overlap of mood disorder symptoms with those of thyroid disease. Thyroid function testing is not helpful as a diagnostic tool for the diagnosis of depression. The thyrotropin-releasing hormone (TRH) stimulation test has shown no differences in the responses of children and adolescents with depression and those of normal controls.

Toxicology screens for drugs of abuse including opioids, cannabinoids, amphetamines, barbiturates, and cocaine are advisable on admission for adolescent psychiatric inpatients. This battery may be useful in clarifying contributing causes of a new onset of psychosis, mood changes, and aggressive behavior. Positive urine toxicology reports should be confirmed by blood testing whenever possible. Toxicology screening may, however, reveal past exposure rather than current use (e.g., positive results several weeks after cannabis or cocaine use and 48 hours after use of amphetamines).

Brain-imaging techniques, including MRI, computed tomography (CT), positron emission tomography (PET), and single photon emission computed tomography (SPECT), are being used investigatively in child and adolescent psychiatry. Clinical indications for the use of MRI or CT are to rule out signs of intracranial pressure or other neurological condition. PET scans and SPECT have been used to study blood flow patterns in children with attention-deficit/hypersensitivity disorder, but there is no current indication for their use diagnostically for children and adolescents.

Electroencephalography is also not of use diagnostically, especially since up to 15 percent of the general population displays abnormal EEG findings.

Laboratory testing is used routinely to monitor medication concentrations and adverse effects of medications. For example, lithium concentrations are used routinely, although the therapeutic range for children and adolescents has not been definitely demonstrated. Saliva testing is not recommended because of the wide range of saliva to serum ratios in children. If the patient is about to begin treatment with a tricyclic drug, regular monitoring of the electrocardiogram (ECG), pulse, and blood pressure is required. Anticonvulsants such as carbamazepine (Tegretol) must not be given until results of a complete blood count with platelets and reticulocyte counts and serum iron concentration are known. Other psychiatric medications including divalproex and clozapine (Clozaril) are also monitored with blood concentrations, although there is no clear-cut therapeutic window.

Psychological testing, including tests of intelligence quotient, are not useful diagnostically, although they may help to identify a child's relative strengths and weaknesses. Projective testing, such as the Thematic Apperception Test (TAT) or the Rorschach test may confirm depressive themes but cannot be used as a diagnostic tool.

DIFFERENTIAL DIAGNOSIS

A striking component of the differential diagnosis of mood disorders in children and adolescents is the likelihood of associated symptoms and comorbid psychiatric diagnoses. Multiple psychiatric disorders coexist more frequently in children and adolescents than in adults. The emergence of major depression in a child or adolescent increases the risk of an additional psychiatric diagnosis ten- to twentyfold.

The main psychiatric disorders that major depression must be distinguished from include anxiety disorders, conduct disorders, and substance use disorders. The main entities requiring clarification from mania and hypomania are attention-deficit/hyperactivity disorder and disruptive behavior disorders.

Many studies have indicated that anxiety symptoms are universal among children and adolescents with major depression and, conversely, children with anxiety disorders also exhibit depressive symptoms. Major depressive disorder in childhood is commonly accompanied by increased irritability, more fighting, and aggression. There are three possible chronologies leading to this constellation that must be elucidated by history. First, a child with a preexisting conduct disorder may develop a major depression. In fact, children with conduct disorder are at increased risk to develop depression. Second, some children and adolescents become irritable and explosive when the major depressive disorder occurs but do not go on to exhibit the typical violations of human rights that characterize conduct disorder. In this case, the conduct disturbance is likely to resolve when the depression is treated. The third scenario consists of full emergence of a conduct disorder concurrent with the depressive episode. This may occur in up to one third of children with major depression. Approximately one half of children with chronic school refusal and separation anxiety disorder meet all of the criteria for major depression. Thus, comorbidity of anxiety disorders and depressive disorders is frequent among youth.

A line must be drawn between acute depression and bereavement reactions and adjustment reactions with depressed mood. No clinical boundary separates these entities. Thus bereavement (as described in DSM-IV) consists of a syndrome that may include sadness, insomnia, and appetite and weight changes. The bereaved individual considers the reaction to be “normal.” A diagnosis of major depressive disorder is usually not given unless the symptoms are still present 2 months after the loss. The presence of certain symptoms—including guilt, thoughts about death, morbid pervasive preoccupation beyond the realm of “survivor” misgivings—may be designated as major depressive disorder. Additional symptoms including pervasive functional impairment, hallucinations, and recurrent desire to die indicate a depressive syndrome beyond bereavement. In adjustment disorder with depressed mood, the symptoms develop within 3 months of the onset of the stressor but do not fulfill criteria for a mood disorder diagnosis. Once the stressor has terminated, the symptoms do not persist for more than another 6 months.

The boundary between attention-deficit/hyperactivity disorder and hypomania or mania in children has been difficult to delineate. Even oppositional defiant disorder can be construed as overlapping clinically with mania and hypomania. Although long-term follow-up studies of children with attention-deficit/hyperactivity disorder do not indicate higher-than-expected rates of mania, the converse is true. That is, data suggest that prepubertal-onset bipolar children exhibit hyperactivity more often

than expected in the population by preschool age, and it continues until mania emerges during the grade school years. Attention-deficit/hyperactivity disorder is more prevalent in prepubertal-onset bipolar disorder children than in adolescent-onset bipolar disorder youths. Among bipolar disorder youths referred for mania, approximately 90 percent of prepubertal and 30 percent of adolescent bipolar disorder youth exhibit attention-deficit/hyperactivity disorder. According to the work of Barbara Geller and coworkers, hyperactivity may be the first developmentally age-specific manifestation of a bipolar disorder in the prepubertal age group, and although bipolar disorders and attention-deficit/hyperactivity disorder may coexist, it is difficult to detect this in young patients with a bipolar disorder.

Similarly, long-term follow-up of conduct disorder has not detected higher-than-expected prevalence rates of bipolar disorder, yet conduct disorder occurs in up to 22 percent of bipolar disorder children and up to 18 percent of bipolar disorder adolescents. Conduct disorder may also be construed as an early manifestation of prepubertal mania or hypomania, since the disturbance of conduct in these patients is linked to grandiose thinking, poor judgment, and poor impulse control.

During adolescence, however, the main differential diagnosis for bipolar disorder is with schizophrenia. This is often difficult in an adolescent who is floridly psychotic. The decision is easier in an adolescent with multiple family members with bipolar illness.

Substance abuse is not an uncommon confounding variable in the evaluation of symptoms that are typical of mania or hypomania. Substance abuse may also coexist with mania, and the effects of both may interact.

COURSE AND PROGNOSIS

Multiple long-term follow-up studies of children who have experienced major depression, among clinical samples and in the community, lead to the conclusion that major depression among youth is a serious, relapsing disorder with lingering social impairment and a high risk of suicide. The work of Maria Kovacs has underscored the long and sometimes tortuous course of various mood disorders in children and adolescents. Children who exhibited dysthymic disorder had a mean episode length of of 3 years, while those with major depression had a mean episode length of 32 months. While 92 percent of prepubertal children diagnosed with a major depressive disorder had recovered at 18 months, 89 percent of those with dysthymic disorder had recovered after 6 years. Sixty-nine percent of children with dysthymic disorder had a major depression within 5 years of diagnosis, and there was a 72 percent cumulative risk of relapse for children initially diagnosed with major depressive disorder. The earlier the age at diagnosis, the more protracted the recovery was for either major depression or dysthymic disorder. In the adult literature, it has been proposed that an initial episode of a depressive disorder sensitizes an individual to future episodes. Episode duration among children and adolescents with mood disorders has been studied with respect to age at onset, gender, psychiatric comorbidity, presence of severity indicators such as delusions, and family history. In adults, presence of dysthymic disorder at the onset of a major depression, “double depression,” predicts decreased length of episode, with the baseline being the dysthymic disorder. In a recent investigation of multiple variables influencing recovery from depression among youths, none of the baseline variables including gender, social class, severity of depression as measured by presence of delusional symptoms, or nonaffective psychiatric diagnoses influenced length of a major depressive episode. When a major depression was superimposed on dysthymic disorder, the length of the episode (recovery back to dysthymic disorder) was shortened, as was found with adults. Recovery from the first episode of dysthymic disorder was influenced by the presence of comorbid conduct disorder, oppositional defiant disorder, or attention-deficit/hyperactivity disorder. Factors that predict relapse and recurrence of major depression among adults include the severity of the index episode, psychotic features, earlier age of onset, multiple past episodes, and a history of double depression. In a recent follow-up of major depressive disorder among youths, the risk of recurrence was highest in the first year after recovery. Most recurrences occurred within 2 years of the index episode. More-severe depressive symptoms, older age, and presence of psychotic features increased the risk of recurrence.

An important possible sequela of major depression among children and adolescents is the development of mania and diagnosis of bipolar illness. The switch from depression to mania among children and adolescents is significantly higher than it is for adults. In the adult literature, between 5 and 18 percent of patients with depression become manic. Among prepubertal children with severe major depression, the rates of eventual mania have been reported to be as high as 32 percent within a 2- to 5-year period. Among adolescents, rates of mania between 20 and 40 percent have been reported within 5 years of the index depression. Among adolescents, severe depression characterized by psychomotor retardation, delusional symptoms, and pervasive anhedonia is associated with the highest risk for future manic episodes. Among young adults, the development of future hypomanic episodes (bipolar II disorder) has been associated with atypical depression, early onset of depression, seasonal affective disorder, and comorbid substance abuse. Among adolescents, a switch to hypomania may sometimes be misconstrued as a disruptive behavior disorder.

Additional complications of major depressive disorder in children and adolescents are the persistence of social impairment and the perception of being socially isolated. While social disturbance is not unique to mood disorders, depression in youth is associated with increased risk of tobacco and substance use and continued negative attributions, all of which affect social functioning.

One recent study of depressed children followed up at 36 weeks found that 50 percent still met criteria for major depressive disorder; of those, 73 percent had remained depressed, and 27 percent had recovered and relapsed. In this sample, major depression at follow-up was predicted by more-severe depression at intake, comorbid obsessive-compulsive disorder, and being older at the time of the index depression.

The course and prognosis of bipolar disorders among children and adolescents are clearer in adolescents than in prepubertal children since diagnosis is more clear-cut. A number of studies have indicated that an earlier age of onset (i.e., adolescent compared with young adult) of bipolar I disorder is associated with an increased frequency of cycles as well as an increased likelihood of mixed states. Other studies have suggested that the relapse rate for bipolar I disorder patients whose first episode is in their teens or early twenties is lower than that of bipolar I disorder patients who experienced a first episode in their thirties. Adolescent bipolar I disorder patients appear to have a higher risk of experiencing psychotic symptoms than patients who developed mania as adults. In a follow-up of adolescent bipolar I disorder patients, those whose index episode was purely manic or mixed recovered from the episode more rapidly than those whose index episode was depressive. Of this sample, 44 percent relapsed within 5 years and 21 percent of the sample manifested at least two relapses following recovery from the index episode. The likelihood of multiple relapses was highest for patients whose index feature was mixed episode or rapidly cycling. Among adolescents who develop manic episodes, those with a premorbid history of psychiatric disorders, including disruptive behavior disorders, had significantly more first-and second-degree family members with a history of affective illness.

Most adolescents with bipolar disorders experience recovery, although the length of episode is influenced by its polarity. Both youths and adults recover more rapidly from a pure manic episode than from a mixed or depressive episode. Among adolescents, however, an index episode of pure depression appears to be associated with a poorer prognosis than it does for adults with bipolar illness. Current evidence suggests that adolescent patients may be slower to relapse than adult patients, at least after their initial diagnosis is made.

Prepubertal-onset bipolar disorder may not follow an acute episode model with interepisode recovery. It has been proposed that prepubertal bipolar disorder may manifest with a continuous, rapid cycling of multiple, brief manic episodes. Adults whose illness is characterized by mixed episodes or rapid cycling have a poorer prognosis than bipolar I disorder adults with more-discrete episodes.

It is not clear whether a switch from a bipolar II to a bipolar I diagnosis in children is similar to that in adults; so far, low rates have been found for both adults and children. The question remains, however, whether an eventual bipolar I disorder diagnosis in children who start out with a bipolar II disorder diagnosis represents a switch of diagnostic category or simply the manifestation of bipolar I disorder that was unexpressed previously for developmental reasons. This distinction could have implications for treatment, since mood-stabilizer treatment might be more appropriate than antidepressant treatment in a child diagnosed with bipolar I disorder.

Suicide is a considerable risk for children and adolescents with any category of mood disorder. In one follow-up study of 28 adolescents hospitalized for an index episode of either depression or mania and followed for 10 years, only 5 of the 16 depressive disorder adolescents recovered and remained well. The other 11 depressive patients had a relapsing course over the 10 years. Of the 12 bipolar I disorder adolescents, 3 committed suicide, and the other 9 continued to have cycling episodes with poor adjustment. Some additional data suggest that the suicide rate among bipolar adolescents is higher than that of adolescents with other psychiatric disorders. The addition of substance use further increases the risk of suicide among adolescents with depressive and bipolar disorders. A review of adolescent suicide reported depressive disorders present in 43 to 76 percent of suicide completers, substance abuse in 26 to 66 percent of completers, and antisocial behaviors in 6 to 70 percent of completers.

A recent study reviewed the different rates and risks of suicide for younger adolescents (below 15 years of age), and older adolescents (15 to 19 years old) in Norway. Of suicide completers under age 15, 43 percent had psychiatric disorders, compared with 79 percent of adolescents over age 15. All but one of the completers under age 15 used hanging as the suicide method, whereas the older adolescents used firearms and hanging equally. Older adolescents were more likely to have expressed suicidal intent. Twenty-nine percent of the younger adolescents and 68 percent of the older adolescents had either expressed suicidal ideation, made a

previous attempt, or left a suicide note. An identifiable precipitating event was present in 49 percent of the older adolescents, compared with only 29 percent of the younger adolescent suicide completers. None of the younger adolescent completers was intoxicated with alcohol or other substances at the time of death, compared with 50 percent of the older adolescent completers. The suicide risk relating to mood disorder and family situation was almost identical for the younger and the older adolescent groups. The lower rates of suicide in the younger group may reflect the decreased likelihood of having an affective disorder, less exposure to stressors, or increased resiliency in the younger group. Overall, lower exposure to the risk factors for suicide, including mood disorders, substance use, and accessibility of firearms, contributes to the relative protection of younger adolescents from completing suicide.

Adolescents Who Kill A disturbing number of reports over the last few years have depicted the tragic killings of both students and adults by adolescents. More than 200 deaths have occurred in school yards across the United States within the last decade. In more recent years, these killings have been more frequent in rural settings and have involved multiple victims. In the early 1990s many of the school-related killings were attributable to gang activities or disputes over money or romantic involvements. In the last few years, multiple homicides perpetrated by adolescents have related more to issues of revenge for being mercilessly teased or have been of enactments scenarios of violence by adolescents who appear to be obsessed with violence and destruction.

While no unifying profiles identify adolescents who may commit murder, some potential similarities can be noted retrospectively in the histories of adolescents who have killed. When pieces of the histories of adolescent murderers are disclosed, it appears that most of these adolescents had long-standing troubles, some with major depression, they were often socially isolated, and they had explicitly displayed a fascination with violence. Some of the adolescent killers disclosed their inner agony fueled by perpetual tormenting by peers who ridiculed and rejected them daily. Some had experienced clear family dysfunction or circumstances of abuse, neglect, or parental psychiatric history that influenced the adolescent negatively. Many of the adolescents who killed had expressed the desire for violent revenge to parents, teachers, or peers. All of the adolescent killers reported ease in obtaining firearms, and many were expert in using them. In most cases, adolescents acquired semiautomatic weapons that were unlocked and kept in their homes within easy reach. Some of the adolescents gave literal warnings a day or a few weeks before the killings, either stating that there would be a violent attack or warning certain peers to stay away from school on a certain day. Others wrote poems, compositions, or journal entries depicting the violent scenes. Most of the adolescents who left notes at the time of the murders were suicidal and assumed that they would also die during the attack.

Suicide rates among youths have been on the increase, and rates of depression have also increased in adolescents in younger age groups. A depressed, demoralized, socially rejected adolescent male preoccupied with violent fantasies, with easy access to semiautomatic weapons, is a potential disaster. Any violent threat must be taken seriously, and families must be taught to keep weapons away from these adolescents.

TREATMENT

Psychological Therapies

Cognitive-Behavioral Therapy Cognitive-behavior therapy is a problem-oriented treatment that targets cognitive distortions, faulty attributions, and poor self-esteem, all of which commonly plague children and adolescents with mood disorders. The treatment aims to diminish punitive attitudes and self-recrimination and focuses increased attention on positive attributes and improved social competence. Although negative attributional style is not unique to depressive disorders, it is pervasive among depressed adults and adolescents. Controlled outcome studies with adults have been promising, suggesting that cognitive-behavioral approaches may be effective in treating major depressive disorder in diminishing rates of relapse, and (perhaps) in preventing recurrence of depression. There have been relatively few efficacy studies of cognitive-behavioral interventions for children and adolescents with mood disorders.

A recent review of controlled cognitive-behavioral studies in children and adolescents led to the following conclusions. To be reviewed, the cognitive-behavioral intervention study must have met the definition of the treatment as an intervention promoting emotional and behavioral change by teaching children new coping styles and new thoughts in a problem-oriented manner. All but one of the six studies reviewed recruited their dysphoric subjects from schools and used a group therapy format. Two of the studies used adolescents formally diagnosed with major depression and used structured interviews as part of the diagnostic process. The period of treatment ranged from six to fourteen sessions, and the time periods ranged from 5 to 8 weeks. The control conditions included relaxation intervention, being placed on a waiting list, family therapy, supportive psychotherapy, self-control therapy, and self-modeling therapy. Each study used one or more of a number of depression rating scales given pretreatment and posttreatment, including the Beck Depression Inventory (BDI), the Reynolds Adolescent Depression Scale (RADS), the Children's Depression Inventory (CDI), the Children's Depression Scale (CDS), the Center for Epidemiologic Studies Depression (CESD) Scale, or the Mood and Feeling Questionnaire (MFQ). The total number of subjects in these studies was 217. The posttest ratings indicated significantly diminished symptoms of dysphoria and depression among adolescents. Follow-up posttreatment ranged from 1 to 24 months. Most studies were followed up for approximately 3 months. The overall posttest ratings at follow-up was also significant, indicating that the improvements made after the cognitive-behavior treatment was completed had been maintained. The study comparing cognitive-behavioral therapy with supportive psychotherapy found cognitive-behavior therapy to be superior. The studies that included a relaxation treatment, a self-control treatment, or a self-modeling treatment found that each of the active treatments was associated with improvement in depressive symptoms. In studies that included a wait-list group, all of the active treatments yielded better results than the wait-list group.

The improvements shown by adolescents after cognitive-behavioral therapy are similar to those of depressed adults. Cognitive therapy aims to modify maladaptive beliefs and attempts to enhance problem-solving abilities, social competency, and the ability to achieve a sense of mastery in social and occupational domains. Although the studies reviewed did not demonstrate a highly specific mode of employing these techniques, the positive yield was notable. The translation of these improvements into units of increased functionality and adaptability requires further study. Most subjects in these studies were “dysphoric,” and thus this review could not aid in determining whether cognitive-behavioral strategies are efficacious in more severely ill adolescents with major depressive disorder or bipolar I disorder. A recent report of cognitive-behavioral therapy for depressed adolescents did ask which patients are more likely to benefit from this modality of treatment. This study found that two variables predicted response to the cognitive-behavioral approach: age and severity. Younger patients and those with less severe depression responded better to a relatively brief trial of cognitive-behavioral therapy. The cognitive-behavioral strategy was not compared with some alternative treatments including pharmacotherapy.

A brief cognitive-behavioral treatment program called Successful Negotiation Acting Positively (SNAP) created by Mary Jane Rotheram-Borus and coworkers has been used in adolescent suicide attempters and is initiated at the time of their emergency room evaluation for the suicidal behavior. The treatment consists of a six-session, outpatient, highly structured program aimed at diminishing cognitive distortions and improving problem-solving strategies in families of adolescents who have attempted suicide. This program uses several operationalized components, including tokens given to each family member to pass to another whenever a positive feeling or an agreement has been reached. Role playing is used to practice problem solving from someone else's point of view. A schematic “feeling thermometer” is also used to help adolescents and their families identify their feelings. The therapist reframes family disagreements into forms that can be solved more effectively. Initiating this form of treatment at the time of the suicide attempt in the emergency room setting and with additional components of urgent treatment improved attendance at follow-up visits. Cognitive-behavioral treatment for adolescents and their families focuses on identifying and clarifying beliefs and notions regarding the roles and responsibilities of both parents and the adolescents.

Group settings often employ cognitive-behavioral techniques to treat depression in children and adolescents. Group therapy situations may focus on social skills improvement and social learning to diminish depressive symptoms. One protocol, called the Coping with Depression course for adolescents also uses a classroom format containing structured social learning tasks with assignments to be done at home. Although not every depressed child is willing to participate in such a treatment, participants have shown improvement, compared with waiting-list controls. A parallel intervention for parents led to a decrease in parental ratings of the child's behavior problems but was not associated with further decrease in depressive symptoms, as rated by the children.

Interpersonal Therapy *Interpersonal psychotherapy*, a brief individual intervention that focuses on the patient's social relationships and current evaluation of these relationships, was originally developed for adults with depression. The therapy aims to decrease depressive symptoms while improving interpersonal functioning. Interpersonal therapy emerges from the notion that depression is expressed within a social context and that interpersonal relationships with significant people influence the course of the depressive disorder. The assumption is that as communication improves within significant relationships, depressive symptoms diminish. At least five published controlled trials of interpersonal therapy used in depressed adults have found it efficacious in acute and maintenance treatment. There are no published controlled trials at this time of interpersonal therapy for depressed adolescents, but one controlled trial is in progress. In this trial, a standardized protocol consisting of 12 weeks of individual sessions is used. The first phase of treatment is aimed at identifying interpersonal patterns that promote depressive cognitions. During this phase, the therapeutic alliance is developed, with specific goals set for the treatment. In the second phase of treatment, strategies are developed to alter interpersonal communications. The final phase of treatment involves practical application of the interpersonal strategies. A recent publication reported the outcome at 1 year of depressed adolescents who received 12 weeks of interpersonal therapy using a manual. Sessions were held individually on a weekly basis. Of the 10 adolescents who were followed up, only 1 met the criteria for a mood disorder. In this follow-up, the improvements that had occurred by the end of the treatment period had been maintained for the next year, with only one subject depressed at follow up. The other subjects reported few depressive symptoms and consistent

social functioning. It remains to be determined whether interpersonal therapy can be protective with regard to responses to adverse life events. As mentioned above, interpersonal therapy presumes that whatever the cumulative causes of depression, improved significant relationships will diminish the depressive symptoms. Whether successful gains in interpersonal relationships are prophylactic for relapse or recurrence of depression remains to be seen.

There have been scant studies of a variety of additional psychotherapeutic interventions including psychodynamic therapy, supportive psychotherapy, and relaxation techniques. In most studies comparing any of the psychosocial intervention groups to a waiting-list control, the psychosocial interventions have been associated with greater improvement in depressive symptoms.

Family-Focused Psychoeducational Treatment The emotional attitudes of family members of psychiatrically ill adults, particularly those with schizophrenic disorder, influence the course and relapse rate of the disorder. Many studies of adults have found that schizophrenics whose families exhibit high expressed emotion (i.e., are highly critical, hostile, or overly involved) are at higher risk for relapse. At least two studies of manic and mixed episode young adults have shown that those residing with families showing high expressed emotion were more likely to relapse within a year and to show lower global functioning in the short term. Thus, family-focused psychoeducational treatment is aimed at helping families find an equilibrium after an affective episode in one family member and reducing the traumatic impact of the illness on the family. The components of this treatment include assessment of the family milieu, psychoeducation for the patient and the family about the course of bipolar disorder, and both communication enhancement and problem-solving-skills training. One model of this treatment proposes 21 outpatient sessions including 12 weekly sessions, 6 biweekly sessions, and 3 monthly sessions after an acute episode. Communication enhancement, the part of the treatment that approaches the expressed emotion in the family, consists of developing mastery in four areas: active listening, delivering positive feedback, delivering negative feedback, and requesting changes in other family members' behavior.

Psychopharmacotherapy Current antidepressant treatment of mood disorders in children and adolescents is still in the early phases of being validated with double-blind efficacy studies. Thus far, no studies have supported the use of tricyclic drugs in children or adolescents, and at least one published placebo-controlled double-blind study exists comparing fluoxetine's efficacy for depressive symptoms in adolescents with that of placebo.

The tricyclic drugs have shown efficacy in multiple adult studies of mood disorder, but since the advent of antidepressants with safer adverse-effect profiles (i.e., minimal risk of cardiac arrhythmias and considerably lower lethal potential), the tricyclic drugs are usually not among the first-choice antidepressants for children and adolescents.

Selective serotonin reuptake inhibitors (SSRIs) have become the drugs of choice for mood disorders in children and adolescents because of promising research findings and the favorable profiles of the drugs. Serotonin reuptake inhibitors include fluoxetine, paroxetine (Paxil), sertraline, fluvoxamine (Luvox), citalopram (Celexa), and nefazedone (Serzone). Nefazedone blocks norepinephrine receptors and serotonin receptors. These drugs all have a relatively mild adverse-effect profile, with low lethality after overdose. Dosage-related adverse effects include headache and nausea. Except nefazedone all these drugs can cause sexual dysfunction including diminished libido, delayed ejaculation, or anorgasmia. There have been no indications thus far that plasma level can be correlated to expected response. In the double-blind study by Graham Emslie, response to the SSRIs seemed to be equivalent in children and older adolescents. Additional potential adverse effects of these drugs include insomnia and a syndrome of restlessness and agitation that can be ameliorated by decreasing the dosage.

Additional antidepressants that can be used include bupropion (Wellbutrin), a dopaminergic drug; venlafaxine (Effexor), which blocks serotonin and norepinephrine reuptake, mirtazepine (Remron), a serotonin plus norepinephrine reuptake inhibitor; and the monoamine oxidase (MAO) inhibitors phenelzine (Nardil) and tranylcypromine (Parnate). Bupropion is an antidepressant with stimulant properties that has few anticholinergic side effects, produces virtually no sedation, and is much safer in overdose than the tricyclic drugs. Because of its stimulant properties, it has also been used to treat attention-deficit/hyperactivity disorder. Venlafaxine is a relatively new antidepressant that largely shares an adverse-effect profile with the SSRIs; the most common adverse effects include nausea, anorexia, nervousness, and sexual dysfunction. Despite its potential to induce nervousness, there have been anecdotal reports of its efficacy in the treatment of anxiety. There have been no studies of mirtazapine, a new antidepressant similar to nefazedone in mechanism of action. Mirtazapine is significantly more sedating than nefazedone.

MAO inhibitors have been sparsely studied in children and adolescents although they have been shown to be effective in adults with major depressive disorder and those with atypical depression. In one published open trial of phenelzine with adolescents who were refractory to treatment with tricyclic drugs the treating physician judged that up to 60 percent of these adolescents were improved by the phenelzine treatment. Since the monoamine oxidase inhibitors require strict adherence to a restricted diet (no cheese [except cottage cheese or cream cheese], processed meats, caviar, cured fish, overripe fruits, avocados, fava beans, yeast extracts, chianti or burgundy wine, beers containing yeast, or over-the-counter cold preparations, especially decongestants and inhalers). Aside from these restrictions (which potentially result in hypertensive crises and death if ingested), other adverse effects of the drugs include orthostatic hypotension and sleep disturbances. The MAO type A inhibitor moclobemide (available in Canada, but not in the United States) is a safer drug with a milder adverse-effect profile, but it has not yet been studied with children or adolescents. In children or adolescents with major depressive disorder there is a risk of precipitating a hypomanic or manic episode during the course of any antidepressant treatment.

Although there has been only one double-blind placebo-controlled study showing the efficacy of lithium for bipolar children and adolescents, pharmacokinetic studies of lithium in children and adolescents with behavior disorders indicate that lithium can be used with the same safety precautions used for adults. As with adults, electrolytes, renal function tests, thyroid function tests, and calcium and phosphorus concentrations should be monitored at least twice a year. Lithium has a shorter half-life in children than in adults, consistent with a higher efficiency of the kidneys in children. Treating hypomania or mania in adolescents applies the same basic strategies used for adults. Initiation of a psychopharmacological treatment in a manic adolescent often requires hospitalization if the patient cannot be contained or follow directions reliably. For adolescent patients with pure mania, lithium is generally the drug of choice. It has demonstrated efficacy in the treatment of manic states in adults and has shown some prophylactic efficacy for future mood episodes in adults, more strongly for manic episodes than for depressed episodes. Since there is controversy regarding inclusion of a variety of cycling and noncycling severe mood and behavioral disturbances in children and adolescents, the studies done in this age group have no doubt included a heterogeneous diagnostic group. Nevertheless, in the published open-trial reviews of lithium treatment of youths, the efficacy of lithium for behavioral improvement has ranged from about 50 to 66 percent. This is substantially lower than reported lithium efficacy in manic adults. Clinical features that may predict poor responsivity to lithium include mixed mood states and atypical features.

Prior to starting lithium treatment, a number of laboratory indexes must be documented to aid in monitoring adverse effects of lithium (Table 45-3). Lithium treatment should be initiated with 300 or 450 mg and be titrated up to achieve a concentration in blood of approximately 0.8 to 1.2 mEq/L. The most common adverse effects are nausea, diarrhea, stomach ache, fine tremor of the hands, polyuria, polydipsia, and weight gain. Signs of lithium toxicity include drowsiness, slurred speech, ataxia, muscle tremors, and hyperreflexia. Children on lithium also run the risk of cognitive slowing. The reliability of the family must be assessed regarding follow-up of lithium concentrations and laboratory studies before lithium treatment is initiated in a child or adolescent.

Laboratory Test	Reason
Serum electrolytes	Hypokalemia may increase risk of cardiac arrhythmias; hyponatremia can decrease lithium excretion resulting in lithium toxicity
BUN, creatinine	If renal function is impaired lithium may not be excreted
Blood count	Lithium may cause leukocytosis
ECG	Lithium can cause T-wave inversion or conduction abnormalities
Thyroid function tests (TSH)	Lithium can cause hypothyroid function or euthyroid goiter
Pregnancy test	Lithium is associated with first-trimester anomalies

Table 45-3 Laboratory Testing Prior to Lithium Treatment

Anticonvulsants have been shown to be efficacious in adult manic patients; there have been no double-blind studies to indicate efficacy in adolescents. Naturalistic follow-up data suggest that anticonvulsants may be efficacious for adolescents with mixed-state episodes or those with rapid cycling disorders. In studies of adult bipolar patients refractory to lithium treatment, a significant subgroup responded to carbamazepine, with or without adjunctive lithium. Hypothetical models of mood cycling exhibiting a brain-kindling effect support the use of anticonvulsants to interrupt this process. While carbamazepine and valproic acid have not been studied

systematically in children or adolescents with mania, they have both been used widely in the treatment of seizure disorders in children. There have been a few anecdotal reports of behavioral problems diminishing in epileptic children treated with carbamazepine. Several open clinical trials of both carbamazepine and valproic acid for adolescents with mania have suggested that these medications may be useful, especially in manic children or adolescents who do not respond to lithium. Both carbamazepine and valproic acid have potentially serious adverse effects. Carbamazepine requires weekly blood counts within the first month of treatment to follow the white blood cell count, since carbamazepine has been associated with aplastic anemia, thrombocytopenia, eosinophilia, and agranulocytosis. Valproic acid may be associated with a variety of adverse effects including nausea, rash, drowsiness, weight gain, and hair loss. Valproic acid is also associated with development of polycystic ovary disease in young women. Data comparing epileptic females taking valproic acid with those on other drugs indicate that the rate of polycystic ovary disease among young women on valproic acid was three times that of women who were not exposed to it, and rates of disease were close to 90 percent. These serious adverse effects must be considered before initiating a trial of valproic acid treatment in females.

Antipsychotic medications are sometimes indicated for mood disorders, including both depressive and bipolar disorders in adolescents with psychotic features. The newer antipsychotic agents such as risperidone (Risperdal), olanzepine (Zyprexa), and clozapine are often chosen as first-line adjunctive drugs, given their diminished extrapyramidal adverse effects. While clozapine has been suggested to be the most effective antipsychotic agent for pure mania, its risk for agranulocytosis and the need for weekly blood count monitoring make it less desirable. Antipsychotics are usually not effective as monotherapy for a mood episode, and additional strategies involving augmentation strategies and additional mood stabilizers are generally the next step for a refractory mood episode.

The use of multiple mood stabilizers has successfully controlled some resistant rapid-cycling disorders in adults. Lithium and carbamazepine have been reported to be adequately tolerated in the above situations. The addition of an anticonvulsant to lithium increases the probability of such effects as drowsiness, dizziness, and ataxia. On the other hand, the combination of carbamazepine with a smaller, adjunctive dose of lithium has also been successful in some patients with refractory bipolar disorders. Two open trials with adolescents have used lithium adjunctively in the treatment of major depressive disorder.

Additional agents reported to have efficacy in treating refractory mania in adults are calcium channel blockers (e.g., nimodipine [Nimotop]), other anticonvulsants (e.g., lamotrigine [Lamictal], gabapentin [Neurontin], and topiramate [Topamax]), and thyroid hormone augmentation. Open trials and several double-blind trials have reported L-type calcium channel inhibitors such as nimodipine to be effective in controlling mania in either refractory or rapid-cycling bipolar illness. Other calcium channel inhibitors such as verapamil (Calan, Isoptin) may not share the same antimanic properties. Lamotrigine has been used adjunctively in patients who are already on other mood stabilizers, and it has also been tried as monotherapy in adults. One double-blind study of lamotrigine as monotherapy reported a 48 percent response rate, with a placebo response rate of 5 percent. Lamotrigine is associated with pruritic rash in up to 10 percent of patients, which may be related to the rate of increase during drug titration. Of more concern is that 1 out of 500 patients may develop a potentially life-threatening skin condition called Stevens-Johnson syndrome or Lyell's syndrome. The metabolism of lamotrigine is influenced by other anticonvulsants being used; combined with valproic acid, lamotrigine's concentration may be doubled, and combined with carbamazepine, lamotrigine's concentration may be reduced by half. Although lamotrigine has been shown to be efficacious in mania, patients who responded to lamotrigine still required augmentation of the lamotrigine.

Gabapentin is an anticonvulsant believed to increase the concentration in the brain g-aminobutyric acid (GABA), the inhibitory neurotransmitter. A short-term double-blind trial of gabapentin indicated that it had an antimanic effect on approximately one third of patients, while placebo response was 5 percent. Gabapentin is reportedly well tolerated and has been used adjunctively as well as a monotherapy. Gabapentin is given in divided doses to maximize absorption and has been reported to improve sleep and anxiety symptoms.

Thyroid augmentation, using liothyronine (Cytomel) or levothyroxine (Levoxyl, Synthroid) has been used in combination with lithium in the treatment of both depressive and bipolar disorders. This strategy has also been used in refractory rapid-cycling bipolar disorder. Investigations are in progress to determine whether thyroid augmentation is effective in the prophylaxis of bipolar disorder. For levothyroxine augmentation, up to 70 percent of women showed improvement with the addition of the agent, compared with 20 percent of men, while for liothyronine augmentation, 44 percent of women responded, compared with 10 percent of men.

However, these augmentation strategies remain to be studied systematically in children and adolescents.

Electroconvulsive Therapy There have been a number of reports of electroconvulsive therapy (ECT) use in the adolescent population during the last decade. It has been described as a potentially efficacious treatment in adolescents for mood disorder characterized by catatonia, psychosis, and melancholic depression. A recent epidemiological survey of ECT among adolescents in the Australian state of South Wales found that only 42 patients between the ages of 14 and 18 years were administered a total of 49 courses of ECT comprising 450 treatments: 29 percent had major depressive disorder without psychosis, 29 percent met criteria for psychotic depression, 6 percent had mania, 12 percent had schizophrenia complicated by depression, 6 percent had schizophrenia without catatonic symptoms, 6 percent had schizophrenia with catatonia, 8 percent had schizoaffective disorder, and 4 percent had psychosis not otherwise specified. Several patients had neuroleptic malignant syndrome during the course of ECT. The number of treatments given ranged from 4 to 21. Electrode placement was unilateral on the right side in 64 percent of the courses, bilateral in 11 percent, and mixed in 25 percent. In patients who responded, response was evident in an average of 4.4 days, or after 2.8 treatments. The overall response rate to ECT in this review was 51 percent. Overall symptoms were rated significantly improved in 38 percent of patients with a diagnosis of major depression, 85 percent of patients with a diagnosis of psychotic depression, 100 percent of those with mania, 16 percent of those with schizophrenia complicated by depression, 33 percent of those with noncatatonic schizophrenia, 66 percent of schizophrenics with catatonia, and 50 percent of those with psychosis not otherwise specified. Depressive symptoms diminished irrespective of a diagnosis of major depression. Psychotic symptoms were more responsive to ECT in patients with mood disorders. Adverse effects consisted of headache in 65 percent of the courses, memory problems in 22 percent, confusion in 18 percent, muscle aches in 12 percent, and manic switch in 4 percent. Adolescents tended to have longer seizures than those reported for adults. Other reviews have indicated similar reports of adverse effects. Case studies of ECT among adolescents over a period of time at several university teaching hospitals corroborate the efficacy and overall safety of ECT in adolescents. Mood disorders in adolescents, characterized by psychomotor abnormalities, psychotic features, catatonia, and marked cognitive impairment may respond favorably to ECT. Given the rapidity of response and the relative safety of the procedure, ECT should be considered for treatment of severe mood disorders among adolescents.

Light Therapy Light therapy has been found effective in the treatment of depressive disorders with seasonal pattern in adults. There have been sparse reports of light treatment for children and adolescents with seasonal or other mood disorders. A recent double-blind placebo-controlled trial of light therapy was administered to 28 children and adolescents. The trial used a crossover design of 1 week of active treatment (1 hour of bright light and then 2 hours of dawn simulation) or placebo (1 hour of clear goggles plus 5 minutes of low-intensity dawn simulation, and then the reverse). Children initially underwent a baseline week in which they wore dark glasses for 1 hour a day. They were then randomized to active treatment or placebo. Following this, they entered another 1- to 2-week phase with dark glasses, before beginning the alternative phase. Active treatment consisted of bright light (2,500 lux for children under 9 years, and 10,000 lux for children older than 9 years) administered between 4 pm and 8 pm and dawn simulation (250 lux) at 6:30 am. Depressive symptom scores were significantly decreased from baseline and washout phases during the active treatment phase, according to parent report. Based on the parent report, 71 percent of the children had at least a 50 percent reduction of depressive symptoms during the active phase, while 25 percent reported reduction in the placebo phase. Although not significant, a trend toward diminished depressive symptoms during the active phase appeared in the child's report. Active treatment was not associated with more-frequent side effects than the placebo phase. At the end of the study, 78 percent of the parents and 80 percent of the children rated the active phase of treatment as the one in which they felt “best,” and almost all of the subjects opted to continue light therapy after the study had finished. The prevalence of seasonal affective disorder has recently been found to be between 3 and 4 percent, so it is important to continue investigation of light therapy.

SUGGESTED CROSS-REFERENCES

Chapter 14 provides an exhaustive discussion of mood disorders, and suicide is discussed further in Section 29.1.

CHAPTER REFERENCES

Beautrais AL, Joyce PR, Mulder RT: Precipitating factors and life events in serious suicide attempts among youth aged 13 through 24 years. J Am Acad Child Adolesc Psychiatry *36:* 1543, 1997.

Birmaher B, Ryan ND, Williamson DE, Brent D, Kaufman J: Childhood and adolescent depression: A review of the past 10 years. Part II. J Am Acad Child Adolesc Psychiatry *35:* 1575, 1996.

*Birmaher B, Ryan ND, Williamson DE, Brent D, Kaufman J, Dahl R, Perel J, Nelson B: Childhood and adolescent depression: A review of the past 10 years. Part I. J Am Acad Child Adolesc Psychiatry *35:* 1427, 1996.

Brent D, Moritz G, Bridge J, Perper J, Cannobio R: Long-term impact of exposure to suicide: A three-year controlled follow-up. J Am Acad Child Adolesc Psychiatry *35:* 646, 1996.

Casat CD, Arana GD, Powel K: The DST in children and adolescents with major depressive disorder. Am J Psychiatry *140:* 503, 1989.

*Emslie GJ, Rush AJ, Weinberg WA, Kowatch RA, Hughes CW, Carmody T, Rintelman J: A double-blind randomized placebo-controlled trial of fluoxetine in depressed children and adolescents. Arch Gen Psychiatry *54:* 1031, 1997.

*Emslie GJ, Walkup JT, Pliszka SR, Ernst M: Nontricyclic antidepressants: Current trends in children and adolescents. J Am Acad Child Adolesc Psychiatry *38:* 517, 1999.

Geller B, Cooper TB, Sun K, Zimerman B, Frazier J, Williams M, Heath J: Double-blind and placebo-controlled study of lithium for adolescent bipolar disorders with secondary substance dependency. J Am Acad Child Adolesc Psychiatry *37:* 171, 1998.

*Geller B, Luby J: Child and adolescent bipolar disorder: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *36:* 1168, 1997.

Goodyer IM, Herbert J, Tamplin A, Secher SM, Pearson J: Short-term outcome of major depression: II. Life events, family dysfunction, and friendship difficulties as predictors of persistent depression. J Am Child Adolesc Psychiatry *36:* 474, 1997.

Gould MS, Shaffer D, Fisher P, Garfinkel R: Separation/divorce and child and adolescent completed suicide. J Am Acad Child Adolesc Psychiatry *37:* 1998.

Groholt B, Ekeberg O, Wichstrom L, Haldorsen T: Suicide among children and younger and older adolescents in Norway: A comparative study. J Am Acad Child Adolesc Psychiatry *37:* 473, 1998.

Jayson D, Wood A, Kroll L, Fraser J, Harrington R: Which depressed patients respond to cognitive-behavioral treatment? J Am Acad Child Adolesc Psychiatry *37:* 35, 1998.

Kaplan SJ, Pelcovitz D, Salzinger A, Mendel F, Weiner M: Adolescent physical abuse and suicide attempts. J Am Acad Child Adolesc Psychiatry *36:* 799, 1997.

Kovacs M, Obrosky DS, Gastonis C, Richards C: First-episode major depressive and dysthymic disorder in childhood: Clinical and sociodemographic factors in recovery. J Am Acad Child Adolesc Psychiatry *36:* 777, 1997.

*Leonard HL, March J, Rickler KC, Allen AJ: Pharmacology of the selective serotonin reuptake inhibitors in children and adolescents. J Am Acad Child Adolesc Psychiatry *36:* 725, 1997.

McMahon FJ, DePaulo R: Genetics and age of onset. In *Mood Disorders Across the Life Span*, KI Shulman, M Tohen, S Kutcher, editors. Wiley, New York, 1996.

Miklowitz DJ: Psychosocial approaches to the course and treatment in bipolar disorder. CNS Spectrums Int J Neuropsychiatr Med *3:* 48, 1998.

Moise F, Petrides G: Case study: Electroconvulsive therapy in adolescents. J Am Acad Child Adolesc Psychiatry *35:* 312, 1996.

Mufson L, Fairbanks J: Interpersonal psychotherapy for depressed adolescents: A one-year naturalistic follow-up study. J Am Acad Child Adolesc Psychiatry *35:* 1145, 1996.

Pataki CS, Carlson GA: Childhood and adolescent depression: A review. Harvard Rev Psychiatry *3:* 140, 1995.

Pfeffer CR, Martins P, Mann J, Sunkenberg M, Ice A, Damore JP Jr, Gallo C, Karpenos I, Jiang H: Child survivors of suicide: Psychosocial characteristics. J Am Acad Child Adolesc Psychiatry *36:* 65, 1997.

Pike A, Plomin R: Importance of nonshared environmental factors for childhood and adolescent psychopathology. J Am Acad Child Adolesc Psychiatry *35:* 560, 1996.

Post RM, Frye MA, Leverich GS, Denicoff KD: The role of complex combination therapy in the treatment of refractory bipolar illness. CNS Spectrums Int J Neuropsychiatr Med *3:* 66, 1998.

*Reinecke MA, Ryan N, DuBois DL: Cognitive-behavioral therapy of depression and depressive symptoms during adolescence: A review and meta-analysis. J Am Acad Child Adolesc Psychiatry *37:* 26, 1998.

Rotheram-Borus MJ, Piacentini J, Von Rossem R, Graae F, Cantwell C, Castro-Blanco D, Miller S, Feldman J: Enhancing treatment adherence with a specialized emergency room program for adolescent suicide attempters. J Am Acad Child Adolesc Psychiatry *35:* 654, 1996.

*Ryan ND, Bhatara VS, Perel JM: Mood stabilizers in children and adolescents. J Am Acad Child Adolesc Psychiatry *38:* 529, 1999.

Shaffer D, Piacentini J: Suicide and attempted suicide. In *Child and Adolescent Psychiatry*, ed 3, M Rutter, L Hersov, E Taylor, editors. Blackwell Scientific, London, 1994.

Shulman KI, Tohen M, Kutcher SP, editors: *Mood Disorders Across the Life Span*. Wiley, New York, 1996.

Stein D, Apter A, Raztoni G, Har-Even D, Avidan G: Association between multiple suicide attempts and negative affects in adolescents. J Am Acad Child Adolesc Psychiatry *37:* 488, 1998.

Stein MA, Szumowski E, Ravitz A, Frey MJ, Leventhal BL: Dexamethasone suppression and childhood depression: Association with categorical and dimensional measure. J Child Adolesc Psychopharmacol *4:* 43, 1994.

Strober M, Pataki C, DeAntonio M: Complete remission of “treatment resistant” severe melancholia in adolescents with phenelzine: Two case reports. J Affect Disord *50:* 55, 1998.

Swedo S, Allen AJ, Glod C, Clark CH, Teicher MH, Richter D, Hoffman C, Hamburger SD, Dow S, Brown C, Rosenthal NE: A controlled trial of light therapy for the treatment of pediatric seasonal affective disorder. J Am Acad Child Adolesc Psychiatry *36:* 816, 1997.

Walter G, Rey JM: An epidemiological study of the use of ECT in adolescents. J Am Acad Child Adolesc Psychiatry *36:* 822, 1997.

*Williamson DE, Birmaher B, Anderson BP, AL-Shabbout M, Ryan ND: Stressful life events in depressed adolescents: The role of dependent events during the episode. J Am Acad Child Adolesc Psychiatry *34:* 591, 1995.

Zametkin A, Ernst M, Silver R: Laboratory and diagnostic testing in child and adolescent psychiatry: A review of the last 10 years. J Am Acad Child Adolesc Psychiatry *37:* 464, 1998.

CHAPTER 46. ANXIETY DISORDERS IN CHILDREN

46.1 OBSESSIVE-COMPULSIVE DISORDER IN CHILDREN

JOHN PIACENTINI, PH.D., AND R. LINDSEY BERGMAN, FH.D.

Until the last decade, obsessive-compulsive disorder (OCD) was thought to be rare in children and adolescents, and only limited attention was focused on this area. Over the past 10 years, however, several factors, including age-of-onset data from adult studies, the development of effective treatments, and more sophisticated assessment and screening methodologies, have led to significantly heightened professional and public awareness of childhood OCD and substantial upward revision in the prevalence rate for the disorder in this age group. Although significant strides have been made in advancing the scientific understanding and treatment of OCD in childhood, pediatric OCD remains less well studied than its adult counterpart. Like the adult disorder, childhood OCD is a chronic and distressing disorder that can severely impair social, academic, and family functioning. However, the continued refinement of current treatment approaches has increased the likelihood that many youngsters with OCD will be able to lead productive, satisfying, and relatively normal lives.

DEFINITION

According to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), *OCD* is defined as the presence of either obsessions or compulsions that "cause marked distress, are time consuming, or significantly interfere with the person's normal routine... or functioning." *Obsessions* are defined as "recurrent or persistent thoughts, impulses, or images" that are experienced as intrusive or inappropriate, are not simply excess worries about real-life problems, and which cause marked anxiety or distress. They must be recognized as products of one's own mind (to differentiate them from thought insertion). Common obsessions in childhood include fears of harm coming to self or others, fear of harming others, and contamination fears. *Compulsions* are repetitive behaviors or mental acts that are performed in response to an obsession or according to some other rigidly applied rules. Compulsions are meant to reduce anxiety or distress or prevent some dreaded event, but they are clearly excessive or are not realistically connected with the triggering stimulus. Typical compulsions in childhood include excessive handwashing, cleaning, checking, counting, and arranging.

HISTORY

Descriptions of one OCD variant, scrupulosity (e.g., groundless fears of moral, religious, or ethical transgression) date back to religious literature from the sixteenth century. In fact, the first survey of OCD symptoms in childhood was conducted in 1927 by the Catholic Church, which assessed 400 female Catholic high school students and identified 17 (4 percent) as overly scrupulous. The first description of OCD in childhood was provided by Pierre Janet in 1903 who reported on a 5-year-old boy with characteristic symptoms. Leo Kanner in 1935 noted the social isolation of OCD youngsters and the high degree of family overinvolvement in their child's rituals. In 1955, Louise Despert noted several key facts that remain characteristic of current thinking on the disorder: a male preponderance, children's recognition of their symptoms as abnormal and unwanted, and the tendency of these youngsters to hide their OCD symptoms from treatment professionals and others.

Early studies of both epidemiological and clinical samples supported the notion that OCD was rare in children and adolescents. The landmark Isle of Wight study of over 2000 English school children in the 1960s reported a 0.3 percent prevalence rate for "obsessional/anxiety disorder," and similar rates were derived from a handful of large, retrospective clinic-based chart review studies. In the mid-1980s, findings from the Epidemiologic Catchment Area (ECA) study that most adults with OCD report an onset of the disorder by mid- to late-adolescence significantly increased both the awareness of the disorder in childhood and research activity in this area. Much of the heightened professional and public awareness of childhood OCD was due to the efforts of a group of researchers at the National Institute of Mental Health (NIMH), led by Judith Rapoport. This group was also responsible for the first systematic studies of the epidemiology, phenomenology, and psychopharmacological treatment of OCD in this age group.

COMPARATIVE NOSOLOGY

The diagnosis of OCD has remained relatively constant from the third edition of DSM (DSM-III) to DSM-IV with minor exception. Compared with DSM-III, the revised third edition (DSM-III-R) provided stricter operationalization of the interference criterion (i.e., symptoms must take more than 1 hour per day). A more significant change was the reclassification of mental acts (e.g., praying, counting, mental organization) in DSM-IV from obsessions to compulsions. Although formerly considered obsessions because of their cognitive nature, these mental activities were reclassified as compulsions in DSM-IV to reflect their intended function as a means of reducing obsession-related anxiety. DSM-IV also added a specifier to code degree of insight and, consistent with the general DSM-IV framework, excluded disturbance directly due to a general medical condition.

The definition of the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) of OCD is phenomenologically similar to that for DSM-IV although less detailed and not as well operationalized. ICD-10 describes five subtypes of OCD: (1) predominantly obsessional thoughts or ruminations, (2) predominantly compulsive acts (obsessional rituals), (3) mixed obsessional thoughts and acts, (4) other obsessive-compulsive disorders, and (5) obsessive-compulsive disorder, unspecified. Unlike DSM-IV, ICD-10 does not preclude excess worries about real-life problems as obsessions or specifically address mental acts (as either obsessions or compulsions). Similar to DSM-IV, ICD-10 describes the function of compulsions as a means of preventing some objectively unlikely event. However, it does not go as far as DSM-IV in also noting that compulsions may further serve to reduce distress not associated with an untoward event (e.g., reducing vague discomfort associated with the "just-right" phenomenon). ICD-10 notes that compulsions are usually recognized as pointless or ineffectual providing some correspondence to the DSM-IV requirement that symptoms be recognized as excessive or unreasonable, but in contrast to DSM-IV, the applicability of this criterion to children is not addressed. Moreover, although symptoms are described as almost always distressing, ICD-10 does not provide other operationalized impairment criteria (e.g., time-consumption or interference with functioning). Finally, although noting that an OCD diagnosis "should be preferred only if ruminations arise or persist in the absence of a depressive disorder," ICD-10 does not specify the range of diagnostic exclusions found in DSM-IV.

EPIDEMIOLOGY

Recent epidemiological studies suggest a lifetime prevalence for OCD in children and adolescents between 2 and 4 percent, indicating that the disorder is approximately as common among youngsters as in adults. One recent community study of adolescents found a prevalence rate of 3 percent for "clinical" OCD, defined as symptoms severe enough to cause distress and interfere with functioning. The most commonly accepted 6-month prevalence rate is between 0.5 and 1.0 percent of the general pediatric population.

Most studies report a mean age of onset for childhood OCD between 6 and 11 years of age. Onset age appears to be bimodally distributed, with peaks in early childhood and early adolescence. Studies conducted at NIMH indicate that youngsters with an early onset of OCD (below age 7) are more likely to be male and to have a family history of OCD than those with later onset, suggesting that genetic factors may be more likely to play a role in development of the early onset subtype than the late-onset subtype. Although clinic-based studies of primarily younger patients suggest that the disorder is more common in males than females, this finding has not been replicated in epidemiological surveys of adolescent samples and may be a function of age or referral bias. Referral bias may also account for the underrepresentation of minorities and youngsters from lower socioeconomic strata in clinical samples of OCD youth, as this pattern has not been replicated in epidemiological studies with adult samples.

Comorbidity Comorbidity in children and adolescents with OCD is common, with up to 80 percent of affected youths meeting diagnostic criteria for an additional Axis

I disorder and as many as 50 percent experiencing multiple comorbid conditions. The most common comorbid disorders include other anxiety disorders (26 to 75 percent, depending on the sample), depressive disorders (25 to 62 percent), and disruptive behavior disorders (18 to 33 percent). The fact that depressive symptoms commonly appear after the OCD suggests that concurrent mood disturbance may be reactive. Obsessive-compulsive personality disorder has been reported in 11 to 14 percent of pediatric OCD samples. Motor and vocal tics are common in children with OCD, and 20 to 30 percent also suffer from Tourette's disorder. Youngsters with tic-related OCD (i.e., OCD with comorbid Tourette's disorder) are less likely to report engaging in ordering, hoarding, and washing compulsions than youngsters with OCD only.

ETIOLOGY

Neurobiological Factors The onset of OCD following encephalitis, head trauma, and epilepsy and the association between OCD and Sydenham's chorea and Tourette's syndrome support the notion of OCD as a neurobiological disorder.

Neurophysiological Factors Pharmacological and biochemical challenge studies suggest that abnormalities in central nervous system (CNS) serotonergic activity and sensitivity are related to at least a significant proportion of OCD cases. Primary support for this hypothesis (the serotonin hypothesis) is the fact that highly potent serotonin reuptake inhibitors are currently the most effective psychopharmacological treatments for OCD. Studies of both adult and adolescent samples have shown a relation between OCD severity and treatment outcome and peripheral markers of serotonergic functioning. A positive relation between decreased platelet serotonin concentration and decreases in OCD symptomatology in serotonin reuptake inhibitor-treated adolescents has been reported. Basal prolactin concentrations, an indirect measure of serotonergic activity, also correlated negatively with both duration and severity of OCD symptoms in a group of youngsters with severe OCD, and serotonin reuptake inhibitor treatment can significantly increase basal prolactin concentration over time.

Neuroanatomical and Neuropsychological Factors A number of neuroimaging studies of adults implicate abnormalities in corticostriatal-thalamocortical pathways in the etiology factors for OCD. Positron emission tomography (PET) studies show increased glucose metabolism in orbital frontal and prefrontal cortex, right caudate nucleus, and anterior cingulate gyrus. Moreover, successful treatment with either behavioral therapy or selective serotonin reuptake inhibitor (SSRI) medication decreases activity in these areas and attenuates the correlations of activity between these regions. Magnetic resonance imaging (MRI) and volumetric computed tomography (CT) scan studies also suggest frontal lobe and caudate nuclei abnormalities, and normalization of frontal and basal ganglia glucose metabolism was observed in OCD adults showing positive response to treatment.

In OCD patients, signals from an overactive orbitofrontal cortex may underscore feelings of danger and lead to ritualistic behavior, while the caudate nucleus and cingulate gyrus are associated with the more visceral and affective aspects (associated dread and worry) of the disorder. Some researchers believe that high intercorrelation of activity along the corticostriatal-thalamocortical pathway may lead to the circuit becoming “stuck,” resulting in an escalating pattern of doubt and checking characteristic of many OCD patients. Successful treatment may relieve symptoms by reducing the correlation in activity rates among these different brain areas.

Higher rates of neurological soft signs have been observed in both adults and children with OCD, and one prospective follow-up study found that the presence of neurological soft signs at age 7 predicted the occurrence of obsessive-compulsive symptoms in adulthood. Collective findings from imaging, neurological, and neuropsychological studies implicate greater right than left hemispheric dysfunction.

Immunological Factors Researchers have recently described a subgroup of youngsters whose OCD symptoms appear to be triggered or exacerbated by group A b-hemolytic streptococcal (GABHS) infection. The obsessive-compulsive symptoms in these individuals appear to stem from caudate swelling caused by a cross-reaction between caudate tissue and antineuronal antibodies formed against GABHS in a process similar to the development of Sydenham's chorea. This variation of OCD, *pediatric autoimmune neuropsychiatric disorders associated with streptococcal infection* (PANDAS), is characterized by sudden, dramatic onset or exacerbation of OCD or tic symptoms, associated neurological findings, and a recent streptococcal infection.

Family and Genetic Factors Several studies suggest that up to 50 percent of childhood OCD cases may be familial. Rates of OCD among parents of youngsters with OCD range from 17 to 19 percent, with some OCD symptoms reported in 52 percent of parents. Relatives of probands with early-onset OCD exhibit higher rates of OCD and tic symptoms. No significant correlation has been found between the type of OCD symptoms experienced by probands and their affected family members, suggesting that modeling may not play a primary role in the intergenerational transmission of the disorder.

Behavioral Factors Most behavioral conceptualizations of OCD are based on Hobart Mowrer's two-factor conditioning theory. According to this theory, a previously neutral stimulus becomes aversive (a conditioned stimulus) as a result of being associated with an unrelated aversive stimulus. Any actions associated with a decrease in the aversive response to the conditioned stimulus are then reinforced through higher-order conditioning (i.e., negative reinforcement). According to operant theory, compulsions become strengthened as a result of their anxiety-reducing properties. Because specific traumatic events responsible for the conditioned response are rarely identified, conditioning theories do not adequately explain the acquisition of obsessive fears. However, these theories are useful for understanding the maintenance of many OCD symptoms.

DIAGNOSIS AND CLINICAL FEATURES

Presentation of the disorder in children and adolescents is generally similar to that for adults. Compared with adults, however, children may be somewhat less likely to consider their OCD symptoms excessive, perhaps because of lower levels of cognitive awareness. Although a criterion for adult OCD, DSM-IV does not require that youngsters recognize their symptoms as senseless or unrealistic for the diagnosis to be made. However, DSM-IV does require that if another Axis I disorder is present, the content of the obsessions or compulsions are not restricted to that disorder. Examples of this include concern with appearance in body dysmorphic disorder, hair pulling in trichotillomania, and fears of harm befalling a loved one in separation anxiety disorder. Disturbance due to the direct physiological effects of a substance or general medical condition is also ruled out (see Table 15.6-10.)

The range of obsessions and compulsions experienced in childhood is quite similar to that found in adults with the disorder, although children are more likely to engage in compulsive reassurance seeking and involve family members in their rituals. Most youngsters with OCD have both obsessions and compulsions. The most common obsessions in this age range tend to focus on germs or contamination, followed by fears of harm to self or others, concerns with symmetry, and excessive moralization or religiosity. The most common compulsions include excessive washing, repeating, checking, touching, counting, ordering, and arranging. Compulsions in the absence of anxiety-related or otherwise distress-inducing obsessions are most commonly found in younger children, who often describe their rituals as being performed in response to an irresistible urge or an otherwise vague sensation (the “just right” phenomenon). Although the pattern and type of symptoms typically shift over time, the absolute number of symptoms generally remains constant. However, the disorder is stress sensitive, and many children experience acute symptom exacerbations during times of psychosocial stress or change (e.g., start of school year, moving to a new home, death of or separation from a family member).

Given the typically strange and senseless nature of many OCD behaviors, most children and adolescents with OCD attempt to camouflage or hide their rituals. With substantial effort, most children can inhibit or control their symptoms for short periods of time. In many cases, parents, teachers, and others close to the child remain unaware of the child's problem for months or years, learning of it only after the child can no longer control the symptoms or becomes too overwhelmed to cope.

Julie is a 10-year-old female who lives with her parents and two younger siblings. Julie reported a past history of OC behaviors including counting rituals dating to about first grade. These symptoms were moderate and remained manageable until 6 months prior to intake, when she began experiencing intrusive thoughts related to harm and subsequent near-constant performance of rituals. Julie's parents recognized her symptoms from a television program on OCD and scheduled an appointment for consultation and evaluation. For the past 8 months, Julie had been attending weekly supportive psychotherapy, although her OCD symptoms had not been addressed in this treatment.

At intake, Julie exhibited a range of obsessions primarily centered on harm befalling herself or her siblings, symmetry, and fears of forgetting things. Her compulsions included always making sure that if something touched one side of her body, she touched it with the other side as well, trying to hear things equally with both ears (causing her to frequently hold her head at an odd angle), and needing to see both feet as she walked. Julie's most severe ritual was related to walking through doorways. She felt that she needed to have a particular good thought in her head when walking through a doorway and would doubt whether she was having the appropriate thought. This would cause her to walk back and forth across doorways several times in an effort to ensure that she was having the good thought. She believed that if she wasn't having the good thought when walking through the doorway, it would be the last time that she would walk through any doorway because she would die. This symptom caused Julie to expend considerable time and energy and led to considerable interference with academic and social functioning at school. When she was unable to continue walking through doorways until satisfied, Julie would experience marked ongoing distress related to instrusive thoughts about her own death. This interfered with her ability to focus on her schoolwork, and her grades declined. In addition, the rituals had become very annoying to other family members and embarrassing to Julie when she was in public. Julie stated that although she still enjoyed being with her friends and playing basketball, her rituals were “getting in the way” and were very upsetting to her.

DIFFERENTIAL DIAGNOSIS

The DSM-IV requirement of functional impairment and distress differentiates obsessive-compulsive symptoms from benign habits or mildly excessive thoughts. Children and adolescents with OCD can be differentiated from those with other anxiety disorders in several ways. Although children with various anxiety disorders are often preoccupied or worried about issues related to their fears (e.g., a child with separation anxiety disorder worrying about harm befalling a parent, a child with dog phobia worrying about the presence of a dog), they do not typically engage in compulsions to reduce this anxiety. Although youngsters with generalized anxiety disorder typically experience recurrent intrusive worries, these worries tend to be focused on everyday matters, such as school, work, friendships, and future events. In contrast, obsessions usually involve some bizarre or seemingly magical content and are recognized by most children as unrealistic. Also, rituals or compulsions do not usually accompany the type of worry characteristic of generalized anxiety disorder.

Youngsters with Tourette's disorder typically exhibit motor and vocal tics that occur frequently and can appear behaviorally similar to compulsive actions. Tics are differentiated from compulsions, however, in usually being avolitional and not preceded by an obsessional thought. On occasion, children with OCD are mistakenly diagnosed with attention-deficit/hyperactivity disorder, since frequent intrusive obsessions can lead to inattentiveness, and ritualistic or avoidant behaviors can make the child appear overactive or impulsive. However, the purposeful nature of the OCD child's movements and their association with obsessive thoughts can be used to distinguish compulsions from motor hyperactivity. Perseverative and stereotypic behaviors are characteristic of pervasive developmental disorders such as autistic and Asperger's disorders. These can usually be differentiated from compulsions, since the former behaviors are neither distressing to the child nor performed in response to an obsession. In some cases, the fixed nature of bizarre or unusual obsessional beliefs can have a delusional quality and thus resemble schizophrenic symptomology. OCD can be distinguished from schizophrenia by the absence of other schizophrenic symptoms (e.g., hallucinations, formal thought disorder) and the ability to engage in reality testing. Several neurological disorders should be considered in the differential diagnosis of OCD, including temporal lobe epilepsy and complications secondary to CNS insults or tumors.

OCD must also be distinguished from developmentally normative ritualistic behaviors. These behaviors, which typically include bedtime or mealtime rituals and rigid or stereotypic rules for many activities, often serve to enhance the preschool child's sense of mastery and control. Although some have speculated that OCD represents the extreme endpoint along a continuum of normal developmental rituals, OCD can be distinguished from normal developmental phenomena on the basis of timing, content, and severity.

COURSE AND PROGNOSIS

Longitudinal studies attest to the chronicity of OCD in childhood. Follow-up studies have revealed that 43 to 68 percent of OCD youngsters continue to meet diagnostic criteria for the disorder up to 14 years after initial identification. Poorer outcome at follow-up is predicted by poor initial treatment response, a lifetime history of tic disorder, and a parental Axis I psychiatric disorder. The largest study to date reevaluated 54 participants in controlled medication trials at NIMH 2 to 7 years postintake. At follow-up, 43 percent of subjects still met diagnostic criteria for OCD, and only 6 percent were classified as complete remitters. On the whole, however, the group was significantly improved, with only 19 percent of subjects rated unchanged or worse. Substantial rates of comorbid psychiatric diagnoses have also been found at follow-up. Relatively few youngsters in the follow-up studies to date received an adequate trial of behavior therapy, which has been associated with positive long-term outcomes in adult samples. As behavior therapy becomes increasingly available for children with OCD, it is reasonable to expect the long-term outcome for these youngsters to improve.

TREATMENT

Several recent controlled trials with children and adolescents demonstrated the efficacy and tolerability of serotonergic agents for OCD in this age group. Fluvoxamine (Luvox), sertraline (Zoloft), and clomipramine (Anafranil) have all been approved by the Food and Drug Administration (FDA) for treatment of OCD in children and adolescents. Open trials of cognitive-behaviorial therapy (i.e., exposure plus response prevention) indicate that this treatment also benefits those with childhood OCD, with response rates similar to, or higher than, those for medication. In contrast, psychodynamic, supportive, and other psychosocial interventions have not proved effective when used alone for the treatment of OCD in either children or adults.

Pharmacotherapy

Selective Serotonin Uptake Inhibitors Large multicenter trials of fluvoxamine, sertraline, and paroxetine (Paxil) for childhood OCD have been completed. Fluoxetine (Prozac) showed similar benefit in two smaller controlled trials and several open trials. Response rates from the controlled medication trials average from 50 to 65 percent, although typical symptom reduction averages only 20 to 50 percent. Thus, the effectiveness of these medications in children is quite similar to that for adults. The most common adverse effects of the SSRIs include nausea, insomnia, hyperstimulation, agitation, headache, and sexual effects. These effects are typically transient, and most youngsters tolerate these medications well.

Clomipramine The earliest U.S. psychopharmacology trials for childhood OCD were conducted with clomipramine, a tricyclic drug. The largest controlled trial for childhood OCD yielded a 60 percent response rate for the medication group versus only 17 percent for placebo. Clomipramine adverse effects were typical of tricyclic drugs: dry mouth, somnolence, dizziness, fatigue, and tremor. Over the long term, however, tachycardia and electrocardiogram abnormalities are common in children treated with clomipramine and other tricyclic drugs; thus pretreatment and periodic electrocardiographic and therapeutic drug monitoring is warranted. Given the preferable safety profile and similar efficacy of the selective serotonin reuptake inhibitors, clomipramine should be considered a second- or third-line treatment option in most cases.

Medication Strategies Up to one-third of patients may not respond to monotherapy with a given serotonin reuptake inhibitor, and the likelihood of responding drops even further after a third serotonin reuptake inhibitor trial. Since many youngsters show a positive treatment response only after 2 to 3 months of treatment, one should wait at least 8 weeks and preferably longer before changing agents, adopting high-dose strategies, or undertaking augmentation regimens. A recent multisite fluvoxamine extension study found that many youngsters showed continuing improvement over the longer term, with maximum symptom relief occurring after 4 to 6 months of treatment (Table 46.1-1).

Generic Name	Brand Name	Typical Daily Dosage Range (mg)
First line		
Fluvoxamine	Luvox	50–200
Fluoxetine	Prozac	20–80
Sertraline	Zoloft	100–300
Paroxetine	Paxil	20–80
Citalopram*	Celexa	10–40
Second line		
Clomipramine	Anafranil	75–200

* Has not been well studied in children and adolescents. Clear dosing parameters for childhood OCD have yet to be established and are extrapolated.

Table 46.1-1 Primary Drugs Used for Childhood OCD

Combination drug therapy can be used for treatment-resistant patients or those with comorbid psychopathology. In most cases, the choice of an augmenting agent is determined by the comorbid condition, that is, busipirone (BuSpar) or clonazepam (Klonapin) with comorbid anxiety, lithium (Eskalith, Lithobid) with comorbid affective symptoms, and neuroleptics with comorbid tic disorder or refractory schizoobsessive symptoms. The use of combined medications for OCD has been poorly studied in children, and most current practice is based on clinical reports from the adult literature. Thus, a conservative approach must be taken when considering these strategies for OCD in children. In most cases, augmentation should only be considered for youngsters who have failed at least two monotherapy SSRI trials and a course of high-quality behavioral therapy.

Behavior Therapy Although both cognitive-behavioral therapy and SSRIs appear to be effective and well-tolerated treatments for childhood OCD, data from controlled trials with adults suggest that behavioral therapy may yield somewhat higher response rates, greater symptom reduction, and more-durable treatment gains than those obtained with medication alone. This, along with concerns regarding the long-term use of medication in children, has led to a preference by many clinicians to begin treatment with cognitive-behavioral therapy.

Behavioral therapy for childhood OCD is closely based on that for the adult disorder and is focused on exposure plus response prevention. The first step in exposure plus response prevention consists of developing a rank-ordered list of all of the child's obsessive and compulsive symptoms. Following this, patients are systematically exposed to these situations (the least difficult ones first), while being instructed not to engage in their ritualistic behaviors. Over repeated exposures, the associated anxiety dissipates through the process of autonomic habituation. In addition, when the feared consequences of not ritualizing fail to occur, patients' heightened expectations of harm disappear, reducing anxiety even further. Response rates from the open behavioral trials average from 75 to 80 percent, with mean symptom reduction in the 45 to 60 percent range. Preliminary data regarding predictors of outcome to behavior therapy for childhood OCD were provided by one recent study that reported an inverse relation between degree of symptom reduction and worse depression, anhedonia, and social functioning at baseline.

A number of developmental considerations complicate the behavioral treatment of childhood OCD. The most important complications stem from the cognitive and motivational limitations of young children for whom the motivating aspect of future improvement is likely to be outweighed by the high anxiety associated with the initial treatment exercises. These issues have been addressed by using psychoeducation to facilitate an understanding of the illness, cognitive strategies for dealing with anxiety, and behavioral rewards for compliance with treatment tasks, as well as an emphasis on graphic feedback of progress and family involvement in treatment.

Childhood OCD can have a significant negative impact on family functioning and relationships through both the disruption of normal family activities and the involvement of family members in rituals. As a result, inclusion of family members in the child's treatment is frequently warranted. Typically, family treatment provides psychoeducation about OCD to reduce blame and help the family distinguish OCD from non-OCD problem behaviors. Subsequent interventions facilitate family disengagement from the child's OCD behaviors and reinforce healthy patterns of family interaction.

Combined Therapy Even though concurrent behavioral and psychopharmacological treatment of pediatric OCD has not been well studied, evidence from the adult literature supports the use of a combined treatment approach. As noted, most medication and some behavior therapy responders experience continued significant OCD symptoms. In adults, combined treatment affords greater symptom relief than exposure plus response prevention or medication alone. Moreover, in adults, medication increases compliance with behavioral therapy, while the addition of behavior therapy minimizes relapse upon medication holiday and breakthrough symptoms associated with prolonged medication use. This latter point is especially significant for youngsters, in whom long-term medication effects are unknown.

Since symptom relapse follows cessation of medication, medication withdrawal should only be considered for youngsters who have remained symptom free for at least 1 year and must be done gradually. The typical strategy involves lowering the dosage in 25-percent increments at 2-month intervals in combination with cognitive behavioral therapy booster sessions to prevent relapse.

Other Interventions Case reports suggest that plasmapheresis and intravenous immunoglobulin are helpful for adolescents with presumed infection-triggered autoimmune OCD. The only published study of psychosurgery for adolescents with OCD reported significant symptom reduction in five adolescents following sterotaxic capsulotomy, although the course of postoperative rehabilitation was difficult. Scattered reports exist describing the use of electroconvulsive therapy, sleep deprivation, acupuncture, and biofeedback for adult OCD; however, no information is available regarding the utility of these techniques with children.

SUGGESTED CROSS-REFERENCES

Chapter 15 covers the anxiety disorders. SSRIs are covered in Section 31.25, and clomipramine is covered in Section 31.30 on tricyclic drugs. Cognitive-behavioral therapy in children is covered in Section 48.3. Compined psychopharmacology-psychotherapy is presented in Section 30.12.

SECTION REFERENCES

*Albano A, Piacentini J, March J: Cognitive behavioral treatment of obsessive-compulsive disorder. In *Handbook of Prescriptive Treatments for Children and Adolescents,* ed , R Ammerman, M Hersen, C Last, editors. Allyn & Bacon, Needham Heights, MA, 1999.

*Allen AJ, Leonard HL, Swedo SE: Case study: A new infection-triggered, autoimmune subtype of pediatric OCD and Tourette's syndrome. J Am Child Adolesc Psychiatry *34:*307, 1995.

Apter A, Ratzioni G, King R: Fluvoxamine open-label treatment of adolescent inpatients with obsessive-compulsive disorder or depression. J Am Acad Child Adolesc Psychiatry *33:*342, 1994.

Baxter LJ, Schwartz JM, Bergman KS: Caudate glucose metabolic rate changes with both drug and behavior therapy for obsessive-compulsive disorder. Arch Gen Psychiatry *49:*681, 1992.

DeVeaugh-Geiss J, Moroz G, Biederman J, Cantwell D, Fontaine R, Greist JH, Reichler R, Katz R, Landau P: Clomipramine hydrochloride in childhood and adolescent obsessive-compulsive disorder—a multicenter trial. J Am Acad Child Adolesc Psychiatry *31:*45, 1992.

Franklin M, Kozak M, Cashman L, Coles M, Rheingold A, Foa E: Cognitive-behavioral treatment of pediatric obsessive-compulsive disorder: An open clinical trial. J Am Acad Child Adolesc Psychiatry *37:*412, 1998.

Hanna G: Demographic and clinical features of obsessive-compulsive disorder in children and adolescents. J Am Acad Child Adolesc Psychiatry *34:*19, 1995.

Hanna G, McCracken J, Cantwell D: Prolactin in childhood obsessive-compulsive disorder: Clinical correlates and response to clomipramine. J Am Acad Child Adolesc Psychiatry *30:*173, 1991.

Hibbs E, Hamburger S, Lenane M: Determinants of expressed emotion in families of disturbed and normal children. J Child Psychol Psychiatry *32:*757, 1991.

Kiessling LS, Marcotte AC, Culpepper L: Antineuronal antibodies: Tics and obsessive-compulsive symptoms. J Dev Behav Pediatr *15:*421, 1994.

Knox L, Albano A, Barlow D: Parental involvement in the treatment of childhood OCD: A multiple-baseline examination involving parents. Behav Ther *27:*93, 1996.

Leonard H, Swedo S, Lenane M, Rettew D, Hamburger S, Bartko J, Rapoport J: A 2- to 7-year follow-up study of 54 obsessive-compulsive children and adolescents. Arch Gen Psychiatry *50:*429,

1993.

Leonard HL, Goldberger EL, Rapoport JL, Cheslow DL, Swedo SE: Childhood rituals: Normal development or obsessive-compulsive symptoms. J Am Acad Child Adolesc Psychiatry *29:*17, 1990.

*Leonard HL, March J, Rickler KC, Allen AJ: Pharmacology of the selective serotonin reuptake inhibitors in children and adolescents. J Am Acad Child Adolesc Psychiatry *36:*725, 1997.

*March J: Cognitive-behavioral psychotherapy for children and adolescents with OCD: A review and recommendations for treatment. J Am Acad Child Adolesc Psychiatry *34:*7, 1995.

March J: *Anxiety Disorders in Children and Adolescents.* Guilford Press, New York, 1995.

March J, Frances A, Kahn D, Carpenter D: Expert consensus guidelines: Treatment of obsessive-compulsive disorder. J Clin Psychiatry *58*(Suppl):1, 1997.

Owens E, Piacentini J: Behavioral treatment of OCD in a boy with comorbid disruptive behavior problems. J Am Acad Child Adolesc Psychiatry *37:*443, 1998.

Pauls DL, Alsobrook JP, Goodman W, Rasmussen S, Leckman J: A family study of obsessive-compulsive disorder. Am J Psychiatry *152:*76, 1995.

Piacentini J, Gitow A, Jaffer M, Graae F, Whitaker A: Outpatient behavioral treatment of child and adolescent OCD. J Anxiety Disord *8:*277, 1994.

*Piacentini J, Graae F: Childhood obsessive-compulsive disorder. In *Obsessive-Compulsive Disorders: Diagnosis, Etiology, and Treatment,* E Hollander, D Stein, editors. Marcel Dekker, New York, 1997.

Piacentini J, Jaffer M, Gitow A, Graae F, Davies S, Del Bene D, Liebowitz M: Psychopharmacologic treatment of child and adolescent obsessive compulsive disorder. Psychiatr Clin North Am *15:*87, 1992.

Rapoport J, Leonard H, Swedo S, Lenane M: Obsessive compulsive disorder in children and adolescents: Issues in management. J Clin Psychiatry *54:*27, 1993.

Rettew D, Swedo S, Leonard H, Lenane M, Rapoport J: Obsessions and compulsions across time in 79 children and adolescents with obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry *31:*1050, 1992.

Riddle M, Scahill L, King R, Hardin M, Anderson G, Ort S, Smith J, Leckman J, Cohen D: Double-blind, crossover trial of fluoxetine and placebo in children and dolescents with obsessive-compulsive disorder. J Am Acad Child Adolesc Psychiatry *31:*1062, 1992.

Riddle M, Scahill L, King R, Hardin M, Towbin K, Ort S, Leckman J, Cohen D: Obsessive compulsive disorder in children and adolescents: Phenomenology and family history. J Am Acad Child Adolesc Psychiatry *29:*766, 1990.

*Rosenberg D, Keshavan M: Toward a neurodevelopmental model of obsessive-compulsive disorder. Biol Psychiatry *43:*623, 1998.

Swedo S, Leonard H, Kiessling L: Speculations on anti-neuronal antibody-mediated neuropsychiatric disorders of childhood. Pediatrics *93:*323, 1994.

Swedo S, Rapoport J, Leonard H, Lenane M, Cheslow D: Obsessive-compulsive disorder in children and adolescents: Clinical phenomenology of 70 consecutive cases. Arch Gen Psychiatry *46:*335, 1989.

Valleni-Basile L, Garrison C, Jackson K, Waller J, McKeown R, Addy C, Cuffe S: Frequency of obsessive-compulsive disorder in a community sample of young adolescents. J Am Acad Child Adolesc Psychiatry *33:*782, 1994.

46.2 POSTTRAUMATIC STRESS DISORDER IN CHILDREN AND ADOLESCENTS

LISA AMAYA-JACKSON, M.D.

Descriptions of child and adolescent victims of disaster and violence flash across television screens and newspapers with such growing frequency that more and more people are questioning the psychiatric consequences of direct, witnessed, and indirect traumatic experiences on the lives of youth. Research studies in recent years have confirmed that posttraumatic stress disorder and its inherent comorbidity occurs not only in adult victims of traumatic life events, but in children across the age spectrum. Delineating the many factors that influence the development of posttraumatic stress disorder as well as the best treatment outcomes within a pediatric developmental framework remains a pressing research agenda.

DEFINITION

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) defines posttraumatic stress disorder in six diagnostic criteria: (1) the person must have experienced, witnessed, or been confronted with an event involving death, serious injury, or a threat to the physical integrity of the self or others; (2) the traumatic event must be persistently reexperienced in the form of distressing images, thoughts, perceptions, dreams, or reliving, and intense psychological or physiological reactivity may be present with reminders of the event; (3) persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness must be present following the trauma; (4) persistent symptoms of increased physiological arousal must be present; (5) the duration must exceed 1 month, and (6) some functional impairment must result. The related disorder, acute stress disorder, is not discussed in this section because the nosology of acute stress disorder is the same for children as for adults.

HISTORY

Accounts of emotional reactions accompanying human trauma can be traced throughout history. Pierre Janet's studies in the early 1880s are heralded as the first that contributed to the core set of symptoms that evolved into the current disorder. Following Janet, Sigmund Freud's conceptualization of psychic trauma had major underpinning for his contribution to psychoanalytic theory. “Traumatic neurosis” can be viewed as the earliest version of posttraumatic stress. The syndrome's evolution was primarily influenced by war trauma studies, particularly those on World War I and World War II veterans who suffered “shell-shock” and “combat fatigue”; Abraham Kardiner described it as a “physioneurosis” because of the notable autonomic hyperphysiology. Studies on the effects of traumatic events on children are primarily recent except for the noted work by Anna Freud and Dorothy Birlinghan that describes the reactions of children to the air raids of London. They noted that children's reactions often were mediated by parental response. Later studies of children's reactions found that when children are direct victims or witnesses to life threat (particularly when their developmental level allows them to understand the degree of threat), level of exposure is a stronger predictor of child distress than parental response.

The core symptoms of posttraumatic stress disorder were not actually applied to children until 1987, in the revised third edition of DSM (DSM-III-R). Research then confirmed that children's exposure across a wide range of traumatic life events can result in posttraumatic stress disorder.

COMPARATIVE NOSOLOGY

Prior to the third edition of DSM (DSM-III), specific reactions to trauma were called “gross stress reactions” (the first edition of DSM [DSM-I]) or “adjustment disorders of adult life,” with no application specified for children or adolescents. Posttraumatic stress disorder entered the nomenclature in DSM-III in 1980 as an anxiety disorder with an established stressor criterion, reexperiencing, numbing, avoidance, and general symptoms of distress. Not until DSM-III-R was there any application to children. DSM-III-R and DSM-IV refined the symptom groupings of the criteria in respect to the phasic nature of the symptoms (relative to both the physiological and psychological distress resulting from reminders of the event). The definition of a traumatic event (criterion A) evolved from something that would “cause distress in most people” (DSM-III), to “an event that is outside usual human experience” (DSM-III-R), to the current emphasis on the event's characteristics and the subjective distress response of the victim (DSM-IV). DSM-IV mentions several childhood stressors and emotional responses in children in criterion A in a nonspecific, nondevelopmentally derived way. Acute stress disorder is introduced in DSM-IV. DSM-IV attempts to be compatible with the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). Disorders comparable to posttraumatic stress disorder and acute stress disorder appear in ICD-10; however, they are placed in a category of stress disorders rather than anxiety disorders.

EPIDEMIOLOGY

Prevalence rates for current and lifetime posttraumatic stress disorder were found to be 0.4 and 1.3 percent, respectively, in the Epidemiologic Catchment Area study. An additional 15 percent experienced posttraumatic stress disorder symptoms but failed to meet full criteria for the disorder. The only epidemiological study that looked specifically at the general population incidence or the prevalence of posttraumatic stress disorder in children and adolescents, done on a rural population of children in the Smokey Mountains, noted a 3-month prevalence rate of 0.6 percent. High rates of posttraumatic stress disorder have been documented in children exposed to such life-threatening events as combat and other war-related trauma, kidnapping, severe illness or burns, bone marrow transplantation, and a number of natural and man-made disasters. Studies on young victims or witnesses to criminal assault, domestic violence, and community violence have revealed high psychiatric morbidity (including posttraumatic stress) following exposure to violence.

Several incidents led to pioneering efforts in landmark studies in the area of sudden incident trauma. Lenore Terr's description of the Chowchilla school bus kidnapping described posttraumatic stress symptoms in all the children involved, regardless of their developmental, psychiatric, or prior trauma histories. Rosemarie Galante and Daria Foa followed young Italian earthquake victims and emphasized the inadequacy of non–trauma-specific measures. Robert Pynoos and investigators noted a higher than 90 percent incidence of posttraumatic stress disorder in child victims of a schoolyard sniper attack, and nearly 60 percent of exposed children continued to meet full criteria 1 year later. Posttraumatic stress disorder prevalence rates of 30 to 70 percent have been reported by other investigators studying disaster, community violence, and postwar adolescent refugee populations. These studies confirmed that posttraumatic stress disorder in children can become chronic and is highly comorbid with anxiety and depression. These and other studies have noted that parents and teachers may significantly underestimate both the intensity and the duration of the stress reactions in children. Furthermore, as in adult research, studies on child trauma reveal that exposure (proximity) is directly related to the risk of severe symptoms of posttraumatic stress disorder.

Only recently have posttraumatic stress disorder researchers turned to the more chronic forms of child trauma, such as domestic violence and child sexual abuse, despite long-standing public health concerns about their prevalence and impact on psychosocial functioning of affected young persons. A meta-analysis of the effects of sexual abuse revealed a mean prevalence of posttraumatic stress disorder in sexual abuse victims of 32 percent, and the rate increased to 53 percent when the analysis included studies of victims of severe ritualistic abuse.

ETIOLOGY

Posttraumatic stress disorder, by its first criterion develops in response to a stressor event. However, other variables contribute to development of the disorder,

including characteristics specific to the children and their environment, characteristics specific to the stressor, and posttrauma sequelae. The roles of risk factors, protective factors, and associated mediating variables of posttraumatic stress symptoms in children and adolescents are not well understood. Child variables of important consideration include demographic factors (e.g., age, sex, socioeconomic status), other life events (positive and negative), social and cultural cognitions, psychiatric comorbidity, and inherent coping strategies. Family factors (e.g., parental psychopathology and functioning, marital status, and education) play key roles in determining child symptoms. Parents' response to traumatic events particularly influence young children who may not completely understand the nature of the trauma or its inherent danger. Social psychologists have noted that traumatized young persons evidence a wide variety of social behaviors considered outside the norms, though which are risk factors and which are outcomes of the traumatic experience is less clear. Developmental psychologists have applied risk-analysis to studies of chronic adversity, and results suggest that multiple traumas, multiple risk factors, or both follow an additive-risk model. This has not yet been applied to posttraumatic stress disorder specifically.

Several frameworks attempt to conceptualize how the stressor actually causes the notable biological and psychological symptoms. While these frameworks have clear application to children and adolescents, none of the three were developed within a developmental psychopathology context. Cognitive-behavioral components probably have the strongest links to child-based developmental theory.

Neurobiological Factors Multiple physiological and neurochemical changes are implicated in the biology of posttraumatic stress disorder. Adult studies found prolonged autonomic sympathetic activity in trauma victims with the disorder. Using trauma-specific cues (e.g., combat films shown to war veterans), investigators found elevations in heart rate, blood pressure, electromyography, and sweat activity in veterans with posttraumatic stress disorder but not in those without the disorder (even when the latter had other anxiety disorders). Abnormal noradrenergic activity was found in studies using yohimbine (Yocon), a type 2 a-adrenergic receptor antagonist. Serotonin has been implicated in the stress response, with evidence suggesting better adaptation in animals with increased central serotonin functioning. Posttraumatic stress disorder patients experienced flashbacks and panic induction following administration of a serotonin receptor agonist. Studies investigating noradrenergic and serotonergic systems have not been replicated in children or adolescents; however, studies evaluating baseline levels of sympathetic tone and fear-enhanced startle responses in children with posttraumatic stress disorder have supported some of these findings.

The hypothalamic-pituitary-adrenal axis exhibits distinct changes in neuroendocrine functioning in patients with posttraumatic stress disorder. Changes in 24-hour urinary cortisol concentrations have been reported, as has dexamethasone (Decadron) suppression of cortisol in adult posttraumatic stress disorder patients. Neuroendocrine changes in maltreated children and adolescents were noted in a study looking at growth hormone and one assessing early pubertal changes.

Stress-induced analgesia in combat veterans exposed to combat films was reversed by naloxone (Narcan) administration, implicating abnormalities in the opioid system. Changes in sleep architecture in adult trauma victims with posttraumatic stress disorder include increased rapid eye movement (REM) latency, decreased REM sleep (reversed findings are noted in studies on patients with depression), and decreased stage 4 sleep (increased in those with panic states and night terrors). These sleep findings, and some of the neuroendocrine findings have important implications in differentiating posttraumatic stress disorder from depression and anxiety states. All these findings require further neurobiological research in children and adolescents to understand the developmental context of the stress response. Children, as developing organisms, may have distinct biological vulnerabilities to these neurobiological changes.

Psychodynamic Factors Sigmund Freud, around 1920, described psychic trauma (then referred to as “traumatic neurosis”) resulting from a traumatic stimulus intense enough to penetrate the stimulus barrier, overwhelm the ego, and render it helpless. This conceptualization, influenced by both Janet and Freud, posited that the dramatic loss of function resulted from the inordinate amount of energy trauma victims expended setting up every defensive mechanism they were capable of. Freud described “repetition compulsions” as unmastered themes of the trauma repeated in actions of everyday life that lead to the “fixation” on the trauma by the victim. Psychoanalytically defined defense mechanisms deemed important in trauma work include denial, dissociation, projection, and identification with the aggressor. Other theorists furthered psychodynamic interpretations by focusing on the (trauma-induced) emotion itself—so overwhelming that the traumatized individual experiences it as uncontainable.

Cognitive-Behavioral Factors Social learning theory applied to posttraumatic stress disorder uses the behavioral two-factor learning process in which both classical (factor one) and instrumental (factor two) conditioning apply. In classical conditioning, the stressor, or traumatic event, acts as an unconditioned stimulus that elicits an unconditioned response characterized by extreme fear and the cognitive perception of helplessness. Children note cognitive, affective, physiological, and environmental cues accompanying the traumatic event as conditioned stimuli. These stimuli become capable of eliciting a conditioned response through the process of stimulus generalization. The conditioned response is posttraumatic stress disorder, full blown or subsyndromal. In the second factor (instrumental conditioning), children quickly learn by trial and error how to reduce posttraumatic stress disorder symptoms through cognitive and behavioral avoidance. This model intertwines with both cognitive and information-processing models that note the inability to process and assimilate highly stressful experiences. Framed either as cognitive components that may follow higher-order conditioning patterns or as fear structures that remain easily activated, the individual adult or child would be unable to process memories beyond active memory storage. Hence these active memories repeatedly resurface, often rendering the trauma victim unable to modulate the arousal linked to the experience. Subsequently, the individual attempts avoidance and numbing.

DIAGNOSIS AND CLINICAL FEATURES

Stressor Diagnosis of posttraumatic stress disorder (see Table 15.6-12) requires meeting a stressor criterion. The stressor criterion A has been somewhat controversial in terms of defining a traumatic event. DSM-IV combines the qualifying aspects of life threat or association with serious injury or threat with the necessary subjective element of extreme fear and helplessness (which children may express as disorganized or agitated behavior). Clearly a child victim or witness to a life threat fulfills criterion A; instances in which chronic adversity plays a substantive role in a child's psychiatric presentation are less clear. Chronic physical and sexual abuse are considered potent stressors; however, their sequelae are not always adequately represented by the posttraumatic stress disorder symptom picture. The massive psychosocial adversities found in the inner city, terrorist countries, and third world child populations are traumatic by all standards, but they do not fit the sudden-incident nomenclature of posttraumatic stress disorder. Typologies such as type I trauma (sudden, single-incident trauma) and type II (ongoing or chronic trauma) have been suggested. Children also suffer posttraumatic stress as the result of “indirect” exposure—that is, the unwitnessed death or injury of a loved one, as in situations of disaster, war, or community violence.

Reexperiencing Children, like adults, reexperience the traumatic event in the form of distressing, intrusive thoughts or memories, flashbacks, and dreams. Reexperiencing phenomena, a unique feature of posttraumatic stress disorder, sets it apart from the other DSM disorders.

Children's nightmares may be linked specifically to a trauma theme or may generalize to other fears. Flashbacks occur in children as well as in their adolescent or adult victim counterparts. “Traumatic play,” a specific form of reexperiencing seen in young children, consists of repetitive acting out of the trauma or trauma-related themes in play. Depending on the nature of the play, it may be viewed as a positive or negative ability of the child to work through the experience of the event. Older children may incorporate aspects of the trauma into their lives in a process termed *reenactment*. Fantasized actions of intervention or revenge are common; adolescents should be considered at increased risk for impulsive acting out secondary to anger and revenge fantasies. Related behaviors in child and adolescent victims of trauma include sexual acting out, substance use, and delinquency.

Avoidance and Numbing Young victims of trauma, like adult victims, make the conscious attempt not to think about the event. Children not only cognitively suppress but also make conscious attempts to distract themselves via motor restlessness or impulsive behavior. This combined with the hyperarousal (criterion D) symptoms makes differentiation from attention-deficit disorder important but difficult.

Children often withdraw and show reduced interest in previously enjoyable activities. Loss of previously acquired skills is mentioned specifically in the C criterion and includes regressive behaviors such as enuresis or fear of sleeping alone. “Sense of foreshortened future” is another specific criterion C symptom that applies to children who may believe they will never grow up or who may lose motivation toward fulfilling a life goal because of a pervading sense of futurelessness.

Hyperarousal At least two symptoms of increased physiological arousal characterized by sleep disturbance, irritability, poor concentration, hypervigilance, and increased startle must be present. These manifest similarly in children and adults. Sleep disturbances are particularly problematic for children.

Associated Features Survival guilt was changed from a symptom to an associated feature in DSM-III-R, yet it remains highly prevalent in children with posttraumatic stress symptoms. DSM-IV struggles to accommodate the complex features of interpersonal violence and abuse (particularly when experienced in a prolonged manner during childhood) in the posttraumatic stress nomenclature. DSM-IV describes noteworthy associated symptoms of impaired affect modulation, self-destructive behavior, dissociative symptoms, somatic complaints, feelings of ineffectiveness and shame, feeling permanently damaged, constantly threatened, and a loss of

previous beliefs.

PATHOLOGY AND LABORATORY EXAMINATION

No laboratory studies presently help differentiate this disorder from other conditions. Psychological testing and the use of rating scales in children may help in data gathering for the differential diagnosis and treatment; however, they do not provide specific indicators that would make the diagnosis.

DIFFERENTIAL DIAGNOSIS

Children exposed to traumatic life events often exhibit symptomatology other than posttraumatic stress disorder. Furthermore, the symptoms of posttraumatic stress disorder may be confounded by their overlap with criteria symptoms for other disorders. For example, difficulty concentrating may be a symptom of posttraumatic stress disorder, depression, or attention-deficit disorder. Differentiating between likely comorbidities is particularly important since treatments may differ.

Anxiety disorders share symptoms (notably the criterion D set) with posttraumatic stress disorder as well as being highly comorbid conditions following trauma. Children experience fears about the safety of family and friends and not infrequently manifest full separation anxiety disorder. Generalized anxiety disorder is one of the most frequent comorbid diagnoses in both children and adults with posttraumatic stress disorder. Overlap with depressive symptoms is seen in the C and D criteria of posttraumatic stress disorder. These must be differentiated from major depression, which both the adult and child trauma literature show to be highly comorbid with posttraumatic stress disorder. Further complicating the picture, grief reactions and trauma-induced demoralization may also occur.

Disruptive behavior disorders or substance use may increase in children exposed to trauma—correlating with adult studies that found increased rates of antisocial behaviors after disaster and war. However, posttraumatic stress disorder must be ruled out as a cause of deteriorating school performance, concentration difficulties, irritability, or aggression that could mistakenly be attributed to disorder or attention-deficit/hyperactivity disorder. Sexualized behavior and suicidal ideation are often found in victims of child abuse. Dissociative symptoms may be particularly pronounced in abuse victims and victims of other interpersonal violence but are not uncommonly seen in child victims of other forms of trauma.

COURSE AND PROGNOSIS

High rates of symptomatology are seen immediately after a traumatic life event, and many children or teenagers are likely to meet criteria for acute stress disorder. A trauma victim is not eligible for the diagnosis of posttraumatic stress disorder until 1 month of symptoms has passed. Then it is not uncommon for rates of full-criteria symptomatology to be in the range of 60 to 90 percent of victims. Clinical lore suggests that trauma victims become more avoidant over time, which may negatively influence their willingness to seek treatment, despite the fact that some individuals' symptoms may worsen or may not fully arise until a substantial period of time has passed. Childhood delayed-onset posttraumatic stress symptoms have mostly been described in victims of severe physical abuse, sexual abuse, or both.

TREATMENT

The mainstay of both adult and child treatment strategies for posttraumatic stress disorder is cognitive-behavioral and psychodynamic psychotherapy used in combination with pharmacotherapy. Pharmacotherapy is less likely to be the primary treatment modality in children than in adults. Many fewer treatment studies have been done in children than in adult victims of trauma, particularly in clinical drug trials. Nevertheless, a growing number of clinicians who work with children are turning to trauma-based treatment strategies because of their increasing awareness of the impact of traumatic life events in their young patients.

Crisis Intervention When a traumatic event affects populations of children and their families, a crisis intervention plan should incorporate triage for acutely exposed victims, identify resources, and focus on supporting and strengthening coping skills necessary for dealing with the grief and trauma symptoms. The goal is preventing the development of chronic symptoms. Crisis centers and often schools serve as screening sites for children and families who may need further evaluations. Sessions called “critical incident stress debriefings” guide victims and caregivers to identify their reactions to the traumatic event. Outcome research on these debriefings have given them mixed reviews. Nevertheless, early preventive interventions are important in all acute trauma, when victims' symptoms are most prominent and they may be most amenable to contact with mental health professionals. Adult caregivers often need help dealing with resurfacing of prior losses or subsequent guilt over not being able to prevent what happened.

Psychotherapy The few treatment studies done in children emphasize cognitive-behavioral principles in targeting posttraumatic stress symptoms. Reexposure to traumatic reminders under safe conditions while incorporating mastery elements in a structured, supportive manner is a central theme for both group and individual trauma treatment. Systematic review of the traumatic event with children in a way that identifies their perception of the event and its meaning to them, their level of exposure, and their sense of “worst moment” becomes part of the working through of psychotherapy. Normalizing the children's reactions, acknowledging any actions that they did or wished they had done, identifying key traumatic reminders (conditioned stimuli), and clarifying their misunderstanding and predictable distortions of the event are important aspects of early treatment. Studies suggest that children may be particularly in need of contextual support in remembering, since proximity to life threat can influence memory. Children's detailed description of the trauma and their trauma response form the basis for narrative exposure that is strategically recalled in a play-by-play or moment-by-moment manner, to minimize avoidance mechanisms and habituate anxiety. Use of psychodynamic therapeutic principles allows development of the therapeutic relationship to sustain the discomfort of the exposure, facilitate an adequate holding environment, and attend to the meaning of the event that must be integrated into the child's experience. Cognitive restructuring, affective processing, and relaxation training are also used.

Psychotherapy must address the effect of the trauma and the resultant symptoms on the child's family life, peer relationships, and school performance and attend to the earlier described comorbidities. Treatment duration varies with level of exposure, and long-term treatment may be warranted for a child exposed to horrific violence, interfamilial homicide or suicide, or chronic abuse or when multiple risk factors (e.g., parental psychopathology, a chaotic living situation, and severe poverty) are present.

Group and family therapy are important treatment approaches that are used adjunctively or in place of individual therapy. Family members may need specific help in dealing with the child's distress and may suffer from trauma symptoms themselves that limit their ability to be in tune to their child's needs.

Parents benefit from psychoeducation about their child's symptoms and how to manage the symptoms. Working with nonoffending parents in maltreatment-related trauma cases has a strong empirical basis for improving a child's treatment outcome.

Group work can be specifically useful not only as a means to avail treatment when individual treatment is not available, but also when interpersonal difficulties need addressing or peer influences are important. Often victims express a need for support and understanding that only a fellow victim of trauma may understand.

Group treatment interventions in school settings tailor a public health approach in identifying children with symptoms and offering treatment when either shared exposure to single-incident violence or disaster occurs or multiple exposures to similar trauma such as widespread community violence occurs.

When a child's posttraumatic symptoms occur in the context of psychosocial adversity, including poverty, parental psychopathology, chronic abuse, lack of community support, and system failure, more systemic approaches that work with the family, community agencies, and the legal system may be necessary before trauma treatment strategies can be successful—given the rubric that most cognitive-behavioral treatment approaches stipulate the child is free from ongoing threat.

Psychopharmacology The neurochemical and physiological changes that accompany both short- and long-term posttraumatic stress symptoms indicate that medications may help to relieve the intense distress that so impairs trauma victims. Children, possibly more so than adults, may be unamenable to trauma-focused psychotherapy because of the intense hyperarousal and subsequent amplified avoidance. This is where psychopharmacological intervention would appear to be helpful. However, the clinical drug trial data on treatment of children and adolescents with posttraumatic stress disorder remains sparse, and clinicians who treat young people must rely on the results of research on adults. Medications are used to decrease intrusions and avoidance as well as hyperarousal symptoms; other target symptoms include impulsivity and sleep disturbance.

Clinical experience in children and adolescents supports many of the empirical findings from adult studies. Both tricyclic drugs (amitriptyline [Elavil] or imipramine [Tofranil]) and selective serotonin reuptake inhibitors (SSRIs) (fluoxetine [Prozac] and sertraline [Zoloft]) have efficacy on core posttraumatic stress disorder symptoms in adults that is not merely attributable to the drugs' antidepressant or anxiolytic effects. These drugs and others in their class have been used in children.

Concern about the lethality of tricyclic drugs and the necessity for careful electrocardiographic (ECG) monitoring may lead the treating physician to prefer the SSRIs over the tricyclic drugs; however, a posttraumatic stress disorder–induced sleep disturbance may be somewhat less responsive. Despite successful controlled trials using phenelzine (Nardil) in adults with posttraumatic stress disorder, clinicians have shied away from using monoamine oxidase inhibitors in children as a first-line drug.

Benzodiazepines, particularly alprazolam (Xanax), have only limited application in child posttraumatic stress disorder; clonazepam (Klonopin) may be an exception to this, although paradoxical disinhibition is a worry with it as with the rest of this class. Buspirone (BuSpar) may target anxiety and possible affective components, but as a single treatment it does not sufficiently relieve even the arousal symptoms on its own. Clonidine (Catapres) and possibly its counterpart, guanfacine (Tenex), target startle response in adults and children and are frequently used to treat intrusive symptoms and those of autonomic arousal. One study, using propranolol (Inderal), suggested that b-adrenergic receptor antagonist (beta-blockers) may reduce arousal symptoms in children. Carbamazepine (Tegretol) has been (case) reported to be helpful. Stimulants are often used unknowingly in child trauma victims to treat attention-deficit/hyperactivity disorder–like symptoms. Untangling the causes and overlapping criteria sets of posttraumatic stress disorder and attention-deficit/hyperactivity disorder is not easy; neither can one predict that a child's response to stimulants will determine whether attention-deficit/hyperactivity disorder is a true comorbidity or a misread stress response. Stimulants may exacerbate anxiety states and physiological hyperarousal. Alternatively, some patients with posttraumatic stress disorder who show the attention difficulties and impulsivity characteristic of attention-deficit/hyperactivity disorder may show improvement in impulse control, ability to stay on task, and activity level while on stimulants, though intrusive symptoms and sleep disturbances are unaffected. Only research that incorporates a developmental psychopathology view can untangle these and other child-specific factors in the psychopharmacological treatment of posttraumatic stress disorder in child and adolescent trauma victims.

B.C., a 5-year-old Asian boy living in New York City was brought by his mother to a university-based psychiatric clinic for evaluation and treatment of speech delays and psychological distress resulting from horrific family violence. B.C. and his two siblings aged 2 and 8 lived with his mother and her widowed father after fleeing their San Francisco home because of her husband's violent outbursts directed at her and her children. She reports that her husband of 12 years had a severe temper and when angered would verbally abuse her, or the children, or both, destroy furniture and the children's toys, and often physically abuse her and (more recently) the children. She stayed with her husband because her Chinese culture looked down on divorce, and she did not want her Chinese community to look down on her children's father. She left the home when the severe violence, which often left her with broken ribs and caused several episodes of loss of consciousness, extended to the children. The culminating episode was when the father—angry with his wife—stomped on the children's toys and then kicked B.C. in the stomach. Mrs. C. left with the children the next morning as soon as Mr. C. left for work. There was a 1-month period when the father took the children back to San Francisco and would not allow her even phone contact with them. When she was able to obtain sole legal custody, the children were returned to her. During the month with the father, the 8-year-old told her, the father told them daily that they must always stay and care for him when they grew up, otherwise his life would be ruined, and that he would take revenge against their mother and her relatives who were all to blame for the family being pulled apart. Mrs. C. reported that on their return from their father, all three children were more verbally and physically aggressive toward each other. Since living with the mother over the last 3 months their aggression has markedly decreased. Mrs. C. was worried about B.C. because of his terror of being separated from her—even if only for her to go to the bathroom—he would cling and cry inconsolably. B.C. had secondary enuresis of 2 years' duration. His mother reported that he had a severe sleep disturbance with frequent awakenings and nightmares throughout the night. School had gone surprisingly well with only mild separation problems upon being dropped off by his mother. Rating scales filled out by his teacher and his mother revealed no other striking behavioral concerns beyond those already described.

B.C. presented as an attractive, expressive 5-year-old whose speech was at times rapid, with some articulation problems. He was well behaved, without hyperactivity. He appeared to have a loving, close relationship with his mother and siblings that was observed both in session and in the waiting area. His mood was bright with no evidence of depression; his affect was clearly anxious. There was no evidence of thought disorder. He would often look out the window “to check”; he was clearly startled when a door down the corridor slammed shut as well as when he knocked several blocks off his worktable. He described nightmares of “blood on my mama and me” from a monster. He talked about liking school but feeling shy with the other children. He said he was not afraid at school “because there is a fence all around and no one could get me.” B.C.'s play over a two-session evaluation was either monotonous, clearly avoided any attempts of discussing family conflicts, or was full of aggressive themes. He wanted to blast the father doll to Mars because his older brother had told him there was no oxygen—“so he can't do any fighting.” He insisted on redrawing the face on his self-portrait multiple times because “I don't want the mouth to be mad.” He admitted he had trouble having fun when he played with his brothers at home because of frequent reminders of when his father would scream at them and “throw our toys at mama's head.” His mother described one particularly poignant episode in which B.C. ran into the yard with his hands covering his ears to try to keep a flashback “out of his head.”

B.C. was diagnosed with posttraumatic stress disorder and separation anxiety disorder. His was given multimodal treatment: individual and family therapy that included educational, cognitive, and behavioral interventions and adjunctive pharmacotherapy. Using stress management strategies (including relaxation training combined with clonidine to target his hyperarousal, startle, and sleep difficulties), B.C. used a narrative exposure task that combined drawing and narration to tell his “Scarytime in San Francisco” storybook—a detailed recounting of his traumatic experiences—in a gradual, desensitizing manner. This allowed him to target traumatic reminders of his father's rages and to habituate to both his painful recall of the traumatic events and his constant fear that his father would explode and kill his mother. Cognitive restructuring allowed him to process associated guilt feelings and horror at his revenge fantasies. Masterful actions taken by him and his family were integrated into his story. He and his mother were assigned homework to desensitize him gradually to being separated from her within, and later outside, their home. The family together designed “safety plans.” Within 6 months B.C.'s most severe symptoms of distress were ameliorated; he no longer met criteria for posttraumatic stress syndrome, and his separation anxiety had decreased significantly. He tolerated discontinuation of his medication without recurring symptoms. B.C. had some remaining self-esteem issues and was assigned a brief 8-week group treatment targeting peer relatedness and self-esteem concerns; afterward he no longer was considered to need treatment beyond occasional participation family therapy. His mother, who continued her own individual therapy (as did his older brother who had been diagnosed with posttraumatic stress disorder and major depressive disorder) was advised that B.C. would potentially need future interventions as he reached developmental milestones that could not help but be affected by his past relationship with his father (e.g., going through puberty or starting to date).

T.R., a 15-year-old white female had asked her maternal uncle to take her to the mall to meet her friends; he agreed, and they stopped at his local bank on the way. They inadvertently walked into a 4 PM bank robbery in progress, and her uncle was shot and killed. The robbers were apprehended 4 blocks away. T.R. had thrown herself to the ground at her uncle's instruction just before he was hit. Police initially thought T.R. had been shot because so much of her uncle's blood was on her clothes. T.R.'s parents kept her home from school several weeks after the funeral, viewing her as too fragile to handle school and fearing that unwanted questions would cause her distress. Her mother, grief-stricken and terrified that something could happen to T.R., would not let her out of her sight. Within a few weeks of resuming school a guidance counselor urged her mother to take T.R. in for a mental health evaluation. While her academics had suffered only mildly as a result of her trauma, the counselor noted that T.R. no longer ate lunch with her friends in the courtyard, cried every day during chorus, and no longer rode the school bus. A child psychiatrist, social worker, and psychologist team evaluated T.R. with rating scales (including parent and teacher child-behavior checklists and general symptom, anxiety, depression, posttraumatic stress disorder, and dissociation self-report checklists by the child) and interviews with both parents and T.R. T.R. had premorbidly been doing well though was temperamentally a bit shy and reserved. Since the traumatic death of her uncle, she displayed major sleep disturbance including difficulty falling asleep and nightmares with death themes. She reported intrusive recall of the robbery and the face of her uncle in death; numerous triggers caused flashbacks of the robbery including the autonomic response of going up the steps at school, which caused a full-fledged panic cascade reminiscent of her state while lying on the floor of the bank. She no longer would go into any commercial public place such as stores or banks. She felt tremendous guilt about her uncle's death—believing he would not have gone to the bank if she hadn't requested a ride to the mall. In fact, this guilt accompanied by the impairments from her anxiety caused tremendous demoralization that seemed to have bloomed into a full-fledged major depressive episode. Cognizant of the overlap in criteria, the team diagnosed T.R. as having posttraumatic stress disorder, major depression, and panic disorder. T.R.'s trauma avoidance strategies (e.g., not leaving her home) were reinforced by her mother, who herself appeared to have developed significant symptoms of posttraumatic stress disorder and separation anxiety. T.R. and her family began a multimodal treatment package consisting of individual therapy for both T.R. and her mother, adjunct psychopharmacological treatment for T.R. and her mother, and four sessions of family therapy. T.R. was guided through a detailed recounting of events preceding, during, and after the bank robbery. T.R. wrote out a timeline of events with accompanying thoughts and feelings describing the event step by step. The therapist carefully noted her current feelings and attributions (e.g., “that was so stupid of me”) about the original events and thoughts as well as stimuli that currently served as traumatic reminders. A series of “worst moments” was detailed along the way, often using drawings and colors in the drawings that identified the most terrifying aspects of the trauma—the ones that would come back to haunt her. Between sessions, exposure-response prevention homework tasks were done (often in conjunction with her mother), beginning with the easiest tasks and moving up a hierarchy to more difficult ones such as separation from her mother and going to a movie with friends. Medication management of T.R. began with paroxetine (Paxil); guanfacine was added after 1 month, and nicely reduced her remaining sleep disturbance and autonomic hyperarousal. Guanfacine dosage was tapered after 3 months of treatment (6 weeks after resolution of symptoms). T.R's individual psychotherapy consisted of approximately 25 to 30 sessions. Her medication (paroxetine) management continued for over 1 year.

SUGGESTED CROSS-REFERENCES

The reader is encouraged to refer to the related anxiety disorders sections on epidemiology (Section 15.2), clinical features (Section 15.6), psychotherapy (Section 15.8) and psychopharmacology. Other related material includes child abuse and neglect (Section 49.4), dissociative identity disorder (Section 49.8), stress and psychiatry (Section 25.9), attention-deficit disorder (Chapter 39), and the child psychiatric treatment section (Chapter 48).

SECTION REFERENCES

Amaya-Jackson L, March JS: Post-traumatic stress disorder in children and adolescents. In *Anxiety Disorders in Children and Adolescents*, JS March, editor. Guilford, New York, 1995.

*American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of post-traumatic stress disorder in children and adolescents. Judith A Cohen, principal author. J Am Acad Child Adolesc Psychiatry *37*(Suppl):4S, 1998.

Cohen JA, Mannerino MP: Factors that mediate treatment outcome of sexually abused preschool children: 6- and 12-month followup. J Am Acad Child Adolesc Psychiatry *37:*44, 1998.

*Davidson JR: Biological therapies for posttraumatic stress disorder: An overview. J Clin Psychiatry *58*(Suppl):29, 1997.

Davidson R, Hughes D, Blazer D, George L: Post-traumatic stress disorder in the community: An epidemiological study. Psychol Med *21:*713, 1991.

*Deblinger E, McLeer SV, Henry D: Cognitive behavioral treatment for sexually abused children suffering post-traumatic stress. J Am Acad Child Adolesc Psychiatry *29:*747, 1990.

*Deblinger E, Steer R, Lippmann J: Maternal factors associated with sexually abused children's psychosocial adjustment. Child Maltreatment *4:*13, 1999.

Eth S, Pynoos R: *Post-traumatic Stress Disorders in Children*. American Psychiatric Press, Washington, DC, 1985.

Famularo R, Kinscherff R, Fenton T: Propranolol treatment for childhood posttraumatic stress disorder, acute type. Am J Dis Child *142:*1244, 1988.

Foa EB, Meadows EA: Psychosocial treatments for post-traumatic stress disorder: A critical review. Annu Rev Psychol *12:*273, 1997.

*Friedman MJ: Posttraumatic stress disorder. J Clin Psychiatry *58*(Suppl):33, 1997.

Garbarino J, Kostelny K, Dubrow N: No place to be a child: Growing up in a war zone. Lexington Books, Lexington, MA, 1991.

Gibbs MS: Factors in the victim that mediate between disaster and psychopathology: A review. J Trauma Stress *2:*489, 1989.

Green B, Korol M, Grace MC, Vary MG, Leonard AC, Gleser GC, Smitson CS: Children and disaster: Age, gender, and parental effects on PTSD symptoms. J Am Acad Child Adolesc Psychiatry *30:*945, 1991.

Horowitz M: *Stress Response Syndrome*. Aronson, New York, 1976.

Jensen JB, Pease JJ, Benset R, Garfinkel BD: Growth hormone response patterns in sexually or physically abused boys. J Am Acad Child Adolesc Psychiatry *30:*784, 1991.

Joseph SA, Brewin CR, Yule W, Williams R: Causal attributions and post-traumatic stress in adolescents. J Child Psychol Psychiatry *34:*247, 1993.

*Kendall-Tackett KA, Williams LM, Finkelhor D: Impact of sexual abuse on children: A review and synthesis of recent empirical studies. Psychol Bull *113:*164, 1993.

Kinzie JD, Sack WH, Angell RH, Manson R, Rath B: The psychiatric effects of massive trauma on Cambodian children: I. The children. J Am Acad Child Psychiatry *25:*370, 1986.

Kramer M, Kinney L, Scharf M: Sleep in delayed stress victims. Sleep Res *11:*113, 1982.

Lonigan CJ, Shannon MP, Finch AJ, Daugherty TK, Taylor CM: Children's reactions to a natural disaster: Symptom severity and degree of exposure. Adv Behav Res Ther *13:*135, 1991.

March JS, Amaya-Jackson L, Murray MC, Schulte A: Cognitive-behavioral psychotherapy for children and adolescents with post-traumatic stress disorder after a single-incident stressor. J Am Acad Child Adolesc Psychiatry *37:*585, 1998.

March J, Amaya-Jackson L, Terry R, Costanzo P: Posttraumatic symptomatology in children and adolescents after an industrial fire. J Am Acad Child Adolesc Psychiatry *36:*1080, 1997.

Martinez P, Richters JE: The NIMH community violence project: II. Children's distress symptoms associated with violence exposure. Psychiatry *56:*22, 1993.

Perry BD: Neurobiological sequelae of childhood trauma: Posttraumatic stress disorders in children. In *Catecholamine Function in Posttraumatic Stress Disorder: Emerging Concepts*, MM Murburg, editor. American Psychiatric Press, Washington, DC, 1994.

Pfefferbaum B. Posttraumatic stress disorder in children: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *36:*1503, 1997.

*Pfefferbaum B, Moore VL, McDonald NB, Maynard BT, Gurwitch RH, Nixon SJ: The role of exposure in posttraumatic stress in youths following the 1995 bombing. J Okla State Med Assoc *92:*164, 1999.

Pittman RK, van der Kolk BA, Orr SP, Greenberg M: Naloxone-reversible analgesic response to combat-related stimuli in post-traumatic stress disorder. Arch Gen Psychiatry *47:*541, 1990.

Pynoos RS, Eth S: Witness to violence: The child interview. J Am Acad Child Psychiatry *25:*306, 1986.

*Pynoos RS, Frederick C, Nader K, Arroyo W, Steinberg A, Eth S, Nunez F, Fairbanks L: Life threat and post-traumatic stress in school-age children. Arch Gen Psychiatry *44:*1057, 1987.

*Pynoos RS, Nader K: Issues in the treatment of post-traumatic stress in children and adolescents. In *The International Handbook of Traumatic Stress Syndromes*, J Wilson, B Raphael, editors. Plenum, New York, 1993.

Saigh PA: The behavioral treatment of child and adolescent posttraumatic stress disorder. Adv Behav Res Ther *14:*247, 1992.

*Shin LM, McNally RJ, Kosslyn SM, Thompson WL, Rauch SL, Alpert NM, Metzger LJ, Lasko NB, Orr SP, Pitman RK: Regional cerebral blood flow during script-drive imagery in childhood sexual abuse-related PTSD: A PET investigation. Am J Psychiatry *156:*575, 1999.

Terr L: Chowchilla revisited: The effects of psychic trauma four years after a school-bus kidnapping. Am J Psychiatry *140:*1543, 1983.

Terr L: *Too Scared to Cry*. Harper Collins, New York, 1990.

van der Kolk BA: The psychobiology of posttraumatic stress disorder. J Clin Psychiatry 58, 1997.

Wolfe VV, Gentile C, Wolfe DA: The impact of sexual abuse on children: A PTSD formulation. Behav Res Ther *20:*215, 1989.

Yule W: Posttraumatic stress disorder in child survivors of shipping disasters: The sinking of the Jupiter. Psychother Psychosom *57:*200, 1992.

46.3 SEPARATION ANXIETY DISORDER AND OTHER ANXIETY DISORDERS

CARRIE SYLVESTER, M.D., M.P.H.

Anxiety disorders have peak ages of onset in childhood and adolescence, although only two, separation anxiety disorder and selective mutism, occur solely in childhood. Anxiety symptoms are also quite common in normal children of normal parents. Thus, early development of these disorders in vulnerable children and adolescents is not surprising. They remain classified as adult disorders, however, because many patients do not come to treatment until they can seek help for themselves as adults.

COMPARATIVE NOSOLOGY

Child psychiatric disorders first appeared in the second edition of *Diagnostic and Statistical Manual of Mental Disorder* (DSM-II) with two child anxiety disorders, withdrawal reaction and overanxious reaction. The former was similar to avoidant disorder in the third edition (DSM-III), the latter to overanxious disorder. Separation anxiety disorder, which first appeared in DSM-III, is the only child anxiety disorder in the fourth edition (DSM-IV). The only significant change in criteria was a duration change from 2 to 4 weeks. Avoidant disorder was subsumed under social phobia and overanxious disorder under generalized anxiety disorder. Elective mutism did not appear as an anxiety disorder of childhood in DSM-III but appeared as selective mutism in DSM-IV. Now viewed as a form of social phobia, it may not survive to another edition of DSM as an independent disorder of childhood.

The International Classification of Diseases (ICD) is a categorical system that has developed parallel to the DSM. Although separation anxiety disorder is essentially the same in both, the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) has some differences from DSM-IV as shown in Table 46.3-1. Phobic anxiety disorder of childhood and generalized anxiety disorder of childhood are specific diagnoses, but descriptions of the disorders are similar to DSM-IV adult diagnoses. Selective mutism retains the older nomenclature of elective mutism in ICD-10.

Table 46.3-1 ICD-10 Diagnostic Criteria for Emotional Disorders With Onset Specific to Childhood

COMORBIDITY

Discussing the anxiety disorders as discrete entities is a heuristic exercise for many patients. Half the children and adolescents with either generalized anxiety disorder or separation anxiety disorder have at least one other anxiety disorder. Specific phobia, in contrast, confers little risk for another disorder. Studies of community samples have found anxiety disorder comorbidity rates from 14 percent in adolescents to 39 percent in children. Depression, attention-deficit/hyperactivity disorder, and conduct disorder or oppositional-defiant disorder are commonly diagnosed in anxious children. Anxiety causes distractibility and inattention so that the diagnosis of attention-deficit/hyperactivity disorder must be approached with caution. Oppositional behavior or the impression of such can be created by the stress of an anxious, impatient parent rushing an anxious child who is unable to decide rapidly or to enter novel situations easily.

Although a specific anxiety disorder diagnosis may not be longitudinally stable, anxious children often develop another anxiety disorder or an affective disorder. Thus, statements about remission of a particular disorder cannot be taken to mean that risk for another anxiety disorder or an affective mood disorder has abated.

SEPARATION ANXIETY DISORDER

Definition The primary feature of separation anxiety disorder is developmentally inappropriate anxiety sufficient to cause clinically significant distress or impairment when faced with separation from home or major attachment figures.

Epidemiology The prevalence rate drops from 4.1 percent in 7- to 11-year-old children, to 3.9 percent in 12- to 16-year-olds, and 1.3 percent in 14- to 16-year-olds. The mean age of onset for this disorder is 7.5 years. Although DSM-IV accepts onset up to 18 years, symptoms of separation anxiety (e.g., school refusal) after 13 years of age usually suggest another major psychiatric disorder. Gender differences in prevalence of separation anxiety disorder are uncertain in children, but in adolescence girls outnumber boys with this disorder. These children commonly come from single-parent homes, which probably explains the association with lower socioeconomic status. No studies specifically implicate genetic factors in the development of separation anxiety disorder discrete from other anxiety disorders or depression.

Etiology Interplay of temperament, attachment, and parenting style is believed key to the development of separation anxiety disorder. A parenting style combining lack of affection and maternal care interacting with overprotection and excessive control has been implicated in school refusal. Lack of parental care and intrusive overcontrol may lead to anxious, insecure attachment as a substrate for separation anxiety disorder and related disorders. Stressful transitions such as moves, changes in schools for other reasons, and hospitalization of children or attachment figures, may precipitate acute onset of separation anxiety disorder.

Diagnosis and Clinical Features Table 46.3-2 lists the eight criteria, of which at least three must be met for at least 4 weeks. Onset before age 6 years warrants the specification of early-onset disorder. The most common age at presentation to a mental health facility is 10.3 years, usually with school refusal. Refusal to participate in other activities that involve separation is more apparent in school-aged children. Younger children are more likely to report preoccupation with thoughts of harm befalling attachment figures and to have somatic complaints that cause them to present to the pediatrician. Parents may complain that the child is intruding on their marriage by insisting on sleeping in their bed.

A. Developmentally inappropriate and excessive anxiety concerning separation from home or from those to whom the individual is attached, as evidenced by three (or more) of the following:
(1) recurrent excessive distress when separation from home or major attachment figures occurs or is anticipated
(2) persistent and excessive worry about losing, or about possible harm befalling, major attachment figures
(3) persistent and excessive worry that an untoward event will lead to separation from a major attachment figure (e.g., getting lost or being kidnapped)
(4) persistent reluctance or refusal to go to school or elsewhere because of fear of separation
(5) persistently and excessively fearful or reluctant to be alone or without major attachment figures at home or without significant adults in other settings
(6) persistent reluctance or refusal to go to sleep without being near a major attachment figure or to sleep away from home
(7) repeated nightmares involving the theme of separation
(8) repeated complaints of physical symptoms (such as headaches, stomach aches, nausea, or vomiting) when separation from major attachment figures occurs or is anticipated
B. The duration of the disturbance is at least 4 weeks.
C. The onset is before age 18 years.
D. The disturbance causes clinically significant distress or impairment in social, academic (occupational), or other important areas of functioning.
E. The disturbance does not occur exclusively during the course of a pervasive developmental disorder, schizophrenia, or other psychotic disorder and, in adolescents and adults, is not better accounted for by panic disorder with agoraphobia.

Specify if:
Early onset: if onset occurs before age 6 years

Table 46.3-2 DSM-IV Diagnostic Criteria for Separation Anxiety Disorder

Tony was a 6-year-old boy referred by his school because of persistent refusal to attend school. He had always been a somewhat “clingy” child, but his difficulties had intensified in the past 4 months. He would follow his parents around the house and refuse to leave their sides in playgrounds or other situations where separating would be age appropriate. That behavior was worse if he could not keep them in clear sight. He said that he might be kidnapped or lost if he ever got out of their sight. He would tantrum whenever they would try to go out for the evening, saying that they were not coming back. When forced to engage in any of the distressing activities, he would complain of various somatic symptoms until one of his parents remained.

He was the product of a pregnancy marred by caesarean section for premature rupture of membranes at 34 weeks gestation. The perinatal course was complicated by moderate respiratory distress necessitating hospitalization for more than 3 weeks. Evaluation of the recent somatic symptoms revealed a child who was normal except for the psychiatric complaints.

His early development had been normal. He attempted preschool, but his mother had removed him because he cried at the beginning of each day, followed by aggression toward smaller peers. She participated as a parent aide in his kindergarten classroom, stating that she was trying to help him adjust to school. He was in grade school with peers who were after-school playmates in his home but would only go to their homes with a parent.

Family history was positive for maternal recurrent major depressive disorder, in remission, and a history of panic disorder in Tony's maternal grandmother. His maternal grandfather and a paternal uncle had alcoholism. Tony's parents were both in their early 30s with some college education. His father worked as an assistant manager in a small service firm; his mother had worked as a bookkeeper until Tony was born. They said that his mother would have returned to work if he had liked preschool. His difficulties were straining their marriage, as he frequently insisted on sleeping in their bed. His mother insisted that she wanted to return to the workforce and that Tony's separation anxiety prevented it. There was no suggestion of domestic violence.

Mental status examination revealed a thin, nicely attired boy who had great difficulty allowing his mother to go beyond a chair outside the office door. He insisted on checking twice whether she was there. He was fidgety and admitted feeling very worried about his mother leaving. He said that kind of worry was bad enough that he has trouble thinking. He denied depressed mood or symptoms including irritability. He denied hallucinations and did not voice any delusional ideas. He said that he enjoyed his friends and that his teacher was “nice” but that he worried about his mother when he was away from her.

Further exploration of the family situation revealed that his maternal grandfather had died 2 years earlier. His grandmother, who lived nearby, had been very agoraphobic until recently and had been dependent on the patient's mother, which had caused significant marital friction. Father had hoped that mother would return to work so that he could complete college and leave a job he hated. When mother tried to force Tony to go to school, grandmother would come to the home to help and would then encourage mother to “go easy on the child.” Mother would become angry, relent, and feel guilty about her interactions with her mother and her child. Grandmother further lectured mother that she needed to stay at home to care for her needy and delicate child because he had been premature.

Treatment focused on bringing the adults together to develop a supportive program for the child and to enlist grandmother in an appropriate role helping her daughter and son-in-law move forward with their lives. Grandmother supported Tony's return to the classroom once she understood the issues and approach. That was accomplished without directly addressing or interpreting her significant identification with her separation-anxious grandchild. Mother returned to work, with grandmother providing after-school care. Tony also responded well to a simple reward program for sleeping in his own room. Unfortunately, he presented again 7 years later as a popular, very good student and athlete who developed major depressive disorder. He responded well to fluoxetine (Prozac) and cognitive therapy.

Differential Diagnosis Social phobia, posttraumatic stress disorder, other anxiety disorders, and adjustment disorder each can present with some features of separation anxiety disorder. Depressed or psychotic children may cling to attachment figures. Inappropriate academic placement, fear of gang violence, or simple truancy due to conduct disorder are other reasons for school refusal. Additionally, culturally determined expectations about separation during the preschool years vary considerably.

Course and Prognosis The 96-percent remission rate for separation anxiety disorder is the highest for a childhood-onset anxiety disorder. The relation between separation anxiety disorder and later onset of panic disorder or depression remains conjectural because the evidence is retrospective. The previously held belief that early separation anxiety disorder is a prodromal symptom of schizophrenia has not been substantiated.

Treatment Parent education is the basis for development of an individualized behavior-modification program. The parents must become comfortable with a structured approach that includes systematic, gradual desensitization and rewards for improvement. Some mildly affected children respond rapidly once parents insist that they attend school or sleep in their own bed with appropriate praise and reward for their efforts. Others need some relaxation training and much support as they spend increasing time at school or move first to a cot in the parents' room and then to their own room.

Children who are unable to respond to support and behavioral management may benefit from pharmacological intervention. Although imipramine (Tofranil) was long the treatment of choice for separation anxiety disorder presenting as school refusal, attempts to replicate the studies have been unsuccessful. Current treatment recommendations include the serotonin reuptake inhibitors in dosages similar to those used for depression. Benzodiazepines or buspirone (BuSpar) may also be useful adjunctive treatments.

GENERALIZED ANXIETY DISORDER

Definition Overanxious disorder in DSM-III and the revised third edition of DSM (DSM-III-R) highlighted preoccupation with future and past events in contrast to generalized anxiety disorder characterized by poorly controlled, excessive worry. Overanxious disorder has been subsumed under generalized anxiety disorder in DSM-IV. Overanxious disorder emphasized concerns about competence and self-consciousness that were vague and commonly found in children of normal parents. Only one of six symptoms of restlessness, fatigue, difficulty with concentration, irritability, muscle tension, and sleep disturbance is required to make the diagnosis of generalized anxiety disorder in children. Affected older children and adolescents have more symptoms than younger children.

Epidemiology Rates of overanxious disorder in children and adolescents ranged from 2.9 to 4.6 percent. Symptoms of generalized anxiety disorder are usually overlooked by parents until their child is in school. Thus, data currently available on peak age of onset are suspect. The sex distribution is probably equal in early childhood, but the ratio of adolescent girls to boys with generalized anxiety disorder is 6 to 1.

Children of anxious parents are more likely to have an anxiety disorder than those with normal parents or parents with other psychiatric disorders. First- and second-degree relatives of children with overanxious disorder have increased rates of generalized anxiety disorder, separation anxiety disorder, and panic disorder. Monozygotic twins are more similar in fearfulness, emotional tone, and approach and withdrawal behavior than dizygotic twins.

Etiology The relations among temperamental aversions to novel situations, attachment, and parenting behaviors as contributors to the development of anxiety

disorders are foci of investigation into the etiology of the anxiety disorders.

Temperament Temperamental traits of passivity, avoidance, fearfulness, and shyness or reticence in novel interpersonal situations may be related to development of anxiety disorders in children, especially when at least one parent has an anxiety disorder or a depressive disorder. These traits have also been correlated with signs of increased autonomic nervous system activity, including tachycardia, laryngeal tension, pupillary dilatation, increased salivary cortisol concentration, and elevated urinary catecholamine concentrations. Also, a genetic trait of high reactivity has been proposed as an animal model for part of the hereditability of anxiety. Rhesus monkeys with that trait react to stressors such as forced separation with retreat, diminished exploration, and increased physiological arousal.

Attachment Infants who are insecurely attached may have more anxiety disorders in childhood and adolescence. Mothers of insecurely attached infants also have poor current and past attachments.

Parenting Parents of anxious children tend to interpret ambiguous situations as threatening; the same response has been shown in anxious children. Their parents provide more-negative responses and are more restrictive; they usually encourage avoidance as a way to handle stressful situations. As a corollary, parenting behaviors may be due to the child's distress. More likely, the family system strikes a dysfunctional balance between child and parent factors.

Diagnosis and Clinical Features These children are referred because of somatic complaints, insomnia, social impairments, or comorbid depression and suicidality. They are often well behaved and eager to please but may intrude on parents with undue concern about finances or minor family problems. Excessive, intrusive reassurance seeking or procrastinating with school work may show their anxiety. They may attempt to limit their families' activities because of difficulty with novel situations.

The two most common measures of childhood anxiety are the Revised Children's Manifest Anxiety Scale (RCMAS) and the State-Trait Anxiety Inventory for Children (STAIC). The RCMAS has an advantage of providing a scale suggesting whether the child is misrepresenting symptoms or misunderstanding the instrument. Neither instrument is useful in making a differential diagnosis. Both are useful in gauging severity and treatment response.

Barney was an 8-year-old boy who was referred by his parents in the first month of third grade because of a 1-year history of increasing anxiety during transitions. He stated his problem as “being nervous about things” and said that he had been “worse” in the past few weeks. He had had difficulty falling asleep and nightmares as long as he could remember. An example was that Frankenstein's monster was chasing him to turn him into a rock. His insomnia increasingly included waking during the night and sleepiness in the morning. He said that he was always a “picky eater” and stated that he did not like to try new things. His parents noted that he cried easily and complained about feeling nervous about going to school. He compared himself unfavorably with other children and expressed much performance anxiety regarding academic achievement. He also described much anticipatory anxiety about not doing well enough at things and worrying about how tired he would be in the morning. He had multiple fears including that “nobody will be nice to me,” spiders, snakes, people breaking into the house, bombs, heights, big dogs, the dark or “stuff under the bed,” getting sick or hurt, and speaking in class. He was preoccupied with fearful thoughts at bedtime. He worried about being kidnapped by robbers and never seeing his parents. However, he could go to camp and tolerate a parent being away for several days on a business trip. The parents usually avoided being away at the same time. There was no suggestion of abuse and no history of recent or significant past losses or trauma.

Barney was the healthy product of a normal pregnancy, labor, and delivery. Early temperament and developmental milestones were normal. His medical history was uneventful, and his most recent physical examination was entirely normal. He changed schools in second grade after having attended a small private school for preschool through first grade. He was in an excellent neighborhood private school and was earning mostly above-average grades. He excelled at basketball and enjoyed competitive team play.

Family history was positive in that mother had recurrent mild major depressive disorder. Maternal grandparents both had severe recurrent major depressive disorder, but their siblings and their children were thus far normal. Barney's parents were successful professionals in their late 40s. Both were in good health and considered this first marriage an important, positive relationship. They enjoyed a lifestyle that included a second home in another region of the country that Barney said he enjoyed.

Mental status examination revealed an attractive, charming, cooperative boy who was developmentally appropriately dressed and groomed. His interactions with his parents were affectionate, and he separated easily from them. His speech was articulate and spontaneous. He described some sad mood, but anxiety predominated. His affect was appropriate to the content of the discussion. He smiled readily and could relax during the interview. The remainder of the examination was entirely normal.

Further discussion over a brief course of therapy revealed that he would read for long periods at night to try to distract himself from his thoughts so he could go to sleep. His parents had stopped reading to him when he achieved early reading proficiency. He also revealed that he would go from one parent to the other when they had differences over his rules. He would then feel overwhelmed by anxiety and guilt about mild parental dissension. He particularly found himself in that situation when he wanted to avoid homework. Treatment included teaching him and his parents relaxation techniques, using basketball game imagery to help him when fears and worries became intrusive. His parents resumed holding him and reading to him at night as a soothing bedtime ritual. He expressed great relief, saying that he could go to sleep easily and that his nightmares dramatically diminished. He and his father, in particular, worked on a variety of cognitive strategies to deal with his fears. An important change in family dynamics was that he adopted a rule of leaving the room if his parents needed to discuss one of his requests. He also stopped making requests of both parents without telling each what the other had said. He admitted to feeling much better having eliminated the main reason his parents would be upset in front of him. Fortunately, his second year in the school went well academically, and his basketball team (on which he had an important starting position) had an excellent season. These external events reinforced the work done in 10 therapy sessions. His parents said that they would return for further evaluation and treatment if new symptoms developed.

Differential Diagnosis Generalized anxiety disorder is distinguished from the other anxiety disorders on the basis of persistent symptoms and a lack of a specific focus for the anxiety. The real worries of children who are ill, poor, or with impaired parents must not be confused with generalized anxiety disorder.

Most substance-related and other medical causes of tension and autonomic reactivity are not seen until adolescence. Relatively common causes are caffeinism and bronchodilator overuse. Uncommon causes include hyperthyroidism, lupus, hypoglycemia, and pheochromocytoma.

Course and Prognosis Eighty percent of children and adolescents with the disorder have remitted at 4-year follow-up. Early age at onset and older age at diagnosis are related to slow remission.

Treatment

Education and Family Therapy Parent education and family therapy are important in the treatment of anxiety disorders in children and adolescents. Understanding which family stressors may be aggravating anxious symptoms in the child is essential. If school performance and test anxiety are prominent features of a child or adolescent's generalized anxiety disorder, consulting with the school may be helpful.

Behavior Therapies Behavioral treatment may be indicated if the child's anxiety is circularly causing and responding to inconsistent limit setting. It also organizes family contexts in which an anxious, depressed, or simply inexperienced parent has difficulty setting priorities for expectations, fails to reward attempts at appropriate behavior, or does not allow a child adequate time to respond. Behavioral programs must emphasize simple, positive statements of desired behaviors rather than prohibitions of undesired behaviors. Negative (“do not”) goals increase ambiguity and anxiety; clear statements of positive expectations coupled with social rewards and encouragement reduce ambiguity and anxiety. Punishment may be included, but the emphasis must be on rewarding even small efforts and one must avoid expecting rapid, error-free improvement.

Pharmacotherapy The current standard of care is that pharmacotherapy must be viewed as adjunctive to other interventions. Results of double-blind controlled studies document the efficacy of selective serotonin reuptake inhibitors (SSRIs) in the treatment of depressive disorders in children and adolescents. Many have wellknown anxiolytic effects and the advantage of treating common comorbid conditions. No comparable data exist for the tricyclic drugs, but they may play a role in management of children with comorbid attention-deficit/hyperactivity disorder or enuresis. Experienced clinicians caution against starting treatment of children and adolescents, especially those with anxiety disorders or risk for bipolar disorders, with more than the equivalent of 2 to 5 mg daily of fluoxetine, with gradual increases

every 2 to 4 days, because younger or anxious patients are particularly at risk for agitation.

The benzodiazepines also have a role in the short-term treatment of anxiety disorders in children, but the risk for behavioral disinhibition is greater in younger patients (Table 46.3-3). Open trials have suggested that buspirone may be efficacious in some young anxious patients without risk for habituation or disinhibition. The starting dosage of buspirone for children and adolescents is 2.5 mg daily or twice a day, with an increase every 3 to 4 days up to a maximum of 20 mg daily for children and 60 mg daily for adolescents. The usual total daily dosage is 0.3 to 0.6 mg/kg. Antihistamines and neuroleptic drugs are no longer recommended anxiolytic agents for children and adolescents.

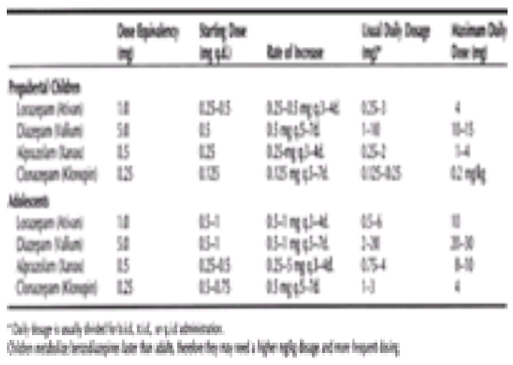

	Dose Equivalency mg	Starting Dose mg q.d.)	Rate of Increase	Usual Daily Dosage mg*	Maximum Daily Dose mg
Prepubertal Children					
Lorazepam (Ativan)	1.0	0.25–0.5	0.25–0.5 mg q.3–4d.	0.25–3	4
Diazepam (Valium)	5.0	0.5	0.5 mg q.5–7d.	1–10	10–15
Alprazolam (Xanax)	0.5	0.25	0.25 mg q.3–4d.	0.25–2	1–4
Clonazepam (Klonopin)	0.25	0.125	0.125 mg q.3–7d.	0.125–0.25	0.2 mg/kg
Adolescents					
Lorazepam (Ativan)	1.0	0.5–1	0.5–1 mg q.3–4d.	0.5–6	10
Diazepam (Valium)	5.0	0.5–1	0.5–1 mg q.5–7d.	1–20	20–30
Alprazolam (Xanax)	0.5	0.25–0.5	0.25–0.5 mg q.3–4d.	0.75–4	8–10
Clonazepam (Klonopin)	0.25	0.5–0.75	0.5 mg q.5–7d.	1–3	4

*Daily dosage is usually divided for b.i.d., t.i.d., or q.i.d. administration.
Children overmetabolize benzodiazepines faster than adults, therefore they may need a higher mg/kg dosage and more frequent dosing.

Table 46.3-3 Benzodiazepine Dosage Regimen for Children and Adolescents

SPECIFIC PHOBIA

Definition Unlike adults, children are not required to acknowledge their fears and worries as unreasonable or excessive to make a diagnosis of a phobic disorder. It is sufficient that the fear causes avoidant behavior that interferes with some aspect of childhood functioning.

Epidemiology Specific phobia occurs in 2.4 to 9.2 percent of children and adolescents, with usual onset between 5 and 13 years of age. Specific phobia occurs equally in young boys and girls, but 9- to 12-year-old girls have more fears than boys. The sex distribution of this disorder favors girls in adolescence. One study of specific phobia probands found that 15 percent of their children had specific phobia; 8 percent of the children of normal controls had it.

Etiology Twin studies suggest that specific phobia has the lowest genetic contribution of any of the anxiety disorders. Psychodynamic theorists extended Sigmund Freud's original conceptualization of phobic displacement of Oedipal urges to emphasize aggressive and social conflicts, but their ideas rest on anecdotal case studies. In contrast, empirical studies have implicated indirect experience of trauma and information transmission as most important in the early development of specific phobias. Direct exposure is a less common pathway for children, in contrast to adults, who more typically acquire phobic avoidance after direct exposure to feared objects.

Diagnosis and Clinical Features Heights, loud noises, injections, insects, dogs, and other animals are among the more common phobias. School phobia, however, is a misnomer for school refusal. School refusal is a symptom of disorders including separation anxiety disorder, social phobia, depressive disorder, and conduct disorder.

Kathy was a 10-year-old girl referred by her pediatrician. Her worsening, nearly lifelong panic response to dogs was interfering with peer relations and apparently causing family conflict. She denied excessive fear of other animals including insects, except for a healthy respect for creatures that her parents had warned her carry rabies. She had a guinea pig for which she provided most of the care. Neither she nor her parents could recall a traumatic incident involving a dog, but barking had always frightened her. She was somewhat afraid of the dark but could fall asleep with a small light in her room.

Medical and developmental history was normal. She was a very good student and was popular despite limits imposed by her fear of dogs. She had, however, recently been subjected to some teasing by peers about her phobia.

Family history was positive in that Kathy's mother recalled that her brother had an excessive fear of the water that took considerable effort to overcome so he could swim. Kathy was the middle of three children of a rural carpenter and homemaker in their mid-30s. Her father and 12-year-old brother wanted a large dog suitable for hunting. Her mother was unenthusiastic, but believed that denying them a dog because of Kathy's phobia was unfair.

Mental status examination revealed a pleasant, attractive, intelligent girl who related easily and had no evidence of a disorder other than specific phobia. She expressed regret that her fear was interfering with her friendships. Kathy was somewhat angry that her father and brother were forcing her to confront this; nevertheless, she added that keeping them from having a dog was unfair and made her sad.

Treatment consisted of a gradual desensitization procedure that began with looking at pictures of dogs while practicing relaxation techniques. We then arranged to have a small, friendly puppy visit three sessions. She gradually could tolerate and finally hold and pet the puppy. Her father and brother agreed to have her work with them to raise a puppy. They chose a gentle, intelligent retriever with a predictable temperament. The remainder of her desensitization included playing with the family puppy and accompanying her father and brother to puppy obedience training. She later attended similar training for dogs with their dog and adult dogs of other breeds so that her reduced fear could generalize. This family fortunately wished to own a large, easily trained dog who provided an optimal experience for Kathy. It also helped that her mother grew to love him too.

Differential Diagnosis Persistence for at least 6 months is required to avoid misdiagnosis of developmentally normal fears. The disorders most easily confused with specific phobia in children include panic disorder, social phobia, obsessive-compulsive disorder, and posttraumatic stress disorder. Delusional disorders mimicking phobias are rare in children.

Course and Prognosis Mild specific phobias are transient, probably because fears are a normal part of childhood. The longitudinal course is affected by the presence of comorbid disorders and the response of important individuals in the child's environment.

Treatment Gradual desensitization is the supportive behavioral treatment of choice. Implosion treatment is contraindicated in children. Guided imagery using relaxation techniques is useful in preparing a child with particularly distressing somatic symptoms for gradual in vivo exposure. Family therapy is useful when parental factors such as shared persistent fears or impatience with systematic desensitization are compromising the child's progress. Medication is rarely recommended in uncomplicated specific phobia. Careful reassessment of the diagnosis and consideration of occult family factors is usually indicated rather than use of an anxiolytic medication.

SOCIAL PHOBIA

Epidemiology Social phobia occurs in 1 percent of children and adolescents, with typical onset in early to midadolescence. Avoidant disorder, now childhood onset social phobia, occurs between 2½ years old and the early school years. Sex distribution is equal in childhood and probably throughout adolescence.

One approach to elucidating familial disposition to social fears is controlled study of the families of behaviorally inhibited children. Although they do not exhibit social phobia when examined at a young age, their parents have very high rates of social phobia, avoidant disorder, and generalized anxiety disorder with considerable comorbidity.

Etiology The factors that lead from avoidant, inhibited temperament to social phobia are poorly understood. As with specific phobias, a combination of real or

vicarious exposure, parental attitudes, and information transmitted to a vulnerable child is a likely pathway to the disorder. No information exists to clarify the relative contribution of genetic variables.

Diagnosis and Clinical Features To subsume DSM-III avoidant disorder, the criteria for DSM-IV social phobia now include fears of unfamiliar people and fear of social evaluation. Children with social phobia report distressing events occurring approximately every other day. They complain of a variety of somatic symptoms and use illness to avoid embarrassing situations. Young children and early adolescents with social phobia report a sense of being scrutinized.

Differential Diagnosis In middle school and high school, children with social phobia may be willing to participate in activities with their close friends while avoiding school when they cannot be in classes with children they know well. Selective refusal of school, while being willing to go away from parents to participate in pleasurable activities with friends, can be easily misinterpreted as oppositional behavior or an aspect of conduct disorder. Misdiagnosis of developmentally normal shyness and normal anxiety engendered by public speaking must be avoided. Severe disorders that may initially seem to be social phobia include major depressive disorder, panic disorder, and (rarely) childhood-onset schizophrenia.

Course and Prognosis Childhood-onset social phobia is the most stable childhood anxiety disorder. The lifelong effects of interference with school attendance or participation in developmentally important activities are obvious. The proclivity for using alcohol to disinhibit shy or avoidant adolescents is another risk of social phobia.

Treatment Systematic desensitization and family therapy are initial treatments aimed at improving support for the child. Social skills training including group therapy and peer mentoring are effective adjunctive treatments.

Medication may have a role in the management of social phobia, especially if the child is refusing to attend school or is self-medicating with alcohol. No placebo-controlled studies of benzodiazepines or antidepressants have been reported for this indication. Studies of children with selective mutism indicate that SSRIs, such as fluoxetine, may benefit these children and adolescents.

PANIC DISORDER

Epidemiology The prevalence of panic disorder in childhood and adolescence is approximately 0.6 percent. There have been anecdotal reports of panic disorder symptoms in children and retrospective reports of childhood onset of panic disorder. It is uncommon before puberty, but the peak age of onset for panic disorder is between 15 and 19 years. The sex distribution is equal in childhood, but the adult female predominance of this disorder rapidly appears after puberty.

Etiology There are no studies that clarify etiological factors for an early onset of panic disorder as opposed to early onset of other anxiety disorders. The genetic and familial nature of panic disorder is well documented.

Diagnosis and Clinical Features The diagnostic criteria for panic disorder are the same for children and adolescents as for adults. As with less psychologically minded adults, children may focus more on somatic symptoms or may express panic as acute separation anxiety if symptoms occur while they are stressed and away from familiar surroundings. They may not tell anyone about the psychological symptoms or discrete panic attacks without systematic inquiry. Because younger children become easily confused about time lines, gathering unbiased interview data is difficult.

Differential Diagnosis Spontaneous panic attacks must be differentiated from specific phobia, social phobia, or generalized anxiety disorder. Associated agoraphobia, fear of going out or fear of crowds, may be confused with separation anxiety disorder in younger patients. Children may develop panic disorder when depressed, especially when separation or loss precipitates depression.

Most medical disorders known to mimic panic disorder have onset in childhood and adolescence. These include asthma, cardiac arrhythmias, thyroid disorders, misuse or abuse of substances including caffeine, and rare tumors such as pheochromocytoma or insulin-releasing tumors. Distinguishing seizure disorder from pseudoseizure secondary to panic is important because brain tumors are the commonest solid tumors of childhood.

Course and Prognosis Panic disorder with onset in childhood and adolescence has the poorest outcome of the childhood-onset anxiety disorders, with a remission rate of 70 percent. Careful monitoring for comorbid depression and associated increased risk for suicide is crucial.

Treatment Education and relaxation training are fundamentals of management of this disorder. Behavioral treatment of associated agoraphobia may be necessary. Supportive or cognitive therapy may be needed to overcome the pathological explanatory models for the frightening experience of spontaneous panic attacks.

Ample open-trial and anecdotal evidence supports the use of SSRIs or high-potency benzodiazepines to treat this disorder in children. Children, adolescents, and individuals with panic disorder are, however, particularly sensitive to rapid increases in either class of medications, and may exhibit agitation, worsening panic, or disinhibition.

SUGGESTED CROSS-REFERENCES

Anxiety disorders are discussed in detail in Chapter 15. Benzodiazepine receptor agonists and antagonists are covered in Section 31.9 and buspirone in Section 31.11. Selective serotonin reuptake inhibitors are explained in Section 31.25. Tricyclic and tetracyclic drugs are further discussed in Section 31.30. Normal child development is covered in Section 32.2. Selective mutism can be found in Section 46.4. Mood disorders of childhood and adolescence are discussed in Chapter 45. Behavior therapy is discussed in Sections 48.3 and Section 30.2 and family therapy in Section 48.5.

SECTION REFERENCES

Anderson JC, Williams S, McGee R, Silva PA: DSM-III disorders in preadolescent children: Prevalence in a large sample from the general population. Arch Gen Psychiatry *44:*69, 1987.

Beidel DC: Social phobia and overanxious disorder in school-age children. J Am Acad Child Adolesc Psychiatry *30:*545, 1991.

*Bender PS: *How to Keep Your Kids From Driving You Crazy: A Proven Program for Improving Your Child's Behavior and Regaining Control of Your Family*. Wiley, New York, 1997.

Benjamin RS, Costello EJ, Warren M: Anxiety disorders in a pediatric sample. J Anxiety Disord *4:*293, 1990.

Bernstein GA, Borchardt CM, Perwien AR: Anxiety disorders in children and adolescents: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *35:*1110, 1996.

*Bernstein GA, Shaw K: Practice parameters for the assessment and treatment of children and adolescents with anxiety disorders. J Am Acad Child Adolesc Psychiatry *36:*69S, 1997.

Black B, Uhde TW: Psychiatric characteristics of children with selective mutism. J Am Acad Child Adolesc Psychiatry *34:*847, 1995.

Bourdon K, Boyd J, Rae D, Burns B, Thompson J, Locke B: Gender differences in phobias: Results of the ECA community survey. J Anxiety Disord *2:*227, 1988.

Bowen RC, Offord DR, Boyle MH: The prevalence of overanxious disorder and separation anxiety disorder: Results from the Ontario Child Health Study. J Am Acad Child Adolesc Psychiatry *29:*753, 1990.

Cantwell DB, Baker L: Stability and natural history of DSM-III childhood diagnoses. J Am Acad Child Adolesc Psychiatry *28:*691, 1988.

*Essau CA, Conradt J, Petermann F: Frequency of panic attacks and panic disorder in adolescents. Depress Anxiety *9:*19, 1999.

Hill SY, Hruska DR: Childhood psychopathology in families with multigenerational alcoholism. J Am Acad Child Adolesc Psychiatry *31:*1024, 1992.

*Hofmann SG, Albano AM, Heimberg RG, Tracey S, Chorpita BF, Barlow DH: Subtypes of social phobia in adolescents. Depress Anxiety *9:*15, 1999.

Kashani J, Orvaschel H: A community study of anxiety in children and adolescents. Am J Psychiatry *147:*313, 1990.

Kearney CA, Silverman W: A critical review of pharmacotherapy for youth with anxiety disorders: Things are not as they seem. J Anxiety Disord *12:*83, 1998.

Kovacs M, Devlin B: Internalizing disorders in childhood. J Child Psychol Psychiatry *39:*47, 1998.

Last CG, Hersen M, Kazdin AE, Finkelstein R, Strauss C: Comparison of DSM-III separation anxiety and overanxious disorders: Demographic characteristics of patterns of comorbidity. J Am Acad Child Adolesc Psychiatry *26:*527, 1987.

Last CG, Hersen M, Kazdin AE, Orvaschel H, Perrin S: Anxiety disorders in children and their families. Arch Gen Psychiatry *48:*928, 1991.

Last CG, Perrin S, Hersen M, Kazdin AE: DSM-III-R anxiety disorders in children: Sociodemographic and clinical characteristics. J Am Acad Child Adolesc Psychiatry *31:*1070, 1992.

*Manassis K: *Keys to Parenting Your Anxious Child*. Barrons, Hauppauge, NY, 1996.

Manassis K, Bradley S: The development of childhood anxiety disorders: Toward an integrated model. J Applied Dev Psychol *15:*345, 1994.

Mineka S, Watson D, Clark LA: Comorbidity of anxiety and unipolar mood disorders. Annu Rev Psychol *49:*377, 1998.

Muris P, Meester C, Merckelbach H, Sermon A, Zwakhalen S: Worry in normal children. J Am Acad Child Adolesc Psychiatry *37:*703, 1998.

Nesse RM, Williams GC: *Why We Get Sick: The New Science of Darwinian Medicine* . Vintage, New York, 1996.

Ollendick TH, King NJ: Empirically supported treatments for children with phobic and anxiety disorders: Current status. J Clin Child Psychol *27:*156, 1998.

Ollendick TH, King NJ, Hamilton DI: Origins of childhood fears: An evaluation of Rachman's theory of fear acquisition. Behav Res Ther *21:*117, 1991.

Parker G: Parental representations of patients with anxiety neurosis. Acta Psychiatr Scand *63:*33, 1981.

*Patterson GR: *Families: Applications of Social Learning to Family Life, Revised.* Research Press, Champaign, IL, 1975.

Reynolds CR, Richmond BO: What I think and feel: A revised measure of children's manifest anxiety. J Abnorm Child Psychol *6:*271, 1978.

Rosenbaum JF, Biederman J, Hirshfeld DR, Bolduc EA, Farone SV, Kagan J, Snidman N, Reznick JS: Further evidence of an association between behavioral inhibition and anxiety disorders: Results from a family study of children from a non-clinical study. J Psychiatr Res *25:*49, 1991.

Spielberger C: *Manual for the State-Trait Anxiety Inventory for Children*. Consulting Psychologists Press, Palo Alto, CA, 1973.

Stein DJ, Bower C: A neuro-evolutionary approach to the anxiety disorders. J Anxiety Disord *11:*409, 1997.

Strauss CC, Last CG: Social and simple phobias in children. J Anxiety Disord *7:*141, 1993.

Thapar A, McGuffin P: Are anxiety symptoms in childhood hereditable? J Child Psychol Psychiatry *36:*439, 1995.

Tracey SA, Chorpita BF, Douban J, Barlow DH: Empirical evidence of DSM-IV generalized anxiety disorder criteria in children and adolescents. J Clin Child Psychol *26:*404, 1997.

Turner SM, Beidel DC, Wolff PL: Psychopathology in the offspring of anxiety disorder patients. J Consult Clin Psychol *55:*229, 1987.

*Ziegenhain U, Jacobsen T: Assessing children's representational attachment models: Links to mother-child attachment quality in infancy and childhood. J Genet Psychol *160:*22, 1999.

46.4 SELECTIVE MUTISM

HENRIETTA L. LEONARD, M.D.

Selective mutism is a disorder of childhood characterized by a total lack of speech in specific situations even though the individual has the ability to speak as demonstrated by speech production in other (usually familiar) settings. Recently, psychodynamic etiological factors have been deemphasized in lieu of a possible reconceptualization of selective mutism as an anxiety disorder. A focus on biologically mediated temperamental and anxiety components currently drives ongoing research on the phenomenology, etiology, and treatment of selective mutism.

DEFINITION

Selective mutism appears under “Disorders Usually First Diagnosed in Infancy, Childhood, or Adolescence” in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). DSM-IV defines selective mutism as a persistent failure to speak in specific social situations (e.g., school) where speaking is expected, despite speaking in other situations. The name was changed from elective mutism to selective mutism with changes from the revised third edition of DSM (DSM-III-R) to DSM-IV. Although the DSM-IV definition is essentially unchanged from that given in DSM-III-R, duration (at least 1 month, not the first month of school) and severity (interfering with social communication or educational or occupational achievement) diagnostic criteria were added in DSM-IV.

HISTORY AND COMPARATIVE NOSOLOGY

The latter part of the nineteenth century saw the first description of a disorder in which people would not speak in some situations, despite having the ability to speak. The use of the term *aphasia voluntaria* in the German literature emphasized the conceptualization of this phenomenon as a voluntary decision not to speak. The nomenclature shifted to *elective mutism* in the 1930s, with the prevailing belief that children were electing not to speak, thereby assigning oppositional characteristics to these children. With the change from DSM-III-R to DSM-IV, the more descriptive term *selective mutism* (implying that children do not speak in select situations) was adopted. This replacement is consistent with new theories of etiology stressing temperamental inhibition, shyness, and anxiety. In the 10th revision of *International Statistical Classifications of Diseases and Related Health Problems* (ICD-10), elective mutism is characterized by a marked, emotionally determined selectivity in speech and is usually associated with social anxiety (see Table 44.1-1).

EPIDEMIOLOGY

Few systematic studies exist to support the claim that selective mutism is a rare disorder. Even with a current (weighted) prevalence rate estimated to be below 1 percent in psychiatrically referred samples, selective mutism may not be so uncommon, as most cases do not come to medical attention and do resolve with age. In one epidemiological study using strict diagnostic criteria, a prevalence rate of 0.06 percent was found in a city cohort of 7-year-olds. Another study reports that transient mutism upon entry to school is much higher, 0.69 percent, with the rate falling to 0.02 percent after 56 weeks.

Selective mutism seems to be more prevalent in females than males, although some clinical reports found only slightly higher frequency in girls. No epidemiological data have been reported on the sex distribution of selective mutism. Onset is usually insidious, but the diagnosis is often first made when the problem is noticed upon entry to school, about age 5 or 6. Boys may present to treatment at an earlier age than girls, which may be attributable to greater tolerance of this behavior in girls.

ETIOLOGY

Various etiological explanations have been proposed. Historically, the disorder has been seen as a child's response to family neurosis, usually characterized by overprotective or domineering mothers and strict or remote fathers. Unresolved psychodynamic conflicts, trauma, divorce, death of a loved one, and frequent moves have all been postulated to play a role in symptom development.

Recently, several investigators have likened selective mutism children to socially phobic adults. Immense discomfiture in social situations and chronic avoidance of interaction seen in socially phobic adults can be compared with selectively mute children's apprehension about speaking to people outside the immediate family or of hearing their own voice. Furthermore, family histories were remarkable for the presence of anxiety disorders, parental public-speaking anxiety, and parental childhood shyness. Nearly all descriptions of selectively mute children have alluded to their shyness, inhibition, or anxiety. Reframing selective mutism as an anxiety disorder or as a symptom of social anxiety may be salient.

Although by definition the diagnosis of selective mutism cannot be made if the symptoms are better ascribed to developmental, organic, or sociocultural obstacles to adequate language acquisition, linguistic factors may be contributory. Despite the lack of systematic speech and language assessments, speech delays or difficulties in clinical samples have been reported. Delayed speech onset, articulation disorders, expressive language disorders, and mixed expressive-receptive language disorders are not uncommon in selectively mute children. The rates of speech and language disorders in this population and the impact of such problems merit further investigation.

Some children who have immigrated into a new linguistic environment have been reported to be selectively mute. Obviously, the diagnosis is precluded in children who have not yet attained fluency in their new language; however, selective mutism may result from resistance to the nondominant language or a sort of culture shock. Children from multilingual homes may display selectively mute behaviors upon starting school where a nondominant language is exclusively used. Familial and community isolation may play a role in such cases.

DIAGNOSTIC AND CLINICAL FEATURES

The hallmark of selective mutism is the failure to use speech in a situation where it is expected despite demonstrable expressive and receptive language competency in other circumstances. Most commonly, children will not speak in school and to unfamiliar adults. Sometimes the child will speak to strangers (e.g., ordering in a restaurant) but not to teachers at school. Some children can speak on the telephone with someone to whom they will not speak face-to-face. Speech is commonly present in the home, and the affected child may be quite talkative around family members. There is usually a distinct hierarchy of people to whom the child will and will not speak.

Usually, selectively mute children are completely silent when mute, but some use whispers or monosyllables in the nonfluent situation. These children vary in their degree of verbal and nonverbal inhibition. In unfamiliar environments, some selectively mute children are shy and anxious, others will interact in some way (e.g., gesturing, nodding their head, smiling) even if they will not talk. The diagnostic criteria for selective mutism are given in Table 46.4-1.

A. Consistent failure to speak in specific social situations (in which there is an expectation for speaking, e.g., at school) despite speaking in other situations. B. The disturbance interferes with educational or occupational achievement or with social communication. C. The duration of the disturbance is at least 1 month (not limited to the first month of school). D. The failure to speak is not due to a lack of knowledge of or comfort with, the spoken language required in the social situation. E. The disturbance is not better accounted for by a communication disorder (e.g., stuttering) and does not occur exclusively during the course of a pervasive development disorder, schizophrenia, or other psychotic disorder.

Table 46.4-1 DSM-IV Diagnostic Criteria for Selective Mutism

Amy is a 5-year-old girl who was brought by her parents because she was not speaking at preschool. The parents reported that this was her second year of preschool and that she had not yet spoken to the teacher or any child at school. She does not use the bathroom at school, because she cannot ask to go. Although she has not spoken to any of the children in her class, she will participate in the class activities. She sits in the group at circle time, and she participates in all the general activities. She will not verbalize if called upon or if she needs anything. In contrast, the parents report that Amy is a “chatterbox” at home. She speaks easily and loudly to her parents and her older and younger sisters. On occasion, she will speak to her grandparents, but she will only consistently talk to one grandmother, whom she knows the best. Amy speaks with her two girlfriends with whom she plays regularly. Even though one of her friends is in her class, she will not speak with her in the classroom. Amy will not answer the phone nor speak out loud in public. She does not order her food in a restaurant nor talk to other children on the neighborhood playground. Her parents describe her as shy with strangers and new situations but quite happy and comfortable with familiar settings. The parents chose this time for seeking an assessment for Amy because they are concerned that she may not speak in kindergarten next fall and that this may be a difficult situation for her.

The parents report that Amy “has always been this way.” Nothing about her behavior at home is particularly concerning. She eats well at home, and she does not have difficulty going to bed at night. She is a “little shy” in new settings, but they don't think that this is particularly unusual. She is not temperamentally difficult, and she is not easily frustrated. On occasion, she may worry about something she has heard or seen. Sometimes she may fear that there is “someone downstairs,” but she has no persistent worries or obsessions. The parents do not feel that Amy is depressed or particularly difficult. They know that she enjoys playing with her good friends. The parents have not noticed any immaturity in her speech and language at home. There was no concern about physical or emotional trauma.

Amy's developmental history was unremarkable. Her developmental milestones were on time, including those for speech and language. She has no significant medical history, and she has never been hospitalized. Amy's mother recalled some separation issues when she was a child, and Amy's father was reportedly “shy as a child.” There is no family history of selective mutism. Amy has never had any formal evaluation for her behavior or her speech. Screening showed that her hearing was within normal limits.

On interview, Amy was inhibited and shy. She did not want her mother to leave the room, and after her mother left the room, she was reluctant to engage in the process, even nonverbally. She did not answer any direct question aloud. Toward the middle of the assessment, she was able to draw pictures with the interviewer. However, she was reluctant to play with the toys in front of the interviewer. There was no evidence of motor hyperactivity, abnormal movements, or thought process or content abnormalities. She appeared shy and slightly apprehensive but was able to draw with the interviewer and participate in some play activities. She appeared more inhibited and shy than one would expect for her age. Her reluctance to engage in the process appeared consistent with her behavior at school. She did not appear depressed or withdrawn.

At the time of the assessment, Amy's presentation was most consistent with a diagnosis of selective mutism.

Associated Features Immaturity, developmental delays, and emotional and behavioral symptoms have been associated with selective mutism, although the older reports may have included a more heterogeneous population than that currently described. Reports on these children in the literature have consistently pointed to their shyness, inhibition, or anxiety. Often emotionally characterized as timid, socially withdrawn, or clinging, selectively mute children exhibit a wide variety of comorbid psychiatric conditions including enuresis, encopresis, obsessive-compulsive features, phobias, school refusal, and depression. The literature reports aggression, hostility, temper tantrums, and oppositional behaviors—more likely manifested at home than elsewhere—but it is most authors' impression that these children are unable to speak in the specific social setting. It is unclear whether an oppositional stance truly exists in even a subgroup of children with selective mutism.

DIFFERENTIAL DIAGNOSIS

A child under consideration for a diagnosis of selective mutism should have a comprehensive evaluation. Many other psychiatric disorders (including pervasive developmental disorders and mental retardation) can incorporate speech inhibition as a secondary symptom, so the differential diagnosis for selective mutism can be complex. Since most selectively mute children will not speak with clinicians, a speech sample recorded on audiotape at home may provide valuable clues. Parents should be interviewed to obtain the child's symptom history, with particular attention to type of onset (sudden or insidious), which may help establish the diagnosis of selective mutism. The child can be evaluated through nonverbal interactions with the diagnostician, direct observations, and adaptive structured interviews and standardized tests to which the child may nod in response.

Other neurological or psychiatric problems (e.g., autistic disorder, and other developmental disabilities, aphasia) should be ruled out in a child presenting with patterns of behavior that deviate from the classic traits of selective mutism, such as not talking to immediate family members, abrupt cessation of speech in one environment, or absence of speech in all settings. Children with Asperger's disorder, right hemisphere deficit disorder, or social-emotional learning disorders may resemble selectively mute children because of their shyness and social isolation, but research suggests that their symptoms are based on an inability to process social cues. A history of neurological insults, developmental delays, neuropsychological deficits, or atypical speech and language difficulties (e.g., prosodic problems) should be sought during the parental interview, as such difficulties may exacerbate speech inhibition. Review of the child's medical and psychiatric history is critical, as physical conditions may underlie the child's mutism. As mentioned, some comorbid diagnoses (e.g., pervasive developmental disorders, schizophrenia) may prelude assignment of the diagnosis of selective mutism. Auditory testing should be performed, especially if the child has a history of frequent ear infections, because hearing problems are sometimes associated with learning and language delays.

A complete family psychiatric and medical history should be obtained. Specifically, the diagnostician should explore with the parents any selective mutism, extreme shyness, or anxiety disorders (e.g., social phobia, panic disorder, obsessive-compulsive disorder) in the family that may predispose the child to developing similar problems.

Distinguishing symptoms due to a communication disorder and symptoms due to selective mutism may be especially problematic. A comprehensive speech and language evaluation is essential; it is a misconception that nonverbal children cannot be evaluated for speech and language functioning. Possible determinants of a child's language delay include a parent with identified speech and language problems or inadequate exposure to the dominant language (as in some bilingual homes), and the influences of these factors should be considered. Children from immigrant families who have had adequate exposure to the language and yet exhibit persistent mutism in school may be given the diagnosis of selective mutism.

Additionally, transient shyness upon entry to school, social phobia, major depressive disorder, and adjustment disorder may produce short-lived speech inhibition in specific social situations; in selective mutism, speech inhibition is protracted.

COURSE AND PROGNOSIS

Selective mutism usually has an insidious onset between 3 and 6 years of age. Parents commonly report that their child has “always been this way.” Identification and referral peak during the first and second grades, when speech expectations at school are continually not met.

Most selective mutism cases are transient and spontaneously resolve. It is not known what (if any) sequelae follow untreated but self-limited selective mutism. In the more severely affected children, disability may last from several months to years. Older children who have not improved by age 10 years may be suffering from a more intractable form of the disorder. Some reports have revealed a few cases in which a chronic, unremitting course ended in pervasive psychiatric disorder.

Children who are symptomatic for longer than 6 months should be evaluated and treated. The long-term outcome of children treated for selective mutism is not clear. Most are improved at follow up, but some speech, behavioral, and emotional symptoms may remain. In light of the new theories on temperamental and anxiety components in selective mutism, systematic research is warranted to determine whether these remnants of shyness or anxiety persist and fulminate into an anxiety disorder in adulthood. Early identification and intervention, with cooperation from both home and school, indicate the best prognosis.

TREATMENT

Although many different approaches have been taken, treatment for selective mutism has long been considered difficult. Case studies, often with only a single subject, dominate the treatment reports. Methodological issues obscure the efficacy and generalizability of various behavioral, psychodynamic, family, speech or language, and pharmacological interventions, since few systematic studies have been conducted with any of these modalities. A literature is now emerging on the pharmacotherapy of selective mutism.

Behavior Therapy Behavioral techniques based on the principles of learning theory have been used most frequently to treat selective mutism. Reinforcement (often combined with an absence of reinforcement for the mute behavior), stimulus fading (gradual stepwise desensitization), shaping (increasing approximations to the desired behavior), and self-modeling (audio or video " *dubbed adaptive* behaviors") have all been tried with variable success. Unfortunately, only one controlled study of behavioral therapy exists to date. Subjects in this study who were randomized to receive active treatment had significantly more vocalizations than untreated subjects at 5 weeks, but the improvement was not significant at 1-year follow-up. Behavioral therapy is usually an important component of the treatment plan.

Psychodynamic Therapy Once the preferred treatment for selective mutism, psychodynamic therapy focused on identifying and resolving underlying intrapsychic conflicts. Art and play were frequently incorporated into sessions to facilitate communication with the selectively mute child, although the therapeutic process was often arduous and time consuming. Presently, cognitive-behavioral approaches are being used with increasing frequency in the psychotherapeutic relationship.

Family Therapy The role of family therapy has shifted somewhat in light of the contemporary clinical perspectives. Family therapy was once used to identify and treat dysfunctional patterns thought to be a causal factor in the development and preservation of selective mutism, but more recently family members have been engaged in designing and implementing the child's treatment plan. Family pathology is no longer seen as the (primary) cause of the symptoms; however, if family problems are affecting the child, a more traditional, insight-oriented family treatment approach could be warranted.

Pharmacotherapy Some medications that have been effective for social phobia have been reported for the treatment of selective mutism. One patient reportedly responded to phenelzine (Nardil), up to 2 mg a day; others reported improvement with 20 mg a day of fluoxetine (Prozac) or fluvoxamine (Luvox). In a double-blind, placebo-controlled 12-week trial of fluoxetine, the six children on active medication showed significant improvement on some, but not all, ratings of mutism and anxiety, but subjects in both groups remained symptomatic at 12 weeks. An open trial of fluoxetine in graduated doses was conducted by other investigators, and 16 of 21 children (76 percent) were improved at the end of the 9-week study. A longer trial, flexible dosing, or multimodal interventions may be appropriate. If pharmacotherapy is prescribed, it is usually accompanied by behavioral therapy. The efficacies of fluoxetine and fluvoxamine for the treatment of selective mutism are currently under study. If anxiety is a prominent factor or if treatment-resistant symptoms exhaust other options, a medication trial should be considered.

Individualized Educational Plan A multidisciplinary, individualized educational plan can be implemented with the cooperation of parents, clinicians, and teachers. Teachers can carry out simple interventions in the classroom to decrease speech anxiety and encourage the student to interact and communicate. With the increased prevalence of speech and language problems noted in the literature, speech and language therapy may help the selectively mute child. Intensive speech therapy focusing on articulation and language training can provide structured practice whether or not the child has an identified speech and language disability. An integrated approach to treatment with the coordinated efforts of all involved may produce the best outcome. Since the comparative efficacies of these different methods are obscure, further systematic research is needed.

The author thanks Mai Karitani and Sara P. Dow for their help.

SUGGESTED CROSS-REFERENCES

Chapter 37 on communication disorders describes some differential diagnoses for selective mutism. Children with pervasive developmental disorders (Chapter 38) and mental retardation (Chapter 34) may exhibit mutism. Lack of speech output may also occur in mood disorders, discussed in Chapter 45, or in anxiety disorders, considered in Chapter 46.

SECTION REFERENCES

Bishop DVM: Developmental disorders of speech and language. In *Child and Adolescent Psychiatry: Modern Approaches,* M Rutter, E Taylor, L Hersov, editors. Blackwell, Oxford, 1994.

*Black B, Uhde TW: Elective mutism as a variant of social phobia. J Am Acad Child Adolesc Psychiatry *31:*1090, 1992.

*Black B, Uhde TW: Treatment of elective mutism with fluoxetine: A double-blind, placebo-controlled study. J Am Acad Child Adolesc Psychiatry *33:*1000, 1994.

Black B, Uhde TW: Psychiatric characteristics of children with selective mutism: A pilot study. J Am Acad Child Adolesc Psychiatry *34:*847, 1995.

Black B, Uhde TW, Tancer ME: Fluoxetine for the treatment of social phobia. J Clin Psychopharmacol *14:*293, 1992.

Blum NJ, Kell RS, Starr HL, Lender WL, Bradley-Klug KL, Osborne ML, Dowrick PW: Case study: Audio feedforward treatment of selective mutism. J Am Acad Child Adolesc Psychiatry *37:*40, 1998.

Boon F: The selective mutism controversy. J Am Acad Child Adolesc Psychiatry *33:*283, 1994.

Brown JB, Lloyd H: A controlled study of children not speaking at school. J Assoc Workers Malad Child *3:*49, 1975.

Browne E, Wilson V, Laybourne PC: Diagnosis and treatment of elective mutism in children. J Am Acad Child Psychiatry *2:*605, 1963.

Calhoun J, Koenig KP: Classroom modification of elective mutism. Behav Res Ther *4:*700, 1973.

Crumley FE: The masquerade of mutism. J Am Acad Child Adolesc Psychiatry *29:*318, 1990.

*Dow SP, Sonies BC, Scheib D, Moss SE, Leonard HL: Practical guidelines for the assessment and treatment of selective mutism. J Am Acad Child Adolesc Psychiatry *34:*836, 1995.

Dummit ES, Klein RG, Tancer NK, Asche B, Martin J: Fluoxetine treatment of children with selective mutism: An open trial. J Am Acad Child Adolesc Psychiatry *35:*615, 1996.

Dummit ES, Klein RG, Tancer NK, Asche B, Martin J, Fairbanks JA: Systematic assessment of 50 children with selective mutism. J Am Acad Child Adolesc Psychiatry *36:*653, 1997.

Fundudis T, Kolvin I, Garside R: *Speech Retarded and Deaf Children: Their Psychological Development*. Academic Press, London, 1979.

Giddan JJ, Ross GJ, Sechler LL, Becker BR: Selective mutism in elementary school: Multidisciplinary interventions. Lang Speech Hear Serv Schools *28:*127, 1997.

Goldwyn DH, Weinstock RC: Phenelzine treatment of elective mutism: A case report. J Clin Psychiatry *51:*384, 1990.

Guna-Dumitrescu L, Pelletier G: Successful multimodal treatment of a child with selective mutism: A case report. Can J Psychiatry *41:*417, 1996.

Hesselman S: Elective mutism in children 1877–1981: A literary summary. Acta Paedopsychiatr *49:*297, 1983.

Kolvin I, Fundudis T: Elective mute children: Psychological development and background factors. J Child Psychol *22:*219, 1981.

Kopp S, Gillberg C: Selective mutism: A population based study: A research note. J Child Psychol Psychiatry *38:*257, 1997.

*Kratochwill TR: *Selective Mutism: Implications for Research and Treatment*. Erlbaum, Hillsdale, NJ, 1981.

Kristensen H: Elective mutism associated with developmental disorder/delay: Two case studies. Eur Child Adolesc Psychiatry *6:*234, 1997.

Kumpulainen K, Rasanen E, Raaska H, Somppi V: Selective mutism among second-graders in elementary school. Eur Child Adolesc Psychiatry *7:*24, 1998.

Labbe EE, Williamson DA: Behavioral treatment of elective mutism: A review of the literature. Clin Psychol Rev *4:*273, 1984.

Lafferty JE, Constantino JN: Fluvoxamine in selective mutism (Letter). J Am Acad Child Adolesc Psychiatry *37:*12, 1998.

Leonard H, Dow S: Selective mutism. In *Anxiety Disorders in Children and Adolescents*, JS March, editor. Guilford, New York, 1995.

Pigott HE, Gonzales FP: Efficacy of video tape self-modeling in treating an electively mute child. J Clin Child Psychol *16:*106, 1987.

*Rupp SN: Haloperidol for Tourette's disorder plus selective mutism. J Am Acad Child Adolesc Psychiatry *38:*7, 1999.

Schneier FR, Chin SJ, Hollander E, Liebowitz MR: Fluoxetine in social phobia. J Clin Psychopharmacol *12:*62, 1992.

Steinhausen HC, Juzi C: Elective mutism: An analysis of 100 cases. J Am Acad Child Adolesc Psychiatry *35:*606, 1996.

Tancer NK: Elective Mutism. In *Advances in Clinical Child Psychology*, vol 14. BB Lahey, AE Kazdin, editors. Plenum, New York, 1992.

Voeller K: Right-hemisphere deficit syndrome in children. Am J Psychiatry *143:*1004, 1986.

Weintraub S, Mesulam M: Developmental learning disabilities and the right hemisphere. Arch Neurol *40:*463, 1983.

Wilkins R: A comparison of elective mutism and emotional disorders in children. Br J Psychiatry *146:*198, 1985.

Wright HH: Early identification and intervention with children who refuse to speak. J Am Acad Child Adolesc Psychiatry *24:*739, 1985.

Wright HH, Cuccaro MD, Leonhardt TV, Kendall DF, Anderson JH: Case study: Fluoxetine in the multimodal treatment of a preschool child with selective mutism. J Am Acad Child Adolesc Psychiatry *34:*857, 1995.

Wright HL: A clinical study of children who refuse to talk in school. J Am Acad Child Psychiatry *7:*603, 1968.

CHAPTER 47. EARLY-ONSET SCHIZOPHRENIA

JON M. MCCLELLAN, M.D.

Schizophrenia is a neurodevelopmental disorder with deficits in cognition, affect, and social relatedness. Although it rarely occurs in children, the incidence of schizophrenia increases steadily after the onset of puberty. It is often a chronic condition, with significant long-term morbidity and functional impairment. Therefore, clinicians who work with children and adolescents need to be familiar with the phenomenology, clinical course, and treatment of this disorder.

DEFINITION

Schizophrenia is diagnosed in children and adolescents using the same criteria as for adults. The disorder is defined by the presence of positive and negative symptoms. Positive symptoms consist of hallucinations, delusions, or bizarre and disorganized thinking and behavior. Negative symptoms consist of paucity of speech, paucity of thought content, apathy, avolition, and flat affect. These symptoms are coupled with a significant deterioration in functioning, including the failure to achieve expected levels of social development.

HISTORY

Although schizophrenia is primarily considered an adult disorder, its occurrence in children has been described since the time of Emil Kraeplin. However, as the concept of childhood psychoses evolved, autistic disorders and pervasive developmental disorders became classified as childhood schizophrenia. As a result, much of the early literature regarding childhood schizophrenia is actually about autistic and psychotic disorders.

Israel Kolvin and others challenged this nosological construct by demonstrating that autistic disorder was a discrete disorder, distinct from schizophrenia. Thus, beginning with the ninth revision of the *International Statistical Classification of Diseases* (ICD-9) and third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III), childhood schizophrenia was differentiated from pervasive developmental disorders and diagnosed using the same criteria as for adults. Subsequent research has generally validated this decision.

There is a paucity of research examining schizophrenia in youth. Treatment studies are generally lacking. Although the incidence of the disorder increases after the onset of puberty, much of the existing literature focuses on childhood onset. Other methodological problems include retrospective designs, lack of standardized assessment tools such as diagnostic interviews, small subject pools, and lack of comparison groups. However, despite these limitations, reasonable conclusions can be drawn from the existing studies regarding the diagnosis of schizophrenia in youth. Furthermore, because schizophrenia in youth appears to be essentially the same heterogeneous disorder as in adults, the adult literature can generally be extrapolated to children and adolescents as long as developmental issues are taken into account.

Although schizophrenia in children under age 13 years has often been described as prepubertal, most studies have determined this by age, not by physical development. To avoid ambiguity, early-onset schizophrenia has been defined as onset prior to 18 years of age, with a subgroup of very early-onset schizophrenia defined as onset before age 13 years.

COMPARATIVE NOSOLOGY

The conceptual notion of schizophrenia, including what are considered characteristic symptoms, remains basically unchanged. Since the recognition that childhood schizophrenia and autistic disorder were distinct conditions, only minor modifications have been made within the specified diagnostic criteria. The revised third edition of DSM (DSM-III-R) required that the active psychotic symptoms last at least 1 week, whereas in the fourth edition of DSM (DSM-IV) these symptoms must persist for at least 1 month. The other significant change is that DSM-IV added negative symptoms as one of the characteristic symptoms of the active phase of the disorder.

The diagnostic criteria defined by the ICD-10 are similar to DSM-IV. The only difference is that ICD-10 only requires a total duration criterion of 1 month, compared to the 6-month duration specified by DSM-IV. A high rate of diagnostic agreement was found between DSM-III-R, DSM-IV, and the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) in hospitalized psychotic adolescents.

EPIDEMIOLOGY

The prevalence of schizophrenia in youth has not been adequately established. Clinical experience suggests that schizophrenia with onset prior to age 12 years is rare. It has been estimated that 0.1 to 1 percent of all schizophrenic disorders present before age 10 years, with 4 percent occurring prior to 15 years of age. The rate of onset increases sharply during adolescence, with the peak ages of onset generally ranging from 15 to 30 years.

A study examining young adults during their first episode of schizophrenia (*N* = 232) found that 47 percent had displayed the first sign of their illness before age 21 years. However, only 21 percent developed psychotic symptoms prior to this age. There are reported cases of onset prior to age 6 years. However, a diagnosis of schizophrenia in young children must be carefully scrutinized.

Schizophrenia in youth, especially in children occurs predominantly in males. As age increases, this ratio tends to even out. Since the adult literature suggests that the age of onset in males is significantly younger than that in females, the male predominance may be a cross-sectional effect.

ETIOLOGY

Schizophrenia is a heterogeneous disorder of unknown cause. Etiological mechanisms are undoubtedly complex, and remain to be elucidated. However, at least three potential causal mechanisms have been identified: genes, neurodevelopmental insults, and viral exposure. It has been hypothesized that a multifactorial polygenic model best explains the development of schizophrenia. In this model a variety of risk factors, genetic and environmental, interact in an additive fashion to produce a vulnerability towards schizophrenia. If a threshold is reached, the individual will develop the illness. Subthreshold cases may present with related conditions, such as schizotypal personality disorder. The relative risks of different etiological mechanisms, and how these factors interact, remain an important area of future research.

Genetic Factors There is substantial evidence that genetics play an important role in the development of schizophrenia, including data from both twin and adoption studies. The lifetime risk of developing schizophrenia is 5 to 20 times higher in first-degree relatives of affected probands when compared to the general population. In early-onset schizophrenia the genetic risk appears to be at least as high as for the adult-onset form. Moreover, some studies have found that an earlier age of onset is associated with an increased familial risk.

Despite this evidence, no single model of genetic inheritance has been identified. Although some studies have identified potential sites on the genome that were linked to schizophrenia, these associations have not been consistently found or replicated. Because schizophrenia is a heterogeneous disorder, it is likely that multiple genes are involved. Moreover, the risk conferred may not be always expressed as schizophrenia, but rather as related traits such as a schizoid personality. Finally, some cases are assumed to be sporadic and as resulting from environmental causes rather than genes. The current evidence suggests that the development

of schizophrenia is best explained by a multifactorial polygenic model.

Neurodevelopmental Factors Schizophrenia may be a neurodevelopmental disorder with early central nervous system lesions affecting normal maturational processes. Thus, premorbid abnormalities and developmental delays may represent the early neuropathological manifestations of the disorder. Perinatal complications, alterations in brain structure and size, minor physical anomalies, and disruption of fetal neural development, especially during the second trimester of pregnancy, have been correlated with schizophrenia. An earlier age of onset (before age 22 years) has been associated with an increased rate of obstetrical complications. It has been hypothesized that early neurological trauma may interact with genetic risk factors to produce schizophrenia in at-risk individuals.

Viral Exposure Viral exposure has been implicated as a risk factor for schizophrenia. Individuals with schizophrenia have a higher than expected rate of being born during winter months, which suggests the involvement of an infectious agent. Furthermore, several studies have found an increase in viral titers, or immunological markers, or both, in schizophrenic patients. Finally, birth cohorts who were in the second trimester of pregnancy during an influenza epidemic have been shown to have an increased risk for schizophrenia. However, efforts to directly link schizophrenia to maternal influenza infections have produced mixed findings. Further research is needed to clarify what role, if any, exposure to viral infections has in the development of schizophrenia. One proposed mechanism is that viral exposure produces the illness in some at-risk individuals by an autoimmune process.

Psychological Factors There is no evidence that psychological factors alone cause schizophrenia. Rather, psychological factors interact with biological risk factors, primarily influencing the course of the disorder. Psychosocial stressors, including expressed emotion within the family setting, influence the onset or exacerbation of acute episodes and relapse rates. However, these interactions are complex and bi-directional. The presence of difficult family interactions may not be “causal”, but rather a reaction to the collection of difficulties the patient brings to the family setting.

DIAGNOSIS AND CLINICAL FEATURES

Diagnosis The diagnostic criteria for schizophrenia are outlined in the DSM-IV, (see Table 12.7-1). Diagnostic tools, such as structured interviews, symptom scales, and diagnostic decision trees, help to ensure the reliability and veracity of diagnosis. Because schizophrenia is a serious disorder with ominous prognosis and social stigma, it is important that the criteria be adequately assessed before a diagnosis is made. However, that is not tantamount to saying that schizophrenia should not be diagnosed when present because that potentially denies the child and family access to appropriate treatment and psychoeducational resources.

Characteristic Symptoms At least two of the following are needed for a period of at least 1 month; delusions, hallucinations, disorganized speech, grossly disorganized or catatonic behavior, and/or negative symptoms. The duration of these symptoms may be less than 1 month if they have been alleviated by treatment.

Social or Occupational Dysfunction The level of social, occupational, and self-care functioning has markedly deteriorated below the highest previously obtained level. In children and adolescents, this includes the failure to achieve expected levels of social development.

Duration The disturbances must be present for a period of at least 6 months. This includes an active phase of overt psychotic symptoms (criterion A) with or without a prodromal or residual phase. A prodromal phase involves the deterioration in functioning prior to the onset of psychotic symptoms, while the residual phase follows the active phase. Symptoms characteristic of both prodromal and residual phases include marked social isolation, deterioration in occupational functioning, peculiar behavior such as food hoarding, poor hygiene, blunted or inappropriate affect, disordered thought processes (tangentiality, circumferentiality, poverty of speech or speech content), odd beliefs or perceptions, and anergia.

Exclusion of Mood Disorder Schizoaffective disorder and mood disorder with psychotic features must be ruled out. If mood episodes are present, their duration must be brief relative to the course of the schizophrenic symptoms. These exclusion criteria are especially pertinent for adolescents with bipolar I disorder, since manic episodes in this age group frequently include schizophrenia-like symptoms at onset.

Substance and General Medical Condition Exclusion The symptoms are not due to either the direct effects of a substance or to a medical condition.

Relationship to a Pervasive Developmental Disorder If there is a history of autism, or another pervasive developmental disorder, prominent delusions or hallucinations (for at least 1 month unless successfully treated) are required to make a diagnosis of schizophrenia.

Clinical Features Although this disorder is diagnosed using the same criteria as are used for adults, some specific developmental characteristics have been noted.

Premorbid Functioning Schizophrenia in youth generally has an insidious onset, with a longstanding history of developmental and personality abnormalities. An age of onset before age 12 years is associated with highest rates of premorbid problems. The abnormalities most frequently described include (1) being socially withdrawn, odd, and isolated; (2) behavioral disorders; and (3) multiple developmental delays, including lags in cognitive, motor, sensory, and social functioning. Autistic disorder and pervasive developmental disorders have been reported.

These premorbid features are not equivalent to a diagnosis of schizophrenia. Many children are considered odd, or have multiple developmental problems. However, most will not have the prerequisite psychotic symptoms required to make the diagnosis.

Symptomatology In studies of early-onset schizophrenia hallucinations, thought disorder, and flattened affect are consistently found; systematic delusions and catatonic symptoms are less common. Developmental differences in language and cognition may influence the spectrum and quality of symptom presentation. Children tend to have less complex delusions, with the content reflecting childhood themes.

Most hallucinations in children lack the persistence and associated symptomatology required for schizophrenia. These diagnostic issues have led to high rates of misdiagnosis, especially in children. A child's report of hallucinations may represent: imagination; misinterpretation of normal intrapsychic experiences; misunderstanding of the clinician's question; developmental phenomena; dissociative phenomena; factitious symptoms; or symptoms of another psychotic illness (i.e., a mood disorder). These issues must be examined prior to diagnosing a child with schizophrenia, given its long-lasting treatment and prognostic ramifications.

Significant rates of formal thought disorder, including loose associations and illogical thinking, have been found in children with schizophrenia. It is important, however, to differentiate the thought disorder of psychosis from that of either developmental delays or language disorders.

Rapidity of Onset Schizophrenia with very early onset generally has an insidious onset. Conversely, reports of onset in early adolescence describe high rates of acute or subacute onset, defined as less than 1 year; other reports found a predominance of insidious onset.

Subtype Reports of schizophrenia with childhood onset vary as to whether the paranoid type or the undifferentiated type are more common. At this time sufficient evidence is lacking to justify categorizing early-onset schizophrenia as a separate diagnostic group.

Cognitive and Language Dysfunction Approximately 10 to 20 percent of youth with this disorder have a low intelligent quotient (I.Q.) scores. Language and communication deficits are common. Children with schizophrenia have deficits in their information-processing capacities, a finding also noted in adults.

Caution must be used in making the diagnosis of schizophrenia in children who have severe language impairments. Since the standard presentations of hallucinations and delusions involve language, assessing a child with severely impaired language for psychotic symptoms can be a diagnostic challenge. In these cases, the clinician is dependent on observations of behavior. Psychotic symptoms may be identified when their emergence is associated with a deterioration in mental status and global functioning.

Mortality In follow-up studies, the risk of suicide or accidental death directly due to psychosis appears to be approximately 5 percent. However, this estimate needs to be further studied given the small number of subjects examined and the limited length of follow-up periods.

An 11-year-old white male presented with a greater than 1-year history of increasingly bizarre behavior, including: confused and tangential thinking, seeing ghosts, hearing voices, and taking on the identity of fictional media characters (Teenage Mutant Ninja Turtles). He had no noted major episodes of mood symptoms. He also had a chronic history of behavior problems, including decreased attention span, poor impulse control, severe oppositional behaviors, aggression, and sexually inappropriate behaviors. He was socially odd and isolated, and avoided interacting with peers unless it was in an aggressive or intrusive fashion.

Developmentally, his first 4 years were chaotic. He was intermittently exposed to domestic violence, possibly physical abuse, and a lack of structure and supervision. There was no documented history of sexual abuse. His developmental milestones were slightly delayed, with onset of speech at age 18 months and of walking at age 15 months. By the time he entered school, cognitive and academic delays were apparent. Intellectual testing at age 8 years displayed a full-scale I.Q. of 71. When retested at age 10 years, his intellectual functioning had deteriorated, with an I.Q. of 58.

He had no significant medical problems. An organic workup was unremarkable, including normal chromosomes, a metabolic urine screen, an electroencephalograph (EEG), and a computed tomography (CT) scan of the head. He had a family psychiatric history of substance abuse in second-degree relatives. However, much of his paternal family psychiatric history was unknown.

His mental status examination was significant for tangential and disorganized thinking, loose associations, and perseverative speech patterns. His moods were labile, his affect silly and inappropriate. He was preoccupied with the belief that he was a Teenage Mutant Ninja Turtle. He also reported auditory hallucinations, for example, hearing God, the devil, and the turtles. Socially, he was awkward and immature; behaviorally, he was oppositional and intermittently aggressive.

Since the initial referral, his psychotic symptoms have partially responded to various trials of antipsychotic medications, in addition to mood stabilizers for his explosive behaviors. However, his intermittent aggression, sexualized behaviors, and developmental delays have necessitated several long-term hospitalizations. At age 15 years he remained severely impaired, with ongoing delusions (e.g., the Teenage Mutant Ninja Turtles have a computer chip controlling his brain), disorganized thinking, and impaired social relationships.

PATHOLOGY AND LABORATORY EXAMINATIONS

There are few studies examining the biological markers of schizophrenia. In an ongoing study of youth with schizophrenia with very early onset at the National Institutes of Mental Health (NIMH), neurobiological abnormalities have been noted. These include deficits in smooth pursuit eye movements and autonomic responsivity, which are similar to those reported in the adult literature. Using magnetic resonance imaging (MRI) subjects with schizophrenia had larger left frontal ventricular horns, larger left caudate regions, and a lack of normal caudate asymmetry when compared to normal controls. A progressive increase in ventricular size was seen in these subjects over a 2-year period. Smaller total cerebral volumes were correlated with negative symptoms. In a separate study, magnetic resonance spectroscopy showed frontal lobe dysfunction in 13 children with schizophrenia.

These neurobiological findings are important research efforts to understand the underlying etiological mechanisms of the disorder; however, none are diagnostic. Laboratory evaluations and neuroimaging techniques do have a role in the standard assessment of the condition, but primarily for ruling out other medical disorders.

DIFFERENTIAL DIAGNOSIS

When assessing a child or adolescent with symptoms suggestive of schizophrenia, a comprehensive diagnostic evaluation is needed to rule out other conditions that present with similar symptomatology. A thorough review of presenting symptoms, course, premorbid functioning, adherence to DSM-IV criteria, how psychotic symptoms present in this age group, and family psychiatric history all will help improve the accuracy of diagnosis. However, discriminating between these various disorders still may be difficult, especially at the initial presentation. Therefore, the diagnosis needs to be periodically reassessed.

Mood Disorders Both schizophrenia and psychotic mood disorders (especially bipolar I disorder) typically present with a variety of affective and psychotic symptoms. This overlap in symptomatology increases the likelihood of misdiagnosis at the time of onset. Approximately one-half of adolescents with bipolar I disorder may be originally misdiagnosed as having schizophrenia. Longitudinal reassessment is therefore needed to ensure accuracy of diagnosis. Family psychiatric history may also be a helpful differentiating factor, although it is important to note that studies also have found an increased family history of depression in young people who have schizophrenia.

Schizoaffective Disorder Early-onset schizoaffective disorder has not been well defined in this age group. Twenty-eight percent of a sample of schizophrenia with childhood onset at follow-up had schizoaffective psychoses, which is an ICD-9 diagnosis that overlaps with DSM-III-R diagnoses of bipolar disorder and schizoaffective disorder. Other follow-up studies of psychotic youth have also found this disorder, but at lower rates.

Nonpsychotic Behavioral or Emotional Disorders (Including Dissociative or Personality Disorders) Youth with conduct and other nonpsychotic emotional disorders may report psychotic-like symptoms, and thus be misdiagnosed as having a primary psychotic disorder. When compared to children with psychosis, these children have lower rates of delusions and thought disorder. At follow-up, an increase in personality dysfunction, including personality disorders, but not psychotic disorders, has been found. When there is a history of abuse or neglect, the psychotic-like symptoms may represent dissociative symptoms. It is important that these children be accurately characterized because a misdiagnosis of schizophrenia may unnecessarily expose them to the long-term adverse effects of antipsychotic medications.

Pervasive Developmental Disorder Autistic and other pervasive developmental disorders are distinguished by the absence or transitory nature of the required positive psychotic symptomatology, that is, hallucinations and delusions, as well as by the predominance of the characteristic deviant language patterns, aberrant social relatedness, and other key symptoms that characterize these disorders. The earlier age of onset and the absence of a normal period of development are also indicative, although some children with schizophrenia do have a lifelong history of developmental delays. However, compared to pervasive developmental disorder the premorbid abnormalities in schizophrenia tend to be less pervasive and severe.

Childhood disintegrative disorder resembles autistic disorder except that the onset occurs after 2 or more years of normal development. Children with Asperger's disorder lack the marked language disturbances associated with autistic disorder but present with deficits in social relatedness, contextual communication (especially with social cues), and a restricted (and possibly bizarre) range of interests. The lack of overt hallucinations and delusions distinguishes both these conditions from schizophrenia.

Communication Disorders Children with developmental speech and language disorders may be mistakenly diagnosed as having a thought disorder. Such children do not, however, have other prerequisite symptoms of schizophrenia such as hallucinations, delusions, or odd social relatedness.

Other Disorders Other disorders that need to be differentiated from schizophrenia include schizotypal and schizoid personality disorders and other psychotic disorders (e.g., delusional disorders and schizophreniform disorder). Finally, there are children with multiple developmental lags, including disturbances in affect modulation, social relatedness, and thinking, whose symptoms do not fit well within the current criteria for schizophrenia.

Organic Disorders It is important that all children and adolescents with a psychosis receive a thorough pediatric and neurological evaluation. The possibility of an organic psychosis needs to be considered when obtaining the history, completing the physical examination, and in selecting initial laboratory investigations. The list of potential organic etiological agents is exhaustive; however, conditions that must be considered include (1) delirium; (2) seizure disorders; (3) central nervous system lesions (e.g., brain tumors, congenital malformations, head trauma); (4) neurodegenerative disorders (e.g., Huntington's chorea, lipid storage disorders); (5) metabolic disorders (e.g., endocrinopathies, Wilson's disease); (6) toxic encephalopathies (e.g., substances of abuse such as amphetamines, cocaine, hallucinogens, phencyclidine, and solvents; medications such as stimulants, corticosteroids, or anticholinergic agents; and other toxins such as heavy metals); and (7) infectious diseases (e.g., encephalitis, meningitis, and human immunodeficiency virus [HIV]-related syndromes). Also, it is important to examine for comorbid medical or developmental disorders. For example, there is an increased risk for psychotic disorders associated with the genetic condition velocardiofacial syndrome.

Given the significant rates of comorbid substance abuse with schizophrenia and psychotic mood disorders in adolescents (as high as 50 percent comorbidity in some studies), it is not uncommon for a history of drug abuse to be obtained at the first onset of the psychotic disorder. If the psychotic symptoms persist for longer than a

few days despite documented detoxification from abused substances, a primary psychotic disorder should be diagnosed, with substance abuse as a comorbid condition and exacerbating agent (rather than a primary etiological agent).

COURSE AND PROGNOSIS

There are few studies examining the course of early-onset schizophrenia. Most are retrospective and none factor out the influence of treatment. The adult literature suggests that schizophrenia is a phasic disorder, although there is a great deal of individual variability. Diagnostic and therapeutic decisions depend on the recognition of these stages. The phases include:

Prodrome Prior to the overt development of psychotic symptoms, most patients will experience some degree of functional deterioration. This may include social withdrawal, odd or schizotypal preoccupations, deteriorating academic performance, worsening hygiene and self-care skills, dysphoria, and idiosyncratic or bizarre behaviors. Some youth will present with an increase in aggressive behaviors or other conduct problems, including substance abuse. These symptoms may confuse the diagnostic picture.

The prodromal phase varies greatly in time, from an acute change (days to weeks) to chronic impairment (months to years). The symptoms may represent a marked change from baseline functioning, or alternatively, a worsening of premorbid personality or behavioral characteristics. Many youth, especially those with schizophrenia with very early onset, have an insidious onset. In such cases it may be difficult to distinguish between the premorbid personality or cognitive abnormalities and the onset of the disorder.

Acute Phase During the acute phase positive symptoms predominate (i.e., hallucinations, delusions, thought disorder and disorganized behavior), and are associated with a significant deterioration in functioning. This phase generally lasts 1 to 6 months, although it may persist for over a year. Symptoms may shift from positive to negative over the course of treatment, and the length of this phase is in part determined by treatment response.

Recovery Phase As the acute psychosis remits, there is generally a period lasting several months in which the patient continues to experience a significant degree of impairment. This is most often due to negative symptoms (flat affect, anergia, social withdrawal), although it is common for some positive symptoms to persist. In addition, some patients will develop a postpsychotic depressive disorder of schizophrenia, characterized by dysphoria and flat affect.

Residual Phase As they recover, patients may have prolonged periods (several months or more) without active positive symptoms. However, most patients will continue to experience some degree of impairment due to negative symptoms.

Chronically Ill Patients Some patients will remain chronically symptomatic, despite adequate treatment, over a period of many years. These patients are among the most severely impaired and require the most intensive treatment resources. The advent of the serotonin-dopamine antagonists (atypical antipsychotic) agents offers some promise for these individuals, since clozapine (Clozaril) has been effective for treatment-refractory schizophrenia.

Longitudinal Course Schizophrenia generally follows a pattern characterized by increasing deterioration after each cycle until, after approximately 10 years, the disorder tends to wane, leaving a residual state of varying disability typified predominantly by negative symptoms. There are insufficient data to determine whether this long-term pattern holds for early-onset schizophrenia. Some youth with schizophrenia may only have one cycle, although this is not common. Recovery is incomplete in approximately 80 percent of cases where patients have had more than one episode.

Long-Term Outcome There are very few studies examining long-term outcome in early-onset schizophrenia; most studies are retrospective. In addition, there are potential problems in interpreting the existing literature. Outcome studies may be influenced by diagnostic errors, given that historically bipolar I disorder was frequently misdiagnosed as schizophrenia. Also, because schizophrenia is a phasic disorder, ratings of outcome will vary depending on the phase during which the assessment occurs.

Premorbid characteristics, treatment response, and adequacy of therapeutic resources invariably predict short-term outcome. Of 113 adolescents and young adults (mean age 18.3 ± 3 years) with schizophrenia, only 14 percent were found to have a complete remission of symptoms during the index hospitalization. The same study also reported on a cohort of 64 individuals with schizophrenia (ages 12 to 22 years) who entered a special rehabilitation program. This sample was followed at 3-month intervals over a 1-year period. Significant improvement was noted in symptomatology and cognitive functioning; subjects with higher levels of cognitive abilities and premorbid functioning did better at outcome.

There are some differences found between studies that have examined outcome over approximately a 5-year period. Two separate retrospective studies found that the majority of subjects displayed moderate to severe impairment at outcome. Eighty to 90 percent had two or more episodes during the follow-up period whereas only a few had a complete remission; outcome was best predicted by premorbid and intellectual functioning.

Eighteen children with very early-onset schizophrenia were followed over 2- to 7-year periods; 44 percent showed minimal improvement or a deteriorating course. After 1 year, only two children displayed no evidence of symptoms. However, between 2 and 7 years, one-third of the sample no longer met criteria for either schizophrenia or schizoaffective disorder; four children had no psychiatric disorders at outcome. The children in this study received extensive pharmacological and psychosocial treatment, which may have influenced the findings.

There are two sets of studies that have examined long-term outcome. One study followed up on 40 subjects with early-onset schizophrenia (mean follow-up period 14.8 years, mean age of onset 14.0 years); only two subjects had a complete recovery. Most (74 percent) were moderately to severely impaired. Premorbid functioning and the severity of positive and negative symptoms during acute episodes best predicted outcome.

Fifty-seven patients with childhood schizophrenia (onset between 7 and 13 years of age) were followed over a mean follow-up period of 16 years. At outcome, 28 percent had schizoaffective disorder. Overall, 50 percent of the sample were significantly impaired at outcome, 30 percent had good to satisfying social adaptation, and 20 percent had a complete remission. Those with schizoaffective disorder had a more favorable course. They also tended to have less premorbid difficulties, which was also a significant prognostic factor. Onset before age 10 (N = 11) was uniformly associated with a poor outcome.

Within this same sample, 44 subjects were reassessed after a mean follow-up period of 42 years. The outcome ratings were similar: 25 percent had complete remission; 25 percent partial remission, and 50 percent had chronic impairment. An insidious onset (over more than a 4-week period) and an age of onset before age 12 years were both associated with greater disability at outcome.

In general, the available follow-up studies are consistent with the adult literature. There are few studies that have compared early-onset schizophrenia versus adult onset. Onset before 15 years of age was found to be associated with higher ratings of negative symptoms in adulthood, while another study reported greater social impairment in patients with onset before age 21 years. These two studies support the clinical observations that the disorder may have a more insidious and chronic course, with less favorable outcome.

TREATMENT

There is little research addressing the treatment of early-onset schizophrenia. Therefore, treatment recommendations must be extrapolated from the adult literature. In general, therapeutic strategies should be tailored to the developmental characteristics of the patient, the needs of the family, and the different stages of the disorder. Comprehensive multimodal interventions are needed that incorporate psychopharmacological as well as psychosocial therapies. An array of therapeutic resources are needed, including inpatient/day patient psychiatric units; medication management; psychoeducational services; intensive case management; family interventions; vocational and rehabilitative assistance; special education programs, and in some cases, residential programs.

Psychopharmacology The efficacy of antipsychotic agents for the treatment for schizophrenia has been firmly established. However, there are few studies examining their use in youth. In children with schizophrenia, 0.02 to 0.12 mg/kg of haloperidol (Haldol) was found to be superior to placebo in reducing symptoms of thought disorder, hallucinations, and persecutory ideation. Loxapine (Loxitane) was superior to placebo in a study of adolescents with schizophrenia. Youth may experience the same adverse effects noted in adults, for example, extrapyramidal symptoms, sedation, tardive dyskinesia, and neuroleptic malignant syndrome. The long-term

use of antipsychotic drugs in this disorder has not been studied.

The most significant recent advance in the pharmacotherapy of schizophrenia is the advent of the serotonin-dopamine antagonists, including clozapine, risperidone (Risperdal), and olanzapine (Zyprexa). These agents are considered atypical since they are not primary type dopamine 2 (D_2) antagonists like the traditional antipsychotic medications (i.e., dopamine receptor antagonists), but instead are serotonin receptor antagonists, with some D_1 activity. Clozapine is more effective than traditional agents in reducing both positive and negative symptoms; however, its use is limited by the potential for neutropenia or seizures. Although the other atypical agents need to be further studied, they all have a lower risk for extrapyramidal symptoms and may be more effective for negative symptoms. These agents are likely to replace the traditional drugs as the primary medication for treating schizophrenia.

There are few studies examining the use of atypical agents in youth. Clozapine has been shown to be effective for treating schizophrenia with childhood onset. In the ongoing NIMH study, clozapine (mean dose 176 mg ± 149 mg a day) was superior to haloperidol (16 ± 8 mg a day) for treating both positive and negative symptoms in 21 youth (mean age 14.0 ± 2.3 years) with schizophrenia. However, while on clozapine, five youth developed significant neutropenia (this resolved spontaneously in three of subjects), and two had seizures. Therefore, although potentially more efficacious, clozapine's apparent increased risk for adverse reactions in youth raises concerns. The other atypical agents have not been systematically studied with this age group and further research is needed.

Other medications with some reported efficacy for schizophrenia in adults include lithium (Eskalith), benzodiazepines, and anticonvulsants. However, the evidence supporting the antipsychotic activity of these agents is limited, and their use in youth with schizophrenia has not been studied.

Treatment varies dependent on the phase of the illness and the patient's history of medication response and adverse effects. Except clozapine, antipsychotic drugs can be considered first-line agents; clozapine is recommended for treatment-resistant cases.

Baseline Assessment Prior to initiating a medication trial, the targeted psychotic symptoms should be adequately documented. The patient and family should be provided with adequate information regarding diagnosis, medication options, and adverse effects.

A thorough medical assessment is also needed. Evidence of neurological dysfunction warrants further evaluation, including consideration of an EEG, neuroimaging studies, and a neurology consultation. Any preexisting abnormal movements should be documented to avoid later mislabeling them as medication adverse effects. Routine laboratory screening tests to be considered include blood counts, serum chemistries, thyroid functions, and urinalysis and toxicology screenings. Testing for HIV should be done if the risk factors are present. Chromosomal analysis may be indicated for patients with clinical presentations or features suggestive of a developmental syndrome.

Acute Phase Antipsychotic drug therapy should be implemented for a period of no less than 4 to 6 weeks, using adequate dosages, before efficacy of the medication choice is determined. The immediate effect of the medication is generally sedative, with the antipsychotic effects becoming more apparent only after the first week or two. Large dosages during the early part of treatment generally do not hasten recovery, and more often result in unnecessarily excessive adverse effects. If no results are apparent after 4 to 6 weeks or if adverse effects are not manageable, a trial of a different medication should then be undertaken.

Recuperative Phase This generally occurs after 4 to 12 weeks provided the acute phase can be controlled. As positive symptoms improve, the patient may have persistent confusion, disorganization, and dysphoria. During this period antipsychotic medication should be maintained. Attempts to gradually lower the dosage may be indicated to decrease adverse effects, including exacerbation of negative symptoms. This is especially true if high dosage was needed to control the acute psychotic phase. However, any lowering of the dosage must be carefully monitored to avoid relapse.

Residual Phase In this phase, antipsychotic therapy has well-documented efficacy in preventing relapse. Approximately 65 percent of adult patients on placebo will have a relapse within 1 year of their acute psychotic phase, compared to 30 percent of patients on psychiatric drugs. However, in newly diagnosed patients who have been stabilized for at least 6 months, a medication-free trial may be advisable because a small percentage of patients will not relapse and therefore should not be exposed to the risk of the long-term adverse effects of drugs. In patients with relapses or chronic illness, or in newly diagnosed patients with persistent psychotic symptoms, the maintenance medication should be maintained indefinitely at the lowest effective dose. Longitudinal monitoring is needed, with gradual lowering of the dosage as tolerated.

Nonresponders to Antipsychotics Many patients with schizophrenia do not respond adequately to traditional antipsychotic drugs. The atypical agents may be more effective for treatment-resistant cases, although at this time clozapine is the only antipsychotic drug with clearly documented superiority in efficacy. Given clozapine's potential adverse effects, it is only used in cases where patients have failed to respond (or have had significant adverse effects) to at least three adequate trials of different antipsychotic agents.

Psychosocial Therapies In the adult literature most traditional psychotherapies have not been effective for treating schizophrenia. However, psychoeducational interventions directed at family functioning, problem-solving and communication skills, and relapse prevention have been shown to decrease relapse rates.

The family interventions stem in part from research regarding expressed emotion. Expressed emotion refers to attributes of overprotectiveness or criticism expressed towards the patient. The relapse rates for patients with schizophrenia are higher when living in families characterized as having high expressed emotion. Family intervention programs, in conjunction with medication therapy, have been shown to significantly decrease schizophrenia relapse rates.

Another important psychoeducational modality is social skills training. These programs focus on improving the patient's strategies for dealing with conflict and avoidance; identifying the correct meaning, content, and context of verbal messages within their families; and enhancing their socialization and vocational skills. The combination of family treatment, social skills training, and medication therapy also has been shown to decrease relapse rates.

The effectiveness of a psychoeducational treatment program was compared to the standard community treatment in a sample of adolescents with schizophrenia (12 subjects per group). The psychoeducational treatment program included parent seminars, problem-solving sessions, milieu therapy (while the subjects were hospitalized), and networks (reintegrating the subjects back into their schools and communities). These interventions were used in conjunction with medication therapy. The standard treatment group received a mixture of individual psychotherapy, milieu therapy, and medications; outcome was assessed after 2 years. The psychoeducational treatment program had lower rates of rehospitalization and was more cost effective. Subjects with poor premorbid psychosocial functioning benefited the most from the psychoeducational interventions. Clinical improvement was associated with the families' expressed emotion ratings changing from high to low. Although more research is needed in this area, family treatment and social skills training should be considered helpful adjuncts to medication treatment for children and adolescents with schizophrenia.

Further research is needed to delineate the course, phenomenology, response to treatment, and outcome of schizophrenia with early onset. Because schizophrenia in youth appears to be the same disorder as in adults, studying the progression of its neuropsychiatric manifestations against the backdrop of normal developmental maturation may potentially provide clues to underlying etiological mechanisms. Moreover, because schizophrenia in youth may be a more homogeneous form of the illness, research examining this population may be more likely to identify potential biological and genetic markers.

SUGGESTED CROSS-REFERENCES

Chapter 12 reviews the adult literature on schizophrenia; antipsychotic medications are discussed in more detail in Section 12.8, Section 31.17, and Section 31.21. Medical conditions that may mimic psychotic disorders are reviewed in Chapter 10. Other pertinent sections include the psychiatric treatment of children (Chapter 48), mood disorders (Chapter 14), and pervasive developmental disorders (Chapter 38).

CHAPTER REFERENCES

Alaghband-Rad J, Hamburger SD, Giedd JN, Frazier JA, Rapoport JL: Childhood-onset schizophrenia: Biological markers in relation to clinical characteristics. Am J Psychiatry *154:*64, 1997.

Alaghband-Rad J, McKenna K, Gordon CT, Albus KE, Hamburger SD, Rumsey JM, Frazier JA, Lenane MC, Rapoport JL: Childhood-onset schizophrenia: The severity of premorbid course. J Am Acad Child Adolesc Psychiatry *34:*1273, 1995.

*American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of children and adolescents with schizophrenia. J Am Acad Child Adolesc Psychiatry

36(Suppl):177S, 1997.

*American Psychiatric Association: Practice guideline for the treatment of patients with schizophrenia. Am J Psychiatry *154*(Suppl):1, 1997.

Armenteros JL, Fennelly BW, Hallin A, Adams PB, Pomerantz P, Michell M, Sanchez LE, Campbell M: Schizophrenia in hospitalized adolescents: Clinical diagnosis, DSM-III-R, DSM-IV, and ICD-10 criteria. Psychopharmacol Bull *31:*383, 1995.

Asarnow JR, Tompson MC, Goldstein MJ: Childhood onset schizophrenia: A follow-up study. Schizophr Bull *20:*647, 1994.

Asarnow R, Asamen J, Granholm E, Sherman T, Watkins JM, Williams ME: Cognitive/neuropsychological studies of children with a schizophrenic disorder. Schizophr Bull *20:*647, 1994.

Caplan R, Guthrie D, Gish B, Tanguay P, David-Lando G: The Kiddie Formal Thought Disorder Scale: Clinical assessment, reliability, and validity. J Am Acad Child Adolesc Psychiatry *28:*408, 1989.

Carlson GA, Fennig S, Bromet EJ: The confusion between bipolar disorder and schizophrenia in youth: Where does it stand in the 1990s? J Am Acad Child Adolesc Psychiatry *33:*453, 1994.

Carlson GA: Child and adolescent mania: Diagnostic considerations. J Child Psychol Psychiatry *31:*331, 1990.

Clark AF, Lewis SW: Treatment of schizophrenia in childhood and adolescence. J Child Psychol Psychiatry *39:*1071, 1998.

Eggers C: Schizo-affective psychosis in childhood: A follow-up study. J Autism Dev Disord *19:*327, 1989.

Eggers C: Course and prognosis in childhood schizophrenia. J Autism Child Schizophr *8:*21, 1978.

Eggers C, Bunk D: The long-term course of early-onset schizophrenia. Schizophr Bull *23:*105, 1997.

*Geddes J: Prenatal and perinatal risk factors for early onset schizophrenia, affective psychosis, and reactive psychosis. BMJ *318:*426, 1999.

Gordon CT, Frazier JA, McKenna K, Giedd J, Zametkin A, Zahn T, Hommer D, Hong W, Kaysen D, Albus KE, Rapoport JL: Childhood-onset schizophrenia: An NIMH study in progress. Schizophr Bull *20:*697, 1994.

Hafner H, Nowotny B: Epidemiology of early-onset schizophrenia. Eur Arch Psychiatry Clin Neurosci *245:*80, 1995.

Hogarty GE, Anderson CM, Reiss DJ, Kornblith SJ, Greenweld DP, Javna CD, Madonia MJ: Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia. I. One-year effects of a controlled study on relapse and expressed emotion. Arch Gen Psychiatry, *43:*633, 1986.

*Hultman CM, Sparen P, Takei N, Murray RM, Cnattingius S: Prenatal and perinatal risk factors for schizophrenia, affective psychosis, and reactive psychosis of early onset: Case-control study. BMJ *318:*421, 1999.

Jacobsen LK, Giedd JN, Castellanos FX, Vaituzis AC, Hamburger SD, Kumra S, Lenane MC, Rapoport JL: Progressive reduction of temporal lobe structures in childhood-onset schizophrenia. Am J Psychiatry *155:*678, 1998.

Kolvin I: Studies in the childhood psychoses. Br J Psychiatry *6:*209, 1971.

Kumra S, Jacobsen LK, Lenane M, Zahn TP, Wiggs E, Alaghband-Rad J, Castellanos FX, Frazier JA, McKenna K, Gordon CT, Smith A, Hamburger S, Rapoport JL: Multidimensionally impaired disorder: Is it a variant of very early-onset schizophrenia? J Am Child Adolesc Psychiatry *37:*91, 1998.

Kumra S, Frazier JA, Jacobsen LK, McKenna K, Gordon CT, Lenane MC, Hamburger SD, Smith AK, Albus KE, Alaghband Rad J, Rapoport JL: Childhood-onset schizophrenia. A double-blind clozapine-haloperidol comparison. Archives Gen Psychiatry *53:*1090, 1996.

Leff J, Vaughn C: *Expressed Emotion in Families: Its Significance for Mental Illness*. Gilford, New York, 1985.

Maziade M, Bouchard S, Gingras N, Charron L, Cardinal A, Roy M, Gauthier B, Tremblay G, Cote S, Fournier C, Boutin P, Hamel M, Merette C, Martinez M: Long-term stability of diagnosis and symptom dimensions in a systematic sample of patients with onset of schizophrenia in childhood and early adolescence. II: Positive/negative distinction and childhood predictors of adult outcome. Br J Psychiatry *169:*371, 1996.

Maziade M, Gingras N, Rodrigue C, Bouchard S, Cardinal A, Gauthier B, Tremblay G, Cote S, Fournier C, Boutin P, Hamel M, Roy M, Martinez M, Merette C: Long-term stability of diagnosis and symptom dimensions in a systematic sample of patients with onset of schizophrenia in childhood and early adolescence. I: Nosology, sex and age of onset. Br J Psychiatry *169:*361, 1996.

McClellan J, McCurry C: Neurodevelopmental pathways in schizophrenia. Semin Clin Neuropsychiatry *3:*320, 1998.

McClellan JM, Werry JS, Ham M: A follow-up study of early onset psychosis: Comparison between outcome diagnoses of schizophrenia, mood disorders and personality disorders. J Autism Dev Disord *23:*243, 1993.

McKenna K, Gordon CT, Lenane M, Kaysen D, Fahey K, Rapoport J: Looking for childhood-onset schizophrenia: The first 71 cases screened. J American Acad Child Adolesc Psychiatry *33:*636, 1994.

Meltzer HY, Lee MA, Ranjan R: Recent advances in the pharmacotherapy of schizophrenia. Acta Psychiatr Scand *90*(Suppl):95, 1994.

Pool D, Bloom W, Mielke DH, Roniger JJ Jr, Galant DM: A controlled evaluation of loxitane in seventy-five adolescent schizophrenia patients. Curr Ther Res Clin Exp *19:*99, 1976.

Rapoport JL, Giedd J, Kumra S, Jacobsen L, Smith A, Lee P, Nelson J, Hamburger S: Childhood-onset schizophrenia: Progressive ventricular change during adolescence. Arch Gen Psychiatry *54:*897, 1997.

Remschmidt H, Martin M, Schulz E, Gutenbrunner C, Fleischhaker C: The concept of positive and negative schizophrenia in child and adolescent psychiatry. In *Positive Versus Negative Schizophrenia*, Marneros A, Andreasen NC, Tsuang MT, editors. Springer-Verlag, Berlin, 1991.

Remschmidt HE, Schulz E, Martin M, Warnke A, Trott G-E: Childhood-onset schizophrenia: History of the concept and recent studies. Schizophr Bull *20:*727, 1994.

Rund BR, Moe L, Sollien T, Fjell A, Borchgrevink T, Hallert M, Naess PO: The psychosis project: Outcome and cost-effectiveness of a psychoeducational treatment programme for schizophrenic adolescents. Acta Psychiatr Scand *89:*211, 1994.

Russell AT: The clinical presentation of childhood-onset schizophrenia. Schizophr Bull *20:*631, 1994.

Spencer EK, Kafantaris V, Padron-Gayol MV, Rosenberg C, Campbell M: Haloperidol in schizophrenic children: Early findings from a study in progress. Schizophr Bull *28:*183, 1992.

Thomas MA, Ke Y, Levitt J, Caplan R, Curran J, Asarnow R, McCracken J: Preliminary study of frontal lobe 1H MR spectroscopy in childhood-onset schizophrenia. J Magn Reson Imaging *8:*841, 1998.

Thomsen PH: Schizophrenia with childhood and adolescent onset—a nationwide register based study. Acta Psychiatr Scand *94:*187, 1996.

*Tsuang MT, Faraone SV: The case for heterogeneity in the etiology of schizophrenia. Schizophr Res *17:*161, 1995.

Verdoux H, Geddes JR, Takei N, Lawrie SM, Bovet P, Eagles JM, Heun R, McCreadie RG, McNeil TF, O'Callaghan E, Stober G, Willinger MU, Wright P, Murray RM: Obstetric complications and age at onset in schizophrenia: An international collaborative meta-analysis of individual patient data. Am J Psychiatry *154:*1220, 1997.

*Werry JS: Child and adolescent (early onset) schizophrenia: A review in light of DSM-III-R. J Autism Dev Disord *22:*601, 1992.

Werry JS, McClellan J: Predicting outcome in child and adolescent (early-onset) schizophrenia and bipolar disorder. J Am Acad Child Adolesc Psychiatry *31:*147, 1992.

Werry JS, McClellan J, Chard L: Early-onset schizophrenia, bipolar and schizoaffective disorders: A clinical and follow-up study. J Am Acad Child Adolesc Psychiatry *30:*457, 1991.

*Werry JS, Taylor E: Schizophrenic and allied disorders. In *Child and Adolescent Psychiatry: Modern Approaches*, M Ruther, E Taylor, L Hersov, editors. Blackwell Scientific, Oxford, England, 1994.

Zahn TP, Jacobsen LK, Gordon CT, McKenna K, Frazier JA, Rapoport JL. Autonomic nervous system markers of psychopathology in childhood-onset schizophrenia. Arch Gen Psychiatry *54:*904, 1997.

CHAPTER 48. CHILD PSYCHIATRY: PSYCHIATRIC TREATMENT

48.1 INDIVIDUAL PSYCHODYNAMIC PSYCHOTHERAPY

OWEN LEWIS, M.D.

Psychodynamic psychotherapy is conceptualized and organized as an individual psychotherapy. What distinguishes psychodynamic psychotherapy from other therapies is that it seeks to treat not only the individual symptoms, disabilities, and vulnerabilities of the child, but the meanings of those symptoms, disabilities, and vulnerabilities and their role in the overall functioning of the child's personality. In its classical format, it was derived from psychoanalysis and did not usually involve other modalities of treatment. In modern applications, it may include a variety of modalities in an overall psychodynamic framework. It is used to treat a wide variety of disorders and developmental circumstances.

Central to psychoanalytic theory is the role of the unconscious fantasy life. Not only are symptoms viewed in regard to the unconscious, but also patterns and quality of relationships, modulation of affects, and personality traits are related to unconscious functioning. Of concern, too, are the patterns of defense mechanisms, the presence of traumatic experiences, the level of psychosexual functioning, and the evolution of the superego and conscience. Modern adaptations also look at the presence of biologically determined disorders (such as some forms of depression or anxiety) and biologically determined cognitive patterns (such as attention-deficit disorders or learning disorders) and the meaning of these to the child. Modern adaptations also address those factors that maintain problematic behaviors of the child in the family.

The inner world of the child is understood through several pathways. Children can and do speak about themselves. Reports about the child from parents and teachers can suggest points of difficulty. Free play serves as an important expressive modality for many children, especially younger children. Older children may be given to playing games, and their conduct in such games is often revealing. Children of all ages may benefit from a variety of creative techniques such as drawing, storytelling, story writing, or playacting. The therapist's own reactions to the patient can also serve as an important source of information about the subjective world of the child.

In contrast to adult psychotherapy in which great emphasis is placed on the importance of interpretations of the unconscious, the child's experience with the therapist in the therapy can exert a powerful influence toward change. The child not only uses the therapist as a transference object, but also readily uses the therapist as a real figure offering important new identifications.

In general, children with internalizing symptoms (e.g., depression or obsessive-compulsive disorder) are more accessible to individual psychotherapy as they become able to locate the source of distress as internal and desire to find help. Children with externalizing symptoms (e.g., conduct disorder or oppositional defiant disorder), on the other hand, tend to blame others, are less able immediately to see their own role in the genesis of their difficulties, and may therefore be less accessible to individual work. This tends to be a problem in the initial phase of treatment. Current research supports the application of individual psychodynamic psychotherapy to a wide variety of disorders.

HISTORY

Child psychotherapy begins with Sigmund Freud's case of Little Hans, a 5-year-old phobic boy. Published in 1909, it was the first description of the psychotherapeutic treatment of a child. The treatment in this case was actually rendered by Han's father who reported to Freud and received guidance from him. Significant interest in the mental and emotional lives of children was generated by Freud's theories of psychosexual development that posited that symptoms in adulthood could be traced to conflicts arising at earlier stages of development. Sandor Ferenczi was one of the first to attempt to analyze children, but he quickly became frustrated with the endeavor, finding that children only wanted to play. Hermione Hug-Hellmuth published the first report of actual play therapy with children suggesting that play and drawings were valid modes of communication used by children.

However, two psychoanalysts, Anna Freud and Melanie Klein, developed the field of child psychoanalysis. While they worked in different ways, they undertook the analyses of many children and spurred the work of many others. Klein understood play to be the actual equivalent of adult free-association. She was interested in the earliest object relations; the importance of unconscious fantasies at each stage of development; the role of primitive defenses such as projection, projective identification, and omnipotent control; the process of early identifications; and the role of envy and guilt in these early relationships. Freud looked at play from a psychoanalytic perspective and learned about the child from play but, for a number of reasons to be elaborated later, did not view play as a substitute for free-association. Her work concerned the development of the ego, the evolution of defenses, and the developmental pathway of various ego functions. In one example of a developmental pathway, she described a continuity from the child's capacity to play to the adult's capacity to work.

There were several other major early contributors to field. Donald Winnicott, a psychoanalyst and pediatrician, emphasized the importance of the mother-infant relationship. His understanding of the transitional object, for instance, as playing a role in the child's ability to separate from the maternal figure infused a new appreciation of the meaning of commonly observed childhood behaviors. Margaret Mahler observed mother-toddler interactions in a more systematized way and described the evolution of early object relations from the perspective of separation and individuation. Winnicott's and Mahler's ideas had great impact on the issues dealt with in children's psychotherapy. August Aichorn first extended the work to delinquent adolescents. Finally, the work of Jean Piaget focused on children's cognitive development and this subsequently influenced the practice of psychotherapy.

Another early influence on the development of child psychiatry and child psychotherapy came in the child guidance clinic movement. William Healey settled in Chicago in 1909 and turned his attention to the care of delinquent, abandoned, and abused children appearing before the Cook County Court. The Juvenile Psychopathic Institute (later becoming the Institute for Juvenile Research) was founded to evaluate and treat these children. Healey stressed the importance of psychologists for testing and social workers for making home visits. By 1917 a number of other child guidance clinics had been established, Healey went to the Judge Baker Guidance Clinic in Boston. Then and through the 1920s, the work with children in these settings was largely environmental. These clinics established the practice of intervention with troubled children. The influx of European psychoanalysts in the 1930s turned this work in a psychodynamic and psychotherapeutic direction.

CLASSICAL PSYCHODYNAMIC PSYCHOTHERAPY

Classical psychodynamic psychotherapy evolved from an approach to child psychoanalysis rooted in drive theory, in which symptoms are viewed as a compromise formation of the various agencies of the mind (id, ego, superego). The approach relies on interpretation of defenses and conflicts and strives to make these unconscious processes conscious. The principle of multiple causation holds that any given behavior, emotion, or communication has multiple meanings. The approach should not be confused with the aims of psychotherapy. Hansi Kennedy has noted that the most basic aim of any child's psychotherapy is to restore that child to a path of normal development; this end-point must be kept in mind.

Freud noted that what distinguishes child psychoanalysis from other forms of therapy is that it relies chiefly on interpretation of resistance and transference and attempts to avoid coercion, the use of authority, suggestion, abreaction, and manipulation. This also holds true for classical psychodynamically oriented psychotherapy. She notes several features that distinguish child psychoanalysis from adult psychoanalysis. First, children, unlike adults, do not as a rule voluntarily bring themselves to therapy; rather, they are brought. Thus, forming a task-oriented working alliance is more difficult. Resentment at having been brought to a psychiatrist may also interfere with the work. In her earlier writing she proposed using many nonanalytic means to induce children into treatment, such as feeding them or promising actual help with difficult parents and teachers. She later reversed this position and advocated instead an approach of interpretation from the

beginning. Such approaches may nonetheless be useful with the resistant child at the start of treatment. On the other hand, the curative tendencies may be more easily tapped since the child's urge to complete development is immeasurably stronger than the adult's.

Many significant differences in technique exist; chief among them is the relative absence of free association in the child. The child tends to act instead of speak and possesses an innate resistance to regression. In addition, children preferentially express aggressive drives over libidinal drives. Play, drawing, and the staging of fantasy games, according to Anna Freud, cannot be accepted in place of free association because they introduce the uncertainty of symbolic interpretation. They serve only as an arena for psychoanalytic observation.

While interpretation, principally of resistances and transferences, remains the primary vehicle of the therapy, verbalization plays a specific role in the process leading to interpretation. First emphasized by Annie Katan, verbalization is an indispensable prerequisite for secondary process thinking, and it promotes reality testing and ego control over id impulses. *Verbalization* refers to the therapist's active narration of the child's play, behavior, moods, and affects. While verbalization is important in the treatment of children of all ages, it is particularly important in the treatment of early school-age and preschool children and children who have more-extreme developmental delays and defects.

While the curative tendencies can be seen as stronger in children, so too can the resistances. The various sources of resistance in children, according to Freud, include the following:

1. Children do not enter treatment on their own and are therefore not bound by the analytic rules
2. Children cannot take a long-term view of treatment
3. Acting-out takes precedence over talking
4. Children's egos are relatively weak and defenses are kept up more rigidly
5. Primitive defenses remain present, even as more mature ones evolve
6. The urge to grow up and reject the past is stronger, so the infantile (conflictual) past is often not accessible
7. Children prefer environmental solutions to internal ones.

These various resistances must be dealt with in the therapy. Because children cannot be counted on to maintain an independent alliance with the therapist, throughout treatment the alliance with the child's parents must be maintained and fortified.

Psychoanalytically oriented treatment has traditionally been divided into a beginning phase, a middle phase, and a termination phase. The beginning phase depends on establishing a strong working alliance with the parents. After this, an initial session must indicate the nature of the work without overwhelming the child with a sense of having problems or with anxiety. Finding a comfortable mode of communication, whether in play, games or being able to talk about one's current life, ensures that conflictual communication can occur in a predominantly nonthreatening setting.

The middle phase involves an intensification of transference reactions as well as play sequences. Here, basic conflicts are worked through via interpretation of both transference and the play.

PLAY

Play has had a central place in the psychotherapy of children from its inception. Play is a natural means of child communication. Play serves many functions in a child's development including socialization, motor development, development of imagination, integration of emotions, and the development of autonomy. Anna Freud described a line of development from a child's capacity to play to the adult's capacity to work. While the use of play in child psychotherapy may draw on these functions, there are special considerations regarding play and psychotherapy. Whether play is used ultimately for the purposes of therapeutic interpretation or whether play itself has a primary therapeutic benefit in the context of ongoing psychotherapy is controversial.

Play in the context of psychotherapy takes on special characteristics. While children's play usually occurs either in a social context (which after the age of 4 usually involves a high degree of reciprocity) or in a solitary setting (in which all the play is elaborated from the child), the context of therapeutic play is neither social nor solitary. While the therapist may choose to be more or less involved in the play, the therapist will attempt to encourage the elaboration of play by children. The therapist can attempt to redirect the play towards areas of conflict and can help children overcome resistance. Because the therapist will show interest in all areas of the play and will not express judgments about what is good or bad, permissible or prohibited, children come to understand that playing with their therapist is different from playing with anyone else. Children are thus encouraged to express more and more material that might otherwise be repressed.

What is also unique about the play therapy situation is the balance between the control children have over the play and the safety provided by the presence of the adult. In a play enactment, for instance, children may assign parts to themselves and other parts to the therapist. While the therapist will make some comments spontaneously to facilitate the play, the therapist will also occasionally defer to the children, asking what a given character should say. The children retain ultimate control over the script and may terminate a scene at will. Yet, the presence of the therapist confers a certain safety. Children who would not dare play out a certain theme alone not only feel encouraged, but also may feel that the presence of the therapist creates what Donald Winnicott has termed the *holding environment*, in which security, if it existed in the child's development in the first place, is ensured.

In this way, children come to communicate more and more in therapeutic play. Facilitation of communication, as described above, depends on the constellation of internal object relations. Security can only develop if there is an internal source of security to draw on. The same consideration applies to all affects and emotions. The seductiveness, punitiveness, or negligence of the primary caretakers comes to bear in the evolution of the therapy.

> A 3½ year-old boy was brought to treatment because he regularly objected to visiting with his father who had left the family 6 months previously. His father practiced a strict fundamentalist religion and required practices and prayers of the boy that he little understood. His father felt that the mother was turning the boy against him. In the first three sessions, he layered the roof of a small doll house with strips of black tape, as if he were reshingling the roof. On the fourth session, he had a father figure walk along the pitch of the roof, stumble, and fall. With each fall, he clapped his hands with glee. At first, he looked to the therapist for permission to go on with this play. When this was given, he repeated the sequence over and over again with delight. Some weeks later, at a session after another child had removed some of the pieces of tape from the roof, the boy looked up with great alarm, saying, “Who's been playing in my house?”
>
> This case illustrates the way a child will express thoughts and feelings not permitted in other contexts, the importance of the permission and facilitation of the therapist, and finally, the sense of privacy and security of the play situation. After the initial anxiety, the child became anxious again only when he perceived that his play space had been entered.

Other Considerations Several other aspects of play in a therapeutic context must be considered. These include play as defense, traumatic play, the use of games, and what Eva Sperling has called “play as a medium of exchange.”

Play can often be used for defensive purposes. In such cases, play becomes repetitious, rote, and unadventurous. Avoidance of interpersonal contact or revelation of personally threatening matters is thus achieved. It is not uncommon for a therapy to go through periods of such defensiveness. At these times, neither therapist nor patient may realize that the patient may be testing the therapist in various ways. Is the therapist really interested? Is the therapist really patient, accepting, or nonjudgmental? Eventually, the therapist must challenge the repetitious nature of the play or undertake maneuvers to introduce new activities.

Repetitious play must be distinguished from posttraumatic play. In the latter, a traumatic situation is repeated with no particular elaboration or use of the imagination. Often the play situation is devoid of expressed emotion. What distinguishes this type of play from defensive play is the nature of the traumatic situation portrayed, which will have been unusually and severely stressful. Examples of such trauma include natural disasters, witnessing or participating in accidents or crimes, or kidnappings. In these cases, the technique differs from that used in defensive play. One aim of the therapy here is to connect the displaced emotion to the traumatic situation.

Latency-age children tend to use an admixture of games and free play. Their interest in games reflects the Piagetian stage of concrete operations in which children are becoming capable of understanding the rules that govern external reality. In this regard, their preoccupation with rule-bound games reflects an exercise of the new

capacity of the developmental stage. Games, however, often flow into play. As John Meeks has noted, when children obviously cheat at a game, they are often at the point of moving from the realm of games into the world of play. If one understands play as a situation in which the unconscious fantasy life dominates reality in the play sequence, the same situation occurs when cheating overtakes the rules of the game. The children are motivated by what they want reality to be, not what they currently perceive it to be. The therapist must note that a new set of rules is being invoked to help the children perceive their cheating and support the development of superego functions, but noting this change in rules is not judgmental.

A 10-year-old boy was brought to treatment because of his disillusionment with his learning disabilities. These had been appropriately diagnosed, and he was receiving adequate remedial work. His emotional reactions to these difficulties had been overlooked for some time. Early in treatment, in the midst of a game of checkers, and somewhat annoyed that the therapist had achieved two kings, he moved his king back across the board, added a third checker on top of the two, and declared that this was an emperor, more powerful than a mere king, and able to move diagonally. Soon his emperor retraversed the board and became an arch emperor with a fourth checker on top. The archemperor had all the emperor's powers plus the ability to enter red squares. So the game went. With pieces soon piled six and seven high, his need for strength and his sense of powerlessness were revealed.

Finally, play in therapy is often used as a medium of exchange between patient and therapist. Here, the play is neither being used in the process of self-disclosure nor is it specifically defensive. Rather, because children may not be able to sustain a conversation for the full session, the play activity serves to structure the interaction between the therapist and the patient. Conversation then need not carry the full weight of the interaction and can start and stop and then proceed again as motivated by a true desire for communication. Drawing, constructing, or games may serve this function.

TRANSFERENCE

Most psychodynamically oriented psychotherapists agree that the emergence of transference is considered central to the process of psychotherapy. Through the interpretation of the emerging transference, the conflicts that are central to both the particular presenting symptoms and the particular organization of the personality can be resolved. Transference has been regarded as a window on psychic organization; through it, past relationships become visible particularly as they color and shadow the current, actual relationship with the therapist.

The application of the concept of transference to the treatment of children needs modification. As first described by Freud, children use the therapist differently from adults. Children's hunger for new experiences is often stronger than the repetition compulsion. Thus, therapists are experienced as new persons in their own right, and the children are drawn to what is new in the relationship with them. This does not preclude repetition of elements of the past, but these exist alongside the new experiences. Child therapists, however, must distinguish between the two. In addition, the fact that a child will approach the therapist as a new object provides other opportunities for the therapist via new identifications.

Other factors are associated with transference in children. Most importantly, most children are living with their parents. They have neither physically nor psychologically separated from them. This, of course, is the psychological task accomplished by the end of adolescence. Because of this retained connection, many other modes of interaction occur that look like transference but must be differentiated from it. Robert Tyson and Dayllis Tyson have distinguished these reactions: habitual modes of relating, transference predominantly of current relationships, transference predominantly of past experiences, and transference neurosis. *Habitual modes of relating* are characteristic ways of relating to others that may occur in the therapy but are not specific to the therapy. *Transference of current relationships* refers to “spill-overs” of interactions currently occurring with the parents. Children and adolescents often see the therapist as an extension of their parents, so they are prone to these types of displacements with the therapist. *Transference of past experiences* refers to a type of limited transference reaction that occurs after some period of treatment. Children are usually aware of some qualitative change in their feelings about, or attitudes toward, the therapist. A *transference neurosis* involves both a qualitative and quantitative shift so that much (if not all) of the patient's interactions with the therapist are dominated by conflicts from the past, however strongly felt in the present. When a transference neurosis is established, the neurotic elements in the children's lives become specifically focused on the therapist, thus freeing the children from the influence of these conflicts in other areas of their life. Not all intensive child psychotherapies, however, progress to the development of a transference neurosis; in fact, some question whether a true transference neurosis can emerge at all with children. A transference neurosis is more likely to occur when there have been discontinuities in the child's essential object relationships. At such times, the children may see the therapist as the basic cause of their difficulties and desire to flee the therapy. Understanding the various sources of the vacillations in children's treatment of their therapist is essential for effective interpretation.

An 8-year-old boy was brought to treatment because of behavioral difficulties in school, infantile behaviors at home, and a recent onset of stealing or, as he called it, “borrowing” toys from his friends. He had been adopted at birth. At the age of 3 his mother contracted cancer and was slowly dying, principally at home, over the next 2 years. The year after her death his father fell into a depression and remained home much of the time with the boy. The boy was indulged throughout the mother's demise and during the year of mourning. A kindly housekeeper saw to all his needs and continued to feed and dress him. At the age of 6, the father married a woman with no children who made it her business to whip the boy into shape. She installed herself as mother and first informed the boy about his original adoption indicating that his mother had not really been his mother. She replaced his bedside picture of his mother with one of herself.

In therapy, a particular play sequence emerged over several weeks and was repeated regularly for almost 6 months. The boy went into an imaginary time machine and crawled out as a baby. The therapist as “Daddy” would make him cookies and bottles out of paper and feed him. Sometimes he did not like swallowing the cookies and spit them out. After this, the therapist had to read him three juvenile stories. Then both would have to feign sleep with the lights dimmed. Soon it was morning, but neither the boy nor the therapist were supposed to feel like getting out of bed and would lounge, either pretending to watch television or reading the stories again. Once this sequence was established, the child's stealing, infantile behaviors at home, and disruptive behaviors at school subsided. Eventually, the child replaced this sequence with a new enacted story that concerned a little puppy who needed training.

These play sequences fit the description of a transference neurosis. The enactments were both intense and comprehensive, with a behavioral improvement outside of the therapy. Clearly the boy needed to revisit the lost, merged relationship he had experienced with his father after his mother's death, which had been abruptly taken away from him. When he was ready to be “trained,” he indicated so.

INTERPRETATION

Interpretation historically referred to a set of communications from analyst to patient to bring about dynamic change in the structure of the personality. Specifically, interpretations were aimed at making the unconscious conscious and were usually centered on either the defenses or an explanation of the transference. The variability of children's development has required modifying the use of these interventions with children.

Melvin Lewis has proposed that interpretation with children must occur over a range of six types of interventions that proceed from less to more comprehensive statements. *Setting statements* are aimed at creating the conditions necessary for therapeutic work; for example, telling children that therapy is unique because it lets them understand their feelings and actions. *Attention statements* focus children on the facts of their actions or statements. This involves direct labeling; for example, a certain character is speaking nicely but doing mean things. *Reductive statements* attempt to identify themes that have occurred in a variety of settings. If children have been reacting with sadness to the departures of friends in their actual life and characters in play, this might be pointed out. This requires that the children's cognitive capacities allow them to generalize across situations. *Situational statements* follow from reductive statements. Building on the ability to recognize an affect or reaction arising in different settings, the children's role in the generation of the affect or reaction can be highlighted. *Transference interpretations* are considered to be the essential interpretation. As noted in the section on transference, transference may not exist as discretely in children as it does in adults. Nonetheless, to the extent that transference is manifested, whether completely as a transference neurosis or incompletely, statements that elucidate the nature of an earlier object relationship are considered to be at the heart of the conflict. *Etiological statements* attempt to connect earlier developmental events with current behaviors and feelings through recall and reconstruction. While the children's cognitive limitations may make this aspect of the work difficult, at termination children can often recall earlier phases of the treatment. This makes some aspects of reconstruction possible.

Paulina Kernberg has described the hierarchy of verbal interventions differently. She discusses various levels of statements moving from surface consciousness to the unconscious, including (1) statements of ordinary social behavior, (2) statements relating to treatment, (3) questions about factual information, (4) supportive statements, (5) facilitating statements, (6) clarifications, (7) confrontations, and (8) interpretations. Psychotherapy, then, is formulated on those interventions that are both possible and desirable in a given situation. Broadly speaking, psychotherapy ranges from supportive-expressive psychotherapy to expressive psychotherapy to

psychoanalysis.

In supportive-expressive psychotherapy, resolution of conflicts is not expected. Symptoms are expected to resolve but without character change. The therapeutic alliance is seen as an end in itself, with emphasis on identification with the therapist and a corrective emotional experience in the therapy. The usual types of interventions include clarifications, confrontations, and supportive interventions. Transference interpretations are generally limited to negative transference to facilitate the treatment. Other interpretations are limited to the manifest content. The therapist actively chooses topics to be discussed and gives information and advice.

In expressive psychotherapy, the goal includes resolution of conflicts from some, but not all, levels of development. Symptoms are expected to resolve but without character change. Children are given greater autonomy in the elaboration of the treatment. Interventions include some supportive statements, clarifications, and confrontations, though some transference and reconstructive interpretations are possible.

In child psychoanalysis, the goal is not only symptom resolution but also character change. Resolution of conflicts from all developmental levels is expected. The pace and unfolding of materials is left entirely to the children. A full transference neurosis is expected to develop, and full use of it is made in interpretation. Other types of verbal interventions are used only as preliminary steps en route to transference and reconstructive interpretations.

RESEARCH

Most research on the evaluation of outcome in intensive psychoanalytic treatment of children has been flawed by poor definitions of the disorders, poor specification of treatment procedures, and unreliable outcome measures. Several studies do support the efficacy of intensive psychoanalytic treatments and have led to a reconsideration of certain aspects of psychoanalytic treatment of children and its application.

Peter Fonagy and George Moran compared two groups of 11 children hospitalized for poorly controlled diabetes. One group received hospital care plus three to four sessions of psychoanalytic psychotherapy a week for an average of 15 weeks; the control group received only hospital care. The treatment group showed significant improvement in diabetic control, which was maintained for over 1 year at follow-up. The control group returned to pretreatment levels of diabetic control after 3 months. In a study with single-case design, Fonagy and Moran investigated growth rate changes in three children who exhibited growth retardation. All three children showed an acceleration in growth and several biological growth indicators.

Mary Target and Fonagy undertook a retrospective chart review of 763 patients treated at the Anna Freud Center. Patients had received either psychoanalysis four or five times a week or psychotherapy one to three times a week. The authors used standardized psychological and psychiatric descriptions of the children to allow comparison and used diagnostic change and changes in social and emotional adaptation as outcome indexes. Three broad studies were undertaken: children with emotional disorders, children with disruptive disorders, and the impact of age.

There were several notable findings. Of the 352 children with emotional disorders, only 24 percent had a diagnosable disorder after 6 months, regardless of the intensity of treatment, and 72 percent showed reliable, clinically significant improvement. In this group, both the more- and less-frequent treatment groups showed more improvement with longer treatment, but the more-frequent treatment group showed greater improvement. Intensive treatment was significantly more helpful for children with more severe disturbance in terms of multiple diagnoses or pervasive impairments.

The 135 children with disruptive disorders showed lower overall rates of improvement than the matched children with emotional disorders, and nearly one-third dropped out of treatment after 1 year. Of those who remained in treatment, 69 percent were no longer diagnosable at termination. Predictors of improvement included younger age, more-intensive treatment, longer treatment, concurrent treatment of mother, and history of foster care placement.

Children were grouped by age: below 6 years to 12 years, and above 12 years. The youngest group showed the most benefit. The 6- to 12-year-olds showed most benefit from more-intensive treatment. Adolescents did not show any benefit from increased frequency of treatment.

These studies have several important results. Younger children benefit more from treatment, which underscores the need for early intervention. Adolescents did not benefit from the more intensive treatment, which suggests that a frequent treatment should not generally be prescribed for this age group. Children with simple or single emotional disorders, the assumed ideal patient for psychoanalysis, appeared to benefit equally from intensive and nonintensive psychotherapy. The children with severe, entrenched, complex psychosocial problems, which included conduct disorder and at least one emotional disorder, only benefited from psychotherapy that was intensive.

CRITIQUE OF CLASSICAL PSYCHOTHERAPY

Classical psychotherapy, according to its own theory, relies predominantly on interpretation as the vehicle of psychic change. Interpretation is a verbal intervention. The emphasis on verbal intervention is based on the concept that verbalization places wishes, affects, and fears under the rational control of the ego. Thus, in Kemberg's schema, treatment progresses from less complete treatment (supportive-expressive) to more complete (psychoanalysis). Complete treatment involves character change. Symptom abatement, change in diagnosis, or improvement in social and adaptive functioning are presumed to follow character change. Character change is made possible through verbally derived insight into conflict.

That children can and do change through psychoanalytically oriented psychotherapy or analysis is substantiated by the studies of Fomagy and Moran and Target and Fomagy. Their demonstration that the more pervasively disturbed children in the emotion disorders group benefited from more frequent (but not less frequent) treatment and that children with disruptive disorders benefited more from intensive, longer treatments suggests that the application of intensive psychoanalytic treatment to these groups (formerly presumed not to benefit from intensive psychotherapy) should be reconsidered. The assumption that these groups did not benefit from intensive psychotherapy derives from their apparent inability to use an interpretative approach. Are these children changing because of verbally mediated interpretations or is some other process at work?

Kennedy, very much in the classical psychoanalytic tradition, described developmental limitations on children's capacity for insight at each developmental stage. The egocentric, non-reality-based orientation of school children precludes their gaining objective insight. Their needs and wishes dominate the moment. The problem for the latency-age child is exactly the opposite. The Piagetian stage of concrete operations of cognition is essentially an interest in exploring the external world and the rules by which it operates. Here, there is a resistance to exploring the inner world. Thus, while objectivity is possible, subjectivity is resisted. For the adolescent, other factors work against the attainment of insight. While adolescents are developing the capacity for abstract thinking, which should facilitate the attainment of insight, their increased tendency toward introspection is based on a defensive narcissism, which (in ways similar to the preschool situation) precludes an adequate perception of the external world and works against the attainment of full insight.

Cognitive studies reveal that many operations that are part of the interpretative process are difficult, if not impossible, for children at certain stages of development. S. Harter notes that while preoperational children can generate free association, their cognition limits their ability to distinguish fantasy from reality, their ability to take in what the therapist may articulate for them, and their difficulty in appreciating connections and inferences. Free association in the latency period is resisted as regressive, and the external orientation is maintained. While the expression and understanding of emotion is basic to the process of psychotherapy (particularly the experience and understanding of ambivalent emotion), the child's capacity both to experience contradictory emotions and to experience these emotions directed toward the same person may not emerge until the end of the latency period. Similarly, an appreciation of the difference between manifest and latent feelings emerges later in latency. Here a belief that emotions cannot be hidden progresses to an understanding that emotions can be hidden, to a realization that mental operations can put feelings out of one's mind, to an understanding, in adolescence, that certain thoughts and feelings are not always consciously available. Much interpretative work requires some understanding of latent feelings. The recognition and appreciation of feelings in parents also develop later. Some emotions (e.g., fear and happiness) lag behind others (e.g., anger and sadness) in being differentiated from parental feelings. The sense of self, often invoked in the language of interpretation, also undergoes development into adolescence. The self is defined first by concrete behaviors and attributes and then by increasingly abstract and differentiated attributes. Finally, an awareness of dreams as the product of the child's own mind again emerges throughout latency. Thus, these various mental operations that are invoked in the interpretative process are not well established at the ages when many children undergo psychotherapy.

In considering the cognitive level of children and attempting to determine what types of interventions are possible for their individual level of understanding, one must remember that when children are confronted with intense emotion or unusual stresses (as often occurs in psychotherapy), they are likely to experience both emotional regressions and cognitive regression. In fact, all regressive behaviors have both emotional and cognitive components. Thus, although a child may have recently achieved more-advanced cognition in a certain area, this achievement will most likely be lost when it is most needed, namely at crucial points in therapy when

emotion reaches a certain pitch. Possibilities of interpretation are then correspondingly diminished.

Finally, as first highlighted by Anna Freud, children use the therapist equally as a new object in their lives and as a transference. They are interested in the new experiences provided by therapy and are open to developing new identifications with the therapist. This means, for instance, that while therapists may think they are interpreting children's anger at a parent, the children perceive an entirely new way of someone reacting to their anger and talking about it. This new experience itself may well be internalized. If and when the anger emerges in the transference, the children can then replace their identifications with their parents (who may themselves have reacted angrily, critically, or punitively to the children's anger) with an identification with the therapist who is reacting in a benign manner.

ALTERNATIVE FORMULATIONS

P>While maintaining a belief in an unconscious determination of many types of symptoms and an unconscious motivation for many behaviors, a number of authors have recognized that even within the process of psychotherapy, factors other than interpretation may account for much change that occurs. While such considerations apply to the resolution of dynamically based conflicts, these authors also attempt to balance dynamic factors with other etiological sources of pathology such as biologically based deficit, biologically based constitutional trends, and social learning.

Enhancing Mentalization Efrain Bleiberg, Peter Fomagy, and Mary Target have presented a partial reformulation of child psychoanalysis that is developmentally based rather than relying on conflict resolution. It attempts to enhance the skills of self-reflection and perception of others, or *mentalization*, and derives from recent research findings on the outcome of intensive psychoanalytic treatments. In the large group of children with complex psychopathology who benefited from intensive psychoanalytic treatment in these studies, there appeared to be two clusters of patients. Cluster A patients showed unstable reality contact and thought disorganization in situations without structure, impoverished relationships, communication difficulties, and poor affective regulation. These patients had diagnoses from the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) that included schizotypal and schizoid personality disorders and milder forms of pervasive developmental disorder. The Cluster B group patients showed intense hunger for social reponsiveness, a variety of reactions to separation, affective lability, and significant self-centeredness. DSM-IV diagnoses included attention-deficit/hyperactivity disorder, conduct disorder, separation anxiety disorder, and mood disorders. Despite the obvious heterogeneity of these groups, they either pervasively (Cluster A) or intermittently (Cluster B) lacked the ability to be aware of their own and other's mental states.

In light of these deficits, the child analytic approach focused on enhancing mentalization includes the following goals: (1) facilitating verbalization and differentiation of feelings, (2) helping the child to master manageable units of anxiety, (3) to appreciate cause and effect within relationships, (4) to develop better reality testing, (5) to understand limit setting, (6) to increase frustration tolerance, (7) establishing internal representations of affect that can lead to self-regulation, and (8) expanding the child's appreciation of ambivalent and conflicting thoughts and feelings.

A number of techniques are proposed to achieve these goals. Enhancing reflective process is achieved by improving children's observation, labeling feeling states, and tracing sequences of behaviors. Strategies that allow children choices before action, consideration of alternative actions, and possibilities of verbal instead of physical action will strengthen impulse control and self-regulation. Becoming aware of others requires attention to the smallest indicators of a child's awareness of other's feelings or thoughts. Attempts to make reparations are often seen in play. The experience of guilt, however displaced, may also indicate that children have some early awareness of their effect on another. Attention to the therapist's countertransference may serve as a guide. Transference considerations differ in the two clusters. Cluster A patients require an active use of the therapist to integrate these new processes. Cluster B children adhere to a rigid, often hostile, organization of the relationship that usually tests the therapist's ability to remain empathic and requires patience, bargaining, and cajoling on the part of the therapist to help these patients find other ways of relating.

Many of the above techniques may appear to be techniques used in cognitive therapy, but this approach to psychoanalytic psychotherapy is less prescriptive and anticipates resistance to change that is psychodynamically based. Furthermore, here the actual therapeutic relationship is the tool of integration, rather than prescriptive exercises or reinforcements.

Integrated Psychodynamic Psychotherapy Owen Lewis presented integrated psychodynamic psychotherapy, an approach to dynamic psychotherapy based on four clinical principles. The first of these, the dual purposes of psychotherapy, holds that psychotherapy proceeds in two sometimes contradictory directions. Psychotherapy must function as a specific treatment for a specific disorder in a child at a specific developmental stage. With these parameters, specific techniques are used to treat the condition. On the other hand, psychotherapy also functions as an open-ended exploration of the unconscious factors that may have a role in the child's problems. Even if the disorder is developmentally or biologically based, it will be understood and incorporated by the child in unique ways that must be elucidated. These factors cannot be known in advance and are discovered in the course of psychotherapy.

The second principle, the use of all modalities of treatment, follows from the first. The complexity of most psychiatric disorders and the inherent limits of any one approach warrant the use of any combination of modalities of treatment. While certain approaches (e.g., classical psychoanalytic treatment) were seen as incompatible with other modalities of treatment, integrated psychodynamic psychotherapy finds a positive synergistic interaction between modalities. Integrating treatment approaches requires innovation to create the best comprehensive treatment plan for a given child.

The various modalities of treatment (e.g., behavioral, cognitive, familial, psychopharmacologic, and educational) are organized through a basic psychodynamic understanding. The third principle involves tracking the psychodynamic process. How these various modalities interact is monitored through shifts in the material presented by the patient. Does alleviating an anxiety symptom through a behavioral intervention make the patient more or less interested in pursuing its cause? Do the effects of an antidepressant make a patient more or less able to talk about troubles? Intervention by intervention, session by session, the flow of material is watched. The therapist also notes positive interventions that soon yield negative results. If the dynamic meaning of change is not appreciated, the patient will resist change.

The fourth principle holds that psychotherapy proceeds through the therapeutic relationship. In this interpersonal perspective, all therapy is ultimately organized within the relationship with the therapist, and without that relationship, there is no therapy. While transference is a necessary component to this relationship, patients are also bound by their developmental capacity for relatedness. This capacity is determined not only by developmental stage, but also by the patient's accumulated experience. A therapist who offers a child help with anxiety symptoms through a relaxation technique can simultaneously and symbolically expose a deficit in care by a parent (transference) and provide new opportunities for care (new identifications). From this point of view, the therapist must equally note transferences and the patient's resistance to these transferences as well as the real relationship and the patient's resistance to this.

Change, for the child, comes from a variety of sources. Specific behavioral techniques will shape behavior. Family interventions can modify parenting approaches. School consultations can help create an environment for children that is more conducive to their learning and social environment. All these interventions (which may or may not be accepted by the child), are processed through the therapeutic relationship and color that relationship in ways that have dynamic meaning. While classic interpretation may be used, interpretation per se often formulates for the child change that has already occurred. This change ultimately evolves from the new emotional, cognitive, and interpersonal integrations achieved within the therapeutic relationship.

A 4½-year-old boy was brought to treatment because of his persistent use of a diaper and other infantile behaviors. Up to the age of 3 he was developing normally, if not precociously, when his sister was born. He now clung to his bottle, diaper, and changing table, and resorted to baby-talk with his parents. After a year and a half of trying various ways to eliminate these behaviors, the parents sought consultation.

A series of behavioral rewards was established for the elimination of the infantile behaviors, the use of the bathroom for changing diapers, and finally sitting on the toilet seat after defecation.

Play initially involved the boy playing the part of a drill sergeant with a shrill whistle and harsh commands. One day the child devised a game in which he climbed inside a cardboard box, declared himself to be “Mr. Duty,” and had the therapist speak to him through a vaccum hose. After a brief discussion of his cramped quarters, he rumbled out, and therapist and patient changed roles. This game went on for several weeks.

One day, he announced to his father that he might want to use the toilet. His father, himself now well trained, gave him his choice. From that point on the child used the toilet.

A new game evolved shortly after this. Playing the part of a mother kangaroo with a baby kangaroo stuffed into his “pouch,” he hopped around the office telling the baby it was not yet time to come out.

Both the behavioral and the dynamic treatments contributed to this boy's improvement. Of particular note, the dynamic meaning of his encopresis was not revealed until after the symptom improved, namely equating the feces with the baby (sibling).

Other combinations of treatments have shown similar synergistic effects. Medication, for instance, often serves as an emotional probe. What does the medicine mean to the child? What does illness mean to the child? What improvement is attributable to the child, what to the pill? Similarly, family therapy and individual therapy have important interactions. Not only can family therapy improve communication, it can also challenge or confirm views the patient has of other family members, and thus hasten improvement.

SUGGESTED CROSS-REFERENCES

Section 7.2 discusses the psychiatric report and medical record, and Section 7.6 discusses psychological and intellectual assessment of children. Section 30.1 covers psychoanalysis and psychoanalytic psychotherapy, Section 30.7 covers interpersonal therapy, and Section 30.12 covers combined psychotherapy-psychopharmacology. Short-term psychotherapy in children and adolescents is discussed in Section 48.2, cognitive-behavioral therapy in children and adolescents is discussed in Section 48.3, and family therapy is discussed in Section 30.5 and Section 48.5. Section 48.10 covers the psychiatric treatment of adolescents.

SECTION REFERENCES

*Ablon S: The therapeutic action of play. J Am Acad Child Adolesc Psychiatry *35:*545, 1996.

*Bleiberg E, Fonagy P, Target M: Child psychoanalysis, critical overview and a proposed reconsideration. Child Adolesc Psychiatr Clin North Am *6:*1, 1997.

Fonagy P, Moran G: Studies of the efficacy of child psychoanalysis. J Consult Clin Psychol *58:*684, 1991.

*Freud A: The relations between child analysis and adult analysis. In *Normality and Pathology in Childhood*. International Universities Press, New York, 1965.

*Harter S: Cognitive-developmental consideration in the conduct of play therapy. In *Handbook of Play Therapy*, S Schaefer, K O'Connor, editors. Wiley, New York, 1983.

Healey W, Bronner AF: The child guidance clinic: Birth and growth of an idea. In *Orthopsychiatry, 1923-1948: Retrospect and Prospect*, L Lowry, editor. American Orthopsychiatric Association, New York, 1948.

Katan A: Some thoughts about the role of verbalization in early childhood. Psychoanal Study Child *16:*184, 1961.

Kennedy H: The role of insight in child analysis: A developmental viewpoint. In *Psychoanalytic Explorations of Technique*, H Blum, editor. International Universities Press, New York, 1980.

Kennedy H, Moran G: Reflections on the aim of child analysis. Psychoanal Study Child *46:*181, 1991.

Kernberg P: Individual psychotherapy. In *Comprehensive Textbook of Psychiatry*, ed 6, H Kaplan, B Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Lewis M: Interpretation in child analysis. J Am Acad Child Adolesc Psychiatry *13:*32, 1974.

*Lewis O: Integrated psychodynamic psychotherapy with children. Child Adolesc Psychiatr Clin North Am *6:*53, 1997.

Marshal R: Countertransference in the psychotherapy of children and adolescents. Contemp Psychoanal *15:*595, 1979.

Meeks J: Children who cheat at games. In *Psychoanalytic Treatment of the Child*, J McDermott, S Harrison, editors. Aaronson, New York, 1977.

*Mordock JB: Some risk factors in the psychotherapy of children and families: Well-established techniques that can put some clients at risk. Child Psychiatry Hum Dev *29:*229, 1999.

*Ritvo R, Al-mateen C, Ascherman L, Beardslee W, Hartmann L, Lewis O, Papilsky S, Sargent J, Sperling E, Steiner G, Szigethy E: Report of the psychotherapy task force of the American Academy of Child and Adolescent Psychiatry. J Psychother Prac Res *8:*93, 1999.

Sperling E: The role of play in child psychotherapy. Child Adolesc Psychiatr Clin North Am *6:*69, 1997.

Target M, Fonagy P: Research on intensive psychotherapy with children and adolescents. Child Adolesc Psychiatr Clin North Am *6:*39, 1997.

Tyson R, Tyson P: The concept of transference in child psychoanalysis. J Am Acad Child Adolesc Psychiatry *25:*30, 1986.

*Weiss B, Catron T, Harris V, Phung TM: The effectiveness of traditional child psychotherapy. J Consult Clin Psychol *67:*82, 1999.

48.2 SHORT-TERM PSYCHOTHERAPY

EUTHYMIA D. HIBBS, FH.D.

Short-term therapy with children and adolescents flourished in the last two decades in response to several factors. Long treatment experience and frequency of premature termination and drop-outs suggested that most children and their parents could not sustain long-term commitments. Second, because short-term therapy addresses a specific problem or behavior and not the overall functioning of the child, a limited number of sessions may suffice to correct the problem. In the last decade managed care limited the number of sessions, based on the results of the efficacy studies, forcing clinicians to short-term treatments regardless of the needs of the patients or the extent of the problem. Finally, medication has replaced long-term treatments to a significant extent, but because of the lack of pharmacological trials, the public is currently concerned about the extensive use of medication for the quick treatment of children.

Short-term therapy may be appropriate for a number of children. Surveys indicate that outpatient treatment with children is by default time-limited for several reasons. Children are usually brought to therapy by adults. When the recommendation to enter treatment comes from outside sources (e.g., schools, courts), parents may not be motivated to follow through. Further, children sometimes may not be available for long-term commitments, or they may not be interested (or cooperative) in entering open-ended therapy. Financial constraints and managed care's limits on treatment length also mitigate against long-term therapy.

Since the late 1980s psychotherapy research with children and adolescents has expanded, especially in modalities that claim to be of short duration, such as behavioral, cognitive, and cognitive-behavioral therapies. These studies have reported success rates, but problems remain. Not many children benefit to the same degree from these short-term therapies, and the benefits they derive are not always stable. Manuals are useful in efficacy studies to standardize treatment across subjects, therapists, and settings and secure the validity of the treatment, but manuals do not ensure the generalizability of treatments. Further, nonspecific effects of treatments such as moderators, mediators, and the role of the therapeutic alliance are not considered in efficacy studies with children. Managed care companies misinterpret the results of studies on the efficacy of short-term treatment and request uniform therapy for all conditions.

However, research on short-term therapy enhanced the scientific theory and understanding about the ways people change and adapt to their environment. Short-term therapy may benefit children during periods of developmental reorganization and may contribute to a sense of continuity, relational patterns, acquisition of particular skills, and environmental disruptions.

This chapter discusses only the four major short-term therapies that are most commonly used in the clinical field: psychoanalytic/psychodynamic psychotherapy, behavior therapy, cognitive therapy, and cognitive-behavioral therapy. These treatments, with the necessary adjustments, may be applied in individual or group therapy with children and adolescents.

DEFINITION

Short-term psychotherapy is defined as any intervention that is guided by a core set of principles, based on the theoretical model of choice, and uses techniques essential for the specific treatment. Short-term psychotherapy is limited to 1 to 30 sessions (as opposed to open-ended treatments), depending on the theoretical basis of the therapy. Psychodynamically oriented psychotherapy aims to develop a limited psychogenic understanding of the focal problem. The major therapeutic technique, interpretation, is adapted to the brief format. Behavioral, cognitive, and cognitive-behavioral therapies aim to identify the primary problem and the discrete behaviors that maintain the disorder. The main therapeutic techniques used are verbal or cognitive mediation to bring about behavior change. Therapists maintain a consistent focus on the goals of therapy and allow little deviation from the problem at hand.

HISTORY

Short-term therapies for children and adolescents are derivations or downward extensions of adult treatments. Therefore, their history is blended with the development of brief therapies for adults. Only a few specific techniques such as play therapy and parent management skills have been specifically developed to treat children.

Short-term therapies began with Sigmund Freud who sought a quick cure for neuroses. He believed that once the etiology of the neurosis was known, it would lead to a quick solution and resolution. For example, Freud treated the composer Gustave Mahler in a single 4-hour session and the conductor Bruno Walter in six sessions. Freud's book *Studies on Hysteria* indicates that in some cases, once the problem was identified and discussed, it was also solved. However, as Freud sought a solution for the problem of resistance, psychoanalytic therapy lengthened considerably.

Otto Rank may be considered the forerunner of contemporary brief therapies. His theory on separation and union postulates that in the course of human development, individuals move between emotional attachment and dependency and separation and independence. He emphasized that in therapy, the patient learns to deal with separation from an ambivalent love object through engagement, attachment, and eventual loss. Thus time-limited therapy, at least with some patients, may be the treatment of choice because it requires them to deal with the dread of time and its implications for loss. This theoretical concept, though, is quite different from the reasons why short-term therapies were established. In reality, the movement toward short-term therapies was principally a response to long waiting lists, limited resources, other public health needs, and (in the last two decades) economic factors.

During World War II the need for short therapeutic interventions became crucial, to address the stress-related symptoms developed by soldiers. The psychosocial changes and traumas resulting from the war pressured the psychotherapeutic community to explore briefer, more direct methods of treatment. The earliest short therapy was crisis intervention aimed at stress reduction, symptom relief, and prevention of greater breakdown. The goal was to restore self-esteem and avoid retreat into more serious maladjustment. These short-term interventions were based on psychoanalytic principles and the transference phenomena, regardless of the length of the therapy. The work of Franz Alexander and Thomas French, which clarified a path toward a briefer therapy, was also based on the psychoanalytic understanding of the personality and the process of psychotherapy. They tried to define the basic principles that make possible a shorter and more efficient means of psychotherapy. They advanced the concepts of “corrective emotional” experience that focused on emotional change based on the reexperiencing of old conflicts in the therapeutic relationship and the “principles of flexibility” in which the therapist creates an interpersonal climate.

Originally, the psychoanalytic thinking was that a specific diagnosis was not necessary to treatment. However, as early as 1941, Abraham Kardiner indicated in his book *The Traumatic Neuroses of War* that a precise diagnosis must be linked to a specific intervention. Kardiner may have been at the vanguard of the diagnostic systems. George Kriegman and Harlan Wright in 1947, for example, studied a specific diagnostic problem—enuresis among soldiers—and evaluated several techniques such as suggestion, reassurance, hypnosis, and reeducation as interventions.

The impetus to use effective treatment methods comes principally from supply-and-demand pressures. World War II generated a need for short-term interventions, and the psychotherapeutic community responded appropriately. The Veterans Administration initiated the development and adoption of short-term therapeutic methods by assigning veterans to private psychiatrists for treatment and contracting for a limited number of sessions. Otto Fenichel described brief psychotherapy as the “child of bitter practical necessity.” Short-term interventions thereafter became a reality. In 1958, Leopold Bellak opened the Trouble Shooting Clinic in Queens, New York. About the same time the emergency clinic at Bronx Municipal Hospital was created. In 1962 the Benjamin Rush Center for Problems of Living was

established in Los Angeles. Other centers and training programs burgeoned in the country in response to lengthening waiting lists of patients. Most short-term therapies at that time were based on psychoanalytic principles, and the therapists were psychoanalysts who tried to adapt to the demands of their time.

Another turnabout that helped to establish short-term therapies was the social movement of the 1960s with its political counterculture activities, availability of drugs, and change in moral standards. Many individuals at that time experienced situational and personal turmoil that often led to a need for psychological help. In response, many counterculture “rap” centers, drop-in clinics, other alternative centers, and free clinics were established. These interventions, however, were often delivered by nonestablishment volunteers who offered nontraditional short-term support and counseling.

While the brief therapies based on psychoanalytic principles were unfolding, behavior therapy was being developed and established. Behavior therapy focused on addressing specific treatment to specific problems. Treatments were established to be short term, with a proposed systematic assessment of the diagnostic status, treatment process, and treatment outcome. Behavior therapies also redefined the role of the therapist by ignoring the alliance issues and the therapist's creativity. The therapist was required to take direct charge, assume an active role in the process, and direct the client.

Short-term therapies for children and adolescents also surfaced in the 1940s. The expense and time-consuming nature of longterm psychiatric care became prohibitive because of broken appointments, dropouts, and the inability of children and parents to sustain long-term commitments. There was also an urgency to offer services to many more children and families who required care. Brief service was developed by the Philadelphia Child Guidance Clinic whose goal was to assess the dynamic elements that could be used in short clinical services. In 1949 the Buffalo Children's Hospital developed a new structure for short-services cases. Specific patients were initially selected to try out the new therapeutic model that was later extended to all possible patients. All child short-term therapies that followed were based on the adult model of psychoanalytic-psychodynamic therapy. During the first two sessions children were helped to clarify their emotions, feelings, concept of self, wishes, desires, and expression of needs via symbolic play, imagery, and fantasy. Today, the first two sessions of almost all therapeutic modalities are used to assess the child and determine the problem to be addressed in the short-term therapy. Several childhood problems (e.g., school phobia, misbehavior, internalized disorders) were then treated with either psychoanalytic psychotherapy or behavior therapy. In the 1970s a major shift toward cognitive therapies took place and a combination of cognitive-behavioral therapies is most commonly used today.

Research on short-term therapies for children and adolescents is scanty. Besides the treatment of Little Hans by Freud, one of the earliest short-term therapy efforts with children appeared in 1946 in a study by Evelyn Alpern. He compared the outcomes of two approaches: seeing parents and children separately or seeing them together. The outcome measures were clinicians' evaluations of the child's progress—no statistical analyses were used. The emphasis on empirical research on the treatment of childhood disorders started in the early 1980s and intensified in the 1990s with the National Plan for Research on Child and Adolescent Disorders, a plan generated in response to a congressional mandate.

TECHNIQUES

Most short-term psychotherapies share some basic characteristics. They are very selective about the problem to be treated and the patients to be involved; that is, the problem or symptom must be circumscribed, with careful selection of inclusion and exclusion criteria. The first two or three sessions are dedicated to the assessment of the child and the problem. In addition, the principles of the short-term therapy are explained, emphasizing the immediacy and urgency of the change process, as well as the time limits, and techniques to be used. From this point on, both therapist and patient concentrate on working on the core symptoms. Each therapeutic modality follows its own theory-based principles of treatment. The therapist takes a direct and active role in promoting client functioning while simultaneously encouraging the child's capability for change. The cognitive-behavioral therapist assigns tasks or homework, and the patient is expected to be an active participant in the therapeutic process.

SELECTION CRITERIA

Children with less severe psychopathology are usually selected for short-term therapy. For short-term psychoanalytic/psychodynamic psychotherapy children need to possess object relational capabilities, basic trust, ego defenses that are functioning adequately, flexibility, and the ability to rapidly enter and then leave a meaningful relationship. Children who can profit from behavior therapy and cognitive-behavioral therapy are those whose problem is circumscribed, such as specific phobias, symptoms of the anxiety spectrum disorders, symptoms of affective disorders, disturbance of conduct and other well-defined childhood problems. Excluded are children with chronic psychopathology, such as psychotic symptoms, developmental disabilities, traumatic abuse, and comorbidity and those who lack sufficient cognitive and verbal skills. Also excluded are children with a history of severe loss, such as parental loss, maternal deprivation, and severe family pathology.

Indications Short-term therapies, in general, are indicated for specific, well-delineated problems. Psychodynamic psychotherapy is recommended for children with relatively simple problems and basically healthy personalities who have difficulty relating to parents or other important figures in their lives or who suffer from mild phobias, mild anxiety disorders, or dysthymic disorder. They must have above average intelligence and be motivated to feel better. For developmental reasons (i.e., childhood plasticity in the unfolding of the personality) children with less severe psychopathology may profit from a short-term intervention to remove the obstacles and permit resumption of the development process. Behavior therapy is indicated for specific, well-delineated behaviors of children and aims to ameliorate the symptom or behavior. Like the other short-term therapies, cognitive therapy is indicated for children with a prescribed problem who are intelligent and verbal and can understand the focus of the therapy. In addition, the child must be aware of the problem and the emotional fluctuations and be able to establish a working relationship with the therapist.

Limitations Short-term therapies have many limitations, but the main limitation is that severe conditions or syndromes are excluded. Short-term psychodynamic psychotherapy is not indicated for highly resistant children or those with antisocial behaviors, developmental disorders, or other serious psychopathology. The same is true for children with severe psychopathology whose families are of low educational level. Behavioral treatments can address only one behavior at a time; therefore, they are not recommended for clinically dysfunctional populations and those with comorbid conditions. Cognitive therapy also may not be the treatment of choice for children with comorbid disorders and those with severe psychopathology. In general, short-term therapies for children may be effective for some distinct problems, but they do not address the overall functioning of the individual.

SHORT-TERM PSYCHOANALYTIC-PSYCHODYNAMIC PSYCHOTHERAPY

Definition Definition of psychoanalytic-psychodynamic psychotherapy is difficult because it has evolved and been modified throughout its 100-year existence. There are some underlying treatment assumptions: (1) a continuity exists between the mental states of “normality” and “abnormality,” as well as the belief of a causal continuity between the mental states of infancy, childhood, and adulthood. Within these continuities, motivational instinctual drives and conflicts affect all aspects of mental life. (2) The observable actions, behaviors, or verbal communications are the sources of information about the subjective experience. (3) Organisms strive to reach a steady state (adaptation, homeostasis) in the face of pressures from external environmental sources, internal physiological sources, and powerful sources within the mental apparatus itself. Defense mechanisms are short- and long-term adaptational forces in the control and modification of conflicting motivational tendencies. (4) Most mental events are outwardly expressed because they originate in multiple converging psychological processes. (5) The unconscious aspects of thought that are for the most part inaccessible to the conscious mind can be brought into consciousness to introspective awareness. These unconscious aspects become accessible only under specific circumstances such as dreams or some psychotic states or are implied in the symbolic meanings of symptoms, parapraxes, and slips of the tongue, and above all they can be constructed by the clinical method of psychoanalysis.

Short-term psychodynamic psychotherapies, therefore, focus on the underlying personality processes, the expression of feelings, the development of insight, and the identification and (eventually exploration) of new ways of functioning. The therapeutic relationship, or therapeutic alliance, is accorded a special role and serves as the central agent for therapeutic change.

Theoretical Issues In psychoanalytic-psychodynamic psychotherapy, developmental impasses are considered to result from the interaction of limitations of structural adaptive capabilities in the child and environmental stressors such as losses, familial inadequacy, illness, or other psychosocial strains. Symptoms may be understood in terms of the child's unsuccessful efforts to cope with the experience of being overwhelmed by traumatic or chronic stress. Psychodynamic concepts addressing personality structure (e.g., ego strength, reality testing, tolerance for frustration) are seen in the context of essential adequate external support. Character structure is deemphasized, and the adaptive fit of child and environment becomes the locus of attention. Short-term psychotherapy relies on formulating a clinical focus to organize the therapeutic activity and keep it on track. The central issue is formulated on the basis of the presenting symptoms, the developmental history, and the current family situation. The termination addresses the child's ambivalence issues of loss and separation.

Therapy uses the basic principles of psychoanalytic theory, which include diagnosing the problem with consideration of the developmental context and the

psychoanalytic concepts of development, including the notions of regressions, fixation, and the possibility of a developmental breakdown. The defense mechanisms used and the concepts of transference and resistance are also examined. This occurs typically in 10 to 30 sessions, and the therapy is focused on the problem at hand. The aim is not to restructure the child's personality but to modify the existing balance of internal and external forces to address some observable life stressors, both in therapy and in the extended environment, to reactivate the developmental imperatives. Therefore, classifying the child in a categorical diagnosis is not as helpful in predicting outcome as the assessment of interpersonal, intrapsychic, and object relational capacities required by short-term therapy. The decision to use a brief treatment must weigh the child's social and familial milieu, since the child remains within the influence of the family, which thus has a great impact on the outcome of treatment.

Techniques Several theoretically based models of long-term psychodynamic psychotherapy techniques have been adapted to short-term treatments with children. They all follow a systematic process. Therapists maintain the focus of the treatment, are more active, sometimes give directives, are flexible, and address termination. During the assessment phase, therapists take a complete history and form a clinical opinion about operating dynamics and symptomatology. In addition, a clinical opinion is also formed on the basis of the history given by the children and their parents concerning the problem; the children's relationships with siblings, peers, and parents are examined; and their behavior during the interview is considered. The children's motivation is also assessed. From the onset, therapists are supportive, empathetic, and reassuring and show that they care about the child's feelings and problem. An effort is made at the initial phase to establish a good working relationship, and the time of termination is decided.

The next step is to help the children acquire insight into what is behind their problem. This can be achieved through the use of dreams, play, or transference interpretations. Sometimes interpretations go beyond the phenomenological material and include unconscious defensive operations, suppressed instinctual tendencies, and hidden meanings of behavior patterns. Interpretations are accompanied by warm, empathetic, and emotionally available therapists who help the children understand the basis of the problem, how it is related to the presenting dysfunction, and how it may affect their development.

The ensuing stage is devoted to helping the children resolve the conflict that brought them to a therapist, by encouraging them to use the new insight or understanding of the problem at hand and then putting the insight into action. That is, to challenge their fears and engage in actions that are anxiogenic but hold promise of rewards. Therapists and children discuss the strategies and possible consequences of the actions, with repetition encouraged until the children feel comfortable doing so, then they test and expand their repertoire to other challenges in life. The final task, termination, is to integrate what has been learned so that the children realize that only they can control anxiety and tension about events precipitated from the environment and recognize that they can find different ways to address a stressful situation.

Some specific therapeutic techniques are used with children such as finger-painting or painting, puppets, and other games. All aim at releasing children's internal or fantasy world to understand it, be less afraid of it, and guide it in reaching age-appropriate functioning.

Play therapy, the technique of choice for young children ages 4 to 12, includes age- and gender-appropriate activities such as games, painting, building, and puppets or dolls that allow children to “play out” their feelings, their fears, anger, loneliness, and their feelings of failure or inadequacy. In play therapy children are given the opportunity to learn about their problems and their relationships with others and work on the issue that brought them to therapy. The aim is to ameliorate clinically relevant domains of functioning and learn to cope better with the problem at hand. Play therapy also promotes the development of trust and confidence in communicating feelings. Several techniques based on different theoretical concepts are used for that purpose.

Mutual story telling is a technique in which first children tell a story. The therapists then infer its psychodynamic meaning and tell a story of their own, that contains the same characters in a similar setting but introduces healthier adaptations and resolves the conflicts exhibited in the children's story.

Story making consists of therapists writing stories using the children's expressed concepts and their knowledge of the children's problems. The aim is to enter the children's world and facilitate the children's expression of their fantasies while suggesting avenues in which the children can overtly express their feelings and involving them in the world beyond fantasy. The stories are adjusted to reflect the individual needs of the children.

No short-term treatment with children is conceptualized without involving the parents or the family as whole. One-on-one sessions with the child are supplemented by family sessions. The child is often best helped by assisting the parents in handling the changes brought about by therapy and the maturation of their child. Therefore, parents are usually actively engaged in individual or conjoined therapy sessions. They are involved at the outset of treatment, both for information gathering and in the development of a therapeutic process. If necessary, parents are helped to recognize and resolve their own conflicts and their conflicts toward their children. The techniques used for parents are mainly psychoeducational approaches, support, clarification, and direction.

> John, a bright 14-year-old, was treated with brief (25 sessions) psychotherapy. The initial complaint was that his grades had dropped during the academic year, he had withdrawn from sports, was anhedonic, had difficulty relating to his peers, and whined a lot. His parents divorced when John was 7 years old. He had two younger sisters, ages 12 and 9. John “hated” his father, therefore he missed many visitations with him, pretending that he was sick or too busy to see him. His sisters kept close contact with the father. Mother had a live-in boyfriend who moved in the year that John became symptomatic. John also “hated” the boyfriend. He felt miserable, and he reproached every one in his environment. During the first two therapy sessions, two issues to be addressed in the brief therapy were delineated: John feeling rejected by his father and mother (who had found another man) and rivalry. During the following treatment sessions, an empathetic and supportive therapist helped John acquire insight into his feelings through interpretations of his defenses, transferential manifestations, and clarifications. John and the therapist discussed strategies and activities to be carried out during the treatment period to alleviate his discomfort. By session 16 John had reestablished regular contact with his father and was able to tolerate mother's boyfriend. The termination phase included the integration of what had been discussed, learned, and practiced, and gave control to John to find ways to understand his internal conflicts and find more appropriate ways of managing them. His rivalry issues had diminished since he was also able to share his father with his sisters. Biological parents and mother's boyfriend were seen in parallel. The sessions consisted of psychoeducational approaches concerning John's developmental level of functioning and the way he perceived and experienced his environment. Parents were helped to recognize and handle John's problems as well as their own conflicts, and strategies were proposed for facilitating John's development. At termination it was agreed that John would return to see the therapist for one follow-up session every 3 months in the first year and every 6 months in the following 2 years. At the 2-year follow-up, it was apparent that John had improved academically and had resumed his outside activities such as sports. He remained sensitive to rejection, but he was able to use the skills he had learned to manage those feelings.

Research and Evaluation To date there are few empirical, manual-driven studies on time-limited psychodynamic psychotherapy with children and adolescents. The few existing studies have methodological problems. Therefore, the efficacy of the short-term psychoanalytic/psychodynamic treatments remains untested. Success with short-term psychodynamic therapy usually appears in case reports.

BEHAVIOR THERAPY

Definition *Behavior therapy* (behavior modification) is a generic term that embraces many different techniques used to treat a variety of clinical problems. Therefore, it is difficult to provide a simple definition. Behavior therapy has eight main principles: (1) it concentrates on behavior rather than an underlying cause; (2) it assumes that maladaptive behaviors are learned; (3) it assumes that learning principles can be effective in modifying behaviors; (4) it sets specific, clearly defined treatment goals; (5) it rejects the classical trait theory; (6) the therapist adapts the method of treatment to the child's problem; (7) it concentrates on the here and now; and (8) therapists place great value in obtaining empirical support for their technique. In general, behavior therapy focuses on promoting adaptive prosocial behaviors and aims to decrease maladaptive behaviors. Behavioral treatments give primary attention to behavior as a means of dealing with clinical problems.

Theoretical Issues Behavioral interventions include many different theories about how clinical problems emerge and how they are maintained. These theories vary in their explanations of behavior and in the role they accord to the influences of processes within the individual (thoughts, beliefs, and perceptions) or processes resulting from events in the environment (cues, feedback, and consequences that behavior produces). Behavior therapies have their roots in the seminal work of Joseph Wolpe, B. F. Skinner, and later on Albert Bandura.

Wolpe's work on the experimental development of neurosis in cats led to a most important and enduring clinical technique known as systematic desensitization. Wolpe presumed that anxiety underlies most neurotic or maladaptive behaviors. His theory relies on the Pavlovian model and assumes that anxiety has been classically conditioned to harmless stimuli. Based on this postulate, several treatment techniques such as systematic desensitization, in vivo desensitization, and reciprocal inhibition have been developed to decondition the autonomic nervous system responses to key stimuli in the individual's environment.

Another influence on behavior therapy was Skinner's work that proposed that the principles derived from his study of operant conditioning and in particular the effect of learning the consequences of behavior could be applied therapeutically. Operant conditioning assumes that most emotional problems are reactions to overcontrolling, punitive environments, lack of control, or both. Therefore, therapies based on operant conditioning aim to change the frequency of overt behaviors and control the consequences that follow behaviors to increase or decrease those behaviors. Behavioral techniques based on this model that are used frequently in short-term behavior therapy with children include positive reinforcement, negative reinforcement, punishment, extinction, shaping, token economy, and contingency contracting.

Bandura's work relies heavily on social learning theory, which greatly influenced behavioral interventions. Bandura emphasized that behavior change involves cognitive symbolic processes, as evidenced by observational learning. The premise is that behavior can be altered by providing new learning experiences. For example, social skills training includes modeling (a form of role playing), behavioral rehearsal (practice duplicating a behavior), correction feedback (from therapist), coaching, and reinforcement. Bandura's work was the forerunner of the cognitive-behavior therapies.

In general, behavior therapy assumes that people become who they are through both learning from interaction with the environment and genetic loading. Difficulties occur when learning is inadequate or faulty. Problems can be corrected by offering necessary learning experiences.

Techniques Behavioral interventions combine treatment techniques from all behavioral schools. The main characteristic is that they focus directly on overt behaviors and give primary attention to behavior as a means of dealing with the clinical problem. They do not rely on the therapeutic relationship or the development of psychological insight. Rather, they rely on directive and active treatments using problem-oriented approaches. Children and parents play an important role in developing the plan of action to change the behavior.

There are five stages in the behavioral interventions.

1. Detailed analysis of the child's problems and factors related to them. That is, the behaviors to be altered are carefully identified and assessed, using mostly (in the case of children) multiinformant techniques. Significant people in the child's life (e.g., parents, teachers, and others affected by the child's behavior) are involved in evaluating the extent and nature of the problem.
2. Determination of the specific goals for treatment. That is, the goals, the means for reaching them, and procedures are discussed and specified with the child and parents.
3. Development of a treatment plan involving the use of the techniques of behavior therapy that are applicable to the child's problem.
4. Implementation of the treatment plan. Explicit training experiences may be prescribed, such as homework assignments, to facilitate learning new behaviors to be used in everyday life. The actions or activities are carried out both within and outside the treatment sessions (rehearsals).
5. Objective evaluation of the results of treatment to determine whether the desired outcome has been realized. Although treatment is time limited, it is possible, depending on the outcome, to extend and modify the treatment plan based on the feedback provided by such evaluation.

The following are some of the behavioral techniques most commonly used in therapy.

Systematic desensitization involves relaxing all the muscles of the body, followed by practice sessions. The therapist and child construct an anxiety hierarchy. During relaxation the therapist asks the child to imagine the least anxiety-producing item. The process continues by introducing higher anxiety level stimuli until anxiety is eliminated.

Self-control desensitization is an active mediational process in which the child learns to interrupt maladaptive thinking through relaxation. It consists of four steps: directing the child's attention to the sensations associated with both tension and relaxation, incorporating a variety of stressful situations in the anxiety hierarchy to promote generalization, having the child continue imagining the anxiety-producing scene while practicing relaxation, and encouraging the child to practice in all anxiety-producing situations.

Exposure consists of imaginal or in vivo desensitization, is highly structured, and involves the actual confrontation of the fear-producing stimuli.

Positive reinforcement refers to the application of a positive reinforcer following a response or behavior to increase the occurrence of that response or behavior.

Token economy involves establishing an environment in which appropriate behaviors are consistently reinforced by generalized conditioned reinforcers (tokens) that can be exchanged for backup reinforcers. It involves defining the target behaviors to be reinforced, selecting the tokens and backup reinforcers, developing a system of monitoring and assessing token economy, and developing a plan for phasing out the contingencies.

Response cost is associated with the token economies and involves the withdrawal of money or tokens for inappropriate behaviors. Response cost is more effective when it is used in conjunction with positive reinforcements for alternative behaviors.

Punishment refers to the application of an adverse stimulus (e.g., withdrawal of privilege) after a nonacceptable response or behavior has occurred. The goal is to reduce the probability that the behavior will recur.

Contingency management refers to a formal agreement between two or more individuals (therapist, child, parent) that defines what behaviors are to be changed and what consequences will follow performance of those behaviors. An effective contingency contract must explicitly define the role of each party and the reward for meeting the responsibilities. The behaviors included in the contract must be capable of being monitored. A system of sanctions for failure to meet the contract's terms should be in place, and the contract should define “bonuses” for consistent compliance. A record-keeping system should be established to provide constant feedback to the parties about the frequency of target behaviors and the delivery of reinforcers.

Time out refers to the removal of positive reinforcement for a specific period of time following the performance of an undesirable behavior. Ideally all sources of reinforcement are removed. Isolation is common (i.e, the child is removed from all sources of positive reinforcements). This technique has been successfully used in classrooms and home environments.

Indications Behavior therapy has been applied with considerable success to a wide variety of problems including phobias, obsessive-compulsive disorders, generalized anxiety, depression, enuresis, encopresis, substance abuse, aggression, delinquency, and eating disorders. One behavioral technique is rarely used for the treatment of childhood disorders. The following examples show how a combination of techniques is used.

Childhood Anxiety Disorders The main therapeutic procedure for the treatment of phobic disorders is exposure. In this technique the youth confronts anxious phobic objects or events so that anxious phobic symptomatology can eventually diminish. Exposure can involve in vivo or imaginal forms or sometimes both. Exposure is usually accompanied by systematic desensitization. Self-control strategies are used to teach the child how to use appropriate cognitive strategies to facilitate exposure or approach behavior toward feared objects.

Since parents are an integral part of the treatments involving children, they are offered training in contingency management that defines what behaviors are to be changed and what consequences will follow performance of those behaviors. A detailed contract is established between therapist, parents, and child, instructing the parents to administer a specific reward to the child when the behavior is performed.

Attention-Deficit/Hyperactivity Disorder Operant and social learning principles are mostly used in the treatment of children with attention-deficit/hyperactivity disorder to shape more appropriate academic and social behaviors and reduce defiant, aggressive behaviors. Self-control procedures, self-monitoring, and anger management (recognition of cues that provoke anger, strategies to counteract anger, and choice of plans to address provocation that results in anger) are the main behavioral techniques used. Parental and school or teacher involvement is a prerequisite. Contingency management procedures are established that use reward and response cost at home and in the classroom. Behavior management procedures are taught to parents and teachers who implement targeted strategies in the child's natural environment (i.e., using positive attending or ignoring skills to promote appropriate behavior, setting a home or classroom token system, and using time out).

Conduct Problems Parent management training involves working primarily with parents, with little or no contact with the children. The aim is to develop specific skills

in the parents so they can implement procedures effectively at home. Parents are trained to identify, define, and observe behaviors in new ways. When the behaviors of interest are identified, parents are taught how to reinforce or punish these behaviors. Methods include use of social praise and tokens for prosocial behaviors or mild punishment, such as time out, and loss of privileges for unacceptable behaviors.

Mary, a 16-year-old girl, presented with social phobia. She was afraid to stand up and talk in front of the other children in her class. She perspired profusely, panicked, and forgot the topic she was supposed to talk about. If she knew in advance that she had to give a presentation, she stayed home sick. For the assessment phase both Mary and her parents were interviewed to determine the nature and extent of the problem. The therapist explained the procedures and the length of the treatment, which consisted of 16 sessions. The technique of choice for social phobia is exposure. Twelve of the sessions were dedicated to gradual desensitization and imaginal and in vivo exposure. Mary and the therapist constructed the public speaking anxiety hierarchy. Four treatment sessions were devoted to systematic desensitization and imaginal exposure. Mary was asked to pay attention to the sensations associated with anxiety and relaxation while imagining anxiety producing scenes and practicing relaxation. The next sessions consisted of gradual in vivo exposure; that is, Mary was asked to give a presentation to a person in the family that she trusted most, then gradually to other family members, then to a stranger, and eventually to more strangers and familial individuals. Sessions 15 and 16 were devoted to termination and evaluation of the treatment results. Mary's symptoms were eliminated. She was able to give class presentations by the end of the 12 sessions. A 12-month follow-up indicated that Mary had maintained the gains and functioned at the premorbid level.

Research and Evaluation By the 1960s behavior therapists had begun systematic research on the applications of behavioral techniques to emotional and behavioral problems of children. Operant conditioning programs have produced dramatic behavior changes in settings such as classrooms, institutions, and at home. Parent management training, in which parents are trained to alter the child's behavior at home, is probably the child treatment technique investigated most in controlled outcome studies. It has been used successfully to treat children with disruptive behavior disorder, attentiondeficit/hyperactivity disorder, pervasive developmental disorders, mental retardation, learning disorders, eating disorders, and other childhood psychological problems. Parent management has been proved beneficial except with dysfunctional families in which parents are unwilling or unable to participate and follow through with the treatment. Many manual-driven, empirical studies exist today that support the use of behavior therapy for the treatment of various childhood problems. However, neither the long-term effects of treatment nor the rates of relapse have been reported.

COGNITIVE THERAPY

Definition *Cognitive therapy* uses an active, directive, time-limited, structured approach in a family of techniques that seeks to produce change of the child's cognitive processes by altering affect and behavior. The approaches differ in both process and procedure. The premise is that thoughts and faulty learning determine the behavior; therefore, children misconceive their environment. The focus of change may include perceptions, self-statements, attributions, and expectations. Some approaches emphasize modifying distorted beliefs. Others attempt to compensate for perceived deficits in cognitive skills.

Theoretical Issues Three theories have strongly influenced the development of cognitive therapy. Albert Ellis suggested in 1962 that psychopathology arose from faulty, maladaptive, and irrational beliefs. This theory led to rational-emotive therapy whose objective is to substitute adaptive thoughts for maladaptive ones. In 1973 Donald Meichenbaum's self-instructional theory emerged. The aim of the therapy based on this theory is to teach individuals to increase their coping skills through education, controlling physiological arousal, rehearsing, and physical relaxation. Following these theories, in 1976 Aaron Beck developed a treatment for depression based on the findings that depressed people have negative views of themselves, their experiences, and the future. This therapy's objective is to alter distorted thinking by helping the individual consider alternatives.

In general, the theoretical premise of cognitive therapy is that cognition is a significant variable in the development and maintenance of psychopathology. Models of cognition maintain that to understand human functioning, one should recognize not only the role of the environment in directing the individual's experiences but also the role of the individual in interpreting that environment. The cognitive models of psychopathology suggest that the individual's construct of the experiences largely determines the emotional reactions to events and future behaviors in similar situations. There is an essential interaction between the way individuals feel and behave, which in turn influences the way they define their world, their self, and their way of viewing the future. The assumption is that children with a deficiency in particular processes and those unable to use or apply cognitive skills, exhibit deviant behaviors.

The cognitive perspective takes into consideration the contributions of innate propensities or prior experience in which thinking plays a role in determining subsequent affect and behavior. Although not exclusively, cognitive theory suggests that thinking plays a causal role in generating and maintaining at least some disorders, in concert with other factors. Cognitive therapy thus seeks to relieve distress by providing new information-processing skills to changing maladaptive beliefs.

Techniques Cognitive therapy is an active and directive short-term intervention. It is a collaborative endeavor in which therapist, child, and parents work as a team. The goal is to help children to discover their dysfunctional and irrational thinking and to reality test their thinking and behavior. The therapist proposes hypotheses, models for coping and adaptation, and strategies for testing them, to alleviate the presenting symptoms and help the patient develop more adaptive and functional behavior. Cognitive therapy focuses on the cognitive processes believed to underlie disordered behavior, and it often uses specific techniques to change these processes. Following are some of the cognitive therapy techniques.

Problem-solving techniques are most frequently used for children who appear unable to generate alternative solutions to interpersonal problems, focus on ends or goals, and foresee the consequences of their actions. Problem-solving training is defined as a process that provides a variety of potentially effective response alternatives for managing problem situations and increases the likelihood of selecting the most effective response. George Spivak and Myura Shure have identified different cognitive processes or interpersonal cognitive problem-solving skills that underlie social behavior: alternative solution thinking (the ability to generate different options for solving problems in interpersonal situations), means-ends thinking (awareness of the intermediate steps required to achieve a particular goal), consequential thinking (the ability to identify potential results of a particular act or solution), causal thinking (the ability to relate one event to another over time and understand why one event led to a particular action by other persons, and sensitivity to interpersonal problems (the ability to detect a problem, identify its interpersonal aspects, and prevent a possible confrontation).

Systematic rational restructuring focuses on the immediate cognitive factors that mediate maladaptive behaviors and emotions. The way children evaluate a situation determines their emotional reaction to the situation. Through imagined scenes or role-playing the child is exposed to an anxiety-producing situation and asked to identify any self-defeating, anxiety-provoking cognitions produced by, or related to, the situation. The therapist helps the child rationally reevaluate these cognitions.

Self-instruction training is a step-by-step approach that includes defining the problem, focusing on the task, and self-evaluation. Therapists perform a task while talking to themselves out loud. The children then perform the same task, first under the therapists' instruction then while instructing themselves aloud. Next the children whisper the instructions to themselves as they perform the task. Finally, the children perform the task while using private, or internalized (and inaudible), speech. The loud dialogue provides a model for the children until they incorporate the process and use private speech.

Other techniques used alone or in variation with others include thought stoppage, which involves stopping irrational beliefs before they begin to cause discomfort and role-playing, in which children are encouraged to see themselves in the perspective of other people and to take the roles of other people.

Indications Cognitive therapy was developed to treat depressive and anxiety disorders in adults. However, the format was adapted to the treatment of children and other problems in addition to depression and anxiety. Because cognitive therapy is time limited, there must be a fit between the child's characteristics and the therapist. The therapist must establish clear goals at the outset, remain focused, and work toward these goals. Parents are also involved in the child's treatment and are trained in problem-solving skills so that they may facilitate and promote the child's skills, enhance communication among family members, and change the child's interpersonal environment.

Cognitive therapy is rarely used today as the sole modality for the treatment of children with mental disorders. Cognitive-behavioral therapy, which combines cognitive and behavioral techniques, is the current treatment of choice for childhood disorders. Therefore, application of cognitive therapy is discussed below in the cognitive-behavioral section.

Research Issues Cognitive therapy has been used successfully in various settings, such as outpatient, inpatient, and school settings, to change maladaptive cognitive processes in children. Outcome studies indicate that variations of cognitive therapy produce greater change than nontreatment or control conditions. The changes are usually demonstrated in measures that reflect behavior in the home, school, or community. There is no information on the effects on long-term

functioning of these children.

COGNITIVE-BEHAVIORAL THERAPY

Definition Cognitive-behavioral therapy resulted from efforts to expand the conceptual base of behavioral techniques by according cognitive processes a more prominent role. Cognitive-behavioral therapy thus has its roots in both behaviorism and cognitive theory. It represents a more comprehensive approach to understanding impairment, etiology, and symptom display and incorporates techniques from both treatment modalities. A combination of treatment techniques has emerged.

Theoretical Issues Cognitive therapy was developed by dynamically trained theorists and tended to emphasize the role of meaning—what a person thinks is not as important as what that person believes. Since behavior therapy could not deal with nonobservable cognitive events, theorists trained as behaviorists developed cognitive-behavioral therapy. In this theoretical concept, thinking tends to be conceptualized in a more concrete fashion and is often regarded as a set of covert self-statements (private behaviors) that can be influenced by the same laws of conditioning that influence other covert behaviors.

The cognitive theorists led the development of strategies for examining the rationality or validity of existing beliefs, whereas the cognitive-behavioral theorists focused on the development of strategies for teaching specific cognitive skills. Both approaches borrowed from each other to the point that today the distinctions between them are blurred. Thus, cognitive-behavioral therapy focuses on the cognitive processes believed to underlie disordered behaviors while simultaneously using specific behavioral techniques such as rehearsal reinforcement to change these processes.

The cognitive-behavioral model includes the role of both affect and social context. Therefore, the cognitive-behavioral analysis of child and adolescent disorders and adjustment problems as well as related analyses of treatment-produced gains includes considerations of the child's internal and external environment and represents a blended perspective. That is, it integrates the cognitive, behavioral, affective, and social strategies to bring about change.

Techniques Selected techniques within behavior modification, particularly those based on operant conditioning (e.g., positive reinforcement, punishment, token economy, response cost), emphasize the role of environmental influences on behavior. On the other hand, social learning theory explains behavior on the basis of an integration of cognitive processes and environmental events. Cognitive processes are believed to be pivotal because they mediate environmental influences and determine what the individual attends to and perceives and how the environment is interpreted.

Indications Cognitive-behavioral therapy is probably the most extensively used therapy today for the treatment of all psychological problems, especially in children ages 9 years and above because of their ability to verbalize their problems. It is used for both group and individual therapy. Following are some examples of the application of cognitive-behavioral therapy to various childhood disorders.

Conduct Disorder Children with conduct disorders may be impulsive, oppositional, defiant, and angry. Assessment sessions delineate the problem to be addressed. The way the child approaches the situation and the thought process is evaluated, and based on this evaluation, therapist and child construct an agenda for systematic therapy. The child is then taught to use a step-by-step approach to solving problems. Problem-solving skills training involves having children make statements about themselves concerning the problem and guiding them to a more appropriate solution. Dysfunctional thought processes are examined, and the therapist helps children generate alternate solutions to problems, focus on goals, see the consequences related to their behavior, and recognize the causes of other people's behavior and become less sensitive to it. This technique may be adapted to include ages 4 to young adults. The treatment uses structured tasks such as games, academic activities, and other age-appropriate tools. The therapist plays an active role by making verbal statements and applying the sequence of statements to a particular problem while encouraging the child to use the skills. Modeling follows with practice and role playing. Punishment or withdrawal of privileges may be applied whenever necessary.

Group Therapy of Adolescent Depression Peter Lewinsohn and his colleagues developed the Adolescents Coping with Depression Course, which consists of 16 2-hour sessions conducted over an 8-week period for groups of up to 10 adolescents. It includes a psychoeducational component that aims to destigmatize the problem, to emphasize skill training to promote control over one's mood, and to enhance the adolescents' ability to cope with problematic situations. The group activities include role playing. Social skills training is distributed throughout the therapy to facilitate and enhance communication by teaching conversational techniques, planning social activities, and developing strategies for making friends. Sessions are designed to increase pleasant activities, based on the assumption that depressed adolescents have little positive reinforcements in their lives. Relaxation training is provided to enhance enjoyability and performance in social situations. Focus on changing depressogenic cognitions is provided by identifying, challenging, and changing negative thoughts and irrational beliefs. Six sessions involve teaching communication skills focusing on the acquisition of positive behaviors such as paraphrasing to verify the message, active responding, and appropriate eye contact. Adolescents are also taught negotiating and problem-solving techniques such as defining the problem without criticism, brainstorming alternative solutions, evaluating and mutually agreeing on a solution, and specifying the agreement with the inclusion of positive and negative consequences.

Obsessive-Compulsive Disorders John March and his colleagues developed the cognitive-behavioral treatment called How I Ran OCD Off My Land. This treatment uses a variety of techniques such as exposure, response prevention, extinction, anxiety-management training, reinforcement, modeling and shaping, habit reversal, individual therapy, group psychotherapy, and family therapy. Following the assessment, the first session is dedicated to educating the child about obsessive-compulsive disorder and categorizing it in the neurobehavioral framework, that is, equating obsessive-compulsive disorder with a medical illness and detailing the treatment. In the second session the aim is to make the disorder the target of treatment. The child is asked to give obsessive-compulsive disorder a disparaging nickname, so it becomes the enemy and allows the child to “boss back” the disorder. In addition, the use of story metaphors allows development of a therapeutic relationship and treatment monitoring through the use of informal symptom diaries. The third session is dedicated to mapping the disorder, that is, describing specific obsessions, compulsions, triggers, avoidance behaviors, and consequences. Sessions 4 to 16 deal with anxiety-management training and exposure and response prevention. Anxiety-management training includes relaxation, diaphragmatic breathing, and constructive self-talk. Exposure involves therapist-assisted imaginal and in vivo exposure. Finally, response prevention consists of exercises coupled to weekly homework. Parents are trained to ally with the child to “boss back” obsessive-compulsive disorder.

Socially Rejected Children Fred Frankel and his colleagues developed the 12-session cognitive-behavioral treatment called Parent Assisted Social Skills Training. First the child's social network is examined taking in consideration the natural environment, family, school, neighborhood to develop an integrated approach to treatment. Children are seen in groups to learn the rules of peer etiquette. They are also offered training skills to expand their peer network, while parents and children are taught how to work together to promote more-successful play dates, improve the child's competence with nonaggressive responses to teasing and conflict with children and adults. Coached play begins in session 2, in which coaches do not participate in the activities but observe the children's play. They only intervene to dispense reinforcements (verbal and token) and if necessary a 2-minute time-out for misbehavior. Child socialization homework begins in session 1, with the aim of setting the mechanics of the telephone call in place. First children call someone from the group, then children outside the group. Also, children may bring a toy from home, so they may engage other children in play. Didactic presentation of parents consists of informing the parents of their part in their child gaining peer acceptance and ensuring that the parents adhere to their assigned roles in the child's socialization homework and provide supportive feedback for the principles being taught.

Patrick, a 15-year-old boy presented with aggressive, violent behavior; used offensive language toward his mother, teachers, and other adults in his environment; and made up stories (my father is the head of the Mafia) to impress his peers, who in turn rejected him. He was under court order to enter psychotherapy because he threw a flower pot from the twenty-fourth floor of the building where he resided and almost injured a pedestrian. Patrick is the only son in a divorced family. His parents divorced when he was 4 years of age. His father then left the country and was never heard from again. His mother was alcoholic but has been sober the last few years. Although highly educated his mother could not hold a job. The communication between mother and son was hostile and charged with threats. Patrick was threatening to physically abuse his mother and she responded by threatening to deprive him outings and television, but she never followed through on her threats. The treatment plan consisted of seeing mother and Patrick separately for 20 sessions plus 6 sessions conjointly at the end. The two first sessions were dedicated to assessment and discussion of the treatment plan. The techniques used for both parties were 6 sessions of *anger management*, that is, to recognize body warnings such as getting flushed and muscle tension; to stop and go away from the distressing situation; to avert their gaze from the other; and to take a few deep breaths and count backward from 15 to 0. Twelve sessions of cognitive restructuring were devoted to reorganizing Patrick's and mother's distorted thinking. Specific self-statements were offered for each of the areas of conflict (e.g., “She or he may have started the tirade but I do not have to answer the same way,” “If she or he asks me to do something she/he may need it or be looking after my interest”). In addition, mother who was more eager to learn was offered psychoeducational training concerning adolescent development, acceptable and unacceptable adolescent behaviors, and ways to respond. At first Patrick resisted this “silly stuff,” but soon he complied because his mother had changed her behavior, which facilitated his acquiescence. Conjoined sessions were used for role playing, practice, and termination. Patrick began to change because he discovered that he was more successful in making his ideas and wishes respected. This change generalized to the peer group.

Research Issues Cognitive-behavioral therapies are the most used and the most manual-driven treatment modalities. Cognitive-behavior therapy has been successful in treating a variety of child and adolescent problems and even more severe conditions than psychodynamic therapy, behavior therapy, or cognitive therapy alone. Some evidence suggests that cognitive-behavior therapy may be useful for the treatment of comorbid conditions; however, it is preliminary and further study is needed. As with the other short-term therapies, there is a dearth of information concerning long-term effectiveness. The existing information on follow-up ranges between 6 to 18 months; thus the long-term effectiveness of cognitive-behavioral therapy is unknown.

INTERPERSONAL PSYCHOTHERAPY

Interpersonal psychotherapy is a developing and promising therapy for adolescents with depression. It has not been tested extensively, so only a summary of its principles is presented. Interpersonal therapy was originally developed by Gerald Klerman and his colleagues for the treatment of depressed adults. It places the depressive episode in the context of interpersonal relationships and focuses on current interpersonal conflicts. It aims at decreasing symptomatology and improving interpersonal functioning.

Laura Mufson and her colleagues have recently adapted this treatment for depressed adolescents. Interpersonal therapy consists of three phases. Phase 1 (sessions 1 to 4) comprises a diagnostic assessment, including evaluating the child's social and familial relationships, identifying the problem areas, and establishing a treatment contract. In phase 2 (sessions 5 to 8) therapist and child begin to work on one or two of the problem areas. The child is encouraged to express feelings and monitors depressive feelings. Techniques used include exploratory questioning, encouragement of affect, linkage of affect with events, clarification of conflicts, communication analysis, and behavior change techniques such as role playing. Parents are also involved in the treatment. Phase 3 (sessions 9 through 12), which is addressed at the beginning of treatment, has as goals giving up the relationship with the therapist and establishing a sense of competence in dealing with future problems. No research results are available at this point, and it has not been tested with younger children.

SUGGESTED CROSS-REFERENCES

Information related to child and adolescent group psychotherapy may be found in Section 48.4 on group psychotherapy, combined individual and group psychotherapy, and psychodrama, in Section 48.5 on family therapy, in other sections of Chapter 49 and in Section 51.6c on ethics in psychiatry. Normal child development is discussed in Section 32.2 and normal adolescent development in Section 32.3.

SECTION REFERENCES

Albano AM, Barlow DH: Transfer of control: A psychosocial intervention model for internalizing disorders in youth. In *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*, ED Hibbs, PS Jensen, editors. American Psychological Association, Washington, DC, 1996.

Alexander F, French TM: *Psychoanalytic Therapy*. Ronald Press, New York, 1946.

Barten HH: *Brief Therapies*. Behavioral Publications, New York, 1971.

Bergin AE, Garfield SL: *Handbook of Psychotherapy and Behavior Change*. Wiley, New York, 1994.

Birmaher B, Brent D: Practice parameters for the assessment and treatment of children and adolescents with depressive disorders. J Am Acad Child Adolesc Psychiatry *37:*63S, 1998.

Brent DA, Holder D, Kolko D, Birmaher B, Baugher MA, Roth C, Iyengar S, Johnson BA: A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry *54:*877, 1997.

Brestan EV, Eyberg SM, Boggs SR, Algina J: Parent-child interaction therapy: Parents' perceptions of untreated siblings. Child Fam Behav Ther *19:*13, 1997.

Fenichel O: Brief psychotherapy. In *The Collected Papers of Otto Fenichel*, H Feniche, D Rapaport, editors. Norton, New York, 1954.

Fisher DJ, Himle JA, Hanna GL: Group behavioral therapy for adolescents with obsessive-compulsive disorder: Preliminary outcomes. Res Soc Work Pract *8:*629, 1998.

Frankel F, Cantwell DP, Myatt R: Helping ostracized children: Social skills training and parent support for socially rejected children. In *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*, ED Hibbs, PS Jensen, editors. American Psychological Association, Washington, DC, 1996.

Freeman A, Greenwood VB: *Cognitive Therapy*. Human Sciences Press, New York, 1987.

Harrington R, Whittaker J, Shoebridge P: Psychological treatment of depression in children and adolescents: A review of treatment research. Br J Psychiatry *173:*291, 1998.

Hibbs ED: Improving methodologies for the treatment of child and adolescent disorders. J Child Abnorm Child Psychol *26:*1, 1998.

Hibbs ED, Clarke GN, Hectman L, Abicoff HB, Greenhill LL, Jensen PS: Manual development for the treatment of child and adolescent disorders. Psychopharmacol Bull *33:*619, 1997.

Hibbs ED, Jensen PS: *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*. American Psychological Association, Washington, DC, 1996.

Kazdin AE: *Treatment of Antisocial Behavior in Children and Adolescents*. Dorsey, Homewood, IL, 1985.

Kazdin AE: *Behavior Modification in Applied Settings*. Brooks/Cole, Pacific Grove, CA, 1994.

Kendall PC, Flannery-Schroeder E, Panicelli-Mindel SM, Southam-Gerow M, Henin A, Warman M: Therapy for youths with anxiety disorders: A second randomized clinical trial. J Consult Clin Psychol *65:*366, 1997.

Kish EH: Brief psychotherapy with children, adolescents and their families. Child Adolesc Psychiatr Clin North Am *6:*137, 1997.

Lewinsohn PM, Clark GN, Rohde P, Hops H, Seeley JR: A course in coping: A cognitive-behavioral approach to the treatment of adolescent depression. In *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*, ED Hibbs, PS Jensen, editors. American Psychological Association, Washington, DC, 1996.

*Lewinsohn PM, Clarke GN: Psychosocial treatments for adolescent depression. Clin Psychol Rev *19:*329, 1999.

March JS, Mulle K: Banishing OCD: Cognitive-behavioral psychotherapy for obsessive compulsive disorders. In *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*, ED Hibbs, PS Jensen, editors. American Psychological Association, Washington, DC, 1996.

Messer SB, Warren SC: *Models of Brief Psychodynamic Therapy*. Guilford, New York, 1995.

Mufson L, Moreau D, Weissman MM: Focus on relationships: Interpersonal psychotherapy for adolescent depression. In *Psychosocial Treatment for Child and Adolescent Disorders: Empirically Based Strategies for Clinical Practice*, ED Hibbs, PS Jensen, editors. American Psychological Association, Washington, DC, 1996.

Schmidt E: Brief psychotherapy with children and adolescents: A developmental perspective. Child Adolesc Soc Work J *13:*275, 1996.

*Silverman WK, Ollendick TH: *Developmental Issues in the Clinical Treatment of Children*. Allyn & Bacon. Boston, 1999.

Skinner BF: *Science and Human Behavior*. Macmillan, New York, 1948.

Small L: *The Briefer Psychotherapies*. Brunner/Mazel, New York, 1979.

Webster-Stratton C: Parent training with low-income families. In *Handbook of Child Abuse Research and Treatment*, JR Lutzker, editor. Plenum, New York, 1998.

Wells RA, Giannetti VJ: *Handbook of Brief Psychotherapies*. Plenum, New York, 1990.

Wolpe J: *The Practice of Behavior Therapy*. Pergamon, New York, 1973.

48.3 COGNITIVE-BEHAVIORAL PSYCHOTHERAPY

JOHN S. MARCH, M.D., M.P.H.

With the emergence of a rich psychopathology literature, child and adolescent psychiatry has moved away from nonspecific interventions toward problem-focused treatments keyed to specific fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) diagnoses. In particular, the past 40 years have seen the emergence of diverse, sophisticated, empirically supported, cognitive-behavioral therapies that cover the range of childhood-onset mental disorders. Hence, many now believe that cognitive-behavioral therapy administered within an evidence-based, disease management model is the psychotherapeutic treatment of choice for many if not all pediatric mental illnesses.

HISTORY AND DEFINITION

Historically, behavior therapy evolved within the theoretical framework of classical and operant conditioning, with cognitive interventions assuming a more prominent role with the increasing recognition that person-environment interactions are powerfully mediated by cognitive processes. Looked at in the context of situational or cognitive processes, behavioral therapy is sometimes referred to as nonmediational (emphasizing the direct influence of situations on behavior) and cognitive therapy as mediational (emphasizing that thoughts and feelings underlie behavior). Behavioral psychotherapists work with patients to change behaviors and thus reduce distressing thoughts and feelings. Cognitive therapists first work to change thoughts and feelings, with improvements in functional behavior following in turn. For the most part these two main conceptual positions and newer contributing theories, such as information processing theory are usually subsumed under the broad umbrella of social learning theory.

For example, two-factor conditioning theory frames posttraumatic stress disorder as a stimulus-driven anxiety disorder in which both classical (factor one) and instrumental, or operant, (factor two) conditioning play important roles. In classical conditioning, the stressor, or traumatic event, acts as an unconditioned stimulus that in turn elicits an unconditioned (reflexive) response in the child, characterized by extreme fear and the cognitive perception of helplessness and terror. Cognitive, affective, physiological, and environmental cues accompanying the traumatic event then constitute conditioned stimuli, often called “traumatic reminders” by pediatric traumatologists. In turn, traumatic reminders become capable of eliciting a conditioned response in the form of posttraumatic stress disorder symptoms, which should decrease (habituate) with prolonged exposure, assuming the absence of real threat. In the second factor (instrumental conditioning), children quickly learn by trial and error how to reduce posttraumatic stress disorder symptoms through cognitive and behavioral avoidance and sometimes anxiety-dampening rituals such as checking or seeking reassurance. These operant anxiety-reducing behaviors preclude the extinction of trauma-based anxiety and foster stimulus generalization, rendering them relatively straightforward as targets for exposure-based interventions.

Like psychoanalytic theories, cognitive-behavioral theories regarding the genesis of psychiatric symptoms also recognize that internal mental phenomena play an important role in mediating behavior. Continuing the posttraumatic stress disorder example, Edna Foa and colleagues note that persons with this disorder develop “fear structures” that are conditioned by both the event and the posttraumatic stress disorder symptom picture. Fear structures (which contain automatic stimulus-response elements such as verbal, somatic, and behavioral cues and information regarding the meaning of the event) are exceptionally sensitive to activation by internal and external cues reminiscent of the initiating trauma, including thoughts and affects incorporated during and after the event. These alterations in meaning determine in part whether the event is traumatic, that is, whether it becomes a stressor. In turn, the cognitively oriented therapist attends closely to the ideational content of the trauma-based fear structure, actively supplying both corrective and behavioral information as required.

Cognitive-behavioral therapy also meshes nicely with developmentally sound neurobehavioral approaches to pediatric mental illness in which psychopathology can be seen as analogous to a learning disability mediated by highly conserved central nervous system (CNS) information processes. As with academic skills, children normally acquire social-emotional (self and interpersonal) competencies over time. The failure to do so, relative to age-, gender-, and culture-matched peers, may reflect capacity limitation, individual difference in the rate of skill acquisition for specific competencies, environmental factors, development of a mental illness, or a combination of these factors. In cognitive-behavioral therapy, the mental health practitioner must understand the presenting symptoms in the context of constraints to normal development and tailor a target-specific cognitive-behavioral treatment program that eliminates those constraints so that the youngster can resume normal developmental trajectory insofar as possible.

Finally, to the extent that symptom relief occurs, it is assumed to reflect concurrent changes (e.g., learning) in the CNS. In this view, the cognitive-behavioral treatment of pediatric mental illness can be thought of as partially analogous to the treatment of juvenille-onset diabetes, with the caveat that the target organ, the brain in the case of major mental illness, requires interventions of much greater complexity. For example, treatment for diabetes and obsessive-compulsive disorder both involve medications, which in diabetes might be insulin and in obsessive-compulsive disorder, a serotonin reuptake inhibitor. Each also involves crucial psychosocial interventions that work in part by biasing the somatic substrate of the disorder toward more normal function. In diabetes, the psychosocial treatment of choice is diet and exercise, and in obsessive-compulsive disorder, cognitive-behavioral therapy. Finally, not everyone recovers completely, so some interventions need to target coping with residual symptoms. Just as diabetic foot care must be targeted in diabetes, patients and their families need help to cope with residual symptoms in obsessive-compulsive disorder.

GENERAL CHARACTERISTICS OF COGNITIVE-BEHAVIORAL THERAPY

Most mental health clinicians are familiar with treatments that assume that psychological distress results from historical and current relationship problems that must be uncovered and addressed in therapy. In contrast to these more “story-oriented” approaches to psychotherapy, cognitive-behavioral therapy uses a problem-solving model in which the clinician acts as a coach to teach the patient a set of adaptive coping skills (and also to unlearn unskillful coping behaviors) for specific symptoms associated with distress and impairment in the present. From the point of view of differential therapeutics—defining different treatment interventions for different treatment targets—this hypothesis-driven process is somewhat analogous to a game of pick-up sticks. Put prosaically, one must correctly identify the targets of treatment at the symptom level (the sticks) before sequencing a set of target-specific interventions (picking up the sticks in the proper order) to help the patient get better. Put experimentally, cognitive-behavioral therapy allows viewing the treatment of each patient through the lens of one of several possible single-case designs.

Behavioral Assessment Vagueness is anathema to expert cognitive-behavioral therapy; that is, the cornerstone of cognitive-behavioral therapy is a careful functional analysis of problem behaviors that is governed by several important assumptions. First, behavior (both normal and problematic) is primarily governed by environmental contingencies (and in cognitive theory, by thoughts and emotions), so the relation of thoughts, feelings, and behaviors is the primary focus of assessment. Second, the antecedents and consequences of target behaviors and the behaviors themselves must be operationally defined and accurately measured. Third, behavior may differ across settings, so multiinformant, multimodal, multidomain assessment is critical. Fourth, treatment planning depends on careful assessment with periodic reassessment of how behaviors have changed and revision of treatment interventions as necessary. Thus, behavioral assessment continues throughout treatment.

In terms of the disease management model, modern cognitive-behavioral therapy requires clear specification of the behavioral/emotional syndrome (e.g., depression); and within the syndrome, the problems (e.g., oppositional behavior); and within the problems, the symptoms (e.g., won't go to bed) targeted for intervention. Consequently, a thorough diagnostic assessment that includes both a clinical interview and a multimethod, multiinformant, multidomain scalar evaluation is needed to design a cognitive-behavioral treatment plan. The overall assessment goal is to move from the presenting complaint through a DSM-IV five-axis diagnosis to a treatment plan tailored to the problems besetting the child or adolescent. This evaluation might begin with review of demographic, developmental, treatment and psychiatric/medical history; review of findings from normed rating scale data; school records and previous mental health treatment records; a clinical (or preferably structured) interview of the child and parents covering Axes I through V of DSM-IV; a formal mental status examination structured interview; and in some cases, a

specialized neuropsychological evaluation.

Of all these assessment tools, gender-, age- and race-normed rating scales are perhaps the most efficient way to collect information about both internalizing and externalizing behavioral disturbances at home and school. Excellent scales with good psychometric properties are now available for self-report of conduct problems, anxiety, and depression. Besides assessing an overall construct (e.g., anxiety), child self-report measures also provide useful information about factors (e.g., physical anxiety symptoms) and items (e.g., suffocation anxiety). This both speeds the interview and begins a dialogue between the doctor and the patient about the patient's most troubling symptoms, which facilitates treatment planning. This procedure is consistent with medical evaluation procedures across other medical specialties and meets goals for guidelines-based practice in managed care.

After the diagnostic evaluation, the clinician inevitably needs detailed data on the patient's specific symptomatology and its impact on the patient's developmental trajectory. These data provide a baseline for evaluating treatment and identifying targets for specific behavioral interventions. To explicate more clearly factors in the family that may be associated with childhood behavioral symptomatology, cognitive-behavioral therapy in pediatric patients depends on a molecular examination of the interactions between children and their parents. Hence, most cognitive-behavioral therapy manuals include detailed mapping of triggers, responses, and problem-maintaining factors (including family, peer, and school problems) as a routine part of treatment planning.

Finally the course and outcome of treatment is monitored with child-specific symptom checklists, standardized rating scales, or scales of global improvement, such as the Clinical Global Impairment and Improvement scales. Why evaluate outcome? First, tracking symptoms requires updating the problem target list periodically, which minimizes the possiblity of missing new or reemerging symptoms. Second, child and parent ratings let the clinician address any discrepant views of the child's progress. Third, rating scales detail how the child is progressing in treatment. In this regard, disorder-specific rating scales, whether standardized or tailored to the patient's problems, provide far more information than global measures, that simply involve therapist ratings of general outcome. With obsessive-compulsive disorder, patient symptomatology can be tracked with the Yale Brown obsessive-compulsive scale (YBOCS) symptoms checklist, obsessive-compulsive disorder symptoms with the YBOCS itself, specific cognitive-behavioral therapy targets with a stimulus hierarchy and anxiety in response to successive exposure to phobic stimuli (e.g., habituation curves), using a fear thermometer.

Cognitive-Behavioral Therapy Is Action-Oriented in the Present Many child and adolescent therapists are used to empathic, nondirective listening and especially to play therapy techniques that focus on uncovering past traumas. While good listening skills and creative play are important to cognitive-behavioral therapy, play by itself works poorly for the cognitive-behavioral psychotherapist who must instead actively focus on action-oriented, present-time change strategies. In other words, cognitive-behavioral therapy is best suited for patients with an identifiable problem and an active desire to change their behavior to reduce their suffering. Because cognitive-behavioral therapy in some ways resembles coaching an athletic team (i.e., the coach instructs, models, insists on practice, and finally stands on the sidelines while the players play the game for keeps), it is not surprising that cognitive-behavioral therapists and their patients do a lot of active communicating, talking as well as listening. In particular, cognitive-behavioral therapists typically use a Socratic dialogue in which they query patients about their experiences (e.g., “Can you be more specific?” “Tell me more about that.” “Can you give me an example of what we're talking about from the last week or so?”) This approach sometimes seems like a videotape replay in which the therapist and patient engage in a playful, two-way, detailed interchange about how the child or parents or both experience being in or successfully resisting the problem. Stylistic considerations, such as greater self-disclosure, therapist modeling, and extensive use of humor, also characterize successful cognitive-behavioral therapies.

Cognitive-Behavioral Therapy Includes Everybody Necessary While most psychiatric disorders tend to run in families, family psychopathology is neither necessary nor sufficient to generate mental illness in youth. Nonetheless, families affect and are affected by a child's symptoms, especially if more than one family member is affected. For example, high expressed emotion exacerbates behavioral problems; a calm, supportive family typically improves outcome. The treatment of choice for children with externalizing disorders is parent training, and for adolescents (especially when substance abuse is an issue), behavioral family therapy. Similarly specific behavioral programs have been developed to handle behavioral and pedagogic problems in the school setting. Cognitive-behavioral treatments for internalizing disorders are more commonly administered in an individual setting; however, a combination of individual and family sessions may be helpful when families are tangled up in anxiety or when treatment must involve transferring control from the therapist to the parent and then to the child. The failure to include parents, teachers, and (if needed) peers is a common source of treatment failure in child-centered therapies. Cognitive-behavioral therapy avoids this trap by including those considered crucial to the intervention. Hence, cognitive-behavioral therapy requires that the therapist include others in treatment as is appropriate at any given stage.

APPLICATION TO CLINICAL SYNDROMES

Although cognitive-behavioral therapy may appear to the casual observer to be a simplistic, unitary treatment, it is actually a diverse collection of complex and subtle interventions that must be mastered and then blended for the patient presenting with more than one mental disorder. Cognitive-behavior interventions for the disruptive behavior disorders concentrate on helping parents issue effective rules and instructions and subsequently manipulate rewards and punishments, with some cognitive training in anger coping for the child. Cognitive-behavioral therapy for depression directly confronts maladaptive depressogenic cognitions (e.g., helplessness, hopelessness, and hostility) and aims behaviorally to reconstitute relationships, whether they are intrapsychic, interpersonal, work, or spiritual. Parenthetically, cognitive-behavioral therapy for depression is the least specific treatment with respect to response; for example supportive, dynamic, and interpersonal psychotherapies also appear effective, and all are associated (not surprisingly) with a large placebo response. With the anxiety disorders, cognitive-behavioral therapy targets cognitions, autonomic arousal, and behaviors in a manner designed to promote habituation or extinction of inappropriate fears. Theoretically, these three treatments involve somewhat different emphases, for example, reward and arousal with the disruptive behavior disorders, interpersonally mediated affects in depression, and habituation theory in anxiety. Few therapists can discuss, much less apply, all the various forms of cognitive-behavioral therapy for pediatric mental disorders.

Despite their seeming differences, virtually all cognitive-behavioral interventions share four qualities: (1) an emphasis on psychoeducation, (2) a detailed behavioral analysis of the problem and the factors that maintain or extinguish it, (3) problem-specific treatment interventions designed to ameliorate the symptoms of concern, and (4) relapse prevention and generalization training at the end of treatment. As evidence-based therapies, each is supported by a more or less robust research literature and manuals are usually available to guide practitioners in using cognitive-behavioral therapy for specific problems. Cognitive-behavioral therapy thus fits nicely into the current medical practice environment that appropriately values empirically supported, brief, problem-focused treatments. Table 48.3-1 illustrates common cognitive-behavioral intervention strategies, most of which are common to a wide range of cognitive-behavioral treatments.

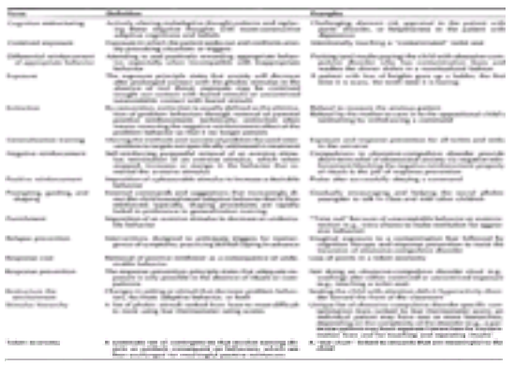

Table 48.3-1 Techniques of Cognitive-Behavioral Therapy

Disruptive Behavior Disorders Cognitive-behavior therapy targeting disruptive behaviors in children with attention-deficit/hyperactivity disorder has wider empirical support than any other area. It also provides a heuristically valuable example of how cognitive-behavioral therapy and medication can and should be combined.

First, exclusive treatment with psychostimulant medication is effective, but may not be maximally effective for attention-deficit/hyperactivity disorder, especially in older children or adolescents. Many children will experience reduced symptoms that nevertheless persist at clinically significant levels. Although many children improve substantially with pharmacotherapy, many do not achieve full normalization with stimulant treatments alone, even at very high dosages. More children may achieve

normalization with combination treatment.

Second, stimulant medication may not affect the full range of symptoms of children with attention-deficit/hyperactivity disorder and other, comorbid conditions. Although the primary symptoms of the syndrome (i.e., attention, impulsivity, and activity level) may be greatly improved, effects on other primary or comorbid characteristics of the syndrome (e.g., oppositional and aggressive behavior, academic underachievement, and poor peer relationships) are often insufficient. Because these conditions (along with family dysfunction) are robust predictors of poor long-term outcome for children with attention-deficit/hyperactivity disorder, treatment with stimulants alone may not significantly improve the outcome for these children. Research suggests the great relevance of the secondary problems of aggressive behavior (and the coercive family process associated with aggression), academic underachievement and associated school behavioral problems, and peer relationships to intervention in attention-deficit/hyperactivity disorder. These arenas are most amenable to psychosocial treatments. Moreover, psychosocial treatments may be the only treatments available for the minority of children with attention-deficit/hyperactivity disorder who do not respond to stimulants, who experience intolerable adverse effects, or whose parents reject the use of medication.

A final limitation of stimulant monotherapy involves its applicability to home behavior problems. Many pharmacotherapists limit the use of stimulants to school hours during the 9 months of the academic year to avoid appetite suppression, sleep disruption, and other adverse effects. This leaves parents to their own devices to manage impulsive, oppositional, and disruptive behavior in the afternoons and evenings, weekends, and summers. When no other treatment is provided, parents frequently become coercive, hostile, and overly punitive, which may exacerbate the child's behavior problems. Child and parent aggression are among the best predictors of poor outcome. For all of these reasons, cognitive-behavioral therapy continues to be used and evaluated, alone and in combination with stimulant medication.

Psychosocial treatment for attention-deficit/hyperactivity disorder optimally involves a clinical, behavioral therapy approach that focuses on the child, the parents, and the school. Intervention with parents typically uses parent training designed to teach parents skills and techniques for managing disruptive, impulsive, and oppositional behaviors in the home and in the community. Intervention in the school involves direct consultation with teachers about establishing behavior management systems in the classroom. In addition, systems that tie the school and home together in a direct, cooperative effort are increasingly being used to improve school behavior and academic performance. Innovative treatment programs can be tailored to the individual child.

In parent training for attention-deficit/hyperactivity disorder parents are taught a comprehensive set of behavior management skills geared toward the behavioral excesses and deficits displayed by these children. Common targets include improving the general family emotional climate, parental skills for positively reinforcing children's prosocial behavior, and parental skills for confronting disruptive child behavior with effective antecedents and consequences. For example, parents are taught how to spend structured, positive time with their children; how to “catch their child being good” and apply reinforcers on these occasions; how to teach their children to engage in independent on-task behavior, give effective instructions, and set appropriate rules and expectations with their children; and how to apply specific consequences to negative child behaviors. Typical child behavior targets are noncompliance with instructions, other oppositional behavior, aggression with sibs and other family members, independent behavior while the parent is busy, and impulsive, disruptive behaviors within the family. Because these youngsters frequently present management problems outside the home, parents are also taught to design mini-behavior-management programs for use on outings outside of home. Finally, more-recent innovative parent training programs for attention-deficit/hyperactivity disorder include stress management for parents as well as training them to become advocates for their child in the school system.

Another major aspect of clinical behavior therapy for attention-deficit/hyperactivity disorder involves the use of behavior management strategies in the classroom. Indeed, there are many more research evaluations of behavioral intervention in the school than in the home. School intervention usually begins with consultation with the teacher about the disorder and social learning and behavior management principles. Explaining the disorder is especially important if the teacher has labeled the child “lazy” or “bad” and has adopted a coercive posture toward the child. Even a teacher knowledgeable about general behavior management may need to learn the special refinements to standard behavior management needed for successful treatment of the child with attention-deficit/hyperactivity disorder.

A cueing system is developed to prompt teachers to increase their monitoring of ongoing child behavior and deliver positive attention at regular, frequent intervals. Teachers also are taught how to use ignoring and effective reprimands for disruptive behavior. A token system may be established with specific academic performance and behavioral targets. The child earns tokens or points at regular intervals (worked out with the teacher) for successfully achieving those targets. For younger children, tokens are usually delivered by the teacher. Self-monitoring and self-reward systems have been used for older children (third grade and above) in which they monitor and reward their own behavior, with occasional reliability checks by the teacher. Token programs for these children also include a response cost component (i.e., points are lost for inappropriate behavior), which research has shown to increase the effectiveness of reward programs with these children. A more-recent innovation in classroom behavior management uses daily report card systems that tie the home and school together in a cooperative effort; teachers monitor behavior at school and deliver points or tokens at regular intervals, and parents deliver backup consequences at home. With guidance from the cognitive-behavior therapist, teacher and parent work together over the course of the school year to refine the system.

While much research supports the efficacy (though not the maximum effectiveness) of clinical behavior therapy including parent and teacher training for children with attention-deficit/hyperactivity disorder (with the exception of anger coping), cognitive therapy with the child, once thought to hold great promise, has proved largely ineffective. Social skills training in the typical outpatient setting has also produced disappointing results. Consequently, innovative clinical researchers have turned to intensive summer day-camp treatment programs to provide academic remediation and address the peer problems and social and athletic skills deficits exhibited by these children. Because these areas of functioning contribute to low self-efficacy and self-esteem and poor long-term outcome and are not amenable to standard treatment, such innovative treatment approaches may be the wave of the future, although firm empirical support is not yet in hand.

William is an 11-year-old white boy from a single-parent, lower-class family, who presented to an attention-deficit/hyperactivity disorder clinic because of mild behavior problems at home and school. Temperamentally, William was high in novelty seeking, low in harm avoidance, and very dependent on high salience rewards. In school, he was generally well liked by his teachers, who noted that he had a great deal of trouble because of poor sustained attention and impulsive behavior, especially when bored. His teacher suspected that he might have a reading disability. William's family history was positive for a past history of depression in his mother and “reading difficulties” in his father, who was in the military and lived overseas. A careful psychoeducational evaluation resulted in a variety of pedagogic (a reading tutor) and behavioral (a classroom points system) interventions. William was then started on psychostimulant therapy, which together with the classroom adaptations normalized his behavior at school. However, he continued to be stubborn and oppositional at home with his mother, who had little patience after being at work and so yelled and nattered, with no follow-through in terms of rewards or consequences. As a result, William's pediatrician recommended that the mother attend an 8-week group parent training course led by the practice psychologist that was to be held on Saturday mornings. Parent training focused on psychoeducation about attention-deficit/hyperactivity disorder and general child management skills, how to issue effective commands, and how to institute a reward schedule for good behavior and time out and other punishment procedures for failing to follow commands. Specific attention was paid to generalizing these procedures from home to other settings, such as shopping and riding in the car. Working with the school guidance counselor, William's mother also worked out a daily report card that specified daily and weekly rewards if William followed the rules at home and at school. One specific problem behavior—provoking his younger sister—required a response-cost procedure in which William lost television minutes for picking fights, but otherwise the program emphasized rewards. With this program in place, William's mother began spending more time with William. Her overall mood improved, as did William's grades at school. When summer vacation came around, William's mother was able to stop his medication entirely, relying solely on the behavioral plan to manage his attention-deficit/hyperactivity disorder.

Anxiety Disorders Recent years have witnessed increased attention to the treatment of anxiety in children and adolescents. Although well-controlled empirical studies for anxiety and depression in youth are just beginning to emerge, the literature validates the effectiveness of specific techniques and a mounting literature supports the use of prescriptive treatment protocols for these disorders. The behavioral treatment of fear and anxiety in children builds on early studies conceptualizing anxiety as a set of classically conditioned responses that can be unlearned or counterconditioned through associative pairing with anxiety-incompatible stimuli and responses. For example, in systematic desensitization, anxiety-arousing stimuli are systematically and gradually paired (imaginally or in vivo) with competing stimuli such as food, praise, imagery, or cues generated from muscular relaxation. Systematic desensitization with children consists of three basic steps: (a) training in progressive muscle relaxation, (b) ordering of fearful situations from lowest to highest, and (c) hierarchically presenting fear stimuli via imagery while the child is relaxed. Systematic desensitization appears to work well with older children and adolescents; however, younger children often have difficulty with both obtaining vivid imagery and acquiring the incompatible muscular relaxation. Strategies such as using developmentally appropriate imagery and adjunctive use of workbooks may boost the effectiveness of these procedures with younger children.

Without encouragement, anxious children and adolescents often find it difficult to remain in the presence of anxiety-arousing stimuli long enough to allow extinction in the natural environment. Following the adult treatment literature, this led to the development of exposure-based interventions for a wide range of pediatric anxiety disorders. Because escape and avoidance behaviors are negatively reinforced by cessation of anxiety, exposure-based procedures require extended presentation of fear stimuli with concurrent prevention of escape and avoidance behaviors to elicit extinction. Unlike systematic desensitization, stimulus presentation is not

accompanied by progressive muscle relaxation. Rather, anxiety to phobic stimuli is attenuated by graduated imaginal or in vivo exposure (or both) to hierarchically presented fear stimuli. Gradual exposure, with the consent of the child, is generally considered to produce less stress for the client (and therapist) and thus is often preferred over more prescriptive techniques, especially flooding.

Cognitive interventions, usually combined with exposure, also have a prominent role in cognitive-behavioral therapy for anxious children and adolescents. Based on the premise that anxious children view the world through a template of threat, automatic questioning (e.g., "What if"), and behavioral avoidance, treatment focuses on providing educational experiences to build a new coping template for the child. Therapists help the children reconceptualize anxietyprovoking situations as problems to be solved, situations to be coped with. A variety of cognitive-behavioral components help the therapist and child build the coping template: relaxation training, imagery, correcting maladaptive self-talk, problem-solving skills, and managing reinforcers. Therapists use coping modeling, role-play rehearsals, in vivo exposure, and a collaborative therapeutic relationship with the child to facilitate treatment. As a rule, parents are active collaborators in all facets of treatment.

For example, when significant others are contributing to the child's anxiety symptoms or maladaptive coping strategies, it is crucial to help the child help the parents to stop this behavior (e.g., participating in avoidance strategies or rituals). To test the hypothesis that adding a family anxiety management component would boost treatment effectiveness, Paula Barrett and her colleagues developed a family program that parallels Kendall's "Coping Cat," based on behavioral family intervention strategies effective for the treatment of externalizing disorders in youth. After each child session with the therapist, the child and parents participate in a family anxiety management session with the therapist. The program aims to empower parents and children by forming an expert team to overcome and master anxiety. Parents are trained in reinforcement strategies, with emphasis on differential reinforcement and systematically ignoring excessive complaining and anxious behavior. Contingency management strategies are the main methods for reducing conflict and increasing cooperation and communication in the family. Parents are also trained in communication and problem-solving skills so that they will function better as a team in solving future problems. Preliminary results suggest that combined treatment is more effective and has lower relapse rates than individual treatment. In much the same vein, tailored family involvement in the treatment of children and adolescents with obsessive-compulsive disorder is the rule rather than the exception. In this regard, unilateral extinction strategies (e.g., a parent returning the school phobic-child to school by force) have significant disadvantages compared with consensual child involvement: (1) parents often have no workable strategy for managing the child's distress; (2) the treatment relationship is disrupted; (3) symptoms out-of-sight for parents and teachers cannot be targeted; and (4) most importantly, these approaches fail to help the child internalize a more-skillful strategy for coping with current and future anxiety symptomatology.

> Sally is a lovely, very shy, 7-year-old white girl from a two-parent upper-middle-class family. Her mother at one time suffered from panic disorder; her father has mild social phobia. About 1 month before coming to the clinic, Sally began to experience stomachaches at school. Many other children were sick with a stomach virus at the time so Sally's symptoms didn't arouse unusual concern. After a visit to the pediatrician, which turned up nothing unusual, Sally soon went back to school. Unfortunately, although the other children were back to normal, Sally continued to have stomachaches; other sick feelings, such as dizziness, began as well. After several more days of this, she began to resist going to school, since she felt better at home. Sally's mother, who was generally sympathetic and protective of Sally, let Sally stay home. Conversely, Sally's father got rather angry when Sally wanted to stay home again and again, and over Mom's objections insisted that Sally go to school, which she did, protesting all the way. Midmorning, Sally experienced her first full panic attack, actually throwing up in class, which caused her mother to come to school to take Sally to the pediatrician, who again found nothing wrong. Sally by this time had become clingy, refused to stray far from home, and repeatedly expressed fears that something might happen to her parents, particularly her mother, who she worried might not be able to help her when she felt sick and scared. By the time she presented to a psychiatrist on the advice of her pediatrician, Sally had been out of school for 2 weeks. By this point, Sally and her family were at war over whether Sally was sick or just being oppositional. On the Multidimensional Anxiety Scale for Children Sally showed markedly elevated *t*-scores for separation anxiety along with anxious coping and somatic-autonomic symptoms. Cognitive-behavioral psychotherapy was started with the possibility of later addition of a medication if she did not respond rapidly. Treatment began with psychoeducation about separation anxiety disorder and a parental intervention to bring both parents into alignment with Sally's gradual reintroduction to school. Individual treatment began with cognitive-restructuring of faulty attributions regarding panic, harm to her parents, and vomiting fears; specific behavioral interventions targeting somatic/autonomic symptoms, including breathing retraining, relaxation training, and exposure to specific interoceptive panic triggers; and gradual reintroduction to the school setting, using a fear hiearchy that the school guidance counselor, Sally, and her mother developed together. While at home, Sally was required to follow the class schedule and to keep pace with her classmates in school work. Within 2 weeks, Sally was back in school for the morning, and full attendance was restored in 6 weeks without pharmacotherapy.

Major Depressive Disorder At any one time, approximately 1 in 20 children and adolescents suffers from major depressive disorder, and rates of depression rise dramatically in adolescents, especially in girls. While the economic burden of depression in youths is uncertain, the human cost is considerable, especially with teenage suicide. The empirical literature shows considerably more support for problem-specific psychotherapies, especially cognitive-behavioral therapy, than for medication management of pediatric depressive disorders. In particular, several controlled trials have now demonstrated that individual or group administered cognitive-behavioral psychotherapy is effective for depressed youth, and some investigators now consider cognitive-behavioral therapy the treatment of choice.

Like other cognitive-behavioral treatment packages, cognitive-behavioral therapy for depression in youth is a skills-based treatment, in this case based on the assumption that depression is either caused or maintained by inadequate social cognitive skills for coping with stress. Personality is an interactive multidirectional system of behaviors, cognitions, and emotions, and depression is manifested in each of these three components of the personality. However, cognitive-behavioral therapy for depression assumes that symptom change is most likely to occur when interventions modify patterns of behavior or cognition, with emotion following in turn. Among the behavioral and cognitive skill deficits that may characterize depressed youths are low involvement in pleasant activities, poor problem-solving and assertion skills, cognitive distortions that negatively bias perceptions, negative automatic thoughts, negative views of self and future, and failure to attribute positive outcomes to internal, stable, or global causes. Thus the therapist must establish a working alliance with adolescents and help them learn new ways of behaving or thinking that will in turn reduce depressive severity and risk of relapse.

Most cognitive-behavioral treatment packages for depressed youths share two salient characteristics: (1) both general, or required, skill-building sessions and optional, or modular, sessions for specific problems and (2) integration of parent and family sessions with individual cognitive-behavioral therapy. Treatment is generally designed to improve the teenager's problem-solving ability when faced with a stressful situation (e.g., parent-child conflict, role transitions, and grief reactions or peer problems). Thus the required aspects of treatment include psychoeducation about depression and its causes, goal setting with the adolescent, and general problem-solving skills. Modules chosen jointly by therapist and adolescent then address the specific skill deficits of the teenager. Parent-child conflict is a risk factor for depression, a poor treatment outcome, and a relapse after treatment, which justifies including a parental component in cognitive-behavioral therapy, and preliminary evidence suggests that parent and child treatment may be somewhat more effective than treatment directed at the teenager alone. Parents are taught contingency management procedures and alternative, effective methods for parenting and creating a more positive family environment. Moreover, family interactions are targeted directly to shape and reinforce effective communication and interactions and increase pleasant activities and positive affect.

Other Applications For heuristic purposes, this chapter focuses on three common applications of cognitive-behavioral techniques to childhood mental disorders. However, as summarized nicely by Euthymia Hibbs, cognitive-behavioral interventions are also effective for impulse-control disorders, such as trichotillomania, tic disorders, autistic disorder and other pervasive developmental disorders, and severe conduct disorders as well as a wide variety of common childhood behavior problems not found in DSM-IV, such as dental and medical fears.

FUTURE DIRECTIONS

Cognitive-behavioral therapy is quintessentially scientific in its theoretical foundations and pragmatic evolution. Thus it fits beautifully in the context of evidence-based medicine, which is a promising paradigm for ensuring competence in clinical practice. Because evidence-based medicine stresses systematic diagnostic assessment technologies and clinical research, it helps the individual clinician incorporate information into everyday clinical work. Compared with other psychotherapeutic approaches, current research in developing and generalizing cognitive-behavioral psychotherapies illustrates the power of using the scientific method in treatment development, namely, using a careful review of the literature to generate a hypothesis regarding a treatment efficacy or mechanism that can be tested experimentally against a control or against other active treatments. Pediatric psychiatry is now moving rapidly toward multimodal treatment research that studies not only the effectiveness of cognitive-behavioral therapy alone, but also the interactions of pharmacological and cognitive-behavioral treatments.

Despite limitations in the research literature with regard to long-term outcome; how best to combine cognitive-behavioral treatments; effectiveness across divergent cultural, age, and ethnic groupings; comparative efficacy with medications; and optimal assessment procedures, the empirical literature generally supports the benefits of short-term cognitive-behavioral psychotherapy for the major neuropsychiatric disorders in children and adolescents. However, the short-term benefits of cognitive-behavioral therapy do not necessarily translate into long-term successful outcome. In some (but not all) cases, psychosocial treatments may substantially improve long-term outcome in the most severely ill young persons when combined with pharmacotherapy. Thus, psychosocial therapies alone, or when combined with

psychotropic medication treatment, is most likely to provide sustained benefit to mentally ill children, adolescents, and adults.

The author wishes to thank John Curry, Ph.D., Karen Wells, Ph.D., and Anne Marie Albano, Ph.D., for their help.

SUGGESTED CROSS-REFERENCES

Cognitive therapy is extensively covered in Section 30.6, and behavior therapy in Section 30.2. Mood disorders are covered in Chapter 14 and Chapter 45, and anxiety disorders in Chapter 15 and Chapter 46. Attention-deficit disorders are presented in Section 39.1, and disrupted behavior disorders are discussed in Chapter 40.

SECTION REFERENCES

Abikoff H: Cognitive training in ADHD children: Less to it than meets the eye. J Learn Disabil *24:*205, 1991.

Barkley R: *Defiant Children: A Clinician's Manual for Parent Training*. Guilford, New York, 1987.

Barrett PM, Dadds MR, Rapee RM: Family treatment of childhood anxiety: A controlled trial. J Consult Clin Psychol, *64:*333, 1996.

Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J: Childhood and adolescent depression: A review of the past 10 years. Part II. J Am Acad Child Adolesc Psychiatry *35:*1575, 1996.

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, Iyengar S, Johnson BA: A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry *54:*877, 1997.

*Brent DA, Kolko DJ, Birmaher B, Baugher M, Bridge J: A clinical trial for adolescent depression: Predictors of additional treatment in the acute and follow-up phases of the trial. J Am Acad Child Adolesc Psychiatry *38:*263, 270, 1999.

*Clarke G, Lewinshoh P, Hops H: *Instructor's Manual for the Adolescent Coping with Depression Course*, ed 4. Castalia Press, Eugene, OR, 1990.

*Clarke GN, Rohde P, Lewinsohn PM, Hops H, Seeley JR: Cognitive-behavioral treatment of adolescent depression: Efficacy of acute group treatment and booster sessions. J Am Acad Child Adolesc Psychiatry *38:*272, 1999.

Conners C: *Conners' Rating Scales*. Multi-Health Systems, Toronto, CA, 1995.

DuPaul G, Stoner G: *ADHD in the Schools*. Guilford, NY, 1994.

Eisen A, Kearney C: *Treating Fear and Anxiety in Children and Adolescents*. Aronson, Northvale, NJ, 1995.

Foa E, Steketee G, Rothbaum B: Behavioral/cognitive conceptualizations of post-traumatic stress disorder. Behav Ther *20:*155, 1989.

*Hibbs E, Jensen P: *Psychosocial Treatments for Child and Adolescent Disorders*. American Psychological Press, Washington, DC, 1996.

Jensen PS: Development and implementation of multimodal and combined treatment studies in children and adolescents: NIMH perspectives. Special section: Design and methodology issues for clinical treatment trials in children and adolescents. Psychopharmacol Bull *29:*19, 1993.

Kazdin AE: A model for developing effective treatments: Progression and interplay of theory, research, and practice. J Clin Child Psychol *26:*114, 1997.

Kendall PC, Flannery-Schroeder E, Panichelli-Mindel SM, Southam-Gerow M, Henin A, Warman M: Therapy for youths with anxiety disorders: A second randomized clinical trial. J Consult Clin Psychol *65:*366, 1997.

Kovacs M: The Children's Depression Inventory (CDI). Psychopharmacol Bull *21:*995, 1985.

Lewinsohn PM, Clarke GN, Rohde P: *Psychological Approaches to the Treatment of Depression in Adolescents*. Plenum, New York, 1994.

Lochman J, Lampron L, Gemmer T, Harris S: Anger coping intervention with agressive children: A guide to implementation in school settings. In *Innovations In Clinical Practice: A Source Book*, vol 6, P Keller, S Heyman, editors. Professional Resource Exchange, Sarasota, FL, 1987.

March J: *Anxiety Disorders in Children and Adolescents*. Guilford, New York, 1995.

March J: *Manual for the Multidimensional Anxiety Scale for Children (MASC)*. MultiHealth Systems, Toronto, 1998.

March J, Amaya-Jackson L, Murry M, Schulte A: Cognitive-behavioral psychotherapy for children and adolescents with post-traumatic stress disorder following a single incident stressor. J Am Acad Child Adolesc Psychiatry *37:*585, 1998.

March J, Frances A, Kahn D, Carpenter D: Expert consensus guidelines: Treatment of obsessive-compulsive disorder. J Clin Psychiatry *58*(Suppl 4):1, 1997.

*March J, Mulle K: *OCD in Children and Adolescents: A Cognitive-Behavioral Treatment Manual*. Guilford, New York, 1998.

O'Donohue W, Kransner L: *Theories of Behavior Therapy: Exploring Behavior Change*. American Psychological Association, Washington, DC, 1998.

Pelham WE Jr, Wheeler T, Chronis A: Empirically supported psychosocial treatments for attention deficit hyperactivity disorder [in process citation]. J Clin Child Psychol *27:*190, 1998.

Sackett D, Richardson W, Rosenberg W, Haynes B: *Evidence-Based Medicine*. Churchill Livingstone, London, 1997.

*Silverman W, Kurtines W: *Anxiety and Phobic Disorders: A Pragmatic Approach*. Plenum, New York, 1986.

Van Hasselt V, Hersen M: *Handbook of Behavior Therapy and Pharmacotherapy for Children*. Longwood, Boston, 1993.

*Wells K: Adaptations for specific populations. In *Helping the Non-Compliant Child: A Clinician's Guide to Parent Training*, RL Forehand, RJ McMahon, editors. Guilford, New York, 1981.

48.4 GROUP PSYCHOTHERAPY

ALBERTO C. SERRANO, M.D.

DEFINITION

During their formative years children and adolescents spend significant amounts of time in groups, beginning as members of the family group, followed by complex group interactions with a growing number of peers and adults. While basic interpersonal skills develop in the context of the family, children learn developmentally appropriate interpersonal skills in a wide range of natural group experiences. This interpersonal learning, essential for successful social functioning, takes place to a large extent in schools and the community. These are natural group settings in which a child acquires basic cultural cues and learns ways of interacting with peers and adults, including skills for negotiating interpersonal relationships and resolving conflicts. Disturbed children and adolescents frequently have a background of faulty early group experiences, typically beginning in the family and extending to school and community. These early deficits often lead to new rejection and failure at the time when the child should be able to develop positive identifications with peers and significant others. In this context maladaptive patterns emerge that may affect all aspects of the child's development and lead to further failure.

Group psychotherapy with children and adolescents is a form of therapy in which typically between three and eight young patients of compatible developmental level meet with one or more therapists to change disturbed behavioral and psychological patterns. Considerable experience has been accumulated over the past decades in applying group psychotherapy techniques to a wide range of psychiatric disorders in this age population. Group modalities have proved efficacious and cost-effective. Unfortunately in the current climate of health care funding their use is often limited to inpatient and residential programs and time-limited groups that focus on such specific clinical conditions as eating disorders, substance abuse, sexual abuse, and children of divorce. Further, training in group psychotherapy for psychiatrists has declined with limited or no support for their being involved as therapists in groups.

This decline is associated with advances in biological psychiatry and at the expense of psychodynamic psychiatry. Fortunately treatment outcome studies are beginning to bring brain and mind back together. Growing evidence indicates that psychopharmacological interventions elicit more compliance and are more effective when used in a well-integrated, psychosocial treatment plan. Group modalities have proved effective alone and in combination with biological therapies.

HISTORY

Several names are associated with pioneering applications of groups in the treatment of children and adolescents. Jacob L. Moreno, a psychiatrist in Vienna, was interested in theater and developed group psychodrama. As early as 1911 Moreno had created a children's theater with children and adolescents. Also in 1911, Sam Slavson in New York had developed the Self-Culture Club which is the precursor of activity group therapy, a technique for which he is best known. By 1918 Alfred Adler was leading groups with children and with parents in child guidance clinics, and August Aichorn reported treating delinquent youths in groups. In the late 1920s Betty Gabriel described her technique of interview group therapy with adolescents, combining activity and group discussion. In the late 1930s Lauretta Bender incorporated the use of puppets and plastic materials in groups with hospitalized children. Also at this time John Levy and Helen Durkin applied relationship therapy principles in preschool and school age groups. In the 1940s Fritz Reidl promoted the use of groups in specialized summer camps for delinquent, early-adolescent youth. In 1961 Haim Ginott published the first textbook on group psychotherapy with children; Slavson, Saul Scheidlinger, and Mortimer Schiffer published several books on this subject during the 1970s. The American Group Psychotherapy Association published two important monographs in the 1980s, *Child Group Psychotherapy—Future Tense*, edited by Albert Reister and Irvin Kraft, and *Adolescent Group Psychotherapy*, with Fern Cramer Azima and Lewis Richmond, editors. By 1991 a specialized journal started to be published quarterly, the *Journal of Child and Adolescent Group Psychotherapy*. Paul Kymissis and David Halperin edited a textbook, *Group Therapy with Children and Adolescents*, published in 1996, that includes some of the most current thinking and applications in this field.

Theoretical Issues Group psychotherapies developed as a way to treat several patients together. Children and adolescents have been involved in therapeutic groups for several decades. Regardless of theoretical orientation, most clinicians would agree that in the hands of trained clinicians, psychotherapy groups are practical and effective agents of behavioral and psychological change. Multiple interactions among group members can be directly observed, including those that involve the therapist. Additional therapeutic factors include the development of group cohesiveness among the members, which leads to mutual support and the modeling of adaptive prosocial behaviors, which have unique value for improving interpersonal skills and reducing symptoms. Irvin Yalom's well-established review of “curative factors” was based on groups of adult patients but is basically applicable to children and youths, with the caveat that some of the factors have different significance because of developmental considerations.

While early contributors distinguished between psychotherapy groups and group applications of psychoeducational or behavioral nature, current practices underline the need for theoretical and technical eclecticism. Psychiatrists should be familiar with the use of a wide range of group approaches, including their indications and limitations in the treatment of specific clinical conditions at specific developmental levels.

Therapist While a comprehensive background in child development and psychopathology is essential, the personal characteristics of the therapist are also of crucial significance. Children and adolescents are very perceptive about adults and often suspicious of their feelings and motivation, particularly those in a clinical population. Children expect the therapist to be able to understand and contain their feelings and behaviors, even when they are not clearly expressed. A detached attitude in a therapist (which might be appropriate with an adult patient population), would typically be interpreted by children and adolescents as a lack of personal interest and evidence of rejection.

Some group therapists become particularly skilled at working with certain age groups, specific patient populations, or special techniques. Supervised training experience is essential, but certain personality characteristics and clinical experiences are needed for a good fit and effective therapeutic group engagement. Capacity for empathy, warmth, high tolerance for frustration, spontaneity, imagination, curiosity, a good sense of humor, and a playful touch are some of the qualities associated with group therapists effective at working with children and adolescents. With the growing ethnic and cultural diversity in the patient population, understanding the significance of cultural factors has become an additional challenge.

Structure The therapist is responsible for maintaining the therapeutic integrity of the group, which is predicated on having the well-defined structure essential for a group to be therapeutic. In this context the therapist has a number of important roles: selecting the membership when possible, setting limits, establishing treatment goals for each group member and changing them as indicated, and modulating the interactions among members to maximize behavioral and emotional learning.

Transferential Issues Transference and countertransference are more complex in groups than in individual therapies. These valuable therapeutic tools can be observed even in groups with a primary psychoeducational or recreational focus. The recent emphasis on short-term groups limits the opportunities to observe and analyze transferential and countertransferential issues. The multiple transferences that emerge in groups are most helpful in understanding and correcting maladaptive patterns of behavior and thinking. Whether the therapist applies a psychodynamic approach, a cognitive-behavioral approach, or psychodrama, to just name a few orientations, transferential issues are always present in patients and therapists and cannot be ignored. A child whose provocative behavior in group seems to be associated with earlier experiences of anticipated rejection is one situation in which a competent clinician, regardless of theoretical orientation, could

make an intervention of therapeutic value for that child and the other group members. The ability of the therapist to contain the anxieties of the group is closely associated with awareness of countertransference, particularly during those periods of emotional and behavioral upheaval in which one notices the difference in how experienced and novice group therapists intervene to set limits and make the experience a therapeutic event.

TECHNIQUES

Being a member of a group is a social phenomenon with roots in the natural world. Groups are essential for the survival of animals and humans alike. Learning to be an accepted member of one's culture starts with the family and continues in the larger social context. Disturbed emotions and behaviors are most often manifested in the interpersonal matrix of family, school, and community. Therapeutic groups used either as a single modality or in combination with other treatment approaches can be effective, but careful selection is needed in matching a child with a group.

Placing children in groups at their developmental levels is a well-established practice, but even more important is the suitability of a specific group for the child. Levels of physical, cognitive, emotional, and social development need to be considered in the selection process to maximize the therapeutic value for both the patient and group members. Will this group be “good” for this patient? And conversely, will this patient be a “good” addition to the group? Finding the proper fit goes beyond matching considerations. In practice and for practical reasons, many of these principles are often overlooked. Most inpatient, partial hospitalization, and residential treatment programs for children and adolescents have a wide range of group applications in their treatment repertoire. Children and adolescents admitted to those institutions are expected to participate in their group programs. Assignment to specific groups may not be preceded by careful selection and preparation in the manner discussed above. Institutions that recognize the value of group therapy have developed well-defined standards for its practice and offer ongoing training and supervision. Unfortunately in too many institutions the clinical staff leading groups has limited training and supervision. The therapeutic value of group therapy is seriously compromised in those programs.

PRESCHOOL AGE GROUPS

Developmental Considerations Children ages 3 to 5 are exposed to an active phase of social exploration. During their early years children observe and experience intimate relationships with mother, father, and siblings as well as extended family and significant others. They now move rapidly into a larger social context in which intense social learning takes place by observing and relating to peers and teachers. From initial imitative behaviors, children soon develop new rules of behavior and new judgment standards. The development of new cognitive and social skills facilitates moving beyond egocentricity. At this time preschool experiences expose children to multiple experiences of competition, jealousy, happiness, sadness, guilt, success, failure, anger, fear, and admiration, to name some of the most common. Growing mastery begins to develop along with better modulation of emotions and behaviors, even when the child's cognitive development has not reached the capacity for organizing and reflecting that will be achieved during the latency years.

Techniques Groups for preschool children are typically heterogeneous and include children with different diagnosis, in contrast to the homogeneous groups in which children with a common problem (e.g., a chronic illness or parental divorce) share a psychoeducational and supportive group experience. Preschool groups are structured around toys, play, artwork and other age-appropriate activities. These groups typically include only three to five same-sex children. Because of the children's short attention span, time and space are important considerations, particularly with hyperactive, overanxious, and aggressive children. Groups should not last more than 45 to 60 minutes, including time for a snack. A medium-sized room with appropriate furniture for the age will help provide good boundaries and a therapeutic climate. It is recommended that play and artwork materials be stored away until group time. The use of multiple therapists, highly recommended for child and adolescent groups, is particularly indicated in the treatment of emotionally disturbed preschoolers. These children have acquired at best a basic repertoire of social skills, while emotional regulation is just beginning. As they do in individual therapy, children in a therapy group reenact their behavioral and emotional conflicts, which can become a challenging management problem. Two therapists are needed to maintain clear and firm limits in a consistent structure and model prosocial behaviors and the development of new masteries. Several authors recommend a male-female cotherapy team as the ideal therapeutic configuration. This is not always possible, and cotherapy teams may be of the same sex, including a senior clinician with a trainee or two trainees, and frequently from different disciplines. Young children's limited verbal and social skills, emotional lability, impulsiveness, and out-of-control behaviors can be most challenging, even to the experienced clinician. Availability of consultation and supervision is essential for a successful group therapy program.

With younger children therapists typically initiate each session's activities around educational and therapeutic themes. Using a warm, supportive approach therapists engage children in play activities with puppets, toys, and art materials. The therapist's knowledge of the patients' individual and family histories permits using the group context for further identification and understanding of children's conflicted behaviors and interactions. Mutual-story-telling techniques can provide the children with corrective models that introduce constructive, prosocial options and new masteries. Emphasis is placed on understanding the children's disturbed behaviors in light of their individual and family histories. Children soon reenact earlier experiences in the group, mostly through play and interpersonal behaviors. Interpretations should be kept at a minimum. Therapist interventions are aimed at providing corrective emotional experiences that help the child reach greater social and emotional mastery. Most frequently these groups are offered in combination with family therapy or parents' groups that tend to be mostly supportive and psychoeducational. Close participation of the family in children's groups greatly facilitates the therapeutic engagement and progress. It also empowers the parents by helping them become more competent in their role.

LATENCY AGE GROUPS

Developmental Issues Latency evolves over several years and can be divided into early (5 to 7); middle (8 to 10) and late childhood (10 to 12). During the early period children demonstrate advances in social skills with improved self-control and better behavioral and emotional modulation. The beginning of a moral conscience can be noticed as the child reaches the stage of concrete operations. Further maturation of the child's nervous system proceeds rapidly in areas of motor, intellectual, emotional, and sensory development. Physical action and competitive games and sports become of growing interest, particularly for boys. Higher mastery of verbal skills, better capacity for concentration, and tolerance for frustration make older children more interested in sports and team activities. Development of friendships and participation in group activities becomes of growing interest in older children and continues into adolescence.

Techniques Latency age groups require a larger room. These groups typically include from six to eight members of the same sex. A larger space is needed to engage them in games and physical activities involving group play and discussion. These groups last from 60 to 90 minutes. The combination of activity with interview techniques, evolved from Slavson's early contributions, incorporates psychodynamic and behavioral concepts and techniques.

Children who have reached latency age are more able to connect with each other, but initially the therapist must actively foster a climate of safety and trust before they can engage effectively as group members. Activities should be introduced by the therapist with the intent of facilitating creative and playful interactions that catch the children's interest. Observation of the children in group also clarifies diagnostic questions. Limit setting is an important consideration, particularly with aggressive children with poor control. Some of the more typical activities include games, storytelling, creating a mural, dramatizations, use of puppets, masks, and physical activities. These activities must be structured to allow enough time for discussing the experiences to connect the group members. As the group becomes more cohesive members can be allowed to initiate games and activities with the therapist intervening to set limits to prevent or to deescalate dangerous situations. Interventions also include interpretations and clarifications that help the group become a therapeutic agent. Going around and group discussions are valuable techniques to further facilitate connecting group issues with individual, school, and family problems. When a group has matured to developing its own culture, limits are set by the group itself. This level is more typical of longer-term inpatient and residential facilities, but it can be achieved in outpatient groups.

> Child psychiatry residents who started a latency-age group with eight boys reported the struggle they had establishing rules for the group, considering the disruptive behaviors of several members. The group met for 20 minutes to discuss and plan the activities of the day followed by 20 minutes of participation in the selected activity. The last 20 minutes were devoted to discussion of the process while they had a snack. The cotherapists reported initially that they complemented each other well. One took a more active direct role and was a limit setter; the other was more passive and willing to trust the group process. After several weeks the more-active therapist discussed in supervision his frustration with his role and a growing resentment toward his cotherapist. His background as a school teacher was compared with his new role as therapist. His cotherapist recognized the need for more active involvement.
> Out-of-control behaviors diminished after the cotherapists became more aware of how their behaviors affected the group process and brought better balance to their leadership styles. The therapeutic value of the group experience emerged even more clearly when parents and school reported significantly improved behaviors in most of the children, even when their level of insight was limited.

EARLY ADOLESCENT GROUPS

Developmental Issues Early adolescence applies to youths between about 13 and 15 years of age who are usually in grades 6 to 8. It is associated with major

physical, psychological, emotional, social, and cultural changes, often taking place unevenly. The onset of puberty, one of its markers, typically occurs 1 to 2 years earlier for girls than for boys. Physical maturation, a second phase of separation and individuation, reaching cognitive capacity for “formal operations,” and developing strong attachments to peers are some of the most striking tasks of this period. Friendships involve mostly children of the same sex. Relationships with parents and adults become strained and distant as the adolescent struggles with dependency issues. While most adolescents do not seem to experience excessive turmoil, the clinical population under consideration usually presents evidence of conflict in more than one area. Early adolescents often behave defensively with their parents and most adults. A therapeutic alliance with them is difficult to achieve and very fragile. Early adolescents are often on an emotional roller coaster, feeling very insecure and vulnerable to criticism. Denial, projection, and intellectualization are the typical defense mechanisms. Early-maturing boys and late-maturing girls seem to have some advantages over their contemporaries in coping with the challenges of going from childhood into adolescence. A loving, intact family and having achieved reasonable levels of competence during previous developmental phases are also considered protective factors.

Techniques Early adolescents are often more willing to participate in a group of same-sex peers than in individual psychotherapy. On the other hand while they may be willing to discuss problems concerning relationships with family and peers, they have only a limited capacity to contain and modulate emotions, both their own and those of others. To maintain an optimal structure, group size should be limited to eight. Early adolescents often become silly and rowdy when anxiety-laden issues come into focus. Therapists need to keep a good structure and firm limits before the group becomes an effective therapeutic tool.

Establishing a well-defined, clear structure is always essential and is even more challenging with early adolescents. They want to retain the advantages of younger children and simultaneously insist on receiving adult privileges. Simple rules must be established from the outset concerning proper behavior, taking turns to listen to each other, and respect for confidentiality. Other issues that frequently need to be addressed are socialization outside the group and the use of food during group meetings. Most experienced clinicians agree that fewer rules with some flexibility for creative options that can be initiated and monitored by the group can have excellent therapeutic value. Considering a visit to a hospitalized member or planning a special celebration are two examples of projects in which group members would participate in all aspects of development and implementation. How much autonomy and responsibility can be delegated to the group is determined by clinical considerations by the therapist who is ultimately in charge.

A valuable tool in working with adolescents is the careful use of humor in a nonthreatening way. At this stage of development their humor is often scatological and crude, but they feel very vulnerable to criticism. It is an art to learn to accept and contain early-adolescent expressions of humor without reciprocating or reinforcing them. Reframing or redirecting with a light touch of humor or an interpretation can focus the group into examining issues they may be avoiding. A valuable use of humor can be illustrated by the therapist sharing an embarrassing experience during adolescence. Verbal games are also valuable tools, as are role playing and psychodramatic techniques. As with younger children, group techniques are only means to help achieve therapeutic goals. Clinical considerations should determine when and how they are prescribed.

Johnny was a high-functioning, 14-year-old boy diagnosed with autistic disorder. He had been in individual and family therapy for several months before he was considered ready for group therapy. Johnny was an awkward-looking adolescent who looked and acted younger than his chronological age. His academic level was above average, but his social development was very limited. A supercilious, hypermoralistic attitude of more recent development contributed considerably to his social isolation, particularly after starting seventh grade. He was assigned to an established group of early adolescents with a mixture of clinical conditions, meeting once weekly for 75 minutes. Initially Johnny limited his participation to monosyllabic answers to direct questions, then he would go back to reading a book on the history of Napoleon, his favorite subject and object of fascination. Group members chose to ignore him after a while. Over a period of several weeks his interest in the book seemed to abate. Johnny brought it, but it remained unopened on his lap. He would make an occasional remark, mostly to criticize another group member for his “vulgarity.” The group laughed at his remarks but scapegoating could be avoided. They seemed to respect his “differentness.” Two months later, Peter, a very shy schizoid 13-year-old boy joined the group. After a few sessions Johnny had developed an unexpected interest in Peter and sat by him and encouraged him to interact with the group. Soon it was noticed that Johnny was not bringing a book any longer and that he was more actively involved with group members. He responded to social cues in a more age-typical and appropriate manner, and though he continued having morbid preoccupations with power and a fascination with Napoleon, the intensity was considerably reduced. Johnny's growing interest in people was clinically evident. Group therapy was used in combination with individual and family therapy and psychotropic medication over 18 months. Although the group experience was only one component of the treatment plan, it became a most significant tool to help Johnny with his interpersonal deficits.

MIDDLE AND LATE ADOLESCENT GROUPS

Developmental Issues Middle adolescents are approximately 15 to 17 years of age, usually in high school in ninth, tenth, and eleventh grades. Late adolescents are ages 17 to 19, typically at the completion of high school and the transition into adulthood. Middle adolescence is punctuated by a more stable involvement with peers and the development of relationships with the opposite sex. Greater autonomy from the parents also emerges, along with a clearer sense of identity and growing vocational interests.

As they move into late adolescence, commitments to further education, training, work and other people become clearer. A more solid sense of self emerges, and the earlier critical attitude about the parents becomes less intense.

Techniques A mixed group of boys and girls with a cotherapy team representing both sexes provides an ideal format. In practice this balance cannot always be achieved. These groups can be larger, 8 to 12, depending on clinical considerations and fit. Older adolescents are capable of higher levels of abstraction, introspection, and emotional and behavioral control. They can typically discuss openly topics of personal and mutual interest. The therapist initially invites each member to express the reasons for being in group and to give a brief background. Going-around, group discussions often facilitate active interaction. In contrast with leading younger age groups, older adolescents typically enjoy the opportunity to discuss feelings, opinions, and concerns that would be harder to elicit in individual therapy. Because of their tendency to intellectualize or avoid dealing with personal issues by focusing on someone else's problems, the therapist should be able to redirect the focus by reflecting, interpreting, and going around, asking each individual's opinions and feelings. Group cohesiveness often develops quickly, and members discover that many concerns that had been a source of worry and shame are not unique. Catharsis is important, particularly in the beginning stages. The therapist's empathetic interventions, support, confrontation, and careful use of interpretations are essential for developing group cohesion. Soon group members can interact with each other more openly and enter a new phase of group work. The therapist's role is maintaining the therapeutic focus with good boundaries. Intense relationships develop in group, and members often want to socialize between sessions or start dating each other. Group members need to be reminded that outside contacts and pairing off affects the integrity of the group, because those members are no longer able to be as honest and open with the rest of the group as in the past. Group members often underestimate the significance of breaking this rule and may blame the therapists, projecting on them the image of overrestrictive parents. Further elaboration in group typically brings new appreciation for the value of this rule and a stronger awareness of how group psychotherapy works. Often the infractors ask the group to allow them to continue as members and promise to stop contact outside the group. Confidentiality is another major issue for careful review. Patients must understand that confidentiality should be respected at all times except when the patient's or someone else's life is at imminent risk. Otherwise, abstinence in discussing any information concerning individual members outside the group is essential for the safety of the group as a therapeutic environment.

PARENT GROUPS

Considerable literature supports the effectiveness of parent groups, particularly those with a psychoeducational supportive orientation with emphasis on learning behavior-management techniques. They are well structured and didactic and are offered in a course format over a limited number of classes which typically include time for questions and less frequently for group discussion.

Psychoeducational formats have become popular in the management of specific conditions such as attention-deficit/hyperactivity and conduct disorders, but are also indicated for parents of children and adolescents with posttraumatic stress disorder, learning disabilities, retardation, schizophrenia, bipolar disorders, Tourette's disorder, autistic and pervasive developmental disorders, eating disorders, substance dependence, and juvenile delinquency, of adolescent sexual offenders; and of children or adolescents with a wide variety of chronic medical conditions such as cystic fibrosis, cancer, diabetes, and various types of transplantation. Hardly a condition exists that lacks an educational/support group. However, self-help groups that are purely educational and led by nonclinicians differ from those whose group leader is a clinician. The latter would be more inclined to use group dynamic principles in their approach. Going beyond the instructional goal allows more-personalized recognition of individual and family issues, particularly those concerning the relevance of expressed emotion in families and its impact on relapse prevention. When the focus shifts from behavioral management of a diagnosed child to exploration of parental emotions and thoughts in response to the child's condition, the intervention becomes more psychotherapeutic. Parent's groups and multifamily groups are valuable tools that have proved effective, but unfortunately they only are widely used in specialized centers.

There are several advantages to conducting parental guidance and support groups. The more-experienced parents can serve as models for resistive parents and those of newly diagnosed children, which benefits both “senior” and “junior” members. An insight-oriented approach is not indicated when the need is providing basic mastery of effective parenting to reduce the sense of guilt or failure that these parents often experience. Parents who are psychotic, paranoid, narcissistic, or extremely rigid and controlling cannot function effectively in a group and should be offered a more traditional individual or family approach. Parallel parents groups have been popular in child guidance centers, particularly with children up to the young adolescent age. Close collaboration between the therapists of the parallel groups often includes a monthly joint session with children and parents, resembling the multiple family groups but not including siblings or other family members.

CLINICAL ISSUES

Indication In the context of a psychiatric evaluation, assessment of the child and family should identify areas of strength and weakness, including protective factors and vulnerabilities. The formulation and treatment plan should be comprehensive enough to include consideration for a group psychotherapy approach in combination with the possible use of psychotropic medication and individual or family techniques or both. Residential care and day program group treatments are often the primary modality for hospitalized children and adolescents, with several different group experiences offered daily. Individual and family interventions, while significant, are offered less frequently. In contrast outpatient groups are more frequently used as adjunctive treatment to deal with specific diagnostic conditions such as attention-deficit/hyperactivity disorder, conduct disorders, victims of abuse, posttraumatic stress disorder, postdivorce adjustment, and adolescent sexual offenders. Groups using cognitive-behavioral therapy have proved effective in the treatment of conduct disorders and depression. The use of cognitive-behavioral therapy is expanding rapidly, and it has become one of the most popular and best-researched psychotherapy approaches.

Deficit or impaired social skills are the most frequent indications for group therapy; they include many clinical conditions (e.g., schizoid personalities, conduct disorders, sexual disorders, social phobias, physical and sexual abuse, and depression).

Limitations As a single treatment modality group psychotherapy is of limited use in children except when combined with other therapeutic interventions. Older adolescents are a possible exception. Late adolescent groups are used effectively as a single treatment modality in counseling centers and colleges for a number of the clinical conditions discussed above.

Complications Problems of breach of confidentiality and of dating or socializing outside the group have been discussed. Additional problems are associated with out-of-control or psychotic behaviors that become too disruptive for adequate group function. Inducing delinquent behaviors or using drugs are reasons for exclusion from the group. With careful assessment these problems are often anticipated before placing a patient in group. Therapeutic termination from the group is at times painful, but needed, when the patient cannot be contained by the group and attempts at corrective intervention prove ineffective. Several authors have discussed strategic exits from group with a more individualized approach as an alternative. Another complication involves a possible cultural mismatch when a member of a cultural or religious minority finds himself or herself isolated and without basic social cues to overcome the barriers. Again, careful diagnostic assessment should anticipate this complication. Adding another member of the same minority as a member or a therapist can turn around this complication. Consultation and supervision are most helpful at preventing these complications and facilitating alternative therapeutic options.

Contraindications Children and adolescents who are extremely anxious, actively psychotic, and disorganized cannot benefit from group therapy before they are stabilized with medication and individual and family techniques. Sociopathic, narcissistic patients with potential for predatory behaviors could make destructive use of a group and should not be included. In juvenile correctional institutions these patients can make good use of group therapy; typically, psychoeducational and behavioral interventions are effective with incarcerated youth.

Goals of Treatment As a single modality or most frequently in combination with other therapeutic interventions, treatment goals should be clearly defined and specifically related to each modality. Typical goals for a group intervention include improved interpersonal and social skills and increased trust in others, empathic sensitivity, and ability to recognize and express feelings in self and others. The case of Johnny (described above) included all of those goals. When combined therapies are used including involvement of two or more clinicians, the treatment plan must include coordination of various modalities, with the treatment team maintaining open lines of communication. Consultation and supervision are particularly valuable in those cases to maximize their therapeutic value and prevent complications. In Johnny's case, the frequency of family sessions was reduced from weekly to twice monthly over a period of several months, followed by once-monthly sessions; weekly group and individual therapies continued and for the final 6 months, group was the primary modality, with the mother in a parallel group. Infrequent family and individual sessions were scheduled periodically to assess and integrate the treatment.

Homogeneous and Heterogenous Groups Early authors recommended that groups include children with mixed clinical conditions. They described the advantage of combining, for example shy, inhibited children with more aggressive or expressive youth. Opportunities for mutual learning can thus be complemented, with a wider range of therapeutic opportunities available for group interaction and for learning. Psychodynamic, activity-discussion groups that include children and youths with mixed clinical conditions follow this orientation.

More recently the use of homogeneous groups has grown in popularity. Patients with diagnoses in a similar category are placed together in homogeneous groups for such conditions as eating disorders, attention-deficit/hyperactivity disorder, posttraumatic stress disorder, and substance dependence. These groups develop strong cohesiveness faster than heterogeneous groups, but they usually do not reach the more intense levels of clinically mixed groups. They have a more psychoeducational design, with a primary focus on behavioral changes. Homogeneous groups have proved effective and are often used on a time-limited basis. Self-help groups are also homogeneous but are led by nonclinicians; they provide excellent support and have an educational focus. They are used less frequently with children and adolescents, with the exception of Alateen groups for children of alcoholics.

SETTING

Outpatient Child and adolescent groups mostly occur in child guidance clinics, community mental health centers, and large clinics. Individual practitioners in private practice find it difficult to gather enough suitable patients with compatible clinical needs to organize groups, particularly because involvement of parents and collateral therapeutic interventions is required. Group practices are more able to organize groups. Space, volume of patients, opportunity for cross referrals, availability of clinicians for cotherapy, consultation, and supervision make it possible. Student health centers frequently use groups for adolescents. Group approaches are also popular in school-based and day programs for emotionally disturbed children and adolescents. Specialized group programs have recent popularity in the outpatient treatment of single clinical conditions.

Inpatient Current practices of inpatient and residential psychiatric care for children and adolescents include a strong group therapy program. Because the length of stay in most institutions has become very short, most payers and institutions emphasize the use of groups 7 days a week, often including several different types of group during the same day, depending on the diagnostic problems and the level of care. Multiple members of the staff or trainees may be involved in one or more group modalities depending on their training and experience. In addition to psychodynamic groups, most inpatient, partial hospitalization, and residential care institutions currently offer a menu of groups to select from. The list includes, but is not limited to, mixed-gender, single-gender, anger management, relaxation and stress management, substance dependence, social skills training, acquired-immune deficiency syndrome (AIDS) education, recreational, art, psychodrama, movement and dance groups, and multiple family groups. Team meetings and a clinician case manager facilitate the integration of these modalities into a treatment plan along with individual, family, and psychopharmacologic interventions.

Schools Counseling groups have been popular in schools. They are typically led by school counselors, with psychiatrists frequently as consultants, coleaders, or supervisors. The groups are held during the school day. The range of options is narrower because of time limitations. More intense, focused group experiences are indicated when a traumatic event affects a school and its community such as a recent suicide of a student, a shooting, or a serious accident on school grounds.

ETHICAL ISSUES

Confidentiality Confidentiality is usually not a significant issue with younger children. Preschoolers and young elementary age children expect parents to have close involvement and to know just about everything about them. A rule of confidentiality concerning group activities could be misinterpreted as a request for secrecy by young children and their parents. Older children should understand the importance of keeping group information in the group, particularly information about members and group material to be kept private. Sharing a personal experience or insight with parents is acceptable, but disclosing someone else's is not. Inevitably children sooner or later run into another group member in school, church, or the neighborhood. They should be cautioned to avoid discussing group experiences, particularly when around other children. Sharing information with parents is another ethical issue, particularly with adolescents. Collaboration with the parents is needed to

maximize the therapeutic benefits. What information should be shared and how is decided on the basis of clinical considerations. Certainly when the child is at risk of harming self or others, the rule of confidentiality does not apply. In all other cases it is best to discuss with the child or adolescent the therapeutic value of bringing a particular piece of information in the open to be shared with their parents. A family session, regularly or specially scheduled, is often the best context for such disclosure which may open or reopen channels of communication between parents and child. Contacts with parents are often regarded with suspicion. Clinical judgment should determine how much of those contacts is discussed in group. Parents are also entitled to confidentiality and should be free to discuss personal issues with the therapist.

A breach of confidentiality by a group member is a serious matter that puts the integrity of the therapist and the group on the line. It must be discussed in group in an open and straightforward way to have trust in the value and safety of the group restored. Risk-taking behaviors including promiscuous sexual activities, use of drugs, abuse of alcohol, and involvement in delinquent behaviors cannot be condoned. Clinical judgment should determine where experimentation ends and self-destruction starts. A mature, caring group often can recognize the problem and put group pressure on that individual to face the problem openly. Disclosure to the parents and the need for additional help are central considerations. Ignoring self-destructive behaviors would be regarded by the group as the therapist condoning and even encouraging them. The group would no longer feel as safe, and the therapist risks losing the trust of its members.

RESEARCH AND EVALUATION

Fewer studies have been conducted on group therapy with children and adolescents than with adult populations. Several explanations for this have been offered, including problems in patient recruitment, ethical issues, funding, family involvement, developmental factors, compliance, and motivation. A suitable pool of patients may be available only in larger institutions. There is a stronger tradition for psychotherapy research with adults, and the number of researchers specialized in the field of child and adolescent group psychotherapy is rather small.

Short-term groups and cognitive behavioral therapies have been better studied than psychodynamic groups Therapists using manualized techniques in university settings have reported better outcomes than therapists in field studies. Despite some of those limitations, outcome studies consistently demonstrate that group psychotherapy with children and adolescents is as effective as individual therapy, and both are significantly more effective than control conditions. Only 20 of approximately 500 studies involved child or adolescent groups, but their effect size was not analyzed separately. Many of the studies examined different areas of outcome, process, and leadership. Unfortunately, standardized research instruments were used in only a few of them, and most groups did not have control or contrast groups. Follow-up studies are limited or absent.

Time-limited group psychotherapy as applied to a number of clinical conditions has received more interest. Eating disorders, including anorexia and bulimia, have responded well to group formats that combine cognitive-behavioral, psychoeducational, and psychodynamic theories and techniques, alone or in combination with psychotropic medication. The use of group therapy modalities in the treatment of chemical-dependent youth has been the subject of several excellent studies.

Unfortunately there has been great diversity in the treatment designs, and comparing studies is problematic. Most treatment approaches for this population predict better outcome when families participate actively in treatment. In general, well-structured, problem-oriented, psychoeducational groups are more effective and have more practical value than psychodynamically oriented, unstructured groups.

Groups for sexually abused children have been an established practice; however, few studies demonstrate their efficacy. Some have victims alone, others have a parallel group for parents, and some combine group with family sessions. Authors conclude that children become less anxious, manifest improved self-esteem, show less suicidal and self-mutilating behavior, are more able to trust and show anger, and present less-sexualized behaviors. Adolescent sexual offenders have also benefited from group therapy approaches. Because treatment takes place most often under court order and conditions of probation, compliance is high. They are problem oriented, including among other goals that the victimizer admit responsibility for the offenses and over time develop empathy for the victim, followed by a formal apology. Social skills training and school and vocational support are also part of the treatment of sexual offenders. Psychodramatic applications and role playing techniques are valuable with this population. The few published studies demonstrate that group techniques are significantly effective, and the “admitters” show a marked reduction in new sexual offenses over a period of 2 years.

TRAINING

While psychiatry and child and adolescence residency programs are required to offer training in group psychotherapy, this training often takes a back seat to other treatment practices. Further, the number of well-trained psychiatric faculty for teaching and supervision of group psychotherapy, particularly child and adolescent psychiatry, is limited. Frequently training is delegated to nonpsychiatrists. Fortunately most training programs expose residents to a variety of group experiences in outpatient, inpatient, and residential settings and schools. Trainees typically participate as coleaders but only rarely over an extended period of time, mostly because of rotation requirements. It is best to have several well-supervised group experiences to become familiar with the power of group dynamics and the value of group psychotherapy in the treatment of children and adolescents, even though opportunities for personal involvement may be limited following residency training.

SUGGESTED CROSS-REFERENCES

Information related to child and adolescent group psychotherapy may be found in Section 48.4 on group psychotherapy, combined individual and group psychotherapy, and psychodrama; in Section 30.5 on family therapy; in Section 30.11 on evaluation of psychotherapy; in other sections of Chapter 48; and in Section 49.13 and Section 54.2 on ethics in psychiatry. Normal child development is discussed in Section 32.2 and normal adolescent development in Section 32.3.

SECTION REFERENCES

Azima FJC, Richmond LH: *Adolescent Group Psychotherapy*. International Universities Press, Madison, CT, 1989.

Bernet W: The technique of verbal games in group therapy with early adolescents. J Am Acad Child Psychiatry *21:*496, 1982.

Bierman KL, Furman W: The effects of social skills training and peer involvement in the social adjustment of preadolescents. Child Dev *55:*151, 1984.

Brandes NS, Gardner ML: *Group Therapy for the Adolescent*. Aronson, New York, 1973.

*Borduin CM: Multi systemic treatment of criminality and violence in adolescents. J Am Acad Child Adolesc Psychiatry *38:*242, 1999.

Bromfield R, Pfeifer G: Combining group and individual psychotherapy: Impact on the individual treatment experience. J Am Acad Child Adolesc Psychiatry *27:*220, 1988.

*Brook D: Adolescents who abuse substances. In *Group Therapy with Children and Adolescents*, P Kymissis, D. Halperin, editors. American Psychiatric Press, Washington, DC, 1996.

*Clarke GN, Rhode P, Lewinshen PM: Cognitive-behavioral treatment of adolescent depression efficacy of acute group treatment and booster sessions. J Am Acad Child Adolesc Psychiatry *38:*272, 1999.

*Corder B, Whiteside L, Haizlip T: A study of curative factors in group psychotherapy with adolescents. Int J Group Psychotherapy *31:*345, 1981.

Ferrara ML: *Group Counseling with Juvenile Delinquents: The Limit and Lead Approach*. Sage Human Services Guides, Newbury Park, CA, 1993.

Gabriel B: An experiment in group therapy. Am J Orthopsychiatry *9:*146, 1939.

Ginott H: *Group Psychotherapy with Children: The Theory and Practice of Play Therapy*. McGraw-Hill, New York, 1961.

Golden J, Damon M, Castaldo L: *Group Treatment for Sexually Abused Children*. Guilford, New York, 1989.

Jenkins J, Karuo M: The meaning of expressed emotion: Theoretic issues raised by cross-cultural research. Am J Psychiatry *141:*9, 1992.

Kaplan HI, Sadock BJ, editors: *Comprehensive Group Psychotherapy*, ed 3. Williams & Wilkins, Baltimore, 1993.

*Kymissis P, Halperin D, editors: *Group Therapy with Children and Adolescents*. American Psychiatric Press, Washington, DC, 1996.

MacLennan BW, Felsenfeld N: *Group Counseling and Psychotherapy with Adolescents*. Columbia University Press, New York, 1968.

Meeks J, Bernet W: Group psychotherapy of the adolescent. In *The Fragile Alliance*, ed 4, J Meeks, W Bernet, editors. Krieger, Malabar, FL, 1990.

Mooney S, Schamess G: Focused time limited interactive group psychotherapy with latency age children: Theory and practice. J Child Adolesc Group Ther *1:*107, 1991.

Moreno JL: *Theatre of Spontaneity: An Introduction to Psychodrama*. Beacon Press, New York, 1973.

Rachman AW: *Identity Group Psychotherapy with Adolescents*. C Thomas, Springfield, IL, 1975.

Redl F: Group emotion and leadership. Psychiatry *4:*573, 1942.

Riester A: Creating the adolescent group psychotherapy experience. In *Group Therapy in Clinical Practice*, A Alonso, H Swiller, editors. American Psychiatric Press, Washington, DC, 1993.

*Riester AE, Kraft IA, editors: *Child Group Psychotherapy: Future Tense*. International Universities Press, Madison, CT, 1989.

Riester AE, Tanner DL: Group counseling: Follow-up viewpoints. Elementary School Guidance Counsel *14:*222, 1980.

Scheidlinger S: Short-term group psychotherapy for children: An overview. Int J Group Psychother *34:*573, 1984.

Schiffer M: *Children's Group Therapy: Methods and Case Histories*. Free Press, New York, 1984.

*Serrano A, Hou S: Culture and ethnicity. In *Group Therapy with Children and Adolescents*, P Kymissis, D Halperin, editors. American Psychiatric Press, Washington, DC, 1996.

Slavson SR: *Child-Centered Group Guidance of Parents*. International Universities Press, New York, 1958.

Slavson SR, Schiffer M: *Group Psychotherapies for Children: A Textbook*. International Universities Press, New York, 1975.

Sugar M: *The Adolescent in Group and Family Therapy*. Brunner/Mazel, New York, 1975.

*Sugar M: Research in child and adolescent group psychotherapy. J Child Adolesc Group Ther *3:*207, 1993.

Tramontana MG: Critical review of research on psychotherapy outcome with adolescents: 1967–1977. Psychol Bull *88:*429, 1980.

Van Schoor E, Schmidt K, Ghurman H: Group psychotherapy with children and adolescents. In *Handbook of Child and Adolescent Outpatient, Day Treatment and Community Psychiatry*, H Ghurman, R Sarles editors. Brunner/Mazel, Philadelphia, 1998.

Yalom ID: *The Theory and Practice of Group Psychotherapy*. Basic Books, New York, 1970.

48.5 FAMILY THERAPY

ALLAN M. JOSEPHSON, M.D.

DEFINITION

Family therapy is any intervention that focuses on altering the interactions among family members and attempts to improve the functioning of the family as a unit of individual members of the family. The clinician conducting family therapy attempts to interrupt rigid intergenerational patterns that cause distress within or between individuals. Family therapy can address the concerns of any family member, yet it is most likely to influence children, whose daily reality is directly affected by family context.

The term *family therapy* is typically used in two contexts: to describe an intervention modality, the focus of this chapter, and as a conceptual framework for intervention. To complicate matters, there is no one type of family therapy but numerous schools of family therapy subsumed under the generic term.

As opposed to *family therapy*, the denotation “family treatment,” may more accurately reflect the numerous ways in which clinicians can intervene with families—with the entire system, with dyadic subgroups, and with individual members. For a generation, the systems view (i.e., all family members must be present for treatment) has been synonymous with family therapy. Eclectic family treatment encompasses behavioral, psychological, and educational interventions with families who have a child or adolescent with a psychiatric disorder. *Family therapy* and *family treatment* are used interchangeably in this chapter.

While most child and adolescent psychiatrists believe that families must be involved in clinical interventions with children and adolescents, views on the family's role in the evolution of a child's psychopathology differ. Families can be categorized as promoting individual disorders, as responding to disorders, or as a system in which an individual disorder indicates family dysfunction. In a biopsychosocial, interactionist model, family interventions often involve a combination of these perspectives. The cumulative impact of family interaction can worsen or ameliorate a child's symptoms.

Family therapy can include concepts familiar to individual therapists, such as developing an alliance and working through and dealing with resistance. Family therapy also uses a behavioral emphasis. Clinicians observe specific behaviors and design plans to alter them by intervening in interactional patterns associated with these behaviors. Family therapy shares characteristics of group therapy. Pragmatically, more than one individual is receiving clinical attention. Conceptually, individuals are viewed as performing functions for the group as a whole. In family therapy, however, roles and hierarchies do not solely evolve out of group interactions but are also assigned by biological relationship of parent to child and sibling to sibling. Further, a shared history of participants exists that is more intense and influential than that experienced by group therapy members. Finally, because family interactions observed in therapy reflect enduring patterns outside the therapy experience, they indicate a pervasive effect on the internalized world of its members in contrast to the more limited effects of group therapy interactions on its members.

HISTORY

Most observers agree that the roots of family therapy originated in the early 1900s when child guidance clinics emphasized that the problems of children were embedded in a family context. The orientation of these clinics was characteristic of the time, a depth psychology approach informed by psychoanalysis. While Sigmund Freud had pointed out the importance of family forces on intrapsychic development, the primary clinical intervention typically focused on the mind of the child. The psychiatrist treated the child, and the parents were seen by the social worker. This division of labor reflected the primacy of the individual approach, a situation that continued through the first half of the century. Even as family treatments ascended, most psychiatrists saw family interviews as ancillary to the treatment of the child's internal conflicts.

Pioneers Nathan Ackerman, arguably the father of family therapy, typified the early family therapy leaders. Many of these clinicians had psychoanalytic backgrounds yet were increasingly dissatisfied with individual treatments. They either believed the treatments ineffective or observed effective treatments undermined by families who were having difficulty adjusting to therapeutically induced changes in their children. Many of these individuals worked alone, outside mainstream psychiatry. These clinicians began to experiment by working with families and, while not disbanding psychodynamic notions entirely, began to take a broader view of the psychodynamic process.

Coincident with the appearance of formal family approaches, modifications of psychoanalytic theory began to emphasize social forces: Erik Erikson's epigenetic theory, Heinz Hartmann's adaptive ego psychology, and Harry Stock Sullivan's interpersonal psychiatry. By taking into consideration the interactional forces and relational matrix that give rise to intrapsychic process, these theorists indirectly fostered the development of family therapy. These perspectives helped pave the way for empirical research by Margaret Mahler and colleagues on the separation-individuation process, which furthered the notion that development does not occur apart from social context. Mahler's seminal work is rarely mentioned in family therapy literature, yet it supports family interventions.

Radical Environmentalism and Family System Theory The next era in the development of family treatments was revolutionary; its impact is still being felt. The family was defined as a system, and this notion dominated family therapy for a generation. For many clinicians, the only way to intervene with a family was through systems therapy. The biologist Ludwig von Bertallanfy was perhaps the most influential theoretician of general systems theory. The first wave of family systems thinkers included Carl Whitaker, Donald Jackson, Gregory Bateson, and Salvador Minuchin, originator of structural family therapy, perhaps the most influential school of family therapy.

While systems therapists avoided individual psychodynamic approaches, they remained vigorous environmentalists. In the systems view, they did not focus on an identified patient, individual differences, individual symptoms, diagnostic classification, and therapeutics related to such classification. Early systems proponents believed all mental illness would be cured through this new approach. Along with this radical view came a rigid adherence to the systemic view. It was not uncommon for a family to come for a session without a family member, have their session canceled and be asked to return with the member at another time. By such actions, systems therapists were making their point: the family is a system, and an absent member crucially alters what happens in a session.

The here-and-now focus of family systems treatment was often demonstrated through live interviews in educational and clinical settings. One-way mirrors and the advent of video technology allowed practitioners to demonstrate their work, which was an enormous educational advantage over the challenge of verbally describing complex family interactions to a supervisor. On the other hand, direct observation appeared to foster an emphasis on techniques and a dramatization of the change process. For example, it was not uncommon for a clinician to teach that a strategic intervention in a stuck system could produce results immediately. Some specificity of family type related to disorder was sought with little success.

Child psychiatry and family systems therapists interacted minimally, in part because of the antimedical and antiindividual stances prominent in the work of family systems therapists. Child psychiatrists experienced cognitive and affective challenges related to the shift in clinical focus from the dyadic patient-therapist relationship to complex group interactive processes. These challenges precluded an easy transition to systems thinking for many psychiatrists.

Rapprochement and Integration The perspective of the system offered a new vantage point yet also possessed limitations as the sole conceptual basis for family therapy. The exclusive emphasis on the system led some therapists to believe that relevant perspectives on individual functioning had been lost in the systemic schools of family therapy. The current status of family treatment (and the orientation of this chapter) is that an approach integrating systems and individual

psychodynamic and behavioral concepts is necessary to treat the psychiatric disorders of children and adolescents. Family treatments now include parent training and education, and individual work with parents and children, as the clinical needs of a family dictate. Integrated family treatments often include aspects of different schools of family therapy.

Two common clinical problems illustrate the need for integrating systems and individual perspectives. When an adolescent with anorexia nervosa restricts food intake and loses weight, she experiences success and self-efficacy. The young girl's success in weight loss is not present in more typical adolescent pursuits and this seems to feed her compulsion to succeed at being thin. The disorder is often accompanied by familial interactions that are characterized by intense family involvement in the child's eating behaviors. Eating, an autonomous event, becomes inextricably involved with interactional processes. The adolescent's symptoms allow the family to maintain a set pattern of interaction, while the experience of weight loss is associated with her struggle for identity. Alternatively, the symptoms can be viewed as performing a defensive function for her and for other family members.

Similarly, individuals who exhibit self-harming behavior draw others toward themselves. Any assessment of self-harm must consider the family system as a potential etiologic factor, while simultaneously appreciating the effect of such behavior on a family. This interactional assessment must be balanced by an assessment of the affective and perceptive state of individuals who harm themselves. Self-harming behavior is often associated with disturbing affects and inadequate individual coping skills.

Several areas of clinical activity, research, and theory building support the integration of family treatments with other interventions, particularly interactions with individuals.

1. Family theorists such as David Reiss have emphasized factors that reside within the individual as well as factors that reside within the system.
2. Child development studies, such as the work of Daniel Stern and Robert Emde, have described the development of mental representations, or the inner world of the child, as generated from a relational context.
3. Clinical child psychiatry research emphasizes that biological vulnerability interacts with environmental demands to determine clinical outcome (“goodness of fit”).
4. Biopsychosocial psychiatry is a useful model for many clinicians because it considers the individual's biological and psychological endowments interacting with the individual's social context.
5. Psychodynamic psychiatry includes interactive concepts in its theoretical base, if not in actual practice. Self psychology uses the concept of parents as self-objects who must provide mirroring and idealizing functions for healthy child development, while object relations theory is based on the internalized object relations derived from life experience.

THEORETICAL ISSUES

Most practitioners of family therapy over the last generation have maintained that family therapy is both a theory of understanding human behavior and a set of strategies to intervene in problematic patterns of family interaction. Three core issues illustrate the current integrative approach to working with families. The effective practitioner understands the biopsychosocial approach to the relationship of family interaction to individual psychopathology, the importance of integrating systems and individual psychodynamic concepts and the need to appreciate the individual narrative, or personal experience, of each family member.

Biopsychosocial Approach Most contemporary approaches to family therapy emphasize that the family never causes psychopathology. This is accurate but also misleading in that biological, psychological, and social contributions to disorders (e.g., genetic vulnerability, inability to trust others, unemployment) do not solely cause disorders but predispose individuals to such disorders. The effective family therapist is open to two possibilities—that a family's interactive problems may primarily be a response to disorder or a cause of disorder. Past inaccurate attribution of familial responsibility for disorders such as autistic disorder appears to have given way to a reluctance to address parental and familial deficiencies. Effective family therapy supports a family that is responding to a disorder within a child and empathically explores family interactions that seem more etiologically important in the onset and maintenance of a child's disorder.

Integration of Individual and Systems Perspectives In family therapy influenced by systems perspectives, pathology results from structural or functional imbalances in the system rather than from individual difficulties. What happens to an individual's development, especially the development of a child, when such familial structural imbalances are persistent and pervasive? The answer to this question brings the focus back to the individual. Persistent, repetitive life experiences shape individual cognitive and emotional development to the point where these external experiences become internalized. This dynamic, developmental perspective is at the core of the contempor ary shift to integrated psychosocial treatments for children and adolescents.

The *systems approach*, a departure from so-called linear theories, uses cybernetics and general systems theory. Cybernetics denotes that systems maintain an equilibrium whereas in general systems theory, all living systems are characterized by tension between homeostasis and change. All things are interdependent, and nothing changes without everything else changing accordingly. In this view, symptoms are not seen as residing within the child but as serving a purpose for the whole family system. These symptoms provide systemic survival, or continuity, and maintain homeostasis. Each family member is presumed to act in a way that opposes symptomatic improvement in the presenting patient. Most family systems theorists hold this view, and therapeutic interventions attempt to counteract this hypothetical process. For example, in this view, one must destabilize the family system to promote change. The formulation for a systems-oriented clinician has less to do with the symptoms residing within a child than with the meaning such symptoms have for the family. Problem-maintaining sequences are observed, and these processes are interrupted. Schools of family therapy seen as systems therapy include strategic and structural schools (Table 48.5-1).

Table 48.5-1 Varieties of Family Therapy

The *dynamic, developmental approach* emphasizes that life experience becomes internalized within the individual, and the internalizations are derived from a relational (typically family) matrix. These internalizations prepare the way for choice of partner and new family formation. When parents' family-of-origin life experiences are negative (e.g., abuse, parental intrusiveness or overprotectiveness), they predispose to pathological family interaction. In some measure, they predict how parents will negotiate the transitions of the family life cycle in the next generation. The family's negotiation of life cycle tasks and the interactions that accompany it give rise to the inner world of the developing child.

In this cumulative view of development, the internalization of interactions over time constrains the way future interactions may develop. The term “dynamic” is used because each relationship can potentially alter inner models of self and others. This approach and its variations have been termed psychodynamic, and intergenerational (Table 48.5-1). Object relations family therapy and contextual family therapy heavily rely on these developmental and dynamic concepts. Experiential and cognitive-behavioral therapies, while not dynamic and developmental, also focus on the individual perspective.

Effective family therapy of child and adolescent psychiatric disorders is not predicated upon a choice between systems and dynamic, developmental views. It does involve a choice in the sequencing of interventions, which over time often involves both perspectives. Effective family therapists consider how these two approaches can be used in the choice of treatment modality and sequencing of treatments.

Narrative Therapy More recently, a dissatisfaction in pure systems approaches has been associated with the development of narrative therapy as an organizing

construct for clinical work with families. At first examination, this approach appears to be a return to the developmental perspective—life is lived in stages, and in each stage one's life story is developed. While there are similarities, there are also distinct differences. This school of therapy is heavily influenced by social constructivism, the idea that there are no essential truths except the ones individuals construct for themselves. In this view, the therapist enters the therapy session without a priori assumptions regarding healthy family life and normal child development but rather joins the family through interpreting their experience.

This approach is similar to the dynamic, developmental approach to family therapy in that the patient and family's descriptions of their lives—their narrative—are key elements to clinical formulation. The clinical data are interpreted quite differently however. Whereas a clinician from a dynamic, developmental perspective seeks to identify generalizable, if not universal, principles of families and child development, the narrative therapist brings no such structures to the interview. A therapist using this perspective helps the family to construct a new, current reality from the spontaneous conversation of the session. Since the systems therapist brings systemic principles to the clinical encounter and attempts to make clinical observations congruent with guiding systemic principles, the unstructured narrative approach departs from the systems perspective as well.

This new theoretical approach will be difficult to research in that, by definition, it rejects generalizable principles and primarily emphasizes the diversity of human functioning.

TECHNIQUES AND THERAPEUTIC PROCESS

Conjoint family therapy refers to sessions in which all family members are treated by a single therapist or cotherapists. *Concurrent family therapy* refers to family therapy and individual therapy conducted concurrently, usually with different clinicians. Finally, *a multiple family group* denotes meetings of several families, all of whom have an individual member with a specific problem. In each of these varieties, family therapy sessions are held regularly, with wide variation in frequency, and can last as long as 2 hours.

The clinician conducting family treatment must consider how therapy differs according to the constellation of ages presented by the family in treatment. The younger the child, the more likely it is that the nature of the dyadic parent-child relationship will determine the clinical presentation. Excluding instances of organic disease, caregiving practices are seen as the prime determinants of a clinical presentation in infants and toddlers. Families of school-age children are dealing with issues of success and competence. Mastery of this stage requires the availability of parents to foster the success of their children. Adolescent children question parental values and test the strength of the parents' marriage and parental self. The adolescent's emotional and geographical moving away from family taxes the resources of both parent and adolescent. The potential serious consequences of high-risk behaviors involving sexuality, drugs, and motor vehicles, brings an urgency to many family treatments of adolescent patients.

Overarching all these specific age issues are the tasks and stages of the family life cycle, which should be an orienting principle for family therapists. The concerns of families with children not yet in school and of families with children on the verge of becoming young adults are almost discontinuous. Consideration of the family life cycle, indeed almost any issue implying normal development, is complicated by the dramatic alterations in family structure influenced by divorce, single parenting, remarriage, and stepfamily formation. Attempts to structure and define the life cycle of stepfamilies are but one example of the complexity of considering the contemporary family life cycle.

Assessment and Formulation A family usually becomes involved in family treatment through the presentation of a symptomatic child or adolescent. There are two basic approaches to assessing the family. Some clinicians request that the entire family be present for the first session and focus this session on interactional factors, the meaning of the symptoms for the family system, and how the symptoms are maintained by the family system. Other clinicians do not request that the entire family be present but use separate interviews with child and parents to gather extensive history before observing the family as a unit.

Once the assessment is complete, a formulation is developed, and a recommendation for treatment offered. Patterns of family interaction arise for systemic and dynamic, developmental reasons. Repeated experience of these interactions makes them habitual for family members. The original motivation for initiating the interactions gradually becomes less influential in maintaining of the interaction. For example, an overprotective parent may initially behave that way to shore up a low sense of self, but the behavior eventually becomes somewhat independent of this underlying factor. This evolution suggests an initial behavioral approach to family treatment that interrupts the interactions that have become independent of motivational factors. Subsequently, the clinician explores more-complex, intrapsychic issues that underlie the presenting problems. The formulation serves as the basis for a primary therapeutic emphasis: supporting and educating a family about the consequences of a child's disorder or empathically exploring and ameliorating the family's contribution to a child's disorder.

These systemic and dynamic, developmental formulations lead to a sequential, integrated approach to family treatment. In the following description of clinical process, family treatments may not use all phases, and each phase (when experienced) may not occur in the order given. The process is presented as an attempt to address comprehensively the complex phenomena of family treatments.

Stabilize the Crisis Clinical disorders may have a sudden onset as a crisis or a less dramatic onset as the culmination of pervasive, ongoing problems. To stabilize a crisis, family and other systems interventions are needed, which may include removing a child from the home for a psychiatric hospitalization for emergency evaluation and treatment of life-threatening conditions. Clinicians may also need to use community resources, such as shelters for runaway children or emergency foster care. Short-term pharmacological interventions may accompany these systems interventions.

Parent Education After stabilization, or in the management of chronic problems, education is the initial family intervention. The clinician offers directives about child management that are easily understood and implemented by many parents. Parent training and education may suffice to interrupt behavioral problems, particularly when the problems are mild or parents are merely inexperienced. Problems related to limit setting, fostering a child's independence, and understanding a child's developmental needs often respond to didactic interventions and the development of behavioral plans. When education is partially effective or completely ineffective, the clinician likely has observed parental psychopathology or marital conflict underlying the inability of parents to respond to education. This information prepares for intervention in family interaction.

Intervention in Family Process At this point (the point of traditional family therapy), the clinician sees the entire family and notes patterns of interaction that are maintaining the clinical problem. The clinician challenges patterns of family interaction, using techniques such as those developed by the various schools of family therapy (Table 48.5-1). A few of the more common interventions follow.

In a *paradoxical intervention*, the clinician makes an unusual or contrary demand of a family, such as a request to exaggerate a symptomatic family pattern. If a family refuses or is unable to meet this demand, new potentials for change are experienced. By asking an overprotective mother to go everywhere with her son, a clinician opens up the topic of individuation in a new way, and healthier adaptations can be considered. Antecedents to the overprotectiveness can be explored, and a desire for liberation from this behavioral pattern is verbalized. If complied with, the paradoxical directive facilitates a family's alliance with the therapist. This technique is likely to be unsuccessful in the therapy of individuals who are cognitively and emotionally unable to process nuance and may be less useful in treating families with young children.

Reframing is a verbal intervention, analogous to *clarification* in individual psychotherapy, in which the therapist provides an alternative way to perceive a behavior. It is often used in taking problematic behaviors and giving an adaptive, positive connotation to them. For example, a therapist could hear of a child's distracting, demanding behavior while his father watches television and say to the father, “Your son sure has an interesting way of saying ‘I want to spend time with you, Dad.' ”

In *family enactment*, the therapist promotes an enactment of a family problem to understand its interactional and structural aspects. By directing an anorexic patient to eat in front of her family and thus enact a family meal, the therapist can observe pathological responses and offer alternatives. In less dramatic fashion, the clinician can direct family members to resolve communication conflicts evident in the consulting room.

The clinician who *prescribes tasks* gives instructions to the family on how to handle a specific conflict area and thus encourages their attempt to solve such conflicts away from the session. The results of the prescribed task, or homework, are brought back to the next family session and reviewed by the therapist. As families review their successes and failures with the therapist, areas for change are further clarified.

By *circular questioning* the therapist does not directly confront the symptomatic individual but explores patterns of communication within the family as various problems are discussed with family members. The therapist does not, for example, challenge a young boy about his aggressive, noncompliant behavior but understands the symptom through interaction with the boy's mother who describes its effects on her, her fears of its consequences on her other children, and her concerns regarding

possible future legal difficulty.

Through *mimesis* the therapist adopts the cultural style and understanding of the family in an effort to gain their acceptance. The therapist makes every attempt to adopt family rules and perspectives as a way of gaining their trust and minimizing resistance to treatment.

Altering interaction may eliminate the problem and provide opportunities for improved adaptation by all family members, yet these interventions are often resisted. Individual family members may have their own reasons for resisting change.

Considering the Individual Perspective Unsuccessful interventions in family interaction are often the result of the motivations of a family member or members. In trying to interrupt a symbiotic mother-child dyad, for example, the mother's attribution of meaning to her child's independence and the affects attendant to separation manifest. The preexisting factors that placed her at risk to overprotect her child begin to be exposed, and she resists this exposure. Concomitantly, the effects of symbiosis on the developing adolescent (e.g., dependency and selflessness) manifest. This conflict, conferred by developmental processes and promoted through therapeutic intervention in family process, is distressing, often to the point of inducing a return to dysfunctional interactions. By considering the internalized world of family members and using individual interviews to explore this world, a clinician can understand resistances and plan effectively for them. At this stage, family treatment may include individual interventions with the symptomatic child, either parent, or the marital unit. In this stage, the clinician must not be drawn in to pathological family processes, such as scapegoating an individual member.

Marital Therapy Parents may consider their marital relationship after, and sometimes alongside, the previous sequences of family intervention. Developmental histories prior to marriage, issues surrounding choice of partner, and the nature of the family life cycle come into new perspective for individuals who are considering the difficulties of their child. Parents more easily reflect on their possible contributions to the clinical problems when their child's disorder is stable. This stage usually occurs later in family treatment and indicates the deepening of family therapy.

Interventions With Siblings From a systems viewpoint, change in family interaction may be accompanied by the onset of symptoms in other members. If a family interactive process has been toxic to the development of one child, this toxicity will have affected other children as well. Individual factors such as a child's temperament, intelligence, and birth order also determine whether siblings develop clinical problems. For example, an overbearing intrusive father may have one son who is dependent and submissive and another son who actively rejects his authority. It is important to determine whether the onset of symptoms in another child is a diverting mechanism away from another problem, such as marital pathology (systems view), or a statement of individual vulnerability (dynamic/developmental view).

Other Interventions The clinical processes just described illustrate an integrated approach to family treatment conducted over a period of time. The chronological sequence is offered as a guide; the actual sequence varies according to clinical need. Some lower-functioning families will at some point in treatment require interventions at all levels, while higher-functioning families may require only a few components of this sequence (e.g., parent training or family process intervention). A coordinated family treatment may involve an intervention with the entire system or work with parts of the system. The duration of family treatment varies from brief interventions for a specific problem to longer-term treatment with intermittent clinical contacts around different problems, often posed by different family members.

Families in treatment often require other supports and interact with other systems of care and other interventions. Group therapy is helpful for children with socialization difficulties and becomes increasingly important with older children. Parents who work collaboratively with schools maximize educational success for children with academic difficulties. Conduct-disordered children frequently are involved with the courts. This involvement can be therapeutic in that it reinforces families in their limit-setting efforts. Families who have difficulty setting limits can benefit from collaboration with probation officers and court workers. Family therapy interventions with such children are strengthened when behavioral consequences for breaking the law are applied by an agency outside the family.

CLINICAL ISSUES

Indications *Family therapy*, broadly defined as working with the family, is always indicated in treating children, because the family is so critically a part of each child's life, whether it is causing problems or responding to problems. The sequencing of family interventions and whether those interventions should involve an individual child or parent, a subsystem such as the marital unit, or the whole group are important considerations. With some children, the family is viewed as etiologically important in the onset and evolution of the disorder. Here the clinician empathically explores and elucidates the role of family conflict in a child's disorder. With other children, family interactions are seen as a response to illness within the child. In this instance, the clinician educates and empathically supports the family. Complex family situations may involve both approaches.

Empathic exploration of family issues is indicated when (1) the clinical problem presents as an interactional problem (e.g., child abuse, a child running away from home), (2) parental psychopathology (e.g., substance abuse) complicates parenting efforts, (3) a child fails to respond to repeated medication trials, and (4) there is empirical and clinical support for the efficacy of specific family approaches with specific disorders (Table 48.5-2).

Table 48.5-2 Family-Oriented Treatment Approaches for Specific Clinical Problems

The empathic support of families is indicated in neurological disorders such as mental retardation and autistic disorder, medical disorders such as cystic fibrosis and diabetes, and psychiatric disorders with strong biological components such as schizophrenia. The stresses faced by the families of such ill children are unique and often overwhelming. Education, family support groups, respite care for children, and linkage with relevant agencies are all important adjuncts to supportive family treatment.

Limitations The limitations of family therapy are related not so much to the techniques and interventions associated with the modality but rather with developments in psychiatry that have impeded family therapy's full acceptance. Four of the most prominent limitations are (1) the reluctance of systems-oriented clinicians to accept current diagnostic nomenclature, (2) the absence of a generally acceptable classification relational disorders, (3) the preeminence of biological approaches to psychiatric disorders, and (4) a health care system that impedes the full utilization of family treatments.

The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) includes an axis delineating psychosocial problems in any diagnostic assessment. This has not satisfied many family therapists because the diagnostic system still mandates that an individual must be diagnosed with a disorder. Systems family therapists resist this notion because it is contrary to the basic systems postulate that symptoms do not reside within an individual but result from a particular family context. Family therapists of this orientation would rather have a relational disorders classification replace or at least supplement the individual-based DSM-IV. Such a classification has yet to be fully developed. This limitation should recede as integrated approaches to treatment emphasize both individual and systems variables.

The rapid development of the neurosciences has improved the diagnostic and treatment capability of the contemporary child and adolescent psychiatrist. At the same time, its emphasis has obscured the clinical reality that most disorders result from a mosaic of biological, psychological, and social variables. Psychosocial strengths can protect children from expressing disorders with biological contributions, while psychosocial stresses can foster clinical expression of a biological vulnerability. Many psychiatric training programs are having difficulty conveying this perspective, which supports the relevance of child and adolescent psychiatrists incorporating

family assessment and treatment in their clinical work.

For family therapists, health care delivery has primarily focused on the needs of the individual. From record keeping to individual diagnoses, the delivery system has tended to deemphasize systemic factors in favor of individual ones. This has been reinforced by many third-party payers who do not reimburse for family therapy. This has intensified recently as rapid changes in the economics of health care delivery have dictated the characteristics of psychiatric treatment and the roles of respective disciplines. These shifts have increasingly fragmented psychiatric treatments, exemplified by the psychiatrist providing biological treatments and the nonmedical professional providing psychosocial treatments, such as family therapy.

Complications and Difficulties Mobilizing the power and influence of the family in the life of a child can be dramatic. When progress does not occur, one or more common impediments to successful family treatment may exist.

1. Lack of appreciation of the cultural context of the family. Norms of expected family development and child development vary according to culture and, within cultures, socioeconomic level. The treating clinician must understand a family's cultural background and adjust treatment approaches accordingly.
2. Minimal father involvement. Mothers may bring children for intervention with fathers either resisting the process of intervention or physically unavailable, the case in many single-parent families. When fathers do get involved in treatment, progress is not ensured, but a major roadblock to success is removed.
3. A poorly defined role for siblings. Many siblings do not understand why they need to come to therapy, and their lack of involvement promotes the idea that there is only one identified problem child in the family.
4. Lack of clinician objectivity. Family problems have a high potential to mobilize countertransference responses. Gender roles, child-rearing practices, divorce, and human sexuality are emotionally laden areas that clinicians may be dealing with in their own lives. These are also common foci of family treatment and, correspondingly, may mobilize clinician attitudes that negatively affect treatment.
5. An inaccurate formulation of the family's role in the child's disorder. Progress is delayed when a family receives a supportive therapy approach when its conflicts should be explored and change expected in them. Similarly, progress is frustrated when families are inappropriately held accountable for a child's biologically based disorder.
6. Inattention to influential nonnuclear family relationships. Relationships that have the potential to be a resource for the child (e.g., an estranged biological father, a mother's boyfriend) and those that may be negatively affecting the child (e.g., an intrusive grandparent) must be included in planning family treatment.

Contraindications Like other psychotherapeutic interventions, family therapy is generally contraindicated for those who are not motivated to change. A therapeutic trial may be necessary to determine the appropriateness of family therapy. The main areas of contraindication are those that involve ethical and legal issues. In family treatment, it is contraindicated to

1. Repeatedly involve individuals who are not regularly involved in a child's life. Some family therapists recommend interviews with extended family members and using the material gathered from them to assist in the assessment and formulation of problems. Regularly involving individuals who do not have generational authority to make caregiving decisions for children is contraindicated.
2. Conduct family therapy that escalates the potential for harm to family members. Some family interactional pathology predisposes to violence. Some parents have so many deficits they can never safely and effectively set limits on their child's behavior. The family therapist must avoid exploring family conflicts with potential to lead to postsession violence.
3. Continue to work jointly with parents who are divorced. A session to clarify both parents' relationship to the clinician may be necessary, including such things as how information about a child's disorder is to be communicated. Ongoing work with each parent and new family unit should be conducted separately to reflect the new legal definition of family and to minimize confusion for the child.

Merrilee, a 15-year-old female, was referred by her maternal grandmother, with whom she was living. She had been expelled from school, been increasingly oppositional, and experienced depressive symptoms with vague suicidal ideation. Merrilee would return to her mother's home for periods of time only to get into conflicts with Mandy (her mother) and Bob, her mother's live-in boyfriend of 8 years' duration. Merrilee's father and Mandy had divorced when Merrilee was 3 years of age.

The initial formulation of family problems included a mother dependent on her daughter for affection, an immature “stepfather” with limited vocational skills and low self-esteem, and a maternal grandmother viewed by the mother as intrusive. Merrilee's adolescent behavior disturbance seemed the culmination of persistent developmental problems. Her two younger half siblings, Bob's sons, aged 6 and 4, were also suspected of having difficulties.

Merrilee was hospitalized to ensure safety, complete a diagnostic evaluation, and initiate pharmacotherapy. This stabilized the crisis, and resulted in a thorough assessment of familial strengths and liabilities. Contacts with community resources—mental health agency, court and school—helped prepare for outpatient treatment. Merrilee was noncommunicative in individual and group therapies but did comply with the behavioral expectations of the unit. She was discharged with noticeably improved affect, although she remained defiant.

Parent education in behavioral management techniques was an early focus in the outpatient course of treatment that followed hospitalization. Bob and Mandy were taught how to provide Merrilee with consistent discipline, and the family therapist discussed how a nurturant, nonpunitive stance can facilitate a child's acceptance of limits. Mandy related to her daughter as a sister would, indulging her when consistent discipline was indicated. Efforts to get Mandy and Bob to work together to set limits on Merrilee were unsuccessful, as she accurately perceived that Bob had no sanctioned role in the family. Merrilee's tyrannical, aggressive behavior toward her 6-year-old half brother necessitated an intervention in family process.

Merrilee's rejection of Bob's authority revealed problems in family structure. Bob was not supported by Mandy, and Merrilee took advantage of this fact, often commenting, “You are not my father.” Mandy frequently joined her daughter in berating Bob, exacerbating family tension. This structural problem made effective behavioral control of Mandy impossible. It appeared that commitment in the couples' relationship was tenuous and that perhaps Mandy was only continuing the relationship because of the income she received from Bob's Social Security disability benefits.

A family intervention attempted to strengthen the parents' relationship and decrease the control of the grandmother, who not only allowed Merrilee to stay in her home but also was caring for Jacob, the 4-year-old brother. Efforts were made to have both Jacob and Merrilee return home, with their parents working together to provide nurture and discipline. Parent education and intervention in family process were frustrated by the significant individual needs of parents.

Mandy was a self-preoccupied woman who had experienced a severe burn when she was a young child. Some facial disfigurement and a father who spent all of her developmental years incarcerated seemed important factors in her low self-esteem. Mandy's choice of husbands further reflected a low self-concept. Merrilee's father, Mandy's first husband, abandoned the family, and Bob, eight years her junior, was often unemployed and had cognitive impairments. Individual sessions with each parent revealed their feelings of being overwhelmed with raising three children. The role of grandmother took on a new perspective. Individual contacts with the grandmother revealed her to be less intrusive and more the family safety net, providing support and structure for a couple that desperately needed it. Merrilee refused individual therapy but received some support from a group intervention. Merrilee became increasingly aggressive toward John, the 6-year-old. It became apparent that the reunification of the family, with all three children in the home, taxed the parents as individuals and their relationship.

Bob fled from the responsibilities of parenting while Mandy undermined Bob, refusing to marry him. This inadequate parental coalition precluded effective parenting. Family process interventions, thus, became secondary to an evaluation of Bob and Mandy's relationship and its future. Little hope was offered for the children's stability without clarification of their relationship. Several episodes occurred in which Bob impulsively left the home or was asked by Mandy to leave. Couples (marital) therapy reviewed these conflicts and their antecedents, with the goal of enhancing Bob and Mandy's skills in resolving conflict.

School personnel initiated contacts with the family therapists (a cotherapy team) to identify problems with siblings, John and Jacob. John was seen as an aggressive, noncompliant boy while Jacob was seen as distractable, preoccupied, and impulsive. They appeared to be in competition with each other for the affections of teachers. Individual assessments, therapy, and parent training were offered for the boys.

The treatment of this disorganized family was fraught with difficulty, and the process described had several recursive loops. Well into treatment, Merrilee required rehospitalization for aggressive behavior and experienced legal detention for destruction of property. The family received in-home, behavioral parent training for several months. Ultimately, Merrilee could not be integrated into her mother's home, and when residential placement was unavailable, she was permanently placed with her grandmother. This boundary clarification seemed to help Mandy and Bob to focus on the needs of John and Jacob, who remained with them. John and Jacob's behavior improved in the school setting, although in individual sessions they both manifested indications of insecurity

The treatment of this disorganized family was fraught with difficulty, and the process described had several recursive loops. Well into treatment, Merrilee required rehospitalization for aggressive behavior and experienced legal detention for destruction of property. The family received in-home, behavioral parent training for several months. Ultimately, Merrilee could not be integrated into her mother's home, and when residential placement was unavailable, she was permanently placed with her grandmother. This boundary clarification seemed to help Mandy and Bob to focus on the needs of John and Jacob, who remained with them. John and Jacob's behavior improved in the school setting, although in individual sessions they both manifested indications of insecurity related to their father's impermanence in the family. The parents made some progress with their individual conflicts but continued to have difficulty with a long-term commitment in their relationship and experienced another physical separation.

The continuing availability of the family therapists, who identified the needs of individual family members and the need for family system alterations, served an organizing function for this multiproblem family. When the therapists identified the need for adjunctive supports outside clinical sessions, they coordinated family therapy with the activities of other agencies.

Goals of Treatment The immediate goal of family therapy is the relief of presenting symptoms of the child or adolescent. Family treatment involves other goals, but all are an outgrowth of the attempt to address the presenting symptom, usually described by the parent or child as residing within the identified patient. The practicality of this approach fosters a treatment alliance with the family and patient and becomes a marker of therapeutic progress. The systems and dynamic, developmental approaches differ in how they approach the goal of symptom removal and in addressing other, secondary goals.

Systems approaches address abnormalities in family structure, communication, and belief. Symptoms are used as markers of abnormalities in these areas, not as residing within a patient, and they have a meaning for the system. Goals of this approach are related to altering these underlying processes and thus ameliorating presenting symptoms.

The dynamic, developmental approach assesses how individual motivations and unfinished developmental tasks engender problematic family interactions. Symptoms are seen as residing within a patient and having been induced by family interactions that have meaning for each family member. Goals of this approach include understanding how internal mental representations of self and others have given rise to family interactions.

A key principle of family therapy is that any intervention in the family can potentially benefit other family members. As therapy proceeds, its scope naturally broadens to where the individual needs of all family members, including asymptomatic members, are understood and addressed. This outcome is an implicit goal in all family treatments, one with prevention implications.

ETHICAL ISSUES

The most important ethical considerations in family treatment are related to issues of therapist-patient relationship and confidentiality. When a family is seen, the simplest of questions—Who is the patient?—becomes problematic. Therapy may initially focus on an identified patient, but as therapy continues, the concerns and, at times, disorders of other members of the family become a focus for treatment. By empathically exploring the concerns of one family member, the therapist is at risk of coming into conflict with the needs of other family members. Healthier families can tolerate the therapist's apparent unfairness and lack of responsiveness to their concerns. Families whose members have not met each other's needs are less likely to be patient when one member appears to be benefiting most directly from the clinician's intervention. The therapist must be aware of this common ethical dilemma and make a deliberate attempt to relate to all family members. Drawing upon a therapeutic alliance with the whole family, the clinician specifies that, over time, each family member will be heard, and each perspective will be taken into account.

The concept of confidentiality changes in family therapy because the therapist is relating simultaneously to other members of the family. The family therapist can comfortably discuss with any family member any material that previously emerged in a family session and is known to other members. It is more problematic when information shared with a therapist in the absence of other family members has relevance to all family members. In this instance, the therapist must encourage the individual to bring up the issue in a future family session or (in rarer instances) initiate sharing the information with the family. In this latter case, discerning clinical judgment is essential.

RESEARCH AND EVALUATION

In the era of rapid growth of family therapy, the development of systems thinking was not accompanied by a parallel research tradition. Family therapy was developing outside the psychiatric community and, for many systems practitioners, the cause-and-effect logic of science was epistemologically inconsistent with the systems approach. Further, lack of agreement on a classification system for families and a wide variation in treatment methods also hampered research into the outcome of family treatments.

In the last several decades, however, as family treatments have become more accepted, empirical approaches have been developed and applied to family treatment outcome and family process evaluation. While many studies have included heterogenous patient samples with respect to diagnosis and severity of clinical disorder, the positive outcomes of family treatment are impressive. Meta-analysis studies also support family treatment efficacy by combining results from numerous studies and weighing methodological variations. These studies have shown family treatments to be better than no treatment, more effective than some alternative treatments, and comparable to behavioral and cognitive-behavioral treatments in a variety of disorders. Outcome research on family treatments for specific disorders provides further perspective on the direction of family treatments.

Schizophrenia For the last two decades, research on the impact of a negative emotional environment, termed “expressed emotion,” on the course of schizophrenia has led to family-based treatments. While controversy remains about whether high expressed emotion in families (e.g., critical, hostile responses) causes relapse or is a response to it, family treatments are now core components of schizophrenia treatment. These treatments provide family education to reduce guilt and blame, assist in communication skills, and help family members detect decompensation.

Mood Disorders Relatively few studies have used family treatments for mood disorders, although there is clear evidence of family characteristics that serve as risk factors for depression. These include weak parent-child attachment, hostility between parent and adolescent, parental depression, and ineffective parenting behaviors. These factors suggest that family treatment studies are indicated and that a family intervention has potential for success.

Eating Disorders A family treatment intervention appears helpful for patients with anorexia nervosa and bulimia nervosa, compared with individual supportive therapy, particularly for adolescent patients (compared with adults with the disorder), and when the disorder is of less than 3 years' duration. Uncontrolled trials have demonstrated such impressive results with anorexia nervosa that family treatment is now routinely part of most treatment programs. Clinical experience and uncontrolled trials suggest that family treatments aid the adjustment of children and adolescents with psychosomatic disorders such as chronic asthma and diabetes.

Anxiety Disorders Little work has been done on family factors associated with anxiety, let alone intervention studies. Studies of adult patients suggest that family processes such as lack of conflict resolution and triangulation are more common in panic disordered patients than in depressed patients and a nonclinical control sample. Studies of childhood anxiety are beginning to identify family factors associated with anxiety to complement a rich clinical literature on separation anxiety. Family treatment in combination with a cognitive-behavioral approach appears promising in the treatment of childhood anxiety disorders.

Conduct and Oppositional Defiant Disorders There is abundant evidence that family factors such as harsh and inconsistent discipline and poor parental supervision are relevant to conduct disorder and oppositional defiant disorders. Further, several treatment programs based on parent management models, behaviorally oriented family treatment, home-based models of intervention, and integrative treatments have all been shown to be successful. Regardless of the model used, working with families is a core component of any treatment of youths with conduct disorder and oppositional defiant disorder.

Substance Abuse Family risk factors for substance abuse mirror those for conduct disorders. Similarly, there is impressive evidence that family treatments have positive outcomes with substance-abusing adolescents.

Status of Research In several areas, there is consensus regarding outcome research on family therapy for children and adolescents:

1. No one method or school has established overall superiority in effectiveness.
2. Treatment groups are categorized by individual diagnosis or clinical problem as there is no universally accepted method of family diagnosis.
3. The epidemiology of family risk factors associated with developmental psychopathology suggests areas in which family interventions are clinically justifiable (e.g., anxiety disorders, mood disorders) even though effectiveness may not yet be empirically verified.
4. Significant limitations remain in the ability of current research efforts to measure the complex components of family treatments.
5. General dysfunctional family processes and characteristics are easier to demonstrate than their diagnostic specificity.

6. Whether it is better to measure a positive outcome in family treatment by change in an individual or change in the family as a unit remains unclear.

Outcome research in family treatments is developing sophistication and breadth. It is supplemented by family treatment process research, which does not examine the results of treatment models but rather assesses how the processes of treatment determine outcome. Linking process measures to outcome studies will encourage the development of specific, effective family treatments.

SUGGESTED CROSS-REFERENCES

Individual psychotherapy is discussed in Section 48.1 and group psychotherapy in Section 48.4. Family, couples, and marital therapy are also covered in Section 30.5.

SECTION REFERENCES

Barker P: *Basic Family Therapy*, ed 4. Oxford University Press, New York, 1998.

Barnes GG: Family therapy. In *Child and Adolescent Psychiatry: Modern Approaches*, ed 3, M Rutter L Hersov, E Taylor, editors. Blackwell Scientific, London, 1994.

*Carter B, McGoldrick M, editors: *The Expanded Family Life Cycle: Individual, Family, and Social Perspectives*, ed 3. Allyn & Bacon, Needham Heights, MA, 1999.

Combrinck-Graham L: A model of family development. Fam Process *24:*501, 1985.

Combrinck-Graham L: Developments in family systems theory and research. J Am Acad Child Adolesc Psychiatry *29:*501, 1990.

*Diamond G, Serrano A, Dicky M, Sonis B: The current status of family based outcome and process research. J Am Acad Child Adolesc Psychiatry *35:*6, 1996.

Falloon I, Lillie F: Behavioral family therapy: An overview. In *Handbook of Behavioral Family Therapy*, I Falloon, editor. Guilford, New York, 1988.

Forehand R, McMahon R: *Helping the Noncompliant Child: A Clinician's Guide to Parent Training*. Guilford, New York, 1981.

Freedman J, Combs G: *Narrative Therapy: The Social Construction of Preferred Realities*. Norton, New York, 1996.

Goodver I, Herbert J, Tamplin A, Secher S, Pearson J: Short-term outcome of major depression: Life events, family dysfunction, and friendship difficulties as predictors of persistent disorder. J Am Acad Child Adolesc Psychiatry *36:*474, 1997.

Gurman AS, Kniskern DP, editors: *Handbook of Family Therapy*, vol 2. Brunner/Mazel, New York, 1991.

Jenkins H, editor: *The Dictionary of Family Therapy*. Blackwell, Cambridge, MA, 1995.

Jensen P, Josephson A, Frey J: Informed consent: Legal content versus therapeutic process. Am J Psychother *93:*378, 1989.

*Josephson A, Moncher F: Family history. In *Handbook of Child and Adolescent Psychiatry*, vol 5. J Noshpitz, editor. Wiley, New York, 1998.

*Josephson A, Moncher F: Observation, interview, and mental status assessment (OIM): Family unit. In *Handbook of Child and Adolescent Psychiatry*, vol 5. J Noshpitz, editor. Wiley, New York, 1998.

Josephson A, Moncher F: Family treatment. In *Handbook of Child and Adolescent Psychiatry*, vol 6. J Noshpitz, editor. Wiley, New York, 1998.

Kazdin A: Parent management training-evidence, outcomes, and issues. J Am Acad Child Adolesc Psychiatry *36:*1349, 1997.

Mahler MS, Pine F, Bergman A: *The Psychological Birth of the Human Infant: Symbiosis and Individuation*. Basic Books, New York, 1975.

Malone C: Child psychiatry and family therapy: An overview. J Am Acad Child Psychiatry *18:*4, 1979.

McGoldrick M: Ethnicity, cultural diversity, and normality. In *Normal Family Processes*, F Walsh, editor. Guilford, New York, 1993.

McGoldrick M, Heiman M, Carter D: The changing family life cycle: A perspective on normalcy. In *Normal Family Processes*, F Walsh, editor. Guilford, New York, 1993.

Najman J, Behrens B, Andersen M, Bor W, O'Callaghan M, Williams G: Impact of family type and family quality on child behavior problems: A longitudinal study. J Am Acad Child Adolesc Psychiatry *36:*1357, 1997.

*Nichols M: *The Self and the System: Expanding the Limits of Family Therapy*. Brunner/Mazel, New York, 1987.

Nichols WC, Everett CA: *Systemic Family Therapy: An Integrative Approach*. Guilford, New York, 1986.

Pinsof W: *Integrated Problem-Centered Therapy: A Synthesis of Family, Individual and Biological Therapies*. Basic Books, New York, 1995.

Reiss D: The represented and practicing family: Contrasting visions of family continuity. In *Relationship Disturbances in Early Childhood: A Developmental Approach*, A Sameroff, R Emde, editors. Basic Books, New York, 1989.

Rutter M, Rutter M: *Developing Minds: Challenge and Continuity Across the Life Span*. Basic Books, New York, 1993.

Sameroff A, Emde R: *Relationship Disturbances in Early Childhood: A Developmental Approach*. Basic Books, New York, 1989.

*Sargent JS: Family therapy in child and adolescent psychiatry. Child Adolesc Psychiatr Clin North Am *6:*151, 1997.

Scharff D, Scharf J: *Object Relations Family Therapy*. Jason Aronson, Northvale, NJ, 1987.

Selekman M: *Solution Focused Therapy With Children: Harnessing Family Strengths for Systemic Charge*. Guilford, New York, 1997.

Sprenger DL, Josephson AM: Integration of pharmacotherapy and family therapy in the treatment of children and adolescents. J Am Acad Child Adolesc Psychiatry *37:*887, 1998.

Stern D: *The Interpersonal World of the Infant: A View from Psychoanalysis and Developmental Psychology*. Basic Books, New York, 1985.

Thomas A, Chess S: *Dynamics of Psychological Development*. Brunner/Mazel, New York, 1980.

Walsh F: Conceptualization of normal family processes. In *Normal Family Processes*, F Walsh, editor. Guilford, New York, 1993.

48.6 PEDIATRIC PSYCHOPHARMACOLOGY

BARBARA J. COFFEY, M.D.

Contemporary pediatric psychopharmacology has been driven by an exponential growth in the clinical investigation and use of psychotropic agents to treat childhood- and adolescent-onset psychiatric disorders. This development makes it essential for today's practitioner to learn the basic principles and indications for use of medications and become up-to-date on results of the growing body of clinical trials in youths. This chapter reviews and summarizes pertinent information about current use of psychopharmacological agents in children and adolescents.

Indications for the use of psychopharmacological agents in childhood and adolescence have increased in the past decade as developments in neuroscience and clinical research have broadened the knowledge base. Psychopharmacological agents are used to treat symptoms of juvenile-onset psychiatric disorders that cause significant distress and impairment, and they can facilitate optimal development and adaptation. From the biopsychosocial perspective, pharmacotherapy represents one component of a comprehensive treatment plan.

Established indications are discussed in the sections on specific agents. General indications for pharmacotherapy in child and adolescent psychiatry include the following disorders:

- Childhood- and adolescent-onset schizophrenia
- Autistic disorder and pervasive developmental disorder
- Tourette's disorder
- Enuresis
- Attention-deficit skill/hyperactivity disorder
- Obsessive-compulsive disorder
- Major depressive disorder
- Bipolar disorder
- Separation anxiety disorder
- Panic disorder
- Sleep disorder

GENERAL GUIDELINES

Comprehensive diagnostic evaluation is the foundation of treatment planning in child and adolescent psychiatry. Use of the biopsychosocial model as a guiding principle for case formulation remains relevant and sound. The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) has contributed to improved reliability and validity in childhood and adolescent psychiatric diagnostic evaluation. The multiaxial system has been particularly useful in child and adolescent settings, as children often have multiple, coexisting problems and need to be evaluated in the context of both developmental stage and family situation.

The multiaxial diagnostic system lends itself to a diagnosis-specific treatment approach, a general guideline in child and adolescent pharmacotherapy. Diagnosis is a foundation for indications for pharmacotherapy and for medication selection, and individual target symptoms determine specific medication choice.

Developmental issues must be considered in the pharmacological treatment of children and adolescents. Growth and development are associated with changes in organ size and proportion and in tissue mass, particularly water and adipose tissue. Children differ significantly from adults in their pharmacokinetic capacities. Children have relatively large livers in proportion to body weight compared with adults and after about 1 year of age reach adult glomerular filtration rates. Thus, because children may metabolize medications more efficiently, higher doses relative to body weights (compared with adults) are often necessary. In addition, children tend to have less adipose tissue and less protein-binding than adults and may have more drug available for bioactivity and potentially adverse effects.

In addition, juveniles likely differ significantly from adults in their central nervous system (CNS) capacities. Neurotransmitter development takes place sequentially and at differential rates during childhood; for example, dopamine and norepinephrine pathways and functions probably develop earlier in life than do serotonin capacities. Hormonal changes during adolescence are also likely to affect CNS responsivity to pharmacological agents.

The children's place in their family context is an essential component of treatment planning. Typically, the child's behavior or emotional state disturbs someone else in the environment, usually an adult, and the child is not always a willing participant in treatment. Parents are a necessary source of information about the child's functioning and response to medication and need to be considered treatment allies. Parental attitudes of anxiety, doubt, or antagonism can significantly affect compliance. The child's views are also crucial in the decision to use pharmacotherapy. All children have fantasies and beliefs about their problems; the use of “pills” or “drugs” to modify behavior or alleviate mental states is likely to be emotionally charged for the child.

Issues of informed consent are another guiding principle. Since many pharmacotherapeutic agents used in child and adolescent psychiatry are not formally approved for pediatric use by the Food and Drug Administration (FDA), a detailed discussion between physician and parents or legal guardians is needed when the decision to use pharmacotherapy is made. Benefits and risks of pharmacotherapy must be weighed and discussed with the parent or legal guardian. Clinically accepted but off-label medications may be used with detailed discussion. Informed consent can be obtained when the patient (if adolescent) or parent or guardian is mentally competent and understands the nature of the condition being treated, the proposed treatment, and its potential risks and consequences. Natural history of the condition if untreated and alternatives should be discussed. The discussion should be noted in the record. Records should also indicate that the patient and parent or guardian had an opportunity to ask questions during the discussion.

Baseline Assessment and Premedication Workup All children and adolescents who are candidates for psychopharmacological agents should undergo a thorough medical and psychiatric evaluation. The clinical workup should include a comprehensive psychiatric evaluation with particular emphasis on previous psychopharmacological treatment including response to medications, dosages, and duration of trials. A family history is necessary, including medication history for family members with psychiatric illness, with responses to medication noted.

A complete physical examination prior to starting medication is important for differential diagnostics and for monitoring physiological effects such as alterations in blood pressure, pulse, or weight. A screening neurological examination should be performed on all patients who are candidates for psychotropic agents, noting overall developmental level, gross and fine motor coordination, cranial nerve functions, sensory-perceptual apparatus, and gait. Before medication is given, abnormal movements, tics, or stereotypic behaviors (Abnormal Involuntary Movement Scale [AIMS]) should be carefully observed as abnormal movements are not uncommon in children with attention-deficit/hyperactivity disorder, tic disorders, and psychotic disorders. Preexisting movements need to be distinguished from medication-related dyskinesias or tics. Basic laboratory workup for most should include screening of hematological, metabolic, hepatic, and renal function and for some medications, cardiovascular function.

Rating Instruments Systematic baseline evaluation is necessary for all child and adolescent candidates for medication. The use of structured diagnostic interviews and rating instruments can help achieve this goal. Structured or semistructured diagnostic interviews such as the Schedule for Affective Disorders and Schizophrenia–Parent Epidemiological version (K-SADS) or Diagnostic Interview for Children (DISC) obtain systematic history of lifetime or current psychopathology or both in juveniles. They function as a systematic anamnesis, reducing errors of omission in psychiatric evaluation of children and adolescents. Severity-rating scales such as the Conners Parent and Teacher Rating Scales, the Achenbach Child Behavior Checklist, and the Kovacs Children's Depression Inventory have an established place in research studies and can be applied in the clinical setting to add quantitative measures to the child's evaluation. These can be particularly useful

when multiple informants' assessments of the child are used.

Careful record keeping is a must. Documentation should include all relevant diagnoses and disorder-specific symptoms that are targets for medication intervention, rationale for medication selection, initial dosage and treatment plan, increases of dosages and response, documentation of discussion with parents and child regarding informed consent issues, any adverse reactions including expected adverse effects or any unexpected allergic or idiosyncratic reactions, and rationale for changing or discontinuing medications.

Appropriate medication selection requires comprehensive information about the child's history and current functioning. Medications available for use in children and adolescents have increased considerably in the past decade, and the clinician has many choices. Medication selection should always be based on efficacy and safety; in general, the best-established and least toxic agent should be tried first. If after an adequate trial the medication has not ameliorated the symptomatology, another medication or category of medications can be tried. An adequate trial is defined as a high enough dosage for a long enough duration, taking into account the child's age and size.

Pharmacotherapeutic Maintenance and Monitoring Dosage regulation requires time and experience. Plasma concentrations of some drugs are readily determined, and therapeutic ranges similar to those of adults have been found for some drugs such as the tricyclic drugs and lithium (Eskalith, Lithobid). In general, except in emergencies, it is best to start with a low dosage and raise it by small increments at regular intervals until therapeutic effect or range is reached or adverse effects appear. Once the child has been stabilized, usually after several weeks on a maintenance dosage, the maintenance phase begins. Information from teachers and other clinicians involved in the child's care must be regularly communicated to the prescribing physician. Periodic monitoring of laboratory and growth parameters is recommended.

Duration of treatment depends on the diagnosis, the child's age, the specific medication, and the clinical situation. Few fixed guidelines exist for long-term treatment, medication-free periods, and indications for discontinuation. In general, children do not respond to an agent within the expected time frame while in the therapeutic range, they are unlikely to respond at a later date. A rational approach to this situation is to switch to another drug in the same category or to another category altogether if indicated.

Most children who respond positively to a medication need a 6- to 12-month trial of the agent, depending on the clinical situation. The academic year is often used as a framework for medication treatment; if the child has improved and a trial off the agent is indicated, summertime is relatively less risky for a medication-free period. However, the child's clinical situation should be carefully reviewed before a decision is made to discontinue a particular trial; many of the disorders that respond pharmacologically in youths persist over time and may only become more symptomatic during the school year. Consideration of treatment discontinuation should weigh the risks of untreated illness.

Adverse Effects Despite their medication-handling efficiency, children and adolescents are no less at risk than adults and perhaps are more at risk in some cases for development of adverse effects (Table 48.6-1). This has both physiological and practical significance: the clinician must know the adverse effect profiles of all medications being prescribed and must know how to manage adverse effects if they arise. In addition, children and adolescents can be particularly sensitive about their bodies and how they work and often get quite frightened or suspicious if a sudden change or dysfunction occurs during pharmacotherapy. Such feelings can interfere with compliance and optimal outcome.

Table 48.6-1 Adverse Effects and Their Management in Children and Adolescents

Many of the adverse effects of antipsychotic drugs, antidepressant agents, and lithium seen in adult patients can be observed in children and adolescents. Of particular concern are the anticholinergic and cardiovascular effects of the tricyclic drugs and the extrapyramidal adverse effects of the neuroleptics, including dyskinesias. Withdrawal dyskinesias seem more frequent than tardive dyskinesia in young children, but tardive dyskinesia has been observed in this age group.

In general, the best therapy for adverse effects is preventive. It is generally best to start with a low dosage and titrate upward in a stepwise fashion as the child can tolerate. Use of the lowest possible maintenance dosages in the therapeutic range can minimize adverse effects. Monotherapy with a broad-spectrum agent is generally preferable. However, targeted combined pharmacotherapy is often necessary in children with multiple comorbid disorders. Judicious use of two or more agents in combination should not precede careful, systematic trials of single agents from two or more classes of medication as indicated by the clinical situation.

SPECIFIC AGENTS

Sympathomimetics The sympathomimetics are the most widely prescribed and best-studied psychopharmacological agents in juveniles. Methylphenidate (Ritalin), dextroamphetamine (Dexedrine), pemoline (Cylert), and a combination of dextroamphetamine and amphetamine (Adderall) are currently available for use in the United States (Table 48.6-2). Sympathomimetics are formally indicated in the treatment of attention-deficit/hyperactivity disorder, including all subtypes. Core symptoms of attention-deficit/hyperactivity disorder occur in two clusters: hyperactivity-impulsivity and inattention. Psychostimulants achieve their therapeutic effect by reducing these core symptoms including motoric hyperactivity, distractibility, inattention, and impulsivity. Stimulants also reduce aggressive behavior and impulses and improve social skills.

Drug*	Oral Dosage Range (mg/day)
Methylphenidate (Ritalin)	2.5–60
Dextroamphetamine (Dexedrine)	2.5–30
Dextroamphetamine-amphetamine (Adderall)	2.5–40
Pemoline (Cylert)	18.75–112.5

* All cause weight loss with extended use.

Table 48.6-2 Sympathomimetics Used for Attention-Deficit/Hyperactivity Disorder in Children and Adolescents

Mechanism of action involves stimulation of catecholamine activity in the CNS and peripheral nervous system; stimulants prevent catecholamine reuptake at the synaptic cleft and inhibit metabolism and breakdown. More than 160 controlled trials of stimulants have demonstrated their efficacy in reducing symptoms in most

school-age children, as well as adolescents and adults with attention-deficit/hyperactivity disorder. Highly effective, they are easy to administer and generally safe. Their effect actually is not paradoxical in that adult men and normal boys also display increased concentration and focus when given a single dose of dextroamphetamine. Stimulants appear to achieve their effects by short-term improvement of attentional difficulties and concentration; they have been reported to improve performance efficiency, work output, and peer interactions.

Because dextroamphetamine and methylphenidate have a relatively short duration of action, with peak effects between 1 and 3 hours, they must be administered in divided doses two or three times a day. They are typically given at breakfast and at noontime (or early afternoon). Pemoline and combination dextroamphetamine-amphetamine have a longer duration of action and thus can be given once a day, before school. Slow-release forms of dextroamphetamine and methylphenidate are available. Shorter-acting forms can be added to slow-release preparations in the afternoon after school to facilitate the completion of homework.

Dosage titration should be individualized. Historically, dosing was adjusted according to body weight, but more-recent research suggests an absence of correlation between weight and methylphenidate dose. Current recommendations are to use fixed-dose stepwise titration starting with low dosages (5 to 10 mg of methylphenidate or equivalent). Maximum dosages for school-age children are typically around 40 to 60 mg of methylphenidate (or equivalent); individual children may require more or less. Adolescents and adults may require higher maximal dosages. Sustained-release preparations are available for methylphenidate and dextroamphetamine.

Common adverse effects include nausea, abdominal discomfort, anorexia, and insomnia. Some have reported adverse effects on growth parameters that are reversible with discontinuation of the medication, but other evidence is contradictory. Several studies have noted decreased height and weight percentiles with long-term treatment and reduced growth velocity; however, growth tends to rebound after stimulants are discontinued. More-recent research suggests that children with attention-deficit/hyperactivity disorder may already be at risk for growth delay. Common short-term adverse effects include reduced appetite, insomnia, dysphoria, and (possibly) tics.

Some reports have documented the development of tics in children on stimulants. However, some of these youths may have had preexisting risk factors for the development of tics with or without the use of stimulants. The use of stimulants in children with Tourette's disorder is controversial. The most recent research suggests that the ameliorative effects of stimulants on the coexisting attention-deficit/hyperactivity disorder may outweigh any theoretical risk of increased tics.

Stimulants may precipitate dysphoria and mood lability in certain vulnerable children. Psychotic symptoms (including hallucinations) are rare.

Antipsychotic Agents

Dopamine Receptor Antagonists Formal indications for use of antipsychotic agents include childhood- and adolescent-onset psychoses and pervasive developmental disorders, Tourette's disorder and severe self-injurious behavior, uncontrollable agitation, or assaultive behavior (Table 48.6-3). Dopamine receptor antagonists include the phenothiazines, butyrophenones, thioxanthenes, dibenzoxazepines, dihydroindolones, diphenylbutylpiperidines, and dibenzodiazepines. These drugs function as dopamine, a-adrenergic, and cholinergic receptor antagonists in the CNS and peripheral nervous system. They have antipsychotic actions in adults and children with psychotic symptoms such as auditory hallucinations and delusions. They reduce agitation, severe aggression, and movements including motor tics and stereotypic movements.

Table 48.6-3 Antipsychotics for Use in Children and Adolescents

Overall, comparable doses of dopamine-receptor antagonists are approximately equally efficacious for psychoses. The low-potency aliphatic and piperidine phenothiazines such as chlorpromazine (Thorazine) and thioridazine (Mellaril), respectively, tend to be sedating and more anticholinergic, whereas the high-potency butyrophenones and piperazine phenothiazines such as haloperidol (Haldol) and trifluoperazine (Stelazine) tend to be less sedating but more likely to induce extrapyramidal effects such as akathisia, acute dystonic reactions, and parkinsonism.

PSYCHOSES Childhood-onset schizophrenia is rare but probably has continuity with later-onset schizophrenia. Children with psychotic disorders not otherwise specified include a more recently described subgroup with complex developmental disorders and fluctuating psychotic symptoms known as multidimensionally impaired (or multiplex developmental) disorder.

There are few controlled studies of antipsychotics in youth. T. Spencer and coworkers found haloperidol (0.5 to 3.5 mg daily) to be superior to placebo in a 10-week double-blind study in children ages 5 to 11. In a controlled study of 75 youths ages 13 to 18, B. Pool and coworkers found no difference between haloperidol (2 to 16 mg; mean, 9.8) and loxapine (Loxitane) (10 to 200 mg; mean 87.5). Those more severely ill showed somewhat greater improvement on active drugs than on placebo.

TOURETTE'S DISORDER Currently the only formally approved treatments for Tourette's disorder are haloperidol and pimozide (Orap). Several controlled studies have demonstrated that these drugs have greater efficacy than placebo. A 6-week controlled trial of haloperidol and pimozide in adults with Tourette's disorder showed equal efficacy but slightly more adverse effects with haloperidol. Since most patients with Tourette's disorder have mild-to-moderate symptoms and antipsychotic agents are associated with a relatively high risk of adverse effects, treatment alternatives should be considered, such as the a-adrenergic receptor agonists. Additionally, most Tourette's disorder patients treated in clinical settings have significant comorbidity that antipsychotic agents do not address adequately.

Common adverse effects of typical neuroleptics include sedation, weight gain, and cognitive blunting, particularly at higher doses. Extrapyramidal effects associated with typical neuroleptics include acute dystonic reactions, parkinsonism, akathisia (motor restlessness), neuroleptic malignant syndrome, rabbit syndrome, and tardive dyskinesia.

There is growing evidence that adolescents and even young children can develop antipsychotic agent–related dyskinesias. Some studies indicate that children are more likely to develop withdrawal-emergent dyskinesias that are reversible on discontinuation of the medication than tardive dyskinesia or late-appearing dyskinesias; nevertheless, tardive dyskinesia has been reported in both young children and adolescents. Recently, in a study of 34 children with childhood-onset schizophrenia who were not responsive to antipsychotic medication, 17 (50 percent) had neuroleptic-related dyskinesias, including 9 withdrawal and 8 tardive. Abnormal movements present in these children before pharmacotherapy must be differentiated from any that develop after medication treatment, with baseline and periodic AIMS examinations.

Serotonin-Dopamine Antagonists Serotonin-dopamine antagonists are indicated for the treatment of childhood and adolescent-onset psychoses. Those currently available for pediatric use include clozapine (Clozaril), risperidone (Risperdal), and olanzapine (Zyprexa). Both risperidone and olanzapine may be used as first-line treatments; clozapine should be used only as third-line treatment following failure of two adequate trials of other antipsychotic agents.

Mechanism of action includes both dopamine- and serotonin-blocking effects. Serotonin-dopamine antagonists are thought to have reduced risk for extrapyramidal effects, particularly tardive dyskinesia.

Clozapine, historically the first available serotonin-dopamine antagonist, was found effective in about 30 percent of patients with childhood-onset psychoses who had not benefited from multiple previous trials of dopamine-receptor antagonists. A double-blind controlled comparison of clozapine and haloperidol in 27 treatment-refractory childhood-onset schizophrenia patients (mean age, 14) demonstrated clozapine's (mean daily dosage 176 mg) superiority for both positive and negative symptoms. Potentially serious adverse effects included neutropenia and seizures.

Risperidone was found effective in 10 adolescents with childhood-onset schizophrenia (ages 11 to 18) at mean daily dosages of 6.6 mg for positive and negative symptoms and in 16 youths ages 9 to 20 at mean dosages of 5.9 mg a day. It was also studied openly in Tourette's disorder and was found effective in a small group of children ages 8 to 16 at mean dosages of 2 mg a day and in 31 children and adults at mean dosages of 9 mg a day. Adverse effects included fatigue and weight gain.

To date few published controlled data exist on olanzapine in psychotic youths, although studies are in progress. Olanzapine's mechanism of action is similar to that of clozapine. Daily dosages up to 20 mg may be helpful. Adverse effects include headache, fatigue, and gastrointestinal symptoms.

Antiparkinsonian Agents In general, except for the acute dystonic reactions, reduction of antipsychotic dosage is recommended before addition of an antiparkinsonian agent. If decreasing antipsychotic dosage is impossible or ineffective, then antiparkinsonian treatment is indicated (Table 48.6-4). With the high-potency antipsychotic agents, since some investigators found that the greatest risks for acute dystonic reactions occur with males under age 30 during the first few weeks of treatment, prophylactic use of antiparkinsonian agents for the first few weeks of treatment may be indicated. Acute dystonic reactions are treated with intramuscular diphenhydramine (Benadryl) or benztropine (Cogentin), followed by a switch to the oral form for 2 to 4 weeks. Any recurrence of extrapyramidal effects is an indication for further treatment.

Generic Name	Trade Name	Indications	Dosage Range (mg)
Diphenhydramine	Benadryl	Acute dystonia	25–50 i.m.
Benztropine mesylate	Cogentin	Extrapyramidol effects: parkinsonism, akathisia	1–2 i.m. 1–6 p.o. daily
Trihexyphenidyl	Artane	Extrapyramidal adverse effects	2–6 p.o. daily

EPS, extrapyramidal adverse effects.

Table 48.6-4 Antiparkinsonian Medications for Use in Children and Adolescents

Antidepressant and Antiobsessional Agents

Tricyclic Drugs Tricyclic drugs are formally established for use in children and adolescents with enuresis, attention-deficit/hyperactivity disorder, and major depressive disorder (Table 48.6-5). Additional indications for use include school refusal based on separation anxiety and sleep arousal. Imipramine (Tofranil), desipramine (Norpramin), and nortriptyline (Aventyl) have been most frequently studied in youths.

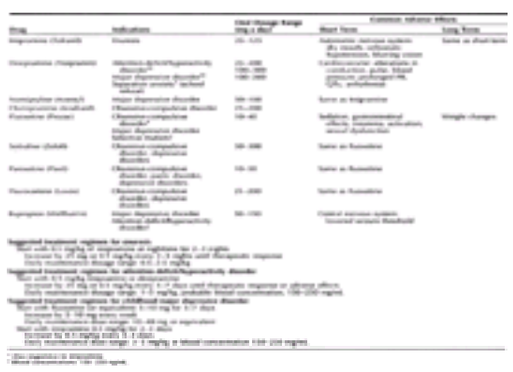

Table 48.6-5 Antidepressants for Use in Children and Adolescents

Tricyclic drugs may have several different mechanisms of action including inhibition of catecholamine or indoleamine reuptake at the receptor level in the CNS and both adrenergic and anticholinergic activity in the CNS and peripheral nervous system. Studies indicate that imipramine's effectiveness in enuresis is related to neither its anticholinergic properties nor its effects on sleep architecture but perhaps some direct effects on bladder muscles. Desipramine has been found more effective than placebo in the treatment of attention-deficit/hyperactivity disorder.

Controlled studies have failed to demonstrate efficacy of tricyclic drugs over placebo in the treatment of major depressive disorder in both children and adolescents. Diagnostic heterogeneity, small sample sizes, hormonal influences, high levels of comorbidity, neurotransmitter immaturity, and pharmacokinetic factors are among the factors proposed to account for this lack of treatment response. Open studies have shown that some children with major depression respond to tricyclic drugs in dosages that achieve concentrations of 150 to 250 ng/mL of blood.

Anticholinergic adverse effects of tricyclic drugs are common but transient and include dry mouth, blurred vision, and constipation. Antihistaminic adverse effects include weight gain and fatigue. In vulnerable individuals with a previous history of seizures, family history of seizures, or history of electroencephalographic (EEG) abnormalities the seizure threshold may be lowered.

Cardiovascular adverse effects are not infrequent, but most are not clinically significant. Small but clinically significant elevations in heart rate and blood pressure occur with tricyclic drugs. However, electrocardiographic (ECG) effects with more serious implications are possible, since tricyclic drugs have a quinidine-like effect and prolong the PR interval and can widen the QTc interval. There have been several case reports of sudden death of school-age and early adolescent children on desipramine, apparently with therapeutic dosages for treatment of attention-deficit/hyperactivity disorder or depression. Unfortunately, no data are available regarding medical, cardiovascular, and family history in these cases so these catastrophic events cannot be informatively interpreted. Nevertheless, until further data are available, it is recommended that children and adolescents who are candidates for tricyclic drugs receive careful cardiovascular evaluation and monitoring. Given the safety concerns with desipramine and the lack of demonstrated efficacy of the tricyclic drugs in general, their use as first-line treatment of major depressive disorder is not supported.

Recommendations for ECG monitoring include the following:

- Careful individual and family history of cardiovascular, neuromuscular, neurological, or other medical problems
- Baseline ECG prior to medication trial
- ECG repeated at 2 mg/kg and with dosage increments up to 5 mg/kg or therapeutic range
- Blood concentration monitoring with serum concentrations not to exceed 150 to 250 ng/mL (depending on the laboratory) ECG parameters: pulse < 120; blood

pressure < 130/90; QTc increased £ 25 percent above baseline; PR < 0.20.

Serotonin Reuptake Inhibitors SSRIs are used in the treatment of depressive disorders, obsessive-compulsive disorder, panic disorder and selective mutism. Fluoxetine, the oldest of the SSRIs, has a long half-life in adults and takes several weeks to reach clinical efficacy. A recent double-blind placebo-controlled study of 96 subjects with major depressive disorder showed significant superiority of fluoxetine over placebo.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS (SSRIs) SSRIs available in the United States include fluoxetine (Prozac), sertraline (Zoloft), paroxetine (Paxil), fluvoxamine (Luvox), and citalopram (Celexa). Open studies have suggested that more than two-thirds of patients treated for major depressive disorder with SSRIs have benefited. The SSRIs act by blocking reuptake of serotonin.

In general, SSRIs are better tolerated than tricyclic drugs and have a significantly reduced risk of cardiovascular toxicity. Common adverse effects include insomnia, gastrointestinal symptoms, and behavioral activation. This class of medications has significant effects upon the hepatic cytochrome P450 oxidase system; thus when SSRIs are used in combination with other agents, a potential for adverse effects through pharmacokinetic interactions exists.

Open trials have supported the efficacy of fluoxetine and fluvoxamine in juvenile-onset obsessive-compulsive disorder. Fluvoxamine was recently found to be more effective than placebo in a multicenter study of 120 children and adolescents with obsessive-compulsive disorder.

NONSELECTIVE SEROTONIN REUPTAKE INHIBITORS Clomipramine (Anafranil) is a nonselective serotonin reuptake inhibitor and tricyclic drug with antiobsessional properties. It has been studied in children and adolescents with obsessive-compulsive disorder and has been found to be more efficacious than placebo. Because of its tricyclic structure, it has the same potential adverse effects as the other tricyclic drugs; cardiovascular monitoring is necessary.

Other Antidepressants Bupropion (Wellbutrin), an aminoketone unrelated to the tricylic drugs, has antidepressant properties. Not formally approved for patients under 18, it has been studied in the treatment of attention-deficit/hyperactivity disorder. Nefazodone (Serzone) was reported to be beneficial in a small open study of depressed children.

Experience in the United States with monoamine oxidase inhibitors in children and adolescents is extremely limited, particularly in the prepubertal child. Due to their potential for toxicity, these medications are difficult to prescribe for most youngsters. Some adolescents with atypical depressive features or phobias may be candidates.

Anxiolytics and Sedative-Hypnotics Historically, anxiolytics and sedative-hypnotics have been used for a number of childhood and adolescent neuropsychiatric disorders such as night terrors, insomnia, and anxiety (Table 48.6-6). Unfortunately, to date, there have been few controlled studies of anxiolytic drugs, despite the prevalence of sleep and anxiety disorders in juveniles. Concerns about the diagnostic validity of pediatric anxiety disorders and adverse effects of benzodiazepines have contributed to the lack of studies of these agents.

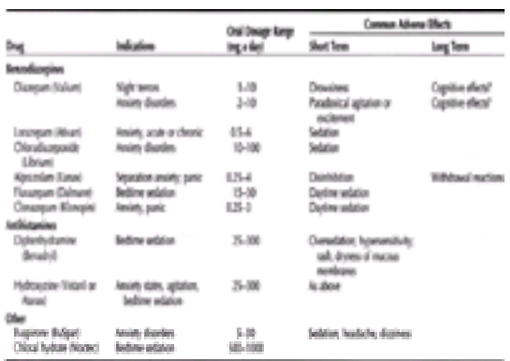

Table 48.6-6 Anxiolytics and Sedative-Hypnotics Drugs for Use in Children and Adolescents

Indications for use of benzodiazepines include sleep disorders, anxiety symptoms or disorders causing significant distress or impairment, and augmentation for other medication used in the treatment of mood and anxiety disorders, such as SSRIs. Night terrors (pavor nocturnus), a sleep disorder in young children during non–rapid eye movement sleep, frequently on arousal from stage 3 or 4 to lighter sleep stages is characterized by awakening in terror with physiological arousal. Although no controlled studies have been conducted, open reports have suggested beneficial response to low-dose diazepam (Valium).

In the nonpsychotic, non–substance-using adolescent population, benzodiazepines can be useful in short-term treatment of anxiety disorders that cause significant distress or impairment. Jovan Simeon and coworkers reported that alprazolam (Xanax) was superior to placebo in a double-blind study of 30 youths with overanxious and avoidant disorders. Though the results did not reach statistical significance, global clinical improvements were observed. K. Graae and coworkers in 1994 treated 12 youths (mean age 9.8 years) with mixed anxiety disorders with clonazepam (Klonopin) or placebo in an 8-week double-blind crossover study. Most children were rated as improved on clonazepam compared with placebo, but the results did not reach statistical significance.

Adverse effects of benzodiazepines include fatigue, dysphoria, and disinhibition. Cognitive effects have not been systematically assessed but are potentially problematic in use of the benzodiazepines.

Buspirone (BuSpar), a nonbenzodiazepine anxiolytic agent, is a partial serotonin (5-hydroxytryptamine [5-HT]) type 1a (5-HT $_{1A}$) agonist. It does not have muscle relaxant or anticonvulsant properties, nor is it likely to cause dependence. It has been used adjunctively with SSRIs in the treatment of obsessive-compulsive disorder and mixed depression-anxiety states. No controlled studies of buspirone in youths have been done, but a few open reports exist of use in separation anxiety disorder and mixed anxiety disorders. Adverse effects include headaches, tiredness, and gastrointestinal symptoms. Cognitive effects and abuse potential are minimal, especially compared with the benzodiazepines.

Mood Stabilizers

Lithium Mood stabilizers are indicated in children and adolescents with bipolar disorders. Lithium, historically the first studied for bipolar I disorder in adults, has been studied far less extensively for bipolar I disorder in youths (Table 48.6-7). Lithium is effective for both treatment of acute mania and for prophylaxis or augmentation in youths at serum concentrations of 0.6 to 1.4 mEq/L. Children with periodic or cyclic disturbances of behavior and mood, children with such disturbances who have relatives who were lithium responsive, and children with severely aggressive behavior disorders may also be lithium responsive. Plasma concentrations should be monitored regularly. Concentrations of 0.6 to 1.2 mEq/L are in the therapeutic range, but some patients may require higher levels for therapeutic efficacy. Kidney, endocrine (thyroid, calcium), and cardiac function must be monitored regularly.

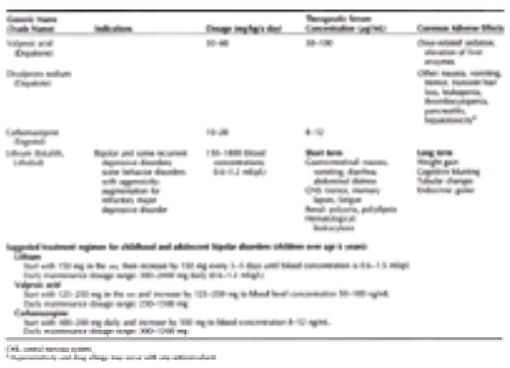

Table 48.6-7 Mood Stabilizers for Use in Children and Adolescents

Anticonvulsants In recent years, valproic acid (Depakene) or divalproex (Depakote) and carbamazepine (Tegretol) have been used to treat bipolar disorders in adults, and divalproex was recently approved by the FDA for bipolar I disorder in adults. As adult data and usage have accumulated, experience with these agents in children and adolescents has also grown. Open studies of valproic acid have shown beneficial effects in youths at dosages that give a serum concentration of 50 to 100 mg/mL. Valproic acid is being used as a first-line treatment for childhood and adolescent bipolar I disorder in clinical practice. Adverse effects include gastrointestinal symptoms and particularly hepatotoxicity, fatigue, tremor, and blood dyscrasias.

Carbamazepine, used for acute mania and rapid-cycling bipolar I disorder in adults, is a third-line option for youths. There are no controlled studies of carbamazepine for bipolar disorder, but it has been used openly in diagnostically heterogeneous children with behavior disorders. Adverse effects include blood dyscrasias, gastrointestinal symptoms, and neurotoxicity.

a-Adrenergic Receptor Agonists and b-Adrenergic Receptor Antagonists In recent years, other drugs available to the adult practitioner have been investigated in the pediatric and adolescent population. The b-adrenergic receptor antagonist propranolol (Inderal) has been found useful in some young patients with uncontrollable aggressive behavior, those with lithium-induced tremors, and those with certain anxiety states that have physiological symptoms such as posttraumatic stress disorder. b-Adrenergic receptor antagonists are contraindicated in diabetic patients, asthmatic patients, and patients with cardiac disease.

The a-adrenergic receptor agonist clonidine (Catapres) has been found useful in some studies in the treatment of Tourette's disorder and attention-deficit/hyperactivity disorder. Some evidence indicates that it is useful in decreasing the motor and verbal tics and associated attentional problems, hyperactivity, and lability of mood. Children who appear to be good candidates for the use of clonidine for attention-deficit/hyperactivity disorder are those with highly aroused impulsive characteristics; clonidine ameliorates these symptoms but may not be quite as efficacious for the cognitive effects of the disorder such as inattention and distractibility. An alternative is guanfacine (Tenex), a longer-acting a-adrenergic agonist that may be less sedating.

SUGGESTED CROSS-REFERENCES

Mental retardation is discussed in Chapter 34, pervasive developmental disorders in Chapter 38, disruptive behavior disorders in Chapter 40, tic disorders in Chapter 42, elimination disorders in Chapter 43, selective mutism in Section 46.4, and stereotypic movement disorder in Section 44.2. Psychiatric treatment of children is covered in Chapter 48. Personality assessment of adults and children is discussed in Section 7.5 and the neuropsychological and intellectual assessment of children in Section 7.4. Schizophrenia is discussed in Chapter 12, mood disorders in Chapter 14, anxiety disorders in Chapter 15, sleep disorders in Chapter 21, and biological therapies in Chapter 31.

SECTION REFERENCES

Alessi N, Naylor M, Ghaziuddin M, Zubieta JK: Update on lithium carbonate therapy in children and adolescents. J Am Acad Child Adolesc Psychiatry *33:*291, 1994.

Allen AJ, Leonard H, Swedo S: Current knowledge of medications for the treatment of childhood anxiety disorders. J Am Acad Child Adolesc Psychiatry *34:*976, 1995.

Biederman J: Sudden death in children treated with a tricyclic antidepressant. J Am Acad Child Adolesc Psychiatry *30:*495, 1995.

Biederman J, Baldessarini RJ, Wright V, Knee D, Harmatz J: A double-blind placebo-controlled study of desipramine in the treatment of attention deficit disorder: I. Efficacy. J Am Acad Child Adolesc Psychiatry *28:*777, 1989.

Birmaher B, Ryan N, Williamson D, Brent DA, Kaufman J: Childhood and adolescent depression: A review of the past 10 years, part II. J Am Acad Child Adolesc Psychiatry *35:*1575, 1996.

*Campbell M, Cueva JE: Psychopharmacology in child and adolescent psychiatry: A review of the past seven years. Part 1. J Am Acad Child Adolesc Psychiatry *34:*1124, 1995.

*Coffey B: Anxiolytics: Traditional and new drugs. J Child Adolesc Psychopharmacol *1:*57, 1990.

Coffey B, Miguel E, Savage C, Rauch S: Tourette's disorder and related problems: A review and update. Harvard Rev Psychiatry *2:*121, 1994.

Dulcan M: Using psychostimulants to treat behavioral disorders of children and adolescents. J Child Adolesc Psychopharmacol *1:*7, 1990.

Emslie G, Rush A, Weiberg A: A double-blind, randomized placebo controlled trial of fluoxetine in depressed children and adolescents. Arch Gen Psychiatry *54:*1031, 1997.

Flament MF, Rapoport JL, Berg CJ, Sceery W, Kilts C, Mellstrom B, Linnoila M: Clomipramine treatment of childhood obsessive-compulsive disorder: A double-blind controlled study. Arch Gen Psychiatry *42:*977, 1985.

Gadow KD, Sverd J: Stimulants for ADHD in child patients with Tourette's syndrome: The issue of relative risk (Review). J Dev Behav Pediatr *11:*269, 1990.

Geller B, Cooper T, Graham D, Marstellar FA, Bryant DM: Double-blind, placebo-controlled study of nortriptyline in depressed adolescents using a “fixed plasma level” design. Psychopharmacol Bull *26:*85, 1990.

*Green WH: *Child and Adolescent Clinical Psychopharmacology*. Williams & Wilkins, Baltimore, 1995.

Greenhill L: Attention deficit hyperactivity disorder: The stimulants. Child Adolesc Psychiatr Clin N Am *4:*123, 1995.

Hunt RD, Capper L, O'Connell P: Clonidine in child and adolescent psychiatry. J Child Adolesc Psychopharmacol *1:*87, 1990.

Jatlow PI: Psychotropic drug disposition during development. In *Psychiatric Pharmacosciences of Children and Adolescents*, C Popper, editor. American Psychiatric Press, Washington, DC, 1987.

Kumra S, Frazier J, Jacobsen LK: Childhood onset schizophrenia. Arch Gen Psychiatry *53:*1090, 1996.

Kumra S, Jacobsen LK, Lenane M: The spectrum of neuroleptic-induced movement disorders and extrapyramidal side effects in childhood onset schizophrenia. J Am Acad Child Adolesc Psychiatry *37:*221, 1998.

Leonard H, March J, Rickler K, Allen AJ: Pharmacology of the selective serotonin reuptake inhibitors in children and adolescents. J Am Acad Child Adolesc Psychiatry *36:*725, 1997.

Mannuzza S, Klein R, Bonagura N, Malloy P, Giampino T, Addalli K: Hyperactive boys almost grown up. Arch Gen Psychiatry *48:*77, 1991.

Mattes JA, Gittelman R: Growth of hyperactive children on maintenance regimen of methylphenidate. Arch Gen Psychiatry *40:*317, 1983.

Nurcombe B: Malpractice. In *The Comprehensive Textbook of Child Psychiatry*, M Lewis, editor. Williams & Wilkins, Baltimore, 1991.

*Riddle M, Claghorn J, Gaffney G, Greist J, Holland D, Ladbloom R, McConville B, Piggott T, Pravetz M, Walkup J, Yaryura-Tobias J, Houser V: Fluvoxamine for OCD in children and adolescents: A controlled trial. Proceedings of the annual meeting of the American Academy of Child and Adolescent Psychiatry, *12:*110, 1996.

Ryan N: Heterocyclic antidepressants in children and adolescents. J Child Adolesc Psychopharmacol *1:*21, 1990.

Ryan N: Pharmacotherapy of adolescent major depression: Beyond TCAs. Psychopharmacol Bull *26:*75, 1990.

Shapiro E, Shapiro AK, Fulop G, Hubbard M, Mandeli J, Nordlie J, Phillips RA: Controlled study of haloperidol, pimozide and placebo for the treatment of Gilles de la Tourette's syndrome. Arch Gen Psychiatry *46:*722, 1989.

Spencer T, Biederman J, Harding M, O'Donnell D, Faraone SV, Wilens TE: Growth deficit in ADHD children revisited: Evidence for disorder-related growth delays. J Am Acad Child Adolesc

Psychiatry *35:*1460, 1996.

Spencer T, Biederman J, Wilens T, Harding M, O'Donnell D, Griffin S: Pharmacotherapy of attention deficit hyperactivity disorder across the life cycle. J Am Acad Child Adolesc Psychiatry *35:*409, 1996.

*Storch DD: Medication-induced hypomania in Asperger's disorder. J Am Acad Child Adolesc Psychiatry *38:*110, 1999.

*Walsh ET, editor: *Child Psychopharmacology*. American Psychiatric Press, Washington, DC, 1998.

48.7 PARTIAL HOSPITAL AND AMBULATORY BEHAVIORAL HEALTH SERVICES

LAUREL J. KISER, PH.D., M.B.A., JERRY D. HESTON, M.D. AND DAVID B. PRUITT, M.D.

Partial hospitalization has been used historically as an umbrella term describing many varieties of alternative care. With the advent of managed care and health care reform, less costly and less restrictive treatment modalities are in increasing demand. To define these alternatives, which stretch along the continuum of psychiatric care between inpatient and outpatient office visits, three levels of ambulatory behavioral health care have been outlined. In this new environment, partial hospital programs (PHPs) are providing short-term, crisis-stabilization services as an alternative to inpatient care. Day treatment programs (DTPs) are serving the needs of children and adolescents with moderate-to-severe disorders who require interventions focused on improved level of functioning, skill-building, and disease management. Intensive outpatient programs (IOPs) function as intermediate or step-up programs from outpatient psychotherapy.

In the behavioral health delivery system, ambulatory behavioral health care (partial hospital, intensive outpatient services, etc.) is analogous to outpatient surgery. These services represent a specific segment along the continuum of care and are more intensive than outpatient services without the iatrogenic effects of hospitalization. Ambulatory behavioral health care services strive to provide intense, highly structured treatment. They do so by utilizing a variety of therapeutic modalities: individual, group, and family therapy; educational or vocational therapy; recreation and activity therapy; as well as medical and nursing services. This multimodal treatment requires an interdisciplinary staff typically drawn from psychiatry, psychology, social work, educational or vocational therapy, occupational or recreational therapy, and nursing. A working definition of ambulatory behavioral health care follows:

> Ambulatory behavioral health care is a time-limited, active treatment offering therapeutically intensive, coordinated, and structured clinical services that incorporate the benefits of a therapeutic milieu within a program and/or within the home/community.

HISTORY

The history of ambulatory behavioral health care for children and adolescents is brief. The first partial hospital programs were for adults in Russia in the late 1930s, Canada in the late 1940s, and England in the early 1950s. The Menninger Clinic introduced the concept to the United States when it opened its unit in the mid-1950s. In 1963 the Community Mental Health Center Act mandated comprehensive services including partial hospital care for children and adolescents. This movement had three main origins: deinstitutionalization, interest in the role of family and community, and movement to cost-effective treatment alternatives

Throughout the 1970s and 1980s and into the present, the modality has been severely underutilized. With only 2 percent of those in need of treatment receiving services in PHPs and the majority of programs operating below 75 percent capacity, underutilization is a major problem for the survival of many programs.

Three clinical issues contribute to underutilization of partial hospitalization. Firstly, providers find it difficult to serve moderately to severely disturbed children and adolescents in an open system with less restrictiveness and a less structured environment. Secondly, families need to make major commitments to family therapy, transportation, and to keeping a difficult child at home. Thirdly, for referral sources who are more accustomed to and better trained to treat severely dysfunctional patients in inpatient programs, a shift in the style and conceptualization of clinical practice is required.

Historically, other factors related to low utilization of partial programs include a lack of definition, criteria, and outcome data with consequent funding limitations. Reimbursement policies serve as a major impediment to wider use of partial hospital services. Traditionally, the vast majority of policies with mental health care benefits had better coverage for inpatient than outpatient services, and no coverage for partial hospitalization services. Therefore, instead of treating the child in the least restrictive, least expensive environment, the clinician and family were forced to utilize more restrictive, more expensive programs. Finally, the nature of the health care industry during the 1980s favored the creation of new hospital beds rather than partial hospital programs. Current industry trends, however, promise major changes in the service delivery system.

In the 1990s the emphasis of health care delivery returned to cost containment with increased interest in community, family-based alternatives, managed care, systems of care, and accountability. Managed care focuses on eliminating unnecessary inpatient services with the efforts directed toward development of systems favoring outpatient practices. Although early indications from managed care pointed to their desire to ensure quality and efficiency, fiscal management frequently translates into either rationing care or providing the cheapest services for the shortest length of time possible regardless of clinical decision-making and effectiveness data.

These trends have created major shifts in the delivery and funding of behavioral health care, including growth in the use of ambulatory behavioral health care, increased utilization of immediate, short-term partial hospital care for patient stabilization, and rapid development of a variety of intensive outpatient treatment options. New funding opportunities for ambulatory services include passage of federal regulations offering partial hospital services to Civilian Health and Medical Program of the Uniformed Services (CHAMPUS) enrollees, and the development of more inclusive and flexible benefit and payment plans.

However, severe pressures for cost containment are pushing some managed care organizations toward viewing partial hospital care in a similar light as inpatient care. These companies have begun applying medical necessity criteria to PHPs that are similar to those used for inpatient and approving only brief lengths of stay (3 to 7 days). Cost containment pressures are also driving the development and utilization of less intensive options, such as IOPs or multi-modal outpatient treatment. Approvals for care at this level are sometimes based on cost without adequate consideration of the treatment needs of the adolescent or child.

THEORETICAL ISSUES

There are six essential principles of care upon which the continuum of ambulatory mental health services derives its foundation:

1. Services are designed for children and adolescents who present with a psychiatric or substance dependence diagnosis and the need for treatment more intensive than outpatient office visits and less restrictive than 24-hour care.
2. Services provide a coordinated array of active therapies determined by an individualized treatment plan and based upon a comprehensive evaluation of patient needs.
3. Patients are treated in a manner that simulates real-life experience with the least amount of disruption to normal daily functioning (allowing for optimal transfer of therapeutic benefits to the natural environment).
4. Services require active involvement of the child or adolescent and the family.
5. Services are available on a consistent basis, augmented with 24-hour crisis back-up.
6. Services are cost-efficient.

From these six essential principles, two basic philosophies of care provide the core for child and adolescent ambulatory behavioral health services. Firstly, these services provide treatment to patients requiring intensive therapeutic intervention with the least amount of disruption to their normal daily functioning. This requires ambulatory services to define the limits between when patients can be treated safely in an outpatient setting and when they require the restrictiveness of hospitalization. Thus, ambulatory behavioral health services are designed to maintain a balance between treatment in the least restrictive environment and risk management.

Secondly, ambulatory behavioral health care encourages providers to take advantage of the open system inherent in the modality. This philosophy emphasizes the use of family supports and strengths and community agencies and programs. This philosophy of care has two major sequelae. Ambulatory services invite the patient and the patient's family to maintain a higher level of functioning than treatment settings that remove the patient from the home. Philosophically, this translates into a program of therapeutic interventions designed to maintain power within the parental or familial subsystem and to view that subsystem as competent in providing care for the child. Interventions are structured to alter dysfunctional patterns of interaction rather than to remove the child from these interactions. Translated into program policy, ambulatory behavioral health programs only use techniques and treatment strategies that a family can also implement at home.

Additionally, ambulatory programs promote reliance on community support systems and programs whenever possible. Benefits of this approach include limiting the development of dependence on the treatment team, expansion of and appropriate use of resources, decreased lengths of stay, and improved adjustment following discharge. This unique combination of providing “security and structure while simultaneously promoting patient responsibility and autonomy” may be the particular advantage of this modality.

The goals underlying ambulatory behavioral health care can be met in programs with a variety of theoretical orientations. Behaviorally oriented programs utilize principles of learning theory. Consequently, the goals of a behavioral program are to teach desirable behaviors and to extinguish undesirable ones. For example, the majority of partial hospital programs, although not operating from a behavioral orientation, use some form of level system for consequent patient behavior. Psychoanalytic programs focus on structuring a therapeutic environment in which behaviors and other symptoms are analyzed and interpreted. Another approach uses a psychoeducational orientation with the classroom as the basic structure for program design. “Students,” rather than “patients,” usually receive psychotherapy in conjunction with classroom instruction. Programs using systems theory emphasize the theoretical proposition that many different areas of a child's life influence functioning: biological, intrapsychic, family, peers, school, work, neighbors, religious affiliation, and others. Within a systems orientation, programs are challenged to assess problems and intervene on a multilevel basis. Some programs are based on a medical model in which all patients receive psychiatric diagnoses and treatment is prescribed by a child and adolescent psychiatrist.

These orientations differ in the manner in which they view the child, the child's problems, and the treatment program. Although many theoretical orientations are possible, within an organization the model used should be consistent across the continuum of care. Conflicting models established between components of care may cause confusion for the staff as well as the patient and family.

GOALS OF TREATMENT

The clinical functions of ambulatory behavioral health care differ according to the level of care offered. Partial hospitalization, and other intensive nonresidential treatment options function to provide short-term crisis intervention as an alternative treatment to hospitalization or an intensive transition setting to shorten a hospital stay. Less intensive ambulatory modalities provide support, maintenance, symptom reduction, and skill-building to prevent relapse, longer-term hospitalization, or residential care. Finally, multimodal outpatient services provide a step up for patients for whom one or two office visits per week are insufficient. Another goal or function of ambulatory behavioral health services can be extensive evaluation involving observation, identification of problem areas, diagnosis, and formulation of treatment plans.

CLINICAL ISSUES

Indications The need for ambulatory behavioral health services to define target populations is essential for continued viability of the modality. The continuum of ambulatory behavioral health care represents three levels of care, each of which targets populations with differing needs. Admission decisions are based on matching interventions and services with the needs of the patient in terms of intensity, structure, and accessibility.

Treatment organizations define their target population by establishing admission criteria. From this programmatic standpoint, evaluation to determine appropriateness for ambulatory behavioral levels of care involves the following five factors: (1) psychiatric signs and symptoms, (2) level of impulse control, (3) level of functional impairment, (4) parental support, (5) physical health, and (6) ability to pay. Specific criteria for the three levels of ambulatory behavioral health care are presented in Table 48.7-1.

Criteria	Level 1	Level 2	Level 3
Psychiatric signs and symptoms	Current diagnosis; severe to disabling symptoms related to acute or chronic condition.	Current diagnosis; moderate to severe symptoms related to acute or chronic condition.	Current diagnosis; moderate symptoms related to acute or chronic condition.
Level of impulse control	Limited control of suicidal, homicidal, runaway ideation, and abuse of drugs or alcohol with high risk for confinement.	Moderate control of suicidal, homicidal, runaway ideation, and abuse of drugs or alcohol with some risk of confinement.	Good control of suicidal, homicidal, runaway ideation, and abuse of drugs or alcohol with low risk for confinement.
Level of functional impairment	Severe difficulty functioning in multiple areas including school, social activities, community, and home.	Marked difficulty functioning in one or more areas including school, social activities, community, and home.	Moderate difficulty functioning in at least one area including school, social activities, community, and home.
Parental support	Able to provide adequate control and support at home during evenings/weekends with close monitoring and frequent assistance from staff.	Able to provide adequate control and support at home during evenings/weekends with monitoring and assistance from staff.	Able to provide adequate control and support at home during evenings/weekends with intermittent monitoring and assistance from staff.
Physical health	No requirement for 24-hour medical care.	No requirement for 24-hour medical care.	No requirement for 24-hour medical care.
Ability to pay	Economic resources available to support care at this level.	Economic resources available to support care at this level.	Economic resources available to support care at this level.

Derived from data supplied by: L.J. Kiser, M.D.; J.D. Heston, M.D., D.C.; Thomas, M.D., and C.M. Hanley, M.D.

Table 48.7-1 Admission Criteria for Child and Adolescent Ambulatory Behavioral Health Care

As pressures increase to manage children and adolescents with unstable, crisis presentations in less restrictive settings, it is critical for providers to develop decision-making tools for patient placement that take risk management into account. For child and adolescent providers, family or community support and structure become a major factor in the decision-making process. A decision matrix, assessing level of impulse control and parental support and structure, is a valuable tool for assessing appropriateness for admission (Fig. 48.7-1). Level of impulse control can be viewed along a continuum with some patients demonstrating no problems and other patients requiring a structured, locked treatment environment due to severe problems with behavior control. On the other axis, the continuum of family functioning (support and commitment to treatment) is evaluated. It is expected that families will work with the treatment team to encourage their child to comply with program limits. Families who are unable or unwilling to participate in this manner are less likely to provide the daily support and structure needed by the patient, who is therefore unlikely to derive significant benefit from treatment in an ambulatory behavioral health care setting.

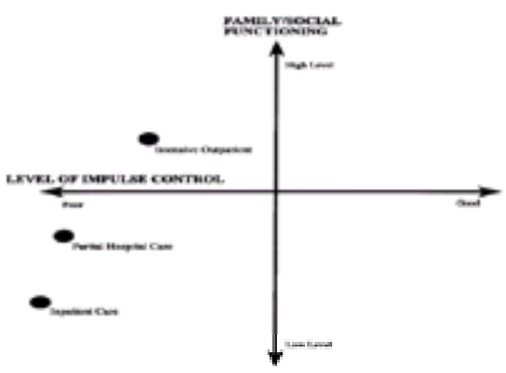

FIGURE 48.7-1 Decision matrix: risk management assessment. (Adapted from Kiser LJ, Heston JD, Millsap PA, Pruitt DB: Testing the limits: Special treatment procedures for child and adolescent partial hospitalization. Int J Partial Hosp *7:*39, 1991.)

Program of Therapeutic Activities The clinical components, that is, each patient's daily activities, are based upon an individual treatment plan developed by the treatment team. Continuous monitoring of the patient's response to treatment dictates alterations in this plan. Therapeutic progress is monitored frequently by the multidisciplinary staff and periodically a treatment update is prepared and the patient's case is reviewed by the entire treatment team. Progress to date on problems is reviewed, new problems are discussed, and progress towards transition to a less intensive level of care or discharge is monitored. A distinguishing characteristic of ambulatory behavioral health care is the intensity of diverse clinical components. These components most often include individual, group, family, medical management, activities, and education.

Individual Individual therapy is offered by most ambulatory behavioral health providers in the United States, usually on a weekly or semiweekly basis with a mean of 1.60 hours per week. Like individual therapy in other settings, it provides the patient with an opportunity to develop a close, interpersonal relationship with an adult other than a parent and an opportunity to work on intrapsychic problems.

Individual therapy in ambulatory behavioral settings differs from outpatient individual psychotherapy because of the wealth of information available to the therapist through contacts with or observations of the patient in other parts of treatment. An example of this occurs in a partial hospital setting when a depressed child denies withdrawal from peers but is frequently observed sitting alone during milieu periods. Differences between individual psychotherapy in ambulatory and inpatient settings are also significant and mainly result from the openness of the ambulatory system. Individual therapy in ambulatory behavioral health care can utilize material from daily encounters with parents, siblings, and peers to address dysfunction in these areas.

Group Group psychotherapy plays an important role in ambulatory behavioral health settings by allowing peers to develop the ability to deal with problems through the expression of feelings and experiences in a safe environment. This therapy also provides an opportunity for work on group dynamics, including cohesion, roles, and norms. As with all group work, it is important that the structure and content of the group be appropriate for the developmental levels of the patients involved. As with individual therapy, therapists in ambulatory behavioral health care have direct knowledge about patient functioning in multiple areas, thus allowing immediate and specific feedback to the group members.

Family The use of family therapies by many programs is consistent with the research demonstrating that family structure is predictive of successful completion of treatment and follow-up. Family involvement in treatment may take a variety of forms, including traditional family therapy, multifamily groups, parent education classes, parent support groups, marital therapy, and in-home services.

Family therapy provides an opportunity to focus on family dynamics in order to facilitate change in the home environment. Within ambulatory behavioral health care this therapy allows intensive work on family conflict and family structure on a day-to-day basis. Because patients remain part of the family while participating in treatment, new skills can be practiced and evaluated continuously. Another important benefit of family therapy in ambulatory behavioral health care is the ability to establish consistent rules and limits for the child both at home and in treatment. Work with the family and community helps to extend the therapeutic benefits of treatment into the patient's overall familial and social milieu. Finally, one powerful difference between family treatment in this modality and that of an inpatient unit is the emphasis on maintaining parental authority. For example, in partial hospitalization parents are involved in every level of decision-making and problem-solving; parents are in effect viewed as co-therapists.

Therapeutic Activities The extent of therapeutic activities, such as recreation and leisure skill-building, and also the availability of a programmatic therapeutic milieu is dependent on the structure and intensity of the services offered. However, at all levels of ambulatory behavioral health care, providers emphasize the development of an active, therapeutic milieu. In ambulatory behavioral health services that do not incorporate a programmatic milieu, the treatment plan prescribes active interventions to extend milieu-based therapeutic benefits to the home, workplace or school, and community, as needed. Recreation, movement, art, occupational, and milieu therapies, as well as community-based activities, provide staff and peer support for appropriate behavior in social settings. Patients learn constructive ways to use leisure time in activities that foster team work, cooperation, and task commitment. Frequent involvement in community events is consistent with the ambulatory behavioral health care belief that children and adolescents should remain active members of the community.

Education Just as much of a child's life revolves around school and related activities, much of the intensive treatment for children and adolescents revolves around school issues and educational therapy. Different ambulatory services meet the needs for educational services in very different ways. Intensive, full-day programs offering a school component provide a minimum of 3 hours of education daily. Short-term programs with lengths of stays under 2 weeks may provide educational experiences based upon a homebound model. Finally, less intensive ambulatory services, such as after-school programs or IOPs, may simply focus on enhancing school performance by offering tutoring or study skills. Regardless of the level of care offered, ambulatory behavioral health care providers utilize the treatment environment to help children to deal appropriately with the frustrations encountered in the school setting, including dealing with authority and assuming responsibility for work. They also understand the importance of schooling in the lives of children and adolescents and thus view active school liaison as a major component of treatment. Obtaining fiscal support for the educational portion of treatment is an area of concern, with piecemeal funding often necessary and with programs thus required to meet several sets of standards and answer to several local and state agencies.

Medical Management The extent of medical involvement in the management of patients in ambulatory behavioral health care varies depending upon the intensity and theoretical background of the program. Medical involvement within ambulatory behavioral health care ranges from medical supervision for patients with unstable, crisis needs to medical consultation for patients with low levels of risk and only moderate symptoms.

Recent trends are toward emphasizing aspects of the medical model in the most intensive ambulatory services such as partial hospital programs. As various forces work toward decreasing rates of hospitalization for children and adolescents, more severely disturbed patients and those with combined medical and psychiatric disorders will be referred to partial hospital programs. In such programs psychiatrists function as attending physicians and members of the treatment team. Under this model the physician can be seen as the team leader, orchestrating the treatment provided. The use of George Engel's biopsychosocial model of medical practice is particularly congruent with the philosophy of ambulatory behavioral health care. Physician involvement is not limited to the specifics of medical management, but includes substantial input into program administration and direction. The cost disadvantage may be counteracted by the preference of insurance companies and Medicaid to reimburse for medically directed services.

Thus, within programs offering short-term symptom reduction and crisis stabilization, medical management includes an assessment of the patient's general medical condition and behavioral health status. Both medical and psychiatric diagnoses are assigned. This may require direct physical examinations done by a physician associated with the program or, in programs that are more community-based, consultation and liaison with the patient's primary care physician or pediatrician. It is the responsibility of nurses and physicians to identify specific health problems that may be important factors in the overall functioning of the patient. These problems may be treated directly or referred to a community physician. Referrals often serve to educate and guide families on accessing needed health care services and on how to use these resources effectively.

Pharmacotherapy, another aspect of medical management, may be practiced within the structure of ambulatory behavioral health care with advantages over other therapeutic settings. As on an inpatient unit, observations from multiple viewpoints are available to determine specific indications for medications. The ability to monitor new medications or changes in regimens is augmented by daily contact with the patient and information about the patient's status in many different activities. This use of medication within a structured environment is frequently seen as an advantage over medication management developed on an outpatient basis. Also, in distinction to an inpatient hospitalization, the family is able to provide or learn to provide ongoing feedback about medication response, adverse effects, compliance, or other concerns about medication.

In programs that are designed as intensive outpatient services, psychiatrists may be contracted to function as consultants to provide specific, limited services. These services typically include initial psychiatric evaluation and diagnosis as well as the monitoring of pharmacotherapy. The psychiatrist is not a part of the treatment team and is usually not involved in the day-to-day care of the patients or administration of the program. Although the advantages of this model are financial and logistical, saving the program the cost of a full-time specialist in a profession that is relatively undersupplied, the main disadvantage is the lack of physician input into the overall treatment of patients.

Finally, as managed care grows and continues to influence psychiatric care, physicians may be expected to help programs, patients, and their families negotiate appropriate financial coverage for treatment. This assistance may come in clarifying diagnoses and outlining treatment options. Frequently physicians in ambulatory behavioral health care are called upon to participate in physician-to-physician reviews of cases that have been tagged as questionable by the managed care company. The psychiatrist will be called upon to document rationale for treatment in ambulatory behavioral health care and should be familiar with medical necessity

criteria and effective arguments for obtaining treatment appropriate to the specific needs of each patient.

Special Populations Proponents of ambulatory behavioral health care suggest that it is a viable treatment mode for many special populations (e.g., juvenile offenders, patients with eating disorders, and abusers of alcohol and drugs) as well as for nonpsychiatric populations (e.g., those with head trauma, epilepsy). Several reports of PHPs designed for adolescent offenders or adolescent antisocial behavior have appeared in the literature. Treatment emphasis for offenders appears to be intensive group therapy and social, recreational, and special educational services with generally positive outcomes noted for these teens who are known to be very difficult to treat.

Another special population described in the literature is preschoolers. Preschool PHPs and therapeutic nurseries for young children place special emphasis upon play, communication and language, socialization, skill-development, and parent-child interaction. Young children benefiting from treatment in a therapeutic preschool include those with developmental delays, victims of physical and sexual abuse, as well as other severe emotional/behavioral disturbances.

Health care reform and the interest in development of cost-effective treatments has resulted in programs that address the psychosocial as well as physical needs of chronically ill pediatric patients. Certain patients with eating disorders, sickle cell disease, asthma, diabetes, and other disorders may benefit from specialized treatment in ambulatory settings that incorporates health care professionals from appropriate pediatric specialties.

Limitations and Complications As various theoretical, ethical, economic, and other factors continue to discourage the use of hospitalization, ambulatory providers will be challenged to manage patients whose behavioral symptoms and impulsivity have escalated to the point that special treatment procedures are indicated. Policies and procedures for dealing with patients displaying aggressive, acting-out behaviors, noncompliance with program rules, and suicidal or runaway ideation are a necessary part of programming. Special treatment procedures designed for use in child and adolescent ambulatory settings must adequately address the issue of safety and involve the family and the community while maintaining the patient in the least restrictive environment.

Management of acutely and severely disturbed patients necessitates policies that address seclusion, physical holding, and restraints. Often PHPs simply do not offer these services on the basis that they are appropriately utilized only on inpatient units. Other child and adolescent PHPs use quiet room and physical holding policies that do not violate the mandate of the program, that is, procedures that can also be used by parents or guardians in the home environment. The quiet room and physical holding are used sparingly with adolescents; immediate transfer to a secure facility may be indicated for an aggressive or acutely psychotic, disruptive adolescent. Manual and chemical restraints are typically not used in PHPs or other ambulatory settings.

Just as programs must develop policies that are compatible with the program's theoretical basis and intensity of treatment to help patients control their behavior, they must address similar issues with other problems both within the treatment setting and at home. For example, PHPs may be able to treat patients with significant suicidal ideation if they are able to maintain appropriate levels of observation and plan for continued supervision after program hours using family therapy, telephone contact, beeper services, or other interventions. Runaway patients may similarly be placed in ambulatory treatment with policies for structure after hours and the ability to liaison with community resources such as runaway homes or other community emergency shelters. Thus, intensive programs that are designed to serve as alternatives to inpatient treatment are able to treat severely disturbed patients using policies that address program structure, family resources, and community involvement.

Because of the intensity of programming, the developmental level and the severity of illness of patients treated, and the lack of restrictiveness, most ambulatory behavioral health care services operate with very low staff-to-patient ratios. Obviously the level of staffing required is dependent on the intensity of care offered and the degree of disturbance within the target population. Within PHPs, a typical ratio is one staff person to fewer than four patients, with a recommended range from 1:3 through 1:6. For afterschool programs or IÒPs, a staff-to-patient ratio of 1:8 to 1:10 is probably sufficient.

Even with small staff-to-patient ratios, the high intensity of the programs results in staff stress and burnout. A fuller explanation lies in the philosophy of the approach: maintaining severely affected patients in the least restrictive environment.

Crisis Management and Hospital Liaison Ambulatory behavioral health care services that attempt to shorten or eliminate hospitalization must have adequate policies and procedures to handle the emergencies that prompt immediate admission to treatment and the inevitable crises that develop during treatment. Specific details of crisis management should be identified by each ambulatory behavioral health service depending upon the level of care offered. At the most intensive level, ambulatory providers should have an integrated emergency system that allows immediate access to current clinical and treatment information. The following recommendations meet these crisis management needs: (1) an on-call child and adolescent psychiatrist available 24 hours per day, 365 days a year; (2) additional on-call services to help families resolve conflicts that do not require hospitalization or major changes in treatment plans, staffed primarily by partial hospital professionals; and (3) an established affiliation with both pediatric and psychiatric inpatient units. In the case of hospital-based programs this should be addressed in policies regarding transfer between PHPs and inpatient units. In free-standing PHPs formal affiliations with hospitals must be made so that admissions can occur quickly and safely for patients with acute emergencies.

Less intense ambulatory programs must also develop emergency plans. These might include policies to admit patients to the PHP within a number of days and consultation with a psychiatrist to assist with decisions regarding emergency management and possible hospitalization. Coverage of crisis situations while a child or adolescent is in care can be handled by a 24-hour crisis call service.

DISCHARGE AND AFTERCARE SERVICES

As integrated systems of care for child and adolescent become more prevalent, planning transitions between levels of care often replaces discharge planning for ambulatory services. Transition or discharge planning remains an integral part of treatment planning, beginning during the initial intake and continuing throughout the treatment process as transition or discharge goals are formulated and clarified. During the course of treatment, progress toward transition or discharge is continually monitored by the treatment team with input from the patient's family.

The following principles govern the planning process: (1) patients receive treatment in the least restrictive environment that provides the structure and intensity necessary; (2) patients receive treatment for the shortest time possible in order to reach maximum treatment benefit; and (3) planning is done in a fashion that maximizes a successful transition to the patient's new setting (either more or less structured). Thus transition or discharge planning involves attention to issues of termination, liaison with community resources (inpatient settings, school settings, etc.), and follow up to encourage compliance with transition recommendations. The case of Tammy, a 16-year-old, African-American girl admitted to a PHP following a suicide attempt, is presented to illustrate the use of partial hospitalization in emergency care.

Tammy was first seen in a pediatric emergency room on a Saturday morning by a psychiatry resident who determined that, although she had continued suicide ideation, she did not have a specific plan and could be treated without hospitalization. Over the weekend the resident contacted Tammy and her family twice by phone and arranged for her to be admitted to a partial hospital program on Monday morning. During her first few days in the program she was monitored closely by the staff and her suicide potential was evaluated at the end of each day by the medical director. During the first week her suicide ideation resolved in response to the support and structure of the program. At the end of her second week in the partial hospital program she was stabilized, started on antidepressant medication, and transitioned to outpatient family and group therapy following stated commitments to follow up by the patient and her parents. The use of partial hospitalization prevented a psychiatric inpatient admission for this acutely symptomatic adolescent.

Other aspects of ambulatory behavioral health care are highlighted in the case of Jamie, a 15-year-old boy who was transitioned to a medium-intensity day treatment program. During a hospitalization for a life-threatening suicide attempt, his family's managed care company decertified his admission saying he was no longer at acute risk for suicide. Due to continued serious depressive symptoms and chronic family dysfunction, he was determined to be inappropriate for routine outpatient management. He was admitted to a day program with a strong family systems orientation. Over the course of his 8-week treatment he was able to develop a therapeutic alliance with his individual therapist, and significant family restructuring was accomplished. The consulting child and adolescent psychiatrist managed Jamie's medication and regularly monitored his suicide potential. At the end of 8 weeks, his depressive symptoms were decreased to the extent that he could transition to outpatient therapy and return to school successfully. This use of the ambulatory behavioral continuum allowed for prompt discharge from the hospital with continued consolidation of progress in a highly structured system.

RESEARCH AND EVALUATION

As partial hospital and intensive outpatient services play a more central role in behavioral health care delivery, it is critical to establish quality standards and to differentiate these services offered from other services currently on the market. In order to provide a summary of ambulatory behavioral health care's indications and limitations, the advantages and disadvantages of the modality are presented in Table 48.7-2.

Table 48.7-2 Advantages and Disadvantages of Ambulatory Behavioral Health Care for Children and Adolescents

Quality of Care Partial hospital and other nonresidential hospital alternatives share many common treatment goals and techniques and differ in important ways from both inpatient treatment and outpatient office visits. Accordingly, the measurement of quality in this modality presents some unique challenges. Using the underlying philosophies, the unique therapeutic aspects, the distinctive treatment procedures as well as the historical problems and promises of the field, specific quality indicators can be formulated.

Several specific quality indicators of effectiveness for partial hospital and intensive outpatient services include valid and accurate admission criteria, standards of care, and low negative, critical treatment incidents. As an alternative treatment modality, it is incumbent upon partial hospital and other ambulatory behavioral health services to establish admission criteria that reliably and validly discriminate between patients who need intervention at a more intensive level than outpatient visits yet at a less restrictive level than 24-hour care. Having targeted an appropriate patient population, the standardization of services rendered as well as special treatment procedures needs to ensure that the care offered provides the safety and security necessary to provide intensive treatment in a nonresidential setting.

Additional unique quality indicators include specific aspects of cost efficiency and patient satisfaction. Providing intensive treatment within a nonresidential environment mandates the use of outside resources and the reliance on strengths within the family and community for augmenting treatment, both of which translate into cost-efficient care. For ambulatory behavioral services, two specific areas of patient satisfaction become applicable: the patient's feelings regarding receiving treatment as an outpatient and the family's sense of contentment with their role in the treatment process.

Finally, for all ambulatory providers viability is a critical factor, both at the level of the individual program and industry wide. Increased utilization, cost containment, and internalization of quality are integral to survival of the modality. It is incumbent upon leaders in the field as well as managed care to improve patterns of utilization if alternative, intermediate forms of behavioral health care are to fulfill a role in the overall continuum of mental health services.

Treatment Outcome Research on treatment effectiveness and outcome in psychiatry involves a complex constellation of factors. Major problems are encountered, including providing an objective, standardized definition of effectiveness and solving methodological constraints such as difficulties of selection procedures and randomization. The scope of the research literature on the effectiveness of child and adolescent partial ambulatory behavioral health care is not yet broad enough to provide definitive information. A growing body of research is available on treatment effectiveness or outcome with child and adolescent samples. A summary of the studies conducted to date on adolescents in partial hospital programs indicates generally positive outcomes with gains in relationships and school performance. Decreased symptoms have been found following treatment in PHPs with effects evident for up to 5 years following discharge. Review of studies on treatment outcome with children suggests that 66 to 90 percent of patients demonstrated improvement and successful return to community-based schools. Some studies suggest that family functioning (structure and stability) are major factors in improvement while others conclude that the younger children showed greater benefits from treatment in partial hospital settings than did older patients. Following a review of outcome studies, improvements were reported in behavioral, academic, and family role functioning following partial hospitalization with a variety of aftercare services needed. In a 5-year follow-up of severely disturbed children treated in a multimodal day program, improvements in global functioning, behavior, self-esteem, peer relations, and academic performance were maintained with parental cooperation of primary importance to positive outcomes.

Movement to partial hospitalization away from traditional treatment modes would be alarming and irresponsible without treatment comparisons. It is essential to look at the treatment effectiveness of partial hospital settings in comparison to the effectiveness of other modalities. Again, published research has not been as extensive as desirable. The literature on child and adolescent populations contains three studies that compare treatment outcomes in partial hospitalization and residential settings and one study comparing partial hospital with inpatient care for substance abusing adolescents. Overall, the findings suggest that partial hospitalization is equally effective with some patients on some variables, is more effective with some patients on some variables, and is less effective with some patients on some variables. Clearly, advocates of partial hospitalization must continue to compare outcomes from different treatment modalities, from various partial program models, as well as compare treatment outcomes of specific populations.

One of the major difficulties encountered in the area of outcome measurement is the lack of standardization of outcome variables and instruments. Six areas of outcome measurement are recommended for ambulatory behavioral health care, including cost of services, utilization of services postdischarge, severity of symptoms, level of functioning, patient ecology (family-community support network), and satisfaction. Standardized assessment of these variables is a goal with the objective of providing industry-wide information regarding expected outcomes.

Ambulatory behavioral health care represents an important segment of the continuum of psychiatric care available to children and adolescents with a wide array of disorders and allows these patients treatment in less restrictive settings than hospitals. In the current era of health care reform, this and other ambulatory mental health services are likely to face increased demand. As policies encourage such options it is necessary to be aware of differences between PHPs, DTPs, and IOPs so that patient needs can be matched with service variables such as theoretical orientation, targeted populations, types of treatments offered within the program, staffing patterns, and the degree of structure provided. Issues related to quality of care, therapeutic outcome, and cost-effectiveness must also be considered. With careful review it is possible to identify intensive PHPs that serve as alternatives to inpatient hospitalization, DTPs or IOPs that function as transitions between highly restrictive treatments and outpatient services, and programs that serve as outpatient alternatives.

Although partial hospitalization is an established concept and has a significant history, the field is in transition as is health care in general. In the years to come several questions will arise as ambulatory behavioral health care providers face the broad challenges brought by health care reform. How will partial hospital programs integrate with other children's services such as special education systems, juvenile courts, and state departments of child protection and human services? How will they fit into large health care organizations to provide prevention and related services? It is hoped that the closing of the twentieth century will provide answers to these important questions and the further development of this exciting mode of treatment.

SUGGESTED CROSS-REFERENCES

Psychotherapies are discussed in Chapter 30, including behavior therapy in Section 30.2, group psychotherapy in Section 30.4, family therapy in Section 30.5, and an evaluation of psychotherapy in Section 30.11. Child psychiatry is discussed in Chapter 32, Chapter 33, Chapter 34, Chapter 35, Chapter 36, Chapter 37, Chapter 38, Chapter 39, Chapter 40, Chapter 41, Chapter 42, Chapter 43, Chapter 44, Chapter 45, Chapter 46, Chapter 47, Chapter 48 and Chapter 49.

SECTION REFERENCES

American Association for Partial Hospitalization: *Partial Hospitalization Industry Overview*, Alexandria, VA, 1994.

American Association for Partial Hospitalization: *Outcomes Measurement Protocol*. Alexandria, VA, 1994.

Baenen RS, Stephens-Parris MA, Glenwich DS: Outcome in psycho-educational day programs: A review. Am J Orthopsychiatry *56:*263, 1988.

*Block BM, Arney K, Campbell DJ, Kiser LJ, Lefkowitz PM, Spear SK: *Standards and Guidelines for Child and Adolescent Partial Hospitalization*. American Association for Partial Hospitalization, Alexandria, VA, 1991.

Comer R: Day treatment of adolescents: An alternative to institutionalization. J Counsel Dev *64:*74, 1985.

Corkey CL, Zimet SG: Relationships with family and friends in young adulthood. Int J Partial Hosp *4:*97, 1987.

Cornwall A, Blood L: Inpatient versus day treatment for substance abusing adolescents. J Nerv Ment Dis *186:*580, 1998.

Engel GL: The clinical applications of the biopsychosocial model. Am J Psychiatry *137:*533, 1980.

Freed PE, Rudolph S: Protecting partial-hospitalization patients from suicide. Psychiatr Serv *34:*14, 1998.

Gavin LA, Roesler TA, Brenner AM: Day treatment for pediatric patients with medical and psychiatric needs. Continuum *3:*95, 1996.

Gaylor ML: Treating the adolescent offender in a rural partial hospitalization program. In *Proceedings of the Annual Conference on Partial Hospitalization*, Boston, MA, Federation of Partial Hospitalization Study Group, 1979.

Glasscote R, Kraft AM, Glassman S, Jepson W: Partial hospitalization for the mentally ill, Washington, D.C. Joint Inf Serv *28:*448, 1977.

Gresenz CR, Liu X, Sturm R: Managed behavioral health services for children under carve-out contracts. Psychiatr Serv *49:*1054, 1998.

Grizenko N, Sayegh L, Papineau D: Effectiveness of multimodal day treatment for children: A one-year follow-up. Continuum *1:*115, 1994.

Grizenko N: Outcome of multimodal day treatment for children with severe behavior problems: A five-year follow-up. J Am Acad Child Adolesc Psychiatry *36:*7, 1997.

*Hoge MA, Farrell SP, Munchel ME, Strauss JS: Therapeutic factors in partial hospitalization. Psychiatry *51:*199, 1988.

Kettlewell PW, Jones JK, Jones RH: Adolescent partial hospitalization: Some preliminary outcome data. J Clin Child Psychol *14:*130, 1985.

Kiser L, Pruitt D: Partial hospital care. In *Managing Care, Not Dollars*, RK Schreter, SS Sharfstein, CA Schreter, editors. American Psychiatric Press, Washington, DC, 1997.

Kiser LJ, Millsap PA, Hickerson SG, Heston JD, Nunn W, Pruitt DB, Rohr M: Results of treatment one year later: Child and adolescent partial hospitalization. J Am Acad Child Adolesc Psychiatry *35:*81, 1996.

Kiser LJ, Wagner BD, Knight MA: Quality indicators for partial hospitalization. Behav Healthc Tomorrow *3:*31, 1994.

*Kiser LJ, Culhane DP, Hadley TR: The current practice of child and adolescent partial hospitalization: Results of a national survey. Continuum *2:*81, 1993.

Kiser LJ, Heston JD, Millsap PA, Pruitt DB: Testing the limits: Special treatment procedures for child and adolescent partial hospitalization. Int J Partial Hosp *7:*37, 1991.

*Kiser LJ, King R, Lefkovitz PM: A comparison of practice patterns and a model continuum of ambulatory behavioral health services. Psychiatr Serv *50:*605, 1999.

*Kiser LJ, Lefkowitz PM, Kennedy LL, Knight M: Visions: The continuum of ambulatory mental health services. Behav Healthc Tomorrow *2:*14, 1993.

Kiser LJ, Pruitt DB, McColgan EB, Ackerman BJ: A survey of child and adolescent day treatment programs: Establishing definitions and standards. Int J Partial Hosp *3:*247, 1986.

Kolko DJ: Multimodal partial/day treatment of child antisocial behavior: Service description and multilevel program evaluation. Continuum *2:*3, 1995.

Kotsopoulos S, Walker S, Beggs K, Jones B: A clinical and academic outcome study of children attending a day treatment program. Can J Psychiatry *41:*371, 1996.

Krizay J: *Partial Hospitalization. Facilities, Cost, and Utilization*. American Psychiatric Association Office of Economic Affairs, Washington, DC, 1989.

Leibenluft E, Leibenluft RF: Reimbursement for partial hospitalization: A survey and policy implications. Am J Psychiatry *145:*1514, 1988.

Leone P, Fitzmartin R, Stetson F, Foster J: A retrospective follow-up of behaviorally disordered adolescents: Identifying predictors of treatment outcome. Behav Disord *11:*87, 1986.

Masters KJ: Using a coordinated treatment system to minimize child psychiatric hospitalization. J Am Acad Child Adolesc Psychiatry *36:*4, 1997.

Matzner FJ, Silvan M, Silva RR, Weiner J, Bendo J, Alpert M: Intensive day program for psychiatrically disturbed truant adolescents. Am J Orthopsychiatry *68:*135, 1998.

*Novello JR: Day hospital treatment of adolescents. In *The Short Course in Adolescent Psychiatry*, JR Novello, editor. Brunner/Mazel, New York, 1979.

Prentice-Dunn S, Wilson DR, Lyman RD: Client factors related to outcome in a residential and day treatment program for children. J Clin Psychol *21:*189, 1981.

Prevost J: Partial hospitalization—dynamics of underutilization. In *Proceedings of the Annual Conference on Partial Hospitalization, San Diego*. American Association for Partial Hospitalization, Boston, 1981.

Raskin R, Novacek J, Bahlinger D, Firth L: A model for evaluating intensive outpatient behavioral health care programs. Psychiatr Serv *47:*1227, 1996.

Rogers SJ, Lewis H: An effective day treatment model for young children with pervasive developmental disorders. J Am Acad Child Adolesc Psychiatry *28:*207, 1989.

Velasquez JS, Lyle CG: Day versus residential treatment for juvenile offenders: The impact of program evaluation. Child Welfare *64:*145, 1985.

Weithorn LA: Mental hospitalization of troublesome youth: An analysis of skyrocketing admission rates. Stanford Law Rev *40:*40, 1988.

48.8 RESIDENTIAL AND INPATIENT TREATMENT

MARK DEANTONIO, M.D.

The use of residential and inpatient treatment facilities for children and adolescents has changed significantly since the mid-1980s. There has been a dramatic increase in use of both facilities, including a tripling of admissions of adolescents to inpatient units. This change has challenged the child and adolescent psychiatrist to rethink the appropriate use of these interventions. The advent of managed care and tighter control of dispersement of mental health resources has further modified inpatient and residential placement utilization. The decision to use these interventions, however, remains powerful and dramatic, for it makes the judgment that a child or adolescent must be separated from parents or caregivers to be given appropriate treatment.

A child and adolescent inpatient psychiatric unit may be located in a general hospital or in a free-standing psychiatric hospital. It must meet the same credential and licensing requirements as other hospital settings. It usually has stricter requirements in terms of patient rights. The unit may be either locked or unlocked. A locked unit can restrain and seclude patients. Due to the higher acuity of involuntary patients, it has psychiatric nursing staffing equivalent to an intensive care unit. While an unlocked unit has less nursing staff, usually it has the same complement of the rest of the interdisciplinary team, which includes psychiatry, psychology, social work, nutrition, occupational therapy, recreational therapy, and school staff. Pediatric and neurological consultations are available. The task of inpatient hospitalization has changed over the years with refined diagnostic and treatment formulation and increasing financial constraints. Currently the goals of inpatient psychiatric hospitalization for children and adolescents are to safely contain patients while intensively evaluating them and implementing treatment that will bring them to a less restrictive setting as quickly as possible.

Residential treatment is less clearly defined. In the literature the terms *residential treatment* and *inpatient psychiatric units* have been used interchangeably, but residential treatment centers should be defined as a separate entity. Such a center is a facility that houses children and adolescents 24 hours a day but is not licensed as a psychiatric hospital. It is not necessarily directed by a physician and does not have the full complement of staff that an inpatient psychiatric unit must have for licensure. A psychiatrist is part of the staff but usually more in a consultative role than as a leader of the treatment team. There are also significantly fewer psychiatric nursing staff. Often social workers and other licensed mental health professionals are the primary staff. Because of the prominence of nonmedical staff, children and adolescents in such facilities are not as acutely disturbed, because interventions such as one-to-one staffing, seclusion, and restraints are not available. Additionally, aggressive medication cannot be managed in such settings. The goals of residential placement are also different. While children and adolescents are often evaluated, the primary focus is to develop better interpersonal group and social skills so that patients can either return to their parents or live independently as adults. The length of stay is in the range of 6 months to years.

HISTORY

By the late 1800s children with severe psychiatric disorders were being described in the literature. However, the concept of hospitalizing children for psychiatric disorders began in the early 1920s in the United States. After World War I, a worldwide influenza epidemic caused a large number of postencephalitic children. These children exhibited emotional lability, agitation, and cognitive deficits requiring containment in a hospital setting. Therefore, pediatric behavioral units were established. These units were run by pediatricians and pediatric neurologists and served a primarily custodial function. The 1950s saw an increased understanding of child psychopathology as well as individual, group, milieu, and family psychotherapeutic interventions. With this new understanding, child psychiatrists became involved in the inpatient units, and their emphasis changed from custodial care to diagnostic evaluations and individually focused treatment interventions.

The development of residential treatment facilities evolved along different lines. In the early 1700s orphanages were organized to attend to orphans and abandoned children. In the mid-1800s, two orientations for residential treatment developed: residential schools for the education and rehabilitation of the cognitively impaired and reform or training schools. With the industrial revolution came urbanization of families and the subsequent need for multiple family members to work, often resulting in fragmentation of families and lack of supervision of children. Issues of abuse and delinquency among children were identified. The field of social work was developed to address chaotic families and abused and delinquent children. It was felt that society and the children would be best served if reform schools were focused on rehabilitation. However, by the 1930s, these institutions changed their focus to detention with a punitive orientation. In the 1950s, the orientation of residential programs was revised to multimodality therapeutic treatment because of new understanding of the psychopathology and psychotherapy of children, paralleling the changes in the inpatient psychiatric settings. Additionally, the psychological and social work literature indicated that custodial care of children without attention to individual issues and needs actually impaired social, emotional, and cognitive development. Orphanages were closed, and a model of foster and group home care was used. By the 1960s the actual treatment in inpatient psychiatric units and residential treatment centers was similar, except for the medical and nursing orientation of the former and the educational and social work orientation of the latter. Analyses of treatment programs in the two settings did not differentiate different diagnostic groups, treatment interventions, or goals, except for a slightly shorter length of stay in the inpatient settings.

In the 1970s another major change occurred, with a shift away from treatment in residential settings to treatment in inpatient psychiatric units in response to societal questioning of putting children and adolescents in long-term residential care. The priority changed to the goal of reintegration into families and communities. Additionally, an understanding of the biological nature of many disorders of childhood developed in child psychiatry. Autistic disorder, attention-deficit disorders, mood disorders, psychotic disorders, and anxiety disorders were being shown to have a biological component with specific medical treatments. As specific medical treatments were articulated for these disorders, both private insurance and federal programs were willing to pay for hospital treatment of them. This change in orientation led to the dramatic increase in inpatient hospitalization of children and adolescents in the 1980s and a decrease in use of residential treatment centers.

In the late 1980s a serious reevaluation of the use of inpatient psychiatric services for children and adolescents began. The cost of inpatient child and adolescent treatment consumed most of the money spent on mental health in the United States. Third-party payors, including the federal government, implemented managed care review to decrease the frequency of hospitalization and length of stay. Capitated contracts with psychiatric institutions were also implemented, which also reduced the use of inpatient hospitalization and length of stay. In the early 1990s it was revealed that some psychiatric hospitals used unethical and even criminal practices in the hospitalization of children and adolescents. Financial and ethical concerns forced a rethinking of the use of inpatient and residential hospitalizations. Critical reviews of uses and goals of inpatient hospitalization began to be published in the child psychiatry literature. The American Academy of Child and Adolescent Psychiatry published a policy statement for inpatient hospitalization. Parental and patient advocacy groups demanded deemphasis of inpatient hospitalization.

Currently the focus in treatment of children and adolescents is to implement the least restrictive intervention for evaluation and treatment. Instead of inpatient treatment being the first intervention, it is now used only as the most restrictive component in an integrated system of care that includes outpatient, partial hospitalization, and residential treatment.

PHILOSOPHY

Inpatient and residential treatment of children and adolescents is now seen as a component of an integrated system of care for psychiatrically disturbed minors. They must interact dynamically with outpatient and partial hospital treatments rather than be exclusive from them or compete with them. Treatment interventions must be individualized to the specific needs of the disturbed child or adolescent and the family. The ultimate goal of both residential and inpatient settings must be to stabilize patients so that they can move to the least restrictive treatment setting in an accountable, timely fashion, with both patient and family engaged in this treatment plan.

ADMISSION CRITERIA

As part of the reevaluation of inpatient hospitalization, specific admission criteria have been developed to ensure appropriate use. The following criteria are derived from recommendations of the American Academy of Child and Adolescent Psychiatry and individual psychiatric facilities. First, the patient must be assessed by a psychiatrist, who must be a child and adolescent psychiatrist for patients 13 years of age or younger. For patients 14 years of age and older, a psychiatrist with documented adolescent training and demonstrated competence in work with adolescents is required. The assessment can also be based on an evaluation by the admitting psychiatrist at the time of admission of the findings of an appropriate clinician. If a psychiatrist does not perform the initial evaluation, such an evaluation must be done within 24 hours of admission. Second, there must be a diagnosis of a psychiatric disorder as defined by the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders*, (DSM-IV). Because of this disorder, the patient must be an imminent risk to self or to others, be gravely disabled, or be experiencing medical complications arising from a psychiatric illness. Imminent risk to self is evidenced by a suicide attempt that presents a severe medical threat; a clear, lethal suicide plan that presents a serious threat to the patient's life; or significant changes in behavior that indicate hopelessness or helplessness (e.g., increased withdrawal, social isolation, decline in school performance accompanied by serious family conflict disruption, or loss of a significant other). Imminent risk to others is evidenced by homicidal or assaultive behaviors that present a serious risk to others; homicidal ideation with a specific, clear, and realistic plan; or significant changes in behavior, loss of judgment, and impulse control accompanied by threatening behavior that cannot be safely contained. Grave disability is evidenced by an impairment in judgment, orientation, memory, and affective functioning. An acute psychotic state, major depressive episode with significant vegetative signs, severely debilitating panic or anxiety, and inability to maintain adequate nutrition because of a mental disorder are clinical conditions that can cause such disability. Medical complications arising from a psychiatric illness are unstable, life-threatening conditions complicating the management, evaluation, or treatment of a psychiatric illness. This may be manifested by one of the following clinical signs and symptoms: serious adverse side effects from psychotropic medications, failure to maintain adequate nutrition, uncontrolled seizure disorder, diabetes mellitus uncontrolled because of the patient's mental disorder, or the use of psychotropic medication requiring constant skilled observation because of a coexisting medical condition. Third, the failure of appropriately intensive treatment in a less restrictive setting must be clearly documented. If a less restrictive setting has not been attempted, it must be documented that either such a setting would put the patient or others at significant risk or the patient's family is severely disrupted and unstable.

The child or adolescent and the family must be informed of the reasons for inpatient admission and why such an admission is the least restrictive intervention. Effort should be made to have the patient and the family voluntarily agree to hospitalization. When hospitalization is involuntary, the patient and family should be so informed.

Admission criteria for residential treatment programs have been less well defined, and there are no consistent standards to date. Prior to admission the child or adolescent must be evaluated by a licensed mental health professional who must diagnose a psychiatric disorder as defined by DSM-IV. The psychiatric disorder must be severe enough to significantly impair daily functioning in interpersonal relations, social interactions, family relations, or school performance. The disorder must be treated in the residential setting because the home environment failed to contain the patient safely and to allow successful outpatient treatment. As in the inpatient admission, the child or adolescent and family must be informed of the indication for admission and why it is the least restrictive option.

GOALS

The primary goal of inpatient hospitalization is to stabilize the issues that formed the basis for admission, so that the patient can be moved to a less restrictive setting. To this end, the hospital uses therapeutic interventions to keep the patient from harming self or others or causing severe milieu disruption. Second, a comprehensive, timely, multidisciplinary diagnostic evaluation must be conducted. This evaluation must address physical, neurological, psychiatric, cognitive, academic, social, interpersonal, and family issues. Third, the family, posthospital caretakers, or both must be actively engaged in understanding how to manage the patient out of the hospital safely. Finally, there must be active networking between the hospital and the outpatient treatment setting to ensure continuity in treatment. For these goals to be achieved in the shortest time, they must be addressed at admission and reviewed on a daily basis throughout hospitalization.

The primary goal of residential treatment is to create a therapeutic living environment that safely contains the patient while allowing for multidisciplinary interventions so that the patient can return to the home or function independently as an adult. While the focus of inpatient care is on containment and stabilization, in residential care the focus is on immersing the patient in an environment that addresses interpersonal, social, educational, and familial deficits. A comprehensive, multidisciplinary diagnostic evaluation must be conducted. This evaluation must also address physical, neurological, psychiatric, cognitive, academic, social, and interpersonal and family issues. Since children and adolescents who enter residential placements have a much higher incidence of psychiatric hospitalizations, school failure, and past out-of-home placements, past records and evaluations must be obtained so that evaluations are neither needlessly repeated nor critical information lost. From the evaluation, specific multidisciplinary treatment goals must be developed so the patient can function successfully in a less restrictive setting. The time frame for meeting these goals can be months to years, though they must be reviewed weekly. It is expected that upon discharge from a residential treatment center, the child or adolescent's issues will be ameliorated but not resolved. Therefore, there must be continuity of care with an outpatient program, family, or other care providers.

ASSESSMENT

A complete multidisciplinary assessment must be done on inpatient and residential treatment patients. A common reason for treatment failure or lack of treatment progress is the lack of a clear diagnostic formulation along biopsychosocial lines. Most children and adolescents have noncontributory physical, neurological, and laboratory findings related to inpatient hospitalization; however, a significant minority have physical issues that need to be addressed. A complete physical and neurological examination is needed during an inpatient stay, and relevant findings must be followed up. There has been discussion in the literature about the necessity of screening laboratory tests and neurological evaluations. With the better understanding of psychopharmacological interventions in inpatient settings, and the resultant more aggressive use of medications, assessment of blood indexes, liver function, renal function, thyroid functions, and cardiac conduction is more relevant. Pregnancy status must be determined when considering medication in adolescent females. Performance of pelvic and rectal examinations must be considered carefully; these examinations should not be routine in children and adolescents. When sexual abuse is of concern, the examination should be done by a physician, nurse practitioner, or physician assistant specifically trained in sexual abuse examinations of children and adolescents. Examinations should be done at other times when specific complaints or behaviors necessitate them. With the continued pressure on decreasing cost and shortening length of stay, comprehensive neurological evaluations come into question. As with laboratory tests, most children and adolescents do not show clinically significant abnormalities on electroencephalogram (EEG) or brain imaging studies. These studies should not be done just to be complete but should be done to resolve a specific clinical concern that would change the treatment.

The human immunodeficiency virus (HIV) epidemic adds to the challenges of psychiatric treatment of children and adolescents in inpatient and residential settings. Because of confusion and controversy around HIV antibody testing in the inpatient setting, the American Academy of Child and Adolescent Psychiatry has made the following recommendations concerning HIV and inpatient psychiatric treatment:

1. HIV infection or risk of HIV infection is not a reason to deny a child or adolescent admission to a psychiatric hospital.
2. All patients admitted to a psychiatric hospital should be treated as if they were HIV infected; even if all patients were tested on admission for antibodies to HIV, one is still not sure of their infection status because of the latency period in antibody response.
3. Because of the convergence of the tuberculosis and HIV epidemics, inpatient staff should be knowledgeable about tuberculosis exposure and infection control procedures.
4. Children and adolescents should *not* be routinely tested for HIV infection; testing should be considered for children and adolescents at risk for infection including, but not limited to, homosexual and bisexual males, those who have taken drugs intravenously, those with HIV-infected mothers, those with histories of sexual abuse, those with multiple sexual partners, and those with sexually transmitted diseases.
5. HIV-positive or potentially positive patients do not require individual rooms or toilet facilities or isolation from the milieu or peer groups; they should participate in all aspects of the inpatient treatment program.
6. All HIV testing requires informed consent. It is not enough just to have a consent form signed; it must be documented that both the patient and the person authorizing testing are fully informed of the consequences of both a positive and a negative result. Specific state laws exist regarding confidentiality of HIV antibody testing results and who authorizes consent for minors. Regardless of legal issues, it is strongly recommended that children and adolescents be actively involved in this consent process.
7. Results of patients' HIV antibody tests must be kept confidential; information about HIV status should be available only to staff who need to know the status to ensure appropriate care of the child or adolescent while an inpatient. These recommendations also apply to residential settings.

The psychiatric assessment combines information from the patient, parents or caregivers, past records, and outpatient clinician and pediatrician with observation and

evaluations of the patient in the inpatient setting. This data collection should focus on delineating the issues leading to the necessity for inpatient hospitalization and making a diagnostic formulation. It is necessary to inquire about frequency, intensity, duration, and circumstances of symptoms and problematic behaviors. To assess the functional impairment caused by the patient's symptoms and difficulties one must inquire about the patient's internal distress, interference with social and academic activities, impact on ongoing development, and impact of the behavior on others. Data collection should concern both the patient's difficulties and symptoms and the patient's strengths, talents, areas of positive adjustment, and support systems. If such strengths are known, they can be integrated into the treatment plan to facilitate stabilization and growth.

The interview with the child or adolescent inpatient provides information about the patient's own perception of the issues leading to hospitalization and an assessment of the patient's overall developmental and mental status. This information is critical because the crisis precipitated by inpatient hospitalization often results from a disparity in the understanding of the relevant issues between the child and adolescent and the parent or caregiver. For example, an adolescent may be admitted to a hospital because of a serious suicide attempt. In interviewing the parents, they may explain that the attempt was due to perceived increasing social withdrawal and school failure. In interviewing the adolescent, he or she may admit for the first time that auditory hallucinations commanded the suicide attempt. For the mental status examination, the clinician observes and assesses the physical appearance; manner of relating; orientation to time, place, person; mood; affect; motor behavior; content and form of thought, including hallucinations and delusions, speech and language; intelligence; attention; memory; neurological functioning; and judgment and insight.

Standardized interviews and rating scales have become powerful tools for consistent assessment of psychiatric disorders and symptoms in clinical research settings. However, it is not known how useful they are in inpatient treatment settings. Symptom rating scales focusing on depressive, psychotic, attentional, anxiety, or obsessive-compulsive symptoms may have utility in quantifying the severity of a symptom and serving as a baseline against which response to a therapeutic intervention can be established. Standardized assessments and rating scales should never be a substitute for face-to-face clinical interviews. They cannot prioritize the significance of symptoms or diagnoses in the context of a child or adolescent's personality style and interpersonal and family relational style, and hence the treatment will not be individualized and focused. Psychological testing, both cognitive and projective, should not be done routinely. It should be performed when it can provide specific answers that cannot be elicited through interview and observation. However used appropriately, such testing can be instrumental in refining treatment strategies. For example, a patient who has a history of failing school and oppositional behavior with new onset of depression and suicidal ideation should be treated with different multimodal strategies depending whether the cognitive level of functioning is in the normal or the borderline range. Projective testing of patients who cannot articulate thoughts and feelings or refuse to do so may add important data to clarify the diagnosis. Speech and language testing can also be used when suspected receptive and expressive language deficits are not documented. Educational assessment is needed if academic failure or disruptive behavior in the school setting is part of the issues around hospitalization. A comprehensive evaluation of the family is also critical. Each family member's strengths and weaknesses in interaction with the patient and whether they add or detract to family stability must be documented. Assessment of family members' communication, ability to resolve conflict, and ability to solve problems is important. Issues of neglect and abuse need to be clarified as well as the family's attitude and response to having their child hospitalized and the degree to which they support this decision.

After the interdisciplinary team has made their assessments, a diagnostic conference should be held. This should take place 5 to 7 days after admission. A diagnostic formulation should use a biopsychosocial model, which entails documenting a psychiatric diagnosis or diagnoses per DSM-IV. Then the biological, psychological, social, and interpersonal issues and patterns that caused the crisis leading to hospitalization must be identified. The patient's physical and psychosocial strengths and resources must be documented for use in promoting stabilization.

This assessment process should be similar in a residential placement. However, more time should be taken in coming to the diagnostic formulation, as the goal of treatment is focused on more complex interpersonal social, educational, and familial deficits, which take more time to clarify. Because of this, a diagnostic conference should take place after 30 days.

TREATMENT PLAN

After the diagnostic formulation is made, the multidisciplinary team documents a treatment plan. Hospital accreditation organizations now require documented problem-oriented treatment plans, which are updated weekly. Essential problems need to be abstracted from the formulation. The problems should not be a diagnosis such as “depression” or a category such as “family,” but specific symptoms, behaviors, emotions, or dysfunctions that must change for the patient to be treated in a less restrictive level of care. These specific problems should address issues from different areas of the biopsychosocial formulation but should be kept to a limited number that can be effectively addressed and ameliorated. For example, a 13-year-old adolescent who is hospitalized after a severe suicide attempt with a 2-week history of increasing depression in the context of parental divorce and conflict, academic failure, and increasing social isolation could have the following essential problems: (1) suicidal ideation with impulse to kill self, (2) feelings of hopelessness and helplessness, (3) conflicted family dynamics, (4) recent academic failure, and (5) increasing social withdrawal. Under each essential problem one notes the treatment intervention, its frequency, and the member of the interdisciplinary team responsible for it. The purpose of the treatment should be briefly stated. Next, a treatment goal for the essential problem is specified that can be achieved in 1 week. An example, for the essential problem of suicidal ideation with impulse to kill self; would be one-to-one observation, 24 hours a day, by nursing staff (identified by name). The purpose would be to keep patient from harming self. The treatment goal for the week would be no self-injurious acts. A second treatment or intervention would be psychotherapy five times a week with the child psychiatrist (by name) responsible for this intervention. The purpose of the treatment would be to have the patient understand why he or she feels like killing self. A treatment goal for the week would be for the patient to develop a written plan to safely manage feelings of wanting to hurt self, with assistance of the child psychiatrist. The same process is completed for each essential problem. Patient and family awareness of, and participation in, the treatment plan must be documented. Formally involving the patient and family in developing the plan increases their participation and motivation to be part of the treatment process.

Every week the multidisciplinary team meets to review the treatment plan. In that meeting they review each essential problem and the progress of the weekly treatment goal. If the problem has not been resolved, a new treatment goal must be formulated for the next week. If new essential problems arise, they can be documented with the proposed intervention and goal as described above. Additionally, the diagnosis, estimated length of stay, discharge plan, and plan for outpatient follow up should be reviewed weekly. Twenty-four hours prior to discharge, a final discharge plan must be documented by the multidisciplinary team, with specific recommendations from each member of the interdisciplinary team. The outpatient follow-up, living arrangement, and discharge medications should be listed. It should also be documented that the parents or other responsible party understood the discharge instructions and recommendations.

In a residential treatment center the treatment planning process functions the same way. The treatment plan should be reviewed weekly, but the weekly goals may reflect more-incremental change.

TREATMENT MODALITIES

For an inpatient hospital or residential treatment center to be effective in treating children and adolescents, it must use the full range of therapeutic modalities. Milieu therapy, psychopharmacotherapy, individual therapy, group therapy, family therapy, recreational therapy, occupational therapy, and a school program must be present. Use of these therapies is individualized by the interdisciplinary treatment plan.

The concept that the actual hospital or residential environment can be seen as a therapeutic instrument has been acknowledged since the 1950s. In the inpatient setting, the milieu primarily needs to be safe and containing. It is the primary responsibility of the nursing staff to manage the milieu. In the past, milieu had specific orientation, often behavioral or psychodynamic, to which the patient had to adapt. Interacting with the ward structure the patient's maladaptive coping strategies would be elicited and therefore be a focus of intervention. While this kind of identified milieu orientation continues to be seen in residential treatment centers, it has been largely revised in the inpatient setting because of the changed goals of hospitalization. The milieu must now be more flexible so that each patient's individualized treatment program can be implemented. The nursing staff develops an individualized treatment plan that addresses the specific treatment interventions articulated in the multidisciplinary treatment plan. Thus, some patients may have a more behavioral treatment program in which they are earning points or stars for specific behaviors while others may discuss feelings with staff for 30 minutes three times a day with no reinforcement beyond the patient-staff interaction. Safety and containment are critical to inpatient units; therefore the use of seclusion and restraints is necessary. When used appropriately and safely, these interventions promote impulse control in the patient and reduce behavior that could be severely dangerous to the patient and others. Research has shown that appropriate use of seclusion and restraints can decrease the length of hospital stay and help engage the patient in the treatment process.

Because the goals of residential treatment address long-term change in interpersonal, social, and family interactions, there is a more specifically defined milieu. It is usually designed to re-create a maturing, consistent environment that contains firm limits and boundaries but tolerates expression of feelings and promotes competency in problem solving. The milieu philosophy can range from behavioral to psychodynamic. Often part of the treatment is seen in the conflict created in

patients as they struggle with the structure of the environment and the specific issues it evokes from their past living situations.

Psychopharmacological interventions have been in use since the beginning of child and adolescent inpatient units in the 1920s. Initially, barbiturates were used to sedate agitated patients. Psychostimulants were used in the 1930s to address symptoms of hyperactivity. Until the 1970s, medication intervention for children and adolescents tended to focus on sedating agitation and improving impulsivity and inattention. Anxiety, mood, and even psychotic symptoms were treated with other psychotherapeutic modalities. Since the 1980s there has been a revolution in the understanding of the use of psychotropic medications for specific disorders in children and adolescents. Protocols have been developed to use medications in children and adolescents effectively. Medication intervention is implemented after the evaluation and recommendation of the treating psychiatrist and written informed consent by the parents. The medication should address specific target symptoms of a diagnosed disorder and should be documented in the interdisciplinary treatment plan. The hospital progress notes should give the indication for the medication and note that the parents and patient have been told the indication and the risk and benefits of the psychopharmacological intervention. Positive effects and adverse effects need to be documented as well as the rationale for dosage changes. Children and adolescents are often resistant to medication interventions. Repeated explanation and using the parents to reinforce compliance may be necessary. In both the weekly interdisciplinary meetings and daily rounds the psychiatrist should get input from all members of the interdisciplinary team about the effects of psychopharmacological interventions.

In residential treatment centers, every resident needs psychopharmacological evaluation. Because in many residential settings, psychiatrists only have a consultative role, some patients may be overlooked unless they exhibit agitated, disruptive, or self-injurious behaviors. A psychiatrist must be part of the regular treatment planning meetings so all cases can be reviewed for medications.

Individual psychotherapy remains an important treatment in inpatient and residential settings. In the inpatient setting the type of psychotherapeutic intervention should be directed by the multidisciplinary treatment plan. Supportive, cognitive, behavioral, insight-oriented, and psychodynamic approaches can all be used to address essential problems. The psychotherapeutic modality should be individualized to the patient's needs at the time. An adolescent who is hospitalized for severe agitation after a rape may initially require a supportive and cognitive behavioral orientation. As the patient is stabilized and moves to an outpatient setting, a psychodynamic orientation may be used if the trauma brings up more dynamic issues. In the residential setting, again the orientation of the individual psychotherapy should reflect the needs and abilities of the child or adolescent. Because of the long-term nature of residential treatment, the therapy may change orientation over time. Initially it may be more supportive and cognitive-behavioral. As time progresses and the child or adolescent matures or develops greater insight, a shift can be made to an insight-oriented or psychodynamic approach.

Family therapy is a powerful and essential part of both inpatient and residential treatment. Active involvement of the family in the therapeutic process decreases both length of stay and frequency of rehospitalization. Often some failure in family interactional dynamics precipitated hospitalization or residential treatment. Because of this, staff tends to view the family in adversarial terms, as family failure is seen as the cause of the hospitalization or residential placement. The challenge is to make the family feel an active part of the treatment team. While familial dysfunctional dynamics must be addressed, equal effort must be made to see the strengths in the family. Intervening in the dysfunctional dynamics and supporting and developing the family strengths further motivates patients in their other modalities of treatment. Seeing that one's family can change instills hope in the child or adolescent and confidence in the therapeutic process. All dysfunctional family issues do not need to be resolved in the inpatient setting. The issues need to be identified and therapeutic interventions initiated. The importance of clear communication of the family assessment and treatment strategies to the outpatient therapist must be stressed.

Residential treatment contains the same issues and goals as the inpatient setting. The treatment with the family must be more intensive and longer term if the goal is having the child or adolescent return to the family, because the need to place an adolescent or child in a residential treatment setting reflects more profound chaotic or dysfunctional family interactions.

Group therapy in an inpatient and residential setting is another powerful tool for addressing social and interpersonal issues in children and adolescents. It also becomes a place where the patient can process issues brought up in the milieu with peers and staff and can mirror appropriate family interaction. Therapy groups can also have specific foci so that patients learn to use peer feedback and support in managing and understanding their issues. Occupational therapy programs are critical in demonstrating and evaluating adaptive functioning of the patients. Recreational therapy demonstrates social interactive skills and deficits and promotes the ability to work together to solve tasks. Both occupational and recreational therapy contribute substantially to the treatment program by being less confrontational and focusing on the patient's strengths.

Educational evaluation is necessary in the inpatient setting, and an educational program during the inpatient stay is ideal. Children requiring inpatient hospitalization have a significant incidence of learning disabilities and school failure. For the patient to progress, reasons for school failure, clarification of learning disabilities, or both must be elucidated. Initiation of an Individual Education Plan, which must be developed by the local school board for severely emotionally disturbed children and adolescents under federal mandate, can be initiated in the inpatient setting. In the residential setting, the educational program is as significant a part of the patient's experience as it would be if the patient were at home. Whether the school setting is part of the residential center or separate, the teaching staff must be part of the multidisciplinary treatment team so they can follow through with treatment interventions and goals.

OUTCOME OF HOSPITALIZATION

Past studies of the effectiveness of inpatient hospitalization may no longer be relevant because of the stricter admission requirements, shorter length of stay, and reevaluation of goals of hospitalization. However, inpatient hospitalization should not be seen as an endpoint of treatment. It should be considered the most restrictive and intense intervention in the context of a comprehensive mental health care continuum that includes outpatient, intensive outpatient, partial hospitalization, inpatient hospitalization, and residential interventions. Patients now move more fluidly between the different treatment settings, so the efficacy of hospitalization must be seen in the context of the entire treatment program. The changes of the late 1980s, while painful and chaotic to clinicians, have had a positive effect in better defining the purpose, goals, and treatment of inpatient and residential treatment centers for children and adolescents.

SUGGESTED CROSS-REFERENCES

Normal child development is discussed in Section 32.2. Normal adolescent development is discussed in Section 32.3. Disruptive behavior disorders are discussed in Chapter 45. Schizophrenia with childhood onset is discussed in Chapter 47. Other aspects of psychiatric treatment of children and adolescents are discussed in the other sections of Chapter 48.

SECTION REFERENCES

American Psychiatric Association: Official actions: Guidelines for psychiatric practice in public sector psychiatric inpatient facilities. Am J Psychiatry *151:*797, 1994.

Baker BL, Blacher J, Pfeiffer S: Family involvement in residential treatment of children with psychiatric disorder and mental retardation. Hosp Community Psychiatry *44:*561, 1993.

Bath H: The physical restraint of children: Is it therapeutic? Am J Orthopsychiatry *64:*40, 1995.

Bickman L, Foster EM, Lambert EW: Who gets hospitalized in a continuum of care? J Am Acad Child Adolesc Psychiatry *35:*74, 1996.

Cotton NS: The developmental-clinical rationale for the use of seclusion in the psychiatric treatment of children. Am J Orthopsychiatry *59:*442, 1989.

Cotton NS: Seclusion as therapeutic management: An invited commentary. Am J Orthopsychiatry *65:*245, 1995.

Dicker R, Morrissey RF, Abikoff H, Alvir JMJ, Weissman K, Grover J, Koplewicz HS: Hospitalizing the suicidal adolescent: Decision-making criteria of psychiatric residents. J Am Acad Child Adolesc Psychiatry *36:*769, 1997.

Erker GJ, Searight HR, Amanat E, White PD: Residential versus day treatment for children: A long-term follow-up study. Child Psychiatry Hum Dev *24:*31, 1993.

Foster EM: Does the continuum of care improve the timing of follow-up services? J Am Acad Child Adolesc Psychiatry *37:*805, 1998.

Geraty R: Administrative issues in inpatient child and adolescent psychiatry. J Am Acad Child Adolesc Psychiatry *28:*21, 1989.

Goldston DB, Daniel SS, Reboussin BA, Reboussin DM, Kelley AE, Frazier PH: Psychiatric diagnoses of previous suicide attempters: first-time attempters, and repeat attempters on an adolescent

inpatient psychiatry unit. J Am Acad Child Adolesc Psychiatry *37:*924, 1998.

Goodman Zimet S, Farley GK, Zimet GD: Home behaviors of children in three treatment settings: An outpatient clinic, a day hospital, and an inpatient hospital. J Am Acad Child Adolesc Psychiatry *33:*56, 1994.

*Harper G: Focal inpatient treatment planning. J Am Acad Child Adolesc Psychiatry *28:*31, 1989.

Irwin M: Literature review. In *Psychiatric Hospitalization of Children*, JL Schulman, M Irwin, editors. Charles C Thomas, Springfield, IL, 1982.

*Jemerin JM, Phillips I: Changes in inpatient child psychiatry: Consequences and recommendations. J Am Acad Child Adolesc Psychiatry *27:*397, 1988.

King CA, Hovey JD, Brand E, Ghaziuddin N: Prediction of positive outcomes for adolescent psychiatric inpatients. J Am Acad Child Adolesc Psychiatry *36:*1434, 1997.

King CA, Hovey JD, Brand E, Wilson R, Ghaziuddin N: Suicidal adolescents after hospitalization: Parent and family impacts on treatment follow-through. J Am Acad Child Adolesc Psychiatry *36:*85, 1997.

King RA: Practice parameters for the psychiatric assessment of children and adolescents. J Am Acad Child Adolesc Psychiatry *34:*1386, 1995.

*Leon SC, Uziel-Miller ND, Lyons JS, Tracy P: Psychiatric hospital service utilization of children and adolescents in state custody. J Am Acad Child Adolesc Psychiatry *38:*305, 1999.

Malone RP, Luebbert JF, Delaney MA, Biesecker KA, Blaney BL, Rowan AB, Campbell M: Nonpharmacological response in hospitalized children with conduct disorder. J Am Acad Child Adolesc Psychiatry *36:*242, 1997.

Masters KJ: Using a coordinated treatment system to minimize child psychiatric hospitalization. J Am Acad Child Adolesc Psychiatry *36:*566, 1997.

Mikkelson EJ, Bereika GM, McKenzie JC: Short-term family-based residential treatment: An alternative to psychiatric hospitalization for children. Am J Orthopsychiatry *63:*28, 1993.

*Miskimins RW: A theoretical model for the practice of residential treatment. Adolescence *25:*867, 1990.

Mundy P, Robertson J, Greenblatt M, Robertson M: Residential instability in adolescent inpatients. J Am Acad Child Adolesc Psychiatry *28:*176, 1989.

*Nurcombe B: Goal-directed treatment planning and the principles of brief hospitalization. J Am Acad Child Adolesc Psychiatry *28:*26, 1989.

Pfeiffer SI, Strelecki BA: Inpatient psychiatric treatment of children and adolescents: A review of outcome studies. J Am Acad Child Adolesc Psychiatry *29:*847, 1990.

Pottick K, Hansell S, Gutterman E, White HR: Factors associated with inpatient and outpatient treatment for children and adolescents with serious mental illness. J Am Acad Child Adolesc Psychiatry *34:*425, 1995.

Ricciuti A, Morton R, Behar D, Delaney MA: Medical findings in child psychiatric patients. J Am Acad Child Adolesc Psychiatry *26:*554, 1987.

*Strauss G, Chassin M, Lock J: Can experts agree when to hospitalize adolescents? J Am Acad Child Adolesc Psychiatry *34:*418, 1995.

Troutman B, Myers K, Borchardt C, Kowalski R, Bubrick J: Case study: When restraints are the least restrictive alternative for managing aggression. J Am Acad Child Adolesc Psychiatry *37:*554, 1998.

Vivona JM, Ecker B, Halgin RP, Cates D, Garrison WT, Friedman M: Self- and other-directed aggression in child and adolescent psychiatric inpatients. J Am Acad Child Adolesc Psychiatry *34:*434, 1995.

48.9 COMMUNITY-BASED TREATMENT

ANDRES J. PUMARIEGA, M.D.

The field of child and adolescent community-based treatment has received increasing attention and significance over the last decade during which time it has gone beyond the traditional community mental health consultation model that was prevalent from the 1950s to the mid-1980s. Community-based treatment now involves the study and implementation of approaches to improve access, utilization, financing, and clinical and cost-effectiveness of mental health services provided to children and adolescents within the context of their home communities, as well as the functioning and effectiveness of systems of care for child mental health.

RECENT TRENDS

A number of recent trends have led to the increasing importance of this field. First, the increasing utilization and demonstrated need for child mental health services in the United States has spurred the rapid growth of such services. Although child and adolescent mental illness and emotional disturbance were once thought to be relatively rare, recent studies suggest overall prevalence rates of 15 to 19 percent, with 3 to 8 percent for serious mental illness and emotional disturbance. A number of morbidities associated with emotional disturbance and mental illness are increasing at alarming rates among children and youth; these include suicide, homicide, substance abuse, child abuse, teenage pregnancy, school drop-out, youth crime, and associated institutionalization and incarceration. A number of health and human service agencies (schools, social welfare agencies, child protective agencies, juvenile justice agencies, public health agencies) have experienced the increasing impact of psychosocial morbidity in children and youth. These agencies typically address pieces of the service system puzzle, with little or no coordination with other agencies that may serve the same youth. However, less than 1 percent of children in the United States receive mental health treatment in hospital or residential settings and another 5 percent in outpatient or community-based settings; the majority of children in need receive either insufficient mental health services or none whatsoever. In contrast, the budgets dedicated to financing public and private mental health services for children have grown exponentially since the 1970s, with total expenditures growing to 4.8 billion by 1990, which is approximately 7 percent of the total national mental health budget. The majority of these expenditures have been dedicated to inpatient and residential treatment, and treatment in such restrictive care settings is viewed with increasing skepticism because of its high cost as well as the limited documentation on its results.

At the same time as these needs are recognized, the resources available to fund child mental health and human services are increasingly limited. Medicaid, the public insurance program for the poor and disabled, funds a significant proportion of child mental health services in the United States through the coverage of Aid to Families with Dependent Children (AFDC) recipients and children with disabilities. Approximately 14 percent of children and adolescents 18 and under are enrolled in the Medicaid program, at least 450,000 of whom have severe emotional disturbances. The great majority of these are poor, underserved children of ethnic minority backgrounds. Children from these populations experience higher levels of stressors, such as poverty, discrimination, immigration, acculturation stress, and exposure to violence and trauma. However, minority populations have been traditionally underserved by mental health, health, and human services, both in terms of overall access to services as well as the cultural competence of the services available to them. The cost of serving these populations of children and adolescents is in contrast to the high cost of the psychosocial morbidity they contend with, including lost productivity and the costs of welfare dependency and institutionalization.

As a result of these trends public child mental health and social service agencies are increasingly pressured to demonstrate improved clinical and cost-effectiveness, and they increasingly turn to managed-care approaches to finance and organize mental health and social services. However, managed-care approaches are relatively new in the field of mental health, and those that have existed were developed with adult and private sector populations in mind. These approaches have relied on *a priori* benefit restrictions based on actuarial formulas that exclude poorer populations, which are high utilizers of services, while including large pools of relatively healthy, minimally impaired populations from higher socioeconomic backgrounds, who are lower utilizers of services. If applied to child mental health services, the usual benefits restriction approach in traditional behavioral managed care has the potential of depriving poor, underserved, and more impaired children of timely and effective intervention and preventive services. Thus, care is fragmented and the burden of services and costs shifts to the other child-serving agencies and systems and potentially increases morbidity significantly. Managed-care models for children's mental health need to emphasize community-based treatment and prevention of such morbidity by enhancing family and community resources and reducing psychosocial stressors for the child as they provide clinical services.

HISTORY

From their earliest origins, mental health services for children have emphasized a community orientation. These services began in the United States in response to the perceived need to counsel juvenile offenders rather than incarcerating them with adult offenders. This was the case in the 1890s when America, much like today, was undergoing rapid cultural changes because of immigration, rapid industrialization, and urbanization. These social strains led to marked increases in juvenile crime. Enlightened reformers saw the need to detain young offenders separately from adult prisoners, adjudicating them in a separate court system—this was the beginning of the juvenile court system—and to provide rehabilitative services to them. The juvenile court clinics in Chicago and Boston gave rise to the first child mental health services in the nation. The success of these clinics led the Commonwealth Foundation to commission a study that recommended the development of child guidance clinics throughout the United States staffed with interdisciplinary teams of professionals to serve the child and the family. At first these clinics were primarily staffed by social workers, but later attracted pediatricians, psychologists, psychoanalysts, and psychiatrists, and subsequently served as the bases of the first child psychiatry programs in the nation. The clinics were quite removed from the specialty-driven medical system that was evolving in tertiary medical centers, and particularly distinct from the practice of hospital-bound care. They provided low-cost services oriented to the needs of the child and the family, with treatment modalities evolving to include individual psychodynamic psychotherapy, family therapy, crisis intervention, and even day treatment programs. Many clinics that have survived to this day have served as the core for the child mental health services in many community mental health centers.

The move towards the medicalization of psychiatry served to move child and adolescent mental health services towards a more hospital-based, tertiary-care model. This left the child guidance clinics and the community mental health centers that followed them without significant psychiatric input as well as generally understaffed and underfunded. In the meantime the United States experienced a rapid increase in the population of poor and minority children, and a greater need for mental health services. Many of these children were placed in the custody of child welfare agencies or placed in residential and detention facilities.

The modern era of community-based systems of care for children was ushered in by the publication of Jane Knitzer's groundbreaking book *Unclaimed Children*, which exposed the aforementioned consequences of neglecting the provision of community-based mental health services for children and their families. Her advocacy and that of others led to the development of the Child and Adolescent Service System Program (CASSP), which assisted all 50 states in the development of an infrastructure for publicly funded community-based services. The CASSP initiative was supported by the conceptual work of Beth Stroul and Robert Friedman, who coined the term *community-based system of care for seriously emotionally disturbed children* and advocated interagency coordination among all the child service agencies in providing mental health services for children and related family support services. They proposed that such services be delivered as close to the child's home and community as possible to avoid more restrictive levels of care that often separated children from natural family and community supports, and that families be involved as partners in the effective treatment of the child. Stroul and Friedman's work spurred the development of various innovative community-based treatment modalities as well as several model demonstration programs that implement these modalities within the context of an organized interagency system of care.

THEORETICAL ISSUES

The CASSP initiative set forth the initial principles inherent in community-based systems of care. The key aspects of these systems include: access to a

comprehensive array of services, treatment individualized to the child's needs, treatment in the least restrictive environment possible (with full utilization of the resources of the family and the community), full participation of families as partners in services planning and delivery, interagency coordination, the use of case management for services coordination, no ejection or rejection from services because of lack of “treatability” or “cooperation” with interventions, early identification and intervention, smooth transition of youth into the adult service system, effective advocacy efforts, and nondiscriminatory, culturally sensitive services.

The principle of cultural competence is an integral aspect of the philosophy of community-based systems of care. It requires that systems of care include practitioner guidelines for a necessary attitudes, skill, and knowledge base to serve minority children and families in their communities, as well as policies and procedures to remove barriers for access to services. Community-based systems of care approaches were found to be consonant with the cultural values of ethnic minority populations, which emphasize strong extended-family involvement in the raising of children and the use of natural community resources first in dealing with the emotional and physical problems of family members. These factors have been shown to be protective against some of the morbidities associated with emotional disturbance, such as substance abuse and suicidality.

Another important principle inherent in this approach is that of the targeting of services to “seriously emotionally disturbed children.” This designation includes the presence of an Axis I diagnosis as specified by the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), but equally emphasizes the child's inability to function in at least one life domain (school, home, socially with peers). This reflects the difficulty that many children and their families have had in accessing necessary services when their diagnosis was not “severe enough,” particularly when the child was diagnosed with a disruptive behavioral disorder. Some studies indicate lack of clarity and validity in clinical child diagnosis, especially in relation to the prognosis of the child, with many children with disruptive behavioral disorders having comorbid serious mental illness when evaluated systematically. These problems are more pronounced in the diagnostic assessment of ethnic minority youth with serious emotional disturbance, with serious underestimation of comorbid mood and substance-related disorders and overestimation of psychotic and cognitive disorders by clinicians. Other studies indicate that the level of care received by children is only partially accounted for by their clinical diagnosis, with much of it accounted for by their level of function and psychosocial stressors.

The movement to managed behavioral health necessitated the application of principles of community-based systems of care to these approaches. These principles actually embody a true managed-care philosophy whereby care management becomes the primary modality for services utilization management rather than arbitrary, *a priori* benefits management. Combined with the principles of least restrictive levels of care, this approach results in higher-quality services and less disruption of children's lives and development, all at a lower cost than when more restrictive, higher levels of care are utilized. Pumariega and colleagues outline guidelines for states developing managed behavioral services for children funded by Medicaid to integrate principles of community-based systems of care into their contracts with managed-care providers, or into their policies and procedures for public managed services.

THERAPEUTIC PROCESS

A community-based treatment approach to services begins with access to services facilitated through multiple portals of entry, including direct referral or referral through child service agencies (such as schools and juvenile justice, child welfare, or public health facilities). A triage system should direct children and families to services and providers appropriate to their special needs. A full interdisciplinary team, including a child and adolescent psychiatrist, should be involved in the development of triage procedures and protocols. Comprehensive mental health assessments should be made by licensed providers who are credentialed and trained to evaluate children and families, in consultation with a child and adolescent psychiatrist. Such assessments should include diagnostic, functional, developmental, family, and cultural or community evaluations of the needs and strengths of the child and family. Standardized measures should be used whenever possible that are reliable, valid, clinically useful, culturally targeted to the population being served, and useful for programmatic evaluation.

Children and families who require time-limited, less intensive services should have an initial care plan developed by a licensed mental health professional in consultation with a child and adolescent psychiatrist. The initial care plan on children who require more extensive treatment should be developed by an interdisciplinary team, including all relevant child mental health and child service professionals. The parents, other relevant family members, and the child, if appropriate, should be members of this team. The care plan should identify problem areas and strengths that are relevant to the functioning of the child and family, as well as community-based interventions to address these needs. Patients and their families should be equal participants with the clinical team in the development and implementation of the plan of care. Plans of care should be individualized to the needs of the child and family, including attention to cultural issues, and should follow the child continuously through all levels of care, being amended as needed. The plan of care should promote continuous coordination and integration of all applicable services and agencies that are responsible for providing services to the child and family (such as primary health care, juvenile justice, schools, and social services).

Treatment and other services should be specific to the unique needs of the child and family and should use approaches that maintain or reintegrate the child into the family and community. Child and adolescent psychiatrists, psychologists, social workers, psychiatric nurses, and other members of the team should participate in the development of service protocols to ensure that the special developmental, physical, mental, and emotional needs of the child are addressed. The system should use as many diverse therapeutic modalities as possible, particularly those that have demonstrated effectiveness with seriously emotionally disturbed children. These include traditional and intensive outpatient services, nonresidential community-based interventions (including family preservation, wrap-around services, day treatment, and school-based services), community-based residential services (such as therapeutic foster homes and group homes), and brief, acute inpatient services. These interventions are aided by the increasing effectiveness of cognitive, behavioral, family-based, and psychopharmacological therapies, which should be integrated into all community-based interventions and levels of care.

Case management services are essential in community-based treatment, including assessment of problem areas, measurement of functioning level, determination of needs, linkage with necessary resources, care coordination, advocacy, and monitoring of services provided and their outcomes. The child and family should be assigned a case manager who has the necessary skills to ensure seamless coordination of care across all levels, different providers, and the full time course of illness or disturbance. The frequency and intensity of case management services is proportional to the clinical and psychosocial needs of the child and family. Patients and their families should be participants in the case management process, sharing responsibility with case managers for using and coordinating services. Case management protocols should facilitate the coordination of multiagency, multisystem interventions, integrating services from the various providers and agencies responsible for serving the child, and ensure coordination of services with primary care physicians and health providers.

ROLES AND FUNCTIONS OF PSYCHIATRISTS

Given the complexity of these new systems of care, the skills that psychiatrists must bring and the roles that they must play go far beyond the circumscribed professional roles of their disciplines. These roles include front-line clinician, clinical consultant to other professionals, clinical team leader, administrative leader in delivery organization of system, quality assurance and improvement consultant, consultant to interagency teams, and outcome evaluator or researcher in systems of care. Given their broad biopsychosocial perspective, child and adolescent psychiatrists are often the professionals who have the broad expertise to integrate and coordinate community-based treatment delivered by multiple professionals with diverse skills. However, given the emphasis on tertiary-care models of care in the training of many child and adolescent mental health professionals, they often lack the skills necessary to function as effective members and leaders of these new systems. A publication from the American Academy of Child and Adolescent Psychiatry, *Guidelines for Training Child and Adolescent Psychiatrists in Community-Based Systems of Care*, provides guidance on the array of skills, knowledge base, attitudes, and training experiences that psychiatrists will need to function effectively within community-based systems of care for children.

CLINICAL ISSUES

Indications Community-based treatment is indicated for the treatment of the majority of children with moderate to serious emotional disturbance or mental illness. It is most indicated for the prevention of out-of-home placement of such children, and for the treatment of multiproblem, multiagency children with limited family resources. It has some limitations in dealing with situations involving extreme danger (to self and others), serious abuse, or neglect, which necessitates removal of children from their environment. Community-based treatment can be contraindicated when the child's family is too fragmented or dysfunctional, as well as when there is imminent danger (suicide, homicide), dangerous symptomatology (e.g., acute psychotic symptoms), or decrease in function that prevents the child from being cared for in the home or community. However, community-based treatment can be resumed when such symptoms or conditions are resolved (usually after brief stabilization), or the family is able to function again such that it can safely participate in the care of the child.

Goals of Treatment The goals of community-based treatment usually include the following: (1) to restore the child's functioning in school, at home, and in the community; (2) to reintegrate the child within the family and community and avert further separation or disruption of community tenure; (3) to individualize interventions to the need of the child and family; and (4) to maximize the strengths and resources of the family and the surrounding community on behalf of the child.

Libby and Shawna were African-American girls aged 13 and 14 when they came to the attention of the interagency community-based team in their rural community. They had been removed from their home at ages 6 and 7 because of their mother's substance abuse, mental illness (bipolar disorder), and resulting neglect, as well as physical abuse by their stepfather. They lived in a succession of foster homes in the region for 2 years, until they were placed in the home of the Browns, an older retired couple. The girls seemed to do well in this setting until Mrs. Brown died from multiple health complications. This left Mr. Brown, a highly religious and rigid man, caring alone for the girls. The girls and Mr. Brown engaged in pitched power-control struggles as they became older, with the girls running away and often ending up in their mother's home. The girls also had difficulties at school with disruptive behavior, especially Shawna, who was quite irritable and agitated and suspected of using drugs; Libby also had significant academic difficulties. Mr. Brown felt that he could no longer handle the girls in his home, and the girls both expressed the desire to be reunited with their mother.

The case of the sisters was referred to the interagency team for consultation and development of a comprehensive care plan. The team first recommended evaluation of the mother's home environment because she had faithfully attended treatment in the mental health center for her substance abuse and bipolar disorder over the past 4 years, and had maintained sobriety for 3 of those years. The family preservation therapist found a stable relationship between the mother and her new husband; the mother was eager to provide for the girls' needs. The consulting psychiatrist was asked to evaluate Shawna, and his impression was that she suffered from major depression, which was successfully treated with a serotonin agent and brief psychotherapy. A therapist in the mental health center and a diagnostician at the school concluded that Libby had a significant learning disorder that required remedial tutoring in a resource class. The team then agreed to stage a gradual plan of reunification, starting with weekend visits with the mother, and later having Libby, then Shawna, rejoin their mother and her husband over a 2-month period. The family preservation therapist, with consultation from the child psychiatrist and assistance from the child welfare case worker, worked with the mother and her husband on behavioral and parenting techniques to address the girls' occasional oppositional and impulsive behaviors, as well as to access support resources from their local African-American congregation. After 6 months of follow-up, the girls were successfully reunited with their mother, with no recurrence of mood or disruptive symptoms or runaway behavior.

ETHICAL ISSUES

Ethical issues can arise in community-based treatment when extremes are pursued, such as when a child is placed in an environment that is too restrictive or insufficiently restrictive. A home with a high level of conflict or abuse can be more restrictive for a child than many residential programs because of the degree to which activities are constrained for the child and family. On the other hand, the child may require a more structured or restrictive environment than the family can provide, such as when the child is exhibiting suicidal, violent, or severe runaway behaviors. Pursuing community-based treatment in such situations can create an ethical dilemma that needs to be addressed by the mental health professionals involved.

The utilization of community-based treatments exclusively to reduce costs without taking the needs of the child and family into account is another serious breach of ethics. Unfortunately, the financial incentives to provide less care inherent in most managed-care approaches often lead to such situations, which can potentially endanger the child. Community-based approaches need to be pursued within organizational and financial structures that do not have arbitrary benefit limits and that rely on on-site case management for the management of care.

RESEARCH EVALUATION

The field of child mental health services research has particularly focused on the effectiveness of community-based treatment. Research in this area has included the evaluation of community-based residential interventions, intensive nonresidential community-based interventions, and the clinical, functional, and cost outcomes of community-based systems of care, as well as factors affecting the need and utilization of mental health services in different populations of children.

Community-Based Interventions A number of innovative techniques that go beyond traditional outpatient services and provide more intensive and individualized services have been developed recently and are undergoing initial evaluation studies. These include different types of day treatment programs, school-based interventions, wilderness programs, crisis mobile outreach teams, therapeutic in-home emergency services, time-limited hospitalization with coordinated community services, intensive case management, and family support services. Studies involving these modalities have demonstrated significantly better outcomes than traditional outpatient or residential services. These include lower levels of externalizing and internalizing symptoms, improved family functioning, and reduced utilization of more restrictive services.

The wrap-around model of care embodies this approach towards multimodal community-based interventions. This model emphasizes aggressive outreach, use of least restrictive treatment options, and individualized, flexible, and unconditional child- and family-centered services for children imminently at risk or already in out-of-home placements. A number of studies of wrap-around programs in different communities across the nation with diverse populations of at-risk children and families have reported positive outcomes in terms of reduction of externalizing behavioral problems, level of function, reduction in out-of-home placement, improved family management skills and function, and consumer/family satisfaction. *Multisystemic therapy*, a therapeutic approach that combines home-based and wrap-around interventions within a systemic context, has been tested with youth at risk of detention and incarceration, with significant results in reducing out-of-home placement, externalizing criminal problem behaviors, rates of arrest and incarceration, and costs.

Community Systems Evaluations Initial outcome data on overall service system function in Vermont, New York, and Ventura County, California have demonstrated reduced hospitalization rates and days, decreased out-of-home placement and less restrictive placements, significantly lower incidence of negative behaviors vulnerable to out-of-home placement, and significantly lower rates of overall problem behaviors. The emerging model of community-based systems of care is being tested in a number of pilot demonstration sites, including 30 pilot community programs funded by the Center for Mental Health Services in different communities with a strong child mental health services' infrastructure and eight community-based systems pilot sites funded by the Robert Woods Johnson Foundation. The Fort Bragg Demonstration Project has been the largest systems of care demonstration project attempted so far. The project developed a comprehensive continuum of care for the children of military dependents at Fort Bragg, North Carolina, with ready access to outpatient and intermediate levels of care and relatively restricted utilization of residential and inpatient services. The outcomes from this system were compared to the usual CHAMPUS-funded services accessed by military dependents at various military installations nationally. Multiple process, quality, and outcome indicators (including clinical indicators, level of function, and cost of care) were evaluated and compared across the experimental site and the usual service sites. The initial results suggest that the program did achieve significantly lower restrictiveness of care, higher consumer satisfaction, greatly increased access to services, and greater funding spent on less restrictive services. However, clinical and functional outcomes were not significantly different overall, with split differences across diagnostic subgroups while overall outcomes were no different. Additionally, costs were not significantly different and possibly even somewhat higher in the demonstration site.

Services for Minority Youth Research on services for underserved minority youth has been sparse, and had initially focused on documenting racial differences in services provided. African-American children tend to remain in foster care longer and to have more foster-care placements than European-American children. Studies have shown that culturally diverse children are underrepresented in mental health institutions and overrepresented in child welfare and juvenile justice settings and placements as compared to nonminority youth, even when they are equally psychiatrically impaired. Ethnic and racial differences in the diagnoses of culturally diverse adolescents have been identified by some investigators, including overdiagnosis of conduct disorder and psychotic disorders and underdiagnosis of mood personality, and substance-related disorders. Significantly lower rates for overall utilization of services and specifically for the treatment of depression in culturally diverse groups have been documented; some studies have found a significant relationship between race and the service needs of racial ethnic minorities.

FUTURE RESEARCH NEEDS

The National Evaluation of the Center for Mental Health Services children's systems of care pilot demonstration sites should provide information on the effectiveness and cost efficiency of community-based systems of care and the effectiveness of our evaluation and research methods. In spite of current research activity, there are still significant deficits in our knowledge of children's mental health services and systems of care. Further research is needed on consumer satisfaction with services, the effectiveness of culturally competent services, the school-mental health interface, the interface with juvenile justice and child welfare, the impact of different cost-containment and reimbursement mechanisms, the effectiveness of interagency collaboration and coordination, and the integration of child mental health and primary care services. The application of mental health services' research findings in service system policy and planning is the ultimate objective and value of such research.

SUGGESTED CROSS-REFERENCES

Section 30.2, Section 30.5, and Section 30.6 contain important background information on behavior, family, and cognitive therapies. Chapter 40 and Section 49.7 contain important information on disruptive behavior disorders and child or adolescent antisocial behavior, including information on comorbidity with other psychiatric

disorders and the relevance of community-based interventions in their treatment. Section 48.7 and Section 48.8 have further details on partial hospitalization and residential and inpatient treatment as they relate to community-based treatment and systems of care. Section 48.9 has relevant information on community psychiatry.

SECTION REFERENCES

Armbruster P, Gerstein S, Fallon T: Bridging the gap between service need and service utilization: A school-based mental health program. Community Ment Health J *33:*199, 1997.

*Atkins DL, Pumariega AJ, Montgomery L, Rogers K, Nybro C, Jeffers G, Sease F: Psychopathology in incarcerated youth: An epidemiological study. J Child Fam Studies *8;*22, 1999.

Berlin I: Development of the subspecialty of child and adolescent psychiatry. In *Textbook of Child and Adolescent Psychiatry*, J Wiener, editor. American Psychiatric Press, Washington, DC, 1991.

Bickman L, Heflinger C, Lambert E, Summerfeldt W: The Fort Bragg Managed Care Experiment: Short-term impact on psychopathology. J Child Family Stud *5:*137, 1996.

Brindis C, Sanghani R: School-based health clinics remaining viable in a changing health care delivery system. Annu Rev Public Health *18:*567, 1997.

Bruns E, Burchard J, Yoe J: Evaluating the Vermont system of care: Outcomes associated with community-based wraparound services. J Child Family Stud *4:*321, 1995.

Burns B, Farmer E, Angold A, Costello E, Behar L: A randomized trial of case management for youths with serious emotional disturbance. J Child Clin Psychol, 1997.

Burns B, Taube C, Taube J: Mental health services for adolescents. In *Adolescent Health volume II: Background and Effectiveness of Selected Prevention and Treatment Services* (OTA-H-466). U.S. Government Printing Office, Washington, DC, 1991.

Cohen R, Parmalee D, Irwin L, Weisz J, Howard P, Purcell P, Best A: Characteristics of children and adolescents in a psychiatric hospital and a correctional facility. J Am Acad Child Adolesc Psychiatry *29:*909, 1990.

Cuffe S, Waller J, Cuccaro M, Pumariega A, Garrison C: Race and gender differences in the treatment of psychiatric disorders in young adolescents. J Am Acad Child Adolesc Psychiatry *34:*1536, 1995.

Eber L, Osuch R, Redditt C: School-based applications of the wraparound process: Early results on service provision and student outcomes. J Child Family Stud *5:*83, 1996.

England MJ, Cole R: Building systems of care for youth with serious mental illness. Hosp Community Psychiatry *43:*630, 1992.

Evans M, Armstrong M, Kuppinger A: Family-centered intensive case management: A step towards understanding individualized care. J Child Family Stud *5:*55, 1996.

Glover S, Pumariega A: The importance of children's mental health epidemiological research with culturally diverse populations. In *Promoting Cultural Competence in Children's Mental Health Services*, Hernandez M, Isaacs M, editors. Brookes, Baltimore, MD, 1998.

Henggeler S, Melton G, Smith L, Schoenwald S, Hanley J: Family preservation using multisystemic therapy: Long-term follow-up to a clinical trial with serious juvenile offenders. J Child Family Stud *2:*283, 1993.

Hyde K, Burchard J, Woodworth K: Wrapping services in an urban setting. J Child Family Stud *5:*67, 1996.

Ichnose C, Kingdon D, Hernandez M: Developing community alternatives to group home placement for seriously emotionally disturbed special education students in the Ventura county system of care. J Child Family Stud *3:*193, 1994.

*Jensen P, Hoagwood K, Petti T: Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model. J Acad Child Adolesc Psychiatry *35:*1055, 1996.

Jordan D, Hernandez M: The Ventura Planning Model: A proposal for mental health reform. J Ment Health Adm *17:*26, 1990.

Kilgus M, Pumariega A, Cuffe S: Influence of race on diagnosis in adolescent psychiatric inpatients. J Am Acad Child Adolesc Psychiatry *35:*167, 1995.

Kiser L, Culhane D, Hadley T: The current practice of child and adolescent partial hospitalization: Results of a national survey. J Am Acad Child Adolesc Psychiatry *34:*1336, 1995.

*Knitzer J: *Unclaimed Children: The Failure of Public Responsibility to Children and Adolescents in Need of Mental Health Services*. Children's Defense Fund, Washington, DC, 1982.

Kupermine G, Cohen R: Building a research base for community services for children and families: What we know and what we need to learn. J Child Family Stud *4:*147, 1995.

Lourie I, Katz-Leavy J: Severely emotionally disturbed children and adolescents. In *The Chronically Mentally III*, W Menninger, editor. American Psychiatric Press, Washington, DC, 1986.

Pumariega A, Atkins L, Montgomery L, Rogers K, Sease F, Jeffers G: Psychopathology in incarcerated youth. In *1996 Annual Meeting: New Research Programs and Abstracts*, Abstract 229. American Psychiatric Association, Washington, DC, 1996.

*Pumariega A, Cross T: Cultural competence in child psychiatry. In *Basic Handbook of Child and Adolescent Psychiatry*, vol 4. J Noshpitz, N Alessi, editors. Wiley, New York, 1997.

*Pumariega A, Glover S: New developments in services delivery research for children, adolescents, and their families. In *Advances in Clinical Child Psychology*, vol 20. T Ollendick, R Prinz, editors. Plenum, New York, 1998.

Pumariega A, Glover S, Holzer C, Nguyen H: Utilization of mental health services in a tri-ethnic sample of adolescents. Community Ment Health J *34:*145, 1998.

Pumariega A, Nace D, England M, Diamond J, Mattson A, Fallon T, Hansen G, Lourie I, Marx L, Thurber D, Winters N, Graham M, Weigand D: Community-based systems approach to children's managed mental health services. J Child Family Stud *6:*149, 1997.

Schoenwald S, Ward D, Henggeler S, Pickerel S, Patel H: Multisystemic therapy treatment of substance abusing or dependent adolescent offenders: Costs of reducing incarceration, inpatient, and residential placement. J Child Family Stud *5:*431, 1996.

Shyne A, Schroeder A: *National Study of Social Services in Children and their Families*. Westat, Rockville, MD, 1978.

*Stroul B, Friedman R: *A System of Care for Severely Emotionally Disturbed Children and Youth*. Georgetown University Child Development Center, CASSP Technical Assistance Center, Washington, DC, 1986.

Tuma J: Mental health services for children: The state of the art. Am Psychol *44:*188, 1989.

Young S, Nicholson J, Davis M: An overview of issues in research on consumer satisfaction with child and adolescent mental health services. J Child Family Stud *4:*219, 1995.

48.10 PSYCHIATRIC TREATMENT OF ADOLESCENTS

CYNTHIA R. PFEFFER, M.D.

The onset of adolescence marks a major transition in life and encompasses developmental changes in cognitive, affective, neurobiological and social features of an individual's profile. Since the 1980s, significant advances in characterizing these features have been associated with the development of new techniques to define and measure them. Consequently, advances in techniques for characterizing normal and pathological features of adolescence have led to innovative strategies for psychotherapeutic and psychopharmacological treatment of adolescents.

Treatment schemas have become more specific to coincide with normal developmental parameters and indicators of developmental psychopathology. Framing treatment needs of adolescents requires understanding the developmental concepts because a particular type of treatment may not be appropriate or applicable for individuals across the life span. Thus, application of treatment for adolescents should be guided, in large measure, by empirical research that identifies the complexities of developmental psychopathological deviation and the efficacy of specified psychotherapeutic and psychopharmacological treatments. Thus the treatment of adolescents presupposes knowledge about the characteristics and timing of onset of psychopathology, its course, rates of recovery, risks for relapse or recurrence, and comorbid states.

DEVELOPMENTAL CHARACTERISTICS OF ADOLESCENTS

Adolescence is a period of life that readies an individual to leave the family of origin and prepare for future adult roles encompassing independent support of self, family, and others. Developmental accomplishments of adolescence involve maturation of cognition, profound physical changes in hormonal balances that eventuate in growth in stature and sexual capacities, and more-complex interpersonal interactions with attendant increases in social roles and identity consolidation. Numerous sociocultural rituals are used to designate the onset of adolescence.

Although the onset of adolescence is marked by enormous states of transition in biological and psychosocial dimensions, the termination of adolescence is less apparent and is associated with significant psychological and social adjustments. In this regard, psychotherapy applied to adolescents who suffer from psychopathological states presents an important opportunity to facilitate intense reorganization in psychological and social functioning that highlights adolescence. Because of the many changes during adolescence, psychotherapy with adolescents must target multiple domains of psychosocial functioning.

According to theoretical concepts of cognitive development, adolescents achieve levels of abstract reasoning that enables them to conceptualize events in terms of causality. These changes suggest that intervention can emphasize cognitive skills designed for independent problem solving.

Although pubertal changes begin at approximately age 9 to 10 years, pubescence is the hallmark of adolescence. Emotional, hormonal, and behavioral changes evident at pubescence are characterized by intense urges for independence and a simultaneous heightened dependency. Rates of pubertal changes are individually determined and differ for boys and girls.

Physical, emotional, and cognitive maturation in adolescence involving advanced levels of spatial, linguistic, and executive functions that control and regulate the integration of information and resultant behavioral responses is instrumental in the progression of the more complex interpersonal relationships formulated in adolescence. Conflicts with parents are juxtaposed with intense needs for guidance from parents and peers. Intensification of peer relationships points the way for future independent living and significant conceptualization of identity. Concerns about autonomy and decreased dependency on parents are major issues in the middle and late phases of adolescence, which are characterized by the slowing of major biological changes, heightened peer relations, increased capacity for abstract and formal operational thinking, identity resolution and consolidation, and preoccupations with future plans. Although emotional and behavioral turmoil is considered a component of adolescence, significant turmoil is not usual and should be considered an indication of psychopathology.

The choice of psychiatric treatment for adolescents is largely determined by the cognitive, emotional, physical, and interpersonal characteristics of individual adolescent patients. Treatment plans must also consider the nature of adolescent psychopathology.

PSYCHOPATHOLOGICAL CHARACTERISTICS OF ADOLESCENTS

Since the 1970s, rates of psychopathology have increased for adolescents. Suicide is the third leading cause of death in those 15 to 24 years of age, and the rate of suicide for children and young adolescents has more than doubled. Adolescents have significantly higher rates of depression, substance abuse, eating disorders, and posttraumatic stress disorders than those born before World War II. Heightened aggression, homicide, and psychotic states have been documented. Comorbidity of these psychopathological states is highly prevalent and challenges any treatment paradigm. Diagnostic formulation of adolescents involves an emphasis on their social context. Development of treatment strategies involves consideration of current psychopathology, features of continuity and discontinuity in psychiatric symptoms, prior psychiatric disorders, physical illness, social stresses, and general adaptive functioning. These issues are classified in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Great advances in understanding the etiology of adolescent psychopathology have been achieved since the late 1970s. For example, etiological parameters are related to acquired and constitutional neurobiological lesions and deficits, psychiatric disorders in the family, family discord, institutional rearing, and unique genetic factors. Given current knowledge about the epidemiology and clinical characteristics of adolescent psychopathology and psychosocial functioning, the most effective treatments for the prevalent psychopathologies of adolescents must be identified.

HISTORY OF PSYCHOTHERAPEUTIC TECHNIQUES FOR ADOLESCENTS

Distinct treatment strategies for adolescents were used before and after the 1970s. Before the 1970s, psychodynamic psychotherapy modeled after the principles of psychoanalysis was the main treatment for adolescents. Among the first instances of psychoanalytic treatment of adolescents was Sigmund Freud's treatment of his patient Dora, who entered psychoanalysis at the age of 16 years. Many pioneers in treating adolescents, such as August Aichorn, Anna Freud, Erik Erikson, and Peter Blos, described interventions based on psychoanalytic-developmental issues. The work of these pioneers centered on characterizing treatments for deviant youth such as delinquents (Aichorn), describing developmental lines for affective and behavioral achievements (Freud), schematizing critical life phases involving psychological achievements and deficits (Erikson), and defining a theory of adolescent development (Blos). These contributions continue to guide clinicians and researchers in developing treatments for adolescents.

Since the 1970s, combinations of psychotherapy and psychopharmacology have been used to treat adolescent psychopathological symptoms and disorders. Pressure to determine the outcomes and cost-effectiveness of psychiatric treatments for adolescents have made certain treatment approaches more common. Currently, health care reform demands methods of cost containment and documentation of the benefit gained by patients who undergo specific treatments. Pretreatment and posttreatment measurement of parameters expected to change as a result of undergoing treatment is advocated. The time needed to achieve clinically significant change in these parameters is considered in treatment planning. Partition of administration of specific types of treatment has become commonplace, so adolescent patients may be treated by multiple providers. This approach makes it necessary to clarify who has oversight for treatment planning.

TYPES OF PSYCHOTHERAPEUTIC TREATMENTS

Psychotherapeutic treatments can be divided into those that focus on the individual and those that involve multiple individuals. Within these classifications, specific types of treatments are offered.

Individual Psychotherapies A therapy that focuses on the individual adolescent intensively evaluates and aims to change patterns of thinking, affect expression, and behavior. Different types of individual psychotherapy emphasize different features of an adolescent's psychosocial functioning and are based on different theoretical paradigms and intervention techniques. Application of such therapies is assumed to have long-term effects that can be generalized to various circumstances encountered by an adolescent. Therefore, multiple interventions administered at varied points in time may be necessary, especially in treatment of adolescents with chronic psychopathology. In general, these therapies do not focus on effecting specific changes in the adolescent's environment. It is assumed that the therapeutic response will foster adaptive means of coping with environmental situations.

Psychoanalytic Treatment and Psychodynamic Psychotherapy Historically, the most widely used individual treatment involves psychodynamic psychotherapy whose theoretical basis is the resolution of unconscious conflict through expression of ideas and emotions and the development of specific insights about the causes of such conflict. The treatment aims to have the adolescent gain insight about problems and acquire skills for more effective coping. This type of treatment is usually not time limited and it requires highly skilled therapists with an active, verbal manner that enables the adolescent to form a real relationship with the therapist. This type of treatment assumes that the adolescent patient has a capacity for self-reflection and can establish a relationship, and it is applied to various types of psychopathology. A spectrum of such therapies exists ranging from psychoanalytic treatment and insight-oriented psychotherapy to supportive psychotherapy.

Psychoanalytic treatment requires multiple sessions each week and uses free association to elicit the material used to identify etiological conflicts and maladaptive coping styles. Various interpretations are offered by the psychoanalyst about the conflicts and their sequelae. Within the psychoanalytic process the patient forms a relationship to the therapist that reflects the unconscious conflicts and their emotional and behavioral sequelae. This transference is used by the analyst to interpret the patient's conflicts.

Psychodynamic psychotherapy uses psychoanalytic principles and is a variant of the psychoanalytic process. It also aims to identify and reduce unconscious conflicts, but it does not require such frequent therapeutic sessions and the therapist offers supportive comments about how the patient can respond to situations and how responses reflect unconscious conflicts.

Psychoanalytic and psychodynamic treatments have not been subjected to extensive systematic research to evaluate the resolution of patient's conflicts and changes in cognitive, emotional, and behavioral profiles. Nevertheless, the principles of such treatments are based on developmental principles that can be readily adapted to the distinct features of unconscious conflict in the adolescent patient. Thus such treatment does not emphasize psychiatric disorder. It has been used to treat adolescents with varied psychiatric disorders.

Cognitive-Behavioral Treatment Recently, cognitive-behavioral treatment has gained popularity and has been a focus of systematic outcome research for a variety of adolescent psychopathologies. This treatment uses cognitive and behaviorally based interventions to effect change in emotional, cognitive, and behavioral profiles of adolescents. It focuses on both the external environment and the adolescent's internal processing of environmental situations. The therapist offers suggestions about coping with the environment and improving problem-solving techniques. This treatment aims to strengthen an adolescent's cognitive skills and behavioral control. A central issue for cognitive-behavioral treatment is how patients respond to their cognitive interpretation of experiences and to a lesser extent the actual environment or experience. It aims to change distorted information processing by changing cognition and methods of behavioral contingency management.

In this treatment cognitive skills are taught and practiced initially within an affectively neutral environment. Once sufficient proficiency has been achieved in neutral settings, cognitive skills are practiced in situations of affective arousal. Thus, cognitive-behavioral therapy aims to develop and strengthen adaptive coping templates. It is not aimed at decreasing unconscious conflicts. Controlled studies have shown cognitive-behavioral treatment to be effective in a time-limited number of sessions to treat adolescents with anxiety, mood, eating, and specific developmental disorders.

Anxious adolescents misperceive environmental demands, which increases situational stress. Cognitive-behavioral therapy aims to identify anxiety-provoking situations, characterize the ideas and emotions associated with such situations, and alter these perceptions and enhance adaptive stress-relieving responses by the use of modeling procedures. Other techniques used include role playing, relaxation, and contigent reinforcement. When skills are learned, they are practiced in a real situation and encouraged by social reinforcement.

> A 13-year-old boy with a recent onset of obsessions about locking the front door of his home and the compulsion to keep his room excessively clean was treated as an outpatient with cognitive behavior therapy. In addition, he complained of worries about getting dirty when he played soccer. Despite this adolescent's talent in athletics, he became very anxious when he attended soccer practice for his school team. Cognitive behavior therapy began with providing a detailed account of his obsessions and compulsions. Desensitization techniques were utilized to reduce these unwanted symptoms. Cognitive paradigms were utilized to reduce his obsessive thinking and to prevent acting out his compulsions. Within 2 months of therapy, he was able to play soccer without interference from the obsessions and compulsions.

Depressed adolescents have cognitive distortions involving global, stable, internal attributions of negative events that lead to low self-esteem, decreased motivation, helplessness, and hopelessness. Cognitive-behavioral treatment aims to restructure thinking and promote adaptive perceptions of self, others, and circumstances as well as strengthen problem-solving skills.

Perceptions of aggressive adolescents are so distorted that they do not use all available information to make decisions. Instead they use biased recall of hostile cues and so, in ambiguous situations they attribute hostile motives to others' behavior. This produces an aggressive, action-oriented behavioral pattern in such adolescents, and they have difficulty planning alternative solutions to problems and selecting appropriate responses. Cognitive-behavioral therapy aims to reorganize concepts and develop less impulsive behaviors.

Behavior Therapy Behavior therapy has a distinct role in effective management of adolescent psychopathology, although research is needed to determine its efficacy for specific disorders or psychiatric symptoms, its comparative efficacy and its use in combination with other interventions for adolescents. When applied to humans, behavior therapy combines approaches of respondent and operant conditioning and reinforcement methods with features that address the complex mental processes, especially the cognitive processes, of humans. Behavior therapy is based on the principles of learning theory. It uses a problem-oriented approach and targets specific symptoms or behavior rather than focusing on a psychiatric disorder.

An important feature of behavior therapy is systematic desensitization, in which an adolescent is exposed to the specific anxiety-provoking stimulus, either in fact or in imagination. This is done gradually and only after the adolescent is taught relaxation as a counterphobic technique. Behavior therapy also involves operant techniques, in which events immediately prior to a behavior (eliciting stimuli) or subsequent to a behavior (reinforcing or punishing contingencies) are systematically manipulated in an effort to change the target behavior. Shaping involves applying positive reinforcement to a behavior targeted for enhancement. Such techniques are used for skill acquisition.

Contingency management involves systematically controlling the consequences of a behavior targeted for change (e.g., by using contracts, token economy, or points). Frequently these techniques are used in conjunction with cognitive techniques and social skills training.

Behavior therapy has been used widely and effectively to treat symptoms prevalent among mentally and developmentally delayed adolescents. Behavior therapy involving contingency management has been widely and effectively used to treat symptoms of attention-deficit disorder. In the case of oppositional defiant and conduct disorders, behavior therapy has shown beneficial long-term effects when applied with behavioral contracting, conflict-resolution skills training, token economies used in the home, and training the parents to restructure the home environment. Methods that incorporate systematic desensitization, social learning, and cognitive methods have been shown effective in treating anxiety disorders, including simple phobias, school refusal, and obsessive-compulsive disorder. Operant conditioning techniques have been helpful in treating eating disorders such as anorexia nervosa and obesity. Behavioral therapy techniques for adolescent substance abuse involve decreasing immediate reinforcing properties of the substance, reinforcing abstinence, teaching alternative successful social behavior, and cognitive approaches that enhance self-help.

Interpersonal Psychotherapy In 1984 interpersonal psychotherapy was developed by Gerald Klerman and colleagues as a time-limited, brief, psychotherapeutic treatment for depressed adults, and subsequently it was applied and studied with depressed adolescents. Its main concept is that depression occurs in the context of

interpersonal relationships that involve discrete problem areas of grief, interpersonal role disputes, role transitions, and interpersonal deficits. One or two of these issues become the main focus of treatment. Treatment is based on the assumption that the onset and continuation of symptoms of depression are associated with these interpersonal situations. This treatment has been studied extensively in controlled clinical trials and shown to be effective with depressed adults alone and in combination with antidepressant medication. It has been adapted and studied in adults for long-term maintenance treatment of recurrent depression, opioid dependence, cocaine abuse, bulimia nervosa, bipolar disorders, and depression due to human immunodeficiency (HIV) disease.

An essential feature of this treatment is a focus on the present and future, rather than the past. Interpersonal treatment has been modified for treatment of nonpsychotic, nonsuicidal, depressed adolescents. In addition to the four issues of focus for treatment of adults, interpersonal treatment for adolescents also focuses on issues relevant to adolescents who live in single-parent families. Specifically, the treatment for adolescents addresses perceptions of authority conflicts with parents and other adults, reactions to separation or divorce of parents, development of interpersonal relationships with adolescents of the opposite sex, peer pressure, and reactions to the death of a relative or friend. An open treatment trial with a relatively small sample of depressed adolescents suggested significant reduction of symptoms of depression at termination of treatment, and 1 year of posttreatment, recovery was generally maintained. Controlled treatment trials are necessary to determine the efficacy of this treatment for depressed adolescents.

A 15-year-old girl complained of intensifying symptoms of sadness, crying, problems falling asleep, difficulty concentrating on academic tasks and thoughts of committing suicide since she broke up with her boyfriend 4 months previously. She was treated as a psychiatric outpatient with time-limited interpersonal psychotherapy that focused on her grief over losing her boyfriend. Treatment educated her about the nature of grief and depression, assisted her to consider her feelings of guilt in causing the loss of her boyfriend, and strengthened skills in adapting to this loss and to develop new satisfying relationships. Follow-up 3 months after treatment indicated that she had maintained improvement in academic work and social relationships with peers and that she reported no symptoms of depression.

Family Treatment The family is the actual and perceived essential interpersonal structure for children and adolescents. It provides the basis for satisfying the dependency needs of the young, protecting against adverse environmental situations and providing models for socialization that significantly determine characteristics for future independent living. Family psychotherapy amplifies these basic constellations and enhances the intrafamilial communication, behavior, and routines to ensure that adolescents' sense of identity, social values, and future goals are identified and discussed.

During early adolescence, a strong antagonistic equilibrium is apparent, as the adolescent begins to move out of the family but remains dependent on parental influences. At this time pubertal changes are most marked and rapid, which heightens the adolescent's wish for privacy and intolerance of parental scrutiny. Often conflicts arise around these developmental events. Family therapy offers a way to guide the parents in planning their responses to their adolescent child.

Midadolescence is significant for its intense peer relationships and identity formation and heightened independence. Family therapy may need to focus on the adolescent and parents to assist the development of these important features. Often parents feel excluded by the adolescent or they feel less empowered to control the behavior of the adolescent. Adolescent problematic behaviors are perceived as insurmountable by the parents, who are often rendered helpless and feel anxious about the adolescent's striving for independence. Family treatment must help parents learn to assess the circumstances of the adolescent and to set appropriate limits while letting adolescents feel that moving toward autonomy and choice is within their domain of responsibility.

In late adolescence important transitions to adult life become evident. Consolidation of identity promotes new choices about school, work, and interpersonal relations, with a view toward family planning. Major biological development has been completed and psychosocial transition is paramount. Family therapy facilitates these advances by helping parents loosen their parental oversight and guide the adolescent out of adolescence and into adult behavior, responsibility, and insights about future life expectations. The availability of parental support is essential, but at lower intensity and less frequently.

Family therapy techniques focus on the family's expression of emotions; the communication styles of individual family members; conflicts; cohesiveness (i.e., the degree to which family members perceive and behave as an emotionally adherent unit), sharing of values, goals, and behavior; the satisfaction of individual family members with family functioning; adaptability; and problem-solving skills. Family treatment also addresses the developmental phase of adolescence, the family developmental status, and the specific psychopathology of individual family members, while maintaining a focus on the interpersonal structure of family functioning. An overview of the family as a unit is essential in conducting family psychotherapy. Thus, the individual is considered within the context of the larger family unit. Specific needs are addressed with regard to the cohesive needs of the family unit, and the individual family members must contribute to satisfying family needs.

Group Treatment Adolescent development involves gaining psychosocial independence from the nuclear family and acquiring the skills to assume adult responsibility. During this process, peer relations let adolescents practice their skills and get feedback in the form of friendships. Developing competence in peer relationships heavily depends on the quality of parent-child relationships. Group treatment provides peer discussion to exchange perceptions, guide problem-solving, and promote a more solid sense of identify. Group intervention is often preferred by adolescents because of the opportunity to be with peers. The guidance offered within the context of the group can be extensive and specific and may be integrated by the adolescent. As in family treatment, focus is maintained on the group as a unit, and progress is estimated by the development of a group cohesion that facilitates active interchange of ideas among its members.

As with treatment of prepubertal children, group psychotherapy of young adolescents (12 to 14 years) often involves a combination of discussion and other activities such as drawing, reviewing pictures, and storytelling. In contrast, group psychotherapy for midadolescents (15 to 18 years) and late adolescents (19 to 21 years) primarily involves discussion, with some dependence on other approaches (e.g., roleplaying, storytelling). Psychoanalytic, psychodynamic, behavioral, cognitive, transactional, and systemic approaches have been used.

Psychopharmacotherapy and Combined Therapy Major advances have occurred in the use of medications. However, few systematic studies of the safety and efficacy of psychotropic medications for adolescents exist, and they are greatly needed. Nevertheless, as with psychiatric treatment of adult patients, treatment of adolescents may be most effective when combined with psychopharmacology, especially for severe symptoms of psychiatric disorders. Studies of adult patients, such as those with major depressive disorder, obsessive-compulsive disorder, and schizophrenia suggest that psychotherapy plus medication is most effective in reducing symptoms. Studies of the combinations of treatment for attention-deficit/hyperactivity disorder, depressive disorders, obsessive-compulsive disorder, and school phobia in children and adolescents offer clues about the most effective clinical management of these disorders in adolescents.

A 17-year-old girl complained of episodes of rapid heartbeat, sweating, trembling, and fears of going out alone to the shopping mall. She had entered her senior year in high school and was considering her choices of colleges and was planning to take her college entrance examination. Her parents wanted her to go to the college from which her mother graduated to maintain the family tradition. Psychoanalytically oriented outpatient treatment and a selective serotonin reuptake inhibitor (SSRI) treatment were instituted to alleviate the panic disorder symptoms. The psychotherapy focused on the patient's conflicts with her parents; highlighted were chronic concerns that she could not maintain parental expectations and fears of her independence. Medication appeared to reduce symptoms of tachycardia, tremulousness, and preoccupations of lack of competence. Psychotherapy was maintained for 8 months during her last year in high school.

Attention-deficit/hyperactivity disorder has been studied most systematically with regard to combinations of psychotherapy and medications. Psychostimulants such as methylphenidate (Ritalin) and dextroamphetamine (Dexedrine) when combined with behavior therapy or cognitive behavioral therapy are most effective in improving social behavior and academic performance. Advances in drug development has widened the choice of medications to treat mood disorders (e.g., SSRIs) and schizophrenia (e.g., serotonin-dopamine antagonists including risperidone [Risperdal], olanzepine [Zyprexa], and clozapine [Clozaril]). These medications have been applied to adolescent disorders. Systematic research is required to determine the efficacy and safety profiles of these medications for treatment of adolescent psychopathology. No empirical studies of combined psychotherapeutic and medication treatment of mood disorders, anxiety disorders, or schizophrenia in adolescents have been published.

Before starting psychopharmacotherapy with adolescents a comprehensive workup is needed, including a physical examination; blood tests to evaluate hematological, kidney, liver, thyroid and other physiological functions; and an electrocardiogram (ECG) to measure cardiac function. Neurological assessment with an electroencephalogram (EEG) is necessary if seizure disorder is suspected or if the medication is likely to lower the seizure threshold.

Treatment Settings Although most of the treatments described are applicable to outpatient settings, treatments may be modified for a variety of treatment settings such as day hospital, partial hospital, and inpatient services. Treatment within these settings implies a disturbance greater than that in adolescents in outpatient services. The expected degree of change when such treatments are applied requires systematic study. Treatments in inpatient, day hospital, and partial hospital services are currently being systematized in terms of indications, type of treatments offered, length of treatment, and assessment of outcome. Much of this is related to

health care reforms with emphasis on cost containment. However, research is needed to determine the most effective treatment approaches in such settings.

Indications for admission of adolescents to inpatient settings are acute danger to self or others, acute psychotic behavior that cannot be modified in less-restrictive treatment settings, and significant inability to function autonomously. A psychiatric inpatient setting offers multidisciplinary and multimodal treatment approaches. An important element is the milieu in which an adolescent inpatient can practice newly developed social and problem-solving skills. Psychiatric inpatient treatment programs usually combine psychotherapeutic and psychopharmacological treatment.

Partial and day hospital treatment of adolescents has received recent increased attention as a cost-limiting and efficacious treatment for severely disturbed adolescents with more-chronic psychiatric problems. Such treatments augment the psychiatric care in psychiatric hospitalization and often are recommended for continuation of the psychiatric treatment received in a psychiatric hospital. Multimodal treatments that combine individual and group psychotherapy with psychopharmacological treatments are applied. Stays in such treatment settings often are longer than those offered in psychiatric hospitalization. The goal of such treatment is to promote significant improvement that will enable the adolescent patient to live in the general community.

SUGGESTED CROSS-REFERENCES

Additional information on normal development is given in Section 32.2. Discussion of normal adolescent development is in Section 32.3, of neurodevelopment and its assessment in Section 1.3, of psychotherapy or psychopharmacology of adults in Section 30 and 31, and epidemiology in Section 5.1.

SECTION REFERENCES

Angold A, Messer SC, Stangl D, Farmer EM, Costello EJ, Burns BJ: Perceived parental burden and service use for child and adolescent psychiatric disorders. Am J Public Health *88:*75, 1998.

Arnett JJ, Taber S: Adolescence terminable and interminable: When does adolescence end. Youth Adolesc *23:*517, 1994.

*Azima FJC, Richmond LH, editors: *Adolescent Group Psychotherapy*. International Universities Press, Madison, CT, 1989.

Barrett PM, Dadds MR, Rapee RM: Family treatment of childhood anxiety: A controlled trial. J Consult Clin Psychol *64:*333, 1996.

Beautrais AL, Joyce PR, Mulder RT: Psychiatric contacts among youths aged 13 through 24 years who have made serious suicide attempts. J Amer Acad Child Adolesc Psychiatry *37:*504, 1998.

Beck AT: *Cognitive Therapy and Emotional Disorders*. International Universities Press, New York, 1976.

*Blos P: *On Adolescence*. Free Press, New York, 1962.

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, Iyengar S, Johnson BA: A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry *54:*877, 1997.

Brent DA, Kolko DJ: Psychotherapy: Definitions, mechanisms of action, and relationship to etiological models. J Abnorm Child Psychol *26:*17, 1998.

Dishion TJ, Patterson GR, Stoolmiller M, Skinner ML: Family, school, and behavioral antecedents to early adolescent involvement with antisocial peers. Dev Psychol *27:*172, 1991.

Dodge KA, Pettet GS, McKlaskey CL, Brown MM: Social competence in children. Monogr Soc Res Child Dev *51:*2, 1986.

Fairburn CG, Jones R, Peveler RC, Carr SJ, Solomon RA, O'Connor ME, Burton J, Hope RA: Three psychological treatments for bulimia nervosa. Arch Gen Psychiatry *48:*463, 1991.

Goenjian AK, Karayan I, Pynoss RS, Minassian D, Najarian LM, Steinberg AM, Fairbanks LA: Outcome of psychotherapy among early adolescents after trauma. Am J Psychiatry *154:*536, 1997.

Greenhill LL, Pine D, March J, Birmaher B, Riddle M: Assessment issues in treatment research of pediatric anxiety disorders: What is working, what is not working, What is missing, and what needs improvement. Psychopharmacol Bull *34:*155, 1998.

Hinde RA, Stevenson-Hinde J, editors: *Relationships within Families: Mutual Influences*. Oxford University Press, New York, 1988.

Holmes WD, Wagner KD: Psychotherapy treatments for depression in children and adolescents. Psychother Pract Res *1:*313, 1992.

*Institute Of Medicine: *Reducing Risks for Mental Disorders: Frontiers for Intervention Research*. National Academy Press, Washington, DC, 1994.

*Kazdin AE: Psychotherapy for children and adolescents: Current progress and future research. Am Psychol *48:*644, 1993.

Kazdin AE, Bass D, Siegel T, Thomas C: Cognitive-behavioral and relationship therapy in the treatment of children referred for antisocial behavior. J Consult Clin Psychol *57:*522, 1989.

*Kendall PC: Cognitive-behavioral therapies with youth: Guiding theory, current status, and emerging developments. J Consult Clin Psychol *61:*235, 1993.

Kendall PC, Flannery-Schroeder EC: Methodological issues in treatment research for anxiety disorders in youth. J Abnorm Child Psychol *26:*27, 1998.

Kendall PC, Kortlander E, Chansky TE, Brady EU: Comorbidity of anxiety and depression in youth: Treatment implications. J Consult Clin Psychol *60:*869, 1992.

Kendall PC, Panichelli-Mindel M: Cognitive-behavioral treatments. J Abnorm Psychol *23:*107, 1995.

*Klerman G, Weissman MM, Rounsaville BJ, Chevron ES: *Interpersonal Psychotherapy of Depression*. Basic Books, New York, 1984.

Larson R, Ham M: Stress and “storm and stress” in early adolescence: The relationship of negative events with dysphoric affect. Dev Psychol *29:*130, 1993.

Leadbeater BJ, Blatt SJ, Quinlan DM: Gender-linked vulnerabilities to depressive symptoms, stress, and problem behaviors in adolescence. J Res Adolesc *5:*1, 1995.

Leonard HL, March J, Rickler KC, Allen AJ: Pharmacology of the selective serotonin reuptake inhibitors in children and adolescents. J Am Acad Child Adolesc Psychiatry *36:*725, 1997.

*Lewinsohn PM, Clarke GN: Psychosocial treatments for adolescents: A guide to diagnosis and treatment. CNS Drugs *11:*181, 1999.

March JS: Cognitive-behavioral psychotherapy for children and adolescents with OCD: A review and recommendations for treatment. J Am Acad Child Adolesc Psychiatry *34:*7, 1995.

March JS, Amaya-Jackson L, Murray MC, Schulte A: Cognitive behavioral psychotherapy for children and adolescents with posttraumatic stress disorder after a single-incident stressor. J Am Acad Child Adolesc Psychiatry *37:*585, 1998.

McGee R, Feehan M, Williams S, Anderson JC: DSM-III disorders from age 11 to age 15 years. J Am Acad Child Adolesc Psychiatry *31:*50, 1992.

Mufson L, Fairbanks J: Interpersonal psychotherapy for depressed adolescents: A one-year naturalistic follow-up study. J Am Acad Child Adolesc Psychiatry *35:*1145, 1996.

Mufson L, Moreau D, Weissman MM, Wickramaratne P, Martin J, Samoilov A: Modification of interpersonal psychotherapy with depressed adolescents (IPT-A): Phase I and II studies. J Am Acad Child Adolesc Psychiatry *33:*695, 1994.

Offer D, Schonert-Reichl KA: Debunking the myths of adolescence. J Am Acad Child Adolesc Psychiatry *31:*1003, 1992.

Olson DH, Russell CS, Sprenkle DH: *Circumplex Model: Systematic Assessment and Intervention*. Haworth, Binghamton, NY, 1989.

Phares V, Compas BE: The role of fathers in child and adolescent psychopathology: Make room for daddy. Psychol Bull *111:*387, 1992.

Rapoport JL, Leonard HL, Swedo SE, Lenane MC: Obesessive compulsive disorder in children and adolescents: Issues in management. J Clin Psychiatry *54:*27, 1993.

Van Engeland H: Pharmacotherapy and behaviour therapy: Competition or cooperation? Acta Paedopsychiatr *56:*123, 1993.

Werry JS, Wollersheim JP: Behavior therapy with children and adolescents: A twenty-year overview. J Am Acad Child Adolesc Psychiatry *28:*1, 1989.

Wood A, Harrington R, Moore A: Controlled trial of a brief cognitive-behavioural intervention in adolescent patients with depressive disorders. J Child Psychol Psychiatry *37:*737, 1996.

CHAPTER 49. CHILD PSYCHIATRY: SPECIAL AREAS OF INTEREST

49.1 PSYCHIATRIC ASPECTS OF DAY CARE

KLAUS MINDE, M.D., F.R.C.P.(C)

Women with small children represent the fastest growing sector of employed mothers in the labor market. In 1991 about 35 percent of all women in the United States and Canada with children aged 0 to 5 years and 45 percent with those aged 6 to 14 worked full time. An additional 37 percent of mothers with infants and toddlers and 34 percent of those with school-aged children worked part time, bringing the total of working mothers to more than 70 percent, almost double the rate seen 20 years ago. Since mothers want to ensure that their children are well cared for during their absence from home, they have triggered a number of studies that examine the effects of maternal employment on children.

Maternal employment does not only affect children because their mothers spend less time at home. In many dual-worker families both parents and children experience increased stress because women generally have continued their traditional role in child rearing and family life. This has led to role conflicts and guilt in more than 40 percent of both employed men and women. Hence children often do not only have less time with their parents, but they may also be affected by the stress their parents experience from trying to do too much.

EFFECTS OF NONPARENTAL CHILD CARE ON CHILD DEVELOPMENT

Child-Adult Relationships Some highly publicized studies in the late 1980s indicated that early initiated nonparental care (i.e., before 12 months of age) affects the security of infant-parent attachment. Specifically, these authors reported that in five homogeneous samples of 491 maritally intact middle- and working-class families, 43 percent of the infants were classified as insecurely attached if they experienced 20 or more hours a week of routine child care during the first year of life. Infants with less-extensive routine child care showed an insecure attachment rate of only 26 percent. This pattern was confirmed by Michael Lamb and Kathleen Sternberg. However, more-recent work suggests that a number of family and maternal personality characteristics affect both the timing and the extent of maternal employment and consequently the duration and type of nonparental care of children. For example, Edward Melhuish and colleagues showed that of 18-month-old children who had either stayed home with their mothers or were cared for by relatives, by caregivers, or in a nursery, only those in the last two categories showed a diminished number of affectionate acts toward the caregiving adult, less overall responsiveness, and less-engaged activities. However, the characteristics of the mothers who chose each nonparental setting also varied and explained much of the overall variability in their children's behavior. This points out the need to assess children both before and after they enter day care to control for these factors. In just such a study, Jaipaul Roopnarine and Michael Lamb found that 3-year-old children whose parents had decided to enroll them in day care were initially more anxious about, and distressed by, a brief separation than those in a control group who would remain at home. After 3 months of day care, however, group differences had disappeared, suggesting that differences observed between home and day-care children cannot automatically be interpreted as effects of nonparental care. Other, more-recent investigations found no significant relationship between child care experience and attachment security.

One possible explanation of the differences between studies from the 1980s and 1990s could be that today's mothers who place their infants in child care are more informed about the possible effects this may have on attachment patterns. This in turn may make them more sensitive and responsive toward their babies when they are with them after work, which compensates for possible deleterious child-care effects. Or perhaps employed mothers receive more social support today than in the past, which allows them to spend more quality time with their infants when they are with them.

To clarify some of these important questions, the National Institute of Child Health and Human Development (NICHD) funded a major 10-site study of early child care, involving 1153 infants and their mothers. This study had various advantages. It was prospective (i.e., infants were identified at birth and followed through their first 3 years). The authors were able to examine the effects of child care in context, (i.e., as it relates to the age at entry and the quality, amount, and stability over time). Early results of this study showed no significant difference in attachment security related to child-care participation. Even enrollment in extensive, unstable, poor-quality day care was not associated with a higher rate of insecure attachment. However, childhood attachment was affected by a combination of child-care and maternal factors. For example, children who received less sensitive care both at home and from their child care providers (i.e., were at dual risk) showed the highest rate of insecure attachment (up to 56 percent). This confirms work by Albert Hausfather and coworkers. The attachment ratings of less-sensitively mothered children were also affected by more-extensive or unstable care arrangements. Furthermore, high-quality, brief day care served a compensatory function for children with unresponsive and insensitive mothers. This suggests that extensive day care, regardless of its quality, did not help compensate for insensitive mothering. The authors suggest that this is due to a dosage effect for maternal sensitivity and involvement. Children with a less-involved mother may need more time with her to develop a sense of her being available, hence they show the highest rate of secure attachment if they are exposed to brief, high-quality day care.

Relationships With Peers Authors who have investigated relationships with peers have reported results similar to those examining the child-adult relationship following nonparental care experiences. Thus, papers published in the 1980s report that infant day care was associated with later aggression toward peers or other problematic behaviors. However, more-recent work emphasizes the relation between the quality of nonparental care and later behaviors with peers. For example, Carrollee Howes found that children who had experienced low-quality care in infancy displayed poor peer relations in kindergarten. In contrast, the more secure the teacher-child relationship had been in day care, the more gregarious, complex, and empathic the play was between these children at age 4.

Long-term follow-up studies confirm this benefit of high-quality care. For example, Tiffany Field and her colleagues found that 7-year-olds with a long day-care history were more popular, more mature, and less aggressive than children who had entered day care only after the age of 2. Benet-Evis Anderson, in a longitudinal study of Swedish children, confirmed the superior social competence of 13-year-old youngsters who had received other than home care since infancy. This supports the notion that the quality of early nonparental care is associated with later relationship skills toward both adult and peers.

Behavior Problems According to attachment theory, children who are insecurely attached to their parents are less compliant as toddlers than those who have established a secure relationship. Since an increased rate of insecure attachment has often been reported in children who have experienced early nonmaternal care, it is not unexpected to find studies that suggest an association between day care in infancy and later inappropriate noncompliance with parental requests. Carrollee Howes and Michael Olenick studied toddlers at home, in their day-care centers, and in a laboratory situation and found that compliance with adult requests at home and in a laboratory were not correlated. This suggests that compliance or noncompliance is not a trait related to child-care history.

Nevertheless, some reports indicate that children in day care show high levels of externalizing problems and overall poor behavior right after school age; others disagree. Overall then, the literature suggests that nonparental care is sometimes associated with behavioral problems. These difficulties tend to involve externalizing symptoms such as aggression and assertiveness and may be similar to the behaviors these children show toward their peers. However, since there are no good studies that use the quality of day care as a possible defining variable, the available data must be interpreted cautiously. The multisite NICHD study may provide more-contemporary data.

COGNITIVE COMPETENCE

The literature dealing with cognitive competence can be divided into studies that describe infants who were enrolled in intervention programs because of specific biological (e.g., prematurity) or social risk factors (e.g., poverty) and reports that explore the overall effects of infant day care on cognitive development. Specific intervention programs are invariably associated with increased cognitive competence, although very small, premature infants benefit less from enrichment than heavier ones, possibly because of neurological constraints. Furthermore, evidence suggests that the effects of cognitive stimulation, even in the bigger premature infants, do not persist after the enrichment program terminates. For example, in the multisite cohort of 986 premature infants studied by Jeanne Brooks-Gunn and colleagues, the special programs for the treatment group continued for 3 years. The cognitive advantages of the children who had attended the child-care program

were highly significant at 3 years of age (14 intelligence quotient [I.Q.] points) but far less so at age 5, and 3 years later all treatment effects had disappeared. Findings are also variable in non-high-risk populations, although children from disadvantaged families seem to benefit most, especially those in high-quality programs. Cognitive gains are usually assessed as school achievement or verbal abilities. In some Scandinavian prospective long-term studies, 8-year-old children with more center-based care before age 3.5 years scored higher than those with less care, irrespective of family background. This may reflect the uniformly high quality of group care in these countries.

QUALITY OF CARE

Quality has been defined and measured by structural and dynamic measures. Structural measures are aspects of the care setting that allow sensitive adult-child interaction. They include staff-to-child ratios (i.e., there should be no more than three infants under age 2 per caregiver in a good-quality day-care situation), quality of physical settings (e.g., furnishings and toys, physical safety features), experience and training of staff and management, group size, and the type of nutrition the children receive. Dynamic measures are designed to assess the quality of the experience provided to the children. They include the developmental appropriateness of the children's experiences and stimulation (e.g. fine and gross motor activities, language and reasoning experiences, creative experiences), the sensitivity of the caregivers' responses to the children, staff-parent interactions, and staff supervision and evaluation. There are also established instruments that assess the overall environment and the actual adult-child interactions.

Unfortunately, only a minority of children receive high-quality child care services in North America at present. For example, only about 2 percent (1800 of 80,000) of eligible child-care facilities in the United States and 10 percent in Canada had achieved accreditation by their respective national organizations in 1990.

ROLE OF THE PSYCHIATRIC CONSULTANT

Psychiatrists can be helpful in various ways. They can help staff understand especially challenging youngsters and provide direct case consultations. They can also be used on a more systemic level. For example, they may provide regular in-service training for staff and administrators, speak at parent meetings on specific developmental topics, or be available in times of crisis (e.g., documented child abuse). Some child psychiatrists may engage in a mix of these activities.

As with any consultative role, the psychiatrist and the child-care agency must clearly state their respective obligations and roles toward the children, their families, and each other. This engenders mutual respect and trust and makes the work of the psychiatrist more effective. Children in nonparental care situations need multifaceted care, and the psychiatrist can have a well-defined role within this multidisciplinary orchestra of caregivers.

Peter was 11 months old when he was referred to a child psychiatrist by his pediatrician and the director of his day-care program because he did not want to go to sleep at night unless his mother laid down with him and entertained him for up to 1 hour. He would wake up at least twice at night and cry until his mother picked him up. He usually fell asleep again in her arms after 30 minutes. At day care, he was unhappy and did not mix with the other children.

Peter's mother was 38 years old and worked as a fashion designer from 9:00 am to 6:00 pm. His father was an accountant. The couple had been married for 8 years and had tried to have children for 3 years prior to Peter's birth. Both had been delighted when Peter was born and Mrs. A. had taken 6 months off work to be with Peter at home. One week before her return to work, she had enrolled him in a day nursery to ease the transition. She had spent some time in the nursery during his first week there but now only delivered him to the nursery every morning. The staff claimed that Peter slept well in the nursery but did not readily mix with the other children and seemed visibly relieved when he was picked up by his father around 5:30 pm. This had not changed since he had entered the nursery 5 months ago.

Mrs. A. had also weaned Peter during the month preceding her return to work. This had been quite difficult, Peter initially refused to take the bottle for 3 days and had cried for many hours. In fact, his sleep had become problematic at this time, and his mother felt that he “may have missed his midnight breastfeed.” However, she had not dared to continue breast-feeding him in the evening and at night, as she feared he would then refuse the bottle during the day.

Mr. A. was 43 years old and felt that Peter was spoiled and needed more discipline. He did not get up at night to care for Peter but did enjoy playing some special “boys' games” with him after he had brought him home from the nursery. Mr. A. also thought Peter should be allowed to take a bottle into bed at night and was convinced that Peter could and would take the bottle on his own and, as a consequence, sleep much better. Mrs. A. was the oldest of three children. Her parents had emigrated from northern Africa, and she had grown up in a family in which her mother's traditional family values had clashed with her father's demands on his children to meet North American expectations of high professional achievements. Mrs. A. saw her “artistic” profession as a compromise between these two cultures, but she had welcomed the opportunity to join her husband in a city far away from her hometown. The only direct family in town was Mr. A.'s 75-year-old father, who had lost his wife 12 years earlier and lived with another woman in a residence for seniors. He visited the A. family three to four times a year.

Mrs. A. brought Peter to the initial interview and cried a great deal, expressing both exasperation and anger with Peter because he kept her up every night and seemed to be so needy. She discussed her work and the pressure she was under with respect to the upcoming season. She discussed her mother, whom she remembered as aloof and lacking understanding when she was a youngster. She frequently mentioned how much she wanted to be a good mother to Peter. The therapist mentioned how much Peter loved and missed her when she went back to work and how hard it was for him to understand the sudden changes in his life. They discussed Peter's inability to judge time and how that might make his days in the nursery seem indeterminably long. After a play session with Peter and his parents during which he was keen to set up games that involved both his father and mother, the following treatment plan was developed.

1. Mother would accept the offer of her trusted cleaning woman to have her sister look after Peter during the afternoon in his own home.
2. Mother would continue to bring Peter to his day care at 9:00 am but alternate with her husband in picking Peter up at noon and bringing him back home to his new baby sitter.
3. Mother would put Peter to bed each night and develop a specific ritual to facilitate his falling asleep (e.g., two songs and one story before leaving the room).
4. Father would attend to Peter at night, if Peter woke up, by simply putting him down when he cried and remaining in his room without engaging in elaborate conversations or games until Peter had calmed down.

This new plan was discussed with Peter and his parents in the therapist's office. Peter obviously did not understand the details of the discussion but he clearly knew that something important was taking place. The therapist promised to call the family every second evening during the following week and arranged another interview about 2 weeks later.

Follow-up Peter accepted his new babysitter readily, especially since she initially came with her sister, the cleaning woman, and enjoyed taking him out for a walk and let him sit on her shoulder while she prepared his supper. While Peter initially seemed unhappy that his father came to his room at night, he ceased to wake up within a week, improving the level of contentment of all family members. He also seemed delighted to welcome his mother home in the evening and show her what he had done with the babysitter during the day. The nursery also reported improvement, and the staff commented on the sensitivity Peter showed toward other children's feelings.

Discussion This case shows that the sudden transition from being a single child at home to being one of 20 children in a nursery can be extremely stressful for a toddler. The associated sudden weaning obviously added to the sense of loss Peter felt and made his poor sleep and behavior at night quite understandable. While his father seemed to be somewhat aloof and distant, he did respond to his new role as the night caregiver and quickly managed to reassure Peter of his parents' faithfulness. In return, Peter could sleep better and enjoy his mother in a less controlling and anxious way.

Research has shown that regular exposure to nonparental caregivers need not have harmful effects on children's development and their relationships with others. The relationships children develop with their caregivers influence their later behavior and may compensate for, or magnify, the effects of the original parent-child attachment. While previous researchers focused on between-group differences, current interest is focused on preexisting differences among children and within-group effects of the day-care experience. Research also shows that the quality of children's interaction with their family members and the quality of care they receive at home exert the most influence on their later development and behavior. Nonparental care, therefore, is most beneficial if it complements the quality of family care. Psychiatrists can play an important role in sensitizing staff and families to the developmental and emotional needs of groups or individual children. They can also use

their knowledge about child development to influence lawmakers to provide high-quality nonparental care to all those who need it.

SUGGESTED CROSS-REFERENCES

Normal child development is described in Section 32.2. Section 25.12 provides an overview of consultation-liaison psychiatry. Public community psychiatry is discussed in Section 52.1.

SECTION REFERENCES

Bates JE, Marvinney D, Kelly T, Dodge KA, Bennett DS, Pettit GS: Child-care history and kindergarten adjustment. Dev Psychol *30:*690, 1994.

*Beaujot R, Gee EM, Rajulton F, Ravanera ZR: *Family Over the Life Course. Current Demographic Analysis*. Statistics Canada, Ottawa, Canada 1995.

Belsky J: Infant daycare: A cause for concern? Zero Three *6:*1, 1986.

Belsky J, Rovine MJ: Nonmaternal care in the first year of life and the security of infant-parent attachment. Child Dev *59:*929, 1988.

Benn RK: Factors promoting secure attachment relationships between employed mothers and their sons. Child Dev *55:*1224, 1986.

Borge AIH, Melhuish EC: A longitudinal study of childhood behavior problems, maternal employment and day care in a rural Norwegian community. Int J Behav Dev *18:*23, 1995.

Bredekamp S, editor: *Accreditation Criteria and Procedures—Position Statement of the National Academy of Early Childhood Programs—a Division of the National Association for the Education of Young Children*. National Association for the Education of Young Children, Washington, DC, 1987.

*Broberg AG, Wessels H, Lamb ME, Hwang CP: The effects of day care on the development of cognitive abilities in eight-year olds: A longitudinal study. Dev Psychol *33:*62, 1997.

Brooks-Gunn J, Klebanov PK, Liaw F, Spiker D: Enhancing the development of low-birthweight, premature infants: Changes in cognition and behavior over the first three years. Child Dev *64:*736, 1993.

Caughy MO, DiPietro JA, Strobino DM: Day care participation as a protective factor in the cognitive development of low-income children. Child Dev *65:*457, 1994.

Fassler D: Psychiatric aspects of day care. In *Comprehensive Textbook of Psychiatry*, ed 6, HI Kaplan, BJ Sadock, editors. Williams & Wilkins, Baltimore, 1995.

Field T: Quality infant daycare and grade school behavior and performance. Child Dev *62:*863, 1991.

Friedman D: *Family Supportive Policies: The Corporate Decision Making Process*. The Conference Board, New York, 1987.

*Gross D, Sambrook A, Fogg L: Behavior problems among young children in low-income urban day care centers. Res Nurs Health *22:*15, 1999.

Harms T, Clifford RM: *Early Childhood Environment Rating Scale*. Teachers College Press, New York, 1980.

Hausfather A, Toharia A, LaRoche C, Engelsmann F: Effects of entry, daycare quality, and family characteristics on preschool behavior. J Child Psychol Psychiatry *38:*441, 1997.

Hoffman LW: Effects of maternal employment in the two-parent family. Am Psychol *44:*283, 1989.

Howes C: The peer interactions of young children. Monogr Soc Res Child Dev *53:*1, 1988.

Howes C, Hamilton CE, Matheson CC: Children's relationships with peers: Differential associations with aspects of the teacher-child relationship. Child Dev *65:*253, 1994.

Howes C, Olenick M: Family and child care influences on toddler compliance. Child Dev *57:*202, 1986.

Lamb M, Sternberg K: Do we really know how day-care affects children? J Appl Dev Psychol *11:*351, 1990.

*Lamb ME: Effects of nonparental child care on child development: An update. Can J Psychiatry *41:*330, 1996.

MacCarton CM, Brooks-Gunn J, Wallace IF, Bauer CR, Bennett FC, Bernbaum JC, Broyles RS, Casey PPH, McCormick MC, Scott DT, Tyson J, Tonascia, J, Meinert CL: Results at age 8 years of early intervention for low-birth-weight premature infants. The Infant Health and Development Program. JAMA *277:*126, 1997.

Melhuish EC, Mooney A, Martin C, Lloyd E: Type of child care at 18 months—I. Differences in interactional experience. J Child Psychol Psychiatry *31:*849, 1990.

*NICHD Early Child Care Research Network: The effects of infant child care on infant-mother attachment security: Results of the NICHD study of early care. Child Dev *5:*860, 1997.

Prodromitis M, Lamb ME, Sternberg KJ, Hwang CP, Broberg AG: Aggression and noncompliance among Swedish children center-based care, family day care, and home care. Int J Behav Dev *18:*43, 1995.

*Ramey CT: High-risk children and IQ: Altering intergenerational patterns. Intelligence *16:*239, 1992.

Roggman L, Langlois J, Huggs-Tait L, Rieser-Danner L: Infant day-care, attachment, and the “file drawer problem.” Child Dev *65:*1429, 1994.

Roopnarine JL, Lamb ME: Peer and parent-child interaction before and after enrollment in nursery school. J Appl Dev Psychol *1:*77, 1980.

Vandell DL, Corasaniti MA: Variations in early child care: Do they predict subsequent social, emotional, and cognitive differences? Early Child Res Q *5:*555, 1990.

Vandell DL, Powers C: Day care quality and children's free play activities. Am J Orthopsychiatry *53:*493, 1983.

*Van Horn ML, Newell W: Costs and benefits of quality child care. Am Psychol *54:*142, 1999.

CHAPTER 49. CHILD PSYCHIATRY: SPECIAL AREAS OF INTEREST

49.2 ADOPTION

STEVEN L. NICKMAN, M.D.

Adoption, the custom whereby children become the legally recognized sons or daughters of adults who are not their biological parents, has existed in human societies throughout recorded history. It has taken various forms in different cultures; in ancient Babylonia its primary purpose was to provide continuity for the transmission of property or skills, while in the Roman Empire it was often used to elevate the status of an individual who was already an adult. In some areas of the Pacific, adoption of young children formed part of an exchange system between related clans. In some traditional adoption systems the care and upbringing of a child was only one part of a complex system of expectations and obligations, and the adults' agendas were seen as more pressing than those of the children.

HISTORY

Other institutions have provided care and nurturance for children. Child abandonment in classical, medieval, and Renaissance times often led to the child's being taken in by a family. Orphanages arose in Europe during the Renaissance; many boys and girls were “oblated” and taken in by monastic institutions that then raised them with the expectation that they would live a life in the Church. Many apprentices to trades or professions were raised primarily as members of their masters' families.

The concerns expressed by adopted persons about not knowing their roots are as ancient as they are contemporary. Euripides' *Ion* contains a touching dialogue between a woman in search of the child she gave up years before and a young priest of Apollo, who does not know he is the woman's son and who says the only mother he knows is Apollo's priestess.

When the social work profession became established in the third quarter of the nineteenth century, one of its tasks was placing children in adoptive homes. This development was made possible by the enactment of laws about the same time that established adoption as the means whereby a child could be made a legal member of a family (though for some time adoptees did not automatically inherit from their adoptive parents). Because of the fear of “hereditary taint,” infants were commonly kept in hospitals or temporary placements for a number of months, sometimes until 1 year of age, to make certain that the child was healthy in all respects and developing normally. These procedures deviate from contemporary practice in two respects: the lack of awareness that children should be placed as early as possible because of concerns about attachment, and the emphasis on the need of adoptive parents to be certain of getting a normal child (as opposed to the current view of adoption as an institution that exists primarily for the child's benefit).

In the second quarter of this century, as research on child development demonstrated the need for early attachment, adoption practice shifted, and infant placements began to occur in the earliest weeks of life. *Closed adoptions* were the rule (i.e., the identities of the birth parents and the adoptive parents were kept secret). This was thought to facilitate the birth mother's ability to get on with her life, to protect her from later intrusion when the child grew up, and to help the adoptive parents feel unencumbered and fully authoritative in the child's upbringing. From the child's standpoint, information about birth parents was not considered essential for healthy development, and adoptive parents were encouraged to give formulaic responses to a child's questions (e.g., “your birth parents had you placed for adoption because they loved you and couldn't care for you”).

After World War II numerous European children were placed in the United States, and domestic transracial adoptions were tried by some agencies about midcentury. After the Korean War needy Korean children were placed in U.S. homes. These trends arose at the time when traditional infant adoption became more difficult because of the availability of abortion and contraceptive pills and increased social acceptability of single parents raising their children developed. Another growing trend by the early 1970s was the adoptive placement of children formerly considered unadoptable: those past infancy who had survived abuse and neglect and been placed in foster homes, and children with physical, mental, and emotional handicaps.

Beginning at midcentury but increasingly forcefully in the last 25 years, a number of adult adoptees, some professionals, birth parents, and a smaller number of adoptive parents have joined to change the laws and customs surrounding adoption—specifically the secrecy involved in traditional, confidential adoption. State legislatures periodically increased the availability of both nonidentifying and identifying information to adoptees at maturity; however, in only two states can adoptees or birth parents now obtain identifying information. In many states, the placing agency serves as an intermediary, and if mutual permission is extended by both adoptee and birth parent, the agency provides each with the other's name and address. This development is seen as part of a general trend toward openness in adoption, which has several facets: actual *open adoption* in which birth parents and adoptive parents (and in some instances the child) have various degrees of contact with each other and a tendency to encourage adoptive parents to value the birth parents, to raise the child with some knowledge of them, and to accept the child's curiosity about them and support a search if such is the child's desire in late adolescence. It has been argued that traditional closed adoption is flawed and that the institution of adoption should be restructured to allow most adoptees to grow up knowing their birth parents, making allowance for the fact that some birth parents are unavailable, uninterested, or too disturbed to participate usefully in an open adoption. This recommendation is controversial but is attracting considerable attention.

In the case of children adopted from abroad, such matters are more complicated, but there is considerable agency support for adoptive families' attempts to search in Korea and Latin American countries. With children adopted from the foster care system, matters are complicated in a different way, but clearly such children (especially if placed after the toddler years) often maintain some attachment to birth parents and need help, especially during adolescence, to solidify their ties with their adoptive families. Occasionally this involves renewed contact with birth families, to aid the adolescent's identity formation. The open adoption movement is currently influencing these placements as well as infant adoption, and causing agencies and courts to reexamine the old policy of total cutoff. Open arrangements in postinfancy adoption require sensitive and creative management by the placing agency as well as good will and mutual respect on the part of all the parents involved; commonly not all of these requirements are met. The obvious concern of those who question open adoption is that the child will be confused and suffer divided loyalties, but a closed adoption in these circumstances may lead to divided loyalties in any case. Later difficulty frequently arises from the agency's failure or inability to facilitate one or more termination meetings between the child and the birth parent, at which the child can receive the parent's permission to be parented by the adopting parents. The usefulness of therapeutic reunions for disturbed adopted adolescents suggests that having some continued contact with the birth parent stipulated in the adoption arrangement might help to avoid or mitigate a troublesome form of attachment disturbance (multiple emotional attachments).

EPIDEMIOLOGY

From 1.5 to 2 percent of children grow up in adoptive placements with parents who are not blood relatives; approximately another 1.5 percent are adopted by relatives or by stepparents married to one of their biological parents. Figures from the National Adoption Information Clearinghouse indicate that in 1992, there were 127,441 adoptions in the United States. Of these, 42 percent (53,525) were by stepparents or relatives and 58 percent (73,916) were unrelated adoptions. Of the unrelated adoptions, 27 percent (19,753) were conducted by public agencies (i.e., adoptions of children from the foster care system); 8 percent (6536) were from other countries, with largest numbers from Korea (1787), Colombia (403), and Chile (403); and 64 percent (47,627) were either arranged by a private agency or were independent adoptions, in approximately equal numbers.

The number of international adoptions varies with world events. In 1995 there were 9384 such adoptions, the increase being brought about by political changes in Russia, Romania, and other former Communist areas of Europe and by the Chinese government's new policy of allowing infants to be adopted in North America and Europe.

The number of completed adoptions from the U.S. foster care system does not give a full picture of the numbers of waiting children. The National Adoption Information Clearinghouse estimates that as of March 1998 there were approximately 500,000 children currently in foster care in the United States. Of these, 110,000 (just over

20 percent) are eligible for adoption. Figures from 1993, when the corresponding number was 86,000 were broken down as follows: 17,000 children were legally free for adoption and waiting in not yet legalized adoptive homes; 21,000 children were legally free but still waiting to be placed in adoptive families; and 48,000 children were not yet legally free for adoption, some living in preadoptive foster placements, but most in ordinary foster homes or residential facilities. Between 1983 and 1993 the number of children in foster care increased 154 percent in California, 158 percent in Illinois, 123 percent in Texas, 120 percent in New York, and 67 percent in Michigan.

PSYCHOLOGICAL ASPECTS OF ADOPTION

Clinical and Research Findings Since the 1940s the literature of child psychiatry has described various problems in adopted children and adolescents, and debate continues about whether they have more difficulty than their nonadopted peers. In 1960 Marshall Schechter found that adoptees were highly overrepresented in his private psychotherapy practice. He and other psychoanalytically oriented observers described difficulties in resolving oedipal conflicts and in formation of the superego, as well as low self-esteem. These adopted children commonly saw themselves as unwanted, and devalued their birth parents, whom they viewed at times as lower-class or debased individuals. Some authors recommended delaying the disclosure of adoption until midlatency to minimize the confusion that might arise from children having to define their psychological place in the family while simultaneously coming to terms with not having originated there.

Several studies found more adoptees referred for externalizing behaviors than for internalizing problems. More-recent studies continue to find adopted children and adolescents overrepresented in outpatient settings. One study found adoptees overrepresented among their hospitalized children but underrepresented among children referred to the juvenile court. They postulated that the relatively high socioeconomic status of adoptive parents led to their children being referred to the psychiatric network rather than the criminal justice system. Another study of adopted adolescents before the juvenile court found that judges were harsher and parents were more rejecting with them than with a group of neglected adolescents also referred to the court. The authors ascribed the difference in the court's disposition to the fact that the parents of neglected children defended them while the parents of adopted children failed to defend them or even requested placement in correctional facilities.

A weakness of many studies is that insufficient distinction is drawn between groups of adoptees that would be expected to differ in the extent of their disturbance (e.g., those adopted in infancy and those adopted at school age following neglect and abuse). The concept of multiple types of adoption has not been well enough recognized. Thus, some of the figures cited above might be taken as reflecting pathogenicity of infant adoptions to an extent that is not warranted.

Adoption Triad With adoption increasingly considered an institution that exists primarily to benefit the child rather than the adopters, activists and child welfare experts have come to speak of the triad, or triangle, that includes birth parents, adopted person, and adoptive parents.

Adoptees' Losses It has been proposed that adoptees experience three general categories of loss and that different types of placement tend to predispose toward certain types of loss. Most adoptees appear to suffer from all three types at certain times during their development. *Overt losses* are those that would be visible to an observer as they happened: changes of caretaker, loss of familiar surroundings, loss of siblings or foster siblings, pets, and belongings; these losses predominate in children adopted from the foster care system and are prominent in children adopted from abroad after the toddler years. *Covert losses* are experienced intrapsychically and are the inner representations of the adoptee's life situation (e.g., “Why wasn't I kept by my birth parents?” “Was there something wrong with me?” “How can I determine whether and when I want to get more information or search for them?”) In addition, some young adoptees have a propensity to fantasize or to engage in activities that represent symbolically the wish to search or the feeling of being devalued. Time spent engaging in fantasy may affect grades in school, and these symbolic activities may be regarded as acting out and lead to consequences for the child or adolescent. Covert losses are the most prominent form of loss in children adopted in infancy by parents of the same ethnic group as themselves, but they are also common in other infant adoptees as well as later-adopted children. *Status losses* are experienced by adoptees who feel stigmatized or singled out in unwanted ways by virtue of some aspect of their adoptive status—most often, appearing different from their adoptive parents or siblings. This type of loss is seen in its clearest form in transracial and international adoptees, but it occurs generally.

Disclosure and Dialogue In the 1960s, child psychiatrists debated the proper age at which to disclose adoption to children. Early disclosure (in the first 3 years), which had been recommended in the 1940s to prevent later traumatic discoveries, was questioned on the ground that children's ability to negotiate oedipal concerns was interfered with and they were confused about family membership and suffered problems identifying with their adoptive parents. Such concerns would probably be viewed somewhat differently now, and in North America and Western Europe there is general belief that children, as well as people in general, deserve the truth about matters of importance concerning themselves. Adoptive parents are still encouraged to tell children quite early about adoption. An important concept in helping them to do so and to carry out a subsequent dialogue about this with the child as needed is *affect tolerance* in the parent (i.e., the ability to bear a degree of anxiety and sadness within themselves and the child, with the confidence that their relationship with the child will survive and that the child will not ultimately be harmed significantly).

Age of Understanding Research has shown that prelatency children, adopted or not, have a fragmentary or distorted understanding of adoption no matter how much information they have been given. As children move into Jean Piaget's stage of concrete operations they begin to understand complex ideas, and for adoptees this includes the fact that being adopted also means being relinquished. From this one can deduce a clinical principle and a caution. The clinical principle is that children of 7, 8, and 9 years of age might be expected to exhibit sad affect and self-doubt as they begin to grasp the sadness implied by relinquishment—and this in fact occurs frequently. The caution is that this research finding should not be used as a reason for not telling the child earlier (“because he wouldn't understand”); later disclosure has its own complications; school-age children who feel that their parents have not been open with them can have consequent difficulties of trust.

> Amy A. was placed with her adoptive parents at age 20 days. She was now 6 and entering first grade, but after a week refused to go. She said, “I miss my old teacher, Miss Z. I don't like going all day. I miss my mommy.” At moments of friction she would say, “I wish I had stayed with my real mother.” Her first-grade teacher, Mrs. Y, was middle-aged and brisk, unlike last year's young teacher Miss Z, whose gentle manner had made school a wonderful experience for Amy. Mrs. A. was taken aback when Amy confided, “I'm afraid Mrs. Y. will steal me from you the way you stole me from my real mother.” In treatment, Amy demonstrated via puppet play her preoccupation with mother-child relationships, specifically her concern with abandonment. Eventually she made it clear that her fantasy went a step further: the kind, gentle Miss Z. was her original mother, and Amy had experienced the changeover to the more work-oriented, demanding first-grade teacher as a reliving of her fantasy of the transition from biological to adoptive family—a version that could be called “paradise lost.” Both Amy and her parents were pleased with the new light that was shed on her school difficulty, and after a few months of weekly psychotherapy, she was attending school happily and successfully.

Adolescence With the onset of adolescence, adopted young people develop a renewed interest in their origins, often specifically in the behavior patterns of their unknown birth parents with respect to sexuality and other aspects of parental behavior (e.g., drugs, alcohol, or illegal activities) that the adoptee believes may have contributed to the birth parent's “badness” or inadequacy and thus to the adoption itself. When embroiled in negative interactions or power struggles with their adoptive parents, adopted adolescents may test illicit behaviors, both as a trial identification with birth parents and as a test of the adoptive parents' loyalty and support. Both teenage pregnancy and alcohol and other substance abuse can be seen in these circumstances.

Another common theme in adolescence is the adopted son or daughter declaring independence from parents with respect to certain types of schooling; athletic, religious, or musical activities or life goals. This can occur without rancor or strife and can represent the adoptee's true or false belief, that his or her abilities and talents differ from those of the adoptive family—perhaps differences that the parents have not sufficiently taken into account. Parents are well advised not to leap to the conclusion that this merely represents rebellion and should be dealt with as a disciplinary matter. In fact, it may be highly adaptive and may give adolescents a solution to the dilemma of defining themselves as a member of their family, but who in some ways differs from them.

Some adopted adolescents are involved for a time in fairly dramatic conflict with their parents. This is almost the rule for later-adopted children from the foster-care system, and it is fairly common in adolescents adopted as infants. When these families—particularly families of later-adopted children—seek help, it is common for professionals to fail to recognize the permanent commitment of adoptive families and to assume that the adoption did not work out or that the child never formed a solid attachment to his or her parents, even though the adoption may have occurred in infancy and vast amounts of love and nurturance may have gone into making a later adoption work. Psychiatrists, psychologists, social workers, and juvenile court judges need to regard the family as a whole as their client and do whatever they can to keep the family intact. Out-of-home placements may be necessary as a temporary measure (e.g., residential treatment, respite care, or placement with other family members), but the adoptee should only reenter the foster-care system as a last resort, because this implies a failure on the part of the family that may not be accurate and may lead to permanent dissolution of the adoption.

Scott, age 14, was referred by his boarding school for cutting his arms and abrading his knuckles with sandpaper, behavior he had engaged in with another adopted student named Kevin. Scott was adopted in infancy but his parents divorced when he was 7. He had average intelligence but also had a reading disorder and other learning difficulties. He saw his parents rarely; his mother was self-absorbed, and his father lived in Europe. “Life's been pretty hard for me,” Scott said. “I wonder if I belong anywhere. I'm worried about who my (birth) mother and father were. I went to the adoption agency on my last vacation, but they wouldn't give me the address.” With feeling he recited the few factual details the worker had been willing to give him. “I wonder why they gave me up for adoption. I was told she had brown hair and brown eyes. I wonder what she looks like now. Whenever I see someone who looks like that I start to wonder. I want to ask if she had a kid and had it adopted. I wonder if my father is mean, or nice, or strict. I lie in my room and get this weird dreamy feeling; then I get mad and start hurting myself. I found out Kevin was adopted too. He refused to follow the rules. He went to a psychologist and found out he was a battered child. He tried to hang himself once. Right after we found out we both were adopted we figured we might be real brothers, so we cut each other to mix the blood. We think about evil sometimes. If you see it from an adopted kid's view, you're trying to find your parents. Evil is what keeps you from finding them. Once you find your parents and the battle's over, you settle down and live a quiet life.” Asked about his goals, he said, “I don't think about that much. I want to stay young.”

Scott was seen 13 times over a 10-month period. He showed his drawings as well as a story he was writing with themes of good, evil, and secret knowledge. Three themes that preoccupied him were his adoption, his learning disabilities, and his parents' divorce. He was encouraged to think of himself as the hero of his own life story and to see himself as having specific gifts and abilities that would serve him well in life. It was clear that his use of fantasy served at least two purposes—to connect him with his birth family and to protect him from the anxiety-provoking task of thinking about how to move forward in his life. He was given support for his feelings of sorrow about the divorce, for how burdened he felt by the learning disability, and for his worries about having a good future; he was also encouraged to view the cutoff from his birth parents as a sad event that was legitimately worthy of being mourned.

Scott was seen again some months after his sessions had ended. His school had written to say that his attitude about school and work had improved, that he was a happier, more active participant in school, and that he had been elected to student council. Scott reported, “Things are better. My old friends left and I have new friends. I'm on the wrestling team. I've seen too many bad things. I'm trying to get my life started good now, to get on the right track.” He indicated that adoption was no longer as great a preoccupation for him. “Before, people would ask, ‘What religion are you?' or ‘What nationality?' and I'd say ‘I don't know', or ‘I was raised as a Protestant,' or I'm adopted.' Now I figure I don't have to bring adoption into it, I can just tell them about what kind of person I am.”

Adoptive Parents and Grandparents Professionals dealing with adoptive families should recognize that it is difficult for parents to impart knowledge of adoption to children and that this process often evokes feelings about parental infertility, if present, and also evokes empathy for the children, whose picture of the world and their place in it now has the added complications implied by this new knowledge.

Adoptive parenthood is stressful for other reasons as well. As the adopted child grows up, the topic must be dealt with repeatedly, and parents must be able to respond to the child or adolescent's distress, which may at times be manifested by externalizing behaviors. In open adoptions, contact with birth parents must be arranged and managed in a way suitable to the child's interest. In addition, adoptive families are perceived by some members of society as different from other families. This is most true for internationally adopting or transracially adopting families and those who adopt older children, but even parents who adopt same-race infants at times feel stigmatized, sometimes in subtle ways.

For these reasons, adoptive parents' relationship with their own parents becomes very important. Grandparents cannot fully support adoptions if they resent the fact that the adopted child is not biologically related to them. Earlier in this century grandparents commonly left adopted grandchildren out of their wills or showed greater affection to biological grandchildren than to adopted ones, and in some families these still occur. Astute clinicians may help grandparents who are exclusively interested in the perpetuation of their biological line to readjust their values on behalf of themselves, their adult children, and their adopted grandchildren.

A couple in their 60s learned of the engagement of their son, in training for what promised to be a prestigious professional career. He shared with them his mixed feelings about telling his fiance that he had been adopted. They advised him strongly to overcome his hesitation, stressing that there was nothing shameful about being adopted and that if the young woman had negative feelings about it, it would be best to find out now. Though aware of these things intellectually, the son needed his parents' guidance under stress. The couple were married and have a charming 12-year-old daughter.

Psychopathology in Adoptees Most adoptions and most adoptees turn out well. The various outcome statistics indicating that child and adolescent adoptees are more vulnerable to developing problems are best understood with reference to the factors outlined above, also bearing in mind that biological factors (e.g., attention-deficit/hyperactivity disorder, bipolar I disorder) may play a part. Overt, covert, and status losses, combined with any problems within the adoptive family that interfere with empathic disclosure or helpful dialogue, contribute to the formation of internalizing or externalizing problems and disorders of personality. In addition, some observers would consider the way adoption is structured in society, with the attendant secrecy, to be a pathogenic factor.

Adopted Adults On reaching adulthood most adoptees are indistinguishable from their nonadopted peers in terms of behavior and achievement; those who showed externalizing symptoms as children and adolescents have largely discarded them. A small number qualify for an Axis II diagnosis and most of these probably fall into two groups: those who were adopted after infancy following abuse and neglect in their biological families and those whose adoptive parents had some psychopathology and did not help the child deal with feelings about adoption. Many adopted adults do have substantial concerns related to their adoption, however.

Difficulty with relationships is common, particularly vulnerability to loss and reluctance to make permanent commitments. Another frequent problem is low self-esteem. Many articulate adoptees express the belief that society treats them as children by not allowing them full access to their original birth records. Occasionally they encounter prejudice (e.g., a potential spouse who is reluctant to marry a person whose family background is not completely known). Adopted adults often long to have a child of their own; this developmental stage has special meaning for them because it will be the first time they will see an individual who is related to them by blood.

Search The *search* is commonly understood to mean that the adoptee wants to meet and get to know the birth parent, but many who search have more-limited goals, such as learning what the birth parent looks like, obtaining a family medical history, or learning about the birth parents' abilities and interests. Under the closed adoption system, most adoptees reach adult life having had no contact with birth parents and little or no knowledge of them. Society does not provide normative standards for whether or not to search, or when or how to do so, because the prevailing attitude is that adoptions are closed and remain so. Adult adoptees have to decide whether they accept this societal standard or reject it as inconsistent with their way of constructing their lives. Some are deterred by the fear of hurting or alienating their adoptive parents, and in fact many parents of adult adoptees are not sympathetic to the search and view it as a negative judgment about themselves or as a threat to their relationship with the adoptee. It is common for adoptees to begin a search following the death of an adoptive parent or other close relative; in some cases because they felt inhibited from doing so before and in others because an old curiosity reasserts itself in the context of loss.

Adoptees who choose to search can obtain support from the adoptee movement (which also includes many birth parents and some adoptive parents), which contains search and support groups that offer moral support and technical assistance to adopted persons who search. Such groups also promote legislation that facilitates contact between adult adoptees and their birth parents.

The closed adoption system is such an accepted part of society that it is difficult for many people to appreciate the losses involved for individuals. Mental health professionals are aware of the widespread family disruptions during World War II and more recently the theft and subsequent placement for adoption of many infants in Argentina. Similarly, earlier in this century many Native-American and Australian-Aborigine children were moved far from their families to large residential schools with the goal of teaching them the culture and values of the majority. The individual tragedies caused by these events are easy to grasp in a context of oppression visited on a group of people: attenuation or loss of personal relationships as well as alienation from one's cultural heritage. Some advocates of adoptees' rights maintain that closed adoption has elements in common with these large-scale social events because it views infants as tabulae rasae with no right to their pasts and views birth parents as without rights to knowledge of, or contact with, their infants. Advocates of the traditional system argue that most adoptees do well and that the risks of openness might offset any advantages it would provide.

Birth Parents Adults in psychotherapy often fail to mention that they have had a child who was placed for adoption. This important fact should be sought specifically, particularly with depressed women of childbearing age and older, because the resulting losses are commonly unresolved and incompletely mourned. Many birth mothers have kept their experience secret because of shame, and this in itself impairs their ability to grieve the loss because they cannot share it with anyone. Birth parents vary in their attitude about being found by a child they had relinquished. Some birth parents, particularly birth mothers, initiate searches themselves, usually after the child is 18 or older. Concerned United Birthparents (CUB) is an activist organization dedicated to improving the way adoption is structured and to increasing openness in adoption; they are also concerned with decreasing the stigma associated with having relinquished a child and encourage individuals to “come out” as

birth parents.

Reunions and Subsequent Relationships There is no formal etiquette in the search and reunion process, though magazine articles, meetings of adoption organizations, and the Internet provide a great deal of relatively recent material that individuals can use as a guide. Some seekers want information and a brief contact rather than a continuing relationship; others wish to convert the cutoff of closed adoption into an ongoing relationship. Uncertainty and anxiety are unavoidable under these circumstances, especially since the two members of the pair may not be seeking the same level of involvement, and in fact one may be reluctant at best. One recently described phenomenon is so-called genetic attraction, in which newly reunited relatives find themselves interested in each other physically. Such interest needs to be understood primarily as nonsexual curiosity, and adoptee organizations caution successful searchers to use restraint when experiencing such curiosity because the normal incest taboo that arises within families over many years does not exist in these dramatic and often emotional reunion situations. Most adoptees and birth parents appear to feel quite positively about their reunion experiences, whether or not they lead to ongoing relationships, and some adoptive parents and adoptees report that their relationships improve after reunion.

Legal Issues Adopted children and adoptive families continue to be vulnerable in a number of ways and this vulnerability motivates the effort to enact legislation. A currently circulated American Bar Association Model State Adoption Act promotes adoptions but focuses largely on independent adoptions and provides few safeguards for children, adoptive parents, or birth parents. The interests of all these parties are better protected by traditional agency adoptions than by the increasingly common independent adoptions. Foci of proposed legislation include:

1. The right of a birth mother to revoke her consent, and after how long a time a consent to adoption is revocable
2. Federal and state subsidy to special-needs adoption and whether subsidies are permitted when adoptions take place across ethnic or racial lines
3. Rights of Native American tribes to reclaim children who were adopted by non-Indian parents and now have existing emotional attachments to those parents
4. Rights of adopted persons 18 years of age or over to have access to nonidentifying or identifying information about their birth parents or to their original birth certificates

The struggle over specific pieces of adoption legislation reflects the fact that the world of adoption contains several constituencies whose interests do not always coincide. The above discussion contains historical background to illustrate this statement. Contemporary child welfare thinking as well as reasoned statements from adult adoptee and birth parent organizations do largely coincide, however, in the belief that the rights of all concerned (child, birth parent, adoptive parent) must be respected to the greatest extent possible in crafting individual adoption arrangements.

SUGGESTED CROSS-REFERENCES

Foster care is the subject of Section 49.3 Mental retardation is discussed in Chapter 34. Attention-deficit disorders are presented in Chapter 39.1, disruptive behavior disorders in Chapter 40, and mood disorders and suicide in Chapter 45. Psychiatric treatment is the focus of Chapter 48 and childhood and adolescent antisocial behavior is the focus of Section 49.7.

SECTION REFERENCES

Boswell J: *The Kindness of Strangers: The Abandonment of Children in Western Europe from Late Antiquity to the Renaissance* . Pantheon, New York, 1988.

Brinich PM: Some potential effects of adoption on self and object representations. Psychoanal Study Child *35:*107, 1980.

Brinich PM, Brinich EB: Adoption and adaptation. J Nerv Ment Dis *170:*489, 1982.

Brodzinsky DM, Schechter MD: *The Psychology of Adoption*. Oxford, New York, 1990.

Brodzinsky DM, Schechter DE, Brodzinsky AB: Children's knowledge of adoption: Developmental changes and implications for adjustment. In *Thinking About the Family: Views of Parents and Children*, RD Ashmore, DM Brodzinsky, editors. Erlbaum, Hillsdale, NJ, 1986.

*Brodzinsky DM, Schechter MD, Henig RM: *Being Adopted: The Lifelong Search for Self*. Doubleday Anchor, New York, 1992.

Euripides: *The Bacchae and Other Plays*, P Vellacot, translator. Penguin, Baltimore, 1954.

Feigelman W, Silverman AR: *Chosen Children: New Patterns of Adoptive Relationships*. Prager, New York, 1983.

Ferguson DM, Lynskey M, Horwood LJ: The adolescent outcomes of adoption: A 16-year longitudinal study. J Child Psychol Psychiatry *36:*597, 1995.

Green T: *A Man and His Mother*. Harper Collins, New York, 1997.

Gritter JL: *The Spirit of Open Adoption*. CWLA Press, Washington, DC, 1997.

Kim DS: How they fared in American homes: A follow-up study of adopted Korean children in the United States. Child Today *6:*2, 1977.

*Kim WJ, Shin YJ, Carey MP: Comparison of Korean-American adoptees and biological children of their adoptive parents: A pilot study. Child Psychiatry Hum Dev *29:*221, 1999.

Kirk HD: *Shared Fate: A Theory of Adoption and Mental Health*. Free Press, New York, 1964.

*Lifton BJ: *Journey of the Adopted Self*. Basic Books, New York, 1994.

Littner N: The importance of the natural parents to the child in placement. Child Welfare *54:*175, 1975.

Melina LR: *Raising Adopted Children: A Manual for Adoptive Parents*. Harper & Row, New York, 1986.

National Adoption Information Clearinghouse website, www.calib.com/naic.

*Nickman SL: Losses in adoption: The need for dialogue. Psychoanal Study Child *40:*365, 1985.

Nickman SL, Lewis RG: Adoptive families and professionals: When the experts make things worse. J Am Acad Child Adolesc Psychiatry *33:*753, 1994.

Pannor R, Baran A: Open adoption as standard practice. Child Welfare *43:*245, 1984.

Pavao JM: *The Family of Adoption*. Beacon Press, Boston, 1998.

Rietz M, Watson KW: *Adoption and the Family System: Strategies for Treatment*. Guilford, New York, 1992.

Rosenberg EB: *The Adoption Life Cycle: The Children and Their Families Through the Years*. Free Press, New York, 1992.

Rosenfeld AA, Pilowsky DJ, Fine P, Thorpe M, Fein E, Simon MD, Halfon N, Irwin M, Alfaro J, Saletsky R, Nickman S: Foster care, an update. J Am Acad Child Adolesc Pscyhiatry *36:*448, 1997.

Schur WM: The ABA State Model Adoption Act: Observations from an agency perspective. Fam Law Quart *19:*131, 1985.

Schechter MD: Some observations on adopted children. Arch Gen Psychiatry *3:*21, 1960.

*Tizard B: Intercountry adoption: A review of the evidence. J Child Psychol Psychiatry *32:*743, 1991.

Triseliotis J: *In Search of Origins*. Routledge & Kegan Paul, London, 1973.

Watkins M, Fisher S: *Talking with Young Children about Adoption*. Yale University, New Haven, 1993.

49.3 FOSTER CARE

MARILYN B. BENOIT, M.D.

It is estimated that foster care costs $10 billion a year with a per capita cost of approximately $17,500, not including the cost of treatment that the foster child may need. Foster care is defined as temporary out-of-home care that some minor children and adolescents need when their families of origin are unable to care for them and the child welfare system intervenes to provide care. As medical and social problems have evolved, so too have the reasons for children's eligibility for foster care placement. Serious and protracted parental illness, parental death, and financial hardship used to be some of the earlier reasons for foster placement. However, since the 1960s child abuse and neglect, and abandonment secondary to parental substance and alcohol abuse have emerged as major reasons for out-of-home placement. Foster care is primarily utilized by the lower socioeconomic stratum, and is linked to issues of teenage pregnancy, substance abuse, family fragmentation, minority status, poverty, and urban residence. Children in the welfare system, Temporary Assistance for Needy Families (TANF)—previously called Aid to Families with Dependent Children (AFDC)—are overrepresented. Although parental abuse is a significant reason for placement, there has been a shift to abandonment and neglect as the predominant reasons for placement in the 1980s and 1990s. This correlates with the epidemic use of drugs since the mid-1980s. Although most foster children return to their families of origin, an increasing number of children are raised in foster care. These vulnerable children are overrepresented in the population that seeks mental health services and are at high risk for development of psychiatric disorders and substance abuse problems as adults. They are also overrepresented in the criminal justice system where the cost to the public is significant. Unfortunately, little research exists in foster care, and much of what is known is descriptive. The field of psychiatry has not traditionally focused on this population as a target of research.

HISTORY

In his book, *Kindness of Strangers*, John Boswell reveals some disturbing information about the abandonment of children. Children were abandoned or “exposed” and sometimes even killed by their parents because the parents felt unable or unwilling to care for them. Illegitimately conceived children also ran the risk of being “exposed.” This was more common in cities than in the rural areas where illegitimacy was less stigmatized. Such was the case in sixteenth-century Basque country in northern Spain where the illegitimacy rate was high, but abandonment of children was low. Abandoned children were at the mercy of strangers. For example,

> The slave places the baby (an illegitimate child), with pork in mouth, in a small covered shelter at the base of a tree in the woods. He is retrieved four days later and reared by a neighbor well aware of his origins, who nonetheless specifically pretends that the child is his own.

Even the wealthy would abandon children because they wanted to preserve their estates. The monasteries and churches played a major role in the care of abandoned children who were often taken into the monastery either as oblates or as servants to the monks. It is believed that many children who were placed in fosterage, as the term was used, ended up being used as child labor. Boswell reported that of the slaves sold in Sicily, Naples, and Venice in the fourteenth century, about one third were under 15. Many abandoned children were housed in congregate care settings, called foundling homes. In France, the practice of fostering came into the mainstream as a way of caring for abandoned children because it proved to be less expensive for the state. Even in the thirteenth century, foster parents could be of either gender and did not have to be married. The term *fostering* was more liberally applied and could actually refer to apprenticeship and godparent relationships. Because the prevailing sentiment was that the abandoned children were the offspring of “debauchery and crime,” there was concern that they could “inherit” such characteristics. Girls seemed to be at particularly high risk for prostitution. Fostering was extended into late adolescence with the expectation that the children would be apprenticed and educated and would become self-supporting. This is not unlike the current practice in the United States where the state will extend support until age 21 to assist foster children with transition into adulthood. In the early 1800s the concept of family reunification was exemplified by the mothers being given the option of presenting themselves to reclaim their children after paying a fine and reimbursing the state for the costs of having raised their abandoned children.

As early as 1853 Charles Loring Brace had implemented a “placing out” system, the New York Children's Aid Society, to rescue children who had been neglected, maltreated, or abandoned by their families and had consequently become a threat to their communities. These children were sent to distant institutional locations without any connections to families. In 1886 Charles Birtwell of the Boston Children's Aid Society attempted to improve upon this approach by establishing foster care in families. His focus was on prevention of long-term out-of-home placement. The first White House conference on children was held in 1909 and led to a national shift from placing children in institutions to family support for child rearing, that is, foster care. Later, a landmark exposure of “foster care drift” in 1959 revealed that many children were not receiving any permanency planning and that the goals of the 1909 White House conference were being neglected, because few efforts were focused on family reunification. In 1961 Title IVA (Aid to Dependent Children) was adopted under the rubric of the Social Security Act, and its intent was to facilitate family preservation. In 1966 the Casey Family Program (long-term out-of-home placement) was initiated in response to the finding that for some children neither adoption nor family reunification was a viable alternative. Although Title IVB (1993, family preservation PL 103-66) supported this placement option, Title IVB-2 encouraged family rehabilitation to improve the chances that children would return home successfully. Reviews of outcomes of foster care children are mixed. Proponents of the system claim that most children have satisfactory outcomes. Critics have declared that foster care is in crisis, citing that it has come under receivership of the courts in several states for failure to provide adequate care, timely placement, and monitoring of the children in care. In 1980 the federal government acted to promote adoption rather than long-term foster care. Congress passed the Adoption Assistance and Child Welfare Act (PL-96-262), which includes foster care but was expanded in 1985 (PL-99-514) to special-needs children who would have additional funding (for provision of services) when adopted (Title IVE of the Social Security Act). In 1997 President Clinton signed into law a new Adoption and Safe Families Act (PL 105-89). This law is designed to improve provisions for child safety and decrease the length of time a child remains in foster care without permanency planning. It also effectively limits the length of time (12 months) parents must undergo rehabilitation before they can resume parenting of children who have been removed from their homes by the child welfare system. Foster care children generally receive medical services by qualifying for Medicaid under Title XIX of the Social Security Act. Another amendment (PL-99-272) made to the Foster Care and Adoption Assistance Bill authorized use of federal funds for Independent Living Assistance to assist foster care “graduates” ages 16 and above (currently up to age 21) with a smoother transition into independent adulthood. These funds were permanently authorized in 1993 with a 75 percent matching rate for state training.

EPIDEMIOLOGY AND DEMOGRAPHICS

Health and Human Services (HHS) numbers for 1994 indicate that 6.5 of every 1000 children are in foster care versus 3.9 of every 1000 in 1982. This represents a 61 percent increase in foster care during that decade. There are more girls than boys; 55 to 69 percent of the clients are girls. Children placed in care as infants are more likely to stay in care. Whites represent 48 percent of children in care and their numbers have been declining since 1982 while the numbers for African-Americans and Hispanics have been increasing. African-Americans represent 40 percent of the approximately half-million children in placement, with 83.4 percent entering foster care at a mean age of 3 years. This finding is compatible with the trend that since the mid-1980s the age of new entrants into foster care has been decreasing. The under-5 age group currently accounts for the fastest-growing segment of the foster care population. Studies reveal that as many as 62 percent of foster children had prenatal drug exposure. A 1989 HHS study revealed that 30 to 50 percent of all drug-exposed children entered foster care, and among African-Americans, parental drug use precipitated almost 80 percent of foster care placements. Parental drug use is positively correlated with maltreatment, with approximately 89 percent of maltreatment cases involving drug abuse. In one California study, 89 percent of infants tested positive for drugs at birth—cocaine being the identified drug in 85 percent of those who tested positive.

Orpheus is an African-American teenager who has been in out-of-home placement for 6 years. Orpheus' mother is a chronic substance abuser who has made multiple unsuccessful attempts (inpatient and outpatient) to become drug free. His father has been incarcerated for most of Orpheus' life. Orpheus has a history of physical abuse and sexual victimization. He has a history of violent episodes during which he is a significant danger to himself and to others. Psychotherapeutic and psychopharmacological interventions have not been successful in stabilizing him. He started out in foster care but was sent to group homes where his out-of-control behavior precipitated admission to a residential treatment center. He was discharged to therapeutic foster care. Since discharge, he has failed two foster care placements and has now returned to a group home. He has also entered the juvenile justice system because of drug involvement. His mother's substance abuse and the unpredictability of her involvement in his life appear to be the most consistent variables that account for his lack of stabilization. Father absence is also contributory.

Mike is a 16-year-old African-American male who was born to an alcohol- and drug-addicted mother. Father was absent from any involvement in Mike's life. Mike was removed from the home because of neglect. He was diagnosed with fetal alcohol syndrome and attention-deficit/hyperactivity disorder. He has been treated with psychostimulants only. He had borderline intellectual functioning and significant mixed learning problems. After repeated, unsuccessful foster care placements because of his highly aggressive behavior, he was referred to residential treatment at age 11. Mike adjusted well to the demands of the program, engaged positively with the staff, and allowed himself to accept the nurturing provided. He had a court-appointed special advocate who was very predictable and consistent in his involvement with Mike throughout Mike's residential stay. This youngster thrived in residential care. His mother was an infrequent visitor. Mike was discharged at age 13 to therapeutic foster care with a single mother. Continued visits with his biological mother and siblings were part of the discharge planning. At yearly follow-up Mike had maintained his foster care placement, was involved with his mother and siblings, had maintained his discharge school placement, and continued involvement with psychotherapeutic and psychiatric services. He was a nonreader when he came into residential placement, and at the most recent follow-up, his foster mother was delighted to report that he had passed the state literacy test.

NEEDS OF CHILDREN IN FOSTER CARE

The late 1980s and 1990s have registered an increase in the complexity of the cases referred to foster care. There has been an increase in the number of children with emotional and developmental disorders. In several studies, mental health services were repeatedly identified as the number one health care need. Studies show that more than 80 percent of foster care children had developmental, emotional, or behavioral problems. Growth abnormalities (including failure to thrive), neurological abnormalities, neuromuscular disorders, language disorders, cognitive delays, and asthma are significant. It is estimated that those AFDC children who are served in the foster care system have health care needs that are six times as costly as matched peers who do not enter foster care. Approximately 28 percent of foster care children need special education services. Children 0 to 5 years of age are eligible for services under PL101-119, which addresses the educational needs of persons with handicapping conditions. About 25 percent are so emotionally damaged that their mental health treatment needs are high. They use the full range of services: outpatient, acute inpatient, day treatment and partial hospitalization, and residential treatment. In recognizing the needs of foster care children, Title IVE Mandatory Protection of Foster Care Children included (under PL-101-239) a statute that mandates reviews of health and education records. Adolescents in foster care are at increased risk for substance abuse, teenage pregnancies, and sexually transmitted diseases, including human immunodeficiency virus (HIV) disease. With public health care moving in the direction of adopting a managed health care system, particular attention has to be paid to the delivery of services to this medically and psychiatrically vulnerable population.

KINSHIP CARE

Although no good data are available on just how much kinship care is used, its use is definitely increasing. Currently, approximately 23 percent of African-Americans are in kinship care nationwide. In New York City it is estimated that as many as 50 percent of African-American children are in kinship care. Kinship caregivers are generally female (mostly maternal grandmothers) with low income, low education, and minority status. The few studies available indicate that outcomes are more positive than those for children in nonkinship care. Children reportedly receive more positive regard from caregivers in kinship care, which consistently results in more stability than nonrelative foster care. As best as can be determined, there is no demonstrable difference in the need for mental health, medical, and educational services. More states are recognizing kinship care as a reasonable placement option and are authorizing licensing and reimbursement to kinship caregivers.

THERAPEUTIC FOSTER CARE

Therapeutic foster care is a treatment placement that has emerged as an alternative to the more expensive and more restrictive traditional residential treatment center. Beth Stroul states:

> Therapeutic foster care is considered the least restrictive option among the range of residential services for severely emotionally disturbed children and adolescents. Therapeutic foster care can be defined as a service that provides treatment for troubled children within the private homes of trained families. The approach combines the normalizing influence of family-based care with specialized treatment interventions, thereby creating a therapeutic environment in the context of a nurturant family home.

Therapeutic foster parents are the agents of therapeutic change, functioning as “extenders” of the clinical treatment team. They are required to have more extensive training than regular foster parents, receive a higher reimbursement because of the special-needs status of the children, are subjected to more intensive monitoring and supervision, and receive more ongoing support from the foster care agency. Although the concept of therapeutic foster care is a promising one, no reliable outcome data are available to determine just how successful it is as an intervention with emotionally troubled foster care youth.

CULTURAL COMPETENCE

According to Anna McPhatter, “Cultural competence denotes the ability to transform knowledge and cultural awareness into health and/or psychosocial interventions that support and sustain healthy client-system functioning within the appropriate cultural context.” The Multiethnic Placement Act of 1994 urged authorities to “consider child's cultural, ethnic and racial background” when making placement decisions. Of approximately 460,000 children in care in 1994 approximately 14 percent were Hispanic, 48 percent white, and almost 38 percent were African-Americans, about three times their representation in the general population. Societal racism has resulted in some children being denied placement with families of a different race, and they ended up in long-term foster care rather than in a permanent adoption placement. Minorities are twice as likely to remain in long-term foster placement, so this has been an issue primarily for African-Americans and Native Americans. As the Hispanic foster care population increases, it can be anticipated that similar issues will emerge for this ethnic group. In 1978 the Indian Child Welfare Act (PL-95-608) transferred to Tribal Courts the power to make placement decisions about Native-American children to reverse the practice of placement in non–Native-American homes. The Association of Black Social Workers went on record as opposing transracial placement of African-American children. Adoption studies have shown that it is not inherently harmful for children to be cross-racially adopted. Congress has passed legislation facilitating transracial adoptions while maintaining the language of cultural awareness in placement decisions. Shawan Gregory and Frederick Phillips of a foster care placement agency in Washington, D.C., Progressive Life Center (PLC), have an Afrocentric approach, “Ntu,” to therapeutic foster care. They describe the “Ntu” perspective as an “Afrocentric, spiritually based, therapeutic framework,” that harnesses “the positive force that is universal to all people.” Their process of admission to therapeutic foster care is carefully orchestrated to include African rituals–a libation and “crossing over.” The principles of Kwanzaa are incorporated into the practices. No research data are yet available from this organization, but a descriptive report states that 74 percent of their clients “improved significantly emotionally” and thus “PLC's faith was justified.” The need for cultural sensitivity, respect, and a capacity to facilitate a foster child's cultural development and identity are well acknowledged. These issues must be addressed in the training of all who provide services to the children in the foster care system. However, no rigorous research has yet identified the essential variables that ensure “success” in cultural competency.

PSYCHOLOGICAL ISSUES IN FOSTER CARE PATIENTS

Foster care children are overrepresented in psychiatric populations. Among those who return home, 40 percent reenter the foster care system. Given the reasons why children enter foster care, the major issues with which they struggle involve abandonment, neglect, rejection, and physical, emotional, and sexual maltreatment. The age of entry into the foster care system and the specific reason for placement affect the emotional issues that any specific child must handle. Early abandonment and neglect of infants and young children can lead to anaclitic depression. Attachment issues are prevalent in this younger population because they have not had an early opportunity to form a secure attachment with a consistent nurturing figure. They are often unprepared for separations, which may be abrupt and repeated. Because of financial constraints, the foster care system simply does not allow time for good relationships to form with foster parents prior to placement.

Joseph is a 13-year-old boy who has been in residential treatment for the past 15 months. He has done extremely well in residential care after five foster care placement failures and psychiatric hospitalizations. He was removed from the parental home because of neglect. He has demonstrated a remarkable capacity for appropriate attachment. He is being discharged to foster care because his mother cannot care for him. A request was made to the department of child welfare that his discharge planning should include a 3-month period of preplacement and graduated home visits with the identified foster mother to promote trust and bonding. This request was denied by the agency and the court. Joseph's emotional and behavioral condition so deteriorated prior to discharge that he required one-to-one suicidal observation, and discharge had to be postponed.

If the child-serving agencies focused on the psychological needs of the child, rather than on the bottom line, thoughtful decision-making about placement of children would lead to better outcomes and ultimately improve the bottom line they so arduously protect. Children who experience foster care drift or foster care bouncing have their capacity to form enduring emotional attachments further compromised and the issue of trust becomes a lifelong challenge. Defense mechanisms tend to be primitive, that is, denial, splitting, projection, and introjection. Characterological problems are not uncommon in adulthood. It is not at all surprising that children who have never had their narcissistic needs appropriately met grow up to become unempathetic adults with pathological narcissism. Children who have experienced traumatic physical and sexual abuse develop behaviors evidenced by dissociation, mistrust, hypervigilance, aggressivity, impulsivity, oppositionality, and avoidance as they attempt to negotiate a world they experience as threatening, hostile, and uncaring. Unmet narcissistic needs render them demanding and unempathetic. When children are raised in a psychosocial environment of trauma, aggression, and lack of empathy from adults, the psychological seeds are sown for later violence against the self and others.

FAMILY PRESERVATION

Much hope was invested in the thrust toward family reunification and family preservation. Estimates on the percentage of children who are reportedly reunited vary from 66 to 90 percent. However, the finding that about 40 percent of reunified children reenter out-of-home care is disappointing. Philosophically, family reunification appears to be the right thing to do, but too little attention has been given to developing discriminating criteria to identify psychosocial profiles of families that would benefit most from family preservation services. In 1996 the Child Welfare League of America acknowledged the failure of family preservation efforts and requested that the child welfare policy makers rethink the current use of intensive family preservation. Unfortunately, no research exists to guide thinking and implementation of sound policy to improve outcomes. One hopes that the relatively new Adoption and Safe Families Act of 1997 (PL-105-89) will permit child welfare agencies to step back from the rather myopic view of family preservation and seriously consider the needs of the child before making subjective decisions.

FOSTER CARE OUTCOMES

The overall quality of available outcome studies is poor; however, some patterns recur across studies. One disturbing finding is that studies reveal that 15 to 39 percent of the homeless are foster care graduates. They are also overrepresented among adult substance abusers and clients in the criminal justice system. It is impossible to conclude that foster care is the single major contributor to such negative outcomes. What is more likely is that the reasons that initially precipitated foster care placement remained over time as etiological factors in the negative adult outcomes. Studies indicate that children entering care who have been victimized, who have substance-abusing parents or parents with major mental illness or high criminality, and who come from homes with a high degree of domestic violence are at greater risk of having poor outcomes. As regards a dose-response model, studies indicate quite consistently that longer time spent in care leads to a more successful outcome. Children returned to their families of origin typically have fared worse than those who have remained in long-term placement. Several studies report findings indicating that poor parental involvement consistently leads to negative outcomes. Multiple placements have also been shown to lead to less successful outcomes.

Federal mandate (PL96-272) requires states to maintain a tracking system for children in foster care. New reporting systems due to be adopted are AFCARS (Adoption and Foster Care Analysis and Reporting System) and SACWIS (Statewide Automated Child Welfare Information System). States are being monitored for compliance, and continued federal funds are contingent upon the implementation of these information systems. With improved data collection, the child welfare system should be better informed about the management of the children in foster care, the needs of the children and their families, and how best to address those needs with the limited resources that can be allocated. Because foster care placement is itself the result of psychosocial environmental failure, fixing the existing system will require more than good information systems. Integration of sound, theory-driven, child-focused services provided by multiple governmental agencies will be essential. Three states—Alabama, Ohio, and North Carolina—have disseminated some early performance outcomes. Through the use of longitudinal, research-based performance measures, reliable data are emerging. The complexity of the impact of ever-changing psychosocial variables makes this type of research challenging, but it is necessary if our welfare dollars are to be spent doing the right thing for needy children and families.

SUGGESTED CROSS-REFERENCES

Sociology and psychiatry are covered in Section 4.2 and adult antisocial behavior and criminality are addressed in Section 27.3. Child psychiatry is addressed in Section 32, Section 33, Section 34, Section 35, Section 36, Section 37, Section 38, Section 39, Section 40, Section 41, Section 42, Section 43, Section 44, Section 45, Section 46, Section 47, Section 48 and Section 49. Specifically, residential and in-patient treatment are addressed in Section 48.8 and adoption is addressed in Section 49.2.

SECTION REFERENCES

American Academy of Pediatrics, Committee on Early Childhood, Adoption, and Dependent Care: Health care of children in foster care. Pediatrics *93:*335, 1994.

*Barth RP, Price A: Shared family care: Providing services to parents and children placed together in out-of-home care. Child Welfare *78:*88, 1999.

*Battistelli ES: *Making Managed Health Care Work For Kids in Foster Care: A Guide to Purchasing Services*. CWLA Press, Washington, DC, 1996.

Benoit M: It's the quality not the category of care. In *When Drug Addicts Have Children*, D Besharov, editor. CWLA Press, Washington, DC, 1995.

Berrick JD: When children cannot remain home: Foster family care and kinship care. Future Child *8:*72, 1998.

Boswell J: *The Kindness of Strangers*. Vintage Books, New York, 1988.

Chamberlain P: *Family Connections: A Treatment Foster Care Model for Adolescents with Delinquency*. Castalia Publishing Company, Eugene, OR, 1994.

*Courtney ME, Barth R: Pathways of older adolescents out of foster care: Implications for independent living services. Soc Work *41:*75, 1996.

Craig C: What I need is a mom. Policy Rev *73:*2, 1995.

*CWLA: *Standards of Excellence for Family Foster Care Services*. CWLA Press, Washington, DC, 1995.

Everett JE: Relative foster care: An emerging trend in foster care placement policy and practice. Smith Coll Stud Soc Work *65:*239, 1995.

Fanshel D: Foster care as a two-tiered system. Children and Youth Serv Rev *14:*49, 1992.

Franck EJ: Prenatally drug exposed children in out of home care: Are we looking at the whole picture? Child Welfare *75:*19, 1996.

Fuchs R: *Abandoned Children: Foundlings and Child Welfare in Nineteenth Century France*. State University of New York Press, Albany, 1984.

Galaway B, Nutter RW, Hudson J: Relationship between discharge outcomes for treatment of foster care clients and program characteristics. J Emot Behav Dis *3:*46, 1994.

Gebel TJ: Kinship care and non-relative family foster care: A comparison of caregiver attributes and attitudes. Child Welfare *75:*5, 1996.

*Genty PM: Permanency planning in the context of parental incarceration: Legal issues and recommendations. Child Welfare *77:*543, 1998.

Gil E: *Foster Parenting Abused Children*, ed 3. National Committee to Prevent Child Abuse, South Deerfield, MA, 1994.

Goerge R, Wulczyn F, Fanschel D: A foster care research agenda for the 90s. Child Welfare *73:*525, 1994.

Gregory SD, Phillips FB: Of mind, body, and spirit: Therapeutic foster care—An innovative approach to healing from an NTU perspective. Child Welfare *76:*127, 1997.

Halfon N, Mendonca A, Berkowitz G: Health status of children in foster care: The experience of the center for the vulnerable child. Arch Pediatr Adolesc Med *149:*386, 1995.

Heath AF, Colton MJ, Aldgate J: Failure to escape: A longitudinal study of foster children's educational attainment. Br J Soc Wk *24:*241, 1994.

Hegar R, Scannapieco M: From family duty to family policy: The evolution of kinship care. Child Welfare *74:*200, 1995.

Henderson J, Wall R, editors: *Poor Women and Children in the European Past*. Routledge, London, 1994.

McCue Horwitz S, Simms MD, Farrington R: Impact of developmental problems on young children's exits from foster care. Dev Behav Pediatr *15:*105, 1994.

*McDonald TP, Allen RI, Westerfelt A, Piliavin I: *Assessing the Long-Term Effects of Foster Care: A Research Synthesis*. CWLA Press, Washington, DC, 1996.

*McNichol T: The impact of drug-exposed children on family foster care. Child Welfare *78:*184, 1999.

McPhatter AR: Cultural competence in child welfare: What is it, how do we achieve it, what happens without it? Child Welfare *76:*255, 1997.

Mallon GP: After care, then where? Outcomes of an independent living program. Child Welfare *77:*61, 1998.

Mech EV: Foster youths in transition: Research perspectives on preparation for independent living. Child Welfare *73:*603, 1994.

Pilowsky DJ, Kates WG: Foster children in acute crisis: Assessing critical aspects of attachment. J Am Acad Child Adolesc Psychiatry *35* (8):1095, 1996.

*Reddy LA, Pfeiffer SI: Effectiveness of treatment foster care with children and adolescents: A review of outcome studies. J Am Acad Child Adolesc Psychiatry *36:*581, 1997.

*Rosenfeld AA, Pilowsky DJ, Fine P, Thorpe M, Fein E, Simms M, Halfon N, Irwin M, Alfaro J, Saletsky R, Nickman S: Foster care: An update. J Am Acad Child Adolesc Psychiatry *36:*448, 1997.

Rosner D, Markowitz G: Race, foster care, and the politics of abandonment in New York City. Am J Public Health *87:*1844, 1997.

*Simms MD, Freundlich M, Battistelli ES, Kaufman ND: Delivering health and mental health care services to children in family foster care after welfare and health care reform. Child Welfare *78:*166, 1999.

Stroul BA: *Series on Community-Based Services for Children and Adolescents who are Severely Emotionally Disturbed, vol III: Therapeutic Foster Care*. CASSP Technical Assistance Center, Georgetown University Child Development Center, Washington, DC, 1989.

Toth J: *Orphans of the Living: Stories of America's Children in Foster Care*. Simon & Schuster, New York, 1997.

*Wells K, Tracy E: Reorienting intensive family preservation in relation to public child welfare practice. Child Welfare *75:*667, 1996.

Wiener JM: Orphanages: An idea whose time has come again? Am J Psychiatry *155:*1307, 1998.

*Wilson L, Conroy J: Satisfaction of children in out-of-home care. Child Welfare *78:*53, 1999.

49.4 CHILD MALTREATMENT

WILLIAM BERNET, M.D.

Child maltreatment is a terrible misfortune for millions of children and families, for communities, and for society. It affects children of all colors, social classes, ethnic groups, and religions. It damages young people of all ages—prior to birth, infants, children, and adolescents. Recognition and diagnosis of child abuse is compromised by the denial that pervades all aspects of the phenomenon: perpetrators routinely deny they did it; abused children sometimes deny that anything bad happened to them; otherwise concerned adults, such as neighbors and teachers and emergency room physicians, may overlook and underreport abuse for any number of personal reasons; complex legal requirements may make it hard to convict abusive parents; and generalized societal denial protects us from awareness of a very unpleasant side of life.

The evaluation, management, and treatment of child abuse requires the cooperation of diverse professional groups, including pediatricians and other primary care physicians, emergency room staff, radiologists, pathologists, attorneys, social service personnel, and a variety of mental health professionals. Psychiatrists and other mental health professionals participate in child abuse cases in several ways. Psychiatrists may evaluate children for a forensic or legal purpose, either in private practice or as part of an interdisciplinary team. They may assist the court in determining what happened to the child and make recommendations regarding placement or treatment. In particular, child psychiatrists may be asked to evaluate the credibility of children who may have been sexually abused. Psychiatrists may provide individual or family therapy for a child who has emotional or behavioral problems caused by abuse and may treat child and adolescent perpetrators of abuse, many of whom were previous victims of abuse. Psychiatrists may evaluate perpetrators of abuse to clarify the psychiatric diagnosis and assess the person's competency to go to trial, the possibility of an insanity defense, the person's potential for treatment, and the appropriateness of termination of parental rights. In civil law suits, psychiatrists may testify about the cause, nature, and extent of the child's psychological injuries. In some forms of abuse, such as factitious disorder by proxy, psychiatric evaluation may be particularly important in devising treatment approaches for both victim and perpetrator. Mental health professionals may deal with these issues on the level of public policy by sharing information and educating attorneys, judges, and legislators about the psychiatric aspects of abuse and the developmental needs of children.

DEFINITIONS

DSM-IV The somewhat terse classification system provided by the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) lists physical abuse of child, sexual abuse of child, and neglect of child. The three classifications appear under and in the section problems related to abuse or neglect, which is grouped with other conditions that may be a focus of clinical attention. These categories should be used when the focus of clinical attention is severe maltreatment of one individual by another. These conditions and problems are coded on Axis I.

DSM-IV does not include detailed definitions or criteria for diagnosis. Circumstances in which the focus of attention is on the perpetrator of child abuse or neglect or on the relational unit in which it occurs have one code. Three different codes cover situations in which the focus of attention is on the victim (for neglect of child, for physical abuse of child, or for sexual abuse of child).

Federal Law The Child Abuse Prevention and Treatment Act was passed in 1974 and has been amended several times. In the federal law “child abuse and neglect” means “the physical or mental injury, sexual abuse or exploitation, negligent treatment, or maltreatment of a child by a person who is responsible for the child's health or welfare, under circumstances which indicate that the child's health or welfare is harmed or threatened thereby.” “Sexual abuse” includes “the rape, molestation, prostitution, or other form of sexual exploitation of children, or incest with children” and the creation of child pornography.

State Law A large mass of legal definitions and guidelines exists at the state level. The legal definitions of terms related to the maltreatment of children vary from state to state, so clinicians should be aware of the definitions used in their own locale.

Neglect The most prevalent form of child maltreatment, *neglect* is the failure to provide adequate care and protection for children. Children can be harmed by malicious or ignorant withholding of physical, emotional, and educational necessities. Neglect includes failure to feed children adequately and protect them from danger. Physical neglect includes abandonment, expulsion from home, disruptive custodial care, inadequate supervision, and reckless disregard for a child's safety and welfare. Medical neglect includes refusal, delay, or failure to provide medical care. Educational neglect includes failure to enroll a child in school and allowing chronic truancy.

Physical Abuse *Physical abuse* may be defined as any act that results in a nonaccidental physical injury, such as beating, punching, kicking, biting, burning, and poisoning. Some physical abuse is the result of unreasonably severe corporal punishment or unjustifiable punishment. Physical abuse may be organized by the site of injury: damage to skin and surface tissue, damage to the head, damage to internal organs, and skeletal damage.

Psychological Abuse *Psychological abuse* occurs when a person conveys to children that they are worthless, flawed, unloved, unwanted, or endangered. The perpetrator may spurn, terrorize, isolate, or berate the child. Emotional abuse includes verbal assaults (e.g., belittling, screaming, threats, blaming, or sarcasm), unpredictable responses, persistent negative moods, constant family discord, and double-message communications. Some authors feel that the terms “psychological” or “emotional abuse” should not be used and that “verbal abuse” more accurately describes the pathological behavior of the caregiver.

Sexual Abuse *Sexual abuse of children* refers to sexual behavior between a child and an adult or between two children when one of them is significantly older or uses coercion. The perpetrator and the victim may be of the same sex or the opposite sex. The sexual behaviors include touching breasts, buttocks, and genitals, whether the victim is dressed or undressed; exhibitionism; fellatio; cunnilingus; and penetration of the vagina or anus with sexual organs or objects. Sexual abuse may involve behavior over an extended time or a single incident. Developmental factors must be considered in assessing whether sexual activities between two children are abusive or normative. In addition to the forms of inappropriate sexual touching, sexual abuse also refers to sexual exploitation of children, for instance, conduct or activities related to pornography depicting minors and promoting or trafficking in prostitution of minors.

Ritual Abuse Cult-based *ritual abuse,* which includes satanic ritual abuse, is physical, sexual, or psychological abuse that involves bizarre or ceremonial activity that is religiously or spiritually motivated. Typically, multiple perpetrators abuse multiple victims over an extended period of time. Ritual abuse is a controversial concept; some professionals believe that ritual abuse is a common, horrible phenomenon in our society, while others are deeply skeptical about most allegations and descriptions of ritual abuse.

Perpetrators of Abuse There is some lack of consistency in who may be defined as an abuse perpetrator. Usually a person must be a parent or designated caregiver to be charged with neglect, physical abuse, or emotional abuse. Another adult (e.g., a stranger) who injures a child would be charged with battery, not with child abuse. On the other hand, a caregiver or any other person could be charged with child sexual abuse. The state laws vary in this regard.

HISTORY Historical and literary sources have documented child maltreatment for thousands of years in cultures from all parts of the earth. Although some forms of

abuse were considered deviant or prohibited in particular cultures, most often the abusive practices were accepted or even encouraged by the societies in which they occurred. The long-term trend in European and American history has been to move from blatant, socially acceptable child abuse to strong societal disapproval. At least four historical periods can be identified.

Ancient Period In the ancient period adults seduced, violated, and murdered their children in an unashamed and socially acceptable manner. Human sacrifice, including of children, was practiced by Greeks, Romans, Celts, Phoenicians, and others. The Old Testament criticizes human sacrifice several times, so it apparently was practiced during that era. Greeks, Romans, and Egyptians commonly practiced both heterosexual and homosexual pedophilia. The most notorious sadistic child abusers were the emperors Tiberius, Caligula, Claudius, and Nero.

Medieval Period Christianity induced guilt, and parents could no longer abuse their children with impunity. When they did, they asserted that they were following God's will. The notion of an incest taboo—a universal prohibition against incest that presumably transcends cultures—is an ironic manifestation of societal denial. Some authors think that the opposite is true, that incest has been universal for most people in most places at most times.

Early Modern Period In the eighteenth, nineteenth, and early twentieth centuries, the sense of guilt and shame increased to such a degree that incest and child abuse could no longer be considered acceptable. However, abusive practices continued in the underworld, out of sight of polite society. (Infanticide through exposure or poisoning was a long-accepted way to dispose of an unwanted or defective child.) Foundling hospitals were established to care for thousands of abandoned babies, but few children survived the dismal routine of institutions. Charles Dickens (*Oliver Twist*, 1837–1839) and Victor Hugo (*Les Miserables,* 1862) laced their story lines with social criticism and graphically portrayed man's inhumanity. Dickens, for instance, recorded the inhuman aspect of industrial society, including child battery, debasement of children for sensual or recreational purposes, and starvation and deprivation of the laboring child.

Toward the end of this period societal awareness of child maltreatment increased. In the 1860s a French forensic pathologist, Ambroise Tardieu, described severe child abuse after performing autopsies on children who had been beaten to death. Child abuse came to public attention in the United States through the case of Mary Ellen, an 8-year-old girl who was severely maltreated. She was discovered by church workers in New York City in 1874, but they found that the only agency available to help was the Society for the Prevention of Cruelty to Animals. Thus, they founded the Society for the Prevention of Cruelty to Children. In 1875 New York became the first state to adopt a child protection law, which became the model for other states.

Late Modern Period In the latter half of the twentieth century child abuse has become the subject of widespread clinical and governmental concern. Abusive practices are actively identified, rather than ignored and denied. Modern recognition of child abuse was signaled in 1946 by a pediatrician and radiologist, John Caffey, who noticed a syndrome of children with multiple skeletal injuries and chronic subdural hematomas. Up until the 1960s physical abuse of children was considered rare, partly because physical discipline of children was generally more acceptable and partly because of societal denial concerning violence toward children. In 1962 Henry Kempe and his colleagues brought child maltreatment to the attention of physicians in a historic article, “The Battered-Child Syndrome.” Within 5 years child abuse reporting laws were adopted by every state.

It took separate societal realization during the 1970s to acknowledge the extent of sexual abuse. Incest was known to occur, but most people thought that it was very unusual and that it happened primarily among very deviant families. It is now known that incest and other forms of sexual abuse are not rare.

In 1974 the federal government passed the Child Abuse Prevention and Treatment Act, which established uniform national standards for identification and management of child abuse cases. Child abuse legislation, originally directed toward case discovery and perpetrator punishment, now focuses on prevention, rehabilitation, and family reunification. Individual states independently define maltreatment, formulate investigative procedures, collect data, and organize service delivery systems.

COMPARATIVE NOSOLOGY

The first and second editions of DSM, (DSM-I and DSM-II) made no mention of child maltreatment. The Group for the Advancement of Psychiatry mentioned child maltreatment among the pathogenic factors of childhood mental disorders: “The deleterious effects ... of open rejection or neglect, occasionally involving willful injury or inadequate feeding, are well documented.” However, the Group for the Advancement of Psychiatry did not include child maltreatment in their proposed system of classification.

When the third edition of DSM (DSM-III) was published, the American Psychiatric Association (APA) Task Force on Nomenclature and Statistics introduced a chapter called “V Codes for Conditions Not Attributable to a Mental Disorder That Are a Focus of Attention or Treatment.” The V Codes, which were modeled on the practice in *International Classification of Disease, 9th Revision, Clinical Modification* (ICD-9-CM), did not include specific categories for child maltreatment. However, it was intended that child abuse would be coded as a “parent-child problem,” the category used “when a focus of attention or treatment is a parent-child problem that is apparently not due to a mental disorder of the individual (parent or child) who is being evaluated.” DSM-III specifically stated that “an example [of parent-child problem] is child abuse not attributable to a mental disorder of the parent. The emphasis of parent-child problem” was apparently on identifying perpetrators of abuse, not the child victims of abuse.

The revised third edition of DSM (DSM-III-R) provided a similar definition for parent-child problem: “This category can be used for either a parent or a child when the focus of attention or treatment is a parent-child problem that is apparently not due to a mental disorder of the person who is being evaluated.” However, the example involving child abuse was replaced by “conflict between a mentally healthy adolescent and her parents about her choice of friends.”

In DSM-IV the name of the chapter changed to “Other Conditions That May Be a Focus of Clinical Attention.” Also, a specific section was introduced, “Problems Related to Abuse or Neglect,” which included terms related to child maltreatment. For the first time it was possible to diagnose clearly “physical abuse of child,” “sexual abuse of child,” and “neglect of child.” Furthermore, DSM-IV made it possible to distinguish whether the person being evaluated is the perpetrator of abuse or the victim of abuse.

DSM-IV terminology is the same as that in the 10th revision of *International Statistical Classification of Diseases and Mental Health Problems* (ICD-10); only the codes are different.

EPIDEMIOLOGY

The National Committee to Prevent Child Abuse collects data each year on the incidence of child maltreatment. The committee estimated that in 1997 almost 3.2 million alleged victims were reported to child protective services. Of those reports about 1 million were substantiated; this represents about 15 of every 1000 children. The substantiated cases were distributed as follows: neglect, 54 percent; physical abuse, 22 percent; sexual abuse, 8 percent; emotional abuse, 4 percent; and other or unspecified cases, 12 percent.

The committee reported that in 1996 over 1000 children died as the result of maltreatment. About 38 percent of these deaths were children under age 1. Many fatalities were linked to parental substance abuse. A large number (41 percent) of the deaths were children who were current, open cases or previous clients of a child protection service agency.

The National Center on Child Abuse and Neglect also collects data each year on child maltreatment. The center estimated that in 1995 the median age of victims of child maltreatment was 7 years. Of the victims, about 53 percent were girls and 47 percent were boys. It was reported that 79 percent of the victims were abused by parents; 10 percent by other relatives; 5 percent by noncaregivers; and 2 percent by foster parents, facility staff, or child-care providers.

These figures are approximations because the actual amount of abuse is unclear. The reporting of abuse has increased in recent years. This increase resulted in large part from greater public awareness and willingness to report child abuse, improved data collection techniques in individual states, and local economic conditions that place a larger number of families under stress.

ETIOLOGY

Physical Abuse Although child abuse occurs at all socioeconomic levels, it is highly associated with poverty and psychosocial stress, especially financial stress.

Child maltreatment is strongly correlated with less parental education, underemployment, poor housing, welfare reliance, and single parenting. Child abuse tends to occur in multiproblem families, that is, families characterized by domestic violence, social isolation, parental mental illness, and parental substance abuse, especially alcoholism. The probability of maltreatment may be increased by risk factors such as prematurity, mental retardation, and physical handicap.

Sexual Abuse Social, cultural, physiological, and psychological factors all contribute to the breakdown of the incest taboo. Incestuous behavior has been associated with alcohol abuse, overcrowding, increased physical proximity, and rural isolation that prevents adequate extrafamilial contacts. Some communities may be more tolerant of incestuous behavior. Major mental disorders and intellectual deficiency have been described in some perpetrators of incest and sexual abuse generally. Depending on one's perspective, incest may be considered a form of child abuse, pedophilia, or a variant of rape.

DIAGNOSIS AND CLINICAL FEATURES

Abused children manifest a variety of emotional, behavioral, and somatic reactions. These psychological symptoms are neither specific nor pathognomonic; the same symptoms may occur without any history of abuse. The psychological symptoms manifested by abused children and the behaviors of abusive parents can be organized into clinical patterns. Although it may be helpful to note whether a particular case falls into one of these patterns, that in itself is not diagnostic of child abuse.

Physically Abused Children In many cases, the physical examination and radiological evaluation shows evidence of repeated suspicious injuries. Abused children display behaviors that should arouse the suspicions of the health professional. For example, these children may be unusually fearful, docile, distrustful, and guarded. They may be wary of physical contact and show no expectation of being comforted by adults. They may be on the alert for danger and continually size up the environment; they may be afraid to go home.

The literature regarding the psychological consequences of physical abuse and neglect indicates a wide range of effects: affect dysregulation, disruptive and aggressive behaviors, insecure and atypical attachment patterns, impaired peer relationships involving either increased aggression or social withdrawal, and academic underachievement. Physically abused children exhibit a range of psychopathology including depression, conduct disorder, attention-deficit/hyperactivity disorder, oppositional defiant disorder, dissociation, and posttraumatic stress disorder.

Physically Abusive Parents Abusive parents typically delay seeking help for the injuries. The history given by the parents is implausible or incompatible with the physical findings. The parents blame a sibling or claim that the children injured themselves. The characteristics of abusive parents include a history of abuse in their own early lives; a lack of empathy for the child; unrealistic expectations of the child; and an impaired parent-child attachment, especially to babies that are defective in some way.

Sexually Abused Children A variety of symptoms, behavioral changes, and diagnoses sometimes occur in sexually abused children:

1. Anxiety symptoms such as fearfulness, phobias, insomnia, nightmares that directly portray the abuse, somatic complaints, and posttraumatic stress disorder
2. Dissociative reactions and hysterical symptoms such as periods of amnesia, daydreaming, trancelike states, hysterical seizures, and symptoms of dissociative identity disorder
3. Depression manifested by low self-esteem and suicidal and self-mutilative behaviors.
4. Disturbances in sexual behaviors, including sexual hyperarousal. Some sexual behaviors are particularly suggestive of abuse such as masturbating with an object, imitating intercourse, and inserting objects into the vagina or anus; other sexual behaviors are less specific, such as showing genitals to other children and touching the genitals of others; a younger child may manifest age-inappropriate sexual knowledge; sexually abused children may display sexually aggressive behavior toward others; in contrast to these overly sexualized behaviors, the child may avoid sexual stimuli through phobias and inhibitions
5. Somatic complaints, such as enuresis, encopresis, anal and vaginal itching, anorexia, obesity, headache, and stomachache

These symptoms are not pathognomonic. Nonabused children may exhibit any of these symptoms and behaviors. For example, normal, nonabused children commonly exhibit sexual behaviors such as masturbating, displaying their genitals, and trying to look at people who are undressing. Approximately one third of sexually abused children have no symptoms. On the other hand, the following factors have been associated with more severe symptoms in the victims of sexual abuse: greater frequency and duration of abuse, sexual abuse that involved force or penetration, and sexual abuse perpetrated by the child's father or stepfather.

Intrafamilial sexual abuse and other sexual abuse that occurs over a period of time is characterized by a particular pattern or sequence of steps. The process evolves through five phases:

1. *Engagement phase,* when the perpetrator induces the child into a special relationship
2. *Sexual interaction phase*, in which the sexual behaviors progress from less to more intimate forms of abuse
3. *Secrecy phase,* during which the perpetrator threatens the victim not to tell
4. *Disclosure phase,* when the abuse is discovered
5. *Suppression phase,* when the family pressures the child to retract his or her statements

Victims of sexual abuse recount a gradual progression of boundary violations by the perpetrator, starting with tiny invasions and escalating to serious, overwhelming intrusions. Healthy, self-confident children rebuff the intrusions either directly (via temper tantrums and verbal disagreements) or indirectly (through silence and distancing maneuvers) or by adopting any strategy that causes the offender to refrain. If a child's personal space is invaded, whether by emotional threats or physical force, the child perceives that violent or intrusive attention is synonymous with interest or affection. Many incest survivors rally around their perpetrators, seeking to capture any modicum of tenderness or interest. At times affection for the perpetrator outweighs the facts of abuse and children recant their statements about sexual assault, regardless of substantiated evidence of molestation.

Incest may be defined strictly as sexual relations between close blood relatives, that is, between a child and the father, uncle, or sibling. In its broader sense, incest includes sexual intercourse between a child and a stepparent or stepsibling. Although father-daughter incest is the most common form, incest can also involve father and son, mother and daughter, and mother and son.

The daughter in father-daughter incest has frequently had a close relationship with her father throughout her childhood and may be pleased at first when he approaches her sexually. As the behavior continues, the abused daughter becomes confused and frightened because she never knows whether her father will be parental or sexual. If the victim tells her mother about the abuse, the mother may or may not be supportive. The mother often refuses to believe her daughter's reports or refuses to confront her husband with her suspicions. Since the father provides special attention to a particular daughter, her brothers and sisters may distance themselves from her. The father, fearful that his daughter may expose their relationship, and often jealously possessive of her, interferes with the girl's development of normal peer relationships.

EVALUATION PROCESS

The evaluation of a child or adolescent who may have been physically or sexually abused depends on its circumstances and context. Practitioners must consider whether they are conducting a forensic evaluation (which has legal implications and may ultimately be used in court) or a clinical evaluation for a therapeutic purpose. A forensic evaluation emphasizes collecting accurate and complete data to determine—as objectively as possible—what happened to the child. Was the injury an accident, self-inflicted, or a result of parental abuse? Was the child actually sexually abused or was she indoctrinated to believe she was abused? The data collected in a forensic evaluation must be preserved in a reliable manner, through audiotape, videotape, or detailed notes. The results of the forensic evaluation are organized into a report that will be read by attorneys, a judge, and others. On the other hand, the emphasis in a therapeutic evaluation is to assess psychological strengths and weaknesses, make a clinical diagnosis, develop a treatment plan, and lay the foundation for continuing psychotherapy. The clinician may also be interested in determining what happened to the child, but it is not so essential to distinguish facts from fantasies. Compared with the forensic evaluation, the psychotherapist does not need to keep such detailed records and ordinarily does not prepare a report for court.

In addition to distinguishing a forensic examination from a therapy meeting, a number of factors may affect the evaluation of a child who was abused or may have been abused: whether one is a pediatrician in an emergency department or a child psychiatrist in an office; whether a parent or another person is suspected of the abuse; the severity of the abuse and the victim's relationship to the perpetrator; whether physical signs of abuse are obvious or absent; the age and sex of the child;

and the degree of anxiety, defensiveness, anger, or mental disorganization that the child exhibits. Often the examiner must be creative and persistent.

From the psychiatric perspective, the interview is usually the primary source of information and the physical examination is secondary. In practice, children who may have been neglected or sexually abused are interviewed first and later given a physical examination and other tests. A child who has been physically abused is more likely to have a physical examination that may be followed by a psychiatric interview.

Parent Interviews The evaluator obtains a history from the parents (separately, in most cases) and other pertinent informants, as well as from the child. The emphasis of the interview depends on the circumstances.

Suspected Physical Abuse The examiner should consider the possibility that the parents are not telling the truth. The physician becomes a detective because the parents who bring the injured child to the emergency room may also be perpetrators of the abuse. To obtain treatment, the caregivers lie about how the injury occurred.

When the child is brought to the emergency room a detailed and spontaneous account of the injury should be obtained promptly from parents or other caregivers before secondary details and rationalizations cloud the information provided. The interviewer should allow the caregiver to explain, expound, derail, or detour the story line. An abuser or codependent parent may claim to have happened on the injured child in a coma or bleeding from some unknown trauma or to have noticed significant bruising, burns, or a crooked extremity while bathing the child. Comparing the parents' histories can provide valuable insight into how power is wielded in the family unit.

> A 30-day-old female was transferred from a rural hospital to a university medical center because of “near SIDS” (sudden infant death syndrome). The child was unresponsive and required mechanical ventilation. A brain scan revealed bilateral subdural hematomas, subarachnoid hemorrhage, and hemorrhage in the parenchyma of the brain. An ophthalmologist observed extensive retinal hemorrhages. After the child was admitted to the Pediatric Intensive Care Unit, the child abuse consultant interviewed the parents separately. The mother, age 28, said that she had recently started a new job. The baby was perfectly fine when she left her in the care of her live-in boy friend, the child's biological father. The father, age 24, said that when he checked on the baby, he found her to be not breathing, blue, and unresponsive. He ran to report this to a neighbor and then called 911. The child abuse consultant told the father that the child must have been injured in some way and asked whether the father had any explanation for this injury. The father said, “I did shake the baby after I found her not breathing.” The consultant concluded that severe child abuse had occurred in the form of shaken baby syndrome. The consultant notified child protective services and the local police department, so that they could initiate and coordinate their investigation.

Suspected Sexual Abuse The examiner should consider the possibility that the parents are not telling the truth. However, this situation is more complex than suspected physical abuse. For example, the mother may wish to avoid the discovery of father-daughter incest by blaming the child's genital injury on another child or a stranger. In another scenario the mother may concoct an allegation of incest when the child had never been abused at all. The first version protects a father who is guilty; the second version implicates a father who is innocent.

The examiner should determine how the allegation originally arose and what subsequent statements were made. Determine the emotional tone of the first disclosure (e.g., whether the disclosure arose in the context of a high level of suspicion of abuse). Determine the sequence of previous examinations, the techniques used, and what was reported. Try to determine whether the previous interviews were likely to have distorted the child's recollections. If possible, review transcripts, audiotapes, and videotapes of earlier interviews. Seek a history of overstimulation, prior abuse, or other traumas. Consider other stressors that could account for the child's symptoms. The examiner should ask about exposure to other possible male and female perpetrators.

In Either Case Whether physical or sexual abuse is involved, a pertinent psychosocial history should be collected and organized including:

1. Symptoms and behavioral changes that sometimes occur in abused children
2. Confounding variables such as psychiatric disorder or cognitive impairment that may need to be considered
3. Family's attitude toward discipline, sex, and modesty
4. Developmental history from birth through periods of possible trauma to the present
5. Family history, such as earlier abuse of the parents, substance abuse by the parents, spouse abuse, psychiatric disorder in the parents
6. Underlying motivation and possible psychopathology of adults involved

Collateral Information The evaluator should consider requesting collateral information from the following, after obtaining authorizations: protective services, school personnel, other caregivers (e.g., babysitters), other family members (e.g., siblings), the pediatrician, and police reports.

Child Interview No hard and fast rules exist for conducting a competent evaluation of a child or adolescent victim. Experience generally supports flexibility and consistent good-hearted behavior by the interviewer. As when seeing any patient, the evaluator must size up the situation and use techniques that are likely to help the youngster become comfortable and communicative. One victim might need a favorite object (e.g., a teddy bear, a toy truck); another might need to have a particular person included in the interview. Some children are comfortable talking; others prefer to draw pictures. An unrelated joke, a shared cookie, or the picture on the evaluator's wall may lead to a disclosure of abuse. Important comments might be made while chatting during the break time, instead of during the structured interviews.

Interview Process The interviewer of the child who may have been abused should remember the following principles:

1. Audiotape or videotape the interview, if possible
2. Use a minimum number of interviews (perhaps two or three), as multiple interviews may encourage confabulation.
3. Avoid repetitive questions, either/or questions, and multiple questions and try to avoid leading and suggestive questions
4. Use restatement, that is, repeat the child's account back to the child (this allows the interviewer to see if the child is consistent and ensures that the interviewer understands the child's report)
5. Conduct the examination without the parent present (if the child is very young, consider having a family member in the room)
6. Use an examination technique that is appropriate to the child's age and developmental level
7. Determine the child's terms for body parts and sexual acts; do not educate or provide new terms

Interview Content The interview should not take the form of an interrogation. Note the child's affect while discussing these topics and be tactful in helping the child manage anxiety. Young children may not be able to report all of the relevant information. The examiner should explore the following 10 items:

1. Whethher the child was told to report or not report anything
2. Who the alleged perpetrator was
3. What the alleged perpetrator did
4. Where it happened
5. When it started and when it ended
6. Number of times the abuse occurred
7. How the child was initially engaging and how the abuse progressed over time
8. How the alleged perpetrator induced the child to maintain secrecy
9. Whether the child is aware of specific injuries or physical symptoms associated with the abuse
10. Whether any photography or videotaping took place

Step-Wise Interview The usual clinical interview may need some modification for evaluating a child who may have been abused. The step-wise interview, which is primarily intended for forensic evaluations, consists of the following components.

BUILD RAPPORT Build rapport and informally observe the child's behavior, social skills, and cognitive abilities.

ASK THE CHILD TO DESCRIBE TWO SPECIFIC PAST EVENTS Asking the child to describe two specific past events assesses the child's memory and models the

form of the interview for the child. Asking nonleading, open-ended questions is the pattern that will follow through the rest of the interview.

ESTABLISH THE NEED TO TELL THE TRUTH Reach an agreement with the child that in this interview only the truth will be discussed, not “pretend” or imagination. Explain that if the child does not know the answer to a question, it is fine to say, “I don't know.” If the child doesn't remember something, it is fine to say so.

INTRODUCE THE TOPIC OF CONCERN Start with general questions such as “Do you know why you are talking with me today?” Proceed, if necessary, to more specific questions such as “Has anything happened to you?” or “Has anyone done something to you?” Drawings may help initiate disclosure; either the child or the interviewer draws an outline of a person and the child is then asked to add and name each body part and describe its function. If sexual abuse is suspected, the interviewer could ask if the child has seen that part on another person and who has seen or touched that part on the child. If physical abuse is suspected, the interviewer could ask if particular parts have been hurt in some way.

ELICIT A FREE NARRATIVE Once the topic of abuse has been introduced, the interviewer encourages the child to describe each event from the beginning without leaving out any details. Children are allowed to proceed at their pace, without correction or interruption. If abuse has occurred over a period of time, the interviewer may ask for a description of the general pattern and then for an account of particular episodes.

POSE GENERAL QUESTIONS The interviewer may ask general questions to elicit further details. These questions should not be leading and should be phrased in such a way that the child realizes that an inability to recall or lack of knowledge is acceptable.

POSE SPECIFIC QUESTIONS If necessary, asking specific questions may yield helpful clarification. For example, the interviewer may follow up on inconsistencies in a gentle, nonthreatening manner. Avoid repetitive questions or appearing to reward particular answers in any way.

USE INTERVIEW AIDS IF NECESSARY Anatomical dolls may be useful in understanding exactly what sort of abusive activity occurred. The dolls are not used to diagnose child abuse, only to clarify what happened.

CONCLUDE THE INTERVIEW Toward the end of the interview, the interviewer may ask a few leading questions about irrelevant issues (e.g., “You came here by taxi, didn't you?”). If the child demonstrates susceptibility to the suggestions, the interviewer must verify that the information obtained earlier was not contaminated. Finally, the child is thanked for participating, regardless of the outcome of the interview. The interviewer should not make promises that cannot be kept.

PHYSICAL EXAMINATION

A forensic medical examination is usually performed by a pediatrician, preferably one with special expertise in child abuse. No part of the human anatomy is unrelated to the question of possible child maltreatment. Fingernails, hair, oral cavity, eardrums, skin surfaces, clothing, and genitals must be assessed, and the results recorded. A comprehensive examination must direct attention to all aspects of the victim's health status including maturity and developmental assessments, dental care, immunizations, and the review of systems.

Damage to the Skin and Surface Tissues Bruises, burns, bite marks, abrasions, and lacerations should be noted. Harsh punishment of the child may produce bruised lower back and buttocks, slapped cheeks, pinched earlobes, and choke marks about the neck. Injuries resulting from abusive toilet training may include pinch bruises on the penis or scrotum and perineal burns. Feeding-related injuries generate lacerations to the oral cavity and labial frenulum tears. Choke marks may be attributed to attempts at cardiopulmonary resuscitation. Sharp-bordered, rectilinear bruises may reflect strap marks; these markings can cover an entire body surface and leave buckle or eyelet impressions.

The human hand leaves various impression patterns. Grab and pinch marks are diagnosed by bruise-colored fingerprints. Negative imprints of fingers are slapping injuries, whereas oval-shaped bruises may result from grabbing during violent shaking. Human bites cause distinctive, paired, crescent-shaped bruises. The age of the biter can be determined by measuring the point-to-point distance between the center of the third tooth on each side (the canines). A distance of more than 3 cm suggests permanent teeth and hence an adult or a child older than 8 years of age.

Dating the onset and duration of bruises is problematic. As bruises age and resolve, the skin coloration passes through several hues, including red, blue, purple, green, yellow, and brown. The timing and coloration may be affected by variables such as the depth of the bruise and the amount of bleeding into the bruise.

Immersion burns generate a saved and scorched pattern, typically designated as stocking/glove or donut distribution burns. The skin pressed against a tub filled with scalding water is protected from the heat and remains unburned. Skin in direct contact with the water is scalded.

> A 16-month-old boy was brought to the hospital emergency room because of burns from scalding water. The child had patches of erythema on his abdomen, down the sides of his legs, and on the medial aspect of his upper arms. The diapered area had been protected from the heat. While the emergency room staff proceeded with their assessment and treatment, the child abuse consultant interviewed the parents separately. The father reported that he had been trying to heat a baby bottle by placing it in the bathtub and running hot water on it. The father returned in a few minutes and found the toddler, who apparently had climbed into the tub, in about two inches of hot water. The mother confirmed that she had been busy and had asked the father to heat the bottle in the bathtub. The consultant thought that the pattern of the burns was consistent with the story; the pattern did not look like that seen in forcible submersion. Although the parents' accounts were unusual, they corroborated each other and were consistent with the injuries. The consultant concluded that this was not likely to be child abuse but might constitute negligence. The consultant notified child protective services.

Damage to the Head Damage to the head includes scalp hematomas, lacerations, brain injuries, whiplash, or shaken baby syndrome. Head trauma is a worrisome injury. Even minor insults to the head prompt most caregivers to seek immediate medical attention. More than 95 percent of serious intracranial injuries sustained in the first year of life result from physical abuse. The cause of injury, typically, is violent shaking, to-and-fro whiplash, or slamming. Subdural bleeding ranks among the most dangerous inflicted injuries, often resulting in death or serious crippling sequelae. Retinal tearing and hemorrhage may be caused by shaking injuries. Falls are often blamed for injuries, and a fall from 1 to 3 feet can result in linear skull fracture and epidural hematoma. Such falls rarely result in subdural hematomas or clavicle or humerus fractures.

Inflicted black eyes are more common than serious eye injuries. Victims smacked about the eyes with an open or closed hand have both eyelids swollen, with massive bruising. Generally, black eyes sustained from accidents involve trauma to one eye. Although other scenarios are possible, the onset of bilateral black eyes immediately following facial trauma, generally indicates intentional injury.

Traumatic alopecia and subgaleal hematomas are caused by pulling the hair. Alopecia areata (noninflammatory hair loss) is characterized by loose hairs at the periphery of the bald area and is easily distinguished from inflammation or boggy swelling of the scalp caused by violent lifting of the child by the hair. Subgaleal hematomas may result when the aponeurosis connection between the occipital and frontalis muscles is wrenched off the calvarium, permitting rapid filling of the remaining space with blood.

Damage to Internal Organs Damage to internal organs includes blunt trauma to the chest; bruised abdominal organs; occult liver lacerations; duodenal hematoma; pancreatic contusion or laceration; superior mesenteric artery laceration; and spleen, kidney, and bladder rupture.

Skeletal Damage Skeletal damage is usually documented with X-rays taken immediately and, if necessary for better definition, repeated about 2 weeks after the injury. The most distinctive radiological features of infant abuse are metaphyseal lesions, called *corner fractures* or *bucket-handle fractures,* depending on the perspective of the X-ray. In addition to metaphyseal lesions, the following are highly specific for child abuse because they are unlikely to occur accidentally: posterior rib fractures, scapular fractures, spinous process fractures, and sternal fractures. The following are considered moderately specific for child abuse: multiple fractures, especially when they are bilateral; fractures of different ages; epiphyseal separations; vertebral body fractures and subluxations; fractures of the fingers; and complex skull fractures. Radiologists are astute about locating evidence of trauma and can provide insight into causality and patterns of old, healing, or new fractures. A child may not report a fracture if more injuries are likely to result after disclosure.

Intoxication and Poisoning Psychotropic and analgesic medication, alcohol, and illicit drugs have all been used by adult abusers to render a child or adolescent more susceptible to assault and less capable of offering accurate witness. There have also been reports of administration of salt, caustic cleaners, laxatives, animal

tranquilizers, blood, or excrement.

Physical Findings of Sexual Abuse Labial and perineal tissues that are susceptible to friction-induced traumas manifest abrasions, contusions, lacerations, petechiae, and edema. Examination may reveal bloodied tissue, genitorectal tears, and infected secretions. Scarring may be minimal or not evident if vaginal penetration has been gentle, consensual, or gradual. Vaginal and rectal tissues heal rapidly after injury and abrasions can be repaired within several days. Thus the lack of physical findings may indicate either a return to normalcy after rapid healing of minor lesions or trauma perpetrated in a manner that left no material evidence.

Major insults to the penile head (tattoos, scaldings, amputation) produce noticeable scars. The more common traumas (e.g., hematomas, abrasions, and lacerations) are superficial injuries and heal rapidly without scarring.

Anal findings include dilation with loss of anal-rectal tone, flattened rectal rugae, lacerations, and wedge-shaped and linear scars. Acute trauma may result in swelling and spasm secondary to submucosal hemorrhage. Most rectal injuries, massive or minor, heal quickly and nearly completely, leaving little evidence of abuse. Minor but persistent irritants, such as pinworms, lichen sclerosis, atopic dermatitis, and rare hematological disorders, can be longer-standing rectal irritants and may confound examination outcomes.

When sexual abuse is suspected, the victim requires evaluation for sexually transmitted diseases. Children may silently carry *Chlamydia trachomatis* infections, gonorrhea, syphilis, the human immunodeficiency virus (HIV), human herpesvirus 1 (HSV-1) or HSV-2, or other infections. The human papillomavirus, which causes condylomata acuminata, is known to be transmitted by direct sexual contact, but infants and young children may acquire anal and genital warts perinatally.

SPECIAL TESTS AND LABORATORY EXAMINATION

Computed tomography scans are invaluable in assessing an unconscious child with head trauma or abdominal injury. Rarely does a child sustain both cranial and abdominal traumas from a single innocent fall. Significant abdominal blunt trauma should be suspected in an atypically still child, because intraabdominal injuries are the second most common cause of death in battered children. A punch or direct kick can compress the organs against the spinal column, rupturing the liver or spleen and rapidly inducing shock from sudden, massive blood loss.

DIFFERENTIAL DIAGNOSIS

Physical Findings Pediatricians who examine children who may have been abused must consider a wide range of congenital, infectious, toxic, and traumatic conditions in their differential diagnoses (Fig. 49.4-1 and Fig. 49.4-2). The following examples are not meant to be exhaustive:

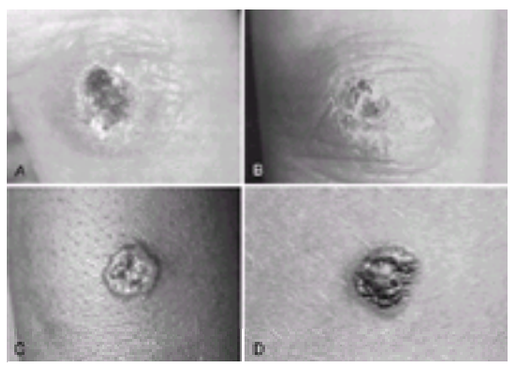

FIGURE 49.4-1 Child maltreatment is frequently manifested by bruises, burns, and other skin lesions. The diagnostic problem is that abusive injuries may be mistaken for medical conditions and vice versa. **A**, Physical abuse, a cigarette burn on the sole of the foot of a 5-month-old infant. **B**, A swimming pool granuloma overlying an interphalangeal joint. **C**, Tinea corporis. **D**, Clark's nevus, a benign, common acquired nevus. [Reprinted with permission from Reece RM: *Child Abuse: Medical Diagnosis and Management*. Lea & Febiger, Philadelphia, 1994 **(A);** Caputo R: *Pediatric Dermatology and Dermatopathology,* vol 4. Williams & Wilkins, Baltimore, 1996 **(B);** Caputo R: *Pediatric Dermatology and Dermatopathology,* vol 2. Lea & Febiger, Philadelphia, 1993 **(C);** Caputo R: *Pediatric Dermatology and Dermatopathology,* Lea & Febiger, Philadelphia, 1990 **(D)**.] (See Color Plate 9.)

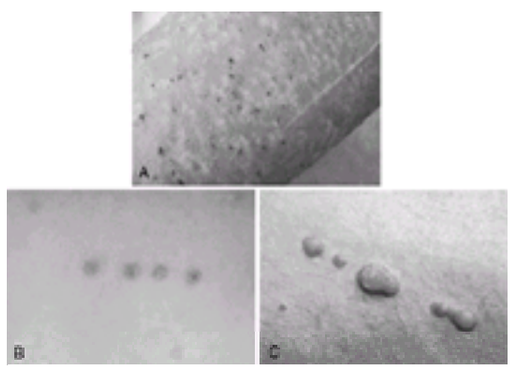

FIGURE 49.4-2 These linear skin lesions illustrate both child maltreatment and other conditions not related to abuse. **A**, Physical abuse, superficial ulcers caused by applying sandpaper to a child's skin. **B**, A series of insect bites distributed in a linear fashion, indicating that a single insect took bites in sequence. **C**, Common warts that are linear because lesions are induced when old lesions are scratched. [Reprinted with permission from Caputo R: *Pediatric Dermatology and Dermatopathology*, vol 2. Lea & Febiger, Philadelphia, 1993 **(A);** Caputo R: *Pediatric Dermatology and Dermatopathology,* vol 4. Williams & Wilkins, Baltimore, 1996 (**B** and **C**).] (See Color Plate 9.)

Conditions manifested by skin lesions: A Mongolian spot, a birthmark, may look like a bruise. Lichen sclerosis may look like sexual abuse. Chicken pox and impetigo may look like cigarette burns.

Conditions manifested by bone abnormalities: Congenital syphilis, cardiopulmonary resuscitation, copper deficiency, and rickets must be considered.

Conditions manifested by bruises: Vitamin K deficiency from liver disease, salicylate toxicity, osteogenesis imperfecta, bleeding disorders, and erythema multiforme must be considered.

Conditions manifested by eye hemorrhages: Valsalva effect and increased intracranial pressure and hypertension must be considered.

Self-inflicted injuries: Scratches from itchy irritations, factitious illness, head-bangers, and Lesch-Nyhan syndrome must be considered.

Psychiatric Considerations The psychiatric evaluation of youngsters who may have been abused may involve assessing the patient's credibility. Although generally children tell the truth when they talk about abuse, sometimes children make false denials (saying they were not abused, when actually they were) or false allegations (saying they were abused, when actually they were not).

Possible Explanations of Denials of Abuse A false denial or retraction may occur for several reasons. The child may have been pressured by the perpetrator or by

family members to recant the allegation. The pressure may consist of bribery, mockery, or threats of injury. The child may be protecting a parent or other family member, even without external coercion (i.e., the child takes on this responsibility through role reversal). A child might be frightened or distressed by the investigation and decide to withdraw participation. For instance, an interviewer could induce a false denial by asking overly challenging questions. A child may be inhibited by shame or guilt; some children mistakenly assume that they were responsible for what happened, and some children accommodate to the abuse, consciously or unconsciously, instead of objecting to it.

Possible Explanations of Allegations of Abuse A false allegation of abuse may occur for several reasons.

ALLEGATIONS ORIGINATING WITH ADULTS Sometimes a false allegation arises in the mind of a parent or other adults and is imposed on the child. The parent may have misinterpreted an innocent remark, a neutral piece of behavior, or a benign physical condition as evidence of abuse and induced the child to endorse this interpretation. This happens in child custody disputes as well as other settings. Sometimes the parent and child share a *folie à deux* or the child may simply give in and agree with a delusional parent. A parent may have fabricated the story and induced the child to collude in presenting it to the authorities.

A false allegation may result from an interviewer's suggestion. Previous interviewers may have asked leading or suggestive questions. An interviewer who believes abuse occurred may unwittingly shape a child's responses until the child validates the interviewer's assumptions.

Group contagion may lead to false allegations. In epidemic hysteria people modify what they have heard to meet their own emotional needs. Thus, rumors may become more convincing as they are retold.

UNCONSCIOUS OR NONPURPOSEFUL MENTAL MECHANISMS IN THE CHILD A younger child may confuse fantasy with reality, and although rare, older children and adolescents may experience delusions about sexual activities in the context of a psychotic illness. Sometimes children misunderstand what happened and later report it inaccurately or they may misunderstand an adult's question and the adult may later misinterpret or take the child's statement out of context. In confabulation, children fill gaps in their memory with whatever information makes sense to them and others at the time.

CONSCIOUS OR PURPOSEFUL MENTAL MECHANISMS IN THE CHILD

Fantasy Lying Children who understand the significance of lying may nonetheless fabricate because of frustration, disappointment, or wishful thinking.

Innocent Lying Children may make false statements because that seems to be the best way to handle the situation they are in. Developmentally, this is more likely to happen with younger children.

Deliberate Lying Children may choose to avoid or distort the truth for some personal advantage. This happens more with older children and adolescents.

PERPETRATOR SUBSTITUTION The child may have actually been sexually abused and exhibits symptoms consistent with abuse but identifies the wrong person as the perpetrator, resulting in a false allegation. The child may do this to protect the actual offender or the child may displace the memories and accompanying affects onto another individual.

Possible Explanations of Allegations of Ritual Abuse Other phenomena may be mistaken for cult-based ritual abuse, including pseudoritualistic abuse, psychopathological repetitive abuse, sexual abuse by pedophilic sex rings, false memories of abuse, and hoaxes.

The parents of a 9-year-old boy were divorced, and he spent the summers visiting his father. When he returned from visitation the mother noticed injuries: a healing cut on the boy's right ear and a collection of superficial, vertical scratches from his umbilicus to his pubis. She had the boy disrobe and took Polaroid photographs of his abdomen. Several days later the mother took the boy and the photographs to the boy's regular psychotherapy appointment. She told the therapist that she was concerned that the boy had been physically and sexually abused. The therapist interviewed the boy individually, who explained in a matter-of-fact manner that his ear was cut accidentally when he had his hair cut and his abdomen was scratched when he went bodysurfing at a North Carolina beach. The therapist concluded the boy had not been abused and reassured the mother. The therapist did not notify child protective services.

SPECIAL ISSUES

Fetal Abuse Fetal abuse can be direct or indirect. Either a mother or a father may directly attack their fetus through the abdominal wall or via the vagina. The risk is greatest during late pregnancy when one of the parents becomes disenchanted with the prospect of a new baby. Indirect fetal abuse by failing to protect the fetus from alcohol, nicotine, or drugs is more common. Fetal alcohol syndrome is an important cause of mental retardation in the United States.

Factitious Disorder by Proxy In factitious disorder by proxy, also called Munchausen syndrome by proxy, a parent, usually a mother, provides a history that fascinates and beguiles family physicians and expert consultants. The mother gives enough data to suggest a serious but poorly understood condition that requires extensive, painful testing. The child is often drawn into the improbable tale out of fear of contradicting the history given by the mother. Professionals become enmeshed in the apparent validity of the history, but no one can verify or witness the extraordinary symptoms that support the pathology of the child's alleged illness. Parents involved in the deception often have narcissistic, histrionic, or borderline personality disorder and thrive on the center-stage position they assume. In families where factitious disorder by proxy is found, more than one child may be involved consecutively, though rarely simultaneously. In extreme cases children can be killed, seriously injured, terrorized, and suffer severe adverse effects from medications or other treatments.

Cultural Variations Parenting practices in American society vary widely. They are partly determined by the cultural heritage of the family, and the cultural context of the alleged act should be considered in evaluating suspected neglect and abuse, including assessing whether the discipline of a child is abusive. Some unusual bruises result from cultural healing practices. Coining or cupping—techniques used in many European and Asian countries—are used to draw toxins from the skin surface and remove feverish mucus from the system. In coining, a coin is heated in hot oil and massaged against the skin. In cupping, a heated glass cup is used to create suction against the skin. Skin scraping and moxibustion may also mimic physical abuse.

Female Genital Mutilation Commonly called “female circumcision,” female genital mutilation is performed on an estimated two million girls worldwide each year. It is practiced across diverse socioeconomic classes and different ethnic, cultural, and religious groups. Commonly, girls are circumcised between 4 and 10 years of age, but the procedure may be performed on infants, postponed until just before marriage, or done after the birth of the first child. In some cultures it is considered part of a ceremonial induction into adult society. Female circumcision, in the mildest form, consists of clitoridectomy, the anatomical equivalent of penile amputation. In its most severe form, total infibulation involves removal of the clitoris and labia minora plus incision of the labia major to create raw surfaces that are then stitched together. This practice has been widely criticized, including opposition by the World Health Organization and other major health care groups. In the United States female circumcision is generally considered child abuse. Efforts have been made to educate immigrant communities on the health risks and legal liabilities of the practice. Compromise may be possible by finding a way to satisfy the cultural requirements without using mutilation such as adopting a nonmutilative ritual incision that results in only small scars on the labia.

Male genital mutilation (circumcision) is not usually considered child abuse, but it does have its critics. Although this simple procedure is one of the most common operations performed worldwide, serious complications may result. These complications include hemorrhage, infection, and penile amputation. In 1999 The American Academy of Pediatrics recommended it not be done as a routine procedure on newborns.

Religious Practices Certain religious practices may be construed as forms of abuse if they are likely to harm children. For example, adherents of Christian Science believe that illness should be treated with prayer instead of medical care. As a result, children may sustain chronic illness, disability, or death. This is an area of flux; some state laws and some courts have allowed parental discretion to withhold conventional medical treatment for legitimate religious reasons, while others have not.

Corporal Punishment Within families, corporal punishment occurs often enough in the United States that it cannot easily be labeled deviant. Most child and adolescent psychiatrists discourage physical discipline but realize that many capable and apparently nurturing parents spank their children. If physical discipline is used as a last resort, parental anger and frustration may exceed acceptable limits. In its most extreme form, a spanking turns into a beating. After 10 to 20 minutes of such abuse, a child will bleed into the muscles and other tissues of the back, buttocks, and legs and may lose enough blood to die from hypovolemic shock.

While sincere professionals may disagree about the propriety of parental spanking of children, there is a consensus against the paddling of children by teachers. Professional organizations (e.g., the American Medical Association, the American Academy of Pediatrics, and the American Academy of Child and Adolescent Psychiatry) have gone on record favoring banning of corporal punishment as a disciplinary measure in schools. Although almost every developed country has banned corporal punishment in schools, it is still allowed in more than 20 states, especially in the South.

Team Work It is draining to establish rapport, delve into hidden agendas, uncover covertly presented data, address multiple human needs, and meet the legal obligations associated with child abuse cases. Clinicians are thus encouraged to participate in a multidisciplinary team. Team participation provides the expert consultation and professional support essential for continued health, hope, and positive social interactions within the grim realities of case discovery.

In cases of suspected child abuse and neglect, the physician should make an assessment based on a history, a physical examination, a skeletal survey (when indicated), and photographs; request appropriate surgical, medical, and psychiatric consultations; diagnose the suspected maltreatment; and perhaps secure the child's safety by admitting the child to a hospital. Usually other members of the multidisciplinary team report the case to the appropriate child protection agency, provide skilled nursing, maintain security, arrange out-of-home placement, arrange a program of care for the child and the parents, and arrange for social service follow-up. These activities, taken together, should be a coordinated team effort. One feature of the team activity is regular meetings of the child abuse committee.

REPORTING LAWS AND PRACTICES

In every state, the District of Columbia, and all areas controlled by the federal government (e.g., military bases), physicians are required to report child abuse to the police or the appropriate child protection agency. The typical statute says that a physician who has “reason to believe” or “reason to suspect” child abuse is required to report it. In most states, other designated individuals, such as nurses, psychologists, school officials, police officers, district attorneys, and providers of child day care and foster care, are also required to report child abuse.

Underreporting A significant factor in reporting errors is the failure of clinicians and other mandated reporters to consider child maltreatment in the differential diagnosis of child injuries. Frequently, mandated reporters fail to pursue the diagnosis of child maltreatment because of diverse cultural and socioeconomic factors. Lack of experience in dealing with the issues and the judicial quagmire that results when reporting many child maltreatment cases is also an important deterrent. Underreporting of abuse also occurs when well-meaning but misguided therapists attempt to monitor and treat their abusive clients without involving the proper authorities. Some therapists think they are above the law; others simply have no idea what the law says about reporting child abuse. Perhaps the most important barrier to seeing and reporting abuse is denial. Reporters, both professionals and lay persons, are repulsed by the reality of child maltreatment and cannot accept the thought that abuse exists or is perpetrated by individuals like themselves, who may or may not regret their painful actions. The most obvious negative consequence of underreporting is that the abused child and siblings are left in dangerous circumstances, vulnerable to further abuse.

Overreporting Although physicians are required to report abuse when they believe or suspect (depending on the jurisdiction) that it occurred, they are not required to report other people's beliefs or suspicions. Some clinicians have the mistaken idea that if a mother expresses a concern that her child may have been abused, the suspected abuse must be reported, even when the clinician has investigated and has no suspicion at all that abuse occurred (e.g., an overly anxious mother who imagines that her child was sexually abused at day care, even though an objective examiner found no indications that abuse had occurred). The negative consequences of overreporting include increasing the load on an already overburdened child protective system and an unnecessary invasion into the lives of families, who are subjected to investigation, suspicion, and criticism.

TREATMENT

Child The first part of the treatment of child abuse and neglect is ensuring the child's safety and well-being. The child may need to be removed from an abusive or neglectful family to ensure protection, but emotionally feel even more vulnerable in an unfamiliar setting. Because of the high risk for psychiatric symptoms in abused and neglected children, a psychiatric evaluation is in order. Along with providing specific treatments for any mental disorders present, the therapist may have to deal with the immediate situation and the long-term implications of the abuse or neglect. Psychotherapeutic issues to be addressed with an abused child include dealing with the child's fears, anxieties, and self-esteem; building a trusting relationship with the therapist, in which the child will not be exploited or betrayed; and gaining a perspective over time of the factors that contributed to the child's victimization at home.

It is usually helpful to have the child share the painful, ugly details of the abusive experiences with the therapist. This may take time and may require therapeutic creativity and flexibility. Although some children may be able to relate the bad things that happened through ordinary conversation, other children may need a variety of play, artistic, or other projective techniques. Some children avoid the painful memories through conscious (suppression) and unconscious (repression) mechanisms. For a child who has been minimally assaulted (e.g., a relatively innocuous single episode of fondling by a stranger), it may be adaptive for the child to forget the incident and move on.

Parents Developing an intervention plan for an abused or neglected child requires assessing the parents' behavior and functioning in several respects. The psychopathology and the prognosis for achieving adequate parenting skills of both the abusive parent and the one who allowed the abuse to occur, should be assessed. The evaluator should determine whether the parent's dysfunction is confined to this child or involves other children and whether the dysfunction is short-term or long-term (reflecting a lifelong pattern). The evaluator should assess both the parent's willingness to participate in the intervention plan and the availability of personnel and physical resources to implement the various intervention strategies. One must consider the risk of additional physical or sexual abuse if the child remains in the home.

On the basis of the information obtained, several options can be selected to improve the parent's functioning: (1) eliminating or diminishing the social or environmental stresses; (2) lessening the adverse psychological effects of the social factors on the parents; (3) reducing the demands on the mother to a level that is within her capacity through day-care placement of the child or provision of a housekeeper or baby-sitter; (4) providing emotional support, encouragement, sympathy, stimulation, instruction in maternal care, and training in planning for, assessing, and meeting the needs of the child (supportive casework); and (5) resolving or diminishing the parent's inner psychic conflicts (psychotherapy).

Incestuous Behavior The first step in the treatment of incestuous behavior is its disclosure. Once the denial and collusion by the family members has been exposed, incest is not likely to recur. When the participants suffer from severe psychopathology, treatment must be directed toward the underlying illness. Family therapy is useful to reestablish the group as a functioning unit and to develop healthier role definitions for each member. While the participants are learning to develop internal restraints and appropriate ways to gratify their needs, the external control provided by therapy helps prevent further incestuous behavior. At times, legal agencies are involved to help enforce external controls.

PREVENTION

Primary Prevention Measures that reduce the occurrence of child abuse, primary prevention, can be classified on the basis of their audience or intended recipient. Universal interventions targeted to the general public include such strategies as a poster campaign admonishing parents not to shake their babies; warnings about consuming alcohol and drugs during pregnancy; and raising public awareness through the media, civic associations, church groups, and parent-teacher organizations. To some extent, children can be taught to avoid or resist being abused. However, ultimately the adults in a community must create safety and security for their children. Selective interventions are targeted to populations at high risk for child abuse. In general, for example, there should be less child abuse in a community with fewer single parents (use of birth control; greater involvement of fathers in families) and greater financial security (local jobs; social support systems). Governmental regulations may be helpful, such as the licensing of day-care centers and regulations concerning the operation of residential programs for children. Finally, one can sometimes intervene with specific persons or families that are at high risk. Counseling and parenting classes may deter parents from being abusive, although this has not been substantiated. Treatment for parents who abuse substances and services for families with a history of spouse abuse may be helpful. Individuals can help friends and relatives who appear stressed and are struggling with their parenting responsibilities.

Secondary Prevention Early detection and prompt intervention when child abuse occurs is secondary prevention. Education of the medical profession, members of allied health fields, and all who come in contact with children will aid early detection. Emergency room personnel, pediatricians, teachers, and counselors have many opportunities to identify abused children. It is hard to know how best to use the limited resources available to child protection agencies. Some critics think that too much effort is put into assessing the millions of referred cases, with relatively little attention given to interventions. For instance, an agency might put many hours into collecting information, arrive at a decision about whether a case is founded or not, engage in expensive legal measures, but offer little or no aid to help a marginal

family function better. In general, child abuse and neglect prevention and treatment programs should try to (1) prevent the separation of parents and children if possible; (2) prevent the placement of children in institutions; (3) help the parents attain self-care status; and (4) help the family become self-sufficient. As a last resort, to prevent further abuse and neglect, children may have to be removed from families who are unwilling or unable to profit from the treatment program.

SUGGESTED CROSS-REFERENCES

Forensic psychiatry is presented in Section 51.6b. Child forensic psychiatry is presented in Section 49.12. A more detailed discussion of the diagnostic process with children and adolescents is in Chapter 44. The various disorders manifested by maltreated children are discussed in the following chapters and sections: posttraumatic stress disorder, Section 46.2; identity problems and borderline disorders, Section 49.10; separation anxiety disorder and anxiety in children, Section 46.3; eating disorders, Chapter 20; feeding and eating disorders of infancy or early childhood, Chapter 41; elimination disorders, Chapter 43; mood disorders and suicide, Chapter 45; and child or adolescent antisocial behavior, Section 49.7.

SECTION REFERENCES

American Academy of Child and Adolescent Psychiatry: Practice parameters for the assessment and treatment of children and adolescents with posttraumatic stress disorder. J Am Acad Child Adolesc Psychiatry *37:*4S, 1998.

American Academy of Child and Adolescent Psychiatry: Practice parameters for the forensic evaluation of children and adolescents who may have been physically or sexually abused. J Am Acad Child Adolesc Psychiatry *36:*423, 1997.

American Academy of Pediatrics Committee on Child Abuse and Neglect: Guidelines for the evaluation of sexual abuse of children. Pediatrics *87:*254, 1991.

American Professional Society on the Abuse of Children: Guidelines for psychosocial evaluation of suspected sexual abuse in children. American Professional Society on the Abuse of Children, Chicago, 1997.

Beitchman JH, Zucker KJ, Hood JE, DaCosta GA, Akman D: A review of the short-term effects of child sexual abuse. Child Abuse Negl *15:*537, 1991.

Bernet W: False statements and the differential diagnosis of abuse allegations. J Am Acad Child Adolesc Psychiatry *32:*903, 1993.

Bernet W, Chang DK: The differential diagnosis of ritual abuse allegations. J Forensic Sci *42:*32, 1997.

Briere JN: *Child Abuse Trauma: Theory and Treatment of the Lasting Effects*. Sage, Newbury Park, CA, 1992.

Britton HL: Perinatal screening for child abuse and neglect. Clin Perinatol *25:*453, 1998.

Caffey J: Multiple fractures in the long bones of infants suffering from chronic subdural hematoma. Am J Roentgenol *56:*163, 1946.

*Ceci S, Bruck M: *Jeopardy in the Courtroom: A Scientific Analysis of Children's Testimony*. American Psychological Association, Washington, DC, 1995.

Coulborn-Faller K, Corwin DL: Children's interview statements and behaviors: Role in identifying sexually abused children. Child Abuse Negl *19:*71, 1995.

Ewing-Cobbs L, Kramer L, Prasad M, Canales DN, Louis PT, Fletcher JM, Vollero H, Landry SH, Cheung K: Neuroimaging, physical, and developmental findings after inflicted and noninflicted traumatic brain injury in young children. Pediatrics *102:*300, 1998.

*Friedrich WN: *Psychotherapy of Sexually Abused Children and Their Families.* Norton, New York, 1990.

Friedrich WN, Grambsch P, Damon L, Hewitt S: Child sexual behavior inventory: Normative and clinical comparisons. Psychol Assess *4:*303, 1992.

Goodman GS, Bottoms BL: *Child Victims, Child Witnesses: Understanding and Improving Testimony*. Guilford, New York, 1993.

Group for the Advancement of Psychiatry: *Psychological Disorders in Childhood: Theoretical Considerations and a Proposed Classification.* Aronson, New York, 1974.

*Heger A, Emans SJ: *Evaluation of the Sexually Abused Child: A Medical Textbook and Photographic Atlas*. Oxford University Press, New York, 1992.

*Helfer ME, Kempe RS, Krugman RD, editors: *The Battered Child,* ed 5. University of Chicago Press, Chicago, 1997.

Kaplan SJ, Pelcovitz D: Child abuse. Child Adolesc Clin North Am *3:*1, 1994.

Kempe CH, Silverman FN, Steele BF, Droegemueller W, Silver HK: The battered-child syndrome. JAMA *181:*17, 1962.

Kini N, Lazoritz S: Evaluation for possible physical or sexual abuse. Pediatr Clin North Am *45:*205, 1998.

Look KM, Look RM: Skin scraping, cupping, and moxibustion that may mimic physical abuse. J Forensic Sci *42:*103, 1997.

Ludwig S, Kornberg AE: *Child Abuse: A Medical Reference*. Churchill Livingstone, New York, 1992.

Melton GB, Barry FD, editors: *Protecting Children from Abuse and Neglect*. Guilford, New York, 1994.

Melton GB, Petrila J, Poythress NG, Slobogin C: *Psychological Evaluations for the Courts,* ed 2. Guilford, New York, 1998.

Monteleone JA, Brodeur AE: *Child Maltreatment: A Clinical Guide and Reference & a Comprehensive Photographic Reference Identifying Potential Child Abuse,* ed 2. Mosby-Year Book, St. Louis, 1997.

Myers JEB: *Evidence in Child Abuse and Neglect Cases,* ed 3. Wiley Law Publications, New York, 1997.

National Committee to Prevent Child Abuse: *Current Trends in Child Abuse Reporting and Fatalities: The Results of the 1997 Annual Fifty State Survey*. National Committe to Prevent Child Abuse, Chicago, 1998.

Ney T, editor: *True and False Allegations of Child Sexual Abuse: Assessment and Case Management*. Brunner/Mazel, New York, 1995.

Nimkin K, Kleinman PK: Imaging of child abuse. Pediatr Clin North Am *44:*615, 1997.

*Nurcombe B, Partlett DF: *Child Mental Health and the Law*. Free Press, New York, 1994.

Quinn KM, White S, Santilli G: Influences of an interviewer's behavior in child sexual abuse investigations. Bull Am Acad Psychiatry Law *17:*45, 1989.

Raskin DC, Esplin PW: Statement validity assessment: Interview procedures and content analysis of children's statements of sexual abuse. Behav Assess *13:*265, 1991.

Reece RM: *Child Abuse: Medical Diagnosis and Management*. Lea & Febiger, Philadelphia, 1994.

*Roosa MW, Reinholtz C, Angelini PJ: The relation of child sexual abuse and depression in young women: Comparisons across four ethnic groups. J Abnorm Child Psychol *27:*65, 1999.

Schetky DH, Benedek EP, editors: *Clinical Handbook of Child Psychiatry and the Law*. Williams & Wilkins, Baltimore, 1992.

Schetky DH, Green AM: *Child Sexual Abuse: A Handbook for Health Care and Legal Professionals*. Brunner/Mazel, New York, 1988.

*Smetana JG, Toth SL, Cicchetti D, Bruce J, Kane P, Daddis C: Maltreated and nonmaltreated preschoolers' conceptions of hypothetical and actual moral transgressions. Dev Psychol *35:*269, 1999.

*Smith M, Walden T: Understanding feelings and coping with emotional situations: A comparison of maltreated and nonmaltreated preschoolers. Soc Dev *8:*93, 1999.

US Department of Health and Human Services, National Center on Child Abuse and Neglect: *Child Maltreatment 1995: Reports from the States to the National Center on Child Abuse and Neglect*. US Government Printing Office, Washington, DC, 1997.

US Department of Health and Human Services, National Center on Child Abuse and Neglect: *Child Abuse and Neglect: State Statute Series*. US Government Printing Office, Washington, DC, 1997.

Van Haeringen AR, Dadds M, Armstrong KL: The child abuse lottery—will the doctor suspect and report? Physician attitudes towards and reporting of suspected child abuse and neglect. Child Abuse Negl *22:*159, 1998.

Whipple EE, Richey CA: Crossing the line from physical discipline to child abuse: How much is too much? Child Abuse Negl *21:*431, 1997.

49.5 CHILDREN'S REACTION TO ILLNESS AND HOSPITALIZATION

MARTIN J. DRELL, M.D., AND TONYA JO HANSON WHITE, M.D.

Illness and hospitalization are all too frequent occurrences in the lives of children and their families. Children 1 to 3 years of age average seven to nine illnesses per year, those 4 to 10 years of age average four to six per year, and those over 10 years of age average four per year. Together, a family with two adults and two children averages 21 illnesses a year. Even though most of these illnesses are minor and cause only transient disruptions of everyday life, it is estimated that 2 to 3 percent of all children have illnesses severe enough to affect growth, development, school performance, or social relationships. With more than four million child and adolescent hospitalizations each year, health care professionals must understand the impact of illness and hospitalization on children to offer appropriate interventions.

The developmental biopsychosocial model, based on work by George Engel, is the best model for understanding the impact of illness on children. This model is used in this section to identify the key developmental, biological, psychological, and social factors that come into play when children become sick (Fig. 49.5-1).

FIGURE 49.5-1 Developmental biopsychosocial model.

DEVELOPMENTAL BIOPSYCHOSOCIAL MODEL

The developmental biopsychosocial model is influenced by general systems theory. It stresses that all medical phenomena, such as a child's reaction to illness, hospitalization and surgery, can best be conceptualized through a thorough understanding of developmental, biological, psychological, and social information concerning all the key persons involved. The model is based on the supposition that each of these sets of information is worthy of study in its own right and no one, individually, can explain the complexity of health and disease. The model proposes that a true understanding can only be achieved by considering the ongoing interplay of these four sets of information. A number of main questions need to be answered in each of these four categories to get a clear picture of a case.

Developmental Sphere The main question to ask in the developmental category is: “What is the current, premorbid developmental stage and functional level of the child?” Although several responses to illness and hospitalization are typical of persons of all ages such as anxiety, regression, mood changes, and changes in bodily functions (e.g., eating, sleeping), some responses are characteristic of an individual developmental level. Different-age children, depending on their developmental level, have different key concerns or tasks that influence how they perceive all events. Thus, their understanding of illness and hospitalization also differs, depending on their developmental level (Table 49.5-1). The child's developmental level can be ascertained through the interview, history taking, and careful observation.

Age (years)	Developmental Tasks	Cognitive View of Illness	Response to Illness and Hospitalization
0 to 3	Establishing primary attachments and bond Learning to regulate attention and bodily functions Increasing opportunities for exploration and learning, yet dependent on their caregivers to protect them from pain and discomfort Limited verbal skills, affective communication predominates	Illness viewed as an external agent inflicting discomfort and pain Almost totally dependent on caregivers for relief of illness	Infants may become difficult to engage and show an intense preoccupation with their environment Sleep and feeding difficulties Procedures often elicit affective interplay—"Why are you taking them (to this or me, Mommy?" answered by "I feel your pain only too well and I'm here to comfort you"
3 to 7	Mastering separation/individuation from parents—fears of abandonment Beginning to use symbols and language Concerns with abandonment and bodily harm	Egocentricity predominates—viewing the illness, hospitalization, or procedures as a result of bad behavior Understands illness as something going on inside Illness defined only via external signs	Focus on an intact body—unrealistic fears visualized with a fantasied aggressive repertoire may escalate anxiety and confusion Can tolerate longer periods of separation Continued need for primary attachment figure
7 to 13	Mastering fears of loss of control and death Need to organize and order experiences Rules and structure have high priority	Increased understanding of illness Can accept necessary procedures on the need for hospitalization Illness is viewed as a result of wrong doing Contagion theories	Illness defined by concrete symptoms More able to understand illness from multiple points of view but unable to entertain hypothetical situations Illness improves through proper care with limited understanding of the body's role in healing itself
13 to 18	Mastering second phase of separation/individuation from parents Dependent on and heavily influenced by peers	Gradual development of an adult concept of illness Feelings of being invulnerable to illness Awareness of the complex and multiple interrelationships of factors such as a "host" and "agent" in causing illness	Attribution of self-blame continues, but decreases with age May exhibit difficulties meeting a chronic illness role that their sense of self-image/interferes to psychosexual development

Table 49.5-1 Developmental Responses to Illness and Hospitalization

Children age 3 years and under are establishing primary attachments. They are learning to regulate attention and other body functions. While they have access to opportunities for exploration and learning, their caregivers protect them from pain and overwhelming stimuli. Children in this age range have little concept of illness. They view it as an external agent that causes discomfort and pain. They depend almost totally on their external caregivers to relieve them of what bothers them.

Children 3 to 7 years old are typically concerned with abandonment, punishment, and bodily harm. They have begun separating and individuating from their caregivers on whom they still remain largely dependent. These children are better able to understand illness as something going on inside themselves, yet often they have unique ideas to explain why they are sick, including the concrete interpretation that the illness is a punishment for real or perceived indiscretions by the child.

Children between the ages of 7 and 13 years are working to master fears of loss of control and fears of death. These children have a greater understanding of illness and themselves, but they still believe that they have done something wrong or that they acquired the illness through contagion. Children at this age can be told more about their illnesses and can be more actively involved in their treatments.

Teenagers between 13 and 18 years are primarily concerned with their physical appearance, self-image, real or potential loss of peers, and fears of dependency. Adolescents understand their illnesses in ways approaching those of mature adults. They can test hypotheses logically and entertain numerous causes for their condition. They depend less on their caregivers, are more able to care for themselves, and generally prefer the company of their peers. Under stress, however, they still tend to consider illness to be a punishment for some indiscretion only known to them.

Dividing these developmental tasks by age is an explanatory artifice. In reality, the changes in both developmental tasks and cognitive abilities vary among children of similar ages and can advance unevenly. Children at all ages do the best they can with the resources they have available. As the developmental biopsychosocial

model predicts, developmental advances depend partially on the interactions among the child, caregivers, and events. Thus, two children close in age with similar caregivers may have entirely different perceptions of illness if one has been chronically ill and the other has previously encountered only minor, transient illnesses. Direct threats to physical development (e.g., illness resulting in growth delays, impaired gross or fine motor functioning, or early or delayed onset of puberty) can influence overall development. Furthermore, illness-specific factors such as disfigurement, isolation, and school absenteeism may disrupt social or identity development. Health professionals must understand developmental concepts enough to ask how the illness is affecting the child and be sophisticated enough to listen and respond to concerns or inexact theories that reflect developmentally based misperceptions.

A polite teenager was consumed with guilt whenever he got the least bit angry with his parents. When he was diagnosed with osteosarcoma, he wondered if this wasn't a punishment for his anger. He told the child psychiatrist during a consultation that he knew his anger was not really the cause, but that he could not get the thought out of his mind.

Knowledge of the impact of separation, especially on young children under stress, has prompted health care providers to change visiting policies at hospitals during the last four decades. Parents are usually allowed 24-hour visitation and at times are even provided with a cot.

Biological Sphere The main questions in the biological category are aimed at determining the biological and organic correlates of the suspected problem. This entails investigating the constitutional strengths and weaknesses of each patient, including their genetic endowment, response to medication, biological manifestations of the illness, perceptual skills, learning capabilities, ability to attach and maintain relationships, temperament, and predisposition to mental illness.

Some more-specific biological questions include: How does a child's intelligence affect the ability to comprehend the illness and comply with treatment? How does a child's temperament affect compliance with treatment? Infant psychiatrists have now described regulatory disorders that are characterized by young children having difficulties regulating such aspects of their functioning as behavior, physiology, affect, attention, and the senses. These regulatory disorders include subtypes such as hypersensitive (overreactive), underreactive, and motorically disorganized or impulsive. Are there biologically based differences in pain thresholds? Are there preexisting physical or mental disorders?

Is there a family history of mental illness? A biological predisposition (diathesis) to mental disorder could become manifest because of the stress of medical illness or the medications used to treat it.

Research shows that a small proportion of children develop mental disorders during the course of their illness and that chronic illness doubles the risk for behavioral and emotional problems. These children usually have a personal or family history of mental disorders. Current practice is to diagnose and treat children if they fulfill the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria for a disorder. One must not automatically dismiss symptoms as a normal reaction to illness (the “I'd be depressed too if I had cancer” response). Diagnoses commonly encountered by consultation-liaison child psychiatrists in pediatric hospitals include adjustment disorder, acute stress reactions, posttraumatic stress disorder, anxiety disorders, and depression.

An 11-year-old boy with cancer was treated with steroids. The nurses reported that he was becoming difficult to care for and that he was refusing treatments and “acting weird.” An evaluation showed that he had become manic. His iatrogenic mania disappeared after a steroid medication adjustment.

A 7-year-old girl with both cancer and generalized anxiety disorder became hysterical during every blood test and medication injection. The doctors called a psychiatric consultant who taught her cognitive-behavioral coping skills. Over time, her hysterical response to medical procedures was significantly reduced but not entirely eliminated.

A 14-year-old girl, newly diagnosed with cancer, became increasingly quiet. She reported a change in energy and appetite and feelings of helplessness and hopelessness. She told her mother that she wanted to die. A psychiatric consultation showed that the girl was depressed. She responded well to antidepressant medication.

Psychological Sphere The main questions to ask in the psychological category are aimed at understanding how persons think about and react to their illness. Individuals have unique perceptions of these events, which are influenced by their past experiences, and their actions will be based on these perceptions. Problems may arise when various persons involved have conflicting or mismatched perceptions about the facts of any given situation. Questions to elicit the child's perception should be asked in a developmentally appropriate manner. Specifically, one might ask the child:

What is going on? Follow up with specific questions until you get a good sense of what the child thinks is going on. Ask specifically what such key individuals as the doctors, nurses, and his parents have told him. Check for discrepancies, however small, between the child and the others involved in the child's care.

How long has this been happening? Remember that young children have a different perception of time than adolescents or adults.

What happened to cause you to be sick? Check especially for idiosyncratic and pseudological explanations, which often occur, especially in young children.

What do you think will make you better? Do you think you will get better? The answers to these questions can greatly increase or compromise compliance with treatment.

Have you ever been sick before? If so, what was that like?

Do you know anyone or have you heard of anyone who was sick? Do you know anyone who was sick with the same thing you have? What happened to them? Specialty wards may contain several children with the same diagnosis, which leads to identification and special relationships. The fate, good or bad, of these children with the same diagnosis can greatly influence how a young patient and caregivers feel.

How are you feeling?

What is the worst part about being sick? Are there any good parts to being sick?

Have there been any changes (in you or your family) since you became sick?

Anything else going on with you or your family? The evaluator wants to know how the illness fits into and changes the life story, or narrative, of the child. In some cases the illness can have a positive impact (e.g., “I am a survivor of cancer!”) or a negative impact (e.g., “I am a victim of cancer!”). For some, especially those who become sick at an early age, the illness can be an organizing theme of their life.

A teenager who had been in remission for years and was considered “cured” wanted to go out for the football team; his mother was reluctant to allow this because of his childhood cancer. Another teenager in long-term remission constantly worried that his cancer would return. His worrying had caused him to restrict his activities and caused him considerable distress.

Social Sphere The main questions in the social category are aimed at understanding the strengths and weaknesses of the caretaking environment including individual, cultural, societal, and national values. Answers to these questions should give the clinician a sense of how the family and other caregivers (including the medical team) are dealing with the ill child. In keeping with the developmental biopsychosocial approach, each family member and caregiver is considered to be a separate, yet interrelated entity with its own developmental, biological, psychological, and social issues. Akin to Russian nesting dolls, the child is supported by caregivers who in turn are supported by others who in turn are supported by still others, and so forth. Like the child confronted by illness, hospitalization, or surgery, the caregivers do the best they can with the resources available to them. Thus, the following questions, similar to those asked of the child, need to be asked of the caregivers as well.

Developmental Questions Where are the caregivers developmentally? Not only children vary in their developmental stages, and adults are not homogeneous in their ability to cope. Many adults have limited ability to abstract and to understand illness.

> The only child of a teenage mother is diagnosed with cancer, and the only child of an older professional mother, who put off having a child because of her career, is diagnosed with cancer. The different developmental levels of these two mothers suggest quite different scenarios, even though the type and prognosis of the cancer may be the same.

Biological Questions Several biologically based questions concern clinicians. One such question attempts to determine the innate intelligence of the primary caregiver. Occasionally caregivers are so compromised in this area that special intervention programs are needed.

> The mother of a 12-year-old girl with cancer had dropped out of school at age 14. She had been engaged in a series of housekeeping positions before becoming pregnant at age 15. After that, she applied for and received welfare payments. The mother had difficulties comprehending the details of her daughter's illness. Special sessions were held to educate the mother, which involved showing her pictures of the medications, alarm watches, and calendars to aid in compliance. The social history showed that the more intelligent 12-year-old daughter regularly acted as an interpreter and caretaker for her intellectually compromised mother at home and in their community.

What is the temperament or personality of the caregiver? Although temperament and personality are not completely biologically determined, evidence suggests that in their development they may serve as a substrate, acted upon by the biological makeup of an individual. Temperament and personality act as a filter through which medical information, decision-making, and communication flows.

> A brilliant oncologist was incapable of taking responsibility for mistakes and always blamed other members of the team. The resulting feelings of the accused team members led to considerable tension and dysfunction on the team.

> The nurses complained bitterly about a father who told them not to give his daughter her medications if she didn't want them. The nurses complained that they were too busy for such nonsense. “He's always so bossy and blows up when he doesn't get his way. He has macho man syndrome.” They were puzzled that this same “macho man,” when initially told of his daughter's diagnosis, sweated profusely, turned white, and became nauseated.

Does the caregiver have a preexisting mental or physical disorder?

> The oncology team was quite concerned when they heard that the mother of one of their patients had schizophrenia. They were relieved when they found her to be quite compliant. The doctor on the case stated that the father made all the decisions but that the mother attended all conferences. He admitted wondering what she understood about her daughter's condition.

Has the stress of the illness triggered a mental disorder in the caregiver? Symptoms of anxiety, depression, and posttraumatic stress disorder have been described in parents of children with chronic medical conditions.

> A father began drinking when his 3-year-old son was diagnosed with leukemia. His condition deteriorated to the point where he needed to be escorted off the hospital ward by security and was verbally abusive to a staff nurse while he was under the influence of alcohol.

Psychological Questions These questions are aimed at understanding how the illness fits into and changes the story, or narrative, of the caregiver's life. How psychologically minded are the caregivers? Can they understand that the illness is a stress for themselves and other persons involved? Most importantly, can they empathize with the plight of their sick child and appreciate the child's needs?

> The mother of a 7-year-old girl with leukemia seldom visited. She stated that she had a major project at work that needed to be completed. When confronted with the impact of this decision on her child and the fact that the child was quite fearful, the mother stated; “She'll be all right. You have to be tough in this world to get along. She is always asking for attention. She's a big baby!” Work with this mother showed that she herself had been raised by emotionally distant parents who tolerated only performance achievement and discouraged any display of emotional weakness. After the mother understood her own anger and sadness about her own upbringing, she became more available to her daughter.

How do the caregivers deal with stress? What coping and defense mechanisms do they use (denial, rationalization, intellectualization)? The professional who deals with children with chronic medical conditions must constantly use more or less healthy ways of coping. The variety and importance of these response styles must be appreciated and dealt with in a manner that allows optimal coping. Individuals are often not aware of their defenses, and attempts to alter them without substituting healthier defenses can lead to worse functioning. Treatment teams have tremendous difficulties with parents who deny aspects of the diagnosis or prognosis of their children. The team often demands that parents face reality in ways they may not be able to handle, with problematic results.

> A father whose child was dying of an inoperable brain tumor seemed oblivious to this fact. The attending physician demanded that the social worker confront the father. During the meeting with the social worker, the father broke down in tears. On his way home he was involved in a minor auto accident. For a week he was unable to work and did not visit his daughter because of his injuries. When he returned, the father seemed just as oblivious. There was no indication that the social worker's intervention had been effective.

> A father from a distant town, whose 2-year-old daughter was diagnosed with leukemia, refused to admit that his daughter was sick. The nurses said that he was in “complete denial.” His wife moved in with her parents to be closer to her daughter during her hospitalizations. The father remained in the family house 50 miles away. The wife told her husband that he had better come to grips with what was going on or she would end up going her own way.

> A first-year child and adolescent psychiatry resident dropped out of his training program after becoming upset on a consultation on a 3-year-old child dying of cancer. The resident said that he couldn't stop thinking about his own 3-year-old child while doing the consultation. He found himself worrying that his own child might have cancer.

> An oncology team with a weekly psychosocial conference discussed deaths as they occurred. The team decided it might be better to discuss all the deaths at once during a monthly meeting. After the first monthly meeting, the plan was changed back to the weekly schedule. Several of the team members said that they were so upset after discussing so many deaths at once that they could not go to their rounds afterward.

Under normal circumstances, how does the caregiver deal with stressful situations? Individuals deal with stress at their own pace in their own way. These normal processes should not be confused with pathological states.

> The clinical team was concerned with a father's denial. “He never asks us any questions about his daughter's condition.” The consultant asked whether the denial impaired the treatment process in any way. The team said no. Within a week the parent was talking openly about his child's serious condition.

Are there any mismatch situations? Various key persons involved may be at different places in their progression of attempts to deal with the illness. Although each person's way of coping is not pathological in itself, problems can arise trying to balance the needs of different individuals in the group, at the same time.

A doctor wanted a parent of a terminally ill child to sign a do not resuscitate (DNR) document. The single parent refused and instead requested that a painful procedure with a small chance for success be performed. The parent wanted to “keep on fighting” and try everything. This angered the doctor who did not want to continue with a treatment that he felt would cause unnecessary pain to the child. In this quite common situation, the needs of the child, the doctor, and the parents were at odds. Yet, both the doctor and the parent are making reasonable requests from their own perspectives.

Do the caregivers have difficulties dealing with children at certain developmental ages (regardless of whether they are sick)? Often such troubles complicate the situation. The classic difficult ages are those of the toddler and the adolescent.

A mother, who was in constant battle with her teenage daughter over how much freedom the daughter should have, continued to have problems after the daughter's diagnosis of cancer. Their battles intensified as the mother became more protective of her daughter. The daughter complained that she was being treated like a baby. A consultation was called when the daughter refused to take her medications. In his interviews with the mother, the consultant found that she had had trouble with her mother at the same age and had in fact run away from home.

What is the impact of past or concurrent illnesses on the caregivers? Just as researchers base their predictions on past data, so do caregivers. The course, intensity, treatment complications, and outcome of other persons with the same illness can influence the perceptions and thus the actions of caregivers.

A mother whose own mother and grandmother had died of cancer within the previous 2 years was sure that her son, recently diagnosed as having osteogenic sarcoma with a good prognosis, would also die. The nurses found themselves arguing with the mother about whether her son would die or not. They were upset that the mother was not more hopeful.

A mother who had had cancer and was successfully treated as a teenager was consumed with guilt when her son was diagnosed with cancer. The mother was sure that her genetics had caused the cancer in her son.

Parents on an oncology unit become attached to the other parents on the unit. They become especially close to those with children similar in age to their own child or with similar diagnoses. They offer support to each other, share the ups and downs of treatment, and are devastated when any of the children dies. When this occurs, the parents go into mourning and become increasingly worried about their own children. This worry intensifies if there is more than one death in a short period.

What are the caregiver's past experiences with health care providers and systems? A bad experience can create anticipatory anxiety, complicate relationships with the health care team, and complicate treatment.

A mother was quite upset after it took several weeks for her local pediatrician to make the difficult diagnosis of leukemia in her daughter. She felt that the pediatrician had been incompetent.

She worried that the delay had compromised her daughter's prognosis. Her distrust of the pediatrician generalized to the pediatric oncologists who took over the case. The mother started to question all the suggestions of the oncologists, who became furious at having their competence challenged. The oncologists' initial efforts to educate the mother on the error of her ways just made matters worse.

Social Questions The main questions in the social category are similar to those asked about the social support system of the ill child. These questions focus on the strengths and weaknesses of the support systems of the caregivers, to describe the larger support context. What are the supports of the caregiver? Is there a supportive spouse or significant other? Is there a supportive family, friends, or a supportive religious community? Is the caregiver happily married? Studies show a slight increase in the divorce rate in families with a chronically ill child. Is the caregiver in a supportive work environment?

The oncology team was upset when it was hard to plan a family conference. The father said he had to work. When confronted with his lack of participation, the father said that his boss had threatened to fire him if he took more time off and that if he were fired he would lose his insurance coverage. Considering this new information, the family meeting was rescheduled to accommodate the father's work schedule.

What are the religious beliefs of the family? Religion plays a pivotal role in the way many families deal with illness. Religion can provide an understanding of the event and place it in a larger spiritual context that may or may not prove helpful. Various religions provide different explanations of the place of illness in life. Because of this, it is important to know the tenets of a family's religion. However, persons interpret religious doctrine in ways that reflect their own personality and past experiences.

An oncology team was challenged each time a patient who is a Jehovah's Witness was admitted because Jehovah's Witnesses refused blood transfusions. The team, through experience, has learned to respect the patient's religion while making it clear to the family that the team's medical responsibilities also need to be respected.

A Baptist minister of a child dying of cancer announced to his congregation that he had a premonition that God would perform a miracle. The family who had acquiesced to their daughter's fate and signed a DNR order, rescinded it in deference to the minister's pronouncement.

During his last visit home before his death, a 5-year-old boy was found packing a suitcase. When asked what he was doing, he smiled and said that he was going to die and was “packing his bags to see God.”

What are the prevailing cultural and societal views on illness and death? Societies and cultures vary in their views of illness and death. These variances can greatly influence the course of the illness and its treatment.

In the span of a year, an oncology team had four Vietnamese patients. All four proved difficult management cases. In a team meeting a nurse stated: “We do not seem to understand them [the Vietnamese] and they do not seem to understand us.” Two patients actually left treatment saying that they were going to seek the advice of their local healers.

Where are the medical team members developmentally, biologically, psychologically, and socially? Each team member has a unique approach to and understanding of illness. Many health care providers chose their specific professions because of their own past experiences with illness. Their understanding of their motivations may be unclear or inexact, leaving the professionals prone to over- or under-reacting to similar situations in their patients.

One of the child psychiatric consultants to the oncology team became interested in oncology because his brother was diagnosed with cancer when they were growing up. He often recounted how his brother's illness changed his family's and his own emotional equilibrium. He made sure, therefore, when consulted, to ask questions about the impact of the illness on the siblings.

What impact does health care reform have on families? In many cases, it has significantly changed the interactions between the child, family, and the health care delivery team.

The shorter lengths of stay mandated by managed care have changed the ability of the team to make relationships, to educate, and to assist patients and their families in dealing emotionally with their illness. The team must depend on other caregivers (e.g., home health care nurses) to perform the duties they themselves once did during the longer hospitalizations.

ILLNESS-SPECIFIC VARIABLES

Variables that depend on the nature of the illness may influence treatment outcome. Several of these variables interact with each other and cannot be viewed apart from the personal characteristics highlighted in the iopsychosocial model.

Chronic Versus Acute Illness A transient illness tends to be less disruptive than chronic illness in the long run. Generally life quickly regains its preillness equilibrium. Chronic illnesses pose an ongoing challenge to children and their caregivers. Most families adapt successfully and find a new equilibrium that balances the needs of the illness, the child, and the caregivers. As developmental, biological, psychological, and social situations change, new adaptations are necessary to prevent future problems.

Although all chronic conditions entail some stress and unpredictability, not all illnesses can be viewed in the same way. Children with cancer, asthma, and cystic fibrosis have dramatically different experiences with illness, hospitalization, and necessary treatments. For example, a meta-analysis of depression in children with chronic medical conditions and controls (Fig. 49.5-2) yields dramatically different results for different illnesses. The lower rates of depression in cancer may result from the propensity of this population to use denial or repression to adapt. On the other hand, the circumscribed nature of treatment may be protective. Whereas the painful crises in sickle cell anemia are intermittent, unpredictable, and long term, most childhood cancers have a characteristic treatment progression and most unpredictable events occur in a window of time rather than throughout life. The unpredictable nature of the illness in early life, as in sickle cell anemia and asthma, is stressful to the child and thus may result in higher rates of depression in these long-term illnesses.

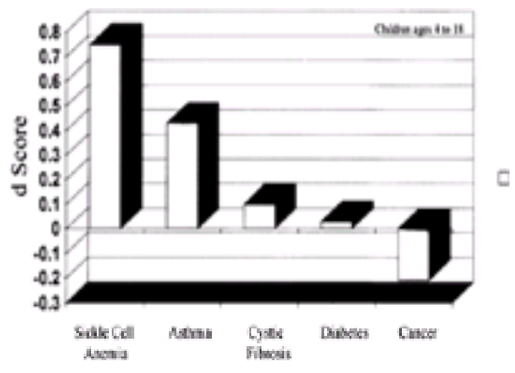

FIGURE 49.5-2 Rates of depression in children and adolescents with chronic medical conditions. A *d* score of 1 indicates depression scores that are one standard deviation apart from the control population.

Age of Onset There are developmental periods when illness can have a greater impact. Early childhood and adolescence are especially sensitive periods. In early childhood, cognition limits the child's ability to comprehend the complexities of the illness and is further complicated by fears of separation from the attachment figure. In adolescence, the illness threatens the normal push toward autonomy and necessitates more dependency on parents and other adults. A review of illness-specific developmental issues (Table 49.5-1) helps one anticipate potential difficulties and prepare preventive interventions.

Etiology The reality or the perception of how the child got the illness makes a difference. If the illness is genetic, the parent on the side of the family with the trait may have to deal with feelings or accusations of guilt or both and concern for siblings or future generations. In addition to the caregivers' concerns, the unique and developmentally influenced theories of etiology of the children can also cause concern.

Diagnosis Delayed diagnosis or misdiagnosis, especially if it is perceived as changing the course of the illness, can produce guilt or affect the family-doctor relationship. Guilt occurs when caregivers perceive that they neglected to do something that could have prevented the illness or modified its severity or course. Tension between the family and the primary care physician or between the primary care physician and the specialist can be exacerbated when the specialist is perceived as a hero for making the diagnosis. In such cases, primary care physicians are often covertly or overtly cast (or cast themselves) as dolts. Obviously, guilt is not the sole domain of the parents.

The perceived severity of the diagnosed illness is also important. The diagnosis of cancer continues to terrify people, despite the numerous advances in medical treatment that have considerably lessened morbidity and mortality. The importance of education in these circumstances cannot be underestimated.

Rare diagnoses can cause distress and bewilderment, especially when the health care worker does not have a working knowledge of the disorder. Families often report feeling isolated by having a child with a disorder that sparks no recognition among peers or medical personnel. This isolation is lessened somewhat by communication with similar families over the Internet.

Deformity and Disability As a rule, children and their caregivers experience increased difficulty when there is more deformity, disability, or both. One must ask how much the illness impairs the lives of the child and the family. Physical deformity is especially difficult for adolescents who are sensitive to issues of body image. The visibility of the deformity is important; those that show tend to be more difficult to deal with. The impact of any disability or deformity must be understood in the context of the lives of the children and their caregivers. For example, a minor disability of the knee that threatens the basketball career of an athletic adolescent is more devastating than the same disability in an adolescent fascinated by computers.

> A 7-year-old girl who was known for her thick, long, blond hair was especially devastated by her hair loss after chemotherapy.

Treatment Regimens What the treatments entail makes a considerable difference. Often treatments (surgery, chemotherapy, bone marrow transplants) with their accompanying pain, deformity, disability, and costs are perceived as worse than the primary illness itself. This is especially true when the illness onset is insidious and involves little pain and discomfort despite a potentially dire prognosis. One should not underestimate the emotional pain of young children who must go to a hospital when they are (by virtue of their stage of development) dealing with separation concerns and greatly attached to their family, home, bedroom, and so on. Young children may view the procedures involved in treatment as more traumatic than the threat of death. The perception of the trauma involved in the treatment is a strong predictor of posttraumatic stress in either the child or parent.

Prognosis The more negative the prognosis, the greater the potential for stress and risk to the family and the child. A prognosis that includes illness-related death or its threat adds a complex set of circumstances for all those involved.

INTERVENTIONS TO MINIMIZE THE RISKS OF ILLNESS AND HOSPITALIZATION

Interventions should be aimed at mitigating risk factors and reinforcing and supporting positive coping mechanisms in the child, family, and caregivers. The goal is to promote the healthiest possible functioning among all concerned and return families to a normal developmental trajectory. Although every case is different, some interventions are productive in most cases.

Be sensitive to the developmental, biopsychosocial complexities of those involved and provde a supportive relationship for the child and family. Effective health care providers need to know not only the developmental biopsychosocial complexities of those involved but also the complexities within themselves that might impede the formation of health-promoting relationships. These providers must inventory their own responses and reactions that could interfere with effectiveness, and they should be especially sensitive to types of patients or families who arouse unusual responses such as underinvolvement, overinvolvement, anger, sadness, or anxiety. If such responses occur repeatedly, talking with a colleague or consultant may help.

Educate all involved parties about the illness, and its assessment, and treatment. Our society promotes openness and informed consent, which helps the child and

family cognitively grasp their situation. Information should be presented sensitively in a developmentally appropriate manner, with an appreciation of the various responses that may be elicited. Defenses such as denial, projection, and intellectualization should be expected and be challenged only when they impede the progress of treatment or cause disproportionate distress.

Realize that children can talk about almost anything (at their developmental level). Many health care providers underestimate the capabilities of young children and make little attempt to communicate with them. This leads to many missed opportunities for both the provider and the child, who often is desperate to know what is happening.

Be available to talk to children and their parents. Make opportunities available for conversation among health care providers, parents, and children. Families may be too shy, too afraid, or too intimidated to initiate conversation. Failure of health care professionals to respond to initial and perhaps feeble inquiries for information may set the stage for a continuing cycle of missed opportunities to communicate.

Speak at the developmental level of all those involved. Everyone needs to have information delivered at a level they can understand. It is helpful to have people repeat what they have been told to ensure that they understand it. Anxiety and stress may impair the ability to hear and process what is heard. A parent may recall hearing “Your child has leukemia” and none of the treatment plan carefully explained to them during the initial hospitalization. Thus, repetition of important material is often appropriate and necessary.

Maintain family and life routines as much as possible. Particular attention should be paid to regression to developmentally earlier behaviors that do not remit over time. Overinvolved caregivers who function inappropriately and reinforce dependent behaviors that were not present before the illness need to be dealt with. A decrease in peer relationships indicates deviation from baseline functioning. If a child is hospitalized, hospital schooling, child life programs, and opportunities to be with peers should be planned. The hospitalization and treatment should not reinforce regressive forces in the child and family any more than necessary.

Identify and encourage larger support systems. Family therapists often speak of all those who are involved who can either cause a problem or help alleviate it. This approach leads to identification of persons such as grandparents, teachers, friends, siblings, and other relatives who can assist in the treatment. This larger network can be enlisted to help with problems such as how the mother will get to the hospital or who will take care of the other children while the parents are at the hospital. Both the family and the medical team must be supported and nurtured to remain effective.

Consider illness-specific support groups. These groups can often provide added, individualized support to patients and nonprofessional caregivers. Other families who have undergone similar experiences with ill children might help them.

Be on guard for psychiatric disorders. The current rule is to treat a disorder if the person has a diagnosable condition that meets DSM-IV criteria. Dealing with a mental illness while dealing with illness in oneself or a loved one is usually detrimental to the outcome of the primary illness. Studies have identified a relation between maternal distress after a child's diagnosis of cancer and the later psychological adjustment of the child. Parents at risk include those with psychiatric conditions such as depression or anxiety, higher trait anxiety, and less family support. Trait anxiety is also associated with higher rates of posttraumatic stress disorder years after the diagnosis of cancer. Thus, identifying parents in distress and providing supportive measures or treatment provides some protection to the child over time.

Treat pain aggressively. The gate control theory considers the perception of pain within a developmental biopsychosocial model. This theory emphasizes that the procedure alone does not define the level of pain, but the combination of the procedure and such variables as cognitive development, previous history with pain, and the perceived reason behind the painful procedure. One study showed that inadequate analgesia for initial procedures in young children could diminish the effect of adequate analgesia in subsequent procedures. A lack of seeing pain through the eyes of the child is thought to contribute to inadequate pain management, which is inversely proportional to the age of the child. Younger children are less able to verbalize the complex experience of pain and may exhibit compensatory behaviors (e.g., fear of separation from parents, tuning out health care workers, or temper outbursts). Children not uncommonly remain silent or deny their pain, fearing that acknowledgment will evoke further painful procedures; thus one should obtain collateral information from the parents and look for subtle signs of pain. Since perception of treatment intensity predicts later distress, therapeutic techniques (e.g., cognitive behavioral interventions or hypnosis) may provide some protection even if the actual number of procedures remains unchanged. Since pain significantly affects a child's perception of hospitalization and illness, it must be identified and treated, with as little additional pain as possible.

If there is a death, the care and support can continue. The health care team becomes an integral part of the family system when children are ill. This relationship continues after the death of the child. Follow-up sessions can often be important to the survivors. Similar attention to medical team members who get attached to patients and their families is also important. Ongoing psychosocial conferences can maintain team mental health and morale.

SUGGESTED CROSS-REFERENCES

Jean Piaget's approach to intellectual functioning is discussed in Section 3.2 and Erik Erikson's stages of development are described in Section 6.2. Normal child and adolescent development are discussed in Section 32.2 and Section 32.3, respectively. Chapter 10 contains cognitive and mental disorders secondary to a general medical condition. Finally, noncompliance with treatment (which is often similar for psychiatric conditions and medical conditions) is described in Section 27.1.

SECTION REFERENCES

Bennett DS: Depression among children with chronic medical problems: A meta-analysis. J Pediatr Psychol *19:*149, 1994.

*Burke P, Elliott M: Depression in pediatric chronic illness: A diathesis–stress model. Psychosomatics *40:*5, 1999.

*Cadman D, Boyle M, Szatmari P, Offord DR: Chronic illness, disability, and mental and social well-being: Findings of the Ontario Child Health Study. Pediatrics *79:*805, 1987.

Cardona L: Behavioral approaches to pain and anxiety in the pediatric patient. Child Adolesc Psychiatr Clin North Am *3:*427, 1994.

Diagnostic Classification: 0-3: *Diagnostic Classification of Mental Health and Developmental Disorders of Infancy and Early Childhood*. Zero to Three/National Center for Clinical Infant Programs, Arlington, VA, 1994.

Drotar D: Relating parent and family functioning to the psychological adjustment of children with chronic health conditions: What have we learned? What do we need to know? J Pediatr Psychol *22:*149, 1997.

Engel G: The need for a new medical model: A challenge for biomedicine. Science *196:*129, 1997.

Fahrenfort JJ, Jacobs EAM, Miedema S, Schweizer AT: Signs of emotional disturbance three years after early hospitalization. J Pediatr Psychol *21:*353, 1996.

Kazak AE, Stuber ML, Barakat LP, Meeske K, Guthrie D, Meadows AT: Predicting posttraumatic stress symptoms in mothers and fathers of survivors of childhood cancer. J Am Acad Child Adolesc Psychiatry *37:*823, 1998.

Lavigne JV, Faier-Routman J: Correlates of psychological adjustment to pediatric physical disorders: A meta-analytic review and comparison with existing models. J Dev Behav Pediatr *14:*117, 1993.

Lock J: Psychosexual development in adolescents with chronic medical illness. Psychosomatics *39:*340, 1998.

*McClowry SG: A review of the literature pertaining to the psychosocial responses of school-aged children to hospitalization. J Pediatr Nurs *3:*296, 1988.

*Mrazek DA: Chronic pediatric illness and multiple hospitalizations. In *Child and Adolescent Psychiatry: A Comprehensive Textbook,* M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Nir Y: Post-traumatic stress disorder in children with cancer. In *Post-Traumatic Stress Disorder in Children*, S Eth, RS Pynoos, editors. American Psychiatric Association Press, Washington, DC, 1985.

Parmelee AH: The child's physical health and the development of relationships. In *Relationship Disturbances in Early Childhood: A Developmental Approach*, AJ Sameroff, RN Emde, editors. Basic Books, New York, 1989.

Perrin EC, Gerrity PS: There's a demon in your belly: Children's understanding of illness. Pediatrics *67:*841, 1981.

Piaget J: *The Essential Piaget*, HE Gruber, JJ Voneche, editors. Basic Books, New York, 1977.

Pittman F: *Turning Points: Treating Families in Transition and Crisis*. Norton, New York, 1987.

Sawer MG, Streiner DL, Antoniou G, Toogood I, Rice M: Influence of parental and family adjustment on the later psychological adjustment of children treated for cancer. J Am Acad Child Adolesc Psychiatry *37:*815, 1998.

*Schonfeld DJ: The child's cognitive understanding of illness. In *Child and Adolescent Psychiatry: A Comprehensive Textbook,* M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Stein REK: Chronic physical disorders. Pediatr Rev *13:*224, 1992.

*Stuber M: Psychiatric sequelae in seriously ill children and their families. Psychiatr Clin North Am *19:*481, 1996.

Von Bertalanffy L: *General Systems Theory*. Brazillier, New York, 1968.

Weisman SJ, Berstein B, Schlechter NL: Consequences of inadequate analgesia during painful procedures in children. Arch Pediatr Adolesc Med *152:*147, 1998.

Woodgate R, Kristjanson LJ: “Getting better from my hurts”: Toward a model of the young child's pain experience. J Pediatr Nurs *3:*233, 1996.

Woolston J: General systems issues in child and adolescent consultation and liaison psychiatry. Child Adolesc Psychiatr Clin North Am *3:*427, 1994.

49.6 PSYCHIATRIC SEQUELAE OF HIV AND AIDS

JENNIFER F. HAVENS, M.D., SHEILA RYAN, C.S.W., AND CLAUDE MELLINS, ĦH.D.

The human immunodeficiency virus (HIV) epidemic has shifted over the last 5 years to affect growing numbers of low-income, minority, inner-city women of childbearing age. While advances in medical treatment may decrease the number of perinatally HIV-infected children and increase the life span and quality of life for HIV-infected adults and children, the sequelae of parental illness and loss will continue to affect thousands of children and adolescents. Most of these children and adolescents must struggle with familial HIV illness in the context of already troubled family backgrounds. Mental health evaluation and interventions for HIV-affected children must be informed by a clear understanding of the mental health sequelae of parental substance abuse as well as the complex and dynamic stressors of HIV illness.

EPIDEMIOLOGY

As of December 1997, there were 98,468 cases of acquired immune deficiency syndrome (AIDS) in women in the United States and 8086 cases of AIDS in children under 12, with 91 percent of the pediatric cases caused by perinatal transmission. The increased incidence of AIDS in women is also reflected in the increasing numbers of children, adolescents, and young adults who are losing one or both parents to AIDS. In 1992, David Michaels and Carol Levine projected that between 72,000 and 125,000 children and adolescents will lose their mothers to AIDS by the year 2000. No estimates are currently available of the number of children who will lose their fathers, or both parents, to AIDS. While combination drug therapies have recently been found to have the potential to reduce AIDS-related mortality, it is unclear whether these new treatments will be broadly effective and accessible, particularly to the population of parents with histories of substance abuse.

In the United States, AIDS in women and children represents a confluence of two major epidemics plaguing American cities, HIV infection and substance abuse. Intravenous drug use is the risk factor for HIV infection in 44 percent of the AIDS cases in women and heterosexual contact with a drug user accounts for a significant percentage of the remaining cases. Reflecting the risk factors for HIV infection in women, 53 percent of children with AIDS were born to women whose risk factor for HIV infection is their own or their partner's intravenous drug use. Recent findings on the efficacy of zidovudine (Retrovir) in reducing the rate of transmission of HIV from mother to infant indicate a mechanism for reducing perinatal infection. However, drug-abusing women, those at highest risk for HIV infection, may be the least likely to access this important treatment advance.

ETIOLOGY

Mental Health Risk Factors Associated With Parental Substance Abuse Risk factors for mental health problems for children with drug-abusing parents cluster in two broad areas: (1) sequelae of parenting deficiencies associated with parental substance abuse (neglect, abuse, discontinuity of attachments and disruption of placement) and (2) the biological risks associated with parental drug addiction (prenatal drug exposure, heritable parental psychiatric disorder).

Neglect, Abuse and Disrupted Attachment Associated With Parental Drug Addiction One of the most devastating effects of substance abuse is the erosion of parenting capacities essential to children's development and well-being. Children living with substance-abusing parents are at increased risk for neglect, abuse, exposure to domestic violence and disruption of attachments, all of which increase children's risk for mental health problems.

Biological Risks Associated With Parental Drug Addiction

PRENATAL DRUG EXPOSURE Several studies have described the developmental and behavioral problems of children prenatally exposed to drugs. Methodological problems in much of the research on these children limit definitive conclusions regarding the link between specific drug exposures and behavioral or cognitive outcomes. However, clearly, maternal drug use during pregnancy with its associated problems, such as inadequate prenatal care, premature birth, and low birth weight may increase children's risk for developmental, cognitive, and mental health problems.

HERITABLE PSYCHIATRIC DISORDERS Adults with substance abuse disorders have high rates of psychiatric comorbidity, in particular, affective and anxiety disorders and childhood histories of attention-deficit/hyperactivity disorder. Children born to substance abusers are at increased risk for mental health problems, because of the heritable nature of the psychiatric disorders prevalent in adult substance abusers.

Furthermore, for reasons that are probably both genetic and environmental, substance abuse in the parental generation frequently clusters in extended families. Extended families with two or more substance-abusing adult siblings are not unusual, it is rare to encounter two or even more adult siblings with HIV infection associated with substance abuse or substance-abusing sexual partners. In these families, material and emotional resources may be too strained to intervene effectively on behalf of grandchildren, nieces, or nephews affected by parental HIV illness; this lack of extended family capability may place the child at further risk.

In summary, HIV illness commonly strikes families already struggling with substance abuse and its associated problems, including impaired parenting capacities and psychiatric disorders. When a child's parent is substance abusing and HIV positive, the HIV disease may play a less significant part in the child's mental health problems than do issues related to the substance abuse.

CLINICAL FEATURES

Neurodevelopmental Sequelae of HIV Infection There have been consistent findings of significant neurological, developmental, and cognitive deficits in HIV-infected children. Two relatively distinct neurodevelopmental patterns have been described: progressive encephalopathy and static encephalopathy. Progressive encephalopathy, corresponding to the AIDS dementia complex in adults, is characterized by the loss of developmental milestones in young children and by declining intelligence quotient (I.Q.) scores and increasing difficulties with language, attention, concentration, and memory in older children. This neurodevelopmental pattern is a direct effect of HIV on the central nervous system and is associated with a poor prognosis.

Static encephalopathy, characterized by nonprogressive deficits in cognitive, motor, or language function, is not directly attributable to HIV and is most likely associated with a non-HIV risk factor, which for many HIV-infected children may involve prenatal drug exposure, prematurity, and low birth weight or heritable or environmentally mediated impairment.

Emotional and Behavioral Disorders There has been little research on emotional and behavioral disorders in HIV-infected children. Clinical reports have suggested that they are at high risk for anxiety, depression, guilt, and low self-esteem. In one of the few studies focusing on psychiatric disorders in HIV-infected children, the first author, Agnes Whitaker, Judith Feldman, and Ankel Ehrhardt found high rates of attention-deficit/hyperactivity disorder. However, these rates did not differ from those in a matched control group of children from similar backgrounds (children living in foster care who had been born to women who had used drugs during pregnancy). Features associated with prenatal drug exposure, such as prematurity, low birth weight, and heritable parental psychopathology, may actually be more potent mediators of mental health problems in children than HIV infection itself.

TREATMENT

In general, age, developmental level, or both dictate children's capacity to understand the meaning of their HIV illness. Preschoolers and young school-age children

cannot grasp the concept of chronic illness, and the strain and anxiety of their diagnosis and care falls on their caregivers. These young children are generally not informed by family members of their HIV infection. For older school-age children with a developing capacity to understand and contain information about HIV infection, a large number of issues emerge. These include learning about one's diagnosis (either through family disclosure or individual deduction); coping with secrecy, stigma and potential rejection by peers; coping with chronic illness, and anxiety.

Parents or other family caregivers often resist discussions with HIV-positive children about their health in an effort to protect them from emotional distress. Often entangled with this motivation are one or more other factors: the adult's own fear of experiencing more intense pain about the child's diagnosis; parental guilt about having transmitted the virus to the child; and concerns that the child will disclose family secrets about HIV in undesirable ways.

While for adult family caregivers and health care providers the specific diagnosis of HIV is often a highly charged, indeed an overwhelming element in communication with the children about health issues, it is in fact only a single component. The children may be more concerned about medications, medical appointments, and procedures associated with a chronic health condition or with impingement on functioning. Establishing effective communication with these children on the broad range of health-related issues is of great importance. When the efforts of family caregivers and health professionals are not achieving this goal, a therapist may need to intervene. The starting point must be listening to the children to ascertain what they know and fear and what information they actually want. Often children are seeking an admixture of medical data and affective material. It is often useful to work with the children on making a list of questions that provides a structure for adult response and helps to ensure that the children's needs frame the communication process. Ten-year-old Teresa, a child with advanced HIV disease, dictated to her therapist the following questions for her pediatrician:

> I know that it is AIDS that I have. How did I get this virus? Why do I have it and Bobby [her younger brother] doesn't have it? Did my mother do something that made me get the virus and Bobby not get it? When can I stop drinking all those pills? I asked Lucy [her nurse] when I am going to get better and she said when I drink all my pills. I drinked all my pills every day, but it's the same thing. Is it true that the pills are going to make me better? When I get married I don't want to be drinking pills in front of my husband, because he will know what I have and be scared.

Adherence to Medical Regimens HIV-infected children, like those with many other chronic illnesses, may need to adhere to complicated medication regimens, attend numerous medical appointments, and undergo frequent diagnostic procedures. Effective communication with these children about their health is the best foundation for securing their participation in these activities. Many families need considerable support in following the complex medication regimens for the child and managing the child's potential opposition to taking medication. Behavior modification techniques (especially, positive reinforcement) may be very helpful to families in dealing with these issues. Offering the child whatever measure of control possible (“Which pill do you want to take first?”) may also minimize struggles over medication.

Promotion of Optimal Development As with adults, the mental health issues for infected children change over the course of the disease. Many HIV-infected children remain relatively symptom free for long periods of time. They may not know their diagnosis, and their lives may be relatively unaffected by illness. They may, however, be stressed by the illness of a parent, substance abuse in the family, disrupted placements and ruptured attachments, or any of the other problems that so often plague the families at highest risk for HIV. Interventions for these problems and attention to promoting the children's optimal development may be more urgent matters for their mental health care during much of childhood than specific focus on HIV.

For adolescents, the normal developmental challenges of this stage, including puberty, sexuality, and the desire to fit in or be normal are seriously complicated by HIV disease. The detrimental effects of HIV on growth and pubertal development pose significant challenges for the growing number of congenitally infected young people living into their teens.

With disease progression, HIV-infected children must confront the physical and mental decline associated with AIDS. As they approach late-stage illness, they and their families must confront the terminal nature of the disease. Often, families are overwhelmed at this stage and have difficulty communicating with the children about issues related to prognosis and death. Sensitive mental health intervention by clinicians with an ongoing relationship with the children and their family can provide a place for these children to express their inevitable anxieties and fears about separation from family members and dying and help the children (if old enough) communicate their wishes about medical treatment.

AIDS Dementia In children with HIV central nervous system involvement, recognition of dementia-related behavioral sequelae and accurate differential diagnosis is clinically important for several reasons. First, parents and medical staff can misinterpret children's increasing fatigue, decreased interest in and resistance to regular activities, and regressive behavior as volitional oppositional behavior or depressive disorder. This can result in efforts to encourage them to alter their behavior when the behavior actually results from the disease process and is beyond their control. Second, early in the presentation of AIDS dementia in children, stimulants such as methylphenidate (Ritalin) can be very helpful in raising their energy level and improving their quality of life. Third, progressive organic involvement in children is an important prognostic sign, generally accompanied by severe immunocompromise and signaling the end stage of the disease. This is an important time to shift gears in the work with children and their families to help them prepare for the final stages of the child's life.

Therapeutic Support as Death Approaches As HIV disease progresses to end stage, children may experience confusion and great anxiety. Parents and other adult family caregivers and perhaps even medical and nursing staff frequently find communicating with the child about the approach of death so daunting a task that they avoid it. Dying children need a venue in which the impending death, unspeakably sad for many of the adults in their environment, can be discussed. The children need an opportunity to express their thoughts and fears and receive reassurance. For most preadolescent children, separation from family, especially from a primary attachment figure, is the most worrisome aspect of death. A therapist may be able to draw on elements in the family's belief system that sustain hope of continuity of attachment between the deceased and surviving family members.

> James, a 7-year-old boy the size of an average 3- or 4-year-old, was aware that he was terminally ill. Neither of his parents, both HIV positive, could bring themselves to speak with him about his imminent death, leaving James isolated within the family in his painful awareness. James was able to share with his therapist his fear that death would be painful and that he would be lonely and scared in the afterlife. His mother's deep religious faith offered her the consolation that James would be reunited in heaven with his maternal grandmother, and that she and his father would join him there in due course. With his mother's permission, his therapist conveyed this information to James. In his last weeks the child frequently asked his therapist to “tell me that story about heaven again” and asked for reiteration of details about the continuity of his care by his grandmother in heaven and his mother's belief that she would join him there after a while.

CHILDREN WITH HIV-INFECTED PARENTS

A conceptual model for mental health issues associated with HIV illness in families must take into account the potential complicating factors as well as the dynamic course of HIV illness, which moves through specific predictable psychosocial stages roughly paralleling the natural history of the disease. These stages begin with diagnosis of HIV infection and continue through illness progression, late-stage illness, death, and reconfiguration of the family with the children in new care arrangements.

At any stage in this process, children and adolescents may exhibit a variety of responses to parental HIV diagnosis, illness, or loss. These responses may be characterized as (1) normative responses requiring supportive counseling, increased access to existing social supports, or both; (2) responses complicated by developmental histories of poor or disrupted attachment or trauma, indicating a need for individual or family psychotherapy; and (3) exacerbations of preexisting psychiatric disorder or precipitation of new-onset disorder, in which syndrome-specific treatment must be among the interventions.

Disclosure of HIV HIV may be diagnosed at any point in the natural history of the infection. Whenever the diagnosis is made, it typically engenders anxiety about health and anticipatory grief for a potentially foreshortened future. When the diagnosed individual is a parent, these issues are compounded by concerns about continuity of care for children in the future and by questions about communicating with children and adolescents about HIV.

Infected parents frequently withhold information about their HIV diagnosis from their children; in fact, most do not tell children prior to early or midadolescence. Even without direct communication many children and teenagers develop their own suspicions or even certitude about the parent's health problems from clues like medications, symptoms, or multiple medical appointments. Children may experience a great deal of distress in the loneliness of fearing or knowing that a parent is struggling with HIV-related illness; yet because they are not supposed to know they remain unable to reach out for adult support and reassurance.

Disclosure of a family member's HIV status to children is often misunderstood to be a discrete event rather than an ongoing process. It is in fact a complex series of cognitive and emotional communications. The content of the information communicated to children, the emotional tone with which it is conveyed, and the quality of the family context all affect their understanding of the parent's health. Moreover, children's developmental level may limit their ability to absorb complex, affectively laden material.

A preschool child's limited cognitive capacity to understand the concept of HIV infection may be overwhelmed by magical thinking about undoing the presence of the virus. A school-aged child will normatively cope with difficult information about a parent's illness by denial and may opt to deny or ignore this threatening fact much of the time. Under stress, the school-aged child may revert to magical thinking that the virus can be wished away. Adolescents do not normatively defend themselves against knowledge of parental HIV by magical thinking or absolute denial, but they may rigidly compartmentalize this information and use conscious methods to distract themselves from thinking about it.

Parents and other adult caregivers of HIV-affected children should be counseled about the importance of developmentally appropriate communication about the parent's health. The child's need for the information and readiness to receive it must be accorded priority in deciding what information to communicate and when to disclose it. It is important for the adult to be honest, while respecting the child's coping mechanisms, including adaptive denial. The family's established style of communication about emotionally charged information may not be flexible enough to allow complete disclosure but may be able to tolerate communication of part of the truth without specifying the precise diagnosis. (“Mommy has an infection in her blood. She's taking medicine to keep her healthy.”) Adults should be counseled to provide children with realistic reassurance about the parent's health and about their own future care and security. HIV-negative children may worry that they themselves are infected because their parent is, and they should be assured that they and their siblings are uninfected when they can be told so honestly. When uncertainty exists about a child's serostatus or when additional reassurance is indicated, HIV testing can be very helpful.

Therapists can intervene to support a child grappling with a parent's HIV infection even when the child is “not supposed to know” about the diagnosis. Although it would be inappropriate to convey information about a parent's health that the parent does not wish to reveal, the therapist can encourage expression of the child's thoughts and feelings in an age-appropriate manner.

> David, an 8-year-old Hispanic child, was brought to a mental health clinic by his mother. She was concerned about a recent incident in which her boyfriend, while angry with her, had told David that his mother had HIV. David exhibited significant separation anxiety and sadness and was experiencing difficulty concentrating in school. His mother refused to discuss her HIV-positive status with the child and did not want the therapist to do so either. Despite the therapist's advice, his mother persisted in refusing to communicate with him about any aspect of her illness. She insisted that he did not know that she was HIV positive.
>
> In a play therapy session, David made himself a famous doctor and the therapist became his patient. David gave the therapist a blood test and announced to her that she was HIV infected and would die. David and the therapist, playing out their respective roles, then discussed the needs of the child of the infected patient (a little boy). The famous doctor contacted the infected patient's grandmother to arrange placement for the child with her after the infected patient's death. The famous doctor then counseled the infected patient against actually telling her child, as he “would be too mad about the drugs.” At the end of the session, the famous doctor (David) and the sick patient (the therapist) took a plane to Hawaii where he cured the therapist of her HIV.

Other therapeutic issues that arise for children and adolescents learning about their parent's HIV diagnosis include fear for the parent's health and anxiety about separation from or loss of their parent. Shame and a sense of social isolation because of the stigmatization of HIV infection are also common responses. Because children's self-esteem often depends on identification with the parent, they may feel that an HIV diagnosis detracts from their own self-worth. The therapist should determine the children's understanding of the meaning of parental illness and provide ongoing support to help them cope effectively with parental HIV infection and illness progression.

Progression of HIV Illness As HIV disease progresses in a parent, children are confronted with deterioration in the parent's health and ability to function as well as separations caused by hospitalizations. In this setting, children and adolescents manifest one or more normative emotional and behavioral responses to parental illness progression. In general, normative responses differ from problematic responses in degree. For example, for a toddler some increase in separation protest may be normative, while a more intensified inability to separate would be problematic; some exacerbation of tantrumming is to be expected, while extremely disruptive behavior is a warning sign. Similarly, school-age children will normatively experience a minimal to moderate decline in school performance, but a precipitous decline indicates more-significant difficulty in coping with the parent's illness.

The importance of communication with children and adolescents about parental illness and its impact on family life increases with disease progression. Normative emotional and behavioral responses can be complicated when children experience a parent's physical, cognitive, or functional decline without accurate information about the illness and prognosis. The normal sadness, anger, and anxiety experienced by children living with an ill parent can be exacerbated by the need to manage these feelings without support from adults who are unable to allow open communication about their illness.

Given the high-risk backgrounds of many HIV-affected families, children's responses to the parent's illness may also be complicated by underlying psychiatric disorder, developmental difficulties in the children's attachment to the parent, and difficulties in family environment. When working with families affected by both drug use and HIV disease, children presenting with symptoms in response to parental illness must be carefully assessed. Attributing symptomatology to HIV-related stressors alone is often clinically inappropriate.

> Jason, age 7 and repeating first grade, had a history of behavioral difficulties in school. Following his mother's hospitalization with an opportunistic infection, these problems escalated. His teacher described Jason as overactive, distracting other children, fighting both on the playground and in the classroom, and refusing to do his schoolwork. Jason's maternal aunt Yadira, who was caring for him during his mother's hospitalization, reported that Jason was also hyperactive and inattentive at home. Yadira believed that Jason's frequent changes of placement during his preschool years and the inadequacies in his care by a foster mother and his paternal grandmother also contributed to his problems.
>
> During psychiatric evaluation, Jason exhibited a short attention span and distractibility. He could not be drawn into the symbolic play with toy hospital equipment or a family of dolls that the therapist had tried to use to elicit material about his mother's illness, but when playing briefly with super heroes and Ninja Turtles, he expressed diffuse anger. Jason told the therapist that he fought with other children at school because they taunted him, saying “Your mother's a crackhead and she's gonna die of AIDS.” After further evaluation, Jason was diagnosed with attention-deficit/hyperactivity disorder. He was given psychostimulant medication and both individual and collateral supportive psychotherapeutic and psychoeducational treatment was initiated with Jason and his caregivers. Over time, treatment of the underlying attention-deficit/hyperactivity disorder increased Jason's ability to cope with HIV-related stressors.

With the pressure of increasing parental illness, the child or adolescent may experience and express intense ambivalence toward the parent, especially if the risk factor for the parent's HIV infection is understood by the child to be a negative behavior (e.g., drug use or selection of a drug-abusing partner). In these circumstances, the child or adolescent may be sad and anxious about the parent, on the one hand, and on the other, angry that the behavior that impaired past parenting will lead to further erosion in capacity as illness progresses and perhaps ultimately to abandonment of the child when the parent dies an early death. Extreme ambivalence toward the parent is perhaps the most difficult psychotherapeutic issue in working with HIV-affected children. This ambivalence may appear when the diagnosis is disclosed and continue as HIV illness progresses in the parent.

Permanency Planning As HIV progresses in the parent, caregivers may require counseling about planning for the placement of children following parental death, commonly known as *permanency planning*. Communication about HIV illness and permanency planning tend to be closely intertwined, with difficulties in one area being reflected in the other. Although it is by no means standard practice, family-based permanency planning that actively involves the children (particularly teenagers) is much more likely to have a successful outcome than planning that excludes the affected young people. Effective permanency planning can be undermined by active substance abuse and untreated psychiatric disorders in family members as well as by family dynamics related to substance abuse.

Optimally, families affected by AIDS have begun the communication and permanency planning process prior to the final stage of parental HIV illness and other adults are available to provide care and support for affected children. However, some parents, particularly those struggling unsuccessfully with mental illness or substance abuse, present to medical care settings in end-stage AIDS without having addressed these issues. In clinical experience, these are usually the most disorganized families with the highest burden of psychiatric and substance use disorder. In these families, special attention to the mental health needs of the children and

adolescents is essential. Both the lack of family preparation and factors associated with that lack of preparation signal increased risk for mental health problems in these children.

Bereavement and Family Reconfiguration Death of a parent in childhood ruptures children's attachment to the parent before they have completed the developmental tasks necessary for functioning as an independent young adult. A wide array of emotional and behavioral problems may be observed over time following the loss, including irritable mood in younger children and sadness in older ones, behavioral difficulties ranging from increased tantrumming and fighting with siblings to conduct disorders, and decreased school performance. Studies of long-term sequelae—in particular, depression in adulthood—have had mixed results, probably because of numerous confounding variables. A positive relationship between the parents prior to the death and a strong surviving parent to keep the family intact protect against elevated risk of mental health problems. Complicated bereavement is predicted by child emotional and behavioral factors predating the loss including prior emotional difficulties, emotional lability, and poor impulse control. Many of the factors associated with complicated bereavement in childhood are present in HIV-affected children and adolescents, including emotional difficulties and negative relationships with parents prior to loss and absence of a surviving parent to provide adequate parenting after death.

Grieving HIV-affected children and adolescents most commonly move into reconfigured families with extended-family members also mourning the loss of the loved one. The social and concrete service supports available to the families by virtue of the parent's AIDS diagnosis generally diminish or disappear following the parent's death. Generally, children and adolescents orphaned by AIDS move from one situation of poverty to another, with responsibility for their care falling on financially limited and often overwhelmed extended-family members. Children and adolescents who move into nonrelative foster care must make the difficult adaptation to new family members and often very different family lifestyles.

Specific issues complicating bereavement in the children of drug-abusing parents include their struggles with complex feelings about their parents, such as anger over abandonment or neglect and internalization of responsibility for parental inadequacies. The following case material shows the interaction of emotional difficulties associated with disrupted attachments and the precipitation of an affective episode in a biologically vulnerable adolescent.

> Nancy was 12 years old when her mother, Denise, died of AIDS. Denise had a long history of severe substance abuse and was diagnosed and treated for bipolar I disorder in the last year of her life. Until her mother's death Nancy was generally high functioning, a bright child with some worrisome outbursts of temper but excellent academic achievement and usually good peer relations. Although most of Nancy's care throughout childhood had been provided by her maternal grandmother, she was deeply injured by her mother's rejection, neglect, and episodes of physical abuse.
>
> In the first several weeks after Denise's death, Nancy displayed the typical mourning signs of an early adolescent: sad mood alternating with normal recreation and socializing. She then reported that her mother was appearing to her in grotesque forms at night. Nancy believed this specter was trying to frighten her. The hallucinations abated briefly with psychotherapeutic interventions designed to allow Nancy to express her intense ambivalence toward her mother. Over the next several weeks the hallucinations returned with the onset of a full-blown manic episode. Over the next year Nancy had three hospitalizations for bipolar I disorder.

A special issue that may arise in assessing a bereaved child is differentiation of illusory phenomena related to grief from hallucinations. Children and adolescents frequently experience visual or auditory perceptions of a deceased parent in the period immediately after the death or sometimes for months or years following. Illusory phenomena associated with bereavement must be differentiated from hallucinations associated with a psychotic or mood disorder. The more benign perceptions may involve the parent calling the child's name or appearing in a familiar location in the home or neighborhood. Particularly when the family's religious or cultural context provides reinforcement for such experiences, these perceptions of the deceased parent may be comforting to the child. When the child lacks cultural supports for these experiences, they may engender concealment and confusion. It often helps the grieving children to have the therapist probe gently to determine whether they are experiencing such perceptions, and if so, to reassure them and assist them in making meaning out of the experience (e.g., a child may explain this as “My mind playing tricks on me,” or “My heart misses daddy so much that it makes my eyes see him.”)

Psychoeducation for grieving children and counseling for their adult caregivers benefit many families. It is helpful to distinguish the usual course of mourning of adults (a sustained period of dysphoria, followed by a gradual return to normal mood) from the mourning that is more typical of children, especially preadolescents (a briefer period of dysphoria punctuated by apparently normal play or socializing, followed by a longer period in which grief returns and then subsides). When a child's deceased parent has been absent or otherwise deficient in their role with the child, adult caregivers may be puzzled by the child's mourning over the loss. They need to understand that the loss of inadequate parenting may in fact complicate rather than minimize the child's grieving process.

SUGGESTED CROSS-REFERENCES

Neuropsychiatric aspects of HIV and AIDS are discussed in Section 2.8. Substance-related disorders are covered in Chapter 11. Death and dying are discussed in Section 28.5.

SECTION REFERENCES

Beardslee WR, Versage EM, Gladstone TRG: Children of affectively ill parents: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*1134, 1998.

Bierut LJ, Dinwiddie SH, Begleiter H, Crowe RR, Hesselbrock V, Nurnberger JI, Porjesz B, Schuckit MA, Reich T: Familial transmission of substance dependence: Alcohol, marijuana, cocaine and habitual smoking. Arch Gen Psychiatry: *55:*982, 1998.

*Boyd-Franklin N, Steiner GL, Boland MG, editors: *Children, Families, and HIV/AIDS.* Guilford, New York, 1995.

Brouwers P, Belman AL, Epstein LG: Central nervous system involvement: Manifestations and evaluation. In *Pediatric AIDS: The Challenge of HIV infection in Infants, Children and Adolescents,* P Pizzo, C Wilfert, editors. Williams & Wilkins, Baltimore, 1991.

Centers for Disease Control and Prevention: HIV/AIDS Surveillance Update *9:*2, 1997.

Cicchetti D, Carlson V, editors: *Child Maltreatment: Theory and Research on the Causes and Consequences of Child Abuse and Neglect*. Cambridge University Press, New York, 1989.

Coles C, Platzman K, Smith IE, James ME, Falek A: Effects of cocaine and alcohol use in pregnancy on neonatal growth and neurobehavioral status. Neurotoxicol Teratol *15:*289, 1993.

P>Connor EM, Spreling RS, Gelber R, Kieslev P, Scott G, O'Sullivan MJ, Van Dyke R, Muhammad B, Sheares W, Jacobsen RL, Jimenez E, O'Nell E, Bazin B, Del Fraissey JF: Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidoviduine treatment. N Engl J Med *331:*1173, 1994.

*Deas-Nesmith D, Brady KT, White R, Campbell S: HIV-risk behaviors in adolescent substance abusers. J Subst Abuse Treat *16:*169, 1999.

Epstein LG, Sharer LR, Goudsmit J: Neurological and neuropathological features of human immunodeficiency virus infection in children. Ann Neurol *23*(Suppl):S19, 1988.

Gonzalez NM, Campbell M: Cocaine babies: Does prenatal drug exposure to cocaine affect development. J Am Acad Child Adolesc Psychiatry *33:*16, 1994.

Goodman R, Stevenson J: A twin study of hyperactivity: II. The aetiological role of genes, family relationships, and perinatal adversity. J Child Psychol Psychiatry *30:*691, 1989.

Griffith DR, Azuma S, Chasnoff I: Three-year outcome of children exposed prenatally to drugs. J Am Acad Child Adolesc Psychiatry *33:*20, 1994.

Haller DL, Knisely JS, Dawson KS, Schnoll SH: Perinatal substance abusers: Psychological and social characteristics. J Nerv Ment Dis *181:*509, 1993.

Harris T, Brown GW, Bifulco A: Loss of parents in childhood and adult psychiatric disorder: The role of lack of adequate parental care. Psychol Med *16:*641, 1986.

Havens J, Whitaker A, Feldman J, Ehrhardt A: Psychiatric morbidity in school-age children with congenital HIV-infection: A pilot study. J Dev Behav Pediatr *15:*S18, 1994.

*Havens J, Mellins CA, Pilowski D. Mental health issues in HIV-affected women and children. Int Rev Psychiatry *8:*217, 1996.

Hutchings D: The puzzle of cocaine's effects following maternal use during pregnancy. Are there reconcilable differences? Neurotoxicol Teratol *15:*281, 1993.

Kelley S: Parenting stress and child maltreatment in drug-exposed children. Child Abuse Neglect *16:*312, 1992.

Kranzler EM, Shaffer D, Wasserman G, Davies M: Early childhood bereavement. J Am Acad Child Adolesc Psychiatry *29:*513, 1990.

McLeer S, Callaghan M, Delmina H, Waller J: Psychiatric disorders in sexually abused children. J Am Acad Child Adolesc Psychiatry *33:*313, 1994.

*Mellins CA, Ehrhardt AA: Families affected by pediatric AIDS: Sources of stress and coping. J Dev Behav Pediatr *15:*S54, 1994.

Mellins CA, Levenson RL, Zawadzki R, Kairam R, Weston M: Effects of pediatric HIV infection and prenatal drug exposure on mental and psychomotor development. J Pediatr Psychol *19:*617, 1994.

Merikangas KR, Stolar M, Stevens DE, Goulet J, Preisig MA, Fenton B, Zhang H, O'Malley SS, Rounsaville BJ: Familial transmission of substance use disorders. Arch Gen Psychiatry *55:*973, 1998.

*Michaels D, Levine C: Estimates of the number of motherless youth orphaned by AIDS in the United States. JAMA 268:3456, 1992.

Najavitz LM, Weiss RD, Shaw SR: The link between substance abuse and posttraumatic stress disorder in women: A research review. Am J Addict *6:*273, 1997.

Pelcovitz D, Kaplan S, Goldenberg B, Mandel F, Lehane J, Quarresa J: Post-traumatic stress disorder in physically abused adolescents. J Am Acad Child Adolesc Psychiatry *33:*305, 1994.

Regier DA, Farmer M, Rae DS, Locke BZ, Keith SJ, Judd LS, Goodwin FK: Co-morbidity of mental disorders with alcohol and other drugs of abuse. JAMA *264:*2511, 1990.

Rodning C, Beckwith L, Howard J: Characteristics of attachment organization and play organization in prenatally drug-exposed toddlers. Dev Psychopathol *1:*277, 1989.

Rutter M: *Children of Sick Parents*. Oxford University Press, London, 1966.

Singer L, Farkas K, Kliegman R: Childhood medical and behavioral consequences of maternal cocaine use. J Pediatr Psychol *17:*389, 1992.

*Speigal L, Mayers A: Psychosocial aspects of AIDS in children and adolescents. Pediatr Clin North Am *38:*153, 1991.

Weissman MM, Gershon ES, Kidd KK, Prusoff BA, Leckman JF, Dibble E, Hamovit J, Thompson WD, Pauls DL, Guroff JJ: Psychiatric disorders in the relatives of probands with affective disorders. Arch Gen Psychiatry *41:*13, 1984.

Wilens TM, Biederman J, Kiely K, Bredin E, Spencer TJ: Pilot study of behavioral and emotional disturbance in the high risk children of parents with opioid dependence. J Am Acad Child Adolesc Psychiatry *34:*779, 1995.

Wiznia AA, Crane M, Lambert G, Samsay J, Harris A, Solomon L: Zidovudine use to reduce perinatal HIV type 1 transmission in an urban medical center. JAMA *275:*1504, 1994.

49.7 CHILDHOOD OR ADOLESCENT ANTISOCIAL BEHAVIOR

HANS STEINER, M.D., AND S. SHIRLEY FELDMAN, ıH.D.

Antisocial behavior, which involves violating the rights of others, is an inevitable part of growing up. Through minor transgressions, at first within families and then outside in the real world, children gradually learn about societal rules and limitations. Although more serious instances of antisocial behavior are viewed as delinquent, they are also relatively commonplace, especially during adolescence. Up to one quarter of youth are apprehended by police and convicted of crimes, and the incidence of self-reported antisocial behavior is much greater than police arrests.

The relationship of antisocial behavior to psychopathology is complex. Not all antisocial behavior is psychopathological and therefore does not require treatment. Although most forms of juvenile antisocial behavior do not progress to criminality, it is quite difficult to distinguish youths with a good prognosis from those who will end up in the justice system and perpetrate severe transgressions. To help in this task, a careful delineation of normative risk-taking behavior and isolated antisocial behavior from syndromal clustering of behavior problems is necessary. Antisocial behavior must also be differentiated from more serious psychopathology, as behavioral disturbances are frequent accompaniments of many psychiatric disorders in children and adolescents, especially in boys. As there appears to be a fairly consistent progression from antisocial behavior to more severe forms of psychopathology, antisocial behavior that comes to the attention of the clinician should be regarded as a behavioral marker or risk factor for more severe problems such as oppositional defiant disorder, conduct disorder, and antisocial personality disorder.

DEFINITION

Antisocial behavior is a descriptive term referring to behavior that is hostile to the rights of others and that shows disregard for the rules and laws of a given society. Basic rights of others (e.g., ownership of property) are violated or age-appropriate societal rules (attendance at school) are broken. The behavior in question is evaluated from the point of view of a given society but is not necessarily adjudicated by the legal system. Antisocial behavior is described by the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as follows:

This category can be used when the focus of clinical attention is antisocial behavior in a child or adolescent that is not due to a mental disorder (e.g., Conduct Disorder or an Impulse-Control Disorder). Examples include isolated antisocial acts of children and adolescents (not a pattern of antisocial behavior).

Defined from a societal perspective, antisocial behavior needs to be distinguished from related or overlapping problems such as *delinquency*, a legal term referring to juveniles committing offenses against the law, and *conduct disorder,* a psychiatric disorder that considers the behavior from a mental health perspective.

There are both advantages and disadvantages to having a clinical category such as antisocial behavior. On the one hand, by serving as a clinical marker it may help bring vulnerable youth to the attention of clinicians who can implement treatment and secondary prevention. Currently, such children are handled by probation and parole systems without adequate input from mental health professionals, despite established links between antisocial behavior and serious psychopathology. On the other hand, there is concern that a psychiatric label might be injurious to antisocial youth, as they are dealt with by the juvenile justice and school systems, or that the psychopathologization of criminal behavior might delay appropriate criminological intervention.

HISTORY

Antisocial behavior was often viewed from two different perspectives: as an internal deficit or as an ecological adaptation to extraordinary circumstances of disrupted and disorganized environments. William Healy, an obstetrician who founded the first juvenile court clinics in the Chicago and Boston areas, described delinquents as having a “constitutional deficiency,” others described them as having genetic deficits. The best representative of the ecological view was August Aichhorn, who used Freudian psychodynamics in an effort to show the adaptive value of some of the symptoms of wayward youth. The establishment early in the twentieth century of clinics attached to juvenile courts increased the tension between those who regarded delinquent youth as budding psychopaths and those who believed their behavior to be reactive to adverse circumstances. *The Mask of Sanity*, a monograph written by Hervey Cleckley that emphasized both hereditary and environmental influences, tried to integrate both these viewpoints.

Subsequently, a developmental perspective on delinquency and antisocial behavior began to replace notions of constitutional inferiority. John Bowlby, for example, revealed that many delinquents had suffered prolonged disruptions of early relationships and as a result, were “affectionless characters.” Work by Lee Robins on the natural history of delinquency established a convincing link between childhood conduct problems and adult antisocial personality disorder. As a result of this developmental perspective, epidemiological studies were carried out and evaluation moved beyond court clinics into community settings.

COMPARATIVE NOSOLOGY

In 1968 the second edition of DSM (DSM-II) differentiated between diverse antisocial reactions. The term “antisocial behavior” established antisocial behavior as a clinical construct, independent from the juvenile justice system classification of delinquency. In 1980 the third edition of DSM (DSM-III) differentiated the construct of “conduct disorder” from that of “antisocial behavior.”

The category of antisocial behavior appears unchanged in DSM since the DSM-III-R, although there is as yet no evidence on its discriminant, convergent, and predictive validity. The 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10) does not contain a similar category.

Prospective studies of youth at risk have shown that there are meaningful subtypes of antisocial behavior. One useful classification describes overt, covert, and authority-defying antisocial acts with different developmental and prognostic implications. For example, the least serious antisocial acts are typically contained in one channel during childhood, most commonly in the authority-defying channel. Covert antisocial behavior, such as stealing, typically emerges in later childhood whereas overt antisocial acts (such as violence) are usually not evident until prepuberty or adolescence. Children with the poorest prognosis tend to show early appearance of overt behaviors and disturbances in all three channels.

EPIDEMIOLOGY

Antisocial behaviors are the most common reason for referrals to child psychiatric clinics, reaching as high as 60 percent of patients seen in an ambulatory setting. They are also one of the most costly disorders, usually involving personal loss to patients, families, victims, and society. Estimates of the incidence of antisocial behavior come from population-based, longitudinal studies of populations at risk, as well as from crime statistics compiled by the National Institute of Justice. Despite the widespread incidence of antisocial behavior, some groups are more at risk for this behavior than others. More so than other age groups, adolescents are especially likely to engage in antisocial behavior, including violent crimes and crimes against property (Fig. 49.7-1). For example, although adolescents only comprise about 8 percent of the population, they commit 50 percent of violent crimes.

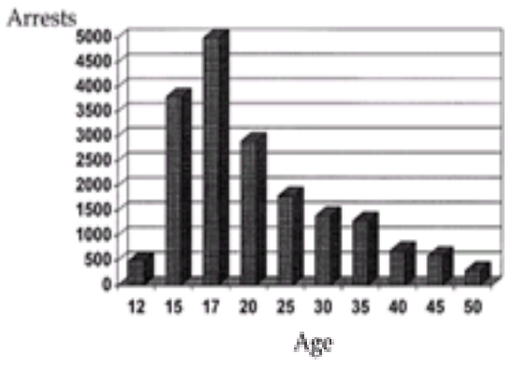

FIGURE 49.7-1 FBI index arrests by age per 100,000 population.

There are gender-specific features of antisocial behavior that become especially apparent in adolescence. First, males are two to three times more likely than females to engage in antisocial acts. Second, boys tend to exhibit more aggression whereas girls tend to commit more covert crimes and prostitution. However, recently there has been an increase in the number of violent crimes perpetrated by females, perhaps due to the inclusion of females in some gangs. Antisocial behavior used to be more frequent in urban than other settings, but this also appears to be changing. Delinquent acts are reported with about the same frequency (10 percent) in rural areas, the suburbs, and the cities.

Although not exclusively a problem of the socially disadvantaged, poverty and low socioeconomic status are common backdrops to antisocial behavior. Antisocial behavior is prevalent in communities characterized by a high level of criminality among adults, drug and alcohol abuse, overt conflict and violence, a high incidence of single-parent families, and an absence of community organization. Such environments have been labeled psychosocially toxic, an apt and especially descriptive term.

ETIOLOGY

Because antisocial behavior ranges widely in its manifestations, especially when age is taken into account, it is unlikely that there is any one set of specific etiological factors. Current data best fit a cumulative risk factor model, where the likelihood of antisocial behavior increases as risk accumulates (Fig. 49.7-2). This model can be expanded to include environmental protective factors and the individual's resilience to improve prediction as to which children succumb to criminal activity or psychopathology and which do not.

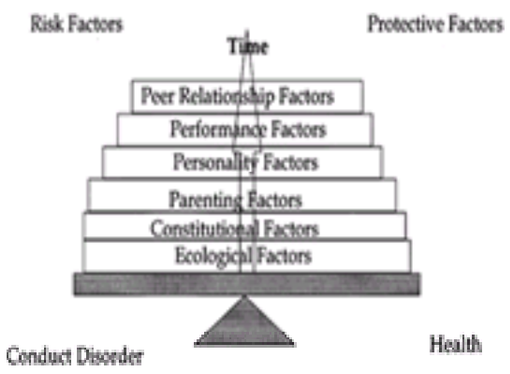

FIGURE 49.7-2 The risk/resilience model for the development of antisocial behavior and conduct disorders.

The model predicts that the greater the number of risk factors and the earlier they appear, the higher the probability that the individual will engage in antisocial acts and be characterized by serious psychopathology. The mechanism by which this occurs seems to be via multiple interactive loops—factors in the family environment interact with characteristics of the child to produce at an early age aggressiveness, impulsiveness, and a disregard for the rights of others. Such children are then rejected by peers and teachers and a new cycle of negative interactions develops.

Risk-Resilience Model It is unlikely that single risk factors operating in isolation make a major contribution to delinquency. It is more productive to think of the accumulation of diverse risks and the interaction between risks that lead to problems. Rolf Loeber has described the gradual stacking of factors in the genesis of antisocial acts. An expanded model might include a parallel pyramid of resilience or protective factors, which accentuates in part the effects of the accumulated risks.

The risk factor model highlights that many different factors, occurring at different stages of life and leaving residuals thereafter, contribute to the onset of antisocial behavior. Risk factors may even begin to operate perinatally, with such factors as the unavailability of prenatal care, maternal drug use, smoking, and poor nutrition all implicated in the development of central nervous system (CNS) impairment. After birth and throughout childhood, risk factors include difficult temperaments, insecure attachments to parents, and unavailable, inconsistent, or harsh parenting. Peer rejection and failure in academic and vocational settings also serve as markers of risk factors for antisocial behavior.

As the child grows up and as risks aggregate and accumulate over time, the child is likely to show increasing problems and deficits. In this developmental model it is increasingly difficult for current protective factors to offset a long history of risk. Once the deficits and problems become manifest as conduct disorder, treatment will be necessary if the outcome is to change.

Biological Factors Since the late 1980s there has been a steady accumulation of data implicating but not definitively proving biological risks for antisocial behavior and criminality. These factors include genetic factors, CNS insult, underarousal of the nervous system, neurotransmitter aberrations, and difficult temperament. The effect of the factors in question is seen most clearly in the face of environmental disadvantage and in the presence of other risk factors.

Heredity is the most controversial of the biological factors, with evidence stronger for adult criminality than for juvenile antisocial acts, and for nonviolent rather than violent criminality. Support for genetic factors comes from studies of family resemblance. For example, monozygotic twins are more concordant for criminality than dizygotic twins who in turn are more concordant than half-siblings. Some family resemblances, however, are not consistent with a genetic hypothesis.

Insult to the CNS either at birth or during childhood has also been implicated as a risk factor for antisocial behavior. For example, children who have perinatal damage are more likely to become criminally active 18 years later, especially if their perinatal risks are paired with maternal rejection. Support for the role of CNS factors in antisocial behavior in adults comes from brain imaging, a technique that is still in its infancy. Two hypotheses have received some support: frontal dysfunction is associated with aggressive crime and temporal lobe dysfunction is associated with sexual offending. The relationship to antisocial behavior in children and adolescents has yet to be shown.

Deficiencies in various channels of the arousal system may be a risk factor for antisocial behavior. The evidence strongly supports low resting pulse as a predictor of future criminality. Evidence from other channels, such as skin conductance and electroencephalography (EEG) is not totally consistent. Overall, the data suggest that there may be a subgroup of antisocial children who are underaroused and not likely to learn readily from and be easily socialized by people in their environment. As a result, these underaroused youths seek high levels of stimulation to raise their low levels of arousal and do this by engaging in risk taking and aggression.

Certain neurotransmitter systems, such as serotonin, noradrenalin, and *g* -aminobutyric acid (GABA), are also implicated in antisocial behavior. Aberrant levels of these neurotransmitters relate to common antecedents of antisocial acts such as aggressivity, impulsivity, and affective lability. Male hormones in humans have

yielded contradictory findings, despite support from animal studies.

Difficult temperament, characterized by irregular rhythmicity, withdrawal from novel stimuli, affective irritability and lability, and hyperactivity, has in prospective studies been shown to place a child at risk for antisocial behavior, especially in interaction with problematic family environments.

Psychological Factors Psychological correlates of antisocial behavior include neuropsychological factors, ease of learning and conditioning, intelligence, and personality factors, to name a few. Whether these are causal factors is not known, nor is it always known how they precisely might operate if they are causal factors.

Multiple neuropsychological deficits, including learning disabilities, frontal lobe dysfunction, left hemisphere dysfunction, and reduced lateralization for linguistic functions have been found in juveniles who commit antisocial acts. On the behavioral level, learning and conditioning deficits have been found under a variety of conditions. For example, antisocial youth exhibit slower classical conditioning and avoidance learning; however, the latter depends on their motivation. There are also higher-order cognitive deficits. In particular, antisocial youth have a lower intelligence quotient (I.Q.) than youth matched in socioeconomic status and age, and the deficit is particularly marked in verbal I.Q. and associated functioning such as moral reasoning. Antisocial youths also show some characteristic personality traits: problems with self-restraint, impulse control, responsibility, and suppression of aggression. These traits involving undercontrol are magnified by posttraumatic stress disorder, which is not infrequent in antisocial youth.

Despite a wealth of data on ways in which antisocial youths differ from others, research on this topic is often methodologically unsophisticated. There is a need for more long-term prospective studies and studies that match antisocial and other youths in terms of important characteristics.

Social Factors Among the many social factors that influence antisocial youths, three are particularly important—family, peers, and extended environment.

Familial Factors Parental criminality is a strong predictor of juvenile antisocial acts. This familial aggregation may reflect genetic influences or well-documented compromised parenting. Family pathology, including abuse, neglect, and sexual molestation, is common in families whose children engage in antisocial activity.

Early clinical studies implicated attachment problems as a precursor of antisocial behavior, but it is only recently that well-designed prospective (rather than retrospective) studies are being carried out. There is stronger evidence that the quality of parenting, including parental absences, lack of supervision, and ineffective parenting styles are strongly linked to aggression, impulsivity, noncompliance, and other behaviors related to antisocial behavior. An impressive body of research has described some mechanisms by which poor parenting contributes to the child's antisocial behavior. For example, parents who are inconsistent and harsh but who eventually capitulate to the coercive behavior of their offspring have undersocialized, aggressive, and impulsive children.

Family influences on children's antisocial behavior are not just limited to parent-child relations. Divorce, separation, and unhappy and conflictful marital relationship are also predictive, but less so than the parenting variables. There is recent evidence that familial variables recede in importance as the child goes through puberty. Instead, internal regulators such as restraint and the quality of peer relationships, which themselves reflect the effect of earlier parenting, become more important.

Peer Factors Negative peer influences assume importance in facilitating antisocial behavior relatively late in the developmental cycle. It is quite unlikely that peer influences alone can be the cause of serious antisocial acts. Peer influences are perhaps more important in later-onset antisocial behavior, when adolescents are particularly prone to conformity pressures. Antisocial peers provide acceptance to those youths whose behavior, personality, and social skills make them outcasts with better-socialized peers. This principle is seen most clearly in gang membership.

Extended Environment The neighborhood or ecological environment in which the child is raised has repeatedly and convincingly been shown to be part of the core risk factors for antisocial behavior. High rates of unemployment in the neighborhood, low parental social class, poor urban neighborhoods with high crime rates, and poor schools all contribute to risk. These environments are often characterized by high levels of community violence, with well-documented injurious effects on children who witness such violence.

Two other well-documented factors deserve mention. First, academic failure in school, even in the early grades, is also associated with children's subsequent antisocial behavior. Second, mass media, particularly extensive viewing of TV with violent content, also have been shown to contribute to the problem: the mindless portrayal of graphic and gratuitous violence, facilitated by the use of sophisticated computer graphics, provides undercontrolled youths with both increased arousal and models for antisocial behavior.

Protective Factors As the model suggests, there are also protective factors that attenuate risk, but much less is known about them. Protective factors refer to more than the simple absence of risk and to more than the polar opposite of stress factors. Instead, they are conceptualized as having pervasive influence on multiple domains of functioning. They exert their effect not through a simple balancing of stress factors, but by strengthening core aspects of individual functioning.

Normal or above-normal intelligence, an easy temperament, an ability to relate to others, good work habits at school, areas of competence outside school, and a good relationship with at least one parent or another important adult all offer protection against antisocial behavior and delinquency in the presence of significant risk. Prosocial peers and a good school that fosters experience of success, responsibility, and self-discipline also emerge as protective factors. The selection of nondelinquent peers, and during young adulthood the selection of a good marital partner (as demonstrated by competent social skills, a stable personality, a good work history, and capacity for good parenting) have been shown to offer protection against further criminal activity and escape from a crime-oriented career path.

It is noteworthy, however, that no studies look at the experimental manipulation of these protective variables. It would be of interest for the clinician to know if some of these protective factors can be induced or augmented and whether such a procedure would lead to a positive outcome in those at risk as well as those who are already symptomatic. To date, only studies of at-risk children who have been adopted into stable families or who as adolescents have been sent to elite boarding schools provide evidence on this issue and these are hardly controlled studies. Much of the therapeutic enterprise is based on the assumption that change is possible, but as of yet, carefully designed studies testing these or similar hypotheses have not been published.

DIAGNOSIS AND CLINICAL FEATURES

Although diagnosing antisocial behavior is relatively easy, the process of obtaining the necessary information is not. Because the patient is usually referred by others under the shadow of some potential legal transgression, it is very difficult to establish a reasonable working alliance with the patient. The sequencing of contacts with patient, parents, schools, departments of social work and probation requires perseverance and sensitivity and the purpose of the evaluation (e.g., forensic versus clinical, management versus treatment) needs to be made clear from the outset.

There are many complexities in working with antisocial children and youth. First, since children and adolescents usually underreport antisocial behavior while those around them usually overreport it, it is necessary to gain information from multiple informants including parents, pediatricians, school personnel, police, and probation workers. Second, the antisocial nature of behavior is not always apparent to the perpetrators, especially if they are relatively young and have lived in a different cultural context. Third, since single antisocial acts are the exception, it is necessary to undertake a careful examination of different domains of functioning in order to gauge the extent of the antisocial behavior. And finally, the definition of antisocial behavior depends on the age of the child: taking money from a mother's purse has a different meaning for a 4-year-old than a 12 year-old; striking one's sister at age 3 is different from hitting her at age 15.

A 9-year-old child was arrested by police for breaking into a local hardware store accompanied by two friends. The three had ridden their bikes in a suburban neighborhood until after dark and engaged in a play of “cops and robbers,” taking turns in pursuing and being pursued. To make the game more lifelike, one of the three suggested that they actually break into this store, whose owner was a somewhat gruff and intimidating man. The three decided that such an adventure would be exciting and they proceeded to smash in the glass door with a brick. Shortly after the glass broke the police arrived and arrested them. Their parents were called to come and collect their children. None of the three had any previous contact with the police or any social service agency. While they previously engaged in some mischief in the neighborhood—such as throwing toilet paper on people's houses and egging cars, none of the three had any serious infractions of societal rules, and none of the three boys had any more of these events cluster in the past few months. One of them had attention-deficit/hyperactivity disorder and some learning difficulties at school, while the other two boys had no particular risk factors for the persistence of antisocial behavior. The child with attention-deficit/hyperactivity disorder progressed in terms of his antisocial activities to the point of fulfilling conduct disorder criteria in early adolescence. Much of his future acting out consisted of drug-related offenses, that is, the use and sale of drugs, stealing, and other covert delinquent activities. The two boys without risks proceeded to develop through a normative, turbulent, and lively adolescence wihout any further legal involvement.

Antisocial behavior is even more common in adolescence, and in most cases resolves without intervention other than feedback to parents and police warnings. As in our child sample though, special risk factors may lead to a more full-blown picture of conduct disorder and multiple comorbidities that portend more poorly for the future of the adolescent.

A midadolescent boy was arrested for possession of drugs by the police and put into a diversion program because this was his first arrest. He had a history of some learning problems at school that seemed to be based on some diagnosable processing problems. These issues were adequately addressed in the school and parental home. The home environment was stable and predictable and there was no family history of a psychiatric or criminal nature. The young man had experimented with cannabis and alcohol, but there was no regular pattern of abuse or dependence. The circumstances of this adolescent's arrest were as follows: the police observed him and a friend sitting in his parent's car on a parking lot on the edge of town that was known to be the site of frequent drug deals. The two boys had gone to this place at 3 in the morning. As it turned out, the arrested youth's friend had just been kicked out from his parental home after getting into a violent fight with his father. He called the first boy to discuss his options and to seek some support. The arrested youth offered to come pick him up in a car and help him bring his belongings to a shelter. In order to do so he sneaked out of the parental home and took the parental vehicle. After collecting the friend's belongings, the boys proceeded to a parking lot that was relatively isolated to discuss the situation and do some planning. The boy who had just been kicked out of his home offered the driver some marijuana in order to calm himself. The driver refused, but stayed to help the friend through the crisis. At that point the police arrested them. The driver was sent to a diversion program, and he and his parents participated in a weeklong drug education program, following which there were no further contacts between the friends and police. The youth graduated satisfactorily and continued to graduate from college. The friend had a more complex progress. His pattern of fighting and of abusing drugs to regulate himself continued and intensified, leading to several arrests and finally to convictions of drug possession and battery. He was referred for clinical evaluation and diagnosed with conduct disorder, a depressive disorder, and intermittent explosive disorder. He was started on a treatment regimen consisting of medication targeting his depression, a remedial school program that assisted him with his difficulties in school work, due mostly to problems with motivation, and a substance abuse program, all of which were made part of his probation.

Distinguishing uncomplicated antisocial behavior from conduct disorder may not always be easy, but there are some key clinical features that indicate more serious diagnoses and that have implications for case management and treatment. These include lack of empathy and concern for others, consistent misperception of the intent of others in ambiguous social situations, lack of guilt or remorse following transgression, and low self-esteem. These features are often hidden behind a tough facade. Recklessness, poor impulse control and low restraint in stressful situations, irritability, and temper outbursts are often part of low frustration tolerance. Extensive risk taking is manifest in status offenses such as early-onset and exploitative sexual behavior, alcohol and other substance abuse, and smoking. Suicidal ideation and attempts are more prevalent than in normal teens and are equal to and sometimes greater than rates found in depressed youth.

The evaluation of a patient with antisocial problems is a complicated process that in many ways parallels the assessment of the child with disruptive behavior disorders. It may take several sessions to obtain sufficient information about the exact nature of the presenting problem. The clinician should be prepared to take a longitudinal perspective and aim at preventive interventions. Little is gained from an attitude that antisocial behaviors are normative and will eventually be outgrown. Careful vigilance and readiness to diagnose other associated conditions that might be causally linked to the behavior is a more prudent strategy.

Clinically Relevant Subtypes of Antisocial Behavior Rolf Loeber described three different forms of antisocial behavior—authority-conflicted, covert (e.g., stealing), and overt (e.g., violence). His typology is important because as antisocial behavior becomes more severe all three types of behavior are manifest with implications for prognosis. Second, inquiry into what behavioral portfolio best describes the antisocial child provides specific targets for prevention and intervention.

PATHOLOGY AND LABORATORY EXAMINATION

There are no pathognomonic medical or psychological tests for the diagnosis of antisocial behavior, but a variety of tests taken in the aggregate can help establish relevant comorbidities and a plan for management and treatment. A thorough physical examination is necessary with specific attention to vision and hearing, the evaluation of neurological conditions such as head injury and seizure disorder, as well as other medical conditions including chronic illnesses. Urine and blood drug screen may be indicated, especially when clinical evidence suggests substance abuse that the patient denies.

Several psychometric instruments, all normed for this population, might be helpful. The most frequently used is the Child Behavior Checklist, which has parent, teacher and self-report versions describing specific externalizing behaviors. The Conners Parent-Teacher Rating Scales contain a Conduct Problem Factor helpful in measuring such problems and tracking response to treatment. The Jesness Inventory and the Hare Psychopathy Checklist were developed in delinquent and criminal populations and assess information relevant to diagnoses of antisocial behavior, criminal behavior, and antisocial personality disorder.

DIFFERENTIAL DIAGNOSIS AND COMORBIDITY

Delineation from disruptive behavior disorders is most important. The diagnosis of oppositional defiant disorder is based on explicit problems with temper, anger, and manifest defiance against adults. These characteristics are not necessarily part of antisocial behavior. In conduct disorder the antisocial acts are quite repetitive and persistent, and the acts themselves are of greater severity. Antisocial behavior can occur in a wide range of other primary psychiatric diagnoses, including substance use disorders, attention-deficit/hyperactivity disorder, learning disorder, mood disorder, bulimia nervosa, posttraumatic stress disorder, mental retardation, delirium, dementia, and impulse-control disorder. Schizophrenia or bipolar I disorder may in infrequent cases also present with antisocial behavior. In older adolescents, antisocial behavior may signal early-onset personality disorders (such as borderline and antisocial personality disorders). Axis III conditions that are often associated with antisocial behavior are seizure disorder, head trauma, and chronic illness affecting CNS functioning.

COURSE AND PROGNOSIS

Longitudinal studies indicate that early onset of antisocial behavior, disturbances in multiple domains, extensive comorbidities, and a high level of ecological risk factors portend poorly for the future, especially when many of these factors are present together. Later onset and the presence of multiple protective factors make it more likely that the youth will avoid developing a conduct disorder or suffer from antisocial personality disorder. Once these diagnoses have set in, the prognosis becomes considerably worse.

Since risk taking is a normative activity in adolescents, isolated occurrences of antisocial behavior in a midteenager who has good levels of premorbid functioning and preserved functioning in the majority of current domains should not be overinterpreted. For such a youth, the prognosis is good, especially if it can be shown that some of the changes in behavior result from changes in the environment such as a move into a delinquent peer group or into a troubled neighborhood, or from a recent significant stressor.

TREATMENT

As antisocial behavior does not represent a psychopathological condition per se, the simple presence of antisocial behavior does not in and of itself call for treatment. There are, however, three initial tasks for the clinician. First, with the goal of delineating the antisocial behavior from other psychopathology, it is necessary to provide a full description of the general context in which the antisocial behaviors occurred, including both risk and protective factors in the child's life. Second, it is

recommended that the patients' progress be charted over 6 to 12 months, with serial assessments made in all major domains of functioning. These include the interpersonal domain, including family, peer, and authority relationships; intrapsychic functioning; basic physical and psychological health; academic and vocational performance; and ability to recreate without inappropriate risk taking. Third, the clinician will need to maintain contacts with parents, teachers, probation personnel, and the pediatrician in order to obtain a balanced picture of the youth's behavior.

Subsequent to a thorough assessment, preventive intervention may be undertaken, although its form will depend on the age of the child. For the family of the preschooler, home visits focusing on parent education might be indicated. During the school years, family sessions dealing with behavioral management along the lines suggested by G.R. Patterson and colleagues can be designed to disrupt coercive and abusive family cycles. During adolescence, individual exploratory sessions with the teenager, with particular attention to peer and authority relationships, may be effective. Topics might include the detailed dissection of complex social situations in which the patient clearly sees few workable alternatives to antisocial behavior and behavioral strategies to deal with aggression.

A major tool in the assessment and prevention of antisocial behavior is the use of structured settings. These are remarkably effective in controlling all but the most disturbed patients, who probably require treatment for conduct disorder and its comorbidities. Options range from augmented classrooms, such as special education or learning disorder settings, day treatment, or residential settings. Hospitalization is only indicated if there are clear and present dangers to self or others. Hospitalizations are mainly crisis driven and should not be regarded as an “inoculation” technique to prevent future recurrence.

The use of psychotropic drugs is helpful primarily when specific comorbid conditions are being addressed. However, it should be assumed that medications are of auxiliary value only and they should not be used as a replacement for more comprehensive treatment. Further, caution with psychotropic drugs is necessary because some antisocial youth live in homes and environments where medication may have street value, and thus the medication may be abused by others as well as the youths themselves.

A special problem is the management of aggression. Behavioral techniques (such as time-outs) work well, especially in young children, but they require consistent effort on the part of the parent. Sometimes, drugs may need to be given to young children to help with the management of aggression. In such cases, the drugs used for older adolescents and adults such as lithium (Eskalith), carbamazepine (Tegretol), propranolol (Inderal), clonidine (Catapres), and phenothiazines may be tried. However, the evidence for their efficacy in young children is very limited, and even in school-age and adolescent populations there is only suggestive evidence that these drugs might be helpful. Treatment plans might be modeled after the ones used for the treatment of conduct disorder. Overall, early intervention is more successful than treatment at a later stage although even the evidence for early intervention programs is not as complete as needed.

SUGGESTED CROSS-REFERENCES

Disruptive behavior disorders are presented in Chapter 40, drug and alcohol abuse is discussed in Chapter 11, attention-deficit/hyperactivity disorders are discussed in Chapter 39. Mood disorders appear in Chapter 14, posttraumatic stress disorders in Section 46.2, impulse-control disorders in Chapter 22, adult criminality and antisocial personality disorder in Section 27.3, child psychotherapy (including behavior therapy) in Chapter 48, child pharmacotherapy in Section 48.6, and development of personality in Chapter 32.

SECTION REFERENCES

Bartol C, Bartol A: *Juvenile Delinquency: A Systems Approach*. Prentice Hall, Englewood Cliffs, NJ, 1989.

Bock G, Goode J, editors: *Genetics of Criminal and Antisocial Behavior*. Ciba Foundation Symposium 194. John Wiley, Chichester, NY, 1996.

Cantwell D, Baker L: *Psychiatric and Developmental Disorders in Children with Communication Disorder*. American Psychiatric Press, Washington, DC, 1991.

Federal Bureau of Investigation: *Uniform Crime Report for the United States*. United States Department of Justice, Washington, DC, 1992.

Feldman S, Weinberger D: Self-restraint as a mediator of family influences on boys' delinquent behavior: A longitudinal study. Child Dev *65:*195, 1994.

Garbarino J: *Raising Children in a Socially Toxic Environment*. Jossey-Bass, San Francisco, 1995.

Hoge R, Andrews D, Leschied A: An investigation of risk and protective factors in a sample of youthful offenders. J Child Psychol Psychiat *37:*419, 1996.

Huizinga D, Loeber R, Thornberry T: *Urban Delinquency and Substance Abuse*. Office of Juvenile Justice and Delinquency Prevention, National Institute of Justice, Washington, DC, 1994.

*Lipsey M: Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In *Meta-Analysis for Explanation,* Cook T, editor. Russell Sage Foundation, New York, 1992.

*Loeber R, Stouthamer-Loeber M: Development of juvenile aggression and violence. Am Psychol *53:*242, 1998.

McCord J, Tremblay R, editors: *Preventing Antisocial Behavior: Interventions from Birth through Adolescence*. Guilford, New York, 1992.

Mendel R: *Prevention or Pork? A Hard-Headed Look at Youth-Oriented Anti-Crime Programs*. American Youth Policy Forum, Washington, DC, 1995.

Mrazek P, Haggerty R, editors: *Reducing Risks for Mental Disorders: Frontiers for Preventative Intervention Research*. National Academy Press, Washington, DC, 1994.

Mukerjee M: Hidden scars: Sexual and other abuse may alter a brain region. Science and the citizen. Sci Am *270* (10):14, 1995.

Patterson G, Narrett C: The development of a reliable and valid treatment program for aggressive young children. Intl J Mental Health *19:*19, 1990.

Rae-Grant N, Thomas F, Offord D, Boyle M: Risk, protective factors, and the prevalence of behavioral and emotional disorders in children and adolescents. J Am Acad Child Adolesc Psychiatry *28:*262, 1989.

Raine A: *The Psychopathology of Crime: Criminal Behaviors as a Clinical Disorder*. Academic Press, San Diego, 1993.

*Raine A, Brennan P, Mednick S: Birth complications combined with early maternal rejection at age 1 year predispose to violent crime at 18 years. Arch Gen Psychiatry *51:*984, 1994.

Raine A, Venables PH, Mednick SA: Low resting heart rate at age 3 predisposes to aggression at age 11: Findings from the Mauritius Joint Child Health Project. J Am Acad Child Adol Psychiatry *36:*1457, 1997.

Responding to the Mental Health Needs of Youth in the Juvenile Justice System. The National Coalition for the Mentally Ill in the Criminal Justice System. Seattle, WA, 1992.

Robins L, Rutter M, editors: *Straight and Devious Pathways from Childhood to Adulthood*. Cambridge University Press, New York, 1990.

Rubinow D, Schmidt P: Androgens, brain and behavior. Am J Psychiatry *153:*974, 1996.

Rutter M, editor: *Studies of Psychosocial Risk: The Power of Longitudinal Data*. Cambridge University Press, New York, 1988.

Rutter M, Casaer P, editors: *Biological Risk Factors for Psychosocial Disorders*. Cambridge University Press, New York, 1991.

Sege R, Dietz W: Television viewing and violence in children: The pediatrician as agent for change. Pediatrics *94:*600, 1994.

*Sholevar P: *Conduct Disorders in Children and Adolescents*. American Psychiatric Press, Washington, DC, 1995.

*Silverthorn P, Frick PJ: Developmental pathways to antisocial behavior: The delayed-onset pathway in girls. Dev Psychopathol *11:*101, 1999.

Steiner H, editor: *Treating Adolescents*. Jossey-Bass, San Francisco, 1996.

Steiner H, editor: *Treating School Age Children*. Jossey-Bass, San Francisco, 1997.

Steiner H, editor: *Treating Preschool Children*. Jossey-Bass, San Francisco, 1997.

Steiner H: Work Group on Quality Issues. Practice Parameters for the Assessment and Treatment of Children and Adolescents with Conduct Disorder. J Am Acad Child Adolesc Psychiatry *36* (Suppl): 122S, 1997.

*Steiner H, Cauffman E, Duxbury E: Personality traits in juvenile delinquents: Relation to criminal behavior and recidivism. J Am Acad Child Adolesc Psychiatry *38:*256, 1999.

*Steiner H, Stone L: Violence and related psychopathology. J Am Acad Child Adolesc Psychiatry *38:*232, 1999.

Steiner H, Garcia I, Matthews Z: PTSD in incarcerated juvenile delinquents. J Am Acad Child Adolesc Psychiatry *36:*357, 1997.

Steiner H, Williams SE, Benton-Hardy L, Kohler M, Duxbury E. Violent crime paths in incarcerated juveniles. In *Biosocial Bases of Violence,* Raine A, Brennan PA, Farrington DP, Mednick SA, editors. Plenum, New York, 1997.

*Steiner H, Cauffman E: Juvenile justice, delinquency and psychiatry. In Berkowitz SJ, Adnopoz J, editors: Child Adolesc Psychiatr Clin North Am *7:*653, 1998.

Tolan P, Guerra N: Prevention of delinquency: Current status and issues. Appl Prevent Psychol *3:*251, 1994.

Weinberger D, Gomes M: Changes in daily mood and self-restraint among undercontrolled preadolescents: A time-series analysis of “acting-out.” J Am Acad Child Adolesc Psychiatry *34:*1473, 1995.

Widom C: The cycle of violence. Science *244:*160, 1989.

Wolfgang M: *Delinquency in a Birth Cohort*. University of Chicago Press, Chicago, 1972.

49.8 DISSOCIATIVE DISORDERS IN CHILDREN AND ADOLESCENTS

NANCY L. HORNSTEIN, M.D.

In the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) the category dissociative disorders contains a group of diagnoses with dissociative symptoms as a core element of their presentation. The five dissociative disorders are dissociative amnesia, dissociative fugue, dissociative identity disorder (known in earlier DSM nomenclature as multiple personality disorder), depersonalization disorder, and dissociative disorder not otherwise specified. Grouping these five diagnoses and placing other diagnoses with dissociative symptoms in different diagnostic categories has perhaps generated more debate in the psychiatric literature than any other DSM-IV diagnostic category. Eating disorders, borderline personality disorder, and posttraumatic stress disorder are the diagnoses most frequently listed by authors who question the clinical validity of the current DSM classification system in terms of its ability to group together diagnoses with similar etiologic, pathological, symptomatic, and psychological/psychodynamic features or interrelatedness. Other authors look to include diagnoses listed in the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10). *The ICD-10 Classification of Mental and Behavioral Disorders,* published in 1992 includes a number of dissociative conversion disorders, and Ganser's syndrome among the dissociative disorders. Some also consider brief reactive psychosis occurring immediately after a trauma to be a psychotic reaction primarily dissociative in nature.

Clinical phenomenology of children who meet adult DSM-IV diagnostic criteria for one of the dissociative disorders often differs from that of adults with the same diagnosis; children's dissociative symptoms are less different than are adults. Other children with extensive dissociative symptoms are thought by authors doing research on dissociative disorders in children to “fall through the cracks” of the adult diagnostic criteria even though they appear to represent the childhood counterparts of adults with dissociative identity disorder. The critical clinical importance of identifying and treating children on their way to developing serious adult psychiatric illnesses led various authors to propose a separate DSM-IV diagnostic category for dissociative disorders in children that would emphasize the unique ways dissociative symptoms manifest themselves on children as compared with adults.

Close examination of the issues underpinning the debate over the diagnostic classification of dissociative disorders reveals roots with enormous clinical importance. Dissociative symptoms appear in situations of extreme stress, fear, or anxiety in individuals across a variety of diagnoses. Some of the diagnoses grouped together as dissociative disorders have little etiologic commonality, whereas some diagnoses grouped elsewhere are etiologically more closely related to an individual dissociative disorder. Many, but not all dissociative disorders occur following trauma. In some the trauma is a single overwhelming episode, others involve repeated traumas in early childhood. All dissociative defenses or symptoms aim to modify overwhelming emotion. Disorders in which the dissociative symptoms do not involve extraordinary child abuse or life-threatening disasters may nevertheless result from psychologically traumatic conflicts. Investigating the clinical and etiological similarities and differences of diagnoses with dissociative symptoms may clarify how these disorders should be grouped in the future.

Developmental issues affect the impact of a trauma on symptom development and presentation. A similar event has enormously different effects depending on the age of the child. Children's developmental capabilities influence their ability to handle a trauma emotionally, the type of event considered traumatic, and the degrees and types of impact a trauma has. The child's environment, available adult nurturing, soothing, and integrative influences affect developmental shaping, symptom modification along with the trauma resolution that occurs over time.

Repetitive trauma, neglect, and chronic deficiencies in parenting interact with a child's ongoing development in complex ways to produce symptoms and other psychopathology (e.g., object relations, coping strategies) that evolve with time. Extreme neglect in infancy causes the ultimate dissociative withdrawal, failure to thrive; less severe deficiencies in infant nurturing and bonding result in developmental attachment disorders. Mary Main and Hillary Morgan described a category of attachment disturbance called the *insecure-disorganized/disoriented attachment category* and noted the phenotypic resemblance of these children's behavior to that typical of dissociative states. Thus, a set of dissociative symptoms in a child affects the child's development, and the child's ongoing development and environmental influences then interact with the need for dissociative defenses, the way these are expressed as dissociative symptoms, and ultimately the presentation of diagnostic pathology. Childhood dissociative disorders must be viewed as occurring along a developmental, symptomatic, and diagnostic continuum.

Further study of dissociation in young children aims to integrate the effects of trauma, neglect, and extreme parental dysfunction to understand their developmental effects over time and thus to better understand both childhood and adult psychopathology. Such increased understanding holds forth better therapeutic interventions to ameliorate the effect of trauma on an identified individual child or adult.

The research and theoretical literature on dissociative disorders in children and adolescents focuses on the DSM-IV diagnoses of dissociative disorder not otherwise specified and dissociative identity disorder. This literature addresses both children who meet the current adult DSM-IV diagnostic criteria for these two diagnoses and children with symptomatic presentations involving significant dissociative pathology thought to be on a continuum of disturbance with these two diagnoses.

HISTORY

A number of cultural changes, “discovery” of numerous victims of sexual abuse after such reports were considered fantasy by early Freudian analysts, and the new openness to investigating these areas resulting from the sexual revolution of the 1960s are a few of the factors that led to recognition of contemporary cases of multiple personality disorder. These cases led to a flurry of recognition of more adult cases after a hiatus of nearly 100 years. Professional skepticism about these case reports, difficulties encountered treating these victims of extreme early child abuse, and the need to develop new theoretical constructs to understand dissociative symptoms led early investigators back to the writings of Jean Charcot and Pierre Janet at the Salpetrière about their work with patients with hysteria. Janet clearly described patients with multiple personality disorder, and reviews of his work and that of his contemporaries helped investigators greatly in their growing theoretical and clinical understanding of multiple personality disorder.

A search for the childhood counterparts of adult patients with multiple personality disorder provided early impetus for research into childhood dissociative disorders, since virtually all adult patients reported that their symptoms began in childhood. This search was motivated by the desire to (1) identify children suffering from extreme abuse and intervening to protect them in the hope that childhood intervention could prevent further suffering from abuse and reduce later symptomatic suffering and (2) the wish to put to rest the taunts of skeptics attacking the multiple personality disorder diagnosis who were asking the question, “If multiple personality disorder starts in childhood, where are the children with this illness?”

Richard Kluft led these efforts by reporting several patients with multiple personality disorder, and he and others developed symptom predictor lists to help clinicians identify such children. He also included description of multiple personality disorder in a child treated by Despine, published in 1940 in his numerous lectures. Kluft's case series was blocked from publication for 6 years, and when finally published, it coincided with a similar case series published by J. Fagan and P. McMahon. Other small case series quickly appeared in the literature, and the interested clinician or researcher had in effect a translation from the adult phenomenological presentation of multiple personality disorder into the ways this disorder would present in childhood. This translation was critical to the field, since it was conceptually difficult to bridge the gap between the adult patients' ability to make their dissociative experiences understood via their developmental sophisticated use or language and abstract symbolization, which children lack. It was necessary to develop ways to understand and assist a child's efforts to describe dissociative experiences without the conceptual capacities of later development.

Some initial reports described children who would have met the diagnostic criteria for dissociative disorder of childhood but not for multiple personality disorders. These authors highlighted the ways in which these children's dissociative problems fit on a developmental continuum with adult multiple personality disorder. A large

clinical series by the author and Frank Putnam further elucidated the clinical similarities of children with multiple personality disorder and those with dissociative disorder not otherwise specified. Several researchers have recently developed sophisticated diagnostic tools to achieve an accurate diagnosis in children with dissociative symptoms.

Researchers studying dissociative disorders in children have shifted the focus in multiple personality disorder research from the drama of adult dissociative identity disorder to the varied developmental effects of repetitive early childhood trauma expression.

EPIDEMIOLOGY

An accurate estimate of the frequency of dissociative disorders in childhood is not possible. They are certainly not as rare as once thought. Urban child psychiatric clinics and child inpatient psychiatric units that evaluate and treat a number of children often have several documented and suspected cases at any given time. In populations of children in protective custody because of abuse, in the juvenile justice system, and in child residential facilities, the incidence increases, particularly when diagnostic screening is used to identify children with dissociative symptoms.

Although the adult literature on dissociative identity disorder states that females outnumber males three to one, the child literature reports a more even distribution. Some authors cite the tendency for women to internalize symptoms while men tend to externalize and act out to explain this female to male ratio. Other researchers have reported an unexpectedly high incidence of dissociative symptoms in men incarcerated in the penal system.

A variety of factors confound research on childhood dissociative disorders. The adult dissociative identity disorder patients who reported childhood onset of their symptoms generally did not come to the attention of child psychiatrists. Most maintained the minimum level of function needed to avoid being targeted for intervention. Hearing some of their case histories one wonders in retrospect how their illnesses could have been missed. An honest appraisal of children who slip through the cracks today is equally shocking. Our current economic structure does not encourage schools to identify children needing treatment unless they present severe behavioral management problems in the classroom; new laws require schools to pay for mental health evaluations of children they identify as needing help. Children in families with ongoing abuse are often carefully schooled not to draw attention to themselves, and their parents do not seek psychiatric attention for fear that the abuse will be discovered—with major legal ramifications.

The children who are diagnosed as having dissociative disorders are likely an atypical subset of children with these disorders. Children come to the attention of mental health professionals most frequently when some aspect of their behavior is problematic to adults. Those who disrupt classrooms, exhibit explosive, aggressive, violent behavior; commit crimes; injure themselves; attempt or threaten suicide; or are so depressed that they disturb authorities are the ones seen. Children whose families are so dysfunctional that they fail to hide abuse or disintegrate so that the child is removed to protective custody may come to the attention of mental health professionals. In overloaded urban welfare systems the usual course of mental health treatment is often too little, too late, only for the most severely dysfunctional.

Comprehensive psychiatric evaluation and treatment are increasingly unavailable for the average American family, whose insurance company simply will not pay for them. Fewer adoptive children with symptoms secondary to abuse prior to adoption from families with middle incomes are treated.

These factors skew research. The equal male to female ratio in childhood dissociative disorder may only mean that more boys are brought into the mental health system because they are more apt to exhibit externalizing symptoms that result in behavioral management problems. An important question awaiting future research is whether aggressive outbursts, self-injury, suicidal gestures, inappropriate sexual behaviors, and antisocial actions are inherent to the dissociative disorders in children, or simply reflect the kinds of disturbances that bring youngsters with dissociative disorders to psychiatric attention.

Research suggests an increased incidence of dissociative disorders in children when a parent carries the diagnosis of dissociative identity disorder. It is not known what roles a genetic tendency or capacity for dissociation, identification with the dissociating parent, or dysfunctional parenting because of dissociative identity disorder symptoms play in this increased incidence. Some family case studies that show the difficulties dissociative identity disorder patients encounter in parenting indicate areas that should be explored for therapeutic intervention while the reason for the increased family frequency is being investigated.

ETIOLOGY

Overwhelming evidence from the literature on dissociative identity disorder in adults and even more powerfully from the literature on dissociative disorders in childhood points to ongoing, severe trauma, sexual abuse, and violence among family members as causal in the development of these disorders. The typical case history details repetitive, exploitative, violent, and merciless physical and sexual abuse beginning before the age of 5. In many childhood cases this abuse is well documented; these cases include children who have been prostituted for drugs, those used to produce pornographic material, and those exposed to sadomasochistic torture. Others live in chaotic, violent households or are exposed to abuse by a series of dysfunctional adults brought into the home by a highly dysfunctional parent. Cases in which violence or sexual abuse began later in childhood are less frequent; the abuse usually began well before adolescence, and predisposing family chaos appears to have led to development of a childhood dissociative disorder.

In rare instances preschool children exposed to the severe upheaval of war, loss of family members, grotesque war trauma or the uprooting that is part of being a refugee have also been given diagnoses of dissociative identity disorder. That these diagnoses are not more common in these populations speaks to the resiliency of children and their ability to use a variety of responsive caregivers to help them achieve integrative self-functions, even in the presence of severe trauma and resultant posttraumatic stress disorder. These children may have the psychogenic amnesias, traumatic flashbacks, and dissociative phenomenon common in posttraumatic stress disorder, but the infrequency with which they develop separate self-states indicates causal factors in the development of dissociative identity disorder that go beyond severe traumatic experiences in early childhood.

Evaluating young children in the midst of war trauma, immediately after natural disasters in which injury occurred or parents were lost, or after some other extreme single trauma, such as witnessing a parent's murder, yields a number of children suffering from acute dissociative symptoms. Trance states, episodes of regression, traumatic reenactment, and traumatic flashbacks are not uncommon. Some may have persistent dissociative symptoms and may exhibit some of the behaviors listed under B and C of the diagnostic criteria for dissociative disorder of childhood. Few ever meet the full diagnostic criteria.

Dissociative identity disorder and dissociative disorder of childhood appear to have at their core not only extraordinary trauma in early childhood, but trauma that occurs at the hands of family members or as a result of extreme parental dysfunction that allows such trauma to occur repeatedly at the hands of others. This trauma is combined with disturbances in the parent's ability to form a healthy attachment to the child and the further absence of adults to help perform the integrative and self-soothing functions that cannot be performed developmentally by the young child. Deficiencies in normal parental attachment and a lack of caregivers to help the young child develop normal self-soothing and integrative functions in early childhood appear to be necessary for dissociative identity disorder or dissociative disorder of childhood to develop in response to childhood trauma occurring after the age of 6. The dynamic conflicts inherent in knowing that the parent who is depended on is also the source of trauma are inversely related to the age and helplessness of the child. The development of separate self-other internalization avoids the experience of overwhelming anxiety. Development of an integrated identity is blocked by the traumatic anxiety it would unleash, and the self-object internalizations evolve into separate personality states.

Inborn developmental abnormalities, inborn or acquired neurological problems, or other organic factors that influence a child's ability to form an attachment, be soothed, or block the development of self-integration may cause children with other diagnoses to present with dissociative symptoms. Inborn difficulties in these children interfere with normal development and may transform some aspects of commonplace occurrences into overwhelming traumas.

A developmental capacity to dissociate is likely neither sufficient nor necessary to cause development of dissociative identity disorder. Research into the capacity to dissociate appears in the literature on hypnotizability. Individual variations in hypnotizability exist, some of which may be inherited. This research also shows that hypnotizability, (correlated with the capacity to dissociate) peaks in early childhood and declines through adolescence. The persistence of high hypnotizability in college students was correlated with corporal punishment in childhood, which suggests that the dissociative capacities of young children can be retained as a result of environmental influences. An inherited capacity to dissociate may provide a protective function for a child exposed to repeated overwhelming trauma rather than represent a genetic vulnerability to develop dissociative identity disorder.

Studies on the neurobiology of trauma indicate that traumatic experiences change neurophysiological functioning and the ways in which memories are stored and retrieved. Dissociation in the face of a posttraumatic stimulus helps prevent dysfunctional panic/anxiety states but may cause traumatic stimuli to maintain their

potency over time. These areas of research promise to further our understanding of the etiology of dissociative disorders.

DIAGNOSIS

According to DSM-IV, “[t]he essential feature of the Dissociative Disorders is a disruption in the normal integrative functions of consciousness, memory, identity or perception of the environment.”

J. was fishing as he often did to help feed his family. Although he had just turned 10, he played a significant role in his family's survival, and fishing was one of the more enjoyable ways he helped. His favorite fishing hole was a short hike through the jungle from his family's village. Suddenly, he heard crashing and shouts of men running through the jungle. He saw a young man desperately run into the water, but not quickly enough to escape a hail of gunfire. He was hit several times in the back and fell, bleeding copiously, rolling as he fell so that J. recognized him as a cousin who was only 5 years older than he was—a boy he sometimes fished with, and one he admired for his stories and his wisdom and popularity in the village. As these events unfolded J. felt as though it all must be a dream. He experienced himself watching the event from high in the trees, far above his own body at risk by the water's edge. This experience lasted for the seconds it took for the young man and his pursuers to appear, the shots to ring out, and the young man to fall, revealing his identity to J. Quickly J. was back in his body diving into a safe hiding place. He had no idea how long he hid in the bushes by the water until long after the men with guns disappeared. He has a foggy image of trying to revive his cousin and realizing he was dead. He trembled as he remembered returning to tell his grandmother, but finding no words, being able only to gesture and lead his uncle to the site. He told this story in English because as he stated, “I've never been able to remember how to speak the language of my village ever since then.” His family escaped the country shortly after this incident, other family members were lost, and there was much hardship in their journey to the United States. J. was 16 when he told this story.

An important, but often difficult distinction is the dividing line between a dissociative defense, a dissociative symptom, and a dissociative disorder.

The story of J. meets the DSM-IV definition of dissociation three times. The first dissociative experience is his “out of body experience,” or brief depersonalization—it is fleeting, occurs at the moment of his greatest fear, and quickly disappears. This is a good example of a dissociative defense thats appears to help a person cope at a moment of extreme fear and stress. The common appearance of this defense in response to trauma is reflected in colloquial phrases such as “I was beside myself with terror,” “I was scared out of my wits,” “My heart leaped out of my chest at that moment,” and “I felt as though the entire experience was unreal.”

J.'s second dissociative experience was the distortion in his sense of time—how long he hid. To this day he has no idea how long he hid. He says it could have been minutes or hours. He also can't recall where the sun was in the sky, how long he had been fishing before the trauma occurred, or where the sun was in the sky when he went back to the village. He knows it was early morning when he set out fishing and that his uncle and others lit torches to find their way back to the village after he led them to his cousin. This persistent distortion in memory and perception represents a dissociative symptom relating to the trauma, but it does not affect J.'s current life enough for a diagnosis of a dissociative disorder. Alteration in time perception is a common enough symptom to generate the colloquial phrase “At that [traumatic] moment time stood still.”

The third dissociative experience, persistent amnesia for his native tongue, has a clear impact on his relationships with family and his ethnic community. He understands words spoken to him in his native tongue but needs a translator to translate from the English he has learned back into his native language. This symptom grants him the diagnosis of dissociative amnesia. Brief traumatic loss of speech is acknowledged in the everyday phrases “scared speechless” and “the cat got my tongue.”

Other common colloquial descriptions of dissociative responses to trauma culturally organize the strange alterations in integrative functions that occur. States of posttraumatic numbness, derealization, depersonalization, and trance (eyes staring blankly ahead, oblivious to surroundings) are called “shell shock,” even when they occur in response to psychologically overwhelming news rather than the traumatic experience of war. “When he first got off the phone having learned his wife was having an affair he was absolutely shell shocked and it took him some minutes to overcome a sense of numbness where he could neither feel nor react.” Dissociative conversion is a dissociative disorder recognized by ICD-10 and in the expression “my legs gave out under me when I heard the news.” Dissociative alterations in identity are known also in their mildest form as a “normal reaction to trauma” in describing a person no longer feeling like themselves or the entire world turning upside down in an instant.

Nonpathological dissociative behaviors are part of our normal human repertoire; an example is being so lost in thought you drive miles past your exit before your consciousness of your surroundings reemerges. Another example involves staring off in a glassy-eyed daydream (in the elevator perhaps), prompting friends to wave their hands in your face and say, “Earth to Mars.”

Developmental Considerations The diagnosis of dissociation in childhood in terms of the DSM-IV definition of a dissociative disorder is complicated by developmental progressions in perceptions of time, identity, and adult reality. Knowing how child development affects conflict resolution is important in assessing whether a child is developmentally on track or significantly delayed in giving up some normative dissociative defenses. Not knowing how long an hour takes to pass is expected in a 6-year-old child but may be a dissociative symptom in a 10-year-old. Young children may not believe they are lying when they insist they did not do something you just saw them do. They can believe it did not happen because they wish it had not, reflecting developmental progressions integrating reality with emotionally satisfying fantasy life.

L. and his brother were roughhousing with their aunt. L., age 5, overly zealous in his desire for physical conquest over the others, suddenly popped his aunt in the nose. The aunt pulled herself free of the boys and went into the bathroom to put ice water on her nose. When she returned to face the raucous L. and explained that he had really hurt her when he hit her in the nose, L. yelled out, “I didn't hit you, you hit me.” L.'s astonished aunt protested his assertion, whereupon he stormed off and locked himself in the bedroom. After a little time, L.'s aunt went to the bedroom door and explained that everybody makes mistakes sometimes and whatever happened she was sure they could have some safe fun together making some artwork. L. emerged and told his aunt it was O.K., they could be friends again. For 5-year-old L., the thought that he had hurt his aunt was overwhelming. In telling him so she “hit” his feelings. He revised reality to go along with his feelings, and he was certain he was telling the truth. He had to flee his aunt's attempts to intrude on this reality because it was too upsetting. He could only come out after he was reassured of their friendship, although she had not behaved angrily or been rejecting. A discussion of lying had no meaning to him at his stage of development.

The knowledge that wishes don't alter reality grows in the early grade school years, as does constancy in the perception of self and others. A 3-year-old child can have the worst mommy in the whole world when she insists the child put on boots to go outside and the best mommy in the world 5 minutes later when she fixes a broken toy. If you remind the 3-year-old of saying the best mommy was the worst 5 minutes before, the child is genuinely confused. Similarly, being scolded as a bad child can feel like the end of the world without hope of ever recovering former good graces without adult assistance. Such fluctuation in object relations in an older child could contribute to the development of borderline personality disorder development with its dissociative shifts between all good and all bad self and other perceptions. Examples of psychologically motivated dissociative confusion between reality and fantasy abound in clinical experience with individuals with dissociative disorders. It is helpful to be cognizant of the age when this is a normative response to internal conflict when devising a therapeutic response.

Z. had an intense emotional relationship with her demanding, self-involved boss. He could do no wrong in her eyes. Her life revolved around his whims and tantrums; she gloried in his praise and never questioned his expectations or her neglect of her own and her family's needs. When her boss got into trouble with the IRS he placed the blame on Z., who in fact had no role in preparing any aspect of his tax reports. The boss fired Z. to “look better.” Z was disconsolate. She was sure there was some mistake and tried to think of ways to force her boss to realize this. Eventually, as her efforts for contact were rebuffed, her dissociate symptoms (which though present, had never brought her to professional attention) worsened, and she was given the diagnosis of dissociative identity disorder. Z. had frequent episodes of amnesia and found herself with items she had no recollection of owning and saw no need for—clothing she saw as frivolous, expensive stuffed animals, male clothing in her size. She found herself in police custody for breaking into her boss's office and stealing files. She could not believe she had done this, although she was told police discovered her at the scene of the crime with the files in her hands. In therapy an alternate personality emerged who identified “himself” as Z.'s teenage friend X. X. and Z. saw themselves as completely distinct individuals. X. denied having female physical attributes. X. explained that he was angry with the boss and that he was trying to get proof that Z. was not involved in the IRS problems. Z. could not in any way relate to X.'s feelings, opinions, or actions. Discovering X.'s “origins” in Z.'s childhood (to fend off sexual abuse by Z.'s father) allowed these two dissociated aspects of Z.'s identity to acknowledge their relationship to each other and the impossible conflict Z. had in coping with her idealized father's sexual abuse of her. Critical to achieving this therapeutic movement toward integration was the therapist's comment that Z. could acknowledge her father's sexual abuse and still admire other things about him. The child's difficulty in integrating good and bad when they are extreme played an important role in Z.'s development of dissociative identity disorder in childhood and its symptomatic worsening when a conflict arose about exploitation by an idealized male in Z.'s adult life.

Imaginary companions are developmentally unremarkable in a 6- or 7-year-old child even though they represent split-off aspects of self. In an adolescent, an imaginary companion may indicate a dissociative identity disturbance. One simply cannot apply adult criteria of alterations in integrative functions of memory, consciousness, identity, and perception to help identify a dissociative disorder in a child of 3, or 6, or 8. This definition has validity only as an adolescent moves closer to adulthood and the clinical phenomenology of a childhood dissociative disorder comes to mirror that seen in adults with these disorders. Dissociative symptoms in childhood represent developmental lags in integrative functioning relative to the child's age when traumatically derived difficulties obtaining or maintaining age-appropriate integrative functions manifest in the face of developmental demands.

The DSM-IV discussion of dissociative identity disturbance includes information not present in the revised third edition of DSM (DSM-III-R) that acknowledges the research findings on dissociative disorders in children and the ways their presentation differs from that of adults, which were published in the years between these two editions. “In preadolescent children, particular care is needed in making the diagnosis because the manifestations may be less distinctive than in adolescents and adults” and “in childhood, the female-to-male ratio may be more even.” Table 49.8-1 lists the DSM-IV criteria for dissociative identity disturbance, the organizing principles behind each criterion, and important considerations for applying the criteria to preadolescent children.

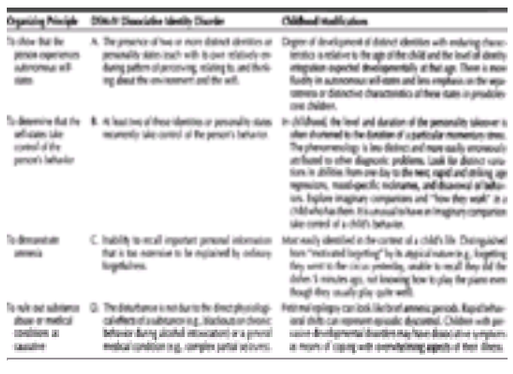

Organizing Principle	DSM-IV Dissociative Identity Disorder	Childhood Modifications
To show that the person experiences autonomous self-states	A. The presence of two or more distinct identities or personality states (each with its own relatively enduring pattern of perceiving, relating to, and thinking about the environment and the self).	Degree of development of distinct identities with enduring characteristics is relative to the age of the child and the level of identity integration expected developmentally at that age. There is more fluidity in autonomous self-states and less emphasis on the organization or distinctive characteristics of these states in preadolescent children.
To demonstrate that the self-states take control of the person's behavior	B. At least two of these identities or personality states recurrently take control of the person's behavior.	In childhood, the level and duration of the personality takeover is often circumscribed to the duration of a particular momentous stress. The phenomenology is less distinct and more readily erroneously attributed to other diagnostic problems. Look for distinct variations in abilities from one day to the next, rapid and striking age regressions, mood-specific incitations, and dissociative behaviors. Explore imaginary companions and “how they work” in a child who has them. It is unusual to have an imaginary companion take control of a child's behavior.
To demonstrate amnesia	C. Inability to recall important personal information that is too extensive to be explained by ordinary forgetfulness.	Most easily identified in the context of a child's life. Distinguished from “motivated forgetting” by its atypical nature (e.g., forgetting they went to the circus yesterday, unable to recall they did the chicken 5 minutes ago, not knowing how to play the piano even though they usually play quite well).
To rule out substance abuse or medical conditions as causative	D. The disturbance is not due to the direct physiological effects of a substance (e.g., blackouts or chaotic behavior during alcohol intoxication) or a general medical condition (e.g., complex partial seizures).	Petit mal epilepsy can look like brief amnesia episodes. Rapid behavioral shifts can represent episodic dissociation. Children with post-traumatic developmental disorders may have dissociative symptoms as means of coping with overwhelming aspects of their illness.

Table 49.8-1 Modifications to DSM-IV Dissociative Identity Disorder Diagnoses to Account for Childhood Symptoms

The DSM-IV diagnosis of dissociative disorder not otherwise specified is given to dissociative individuals who do not fit criteria for other dissociative disorder diagnoses. The diagnostic criterion reads as follows: “Disorders in which the predominant feature is a dissociative symptom (i.e., a disturbance in the normally integrative functions of identity, memory, or consciousness) that does not meet the criteria for a specific dissociative disorder.”

Children with dissociative disorders who do not meet the criteria for a diagnosis of dissociative identity disorder currently fall into this category. Researchers consider this inadequate for two reasons. First, often the failure to meet criteria for dissociative identity disorder is a function of the child's age and developmental level. Given the level of personality organization in the preschool-age child, a child of this age can rarely be diagnosed with dissociative identity disorder despite a complex syndrome of dissociative symptoms and difficulty developing integration of self-states. Second, many researchers argue that a meaningful constellation of dissociative symptoms occurs in children with histories of extreme trauma and abuse that although not organized well enough to meet the criteria for dissociative identity disorder, nevertheless deserves attention. Many consider these children to be at risk for developing dissociative identity disorder as they get older and some have documented cases of children with dissociative disorder not otherwise specified who meet the criteria for dissociative identity disorder later in childhood. Gary Peterson and Frank Putnum organized the symptomatic constellation of children diagnosed with dissociative disorder not otherwise specified from their own and others' research and coined the diagnostic concept “dissociative disorder of childhood.” Although not included in DSM-IV, most in this field hope to have this diagnosis included in later editions of DSM. The proposed diagnostic criteria for dissociative disorder of childhood are listed in Table 49.8-2.

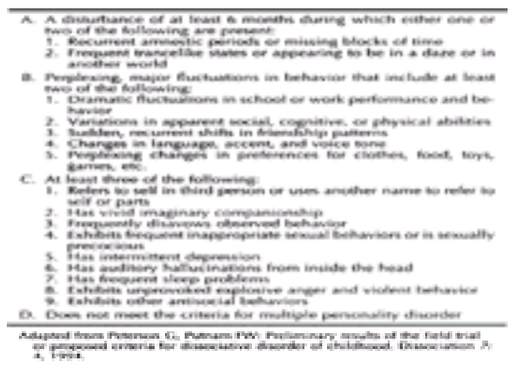

A. A disturbance of at least 6 months during which either one or two of the following are present:
   1. Recurrent amnestic periods or missing blocks of time
   2. Frequent trancelike states or appearing to be in a daze or in another world
B. Perplexing, major fluctuations in behavior that include at least two of the following:
   1. Dramatic fluctuations in school or work performance and behavior
   2. Variations in apparent social, cognitive, or physical abilities
   3. Sudden, recurrent shifts in friendship patterns
   4. Changes in language, accent, and voice tone
   5. Perplexing changes in preferences for clothes, food, toys, games, etc.
C. At least three of the following:
   1. Refers to self in third person or uses another name to refer to self or parts
   2. Has vivid imaginary companionship
   3. Frequently disavows observed behavior
   4. Exhibits frequent inappropriate sexual behaviors or is sexually precocious
   5. Has intermittent depression
   6. Has auditory hallucinations from inside the head
   7. Has frequent sleep problems
   8. Exhibits unprovoked explosive anger and violent behavior
   9. Exhibits other antisocial behaviors
D. Does not meet the criteria for multiple personality disorder

Adapted from Peterson G, Putnam FW. Preliminary results of the field trial of proposed criteria for dissociative disorder of childhood. Dissociation 7: 4, 1994.

Table 49.8-2 Proposed Diagnostic Criteria for Dissociative Disorder of Childhood

The criteria in Table 49.8-2 perform several functions important to the clinician evaluating a child with dissociative symptoms. Criterion A identifies a child who has significant dissociative pathology; neither recurrent amnesia nor frequent trancelike states are normal at any developmental age. Criterion B is sensitive to difficulties in a child's development of integrative functions. Criterion C refers to symptoms and behaviors frequently representing dissociative phenomenon that in the absence of a dissociative disorder might implicate other problems or diagnoses. Multiple personality disorder, mentioned in criterion D, was the DSM name for dissociative identity disorder prior to DSM-IV. The aim of this new diagnostic category is to identify children with a dissociative syndrome that is on a developmental continuum with dissociative identity disorder. This aim continues the historical aim that fueled the early development of research on childhood dissociative disorders. The opportunity to prevent further trauma, intervene, and treat the disorder early enough to prevent a serious adult mental illness and to develop further understanding of the impact of childhood trauma is a powerful motivator of research in this field.

CLINICAL FEATURES

Several instruments are now available to assess dissociative symptoms in children. They are enormously helpful for systematic evaluation of children suspected of

having dissociative disorders or for determing the extent of the dissociative pathology present. Some are brief enough to be used as screening tools in populations of children at risk for these disorders. In an ideal world this type of screening would be a part of the intake evaluation of a child in the child welfare or juvenile justice system and would be conducted during any investigation of child abuse. They are invaluable aids for clinicians seeking to gain experience interviewing children about their dissociative experiences, since the questions are to a child's developmental level.

Interviewing children about dissociative experiences is initially difficult. It requires above-average attunement to how children conceptualize their experiences. Children younger than preadolescents have difficulty with abstract concepts, early-school-age children have no sense of time as an organizing factor for placing life events, and few children have any experience describing any aspect of their sense of self. How does one begin asking about amnestic episodes for example? Anchoring inquiry in the events of the child's daily life is useful and examples are helpful. Often the clinician is modeling for the first time a way to discuss different types of experiences.

> *Sometimes kids have things happen to them that seem kind of strange, I'm wondering if any of these kinds of things ever happened to you? One kid told me that sometimes his teacher was doing one thing, like math problems, and then all of a sudden she was asking him to read aloud in his book. All the other kids knew where they were in the book, but he didn't even know he was supposed to have his book out, or what happened to the math problems. Did you ever have anything like that happen? What? Tell me about it*. Be aware of some children's natural tendency to want to answer these new questions—especially if they like you and want to please you—by acknowledging similar experiences. Always have them describe the experience before you accept it. A yes responder says, “Oh, yeah, I got in trouble for that yesterday.” *What happened?* “I didn't know where we were supposed to be in the reading book.” *Were you fooling around or anything that got you lost?* “Yeeees, well Bobby poked me with his pencil and then I was trying to kick his foot under the chair.” *Yeah, it's easy to get lost with that stuff going on*.
>
> The dissociative disorder child (same question): “Well no, um, maybe yes.” *Tell me a time you think might be like that*. “Uh, everybody got up for recess, but I didn't know, and then the teacher was kind of yelling my name a few times, why wasn't I getting in line. I didn't know, I thought we were still doing social studies.” *Were you fooling around with anything so you didn't know?* “I think, daydreaming?” *What were you dreaming about?* “I don't know, sometimes, I just go away somewhere.” *Where do you go?* “Um, well someone tells me I go to Hawaii.” *While you're still in your seat in class? Who tells you that?* “Boney” *Who is Boney?* “Some guy.” *Where is he from?* “Hawaii, except sometimes he lives in my head.” *How do you know he's in your head?* “He talks to me.” *Does Boney ever do things besides talk?* “In my head?” *Out of your head, in the world*. “How did you know that? He does, people blame me for it, but I didn't do it. They say they saw me but it wasn't, it was Boney.” *How does that work? How can people see you do it but really it's Boney?* “Boney just gives me a strong feeling and I have to shut my eyes and then he comes over me, out of my mouth and nose and eyes. Then I'm walking and my hands are doing things but I don't know what they're doing, because it isn't me, it's Boney.”

These interview excerpts are essentially verbatim. They illustrate interviews regarding dissociation with children who are still in the cognitive stage of concrete operations and give an example of a child's experience of having an alternate personality take over his behavior. Besides the concrete nature of their descriptions, children frequently have not elaborated their dissociative experiences in the ways adults do; this child readily acknowledges it is his body that did the action, which is not unusual. In late adolescence and adulthood, the alter is described as having a separate body. As cognitive abilities develop further and a child's sense of identity becomes more fully established and developed, the individual with dissociative identity disorder needs greater distinctness from the alternate personality states to avoid threatening anxiety. The child with dissociative disorder of childhood with fluid, transient self-states becomes the child with dissociative identity disorder as greater organization is possible developmentally or needed developmentally to fend off anxiety. Children have little opportunity to elaborate and express “separateness” in dissociated self-states; the clinician must look for more subtle manifestations of split-off, dissociated personality states.

Children with dissociative disorders are often amnestic to emotionally laden events. Schoolyard fights, punishments, explosive outbursts, and separations from caregivers trigger dissociative episodes. Older children may confabulate memories for these or lie about remembering because they know they are expected to remember and have often been punished for lying when they earlier told their accusers they did not remember committing some infraction. Patient inquiry into the details of the experience often helps reveal the amnesia the child is covering up, and some children who realize that you are interested in their experience of what happened rather than seeking a confession and remorse tentatively open the door to disclosing more about their dissociative experiences.

Children who have dissociated aspects of self, even when these are not elaborated into separate identities, lack integration across spheres. Knowledge, skills, relationship preferences, clothing, toy and food preferences, personality style, voice and speech patterns, and accents may vary between these dissociated aspects of self. Observational data in a variety of settings are thus important for differential diagnosis. The lack of integration between these dissociated aspects of self in children with either dissociative identity disorder or dissociative disorder of childhood goes beyond defensive imaginary constructs. Psychological testing may show wide fluctuations in the same skill from day to day. Hearing tests and eye examinations can also yield different results for different personalities or alters.

DIFFERENTIAL DIAGNOSIS

Differential diagnosis in children with dissociative disorders is complicated both by dissociative symptoms mimicking those of other diagnoses and by the variety of symptomatic disturbances brought about by trauma. Difficulties in affect regulation produce symptoms of irritability and affect lability. The psychodynamic legacy of child abuse includes fragile self-esteem, rejection hypersensitivity, self-blame, and moods that plummet into helpless, hopeless dysphoria at the slightest alteration in the caregiver's responsiveness. Some exhibit identifiable alternate personalities who blame themselves for the abuse and are chronically suicidal and intensely depressed. Small events that are linked in some way to past trauma can trigger the emergence of these personalities bent on self-destruction.

> For example, anytime M., age 14, wonders why I seem to care about her, this thought triggers a belief that I could not possibly care if her parents did not, and M. falls into a deep suicidal depression. Movies about child abuse also trigger M.'s depressed states even when she only reads a review. If M. admires some accomplishment or attribute in another, she is certain she's inadequate and forever doomed to rejection. She recognizes that she can be demanding but cannot recognize the traits that cause others to like her despite her problems. It's difficult to progress in therapy when efforts to discuss problem areas trigger violent self-retribution for having problems.

Not surprisingly, children, like adults diagnosed with dissociative disorders, frequently have held previous diagnoses of depression or bipolar disorder and show little or no response to antidepressant medication; moods that may be down or euphoric for weeks shift to the opposite in a matter of seconds in response to external events, a new relationship, or special recognition and change depression into giddiness. Conduct and behavioral problems are common in children diagnosed with dissociative identity disorder and dissociative disorder of childhood. These children (and adults with this disorder) often have no history of behavioral problems before they were removed from the abusive situation. In part this is related to their failure to develop internal controls for behavior. Survival meant tight responsive attunement to abusive caretakers. Similarly, adults with dissociative identity disorder become more symptomatic in good relationships. Out-of-control behavior is often managed initially through intensive therapeutic focus on establishing boundaries, a type of nonabusive external control that allows time for internal controls to develop.

An abused child often has difficulty understanding others' intentions and actions and may respond violently to perceived threats, have low frustration tolerance, or react strongly to manage feelings of helplessness. Children with dissociative disorders are frequently accused of lying because of their amnestic episodes and may have separate personality states that suddenly emerge to handle a situation through violence. Identification with the aggressor often plays a role in the dynamics of aggressive symptoms. Diagnoses of episodic dyscontrol are at times made erroneously because of the sudden, at times inexplicable, temper outburst. These children do not have the classic triad of enuresis, cruelty to animals, and fire setting associated with conduct disorder children on their way to developing antisocial personality disorder.

Because of dissociative symptoms, children with dissociative indentity disorder and dissociative disorder of childhood often have difficulty regulating attention and concentration as they drift in and out of dissociative states. Many have comorbid diagnoses of posttraumatic stress disorder. The combination of inattention, posttraumatic hypervigilance, and increased startle reactions often lead to misdiagnoses of attention-deficit/hyperactivity disorder. Dissociative states may also appear to be partial complex seizures. In the presence of dissociation, do not attempt to treat “presumptive seizures” without an electroencephalogram (EEG) showing epileptiform waves.

Dissociative problems with maintaining focus and dissociative “scrambling” of incoming information, along with fluctuations in knowledge and abilities associated with emergence of disparate personality states mimic a variety of learning disorders. For example, it was not unusual for the dissociative youngsters on the author's

inpatient unit to test as having auditory processing problems that fluctuated from day to day.

The traumatic experiences and the instability of relationships that produce dissociative identity disorder and dissociative disorder of childhood also produce anxious, traumatized children. Constant anxiety is a hallmark of these conditions. This can be relationship anxiety, where nothing can be taken for granted from one moment to the next; marked anxiety at bedtime with difficulty going to sleep—never knowing what might happen (often the children's abuse occurred in the night), or anxiety about certain activities reminiscent of emotionally difficult or traumatic experiences. Panic reactions are common as is invasive, traumatic imagery. Self-soothing abilities are often virtually absent.

> An 11-year-old boy with dissociative identity disorder experienced severe bowel impaction as a result of avoiding use of the bathroom. When he did have to enter the bathroom he often emerged in a confused, disorganized state. At times he was discovered huddling in the corner of a stall. It eventually came out that his father had repeatedly and violently sodomized him and he had scarring and resultant pain with defecation that triggered severe flashbacks to his abuse.

Dissociative phenomenon can also appear as psychotic states. Extreme anxiety, recent trauma, or times when their minds are flooded with traumatic material and flashbacks can cause these children to switch rapidly between different self-states, causing disorganization or catatoniclike withdrawal into a trance state. Unlike psychotic illness, fluctuations occur between these states and better organization, and stresses can be identified that trigger these states. Traumatic relationship anxiety can also occur in these children, producing a remote, glassy appearance that can mimic flat affect but is in fact dissociative regulation of extreme affect. Alternate personality states are frequently experienced as “voices” in the head. At times the hallucinations can appear to be coming from outside and be accompanied by visual hallucinations of abusers or alter personalities. One clear distinction between dissociative and psychotic disorders is the rapid response to establishing a safe, soothing environment. The key to successful differential diagnosis lies in observing children for behaviors that “don't fit” into the present diagnostic paradigm, seeking to understand the underpinnings of symptoms and behaviors, and keeping in mind that extreme measures are called for in extreme circumstances; thus bizarre dissociative phenomena may make perfect sense when they are seen in the context of the experiences that produced them.

COURSE AND PROGNOSIS

The literature is optimistic about treatment of dissociative disorders in childhood. The greater fluidity of the dissociative symptoms and the absence of more-solidified defensive organization makes therapeutic intervention more successful and quicker and often produces better mental health than is possible for adults with dissociative identity disorder. Follow-up at new developmental stages is recommended, as they may reveal conflicts that could not be resolved or were not active at the developmental stage when treatment initially occurred. Without treatment, some children with dissociative disorder of childhood will go on to develop dissociative identity disorder. Perhaps many will, but this is not an ethical experiment. There are documented instances of children with dissociative disorder of childhood developing more-classic dissociative identity disorder during treatment, as new developmental milestones are reached. Children with dissociative identity disorder are not likely to experience resolution of the disorder without treatment.

TREATMENT

Successful treatment involves first securing a safe, nurturing environment for the child. Without this, they will need their dissociative defenses and any therapy should be limited to supportive interaction. Safety and nurturing are needed at their highest level. These children need an environment of emotional involvement with them and if at all possible, love. The dysfunctional object relationship experiences of these children are combined in equal measure with trauma to create disturbances in their integrative functions of memory, perception, self, consciousness, affect, and object constancy.

Therapy combines an object-relational psychodynamic focus with cognitive-behavioral interventions. Active assistance in developing self-soothing skills is provided through imagery and suggestion. When behavioral acting out of aggressive and self-destructive impulses occurs, the therapist works to establish structural boundaries in the therapy to provide external regulation while internal regulation of behavior is developing. Use of medication, is not recommended, except for producing sedation in crisis situations.

SUGGESTED CROSS-REFERENCES

Chapter 15 on posttraumatic stress disorder and other anxiety disorders contains important background information, as does Chapter 18 on dissociative disorders. Psychiatric treatment of children is discussed in Chapter 48.

SECTION REFERENCES

Benjamin LR, Benjamin R: Utilizing parenting as a clinical focus in the treatment of dissociative disorders. Dissociation *7:*4, 1994.

Cole PM, Alexander PC, Anderson CL: Dissociation in typical and atypical development: Examples from father–daughter incest survivors. In *Handbook of Dissociation,* LK Michelson, WJ Ray, editors. Plenum, New York, 1996.

Coons PM: Clinical phenomenology of 25 children and adolescents with dissociative disorders. Child Adolesc Psychiatr Clin North Am *5:*2, 1996.

Ellenberger HF: *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*. Basic Books, New York, 1970.

Fagan J, McMahon P: Incipient multiple personality in children: Four cases. J Nerv Ment Dis *172:*26, 1984.

Goodwin JM, Sachs RG: Child abuse in the etiology of dissociation. In *Handbook of Dissociation*, LK Michelson, WJ Ray, editors. Plenum, New York, 1996.

Hilgard ER: *Hypnotic Susceptibility*. Harcourt Brace Jovanovich, New York, 1965.

*Hornstein N, Putnam FW: Clinical phenomenology of child and adolescent dissociative disorders. J Am Acad Child Adolesc Psychiatry *31:*1077, 1992.

*Hornstein NL: Complexities of psychiatric differential diagnosis in children with dissociative symptoms and disorders. In *The Dissociative Child: Diagnosis, Treatment, and Management,* ed 2, JL Silberg, editor. Sidran Press, Lutherville, MD, 1998.

Kluft RP: Multiple personality in childhood. Psychiatr Clin North Am *7:*121, 1984.

Lewis DO: Diagnostic evaluation of the child with dissociative identity disorder/multiple personality disorder. Child Adolesc Psychiatr Clin North Am *5:*303, 1996.

*Main M, Morgan H: Disorganization and disorientation in infant strange situation behavior: Phenotypic resemblance to dissociative states. In *Handbook of Dissociation,* LK Michelson, WJ Ray, editors. Plenum, New York, 1996.

*Peterson G: Diagnostic taxonomy: Past to future. In *The Dissociative Child: Diagnosis, Treatment, and Management,* ed 2, JL Silberg, editor. Sidran Press, Lutherville, MD, 1998.

Peterson G, Putnam FW: Preliminary results of the field trial of proposed criteria for dissociative disorder of childhood. Dissociation *7:*4, 1994.

*Putnam FW: Child development and dissociation. Child Adolesc Psychiatr Clin North Am *5:*285, 1996.

Putnam FW, Helmers K, Trickett PK: Development, reliability, and validity of a child dissociation scale. Child Abuse Negl *17:*731, 1993.

Putnam FW, Hornstein N, Peterson G: Clinical phenomenology of child and adolescent dissociative disorders: Gender and age effects. Child Adolesc Psychiatr Clin North Am *5:*351, 1996.

Ross CA: Epidemiology of dissociation in children and adolescents: Extrapolations and speculations. Child Adolesc Psychiatr Clin North Am *5:*2, 1996.

Siegel DJ: Cognition, memory and dissociation. Child Adolesc Psychiatr Clin North Am *5:*509, 1996.

Silberg JL, editor: *The Dissociative Child: Diagnosis, Treatment, and Management,* ed 2. Sidran Press, Lutherville, MD, 1998.

Smith SR, Carlson EB: Reliability and validity of the Adolescent Dissociative Experiences Scale. Dissociation *9:*2, 1996.

Steinberg M: Diagnostic tools for assessing dissociation in children and adolescents. Child Adolesc Psychiatr Clin North Am *5:*333, 1996.

Waters FS, Silberg JL: Therapeutic phases in the treatment of dissociative children. In *The Dissociative Child: Diagnosis, Treatment, and Management*, ed 2. JL Silberg, editor. Sidran Press, Lutherville, MD, 1998.

Waters FS, Silberg JL: Promoting integration in dissociative children. In *The Dissociative Child: Diagnosis, Treatment, and Management*, ed 2. JL Silberg, editor. Sidran Press, Lutherville, MD, 1998.

Yeager CA, Lewis DO: The intergenerational transmission of violence and dissociation. Child Adolesc Psychiatr Clin North Am *5:*393, 1996.

49.9 GENDER IDENTITY AND SEXUAL ISSUES

ALAYNE YATES, M.D.

SEXUAL DEVELOPMENT

Sexual growth is composed of several developmental lines: (1) psychosexual growth, which contributes to personality formation; (2) gender, which contributes to the sense of self, self-worth, role, and identity; and (3) eroticism, which contributes to sexual responsiveness. The three components interrelate, complement, and sometimes conflict with one another.

EROTIC DEVELOPMENT

Sexual interest or activity in prepubescent children is often assumed to be abnormal, potentially dangerous, or indicative of child abuse. Historically, these concerns are related to the Calvinist perception of sexuality as an evil force spawned by the devil and the somewhat later romantic perception that children are innocents whom parents and other adults must protect from evil.

The sexual revolution of the 1960s and 1970s recognized that adult sexual dysfunction was probably related to earlier sexual inhibition or trauma but the revolution never addressed early erotic development. In the 1980s and 1990s exploration of normal development was further compromised by the association with child sexual abuse. New rules were formed to protect children; for instance, preschools, which had permitted some age-appropriate exploration, no longer allowed boys and girls to visit the toilet at the same time. Eventually the concept of sexual abuse came to include anything sexual that could involve children, casting doubt upon events such as sibling sex play, children bathing with parents, and teachers hugging pupils.

Research in Child Development Research in early erotic development is sharply limited by social constraint. Parents, schools, and review boards are unlikely to permit the direct questioning of children about their sexual interests and experiences. Therefore, most of the information available comes from questionnaires administered to caregivers or from the direct inquiry of children who have been or are thought to have been sexually abused. Parent questionnaires can yield valuable information, but most children over the age of 5 or 6 conceal their sexual activities from adults; those who do not are likely to have behavior problems.

Research on sexually abused children inevitably associates sexual activity with emotional and behavioral problems and therefore is of little use in understanding normal development. However, the study of sexually abused children does illustrate the process of eroticization, which occurs to some extent in all children. Young, sexually abused children can become highly eroticized, which most reliably differentiates them from nonabused youngsters. They may initiate fondling and oral sex with others and they often react to nonsexual situations as if they were sexual. Once established, this pattern is remarkably resistant to change.

Researchers ascribe the hypereroticism of sexually abused children to re-enactment of trauma, a symptom of posttraumatic stress disorder, which results from the fact that the sexual behavior is inflicted on children and that they have no control over it. Yet many abused children without posttraumatic stress disorder have also become highly erotic, perhaps because they have simply learned to be very sexual. Abused children receive more erotic stimulation and less punishment for being sexual and they may have learned to enjoy it more. This does not change their status as victims or make the sexual abuse less of a crime, but it does challenge society's notion of the innocent child.

Assuming that children are sexual and that early eroticism has something to do with adult sexual function or dysfunction, the obvious next step would be to study how normal children learn to be sexual or, conversely, how they acquire sexual inhibitions. A comprehensive longitudinal study of normal erotic development would answer some of these questions and determine which biopsychosocial factors add to or detract from healthy sexual function in adult life. Such a study could define risk and protective factors for posttraumatic stress disorder when children are exposed to sexual events or sexual trauma, and could provide information on whether current educational programs to prevent sexual abuse help or harm normal erotic development.

Virtually all children are born sexual. Infant boys have erections from the time they are in utero; infant girls lubricate vaginally when their genitals are cleansed shortly after birth. Infants of both sexes stimulate themselves and, in some cases, experience orgasm in the first year of life, although the peak of masturbation is somewhat later, at age 2 to 3 years. From the age of 4 months, infants anticipate diapering with pleasure, gurgling, and chortling, grabbing the parent's hand and placing it on the genitals. Toddlers enjoy exhibiting their genitals to parents and they often initiate bathtub or toilet sex play. Activities such as these help children define themselves as boys or girls but, as most parents know, they are also extremely pleasurable. Half of all preschool children engage in some form of heterosexual or homosexual sex play, mostly consisting of looking at or touching the genitals, tickling, or rubbing against each other, as in the games “Doctor” and “Barbie and Ken.” Activities such as undressing and showing off the genitals, touching one's own genitals or those of another child, scratching one's crotch, touching women's breasts, and peeking at others who are naked are commonly reported by parents.

Self-pleasuring and sex play is associated with emotional health and good parenting in the preschool years. After age 6 the presence of observed sexual activity is more likely to be associated with acting out or “high-risk” behaviors, probably because internalizing youngsters do not engage in, or hide, sexual activity whereas more aggressive, externalizing youngsters demonstrate many upsetting behaviors, some of which are sexual. Nothing is known about the sexual interests or activities of school-aged internalizing children and so they are often presumed to be asexual. By the same token, little is known of the sexual adjustment of internalizing adolescents whereas a great deal is known about adolescents who display a variety of high-risk behaviors.

Biological or temperamental factors must play a role in erotic development. Inhibited children are more likely to suffer from separation anxiety when they are young and from depression as they mature; this inhibition and depression could contribute to sexual problems in adult life. However, biological factors alone cannot account for most sexual problems in adults; social and psychological factors are usually involved.

Sexual development is influenced by the confusing messages children receive from their parents and the media. Sexual activity is presented as fascinating and great fun on television but as naughty or dangerous by parents and teachers. Bodies are idealized on magazine covers but youngsters are chastised if they show their underpants. Children who are not helped to reconcile these messages may conclude that values and moral standards change depending on who is talking or the channel to which the television is tuned.

An overriding effect on erotic development is society's critical attitude toward the body, which focuses on how it should look or perform instead of how good it appears and how pleasurable it can feel. Many preschool children and almost half of grade-school children are already dissatisfied with their body and attempt to change it through diet or exercise. Body dissatisfaction escalates after the onset of puberty in girls and remains high into the adult years; it is often accompanied by self-criticism and irrational fears about social and vocational failure. These elements can contribute to the development of an eating disorder and persons with eating-disorders frequently suffer from sexual dysfunction.

Body dissatisfaction is usually attributed to the media's emphasis on diet and slimness. However, the media could just as well be responding to the demands of its audience. The reason why body dissatisfaction has increased sharply in recent years is not known but could be related to sociocultural changes such as the rapid increase in dual-wage families. Children are often separated from their parents at an early age and are expected to become relatively independent. This is a normative crisis of considerable magnitude that tests the limited resources of toddlers. Because children experience the need to be held and fed in the body, they may

reject the body in order to achieve greater autonomy.

SEXUAL ANATOMY: MALE AND FEMALE

Genetic influences determine whether a person is born male or female. The fetus is initially and intrinsically female; an anatomical male is produced if the Y chromosome is present to promote testicular development and the production of testosterone which is then converted to dihydrotestosterone. If the gonads are removed before the eighth week of fetal development, the embryo will develop as a female. The androgen-estrogen balance determines gender differences in the size and number of nuclei contained in brain neurons, synaptic development, and dendritic branching patterns.

However, the relationship between chromosomes and clinical outcome is complex. Individuals with sex chromosome abnormalities such as 47, XXY, 47, XYY, and 47 XXX do not present with gender identity disorder, requests for sex change, or unconventional sexual orientation. Other mechanisms than the action of gonadal steroids could determine sexual dimorphism. In mammals, differentiation is fostered by gene products such as SRY, a testis determining gene on the Y chromosome. Other factors such as gene linkage disequilibrium may play a role.

Fetal testosterone permanently alters the developing brain through the migration of neurons, exerting an organizing influence on the hypothalamus, the preoptic area, and the amygdala during the sixth to the twelfth week of fetal development. Brain organization takes place only within brief critical periods, but its affective, behavioral, and cognitive effects last a lifetime. Testosterone-affected brains become larger and heavier due to an increase in cortical gray matter. Sex difference or dimorphism is also present in the corpus callosum, cerebellum, amygdala, and planum temporale. Perinatal testosterone exposure causes a nucleus located in the preoptic area of the hypothalamus to become two and a half times larger in males, with twice the number of neurons; this nucleus is probably involved in sexual orientation. The development of the female fetus is not as well understood as that of the male. Follicle-stimulating hormone is present in the pituitary of males and females by the tenth week of gestation. It increases markedly in females between the twelfth and twentieth week of gestation, the same time that males are experiencing a surge in testosterone. Follicle-stimulating hormone fosters the development of ovaries and the production of ovarian hormones.

Gender-specific effects of sex hormones on the brain include: (1) language organized primarily on the left side of the brain in males and on both sides of the brain in females, (2) increased language and fine motor skill in females, (3) greater visual-spatial ability in males, and (4) gender-specific repertoire of sexual and nonsexual behaviors. In addition, sex hormones regulate and are regulated by neurotransmitters, neuropeptides, growth factors, and neuroactive steroids.

Sex hormones continue to affect mood, cognition, and behavior throughout life. Testosterone fosters aggressiveness and activity in males and intensifies sexual thoughts, feelings, and behaviors. In the past, high levels of testosterone were thought to promote impulsive aggression and social maladjustment in males; recent work indicates an association instead with social dominance and success. A reciprocal feedback system operates whereby perceived success in social or competitive arenas increases testosterone levels whereas stress decreases testosterone levels.

Although estrogen is thought to be a female hormone, cross effects within the brain are currently being investigated. Male animals who are unable to respond to estrogen become uncoordinated in sexual and aggressive pursuits, and females unable to respond to estrogen exhibit uncharacteristically fierce territorial aggression, including attacks on males.

Normal youngsters experience an increase in sex hormones during the first 2 to 4 years of life, in conjunction with relatively robust erotic interest. Then hormone levels drop and they remain relatively low until several years before puberty. This juvenile phase, roughly equivalent to latency, seems to be secondary to cortical inhibition. Puberty begins with a dramatic rise in gonadotrophins, growth hormone, estrogen, and testosterone.

Prepubescent girls require a certain percentage of body fat for menarche to occur. Extremely lean gymnasts and dancers retain their prepubescent habitus. Leptin is a recently discovered hormone that resides in fat cells and that signals the brain when there is sufficient body fat to accomodate puberty. Leptin increases in girls but decreases in boys during puberty, as testosterone appears to supress leptin. When leptin is injected in immature female animals, they mature much earlier than others.

GENDER IDENTITY

Gender identity is how a person perceives the self as male or female. The human infant is born with a bias toward a certain gender identity, one that is created or reinforced by the maleness or femaleness of the brain. Brain bias usually matches the genital or anatomical sex as both develop in utero in response to fetal and maternal hormones. By the age of 2½, children can clearly and consistently state whether they are male or female, and they recognize others as male or female. By age 6 to 10 they know if their gender identity matches their anatomical sex.

Sex Assignment When the genitals are ambiguous, a decision to rear the child as male or female is usually made shortly after birth, with completion of surgical reconstruction by the child's third birthday. Decisions are based on the presence and adequacy of the genital apparatus, expected socio-sexual functioning, and the parents' preferences. The objective is to achieve, as far as possible, a gender-congruent physique. The female sex is favored because it is easier for the surgeon to construct a vagina than a functional phallus.

Most children adapt to the assigned sex even when it does not match the original anatomical sex. Contradictory feelings and behaviors resulting from prenatal hormonal influences are usually accommodated as long as the child's physical appearance is gender appropriate and the parents are not unduly confused or conflicted about the decision. However, clinicians have begun to question routine sex assignment because the sex chosen may be at odds with brain programming and the brain may have the final word. A boy whose penis was accidentally destroyed at the age of 8 months had his testicles removed, his perineum reconstructed, and was reared as a female. Yet he refused to play with girls' toys and insisted on standing to urinate. At age 14, he refused to live as a female, even though he was still unaware that his sex had been reassigned; he was eventually surgically reassigned once again as male.

Sex assignment is completed early in order to facilitate the formation of gender identity and to avoid gender confusion. Once acquired, gender identity was thought to be set for a lifetime; however, gender may be more plastic than originally thought. Genetic males with 5-alpha-reductase deficiency who appear to be female and who are raised as females usually make a rapid, decisive transition to the culturally favored male gender when they virilize at the time of puberty. If they had been raised unassigned, as boy, girl, or intersex, reassignment would not have been necessary. In recent years, society has become more tolerant of transgender behavior. A small but growing number of adults live in a mixed or intersex state, sometimes receiving hormonal but not surgical treatment. A provisional assignment as male or female, without surgery, may be an option for some children in the future.

Caution is advised in assigning gender before the child is old enough to state a clear preference and to demonstrate gender-role behavior. However, if after receiving counseling the parents are unable to tolerate sexual ambiguity, early assignment is still indicated. Sex assignment ultimately rests upon our ability to assess the maleness or femaleness of the human brain. This may be possible within the next decade, enabling early, accurate decisions on sex assignment.

GENDER ROLE

Gender role refers to a collection of attitudes and behaviors that are typically male or female. Boys and girls display different play and work preferences, extracurricular interests, and courting patterns. Preferences are based on differing patterns of information processing and emotional responsiveness.

Although gender role is largely culturally determined, genetic or biological factors play a part. Across many cultures men are more active, aggressive, and competitive from an early age. Boys gravitate toward action toys and play out battle-and-capture scenes. Similarly, girls choose dolls and play out the desire to care for and teach children. When angry, girls are more likely to lash out verbally while boys are more likely to lash out physically. Girls worry about the people they care about whereas boys worry about performance.

Although gender identity is usually set from age 2, gender role often remains malleable until age 5, especially in females. By the time children enter kindergarten, most boys and girls feel compelled to act according to gender-specific rules and expectations. These are enforced primarily by peers who tease and ostracize effeminate boys and, to a lesser extent, girls who act like boys. Social sanctions are even more apparent in adolescence when appearance, courtship, and sexual performance are dictated by gender-specific scripts. Most adolescents readily accept these constraints because they enjoy the praise of peers and family and the

security of group membership.

Prenatal Influences The brain is the ultimate discriminator between male and female and it provides the biological substrate for gender role. Women are more proficient in verbal and affective communication and men display greater logical, analytical, and spatial abilities. These gender-specific patterns are related to male fetal development. Higher testosterone levels in the male fetus foster a delay in right-hemispheric development and greater left-hemispheric specialization.

In humans, prenatal hormones influence gender role most clearly in the choice of toys and pattern of play. Less clear, but still fairly convincing, is the case for hormonal influence on aggression and sexual orientation. Girls with congenital adrenal hyperplasia are exposed to increased androgens before and after birth and they often prefer playing with boys' toys. If they are not treated until age 8, they display a masculine pattern of arousal in that they quickly respond to visual erotic stimuli, experiencing strong, genitally located erotic arousal. They report fewer-than-expected fantasies of marriage, pregnancy, or motherhood and they are more likely than controls to become lesbians. Boys exposed to antiandrogens display lower energy levels, less aggression, and less heterosexual activity in adolescence, whereas girls exposed to prenatal antiandrogens appear more traditionally feminine than other girls.

In spite of these and other positive findings, there is a vast body of literature that is not only confusing but also reveals many negative findings about the effects of prenatal exposure to hormones. This may be because insufficient attention has been paid to the brief critical periods during which the hormones exert an effect. Based on current research, the overall influence of prenatal hormones seems to be less than the overall influence of the environment, whether this is translated to the child by social reinforcement, role modeling, or through dynamic, intrapsychic forces. However, today biological factors are viewed as more important than they were in the past.

CROSS-GENDER IDENTIFICATION

Symptoms of cross-gender identification are usually recognized in the preschool years, typically in boys who dress in women's clothes, prefer to play house, and avoid rough play. They may sit to urinate and openly express the wish to become a girl. Girls with a cross-gender identification dislike wearing dresses and prefer to play with boys. They may expect to grow a penis and become a man. Cross-gender identification extends to fantasy, with boys enjoying romantic comedies and girls gravitating toward action heroes.

Cross-gender identification is the conviction of belonging or the desire to belong to the other sex. It is almost always reflected in behaviors such as cross-dressing. Male cross-sex-typed behavior exists across various cultures regardless of how the culture views homosexuality. The behavior surfaces at approximately the same age and is resistant to change, all of which suggests that biological factors are its chief determinants.

Boys with cross-sex-typed behavior are likely to be shy and inhibited. Like other shy and inhibited boys, they tend to be anxious in the early years and to suffer from separation anxiety disorder, often becoming depressed later in life. This suggests that cross-sex-typed behavior and gender identity issues are not an isolated finding but part of a broader picture that includes sensitivity to attachment disruption and a predisposition to affective disorder. The cross-gender behavior could be the child's way of managing separation anxiety by becoming more like the mother.

Gender identity disorder is distinctly uncommon in males and extremely rare in females. Most young children with cross-sex-typed behavior do not qualify for a diagnosis of gender identity disorder because the behavior is not extreme or they are not distressed enough about their situation. Those who develop gender identity disorder represent a more disturbed subgroup with marked behavioral and peer problems; more than half qualify for a diagnosis of *separation anxiety disorder* perhaps related to temperamental variables or an overly close and ambivalent relationship to the master.

Boys with gender nonconformity are perceived by adults and girls as more attractive or handsome than other boys, which suggests that they are well accepted by parents, teachers, and female peers. However, in grade school and junior high they are likely to be labeled “sissy” or “fag” by male peers and can become isolated and depressed. Most learn how to moderate their behavior so that they are more accepted by peers and this favorable adaptation prevents them from developing gender identity disorder. A few youngsters, usually those who are markedly different from their classmates, become so affectively and cognitively preoccupied with their gender discordance that they cannot function socially or academically. In adolescence they dwell on their appearance, which is a preoccupation more typical of adolescent females. If this persists, it often leads to depression and sometimes suicide.

Three quarters of boys with cross-sex-typed behavior later become homosexual or bisexual. This is the case whether or not they have had therapy, have been able to suppress the incongruent behavior, or have developed gender identity disorder. Later in life a tiny minority, convinced that they are really female, take hormones, live as females and may request sex change surgery. The outcome for girls with cross-sex-typed behavior is unclear and there are no prospective studies.

SEXUAL ORIENTATION

Sexual orientation refers to an individual's overall sexual responsiveness to men, women, or both. Once established, sexual orientation is usually but not always constant throughout life. Approximately 4 to 6 percent of men and 2 to 3 percent of women are exclusively homosexual; a smaller percentage describe themselves as bisexual. Other cultures reveal a similar percentage of homosexuals. Homosexuality is not a disorder of gender identity and does not connote psychopathology; homosexual men may or may not be effeminate. They differ from the rest of the population in being somewhat less devoutly religious and better educated.

Gay men describe themselves as having felt different from their same-sex peers from early childhood. Lesbian women report having felt less feminine and less beautiful than peers, imagining that they were male, preferring boy's games, and being called “tomboy.” However, many women who are not gay also report these experiences.

Gay men usually become aware of homoerotic fantasies in early adolescence. At the same time they may have erotic fantasies about women. This generates confusion and considerable conflict, which in turn contributes to the high rate of suicide and school drop out among gay youth. Indecision and bewilderment often persist until a clear homosexual identity is established in early adulthood.

Compared to gay men, bisexual men are more conflicted, alienated, and depressed. This could be related to greater biological discordance or to greater difficulty in defining an identity in a society that prefers neat categories. Also, bisexual men with deep-seated problems may be less able to define themselves as male or female.

Compared to gay men, lesbians are more flexible and less conflicted about adopting a lesbian role and identity. Lesbian adolescent girls are less likely to be rejected by peers and to have more social support available. It is fairly common for girls to experiment sexually with female peers and for heterosexual married women to enter a homosexual relationship following divorce. Also, more women than men claim to be bisexual.

Etiology Certain animals are known to engage in sexual activity with members of the same sex; a smaller number are exclusively homosexual. Homosexual relationships can help the individual survive or foster the survival of the group. In some species, females who engage in sexual activity with each other are more willing to share food, to forage together, and to groom each other.

Psychosocial Factors Psychosocial forces cited in the literature and thought to further male homosexuality include: (1) separation-individuation issues, with an overly close or hostile-dependent relationship with the mother; (2) mothers' fear and devaluation of men; (3) parental preference for a girl; (4) trauma, including death of a caregiver; and (5) an emotionally distant or absent father.

Analytic theorists suggest that lesbian relationships are often a search for the bodily pleasure found in the early mother-daughter union. Early erotic experiences that can bias sexual orientation may also include frequent enemas, prolonged molestation, and sex play. Homosexual men and women report more early anal-erotic and homoerotic experiences within the family than do heterosexual individuals.

Genetic Factors Genes are strong contenders as progenitors of sexual orientation. Genes may influence sexuality by changing the organization or the structure of the brain or by affecting the production of neurohormones and transmitters. Minute changes in endogenous hormones could subtly recast attitudes, thoughts, emotions, and behaviors. Alone, or in conjunction with environmental influences, this could bias the individual toward homosexuality.

Family studies support a genetic basis for homosexuality. Homosexual men are more likely than other men to have homosexual brothers. Monozygotic homosexual

twins are more than twice as likely as dizygotic twins to be homosexual. Genetic influences are noted in lesbian women also. However, other variables must play a part because a significant number of monozygotic twins are discordant for homosexuality. More than one gene, as well as other biological and environmental influences, are likely to be implicated.

When the extended family of known homosexual persons is examined, more than the expected number of homosexual persons is found, but only on the mother's side of the family. This suggests transmission via the X chromosome. Indeed, researchers have correlated homosexuality with the inheritance of five markers at the tip of the long arm of the X chromosome, in the Xq28 region.

If homosexuality is largely genetic in origin, then children of heterosexual parents adopted by homosexual partners would not be more likely than the rest of the population to be homosexual. Although other factors such as role modeling and these parents' strong wish to raise a heterosexual child are also involved, these children are no more likely than the rest of the population to become gay or lesbian.

Brain Structure Differences in the brain structure between heterosexuals and homosexuals would ultimately result from genetic influences. Differences reported include an increase in the size of the suprachiasmatic nucleus of the hypothalamus, diminished size of one of the sex dimorphic interstitial nuclei of the anterior hypothalamus, and an increase in the size of the anterior commissure. Genetically male transsexuals who from an early age have been convinced that they are female demonstrate a smaller nucleus in the preoptic area of the hypothalamus, a nucleus that is typically larger in males and that plays an essential role in male sexual behavior. The auditory system of lesbian and bisexual women seems to function more like the auditory system of men. This finding must be replicated.

Brain Hormones Steroid hormones, including estrogens and androgens, were thought to be produced only in the adrenals and the gonads. Recently, endogenous steroid hormones have been found in localized areas of the brain. These hormones are called *neurosteroids* because they are formed by neurons in the same way that the gonads and the adrenals produce steroids. Neurosteroids may organize localized areas of the brain toward maleness or femaleness in the prenatal period and may affect postnatal behavior directly or through the earlier organizational changes. The existence of a third, independent hormonal pathway could explain why some men who look and behave very masculine are homosexual or why some persons are able to maintain sexual function following castration.

Another possibility that is currently under investigation is that prenatal stress and a deficiency of enzymes involved in steroid synthesis could predispose the fetus toward homosexuality. Corticotropin (ACTH) administration significantly increases production of the cortisol precursor 21-deoxycortisol in homosexual men and women, compared to nonhomosexual subjects; a similar increase occurs in the mothers of homosexual men.

Peripheral Hormones Earlier research tried to establish a link between homosexuality in males and lower plasma testosterone levels but no such link has ever been demonstrated in males. About one third of lesbian women, however, do present with higher levels of testosterone than the average woman but these levels are still well below the levels found in men. Small increases could result from differences in social dominance or self-efficacy.

Immunological Factors Homosexual men are far more likely than other men to be born late in the birth order and to have more older male siblings. Conversely, parents who produce many sons are more likely to produce extremely effeminate, homosexual sons. Effeminate boys with gender identity disorder also tend to be born late and to have older male siblings. Several explanations are possible, one of which is based on immunology. Long before a mother gives birth to a male fetus, fetal cells are circulating in her bloodstream. These persist for as long as 27 years following the birth of the child. The presence of male cells could provoke an immune reaction against Y chromosomal material, which theoretically could lower prenatal brain testosterone in later-born males, thus affecting the structure and function of the brain. Later-born male infants are known to have lower levels of testosterone than first-born males, and infants spaced closely have lower testosterone than infants born 4 or more years apart.

Homophobia A strong antigay bias, called *homophobia,* exists in our society. Homophobia is based on social sanctions by which anyone who deviates from expected norms of gender role may be criticized or rejected; sanctions are stricter for males than for females. Individuals who repress homosexual inclinations are more strongly homophobic than those who are less inclined toward homosexual impulses. In essence, they reject the homosexual parts of the self by rejecting all that is homosexual. If a biological basis for homosexuality is established, homophobia should diminish and it would be easier for individuals to accept themselves and to appreciate the uniqueness and diversity of human nature.

SUGGESTED CROSS-REFERENCES

Related discussions include Section 19.1a on normal human sexuality and sexual dysfunctions, Section 19.1b on homosexuality, Section 32.2 on normal child development, and Section 32.3 on normal adolescent development. Gender identity disorders are presented in Section 19.3.

SECTION REFERENCES

Arnold AP: Genetically triggered sexual differentiation of brain and behavior. Horm Behav *30:*495, 1996.

Blanchard R, Zucker KJ, Cohen-Kettenis PT, Gooren LJG, Bailey JM: Birth order and sibling sex ratio in two samples of Dutch gender-dysphoric homosexual males. Arch Sex Behav *25:*495, 1996.

Blanchard R, Bogaert AF: Homosexuality in men and number of older brothers. Am J Psychiatry *153:*27, 1996.

Blum WF, Englaro P, Hanitsch S, Juul A, Hartel NT: Plasma leptin levels in healthy children and adolescents: Dependence on body mass index, body fat mass, gender, pubertal stage, and testosterone. J Clin Endocrinol Metab *82:*2904, 1997.

Bradley SJ, Zucker KJ: Gender identity disorder: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *36:*872, 1997.

Coates S, Person ES: Extreme boyhood femininity: Isolated behavior or pervasive disorder? Annu Prog Child Psychiatry Child Devel *73:*197, 1986.

Coates SW, Wolfe SM: Gender identity disorder in boys: The interface of constitution and early experience. Psychoanal Inq *15:*6, 1995.

Diamond M: Sexual identity and sexual orientation in children with traumatized or ambiguous genitalia. J Sex Res *34:*199, 1997.

*Diamond M, Binstock T, Kohl JV: From fertilization to adult sexual behavior. Horm Behav *30:*333, 1996.

Finkelhor D, Berliner L: Research on the treatment of sexually abused children: A review and recommendations. J Am Acad Child Adolesc Psychiatry *34:*1408, 1995.

Friedrich WN, Grambsch P, Broughton D, Kuiper J, Beilke RL: Normative sexual behavior in children. Pediatrics *88:*456, 1991.

*Green R: *The "Sissy Boy Syndrome" and the Development of Homosexuality*. Yale University Press, New Haven, 1987.

Harmon RJ, Bender BG, Linden MG, Robinson A: Transition from adolescence to early adulthood: Adaptation and psychiatric status of women with 47 XXX. J Am Acad Child Adolesc Psychiatry *37:*286, 1998.

Kagan J, Reznick JS, Snidman M: Biological bases for childhood shyness. Science *240:*141, 1988.

Kim K, Smith PK: Childhood stress, behavioural symptoms and mother-daughter pubertal development. J Adolesc *21:*231, 1998.

Koestner R, Aube J: A mulfactorial approach to the study of gender characteristics. J Pers *63:*681, 1995.

McFadden D, Pasanen EG: Comparison of the auditory systems of heterosexuals and homosexuals. Proc Natl Acad Sci USA *95:*2709, 1998.

Money J: The concept of gender identity disorder in childhood and adolescence after 39 years. J Sex Marital Ther *20:*163, 1994.

Money J, Russo AJ: *Homosexual Outcome of Discordant Gender Identity/Role in Childhood: Longitudinal Follow-up*. Plenum Press, New York, 1993.

Okami P, Olmstead R, Abramson PR, Pendelton L: Early childhood exposure to parental nudity and scenes of parental sexuality ("primal scenes"): an 18-year longitudinal study of outcome. Arch Sex Behav *27:*361, 1998.

Patterson CJ: Children of lesbian and gay parents. Child Dev *63:*1025, 1992.

Phillips G, Over R: Differences between heterosexual, bisexual, and lesbian women in recalled childhood experiences. Arch Sex Behav *24:*1, 1995.

Pillard RC, Bailey JM: A biologic perspective on sexual orientation. Psychiatr Clin North Am *18:*71, 1995.

Rosenblatt EA: Emerging concepts of women's development: Implications for psychotherapy. Psychiatr Clin North Am *18:*95, 1995.

*Rubinow DR, Schmidt PJ: Androgens, brain and behavior. Am J Psychiatry *153:*974, 1996.

Sacman MV: Psychopathology in women and men: Focus on female hormones. Am J Psychiatry *154:*1641, 1997.

Sanders AR, Gershon ES: Clinical genetics, VIII: From genetics to pathophysiology-candidate genes. Am J Psychiatry *155:*162, 1998.

Schaal B, Tremblay RE, Soussignan R, Susman EJ: Male testosterone linked to high social dominance but low physical aggression in early adolescence. J Am Acad Child Adolesc Psychiatry *34:*1322, 1996.

*Szkrybalo J, Ruble DN: “God made me a girl”: Sex-category constancy judgments and explanations revisited. Dev Psychol *35:*392, 1999.

*Tasker F, McCann D: Affirming patterns of adolescent sexual identity: The challenge. J Fam Ther *21:*30, 1999.

Tiefer L, Kring B: Gender and the organization of sexual behavior. Psychiatr Clin North Am *18:*25, 1995.

Turner WJ: Homosexuality, type 1: An Xq28 phenomenon. Arch Sex Behav *24:*109, 1995.

Wagner CK, Nakayama AY, De Vries GJ: Potential role of maternal progesterone in the sexual differentiation of the brain. Endocrinology *139:*3658, 1998.

Yates A: *Sex Without Shame: Encouraging the Child's Healthy Sexual Development*. William Morrow, New York, 1978.

Yates A: Differentiating hypererotic states in the evaluation of sexual abuse. J Am Acad Child Adolesc Psychiatry *30:*791, 1991.

*Yates A, editor: *Sexual and Gender Identity Disorders. Child and Adolescent Psychiatric Clinics of North America*. WB Saunders, Philadelphia, 1993.

Zhou JN, Hofman MA, Gooren LJ, Swaab DF: A sex difference in the human brain and its relation to transsexuality. Nature *378:*68, 1995.

*Zucker KJ, Bradley SJ: *Gender Identity Disorder and Psychosexual Problems in Children and Adolescents*. Guilford Press; New York, 1995.

Zucker KJ, Bradley SJ, Lowry Sullivan CB: Traits of separation anxiety in boys with gender identity disorder. J Am Acad Child Adolesc Psychiatry *35:*791, 1996.

Zucker KJ, Green R, Coates S, Zuger B, Cohenkettenis PT: Sibling sex ratio of boys with gender identity disorder. J Child Psychol Psychiatry *38:*543, 1997.

49.10 IDENTITY PROBLEM AND BORDERLINE DISORDERS

VICTOR M. FORNARI, M.D., AND DAVID PELCOVITZ, ꟻH.D.

IDENTITY PROBLEM

The development of an identity has been the subject of much debate, and numerous efforts have been made to describe the process that leads to identity formation. Failure to negotiate an identity is described in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) as an “identity problem” in the section entitled “Other Conditions That May Be a Focus of Clinical Attention.” As defined in the DSM-IV, the criteria for identity problem make it very difficult to use rigorous research methodology to clarify this condition's validity.

Over the past four decades the syndrome of identity confusion was described by Erik Erikson in his classic paper “The Problem of Ego Identity.” Erikson described a group of adolescents who failed in the negotiation of the moratorium provided by society between childhood and adulthood. Adolescents who experienced difficulties with the formation of an identity shared clinical features with individuals with borderline personality disorder. Identity formation has been described by Erikson as a control task of adolescence.

> The young person in order to experience wholeness, must feel a progressive continuity between that which he has come to be during the long years of childhood and that which he promises to become in the anticipated future; between that which he conceives himself to be and that which he perceives others to see him and to expect of him ... Identity is a unique product, which now meets a crisis to be solved only in new identifications with age mates and with leader figures outside the family.

Definition The concept of identity problem represents the DSM-IV current nosology “when the focus of clinical attention is uncertainty about multiple issues relating to identity such as long-term goals, career choice, friendship patterns, sexual orientation and behavior, normal values, and group loyalties.” Identity problems may appear as a feature of a variety of other psychopathology during adolescence, such as mood disorders, borderline personality disorder, posttraumatic stress disorder, and schizophrenia.

History Identity crisis during adolescence gained widespread attention during the late 1960s and early 1970s when many adolescents rejected their parents' mainstream values, and ideals and chose alternative lifestyles.

Identity disorder was considered a childhood variant of borderline personality disorder. Identity disorder was included in the third edition of DSM (DSM-III) and retained in the revised third edition (DSM-III-R) without any systematic research. There have been no systematic empirical research studies to date regarding the diagnosis of identity disorder. Some believe that in certain individuals the symptoms may reflect normal development and in others be a manifestation of established mental disorder.

Comparative Nosology DSM-III and DSM-III-R included the category identity disorder as one of the disorders usually first evident in infancy, childhood, or adolescence. The disorder was described for children and adolescents with “severe subjective distress regarding uncertainty about a variety of issues relating to identity,” which resulted in social or school functional impairment. These symptoms were required to be present for at least 3 months and not result from a mood disorder or a psychotic disorder.

The relation between identity disorder and borderline personality disorder has not been established. Most adolescents who struggle with issues regarding their identity and values do not develop borderline personality disorder. In DSM-IV, however, identity disorder was deleted from the classification of psychiatric disorders. As David Shaffer and his colleagues describe in their review of DSM-IV childhood disorders, it was decided that identity disorder could not be substantiated as a valid psychiatric syndrome, defined by reliably ascertained signs and symptoms and clearly distinct from other psychiatric disorders. In addition, it was difficult to distinguish identity disorder from the normal crisis of adolescence.

Epidemiology No reliable information exists on the frequency of occurrence, age of onset, or prevalence of identity problem. Identity problem does not qualify as a mental disorder but is related to identity concerns specific to a developmental phase. Therefore, it may be impossible to determine the true epidemiology of identity problem.

In *From Teenage to Young Manhood: A Psychological Study,* Daniel Offer and Judith Offer describe approximately one third of a model sample of adolescent boys who experienced a tumultuous course during their adolescence. This description resembles the features of identity problem. It is generally assumed that identity problems are more prevalent in industrial societies than in developing countries. This is based on the belief that traditional cultures remain bound by a set of stable values, ideals, and norms. Authority and hierarchy are clearly defined and leave little room for personal choice. Industrial societies offer a wide array of choices in occupations, relationships, lifestyles, and ideals. Greater emphasis is placed on personal, educational, and occupational achievement, with adolescents often leaving home by age 18 to attend university. In contrast, in developing countries, greater emphasis is placed on the extended family and the importance of traditions.

Etiology As in all human development, a variety of factors may contribute to the ultimate outcome. Identity problem may have predisposing factors, which may be biological, psychological, or social. The cause of identity problems is generally multidimensional and can best be evaluated on the basis of the development tasks outlined by Erikson in 1968 and Peter Blos in 1975. The task challenging the adolescent is to sort out, integrate, and develop a sense of identity following the dramatic transformation brought about by puberty and its concomitant physical changes. Following these changes, there is heightened sexuality with reproductive capacities. This brings with it greater demand for psychosocial separation from parents to find greater independence and the capacity for adult sexual intimacy. Erikson states that achieving a sense of inner identity results in persons experiencing themselves as “whole.” According to Blos, a central task of adolescence is to elaborate on the separation.

Diagnosis and Clinical Features The category of identity problem can be used when the focus of clinical attention is uncertainty about multiple issues relating to identity such as long-term goals, career choice, friendship patterns, sexual orientation and behavior, moral values, and group loyalties.

Friendship Patterns In a recent review of over 80 empirical studies investigating the role of friends in child development, Andrew F. Newcomb and Catherine L. Bagwell present impressive evidence regarding the crucial part that friends play in forming a healthy identity, fostering a positive self-concept, and promoting healthy psychological functioning. In recent years, a number of studies have also demonstrated the existence of a strong positive correlation between healthy friendships and school success and effective problem-solving skills. Rejected children tend to fall into two groups: disruptive children with high levels of aggression and socially withdrawn children who make easy marks for their peers. The work of Susan Harter and her colleagues has demonstrated that in both childhood and adolescence friends tend to be similar to one another in abilities and outlook. Rejected children tend to become friends with other rejected children, and aggressive children with other acting-out children. It is therefore not surprising that studies find that a major predictor of antisocial behavior is a situation in which an adolescent's peer network is composed of friends who smoke, drink, use drugs, or have a negative attitude toward education. Conversely, children who are attracted to friends with positive outlooks are more likely to develop positive school-related attitudes, career aspirations, and achievement.

A child's ability to interact appropriately with friends is a major contributor to formation of a healthy identity. Although infants may have regular contacts with other infants, peer relations are thought to exist in infancy in only a very loose sense. During the second year, for the first time, toddlers engage in simple reciprocal play that gradually gives way to parallel play, typically between ages 2 and 3. Social interaction during this period is typified by an awareness of each other's play without an accompanying sharing in a common goal. Cooperative play typically begins by age 3 and is consolidated by increasing interaction with peers in preschool

programs and other social activities arranged by parents. Between the ages of 3 and 6, most children begin to seek out the company of peers, prefer to associate with a member of the same sex, and actively engage in play that involves imitating adult roles with which they are trying to identify. Between the ages of 6 and 12, the children's circle of friends gradually widens, and social interactions become better organized. With the onset of adolescence, the range of acquaintances broadens, and friends are typically recruited from a wider pool on the basis of common values, interests, and styles. During adolescence, intimacy becomes an increasingly important component of friendship. Sharing thoughts and feelings with peers in a supportive and validating atmosphere is a key ingredient in the formation of identity.

In evaluating possible identity problems in children, the following aspects of the child's relationships with friends should be considered: (1) the number of friends (a consistent finding in the literature is that the amount of time the child spends with peers relative to that with family members increases from infancy through adolescence); (2) the quality of friendships (e.g., most girls show an increasing capacity for self-disclosure and intimacy in their friendships during the preadolescent years); (3) the behavioral and emotional difficulties of friends (since a major contributor to a child's identity, values, and behavior is shaped by friendships, a child or adolescent who interacts mainly with behaviorally disordered children is at increased risk for a variety of emotional and academic difficulties).

Sexual Orientation or Behavior Evaluation of normal and sexual behavior in children must be informed, in large part, by prevailing cultural, religious, and ethnic values. Since the norms for such behavior are constantly shifting, the mental health professional must be familiar with the standard in each child's neighborhood, school, and culture. Calvin Friedrich and his colleagues conducted the most widely cited empirical studies on normative child sexual behavior. He concluded that overt sexual behavior in boys increases from ages 2 to 4 and then gradually declines until age 12. In contrast, in girls overt sexual behavior steadily declines from ages 3 to 12. During the early school years, behaviors associated with adult sexuality (e.g., masturbating with objects, French-kissing, placing the mouth on another child's sex parts) are relatively rare in American children and should raise the possibility that the child might be sexually abused or overstimulated by inappropriate exposure to adult sexuality. Differences in the behavior of infant boys and girls tend to be subtle. Although some empirical evidence indicates that before the toddler years, boys tend to be more active and girls more vocal, it is not until approximately the age of 2, that clear differences in behavior emerge. Boys are more likely to prefer blocks, transportation toys, and rough-and-tumble play, while girls prefer more sedentary activities such as doll play and art work. The research of Kenneth Zucker, Richard Green and other experts in gender identity points to the importance of not viewing marked cross-gender behavior in young children as “just a phase.” A noticeable difference in patterns of sex-typed behavior during the preschool years should raise the possibility of a diagnosis of gender identity disorder. Many children with gender-atypical behavior will not display the persistent, intense distress required for this diagnosis, nor will they repudiate their basic sense of being male or female. Even if the child does not meet full criteria for gender identity disorder, a pattern of cross-gender interests in toy choices, fantasy play, or sex of playmates should raise the possibility of atypical gender development. An association between childhood gender nonconformity and adult sexual orientation has long been documented by the research of Zucker, Green, and others. Many, though not all, gay or lesbian adults recall a childhood marked by preference for cross-gender play. Adult retrospective research has also determined that although same-sex sexual orientation is often recognized by the child at an early age, the process leading to the insight that this is not just a phase typically does not emerge until adolescence or later. In many adolescents, this process is accompanied by a painful struggle that may entail periods of denial, suppression, and failed attempts at developing romantic or sexual relationships with the opposite sex. Achieving positive self-identification and ultimate integration and acceptance of sexual orientation may require the assistance of a mental health professional who is sensitive to the complex challenges to identity formation faced by these children and adolescents.

Values and Moral Development The gradual development of a child's sense of what is right or wrong, independent of the presence of parents or other adult authority figures, is a major component of overall identity. The pioneering work of Jean Piaget and Lawrence Kohlberg on the developing sense of morality in children has been expanded in recent years through a series of longitudinal studies conducted by (among others) Nancy Eisenberg and her colleagues and by the research of Barbara Stillwell and her associates.

Toddlers have only the most rudimentary conception of morality. This is perhaps best illustrated in their play, which typically involves no formal rules. As the child approaches the preschool years, a rigid set of rules emerges, albeit in a manner that is viewed by the children as inflexible and as fixed in stone as the laws of physics. The external conscience, which is present in most preschoolers, is characterized by an ability to label behaviors “good” or “bad” as derived from the outside authority of parents or other adults. This view gradually gives way to the child's conception of right or wrong as determined by self-interest; they obey rules or cooperate with peers as long as they get something in return. During the elementary school years, children's values reflect concern with the approval of others and an accompanying desire to display stereotypically “good” behavior. As development proceeds, a more sophisticated view of morality emerges. In sophisticated adolescents, moral reasoning is based on a belief that society must be based on laws and regulations. Individual needs are not considered more important than societal needs. Stillwell's research has demonstrated that during early adolescence, children acknowledge the importance of rules and regard for the feelings of others as important determinants of morality. Their conscience is viewed as a component of their personality, which they at times listen to, at times argue with, and at times ignore. In midadolescence, the “confused conscience” is described by Stillwell and her colleagues as a period of conflict and indecisiveness with the “gray” zones of good and evil. Conflicting messages adolescents receive from their peer groups and adult authority figures are reflected by adolescents' confusion regarding what they consider their true values. In late adolescence this period of uncertainty gives way to a more integrated conscience marked by flexibility, increasing ability to handle ambiguous moral situations, and a return of confidence regarding moral values.

When a child or adolescent displays values with sharp discrepancies from those of the family, culture, or society, the clinician should determine whether this reflects atypical identity development, risk for conduct problems, or part of a more normative process of identity development.

Career Choice Vocational identity is an important component of overall identity. In adults, having a job that is performed competently and is valued by society is an important contributor to a stable identity and a sense of general well-being. Recent longitudinal research has documented better long-term psychosocial and educational outcomes in at-risk children who can maintain a part-time job through adolescence. Research conducted by the second author and his colleagues for the DSM-IV posttraumatic stress disorder field trials on the sequelae of interpersonal victimization found that victims of child abuse are likely to have lower career aspirations than their nonabused counterparts.

Wim Meeus and other researchers have documented various approaches that adolescents may take in developing their occupational identity. Good outcomes are associated with adolescents who are satisfied with their occupational choice after openly exploring a variety of alternatives. Choosing a moratorium, in which options regarding occupational choice are left open and various alternatives are explored as part of a long-term process of defining ultimate occupational identity, is also viewed as a generally adaptive approach to defining this aspect of identity. Difficulties with the adolescent's emerging identity may be present when the adolescent indicates premature foreclosure by dealing with anxiety regarding the future by chosing a career course without exploring any options. Difficulty with vocational choice may also be indicated by adolescents who show signs of identity diffusion that may be indicated by a total lack of concern regarding planning for a future vocation or a view that luck will take care of finding a satisfactory vocation. Adolescents who take either of these two paths should be evaluated, as studies have found this approach to vocational choice to be associated with more difficulties with self-concept, educational achievement, and social adjustment.

Jane was a 16 ½-year-old white girl in the 11th grade, the middle of three children and resided with her parents. She had an older brother who was a junior in college and a younger sister in the 8th grade. Jane was referred for psychotherapy because of a concern about “poor school performance and poor choice of friends.” Jane reported to her therapist that she had lost interest in school because she was not clear what her future career goal would be. Feeling that she could not be as good a student in high school as she had been in grade school, she had withdrawn from academic work, and her school performance deteriorated. Previously very social, Jane began to withdraw from friends and hang out with peers with behavior problems. Jane's parents were concerned about the company she was keeping, and they would frequently argue about where she was going. Jane began to use marijuana to make herself relax and described a sense of uncertainty about her future goals and her friends. Jane's older brother, an accomplished student, was successfully completing his junior year in college, and Jane felt that she could never live up to his standards. A younger sister was in honor classes, and Jane felt inadequate compared with her. Although in good physical health, with no history of medical problems and no prior mental health contact, Jane exhibited significant distress.

On mental status examination, she appeared her stated age and related in a friendly and cooperative manner. Her mood appeared neutral, and her affect was wide range and appropriate. She stated, “I don't know why I'm here. I don't think I have a problem. My parents think I have a problem.” There was no evidence of suicidal or homicidal ideation. She was oriented to person, place, and time and appeared of at least average intelligence. Her fund of information was commensurate with her educational background. Jane did not meet criteria for major depressive disorder, an anxiety disorder, alcohol or other substance use disorder or any personality disorder.

Jane was struggling with her long-term career goals, her friendship patterns, and her position in her family. Jane struggled to accept her capacities as compared to those of her older brother and younger sister. The focus of therapy was on issues related to her emerging identity rather than any Axis I or II disorder.

Pathology and Laboratory Examination There are no specific laboratory tests or results for identity problem. No studies have established any consistent neuropathological or imaging techniques specific for this clinical entity.

Differential Diagnosis Identity problem can be differentiated from the normal conflicts of adolescence. Normal adolescence is not generally associated with deterioration in occupational, school, or social functioning.

Normal adolescents may argue with their parents, but their relationship with them remains intact. Adjustment disorders by definition are secondary to a specific stressor and are time limited. Although precipitating stressors may often occur at the onset of an identity problem, concerns with identity issues such as career choice, gender identity, and future plans are far less prominent in the adjustment disorders.

Identity problem must be differentiated from identity concerns that may represent the prodromal manifestations of schizophrenia, schizoaffective disorder, schizophreniform disorder, and mood disorders. Psychotic symptoms such as hallucinations and delusions or disturbances in thinking processes such as thought blockage, tangentiality, or flight of ideas are not present in identity problem. Adolescents who are struggling with identity issues may have subjective anxiety and confusion, which often makes the initial differential diagnosis difficult.

The most difficult differential diagnosis is that with borderline personality disorder. Confusion regarding identity concerns may be present in both conditions. In borderline personality disorder, the dramatic clinical picture includes a much more complex set of diagnostic criteria such as chaotic, unstable sexual and interpersonal relationships, alternating idealization and devaluation, intense anger, self-destructive behaviors, and chronic dysphoria. Borderline personality disorder does not resolve in several months to a year but generally persists throughout adulthood, often with significant morbidity and mortality.

Course and Prognosis Because there has been no systematic research to validate identity problem as a mental disorder, no data exist regarding the natural course of this disorder with and without treatment.

Treatment Because identity problem remains a poorly defined clinical entity, little is known about the most effective intervention strategies. There have not been any systematic studies testing the efficacy of treatment approaches; thus the literature includes only untested clinical accounts. The clinical literature emphasizes establishing a therapeutic alliance as a critical step in psychotherapy with adolescents with identity problem (i.e., developing an empathic and supportive relationship with a therapist who can engage the adolescent to clarify the conflicts). The capacity to collaborate productively allows the adolescent to address maladaptive coping mechanisms.

Family therapy can be clinically useful despite the lack of systematic studies demonstrating significance. Family therapy may help parents understand their adolescent child's new developmental level. Parents can then attempt to communicate more effectively without excessive struggle for control.

Group psychotherapy may offer the adolescent an opportunity to minimize a sense of isolation and provide a way of relating to peers and receiving peer feedback. Psychopharmacotherapy is not indicated for identity problem. Medications may be used to address specific symptoms such as anxiety or depressed mood.

BORDERLINE DISORDERS

Perhaps no disorder evokes as much negative emotion, intense hatred, or images of rage as borderline personality disorder. As adolescents, these young people often elicit contempt if not frankly murderous impulses; however, few if any disorders have been as controversial. Many consider it a legitimate diagnostic entity; others regard it as a “waste basket” for individuals who do not fit neatly into any other single category. Some believe that many individuals who receive this diagnosis may have a form of complex posttraumatic stress disorder, particularly following chronic child sexual abuse. The existence of borderline personality disorder during childhood and adolescence remains hotly debated because, it has been argued, adolescence is an active period of personality development. Borderline disorders often do not reflect the currently accepted DSM-IV diagnosis of borderline personality disorder, but rather are a variety of childhood manifestations often characterized by severe forms of psychopathology on the border between neurosis and psychosis.

The concept of borderline disorders in children and adolescents remains controversial. In 1990, Paulina Kernberg wrote that borderline personality disorder exists in children under age 12. Indeed, Kernberg argues for acceptance of the concepts of childhood personality and childhood personality disorder. Theodore Shapiro argues against such concepts; he believes it is problematic to use a designation designed for adults that does not take into consideration the developmental issues of childhood.

Definition Borderline disorders may reflect the diagnostic criteria outlined by DSM-IV for borderline personality disorder: a pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity beginning by early adulthood and present in a variety of contexts, indicated by at least five of nine specific symptoms. Borderline conditions may also represent something quite distinct. They may be characterized by a wide variety of severe childhood psychopathology.

History During the past 50 years, clinicians have described seriously troubled children whose disturbance in ego functions was less severe than that in psychotic children but more serious than that displayed by neurotic children. In 1949 Margaret Mahler and her colleagues categorized these children at the mild end of a clinical and developmental range that extends to the most severe psychotic conditions, the autistic and symbiotic psychoses of childhood. Mahler proposed the terms “benign” or “borderline” psychosis, a precursor to the idea of a schizophrenic spectrum in which borderline represents a less severe form.

In 1954 Rudolf Ekstein and Judith Wallerstein proposed the term “borderline” to describe children who were not on the way to psychosis but rather behaved in a “characteristic pattern of unpredictability which is paradoxically one of their most predictable aspects.” Thus, the concept of “borderline” as a stable clinical entity was defined by ongoing and rapid shifts in ego functioning. These early efforts generated an interest among psychoanalysts that led to further speculation about the developmental and clinical aspects of borderline children. Between the late 1950s and the late 1960s a number of authors described children with a wide range of difficulties. These children displayed low frustration tolerance, uneven development, a tendency to withdraw into fantasy or to regress into primary process in response to stress, and multiple somatic and anxiety symptoms.

By 1975, the important and influential work of Otto Kernberg defined borderline as a level of personality organization. Several types of personality disorders including schizoid, paranoid, antisocial, and narcissistic generally function at a borderline level of personality organization, according to Kernberg.

In DSM-IV, borderline personality disorder is designated as one of the Cluster B (or dramatic) personality disorders. This is consistent with the empirical, atheoretical approach of DSM-IV, which also includes narcissistic, antisocial, and histrionic personality disorders in Cluster B.

Comparative Nosology Personality disorders have been defined as relatively enduring and pervasively maladaptive patterns of experiences, relationships, and coping styles. Children and adolescents are involved in fluid developmental processes in which their personalities and bodies are changing at different rates, creating equilibrium changes in their relationship to their family and friends and within themselves. Is it valid to assign the diagnosis of a personality disorder to a child or adolescent? Are there enough data to support a distinct diagnostic entity of borderline or borderline personality disorder in childhood? Is there a clinical continuity between borderline children, borderline adolescents, and borderline adults? Does *borderline* refer to a discrete disorder as DSM-IV suggests or a level of developmental attainment as Kernberg describes? What is the relationship between borderline disorders and common Axis I diagnoses of children such as conduct disorder, attention-deficit disorders, anorexia nervosa, substance abuse, posttraumatic stress disorder, and depressive disorders? These and other important questions remain unanswered.

The common comorbid association of Axis I disorder raises the question whether borderline is really a complicated form of Axis I disorders. The common finding of a history of sexual abuse raises the question whether borderline is a pejorative designation for individuals who suffer a complex posttraumatic stress disorder as a consequence of protracted abuse as suggested by the second author and Sandra Kaplan and their colleagues. The diagnosis of complex posttraumatic stress disorder requires that an external stressor cause the pattern of symptoms seen in this syndrome. This represents a departure from the atheoretical approach taken by DSM-IV.

Developmental research has begun to generate some data concerning the notion of personality disorders in childhood. Prospective studies have shown how deviant

attachment patterns, tempermental constellations, and disturbances in early childhood relationships can evolve into extreme rigidity in relationships, coping mechanisms, and experience. Kernberg has contended that children can indeed display personality disorders. Basing her assumptions largely on clinical experience, Kernberg has advocated the validity of the concept of borderline personality in children under 12 years of age. Theodore Shapiro has cautioned that a good deal of research is needed to establish the validity and reliability of borderline personality disorder in childhood and adolescence.

One retrospective study has questioned the ability to discriminate between borderline and nonborderline children. More-recent studies by Pamela S. Ludolph and colleagues and Stuart J. Goldman and colleagues suggest that semistructured interviews such as Gunderson's Diagnostic Interview of Borderlines and DSM-III-R criteria can be applied to children. Longitudinal follow-up studies are needed to clarify whether children diagnosed as borderline grow into adolescents and adults with borderline personality disorder.

Epidemiology The first longitudinal study examining the childhood antecedents of a range of DSM-III-R personality disorders in a community-based sample was reported by David Bernstein and his colleagues in 1996. Although recent evidence supported the view that adolescence is a period of risk for the onset of personality disorders, little was known about childhood antecedents of adolescent personality disorders. Two well-controlled studies support the continuity of borderline symptoms from childhood through adolescence and adulthood. Bernstein and his colleagues reported in 1993 that personality disorders in adolescents peaked at age 12 in boys and age 13 in girls and declined thereafter.

The prevalence of personality disorders in adults was reported by the National Institute of Mental Health Epidemiologic Catchment Area program in the second stage of the Eastern Baltimore Mental Health survey in 1981, published in 1994. The adult subjects were directly examined by psychiatrists using a semistructured method that allowed diagnosis of all DSM-III personality disorders as well as comorbid DSM-III psychiatric disorders. The prevalence of personality disorders was reported to be 5.9 percent (9.3 percent when provisional cases were included). Men were reported to have higher rates than women. Individuals who were separated or divorced had the highest rates. Reliable epidemiologic data regarding borderline personality disorders in children and adolescents are limited.

John F. Clarkin and Cynthia Sanderson estimated the prevalence of borderline personality disorder in adults to be about 2 percent of the general population, about 8 to 10 percent among individuals seen n outpatient mental health clinics, and about 15 to 20 percent among psychiatric inpatients. It ranges from 30 to 60 percent among clinical populations with personality disorders. Borderline personality disorder is about five times more common among first-degree biological relatives of those with the disorder than in the general population. There is also an increased familial risk for substance-related disorders, antisocial personality disorder, and mood disorders. Borderline personality disorder is three times more frequent in women than men.

The pattern of behavior seen in borderline personality disorder has been identified in many settings around the world. Adolescents and young adults with identity problems, especially those accompanied by substance use, may transiently display behaviors that give the impression of borderline personality disorder. Such persons display emotional instability, existential dilemmas, uncertainty, anxiety-provoking choices, conflicts about sexual orientation, and competing social pressures to decide on careers.

Etiology A variety of theories exist regarding the etiology of borderline personality disorder. Each theory presents a different perspective, with varying amounts of theoretical and empirical support. Currently, a combination of genetic, biological, and environmental factors is believed to contribute to the abnormal behaviors related to deviant personality development.

Biological Factors Several studies have examined the incidence of psychopathology in first-degree relatives of individuals with borderline personality disorder. Borderline personality disorder patients have significantly higher rates of depressive disorders, alcohol use disorders, and antisocial personality disorder than schizophrenia patients. This and other family data suggest that borderline personality disorder is related to the affective spectrum of psychopathology and not to the schizophrenic spectrum. However, the relation of borderline personality disorder with concurrent or comorbid schizotypal personality disorder may be part of the schizophrenic spectrum and may exist in a subgroup of patients.

Evidence for biological and neurotransmitter causality in borderline personality disorder is compelling. The serotonin system, in particular, has been associated with low levels of behavioral inhibition in individuals with antisocial behaviors. Biological abnormalities in individuals with borderline personality disorder have been associated with affective instability, impulsivity, anxiety, and cognitive disorganization. Several small open trials of the selective serotonin reuptake inhibitor fluoxetine (Prozac) in individuals with borderline personality disorder demonstrated improvement in such core symptoms of the disorder as impulsive aggression following treatment. These and other findings support a role for 5-hydroxyindoleacetic acid in borderline personality disorder and have generated interest in the use of specific pharmacological probes to clarify further the role of serotonin in this disorder. In one study, subjects with borderline personality disorder were challenged with oral methchlorophenylpiperazine (mCPP) or placebo. Following mCPP the patients experienced decreased anger and fear. The male patients with borderline personality disorder had higher cortisol concentrations than the normal male subjects and marginally blunted prolactin responses after receiving mCPP. These results suggest serotonergic dysfunction in borderline personality disorder.

Positron emission tomography has been used over the past decade to probe for biological differences in regional cerebral blood flow in persons with borderline personality disorder. These individuals were reported to have a significant decrease in frontal cortex metabolism in one area above the canthomeatal line and significantly increased metabolism in another area above the canthomeatal line. These and other studies have investigated the hypothesis of a relation between frontal lobe cerebral rates of glucose metabolism and a lifetime history of aggressive impulsive difficulties.

There has been increased interest in the role of biological markers of borderline personality disorders. The dexamethasone suppression test, the thyrotropin-releasing hormone test, and sleep studies no longer indicate that borderline personality disorder is related to depression, although there is probably an affective subgroup. There may be new evidence that some subgroups of borderline personality disorder overlap with schizophrenia. There is new evidence regarding serotonin studies and new directions for understanding suicidal, aggressive, and impulsive traits. The tendency to mood lability, anger, and irritability in individuals with borderline personality disorder clearly interferes with development of a cohesive sense of self. Parents of these individuals, often vulnerable themselves to mood and personality disorders, and their children become involved in coercive cycles of anger and rejection followed by desperate attempts at resolution. The child psychiatric literature has long described the atypical ego development, impulsivity, and learning problems of many borderline children and adolescents.

Strong clinical overlap exists between the symptoms of the disruptive behavior disorders, particularly attention-deficit/hyperactivity disorder and conduct disorder, and those of borderline personality disorder. Other investigators report a link between learning disabilities and disruptive behaviors in children and borderline personality disorder in adulthood. Attention-deficit disorders and reading disorders can both affect a child's development in a number of ways. Because the actions of inattentive and impulsive children seem to occur rather than result from choice, they experience little sense of guilt or control. Studies have found that juvenile delinquents whose behavior is dominated by borderline and impulsive symptoms constitute two of four overlapping psychological subtypes. These individuals often have little understanding of their own inner psychological world, and these impulsive angry youngsters often fuel the chaos in their families. A direct linear relation between constitutional vulnerability and development of borderline personality disorder in children and adolescents cannot be drawn. Most children with attention-deficit/hyperactivity disorder, mood disorder, or separation anxiety disorder do not go on to develop borderline personality disorder. In addition, there are children who develop borderline personality disorder who do not have attention-deficit/hyperactivity disorder, mood disorder, or separation anxiety disorder. Biological factors in the development of borderline personality disorder can be related to the disorder in at least two ways. A vulnerability may increase the likelihood of other problems, thus creating an environment that can result in borderline personality disorder, or the biological vulnerabilities may increase the symptoms or their severity, increasing the chance of bringing children with the disorder to diagnosis and treatment. Thus, clinical surveys of children with borderline personality disorder may need caution, because they represent an unclear reflection of biological factors in these disorders.

Psychodynamic Factors Most psychodynamic theories related to the development of borderline personality disorder have suggested significant deficits in the area of interpersonal relationships, particularly in separation individuation. Failing to internalize early mother-child caring because of hypothesized difficulties in normal separation individuation, the borderline child or adolescent often feels empty and anxious and lacks a stable sense of self. Psychoanalytic theories include a range of hypotheses to explain the cause of borderline personality disorders in children and adolescents.

Mahler's ideas regarding separation individuation and Otto Kernberg's theories regarding splitting have provided the most important framework for psychodynamic clinicians. Children between the ages of 1 and 3 go through a series of developmental stages during which they develop the capacity to internalize some of the soothing functions previously performed exclusively by caregivers and they develop an awareness about themselves and begin to integrate the good and bad images of the self and the object. These developmental processes allow children to begin to accept the reality of their existence as separate individuals and to develop object constancy. *Object constancy* refers to the ability to maintain relationships in spite of separation from or frustration with their loving caregivers. Both Kernberg and Mahler assert that difficulties in these developmental processes result in borderline psychopathology. According to Kernberg, the central feature of borderline psychopathology is the ongoing effort to hold onto all good idealized images of the self, which has to be protected from constantly lurking all-bad introjects activated

by separation, frustration, or the object's failure to live up to idealized expectations.

James Masterson describes certain patterns of mother-infant interaction that interfere with the separation individuation process and lead to borderline psychopathology. In this view, mothers of future borderline individuals find gratification in their child's dependency and reward passive dependent, “clingy” behavior. Gerald Adler postulated that the central feature of borderline psychopathology is the individual's inability to evoke the memory of a soothing, comforting object when facing distress. That deficit in internalization is attributed to the failure of the parent to provide an adequate holding environment as described by Donald Winnicott. As a result, an inner sense of emptiness, reliance on transitional objects such as drugs or food to provide comfort and soothing, and manipulative efforts to produce the attention and involvement of others develop.

Glen Gabbard summarized the various psychodynamic models of the development of borderline personality disorder and stressed the importance of the various early developmental theories regarding both the separation individuation process and constitutional vulnerabilities. These psychodynamic hypotheses tend to exaggerate maternal responsibility and ignore the role of others, particularly abusive and neglectful parents, in the pathogenesis of borderline personality disorder in children and adolescents.

Environmental Factors In recent years, dissatisfaction with the validity of applying the construct of borderline personality disorder to childhood victims of physical or sexual abuse has increased. At the same time, clinical and empirical studies of childhood victims of interpersonal trauma have demonstrated a subset of individuals who, following exposure to traumatic events, manifest a constellation of symptoms not fully captured by the reexperiencing, arousal, and avoidance symptoms that constitute posttraumatic stress disorder. Research during the past decade has shown that the impact of trauma on psychological adaptation differs at different stages of development and that earlier trauma affects subsequent maturational processes. The research of Bessell van der Kolk and his colleagues has shown that traumatic experiences, particularly if they occur in childhood or adolescence, interfere with the development of self-regulatory processes and with the capacity to manage subsequent stresses. In particular, the research of Richard Famularo and other experts in the field of child abuse and neglect demonstrated that childhood victims of physical and sexual abuse manifest a variety of symptoms more traditionally associated with a diagnosis of borderline personality disorder.

Profound changes in affect regulation and self-identity have long been noted in the clinical and research literature on childhood victims of interpersonal trauma. Most noteworthy is the consistent finding that childhood victims of neglect, physical abuse, and sexual abuse or a combination are at much greater risk for extended periods of anger, unmodulated aggression, apprehension, guilt and fear, alterations in their relationships to caregivers, and difficulty with intimacy. Research has also shown that age-appropriate psychological defenses used to cope with early traumatic experiences, such as denial and dissociation, are frequently used by traumatized individuals during periods of subsequent stress.

Virtually all of the nine criteria for borderline personality disorder have been noted in childhood victims of abuse. Affect dysregulation is described in the four criteria that include difficulty controlling anger, affective instability, impulsivity, and suicidal behavior, (criteria 4, 5, 6, and 8); alterations in relations with others is seen in the sensitivity to abandonment and unstable interpersonal relationships (criteria 1 and 2); disturbance in the sense of self is described in the identity disturbance criterion (criterion 3), and one of the most researched sequelae of abuse, dissociation, is the ninth borderline criterion. It is therefore not surprising that a number of recent investigations have documented histories of early interpersonal trauma in most adults and children diagnosed with borderline personality disorder.

A number of recent clinical and research studies have investigated the suitability of the diagnosis “complex posttraumatic stress disorder” to capture the broad range of sequelae of interpersonal victimization. As described by Pelcovitz and his colleagues, in preparation for the DSM-IV posttraumatic stress disorder field trials, which had an investigation of the validity of such a diagnosis as one of its goals, a list of symptoms typically seen in trauma survivors was generated by a survey of experts in the field and systematic review of the literature on the emotional and behavioral sequelae of childhood sexual abuse, physical abuse, crime, rape, incarceration in concentration camps, torture, and spouse abuse. This process led to a consensus definition of complex posttraumatic stress disorder that was in turn investigated in the field trials. Ironically, although not the intention of this team of investigators, these symptoms overlapped significantly with the diagnosis of borderline personality disorder. The symptoms of complex posttraumatic stress disorder include alterations in the following domains of functioning: alterations in regulation of affect and impulses, including difficulty modulating anger and sexual impulses, suicidal behavior, and risk taking; alterations in attention and consciousness, including amnesia and transient dissociative episodes; alterations in self-perception, including feelings of ineffectiveness, shame, damage, guilt, and responsibility for the trauma; alterations in relations with others, including difficulties with basic trust, a tendency to enter situations in which revictimization is likely, or responding to the trauma by victimizing others in the same manner that one was victimized; somatization, somatic symptoms for which no physical cause has been found; and alterations in systems of meaning, including feelings of despair and hopelessness and loss of previously sustaining beliefs. In DSM-IV, the symptoms of complex posttraumatic stress disorder are described under the “Associated Features” of the posttraumatic stress disorder diagnosis. In the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), a virtually identical diagnosis is included in the category of lasting personality changes following catastrophic stress.

In recent years a series of empirical studies conducted by van der Kolk, the second author, and Susan Roth, have provided support for the validity of the complex posttraumatic stress disorder diagnosis. As predicted by trauma theory, complex posttraumatic stress disorder is more prevalent in victims of early trauma than of adult trauma, is more likely to be diagnosed in survivors of interpersonal than of man-made trauma, and has shown promise as a construct likely to generate more-effective treatments for trauma survivors than more traditional treatment approaches to borderline personality disorder.

Diagnosis and Clinical Features According to the DSM-IV, diagnostic criteria include the following; borderline personality of disorder is a “pervasive pattern of instability of interpersonal relationships, self-image, and affects and marked impulsivity, beginning by early adulthood and present in a variety of contexts...”

John was 17 years 9 months old when he presented for urgent consultation 6 weeks prior to high school graduation, referred by his high school psychologist because of concern that he was depressed. On presentation, John had made six cigarette burns along his left arm and three on his right arm. In addition, John reported having made several scratches with an Exacto knife 1 year before on his left arm. Blond and fair, John presented with blue-green hair. Prior to that, he reported that he had dyed it black several months before and then “cut it all off.” Subsequently, he dyed it orange, then purple, then pink. He was currently failing two subjects and getting a D in a third, and he reported that he was “smart but not interested in doing the work.” He had no prior contact with mental health professionals. John's school psychologist had referred him to the emergency room several days before, but John ran away, and this consultation was an effort to see “if there was any serious problem.” John's parents divorced when he was 18 months old, and he reported no subsequent contact with his father. John was the product of a full-term planned pregnancy with no complications in the newborn period. He was weaned from the bottle at 12 months of age, and other developmental milestones were reported normal. John lives with his maternal grandmother, mother, and older sister. In school, John did well through the 9th grade, but in the 10th grade his grades began to drop; in the 11th grade he did very poorly. As a senior, he has done hardly any work and there is concern that he may not graduate. He has no physical problems but does report allergies to pollen.

John reports chronic suicidal ideas for the past 2 years and cutting and burning himself on and off. He describes these self-mutilating behaviors as punishing himself for something wrong he has done and stops when he feels satisfied. John reports feeling comfortable only when depressed and feeling uncomfortable when not depressed and describes intense periods of crying that may shift into hysterical laughter. He describes several instances when he opens his eyes and does not know where he is for several minutes and then hears what people are saying but doesn't seem to make sense of it. He denies any history of physical or sexual abuse and denies drug or alcohol use.

A full-scale I.Q. of 116 placed him in the high average range and may underestimate his true potential. John is preoccupied with senses of worthlessness and guilt and accepts his sad fate of being chronically unhappy and empty as inevitable. He has significant difficulty with interpersonal relationships and lacks comfort with social skills. His social communications are likely to be strained. Interactions that involve sexual functioning are particularly distressing. He clings to avoidance stances to limit anticipated rejection. John reports “I am best when I am by myself.” He struggles with an ongoing sense of confusion regarding his life and his future.

Steven was a 3 ½-year-old white boy attending a therapeutic preschool, who lived with his mother, twin brother, and two older sisters. His mother was legally separated from his father. He was referred to a neurologist to rule out any neurological problems. The neurologist then referred Steven to the child and adolescent psychiatry outpatient department for psychiatric assessment. Steven and his mother were interviewed for 2 hours, and telephone contact was made with mother's therapist. The social worker at Steven's therapeutic preschool was contacted by telephone and forwarded clinical evaluations and psychological testing results.

“I don't know how to manage my son Steven. He has violent and aggressive behavior, like killing animals, kicking, punching, biting others, breaking furniture, smearing feces, banging his head and smashing things.” Mother said he had been aggressive from the beginning. Mother was married for 9 years and for the last 2 years has been legally separated. There had always been verbal and physical fights between his parents. Steven witnessed much aggression in the home. When he was 4 months old, his mother left him with his father and went to work; when she came back, he was “black and blue.” A similar incident occurred again. Mother described Steven as having a history of aggressive, impulsive, out-of-control and destructive behaviors. He spoils furniture at home, throws things, soils walls, and fights with his siblings. Steven killed some small animals such as hamsters and walked away without any concern or remorse. Mother also reported that Steven played with his feces and smeared it on the walls. When angry, Steven bangs his head. Mother was embarrassed to take him anywhere because he might fight with others or destroy property. According to the mother, Steven had been a difficult child from the beginning. He was attending a therapeutic preschool for his behavioral problems and, according to the mother and social worker, was not showing much improvement there.

He was physically healthy with no known medical problems. He was not on any medications. Steven had a history of head trauma at age 3. Computed tomography indicated a fractured skull. Steven's parents were legally separated, and for the last year his father had no contact with Steven or his siblings. Steven's father does not pay child support consistently. Mother disciplines Steven by yelling and restraining him by holding him. One maternal grandmother had diabetes. One of mother's uncles had multiple sclerosis. Steven's father is a substance abuser (alcohol, cocaine, marijuana, and heroin). The paternal grandmother was also an alcoholic. Patient's twin brother Frankie has similar but less severe problems. One sister has attention-deficit/hyperactivity disorder.

Steven's mother worked as a night shift cleaner and received food stamps. She was cooperative but very overwhelmed by Steven's behavior. Steven related age appropriately to interviewer. He participated in play for a while. He was cooperative and tolerated separation from his mother well. The therapist noted the boy's general appearance: casually dressed and looked stated age. His motoric behavior and coordination was mildly hyperactive, with no obvious incoordination. His attitude toward the examiner was that he related well and he was friendly. Steven made fair eye contact. He was alert, oriented to person, place, and time. Speech and language consisted of short sentences. His mood appeared neutral, with a full-range affect. Form of thought was coherent. Content of thought was violent: “When I get big, I am gonna kill you.” Child's perception of the problem was vague. Three wishes and sense of self could not be elicited. The patient was aggressive and impulsive. No obvious perceptual abnormality was observed. Apparent intellectual functioning was average.

Attention and concentration were limited. Memory was intact. His general fund of knowledge was age appropriate, and his abstraction abilities were concrete; however, his insight and judgment were limited. Steven exhibits aggressive, impulsive, hyperactive behavior, and poor impulse control. He witnessed and experienced severe family violence in the past and may be reenacting that behavior.

Pathology and Laboratory Examination Biological abnormalities have been described in individuals with personality disorders. Decreased 5-hydroxyindoleacetic acid concentrations in cerebrospinal fluid were correlated with increased impulsive-aggressive behavior. Challenge studies have suggested that serotonin dysregulation is associated with specific core symptoms found in patients with personality disorders. Results of research with pharmacological probes suggest serotonergic dysregulation in borderline personality disorder.

Differential Diagnosis Because borderline disorders lack specific criteria, a definition is needed. If the adult criteria for borderline personality disorder are being used, the differential can be made. Borderline personality disorder is often comorbid with depressive disorders, and when criteria for both are met, both may be diagnosed. Because the cross-sectional presentation of borderline personality disorder can be mimicked by an episode of depressive disorder, the clinician should avoid giving an additional diagnosis of borderline personality disorder on the basis of only cross-sectional presentation determining the pattern of behavior to be long-standing.

Other personality disorders may be confused with borderline personality disorder because they have certain features in common; thus these disorders must be distinguished on the basis of differences in their characteristic features. However, if an individual has personality features that meet criteria for one or more personality disorders in addition to borderline personality disorder, all can be diagnosed. Although histrionic personality disorder can also be characterized by attention seeking, manipulative behavior, and rapidly shifting emotions, borderline personality disorder is distinguished by self-destructiveness, angry disruptions in close relationships, and chronic feelings of deep emptiness and loneliness. Paranoid ideas or illusions may be present in both borderline personality disorder and schizotypal personality disorder, but these symptoms are more transient, interpersonally reactive, and responsive to external structuring in borderline personality disorder. Although paranoid personality disorder and narcissistic personality disorder may also be characterized by an angry reaction to minor stimuli, the relative stability of self-image as well as the relative lack of self-destructiveness, impulsivity, and abandonment concerns distinguish these disorders from borderline personality disorder.

Antisocial personality disorder and borderline personality disorder are both characterized by manipulative behavior, but individuals with antisocial personality disorder are manipulative to gain profit, power, or some other material gratification, whereas those with borderline personality disorder are directed more toward gaining the concern of caretakers.

Both dependent personality disorder and borderline personality disorder are characterized by fear of abandonment; however, the individual with borderline personality disorder reacts to abandonment with feelings of emotional emptiness, rage, and demands, whereas the individual with dependent personality disorder reacts with increasing appeasement and submissiveness and urgently seeks a replacement to provide caregiving and support. Borderline personality disorder can be further distinguished from dependent personality disorder by the typical pattern of unstable and intense relationships.

Borderline personality disorder must be distinguished from personality change due to a general medical condition, in which the traits emerge because of direct effects of a general medical condition on the central nervous system. It must also be distinguished from symptoms that may develop in association with chronic substance

use (e.g., cocaine-related disorder not otherwise specified).

Borderline personality disorder should be distinguished from identity problem, which is reserved for identity concerns related to a developmental phase (e.g., adolescence) and does not qualify as a mental disorder.

Disruptive disorders often need to be differentiated from borderline personality disorders. Attention-deficit/hyperactivity disorder appears to be a constitutional vulnerability or risk factor for development of borderline personality disorder. Physical and sexual abuse are common antecedents to both borderline personality and conduct disorders. Conduct disorders can also be associated with other personality disorders as well as depressive disorders and learning disorders.

Course and Prognosis The course of borderline personality disorder varies considerably. Most commonly there is chronic instability in early adulthood, with episodes of serious affective and impulsive dyscontrol and extensive use of health and mental health resources. Impairment from the disorder and the risk of suicide are greatest in the young-adult years and gradually wane with advancing age. During their 30s and 40s, most individuals with this disorder gain stability in their relationships and vocational functioning.

Treatment Borderline personality disorder in children and adolescents remains controversial and poorly understood despite efforts by clinicians to study this disorder. The heterogeneous symptomatology in children and adolescents with borderline personality disorder makes optimal treatment planning difficult. The entire range of psychiatric treatments available to child and adolescent psychiatrists have been adapted for youngsters with borderline personality disorder, including short-term inpatient psychiatric hospitalization, residential treatment, day hospital, psychopharmacotherapy, individual psychotherapy, group psychotherapy, and family therapy. Most treatment planning is based on clinical experience rather than data demonstrating efficacy for a particular treatment modality. Often youngsters with borderline personality disorder require brief psychiatric hospitalizations followed by a variety of outpatient psychiatric treatments. It is not uncommon for children and adolescents to be involved with a variety of systems of care including the mental health system, the special education system, the social service system, and the juvenile justice system, as well as numerous private practitioners.

To effectively plan for treatment of individuals with borderline personality disorder requires individualizing the treatment approach for each youngster. Some children and adolescents with borderline personality disorder may require extended treatment because short-term care is inadequate. Often the systems of care involved with an individual are poorly coordinated and there is growing awareness of the need to provide seamless systems of care with close coordination and case management to optimize outcome. Rigorous research in this area is lacking.

The primary goal of treatment of a child or adolescent with borderline personality disorder is to provide a safe therapeutic environment in which youngsters can work on their maladaptive defensive operations, continue their education, address their specific psychiatric symptoms, and improve their family functioning.

Inpatient Psychiatric and Residential Treatment Children and adolescents with borderline personality disorder often require inpatient psychiatric treatment because of concern about self-destructive, suicidal behavior. Psychiatric hospitalization affords an opportunity to provide a safe, therapeutic environment while comprehensive evaluation determines the most appropriate interventions and level of care. These youngsters can be particularly complex, and the therapeutic environment must be sensitive to their multiple needs and potential risks of self-destructive behavior.

Inpatient psychiatric care of these complex youngsters requires an empathic milieu with firm and consistent limits to contain the impulsive and maladaptive behaviors they often exhibit. The level of care that a child or adolescent with borderline personality disorder requires is often determined by the organization of the family, the capacity of the youngster to be safe, and the availability of ommunity resources and services to work with the family. Multiple treatment modalities are often required, and case management ensuring coordinated and integrated care has been clinically useful. Failure to coordinate the care results in a fragmented attempt at healing that often reflects the individual's own sense of personal fragmentation.

Children and adolescents with borderline psychopathology may unwittingly undermine every treatment attempt. Testing limits, splitting the treatment staff, and devaluating the treatment team present particular challenges for those involved with caring for these youngsters and their families. The therapeutic milieu must provide support, structure, and firm limits. Psychopharmacology, special education, and therapeutic activities all provide important adjuncts to the therapeutic program in hospitals. Individual, group, and family therapy become an important opportunity for therapeutic change.

The primary goal of psychiatric hospitalization is to prevent self-destructive behaviors. In addition, confronting the individual's maladaptive coping mechanisms can be done in the hospital while providing a safe environment. Hospitalization may present an opportunity to educate families about the severity of the symptoms and the need for treatment and to engage the child or adolescent in the therapeutic process.

Shorter stays, managed mental health care, and other financial issues are obstacles to obtaining appropriate inpatient psychiatric care for children and adolescents with all forms of psychopathology, including borderline personality disorder. Inpatient treatment may determine the need for ongoing residential care. The lack of available residential care often makes premature discharge to the community necessary. Certain youngsters who benefit from inpatient care and are able to return to the community to continue their treatment in less restrictive settings may benefit from day treatment or individual, group, and family psychotherapies. For most children and adolescents, inpatient psychiatric hospitalization provides an opportunity to stabilize a crisis, complete a thorough assessment, develop an alliance with the child and the family, and initiate appropriate treatment.

Individual Psychotherapy The literature describes the use of individual psychotherapy for the child or adolescent with borderline personality disorder. The primary task for the therapist is development and maintenance of a therapeutic relationship with the child. Confronting the child's maladaptive behaviors is generally avoided in an effort to maintain therapeutic alliance. Most of the literature discusses psychoanalytically oriented individual treatment, emphasizing that the psychotherapeutic approaches should be ego supportive. Others describe the importance of the therapist as an auxiliary ego and a stable figure available for internalization.

The beginning phase of individual psychotherapy aims at having the child or adolescent develop the notion that collaborative activity with the therapist may be useful. Clinical experience rather than empirical research suggests that one to two sessions per week may be necessary to foster the therapeutic alliance. Children and adolescents with borderline personality disorder may idealize their therapist and feel that they have found the perfect therapist. However, the initial phase of treatment may be marked by devaluation of the therapist and the child or adolescent's need for control of these treatments. It is generally accepted that therapists must clearly define the limits from the outset of treatment and provide the structure necessary to maintain the therapeutic alliance. Premature interpretation of negative affects is discouraged. Reality-oriented and supportive therapy is recommended. Working with children and adolescents with borderline personality disorder is particularly challenging for the therapist because of the strong feelings that are elicited in the countertransference. Therapists may feel manipulated, idealized, and seduced and must work against their own reaction to avoid acting out against their patients.

Kernberg underscored the importance of understanding these countertransference reactions as clues to the patient's internal world. Therapists can facilitate development of the therapeutic alliance by allowing the child or adolescent to save face. It is often helpful for the therapist to acknowledge the difficulty in relinquishing the familiar coping mechanisms. This may allow children and adolescents to begin to regain some control over their own lives.

The work of Marsha Linehan in adults has gained wide acceptance. Dialectical behavior therapy is, for the most part, the application of a broad array of cognitive and behavior therapy strategies to the problems of borderline personality disorder. This approach may have practical benefits when adapted to children and adolescents with borderline personality disorder.

Individual psychotherapy for children and adolescents with borderline personality disorder is an important mainstay in their overall recovery. The individual therapy must be incorporated into the appropriate level of care, based on the acuity of the patient's current clinical symptoms.

Family Therapy Family therapy for children and adolescents with borderline personality disorder can be important to the overall recovery of these children and adolescents. Maladaptive family relationships that may have led to the difficulties need to be pointed out in an effort to develop more-adaptive ways for the family members to relate to one another. In addition, a child or adolescent with borderline personality disorder produces enormous dysfunction in a family because of the stress that this disorder generates.

Cognitive behavioral approaches have been used to help parents learn to set more-consistent limits. Despite various theoretical and technical differences among family therapists, there is general agreement on strengthening parental competence. Parents often need assistance to enhance their ability to nurture and care for their children as well as to set effective and consistent limits. The family therapy of borderline youngsters during psychiatric hospitalization is important because this is

a time when parents generally feel out of control and incompetent. The therapeutic staff must develop an alliance with the parents and promote the relationship between the clinical staff and the parents to work together to facilitate the child's development. Failure of the therapeutic staff and the parents to work together may aggravate the child's pathological symptoms.

The family therapy that goes on as part of a multimodal treatment informs both the individual psychotherapist and the therapeutic milieu staff about the context in which the child or adolescent lives. Family involvement in the treatment can be crucial. Empirical studies on the efficacy of family therapy in the treatment of children and adolescents with borderline personality disorder are needed.

Psychopharmacotherapy Children and adolescents with borderline personality disorder have been treated with a variety of psychopharmacological agents addressing various key psychiatric symptoms. Antidepressant medication, stimulants, anxiolytic agents, antipsychotic agents, mood stabilizers, and anticonvulsants have all been used. Controlled studies have not been conducted with borderline children, although several have been conducted with borderline adults. The clinical literature suggests a rather broad consensus that specific pharmacological agents alone or in combination may be beneficial in the treatment of children and adolescents with borderline personality disorder. Some researchers advocate targeting specific symptoms; others have found the use of antipsychotic agents alone or in combination with other medications to be clinically useful. Currently, psychopharmacotherapy for children and adolescents with borderline personality disorder lacks empirical data. Clinically, psychotropic medication can decrease impulsivity and anxiety and elevate mood. Like other elements of the treatment of children and adolescents with borderline personality disorder, the meaning of medications for the child and the family must be examined. The paucity of empirical data to guide pharmacological intervention calls for more studies to determine the most effective medication treatment for children and adolescents with this disorder.

SUGGESTED CROSS-REFERENCES

Section 12.7 discusses the clinical features of schizophrenia. Section 14.6 discusses the clinical features of mood disorders. Section 46.2 discusses posttraumatic stress disorders. Section 19.3 discusses the clinical features of gender identity disorder. Chapter 22 and Chapter 23 discuss impulse-control disorders and adjustment disorders. Chapter 35 discusses learning disorders. Chapter 39 discusses attention-deficit disorders. Chapter 40 discusses disruptive behavior disorders. Chapter 45 discusses mood disorders and suicide in children, and Chapter 47 discusses schizophrenia with childhood onset. The psychiatric treatment of children and adolescents is discussed in Chapter 48. Section 49.4 discusses child sexual abuse.

SECTION REFERENCES

Adler G: *Borderline Psychopathology and Its Treatment*. Jason Aronson, New York, 1985.

*Bemporad JR, Smith HF, Hanson G, Cicchetti D: Borderline syndromes in childhood: Criteria for diagnosis. Am J Psychiatry *139:*596, 1982.

*Bernstein DP, Cohen P, Skodol A, Berziganian S, Brook JS: Childhood antecedents of adolescent personality disorders. Am J Psychiatry *153:*907, 1996.

Biederman J, Newcorn JH, Sprich S: Comorbidity of attention-deficit/hyperactivity disorder. In *DSM-IV Sourcebook,* vol 3, TA Widiger, American Psychiatric Association, editors. American Psychiatric Press, Washington, DC, 1997.

Blos P: The second individuation process of adolescence. Psychoanal Study Child *22:*162, 1967.

Clarkin JF, Sanderson C: Personality disorders. In *Psychopathology in Adulthood,* ed 2, AS Bellack, M Herson, editors. Allyn & Bacon, Boston, 1997.

Eisenberg N, Carlo G, Murphy B, Van Court P: Prosocial development in late adolescence: A longitudinal study. Child Dev *66:*1179, 1995.

Ekstein R, Wallerstein J: Observations on the psychology of borderline and psychotic children. Psychoanal Study Child *24:*307, 1956.

Erikson EH: The problem of ego identity. J Am Psychoanal Assoc *4:*428, 1956.

Famularo R, Kinscherff R, Fenton T: Posttraumatic stress disorder among children clinically diagnosed as borderline personality disorder. J Nerv Ment Dis *179:*428, 1991.

Friedrich WN: Normative sexual behavior in children. Pediatrics *88:*456, 1991.

Gabbard GO: An overview of countertransference with borderline patients. J Psychother Pract Res *2:*7, 1993.

Goldman SJ, D'Angelo EJ, DeMaro DR: Psychopathology in the families of children and adolescents with borderline personality disorder. Am J Psychiatry *150:*1832, 1993.

Goldstein WN: The borderline patient: Update on the diagnosis, theory, and treatment from a psychodynamic perspective. Am J Psychother *49:*317, 1995.

Goyer PF, Andreason PJ, Semple WE, Clayton AH, King AC, Compton-Toth BA, Schulz SC, Cohen RM: Positron-emission tomography and personality disorders. Am Coll Neuropsychopharmacol *10:*21, 1994.

Gunderson JG, Sabo AN: The phenomenological and conceptual interface between borderline personality disorder and PTSD. Am J Psychiatry *150:*19, 1993.

*Guzder J, Paris J, Zelkowitz P, Marchessault K: Risk factors for borderline pathology in children. J Am Acad Child Adolesc Psychiatry *35:*26, 1996.

*Guzder J, Paris J, Zelkowitz P, Feldman R: Psychological risk factors for borderline pathology in school-age children. J Am Acad Child Adolesc Psychiatry *38:*2, 1999.

Hartup WW: The company they keep: Friendships and their developmental significance. Child Dev *67:*1, 1996.

Hollander E, Stein DJ, Stein MB, Concetta M, DeCaria MS, Cohen L, Saoud MS, Skodol AE, Kellman D, Rosnick L, Oldham JM: Serotonergic sensitivity in borderline personality disorder: Preliminary findings. Am J Psychiatry *151:*277, 1994.

Kernberg O: *Borderline Conditions and Pathological Narcissism*. Aronson, Northvale, NJ, 1975.

Kernberg P: Resolved: Borderline personality exists in children under twelve. J Am Acad Child Adolesc Psychiatry *29:*478, 1990.

Linehan MM: *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. Guilford, New York, 1993.

Ludolph PS. Westen D, Misle B: The borderline diagnosis in adolescents: Symptoms and developmental history. Am J Psychiatry *147:*470, 1990.

Mahler M: Remarks on psychoanalysis with psychotic children. Q J Child Behav *1:*18, 1949.

Meeus W: Occupational identity development, school performance and social support in adolescence: Findings of a Dutch study. Adolescence *28:*809, 1993.

Nass ML: The superego and moral development in the theories of Freud and Piaget. Psychoanal Study Child *21:*51, 1966.

Newcomb AF, Bagwell C: Children's friendship relations: A meta-analytic review. Psychol Bull *117:*306, 1995.

Offer D, Offer JB: *From Teenage to Young Manhood: A Psychological Study*. Basic Books, New York, 1975.

Pelcovitz D, van der Kolk B, Roth S, Mandel F, Kaplan S, Resick P: Developmental of a criteria set and a structured interview for disorders of extreme stress (SIDES). J Trauma Stress *10:*3, 1997.

*Petti TA, Vela RM: Borderline disorders in childhood: An overview. J Am Acad Child Adolesc Psychiatry *29:*327, 1990.

Roth S, Newman E, Pelcovitz D, van der Kolk B, Mandel FS: Complex PTSD in victims exposed to sexual and physical abuse: Results from the DSM-IV field trial for post-traumatic stress disorder. J Trauma Stress *10:*539, 1997.

Shaffer D, Widiger TA, Pincus HA: Disorders usually first diagnosed in infancy, childhood, or adolescence (part II). In *DSM-IV Sourcebook,* vol 3, American Psychiatric Association, editors. American Psychiatric Press, Washington, DC, 1997.

Stein DJ, Simeon D, Frenkel M, Islam MN, Hollander E: An open trial of valproate in borderline personality disorder. J Clin Psychol *56:*506, 1995.

Stilwell BM, Galvin M, Kopta SM, Padgett RJ: Moral valuation: A third domain of conscience functioning. J Am Acad Child Adolesc Psychiatry *35:*230, 1996.

van der Kolk BA, Pelcovitz D, Roth S, Mandel FS, McFarlane A, Herman H: Dissociation, somatization and affect dysregulation: The complexity of adaptation to trauma. Am J Psychiatry *153*(Suppl):83, 1996.

Winnicott DW: *The Maturation Process and the Facilitating Environment*. International Universities Press, New York, 1991.

Zucker KJ, Green R: Psychological and familial aspects of gender identity disorder. Child Adolesc Psychiatric Clin North Am *2:*513, 1993.

49.11 ADOLESCENT SUBSTANCE ABUSE

OSCAR G. BUKSTEIN, M.D., M.P.H.

Psychiatrists, other mental health professionals, educators, and politicians are increasingly identifying substance use and abuse by children and adolescents as a critical public health problem. Despite attempts to limit access to psychoactive substances by young persons, the use of such substances is common among adolescents and growing in some groups. Substance use can have profound acute and chronic effects on the behavior and emotional status of adolescents. Substance abuse can affect the psychosocial development of adolescents. However, clinicians and researchers are only now developing an understanding of substance use disorders in adolescents and attempting to develop effective methods of treatment and prevention.

DEFINITION

Both substance abuse and dependence require a maladaptive pattern of substance use. Substance abuse criteria include use resulting in inability to meet major role obligations, legal problems, and an increase in risk-taking behaviors or exposure to hazardous situations. Criteria for substance dependence include the physiological criteria of tolerance and withdrawal symptoms, use resulting in abandonment of important activities, spending increasing time in activities related to substance use, using substances for a longer time than planned, and use despite the existence of worsening problems due to the substance use. The most common manifestation of substance use problems in adolescents is impairment in functioning, usually interpersonal or family conflict and academic failure. Withdrawal symptoms are much less common in adolescents than in adults. Other common symptoms include a reduction in activities because of use and using more than intended and tolerance.

COMPARATIVE NOSOLOGY

Like its predecessors, the third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III) and the revised third edition (DSM-III-R), the fourth edition (DSM-IV) substance use disorder criteria were developed for adults, with few empirical data to support their use for adolescents. The DSM-III diagnosis of substance dependence, especially alcohol dependence, could rarely be applied to adolescents because of the infrequent occurrence of clinically significant withdrawal symptoms. DSM-III substance abuse was defined in terms of pathological use and impairment in social or occupational functioning. The attribution of problems or consequences specifically to substance use is particularly problematic for adolescents whose coexisting psychiatric disorders and other psychosocial problems could also explain the impairment in functioning.

DSM-III-R and DSM-IV represent a substantial change from DSM-III. The definition of dependence was broadened to include clinically significant behaviors, cognitions, and nonphysiological symptoms that indicate substantial involvement with substance use. The broadened definition likely increased the number of adolescents meeting dependence criteria. DSM-IV differs from DSM-III-R in requiring clinically significant impairment or distress, switching the criteria of substance use resulting in failure to fulfill major role obligations from a diagnosis of dependence to one of abuse, and adding recurrent substance-related legal problems to the criteria for abuse.

The attribution of the many problems observed in adolescents who use substances remains a critical problem in the diagnosis of substance use disorders. Substance use per se should not be considered evidence of a substance use disorder. Almost all psychoactive substances are illegal for adolescents, and certain negative consequences may result from legal or cultural proscription rather than directly from the substance use or related behaviors.

EPIDEMIOLOGY

Alcohol is the substance of choice for most adolescents. According to the University of Michigan Monitoring the Future 1996 survey of high school students, 79.2 percent of twelfth graders, 71.8 percent of tenth graders and 55.3 percent of eighth graders report lifetime use. Although 26.8 percent of eighth graders report ever having been drunk, by the tenth and twelfth grades this percentage had risen to 48.5 and 61.8 percent, respectively. In terms of regular drinking, 30.2 percent of twelfth graders, 24.0 percent of tenth graders, and 15.6 percent of eighth graders report having five or more drinks on a single occasion in the preceding 2 weeks.

From 1991 to 1996 the proportion of high school students reporting the use of any illicit drug in the past 12 months increased from 27 to 40 percent for twelfth graders, from 20 to 38 percent for tenth graders, and from 11 to 24 percent for eighth graders. Marijuana accounted for much of this increase during the early 1990s. Almost 5 percent (4.9 percent) of twelfth graders reported daily marijuana use, and 36 percent reported use during the past year. Among eighth graders, 18 percent reported marijuana use within the past year.

The annual prevalence in the use of a number of other substances also rose in the early 1990s. In 1996, 4 percent of eighth graders and 9 percent of twelfth graders reported lysergic acid diethylamide (LSD) use in the preceding year. Annual prevalence rates for amphetamines were 9 percent for eighth graders, 12 percent for tenth graders, and 10 percent for twelfth graders. In 1996 2.8 percent of twelfth graders reported the use of methamphetamine (“ice”) and nearly 5 percent reported the use of methylenedioxymethamphetamine (MDMA, “ecstasy”) during the past year. The annual prevalence of other substances remained low in 1996. Heroin was used by 1.6 percent of eighth graders and 1 percent of twelfth graders. Cocaine was used by 4.9 percent of twelfth graders and 3 percent of eighth graders, while crack cocaine was used by 2.1 percent of twelfth graders and 1.8 percent of eighth graders during the preceding year.

The percentage of students reporting they had smoked cigarettes during the past 30 days rose by almost 50 percent (14 to 21 percent) in eighth graders from 1991 to 1996. Thirty percent of tenth graders and 34 percent of twelfth graders reported a 30-day prevalence of cigarette smoking in 1996. Daily cigarette use was reported by 22.2 percent of twelfth graders. Thirteen percent of twelfth graders reported smoking half a pack a day or more.

In general, male adolescents use substances of all kinds more than females do. Overall, white and Hispanic students are more likely than African-American students to report lifetime alcohol use and heavy episodic use. African-American students are more likely to report both lifetime and current marijuana use.

Few studies have examined the prevalence of substance use disorders in general population samples. Rates of lifetime alcohol abuse or dependence range from 5.3 percent in 15-year-olds to 32.4 percent in 17- to 19-year-olds. The lifetime prevalence of drug abuse or dependence ranges from 3.3 percent in 15-year-olds to 9.7 percent in 17- to 19-year-olds.

ETIOLOGY

While most adolescents will try alcohol and many will try other substances at some point during their adolescence, the risk for substance use and potential development of substance use disorders and related problems differs among adolescents and appears to be related to any number of risk factors. The age of onset of substance use is a prominent risk factor for early onset of substance use and rapid progression to the use of various illicit substances.

Risk factors for substance use and abuse can be related to individual characteristics such as the presence of early childhood disruptive behavior problems, aggressive behavior, poor academic performance, risk-taking behaviors, and favorable beliefs and attitudes about substance use. A number of biological characteristics also appear to place adolescents at risk for substance use and abuse, including increased latency of the P3 component of evoked potentials and

differences in static ataxia (body sway).

Temperamental deviations are associated with an increased risk for psychopathology and substance abuse. For example, children with a difficult temperament manifest externalizing and internalizing behavior problems by middle childhood and in adolescence more commonly than children whose temperament is normative. Increased behavioral activity is noted in both youths at high risk for substance abuse and those having a substance use disorder. Other temperamental trait deviations found in high-risk youth include shorter attention spans, increased impulsivity, and such negative affect states as irritability and emotionality.

The environment may account for many risk factors. Parental or family risk factors include parental beliefs and attitudes about substance use; tolerance of substance use; lack of closeness and attachment between parent and adolescent; lack of involvement, supervision, or discipline of the adolescent; and parental substance use. Peer-related variables such as peer substance use, peer attitudes, and a greater orientation of the adolescents to peers (as opposed to their parents) are also important risk factors. Community characteristics such as low socioeconomic status, high population density, physical deterioration, and high crime are also associated with higher levels of substance use in young persons.

The risk and natural history of adolescent substance use and related problems are best viewed from a developmental perspective. Several studies have established a common sequence, or stages, of substance use among adolescents. Most adolescents eventually try “gateway” drugs such as alcohol or cigarettes, which are legal for adults and thus easily accessible by adolescents. Progressively fewer adolescents progress to use of marijuana or other illicit substances and more serious substance use. Different risk factors appear to be more critical at different stages of substance use. For example, the adolescent's initial experiences with substances usually take place in a social context and involve peer-related factors. Later stages of substance involvement are determined by other major categories of risk factors such as parental and individual risk factors and psychological distress. Substance use and deviant and other risk-taking behaviors commonly occur together in adolescents. Problem behavior theory suggests that substance use, deviant behavior, deviant attitudes, and environments that promote these behaviors cluster together.

DIAGNOSIS AND CLINICAL FEATURES

The optimal assessment of the adolescent suspected of having a substance use disorder involves not only substance-use-related behaviors but also other domains of psychosocial functioning, including psychiatric and behavior problems, school or vocational functioning, family functioning, social competency and peer relationships (including social skills), leisure and recreational activities, and current medical status.

Substance use is a multidimensional behavior that requires thorough evaluation of several dimensions of substance use behavior in addition to quantity and frequency of substance use. Critical dimensions of substance use behavior include *pattern of use* (i.e., quantity, frequency, onset, and types of agents used), *negative consequences* (i.e., school/vocational, social/peer/family, emotional/behavioral, legal, and physical), *context of use* (i.e., time/place, peer use/attitudes, mood antecedents, consequences, expectancies, and overall social milieu) and *control of use* (i.e., view of use as a problem, attempts to stop or limit use, other DSM-IV dependence criteria).

Because substance use is usually an overt behavior, the clinician must often depend on the youth's self-report of substance use, related behaviors, and resulting problems. Although self-reports appear to be reliable in some populations, specific populations such as extremely antisocial youths may report less use than drug clinic samples. The clinician may attempt to substantiate suspected use by reports from third parties or through the use of urine or blood toxicology. Toxicology tests detect the presence of a substance, not a pattern of abuse or dependence, but they may be a useful check on the truthfulness of an adolescent's self-report as well as an ongoing check of substance use during and after treatment.

A variety of instruments are available and others are being developed to assist in screening and detailed assessment of substance use and related behaviors and problems. The common use of instruments developed for adults or those that lack standardization and validation may be inappropriate for adolescents.

DIFFERENTIAL DIAGNOSIS AND COMORBIDITY

A number of psychiatric disorders are commonly associated with substance use disorders in youths (Table 49.11-1). Many adolescents who use substances also display a variety of other deviant behaviors. It may be difficult to attribute dysfunction to a particular deviant behavior such as substance use rather than to other deviant behaviors, the entire syndrome of deviant behavior (i.e., conduct disorder), or other psychiatric disorders. Conduct disorder and constituent criteria such as aggression usually precede and accompany adolescent substance use disorders. Clinical populations of adolescents with substance use disorders show rates of conduct disorder ranging from 50 to almost 80 percent. Although attention-deficit/hyperactivity disorder is commonly noted in substance-using and substance-abusing youths, the observed association is likely due to the high level of comorbidity between conduct disorder and attention-deficit/hyperactivity disorder. Earlier onset of conduct problems and aggressive behavior in addition to the presence of attention-deficit/hyperactivity disorder may increase the risk for later substance abuse.

Disruptive behavior disorders
Conduct disorder
Oppositional defiant disorder
Attention-deficit/hyperactivity disorder
Mood disorders
Dysthymic disorder
Cyclothymic disorder
Major depressive disorder
Bipolar disorder
Anxiety disorders
Posttraumatic stress disorder
Social phobia
Bulimia nervosa
Schizophrenia

Table 49.11-1 Psychiatric Disorders Commonly Comorbid With Substance Use Disorders in Adolescents

Onset of mood disorders, especially depressive disorders, frequently precedes or follows the onset of substance use and substance use disorders in adolescents. The point prevalence of depressive disorders in these studies ranged from 24 to over 50 percent.

The literature supports substance use disorders among adolescents as a risk factor for suicidal behavior, including ideation, attempts, and completed suicide. Possible mechanisms for this relation include acute and chronic effects of psychoactive substances. Adolescent suicide victims are frequently using alcohol or other drugs at the time of suicide. Short-term substance use may produce transient but intense dysphoric states, disinhibition, impaired judgment, and increased level of impulsivity or may exacerbate preexisting psychopathology including depression or anxiety disorders.

Aggressive behaviors are present in many adolescents with substance use disorders. Consumption of such substances as alcohol, amphetamines, and phencyclidine (PCP) may increase the likelihood of subsequent aggressive behavior. The direct pharmacological effects resulting in aggression may be further exacerbated by the presence of preexisting psychopathology, the use of multiple agents simultaneously, and the frequent relative inexperience of the adolescent substance user.

A number of studies of clinical populations show high rates of anxiety disorders among youth with substance use disorders. In clinical populations of adolescents with substance use disorders, the prevalence of anxiety disorder ranged from 7 to over 40 percent. The order of appearance of comorbid anxiety and substance use disorders appears to vary, depending on the specific anxiety disorder. Social phobia usually precedes abuse, while panic and generalized anxiety disorder may more often follow the onset of substance use disorders. Adolescents with substance use disorders often have a history of posttraumatic tress disorder. Bulimia nervosa is also common in adolescents with substance use disorders. As suggested by studies showing language deficits in youth affected by or at high risk for substance use disorders, learning disorders may also show an increased incidence of comorbidity. Multiple psychiatric disorders, both internalizing and externalizing types, are often noted in populations of adolescents with substance use disorders.

Psychiatric disorders in childhood, disruptive behavior disorders as well as mood or anxiety disorders, increase the risk for development of substance use disorders in

adolescence. The etiological mechanisms have not been systematically researched. A number of possible relationships exist between substance use disorders and psychopathology. Psychopathology may precede substance use disorders, may be a consequence of preexisting substance use disorders, may influence the severity of substance use disorders, may not be related, or may originate from a common vulnerability.

COURSE AND PROGNOSIS

Most adolescents who use substances do not go on to develop substance use disorders in adolescence or adulthood. Generally, levels of substance use peak in late adolescence or early adulthood. Life events such as education, career attainment, marriage, and parenthood tend to decrease or stop substance use. However despite such maturational processes, substance use can disrupt the ability of adolescents to negotiate the developmental tasks of adolescence and young adulthood.

Psychoactive substances by virtue of their pharmacological properties affect the mood, thought processes, and sensory perception of their users. Adolescent substance users are more likely to be novice users with minimal tolerance, and they may experience more noxious or adverse reactions to substances than more experienced adult substance users. Inexperienced adolescent substance users may not appreciate the extent of their impairment from use or intoxication.

Substance use by adolescents has a number of short-term consequences. Accidents and trauma, including driving motor vehicles or bicycles while intoxicated, and drownings are common in adolescents who are using substances. Adolescent perpetrators of violence and their victims are commonly under the influence of, or have histories of, substance use disorders. Early onset of sexual behavior, another high-risk behavior, is more common among adolescents with early-onset substance use. Although adolescents constitute a relatively small percentage of individuals with human immunodeficiency virus (HIV) disease, the frequently long latency between HIV infection and the onset of symptoms suggests that many may have become infected as adolescents.

TREATMENT

Research on the outcomes of adolescent substance abuse treatment has lagged behind research on the predictors, course, and correlates of treatment outcome among adults. Unfortunately, data indicate that most adolescents return to some level of alcohol or other drug abuse following treatment. Despite these outcomes, studies have identified specific predictors of treatment outcome, including patient or adolescent characteristics, social support system variables, and program characteristics. Adolescents in substance abuse treatment begin substance use at an earlier age, progress rapidly to the use of “hard” drugs, and usually use multiple drugs. Other clinical features of adolescents entering treatment include often high levels of coexisting psychopathology or early personality difficulties, deviant behavior, school difficulties including high levels of truancy, and family disruption and substance abuse. Several pretreatment characteristics predict completion of treatment by adolescents: more severe alcohol problems; greater use of drugs other than alcohol, marijuana, and tobacco; a higher level of internalizing problems; and lower self-esteem. Premorbid psychopathology such as conduct disorder is negatively correlated with treatment completion and future abstinence. Although factors such as severity of substance use may predict short-term treatment outcomes, most longer-term outcomes may depend on social and environmental factors. This is consistent with studies that suggest that relapse in adolescents is more often associated with social pressures to use rather than situations involving negative affect, as is usually found in adult relapse. Attendance at self-support or aftercare groups is associated with higher rates of abstinence and other measure of improved outcome than those in adolescents who did not attend such groups. Table 49.11-2 outlines current approaches for treating alcoholism in adolescents.

Approach	Components and/or Focus
12-Step model	Group meetings (e.g., Alcoholics Anonymous) Group counseling Worksheets Family sessions or groups
Behavioral intervention	Relapse prevention Development or improvement of refusal skills, social skills, and problem-solving skills Anger control training Leisure-time management
Family therapy Behavioral	Parent management training Contingency contracting
Strategic-structural	Restructure maladaptive patterns
Systemic	Address behavioral limit-setting and intergenerational conflicts
Contextual	Integration of emotional, behavioral, cognitive, and spiritual aspects of family
Mixed	A combination of various family therapy components
Educational and vocational assistance/rehabilitation	Achieve basic level of academic skills and completion of a secondary education
Medications for coexisting psychiatric disorders Depression and other mood disorders	Antidepressants and lithium
Attention-deficit/hyperactivity disorder	Stimulants

Table 49.11-2 Current Approaches and Their Components for Alcoholism Treatment in Adolescents

Despite a higher level of return to substance use among adolescents after treatment, abstinent teens may expect decreased interpersonal conflict, improved academic functioning, and increased involvement in social and occupational activities. Patterns of substance abuse among adolescents appear to become more stable between 6 and 12 months after treatment. Reviews of treatment outcome conclude that treatment can be effective and is certainly better than no treatment.

A number of treatment program characteristics are associated with improved abstinence and lower levels of relapse. They include longer duration of treatment, available follow-up or after-care treatment, family involvement, and the availability of social support services.

Until recently, there were few data to establish the superior effectiveness of a particular treatment modality over alternative treatment modalities. Several investigators have demonstrated the effectiveness of family-based approaches—in comparison with nonfamily-based interventions—in treating adolescent substance use disorders. Scott Henggeler and associates' multisystemic therapy represents a family-ecological systems approach with interventions targeting family functioning and communication, school and peer functioning, and community functioning. In addressing the frequently multidimensional adolescent substance use disorders and associated problems, multisystemic therapy focuses on multiple targets and using many treatment techniques including family therapy and behavioral therapies. Multisystemic therapy is integrative and comprehensive but targets specific, identified areas of dysfunction.

A number of cognitive-behavioral treatments, including many previously used with adults (e.g., relapse prevention), may show promise as treatment modalities for adolescents with substance use disorders. Problem-solving and social skills training appear to be effective for populations of conduct-disordered youth.

Many treatment programs for adolescents continue to be based on the 12 steps of Alcoholics Anonymous (AA) or Narcotics Anonymous (NA). Attendance at self-support groups appears to predict abstinence and other measures of improved treatment outcome.

Pharmacotherapy can potentially treat withdrawal, counteract or decrease the subjective reinforcing effects of illicit substance use, and treat comorbid psychopathology. Unfortunately, there has been no systematic research evaluating the efficacy and safety of any psychotropic medication in the treatment of adolescents with substance use disorders. Although clinically significant withdrawal symptoms appear to be rare in adolescents, little rationale exists for using detoxification protocols different from those used for adults. The use of agents to block the reinforcing effects of various substances, aversive agents (e.g., disulfiram [Antabuse]) or those that relieve craving during and after acute withdrawal has been studied in adults but has received scant attention in adolescents. Aversive pharmacological treatment with agents such as disulfiram is rare in adolescents.

The high prevalence of coexisting psychiatric disorders in adolescents with substance use disorders presents additional targets for pharmacological agents. Potential targets for pharmacological treatment include depression and other mood problems, attention-deficit/hyperactivity disorder, severe aggressive behavior, and anxiety disorders. Unfortunately, few data in the literature demonstrate the efficacy of pharmacological agents prescribed for adolescents with substance use disorders and comorbid psychiatric disorders. However, a recent study indicates that these agents have limited clinical utility. In general, clinicians should use the same caution in considering pharmacological treatment for adolescents with comorbid substance use disorders and psychiatric disorders as they do with adolescents with psychiatric symptoms alone.

Substance use disorders may increase the potential for intentional or unintentional overdose. Some pharmacological agents may have inherent abuse potential. Critical issues in the use of pharmacotherapy include (1) avoiding the precipitation or exacerbation of psychiatric symptoms by the abused substances, (2) the need to achieve some level of abstinence or control of substance use before making more optimal assessment of symptoms and starting pharmacological treatment, (3) the potential of acute drug effects resulting in intentional or unintentional overdose, and (4) the potential abuse of the pharmacotherapeutic agents themselves. The treatment of attention-deficit/hyperactivity disorder in those with substance use disorder remains problematic because of the abuse potential of central nervous system

(CNS) stimulants by the patient, family, and peers. In addition to close supervision of medication compliance, clinicians should consider the use of effective agents with much lower abuse potential such as the tricyclic drugs, bupropion (Wellbutrin), and pemoline (Cylert).

Although not specifically studied, the multiple areas of possible dysfunction in adolescents with substance use disorders and the many available treatment modalities suggest a multimodal approach. Treatment matching, or matching patients with specific characteristics with appropriate levels of care and types of treatment modalities, received much attention in the adult literature. Psychiatric severity may be the best-identified guide to matching patient characteristics with specific treatment modalities.

Prototypical Treatment Increasingly, over the past decade, the previous emphasis on residential treatment for adolescents with substance use disorders has been replaced by less costly and less intensive alternatives such as outpatient treatment or intensive outpatient treatment. Residential stays are becoming much more variable, with significantly more-limited goals for treatment. In intensive outpatient treatment, the adolescent attends a program several afternoons or evenings a week while continuing to reside at home and attend school. The predominant treatment approach continues to be based on the 12-step approach of AA or NA. Each treatment setting offers the same modalities, but at a different level of intensity or frequency. Residential treatment, while more costly, offers more structure and total control of the adolescent's environment and thus access to substances of abuse. Further research is needed to determine whether residential treatment has any advantage over community-based treatments.

Typically, 12-step programs focus on several issues, confronting the adolescent's denial about substance use as a problem and teaching the adolescent about the process of recovery. The schedule is usually highly structured. Most of the treatment is carried out in group formats where adolescents may challenge one another about the veracity of their statements or therapeutic work. The content of group therapy can include 12-step work, family issues, and psychoeducation about the effects of substances and substance use and the process of recovery including relapse and relapse prevention. Attendance at AA or NA meetings and urine screening for the presence of illicit substances are common features. Adolescents may spend some time in individual therapy either working on 12 steps or other assignments or on specific problems such as depression or physical or sexual abuse victimization. Family therapy in both individual and multifamily formats is part of most programs. Increasingly, cognitive-behavioral elements such as anger control or social skills are being included in treatment, under the general guise of coping skills. Many, but not all programs have psychiatric consultation available to assess other possible comorbid psychopathology and to provide psychopharmacotherapy for such problems as attention-deficit/hyperactivity disorder, depression, or bipolar I disorder.

Another treatment controversy is the use of toxicological methods on bodily samples, usually urine, to detect substances. Technology being developed may extend screening to hair samples and using urine screening to detect recent alcohol use. The use of urine screening at initial assessment may be limited, as a positive result only indicates that the adolescent has used one or more substances and not whether significant problems with use exist or how much or frequently an adolescent is using one or more substances. Urine drug screens, especially relatively inexpensive radioimmune assay kits, may be valuable adjuncts in follow-up to detect ongoing substance use by the adolescent. Usually screens are requested by therapists at random intervals for at least several months following the adolescent's entry into treatment. Adolescents, therapists, and family should be aware of the consequences of both “dirty” (those indicating the detection or presence of one or more illicit substance) and “clean” (absence of any illicit substance) urine samples prior to the onset of a screening program.

Fred is a 16-year-old male admitted to substance abuse treatment for the second time, following a relapse and threats of suicide. He was initially admitted to an inpatient program following a serious suicide attempt. He reported a long history of disruptive behavior and academic failure since childhood. He was increasingly truant and difficult for his family to control. During his first treatment episode, he reported an onset of substance use at age 11 years, a rapid progression in substance involvement since age 13 years, then current use of marijuana on a daily basis, drinking alcohol up to several times a week, frequent trips on LSD, and experimentation with a variety of substances. Fred attended group sessions focusing on his initial denial of a substance use problem and then learned the process of recovery while attending other groups and AA and NA meetings. Family group sessions showed him and his parents the need for better communication and more adaptive interactions. Fred gradually responded to the structure of the treatment program, although he had frequent problems with anger control when confronted by peers or staff or when frustrated. Depressive symptoms failed to remit following 2 weeks of abstinence, and Fred was given fluoxetine (Prozac). He showed a rapid improvement in mood and treatment compliance. Upon discharge, he was attending NA meetings and outpatient therapy. However, family conflict soon recurred, and Fred became noncompliant with outpatient treatment, medication, and meetings. He resumed old relationships with deviant peers and relapsed into daily marijuana use and occasional alcohol use.

Prevention The most effective intervention for adolescent substance use or use disorders is preventing the initial development of substance use or pathological patterns of use. Although a consensus exists regarding the importance of prevention efforts, there is a lack of agreement of a conceptual goal for prevention; that is, what should be prevented—use, abuse, or dependence? While some would have interventions directed at the prevention of various use patterns or consequences of substance use, a broader view of prevention would target the risk factors for the development of substance use or substance use disorders. Targeting risk factors likely requires a host of measures involving the educational, mental health, and welfare systems and is certainly more costly than more direct interventions for substance use behaviors. However, primary prevention efforts aimed at risk factors may prove more effective and influence the development of a wide range of problems and psychosocial dysfunction in children, adolescents, and their families.

Prevention efforts are based on various theoretical models of adolescent substance use and abuse development. Most prevention interventions are based on social learning models. If one can change what young persons are exposed to and what they learn from their environment, then behavioral changes will follow. These interventions include educational approaches, family-based interventions, and community-based projects. Educational approaches include three basic methods: (1) knowledge and attitude, (2) values and decision making, and (3) social competency or skills. Although educational, individually focused, family-focused (e.g., parent training), and community-focused (e.g., advocacy groups, media campaigns and regulatory changes) prevention interventions are an increasing part of the total prevention effort, the variety of critical risk factors for the development of adolescent substance use and abuse suggests the targeting of multiple risk factors or influences as part of comprehensive prevention efforts. The successful prevention programs appear to be those that target salient risk factors, are skills oriented, have sufficient intensity and duration, have follow-up, and respect the socioeconomic and cultural realities of the targeted communities.

SUGGESTED CROSS-REFERENCES

Specific substance-related disorders with diagnostic tables are covered in Chapter 11. Conduct disorder is discussed in Chapter 40, and childhood mood disorders are discussed in Chapter 45. Family therapy is covered in Section 48.5, and cognitive-behavioral treatments are covered in Section 48.3.

SECTION REFERENCES

*Biederman J, Wilens T, Mick E, Faraone SV, Weber W, Curtis S, Thornell A, Pfiser K, Jetton JG, Soriano J: Is ADHD a risk factor for psychoactive substance use disorders? Findings from a four-year prospective follow up study. J Am Acad Child Adolesc Psychiatry *36:*21, 1997.

*Blood L, Cornwall A: Pretreatment variables that predict completion of an adolescent substance abuse treatment program. J Nerv Ment Disease *182:*14, 1994.

Brook JS, Cohen P, Brook DW: Longitudinal study of co-occurring psychiatric disorders and substance use. J Am Acad Child Adolesc Psychiatry *37:*322, 1998.

Brown SA: Recovery patterns in adolescent substance abuse. In *Addictive Behaviors across the Life Span,* JS Baer, GA Marlatt, RJ McMahon, editors. Sage, Newbury Park, CA, 1993.

Bukstein OG: Substance abuse. In *Handbook of Aggressive and Destructive Behavior in Psychiatric Patients,* M Hersen, RT Ammerman, LA Sisson, editors. Plenum, New York, 1994.

*Bukstein OG: *Adolescent Substance Abuse: Assessment, Prevention, and Treatment*. Wiley InterScience, New York, 1995.

Bukstein OG, Brent DA, Kaminer Y: Comorbidity of substance abuse and other psychiatric disorders in adolescents. Am J Psychiatry *146:*1131, 1989.

Bukstein OG, Kaminer Y: The nosology of adolescent substance abuse. Am J Addict *3:*1, 1994.

Bukstein OG, Van Hasselt VB: Alcohol and drug abuse. In *Handbook of Behavior Therapy in the Psychiatric Setting,* AS Bellack, M Hersen, editors. Plenum, New York, 1993.

Bulik CM, Sullivan PF, Epstein LH, Weltzin T, Kaye W: Drug use in women with anorexia and bulimia. Int J Eating Disord *11:*213, 1992.

Clark DB, Kirsci L, Moss HB: Adolescent versus adult onset and the development of substance use disorders in males. Drug Alcohol Depend *49:*115, 1998.

Clark DB, Pollack N, Bukstein O, Mezzich AC, Bromberger J, Donovan JE: Gender and comorbid psychopathology in adolescents with alcohol dependence. J Am Acad Child Adolesc Psychiatry

*36:*1195, 1997.

Clark DB, Sayette MA: Anxiety and the development of alcoholism. Am J Addict *2:*56, 1993.

Crumley FE: Substance abuse and adolescent suicidal behavior. JAMA *263:*3051, 1990.

Fleisch B: *Approaches in the Treatment of Adolescents with Emotional and Substance Abuse Problems*. DHSS publ no. (ADM) 91-1744, U.S. Government Printing Office, Washington, DC, 1991.

Hawkins JD, Catalano RF, Miller JY: Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. Psychol Bull *112:*64, 1992.

Henggeler SW, Melton GB, Smith LA: Family preservation using multi-systemic therapy: An effective alternative to incarcerating serious juvenile offenders. J Consult Clin Psychol *66:*953, 1992.

Kaminer Y: Issues in the pharmacological treatment of adolescent substance abuse. J Child Adolesc Psychopharmacol *5:*93, 1995.

Kandel D: Stages in adolescent involvement in drug use. Science *190:*912, 1975.

Kumpfer KL: Prevention of alcohol and drug abuse: A critical review of risk factors and prevention strategies. In *Prevention of Mental Disorders, Alcohol and other Drug Use in Children and Adolescents,* D Shaffer, I Phillips, NB Enzer, editors. U.S. Department of Health and Human Services, Rockville, MD, 1989.

Liddle HA, Dakof GA: Family-based treatment for adolescent drug use: State of the science. In *Adolescent Drug Abuse: Clinical Assessment and Therapeutic Interventions,* E Rahdert, D Czechowicz, editors. National Institute on Drug Abuse Research Monograph 156, NIDA, Rockville, MD, 1995.

Loeber R: Natural histories of conduct problems, delinquency and associated substance use. In *Advances in Clinical Child Psychology,* vol 11, BB Lahey, AE Kazdin, editors. Plenum, New York, 1988.

Martin CS, Kaczynski NA, Maisto SA, Bukstein OG, Moss HB: Patterns of DSM-IV alcohol abuse and dependence symptoms in adolescent drinkers. J Stud Alcohol *56:*672, 1995.

*Martin CS, Winters KC: Diagnosis and assessment of alcohol use disorders among adolescents. Alcohol Health Res World *22:*95, 1999.

Newcomb MD, Bentler PM: *Consequences of Adolescent Drug Use*. Sage, Newbury Park, CA, 1988.

Reinherz HZ, Giaconia RM, Lefkowitz ES, Pakiz B, Frost AK: Prevalence of psychiatric disorders in a community population of older adolescents. J Am Acad Child Adolesc Psychiatry *32:*369, 1993.

Tarter RE, Kirisci L, Hegedus A, Mezzich A, Yanyukov M: Heterogeneity of adolescent alcoholism. Ann NY Acad Sci *708:*172, 1994.

University of Michigan: *1996 Monitoring the Future Survey*. Institute for Social Research, Ann Arbor, MI, 1996.

*Wagner EF, Brown SA, Monti PM, Myers MG, Waldron HB: Innovations in adolescent substance abuse intervention. Alcohol Clin Exp Res *23:*236, 1999.

*Weinberg NZ, Rahert E, Colliver JD, Glantz MD: Adolescent substance abuse: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *37:*252, 1998.

Wilens TE, Biederman J, Abrantes AM, Spencer TP: Clinical characteristics of psychiatrically referred adolescent outpatients with substance use disorders. J Am Acad Child Adolesc Psychiatry *36:*941, 1997.

Winters KC, Stinchfield RD: Current issues and future needs in the assessment of adolescent drug abuse. In *Adolescent Drug Abuse: Clinical Assessment and Therapeutic Interventions*, E Rahdert, D Czechowicz, editors. National Institute on Drug Abuse Research Monograph 156, NIDA, Rockville, MD, 1995.

49.12 FORENSIC CHILD AND ADOLESCENT PSYCHIATRY

STEPHEN P. HERMAN, M.D.

Forensic child and adolescent psychiatry has come of age. Born of an inchoate juvenile justice system a century ago, tempered by a profound realism about the limits of psychoanalysis to cure the ills of society, and matured by endless requests for psychiatric assistance in courts throughout the land, forensic child psychiatry has finally come into its own as an important and exciting subspecialty. Child and adolescent psychiatrists are now frequently involved in conducting medico-legal evaluations in custody and other child placement cases, criminal matters, allegations of sexual and physical abuse, competence of child witnesses, damages arising from psychological and physical trauma, school-related cases, and other circumstances. Although most child and adolescent psychiatrists prefer to avoid this work and all its complications and special anxieties, few escape involvement in some aspect of the legal system, and all at some time or another will have patients who themselves are caught in its grasp.

Organized medicine and medical educators have accepted the need for training and demonstrated competence in forensic psychiatry. In 1988 the Accreditation Council for Graduate Medical Education formally included forensics as a required part of training in child and adolescent psychiatry. In 1992 the field was officially recognized as a subspecialty of psychiatry by the American Board of Medical Specialties. The American Board of Psychiatry and Neurology held its first examination for added qualifications in forensic psychiatry in 1995. Some knowledge of this area is required of all child and adolescent psychiatrists, whether or not they choose to become forensic practitioners.

HISTORY

Forensic child and adolescent psychiatry, initially a stepchild of the child guidance movement of the late nineteenth and early twentieth centuries, evolved as a natural response to the public's concern about juvenile crime. Until this enlightened era, children had no special rights under law, were easily exploited, were subject to criminal proceedings identical to those of adults, and had no special legal venue. The country's first juvenile court was established in Chicago in 1899. Maine, the last state to create such a court, did not do so until 50 years later. Thus, a system of juvenile and family courts, which is taken for granted today in the United States, is actually a relatively recent development.

For most of this century, those psychiatrists working as consultants within family courts had little or no formal training in child psychiatry and the law. They learned by doing; there were no published standards for performing various forensic examinations, little attention was paid to ethical considerations, and scholarly publications were scarce.

Fortunately, the field has matured, primarily in the past 25 years, into a discipline with a significant body of literature, practitioners who have received advanced training in forensic fellowship programs, and published standards and practice parameters that have greatly improved the quality of the work done in this area. Unfortunately, of the nearly 30 national fellowship programs offering advanced training in forensic psychiatry, only a few specialize in children and adolescents. And although all child and adolescent psychiatry training programs must now provide some exposure to forensics, limited availability of experienced teachers and mentors and few working liaisons with the legal system mean that most child psychiatrists have extremely limited exposure to this field.

LEGAL SYSTEM

Forensic child psychiatry operates at the interface of clinical psychiatry and the legal system. The field removes child psychiatrists from the safe haven of the office and often places them in the arena of the adversarial process—a setting quite foreign, sometimes confusing, and often anxiety provoking. Clinicians choosing to practice forensics, therefore, must have some background knowledge of the legal system.

In the United States sources of law are federal and state statutes and case law stemming from court decisions. State courts determine the common law for each state, and federal courts are empowered by the Constitution and any additional laws passed by Congress. The federal court system comprises district courts (which conduct trials), federal courts of appeal, and the United States Supreme Court. State court systems vary but usually comprise trial courts, an appellate division, and the state's supreme or highest court. States may also have specialized courts, such as courts of probate or family court.

Child and adolescent psychiatrists engaged in forensic practice are frequently involved in judicial proceedings that are either civil or criminal, decided by a judge or a jury. Criminal cases involve serious offenses punishable by loss of liberty (e.g., imprisonment, execution) or by a fine. Because of the serious ramifications of having the state restrict personal liberty, criminal convictions require the highest standard of proof, *beyond a reasonable doubt*. Civil litigation includes negligence or malpractice cases and family law issues such as child custody disputes and termination of parental rights. The burdens of proof in civil cases are lower than in criminal cases: *a preponderance of evidence* and *clear and convincing evidence*.

Within this legal framework, the rights of children have evolved slowly but surely in the United States. From being originally viewed as a valuable and productive asset of their parents in the labor they could perform, to being protected as needy individuals by the strength of the government *(parens patriae),* to having the United States Supreme Court finally recognize that they are protected by the Bill of Rights and the Fourteenth Amendment, children have at last been recognized as having fundamental rights that must be protected and maintained. Against this backdrop the forensic child psychiatrists practice their specialty.

Finally, forensic practice differs from therapeutic practice in a significant way: in general, the patient's rights of confidentiality and privilege, usually protected (with some exceptions) during psychotherapy, are waived during a forensic evaluation. By definition, the forensic evaluation is part of a legal process. If a report is written, others will read it. The forensic evaluator may have to testify. The forensic psychiatrist explains this to the subject of the evaluation before it begins.

FAMILY LAW

Child Custody Child custody disputes are probably the most common legal proceedings involving child and adolescent psychiatrists. Of the one million children a year who experience divorce in the United States, approximately 10 to 15 percent are in litigated cases. A child psychiatrist who is not routinely evaluating such cases has most likely treated a child or adult going through this complex and emotionally draining process.

Child custody disputes throughout recorded history have reflected a society's view of the child in the family. The practice of courts becoming involved in such private family affairs is fairly recent. The history of this issue traces movement from seeing children as essentially owned by their father to considerations of what is in their best interests. Judicial decisions have been informed by various presumptions, such as “the tender years” and “the best interests of the child.” The tender years presumption existed well into the twentieth century. It held that young children (from birth to about age 7) were usually better off with their mothers, who were generally assumed to be better skilled in nurturing and raising their offspring. This presumption was replaced in the last third of the twentieth century by “the best interests of the child,” which is the current law in the United States. This presumption holds that the focus of a child custody case should be the child, and courts ought not lean toward one parent or the other strictly on the basis of sex. Further, there may be situations in which custody should reside with a nonparent rather than a parent, if that represents the best interests of the child. To help courts determine a child's best interests, judges have increasingly turned to child psychiatrists to evaluate families and make recommendations.

Child custody cases, complex and emotion-laden in themselves, have become even more complicated by special issues that commonly arise. These can include a mentally ill parent, a homosexual parent, a stepparent or grandparent seeking custody, parental kidnaping, allegations of sexual abuse, relocation cases, and cases arising from advances in reproductive technology.

In general, child psychiatrists ought to be court-appointed, impartial experts when evaluating such cases. Serving as a neutral evaluator avoids the shadowy status of hired gun and gives the clinician the greatest credibility in the eyes of the court. Recently, the American Academy of Child and Adolescent Psychiatry published Practice Parameters for Child Custody Evaluation, which sets forth guidelines for this forensic evaluation. The psychiatrist sees the litigating adults and the children, separately and in conjoint interviews, holds collateral interviews with important figures such as therapists and teachers, writes a cogent and complete report setting forth the data gathered and how that data led to the clinician's conclusions, and sometimes defends the report by testifying in court.

Adoption Adoption continues to be a major avenue for becoming a parent; and, as with society itself, this process has undergone much change. Forensic evaluations may be required in complex adoption cases, particularly when there is a question of whether an adoption is or was fraudulent or when a birth parent who had apparently given up a child is demanding the return of the child. These cases often cross state lines and international boundaries. Debate continues in the United States about the virtues or drawbacks of transracial adoption. The child and adolescent psychiatrist may be called as an expert witness to testify about the best interests of a child who is the subject of such proceedings.

Termination of Parental Rights Common law presumes that parents are competent to provide appropriate care and nurturing to their children. The state does not interfere unless there are egregious deviations from acceptable parenting behavior. When that is thought to occur, usually after child protective services have been involved with a family, a state may petition to terminate parental rights. Such termination ends the parental-child relationship in every way. It also may free the child from the uncertainties of foster care and allow for permanent placement and adoption. In the United States Constitutional law, as interpreted by the United States Supreme Court in 1982, requires the burden of proof in such cases to be clear and convincing evidence. Some states go further and use the standard of beyond a reasonable doubt.

When such cases arise, agencies commonly consult with child psychiatrists, who are asked to evaluate the parents and sometimes the children to give an opinion on the validity of a petition for termination. The psychiatrist reviews all of the historical data and assesses the parents to see whether they have a serious psychiatric disorder, whether they demonstrate any hope of rehabilitation, and whether their children would be in danger if returned to them. The forensic evaluation leads to a report and, frequently, courtroom testimony.

Two parents were locked in a bitter custody dispute over their 4-year-old son. Throughout the litigation, the father made numerous allegations that his wife was physically abusing the boy, burning him with cigarettes, sticking him with pins, poisoning his milk, and terrifying him by holding his head underwater while giving him a bath. The father had called Child Protective Services numerous times, had taken his son to every emergency room in the city and had sent out milk samples for toxicological analysis. All physical examinations and laboratory tests were normal, and no doctor or city social worker had ever confirmed signs or symptoms of abuse. Yet, the father persisted in his accusations.

The court-appointed child psychiatrist recorded the father's complaints but also failed to confirm abuse. In the joint interview with father and child, the boy was playing at the dollhouse. His father picked up a miniature bathtub and told him, “Billy, show the doctor how Mommy pushes your head underwater and tries to scare you. Go on, Billy, show the doctor.” The child refused and became quite anxious. He moved away from the dollhouse.

The child psychiatrist concluded that the father suffered from delusions about abuse and that these delusions were harmful to the child, because he acted upon them and caused the child to undergo many unnecessary examinations and tests. He recommended that for these and other reasons, custody be awarded to the mother. He also recommended supervised visitation for the father. The judge granted custody to the mother but initially did not order that the father have supervised visitation. When he was with the child, the father continued to take him to emergency rooms. After another year, the judge stopped the visits altogether.

Child custody cases show no sign of abating. Attempts at mediation and arbitration have sometimes been successful in localities where they may be mandated or available on a private basis. However, parents will continue to seek relief within the adversarial system. There will most likely be more cases involving relocation issues. For example, a custodial parent may wish to move to another state because of employment not available locally. The noncustodial parent, still very much involved with the children, sues to prevent the exspouse from moving with the children. Most likely the best-interests presumption will be applied here as well.

There will be more cases involving allegations of sexual abuse and consequently evaluations-within-evaluations, in which the psychiatrist will be asked to assess the validity of the allegations as part of the custody evaluation. To do so requires specialized skills, discussed elsewhere in this book.

Finally, there will be more litigation born of the astonishing advances in reproductive technology. Children may be conceived today by such techniques as in vitro fertilization, artificial insemination of sperm, oocyte donation, and surrogacy. Custody disputes have already arisen over frozen embryos. With tens of thousands of potential viable frozen embryos being maintained currently in the United States, such cases are inevitable. The role of the forensic child psychiatrist in these complex cases will be hotly debated.

CIVIL LAW

Civil litigation, in which there are issues of psychological damages arising from negligence, frequently requires a child or adolescent psychiatrist to serve as an expert witness. Because this is an adversarial process, forensic evaluations may be performed for the plaintiff (the individual bringing the law suit) or the defendant (the person accused of negligence). Child psychiatrists may be asked by an attorney to evaluate a child represented by that attorney or a child who has been examined by a child psychiatrist for the adversary. In the latter case, the forensic child psychiatrist will be asked by the attorney to refute, if possible, the diagnostic conclusions made by the expert for the other side.

Posttraumatic Stress Disorder A common diagnosis, often the centerpiece of a civil trial, is posttraumatic stress disorder. In a typical lawsuit, a plaintiff child (or the parents) allege psychological damages, specifically posttraumatic stress disorder, resulting from the negligence of the defendant. In a civil case such an allegation of personal injury falls under tort law. Since 1970 case law has become more sympathetic to claims of psychic injury. Such injury is recognized as occurring concomitantly with physical injury, when a person is nearby and in danger of physical injury (the zone of danger), or when physical injury to another is witnessed.

A forensic child psychiatrist examining a plaintiff child for that child's attorney must first make the diagnosis of posttraumatic stress disorder, satisfying the criteria of the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Beyond that, however, the psychiatrist will be asked to give an opinion about whether the psychiatric disorder resulted from specific actionable behavior by the defendant. A child psychiatrist hired by the defense, on the other hand, would be asked to refute the diagnosis of posttraumatic stress disorder, or conclude that (whatever it is) the disorder did not result from the defendant's alleged negligence.

Such evaluations by child and adolescent psychiatrists are fraught with problems, especially if clinicians allow themselves to be used by attorneys to benefit their client. As in all areas of clinical practice, psychiatrists are first and foremost physicians, advocating what they believe to be clinically accurate and truthful, not what neatly fits into a lawyer's strategy.

Other Issues Other civil cases in which forensic child psychiatrists might be involved include the psychiatric hospitalization of children and assessing psychic damages from physical or sexual abuse. In evaluating such cases for either side, the child psychiatrist must carefully assess the degree of impairment in the child, offer hypotheses about causation of the psychic injury, discuss the prognosis with special attention to further growth and development, and consider any necessary treatments and their costs.

A defendant's attorney asked a child psychiatrist to examine a 7-year-old boy. Two years before, his client had driven his car into another car, which then jumped a curb, pinning the child beneath the wheels for nearly a half hour. Miraculously, the child suffered no broken bones or neurological injuries. However, for months after the accident, he had nightmares, developed phobias about being near automobiles or on sidewalks, cried frequently, complained of multiple somatic problems, and refused to play outside with his friends after school.

His parents sued the driver of the first car and claimed the accident caused their son to suffer posttraumatic stress disorder. The defense attorney secured the child's records from kindergarten, first grade, and part of second grade. The child was doing well in school. The defense held that the child did not suffer from posttraumatic stress disorder because of his excellent school performance.

The defendant's child psychiatrist expert reviewed the school records, examined the child, and met with his parents. He concluded the child did have posttraumatic stress disorder but continued to work very hard in school. He found the parents' report of the child's behavior to be credible, and he concluded that the symptoms were a direct result of the accident. The defendant's lawyer settled the case out of court.

Cases involving allegations of posttraumatic stress disorder in children will continue. The forensic child psychiatrist will have to follow the literature closely and keep up with developments in the field regarding the diagnosis and etiology of this disorder. Forensic child psychiatrists may also be asked to participate in cases involving psychic damage from alleged sexual harassment. Several cases have already emerged from school districts across the country, accused by parents of not intervening adequately when one student allegedly sexually harasses another. Child psychiatrists will be asked both to testify about normative behaviors at various developmental stages and to assess supposed damages from behaviors alleged to be harmful–as well as harassing.

Finally, cases concerning psychological trauma from physical and sexual abuse will continue. Child psychiatrists will evaluate psychic damages from abuse contemporaneous with the injury. Experts will also continue to be involved in cases arising later in childhood or in adulthood in which plaintiffs allege psychic injury from sexual abuse alleged to have been committed against them when they were children. Such cases may include so-called repressed-memory cases. In all such circumstances, the forensic expert will need to stay abreast of both current scientific developments and state and federal laws.

CRIMINAL LAW

Child and adolescent psychiatrists commonly perform forensic evaluations in criminal cases. These can involve assisting the court as a neutral evaluator or working with either prosecutors or defense attorneys. In addition to trial-related evaluations, child psychiatrists can be involved in the preadjudication phase by assessing immediate psychiatric needs of a juvenile detainee, participating in a waiver hearing (to determine whether the child will be tried as an adult), or evaluating competence to stand trial. In the postadjudication phase, the psychiatrist can assist the court in sentencing recommendations, long-term treatment alternatives, or the determination of probation.

Knowledge of the juvenile court system and its strengths and weaknesses is essential for the forensic child psychiatrist working in this area. Likewise, the clinician must appreciate the indications for waiver hearings and the grounds for judicial determination that a minor be tried as an adult. Although the juvenile court system operated for decades as supposedly child oriented and protective, the United States Supreme Court, in the landmark case *In re Gault,* determined that the system sometimes did not accord juveniles rights equal to those of adults. This decision held that in delinquency cases, juveniles must be accorded basic due process rights: the right to notice of charges, the right to legal counsel, the Fifth Amendment privilege against self-incrimination, and the right to confront witnesses. Another important Supreme Court case, *In re Winship,* held that the standard “beyond a easonable doubt” must be followed in delinquency hearings.

The forensic child psychiatrist evaluating a child defendant in a criminal case performs a comprehensive examination regardless of the side making the referral. All previous records must be obtained and both family members and the child who is charged must be interviewed. The report is tailored to the needs of the attorney or the court or both. For example, the clinician might be asked to present diagnoses and their prognoses or to assess dangerousness or recidivism. The child psychiatrist may also be asked to evaluate the juvenile in the context of a possible insanity defense. Clearly, the forensic evaluator of the delinquent child must not only appreciate local statutes and legal principles, but must also be grounded in developmental issues, factors correlated with antisocial behavior, clinical diagnosis, and appropriate treatment alternatives.

A 13-year-old boy was charged with the brutal murder of a 4-year-old child. In his signed confession, the teenager said that for no apparent reason, except that he felt angry, he choked the child, fractured his skull with a large rock, and then sodomized him. The public defender asked a child psychiatrist to evaluate the accused in the hopes of shedding some light on this seemingly senseless act.

The child psychiatrist interviewed the adolescent and all members of his immediate family. He reviewed school and medical records and ordered psychological, endocrinological, and neurological evaluations. The child had stigmata of a congenital dysmorphic syndrome caused by an anticonvulsant drug taken by the mother while pregnant. The boy had breath-holding spells as an infant and toddler and intermittent rage attacks throughout his life. He had been known to torture animals. He had few friends.

Under state law, the child was tried as an adult. The forensic child psychiatrist testified that the teenager suffered from intermittent explosive disorder and should be held in a psychiatric institution for many years. He recommended intensive group and individual psychotherapy along with medication to control the rage. The jury, however, found the young man guilty of second-degree homicide, and the judge sentenced him to the maximum allowed by state law, 9 years to life. The child was remanded to a juvenile detention facility until his eighteenth birthday, at which time he will be transferred to an adult prison. Inside the detention facility, he receives no specific psychiatric treatment.

With United States government predictions of more juvenile arrests for serious crimes in the next decade, forensic child psychiatrists will be busy. They will most likely be participating in more waiver hearings, assessing more violent children, making more predictions of future violence, and working in and evaluating juvenile detention centers. In their reports and trial testimony they will be called upon to offer hypotheses about the causes of violent antisocial behavior in children and adolescents, and they will need to stay current with new research.

OTHER FORENSIC ISSUES

Child and adolescent psychiatrists may be involved in forensic evaluations outside the more common venues of the family or criminal court. For example, they may be called upon to make certain recommendations for a student with special needs under the landmark Education for All Handicapped Children Act of 1975. This law requires that all handicapped children, regardless of the severity of their condition, be provided a free and appropriate public education in the least restrictive environment. Handicapped children are defined as those who are mentally retarded, learning disabled, physically disabled, or emotionally disturbed. A treating or evaluating child psychiatrist may be called upon to testify at various hearings required under the law. Knowledge of the law's implications as well as the particular child will be critical to performing a proper evaluation.

Occasionally, a child and adolescent psychiatrist may be asked to become involved in a malpractice suit against another child psychiatrist. This involvement may be for the plaintiff or the defendant's side. Malpractice cases in child psychiatry may arise for a number of cited reasons including failure to give informed consent to parents, medication errors, inappropriate behavior toward a child, failure to warn a third party of endangerment, abandonment, diagnostic error, or disclosure of private information. The forensic child psychiatrist evaluates the data carefully before offering an expert opinion. Malpractice cases are complicated and arouse strong emotions. Forensic psychiatrists must monitor their own potential for countertransference under these circumstances.

From time to time, a child psychiatrist may be asked to testify in general or regarding a specific case about children as witnesses. Such cases may arise when there are allegations of sexual abuse or when a child either witnessed or is the victim of a crime. The forensic expert's role is to educate the court about what is known about children's memory and recall. This area is currently under intensive investigation.

FUTURE DIRECTIONS

The future will be busy and challenging for forensic child and adolescent psychiatry. Courts will continue to grapple with calendars clogged with child placement cases. Juvenile crime rates are expected to remain high, and judges will continue to turn to forensic experts for guidance in these complex cases. As the field

continues to grow in terms of practitioners and an increasing body of scholarly literature, some changes and improvements are in order.

Forensic child psychiatry needs more standards for various kinds of evaluations. The Practice Parameters on Evaluating Allegations of Sexual Abuse and Child Custody Evaluation published by the American Academy of Child and Adolescent Psychiatry are steps in the right direction. The field needs better clinical consensus on evaluating criminal behavior, participating in civil litigation, and the uses and misuses of the child psychiatrist expert witness.

Although more forensic psychiatry fellowships are becoming available, few offer extensive experience in forensic child and adolescent cases. More training opportunities are needed in this field. As an additional way to teach younger clinicians, more experienced child psychiatrists should offer to serve as mentors, guiding and acting as role models for those new to the field.

In addition, the time may now be right for establishment of an international society for forensic child and adolescent mental health professionals. Forensic clinicians who are psychologists, social workers, and psychiatrists have much to learn from one another.

Finally, the revolution in communications and information can benefit forensic practitioners, as it has so many other professionals. The Internet is a repository of numerous World Wide Web sites devoted to forensics as well as legal sites of interest to psychiatrists. Both major recent court decisions and historic, landmark rulings are available on the Internet. Law schools have their own websites, as do associations such as the American Bar Association. There is even a website providing the audio from oral arguments before the United States Supreme Court. Thus, information pertinent to forensic practice will become more accessible and it will be easier to communicate with colleagues all over the world.

SUGGESTED CROSS-REFERENCES

Forensic and ethical issues are discussed in Chapter 54. Posttraumatic stress disorder and other anxiety disorders are covered in Chapter 15. Chapter 22 covers intermittent explosive disorder and other disorders of impulse control. Child abuse is discussed in Section 49.4. Section 49.2 discusses adoption.

SECTION REFERENCES

Ackerman MJ: *Clinician's Guide to Child Custody Evaluations*. Wiley, New York, 1995.

*American Academy of Child and Adolescent Psychiatry: Practice parameters for child custody evaluation. J Am Acad Child Adolesc Psychiatry *36* (Suppl):57S, 1997.

Armitage DT: Parental competence and termination of parental rights. In *Principles and Practice of Forensic Psychiatry*, R Rosner, editor. Chapman & Hall, New York, 1994.

Ash P, Derdeyn AP: Forensic child and adolescent psychiatry: A review of the last 10 years. J Am Acad Child Adolesc Psychiatry *36:*1493, 1997.

Borum R, Grisso T: Establishing standards for criminal forensic reports: An empirical analysis. Bull Am Acad Psychiatry Law *24:*297, 1996.

Cancian M, Meyer DR: Who gets custody? Demography *35:*147, 1998.

*Ceci S, Bruck S: *Jeopardy in the Courtroom—Scientific Analysis of Children's Testimony*. American Psychological Association, Washington, DC, 1995.

Crespi TD, Rigazio-DiGilio SA: Adolescent homicide and family pathology: Implications for research and treatment with adolescents. Adolescence *31:*353, 1996.

Derdeyn A: Child custody conflicts in historical perspective. Am J Psychiatry *133:*1369, 1976.

Derdeyn AP: Adoption. In *Child Psychiatry and the Law,* DH Schetky, EP Benedek, editors. Brunner/Mazel, New York, 1980.

Feldman-Schorrig S: Need for expansion of forensic psychiatrists' role in sexual harassment cases. Bull Am Acad Psychiatry Law *23:*513, 1995.

Griffith EH: Forensic and policy implications of the transracial adoption debate. Bull Am Acad Psychiatry Law *23:*501, 1995.

Guyer MJ: Commentary: The juvenile justice system. In *Emerging Issues in Child Psychiatry and the Law,* DH Schetky, EP Benedek, editors. Brunner/Mazel, New York, 1985.

Heide KM: Juvenile homicide in America: How can we stop the killing? Behav Sci Law *15:*203, 1997.

Herman SP: Forensic child psychiatry. J Am Acad Child Adolesc Psychiatry *29:*955, 1990.

Herman SP: Special issues in child custody evaluations. J Am Acad Child Adolesc Psychiatry *29:*969, 1990.

*Herman SP: Child custody evaluations. In *Clinical Handbook of Child Psychiatry and the Law*, DH Schetky, EP Benedek, editors. Williams & Wilkins, Baltimore, 1992.

In re Gault 387 U.S. 1, 1967.

In re Winship 397 U.S. 358, 1970.

*Kalogerakis MG: Juvenile delinquency. In *Clinical Handbook of Child Psychiatry and the Law*, DH Schetky, EP Benedek, editors. Williams & Wilkins, Baltimore, 1992.

Kruh IP, Brodsky SL: Clinical evaluations for transfer of juveniles to criminal court: Current practices and future research [published erratum appears in Behav Sci Law 16(1):155, 1998]. Behav Sci Law *15:*151, 1997.

Loeber R, Stouthamer-Loeber M: The development of offending. Crim Justice Behav *23:*12, 1996.

Loeber R, Stouthamer-Loeber M: Development of juvenile aggression and violence. Some common misconceptions and controversies. Am Psychol *53:*242, 1998.

Loftus E, Ketcham K: *The Myth of Repressed Memory*. St. Martin's Press, New York, 1994.

Mason MA: *From Father's Property to Children's Rights—The History of Child Custody in the United States*. Columbia University Press, New York, 1994.

*Nurcombe B, Partlett DF: *Child Mental Health and the Law*. Free Press, New York, 1994.

Pfefferbaum B: Posttraumatic stress disorder in children: A review of the past 10 years. J Am Acad Child Adolesc Psychiatry *36:*1503, 1997.

Puri BK, Lambert MT, Cordess CC: Characteristics of young offenders detained under section 53(2) at a young offender's institution. Med Sci Law *36:*69, 1996.

Quinn KM: Legal issues and the schools. In *Emerging Issues in Child Psychiatry and the Law,* DH Schetky, EP Benedek, editors. Brunner/Mazel, New York, 1985.

Quinn KM: Ethical dilemmas in forensic child and adolescent psychiatry. Child Adolesc Clin North Am *4:*809, 1995.

Rampe D: Your own law library (no shelves required). New York Times, May 12, 1997:D6.

Rosenberg JE, Eth S: Posttraumatic stress disorder in children: Clinical and legal issues. In *Principles and Practice of Forensic Psychiatry,* R Rosner, editor. Chapman & Hall, New York, 1994.

Schetky DH, Benedek EP: Historical development of forensic child psychiatry. In *Child Psychiatry and the Law,* DH Schetky, EP Benedek, editors. Brunner/Mazel, New York, 1980.

Schetky DH, Guyer M: Civil litigation and the child psychiatrist. J Am Acad Child Adolesc Psychiatry *29:*963, 1990.

Schetky DH: Psychic trauma and civil litigation. In *Clinical Handbook of Child Psychiatry and the Law,* DH Schetky, EP Benedek, editors. Williams & Wilkins, Baltimore, 1992.

*Schwab-Stone M, Chen C, Greenberger E, Silver D, Lichtman J, Voyce C: No safe haven II: The effects of violence exposure on urban youth. J Am Acad Child Adolesc Psychiatry *38:*359, 1999.

Sikorski JB: Forensic psychiatry. In *Textbook of Child and Adolescent Psychiatry,* ed 2, JM Wiener, editor. American Psychiatric Press, Washington, DC, 1997.

Steiner H, Garcia IG, Matthews Z: Posttraumatic stress disorder in incarcerated juvenile delinquents. J Am Acad Child Adolesc Psychiatry *36:*357, 1997.

Straus MB: *Violence in the Lives of Adolescents*. Norton, New York, 1994.

Viemero V: Factors in childhood that predict later criminal behavior. Aggressive Behav *22:*87, 1996.

Wynne EE: Children's rights and the biological bias in biological parent versus third-party custody disputes. Child Psychiatry Hum Dev *27:*179, 1997.

49.13 ETHICAL ISSUES IN CHILD AND ADOLESCENT PSYCHIATRY

DIANE H. SCHETKY, M.D.

Ethical codes are characteristic of all professions and serve to regulate behavior, safeguard the public, and promote trust. The basic ethical theories and principles discussed in Section 49.12 were first codified by the American Medical Association (AMA) in 1847. The AMA Code of Ethics underwent many revisions and in 1973 was incorporated into the American Psychiatric Association's (APA) Code of Ethics with Added Annotations Applicable to Psychiatry. The American Academy of Child and Adolescent Psychiatry (AACAP) developed its Code of Ethics in 1980. This Code addresses the relationship between the child and parent and the potential for conflicting interests between child, parent, and society. It also emphasizes developmental considerations that enter into ethical decisions and special issues related to the confidentiality of child and adolescent patients. In 1995 the AACAP added Annotations to the AACAP Ethical Code with Special Reference to Evolving Health Care Delivery and Reimbursement Systems.

In the United States physicians are expected to adhere to the ethical code of the AMA. Child and adolescent therapists who are not physicians are expected to follow the ethical codes of the professional organizations to which they belong. Neither the APA nor AACAP Codes of Ethics address forensic issues. The child and adolescent psychiatrist who practices in this area should be acquainted with the Ethical Guidelines for the Practice of Forensic Psychiatry published by the American Academy of Psychiatry and Law.

SPECIAL ISSUES

Agentry Special ethical issues arise in treating child and adolescent patients because of their developmental immaturity and minority status. Even though a child is the identified patient, he rarely presents on his own and is more likely to be seen because of behavior that is troubling to others. Referral sources, be they parents, guardians, physicians, schools, or the courts, often have their own agenda in making a referral, an agenda that may or may not coincide with the child's agenda. The therapist needs to refrain from colluding with agendas and focus on what lies in the patient's best interest.

> A typical scenario is the Green family, who come to Dr. Riley telling him that he came highly recommended by their pediatrician and pastor. They complain that their 15-year-old son, Josh, talks back, won't communicate with them or inform them of his whereabouts, is obsessed with his rock band, dresses like a street person, and refuses to attend church with them. The final straw that led them to seek professional help is the ring that Josh now wears in his eyebrow. Mr. and Mrs. Green are hoping that Dr. Riley will bring back the son they used to know. Josh doesn't see what the big deal is and tells Dr. Riley that everything would be fine if only his parents would get off his back, trust him, and allow him to make some of his own decisions. He is doing fine in school and is enrolled in honors courses. He reassures Dr. Riley that, contrary to his parents' fears, his interest in rock music will not cause him to take drugs and drop out of school, which is what happened to his mother's younger brother.
>
> Dr. Riley decides to ally himself with neither the parents nor Josh's agendas, choosing rather to focus on issues of separation-individuation. Yet another issue for him to consider is whether Josh is even in need of treatment and whether he can justify seeing him under his managed-care plan for what appears to be a parent-child problem. He realizes the decision for treatment also lies in Josh's hands and that Josh's parents forcing Josh to return against his will is likely to be counterproductive. He might choose to work just with Mr. and Mrs. Green, but must weigh this against whether it would preclude Josh from seeing him if he chooses to return at a later time; alternatively, he might consider a brief course of family therapy.

Collaterals Given that children are less able than adults to provide past history and family history, the child therapist must often rely on collateral sources of information to piece together a history and formulation. In addition to meeting with parents, this often involves seeking information from collateral sources such as schools, courts, and pediatricians. These contacts may also be important in implementing treatment plans. The therapist needs to be sensitive to the concerns of both patient and family in initiating these contacts and must seek proper consent for release of information. Some parents may willingly consent whereas others fear stigmatization or do not trust that recipients of the shared information will maintain confidentiality.

How aggressive the child therapist should be about pursuing contact with a school might depend on the extent to which a child's problem is affecting his ability to function in school.

> For instance, Ryan, age 10, is falling behind in school because his morning rituals prevent him from getting to school on time. Once there, he constantly checks and rechecks his paperwork and tests for errors. He has difficulty completing assignments, and excessively questions and seeks reassurance. Ryan's school needs to know that obsessive-compulsive disorder underlies his indecisiveness. School, family, and therapist need to work together towards developing a plan to help Ryan deal more constructively with his anxiety. Most schools also want to be informed if a child is taking medication and about possible adverse effects. The clinician needs to be cautious about how much family history to share with the school and should generally limit it to that which is directly relevant to the student's problem.

Countertransference Working with children and adolescents brings out special countertransference issues. It is easy for the inexperienced therapist to overidentify with the child and to blame the parents for a child's problem. Therapists may also overidentify with adolescent patients and unwittingly encourage their acting out or anger with parents if they have not worked through their own adolescent issues. Fantasies of rescuing a child from a bad home or from parents who do not appreciate her may also occur and impede work with the family and child. Fortunately, blaming the parent has become less common as we have come to appreciate that many childhood disorders are biologically based. Another antidote for this sort of countertransference is parenthood, which inevitably instills instant humility and empathy in the therapist. Some childless therapists may come to view a child patient as a longed-for child and give her special treatment. Child therapists may also experience erotic countertransference either towards patients or their parents and have great difficulty processing these feelings because of their forbidden nature. Therapists must deal with countertransference issues so that they not lead to acting-out behavior, which is potentially harmful to the child and can impede treatment.

CONSENT

Obtaining Informed Consent Obtaining consent with minors is complex because generally they are not considered competent to make legal decisions until age 18. Exceptions exist in most states for emancipated minors, those seeking birth control, treatment of sexually transmitted diseases or substance abuse, and in emergency situations. Some states allow mature minors to contract for treatment. Therapists who choose to treat an adolescent without parental permission should be aware of the laws in their state and the risks they might incur. By about age 11 or 12, children reach the stage of formal operational thinking and are capable of comtemplating various courses of action, understanding contingencies and consequences, and considering hypotheticals. Studies have demonstrated that adolescents are capable of making complex decisions; indeed, courts afford considerable weight to the wishes of adolescents in custody and visitation decisions.

In most instances it is the parent or guardian who will make treatment decisions for the child; nonetheless, children and adolescents should be asked for their consent to the proposed treatment. Consent is not legally binding but conveys to the child that his input is valuable; it also allows the child to express veto power.

Conditions of informed consent require that one: (1) be given information about the nature and extent of the psychiatric problem, (2) be made aware of risks and benefits of the proposed treatment, (3) be told what treatment alternatives are available, and (4) be capable of voluntary consent. How much information is disclosed is in most states based on the materiality of information standard, which holds that the information disclosed should be based on what a reasonable person would

want to know in order to make a decision under similar circumstances. Parents and adolescents also need to know if drugs are proposed that do not have approval from the Food and Drug Administration (FDA) for use in children and the rationale for using them. Very few psychiatric drugs have been tested in children, hence their lack of FDA approval. However, the FDA does not limit the physician's prescribing habits and there is a body of research emerging on new uses of drugs for conditions and age groups for which they were not intended (e.g., quanfacine [Tenex] for attention-deficit/hyperactivity disorder and selective serotonin reuptake inhibitors [SSRIs]) prescribed for children. Discussions about treatment and potential risks or side effects should be documented in charts. Written consent is also advisable when using antipsychotic agents or medications that are likely to carry serious risks.

Ethical Conflicts Pertaining to Consent Ethical conflicts arise when parents disagree on proposed treatment.

> Mrs. Schulz, who is recently divorced, seeks therapy for her 6-year-old son, Aaron. Mr. Schulz, who has shared custody and is liable for the costs of treatment, insists that Aaron is fine when he is with him and sees no need for treatment. He suggests that if there is a problem it lies with his ex-wife. Dr. Jain, who has evaluated Aaron and believes he needs treatment for depression, now finds herself in a bind. Legally, she only needs the consent of one custodial parent to see a child but she recognizes that without the consent of the father he is likely to undermine therapy. She must decide what is in Aaron's best interest in the long run. She later wishes she had chosen to involve the father in the initial evaluation or perhaps not have conducted the evaluation without his permission.

Conflicts may arise between patient, parent, and therapist over medication compliance.

> Twelve-year-old Jennie, who has bipolar I disorder, has done well on lithium over the past 5 years. She suddenly announces to Dr. Daniello that she wants to stop taking her lithium. On exploring the reasons for this, he learns that Jennie's classmates are calling her a “druggie” for having to take lithium and that she still experiences occasional lithium-induced enuresis and fears embarrassment if she goes to a friend's house overnight. Jennie's anxious mother proposes bribing her. The more she pushes, the more Jennie balks. Dr. Daniello ponders how much autonomy he should give this preadolescent regarding the decision to discontinue medication. He appeals to her wish for autonomy and her intellect and reminds her of studies they have previously discussed showing decreased effectiveness of lithium when it is stopped and then reintroduced. He suggests a compromise in which she skips her evening lithium if she is going on an overnight visit. He reminds her that taking medication is her decision, that neither he nor her mother will force it upon her, and suggests she think about what she wants to do. Jennie returns for her next appointment feeling more empowered and says she's decided to stay on her lithium.

Release of Information to Third Parties Older children and adolescent patients as well as their parents should be consulted in decisions about release of information to others. They need to know what is going to be released, to whom, and for what purposes. Reports should be written such that they can be read by parents or adolescents, which requires sensitivity and tact. Once a report is released, the author of it often loses control over who sees it. Particularly in forensic matters, reports are likely to be shared with parents. In custody evaluations children need to be informed at the onset about the lack of confidentiality.

Child and adolescent psychiatrists often work in collaboration with other therapists. This may involve medication management, supervision, or consultation. The success of these relationships often hinges on good communication. It is recommended that parents and collaborative therapists sign a consent form that spells out respective responsibilities and lines of communication.

Consent for Research As the field of child and adolescent psychiatry becomes more research based, children and adolescents are increasingly being sought out as research subjects. Factors to be considered include age of the child, risk the research poses to the child, invasiveness of the research, and the prospect that the research will benefit the child. Levine notes that, “Past policies and practices designed to protect vulnerable persons from unwanted or unwarranted burdens actually deprived them of important benefits.” Some authors propose a system of age-graded consent for moderate-risk protocols.

CONFIDENTIALITY

Defining Confidentiality Confidentiality refers to the clinician's obligation to keep material shared in the course of a professional relationship from third parties, unless confidentiality is waived by the patient or guardian. The term *confidentiality* is often confused with the term *privilege,* which refers to the patient's right to bar the clinician from testifying about professional material in a judicial or quasi-judicial proceeding. Confidentiality is the clinician's obligation whereas privilege is the patient's right. Confidentiality is the cornerstone of therapy, without which patients would be reluctant to share information necessary to diagnosis and treatment.

Guidelines governing the minor's right to confidentiality are less clear than those for adult patients. Parents may feel that they are entitled to information about their child, especially on such topics as sexuality and substance abuse. The therapist, on the other hand, needs to respect the adolescent patient's need for confidentiality and to preserve the therapeutic alliance. Compounding this dilemma is the fact that parents may waive the therapist-patient privilege on behalf of the child, even though the waiver is not truly informed because they do not know what is in the child's record. This may occur in custody battles where one parent thinks information from the child's therapist will aid their side. The therapist may feel that release of sensitive information is detrimental to the child or to the parent-child relationship. In such a situation, the therapist should appeal to the attorney requesting the information or to the judge. A compromise proposed by the APA is that the petitioner must first demonstrate why the information is relevant to a custody determination and why it cannot be obtained from other sources. If this threshold is crossed, the judge or another professional could then review the material and decide on its relevance or appropriateness.

Limits of Confidentiality Most child and adolescent therapists at the onset of therapy spell out the limits of confidentiality to the families they are working with and clarify how communications with parents are to be handled; they also need to discuss the limits of confidentiality in regard to insurance plans. Additional limitations to confidentiality include behaviors harmful to self or others. When disclosures need to be made to parents, the patient may be given the option of telling them himself, or of discussing it with parents and therapist together. Other exceptions to confidentiality include laws on reporting child abuse, which are mandatory in all states. A physician who fails to report suspicion of abuse may be liable for civil as well as criminal sanctions. If an abuse report needs to be filed, parents should be so informed. In some states physicians may also have a duty to protect third parties under *Tarasoff I* and *Tarasoff II* (i.e., the 1976 and 1982 rulings of the California Supreme Court in the two cases of *Tarasoff v. Regents of University of California*).

BOUNDARIES

Purpose of Boundaries The APA's *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry,* Section 1, stresses the need for the patient to be able to trust a psychiatrist knowing that professional ethics preclude the exploitation of patients for personal gain. The need for continued vigilance about boundaries in the doctor-patient relationship is also emphasized. Good boundaries promote good therapy and greatly diminish the likelihood of a malpractice suit. Boundaries maintain professional roles, define limits, provide security to the patient and therapist, and lessen the risk of exploiting patients for personal gain. Numerous forces threaten to derail the therapist's boundaries, including countertransference, the primal productions of children, dual relationships, undue pressure from others, greed, and psychiatric impairment in the therapist.

Maintaining Boundaries Maintaining boundaries is an ongoing process that begins prior to accepting a patient and continues through therapy and after termination. It is important to screen new patients for potential conflicts of interest. Treating the children of friends, colleagues, or employees is rife with difficulties. The therapist is rarely neutral at the onset and usually has access to information from outside therapy that she cannot use in therapy, but that may color her perceptions. Also, the therapist may become privy to information about her friends and colleagues that she would rather not know and will need to compartmentalize and not use it outside the office. For instance, she may learn through her young patients that her dentist beats his wife, that Dr. Jones down the hall cross-dresses, or that the neighbor with whom she car pools is a closet alcoholic. It is usually preferable, although often difficult, to turn down these tainted requests for professional help that come from friends and colleagues and to refer them elsewhere. In communities with few resources, the therapist may have no alternative but to see the family. In cases where the pathology is minimal (which is impossible to know at the onset) and where boundaries are maintained vigilantly, the therapist may be able to pull off a successful intervention. However, the therapist's relationship with the family in question will never be the same again.

Dual Relationships Child and adolescent therapists often find themselves in dual relationships with their young patients such as when their own children participate in soccer, scouts, and youth groups with their patients. It may be awkward if a therapist's daughter wants to have a play date with one of her parent's patients. The older one's children become, the less control one has over the friends with whom they choose to associate. Difficulties also arise when spouses have no knowledge of the therapist's case load and may inadvertently have contacts with patients or their families. Confidentiality needs to be preserved unless the situation becomes untenable.

In small towns, out-of-office contacts with patients and their families are daily occurrences that the therapist has to learn to deal with. Successful management of these encounters requires that the therapist respect patient boundaries, process these encounters in therapy, and not do anything in public that she would not want her patients to see. Some of these encounters may even be of therapeutic value and encourage the young patients to see the therapist as a real person who has a life outside of the office.

Forensic Issues Often therapists are pressured to testify in court on behalf of their patients. This may occur at the request of patients or parents who believe that such testimony will help their case or in response to subpoenas from either party. It is common in child custody cases where the therapist has been treating the child of divorcing parents. Therapists may agree to testify expecting their testimony to be helpful, but it may just as likely end up being harmful. Testifying on behalf of a patient is rife with pitfalls. Therapists lack the objectivity of a forensic evaluator and often have a more limited data base. Patients or parents of patients may waive privilege unaware of what is in their records or how it might be used to discredit them. Once on the witness stand, therapists lose control over the patient's record and may have to disclose things that were said in confidence. They also risk having to say things that may jeopardize their therapeutic alliance with the patient and the patient's parents or that may harm the child-parent relationship.

The importance of not wearing two hats cannot be stressed enough and therapists should avoid rendering forensic opinions on patients they are treating or have treated in the past. There are several alternatives, including recommending an outside forensic evaluation when legal issues arise. The therapist may share information with the forensic evaluator but does not have to render recommendations on sensitive matters such as which parent a child should live with or commitment of a patient to a youth facility. This arrangement also makes it less likely that forensic issues will derail therapy. If faced with a subpoena from the patient's attorney, the therapist may appeal to the attorney issuing the subpoena and point out how testimony might be harmful to the patient and to therapy. The psychiatrist may also appeal to the court to quash a subpoena. If forced to testify, therapists should do so in as narrow a fashion as possible, limiting themselves to factual data such a diagnosis, dates of treatment, and treatment needs rather than giving opinion testimony. A useful response to being questioned in court is: “Having observed the patient only from the vantage point of a treating clinician, I have no objective basis for rendering an expert opinion, with a reasonable degree of medical certainty on a legal as opposed to a clinical question.”

If the legal issue concerns tort litigation, the litigant automatically tenders medical records in bringing about a claim for damages. In such cases, the therapist may be compelled to testify more broadly on the litigant's mental functioning prior to or subsequent to the action for which relief is being sought.

Boundary Violations Common boundary violations include financial conflicts of interest, exploitation of fame and notoriety, exploitation through writing, double agentry, socializing with patients, and sexual exploitation. These are not unique to child and adolescent psychiatry. It is prudent to adhere to the maxim “once a patient, always a patient” and consider each new patient who enters the office as one less potential friend. Transference rarely ends with treatment and by not socializing with former patients and their families the therapist leaves the door open should they choose to return to treatment at a later date.

Minor boundary violations, such as giving special treatment to a particular patient, are common and can be easily corrected if recognized; more ominous is a repeated pattern of boundary violations. Typically, in these scenarios the therapist uses the patient or his family to meet her own needs or fails to differentiate her own needs from those of the patient. Where there is a pattern of repeated boundary violations, therapists are advised to seek consultation or therapy.

ETHICAL ISSUES WITH MANAGED-CARE ORGANIZATIONS

Conflicts of Interest With the advent of managed care, some physicians are confused about where their allegiances lie. For instance, a child psychiatrist advises a managed-care organization for whom he works that it should not insure a particular child because the child requires such extensive services. Ethics demand that the patient's interests are foremost, but new economic pressures threaten to corrupt many physicians. Examples of policies designed to shape physicians' behavior are financial penalties for giving too much treatment or making too many referrals and the threat of being delisted if one does not comply with managed-care policies or advocates too vigorously for a patient.

The phenomenon of gag orders amply demonstrates how physicians cannot serve two masters. Fortunately, most states have now enacted legislation that forbids withholding from patients information that they need in order to make informed treatment decisions. This situation exemplifies the inherent conflict between business ethos, which stresses profit for investors, and medical ethics, which stresses doing no harm and advocating for the patient's well-being. Child and adolescent psychiatrists who chose to contract with managed-care organizations may refuse to sign gag orders and insert in their contracts statements such as “First and foremost, I will follow the ethical code of The American Academy of Child and Adolescent Psychiatry.” Patients should be informed if their therapist is practicing under an incentive plan or has any conflicts of interest in regard to the care delivered to them.

Advocates of managed care argue that this system of health care delivery is ethical because it stresses the concept of justice for all in an era of limited health resources. This would be plausible if report cards on managed-care companies showed that they were providing better care to the poor, elderly, and chronically ill; however, such has not been the case. Rather, surveys demonstrate that managed care prefers to enroll the cream of the crop; patients with more serious or chronic problems are relegated to Medicaid and Medicare or remain uninsured. Currently 40 million Americans, including 10 million children, are uninsured.

Many child therapists become frustrated in their efforts to advocate for their young patients' health care needs under managed care. They feel forced into giving less than optimal service or face the ethical temptation of exaggerating symptoms in order to gain needed approval of services.

It has been noted that inpatient units face many dilemmas and the gatekeepers encounter a barrage of ethical conflicts with each referral. Typically, the medical director is faced with highly complex cases that pose a high degree of risk. Children have the right to be treated in the least restrictive environment, yet this may not be available. In some cases the patient's needs may exceed the resources of a particular inpatient unit. The insurance company may be unwilling to pay for a higher level of service or will only consent to admission to a facility that is at a great distance from the patient's home.

> For instance, one medical director of a managed-care company from a distant state, with no knowledge of Maine geography, denied an inpatient admission for a suicidal, syncopal teenager with an eating disorder and would only authorize a day treatment program 2 hours from home to which she would have to drive herself on icy roads. His rationale for denial of inpatient care was that she was not yet in medical danger.

The staff of inpatient units are pressured daily by case managers to accelerate a patient's care (e.g., to initiate medication before an evaluation is complete, or to discharge a patient as soon as she is no longer suicidal). Staff must also deal with anxious parents worried about the cost of treatment once insurance benefits are exceeded and about their ability to handle a child who is prematurely allowed to go home. Should a therapist even accept patients whose insurance will not pay for minimally acceptable treatment? What is the therapist's liability and responsibility once benefits are cut off? Amidst the scramble for patients and the atmosphere of diminishing reimbursement, it is easy to be swayed by the need to keep an inpatient unit full or to maximize one's outpatient practice. However, patient care must not be compromised, nor should therapists jeopardize their integrity by contracting with plans that offer substandard care.

Need for Child Advocacy Subscribers are also confused by new systems of health care delivery and are often unclear about whether a case manager is looking out for their interests or those of the managed-care company. Many people are ill informed about the constraints of managed care and their rights and need their therapist's help with advocacy for their children. The 1995 AACAP Annotations to the Ethical Code stresses avoiding actions detrimental to the child and the need for the child and adolescent psychiatrist to reduce the deleterious effects of actions by others on the child. This translates to aggressive child advocacy, appealing managed-care decisions that are detrimental o the child, and demanding publicly accessible justifications for decisions.

Providers who choose to work within a managed-care network can educate policymakers about the unique treatment needs of children and that working with parents and children takes time. Providers can pressure companies to respect professional ethics and to develop their own code of ethics regarding patient care. Good managed care should include choice, competence, compassion, and doctor-patient relationships that are free of conflict and that promote continuity of care. Child therapists also have an important role to play in working with their professional organizations and legislators to come up with better solutions for equitable health care. Medical ethics existed long before the advent of managed care and will continue to exist long after managed care as long as physicians refuse to capitulate to financial pressure.

REPORTING ETHICAL VIOLATIONS

Section 2 of the APA's Principles of Medical Ethics states “A physician shall deal honestly with patients and colleagues, and strive to expose those physicians deficient in character or competence, or engage in fraud or deception.”

It is never easy to report a colleague and yet it is a task that is sometimes necessary to protect patients and the profession and to rehabilitate colleagues when possible. Concerns about ethical violations by psychiatrists should be reported to the district branch of the APA. If the clinician in question is not a psychiatrist, complaints should be filed with the relevant professional organization. Ethical complaints must be filed in writing and, once received, they are carefully investigated and, if warranted, go to a hearing. Ethical violations may result in sanctions such as admonishment, reprimand, suspension, or expulsion. Decisions regarding licensure are made by state licensing boards who conduct their own independent investigations. In some instances, violations may result in complaints being made with both the APA district branch and the state licensing board. Members who have committed ethical violations may face criminal and civil complaints filed by the plaintiff as well. If there is question of substance abuse or mental disorder, referral to the Impaired Physicians Committee may be appropriate.

A frequently asked question is whether a therapist must report abuse by a physician or former therapist when it is disclosed in therapy. Any abuse disclosed by a child to a mandated reporter must be reported to protective services if the abuse is within the family, and to the district attorney or appropriate law enforcement agency if it occurred outside the family. In contrast, if the patient bringing the complaint is an adult, it is usually the patient's decision whether or not to file a complaint and with which agency or organization, although this may vary from state to state. If the patient chooses not to report the abuse the therapist can be placed in a very awkward position regarding the risk that the alleged offender poses to other patients. It also pits the therapist's ethical duty to report a physician's alleged misconduct against his duty to maintain the patient's confidentiality. As with many ethical dilemmas, the therapist is left weighing competing principles.

SUGGESTED CROSS-REFERENCES

The reader is referred to Section 49.12 for a discussion of forensic child and adolescent psychiatry and Section 54.2 on ethics in psychiatry. For a further discussion of the Tarasuff decisions see Section 54.1, Legal Issues in Psychiatry.

SECTION REFERENCES

*American Academy of Child and Adolescent Psychiatry: *Annotations to AACAP Ethical Code with Special Reference to Evolving Health Care Delivery and Reimbursement Systems*. American Academy of Child and Adolescent Psychiatry, Washington, DC, 1995.

American Academy of Psychiatry and Law: *Ethical Guidelines for the Practice of Forensic Psychiatry*. American Academy of Psychiatry and Law, Washington, DC, 1991.

American Psychiatric Association: *Report of the Task Force on Disclosure of Psychiatric Treatment Records in Child Custody Disputes*. American Psychiatric Association, Washington, DC, 1991.

*American Psychiatric Association: *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry*. American Psychiatric Association, Washington, DC, 1995.

Arnold LE, Stoff D, Cook E, Cohen D, Kruesi M, Wright C, Hattab J, Graham P, Zametkin A, Castellanos X, McMahon W, Leckman J: Ethical issues in biological psychiatric research with children and adolescents. J Am Acad Child Adolesc Psychiatry *34:*929, 1995.

Benedek EP: Ethical issues in practice. In *Child Psychiatry and the Law,* DH Schetky, E Benedek, editors. Williams & Wilkins, Baltimore, 1991.

*Epstein RS: *Keeping Boundaries*. American Psychiatric Press, Washington, DC, 1994.

Koocher G, Keith-Spiegel P: *Children, Ethics, and the Law*. University of Nebraska Press, Lincoln, 1990.

Levine RJ: Children as research subjects: Ethical and legal considerations. In *Ethics: Child and Adolescent Psychiatric Clinics of North America,* DH Schetky, editor. Saunders, Philadelphia, 1995.

Natanson v. Kline, 186 Kan 393, 350 P.2d 1093 (1960).

*Rupert PA, Kozlowski NF, Hoffman LA, Daniels DD, Piette JM: Practical and ethical issues in teaching psychological testing. Prof Psychol Res Pr *30:*209, 1999.

Schetky DH, Devoe L: Countertransference issues in forensic child psychiatry. In *Child Psychiatry and the Law,* DH Schetky, E Benedek, editors. Williams & Wilkins, Baltimore, 1991.

Schetky DH: Ethical issues in forensic child psychiatry. J Am Acad Child Adolesc Psychiatry *31:*403, 1992.

*Schetky DH: Boundaries in child and adolescent psychiatry. In *Ethics: Child and Adolescent Psychiatric Clinics of North America,* DH Scheyky, editor. Saunders, Philadelphia, 1995.

Schetky DH: Ethics and the clinician in custody disputes. Child Adolesc Psychiatr Clin North Am *7* :455, 1998.

Strasburger L, Gutheil T, Brodsky A: On wearing two hats: Role conflict in serving as both psychotherapist and expert witness. Am J Psychiatry *154* :454, 1997.

Towbin KE, Campbell P: Ethical conflicts and their management in inpatient child and adolescent psychiatry. In *Ethics: Child and Adolescent Psychiatric Clinics of North America,* DH Schetky, editor. Saunders, Philadelphia, 1995.

Weithorn LA, Campbell S: The competency of children and adolescents to make informed treatment decisions. Child Dev *53:*1589, 1982.

49.14 SCHOOL CONSULTATION

RICHARD E. MATTISON, M.D.

HISTORY

School consultation has long been a common component in the clinical practices of child psychiatrists. School consultants have frequently focused on advising special educators who are teaching problem students. The need for mental health services by the 400,000 to 500,000 students receiving special education for their school behavioral or emotional problems is being emphasized anew. Epidemiological studies have begun to demonstrate that the majority of youth in the United States who need mental health intervention do not receive such care. When they do, a large proportion receive services in the schools, although the exact nature of how such services are delivered is not yet clear. Consequently, school mental health clinics are being investigated to reach the many unserved students with mental health needs. Thus, as school consultation grows and matures, the need is also increasing for more objective research and for increased collaboration with educators. Also, child psychiatry has begun to re-examine the instruction of trainees in school consultation to ensure their adequate preparation.

SCHOOL CONSULTATIVE PROCESS

Models for school consultation have largely been derived from the work of Gerald Caplan. A psychiatrist, usually under contract with a school district or a special education agency, is usually asked to perform a client-centered consultation, that is, the consultees are school personnel who have questions about a specific student. However, the principles of interaction with school staff are also applicable to a psychiatrist meeting with school personnel concerning a private patient. One model for a desirable sequence of steps is: information acquisition, initial meeting with school staff, student interview, parent interview, feedback meeting with school staff, feedback meeting with student and parent, and report.

Information Acquisition Prior to consultation the psychiatrist should receive from the school any formal reports that are in the student's file. Such history will provide the consultant with some initial understanding of the student prior to the meeting with the school personnel, thus helping the consultant to gather a more informed history from the school staff. Additionally, a teacher rating scale, such as the Achenbach Teacher Report Form, can help the psychiatrist gain an objective understanding of the student's current functioning in school.

Initial Meeting With School Staff A meeting with school personnel should precede interviews of the student and the parents. As the consultee, the teacher or school staff should first be encouraged to voice their concerns about the student and then to identify the specific questions that they wish the consultant to address. The psychiatrist will then expand and clarify this information, as well as any issues that previous reports may have raised. Although the staff may focus on one area of school dysfunction, the consultant must gain a comprehensive understanding of all the key areas of school functioning: attendance, academic (achievement levels and learning skills), social (peer and adult), behavioral, and emotional. School personnel should also be asked for their knowledge of the student's functioning at home and in the neighborhood, as well as environmental stressors. Finally, the staff should be questioned for their reasoning about causes of the student's school problems.

Student Interview Interviewing the student first will usually help to relieve the student's anticipatory tension about the meeting being important to the student's school future. The pupil should be told that the psychiatrist has been asked by the school staff for advice about the student's difficulties in school, but that confidentiality will be maintained. The youth's view of school functioning (in all major areas) should then be ascertained, including thoughts on the school's concerns if the student does not spontaneously remark on those issues. After this initial discussion of school matters, a standard comprehensive interview or mental state examination should then be conducted, beginning with any problems or worries that concern the student outside of school.

Parent Interview Next, the parents should be interviewed; the consultation process should be explained and their knowledge and thoughts about the school staff's current concerns should be ascertained. During this initial segment the parent should also provide the student's past school history because the school staff may have an incomplete educational history, especially if the child is in secondary school. A standard history should also be gathered from the parent, as would be obtained during an initial evaluation in the psychiatrist's office. Particularly important are parental concerns about the child outside of school, any past psychiatric problems the student may have had (including past and current mental health intervention), and family atmosphere.

Feedback Meeting With School Staff The consultant should begin with a summary of the interviews with the student and the parent(s), carefully protecting confidentiality. Diagnostic hypotheses should be given next, which will provide the school staff with a rationale for the consultant's final feedback step, that is, answers to the educators' initial consultation questions. The psychiatrist should then facilitate discussion by the school team, a multidisciplinary debate during which the consultant's own thinking on the case can be refined, especially to be sure that the initial questions of the staff have been addressed and that issues that were not raised at first have been identified.

The feedback meeting to the school staff should conclude with collaborative treatment planning for the student to include not only intervention steps in the school but also treatment needs to be met by community services. Team members can be especially helpful with actualizing treatment suggestions that the consultant may suggest for the classroom and school. At the conclusion of this feedback, staff should have a clear understanding of what school interventions will be undertaken with the student, and one member should become the case coordinator. They should also be aware of what community mental health services will be suggested to the student and the parent; the identified coordinator can follow through with the family to ensure that the outside help is obtained.

Feedback Meeting With Student and Parent Optimally, the case coordinator identified from the school team should participate in this feedback, as that staff member will be responsible for moving the case along both in and out of school. Separate feedback to the parents and then to the student is desirable in order to promote unimpeded reactions and questions that might not occur in joint feedback. Case coordinators should be identified as such and permission should be asked for their participation. Consultants should then explain their reading or diagnosis of the student's situation, followed by recommendations for school and community services that the psychiatrist and team have developed. Questions should be encouraged from the family to ensure their maximum understanding. Planning should begin for community services, especially to determine if the consultant is needed further (and thus proper written permission obtained). For example, the psychiatrist may need to phone the student's pediatrician to discuss a trial of stimulant medication, contact the child's current therapist to share the consultation and thus assist in treatment direction, or contact a mental health professional selected by the family to discuss the student's working diagnosis and probable treatment needs.

Report Preparation A good consultative report can not only influence the care of the student positively in the period after the consultation, but also potentially in subsequent school years by new school staff. Current school staff will return to the report to further develop their understanding of the student as well as the treatment rationale and school components. The report can also be very helpful to community mental health professionals who become involved in the case, and thereby promote their collaborative effort with the school team.

Consequently, a good report of a school consultation should provide a clear description of the student's symptoms in and out of school, both past and present, including a differential review. The student's medical and developmental history should be summarized, as well as any mental health treatments to that point of time. Pertinent family history should be highlighted, especially stressors. The mental state should be thoroughly described. Diagnoses from the fourth edition of *Diagnosis and Statistical Manual of Mental Disorders* (DSM-IV) should be stated as well as the consultant's formulation. Most importantly, at the end of the report each original question from the consultee should be answered, accompanied by any appropriate treatment suggestion for both in and out of school.

Finally, release of the consultation report must be strictly controlled. Parent or student permission, or both, should be gained before the report is sent to any external resource by the school or the consultant. Internally, the school must closely control which school personnel may have access to the report because it is a confidential, professional document.

COMMON PSYCHIATRIC DIAGNOSES

Attention-Deficit/Hyperactivity Disorders In general, school personnel have now become much better educated about the clinical features of attention-deficit/hyperactivity disorders through workshops, courses, and readings targeted for teachers. However, questions continue for consultants concerning treatment, especially medication and classroom interventions. Consultants must be careful to continue to teach or remind teachers of what medicines can and cannot do. For example, psychostimulants can positively affect concentration and hyperactivity, but they do not impart socialization skills or anger management, which must often be taught in conjunction with medication. Thus, psychiatric consultants can help teachers to become good observers of children's responses to these medicines, which will greatly help the physicians who regulate the medication.

School staff are often frustrated by a lack of communication with the doctor who is regulating a child's medicine. In such a case, the staff should first obtain written parental permission to communicate with the physician about the student's medicine (i.e., when necessary, the school should initiate and maintain communication with a child's doctor to increase the certainty of proper medication management). Then staff can select an objective teacher rating scale for attention-deficit/hyperactivity disorder, complete the instrument regularly for the student, and mail it to the physician. These instruments usually contain extra space where staff can briefly write their observations about the medicine's effects on a child, note any adverse effects, and ask specific questions of the doctor. Such a step by school staff can often stimulate proper communication with a physician.

Although teachers are now more knowledgeable about the clinical picture of attention-deficit/hyperactivity disorder, their skills in classroom intervention with such students are often not as advanced. When a consultant has determined that a child with the disorder needs classroom intervention, the capabilities of teachers and their support staff in the school must be ascertained. Depending on the answers to these key questions, the consultant will better understand what to advise.

The need to improve teachers' skills in addressing the specific problems of students with attention-deficit/hyperactivity disorder is now being recognized and addressed in different manners. Self-taught teachers can continue to increase their skills with a growing practical literature. Pressure is increasing on teachers' colleges to include basic instruction on this disorder in course work. Certain universities have begun to offer consultation packages to school districts, especially to train and maintain local personnel as resident experts in this condition. Finally, recognizing that very often even knowledgeable teachers do not have time to properly address these student's needs (especially children with complex or serious clinical pictures), some school districts are beginning to train special paraprofessionals to work in classrooms containing children with attention-deficit/hyperactivity disorder and thereby to more effectively provide intervention. For example, the Child Development Center at the University of California-Irvine has been at the forefront of establishing models of school-based intervention for such students (parallel teaching, paraprofessional, and multicomponent treatment).

Consultants must be versed in the interventions that have proven effective with attention-deficit/hyperactivity disorder in the classroom. Although many techniques appear promising, the actual supportive literature is small, partly because much of this research has been conducted in clinical treatment settings rather than in schools (where the prerequisite of data-based intervention is less easy to attain and where teacher resistance is an important consideration). Most research has focused on contingency management, followed by self-management and lastly actual academic instruction. Positive or token reinforcement, often in conjunction with response cost or time out, has proved to be effective in reducing activity level, increasing time on task, and improving academic performance. Manipulation of academic instruction or materials has also been shown to enhance academic performance. In contrast, cognitive-behavioral therapy has not proven as effective as anticipated in reducing core symptoms of the disorder. Efficacy may be compromised by the lack of a specific home or parent component and is usually enhanced when used in conjunction with medication, which is the most effective single treatment. Thus, multimodal treatment has been the desired standard of practice and is currently the focus of a multisite, longitudinal study of efficacy. However, at this time school consultants should remember that proven, cohesive educational programming for students in actual school settings is in the development stage. Although some necessary components appear to be understood, their application in real classrooms by regular teachers must be further developed, a task with which school consultants can be helpful.

Conduct Disorders This diagnostic category has proven very difficult to treat in any setting. Thus, psychiatrists consulting on students with conduct disorder will frequently meet school staff who are exasperated and pessimistic. Nevertheless, several important steps may improve such tense situations. For example, in a student with a conduct disorder, the consultant must accomplish a thorough differential to rule out common comorbid conditions such as attention-deficit/hyperactivity disorder, learning disorders, substance abuse, and mood disorders. These disorders must not be overlooked, especially because established treatments can improve their symptoms, which may in turn decrease conduct symptoms or allow more effective concentration on the conduct disorder.

Relatedly, school personnel will hope that a medication exists for quick assistance of students with conduct disorder. Unfortunately, consultants must instruct teachers that medicines have not proven very helpful with this disorder, except possibly lithium (Eskalith) or another mood stabilizer when unpredictable, uncontrollable aggression exists. However, an important medication issue often does exist in students who have conduct disorders. When such students have a comorbid condition that is responding well to a medication, staff must ensure that the medicine is given regularly. If the parents are unreliable (a common finding in the parents of children with conduct disorder), then staff members should get parental permission to give morning, noon, and even afternoon doses at school. Teachers should also alert the prescribing doctor about this need; the doctor may in turn be able to prescribe a long-acting version of the medicine to address the situation even better.

Commonly, consultants will be presented with students who have conduct disorder (without comorbidity) who are responding poorly to school interventions, but are also receiving little help for a stressful home environment (which may or may not have been defined). Students with conduct disorder are at risk for the experience of ongoing exposure to physical abuse or other violence, parent psychiatric illness, and disruptive parenting, which may be counterbalancing school measures. Teachers may be frustrated in dealing with such families, which in turn affects their work with the child. Therefore, a consultant must be sure that home stressors are defined and addressed as best as possible. This may necessitate extra time with the family during the consultation to clarify the issues, or a home visit to determine the actual atmosphere in which the student lives (such visits can also effectively be conducted by staff). Such extra time may well disclose situations that may be remedied: a woman who needs protection from her abusive boyfriend, a father who needs referral for his alcoholism, a mother who needs immediate treatment for her depression as well as in-home instruction on adequate parenting, a juvenile officer who has not yet contacted the school, or a student in need of a constructive after-school program. Thus, consultants must be ready in cases of conduct disorder to expend extra energy to ensure that comprehensive treatment is occurring as optimally as can be accomplished in that community.

Advice to school staff on school intervention must be realistic, that is, with the appreciation that conduct disorder is too often treatment resistant (especially with increasing patient age) and that no entirely successful protocol has been established for use in schools. However, much is known but too seldom is consistently applied by school staff, which is one roadblock preventing progress in students with conduct disorder. Some general principles have evolved that appear to be necessary for any progress: long-term application, home-school collaboration, peer involvement, and a systematic, integrated school program. Specific school components include social skills training, adult praise, individual and group positive reinforcement contingencies, time-out, and cost contingency. This difficult work by school staff can be greatly assisted by a recent landmark book on school interventions for students with conduct disorder: *Antisocial Behavior in School: Strategies and Best Practices*.

Finally, school consultants may also be asked about the general problem of school violence, a topic which requires a different working knowledge than the issue of conduct disorder. In order to offer credible help to school staff, consultants must keep familiar with this growing area—for example, by reading current reviews on school violence, which will then lead to more specific references of interest to the consultant. Any psychiatrist beginning to address this problem area should understand that it is a large and difficult undertaking, which involves early intervention (including preschool), multiple interventions for multifactorial causes, participation of the entire school staff, and collaboration by the school with parents and with community agencies, including juvenile authorities.

Ricky was a 9-year-old third grader who had recently been suspended from school after bringing a Swiss Army knife to school and showing it to classmates in the bathroom. The consultant to the special education system was asked to assess the boy's violence potential. He was also asked to help the multidisciplinary team decide if Ricky had conduct disorder and needed a special education classroom because of deteriorating behavior, including defiance and fighting.

This boy had shown some difficulties since kindergarten such as talking in class, disrupting others, being away from his desk, and not staying on task. Previous teachers had tolerated these problems because he generally redirected well, was usually in a good mood, and was liked by the other kids even though he often could not keep his hands to himself.

However, Ricky had changed since summer vacation. He seemed more irritable which caused him problems with everyone. With teachers, he redirected less well and began yelling that they were picking on him. He got increasingly angry when peers joked with him and began to swing at them. After such interactions he would often cry but not talk.

Testing by the school psychologist found Ricky to be rather closed about home, although he implied that there were some problems there. He was more open about his distress at school. He no longer liked school because “I'm always getting into trouble,” “Nobody likes me anymore,” and “I'm dumb.” Basic psychoeducational testing showed that his concentration was poor and his reading comprehension was at least one grade level behind, but that his intelligence quotient (I.Q.) was normal. His teacher completed a checklist for attention-deficit/hyperactivity disorder which scored positive for the disorder.

During the consultant's interview with the mother, she described the same problems at home. Although lovable, Ricky had always been a “handful,” and the family doctor was “watching him for hyperactivity.” However, since the summer the two of them were arguing much more. He had also begun to push around his little brother and was staying inside more. Her opinion of the knife incident was “one of the impulsive dumb things which he does from time to time.” She attributed his worsened behavior during the school year primarily to her separation from Ricky's father just before school started. Her husband's drinking problem had recurred this summer. When he began threatening her, although he did not physically abuse her or the children, she threw him out. His communication and visiting with the family had been erratic since that time, and he had not yet sought help for his alcoholism. She had been through this before with him, and although she was sad and edgy, she still hoped he would “come around.” The mother worried that without help Ricky was going to become a “truly bad boy,” which she did not feel was currently the case.

The consultant interviewed Ricky and found him rather open. Despite his interpersonal problems, he still liked his teacher and classmates. His solution to his school problems was for everyone to leave him alone if they saw he was “grumpy,” which he said was his most common mood lately. (He denied suicidal or homicidal ideation and vegetative symptoms.) He also wished he did not have to read so much in third grade because it was hard. As for the knife incident, he had not gotten into much trouble before for bringing toys to school; however, he should not have done it this time. He pointed out that, as usual, he wasn't thinking well; he remembered hoping it might win him a new friend, with no intention of using it as a weapon. Although he often felt grumpy or sad, he felt happy when he was in his room watching television or funny videos or playing with his computer (which did not get mad at him). There he was away from his brother and neighborhood friends and staying out of trouble for fighting with them. He was also away from his mother because she “was more grumpy, too.” However, he noted that they both worried about each other, and about how his father was doing. Ricky was also angry at his father for drinking again and not being around, and he worried that his parents might get divorced.

In his feedback to the school staff, the consultant did not feel the knife incident was indicative of potential violence, but more likely an impulsive occurrence. Rather than a conduct disorder, the psychiatrist felt Ricky had a chronic attention-deficit/hyperactivity disorder (combined type) and comorbid reading disorder—both previously untreated. These disorders were interacting adversely with an adjustment disorder with mixed disturbance of emotions (increased irritability) and conduct (defiance and fighting), related to the summertime occurrence of the father's drinking and the parents' subsequent separation. The knife incident was especially disconcerting to the school staff in light of another boy bringing a gun to school and accidentally wounding a classmate earlier in the year at a different city school.

As for school intervention, the consultant and school staff decided to further test Ricky for a possible reading disorder and then provide the appropriate help. The teacher was able to begin specific classroom interventions for attention-deficit/hyperactivity disorder, and would help Ricky assess his mood state each morning (related to home life). If he was especially grumpy then or during the day, the school guidance counselor would make herself available to talk to him. Fighting would continue to collect the school's regular punishment. In feedback to the mother, the psychiatrist suggested that he review with her family doctor the probable diagnosis of attention-deficit/hyperactivity disorder and need of a psychostimulant trial. He also suggested that individual or family therapy could help Ricky (and her) deal with the separation experience. The mother agreed with both suggestions. The consultant, family doctor, and school staff would work out the best therapy resource and get back to the mother. Finally, the school staff would update the consultant in approximately 1 month to reassess the boy's progress.

Depressive Disorders School staffs have become more knowledgeable about depressive disorders as they have learned about the common relationship of suicidal behavior and depressive illness, and as psychiatry has increased its efforts at educating the public about mood disorders. Nevertheless, consultants still face common situations in the diagnosis and treatment of depressive disorders in children and adolescents.

Teachers are now more likely to be concerned about a diagnosis of depression in a student who shows withdrawal, decreased performance, or persistent sad affect. However, they may still remain less likely to consider the possibility of depression in a student with new irritability, agitation, loss of concentration, or disruptive behavior. Teachers must be reminded that these acute externalizing symptoms can be associated with major depressive disorder in young students, and therefore they must consider, and observe for, less overt symptoms of depressive disorder. This step by an educator can lead to early diagnosis and treatment, before the externalizing symptoms receive growing negative attention and exacerbate the student's clinical state.

In contrast to underdiagnosis, consultants will encounter teachers who overdiagnose depressive disorders with the hope of a student's quick improvement through medication. A psychiatric consultant may then thoughtfully help the teacher begin to discern symptoms of major depressive disorder from depressive symptoms associated with stressful adjustments, comorbid disorders such as learning disorders and chronic dysthymic disorder. Also, teachers must be reminded that although the ability to treat depression in children and teenagers with medication is improving through the use of selective serotonin reuptake inhibitors (SSRIs), the appropriate clinical indication for use of antidepressants is primarily major depressive disorder.

Therefore, psychotherapeutic intervention is often the first treatment of choice in children with depressive disorder. Research in this therapeutic modality is limited in children and adolescents, but indications are promising for cognitive-behavioral therapy. In particular, consultants should be aware of school-based, group cognitive-behavioral approaches that are being developed for the treatment of depression in adolescent students. One well-advanced promising example is the Adolescent Coping with Depression Course, in which a small group of adolescents with depression is taught important skills for surmounting depression during 16 2-hour, after-school sessions. Different methods of instruction (lectures, discussions, role-playing, and homework) are used to teach relaxation, evoking pleasant events, dispelling irrational and negative thoughts, social skills, communication, and problem-solving; a parallel course can also be conducted with parents.

In association with depression, consultants will be asked about suicide-related issues. Recent school-based efforts at primary (courses) and secondary (screening) prevention have not yet proven very productive, except to push schools toward dealing with student mental health issues in a more serious and coordinated manner and with emphasis on both primary and secondary components. Some research has proven helpful; for example, when a student suicide has occurred, schools should especially concentrate on the evaluation of close friends of the student and pupils with known psychopathology, especially affective illness and previous suicidality. Thus, schools should have a protocol established to deal with the aftermath of a suicide by one of their students, as well as active suicidality in one of their students; psychiatric consultants with their knowledge of this specific literature can assist school personnel.

MENTAL HEALTH SERVICES IN SCHOOLS

Mental health services have traditionally been delivered in schools in a standard manner, with which consultants interface. Most public schools have designated staff to assist children with mental health problems; a school centrally located in a school district may also have special educators to work with students classified with behavioral or emotional problems. However, newer mental health services are evolving, which include the screening and monitoring of students at high risk for behavioral or emotional disorders, and the development of school-based mental health clinics.

School Staff Most regular classroom teachers can deliver modest mental health interventions to students; for more intensive programming, they need additional

trained manpower in their classes or frequent coaching from another specialized member of the school staff. When a teacher is concerned about the behavior or emotional state of a student, advice will usually be sought from the school guidance counselor (often a former teacher with a Master of Arts degree in Counseling). The counselor can screen the child for the seriousness of the problem, and then deliver help to the student with or without the teacher if the situation is not too serious. However, if the problem is more major, the counselor will suggest community mental health evaluation to the parents. School guidance counselors often also conduct group counseling for common problems experienced by schoolchildren, such as parental divorce or poor socialization skills. Finally, counselors can frequently assist consultants as case coordinators both in preparation for a consultation and in the later management of a treatment plan.

School psychologists usually have the most sophisticated knowledge among school staff on children's psychological problems. Thus, they can serve as invaluable resources for school psychiatric consultants, for example, in the design and execution of a behavior modification treatment plan. Unfortunately, the availability of school psychologists is often compromised by demands on their time to conduct primary testing or retesting of students. Many leaders in school psychology are currently trying to reverse this trend by advocating that school psychologists become more involved in the school-based assessment and treatment of children with psychiatric disorders. For example, with a carefully selected psychoeducational battery, a properly trained school psychologist can expertly assess a student for attention-deficit/hyperactivity disorder, refer a student for diagnostic confirmation and treatment planning, and then collaborate with the family's doctor to assess medication effects in school and to help the teacher deliver specific cognitive-behavioral techniques in the classroom.

Screening for Psychopathology School-based empirical procedures are being developed to economically and efficiently screen schoolchildren for the presence or risk of psychiatric disorders, both generally and specifically (e.g., disruptive and depressive disorders). Two important general models are emerging. First, the Achenbach battery of instruments (child interview or self-report plus parent and teacher behavior checklists) can be used to identify schoolchildren with significant psychopathology and then to plan and monitor their treatment. Initial work has begun to demonstrate how this battery can distinguish students who need special education.

A second model, the Systematic Screening for Behavior Disorders (SSBD) has been more thoroughly tested and utilized. The SSBD can screen children in kindergarden through grade 6 to identify students at risk for developing behavior disorders. The multigate screening process consists of three stages. In the first stage classroom teachers rank their pupils according to externalizing and internalizing problems (from specific lists of characteristics), and the top three in each dimension are considered to have passed gate one. In the second stage, these top children are then rated by their teachers on the Critical Events Index (a list of serious specific behaviors), and on the Combined Frequency Index for adaptive and maladaptive behaviors. Those students who exceed the normative-based cutoffs of the two instruments pass through this second gate and enter the third stage. In this last stage the designated students are observed in academic and playground (peer) settings and coded with specific observation protocols for their academic and social skills. The children who exceed these cutoffs and pass through this last gate can then be referred to a school's child study team for further monitoring, pre-referral intervention, consideration for community mental health services, or even special education evaluation. So far, the psychometric properties of the SSBD have proven sound, especially its ability to identify grade-school students with significant behavioral or emotional problems. Nationally, many school districts are investigating this screening system for standard implementation. Thus, consultants should be aware of the SSBD and other screening batteries if a school district asks for advice about such procedures.

School Mental Health Clinics In the past, rare community programs linked to schools were developed to actively evaluate and treat problematic students referred by teachers, but in the past, community mental health agencies and clinicians have primarily cared for only those students referred by schools who actually made office appointments. In addition, collaboration and communication between school staff and mental health professionals have generally been poor in the joint treatment of youth with psychiatric disorders. However, this undesirable situation now shows some promise of change with a movement toward establishing mental health clinics in schools.

Various reasons have stimulated this new direction. Epidemiological evidence has been accumulating that indicates that the majority of youth with psychiatric disorders do not receive any mental health treatment, and if they do, the intervention is commonly school based. Simultaneously, as school-based clinics have grown to provide for the general medical needs of adolescents (who commonly do not seek medical services in the community), clinical experience and research have begun to show that 30 percent or more of students attending school-based clinics seek help for psychological problems, for which such clinics are not properly staffed. Finally, pressure has grown on schools to provide counseling for special-education students who have dysfunctional behavioral and emotional problems. Guidance counselors may not be sufficiently licensed to deliver this mental health service or may have burdensome administrative demands, thus forcing schools to contract with community mental health professionals. Thus, driven by these intersecting reasons, a government-funded movement is beginning to develop mental health clinics in schools.

Currently such programs are just being developed and described, and the literature is limited. Some models already exist in a limited number of schools, from a modest satellite of a local community mental health center, to a full-service mental health clinic that also educates teachers as well as psychiatric trainees. School psychiatric consultants will have various roles in such clinics, from traditional consultation to actual service delivery (with the advantage of easier access to teachers and other school personnel for better coordination of treatment).

SPECIAL EDUCATION

Psychiatrists frequently serve as consultants to specialized school staffs serving students classified by the special-education category of *serious emotional disturbance*. Since 1975 through the Education for All Handicapped Children Act (EAHCA) (Public Law 94-142), children with disabilities, including behavioral and emotional, have been entitled to a free and appropriate public education. This basic educational right has been expanded through Public Law 99-457 (which in 1986 amended the previous EAHCA) and Public Law 102-119 (the 1991 Individuals with Disabilities Education Act [IDEA], which was reauthorized in 1997). The latter defines *serious emotional disturbance* as a condition having one or more of the following over a long period of time and to a marked degree and adversely affecting educational performance: an inability to learn that cannot be explained by intellectual, sensory, or health factors; an inability to build or maintain satisfactory interpersonal relationships with peers and teachers; inappropriate types of mood or behavior under normal circumstances; a general mood of unhappiness or depression; and a tendency to develop physical symptoms or fears associated with personal or school problems. A recent controversial issue is the consideration of conduct disorder by many states as a exclusionary criterion for serious emotional disturbance classification.

A consultant may be asked to help school staff determine if a student meets the criteria for serious emotional disturbance. If a student is already classified as having the condition, the consultant may advise on a revised treatment plan (also termed *individualized education plan*, which is required for any special-education student). Finally, although a student may not meet criteria for serious emotional disturbance, the student may show enough school dysfunction to qualify for educational assistance under Section 504 of Public Law 93-112 (Rehabilitation Act of 1973).

Students with serious emotional disturbance may be served for a flexible part of their school day in a regular school's resource room, which usually also serves students in other special-education categories. Or the setting may be more intensive in a separate self-contained classroom of the regular school or in a separate public or contracted school for children with serious emotional disturbance. In the more intensive settings, typically a special education teacher and trained aide educate 8 to 12 students with this condition. Token reinforcement programs are usually employed, with individualized tailoring to meet each student's academic, behavioral, and emotional needs. Groups are often used, for example, to improve socialization skills. Specific protocols on techniques may be utilized with targeted students or subgroups to treat such issues as depression, self-esteem, or anger. The staff are supported by school social workers and psychologists, in varying degrees of allotted time.

Empirical knowledge about students with serious emotional disturbance and their outcomes is gradually being established to assist educators and psychiatric consultants. Over 400,000 students (approximately 1 percent of the nation's student population) are classified as having serious emotional disturbance, predominantly boys (85 percent). They commonly have high rates of family stressors (divorce, abuse, or parent psychiatric illness) and externalizing psychiatric disorders, and the majority are underserved by community mental health professionals. Parent and teacher checklist ratings usually both indicate serious dysfunction, often two standard deviations above normal. When compared to students who are evaluated but subsequently not recommended for serious emotional disturbance classification, these students show significant differences: lower socioeconomic level, more abuse experience, more family stressors, greater diagnostic comorbidity, more serious clinician ratings, and higher externalizing ratings by teachers and clinicians.

Thus, these students present serious, complex clinical pictures and are among the most psychologically dysfunctional youth in any community. Their school staff clearly needs the consultative support of psychiatrists; indeed, these students should be one of the primary patient populations served by child psychiatrists. When working with such teams, consultants must be prepared to help teachers refine their cognitive and behavioral interventions with students, to communicate classroom observations to community physicians and psychiatrists to help them adjust medications for students with serious emotional disturbance, and to meet with and stimulate families to seek community interventions. The latter point is of primary importance because many families of such students need wrap-around services from

different agencies as an optimal treatment plan, although too often such coordinated care does not exist.

Past outcome literature for these students has been limited and pessimistic. These students annually have the highest dropout rate among special education children, approximately 35 to 40 percent. Follow-up studies of students after high school have shown poor employment history and high arrest rates. However, a recent longitudinal investigation has demonstrated a more optimistic picture for younger students in terms of successful completion of programming (high-school graduation or transfer back to complete regular programming). Furthermore, in a related study of the predictive value of baseline factors, four objective enrollment variables have been identified that significantly predict unsuccessful outcome for new students newly diagnosed with serious emotional disturbance (dropping out or transfer to an intensive community juvenile or psychiatric program): increasing age, the presence of a conduct or oppositional disorder, the presence of a verbal intelligence quotient (I.Q.) significantly lower than the performance I.Q., and the absence of a depressive or anxiety disorder. The converse is true for the prediction of successful outcome. Thus, prognostic information is beginning to emerge to help consultants and special educators understand which characteristics of newly diagnosed students are most predictive of success and which are indicative of poorer prognosis and the subsequent need for very careful treatment planning.

Consultants should be aware of research directions within the field of special education. Leaders are emphasizing the need to establish a more empirical base and to rigorously apply and study emerging successful treatment interventions (established in clinical populations not in special education) in these students in actual classrooms for children with serious emotional disturbance. Faculties are re-examining the interventions taught to prospective teachers and school psychologists to ensure the practicality of the techniques and the ability of the trainees to execute these methods with students. Collaboration with other disciplines is also being encouraged both in the clinical care and in the research of these students.

FUTURE DIRECTIONS

The traditional practice of school consultation by psychiatrists may well change in the future. Since children with dysfunctional psychopathology currently often receive mental health services in schools, such services will most likely continue to expand in order to reach more students. Thus, more psychiatrists will be doing increased school-based work, and must understand and be appropriately trained for collaborative intervention in the school environment. With this opportunity for easier multimodality treatment, they must be knowledgeable about school interventions that are efficacious for the specific child psychiatric disorders. Also, unlike the current primary focus on case consultation, schools may wish for more program consultation in the design and delivery of mental health services in the school, for which school consultants must be well prepared to advise.

Consultation to specialized programs may also change. These special education students already do not receive sufficient mental health services, and their numbers may grow secondary to the current insurance limitations on psychiatric hospitalization and residential treatment. Thus, the need for psychiatric consultation to staff for children with serious emotional disturbance will increase, along with the nature of the consultations as these special classrooms address the new short-term and long-term needs of their increasingly diverse students. Indeed, school programs for children with serious emotional disturbances may receive new interest as a therapeutic modality by mental health professionals.

Finally, as school consultation grows and changes, its empirical base must improve from its current initial stage. By its nature, such studies must be collaborative with educational investigators. Although some basic instrumentation exists to conduct such research, other important methodological components do not; for example, a structured interview of teachers for DSM child psychiatric disorders. Such increased objective studies will not only benefit school consultants, but also potentially benefit the treatment of child psychiatric disorders in general as current treatment studies rarely study school interventions. With increased knowledge about effective intervention with children in the very important environment of the school, the overall care of children and adolescent psychiatric disorders should benefit.

SUGGESTED CROSS-REFERENCES

Attention-deficit disorders are discussed in Section 39.1, disruptive behavior disorders are discussed in Chapter 40, and mood disorders in Chapter 14. In addition, learning disorders are presented in Chapter 35.

SECTION REFERENCES

Achenbach TM, McConaughy SJ: *School-based Practitioners' Guide for the Child Behavior Checklist and Related Forms*. University of Vermont Department of Psychiatry, Burlington, VT, 1998.

Adelman HS, Taylor L: Mental health in schools: Moving forward. Sch Psychol Rev *27:*175, 1998.

American Psychiatric Association: *Psychiatric Consultation in Schools*. American Psychiatric Association, Washington, DC, 1993.

Birmaher B, Ryan ND, Williamson DE, Brent DA, Kaufman J: Childhood and adolescent depression: A review of the past 10 years. Part II. J Am Acad Child Adolesc Psychiatry *35:*1575, 1996.

*Bostic JQ, Rauch PK: The 3 R's of school consultation. J Am Acad Child Adolesc Psychiatry *38:*339, 1999.

Brent DA, Holder D, Kolko D, Birmaher B, Baugher M, Roth C, Iyengar S, Johnson BA: A clinical psychotherapy trial for adolescent depression comparing cognitive, family, and supportive therapy. Arch Gen Psychiatry *54:*877, 1997.

Brent DA, Kerr MM, Goldstein C, Bozigar J, Wartella M, Allan MJ: An outbreak of suicide and suicidal behavior in high school. J Am Acad Child Adolesc Psychiatry *28:*918, 1989.

Caplan G: *The Theory and Practice of Mental Health Consultation*. Basic Books, New York, 1970.

*Clarke G, Lewinsohn P, Hops H: *Leader's Manual for Adolescent Groups. Adolescent Coping with Depression Course*. Castalia, Eugene, OR, 1990.

Committee on Community Psychiatry and Consultation to Agencies: *A Core Curriculum for Training in Community Child Psychiatry and Consultation to Agencies*. American Academy of Child & Adolescent Psychiatry, Washington, DC, 1995.

DuPaul GJ, Eckert TL: The effects of school-based interventions for attention deficit disorder: A meta-analysis. Sch Psychol Rev *26:*5, 1997.

*DuPaul GJ, Stoner G: *ADHD in the Schools: Assessment and Intervention Strategies*. Guilford, New York, 1994.

Erchul WP, Martens BK: *School Consultation: Conceptual and Empirical Bases of Practice*. Plenum, New York, 1997.

Jensen PS, Hoagwood K, Petti T: Outcomes of mental health care for children and adolescents: II. Literature review and application of a comprehensive model. J Am Acad Child Adolesc Psychiatry *35:*1064, 1996.

*Kauffman JM, Lloyd JW, Hallahan DP, Astuto TA: *Issues in Educational Placement. Students with Emotional and Behavioral Disorders*. Erlbaum, Hillsdale, NJ, 1995.

Knitzer J, Steinberg Z, Fleisch B: *At the Schoolhouse Door: An Examination of Programs and Policies for Children with Behavioral and Emotional Problems*. Bank Street College of Education, New York, 1990.

*Mattison RE: School consultation. In *Child and Adolescent Mental Health Consultation in Hospitals, Schools, and Courts,* GK Fritz, RE Mattison, B Nurcombe, editors. American Psychiatric Press, Washington, DC, 1993.

Mattison RE, Gamble AD: Severity of socially and emotionally disturbed boys' dysfunction at school and home: Comparison with psychiatric and general population boys. Behav Dis *17:*219, 1992.

Mattison RE, Morales J, Bauer MA: Adolescent schoolboys in SED classes: Implications for child psychiatry. J Am Acad Child Adolesc Psychiatry *32:*1223, 1993.

Mattison RE, Spitznagel EL, Felix BC: Enrollment predictors of the special education outcome for students with SED. Behav Disord *23:*243, 1998.

Miller DN, DuPaul GJ: School-based prevention of adolescent suicide: Issues, obstacles, and recommendations for practice. J Emot Behav Dis *4:*221, 1996.

Nastasi BK, Varjas K, Bernstein R, Pluyment K: Mental health programming and the role of school psychologists. Sch Psychol Rev *27:*217, 1998.

Pfiffner LJ, Barkley RA: Treatment of ADHD in school settings. In *Attention-Deficit Hyperactivity Disorder. A Handbook for Diagnosis and Treatment,* ed 2, RA Barkley, editor. Guilford, New York, 1998.

Shapiro ES: Training school psychologists for service delivery to children with severe emotional disturbance. Sch Psychol Rev *20:*485, 1991.

Stark KD: *Childhood Depression: School-Based Intervention*. Guilford, New York, 1990.

Swanson JM: *School-Based Assessment and Interventions for ADD Students*. K. C. Publishing, Irvine, CA, 1992.

Sylva K: School influences on children's development. J Child Psychol *35:*135, 1994.

*Walker HM, Colvin G, Ramsey E: *Antisocial Behavior in School: Strategies and Best Practices*. Brooks/Cole, New York, 1995.

Walker HM, Severson HH: *Systematic Screening for Behavior Disorders (SSBD) Technical Manual*. Sopris West, Longmont, Colorado, 1992.

Weist MD: Expanded school mental health services: A national movement in progress. In *Advances in Clinical Child Psychology,* vol 19, TH Dendick, RH Prinz, editors. Plenum, Washington, DC, 1997.

Zahner GEP, Pawelkiewicz W, DeFrancesco JJ, Adnopoz J: Children's mental health service needs and utilization patterns in an urban community: An epidemiological assessment. J Am Acad Child Adolesc Psychiatry *31:*951, 1992.

49.15 PSYCHIATRIC PREVENTION IN CHILDREN AND ADOLESCENTS

NORBERT B. ENZER, M.D., AND STEPHANIE L. HEARD, M.D.

Astounding, virtually inestimable successes have been realized in child health through prevention. However, despite the fact that many identify the preventive work of William Healy in Chicago during the first decade of the twentieth century and the subsequent spread of the child guidance movement as one of the important historical roots of American child psychiatry, prevention in child psychiatry has been characterized more by hope and rhetoric than by investment and success. Perhaps, as the century draws to a close there is reason for a somewhat more optimistic outlook in the future.

Historically, prevention in regard to general child health focused on infectious diseases and drew heavily on the sciences of epidemiology, microbiology, and immunology. Infectious diseases yielded to epidemiological study and the perception that the emergence of disease in individuals could be explained by factors in the host, the agent, and the environment. This conceptualization contributed enormously to the prevention of communicable diseases in children via immunization and the reduction of such diseases as typhoid fever and cholera via sanitation and protection of the water and food supply.

Work on infectious disease taught some useful lessons. William Foege's strategy of “search and containment” rather than mass vaccination alone for smallpox may be the precursor to the contemporary concept of at-risk populations. Immunization to increase the resistance in particular populations proved incredibly successful, and the last human case of smallpox was reported in 1977.

Unfortunately, despite the splendid success of the battle against smallpox, generalizations are limited. There is no animal reservoir for the smallpox virus and no human carrier state. Transmission is only from person to person and only during the active phase of the rash. Immunity is long lasting after recovery (if it occurs) or vaccination and is evident from visible scars. Not only has the disease disappeared from the world, so has the variola virus. Although achieving similar success with other diseases is unlikely, it is evidence that concerted, massive effort can have astounding results. When faced with an endemic necessary agent, enhancing host resistance is clearly a wise choice. The successes of poliomyelitis, rubeola, tetanus, and other vaccines speak for themselves.

But the history of medicine has examples of primary prevention efforts that succeeded without the knowledge of a necessary offending agent. In 1847, Ignaz Semmelweis showed that the incidence of puerperal fever was reduced if physicians and midwives washed their hands between deliveries—decades before Koch and Pasteur and the germ theory of disease. Puerperal fever had a necessary offending agent but crude antiseptic practice reduced morbidity without any specific knowledge of the cause. As unthinkable as it seems today, Semmelweis' approach was controversial at the time, and he was the target of ridicule.

Later, other factors were more fully appreciated. The tubercule bacillus was clearly the necessary agent in tuberculosis but was not always sufficient to produce disease. Host factors such as adequate nutrition and environmental factors such as better housing seemed to increase resistance to the organism well before there was a clear understanding of the body's immune system. Our understanding of these multiple factors and the development of effective pharmacological agents clearly contributed to the major decline several decades ago in the prevalence of tuberculosis. The current increase in incidence can be explained by impaired host immunity (due to acquired immune deficiency syndrome [AIDS]), changes in the agent (drug resistance) that decreased its susceptibility to antitubercular medications, and such factors as homelessness and poor nutrition in some.

Some diseases are largely, if not wholly, genetically determined (e.g., Huntington's disease, sickle cell anemia, and hemophilia). In genetic disorders, unlike infectious disease, a single host factor is necessary and sufficient to produce disease. Education and genetic counseling have proved useful in many of these disorders, and more-specific interventions have reduced morbidity and mortality.

Although the number of disorders in which specific genetic patterns are the necessary and sufficient cause may well be limited, one's genetic makeup can increase the risk of disease in virtually every system, including cardiovascular disease, cancer, arthritis, and some mental illness, particularly affective disorders and schizophrenia. The potential of genetic alterations as preventive interventions no longer seems to be futuristic science fiction.

Along with the remarkable advances in molecular genetic knowledge has come a similar recognition of the etiologic influences of environmental toxins—carcinogens in air and water, lead, and perhaps even certain foods. Equally toxic to the developing child are life stresses, inadequate nutrition, poverty, homelessness, trauma, inadequate parenting, prejudice, violence, and other noxious influences to which a child may be exposed at critical times in development. Reducing the potency and prevalence of these factors has proved to be extremely difficult and complex. Nevertheless, all of medicine now looks to multiple causality, risk factors, and complex interactions among them to further explain disease and broaden our understanding of illness.

RISK FACTORS IN CHILD AND ADOLESCENT PSYCHIATRY

Prevention in child and adolescent psychiatry is focused strongly on identifying populations at risk and reducing the factors that contribute to increased risk. Table 49.15-1 lists some of the many risk factors identified in the literature. Risk factors correlate with poorer mental health outcomes in populations but not necessarily in individuals, and although some factors may appear to have higher rates of correlation than others and thus appear more potent, certain factors may carry the likelihood that others will also be present. Risk does seem to increase when the number of risk factors increases, but it is not the number of factors alone that seems to increase negative outcomes. Context, the point in development when an event occurs, and any protective factors that may be present also influence outcomes.

Chronic illness or disability, particularly those involving the central nervous system
Limited academic skills
Emotional abuse or neglect
Physical or sexual abuse
Death in the family
Psychopathology in parents
Conduct disorders in parents
Substance abuse in family or child
Suicidal behavior in family
Separations and losses
Family discord and divorce
Economic distress in family
Poverty
Racism
Neighborhood violence
Limited social and educational opportunity
Environmental toxins
Natural disasters
Accidents

Table 49.15-1 Risk Factors for Children and Adolescents

Although child-oriented professionals are often tempted to focus on risk factors in the social environment or family, the child's contribution must not be ignored. Central nervous system problems that result in obvious disabilities (e.g., motor, perceptual, speech, or language problems), seizures, or mental retardation clearly

increase risk. However, more-subtle problems of temperament, appearance, and even interest or talents that depart from parental expectations or values may create a high level of risk for some.

Certain risk factors are more potent at particular times in the development of the child or the life of the family. These are dynamic, constantly changing interactions. Perhaps this is why the identification of patterns or clusters of risk factors that predict specific disorders, problematic behaviors, or maladaptations has been elusive. Similar clusters of risk factors may be associated with different outcomes, and analysis of the histories of individuals with similar disorders often yields only limited recognizable predictive patterns. Perhaps in the future certain discrete causality patterns will aid prevention or amelioration of mental and behavioral disorders, but this seems unlikely, at least in the near future.

Our current knowledge of mental disorders demands that prevention efforts recognize an enormous array of risk factors whose potency depends not solely on the strength of each factor alone or in combination, but on the number of factors present, the response of individual children and their families to the various events or interactional patterns in their environment, the time in the development or life cycle when the factors occur, and the personal, interpersonal, and material resources that dictate whether the response will be adaptive or maladaptive.

PROTECTIVE FACTORS

Less attention has been given to protective factors—those that seem to create resistance or counterbalance risk factors. Nevertheless, there is a growing recognition that they are important and may help explain why similar risk patterns do not affect all children equally. Protective factors can only be defined in relation to risk factors occurring in similar social or environmental contexts. Among the protective factors that have been identified are warm and supportive parenting; effective social support systems; effective limit setting; competent academic, interpersonal, and social skills; and adequate social and educational opportunities.

TARGETS OF INTERVENTIONS

The obstacles to major investments in preventive child psychiatry are formidable and the results of efforts intended to be preventive are often inconclusive. Until recently, the lack of a widely accepted nosology has been a major handicap. In addition, etiologic knowledge is limited and disagreements are based more on theory and belief than on scientific information. The target of a particular preventive intervention is often unclear. Is it a specific, diagnosable disorder occurring during childhood or adolescence or some general maladaptive behavior that interferes with maximum functioning? Is it a disorder that occurs in adulthood, a specific disorder or a pattern of behavior that interferes with social functioning and success? Or is the intervention intended to increase general adaptability and resilience?

Efforts aimed at promoting general well-being and developmental success have often been associated with prevention. Many of society's ills clearly affect children directly and indirectly; some negative consequences may be fairly subtle and may not appear for years while others are more immediate. Table 49.15-1 shows that the list of risk factors is lengthy, and many are widespread. Because some of these problems are so pervasive and heart wrenching, they have often been the focus of much political rhetoric and, at times, action. Unfortunately, often in the lives of many children these factors occur in combination, and the interactions between them are not well understood. Consequences differ from one individual to the next, with some suffering little while others experience various patterns of profound maladaptation.

None of these social factors can be considered a necessary and sufficient cause for any subsequent specific, definable psychopathology. Nevertheless, numerous statistical correlations persuasively suggest causal relationships and motivate many to call for action.

Addressing society's ills in deliberate and organized ways has often proved fruitless. Proposed solutions are frequently controversial and lack evidence that they will be effective. Some programs in the past that were enthusiastically supported have clearly not achieved the successes that were expected, particularly when they attempted to change behavior in large population groups (e.g., reducing smoking and drug and alcohol abuse use among teenagers).

In contrast, promoting the use of car seats and bicycle helmets for young children seems to have reduced accidental injuries and deaths. Like efforts to reduce smoking and alcohol consumption, these efforts have been supported by law. The difference is that the use of car seats and bicycle helmets focuses on specific, immediate injuries in young children and actions by parents, whereas smoking and alcohol and drug consumption involve the behavior of adolescents and often, delayed consequences.

Targeted Social Forces Despite uncertainty, some large-scale social programs seem to have achieved enough success to receive sustained support. Perhaps the most widely known is the Head Start Program, begun in the 1960s. Originally designed to enhance the early school performance and social development of underprivileged children, Head Start involved summer programs for 4- and 5-year-old children and the early results were positive.

A follow-up study to evaluate its effectiveness was awarded to the Westinghouse Learning Corporation 4 years after the start of the program. This study found that certain early results were not sustained and raised serious questions about the program, particularly its cost-benefit. The particular finding that increases in intelligent quotient (I.Q.) were not sustained into elementary school led to the conclusion that the program was a failure. Later analyses of the methods used and the particular outcomes that were assessed concluded that the study was flawed in several ways. Changes in I.Q. were used instead of school achievement; all local programs were assessed equally, with no adjustment for differences in quality or early results; and little consideration was given to of less specific outcomes. Much later evaluations demonstrated broad, general effectiveness with better school progress, fewer school dropouts, less delinquency, and better work records after high school. In retrospect, analyses suggested that the lack of even more sustained benefit was due to the brevity of the intervention and the ongoing negative influences of the social environment.

Evaluations of a number of other early intervention programs establish that there are lasting effects such as increased graduation from high school, reduced grade failures, and fewer children referred to special education. However, the risk factors are exceedingly complex, cumulative, and often not outweighed by opportunity. More children and families are participating in early intervention programs that have broad goals not too dissimilar from those of Head Start, although they may be conceptualized differently and involve other specific intervention techniques. The risk factors are at least the same and for many even more severe and complex, so positive effects are more limited, and public support may be less enthusiastic. Unfortunately, expectations can be very high, and positive outcomes that fall short of those expectations are often seen as failures. Well-intentioned programs, even some with positive early results have been abandoned without sufficient time for full intervention or to assess long-term outcomes.

Damaging Agents Early interventions that address highly specific factors not directly related to social forces have been more clearly effective. These often involve the avoidance of particular damaging agents. The effects of phenylketonuria, lead poisoning, and fetal alcoholism can be reduced by specific early intervention. Neonatal screening and the elimination of phenylalanine from the young child's diet has reduced the devastating behavioral problems and early mortality associated with this inborn error of metabolism. Reduction in lead concentration in water supplies and removal of lead from paints has reduced lead poisoning in children with its subsequent influence on central nervous system functioning and learning.

Lead-based paints in old homes are a source of lead intake for some children. Early and regular screening with adequate treatment if necessary can reduce the risks of biological damage but not the initial risk. The control of lead in the environment and the required removal of lead paint from old, inhabited buildings has raised arguments about cost-benefit of this approach compared with the cost of screening and treatment. Furthermore, since lead poisoning is less frequent now, is true primary prevention worth the cost? This issue is of concern in other areas of prevention in child mental health. Even when knowledge is sufficient and methods of prevention are proved, is society willing to bear the cost of preventing relatively rare disorders when effective treatment is available?

Fetal alcoholism raises a different issue. Even modest consumption of alcohol during pregnancy can damage the fetus and result in the syndrome of fetal alcoholism. Unlike the situation with lead poisoning, the elimination of fetal alcoholism does not depend upon community action nor can it be reduced by legislation as lead poisoning could. However, the lack of will of society, individuals, or both clearly limits the effectiveness of prevention in these specific disorders as well. Lead concentration in water supplies is still too high in some communities and far too many women consume alcohol during pregnancy despite public education efforts and prenatal care.

Resistance to Intervention It is perhaps not surprising that some efforts at early intervention have such limited success even though evidence suggests that the potential is much greater. Resistance to intervention at a community or individual level is complex, and efforts to reduce resistance are often simplistic. Although a lack of funding for prevention is frequently cited as the principal limiting factor, even that is not a simple matter, and programs with low cost-benefit ratios still meet resistance. Far too many children are still not being fully immunized, and the number may be increasing among some groups of children. Programs requiring greater

funding with less-specific, proved results; more-complex interventions involving changes in attitudes, beliefs, and behaviors; and those that threaten values or other community priorities are clearly more difficult to implement and sustain.

Much early intervention is focused on education, enrichment, and socialization with the assumption that the intervention will compensate for, or increase resistance to, risk factors. These programs often engender great enthusiasm and high expectations. But the increasing frequency and intensity of real world risk factors and their complexities mitigate against success, strengthen skepticism, and diminish public support.

Scientific evidence of benefit is not always the critical factor in convincing the public and its leaders of the value of preventive or therapeutic programs in health or mental health. Groups have embraced approaches with no evidence of success. Nevertheless, positive outcomes from carefully planned and studied interventions are of inestimable value, and evidence does exist for some programs.

CURRENT APPROACHES

In 1964, coincident with the federally supported Community Mental Health movement, Gerald Caplan published the first major American textbook exclusively devoted to prevention in child psychiatry. He defined three categories of preventive interventions in mental health. *Primary prevention* focused on preventing disorders or problems in the future, and usually the targets were normal populations. *Secondary prevention* addressed problems in their very early stages and was intended to reduce or delay the development of related or more serious problems. *Tertiary prevention* was largely defined as intervention aimed at reducing the duration or impairment of disorders or problems. Looking back at Caplan's work shows just how far prevention has come in the past quarter century. Enthusiasm for primary prevention, as defined by Gerald Caplan, has waned because of costs, lack of sustainability, methodological difficulties, and the lack of evidence of sufficiently positive results. But there has been growing interest in other types of intervention. Although more recent programs are often far less inclusive, they are more focused and more thoughtfully conceptualized and they have better follow-up. Research has expanded enormously; over 30 percent has been published since 1990.

More recently, following a study by the Institute of Medicine, prevention has been conceived as including only those interventions prior to the onset of significant problems. This discussion considers only efforts that can be considered primary or secondary prevention. The Institute of Medicine suggested new terminology—*selective preventive intervention* for primary intervention targeted at high-risk populations, such as families with parental substance abuse and *indicated preventive intervention,* which focuses on children with mild behavioral symptoms. *Universal intervention* is often used as a substitute for *primary prevention*.

PRIMARY PREVENTION (UNIVERSAL INTERVENTION)

Many programs that can be classed as primary prevention have been tried, and they often differ substantially in their approaches. Some attempt to intervene with all children or parents in particular groups, irrespective of the presence of risk factors, for example, all children in a particular school, all at a particular grade level in a school district, all kindergarten parents or families living in a geographic region. This approach can be considered to have a universal target population that is normal, at least in theory. Among such efforts are programs to reduce teenage smoking and substance abuse, adolescent pregnancy and sexually transmitted disease, and programs intended to enhance social and interpersonal development.

Opposition Opposition to primary prevention programs can be substantial, at times spilling over into contentious exchanges between opponents and supporters in the community. The introduction of drug and sex education programs in schools has been particularly troublesome, but programs on social skills, conflict resolution, and affect sensitivity training have met with resistance as well. The resistance has taken many forms—the belief that such programs intrude into the prerogatives or values of parents, that the outcomes will result in promoting the very behaviors the programs intend to diminish, that the resources invested in the prevention are better devoted to teaching traditional academic skills and knowledge, or that there is little evidence that the intervention will be successful. Nevertheless, community efforts to promote the welfare and protection of children, reduce violence, and enhance opportunities continue, largely outside the purview of psychiatry. Some evidence suggests that such efforts as the direct involvement of children in school-based education can reduce the risk of trauma to children from sexual abuse and violence. Do not underestimate the intensity of the controversy. Its ripple effect has influenced school board composition; the curriculum in science, social studies, and literature; and library holdings. Prevention can be truly controversial, and community understanding and support is critical.

Although multilevel interventions are more likely to show positive results, single-level programs have been successful. Teaching parenting skills to parents at risk for child abuse has reduced the critical initial episodes of abuse. Likewise, parental monitoring of adolescents after educational programs has reduced teenage drug abuse.

Some programs, previously considered primary prevention, may now be better classified as selective preventive intervention in keeping with the Institute of Medicine definition. These programs address target populations that are at risk for problems yet to emerge, such as children whose parents have been diagnosed with mental disorders, who have been abused, or who are homeless or living in severe poverty. Other target populations include children or families who are at the point of critical and often especially stressful transition, such as divorce, death in the family, or entrance into a new and unfamiliar school, and those who require substantial medical and surgical care.

Levels of intervention also differ. Some can be considered environmental, involving the community as a whole, a school, peer groups, or the family; others may focus on the individual. Interventions frequently include more than one level, and multilevel interventions seem more likely to have positive results. Children facing substantial surgical procedures may be the focus of individualized specific techniques to inform, reduce apprehension, and increase compliance. Their parents may participate in education, support group activities, and training in the care and emotional support of the child; and the activities of hospital staff are changed specifically to attend to the psychosocial needs of child patients and their parents.

Prevention programs require recognition that many individuals and factors interact with one another in dynamic ways. These transactions are multidimensional. Parent affects child—child affects parent; school affects child—child affects parent; parent affects school—school affects peers; peers affect child, and so on. The transactions are complicated, changing, and powerful.

Evaluation Evaluating primary prevention programs is difficult. Establishing the existence of lasting positive effects is hampered by all of the problems inherent in long-term interventions and outcome studies. Positive effects must be defined as well as changes that should not be attributed to the intervention. Nevertheless, studies have shown successes; and some metanalyses of multiple studies support effectiveness in follow-up. Notable specific universal programs include Head Start; a substance abuse prevention program with life skills training of seventh graders with reinforcement in the eighth and ninth grades; and a 6-year school-based program to reduce substance abuse, delinquency, and school failure among low-income urban children. Two examples of successful selective preventive intervention programs are the Comprehensive Child Development Program targeted at low-income families and the Strengthening Families Program aimed at children of families with substance abuse problems.

SECONDARY PREVENTION INDICATED PREVENTIVE INTERVENTION

Secondary prevention is intended to reduce the probability that children and adolescents with early or minimal signs of maladaptation will develop more serious dysfunction or specific clinical disorders in the future. It depends upon three fundamental assumptions: (1) those with early disorders can be identified, (2) they are at greater risk for more serious problems in the future, and (3) secondary preventive intervention will improve later adaptation and function.

Although one cannot predict which of those children with early or minimal problems will develop more serious problems, it is clear that as a group they are more likely to have greater difficulties in the future. They are at greater risk than their normal peers. Although one cannot predict specific future problems, patterns or sequences of compounding problems often appear in specific children.

Poor academic performance in early grade school may be followed by discouragement, damaged self-concept, poor peer relationships, behavior problems in school, a reputation of being a troublemaker, and later depression and perhaps rebelliousness and substance abuse in adolescence. Interpersonal aggressiveness in a preschool child may lead to harsh discipline, punishment, increasing and more-generalized anger, social isolation, contact with law enforcement, and later, more-specific delinquent behavior.

The complexities of these sequences and the interactions between child, parents, peers, teachers, and others should not be underestimated. The child who is not doing well in school may be under considerable pressure from parents or teachers to work harder or get special assistance and may, as a result, be prevented from

participating in a variety of nonacademic activities that can be so useful in social development and peer interactions. Labeling a child “stupid,” “lazy,” or a “troublemaker” may actually create a self-fulfilling prophecy and reinforce the problematic behavior. The act of identifying early or minimal problems in a child runs a similar risk, particularly if appropriate intervention is not available or accessible.

Selection and Assessment Although space does not permit a full discussion of the many varied approaches to screening to identify children with early or minimal problems, it is the first step, and the approaches usually involve informal reports from parents, teachers, peers, and others in contact with the child. Unfortunately, such early reports often do not receive sufficient concern, even from expert professionals. Such reactions not only can delay further evaluation, they may reassure parents and create greater expectations, perhaps further compounding the problems. Formal screening using psychometric or sociometric instruments and behavioral inventories has proved useful. Whatever screening methods are used to identify children who may benefit from a preventive intervention, a more complete assessment is then necessary. Similar problems may have different origins, and both the strengths and protective factors and the problems and risk factors must be considered.

Multiple levels of assessment are most useful. The first level may involve informal reports or more-formal simple screening instruments such as peer rating scales or brief behavior check lists completed by parents, teachers, or both. The next level may involve direct observations, psychological tests, or interviews with children, parents, and teachers. The third level involves more-complete individual evaluation of those who have passed certain thresholds at the earlier levels. Such approaches must be carefully implemented to avoid outcomes that are overly inclusive and yet be sensitive enough to avoid significant underidentification.

Resistance Along with resistance similar to that generated by primary prevention programs, secondary prevention programs may generate resistance because of the continuing stigma regarding emotional and behavioral problems and psychiatric disorders. The identification of problems or even potential problems may threaten parents. Even if parents recognize a problem, they may consider it a disciplinary problem, laziness, or a variation of normal (“that's just the way the child is”) or believe that the problem is not in the child but rather in the school or the relationship with the teacher. Since secondary prevention almost always requires the active participation of parents, their views of the problem and the proposed interaction and their capacity or willingness to be involved limit many efforts.

Outcomes Beginning with the Primary Mental Health Project at the University of Rochester in the mid-1950s, a great many school-based secondary prevention programs have been reported. Overall, some of these programs have effectively enhanced the adjustment of the children or adolescents in the target populations. The outcomes of multiple secondary prevention programs that addressed different target populations, with different types of problems, using different intervention techniques have been reviewed. In general, children with externalizing behavior problems, internalizing problems, poor peer elationships, and poor school achievement all showed positive effects. Those with externalizing behavior problem showed the most-positive change, and those with poor school achievement had the least-positive results. Of the various techniques used, behavioral and cognitive-behavioral approaches were the most effective. Although outcome studies of school-based secondary prevention programs aimed at reducing the frequency of specific disorders are often inconclusive, evidence suggests that better results may be possible. Several studies have shown that episodes of affective disorders can be reduced by school-based programs.

School-based behavioral and cognitive-behavioral programs have been less effective for children with academic problems than for others, but numerous programs have demonstrated that children with learning problems can be helped if interventions are implemented early enough—often before entry into elementary school—and if the interventions are comprehensive and of sufficient duration. Effectiveness may require changes in the schools themselves, including staff development, collaborative decision-making, social skills training, and parental participation.

Prevention Outside School Secondary prevention efforts go well beyond those in schools. Home-based interventions often are part of broader community efforts to reduce child abuse, and they have been shown to enhance the effectiveness of parenting programs. Such programs have also been effective in enhancing cognitive development and in promoting general health care. However, these programs are often considered costly by some and intrusive by others, and parental involvement is critical whether the interventions are within the home or elsewhere.

The problem of foster care is particularly important especially when it results from a judgment that the natural home of a child is unsuitable or unsafe. Truly professional foster homes are useful but unfortunately not readily available. Foster parents supported by readily available and regular professional support and consultation are a positive alternative. However, the experience of many children remains one of multiple foster homes and increasing difficulties.

ROLE OF PSYCHIATRISTS IN PREVENTION

Although few general and child and adolescent psychiatrists have been involved in the design, implementation, and evaluation of primary and secondary prevention programs, others have assessed or given therapy to children or adolescents and their families who needed more-individualized attention. Others have functioned as advocates for prevention in schools, communities, and the political arena. However, most general or child and adolescent psychiatrists have not devoted much attention to prevention, and many find it difficult to define a role for themselves. Many may feel ill prepared, for although accredited training programs must provide some teaching in prevention, few offer more than some lectures on the subject.

Individual practitioners can have a preventive influence within their own practices without substantial changes. Recognizing potential risk factors and inquiring about these factors into the assessment of all referred children may identify children at greater risk who might benefit from some intervention short of a full treatment program for a particular disorder. Obviously, general and child and adolescent psychiatrists are not the only physicians for whom such behavior is appropriate. In fact, it may be even more appropriate for pediatricians, family physicians, and other primary health professionals. Further, such attention should not be limited to physicians who see children or adolescents. All need to raise the question “what about the children?” in their contacts with adult patients who are parents or potential parents.

Of particular importance is the presence of psychopathology in parents, especially psychosis, depression, substance abuse, alcoholism, and antisocial behavior, or a history of abuse or neglect in their own childhood or current family discord, disruption, or abuse. In such cases, comprehensive inquiry into the development and functioning of the children is critical, and a direct assessment of the children may be required. Intervention will depend upon the findings of assessment, but no intervention is possible unless the questions are asked. The presence of diagnostic criteria for a specific disorder must not be the sole basis for preventive intervention. Likewise, reports from parents, teachers, or others that suggest deviations in development or socialization or academic difficulties should be taken seriously and not dismissed as variants of normal without sufficient clinical evidence. All too often a wait-and-see approach is taken without adequate assessment or specific follow-up planning. Careful documentation of clinical findings is essential to permit follow-up comparisons and assess progress or the lack of it.

Practitioners must be aware of the resources in the community that are available to families who need a preventive intervention in addition to, or instead of, a therapeutic one. Support groups for divorcing parents, the spouses and children of substance- or alcohol-abusing parents, and parents of children with attention or learning disorders, and classes on child development and parenting are examples of the resources available in many communities. Enrichment and socialization opportunities for children and adolescents often available through school systems, the YWCA and YMCA, Boys and Girls Clubs, and others can increase self-esteem, skill development, and the productive use of time.

Practitioners need to have a preventive attitude—a belief that early recognition and intervention can be helpful and that the effects of risk factors can be minimized, protective factors can be enhanced, and the path to psychopathology or maladaptive behavior can be altered.

OUTCOMES RESEARCH

Research into the outcomes of prevention programs is difficult, and results are often inconclusive. Nevertheless, meta-analyses of multiple studies do suggest that prevention in child and adolescent mental health can yield positive results. The conceptualization of intervention and the methods of evaluation are as critical in prevention as they are in any outcome research. Large sample sizes are often required because many disorders or problems have very low base rates, but large samples run the risk of subject attrition. The assessment must occur over a long enough time to determine the duration of effects. However, time can work against conclusive findings because children continue to develop, families and social environments change, and new risks or protective factors may develop.

SOCIAL POLICY REGARDING PREVENTION

The need for prevention in child and adolescent mental health seems obvious. In 1990 the National Advisory Mental Health Council estimated that 12 to 22 percent of this nation's children and youth needed mental health services. Estimates suggest that less than one in four receives any formal mental health intervention, and for

those who do, it is often inadequate or inappropriate.

Certain risk factors increase the incidence of disturbances and the gap between the need for service and appropriate intervention. Poverty is the single most important environmental risk factor, and the poor are less likely to receive appropriate intervention. The Bureau of the Census reported in 1994 that 23 percent of the nation's children lived below the poverty level. If current trends continue, one third of children will live in poverty by the year 2000. Severe or chronic psychopathology in parents is associated with the mental health of their children. With estimates that 20 percent of the adults in the United States have active mental disorders and markedly decreased isolation of the mentally ill, this must be considered the most important family risk factor. Children are less likely to receive appropriate intervention, and when combined with disadvantaged minority status, race creates a powerful risk. It is unrealistic to assume that the needs of these children and adolescents can be met solely by treating those with defined psychiatric disorders.

As a nation we suffer from certain simplistic notions—that prevention is much less expensive than treatment; that mental illness and behavioral disorders can be prevented through education; that family values will offset immoral, self-destructive or aggressive behaviors and dysfunction; and that personal strength and hard work will overcome social adversity. There are contrary views as well—that the problems are so severe and widespread that nothing will make a difference. Unfortunately, grandiose promises in the past have created a pervasive disbelief in preventive mental health and psychiatry that makes it difficult to sustain even those programs that have shown positive results.

In the mental health arena, stigma clearly plays a role in limiting society's commitment to prevention in mental health. Despite some pockets of resistance, prevention in other areas (e.g., cancer, heart disease, infectious disease, and drunk driving) seems to have more public support and commitment. Along with the stigma associated with the mentally ill, it is an all too prevalent view that mental disorders or maladaptive behavior results from inherent weakness, lack of will power, poor motivation, or a lack of responsibility for one's behavior.

The economics of prevention are controversial and it is often difficult to provide cost-benefit data. Agreed-upon definitions of cost and benefits and standards for comparative analyses are lacking in most cases. But even if more information existed, there would likely be controversies regarding community, state, and national priorities, and these priorities are influenced by social and political philosophies, religious and ethnic belief systems, and individual values—all of which are diverse and often controversial.

Few credible cost-benefit analyses of prevention programs are available. Credibility depends on careful design that includes all costs and relevant benefits and enough time to show that any benefits are durable. Even with these standards, evidence indicates that some preventive programs for children can be cost-effective. Despite this evidence, it is likely that prevention will continue to have lower priority than the treatment of individuals with defined mental disorders. This follows the pattern established in most states that assigns top priority for mental health treatment to those with the most severe or chronic disorders even though it is possible to produce a greater change in many of those with less severe or more acute problems.

The drive to reduce health care costs has created a situation in which the assessment of a child or adolescent with minimal symptoms and in whom a formal diagnosis is impossible must often be financed outside the provisions of a health insurance plan or a health maintenance organization, and interventions in such cases are seldom supported at all.

Several states are moving to privatize mental health services, and more seem likely to follow. Whether any preventive services at one time required in community mental health centers will survive remains to be seen. The United States' commitment to prevention in child and adolescent mental health must be considered in the context of the general needs of children and their families. The effects of welfare reform on children remain to be seen as the number living in poverty increases. There is a lack of sufficient and competent out of home care (day care) as the number of single-parent and both-parents-employed households increases and the local and national debate regarding public education continues while school buildings crumble and school bonds fail to be approved by the public. The possibility that these and other social risk factors may become even more pervasive and affect even more children and families cannot be ignored.

Social and political controversies surround many issues that affect the lives of children, and in many areas the United States does far less well by its children than other developed nations despite its wealth. Clearly, perhaps more now than before, as the proportion of elderly adults in the population increases, the nation needs national policy that truly addresses the needs of all children and their families. In the arena of child mental health and child psychiatric disorders, there is a need for balanced, comprehensive approaches including direct clinical services, universal programs to reduce risk factors, strengthen protective factors, and enhance the general welfare and adaptability of all children and targeted preventive interventions for those who do not benefit sufficiently from universal programs but as yet are not in need of direct clinical service. The burden of suffering from child psychiatric disorders is very high and its costs are enormous. Simplistic solutions are unrealistic.

However, major advances in knowledge through research can produce greater public support. Community-based coalitions of several agencies, voluntary organizations, and an apparent growing involvement by business suggest the possibility of a brighter future, but the real world risk factors of poverty, racism, violence, family disruption, and limited opportunities for successful development of all children remain major obstacles.

SUGGESTED CROSS-REFERENCES

Section 4.2 discusses sociology and psychiatry, while Section 5.1 discusses epidemiology. Section 7.8 covers psychiatric rating scales, and Section 28.4 covers genetic counseling. Family therapy is discussed in Section 30.5 and Section 48.5, and public and community psychiatry is discussed in Section 52.1. Section 49.1 covers psychiatric aspects of day care, Section 49.11 covers adolescent substance abuse, and Section 49.14 covers school consultation. Mood disorders are discussed in Chapter 14 and Chapter 45. Chapter 35 discusses learning disorders, Chapter 37 discusses communication disorders, Section 39.1 discusses attention-deficit disorders, and Chapter 40 discusses disruptive behavior disorders. The future of psychiatry is covered in Section 55.3.

SECTION REFERENCES

Albee GW, Gulotta TP, editors: *Primary Prevention Works,* vol 6. Sage, Thousand Oaks, CA, 1997.

*Ammerman RT, Hersen M: *Handbook of Prevention and Treatment with Children and Adolescents*. Wiley, New York, 1997.

Anglin TM, Nayler KE, Kaplan DW: Comprehensive school-based health care: High school students use of medical, mental health and substance abuse services. Pediatrics *97:*318, 1996.

Aronen ET, Kurkela SA: Long-term effects of an early home-based intervention. J Am Acad Child Adolesc Psychiatry *35:*1665, 1996.

Beardslee WR: Prevention and the clinical encounter. Am J Orthopsychiatry *68:*521, 1998.

Beardslee WR, Salt P, Versage EM, Gladstone TR, Wright EJ, Rothberg PC: Sustained change in parents receiving preventive intervention for families with depression. Am J Psychiatry *154:*510, 1997.

Beardslee WR, Wright EJ, Salt P, Drezner K, Gladstone TR, Versage EM, Rothberg PC: Examination of children's responses to two preventive intervention strategies over time. J Am Acad Child Adolesc Psychiatry *36:*196, 1997.

Bierman KL: Implementing a comprehensive program for the prevention of conduct problems in rural communities: The Fast Track experience. The Conduct Problems Prevention Research Group. Am J Community Psychol *25:*493, 1997.

Botvin GJ: Substance abuse prevention through life skills training. In *Preventing Childhood Disorders, Substance Abuse and Delinquency,* RD Peters, RJ McMahon, editors. Sage, Thousand Oaks, CA. 1996.

Caplan G: *The Principles of Preventive Psychiatry*. Basic Books, New York, 1964.

Chilcoat HD, Anthony JC: Impact of parent monitoring on initiation of drug use through late childhood. J Am Acad Child Adolesc Psychiatry *35:*91, 1996.

Clark GN, Hawkins W, Murphy M, Sneeber LB, Lewinsohn PM, Seeley JR: Targeted prevention of unipolar depressive disorder in an at risk sample of high school adolescents: A randomized trial of a group cognitive intervention. J Am Acad Child Adolesc Psychiatry *34:*312, 1995.

Conduct Problems Prevention Research Group: A developmental and clinical model for the prevention of conduct disorder: The Fast Track Program. Dev Psychopathol *4:*509, 1992.

*Costello EJ, Angold A, Keller GP: Adolescent outcomes of childhood disorders: The consequences of severity and impairment. J Am Acad Child Adolesc Psychiatry *38:*121, 1999.

Durlak JA: Primary prevention mental health programs for children and adolescents are effective. J Ment Health UK *7:*479, 1998.

*Durlak JA: *Successful Prevention Programs for Children and Adolescents*. Plenum, New York, 1997.

Durlak JA, Wells AM: Evaluation of indicated preventive intervention (secondary prevention) mental health programs for children and adolescents. Am J Community Psychol *26:*775, 1998.

Ellis RA: Filling the prevention gap: Multi-factor, multi-system, multi-level intervention. J Prim Prev *19:*57, 1998.

Fawcett SB, Lewis RK, Andrews A, Francisco VT, Richter KP, Williams EL, Copple B: Evaluating community coalitions for prevention of substance abuse: The case of Project Freedom. Health Educ Behav *24:*812, 1997.

Fawcett SB, Paine-Andrews A, Francisco VT, Shultz JA, Richter KP, Lewis RK, Williams EL, Harris KJ, Berkley JK, Fisher JL: Using empowerment theory in collaborative partnerships for community health and development. Am J Community Psychol *23:*667, 1995.

Gans JE, Alexander B, Chu RC, Elster AB: The cost of comprehensive preventive medical services for adolescents. Arch Pediatr Adolesc Med *149:*1226, 1995.

Goldston SE: Cost analysis and primary prevention: A sound idea whose time has come. J Ment Health UK *7:*505, 1998.

*Group for the Advancement of Psychiatry, Committee on Preventive Psychiatry: Violent behavior in children and youth: Preventive intervention from a psychiatric perspective. J Am Acad Child Adolesc Psychiatry *38:*235, 1999.

*Haggerty RJ, Sherrod LR, Garmezy N, Rutter M: *Stress, Risk and Resilience in Children and Adolescents: Stresses, Mechanisms, and Interventions*. Cambridge University Press, Cambridge, 1994.

Hardy JB, Shapiro S, Mellits ED, Skinner EA, Astone NM, Ensminger M, LaVeist T, Baumgardner RA, Starfield BH: Self-sufficiency at ages 27 to 33: Factors present between birth and 18 that predict educational attainment among children born to inner-city families. Pediatrics *99:*80, 1997.

Harrington R, Clark A: Prevention and early intervention for depression in adolescence and early adult life. Eur Arch Psychiatry Clin Neurosci *248:*32, 1998.

Harrington R, Kerfoot M, Dyer E, Gill J, Harrington V, Woodham A, Byford S: Randomized trial of home-based intervention for children who have deliberately poisoned themselves. J Am Acad Child Adolesc Psychiatry *37:*512, 1998.

Institute of Medicine: *Research Children and Adolescents with Mental, Behavioral and Developmental Disorders*. National Academy Press, Washington, DC, 1989.

Kendrick T, Tylee A, Freeling P, editors: *The Prevention of Mental Illness in Primary Care*. Cambridge University Press, Cambridge, 1996.

Kumpfer KL, Molgaard V, Spoth R: The strengthening families program for the prevention of delinquency and drug use. In *Preventing Childhood Disorders, Substance Abuse and Delinquency,* RD Peters, RJ McMahon, editors. Sage, Thousand Oaks, CA, 1996.

LaFreniere PJ, Capnano F: Preventive intervention as means of clarifying direction of effects in socialization: Anxious-withdrawn preschoolers case. Dev Psychopathol *9:*551, 1997.

MacMillan HL, MacMillan JH, Offord DR, Griffin L, MacMillan A: Primary prevention of child physical abuse and neglect: A critical review, part I. J Child Psychol Psychiatry *35:*835, 1994.

MacMillan HL, MacMillan JH, Offord DR, Griffin L, MacMillan A: Primary prevention of child sexual abuse: A critical review, part II. J Child Psychol Psychiatry *35:*857, 1994.

McCroskey J, Meezan W: Family-centered services: Approach and effectiveness. Future Child *8:*54, 1998.

*Mrazek PJ, Haggerty RJ, editors: *Reducing Risks for Mental Disorders: Frontiers for Preventive Intervention Research*. National Academy Press, Washington, DC, 1994.

Offer DR, Kraemer HC, Kuzdiu AE, Jensen PS, Harrington R: Lowering the burden of suffering from child psychiatric disorders: Trade-offs among clinical, targeted, and universal interventions. J Am Acad Child Adolesc Psychiatry *37:*686, 1998.

Offord DR, Bennett KJ: Conduct disorder: Long-term outcomes and intervention effectiveness. J Am Acad Child Adolesc Psychiatry *33:*1069, 1994.

Olds DL, Eckenrode J, Henderson CR, Kitzeran H, Powers J, Cole R, Gidora K, Morris P, Pettit LM, Larkey D: Long-term follow-up of home visitation on maternal life course and child abuse and neglect. JAMA *278:*673, 1997.

Pizzolongo PJ: The comprehensive child development program and other early intervention program models. In *Preventing Childhood Disorders, Substance Abuse, and Delinquency,* RD Peters, RJ McMahon, editors. Sage, Thousand Oaks, CA, 1996.

Reddy LA, Pfeiffer SI: Effectiveness of treatment foster care with children and adolescents: A review of outcome studies. J Am Acad Child Adolesc Psychiatry *36:*581, 1997.

Rispens J, Aleman A, Gondena PP: Prevention of child sexual abuse victimization: A meta-analysis of school programs. Child Abuse Negl *21:*975, 1997.

Tremblay RE, Pagan-Kurtz L, Masse LC, Vitaro F, Pihl RO: A bimodal preventive intervention for disruptive kindergarten boys: Its impact through mid-adolescence. J Consult Clin Psychol *63:*560, 1995.

*US Department of Health and Human Services, Public Health Services, Office of Substance Abuse Prevention: *Prevention of Mental Disorders, Alcohol and Other Drug Use in Children and Adolescents*. DHHS publ no. (ADM): 89–1646, Washington, DC, 1989.

Willis DJ, Holden EW, Rosenberg ML, editors: *Prevention of Child Maltreatment: Developmental and Ecological Perspectives*. Wiley, New York, 1992.

Zeanah CH, Boris NW, Larrieum JA: Infant development and developmental risk: A review of the past 10 years. J Am Acad Child Adolesc. Psychiatry *36:*165, 1997.

CHAPTER 50. ADULTHOOD

CALVIN A. COLARUSSO, M.D.

As they leave childhood and adolescence behind and journey through young adulthood and midlife, men and women attempt to achieve the following:

1. Separate psychologically from the parents of childhood and achieve self-sufficiency in the adult world
2. Find a gratifying place in the world of work
3. Experience sexual and emotional intimacy within a committed relationship
4. Become a parent
5. Accept the aging process in the body
6. Integrate the growing awareness of time limitation and personal death
7. Maintain physical and emotional intimacy in the face of the powerful physical, psychological and environmental pressures of midlife
8. Facilitate the emergence of children into adulthood
9. Develop and sustain friendships with individuals of different ages and backgrounds
10. Continue to play
11. Leave a legacy for future generations by facilitating the development of younger individuals

LEGACY OF THE PAST

There is little new under the sun when the subject is man himself, so it is not suprising that ancient civilizations contained rather sophisticated ideas about the life course. In the sixth century BC, Solon described six periods of adulthood and assigned appropriate developmental tasks to each. In the following century Plato described the educated man as one who used enlightened reason to contemplate the realities of human life and the reasons for his existence. The developmental flowsheet of Confucius described how at 15 he set his heart upon learning, by 30 he planted his feet firmly on the ground, by 40 he no longer suffered from perplexities, at 50 he knew the biddings of heaven, and at 60 he allowed himself to follow the dictates of his heart. Concerns with time, age, and change pervade ancient Japanese thought as well. There adulthood was viewed as a time of life requiring effort and attention as the individual strove to become freer of the self to gain the highly respected wisdom of old age.

The ancient writings describe a chronological life cycle, the adult in a constant state of dynamic change and flux, a continual need to define the adult self, a preoccupation with time as an expression of awareness of a limited life span and individual mortality, and a struggle between love of self and responsibility to the society in which one lives.

PIONEER ADULT DEVELOPMENTALISTS

The scientific study of adulthood is essentially a phenomenon of the twentieth century, stimulated by the writings of four men (Table 50-1). In 1908 Arnold Van Gennep (1873–1957), a cultural anthropologist, described the importance and meaning of the universal rituals that surround such events as pregnancy, childbirth, menarche, betrothal, and death. The rituals are ceremonies whose essential purpose is to enable individuals to pass from one stage in life to another.

Pioneers	Contemporaries
Erik Erikson (1902–1994)	Calvin Colarusso
Sigmund Freud (1856–1957)	Daniel Levinson
Carl Jung (1875–1961)	Robert Nemiroff
Arnold Van Gennep (1873–1957)	Bernice Neugarten
	George Vaillant

Table 50-1 Developmental Theoreticians

Sigmund Freud (1856–1937) formulated the first modern theory of child development. Intrinsic to the predictable developmental progressions were basic transformations, regressions, and advances. Essential to Freud's theory of development is the idea that the mind changes constantly as a result of the ongoing interaction among mind, body, and environment.

The pioneer who focused primarily on the adult years was Carl Gustav Jung (1875–1961). He described a continuing process of physical and psychological separation from parents during the 20s and 30s, and viewed age 40 as a time of significant psychological change, growth, and transition. The archetypes “puer” (young) and “senex” (old) that Jung saw as fundamental polarities stimulated modern adult developmentalists.

The fourth major pioneer, Erik Erikson (1902–1994) provided the first integrated psychosocial view of individual development through eight stages from birth to death. For him, development lies not in stability per se, but in the changes necessary to transcend successive contradictions, conflicts, and states of disequilibrium. Current-day developmentalists owe a profound debt to the work of Erikson.

Erikson's theory has recently been used to examine the developmental and sexual issues facing seropositive gay males with particular reference to the polarities of identity versus role confusion and intimacy versus isolation. Other investigators, studying the adult development of women, have found that women do not conform to the sequencing aspect of Erikson's stage model because in our society they tend to be more relationship oriented. Thus intimacy may occur as a dominant theme earlier in development. In addition, identity and intimacy are conceptualized as concurrent processes for both men and women. The unfolding of one leads to the further delineation of the other and affects the related polarities of self and other, masculine and feminine, and agency and communion.

CONTEMPORARY ADULT DEVELOPMENTALISTS

A comprehensive understanding of adult development is beginning to emerge from studies conducted over the past 30 years. The study of childhood preceded the study of adulthood because of social and psychological factors such as the spread of compulsory education and Freud's discovery of the influence of childhood experiences on adult psychopathology. The recent shift of interest to the adult years builds on this knowledge and may be due to the increased life span and the need to understand and accommodate the rapidly growing number of middle-aged and elderly individuals. Among the most influential current theoreticians are the

following.

Daniel Levinson and his associates devised a psychosocial theory of male development that proposes a life cycle consisting of distinctive, identifiable eras of approximately 20 years extending from birth to death. Within these eras are alternating periods of 6 to 7 years of stability followed by 4- to 5-year intervals of transition, each with its own tasks to be mastered. These concepts are clinically useful since many patients, particularly the healthier ones, tend to present themselves during periods of transition when internal and external conflicts are increased.

Levinson's theory is currently being used as a framework for the study of various groups. Application of his concepts of developmental periods to the existing data on older gay men indicates that the stereotype of lonely, depressed, sexually frustrated, aging gay men may not be a valid picture of the cohort.

Gender-oriented research has begun to question whether adult transitions are the same for men and women. Levinson's alternating intervals of stable structure and transition have been used to delineate the developmental and transitional periods experienced by Army officers from preadulthood to midlife. The study examined the impact of living and working within a highly ordered, systematic organizational structure such as the military on the way in which these adult lives unfold. Levinson's outline of adult phases was found very useful in understanding intellectual and emotional life within a career military setting. The impact of regimentation and a clearly delineated future for young adulthood and midlife tended to mute the distinctions between periods of transition and stable structure. The effect of a military career on mid and late adulthood postmilitary life remains a fertile field for investigation.

The Harvard Grant Study has followed longitudinally the life course of 268 undergraduate students from 1939 to the present. The current director, George Vaillant, has used these data to study adaptation in adulthood, particularly ego mechanisms of defense. Illustrating the developmental nature of intrapsychic processes in the adult, Vaillant demonstrated that the psychologically healthier members of the research sample used mature defenses such as suppression, altruism, sublimation, anticipation, and humor more frequently in midlife than they had in late adolescence and young adulthood, thus illustrating the effect of ongoing, dynamic developmental processes. In agreement with all other major dynamic theories, Vaillant found conflict to be an integral, inescapable aspect of normal development.

More recently, since the subjects of the Grant study are now in the developmental phase of late adulthood, Vaillant has used a modified version of Erikson's stage model to organize the process of psychosocial maturation in the latter third of life, focusing on the Eriksonian tasks of generativity and integrity by rating longitudinal case histories of the Grant Study subjects. Success or failure to engage and master these developmental themes was closely correlated with patterns established earlier in development in this select group of men who were originally chosen for their emotional and physical health and have now been followed for more than half a century (ages 18 to 65).

Through the study of adults in nonclinical settings Bernice Neugarten and her colleagues have emphasized the psychological importance of an increased awareness of aging and the personalization of death, as expressed in body monitoring and a tendency to view time in terms of time left to live, rather than time since birth. Middle-aged adults develop a sense of competence that was unrealized earlier in life and have a unique perspective on the younger and older generations. As middle age progresses, people become more introspective and develop an increased sense of interiority.

On the basis of their experience as clinicians and psychoanalysts, Calvin Colarusso and Robert Nemiroff propose a broad theoretical foundation for adult development (Table 50-2) by suggesting that the developmental process is basically the same in the adult as in the child, because, like the child, the adult is always in the midst of an ongoing dynamic process, continually influenced by a constantly changing environment, body, and mind. Whereas child development focuses primarily on the formation of psychic structure, adult development is concerned with the continuing evolution of existing psychic structure and with its use. Although the fundamental issues of childhood continue in altered form as central aspects of adult life, attempts to explain all adult behavior and pathology in terms of the experiences of childhood is considered reductionistic. The adult past must be taken into account in understanding adult behavior in the same way that the childhood past is considered. The aging body is understood to have a profound influence on psychological development in adulthood, as is the growing midlife recognition and acceptance of the finiteness of time and the inevitability of personal death.

Development is a life-long, dynamic process that is basically the same in childhood and adulthood
Basic themes from childhood continue to affect psychic development in adulthood, but adult functioning and symptomatology are an amalgam of both childhood and adult experiences
Recognition and acceptance of the finiteness of time and the inevitability of personal death are major psychic organizers in adulthood, both in the promotion of normal development and the formation of symptomatology

Table 50-2 Hypotheses About Development in Adulthood

Although Colarusso and Nemiroff, following Erikson, divide adulthood into 20-year periods, developmental phases are not considered to be as useful in adulthood as they were in childhood because developmental themes do not appear with the chronological precision or the phase specificity that they have in childhood. For example, the developmental milestones of infancy appear within a span of a few weeks or months. The phases of Margaret Mahler's separation individuation theory and Jean Piaget's cognitive developmental sequences cover months or a few years at most. By contrast, the experience of biological fatherhood can occur over a span of 60 years or more, and the intrapsychic effects vary markedly, depending on the phase of development in which parenthood occurs. A more precise understanding of adult experience may be achieved by using the concept of “adult developmental tasks.” They are the psychological responses to such major life experiences as work achievement, parenthood, and retirement, which, as the result of actual occurrences or psychological considerations, produce intrapsychic changes in all persons in a particular age group. Used together, the concepts of adult developmental phases and tasks can deepen the clinician's understanding of the effect of past and present experiences on the evolving mind.

YOUNG ADULTHOOD (AGES 20 TO 40)

Shift From Biological Growth to Aging The shift from physical progression to retrogression, from growth to aging, occurs in young adulthood. This strikingly obvious and developmentally significant event has scarcely been addressed in the literature on adulthood. The effect of biological maturation on mental evolution has been extensively described in the developmental theory of childhood. For instance, Freud described the libidinal progression from erogenous zone to erogenous zone and built his conceptualization of oral, anal, oedipal, latency, and adolescent stages upon this biologically determined sequence.

There is no interval between the end of maturational progression and the beginning of aging. The two overlap during the young adult years, but aging gradually replaces growth as the dominant biological influence. Evidence of this shift includes slowing of reflexes, loss of skin tone, early signs of balding (in some during the 20s), and the far more dramatic loss of the procreative function in women in the late 30s and early 40s.

Transition From Adolescence to Young Adulthood The transition from adolescence to young adulthood is characterized by real and intrapsychic separation from the family of origin and the engagement of new, phase-specific tasks (Table 50-3). Various authors have described this process as the shedding of family dependencies to become a member of society at large, the shift from preadolescent idealized parental images to postadolescent idealized ethics and values, and the gradual shift from the family of origin to the family of procreation.

To develop a young-adult sense of self and other: the third individuation
To develop adult friendships
To develop the capacity for intimacy; to become a spouse
To become a biological and psychological parent
To develop a relationship of mutuality and equality with parents while facilitating their midlife development
To establish an adult work identity
To develop adult forms of play
To integrate new attitudes toward time

Table 50-3 The Developmental Tasks of Young Adulthood

The Berkeley Guidance and the Oakland Growth Studies are long-term investigations of normal development that were begun in 1929 and 1932, respectively. They indicate that psychological adjustment in adulthood is not easily predicted from adjustment during adolescence. This important observation supports the idea that conflicts in adulthood can lead to the reworking of past issues with better resolutions or to the onset of new forms of pathology.

The same idea is incorporated in Levinson's characterization of transitions, intervals of approximately 5 years of change and turmoil that occur between periods of relative stability. The transition from late adolescence to young adulthood occurs between ages 17 and 22. During these years the individual resolves the issue of childhood dependency enough to establish self-reliance and begins to formulate new, young-adult goals that eventually result in creation of new life structures that promote stability and continuity.

Another important intrapsychic aspect of the transition is a redefinition of the childhood and adolescent past. For the first time, an entire phase of life is consigned to the past. This gradual, painful process brings closure to an epoch of life and forces redefinition of aspects of psychic structure, particularly the superego and ego ideal.

Developmental Tasks of Young Adulthood

Developing a Young-Adult Sense of Self and Other: The Third Individuation Establishing a self separate from parents is a major task of young adulthood. For most individuals the emotional detachment from parents that takes place in adolescence and young adulthood is followed by a new inner definition of themselves as comfortably alone and competent, able to care for themselves in the real and intrapsychic worlds. This shift away from the parents continues long after marriage and parenthood result in the formation of new relationships that replace the progenitors as the most important individuals in the young adult's life.

Psychological separation from the parents is followed by synthesis of mental representations from the childhood past and the young adult present. For instance, as their children grow, young parents reengage memories of their own childhoods and fuse their experiences as parents with memories of their own progenitors from a generation ago. The separation-individuation process in infancy is responsible for establishing a stable sense of self and the capacity to relate to others. The psychological separation from parents in adolescence has been called “the second individuation” and the continued elaboration of these themes in young adulthood, “the third individuation.” The continuous process of elaboration of self and differentiation from other that occurs in the developmental phases of young (20 to 40) and middle (40 to 60) adulthood is influenced by all important adult relationships. At its core are ties to children, spouse, and parents (i.e., the family), the same psychological constellation that shaped the first and second individuations.

Adult developmental theory postulates a growing complexity in relationships as the individual moves from developmental phase to developmental phase. The first individuation is a rather exclusive affair between infant, mother, and father. During the second individuation relationships expand to include important nonfamilial relationships such as friends and mentors. The transition from the second to the third individuation is a young-adult experience, stimulated by growing intrapsychic separation from the parents and the loneliness that follows in its wake. Sooner or later most young adults attempt to fill the real and intrapsychic voids left by separation from the parents of childhood by establishing a family of procreation. Other aspects of the third individuation deal with the developmental tasks of intimacy, parenthood, and separating from parents.

Developing Adult Friendships Freud described friendship as an expression of libido, stemming from the same source as sexual love that has sexual union as its aim. In the relationship between the sexes, the impulses force their way toward sexual union, but in other circumstances they are diverted from this aim or are prevented from reaching it though always preserving enough of their original nature to keep their identity recognizable. But human relationships, including friendships, are not based on love alone. Like all other human interactions, friendships are also based on aggression. The character of friendship is determined by the aim-inhibited expression of the aggression, not by its absence.

Using a psychodynamic framework, friendship may be defined as an extrafamilial object relationship based on mutuality, equality, and freedom of choice, in which the expression of sexual and aggressive impulses is predominantly aim inhibited. However, because of the power of emotion, friends can be transformed into lovers or enemies and sometimes back again into friends. This tendency toward fluidity is common in childhood, whereas the maintenance of a stable state of friendship over an extended period of time is more characteristic of adulthood.

From the latency period onward, friendships are an integral part of human experience, a vital form of relationships. At each subsequent developmental phase, including adolescence and young adulthood, the character and substance of healthy friendships are determined in part by the mutual need to engage and resolve major phase-specific developmental tasks.

At times developmental pressures strain the capacity for friendship to the limit. In early adolescence, in particular, the strength of the drives vis-a-vis the ego often breaks down aim inhibition, as observed in both homosexual and heterosexual experimentation between peers. But, generally speaking, at no other time in life do friendships play such a prominent role in the developmental process. At this time, they facilitate the engagement and resolution of such developmental tasks of the phase as separation from parents and beginning integration of adult sexual and work identities.

In late adolescence and young adulthood, before marriage and parenthood, friendships are often the primary source of emotional sustenance. In the years between the family of origin and the family of procreation, the young adult has little opportunity to gratify impulses within a committed relationship and experiences the loneliness of young adulthood. Roomates, apartment mates, sorority sisters, and fraternity brothers, as indicated by the names used to describe them, are substitutes for parents and siblings, temporary stand-ins until more permanent replacements are found.

The emotional needs for closeness and confidentiality are largely met by friendships. All major developmental issues are discussed with friends, particularly those in similar circumstances. As marriages occur and children are born the central emotional importance of friendships diminishes. Some friendships are abandoned at this point because the spouse objects to the friend, recognizing at some level that they are competitors. Gradually there is movement toward a new form of friendship, couples friendships. They reflect the newly committed status but are more difficult to form and maintain because four individuals must be compatible, not just two.

As children begin to move out of the family into the community, parents follow. Dance classes and Little League games provide the progenitors with a new focus and the opportunity to make friends with others who are at the same point developmentally and receptive to the formation of relationships that help explain, and cushion, the pressures of young-adult life.

Developing the Capacity for Intimacy: Becoming a Spouse Erikson defined the major developmental dichotomy of young adulthood as intimacy versus stagnation. Intimacy includes the capacity to experience others' needs and concerns as being as important as one's own. The ability to be intimate has its origins in the early parent-child relationships, the successful resolution of the oedipal complex, and adolescent sexual experimentation. But it does not become a sustainable capacity until young adulthood, apparent in the highest functioning individuals by the fourth post–high school year.

The developmental shift from sexual experimentation to the desire for intimacy is experienced in young adulthood as an intense loneliness, resulting from the

awareness of an absence of committed love similar to that experienced in childhood with the parents. Brief sexual encounters in short-lived relationships no longer significantly boost self-esteem. Having mastered the mechanics of capable sexual performance, mere repetition no longer provides emotional satisfaction. Increasingly, the desire is for emotional involvement in a sexual context. The young adult who fails to develop the capacity for intimate relationships runs the risk of living in isolation and self-absorption in midlife.

Significant intrapsychic change occurs when intimacy is achieved. Through the repeated fusion of sex and love, the self is increasingly linked to the partner. As sexual thoughts, feelings, and practices are repeated in relation to the loved one the superego becomes more tolerant and flexible. Sexual identity is refined as feminine or masculine aspects of the self are projected onto, and accepted and loved in, the partner. The ego ideal is altered by inclusion of the partner's goals for the couple's future, particularly in regard to such young-adult issues as the desire for children and career ambitions.

The development of the capacity for intimacy encourages the acceptance of the equal status and complementary nature of the male and female genitals. Repeated experiences with foreplay, intercourse, conception, pregnancy, childbirth, and psychological parenthood provide the optimal environment in which to abandon infantile notions of phallic superiority and replace them with the recognition that female and male genitalia are equal and interdependent for sexual pleasure, intimacy, and reproduction.

For most individuals in Western culture, the experience of intimacy increases the desire for marriage. Longitudinal developmental studies indicate that no single variable predicts mental health as clearly as the capacity to remain happily married over time.

Becoming a Biological and Psychological Parent Although biological parenthood is not limited to young adulthood, it is experienced by most individuals for the first time during this phase of development. Consideration of the developmental line of sexual identity during childhood will place parenthood in a developmental context. During the first 18 months of life a basic sense of maleness or femaleness called “the core gender identity” is established. By age 3, children become aware that there are two sexes; then during the oedipal phase, sexuality is explored within relationships. During latency subjective family, community, and cultural attitudes about masculinity and femininity are integrated. During adolescence the maturational changes of puberty stimulate interest in the use of the adult body as a sexual instrument, first with the self, then with others. Then in young adulthood the developmental experiences of intimacy and parenthood are added.

Biological parenthood initiates the process of psychological parenthood, the mental state in which healthy young adults become increasingly attached to and involved with their children. For both parents-to-be, pregnancy adds a new dimension to sexual identity by confirming that their sexual apparatus can perform the primary function for which they were intended. After birth, each interaction with the infant enhances the heightened sense of sexual completeness and stimulates the desire to lovingly engage the baby who is so strongly identified with the self. When a young couple become parents for the first time, a family is created. Its structure is identical to the family of origin except for the reversal of roles. As the former children, now the parents, minister to their creation, they undergo profound intrapsychic change as a result of the simultaneous reexamination of their own experiences as children and the growing sense of themselves as adults.

Parenthood intensifies the relationship between the new parents. Through their physical and emotional union the couple has produced a fragile, dependent being who needs them in the interlocking roles of father and mother. This recognition expands their internal images of each other to include thoughts and feelings emanating from the role of parent. The superego and ego ideal are expanded to include the role of parenthood, and as they live together as a family, the lovers' relationship to each other changes. They become parents relating to one another and to their children.

Becoming a parent also stimulates further individuation from the members of the family of origin. Assuming the roles that were formerly the exclusive prerogative of their progenitors, the new parents develop parity. Further, constant conscious and unconscious comparisons of child-rearing practices heighten an internal sense of difference while paradoxically reinforcing connectedness and continuity between the generations.

Finally, the ability to be instrumental in the midlife development of one's parents by providing them with the experience of grandparenthood is striking evidence of the shift in the power balance between the generations, foreshadowing the not-too-distant future when the elderly parents may become dependent upon their children for physical and psychological well-being.

Separating Psychologically From Parents and Developing a Relationship With Them of Mutuality and Equality While Facilitating Their Midlife Development Real and intrapsychic relationships with parents undergo dramatic changes during young adulthood. This experience may be divided into three phases that build upon one another.

PHASE I Psychological separation from parents continues. Not all aspects of the first and second individuations are resolved by the end of adolescence. Both, but particularly the more-recent second individuation, continue to be engaged during young adulthood. The ability to function relatively independently of parents, without using them as a major source of comfort, security, and direction, comes to fruition during the young-adult years.

Building on the developmental achievements of adolescence, this process is greatly facilitated by marriage and parenthood. As the spouse is internalized as a significant partner and children assume roles of increasing intrapsychic importance, the power of earlier parental introjects is gradually diminished. These childhood memories and images are further transformed by their fusion with current representations of the spouse and the self in conjunction with the functions of both as lover and parent. Every young adult brings from childhood a detailed plan (largely unconscious, but firmly institutionalized in the superego/ego ideal) of how a husband or wife and parent should act. These idealized expectations must be modified by current reality if they are to be adaptive.

PHASE II Once the roles of spouse-parent-provider have been assumed, the stage is set for the establishment of an inner sense of equality and mutuality with parents. This occurs as young adults marry, become parents, work, develop adult friendships, and become part of a community. As these adult experiences become the substance of everyday life they transform the intrapsychic relationship with the parents of childhood from one of dependency and need to one of mutuality and equality.

However, parents remain important as long as they live, since only parents and children place one in the center of a genetic continuity that spans three generations. As middle age approaches with its preoccupation with time limitation and personal death, the intrapsychic importance of this genetic immortality increases.

PHASE III Equality and mutuality with parents may continue for many years or may be short-lived, depending on the mental and physical health of the parents. At some point the adult child will be confronted with the psychological and possibly physical task of caring for vulnerable, dependent parents who no longer function independently. When this occurs, the memories from childhood of the dependent, immature self and the aging parent of the present stimulate the child to reenter the parent-child dyad, now reversed, and assume the role of caretaker.

Simultaneously, through their interactions, living parents and grandparents provide examples of how the developmental tasks of middle and late adulthood may be engaged. As the young adult parents and their children internalize these examples the foundation is laid for interactions in the years to come when their roles will be reversed.

Longitudinal studies that have traced personality development through childhood and into adulthood indicate that the adult personalities of children closely resembled those of one or both parents. This similarity may be attributed not only to the earlier adult-child interaction but also to their continuing relationships as adult to adult.

Both parents have an effect, not just the parent of the same sex. In fact, identification with the parent of the opposite sex tends to increase in young adulthood because heterosexual identity is more secure, strengthened by marriage and psychological parenthood. The young adult is then freer to incorporate valued aspects of the character of the opposite-sex parent, particularly in regard to the engagement of the developmental tasks of middle and late adulthood.

Establishing an Adult Work Identity Establishing a work identity is a critical developmental task of young adulthood. When this developmental line is relatively conflict free, there is a smooth progression from high school to on-the-job training or from high school to college and graduate school. In most advanced countries opportunities to change jobs or acquire further training to enhance career advancement remain open throughout the young-adult years. In more closed or class-oriented societies, such as China or the United Kingdom, decisions about career choice are made in adolescence on the basis of intellectual potential and academic achievement. Once excluded from the educational track the individual has little subsequent opportunity for formal academic advancement.

The transition from learning and play to work may be gradual or abrupt, but at some point, usually in the late teens or 20s, work becomes a central intrapsychic and real activity. Then, the pleasures of play or learning or both must be subordinated to the temporal and emotional demands of job or career. Depending upon choice of

career and opportunity, work may become a source of ongoing frustration or an activity that enhances self-esteem and gradually leads to a shift in identity from child to adult and from player to worker.

> A young-adult female patient had greatly enjoyed her 5 years in college and only reluctantly accepted a job with a large real estate firm. During college she had had limited interest in her appearance, and she began work in clothing borrowed from family and friends. She scoffed when her boss began to criticize her dress and gave her an advance to buy an upscale wardrobe; but she began to enjoy the fine clothing and the respect engendered by her appearance and position. As her income began to rise, work became a source of pleasure and self-esteem and the way to acquire some of the trappings of adulthood.

Young adults, regardless of career choice, must assume the role of mentee and work with mentors to acquire the information and skills necessary to forge a work identity. The relationship with the mentor is based on infantile identifications, but this adult process is not a duplication of the parent-child experience. The relationship with the mentor normally goes through three phases. The first phase, a psychological fusion with the mentor, is followed by an internalization of aspects of the mentor's ideas and attitudes. Eventually, psychological and physical separation from the mentor leads to further individuation.

Clinicians who recognize the importance of this critical developmental process will be extremely sensitive to the presence (or absence) of material about work and relationships. In some patients, problems with work may be a major symptom, rooted in significant internal conflict; in others work may be the unrecognized source of interference with other developmental themes.

> A 28-year-old lawyer who worked 80 hours a week did not recognize that his social isolation was of his own making. Chronic fatigue and lack of time were the rationalizations he used to avoid considering more deeply seated needs to avoid sex and intimacy.

Developing Adult Forms of Play Young adulthood is when the ability required to perform most of the physical games learned in childhood and adolescence reaches a peak and then declines. Most individuals do not realize their childhood fantasies about athletic fame and fortune and must realistically appraise their capability and mourn what was not achieved. Aspirations for athletic fame are not easily relinquished, however. Some are maintained through identification with sports figures who are now chronological contemporaries. These identifications may continue into middle and late adulthood—these are the true fans. For others, however, new interests and acceptance of the aging process stretch the credibility of comparisons between youthful heroes and the middle-aged self until such identifications become impossible, which leads to a diminished interest in spectator sports and a search for new forms of play.

Some patterns of childhood play continue into the adult years, existing side by side with new forms of play that reflect the growing importance of spouse and children. These range from tennis or jogging with one's partner to mental games, such as bridge, with friends. Involvement in the play activities of children is a powerful source of narcissistic gratification for some parents, providing a second chance to realize childhood goals through identification with the success of sons or daughters. In the psychotherapeutic setting, material about childhood play is a rich source of information, whose exploration may lead to unexpected revelations and insight. The exploration of adult play activities may have similar benefits. For example, the failure to moderate physical activity in keeping with the aging process in the body may result in physical injury that precipitates anxiety or depression and exposes conflicts about aging and other young-adult issues.

Integrating New Attitudes Toward Time

TIME SENSE IN YOUNG ADULTHOOD Time sense in the 20s rests on psychic structures built in childhood and adolescence. In the early 20s, consignment of childhood to the past can produce a brief, latency-like temporal calm that is built upon both the resonance between conscious aspirations and unconscious expectations and the realization that the future is long enough to postpone some decisions, undo mistakes in career or object choice, and start again if necessary. But by the mid-20s, time sense is increasingly influenced by the search for adult structure and new objects—career definitions, friends, lovers, and spouse and children—to replace the temporal organizers of childhood, mother and father.

Subjective time sense in the 30s differs qualitatively from that in the 20s because midlife issues emerge and gradually become dominant. For example, as signs of physical aging become more apparent, the phenomenon of loss is translated into a growing awareness of personal time limitation. This is particularly true for women, who must confront the approaching loss of procreative function, which affects subjective time sense in as dramatic a manner as the onset of puberty did in early adolescence. The “30-something” individual must also assess relationships and work achievement differently. Significant progress toward establishing the family of procreation to replace the family of origin is expected, as is work and financial achievement.

RELATIONSHIPS Relationships significantly influence subjecting time sense in young adulthood. The relationship to the spouse is one of the most important intrapsychic organizers of time sense in adulthood. This temporal commitment is made in the most demanding terms, and like no other mutually consensual relationship in life, marriage defines the expectation of how time will be used as well as conceptualized consciously and unconsciously. Over time, a long-lasting, mature love relationship becomes as significant an organizer of time sense in adulthood as the relationship between parent and child was in childhood.

However, becoming a biological parent is the quintessential temporal experience of young adulthood. The ability to reproduce provides humans with their only immortality; through genetic transmission a part of the self will live on after death. Biological parenthood coupled with the capacity to love and remain invested in one's offspring leads to major psychological and more specifically temporal reorganization. Future sense is expanded, and the past is extensively reworked as the parent consciously and unconsciously relives each developmental stage with the child.

NORMAL TEMPORAL FUNCTIONING AT THE END OF YOUNG ADULTHOOD In temporal terms, normal functioning implies that as individuals approach 40, they have accepted the aging process in the body, integrated the use of time with productivity, separated psychologically from Mother and Father time, begun to deal with the notion of personal death, and forged new ties with contemporaries and children who give new meaning to the present and future. So equipped, the young adult approaches middle age ready to engage, master, and integrate the monumental temporal challenges that lie just ahead.

One of the most profound influences on development in adulthood is the increasingly conflictual awareness that personal time is limited and one will die. This painful realization is both a source of conflict and psychopathology and a developmental stimulus that forces a new appreciation of the preciousness of time and a reordering of priorities.

Coming to terms with time limitation is a central psychological task of middle age. Implicit in this statement is the conviction that temporal awareness and management in midlife has a greater impact on development, healthy and pathological, than at any other point in the life cycle.

Changes in Psychic Structure in Young Adulthood In young adulthood the ego increases its ability to control impulses consistently, delay gratification, limit and control aggression, channel energy into work and other sublimated activities, and separate from parents and form new loving relationships.

Defenses The longitudinal Grant Study of a population of Harvard undergraduates determined that these men were twice as likely to use the mature defenses of sublimation, suppression, anticipation, and humor in young adulthood as they were in adolescence. In middle adulthood the change was even more dramatic; mature defenses were used four times more frequently than in adolescence. In addition, by midlife the subjects used relatively fewer immature defense mechanisms (i.e., acting out, projection, hypochondriasis, and masochism). A defensive shift toward optimism and idealism has also been observed in young adults, reflecting the need to cope with growing preoccupations with two fundamental features of human life, inevitable death and the existence of hate and destructive wishes inside each person.

Intelligence and Cognition Studies in the 1940s proposed that the intelligence quotient (I.Q.) increased into the 20s and then declined throughout the remainder of life. This notion has persisted despite detailed longitudinal studies that convincingly demonstrate that in the absence of a cognitive disorder, intelligence increases throughout young and middle adulthood.

This conclusion is confirmed by new post-Piagetian ideas about intelligence in adulthood. Like Freud, Piaget ended his theory with adolescence, describing attainment of the capacity for abstraction as the highest order of thinking. More-recent studies indicate that the nature of thought in adulthood is far more complex. Piaget's constructs have been extended to include a stage beyond formal operations, called “structured analytic thought.” Going beyond simple abstraction, the young-adult thinker begins to grasp the connectedness of sets of relationships, that is, the relationship between systems.

Progressing beyond formal operations, some adults develop the capacity for dialectical thinking. While formal thinkers focus on universals, dialectical thinkers are

relativists, recognizing the relation between ideas and facts that creates “truth” within a particular system of thought or historical period. Contradiction and paradox are integral aspects of adult thinking that lead to new ideas and synthesis.

Some young adults and middle-aged adults may progress beyond formal operations and dialectical thought to autonomous thought. Whereas the formal operational thinker analyzes relations within a closed system and the dialectical thinker understands that some truths are relevant, the autonomous thinker integrates logical and irrational aspects of experience and recognizes how the thinker participates in the creation of truth. Other studies support the notion that thinking becomes increasingly relative in adulthood because of the newly emergent ability to synthesize experience and emotional and contradictory ideas. Biologically intact individuals can continue this developmental evolution in thinking into late adulthood.

REALITY TESTING The ability to test and comprehend reality grows in complexity in young adulthood for a number of reasons. Greater comfort with the inner world lowers the tendency to distort internal and external stimuli. Greater experience in living increases the capacity to manage everyday events autonomously and with less stress. The ability to think and reason increases qualitatively; phase-specific developmental tasks such as the growing awareness of time limitation and personal death are engaged.

For many, the continued elaboration of these ego functions during young and middle adulthood results in an enhanced capacity for introspection and the emergence of wisdom, that special ability to understand and interpret human experience from the viewpoint of the second half of life.

Female Development A radical change in the theory of female development was crystalized in 1976 when the *Journal of the American Psychoanalytic Association* devoted an entire supplement to female psychology. Female development was conceptualized as following its own line from birth onward, which differed from Freud's proposal that male and female psychic development was identical until the oedipal phase. Further, penis envy was no longer recognized as a major component of female development during the oedipal phase. The supplement was also noteworthy in extending the consideration of female development into adulthood.

Aspirations for a career and motherhood are particularly pressing issues for women in young adulthood. The absence of a single, socially prescribed role for them provides opportunities for both growth and considerable inner turmoil. Therapists are treating increased numbers of disillusioned young-adult women—and men—who concentrated on their careers and did not marry and have children. Accomplished in the workplace, they complain of loneliness, depression, and isolation.

The loss of the ability to have children in their 30s and early 40s is the most striking developmental difference between the sexes in young adulthood, and it affects almost every other developmental line.

> A childless 38-year-old, divorced lawyer sought treatment for herself and her live-in boyfriend of 2 years. She was anxious to marry and have children, but he was hesitant. She had become increasingly depressed over the past several years, as her awareness that the passage of time was diminishing her chances to have children deepened. After several months of therapy the boyfriend decided to leave the relationship.

Developmental theory suggests that psychopathology results from the failure to engage and master major developmental tasks during the phase in which they are central themes. This remains true throughout the life cycle. An awareness of the basic biological and psychological forces that underlie normal young-adult development can provide individuals and clinicians with the knowledge needed to prevent such young-adult psychopathology.

MIDDLE ADULTHOOD (AGES 40 TO 60)

Concepts of time, aging, and death are the currency of midlife development. Consequently, therapeutic attention can no longer focus entirely on the past. It must also include current reality and the struggle to adapt to ever-changing time, since these issues are at the core of understanding and treating individuals between the ages of 40 and 60. Despite growing awareness of aging and death, the middle years are the best time in life for many. Middle adulthood is the golden age of adulthood, similar to the latency years in childhood but much longer. Physical health, emotional maturity, competence and power in the work situation, and gratifying relationships with spouse, children, parents, friends, and colleagues all contribute to a normative sense of satisfaction and well-being.

Transition From Young to Middle Adulthood The transition from young adulthood to middle adulthood is slow and gradual, with no sharp physical or psychological demarcation. The aging process picks up speed and becomes a powerful organizing influence on intrapsychic life, but the change is gradual, unlike adolescence. Mental change is experienced in a similar fashion, slow and imperceptible, without a sense of disruption.

Development in young adulthood is embedded in close relationships. Intimacy, love, and commitment are related to the mastery of the relationships most immediate to personal experience. The transition from young adulthood to middle age includes widening concern for the larger social system and differentiation of one's own social, political, and historical system from others. Authors have described middle adulthood in terms of generativity, self-actualization, and wisdom.

Developmental Tasks

Integrating the Potential for Attachment and Loss: The Fourth Individuation The fourth individuation refers to the elaboration of separation-individuation processes in middle adulthood (ages 40 to 60). One of the most powerful influences on these processes in midlife is the ironic awareness that one will die and be deprived of involvement with loved ones, at the very time that a mature understanding of the importance of others for one's health, happiness and security is at its peak. This paradox is particularly poignant in middle adulthood because fulfillment is based in no small measure on the incredible richness of involvements that grow by leaps and bounds between 40 and 60. Unlike the 20s and some of the 30s, when in the midst of the third individuation individuals have left their family of origin and not yet created a family of procreation, midlife men and women are immersed in relationships with spouse, children, elderly parents, in-laws, friends, and colleagues and are forging new ties with new individuals of great personal importance, grandchildren. At no other point in life is the potential for attachment— and loss—so great. Accepting this juxtaposition of interdependence with others and the inevitability of total separation and loss is a central developmental task of middle adulthood that must be engaged and mastered if developmental progression is to continue (Table 50-4)

To integrate the potential for attachment and loss: the fourth individuation
To accept the aging body
To accept time limitation and personal death; time sense in middle adulthood
To reappraise relationships; to let children go, achieve a relationship of equality with them, and integrate new members into the family
To accept the reversal of roles with elderly parents
To develop midlife friendships
To become a generative mentor and plan for retirement
To give play new meanings and purposes
To become a grandparent

Table 50-4 The Developmental Tasks of Middle Adulthood

Accepting the Aging Body Physical decline begins to affect psychological development in the 20s and begins to have a significant effect in the 30s. By midlife, however, because of the universal awareness of aging and the marked increase in major illnesses and peer group death that occur then, thoughts and feelings about the aging body become a major, sometimes dominant influence on mental life. The appearance of the midlife body takes on a different significance. Efforts to remain trim and fit are no longer made to develop a sense of identity or to separate and individuate as they were in young adulthood. The struggle in midlife is to maintain body integrity in the face of anxieties about aging, failing health, and the potential loss of independence. Awareness of change in physical appearance and function is constant. In addition to changes in public presentation such as vision, hair color, and skin tone, more private aspects of physical functioning change, such as cessation of menstruation, altered sexual functioning, increased urinary frequency, and diminished force of the urinary stream.

The reaction to these dramatic biological changes varies considerably from individual to individual, as evidenced by the responses to the menopause. Formerly thought to be a direct result of decreased production of estrogen, more-recent studies indicate that emotional lability, depression, irritability, and other somatic and psychological complaints are individual responses that are present to varying degrees in a minority of women. Only cessation of menstruation and hot flashes occur universally, because they are the direct result of diminished estrogen production.

The physical changes just described are experienced mentally in the form of body monitoring—a continual, conscious and unconscious, comparison of the midlife body with the body of youth. This painful process leads to a conflict between wishes to deny the effects of aging and the need to accept the loss of a youthful body. The normative result is mourning for the lost body of adolescence and young adulthood. Attempts to deny aging include the inappropriate use of plastic surgery, fusion with younger bodies, and exaggerated physical competition with younger individuals. The acquisition of possessions such as rare art, expensive automobiles, and fine clothing is also used as substitutes for the body. Since the aging process continues for the rest of life, there is no resolution of this conflict, but in the healthy individual, the gradual acceptance of aging produces a major change in the body image and increases the pleasures that the midlife body can provide, particularly if it is cared for properly.

Accepting Time Limitation and Personal Death: Time Sense in Middle Adulthood Because of the aging process in the body, the death of parents and contemporaries, the growth of children into adulthood, and grandparenthood and the approach of retirement, midlife individuals come face to face with their mortality and with the painful, but unavoidable, realization that the future is limited and they will die. An extremely powerful psychic organizer, acceptance of the inevitability of personal death precipitates reexamination of all aspects of past and present life and reassessment of how the time left to live will be used.

DEATH AWARENESS Freud observed that we show an unmistakable tendency to put death aside, to eliminate it from life. Our instincts do not respond to a belief in death, and in the timelessness of our unconscious, we are convinced of our immortality. An understanding of a developmental line of time sense from childhood and young adulthood helps explain how experience counters this unconscious (and conscious) wish and false belief in immortality and culminates in midlife in a poignant, painful, monumental conflict between irrationality and reason that, more than any other, defines the mature adult and the essence of the human condition.

Childhood is characterized by a tendency to deny the inevitability of personal death because of the immaturity of the psychic apparatus, the forward thrust of maturation, and the limited understanding of the concept of time. Little in the anabolic thrust of the developmental process indicates a personal end.

Then, in late adolescence, the loosening of intrapsychic ties to the parents elicits a sense of history that includes the realization that a part of the self and an entire epoch of life (childhood) is in the past and irretrievably lost. But this dawning recognition is quickly defended against by the optimism and idealization of youth, a semisuccessful attempt to deny the two fundamental features of human life—the inevitability of death and the existence of hate and destructive impulses inside each person. Gradually, a more contemplative pessimism replaces the youthful belief that the power of stasis, the status quo, and catabolism can be overcome. With youth behind them, normal midlife adults must confront their finiteness as the defenses against the acceptance of time limitation and personal death crumble before the power of new physical, psychological, and environmental experiences.

Then, too, parents, friends, and contemporaries die. The death of a parent, in particular, undercuts the childhood sense of unending continuance and safety that was provided by the good-enough parent. One is left alone, unprotected by the infantile notion of parental omnipotence, with the staggering realization that the child must die, just as the parent did. As the older generation disappears from the scene, the younger generation approaches adulthood. The transformation of children into physically and sexually mature adults also shatters the parent's sense of perpetual youth, because identification with young, immature children is no longer possible as they become bigger and stronger than their progenitors.

Through genetic extension, through the process of identification of parent-child-grandchild, middle-aged grandparents, for the second time in their lives, participate in the creation of new life—the creation of time. Through genetic continuity, children and grandchildren provide the only form of physical immortality that exists. The dynamic relationship between grandparents and grandchildren cannot be fully understood without comprehending the temporal bonds that define this relationship. Grandchildren draw grandparents toward the beginning of life, toward tender years when time itself seemed to exist in unlimited supply, and away from the painful awareness of old age, time limitation, and death.

The realization that time is running out is also painfully thrust into consciousness by the growing awareness that many cherished ambitions and goals will never be realized and that there is not enough time left to achieve new ones of equal importance. Further, a personal end is also forcefully brought into focus by the loss of power and prestige in the workplace or the realization that the highest level of achievement possible has been reached. Power belongs to the young, who have the time to begin new ventures and bring them to fruition in the distant future. As one patient, a highly successful professional, put it: “I've become redundant. The heirs apparent can do everything I can and a lot that I can't. My time in the spotlight is just about over. Soon it will be lights out!”

TIME SENSE IN MIDDLE-AGED WOMEN Puberty focuses and defines gender differences in temporal attitudes, which continue to expand during the remainder of adolescence and adulthood. With the beginning of menstruation, a woman's sense of time is influenced by her “monthly,” her “cycle.” Later, in sexual relations, the female is less constricted temporally; she has no refractory period following climax.

Pregnancy produces another sexually defined cycle for the female. This quintessential psychobiological event that spans 9 calendar months is divided into temporal phases—early, middle and late—and has a clear beginning (intercourse) and end (labor and delivery).

In young and middle adulthood gender differences continue to manifest themselves, most notably through the loss of the procreative function in the late 30s and 40s. The running down of the biological clock leads to the end of this interval and produces two paradoxical effects on middle-aged time sense. The cessation of menstruation is a temporal watershed, forcing a mourning reaction related to the loss of both procreative capacity and its power to imbue time with a sense of youthful purpose and productivity and the awesome ability to create time through the creation of new life. The inability to engage and master this mourning process is a significant dynamic factor in the problematic and pathological responses to the menopause that are seen in some women.

However, after the mourning process related to the menopause is successfully worked through, a major developmental transformation in subjective time sense occurs. No longer dominated by the time tables of the monthly menstrual cycle or the possibility of pregnancy, women are free to use time in new, egocentric ways. This sense of liberation is one of the dynamic factors supporting the contention that a problematic reaction to menopause and the empty nest are not usual responses.

A significant developmental task of middle adulthood for both sexes is the ability, free of significant conflict and guilt, to use increasing amounts of time for egocentric purposes. For women, this task is thrust into the forefront of the developmental process and driven by the menopause. But in both men and women, the newly found temporal freedom and sense of exhilaration that accompanies it become quickly fused with generative activities. In other words, the self is aggrandized by giving to, and caring for, younger generations and identifying with their abundant future and temporal riches. When this fusion occurs, the focus on giving to, and caring for, the young continues a pattern that was begun with parenthood but is now driven by a different dynamic force, namely, the need to react to the midlife confrontation with awareness of time limitation and personal death.

Maintaining Intimacy in the Face of Significant Physical, Intrapsychic, and Environmental Interferences Whereas the young adult is preoccupied with developing the capacity for intimacy, the midlife individual is focused on maintaining intimacy in the face of deterring physical, psychological, and environmental pressures. In a long-standing relationship, these pressures include real and imaginary concerns about diminished sexual capability, emotional withdrawal due to preoccupation with developmental tasks, and the realistic pressures related to work and providing for dependent children and sometimes elderly parents as well. In relationships that begin in midlife all of the above factors may operate, but in addition, the maintenance of intimacy may be compromised by the absence of a common past, age and generational differences in interests and activities, and the difficulties involved in forming a stepfamily.

For sexual intimacy to continue, the participants must accept the appearance of the partner's middle-aged body and continue to find it sexually stimulating and accept the normative changes that occur in sexual functioning. For those who master these developmental issues, the partner's body remains sexually stimulating. Diminished sexual ability is compensated for by feelings of love and tenderness generated over the years by a satisfying relationship. Those who cannot accept the changes in the partner's body or their own stop having sex, begin affairs, or leave the relationship, usually in search of a younger partner.

Normative changes in midlife sexual functioning include diminished sexual drive and an increase in mechanical problems. Men have greater difficulty getting and sustaining erections and experience a longer refractory period after ejaculation. Because of diminished estrogen production women experience a thinning of the vaginal mucosa, a decrease in secretions, and fewer contractions at the time of orgasm. These physical changes are powerful psychological stimuli that produce

normative and pathological responses in every patient in this age group. Because the subject matter is embarassing and difficult, many patients avoid this area entirely and resist attempts by the therapist to introduce the subject.

The demands of raising children interfere with the privacy and emotional equilibrium required for intimacy, as do the pressures and responsibilities of work. Fatigue and diminished interest are common denominators in these circumstances. Patients with deeply rooted problems with sexuality or relationships may use aging, work, and relationships with children or elderly parents as a means of rationalizing their conflicts and refusing to analyze them.

Reappraising Relationships Midlife is a time of serious reappraisal of marriage and committed relationships. In the process individuals struggle with the question of whether to settle for what they have or search for greater perfection with a new partner. For some the conflict rages internally and is kept from others; others express it through action that takes the form of affairs, trial separations, and divorce.

Recent research on happy marriage indicates that these couples, despite internal and real conflict, have found or achieved a special goodness-of-it between their individual needs, wishes, and expectations. They regard this fit as unique and probably irreplaceable. In the eyes of these couples, marital success is based on the ongoing, successful engagement of a number of psychological tasks. Among the most important are providing a safe place for conflict and difference, holding a double vision of the other, and maintaining a satisfying sexual life.

The decision to leave a long-standing, committed relationship has great consequences not only for the two individuals involved but also for their friends and loved ones. The effect on children, in particular, is especially profound, extending far beyond childhood. The effects on the abandoned spouse, parents, and close relatives may be nearly as severe.

Various forms of therapeutic intervention such as marital counseling, individual psychotherapy, and psychoanalysis can be extremely effective in helping uncertain individuals decide what to do or helping those who leave deal with the consequences of their decision on the abandoned partner, children, and other loved ones. Problems relating to intimacy, love, and sex can occupy a prominent position in an outpatient practice.

A 55-year-old patient, Mrs. A., sought treatment “in order to leave my marriage. We've been married for 35 years but I haven't loved my husband for the last 20. I've been so dependent on him all of my adult life that I don't know if I have the courage to leave.” Twice-weekly psychotherapy that lasted 15 months helped her leave her husband, start a business, and begin a new relationship. “I have less money and I'm scared about the future, but I feel alive and in control of my life. I think Bill is happier, too.”

A 43-year-old patient, Mr. S. was continually preoccupied with his marriage during his 4-year analysis. Sexually inhibited during adolescence, he “married the only girl in the world who knew less about sex than I did.” Exploration of his sexual inhibitions led to a decision to stay in the marriage. “I've learned in this analysis that sex is not the rare, extraordinary thing I thought it was as a kid; billions of people do it every day. I know I could go out and sleep with a lot of different women, but how different, or better, would it actually be? Jane and I have built a pretty good life together. She's changed a lot and so have I. I think we can make the next 20 years better than the last 20.”

Mr. V. was distraught when his wife of 21 years left him for another man. During the first 2 years of psychotherapy he worked through his feelings of abandonment. Then at age 52, in a compulsive attempt to reassure himself about his deeply damaged sense of masculine competence, he began sleeping with many different partners. Repeated interpretation diminished the pattern, but 4 years after the divorce he was still unable “to take a chance on getting close to a woman again.” Six years after the divorce, after detailed consideration in therapy, he remarried.

Fifty-year-old Mrs. T. left her “wonderful” husband because “I've missed something. I just have to get out on my own.” Married at age 18, “after going from my parent's home to his home,” she recognized that her rage at her husband for “not being all the other men I could have married, for closing off all the living I could have done” was irrational but uncontrollable. “I have to live on my own for awhile, to see if I can do it, before it's too late.” Fully intending to return to her husband, she continued exploring the infantile and adult issues that precipitated the separation, leaving the future of the marriage in doubt.

Letting Children Go, Achieving Equality, and Integrating New Members Into the Family During middle adulthood children grow into adolescents and young adults. Their inevitable progression through the developmental phases from childhood to adulthood affects every aspect of the parent's life. How parents facilitate their children's individuation and relate to them as young adults sets the stage for a new relationship based on equality and mutuality. The integration of in-laws and grandchildren into their lives can make the difference between a rich, full late adulthood or one characterized by rancor and emptiness.

LETTING GO Young-adult parental vigor and the control of young children go hand in hand, but so do middle-age awareness of physical decline and time limitation and the inevitable loss of control of adolescent and young-adult offspring. The shift in the balance of power between them is gradual, taking place over many years as both engage their separate but interlocking developmental tasks.

Facilitating the developing sexuality of adolescent children is difficult for the midlife parents because of the juxtaposition between their waning sexuality and the adolescent's sexual promise. The liberalization of sexual mores and attitudes since the parent's adolescence may also add to the discomfort by making the contrast between the parent's experience and the adolescent's opportunities quite distinct, particularly for women. Healthy parents struggle with their conscious and unconscious envy and concerns but gradually accept the fact that sons and daughters are likely to begin dating and become sexually active in mid- or late adolescence.

Conscious concerns about their children's sexual behavior and the reactivation of unresolved sexual conflicts from the childhood and adult past are frequent presenting themes in the treatment of middle-aged patients. Others find this area a source of great resistance because of excessive interest in, or undue restriction of, their children's sexual development and the arousal of unconscious incestuous feelings. A fundamental knowledge of child and adult development will help the clinician understand the interplay of the infantile, adolescent, and adult sexual issues that determines the patient's symptomatology and interaction with sons and daughters.

Each developmental transition brings with it the prospect of significant intrapsychic change and a shifting balance in relationships. In middle adulthood that transition is characterized by physical aging and disappearance of the parental function. Cross-cultural studies of older men and women indicate that normal midlife development and late-onset pathology are driven by the same forces, energies released in men and women in the course of the postparental transition toward androgeny. As the gender distinctions between middle-aged parents blur, postparental men become more nurturing and tender, and their wives adopt some of the ascendant, competitive qualities that their husbands have relinquished. As each postparental spouse becomes as the other used to be, the couple moves toward the normal androgeny of later life. Because of its linkage to the genetic requirements of parenthood, this contrasexual transition is, like paternity and maternity themselves, a quasi-universal event. As such, it usually precedes a developmental advance. After some period of psychic dislocation, most men and women accommodate to the changes in themselves and in their spouses and shape the energies liberated by the postparental reversal into new executive capacities of the personality without losing their identities as men or women, in an expanded sense of self.

But the relationship between grown children and their parents does not end here. The potential for further transformation remains as the young-adult child and the middle-aged parent work together to structure a new relationship that is in some ways more complex than the one that existed when they were younger.

ACHIEVING EQUALITY As parent and child move toward the latter stages of middle and young adulthood, respectively, all vestiges of the child disappear. The youthful appearance of adolescent and the early 20s is gone, replaced by the mature young adult who shows signs of aging. The dependent child has been replaced by the independent adult who lives away from home and is self-supporting and significantly involved with others, sexually and emotionally. These adult relationships and capabilities push the healthy parent-child relationship toward equality, but they are not in and of themselves evidence that equality has been achieved.

Ron, a 25-year-old man, complained bitterly in his therapy that his 57-year-old mother continued to treat him like a child. On a recent visit home mother insisted that Ron change his slacks and sport coat, considering them inappropriate for the occasion planned. When an argument ensued, Ron's father intervened (as he always did) and encouraged his son to change his clothes, which he did. Ron readily recognized his parent's need to infantalize him but was unaware of his own struggle between the wish to appease mother and so remain emotionally involved and dependent on her and the desire to be independent. Over the course of 2 years of psychotherapy, as he came to recognize his own contribution to the continuation of an unequal relationship, he began to make more-independent decisions. At first the changes in Ron's behavior were treated as provocations and intensified mother's attempts to control him. Gradually she worked through her feelings of rage, impotence, and loss of control and begrudgingly moved toward a more equal relationship.

In the more normative situation healthy parents not only accept their child's desire for independence and autonomy but also actively encourage moves in that direction whenever possible. The rationale for such behavior is not entirely altruistic, since it is based in part on the recognition of how involvement with this very special adult and his or her loved ones enhances the parent's mid- and late-life development.

INTEGRATING NEW MEMBERS INTO THE FAMILY Mother-in-law jokes are more than the sarcastic expression of some comic's humor. They also illustrate a universal tension between the newly married and their in-laws. The task for the in-laws is clear enough, if difficult to achieve: give up your claim to the position of primary love object in your child's life, then accept the new partner, initially experienced to some degree as an interloper, and work to cultivate his or her friendship. Once again the motivation for such behavior is not selfless. It is based on a desire to continue to occupy a central, although less important, position in the life of their child and to form a relationship with someone who may add a new dimension to life but who will also exercise some control over future involvement with one's child and grandchildren.

Because spouse and parent of the same sex love the same person a triangle is created, similar to the infantile oedipal one. Consequently, all three individuals are forced to reengage and rework their infantile experience in this new adult context. Like the infantile version, this current rendition is only resolved gradually and partially; the outcome is determined by the emotional capabilities of the participants. The solution for midlife parents may lie in the birth of grandchildren, those marvelous creatures who provide such a significant boost to mid- and late-life development by serving as partial replacements for lost children, connecting midlife development to childhood themes, and providing the only mastery of death that is available, the opportunity to continue living through genetic continuity.

For many midlife parents, particularly mothers, the attraction to grandchildren is not all-consuming, nor need it replace in importance or preference other, extrafamilial, interests. The failure of young adults to recognize that their parents have other interests than them and their children can be a source of conflict between the generations.

Accepting the Reversal of Roles With Elderly Parents At some point, as elderly parents become less able to care for themselves, a role reversal occurs. The child becomes the parent of the parent, increasingly fulfilling the functions of physical and mental caregiver.

A middle-aged patient described the change in her relationship with her 83-year-old mother. “It's sad. She used to be so vital. I remember her being so strong. When I was a child she worked from day to night. Now I have to help her when she walks. Yesterday I had to cut her meat. I felt like I was taking care of a 2-year-old. But every once in awhile she gets that spark back in her eyes, and I remember what she was like for so many years. My daughter was there at the time. Neither one of us said anything, but I knew we were seeing ourselves in the same situation, 20 or 30 years in the future.” For this woman the relationship with her mother was a central aspect of her psychotherapy, stimulating both a reengagement of issues from the childhood and adolescent past and the consideration of phase-specific midlife themes.

The acute reversal of roles that occurs when elderly parents cannot care for themselves forces the middle-aged child to anticipate the parent's demise, thus stimulating the ongoing process of psychological separation from them. Caring for aging parents is one of the most difficult and most frequently avoided developmental tasks of middle adulthood. In addition to presenting difficult financial and management problems, it forces a reworking of childhood themes; focuses attention on time limitation and personal death; and anticipates the inevitable role reversal that will occur with one's own children.

Avoidance of this developmental task has considerable psychological consequences, including the occurrence of late-onset depression. Clinicians who are aware of the enormous psychological power of this developmental process will pay close attention to their patients' experience in this area and will address the resistances to such material when it is absent.

When an elderly parent dies, no matter how expected or anticipated, a mourning process ensues. Long after the acute phase of this process is over intrapsychic relationships with dead parents remain dynamic, emotionally charged, and an important subject for psychotherapeutic exploration.

Developing Midlife Friendships Unlike friendships in latency and adolescence and, to some extent, in young adulthood, midlife friendships do not usually have the sense of urgency or the need for frequent or nearly constant physical presence of the friend. Midlife individuals have neither the need to build new psychic structure (as do latency-age children and adolescents) nor the pressing need to find new objects (as do young adults). They may have many sources of gratification available through relationships with spouse, children, and colleagues.

As their firstborn sons progressed through high school, two women in their mid-40s became fast friends. In addition to raising money for the school activities in which their sons were involved, thus maintaining a close involvement with the boys, they spent many hours talking about the boys' activities, girlfriends, and plans for college. Their husbands, who liked each other, became acquaintances, not friends. They directed their own feelings about their sons into other relationships. After the boys left for college the intensity of the friendship diminished, tending to peak again during vacation periods.

Because of their unique position in the life cycle, these individuals are easily able to initiate and sustain friendships with individuals of different ages as well as chronological peers. It need not be assumed that the unconscious motivation underlying these relationships is different from any other. However, the capacity for sublimation, particularly of aggressive impulses, may be considerable. Friendships with adolescents may be based in part on an identification with their youth and an envy of their abundant future in an attempt to bolster a sagging sexuality and painful feelings about aging. Friendship with young adults may serve the same psychic aims as well as provide an outlet for sexual and aggressive impulses related to parenthood and work. Friendships with older individuals may have multiple determinants, including longing for preoedipal parenting, oedipal sexual and aggressive gratification, or the passive gratifications of a mentee relating to an older mentor.

However, as at all other points in the life cycle, Freud's recognition that friends can rapidly become lovers or enemies remains completely valid. In the face of a disrupted marriage or intimacy or the pressure of other midlife developmental themes, friendships may quickly become vehicles for the direct expression of impulses.

In the face of his wife's aging, a man in his early 50s turned to a younger woman. In a spirit of locker room comaraderie, he spoke openly to his closest friends of the sexual details of the affair, bragging about sexual prowess and adventure. As it appeared to his therapist (and partly to the patient as well), the woman seemed more interested in the gifts and money that he lavished on her than his sexual prowess. He did not tell his friends of his occasional impotence with the girlfriend and with his wife.

At first, the friends listened to his war stories and vicariously enjoyed the affair. However, their feelings turned to concern when he expressed a desire to leave his wife and business and move away with his girlfriend. Concurrent therapeutic efforts to understand and temper the desire for abrupt action were stymied by the excitement of the affair and an urgent wish for a new life. Dynamically, the behavior was understood as a flight from fears of aging, loss of sexual prowess, impotence, and multiple difficulties in the marriage. These midlife conflicts also had many, readily apparent, infantile determinants that were being presented concurrently in the transference.

The therapist watched as the patient's best friend confronted the patient, telling him his behavior was stupid and inappropriate and that his girlfriend did not really love him and was taking him for a ride. He was told that he was acting like “an old fool” and had better stop before he threw away his life. After initial embarassment and rage, the patient analyzed his response and eventually came to admire his friend's courage and appreciate the depth of his concern. The confrontation produced a postponement of the move and greater willingness to use the therapeutic process. Eventually, after a miniscandal, the patient broke off the affair and began to approach the complicated and painful issues in his marriage and in himself.

PRESSURES ON THERAPIST'S FRIENDSHIPS The relationship between friends and between therapist and patient are alike in that both are primarily aim inhibited. Neither friend nor therapist is expected to use the relationship as an avenue for sexual or aggressive actions. Sometimes, the aim-inhibited nature of the relationship breaks down, and therapist and patient become lovers or enemies. Much more frequently the feelings generated in therapists by their work are displaced onto nontherapeutic relationships such as those with spouse and children or friends.

Although the family is a constant repository of such displacement, friendships may be unconsciously singled out even more because they do not have the central importance of relationships within the family and therefore can be disrupted with less realistic consequences. In many instances, disruption of a friendship causes less intrapsychic pain because it can be more easily rationalized and more easily replaced than a relationship with either a family member or a patient.

In addition, like all other human beings in the second half of life, therapists are subject to the normative conflicts engendered by the incessant pressures of the adult developmental process. If these conflicts become particularly severe or if therapists become subject to an unusual number of traumatic acts of fate such as severe personal illness or the premature loss of loved ones through death, both their work and their friendships may suffer even more.

Becoming a Generative Mentor and Planning for Retirement Work is a psychic organizer of major importance: organizing the use of time, providing meaning and purpose, enhancing relationships, and ensuring financial well-being. Midlife is the time of achievement and exercise of power, the result of years of effort during young and middle adulthood when skills were mastered and seniority acquired.

The narcissistic gratifications related to work may be considerable, compensating for the painful realities of everyday life. Sometimes the imbalance is such that work becomes the main source of emotional gratification, resulting in a relative failure to engage more-difficult developmental issues of midlife and neglect of relationships with spouse and children. As a result the eventual loss of work as a central organizing function and source of gratification may be ignored. The recognition of the juxtaposition of maximum achievement and power in the workplace and the acceptance of eventual displacement by the next generation is at the core of the midlife worker's intrapsychic experience.

Recognition of the conflict may be facilitated by plateauing (lateral movement in the workplace instead of promotion), indicating that the highest level of achievement possible has been attained, or by relationships with subordinates.

As one company president put it, “My toughest job is not running the company and making tough decisions, it's finding and training somebody to succeed me. I know whomever I choose will be grateful and appreciative, but they'll also be anxious to get me the hell out of there.”

The conflict of the midlife mentor is to pass on knowledge and power to the next generation while recognizing that this behavior will lead to eventual displacement. The healthy individual does not act on the anger and envy generated by such understanding to any significant degree and instead sublimates these feelings into generativity. Others, however, may display cruel, sadistic verbal attacks or actions intended to impede the development and progression of the mentee.

Midlife attitudes toward money are closely related to success or failure in the workplace. For almost all individuals except the very wealthy the issue is the same: how to earn enough money to meet the simultaneous demands to cover daily expenses, provide children with the opportunity for higher education or the transition to independent status, care for aging parents, and enjoy life in the present while providing for a secure old age.

In the treatment process money may be a therapeutic preoccupation or be defensively ignored. In either circumstance, the developmentally oriented therapist will help the middle-aged patient deal with the powerful thoughts and feelings about finances and develop realistic plans regarding money management.

Giving Play New Meanings and Purposes Play is a lifelong human activity, an amalgam of physical abilities and limitations and mental capabilities and preoccupations. Thus the unique features of midlife play, not suprisingly, reflect the phase-specific preoccupation with the acceptance of time limitation and personal death. The structure of play changes little beyond adolescence. The games that most adults and children prefer are learned before adulthood begins.

By midlife, aging forces the abandonment of most contact sports and the modification of other physically demanding ones. Individuals who refuse to modify their physical expectations of themselves often present themselves in therapy with pathological responses to physical injuries or an exaggerated preoccupation with physical activity.

The psychological meaning of play also changes. Instead of being a joyous expression of mental and physical capability as it was in childhood and young adulthood, physical play is increasingly associated with the maintenance of physical integrity and the enhancement of the aging body.

All forms of play are increasingly used (consciously and unconsciously) to master the developmental tasks of accepting time limitation and personal death. Golf and gin rummy are examples. Golf is a game full of brief intervals with numerous beginnings and endings; in other words, opportunities to conquer time and imperfection by beginning over and over. There is always another shot, an approaching hole, tomorrow's match; unlike one's life, which has one beginning and, more to the midlife point, an approaching end. Gin rummy and similar mental games eliminate the necessity to use the increasingly imperfect body at all and offer the same concept of inexhaustable rhythmic beginnings.

> The musing of an avid golfer patient who was analyzing his game with great seriousness eventually led to the following insight. “I got so upset on the course yesterday, I considered quitting for good. Then I thought to myself this is ridiculous. This isn't life or death, it's just a game. A month from now I won't even remember what I shot today. I don't want to stop; I love the game. All I have to do is accept a higher handicap as I get older. On second thought, no I don't, when I'm eighty, I want to shoot my age.”

Becoming a Grandparent In the mid twilight of their lives and because of their position in the life cycle, grandparents tend to idealize their grandchildren. The grandparental tendency to engage their grandchildren with intense love and devotion is similar to the toddler's undeterrable need for mother during the rapprochement crisis. The similarity exists because both toddler and grandparent have an intense developmental need for fusion with the objects of their attention as they face an immense developmental challenge. The toddler needs to be refueled before venturing out into the ever-expanding world beckoning beyond the symbiotic membrane, and the grandparent faces the collapse of the world and the great unknown void beyond the end of mortal existence. Thus, the intense investment in, and idealization of, grandchildren serves several defensive and developmental purposes: (1) a narcissistic buffer against the stings of old age and the inevitability of death, (2) a chance for magical repair of one's own life through genetic immortality, and (3) a denial of unalterable imperfections in the self through selective identification with particular qualities in the grandchild. The healthy grandparent sees the falseness in this idealization but deeply enjoys the intensity of the rapprochement that accompanies it.

MIDLIFE TRANSITION AND CRISIS

Although the terms “midlife transition” and “midlife crisis” have become part of the pop culture, they are definable syndromes that have considerable clinical relevance (Table 50-5). The midlife transition is a quasi-universal, normative developmental phenomenon. The midlife crisis, on the other hand, is a pathological state experienced by only a few.

Midlife transition—A searing intrapsychic appraisal of all aspects of life
Midlife crisis—A major pressured upheaval in which long-standing relationships and achievements are abandoned impulsively, without insight

Table 50-5 Definitions

The midlife transition has been defined as an intense reappraisal of all aspects of life precipitated by the growing recognition that life is finite and approaching an end. It is characterized by mental turmoil, not action. For 80 percent of those studied, the examination of relationships, achievements, failures, and future plans was preoccupying, painful, and searing. For the others, this midlife critique was less conscious and much less painful. The common denominator was the need to reappraise all aspects of life and make decisions about them—while there was still time to change. For most people the reappraisal results in decisions to keep most life structures, such as marriages and careers, which have been painstakingly built over time. When major changes are made they are thoughtful and considered, even when they include such major shifts as divorce or a job change. The developmentally aware clinician recognizes that every patient in this age group is engaged in a midlife transition (whether the patient is talking about it or not) and facilitates the process by making it conscious and verbal.

A true midlife crisis is a major, revolutionary turning point in life, involving changes in commitments to career or spouse or both and accompanied by significant, ongoing emotional turmoil for both the individual and others. It is an upheaval of major proportions. A period of internal agitation is followed by a flurry of impulsive actions; for example, leaving spouse and children, becoming involved with a new sexual partner, and quitting a job, all within days or weeks of each other. Although there may have been unrecognized warning signs, those who are left behind are often shocked by the suddenness and abruptness of the change.

Efforts by family members or therapists to get the individual to stop and reconsider usually fall on deaf ears. The overwhelming need is to avoid anyone who counsels restraint and to ignore therapists who recommend examining motivations and feelings before making such major decisions. Usually, in the midst of the crisis, the therapist is left with the painful job of helping those who have been left deal with their shock and grief.

Dr. C., a prominent, 45-year-old surgeon, had been married for 22 years and was the father of three adolescent children. One Saturday he left home for the golf course and did not return. Although he had been somewhat bored by his wife for some time, he did not have any conscious plans to leave her. As he reconstructed his thoughts, he was standing on the tenth tee, after making a par on the ninth, when the thought occurred to him “I'm never going home again.” After showering and lingering over a drink with his friends, he drove for several hours and checked into a motel. The next day was spent in thought “about my life, its meaning and purpose. I suddenly knew it wasn't right for me and I had to change it.” On Monday morning he returned home, left his wife a note in the mailbox telling her, without any explanation, that he was leaving, and drove to his office. There he saw his patients, went to the bank and cashed in a $100,000 certificate of deposit from his pension plan, and got on a plane for a distant city. Upon arriving he left a message for one of his partners asking him to take over his practice. For the next 2 months he did nothing but exercise and think. Eventually he wrote his wife a letter informing her of his whereabouts and of his decision to remain away.

When he came for treatment 16 months later he was working in an emergency room and had begun a relationship with a 43-year-old divorcee with two teenage sons, “not so very different from the family I left,” he said.

Therapeutic efforts to reconstruct the patient's thinking during the acute phase of the crisis were limited by his inability to remember. He just knew he had to change, now.

Therapeutic exploration of the probable dynamics behind the crisis—the recent death of his father, occasional impotence and dissatisfaction with his wife—were illuminating but did not change Dr. C.'s determination to avoid his former life. He refused to meet with his wife and eventually agreed through a lawyer to a divorce settlement that gave her almost all of their assets. He did agree to see his children when they sought him out but did not initiate contacts with them. All of his energies were directed toward “doing what I want to do, meeting new people, and finding myself.”

The therapy seemed to meet the patient's need to understand himself and build a new life. He eventually remarried and supported his new family by continuing to work in emergency rooms, making far less money than he had in his surgery practice. He even took up golf again.

ADULT DEVELOPMENTAL DIAGNOSTIC EVALUATION

The purpose of doing an adult developmental diagnostic evaluation is to collect relevant data about the patient that will allow the clinician to formulate a clear understanding of the normal and deviant developmental influences from childhood and adulthood that underlie the patient's healthy growth and symptomatology and to formulate a treatment plan (Table 50-6).

Traces the patient's life experience from birth to the adult present
Provides the information necessary to understand the meaning of presenting symptoms
Describes the influence of important figures such as parents, siblings, and grandparents in adulthood

Table 50-6 The Adult Developmental History

The components of the evaluation are the same as those in any standard psychiatric evaluation; however, added emphasis is placed on obtaining a developmental history that extends beyond childhood and adolescence to the chronological present. The child and adult developmental histories provide the information necessary to understand the meaning of the presenting symptoms and appreciate their effect. Without detailed knowledge of the patient's life experiences, the diagnostician is reduced to making educated guesses in an impersonal context. Relying on knowledge of child and adult development, the clinician traces the individual's life experience from conception to the present, relating the findings to the information obtained in the history of the present illness.

The influence of important figures such as parents, siblings, and grandparents does not end with adolescence. Consequently, the course of these interactions may be traced throughout life, focusing on such critical issues as the interactions between newly married children and their middle-aged parents. Even after death, parents remain important intrapsychic objects and may be the subject of meaningful therapeutic dialogue when the clinician recognizes their importance to the older patient and inquires about them.

In gathering information about adult developmental processes the historian inquires systematically about all major areas of adult development. Thinking in terms of the adult developmental tasks described earlier in the chapter is one way to organize the process. For example, in covering the young-adult years, the clinician would inquire about how patients feel about the aging process in their body. Is the body being cared for or neglected? In another area questions would be asked to determine whether gradually increasing sexual experience had led to a sense of comfort with the body as a sexual instrument and to the emergence of the capacity for intimacy. Other developmental tasks of young adulthood are addressed in similar ways.

If patients are in middle adulthood, ages 40 to 60, examples of the questions to be asked are as follows: Has an adult-to-adult relationship been forged with grown children and their spouses? Have grandchildren been recognized and enjoyed? Has generativity become a central aspect of relationships with young individuals, particularly in the workplace? Have intimacy and an active sexual life been maintained in the face of diminished sexual drive, menopause, and the environmental

pressures of midlife?

CASE OF JOHN L.

The following detailed case report and discussion illustrate the use of adult developmental concepts in the diagnosis and treatment of a young-adult man, on the cusp of middle age.

John L. entered therapy at age 38, a trim, well-dressed man with streaks of gray in his hair. He was self-referred because despite a happy marriage and a successful law practice, he was increasingly anxious to the point of panic and had been experiencing fleeting thoughts of running away. John L. had little idea what was upsetting him but knew it was “time to talk to somebody.”

John L. was married for the second time 5 years earlier to a woman who had two sons, currently aged 12 and 14. He had a 15-year-old son of his own from his first marriage who lived with his mother in a nearby city. John L.'s own mother was dead, but his father was alive and well.

Initially John L. attributed his symptoms primarily to work. His solo practice, in a city overflowing with lawyers, was very successful but demanded 80+ hours a week of his time. Good help was difficult to find and keep, and his entire existence seemed to revolve around his law practice. His wife thought he worked entirely too hard but had demanded little of his time or energy until recently, when her 14-year-old son began to manifest behavioral problems at home and in school. John L. did what he could to help but felt increasingly guilty as he spent more and more time with his stepson and even less with his natural son, whom he saw only sporadically at best.

The mention of his natural son was accompanied by a striking increase in free-floating anxiety and an unrecognized revelation of important conflictual material. John L.'s girlfriend had become pregnant with his son when they were college students. Both of them had been virgins prior to dating each other and had only engaged in intercourse three or four times when the pregnancy occurred. His parents were dismayed by the pregnancy since they did not like his girlfriend and wanted John L. to complete his education before marrying.

When John L.'s father announced that he would no longer pay for his son's college education, John L. said one word (“Oh?”) and left. The two men never discussed the matter again. John L. married his girlfriend and was cut off financially. He in turn withdrew emotionally. In the past 20 years, contact with his parents had consisted of four or five frosty phone conversations initiated by his mother and a false display of family togetherness at her funeral 5 years ago.

The marriage was difficult from the start, beginning as it did with a pregnancy and an absence of family blessing or financial resources. John L. managed to stay in college—his wife dropped out to care for the baby—but within a short time he knew that the marriage was a mistake. He loved his young son desperately but was increasingly distant from his wife who seemed “superficial, boring, and fat.” After graduation and a brief trial of marriage counseling, he took off, feeling “exhilarated and free. I finally had all of it off my back.”

After working for a while, John L. put himself through law school. His former wife followed him, hoping for a reconciliation, but to no avail. She eventually married a man who made little money but was a decent stepfather. John L. was painfully aware of the difference between the very comfortable lifestyle that he provided for his two stepchildren and the rather meager circumstances under which his natural son lived.

When John L. met his second wife, he felt exhilarated and happy for the first time in many years. She was outgoing, uninhibited sexually, and “made me the center of her life.” They were married within 6 months. His new stepsons, quiet and unassuming latency-aged boys, “seemed like nice kids, I didn't give them much thought.”

John L. did not remember much about his early years, but gradually a picture emerged of a quiet, somewhat shy child who was physically healthy and who did not manifest any obvious psychological problems or developmental delays or deviations. His mother was at home on a full-time basis and seemed to enjoy raising her son.

John L.'s memories of latency were clear and not particularly happy. Although an excellent student, he was short and frequently teased. He had a few friends but felt “on the outside looking in.” His father pushed him into sports where he did not excel and seemed disappointed in him.

Puberty occurred late, in tenth grade. “I remember erections, wet dreams, and thoughts about girls.” When I commented on the absence of masturbation from his list, John L. blushed and said, “that was so good I knew it must be terrible in some way.”

College was a time of greater acceptance socially, a sense of “liberation” from his father, and the beginning of dating and a sexual life. Unfortunately, this heady sense of exhilaration and freedom was short-lived, quickly replaced by premature work, marriage, and parenthood.

The clinician made a diagnosis of anxiety neurosis and presented a recommendation for therapy. After a discussion about schedule and finances, John L. agreed to begin twice-weekly psychotherapy.

John L.'s sexual inhibitions emerged as a central theme in the treatment fairly early on. When his masturbation was analyzed, strongly repressed phallic and exhibitionistic fantasies emerged. For example, he pictured himself as an Arab sheik lounging on silk pillows in a richly decorated palace, choosing his sexual companion for the night from his harem girls, who paraded suggestively before him. As he anticipated the sexual adventure ahead, his erection began to grow, and grow, and grow, breaking through the ceiling and continuing skyward, not unlike Jack's beanstalk. The gradual disappearance of John L.'s sexual inhibitions led to a return of his masturbation and an enhanced sex life with his wife.

The analysis of the transference, specifically the interpretation of John L.'s great concern about the therapist's reaction to his phallic aggressiveness, led to many associations about his father. Gradually he tested the therapist's ability to accept his rage at the father/therapist, eventually spewing forth blast after blast at his father for his coldness and insensitivity, particularly in regard to the pregnancy. Signs of softening and wishes for reconciliation followed, stimulated by the therapeutic work, to be sure, but also propelled by the patient's position on the threshold of midlife. As he approached the age of 40 and his father, 70, he became increasingly aware that if a reconciliation were to take place, it better occur soon, and it did. John L. called his father, who was mildly receptive, and eventually went, along with his son, to visit him. John L. was surprised at the pleasure and inner comfort that he experienced during the visit but was totally unprepared for—and consciously jealous of—the warmth of the rapport that developed almost instantly between grandfather and grandson, strangers just getting to know each other. John L. and his father continued to develop their relationship, which took the form of a reversal of generations as John increasingly cared for his aging father—not without a sense of superiority for being there for his father in his time of need.

In the months following the rapprochement between son and father, John L.'s associations increasingly turned to his relationship with his own son. Following the acceptance of the painful interpretation that he had abandoned his son just as his father had abandoned him, John L. increased their contact and eventually invited his son to come and live with him, “during the few years that are left before he becomes a man.” The transition took place after several months of patient negotiation with both wives.

Watching John L.'s guilt diminish and self-esteem grow as he incorporated the role of full-time father was a rewarding experience for patient and therapist alike. Although personally satisfying, detailed contact with his son also led to the realization that the teenager was struggling, primarily with social relationships. In a bittersweet burst of newly found similarity, John L. recognized that “he looks like me, is beginning to talk like me, and seems to have the same problems I did. At least he'll get help when he's 15, not 40.”

His wife's kindness to his son suprised him and precipitated more guilt since John L. had accepted her children but not the role of their stepfather. Gradually he came to accept the realization that he had three teen-aged sons to raise and support.

The development of his relationship with the three boys and his father as well as the more intimate, less inhibited relationship with his wife gave John L. a sense of integrity and competence that he had never experienced before. “For the first time in life I'm doing what I ought to be doing—I'm being a good husband, father, and son. I wish my Mom were alive to see this. I think she'd be very proud of me.”

he came to accept the realization that he had three teen-aged sons to raise and support.

The development of his relationship with the three boys and his father as well as the more intimate, less inhibited relationship with his wife gave John L. a sense of integrity and competence that he had never experienced before. “For the first time in life I'm doing what I ought to be doing—I'm being a good husband, father, and son. I wish my Mom were alive to see this. I think she'd be very proud of me.”

But the joy and pleasure in these relationships, the shift from a life almost devoid of closeness to one overflowing with intensity, led to the realization that John L. was responsible for his loved ones. Increasingly, he began to recognize and accept that as an involved midlife husband, father, and son, he had assumed the “burden” of caring for his father as he aged and eventually died, of raising his sons, including the probability of putting all three of them through college, and of remaining emotionally and sexually intimate with his wife in a sustained relationship. “It's tough,” he said, “but it's great. Life does begin at 40!”

Consideration of the Clinical Material From a Developmental Standpoint

The Third Individuation The transition from the second to the third individuation is a young-adult experience, stimulated by the growing capacity for intrapsychic separation from the parents and engagement of the phase-specific developmental tasks of young adulthood. John L. had been individuating rather easily—as evidenced by the growing sense of emotional self-sufficiency at college, academic success, and the beginning of an adult sexual pattern—until the unexpected pregnancy and his parents' reaction to it arrested the process. Although John L. appeared to have mastered the third individuation (he married, became a father, had a successful career, and moved to another state), in reality he repressed his primarily neurotic conflicts around sexuality and aggression and ran away from his wife, son, and parents. This state of affairs continued until the death of his mother, the approaching old age of his father, and the adolescence of his son upset the tenuous emotional equilibrium that had existed through much of his young adulthood and resulted in the outbreak of intense anxiety. The unconscious command was do something about these relationships before time runs out.

Developing the Capacity for Intimacy John L. married his first wife out of need and spite. As their relationship was compromised by pregnancy, parenthood, and near poverty, he emotionally withdrew from her. It took more than 10 years before he allowed himself to become emotionally involved with another woman. For a second time, he began to fuse sexual and loving feelings, this time in an environment far removed from parental influence and with a woman much more advanced than he along the developmental line of the capacity for intimacy. In fact, some of the anxiety that precipitated the search for treatment was related to a growing conscious awareness that he could not match his wife's level of sexual freedom and comfort or her desire for emotional closeness. The removal of many of his sexual inhibitions through treatment made it possible for him to sustain an intimate relationship with a woman for the first time in his life. To quote him, “life does begin at forty.”

Becoming a Father The experience of impregnating a woman adds a new dimension to a young man's sexual identity by confirming that his penis and testicles can perform the primary function for which they are intended. For most males in their teens and early 20s, the goal is avoiding this state. When an active, purposeful attempt is made to become a father, a powerful psychological process is begun. One newly married man described the conscious aspect of the process as follows: “I've always enjoyed sex, but now that my wife and I are trying to have a baby it's totally different. I hope we can do it. It will be pretty awful if we can't.”

John L. became a biological father in his early 20s but avoided a psychological commitment to his son for over 10 years. His failure to become a committed psychological parent was based on his conflicted identification with his father. He attempted to master the devastating psychological abandonment by his father when his girlfriend became pregnant by doing to his son what was done to him. A strong positive aspect in this relationship with his father facilitated his masculine development during childhood and became internalized in his superego and ego ideal. Thus, he could not rid himself of the internalized father or the internalized son. Eventually, without conscious awareness, he had to seek help to rid himself of his guilt by analyzing his conflicts about the roles of father and son relationships and then reestablish his ties with the two most important men in his life while there was still time to affect their adolescent and late-life development.

Developing a Relationship of Mutuality and Equality With Parents While Facilitating Their Mid- and Late-Life Development As John L. neared 40, he was deeply engaged in this process with his father. His reward was a deep inner sense of integrity and authority and, for the first time in many years, the absence of guilt.

When their normal development is not significantly impeded, adolescent males enter young adulthood (that marvelous time of life when masculinity reaches mature fruition), and in the process they become independent adults, loving husbands, and last but not least, fathers to their sons and eventually fathers to their fathers.

WISDOM, MATURITY, AND FULFILLMENT

Success and happiness in adulthood are made possible by achieving a modicum of maturity—a mental state, not an age. However, the capacity for maturity is a direct outgrowth of the engagement and mastery of the developmental tasks of young and middle adulthood. From a developmental perspective, *maturity* may be defined as a mental state found in healthy adults that is characterized by a detailed knowledge of the parameters of human existence, a sophisticated level of self-awareness based on an honest appraisal of one's own experience within those basic parameters, and the ability to use this intellectual and emotional knowledge and insight caringly in relation to one's self and others.

The achievement of maturity in midlife leads to the emergence of the capacity for wisdom. Those who possess wisdom have learned from the past and are fully engaged in life in the present. Just as important, they anticipate the future and make the necessary decisions to enhance prospects for health and happiness. In other words, a philosophy of life has been developed that includes an understanding and acceptance of the person's place in the order of human existence. That worldview, which varies widely from individual to individual, produces fulfillment when the following aspects of the human condition are accepted and integrated.

1. The body must be cared for—in health through regular checkups, exercise, and a proper diet, and in sickness through prompt treatment and proper care. Caring for the body is not an end in itself, but it is critically important because sentience, the essence of human experience, springs from a healthy brain and body.
2. Human beings are individuals alone with themselves, separated and individuated from all others.
3. Paradoxically, human beings cannot survive or develop on their own. Helpless at birth and relatively dependent throughout childhood, even the most self-sufficient individuals require the sustaining presence of others. We exist in a framework of interdependence, a basic characteristic of all human relationships, be it the parental view of the child as a confirmation of his or her sexuality, the need of the child for the parents' loving care and protection, or the reversal of these roles between adult child and aging progenitor. The mature adult (unlike the child who takes, uses, controls, and dominates) mutes the grandiose expectations of childhood and propels the self toward interactions characterized by caring and mutuality, thus striking a balance between personal needs and those of others.
4. Change is a constant in life. A basic aspect of this change is the shifting nature of significant relationships. Adult involvement with loved ones such as children, parents, colleagues, and friends is in constant realignment. Healthy marriages deepen in significance while others break up on the shoals of midlife development. Parents die. Children grow, leave, and return with new family members. As opposed to old age and in some respects childhood as well, the task for the mature person is to sort out, categorize, and set priorities among relationships, in the process balancing emotional needs and realistic demands and responsibilities. The shifting nature of relationships stimulates the achievement of greater maturity by forcing a constant redefinition of who one is in relationship to others. Mature individuals mourn lost relationships but can remain focused on current and future ones.
5. All human beings—regardless of wealth, position, power, achievement, appearance, or cultural background—are on the same developmental course. All are born and will die. All have a body with the same functions. All have the same emotional needs for closeness and love and the same vulnerability to loss and deprivation.
6. Few individuals have an exaggerated importance. The wish for such grandiose prominence is universal, the result of the untempered narcissism of infancy and childhood. In reality most individuals are important to themselves and a relatively limited number of others who know and love them. The mature individual accepts this fact without despair and uses the knowledge to set realistic goals and priorities that will increase the chances for personal happiness and fulfillment.
7. Personal time is limited. Everyone will die. Young children do not have the cognitive capacity to understand the notion of personal death. Spurred on by the thrust of physical growth and a seemingly endless future, adolescents and young adults think and act as though they are immortal. The true acceptance of time limitation and personal death occurs in midlife. Then the mature individual, religious or not, stimulated by an awareness of the aging process in the body, the maturation of children, the death of parents and friends, and the arrival of grandchildren, accepts the inevitability of a personal end. As with the realization of the limited importance of each individual, this painful recognition, which precipitates midlife crises in some, stimulates the mature individual to seek fulfillment in each moment, to define what is truly important, and to plan the future to actualize those priorities.
8. Money and possessions have limited intrinsic value. They are a means to an end, tools for enriching life and improving the human condition of loved ones and the broader community. Ownership of tangible objects is temporary; sooner of later they will be lost, left behind, or given to others.
9. Work occupies a central position in adult life. Considered a drudgery by some, the wise person recognizes its extraordinary value. In addition to the obvious function of earning a living, work is organizing, an activity that provides purpose and direction, a meaningful way to manage time, and an environment in which to form sustaining relationships. Satisfied midlife workers, at the peak of power and position in the workplace, facilitate the development of the skills and capabilities of younger colleagues while fully realizing that these individuals will, sooner or later, replace them and assume control of the levers of power. Wishes to hold down and attack younger workers are sublimated into generativity, Erikson's term for enhancing the development of the next generation.
10. Midlife is the phase of life in which the experience of being human can be realized and enjoyed most fully. It is when the combination of physical health and

vigor, power and prestige in the workplace, accumulation of wealth and possessions, and meaningful relationships from within the midst of three or four generations provide the potential for a life overflowing with richness and complexity. The mature life is one in which the triumvirate of human experience—love, work, and play—are successfully balanced to bring true fulfillment.

11. Unfortunately, the joys of midlife do not last forever. Old age lies ahead. Although the hope and statistical expectation is for many years of mental competence and independence, physical and mental decline, increased dependence, and eventually death must be anticipated. Late adulthood has its own great pleasure when there is a focus on continued mental and physical activity, a dominant preoccupation with the present and the future, and involvement with and facilitation of the young. Then death can be met with feelings of satisfaction and acceptance, the natural endpoint of human existence that follows a life lived and well loved.

SUGGESTED CROSS-REFERENCES

Anthropology is discussed in Section 4.1; psychoanalysis in Section 6.1; and Erik Erikson in Section 6.2. Other theories of personality may be found in Chapter 6; child development and psychopathology in Chapter 32. Sexual functioning, normal and pathological is discussed in Chapter 19. Geriatrics and normal aging are discussed in Chapter 51.

CHAPTER REFERENCES

Bee HL: *The Journey of Adulthood*, ed 2. Macmillan, New York, 1992.

Block J: *Lives Through Time*. Bancroft, Berkeley, CA, 1971.

Butler RN: The life review: An interpretation of reminiscence in the aged. Psychiatry *26:*65, 1963.

*Colarusso CA: *Child and Adult Development: A Psychoanalytic Introduction for Clinicians*. Plenum, New York, 1992.

Colarusso CA: A developmental line of time sense: In late adulthood and throughout the life cycle. Psychoanal Study Child *53:*113, 1998.

Colarusso CA: Play in adulthood. Psychoanal Study Child *48:*225, 1993.

Colarusso CA: Separation-individuation processes in middle adulthood: The fourth individuation. In *The Seasons of Life*, S Akhtar, S Kramer, editors. Aronson, Northvale, NJ, 1997.

*Colarusso CA: The development of time sense in middle adulthood. Psychoanal Q 1999.

Dederick JG, Miller HL: Transitions into adulthood: Are they the same for women and for men? In *Gender Issues Across the Life Cycle*, BR Wainrib, editor. Springer, New York, 1992.

Emde RN: From adolescence to midlife: Remodeling the structure of adult development. J Am Psychoanal Assoc *33:*59, 1985.

Erikson EH: *Childhood and Society*, ed 2. Norton, New York, 1963.

Fiske M, Chiriboga DA: *Change and Continuity in Adult Life*. Jossey-Bass, San Francisco, 1990.

Freud S: Three essays on the theory of sexuality. In *Standard Edition of the Complete Psychological Works of Sigmund Freud*, vol 7. Hogarth Press, London, 1968.

Gould R: Transformational tasks in adulthood. In *The Course of Life*, vol VI, *Late Adulthood*, G Pollock, S Greenspan, editors. International Universities Press, Madison, WI, 1993.

Greenspan SI, Polk WJ: A developmental structuralist approach to assessment of adult personality functioning and psychopathology. In *The Course of Life*, vol V, *Early Adulthood*, G Pollock, S Greenspan, editors. International Universities Press, Madison, WI, 1993.

*Guttmann DL: The country of old men: Cultural studies in the psychology of later life. In *Occasional Papers in Gerontology*, ed 5. Institute of Gerontology, University of Michigan, Ann Arbor, 1969.

Heath R: *Princeton Retrospective: The Class of 1954*. Princeton University Press, Princeton, 1979.

Jacques E: The midlife crisis. In *The Course of Life*, vol V, *Early Adulthood*, G Pollock, S Greenspan, editors. International Universities Press, Madison, WI, 1993.

Jarvik LJ, Eisdofer C, Blum JE: *Intellectual Functioning in Adults*. Springer, New York, 1973.

Jung CG: *Modern Man in Search of a Soul*. Harcourt Brace, New York, 1933.

Kimmel DC: Adult development and aging: A gay perspective. In *Psychological Perspectives on Lesbian and Gay Male Experience*, L Garnets, D Kimmel, editors. Columbia University Press, New York, 1993.

*Levinson DJ, Darrow CN, Klein EB: *The Seasons of a Man's Life*. Knopf, New York, 1978.

*Lyketsos CG, Chen L, Anthony JC: Cognitive decline in adulthood: An 11.5-year follow-up of the Baltimore Epidemiologic Catchment Area Study. Am J Psychiatry *156:*58, 1999.

McNally JA: *The Adult Development of Career Army Officers*. Praeger, New York, 1991.

Michels R: Adulthood. In *The Course of Life*, vol V, *Early Adulthood*, G Pollock, S Greenspan, editors. International Universities Press, Madison, WI, 1993.

Nemiroff RA, Colarusso CA: *The Race Against Time: Psychotherapy and Psychoanalysis in the Second Half of Life*. Plenum, New York, 1985.

Neugarten BL, Berkowitz H, Crotty WJ, Gruen W, Guttman DL, Lubin MI, Miller DL, Peck RF, Rosen JL, Shukin A, Tobin SS: *Personality in Middle and Late Life*. Atherton, New York, 1964.

Person EP: The “construction” of femininity: Its influence throughout the life cycle. In *Course of Life*, vol V, *Early Adulthood*, G Pollock, S Greenspan, editors. International Universities Press, Madison, WI, 1993.

Pollock GH, Greenspan SI, editors: *The Course of Life*, vol VII, *Completing the Journey*. International Universities Press, Madison, WI, 1998.

Rudinger G, Thomas H: The Bonn longitudinal study of aging: Coping, life adjustment, and life satisfaction. In *Successful Aging: Perspectives from the Behavioral Sciences*, PB Baltes, MM Baltes, editors. Cambridge University Press, Cambridge, 1990.

Smolak L: *Adult Development*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

*Stevens-Long J: Adult development: Theories past and present. In *New Dimensions in Adult Development*, RA Nemiroff, CA Colarusso, editors. Basic Books, New York. 1990.

*Vaillant G: *Adaptation to Life*. Little, Brown, Boston, 1977.

Vaillant G: Natural history of male psychological health: 17. A forty-five year study of predictors of successful aging at age 65. Am J Psychiatry *147:*31, 1990.

*Van Gennep A: *The Rites of Passage*. University of Chicago Press, Chicago, 1960.

Wallerstein J: The psychological tasks of marriage: II. Am J Orthopsychiatry *66:*217, 1996.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.1 OVERVIEW

51.1A GERIATRIC PSYCHIATRY: INTRODUCTION

LISSY F. JARVIK, M.D., PH.D. AND GARY W. SMALL, M.D.

The number of individuals over age 65 is rapidly expanding. In 1900, for example, 4 percent of the U.S. population was older than 65 years. By 1990 it was 12.5 percent, and by 2030 it is projected to be 20 percent. That increase far exceeds the general population growth—10-fold compared to just over 3-fold between 1900 and 1990—and is projected to continue (e.g., 2½ times vs. just over 1½ times between 1990 and 2050) (Table 51.1a-1).

Year	Median Age	Mean Age	Population, in Millions and as a Percentage of Total Population: All Ages (N)	65 and over (N)	65 and over (%)	85 and over (N)	85 and over (%)
1900			76.0	3.1	4.1%	0.1	0.1%
1950			150.1	12.3	8.2%	0.6	0.4%
1990			248.7	31.1	12.5%	3.0	1.2%
2000	35.7	36.5	276.2	35.3	12.8%	4.3	1.6%
2010	37.2	37.8	300.4	40.1	13.3%	6.0	2.0%
2030	38.5	39.9	350.0	70.2	20.1%	8.8	2.5%
2050	38.1	40.3	392.0	80.1	20.4%	18.9	4.8%

Population: U.S. Bureau of the Census: *Current Population Reports, Special Studies, P23-190, 65+ in the United States.* U.S. Government Printing Office, Washington, DC, 1996.
Mean/Median Age, 2000–2050: Day JC: Population projections of the United States by age, sex, race and hispanic origin: 1995 to 2050. In *U.S. Bureau of the Census, Current Population Reports, P25-1130.* U.S. Government Printing Office, Washington, DC, 1996.

Table 51.1a-1 Aging Population of the United States: 1900–2050

Survival curves illustrate the trend: in 1900 to 1902, for example, about 50 percent survived to age 55; by 1991, about 50 percent survived to age 80 (Fig. 51.1a-1). It is notable that life expectancy at age 85 has consistently been longer for blacks than for whites, both men and women (Table 51.1a-2).

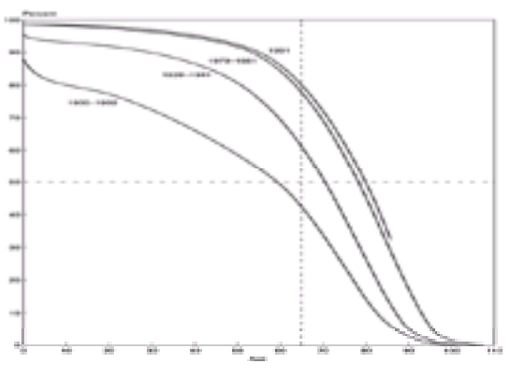

FIGURE 51.1a-1 Percentage of persons surviving to each exact age according to life tables: 1900–1902 to 1991. (U.S. Bureau of the Census. Current Population Reports, Special Studies, P23–190, 65+ in the United States. U.S. Government Printing Office, Washington, DC, 1996. Source: Data for 1901–1902 from U.S. Bureau of the Census, United States Life Tables 1890, 1901, 1910, and 1901–1910, 1921, table 1; 1979–1981 data are from National Center for Health Statistics: United States Life Tables, U.S. Decennial Life Tables for 1979–1981, vol 1, no 1, DHHS Pub. No. (PHS) 85-1150-1, Public Health Service, 1985, table 1, U.S. Government Printing Office, Washington, DC; data for 1991 are from Public Health Service, National Center for Health Statistics, unpublished data from Interpolated Abridged Life Table, 1991.)

Year	Male: White	Male: Black	Female: White	Female: Black
1900–1902	3.8	4.0	4.1	5.1
1909–1911	3.9	4.5	4.1	5.1
1919–1921	4.1	4.5	4.2	5.2
1929–1931	4.0	4.3	4.2	5.5
1939–1941	4.0	5.1	4.3	6.4
1949–1951	4.4	5.4	4.8	6.2
1959–1961	4.3	5.1	4.7	5.4
1969–1971*	4.6	6.0	5.5	7.1
1979–1981*	5.1	5.7	6.3	7.2
1991*	5.3	5.1	6.5	6.3

*Deaths of nonresidents of the United States were excluded beginning in 1970.
From U.S. Bureau of the Census: *Current Population Reports, Special Studies, P23–190, 65+ in the United States.* U.S. Government Printing Office, Washington, DC, 1996.

Table 51.1a-2 Life Expectancy at 85 Years by Sex and Race: 1900–1902 to 1991 (Average Number of Additional Years of Life Remaining)

The most rapidly growing segment of the population is the age group 85 years and older, the group with the highest morbidity and the highest rate of psychiatric and medical comorbidities. This age group grew 30-fold, from 100,000 in 1900 to 3 million in 1990, and is projected to reach 18.9 million by 2050—an increase of over 600 percent (Table 51.1a-1).

The accuracy of the above projections depends on the accuracy of predictions concerning birth rates, immigration, and emigration—all of which are more difficult to gauge for the future than the remaining variables, death rates or life expectancies. Projections concerning life expectancy, for example, can change substantially within a single decade. Thus, on the basis of the 1980 census, the life expectancy for women at birth was projected to continue to exceed that for men by 6.9 years until the year 2050. After the 1990 census, however, new projections show the gap in life expectancy gradually diminishing to only 4.6 years by 2050 (Table 51.1a-3). The corresponding difference in life expectancy in 2050 at age 65 declined from 5.0 years to 2.1 years between the two projections (Table 51.1a-3). By 2050, therefore, the composition of the U.S. population by age and sex is estimated to differ markedly from that of 1990. Such changes are bound to influence income and marital statistics, the percentage of elderly persons living alone or in long-term care facilities, and other aspects of the social network. The social structure of different

ethnic groups will also be affected by changes in the percentage of elderly persons in the population (Fig. 51.1a-2).

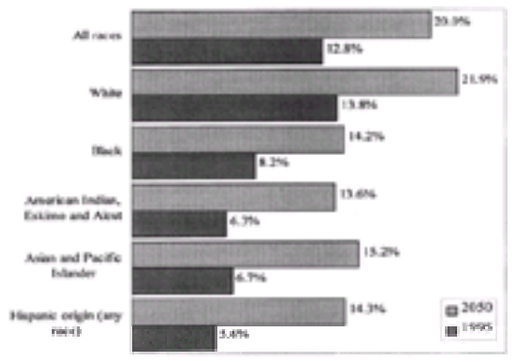

FIGURE 51.1a-2 Percentage of elderly persons, by race and Hispanic origin: 1995 and 2050. (From Day JC: Population Projections of the United States by Age, Sex, Race and Hispanic Origin: 1995 to 2050. U.S. Bureau of the Census, Current Population Reports. U.S. Government Printing Office, Washington, DC, 1996.)

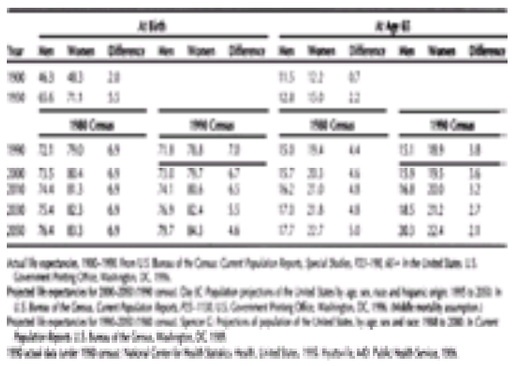

Table 51.1a-3 Life Expectancies at Birth and at Age 65 According to Two Census Reports

Prevalence data for mental disorders in elderly persons vary widely, but a conservatively estimated 25 percent of that population have significant psychiatric symptoms. Thus, the number of elderly mentally ill individuals will total about 9 million by the year 2000 and 20 million by 2050. In recognition of this trend, in April 1991 the first examination was offered for added qualifications in geriatric psychiatry, the first area of specialization recognized within psychiatry in three decades.

In 1988 it was estimated that the geriatric psychiatrist of the twenty-first century will have, at best, about 20 minutes to spend with each patient each year. That prediction assumed that 34 geriatric psychiatrists would complete training annually. Although some new positions have been funded by the Department of Veterans Affairs (beginning with 12 in 1991 and doubling in 1992) and some hospitals are funding their own fellowships, that is still not enough to fulfill the prediction of even the inadequate 20 minutes for each patient each year.

When considering how variations in the geographic distribution of the older population (Fig. 51.1a-3) influence the availability of geriatric psychiatrists, one must remember that the states in which elderly persons constitute a higher percentage of the population are not necessarily the states with the largest numbers of persons over the age of 65. For example, California, with the largest number of elderly individuals (over the age of 65—3.1 million in 1990—ranked only 46th among the 50 states in terms of the percentage of the population that is elderly (10.5 percent). The leader, Florida (18.2 percent), had only 2.4 million elderly individuals in 1990. Nonetheless, clearly, there will be so few geriatric psychiatrists that most will be needed as clinical consultants and academic teachers, leaving the major burden of providing mental health care for elderly persons to their professional colleagues. Consequently, every psychiatrist needs to know the basics of geriatric psychiatry.

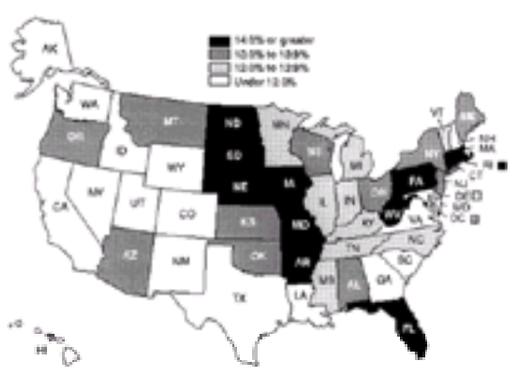

FIGURE 51.1a-3 Percentage of the population that is aged 65 years and over, by state: 1993. (From U.S. Bureau of the Census: *Current Population Reports, Special Studies, P23–190, 65+ in the United States*. U.S. Government Printing Office, Washington, DC, 1996.)

Why is specialized training in geriatric psychiatry emphasized? After all, is not geriatric psychiatry just general adult psychiatry applied to people who are elderly? True, many patients grow old with their psychoses and neuroses. Others remain well until their years and life circumstances create stresses that eventually exceed their ability to cope. It is, perhaps, the frequency and severity of the stressors associated with aging that make geriatric psychiatry unique.

STRESSORS

High-ranking stresses of aging include acute and chronic medical illnesses (Table 51.1a-4), the concomitant use of therapeutic drugs, and the complicating drug-drug and drug-disease interactions. Thus, geriatric psychiatrists must be able to recognize both the physical and mental ills of their patients, as well as having skills in the social sciences, knowledge of the health care delivery system, and information about the availability of financial and social supports, especially nursing homes (Table 51.1a-5). Medical illness connotes physical loss and changes in body image, but the loss of one's job, including voluntary and involuntary retirement, carries with it the loss of financial resources, social status, and much of the social network (Fig. 51.1a-4, Fig. 51.1a-5, Fig. 51.1a-6 and 51.1a-7). The loss of contemporaries through death, illness, and migration brings not only psychological deprivation of an intimate love object but also a void that usually remains unfilled; forming new friendships is difficult in old age. Physical limitations and the loss of friends are frequently associated with restricted mobility, which leads to further social isolation and increased difficulty in pursuing the tasks of daily living, such as procuring food and clothing and maintaining one's shelter. Often, homes are lost because of financial strains and the inability to perform home upkeep. Many widows, for example, have had to move from the 5- to 10-room family homes which they occupied for most of their lives to half a room in a residential extended-care facility for the elderly. In addition to losing most of their wordly possessions and social support, they also lose their privacy and their sense of self-worth.

Condition	Age				Race (65+)		
	65+	65 to 64	65 to 74	75+	White	Black	Black as % of White
Arthritis	480.0	253.8	437.3	554.5	483.2	522.6	108
Hypertension	380.4	229.1	343.8	375.6	367.4	517.7	141
Hearing impairment	286.5	127.7	239.4	360.3	297.4	174.5	59
Heart disease	278.9	118.9	231.6	353.0	286.5	220.5	77
Cataracts	156.8	16.1	107.4	234.3	162.7	139.8	87
Deformity or orthopedic impairment	155.2	135.5	141.4	177.0	136.2	150.8	97
Chronic sinusitis	152.4	173.5	151.8	135.8	157.1	125.2	80
Diabetes	80.2	58.2	89.7	85.7	80.2	165.9	207
Visual impairment	81.9	45.1	69.3	101.7	81.1	77.0	95
Varicose veins	78.1	57.8	72.6	86.6	82.3	64.0	80

From National Center for Health Statistics: Current estimates from the National Health Interview Survey, 1989. In *Vital and Health Statistics*. National Center for Health Statistics, Washington DC, 1990.

Table 51.1a-4 Top 10 Chronic Conditions for Persons 65+, by Age and Race: 1989 (Number per 1,000 Persons)

	1985	1995
Nursing Homes (N)	19,100	16,700
Beds		
(*N* × 1,000)	1,624	1,771
per nursing home	85	106
Current residents		
(*N* × 1,000)	1,491	1,549
Occupancy rate*	91.8%	87.5%
Full-time equivalent employment		
Administrative, medical & therapeutic		
(*N* × 1,000)	NA	20.1
rate per 100 beds	NA	1.1
Nursing		
(*N* × 1,000)	NA	914
rate per 100 beds	NA	51.6

* number of residents divided by the number of available beds.
Adapted from U.S. National Center for Health Statistics. *Advance Data*, No. 280, January 23, 1997.

Table 51.1a-5 Nursing Homes—Selected Characteristics: 1985 and 1995

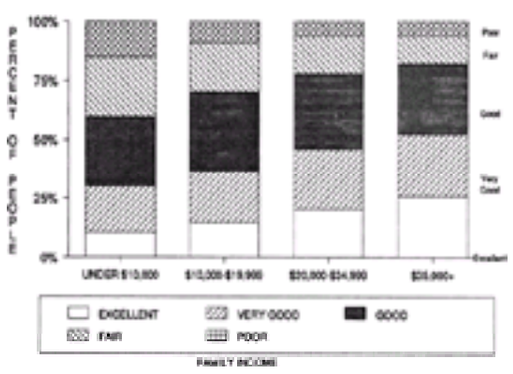

FIGURE 51.1a-4 Self-assessment of health by income for people 65+: 1989. (From National Center for Health Statistics: Current estimates from the National Health Interview Survey, 1989. In *Vital and Health Statistics*. U.S. Government Printing Office, Washington, DC, 1990.)

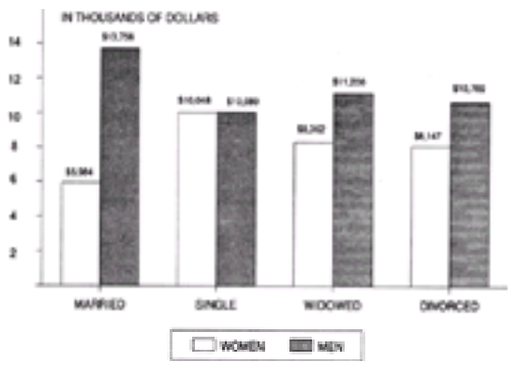

FIGURE 51.1a-5 Median income of elderly men and women, by marital status: 1989. (From the Congressional Research Service: *Current Population Survey*. U.S. Government Printing Office, Washington, DC, 1990.)

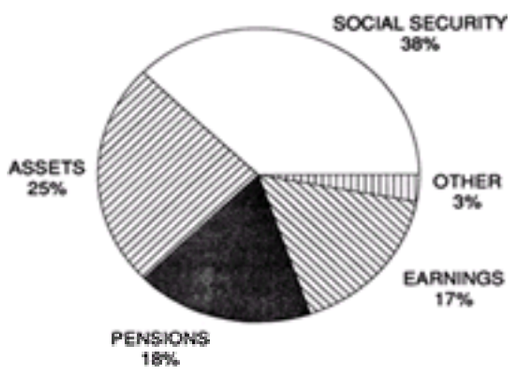

FIGURE 51.1a-6 Income sources of persons age 65+: 1988. (From Grad S: *Income of the Population 65 or Over, 1988*. U.S. Government Printing Office, Washington, DC, 1990.)

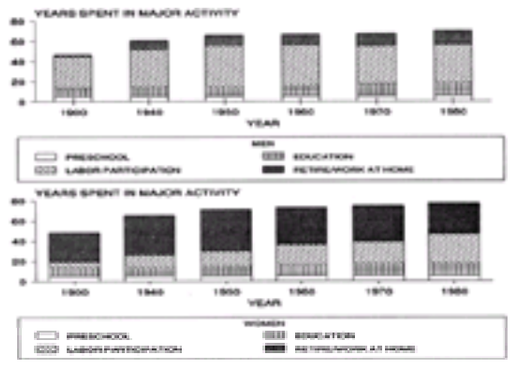

FIGURE 51.1a-7 Life cycle distribution of major activities: 1900–1980. (From U.S. Bureau of the Census: Educational attainment in the United States: March 1981 and 1980. In *Current Population Reports*. [median years of school for persons 25 years or older, 1940–1980]. U.S. Government Printing Office, Washington, DC, 1984; Best F: *Work Sharing: Issues, Policy Options, and Prospects*. Upjohn Institute for Employment Research, Kalamazoo, MI, 1981; National Center for Health Statistics: Life tables. In *Vital Statistics of the United States*. U.S. Government Printing Office, Washington, DC, 1990; U.S. Department of Labor, Bureau of Labor Statistics: *Worklife Estimates: Effects of Race and Education*. U.S. Government Printing Office, Washington, DC, 1986.)

HETEROGENEITY

Generalizations concerning the stresses of aging must be viewed in light of the heterogeneity of the aged as a group as well as the uniqueness of each older person. Individual variability characterizes the geriatric age group not only in terms of chronological age, physical and mental vigor, socioeconomic resources, educational background, ethnocultural heritage, spirituality, and life experience, but also in terms of physiological, anatomical, genetic, and psychodynamic elements. At one end of the spectrum is the healthy, active, involved person, usually under age 85, whose lifestyle differs imperceptibly from that of the middle-aged adult despite changed responsibilities (e.g., grandchildren, management of increased assets and estate planning). Many in this group are retirees who are busier now than they were before retirement. At the other end of the spectrum is the frail octogenarian, nonagenarian, or centenarian plagued by chronic illness, sensory loss, and debility, who has difficulty in processing information and performing the essential activities of daily living. Age alone is not the sole determining factor, as illustrated by the space flight of Senator John Glenn while in his mid-70s.

The geriatric psychiatrist must be aware of the vast individual variability in this age group and must be sensitive to the possibility that medical comorbidity as well as atypical presentation of psychiatric illness may complicate diagnosis and treatment of frail elderly persons. Subtle cues may have to be pursued in detecting such diverse disorders as major depression, subdural hematoma, or infectious or hematological disease.

Attention to psychiatric symptoms among medically ill patients has just begun to become a focus for geriatric psychiatrists. Traditionally, these patients were excluded not only from clinical trials but from other research targeting uncomplicated patients with a single disorder (i.e., the disorder under study). Psychiatric comorbidities are emerging as important factors in increasing length of acute inpatient stays and frequency of medical complications as well as mortality. Therefore, it is expected that geriatric psychiatrists will participate much more in the care of these patients. Further, the geriatric psychiatrist must recognize that the impaired functioning produced by psychiatric or medical illness may be only temporary or may persist for prolonged periods. In either case, the goal is to maintain the individual in the least restrictive environment. Many older individuals are embedded in a system: family members, friends, or neighbors are closely involved in activities of daily living. The psychiatrist must be skilled in assessing caregivers: is the caregiver at risk for medical or psychiatric illness or for substance abuse? Do caregivers need respite, or are they unable to continue caring for the older individual for other reasons? Are there appropriate resources in the community such as day programs, assistance in the home, or group living situations? Does the community have accessible and affordable options for transportation for older individuals? Is elder abuse present, whether physical abuse, neglect by self or others, or fiduciary abuse? Ethical and legal dilemmas that arise as people age mandate additional expertise in assessment and the ability to make appropriate referrals.

The psychiatrist who treats older adults must be comfortable collaborating with primary care physicians and various subspecialty physicians. The ability to work with a multidisciplinary team is essential. Education of physicians, nonphysicians, and older individuals and their families is an important role for the geriatric psychiatrist. The older adult must be able to access psychiatric care. Barriers to care include such beliefs on the part of health providers and individuals in the community as “I'd feel depressed too, if I were old and sick,” “Medicines won't do any good, they'll only make her sicker,” and “That person is too old for therapy.” Older individuals themselves may fear being labeled “crazy” if they seek psychiatric help. Knowledge of beliefs of different cultures regarding aging and psychiatric illness will contribute to improved access to evaluation and treatment.

Advocacy for the older adult by the psychiatrist and members of other disciplines will become increasingly important as guidelines for care, reimbursement policies, and measurement of outcomes are developed in both the public and the private sectors. Organizations such as the National Alliance for the Mentally Ill (NAMI), the American Association of Retired Persons (AARP), the American Association for Geriatric Psychiatry, and the Alzheimer's Association will have input in many areas, including care of aging chronically mentally ill patients. As managed care organizations enroll increasing numbers of older adults, the need for adequate access to psychiatric care will emerge as will the need for collaboration between psychiatrists and other physicians in the diagnosis and treatment of psychiatric disorders.

The quiet apathetic 86-year-old individual brought to the office by relatives convinced that “something is wrong” is readily dismissed as demented, probably suffering from Alzheimer's disease and best left undisturbed. In collaboration with the geriatric psychiatrist who makes the diagnosis of coexisting myocardial infarction and depressed mood, this type of patient can receive appropriate treatment for both conditions. All psychiatrists, all physicians, all mental health specialists, and all health care professionals should bear in mind that most elderly patients are neither demented nor suffering from any other psychiatric disorder; and, they must carry out their professional mandate to assess each patient carefully, regardless of age, before arriving at a diagnosis. This means giving thoughtful consideration to the differential diagnosis, including the diagnosis of no mental disorder.

BURDEN OF CARE

As the cost for extended-care facilities continues to skyrocket, fewer and fewer persons will be able to afford them; thus the burden of care is likely to increase for the adult children of mentally ill and physically frail elderly parents. Both society and psychiatry are ill prepared. As the “sandwich” generation—the adult children sandwiched between the needs of their parents and the needs of their children—becomes more prominent and more squeezed, more facilities at affordable rates will be needed to treat the major mental disorders of old age, such as Alzheimer's disease (called dementia of the Alzheimer's type in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV]), depression, and anxiety. Those disorders are frequently reasons for the dependence of elderly persons.

Most dependence needs of elderly parents are being fulfilled by adult children, especially daughters (Fig. 51.1a-8). Caring for elderly parents raises numerous psychological issues for the adult children. Geriatric psychiatrists may often be the most appropriate professionals to work with those caregivers—either directly or in conjunction with other mental health professionals—because their expertise encompasses the psychological, medical, and sociological databases required for truly informed decision making.

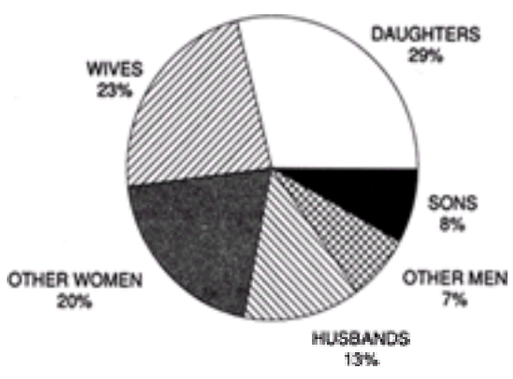

FIGURE 51.1a-8 Caregivers and their relationship to the elderly care recipient: 1982. (From Select Committee on Aging, U.S. House of Representatives: *Exploding the Myths: Caregiving in America*. U.S. Government Printing Office, Washington, DC, 1987.) (Caregiver population includes primary and secondary caregivers.)

TREATMENT

Psychiatrists are familiar with diverse treatment approaches and can use pharmacotherapy, psychotherapy, or both, as well as other somatic therapies. Elderly persons require special techniques in all forms of therapy. The psychiatrist of an 87-year-old patient suffering from heart disease, arthritis, and depression must ask a number of questions: What is the best treatment? Pharmacotherapy? Psychotherapy? Electroconvulsive therapy? If pharmacotherapy, what is the most appropriate drug? Balancing adverse effects and efficacy, what is the best dosage? How soon will the patient's symptoms decrease? If the drug is effective, how long will the improvement last? If the drug seems ineffective, how long should the wait be before changing the treatment? Will the psychiatrist and patient have to settle for improvement without remission?

Currently, psychopharmacology is the mainstay of treatment in geriatric psychiatry. Advantages include acceptability to this cohort of patients accustomed to medication treatment of other disorders, ability of nonpsychiatrists to prescribe psychotropics, and ability to obtain reimbursement. Disadvantages include adverse effects (even with new medications), drug-drug interactions, and interaction of medication with existing disease processes.

Psychosocial treatments have not yet gained acceptance for several reasons: results become evident later than with pharmacotherapy; it is difficult to quantify evaluations of treatment effectiveness; existing data specifically addressing older adults are very limited; treatments tend to be more expensive than pharmacotherapy, even if administered in group settings; and reimbursement is often problematic.

Electroconvulsive therapy (ECT) is an effective treatment for depression in elderly patients. However, its effectiveness and safety have not been widely recognized because of politically charged issues. ECT should be considered especially in frail elderly individuals whose nutritional status and activity level have deteriorated significantly, since it may result in rapid improvement. The risk of memory loss must be weighed against the adverse effects and drug interactions of psychotropic medication.

FUTURE DIRECTIONS

Little is known about the needs of elderly patients in the twenty-first century. What will happen to the Vietnam-era and Gulf War veterans who suffer from posttraumatic stress disorder and to similarly afflicted victims of crime and survivors of genocide in Africa, Asia, and Eastern Europe? Substance abuse and its long-term effects will undoubtedly assume increased importance as the youths of the 1960s reach geriatric age. Will the trend toward decreasing alcohol and nicotine abuse continue among elderly persons? (Fig. 51.1a-9). Clearly, psychiatrists of the twenty-first century will have to be alert to acquired immune deficiency syndrome (AIDS) dementia among elderly persons (Fig. 51.1a-10) as improved treatments prolong the lives of those testing positive for the human immunodeficiency virus. Further, genetic factors will probably assume increasing importance in elucidating susceptibility and resistance not only to AIDS but also to Creutzfeldt-Jakob disease, bovine spongiform encephalopathy (mad cow disease), and other transmissible diseases. Expertise in molecular genetics will help geriatric psychiatrists participate actively in exploring the causes and cures of other dementing (e.g., Alzheimer's disease) and nondementing (e.g., mood disorders) illnesses afflicting the elderly population.

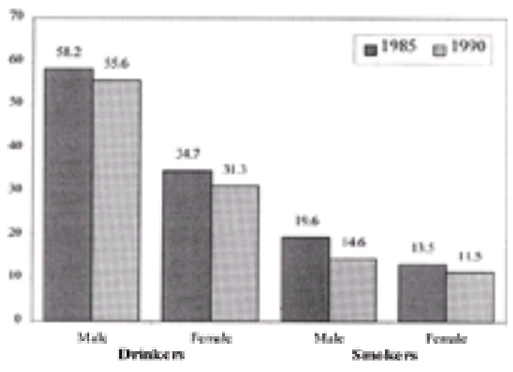

FIGURE 51.1a-9 Alcohol use and cigarette smoking in the elderly. Percentages of people surveyed reporting that they were current smokers or drinkers. (Data from U.S. Bureau of the Census: *Current Population Reports, Special Studies, P23–190, 65+ in the United States*. U.S. Government Printing Office, Washington, DC, 1996.)

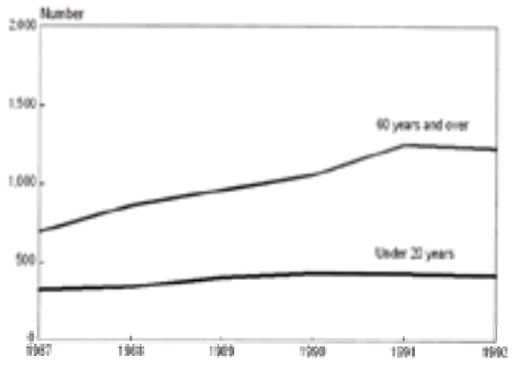

FIGURE 51.1a-10 AIDS deaths under 20 years and 60 years and over: 1987 to 1992. (From U.S. Bureau of the Census. *Current Population Reports, Special Studies. 65+ in the United States*. U.S. Government Printing Office, Washington, DC, 1996.)

Although psychiatrists delight in the advances made in neuroimaging, brain mapping, and other technological approaches to the neurocognitive aspects of Alzheimer's disease and look forward to the discovery of additional genes for the major psychiatric illnesses, progress has been slow in converting the knowledge gained in these areas into effective treatments or prevention. Geriatric psychiatrists entering the twenty-first century will not only increase their collaboration with primary care physicians as increasing numbers of frail elderly persons are maintained in the community, but they will find themselves in new settings as aged individuals increasingly become part of the penal system. Regardless of how diverse the area of geriatric psychiatry, when it comes to treating the individual patient, it will still be the psychiatrist's clinical skills—grounded in scientific knowledge—that will ensure that the patient receives the best care medicine can offer, and it is likely that the psychiatrist will still assume the roles of family doctor, friend, parent, and child to the geriatric patient. It is an awesome responsibility for which psychiatrists must be prepared.

SUGGESTED CROSS-REFERENCES

Normal aging is discussed in Section 51.2c, neuroimaging in Section 51.2e, Alzheimer's disease and other dementing disorders in Section 51.3e, psychiatric problems in medically ill geriatric patients in Section 51.3a, sociocultural and minority issues in Section 51.6d and Section 51.6e, financial issues in Section 51.5a,

AIDS in Section 2.8, health care delivery systems in Section 51.5, and posttraumatic stress disorder in Section 51.3c.

SECTION REFERENCES

Alexopoulos GS, Young RC, Meyers BS: Geriatric depression: Age of onset and dementia. Biol Psychiatry *34:*141, 1993.

*Colenda CC, Streim J, Greene JA, Meyers N, Beckwith E, Rabins P: The impact of OBRA '87 on psychiatric services in nursing homes. Joint testimony of the American Psychiatric Association and the American Association for Geriatric Psychiatry. Am J Geriatr Psychiatry. *7:*12, 1999.

Devanand DP, Nobler MS, Singer T, Kiersky JE, Turret N, Roose SP, Sackeim HA: Is dysthymia a different disorder in the elderly? Am J Psychiatry *151:*1592, 1994.

Iqbal K, Winblad B, Nishimura T, Takeda M, Wisniewski HM, editors: *Alzheimer's Disease: Biology, Diagnosis and Therapeutics*. Wiley, New York, 1997.

Jarvik LF, Small GW: *Parentcare*. Bantam, New York, 1990.

*Krasucki C, Howard R, and Mann A: Anxiety and its treatment in the elderly. International Psychogeriatr. *11:*25, 1999.

*Nyenhuis DL, Yamamoto C, Luchetta T, Terrien A, Parmentier A: Adult and geriatric normative data and validation of the profile of mood states. J Clin Psychol *55:*79, 1999.

Relkin NR, Tanzi R, Breitner J, Farrer L, Gandy S, Haines J, Hyman B, Mullan M, Poirer J, Strittmatter W, Folstein M, Mayeux R, Petersen R, Roses A, Schenk D, Small G, Vangool W, Cook-Deegan R, Fleck L, Kapp M, Karlinsky H, Pericak-Vance M, Post S, Wolpert C: Apolipoprotein E genotyping in Alzheimer's disease: Position statement of the National Institute on Aging\Alzheimer's Association Working Group. Lancet *347:*1091, 1996.

Roses AD, Weisgraber KH, Christen Y, editors: *Apolipoprotein E and Alzheimer's Disease*. Springer, New York, 1996.

*Salzman C, editor: *Clinical Geriatric Psychopharmacology*, 3rd ed. Williams & Wilkins, Baltimore, 1998.

Schneider LS, editor: *Updates in Geriatric Psychiatry*, no. 76. Jossey-Bass, San Francisco, 1997.

*Schneider LS, Reynolds CF, Lebowitz BD, Friedhoff A, editors: *Diagnosis and Treatment of Depression in Late Life: Results of the NIH Consensus Development Conference*. American Psychiatric Press, Washington, DC, 1994.

Seeman TE, Berkman LF, Charpentier PA, Blazer DG, Albert MS, Tinetti ME: Behavioral and psychosocial predictors of physical performance: MacArthur studies of successful aging. J Gerontol Med Sci *50A*:M177, 1995.

Small GW: Geriatric psychiatry fellowship recruitment: Crisis or opportunity? Am J Geriatr Psychiatry *1*:67, 1993.

Small GW: Recognizing and treating anxiety in the elderly. J Clin Psychiatry *58*(Suppl):41, 1997.

Small GW, Birkett M, Meyers BS, Koran LM, Bystritsky A, Nemeroff CB, and the Fluoxetine Collaborative Study Group: Impact of medical burden on quality of life and antidepressant response in geriatric major depression. J Am Geriatr Soc *44:*1220, 1996.

Small GW, Fong K, Beck JC: Training in geriatric psychiatry: Will the supply meet the demand? Am J Psychiatry *145:*476, 1988.

Small GW, Howard RJ, editors: Issue on treatment of late-life depression. Int Psychogeriatr *7*(Suppl):1, 1995.

*Small GW, Rabins PV, Barry PP, Buckholtz NS, DeKosky ST, Ferris SH, Finkel SI, Gwyther LP, Khachaturian ZS, Lebowitz BD, McRae TD, Morris JC, Oakley F, Schneider LS, Streim JE, Sunderland T, Teri LA, Tune LE: Diagnosis and treatment of Alzheimer disease and related disorders: Consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society. JAMA *278:*1363, 1997.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.1 OVERVIEW

51.1B EPIDEMIOLOGY OF PSYCHIATRIC DISORDERS

A. SCOTT HENDERSON, M.D., D.SC., F.R.A.C.P., F.R.A.N.Z.C.P., F.R.C.P., F.R.C.PSYCH.

Epidemiology is concerned with the mass aspects of health and disease. For elderly persons as for other age groups, it has eight purposes: to identify changing patterns of disease; to estimate the prevalence and incidence of specific categories of morbidity; to study the operation of health services and their impact; to estimate the risk for individuals of developing a disorder according to the exposures they have experienced; to complete the clinical picture by determining all the presentations of a disorder and its course at the general population level; to identify syndromes from the distribution of clinical phenomena in the population; to search for causes or risk factors; and finally, prevention, which Ernest Gruenberg said was its ultimate service.

A clinician who deliberately develops an epidemiological style of thinking can gain an appreciation of what contribution is being made, or not made, by a given environmental exposure. Here, the key concept is the population at risk. For example, a patient may attribute mental symptoms to a disturbing experience, recent or past. In ideographic terms, this may be plausible, and the clinician may accept the explanation as at least one causal factor. But the epidemiological thinker must know how many persons have had the same experience but not developed the symptoms and how many had the exposure, developed symptoms, but did not reach professional services and thus were not subsequently referred to a specialist. Only in this way can one determine the probability that the experience in question is indeed causally related to the onset of a disorder.

Global Demographic Changes The need for epidemiological information about elderly persons has taken on some urgency because of an unprecedented change in the world population. Not only is the global population increasing rapidly, but many more people are living into old age, and that change applies particularly to the very old. The change is due to a shift from high fertility and high mortality to low fertility and low mortality. Many more people than in the past live to an age when the likelihood of dementia greatly increases, so that the absolute number of persons with dementia is now expanding rapidly throughout the world. The predicted situation in six economically and culturally diverse countries in shown in Figure 51.1b-1, which is based on data published by the United Nations. It illustrates the growth in the total population, in the elderly population, and in persons with dementia in each country. In some countries, the proportion with a dementia is increasing much faster than the size of the elderly population. This is because of the greater increase in the very old. For some countries, the public health consequences are considerable. According to predictions published by the United Nations, the United States population aged 60 years and older will rise from 43.3 million in 1995 to 92.4 million in 2050, an increase of 213 percent. Even more significant for health services is the projected increase in persons 80 years and older from 8.0 million in 1995 to 24.9 million in 2050, an increase of 311 percent. The social, economic, and public health consequences are profound. Cost-containment alone will not restrain the projected increase in health service use. The prevention and effective treatment of age-dependent disease has become a compelling need. One contribution to this comes from epidemiology.

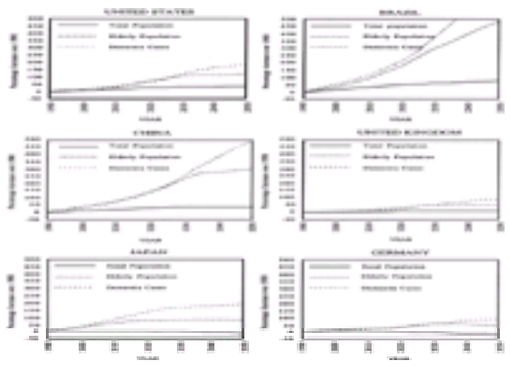

FIGURE 51.1b-1 Dementia cases predicted in six countries. (Adapted from Jorm AF: *The Epidemiology of Alzheimers Disease and Related Disorders*. Chapman & Hall, London, 1990.)

P>**Levels of Inquiry** Psychiatric epidemiology conducts its inquiries at one of three levels: morbidity as it occurs in the community or general population; morbidity as it occurs in primary health care (including family physicians); and morbidity as it occurs in psychiatric or medical services. Research on etiology may be carried out with advantage at the community level rather than on treated series because bias is less likely in a community sample, particularly in relation to the likelihood of exposure to putative risk factors. This can be important in avoiding Berkson's bias, whereby some other variable, unrelated etiologically, may influence the chances of a person's reaching services. The unaware researcher might then mistakenly conclude that the unrelated variable contributed causally. Such biases are usually absent in community samples, which also provide a wider range of severity, whereas hospital or clinic series will, for example, only have depressive disorder of the severe type. But community samples have their own problems. They are laborious and costly, requiring trained personnel to contact and interview hundreds and commonly thousands of persons in their own homes. Invariably refusals and persons who cannot be contacted lead to significant sample bias. Another important but unavoidable bias applies in prospective longitudinal studies in which the focus of interest is the course and outcome of morbidity. When a sample of elderly persons is followed over several years, those who subsequently refuse, are not available, or have died during the study are often more likely to have had the condition in question, such as a dementia, or to have developed it during the course of the inquiry. This means that those who are successfully reexamined are a survival elite who may differ in important ways from the original cohort. These distortions could lead to mistaken conclusions if the nonrandom losses are not allowed for.

Instruments Case ascertainment might be assumed to be the sine qua non for any progress in the epidemiology of mental disorders, namely to identify individuals who have the disorder. But using the traditional expression “case ascertainment” nicely illustrates a problem that has to be overcome: it implies a categorical structure in the morbidity being studied. In a population, there are traditionally cases and noncases, but some have recently argued that many neuropsychiatric conditions have dimensional properties and that there are advantages for some research purposes in not only identifying persons with cases by criteria in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) or the 10th revision of *International Statistical Classification of Diseases and Related Health Problems* (ICD-10), but also determining the score on relevant scales for all members of the sample. The resulting frequency distribution of symptoms such as depression or cognitive impairment is usually highly skewed, with most people having none or only a few symptoms and progressively fewer persons having higher counts. The same applies to measures of disability in the activities of daily living (ADLs), which are an important part of the assessment of older persons. For research on etiology or on the effect of interventions, the inclusion of continuous measures is to be recommended because they lose less information in statistical analyses.

Epidemiological research on mental disorders in late life calls for specially built instruments to detect what is clinically important in this age group: cognitive impairment, cognitive decline, dementia, anxiety and mood disorders, psychotic states, and functional capacity in ADLs. There are two main types of instrument: scales and standardized interviews. The Geriatric Depression Scale is specifically designed for elderly persons.

The standardized research interviews in current use typically have a section for assessing the elderly person and a separate section for an informant, who is usually a relative or close friend. The Geriatric Mental State Examination has been used extensively in diverse countries, allowing comparisons on the prevalence of dementia and depression. The CAMDEX was originally intended as a clinician's instrument but can be used by lay persons after some training. The Consortum to Establish a Registry for Alzheimer's Disease (CERAD) battery is a standardized clinical assessment of persons likely to have dementia of the Alzheimer's type. It is principally a research instrument for use by clinicians. The Canberra Interview for the Elderly was designed for use by lay interviewers and was developed by constructing items to tap each element in the revised third edition of DSM (DSM-III-R), DSM-IV, and ICD-10 criteria for dementia and depressive disorders. The coverage of cognitive function is wide and includes the Mini-Mental State Examination. It also has a depression scale derived from all the depressive symptoms in DSM-IV and ICD-10. Subsequently, using latent trait analyses, the Psychogeriatric Assessment Scales were developed as an optimally parsimonious screening instrument for both clinicians and nonclinicians to detect individuals likely to have depression, cognitive decline, or stroke. These scales are now in wide use for both routine assessment and research.

Matrix for Organizing Epidemiological Knowledge To organize what is known about the epidemiology of mental disorders in the elderly population, it is useful to use a matrix in which the columns consist of the main categories of mental disorders and the rows are made up of the variables that may contribute to the onset or course of morbidity. This matrix helps to organize the information already available, and as inspection will show, it also acts heuristically by proposing associations that otherwise might not have been considered. Table 51.1b-1 sets out such a matrix for the main psychiatric disorders of later life.

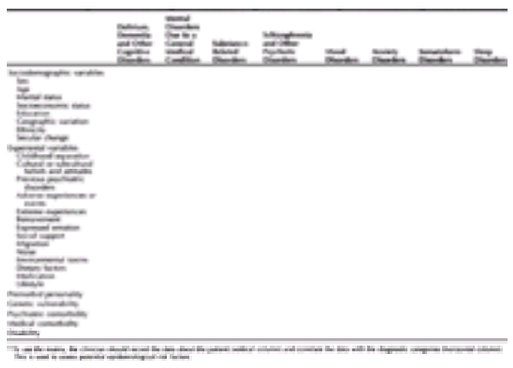

Table 51.1b-1 A Matrix for Organizing Epidemiological Knowledge on Psychiatric Disorders in the Elderly*

DEMENTIA

Incidence Studies There are many fewer studies of the incidence than of the prevalence of dementia, but enough exist to allow some estimate of age-specific rates in the community. As an approximate guide, for dementia of all types the incidence is about one new case per year for every hundred persons aged 65 years and over. What is clear is that the age-specific incidence of both dementia of the Alzheimer's type and vascular dementia rises steeply from the 60s. It is uncertain whether this continues to extreme old age, reaches a plateau, or even decreases in those who survive until then. Data from the Framingham study suggest that the annual incidence of dementia of the Alzheimer's type is about 1.0 percent for men and women aged 75 years and over.

Prevalence There are published reports on the prevalence of dementia in diverse areas of the world including North America, Europe, Scandinavia, Russia, Japan, China, Singapore, India, Africa, and Australasia. Prevalence studies are currently under way in several Latin American countries. As in the incidence studies, only a proportion refer to the prevalence of specific dementias; the remainder are estimates for the dementia syndrome. An integrative analysis of 47 surveys across 17 countries has suggested approximate rates under 1 percent for dementia from any cause in persons aged 60 to 69 years, rising to about 39 percent in persons 90 to 95 years old. The prevalence doubles with every 5 years of age within that range. For both cognitive impairment and all the psychiatric disorders of the elderly considered here, the findings of the large National Institute of Mental Health Epidemiologic Catchment Area (ECA) studies in the United States are summarized in Table 51.1b-2. In the studies in five centers, 4.9 percent of persons aged 65 years and older were found to have severe cognitive impairment, as evidenced by a score of 17 or less on the Mini-Mental State Examination. That test was used instead of the much more demanding task of making a diagnosis of dementia, let alone a specific dementia. The annual prevalence of cognitive disorder in the ECA studies was 4.6 percent in persons 65 years or older but only 0.4 percent in young adults. For the prevalence of dementia of the Alzheimer's type specifically, a pooled analysis from six European surveys suggests rates of 0.02 percent, 0.3 percent, 3.1 percent, and 10.8 percent for the age groups 30 to 59, 60 to 69, 70 to 79, and 80 to 89 years, respectively. Age-specific prevalence rates rising to 47.2 percent in those over 85 years have been reported in an East Boston study. For vascular dementia the prevalence in the United States has been estimated at 3.5 percent for men and 0.8 percent for women 65 years and older.

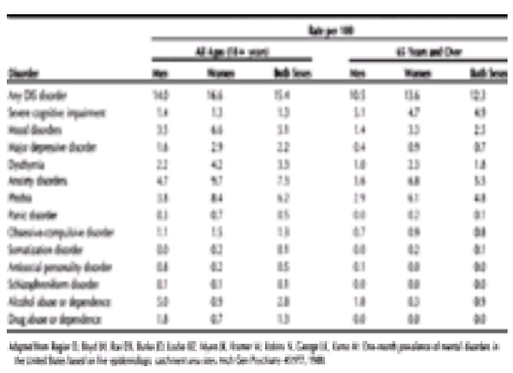

Table 51.1b-2 One-Month Prevalence Rates of Diagnostic Interview Schedule (DIS)/DSM-III Disorders in Five Sites in the United States

Secular Change Is there more dementia now than some decades ago? Because there are larger numbers of elderly persons now than there were only a few decades ago and more of the very elderly in particular, the number of persons suffering from a dementia has inevitably increased. Furthermore, an elderly person with a dementia is now likely to survive longer because of the greater availability of medical and nursing care—what has been called the failure of success. Even with no change in incidence, this would tend to increase prevalence. Studies in Sweden and across Western Europe have found no change in incidence, but it is intuitively plausible that vascular dementia may be decreasing in many countries in parallel with a general decrease in vascular disease.

Age and Sex The incidence of dementia increases steeply with age from the 60s. Some have suggested that the age-specific incidence may plateau or even fall by the ninth decade, so that persons reaching late old age without a dementia may have a reduced risk. But to answer this requires longitudinal data on a representative sample of very elderly persons assessed cognitively on at least two occasions a few years apart. Because the numbers dwindle so markedly, the estimates of incidence tend to be very unstable. As a consequence, the possibility of a late decline in incidence remains uncertain. The overall prevalence of dementia is the same in men and in women, but the prevalence of dementia of the Alzheimer's type is higher in women. This may be partly explained by the fact that those with dementia of the Alzheimer's type are more likely to survive if they are women.

Education A recurrent finding in field surveys is that rates for dementia and cognitive impairment are higher in elderly persons who have had little education. This may be partly due to bias in ascertainment, whereby the tests are done better by persons who are more literate. While such bias may be present in the detection of mild impairment in surveys, it is much less likely to influence the diagnosis of a fully developed dementia. There may indeed be a true gradient in the incidence of dementia, including dementia of the Alzheimer's type, across educational levels. In this way, lack of education could be seen as an exposure that may confer increased risk. One interpretation is that education may delay the point at which a developing dementia becomes clinically manifest. In interpreting results of a large

survey in Shanghai, the authors raised the possibility that having no education may lower brain reserve, allowing the earlier appearance of symptoms of dementia. It is also possible that education is a proxy for other beneficial factors, as in diet or lifestyle. But a recent longitudinal study of American nuns suggests that education and intelligence may actually protect against the neuropathological processes in dementia of the Alzheimer's type. This means that exposure to education in childhood may conceivably have a protective effect many decades later. If such a protective effect does apply, the implications for public health could be appreciable, because education has become much more widely available to young persons during the twentieth century. The incidence of dementia may therefore show a modest decrease in the future.

Geographic and Ethnic Comparisons Dementia of the Alzheimer's type and vascular dementia may occur at different rates in different parts of the world and in different ethnic groups. But it is not easy to be certain what factors are operating. There may be different patterns of mortality for dementia patients with different living conditions, different genetic vulnerabilities, or different environmental exposures. A general population study in Mississippi and North Carolina found more dementia of the Alzheimer's type in blacks than in whites. A study in Israel suggests that dementia of the Alzheimer's type is more common in Ashkenazi than Sephardic Jews. A fairly consistent finding is that the ratio of dementia of the Alzheimer's type to vascular dementia is well below 1.0 in Japan and elsewhere in the Far East, but significantly above 1.0 in white populations. Although the overall prevalence of dementia in elderly Cree Indians is the same as in whites, dementia of the Alzheimer's type is very rare in that group. No explanation has yet been found. In an attempt to disentangle possible genetic, dietary, and lifestyle factors, dementia was studied in Japan and in ethnic Japanese in Hawaii. This study showed that when Japanese men migrate to Hawaii, they acquire a higher risk of developing dementia of the Alzheimer's type. They also acquire a slightly lower risk for vascular dementia. Although these different rates for dementia could have other explanations, this study has yielded some of the best data yet available on the hypothesis that significant environmental factors are involved in the causation of dementia of the Alzheimer's type. A further finding of potentially great significance comes from a comparison of prevalence rates for dementia and specifically for dementia of the Alzheimer's type in 2494 Nigerian Africans and 2212 African-Americans in Indianapolis, both aged 65 and over, examined by the same methods. The prevalence was 2.3 percent and 1.4 percent for dementia and dementia of the Alzheimer's type, respectively, in the Nigerians, but 8.2 percent and 6.2 percent in the combined community and nursing home sample of African-Americans in Indianapolis. This demonstration of significant differences in rates in two different communities with similar ethnic origins again points to the possibility that some unknown environmental factors promote the neuropathological processes in dementia of the Alzheimer's type.

Prevalence of Dementia in Institutional Settings A matter of considerable importance for health economics and services is the proportion of persons with a dementia who are retained in the community, as opposed to being in nursing homes or sheltered accommodation for the elderly. This proportion is likely to vary considerably, since it is dictated by cultural and socioeconomic conditions. In a large prevalence study in Beijing, all persons with dementia lived in the community. This is in marked contrast to the pattern in most Western countries, where as many as two thirds of all patients typically live in institutional settings. In a Cambridge, U.K., study for example, 41 percent of those with moderate or severe dementia were in residential care, while one third of the severely demented were being cared for at home. A related matter is the proportion of nursing home residents who have a dementia. In the United States, a remarkably consistent average prevalence of 52 percent for nursing homes has been found. This estimate agrees closely with the U.S. Department of Health and Human Services' finding of a total of 696,800 persons with organic brain syndrome in U.S. nursing homes, representing 47 percent of all residents. In marked contrast, only 9000 persons with organic disorders were inpatients in psychiatric hospitals. This represents a ratio of 77 to 1 for numbers of persons with dementia in nursing homes compared with numbers in psychiatric hospitals.

Risk Factors For over two decades, there has been intensive epidemiological research to identify risk or protective factors for the most common dementias, so that preventive action can be taken against this serious social, health, and economic problem for many countries.

Vascular Dementia Risk factors are assumed to be similar to those for stroke. These are age, being male, a family history of stroke or dementia, hypertension, diabetes mellitus, transient ischemic attacks, elevated cholesterol or lipid concentrations, cigarette smoking, heavy alcohol use, oral contraceptives, physical inactivity, obesity, and ethnicity.

Dementia of the Alzheimer's Type The power of individual case-control studies of dementia of the Alzheimer's type has been greatly enhanced by an initiative undertaken by the European Consortium on Dementia (EURODEM) established by the European community. They carried out a collaborative reanalysis of 11 case-controlled studies, six from the United States and one each from Australia, Finland, Italy, Japan, and the Netherlands. That analysis revealed risk factors that had hitherto been only speculative. From this and other sources, only four risk factors can now be regarded as confirmed.

AGE As for dementia in general, the incidence rises steeply with age, making it the strongest of all risk factors.

FAMILY HISTORY Having a parent or sibling with dementia of the Alzheimer's type increases the risk of developing the disease about 3.5 times. The risk is greater for relatives of early-onset patients than later-onset patients. In interpreting the epidemiological data for individual patients though, the clinician needs to emphasize that the risk conferred by a positive family history depends on how long that person lives. Those who do not reach old age have a low risk. Even for relatives who live to age 90, the probability that they themselves will develop the disease is only about 50 percent.

APOLIPOPROTEIN E GENOTYPE The much rarer, early-onset dementia of the Alzheimer's type is caused by single genes such as a mutation of the amyloid precursor gene on chromosome 21 or the presenilin genes on chromosomes 1 and 14. But in most cases of dementia of the Alzheimer's type, onset is not until the 70s or 80s. In this group, of much greater public health importance, there are multiple genetic and environmental influences. One of the most exciting discoveries is that the apolipoprotein E e4 allele on chromosome 19 affects the risk of developing the disease. The e4 allele of this gene increases risk and the e2 allele may reduce it. Although early research with clinical samples showed a very strong relationship between apolipoprotein E e4 genotype and dementia of the Alzheimer's type, more-recent studies with general population samples show a weaker relationship. Currently much interest exists in preliminary findings that a combination of having the e4 allele and being infected with the herpes simplex type 1 virus confers a very high risk. It is now clear that although all individuals with the e4 allele are at increased risk, even homozygotes can live to age 90 with only a 50 percent chance of developing a dementia. An interesting proposal is that the apolipoprotein E genotype predicts when (not whether) a person is predisposed to develop this dementia. These epidemiological findings may in time lead to the development of pharmacological methods to slow the deposition of b-amyloid.

DOWN SYNDROME Persons with Down syndrome develop the brain changes of dementia of the Alzheimer's type before age 40. This is believed to be related to their having an extra copy of the amyloid precursor gene on chromosome 21. Interestingly, they do not usually develop dementia until much older, which may mean that accumulation of amyloid is necessary but not sufficient for the development of dementia of the Alzheimer's type.

POSSIBLE RISK FACTORS A number of possible risk factors exist for which the evidence is presently not conclusive. The importance of ethnicity and geographical region is described above.

Head Injury An integrative analysis of early studies showed that a history of head injury increased the risk of developing dementia of the Alzheimer's type by 80 percent. However, most of these studies assessed history of head injury through reports from relatives rather than from medical records, which could lead to biased reporting. A recent study suggests that head injury may be a risk factor only in individuals who carry the apolipoprotein E e4 allele.

Aluminum Because aluminum, known to be neurotoxic, occurs in neuritic plaques, evidence has been sought for an association between exposure to this metal and the development of dementia of the Alzheimer's type. Aluminum is ingested in food, drinking water, antacids, and toothpaste. It is used in kitchen utensils, and it is applied to the body in antiperspirants. The widespread use of aluminum as a flocculant in water supplies has led to public concern, although drinking water provides only a tiny percentage of dietary aluminum. The amount absorbed depends on its bioavailability, and considerable uncertainty exists about its subsequent deposition in the brain. From the epidemiological evidence, involvement of aluminum from drinking water or other sources in causing dementia of the Alzheimer's type remains unproven.

a2-Macroglobulin Inheritance of a polymorphism in the a2-macroglobulin (A2M) gene has been found to confer increased risk of Alzheimer's disease, but this has yet to be confirmed in other studies.

Previous Depression Interest has been shown in a possible link between previous psychiatric disorders or their treatment and the development of dementia of the Alzheimer's type. Examination of such an association requires clear determination of the time of onset of the previous disorder, because disorders such as depression could be early manifestations or prodromata of dementia of the Alzheimer's type itself. In the EURODEM reanalysis based on four studies in which depression had been assessed before the onset of dementia, a relative risk of 1.8 (95 percent confidence intervals, 1.2 to 2.9) was found for a previous depressive episode. This held for episodes more than 10 years before onset of dementia of the Alzheimer's type, but the effect was confined to late-onset cases. No association was found with

exposure to antidepressant treatment or to adverse life events such as death of a spouse, death of a child, or divorce.

Other Possible Risk Factors Three of the unconfirmed risk factors are possibly linked to the association with having Down syndrome: a history of Down syndrome in a first-degree relative, advanced maternal age at one's birth, and an excess of ulnar loops in fingerprints. Lastly, hypothyroidism and a long-standing history of physical inactivity have been reported in some case-controlled studies. Importantly, no evidence links dementia of the Alzheimer's type and antidepressants or electroconvulsive therapy (ECT), neurotropic viruses, allergic disorders, or general anesthetics or blood transfusions. The EURODEM reanalysis found that risk was not increased by any level of alcohol consumption or by occupational exposure to solvents or lead.

Possible Protective Factors Some factors may confer protection against dementia, specifically that of the Alzheimer's type. For the latter, four have been proposed, all identified through epidemiological research: education, anti-inflammatory drugs, estrogen replacement therapy, and moderate amounts of red wine. Each of these suggests the possibility of prevention; prevention can precede an understanding of etiology. The contribution of education is described above.

Anti-Inflammatory Drugs Since an inverse association between rheumatoid arthritis and dementia of the Alzheimer's type was first observed, over 20 publications have examined the possibility that persons who have taken steroids, aspirin, or other nonsteroidal anti-inflammatory drugs (NSAIDs) over long periods have a lower risk of dementia or have slower cognitive decline in late life. Some of these studies have reported a protective effect, but many have deficient designs, and there may be publication bias (i.e., papers with negative evidence for such an effect are less likely to be submitted and accepted for publication). Yet comprehensive and balanced information on the topic is needed. The most recent information is that anti-inflammatory drugs probably do prevent or attenuate the symptoms of dementia of the Alzheimer's type. This effect is biologically plausible in terms of the action of these drugs to inhibit the immune and chronic inflammatory pathology suspected to apply in dementia of the Alzheimer's type. But it is premature for physicians to prescribe anti-inflammatory drugs for dementia of the Alzheimer's type before their effect is established in a randomized controlled trial and the findings balanced with their risks.

Estrogen Replacement Therapy Case-controlled studies suggest possible protection against dementia of the Alzheimer's type afforded to women who take estrogen. But since these women also tend to be better educated and to differ in other lifestyle factors, this finding could be misleading. Again, only randomized controlled trials will establish whether a protective effect exists.

Red Wine A large population-based prospective study in France has found evidence that moderate consumption of red wine protects against the onset of dementia. The work was conducted in Bordeaux, France.

Other Causes of Dementia A third type of dementia has recently been recognized for which the name dementia with Lewy bodies has been proposed. Epidemiological data are not yet available, but it has been claimed to account for 15 to 25 percent of cases of dementia at autopsy, making it possibly the most common pathological group after dementia of the Alzheimer's type. The dementia is diagnosed neuropathologically by the presence of concentric hyaline inclusion bodies in the cytoplasm of neurons in the brainstem and subcortical nuclei. The clinical presentation has the following main features: progressive cognitive impairment often with deficits in tests of attention, problem solving, and visuospatial tasks; fluctuating cognition; persistent well-formed visual hallucinations; sensitivity to antipsychotic drugs; and motor features of parkinsonism. No community-based surveys have yet been able to estimate the prevalence of or the risk factors for this common dementia of unknown etiology.

Three other causes of dementia are Pick's disease, Huntington's disease, and Creutzfeldt-Jakob disease. Because Pick's disease cannot be accurately differentiated from Alzheimer's disease without a neuropathological study and because only a small proportion of all patients come to brain autopsy, little is known about the age-specific prevalence of Pick's disease. In a comprehensive autopsy series in Minnesota, 5 percent of dementia cases in the elderly were due to Pick's disease and 70 percent were due to the Alzheimer's type.

A large Maryland survey established a minimum prevalence estimate for Huntington's disease of 5.9 per 100,000 persons over 10 years old. The disorder can start in persons in their 60s, and one case is known to have begun at the age of 80.

The annual incidence of Creutzfeldt-Jakob disease is 0.1 to 0.2 per 100,000 population. Men and women are equally affected. The age of onset ranges from the late teens to the 80s, with a peak in the 50-to-59 age group. Because the incidence drops sharply in the eighth and ninth decades, the disorder is not related to aging.

Age-Related Cognitive Decline There seems to be a need to describe a state of mild cognitive decline associated with aging but short of dementia in both clinical features and severity. Terms include age-related cognitive decline in DSM-IV and age-associated memory impairment and mild cognitive disorder in ICD-10. The epidemiological evidence indicates that considerable caution is needed before new diagnostic entities such as these are established and their treatment then proposed. In a cohort study of a community sample of elderly persons, those diagnosed as having mild cognitive disorder by ICD-10 criteria were reassessed 3 to 4 years later. The diagnosis proved to depend on the presence of subjective complaints about memory and thinking, to be associated with anxiety and depression, and to have no prognostic value. From this and other studies, the epidemiological evidence is that subjective complaints of memory problems are more highly associated with mood state than with objective decline in cognitive function.

DELIRIUM

Delirium, a term introduced by Celsus 2000 years ago, is derived from the Latin *de* (from, away) and *lira* (track, furrow). Despite its long-established place in nosology, delirium must number among the most underdiagnosed conditions in medicine. Indeed, the most notable feature in the epidemiology of delirium is the marked discrepancy between its true prevalence and the frequency of its recognition. The point prevalence of delirium in elderly persons in general hospitals is conservatively estimated as about 5 percent. In community surveys of the elderly, delirium is rarely reported, presumably because it is a transient disorder with a high mortality and because many cases may be missed because the person was too ill to be seen by survey interviewers.

A classic report of 106 hospital cases of delirium collected in Baltimore, New York, and London noted the variability of both the causes and the clinical features of the disorder; yet no specific relationship was found between the clinical features and exposure to particular noxious agents. Delirium occurs most commonly at the extremes of life. Its highest incidence is in elderly persons, in whom it indicates the presence of systemic or cerebral disease or drug toxicity.

Despite its grave clinical significance, the marked disruptive effects of delirious behavior, and the high mortality attending it, delirium often goes unrecognized in medical and surgical wards. A study of considerable significance found that nurses had often recorded the signs of delirium in patients' notes, but those notes had been overlooked by physicians who tended either not to consider the diagnosis or to misdiagnose it. This is a nice example of how clinical epidemiology can contribute to areas of need in medical education.

MOOD DISORDERS

Elderly persons might be expected to have higher rates of depressive disorders than younger adults because of their social, economic, intrapsychic, and medical circumstances. Information based only on service use or treated prevalence of depression could be misleading: on the one hand, elderly persons may underconsult physicians; and on the other, comorbidity with physical disorders could distort the estimates. Only data from field surveys of all adult age groups can determine whether elderly persons have more or fewer depressive disorders than younger adults. Here a puzzle emerges. For all five centers in the ECA studies, both the lifetime and the 1-month prevalence of major depressive episode was appreciably lower in the elderly than in other age groups (Table 51.1b-2). The 1-month prevalence for major depressive episode was 2.2 percent for adults aged 18 to 64 but less than 1 percent in those aged 65 years or more. For dysthymia, the rates were 3.3 and 1.8 percent, respectively. A large Australian survey has also found that symptoms of depression and symptoms of anxiety become less prevalent as people age. And so do many risk factors for depression such as adverse life events and neuroticism. Some think these findings may be attributable to an error in ascertainment, namely, inappropriately discounting symptoms that could also be due to physical illness or to age itself. Alternatively, a cohort effect has been proposed, in which those aged 65 to 85 years in the ECA studies may have been protected against depression. Selective mortality could also contribute. It seems likely that the risk of cancer and cardiovascular disease increases in persons who have had a depressive disorder.

This raises a further intriguing possibility that is at first counterintuitive, namely that aging protects against depression and indeed against a range of other mental disorders. In the ECA studies, the 1-month prevalence rates for all disorders except cognitive impairment were lower in elderly persons than in any other age group. It has been suggested that exposure to adverse experience over a lifetime may induce increased resistance to depression and anxiety through psychological immunization. Whether the incidence of depression does decrease with age must be further tested, and if it does, the basis needs to be elucidated.

Geographical Distribution The prevalence rates of depression determined by exactly the same methods in elderly persons in New York and London were compared as part of the United States-United Kingdom Cross-National Diagnostic Project. The rates for pervasive depression were remarkably similar in New York (13.0 percent) and London (12.4 percent), despite the arguably different social environments for elderly persons in the two cities. Evidence from Japanese surveys of depression in elderly persons points to strikingly low rates of below 2 percent. That difference is attributed by some to the different social conditions for elderly persons in Japan. For example, even in urban Tokyo, three-generation families constitute 42 percent of the population (versus 3.3 percent in the United States), and the elderly are accorded much respect in Japan. But the existence of truly different incidence or prevalence rates of depression in elderly persons, free of artifact in diagnostic practice, remains to be demonstrated. Explanations that invoke factors in the social environment are premature.

Etiology When elderly persons have a depressive disorder, one may reasonably ask whether or not it is a recurrence. A study in London found that 37 percent of patients in a hospital series had had a previous episode. Several studies have noted that severe life events precede the onset of depression in at least half of all patients; conversely, depressive disorder may emerge in elderly persons with relatively untrammeled lives. Risk factors for depression in late life are a past history of such a disorder, physical illness or chronic pain, recent adverse experiences, and the absence of a confiding relationship. Contrary to expectation, elderly persons who live alone have not been found to be at risk for depressive disorder. A consistent finding in community surveys is that the complaint of loneliness is associated more with depression than with actual social isolation. A recent study of risk factors in a community sample of nearly 4000 elderly Americans found that on simple bivariate analysis, depressive symptoms were associated with increased age, being female, having a low income, physical disease, disability, and poor social support. But when these other variables were controlled in a multiple regression analysis, the association between depression and age reversed: the oldest old had fewer depressive symptoms than the younger old after allowing for confounding by these other factors. In a prospective longitudinal survey of 1045 elderly Australians assessed twice over 3 to 4 years, the best predictors for the onset of depressive symptoms were having had such symptoms at the start, followed by deterioration in health and ADL function, high neuroticism, poor social support, inactivity, and high use of health services. Depressive symptoms were associated with a higher mortality but did not predict subsequent cognitive decline or dementia.

Bipolar Disorders Although bipolar disorders continue into late life, no cases of mania were observed in the ECA 1-month prevalence estimates. Epidemiological information is limited to clinical series assembled at major hospitals, which show that manic episodes can emerge for the first time in elderly persons, usually those who have had previous episodes of depression.

SUICIDE

Elderly men have consistently higher rates for suicide than those in other age groups, in marked contrast to attempted suicide, which has its highest rates in young women. In the United States in 1993, the suicide rate for all ages was 12.1 per 100,000; but it was 22.3 per 100,000 in those 75 to 84 years. Elderly white men had markedly higher suicide rates than elderly black men: 52.1 and 16.3 per 100,000, respectively, in 1993 for persons 75 to 84 years. By contrast, the rates for elderly white and elderly black women were only 6.1 and less than 1 per 100,000. Factors associated with death by suicide in the elderly include depressive disorder, disabling physical illness, chronic pain, and social isolation.

ANXIETY DISORDERS

Anxiety disorders in late life have been more overlooked than depression as a significant cause of morbidity. In the ECA studies the 1-month prevalence for anxiety disorders in persons aged 65 years and older was 5.5 percent (Table 51.1b-2), which is broadly similar to the prevalence in young adults and to estimates from other industrialized countries. By far the most common syndrome was phobia (4.8 percent); the rate for panic disorder was only 0.1 percent. Elderly women had somewhat higher rates than elderly men and those in all other age groups. In an integrative review of eight community-based surveys of persons aged 60 years and above, anxiety disorders had a lower prevalence than in younger adults. The explanation is unknown. Agoraphobia often occurred for the first time in late life, whereas most other anxiety disorders had either persisted from earlier years or had arisen alongside another psychiatric or medical disorder. Depression frequently accompanied generalized anxiety disorder or phobias, and the depression was usually inappropriately treated with a benzodiazepine. Anxiety symptoms may also occur in the early stages of dementia and in association with depressive disorders. Little information is available on the prevalence of posttraumatic stress disorder, risk factors for it, or its natural history in older persons. Obsessive-compulsive disorder is infrequent in elderly persons, as shown in the ECA data (Table 51.1b-2).

SOMATIZATION DISORDER AND ANTISOCIAL PERSONALITY DISORDER

In the ECA studies, all the observed cases of somatization disorder in any age group were in women. Antisocial personality disorder was strikingly rare in elderly persons and was not observed at all in elderly women. The rate for elderly men was one fifteenth of that for men aged 18 to 24 years. That difference could be attributed to a maturation effect, a cohort effect, or differential survival.

SCHIZOPHRENIA AND SCHIZOPHRENIFORM DISORDER

DSM-IV and ICD-10 recognize that schizophrenia can begin in later life and that a number of distinguishing features may be present: a higher ratio of women, a better occupational history, and a higher frequency of marriage than in younger patients, with sensory loss being more common in those with onset at a more advanced age. DSM-IV recognizes that delusional disorder generally has its onset in middle or late adult life. When both affective and schizophrenic symptoms are present, schizoaffective disorder may be diagnosed in both DSM-IV and ICD-10, but this disorder in elderly persons has not attracted much attention. DSM-IV provides a subcategory of dementia of the Alzheimer's type for presentations with delusions as the predominant feature. In the Scandinavian classification, emphasis is placed on paranoid states as a reaction to life events and interpersonal difficulties.

Information about the epidemiology of schizophrenia and schizophreniform disorder in late life comes from two sources: studies of persons who have reached psychiatric services and surveys of elderly persons sampled from the general population. A review of published data from both sources leads to a number of conclusions.

1. Functional psychotic disorders in elderly persons living in the community have lower base rates than the dementias.
2. The prevalence of schizophrenia in late life may be lower than the approximate value of 1 percent generally accepted for the general adult population. In the ECA studies the 1-month prevalence of schizophreniform disorder in elderly persons was 0 (Table 51.1b-2) as were estimates of lifetime prevalence. The latter observation is particularly puzzling, since the lifetime risk of schizophrenia is about 1 percent. One interpretation is that schizophrenia, including paranoid states, does occur in elderly persons (as every clinician knows), but it may often fall short of the third edition of DSM (DSM-III) and DSM-III-R criteria that were used in the epidemiological surveys. Selective mortality, remission of symptoms, or nonavailability of respondents could also explain this result. If there is indeed a lower prevalence in late life, the reasons need to be sought.
3. Among elderly persons, the prevalence of schizophrenia and delusional disorder probably increases with age.
4. Psychotic symptoms are appreciably more common than formal diagnostic categories.
5. In studies that provide separate rates, women have higher rates than men.
6. The reported prevalence rates are likely to be an underestimate due to underreporting and the possible overrepresentation of psychotic persons among those not interviewed.
7. In those studies that inquired about it, impaired hearing and vision were both associated with psychotic symptoms.
8. Where it has been assessed, social isolation is a further associated factor, but it is not known whether this is a premorbid characteristic, a causal factor, or a consequence.
9. Self-report of poor physical health is more common in the psychotic group.
10. Few studies have determined the proportion of persons with psychotic symptoms who were receiving health or social services. Although one might intuitively expect paranoid or deluded individuals to be less likely to come to the attention of or accept professional services, the findings suggest that such persons do not differ from others in their contact with services, at least in some communities.
11. Population-based surveys that assessed both psychotic symptoms and cognitive function consistently found paranoid or delusional symptoms to be more frequent in persons with cognitive impairment. Further, the well-established association between functional psychotic syndromes and dementia in clinical series may be the same association now reported in community samples, that is, paranoid symptoms and cognitive impairment occur together.
12. Field studies have not often implicated variables found to be relevant in clinical series. These include gender, marital status, number of offspring, social relationships, premorbid personality, physical health, impaired hearing or sight, stressful life events and long-standing difficulties, and a family history of mental disorders and mortality.

ALCOHOL AND OTHER SUBSTANCE-RELATED DISORDERS

The drinking habits and associated psychiatric diagnoses of elderly persons are only part of the overall picture. The ECA studies showed that alcohol abuse or dependence was clearly lower in the elderly than in any other age group and that the rate in elderly women was one sixth that in elderly men. Heavy alcohol intake can contribute to cognitive decline and dementia, possibly because of its effects on key forebrain nuclei. An epidemiological study in Liverpool, England, found a clear association between a past history of heavy drinking and the development of either dementia or a depressive disorder in late life. Insofar as alcohol abuse in early adulthood is a potentially modifiable risk factor, the implications for prevention are considerable.

Substance abuse occurs only rarely among elderly persons; indeed, it was absent in the ECA elderly sample. Such data do not include elderly persons who are dependent on prescribed medications, a group that is likely to be far from negligible in size.

SLEEP DISORDERS

Elderly persons are commonly believed to have more sleep problems than younger adults. Data to confirm or refute this are scanty. In a recent community survey of 1485 persons aged 50 to 93 years in the Netherlands, more women had subjective complaints of sleep difficulties, but more men complained of excessive daytime sleepiness. With increasing age, only two variables changed: more had difficulty getting to sleep and more time was spent in bed. The most commonly reported cause of initial insomnia was worrying and nocturia was the main cause of broken sleep. In the National Institute on Aging's Publication “Established Populations for Epidemiologic Studies of the Elderly” covering over 9000 persons aged 65 and over, less than one fifth had sleep problems only rarely, while over half had some regular difficulty. Sleep complaints were linked to respiratory symptoms, physical disabilities, use of nonprescription medication, symptoms of depression, and poor self-rated health. An Australian survey found a prevalence of persistent insomnia of 16 percent in an elderly community sample and 12 percent in nursing home residents. In the community sample, 14 percent were regularly taking a hypnotic drug, compared with 40 percent in the nursing homes. Of those without the complaint of insomnia, 10 percent in the community but over a third in institutions were using a hypnotic agent. Insomnia was again found to be associated with depression, pain, and poor physical health. These findings lead to two conclusions. First, persistent insomnia in the elderly population, as in other age groups, is strongly associated with depressed mood and physical disease. Because of this, insomnia should not be dismissed as a normal accompaniment of aging and thus be ignored as a significant symptom. Second, continued surveillance is needed in family practice, geriatric services, and nursing homes of the routine use of hypnotic drugs by elderly persons.

SERVICES FOR ELDERLY PERSONS

The extent to which elderly persons with mental disorders receive services can itself be examined epidemiologically. The notion of filters proposed by David Goldberg and Peter Huxley proves useful. The first filter determines which persons with mental disorders reach a physician, whether they initiate the consultation themselves or have it initiated by their families or the physician. The second filter determines whether the physician recognizes the significant mental disturbance. The third filter determines referral to a specialist or to specific services for the elderly population. Although family physicians recognize early dementia quite efficiently, they often require further training in the diagnosis of depressive disorders in elderly persons, in whom tiredness, poor appetite, insomnia, and apathy may be mistakenly ascribed to old age. The benefits of correct diagnosis for both the patient and the community are self-evident.

DISABILITY

From a public health perspective, information about elderly persons and the services they require must include an assessment of disability; impaired functioning in daily life, rather than symptoms or diagnosis, largely indicates the need for services. Dementia is second only to arthritis as a cause of disability in this population. Less is known about impaired functioning caused by depressive disorders and depressive symptoms, but since they often occur with physical disease, their interactive effect is considerable. A community-based study conducted by an experienced geriatrician found that neuropsychiatric disorders of cognition, behavior, and mood together with extrapyramidal gait disorders are major predictors of disability in late life. Somatic disorders have less impact.

SUGGESTED CROSS-REFERENCES

A general discussion of epidemiology appears in Section 5.1, of normal aging in Section 51.2c, of sociocultural sciences in Chapter 4.1, of dementia and delirium in Chapter 10, of mood disorders in Chapter 14, of suicide in Section 29.1, of anxiety disorders in Chapter 15, of personality disorders in Chapter 24, of schizophrenia in Chapter 12, of schizophreniform disorder in Section 13.1, of substance-related disorders in Chapter 11, and of sleep disorders in Chapter 21. Psychiatric disorders of late life are discussed in Section 51.3. Psychiatric Rating Scales are discussed in Section 7.8.

SECTION REFERENCES

Birge SJ: Is there a role for estrogen replacement therapy in the prevention and treatment of dementia? J Am Geriatr Soc *44:*865, 1996.

Blacker D, Wilcox MA, Laird NM, Rodes L, Horvath SM, Go RC, Perry R, Watson B Jr, Bassett SS, McInnis MG, Albert MS, Hyman BT, Tanzi RE: a $_2$-Macroglobulin is genetically associated with Alzheimer disease. Nat Genet *19:*357, 1998.

Blazer D, Burchett B, Service C, George LK: The association of age and depression among the elderly: an epidemiologic exploration. J Gerontol *46:*M210, 1991.

Brayne C, Gill C, Paykel ES, Huppert F, O'Connor DW: Cognitive decline in an elderly population—a two wave study of change. Psychol Med *25:*673, 1995.

Breitner JCS: Inflammatory processes and anti-inflammatory drugs in Alzheimer's disease: A current appraisal. Neurobiol Aging *17:*789, 1996.

Christensen H, Henderson AS, Jorm AF, Mackinnon AJ, Scott R, Korten AE: ICD-10 mild cognitive disorder: Epidemiological evidence on its validity. Psychol Med *25:*105, 1995.

Cullen JS, Grayson DA, Jorm AF: Clinical diagnoses and disability in cognitively impaired older persons. Int J Geriatr Psychiatry *11:*411, 1996.

Doll R: Alzheimer's disease and environmental aluminium. Age Ageing *22:*138, 1993.

Eastwood MR, Rifat SL, Roberts D: The epidemiology of dementia in North America. Eur Arch Psychiatry Clin Neurosci *240:*207, 1991.

Flint AJ: Epidemiology and comorbidity of anxiety disorders in the elderly. Am J Psychiatry *151:*640, 1994.

Henderson AS: *Epidemiology of Mental Disorders and Psychosocial Problems: Dementia*. World Health Organization, Geneva, 1994.

Henderson AS: Does ageing protect against depression? Soc Psychiatry Psychiatr Epidemiol *29:*107, 1994.

Henderson AS, Kay DWK: The epidemiology of functional psychoses of late onset. Eur Arch Psychiat Clin Neurosci *247:*176, 1997.

Hendrie HC, Hall KS, Pillay N, Rodgers D, Prince C, Norton J, Brittain H, Nath A, Blue A, Kaufert J, Shelton P, Postl B, Osuntokun B: Alzheimer's disease is rare in Cree Indians. Int Psychogeriatr *5:*5, 1993.

Jorm AF, Mackinnon AJ, Christensen H, Henderson AS, Jacomb PA, Korten AE: The Psychogeriatric Assessment Scales (PAS): Further data on psychometric properties and validity from a longitudinal study of the elderly. Int J Geriatr Psychiatry *12:*93, 1997.

Jorm AF, Mackinnon AJ, Henderson AS, Scott R, Christensen H, Korten AE, Cullen JS, Mulligan R: The Psychogeriatric Assessment Scales: A multi-dimensional alternative to categorical diagnoses of dementia and depression in the elderly. Psychol Med *25:*447, 1995.

Koenig HG, Blazer DG: Epidemiology of geriatric affective disorders. Clin Geriatr Med *8:*235, 1992.

Mackinnon A, Christensen H, Cullen JS, Doyle CJ, Henderson AS, Jorm AF, Korten AE, Scott LR: The Canberra Interview for the Elderly: Assessment of its validity in the diagnosis of dementia and depression. Acta Psychiatr Scand *87:*146, 1993.

Martin GM, Kukull WA: Do cultural differences affect Alzheimer's disease? JAMA *276:*993, 1996.

Meyer MR, Tschanz JT, Norton MC, Welsh-Bohmer KA, Steffens DC, Wyse BW, Breitner JC: APOE genotype predicts when—not whether—one is predisposed to develop Alzheimer disease. Nat Genet *19:*321, 1998.

Orgogozo JM, Dartigues JF, Lafont S, Letenneur L, Commenges D, Salamon R, Renaud S, Breteler M: Wine consumption and dementia in the elderly: A prospective community study in the Bordeaux area. Rev Neurol (Paris) *153:*185, 1997.

Regier DA, Boyd JH, Rae DS, Burke JD, Locke BZ, Myers JK, Kramer M, Robins LN, George LK, Karno M: One-month prevalence of mental disorders in the United States based on five epidemiologic catchment area sites. Arch Gen Psychiatry *45:*977, 1988.

Regier DA, Farmer ME, Rae DS, Myers JK, Kramer M, Robins LN, George LK, Karno M, Locke BZ: One-month prevalence of mental disorders in the United States and sociodemographic characteristics: The Epidemiologic Catchment Area study. Acta Psychiatr Scand *88:*35, 1993.

Rose G: *The Strategy of Preventive Medicine*. Oxford University Press, Oxford, 1992.

Roth M, Huppert FA, Tym E, Mountjoy CQ: *CAMDEX: the Cambridge Examination for Mental Disorders of the Elderly*. Cambridge University Press, Cambridge, 1988.

Social Psychiatry Research Unit: The Canberra Interview for the Elderly: A new field instrument for the diagnosis of dementia and depression by ICD-10 and DSM-III-R. Acta Psychiatr Scand *85:*105, 1992.

United Nations: *The Sex and Age Distribution of the World Populations*. United Nations, New York, 1994.

*van Duijn CM, Stijnen T, Hofman A: Risk factors for Alzheimer's disease: A collaborative re-analysis of case-control studies. Int J Epidemiol *20:*S4, 1991.

*White L, Petrovich H, Ross W, Masaki KH, Abbott RD, Teng EL, Rodriguez BL, Blanchette PL, Havlik RJ, Wergowske G, Chiu D, Foley DJ, Murdaugh C, Curb JD: Prevalence of dementia in older Japanese-American men in Hawaii. JAMA *276:*955, 1996.

Yesavage JA, Brink TL: Development and validation of a geriatric screening scale: A preliminary report. J Psychiatr Res *17:*37, 1983.

51.2 ASSESSMENT

51.2A PSYCHIATRIC EXAMINATION OF THE OLDER PATIENT

ELEANOR P. LAVRETSKY, M.D., PH.D. AND LISSY F. JARVIK, M.D., PH.D.

In general, the psychiatric evaluation of the elderly patient, whether in residential settings, institutions, or in the community, is similar to that of younger adults. The major principles remain the same, that is, establishing rapport, minimizing discomfort, maximizing privacy and confidentiality, and yet eliciting the information necessary to assess the patient's mental and physical status as well as pertinent psychosocial influences. The psychiatrist must go beyond the diagnostic assessment and consider the patient and the setting as a whole, evaluating the influence of internal and external destabilizing factors and assisting in the development of treatment plans as well as maintenance strategies.

When approaching the examination of the older patient, one must remember that older adults differ markedly from one another. Just as the approach to examining a child will take into account whether the child is 3 years old or 12 years old, so the approach to examining the older patient must take into account whether the person is a healthy 75-year-old who recently retired from a second career or a frail 96-year-old who just lost the only surviving relative with the death of the 75-year-old caregiving daughter.

METHODOLOGICAL ASPECTS OF INTERVIEWING OLDER PATIENTS

Assessment of the geriatric patient differs from that of younger adults primarily in regard to greater use of ancillary sources of information and greater attention to concomitant medical problems; prescription and over-the-counter medications; external destabilizing factors such as loss of family and friends (whether through death or relocation), socioeconomic losses, and recent changes in environment; cultural and religious influences; and sensory impairments, especially vision and hearing.

The psychiatrist needs to take a more active stance with geriatric patients than with younger patients, involving interaction with family and other caregivers, referral sources, social services, cultural and religious agencies, and outside hospitals and physicians. Also, the psychiatrist needs to schedule extra time for geriatric patients, since they have longer life histories, may be slower, may fatigue easily, and may be distracted by reminiscing. It is usually not possible, therefore, to complete the psychiatric evaluation of the geriatric patient in a single session; two or three sessions are often needed. Even that may not be sufficient in a slowly progressing disorder. For example, a patient with dementia of the Alzheimer's type may require several visits over a period of months to gauge progress before a diagnosis can be made.

While gathering information from informants other than the patient is always desirable, it may be mandatory with cognitively impaired patients. Often, one must consult more than one informant, and every effort should be made to get in touch with persons in close contact with the patient, be they family members, friends, or neighbors. Information from one or more independent sources intimately acquainted and in close contact with the patient is often vital not only when cognitive impairment is suspected, but also to corroborate changes in personality, mood, and activities of daily living. The reliability and validity of the information obtained from third parties should always be carefully assessed. Psychiatrists need to consider how well informed third parties are and what biases they may have that either favorably or unfavorably distort the information about the patient.

The psychiatrist must see the patient alone, even if there is clear evidence of cognitive impairment. Our recommendation is that the psychiatrist divide the first visit into the following four segments.

Interview With the Patient Alone Interviewing the patient alone shows clearly that the patient is the key individual, that the psychiatrist attaches importance to the doctor-patient relationship, and that the psychiatrist wants to hear what the patient has to say, and it provides the privacy necessary for patients to express opinions freely and provide data that they are not ready to share with family members or other third parties. Information concerning undesirable behavior of significant others may not be elicited in their presence or the presence of someone who might communicate it to them, especially if the patient depends upon them and fears reprisal. Patients may also minimize depressive symptoms and suicidal thoughts in the presence of others because of shame, embarrassment, or fear of burdening the family. The same applies to paranoid ideation, especially if it involves family members or others significant in the patient's daily life.

When interviewing elderly patients one must determine whether they understand the nature and purpose of the examination. One can make sure that perceptual changes are not severe enough to preclude the patient's understanding, for example, by sitting close, selecting the side of the patient's good ear, talking slowly and loudly without shouting, procuring hearing aid devices, providing ample time for responses, and, if there is any doubt, having the patient repeat the question. Writing key words in large print and showing them to the hearing-impaired patient may be valuable.

As with middle-aged and younger adults, one must pay attention to culturally derived attitudes toward psychiatry and toward symptoms of mental illness and their expression. Many elderly patients are reluctant to see a psychiatrist because they fear they will be labeled “crazy.” Stress interviews are rarely indicated in the geriatric patient, and the same is true of direct confrontation. Geriatric patients tend to withdraw and discontinue treatment rather than being stimulated to explore feelings aroused by a challenging or probing confrontation. For example, cognitively impaired patients, aware of their deficits, may refuse to cooperate and thus prevent full evaluation if their deficits are clearly exposed, especially in the presence of others. Indirect assessment may yield far richer data.

Family or Significant Others Without the Patient The psychiatrist then tells the patient that the others will now be seen in private and then the patient will be called back in. The absence of the patient makes it easier to explore subtle signs of cognitive impairment and paranoid ideation as well as course of illness. The psychiatrist must not reveal information received from the patient to the family or others unless authorized by the patient to do so, but without revealing any particulars, information obtained from the patient can be verified or refuted by the family. However, the psychiatrist must always bear in mind that even the cognitively impaired patient's account may be the correct one, not the relatives' account. Talking with the family members will help the psychiatrist ascertain if the patient is being scapegoated, neglected, or abused, or became the identified patient for reasons of family dynamics rather than patient psychopathology. The frail elderly patient often depends on the family, just as the frail young child does and the psychiatrist needs to be as alert to elder abuse as to child abuse.

Growing numbers of elderly adults are residing in retirement hotels, board and care homes, or nursing care facilities, depending on their level of functioning and type of impairment. The transition from independent living in their own apartments or homes or living with family members to such other facilities is a significant source of stress and tends to create intrafamily conflicts. Sometimes adult children decide on the move without the patient's full involvement or even awareness. Subsequently, a wide range of negative emotional reactions can be seen in both the patient and the family. Patients feel rejected, abandoned, manipulated, and taken advantage of by the family, become angry, and blame the family for their declining health and depression. Even if there were no other options, family members frequently feel guilty, helpless, frustrated, or angry for having made this decision, and they can turn their anger and frustration against the patient as well as against treating physicians and facility administrators, demanding better care no matter how good the care is. Many of these processes tend to be revealed in separate sessions with the patient and family.

Patient With the Family or Significant Others The patient returns for the third portion of the session, and the psychiatrist gets an opportunity to gauge both the interaction between family members in the presence of the patient and the interaction between the patient and the others. Further, the psychiatrist can discuss issues about which contradictory information was given, always respecting confidentiality. Thus, for example, disagreement over driving (e.g., patient insists on driving

despite poor eyesight or tendency to get lost) can be elicited, and the psychiatrist can get an idea of the family dynamics and approach a realistic estimate of the disputed facts.

Patient Alone Again During the fourth and last portion of the session, the patient is again seen alone. The psychiatrist can now reevaluate the patient in light of the additional information obtained, and the patient has another opportunity to bring up matters of concern and to clarify interactions that took place with the others. Also, both patient and family can see that the attention has been focused on the patient. (Depending on the particular family situation, subsequent visits may have to focus on one or more family members or the family as a unit; however, this should not be done in a way that excludes or neglects the patient.)

Many geriatric patients are plagued by the fear of going crazy and will ask the psychiatrist at the end of the visit, “Am I losing my mind?” or some variant thereof. Instead of bland denial (which the psychiatrist may mistakenly consider reassurance), these patients need an opportunity to describe what they mean by going crazy and what leads them to believe that it may happening to them. Although the geriatric patient may not always find consolation in the distinction, “It is not your age, it is your hypertension that's slowing you down,” there are clear differences in potential treatment strategies. Treatment of hypertension or diabetes may indeed improve mental functioning. Symptomatic treatment can restore the individual to a prior level of functioning—physiological, psychomotor, social, and occupational. If the psychiatrist adopts an optimistic stance, it will encourage the patient and the family and will help maintain rapport and further the therapeutic relationship. At the same time, one must consider the patient's mood and outlook and be realistic in treatment outcome expectations. A profoundly depressed 85-year-old person with feelings of hopelessness and helplessness is likely to perceive exuberant optimism and anticipation of a quick cure as a lack of understanding and professional skill. Regardless of the patient's age, the skilled psychiatrist treads a thin line between reasonable expectations based on realistic assessment and the detrimental effects of dashing the patient's hopes in the mistaken belief that brutal bluntness is the only honest way.

HISTORY

The psychiatric history includes the identifying data (i.e., name, age, sex, marital status); chief complaint; history of present illness; past psychiatric problems; family, medical, and psychosocial history; and information about the patient's habits, particularly any substance abuse or medication misuse.

History of Present Illness The history of present illness usually includes a description of the temporal sequence of symptom development, the relation to daily activities, medication usage, and treatment outcome. The accepted teaching that the history is the most important element in the diagnosis and that “unless you think of it first, you won't get the information” is as true in elderly patients as in younger ones. One should know about differences in presenting symptoms between old and young patients; for example in depression, somatization tends to be more common in older patients and expressed feelings of guilt and self-reproach more frequent in the young.

The psychiatrist must also realize that many older patients may have adjusted to a variety of symptoms, learned to tolerate them, and even accept them as part of ordinary existence, so that they may not consider them important enough to mention. This applies to both psychiatric and somatic symptoms. They may also dismiss somatic symptoms, such as insomnia, anorexia, and general malaise as part of normal aging. Finally, medications may have become so routine that they are not mentioned (e.g., thyroid preparations that patients have taken for decades).

The contribution of life events to disease development needs to be carefully examined. At a time when the body's adaptive physiology, immune system, and other host defenses are compromised by the ravages of old age, life stresses such as loss of family and friends, loss of income, and loss of status and social supports tend to mount. Death of a spouse is considered one of the highest stressors, and clinical experience suggests that death of an adult child is among the most stressful events for older patients; parents do not expect to outlive their children. Minor stresses (e.g., a change in daily routine) may contribute to the decompensation of the fragile stability of geriatric patients.

The older patient is clearly more vulnerable than the younger adult to the stress of hospital admission. Thus, delirium is a frequent complication in elderly individuals but not in younger adults. In a prospective study of the adverse consequences of hospitalization, 30 percent of those over age 70 (vs. 3.6 percent of those under age 70) exhibited confusion within 6 weeks of their admission to general medical wards. Reports from psychiatric wards have given frequencies of delirium between 10 and 40 percent. The mnemonic *sundowners* summarizes the risk factors for delirium: *s*ick, *u*rinary retention/fecal impaction, *n*ew environment, *d*emented, *o*ld, *w*rithing in pain, *n*ot adequately evaluated, *e*yes and ears, *r*x–therapeutic drug intoxication, *s*leep deprived.

In frail elderly individuals, transition from home to an institution may become life's last tragedy. Careful interview reveals what kind of losses and stresses this change of environment brings to the older patient: loss of independence (the most frequent complaint); feeling unnatural among only old and sick people; being constantly reminded about upcoming death; humiliation of being treated by staff as a kid or a demented person; and giving up driving.

In some institutions elderly persons with different levels of functioning reside in close proximity to each other. For example, when board and care and nursing homes (usually owned by the same agency) are located on the same premises and use the same dining and activities rooms, multiple daily contacts are the rule. In this situation, many high-functioning individuals feel depressed, frightened, and overwhelmed by constant interaction with those who are severely impaired mentally or physically, moribund, or bemoaning their fate. Understanding such stresses may help one uncover the source of depressive symptoms and lead to improved treatment.

Past Psychiatric History Psychiatric history should include careful documentation of dates as well as description of specific psychiatric symptoms, details of treatment modalities (e.g., names and dosages of drugs; bilateral or unilateral, as well as number of electroconvulsive treatments [ECTs]), treatment outcomes, remissions, and exacerbations. Often, documentation is unavailable because old records have been destroyed or lost, and the psychiatrist may have to resort to best guesses.

Personal and Psychosocial History As in the middle-aged and younger adult, the personal history indicates the level of premorbid functioning and provides clues to the patient's coping strategies. Knowing premorbid personality characteristics may clarify current psychopathological symptoms. The stability of personality throughout life remains controversial. Some assert that personality does not change with age, although certain traits may become exaggerated. Thus, the well-adjusted, assertive, liberal 30-year-old is likely to turn into the well-adjusted, assertive, radical, 80-year-old, with Benjamin Spock the prototypical example. The hostile, unfriendly 30-year-old may become the even more hostile, crotchety, misanthropic 80-year-old. Above all, the psychiatrist needs to appreciate the range of individual differences, which generally increases in old age.

The personal history may also indicate how the patient is likely to adapt to certain losses. For example, energetic, independent, ambitious individuals may find it more difficult to accept physical limitation and dependency than their dependent, relaxed, leisure-oriented, age-matched peers. Early life history and developmental data may shed light on the patient's interpersonal relations when they must depend upon others, particularly at a time of life when dependence upon others looms large again. Religion is often of major importance to older patients, and religious beliefs may contribute to the patient's ongoing problems or ameliorate them.

Sexual History The geriatric patient usually has been married and, if a man, likely is married. Psychiatrists are often reluctant to take a sexual history from a geriatric patient because they identify the patient with a parent or a grandparent or they subscribe to the popular stereotype of old persons as asexual. The geriatric patient, however, may be eager to discuss sexual dysfunctions (e.g., increased time needed to achieve erection or ability to achieve only partial erection in the man, physical discomfort due to vaginitis in the woman, performance anxiety in the newly remarried man, inability to satisfy sexual urges in the widow without a sexual partner, extramarital activity in the spouse of a demented or physically incapacitated patient). At times new onset of neurosis in old age may result. Even though sexual history needs to be taken in the privacy of the doctor-patient dyad, information from significant others is also needed, not only for corroboration of the patient's report but also because they reveal, for example, extramarital affairs or aberrant sexual interests and activities (e.g., paraphilias) that the patient failed to mention. The inquiry may also open subjects for discussion by family members who were reluctant to bring them up, whether in the realm of physiological, psychological, or social functioning. Many children, for example, find it unacceptable when their parent has affairs or when the 83-year-old mother decides to move in with her male friend of many years' standing—without benefit of marriage. A decision to get married may carry equal disapproval!

Family History The older the patient, the more difficult it is to verify the family history unless such documents as family bibles, diaries, and other vital records are available. Patients are usually the oldest surviving family members, so there is no one in their own or preceding generations who can contribute. Succeeding generations are generally not well informed about their ancestors, unless someone has taken a special interest in the family history. Clearly, family history is important in genetic disorders such as Alzheimer's, Creutzfeldt-Jakob, Gerstmann-Straussler-Scheinker, Huntington's, and Pick's diseases. Elderly persons are at high risk for suicide, especially white men, but information is inadequate to assess the importance of prior suicides in the family.

Medical History Elderly patients have more concomitant chronic and multiple medical problems and take more medications than younger adults; many of these medications can influence their mental status. The past medical history includes all major illnesses, traumata, hospitalizations, and treatment interventions. The psychiatrist should also be alert to underlying medical illness. Infections, metabolic and electrolyte disturbances, and myocardial infarction and stroke may first be manifested by psychiatric symptoms. Depressed mood, delusions, and hallucinations may precede other symptoms of Parkinson's disease by many months. On the other hand, psychiatric disorder can also be the cause of such somatic symptoms as weight loss, malnutrition, and inanition of severe depression.

Careful review of medications (including over-the-counter medications, laxatives, vitamins, tonics, and lotions) and even substances recently discontinued is extremely important. Drug effects may be long lasting and may induce depression (e.g., antihypertensives), cognitive impairment (e.g., sedatives), delirium (e.g., anticholinergics), or seizures (e.g., neuroleptics). Review of medications must include sufficient detail to identify misuse (overuse, underuse) and relate medication use to special diets. A dietary history is also important; deficiencies and excesses (e.g., protein, vitamins) may influence physiological function and mental status.

History of Alcohol and Substance Use and Abuse One must inquire into alcohol and substance use and abuse. Geriatric patients—even women in their 70s and 80s who only used alcohol socially throughout life—may develop serious alcohol problems in old age. Lifelong alcohol abusers and those who stopped or markedly decreased their alcohol intake before age 65 may exhibit residual physical complications from their past alcoholism. Data on substances other than alcohol are notably absent. Some prescription medications (e.g., benzodiazepines, other sedatives and hypnotics) are often misused and overused by geriatric patients who come to clinical attention.

MENTAL STATUS EXAMINATION

The goal of the mental status examination is to obtain precise, detailed information about the patient's functioning—how the patient behaves, feels, and thinks. Geriatric patients may require serial mental status examinations to evaluate fluctuating and changing behavior, mood, and ideation.

General Description General observations of the geriatric patient are those used with patients of any age, that is, appearance, behavior, functioning, motor activity, speech, and attitude toward the examination and the examiner.

Appearance Caution is indicated with regard to appearance. The disheveled, cachectic, unkempt patient, smelling of urine, is readily diagnosed as demented, while the well-groomed, poised, quiet patient is readily categorized as depressed. In actuality, the situation is often the reverse, the appearance of the demented patient being the product of loving care bestowed by a relative or other concerned caregiver. One must also note the presence and absence of hearing aids and corrective lenses to ascertain to what extent the patient suffers from one or more of the sensory deficits so common in old age.

Behavior The patient's attitude toward the examiner (e.g., cooperative, attentive, interested, guarded, suspicious, defensive, apathetic, withdrawn) needs to be documented. Disturbances of motor activity should be described (e.g., body asymmetry, unsteady gait, retarded movements, stooped posture, hyperactivity, mannerisms, combativeness, tremor, involuntary movements of the mouth and tongue). Many depressed, apathetic, and demented patients may exhibit slow movements and speech. Masklike facies, resting tremor, and pill-rolling movements may indicate Parkinson's disease or adverse effects of neuroleptics.

Speech and Language Quality, quantity, and fluency of speech need to be recorded. Speech may be described as normally responsive, rapid, pressured, loud, or emotional; it can be slow, monotonous, unspontaneous, slurred, or mumbled. Mutism, excessive latency of response, paucity of speech, anomia, agrammatism, and paraphasic errors are all important symptoms for diagnostic assessment.

A patient's language reflects thought processes and may reveal verbal perseveration, circumstantiality, loosening of associations, and other disturbances. Different types of aphasia, often seen in elderly patients with a history of strokes, may lead to the mistaken classification of the aphasic patient as demented and should always be taken into consideration.

Mood and Affect Elderly patients are generally less psychologically minded than younger adults and experience greater difficulty in describing their mood and affect. Repeated inquiries about mood are often appropriate, since denial and suppression of affect are prevalent in geriatric depressed patients, who may say, “I don't feel sad, I'm not down” as they cry. Some patients who deny experiencing sadness and depressed mood do describe the somatic and vegetative symptoms of depression (i.e., significant changes in appetite and weight, insomnia or hypersomnia, fatigue, and diurnal variations in general well-being) and lack interest in family affairs and wordly events. Somatization is common, for example, in the patient who attributes to abdominal pain all somatic and dysphoric symptoms: “If only you could help me with my abdominal pain, or my back pain, or... I would be perfectly well and have no problems.”

Some standardized scales may be helpful in documenting the symptoms of depression and estimating severity (e.g., Hamilton Depression Scale or Geriatric Depression Scale [Table 51.2a-1]). The psychiatrist should ask the patient directly about suicidal ideas—passive death wishes—and then about intent, plans, and definite methods. Past suicide attempts increase the risk of a new and successful suicide attempt.

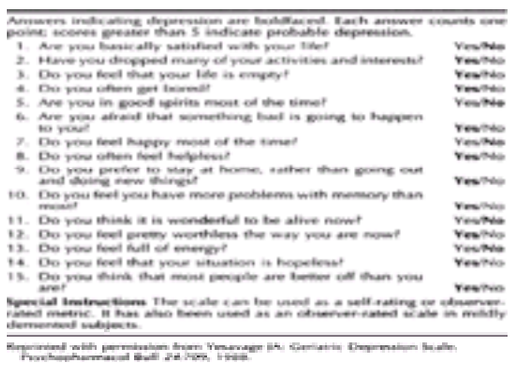

Answers indicating depression are boldfaced. Each answer counts one point; scores greater than 5 indicate probable depression.

1. Are you basically satisfied with your life? **Yes**/**No**
2. Have you dropped many of your activities and interests? **Yes**/**No**
3. Do you feel that your life is empty? **Yes**/**No**
4. Do you often get bored? **Yes**/**No**
5. Are you in good spirits most of the time? **Yes**/**No**
6. Are you afraid that something bad is going to happen to you? **Yes**/**No**
7. Do you feel happy most of the time? **Yes**/**No**
8. Do you often feel helpless? **Yes**/**No**
9. Do you prefer to stay at home, rather than going out and doing new things? **Yes**/**No**
10. Do you feel you have more problems with memory than most? **Yes**/**No**
11. Do you think it is wonderful to be alive now? **Yes**/**No**
12. Do you feel pretty worthless the way you are now? **Yes**/**No**
13. Do you feel full of energy? **Yes**/**No**
14. Do you feel that your situation is hopeless? **Yes**/**No**
15. Do you think that most people are better off than you are? **Yes**/**No**

Special Instructions The scale can be used as a self-rating or observer-rated metric. It has also been used as an observer-rated scale in mildly demented subjects.

Reprinted with permission from Yesavage JA: Geriatric Depression Scale. Psychopharmacol Bull 24:709, 1988.

Table 51.2a-1 Geriatric Depression Scale (Short Version)

The examiner must record the predominant affect during the interview. An incongruous, labile affect that changes readily from laughing to crying may indicate cerebrovascular disease, vascular dementia, or other brain damage.

Perception Hallucinations vary from simple and unformed to highly complex and organized. They may be transitory and associated with vision and hearing impairment in the elderly patient. They also often occur in patients with organic mental disorder and may be associated with alterations of consciousness, disorientation, or confusion. Some hallucinations indicate focal brain or peripheral perceptual organ pathology, and their appropriate description may lead to an accurate diagnosis. For example, focal lesions in the visual pathway are often associated with release visual hallucinations, complex formed images that tend to occur in the area of field deficit.

Other types of perceptual impairment that may be found in older patients with organic brain disorders include visual agnosia (inability to recognize visible objects); prosopagnosia (inability to recognize familiar faces); astereognosis (inability to recognize objects by touch); autotopagnosia (inability to recognize parts of one's own body); micropsia (objects seem smaller than their real size), or macropsia (objects seem larger than their real size).

Thought Process and Content As with younger patients, this section of the examination deals with the logic and coherence of thought processes. Can the patient understand and answer questions appropriately? Is thinking concrete and literal? What are the patient's preoccupations or obsessions? In depressed elderly patients preoccupation with health problems often replaces the complaint of depressed mood.

The content of any delusions should be described in detail, and whenever possible, in the patient's own words. Are they persecutory, guilty, nihilistic, somatic, erotic,

grandiose? Are they bizarre and diffuse or focused on the immediate environment? Are they stable or transient? What is their impact on the patient's life?

Elderly individuals often display significant suspiciousness. They become concerned about external forces or individuals controlling their lives. These concerns may be diffuse or (more often) focus upon a limited number of targets, usually persons in the patient's closest environment—children, other relatives, neighbors, landlords. Ideas that things are being stolen are very common and are generally interpreted as a response to the losses accompanying old age, when beauty, youth, health, and vigor are indeed being stolen. Sometimes, however, these ideas reflect reality rather than delusional thinking, and clothing, jewelry, or cash have actually been stolen.

Hypochondrial delusions often center around the functions of the bowel or brain. Patients may believe that their bowels are paralyzed or completely destroyed so they cannot digest any food. Use of rating scales (e.g., University of Washington Paranoia Scale) may help define the range and severity of the delusional disorder.

Sensorium and Cognition Because of the wide prevalence of memory and cognitive impairment, this part of the mental status examination is specially important with geriatric patients. The psychiatrist can learn a lot about the patient's cognitive functioning throughout the entire interview. Careful listening and asking specific questions should enable the psychiatrist to comment on level of alertness and consciousness, attention and concentration, orientation, memory (recent, remote), language (spontaneous, naming, comprehension), constructional abilities and praxis, and frontal lobe tasks.

Impaired consciousness needs to be assessed first, since it may preclude obtaining other information. Alterations of consciousness are characteristic of organic mental disorder in elderly persons; patients are not alert, display fluctuations in level of awareness, and are somnolent or lethargic. Disorientation to time, place, person, and situation may be present in patients with impairment of consciousness but also may occur in those with cognitive deficits without impaired consciousness.

Memory—immediate, recent, and remote—is evaluated by asking the patient about recent and past life events and by administering special tests. A popular way to test immediate recall is to give the patient three items or six digits to repeat forward and backward. Errors can reflect attentional or memory deficits. Impaired performance may occur with depression, anxiety, and agitation. Patients with attentional disorders also perform poorly in tests of verbal fluency. For example, they enumerate only a few items when asked to name everything that can be found in a supermarket or as many words as they can think of beginning with the letter ‘K'.

Recent memory, the ability to learn and recall new information; may be elicited in various ways. One might ask patients to repeat the three items they were told to remember 5 minutes earlier or to repeat a short story told 5 to 10 minutes earlier. Orientation to time and place and questions about events of that morning or the day before help to explore recent memory; it tends to deteriorate early in Alzheimer's disease and other dementing processes. Confabulations may also be revealed during such testing. Remote memory can be tested by asking about significant events in the patient's life (e.g., birthdates of children, years of graduation from school, date of retirement) or well-known historical events (e.g., wars, holidays). One must keep in mind the premorbid state of this individual's memory, and cross-checking information is advised. Age-related changes may also impair memory, but cross-sectional and longitudinal studies have shown that such decline in memory, learning, and problem solving is not substantial for groups of normal individuals until they are well past their 70s.

When testing intellectual functioning and general knowledge, educational level and social status are considered. Vocabulary, complexity of concepts that the patient uses, knowledge of current political, sporting, and cultural events, as well as names of the president, vice president, and governor may provide a preliminary estimate of cognitive functions. Testing reading and writing abilities and clock drawing may also reveal memory and speech deficits.

Constructional ability can be tested by asking the patient to draw or copy a circle, a cross, a cube, a clock, or other two- or three-dimensional figures. Praxis is tested by asking the patient to perform volitional actions on command; for example, “with your right hand show me how you would comb your hair,” “drink through a straw,” “hammer a nail,” “stand as a boxer.”

The Mini-Mental State Examination (MMSE) is a simple tool for general assessment of cognitive functions. Other standardized scales currently used in evaluating geriatric patients include the Blessed Dementia Index, the Progressive Deterioration Scale (PDS), and the Alzheimer's Disease Assessment Scale (ADAS) (Table 51.2a-2).

Table 51.2a-2 The Alzheimer's Disease Assessment Scale (ADAS)

Frontal system tasks are used to reveal frontal lobe dysfunction. The frontal lobes are involved in recruiting and engaging attention, inhibiting inappropriate behavior, and initiating, sequencing, and planning complex behaviors. Disturbance of attention and initiation and apathy and withdrawal may be found in patients with lesions of the frontal lobe. Disinhibition of aggression and sexuality, impairment of planning functions, and concreteness are characteristic of patients with lesions in orbital areas of the frontal lobe. Tests for perseveration (the pathological repetition of speech, actions) and concrete thinking (e.g., similarities, metaphorical speech, proverbs) are often used to reveal frontal lobe dysfunction.

Insight and Judgment The ability of patients to understand the situation (e.g., their own disease, its causes and symptoms, and reasons for examination) can clarify their level of insight. Judgment includes the ability to act appropriately, in accordance with a given situation. It can be assessed by asking about the patient's proposed action in a specific hypothetical situation.

Functional Assessment Functional status is an important part of assessment in geriatric psychiatry. It includes characteristics of everyday behavior and activities in the community and at home. Assessing how well an elderly person maintains independence and performs activities such as dressing, grooming, eating, toileting, and preparing food, enhances the psychiatrist's knowledge of the patient's competence in dealing with the tasks of daily life.

Table 51.2a-3 lists characteristics of physical and instrumental activities of daily living. This scale typifies those now widely used in geriatric psychiatry for functional assessment of the elderly patient. It may be completed by caregivers even before the formal interview, to give important information about the level of the patient's functioning. Later corroborative evidence from the patient, and if indicated, from others, then allows psychiatrists to form their own judgment.

Table 51.2a-3 Functional Assessment

Competence The assessment of competence stands in close relationship to the mental status examination. Can the patient give consent for medical treatment, give power of attorney, sign or change a will, manage personal finances? Information gained throughout the interview should help the psychiatrist to answer these questions. Is the patient aware of the illness and the need for treatment? How much does the patient's memory impairment interfere with the ability to remember financial transactions? Is judgment so impaired that the patient's finances are in jeopardy? Such issues may assume substantial legal importance and may be why a psychiatric evaluation of the older person was requested.

SUMMARY, DIAGNOSIS, RECOMMENDATIONS

The psychiatric examination concludes with a short summary of the major findings and differential diagnostic considerations. These data yield the rationale for the initial diagnosis and recommended treatment strategies.

Mr. S. is an 89-year-old white, retired businessman, married, living with his 90-year-old wife in a board and care home located in a well-designed and well-equipped retirement complex that holds several board and care and nursing homes on the same park territory. Mr. S. was referred for psychiatric evaluation for symptoms of depression, social isolation, disinterest in his usual activities, and poor sleep and appetite.

Mr. S. has no previous psychiatric history. He is described by his family as an active, enthusiastic, and warm person. He moved to this retirement complex approximately 7 months prior to his first clinic appointment from his own condominium, where he and his wife had been residing for 20 years. Serious damage to the condominium building from the 1994 Northridge earthquake had necessitated having all residents move elsewhere for 1 to 2 years during reconstruction. The couple's four children decided that placement in a comfortable retirement setting with medical and nursing care available would be better for their parents than relocation to a new rented apartment. Their decision was based on their parent's advanced age, their mother's declining physical health and memory, and their father's inability to take care of his wife and household. The patient initially resisted his children's recommendation, but his wife agreed, and finally he too decided to accept it. Mr. S. and his wife moved into separate rooms on the same floor, at the patient's request. They had always had different interests; his wife would spend many hours a day watching TV or talking on the phone to her friends, and he preferred to work on his computer or write notes for his autobiography, which he has been planning for many years. He felt they both would be most comfortable to be next to each other but in separate rooms.

One month after they moved to the complex, Mr. S. became depressed. He ruminated about the “mistake I made by moving to an institution,” which entailed loss of his condominium, financial independence, car, activities in clubs and volunteer organizations, and daily meetings with friends. More important, he lost self-respect; he felt that from now on he would be treated as an “institutionalized,” potentially demented person, dependent on the board and care staff for care and decision making. He felt “unnatural” being among only elderly people, many of whom were ill, with impaired mobility and cognition. This forced him to acknowledge his own age and the possibility of similar impairments and death in the near future. He was angry with his children who, he felt, had used his wife's declining intellectual capacity to force him to move here. His children were unwilling to discuss an alternative living situation, such as an apartment with a 24-hour caregiver for his wife.

Mr. S. described the situation: “I stayed in my room, trying to work on my computer and write, but I could not concentrate or enjoy it. I no longer was interested in doing volunteer work. I couldn't sleep and began losing weight (18 pounds by the time of his first appointment, from an initial weight of 168 pounds). Finally, I decided to kill myself. It was a difficult decision. I have a big family: 4 children, 13 grandchildren, and 12 great-grandchildren. I knew that my suicide would be a tragic blow to my family. I struggled with this decision, constantly reviewing in my mind all the pros and cons. I decided that I had only two options: death or long-lasting suffering.”

Mr. S. worked out a plan. He invited all his family for a four-day cruise. He would see and hug them all for the last time and, upon return, would obtain a gun and shoot himself. He made necessary preparations, paid all the family's cruise expenses, and wrote letters or called his grandchildren living out of state to invite them for this “good-bye trip.” Everybody came. He almost revealed his plan to one of his grandsons who always was close to him and seemed to understand him better than others. This grandson shared his concern about his grandfather's depression and possible suicidal thoughts with two of Mr. S.'s children. However, it was not taken seriously.

After the trip the patient became even more depressed, isolating himself in his room and avoiding any activities. During his regular meeting with the social worker he was tearful, unable to discuss any issues. At this time he was referred for psychiatric consultation.

Mr. S. had never before visited a psychiatrist. He always was very active, enthusiastic, productive in his business, warm with his family and friends. In the last 15 to 20 years he had had difficulty sleeping but refused to take any medications and preferred to work on his computer several hours every night until he felt tired and able to fall asleep.

History of psychiatric disorder in the family included reports of depression in one of the patient's brothers and some emotional instability and increased level of anxiety in two of his grandsons.

Prior surgical history included an appendectomy at age 21 and hernia repair at age 29. Medical history included pneumonia at age 65 and a recent diagnosis of degenerative disease of the lower spine. The patient has partial vision loss. He takes aspirin (81 mg once a day) for cardiovascular prophylaxis. He takes Tylenol (65 mg once or twice a day as needed) for back pain and headaches. The patient denied any history of substance abuse. He drinks wine socially, one to two drinks every 2 or 3 weeks. He never smoked.

Mr. S. was born and raised in Chicago, the youngest of three brothers. His father died of heart disease at age 78 and his mother of cancer at age 83. One brother died of brain tumor at age 69, and one brother is still alive. After graduating from high school, Mr. S. worked for a management company in Chicago. He married his current wife at age 27 and moved to Los Angeles, where he started his own business managing office buildings. He was relatively successful and financially independent. The couple had two sons and two daughters, all college educated. According to Mr. S. and his children, the marriage was not a very happy one, but the couple never considered divorce. They had close relationships with their children, grandchildren, and great-grandchildren, many of whom they supported financially, even recently. Mr. S. enjoyed his club memberships, activities in charitable organizations, and his many friends.

MENTAL STATUS EXAMINATION

Appearance. The patient was an 89-year-old man, neatly and casually dressed, with a sad facial expression and minimal body movements.

Behavior. Eye contact was poor, as patient primarily stared at his hands or shoes. Although he initially resisted discussing his “private and difficult situation which nobody can understand or change,” he became more open and talkative as the interview progressed, seeking understanding of his position and decision.

Speech. Was soft and slow initially, but toward the end it reached almost normal volume and speed.

Mood. Mr. S. was depressed and stated “I've lived enough.” He did not want to become demented and sick before dying of natural causes; he wanted to make his own decisions. He felt that his life had become meaningless and his days “long and empty” since moving to his current residence. He had lost his interests and had no energy to get up in the morning or eat. He showed his “good-bye” note: “Last year has been too hard for me. I had enough. I love you all.” He revealed his suicide plan. Hamilton Rating Scale for Depression score was 33, which indicates serious depression.

Affect. Had decreased range.

Neurovegetative symptoms. Mr. S. had no energy to get up in the morning or to eat. He complained of slowed thinking and poor concentration. His movements were slowed.

Perceptual disturbances. Mr. S. denied auditory, visual, and other hallucinations. He said he felt detached from his environment, which had become less bright and real for him.

Thought processes. Were logical and goal directed. He complained that his thoughts were slow and that it was “hard to think.”

Thought content. Mr. S. had active suicidal ideation, with a specific plan for killing himself. He denied delusions or homicidal ideation. He was focused on the change in his living situation, his current dependence on other people's decisions, and control by the “institution's rules and regulations.” He was distressed by changes in his sleeping and eating routines imposed by the retirement home—such as rising at 6 AM to be ready for breakfast at 7—as well as by his association at mealtimes with other residents whom he did not know, many of whom were cognitively impaired. “It is what I am going to be soon, if I survive.”

Sensorium. The patient was alert and oriented to time, place, and person.

Attention. He repeated five of six digits forward and four backward and performed serial 7's with one error.

Memory. The patient's memory was not significantly impaired. He had some difficulties with concentration and learning new material. However, he did remember most important dates of his life; names of his grandchildren and great-grandchildren; their education, occupations, and achievements; names of his recent doctors; and his family's visits. He recalled all three objects after 1 and 5 minutes.

Abstract thinking. He had good abstract thinking. He had no difficulty explaining similarities or proverbs.

Knowledge and intelligence. The patient had a good fund of knowledge. His Mini-Mental State Examination score was 29/30; he made one mistake in serial 7s.

Learning. The patient was able to learn 4 of 10 words after the first and second trials, 6 words after three trials, and 7 after four trials.

Insight and judgment. Were fair. The patient believed he had no mental problems that could be treated and that he had reason to be depressed secondary to multiple losses and complete change in his life. He did not realize his growing limitations and stated that he could still drive a car (even with his impaired vision), live with his wife in an apartment, and take care of her with help of a part-time housekeeper (which was not realistic in light of his physical limitations). However, he was interested in finding out if there was something that could help him with his depressed mood and constant thinking about suicide (note the contradiction with his statement that he had no mental problems to be treated).

Impulse control. Was preserved despite intrusive suicidal thoughts. He could continuously control them, trying to find solutions, talking to his children, and viewing different perspectives.

FUNCTIONAL ASSESSMENT

The patient remained independent in most activities of daily living. His limitations in driving and reading were related to declining vision. Difficulties lifting were due to frequent back pain. Loss of interests and energy were due to depression. There was no change in physical self-maintenance scale (Table 51.2a-3).

Physical examination and laboratory tests revealed only mild benign prostate hypertrophy, degenerative joint disease in the lower spine; impaired vision; and increased concentrations of cholesterol (248), low-density lipoprotein (LDL, 212) and high-density lipoprotein (HDL, 33).

Discussion This 89-year-old patient with no prior psychiatric history meets fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) criteria for major depressive disorder with approximately 5-month history of increasingly depressed mood; loss of interest and pleasure in usual activities; feelings of hopelessness and helplessness; and serious suicidal ideation and plans; as well as low energy, insomnia, poor appetite, and weight loss. An interesting feature of this depression is that it developed late in life and was precipitated by placement at a board and care facility, which the patient considered a dramatic loss of the major values of his life—independence, control, self-respect, and the ability to be productive and give help rather than to get it from others.

This patient did not have significant cognitive deficits or medical problems, which ruled out such diagnoses as dementia with mood disorder or mood disorder due to medical condition. Differential diagnosis includes adjustment disorder with depressed mood.

WORKING DIAGNOSIS

Axis I Major depressive disorder, single episode, severe, without psychotic features

Axis II None

Axis III Medical; benign prostatic hypertrophy, degenerative spine disease

Axis IV Severe psychosocial and environmental problems; change of environment due to institutional placement

Axis V Global assessment of functioning, 23; last year, 75

TREATMENT AND OUTCOME

The patient was initially seen daily to ensure his safety. A combination of medication and psychotherapy (individual and family therapy) was used to treat this patient, with the patient confirming that he would not attempt suicide. Treatment was started with paroxetine (Paxil) for depression (10 mg a day; after 3 days the dosage was increased to 20 mg a day) and zolpidem (Ambien) 5 mg at bedtime for insomnia with difficulty initiating sleep. Sessions with the patient alone, family alone, and patient with family helped to educate them about depressive disorders, suicide, current concepts, and treatments and to clarify family behavior that made Mr. S. believe that he was “forced, placed” without consulting with him.

As a result, the patient was given the right to reconsider his current living situation, see other places, including separate apartments with assisted living, and make his own choice. The patient's children took him to these places, discussed with him all pros and cons, and explained their reasons for believing that the current living arrangement was the best for both him and his wife. He was given time and responsibility for making the final decision. The patient finally concluded that he could not find a better choice and decided to stay with the current arrangement. His depression gradually resolved; he became more active and joined volunteer organizations in his retirement complex. He remained on maintenance treatment with paroxetine (20 mg a day) for approximately 6 months, then he was weaned from this medication. At 5-months follow-up he remained stable.

SUGGESTED CROSS-REFERENCES

The psychiatric examination is discussed in Chapter 7. Dementia and Alzheimer's disease are discussed in Chapter 10 and Section 51.3e, substance-related disorders in Chapter 11 and Section 51.3h, and psychiatric problems in the medically ill in Section 51.3a.

SECTION REFERENCES

American Psychiatric Association: *Practice Guidelines*, ed 1. American Psychiatric Association, Washington, DC, 1996.

Appelbaum PS, Gutheil TG: *Clinical Handbook of Psychiatry and the Law*, ed 2. Williams & Wilkins, Baltimore, 1991.

Benedict RM, Goldstein MZ, Dobraski M, Tannenhaus J: Neuropsychological predictors of adaptive kitchen behavior in geriatric psychiatry inpatients. J Geriatr Psychol Neurol *10:*146, 1997.

*Berkman B, Chauncey S, Holmes W, Daniels A, Bonander E, Sampson S, Robinson M: Standardized screening of elderly patients' needs for social work assessment in primary care: Use of the SF-36. Health Soc Work *24:*77, 1999.

Blessed G, Tomlinson BE, Roth M: The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects. Br J Psychiatry *114:*797, 1968.

Brocklehurst JC, editor: *Textbook of Geriatric Medicine and Gerontology*, ed 4. Churchill Livingstone, Edinburgh, 1992.

Coffey CE, Cummings JL, editors: *Textbook of Geriatric Neuropsychiatry*. American Psychiatric Press, Washington, DC, 1994.

Folstein MF, Folstein ME, McHugh PR: Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. J Psychiatry Res *12:*189, 1975.

La Rue A: *Aging and Neuropsychological Assessment*. Plenum, New York, 1992.

Lawton MP, Brody EM: Assessment of older people: Self-maintaining and instrumental activities of daily living. Gerontologist *9:*179, 1969.

Levkoff SE, Besdine RW, Wetle T: Acute confusional states (delirium) in the hospitalized elderly. In *Annual Review of Gerontology and Geriatrics*, vol 6, C Eisdorfer, editor. Springer-Verlag, New York, 1986.

Post F: Depression, alcoholism, and other functional syndromes. In *Psychogeriatrics: An International Handbook*, M Bergener, editor. Springer-Verlag, New York, 1987.

Rosen WG, Mohs RC, Davis KL: A new rating scale for Alzheimer's disease. Am J Psychiatry *141:*1356, 1984.

Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A: Pathological verification of ischemia score in differentiation of dementias. Ann Neurol *7:*486, 1980.

Sadavoy J, Lazarus LW, Jarvik LF, Grossberg GT, editors: *Comprehensive Review of Geriatric Psychiatry-II*, ed 2. American Psychiatric Press, Washington, DC, 1996.

Shulman KI, Shedletsky R, Silver IL: The challenge of time: Clock-drawing and cognitive function in the elderly. Int J Geriatr Psvchiatry *1:*135, 1986.

Streim JE, Oslin D, Katz IR, Parmalee PA: Lessons from geriatric psychiatry in the long-term care setting. Psychiatr Q *68:*281, 1997.

Weiner MF, editor: *The Dementias: Diagnosis, Management and Research*, ed 2. American Psychiatric Press, Washington, DC, 1996.

Yesavage JA: Geriatric Depression Scale. Psychopharmacol Bull *24:*709, 1988.

*Yousef C, Ryan WJ, Lambert T, Pitt B, Kellett J: A preliminary report: A new scale to identify the pseudodementia syndrome. Int J Geriatr Psychiatry *13:*389, 1998.

51.2 ASSESSMENT

51.2B CENTRAL NERVOUS SYSTEM CHANGES WITH NORMAL AGING

JEFF VICTOROFF, M.D.

An understanding of the expected impact of aging on the central nervous system (CNS) is important to geriatric psychiatry. Such an understanding alerts psychiatrists to typical changes in neurological and neurochemical function that may impose or alter senescent vulnerability to psychiatric disorders and response to medications and leads them to consider the role of sensory and signal-processing decrements in psychiatric disease. Ultimately, an understanding of the neurobiology of normal aging may provide insight into the etiology of mental illness among aged individuals, not only by determining the changing neural substrate of mental life but also by explaining how altered capacity and strategies for adapting to the environment may change the impact of psychological stresses on these individuals.

Aging is often conceptualized as a final stage of life, marked by deterioration from a prior level of functioning. However, several alternate concepts may be much more useful in considering the changes in the CNS associated with normal aging. First, although many aspects of neurobiology clearly change over the life span, there is no simple dichotomy of age versus youth, since a continuum exists for most structures and functions across the age spectrum. Second, there are great individual differences in the impact of aging and cross-sectional studies of the aging nervous system show a large variance in values at any given age. Third, aging of the nervous system can be seen not as a late stage or a critical threshold, but as a feature of development that begins with conception and unfolds throughout the life span. Fourth, many age-related aspects of physiology (e.g., oxidative stress, excitatory neurotoxicity, or glucocorticoid-induced neurotoxicity) might be considered both causes and effects of aging. Fifth, some age-related changes in the nervous system may result from evolutionary processes with neutral or even healthfully adaptive significance for human interaction with the environment. Finally and most critical to the goal of examining the evidence of normal aging of the CNS, neuroscientists are seriously reexamining the boundaries between normal and abnormal aging. At the extremes of the spectrum of neuropathologic findings and behavioral functions it is easy to say that major macroscopic, microscopic, and physiological changes have or have not occurred in a nervous system between sexual maturation and senescence or that, behaviorally, an individual has or has not lost the ability to carry out the activities of daily life. The most difficult single task in discussing the normal aging of the human CNS is specifying a point on any of the multiple continua of age-related changes where it is biologically or philosophically meaningful to say normal ends and abnormal begins. In particular, with regard to the significantly overlapping diseases called neurodegenerative (e.g., Alzheimer's disease and related disorders), despite remarkable progress in untangling their molecular genetics, neurobiology, treatment, and even prevention, arbitrary operational definitions are still widely used to distinguish the effects of aging per se from those of mild disease.

It is currently unknown whether everyone would experience some or all of the cerebral changes that are currently called Alzheimer's disease were they to live long enough, but evidence suggests they would. Thus one must carefully avoid equating age-related normal CNS changes with whatever changes may be typically found among the aged or assuming that age-related neurodegeneration differs from precocious presentation of a nearly universal combination of genetically programed human nervous system change and common environmental influences. Instead, as neuroscientists study the rich evidence of age-related changes in the CNS, they must attempt to critically examine the validity of apparent dichotomies between normal and abnormal neurodegeneration. Overall, three exciting themes have emerged at the very end of the twentieth century: (1) age-related neuronal loss may be less than was formerly thought; (2) typical aging and neurodegeneration blend imperceptibly into one another, such that, for instance, so-called nonfamilial Alzheimer's disease is probably on a continuum with the universal changes expected with aging; and (3) that the no-new-neurons dogma may be incorrect, and perhaps, multipotential stem cells and some capacity for continued division and differentiation of neurons may persist in the adult human brain. Table 51.2b-1 outlines age-related CNS changes with aging.

Gross brain atrophy
Ventricular enlargement
Selective regional neuronal loss
Remodeling of dendrites, axons, and synapses
Appearance of intraneuronal lipofuscin
Selective regional decrease in neurotransmitters and neuropeptides
Selective modification of neurotransmitter metabolism
Possible dysregulation of gaseous neurotransmitters
Glucocorticoid neurotoxicity
Changes in receptors
Changes in neurotrophins
Changes in signal transduction
Impairment of calcium homeostasis
Possible changes in cell cycle regulators (e.g., cyclins)
Possible changes in extracellular matrix proteins (e.g., laminins, proteoglycans)
Probable regional decline in cerebral blood flow
Probable regional decline in cerebral metabolic rate
Appearance of senile plaques and neurofibrillary tangles

Table 51.2b-1 Summary of Age-Related Changes in the CNS

GROSS CHANGES IN THE BRAIN

Brain atrophy is often reported to accompany aging. Neuropathological examination usually reveals loss of cortical tissue and white matter and enlargement of the ventricles. However, as reviewers of this subject have pointed out, there are five problems with interpreting the results of pathological studies of normal aging and brain atrophy. First, since many such studies do not prospectively determine the cognitive status of the subjects, a large pathology series may include a significant number of subjects with dementia or other neurological disorders. Second, because of changes in general health and total body weight, persons born earlier in the century may have had a lower average brain weight than those born later, which might confound age-related measurements of atrophy. Third, normal brain weight varies widely, which may obscure specific effects of aging. Fourth, most large neuropathology series do not measure the brain weight in a standardized way. Finally, and possibly most interesting, currently available estimates reflect average brain changes derived from cohorts born before the discovery of environmental factors that may accelerate or decelerate brain aging. Some brains show no evidence of atrophy. As more is learned about the factors that alter the pace of atrophy and this knowledge is applied earlier in the life course, later cohorts may better reflect the maturation of the normally aging human CNS.

Despite these caveats, several findings have held up consistently. Robert Boyd reported in 1860 that brain weight in males increases from 493 grams at 3 months of age to a peak 1374 grams between 14 and 20 years. By age 80, the brain has lost 90 grams. Subsequent studies have shown a mean loss of 100 grams in both sexes between ages 25 and 70. This represents about 7 percent loss of weight over the life span or a mean loss of about 3 grams a year for both sexes from age 40 to 90. This weight loss may not be linear, since brain atrophy appears to accelerate after age 60. In addition, there may be a sex difference in brain atrophy, with women showing loss by the sixth decade and men showing loss beginning in the seventh decade. There is typically some regional variation in lobar loss, with greater average loss in the frontal lobes and less in the occipitoparietal region. Furthermore, atrophy progresses differently in different types of brain tissue. Some evidence suggests that cortex is lost more rapidly than white matter between ages 20 and 50 and that white matter atrophies more than cortex between ages 70 and 90. Still, this loss of gray and white matter has recently come under new scrutiny. Recent evidence suggests that age-related brain atrophy may be due to much less cortical loss and much more white matter loss than previously thought.

Neuroimaging of Age-Related CNS Atrophy The gross atrophy of the brain can be visualized in the living patient by neuroimaging techniques. Both lobar shrinkage and increase in ventricular size are apparent on computed tomography (CT) scans, although quantifying the changes in the ventricles is much easier than measuring the multiple convolutions of expanded sulci. Changes in ventricular size with age have been demonstrated by linear measurements, planimetric area measurements, and volumetric measurements. Linear measures have shown an 11 to 220 percent difference in the width of the lateral ventricles from the third to the eighth decade. Planimetry results are usually reported as ventricular area divided by brain area at a single CT scan level (ventricular brain ratio [VBR]). The VBR is reported to increase from 2.2 to 14.2 percent over the life span, and in several studies, increased VBR has been inversely correlated with cognitive function. Volumetric measurements are made either by outlining ventricular area and taking into account the thickness of the CT slice or by counting pixels with cerebrospinal fluid density. Studies using these methods have consistently reported an age-related, progressive increase in ventricular volume. These changes with age may not occur in a linear progression; most studies report a relatively stable ventricular volume during young adulthood, followed by a rapid increase that has been variously reported to occur in the 40s, 50s, or 60s. Evidence from both CT and magnetic resonance imaging (MRI) studies suggests that men undergo more age-related brain atrophy than women. One investigator reported that men experience this increase in ventricular volume at a younger age than women. Studies of the relation of cognitive status to ventricular volume do not permit definite conclusions, and again, recent MRI studies suggest that most of the age-related atrophy of the cerebrum occurs in white matter, not in cortex.

NEURONAL LOSS

Since the publication of a classic paper by Harold Brody in 1955, it has been widely assumed that loss of cortical neurons was a striking and inevitable feature of aging. Brody reported neuronal loss in human cerebral cortex most marked in the superior temporal gyrus and—among the cortical layers—most apparent in the small neurons of the external and internal granular cell layers. Since then, many studies have reported up to a 40 percent loss of cortical neurons, even among subjects selected for antemortem normal mental status. However, this widely held assumption of significant cortical neuronal loss with age has recently come into question.

Humans probably possess their full complement of neurons at birth. After this point, it has been generally thought that neurons, which are postmitotic cells, can no longer divide and that lost neurons cannot be replaced. However, this traditional belief has recently come under question. It was shown in 1997 that multipotential stem cells could be identified in the brains of adult mice and—perhaps most interesting because of the implications for eventual therapies—that such cells can be transplanted to various brain regions where they differentiate to become mature, working components of neuronal systems. A report published in 1998 indicates that adult marmoset monkeys generate new hippocampal neurons. Although a previous comparable study had failed to find similar evidence of adult neurogenesis in rhesus monkeys (a species more closely related to humans), another 1998 report suggested that, in fact, adult rhesus brains do exhibit new neurons. Neuronal progenitor cells have been cultured from the subependymal zones of temporal lobes obtained from adult human epileptics, suggesting a possible lifelong capacity for differentiation. Finally, in 1998 Peter Eriksson of Goteborg, Sweden, and Fred Gage of the Salk Institute in San Diego reported evidence of adult neurogenesis in the dentate gyrus of the hippocampus in humans.

Estimates of the rate of neurogenesis in adult mammalian brains vary greatly, and some theorists suggest that thousands of new neurons are created weekly. One interesting possibility is that adult neurons may enter the cell division cycle, but ordinarily this process is arrested before the G2 phase. In advanced age or in Alzheimer's disease, perhaps regulatory control is lost, and cells proceed further but maladaptively into the cell division cycle and then become targets of apoptosis. Thus, adult human neurogenesis has created a wave of excitement in the neuroscience community, both because it could explain some aspects of age-related neurodegeneration and because it suggests some previously unsought therapeutic options, but it remains unclear how widespread this phenomenon may be in humans and how it relates to or is altered in typical aging or neurodegeneration.

The greatest loss of cortical neurons occurs in the perinatal period, when pruning via genetic programs of cell death—apoptosis—interacts with environmental cues to sculpt a working CNS. After the perinatal period and proceeding through maturation there may be a much slower decrease in neuron counts and neuronal densities. Even if age-related neuronal loss is assumed, it cannot necessarily be concluded that these losses lead to dysfunction; apoptosis may be a common feature of normal adult aging and conceivably serves an adaptive role, paring neuronal populations in a way that optimizes cerebral function for the needs of mature life.

Table 51.2b-2 lists the brain regions that have been reported to undergo or not undergo age-related neuronal loss. Table 51.2b-3 compares the loss reported in normal aging with that reported in persons in Alzheimer's disease; the considerable overlap found in comparable studies suggests that aging and Alzheimer's disease, may, in important respects, be on a continuum. Although the likelihood of age-related neuronal loss probably differs in different brain regions and between those with greater or lesser degrees of neurodegenerative change, one should be very cautious about accepting these reports uncritically. Many studies that survey age-related brain changes fail to screen adequately for subclinical disease, including early neurodegenerative processes that lack a good biological marker and conditions such as diabetes or hypertension that might have produced chronic cerebrovascular compromise without overt stroke. Perhaps more importantly, most surveys of normal brain aging have failed to take into account genetic diversity and pool results, for instance, from subjects with different apolipoprotein E genotypes. There is no easy way to bypass this problem. When one asks what is normal age-related brain change in a human population, one is by necessity attempting to describe norms in a highly heterogeneous group. For similar reasons, one must be wary of studies contrasting normal aging with Alzheimer's disease. Too much uncertainty about a coherent strategy remains to dichotomize these two, and indeed, a dichotomy between late onset nonfamilial Alzheimer's and typical aging may be more a difference of degree than of kind.

Regions that may undergo neuronal loss with normal aging
Cortical
Frontal cortex: superior and middle frontal gyri, precentral gyrus, gyrus rectus
Temporal cortex: superior, middle, and inferior temporal gyri
Parietal cortex: postcentral gyrus, inferior parietal cortex
Occipital cortex: striate and peristriate cortex
Limbic lobe: subiculum, hippocampus, amygdala, cingulate gyrus
Cerebellum: Purkinje cells
Subcortical
Hypothalamus: suprachiasmatic and sexually dimorphic nuclei
Locus ceruleus
Substantia nigra
Nucleus basalis of Meynert (possible)
Putamen (possible)
Inferior olive (possible)
Regions that may not undergo neuronal loss with normal aging
Cortical: none reported
Subcortical
Hypothalamus: supraoptic and paraventricular nuclei
Ventral cochlear nucleus
Motor nucleus of facial nerve
Dorsal tegmental nucleus
Abducens nucleus
Trochlear nucleus
Mammillary bodies

Table 51.2b-2 Locations of Neuronal Loss

Brain Region	Age Range (years)	Percentage Decrease	
		Normal Aging	AD
Frontal lobe			
Precentral gyrus	18–95	22–44	19–37
Association areas	19–95	15–51	22–40
Anterior cingulate gyrus	68–95	15–16	22
Temporal lobe			
Superior gyrus	18–95	34–57	33–53
Middle gyrus	19–97	23–50	25–67
Inferior gyrus	18–95	18–36	11–63
Parietal lobe			
Inferior parietal cortex	19–NA	15	9–37
Occipital lobe			
Area 17	18–95	31–54	17–22
Area 18	19–NA	14	14–30
Hippocampal cortex	45–90	27–80	47
Cerebellum			
Purkinje cells	60–100	25	NA
Subcortical areas			
Locus ceruleus	60–80	43	60–70
Substantia nigra	20–65	35	35–40
Nucleus basalis of Meynert	50–66	Unchanged?	75–80

AD, Alzheimer's disease; NA, not available.
Reprinted with permission from Timiras PS: Aging of the nervous system: structural and biochemical changes. In *Physiological Basis of Aging and Geriatrics*, PS Timiras, editor. CRC Press, Boca Raton, FL, 1994.

Table 51.2b-3 Neuronal Loss in Selected Regions of the Human Brain During Normal Aging and in Alzheimer's Disease

Cortical Regions In multiple studies of human cerebral cortex, a 10 to 60 percent loss of neuronal density has been reported from young adulthood to advanced age in the superior, middle, and inferior temporal gyri; superior and middle frontal gyri; pre- and postcentral gyri; cingulate gyrus; gyrus rectus; striate and peristriate

cortex; subiculum; and inferior parietal cortex. Such losses would amount to roughly a 1 percent decrease per annum in the neocortex between ages 69 and 95. In contrast to Brody's early report, subsequent studies find the greatest age-related loss among the larger neurons. The hippocampus of the temporal lobe was reported to lose 3.6 to 6.2 percent of pyramidal cells per decade; these hippocampal losses may be most prominent in specific regions such as the subiculum and dentate gyrus and have been speculated to be the basis of memory loss with aging. Several nuclei of the amygdala have also been reported to lose cell density, and about 2.5 percent of cerebellar Purkinje cells are reportedly lost per decade.

The question currently under debate is how much, if any, of this cortical neuronal loss is truly normal, that is, an inevitable, universal feature of human aging absent disease. The usual methods of brain tissue examination may confound the measurement of cortical loss. Several investigators have suggested that less shrinkage of older than younger brains in postmortem processing leads to an apparent, rather than a real loss of neuronal density. One careful study at the University of California at San Diego indicated that neuronal density does not actually decrease with age; decreased average neuron size creates the appearance of reduced density. Others proposed that since the cortex shrinks as a whole, which would tend to increase cell density, reports of decreased neuronal density may actually underestimate the neuronal loss that occurs with aging. Another issue is whether sufficient care has always been taken to exclude persons with mild dementia from populations described as normal. Collaborators from Harvard University and Washington University in St. Louis reported that among 10 subjects carefully selected to rule out even very subtle cognitive deficits, there was no neuronal loss in the entorhinal cortex (a part of the temporal lobe important for memory functions) between ages 60 and 90. Similarly, Danish investigators, using a new method that determines total neuron counts rather than neuron density, found no neuronal loss in the memory-associated Ammon's horn pyramidal cell regions of the temporal lobe in cognitively normal elderly humans but found significant losses in the subiculum. A third issue is that white matter loss may be greater than previously thought and may account for much of the observed age-related cerebral atrophy. While most animal studies have reported an increase in glial cells with age, several recent studies in rhesus monkeys suggest that cortical neurons are preserved with age, but that myelin loss is significant, suggesting decreases in the oligodendroglia that envelope axons, which would be consistent with recent MRI observations of disproportionate white matter loss. Taken together, these recent findings call into question the long-accepted assumption of normal age-related cortical neuronal loss; improved methodology may be required to clarify this issue.

If there is neuronal loss with aging, when does it begin? The decrease in cell density observed before the middle teen years probably reflects growth of the cerebrum. However, reports differ regarding whether true adult-era neuronal loss begins as early as age 20, not until age 50, or much later in life. This wide variation may represent either substantial individual variation in maturation of the CNS or considerable variation in the impact of subclinical neurodegeneration.

Subcortical Regions Of particular interest to psychiatry may be the apparent cell losses with aging among subcortical nuclei. In general, the sensory and motor nuclei of the brainstem are unchanged with age, as are the mammillary bodies. However, several nuclei associated with the production of neurotransmitters have been reported to lose neurons. The locus ceruleus (norepinephrine) may lose 40 percent of its pigmented cells after age 63. The substantia nigra (dopamine) may lose 35 percent of pigmented cells after age 65. Data regarding cell density in normal aging of the nucleus basalis (acetylcholine) are less consistent; some studies report losses and others report no significant change with age. Further, several hypothalamic nuclei with potential behavioral significance exhibit cell loss: the suprachiasmatic nucleus (implicated in biological rhythms) and the sexually dimorphic nucleus both reportedly lose neurons with age.

CHANGES IN NEURONAL STRUCTURE

Dendrites form the receptive network of interneuronal communication. Like neuronal cell bodies, the dendritic tree exhibits changes with age. The loss and replacement of dendrites takes place throughout life, a process of neural network remodeling that may underlie the ongoing adaptation of the CNS to the environment. Golgi studies show a net age-related loss of dendritic branching and regularity in several regions of the aged human cortex, including superior frontal and superior temporal gyri, hippocampus, and entorhinal cortex. This loss is most apparent in the pyramidal cells of the third and fifth layers of cortex, which may play an important role in the processing of information within association cortex. Dendritic regression is most common among the oldest of the old in both rodents and primates, and dendritic loss appears to follow a centripetal sequence, with the first losses at the extremes of branching. The number of dendritic spines per unit length of the dendritic branches may also decrease. Axons, which carry signals from the cell body to the target cell, may undergo demyelination, swelling, and changes in cytoskeletal components such as neurofilaments and neurotubules. Synapses, the points of interneuronal contact, are also altered with age, which may reflect the loss of dendritic spines or be an independent process. A 13 percent loss of synapses in the frontal cortex was reported among elderly persons compared with a younger group.

Despite these findings, there is an intriguing variability regarding the reports of age-related dendritic atrophy. Investigators have reported an increase in the number and length of terminal dendritic segments in layer II of the parahippocampal gyrus with advanced age at the same time as a decrease was found in the middle frontal gyrus. It was thus proposed that some neurons are dying and losing their dendrites while others are actually growing with age, expanding dendritic arborization and neuronal connectivity to compensate for losses of neighboring neurons or even that individual neurons may both lose and gain dendrites in the lifelong process of adaptive neural remodeling. This may be one aspect of the potential for plasticity in the aging human brain. Research in macaques suggests that the ability to repair dendritic loss, like the loss itself, is probably both regional and age dependent. Therefore, rather than an inevitable decrease in a neuron's arborization with time, the apparent overall loss of dendrites and dendritic spines with age may represent a slower rate of replacement relative to the rate of loss.

Further evidence of plasticity in the aging CNS comes from animal studies of axon sprouting and reactive synaptogenesis. *Axon sprouting* is the phenomenon in which, when nerve fibers are damaged or lost, fibers from undamaged neurons sprout to form new connections with the target. This process was first noted in the peripheral nervous system and is now recognized to occur in the CNS as well. *Reactive synaptogenesis* is the process by which synapses are replaced by new synapses. This can occur in response to damage, chemical stimulus, or behavioral changes. Both axon sprouting and reactive synaptogenesis occur in the aged rodent brain, although reinnervation is slower in the aged animal. Conceivably, the preservation of these mechanisms with aging permits continued adaptability to novel circumstances, although that adaptability may be lower in aged individuals than in those younger.

Other Structural Changes One of the most consistent findings in the biology of aging is the increase in *lipofuscin*, or aging pigment, in cells. Lipofuscin, which contains lipid, carbohydrate, and protein, is found in granules distributed through the cytoplasm or clustered near the nucleus. It accumulates not only in neurons and glia but also in the cells of multiple organs and multiple species, so that its appearance has been proposed to be a basic function of cellular aging. Although it has been proposed that lipofuscin is produced during cellular self-digestion, it remains unknown whether lipofuscin in neurons merely marks aging or is causally related to neuronal deterioration. To date, there is little evidence that neuronal lipofuscin is deleterious to neuronal function.

Other age-related structural changes include granulovacuolar degeneration and Hirano bodies. *Granulovacuolar degeneration* refers to small granules that stain with silver; *Hirano bodies* are larger and stain with eosin. Both are found in hippocampal pyramidal cells in aging and neurodegenerative disorders. Both are of unknown significance.

NEUROCHEMICAL CHANGES

Neurotransmitters The neurochemistry of aging has been extensively investigated and evidence indicates that several established neurotransmitters undergo significant changes in production, metabolism, or both with age. Some of the observed decrease in neurotransmitter concentrations may be due to loss of the pigmented cells of the locus ceruleus and substantia nigra (which respectively produce norepinephrine and dopamine) and possibly in the nucleus basalis (which produces acetylcholine). However, aging also affects specific points in the metabolic pathways for neurotransmitters.

Monoamines There is an age-related decline in tyrosine hydroxylase activity, the enzyme that converts tyrosine to dopa, although much of this decline may occur before young adulthood. There are inconsistent changes in different brain regions for dopa decarboxylase activity, the enzyme that converts dopa to dopamine. Still, dopamine content of the midbrain and basal ganglia may be 50 percent lower by age 75. Aging may have different effects on the two classes of dopamine receptors, the type 1 (D_1) receptors linked to adenylate cyclase, and the D_2 receptors, which appear to be the binding site associated with clinical efficacy of neuroleptics. Autopsy studies suggest that D_1 receptors may be unchanged, while D_2 receptors in the striatum are progressively lost with aging. However, one noninvasive study using positron emission tomography (PET) scanning with a selective D_1 ligand found an age-related decrease in D_1 binding in both frontal cortex and striatum. In addition, evidence suggests that the concentration of messenger ribonucleic acid (mRNA) for the dopamine transporter (a cell-specific protein that regulates dopamine reuptake into presynaptic terminals) decreases significantly with age, suggesting that age-related decreases in dopaminergic transmission could be due to altered gene expression.

Enzymes that degrade catecholamines are altered regionally. Monoamine oxidase activity increases with age in the frontal cortex, basal ganglia, and substantia nigra, leading to increased catabolism of dopamine and norepinephrine. Catechol-O-methyltransferase activity increases in the hippocampus, again suggesting localized

acceleration of monoamine catabolism.

Acetylcholine Since acetylcholine degrades rapidly in postmortem tissue, inferences about the impact of aging on cholinergic activity come primarily from measurements of the synthetic enzyme cholineacetyltransferase. Cholineacetyltransferase activity reportedly declines in the human cortex with age but appears to remain unchanged in the striatum. There is an age-related decrease in cortical uptake of choline, the precursor that combines with acetylcoenzyme A (acetyl-CoA) to form acetylcholine. Both changes suggest decreased cholinergic function with normal aging, although not of the magnitude seen in Alzheimer's disease. The activity of the degradative enzyme acetylcholinesterase also decreases with age.

GABA Glutamate functions as an excitatory neurotransmitter. However, glutamate is converted to the inhibitory neurotransmitter g-aminobutyric acid (GABA) by the action of the enzyme glutamic acid decarboxylase. GABA is widely distributed in the brain and may act at as many as one third of all synapses. Both GABA concentration and its glutamic acid decarboxylase activity have been reported to decline with age in human cortex. On the other hand, based on inconsistent observations of increased GABA concentration in the brains of aged rats and Alzheimer's patients, a GABAergic theory of aging and Alzheimer's disease postulates that adenosine triphosphate (ATP)-mediated negative control of GABA synthesis declines with age, unleashing GABA production and leading to excessive concentrations of this inhibitory transmitter. Better assays of age-related changes in GABA synthesis will be needed to resolve these inconsistencies.

Neuropeptides and Amino Acids An emerging area of research in neurobiology involves the presence in brain of neuropeptides that function as neurotransmitters. With several exceptions, little is known about the specific role of these neuropeptides in human behavior. For example, beta endorphin may mediate and modify responses to pain and stress, and cholecystokinin may be colocalized with dopamine in limbic dopaminergic neurons that may play a role in psychosis. As with other neurotransmitters, some regional declines in neuropeptide activity have been reported with age. Somatostatin concentrations are reportedly unchanged with normal aging in the frontal cortex, basal ganglia, or substantia nigra, while diminished somatostatin immunoreactivity has been found in the cortex and hippocampus of Alzheimer's patients. Brain corticotrophin-releasing hormone concentration may decrease in neurodegenerative disease. Substance P concentration is reportedly lower with normal aging in the putamen and hippocampus but not in frontal cortex. Neurotensin concentrations have been reported to decline by 40 percent in the human substantia nigra but remain stable in frontal cortex and basal ganglia. Cholecystokinin concentration appears to be unchanged. Vasoactive intestinal polypeptide (VIP) concentration in the temporal lobe increases with old age. More must be learned about the biological role of these neuropeptides to assess the impact of these changes on function in elderly adults.

Perhaps most pertinent to aging and cognition is the apparent functional decline in the excitatory glutaminergic amino acid neurotransmitter system. Glutaminergic activity decreases with age, and more so with Alzheimer's disease. Multiple reports exist of decreased activity of the *N*-methyl-D-aspartate (NMDA) subtype of glutamate receptors, which are important to hippocampal memory function; how much of this decline is presynaptic—decreased production of glutamate—and how much is due to the loss of postsynaptic receptors remains uncertain. A well-established association exists between excess glutaminergic stimulation and cell death. One theory attempts to reconcile the apparent paradox of decreased glutaminergic function with age and increased excitotoxic neuronal injury as follows: aging causes a loss of inhibitory neurons bearing NMDA receptors, which results in a relative increase in excitatory neurons bearing such receptors and unleashes chronic, low-level excitotoxic cell damage.

Gaseous Transmitters Nitric oxide (NO) and carbon monoxide, both soluble gases, were recently reported to function as neurotransmitters in the CNS. NO in particular is emerging as an important new component of the neuronal signaling system. NO is produced when stimulation of the NMDA receptor leads to calcium influx that activates nitric oxide synthase. NO and its congeners have been associated with (1) NMDA-mediated long-term potentiation, thought to be important in learning; (2) NMDA-mediated neurotoxicity, thought to play a role in neurodegenerative disorders including Alzheimer's disease and amyotrophic lateral sclerosis; (3) generation of potentially toxic free radicals; and paradoxically (4) neural protection against excitatory neurotoxicity. The relation between normal aging and gaseous transmitters is being investigated. Aging may be associated with dysregulation of the NO system, which might both lead to neuronal loss due to increased excitatory neurotoxicity and induction of apoptosis and specifically impair the plasticity of neuronal electrophysiology that underlies memory.

Glucocorticoids and CNS Aging A glucocorticoid hypothesis of brain aging was proposed in the late 1970s as a unifying explanation for neuronal loss with advancing age. While the sequence of molecular events culminating in neuronal loss cannot be attributed to this single factor, much evidence supports a role for glucocorticoids in age-related CNS change. The hippocampus is particularly rich in glucocorticoid receptors and particularly vulnerable in aging. Even normal physiological concentrations of corticosteroids are associated with age-related hippocampal pyramidal cell loss and reactive astrocytosis. Stress increases both glucocorticoid concentrations and glucocorticoid-induced neuronal loss. These losses are reportedly mediated by dysregulation of calcium homeostasis. While the evidence is mixed regarding age-related changes in brain glucocorticoid receptors, it is currently hypothesized that aging produces increased neuronal vulnerability to corticosteroids, perhaps because loss of a protective downregulatory mechanism leads to excessive calcium influx. Like certain other aspects of age-related CNS physiology (e.g., oxidative stress, excitatory neurotoxicity, and impaired calcium homeostasis), the effect of glucocorticoids on neurons may be regarded as both a physiological and a pathological occurrence and both an effect of the organism's age and a cause of CNS aging.

Receptors Complicating the issue of neurotransmitter loss is the question of changes in neuronal responsiveness due to age-related changes in receptors. Only a few studies determine receptor number, an anatomical issue; most studies measure receptor binding, a physiological process that could be affected by multiple factors. Receptors may be divided into (1) those that have enzymatic activity, (2) those that couple to G-proteins, (3) those that act as transcription factors, and (4) those whose action is unknown. Depending on type, both increased and decreased receptor function with age have been reported. Many animal studies confirm a loss of hippocampal receptor binding associated with age, stress, or both. One study reported a 30 percent age-related loss of NMDA receptors in the hippocampus. Since this excitatory neurotransmitter facilitates memory storage, changes in its cellular reception might account for memory dysfunction even without cell loss. Neocortical serotonin (5-hydroxytryptamine [5-HT] type 2A (5-HT_{2A}) receptor binding, studied in vivo with PET, also reportedly decreases with age. Another study reports increased GABA receptor sensitivity in aged rats, possibly compensating for the loss of presynaptic production of this inhibitory transmitter.

This again raises the issue of age-related changes in plasticity. Up- or downregulation of receptors modifies neuronal responsiveness. In addition to the apparent decline in reactive synaptogenesis—a question of receptor structure—age-related changes have also been reported in the postsynaptic neuron's capacity to up- or downregulate receptivity—a question of function. It is extremely difficult to study this issue in humans, so one must attempt to extrapolate from the available animal literature. One of the best-established findings is that aged rats show impaired muscarinic receptor regulatory capacity. Different neurotransmitter receptor systems may exhibit different degrees of change, so the net affect on signaling is unclear. Overall, these data suggest that the final impact of normal aging on neurotransmission may be a complex result of simultaneous processes that may in fact have opposing effects.

Synapsins A related issue is the question of the effects of aging on the terminal membrane–associated vesicle proteins, the synapsins. The four synapsins, of which the best studied are synapsin I and synaptophysin, are associated with the regulation of neurotransmitter release. Synapsin concentrations appear to be associated with the ability of axons to form synaptic vesicle clusters in response to contact with neighboring neurons; hence, synapsin production indicates the capacity for intercellular signaling. Several studies in rats have demonstrated an age-related decline in synapsin concentrations.

Neurotrophins Neurotrophins play a critical part in the growth, differentiation, and maintenance of neurons. These substances include nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), glial-cell-line-derived neurotrophic factor (GDNF), and neurotrophins 1,2,3, 4/5, and 6 (NT-1, NT-2, NT-3, NT-4/5, NT-6). Neurotrophins bind to a number of specific membrane receptors; for example, BDNF and NT-3 bind to tyrosine kinase A and B receptors (trkA, trkB). Another neurotrophin receptor, the $p75^{NTR}$, was recently shown to bind many neurotrophins, including NGF, BDNF, NT-3, NT-4/5, and NT-6. Since neurotrophins control nerve growth and may protect against neuronal injury, they are being actively investigated both as factors in aging and neurodegeneration and as potential pharmaceutical agents.

Studies have revealed multiple aging effects on the neurotrophin system. Aging is reportedly associated with decreased trkA, trkC, and p75 immunoreactivity and decreased concentrations of mRNA for trkA, trkB, trkC, and p75 in various parts of the rat CNS. Aging has also been associated with reduced neuronal response to neurotrophins and (possibly pertinent to the neurochemistry of Alzheimer's disease) with reduced transport of NGF in basal forebrain cholinergic neurons. One study reports that the neuroprotective effect of BDNF against ischemic damage is markedly lower in adult rats than in developing rats, although whether this protective capacity continues to decline into senescence is not known. Perhaps most intriguing for understanding programmed cell death, stimulation of the $p75^{NTR}$ by NGF, BDNF, or NT-3 reportedly increases conversion of sphingomyelin to ceramide and may cause apoptotic neuronal death.

Transducers and Second-Messenger Systems Regulation of the transmission of messages from the cell surface to intracellular elements permitting control of transcription is known as *signal transduction*. Some membrane signals may be carried to the nucleus of neurons by a sequence of events that begins with binding of the first messenger (the neurotransmitter) to membrane receptors, which activates a family of transducer molecules known as guanine nucleotide binding (G) proteins. These in turn activate enzymes such as phospholipases that produce molecules that serve as second messengers and activate the phosphorylation and

dephosphorylation events that finally control deoxyribonucleic acid (DNA) transcription and gene expression. In one such cascade, extracellular events trigger activation of phospholipases C and D that hydrolyze glycerophospholipids such as phosphotidylinositol to produce diacylglycerol, a second messenger that activates protein kinase C. Both the G proteins and the enzymes involved in transduction, such as cyclic-adenosine monophosphate (AMP)-dependent protein kinase and protein kinase C, may be altered with aging. Recently (at least in replicating cells) cellular senescence has been specifically associated with decreased diacylglycerol concentration (and hence, protein kinase C activation) in response to external stimuli. Such changes in kinases would perhaps hinder phosphorylation of hippocampal membrane proteins, considered possibly an important step in the consolidation of new memories. Although the extent to which these age-related changes occur in postmitotic neurons is still being investigated, it is thought that these changes may alter gene expression and lead to lasting changes with age in the structure or function of neurons or both.

Calcium Homeostasis Many neuronal functions of behavioral significance—including the regulation of neurotransmission and the production of growth cones that may be one structural basis of long-term memory—depend on careful regulation of calcium ion (Ca^{2+}) concentrations in different compartments. Normal Ca^{2+} influx is controlled by voltage-gated calcium channels, the NMDA receptor channel complex, and the sodium ion (Na^{+})/Ca^{2+} exchanger. Ca^{2+} efflux is controlled by Ca^{2+}–adenosine triphosphate (ATP) pumps and the Na^{+}/Ca^{2+} exchanger. Ca^{2+} may also be released within the neuron by the Ca^{2+}-binding proteins calmodulin, calbindin, calretinin, and parvalbumin. In healthy neurons, after Ca^{2+} influx (e.g., through voltage-gated channels following depolarization), it is expected that Ca^{2+} homeostasis will be restored by Ca^{2+}-dependent outward potassium currents that produce feedback inhibition of the Ca^{2+} influx.

Evidence is emerging that dysregulation of Ca^{2+} influx, perhaps causing excessive intracellular free Ca^{2+}, is a feature of aging in the CNS. Aged rats reportedly exhibit excess Ca^{2+} influx through L-type calcium channels in hippocampal neurons, leading to prolonged afterhyperpolarization and decreased neuronal responsiveness. Concentrations of calbindin and calretinin, two of the Ca^{2+}-binding proteins that would buffer free Ca^{2+}, are reportedly decreased in the hippocampus of aged rats. Uptake of free Ca^{2+} into synaptosomes also decreases with age. However, controversy remains regarding intraneuronal calcium and aging. On one hand, John Disterhoft and colleagues at Northwestern University Medical School reported that aged rabbits have decreased excitability of CA1 hippocampal neurons (neurons in Ammon's horn known to be important for memory consolidation), a likely cause of impaired learning, and they attribute this change to excessive Ca^{2+} influx causing enhanced afterhyperpolarization of the membrane. On the other hand, Hans Forstl's group in Mannheim, Germany, reported both decreased basal intracellular Ca^{2+} and decreased poststimulation Ca^{2+} concentrations in aged rats' brains. Perhaps Ca^{2+} influx may both increase or decrease with aging, depending on regional factors and which receptor system is activated. Further, Ca^{2+} sensitivity may perhaps be enhanced or diminished, depending on other circumstances of aging, such as the co-occurrence of ischemic or glucocorticoid stress. Nonetheless, substantial evidence indicates that excessive Ca^{2+} influx acts in neuronal death by (1) mediating glutamate excitatory neurotoxicity; (2) activating calpain, a protease that can damage neuronal proteins; (3) activating phospholipase C, which activates protein kinase C, which can both amplify Ca^{2+} influx and alter DNA transcription; (4) activating NO synthase, generating NO as part of the excitatory neurotoxic cascade, and (5) activating endonucleases that degrade DNA. Moreover, there may be a tendency toward a self-perpetuating cycle in which excessive Ca^{2+} influx leads to loss of normal feedback controls and further excess influx. In this way, even small aging changes in Ca^{2+} homeostasis may effect age-related changes in the CNS that are pertinent to cognition. The finding that b-amyloid deposition may increase Ca^{2+} influx provides a potential link between altered calcium homeostasis and Alzheimer's-type neuropathology.

METABOLIC CHANGES

The human brain undergoes a variety of changes in chemical constituents and metabolic functions with aging. There are regional decreases in protein content and synthesis. Although total DNA content appears to be preserved, there are regional changes in RNA concentration, which declines in cortex but may actually increase with advanced age in the subicular region of the hippocampus. Brain glucose metabolism may be altered in part by decreased sensitivity of neuronal insulin receptors. Total brain lipid decreases after age 50, although the significance of this is unclear, since decreased brain weight and changes in the proportion of gray and white matter may keep relative lipid content unchanged. Lipid synthesis also decreases, including membrane phospholipids and specific lipids important to the myelin sheath.

Noninvasive neuroimaging methods have recently revealed metabolic changes in the human brain in vivo. The effect of age on cerebral metabolic rate for glucose was measured by the [^{18}F]fluorodeoxyglucose method using PET. One early report found a mean 26 percent decrease in global cerebral metabolic rate for glucose when comparing human volunteers without evidence of dementia over the age range of 18 to 78 years. Several subsequent PET scan studies have failed to confirm this global decline in cerebral metabolism with age, especially after the metabolic measurements have been corrected for ventriculomegaly. While global metabolic rates may remain stable through the life span, regional metabolism may be altered; a relative hypofrontality of glucose metabolism with increasing age has been reported. One factor that complicates these PET measurements is that atrophy may cause underestimation of regional metabolism, although metabolic decline has been reported in brain regions that should be little affected by atrophy or ventricular enlargement. Further, the sensory deficits of the aged may lower their basal level of stimulation and reduce the cerebral metabolic rate. Determining the impact of normal aging on global or regional brain metabolism requires further study.

Magnetic resonance spectroscopy (MRS) is another noninvasive way to measure certain aspects of cerebral metabolism. Proton MRS has been used to measure *N*-acetylaspartate (NAA) concentrations, regarded as a neuronal marker, and ^{31}P-MRS has been used to measure high-energy phosphate concentrations in aged controls and subjects with dementia. Preliminary results suggest a possible decline in NAA concentrations with age or dementia, but this research is still in an early stage.

Oxidative Stress At the molecular level, much recent research suggests that one cost of normal human aerobic metabolism may be oxidative stress-induced neurodegeneration. As with many aspects of age-related change, it is quite difficult to determine how much of this oxidative injury should be considered an inevitable physiological consequence of aging and how much should be considered pathological. It appears that physiological oxidative phosphorylation—the mitochondrial process of reducing O_2 to H_2O—produces free radicals including superoxide anion, hydrogen peroxide, and hydroxyl radicals that cause age-related damage to DNA (especially mitochondrial DNA), proteins, and lipids. Perhaps because free radical production is maximal near the mitochondrial inner membrane or because of a lack of protective histones, mitochondrial DNA is especially sensitive to oxidative injury; several studies report strongly age-related deletional mutations in mitochondrial DNA at about ten times the mutation rate in nuclear DNA. One study of human brain tissue found a marked age-related increase in 8-hydroxy-2-deoxyguanosine concentration (a marker of oxidative DNA damage), which was also much higher in mitochondrial DNA than in nuclear DNA. Most of the protein complexes involved in oxidative phosphorylation are coded by mitochondrial DNA, and these proteins also seem to be particularly vulnerable to oxidative stress, with several studies reporting increased damage to these proteins with age. Lipid peroxidation generates potentially mutagenic free radicals such as peroxyl radicals and singlet oxygen and may also damage neighboring membrane proteins; lipid peroxidation reportedly also increases with age.

Endogenous defense mechanisms exist for all these types of oxidative stress. Superoxide dismutase detoxifies the superoxide radical. Repair-glycolases excise damaged DNA. However, in aging, the rate of repair seems to fall behind the rate of oxidative injury. Perhaps the best specific demonstration of the connection between age-related neurodegeneration and oxidative injury is the finding that familial amyotrophic lateral sclerosis is associated with mutation of the Cu/Zn-superoxide dismutase gene, with decreased enzyme production leading to loss of defense against superoxide radical damage. This may be a model for the more general process by which age-related deficiency in repair processes (especially in mitochondrial DNA, proteins, and lipids) leads to neuronal loss.

VASCULAR CHANGES

As with other aspects of age-related change in the CNS, it is very difficult to dissect the consequences of aging from the consequences of vascular pathologies. However, independent of atherosclerosis, intracerebral arterioles may undergo hyaline arteriosclerotic changes, and become increasingly tortuous and probably stiffer with aging. Possibly, changes occur in laminins in vascular basement membranes. Capillary diameter and length per unit volume also reportedly increases in aged persons. Vascular amyloid, identical to the amyloid found in Alzheimer's disease, has been found in the walls of arteries and arterioles of the aging human brain, although in smaller amounts than seen in Alzheimer's patients. The effect of these microvascular changes on cerebral circulation and brain function in aged persons remains uncertain.

Under normal circumstances, cerebral blood flow directly reflects cerebral metabolism and responds to the metabolic demands of perfused brain tissue. Therefore, cerebral blood flow should be a good measure of neuronal activity. Relative cerebral blood flow has been measured noninvasively using the tracer ^{99m}Tc-*d*,*1*-hexamethylpropylenamineoxime (HMPAO) monitored with single photon emission computed tomography (SPECT) scanning and absolute regional cerebral blood flow has been measured using the xenon-133 inhalation technique in subjects of different ages. Results of these studies are highly variable; some investigators report significant decline in global cerebral blood flow with age and others report no correlation between age and cerebral blood flow. PET scanning has also been used to assess changes in cerebral blood flow with age using such tracers as ^{15}O-labeled water, also with varying results. While a few PET studies report no age-related change, most report that regional cerebral blood flow decreases with age bilaterally, especially in cingulate gyri, parahippocampal, superior temporal,

medial frontal, and posterior parietal cortices. Thus, selective decline in cerebral blood flow probably occurs, especially in limbic and associated areas of cortex and periventricular regions of white matter. Aging may also decrease the capacity for cerebral blood flow autoregulation, perhaps in part because of reduced endothelium-dependent relaxation in response to b-adrenergic stimulation.

It is unclear whether age-related changes in cerebral blood flow may be a factor causing the cognitive changes of normal aging or may result from decreased cognitive activity. However, emerging evidence indicates that CNS vascular changes may contribute to neurodegeneration of the Alzheimer type. Further, recent evidence from PET and functional MRI studies suggests age-related reductions in cortical cerebral blood flow responsiveness to perceptual and memory challenges.

Periventricular White Matter Changes *Leukoaraiosis* refers to rarefaction of the periventricular white matter. This change is visible on head CT scans as a periventricular halo of hypodensity, and visible on both proton density or T2-weighted brain MRI scans as periventricular white matter hyperintensities (PVWMHs). In one study, leukoaraiosis on CT was reported in 22 percent of normal volunteers, 52 percent of patients with vascular dementia, and 61 percent of patients with Alzheimer's disease. In a study of nondemented elderly volunteers, those with leukoaraiosis exhibited slightly lower scores on psychometric testing. PVWMHs have been reported in as many as 93.5 percent of unselected MRI scans and appear on 20 to 60 percent of scans of cognitively normal adults over age 65. The pathophysiological changes underlying PVWMHs comprise a spectrum, including mild ependymal disruption, subependymal gliosis, and myelin loss. However, changes that extend further from the ventricles into the deep white matter probably involve more extensive ischemic microangiopathy. It has been proposed that many of these pathological changes are due to acute or chronic white matter ischemia, and some evidence suggests a correlation between the extent of PVWMHs and (1) cerebrovascular risk factors and (2) degree of dementia among patients with vascular dementia. Still, PVWMHs are so common among nondemented elderly adults that they cannot be taken to indicate the presence or cause of a cognitive disorder. A threshold effect has been reported by several investigators in which PVWMHs and cognitive loss are associated only in the most severe stages. However, the vascular pathophysiology and cognitive significance of PVWMHs remain uncertain.

NORMAL AGING, NEURODEGENERATION, AND ALZHEIMER'S DISEASE

The most difficult problem in characterizing the normal aging of the human CNS is identifying a meaningful boundary between normal aging and abnormal neurodegeneration.

Plaques and Tangles Senile plaques and neurofibrillary tangles are associated with Alzheimer's disease, Down syndrome, and other neurodegenerative diseases. Tangles are fibrous bands of intracytoplasmic inclusions consisting of paired helical filaments that seem to result from abnormal hyperphosphorylation and perhaps glycosylation of the microtubule-associated protein tau. Plaques consist of an b-amyloid core surrounded by microglial processes, dendritic processes, and sometimes neurites that may also contain paired helical filaments. While they are often called “lesions,” implying pathology, both plaques and tangles occur in the brains of normal aged persons. Plaques may be found, in lower average numbers than in Alzheimer's disease, throughout the neocortex and in the hippocampus. Tangles accumulate with age first in layer II of the entorhinal cortex, later in the hippocampus and nearby temporal lobe areas, and later still in neocortex. There is a rough correlation between plaque counts (especially the neuritic type) and cognitive loss, but there is so much overlap between the counts in elderly adults without cognitive loss and in Alzheimer's patients that conventional neuropathological criteria for Alzheimer's disease require age-adjusted plaque counts. There is a better correlation between tangle counts and cognitive loss. Nonetheless, as Arnold Scheibel of the University of California, Los Angeles has noted, “It is not possible to provide parametric limits to the ‘allowable number' of plaques and tangles in the healthy aging brain.” While the number of plaques, the proportion of plaques containing neurites, the presence of hyperphosphorylated tau, and the number of tangles are greater, on average, in individuals with impaired cognition, there is no absolute structural or biochemical difference in these entities when they are found in the brains of apparently normal individuals and in those considered to have Alzheimer's disease, suggesting that plaques and tangles are not specific disease markers but are probably a consequence of multiple age-related processes.

A subgroup of aged individuals has been reported with no evidence of dementia despite numerous neocortical plaques. Since this group also had greater brain weight and neuronal density than controls, possibly individuals with a greater brain reserve are more resistant to age-related neurodegeneration. This phenomenon may be important to the conceptualization of Alzheimer's disease, because it suggests that the same pathological process may have significantly different effects on cognition and that the appearance of dementing disorder is not an all-or-none occurrence but may depend upon crossing a critical threshold of cerebral dysfunction.

Synapsins Synapsin concentrations may decline in normal aging. Robert Terry and colleagues at the University of California at San Diego have also reported that human synaptophysin concentrations are the best correlate of cognitive loss in Alzheimer's disease. Possibly aging leads to deterioration in the neuron's capacity to establish and maintain transmitter-dependent connectivity with other neurons—a form of diminished plasticity—and beyond a certain point, such loss of connectivity impairs cognition enough to be clinically apparent. Again, this model might imply that another aspect of what is currently called Alzheimer's disease may differ from normal aging in degree and timing rather than in kind.

Apolipoprotein E There is a definite association between the number of e4 alleles (zero, one, or two) of the cholesterol transport protein apolipoprotein E and the risk of developing sporadic Alzheimer's disease. e4 is associated with increased b-amyloid in brains of patients diagnosed with Alzheimer's disease; pathophysiologically, e4 may be associated with enhanced production or decreased degradation of neurotoxic amyloid fibrils. Gary Small of the University of California, Los Angeles reported that cerebral metabolism may decrease in a pattern suggesting Alzheimer's disease in people with the e4 allele, even prior to cognitive decline. Recently, a group from the University of Barcelona showed that among elderly patients complaining of memory loss without dementia (age-associated memory decline), the presence of e4 alleles was associated with slightly lower cognitive scores and subsequent progression to dementia. These findings do not resolve the question of whether normal aging and Alzheimer's disease are on a continuum, but they suggest that a genetic risk factor strongly associated with eventual development of Alzheimer's disease can influence the cerebral function of people who appear to be aging normally.

Cerebrovascular Contributions Another problematic area is the relation between age-related cerebrovascular changes, CNS aging, and Alzheimer's disease. The convention is to distinguish between vascular dementias and neurodegenerative dementias such as Alzheimer's disease. However, blood vessels serving the CNS may change throughout life with no overt clinical effect but with an unknown—but probably accelerant—effect on the rate of many aspects of CNS aging. Recent evidence suggests that distinguishing vascular and neurodegenerative dementias may be difficult, since cerebrovascular change may significantly influence the expression of Alzheimer's disease, b-amyloid deposition on cerebral vessels is associated with stroke, and ultrastructural studies show angioarchitectural distortions in cerebral capillaries in brains of persons with Alzheimer's disease. Perhaps the most persuasive evidence comes from the Nun Study, in which David Snowdon and colleagues at the University of Kentucky showed that dementia was more likely to occur in subjects with the neuropathology of Alzheimer's disease if they also exhibited subcortical lacunar infarcts, suggesting an interaction between vascular and Alzheimer's pathologies. Alzheimer's disease may in many ways be likened to atherosclerosis, a pathological process that (1) is more or less universal, (2) becomes clinically apparent only after significant progression, and (3) may be accelerated by and exhibit phenotypic heterogeneity due to many genetic and environmental factors. Recent studies suggest that age-related changes in the CNS, as well as Alzheimer's disease, are not only analogous to vascular disease but are in part due to it.

Clearly, Alzheimer's disease is not just accelerated normal aging of the entire human organism, because little evidence suggests that it is associated with accelerated degeneration of systems outside the CNS. Further, there are familial forms of this disorder in which the outstanding pathophysiology is a mutation in the amyloid precursor protein leading to increased production of neurotoxic b-amyloid. This suggests that in at least some cases, dysregulation of a specific single neurobiological system initiates a cascade of events that causes the phenotype, as might be expected for a true genetic disease. This raises the question of the relation between the genetics of aging and the occurrence of neuronal loss and Alzheimer's disease.

MOLECULAR GENETICS OF CNS AGING

Study of the genetics of CNS aging is just beginning, so any conclusions based on the current data must be considered highly tentative. In addition, many of the phenomena currently under study have been observed in yeast, nematodes, or rodents but not yet in humans. For example, a well-established age-related genetic change in rodent brain is the increase in the transcription and expression of the gene for glial fibrillary acidic protein (GFAP), which is conceivably related to increased astrocytic activity in aged humans. Perhaps most pertinent to both aging and Alzheimer's disease is the fact that across species, neurons possess the cellular machinery of apoptosis.

Apoptosis is a microscopic morphological pattern of a dying neuron that exhibits blebbing of the membrane, clustering of nuclear chromatin, and finally DNA fragmentation without membrane dissolution or the inflammatory features associated with necrosis. This pattern, usually attributed to genetically programed cell death, serves important physiological purposes such as deleting self-recognizing thymocytes and probably the developmental pruning of neurons. However, apoptosis followed by secondary necrosis may be the principal mechanism for loss of neurons with normal aging and Alzheimer's disease. Apoptosis apparently requires the presence of certain death domains in cell receptors and transcription factors, and multiple stimuli may act on cells with these domains to initiate the cascade of apoptosis. For example, NGF appears to induce terminal differentiation of the neuron to its postmitotic state in part by activating sphingomyelinase, which then

generates ceramide, a second messenger probably involved in growth arrest. NGF may also activate the cascade leading to apoptosis, and it is likely that ceramide is a transducer in this process. Hence, the effects of NGF appear to be rather complex, potentially facilitating both the terminal differentiation of the neuron into its mature functional role and the apoptotic death of the neuron. Several other mechanisms may link aging or Alzheimer's disease (or both) to genetically programed neuronal death: age-associated dysfunction of cytochrome c oxidase (complex IV of the oxidative phosphorylation pathway) may enhance apoptosis; b-amyloid may also initiate the apoptotic cascade—a potential explanation for the toxicity of amyloid in neurodegeneration; cyclins (cell cycle regulators) may participate in excitatory neurotoxicity and apoptosis; impaired energy metabolism may lead to production of 4-hydoxynonenal, which mediates apoptosis related to oxidative stress; and caspases may cleave presenilin to produce fragements that enhance vulnerability to apoptosis. Furthermore, ceramide production in aging neurons may be involved not only in the intracellular cascade leading to apoptosis, but also in the loss of neuronal capacity to deal with DNA damage; by activating caspases, ceramide may cause the inactivation of poly-adenosine diphosphate-(ADP)-ribose-polymerase, an enzyme thought to be valuable in DNA repair. This might diminish the neuron's capacity to recover from age-related DNA damage.

However, as Kalluri Subba Rao of the University of Hyderabad in India has pointed out, evidence is mixed regarding age-related decline in DNA repair processes in neurons. Several syndromes thought to be associated with elevated genomic damage or failed repair, such as Down syndrome, ataxia telangiectasia, and Cockayne's syndrome, are also associated with both premature aging and neurodegeneration. Others, such as Werner's syndrome or Hutchinson-Gilford's syndrome, are associated with striking premature aging, or *progeria*, of multiple organ systems but do not exhibit early neurodegeneration, which suggests that the features of aging may be dissociable. The gene for Werner's syndrome, called *WRN*, was recently located on chromosome 8 by positional cloning. *WRN* encodes a protein that is very similar to the helicases involved in DNA unwinding, which may explain why DNA repair capacity is altered in this syndrome. Yet, even though Werner's syndrome is associated with both diminished DNA repair and increased production of oxidized proteins, patients with Werner's syndrome do not typically develop either Alzheimer's disease or hypertension. Down syndrome, on the other hand, is associated with both diminished DNA repair and premature Alzheimer's changes, which hints relation between the genetics of failed DNA repair and the occurrence of neurodegeneration. Williams syndrome, another hereditary condition associated with cognitive deficits and early death (though not thought of as a progeria), has also been associated with premature Alzheimer's-type changes in the brain of a 35-year-old patient. The examples of Werner's, Down, and Williams syndromes suggest that different genes control different aspects of aging or the same mechanisms of aging in different organ systems; hence genes influencing the rate of neurodegeneration may act independently of genes that influence other features of aging.

One recent line of research suggests that even in the adult mammalian CNS, multipotential stem cells exist that retain the capacity to differentiate into neurons. This, combined with the recent report that new neurons arise in the adult marmoset hippocampus, might represent another aspect of adult plasticity and raises the possibility that even in aging, the CNS has some potential to compensate for lost neurons by the differentiation of new neurons from neuroepithelial tissue in response to currently unknown extracellular stimuli. Again, rather than thinking of CNS aging as a one-way street, this finding strengthens the concept of CNS aging as a dynamic to-and-fro of generation and degeneration.

CLASSIFICATION OF AGE-RELATED NEURODEGENERATION

Whether the neurobiological and neuropsychological changes associated with so-called normal CNS aging are best regarded as a later, slower, or less severe presentation of the changes currently called sporadic Alzheimer's disease or to what extent both reflect the same underlying processes is currently unknown. However, excluding the small proportion of familial Alzheimer's cases with mutations in the genes for presenilins 1 or 2 or amyloid precursor protein or other as-yet-unidentified mutations, no evidence currently exists for a pathophysiological process that occurs in Alzheimer's disease but not in aging with normal cognition. Aging is associated with neurodegeneration, but how much is normal? One way to conceptualize the difference between normal and abnormal neurodegeneration might be as follows. There are two categories of age-related neurodegeneration: (1) changes clearly related to a single gene mutation, such as familial Alzheimer's disease, familial Parkinson's disease, familial amyotrophic lateral sclerosis, and perhaps the chromosome 17–linked dementias, and (2) the heterogeneous, overlapping, multidetermined changes of CNS aging affected by interdigitating genetic and environmental factors that may accelerate the process, including apolipoprotein e4 alleles, excitatory neurotoxicity, oxidative stress, abnormal processing of amyloid precursor protein, altered production and function of neuronal proteins (e.g., b-amyloid alpha-synuclein, and ubiquitin), calcium dysmetabolism, enhanced apoptosis, cerebrovascular disease, stress-related glucocorticoid neurotoxicity, and head trauma, or decelerate the process, such as endogenous antioxidation mechanisms and DNA repair processes, cerebral reserve possibly related to developmental stimulation and educational and occupational challenge, and possibly interventions such as postmenopausal estrogen replacement therapy, vitamin E supplementation, and antiinflammatory agents. Should they live long enough to experience these problems, all humans will develop neurodegenerative changes and cognitive loss with age due to cumulative neuronal apoptosis and necrosis reflecting the sum of these factors and ultimately leading to a critical loss of neuronal connectivity that becomes clinically apparent. Some aspects of these interrelated processes may involve vicious cycles, such as mutations of mitochondrial DNA that compromise oxidative phosphorylation, and thus generate reactive oxygen species that further damage DNA. Due to lost regulation of feedback-inhibited control systems, this process might be somewhat analogous to the self-directed criticality of chaos theory, like a gradually fraying of a rope, in which initial fraying accelerates subsequent fraying until the rope inevitably breaks. A subgroup of people will experience a disproportionate acceleration in the aging of the CNS relative to other organ systems—a kind of segmental progeria which is currently called “Alzheimer's disease” but difficult to separate from the universal processes of aging except in arbitrary quantitative terms. Thus, pending the discovery of a definitive genetic or physiological discriminator between aging and sporadic so-called Alzheimer's disease, it might be conceptually and semantically useful to make the following distinction: the term *Alzheimer's disease* should perhaps be reserved for cases of age-related neurodegeneration due to a specific cause such as a heritable mutation. On the other hand, when similar segmental CNS progeria with plaques and tangles that instead results from an interaction between multiple genetic and environmental accelerants and decelerants of aging, and produces a spectrum of overlapping pathophysiologies and phenomenologies with no neurobiologically clear-cut differentiation from aging might be more appropriately labeled *typical, sporadic, age-related neurodegeneration*. In the near future, specific molecular pathophysiologies may be discovered that will enable better isolation of parts of this spectrum, leading to both a better understanding of aging and an expanded causal system of classification of neurodegeneration.

CONCEPTUAL SUMMARY AND UNRESOLVED QUESTIONS

A review of the evidence that the human CNS exhibits significant structural and functional changes with age that tend to have a selective impact on different regions, systems, and molecular processes in the brain reveals a recurrent theme, namely, CNS aging and neurodegeneration may be associated with both an increase in neurotoxic processes and a decline in molecular repair mechanisms. Rather than thinking of brain aging as separate from brain development, one might regard it as a variation on a theme. Over the course of the life span, certain aspects of brain physiology shift in purpose and degree. For instance, in fetal life, apoptosis vigorously selects neurons with favorable connectivity, whereas in postmaturational life, apoptosis is less active and may primarily remove damaged neurons. Neurotrophin activity, cellular surveillance, DNA repair, neurotransmitter production and receptor modification, activity of transducer systems, and many other processes that change over the life span similarly may serve somewhat different purposes at different stages of life. Aging of neurons ultimately might be thought of as reflecting the results of a number of self-reinforcing cycles. For instance, mitochondrial DNA is particularly vulnerable to age-related mutations, which decrease the efficiency of oxidative phosphorylation. This not only deprives cells of energy in the form of high-energy phosphates, but also increases the relative production of reactive oxygen species—free radicals. Reactive oxygen species are notoriously hard on phospholipid components of neuronal membranes, peroxidating them. Lipid peroxidation of neuronal membranes threatens neurotransmission by damaging receptor/transducer integrity, a direct threat to cell-to-cell signaling, and also specifically reduces their capacity to bring in glucose, a form of insulin resistance that some have referred to as diabetes of the brain. Neuronal insulin resistance compromises delivery of glucose, the substrate for oxidative phosphorylation, which then reinforces the cycle by further decreasing the efficiency of energy production. The resulting excessive production of reactive oxygen species may very well increase the rate of mutation of mitochondrial DNA—completing the cycle. The consequences are manifold, since all of the ATP-dependent processes of the neuron might be compromised by this loss of efficiency, from intracellular signaling to transcription and translation of new proteins to DNA repair to alterations in NMDA receptors that increase vulnerability to excitotoxic injury. It is even conceivable that another consequence of excessive reactive oxygen species activity and lipid peroxidation is the upregulation of b-amyloid production, a protein best known as a marker for Alzheimer's disease which might have a physiological function of cellular repair. As an intermediate stage, the neuron is floundering somewhat in its many tasks. The final result would be the evocation of apoptosis and neuronal death, as the injured neuron is removed from the system by setting off the signals for self-destruction. While some aspects of this scenario remain to be confirmed as universal, if this is a fair representation of age-related neuronal change, it is easy to see how it blends into the condition currently referred to as Alzheimer's disease.

Questions that remain to be answered include why do neurons die? Or, more specifically, what is the chain of genetic and environmental causality that leads to the age-related neurodegeneration that seems to involve excitatory neurotoxicity, oxidative damage, calcium dysmetabolism, apoptosis, and necrosis? What makes particular neurons more susceptible? Is it location, vascular supply, connectivity, neurotransmitter type, metabolic rate, DNA repair capacity, free radical detoxification mechanisms, or some combination of these factors? To what extent are neuronal losses due to pathological events as opposed to universal genetically programed effects of aging? To what extent are CNS changes such as dendritic atrophy reversible once they have begun? How does the recent discovery of a potential master genetic clock for aging relate to human brain aging? To what extent are the age-related changes in the CNS deleterious as opposed to adaptations that crystallize the learning of a lifetime? How do the changes reported in cerebral structure, physiology, and chemistry relate to the epidemiology of psychiatric disorders among elderly adults? What role do life events and their psychological impact play in accelerating or decelerating the normal changes of age? What preventive interventions are

most likely to delay age-related neurodegeneration? These and other questions will be the focus of the next generation of research on the neurobiology of human aging.

SUGGESTED ROSS-REFERENCES

Sections in the neural sciences chapter relevant to the discussion of changes in the aging brain include Section 1.3, which discusses specific functional systems that may be altered with age, Section 1.4 on the monoamine neurotransmitters, Section 1.5 on amino acid neurotransmitters, Section 1.6 on neuropeptide biology and regulation, and Section 1.8 on intraneuronal signaling pathways. Section 1.15 and Section 1.16 introduce the basic concepts and psychiatric significance of CT, MRI, SPECT, and PET scanning. Most pertinent for correlation of brain aging with clinical psychiatry are other sections in the geriatric psychiatry chapter, including the introduction and overview in Section 51.1, and Section 51.2b and Section 51.2c on normal aging and physiological aspects. Section 51.3e on Alzheimer's disease and other dementing disorders discusses the important clinical consequences of neurodegeneration of the CNS.

SECTION REFERENCES

Ames BN, Shigenaga MK, Hagen TM: Oxidants, antioxidants, and the degenerative diseases of aging. Proc Natl Acad Sci USA *90:*7915, 1993.

Bowling AC, Beal MF: Aging, energy and Alzheimer's disease. In *Amyloid Protein Precursor in Development, Aging and Alzheimer's Disease*, CL Masters, K Beyreuther, M Trillet, Y Christen, editors. Springer-Verlag, Berlin, 1994.

*Brody H. The aging brain. Acta Neurol Scand Suppl *137:*40, 1992.

Cameron HA, Hazel TG, McKay RD: Regulation of neurogenesis by growth factors and neurotransmitters. J Neurobiol *36:*287, 1998.

Coffey CE, Lucke JF, Saxton JA, Ratcliff G, Unitas LJ, Billig B, Bryan RN: Sex differences in brain aging: A quantitative magnetic resonance imaging study. Arch Neurol *55:*169, 1998.

Cotman CW: Synaptic plasticity, neurotrophic factors, and transplantation in the aged brain. In *Handbook of the Neurology of Aging*, ed 3, EL Schneider, JW Rowe, editors. Academic Press, San Diego, 1990.

Cotman CW, Whittemore ER, Watt JA, Anderson AJ, Loo DT: Possible role of apoptosis in Alzheimer's disease. Ann NY Acad Sci *747:*36, 1994.

Davis PC, Mirra SS, Alazraki N: The brain in older persons with and without dementia: Findings on MR, PET, and SPECT images. AJR *162:*1267, 1994.

de la Torre JC: Cerebromicrovascular pathology in Alzheimer's disease compared to normal aging. Gerontology *43:*26, 1997.

Disterhoft JF, Moyer JR, Thompson LC: The calcium rationale of aging and Alzheimer's disease. Ann NY Acad Sci *747:*382, 1994.

*Drachman DA: Aging and the brain: A new frontier. Ann Neurol *42:*819, 1997.

Eriksson PS, Perfilieva E, Bjork-Eriksson T, Alborn AM, Nordborg C, Peterson DA, Gage FH: Neurogenesis in the adult human hippocampus. Nature Med *4:*1313, 1998.

Fazekas F, Schmidt R, Kleinert R, Kapeller P, Roob G, Flooh E: The spectrum of age-associated brain abnormalities: Their measurement and histopathological correlates. J Neural Trans Suppl *53:*31, 1998.

Finch CE: Neuron atrophy during aging: Programmed or sporadic? Trends Neurosci *16:*104, 1993.

Finch CE, Tanzi RE: Genetics of aging. Science *278:*417, 1997.

Francis PT, Webster MT, Chessell IP, Holmes C, Stratmann GC, Procter AW, Cross AJ, Green AR, Bowen DM: Neurotransmitters and second messengers in aging and Alzheimer's disease. Ann NY Acad Sci *695:*19, 1993.

Gage FH: Stem cells of the central nervous system. Curr Opin Neurobiol *8:*671, 1998.

Gomez-Isla T, Hollister R, West H, Mui S, Growdon JH, Parisi JE, Hyman BT: Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer's disease. Ann Neurol *41:*17, 1997.

*Ihl R, Brinkmeyer J: Differential diagnosis of aging, dementia of the Alzheimer type and depression with EEG-segmentation. Dement Geriatr Cog Disord *10:*64, 1999.

Joseph JA, Cutler RC: The role of oxidative stress in signal transduction changes and cell loss in senescence. Ann NY Acad Sci *738:*37, 1994.

Landfield PW, Eldridge JC: Evolving aspects of the glucocorticoid hypothesis of brain aging: Hormonal modulation of neuronal calcium homeostasis. Neurobiol Aging *15:*579, 1994.

Manolio TA, Kronmal RA, Burke GL, Poirier V, O'Leary DH, Gardin JM, Fried LP, Steinberg EP, Bryan RN: Magnetic resonance abnormalities and cardiovascular disease in older adults. The Cardiovascular Health Study. Stroke *25:*318, 1994.

Marczynski TJ: GABAergic deafferentiation hypothesis of brain aging and Alzheimer's disease revisited. Brain Res Bull *45:*341, 1998.

McKay R: Stem cells in the central nervous system. Science *276:*66, 1997.

Meyer JS, Terayama Y, Takashima S: Cerebral circulation in the elderly. Cerebrovasc Brain Metab Rev *5:*122, 1993.

Moeller JR, Ishikawa T, Dhawan V, Spetsieris P, Mandel F, Alexander GE, Grady C, Pietrini P, Eidelberg D: The metabolic topography of normal aging. J Cereb Blood Flow Metab *16:*385, 1996.

Moody DM, Brown WR, Challa VR, Ghazi-Birry HS, Reboussin DM: Cerebral microvascular alterations in aging, leukoaraiosis, and Alzheimer's disease. Ann NY Acad Sci *826:*103, 1997.

Mrak RE, Griffin ST, Graham DI: Aging-associated changes in human brain. J Neuropathol Exp Neurol *56:*1269, 1997.

Muller WE, Hartmann H, Eckert A, Velbinger K, Forstl H: Free intracellular calcium in aging and Alzheimer's disease. Ann NY Acad Sci *15:*301, 1996.

Nagy Z, Esiri MM, Smith AD: The cell division cycle and the pathophysiology of Alzheimer's disease. Neuroscience *87:*731, 1998.

Obeid LM, Venable ME: Signal transduction in cellular senescence. J Am Geriatr Soc *45:*362, 1997.

Olney JW, Wozniak DF, Farber NB: Excitotoxic neurodegeneration in Alzheimer disease: New hypothesis and new therapeutic strategies. Arch Neurol *54:*1234, 1997.

Pantoni L, Garcia JH: Pathogenesis of leukoaraiosis: A review. Stroke *28:*652, 1997.

Pedigo NW: Neurotransmitter receptor plasticity in aging. Life Sci *55:*1985, 1994.

Rao KS: Genomic damage and its repair in young and aging brain. Mol Neurobiol *7:*23, 1993.

Reiter RJ: Oxidative processes and antioxidative defense mechanisms in the aging brain. FASEB J *9:*526, 1995.

*Roth M: The relationship between dementia and normal aging of the brain. In *Dementia and Normal Aging*, FA Huppert, C Brayne, DW O'Connor, editors. Cambridge University Press, Cambridge, 1994.

*Scheibel AB: Structural and functional changes in the aging brain. In *Handbook of the Psychology of Aging*, ed 4, JE Birren, KW Schaie, editors. Academic Press, San Diego, 1996.

Snowdon DA, Greiner LH, Mortimer JA, Riley KP, Greiner PA, Markesbery WR: Brain infarction and the clinical expression of Alzheimer disease. JAMA *277:*813, 1997.

Strong R: Neurochemical changes in the aging human brain: Implications for behavioral impairment and neurodegenerative disease. Geriatrics *53*(Suppl 1):S9, 1998.

*Sulger J, Dumais-Huber C, Zerfass R, Henn FA, Aldenhoff JB: The calcium response of human T lymphocytes is decreased in aging but increased in Alzheimer's dementia. Biol Psychiatry *45:*737, 1999.

*Timiras PS: Aging of the nervous system: Structural and biochemical changes. In *Physiological Basis of Aging and Geriatrics*, PS Timiras, editor. CRC Press, Boca Raton, FL, 1994.

51.2 ASSESSMENT

51.2C PSYCHOLOGICAL CHANGES WITH NORMAL AGING

JENNIFER J. DUNKIN, PH.D. AND JULIA E. KASL-GODLEY, PH.D.

The demographics of aging are by now well known. Advances in medicine and improvements in living conditions over the past century have resulted in greater and greater numbers of persons reaching old age. This graying of the population is expected to reach its peak over the next 60 years, as the baby-boom cohort reaches old age. By the year 2050 79 million persons are expected to be over the age of 65 (20.6 percent of the population), versus 31 million persons (12.5 percent of the population) over the age of 65 in 1990. Furthermore, the oldest-old population—those 85 years of age and older—is expected to grow at an even faster rate and will comprise 4.6 percent of the population (18 million persons) in 2050, versus 1.3 percent (3.1 million) in 1990.

While on the surface these demographic trends reflect favorably on society's attempts to eradicate disease, reduce infant mortality, and improve standards of living, recent debates (especially in political arenas) have focused on the hidden costs of these changes. For example, aging is associated with increased risks of physical illness, personal losses, and cognitive impairment, among other factors. If these risks translate into significant impairment for most older people, costs to society will increase as more people reach old age. While many view a future with many aging individuals as a contingent of happy, well-adjusted retirees on the golf course, others picture a bleak future with millions of debilitated elderly adults in overflowing nursing homes or living in poverty and isolation. Which picture will predominate in the future? It may be wise to make such predictions with caution; possibly the younger one is, the more susceptible one is to aging stereotypes and the less favorable the process of aging appears. While unique life challenges face the elderly, it may surprise many that most research suggests that most elderly people have high levels of life satisfaction, cope well with the challenges presented them in late life, and demonstrate considerable resilience in the face of functional and personal losses. This chapter reviews these studies and discusses their implications for those working with older adults.

RESEARCH ISSUES

The study of psychological issues and aging is relatively young, with most theoretical developments taking place after World War II. Nevertheless, by now a growing body of literature addresses life span development, mental health issues, cognitive aging, and coping and adjustment in old age. However, these studies vary considerably in design and sample used, which often limits the study's overall impact. A brief review of design and sampling issues provides a context in which to view studies of psychology and aging.

Design The design of a study markedly determines how its results can be interpreted, and erroneous conclusions can be drawn based on misinterpretations of a study's design and its limitations. For example, most studies in the field of aging are cross-sectional, that is, groups of individuals of different ages are sampled at one point in time. While meaningful conclusions about differences between age groups may be drawn from cross-sectional studies, it is inappropriate to draw inferences about change over time from such studies, since individuals are not studied prospectively and cohort differences can be mistaken for longitudinal change. For example, there are vast differences between the average 30-year-old and the average 70-year-old besides age, including average number of years of education, exposure to historical events such as the Great Depression and World War II, and comfort with revealing personal information. Thus, what may resemble change over time may instead be baseline cohort differences in the variable of interest. Because of these issues, it is likely that different trajectories of change will be necessary for each individual age cohort.

Sampling Until recently, most studies of psychological issues in aging have used predominantly white, well-educated, healthy, and financially comfortable samples of older adults. Studies have primarily used samples of subjects in the so-called young-old age group (65 to 75 years of age) or, less frequently, the old-old age group (75 to 85) but rarely the oldest-old group (85 and older). Furthermore, many longitudinal studies of personality development and adjustment have used only male samples. The result is that comparatively little is known about minority elderly adults, elderly women, inner-city and lower-income elderly adults, and the oldest-old population. Many psychometric and neuropsychological tests that measure cognitive functioning do not have adequate normative data for the oldest age groups. Attrition rates are also important phenomena; studies rarely compare individuals who drop out of studies with those who remain in the study. This may overestimate the amount of stability and underestimate the amount of decline in psychological functions, as those who decline in functioning tend to have higher attrition rates than those who do not. In terms of cognitive aging, studies have varied considerably in whether individuals with medical conditions known to affect brain function were included in the sample. While such individuals may be typical of their age cohort, their inclusion in studies makes it difficult to determine which cognitive changes can be expected as a function of the aging process itself and which may be a function of epiphenomena associated with aging such as medical conditions.

MENTAL HEALTH AND PERSONALITY

Mental Health Over the Life Span With the exception of dementing disorders and delirium, mental disorders are no more common in elderly adults than in younger age groups. Epidemiological studies have found that prevalence estimates of common disorders such as major depression, anxiety disorders, substance abuse, and psychotic disorders are generally lower for those over the age of 65 than for younger age cohorts. Furthermore, recent longitudinal studies that have followed individuals for up to 25 years found no evidence that rates of emotional distress increase with age.

Age differences in the experience and expression of both positive and negative emotions have also been studied to determine if older adults report more negative emotional experiences than younger adults. In a recent study using a series of large ethnically diverse samples, older adults reported experiencing fewer negative emotions such as sadness, anger, and fear but more positive emotions such as happiness than those in the younger age cohorts. In addition, older adults reported a greater sense of emotional control over negative internal emotional states than younger subjects. To explain their data, the authors proposed an emotional control model suggesting that aging is associated with an increased ability to inhibit negative emotional states and maintain positive emotional ones.

Stage Theories of Personality Development Early personality theorists proposed that development was completed by the end of childhood or adolescence. One of the first development theorists to propose that personality continues to develop and grow over the course of the life span was Erik Erikson. Erikson believed that development proceeded through a series of psychosocial stages, each with its own conflict that is resolved by the individual with greater or lesser success. Erikson termed the crisis of the last epoch of life *integrity versus despair* and believed that successful resolution of this crisis involved a process of life review and achieving a sense of peace and wisdom through coming to terms with how one's life was lived. For example, Erikson proposed that successful resolution of this crisis would be characterized by a sense of having lived one's life well, while a less successful resolution would be characterized by feeling that life was too short, that one did not choose wisely, and bitterness that one will not have a chance to live life over.

Several studies have attempted to validate aspects of Erikson's theory. In one study, a sample of over 400 men was studied prospectively, and the highest Eriksonian life stage each achieved was rated according to data gathered on the circumstances of his life. For example, if a man had achieved independence from his family of origin and was self-sufficient but was unable to develop an intimate relationship, the highest life stage achieved would be the identity stage, not the intimacy stage. This study found that Eriksonian stages are passed through in sequential order, although often not at the same age for different individuals, and that the stages are surprisingly universal in populations that are ethnically and socioeconomically diverse. However, in another study of approximately 150 elderly subjects that included both men and women in the sample, important gender differences emerged. For elderly women, scores on a self-report measure of identity appeared to bear the strongest relationship to a measure of self-realization, while for elderly men, scores on a measure of trust bore the strongest relationship to self-realization. In addition, a recent longitudinal study of approximately 500 subjects from two age cohorts found not only that the earlier age cohort scored significantly higher on a measure of integrity than the later age cohort, but also that scores for both age cohorts on this measure had declined significantly by the final time of testing. These

data suggest that the conflict of integrity versus despair may have a more favorable outcome in earlier age cohorts than in later ones, raising the possibility that changing societal values have had a negative impact on the struggle for integrity. Furthermore, a recent study found that *wisdom*, a construct related to integrity, bore a stronger relation to life satisfaction in elderly adults than other variables, including finances, health, and living situation.

Personality Over the Life Span: Stability or Change? While Erikson and other stage theorists focused on unique developmental tasks and stages central to each phase of life, other theorists have focused on defining core personality traits within the individual and determining their course over the life span. For example, do those who are gregarious or extraverted during early childhood and adolescence remain extraverted through midlife and old age? Several well-designed longitudinal studies that have followed individuals over periods ranging from 10 to 50 years have found strong evidence for stability in five basic personality traits: extraversion, neuroticism, agreeableness, openness to experience, and conscientiousness. Some studies found slight decreases in extraversion and slight increases in agreeableness as individuals move into the oldest-old category, which contrasts with early theories that proposed that personality rigidifies as individuals age.

Is the fact that personality appears to have considerable stability over time inconsistent with the basic tenets of stage theories? Perhaps not. It may be that while individuals are consistent over time in their basic personality structure, the themes and conflicts with which they struggle change considerably over the life span, from concerns about developing identity and a stable sense of self, to finding a life partner, to issues related to life review, as hypothesized by the stage theories. In addition, in developing theories about personality change, few studies have examined the impact of significant historical events on personality; thus the ways in which these events may result in personality change has not been systematically studied.

ADJUSTMENT AND COPING IN ELDERLY ADULTS

Individuals can expect to experience periods of change, disruption, and even crisis throughout the life span. Older adults experience significant challenges as they enter and negotiate the last phase of life, including bereavement, loss of socially valued roles (e.g., worker), and chronic, accumulating health problems. Younger individuals, when viewing the life circumstances of many elderly adults, often wonder how older individuals maintain self-esteem and joy of living in the face of deteriorating health, cognition, and functioning. Questions have been raised about how age affects the process of managing change and discontinuity, with some authors proposing that people become less effective as they age, because of depleted resources, while others have hypothesized that people become better at managing stressful life events over time, having learned more-effective strategies. These questions have been investigated from a wide range of theoretical perspectives; a discussion of several such perspectives follows.

Defensive Processes and Aging Much of the early work in adaptation and coping was generated by theorists working within a psychoanalytic framework. This model is exemplified by George Vaillant's now classic study in which he outlined a hierarchy of defensive processes and assessed age-related changes in them. Using small samples of men studied over 25 years, Vaillant measured the level of defense by blind ratings of numerous vignettes that described a particular conflict or stressor from several different periods in each man's life. Independent overall life-adjustment scores were made on the basis of history, including occupational and interpersonal functioning. Vaillant found that individuals who received the highest life-adjustment ratings were those who used primarily mature defenses (e.g., humor), and corresponding decreases in life adjustment scores were correlated with the use of neurotic defenses (e.g., intellectualization) and immature defenses (e.g., projection). For example, individuals receiving the highest life-adjustment score showed an average of 7.3 (±2.6) vignettes illustrating use of mature defenses, 7.8 (±3.3) using neurotic defenses, and only 0.6 (±0.6) vignettes using immature defenses. In contrast, individuals with only fair life-adjustment scores had an average of 5.2 (±2.6) vignettes using immature defenses, 12.3 (±2.3) using neurotic defenses, and 4.6 (±2.8) using mature defenses. Vaillant also found increased use of mature defensive processes as individuals age, as well as increased use of mature defenses as psychopathology decreased. Although Vaillant's work can be criticized on the basis of possible selection bias, it was one of the first to use prospective methodology in the study of coping and adaptation and the first study to suggest that coping abilities become more adaptive (rather than more regressive) over the life span.

Age Differences in Coping Strategies and Coping Effectiveness A cognitive conceptualization of adaptation is provided by the more recent stress and coping literature. *Coping* is a general term that refers to the actual strategies individuals use to manage stressful life situations that involve perceived or actual threats. It includes strategies designed to reduce emotional distress (e.g., distraction, passive-avoidance, and positive reappraisal) and strategies designed to solve the problem (e.g., direct action, confrontation, and information seeking). Cross-sectional studies of age differences in coping strategies have yielded somewhat mixed results; some studies find no differences between age groups in their use of coping strategies and other studies find that older adults preferentially use strategies designed to reduce emotional distress, while younger adults use more problem-focused coping, although the effect sizes are relatively small. This finding has often led to the conclusion that older adults are more passive in their coping style, yet when the type of stressor is controlled for, these differences are often eliminated or reduced. In addition, older adults are often attempting to cope with situations that are less changeable than those faced by younger individuals, and therefore using coping techniques that attempt to reduce emotional distress rather than solve the problem may be a more effective approach.

Coping effectiveness across the life span is also of interest to gerontologists. The effectiveness of a coping strategy is not defined a priori, but rather by its outcome, usually measured by level of distress, depression, or health symptomatology. Clearly, if individuals become less effective at coping with the demands of life as they age, this has considerable implications for clinical practice and the kinds of interventions that might be used with older adults. However, studies of coping effectiveness have generally not revealed age differences; that is, the old appear to cope as well as the young, and their outcome, in terms of depression and other stress-related symptomatology, is just as favorable. For example, a recent study investigating age differences in coping across a broad range of stressors using a sample of over 2000 men ranging in age from the late 40s to over 90 found no significant age differences in the reporting of negative emotional states in response to stressors or perceived efficacy in coping.

Role Transitions Role transitions are important in the aging process, and in no domain is it more important to understand issues of coping and adaptation than during transitional phases of life. Many transitions are experienced by elderly adults but two major role transitions are especially likely: retirement, or the transition from worker to nonworker, and bereavement, or the transition from spouse to widow. These major transitions challenge the individual to redefine identity and create new roles to replace those that have been lost.

Retirement Retirement is an age-graded event; that is, individuals in our society are expected to retire at a particular age, regardless of health status or personal preference. This is almost entirely a phenomenon of Western, industrialized, and postindustrialized societies; prior to the late nineteenth century, it was largely a personal decision. The transition and adjustment to retirement has been characterized by social gerontologists as a psychosocial crisis, and the image of the successful executive deteriorating or dying soon after retirement has been popularized in the press. However, recent studies have suggested that most recent retirees rank retirement quite low on a list of psychosocial stressors, with 17 percent of men and 11 percent of women reporting dissatisfaction with retirement. Nevertheless, some heterogeneity exists among retirees, and the factors that predict a poor adjustment to retirement include an involuntary or unexpected retirement, low socioeconomic status, health problems, a competitive and highly achievement-oriented personality (the so-called type A behavior pattern), inability to set future goals, and general difficulty adjusting to change. Further, despite early data suggesting that an increased risk of death in the first 2 years after retirement, recent studies controlling for health status at retirement have failed to reveal any associations between retirement and mortality. In addition, when preretirement mental and physical health is controlled for, there are no clear associations between retirement and subsequent physical or mental health consequences. Thus, retirement must be considered in the context of an individual's life and other changes that may be occurring rather than in isolation.

Spousal Bereavement Demographic data suggest that 51 percent of women and 14 percent of men over the age of 65 will be widowed at least once. Bereaved spouses consistently rank spousal loss as among the most stressful of all life experiences. Early work on bereavement and elderly adults hypothesized that older adults might be less able to cope with the loss of a spouse than younger adults, because of declining physical capabilities and subsequent decrements in the ability to manage stress. Subsequent research has revealed age differences in coping with bereavement but in the opposite direction; younger bereaved spouses report greater symptomatology on measures of physical and mental health following bereavement than older spouses. Several factors might account for this difference, including cohort effects, with younger cohorts more comfortable with acknowledging symptoms, or whether the death was expected or unexpected, with unexpected deaths causing greater symptomatology and occurring with greater frequency in younger age groups.

Other recent data are consistent with the idea that as a group, older adults appear to have a more favorable outcome than expected following the death of a spouse. A recent longitudinal study comparing over 120 older bereaved spouses with age-matched nonbereaved controls found that depressive symptoms peak within the first few months after a death but decline significantly within a year, and after 2 or 3 years of follow-up, while bereaved groups still report significantly higher levels of depressive symptoms than age-matched controls, these differences are very small. For example, at 30 months following the death, female bereaved subjects obtained an average score of 7.22 (±6.31) on the Beck Depression Inventory (BDI), a commonly used self-report scale to measure depressive symptoms, compared with an average score of 6.62 (±5.83) for age-matched female control subjects. Both scores are in the nonclinical range. Even during the first few months following bereavement, few older adult widows (between 5 and 17 percent) score in the severely depressed range on depression rating scales. While elderly bereaved spouses report that loneliness and performing activities of daily living complicate adjustment, many older spouses also report positive feelings following a death, such as relief, in contrast to younger samples. Further, the relationship between spousal loss and subsequent mortality has only been found to be significant in younger samples.

Nevertheless, important exceptions exist; studies suggest that elderly survivors of spouses who committed suicide, who have low social support, and who report greater psychopathology at the time of death are all at higher risk of poorer long-term adjustment.

Resilience of the Aging Self Older adults by and large cope well with adversities and challenges, despite often facing more-chronic and accumulating risks. The concept of *resilience*, offered by several theorists to explain this phenomenon, is defined as the process of both maintaining adaptive behavior in the face of stress and recovering from adversity. Rather than viewing the last epoch of life as a period of multiple losses that deplete the individual, resilience theory suggests that maintenance of adaptive behavior despite these losses is a example of the resilience of the aging self. Resilience may involve a process of *selective optimization*, an accommodative process of rescaling goals to match one's limitations and the demands of the environment, and investing the fewer resources available in personally salient activities that maintain self-esteem.

SOCIAL FUNCTIONING IN LATE LIFE

Research over several decades has consistently validated the importance of social support in buffering the effects of stress; given equivalent stressors, individuals with strong social ties appear to have better outcomes than those who are isolated or who perceive that support is unavailable or unhelpful from those in their network. In fact, a recent study of over 1000 elderly adults found that the anticipation of family support in case of financial hardship was effective in reducing depressive symptoms. Yet, elderly adults appear to have reduced social networks and fewer social contacts than younger adults, which may be especially true for the oldest-old individuals, whose social networks are one-half the size of those in the young-old group. This finding was previously thought to be evidence for the *disengagement* hypothesis of aging (i.e., aging involves a process of withdrawal from society and others) and the development of a more inward focus. However, for many older adults, the reduction in social contact is not voluntary and may be more related to disability and reduced mobility than preference. The disengagement hypothesis has been recently reframed as *socioemotional selectivity*, based on the finding that while overall social contacts do decrease in late life, only contact with acquaintances and other peripheral individuals decreases; social contacts with close friends and family remain stable, and older adults' satisfaction with these close relationships remains high. It may be that as individuals age and resources become fewer, those resources are adaptively invested in maintaining close relationships rather than in maintaining acquaintanceships. However, this phenomenon may depend at least partially on the external circumstances of older people's lives. For example, in a recent study of over 500 elderly individuals over the age of 70, the size of social networks and the feelings of connectedness to that social milieu were related to the presence and availability of nuclear family. Older adults with nuclear family members tended to have social networks that were both larger and more emotionally close than those of older adults without nuclear family. On the other hand, those without nuclear family reported greater feelings of closeness to friends than those with nuclear family, suggesting that older adults can adapt to the absence of family by creating satisfying relationships with others.

The social structure of oldest-old individuals has recently been studied, and data suggest that this group may be particularly vulnerable to social isolation because of multiple losses. In a recent study, most individuals over the age of 85 had at least one surviving relative, but approximately one-quarter had little or no contact with that relative. Most social contact was derived from children, and for those individuals without children, social contact and support was not provided by more-distant relatives, except within the African-American community, where other relatives and friends were available to provide help and support. For the oldest-old group then, the family (especially adult children) appears to play a crucial role in providing social support.

COGNITIVE AGING

The study of cognition and aging has been extraordinarily productive over the past several decades. Nevertheless, the existing literature is difficult to interpret, with confusion or disagreement about the inclusion or exclusion of subjects with medical problems and multiple medications and the inadvertent inclusion of subjects with early dementia, which tends to overestimate the effects of age by lowering the overall mean of elderly groups. Furthermore, before magnetic resonance imaging (MRI) scanning was widely used, studies unknowingly included subjects with subcortical ischemic disease, confounding the effects of age with those of disease. In addition, design issues are important; cross-sectional studies that compare individuals of different age groups at one point in time almost always show broader and sharper age effects than longitudinal studies, because of possible confounding of age effects with cohort effects.

Summary of Age Effects on the Major Cognitive Domains The broad domain of cognition has been divided into subdomains that include attentional processes, language functions, visuoperceptual abilities, learning and memory, processing speed, and the executive functions. Age effects on each domain are reviewed below.

Attentional Processes Attentional processes are a complex group of behaviors by which individuals select information for cognitive processing while ignoring competing information, concentrate on information over a sustained period of time, divide attentional focus between two or more aspects of a task at a time, and shift attentional focus from one set of stimuli to another and back again. Basic attention span (the amount of information that can be apprehended at one time) demonstrates little or no relationship to age. Similarly, sustained concentration or vigilance is not affected by the aging process. However, complex attentional tasks that place more demand on cognitive capabilities (e.g., divided attention) show significant declines with advancing age.

Language Functions and Verbal Abilities Language functions remain remarkably stable with age, with few exceptions. Vocabulary and expressive word knowledge may even increase over the life span. Verbal expression and verbal reasoning skills may demonstrate subtle changes with age; some studies have shown that spontaneous language may become less precise and more repetitious over time, although most of these conclusions were drawn from cross-sectional studies and may be confounded with cohort effects. Finally, the ability to retrieve names of objects from the lexicon spontaneously appears to show reliable but small decreases with age, although these effects are heavily influenced by educational achievement.

Visuoperceptual and Visuoconstructive Abilities Visuoperceptual abilities are those functions that involve the recognition and appraisal of nonverbal relationships, specifically those involving shapes, patterns, and objects that are spatially related. Visuoconstructive abilities are those involving the reproduction of nonverbal material, such as the direct copy of simple and complex designs. Visuoperceptual abilities demonstrate relatively consistent age effects across studies, with older adults generally performing as well as younger adults on tests of simple object and pattern recognition but less well on measures of visual closure and spatial orientation. Similarly, in terms of visuoconstructive functions, older adults perform about as well as younger adults when copying simple two-dimensional designs but less well when copying three-dimensional complex designs and when planning and strategy are necessary for good performance. Age effects are largest on visuoconstructive tasks that are timed, and removing the speed component of the task often considerably reduces (but does not remove) the age difference.

Learning and Memory Effects of age on learning and memory are among the best-documented and best-studied effects in the neuropsychological literature. However, memory is an extremely complex set of functions and systems, and the aging process affects some components of memory more than others. Sensory memory (the extremely brief sensory record of a stimulus), primary memory (exemplified by the ability to recall a string of digits long enough to repeat them), and tertiary memory (memories from childhood) show no significant age effects or mild and subtle effects that can be detected in the laboratory. Studies have consistently shown that secondary memory exemplified by the ability to learn information and recall it after a time delay is most strongly affected by age. Nevertheless, the age-related effects on secondary memory are relatively small; studies suggest that performance declines at a rate of 10 percent per decade of life beginning in midlife. However, once older adults learn information, it is retained well; studies suggest that older adults retain information at rates only slightly below that of younger adults. In addition, available data suggest considerably stronger age effects in spontaneous recall than in recognition memory, which is the ability to recognize information previously presented.

Processing Speed Speed of information processing and pure motor speed demonstrate the most reliable age differences and decrements with age, whether measured cross-sectionally or longitudinally. Indeed, reduced processing speed in aging may be a crucial component of many age-related effects on cognition; a recent study found that age effects on a number of neuropsychological tests were considerably diminished after processing speed was partialled out, even on tests where speed of performance was not measured. Nevertheless, it is doubtful that processing speed accounts for all age effects on cognition, as age effects on memory performance appear to be relatively independent of processing speed.

Executive Functions The term “executive functions” refers to those cognitive skills involved in abstract reasoning, problem solving, planning, response inhibition, and the ability to shift cognitive set. Early studies suggested that executive functions decline with age. However, recent studies that use carefully selected samples and exclude subjects with histories of chronic medical conditions known to affect central nervous system (CNS) functioning or documented evidence of subcortical white or gray matter lesions on MRI scanning reveal very few age differences on executive tasks. Because of the relation of performance on tests of executive functioning to the presence of medical conditions, especially diseases of the cardiovascular system, these studies suggest that some of these age effects may be preventable.

Models of Cognitive Aging Taken together, studies suggest a profile of cognitive changes with old age that includes significantly reduced information processing speed and pure motor speed and mild decrements in spontaneous recall, executive skills, complex attentional processes, and complex visuoperceptual and visuoconstructive abilities. While age-related changes in cognition are statistically significant, most elderly persons still perform within normal limits on cognitive

testing. The extent to which these observed changes are related to the process of aging itself, the effects of particular medical conditions on the brain and CNS functioning, or cohort differences in education and experience is not well understood at present, although a recent longitudinal study with a 10-year follow-up interval revealed significant decline in cognitive functions (excluding psychomotor speed) only in individuals suffering from poor health. A number of models proposed to account for these changes include both psychosocial and biological interpretations. First, it is unlikely that all age-related effects on cognition are secondary to cohort differences, as cohort-sequential studies, in which individuals from different age cohorts are followed prospectively, reveal clear longitudinal decline across age cohorts. Unfortunately, few studies have specifically addressed this important question, and they have only investigated intellectual abilities, not functions such as memory or other neuropsychological variables. Cohort effects probably exert the strongest influence on those abilities most dependent on formal education, but this hypothesis has not been formally tested. It is also unlikely that simple disuse accounts for significant aspects of age-related cognitive decline.

The Cattell-Horn model of intellectual changes over the life span proposed that age affects different cognitive processes differently. The model proposes that general intelligence can be subdivided into *crystallized intelligence* (referring to abilities and information learned through exposure to education and life experience) and *fluid intelligence* (referring to abilities involved in novel problem-solving). The model proposes that crystallized intelligence remains stable or increases over the life span, and may even increase into old age, because of continuing exposure to information. In contrast, fluid intelligence, thought to depend on the integrity of brain function, was hypothesized to decline with advanced age. Available data have generally supported this model, with studies showing little evidence of decline on measures of such crystallized abilities as word knowledge or general fund of information, while measures of fluid abilities such as novel problem-solving generally reveal some decline. However, the model as originally proposed is limited in that it only discusses changes in intellectual functions, while excluding other cognitive functions, such as learning and memory. Furthermore, the limitations of studies of executive functions discussed above apply to the fluid intelligence concept.

Neuropsychological models of age-related decline have attempted to integrate findings from cognitive testing with findings from brain imaging and neuropathological studies to suggest possible causes of cognitive changes. The most promising of these is a model called the frontal deficit hypothesis or the frontal subcortical hypothesis, based on the finding that the pattern of cognitive performance in normal aging, although less pronounced, is similar to that seen in many dementing processes affecting primarily subcortical-frontal circuitry. In addition, neuropathological data suggest that frontal-subcortical structures are the most vulnerable to aging, and many studies reveal that neuronal loss and degeneration of dendritic arborization is most prominent in the frontal cortex and in some subcortical areas. Further, definite changes in the vasculature of the brain with age result in ischemic white and subcortical gray matter disease that predominantly affects cognitive processes mediated by frontal-subcortical circuits. Finally, a recent study found that age-related changes in the dopamine system, particularly declines in dopamine type 2 (D_2)-receptor availability in some basal ganglia structures, were related to age-related declines in performance on both motor and executive systems tasks.

The line between a “normal” process of cognitive aging and a pathological process is not well drawn at present. For example, there is considerable disagreement in the field about the inclusion or exclusion of elderly subjects with medical conditions known to affect brain function or individuals with white matter lesions in studies of normal aging and cognition. Since these conditions affect such a high proportion of functional elderly persons, their exclusion may lead to erroneously high estimates of normal expectations of performance and subsequent classification of many individuals as impaired. In addition, many older individuals without known cerebrovascular risk factors or other medical conditions have periventricular white matter lesions on MRI scanning, confusing further the line between disease processes and normal aging. Until this demarcation is better defined, the study of typical aging, which attempts to characterize most elderly people, as well as the study of optimal aging, which attempts to characterize the ideal outcome of both the aging process and changes that are a function of aging itself and not epiphenomena of other conditions, are both important, as long as the appropriate sample is used.

IMPLICATIONS FOR CLINICAL PRACTICE

One of the major points of this chapter is that with the exception of cognitive processes, age per se is rarely an important predictor of outcome. Studies of mental health over the life span, coping and adjustment, and personality often fail to reveal any significant age effects when other factors are controlled. In addition, research has revealed often surprising resilience in older adults. Therefore, the clinician working with older adults should not expect age to be a risk factor for poor outcome. Unfortunately, this notion has probably been partially responsible for the underidentification of depression in many elderly cohorts; many clinicians used to believe that depression was a normal response to aging. Clearly, the life themes older adults struggle with differ from those of younger adults, and the clinician should be prepared to deal with issues of loss, increasing dependency, and the struggle to come to terms with the past. In addition, the concept of resilience may be useful in psychotherapeutic work with older adults; clinicians can encourage patients to selectively invest resources in activities they find especially gratifying.

The important exception to some of these conclusions may be individuals in the oldest-old category (those over the age of 85). Research has shown that these individuals may be at higher risk of social isolation than younger age groups, especially those without living children. Clinicians should be alert to the unique life situations faced by individuals in this age group, who may be especially vulnerable not only to isolation but also to increasing frailty and the cumulative deleterious effects of chronic medical conditions. Interventions that both provide social support and encourage alternative sources of social support such as senior centers may be especially helpful for individuals in this age group without living family.

In terms of cognitive processes, clear decrements exist in some, but not all, cognitive processes over the life span. However, with the exception of processing speed, these changes are generally small. The typical changes in cognition should not have a significant impact on the clinical situation, either the psychotherapeutic encounter or the psychiatric consultation. Some older adults may require more repetition and explanation of material, especially novel and highly abstract material, but this can easily be managed with modified psychotherapeutic approaches. The pattern of cognitive decline seen in normal aging discussed in this chapter should alert the clinician to changes that deviate from this pattern that might be cause for concern, especially such things as difficulties with recalling information even after cueing, problems with basic attentional processes, problems with constructing basic two-dimensional designs, or language impairment.

Finally, since research findings are based on the central or average tendencies of groups of individuals, the conclusions drawn often mask underlying heterogeneity. This may be especially true with elderly adults, as research has shown that the variability of almost every construct studied (e.g., personality, cognition, health status, and income) increases over the life span. This increased heterogeneity is probably a complex interactive process of increasingly individualized expressions of genetic predispositions in response to accumulating demands from the environment over the life span. Thus, the clinician working with older adults may have less precise guidelines or standards for normative expectations for elderly patients, since they can display a much wider range of behavior, personality, and life experience than younger adults.

SUGGESTED CROSS-REFERENCES

Section 51.2d describes neuropsychological evaluation in the older adult. Section 51.1 is a useful overview of many relevant issues in geriatric psychiatry. Section 51.6e contains more information on ethnic and cultural issues relevant to working with older adults. Section 51.1b contains more information on the epidemiology of psychiatric disorders in late life. Section 51.2b and Section 51.2e describe basic physiological and CNS changes with age. Psychotherapy with the older adult is described in more detail in Section 51.4i. Life span development is discussed in Chapter 50 on adulthood, and in Section 6.2, which describes the work of Erik Erikson in greater detail. Section 3.1 contains detailed information on perceptual and cognitive processes in humans.

SECTION REFERENCES

Aldwin CM, Spiro A, Levenson MR, Bosse R: Longitudinal findings from the normative aging study: I. Does mental health change with age? Psychol Aging *4:*295, 1989.

Aldwin CM, Sutton KJ, Chiara G, Spiro A: Age differences in stress, coping, and appraisal: Findings from the Normative Aging Study. J Gerontol Psychol Sci *51B:*P179, 1996.

Anthony JC, Aboraya A: The epidemiology of selected mental disorders in later life. In *Handbook of Mental Health and Aging*, ed 2, JE Birren, RB Sloane, GD Cohen, editor. Academic Press, New York, 1992.

Ardelt M: Wisdom and life satisfaction in old age. J Geronol B Psychol Sci Soc Sci *52B:*15, 1997.

*Bisconti TL, Bergeman CS: Perceived social control as a mediator of the relationships among social support, psychological well-being, and perceived health. Gerontologist *39:*94, 1999.

Blau ZS, Oser ST, Stephens RC: Patterns of adaptation in retirement: A comparative analysis. In *Coping with Medical Issues: Aging*, A Kolker, PE Ahmed, editors. Elsevier Biomedical, New York, 1982.

Boone KB, Miller BL, Lesser IM, Hill E, D'Elia L: Performance on frontal lobe tests in healthy, older individuals. Dev Neuropsychol *6:*215, 1990.

Bosse R, Aldwin CM, Levenson MR, Workman-Daniels K: How stressful is retirement? Findings from the Normative Aging Study. J Gerontol Psychol Sci *46:*P9, 1991.

Brandtstadter J, Wentura D, Greve W: Adaptive resources of the aging self: Outlines of an emergent perspective. Int J Behav Dev *16:*323, 1993.

Cleiren M: *Bereavement and Adaptation: A Comparative Study of the Aftermath of Death*. Hemisphere Publishing, Washington, DC, 1993.

Cornman JM, Kingson ER: Trends, issues, perspectives, and values for the aging of the baby boom cohorts. Gerontologist *36:*15, 1996.

Costa PT, Yang J, McCrae RR: Aging and personality traits: Generalizations and clinical implications. In *Clinical Geropsychology*, IH Nordus, GR VandenBos, S Berg, P Fromholt, editors. American Psychological Association, Washington, DC, 1998.

Domino G, Hannah MT: Measuring effective functioning in the elderly: An application of Erikson's theory. J Pers Assess *53:*319, 1989.

Erikson E: *Childhood and Society*. Norton, New York, 1950.

Field D, Millsap RE: Personality in advanced old age: Continuity or change? J Gerontol Psychol Sci *46:*P299, 1991.

*Folkman S: Coping across the life span: Theoretical issues. In *Life Span Developmental Psychology: Perspectives on Stress and Coping*, EM Cummings, AL Greene, KH Karraker, editor. Erlbaum, Hillsdale, NJ, 1991.

*Freund AM, Smith J: Content and function of the self-definition in old and very old age. J Gerontol B Psychol Sci Soc Sci *54B:*55, 1999.

Gross JJ, Carstensen LL, Tsai J, Skorpen CG, Hsu AYC: Emotion and aging: Experience, expression and control. Psychol Aging *12:*590, 1997.

Horn JL: The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In *Aging and Cognitive Processes*, FIM Craik, S Trehub, editors. Plenum, New York, 1982.

Krause N: Anticipated support, received support, and economic stress among older adults. J Geronol B Psychol Sci Soc Sci *52B:*284, 1997.

Lang FR, Carstensen LL: Close emotional relationships in late life: Further support for proactive aging in the social domain. Psychol Aging *9:*315, 1994.

Lang FR, Staudinger UM, Carstensen LL: Perspective on socioemotional selectivity in late life: How personality and social context do (and do not) make a difference. J Gerontol Psychol Sci *53B:*P21, 1998.

*LaRue A: *Aging and Neuropsychological Assessment*. Plenum, New York, 1992.

Light JM, Grigsby JS, Bligh MC: Aging and heterogeneity: Genetics, social structure, and personality. Gerontologist, *36:*165, 1996.

*Lund DA: Conclusions about bereavement in later life and implications for interventions and future research. In *Older Bereaved Spouses: Research with Practical Applications*, DA Lund, editor. Hemisphere Publishing, New York, 1989.

*Maier H, Smith J: Psychological predictors of mortality in old age. J Gerontol Psychol Sci *54B:*44, 1999.

Perry CM, Johnson CL: Families and support networks among African-American oldest-old. Int J Aging Hum Dev *38:*41, 1994.

*Salthouse TA: Age related changes in basic cognitive processes. In *The Adult Years: Continuity and Change*, M Storandt, GR VandenBos, editors. American Psychological Association, Washington, DC, 1989.

Salthouse TA, Fristoe N, Rhee SH: How localized are age-related effects on neuropsychological measures? Neuropsychology *10:*272, 1996.

Sliwinski M, Lipton RB, Buschke H, Stewart W: The effects of preclinical dementia on estimates of normal cognitive functioning in aging. J Gerontol Psychol Sci *51B:*P217, 1996.

*Staudinger UM, Marsiske M, Baltes PB: Resilience and reserve capacity in later adulthood: Potential and limits of development across the life span. In *Developmental Psychopathology II: Risk, Disorder, and Adaptation*, D Cicchetti, DJ Cohen, editors. Wiley, New York, 1995.

Swan GE, Dame A, Carmelli D: Involuntary retirement, type A behavior, and current functioning in elderly men: 27-year follow-up of the Western Collaborative Group Study. Psychol Aging *6:*384, 1991.

Thompson LW, Gallagher-Thompson D, Futterman A, Gilewski MJ, Peterson J: The effects of late-life spousal bereavement over a 30-month interval. Psychol Aging *6:*434, 1991.

Tranel D, Benton A, Olson K: A 10-year longitudinal study of cognitive changes in elderly persons. Dev Neuropsychol *13:*87, 1997.

Troll LE: Family-embedded versus family-deprived oldest-old: A study of contrasts. Int J Aging Hum Dev *38:*51, 1994.

Vaillant GE: Theoretical hierarchy of adaptive ego mechanisms: A 30-year follow-up of 30 men selected for psychological health. Arch Gen Psychiatry *24:*107, 1971.

Vaillant GE, Milofsky E: Natural history of male psychological health: IX. Empirical evidence for Erikson's model of the life cycle. Am J Psychiatry *137:*1348, 1980.

Van Gorp WG, Mahler M: Subcortical features of normal aging. In *Subcortical Dementia*, JL Cummings, editor. Oxford University Press, New York, 1990.

Volkow ND, Gur RC, Wang G, Fowler JS, Moberg PJ, Ding YS, Hitzemann R, Smith G, Logan J: Association between decline in brain dopamine activity with age and cognitive and motor impairment in healthy individuals. Am J Psychiatry *155:*344, 1998.

Whitbourne SK, Zuschlag MK, Elliot LB, Waterman AS: Psychosocial development in adulthood: A 22-year sequential study. J Pers Soc Psychol *63:*260, 1992.

Zelinski EM, Burnight KP: Sixteen-year longitudinal and time lag changes in memory and cognition in older adults. Psychol Aging *12:*503, 1997.

51.2 ASSESSMENT

51.2D NEUROPSYCHOLOGICAL EVALUATION

KYLE BRAUER BOONE, PH.D.

Elderly individuals comprise an ever larger percentage of the general population. Older individuals are particularly at risk for the development of diseases that compromise cognitive function, leading to substantial declines in quality of life and an economic burden to families and society at large. Thus, for each geriatric patient, one must ascertain whether the observed cognitive abnormalities are treatable. A neuropsychological evaluation is a cost-effective and efficient means of determining if a significant cognitive abnormality is present, and if so its impact on activities of daily living and whether the pattern of cognitive scores points to a specific disease process.

EVALUATION AND DIAGNOSIS

The field of neuropsychological testing initially developed from the observation that different brain areas are associated with specific types of thinking skills and that measuring various cognitive abilities can yield information regarding possible areas of cerebral dysfunction. A neuropsychological evaluation involves standardized paper-and-pencil testing of a broad range of cognitive functions, including memory and learning, language, visuospatial skills, executive/problem-solving skills, motor dexterity, attention, information-processing speed, and overall intelligence.

Basic attention is typically assessed by having patients repeat spoken numbers in both forward and reverse order; significantly lowered attentional ability often signals a delirium or confusional state. Information-processing speed can be measured by having patients rapidly complete rote tasks such as drawing lines between numbers in sequential order, copying codes from a template, reading words, or naming colors. Mild-to-moderate losses in mental speed are found in depressed patients, but marked abnormalities in information-processing speed are prominent in patients with Parkinson's disease and other subcortical dementias.

In most individuals, language skills are tied to left hemisphere functioning. The typical language skills assessed in a neuropsychological evaluation include word-retrieval, word generation, and vocabulary range. Word-retrieval is measured by having patients provide precise names for pictured objects, and word generation is quantified by having them generate words beginning with particular letters (e.g., *F*, *A*, *S*) or in specific categories (e.g., animals) within a circumscribed time period. Measurement of vocabulary range usually involves having patients define various words. Mild declines in word-retrieval and word-generation are often observed in early Alzheimer's disease, multi-infarct dementia, and subcortical dementias, although a marked loss in word generation, a skill associated with left frontal lobe functioning, may signify frontotemporal dementia. For example, while Alzheimer's patients in the early stages of the disease may be able to generate only 20 words beginning with the letters *f*, *a*, and *s* within 3 minutes, a performance 50 percent below that of normal individuals, patients with frontotemporal dementia often generate fewer than 10 words.

Visual perceptual and constructional ability are skills primarily associated with right posterior cerebral functioning. Visual perception can be sampled by having patients identify missing parts of pictured objects. Visual spatial/constructional ability has traditionally been assessed by having patients copy line drawings with paper and pencil or reproduce designs with colored blocks. Low scores on visual perceptual/spatial tasks often suggest Alzheimer's disease or other conditions (e.g., right hemisphere stroke) that interfere with right parietal lobe functioning.

Memory testing usually involves measuring the ability to learn both verbal and nonverbal information. Verbal memory, an ability primarily tied to the left temporal lobe, is typically measured by reading patients a list of words, and then having them repeat the words recalled. Four to five learning trials are generally administered to assess the presence of a learning curve. After a delay of 15 to 30 minutes, patients are requested to recite again the words recalled. The use of an 8- or 10-word list (rather than the more standard 15- or 16-item list used in younger populations) appears adequate to detect the memory difficulties exhibited by older patients. Verbal memory can also be assessed by reading patients a short story and having them recall story gist and details, both immediately and after a delay. Visual memory, a skill that may be primarily related to right temporal lobe function, is typically measured by having patients reproduce or recognize pencil-and-paper designs viewed or copied previously.

The ability to retain newly learned information over a delay is especially correlated with medial temporal lobe functioning, and marked impairment in memory scores is found in patients with prominent hippocampal dysfunction (e.g., those with Alzheimer's disease). For example, many patients with Alzheimer's disease cannot recall any word list items, story details, or components of a geometric design after a short delay.

Executive/problem-solving ability is used to describe higher-level organizational skills requiring planning and monitoring of behavior. These abilities are assessed by (1) divided-attention tests in which patients must keep track of two streams of information simultaneously (e.g., Trails B of the Trailmaking Test, in which patients draw lines alternately between numbers and letters in sequential order), (2) tests involving inhibition of overlearned responses (e.g., the Stroop Test, which contains names of colors printed in an incongruous color ink; patients are to name the ink color rather than read the word), and (3) tests requiring behavioral change in response to environmental feedback (e.g., the Wisconsin Card Sorting Test in which patients must deduce the correct strategy for matching cards by listening to feedback to initial sorting attempts). Executive abilities tend to be particularly sensitive to frontal lobe functioning, and while most patients with Alzheimer's disease, vascular dementia, and subcortical dementia show declines in this sphere, pronounced deficits are seen in patients with frontotemporal dementia. For example, many patients with frontotemporal dementia are completely unable to stop themselves from reading the color names on the Stroop Test.

Table 51.2d-1 lists the general cognitive domains assessed in a neuropsychological evaluation with the tests used to measure that skill and a description of the specific behaviors measured by each test. The tests listed in the table constitute a comprehensive test battery generally appropriate for use with a geriatric population. Use of a comprehensive battery is preferable for confident determination of presence and type of dementia or other cognitive disorder in elderly persons, but in some circumstances, administering a several-hour battery is not possible. Tests marked with an asterisk are the core tests that are most sensitive for detection of a dementia.

Table 51.2d-1 Cognitive Domains

Test score interpretation is a multi-step process. First, patient scores are compared with normative data from same-age peers. Second, an attempt is made to compare the patient's scores with estimates of premorbid levels of function based on educational level, grades in school, and occupational attainment. Cognition is generally judged to be impaired when scores fall 1½ to 2 standard deviations below normative means. However, more minor losses warrant longitudinal monitoring because they may signal the earliest stages of a disease process. In addition, in some cases, current scores may be technically within the normal range but still be judged abnormal if evidence indicates the patient functioned at an above-average level premorbidly. Alternatively, although a patient's scores may be significantly depressed, they have less clinical relevance if the patient functioned at a relatively low level premorbidly.

Finally, the pattern of observed deficits is analyzed to determine whether the lowered scores reflect a viable pattern of cognitive decline or a more random pattern associated with diversity in normal functioning. Recent research has revealed that three fourths of healthy, normal individuals aged 50 and older earn a borderline or impaired score on at least one measure in a comprehensive neuropsychological test battery, while 20 percent of normal older individuals display at least two scores in the impaired range in separate cognitive domains. However, the number of persons showing multiple lowered scores within a cognitive domain is much lower. These findings indicate that isolated lowered scores may simply reflect normal variability.

Periodic neuropsychological reevaluations, typically at yearly intervals, can be used to track disease course and document efficacy of treatment. In addition, sequential testing can help determine when conservatorship or durable power of attorney become necessary and when independent living is no longer a viable choice. Prominent memory difficulties can be expected to compromise effective management of safety and financial matters; for example, the patient may forget to lock doors, take medications, turn off the stove, pay bills.

DIFFERENTIAL DIAGNOSIS

Various neurological, psychiatric, and medical diseases have relatively distinct patterns of cognitive deficits, thus neuropsychological testing can aid in differential diagnosis. However, detecting cognitive abnormalities in elderly persons first requires understanding the cognitive decline associated with normal aging.

Normal Aging Many early studies on cognitive losses associated with aging presented a rather dismal picture of marked deficits in most cognitive domains. However, more recent studies that have appreciated the importance of controlling for substance abuse, psychiatric illness, and medical disease have shown that cognitive losses with advancing age are much more circumscribed than previously observed.

Short-term memory, mental speed, and visual perceptual and spatial skills do appear to decline gradually with advancing age. However, verbal intelligence, most language skills (e.g., vocabulary range, reading, spelling), calculation ability, and basic attention appear to be generally well maintained across the life span of healthy individuals. Even executive abilities, which initially were thought to be the cognitive skills most disrupted by aging, are generally found to be spared in healthy persons. For example, several recent studies found no significant declines in word generation and ability to shift set or increases in perseverative responses in healthy, older individuals.

Men and women display subtle differences in cognitive skills, which may become more pronounced with advanced age. Men appear to have a slight advantage in math calculations, fund of general information, and select visual-perceptual-spatial tasks, while women may show a minor advantage on rote verbal memory tasks, select executive skills, and speeded eye–hand coordination. The superiority of women in rote verbal memory and select executive skills appears to be either confined to or more pronounced in elderly women versus elderly men. Biological environmental factors may play a role in the greater gender differences in cognition observed in advanced age. For example, estrogen enhances and protects verbal skills, and verbal fluency and verbal memory skills are higher in older women on estrogen replacement than in older women without hormonal supplements and older men. In addition, comparison of studies conducted since the 1970s with data collected in prior decades has suggested that gender differences in cognition have declined over the last 30 years, perhaps due to changes in parental socialization practices or more equitable school environments. As a result, measurement of cognitive skills in older cohorts could be expected to reveal larger gender differences than those observed in younger generations.

Medical Illness Chronic cardiovascular illness in particular appears to be related to disruption in select cognitive abilities. Hypertensive patients have shown declines primarily on executive, attentional, mental speed, and memory tests. Similarly, patients with mild-to-moderate cardiovascular disease show lowered scores on speed and executive tasks. One recent study found that cardiovascular health status was a potent predictor of performance on an executive categorization measure, accounting for 28 percent of test score variance.

Factors that positively influence the cardiovascular system, such as regular exercise, low cholesterol and triglyceride levels, and antihypertensive medications, may improve cognition, especially executive abilities. For example, vigorous exercise is associated with superior reasoning, concept formation, and abstraction, as well as rapid information processing. Similarly, decreases in serum cholesterol concentration and hypertriglyceridemia, obtained either through diet or medication, improve cognition, particularly problem-solving ability and concept formation. While most studies have failed to detect any neuropsychological differences between patients treated with antihypertensive medications and those who fail to take medications, a few investigations have shown some evidence of improvement in executive skills after medication treatment.

Diabetes and chronic obstructive pulmonary disease also appear to have a negative impact on cognitive ability. Diabetic patients reportedly perform particularly poorly on tests requiring such executive abilities as complex problem-solving, abstract reasoning, categorization, and divided attention. Similarly, chronic obstructive pulmonary disease has been reported to be associated with decrements in perceptual-motor learning, problem-solving, and motor speed and strength in the context of intact verbal intellectual abilities, alertness, and psychomotor speed.

Taken as a whole, the available literature suggests that executive skills may be particularly sensitive to chronic systemic illness. The physiological mechanism underlying this sensitivity is not known, but in some patients it may be related to decreased cerebral blood flow in anterior brain regions. For example, the relative impairment in executive skills in hypertension is consistent with reported declines in regional cerebral blood flow (rCBF) in the frontal and temporal regions of hypertensive patients. Cognitive declines associated with hypertension and diabetes may also be due to “silent” cerebrovascular lesions in brain white matter. Hypertension and diabetes are risk factors for these so-called white matter hyperintensities observed on brain magnetic resonance imaging (MRI) in older individuals, and recent research has suggested that once a threshold of white matter abnormalities is exceeded, significant deficits in attention and executive abilities are observed.

Depression Depressive symptoms, the most common psychiatric symptoms observed in the elderly population, are estimated to occur in over one fourth of community-residing older individuals, with 1 to 2 percent actually meeting criteria for a major depression.

The term “pseudodementia” has been used to refer to cognitive changes that occur during a depressive episode. The incorporation of *dementia* in the name “pseudodementia” implies that advanced age and depression interact to cause cognitive decline that mimicks that observed in a frank dementia such as Alzheimer's disease. The use of *pseudo* reflects the belief that the cognitive abnormalities are not permanent cognitive impairments and that they normalize when the depression remits. While *pseudodementia* has been useful in alerting mental health professionals to the cognitive sequelae of depression, the underlying assumptions embodied in its name have no empirical support.

Examination of neuropsychological functions of older depressed patients with major depression has revealed evidence for subtle declines in *right hemisphere skills* (e.g., performance I.Q., visual constructional skill, visual nonverbal memory). In addition, increasing severity of depression is associated with additional decrements in mental speed and executive skills. The pattern of cognitive scores, suggesting right hemispheric and anterior cerebral dysfunction, matches evidence from some studies of reduced right hemisphere and bilateral frontal lobe rCBF in depression.

While discrete cognitive difficulties are detectable in depression, the abnormalities are mild and do not approximate the deficits in short-term memory, speed, executive skills, and I.Q. found in degenerative or vascular dementias. In addition, advanced age does not appear to potentiate the effect of depression on cognition; in fact, while subtle differences are documented between depressed and nondepressed subjects under age 70, the cognitive abilities of depressed patients age 70 and older actually appear to converge with those of nondepressed elderly persons. Of interest, the cognitive abnormalities in depression appear to be confined to patients with primarily vegetative symptoms; patients with primarily psychological symptoms (e.g., dysphoria, feelings of guilt, suicidal ideation, worry) appear to be cognitively intact. The abnormalities in mental speed and executive skills appear to be “state” cognitive characteristics limited to acute depressive episodes, and the executive deficits in particular have been found to be predictive of treatment response; recent research suggests that the more severe the executive deficits at baseline, the worse the prognosis for recovery from depression. In contrast, the deficits in visual memory, visual spatial skills, and nonverbal intelligence may be “trait”

cognitive characteristics of depression, remaining after recovery from depression and possibly signaling a chronic underlying abnormality in the right hemisphere.

The above research was limited to patients with major depression of mild-to-moderate severity, and very severe depression may be associated with more marked cognitive impairment. However, if pseudodementia exists, it is rare. It is more likely that if significant cognitive impairment is seen in older depressed patients, it is due to the independent effects of concurrent dementia or some collateral factor such as chronic or acute medical illness or white matter hyperintensities, or to an interaction between depression and medical or neurological illness.

Dementia Diagnosis of dementia requires that the patient meet the criteria outlined in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Specifically, the patient must show evidence of a significant short-term memory impairment, as well as evidence of at least one of the following: (1) language impairment, (2) apraxia, (3) agnosia, and (4) executive dysfunction. Several types of dementia exist; the cognitive characteristics of some of the most common forms are discussed below.

Alzheimer's Disease Alzheimer's disease is a progressive degenerative disorder characterized by accumulation of neurofibrillary plaques and neuritic tangles in the hippocampus, amygdala, and temporoparietal and frontal cortex. Neuroimaging studies in the initial stage of Alzheimer's disease reveal prominent hypometabolism and hypoperfusion of bilateral temporal and parietal lobes. The associated cognitive findings are significant loss in short-term memory for both verbal and nonverbal information and mild-to-moderate losses in visual spatial/constructional skill, word-retrieval, mathematical calculations, mental speed, executive skills, and overall intelligence. While recognition memory is generally superior to free recall, it is still below normal. In contrast, basic attention (as measured by forward digit span), motor dexterity, remote memory, and vocabulary range are generally intact.

The middle stage of Alzheimer's disease is marked by diffuse cerebral hypometabolism and hypoperfusion, with associated marked declines in most standardized test scores. In the final stage of the illness, profound cognitive changes render patients essentially untestable.

The various neuropsychological test scores vary in their utility for diagnosis versus disease staging. Learning and recall scores are the best measures for initial diagnosis of dementia of the Alzheimer's type, since they decline precipitously early in the disease but then plateau. In contrast, because word-retrieval, word-generation, visual spatial skills, and recognition memory decline gradually from onset to the middle stage of the illness, these abilities are better for staging the disease.

Vascular Dementia Vascular dementia is typically associated with multiple infarcts in cerebral gray matter that are readily observed on brain imaging. Given the different etiologies for Alzheimer's disease and vascular dementia, one might expect the two diseases to have distinct neuropsychological profiles. However, empirical comparisons of cognitive scores of the two groups generally have revealed no significant differences, although there are a few exceptions. Some studies have suggested that patients with Alzheimer's disease tend to have intact motor dexterity, while patients with vascular dementia display motor dexterity impairment comparable to their cognitive impairment. This occurs because the infarctions in vascular dementia impinge on motor areas, whereas in Alzheimer's disease, motor function is spared until late in the disease. In addition, Alzheimer's patients may have lower memory scores, especially in visual memory, and show more scatter across I.Q. subtest scores than patients with vascular dementia. Finally, some reports indicate that patients with vascular dementia may exhibit more depression and also more insight into their impairments in activities of daily living, more focal motor and sensory abnormalities and gait disturbance, and more pronounced frontal system impairment (e.g., worse performance on Wisconsin Card Sorting Test, initiate less on unstructured cognitive tasks) than patients with Alzheimer's disease.

Frontotemporal Dementia Postmortem examination of patients with frontotemporal dementia reveals neuronal loss, gliosis, and sometimes spongiosis and Pick bodies but an absence of the neurofibrillary tangles and amyloid plaques seen in Alzheimer's disease. Frontotemporal dementia has traditionally been considered a rare degenerative dementia; however, it is often misdiagnosed as Alzheimer's disease. In fact, recent research has suggested that it may represent up to 25 percent of primary progressive dementia cases.

In contrast to the relative preservation of social skills observed in the early stages of Alzheimer's disease, frontotemporal dementia is characterized by prominent behavioral changes including loss of personal awareness, hyperorality and changes in eating habits, stereotyped and perseverative behavior, progressive reduction of speech, and aggressive and antisocial actions.

Neuroimaging studies reveal distinct cerebral abnormalities in the early stage of the disease involving atrophy and hypoperfusion of frontal lobe or anterior temporal areas. Cognitive performance reflects the underlying cerebral dysfunction and is characterized by significant declines in executive function, with more variable difficulties in memory, word-retrieval, and overall intelligence. As the disease progresses, the cerebral dysfunction spreads posteriorly, and a diffuse cognitive disturbance emerges that is indistinguishable from moderate-to-severe Alzheimer's disease.

Despite the unique neuroimaging findings that distinguish Alzheimer's disease from frontotemporal dementia early in the disease course, empirical studies have suggested that the cognitive differences between the two are relatively subtle. Both groups perform significantly below controls on verbal and nonverbal memory, executive skills, and complex constructional skills, with Alzheimer's patients showing more widespread memory deficits. The two patient groups do not differ significantly from each other except that Alzheimer's patients score lower on select memory variables and frontotemporal dementia patients may exhibit more perseverative behavior and poorer judgment regarding social dilemmas. Thus, examination of absolute scores is of limited use for distinguishing between the two disorders. However, analysis of the relative pattern of scores across cognitive domains appears to show more promise in differential diagnosis. Specifically, Alzheimer patients demonstrate relatively better performance on executive than memory tasks, while frontotemporal dementia patients typically show the opposite pattern. For example, in a recent study, 100 percent of Alzheimer's patients obtained word generation (FAS test) *z*-scores that were superior to Rey-Osterrieth delay *z*-scores, while only 25 percent of frontotemporal dementia patients showed this pattern.

In addition, while both groups demonstrate low scores on pencil-and-paper copy of complex line drawings, they appear to fail for different reasons; Alzheimer patients score poorly because of frank constructional apraxia, while frontotemporal dementia patients obtain substandard scores because of poor organization rather than any actual constructional impairment. This author has also observed that in copying simpler designs, the performance of frontotemporal dementia patients closely approximates that of normal individuals, while Alzheimer's disease patients continue to demonstrate evidence of a constructional disability even on this easier material.

Emerging research is suggesting that cerebral dysfunction in frontotemporal dementia is often asymmetrical, with unique cognitive and behavioral profiles associated with primarily right anterior hypoperfusion and primarily left anterior hypoperfusion. Left frontotemporal dementia involves prominent language abnormalities including anomia and reduced word generation, losses in expressive vocabulary, inability to sequence verbal material, and lowered verbal I.Q., in the context of intact social behavior. In contrast, patients with right frontotemporal dementia frequently display bizarre and socially disinhibited behavior, flat affect and a remote and distant demeanor, lack of facial expression, and intrusive staring. In addition, the cognitive profile is characterized by lowered performance I.Q., word generation superior to design generation, verbal sequencing superior to pictorial sequencing, more-pronounced perseverative behavior, and relatively intact word-retrieval. A frontotemporal dementia subtype has recently been identified, the *temporal lobe variant*, in which degeneration targets anterior temporal and basal frontal lobes, leaving dorsolateral areas intact. The left temporal lobe variant is characterized by pronounced aphasia, while patients with the right temporal lobe variant display marked behavioral aberrations including impulsiveness, bizarre alterations in dress, decreased facial expression, and increased visual alertness. While dementia is typically viewed as leading to global cognitive devastation, a particularly fascinating finding in patients with frontotemporal dementia, especially those with the temporal lobe variant, has been the observation of development of visual artistic skill after disease onset. It is hypothesized that emergence of artistic talent in a subset of patients with frontotemporal dementia is due to (1) sparing of the posterior brain regions tied to visual-spatial and constructional ability and (2) interruption of the inhibitory control of the degenerated anterior brain regions on posterior visual systems. Table 51.2d-2 summarizes the cognitive patterns of the various clinical disorders.

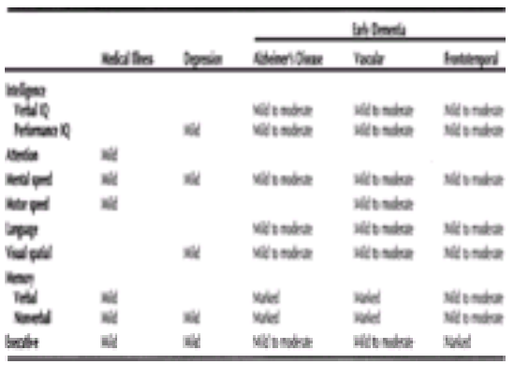

Table 51.2d-2 Summary of Cognitive Deficits for Various Clinical Disorders

SUGGESTED CROSS-REFERENCES

Chapter 3 contains discussions of the biological underpinnings of memory and other thinking processes. For another perspective on neuropsychological assessment, cognitive testing of adults is described in Section 7.4. Neuroimaging findings as they relate to neuropsychiatry and behavioral neurology and geriatric psychiatry are addressed in Section 2.13, Section 51.2e, and Section 51.2f. An overview of normal aging, mood disorder in the elderly, and Alzheimer's disease and other dementing disorders is provided in Chapter 51. For a discussion of cognitive disorders including dementia caused by medical conditions, refer to Chapter 10.

SECTION REFERENCES

*Boone KB: Neuropsychological assessment of executive functions: Impact of age, education, gender, intellectual level, and vascular status on executive test scores. In *The Human Frontal Lobes*, BL Miller, JL Cummings, editors. Guilford, New York, 1999.

Boone KB, Lesser IM, Miller B, Wohl M, Berman N, Lee A, Palmer B: Cognitive functioning in a mildly to moderately depressed geriatric sample: Relationship to chronological age. J Neuropsychiatry Clin Neurosci *6:*267, 1994.

*Boone KB, Lesser IM, Miller BL, Wohl M, Berman N, Lee A, Palmer B, Back C: Cognitive functioning in older depressed outpatients: Relationship of presence and severity of depression to neuropsychological test scores. Neuropsychology *9:*390, 1995.

Boone KB, Miller BL, Lee A, Berman N, Sherman D, Stuss D: Neuropsychological patterns in right versus left frontotemporal dementia. J Int Neuropsychol Soc, in press.

Boone KB, Miller BL, Lesser IM: Frontal lobe cognitive functions in aging: Methodologic considerations. Dementia *4:*232, 1993.

Boone KB, Miller BL, Lesser IM, Hill E, D'Elia L: Performance on frontal lobe tests in healthy, older individuals. Dev Neuropsychol *6:*215, 1990.

Boone KB, Miller BL, Lesser IM, Mehringer CM, Hill-Gutierrez E, Goldberg MA, Berman NG: Neuropsychological correlates of white-matter lesions in healthy elderly subjects: A threshold effect. Arch Neurol *49:*549, 1992.

Corey-Bloom J, Galasko D, Hofstetter CR, Jackson JE, Thal LJ: Clinical features distinguishing large cohorts with possible AD, probable AD, and mixed dementia. J Am Geriatr Soc *41:*31, 1993.

Croog SH, Levine S, Tests MA, Brown BN, Bulpitt CJ, Jenkins CD, Klerman DL, Williams GH: The effect of antihypertensive therapy on the quality of life. N Engl J Med *314:*1657, 1986.

*Cummings JL, Benson DF: *Dementia: A Clinical Approach*. Butterworth-Heinemann, Boston, 1992.

DeBettignies BH, Mahurin RK, Pirozzolo FJ: Insight for impairment in independent living skills in Alzheimer's disease and multi-infarct dementia. J Clin Exp Neuropsychol *12:*355, 1990.

Dunkin J, Cook IA, Leuchter AF, Kasl-Godley J, Abrams M, Betz B, Anderson-Hanley C: Executive systems dysfunction predicts nonresponse to fluoxetine in major depression. J Int Neuropsychol Soc *4:*5, 1998.

Dywan J, Segalowitz SJ, Unsal A: Speed of information processing, health, and cognitive performance in older adults. Dev Neuropsychol *8:*473, 1992.

Edwards-Lee T, Miller BL, Benson DF, Cummings JL, Russell G, Mena I: The temporal lobe variant of frontotemporal dementia. Brain *120:*1027, 1997.

Elias MF, Robbins MA, Schultz NR, Streeten DHP: A longitudinal study of neuropsychological test performance of hypertensive and normotensive adults: Initial findings. J Gerontol *41:*503, 1986.

Elsayed M, Ismail A, Young RJ: Intellectual differences of adult men related to age and physical fitness before and after an exercise program. J Gerontol *35:*383, 1980.

Gregory CA, Orrell M, Sahakian B, Hodges JR: Can frontotemporal dementia and Alzheimer's disease be differentiated using a brief battery of tests? Int J Geriatr Psychiatry *12:*375, 1997.

Karbe H, Kertesz A, Polk M: Profiles of language impairment in primary progressive aphasia. Arch Neurol *50:*193, 1993.

*La Rue A: *Aging and Neuropsychological Assessment*. Plenum, New York, 1992.

Lesser IM, Boone KB, Mehringer CM, Wohl MA, Miller BL, Berman NG: Cognition and white matter hyperintensities in older depressed patients. Am J Psychiatry *153:*1280, 1996.

Lesser IM, Mena I, Boone KB, Miller BL, Mehringer CM, Wohl M: Reduction of cerebral blood flow in older depressed patients. Arch Gen Psychiatry *51:*677, 1994.

Lindman K, Boone KB, Lesser I, Miller B: Does estrogen replacement therapy protect cognitive ability in postmenopausal women? J Int Neuropsychol Soc *4:*54, 1998.

*Locascio JJ, Growdon JH, Corkin S: Cognitive test performance in detecting, staging, and tracking Alzheimer's disease. Arch Neurol *52:*1087, 1995.

Mendez MF, Ashla-Mendez M: Differences between multi-infarct dementia and Alzheimer's disease on unstructured neuropsychological tasks. J Clin Exp Neuropsychol *13:*923, 1991.

Miller BL, Chang L, Mena I, Boone K, Lesser I: Progressive right frontotemporal degeneration: Clinical, neuropsychological and SPECT characteristics. Dementia *4:*204, 1993.

Miller BL, Cummings J, Mishkin F, Boone K, Prince F, Ponton M, Cotman C: Emergence of artistic talent in frontotemporal dementia. Neurology *51:*978, 1998.

Miller BL, Cummings JL, Villanueva-Meyer J, Boone KB, Mehringer CM, Lesser IM, Mena I: Frontal lobe degeneration: Clinical, neuropsychological, and SPECT characteristics. Neurology *41:*1374, 1991.

Miller BL, Darby A, Benson DF, Cummings JL, Miller MH: Aggressive, socially disruptive and antisocial behaviour associated with fronto-temporal dementia. Br J Psychiatry *170:*150, 1997.

Miller BL, Ikonte C, Ponton M, Levy M, Boone K, Darby A, Berman N, Mena I, Cummings JL: A study of the Lund-Manchester research criteria for frontotemporal dementia: Clinical and single-photon emission CT correlations. Neurology *48:*937, 1997.

*Mitrushina MN, Boone KB, D'Elia LF: *Handbook of Normative Data for Neuropsychological Assessment*. Oxford University Press, New York, 1999.

*Pachana NA, Boone KB, Miller BL, Cummings JL, Berman N: Comparison of neuropsychological functioning in Alzheimer's disease and frontotemporal dementia. J Int Neuropsychol Soc *2:*505, 1996.

Palmer BW, Boone KB, Lesser IM, Wohl MA: Base rates of “impaired” neuropsychological test performance among healthy older adults. Arch Clin Neuropsychol *13:*503, 1998.

Palmer BW, Boone KB, Lesser IM, Wohl M, Berman N, Miller BL: Neuropsychological deficits among older depressed patients with predominantly psychological or vegetative symptoms. J Affective Disord *41:*17, 1996.

Rogers RL, Meyer JS, McClintic K, Mortel KF: Reducing hypertriglyceridemia in elderly patients with cerebrovascular disease stabilizes or improves cognition and cerebral perfusion. Angiology *40:*260, 1989.

Russell EW, Polakoff D: Neuropsychological test patterns in men for Alzheimer's and multi-infarct dementia. Arch Clin Neuropsychol *8:*327, 1993.

Skenazy JA, Bigler ED: Neuropsychological findings in diabetes mellitus. J Clin Psychol *40:*246, 1984.

Spieth W: Cardiovascular health status, age, and psychological performance. J Gerontol *19:*277, 1964.

Spirduso WW: Reaction and movement time as a function of age and physical activity level. J Gerontol *30:*435, 1975.

Swartz JR, Miller BL, Lesser IM, Booth R, Darby A, Wohl M, Benson DF: Behavioral phenomenology in Alzheimer's disease, frontotemporal dementia, and late-life depression: A retrospective analysis. J Geriatr Psychiatry Neurol *10:*67, 1997.

Tei H, Miyazaki A, Iwata M, Osawa M, Nagata Y, Maruyama S: Early-stage Alzheimer's disease and multiple subcortical infarction with mild cognitive impairment: Neuropsychological comparison using an easily applicable test battery. Dementia Geriatr Cogn Disord *8:*355, 1997.

Waldstein S, Manuck SB, Ryan CM, Muldoon MF: Neuropsychological correlates of hypertension: Review and methodologic considerations. Psychol Bull *110:*451, 1991.

Willis L, Yeo RA, Thomas P, Garry PJ: Differential declines in cognitive function with aging: The possible role of health status. Dev Neuropsychol *4:*23, 1988.

51.2 ASSESSMENT

51.2E NEUROIMAGING: OVERVIEW

ERIC M. REIMAN, M.D.

Neuroimaging techniques provide a window through which clinicians and researchers can peer into the head and study the structure and functions of the living human brain. Structural imaging techniques include X-ray, computed tomography (CT), and magnetic resonance imaging (MRI). Functional imaging techniques include positron emission tomography (PET), single photon emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI), and magnetic resonance spectroscopy (MRS).

As clinical tools in geriatric neuropsychiatry, structural imaging techniques are commonly used to rule out potentially treatable space-occupying lesions in patients with dementia and those who develop behavioral disorders in conjunction with the onset of focal neurological signs. Functional imaging techniques are sometimes used to support the diagnosis of Alzheimer's dementia, frontotemporal dementia, and vascular dementia. As research tools in the behavioral neurosciences, imaging techniques have begun to help characterize the alterations in brain structure and function that are associated with normal aging and those that are involved in the predisposition to, clinical course, behavioral features, treatment, and possible prevention of age-related disorders.

STRUCTURAL IMAGING TECHNIQUES

Computed Tomography CT provides images of tissue density. This technique uses a rotating source of radiation to transmit X-rays through the head, scintillation detectors to characterize the attenuation of radiation as it passes through tissues in the head at different trajectories, and a computer algorithm to transform attenuation values into horizontal sections that reflect the electron and physical density of tissues. (A recently developed technique known as spiral CT can provide an image consisting of volumetric sections in about 10 seconds.) Conventionally, cerebrospinal fluid (CSF), which has low radiation-attenuation values, is black; bone, which has high radiation-attenuation values, is white; and the brain, which has intermediate radiation-attenuation values, is gray. In addition, an image can be acquired following the intravenous administration of an iodinated or noniodinated contrast agent. The intravascular contrast agent can help visualize vascular abnormalities, such as arteriovascular malformations and aneurysms, and it can enhance abnormalities associated with a breakdown in the blood-brain barrier, including tumors, abscesses, recent hemorrhage, and inflammation. Advantages of CT include its affordability, its ability to produce an image quickly (an important consideration in detecting lesions that require acute surgical intervention, such as a subarachnoid hemorrhage, and in studying patients who have difficulty remaining still during the imaging procedure), its ability to detect bony abnormalities and calcified tumors, its safety in patients with pacemakers and certain ferromagnetic foreign bodies, and its freedom from small spatial distortions (not usually a consideration in clinical psychiatry). Disadvantages of CT include “beam-hardening artifacts,” which make it difficult to characterize tissues in close proximity to bone (such as inferior frontal and anterior temporal cortex), relatively poor sensitivity in detecting white matter lesions, relatively poor contrast between white matter and gray matter, the inability to acquire sagittal or coronal sections (an advantage for certain lesions), and relatively uncommon but potentially dangerous allergic reactions to the iodinated contrast agent.

Magnetic Resonance Imaging MRI provides information about the response of hydrogen protons in water to perturbations in an externally applied magnetic field. This technique uses a strong magnet (typically 0.5 to 1.5 Tesla), causing protons in water to precess in the same frequency along the axis of a strong magnetic field; brief radio-frequency pulses to tip these protons away from this axis; and radio-frequency receivers to measure one of the three following parameters: T1-relaxation, which reflects the rate at which the protons in water return to the axis of the magnetic field; proton density; and T2-relaxation, which reflects the rate at which the protons decline to precess at the same frequency because of nonuniformities in surrounding protons. Using different radio-frequency pulse sequences, MRI can provide different kinds of images. T1-weighted images are commonly used to characterize normal anatomy: conventionally, gray matter is gray, white matter is white, and CSF is black. T2-weighted images are commonly used to detect brain pathology because cerebrospinal fluid (CSF) and brain lesions (which typically are associated with increased water) are white. Proton density images are intermediate in density. In addition, an MRI image can be acquired following intravenous administration of a paramagnetic contrast agent (most commonly gadolinium diethylenetriamine pentaacetic acid [GdDTPA]). As in CT, the intravascular contrast agent can help visualize vascular abnormalities, such as an arteriovascular malformation or aneurysm, and it can enhance abnormalities associated with a breakdown in the blood-brain barrier, including certain tumors and abscesses, a recent hemorrhage, and inflammation. Its advantages include the ability to characterize potentially complementary tissue parameters, its sensitivity to detect white matter (T2 hyperintensity) lesions, excellent contrast between gray matter, white matter, and CSF, the ability to study brain tissues in close proximity to bone (which is not visualized), the ability to acquire images in any orientation, the ability to acquire volumetric images containing contiguous sections with extremely high spatial resolution, the absence of allergic reactions to the paramagnetic contrast agent, and the absence of ionizing radiation. Disadvantages of MRI include its expense, the inability to detect bony abnormalities and calcified tumors, the duration of the imaging procedure (an important consideration in detecting lesions that require acute surgical intervention and in studying patients who are unable to remain still), claustrophobic anxiety, contraindications to the magnet (e.g., a pacemaker and certain ferromagnetic foreign bodies), and small spatial distortions and signal losses due to inhomogeneities in the magnetic field (a consideration in quantitative, morphometric studies that seek to measure the volume of different brain regions). With only a few exceptions, MRI is the structural imaging technique of choice in clinical psychiatry and the behavioral neurosciences.

FUNCTIONAL IMAGING TECHNIQUES

Positron Emission Tomography PET is an imaging technique that provides information about physiological and biochemical processes. This technique capitalizes on the synthesis and administration of a positron-emitting radiotracer; an imaging system to record PET counts; attenuation-correction and image-reconstruction algorithms to transform PET count records into horizontal images of tissue activity; tracer-kinetic models (mathematical equations that account for the behavior in the body) to convert tissue activity data into quantitative measurements of biochemical or physiological processes; and increasingly sophisticated image-analysis techniques. These still-developing techniques can be used to coregister the blurry images of brain function onto each subjects' volumetric MRI; define anatomical regions of interest; segment the brain into gray matter, white matter, and CSF tissues; correct PET images for the combined effects of atrophy and partial-volume averaging (i.e., contributions from neighboring CSF and white matter regions to measurements in a gray matter region of interest, which are expected to be greater in subjects with brain atrophy); deform brain images into the spatial dimensions of a standard brain atlas; and compute statistical maps that compare PET data acquired in different subjects.

PET measurements of cerebral blood flow and the cerebral metabolic rate for glucose are markers of local neuronal activity; these commonly used measurements have the potential to characterize regions of the brain that are preferentially involved in normal human behaviors, those that are preferentially affected during normal human aging, and those that are involved in the pathophysiology, clinical course, possible diagnosis, and possible treatment of age-related disorders. Because of the unusually short radioactive half-life of the cerebral blood flow tracer [^{15}O] water, cerebral blood flow measurements are usually used in studies that involve multiple scans during a single imaging sessions (e.g., scans acquired during memory and control tasks); because of the quality of images acquired over a 30- to 60-minute period following the uptake of the cerebral metabolic rate for glucose tracer [^{18}F]-fluorodeoxyglucose, these measurements are usually used in studies that involve comparisons across imaging sessions (e.g., clinical studies, comparison of patients with controls, and longitudinal studies performed during the resting state or a single, psychophysiologically stable behavioral task). In addition to these measurements, PET measurements of cerebral blood volume, the cerebral metabolic rate for oxygen, and the oxygen extraction ratio have been used to investigate the viability of ischemic brain regions in patients with cerebrovascular disease. While PET measurements of the cerebral metabolic rate for glucose have the potential to grade tumors and predict survival time with greater accuracy than the histological grade of sampled tissues, PET measurements of amino acid uptake are increasingly used to define the spatial extent of these tumors. PET measurements of dOPA decarboxylase activity (using [^{18}F]6-fluoro-dOPA) and the density of dopamine transporter sites provide information about the functional integrity of dopaminergic

neuronal terminals in the striatum, which is reduced in patients with Parkinson's disease. PET has the potential to characterize a wide range of neuroreceptor and neurotransmitter processes. Techniques have been developed for estimating the density of dopamine D_1, D_2, D_3, D_4, and transporter receptors, serotonin 1A, 2A, and transporter receptors, glutamate *N*-methyl-D-aspartate (NMDA) receptors, acetylcholine muscarinic and nicotinic receptors, μ-opioid and d-opioid receptors, benzodiazepine omega-1 and omega-2 receptors, and estradiol receptors; for characterizing how synaptic concentrations of neurotransmitters (including dopamine, serotonin, acetylcholine, and endorphins) are affected in relation to pharmacological challenges and behavioral tasks; for characterizing the synthesis of serotonin and dopamine and the activity of monamine oxidase types A and B. Techniques are now being developed to investigate second-messenger processes.

Compared with other functional imaging techniques, PET has several advantages. Compared with SPECT and MRS, PET is more sensitive in detecting processes that occur in minute concentrations (such as neurotransmitter and neuroreceptor processes). Compared with SPECT, PET also has the potential to measure a wider range of biochemical and physiological processes (partly because the positron-emitting radiosotopes ^{15}O, ^{11}C, ^{13}N, and ^{18}F can be incorporated into physiological and pharmacological compounds without affecting their behavior in the body) and has fewer artifacts from radiation scatter and attenuation (because of the current availability of attenuation-correction procedures). Compared with fMRI, advantages of PET include a better-established role in comparing images acquired in the resting state during different scanning sessions (e.g., in patients versus controls, before and after treatment, and in longitudinal studies of aging and age-related disorders); less-severe artifacts from head movement, claustrophobic anxiety, and ambient noise; the absence of signal dropout in brain regions in close proximity to sinuses (e.g., the hippocampus); the ease with which one can provide sensory stimuli and acquire ancillary measurements during the imaging session; and the ability to characterize a wide range of neurochemical processes. Disadvantages of PET include its relatively low spatial resolution (limiting its ability to characterize changes in extremely small regions, such as brainstem nuclei), the time required to acquire an image (requiring behavioral and baseline tasks to be studied in blocks), radiation exposure (limiting the number of scans that can be acquired in normal volunteers), expense (typically, about $1500 to $2000 per scan), and limited availability.

Single Photon Emission Computed Tomography Like PET, SPECT is an imaging technique that provides information about biochemical and physiological processes. This technique capitalizes on radiosotopes (e.g., ^{123}I, ^{99m}Tc, ^{133}Xe) with longer radioactive half-lives than those used in PET studies, permitting their manufacture at facilities far away from the imaging system; while more powerful systems continue to be developed, the most widely available systems use one to three gamma cameras that rotate around the head. SPECT is commonly used to provide measurements of cerebral blood flow (e.g., using the radiotracer ^{99m}Tc-hexamethylpropyleneamine oxime [HMPAO]). SPECT also has the potential to measure neuroreceptor processes. Techniques have been developed for estimating the density of acetylcholine presynaptic and muscarinic receptors and dopamine type 1 (D_1), D_2, and transporter receptors. Compared with PET, SPECT studies are less expensive and more widely available. However, SPECT provides a smaller range of biochemical and physiological measurements; current imaging systems have a slightly lower spatial resolution, have lower sensitivity for detecting processes that occur in minute concentrations, and cannot acquire transmission images for the correction of radiation attenuation. The longer-lived radioisotopes tend to be associated with slightly higher levels of radiation exposure.

Functional Magnetic Resonance Imaging First demonstrating its brain mapping potential in 1991, fMRI is a rapidly developing imaging technique that can provide information about regional cerebral blood volume and perfusion in the baseline state and state-dependent within-session changes in cerebral blood volume, cerebral blood flow, and blood oxygenation. fMRI techniques typically use conventional MRI systems (typically 1.5 to 4.7 Tesla) and either an echo planar imaging (EPI) platform (an ultrafast imaging technique usually requires specialized gradient coils for the whole body or head) or a spiral multishot imaging platform. These techniques capitalize on an exogenous or endogenous paramagnetic material to produce a measurable magnetic susceptibility contrast.

To measure cerebral blood volume, one technique capitalizes on an intravenous paramagnetic contrast agent, such as Gd(DTPA), and EPI to dynamically track its passage through small cerebral blood vessels. This technique has been used to map regions of the brain that are involved in normal human behaviors (though it is less commonly used for this purpose than techniques that do not require an exogenous contrast agent) and to identify reductions in temporoparietal cerebral blood volume in patients with Alzheimer's disease. To measure CBF, techniques generally capitalize on arterial spin labeling, in which arterial blood is magnetically tagged prior to its arrival in the imaging plane and EPI is used to track its passage dynamically through small blood vessels. Recently, researchers have begun to develop techniques to generate quantitative cerebral blood flow images, acquire data from a larger number of brain slices, and identify reductions in temporoparietal blood flow in patients with Alzheimer's dementia. Brain oxygen-level dependent (BOLD) contrast capitalizes on the observation that increased local neuronal activity is associated with increases in cerebral blood flow, cerebral blood volume, and oxygen delivery that exceed the cells' need for oxygen, leading to an increased concentration of oxygenated hemoglobin and a decreased concentration of the paramagnetic material deoxygenated hemoglobin in venules and veins. While BOLD is extremely well suited for investigating state-dependent changes in local neuronal activity (e.g., mapping regions of the brain that are involved in normal human behaviors) and has the potential to investigate how these state-dependent changes are affected by aging and age-related disorders, it cannot be used to characterize differences in baseline measurements of brain oxygenation (e.g., task-independent, between-session differences in patients versus normal control subjects, the treated versus nontreated state, and different stages of aging and age-related disorders).

Advantages of fMRI include its spatial resolution (although the spatial relationship between neuronal activity and the hemodynamic response is a limiting factor), its ability to acquire images quickly (although its ability to characterize the sequence of neuronal events may be limited by the lag time between neuronal activity and the hemodynamic event), and the absence of ionizing radiation. Disadvantages include the confounding effects of head movement, claustrophobia, practical challenges providing sensory stimuli or ancillary measurements during the scan, signal dropout and spatial distortions in certain brain regions (e.g., the anterior temporal lobe), and the inability to use BOLD contrast to compare data acquired in the baseline state in different subjects or during different scanning sessions. The last limitation may be addressed by further improvements in the application of fMRI to quantitative perfusion imaging.

Magnetic Resonance Spectroscopy MRS provides information about certain biochemical and physiological processes. This technique capitalizes on the extent to which the response of atoms with an odd number of protons and neutrons (e.g., ^{1}H, ^{7}Li, ^{13}C, ^{19}F, ^{23}Na, and ^{31}P) to perturbations in an externally applied magnetic field is affected by surrounding atoms. MRS uses highly homogeneous, high-field (e.g., 1.5 to 4.7 Tesla) MRI systems, radio-frequency pulses that selectively resonate a particular element, and a localized surface coil to provide information about the relative concentrations of compounds that incorporate that element. For instance, ^{1}H MRS can provide regional information about the relative concentration of *N*-acetyl-L-aspartate (NAA), which is thought to be localized in neurons and is thus a measure of viable neurons; inositol, which is involved in the inositol phospholipid second-messenger cascade; choline-containing compounds, some of which participate in the synthesis and degradation of membrane phospholipids. While patients with Alzheimer's dementia appear to have reduced concentrations of NAA in temporoparietal cortex and other brain regions, additional research is needed to determine the extent to which reduced NAA concentrations distinguish Alzheimer's disease from other forms of dementia, preferentially affect certain brain regions, are related to reductions in neuronal number or the combined effects of atrophy and partial volume averaging, and are related to the density of neuritic plaques, neurofibrillary tangles, and clinical course. ^{31}P MRS can provide information about the relative concentrations of phosphomonoesters and phosphodiesters, precursors and degradation products of membrane phospholipids, respectively; adenosine triphosphate (ATP), adenosine diphosphate (ADP), inorganic phosphate, and phosphocreatine, which are involved in high-energy-phosphate metabolism. Although findings from ^{31}P MRS in patients with Alzheimer's disease have not been entirely consistent, one study found that the phosphomonoester concentration increased in the early stage of the dementia and decreased in the later stages. Since MRS provides relatively unique information about brain chemistry, it has the potential to complement other neuroimaging and neuroscientific methods. Its limitations include poor sensitivity for detection of processes that occur in minute concentrations (e.g., neurotransmitters and neuroreceptors), poor spatial resolution, and an inability to compare measurements throughout the brain. Further research is needed to determine the role of MRS in the differential diagnosis of dementia and other age-related disorders.

NORMAL AGING

Understanding how alterations in brain structure and function are related to the pathophysiology, treatment, natural history, and diagnosis of age-related neuropsychiatric disorders requires understanding the alterations associated with normal aging. Studies of usual aging, which seek to characterize how the brain is affected in a representative sample of older subjects, must be distinguished from studies of “successful aging,” in which researchers attempt to characterize how the brain is affected independent of the potentially confounding effects of psychiatric, substance use, neurological and other medical disorders, and cardiovascular risk factors. It is also important to distinguish between cross-sectional studies of aging, which could be influenced by cohort effects (such as increases in brain volume, intracranial volume, or both in successive generations), and longitudinal studies, which are time-consuming, subject to potentially selective attrition effects, and vulnerable to the possibility that the study might outlive the usefulness of the original technology as improved image-acquisition techniques become available.

Postmortem studies suggest that normal aging is associated with reductions in brain volume and the number of large neurons (due more to cell shrinkage than to cell loss) and white matter changes (including a loss of myelin content and an increase in interstitial fluid). The reductions in brain volume and number of large neurons appear to be greatest in certain regions of prefrontal cortex. Imaging studies of the living human brain have supported some of these postmortem findings independent of the potentially confounding effects of selection bias (e.g., different preexisting illnesses and causes of death in younger and older age groups) and brain-fixation artifacts.

Atrophy Structural brain-imaging studies consistently find that aging is associated with reductions in whole brain volume; cortical thinning; enlargement of lateral ventricles, the third ventricles, and extraventricular cerebrospinal fluid spaces; and more-prominent sulci. These changes appear to begin at about 40 years of age, become more pronounced after age 60, and tend to be most pronounced in certain regions of prefrontal cortex (although other cortical and subcortical gray matter structures are also affected). In general, anatomical measurements become more variable with aging, and this variability is at least partly attributable to the presence of age-related disorders. So far, studies have not found a consistent relationship between age-related reductions in brain volume and neuropsychological test scores. While volumetric methods provide the most sensitive measure of brain atrophy, qualitative rating scales have been used to demonstrate that atrophy is not an inevitable consequence of aging (since the rating of mild atrophy is found in only half of persons by approximately age 70), that in the absence of age-related disorders the atrophy associated with normal aging is rarely rated more severely than mild, and conversely, that the identification of mild atrophy in a clinical study should not be used to confirm the diagnosis of an age-related disorder.

White Matter Lesions Aging is associated with an increased likelihood and severity of focal hyperintensity lesions on T_2-weighted MRIs. The lesions are found most commonly in deep white matter regions and less commonly in the pons, periventricular white matter regions, and the basal ganglia. It may be important to distinguish hyperintensity lesions that appear as periventricular caps or rims from those that appear as periventricular halos and those that extend more deeply into white matter regions or basal ganglia. Lesions in the form of periventricular caps or rims are commonly found in older persons independent of age-related neuropsychiatric disorders, diabetes, or cardiovascular risk factors, and are probably related to decreased myelin content, increased interstitial fluid, and other factors. The other lesions are more likely to be associated with age-related neuropsychiatric disorders (e.g., vascular dementia, dementia of the Alzheimer's type, late-onset depression, and bipolar I disorder), diabetes, or cardiovascular risk factors; these lesions could reflect a range of histopathological changes, including vascular ectasia, dilated perivascular spaces, edema, decreased myelin content, meningeal thickening, atherosclerosis, amyloid angiopathy, and lacunar infarcts.

Neurophysiological Changes When subjects with age-related disorders and cardiovascular risk factors were rigorously excluded from study, successful aging was not associated with significant reductions in whole brain cerebral blood flow or cerebral metabolic rates for oxygen or glucose in the resting, baseline state. When subjects with these disorders were not rigorously excluded from study, usual aging was associated with modest reductions in these measures in the resting baseline state, which could be at least partly related to the combined effects of atrophy and partial-volume averaging. Consistent with MRI and postmortem histological studies, PET studies suggest that normal aging preferentially affects regions in prefrontal cortex. After normalizing data for the variation in absolute measurements, successful and usual aging are associated with significant reductions in prefrontal cerebral blood flow and cerebral metabolic rates for oxygen and glucose in the resting baseline state. These reductions appear to be magnified during visual stimulation (though the use of standard tasks introduces the potentially confounding effects of differential task performance, such as age-dependent reductions in visual acuity). Although normal aging did not affect cerebral blood flow increases in the hippocampal formation during a memory retrieval task, it did affect the location of blood flow increases in prefrontal cortex, suggesting that aging alters some of the executive strategies that participate in tasks, such as memory retrieval. The finding that cognitively normal persons homozygous for the apolipoprotein E (APOE) type 4 allele have a reduced prefrontal cerebral metabolic rate for glucose in late middle age supports the possibility that this Alzheimer's disease susceptibility gene accelerates certain aging processes, and it illustrates the potential of imaging techniques to characterize the extent to which putative risk factors and therapeutic interventions might accelerate or retard aspects of neural aging.

Neurochemical Changes A few studies have investigated the effects of normal aging on neurotransmitter and neuroreceptor processes. While one PET study suggests that usual aging is associated with reduced uptake in the striatum of 6-fluoro-dopa (a marker of dopaminergic, nigrostriatal neuronal terminals), a PET study of more rigorously screened subjects suggests that successful aging is not associated with a significant decline in this measurement. Other PET studies find that normal aging is associated with reductions in the density of available presynaptic dopamine transporter sites, D_1 receptors, and D_2 receptors, muscarinic (M_1, M_2) receptors, and serotonergic (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptors. While these findings are consistent with data from postmortem studies, subjects with age-related neuropsychiatric disorders and cardiovascular risk factors were not routinely excluded, the validity of the radiotracer technique used to characterize D_1 receptors remains to be established, and reductions could be at least partly related to the combined effects of atrophy and partial-volume averaging.

DEMENTIA

Structural brain-imaging techniques are indicated in the diagnostic evaluation of dementia to rule out potentially treatable space-occupying lesions, including brain tumors, cerebrovascular lesions (e.g., subdural hematomas, other hemorrhages, infarcts, and pathological white matter lesions), abscesses, a chronic meningitic process, and normal pressure hydrocephalus and to provide evidence supporting several forms of dementia (such as vascular dementia, Pick's disease and other frontotemporal dementias, an asymmetric cortical degeneration syndrome, and a prior traumatic encephalomalacia). This section briefly considers characteristic brain imaging abnormalities in groups of patients with dementia of the Alzheimer's type, vascular dementia, and other forms of dementia; the promising but not fully established role that functional brain imaging techniques may play in the clinical diagnosis of patients with mild or atypical patterns of cognitive impairment and no known reversible causes; and the emerging use of brain imaging techniques to characterize abnormalities in cognitively normal persons at risk for Alzheimer's disease and to test the ability of treatments to prevent the onset of dementia.

Alzheimer's Disease Alzheimer's disease is the most common form of dementia, accounting for 50 to 75 percent of dementia cases. It is characterized clinically by a gradual but progressive decline in memory and other cognitive abilities and histopathologically by neuritic plaques, neurofibrillary tangles, and a loss of neurons and synapses. The clinical diagnosis of dementia of the Alzheimer's type (or probable Alzheimer's disease), which depends on the exclusion of other causes of dementia through a history, physical and neurological examinations, and laboratory tests, conforms to the histopathological diagnosis in about 70 to 90 percent of cases but is least reliable in the earliest stages of the illness.

Although brain-imaging studies are not routinely indicated to confirm the diagnosis of Alzheimer's disease, they have identified statistical differences between groups of patients with this disorder and normal control subjects. Further research is required before imaging techniques are routinely indicated to confirm the diagnosis of Alzheimer's dementia. Since these techniques cannot predict whether or when persons at risk for Alzheimer's disease will develop dementia, can cause false alarm or false reassurance, and provide no guidance about strategies to prevent the onset of dementia, they are not clinically indicated to predict a person's risk for this catastrophic disorder. Still, some of these techniques could be used scientifically to help test the ability of treatments to retard disease progression (rather than compensate for disease progression) in patients with Alzheimer's disease, persons with mild cognitive impairment (who are at increased risk for dementia of the Alzheimer's type), and cognitively normal persons at genetic risk for Alzheimer's disease.

Volumetric MRI studies find reductions in hippocampal volume and increased rates of cortical atrophy in patients with Alzheimer's disease, which are correlated with the severity of cognitive impairment. Although not clinically indicated, studies of persons at genetic risk for Alzheimer's disease suggest that hippocampal volume reduction tends to begin in conjunction with memory decline, shortly before the onset of Alzheimer's dementia. Techniques for the reliable measurement of hippocampal volumes are labor intensive. Characterizing increased rates of cortical atrophy requires sequential studies (typically, at least one annual follow-up study), some manual processing in addition to an automated image-subtraction algorithm, and confirmation of their predictive value. Additional research is required to establish the sensitivity, specificity, prognostic value, and therapeutic implications of structural brain-imaging techniques in the clinical diagnosis or prediction of Alzheimer's disease.

While some MRI studies fail to find an increased number of T2-hyperintensities in rigorously screened patients with Alzheimer's dementia, others find an increased number of hyperintensities in periventricular regions. One postmortem MRI study found that these hyperintensities were related to loss of myelin content, arterial changes, and a breakdown of the ventricular lining. Clinically, the presence of periventricular white matter changes should neither confirm nor exclude the diagnosis of dementia of the Alzheimer's type.

PET studies find characteristic abnormalities in cerebral blood flow and cerebral metabolic rates for oxygen and glucose in patients with Alzheimer's dementia, including reductions in posterior cingulate, posterior parietal, temporal, and prefrontal regions. Reduced measurements in posterior cingulate and cinguloparietal transitional cortex appear to provide the earliest index of Alzheimer's disease pathology, reductions in posterior parietal and temporal regions also appear relatively early, and reductions in prefrontal and whole brain measurements appear to develop in more advanced stages of the disorder. While the reductions in regional metabolic rates for glucose are typically bilateral, they are sometimes asymmetrical (especially in the earlier stages of the illness) and more-prominent reductions are sometimes found to correspond to salient clinical features (e.g., a reduction in left temporal metabolic rate for glucose in a patient with aphasic symptoms). The metabolic rate for glucose is relatively spared in primary visual and somatosensory cortex and may be least affected in the pons. While reduced metabolic rates for glucose in parietal and temporal regions have also been reported in Parkinson's dementia and dementia with Lewy bodies, the latter disorder may be distinguished from most cases of Alzheimer's disease by an additional reduction in occipital regions containing primary visual and visual association cortex. Additional studies are required to further differentiate these forms of dementia by the pattern of regional reductions in the metabolic rate for glucose throughout the brain. Although the reductions observed in heteromodal sensory association areas in patients with Alzheimer's dementia appear to be greater than those observed in the hippocampal

formation (which histopathological and structural brain-imaging studies find to be affected in the early stages of Alzheimer's dementia), relatively few studies have examined hippocampal measurements using coregisteredMRIs and reliable methods for characterizing the hippocampal region of interest. The observed PET abnormalities could reflect reduced density, viability, or metabolic function in the terminal neuronal fields that innervate the implicated regions (rather than the cell bodies that arise there), the combined effects of atrophy and partial-volume averaging, or a combination of these factors.

Automated algorithms to generate surface-projection three-dimensional maps appear to be more powerful than visual inspection in the detection of PET abnormalities. As research tools, these still-developing algorithms have been used to demonstrate that cognitively normal persons at risk for Alzheimer's disease have reduced cerebral metabolic rates for glucose in the same brain regions as patients with Alzheimer's disease. If the PET abnormalities prove to be progressive, PET could be used to test the ability of treatments to prevent Alzheimer's disease without having to wait many years to determine whether or when treated persons at risk go on to develop symptoms. Using the same algorithms, preliminary studies suggest that PET is about 94 percent sensitive and 99 percent specific for Alzheimer's disease in patients with questionable or mild dementia. While additional studies are required to confirm the ability of PET to distinguish Alzheimer's disease from other conditions (e.g., Parkinson's dementia and dementia with Lewy bodies) in the early stages of cognitive impairment, provide information about a patient's prognosis, and influence treatment decisions, some clinicians use this technique to confirm the diagnosis in patients who have mild or atypical cognitive impairment, no known reversible causes, and family members who are otherwise reluctant to accept the diagnosis. While PET scans are most commonly performed in the resting state, unconfirmed reports raise the possibility that scans acquired during the performance of visual, audiovisual, memory, or other behavioral tasks might enhance the ability to detect PET abnormalities in the earliest stages of Alzheimer's disease.

All patients with trisomy 21 (Down syndrome) have neuritic plaques by the age of 40, and about 75 percent have Alzheimer's disease by the age of 60. PET studies find that older patients with this disorder have a pattern of reductions in cerebral metabolic rate for glucose similar to that observed in Alzheimer's dementia (e.g., reductions in temporoparietal measurements). While one report suggests that abnormalities in preselected temporoparietal regions of interest are enhanced if the subjects are studied during the performance of an audiovisual task, it would be helpful to compare the sensitivity and specificity of PET to detect characteristic reductions when automated brain mapping algorithms are used to compare data acquired during the resting state, audiovisual tasks, and other behavioral tasks.

SPECT cerebral blood flow studies and still-developing fMRI cerebral blood flow and blood volume studies find that patients with Alzheimer's dementia have a pattern of regional abnormalities similar to that observed in PET studies in the resting state. Additional research is needed to refine MRI methods for the acquisition of quantitative cerebral blood flow images and directly compare the sensitivity and specificity of PET metabolic rate for glucose, SPECT cerebral blood flow, and fMRI cerebral blood flow imaging in the diagnosis of Alzheimer's disease. ^{1}H MRS studies find that patients with Alzheimer's dementia have reduced concentrations of NAA (a marker of intact neurons) in parahippocampal, parietal, temporal, and frontal regions, some of which appear to be correlated with the local density of neuritic plaques and neurofibrillary tangles. A ^{31}P MRS study suggests that patients with Alzheimer's disease have increased levels of phosphomonoesters (membrane phospholipid precursors), increased cerebral metabolic rate for oxygen, and reduced levels of phosphocreatinine and ADP (ATP precursors) during the early stages of dementia of the Alzheimer's type and reductions in each of these measures during the subsequent stages of the illness. A placebo-controlled study by the same group suggests that the compound acetyl-L-carnitine (a compound reported to normalize membrane and energy metabolism and increase the effects of nerve growth factor in animals) normalized phosphomonoester concentration and attenuated cognitive deterioration in a small group of patients with Alzheimer's disease. While this finding requires independent confirmation in larger subject groups, it illustrates the promise of MRS and other imaging techniques in the study of this disorder and its treatment. While MRS promises to complement other research methods, it is not yet clinically indicated in the diagnosis of Alzheimer's disease.

Dementia With Lewy Bodies Though well characterized only recently, dementia with Lewy bodies may be the second most common form of dementia. While this form of dementia shares several clinical features of Alzheimer's disease, it tends to begin at an earlier age, fluctuate in severity, and progress more rapidly, and it is associated with more prominent parkinsonian features, particular sensitivity to the extrapyramidal side effects of antipsychotic medications, and visual hallucinations. Histopathologically, the disorder is distinguished by the presence of Lewy bodies in the cerebral cortex; it also displays Lewy bodies in subcortical regions (as does Parkinson's disease) and neuritic plaques (as does Alzheimer's disease). With or without dementia, patients with diffuse Lewy body disease appear to have reductions in posterior cingulate, parietal, temporal, and prefrontal metabolic rates for glucose similar to those observed in patients with dementia of the Alzheimer's type. However, they appear to be distinguished from patients with Alzheimer's disease (except, perhaps, in those Alzheimer's disease cases associated with Balint's syndrome and other complex visual disturbances) by additional reductions in occipital areas containing primary visual and visual association cortex. Additional research is needed to determine the extent to which PET and SPECT can distinguish dementia of the Alzheimer's type from dementia with Lewy bodies in the earliest stages of the illness.

Dementia Due to Parkinson's Disease Some 20 to 60 percent of patients with Parkinson's disease develop dementia, usually late in the course of the illness. Analyzing preselected regions of interest, dementia due to Parkinson's disease is associated with reductions in parietal and temporal metabolic rates for glucose—indeed, this pattern was shown early in the course of one patient with histologically confirmed dementia due to Parkinson's disease (i.e., Lewy bodies confined to the substantia nigra and no neuritic plaques) prior to the onset of extrapyramidal features, suggesting that this pattern is not specific for Alzheimer's disease. Additional studies using automated brain-mapping algorithms and histopathological confirmation are needed to determine the extent to which PET and SPECT can distinguish Parkinson's disease from Alzheimer's disease and dementia with Lewy bodies in the earliest stages of the illness. PET studies of 6-fluoro-dopa uptake and PET and SPECT studies characterizing the density of dopamine transporter sites reveal abnormal reductions in the striatum of patients with Parkinson's disease. Additional studies are needed to confirm the extent to which these reductions distinguish dementia with Parkinson's disease from Alzheimer's disease and dementia with Lewy bodies and the extent to which these differences might predict outcome or influence treatment decisions.

Pick's Disease and Other Frontotemporal Dementias Whether or not patients have Pick bodies at autopsy, frontotemporal dementias are characterized by declines in executive function, personality, social skills, emotional expression, and language abilities in the early stages of the disorder. Other features of dementia become apparent as the illness progresses. Structural brain-imaging studies find atrophy preferentially affecting the frontal or temporal lobes. When used in conjunction with automated brain-mapping algorithms, PET and SPECT appear to be even more sensitive in the confirmation of frontotemporal dementias. The reductions in frontal and temporal measurements (which are relatively selective in the earlier stages of the illness) could reflect reductions in the density, viability, or metabolism of terminal neuronal fields that innervate these regions, the combined effects of atrophy and partial-volume averaging, or a combination of these factors.

Vascular Dementia Vascular dementia is typically characterized by an abrupt onset, stepwise progression, focal neurological signs and symptoms, imaging evidence of cerebrovascular disease with CT or MRI, risk factors for vascular disease (e.g., hypertension, diabetes, and hypercholesterolemia), and histopathological evidence of cerebrovascular disease obtained from biopsy or autopsy specimens. CT and MRI findings include evidence of multiple large vessel strokes; a single, strategically situated infarct (e.g., in the distribution of the carotid, anterior, middle, or posterior cerebral arteries and affecting certain regions, such as the thalamus and basal forebrain); CT lucencies or MRI T2-hyperintensities in deep white matter and basal ganglia; extensive periventricular white matter lessions (which are not unique to this disorder); or a combination of these features. Using CT, the probability of finding an infarction (i.e., one or more lucent areas) in patients with vascular dementia is highly variable (20 to 86 percent). Using MRI, patients with vascular dementia invariably have one or more T2-hyperintensities in periventricular regions, deep white matter (leukoaraiosis), the basal ganglia, or the thalamus. The clinical significance of the periventricular hyperintensities in the form of caps or rims is questionable; while extensive white matter hyperintensities are commonly found in patients with vascular dementia, some studies suggest that they are present in dementia of the Alzheimer's type, late-onset depressive and psychotic disorders, bipolar disorders, and other conditions. Indeed, the relationship between probable vascular dementia and probable Alzheimer's disease (i.e., diagnoses of exclusion made in the absence of histopathological confirmation) can be confounded by the observations that the disorders sometimes co-exist, that white matter lesions may be present in both disorders (though much more commonly in patients with vascular dementia), and that cerebrovascular disease may be a risk factor for dementia of the Alzheimer's type.

Vascular dementia is typically characterized by a seemingly random pattern of focal reductions in PET images of cerebral blood flow and metabolic rates for oxygen and glucose and in SPECT images of cerebral blood flow. This pattern, which differs from that typically observed in Alzheimer's dementia, might influence decisions about the treatment of cerebrovascular disease. Common treatments include hypertensive therapy, aspirin, and cessation of cigarette smoking. In uncommon cases, treatment may involve a surgical vascularization procedure or, in patients with signs of collagen-vascular disease, aggressive treatment of inflammatory vasculitis. Additional studies are needed to determine how well structural and functional brain-imaging studies correspond to histopathological diagnosis and histopathological characterization of the imaging abnormalities, influence treatment, and predicted outcome.

Other Neurodegenerative Disorders *Focal (asymmetrical) cortical degeneration* is characterized by gradually progressive symptoms such as apraxia, aphasia, agnosia, and cortical blindness. Histopathological features may be consistent with Alzheimer's disease, Pick's disease, or other conditions. Structural brain imaging studies typically find atrophy preferentially affecting the temporal or parietal region to which the symptoms are related. PET cerebral metabolic rate for glucose and SPECT studies find localized reductions in these areas. A study of identical twins who were clinically discordant for asymmetrical cortical degeneration suggests that imaging can detect localized reductions prior to the onset of symptoms and that the pattern can resemble that found in patients with Alzheimer's disease. In patients with *Huntington's disease*, CT and MRI studies find atrophy of the caudate nucleus and whole brain; PET studies find characteristic reductions in the metabolic rate for glucose in the caudate and putamen, which are apparent in genetically predisposed persons prior to the onset of motoric or cognitive symptoms. In later stages of

Huntington's dementia, the metabolic rate for glucose is also reduced in parietal, temporal, and frontal cortex. In patients with *Creutzfeldt-Jakob disease*, structural brain-imaging studies are either normal or show nonspecific atrophy; some PET studies find a patchy pattern of metabolic rate for glucose reductions, while others find asymmetrical reductions in the temporal lobe.

LATE-ONSET DEPRESSIVE DISORDERS

In about two thirds of cases, late-onset depressive disorders may be attributable to an identifiable neurological or medical condition. While structural and functional imaging studies have been used to investigate regions of the brain that are involved in the pathophysiology and treatment of major depressive disorder, they have been used in relatively unique ways to study primary and secondary depressive disorders that occur in late life.

Compared with elderly control subjects, patients with late-onset depressive disorders have an increased number of MRI T2-hyperintensity lesions in deep white matter (especially in the frontal lobe) and the basal ganglia (especially in the caudate nucleus). While this finding has also been reported in patients with bipolar I disorder irrespective of age, it has not been reported in patients with major depressive disorder who develop their first episode at an earlier age, raising the possibility of etiological and pathophysiological distinctions between early-onset and late-onset major depressive disorders. Additional studies are needed to confirm the suggestion that these abnormalities distinguish patients with major depressive disorder who develop their first episode in late life from those who develop their first episode at a younger age; to determine the extent to which patients with an excessive number of T2-hyperintensities differ from other patients with major depressive disorder in their clinical course or response to treatment; and to characterize the histopathological significance of these lesions.

In 20 to 50 percent of cases, patients develop depression following a stroke. Several structural brain-imaging studies find a significant association between poststroke depression and lesion location. Several studies find that poststroke depression is more likely to be associated with lesions in the left dorsolateral prefrontal cortex and left caudate nucleus. These regions have been implicated in functional imaging studies of major depressive disorder and have been postulated to participate in a circuit involved in the pathophysiology and treatment of depression. Several studies of poststroke depression also find a significant relation between the proximity of lesions to the frontal pole and the severity of depression. According to one theory, these lesions interfere with bioaminergic pathways that arise in the brainstem and pass through frontal regions on their way to more posterior regions of the brain.

LATE-ONSET PSYCHOTIC DISORDERS

Late-life psychosis can be attributable to a spectrum of disorders, including early-onset schizophrenia persisting into late-life, late-onset schizophrenia (sometimes called paraphrenia), late-onset delusional disorder (characterized by persistent paranoid delusions), dementias, delirium (from drug use and other causes), cerebrovascular disease in the absence of dementia, and other identifiable neurological disorders. Initial MRI studies of late-onset schizophrenia find an increased number of T2-hyperintensity lesions in deep white matter and basal ganglia regions that cannot be reduced to a specific histopathological lesion or location. Although findings are not entirely consistent, CT and MRI studies tend to find slightly enlarged ventricles in patients with late-onset psychotic disorders, more so for late-onset psychotic disorders other than schizophrenia. Initial SPECT studies find multiple areas of reduced cerebral blood flow in most patients with late-onset psychotic disorders. One PET study found that neuroleptic-naive patients with late-onset schizophrenia had an increased density of dopamine D_2 receptors in the striatum. While the researchers also found an increased density of D_2 receptors in patients with early-onset schizophrenia, this finding was not confirmed in an independent group of patients with early-onset schizophrenia when a different radiotracer technique was used.

Findings from T2-weighted MRI and SPECT studies raise the possibility that microvascular disease contributes to some of the late-onset psychotic disorders. Still, research is needed to further distinguish different late-life psychotic disorders using imaging techniques to identify distinctive structural and functional brain abnormalities in patients with these disorders and determine how these abnormalities are related to the patients' clinical course, response to treatment, and histopathology.

CLINICAL INDICATIONS

As part of the laboratory dementia workup, a structural brain image is routinely indicated to rule out potentially treatable, space-occupying lesions such as brain tumors (e.g., frontal lobe meningiomas and temporal lobe gliomas), cerebrovascular lesions (e.g., subdural hematomas, other hemorrhages, infarcts, and pathological white matter lesions), a chronic meningitic process (as suggested by meningeal enhancement), normal pressure hydrocephalus (suggested by pronounced ventriculomegaly without commensurate cortical atrophy), and a prior traumatic encephalomalacia. A structural brain image can also be used to identify clinically relevant patterns of focal atrophy, supporting the diagnosis of certain neurodegenerative disorders (e.g., frontotemporal dementia and asymmetric cortical degeneration syndromes), and it can be used to provide evidence for vascular dementia. While CT is less expensive and less confining, MRI offers several advantages including superior sensitivity in detecting lesions in white matter and meninges and the ability to study brain regions in the posterior fossa and brainstem without artifacts from bone.

Structural and functional brain-imaging studies find significant differences between groups of patients with dementia of the Alzheimer's type and normal controls, but these techniques are not routinely used to confirm this diagnosis. It remains to be established how well these techniques improve the clinician's ability to detect Alzheimer's disease in an individual patient with mild cognitive impairment (when the clinical diagnosis is least reliable), distinguish this disorder from normal aging and other forms of cognitive impairment (e.g., dementia with Lewy bodies), predict a patient's clinical course and postmortem histopathology, and make decisions about treatment.

While the sensitivity, specificity, and predictive value of imaging techniques in the diagnosis of Alzheimer's disease remain to be established, some groups have begun to use either PET cerebral metabolic rate for glucose or SPECT cerebral blood flow images in conjunction with automated brain-mapping algorithms with improved power to identify characteristic abnormalities in the diagnostic evaluation of patients with mild or unusual forms of cognitive impairment and no known reversible causes. To the extent that these techniques identify patterns that are consistent with the diagnosis of Alzheimer's disease, they may help family members who are reluctant to accept this catastrophic diagnosis take the proactive steps needed to maximize the patients' safety, address their financial and legal needs, and ensure future access to assisted living programs.

In addition, some clinicians use either PET metabolic rate with glucose or SPECT cerebral blood flow images to support the diagnosis of vascular dementia (e.g., identify a patchy pattern of focal reductions), which could affect decisions about risk factors for cerebrovascular disease. Some clinicians use PET metabolic rates for glucose and oxygen or SPECT cerebral blood flow images in conjunction with automated brain-mapping algorithms to support the diagnosis of certain neurodegenerative disorders other than Alzheimer's disease (e.g., frontotemporal dementia and asymmetrical cortical degeneration), which might benefit from unique behavioral management strategies. Although its ability to predict outcome, correspond to histopathology, or influence treatment decisions remains to be established, some clinicians recommend PET 6-fluoro-dopa, PET dopamine transporter site, or SPECT transporter site images to support the diagnosis of Parkinson's disease in its earliest stages.

While structural and functional brain-imaging studies find statistical differences between groups of patients with psychiatric disorders other than dementia (e.g., depression, schizophrenia) and normal controls, these techniques are not yet indicated in the clinical diagnosis of these disorders. It remains to be established how well these techniques distinguish the psychiatric disorder of interest from other disorders (e.g., distinguish depressive pseudodementia from other forms of dementia), predict clinical course, correspond to postmortem histopathology, or influence decisions about treatment.

FUTURE DIRECTIONS

Improvements in image-acquisition and image-analysis techniques promise to improve the ability of researchers to characterize the brain regions, pathways, and chemical processes that are affected during normal aging and those that are involved in pathophysiology of age-related disorders and to strengthen the role of imaging techniques in the diagnosis of dementia of the Alzheimer's type and other dementias. By providing surrogate markers of illness in intervention studies, imaging techniques promise to help distinguish treatments that slow the progression of Alzheimer's disease and other disorders in patients with dementia from those that compensate for these disorders and to help establish the efficacy of prevention treatments in healthy persons at genetic risk for these disorders. As a complement to other neuroscientific studies, imaging techniques can potentially help to clarify how identified brain abnormalities are related to specific behavioral features, clinical course, response to treatment, and histopathology in geriatric patients with psychiatric disorders.

SUGGESTED CROSS-REFERENCES

The basic principles of neuroimaging techniques, including CT, MRI, PET, SPECT, MRS, and fMRI, are described in Section 1.15 and Section 1.16. The application of neuroimaging techniques to the clinical evaluation of dementia, other degenerative disorders, vascular disease, head trauma, tumors, and epilepsy is discussed in Section 2.13. Electrophysiological recording techniques, including electroencephalography (EEG), quantitative EEG, event-related potentials (ERP), and magnetoencephalography (MEG) are discussed in Section 1.9. Functional neuroanatomy is discussed in Section 1.2. The use of CT and MRI is further discussed in Section 51.2f.

SECTION REFERENCES

Bohnen NI, Minoshima S, Kuhl DE, Frey KA: The roles of FDG PET and MRI in the diagnosis of Alzheimer's disease. Neurosci News *1:*26, 1998.

Botteron KN, Figiel GS: The neuromorphometry of affective disorders. In *Brain Imaging in Clinical Psychiatry*, KR Krishnan, PM Doraiswamy, editors. Marcel Dekker, New York, 1996.

*Coffey CE: Anatomic imaging of the aging human brain: Computed tomography and magnetic resonance imaging. In *Textbook of Geriatric Neuropsychiatry*, CE Coffey, JL Cummings, editors. American Psychiatric Press, Washington, DC, 1994.

*Doder M, Jahanshahi M, Turjanski N, Moseley IF, Lees AJ: Parkinson's syndrome after closed head injury: A single case report. J Neurol Neurosurg Psychiatry *66:*380, 1999.

Duara R: Neuroimaging with CT and MRI in Alzheimer disease. In *Alzheimer Disease*, RD Terry, R Katzman, KL Bick, editors. Raven, New York, 1994.

Figiel GS, Krishnan KRR, Doraiswamy M, Rao VP, Nemeroff CB, Boyko OB: Subcortical hyperintensities on brain magnetic resonance imaging: A comparison between late age onset and early onset elderly depressed subjects. Neurobiol Aging *12:*245, 1991.

*Forstl H, Hentschel F: Electrophysiology and neuroimaging. In *Psychiatry in the Elderly*, ed 2, R Jacoby, C Oppenheimer, editors. Oxford University Press, Oxford, 1997.

Harris GJ, Lewis RF, Satlin A, English CD, Scott TM, Yurgelun-Tood DA, Renshaw PF: Dynamic susceptibility contrast MRI of regional cerebral blood volume in Alzheimer's disease. Am J Psychiatry *153:*721, 1996.

Hatazawa J, Uemura K: The aging brain. In *Principals of Nuclear Medicine*, ed 2, HN Wagner, Z Szabo, JW Buchanan, editors. Saunders, Philadelphia, 1995.

Heiss WD, Podreka I: Cerebrovascular disease. In *Principals of Nuclear Medicine*, ed 2, HN Wagner, Z Szabo, JW Buchanan, editors. Saunders, Philadelphia, 1995.

Hoffman JM: Positron emission tomography studies in dementia. In *Brain Imaging in Clinical Psychiatry*, KR Krishnan, PM Doraiswamy, editors. Marcel Dekker, New York, 1996.

Keshavan MS, Kapur S, Pettegrew JW: Magnetic resonance spectroscopy in psychiatry: Potential, pitfalls, and promise. Am J Psychiatry *148:*976, 1991.

Krishnan KRR: Neuroanatomic substrates of depression in the elderly. J Geriatr Psychiatry Neurol *6:*39, 1993.

Kumar A, Newberg A, Alavi A, Berlin J, Smith R, Reivich M: Regional cerebral glucose metabolism in late-life depression and Alzheimer's disease: A preliminary positron emission tomography study. Proc Natl Acad Sci USA *90:*7019, 1993.

Lesser IM, Miller BL, Swartz JR, Boone KB, Mehringer CM, Mena I: Brain imaging in late-life schizophrenia and related psychoses. Schizophr Bull *19:*773, 1993.

Margolin R, Moran P: Neuroimaging. In *Comprehensive Review of Geriatric Psychiatry—II*, ed 2, J Sadavoy, LW Lazarus, SF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1997.

*Minoshima S, Frey K, Koeppe RA: A diagnostic approach in Alzheimer's disease using three-dimensional stereotactic surface projections of fluorine-18-FDG PET. J Nucl Med *36:*1238, 1995.

Muller-Gartner HW: Dementia. In *Principles of Nuclear Medicine*, ed 2, HN Wagner, Z Szabo, JW Buchanan, editors. Saunders, Philadelphia, 1995.

Nasrallah HA, Pettegrew JW, editors: *NMR Spectroscopy in Psychiatric Brain Disorders*. American Psychiatric Press, Washington, DC, 1995.

*Pearlson GD, Petty RG: Late-life-onset psychoses. In *Textbook of Geriatric Neuropsychiatry*, CE Coffey, JL Cummings, editors. American Psychiatric Press, Washington, DC, 1994.

Pearlson GD, Tune LE, Wong DF, Aylward EH, Barta PE, Powers RE, Tien AY, Chase GA, Harris GJ, Rabins PV: Quantitative D_2 dopamine receptor PET and structural changes in late-onset schizophrenia. Schizophr Bull *19:*783, 1993.

Pettegrew JW, Klunk W, Panchalingam K, Kanfer JN, McClure RJ: Clinical and neurochemical effects of acetyl-L-carnitine in Alzheimer's disease. Neurobiol Aging *16:*1, 1995.

*Reiman EM, Caselli RJ, Yun LS, Chen K, Bandy D, Minoshima S, Thibodeau S, Osborne D: Preclinical evidence of a genetic risk factor for Alzheimer's disease in apolipoprotein E type 4 homozygotes using positron emission tomography. N Engl J Med *334:*752, 1996.

Reiman EM, Uecker A, Caselli RJ, Lewis S, Bandy D, de Leon MJ, De Santi S, Convit A, Osborne D, Weaver A, Thibodeau SN: Hippocampal volumes in cognitively normal persons at genetic risk for Alzheimer's disease. Ann Neurol *44:*288, 1998.

Sandson TA, O'Connor M, Sperling RA, Edelman RR, Warach S: Noninvasive perfusion MRI in Alzheimer's disease: A preliminary report. Neurology *47:*1339, 1996.

*Simpson S, Baldwin RC, Jackson A, Burns A: The differentiation of DSM-III-R psychotic depression in later life from nonpsychotic depression: Comparisons of brain changes measured by multispectral analysis of magnetic resonance brain images, neuropsychological findings, and clinical features. Biol Psychiatry *45:*193, 1999.

Small GW, Mazziotta JC, Collins MT, Baxter LR, Phelps ME, Mandelkern MA, Kaplan A, LaRue A, Adamson CF, Chang L: Apolipoprotein E type 4 allele and cerebral glucose metabolism demonstrated by positron emission tomography in individuals at risk of familial Alzheimer's disease. JAMA *273:*942, 1995.

Steffens DC: MRI and MRS in dementia. In *Brain Imaging in Clinical Psychiatry*, KR Krishnan, PM Doraiswamy, editors. Marcel Dekker, New York, 1996.

Symonds LL, Jeste DV: Neuroimaging in late-life psychosis. In *Brain Imaging in Clinical Psychiatry*, KR Krishnan, PM Doraiswamy, editors. Marcel Dekker, New York, 1996.

Tedeschi G, Bertolino A, Lundbom N, Bonavita S, Patronas NJ, Buyn JH, Metman LV, Chase TN, Di Chiro G: Cortical and subcortical chemical pathology in Alzheimer's disease as assessed by multislice proton magnetic resonance spectroscopic imaging. Neurology *47:*696, 1996.

Wong DF, Wagner HN, Dannals RF, Links JM, Frost JJ, Ravert HT, Wilson AA, Rosenbaum AE, Gjedde A, Douglass KH, Petronis JD, Folstein MF, Toung JKT, Burns HD, Kuhar MJ: Effects of age on dopamine and serotonin receptors measured by positron tomography in the living human brain. Science *226:*1393, 1984.

Zubenko GS, Sullivan P, Nelson JP, Belle SH, Huff J, Wolf GL: Brain imaging abnormalities in mental disorders of late life. Arch Neurol *47:*1107, 1990.

51.2 ASSESSMENT

51.2F NEUROIMAGING: SPECIAL ISSUES

ANAND KUMAR, M.D.

The development of sophisticated imaging techniques such as computerized tomography (CT) and magnetic resonance imaging (MRI) permit examination of the structural correlates of the major mental disorders in elderly adults. These techniques together with magnetic resonance spectroscopy (MRS) can provide meaningful information on neurochemistry and neurotransmitter functions as they relate to major psychiatric disturbances. In addition to cross-sectional information, they also provide an opportunity to study brain function longitudinally and to examine the relations between clinical phenomena and anatomical and biochemical perturbations. Neuroanatomical indexes may also serve as early preclinical markers of disease and thereby provide a more holistic picture of human neurobiology.

METHODOLOGY

CT provides an example of the early use of computers to reconstruct cross-sectional images of the body from information gathered by a narrow beam of X-rays. The X-ray beam enters the edge of the cross-section of the body being imaged. After traversing the body the emerging X-rays are counted, and the number lost during passage through the body is calculated. During a scan, the position of the beam is changed many times, so that any one cross-section of the body may be examined many times from several different directions. The computer then reconstructs the density of the tissue in the cross-section being examined. A CT image thus reflects tissue density, which can be quantified by the computer and obtained as CT attenuation numbers (also referred to as density numbers).

MRI uses the magnetic properties of living tissues to produce an image. Since hydrogen has the most magnetic nucleus and makes up two thirds of the atoms in living tissue, it is largely the hydrogen in tissue water that is imaged. The hydrogen nucleus has only one proton; consequently, imaging of hydrogen is often referred to as proton imaging. In clinical MRI, the protons are aligned by placing the subject in a strong magnetic field. Tissues are exposed to a single or multiple excitation pulses (radio waves), which causes the hydrogen atoms to absorb energy and precess (spin) around their common axes. The protons then release energy and return to their resting orientation. The released energy is the signal that forms the basis of an MR image. A standard MRI scan provides an image of hydrogen distribution in the organ being examined and it displays a slice of tissue as regions of varying brightness. Typically two or more pulses are used for better characterization of tissues. By varying the time interval between two successive pulse cycles (TR) and the interval from a pulse to the measurement of the MR signal (TE), one can obtain MR images with different tissue contrasts (T1- and T2-weighting and proton density). A T1-weighted image, in which cerebrospinal fluid (CSF) appears black, provides better gray:white parenchymal discrimination, while a T2-weighted image (in which CSF is bright) helps distinguish brain parenchyma from CSF. Proton density images provide tissue contrasts of intermediate brightness.

Early studies using CT to examine cerebral anatomy focused on indirect measures such as ventricular-brain ratios (a measure of CSF volume relative to brain parenchymal volume) and CT attenuation numbers as indicators of parenchymal damage. The increased spatial resolution and sensitivity of MRI permitted more-direct study of neuroanatomy. Quantitative estimates of total and hemispheric brain and CSF volumes provided global estimates of neuroanatomical integrity. Regional measures of neocortical and subcortical structures can be obtained by MRI, and these indexes facilitate the study of more anatomical aberrations in psychiatric disorders. The increased sensitivity of MRI also allows qualitative and quantitative examination of the high-intensity-signal lesions that are typically found in the periventricular, deep white matter, and subcortical regions of the brain. These lesions appear as bright spots on conventional T2-weighted MR brain images and were previously referred to as unidentified bright objects (UBOs). These measurements, together with segmentation techniques, make possible individual examination of the structural integrity of gray and white matter, thereby providing the clinical neuroscience disciplines with the necessary tools for more precise examination of the relation between anatomical perturbations and major mental disorders.

NEUROANATOMIC CHANGES WITH NORMAL AGING

Before considering changes in cerebral anatomy that may be associated with specific psychopathological states, one must first appreciate the neuroanatomical changes that occur in aging per se, unassociated with any mental disorder. Several studies have reported MRI-determined structural changes associated with normal aging. These studies demonstrate an increase in whole brain CSF and high-intensity-lesion volumes with increasing age. Aging is also associated with decreased volumes of cerebral hemispheres, frontal and temporal lobes, associated with increased ventricular volumes. Volumes of specific diencephalic structures and the caudate and lenticular nuclei have also been reported to decrease with age (Table 51.2f-1). Sex differences in brain structure have been reported by several (though not all) groups using volumetric MRI. A few studies have found that the brains of men age faster or earlier than those of women. Sex differences in aging may be apparent at both a hemispheric and a lobar level, with the frontal lobes showing more marked sex differences with aging (with men showing more atrophy) than the temporal lobe.

Condition	Observations	Caveats
Healthy aging	Increase in volume of CSF and high-intensity lesions Decrease in brain volumes	Gender differences present; structural changes amplified by superimposed disease
Mood disorders	Smaller brain volumes Larger lesion and CSF volumes than controls	Differences depend on health status of control group; lesion volume correlates with overall medical burden
Alzheimer's disease	Progressive atrophy; early changes detected in the hippocampus Atrophy widespread	Lesion volume increases with cerebrovascular risk factors; despite consistent group differences, diagnostic utility is minimal
Psychoses	Larger CSF and ventricular volumes together with lesion volumes	Atrophic changes more striking than those observed in younger patients with psychoses

Table 51.2f-1 Summary of Findings in Late-Life Mental Disorders

NEUROIMAGING IN LATE-LIFE DEPRESSION

Computerized Tomography Studies The application of CT to the study of brain disorders has permitted examination of neuroanatomical aberrations that occur with aging and those associated with specific psychiatric disorders. CT studies demonstrated relative abnormalities in CT attenuation numbers (suggesting altered parenchymal density and consequently brain pathology) in patients with late-life major depressive disorder, compared with controls. Some studies suggest that CT density counts in patients with major depressive disorder lie midway between nondepressed elderly controls and patients with probable dementia of the Alzheimer's type. A subgroup of depressive persons with cognitive impairment had CT counts that approached those of patients with dementia of the Alzheimer's type. Some investigators correlated abnormal CT numbers with poor prognosis and mortality in patients with major depressive disorder. Other groups examining semiquantitative measures, such as ventricular-brain ratios, demonstrated global abnormalities in subjects with late-life depression, compared with nondepressed controls. While CT provided an early tool for examining neuroanatomical abnormalities, MRI permits examination of neuroanatomy with much more precision than previously possible.

Superior resolution and increased sensitivity to brain changes have made MRI extremely useful for psychiatric research.

MRI Studies—High-Intensity Signals in Depression High-intensity lesions have received considerable scrutiny in psychiatric research. Earlier reports using MRI in late-life depression focused on the presence and severity of high-intensity signals in the brain in T2-weighted MR images. Elderly patients with major depression exhibited more frequent and more severe high-intensity signals in the periventricular, deep white matter, and subcortical regions of the brain than nondepressed controls. Some of these studies focused on depressed inpatients and those referred for electroconvulsive therapy (ECT) for failure to respond to antidepressant trials. Qualitative ratings were used to estimate high-intensity-lesion volumes in patients and controls in most of the early studies. Despite these caveats, the increased severity of these lesions in patients diagnosed with major depressive disorder (compared with controls) was widely interpreted as evidence supporting a “structural” basis for depression in late life.

Several investigators have examined the clinical and biological correlates of high-intensity signals in humans. Results from community-based studies, such as the Cardiovascular Health Study, indicate that the most frequent correlates of these lesions include increasing age, prior stroke, hypertension, and forced expiratory volume in 1 second (FEV_1). The prevalence of high-intensity signals in the brain has been consistently low in studies that screened subjects to exclude those with major vascular risk factors. However, not all studies find strong correlations between high-intensity signals and specific vascular risk factors such as hypertension and diabetes in depression. It therefore appears likely that other, yet unidentified, biological factors may also be responsible for these lesions. Pathologically, they reflect a spectrum of changes, including arteriolar ectasia and myelin pallor associated with arteriosclerosis of perforating arterioles, more pronounced in the subcortical white matter. The location of these lesions together with their frequent pathological correlates suggests that white matter compromise from medical or vascular disease may contribute to late-life mood disorders. Some investigators have argued that high-intensity signals represent silent infarcts of the brain in depression. The increased severity of high-intensity signals in major depressive disorder and their relation to cerebrovascular risk factors have led other investigators to suggest a vascular etiology for late-life depression. While a vascular component to major depressive disorder remains plausible, evidence supporting this claim remains largely indirect at the present time. These complexities should be seriously considered in designing and interpreting the results of neuroimaging studies in mental disorders in late life. Interpretations of data on specific psychopathologic states should always be tempered by comparisons with identical measures obtained from control subjects matched for age, gender, and appropriate medical comorbidity.

Volumetric studies of the brain in late-life depression have focused on both global and regional measures of brain and CSF volumes. Global measures examined include total brain volume, ventricular volume, and whole-brain CSF volume. Comparisons have typically been made between major depressive disorder patients with several comorbid medical disorders and relatively healthy elderly control subjects recruited from the community. An earlier study from the author's laboratory demonstrated that in a sample of patients with major depressive disorder and diverse medical comorbidity, whole-brain CSF volume and total ventricular volumes were significantly larger than those of healthy nondepressed controls recruited from the community. In that study, high whole-brain CSF volume (normalized to intracranial volume) and overall medical comorbidity (quantified using the cumulative illness rating scale) greatly increased the odds of patients developing major depressive disorder.

Other investigators have reported smaller frontal lobe volumes in samples of mixed-age depressive patients referred for ECT. Studies have also demonstrated smaller caudate volumes in a mixed-age sample of patients diagnosed with major depressive disorder, which has been interpreted as evidence of a disruption to the caudate-prefrontal pathway in depression. Y. I. Sheline and coworkers reported smaller hippocampal volumes in a small sample of medically healthy women in remission from major depressive disorder than in age- and sex-matched controls, but other investigators find no consistent neuroanatomical differences between depressed patients and controls. A recent study by the author's group, focusing on a group of elderly patients with major depressive disorder and nondepressed elderly controls better matched for overall medical comorbidity, demonstrated that smaller prefrontal lobe brain volumes and greater high-intensity-signal-lesion volumes independently increased the odds of developing depression. These findings suggest that atrophy and high-intensity lesions may represent complementary pathways to late-life depression. The group also reported on the neuroanatomical correlates of patients with late-onset minor depression diagnosed using modified research criteria from the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) for minor depressive disorder. Patients with minor depressive disorder were compared with an age-matched group of nondepressed controls with comparable medical comorbidity. The group with minor depressive disorder had smaller normalized prefrontal lobe volumes than the controls. In addition, prefrontal lobe volumes were linearly related to severity of illness (i.e., volumes decreased from the controls to patients with minor depressive disorder to those with major depressive disorder). While issues of nosology and clinical definitions of major versus minor depressive disorders complicate interpretation of these results, the data suggest a general neuroanatomical basis for mood disorders in late life that transcends traditional diagnostic boundaries. Collectively these results indicate that MRI detects structural correlates to most forms of mood disorders in late life. The pathways leading from neuroanatomical changes to mood disorders are likely to vary in different clinical populations and need better elucidation.

PSYCHOTIC DISORDERS OF LATE LIFE

There has been a resurgence of interest in examining the clinical and biological characteristics of psychosis occurring in late life. Nosologic disagreements and overlap between conditions such as late-onset schizophrenia, late paraphrenia, and other psychotic states in late life have limited the interpretation of results in certain instances. Several investigators have reported higher ventricular-brain ratios in patients with late-onset schizophrenia than in age-matched control subjects. More recently, a quantitative MRI study of patients with late-onset schizophrenia reported that these patients had larger ventricular volumes than controls. In this study, thalamic volumes were larger in the late-onset group than in elderly schizophrenia patients with an earlier onset. Other investigators have reported that patients with late paraphrenia have larger lateral ventricular volumes than both elderly patients with schizophrenia and nonpsychotic age-matched controls. They argue that paraphrenia may have a more striking neuroanatomical basis than other psychotic disorders in late-life. They additionally assert that part of the confusion regarding the neuroanatomical correlates of psychotic disorders in late life is an inability to distinguish clinically between late paraphrenia and schizophrenia in elderly patients. More recently, investigators using MRI to examine patients with all forms of late-onset psychoses reported an increased prevalence of high-intensity signals in neocortical, subcortical, and brainstem regions compared with age-matched controls without psychosis. In summary, patients with late-onset psychoses (both schizophrenia and other psychotic states) also display neuroanatomical abnormalities on CT and MRI (Table 51.2f-1). These abnormalities are also nonspecific, and the mechanisms by which they contribute to psychoses remain unclear.

NEUROIMAGING STUDIES IN DEMENTIA

Dementia of the Alzheimer's type has been studied quite extensively by both CT and MRI. The findings from neuroimaging studies in dementia closely parallel the more established pathological hallmarks of the disorder. Cross-sectional neuroimaging studies demonstrate smaller brain volumes and increased CSF volumes in these patients than in age-matched controls without dementia. Longitudinal studies using CT demonstrate progressive ventricular enlargement in patients with dementia, significantly greater than the minimal progressive dilatation seen in nondemented controls. Also, while ventricular enlargement was noted relatively early in the course of the illness (when memory impairment dominates the clinical picture), the rate of enlargement appears to increase substantially when other nonmemory cognitive aspects of the dementia become clinically apparent. In addition to the more global measures of atrophy, investigators have demonstrated focal volumetric reductions in both medial and lateral temporal lobe structures in patients with dementia (compared with controls). Even in patients with mild dementia, hippocampal volume reductions between patients and controls were striking. Other reports indicate that volumes of the left amygdala and the entorhinal cortex best discriminate patients with dementia of the Alzheimer's type from controls. In an effort to examine the neuroanatomical correlates of patients in the predementia stage, one study found that patients with minimal cognitive impairment and CT evidence of hippocampal atrophy progressed to clinical dementia on follow-up. Thus focal hippocampal atrophy may serve as a marker for dementia of the Alzheimer's type. Focal neuroanatomical perturbations captured by CT and MRI early in the course of the disease may reflect the early pathological changes that occur in circumscribed brain regions in the limbic/mesial temporal areas. Progression of Alzheimer's disease is clearly associated with progressive neuronal loss, and the neuroimaging evidenceclearly corroborates these changes.

High-Intensity Signal Changes MRI studies have demonstrated high-intensity signals in patients with dementia of the Alzheimer's type in the periventricular, subcortical, and other neocortical regions of the brain. The prevalence of high-intensity signals in dementia of the Alzheimer's type was found to be low and comparable to that in controls in selected samples of patients prescreened to exclude those with major cerebrovascular risk factors. Research also suggests that in certain patient groups, neuropsychological performance may be more impaired in patients with dementia of the Alzheimer's type who have these lesions than in those without them. High-intensity signals may attenuate the relation between neuropsychological functions and glucose metabolism that is typically seen in dementia of the Alzheimer's type. Other investigators found no difference in either cognitive or behavioral features in these patients with or without high-intensity signals. One recent study showed that reduced gray matter volume and an increase in abnormal white matter high-intensity signals independently compromised cognitive functioning in patients with dementia of the Alzheimer's type. Collectively, these data appear to suggest that the presence of high-intensity signals in patients with dementia of the Alzheimer's type indicates an additional pathophysiological process that is predominantly vascular. Some investigators assert that these lesions exert their effects by disrupting corticocortical connections in the brain.

Pathologic and Behavioral Correlates The frequent pathological correlates of MRI high-intensity signals are demyelination, white matter ischemia, and edema in deep white matter and periventricular regions of the brain. These pathological descriptors were obtained from studies of the brains of patients who died from a variety

of neurological and nonneurological illnesses. These findings suggest that a variety of neuropathological and biochemical processes can result in brain changes that are captured by MRI as high-intensity (bright) signals in different brain regions. While high-intensity signals in the periventricular regions may reflect the transudation of ventricular CSF into the surrounding white matter, vascular factors may be responsible for these signals in the deep white matter and subcortical regions. Though these signals do not necessarily provide any diagnostic specificity, they do provide information about possible underlying pathological changes that lead to their development. The neurochemical correlates of high-intensity signals are unknown and have not been investigated in a direct, reproducible manner. There is also some disagreement about the behavioral consequences of these signals in relatively healthy elderly subjects and patients with dementia of the Alzheimer's type. Some investigators have reported that large high-intensity lesions are associated with reduced frontal glucose metabolism and impaired frontal lobe–mediated neuropsychological functions in healthy elderly subjects. Others, however, find no specific cognitive compromise attributable to high-intensity signals in either healthy elderly populations or patients with dementia of the Alzheimer's type. Therefore, while high-intensity signals are consistently associated with certain risk factors in elderly subjects, their behavioral impact remains controversial.

FUTURE DIRECTIONS

High-resolution techniques such as MRI have made it possible to study the anatomical and functional integrity of specific brain structures believed to be responsible for regulating normal and abnormal human behavior, which has helped advance our understanding of the biological basis of certain mental disorders. Several neuroanatomical correlates have been identified in elderly patients with major mental disorders. Many of these abnormalities, however, are nonspecific and may reflect a spectrum of pathological changes not pathognomonic of any single clinical entity. Neuroimaging studies can potentially yield useful information on the mechanisms that lead to major mental disorders in elderly adults. Therefore these technologies are better conceptualized as powerful research tools for the clinical neurosciences than as diagnostic instruments in clinical psychiatry.

Newer techniques such as MRS and functional MRI (fMRI) have rendered historical distinctions between neuroanatomical and physiological imaging somewhat obsolete. MRS can potentially provide information on neurotransmitters and important neuronal markers such as *N*-acetylaspartate, choline, lecithin, creatine, and glutamate. Phosphorous spectroscopy can help us study the role of several metabolites such as adenosine triphosphate and diphosphate in focal brain regions. These tools should be used in carefully designed human studies with appropriate clinical and control groups. In vivo pharmacological challenges specifically designed to probe neurotransmitter systems may elucidate the relation between neuroanatomical abnormalities and dysfunction in individual neurochemical systems. Postmortem MRI together with neurochemical and neuroanatomical analysis of brain tissue in patients with well-characterized antemortem psychiatric and medical illness may further clarify some biological issues relevant to mental disorders in elderly patients. In conclusion, the early results are encouraging and certainly suggest a neuroanatomical basis for late-life mental disorders. The next generation of studies must be more focused and should further probe the role of specific neurochemical and neurobiological systems in depression.

SUGGESTED CROSS-REFERENCES

Section 2.13 discusses neuroimaging in clinical practice. Section 51.2b discusses central nervous system changes in the aging brain. Section 51.3 covers psychiatric disorders of late life. Section 51.2e provides an overview of neuroimaging in geriatric patients.

SECTION REFERENCES

Alexopoulas GS, Young RC, Shindledecker RD: Brain computed tomography findings in geriatric depression and primary degenerative dementia. Biol Psychiatry *31:*511, 1992.

Almkvist O, Wahlund L, Andersson-Lundman T, Basun H, Backman L: White-matter hyperintensity and neuropsychological functions in dementia and healthy aging. Arch Neurol *49:*626, 1992.

Awad IA, Spetzler RF, Hodak JA, Awad CA, Carey R: Incidental subcortical lesion identified on magnetic resonance imaging in the elderly: I. Correlation with age and cerebrovascular risk factor. Stroke *17:*1084, 1986.

Boyko OB, Alston SR, Fuller GN, Hulette CM, Johnson GA, Burger PC: Utility of postmortem magnetic resonance imaging in clinical neuropathology. Arch Pathol Lab Med *118:*219, 1994.

Braffman BH, Zimmerman RA, Trojanowski JQ, Gonatas NK, Hickey WF, Schlaepfer WW: Brain MR: Pathologic correlation with gross and histopathology. 2. Hyperintense white-matter foci in the elderly. AJNR *9:*629, 1988.

*Camicioli R, Moore MM, Sexton G, Howieson DB, Kaye JA: Age-related brain changes associated with motor function in healthy older people. J Am Geriatr Soc *47:*330, 1999.

Coffey CE, Wilkinson WE, Weiner RD, Parashos IA, Djang WT, Webb CM, Figiel CS, Spritzer CE: Quantitative cerebral anatomy in depression. A controlled magnetic resonance imaging study. Arch Gen Psychiatry *50:*7, 1993.

Corey-Bloom J, Jernigan T, Archibald S, Harris MJ, Jeste DV: Quantitative magnetic resonance imaging of the brain in late-life schizophrenia. Am J Psychiatry *15:*447, 1995.

Cowell PE, Turetsky BI, Gur RC, Grossman RI, Shtasel D, Gur R: Sex differences in aging of the human frontal and temporal lobes. J Neurosci *14:*4748, 1994.

DeCarli C, Murphy DGM, Tranh M, Grady CL, Haxby JV, Gillette JA, Salerno JA, Gonzales-Aviles A, Horowitz B, Rapoport SI, Schapiro MB: The effect of white matter hyperintensity volume on brain structure, cognitive performance, and cerebral metabolism of glucose in 51 healthy adults. Neurology *45:*2077, 1995.

*DeLeon MJ, Golomb J, George AE, Convit A, Tarshish C, McRae T, DeSanti S, Smith G, Ferris SH, Nox M, Rusinek H: The radiologic prediction of Alzhiemer's disease: The atrophic hippocampal formation. AJNR Am J Neuroradiol *14:*897, 1993.

Fazekas F, Chawluk JB, Alavi A, Hurtig HI, Zimmerman RA: MR signal abnormalities at 1.5 T in Alzheimer's dementia and normal aging. AJNR *8:*421, 1987.

Fein G, Van Dyke C, Davenport L, Turetsky B, Brant-Zawadzki M, Zatz L, Dillon W, Valk P: Preservation of normal cognitive functioning in elderly subjects with extensive white-matter lesions of long duration. Arch Gen Psychiatry *47:*22, 1990.

Greenwald BS, Kramer-Ginsberg E, Krishnan KRR: MRI signal hyperintensities in geriatric depression. Am J Psychiatry *153:*1212, 1996.

Gur RC, Mozley PD, Resnick SM, Gottlieb G, Kohn M, Zimmerman R, Herman G, Atlas S, Grossman R, Berretta D, Erwin R, Guin RE: Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. Proc Natl Acad Sci USA *8:*2845, 1991.

Hounsfield GN: Nobel Award address. Computed medical imaging. Med Phys *7:*283, 1980.

Howard R, Almeida O, Levy R, Graves P, Graves M: Quantitative magnetic resonance imaging volumetry distinguishes delusional disorder from late-onset schizophrenia. Br J Psychiatry *165:*474, 1994.

Jernigan TL, Archibald SL, Berhow MT, Sowell ER, Foster DS, Hesselink JR: Cerebral structure on MRI, Part I: Localization of age-related changes. Biol Psychiatry *29:*55, 1991.

Jernigan TL, Press GA, Hesselink JR: Methods for measuring brain morphologic features on magnetic resonance images: Validation and normal aging. Arch Neurol *47:*27, 1990.

Killiany RJ, Moss MB, Albert MS, Sandor T, Tierman J, Jolesz F: Temporal lobe regions on magnetic resonance imaging identify patients with early Alzheimer's disease. Arch Neurol *50:*949, 1993.

*Krishnan KRR: Neuroanatomic substrates of depression in the elderly. J Geriatr Psychiatry Neurol *6:*39, 1993.

Kumar A, Yousem D, Souder E: High intensity signals in Alzheimer's disease without cerebrovascular risk factors: A magnetic resonance imaging evaluation. Am J Psychiatry *149:*248, 1992.

*Kumar A, Miller D, Ewbank D, Yousem D, Newberg A, Samuels S, Cowell P, Gottlieb G: Quantitative anatomic measures of comorbid medical illness in late-life major depression. Am J Geriatr Psychiatry *5:*15, 1997.

Kumar A, Schweitzer E, Zhisong J, Miller D, Bilker W, Swan LL, Gottlieb G: Neuroanatomical correlates of late-onset minor depression. Arch Neurol *54:*613, 1997.

*Kumar A, Zhisong J, Bilker W, Udupa J, Gottlieb G: Late-onset minor and major depression: Early evidence for common neuroanatomical substrates detected by using MRI. Proc Natl Acad Sci USA *95:*7654, 1998.

*Lesser IM, Miller BL, Swartz JR, Boone KB, Mehringer CM, Mena I: Brain imaging in late-life schizophrenia and related psychoses. Schizophr Bull *19:*773, 1993.

Longstreth WT, Manolio TA, Arnold A, Burke GL, Bryan N, Jungreis CA, Enright PL, O'Leary D, Fried L: Clinical correlates of white matter findings on cranial magnetic resonance imaging of 3301

elderly people: The cardiovascular health study. Stroke *27:*1274, 1996.

Lopez OL, Becker JT, Rezek D, Wess J, Boller F, Reynolds CF, Panisser M: Neuropsychiatric correlates of cerebral white-matter radiolucencies in probable Alzheimer's disease. Arch Neurol *49:*828, 1992.

Miller D, Kumar A, Yousem DM, Gottlieb GL: MRI high intensity signals in late-life depression and Alzheimer's disease. A comparison of subjects without major vascular risk factors. Am J Geriatr Psychiatry *2:*332, 1994.

Murphy DG, DeCarli C, Schapiro MB, Rapoport SI, Horwitz B: Age-related differences in volumes of subcortical nuclei, brain matter, and cerebropinal fluid in healthy men as measured with magnetic resonance imaging. Arch Neurol *49:*839, 1992.

*Namba H, Iyo M, Fukushi K, Shinotoh H, Nagatsuka S, Suhara T, Sudo Y, Suzuki K, Irie T: Human cerebral acetylcholinesterase activity measured with positron emission tomography: Procedure, normal values and effect of age. Eur J Nucl Med *26:*135, 1999.

Pearlson GD, Rabins PV, Kim WS, Speedie LJ, Moberg PJ, Burns A, Bascom MJ: Structural brain CT changes and cognitive deficits in elderly depressives with and without reversible dementia (‘pseudodementia'). Psychol Med *19:*573, 1989.

Rabins PV, Pearlson GD, Aylward E, Kumar AJ, Dowell K: Cortical magnetic resonance imaging changes in elderly patients with major depression. Am J Psychiatry *148:*617, 1991.

*Rose SE, de Zubicaray GI, Wang D, Galloway GJ, Chalk JB, Eagle SC, Semple J, Doddrell DM: A 1H MRS study of probable Alzheimer's disease and normal aging: Implications for longitudinal monitoring of dementia progression. Magn Reson Imaging *17:*291, 1999.

Sheline YI, Wang PW, Gado MH, Csernanasky JG, Vannier MW: Hippocampal atrophy in recurrent major depression. Proc Natl Acad Sci USA *93:*3908, 1996.

Sze G, DeArmand SJ, Brant-Zawadzki M: Foci of MRI signal anterior to the frontal horns: histologic correlation of a normal finding. AJR *147:*331, 1986.

Tupler LA, Coffey CE, Logue PE, Djang WT, Fagan SM: Neuropsychological importance of subcortical white matter hyperintensity. Arch Neurol *49:*1248, 1992.

Ylikoski A, Erkinjuntti T, Raininko R, Sarna S, Sulkava R, Tilvis R: White matter hyperintensities on MRI in the neurologically nondiseased elderly. Analysis of cohorts of consecutive subjects aged 55 to 85 years living at home. Stroke *26:*1171, 1995.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3A Psychiatric Problems in the Medically Ill

SOO BORSON, M.D. AND JURGEN UNÜTZER, M.D., M.P.H.

Increases in life expectancy during the twentieth century have produced an aged population of unprecedented size and longevity and altered the scope of clinical practice in all medical specialties. Demographic forecasters predict that by 2030 one person in five will be over age 65, and reductions in mortality among the very old (persons over age 80) already exceed those in other age groups. In the United States the average 65-year-old will remain independent for another 10 years and will live an additional 6.5 years in need of assistance; the average 85-year-old is likely to be independent until age 88 and live to age 89 with help. Persons over 65 account for more than one third of the country's total personal health expenditures and use three times the national average hospital bed days, twice the average number of prescription drugs, and one and a half times the number of physician office contacts. Chronic diseases and disabilities account for most of this increased health service utilization in later life, as most persons over age 65 suffer from one or more conditions that significantly impair health or function. Degenerative arthritis, primarily osteoarthritis, affects 50 percent, hypertension 40 percent, hearing loss 30 percent, urinary incontinence up to 30 percent, heart disease 28 percent, diabetes mellitus 15 percent, and significant impairment of vision 13 percent. Medical comorbidities aggregate with aging; 40 percent of men and over 50 percent of women who live beyond the age of 70 have two or more chronic diseases, and most of those over 80 have multiple health problems requiring professional care. These profound changes in the demographics of chronic disease have significant implications for the practice of psychiatry. For the psychiatrist, medical comorbidity is the single most important factor that distinguishes older and younger patients. The impact of comorbidity has increased demand for psychiatric care of patients with complex mental, emotional, and behavioral complications of medical illness. New research has begun to quantify these effects and focused attention on an emerging need for models of practice based on collaboration with primary and specialty medical providers and for interventions that demonstrably improve quality of life.

Estimates of the prevalence of psychiatric problems in the medically ill and disabled elderly population vary widely with clinical setting, but reliable data on comorbidities are still sparse. For example, 30 percent of older adults in primary care have at least one active psychiatric diagnosis; 10 to 15 percent have clinically significant depressive syndromes such as major depressive, dysthymic, or minor depressive disorders; 5 percent have dementia; and 2 to 5 percent have alcohol use or psychotic disorders. As a rule, the lowest rates of psychiatric symptoms and disorders are found among persons living independently in the community, and prevalence increases with medical acuity and severity of functional dependency (Fig. 51.3a-1). The relation between functional health and psychiatric comorbidity has been most clearly established for depression, and emerging data suggest similar trends for anxiety disorders, sleep disorders, and psychoses.

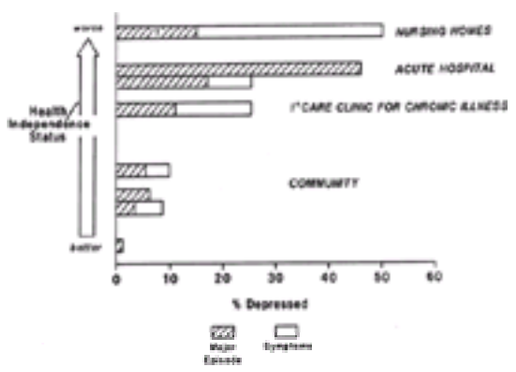

FIGURE 51.3a-1 Health, independence, and the prevalence of depression in late life.

Growing recognition of the association of chronic ill health and mental disorder among aged persons has spawned the development of specialized geropsychiatric services in many larger cities across the United States, but regular access to these services remains limited by financial, attitudinal, and organizational constraints. As a result, most treatment for comorbid psychiatric conditions in the elderly population in all medical settings is provided by primary care physicians and emphasizes the more acute, more severe, and more easily recognized disorders. Designing a full range of geropsychiatric health services, lowering barriers to their utilization, integrating primary care and specialty mental health care, and validating the role of the geropsychiatrist in optimal health care for the elderly are important priorities for the decades to come.

Clinical care of psychiatrically disturbed medical patients poses several special challenges for the psychiatrist. Diagnosis is frequently difficult, as clinical syndromes may vary in presentation from classical disorders with well-established treatment guidelines, and common disorders such as depression and anxiety may present in an atypical fashion. Key etiologic factors may become evident only as treatment unfolds, and physiological vulnerabilities arising from medical illness or its treatment may limit therapeutic choices or complicate therapy. Disparate explanatory models used by patients, their families, physicians, and other professionals must be reconciled before effective treatment can be implemented.

In most situations, psychiatrists initially become involved in the care of medically ill older patients as consultants to other physicians, but in a given episode of care, their role may be variously that of consultant, educator, advocate, mediator, counselor to caregivers, and traditional psychotherapist or pharmacotherapist. In some cases, psychiatrists find themselves becoming primary health care coordinators for complex elderly patients. These multiple tasks require a working familiarity with the common medical and functional problems of the elderly population as well as skill in the full range of psychotherapeutic and somatic treatment modalities unique to the profession of psychiatry. In spite of these inherent challenges, most psychiatrically disturbed older persons can be substantially helped when approached within a systematic framework.

APPROACH TO THE PATIENT

Mastery of a few central concepts helps to integrate the variety of clinical data critical for the successful evaluation and management of psychiatrically disturbed, medically ill geriatric patients (Table 51.3a-1). Aging is accompanied by a gradual diminution in adaptive reserve of all vital organs, including the brain, even in the absence of manifest disease. While the magnitude of these changes can be measured in aging populations, loss of physiological resilience in individual patients can generally be detected only after acute medical or psychosocial stress prompts a sudden, unexpected deterioration. Such age-related reductions in adaptive reserve underlie the commonly observed downward spiral and multiple organ failures of many sick elderly patients. They are responsible for the narrowed margin of safety observed for many drugs; they may also alter the presentation of disease and intensify the morbidity of medical and psychiatric disorders. For example, acute myocardial infarction in an octogenarian may present as painless delirium, while a mild chronic anxiety disorder may become fulminant panic in a patient with minor diverticular bleeding. Such complex patterns of causation are typical of psychiatric symptoms in very old persons and in those with multiple medical disorders. Sometimes causal factors emerge clearly only after successful intervention or not at all.

Diminished organ reserve alters response to illness, treatment, and social stressors
Comorbid medical illness may alter presentation and influence treatment choices
Nondrug treatments are preferable when effective
Combining treatment modalities optimizes gains
Comfort, function, and safety are the major goals of treatment
Coordinated care is better care

Table 51.3a-1 Key Concepts in Psychiatric Care of the Medically Ill Elderly

A 78-year-old man with bipolar I disorder since age 40 became manic while residing in a nursing home. His mental illness had been complicated by chronic dependence, and he was now mildly demented. He had no history of renal disease, and his serum creatinine was normal. After an initial period of typical manic grandiosity, when he was often seen near a local tavern, he soon deteriorated to crawling on his hands and knees, growling and barking like a dog, and striking out at all who approached. His lithium concentration was 1.2 mEq/L with a maintenance dosage of 750 mg a day. After reducing the lithium (Eskalith, Lithobid) dosage to achieve a concentration of 0.6 mEq/L, temporary addition of low-dose haloperidol (Haldol), and close attention to adequate fluid intake and preventing access to alcohol, both the manic episode, and the superimposed delirium cleared within 3 weeks without need for another psychiatric hospitalization.

Medical causes should be considered whenever an elderly patient develops a new psychiatric symptom or disorder. Whenever possible, major changes in management should be evaluated singly, especially when new drug treatments are added. When psychotropic medications are used, clinicians should remember to "start low and go slow," follow target symptoms, carefully monitor for the emergence of adverse effects, and allow sufficient time for medications to reach effective plasma and receptor site concentrations. Close collaboration with other clinicians and caregivers is essential at every stage of treatment to ensure its maximal safety and effectiveness. Combining psychotherapeutic, social, environmental, and somatic treatment modalities usually yields the greatest overall therapeutic gains at the smallest cost in treatment-related adverse events. Achieving therapeutic goals with elderly patients should be framed as an ongoing process that takes shape over time.

An 83-year-old woman with long-standing diabetes and ischemic and hypertensive cardiovascular disease had gradually become bedfast over several years in a nursing home. Two months before psychiatric consultation was requested, she developed a psychotic illness characterized by delusions of being beaten during the night, aggressive outbursts, and pervasive despondency, which led to frequent and disruptive nocturnal phone calls to her daughter and to the local fire station. Initial clinical assessment disclosed a moderately severe dementia in a personable but angry, frightened, and desperate woman eager for relief. Collateral history confirmed progressive cognitive deterioration over the course of her chronic medical illness, on a background of dysthymia of many years' duration. Trials of several antipsychotic and antidepressant medications had exacerbated both her symptoms and the problems of nursing staff in managing her care. A review of her medications showed that 4 of the 12 drugs she was taking on a regular basis were not essential for her medical management. A head computed tomographic (CT) scan showed mild cerebral cortical atrophy, moderate periventricular white matter hyperintensities, and numerous lacunar infarcts. A search for medical causes of her psychosis eventually associated its onset with the initiation of aggressive treatment of arterial leg ulcers, including round-the-clock debridement dressings, oral antibiotics, and narcotic analgesics. Control of her diabetes had deteriorated (as expected with this medical complication), and her insulin dose had been increased. Nursing staff members had provided excellent physical care, but neglected her emotional needs because they resented her ingratitude, racist attitudes, and physical aggressiveness, which exacerbated her loneliness and paranoia. Treatment of her psychiatric illness included the following interventions: discontinuation of nightime debridements (allowing elimination of narcotics for pain), simplification of her drug regimen, preventing nocturnal hypoglycemia (insulin reaction), installation of a nightlight in her room, problem-solving discussions with staff, and weekly psychotherapeutic conversations and wheelchair walks, which represented her first opportunity to go outdoors for several months. Her distressed phone calls ended within the first week. Her mood, sleep, and interactions with staff improved more gradually, and her delusions and racist remarks diminished and then disappeared altogether. Foci of psychotherapy were discovery of traumas she experienced in Cossack Russia (the origin of the delusional content), explanation of the delirium that led to her psychotic episode, and frank discussions of her desire to die rather than to continue to live in a physically deteriorated state. She developed a firm therapeutic alliance and regained the dignified, compassionate, and self-contained manner that had characterized her response to adversity before its erosion by her lengthy illness.

PSYCHIATRIC SYNDROMES IN MEDICALLY ILL PATIENTS

General Principles Almost any psychiatric syndrome may be mimicked by systemic medical illnesses with the capacity to affect central nervous system function. The fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) has simplified the classification of psychiatric disorders due to medical diseases (Table 51.3a-2). These syndromes are usually accompanied by some intellectual impairment, and a careful cognitive assessment provides valuable clues to their possible medical etiology. The cognitive component may be as subtle as confusional episodes in response to complex or changing environments or as obvious as syndromal delirium or dementia. Most serious medical disorders can cause psychiatric symptoms by more than one pathophysiological mechanism. Secondary psychiatric syndromes commonly lack some features of the typical primary disorders they resemble, particularly in patients with no history of psychiatric illness earlier in life. For example, in depressions due to central nervous system effects of medical disorders, guilty rumination is uncommon, and sleep disorder, when present, is often atypical. In patients with a history of psychiatric disorder, the onset or exacerbation of medical illness may precipitate relapse.

Cognitive disorders
Delirium
Dementia
Amnestic disorder
Psychotic disorders
With hallucinations, delusions, or mixed features
Mood disorders
With manic, major depressive, or mixed episode or features
Anxiety disorders
With generalized anxiety or panic attacks, or obsessive-compulsive, social phobic, or agoraphobic symptoms
Catatonic disorder
Personality change
Labile, disinhibited, aggressive, apathetic, paranoid, combined, other
Sexual dysfunction
Sleep disorder
Mental disorder not otherwise specified (e.g., dissociative symptoms due to complex partial seizures)

Table 51.3a-2 Psychiatric Disorders Due to a General Medical Condition

The interaction between psychiatric disorders and chronic medical conditions is complex. For example, a number of medical disorders cause significant depressive symptoms, but depression amplifies somatic symptoms of medical illnesses and can promote despair about the efficacy of treatment, leading to poor adherence and outcomes. On the other hand, comorbid medical illness partially explains why depression and other treatable psychiatric disorders are frequently unrecognized and untreated in primary care settings. Also, medical illness, polypharmacy, and changes in drug metabolism in older adults make adverse effects from psychotropic medications more likely.

A number of clues can alert the treating psychiatrist that an underlying physical disorder or a medication may be responsible for psychiatric symptoms: an abnormal level of alertness, new onset of psychiatric symptoms at a late age, lack of family history, psychiatric symptoms or personality changes that are more sudden or more severe than expected, coexisting physical illness, changes in medications, or poor response to usual psychiatric treatments.

Medications as a Cause of Psychiatric Symptoms Most drugs, in sufficient doses, can affect central nervous system function in a vulnerable host. Drugs affecting

the brain, either as their primary site of therapeutic action (e.g., all psychotropic drugs, opioids, antiparkinsonism agents, and many antihypertensive agents) or as a secondary site not essential to their therapeutic effect (e.g., digitalis [Lanoxin], cimetidine [Tagamet], glucocorticoids, and nonsteroidal anti-inflammatory drugs (NSAIDs) such as indomethacin [Indocin]), can cause psychiatric symptoms in medically ill elderly patients. When a pharmacological cause is suspected, a detailed search is indicated for associated systemic or neurological signs and symptoms, such as recent loss of appetite, reduced sense of well-being, blurred vision, slurred speech, dizziness, gait instability, or somnolence. A patient's complaint of being drugged, while often apparently delusional, should always prompt consideration of drug toxicity. A drug can be reasonably established as the primary cause in several ways. Some drugs produce classical symptoms in overdose, such as nausea and distorted color vision in digitalis overdose, or tinnitus and dizziness with aspirin. Certain clinical pictures with a common pathophysiology may be produced by drugs with different primary indications but similar pharmacological properties. An example is the central anticholinergic syndrome, characterized by tachycardia, pupillary dilation, visual hallucinations, and delirium in patients receiving tricyclic antidepressants, some antipsychotic agents, or antihistaminic drugs for sinusitis or gastric hypersecretion. For the relatively few drugs whose blood concentrations are readily determined and therapeutic ranges are well established, a markedly elevated concentration may be diagnostic. However, improvement when dosage of a suspect drug is decreased or discontinued is the most common and often most crucial test of its role in pathogenesis.

Drug and Disease Interactions Psychiatric symptoms arising from the interactive effects of medications and underlying disease may be severe enough to lead to psychiatric hospitalization. In a series of nearly 16,000 consecutive adult psychiatric admissions, about 1 percent were attributable to adverse drug reactions. Elderly patients were most often affected, particularly those with neurological disease or receiving multiple medications. Nearly 30 percent of elderly persons develop delirium after hospitalization for acute medical illness. Particularly at risk are those with cognitive impairment at the time of admission, the very old (over 80 years), patients with fractures or symptomatic infection, and those receiving antipsychotic drugs or opioids. Early detection of mental disorders arising from drug-disease interactions depends upon familiarity with the common medical illnesses of older people, the mechanisms by which they may affect brain function, and the pharmacological properties of drugs used in medical management and their special potential for interactive effects. Table 51.3a-3 gives examples of some drug-drug and drug-disease interactions that may be encountered in clinical practice.

Drug	Indication	Adverse Effect
Cimetidine	Gastroesophageal reflux	Delirium; elevated drug concentrations (e.g., tricyclic drugs, benzodiazepines, neuroleptic agents, propranolol [Inderal], theophylline [Theo-Dur]
Indomethacin (Indocin)	Cardioprotection	Parkinsonism, exaggerated reaction of levodopa
Tricyclic drugs	Depression, anxiety, insomnia, chronic pain	Confusion, constipation, tachycardia, cardiac conduction defect, hypotension, altered effect of antihypertensive agents (Lenox effect of digoxin, ↓ effect of α2-adrenergic receptor agonists, e.g., clonidine [Catapres])
SSRIs	Depression, anxiety, obsessive-compulsive symptoms	Agitation, insomnia, restlessness, agitation, ↑ acid hydrolysis, ↑ potential for diverse drug interactions, dysadaptation of workers (Lexapro), nausea, gastrointestinal distress, diarrhea
NSAIDs	Arthritis	Cognitive impairment, ↑ serum lithium concentration, ↑ blood pressure, ↓ hypertension control
Antipsychotic drugs	Dementia with psychosis	↑↑ cognitive impairment, extrapyramidal symptoms, falls, aspiration pneumonia
Alprazolam (high dose)	Insomnia	Delirium
Digoxin (Lanoxin)	Memory, arterial insufficiency	↑↑ falling time; restlessness (high dose)
Glucocorticoids	Inflammation, COPD	Depression, mania, psychosis (especially at high doses)

Table 51.3a-3 Examples of Drug-Drug Interactions in Elderly Persons

Differential Diagnosis Differential diagnosis begins with a psychiatric examination of the patient and interviews with competent informants, particularly about the timing of onset of the problem and its association with changes in medical or treatment status. This is followed by a thorough evaluation of the patient's current medical condition, emphasizing disease of major vital organs, painful conditions, and disorders that may leave few obvious clues (e.g., hip fracture or subdural hematoma in a patient who has fallen), a complete medication history; and a focused physical examination, including at a minimum a review of vital signs and a screening neurological examination to check for abnormal gait or movements, focal areas of weakness, or abnormal deep tendon reflexes. These data are used to derive a set of clinical probabilities and testable hypotheses. The judicious use of appropriate laboratory tests may clarify the possibilities in complex situations (Table 51.3a-4). For example, in a patient with blood pressure in the low normal range, pitting ankle edema, apathy, weakness, and depressed mood, who has been taking a high-potency diuretic for many years, an electrocardiogram (ECG) and echocardiogram may be needed to determine whether congestive heart failure or the diuretic is the most salient contributor to the depressive syndrome, whether it is safe to try discontinuing the diuretic, and what special precautions are required in prescribing an antidepressant. In doubtful cases, when the differential diagnosis includes a primary psychiatric disorder not caused by the physical effects of illness, an electroencephalogram (EEG) may be helpful. The EEG in patients with psychiatric syndromes due to metabolic, toxic, or circulatory causes typically shows diffuse slowing (although it cannot implicate specific diseases or drugs, and the interpretation may be confounded when an underlying dementing illness is present). An increase in dominant frequency accompanying recovery lends weight to the formulation.

Establish rapport quickly
Establish the reason for the consultation or visit
Learn and use a few rapid screening instruments
Cognition
Neuropsychiatric symptoms
Functional ability
Expect more than one psychiatric diagnosis (e.g., delirium and chronic depression)
Make a complete problem list
Medical
Psychiatric
Social
Functional
Identify all drugs in current use
Prescription drugs
Over-the-counter medications
Vitamins, minerals, botanicals, and "natural" remedies
Alcohol and tobacco
Include those used infrequently
Ask patients or caregivers to bring all medications from home in a paper bag for your inspection
Interview capable collateral sources
Family or other primary caregivers
Primary physician
Staff of residential facilities
For acute exacerbations of chronic problems, ask "why now?"
If problem unexplained, search for
Undiagnosed medical illness
Drug effect or interaction
Occult injury
Undetected social problem (abandonment, abuse)
Conduct a screening physical examination: vital signs, gait, neurological signs
Obtain appropriate corroborating laboratory studies; consider broad-spectrum tests to screen for subjacent disease (e.g., complete blood cell count [CBC], blood glucose, serum electrolytes, blood urea nitrogen and creatinine, liver enzymes, thyroid-stimulating hormone, syphilis serology, vitamin B_{12} and folate, urinalysis, electrocardiogram, chest X-ray, EEG for delirium, erythrocyte sedimentation rate [ESR] for occult systemic disorders, [sp] scans for a deficient patient who will not walk).

Table 51.3a-4 Practical Evaluation of Psychiatrically Disturbed Medically Ill Older Patients

When a drug is believed to cause or contribute to a psychiatric disorder, all drugs should be sorted into those clearly required for management of serious disease, and those that may be modified or stopped on a trial basis. Familiarity with drug adverse effects, interactions, and the pathways for their elimination may assist in the detection of otherwise occult medication toxicity syndromes. The authors estimate that assiduous adherence to the rule of maximal simplification of medication regimens will by itself lead to clinical improvement in perhaps half of all psychiatrically disturbed medically ill older persons.

A 68-year-old obese man with chronic obstructive lung disease and a history of alcohol dependence in remission came for treatment of an anergic depression with insomnia and daytime sleepiness following relinquishment of a long-standing extramarital affair. He responded initially to nortriptyline (Pamelor), but within 3 months began to deteriorate, with irritability, confusion, worsening depression, marked weight gain, edema, and eventual frank congestive heart failure. The plasma nortriptyline concentration was 75 ng/mL. Digitalis and diuretic treatment resulted in a 40-pound diuresis but no improvement in his psychiatric status. Measurement of (active) hydroxylated metabolites of nortriptyline showed concentrations above 300 ng/mL. His mental state cleared rapidly after discontinuation of the drug. Depression did not return, but he continued to require treatment for chronic congestive heart failure. The relative importance of psychological and medical factors in causing his initial depression was never clarified nor whether accumulation of nortriptyline metabolites (not measured in routine clinical assays) was the cause or consequence of his myocardial insufficiency.

Undiagnosed elderly patients who present with an undiagnosed dementia warrant aggressive assessment. In a review of 32 studies of causes of dementia, Alzheimer's disease accounted for 56 percent; vascular dementia, 13 percent; depression, 4.5 percent; alcoholic-induced persisting dementia, 4.2 percent; and substance-induced dementia, 1.6 percent. In 13.2 percent of cases, potentially reversible causes of cognitive impairment could be identified. In a subset of cases for which adequate follow-up information was provided, 11 percent of dementias resolved. Within this reversible group, drugs and depressive illnesses each accounted for approximately 25 percent of cases, and metabolic encephalopathies, 15 percent. These figures emphasize the value of psychiatric evaluation in detecting potentially reversible causes of cognitive impairment. In many cases, the initial evaluation may require more than one visit, particularly with frail elderly patients who

have difficulty providing relevant historical details.

PSYCHIATRIC SYMPTOMS ASSOCIATED WITH SPECIFIC MEDICAL DISORDERS

Cerebrovascular Disease Ischemic and hemorrhagic strokes cause a variety of disturbances of behavior and emotionality related both to effects of the lesion on brain function and to psychological problems in adaptation to disability. Significant depressions are common, occurring in up to 50 percent of all patients following an acute cerebrovascular accident. Depressive disorders after neocortical strokes resemble primary depressions in both clinical presentation and response to antidepressant pharmacotherapy, although the incidence of drug intolerance may be somewhat increased. Although depressions can occur after stroke anywhere in the brain, larger lesions in the anterior left hemisphere are associated with a high likelihood of depression in the early recovery period. Persistence of milder depressive syndromes after the acute phase has been associated with factors related to functional impairment, including impaired cognition, weak social supports, and inadequate living environment. Manic episodes (which may respond to lithium or antimanic anticonvulsants) and apathetic behavioral states without depressive affect or mood appear especially likely after right hemisphere strokes. *Subcortical dementia* is a clinically useful concept designating a syndrome of psychomotor slowing, depressed mood, inattentiveness, forgetfulness, motor impairment, and seemingly disproportionate functional dependency occurring in patients with diseases of the thalamus, basal ganglia, and upper brainstem. This syndrome commonly results from subcortical ischemic disease (vascular depression) as well as from classic neurological diseases such as parkinsonism. Its chief importance for psychiatrists is its frequent confusion with primary depressions in elderly persons, which leads to neglect of the associated intellectual deficits in formulating an overall management plan, and its contribution to depressive chronicity and poorer antidepressant treatment outcome. However, when depressive elements are prominent, antidepressant treatment may provide decisive benefits, and the apathetic state that frequently accompanies subcortical vascular brain disease can respond to pharmacotherapy with psychostimulants.

Acute confusional states frequently occur after stroke and may be the only presenting sign in a small minority of patients, particularly those with basal ganglia and other subcortical infarcts. Permanent cognitive impairment can result from strategically placed single strokes, and progressive dementia is associated with repeated events. Patients recovering from cerebrovascular accidents and those with vascular dementia may be extremely sensitive to substance induced delirium and to deterioration as a result of either uncontrolled hypertension or hypotensive episodes due to antihypertensive medications and drug interactions.

Psychotic symptoms with delusions, hallucinations, or a full schizophrenia-like syndrome may result from focal brain injury, particularly after lesions of limbic cortical and subcortical structures. Paranoid and persecutory delusions and ideas of reference are common and may be circumscribed or highly elaborate and pervasive. Personality changes, particularly those producing inappropriate aggressive or sexual behaviors, are encountered with regularity in patients with strokes, especially men with frank dementia, and frequently lead to psychiatric consultation and psychotropic drug treatment and sometimes to permanent psychiatric hospitalization. The relationship of these disorders to premorbid character pathology is uncertain, but families often report long-standing socially dysfunctional behavior in these patients before the onset of cerebrovascular disease further weakened the capacity to implement behavioral controls.

Parkinson's, Alzheimer's, and Other Neurological Diseases Depression occurs with disproportionate frequency in Parkinson's disease, affecting up to half of all patients. This depression is believed to be caused, in large part, by the neurodegenerative process itself; its presence is associated with evidence of damage to noradrenergic and serotonergic systems and is thought to predict more rapid overall neurological deterioration. Nevertheless, it typically responds to antidepressant drugs or electroconvulsive therapy. Selegiline (Eldepryl), an antiparkinson agent that selectively inhibits monamine oxidase type B, may be a useful addition to treatment in depressed Parkinson's patients, but its role in psychiatric management is not yet well characterized. Most depressed Parkinson's disease patients have cognitive deficits that also occur without mood disorder and influence learning, initiative, planning, and capacity for independent functioning, beyond the constraints imposed by the motor disability alone. These cognitive deficits tend to persist after successful treatment of mood disorder and are not substantially improved by treatment with levodopa (Larodopa) or other antiparkinsonian agents. Cognitively impaired Parkinson's disease patients are particularly vulnerable to delirium and psychosis induced by antiparkinsonian medications, including levodopa, other dopaminergic agonists, and anticholinergic agents. These syndromes may respond to lowered drug dosages. If not or if motor disability worsens to an unacceptable degree, low-potency dopamine receptor antagonists such as thioridazine (Mellaril) or serotonin-dopamine antagonists may be required. A large body of literature supports the use of clozapine (Clozaril) in relatively low dosages (up to 50 mg a day) for treatment of dopaminergic psychosis in Parkinson's disease, and may improve the motor disturbances as well. Olanzapine (Zyprexa), in dosages of 2.5 to 15 mg a day, is a useful alternative, but its onset of antipsychotic action is typically delayed by 2 to 3 weeks.

In Alzheimer's disease, depression affects up to one third of patients early in the course and may be the major presenting problem. Clear-cut major depressive episodes generally respond well to antidepressant pharmacotherapy with selective serotonin reuptake inhibitors (SSRIs) or secondary amino tricyclic antidepressants, but for apathetic states without prominent mood or sleep disorder, donepezil (Aricept), another cholinergic agonist, or a stimulant such as methylphenidate (Ritalin) may be more effective. As dementia worsens, classic depressions become less frequent, while agitation, psychosis, and aggressive behavior with or without dysphoria become more common. Assaultive or threatening behavior in Alzheimer's disease patients demands thorough medical and socioenvironmental assessment to identify factors that provoke or sustain these problems, but pharmacological interventions are usually required. Delirium induced by drugs or acute medical illness is a risk at all stages of the disease, as it is in other neurodegenerative disorders.

> A 70-year-old woman with moderately advanced Alzheimer's disease was brought for psychiatric treatment because of intolerable agitation and aggressive outbursts. Haloperidol, prescribed by the primary physician, seemed to worsen her behavior. During the interview, markedly impeded by the patient's severe aphasia, she acknowledged depression and tearfully expressed worry about her husband's drinking. After several visits and insistence on interviewing a son, it became obvious that her husband was severely alcoholic and was abusing benzodiazepines, that he had been emotionally and physically abusive throughout the marriage, and that the patient's agitation and aggression were defensive responses to his threats and assaults. Emergency nursing home placement solved her behavior problems, including her depression.

Posttraumatic and postanoxic encephalopathies, hydrocephalus, central nervous system (CNS) tumors and infections, multisystem atrophy (Shy-Drager syndrome), Huntington's disease, demyelinating diseases such as multiple sclerosis, idiopathic inflammatory diseases such as systemic lupus erythematosus, and temporal arteritis and occupational neurodegenerations such as chronic organic solvent encephalopathies may all be associated with depressive, manic, delirious, and psychotic syndromes. For depressions in patients with these brain diseases, antidepressant trials are generally warranted, though drug intolerance may be problematic and require the use of innovative strategies. The management of other psychiatric complications is less well established.

Cardiovascular Disease Acute myocardial infarction and chronic congestive heart failure may both be accompanied by psychiatric disturbances. Confusional states due to cerebral hypoperfusion or complications of treatment (e.g., cardiovascular drugs, electrolyte imbalances due to diuretics, or repeated resuscitations for ventricular arrhythmias) are best managed by primary efforts to improve compromised cardiac function and metabolic status. Brief use of benzodiazepines or low-dose neuroleptic agents is indicated for very fearful patients and to prevent injury to patients or professional staff when agitation poses a safety risk. About 10 percent of patients who undergo open heart surgery suffer lasting neuropsychological impairment as a result. Barring major intraoperative complications, most of these impairments are mild; most affected patients are elderly.

Depression and anxiety disorders following myocardial infarction or heart failure require attention to both possible medical causes and psychological factors, especially patients' concerns about risk of death and disability and the adequacy of supportive care. Untreated or inadequately treated major depression worsens prognosis in heart disease, and effective psychiatric treatment often requires both psychotherapeutic and pharmacological management. The safety and efficacy of antidepressant drugs in patients with chronic heart disease have been evaluated in several studies. SSRIs are usually the drugs of choice. These drugs occasionally cause bradycardia, by a central medullary mechanism, but lack primary cardiotoxicity. Patients who fail to respond to these agents and do not have serious cardiac conduction defects or severe chronic congestive heart failure can generally be safely treated with secondary amine tricyclic agents (nortriptyline or desipramine [Norpramine]) provided the potential for interaction with cardiovascular drugs is kept in mind. Conduction defects and chronic low-output states with hypotension increase the risk of tricyclic treatment; in these situations, bupropion (Wellbutrin) or psychostimulants are preferable. While electroconvulsive therapy is an undisputed myocardial stressor, treatment can usually be given without major complications if careful attention is given to anesthetic management and to individualizing the treatment schedule.

Chronic Diseases of the Lung, Kidneys, and Liver Obstructive lung disease, caused in large part by cigarette smoking, is the most prevalent of the chronic respiratory diseases of the second half of life. Psychiatric complications include acute and chronic encephalopathies related to acute infection or respiratory failure, to sustained hypoxemia and hypercapnia, or to the CNS effects of drugs used in management (oral or inhaled bronchodilators, high-dose prednisone, cough suppressants, or benzodiazepines). Depressions are common, as in other chronic medical illnesses, but tend to be characterized by prominent anxiety and paniclike episodes, accounting for frequent prescription of anxiolytics. Psychotic states may occur and are usually related to acute impairment of brain function. Psychological complications are frequent and include avoidance of therapeutic physical activity because of panic in response to dyspnea, fear of sudden death by asphyxiation, excessive self-blame in patients whose illness is related to smoking, and avoidance of appropriate social activities because of negative body image. Untreated

depression in chronic obstructive pulmonary disease increases functional disability. For milder dysphoric states not meeting criteria for a depressive diagnosis, a multidisciplinary rehabilitation and counseling approach improves mood in many patients. For patients with syndromal depressions who do not have severe left ventricular disease or hypercapnia, antidepressant pharmacotherapy with nortriptyline is effective and markedly improves anxiety and severity of somatic symptoms. Important nonpharmacological interventions for anxious depressed patients with chronic obstructive pulmonary disease provide individualized education about the disease and its management, demonstrate to patients how they can modify their own symptoms with appropriate techniques such as breathing retraining, and improve overall physical condition through supervised physical training. Patients with pulmonary disability and those who require home oxygen therapy may benefit from self-help groups available nationwide. Patients with lung cancer may develop psychiatric complications similar to those occurring in chronic obstructive pulmonary disease. Paraneoplastic psychiatric syndromes have also been reported.

Chronic renal and hepatic disease produce encephalopathic states through their effects on nutrition, metabolism, excretory clearance, and drug disposition. The prevalence of other psychiatric disorders in these conditions approximates those for chronic disease in general; treatment considerations include evaluating the effects of disease on effective plasma drug levels and half life. Chronic hemodialysis has been associated with dementia due to brain accumulation of aluminum present in dialysis solutions, a discovery that has led to its exclusion from the formula.

Arthritis The two major forms of arthritis in the elderly, osteoarthritis and rheumatoid arthritis, are important to the psychiatrist chiefly because together they constitute the leading cause of chronic disability in old age, affecting one of every two persons over 65. Mood disorders, the most prevalent psychiatric complications of degenerative arthritis, occur in up to 25 percent of patients and are primary determinants of the severity of functional disability. Because of the widespread comorbidity of arthritis and depression, recognition and treatment of mood disorder can contribute significantly to improving the functional health of the elderly population. Occasionally, psychiatric symptoms arise as a complication of treatment with NSAIDs, or prednisone.

Thyroid Disease, Malnutrition, and Anemia Thyroid disease, malnutrition, and anemia are relatively common, particularly in older adults with significant comorbid medical illness. They may be responsible for such psychiatric symptoms as lethargy, weakness, confusion, and behavioral changes and should be actively sought and treated.

SPECIAL ISSUES IN THE CARE OF ELDERLY PATIENTS

Impaired Sight and Hearing The importance of sensory impairments in patients who develop psychiatric disorders in later life justifies special attention to these issues during evaluation and treatment. Degenerative diseases of the eye and ear impair responsiveness to the interpersonal and social environment, increase feelings of vulnerability among older persons, and cause hallucinations in the affected sensory modality in some individuals. In addition, deafness originating in childhood disease but exacerbated by presbycusis appears to be a risk factor for late-onset paranoid illness, and hearing and visual loss both contribute to the onset and persistence of depressions. Treatment of eye disease, improvement in corrective lenses, appropriate prescription of hearing aids, and attention to environmental lighting and sound may enhance the efficacy of specific interventions for psychiatric disorder in sensory-impaired patients.

Falling Annually, falling and fall-related injuries occur in 30 percent of older adults living in the community and in more than 50 percent of those living in nursing homes. Complications of falling include hip fracture (25 per 1000 women over age 85), other fractures, lingering painful soft-tissue injury, and head injuries with behavioral complications. Such injuries are widely feared by the old as harbingers of death. Risk factors for falling include conditions that cause postural instability, slow response to unexpected circumstances, general muscular deconditioning, and environmental hazards. Visual, auditory, vestibular, proprioceptive, and central neurological impairments; musculoskeletal dysfunction arising from disease; a sedentary lifestyle; illness-related bed rest; painful or deforming foot conditions; systemic diseases; and medications all increase the risk of falls. Poor lighting, obstacles, slippery surfaces, too much or too little friction of shoe with ground, and inadequate ambulatory supports contribute. The use of sedative medications is the single most powerful predictor of falling in elderly persons, all other things being equal. These include drugs used to manage insomnia, anxiety, depression, and agitation (both prescribed and over-the-counter); alcohol; and drugs used in the treatment of many medical diseases that sedate by themselves or enhance the sedative effects of other drugs. The presence of cognitive disorders (dementia and delirium) and all causes of ambulatory disability substantially increase fall risk, and the presence of multiple risk factors geometrically increases the hazard. Psychiatric complications of falls include acute cognitive disturbance due to head injury, and a common but rarely recognized phobic anxiety state (fear of further falls, which leads to reduction in activity, curtailment of opportunities for social exchange, and sometimes permanent retirement to a wheelchair). Persistent fall phobias may become less common with aggressive rehabilitation after hip fracture and dissemination of exercise programs for frail elderly persons.

Urinary Incontinence Urinary incontinence affects up to one third of older adults living in the community and can cause significant suffering and functional impairment. Patients may not mention this problem to their physician unless specifically asked and may suffer silently, limiting their trips away from home and their opportunities for social activities. There are a number of causes for urinary incontinence, including urinary tract infections, altered mental status, restricted mobility, medications such as diuretics, or more chronic anatomical changes. Patients should be referred to their primary care physician or to an appropriate specialist for a thorough workup.

Over-Reliance on Drugs for Managing Mental Disorders Psychiatrists and other physicians tend to underuse psychotherapeutic modalities in the care of sick or frail older patients. While properly targeted biological interventions have an undisputed role, a psychotherapeutic orientation to care is an indispensable part of optimal diagnosis and clinical decision-making and enhances the efficacy of other treatment modalities regardless of clinical setting or the patient's diagnosis. Formal office-based individual psychotherapy is now, and will likely remain, a relative rarity for the medically ill and frail elderly population. Most psychiatric care for this population will continue to occur in the context of consultation to primary care physicians, psychiatric inpatient care, and liaison work with long-term care facilities. However, empirical studies have demonstrated the value of psychologically oriented, disease-focused, group treatment approaches to chronic heart, lung, and malignant diseases and of problem-solving group therapy for depression in elderly patients in the primary care setting. For all patients, the psychiatrist's capacity to develop a rapid and sustained emotional connection; the flexible use of techniques borrowed from psychodynamic, interpersonal, cognitive, and behavioral treatment paradigms; and empathic integration of the needs of patients, their families, and professional caregivers into a comprehensive and workable treatment plan are fundamental skills that make substantial contributions to both medical management and patients' comfort and capacity to persevere in the face of serious and ongoing illness.

HEALTH-SYSTEMS APPROACH TO THE CARE OF MEDICALLY ILL GERIATRIC PATIENTS

Most older adults receive care for mental and emotional problems exclusively in the general medical sector, and many who could benefit from treatment of psychiatric disorders do not get it. There is ample evidence, for example, that many if not most older adults with significant depression are not recognized and diagnosed in primary care. Even if they are recognized, only a minority of patients receive guideline-level treatment for their depression, despite good evidence for the efficacy of treatments such as antidepressant medications and psychotherapy in old age. Older adults are less likely than younger patients to seek and receive specialty mental health services on their own and, when referred by their primary care providers, are less likely than younger patients to follow through. Obstacles to effective treatment include the fact that older adults and their primary care physicians may misattribute treatable psychiatric symptoms to normal aging, stressful life events, grief, social isolation, caretaking responsibilities, dementia, or medical illness. These misattributions can entrench hopelessness and therapeutic nihilism in patients and providers. Older adults may also fear stigmatization associated with mental disorders or psychiatric treatments. Other barriers to cost-effective care include the fact that older adults may have difficulty accessing specialty mental health care, especially when provided only in off-site specialty clinics. Many mental health insurance plans frequently fragment care by restricting psychiatric services to providers who work in settings where patients do not receive their primary care. Further obstacles include higher copayments for outpatient mental health services under Medicare and limitations in medication coverage and restricted access to new and more expensive psychotropic medications that are often better tolerated by older adults.

In recent years a number of systematic approaches have been developed to improve the treatment of patients with mental disorders in primary care. Such nterventions have included systematic case finding using simple screening instruments such as the Primary Care Evaluation of Mental Disorders (Prime-MD), patient and family education and activation, clinician education and support, systematic assessment and treatment matching, proactive outcomes tracking, and most importantly the integration of primary care and specialty mental health services. These models may use a geriatric nurse specialist with special expertise in recognizing and treating common mental disorders or a mental health specialist such as a psychologist or psychiatrist working in a primary care clinic to provide consultation and treatment on site, thus reducing barriers to access and communication between primary and specialty providers. This kind of integration may be especially important for older adults who have multiple medical problems and who will not seek care in geographically separate psychiatric clinics. In a similar fashion, geriatric psychiatrists have also developed successful models for integrated care provided on site in nursing homes. Psychiatrists and other mental health specialists working in these general medical settings can also help refer patients to further specialty mental health care as needed or link them to helpful community services. Examples of such integrated systems of care have been shown to be cost-effective for mixed-aged populations and should now be formally evaluated for older adults.

SUGGESTED CROSS-REFERENCES

Alzheimer's disease and other dementing diseases are discussed in Section 51.3e. Psychiatry and medicine are discussed in Section 28.1. Cognitive disorders are discussed in Chapter 10.

SECTION REFERENCES

*Alessi CA, Cassel CK: Medical evaluation and common medical problems. In *Comprehensive Review of Geriatric Psychiatry–II*, ed 2. J Sadavoy, Lazarus L, Jarvik LF, Grossberg GT, editors. American Psychiatric Press, Washington, DC, 1996.

*Alexopoulos GS, Meyers BS, Young RC, Campbell S, Silversweig D, Charlson M: "Vascular depression" hypothesis. Arch Gen Psychiatry *54:*915, 1997.

Applegate WB, Blass JP, Williams TF: Instruments for the functional assessment of older patients. N Engl J Med *322:*1207, 1990.

Arean PA, Perri MG, Nezu AM, Schein RL, Christopher F, Joseph TX: Comparative effectiveness of social problem-solving therapy in older adults. J Consult Clin Psychol *61:*1003, 1993.

Arranz FJ, Ros S: Effects of comorbidity and polypharmacy on the clinical usefulness of sertraline in elderly depressed patients: An open multicentre study. J Affect Disord *46:*285, 1997.

Auzou P, Ozsancak C, Hannequin D, Moore N, Sugustin P: Clozapine for the treatment of psychosis in Parkinson's disease: A review. Acta Neurol Scand *94:*329, 1996.

*Bennett DA, Shannon KM, Beckett LA, Wilson RS: Dimensionality of parkinsonian signs in aging and Alzheimer's disease. J Gerontol A Biol Sci Med Sci *54:*M191, 1999.

*Borson S, editor: Behavioral syndromes in neurodegenerative disease. Semin Clin Neuropsychiatry *1:*1, 1996.

Borson S, McDonald GJ, Gayle T, Deffebach M, Lakshminarayan S, Van Tuinen C: Improvement in mood, physical symptoms, and function with nortriptyline for depression in patients with chronic obstructive pulmonary disease. Psychosomatics *33:*190, 1992.

Clarfield AM: The reversible dementias: Do they reverse? Ann Intern Med *109:*476, 1988.

*Coulehan JL, Schulberg HC, Block MR, Madonia MJ, Rodriguez E: Treating depressed primary care patients improves their physical, mental, and social functioning. Arch Intern Med *157:*1113, 1997.

Franco-Bronson K: The management of treatment-resistant depression in the medically ill. Psychiatr Clin North Am *19:*329, 1996.

Hocking LB, Koenig HG: Anxiety in medically ill older patients: A review and update. Int J Psychiatry Med *25:*221, 1995.

Katon WJ, Von Korff M, Lin E, Unützer J, Simon G, Walker E, Ludman E, Bush T: Population-based care of depression: Effective disease management strategies to decrease prevalence. Gen Hosp Psychiatry *19:*169, 1997.

*Katz IR: On the inseparability of mental and physical health in aged persons. Lessons from depression and medical comorbidity. Am J Geriatr Psychiatry *4:*1, 1996.

Katz IR, Sands LP, Bilker W, DiFilippo S, Boyce A, D'Angelo K: Identification of medications that cause cognitive impairment in older people: the case of oxybutynin chloride. J Am Geriatr Soc *46:*8, 1998.

Kunik ME, Benton CL, Snow-Turek AL, Molinari V, Orengo CA, Workman R: The contribution of cognitive impairment, medical burden, and psychopathology to the functional status of geriatric psychiatric inpatients. Gen Hosp Psychiatry *20:*183, 1998.

Lebowitz BD, Pearson JL, Cohen GD: *Clinical Geriatric Psychopharmacology*. Williams & Wilkins, Baltimore, 1998.

Lebowitz BD, Pearson JL, Schneider LS, Reynolds CF III, Alexopoulos GS, Bruce ML, Conwell Y, Katz IT, Meyers BS, Morrison MF, Mossey J, Niederehe G, Parmelee P: Diagnosis and treatment of depression in late life: Consensus statement update. JAMA *278:*1186, 1997.

Lindley CM, Tully FP, Paramsothy V, Tallis RC: Inappropriate medication is a major cause of adverse drug reactions in elderly patients. Age Ageing *21:*294, 1992.

Molaparast Vali F, Walkup J: Combined medical and psychological symptoms: Impact on disability and health care utilization in patients with arthritis. Med Care *36:*1073, 1998.

Nebes RD, Pollock BG, Mulsant BH, Kirshner MA, Halligan E, Zmuda M, Reynolds CF III: Low-level serum anticholinergicity as a source of baseline cognitive heterogeneity in geriatric depressed patients. Psychopharmacol Bull *33:*715, 1997.

Nemeroff CB, DeVane CL, Pollock BG: Newer antidepressants and the cytochrome P450 system. Am J Psychiatry *153:*311, 1996.

Ormel J, Kempen GIJM, Penninx RWJH, Brilman El, Beekman ATF, Van Sonderen E: Chronic medical conditions and mental health in older people: Disability and psychosocial resources mediate specific mental health effects. Psychol Med *27:*1065, 1997.

*Pollock BG, Mulsant BH: Antipsychotics in older patients. A safety perspective. Drugs Aging *6:*312, 1995.

Preskom SH: Recent pharmacologic advances in antidepressant therapy for the elderly. Am J Med *94*(5A):2S, 1993.

Roose SP, Dalack GW, Woodring S: Death, depression, and heart disease. J Clin Psychiatry *52*(Suppl):34, 1991.

Schor JD, Levkoff SE, Lipsitz LA, Reilly CH, Cleary PD, Rowe JW, Evans DA: Risk factors for delirium in hospitalized elderly. JAMA *267:*827, 1992.

Speechley M, Tinetti M: Falls and injuries in frail and vigorous community elderly persons. J Am Geriatr Soc *39:*46, 1991.

Spigset O, Hedenmalm K: Hyponatremia in relation to treatment with antidepressants: A survey of reports in the World Health Organization database for spontaneous reporting of adverse drug reactions. Pharmacotherapy *17:*348, 1997.

Tandberg E, Larsen JP, Aarsland D, Laake K, Cummings JL: Risk factors for depression in Parkinson disease. Arch Neurol *54:*625, 1997.

Townes BD, Bashein G, Hornbein TF, Coppel DB, Goldstein DE, Davis KB, Nessly ML, Bledsoe SW, Veith RC, Ivey TD: Neurobehavioral outcomes in cardiac operations. A prospective controlled study. J Thorac Cardiovasc Surg *98:*774, 1989.

Unützer J, Katon WJ, Sullivan M, Miranda J: Treating depressed older adults in primary care: Narrowing the gap between efficacy and effectiveness studies. Milbank Q *77:*225, 1999.

Unützer J, Patrick DL, Simon G, Grembowski D, Walker E, Rutter C, Katon W: Depressive symptoms and the cost of health services in HMO patients aged 65 years and older. A 4-year prospective study. JAMA *277:*1618, 1997.

Unützer J, Small G: Geriatric consultation-liaison psychiatry. In *Comprehensive Review of Geriatric Psychiatry–II*, ed 2, J Sadavoy, L Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

*Walsh JD, Blanchard EB, Kremer JM, Blanchard CG: The psychosocial effects of rheumatoid arthritis on the patient and the well partner. Behav Res Ther *37:*259, 1999.

Wise MG, Griffies WS: A combined treatment approach to anxiety in the medically ill. J Clin Psychiatry *56*(Suppl 2):14, 1995.

Wolf B, Grohmann R, Schmidt LG, Ruether E: Psychiatric admissions due to adverse drug reactions. Compr Psychiatry *30:*534, 1989.

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3B SLEEP DISORDERS

PATRICIA N. PRINZ, PH.D., MICHAEL V. VITIELLO, PH.D. AND SOO BORSON, M.D.

Sleep and wake disturbances can result from physiological changes that are apparently part of normal aging or from poor sleep hygiene or one or more specific sleep disorders. Such rhythm disturbances are common in elderly adults and often lead to complaints of disturbed sleep that may baffle physicians. Specialized diagnostic aids are frequently helpful. A history of sleep problems, the patient's current sleep hygiene, and a comprehensive drug-use history should be sought. Spousal reports, sleep logs, and (where necessary) all-night polysomnograms from an accredited sleep disorder center can add further clues to help unravel the complex array of potential factors. Age-related physiological sleep changes, poor sleep hygiene, specific sleep disorders, and a variety of special medical, pharmacological, situational, and environmental factors can all affect sleep quality. A thorough diagnostic evaluation cannot be replaced by prescribing a hypnotic drug, and long-term use of a hypnotic agent is almost always contraindicated. Effective treatment, usually possible after thoughtful appraisal of the patient's needs, can range from patient education to surgical intervention.

SLEEP CHANGES IN NORMAL AGING

Age-related changes in the amount and pattern of the various stages of sleep and wakefulness are well described. Elderly people spend more time in bed and less time asleep and are more easily aroused from sleep than are young people. The most striking changes include a reduction in slow-wave sleep (particularly stage 4 sleep), increased nighttime wakefulness, and increased fragmentation of sleep by periods of wakefulness. Less striking age reductions in rapid eye movement (REM) sleep and total nighttime sleep also occur. The age-related impairments in sleep depth and maintenance seem to be accompanied by an age-related increase in sensitivity to environmental stimuli that disturb sleep; for example, elderly people are more easily aroused from nighttime sleep by auditory stimuli than are young people.

Many age-related changes in sleep patterns suggest that aging may disrupt the circadian sleep-wake rhythm. Increased nighttime wakefulness in elderly persons is mirrored by more daytime fatigue, more daytime napping, and a greater likelihood of falling asleep during the day. Advancing age has also been associated with a tendency to fall asleep and awaken earlier than in earlier years, and older people are less tolerant of phase shifts of the sleep-wake schedule, such as those caused by shift work and transmeridian flight (jet lag). Because these changes are also observed in healthy seniors, they are attributed to normal, age-related neuronal alterations in brain areas controlling sleep physiology rather than pathological processes.

SLEEP HYGIENE

Sleep disturbances may also derive from such treatable factors as poor sleep hygiene and specific sleep disorders. The clinician must fully evaluate an aged patient's sleep complaints, not simply attribute them to getting older. Poor sleep hygiene can reflect inappropriate expectations about sleep or sleep schedules (e.g., excessive expectations of sleep need that result in daytime napping and excessive time in bed or irregular bed and rise times), the bedroom environment (including acoustics, lighting, bed partner, and such inappropriate activities as worrying), dietary habits, lack of regular exercise, and inappropriate use of caffeine, alcohol, or medications in amounts or at times that affect sleep adversely. These factors should be evaluated and addressed when dealing with a sleep complaint, even when a specific disorder is suspected. Table 51.3b-1 contains a checklist of major issues to be explored when assessing sleep hygiene.

Table 51.3b-1 Sleep Hygiene Checklist

SLEEP DISORDERS

Sleep disturbances in the elderly can result not only from nonpathological age-related changes and poor sleep hygiene but also from specific sleep disorders.

Breathing-Related Sleep Disorder (Sleep Apnea) Sleep apnea is characterized by repeated episodes of breathing cessation and resultant hypoxemia, multiple brief awakenings from sleep, complaints of excessive daytime sleepiness, and impaired daytime functioning. Sleep apnea usually presents as a complaint of excessive sleepiness that interferes with normal daytime functioning. Apneic events occur more frequently in men as they age, particularly if they snore. The apneic episodes can be exacerbated by the use of depressant drugs and alcohol. A number of epidemiological studies have linked snoring with sleep apnea, hypertension, and cardiovascular disease. Untreated severe sleep apnea syndrome can compromise cardiac function and increase mortality.

The treatment of obstructive sleep apnea includes (from least to greatest risk of adverse effects) minimizing sleeping on the back, using a dental appliance, weight loss, avoiding respiratory depressant drugs (hypnotics and alcohol), using respiratory stimulants such as acetazolamide (Diamox), nasal continuous positive airway pressure, and upper airway surgery. Continuous positive airway pressure is initially well tolerated by most adult patients, and it almost immediately consolidates nighttime sleep, attenuates daytime sleepiness and fatigue, and improves cognitive function. Specific surgical procedures developed to modify the upper airway (e.g., uvulopalatopharyngeoplasty) have shown limited efficacy in follow-up studies. Tracheostomy has long been a viable treatment option for severe apnea. Both types of surgical procedures have obvious limitations including risk of morbidity and mortality.

The decision to treat and the choice of treatment depend on the relative morbidity of the sleep apnea and the treatment. Sleep apnea morbidity assessment includes the frequency and severity of the sleep disturbances, the degree of daytime sleepiness and cognitive impairment, and medical sequelae (cardiorespiratory abnormality). Little evidence supports the treatment of mild obstructive sleep apnea in the elderly in the absence of excessive daytime sleepiness, cognitive impairment, or associated cardiorespiratory abnormality.

Restless Leg Syndrome and Periodic Leg Movements Elderly persons can develop two sleep-related neuromuscular dysfunctions: *restless leg syndrome* (a very strong presleep urge to move one's legs repeatedly, resulting in sleep-onset insomnia) and *periodic leg movements* during sleep, commonly known as *nocturnal*

myoclonus (stereotypical and periodic rapid flexion of the leg and foot, associated with repeated awakenings throughout the night). The prevalence of these disorders appears to increase with age and may be related to metabolic, vascular, and neurological causes. Leg movements are often present in conjunction with sleep apnea.

Current treatments for these disorders are not ideal. Commonly used benzodiazepines (clonazepam [Klonopin] and temazepam [Restoril]) ameliorate repeated arousals from sleep but may have minimal effects on the leg movements and may cause daytime sedation. Opioids, levodopa (Larodopa), clonazepam, carbamazepine (Tegretol), and dopamine receptor antagonists have also been successful in some clinical trials.

Persistent Psychophysiological Insomnia Psychophysiological insomnia results from arousal or anxiety states that interfere with sleep. Also called learned or behavioral insomnia, this disorder is diagnosed when the insomnia is maintained by maladaptive learning or sleep-incompatible behaviors in the bedroom, such as excessive problem solving, worrying, and negative expectations about poor sleep and tiredness the next day. Among frail elderly patients living alone, fears and feelings of vulnerability may be exacerbated at night and lead to difficulty sleeping. Sometimes a move to some form of shared housing is the most effective intervention. Effective treatment includes behavioral therapies to curtail and regularize in-bed hours and to modify sleep-incompatible behaviors, biofeedback, and progressive muscle relaxation, always in conjunction with education about good sleep hygiene (Table 51.3b-1).

Circadian Rhythm Sleep Disorder Rest-activity rhythms and other circadian rhythms may diminish in amplitude with advancing age. Considerable circumstantial evidence indicates that shallow circadian rhythms are associated with impaired sleep. Increased nocturnal temperatures are common in those having disturbed sleep. Attenuation of the rest-activity cycle also occurs and adversely affects sleep in bedridden or sedentary adults, suggesting that the declining physical activity of old age may impair sleep. Possibly related to this attenuation is the tendency to both fall asleep and awaken earlier (i.e., a tendency for elderly people to be “larks” rather than “owls”). Numerous studies have suggested that sleep quality can be improved by interventions that foster a more robust rest activity rhythm by synchronizing or enhancing endogenous circadian rhythms. Interventions such as optimally timed social or physical activity, regularized sleep schedules, and optimally timed bright light, melatonin, or passive heating treatments have successfully improved sleep quality. Use of such interventions to treat age-related circadian sleep-wake disturbances is a new field receiving much attention.

Sleep Disorder Due to a General Medical Condition Sleep is often disturbed because of a medical illness, including arthritic and other pain syndromes, respiratory and cardiac diseases, and neurological diseases. Appropriate treatment can help ameliorate the sleep disturbance (e.g., analgesics for pain). Elderly persons are likely to have progressive cardiac, pulmonary, and neurological disorders that can affect respiration during sleep. Poorly controlled congestive heart failure may be associated with orthopnea and frequent awakenings, and sleep may also be disrupted by nocturnal diuresis. Some patients with irritable airways and gastroesophageal reflux develop acute bronchospasm that disturbs sleep; this nocturnal asthma can often be largely prevented by bedtime doses of histamine type 2 (H_2)-receptor blockers, such as cimetidine (Tagamet). Patients with chronic obstructive pulmonary disease are at greater risk for sleep apnea than healthy older adults. Patients with chronic obstructive pulmonary disease and sleep complaints should generally be considered candidates for sleep apnea screening. A Cheyne-Stokes pattern of breathing associated with fluctuations in blood oxygen saturation due to cardiac, pulmonary, or cerebral disorders can be detected by simple portable oximetry, polysomnography, or other investigative means and can lead to treatment by respiratory stimulants or nocturnal oxygen therapy. Sleep disorder affects most patients with Parkinson's disease and appears to be a function of disease severity. Disrupted sleep in Parkinson's disease is associated with nocturia, pain, stiffness, and difficulty turning over in bed and may also be caused by levodopa therapy. Substitution of sustained-release levodopa preparations, use of nasal vasopressin to control nocturia, temperate use of benzodiazepines, and antidepressants for depressed patients with parkinsonism have been suggested. Nonpharmacological approaches to managing sleep disorder in medically ill patients have not been experimentally investigated.

Recent large epidemiological studies concur that insomnia or daytime sleepiness in elderly persons is associated with impaired health, rated by both subjective and objective measures. In one study, incident self-reported insomnias that developed over a 3-year period were associated with incident medical and psychosocial burden, while the converse also held true: improved health over the 3-year study interval was associated with resolution of the original insomnia complaint. This ongoing study illustrates the secondary nature of many insomnia complaints.

Sleep Disorder Related to Another Mental Disorder Depression and Anxiety Disorders Complaints of insomnia in elderly persons should prompt diagnostic screening and appropriate therapies for depression and anxiety disorders, which are well-known causes of poor sleep. Attention to good sleep hygiene is an important part of any therapeutic regimen. Subjective complaints of sleep disturbance were shown to be a robust predictor of future depression in a large sample of nondepressed older persons.

Dementing Conditions and Delirium Sleep is also disturbed in dementing disorders. Episodes of nocturnal wandering and delirium can occur despite normal daytime functioning. The sleep of patients with Alzheimer's disease is marked by the increased duration and frequency of awakenings, decreased slow-wave and REM sleep, and (in more advanced stages) daytime napping. Similar changes also occur in other dementing disorders. Currently, no single, predictably effective treatment for these dementia-related sleep disturbances is available, but a search for contributing causes is an essential part of the clinical evaluation. In individual cases, frightening environmental stimuli, psychiatric complications of dementia, and associated medical disorders may impair sleep and respond well to specific interventions. Attention to good sleep hygiene in an effort to consolidate nighttime sleep (including discouraging napping, encouraging daytime exercise, and light therapy) and making the patient's environment as safe as possible for nighttime forays may help.

Sundowning Sundowning, defined as worsening of agitation and cognitive impairment in demented patients during the late afternoon and evening, is commonly observed by caregivers for patients with Alzheimer's disease and is associated with significant caregiver distress. The prevalence of sundowning in Alzheimer's disease is uncertain, but it seems to occur in about one-quarter of severely demented inpatients. Investigators have recently developed a valid and reliable method of quantifying sundowning and have shown an increased incidence during the winter months of low illumination in demented residents of nursing homes. Sundowning is commonly associated with nocturnal sleep disturbance.

Vulnerability to sundowning is believed to result from damage to the function of circadian systems by the neuropathologic process, responsible for the underlying dementia but some data suggest that caregiver factors may be important. Lowered tolerance for agitation because of caregiver fatigue and involvement in necessary tasks that reduce interaction with demented persons late in the day may increase agitation or bias reporting of diurnal behavioral rhythms. A decreased capacity of the impaired patient to adapt to normal rhythms of light, ambient activity, and sensory stimulation may play a role.

Numerous uncontrolled reports recommend the use of medications, particularly antipsychotic medications, for the management of sundowning, particularly when it is associated with fearfulness, hallucinations, delusions, and disruptive behavior. Low-dose haloperidol (Haldol) or serotonin-dopamine antagonists such as risperidone (Risperdal), which are less likely to increase cognitive impairment or cause extrapyramidal adverse effects, are often helpful when given before the patient's typical time of onset of sundowning behavior. However, systematic trials suggest that cholinesterase inhibitors (e.g., donepezil [Aricept]), selective serotonin reuptake inhibitors (e.g., sertraline [Zoloft]), and other nonneuroleptic drugs deserve consideration. In clinical care, approaches combining environmental modification (increased light late in the day, pleasurable distractions and activities, and increased attention from caregivers), more support for caregivers during vulnerable periods of the day, and serial trials of medications are frequently required to achieve satisfactory improvement. Bright light therapy is a promising but still experimental approach.

Substance-Induced Sleep Disorder Elderly patients may be taking a variety of prescribed medications (often from a number of physicians), herbal remedies, and over-the-counter medications. Some of these medications may cause secondary insomnia or may exacerbate sleep apnea. Inquiries about fatigue and daytime sleepiness can help detect underlying drug-induced sleep apnea, and physicians should make liberal use of the “brown bag” drug history, asking patients to bring all medications with them to their next appointment. This time-honored technique in geriatric medicine provides clues to potential adverse drug effects and is a useful tool in educating patients. The long-term use of centrally acting agents (e.g., antianxiety agents, antidepressants, sleeping medications, and alcohol) can induce a drug-related insomnia. In elderly persons, drug metabolism may be slowed, and long-term use of hypnotic drugs may result in excessive daytime sleepiness, impaired memory, and impaired psychomotor functioning. These effects may be subtle and must be specifically sought. Although both use and the abuse of alcohol decline with age, unsuspected alcohol abuse or even self-medication with alcohol as a sleep aid can impair sleep. The clinician must educate elderly insomniacs about the sleep-impairing effects of long-term sedative and alcohol use, encourage good sleep hygiene, and use behavioral techniques such as relaxation and stress reduction.

Ill-timed or excessive use of coffee, various teas, and soft drinks that contain significant amounts of caffeine or other xanthines can also produce disturbed sleep because of both the stimulant and the diuretic effects of these compounds. When assessing a sleep complaint, especially in elderly patients who may have increased sensitivity to drugs, the clinician must obtain comprehensive information among prescribed medications, over-the-counter medications, and recreational drugs.

SEDATIVE-HYPNOTIC DRUGS AND OTHER SLEEP AIDS

Sedative-hypnotic drugs can be indicated for short-term use to relieve transient insomnia; their long-term use usually results in habituation, loss of efficacy, a drug-dependency insomnia, and rebound insomnia and nightmares when drug use is discontinued. In the presence of undiagnosed sleep apnea, hypnotic drug use can increase the frequency, duration, and severity of apneic episodes. The adverse daytime effects of hypnotic drugs include impaired cognition, slowed psychomotor functioning, and injuries due to falls. A recent National Institutes of Health consensus paper on drugs and insomnia urges great restraint in the use of hypnotic drugs for other than temporary, situational, or intermittent conditions.

Melatonin, a naturally occurring, endogenous hormone, is gaining considerable popularity as a putative sleep aid. Preliminary studies of exogenous melatonin indicated sleep-promoting effects. However, use of melatonin at this time is extremely premature. Several recent clinical trails have failed to confirm a beneficial effect of melatonin on the sleep of elderly patients. Indeed, exogenous melatonin may suppress the endogenous melatonin rhythm. Finally, the adverse effects of chronic use and indeed the optimum dose of melatonin are unknown, and a recent case report documented delirium in a patient taking high doses.

Long-Term Use of Anxiolytic and Sedative-Hypnotic Drugs in Elderly Persons Large European and American surveys have confirmed high rates of long-term anxiolytic and sedative-hypnotic drug use in elderly persons with chronic sleep complaints. While most experts and many elderly insomniacs and their families agree that long-term use of sedative-hypnotic agents is undesirable, pharmacotherapy remains the most frequently used approach in primary care practice. The more important reasons for long-term sedative prescribing are unfamiliarity with either the risk of long-term hypnotic drug use or safe treatment alternatives. Focused, practice-based drug surveillance programs are useful ways to identify both physician targets for specialized education and patients in need of thoughtful clinical reassessment.

Drug reduction strategies, designed to protect elderly patients from the adverse cognitive and motor consequences of long-term drug use and to identify and ameliorate rebound insomnia, have been tested in nursing home populations. Education of caregivers in the effects of sedative-hypnotic agents in geriatric patients can reduce their use.

Use of low-dose sedative antidepressants, such as trazodone (Desyrel), doxepin (Sinequan), and amitriptyline (Elavil) is a popular approach to chronic insomnia in nondepressed older patients but has not been formally tested for safety and effectiveness in controlled trials.

Significant advances have occurred in the nonpharmacological treatment of chronically disturbed sleep in elderly persons. Cognitive-behavioral treatment, sleep-restriction therapy, and relaxation training have all proved acceptable to elderly persons and effective in small controlled trials. All can improve sleep latency and waking after falling asleep, and cognitive-behavioral and sleep-restriction therapies can produce sustained improvement in sleep efficiency. Aerobic exercise is a useful treatment for moderate sleep complaints in healthy, sedentary older adults. Endurance training consisting of low-impact aerobics and brisk walking can also improve sleep quality, as can good sleep hygiene and daytime bright light exposure.

SUGGESTED CROSS-REFERENCES

A general discussion of sleep disorders appears in Chapter 21, and an explanation of the basic science of sleep appears in Section 1.19. Normal aging is discussed in Section 51.2c. Alzheimer's disease is discussed in Section 51.3e. Anxiety and mood disorders are discussed in Chapter 14 and Chapter 15.

SECTION REFERENCES

Asplund R: Daytime sleepiness and napping amongst the elderly in relation to somatic health and medical treatment. J Intern Med *239:*261, 1996.

Avorn J, Soumerai SB, Everitt DE, Ross-Degnan D, Beers MH, Sherman D, Salem-Schatz SR, Fields D: A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. N Engl J Med *327:*168, 1992.

Bliwise DL: Sleep in normal aging and dementia. Sleep *16:*40, 1993.

*Bliwise DL: Sleep disorders. In *Psychiatric Care in the Nursing Home*, WE Reichman, PR Katz, editors. Oxford University Press, New York, 1996.

Campbell SS, Terman M, Lewy AJ, Dijk DJ, Eastman CI, Boulos Z: Light treatment for sleep disorders: Consensus report. V. Age-related disturbances. J Biol Rhythms *10:*151, 1995.

Dorsey CM, Lukas SE, Teicher MH, Harper D, Winkelman JW, Cunningham SL, Satlin A: Effects of passive body heating on the sleep of older female insomniacs. J Geriatr Psychiatry Neurol *9:*83, 1996.

Foley DJ, Monjan AA, Brown SL, Simonsick EM, Wallace RB, Blazer DG: Sleep complaints among elderly persons: An epidemiologic study of three communities. Sleep *18:*P425, 1995.

*Gillin JC: Psychiatric disorders. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Haimov I, Lavie P: Melatonin—a chronobiotic and soporific hormone. Arch erontol Geriatr *24:*167, 1997.

Hauri P: Primary insomnia. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Hublin C, Kaprio J, Partinen M, Heikkila K, Koskenvuo M: Daytime sleepiness in an adult, Finnish population. J Intern Med *239:*417, 1996.

Hudgel DW: Treatment of obstructive sleep apnea. A review. Chest *111:*528, 1997.

*King AC, Oman RF, Brassington GS, Bliwise DL, Haskell WL: Moderate intensity exercise and self rated quality of sleep in older adults. A randomized controlled trial. JAMA *277:*32, 1997.

Krieger J: Breathing during sleep in normal subjects. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Little JT, Satlin A, Sunderland T, Volicer L: Sundown syndrome in severely demented patients with probable Alzheimer's disease. J Geriatr Psychiatry Neurol *8:*103, 1995.

Livingston G, Blizard B, Mann A: Does sleep disturbance predict depression in elderly people? A study in inner London. Br J Gen Pract *43:*445, 1993.

*Lugaresi E, Cirignotta F, Montagna P: Snoring: Pathogenic, clinical, and therapeutic aspects. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

McCurry SM, Logsdon RG, Vitiello MV, Teri L: Successful behavioral treatment of reported sleep problems in elderly caregivers of demential patients: A controlled study. Gerontol B Psychol Sci Soc Sci *53B:*122, 1998.

Miles LE, Dement WC: Sleep and aging. Sleep *3:*119, 1980.

Montplaisir J, Godbout R: Restless legs syndrome and periodic movements during sleep. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Morin CM, Wooten V: Psychological and pharmacological approaches to treating insomnia: Critical issues in assessing their separate and combined effects. Clin Psychol Rev *16:*521, 1996.

*Morin CM, Mimeault V, Gagne A: Nonpharmacological treatment of late-life insomnia. J Psychosom Res *46:*103, 1999.

Myers BL, Badia P: Changes in circadian rhythms and sleep quality with aging: Mechanisms and interventions. Neurosci Biobehav Rev *19:*553, 1995.

Newman AB, Enright PL, Manolio TA, Haponik EF, Wahl PW: Sleep disturbance, psychosocial correlates, and cardiovascular disease in 5201 older adults: The Cardiovascular Health Study. J Am Geriatr Soc *45:*1, 1997.

Nicholson AN, Bradley CM, Pascoe PA: Medications: Effect on sleep and wakefulness. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Prinz PN, Poceta JS, Vitiello MV: Sleep in the dementing disorders. In *Handbook of Neuropsychology*, vol 4, F Boller, J Grafman, editors. Elsevier, Amsterdam, 1990.

Prinz PN, Vitiello MV, Raskind MA, Thorpy MJ: Geriatrics: Sleep disorders and aging. N Engl J Med *323:*520, 1990.

Robinson RW, Zwillich CW: Drugs and sleep respiration. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Wetter TC, Pollmacher T: Restless legs and periodic leg movements in sleep syndromes. J Neurol *244:*S37, 1997.

Winget C, Vernikos-Denellis J, Cronin S, Leach CS, Rambaut PC, Mack PB: Circadian rhythm asynchrony in man during hypokinesis. J Appl Physiol *33:*640, 1982.

Wooten V: Medical causes of insomnia. In *Principles and Practice of Sleep Medicine*, ed 2, MH Kryger, T Roth, WC Dement, editors. Saunders, Philadelphia, 1994.

Wooten V: Sleep disorders in psychiatric illness. In *Sleep Disorders Medicine*, S Chokroverty, editor. Butterworth-Heinemann, Boston, 994.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3C ANXIETY DISORDERS

IRA M. LESSER, M.D.

The clinical presentation of anxiety disorders in elderly populations may not be as dramatic as that of mood, cognitive, and psychotic disorders. However, anxiety symptoms and syndromes in elderly persons are not uncommon, often occur comorbidly with other psychiatric and medical conditions, and are frequently under- or misdiagnosed.

An erroneous belief that anxiety may be a natural accompaniment of aging has probably contributed to a lack of serious attention for these conditions. Relatively few studies have focused specifically on the phenomenology, etiology, and treatment response of anxiety disorders in elderly patients; much of what is known has been extrapolated from studies with younger patients. The prevalence rates of full syndromal anxiety disorders in the elderly are not high; however, subsyndromal anxiety states are very common and like subsyndromal mood disorders, they can adversely affect health and quality of life.

EPIDEMIOLOGY

The Epidemiologic Catchment Area Study (ECA) reported that the prevalence of anxiety disorders in older groups is generally lower than in younger ones, with few subjects reporting an anxiety disorder beginning after age 65. These findings have been confirmed in epidemiological studies in Canada and in England, although differences in instruments, populations studied, and definition of a “case” have led to somewhat different rates. For example, in the ECA study, the prevalence of all anxiety disorders in those over 65 years of age was reported to be 5.5 percent; in Edmonton, Canada, the prevalence was 3.5 percent; and studies in London have found the prevalence rate to range from 1.7 to nearly 10 percent. The low prevalence rates may reflect several factors: the true clinical picture, poorer recall (especially for events such as panic attacks), reluctance to discuss symptoms, or a cohort effect in which the generation of elderly persons may have been at less risk of an anxiety disorder than more-recent generations.

Adults over age 65 years have a lower prevalence of panic disorder than all other age groups; this is true for whites and African-Americans and for men and women. However, for Hispanic women, rates increase linearly with age, while panic is rare for older Hispanic men. New cases of panic disorder in this age group are quite rare, and data show that panic disorder is less likely to be comorbid with depressive disorders and alcohol use disorders in those over 65 years of age than in younger people. For phobias, although individuals over age 65 have lower rates than those in younger age groups, the differences are not as striking, with substantial rates reported in many studies (up to 12 percent of community-residing adults).

For generalized anxiety disorder, rates are lower in older age groups, but there is still a considerable 1-year prevalence rate in persons over 65 years (1 to 7 percent depending upon the strictness used in defining generalized anxiety disorder). Women have a higher prevalence rate than men; the rates are approximately equal for whites and African-Americans, but Hispanics have a low prevalence rate. Obsessive-compulsive disorder has a lower prevalence in the over-65 age group than in those below age 45 and a prevalence approximately equal to that for people between 45 and 65. The lowest prevalence was found in older Hispanic men. Few data exist regarding the incidence and prevalence of posttraumatic stress disorder and social phobia in elders, although symptoms of both disorders may persist into advanced years in those with an earlier onset. There is a suggestion that rates for phobias and general anxiety disorder are higher in nondemented institutionalized adults than in those residing in the community.

ETIOLOGY

Understanding the determinants of an elderly person's anxiety requires giving careful consideration to coexisting medical and psychiatric illnesses. Beyond that, one must search for psychosocial determinants. Some authors have suggested that rather than being a signal affect (warning the individual of intrapsychic conflict), anxiety in the elderly person is largely related to object loss and the fear of depletion of external resources.

The relationship of anxiety to bereavement must be considered. Studies of surviving spouses have shown that they have appreciable anxiety for at least several years after the death of a spouse. To what degree this relates to resulting social isolation or to fears of separation or abandonment is not clear, but the increasing isolation of elderly adults is a major problem. Retirement can be a particularly difficult milestone for elderly persons, leaving them with little direction or definition of self. Real-life burdens such as financial problems, crime and fears for personal safety, caring for an ill spouse, poor housing, and difficulty in obtaining medical care are significant stressors that can precipitate anxiety or worsen preexisting anxiety disorders.

Questions have been raised about the role of death anxiety in this age group. Some authors have suggested that younger people are more concerned by this than their elders, who may have come to a peaceful resignation regarding their mortality. How one deals with this in later life is related to the individual's mastery of earlier developmental phases. In Erik Erikson's epigenetic model, the outcome of the last stage, integrity versus despair, is a culmination of many previous intrapsychic and interpersonal challenges. Many elderly persons, unhappy with their lives and their legacy, may struggle with an existential angst, facing death frightened and alone, yet having difficulty articulating this struggle.

No data suggest that older people are more predisposed to anxiety from a neurobiological perspective. In fact, it has been hypothesized that the lower incidence and prevalence of panic and other anxiety disorders in the elderly population may relate to the reduced activity of several neurotransmitter systems that occurs with aging, thus increasing the threshold for developing panic attacks. There have been no reports of challenge studies using substances such as lactate or yohimbine (Yocon), which have been reported to precipitate panic attacks in panic-prone younger individuals. Likewise, neuroimaging investigations, including studies of cerebral blood flow and metabolism, have not been conducted with older subjects. More biologically based studies of anxiety in elderly patients are needed to understand the contribution of biological factors.

DIAGNOSIS AND CLINICAL FEATURES

One current conceptual approach to anxiety divides symptoms into the cognitive and the somatic. Clinical experience suggests that older patients often focus more on the somatic symptoms than on the cognitive aspects, which may be minimized. Since elderly persons have more physical illnesses than do younger persons, their emphasis on somatic symptoms can cause difficulty in differential diagnosis, particularly for primary care physicians. In the few studies that have compared the symptom profiles and configurations between young and elderly patients with anxiety disorders, occasional small differences were found in severity (e.g., fewer symptoms of panic attacks, less avoidance behavior) rather than in presentation of specific symptoms. Because of the paucity of data, diagnoses of anxiety disorders in the geriatric age group must rest on the use of criteria that have been validated on younger samples, however imperfect the fit may be.

Comorbid Conditions Anxiety disorders often coexist with other psychiatric disorders and with medical disorders. Whether the anxiety disorder preceded or followed the other disorder, the comorbid conditions typically lead to more severe symptoms, a protracted course, and possible poorer response to treatment interventions.

Anxiety typically accompanies serious physical illness, particularly when the conditions are chronic, associated with pain, and possibly terminal. Although data are lacking about the true prevalence of anxiety disorders among physically ill elderly patients, symptoms of anxiety are frequent. Significant comorbid illnesses include central nervous system disorders (e.g., Parkinson's disease, dementia, delirium, stroke), cardiovascular disease (e.g., angina, arrhythmia), respiratory disease (e.g.,

asthma, pulmonary embolus), endocrine disorders (e.g., hyperthyroidism, hypothyroidism, parathyroid disease, diabetes), and sensory impairment. Numerous medications can cause anxiety-like symptoms (e.g., bronchodilators, calcium channel inhibitors, a-adrenergic receptor agonists). Finally, intoxication and withdrawal states from alcohol and other substances (e.g., benzodiazepines, barbiturates) can have anxiety as the presenting symptom. Disentangling the symptoms in these cases can be very difficult, and treatment for one disorder may worsen the other.

> A 65-year-old woman was seen in consultation for episodes of severe anxiety. Several years before, she developed asthma, became steroid dependent, and spent most of her time in a wheel-chair. She had a devoted husband, but having primary custody of their 7-year-old grandchild and severe financial difficulties were major stressors. Soon after the asthma began, she developed classic symptoms of spontaneous panic attacks, with subsequent avoidance behavior. With each panic episode, her respiratory condition worsened; she often needed emergency care. Her physician found it increasingly difficult to ascertain whether she was having a panic attack with ensuing breathing difficulties or an asthmatic attack which frightened her so that she panicked. A further complication was the number of medications she needed to treat the asthma, many of which also could have caused increased anxiety. She obtained moderate relief of her panic episodes with low-dose benzodiazepines; however, ultimately she died of status asthmaticus.

Anxiety disorders frequently present with depressive disorders. Comorbid anxiety symptoms are very prominent in older patients who have depressive disorders. For example, studies have reported that one third to one half of elderly patients with a major depressive disorder have significant anxiety, and elderly patients with phobias or generalized anxiety disorder or both also have a high prevalence of depression. A particularly high comorbidity of anxiety and depression (80 percent) has been noted in institutionalized elderly patients, which suggests that differentiating these two conditions may not be possible in frail elderly patients. Recent data from younger age patients indicate that suicide risk is greater when depression and anxiety occur together than when either occurs alone. As the rates of suicide continue to increase well into the late decades of life, assessment of this co-occurrence is of utmost importance.

Increasing evidence indicates that alcohol-dependent patients have a high prevalence of anxiety disorders, particularly panic disorder and social phobia. Alcohol abuse and dependence in the older population is a significant problem. One-year prevalence rates for alcoholism are over 3 percent for men over 65 years, making it the third most common psychiatric disorder in this group.

Unique to the elderly population is the association of anxiety and states of cognitive decline. Symptoms of anxiety and agitation often accompany the early stages of dementing illnesses and, in some cases, are the major behavioral disturbances that cause families to bring patients for evaluation. Particularly in patients with frontotemporal dementia, early symptoms may be irritability, agitation, and compulsive behavior. For those with age-associated memory impairment and those with early Alzheimer's disease, recognition that memory may be failing and that other cognitive tasks cannot be accomplished easily often evoke feelings of loss of control and anxiety. Because the symptoms of cognitive decline and the behavioral disturbances can be so pronounced, the contribution of anxiety symptoms is easily overlooked.

DIFFERENTIAL DIAGNOSIS

In addition to understanding the relative contributions of comorbid conditions to anxiety disorders, the clinician must consider other conditions in the differential diagnosis. Whenever an anxiety disorder presents for the first time in an elderly person, the clinician must undertake a careful search for medical illness. In the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), anxiety disorder due to a general medical condition and substance-induced anxiety disorder have replaced organic anxiety disorder. General medical conditions that are thought to be etiologically and temporally related to anxiety conditions would be listed in these diagnoses, along with a specifics indicating the type of anxiety (e.g., anxiety disorder due to hyperthyroidism, with panic attacks). Because patients have more illnesses and take more medications with advancing age, elderly individuals are more prone to have the symptoms of an anxiety disorder be a consequence of their illness or their treatment.

Another new DSM-IV disorder is acute stress disorder. This is related to posttraumatic stress disorder in that a severe stressor must be recognized. In elderly adults, war, concentration camp, refugee, and acculturation experiences as well as early abuse histories are among the stressors that can cause posttraumatic stress disorders. In acute stress disorder, the symptoms are experienced during or immediately after the trauma and last anywhere from 2 days to 1 month (as opposed to a longer time frame for posttraumatic stress disorder). This diagnosis may have considerable relevance to elderly adults, given the many traumas to which they are vulnerable. A diagnosis of mixed anxiety-depressive disorder is included in DSM-IV but is listed as a disorder in need of further study. Its relevance to the elderly population is undetermined. Adjustment disorder with anxious mood should be considered when the anxiety is of short duration and a stressor can be identified. The somatoform disorders, particularly somatization disorder and hypochondriasis, should be considered when the presentation is strongly focused upon somatic symptoms. When the symptoms seem focused upon sleep problems, the primary sleep disorders should be considered. Alcohol and other substance abuse and dependence can be significant problems in elderly adults and should be considered in the differential diagnosis.

TREATMENT

Treatment should follow a thorough evaluation of medical illnesses, and assessments of medications being taken, cognitive function, psychiatric symptoms, current stressors, and characteristic modes of functioning. Whenever possible, the clinician should first use nonpharmacological interventions: education, environmental manipulations, and the full range of psychotherapeutic approaches. Relaxation techniques can decrease anxiety in elderly adults. Preliminary data show that cognitive and behavioral approaches are as successful with the older anxious patient as with younger patients. Traditional psychotherapy or such modifications as the life review or reminiscence can have great benefit. Exploration of the spiritual aspects in an older person's life can also be useful and provide additional support.

Pharmacological treatment can be effective but must be approached with caution in geriatric populations. Altered pharmacokinetics, altered pharmacodynamics, interactions with other medications and with other illnesses, and poor compliance must all be considered. As a general rule, starting with one quarter to one half of the dosage used for a younger patient is a good idea, regardless of the class of medication chosen.

Benzodiazepines may be effective for the short-term treatment of anxiety. Preference should be given to short half-life medications (e.g., alprazolam [Xanax], oxazepam [Serax], lorazepam [Ativan]), particularly those metabolized by direct conjugation in the liver, as they are less likely to accumulate; however, the half-life may still be longer than that seen in younger people. The very short-acting benzodiazepines should probably be avoided because their use in elderly patients has been associated with confusion, anterograde amnesia, and agitation. Short-term use is rarely associated with the need for increasing dosages, but longer-term use may result in dependence. These medications should be used with caution because of possible adverse effects. The major potential adverse effects are those related to sedation, cognitive dysfunction (e.g., reduced short-term memory and attention), psychomotor performance, and cerebellar dysfunction. Taking sedative or hypnotic medications also increases the risk of falling; therefore, caution is indicated when using benzodiazepines. When the medications are discontinued, symptoms of rebound or withdrawal may occur, but these symptoms do not seem to be more severe in elderly patients than in younger patients. Buspirone (Buspar) is reportedly effective in elderly adults, with few adverse effects, but the data are limited.

Antidepressants should be considered, particularly when mixed anxiety-depression or panic disorder are present. Among the tricylic drugs, desipramine (Norpramin) and nortriptyline (Pamelor) are the most widely used for elderly adults. They may have effectiveness in generalized anxiety disorder, panic disorder, and mixed states. Because of an even more favorable side-effect profile, increasing clinical experience suggests the use of the selective serotonin reuptake inhibitors (citalopram [Celexa], fluoxetine [Prozac], fluvoxamine [Luvox], paroxetine [Paxil], sertraline [Zoloft]) for panic disorder, obsessive-compulsive disorder, and mixed states. Limited data exist on treating anxiety disorders with other, newer antidepressants (e.g., bupropion [Wellbutrin], nefazodone [Serzone], venlafaxine [Effexor]). Trazodone (Desyrel) has been used for its sedating properties, although again, no data exist regarding its anxiolytic properties. Although care needs to be taken with their use, the monoamine oxidase inhibitors (MAOIs) can be useful medications and may be tried in refractory patients with panic disorder. There are few studies specific to the elderly population on which to base recommendations.

Antipsychotic medications should not be used routinely in anxious patients unless the anxiety or agitation is secondary to dementia, delirium, or psychosis or is refractory to first-line approaches. If antipsychotic medications are necessary, the best choices are medications in either the high-potency, conventional dopamine receptor antagonists (e.g., haloperidol [Haldol], thiothixene [Navane], fluphenazine [Prolixin]) or the serotonin-dopamine antagonists (e.g., risperidone [Risperdal], olanzapine [Zyprexa], and quetiapine [Seroquel]). Emerging evidence indicates that the novel antipsychotics have the most benign adverse-effect profile regarding extrapyramidal symptoms, are possibly the least likely to cause tardive dyskinesia in elderly individuals, and may have the fewest detrimental effects upon cognitive performance.

SUGGESTED CROSS-REFERENCES

A full discussion of anxiety disorders can be found in Chapter 15. Further discussion of geriatric comorbid states and the differential diagnosis of depression appears in Section 51.3d, of drug and alcohol use in Section 51.3h, of sleep disorders in Section 51.3b, and of dementing disorders in Section 51.3e. Individual psychotherapy for elderly adults is discussed in Section 51.4i, and psychopharmacology for elderly adults is discussed in Section 51.4f.

SECTION REFERENCES

Beck JG, Stanley MA: Anxiety disorders in the elderly: The emerging role of behavior therapy. Behav Ther *28:*83, 1997.

Bland RC, Newman SC, Orn H: Prevalence of psychiatric disorders in the elderly in Edmonton. Acta Psychiatr Scand Suppl *338:*57, 1988.

*Blazer DG: Generalized anxiety disorder and panic disorder in the elderly: A review. Harv Rev Psychiatry *5:*18, 1997.

Copeland JRM, Gurland BJ, Dewey ME, Kelleher MJ, Smith AMR, Davidson IA: Is there more dementia, depression and neurosis in New York: A comparative study of the elderly in New York and London using the computer diagnosis AGECAT. Br J Psychiatry *151:*466, 1987.

Erikson E: *Childhood and Society*. Norton, New York, 1950.

*Flint A: Epidemiology and comorbidity of anxiety disorders in the elderly. Am J Psychiatry *151:*640, 1994.

Flint AJ, Cook JM, Rabins P: Why is panic disorder less frequent in late life? Am J Geriatr Psychiatry *4:*96, 1996.

*Hocking LB, Koenig HG: Anxiety in medically ill older patients: A review and update. Int J Psychiatry Med *25:*221, 1995.

Jackson CW: Obsessive-compulsive disorder in elderly patients. Drugs Aging *7:*438, 1995.

Jacobs S, Hansen F, Kasl S, Ostfeld A, Berkman L, Kim K: Anxiety disorders during acute bereavement: Risk and risk factors. J Clin Psychiatry *51:*269, 1990.

*Jeste DV, Rockwell E, Harris MJ, Lohr JB, Lacro J: Conventional vs newer antipsychotics in elderly patients. Am J Geriatr Psychiatry *7:*70, 1999.

Kohn R, Westlake RJ, Rasmussen SA, Marsland RT, Norman WH: Clinical features of obsessive-compulsive disorder in elderly patients. Am J Geriatr Psychiatry *5:*211, 1997.

Levendusky PG, Hufford MR: The application of cognitive-behavior therapy to the treatment of depression and related disorders in the elderly. J Geriatr Psychiatry *30:*227, 1997.

Lindesay J, Briggs K, Murphy E: The Guy's/Age Concern survey. Prevalence rates of cognitive impairment, depression and anxiety in an urban elderly community. Br J Psychiatry *155:*317, 1989.

Manela M, Katona C, Livingston G: How common are the anxiety disorders in old age? Int J Geriatr Psychiatry *11:*65, 1996.

Mulsant BH, Reynolds CF, Shear K, Sweet RA, Miller M: Comorbid anxiety disorders in late-life depression. Anxiety *2:*242, 1996.

Parmelee PA, Katz IR, Lawton MP: Anxiety and its association with depression among institutionalized elderly. Am J Geriatr Psychiatry *1:*46, 1993.

Raj BA, Corvea MH, Dagon EM: The clinical characteristics of panic disorder in the elderly: A retrospective study. J Clin Psychiatry *54:*150, 1993.

Robins LN, Regier DA, editors: *Psychiatric Disorders in America: The Epidemiologic Catchment Area Study*. Free Press, New York, 1991.

*Sadavoy J, LeClair KJ: Treatment of anxiety disorders in late life. Can J Psychiatry *42*(Suppl):28S, 1997.

Salzman C: Treatment of anxiety and anxiety-related disorders. In *Clinical Geriatric Psychopharmacology*, ed 3, C Salzman, editor. Williams & Wilkins, Baltimore, 1998.

*Salzman C, Lebowitz B, editors: *Anxiety in the Elderly: Treatment and Research*. Springer, New York, 1991.

Salzman C, Sheikh JI: Diagnosis of anxiety and anxiety-related disorders. In *Clinical Geriatric Psychopharmacology*, ed 3, C Salzman, editor. Williams & Wilkins, Baltimore, 1998.

Schneider LS: Overview of generalized anxiety disorder in the elderly. J Clin Psychiatry *57*(Suppl):34, 1996.

Schweizer E, Case WG, Rickels K: Benzodiazepine dependence and withdrawal in elderly patients. Am J Psychiatry *146:*529, 1989.

Sheikh JI, King RJ, Taylor CB: Comparative phenomenology of early-onset versus late-onset panic attacks: A pilot survey. Am J Psychiatry *148:*1231, 1991.

Shorr RI, Robin DW: Rational use of benzodiazepines in the elderly. Drugs Aging *4:*9, 1994.

Smith SL, Sherill KA, Colenda CC: Assessing and treating anxiety in elderly persons. Psychiatr Serv *46:*36, 1995.

Stoudemire A, Moran MG: Psychopharmacologic treatment of anxiety in the medically ill elderly patient: Special considerations. J Clin Psychiatry *54*(Suppl): 27, 1993.

Swales PJ, Solfvin JF, Sheikh JI: Cognitive-behavioral therapy in older panic disorder patients. Am J Geriatr Psychiatry *4:*46, 1966.

Zisook S, Schneider D, Schucter SR: Anxiety and bereavement. Psychiatr Med *2:*83, 1990.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3D MOOD DISORDERS

GEORGE S. ALEXOPOULOS, M.D.

Geriatric mood disorders have important medical, social, and financial consequences. Mood disorders cause suffering to elderly patients and their families, exacerbate medical illnesses, and contribute to disability and require expensive support systems. Geriatric mood disorders can be effectively treated in most cases; therefore, it is crucial to identify mood syndromes and treat them appropriately.

Geriatric depression is underrecognized. Clinicians and patients often attribute depressive symptoms to the aging process. Another reason is that older persons emphasize somatic symptoms and underreport depressed mood. Finally, geriatric depression often occurs in the context of medical or neurological brain diseases whose symptoms are similar to the symptoms of depression. In some cases the overlap of symptomatology is such that depression can only be diagnosed after successful treatment with antidepressant medications. In patients at low risk for adverse effects, it is best to offer a trial of antidepressant medications rather than to perpetuate diagnostic ambiguity. The diagnosis of depressive disorder may be difficult in patients who have dementing disorders. Unreliable reports by the patients and fluctuation of depressive symptomatology over time may interfere with diagnosis because the examiner may see the patients during a well period. Caregiver reports can be helpful in such cases because caregivers observe the patients for long periods.

EPIDEMIOLOGY

The Epidemiologic Catchment Area study documented that major depressive disorder is less frequent in elderly compared to younger adults. The overall prevalence of major depressive disorder among persons aged 65 years or older was estimated to be 1.4 percent in women and 0.4 percent in men with an overall prevalence of 1 percent. This prevalence rate is approximately one fourth of that of younger adults. These findings were confirmed by other investigators. Approximately 2 percent of the elderly population suffer from dysthymic disorder and 4 percent have an adjustment disorder with depressed mood. A rather large percentage of elderly persons, approximately 15 percent, have depressive symptomatology that does not meet criteria for a specific depressive syndrome. Although subsyndromic depression is more frequent than syndromic depression, even subsyndromic depression does not appear to increase with age.

The relatively low prevalence of depressive syndromes identified in elderly populations may in part be due to methodological problems, including the tendency of elderly persons to express psychiatric symptoms in somatic terms, reluctance to recall and report psychiatric symptoms, and use of diagnostic categories unsuitable for the elderly. In addition to methodological problems, a cohort effect has been suggested as a reason for the low prevalence of depression in the elderly. It has been observed that the prevalence of major depressive disorder has been increasing in recent birth cohorts compared to cohorts born in the early part of the twentieth century. The cohort effect is also supported by suicide studies showing that white men born after 1922 had lower suicide rates than cohorts born earlier. Depression appears to have an increasingly earlier onset in cohorts born in recent years and there is a narrowing in the differential risk for depression between genders. Finally, depression appears to be more frequent between 1960 and 1975 than in earlier years. This increase in the rates of major depressive disorder over time has recently been confirmed in large cross-national samples.

Depression frequently occurs in medical populations. In patients treated in primary care settings, depression was identified in 17 to 37 percent of patients. Approximately 30 percent of these patients have major depressive disorder while the remainder have a variety of depressive syndromes that could benefit from medical attention. In medically hospitalized patients major depressive disorder occurs in 11 percent, and less severe, yet clinically significant depressive symptomatology is identified in 25 percent of the population. The corresponding figures for patients treated in long-term care settings are 12 percent and 30 percent.

Geriatric depression is associated with female sex, divorced or separated marital status, low socioeconomic level, poor social support, and recent adverse and unexpected life events. Severe impairment in medical health resulting in disability constitutes an important risk factor. Neurological and endocrinological disorders as well as chronic obstructive pulmonary disease, myocardial infarction, and malignancies are associated with an increased incidence of depression.

Depressive syndromes occur in the older geriatric population at a rate lower than that in younger adults. However, very high rates of depression occur in medically ill, disabled, or institutionalized elderly persons. With the increase of the elderly population and the aging of birth cohorts who appear to be particularly prone to depression an unprecedented number of elderly patients are expected to be in need of psychiatric attention.

ETIOLOGY

Although the etiology of geriatric depression remains unclear, some differences have been noted in measures of brain structure and function as well as in neurochemical and neuroendocrine measures in the periphery.

Studies of monoamine metabolites in peripheral fluids yielded relatively inconsistent findings. In mixed-age candidates with depression a significant correlation was noted between age of onset and plasma 3-methoxy-4-hydroxyphenylglycol (MHPG). Studies using agonists of the platelet a $_2$ adrenoreceptor have shown increased binding in mixed age and elderly depressed patients. Increased a $_2$-adrenergic receptor binding was associated with early-onset depression. A weak correlation was also noted between age of onset and secondary MHPG. Geriatrics studies are needed to replicate these findings and examine if low urine and plasma MHPG and high a_2-receptor binding can identify early-onset, perhaps familial, depression from late-onset depression, an entity associated with medical or neurological disorders.

Platelet monoamine oxidase (MAO) activity has been thought to be a marker of vulnerability for bipolar I disorder, alcohol use disorder, and perhaps other disorders. In elderly patients with depression, high platelet MAO activity has been associated with presence of medical illness. Platelet MAO activity appears to be elevated in patients with dementia of the Alzheimer's type with and without depression. Hospitalized women with late-onset depression were found to have higher platelet MAO activity than elderly women with early-onset depression. Studies are needed to examine if high platelet MAO activity can identify subgroups of late-onset depression patients with high medical morbidity or with subclinical Alzheimer's disease. Elevated platelet MAO activity was reported in elderly patients with depression with “reversible dementia” compared to nondemented patients who had depression. Future research may examine if high platelet MAO activity can identify those with reversible dementia who develop permanent dementia on follow-up.

Platelet ^{3}H-imipramine binding is thought to reflect the function of a modulating system of serotonin uptake. With one exception, low platelet ^{3}H-imipramine binding has been found in geriatric patients who have depression compared to controls. Platelet ^{3}H-imipramine binding is lower in primary depressive disorder compared to depressive disorder occurring in the context of medical illness. Furthermore, reduced platelet ^{3}H-imipramine binding is associated with poor response to treatment with antidepressant drugs. Since platelet ^{3}H-imipramine binding is within normal range in patients with dementia of the Alzheimer's type it is important to study if this test can identify patients with clinical or subclinical states of dementia.

Plasma cortisol escape from dexamethasone (Decadron) suppression (abnormal dexamethasone supression test [DST]) is more frequent in geriatric compared to younger patients with depression. Therefore the DST is more sensitive in geriatric than in younger patients. However, abnormal DST occurs in one third of patients with dementia, thus suggesting a low specificity for geriatric depression. Although inadequate geriatric data exist, studies of younger adults suggest that lack of

normalization of DST after adequate clinical response is associated with early relapse of depression.

Blunted thyrotropin response to thyrotropin-releasing hormone (TRH) has been observed in 25 percent of depression patients and this abnormality may be trait-related in at least some patients. Blunted thyrotropin response has been reported in geriatric patients with major depressive disorder but also in patients with dementia of the Alzheimer's type. Therefore, blunted thyrotropin response to TRH is not specific for geriatric depression.

Recent development of brain imaging techniques has allowed studies of brain structure and function. Brain computed tomography (CT) studies observed enlargement of lateral brain ventricles in geriatric depression. Ventricular enlargement is more pronounced in late-onset compared to similarly aged early-onset patients with depressive disorder; the ventricles of late-onset depression patients appear to be comparable to those of patients with dementia of the Alzheimer's type. The biological meaning of ventricular enlargement is unclear but it appears to be associated with poor response to antidepressant treatment, as well as functional abnormalities of depression including hypercortisolemia, hypothyroidism, decreased dopamine beta hydroxylase, and increased concentrations of 5-hydroxyindoleacetic acid (5-HIAA) in cerebrospinal fluid (CSF).

Brain magnetic resonance imaging (MRI) studies showed white matter hyperintensity in elderly patients with depressive disorder. This finding was more frequent and more pronounced in late- compared to early-onset depression and was associated with poor response to drug treatment. Other investigators observed a positive correlation between age of illness onset and cortical atrophy and raised the question whether late-onset depression patients have a high risk for dementia of the Alzheimer's type. This suggestion is supported by the observation that geriatric depressed patients who develop dementia on an average 38-month follow-up had later age of illness onset and more sulcal widening than those who did not develop dementia.

Functional neuroimaging studies of major depressive disorder suggest a prefrontal dysfunction in depression. The prefrontal cortex integrates sensory, emotional, and neuroendocrine functions that often are impaired during depression. The frontal lobe is connected to the basal ganglia through five contiguous, nonoverlapping parallel zones (cortico-striato-pallido-cortical pathways). Studies of major depressive disorder have shown variable results, but most have suggested hypoactivity in the left or bilateral frontal regions, including the dorsolateral, inferior, and medial/anterior cingulate. Some studies have shown hyperactivity in these brain areas suggesting that heterogeneous functional abnormalities may exist in depression. Decreased metabolic rate in the caudate nucleus and left-side decrease in caudate blood flow have been reported.

Three cortico-striato-pallido-cortical pathways may be relevant to depression. Damage of the orbitofrontal circuit may lead to disinhibition, irritability, and diminished sensitivity to social cues. Damage of the anterior cingulate may result in apathy and reduced initiative. Damage of the dorsolateral circuit may result in difficulties in set shifting, learning, and word-list generation. These behavioral abnormalities resemble in part the depressive syndrome. Resting hypoactivity in the left dorsolateral prefrontal cortex and the angular gyrus have been correlated with psychomotor retardation and depressed mood. Increased activity in the posterior cingulate and the inferior parietal lobe has been associated with anxiety. Hyperactivity in the medial prefrontal cortex has been correlated with cognitive deficits in depression.

It has been proposed that cerebrovascular disease may predispose, precipitate, or perpetuate some geriatric depressive syndromes. The vascular depression hypothesis is supported by the comorbidity of depression and vascular disease and risk factors and the association of ischemic lesions to distinctive behavioral symptomatology. Elderly patients with vascular depression have greater overall cognitive impairment and disability than the nonvascular depression group. Fluency and naming are most impaired in vascular depression. Patients with “vascular depression” have more apathy and retardation and less agitation as well as less guilt and greater lack of insight. Disruption of prefrontal systems or their modulating pathways by single lesions or by an accumulation of lesions exceeding a threshold are hypothesized to be central mechanisms in “vascular depression.”

In addition to frontal system's abnormalities, dysfunction in anterior temporal-limbic regions has also been described. Resting single photon emission computed tomography (SPECT) studies of elderly patients with depression have demonstrated reduced global cerebral blood flow (CBF) and one study observed decreased orbitofrontal and anterior temporal regional CBF. A resting positron emission tomography (PET) study of depressed geriatric patients has shown global reduction in CBF as well as reduction in all cortical and subcortical regions studied. Abnormalities in postrolandic cortex have also been described.

DIAGNOSIS AND COMORBIDITY

Depression in Medically Ill Patients Geriatric depression, as a rule, occurs in the context of medical disorders. Sometimes, these conditions predispose or trigger depression, for example, depression occurring after the onset of hypothyroidism or after a myocardial infarction. In other cases the relationship of medical or neurological illness to depression is less clear. However, the overall medical severity and disability appear to be important risk factors for depression. For these reasons, geriatric depression is viewed as a heterogeneous entity.

Depressed medical patients have more medical illnesses than nondepressed patients. A recent study of 15,186 primary care patients showed that the total mean number of medical diagnoses in depressed patients was 7.9 compared to 3 medical diagnoses in nondepressed patients. These differences persisted when the elderly group was examined separately. Depressed medical patients stay in bed more days compared to patients with chronic diseases, such as chronic lung disease, diabetes, arthritis, and hypertension.

In primary care patients, depression is associated with significant functional impairment. Depression impairs social function more than heart and lung disease, arthritis, hypertension, and diabetes, and leads to greater impairment in physical function than arthritis, hypertension, and diabetes. It has been shown that antidepressant treatment can improve mood, physical symptoms, and function in depressed patients with emphysema. Therefore, effective antidepressant treatment has the potential to enhance the quality of life of patients and their families and to lower the cost of care.

In addition to high medical morbidity, depression increases the perception of poor health and utilization of medical services. In primary care clinic populations, 75 percent of distressed overutilizers have clinically significant depressive symptomatology. Depressed patients had almost twice the number of appointments per year compared to nondepressed patients. Depressed patients had more than twice the number of hospital days over the expected length of stay compared to nondepressed patients. Finally, 65 percent of depressed patients received more than five medications compared to 35.6 percent of nondepressed patients. Data from a large health maintenance organization population suggest that the health care cost of depressed medical patients is twice that of nondepressed patients with similar levels of medical morbidity.

Depression is underdiagnosed in primary care settings. However, depression is more often recognized in older compared to younger adults. Even when diagnosed, depression is undertreated in primary care settings. Approximately 11 percent of depressed patients who use primary care services receive adequate antidepressant treatment, while 34 percent receive inadequate treatment, and 55 percent receive no treatment. While primary care physicians appear to identify depression more often in elderly than in young patients, there is little difference in the percentage of treated patients.

Depression in Patients With Dementia Dementia patients develop depression at a rate higher than that of the general population. Depressive manifestations of various intensity occur in approximately 50 percent of demented patients. Reports of the rate of major depressive disorder in patients with dementia of the Alzheimer's type have been highly discrepant, ranging from 0 to 87 percent, with most studies showing a range of 17 to 31 percent. Demented patients attending psychiatric centers as well as patients with mild to moderate dementia syndromes are more likely to report depression. Depressive symptomatology appears to be elicited more often when reports of relatives are used than when evaluation relies exclusively on the interviewer's impressions. There is some evidence that family history of mood disorder predispose these patients to develop depression.

Depression occurs in approximately 25 percent of patients with cerebrovascular disease. Cortical and lacunar infarct have the highest comorbidity with depression. Left hemisphere lesions, especially close to the frontal pole, are most frequently associated with poststroke depression while subcortical atrophy appears to be a predisposing factor. However, some epidemiological studies failed to confirm the high incidence of poststroke depression.

Patients with dementia due to Parkinson's disease appear to develop depression at a rate of up to 50 percent. The severity of depression does not appear to be related to the severity of motor disability.

It has been suggested that depression with onset in late-life includes a large subgroup of patients with neurological brain disorders that may or may not be evident when the depression first appears. This assertion is supported by differences observed between late- and early-onset geriatric depression. Compared to early-onset geriatric depression patients, patients with late-onset depression appear to have a less frequent family history of mood disorders, a higher prevalence of dementing disorders, more impairment in neuropsychological tests, a higher rate of dementia development on follow-up, more neurosensory hearing impairment, greater

enlargement in lateral brain ventricles, and more white matter hyperintensities. Further research may identify clinical and biological parameters that can characterize subgroups of late-onset depression and clarify their pathogenesis.

The similarity of depressive manifestations with symptoms and signs of dementing disorders often poses diagnostic problems. Vegetative features seem to overlap while patients with early-stage dementia of the Alzheimer's type show loss of interest, decreased energy, difficulty in concentration, agitation, or retardation. Apathy, a characteristic of frontal lobe syndrome, may be misidentified as retarded depression. Sad, downcast mood and psychic rather than vegetative features have been found useful in distinguishing depressed-demented patients from patients with dementia alone. It remains unclear whether depression is associated with the degree of cognitive impairment in patients with dementia. These patients may be made to identify and express dysphoric feelings. For this reason, examination should rely on caregiver reports as well as examination of the patient. Identification of depression in dementia patients, Parkinsonian patients, and patients with cerebrovascular disease is important because they may respond to drug therapy or ECT.

Depression With Reversible Dementia Some elderly depressed patients develop a dementia-like syndrome that improves or completely subsides after remission of depression. This syndrome has been termed *pseudodementia of depression*, or *depression with reversible dementia*, and attracted considerable clinical and heuristic interest. Depressed elderly patients who remain with some cognitive impairment even after improvement of depression usually have an early-state dementing disorder whose cognitive manifestations are exaggerated when the depressive syndrome is superimposed. Although some disagreement exists it appears that even patients with more or less complete cognitive recovery develop high rates of irreversible dementia (about 20 percent per year) on follow-up. Many of these patients may even remain cognitively unimpaired for 1 to 2 years before the development of irreversible dementia. The heterogeneity of the course of depression with reversible dementia suggests that patients suffering from this syndrome can be ordered along a continuum. At the one end of the continuum lie cases in whom the cognitive disturbance results predominantly from the depressive syndrome itself. On the other end are cases in whom the intellectual impairment originates from a progressive subclinical dementing disorder, with some contribution by the depressive syndrome. Therefore, identification of a “reversible dementia” syndrome in elderly depressives constitutes an indication for thorough diagnostic work-up and frequent follow-up aimed at the identification of treatable neurological disorders.

Psychotic Depression Psychotic depression occurs in 20 to 45 percent of hospitalized elderly depression patients and in 3.6 percent of elderly depression patients living in the community. As a rule, patients with psychotic depression have delusions whereas hallucinations are less frequent. The usual themes of depressive delusions are guilt, hypochondriasis, nihilism, persecution, and sometimes jealousy. Depressive delusions can be distinguished from delusions of demented patients in that the latter are less systematized and less congruent to the affective disturbance. Sometimes it is difficult to distinguish depressive delusions from overvalued ideas of worthlessness and hopelessness. Nondelusional depressed patients are usually able to recognize the exaggerated nature of overvalued ideas although they are unable to free themselves from their excessive concerns. Psychotic depression requires treatment with combinations of tricyclic drugs and psychiatric medications or ECT since it rarely responds to antidepressant drugs alone.

Geriatric Mania The risk of mania declines in late life. Nonetheless, mania and hypomania affect 5 to 10 percent of psychiatric patients. Late-onset mania often occurs in patients who have had recurrent depression since early life. Cerebrovascular disease and particularly right-sided vascular lesions may predispose to late-onset mania. Steroids may precipitate a late-onset manic syndrome. Unipolar mania is an early-onset disease.

Bereavement Loss of a spouse often is followed by depressive symptoms or syndromes. Significant distress remains during the first 2 years of widowhood. Older persons appear to be at lower risk for developing depressive symptoms or syndromes than younger adults during the first months of widowhood. However, by the end of the second year, older and younger individuals have the same rates of major depressive disorder. The prevalence of major depressive disorder continues to increase during the second year of bereavement. At the end of the second year after loss, 14 percent of bereaved individuals have major depressive disorder, a percentage much higher than the prevalence of major depressive disorder in the elderly community population (1 percent). Bereaved elderly individuals who do not meet criteria for major depressive disorder often have significant depressive symptomatology that can contribute to compromised function, disability, and impaired quality of life.

SUICIDE RISK

Elderly persons have a higher risk for suicide than any other population. The suicide rate for white men over the age of 65 is five times higher than that of the general population. One third of elderly persons report loneliness as the principal reason for considering suicide. Approximately 10 percent of elderly individuals with suicidal ideation report financial problems, poor medical health, or depression as reasons for suicidal thoughts. Suicide victims are demographically different than individuals who attempt suicide. About 60 percent of those who commit suicide are men; 75 percent of those who attempt suicide are women. Suicide victims as a rule use guns or hang themselves whereas 70 percent of suicide attempters take a drug overdose and 20 percent cut or slash themselves. Psychological autopsy studies suggest that most elderly persons who commit suicide have had a psychiatric disorder, most commonly depression. However, often psychiatric disorders of suicide victims do not receive medical or psychiatric attention. More elderly suicide victims are widowed and fewer are single, separated, or divorced than is true of younger adults. Violent methods of suicide are more common in the elderly and alcohol use and psychiatric histories appear to be less frequent. The most common precipitants of suicide in older individuals are physical illness and loss whereas problems with employment, finances, and family relationships are more frequent precipitants in younger adults. Most elderly persons who commit suicide communicate their suicidal thoughts to family or friends prior to the act of suicide.

Older patients with major medical illnesses or a recent loss should be evaluated for depressive symptomatology and suicidal ideation or plans. Thoughts and fantasies about the meaning of suicide and life after death may reveal information that the patient is unable to share directly. There should be no reluctance to question patients about suicide since there is no evidence that such questions can increase the likelihood of suicidal behavior.

COURSE AND PROGNOSIS

Knowledge of risk factors and periods of high risk for particular adverse outcomes enable the clinician to use appropriate diagnostic methods and offer preventive treatments with favorable benefit-to-risk ratios. Despite advances in antidepressant therapies, longitudinal studies of 1 to 6 years duration suggest that 7 to 30 percent of geriatric patients have a chronic major depressive disorder. If partially remitted subjects are considered chronic, the rate of chronicity reaches 40 percent. Chronicity of depression may be predicted by history of a long current episode or long previous episodes, coexisting medical illness, high severity of depression, nonmelancholic presentation, and delusions.

Naturalistic treatment studies of geriatric populations observed a 13 to 19 percent rate of relapse or recurrence at 1 year whereas a 15 percent relapse rate is observed in patients receiving controlled antidepressant treatment. This percentage is lower than the relapse or recurrence rate (34 percent) reported in younger adults. However, if the follow-up of elderly patients with depression is extended to 3 to 6 years, the recurrence rate increases to 38 percent. A 20-year follow-up study of mixed-age depressed patients showed that 95 percent had a recurrence. Relapse or recurrence appears to have different predictors than recovery of depression. The likelihood of relapse or recurrence may be high in patients with history of frequent episodes, late age of illness onset, history of dysthymia, intercurrent medical illness, and possibly high severity and chronicity of the index depressive episode.

Dementia is not an outcome of depression of intact elderly patients. However, a large proportion of geriatric depression patients have cognitive dysfunction or dementia. Approximately 40 percent of geriatric individuals with major depressive disorder consecutively hospitalized in a tertiary care academic psychiatric center also met criteria for dementia. The large percentage of elderly depressed patients with evidence of brain disease suggests that it is clinically meaningful to examine predictors of dementia in these populations. With some exceptions, recent follow-up data observed that the elderly depressed who initially had “reversible dementia” develop permanent dementia at a rate of 9 to 25 percent per year. Presence of an initially “reversible dementia” leads to 2-yearly rate of irreversible dementia that is 2.5 to 6 times higher than that of the general geriatric population. In nondemented depressives late age of depression onset and cortical brain atrophy were significant predictors of development of dementia over a 38-month average follow-up period.

TREATMENT

Depression Geriatric depression can be effectively treated with pharmacotherapy, psychotherapy, or both. Interpersonal therapy and cognitive-behavioral therapy have been found effective in depressed elderly patients. Psychotherapy may play an essential role in the care of bereaved depressed elderly patients or in patients who have significant life crises, lack social support, develop disabilities, or lack coping skills to deal with their life situations. Psychotherapy can be helpful in addressing the hopelessness often experienced by depressed elderly patients, especially those with a history of suicide attempts. Psychotherapy may be helpful in supporting elderly patients to cope with their symptoms and adhere to treatment before antidepressant drugs take effect. Psychosocial therapies should be considered in patients who cannot tolerate biological treatments.

The objectives of geriatric depression treatment include (1) remission of depression and (2) reduction in the risk of relapse and recurrence. Because depression often contributes to disability and excess medical morbidity, clinicians should expect an improvement in both these areas. Appropriate behavioral rehabilitation techniques should be combined with antidepressant treatment in order to enable elderly patients to regain function as the depressive syndrome subsides.

Drugs, medical illnesses, and dementing disorders may lead to depression. Steroids, reserpine (Serpasil), methyldopa (Aldomet), antiparkinsonian drugs, and b-adrenergic receptor antagonists can cause depression. Viral infections, endocrinopathies such as thyroid and parathyroid abnormalities, and malignancies such as lymphoma and pancreatic cancer often are complicated by depression. While it is essential to diagnose and treat the underlying disease, depression may not remit until an antidepressant agent is used. For example, depression in a hypothyroid elderly patient rarely responds to thyroid supplementation alone. Similarly, an antidepressant trial often is ineffective before hypothyroidism is corrected. However, thyroid supplementation and an antidepressant drug offer the highest probability of antidepressant response.

Elderly patients benefit from the same psychopharmacological agents as younger patients. However, the clinician must be aware that aging and medical conditions associated with aging have an impact on pharmacokinetics and increase the sensitivity to adverse effects even at low plasma concentrations of antidepressant drugs. Changes in hepatic metabolism prolong the clearance of most drugs in older people, increasing the likelihood that the parent drugs and their active metabolites will accumulate and cause toxicity. Older patients are more likely than younger patients to develop delirium, constipation, urinary retention, dry mouth, and orthostatic hypotension. For this reason, the dosages of antidepressant medication should be increased at a slower pace in older patients than in younger adults. Four families of antidepressant drugs are available for the treatment of geriatric depression. These are tricyclics, selective serotonin reuptake inhibitors (SSRIs), monoamine oxidase inhibitors (MAOIs), and atypical antidepressants. It is crucial that antidepressant drugs are given at adequate plasma concentrations or dosages for a sufficient length of time. An adequate antidepressant trial in the elderly is longer than that of younger adults and should last at least 7 to 9 weeks.

Tricyclic Antidepressants The secondary amines nortriptyline (Aventyl, Pamelor) and desipramine (Norpramin) are the most frequently used tricyclic antidepressant drugs in geriatric depression. They have lower anticholinergic and sedative effects than the tertiary amines amitriptyline (Elavil), doxepin (Adapin), and imipramine (Tofranil). Nortriptyline appears to have a lower potential for orthostatic hypotension than do other tricyclic antidepressant agents. The plasma concentrations of antidepressant medications required for the treatment of depressed elderly patients are similar to those needed by young adults (nortriptyline plasma concentrations of 60 to 150 ng/mL and desipramine above 115 ng/mL). Elderly patients often develop therapeutic blood concentrations of nortriptyline or desipramine while on low daily dosages. Most elderly patients develop therapeutic plasma concentrations with daily dosages of nortriptyline 1 to 1.2 mg/kg of body weight and desipramine 1.5 to 2 mg/kg. Elderly patients on nortriptyline often develop higher plasma levels of the nortriptyline metabolite 10-hydroxy-nortriptyline than do young adults, even when the plasma levels of the parent compound are similar. The reason for the increased levels of hydroxynortriptyline may in part be due to the lower renal clearance in elderly compared to younger subjects. High 10-hydroxynortriptyline levels may contribute to cardiac conduction defects in elderly patients.

High intellectual functions, orthostatic blood pressure, the electrocardiogram (ECG), and the ability to urinate should be monitored frequently in depressed elderly patients receiving nortriptyline or desipramine. Pretreatment systolic orthostatic hypotension has been found to correlate with antidepressant response to nortriptyline in some elderly patients. Therefore, nortriptyline rather than desipramine should be considered in such patients and orthostatic blood pressure and subjective symptoms of orthostasis (feeling lightheaded upon standing) should be monitored carefully.

Tricyclic medications have anticholinergic properties so these drugs should be avoided in patients with prostatic hypertrophy or patients with narrow-angle glaucoma. Nortriptyline and desipramine have properties similar to those of Type 1_A antiarrhythmic drugs (quinidine-like drugs). When administered to patients with right or left bundle branch block, tricyclic agents may cause second-degree block in approximately 10 percent of cases. For this reason, an ECG should always precede the use of tricyclic drugs in the elderly. The Type 1_A properties of tricyclic medications necessitate cautious use of these drugs in patients with ischemic heart disease. A multicenter study demonstrated that Type 1_A antiarrhythmics increase cardiac mortality in postmyocardial infarction patients. Although tricyclic medications were not used in this study, the Type 1_A antiarrhythmic properties of tricyclic drugs suggest cautious use in patients with depression and ischemic heart disease.

Selective Serotonin Reuptake Inhibitors Studies of elderly outpatients observed that SSRIs are equally effective with tertiary amine tricyclic agents in the short-term treatment of depressive disorder. Recently, sertraline (Zoloft) was shown to be as effective as nortriptyline and fluoxetine (Prozac). It is unclear whether SSRIs can be helpful in hospitalized elderly depression patients with severe, melancholic depression. A study that did not use random assignment suggested that cardiac inpatients with severe, melancholic depression have a poor response to fluoxetine while they have a robust response to nortriptyline at therapeutic plasma concentrations. Although research inpatient studies of SSRIs are lacking, clinical experience suggests that these agents are effective in inpatients with a broad spectrum of depressive syndromes.

The dosages of SSRIs should be increased gradually. The starting daily dosages may be fluoxetine 5 to 10 mg, paroxetine (Paxil) 5 to 10 mg, sertraline (Zoloft) 25 mg, and citalopram (Celexa) 10 mg. For most patients, daily dosages of fluoxetine 20 mg, paroxetine 20 mg, sertraline 75 mg, and citalopram 20 to 30 mg are sufficient, although higher dosages are required by some. The SSRI fluvoxamine (Luvox) may be effective in the treatment of geriatric depression although research data are still sparse. The SSRIs have fewer cardiac adverse effects than tricyclic antidepressant agents and often are used as drugs of first choice, especially in patients with mild, nonmelancholic depression or in patients with cardiac disease. The most frequent adverse effects of SSRIs are insomnia, akathisia, nausea, anorexia, pseudoparkinsonism, and inappropriate secretion of antidiuretic hormone leading to hyponatremia.

Drug interactions should be considered in elderly patients receiving SSRIs. Fluoxetine and paroxetine inhibit the cytochrome P450 (CYP) 2D6 liver cytochrome isoenzyme. Sertraline is a weaker inhibitor of the 2D6 isoenzyme and citalopram has an insignificant inhibiting effect. CYP2D6 isoenzyme is essential for the hydroxylation of nortriptyline and desipramine and the metabolism of antipsychotic agents and Type 1_A antiarrhythmic drugs (encainide, flecanide [Tambocor]), beta-blocking drugs, and verapamil (Calan) whose plasma concentrations may be raised in SSRI-treated patients. Therefore, reduction of dosages and monitoring of plasma levels of tricyclic antidepressant, antipsychotic, and antiarrhythmic drugs is required in patients treated with fluoxetine or paroxetine (Paxil). An exception is the SSRI fluvoxamine, which inhibits the CYP 3A4 and the IA2 cytochrome isoenzyme but does not significantly inhibit the 2D6 isoenzyme. CYP 3A4 is responsible for the metabolism of alprazolam (Xanax), triazolam (Halcion), carbamazepine (Tegretol), quinidine (Cardioquin), erythromycin, terfenazine, and astemizol (Hismanal) and may lead to an increase in the plasma concentration of these agents. These drugs should be avoided in patients receiving fluvoxamine. Similarly, theophylline (Aerolate) should be cautiously prescribed in fluvoxamine-treated patients because fluvoxamine may produce a threefold decrease in theophylline clearance by inhibiting the CYP 1D12 isoenzyme.

MAO Inhibitors Limited information exists on the use of these agents in the old literature. However, MAOIs have been studied in depressed patients who are in their late 60s to early 70s. MAOIs are effective in patients with major depressive disorder; they are also effective in depressed patients with panic attacks. Low dosages of MAOIs, for example, phenelzine (Nardil) 30 to 45 mg daily or tranylcypromine (Parnate) 20 to 30 mg daily, should be used in the elderly. Orthostatic hypotension is the most frequent adverse effect of MAOIs. This adverse effect is of concern in the elderly because it may lead to falls and fractures, especially of the hip or the humerus. Other adverse effects include weight gain, lack of energy, and insomnia; lack of energy and daytime somnolence in phenelzine-treated patients; and nervousness, insomnia, and excessive perspiration in tranylcypromine-treated patients. Peripheral neuropathy occurs in a small percentage of patients on MAOIs and often responds to pyridoxine. Sympathomimetic amines, monoamine precursors, tricyclic antidepressants, SSRIs, venlafaxine (Effexor), concomitant administration of two MAOIs and tyramine-rich food may cause a hypertensive crisis and should be avoided in patients on MAOIs. Drug interactions and dietary restrictions often prevent the use of MAOIs in the elderly.

Other Antidepressant Agents Venlafaxine was found effective in hospitalized depressed patients as well as in drug-resistant depressed patients and in depressed patients with chronic pain. For this reason, venlafaxine should be considered in severe depression and in depression unresponsive to other agents. Elderly patients appear to require dosages comparable to those of younger adults. Daily dosages above 100 to 200 mg are adequate for the majority of elderly patients. Nausea, a rather frequent adverse effect of venlafaxine, can be minimized by a slow increase of the daily dosage. Blood pressure should be monitored, especially in patients receiving dosages above 225 mg daily, since venlafaxine can increase blood pressure. Venlafaxine has limited drug-drug interactions and may be selected in elderly patients who require treatment with drugs interacting with other antidepressant drugs. Venlafaxine needs to be administered in three divided doses daily; for this reason, compliance should be monitored carefully.

Bupropion (Wellbutrin) has comparable efficacy with tricyclic antidepressant agents in younger adults. There is limited research information on the use of bupropion in the elderly. Bupropion has fewer adverse effects than tricyclic drugs, including lack of cognitive impairment or sedation, safety in overdose, and lack of cardiotoxicity. Bupropion was found to be safe in patients with heart disease, and it should be considered in elderly cardiac patients. Bupropion has few drug interactions, but it should not be prescribed in patients receiving MAOIs. Bupropion may exacerbate pre-existing hypertension. For this reason, monitoring of blood pressure is required. Seizures have been reported in 0.4 percent of patients treated with bupropion. The risk of seizures can be minimized by the use of sustained-release preparations,

slow introduction of bupropion (75 mg daily), use of divided daily dosages, and restriction of the total daily dosage to 450 mg. Most elderly patients require 75 to 100 mg of bupropion 3 times daily.

Nefazodone (Serzone) promotes sleep, has anxiolytic effects, is well tolerated, and safe in overdose. Nefazodone does not influence the sleep architecture and does not cause sexual dysfunction. However, sedation and need for two daily doses may be a problem for some elderly patients. A total daily dosage of 300 to 500 mg may be required.

Psychostimulants may be ineffective in elderly patients with primary major depression. However, dextroamphetamine (Adderall) and methylphenidate (Ritalin) appear to improve apathy and anergy in medical patients or cognitively impaired patients. Dextroamphetamine or methylphenidate at doses of 5 to 10 mg in the morning and at noon are sufficient. Psychostimulants have rapid onset of action, minimal adverse effects, limited potential for tolerance, and low risk for addiction.

Augmentation of antidepressant drugs may improve the response of partially remitted geriatric depression. Lithium (Eskalith) may augment tricyclic antidepressant response in elderly patients. The dose of lithium required by depressed elderly patients receiving tricyclics may be one third to one half that of younger adults. In younger patients, combinations of tricyclics with SSRIs have led to an antidepressant response sooner than tricyclics alone. Other augmentation techniques include combinations of tricyclics or SSRIs with thyroid hormones, psychostimulants, bupropion, pindolol (Pindolol), and other agents. Clinical experience suggests that augmentation techniques can be effective in some depressed geriatric patients with incomplete response to a single antidepressant agent. However, the role of augmentation techniques is limited in geriatric depression because elderly patients often are unable to tolerate combination therapies.

Continuation and Maintenance Treatment Following remission of depression, the antidepressant medication should continue to be administered for at least 6 months (continuation phase). Continuation treatment is used to prevent relapse of the episode that the patient most recently suffered. After having been well for 6 months, patients are at risk for a new episode of depression (recurrence). History of three or more episodes is the strongest predictor of recurrence. Other predictors are high severity of the initial episode and persisting anxiety. In mixed-age populations, late age of illness onset is a strong predictor of recurrence. Patient at high risk for recurrence should have maintenance treatment for at least 1 to 2 years. Unlike continuation treatment, which is used to prevent relapse of the initial depressive episode, maintenance treatment seeks to prevent new episodes of depression. The dosage and plasma concentration used for continuation and maintenance treatment should be the same as those used for effective acute treatment of depression.

Compliance Approximately 70 percent of patients fail to take 25 to 50 percent of their medications. Older patients and their families may not understand depression and its course and the importance of taking medications as prescribed. Concurrent medical illnesses can interfere with antidepressant response or attainment of adequate dosages. Alcoholism and other substance abuse may undercut pharmacotherapy. Difficulties accessing health care may hinder the ability of the elderly, especially the functionally dependent elderly, to obtain adequate treatment.

Bereavement The treatment of a bereaved elderly person depends on the severity of symptomatology. If major depressive disorder develops during the bereavement period, antidepressant treatment should be considered. Many clinicians suggest that psychotherapy, antidepressant drugs, or both should be administered if the depression lasts longer than 1 year. The timing of intervention and the decision about drug therapy depend on the severity of symptomatology.

Bereaved persons with depressive symptomatology that does not meet the criteria for major depressive disorder may benefit from brief focused psychotherapy. Interpersonal psychotherapy lends itself well for this purpose because it addresses issues related to loss, depression, and change of roles. Although antidepressant drugs can conceivably be effective in the bereaved elderly who have depressive symptoms, clear guidelines are not available because of a lack of research.

Most elderly bereaved individuals do not develop depressive syndromes or clinically significant depressive symptoms. Therefore, prophylactic treatment is not recommended for the whole elderly bereaved population. However, several interventions may be helpful and they should be considered for individuals with an intense reaction to loss or history of depression. These include self-help groups, counseling, group therapy, and individual dynamic psychotherapy.

Mania The management of late-life mania follows the same principles as that of younger adults. However, geriatric manic patients should be examined for drugs and medical or neurological diseases that can predispose or precipitate mania.

Lithium Lithium is the most investigated drug in geriatric mania. Lithium may be less effective when mania is comorbid with neurological and medical disorders than in uncomplicated mania. Age-associated decline in renal clearance in elderly patients leads to high lithium plasma levels even when low dosages are used. Lithium should be introduced slowly, starting with daily dosages of 150 mg. About one half of the dosage required for young adults is sufficient for the elderly. The half-life of lithium is about 24 to 36 hours in patients who are in their 70s. Therefore, steady-state pharmacokinetics are anticipated 5 or more days after the stabilization of the daily dosage. In some elderly patients, lithium plasma levels of 0.3 to 0.6 mEq/L were found to be effective. However, controlled studies are needed to establish the efficacy of low lithium plasma levels in the elderly. The onset of action of lithium is slow and may require several days or weeks.

The adverse effects of lithium are pronounced in geriatric patients. Elderly patients often develop a fine tremor or myoclonus at lithium plasma levels considered nontoxic in young adults. Lithium can produce Parkinsonian symptoms or worsen pre-existing Parkinson's disease. Delirium can develop at plasma concentrations below the therapeutic range. Psychiatric drugs increase the risk of delirium in elderly patients on lithium. Lithium may worsen cognitive function, especially in patients with dementia. Sinoatrial block may be caused by lithium in predisposed patients; digitalis and beta-blocking agents increase the risk for sinoatrial block. Salt depletion, thiazide diuretics, and nonsteroidal anti-inflammatory drugs may raise the lithium plasma concentrations and result in toxicity.

Anticonvulsants The anticonvulsant valproate (Depakene), divalproex (Depakote), and carbamazepine are effective antimanic agents. Anticonvulsants drugs appear to be effective in both lithium responders and lithium nonresponders. Anticonvulsants may be more effective than lithium in rapidly cycling manic patients and in those with dysphoric mania. Patients of mania with neurological brain disease may be more responsive to valproate than patients without brain disease. Although there is limited research on the use of these drugs in elderly populations, they should be considered especially in patients who are at high risk for lithium toxicity or who have a history of poor response to lithium.

Drowsiness and sedation are frequent adverse effects of valproate and carbamazepine; both are dose-dependent adverse effects. Many patients are habituated to sedation. Therefore, slow introduction of valproate or carbamazepine may reduce this problem. In the elderly, the principal adverse effects of valproate are gastrointestinal distress and ataxia; frequent adverse effects of carbamazepine are ataxia and confusion. Rarely, carbamazepine can cause hyponatremia. Approximately 0.4 of patients receiving valproate and 2 percent of patients receiving carbamazepine develop leukopenia with a white count below 4000/mm ^{3}. Half of the patients who develop leukopenia on carbamazepine have a reduction in their white blood cell count during the first 2 weeks. For this reason, periodic monitoring of the white blood cell count is indicated, especially during the early phases of treatment.

Other Drugs Agitated elderly patients with mania may be treated with low dosages of lorazepam (Ativan). Lorazepam is favored over other benzodiazepines because it undergoes only phase II metabolism (glucuronidation), a process minimally affected by aging. Alternatives to lorazepam are high-potency antipsychotics at low dosages, for example, haloperidol (Haldol) 0.5 to 5 mg daily. These drugs should be used early in the treatment of acute mania and should be withdrawn when the agitation is controlled and therapeutic plasma levels of a mood stabilizer are established.

Electroconvulsive Therapy Electroconvulsive therapy (ECT) should be considered in both geriatric depression and mania because of its efficacy, rapid onset of action, and safety. ECT may be chosen in patients with severe mood syndromes unable to tolerate the long waiting imposed by the gradual introduction of antidepressant drugs or mood stabilizers and the slow onset of drug action in the elderly. Approximately 80 percent of the population in the seventh or eighth decade of life have one or more medical illnesses. Medical illness and depression tend to occur in the same persons. Therefore, the percentage of medical illness in elderly patients with depression is even higher than that of the general population. Cardiac conduction defects, prostatic hypertrophy, glaucoma, and other conditions increase the risk of drug therapy. Moreover, even when antidepressant drugs are tolerated the need for gradual introduction of antidepressants and the slow onset of antidepressant action in the elderly prolong the time during which patients remain depressed, thereby increasing the risk for suicide, debilitation, dehydration, and electrolyte disturbances.

Controlled studies demonstrated that ECT is an established acute treatment for depression and mania. Administration of three ECTs a week appears to produce a more rapid recovery than ECT given twice or once a week. Less frequent ECT is associated with less cognitive impairment, but no differences were found 1 month after ECT completion between the groups that had ECT three times or twice weekly.

It appears that electrode placement and electrical dose influence both efficacy and memory adverse effects. Low–electrical-dose (just above seizure threshold)

unilateral ECT was found less effective than high-dose (2.5 times the threshold) unilateral ECT. However, even high-dose unilateral ECT may not be as effective as bilateral ECT, although disagreement exists. Use of high- or low-dose electrical stimulus does not appear to influence the efficacy of bilateral ECT. Regardless of electrode placement, high dosage increases the speed of improvement. These recent data challenge earlier findings suggesting that long duration of seizures is associated with high treatment efficacy. High-dose unilateral ECT is associated with prolonged time to recover orientation after seizure induction. Bilateral ECT requires longer recovery time than high-dose unilateral ECT. Compared to unilateral ECT, bilateral ECT leads to greater retrograde amnesia within a short period after completion of ECT, but there may not be significant memory differences after 2 months. Most of these observations used mixed-age depressed populations.

Anecdotal literature suggests that ECT is an effective continuation or maintenance antidepressant treatment. Controlled studies of continuation and maintenance ECT are needed since there is evidence of high relapse-recurrence rate in ECT responders who previously had a medication-resistant depressive disorder.

The mortality rate following ECT is 0.01 percent. The majority of deaths (67 percent) related to ECT are due to cardiac complications and occur immediately after ECT or within a few hours from treatment. Compared to younger adults, elderly patients and patients with pre-existing cardiac disease are more prone to cardiovascular events occurring in the context of ECT, including ischemic syndromes, arrhythmias, decompensation of heart failure, and transient severe rise in blood pressure. However, with adequate medical evaluation, monitoring during and after ECT, and appropriate intervention most cardiovascular events related to ECT have a benign outcome.

ECT alters cerebrovascular circulation temporarily. ECT leads to cerebrovascular contraction, heightened blood-brain barrier permeability, and increased brain oxygen consumption. Nonetheless, ECT appears to be reasonably safe 1 month after a cerebrovascular accident. ECT is generally contraindicated in patients with intracranial tumors because it may lead to delirium, major neurological events, or death. However, case reports have documented that patients with small asymptomatic tumors without surrounding edema or obstruction of CSF flow may tolerate ECT.

Geriatric patients require a longer time to recover their memory than younger adults, especially after bilateral ECT. Moreover, geriatric patients sometimes develop prolonged confusion after ECT; falls have been reported in up to 15 percent of patients. However, controlled studies are needed in order to examine whether and to what extent ECT-related confusion is responsible for these falls. The practice of favoring ECT over antidepressant drugs in medically ill depressed patients may have led to inflated reports of ECT-related morbidity. Anecdotal literature suggests that ECT was used uneventfully in patients over 100 years old and in elderly patients with aortic aneurysm cardiac pacemakers, or history of myocardial infarction, stroke, transient ischemic attacks, severe hypertension, and arrhythmias or patients in need of anticoagulant drug therapy.

SUGGESTED CROSS-REFERENCES

Delirium and dementia are discussed in Chapter 10, delusions in Section 13.2, and mood disorders in the general population in Chapter 14. Suicide is discussed in Section 29.1 and Alzheimer's disease in Section 51.3e. Treatment of the elderly is discussed in depth in Section 51.4.

SECTION REFERENCES

Alexopoulos GS: Geriatric depression in primary care. Int J Geriatr Psychiatry *11:*397, 1996.

Alexopoulos GS: Methodology of treatment studies in geriatric depression. Am J Geriatric Psychiatry *3:*280, 1995.

*Alexopoulos GS: Clinical and biological findings in late-onset depression. In *American Psychiatric Press Review of Psychiatry*, vol 9, A Tasman, SM Goldfinger, CA Kaufmann, editors. Washington, DC, 1990.

Alexopoulos GS, Meyers BS, Young RC, Kakuma T, Feder M, Einhorn A, Rosendahl E: Recovery in geriatric depression. Arch Gen Psychiatry *53:*305, 1996.

Alexopoulos GS, Meyers BS, Young RC, Kakuma T, Silbersweig D, Charlson M: Clinically defined vascular depression. Am J Psychiatry *154:*562, 1997.

*Alexopoulos GS, Meyers BS, Young RC, Mattis S, Kakuma T: The course of geriatric depression with “reversible dementia”: A controlled study. Am J Psychiatry *150:*1693, 1993.

Alexopoulos GS, Vrontou C, Kakuma T, Meyers BS, Young RC, Klausner E, Clarkin J: Disability in geriatric depression. Am J Psychiatry *153:*877, 1996.

Alexopoulos GS, Young RC, Meyers BS: Geriatric depression: Age of onset and dementia. Biol Psychiatry *34:*164, 1993.

Alexopoulos GS, Young RC, Shindledecker RD: Brain computed tomography findings in geriatric depression and primary degenerative dementia. Biol Psychiatry *31:*591, 1992.

The Cardiac Arrhythmia Suppression Trial (CAST) Investigators: Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. N Engl J Med *327:*227, 1992.

Gurland BJ, Cross PS, Phil M, Katz S: Epidemiological perspectives on opportunities for treatment of depression. Am J Geriatr Psychiatry *4:*S7, 1996.

Katz IR: Diagnosis and treatment of depression in patients with Alzheimer's disease and other dementias. J Clin Psychiatry *59*(Suppl):38, 1998.

Katz IR, DiFilippo S: Neuropsychiatric aspects of failure to thrive in late life. Clin J Geriatr Med *13:*623, 1997.

Koenig HG, Blazer DG III: Minor depression in late life. Am J Geriatr Psychiatry *4:*S14, 1996.

Krishnan KRR, Gadde KM: The pathophysiologic basic for late-life depression: Imaging studies of the aging brain. Am J Geriatr Psychiatry *4:*S22, 1996.

Lebowitz BD, Pearson JL, Schneider LS, Reynolds CF III, Alexopoulos GS, Bruce ML, Conwell Y, Katz IR, Meyers BS, Morrison MF, Mossey J, Niederehe G, Parmelee P: Diagnosis and treatment of depression in late life: Consensus statement update. JAMA *278:*1186, 1997.

Little JT, Reynolds CF III, Dew MA, Frank E, Begley AE, Miller MD, Cornes C, Mazumdar S, Perel JM, Kupfer DJ: How common is resistance to treatment in recurrent, nonpsychotic geriatric depression? Am J Psychiatry *155:*1035, 1998.

Meyers BS: Psychiatric interventions to improve primary care diagnosis and treatment of depression. Am J Geriatr Psychiatry *4:*S91, 1996.

Miller MD, Curtiss EI, Marino L, Houck PR, Paradis CF, Mazumdar S, Pollock BG, Foglia J, Reynolds CF III: Long-term ECG changes in depressed elderly patients treated with nortriptyline. A double-blind, randomized, placebo-controlled evaluation. Am J Psychiatry *6:*59, 1998.

Mirchandani IC, Abrams RC, Young RC, Alexopoulos GS: One year follow-up of continuation convulsive therapy prescribed for depressed elderly patients. Int J Geriatric Psychiatry *9:*31, 1994.

Mukhergee S, Sackheim HA, Schnur DB: Electroconvulsive therapy of acute manic episodes: A review of 50 years' experience. Am J Psychiatry *151:*169, 1994.

Mulsant BH, Sweet R, Rifai AH, Pasternak RE, McEachran A, Zubenko GS: The use of the Hamilton rating scale for depression in elderly patients with cognitive impairment and physical illness. Am J Geriatr Psychiatry *2:*220, 1994.

Niederehe GT: Psychosocial treatments with depressed older adults: A research update. Am J Geriatr Psychiatry *4:*S66, 1996.

*Nyenhuis DL, Yamamoto C, Luchetta T, Terrien A, Parmentier A: Adult and geriatric normative data and validation of the profile of mood states. J Clin Psychol *55:*79, 1999.

Pollock BG, Mulsant BH, Nebes R, Kirshner MA, Begley AE, Mazumdar S, Reynolds CF III: Serum anticholinergicity in elderly depressed patients treated with paroxetine or nortriptyline. Am J Psychiatry *155:*1110, 1998.

Rabins PV: Barriers to diagnosis and treatment of depression in elderly patients. Am J Geriatr Psychiatry *4:*S79, 1996.

Robinson RG, Starkstein SE: Current research in affective disorders following stroke. J Neuropsychol Clin Neurosci *2:*1, 1990.

Roose SP, Glassman AH, Giardina EGV: Tricyclic antidepressants in depressed patients with cardiac conduction disease. Arch Gen Psychiatry *44:*273–275, 1987.

Schneider L, editor: *Diagnosis and Treatment of Depression in Late Life: Results of the NIH Consensus Development Conference*. American Psychiatric Press, Washington, DC, 1994.

Schneider LS: Pharmacologic considerations in the treatment of late-life depression. Am J Geriatr Psychiatry *4:*S51, 1996.

*Steffens DC, Hays JC, Krishnan KR: Disability in geriatric depression. Am J Geriatr Psychiatry *7:*34, 1999.

Stoll AL, Banov M, Kolbrener M, Mayer PV: Neurologic factors predict a favorable valproate response in bipolar and schizoaffective disorders. J Clin Psychopharmacol *14:*311, 1994.

Wells KB, Stewart A, Hays RD: The functioning and well-being of depressed patients. Results from the Medical Outcomes Study. JAMA *262:*914, 1989.

Wiener P, Alexopoulos GS, Kakuma T, Meyers BS, Rosendhal E, Chester J: The limits of history taking in geriatric depression. Am J Geriatr Psychiatry *5:*116, 1997.

Williams-Russo P: Barriers to diagnosis and treatment of depression in primary care settings. Am J Geriatr Psychiatry *4:*S84, 1996.

Young RC, Alexopoulos GS, Kent E, Shamoian CA, Dhar AK, Kutt H: Plasma 10-hydroxynortriptyline and ECG changes in elderly depressed patients. Am J Psychiatry *142:*866, 1985.

*Young RC, Klerman GL: Mania in late life: Focus on age at onset. Am J Psychiatry *4:*73, 1992.

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3E ALZHEIMER'S DISEASE AND OTHER DEMENTIAS

GARY W. SMALL, M.D.

Alzheimer's disease (called *dementia of the Alzheimer's type* in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV]) has received increased attention in recent years, along with the rapidly growing number of elderly persons most vulnerable to it and related dementias. Except for genetic mutations causing the rare early-onset familial cases, the cause or causes of this form of dementia are unknown. Basic and clinical research efforts, however, offer the hope of discovering the cause, which should eventually lead to specific treatments or cures. Until then, accurate diagnosis, symptomatic treatments, and family interventions are essential to minimize the suffering of patients and family members and to maximize the patient's functional level.

DEFINITION

According to DSM-IV, the essential features of dementia include memory impairment, impairment in at least one other cognitive domain (e.g., language and visual-spatial skills), and significant disturbance of work or social functioning or both resulting from cognitive deficits. These features cannot occur exclusively during the course of a delirium. Dementia of the Alzheimer's type is defined as a dementia syndrome that is gradual in onset and progression and without another identifiable and sometimes treatable cause. A definite diagnosis of dementia of the Alzheimer's type can be made only by histopathological examination of brain tissue, generally after the patient's death. A work group on the diagnosis of this disorder established by the National Institute of Neurological and Communicative Disorders and the Alzheimer's Association specified criteria that provide a guide for the diagnosis of possible and probable Alzheimer's disease. Possible Alzheimer's disease includes dementia syndromes in which an additional disease (e.g., tumor or cerebral thrombosis) may be implicated. Probable Alzheimer's disease is similar to progressive dementia of the Alzheimer's type.

HISTORY

In 1906 the German psychiatrist and neurologist Alois Alzheimer first described a middle-aged patient who had suffered from a progressive dementia that affected language, memory, and behavior. After the patient's death at age 55 years, Alzheimer applied new staining techniques to the patient's brain tissue and demonstrated the presence of what is now termed *neurofibrillary tangles* and neuritic plaques in the neocortex and other brain regions. For years, Alzheimer's disease was considered a presenile dementia, partially because some plaques and tangles occurred in elderly persons without dementia and some elderly persons with dementia had few plaques and tangles. Such conflicts were resolved, however, in the late 1960s, when the degree of dementia was shown to correlate with the number of neuritic plaques in neocortical association areas. Moreover, causes of senile dementia other than Alzheimer's disease were also recognized.

EPIDEMIOLOGY

Dementias are the most common causes of psychopathology in elderly persons. Dementia of the Alzheimer's type accounts for about 60 percent of old-age dementias. It is estimated to afflict from 5 to 10 percent of people age 65 and accounts for the most striking rise in dementia incidence in the very old. Dementia of the Alzheimer's type and related dementias cost society each year an estimated $100 billion, which includes both direct costs (i.e., actual dollar expenditures) and indirect costs (i.e., resource losses not involving dollar expenditures). The greatest risk factors for developing dementia is age—the incidence and prevalence of the disease double every 5 years after age 60. An estimated 4 million U.S. citizens suffer from Alzheimer's disease. As the baby-boom generation (born in the 1950s and 1960s) moves into the age group of 60 years and older, Alzheimer's disease will become an even greater public health problem. By the year 2040, 14 million Americans could be affected if current trends remain unabated.

ETIOLOGY

Investigators around the world are searching for possible causes for Alzheimer's disease, which should lead to treatments that may delay or even halt the dementing process. Efforts have focused on various neurotransmitter systems, especially the cholinergic system, because of such findings as loss of cholinergic neurons in the basal nucleus of the forebrain. Some of the latest discoveries have been in the area of molecular biology and genetics. The increased frequency of Down syndrome (trisomy 21) patients who invariably develop Alzheimer's disease neuropathology and dementia by their 40s instigated an initial focus on chromosome 21, and significant linkage to chromosome 21 was discovered in some early-onset families but not in others. Excitement in the area built, however, when the gene coding for the amyloid precursor protein (APP) found in senile plaques was localized to the same chromosome 21 region. APP mutations, however, have been identified in only 20 families. More recent studies have shown that most early-onset pedigrees not segregating APP mutations show mutations of a chromosome 14 gene *(presenilin 1)*. One other form of early-onset Alzheimer's disease is represented by families of Volga German origin. For these families chromosome 1 mutations *(presenilin 2)* have been identified.

Studies of the common late-onset disease (dementia beginning after age 60 years) showed evidence for linkage or association or both for a chromosome 19 region. Apolipoprotein E (APOE) was a candidate Alzheimer's disease susceptibility gene partly because APOE and cerebrospinal fluid bound to immobilized amyloid b peptide and antisera to APOE stain senile plaques and neurofibrillary tangles. In addition, APOE localized to the same region of chromosome 19 identified in linkage studies.

APOE has three allelic variants (2, 3, and 4). Everyone inherits one allele from each parent so that five common genotypes are possible (2/3, 3/3, 2/4, 3/4, and 4/4). The sixth genotype (2/2) is rare. In the general population, approximately 3 percent of persons have the 4/4 genotype, 20 percent have the 3/4 genotype, and most people have the 3/3 genotype. The APOE-4 allele increases risk and decreases age at dementia onset in a dose-related fashion, that is, Alzheimer's disease risk is lowest for the 3/3 genotype, higher for the 3/4 genotype, and highest for the 4/4 genotype. By contrast, the APOE-2 allele has a protective effect so that Alzheimer's disease risk is lower for people with the 2/3 genotype than with the 3/3 genotype. Susceptibility to Alzheimer's disease from APOE affects many races and has been confirmed worldwide.

One strategy used to facilitate searches for causes recognizes that Alzheimer's disease may be a heterogeneous condition. Investigations comparing familial (one or more affected first-degree relatives of proband or large autosomal dominant kindreds) and sporadic Alzheimer's disease patients, as well patients with early-onset (65 years or younger) and late-onset dementia, have found clinical and biological differences among them. Other studies suggest sex differences according to disease expression, cerebral glucose metabolism, and neuroendocrine function and a possible interaction among age at onset, sex, and familiarity. Case-control studies have explored potential risk factors for Alzheimer's disease. Although the studies have demonstrated such risks as prior head trauma and depression, additional study is recommended. Factors that may reduce the risk for developing Alzheimer's disease are also being studied, including higher educational level, larger brain size, and postmenopausal estrogen therapy.

In a sense, research in Alzheimer's disease is like solving a huge jigsaw puzzle: although researchers are beginning to piece together a few corners, the puzzle is far

from solved. In some cases, the cause may be a genetic defect on chromosome 14; in others a mixture of genetic and environmental factors may play a role.

DIAGNOSIS AND CLINICAL FEATURES

The hallmark of dementia is memory impairment. In progressive dementias, short-term memory is affected initially. Patients have particular difficulty in learning new information, as shown by their inability to recall three objects after 5 minutes or even after just two distractions. Long-term memory may be affected, and patients may not recall important past personal information. Other cognitive areas may be impaired, including abstract thinking, judgment, personality, and language. In a typical case of Alzheimer's disease, the onset is so insidious that family members have difficulty estimating when the impairment began.

A complete history from someone who knows the patient well, physical and neurological examinations, and a mental status examination are essential in the diagnostic evaluation. Brief standardized mental status tests (e.g., the Mini-Mental State Examination) are useful to quantify the degree of cognitive impairment, and more extensive neuropsychological batteries provide details on the nature of the cognitive deficits. On examination, the patient may show *apraxia* (inability to carry out motor commands, even though comprehension and motor function are intact), *agnosia* (inability to recognize objects, despite intact sensory function), or difficulties in visual-spatial skills (inability to copy two-dimensional and three-dimensional figures or to assemble blocks). A variety of behavioral changes may accompany the cognitive deficits, including paranoia, agitation, insomnia, anxiety, and depression.

PATHOLOGY AND LABORATORY EXAMINATION

The key histopathological features include neuritic plaques (extracellular deposits of amyloidogenic proteins) and neurofibrillary tangles (abnormal intracellular cytoskeletal filaments). Plaque density has been used as a criterion for the postmortem diagnosis of Alzheimer's disease. Reduction in neuron number and synapses and an accumulation of plaques and tangles occur in the frontal, temporal, and parietal lobes; that finding is consistent with the involvement of association and limbic brain structures. Because of neuronal loss, the brains of Alzheimer's disease patients demonstrate atrophy greater than expected for their ages.

Many clinicians recommend routine use of structural imaging scans (computerized tomography [CT] or magnetic resonance imaging [MRI]) as part of the laboratory evaluation for dementia. Positron emission tomographic (PET) scans can identify patterns of glucose use (the brain's sole source of nutrition in nonstarvation states) that differentiate Alzheimer's type dementia from other dementias (Figs. 51.3e-1 and 51.3e-2), even before the diagnosis can be confirmed clinically. Such parietal and temporal hypometabolism findings observed on a PET scan may be asymmetric and correlate with hemisphere-specific neuropsychological deficits. Moreover, combining genetic risk assessment with PET is a potential strategy for early detection of Alzheimer's disease. Single photon emission computed tomographic (SPECT) scan measures cerebral blood flow and demonstrates biparietal hypoperfusion in Alzheimer's disease, but it has poorer spatial resolution compared with PET.

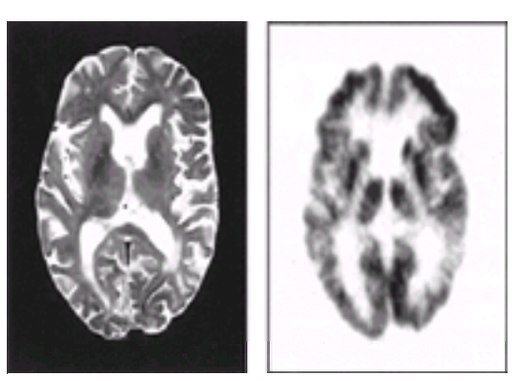

FIGURE 51.3e-1 Brain scans of an 82-year-old woman with a 4-year history of gradual memory loss and a clinical diagnosis of dementia of the Alzheimer type. The MRI (left) shows moderate atrophy with enlarged ventricles and periventricular white matter hyperintensities. The PET (right) shows atrophy but also the temporal hypometabolism characteristic of Alzheimer's disease. Higher planes also showed parietal hypometabolism. The white matter hyperintensities and atrophy are nonspecific findings in this case; by contrast, the parietal and temporal hypometabolism help with the differential diagnosis of dementia.

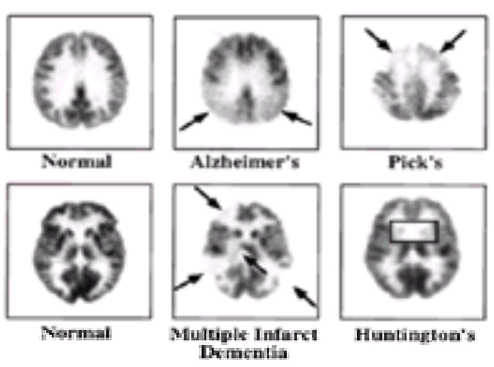

FIGURE 51.3e-2 Characteristic PET patterns (transaxial images) for a cognitively normal person and patients with different dementias. The following metabolic deficits are demonstrated (*arrows*): Alzheimer's disease—bilateral parietal deficits; multiple infarct (vascular) dementia—multiple focal deficits; Pick's disease—frontal deficits with retained parietal metabolism; Huntington's disease—bilateral caudate deficits.

The laboratory evaluation for dementia involves a search for potentially treatable medical conditions. The following blood tests have been recommended as routine: complete blood count; serum urea nitrogen and glucose; serum electrolytes (sodium, potassium, carbon dioxide content, chloride, calcium, phosphorus); serum vitamin B_{12}; thyroid function tests; serological test for syphilis; and liver function tests. The combination of increased tau and decreased beta-amyloid has been found in the cerebrospinal fluid of patients with Alzheimer's disease, but the sensitivity and specificity of this profile has not been studied in large clinical populations.

DIFFERENTIAL DIAGNOSIS

The differential diagnosis of Alzheimer's disease can be a problem. Most potential biological markers are still in a development phase, and the clinician is left with conventional methods of evaluation: the patient's history, a physical examination, and laboratory tests. The history provides important information, and several informants should be sought for corroboration. Even with the most thorough evaluation, diagnostic errors do occur. Clinical researchers who have neuropathological confirmation of diagnosis report accuracy rates ranging from 65 percent to more than 90 percent, even though other progressive dementias (such as Pick's disease) can rarely be distinguished on clinical grounds alone.

Vascular dementia is estimated to account for 10 to 20 percent of the dementias of old age. Classically, the disorder is characterized by a sudden onset of illness and stepwise decline in cognitive function, rather than the insidious onset and gradually progressive course of Alzheimer's disease. Focal neurological signs and association with hypertension in many patients may also help differentiate between these two dementias. The clinical signs and symptoms, coded and given weighted scores, have been used widely to help differentiate multi-infarct dementia from dementia of the Alzheimer's type. Although the validity of the scale has been questioned, searching for these features can help clarify the patient's clinical profile (Table 51.3e-1). An expanded scale has been developed aimed at correcting the methodological shortcomings of the early instrument. Nonetheless, the frequent coexistence of these two dementias (estimated at about one third of all dementias) impairs diagnostic accuracy.

Feature	Point Values*
Abrupt onset	2
Stepwise deterioration	1
Somatic complaints	1
Emotional incontinence	1
History or presence of hypertension	1
History of strokes	2
Focal neurological symptoms	2
Focal neurological signs	2

* A score of 4 or more is consistent with vascular dementia.
Adapted from Rosen WG, Terry RD, Fuld PA, Katzman R, Peck A: Pathological verification of ischemic score in differentiation of dementias. Ann Neurol 7: 486, 1980.

Table 51.3e-1 Clinical Features of the Modified Ischemic Score

Even though Alzheimer's disease and vascular dementia constitute the bulk of all dementias, numerous other conditions must be considered and sometimes can be differentiated from Alzheimer's disease by their clinical presentation. For example, frontotemporal dementia is characterized by marked personality changes, relative preservation of visuospatial skills, and executive dysfunction. By contrast, patients with Lewy body dementia may evidence visual hallucinations, delusions, fluctuating mental status, and sensitivity to antipsychotic drugs.

Alcohol and drugs may cause memory loss and other symptoms of dementia. For elderly patients, drugs deserve special emphasis as a cause of memory impairment because of widespread polypharmacy and age-related changes in body function and pharmacokinetics. An elderly person's brain can be extremely sensitive to change in its environment, and almost any physical illness can effect such a change and produce a dementia-like syndrome. Numerous potentially treatable physical and mental conditions must be considered whenever the diagnosis of dementia is entertained (Table 51.3e-2).

Substances
Anticholinergic agents
Antihypertensives
Antipsychotics
Corticosteroids
Digitalis
Narcotics
Nonsteroidal anti-inflammatory agents
Phenytoin
Polypharmacotherapy
Sedative-hypnotics
Psychiatric Disorders
Anxiety
Depression
Mania
Delusional (paranoid) disorders
Metabolic and Endocrine Disorders
Addison's disease
Cushing's syndrome
Hepatic failure
Hypercarbia (chronic obstructive pulmonary disease)
Hypernatremia
Hyperparathyroidism
Hyperthyroidism
Hypoglycemia
Hyponatremia
Hypothyroidism
Renal failure
Volume depletion
Miscellaneous Conditions
Fecal impaction
Hospitalization
Impaired hearing or vision

Table 51.3e-2 Some Potentially Reversible Conditions That May Resemble Dementia

The clinical presentation of geriatric depression may mimic dementia. The earlier term " *depressive pseudodementia*" is no longer considered useful because the dementia is real; instead, the term *dementia syndrome of depression* often is used. The reported frequency of mistaking depressed for demented patients ranges from 5 to 15 percent. Psychomotor retardation, impaired concentration, and apathy characteristic of the depressed patient may be mistaken for symptoms of dementia, especially when these symptoms are accompanied by memory loss. In the dementia syndrome of depression, however, the complaints of memory tend to exceed the actual memory impairment observed on psychological testing. The fact that dementia and depression often coexist further clouds the diagnostic picture. About 30 percent of patients with a progressive dementia also suffer from some form of depression.

Many people in their 50s and older complain of forgetfulness and worry about its meaning. *Age-associated memory impairment*, as defined by a work group convened by the National Institute of Mental Health, applies to healthy persons 50 years and older who have subjective complaints of memory loss and mild deficiencies in memory performance compared to young adults. About 40 percent of persons in their 50s and up to 85 percent of those age 80 and older suffer from age-associated memory impairment. Preliminary data suggest that the condition is stable over intervals of up to 4 years, and thus is presumed to reflect normal aging and does not usually progress to dementia. Although it may take them longer to remember, most elderly people continue to recall information they have learned well and important events from the past.

COURSE AND PROGNOSIS

Dementia of the Alzheimer's type is a progressive disease that eventually leads to death, not directly from the dementing process but usually from intercurrent illnesses. The rate of progression varies, and it is difficult to estimate how rapid a decline a patient may have. Some patients become extremely demented within a year; others experience plateaus for several years.

TREATMENT

The first step in treating dementia is to attempt to identify any underlying illnesses that may be causing memory impairment and other symptoms of the syndrome. If none is found or if the symptoms persist after appropriate specific treatment, then several symptomatic treatments are available.

Psychopharmacotherapy A variety of pharmacological treatments is currently available for Alzheimer's disease (Table 51.3e-3 and Table 51.3e-4). Many of these approaches will diminish symptoms and increase quality of life, but none of them can halt the dementing process. Most pharmacological agents currently available or in development target a specific symptom (e.g., agitation or memory loss) and are derived from the known neurobiology of the disease (e.g., a specific neurotransmitter deficit) or hypothesized antidementia approaches (e.g., anti-inflammation or antioxidation).

Cholinesterase Inhibitors
Tacrine*
Donepezil*
Metrifonate
Galanzamine
Rivastigmine
Sustained-release physostigmine
Eptastigmine
Cholinergic Agonists
Xanomeline
Milameline
Antioxidants
Vitamin E (alpha-tocopherol)
Hormones
Conjugated estrogens
Other Potential Cognitive Enhancers
Nonsteroidal anti-inflammatory drugs (NSAIDs)
Ginkgo biloba
Acetyl-L-carnitine

* FDA approved.

Table 51.3e-3 Medications Used or Under Investigation for the Treatment of Dementia of the Alzheimer's Type

Antipsychotics	**Anxiolytics**
Haloperidol	Buspirone
Thioridazine	Lorazepam
Risperidone	Clonazepam
Olanzapine	**Antiparkinsonian Drugs**
Quetiapine	Selegiline
Anticonvulsants	***β*-adrenergic receptor antagonists**
Carbamazepine	Propranolol
Divalproex sodium	Pindolol
Antidepressants	
Fluoxetine	
Sertraline	
Paroxetine	
Trazodone	

Table 51.3e-4 Some Drugs Used to Treat Behaviors Associated With Dementia

One approach to improving memory function has been to enhance the cholinergic activity in the brain. Autopsies of patients with Alzheimer's disease have found reduced levels of choline acetyltransferase in the brain, a finding consistent with a central cholinergic deficit. Trials of cholinergic agonist agents have temporarily improved memory in young and old normal volunteers. Trials of several agents—including physostigmine (Antilirium), choline (Mega-B), and lecithin (PhosChol) have yielded contradictory results (some effects versus no measurable effects on memory in demented patients) or else have not yet been replicated. The Food and Drug Administration has approved for short-term treatment two reversible cholinesterase inhibitors, tacrine (Cognex) and donepezil (Aricept), which have modest effects on memory and other aspects of cognition. They also may reduce behavioral symptoms in some patients, and prolonged cholinergic therapy may even delay disease progression. Tacrine has the disadvantage of causing reversible elevations of serum transaminase levels. Additional cholinergic agents are under study.

Other promising agents for cognitive enhancement or protection currently under study include vitamin E, deprenyl (Eldepryl), estrogen, and nonsteroidal anti-inflammatory drugs. As scientists uncover the basic pathogenetic mechanisms of Alzheimer's disease, additional antidementia treatments, designed to perform such functions as inhibition of b-amyloid production or accumulation or interference with apolipoprotein E physiology, are likely to emerge. Additional approaches may target other neurotransmitter systems (e.g., noradrenergic) or use drug combination strategies.

Clinical experience suggests that available therapies for behaviors associated with dementia (e.g., depression, agitation, anxiety) are sometimes effective. Patients who have dementia with concurrent depression may improve following treatment with antidepressant medications. Antidepressant drugs with minimal anticholinergic effects (e.g., selective serotonin reuptake inhibitors) are preferred over tricyclic drugs. Lithium (Eskalith) can be an effective antidepressant for geriatric depression and bipolar I disorder. However, patients with underlying neurological diseases have been reported to do poorly with lithium treatment, so it should be used with caution in patients with dementia.

Preliminary data point to the utility of anticonvulsant drugs (e.g., carbamazepine [Tegretol], valproate [Depakene]) for agitation as well. Psychotherapies aimed at enhancing cognition are ineffective for dementia. Clinical experience suggests the efficacy of nonpharmacological approaches such as optimizing sensory input and maintaining daily routines in minimizing agitation.

Antipsychotic drugs are effective in treating psychotic symptoms and agitation, with the choice of a specific agent depending on its adverse effect profile. A meta-analysis of controlled trials of antipsychotic treatment in dementia indicated that antipsychotic drugs had a significantly greater effect than did a placebo, but the degree of the effects was small. Moreover, efficacy differences between such agents as haloperidol (Haldol) and thioridazine (Mellaril) were not found. Those studies, along with clinical experience, seem to support the judicious use of antipsychotic medications in managing some behavioral complications of dementia. Clinicians should be aware of the risk of tardive dyskinesia in elderly patients. One study of elderly psychiatric patients showed that the greatest risk is during the first 2 years of antipsychotic treatment. Because of such adverse effects, alternative drugs have been used to treat agitation, including beta-blocking agents and sedating antidepressant drugs, such as trazodone (Desyrel).

Benzodiazepines have also been used to treat the agitation that accompanies dementia. However, they have undesirable adverse effects, including some particularly noxious to the demented elderly patient, such as confusion, memory impairment, disorientation, dysarthria, and agitation complicated by ataxic gait. Short-acting benzodiazepines that do not require oxidative metabolism in the liver and have no active metabolites are safer than long-acting agents. Long-acting benzodiazepine agents tend to accumulate in the blood and are best avoided, as are very short-acting compounds, which tend to reach high peak levels rapidly. Relatively short-acting benzodiazepines may be useful for treating insomnia, although the clinician should evaluate the specific causes of the insomnia, such as restless leg syndrome, obstructive apnea, urinary frequency caused by prostatic disease, lack of daytime exercise, use of caffeine, and prolonged stays in bed. Moreover, attempts at reducing daytime sleep, avoiding nighttime stimulants, and regulating the timing of meals and activities should be made before using pharmacological agents for sleep.

OTHER INTERVENTIONS

Attention to the Environment Patients with cognitive losses are sensitive to their surroundings and seem to do best with optimal stimulation. Understimulation may cause withdrawal; overstimulation may cause confusion and agitation. Familiar and constant surroundings maximize the patient's existing cognitive functions. Daily routines often increase a patient's sense of security; memory and orientation can be facilitated by prominent displays of clocks and calendars, a night light, checklists, and diaries. Medication schedules should be simplified, if possible. If moves cannot be avoided, it helps to place familiar objects (e.g., photographs and furniture) in the new environment and to create a homelike atmosphere. The availability of newspapers, radio, and television can be useful in maintaining a patient's contact with and awareness of the outside world.

Family Intervention Psychotherapeutic intervention with family members is a critical aspect of treatment. Education and counseling about the nature of the patient's illness help relatives cope with the anger and puzzlement they often experience when the patient with dementia behaves in peculiar, disturbing, and uncharacteristic ways. Relatives may need reassurance that their emotional reactions are common and that talking about them can bring relief. Many relatives also need help in grieving the loss of the patient who now behaves like a stranger rather than the person they once knew. The Alzheimer's Association, a national organization of family members that has local chapters throughout the United States, has been at the forefront of providing educational and emotional support for family members.

Dementia of the Alzheimer's type has captured the attention of the mass media. Consequently, the public's concern and anxiety have led to a heightened awareness of memory changes and a tendency to overinterpret normal age-related memory impairment as dementia of the Alzheimer's type. Family members of Alzheimer's victims are generally the most anxious about any memory changes they observe in themselves, given the likelihood of genetic components in the disease. Sometimes education and a full physical and neuropsychological evaluation allay, at least temporarily, any unfounded anxiety about the disease. At other times, such an evaluation will uncover the early signs of progressive dementia.

In many situations the elderly person with dementia is the identified patient. But interpersonal conflicts among the family members require resolution to help the geriatric patient. Moreover, the adult child often makes the initial contact with the geriatric psychiatrist, and considerable skill and sensitivity are necessary to maintain an alliance with the adult child while respecting the elderly parent's autonomy, dignity, and privacy.

SUGGESTED CROSS-REFERENCES

For additional discussion on differential diagnosis, etiology, and pathology, see Chapter 10 on dementia. Section 51.4f provides details on the use of psychotropic medications for elderly patients. Section 51.6b and Section 51.6c on medical-legal and ethical considerations offer relevant issues for mentally impaired elderly persons. Section 51.2d on neuropsychological examination of elderly patients provides details on the assessment of cognitive impairment. Other relevant sections from the geriatric psychiatry chapter include Section 51.2b on central nervous system changes, Section 51.2e and Section 51.2f on neuroimaging, Section 51.6a on long-term care, Section 51.5d on community services, and Section 51.5a on financial issues.

SECTION REFERENCES

Aarsland D, Cummings JL, Yenner G, Miller B: Relationship of aggressive behavior to other neuropsychiatric symptoms in patients with Alzheimer's disease. Am J Psychiatry *153:*243, 1996.

*Alzheimer's Disease and Related Dementias Guideline Panel: *Recognition and Initial Assessment of Alzheimer's Disease and Related Dementias*. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. AHCPR Publication No. 97-0702, Rockville, MD, November 1996.

*American Psychiatric Association: Practice guideline for treatment of patients with Alzheimer's disease and other dementias of late life. Am J Psychiatry *154*(Suppl):1, 1997.

Blessed G, Tomlinson BE, Roth M: The association between quantitative measures of dementia and of senile change in the cerebral gray matter of elderly subjects. Br J Psychiatry *114:*797, 1968.

Brayne C, Gill C, Huppert FA, Barkley C, Gehlhaar E, Girling DM, O'Connor DW, Paykel ES: Incidence of clinically diagnosed subtypes of dementia in an elderly population. Cambridge Project for Later Life. Br J Psychiatry *167:*255, 1995.

Clark LM, McDonald WM, Welsh-Bohmer KA, Siegler IC, Dawson DV, Tupler LA, Krishnan KR: Magnetic resonance imaging correlates of depression in early- and late-onset Alzheimer's disease. Biol Psychiatry *44:*592, 1998.

Hulette CM, Welsh-Bohmer KA, Murray MG, Saunders AM, Mash DC, McIntyre LM: Neuropathological and neuropsychological changes in “normal” aging: Evidence for preclinical Alzheimer disease in cognitively normal individuals. J Neuropathol Exp Neurol *57:*1168, 1998.

Hunt L, Morris JC, Edwards D, Wilson BS: Driving performance in persons with mild senile dementia of the Alzheimer type. J Am Geriatr Soc *41:*747, 1993.

Khachaturian ZS, Radebaugh TS, editors: *Alzheimer's Disease: Cause(s), Diagnosis, and Care*. CRC Press, New York, 1996.

Larrabee GJ, Crook TH: Estimated prevalence of age-associated memory impairment derived from standardized tests of memory function. Int Psychogeriatr *6:*95, 1994.

Levy ML, Miller BL, Cummings JL, Fairbanks LA, Craig A: Alzheimer disease and frontotemporal dementias: Behavioral distinctions. Arch Neurol *53:*687, 1996.

*Lyketsos CG, Chen LS, Anthony JC: Cognitive decline in adulthood: An 11.5-year follow-up of the Baltimore Epidemiologic Catchment Area Study. Am J Psychiatry *156:*58, 1999.

*Lyketsos CG, Steele C, Galik E, Rosenblatt A, Steinberg M, Warren A, Sheppard J-M: Physical aggression in dementia patients and its relationship to depression. Am J Psychiatry *156:*66, 1999.

Mace NL, Rabins PV: *The 36-Hour Day. Johns Hopkins University Press, Baltimore, 1991*.

Max W: The economic impact of Alzheimer's disease. Neurology *43*(Suppl):S6, 1993.

Mayeux R, Ottman R, Maestre G, Ngai C, Tang MX, Ginsberg H, Chun M, Tycko B, Shelanski M: Synergistic effects of traumatic head injury and apolipoprotein-epsilon 4 in patients with Alzheimer's disease. Neurology *45:*555, 1995.

McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM: Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of department of health and human services task force on Alzheimer's disease. Neurology *34:*939, 1984.

Mori E, Hirono N, Yamashita H, Imamura T, Ikejiri Y, Ikeda M, Kitagaki H, Shimomura T, Yoneda Y: Premorbid brain size as a determinant of reserve capacity against intellectual decline in Alzheimer's disease. Am J Psychiatry *154:*18, 1997.

Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, Mellits ED, Clark C, and the CERAD Investigators: The consortium to establish a registry for Alzheimer's disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology *9:*1159, 1989.

Moye J, Robiner WN, Mackenzie TB: Depression in Alzheimer patients: Discrepancies between patient and caregiver reports. Alzheimer Dis Assoc Disord *7:*187, 1993.

Price BH, Gurvit H, Weintraub S, Geula C, Leimkuhler E, Mesulam M: Neuropsychological patterns and language deficits in 20 consecutive cases of autopsy-confirmed Alzheimer's disease. Arch Neurol *50:*931, 1993.

Relkin NR, Tanzi R, Breitner J, Farrer L, Gandy S, Haines J, Hyman B, Mullan M, Poirer J, Strittmatter W, Folstein M, Mayeux R, Petersen R, Roses A, Schenk D, Small G, Vangool W, Cook-Deegan R, Fleck L, Kapp M, Karlinsky H, Pericak-Vance M, Post S, Wolpert C: Apolipoprotein E genotyping in Alzheimer's disease: Position statement of the National Institute on Aging/Alzheimer's Association Working Group. Lancet *347:*1091, 1996.

*Roses AD, Weisgraber KH, Christen Y, editors: *Apolipoprotein E and Alzheimer's Disease*. Springer, New York, 1996.

*Schneider LS, Tariot PN, Small GW: Update on treatment for Alzheimer's disease and other dementias. In *Psychiatric Clinics of North America: Annual of Drug Therapy*, ed 4, Dunner DL, Rosenbaum JF, editors. Saunders, Philadelphia, 1997.

Small GW, Donohue JA, Brooks RL: An economic evaluation of donepezil in the treatment of Alzheimer's disease. Clin Ther *20:*838, 1998.

Small GW, La Rue A, Komo S, Kaplan A, Mandelkern MA: Predictors of cognitive change in middle-aged and older adults with memory loss. Am J Psychiatry *152:*1757, 1995.

Small GW, Mazziotta JC, Collins MT, Baxter LR, Phelps ME, Mandelkern MA, Kaplan A, La Rue A, Adamson CF, Chang L, Guze BH, Corder EH, Saunders AM, Haines JL, Pericak-Vance MA, Roses AD: Apolipoprotein E type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease. JAMA *273*:942, 1995.

Small GW, Rabins PV, Barry PP, Buckholtz NS, DeKosky ST, Ferris SH, Finkel SI, Gwyther LP, Khachaturian ZS, Lebowitz BD, McRae TD, Morris JC, Oakley F, Schneider LS, Streim JE, Sunderland T, Teri LA, Tune LE: Diagnosis and treatment of Alzheimer disease and related disorders: Consensus statement of the American Association for Geriatric Psychiatry, the Alzheimer's Association, and the American Geriatrics Society. JAMA *278:*1363, 1997.

Stern Y, Folstein M, Albert M, Richards M, Miller L, Bylsma F, Lafleche G, Marder K, Bell K, Sano M, Devanand D, Loreck D, Wootten J, Bello J: Multicenter study of predictors of disease course in Alzheimer disease (the “predictors study”): I. Study design, cohort description, and intersite comparisons. Alzheimer Dis Assoc Disord *7:*3, 1993.

Sultzer DL, Mahler ME, Cummings JL, Van Gorp WG, Hinkin CH, Brown C: Cortical abnormalities associated with subcortical lesions in vascular dementia: Clinical and positron emission tomographic findings. Arch Neurol *52:*773, 1995.

Sweet RA, Mulsant BH, Gupta B, Rifai AH, Pasternak RE, McEachran A, Zubenko GS: Duration of neuroleptic treatment and prevalence of tardive dyskinesia in late life. Arch Gen Psychiatry *52:*478, 1995.

Terry RD, Katzman R, Bick KL, editors: *Alzheimer Disease*. Raven, New York, 1994.

*Vellas B, Fitten LJ, Dubois B, Albarede JL, editors: *Facts and Research in Gerontology: Alzheimer Disease*. Springer, New York, 1996.

*Zubenko GS, Winwood E, Jacobs B, Teply I, Stiffler JS, Hughes HB, Juff RJ, Sunderland T, Martinez AJ: Prospective study of risk factors for Alzheimer's disease: Results at 7.5 years. Am J Psychiatry *156:*50, 1999.

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3F SCHIZOPHRENIA AND DELUSIONAL DISORDERS

M. JACKUELYN HARRIS, M.D. AND DILIP V. JESTE, M.D.

Beginning with Emil Kraepelin's description of dementia precox, schizophrenia has usually been considered a disorder with onset during adolescence or early adulthood and a progressively downhill course. Over the past century, most studies of schizophrenia have been restricted to young adult patients. Relatively little research outside the United Kingdom and certain other European countries has examined the possible onset of schizophrenia in middle or old age or evaluated the course of schizophrenia in elderly patients. While the European research in this area has been much richer than that done elsewhere, it is limited by variable terminology and unclear diagnostic criteria. Recent investigations seem to challenge the original Kraepelinian notion of dementia precox in terms of both age of onset and long-term course of schizophrenia.

HISTORY AND COMPARATIVE NOSOLOGY

Psychotic symptoms with onset in middle-age or later life that were not related to a mood disorder or to a mental disorder due to a general medical condition were previously grouped as one disorder (paraphrenia) or as several different disorders (paraphrenia, paranoia, schizophrenia). Kraepelin used the term “paraphrenia” to describe a group of patients who had many of the symptoms of dementia precox but with less disturbance of emotion and volition. The term “paranoia” described a chronic well-systematized delusional state without hallucinations. Manfred Bleuler described late-onset schizophrenia as presenting after the age of 40 and otherwise similar to earlier-onset schizophrenia. The first edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I) included the category involutional psychotic reaction for psychotic symptoms occurring during the involutional period. Some of the cases were characterized by depression and others by paranoia.

Martin Roth and colleagues used the terms “paraphrenia” and “late paraphrenia” (onset of the disorder after the age of 60). They divided their paraphrenic patients into three groups: (1) those whose symptoms appeared to be an extension of long-standing personality disorders; (2) those who had delusions that seemed related to their particular stressors or circumstances; and (3) those considered to have “endogenous” paraphrenia (i.e., not attributable to a long-term personality disorder or any particular stressors). Kurt Leonhand and other German authors of the time used the term *paraphrenia* to refer to paranoid schizophrenic patients without specifying age of onset. Felix Post categorized nonaffective paranoid disorders beginning after the age of 50 into three groups: (1) paranoid hallucinosis, with no psychotic symptoms other than auditory hallucinations associated with persecutory delusions; (2) schizophreniform syndrome, with somewhat more understandable paranoid symptoms, and (3) schizophrenic syndrome, with Schneiderian first-rank symptoms. The second edition of DSM (DSM-II) used the term “involuntal state” (involutional paraphrenia) to describe a delusional state with onset during the involutional period. No age cutoff was specified for the diagnosis of schizophrenia. The third edition of DSM (DSM-III) stipulated that a diagnosis of schizophrenia could be made only if the symptoms began before the age of 45. It also included the category “paranoid disorder,” which referred to persistent persecutory delusions without prominent hallucinations or other major symptoms of schizophrenia. The revised third edition of DSM (DSM-III-R) recommended the specification of late-onset schizophrenia if the symptoms began after age 45. It replaced the term “paranoid disorder” with the term “delusional disorder,” modified the diagnostic criteria to include a minimum duration of 1 month, and broadened the scope of delusions to include grandiose, erotomanic, somatic, jealous, and unspecified types. The fourth edition of DSM (DSM-IV) neither restricts the age of onset for the diagnosis of schizophrenia nor specifies the subcategory of late-onset schizophrenia. In this section *late-onset schizophrenia* refers to the onset of prodromal symptoms of schizophrenia after the age of 45 years. Notwithstanding the DSM-IV, late-onset schizophrenia is both a clinically and theoretically important condition that may or may not be a neurobiologically distinct subtype of schizophrenia. The criteria for delusional disorder in DSM-IV have not changed from those in DSM-III-R.

Thus schizophrenia and delusional disorder with onset in middle age and later life are considered separate diagnoses, just as schizophrenia is considered to be a separate disorder in younger patients. More research is needed, however, to establish independent diagnostic validity of those two disorders in late life.

LATE-ONSET SCHIZOPHRENIA

Epidemiology The prevalence of late-onset schizophrenia is not known. A 1988 review of studies of late-onset schizophrenia found that approximately 23 percent of patients with schizophrenia were reported to have experienced the onset of the disorder after age 40, with 13 percent in the fifth decade of life, 7 percent in the sixth decade, and 3 percent thereafter. A 1993 study found that 28 percent of first-contact patients (inpatients and outpatients) had onset of illness after the age of 44, and 12 percent after the age of 64. The exact prevalence of schizophrenia in the elderly population is not known. It has been suggested, however, that of the older patients with schizophrenia, approximately 90 percent have had onset during adolescence or early adulthood, while the remaining 10 percent have later-onset schizophrenia.

Studies of late-onset schizophrenia that have included both men and women have consistently found a predominance of women, which contrasts with the preponderance of men among schizophrenic patients between the ages of 15 and 34. It has been suggested that the relative excess of dopamine type 2 (D_2) receptors in young men (compared with young women) and elderly women (compared with elderly men) may help explain the age-related gender differences seen in schizophrenia. Alternatively, estrogen may help delay the onset of schizophrenic symptoms in predisposed women until after menopause. A large proportion of the patients with late-onset schizophrenia are married or were in the past. Many had been satisfactorily employed or had worked as homemakers prior to the onset of their illness.

Etiology The etiology of late-onset schizophrenia remains to be clarified. The data regarding family history of schizophrenia in the late-onset patients are conflicting. Some European studies found that late-onset schizophrenia was more weakly associated with a family history of schizophrenia than the early-onset form. On the other hand, a recent U.S. study reported that the prevalence of schizophrenia was similar in the early-onset and late-onset patients. It is not clear, however, whether the age of onset of schizophrenia in succeeding generations decreases as a function of the age of onset of the proband or whether the relatives also have late-onset schizophrenia. Abnormal premorbid personality traits (especially paranoid and schizoid) have been noted in a sizable proportion of patients with the diagnosis of late-onset schizophrenia. The pattern of impairment on neuropsychological testing is similar in early-onset and late-onset schizophrenia patients when corrected for age, education, and gender, but the late-onset group has less impairment in learning and abstraction and flexibility of thinking. A number of studies have reported an association between late-onset schizophrenia and sensory (visual and auditory) deficits, although sensory impairment has not been shown to be a risk factor for late-onset schizophrenia. A recent study suggested that older psychiatric patients might have difficulty obtaining optimal correction of their age-related visual or hearing deficits. Brain imaging studies have shown nonspecific structural abnormalities such as slightly enlarged ventricles and increased white matter abnormalities. One preliminary study found that older age at onset of schizophrenia was associated with a larger volume of thalamus on magnetic resonance imagery (MRI).

In general, patients with late-onset forms of schizophrenia and those with early-onset form are similar in terms of positive symptoms, the overall pattern of neuropsychological impairment (e.g., relatively normal retention of information), nonspecific abnormalities on structural brain imaging (e.g., slightly enlarged ventricles and increased white matter hyperintensities), chronic course, and qualitative response to neuroleptic drugs. Such similarities support the notion that the patients properly diagnosed with late-onset schizophrenia indeed have a schizophrenic disorder. However, important differences exist between the two forms. For example, patients with the late-onset variety are more likely to be female, be of paranoid subtype, exhibit fewer negative symptoms and less-severe deficits in learning and abstraction, possibly have somewhat different specific structural brain abnormalities, and need for lower daily dosages of antipsychotic drugs than age-comparable early-onset patients. These differences suggest that late-onset schizophrenia is probably a neurobiologically distinct form of schizophrenia.

Diagnosis and Clinical Features Patients with late-onset schizophrenia have symptoms (especially positive symptoms) largely similar to those of patients with early-onset schizophrenia. A comparative study of early-onset (age of onset 44 years or less) and late-onset (age of onset 45 years or more) schizophrenic patients found no difference between the two groups in the prevalence of delusions of reference, bizarre delusions, or lack of insight. Patients with the late-onset form,

however, tended to have more persecutory delusions with and without hallucinations, organized delusions, and abusive auditory hallucinations or hallucinations with a running commentary. The most common symptoms in the late-onset group were delusions, and approximately two thirds had bizarre delusions. Many of the patients have hallucinations, which are usually auditory. Late-onset schizophrenia patients tend to differ from those with earlier onset in having a lower prevalence of looseness of associations, inappropriateness of affect, and other negative symptoms. Consistent with these findings, late-onset schizophrenia patients typically meet criteria for the paranoid or undifferentiated subtype of schizophrenia.

The patient was a 57-year-old divorced white man who required psychiatric hospitalization for the first time at the age of 49, after presenting with a 6-month history of being under surveillance by the police, Central Intelligence Agency (CIA), and detectives who were following him and monitoring his actions. He became very distraught and claimed to have gotten “tired of dodging them” and turned himself in to the police. He was taken by the police to the hospital for admission. He denied hearing voices but reported a ringing in his ears warning him that something harmful would soon happen. The patient admitted seeing faces in pictures that changed from time to time while he looked at them. At the time of admission the patient was working full-time as a painter and lived with his elderly father as he had since his divorce 19 years earlier. He frequently saw his three sons who lived in the area. The family reported no previous history of mental illness in the patient. The patient's hobby was collecting model trains. His medical history was unremarkable, but the family history was significant for a maternal aunt who spent many years in a psychiatric hospital and his mother who required state hospitalization for 7 years for a “nervous breakdown.” The neurological and other physical examination was within normal limits as were routine laboratory tests and MRI of the brain. The patient was treated with haloperidol up to 4 mg a day with resolution of his delusional system except for episodic paranoid ideation and hearing voices. The patient was no longer employed but helped care for his 90-year-old father.

Differential Diagnosis The principal disorders in the differential diagnosis of late-onset schizophrenia include earlier-onset schizophrenia, mood disorders with psychotic features, delusional disorder, psychosis associated with cognitive disorders, psychotic disorder due to a general medical condition, and substance-induced psychotic disorder.

Early-Onset Schizophrenia Dating the onset of schizophrenia can be difficult. Many schizophrenic patients have an insidious onset of symptoms or may have premorbid personality traits that are difficult to differentiate from prodromal symptoms. Earlier onset and thus longer duration of schizophrenia are generally associated with more severe negative symptoms and greater cognitive deficits (especially in the areas of learning and abstraction).

Mood Disorder Major depression and bipolar disorder may be accompanied by psychotic symptoms such as delusions and hallucinations. Some, but not all, studies suggest that psychotic depression is more likely to occur in older patients than in younger ones. Mood-congruent or mood-incongruent psychotic features may occur with either diagnosis, making it difficult to differentiate them from schizophrenia. A careful psychiatric history of symptoms consistent with a mood disorder as well as a past history of mood symptoms may help to differentiate the two conditions. In addition, a positive family history of mood disorder may be helpful but is not necessarily of diagnostic value.

Delusional Disorder Delusional disorder is differentiated from schizophrenia by a lack of prominent auditory or visual hallucinations and an absence of deterioration in areas of functioning outside the delusional sphere. The delusions are necessarily nonbizarre (i.e., they involve situations that may occur in real life, such as theft, suffering from a disease, or spousal infidelity), while those in schizophrenia could be bizarre (e.g., having a computer installed in the patient's brain by a neighbor who wants to read the patient's mind).

Psychotic Symptoms Associated With Cognitive Disorders or Due to a General Medical Condition A patient who presents with psychotic symptoms after the age of 45, particularly if the symptoms present after the age of 60, requires a careful evaluation for underlying neurological and other medical disorders. Dementia is a common disorder in elderly persons, and approximately one third of patients with a diagnosis of dementia of the Alzheimer's type may present with psychotic symptoms at some point during the course of their dementia. Other types of dementia, such as alcohol-induced persisting dementia and vascular dementia, also need to be ruled out. The abuse, and more commonly misuse, of prescription and over-the-counter medications should also be evaluated.

Course and Prognosis Late-onset schizophrenia tends to follow a chronic course, which may be punctuated by partial remissions and exacerbations. The prognosis is somewhat better than that in early-onset schizophrenia. Nonetheless, mortality (especially from suicide) exceeds that in the general population and is probably comparable to that in early-onset schizophrenia.

Treatment Few systematic studies of the treatment of late-onset schizophrenia exist. It is clear, however, that a number of patients with late-onset schizophrenia respond well to low-dose antipsychotic agents. Maintenance therapy is frequently required since many of these patients relapse if they discontinue the medication. Elderly patients are more susceptible to adverse effects such as sedation, anticholinergic toxicity, and extrapyramidal symptoms. Similarly, the risk of neuroleptic-induced tardive dyskinesia is several times greater in older patients. Whereas the cumulative annual incidence of tardive dyskinesia in younger adults is 4 to 5 percent, that in middle-aged and elderly patients is 25 to 30 percent.

Recently, some of the “atypical” antipsychotic agents with serotonergic and dopaminergic blocking activity have been introduced for the treatment of schizophrenia including clozapine (Clozaril), risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine (Seroquel). So far, the data on the use of these drugs in elderly patients are quite limited. These drugs have fewer extrapyramidal adverse effects, and may be more effective against negative symptoms. Some recent data suggest that the newer agents may also carry a lower risk of tardive dyskinesia and have a better effect on cognitive deficits in schizophrenia, although these findings need confirmation. These drugs do have their own adverse effects, and much lower dosages (about a third to a quarter) are recommended for elderly patients than for younger adults.

Psychosocial treatment is a vital aspect of the management of a late-onset schizophrenia patient. Establishing a good therapeutic relationship in a supportive setting is important. Although therapists need not agree with the patients' delusional system, they should convey a sense of understanding and empathy. Clinicians may seek the help of the patients' caregivers, family members, friends, neighbors, and priests or other religious figures in the community to facilitate the overall management. Such a supportive network may result in the early detection and resolution of emergency situations and thus avoid unnecessary hospitalizations as well as legal problems. Therapists may also use social agencies to help patients obtain appropriate financial, medical, nutritional, or transportational assistance. Clinicians may recommend conservatorship or guardianship for patients, with a view to ensuring proper health care and management of the patient's finances.

Course and Prognosis Only a handful of studies have followed the course of early-onset schizophrenia into old age. The findings from these studies challenge the notion implied in the term *dementia precox* according to which schizophrenia is associated with progressive deterioration. Overall, approximately 20 percent of patients with schizophrenia, as they age, begin to experience improvement in both positive and negative symptoms and eventually have psychosocial remission. Another 20 percent of patients have a worsening of symptoms, and the course is relatively unchanged in the remaining 60 percent. Thus, while schizophrenia continues to be symptomatic in older ages in most patients, a possibility of remission exists in a small proportion of the subjects. Accurate prediction of the long-term course of the illness in an individual patient is not possible, but the general factors associated with eventual positive outcome include female gender, later onset of symptoms, paranoid subtype, less severe negative symptoms, and better premorbid functioning.

The required (and tolerated) neuroleptic drug dosage tends to decrease with aging as a result of both pharmacokinetic and pharmacodynamic factors. At the same time, the marked heterogeneity among the elderly patients with schizophrenia indicates the need to individualize dosage regimens.

DELUSIONAL DISORDER

Epidemiology Studies suggest that between 2 and 8 percent of the elderly psychiatric population suffers from some type of paranoid symptoms. One study indicated that although only about 2 percent of elderly psychiatric patients satisfied DSM-III criteria for paranoid disorder, another 13 percent had paranoid ideation. DSM-IV estimates that the population prevalence of delusional disorder is approximately 0.03 percent. Delusional disorder usually first appears in middle-to-late adulthood. The overall prevalence of delusional disorder is slightly higher among women than among men, and the average age of onset is earlier for men (40 to 49) than for women (60 to 69).

Etiology Several factors have been postulated to contribute to development of delusional disorder: a family history of schizophrenia; avoidant, paranoid, or schizoid personality disorder; hearing loss; or being an immigrant or of low socioeconomic status. The available evidence implicating each of these factors needs confirmation. From the viewpoint of psychological defense mechanisms, delusions are generally attributed to an excessive and inappropriate use of projection. The patient tends to hold others responsible for any psychobiosocial losses that may accompany normal aging.

Diagnosis and Clinical Features Persecutory, somatic, or erotomanic delusions in the elderly patient are usually secondary to another neuropsychiatric disorder (e.g., schizophrenia, mood disorder, or dementia). Occasionally, however, a primary delusional disorder is implicated. The conceptualization and study of this disorder, like that of late-onset schizophrenia, has been hampered by inconsistencies in nomenclature and diagnostic criteria.

The essential feature of delusional disorder is a persistent delusion. In addition, the behavior is not obviously bizarre, and auditory and visual hallucinations, if present, are not prominent. The DSM-IV describes the following types of delusional disorder: erotomanic, grandiose, jealous, persecutory, somatic, mixed, and unspecified.

> The patient was a 74-year-old divorced white man referred by his primary care physician for paranoid delusions. The patient reported that a black gang was after him, especially “Mad Dog,” the leader of the 5000-member gang. The patient and his son said that the onset of the patient's concerns began 4 years earlier, and this resulted in the patient changing residences four times during this period and retiring from his job as a sales clerk in a clothing store 2 years ago. He called the police several times to show them the gang's footprints in the grass and to report noise outside his window. The patient denied depressed mood, change in sleep or appetite, memory impairment, or problems with concentration. He denied auditory or visual hallucinations or any past psychiatric history. The patient lived alone in his apartment and took care of his finances and other activities of daily living. He had a high school diploma, was divorced, and had worked as a sales clerk “all his life.” Medical problems were significant for hypercholesterolemia treated with medication. His Mini-Mental State Examination score was 30/30. His physical (including neurological) examination and routine laboratory tests were unremarkable. The patient refused to take the prescribed antipsychotic drug but did come to the clinic to discuss his concerns.

Common types of delusions seen in elderly patients with delusional disorder are skin infestation by insects; organs such as liver are cancerous (despite a lack of evidence); the person is being spied on, followed, or poisoned; and the spouse is unfaithful (without a real factual support).

Differential Diagnosis The considerations in the differential diagnosis of delusional disorder are similar to those for schizophrenia with onset late in life. One study noted more cerebral infarction on computed tomography (CT) scan than in normal control subjects. Another study reported somewhat milder cognitive deficits in older delusional disorder patients than in similarly aged patients with late-onset schizophrenia.

Course and Prognosis The course tends to be chronic, especially in those with the persecutory type of delusional disorder. Others may have partial remissions and relapses. Generally, both intellectual performance and occupational function are preserved, but marital and social functioning is compromised. The data on prognosis are limited, although the likelihood of complete recovery seems very low.

Treatment Antipsychotic agents are often effective, especially in agitated delusional patients, but they must be administered cautiously in the elderly patient. Noncompliance is a common problem in the treatment of patients with delusional disorder. Establishing rapport with the patients is critical for continued treatment.

SUGGESTED CROSS-REFERENCES

Schizophrenia is discussed in Chapter 12, and delusional disorder is discussed in Section 13.2. Antipsychotic drugs are discussed in Section 31.26 on atypical antipsychotics, and Section 31.17 on dopamine receptors antagonists. Psychopharmacological treatment for elderly patients is presented in Section 51.4b.

SECTION REFERENCES

Bleuler M: Late schizophrenic clinical pictures. Fortschr Neurol Psychiatr *15:*259–290, 1943.

Castle DJ, Murray RM: The epidemiology of late-onset schizophrenia. Schizophr Bull *19:*691–700, 1993.

Castle DJ, Wessely S, Howard R, Murray RM: Schizophrenia with onset at the extremes of adult life. Int J Geriatr Psychiatry *12:*712, 1997.

Corey-Bloom J, Jernigan T, Archibald S, Harris MJ, Jeste DV: Quantitative magnetic resonance imaging in late-life schizophrenia. Am J Psychiatry *152:*447, 1995.

Eastham JH, Jeste DV: Treatment of schizophrenia and delusional disorder in the elderly. Eur Arch Psychiatry Clin Neurosci *247:*209, 1997.

Evans JD, Paulsen JS, Harris MJ, Heaton RK, Jeste DV: A clinical and neuropsychological comparison of delusional disorder and schizophrenia. J Neuropsychiatr Clin Neurosci *8:*281, 1996.

Flint AJ, Rifat SI, Eastwood MR: Late-onset paranoia: Distinct from paraphrenia? Int J Geriatr Psychiatry *6:*103, 1991.

Gurian BS, Wexler D, Baker EH: Late-life paranoia: Possible association with early trauma and infertility. Int J Geriatr Psychiatry *7:*277, 1992.

Hafner H, Hambrecht M: The elderly with schizophrenia. In *Functional Psychiatric Disorders in the Elderly*, E Chiu, D Ames, editors. Cambridge University Press, Cambridge, 1994.

Harding CM, Brooks GW, Ashikaga T, Strauss JS, Breier A: The Vermont longitudinal study of persons with severe mental illness, I: Methodology, study sample, and overall status 32 years later. Am J Psychiatry *144:*718, 1987.

Harding CM, Brooks GW, Ashikaga T, Strauss JS, Breier A: The Vermont longitudinal study, II: Long-term outcome of subjects who once met the criteria for DSM-III schizophrenia. Am J Psychiatry *144:*727, 1987.

Harris MJ, Heaton RK, Schalz A, Bailey A, Patterson TL: Neuroleptic dose reduction in older psychotic patients. Schizophr Res *27:*241, 1997.

*Harris MJ, Jeste DV: Late-onset schizophrenia: An overview. Schizophr Bull *14:*39, 1988.

Harvey P, Lombardi J, Leibman M, White L, Parrella M, Powchik P, Mohs R, Davidson M: Performance of chronic schizophrenic patients on cognitive neuropsychological measures sensitive to dementia. Int J Geriatr Psychiatry *11:*621, 1996.

*Harvey PD, Silverman JM, Mohs RC, Parrella M, White L, Powchik P, Davidson M, Davis KL: Cognitive decline in late-life schizophrenia: A longitudinal study of geriatric chronically hospitalized patients. Biol Psychiatry *45:*32, 1999.

*Henderson AS, Kay DWK: The epidemiology of functional psychoses of late onset. Eur Arch Psychiatry Clin Neurosci *247:*176, 1997.

Howard R, Castle D, Wessely S, Murray R: A comparative study of 470 cases of early-onset and late-onset schizophrenia. Br J Psychiatry *163:*352, 1993.

Howard R, Graham C, Sham P, Dennehey J, Castle DJ, Levy R, Murray R: A controlled family study of late-onset non-affective psychosis (late paraphrenia). Br J Psychiatry *170:*511, 1997.

Jeste DV, Caligiuri MP: Tardive dyskinesia. Schizophr Bull *19:*303, 1993.

Jeste DV, Caligiuri MP, Paulsen JS, Heaton RK, Lacro JP, Harris MJ, Bailey A, Fell RL, McAdams LA: Risk of tardive dyskinesia in older patients: A prospective longitudinal study of 266 patients. Arch Gen Psychiatry *52:*756, 1995.

Jeste DV, Harris MJ, Krull A, Kuck J, McAdams LA, Heaton R: Clinical and neuropsychological characteristics of patients with late-onset schizophrenia. Am J Psychiatry *152:*722, 1995.

*Jeste DV, McAdams LA, Palmer BW, Braff D, Jernigan TL, Stout J, Paulsen JS, Symonds L, Bailey A, Heaton RK: Relationship of neuropsychological and MRI measures with age of onset of schizophrenia. Acta Psychiatr Scand *98:*156, 1998.

*Jeste DV, Symonds LL, Harris MJ, Paulsen JS, Palmer BW, Heaton RK: Non-dementia non-praecox dementia praecox? Late-onset schizophrenia. Am J Geriatr Psychiatry *5:*302, 1997.

Kraepelin E: *Dementia Praecox and Paraphrenia*, 1919, RM Barclay, translator. Krieger, Huntington, NY, 1993.

Lesser IM, Miller BL, Swartz JR, Boone KB, Mehringer CM, Mena I: Brain imaging in late-life schizophrenia and related psychoses. Schizophr Bull *19:*773, 1993.

*Pearlson G, Rabins P: The late-onset psychoses: Possible risk factors. In *The Psychiatric Clinics of North America: Psychosis and Depression in the Elderly*, DV Jeste, S Zisook, editors. Saunders,

Philadelphia, 1988.

Pearlson GD, Tune LE, Wong DF, Aylward EH, Barta PE, Powers RE, Chase GA, Harris GJ, Rabins PV: Quantitative D $_2$ dopamine receptor PET and structural MRI changes in late onset schizophrenia. Schizophr Bull *19:*783, 1993.

Prager S, Jeste DV: Sensory impairment in late-life schizophrenia. Schizophr Bull *19:*755, 1993.

*Riecher-Rossler A, Loffler W, Munk-Jorgensen P: What do we really know about late-onset schizophrenia? Eur Arch Psychiatry Clin Neurosci *247:*195, 1997.

Rockwell E, Krull AJ, Dimsdale J, Jeste DV: Late-onset psychosis with somatic delusions. Psychosomatics *35:*66, 1994.

Salzman C: *Clinical Geriatric Psychopharmacology*, ed 3. Williams & Wilkins, Baltimore, 1998.

Symonds LL, Olichney JM, Jernigan TL, Corey-Bloom J, Healy JF, Jeste DV: Lack of clinically significant structural abnormalities in MRIs of older patients with schizophrenia and related psychoses. J Neuropsychiatr Clin Neurosci *9:*251, 1987.

Yassa R, Dastoor D, Nastase C, Camille Y, Belzile L: The prevalence of late-onset schizophrenia in a psychogeriatric population. J Geriatr Psychiatry Neurol *6:*120, 1993.

Yassa R, Suranyi-Cadotte B: Clinical characteristics of late-onset schizophrenia: Comparison with delusional disorder with and without hallucinations. Schizophr Bull *19:*701, 1993.

*Yorston G: Aged and dangerous: Old-age forensic psychiatry. Brit J Psychiatry *174:*193, 1999.

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3G PERSONALITY DISORDERS

ROBERT M. ROHRBAUGH, M.D.

Personality disorders are enduring patterns of inner experience and behavior that markedly differ from cultural norms, are pervasive and inflexible, and lead to subjective distress or social and occupational impairment. Longitudinal studies have shown personality functioning, both normal and abnormal, to be remarkably constant across the life cycle; the major change in personality with aging tends to be increased introversion.

EPIDEMIOLOGY

Personality disorders manifest in early adulthood and are thought to be relatively stable over time. Despite this stability over time, cross-sectional comparisons show a lower prevalence of personality disorders in United States and European community samples of older adults than in younger individuals. This decrease in prevalence of personality disorders in the community-dwelling elderly population is subject to a number of interpretations. One factor contributing to the difference may be an age bias in diagnostic criteria; many of the examples of personality disorder in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) and in commonly administered personality inventories highlight activities like work and sexual function, activities likely to be more prevalent in younger patients than in older ones. Other potential explanations for the decrease in prevalence with age include diminished or biased recall of events by elderly persons, selective mortality due to impulsive acts by individuals with personality disorders, fewer impulsive actions by elderly persons (“burn out”), and other cohort effects caused by different environmental or cultural experiences. The burn-out hypothesis is supported by an analysis by George Vaillant of a 40-year prospective study of male college students that demonstrated that the internal distress and impairment caused by personality disorder diminished by age 60. Other recent studies have suggested that after age 60 some patients with personality disorder may experience an intensification of symptoms. Thus, some patients with minor dysfunction during middle adulthood may become more impaired in late life and require clinical attention for the first time.

Most studies have found that between 15 and 30 percent of elderly psychiatric inpatients have personality disorders; rates in older age inpatient samples appear to be comparable to those in younger age samples. The proportion of elderly inpatients with a personality disorder varies, depending on the Axis I disorder of the patient. A recently published study using criteria from the revised third edition of DSM (DSM-III-R) found that only 6 percent of patients with an organic mental disorder met criteria for a personality disorder, while 24 percent of those with major depression also met criteria for personality disorder.

Comorbidity Much recent research on personality disorders in the elderly population has focused on the relationship between personality disorder and mood disorders. About 25 percent of mixed-age, community-dwelling patients in American, European, and Japanese populations were diagnosed with both major depressive disorder and a personality disorder. While these patients showed improvement in depressive symptoms after treatment, they remained more depressed and had poorer social functioning than the group without comorbid personality disorders. Both Cluster A (in a mixed-age sample) and Cluster B disorders (in an elderly sample) have been reported to predict significantly worse psychosocial and symptomatic outcomes after treatment of major depressive disorder. A prospective study of outpatients initially evaluated in middle age suggested that the negative effect of comorbid personality disorder on outcome of major depressive disorder may continue for 15 to 20 years after the index episode of major depression. However the presence of a comorbid personality disorder has not been shown to affect 1-year relapse rates in a group of elderly patients with major depressive disorder.

Elderly patients with an early age of onset of major depressive disorder have significantly higher scores on avoidant and dependent personality dimensions than do those with late onset of major depressive disorder. Whether a personality disorder predisposes patients to develop early-onset major depressive disorder or is caused by a major depressive disorder early in life is of clear interest to both researchers and clinicians.

Like those with depressive disorders, approximately 20 percent of a mixed-age sample of patients with bipolar I disorder exhibited a comorbid personality disorder. While one small study of elderly patients with bipolar I disorder has suggested a 65 percent prevalence of personality disorder in elderly patients with bipolar I disorder, this finding requires further validation. Like patients with major depressive disorder, elderly patients with bipolar I disorder and a personality disorder were more symptomatic and had poorer social adjustment than patients without a personality disorder.

CLINICAL FEATURES AND COURSE

In DSM-IV the personality disorders are grouped into three clusters based on descriptive similarities. Cluster A consists of odd or eccentric personalities including paranoid, schizoid, and schizotypal personality disorders. Cluster B consists of dramatic, emotional, and erratic personalities including antisocial, borderline, histrionic, and narcissistic personality disorders. Cluster C consists of anxious and fearful personalities including the avoidant, dependent, and obsessive-compulsive personality disorders.

Cluster A While individuals with Cluster A disorders do not appear to become more or less symptomatic from youth to midlife, a number of biological and psychosocial factors may place these patients at risk for becoming more symptomatic after age 60. Cluster A disorders have been identified as risk factors for development of late-onset schizophrenia; the proportion of individuals with Cluster A disorders who develop schizophrenia is not known. Common health problems in the elderly population, such as hearing loss, visual problems, or dementia, may exacerbate underlying tendencies in these individuals to be suspicious or paranoid. Finally, individuals with Cluster A personality disorders typically have significant difficulty tolerating intimacy with others, often reacting with hostility and suspiciousness. These individuals usually have very few supports in the community. If their physical health declines, professional caregivers may be upset by the patient's tendency to respond in an idiosyncratic or suspicious manner or to reject assistance altogether. Further difficulty may ensue if caregivers must provide personal care (e.g., helping the individual bathe) that these patients perceive as an invasion of their personal space. When faced with such activities, the patient may respond with agitation or even frank delusions. Institutional life, with its lack of privacy and forced intimacy, is often especially problematic for these individuals.

> Mr. H. is an 84-year-old single man who, neighbors reported, had always been a loner. Mr. H. worked at a filling station, dressed oddly, and would gruffly say hello when others initiated contact. Neighbors knew of no friends or family contacts. His only previous difficulties had occurred when children in the neighborhood would intrude on his fenced yard; he would become irate and chase them off. His neighbors had notified town officials when they noticed he was no longer taking care of his lawn. When a caseworker came to investigate, Mr. H. refused to allow him to enter, saying that all was well and that he didn't need any help. When the caseworker persisted, Mr. H. became agitated, yelling that the government didn't have any business interfering with his life. On his next visit, the caseworker arrived with a police escort. Mr. H. reluctantly allowed them to enter. The house was filled to the ceiling with newspapers that Mr. H. had saved for the last 20 years. Narrow paths led to the kitchen and bathroom, both of which were in serious disrepair. When asked why he kept the newspapers, Mr. H. replied that he believed if he didn't keep the newspapers he would “lose his luck.” Mr. H. was also found to have severely infected ulcers on his feet, which made it very difficult to walk or care for the house. Mr. H. refused the help offered by the caseworker. Because of the serious health concerns and the presence of a schizotypal personality disorder, a probate court judge granted a petition for a conservator of person and estate. While Mr. H. was hospitalized to treat his medical problems, the conservator authorized repairs to the house. Because Mr. H. continued to refuse entry, housekeeping services were provided by giving the workers a key to the house. With this support, Mr. H. has been able to continue to live in the community.

Cluster B Individuals with Cluster B disorders generally demonstrate improvement, at least through middle age, in the overt, high-energy, impulsivity that is a feature of these disorders. Antisocial personality disorder decreases from a point prevalence of 2 percent in younger community-dwelling individuals to 0.8 percent after age 65. Antisocial personality disorder was diagnosed in 4 percent of a group of male patients older than 65 who were treated in a primary care setting; the high

prevalence of antisocial personality disorder in this population was linked to health problems caused by substance abuse. Follow-up studies of patients hospitalized with borderline personality disorder suggest that these patients may have considerable improvement in both interpersonal and occupational functioning, especially in the second decade after hospitalization; however, after age 60 functioning may again begin to decline.

The core difficulties of individuals with Cluster B disorders include inability to tolerate or modulate strong emotion, fear of abandonment and of being alone, and real or fantasized narcissistic injury. Life events common to aging including bereavement, separation from family or friends, changes in physical appearance, overt health problems, and loss of social or occupational roles that may lead to a resurgence of previous pathological coping mechanisms. These coping mechanisms may be expressed differently by elderly and younger Cluster B patients. Impulsivity in an elderly patient may be manifested by taking unsafe risks like trying to walk shortly after a hip operation or abusing prescription drugs. Identity disturbance may be manifested by an inability to formulate future plans or pursue goal-directed activities. At other times coping mechanisms will be familiar to those who treat younger Cluster B patients. Settings such as prolonged inpatient medical hospitalization or nursing home placement are especially at risk for disruption because of the intense emotions likely to be engendered and the multiple staff working with the patient. Patients with Cluster B disorders may denigrate one group of staff or strongly express how disappointed they are with the quality of the care being provided, while cultivating care from other staff members.

Several studies have investigated the relationship between Cluster B disorders and attempted or completed suicide. Premorbid personality disorders were found in 76 percent of a mixed-age group of completed suicides at psychological autopsy; Cluster B disorders accounted for 82 percent of the personality disorders diagnosed. Patients meeting criteria for more than one personality disorder and those with personality disorder and a mood disorder were at especially high risk for completed suicide. Recent life events, especially interpersonal and financial difficulties judged to result from the individual's own behavior, were more common in a group of mixed-age suicide victims with personality disorders than in a group of age- and sex-matched controls without personality disorders. The association between life events and suicide in patients with Cluster B disorders may be mediated by impulsive reactions to stressful life events. Impulsivity was highly associated with the number of lifetime suicide attempts in a mixed-age population of patients diagnosed with borderline personality disorder.

Mrs. A. was an 80-year-old widowed female who entered a nursing home after developing unstable angina. She reported having had a “perfect life,” always living in the best neighborhoods, having beautiful clothes and jewelry, and having an adoring husband and children. Her children reported that she was emotionally aloof and always demanded “the best,” causing the family to live well beyond their means. While she often talked about her dissatisfaction with having to enter a nursing home, she was pleased that this home was “the best in the country.” She was very critical of certain staff she considered insufficiently attentive to her needs, referring to them as “the stupid ones.” Despite these difficulties she managed relatively well until her private funds were exhausted and she had to move to a semiprivate room. She then became very irritable and depressed, wondering “How am I supposed to live with another person? I've never had to share anything.” She felt that staff tended to the needs of her roommate immediately and forced her to wait for care. She attributed this to the roommate having a “special connection” with someone important at the home, even though there was no evidence this was true. Mrs. A. would loudly complain about having to deal with the physical problems of the roommate while the roommate was in the room. Mrs. A. believed the roommate envied her because of the close relationship she enjoyed with her children. Mrs. A. was quite dissatisfied with the professional staff, feeling they did not have sufficient power to provide her with a private room; she requested a change to the physician who was also the medical director at the nursing home. A change in roommate and treatment with psychotherapy, antidepressants, and antipsychotic drugs were minimally helpful. After finding her at the entrance telling visitors her view of conditions in the nursing home, administrators permitted Mrs. A. to move to a private room. Mrs. A.'s narcissistic need to be recognized as a person worthy of a private room was fulfilled, and the difficulties ended.

Cluster C Avoidant and dependent personality disorder may be more prevalent in older populations than in younger ones, though small sample size and methodological issues severely limit these studies. Moreover, other investigators have posited that the social withdrawal and difficulty with decision making that often accompany mild cognitive impairment from neurological illness may account for the apparent increased prevalence of Cluster C disorders.

Cluster C disorders may also disrupt the elderly patient's ability to interact with the health care system. Individuals with avoidant personality may interpret professional advice as criticism of their behavior and so avoid follow-up. Elderly patients with dependent personality may indefinitely defer making decisions on medical care or exhibit marked passivity in their relationship with health professionals; interestingly, this passive dependent stance may facilitate their adjustment to institutional life. Individuals with obsessive-compulsive personality may have extreme difficulty managing inactivity or changes in their routine caused by convalescence from surgical interventions or medical disability.

Ms. F. was a 77-year-old female who lived with an older sister and her husband. She had never married because she was always uncertain whether she had found the right person. Ms. F. had relied on her older sister to help her make most decisions in her life, even those involving what to wear or what food would be best for her to eat in restaurants. She had worked in relatively menial clerical positions even though she had been offered promotions. After retirement Ms. F. and her sister volunteered at a infant and child care center where the sister would feed and play with the children while directing Ms. F. to clean and change the infants. After her sister died suddenly, Ms. F. became very anxious and fearful about being left alone. Her relationship with her sister's husband, always somewhat conflicted because of her competition with him for her sister's attention, became increasingly problematic. She began to ask his advice on all decisions and to accompany him whenever he went out. Because he had promised his wife he would care for her sister if she were to die, he felt obligated to keep Ms. F. in the house. However, he frequently became irritable and verbally abusive toward her. She became increasingly depressed, anorexic, and withdrawn, rarely leaving her room. Another relative became concerned about Ms. F. and initiated medical consultation. Treatment with an antidepressant helped relieve her depression and anorexia, but she continued to constantly ask her brother-in-law for advice. After some counseling he was able to make the decision to place Ms. F. in a boarding home. She agreed with his decision and felt reassured after meeting the manager of the boarding house, who reminded Ms. F. of her sister.

DIFFERENTIAL DIAGNOSIS

When considering the diagnosis of a personality disorder in an elderly individual, clinicians must attempt to document enduring difficulties typical of the disorder since early adulthood. Failure to do so should prompt the clinician to consider alternative diagnoses. Even if these enduring patterns are reported by the patient or other informants, the clinician must consider whether other disorders may be exacerbating the personality disorder at this time.

Clinicians should carefully assess the elderly patient for medical, neurological, or Axis I psychiatric syndromes that may cause misdiagnosis of a personality disorder. Perceptual disorders such as hearing or vision loss may significantly worsen Cluster A disorders. Medications can produce erratic, volatile, disinhibited behavior typical of Cluster B disorders or the withdrawal and passivity typical of some Cluster C disorders. Personality changes are common in cortical and subcortical dementias and may precede cognitive impairment. Mood and anxiety disorders, including subsyndromal disorders common in elderly patients with medical disorders, should be evaluated closely, as they may mimic personality disorder symptoms or be comorbid conditions with significant added disability. Substance abuse, including prescription drug abuse (increasingly recognized as a major problem in the elderly population) could either be a symptom of a personality disorder or cause behaviors mistaken for a personality disorder, including the suspicious, hostile presentation of many Cluster A disorders.

TREATMENT

Patients presenting with symptoms consistent with a personality disorder should have a full medical evaluation with special attention paid to the neurological and cognitive examinations. Any medical or Axis I disorder should be fully evaluated and treated with somatic therapies before attempts are made to specifically treat the Axis II disorder.

Treatment plans should involve all medical, family, and community supports involved in the patient's care, especially with Cluster B patients, with whom miscommunication can contribute to disrupted care plan implementation. Limited and realistic goals should be set and agreed upon by the patient and collaterals. The treatment plan should be reviewed frequently, especially at the initiation of treatment.

A psychotherapeutic relationship may be useful if the patient can establish a therapeutic alliance. By focusing on the patient's present reality, this relationship may help address the splitting, interpersonal manipulativeness, and denial of unpleasant realities common in patients with personality disorders. At times conjoint meetings with family or caregivers may be useful to clarify aspects of the treatment plan or to moderate the effect of the patient on others; however the therapist must be explicit about issues of confidentiality when interacting with the patient's larger social network.

Pharmacotherapy for elderly patients with personality disorders may be useful, although data supporting the use of medications are largely extrapolated from studies

with younger patients. Antipsychotic drugs have been shown to be superior to placebo in treating the psychotic-spectrum symptoms in borderline personality disorder. A trial with an antidepressant should be considered for patients with prominent mood symptoms. In a group of mixed-age patients with major depressive disorder, fluoxetine (Prozac) significantly reduced the frequency of individual personality diagnoses after 8 weeks of treatment. Lithium (Eskalith, Lithobid) or an anticonvulsant may be indicated for patients with prominent irritability or mood lability. Clinicians considering a medication trial should proceed with caution, given the potential for adverse effects and the lack of data establishing the efficacy of psychopharmacological treatment of personality disorders in the elderly.

While treatment cannot be expected to resolve the sequelae of a lifetime of difficult interpersonal relationships, problematic social functioning, and intrapsychic distress, treatment may facilitate improved relationships with significant others, maximize function, minimize environmental disruption, and reduce the distress the individual feels while contending with the vicissitudes of aging.

SUGGESTED CROSS-REFERENCES

Personality disorders in general are discussed in Chapter 24. Mood disorders are discussed in Chapter 14 and Section 51.3d. The topic of psychotherapy and psychopharmacological treatment of the elderly is discussed in Section 51.4.

SECTION REFERENCES

Abrams RC: Management. In *Principles and Practice of Geriatric Psychiatry*, JRM Copeland, MT Abou-Saleh, DG Blazer, editors. Wiley, New York, 1994.

*Abrams RC, Rosendahl E, Card D, Alexopoulos GS: Personality disorder correlates of late and early onset depression. J Am Geriatr Soc *42:*727, 1994.

Abrams RC, Spielman LA, Alexopoulos GS, Klausner E: Personality disorder symptoms and functioning in elderly depressed patients. Am J Geriatr Psychiatry *6:*24, 1998.

Ames A, Molinari V: Prevalence of personality disorders in community-living elderly. J Geriatr Psychiatry Neurol *7:*189, 1994.

Barry K, Fleming M, Manwell L, Copeland L: Conduct disorder and antisocial personality in adult primary care patients. J Fam Pract *45:*151, 1997.

Brodsky B, Malone K, Ellis S, Dulit R, Mann J: Characteristics of borderline personality disorder associated with suicidal behavior. Am J Psychiatry *154:*1715, 1997.

Carpenter D, Clarkin JF, Glick ID, Wilner PJ: Personality pathology among married adults with bipolar disorder. J Affect Disord *34:*269, 1995.

Cheng A, Mann A, Chan K: Personality disorder and suicide. A case-control study. Br J Psychiatry *170:*441, 1997.

Falk B, Segal DL: Personality disorders. In *Handbook of Clinical Geropsychology*, M Hersen, VB Van Hasselt, editors. Plenum, New York, 1998.

Fava M, Alpert JE, Borus JS, Nierenberg AA, Pava JA, Rosenbaum JF: Patterns of personality disorder comorbidity in early-onset versus late-onset major depression. Am J Psychiatry *153:*1308, 1996.

Fava M, Bouffides E, Pava JA, McCarthy MK, Steingard RJ, Rosenbaum JF: Personality disorder comorbidity with major depression and response to fluoxetine treatment. Psychother Psychosom *62:*160, 1994.

Fogel BS, Sadavoy J: Somatoform and personality disorders. In *Comprehensive Review of Geriatric Psychiatry*. J Sadavoy, LW Lazarus, LF Jarvik, editors. American Psychiatric Press, Washington, DC, 1996.

*Golomb M, Fava M, Abraham M, Rosenbaum JF: The relationship between age and personality disorders in depressed outpatients. J Nerv Ment Dis *183:*43, 1995.

Heikkinen M, Henriksson M, Isometsa E, Marttunen M, Aro H, Lonnqvist J: Recent life events and suicide in personality disorders. J Nerv Ment Dis *185:*373, 1997.

Kunik ME, Mulsant BH, Rifai AH, Sweet RA, Pasternak R, Zubenko GS: Diagnostic rate of comorbid personality disorder in elderly psychiatric inpatients. Am J Psychiatry *151:*603, 1994.

McGlashan TH: The Chestnut Lodge Follow-up Study III. Long-term outcome of borderline personalities. Arch Gen Psychiatry *43:*20, 1986.

Molinari V, Ames A, Essa M: Prevalence of personality disorders in two geropsychiatric inpatient units. J Geriatr Psychiatry Neurol *7:*209, 1994.

*Molinari V, Marmion J: Relationship between affective disorders and axis II diagnoses in geropsychiatric patients. J Geriatr Psychiatry Neurol *8:*61, 1995.

Monfort JC: The difficult elderly patient: Curable hostile depression or personality disorder? Int Psychogeriatr *7*(Suppl):95, 1995.

Patience DA, McGuire RJ, Scott AI, Freeman CP: The Edinburgh Primary Care Depression Study: Personality disorder and outcome. Br J Psychiatry *167:*324, 1995.

Quinton D, Gulliver L, Rutter M: A 15–20 year follow-up of adult psychiatric patients. Psychiatric disorder and social functioning. Br J Psychiatry *167:*315, 1995.

Rosowsky E, Gurian B: Borderline personality in late life. Int Psychogeriatr *3:*39, 1991.

Sato T, Sakado K, Sato S, Morikawa T: Cluster A personality disorder: A marker of worse treatment outcome of major depression? Psychiatry Res *53:*153, 1994.

Sato T, Sakado K, Uehara T, Sato S: Personality disorder diagnoses using DSM-III-R in a Japanese clinical sample with major depression. Acta Psychiatr Scand *95:*451, 1997.

*Sibert T, Swartz M: Aetiology and genetics. In *Principles and Practice of Geriatric Psychiatry*, JRM Copeland, MT Abou-Saleh, DG Blazer, editors. Wiley, New York, 1994.

*Vaillant GE: *Adaptation to Life*. Little, Brown, Boston, 1977.

Vine RG, Steingart AB: Personality disorder in the elderly depressed. Can J Psychiatry *39:*392, 1994.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.3 PSYCHIATRIC DISORDERS OF LATE LIFE

51.3H DRUG AND ALCOHOL ABUSE

EVE J. WISEMAN, M.D.

DRUG ABUSE

Most elderly drug abusers appear to prefer legal substances used abusively to illicit drugs. Nicotine dependence is the most common substance use disorder in the elderly population in the United States. Elderly people also abuse over-the-counter and prescription medication, especially medications with sedative, hypnotic, anxiolytic, or analgesic properties. Courses of benzodiazepine therapy are considerably longer in elderly persons than in younger ones, and prescription of benzodiazepines for the institutionalized elderly population is widespread. Illicit substance abuse among elderly persons could become a future public health problem, based on the increased prevalence of illicit substance use in current middle-aged people.

Epidemiology Elderly persons are the largest users of legal drugs in the United States, and they tend to use many drugs.

The U.S. Department of Health and Human Services reported that 20.2 percent of men and 18 percent of women between age 65 and 74 were current smokers. Elderly persons in several studies averaged more than four prescriptions and three over-the-counter medications per person. One community program for persons aged 60 or older found prescription drug abuse in 5 percent of cases. Almost two thirds of those identified started abusing prescription drugs before age 65 (early onset), and more than 90 percent had abused prescription drugs for longer than 5 years. They tended to abuse diazepam (Valium), codeine, meprobamate (Miltown), and flurazepam (Dalmane).

The Drug Abuse Warning Network collected data on drug abuse–related emergency episodes (overdoses, unexpected reactions, and chronic effects). Acetaminophen (Tylenol), alprazolam (Xanax), diazepam, and aspirin were the drugs most often reported in suicide attempts. People age 50 and older represented 13 percent of the suicide cases involving benzodiazepines.

Some elderly persons abuse illicit drugs. People who meet the lifetime criteria for alcohol dependence are most likely to have a diagnosis of illicit drug dependence. Elderly abusers of cocaine or heroin are also more likely to be cigarette smokers. Increasing age appears to be associated with a decreased prevalence of opioid abuse and dependence, a phenomenon termed “maturing out.” However, many persons remain opioid dependent for 50 years or longer.

Etiology Altered pharmacokinetics and pharmacodynamics place elderly persons at risk for more frequent and more severe adverse drug reactions. The relative safety of over-the-counter drugs is often established by testing younger people, leaving the elderly population vulnerable.

Advertising of over-the-counter products, the cost of health care, and reclassification of prescription drugs all contribute to drug abuse by elderly persons. A declining percentage of elderly people who abstain from alcohol may reflect an increased willingness to use and abuse drugs, and elderly persons may depend on drugs as a way to solve problems. Benzodiazepines are often prescribed for long-term anxiety and insomnia, and analgesics are prescribed for chronic pain. Illicit drugs may be available even to the institutionalized elderly population: employees of some nursing homes have been caught selling marijuana and cocaine to the more cognitively intact residents.

Diagnosis The clinician's index of suspicion must be high. When medications considered unnecessary or harmful are discontinued (often in an attempt to fulfill federal regulations regarding prescribing), withdrawal reactions can occur. Caffeine intoxication and withdrawal have been associated with irritability in elderly persons, as has nicotine withdrawal. These conditions may first become apparent after admission to a hospital or nursing home.

The phenomenon of sundowning (increased agitation or actual delirium at night) can result from benzodiazepine, anticholinergic, or other drug toxicity. Benzodiazepines produce sedation or paradoxical excitement during intoxication and anxiety during withdrawal. Depending on the drug's half-life, withdrawal symptoms and signs appear within hours or more than a week after cessation of drug use. Benzodiazepine withdrawal may be misdiagnosed as myocardial infarction, hypertensive crisis, or infection.

Severe overdose involving benzodiazepines can be associated with loss of consciousness, respiratory depression, hypotension, and eventual cardiac arrest. Urine and serum drug screens usually confirm the diagnosis. Family members need to list or bring the clinician the patient's medications, including over-the-counter products.

Course Smoking increases the risk of several cancers, vascular disease, chronic obstructive lung disease, peptic ulcer disease, and osteoporosis. Elderly people who continue to smoke also have higher rates of psychiatric symptoms and syndromes. People with previous inpatient treatment for a nonnicotine addiction have an increased cumulative mortality as they age, largely from tobacco-related causes.

The inability of elderly people to manage their own medications frequently causes nursing home admissions, emergency room visits, and hospitalizations. Medications are also implicated in hip fractures and long-term disability (e.g., sedative-, hypnotic-, or anxiolytic-induced persisting dementia).

Over-the-counter drugs can provoke serious complications in elderly persons. Elderly patients with mild dementia have developed delirium after taking a single 25-mg dose of diphenhydramine (Benadryl) for sleep. The combination of diphenhydramine and other medications with anticholinergic adverse effects can cause urinary retention, fecal impaction, confusion, and delirium. The use of ibuprofen (Motrin, Advil) can lead to gastric hemorrhage.

Tolerance can develop from benzodiazepine use, with the concomitant risks of withdrawal. Patients who overdose with benzodiazepines may require intubation, and aspiration pneumonia is a frequent complication. Other complications of benzodiazepine overdose include pressure injury to the skin and muscles, hypoxic brain damage, and irreversible cardiac arrest. Even when administered in therapeutic dosages, benzodiazepines may produce increased sedation and psychomotor impairment in elderly persons. Patients with dementia are more likely to develop intoxication and impaired physiological functioning at lower dosages of sedative, hypnotic, and anxiolytic drugs.

In recent research, cigarette smoking among community-dwelling older persons was associated with albuminemia or progressively lower albumin concentrations. Avoiding or quitting smoking may reduce the likelihood of becoming hypoalbuminemic, which is commonly regarded as a marker of primary protein-energy undernutrition. Cigarette smoking may also be a marker of malnutrition among older persons.

Treatment Cigarette smokers age 50 years and older have often smoked for 40 or more years. Thus, older smokers are usually long-term smokers, and they also tend to smoke more cigarettes per day compared to younger smokers. Although older smokers frequently receive advice to quit during medical visits, they consistently express attitudes unfavorable to smoking cessation. They generally do not think that smoking affects their health, that there is a strong possibility of serious health problems from smoking, that smoking is addictive, that smoke from other people's cigarettes is annoying, or that lung cancer is a terminal illness. Older smokers with realistic health consequences of smoking and those who perceive smoking as addictive may be more likely to be ready to quit. The effectiveness of smoking cessation programs might be improved by including age-sensitive recruitment strategies, program content, and modes of delivery. More emphasis about the consequences of smoking and benefits of quitting may be needed for older adults.

Smoking-cessation strategies include behavioral therapies (rapid smoking and satiation), cognitive therapies, relaxation training, social support, coping skills training, hypnosis, acupuncture, and group or individual counseling. Elderly patients with heart disease, hypertension, or peptic ulcer disease are probably poor candidates for the rapid smoking approach. Nicotine polacrilex (Nicorette) and transdermal nicotine (NicoDerm CQ, Nicotrol) are available as over-the-counter pharmacological treatments to reduce nicotine withdrawal symptoms. Short-term efficacy has been demonstrated for nicotine replacement systems.

Older adults who frequently attend religious services or engage in private religious activities may be significantly less likely to smoke cigarettes, and if they do smoke, may smoke fewer cigarettes per day than less religiously active older people. Religious groups may model prosocial behaviors, proscribe the use of addictive or physically harmful substances, or offer the person an opportunity to form social bonds with others who share positive beliefs concerning health habits. In addition, older adults who participate in religious activities may cope better with stress and experience less of a need to smoke. Research is needed to determine whether older people who participate in secular social group activity where smoking is prohibited report similar differences in smoking to those reported by older people who participate in religious activities.

Sedative, hypnotic, and other psychoactive medications may affect mobility or increase fall risk in older persons. All prescription and over-the-counter drugs should be reviewed routinely in terms of their risks and benefits, especially in frail individuals who may be at great fall risk because of medication adverse effects. Discontinuing unnecessary or ineffective medication may reduce the number of falls among the elderly.

Hospitalization may be necessary to treat acute benzodiazepine-related toxicity and to prevent withdrawal. Flumazenil (Romazicon), a specific benzodiazepine antagonist, may be appropriate for patients with benzodiazepine overdose. Patients with recent overdose frequently require gastric evacuation, lavage, and administration of activated charcoal to prevent further absorption. The airway must be protected during these procedures since elderly patients are at high risk for aspiration. Intensive monitoring for 24 to 48 hours and dialysis may also be necessary.

Gradual benzodiazepine discontinuation (over 4 to 8 weeks) is indicated for benzodiazepine-dependent patients who have not overdosed. Gradual reduction of the benzodiazepine dosage prevents the occurrence of seizures and attenuates withdrawal symptoms. Elderly patients may have more difficulty completing benzodiazepine withdrawal protocols and an increased likelihood of relapse despite completion of a cessation program.

Elderly opioid abusers may benefit from clonidine (Catapres) treatment during inpatient withdrawal. Clonidine reduces withdrawal severity when it is given in gradually decreasing dosages over 2 weeks. However, clonidine has potentially significant hypotensive and sedative effects and has not been specifically studied in elderly patients, so it should be used with caution.

Elderly persons who require opioid medications to control pain may meet criteria for the diagnosis of opioid dependence. They may be at risk to develop opioid withdrawal following either abrupt dosage reduction or administration of an antagonist and should be warned to avoid abrupt discontinuation of the drug. A tapering schedule should be used if treatment cessation is indicated, and opioid antagonist drugs (including mixed opioid agonist—antagonists) should be avoided. The medical use of opioids is very rarely associated with use to achieve euphoria, unsanctioned dose escalation, acquisition of the drug from multiple providers, or use of illicit drugs. Some elderly patients who experience unrelieved pain may manifest drug-seeking behaviors that cease once pain is relieved. Misunderstanding of this phenomenon may lead the clinician to inappropriately stigmatize the patient and to inadequately treat the patient's pain.

The patient and family should be educated regarding the potential complications of drug abuse. Involvement of pharmacists and home health care nurses may resolve medication management issues, and abusers of illicit drugs can be referred to self-help groups such as Narcotics Anonymous.

Pain management, biofeedback, and other nonmedication interventions may be appropriate. The physician should discourage elderly drug abuse by limiting prescriptions for benzodiazepines and analgesics and by regularly monitoring the patient's use of prescription and nonprescription drugs.

ALCOHOL ABUSE

Elderly people with alcoholism are not consistently identified. Chronic, heavy use of alcohol often results in impairments that are not concomitants of normal aging and are reversible with abstinence. Treatment of alcohol abuse and dependence in elderly persons may improve neuropsychiatric function and prevent alcohol-induced persisting dementia.

Epidemiology According to a recent investigation, lifetime prevalence rates for alcohol dependence in people 65 and older were 14 percent for men and 1.5 percent for women. However, these data excluded the homeless, nursing home residents, and hospitalized patients. Higher prevalence rates of alcohol abuse have been noted in elderly medical and psychiatric patients than in the general community. One study of substance abuse in a psychiatric evaluation and admissions unit reported a 40 percent prevalence rate of alcohol abuse in schizophrenic patients and a 65 percent rate in those with major mood disorders. About one fifth of the dual-diagnosis patients were aged 50 or older.

Elderly persons may have abused alcohol for a lifetime, intermittently, or only after a certain age (late onset, usually defined as onset after age 40). Those who meet the criteria for late onset represent 10 percent of the elderly admitted for alcohol treatment and are more likely to be women than men.

Etiology Alcohol is distributed to a smaller volume of blood with aging, so older people may experience problems without altering the amount of alcohol consumed. The aging brain may be more sensitive to the effects of alcohol because of loss of brain mass or impaired cortical blood flow. Inherited factors may regulate the development of tolerance to alcohol as well as susceptibility to the neurotoxic effects of alcohol. Some investigators have speculated that alcohol-related brain changes may contribute to the risk of relapse.

Particular classes of life events may interact with characteristics of the elderly individual and result in increased drinking. Many elderly persons do not classify alcohol as a potentially harmful drug and do not realize that combination of alcohol with medication can produce deleterious effects. Nursing homes may offer happy hours or use wine or beer as appetite stimulants.

Diagnosis Elderly persons with alcohol dependence may display less prominent alcohol use (quantity and frequency of alcohol used), less antisocial behavior, and less impulsivity than younger persons with alcohol dependence. However, older persons with alcohol dependence usually have more levels of cognitive dysfunction, including acquired intellectual impairment, impaired sensorium, self-neglect, and poor concentration. Cognitive impairment in older persons with alcohol dependence may not result solely from a longer duration of alcohol use. In a comparison of older and younger persons with a similar duration of alcohol dependence, the older group still exhibited different cognitive profiles than the younger group, especially in tests of cognitive flexibility.

Older community-dwelling persons with alcohol dependence may utilize health care inappropriately. They seldom seek treatment for alcohol dependence or treatment by a primary care physician, despite frequent emergency room visits and hospital admissions. Alcohol dependence among rural elderly women may be especially underdiagnosed without a thorough and ongoing assessment.

Collaborative sources of information and such indicators as housing problems, falls or accidents, poor nutrition, and inadequate self-care may facilitate the diagnosis. Often family members attribute alcohol-related confusion to advancing age. Because there are no validated instruments for alcoholism screening in cognitively impaired elderly patients, evaluation might include a search for alcoholic beverages at a home visit. The patient may have mental status changes, peripheral neuropathy, ataxia, and elevated values on liver function tests. Computed tomography (CT) and magnetic resonance imaging (MRI) may show cortical shrinkage and ventricular dilatation, which are difficult to separate from aging effects. Positron emission tomography (PET) may show reduced cerebral metabolism, and neuropsychological deficits have been correlated with frontal hypometabolism in PET studies of elderly patients with alcohol dependence. Single photon emission computed tomography (SPECT) scans may reveal perfusion patterns typical of comorbid problems, including head injury, human immunodeficiency virus (HIV) encephalopathy, dementia of the Alzheimer's type, or depression. Brain samples at autopsy may reveal widespread neuronal loss, degeneration of the mamillary bodies, and deficits of cholinergic and aminergic function. Wernicke's encephalopathy is often first recognized at autopsy.

Alcohol withdrawal delirium (delirium tremens [DTs]), Wernicke's encephalopathy, and Korsakoff's syndrome (alcohol-induced persisting amnestic disorder) always require prompt recognition and treatment, especially in frail elderly patients. DTs occur frequently in elderly alcoholic patients, especially in those with a previous history of DTs. The syndrome usually presents 24 to 72 hours after the last drink, but it may occur after as long as a week of abstinence. Characteristically, the patient

is tremulous with elevated vital signs (fever, tachycardia, hypertension) and diaphoresis. Mental status changes include wakefulness, agitation, and hallucinations.

Patients with Wernicke's encephalopathy present with ophthalmoplegia and ataxia, followed by mental status changes. Visual difficulties may include paresis or paralysis of the external ocular muscles with associated nystagmus and dysconjugate gaze. The patient may appear sleepy, confused, and lethargic. If untreated, Wernicke's encephalopathy can evolve into Korsakoff's syndrome.

Korsakoff's syndrome has an insidious onset and a long-term course. The patient is typically alert and responsive, with a profound memory deficit. Amnesia associated with Korsakoff's syndrome has both retrograde (inability to recall the past) and anterograde (inability to retain new information) components. Although the thiamine deficiency causing Wernicke's encephalopathy is probably related to Korsakoff's syndrome, alcohol may also have a direct, toxic role in Korsakoff's syndrome.

Alcohol-induced persisting dementia is generally a diagnosis of last resort. Before the diagnosis can be determined, patients must undergo (1) careful detoxification, (2) serial evaluation of cognitive function after detoxification, (3) nutritional support during detoxification and rehabilitation, and (4) long-term follow-up of cognitive status at regular intervals.

Course Alcohol use disorders in elderly persons are associated with increased mortality from accident, suicide, and illness. Alcohol use is a risk factor for unsafe driving in elderly adults. In a group of elderly psychiatric inpatients with major depression, 13 percent of recent suicide attempters had a diagnosis of alcohol abuse, compared with 2 percent of nonattempters. Medical problems related to alcohol include cirrhosis, one of the leading causes of death in people 65 and older. Alcohol ingestion also increases the likelihood of gastritis, already common in the elderly population, and the concomitant risk of gastrointestinal hemorrhage. Since people with alcohol dependence obtain as much as half of their calories from ethanol, serious deficiencies of protein, thiamine, folate, and niacin can develop.

When DTs occur, it usually lasts 2 to 3 days. Some patients have intermittent episodes of delirium for up to 2 weeks after drinking cessation, and a minority die from DTs. If Wernicke's encephalopathy is identified, ocular palsies improve within hours of treatment and normalize within days or weeks. Confusion from Wernicke's encephalopathy may resolve within a week or persist for as long as 2 months. Most patients with Korsakoff's syndrome show at least a partial recovery of memory with treatment.

As little as one drink a day may increase cognitive impairment in frail elderly persons. Elderly patients with alcoholism tend to have difficulties with memory, information processing, visuospatial analysis, nonverbal abstraction, and set flexibility. Age is the most important risk factor for alcohol-induced persisting dementia; however, prognosis improves with abstinence.

Treatment Hospitalization is usually necessary to detoxify elderly alcohol abusers. Parenteral thiamine administration helps prevent Wernicke's encephalopathy. When the diagnosis of Wernicke's encephalopathy is suspected, the patient should receive 50 to 100 mg of thiamine intravenously before any dextrose-containing solution is administered. Otherwise, carbohydrate metabolism will deplete whatever marginal stores of thiamine exist. Similar daily dosages of thiamine should be administered intramuscularly to all alcoholic patients until a normal diet can be tolerated; then the thiamine should be administered orally.

Patients with DTs require 24-hour monitoring to optimize their medical status and prevent dangerous behavior. Hourly administration of a medication cross-tolerant with alcohol may be required to control agitation. Elderly patients undergoing less-complicated withdrawal may also be confused and must be monitored for wandering and self-care deficits. Short-acting and intermediate-acting benzodiazepines, at one half to one third the middle-aged requirement, prevent withdrawal in elderly patients. A typical geriatric patient with alcoholism may be withdrawn using oxazepam (Serax) 15 to 30 mg up to four times a day and then tapered over a week to 10 days.

The elderly patient may also need hydration, correction of electrolyte imbalances, and nutritional supplementation. An initial workup should include the history, a physical examination, a blood alcohol concentration determination, a drug screening, glucose and electrolyte tests, tests of renal function, liver function tests, an amylase test, a complete blood count with a differential white blood count, a urinalysis, a chest X-ray, an electrocardiogram, and tests for calcium, magnesium, albumin, total protein, and stool guaiac. Brain imaging is essential for patients with a history of falls and neurological findings, since subdural hematomas are frequently diagnosed in elderly patients. Patients with persistent cognitive impairment also need brain imaging, as well as an evaluation of thyroid functions, syphilis screening, vitamin B_{12} and folate concentrations, and possibly assays for heavy metals and HIV.

Coexisting disorders should be addressed with nonaddicting medications. An elderly patient must be abstinent 2 to 3 weeks before a diagnosis of depression, anxiety, or persistent cognitive impairment can be established. Improvement in alcohol-induced memory impairment has been noted when patients were treated with serotonin-specific antidepressants or antidepressants that increase serotonin reuptake. Naltrexone (ReVia) can be used to reduce relapse in conjunction with psychosocial therapy. Naltrexone is contraindicated in opioid-dependent patients and in patients with acute hepatitis or liver failure.

The patient and family should be educated regarding the prognosis of continued drinking. Rehabilitation may include psychotherapy and Alcoholics Anonymous. Some studies have indicated more-favorable treatment outcomes for older patients in groups with similarly aged peers than in mixed-age groups. Programs emphasizing social relationships and positive aspects of an older patient's life have also demonstrated better outcomes than programs using confrontation and focusing on past failures.

Difficulties in getting older persons treatment may include patient denial or refusal, lack of transportation to treatment services, a paucity of local services, or a poor fit with existing services. Treatment programs may be targeted to a younger, healthier, better-educated, and more mobile population. Elderly persons with concomitant medical conditions may be excluded from programs focusing on psychosocial issues.

Although hospitalization is usually recommended for elderly persons undergoing detoxification from alcohol, some may be appropriately detoxified in the outpatient setting. These include older persons with no history of DTs or seizures, no comorbid problem requiring an inpatient level of care, mild to moderate symptoms of withdrawal, a negative breath or blood alcohol level, and adequate social support (assistance with transporation and alcohol-free housing). The preferred detoxification protocol in outpatient settings has been a standard course of benzodiazepine, which is gradually tapered and discontinued. However, older patients may be overmedicated by a standard course of benzodiazepine. Symptom-triggered detoxification, which incorporates a standardized withdrawal scale to determine benzodiazepine requirements during alcohol withdrawal, may be the more appropriate protocol for older persons. This protocol individualizes pharmacotherapy, enabling the clinician to limit the benzodiazepine dosage to the smallest effective amount in older patients.

SUGGESTED CROSS-REFERENCES

Substance-related disorders in the general population are discussed in Chapter 11. Cognitive disorders are discussed in Chapter 10. Neuroimaging is discussed in Section 1.15, Section 1.16, Section 2.13, Section 51.2e, and Section 51.2f.

SECTION REFERENCES

Adams EH: Prevalence of prescription drug abuse: Data from the National Institute on Drug Abuse. NY State J Med *91:*32S, 1991.

Adams KA, Gilman S, Koeppe RA, Kluin KJ, Brunberg JA, Dede D, Berent S, Kroll PD: Neuropsychological deficits are correlated with frontal hypometabolism in positron emission tomography studies of older alcoholic patients. Alcohol Clin Exp Res *17:*205, 1993.

*American Medical Association Council on Scientific Affairs: Alcoholism in the elderly. JAMA *275:*797, 1996.

*American Psychiatric Association: Practice guidelines for the treatment of patients with substance use disorders: Alcohol, cocaine, opioids. Am J Psychiatry *152S:*33, 1995.

*Atkinson RM, Ganzini L: Substance abuse. In *Textbook of Geriatric Neuropsychiatry*, CE Coffey, JL Cummings, editors. American Psychiatric Press, Washington, DC, 1994.

Barry PP: Chemical dependency. In *Principles of Geriatric Medicine and Gerontology*, ed 3, WR Hazzard, EL Bierman, JP Blass, WH Ettinger, JB Halter, R Andres, editors. McGraw-Hill, New York, 1994.

*Beresford TP, Gordis E: Alcoholism and the elderly patient. In *Oxford Textbook of Geriatric Medicine*, JG Evans, TF Williams, editors. Oxford University Press, Oxford, 1992.

Clark MA, Rakowski W, Kviz FJ, Hogan JW: Age and stage of readiness for smoking cessation. J Gerontol *52:*S212, 1997.

Closser MH, Blow FC: Special populations: Women, ethnic minorities, and the elderly. Psychiatr Clin North Am *16:*199, 1993.

Cornelius JR, Fabrega H, Maher PJ, Jones-Barlock A, Salloum IM, Ulrich RF, Mezzich JE: Age effects on the clinical presentation of alcoholics at a psychiatric hospital. Comprehensive Psychiatry *38:*213, 1997.

Diamond I, Messing RO: Neurologic effects of alcoholism. West J Med *161:*279, 1994.

*Franklin JE, Frances RJ: Alcohol-induced organic mental disorders. In *Textbook of Neuropsychiatry*, SC Yudofsky, RE Hales, editors. American Psychiatric Press, Washington, DC, 1992.

Gambert SR: Substance abuse in the elderly. In *Substance Abuse: A Comprehensive Textbook*, JH Lowinson, P Ruiz, RB Millman, editors. Williams & Wilkins, Baltimore, 1992.

George MS: The clinical use of SPECT in depressive disorders. In Difficult Differential Diagnoses in Psychiatry: The Clinical Use of SPECT, MA Schuckit, chairperson. J Clin Psychiatry *56:*539, 1995.

Glass TA, Prigerson H, Kasl SV, Mendes de Leon CF: The effects of negative life events on alcohol consumption among older men and women. J Gerontol Soc Sci *50:*S205, 1995.

Hurt RD, Offord KP, Croghan IT, Gomez-Dahl L, Kottke TE, Morse RM, Melton LJ: Mortality following inpatient addictions treatment. JAMA *275:*1097, 1996.

Jinks MJ, Raschko RR: A profile of alcohol and prescription drug abuse in a high-risk community-based elderly population. DICP Ann Pharmacother *24:*971, 1990.

Jobst KA: SPECT in the diagnosis of dementia, p 539. In Difficult Differential Diagnoses in Psychiatry: The Clinical Use of SPECT, MA Schuckit, chairperson. J Clin Psychiatry *56:*539, 1995.

Khalsa HK, Shaner A, Anglin MD, Wang J: Prevalence of substance abuse in a psychiatric evaluation unit. Drug Alcohol Depend *28:*215, 1991.

Koenig HC, George LK, Cohen HJ, Hays JC, Larson DB, Blazer DG: The relationship between religious activities and cigarette smoking in older adults. J Gerontol *53:*M426, 1998.

Lehman LB, Pilich A, Andrews N: Neurological disorders resulting from alcoholism. Alcohol Health Res World *17:*305, 1993.

Martens F, Koppel C, Ibe K, Wagemann A, Tenczer J: Clinical experience with the benzodiazepine antagonist flumazenil in suspected benzodiazepine or ethanol poisoning. Clin Toxicol *28:*341, 1990.

*Nixon SJ, Phillips JA: Neurocognitive deficits and recovery in chronic alcohol abuse. CNS Spectrums *4:*95, 1999.

*O'Connell MB: Prevention and treatment of osteoporosis in the elderly. Pharmacotherapy *19:*7S, 1999.

Oxman TE: Antidepressants and cognitive impairment in the elderly. J Clin Psychiatry *57:*38, 1996.

Rains VS, Ditzler TF: Alcohol use disorders in cognitively impaired patients referred for geriatric assessment. J Addict Dis *12:*55, 1993.

Reuben DB, Moore AA, Damesyn M, Keeler E, Harrison GG, Greendale GA: Correlates of hypoalbuminemia in community-dwelling older persons. Am J Clin Nutr *66:*38, 1997.

Rifai AH, Mulsant BH, Sweet RA, Pasternak RE, Rosen J, Zubenko GS: A study of elderly suicide attempters admitted to an inpatient psychiatric unit. Am J Geriatr Psychiatry *1:*126, 1993.

Robins LN, Helzer JE, Przybeck TR, Reiger DA: Alcohol disorders in the community: A report from the Epidemiologic Catchment Area. In *Alcoholism: Origins and Outcome*, RM Rose, J Barrett, editors. Raven, New York, 1988.

Solomon K, Manepalli J, Ireland GA, Mahon GM: Alcoholism and prescription drug abuse in the elderly: St. Louis University grand rounds. J Am Geriatr Soc *41:*57, 1993.

*Taylor S, McCracken CFM, Wilson KCM, Copeland JRM: Extent and appropriateness of benzodiazepine use. Results from an elderly urban community. Br J Psychiatry *173:*433, 1999.

Tejera CA, Saravay SM, Goldman E, Gluck L: Diphenhydramine-induced delirium in elderly hospitalized patients with mild dementia. Psychosomatics *35:*399, 1994.

Thibault JM, Maly RC: Recognition and treatment of substance abuse in the elderly. Prim Care *20:*155, 1993.

Tideiksaar R: Medical strategies to reduce fall risk. In *Falling in Old Age: Its Prevention and Management*, ed 2, P Lankas, editor. Springer, New York, 1997.

Van Heertum RL: Placing SPECT in clinical perspective. In Difficult Differential Diagnoses in Psychiatry: The Clinical Use of SPECT, MA Schuckit, chairperson. J Clin Psychiatry *56:*544, 1995.

Wiseman EJ, Souder E: The older driver: A handy tool to assess competence behind the wheel. Geriatrics *51:*36, 1996.

Wiseman EJ, Souder E, O'Sullivan P: Neuropsychological test performance of older and younger patients with alcohol dependence. Clin Gerontol *17:*66, 1997.

Wiseman EJ, Henderson KL, Briggs MJ: Individualized treatment for outpatients withdrawing from alcohol. J Clin Psychiatry *59:*289, 1998.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4A INTRODUCTION AND OVERVIEW

LISSY F. JARVIK, M.D., PH.D. AND FREDDA L. LEITER, M.D.

Psychopharmacology has become a major form of treatment for psychiatric disorders. In the geriatric age cohort, medications are a readily accepted mode of treatment, and psychotropic medications can be prescribed in the primary care clinic, which enables elderly individuals to obtain treatment without leaving the familiar medical setting. Psychiatrists tend to be involved in more complex situations (e.g., patients failing to respond to medications, developing side effects, presenting diagnostic problems), usually providing consultation to the primary care physician.

PSYCHOPHARMACOLOGICAL AND PSYCHOTHERAPEUTIC INTERVENTIONS

While both psychopharmacological and psychotherapeutic interventions have been shown to be clearly superior to placebo, psychopharmacological interventions have some advantages: they are comparatively inexpensive, reimbursement is easier to obtain, and they act relatively quickly. In depression for example, 50 percent remission of symptoms within 3 to 4 weeks is the standard response, and responders show some improvement of symptoms within 2 weeks. Choice of specific antidepressants is not based on efficacy, as clinical trials fail to reveal differences, but on adverse effects. There are few comprehensive studies of adverse effects in older adults, and the lack of standardized reporting and methodology makes the literature difficult to evaluate. A recent meta-analysis concluded that there is a clear need to supplement the existing database with results derived from well-controlled, rigorously designed, prospective research investigations since there is inadequate evidence with regard to clinically meaningful differences in side effects of antidepressant drugs. Nonetheless, there is a much more robust database for psychopharmacological than for psychotherapeutic interventions; since response rates to existing antidepressants hover around 50 to 60 percent, other approaches are clearly needed.

Psychiatrists may use a wide variety of effective treatments in conjunction with medication. Techniques may need to be altered to take into account physical or cognitive conditions or other issues pertinent to the older individual. Flexibility in location of the treatment setting is essential to provide access for the patient. Nonpsychopharmacological interventions include individual psychotherapy, specifically cognitive-behavioral therapy, family therapy, and group therapy.

A current trend in the treatment of psychiatric disorders in older adults is replacement of psychiatrists with other mental health professionals in an attempt to decrease cost for psychosocial treatments. In the future, mental health professionals may be supplemented or replaced by peer counselors and others with limited training in mental health, as a less expensive and more readily available option. The effects of such substitutions on treatment outcome remain to be evaluated.

ELECTROCONVULSIVE THERAPY

Electroconvulsive therapy (ECT) has been controversial in the United States. It can be the most effective treatment option, with the lowest risk of complications for older individuals with comorbid medical conditions likely to produce drug-disease and drug-drug interactions. ECT can provide a rapid response, which is vitally important in seriously ill patients, those at risk due to malnutrition or agitation related to psychiatric illness, and those at high risk for suicide. Nonetheless, ECT remains a politically charged topic. The current use of general anesthesia and medications to prevent musculoskeletal seizures have changed the procedure so that ECT is now generally considered as safe as, if not safer than, medication for use in frail elderly patients.

ANTIPSYCHOTIC MEDICATIONS

Antipsychotic medications used judiciously are of value in the treatment of elderly patients with psychotic symptoms due to schizophrenia, depression, mania, and dementia. The newer antipsychotics now available provide increased options for treatment with the likelihood of fewer adverse effects. Future research directions include studies on newer antipsychotic agents for the elderly population; inclusion of patients with comorbid medical conditions in studies of efficacy and adverse effects; and further research on the course and treatment of long-term adverse effects such as tardive dyskinesia. Reimbursement policies in both the private and public sectors must be examined to make newer antipsychotics easily available to elderly individuals.

ANTIANXIETY AGENTS

Anxiety is the most frequent mental disorder among persons over the age of 65. However, anxiety disorders and symptoms are not as well recognized in older adults as cognitive and mood disorders. Symptoms of anxiety often coexist with depression in older persons, and for these patients, antidepressants (rather than antianxiety agents) constitute the first-line treatment. Many older individuals are given prescriptions for antianxiety agents in the primary care setting, and education of the primary care physician in the diagnosis, course of illness, and treatment of anxiety and mood disorders is essential. According to a recent survey, antidepressants (specifically the selective serotonin reuptake inhibitors [SSRIs]) have already displaced benzodiazepines as the psychoactive agents most frequently prescribed by primary care physicians. The publication of guidelines for the treatment of mental disorders in the primary care setting is an important advance, and outcome measurement has received increasing attention as a means of assessing guideline implementation.

POSTTRAUMATIC STRESS DISORDER

Posttraumatic stress disorder may not be considered and may thus be missed in older individuals. War veterans may have a long history of undiagnosed posttraumatic stress disorder or may manifest symptoms only later in life, after having experienced multiple psychosocial stressors. Elderly individuals from other cultures who have experienced trauma prior to entering the United States (e.g., Holocaust survivors; refugees from Vietnam, Cambodia, or Eastern Europe) may experience posttraumatic stress disorder. In addition, older persons who have survived disasters such as floods, fires, or earthquakes or who have been the victims of crimes may suffer from unrecognized posttraumatic stress disorder. Research concerned with treatment of these individuals is just beginning.

COGNITION

To effectively prescribe psychotropic medication, the psychiatrist must be skilled in the thorough assessment of the older patient, including evaluation of cognition. Cognitive function may be affected by psychiatric conditions such as depression and may improve with appropriate treatment. Conversely, drug treatments may impair cognitive function because of sedation, slowing of cognitive processing, or other psychopharmacological effects. Dementia may coexist with depression, anxiety disorders, or schizophrenia. Accurate diagnosis facilitates the proper selection and monitoring of treatments.

DEMENTIA

The treatment of dementia can only grow in importance as the population ages. Both the cognitive and the behavioral manifestations of dementias must be addressed, as behavioral symptoms cause significant distress in patients and their families and may hasten institutionalization. Since behavioral symptoms may

reflect the emergence of a medical condition such as myocardial infarction or stroke, a careful medical evaluation is the first intervention for demented patients with an acute change in behavior.

Nonpharmacological Interventions for Dementia Behavioral disturbances include sleep, appetite, and sexual disturbances; incontinence; verbal or physical aggressiveness; restlessness; pacing; screaming; or repetitive requests. The psychiatrist should be familiar with nonpharmacological interventions such as adaptation of the environment to the patient's needs; elimination of excess visual and auditory stimuli; maintenance of a regular schedule; verbal reinforcement of positive behaviors; and minimization of negative behaviors with distractions such as music or other activities. Restraints should be avoided because patients may fall; damage skin, joints, or bone; become deconditioned; become incontinent of urine or feces; or have fecal impaction. Further research is necessary on treatment of behavioral symptoms in both community and institutional settings.

Caregivers often play a key role in the treatment of elderly patients. If the caregiver does not fully understand the plan and its rationale, treatment may be unsuccessful. Caregivers may develop burn-out, medical or psychiatric symptoms, or substance use that compromises the patient's care. Multidisciplinary teams can assess caregivers and harness the necessary psychological, medical, financial and other resources.

PSYCHIATRIST AS A HEALTH PROFESSIONAL

The psychiatrist, primary caregiver and subspecialty physicians, and other appropriate health professionals must be in close communication when treating elderly individuals. Otherwise, older adults may fall through the cracks between multiple systems of care and not receive the proper evaluation, treatment, or follow-up. Any one of the health professionals, including the psychiatrist, may at times have to act as patient advocate, and the geriatric psychiatrist may need to ensure access to treatment and continuity of care in addition to providing psychiatric treatments.

SUGGESTED CROSS-REFERENCES

General principles of psychopharmacology are discussed in Section 31.3, antidepressants in Section 51.4c, antianxiety drugs in Section 51.4d, antipsychotic medications in Section 51.4e, antidementia drugs in Section 51.4f, electroconvulsive therapy in Section 31.31, psychosocial treatments in Section 51.4h, individual psychotherapy in Section 51.4i, cognitive-behavioral therapy in Section 51.4j, family therapy in Section 30.5, and group therapy in Section 51.4l.

SECTION REFERENCES

*Gerson S, Belin TR, Kaufman A, Mintz J, Jarvik LF: Pharmacologic and psychologic treatments for depressed older patients: A meta-analysis and overview of recent findings. Harv Rev Psychiatry *7:*1, 1999.

*Pincus HA, Tanielian TL, Marcus SC, Olfson M, Zarin DA, Thompson J, Magno Zito J: Prescribing trends in psychotropic medications: Primary care, psychiatry, and other medical specialties. JAMA *279:*526, 1998.

*Ruskin PE, Talbott JA, editors: *Aging and Posttraumatic Stress Disorder*. American Psychiatric Press, Washington, DC, 1996.

*Swanwick GR: Nonpharmacological treatment of behavioral symptoms. In *Behavioral Complications in Alzheimer's Disease*, BA Lawlor, editor. American Psychiatric Press, Washington, DC, 1995.

*Winograd CH, Jarvik LF: Physician management of the demented patient. J Am Geriatr Soc *34:*295, 1986.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4B PSYCHOPHARMACOLOGY: GENERAL PRINCIPLES

BRUCE G. POLLOCK, M.D., FH.D.

Patients older than age 65 currently represent approximately 13 percent of the U.S. population yet they receive up to 35 percent of all prescription drugs. As many as 20 percent of these patients suffer from multiple health problems and thus require substantial medical treatment. The average older American uses four prescription drugs and an additional four over-the-counter medications, and the typical nursing home resident is taking nine medications. In older patients, the medical complications of pharmacotherapy alone constitute a highly significant, treatable health problem. Adverse reactions to drugs of all types are seven times higher in those aged 70 to 79 years than in those 20 to 29 years old. One sixth of all hospital admissions of patients over the age of 70 have been attributed to adverse drug reactions. By comparison, adverse drug reactions account for only 1 in 35 admissions in the rest of the population. Precise estimates of adverse reactions due specifically to psychotropic drugs are lacking. Nonetheless, psychotropic drugs figure prominently with cardiovascular medications and nonsteroidal anti-inflammatory drugs as a common cause of drug-related problems in the elderly.

Until recently, older subjects, especially those with multiple medical problems, were systematically excluded from clinical drug trials. In 1993, a review of English language publications of any type (including case reports) relevant to treatment of late-life depression identified only 171 patients who were over age 75. An additional limitation in corporate drug trials is that the new entity is usually not compared with a currently recommended treatment standard. For example, amitriptyline (Elavil), which is not considered the antidepressant of choice in older or younger adults, is often compared with new agents in medication trials.

Although elderly patients frequently require lower doses of medications, age alone does not indicate the need for dosage modification. One must consider individual patient characteristics such as cardiovascular or renal function and genetic and acquired differences in body composition or drug metabolism, together with a drug's physicochemical properties. In addition to using rational pharmacotherapy, the physician must form an alliance with patients and their families to ensure adherence to treatment. Noncompliance with therapy is a major problem for psychiatric patients, and this dilemma is exacerbated with age. Age-related health problems combine with physiological changes to increase the probability of adverse effects from medication, which in turn increase the likelihood of noncompliance. Moreover, the complexity of medication regimens, inherent to the multiple illnesses of aged adults, are further complicated by communication difficulties arising from impaired hearing, cognitive impairment, language and cultural differences.

PHARMACOKINETIC CONSIDERATIONS

Pharmacokinetics provides a way of describing and predicting drug concentrations in plasma and various tissues over time. Although single- and multiple-dose pharmacokinetic studies are a standard part of phase I drug development, they are almost exclusively carried out on young, physically healthy, unmedicated males. If pharmacokinetic data are subsequently obtained in aged subjects, they are usually derived from single-dose studies. Single-dose studies are usually not adequate to rule out the possibility of altered pharmacokinetics in elderly adults. This is particularly true if there is systemic illness or interacting medications. Multiple-dose kinetic studies are especially important for high-clearance drugs to detect nonlinear phenomena wherein metabolism is not proportional to dose. Age-associated pharmacokinetic differences may be due to changes in absorption, distribution, metabolism, or elimination of a drug.

Absorption The rates of gastric emptying and intestinal blood flow decrease with age. Most studies evaluating drug absorption, however, have shown no difference in the rate or extent of absorption in geriatric and younger patients. Nevertheless, antacids, high-fiber supplements, and the cholesterol-binding resin cholestyramine (Questran) may significantly diminish the absorption of medications such as tricyclic drugs. Drugs with anticholinergic properties may decrease absorption of levodopa (Larodopa) because of delayed gastric emptying in conjunction with the gastric metabolism of levodopa. The potential gain in compliance achieved by administering medications at the same time should always be weighed against the possibility of drug interactions in the absorption phase.

Distribution The fact that half-life is directly proportional to a drug's apparent volume of distribution is frequently overlooked by clinicians. The proportion of adipose tissue increases from 33 percent in younger women to 48 percent in the old and from 18 to 36 percent in men. For drugs that distribute in body fat (the vast majority of psychotropics), the volume of distribution increases with age, leading to a longer half-life and drug accumulation. For water-soluble drugs such as lithium (Eskalith) or digoxin (Lanoxin), volumes of distribution decrease in older patients, reducing the margin of safety, that is, transient increases of the plasma concentrations of these drugs will not be buffered to the same degree in older adults. The volume of distribution is also affected by the extent of binding to plasma proteins. Age-associated illness may reduce serum albumin concentration and possibly increase a_1 acid glycoprotein. Most psychotropic drugs are transported in the blood extensively bound to these plasma proteins. Drug displacement from plasma proteins was previously of considerable concern because the protein-bound drug represents an enormous, pharmacologically inactive reservoir. It is now recognized that not only is free drug pharmacologically active, but it is also more available for metabolism and tissue distribution. Therefore, an increase in the absolute concentration of free drug (e.g., by reduction in plasma proteins or a displacement interaction) would be immediately buffered, and a new equilibrium rapidly established. Binding-mediated changes alone are now considered unlikely to be clinically important. However, when plasma drug monitoring is used to adjust tricyclic antidepressant or anticonvulsant dosages, the total drug concentrations (free plus protein bound) are usually reported. Although a change in absolute drug concentration caused by a change in protein binding is transient, the new equilibrium in the proportion of bound and free drug may appear to be an alteration in the drug's plasma concentration. In practice, this might be of concern if a severe intercurrent illness increased a $_1$ acid glycoprotein concentration in patients who were well maintained on a monitored medication. In these circumstances, an increase in bound drug might be interpreted as an increase in the drug's plasma concentration. Determining unbound drug concentrations in these situations is useful for lidocaine (Xylocaine), theophylline (Theo-Dur), phenytoin (Dilantin), and digitoxin (Crystodigin).

Metabolism Drug metabolism transforms lipid-soluble compounds into more water-soluble derivatives that can be more effectively eliminated and may be less active pharmacodynamically. Biotransformation of drugs is categorized as either phase I (oxidation or reduction) or phase II (conjugation). A large group of heme-containing enzymes that vary in substrate preferences, collectively called cytochrome P450, is responsible for oxidative (phase 1) drug metabolism. At least 16 human P450 isoenzymes have been identified, each with distinct, although partly overlapping, substrate specificity. As more is learned about both the diversity and specificity of the cytochrome P450 enzymes, a structure is emerging for categorizing potential drug-drug interactions and age-associated changes. Recent evidence suggests that there is not a uniform, age-associated decline in liver metabolism by the cytochrome P450 isozymes. In old adults, however, reductions in hepatic mass and blood flow place greater emphasis on interindividual differences in drug metabolic capacity. These metabolic differences may be genetic or may result from interactions of multiple medications. Isoenzyme specificity suggests that inhibition or induction of a given P450 isoenzyme will affect all drugs metabolized by that specific enzyme. A drug may be a substrate for one cytochrome P450 isoenzyme at a time or simultaneously for more than one. A drug can be both a substrate and a competitive inhibitor for the same isozyme or it can be a noncompetitive inhibitor of a different enzyme.

Cytochrome P450 (CYP) isoenzyme 2D6 (CYP 2D6) is the isozyme specifically responsible for hydroxylation of many antidepressants (e.g., nortriptyline [Pamelor, Aventyl], desipramine [Norpramin, Pertofrane], paroxetine [Paxil], venlafaxine [Effexor]), and several antipsychotics (perphenazine [Trilafon], thioridazine [Mellaril], and risperidone [Risperdal]). Approximately 7 percent of the white population are genetically poor CYP 2D6 metabolizers. To date, studies conducted in older, unmedicated volunteers have not found an age-associated decline in CYP 2D6 activity. Nevertheless, many drugs commonly used by older patients competitively (e.g., perphenazine) or noncompetitively (e.g., quinidine [Quinora]) inhibit this enzyme, leading to significant interactions. Genotype analyses have consistently found an association of mutations impairing CYP 2D6 activity with Parkinson's disease.

CYP 2C19 is responsible for the metabolism of diazepam (Valium), mephenytoin (Mesantoin), omeprazole (Prilosec), and several barbiturates. In addition, the demethylation of tertiary tricyclic antidepressants is also partly dependent on CYP 2C19. The incidence of the poor metabolizer phenotype in a young, white population is approximately 3 percent but was 18 percent in a Japanese sample. Some preliminary evidence suggests that CYP 2C19 functioning declines with age.

The CYP 3A4 isoenzyme metabolizes numerous medications including lidocaine, diltiazem (Cardizem), verapamil (Calan), erythromycin (Eryc), sertraline (Zoloft), and alprazolam (Xanax). Serious toxicity has occurred when the CYP 3A4–mediated clearance of terfenadine (Seldane), astemizole (Hismanal), cisapride (Propulsid), midazolam (Versed), and triazolam (Halcion) is diminished. Clearance of drugs specifically metabolized by CYP 3A4 appears to decline with age; for example, alprazolam and triazolam clearances in elderly adults are only 50 to 80 percent of those in young adults. In particular, evidence indicates that CYP 3A4 activity is greater in younger women than in men and postmenopausal women and that this enzyme has a substantial role in estrogen metabolism. These gender- and age-related pharmacokinetic differences have been demonstrated for several 3A4-type drugs such as nefazodone (Serzone), whose plasma concentrations are substantially higher in older women than in younger women. Unlike CYP 2D6 activity, CYP 3A4 activity may be increased (induced) by other drugs, such as carbamazepine (Tegretol), barbiturates, and steroids.

Tacrine (Cognex), clozapine (Clozaril), fluvoxamine (Luvox), theophylline, and caffeine are metabolized by CYP 1A2, which is deficient in approximately 12 percent of the population, while 20 to 40 percent are rapid metabolizers. CYP 1A2 also contributes to the demethylation of some tertiary tricyclic drugs and readily undergoes induction by cigarette smoking, cruciferous vegetables, charcoaled meats, and certain medications such as omeprazole and phenobarbital (Luminal). Inhibition and induction magnify an already marked, 60-fold variation in CYP 1A2 activity. This metabolic variation mandates the close monitoring of theophylline treatment and may be important in interpreting results from fixed-dose medication trials such as those conducted with tacrine.

Beyond oxidative metabolism, systematic study of phase II metabolism in older patients has been neglected. In general, drugs such as lorazepam (Ativan) or oxazepam (Serax) accumulate less than more lipid-soluble drugs and undergo single-step conjugative elimination. Glucuronidation, however, may not be an immutable process; it is significantly affected by malnutrition. Despite the common assertion that phase II metabolism does not change with age, significant reductions in oxazepam's elimination have been found in those older than 80.

Excretion The average glomerular filtration rate decreases by 50 percent between the ages of 20 and 70. Age alone, however, cannot be used to predict renal function, because at least one third of older patients may maintain relatively stable renal function. Serum creatinine concentration may not appear elevated in the face of declining renal function, since there is often a decline in muscle mass (the major source of creatinine) with age and debilitation. Decreased renal clearance may affect excretion of psychotropic drug metabolites and lithium in older patients. Accumulation of active hydroxylated metabolites of antidepressants such as nortriptyline, bupropion (Wellbutrin), and venlafaxine has been observed in older patients, with uncertain clinical consequences. Risperidone is metabolized by CYP 2D6 to a 9-OH metabolite that possesses dopamine antagonism comparable to that of risperidone. Since this metabolite may accumulate in older patients, unanticipated extrapyramidal adverse effects may occur because of variable renal as well as hepatic function.

The fact that nonsteroidal anti-inflammatory drugs and angiotensin-converting enzyme inhibitors may reduce lithium's clearance (possibly through effects on renal blood flow) deserves particular emphasis given their widespread use in the elderly. An interesting and important example of an age-related interaction putatively mediated by diminished renal clearance is the finding that 1 mg of alprazolam dramatically increased serum digoxin concentrations in those older, but not younger, than age 65. Low-salt diets and the possibility that selective serotonin reuptake inhibitors (SSRIs) may cause inappropriate antidiuretic hormone secretion with concomitant hyponatremia (particularly in older patients) may be additional sources of interactive complications with lithium.

PHARMACODYNAMICS

Underlying differences among patients related to age or illness require careful consideration with regard to pharmacotherapy. As patients age they exhibit a general reduction in homeostatic mechanisms (e.g., postural control, orthostatic circulatory responses, thermoregulation, visceral muscle function, laryngeal reflexes, hypoxic responses, higher cognitive function). This may interfere with the ability to adapt to changes in the environment and may be manifest as adverse drug experiences. For instance, all psychotropic drugs may increase the risk of falls and hip fracture and may increase cognitive impairment. Similarly, reduced b-adrenergic receptor sensitivity contributes to diminished baroreceptor reflex responses, while coronary artery disease, reductions in the heart's intrinsic sinus rate, and damage to conduction pathways predispose to arrhythmias. Orthostatic hypotension is a major risk in elderly adults that mitigates against the use of lower-potency antipsychotic drugs and tricyclic drugs. Orthostatic hypotension may be detected by measuring blood pressure immediately upon, and 2 minutes after standing. One should check for pretreatment orthostatic hypotension before prescribing any drug known to affect blood pressure. Careful assessment is mandatory whenever an antihypertensive drug is combined with a psychotropic drug. Membrane-stabilizing (quinidine-like) properties of some psychotropic drugs may delay cardiac conduction and seriously interact with either underlying pathophysiology or antiarrhythmic medication. Electrocardiographic (ECG) changes have been noted most frequently for the tricyclic drugs and antipsychotics of the diphenylbutylpiperidine (i.e., pimozide [Orap]) and piperidine (e.g., thioridazine) types. The risk of antipsychotic and tricyclic drugs interacting with cardiovascular drug is considerably amplified, since many of these medications intersect both metabolically and dynamically. For example, quinidine and diltiazem not only inhibit cardiac conduction but CYP 2D6 as well. Patients with preexisting conduction disturbance require increased watchfulness. In older patients treated with antipsychotics long term, obesity, poor nutritional status, and increased triglyceride concentrations may contribute additional cardiovascular risk factors.

Specific receptor and neurotransmitter changes associated with senescence, such as reductions in dopamine type 2 (D_2) receptors or acetylcholine content, result in increased sensitivity to D_2 antagonists and antimuscarinic agents. Neuroleptic-induced adverse effects are common and frequently troubling in older patients. In particular, the association of neuroleptic use with significant cognitive impairment, declines in activities of daily living, and significant falls (with concomitant risk of hip fracture) is of grave concern in frail elderly patients. Drug-induced parkinsonism appears to be more common and more persistent after medication withdrawal in older patients than in younger patients. Prospective studies demonstrate a very rapid and high incidence of onset of tardive dyskinesia in older patients who take antipsychotics for relatively brief periods of time (cumulative incidence reported to be 26 to 45 percent over 1 year).

Two groups of drugs cause anticholinergic effects: those used therapeutically for their anticholinergic properties (e.g., oxybutynin [Ditropan], benztropine [Cogentin]), and those with anticholinergic effects that are incidental to their main actions (e.g., amitriptyline). Peripheral manifestations of anticholinergicity include dry mouth, tachycardia, blurring of vision, urinary retention, and constipation. Common difficulties in older patients such as constipation, xerostomia, glaucoma, and urinary retention may be markedly exacerbated by even modestly anticholinergic medications. Tachycardia in patients with preexisting myocardial ischemia is of particular concern, since these effects appear not to be ameliorated by time. Anticholinergic effects on the central nervous system may range from subtle changes such as apparent worsening of depression, mild confusion, and impairment of recent memory to frank delirium.

Older patients require smaller doses of anticoagulants to achieve the same degree of anticoagulation obtained in younger individuals. Although confounded by effects of other medications and hepatic dysfunction, older adults are probably more sensitive to anticoagulants. Pharmacokinetically, fluvoxamine and sertraline increase warfarin (Coumadin) serum concentration. However, increased bleeding times observed when either paroxetine or fluoxetine were combined with warfarin have been attributed to a pharmacodynamic interaction. Concurrent therapy need not be avoided, but close monitoring is advisable whenever an SSRI is combined with warfarin.

In addition to problems that can be anticipated, the fact that most of the data on a new drug's effects were obtained in younger patients means that unexpected adverse effects are not infrequent in older patients. For example, fluoxetine has been associated with problematic weight loss in older patients and with the syndrome of inappropriate antidiuretic hormone secretion (SIADH). While SIADH is a rare adverse effect of many psychotropic medications including tricyclic drugs and all SSRIs, most reports of antidepressant-induced SIADH have involved patients older than 65 years and have implicated fluoxetine. In an analysis of 760 reports of hyponatremia associated with SSRIs, the median time to onset of the hyponatremia was 13 days (range, 3 to 120 days). In older patients, fluoxetine has also been associated with insomnia, nervousness, delirium, apathy, agitation, and parkinsonism or worsening of motor disability in patients suffering from idiopathic parkinsonism. Although all of these unanticipated effects have now been reported with other SSRIs, it is not possible, at present, to determine the differential geriatric adverse-effect profile of the more recently introduced medications, given the much longer and more extensive use of fluoxetine in the United States and the lack of comparative trials.

The greatest threats to successful pharmacotherapy in older patients are patient nonadherence with medication regimens and drug-drug interactions. High-risk medications, including anticoagulants, antihypertensives, and antiarrhythmics (including digitalis) have been consistently emphasized in long-term, comprehensive surveys of adverse drug reactions in older patients. Physicians need to be particularly vigilant and knowledgeable when these high-risk medications are prescribed. As patients age, pharmacodynamic changes and concomitant illness significantly alter their response to medication. Psychiatrists treating older patients should use a set of routinized procedures: (1) identify patient risk factors such as medical status, use of high-risk medications, diet, and alcohol use; (2) take a thorough drug history, involving caregivers and relatives, be sure to include over-the-counter medications, and update patient's record regularly in a prominent place; (3) educate patients and caregivers about their medications and dosing regimens and monitor compliance; (4) be knowledgeable about drug actions and clearance; (5) use the lowest effective dosages and avoid unnecessary polypharmacy, but don't undertreat; and (6) if the patient's physical or cognitive state deteriorates, always suspect an

adverse drug reaction.

SUGGESTED CROSS-REFERENCES

Biological therapies are covered in Chapter 31. Section 51.4c discusses antidepressants and mood stabilizers in the elderly, Section 51.4d anxiolytics in the elderly, Section 51.4e antipsychotics in the elderly, and Section 51.4f drugs for dementia in the elderly.

SECTION REFERENCES

Beard K: Adverse reactions as a cause of hospital admission in the aged. Drugs Aging *2:*356, 1992.

Benet LZ, Kroetz DL, Sheiner LB: Pharmacokinetics: The dynamics of drug absorption, distribution, and elimination. In *The Pharmacological Basis of Therapeutics*, ed 9, JG Hardman, LE Limbard, PB Molinoff, RW Ruddon, AG Gilman, editors. McGraw Hill, New York, 1996.

Benrimoj SI, Langford JH, Bowden MG, Triggs EJ: Switching drug availability from prescription only to over-the-counter status. Are elderly patients at increased risk? Drugs Aging *7:*255, 1995.

*Feinberg M: The problems of anticholinergic adverse effects in older patients. Drugs Aging *3:*335, 1993.

Foglia JP, Pollock BG: Medication compliance in the elderly. Essential Psychopharmacology *1:*243, 1997.

Gershon S, Ferris S, Kennedy J, Kurtz N, Overall J, Pollock BG, Reisberg B, Whitehouse P: Methods for the evaluation of pharmacologic agents in the treatment of cognitive and other deficits in dementia. In *Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines*, RF Prien, DS Robinson, editors. Raven, New York, 1994.

Gomez-Vargas M, Ogawa N: Clinical applications of neurotransmitter-receptor studies in geriatric neuropharmacology. Acta Med Okayama *50:*173, 1996.

*Greenblatt DJ, Harmatz JS, Shader RI: Clinical pharmacokinetics of anxiolytics and hypnotics in the elderly: Therapeutic considerations. Clin Pharmacokinet *21:*165, 1991.

Hardy BG, Shulman KI, Mackenzie SE, Kutcher SP, Silverberg JD: Pharmacokinetics of lithium in the elderly. J Clin Psychopharmacol *7:*153, 1987.

Institute of Medicine: *Pharmacokinetics and Drug Interactions in the Elderly and Special Issues in Elderly African-American Populations*. National Academy Press, Washington, DC, 1997.

Kalmijn S, Feskens EJM, Launer LJ, Kromhout D: Cerebrovascular disease, the apolipoprotein E4 allele, and cognitive decline in a community-based study of elderly men. Stroke *27:*2230, 1996.

Lebowitz BD, Pollock BG, Caligiuri MP, Kluger A, Laghrissi-Thode F: Instrumental measures in geriatric psychopharmacology. Psychopharmacol Bull *31:*641, 1995.

Lindeman RD: Changes in renal function with aging: Implications for treatment. Drugs Aging *2:*423, 1992.

Liu BA, Mittmann N, Knowles SR, Shear NH: Hyponatremia and the syndrome of inappropriate secretion of antidiuretic hormone associated with the use of selective serotonin reuptake inhibitors: A review of spontaneous reports. Can Med Assoc J *155:*519, 1996.

Lou GL, Montgomery PR, Sitar DS: Bioavailability and pharmacokinetic disposition of tacrine in elderly patients with Alzheimer's disease. J Psychiatry Neurosci *21:*334, 1996.

Mulsant BH, Pollock BG, Nebes RD, Hoch CC, Reynolds CF: Depression in Alzheimer's dementia. In *Progress in Alzheimer's Disease and Similar Conditions*, LL Heston, editor. American Psychiatric Press, Washington, DC, 1997.

Nebert DW: Polymorphisms in drug-metabolizing enzymes: What is their clinical relevance and why do they exist. Am J Hum Genet *60:*265, 1997.

Oxman TE: Antidepressants and cognitive impairment in the elderly. J Clin Psychiatry *57:*374, 1996.

Pantuck EJ, Pantuck CB, Kappas A, Conney AH, Anderson KE: Effects of protein and carbohydrate content of diet on drug conjugation. Clin Pharmacol Ther *50:*254, 1991.

Pizzolato G, Chierichetti F, Fabbri M, Cagnin A, Dam M, Ferlin G, Battistin L: Reduced striatal dopamine-receptors in Alzheimer's disease. Neurology *47:*1065, 1996.

*Pollock BG: Recent developments in drug metabolism of relevance to psychiatrists. Harv Rev Psychiatry *2:*204, 1994.

*Pollock BG, Mulsant BH: Clinical use of antipsychotic drugs in the elderly. In *Handbook of Experimental Pharmacology* vol 120: *Antipsychotics*, JG Csernansky, editor. Springer, New York, 1996.

Pollock BG, Peril J, Everett G: Comparative cardiotoxicity of nortriptyline and its isomeric 10-hydroxymetabolites. Neuropsychopharmacology *6:*1, 1992.

Ray WA, Griffin MR, Avorn J: Evaluating drugs after their approval for clinical use. N Engl J Med *329:*2029, 1993.

Rolan PE: Plasma protein binding displacement interactions—Why are they still regarded as clinically important. Br J Clin Pharmacol *37:*125, 1994.

Salzman C, Schneider LS, Lebowitz B: Antidepressant treatment of very old patients. Am J Geriatr Psychiatry *1:*21, 1993.

Schneider LS: Pharmacologic considerations in the treatment of late-life depression. Am J Geriatr Psychiatry *4:*S51, 1996.

Smith CM: Clinical geriatric psychopharmacology—drugs and neuropsychiatric symptoms. J Clin Pharmacol *36:*683, 1996.

Stewart RB, Cooper JW: Polypharmacy in the aged. Drugs Aging *4:*449, 1994.

Sweet RA, Pollock BG: New atypical antipsychotics: Experience and utility in the elderly. Drugs Aging *12:*115, 1998.

Thapa PB, Gideon P, Cost TW, Milam AB, Ray WA: Antidepressants and the risk of falls among nursing-home residents. N Engl J Med *339:*875, 1998.

Tune L, Carr S, Hoag E, Cooper T: Anticholinergic effects of drugs commonly prescribed for the elderly: Potential means for assessing risk of delirium. Am J Psychiatry *149:*388, 1992.

Tumer N, Scarpace PJ, Lowenthal DT: Geriatric pharmacology: Basic and clinical considerations. Annu Rev Pharmacol Toxicol *32:*271, 1992.

*von Moltke LL, Greenblatt DJ, Shader RI: Clinical pharmacokinetics of antidepressants in the elderly: Therapeutic implications. Clin Pharmacokinet *24:*141, 1993.

Willcox SM, Himmelstein DU, Woolhandler S: Inappropriate drug prescribing for the community-dwelling elderly. JAMA *272:*292, 1994.

*Wylie ME, Mulsant BH, Pollock BG, Sweet RA, Zubenko GS, Begley AE, Gregor M, Frank E, Reynolds CF, Kupfer DJ: Age at onset in geriatric bipolar disorder—Effects on clinical presentation and treatment outcomes in an inpatient sample. Am J Geriatr Psychiatry *7:*77, 1999.

Young RC: Hydroxylated metabolites of antidepressants. Psychopharmacol Bull *27:*521, 1991.

*Young RC, Kalayam B, Nambudiri DE, Kakuma T, Alexopoulos GS: Brain morphology and response to nortriptyline in geriatric depression. Am J Geriatr Psychiatry 1:147, 1999.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4C PSYCHOPHARMACOLOGY: ANTIDEPRESSANTS AND MOOD STABILIZERS

CHARLES F. REYNOLDS, III, M.D.

GOALS AND RATIONALE FOR THE PHARMACOLOGICAL TREATMENT OF DEPRESSION IN LATER LIFE

The goals of pharmacological treatment of major depressive disorder in later life include not only the decrease and resolution of depressive symptoms but also the prevention of relapse and recurrence. Hence, long-term use of these agents is probably the rule, not the exception. Treatment goals also include restoring psychosocial functioning, relieving excess disability, improving patient compliance with other medical therapies that they might otherwise refuse because of hopelessness or pessimism caused by depression, and easing the adaptation to irreversible losses.

Depression in later life should be treated because it usually persists in the absence of treatment and causes distress to patients and burden to caregivers, leads to excess health care use, and is associated with substance abuse. Treatment of depression in elderly adults improves day-to-day functioning and self-care for maintaining and improving health. Treatment benefits outweigh the risks, and the risks of not treating are great. The risks of no treatment or of inadequate treatment include chronic depression, cognitive impairment, worsening of physical illness, poor compliance with medical treatment, social impairment, reduced quality of life (satisfaction, ability to cope), and suicide. Most elderly individuals who commit suicide have late-onset major depressive disorders. Mood disorders play a greater role in suicides of elderly patients than in those of younger patients; and the group at highest risk for suicide is white males aged 80 to 84 years old. Geriatric depression frequently occurs in a complex medical and psychosocial matrix. The major causes of late-life depression include medical disease and attendant disability, psychosocial factors, and genetic factors. The consequences of depression in later life include excess mortality and suicide, anxiolytic dependence and alcoholism, cognitive impairment, disability, medical symptoms, and increased health care use.

With respect to the distribution of depression diagnoses in elderly adults, major depression is just the tip of the iceberg. Other forms of clinically significant depression include minor depression, dysthymia, brief recurrent depression, and mixed anxiety-depression. The estimates of prevalence vary widely depending upon the clinical setting. Prevalence rates for all forms of late-life depression are highest in cognitively intact nursing home patients (about 60 percent), intermediate in chronically ill primary care patients and hospitalized patients (about 25 percent), and lowest in community residents (about 10 percent). In such settings the clues to depression include help seeking with persistent complaints of pain, headache, fatigue, and insomnia; frequent calls and visits; and high utilization of services. Additional clues to depression in elderly hospitalized patients include delayed recovery, treatment refusal, and problems with discharge. In the nursing home patient, depression should be suspected in the presence of apathy, failure to thrive, agitation, and delayed rehabilitation.

Antidepressants with particular relevance to elderly adults include the secondary amine tricyclic drugs, such as nortriptyline (Aventyl, Pamelor) and desipramine (Norpramine); the selective serotonin reuptake inhibitors (SSRIs), citalopram (Celexa), fluoxetine (Prozac), fluvoxamine (Luvox), paroxetine (Paxil), and sertraline (Zoloft); the monoamine oxidase inhibitors (MAOIs) such as phenelzine (Nardil) and tranylcypromine (Parnate), and miscellaneous agents including bupropion (Wellbutrin), trazodone (Desyrel), nefazodone (Serzone), and venlafaxine (Effexor). Agents considered unsuitable for use in elderly adults because of significant adverse effects include amitriptyline (Elavil), imipramine (Tofranil), and clomipramine (Anafranil). The amount of information available from controlled clinical trials in elderly adults is relatively meager; hence, drug recommendations are extrapolated from small numbers of older patients (mostly under age 75) or from younger and middle-aged adults.

TRICYCLIC DRUGS

The advantages of tricyclic drugs include decades of use, superior efficacy for inpatients with melancholia, generally predictable kinetics, and once-daily dosing. Limitations include orthostasis, constipation, tachycardia, urinary retention, dry mouth, cognitive impairment, delirium, and sedation. In addition, an electrocardiogram is required prior to therapy and may be necessary during therapy.

Therapeutic blood concentrations of 80 to 120 ng/mL of nortriptyline and more than 115 ng/mL of desipramine are to be targeted. Starting dosages of both nortriptyline and desipramine are 10 to 25 mg at bedtime. Steady state occurs in about 5 days. An overall efficacy rate of 80 percent was reported for nortriptyline in combination with interpersonal psychotherapy for elderly patients with recurrent major depression. Similar efficacy rates have been reported for desipramine monotherapy. The use of tricyclic drugs such as nortriptyline or desipramine allows compliance to be verified via monitoring of blood concentrations. At the same time, their use is less straightforward than that of the SSRIs. For example, a wide dosage range necessitating individual titration is characteristic. In frail, medically ill, elderly patients, lower dosages of nortriptyline or desipramine may suffice to yield therapeutic steady-state levels. Even at lower dosages, however, frail, medically ill elderly patients may not tolerate tricyclic drugs.

SELECTIVE SEROTONIN REUPTAKE INHIBITORS

The SSRIs are generally as effective in elderly patients as tertiary tricyclic drugs. Many clinicians prefer newer SSRIs because of their lower adverse-effect burden, which allows more-adequate dosing and longer-term use. SSRIs are associated with less orthostasis than tricyclic drugs, are less likely to cause cognitive impairment, and have reduced anticholinergic and cardiovascular adverse effects. They can be prescribed with once-daily dosing, require limited dosage adjustment, and are safe in overdose. The limitations of these agents include nausea, headache, restlessness, sexual dysfunction, decreased heart rate, occasional insomnia, inappropriate antidiuretic hormone secretion, and variable drug-drug interactions related to inhibition of the P450 enzyme system, which is important to the metabolism of many drugs. A number of antidepressants, SSRIs and nefazodone in particular, are highly metabolized by the cytochrome P450 enzyme system in the liver. Of the 30 or more proteins in the cytochrome P450 (CYP) family, CYP 2D6, 3A4, 1A2, 2C9, and 2C19 are principal enzymes in the oxidative metabolism of antidepressants. Potential drug interactions should always be considered.

A low starting daily dosage of paroxetine (10 mg), sertraline (25 mg), citalopram (10 mg), or fluoxetine (10 mg) should be considered, together with more-gradual dosage increases in elderly patients, particularly those above age 75. Final dosages will range from 20 to 60 mg a day of paroxetine, 50 to 200 mg a day of sertraline, and 10 to 60 mg a day of fluoxetine or of citalopram. Titration is thought to minimize adverse effects. Therapeutic blood concentrations are not established for clinical use and testing is not widely available. A recent placebo-controlled, multisite study of fluoxetine in geriatric depression showed rather low efficacy. Moreover, the work of Steven Roose and Alexander Glassman showed that older cardiac patients with severe melancholic depression respond better to nortriptyline than to fluoxetine. Although SSRIs have fewer cardiac adverse effects, they may be effective only in some types of geriatric depression. This issue requires further investigation in controlled clinical trials because of the meager database available at this time.

MONOAMINE OXIDASE INHIBITORS

The MAOIs have been found effective and safe in depressed patients in their 60s but tend to be underused in elderly adults. Their advantages include few anticholinergic effects and no effects on cardiac conduction. For most elderly patients, problem interactions with food and drugs can be minimized through patient education and collaboration among treating physicians. Limitations in their clinical utility include both hypotension and hypertension, insomnia, and somnolence. The

starting dosages of tranylcypromine and phenelzine are 10 mg a day and 15 mg a day in two doses, respectively. In general, maximum dosages do not typically exceed 40 to 50 mg a day for tranylcypromine and 75 to 90 mg a day for phenelzine. A recent study from the National Institute of Mental Health by Trey Sunderland and coworkers supported the efficacy of high-dosage selegiline (Eldepryl) in treatment-refractory elderly depressed patients.

BUPROPION, NEFAZODONE, AND VENLAFAXINE

Bupropion, nefazodone, and venlafaxine are important additions to the pharmacotherapeutic armamentarium for treating late-life depression, although information on their use in elderly adults is limited. Advantages of bupropion include efficacy equal to that of tricyclic drugs and SSRIs in the treatment of major depressive disorder, fewer adverse effects than with tricyclic drugs, no adverse effects of cognitive impairment or sedation, minimal cardiotoxicity, and relative safety in overdose. Precautions in the use of bupropion include the need to use divided dosing and/or a sustained-release formulation to minimize the propensity to induce seizures and the need to administer the last daily dose well before bedtime to avoid insomnia. In geriatric patients, bupropion may be initially prescribed at 50 to 75 mg twice daily with the dosage gradually increased to approximately 300 mg in two or three divided doses. Drug interactions are generally not a significant problem with bupropion, but it should not be administered concomitantly with an MAOI antidepressant. Bupropion may be helpful in depressed elderly patients with Parkinson's disease.

Nefazodone has several advantages for the treatment of geriatric depression: sleep-promoting and early anxiolytic effects, few adverse effects, safety in overdosage, and absence of sexual adverse effects. Higher dosages (300 to 500 mg a day) are required for clinical efficacy and twice-daily dosing is required, which may limit compliance. Its sedating property may make it beneficial or less acceptable, depending upon the patient.

Venlafaxine selectively inhibits both noradrenaline and serotonin reuptake and reportedly has efficacy in adult treatment-refractory patients with major depression. Patients respond to a relatively wide range of daily doses, and no dosage adjustment is necessary based upon age. The use of venlafaxine entails several precautions. Nausea may be minimized by beginning with lower dosages of 25 mg daily and gradually titrating up to a therapeutic dosage. Blood pressure must be monitored because higher dosages (in general, above 225 mg a day) can be associated with sustained increases in blood pressure. The drug must be administered in two to three divided doses a day, which may hinder compliance. A sustained-release formulation has recently become available.

ESSENTIALS OF PATIENT MANAGEMENT

Managing the depressed geriatric patient includes educating the patient and family about the nature of the illness and the fact that it is treatable. Patients and their families need to understand that treatment takes time (4 to 12 weeks), that treatment adherence is critical, and that adverse effects can be expected but are amenable to supportive countermeasures. Weekly contact during the early phases of short-term treatment is recommended for individualizing treatment, providing encouragement and support, reviewing questions and concerns, and monitoring safety (including suicidal potential). Rating scales may be useful for documenting patient response. Common errors in practice include underdosing, discontinuing treatment too soon, failure to consider possible drug interactions, and polypharmacy. If a patient has not begun to respond by 6 to 8 weeks, compliance should be reevaluated, dosage or plasma concentration should be optimized, and referral to a geriatric psychiatrist experienced in treatment resistance should be considered.

One must bring a longitudinal model of depression to treatment. Short-term treatment is intended to induce remission. Incomplete remissions do occur and promote relapse. Recurrences after a longer period of recovery are also common. Treatment should be maintained for at least 6 months after remission from an uncomplicated first episode of major depression and for longer periods after a second or third episode. Many older patients have chronic depression that requires indefinite maintenance therapy. Long-term prevention of new episodes is best achieved by maintaining patients on the same dosage of antidepressant medication that was used successfully to treat the acute episode. The Pittsburgh study of maintenance therapies in late-life depression found that recurrence rates over three years was 20 percent in patients maintained on the combination of nortriptyline (80-120 ng/mL) and monthly interpersonal psychotherapy, 43 percent on nortriptyline, a year later, 64 percent on monthly interpersonal psychotherapy, and 90 percent on placebo.

TREATMENT OF PSYCHOTIC DEPRESSION IN LATER LIFE

Psychotic depression is a severe illness with profound depressive symptomatology accompanied by delusions and, less frequently, by hallucinations. Psychotic depression poses a risk for suicide, usually by violent means. For this reason, recognition and treatment of psychotic depression can be life saving. Both young and elderly patients with psychotic depression respond poorly to short-term treatment with tricyclic drugs or placebo. Combinations of tricyclic drugs and antipsychotics appear to be effective in psychotic depression in both controlled studies of younger patients and naturalistic treatment studies of elderly patients. The most frequently prescribed antidepressants are nortriptyline and desipramine at plasma concentrations comparable to those of younger adults (80 to 120 ng/mL for nortriptyline or more than 115 ng/mL for desipramine). Perphenazine (Trilafon) is the antipsychotic used in most studies of psychotic depression, and in studies conducted at the University of Pittsburgh, it appears to be effective at daily doses of approximately 32 mg. It is unclear if elderly patients respond to lower dosages of antipsychotics or whether risperidone (Risperdal) is effective and better tolerated than perphenazine. Because combined drug therapy may not be tolerated by elderly patients, electroconvulsive therapy is frequently the treatment of choice in geriatric psychotic depression.

TREATMENT OF BIPOLAR DISORDERS IN LATER LIFE

Bipolar disorders constitute 5 to 10 percent of the diagnoses of elderly psychiatric inpatients. In principle the management of mania in elderly patients is similar to that in younger patients; however, an effort should be made to identify patients with mania induced by drugs or medical or neurological disorders and treat the underlying condition.

Lithium (Eskalith) and anticonvulsants remain the first-line short-term treatment for geriatric mania, although some experts and clinicians feel that valproate (Depakene) should be the first-line treatment. Elderly patients require 30 to 60 percent lower lithium dosages than younger adults to develop equivalent plasma concentrations because age-induced decrements in glomerular filtration rates and renal blood flow reduce lithium excretion. Starting dosages of 75 to 150 mg a day may reduce adverse effects and improve compliance. The half-life of lithium is about 24 to 30 hours in the seventh decade of life. Therefore, steady-state pharmacokinetics are anticipated 5 or more days after stabilization of the daily dosage. Lithium plasma concentrations of 0.8 to 1.2 mEq/L often are not tolerated by elderly manic patients. The onset of lithium action is slow and may require several days or weeks. Lorazepam (Ativan) or low dosages of high-potency antipsychotics may be used in the early treatment of elderly patients who are severely agitated.

Elderly patients may be particularly vulnerable to lithium adverse effects. For example, delirium can develop at lithium concentrations at or even below the therapeutic lithium plasma concentration range. Delirium and cerebellar dysfunction may last for weeks after discontinuation of lithium use. Lithium-induced tremor is more frequent in older patients than in younger patients. Lithium can worsen Parkinson's disease or produce parkinsonism in neurologically intact elderly patients. Lithium can worsen cognitive function, especially in demented patients. Sinoatrial block can be caused by lithium, a risk that is increased by the concurrent use of digitalis and b-adrenergic receptor antagonists (beta-blockers). Elderly patients often received drugs that increase lithium plasma concentrations, including thiazide diuretics, nonsteroidal anti-inflammatory drugs, or angiotensin-converting enzyme inhibitors.

The anticonvulsants carbamazepine (Tegretol) and valproate are effective antimanic agents in younger adults. Case reports suggest that both agents are effective in geriatric mania, including organic mania. Anticonvulsants appear to be effective in both lithium responders and nonresponders. Some evidence suggests that anticonvulsants are more effective than lithium in rapidly cycling patients and in patients with dysphoric mania. Hematological and liver function tests are needed prior to treatment with carbamazepine and valproate. Carbamazepine can cause sedation, confusion, and ataxia in a dose-dependent fashion. Complete blood counts are needed frequently, since carbamazepine can cause leukopenia white blood cell (WBC) counts below 4000/mm ^{3} in approximately 2 percent of cases. Half of the patients who develop carbamazepine-induced leukopenia have a drop in white count within the first 16 days of treatment. Valproate causes leukopenia in 0.4 percent of patients, which is comparable to that with tricyclic drugs.

Continuation and maintenance treatment is necessary in elderly bipolar patients who achieve remission of a manic or depressive episode. Although studies of geriatric populations are lacking, studies of younger patients suggest that lithium, carbamazepine, and valproate are effective in preventing both depressive and manic episodes. For this reason, continuation and maintenance treatment with these drugs should be considered for elderly patients.

TREATMENT OF DEPRESSION ASSOCIATED WITH MEDICAL ILLNESS AND WITH COGNITIVE IMPAIRMENT

The NIH Consensus Conference on the Diagnosis and Treatment of Depression in Late Life stated that “the hallmark of depression in older people is its comorbidity

with medical illness.” Compelling evidence from placebo-controlled clinical trials indicates that antidepressant medications are efficacious in patients with common comorbid medical disorders as diverse as stroke, Parkinson's disease, Alzheimer's disease, cancer, ischemic heart disease, chronic obstructive pulmonary disease, and arthritis. Analogous results have been obtained in placebo-controlled studies of tricyclic drugs in an inpatient rehabilitation unit, where treatment of depression could facilitate recovery from acute illness or injury and in a residential care setting, where patients were characterized by extreme old age and the interacting effects of many of the common disabling medical disorders of later life. However, drug-drug and drug-disease interactions can make older patients highly vulnerable to the adverse effects of tricyclic drugs; approximately 30 percent of a residential care population, average age 85, could not tolerate treatment with nortriptyline at therapeutic plasma concentrations. Thus, careful monitoring of both potential adverse effects and plasma concentrations is necessary to optimize safety of these agents in frail elderly patients.

There is a growing anecdotal literature on the use of psychostimulant medications such as methylphenidate (Ritalin) for the short-term treatment of depressive symptoms in elderly patients with significant medical illness. This is supported by suggestive results from a recent small-scale placebo-controlled crossover study.

In treating depression associated with cognitive impairment, one must recall that cognitive impairment can result from depression or from a separate dementing disorder. In either case the depression should be treated, following the basic principles that apply with older patients. Medication use should be supervised, and caregivers should assess the benefits and monitor for adverse effects of treatment. Patients with dementia of the Alzheimer's type are highly sensitive to increased cognitive impairment from anticholinergic medications. Therefore, antidepressants with the lowest anticholinergic effects are preferred, which include the SSRIs, nortriptyline, bupropion, nefazodone, and venlafaxine.

Depression in elderly adults is widespread, particularly in primary care settings, acute medical and surgical wards, and long-term care settings. Geriatric depression has serious consequences if not recognized and adequately treated, including suicide. Depression in later life can be accurately diagnosed and effectively treated. It typically requires long-term treatment because it tends to be a relapsing illness. Elderly patients benefit as much from treatment as younger patients, although response tends to be slower and more brittle, with higher relapse rates particularly in patients over age 70. Augmentation strategies such as lithium use may be helpful. Both antidepressant medication and psychotherapy have long-term efficacy in prevention of recurrence, with combined treatment being superior to medication alone, psychotherapy alone, or placebo.

SUGGESTED CROSS-REFERENCES

Mood disorders are covered in Chapter 14 and Section 51.3d. Combined psychotherapy-psychopharmacology is discussed in Section 30.12, and biological therapies are discussed in Chapter 31.

SECTION REFERENCES

Alexopoulos GS, Borson S, Charlson ME, Kaz IR, Reynolds CF III: *Diagnosis and Treatment of Late-life Depression: Making a Difference*. Geriatric Psychiatry Alliance speaker's guide. American Association of Geriatric Psychiatry, Bethesda, MD, 1996.

Borson S, McDonald GJ, Gayle T, Deffebach M, Lakshminarayan S, Vantuinen C: Improvement in mood, physical symptoms and function with nortriptyline for depression in patients with chronic obstructive pulmonary disease. Psychosomatics *33:*190, 1992.

Bowden CL, Brugger AM, Swann AC, Calabrese JR, Janicak PG, Petty F, Dilsaver SC, Davis JM, Rush AJ, Small JG, Garza-Trevino ES, Risch SC, Goodnick PJ, Morris JD. Efficacy of divalproex vs lithium and placebo in the treatment of mania. The Depakote Mania Study Group. JAMA *271:*918, 1994.

Bump GM, Reynolds CF, Smith G, Pollock BG, Dew MA, Mazumdar S, Geary M, Houck PR, Kupfer DJ: Accelerating remission in geriatric depression: A pilot study of sleep deprivation. Depress Anxiety *6:*113, 1997.

Buysse DJ, Reynolds CF, Houck PR, Perel JM, Frank E, Begley AE, Mazumdar S, Kupfer DJ: Does lorazepam impair the antidepressant response to nortriptyline and psychotherapy? J Clin Psychiatry *58:*426, 1997.

Cohn CK, Shrivastava R, Mendels J, Cohn JB, Fabre LF, Claghorn JL, Dessain EC, Itil TM, Lautin A. Double-blind, multicenter comparison of sertraline and amitriptyline in elderly depressed patients. J Clin Psychiatry *51*(Suppl):28, 1990.

Dew MA, Reynolds CF, Houck PR, Hall M, Buysse DJ, Frank E, Kupfer DJ: Temporal profiles of the course of depression during treatment: Predictors of pathways toward recovery in the elderly. Arch Gen Psychiatry *54:*1016, 1997.

Dunner DL, Cohn JB, Walshe T III, Cohn CK, Feighner JP, Fieve RR, Halikas JP, Hartford JT, Hearst ED, Settle EC, Frank J, Menolascino FJ, Muller DJ: Two combined, multicenter double-blind studies of paroxetine and doxepine in geriatric patients with major depression. J Clin Psychiatry *53*(Suppl):57, 1992.

Gerner R, Estabrook W, Steuer J, Jarvik L: Treatment of geriatric depression with trazodone, imipramine, and placebo: A double-blind study. J Clin Psychiatry *41:*211, 1980.

Georgotas A, McCue RE, Cooper TB, Nagachandran N, Chang I: How effective and safe is continuation therapy in elderly depressed patients? Factors affecting relapse rate. Arch Gen Psychiatry *435:*929, 1988.

*Georgotas A, McCue RE, Hapsworth W, Freeman E, Kim OM, Wellkowitz J, Chang I, Cooper TB: Comparative efficacy and safety of MAOIs versus TCAs in treating depression in the elderly. Biol Psychiatry *21:*1155, 1986.

Katz IR: Drug treatment of depression in the frail elderly: Discussion of the NIH Consensus Development Conference on the Diagnosis and Treatment of Depression in Late Life. Psychopharmacol Bull *29:*101, 1993.

*Katz IR, Simpson GM, Curlik SM, Parmelee PA, Muhly C: Pharmacologic treatment of major depression for elderly patients in residential care settings. J Clin Psychiatry *51*(Suppl):41, 1990.

Lakshmanan M, Mion LC, Frengley JD: Effective low-dose antidepressant treatment for depressed geriatric rehabilitation patients: A double-blind study. J Am Geriatr Soc *34:*421, 1986.

*Lebowitz BD, Pearson JL, Schneider LS, Reynolds CF, Alexopoulos GS, Bruce ML, Conwell Y, Katz IR, Meyers BS, Morrison MF, Mossey JF, Neiderehe G, Parmelee PA: Diagnosis and treatment of depression in late-life: Consensus statement update. JAMA *279:*1186, 1997.

Lipsin B: Treatment of mania. In *Clinical Geriatric Psychopharmacology*, C Salzman, editor. Williams & Wilkins, Baltimore, 1992.

McFarland BH, Miller MR, Straumford AA: Valproate use in the older manic patient. J Clin Psychiatry *51:*479, 1990.

Meyers BS: Geriatric delusional depression. Clin Geriatr Med *8:*299, 1992.

Nierenberg AA, Feighner JP, Rudolph R, Cole JO, Sullivan J: Venlafaxine for treatment-resistant unipolar depression. J Clin Psychopharmacol *14:*419, 1994.

NIH Consensus Development Panel: Diagnosis and treatment of depression in later life. JAMA *268:*1018, 1992.

Nyth AL, Gottfries CG, Lyby K, Smedegaard-Andersen L, Gylding-Sabroe J, Kristensen M, Refsum HE, Ofsti E, Eriksson S, Syversen S: A controlled, multi-center clinical trial of citalopram and placebo in elderly depressed patients with and without concomitant dementia. Acta Psychiatr Scand *86:*138, 1992.

*Reynolds CF, Alexopoulos GS, Katz IR, Mulsant BH: Treatment of geriatric mood disorders. Curr Rev Mood Anxiety Disord *1:*189, 1997.

Reynolds CF, Frank E, Houck PR, Mazumdar S, Dew MA, Cornes C, Buysse D, Begley A, Kupfer DJ: Which remitted elderly depressed patients remain well with continued interpersonal psychotherapy after discontinuation of antidepressant medication? Am J Psychiatry *154:*958, 1997.

Reynolds CF, Frank E, Kupfer DJ, Thase ME, Perel JM, Mazumdar S, Houck PR: Treatment outcome in recurrent major depression: A post-hoc comparison of elderly and midlife patients. Am J Psychiatry *153:*1288, 1996.

*Reynolds CF, Frank E, Perel JM, Imber SD, Mazumdar S, Kupfer DJ: Maintenance therapies for late-life recurrent major depression: Research and review circa 1995. Int Psychogeriatr *7*(Suppl):27, 1995.

Reynolds CF, Frank E, Perel JM, Mazumdar S, Dew MA, Begley A, Houck PR, Hall M, Mulsant B, Shear MK, Miller MD, Cornes C, Kupfer DJ: High relapse rates after discontinuation of adjunctive medication in elderly patients with recurrent major depression. Am J Psychiatry *152:*1418, 1996.

*Reynolds CF III, Frank E, Perel JM, Imber SD, Cornes C, Miller MD, Mazumdar S, Houck PR, Dew MA, Stack JA, Pollock BG, Kupfer DJ: Nortriptyline and interpersonal psychotherapy as

maintenance therapies for recurrent major depression: A randomized controlled trial in patients older than 59 years. JAMA *281:*39, 1999.

*Reynolds CF, Miller MD, Pasternak RE, Frank E, Perel JM, Cornes C, Houck PR, Mazumdar S, Dew MA, Kupfer DJ: Treatment of bereavement-related major depressive episodes in later life: A controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry *156:*202, 1999.

Roose SP, Glassman AH, Attia E, Woodring S: Comparative efficacy of selective serotonin re-uptake inhibitors and tricyclics in the treatment of melancholia. Am J Psychiatry *151:*1735, 1995.

Salzman C: Pharmacological treatment of depression in elderly patients. In *Diagnosis and Treatment of Depression in Late Life: Results of the NIH Consensus Development Conference*, LS Schneider, CF Reynolds, BD Lebowitz, AJ Friedhoff, editors. American Psychiatric Press, Washington, DC, 1994.

Schneider LS, Reynolds CF, Friedhoff AJ, editors: *Diagnosis and Treatment of Depression in Late Life*. American Psychiatric Press, Washington, DC, 1994.

Sunderland T, Cohen RM, Molchan S, Lawlor BA, Mellow AM, Newhouse PA, Tariot PN, Mueller EA, Murphy DL: High-dose selegiline in treatment-resistant older depressive patients. Arch Gen Psychiatry *51:*607, 1994.

Toleffson GD, Holman SL: Analysis of the Hamilton Depression Rating Scale factors from a double-blind, placebo-controlled trial of fluoxetine in geriatric major depression. Int Clin Psychopharmacol *8:*253, 1993.

Wallace AE, Kofoed LL, West AN: Double-blind, placebo-controlled trial of methylphenidate in older, depressed medically ill patients. Am J Psychiatry *152:*929, 1995.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4D PSYCHOPHARMACOLOGY: ANTIANXIETY DRUGS

JAVAID I. SHEIKH, M.D., M.B.A., AND CYNTHIA T. M. H. NGUYEN, M.D.

Anxiety disorders as a group are the most common psychiatric conditions in late life. Because few controlled studies of the management of anxiety in older patients exist, much of the knowledge of the use of anxiolytics in the elderly is necessarily based on evidence from clinical trials in younger populations and on clinical experience.

Various classes of antianxiety medications have traditionally been used to manage anxiety disorders in late life. These include benzodiazepines, azapirones, tricyclic antidepressants, monoamine oxidase inhibitors (MAOIs), dopamine receptor antagonists, b-adrenergic receptor antagonists, barbiturates, anticonvulsants, and antihistamines. Newer antidepressants such as selective serotonin reuptake inhibitors (SSRIs), selective serotonin-noradrenaline reuptake inhibitors (SNRIs), and serotonin (5-hydroxytryptamine [5-HT]) type 2 (5-HT_2) receptor antagonists have recently emerged as the first-line treatment for several anxiety disorders and mixed anxiety-depression syndromes in younger patients. Although investigations of these medications in the treatment of anxiety in older adults are lacking, clinical experience suggests similar efficacy in late life. Because of their adverse-effect profiles, the newer antidepressants seem to be much better tolerated by older people than the anxiolytic medications used in the past. Some older anxiolytic medications (e.g., barbiturates, chloral hydrate [Noctec]) are now rarely used, while others (e.g., MAOIs, tricyclics, b-adrenergic receptor antagonists, anticonvulsants, neuroleptic compounds) may be used as alternatives in treating patients refractory to the SSRIs, SNRIs, and 5-HT_2 antagonists. Newer medications are being investigated (e.g., bretazenil and abecarnil, partial benzodiazepine receptor agonists) that may prove useful in treating anxiety disorders in older patients. In prescribing any medication, however, the special physiological vulnerabilities of anxious older patients should be borne in mind.

GENERAL PRINCIPLES

Several factors complicate psychopharmacological management of the anxious older patient, including comorbid conditions, polypharmacy, and physiological changes associated with aging. The risks of adverse effects of psychopharmacology are greater in older adults than in younger patients. Older patients are more likely than younger people to have concurrent medical illnesses that may be exacerbated by the adverse effects of medications. Similarly, several medical illnesses can mimic or produce symptoms of anxiety. These potentially treatable physical illnesses need to be sought and ruled out before antianxiety medications are administered.

The average older person typically takes six to eight different medications a day. That is, although elderly persons constitute 12 percent of the population, they are prescribed 25 percent of all medications and as a group are the largest consumers of over-the-counter drugs. Thus, the adverse effects of drug-drug interactions occur more commonly in older patients because the risk of adverse drug interactions increases substantially with polypharmacy.

Age-related changes in pharmacokinetics put older patients at higher risk of adverse drug effects. The most important of these changes are decreased absorption, decreased protein binding, increased volume of distribution of lipophilic drugs, decreased hepatic function, and decreased renal excretion. The overall effect of these changes associated with age is that many medications are eliminated more slowly and metabolized less efficiently in older patients than in younger ones. Chronic kidney, liver, or heart disease can further delay drug clearance.

ANXIOLYTICS

Several classes of compounds are used as anxiolytics. Some medications (e.g., barbiturates, chloral hydrate) used as anxiolytics in the past with higher risks of adverse effects in older patients are omitted.

Benzodiazepines Over the past three decades, benzodiazepines have been the most frequently prescribed anxiolytic agents for both young and older patients. Benzodiazepines remain the mainstay of pharmacological therapy for acute symptomatology of generalized anxiety disorder. In addition, the efficacy of benzodiazepines in panic disorder and obsessive-compulsive disorder has been shown in several studies.

As a general principle, older patients are more sensitive to both the therapeutic and the toxic effects of the benzodiazepines. Low doses are recommended, as higher doses that are therapeutic in younger patients may prove toxic in geriatric patients. These medications are beneficial because they have a rapid onset and little effect on cardiovascular status. However, potential complications of long-term benzodiazepine use in older patients include excessive daytime drowsiness, cognitive impairment and confusion, psychomotor impairment and risk of falls, depression, paradoxical reactions, amnestic syndromes, respiratory problems, abuse and dependence potential, and breakthrough withdrawal reactions. Because of these complications, use of benzodiazepines for more than a few weeks is generally not recommended.

Deciding which benzodiazepine to use depends on half-life, potency, and metabolic pathway. Benzodiazepines undergo stepwise hepatic metabolism. The long half-life benzodiazepines such as chlordiazepoxide (Librium), chlorazepate (Tranxene), diazepam (Valium), and flurazepam (Dalmane) undergo both phase I and phase II reactions. Phase I reactions tend to be prolonged in older patients, leading to drug accumulation. These long half-life benzodiazepines are metabolized via oxidative pathways into active metabolites that stay in the system even longer. Phase II (conjugative metabolism) reactions are not affected by age. Therefore for older patients, short half-life benzodiazepines (e.g., lorazepam [Ativan], oxazepam [Serax], and temazepam [Restoril] are preferred because they require only phase II metabolism and are inactivated by direct conjugation in the liver.

Although benzodiazepines are very similar in efficacy, high-potency benzodiazepines such as clonazepam (Klonopin) and alprazolam (Xanax) are preferable in panic disorder because low-potency ones would be extremely sedating at therapeutic doses. Likewise, clonazepam is preferred in obsessive-compulsive disorder.

The use of benzodiazepines as the primary treatment for anxiety disorders on an ongoing basis is problematic because of the many possible complications of long-term use. However, in clinical situations, it is not unusual to encounter patients who have been on these medications for a prolonged (more than 6-month) period. Since newer antidepressants such as SSRIs, SNRIs, and 5-HT_2 antagonists are increasingly being considered as the first-line treatment in various anxiety disorders, older patients on long-term benzodiazepine therapy should be assessed for appropriateness of treatment with these antidepressants. Benzodiazepines should then be slowly tapered to avoid severe withdrawal symptoms.

Conversely, because of the delayed onset of the therapeutic effects of antidepressants, concurrent use of a benzodiazepine in the first few weeks of treatment may help to alleviate severe symptoms of anxiety. A general approach for a patient with panic disorder, with severe anticipatory anxiety, for example, is to start administering alprazolam at 0.25 mg orally three times a day and titrate up to therapeutic levels as tolerated. Clonazepam is almost as potent as alprazolam, but it has a longer half-life. Older anxious patients can be given as little as 0.25 mg twice a day initially and increased gingerly. After a few weeks, benzodiazepine use is tapered slowly while the patient continues to receive the chosen antidepressant.

In treating patients with a mixed picture of depression and anxiety, which is one of the most frequently encountered presentations in geriatric psychiatric practice, short-term use of a benzodiazepine concurrent with an SSRI and SNRI may be warranted if immediate relief is needed for severe symptoms of anxiety.

Benzodiazepines have moderate success in managing anxiety and agitation associated with dementia but can cause sedation, cognitive worsening, and ataxia, as

well as paradoxical disinhibition and agitation.

A summary of pharmacokinetics and average daily doses in geratric patients of commonly used benzodiazepines can be found in Table 51.4d-1.

Drug	Half-Life (h)	Active Metabolites	Average Daily Dosage (mg)	
			Adult	Elderly
Alprazolam	12–15	Yes (minor)	0.5–6	0.25–3
Chlordiazepoxide	7–28	Yes (many)	25–100	5–50
Clonazepam	18–56	Yes	1–8	0.5–4
Clorazepate	30–200	Yes	15–60	7.5–30
Diazepam	20–60	Yes	5–30	2–15
Halazepam	15–50	Yes	20–160	20–80
Lorazepam	10–20	None	1–6	0.5–3
Oxazepam	5–15	None	15–90	10–45
Prazepam	25–200	Yes	20–60	10–20

Adapted from Sheikh JI: Anxiety disorders. In *Textbook of Geriatric Neuropsychiatry*. CE Coffey, JL Cummings, editors. American Psychiatric Press, Washington, DC, 1994.

Table 51.4d-1 Pharmacokinetics and Average Daily Dosage of Commonly Used Benzodiazepines

Azapirones Buspirone (Buspar), an azapirone derivative and a 5-HT $_{1A}$ partial agonist, has been shown in clinical trials to be as efficacious as benzodiazepines and better than placebo for generalized anxiety disorder. Buspirone is well tolerated and is devoid of the sedation, psychomotor impairment, and potential for abuse seen with the benzodiazepines. Although its adverse-effect profile is suitable for use in geriatric patients, its effectiveness for generalized anxiety disorder remains somewhat inconsistent in clinical situations.

In subsyndromal mixed anxiety-depression, if the patient presents with predominant anxiety, buspirone may be beneficial. Buspirone has also shown modest success in managing anxiety and agitation associated with dementia in some case studies, but controlled investigations are lacking.

Antidepressants Several classes of compounds initially approved for use as antidepressants have increasingly been used as antianxiety drugs during the last decade. Recent studies have provided empirical support for their use for various anxiety disorders.

Tricyclic Drugs Several studies have documented the efficacy of various tricyclic drugs in panic disorder. Imipramine (Tofranil) has been used the longest and has been extensively studied. At least one study has shown a tricyclic drug to be superior to a benzodiazepine for generalized anxiety disorder. Clomipramine (Anafranil) is being studied and may prove efficacious in generalized anxiety disorder. Results of controlled trials show clomipramine to be superior to placebo for treatment of obsessive-compulsive disorder.

Use of tricyclic drugs has been declining, especially in the older patient population, because of the risk of adverse effects (e.g., anticholinergic toxicity, orthostasis, cardiac toxicity, memory impairment, delirium, seizures) and the increased use of the newer antidepressants (SSRIs, SNRIs, and 5-HT $_2$ antagonists) with their more favorable adverse-effect profiles.

Monoamine Oxidase Inhibitors Phenelzine (Nardil) has proved effective in treating panic disorder and agoraphobia. In addition, phenelzine was found to be equal to clomipramine in efficacy for obsessive-compulsive disorder, but adverse effects and dietary restrictions that go along with MAOIs make use of these medications more problematic in older patients. Sudden hypotensive episodes observed in patients on MAOIs can be particularly dangerous in geriatric patients because of the risk of falls.

Selective Serotonin Reuptake Inhibitors Several SSRIs, including citalopram (Celexa), fluoxetine (Prozac), fluvoxamine (Luvox), paroxetine (Paxil), and sertraline (Zoloft), have shown efficacy superior to that of placebo and similar to that of tricyclic drugs for panic disorder in controlled trials.

Given their efficacy and favorable adverse-effect profile, the SSRIs should now be considered the drugs of choice for panic disorder. These medications may be given in a single daily dose to maximize compliance. However there is a 2- to 4-week delay in onset of action, so medication education must be emphasized when prescribing. In addition, the SSRIs, especially fluoxetine, may be activating and may cause increased anxiety. Some anxious geriatric patients who already tend to overemphasize somatic symptoms may find the gastrointestinal distress commonly associated with SSRI use intolerable. Older patients therefore should initially be given these medications in very small daily doses (e.g., a half of a 25-mg tablet of sertraline or a half of a 20-mg tablet of citalopram). Dosage can be increased as tolerated.

Fluoxetine, sertraline, fluvoxamine, and paroxetine have all been found to be superior to placebo for treatment of obsessive-compulsive disorder. Comparisons found fluvoxamine, sertraline, and paroxetine to be equal in efficacy to clomipramine. Most older patients tolerate SSRIs better than clomipramine, so they should be considered first-line medication treatment for obsessive-compulsive disorder in geriatric patients.

The SSRIs are presently being studied in clinical trials for their efficacy in generalized anxiety disorder. In mixed anxiety-depressive disorder, if the patient has a predominant depression, or depressive and anxious symptoms in equal proportions, SSRIs or newer antidepressants (SNRIs, 5-HT $_2$ antagonists) are the treatment of choice. In the management of anxiety and agitation in older patients with dementia, controlled studies on efficacy have shown mixed results.

Other Antidepressants There is emerging evidence that SNRIs such as venlafaxine (Effexor) and 5-HT $_2$ antagonists such as nefazodone (Serzone) may be efficacious for panic disorder and generalized anxiety disorder. Further, nefazodone has also demonstrated efficacy in open-label studies with younger patients suffering from posttraumatic stress disorder. Several case reports show efficacy in treating anxiety and agitation associated with dementia with trazodone or nefazodone.

Antipsychotic Agents Clinical experience suggests that low-dose, high-potency dopamine receptor antagonists such as haloperidol (Haldol) and fluphenazine (Prolixin) (e.g., 0.25 to 0.5 mg twice a day of haloperidol) can be modestly effective in treating anxiety and agitation associated with dementia. These drugs can cause sedation, cognitive worsening, extrapyramidal reactions, or late-developing adverse effects such as tardive dyskinesia. Their use should be limited to acute situations and for brief periods, with frequent evaluations and documentation of need. Serotonin-dopamine antagonists, such as olanzapine (Zyprexa), risperidone (Risperdal), or quetiapine (Seroquel) have a lower incidence of treatment-related adverse effects and may be preferable in the older patient population. Recent evidence from a controlled trial shows that risperidone is superior to placebo in the management of agitation associated with dementia. Neuroleptic drugs are also used as potentiating agents with SSRIs in the management of obsessive-compulsive disorder. Pimozide (Orap) is used as an adjunct in patients with obsessive-compulsive disorder and Tourette's disorder. In older patients, antipsychotic drugs should generally be used in low dosages and with extreme caution.

Anticonvulsants Valproate (Depakene) and carbamazepine (Tegretol) show moderate effectiveness in the treatment of anxiety and agitation associated with dementia. However, carbamazepine has a high risk of potential toxicity and may not be desirable for use in older patients.

b-Adrenergic Receptor Antagonists Some reports suggest that b-adrenergic receptor antagonists may be efficacious in some geriatric patients with anxiety and agitation. Propranolol (Inderal) showed moderate success in several studies, but due to the risk of hypotension and a possible risk of developing congestive cardiac failure, they are rarely used with older patients. Controlled studies in younger patients suggest that b-adrenergic receptor antagonists may have some efficacy in treating autonomic symptoms associated with generalized anxiety disorder and panic disorder but not the cognitive aspects of anxiety.

Antihistamines Antihistamines such as hydroxyzine (Atarax, Vistaril) and diphenhydramine (Benadryl) are sometimes used in clinical settings to manage mild anxiety, in spite of the dearth of studies demonstrating efficacy or safety. Prolonged use of these medications, however, is fraught with potential problems in older patients, such as anticholinergic effects, excessive sedation, confusion, and disorientation.

SUGGESTED CROSS-REFERENCES

The epidemiology of disorders of late life is covered in Section 51.1b. Section 51.4b, Section 51.4c, Section 51.4c, Section 51.4d, Section 51.4e, Section 51.4f, Section 51.4g, Section 51.4h, Section 51.4i, Section 51.4j, Section 51.4k, Section 51.4l, Section 51.5a, Section 51.5b, Section 51.5c, Section 51.5d, Section 51.6a cover psychological treatments in the elderly. Anxiety disorders are extensively discussed in Chapter 15, and anxiety disorders in late life are discussed in Section 51.3c. Biological therapies are covered in Chapter 31.

SECTION REFERENCES

Aguiar LM, Haskins T, Rudolph RL, Pallay A, Derivan AT: Double blind, placebo controlled study of once daily venlafaxine extended release in outpatients with GAD (NR-643). Proceedings of the 151st annual meeting of the American Psychiatric Association, Toronto, Canada, June 1998.

Ballenger JC: Alprazolam in panic disorder and agoraphobia: Results from a multicenter trial. Arch Gen Psychiatry *45:*413, 1988.

Blazer DG: Generalized anxiety disorder and panic disorder in the elderly: A review. Harvard Rev Psychiatry *5:*18, 1997.

*Boyer W: Serotonin uptake inhibitors are superior to imipramine and alprazolam in alleviating panic attacks: A meta-analysis. Int Clin Psychopharmacol *10:*45, 1995.

Coplan JD, Pine DS, Papp LA, Gorman JM: An algorithm-oriented treatment approach for panic disorder. Psychiatr Ann *26:*192, 1996.

Cross-National Collaborative Panic Study, Second Phase Investigators: Drug treatment of panic disorder: Comparative efficacy of alprazolam, imipramine, and placebo. Br J Psychiatry *160:*191, 1992.

*Davidson JR: PTSD: Who responds to what treatment? Symposium #23D conducted at the 151st annual meeting of the American Psychiatric Association, Toronto, Canada, June, 1998.

Deveaugh-Geiss J, the Clomipramine Collaborative Study Group: Clomipramine in the treatment of patients with obsessive-compulsive disorder. Arch Gen Psychiatry *48:*730, 1991.

Geracioti TD Jr: Venlafaxine treatment of panic disorder: A case series. J Clin Psychiatry *56:*408, 1995.

Hertzberg MA, Feldman ME, Beckman JC, Moore SD, Davidson JRT: Open trial of nefazodone for combat-related posttraumatic stress disorder. J Clin Psychiatry *59:*460, 1998.

Hoehn-Saric R, McLeod DR, Zimmerli WD: Differential effects of alprazolam and imipramine in generalized anxiety disorder: Somatic versus psychiatric symptoms. J Clin Psychiatry *49:*293, 1988.

*Katz I, Brecher M, Clyde C: Risperidone in the treatment of psychosis and aggressive behavior in patients with dementia. Presented at the annual scientific meeting of the American College of Neuropsychopharmacology, Hawaii, December 1997.

*Katz I, Jeste D, Mintzer J, Clyde C, Napolitano J, Brecher M: Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: A randomized double-blind trial. J Clin Psychiatry *60:*107, 1999.

Mavissakalian M, Michaelson L: Agoraphobia: Relative and combined effectiveness of therapist-assisted in vivo exposure and imipramine. J Clin Psychiatry *47:*117, 1986.

Napoliello MJ: An interim multicentre report on 7677 anxious geriatric outpatients treated with buspirone. Br J Clin Pract *40:*71, 1986.

Oehrberg S, Christiansen PE, Behnke K, Borup AL, Severin B, Soegaard J, Calberg J, Judge R, Ohrstrom JK, Manniche PM: Paroxetine in the treatment of panic disorder: A randomized, double-blind, placebo-controlled study. Br J Psychiatry *167:*374, 1995.

*Pohl RB, Walkow RM, Clary CM: Sertraline in the treatment of panic disorder: A double-blind multicenter trial. Am J Psychiatry *155:*1189, 1998.

Sheehan DV, Ballenger J, Jacobsen G: Treatment of endogenous anxiety with phobic, hysterical, and hypochondriacal symptoms. Arch Gen Psychiatry *37:*51, 1980.

*Sheikh J, Salzman C: Anxiety in the elderly: Course and treatment. Psychiatr Clin North Am *18:*871, 1995.

*Silverstone PH, Ravindran A: Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. J Clin Psychiatry *60:*22, 1999.

Stokes PE, Holtz A: Fluoxetine tenth anniversary update: The progress continues. Clin Ther *19:*1135, 1997.

Thompson PM: Generalized anxiety disorder treatment algorithm. Psychiatr Ann *26:*227, 1996.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4E PSYCHOPHARMACOLOGY: ANTIPSYCHOTIC DRUGS

DAVID L. SULTZER, M.D., AND HELEN LAVRETSKY, M.D.

The use of antipsychotic medications to treat psychiatric syndromes in the elderly can be clinically challenging. Older patients demonstrate heterogeneous individual responses to antipsychotic medications, are sensitive to adverse effects, often have comorbid medical conditions, and are at high risk for adverse drug interactions. In addition, antipsychotic medications are prescribed to treat a variety of psychiatric syndromes in the geriatric population.

Despite the clinical complexities, antipsychotic medications are frequently prescribed to elderly persons. Most studies find that at least 30 percent of nursing home residents have an antipsychotic medication ordered, frequently without clear evidence of a psychiatric symptom or diagnosis and often with limited monitoring of medication effects over time. Concern about such practices has resulted in federal legislation (Omnibus Budget Reconciliation Act) that requires documentation of the specific indication for neuroleptic treatment in nursing homes.

Until recently, there has been little systematic study of antipsychotic medications in the geriatric population. Dosing strategies, patterns of short-term adverse effects, predictors of clinical response, long-term effectiveness, prophylactic value, and adverse effects resulting from prolonged use are not well understood. Fortunately, studies focusing on these clinical parameters in older patients are beginning to help provide better treatment guidelines.

TREATMENT PRINCIPLES

Prescribe Antipsychotic Medications Only for Specific Clinical Symptoms Thorough clinical evaluation and accurate diagnosis provide the framework for effective treatment. An older patient with several medical conditions, neurological illness, and mixed symptoms that include delusions, dysphoria, anxiety, and behavioral disturbance requires careful evaluation that leads to a clear *Diagnostic and Statistical Manual of Mental Disorders* fourth edition (DSM-IV) differential diagnosis and an individualized treatment plan. Specific target symptoms should be clearly identified prior to antipsychotic treatment and carefully monitored over the course of treatment. Medication intervention for poorly defined behavioral eccentricities often provides limited clinical benefit, obscures the treatment endpoint, and unnecessarily exposes the patient to medication risks.

Dose-Response Relationships Vary Widely Across Individuals Pharmacokinetic and pharmacodynamic changes associated with aging have a prominent impact on the relationship between dose and clinical response in older patients. Reduced protein binding, reduced hepatic metabolism, and alterations in neuroreceptor densities and neurotransmitter concentrations all likely contribute to enhanced antipsychotic medication effects. Studies measuring neuroleptic plasma concentrations indicate that patients over age 60 achieve concentrations in blood that are 2- to 3-fold higher than those of younger patients at a given oral dose, although not all studies agree. There is also considerable heterogeneity in age-related pharmacokinetic and neuroreceptor alterations. Thus, the efficacious and adverse effects of a particular antipsychotic dosage may vary considerably among older patients. In addition, the time required for antipsychotic response is typically longer than in younger patients. Initial dosage should be relatively low and dosage increases should be titrated cautiously to determine the optimal dosage for an individual patient.

Anticipate Adverse Effects Although antipsychotic treatment is usually effective for elderly patients with specific target symptoms, many experience an unwanted consequence of antipsychotic treatment. Older patients are particularly sensitive to anticholinergic, hypotensive, and extrapyramidal effects of neuroleptic drugs because of age-related changes in neurotransmission and reduced homeostatic mechanisms and are at high risk for tardive dyskinesia. Some atypical antipsychotic medications may be efficacious with fewer adverse effects than dopamine receptor antagonists, but there has been limited experience with their use in the elderly population and in those with comorbid medical conditions.

Consider Potential Drug Interactions Approximately 75 percent of community-dwelling elderly persons in the United States take prescription drugs, and each takes an average of three to five different prescription and nonprescription medications. Even more medications are consumed in nursing homes. Such frequent use of multiple medications, coincident with age-associated pharmacokinetic changes, raises the likelihood of additive medication effects and adverse interactions with antipsychotic drugs. Reviewing the specific prescription and over-the-counter medications that a patient is taking, eliminating unnecessary drugs, and reviewing potential interactions prior to prescribing an antipsychotic agent is crucial.

Monitor Medication Effects Over Time and Adjust the Treatment Plan The heterogeneity of clinical responses to antipsychotic treatment that can be expected in the elderly population provides a strong rationale for careful longitudinal monitoring. Some beneficial effects of antipsychotic medication treatment such as improved sleep may occur within a few days, whereas gradual improvement in delusions and thought disorder may occur over 6 to 8 weeks. Similarly, excessive sedation, hypotension, or acute extrapyramidal signs may develop within the first day of neuroleptic treatment, yet rigidity or gait disturbance may emerge after 6 weeks of neuroleptic treatment at a stable dosage. In addition, noncompliance with a medication regimen among older patients has been reported to vary between 40 and 75 percent, which indicates that monitoring medication use is essential.

ANTIPSYCHOTIC MEDICATIONS

Traditional antipsychotic agents, also referred to as neuroleptics or dopamine receptor antagonists, are thought to ameliorate psychosis via blockade of the dopamine type 2 (D_2) receptors in the mesolimbic system. Serotonin-dopamine antagonists (SDAs), also called “atypical antipsychotics,” have been introduced more recently. Although the term “atypical” is variably defined, these medications are generally associated with few extrapyramidal effects, improvement in negative symptoms of schizophrenia (apathy, blunted affect, emotional withdrawal), and prominent interactions with serotonergic receptors in the brain. SDAs include clozapine (Clozaril), risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine (Seroquel). Antipsychotic medications are listed in Table 51.4e-1, with usual dosages in elderly persons, adverse-effect profiles, and relative affinity for dopamine type 2 (D_2) and serotonin (5-hydroxytryptamine [5-HT]) type 2(5-HT_2) neuroreceptors. The dosage ranges are broad, underlining the need for careful titration in individual patients. Optimal dose ranges and adverse effects of the SDAs are not well characterized in the geriatric population, particularly among those with dementia or other neuropsychiatric illness. However, the adverse-effect profiles of some of these agents are encouraging and they are increasingly prescribed for older patients. Age-related pharmacokinetic changes may have a smaller impact on clearance of olanzapine than other antipsychotic agents, although pharmacodynamic alterations of aging warrant use of lower doses of both the SDAs and dopamine receptor antagonists.

Medication	Brand Name	Dosage Range[a] in Elderly Patients (mg/day)	Adverse Effects[b]			Receptor Affinity In Vitro	
			Anticholinergic Effects	Extrapyramidal Effects	Hypotensive Effects	Dopamine Receptor (D_2)	Serotonin Receptor (5-HT_2)
Chlorpromazine	Thorazine	10–300	+++	+	+++	++	++
Thioridazine	Mellaril	10–300	+++	+	+++	++	+
Loxapine	Loxitane	5–75	++	++	++	++	++
Molindone	Moban	5–75	++	++	+	++	+
Perphenazine	Trilafon	2–32	++	++	++	++	+
Thiothixene	Navane	1–15	++	+++	++	+++	+
Fluphenazine	Prolixin	0.5–5	+	+++	+	+++	+
Haloperidol	Haldol	0.5–5	+	+++	+	+++	+
Clozapine[c]	Clozaril	12.5–300	+++	-	+++	+	++
Risperidone[c]	Risperdal	0.5–5	+	+	+	+++	+++
Olanzapine[c]	Zyprexa	2.5–20	++	+	+	++	++
Quetiapine[c]	Seroquel	25–250	+	-	+	+	+

[a] medication effects: -, none; +, mild; ++, moderate; +++, marked.
[b] Routine clinical dose ranges are listed. Higher or lower doses may be indicated in some cases.
[c] Optimal dose ranges and adverse effects listed are based on limited experience.

Table 51.4e-1 Antipsychotic Medications

Outcome studies indicate that the overall efficacy of antipsychotic medications is similar, although SDAs appear to have a beneficial effect on the negative symptoms of schizophrenia and clozapine may be particularly effective in treatment-resistant patients. Previous history of clinical response to a particular agent and an individual's sensitivity to specific adverse effects are the most important factors in choosing a medication. In general, low-potency antipsychotics and multiple antipsychotic medications are avoided. Many older patients tolerate oral antipsychotic doses once a day, although some patients have fewer adverse effects with more-frequent smaller doses and lower peak blood concentrations. Depot preparations of haloperidol (Haldol) or fluphenazine (Prolixin) may benefit some patients with schizophrenia who require maintenance treatment, particularly when compliance is likely to be poor. However, low initial dosages and careful dosage titration are required because of erratic intramuscular absorption and prolonged clearance rates in elderly persons.

Adverse Effects Adverse effects of antipsychotic medications in the elderly are listed in Table 51.4e-2 and the relative frequency of anticholinergic, extrapyramidal, and hypotensive effects among different agents is shown in Table 51.4e-1.

Anticholinergic effects	Cognitive impairment,* delirium,* dry mouth, blurred vision, constipation,* urinary retention,* tachycardia,** exacerbation of narrow-angle glaucoma
Acute extrapyramidal effects	Rigidity,* bradykinesia,* tremor,* dystonia, akathisia
Tardive movement disorders	Tardive dyskinesia,* tardive dystonia, tardive akathisia
Cardiovascular effects	Hypotension,* tachycardia,* electrocardiogram changes (nonspecific *T*-wave changes, increased QT interval)
Other adverse effects	Sedation,* falls,* cognitive impairment,* elevated prolactin concentration, sexual dysfunction, weight gain, neuroleptic malignant syndrome, elevated hepatic enzyme values, jaundice, hyponatremia, seizure, skin photosensitivity, retinopathy, agranulocytosis, nasal congestion

* Effects that are particularly common or severe among older patients, even with treatment at low dosages.

Table 51.4e-2 Adverse Effects of Antipsychotic Medications in Elderly Patients

Anticholinergic Effects Peripheral anticholinergic symptoms such as constipation and urinary retention are particularly common in older patients, as many are bothered by these problems prior to medication intervention. Central anticholinergic symptoms include poor attention, impaired memory, and behavioral disturbances. These symptoms can result in diagnostic confusion and can be particularly debilitating in patients with preexisting cognitive impairment.

Extrapyramidal Effects Older patients are sensitive to acute extrapyramidal effects, which include parkinsonism, dystonia, and akathisia. Several studies have suggested that the geriatric population is more susceptible to neuroleptic-induced tremor and less susceptible to acute dystonia than younger age groups. Older patients with rigidity or tremor prior to medication treatment are most likely to develop neuroleptic-induced parkinsonism. Overall, extrapyramidal signs are evident in up to 80 percent of older patients treated with neuroleptics, they may emerge after months of treatment at a stable dosage, and can lead to difficulty in swallowing and falls. Acute extrapyramidal symptoms do not generally occur with SDAs, although risperidone may induce these symptoms at low dosages (e.g., 1 to 2 mg a day) in older patients, particularly those with dementia. Akathisia due to neuroleptics may be mistaken for increased agitation in a patient with dementia and behavior disturbance, which the clinician may erroneously respond to by raising, rather than lowering, the dose. Treatment with anticholinergic agents such as benztropine (Cogentin) or trihexyphenidyl (Artane) to ameliorate extrapyramidal symptoms is best avoided to reduce risk of anticholinergic toxicity, although low dosages (e.g., benztropine 0.5 mg twice daily) may be prescribed cautiously. Low-dose propranolol (Inderal) may improve akathisia, and amantadine (Symmetrel) may also help reduce rigidity, but these medications can also cause adverse effects.

Tardive Dyskinesia Advanced age is one of the strongest risk factors for development of tardive dyskinesia. A recent study found that the cumulative incidence of tardive dyskinesia among patients over age 45 was 26 percent after 1 year, 52 percent after 2 years, and 60 percent after 3 years of neuroleptic treatment. Older patients may develop tardive movements after a shorter period of neuroleptic exposure than young patients, and the movement disorder remits less frequently after neuroleptic discontinuation. Risk factors for tardive dyskinesia remain incompletely understood, but the research literature indicates that duration of neuroleptic exposure, female sex, mood disorder, medical illness, history of extrapyramidal symptoms or anticholinergic drug use, and subtle movement disorder prior to neuroleptic treatment, in addition to advanced age, confer greater risk. Most of the SDAs are not expected to cause tardive dyskinesia, although longer outcome studies are needed. A few cases of tardive dyskinesia in patients treated with risperidone have been reported.

Cardiovascular Effects The principal cardiovascular effects of antipsychotic medications are hypotension and tachycardia. Orthostatic hypotension is a particular concern in elderly persons because of age-related limitations in vascular regulation and postural reflexes that can increase the risk of falls. Agents that most frequently cause hypotension (Table 51.4e-1) should be avoided, and blood pressure, pulse, and orthostatic symptoms carefully monitored.

Other Adverse Effects Several other adverse effects may occur with antipsychotic treatment (Table 51.4e-2). Elderly persons tend to be susceptible to these because of reduced homeostasis, medical illness, or concurrent medications. SDAs are not generally associated with the adverse motor and anticholinergic effects of traditional neuroleptic drugs, although excessive sedation, hypotension, and weight gain can occur with these agents, and the profile of effects in the geriatric population is not well characterized. Dose-related extrapyramidal symptoms may occur with risperidone treatment, and anticholinergic effects can be prominent with clozapine treatment. Olanzapine may also have anticholinergic effects. Clozapine treatment often causes sedation and hypotension in addition to anticholinergic effects and is associated with a dose-related seizure risk. Potentially fatal agranulocytosis occurs in approximately 1 percent of patients treated with clozapine, which necessitates frequent blood cell monitoring. Clozapine is thus reserved for patients with prominent psychosis who respond poorly to other antipsychotic agents or have severe tardive dyskinesia. Initial dosages should be very low and the dosage increased slowly; many elderly patients respond to less than 100 mg a day.

Medication Interactions Because older patients frequently take multiple medications, adverse medication interactions are an important clinical concern. Medications that often influence antipsychotic medication effects are listed in Table 51.4e-3. The clinical impact of these interactions varies considerably, depending on the extent of pharmacokinetic effects, genetic polymorphism of metabolic enzyme systems, and involvement of single or multiple metabolic pathways. Inhibitory interactions can lower antipsychotic clearance by 50 percent, and inducing interactions may double antipsychotic clearance. Antipsychotic drugs can also influence the effects of a coadministered medication. For example, concentrations of tricyclic drugs in blood may be higher when neuroleptic medications are concurrently taken.

Medication	Impact on Antipsychotic Medication Kinetics or Clinical Effects
Serotonin reuptake inhibitors, tricyclic drugs, monoamine oxidase inhibitors	Increased concentrations in blood due to reduced metabolism
Benzodiazepines, barbiturates	Increased sedation and risk of falls Possible increased hypotension
Lithium	Increased extrapyramidal symptoms Possible increased risk of delirium Possible increased risk of neuroleptic malignant syndrome
Anticholinergic medications	Delayed gastrointestinal absorption
Carbamazepine	Enhanced hematological toxicity with clozapine
Carbamazepine, phenytoin, phenobarbital	Reduced blood level due to enhanced metabolism; less likely for olanzapine and risperidone
Antacids	Delayed gastrointestinal absorption
Cimetidine	Increased concentration in blood due to reduced metabolism Possible decreased gastrointestinal absorption
Erythromycin, chloramphenicol, ketoconazole	Increased concentration in blood due to reduced metabolism
Propranolol	Increased concentration in blood due to reduced metabolism Hypotension
Opioids	Increased sedation and hypotension
Ethanol	Increased sedation
Cigarette smoking	Reduced concentration in blood due to enhanced metabolism

Table 51.4e-3 Antipsychotic Medication Interactions in Elderly Patients

THERAPEUTIC INDICATIONS

Early-Onset Schizophrenia, Now Older The course of illness among patients with typical onset of schizophrenia in the second to fourth decade of life is heterogeneous. Positive symptoms such as delusions and hallucinations as well as symptoms of thought disorder may be less severe among patients over age 65, and antipsychotic dosages are typically lower than those in younger patients. If negative symptoms are prominent, trial of an SDA may be indicated. Some patients who have been maintained for many years on antipsychotic drugs with prolonged remission of psychotic symptoms may warrant a trial off medication. Relapse may occur, however, and predictors of relapse are not well understood. Some studies find older patients and those with later onset of illness to be at lower overall risk for relapse, but careful monitoring is important.

Late-Onset Schizophrenia Approximately 15 percent of patients with schizophrenia develop their first psychotic episode after age 45. New onset of psychotic symptoms in an older person demands careful evaluation for medical, neurological, or pharmacological causes. Some evidence indicates that patients with late-onset schizophrenia may respond to lower neuroleptic dosages than those with early-onset schizophrenia who are now older.

Delusional Disorder Delusional disorder is characterized by nonbizarre delusions without prominent hallucinations, disorganized behavior, or functional deterioration and tends to occur in mid to late adulthood. Delusional thoughts are often circumscribed and may not require intervention, but a trial of antipsychotic medication is indicated when the delusions have a significant adverse impact on social relationships or behavior. Treatment is often complicated by the patients' lack of insight regarding their beliefs or need for treatment.

Mood Disorder With Psychosis Elderly patients with major depression or mania may present with psychotic symptoms. Short-term treatment with antipsychotic medication may be beneficial, although antidepressant or mood-stabilizing pharmacotherapy is primary. Antipsychotic medication should generally be discontinued following remission of psychosis. Electroconvulsive therapy may be indicated if psychosis, appetite disturbance, or mood symptoms are particularly severe.

Dementia and Delirium Delusions, hallucinations, and agitated or aggressive behavior are major sources of morbidity among patients with Alzheimer's disease and other dementias. Approximately 30 to 40 percent of patients with Alzheimer's disease have delusional thoughts, and at least 25 percent of patients exhibit aggressive behavior. Delusions among patients with dementia are often transient and easily modified by suggestion and thus may not warrant pharmacological intervention. However, delusional thoughts of persecution or misidentification may be severe, and treatment is indicated if the patient is significantly distressed or the delusions contribute to a behavior disturbance.

The term “agitation” refers to a broad range of inappropriate verbal or motor behaviors, and pharmacological intervention is again reserved for situations in which distress is prominent or safety or basic care is compromised. Medical problems that cause behavior disturbance in dementia should be identified: dehydration, infection, urinary retention, pain, or adverse medication effects. Behavioral management strategies should be instituted before medication intervention. Adjusting the physical environment, redirecting distressed patients, talking calmly with patients, and promoting appropriate social interactions and activities, combined with caregiver education and support, can significantly reduce problem behaviors and their impact.

Many studies have explored the efficacy of antipsychotic medication treatment for psychosis and agitation in patients with dementia, although methodological flaws limit the conclusions that can be drawn. Most studies indicate that delusions usually respond to antipsychotic treatment. Dosages are low to moderate. Approximately two thirds of patients with agitated behavior improve with neuroleptic treatment, although the response is generally modest, and some patients do not benefit from the medication. No difference in efficacy among neuroleptic drugs has been consistently demonstrated. High-potency agents are recommended and most patients respond to low dosages (e.g., haloperidol 1 to 2 mg a day). Some patients may require only a very low dose (e.g., haloperidol 0.25 to 0.5 mg a day), and adverse effects are often prominent with haloperidol dosages above 2 mg a day. The concentration of neuroleptic drug in blood required to improve agitation appears to be considerably lower than that required to improve schizophrenia. Cognitive impairment and subtle extrapyramidal symptoms prior to treatment indicate greater risk for neuroleptic-induced parkinsonism.

Patients with delirium require careful medical, neurological, and laboratory evaluation to identify the conditions that are contributing to the mental status change. Antipsychotic medications should be prescribed only when necessary to maintain safety and basic care, as these medications can contribute to the mental status changes and cloud the clinical picture.

SDAs may have an important role in managing psychosis or aggression in patients with dementia, because of their favorable adverse-effect profiles and their impact on the serotonergic system, which may mediate aggressive behavior. Although clinical experience is growing, only a few structured trials in mixed populations have been reported. Risperidone has been shown effective, but dosages are lower than those prescribed for elderly patients with schizophrenia, and much lower than dosages for young patients. Initial risperidone dosage is 0.5 mg a day, and the total daily dose rarely exceeds 3 mg. Extrapyramidal symptoms may occur within this dosage range. Clozapine may ameliorate psychosis or aggression in some elderly patients with dementia, although adverse effects can be prominent, even at very low dosages and many patients are unable to tolerate clozapine. A range of 12.5 to 100 mg a day is recommended in this population, with very cautious titration. Newer SDAs, including olanzapine and quetiapine, may provide benefit with fewer adverse effects. Specific effects and guidelines for their use will be better understood with results from structured trials. Trazodone (Desyrel), other serotonergic medications, and anticonvulsants also ameliorate agitated behavior in some patients and are often well tolerated.

There has been little systematic study of long-term antipsychotic medication use by cognitively impaired patients with psychosis or behavior disturbance, despite concern for potentially excessive use. However, several investigators have examined the consequences of neuroleptic drug discontinuation or dosage reduction in elderly nursing home residents. Results indicate that many patients without active psychotic symptoms or specific behavioral disturbances do not deteriorate following antipsychotic reduction, and some patients demonstrate improvement in blunted affect and social withdrawal. However, a significant minority (5 to 45 percent in different studies) do deteriorate clinically, which supports the importance of careful monitoring.

Other Challenging Circumstances Elderly patients with psychosis and parkinsonism due to Parkinson's disease, diffuse Lewy body disease, or cerebrovascular disease involving subcortical structures are often extremely sensitive to extrapyramidal adverse effects of neuroleptic medications. Dopaminergic medications should be reduced to as low a dosage as possible, as these agents frequently contribute to psychosis or behavior disturbance. If antipsychotic intervention is necessary, low dosages of midpotency agents (e.g., loxapine [Loxitane], perphenazine [Trilafon]) may be helpful without marked increase in parkinsonism. Alternatively, SDAs may improve psychosis without exacerbating parkinsonism. Low dosages of clozapine (e.g., 12.5 to 100 mg a day) or olanzapine (2 to 15 mg a day) benefit some patients. Risperidone may exacerbate parkinsonism, even at low dosages. Newer atypical agents offer a distinct theoretical advantage in this population, but clinical experience is limited. Electroconvulsive therapy is often effective for mood and psychotic symptoms in patients with Parkinson's disease and also ameliorates the movement disorder in most patients.

Older patients with psychosis and a long history of neuroleptic treatment often have tardive dyskinesia. The natural history of tardive dyskinesia varies considerably, and continued neuroleptic treatment may not worsen the dyskinesia. However, the lowest effective dosage should be prescribed to reduce the likelihood of all adverse effects, including exacerbation of tardive movements, and anticholinergic medication should be discontinued if possible. A trial of an SDA may be indicated, as the risk for worsening the dyskinesia is likely lower, although longitudinal clinical data are lacking. Interestingly, several cases of improved tardive dyskinesia following clozapine treatment have been reported.

SUGGESTED CROSS-REFERENCES

Section 51.3e and Section 51.3f describe diagnostic considerations, clinical phenomenology, and overall treatment strategies for elderly patients with psychosis due to schizophrenia, delusional disorder, or dementia. Section 51.4b presents the general principles of psychopharmacology in elderly patients. Section 51.6a describes the psychiatric aspects of long-term care. A general description of somatic treatment of schizophrenia can be found in Section 12.8. Other psychotic disorders are covered in Chapter 13. Cognitive disorders, including delirium and dementia, are covered in Chapter 10. Mood disorders are presented in Chapter 14. Dopamine receptor antagonists are presented in Section 31.17, and serotonin-dopamine antagonists are presented in Section 51.26. Medication-induced movement disorders are covered in Section 31.4.

SECTION REFERENCES

Avorn J, Soumerai SB, Everitt DE, Ross-Degnan D, Beers MH, Sherman D, Salem-Schatz SR, Fields D: A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. N Engl J Med *327:*168, 1992.

Casey DE: Side effect profiles of new antipsychotic agents. J Clin Psychiatry *57*(Suppl):40, 1996.

Ereshefsky L: Pharmacokinetics and drug interactions: Update for new antipsychotics. J Clin Psychiatry *57*(Suppl):12, 1996.

Harris MJ, Heaton RK, Schalz A, Bailey A, Patterson TL: Neuroleptic dose reduction in older psychotic patients. Schizophr Res *27:*241, 1997.

Harvey PD, Howanitz E, Parrella M, White L, Davidson M, Mohs RC, Hoblyn J, Davis KL: Symptoms, cognitive functioning, and adaptive skills in geriatric patients with lifelong schizophrenia: A comparison across treatment sites. Am J Psychiatry *155:*1080, 1998.

Harvey PD, Lombardi J, Leibman M, Parrella M, White L, Powchik P, Mohs RC, Davidson M, Davis KL: Age-related differences in formal thought disorder in chronically hospitalized schizophrenic patients: A cross-sectional study across nine decades. Am J Psychiatry *154:*205, 1997.

Health Care Financing Administration: Medicare and Medicaid: Requirements for longterm care facilities. Final rule with request for comments. Federal Register *54:*5316, 1989.

Horwitz GJ, Tariot PN, Mead K, Cox C: Discontinuation of antipsychotics in nursing home patients with dementia. Am J Geriatr Psychiatry *3:*290, 1995.

Jeste DV, Caligiuri MP, Paulsen JS, Heaton RK, Lacro JP, Harris MJ, Bailey A, Fell RL, McAdams LA: Risk of tardive dyskinesia in older patients: A prospective longitudinal study of 266 outpatients. Arch Gen Psychiatry *52:*756, 1995.

Jeste DV, Eastham JH, Lacro JP, Gierz M, Field MG, Harris MJ: Management of late-life psychosis. J Clin Psychiatry *57*(Suppl):39, 1996.

*Jeste DV, Harris MJ, Paulsen JS: Psychoses. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadovoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

*Jeste DV, Rockwell E, Harris MJ, Lohr JB, Lacro J: Conventional versus newer antipsychotics in elderly patients. Am J Geriatr Psychiatry *7:*70, 1999.

*Katz IR, Jeste DV, Mintzer JE, Clyde C, Napolitano J, Brecher M: Comparison of risperidone and placebo for psychosis and behavioral disturbances associated with dementia: A randomized, double-blind trial. J Clin Psychiatry *60:*107, 1999.

Lanctot KL, Best TS, Mittmann N, Liu BA, Oh PI, Einarson TR, Naranjo CA: Efficacy and safety of neuroleptics in behavioral disorders associated with dementia. J Clin Psychiatry *59:*550, 1998.

Lantz MS, Marin D: Pharmacologic treatment of agitation in dementia: A comprehensive review. J Geriatr Psychiatry Neurol *9:*107, 1996.

Lavretsky H, Sultzer DL: A structured trial of risperidone for the treatment of agitation in dementia. Am J Geriatr Psychiatry *6:*127, 1998.

*Lohr JB, Jeste DV, Harris MJ, Salzman C: Treatment of disordered behavior. In *Clinical Geriatric Psychopharmacology*, ed 2, C Salzman, editor. Williams & Wilkins, Baltimore, 1992.

Marder SR: Clinical experience with risperidone. J Clin Psychiatry *57*(Suppl):57, 1996.

Mazure CM, Nelson JC, Jatlow PI, Bowers MB: Acute neuroleptic treatment in elderly patients without dementia. Am J Geriatr Psychiatry *6:*221, 1998.

Meltzer HY: Atypical antipsychotic drugs. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Rich SS, Friedman JH, Ott BR: Risperidone versus clozapine in the treatment of psychosis in six patients with Parkinson's disease and other akinetic-rigid syndromes. J Clin Psychiatry *56:*556, 1995.

*Richelson E: Preclinical pharmacology of neuroleptics: Focus on new generation compounds. J Clin Psychiatry *57*(Suppl):4, 1996.

Schneider LS, Pollock VE, Lyness SA: A metaanalysis of controlled trials of neuroleptic treatment in dementia. J Am Geriatr Soc *38:*553, 1990.

Soares JC, Gershon S: Therapeutic targets in late-life psychoses: Review of concepts and critical issues. Schizophr Res *27:*227, 1997.

Sultzer DL, Gray KF, Gunay I, Berisford A, Mahler ME: A double-blind comparison of trazodone and haloperidol for treatment of agitation in patients with dementia. Am J Geriatr Psychiatry *5:*60, 1997.

Sunderland T: Treatment of the elderly suffering from psychosis and dementia. J Clin Psychiatry *57*(Suppl):53, 1996.

*Sweet RA, Pollock BG: New atypical antipsychotics. Experience and utility in the elderly. Drugs Aging *12:*115, 1998.

*Tariot PN: Treatment strategies for agitation and psychosis in dementia. J Clin Psychiatry *57*(Suppl):21, 1996.

Verma SD, Davidoff DA, Kambhampati KK: Management of the agitated elderly patient in the nursing home: The role of the atypical antipsychotics. J Clin Psychiatry *59*(Suppl):50, 1998.

Wirshing WC, Marder SR, Van Putten T, Ames D: Acute treatment of schizophrenia. In *Psychopharmacology: The Fourth Generation of Progress*, FE Bloom, DJ Kupfer, editors. Raven, New York, 1995.

Woerner MG, Alvir JMJ, Saltz BL, Lieberman JA, Kane JM: Prospective study of tardive dyskinesia in the elderly: Rates and risk factors. Am J Psychiatry *155:*1521, 1998.

Young RC, Meyers BS: Psychopharmacology. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

Zayas EM, Grossberg GT: The treatment of psychosis in late life. J Clin Psychiatry *59*(Suppl):5, 1998.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4F PSYCHOPHARMACOLOGY: ANTIDEMENTIA DRUGS

LON S. SCHNEIDER, M.D.

So-called antidementia drugs are by definition medications intended to improve cognitive function in patients with dementia. The term has regulatory significance with the Food and Drug Administration (FDA) since it differentiates these medications from cognition enhancers, medications that might be expected to improve cognition without reference to a particular illness. Worldwide, a broad range of putative cognition-enhancing agents is available for prescription, including piracetam derivatives, ergoloid mesylates (Hydergine), cholinesterase inhibitors, cholinergic agonists, and selegiline (Eldepryl), antioxidants, vitamins, food supplements, so-called neutraceuticals, and various plant derivatives, including extracts of *Ginkgo biloba*. In the United States the latter four groups of the products are regulated by the FDA only with respect to apparent safety and inappropriate advertising claims; efficacy for the conditions being treated need not be addressed.

The medications available for prescription in the United States that are indicated for treating cognitive impairment currently include only ergoloid mesylates and the cholinesterase inhibitors tacrine (Cognex), donepezil (Aricept), and revastigmine (ENA 713, Exelon, Novartis). The three cholinesterase inhibitors are specifically indicated for dementia of the Alzheimer's type, and thus are considered antidementia. Results of clinical trials performed throughout the world, however, suggest that other medications available in the United States might have efficacy in patients with Alzheimer's disease, including selegiline, various estrogen preparations, vitamin E, and anti-inflammatories.

Alzheimer's disease is the most common idiopathic dementia associated with aging and hence the most common target for antidementia therapeutics. Vascular dementia is the next most common. Although the causes of Alzheimer's disease are unknown (except for rare early-onset familial/genetic cases), a substantial part of the pathophysiology is sufficiently understood to allow rational development of symptomatic or potentially disease-modifying treatments. The pathophysiology of Alzheimer's disease is characterized, in part, by the deposition of insoluble b-amyloid, early loss of synapses, and death of predominantly acetylcholine-containing neurons, along with an atypical inflammatory response and excess production of toxic, oxidizing free radicals. The framework for clinical trials in Alzheimer's disease developed during the last decade after the general acceptance of research-based diagnostic criteria for the dementia of Alzheimer's type (e.g., National Institute of Neurological Communicative Disorders and Stroke [NINCDS] Alzheimer's Disease and Related Disorders Association [ADRDA] and the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* [DSM-IV] criteria) provided a better understanding of its underlying pathology, and the development of mechanism-based pharmacological therapeutics.

Until the 1980s medications were proposed for dementia treatment based on clinical experience or prevailing theories of dementia and aging. Examples of such earlier claimed treatments included psychostimulants (to increase attention) and vasodilators (to increase blood flow) and hyperbaric oxygen (to enhance metabolism). Only one medication, ergoloid mesylates, was approved in the United States for the ill-defined condition, senile mental decline, which was based on unvalidated diagnostic constructs such as “cerebral insufficiency” or “cerebral deterioration.” The mechanism of action of ergoloid mesylates is unknown but may involve complex neurotransmitter and vascular effects.

REGULATORY AND METHODOLOGICAL CONSIDERATIONS

FDA guidelines for establishing that a drug has antidementia efficacy require that patients meet accepted diagnostic criteria for a primary, well-accepted dementia such as Alzheimer's type, that trials use double-blind, placebo-controlled, parallel-group designs; that at least one trial be of 6 months' duration; and that improvement be obtained on both an accepted cognitive instrument and a clinician's global assessment, thus ensuring a clinically meaningful effect rather than a measurable but trivial cognitive effect.

From the FDA perspective, behavioral or functional (e.g., activities of daily living) changes that result from treatment are of secondary importance, even though improved functional status may have a major impact on maintaining independence and quality of life. However, improvement in behavior or functional activities in the absence of improved cognition does not allow a drug to be labeled antidementia.

For clinical trials in Alzheimer's disease, the Alzheimer's Disease Assessment Scale (ADAS) cognitive subscale has become the standard cognitive change measure and a clinician's interview–based impression of change instrument the measure of “clinical meaningfulness.” The ADAS cognitive subscale provides a summary score that reflects measures of word recall, naming, following commands, constructional and ideational praxis, orientation, word recognition, spoken language, comprehension of spoken language, word finding, and recall of test instructions.

Regulatory guidelines do not exist for the treatment of other dementias (e.g., vascular dementia, Parkinson's disease, Lewy body variant of Alzheimer's disease), other cognitive disorders, age-associated cognitive decline (DSM-IV), or cognitively unimpaired persons, in part, because of a lack of general expert consensus on the validity of the proposed diagnostic criteria. Although there are efforts to gain recognition of functionally significant but mild cognitive loss such as age-associated memory impairment (age-associated cognitive decline) or minimal cognitive impairment, the vagueness of the definitions and underlying pathology limit their general acceptance and hence the acceptance of treatments for such unrecognized conditions. The appropriateness of treatment of unimpaired persons with cognitive enhancers is controversial. Similarly, no guidelines exist for the treatment of cognitive deficits associated with schizophrenia, attention-deficit disorders, or depression, despite the existence of diagnostic criteria.

PHARMACOLOGICAL TREATMENT OF SYMPTOMATIC BEHAVIORS

Patients with Alzheimer's disease often experience such behavioral symptoms as depression, anxiety, psychosis, agitation, aggression, apathy, withdrawal, and vegetative disturbance (e.g., sleep or appetite change). These symptoms frequently lead to psychotropic medication treatment (both appropriate and inappropriate), physical restraint, and institutionalization. Behavioral symptoms burden caregivers and are considered among the most stressful aspects of caregiving. Various psychotropic medications are effective for pharmacological intervention and may improve many of these symptoms.

The treatment of symptomatic behaviors begins with a careful characterization of the target symptoms. Symptoms such as agitation, aggression, or delusions can result from chronic or recurrent psychiatric disorders; can represent the first manifestation of a new psychiatric disorder; or can indicate a change in medical status. A comprehensive medical evaluation will uncover potentially treatable underlying physical conditions that can exacerbate behaviors. Moreover, patients with dementia are at increased risk for developing superimposed deliria. (“Agitation” is often used to describe a range of nonspecific behaviors in elderly adults, including wandering, pacing, uncooperativeness, and verbal outbursts.)

Patients with dementia are sensitive to a variety of factors that may influence behavior, including sensory impairment, environmental disruption, or altered daily routines. Consequently, environmental manipulations may be helpful. Many patients with dementia experience anger, sadness, and denial, which may improve with nonpharmacological interventions. Various behavior-management techniques, including cognitive-behavioral management, may be helpful for moderately to severely impaired individuals. Supportive psychotherapy and counseling may benefit patients with mild cognitive impairments. Caregivers often need greater awareness that

their own behaviors influence the patient. Counseling about the nature of the patient's illness often helps relatives cope better.

DRUGS TO TREAT DEMENTIA

Drugs to treat dementia can be grouped into two broad categories: those intended to improve cognitive function symptomatically and those intended to modify the course of illness. Table 51.4f-1 lists available drugs for treating the cognitive symptoms of dementia.

Cholinesterase inhibitors
FDA-approved for dementia of the Alzheimer's type
Tacrine
Donepezil
Investigational
Rivastigmine
Metrifonate
Galantamine
Other medications available on the U.S. market
Anti-inflammatories (e.g., ibuprofen)
Estrogens
Conjugated estrogens (Premarin)
Estradiol
Vitamin E
Selegiline

Table 51.4f-1 Medications for Dementia

CHOLINERGIC DRUGS

The cholinergic hypothesis of dementia is based on observations of massive decline of the cholinergic basalocortical projections; reductions in the activity of choline acetyltransferase, the key enzyme in acetylcholine synthesis; a marked loss of cholinergic cell bodies in the nucleus basalis; correlations between choline acetyltransferase reduction or nucleus basalis cell reduction and neuritic plaque densities in cortical areas; and correlation of the cholinergic deficits with the decline in cognitive test performance. Many other neuronal systems are affected besides the cholinergic one, including an equally pronounced decline in the noradrenergic projections from locus ceruleus to cortex, and there is currently greater research emphasis on the classic, amyloid neuritic plaques and neurofibrillary tangles.

Nevertheless, the observations that cholinergic deficits represent the most pronounced transmitter deficit and one of the early events in the dementia process evoked considerable focus on cholinergic restitutive strategies. The cholinergic hypothesis of dementia treatment implies that if the cholinergic deficit is responsible for cognitive or behavioral changes, then potentiation of central cholinergic function should improve cognitive and behavioral impairment associated with Alzheimer's disease. Thus far there have been three main cholinergic approaches to Alzheimer's disease treatment: the use of precursor loading, cholinesterase inhibition, and direct, cholinergic agonists. Cholinergic treatment approaches include precursor loading, direct cholinergic receptor stimulation, cholinesterase inhibition, and (potentially) the use of cholinotrophic growth factors.

Precursor Loading Early investigations focused on the use of acetylcholine precursors, including choline and phosphatidylcholine (lecithin), in analogy to the use of dopamine precursors in Parkinson's disease. Numerous trials with precursors, however, have been unimpressive and have not reliably improved cognitive function.

Direct Cholinergic Stimulation Direct cholinergic agonists enhance responsiveness to released acetylcholine and minimize the presynaptic inhibition of further release by intrasynaptic acetylcholine concentrations. Thus far, however, direct muscarinic or nicotinic receptor agonists have not proven useful in enhancing cognition in Alzheimer's disease. Neither the older nor the newer relatively more muscarinic receptor subtype-specific investigational medications such as SB 202026, melameline, xanomeline, AF102B, and Lu25-109 have proved useful under investigation. Marketed cholinergic agonists such as arecholine or bethanechol (Urecholine) cannot be recommended because they have inadequate central nervous system (CNS) penetration, a high frequency of adverse cholinergic effects, and minimal efficacy. Although nicotine and nicotinic agents, including transdermal nicotine, and newer nicotinic receptor modulators do improve attention, arousal, and concentration in adults, their usefulness in cognitive impairment syndromes such as Alzheimer's disease has not been adequately demonstrated. Although there has been some interest in the possibility that cigarette smoking may prevent the onset of Alzheimer's disease or may mitigate symptoms, the evidence is conflicting, and it cannot be recommended.

Cholinesterase Inhibitors The most common and successful approach thus far to symptomatic treatment of Alzheimer's disease has been cholinesterase inhibition. In early studies, physostigmine (Synapton) elicited modest cognitive improvements but required individualized and frequent dosing because of its short duration of action (1 to 2 hours), which limited any potential clinical usefulness. Cholinesterase inhibitors inhibit CNS acetylcholinesterase in the synaptic cleft, thus preventing the degradation of endogenously released acetylcholine. However, progressive presynaptic cholinergic neuron loss continues, so production of acetylcholine diminishes progressively over time.

The several cholinesterase inhibitors that are, or soon will be, available for treatment of the cognitive symptoms of Alzheimer's disease can be grouped into three classes according to their structure and mode of inhibition: (1) reversible tertiary and quaternary amines (e.g., tacrine, donepezil, or galantamine); (2) pseudo-irreversible carbamates (e.g., sustained-release physostigmine and rivastigmine); and (3) the irreversible organophosphate inhibitor, metrifonate (Memron). Currently, tacrine, donepezil, and rivastigmine have been approved for marketing, and New Drug Applications are pending for three (metrifonate, sustained-release physostigmine, and galantamine).

Tacrine Tacrine is a centrally acting, reversible, cholinesterase inhibitor. It was approved by the FDA in September 1993 for the symptomatic treatment of Alzheimer's disease.

CLINICAL EFFICACY Tacrine has significant effects (compared with placebo) on the ADAS cognitive subscale, clinical global assessment, and measures of daily function. In one 12-week parallel-group trial, patients receiving 80 mg improved significantly on the ADAS cognitive and clinical global ratings. In another 30-week study, 663 patients randomized to three dosage treatments showed significant treatment effects in the 120-mg and 160-mg daily dosage groups on the ADAS cognitive subscale and clinical global ratings. Other trials have generally supported these effects.

PHARMACOKINETICS Tacrine is readily absorbed orally and is cleared by the liver through first-pass metabolism. Maximum concentrations are reached within an hour. Bioavailability is approximately 17 percent but is highly variable. Half-life is approximately 2 to 4 hours, and steady-state levels are achieved in about 24 hours. Higher doses and multiple dosing can prolong its elimination half-life, and plasma concentrations are not proportional to dose. At least three active metabolites are produced, mainly by cytochrome P450 isoenzyme 1A2 (CYP 1A2) hydroxylation of the ring positions 1, 2, and 4, and subsequently undergo glucuronidation and elimination. However, the enzyme is readily saturable, so at higher, more therapeutic doses, more of the drug avoids hepatic elimination. Coadministration with food decreases bioavailability by 30 to 40 percent. Men have lower plasma concentrations than women; cigarette-smoking lowers plasma concentrations.

DOSING Optimal response occurs between 120 and 160 mg a day. Treatment is initiated at 10 mg four times a day for 6 weeks, with weekly monitoring of transaminase enzymes (specifically, alamine aminotransferase [ALT]). If after 6 weeks the patient is tolerating the treatment and ALT is less than three times the upper limit of normal (3× ULN), the dosage should be increased to 20 mg four times a day for a second 6-week period. If tolerability remains satisfactory and ALTs are below 3× ULN, then the dosage should be increased to 30 mg four times a day for a third 6-week period. Finally, the dosage is increased to 40 mg four times a day. ALT activity should be monitored within 6 weeks of the last dosing change. After this period, monitoring may be reduced to every 3 months. The management of elevated ALT activity is detailed in the package insert, but basically it requires decreasing or temporarily discontinuing medication and raising or starting it again after ALT activity returns to normal.

DRUG INTERACTIONS Since tacrine is eliminated principally by CYP 1A2, drugs that share this elimination pathway may interact with it. Theophylline (Theo-Dur) concentrations may double when tacrine is added, and cimetidine (Tagamet) will increase the peak concentrations of tacrine by about 50 percent. Importantly, the use of agents with anticholinergic actions counteracts the effects of cholinesterase inhibitors such as tacrine. These include amitriptyline (Elavil) and other tricyclic drugs,

many antipsychotics, benztropine (Cogentin), and diphenhydramine (Benadryl) among others.

Donepezil Donepezil is a piperidine-based, specific, reversible inhibitor of acetylcholinesterase, specific to brain acetylcholinesterase, with little to no effect on peripheral butyrylcholinesterase. The advantages of donepezil over tacrine are no increased transaminase overactivity, once-daily dosing, and possibly milder cholinergic adverse effects.

CLINICAL EFFICACY Two multicenter, placebo-controlled studies comparing donepezil 5 and 10 mg a day with placebo for 12 and 24 weeks, respectively, and another 24-week trial conducted in Europe together involved about 1350 patients. Both studies found statistically significant benefit in both the ADAS and clinicians' rated improvement. The 10 mg a day dosage tended to show a greater effect than the lower dosage early in the course of treatment; but both dosages are more effective than placebo. In both U.S. studies neuropsychological scores approached those in the placebo group 3 to 6 weeks after discontinuation of donepezil.

PHARMACOKINETICS Oral bioavailability approaches 100 percent, with peak concentrations in 4 hours and linear pharmacokinetics. The drug accumulates at a constant rate, reaching steady-state in about 2 weeks. It has a relatively slow clearance, with a long elimination half-life of 70 hours. It is extensively bound to plasma proteins including a_1-acid glycoproteins. It is both excreted unchanged in the urine and extensively metabolized by CYP 2D6 and CYP 3A4 hepatic enzymes to active and inactive metabolites.

DOSING Treatment should be started with 5 mg a day for 2 to 4 weeks and if tolerated raised to 10 mg a day. The efficacy or appropriateness of higher dosages has not been established.

DRUG INTERACTIONS Inhibitors of CYP 3A4 and CYP 2D6 (such as ketoconazole [Nizoral] and quinidine [Duraquin]) and inducers of CYP 2D6 and CYP 3A4 (such as phenytoin [Dilantine], carbamazepine [Tegretol], dexamethasone [Decadron], and phenobarbital [Solfoton]) could either inhibit donepezil metabolism or increase its rate of elimination, but this is not known and may not be clinically significant in view of its high protein binding and very low therapeutic concentrations of 2 to 12 ng/mL.

Rivastigmine Rivastigmine is a pseudo-irreversible, carbamate, selective acetylcholinesterase subtype inhibitor. It is characterized by its selective binding and inactivation of acetylcholinesterase subtypes in specific regions of the brain. It is approved in several countries and is currently being reviewed by the FDA for marketing approval in the United States.

CLINICAL EFFICACY Four placebo-controlled trials of 6-month duration, using both fixed and adjustable doses and involving 2800 patients, many with concomitant illnesses and taking multiple medications, showed significant effects on cognition, ADAS, clinical global impression, and functional activities.

PHARMACOKINETICS Rivastigmine is well absorbed; very little binds to protein. Although the elimination half-life is less than 2 hours, enzyme inhibition lasts approximately 9 hours. Rivastigmine is not metabolized by the hepatic microsomal system. Rather, after binding to acetylcholinesterase the carbamate portion of rivastigmine is slowly hydrolyzed (hence, the pseudo-irreversibility), conjugated to a sulfate, and excreted renally.

DOSING The recommended starting dosage is 1.5 mg, increased to 3 mg twice a day after 2 weeks and to 4.5 mg twice a day if necessary.

DRUG INTERACTIONS Rivastigmine's metabolism does not depend on liver P450 enzymes; therefore no drug interactions related to the P450 system have been observed.

Metrifonate Metrifonate is a prodrug without intrinsic activity that is nonenzymatically converted to 2,2-dimethyldichlorovinyl phosphate (dichlorvos, [DDVP]), which has a long duration of cholinesterase inhibition. Since 1967 it has been used to treat schistosomiasis, largely in developing countries. It was approved for marketing in July 1998. DDVP is marketed as Trichlorfon, an agricultural insecticide.

CLINICAL EFFICACY Two 6-month, placebo-controlled, phase III studies showed significant cognitive effects and improvements in severity and clinicians' global impressions, with the typical expected cholinergic symptoms.

PHARMACOKINETICS Metrifonate is readily absorbed and has linear pharmacokinetics. Clinical response has been associated with extent of cholinesterase inhibition rather than drug concentration. Approximately 2 percent of the maximal concentration of the active drug (DDVP) reaches the CNS, suggesting a considerable blood-brain barrier or a rapid metabolism that limits accumulation in the CNS. The red blood cell cholinesterase inhibition half-life is nearly 2 months, and thus it is considered essentially irreversible.

DOSING No recommended approved dosage has been determined, but it is likely to be approximately 50 mg a day. Metrifonate is currently under review by the FDA. Recently, it was discontinued from all of the approximately 3000 subjects receiving it because of concern about the development of myasthenia and respiratory failure in a small number. It remains to be seen whether or not metrifonate will be approved.

Galantamine Galantamine (Reminyl) is an investigational cholinesterase inhibitor that also has nicotinic modulating effects. Early studies suggest that it is about as effective as other cholinesterase inhibitors. It is marketed in several European countries.

Adverse Effects of Cholinesterase Inhibitors As expected, the most common adverse effects of cholinesterase inhibitors involve those related to cholinergic excess, including nausea, vomiting, diarrhea, dyspepsia, anorexia, fatigue, and weight loss. In addition myalgia or muscle cramps occur in approximately 7 to 9 percent of patients, indicating direct myoneural effects (but not with rivastigmine, which specifically inhibits only certain CNS acetylcholinesterase subtypes). The frequency of the cholinergic adverse effects generally is less than 15 percent and often no greater than that with placebo in clinical trials. In general, they tend to be mild and short-lived (less than 2 or 3 days) and tend to occur when dosage is increased. Medication dosage can usually be maintained or temporarily lowered and then raised. Fatigue also tends to occur and occasionally there is an observable decrease in heart rate of 4 to 7 beats per minute, possibly due to enhanced vagal effects.

Direct hepatotoxicity is associated only with tacrine and is the most common abnormal finding with this drug. Approximately 30 percent of patients have aminotransferase activity exceeding three times the upper limits of normal, with the extent of the elevation correlating with the degree of hepatocyte injury. The elevations tend to occur about 5 to 7 weeks after the start of treatment; 95 percent occur within the first 12 weeks. The ALT elevations tend to be transient and return to normal within 6 weeks. Women are more likely to develop this than are men. Although no hepatic failure or hepatic deaths have been attributed to tacrine, patients taking tacrine need to be monitored weekly during the dose-titration phases.

Contraindications There are relatively few absolute contraindications to cholinesterase inhibitors, and they relate to the unique chemical properties of each drug. Donepezil is contraindicated in patients with hypersensitivity to piperidine derivatives. Information on interactions and contraindications of metrifonate are unavailable currently. Relative contraindications include potentially any illness that might be exacerbated by a cholinergic drug, such as asthma, pulmonary disease, and gastrointestinal ulcer disease.

Clinical Application and Treatment Implementation for Cholinesterase Inhibitors Currently, donepezil is probably the medication of choice since it has a simpler titration schedule than tacrine and no potential for hepatotoxicity. Extensive postmarketing experience with donepezil has revealed no serious adverse events. As newer cholinesterase inhibitors become available their differential pharmacology will require more-complex treatment decisions. These medications have only been demonstrated to be effective in outpatients with mild-to-moderate Alzheimer's disease (Mini-Mental State Examination score between 10 and 26). Many experts believe that efficacy is also apparent in both the more mildly and more severely cognitively impaired.

DURATION OF TREATMENT AND DISCONTINUATION Clinical trials with cholinesterase inhibitors have lasted only 6 months. In some trials, when medication is discontinued patients return within 3 to 6 weeks to the cognitive level of those who received placebo, indicating an overall symptomatic effect. Thus treatment with cholinesterase inhibitors needs to be somewhat longer than 6 months, but the overall duration is not known. One observational study with tacrine suggested that patients receiving more than 30 mg four times a day survived in the community and avoided nursing home placement for considerably longer than patients receiving 20 mg four times a day. Discontinuation of cholinesterase inhibitors has been associated with marked worsening of cognition and occasional episodes of confusion;

therefore tapering is recommended.

OTHER COGNITIVE ENHANCERS

Although cholinesterase inhibitors have been the most studied class of medications for Alzheimer's disease, a broad range of other drugs and supplements are either available for prescribing in some countries or currently undergoing clinical trials.

MONOAMINE OXIDASE TYPE B INHIBITORS

Propentophylline Propentopylline is a xanthine derivative that inhibits both adenosine uptake and phospodiesterase. Increased intrasynaptic adenosine concentration may reduce excitatory amino acid transmission associated with neurotoxicity. In addition, it may stimulate nerve growth factor (NGF), a cholinergic trophic factor, and inhibit free radical production that may be neurotoxic. In clinical trials over 6 and 12 months, patients with Alzheimer's disease or vascular dementia showed better in cognitive and global function than those receiving placebo. Phase III trials are ongoing.

Ergoloid Mesylates Ergoloid mesylates, the longest-used, putative cognition-enhancing drug, is still one of the more commonly prescribed drugs worldwide. It is not commonly prescribed in the United States because of lack of efficacy. It is currently used primarily for patients with dementia, age-associated cognitive symptoms, or just memory complaints. The agent has mixed and complex actions on central a-adrenergic, dopaminergic, and serotonergic receptors.

The FDA reviewed and approved ergoloid mesylates in 1978 and previously for “idiopathic decline in mental capacity” at dosages of 3 mg a day. Although numerous randomized, double-blinded, placebo-controlled clinical trials in a variety of elderly patient populations generally yielded positive effects, the studies are difficult to interpret because of such methodological limitations as inadequately diagnosed and clinically heterogeneous subjects, overreliance on a particular outcome instrument (the Sandoz Clinical Assessment Geriatric), a focus on improvement of individual behavioral symptoms, lack of attention to cognitive outcome, and variable doses and study durations. The clinical meaning of the results is unclear.

Nootropics The term nootropic is derived from the Greek *noos*(mind) and *troepin* (forward). The nootropics are a drug class with diverse actions and are structurally related to piracetam. They are cyclic analogues of g-aminobutyric acid (GABA) and are also related to glutamate, an excitatory amino acid. In addition to piracetam, they include oxiracetam, pramiractam, and aniracetam. These agents may protect the CNS from potential damage from hypoxia or drug intoxication. Nootropics may enhance the CNS microcirculation by reducing platelet activity and adherence of red blood cells to vessel walls. They may improve memory and learning, but no antidementia mechanism has been established. Clinical trials of nootropics in patients with Alzheimer's disease cannot be interpreted because they have not used accepted rigorous diagnostic criteria. Nootropics, however, are marketed for the treatment of dementia and cognitive impairment in many countries other than the United States. Some evidence suggests efficacy in attention-deficit disorder.

MEDICATIONS TO SLOW THE CLINICAL PROGRESSION OF ALZHEIMER'S DISEASE

Treatments for dementias exist that might slow neurodegenerative processes. Atypical inflammatory responses, the production of oxidizing free radicals, and deposition of insoluble and possibly toxic b-amyloid are part of the neurodegenerative process. Therefore, anti-inflammatories and antioxidants might have a role in the treatment of progressive dementias. The neurotrophic and cholinergic properties of estrogens may be useful in women.

Anti-Inflammatories Patients with rheumatoid arthritis who received long-term anti-inflammatory treatment had a lower risk for Alzheimer's disease than controls without rheumatoid arthritis. Twins who used anti-inflammatory medications had a lower risk for Alzheimer's disease than their cotwins: The risk of Alzheimer's disease was reduced in a cohort of people using anti-inflammatories. A placebo-controlled trial of indomethacin (Indocin) supports a potential role for at least nonsteroidal anti-inflammatory drugs. A 1-year, placebo-controlled trial of low-dosage prednisone (10 mg a day) by the National Institute on Aging (NIA) Alzheimer's Disease Consortium showed no advantageous effect, however. Because of gastrointestinal and other potential adverse effects from anti-inflammatory agents, the newer anti-inflammatory cyclo-oxygenase-2 inhibitors may offer promise, and such trials are ongoing.

In the absence of efficacy trials and considering their adverse effects nonsteroidal anti-inflammatory drugs (NSAIDs) cannot be recommended for either prophylaxis or treatment of Alzheimer's disease. Because of media attention and their availability over-the-counter and by prescription, some patients or caregivers may use or request them from physicians, but no clear guidance can be given. Effective dosage ranges are not known. Adverse effects including gastrointestinal bleeding are common, especially with long-term use.

Estrogens Estrogens may improve cognitive function through cholinergic neuroprotective and neurotrophic effects. For example, estradiol replacement enhances learning in ovariectomized rats; it also increases choline acetyltransferase and acetylcholine concentrations after ovariectomy. Clinical evidence of a possible effect of estrogen replacement therapy includes an inverse relationship between estrogen replacement therapy and death-certificate diagnoses of dementia, several case-control and cohort studies in which 40 to 50 percent reductions in the risk of developing Alzheimer's disease were reported, and several clinical reports suggesting that estradiol, estrone, and conjugated estrogens enhance cognitive function in patients with Alzheimer's disease. Furthermore, estrogen replacement therapy might augment the response to cholinesterase inhibitors such as tacrine. Adverse effects of estrogen replacement therapy are generally dose dependent and include endometrial hyperplasia and cancer, uterine bleeding (when taken with a progestin), breast tenderness, and nausea, among others. Because of the enhanced relative risk for endometrial cancer, periodic administration of progestins is recommended for women with intact uteri.

The evidence for estrogen replacement therapy is stronger for a protective effect than for a symptomatic one. An ongoing NIA-sponsored clinical trial is assessing the effect of conjugated equine estrogens (Premarin) on cognition in women with Alzheimer's disease. Another ongoing trial is assessing whether hormone replacement therapy delays onset of Alzheimer's disease (National Institute of Health Women's Health Initiative). In summary, a possible role for estrogen replacement therapy appears promising but is not substantiated.

Antioxidants Vitamin E (a-tocopherol, 1000 IU a day) or selegiline (5 mg twice a day) given alone to a group of moderately impaired patients with Alzheimer's disease delayed the time until patients required nursing home placement, died, or lost their ability to perform activities of daily living. Overall the time to these endpoints was approximately 7 months longer than with a placebo. There were no effects on cognition on this group of patients who were more severely impaired than those in other studies. Adverse effects included falls and syncope in selegiline patients (9 and 10 percent, respectively), a-tocopherol (14 and 7 percent), and the combination (22 and 16 percent). A subsequent 12-month placebo-controlled trial of transdermal selegiline in patients with milder Alzheimer's disease however, revealed no significant effects overall. The former is the only study assessing the potential efficacy of Vitamin E, and the sample size was small. Nevertheless, these results are promising, especially in light of the few adverse effects reported. Currently, no information exists concerning any prophylactic effects of these interventions.

Adjunctive and Combination Therapy It is not known whether combining medications such as cholinesterase inhibitors with other drugs increases efficacy. Pilot studies comparing combinations of cholinesterase inhibitors with drugs such as selegiline suggest only their feasibility, not efficacy. The combination of selegiline and vitamin E did not confer any added benefits over either agent alone in one study. Theoretically combining cognition-enhancing drugs with complementary mechanisms of action might be more useful than monotherapy. Consequently, many experienced clinicians may judiciously apply combination therapy.

VASCULAR DEMENTIA

Vascular dementia is a heterogeneous disorder and etiology must be considered. Some vascular dementia may be due to larger vessel stroke and sequellae, other to small vessel disease and ischemia. Potential interventions can be conceptualized as preventutive, disease modifying, or symptomatic. Hypertension is a significant risk factor for stroke and vascular dementia. Effective treatment of hypertension and hypercholesterolemia reduces cerebrovascular morbidity and mortality. Warfarin (coumadin) and aspirin seem to reduce the risk of stroke in patients with atrial fibrillation. Ticlopidine (Ticlid), aspirin, and dipyridamolc (Persantine) may prevent future stroke in patients with transient ischemic attacks or ischemic stroke. Antiplatelet agents may prove effective in preventing further vascular damage preceding and during dementia. Agents that reduce damage after stroke may be relevant to vascular dementia, including thrombolytic agents, calcium- and sodium-channel antagonists, *N*-methyl-D-aspartate (NMDA) receptor antagonists, glutamate synthesis inhibitors, glutamate-release inhibitors, GABA antagonists, gangliosides, aminosteroids, antioxidants, growth factors, and antiapoptotic agents. The efficacy of these agents in vascular dementia patients remains to be determined. Nimodipine (Nimotop), a calcium channel blocker, may improve the outcome of ischemic stroke when administered early in its course and may also improve cognitive function in some patients with dementia.

Treatment Once vascular dementia occurs, control of risk factors for stroke could help prevent further cognitive decline. Clinical observations suggest that among vascular dementia patients with hypertension, those with better cognition had systolic blood pressure above 135 mm Hg, and those who deteriorated had lower systolic blood pressure, and daily aspirin (325 mg) improved or stabilized cognition. Cessation of smoking and control of blood lipids may also be associated with improved cognition.

Metabolic or anti-ischemic approaches may be helpful. These include pentoxifylline (Trental), a hemorrheological agent; propentopylline, a phosphodiesterase inhibitor with neuroprotective properties, and nimodipine, a calcium channel blocker. Other classes of agents currently under consideration include glutamate antagonists, cholinergic enhancers, and cholinesterase inhibitors. The rationale for the latter involves the commonality of mixed vascular domentia and Alzheimer's disease and shared pathophysiological mechanisms between the two.

FUTURE DIRECTIONS

Current treatment options for Alzheimer's disease include the cholinesterase inhibitors tacrine and donepezil. Rivastigmine and galantamine may be available soon. Cholinergic agonists have not proved successful. Other drugs that affect the cholinergic system are in development. Antioxidants, selegiline, anti-inflammatories, and estrogens are marketed worldwide, not for the treatment of cognitive deficits and are not indicated for dementia, and their efficacy remains unproved. Although no currently available antidementia treatment has dramatically improved the cognitive impairments of Alzheimer's disease, the improvements are nonetheless clinically meaningful and considerable. The efficacy of these treatments for other cognitive disorders remains to be established. Worldwide research and development efforts offer promise of more effective approaches in the future.

SUGGESTED CROSS-REFERENCES

Biological therapies are the subject of Chapter 31. Other aspects of geriatric psychiatry can be found in the other sections of Chapter 51. Section 51.3e provides a broad discussion of Alzheimer's disease and other dementing disorders.

SECTION REFERENCES

Bellavance A: Efficacy of ticlopidine and aspirin for prevention of reversible cerebrovascular ischemic events: The ticlopidine aspirin stroke study. Stroke *24:*1452, 1993.

Chandra B: Treatment of multi-infarct dementia with citicholine. J Stroke Cerebrovasc Dis *2:*232, 1992.

European Pentoxifyline Multi-infarct Dementia (EPMID) Study Group: European pentoxifylline multi-infarct dementia study. Eur Neurol *36:*315, 1996.

Gelmers HJ, Hennerici M: Effect of nimodipine in acute ischemic stroke: Pooled results from five randomized trials. Stroke *21*(Suppl):81, 1990.

Gorelick PB, Brody JA, Cohen DC. Risk factors for dementia associated with multiple cerebral infarcts: A case-control analysis in predominantly African American hospital-based patients. Arch Neurol *50:*714, 1993.

*Kannel WB, Wolf PA, Verter J, McNamara PM: Epidemiologic assessment of the role of blood pressure in stroke: The Framingham Study. JAMA *276:*1269, 1996.

Lindenstrom E, Boysen G, Nyboe J: Influence of systolic and diastolic blood pressure on stroke risk: A prospective observational study. Am J Epidemiol *42:*1279, 1995.

Marcusson J, Rother M, Kitlner B, Rossner M, Smith RJ, Babic T, Folnegovic-Smalc V, Moller HJ, Labs KH: 12-month, randomized, placebo-controlled trial of propentofyiline (HWA 285) in patients with dementia according to DSM III-R. Dement Geriatr Cogn Disord *8:*320, 1997.

Meyer JS, Judd BW, Tawakma T, Rogers RL, Mortel KF: Improved cognition after control of risk factors for multi-infarct dementia. JAMA *256:*2203, 1986.

Meyer JS, Rogers RL, McClintic K, Mortel KF, Lotf J: Randomized clinical trial of daily aspirin therapy in multi-infarct dementia. J Am Geriatr Soc *37:*549, 1989.

Pantoni L, Carosi M, Amigoni S, Mascalchi M, Inzitari D: A preliminary open trial with nimodipine in patients with cognitive impairment and leukoaraiosis. Clin Neuropharmacol *19:*497, 1996.

*Sano M, Ernesto C, Klauber MR, Schafer K, Woodbury P, Thomas R, Grundman M, Growdon J, Cotman CW, Knopman D, Pfeiffer E, Schneider L, Thal L: A two-year, double-blind randomized multicenter trial of selegiline and a-tocopherol in the treatment of Alzheimer's disease. N Engl J Med *336:*1216, 1997.

*Schneider LS: New therapeutic approaches to Alzheimer's disease. J Clin Psychiatry *57*(Suppl 14):30, 1996.

*Schneider LS, Olin JT: Overview of clinical trials of Hydergine in dementia. Arch Neurol *51:*787, 1994.

Schneider LS, Tariot PN, Small GW: Update on treatment for Alzheimer's disease and other dementia. In *Psychiatric Clinics of North America. Annual of Drug Therapy*, J Rosenbaum, D Dunner, editors. Saunders, Philadelphia, 1997.

Sivenius J, Riekkinen PJ, Laasko M, Smets P, Lowenthal A: European Stroke Prevention Study: Antithrombotic therapy is also effective in the elderly. Acta Neurol Scand *87:*111, 1993.

Silver B, Weber J, Fisher M: Medical therapy for ischemic stroke. Clin Neuropharmacol *19:*101, 1996.

Stroke Prevention in Atrial Fibrillation Investigators: Warfarin versus aspirin for prevention of thromboembolism in atrial fibrillation: Stroke Prevention in Atrial Fibrillation Study II. Lancet *343:*687, 1994.

Tariot PN, Schneider LS: Contemporary treatment approaches to Alzheimer's disease. Consult Pharm *11:*16, 1996.

*van Gool WA: Efficacy of donepezil in Alzheimer's disease. Fact or artifact? Neurology *52:*218, 1999.

Walzi M, Walzl B, Lechner H: Results of a two-month follow-up after a single heparin-induced extracorporeal LDL precipitation in vascular dementia. J Stroke Cerebrovasc Dis *4:*179, 1994.

*Whisnant JP: Effectiveness versus efficacy of treatment of hypertension for stroke prevention. Neurology *46:*301, 1996.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4G ELECTROCONVULSIVE THERAPY

DONALD P. HAY, M.D., ELSA M. ZAYAS, M.D., AND GEORGE T. GROSSBERG, M.D.

First described by Ugo Cerletti and Lucio Bini in 1938, electroconvulsive therapy (ECT) has been used effectively to treat a variety of mental disorders in elderly adults. ECT is frequently misunderstood and feared by both health professionals and the lay public. In elderly adults, ECT is a safe and effective treatment for affective disorders, major depression (especially delusional depression), mania, schizoaffective disorders, and movement disorders and dementia with behavioral disturbances.

In the 1940s and 1950s ECT was used widely to treat many disorders. Its use declined with the advent of psychopharmacological agents. The last decade has seen a resurgence in the use of ECT. It is safe and effective, especially in elderly persons, who may be particularly sensitive to medication adverse effects and who may have comorbid medical problems.

HISTORY

ECT has undergone modifications since it was first introduced. These include the use of anesthesia, neuromuscular blocking agents, oxygenation, and electroencephalographic (EEG) and cardiovascular monitoring. The technique itself has been modified by the use of variable electrode placement and pulsed square-wave stimulation apparati. Successful use of this high-technology treatment now requires a team approach that includes the patient and family, the treating psychiatrist, the primary care physician, the anesthesiologist, and other specialists when appropriate.

EDUCATION

Mental illness in elderly adults poses both diagnostic and treatment challenges. Patients and their families must understand that ECT is safe and effective. The advantages of ECT should be discussed with the patients and their families, who must understand that concomitant medical illnesses and conditions can interfere with the safe and effective use of medications, and that severe and debilitating psychopathology necessitates a rapid treatment response. A psychiatrist or other well-informed mental health professional with expertise in the field can answer questions and alleviate concerns and misconceptions.

THEORETICAL ISSUES

The mechanism of action of ECT remains theoretical. Induction of a generalized seizure by electrical stimulation of the brain has been considered necessary for therapeutic efficacy. Traditionally it was thought that to be effective seizures needed to last at least 25 seconds peripherally and 40 seconds centrally, and although generally clinically correct, this concept is being reexamined.

Noradrenergic, serotonergic, and g-aminobutyric acid (GABA)-ergic neurotransmitters are believed to play a role in the efficacy of ECT. Animal studies with electroconvulsive shock have noted changes in serotonergic, GABAergic, and noradrenergic neurotransmission. Dopaminergic system involvement is indicated by the positive response of motor symptoms of Parkinson's disease and other movement disorders to ECT.

TECHNIQUES

Pre-ECT Evaluation The elderly individual with comorbid psychiatric and medical conditions must be carefully evaluated when one is considering ECT. Assessment includes the patient's medical history, the presence of a prior psychiatric illness, a prior response to medications, and a prior response to treatment with ECT.

The patient's cognitive status should be carefully assessed initially and monitored during the course of treatment. Cognitive functioning (especially short-term memory) should be evaluated prior to subsequent treatments in a series. One can use any standardized cognitive assessment scale such as the Mini-Mental State Examination. Any significant cognitive changes from baseline may require discontinuation of treatment. Cognitive impairment per se should not be a contraindication to the use of ECT. An ECT record sheet may be helpful to record and monitor the cognitive status of patients during their index series and continuation or maintenance ECT (Table 51.4g-1). The ECT record sheet should include the patient's identifying information, diagnosis, current medications and dosages (psychotherapeutic and nonpsychotherapeutic medicine), NPO status of the patient from the night before, removal of dentures, any recent acute medical problems, and a notation of heart and lung auscultation. The record sheet should also include the medications used before, during, and after the procedure, and it should have a place to record vital signs (blood pressure and pulse) prior to stimulus application and during the seizure as well as the peak vital signs and the posttreatment values.

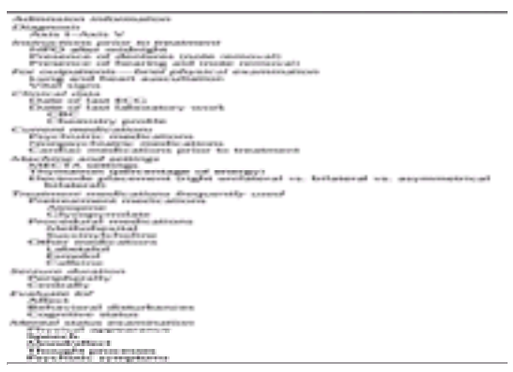

Table 51.4g-1 Components for ECT Record Sheet

The ECT record sheet should also include monitoring of the affective symptoms (A), behavior (B), and cognitive functioning (C) (termed the ABCs of geriatric psychiatry) to guide the treatment course and monitor response.

Evaluating the risk-benefit ratio by assessing affect, behavior, and cognitive symptoms helps the physician determine when it is advisable to initiate, continue, or

terminate treatment. One should evaluate the benefits and the potential risk for each patient prior to a treatment. For example, if a patient is at high risk because of cardiac disease or an arrhythmia but the benefits are high, one would proceed with the treatment after consultation with a cardiologist and anesthesiologist. If a patient is at high risk and the benefits of treatment are low, one would not proceed. Not giving the treatment may at times be more potentially dangerous than giving it.

INDICATIONS

The indications for ECT in elderly adults are the same as those for younger patients. The principal indications for ECT described in the American Psychiatric Association (APA) Task Force Report are mood disorders, schizophrenia and disorders with an affective component, catatonia, and schizoaffective disorder. Mood disorders including depressive disorders with or without melancholic features, depressive disorders with psychotic symptoms, and mania also all respond to ECT. ECT is effective and safe in elderly persons with comorbid medical conditions that may place them at a higher relative risk. Catatonia has been reported to be successfully treated with ECT and should be the treatment of choice for patients who do not respond favorably to antipsychotic agents or antidepressants. In addition, ECT can be effective treatment for the signs and symptoms of the movement disorder of Parkinson's disease, with or without comorbid depression.

Patients who are suffering from depression with comorbid cognitive deficits have also been reported to respond favorably to ECT. When the cognitive impairment is a direct result of the affective disorder, both the affective and the cognitive symptoms frequently respond to a course of ECT without a worsening of the patient's cognitive functioning. Patients suffering from poststroke depression may benefit from a course of ECT. ECT is an effective treatment for patients with agitation secondary to dementia.

RISKS

There is no absolute contraindication to the use of ECT. In considering the use of ECT, one must weigh the risk-benefit ratio of giving the treatment versus the risk-benefit ratio of not giving the treatment. Increased intracranial pressure is a relative contraindication to the use of ECT.

Elderly patients frequently have comorbid cardiovascular disease and may be at increased risk of suffering adverse effects from psychotropic medication. Despite the presence of comorbid cardiac disease, ECT has reportedly been used safely in this population. The most frequent complications found in patients with cardiovascular risk factors who received ECT were ventricular arrhythmias, ischemic events, atrial arrhythmias, and sinus bradycardia.

ECT can be safely administered to patients receiving anticoagulant therapy, and those with heart valve prostheses and preexisting cardiac arrhythmias, but a cardiology consultation should be obtained. These patients should continue taking their cardiac medications and should take their blood pressure medications before a treatment. Medications administered during the treatments are similar to those used in younger adults: methohexital (Brevital), succinylcholine (Anectine), glycopyrrolate (Robinul), esmolol (Brevibloc), labetalol (Normodyne), or caffeine. The concurrent medications that the patient is receiving should be evaluated. Some medications should not be administered during a course of ECT. It is recommended that lithium (Eskalith) be avoided during a course of ECT because of the increased risk of delirium. Theophylline (Theo-Dur) has been reported to prolong seizure duration, leading to status epilepticus. It has been recommended that patients on theophylline be switched to inhaled bronchodilators prior to treatment. One study reported seven patients who received ECT safely while receiving theophylline; however, these authors recommended that theophylline concentrations should be low or that theophylline not be coadministered with ECT.

Benzodiazepines have anticonvulsant properties and the literature suggests that they can shorten seizure duration and decrease treatment efficacy, particularly in unilateral ECT. Benzodiazepines should be avoided whenever possible. If a patient requires benzodiazepines, consider using short-half-life benzodiazepines and do not administer them immediately before treatment. Benzodiazepines should be tapered as much as possible before and during treatment series.

ASSESSMENT

Patients being considered for ECT should undergo the same evaluations as their younger counterparts, outlined in the American Psychiatric Association Task Force Report. These recommendations include (1) a psychiatric evaluation to establish the indication for ECT and (2) a complete medical evaluation including the history of the present illness, a complete physical examination, chemistry profile, cellular blood count, chest X-ray film, and electrocardiogram. Neuroimaging or a neurology consultation should be obtained when indicated by clinical findings.

Careful attention should be paid to orthopaedic status of elderly adults. Spinal films, once routinely recommended, are now obtained only when an adequate reason for concern exists, such as the possibility of compression or other fractures, osteoporosis, or spinal degeneration, especially if accompanied by pain. These conditions may require special considerations (including increasing the dosage of succinylcholine and carefully evaluating the degree of muscle blockade). Elderly patients frequently require an anesthesia consultation before treatment is initiated. In addition, comorbid medical conditions may prompt the treating psychiatrist to obtain other consultations (e.g., a cardiology consultation for an apparent cardiovascular risk or a pulmonary consultation for a patient with chronic obstructive pulmonary disease [COPD] or asthma). Elderly patients frequently have poor dentition (e.g., missing or loose teeth) that may pose potential treatment problems. A dental consultation may be warranted. Dentures should be removed prior to ECT unless they are sturdy and their removal would be more damaging or pose a greater danger to unprotected single or few teeth.

Informed consent is frequently a most difficult and challenging aspect of the treatment of elderly patients. Patients should be assessed for their capacity to give informed consent, since an elderly patient's cognitive status may be limited by either depressive symptomatology or by comorbid cognitive impairment. The patient and family need to be told about ECT, including the procedure, the risks and benefits, the possible adverse effects, and the mortality risk (see below). The family and the patient should be encouraged to openly discuss their fears and concerns. In most states in the United States, if the patient cannot give informed consent, a court order is required. Treating psychiatrists should become familiar with the regulations that govern ECT in their state.

ADVERSE EFFECTS

The mortality rate associated with ECT is approximately the same as that for anesthesia alone, approximately three or four deaths per 100,000 treatments or one death per 10,000 patients treated (or per 40,000 treatments). Cardiac complications with ECT include arrhythmias, T wave and ST segment changes, ischemic events, atrial arrhythmias, and sinus bradycardia. Post-ECT arrhythmias can be treated with lidocaine, and atropine can be use to treat bradyarrhythmias. Lidocaine (Decadron) should be avoided pre-ECT as it reportedly increases seizure thresholds, thereby inhibiting seizure induction.

Although relatively rare, central nervous system adverse effects can include interictal confusion that can persist on the days that patients do not have a treatment. Patients most likely to develop interictal confusion have been found on magnetic resonance imaging (MRI) to have subcortical white matter and basal ganglia lesions that preceded the ECT. Patients also may infrequently experience prolonged seizures or status epilepticus that may be treated with either further administration of the same barbiturate anesthetic used during ECT or benzodiazepines such as diazepam (Valium) and or lorazepam (Ativan), which can be given intravenously. However, benzodiazepines increase the risk of falls. Postictal excitement characterized by motor agitation, disorientation, and poor response to commands is reported to occur in 12.5 percent of patients. The usual treatment is intravenous benzodiazepine, although this may be avoided by increasing the pretreatment anesthetic dosage. ECT produces no structural brain changes.

Cognitive impairment or memory loss is a more frequently cited adverse effect of ECT, by far the adverse effect most feared by patients and family. Cognitive changes in elderly patients involve decreased orientation, attention, calculation, and recall. Cognitive adverse effects are frequently related to the form of ECT administered, including electrode placement, electrical dosage, stimulus wave form, and preexisting cognitive impairment. Bilateral electrode placement is associated with more persistent and severe memory impairment. In contrast, unilateral electrode placement is associated with fewer complaints of memory loss. Memory impairment increases in severity as a function of the number of treatments received. ECT may interfere with memory consolidation, with the most recently acquired information being the most vulnerable.

A newer technique for electrode placement, described for use especially in elderly patients, is a position termed “asymmetrical bilateral.” In this method, one electrode is placed above the left eye, with its lateral edge bordering the intersection between the forehead and the temple, while the right-sided electrode is placed in the standard frontotemporal position. This method has been used effectively and safely in elderly patients, with minimal cognitive adverse effects.

There may be an increased risk of post-ECT falls in elderly patients, which may result in other complications such as hip fractures. Patients should be closely monitored for post-ECT gait dysfunction or falls. Post-ECT headaches have also been noted. One retrospective study found that 5.5 percent patients developed

new-onset headaches, while 31 percent of patients who had a previous history of headaches reported worsening. Headaches can be treated symptomatically.

MEDICATIONS

Pretreatment with an anticholinergic medication is recommended to decrease the effect of parasympathetic discharge on the heart immediately after the delivery of the electrical stimulus. In elderly adults, b-adrenergic receptor antagonists (e.g., esmolol and less frequently labetolol [Normodyne]) are frequently used during a treatment to minimize changes in heart rate or blood pressure. Administration of a b-adrenergic receptor antagonist without prior administration of an atropinic medication may result in bradyarrhythmias. The most frequently used atropinic agents are glycopyrrolate and atropine. An atropine synthetic (glycopyrrolate) is preferable because it does not cross the blood-brain barrier and is thus reportedly less likely to cause postictal confusion in this potentially vulnerable population.

b-adrenergic receptor antagonists may be used pre-ECT in patients with known cardiac risk factors. Some authors have concluded that esmolol can effectively control the hyperdynamic response to ECT. Of concern is the rate-pressure product (i.e., the cardiac rate times the systolic blood pressure). A rate-pressure product above 20,000 may indicate ischemia in a significant number of elderly patients. The rate-pressure product can be controlled by lowering either component. Esmolol (duration of action, 3 to 5 minutes) can lower the rate as effectively as labetolol (duration of action, 3 to 5 hours).

The anesthetic agent of choice for ECT in elderly adults is methohexital, usually given at a dosage of 1 mg/kg. Methohexital reportedly results in rapid recovery of consciousness, is 2.5 times more potent than thiopental, and induces fewer arrhythmias than thiopental. Etomidate is safe and well tolerated when used for ECT. Seizure duration is longer with etomidate than with thiopental but is equivalent to that with methohexital.

Succinylcholine is the muscular blockade agent used most frequently in ECT. After the anesthetic induction is completed, a lower extremity should be cuffed off with a blood pressure cuff prior to the administration of succinylcholine to be able to visually monitor seizure activity. One should observe the patient for fasciculations, which may take longer in elderly patients whose circulation time is frequently slower. Complete paralysis can be determined by testing for absence of a patellar deep tendon reflex or absence of facial or radial nerve reflexes with a hand-held nerve stimulator.

Seizure duration during a course of ECT may become too brief for clinical benefit. Using higher-energy stimulation in elderly patients may increase the risk for adverse events. Caffeine given as an intravenous bolus can prolong the seizure. Caffeine augmentation has not been associated with significant cardiovascular or cognitive effects. During each treatment the patient should be well oxygenated. Patients should be continuously monitored cardiovascularly with an electrocardiogram recording, blood pressure monitoring, and a pulse oximeter.

Electrode Placement Electrode placement can be bilateral or unilateral nondominant, which in most patients is right unilateral, or asymmetrical bilateral, sometimes used in elderly patients to try to minimize the cognitive effects of ECT. Bilateral ECT is more powerful than right unilateral ECT and is associated with more memory impairment. The seizure threshold increases with age and reportedly increases 47 percent on average from the first to the sixth treatment, which requires adjustment of energy settings. Older patients have shorter seizure durations. Debate currently exists in the literature concerning whether other parameters of ictal EEG such as the postictal amplitude and immediate poststimulus spectral amplitude may be more indicative of efficacy than seizure duration.

In 1990 the APA Task Force on ECT recommended that stimulus intensity be adjusted to the individual needs of each patient. Two different empirical procedures have been used to arrive at the stimulus intensity, the fixed dose and the stimulus titration methods. The fixed dose is empirically derived on the basis of the patient's age. In the stimulus titration method a series of stimuli is given to estimate seizure threshold. Both methods have been used in elderly adults, but there are concerns that the titration method may induce subconvulsive seizures that could potentially result in such cardiac effects as bradyarrythmias or asystole. This is of special concern in those with comorbid cardiovascular risk factors.

FREQUENCY OF TREATMENT

The frequency of treatments varies from three times per week in North America to twice weekly in Great Britain. Usually between four and nine treatments are required for efficacy of an index series in elderly patients. The number of treatments in a series is determined by the patient's response. Relief of symptomatology is weighed against any evidence of confusion or cognitive impairment.

CONTINUATION AND MAINTENANCE ECT

After patients have had an effective course of ECT, one should consider the option of continuation ECT (given in the first 6 months after an index series) or maintenance ECT (continuing longer than 6 months after the index series). Frequently these patients have not responded to medication or have a higher risk of experiencing adverse effects. Harold Sackeim found in a prospective study that the patients more likely to relapse after ECT were those treated after ECT with the medication they had previously failed to respond to or they had received an inadequate trial. The literature reports that 50 percent of patients not treated with medications post-ECT relapse within the first 6 months following an index series of therapy. This relapse rate is not related to the length of the initial index episode or to the number of prior episodes of affective disorder. Patients who received continuation ECT following an index episode were less likely to have relapsed at 5-month follow-up. A course of continuation ECT is recommended for treatment-resistant patients or those who may be at an increased risk for medication adverse effects. Maintenance ECT is reported to be a safe, well-tolerated, efficacious treatment with minimal adverse effects, especially for elderly or psychotic patients.

Ambulatory ECT is becoming a primary mode of treatment, with increasing numbers of patients receiving ECT in an outpatient setting. The Association for Convulsive Therapy Task Force on Ambulatory ECT has established criteria recommending inpatient ECT based on diagnosis and for patients who may be at increased risk for medical complications.

The first index treatment is frequently administered to an elderly patient while the patient is hospitalized, to minimize the possibility of medical complications. The procedure for ambulatory ECT should be the same as that for inpatient ECT. Ambulatory ECT should involve a team consisting of the treating psychiatrist, an anesthesiologist, and the nursing staff. Caregivers for elderly individuals are integral partners in ambulatory ECT; they are responsible for communicating to the team any pretreatment considerations, administering pretreatment medications and managing the patient posttreatment (including driving). Caregivers should accompany the patient for the treatment because patients are not allowed to drive to and from the treatment.

RESEARCH AND EVALUATION

The safety and efficacy of ECT in elderly patients has been demonstrated repeatedly. Research continues in the areas of treating patients with comorbid medical conditions, and novel uses for ECT (e.g., treatment of chronic pain). Research and development of new protocols for ambulatory ECT is ongoing. Furthermore, interest is growing in the legal aspects of ECT, especially in the areas of physician privileging, the patient's capacity to give informed consent, and the legal statutes that govern treatment.

SUGGESTED CROSS-REFERENCES

Schizophrenia is presented in Chapter 12 and schizophrenia in the elderly in Section 51.3f. Mood disorders are covered in Chapter 14 and mood disorders in elderly patients in Section 51.3d. A more general discussion of ECT is presented in Sections 31-33. Basic electrophysiology is discussed in Section 1.9.

SECTION REFERENCES

*Abrams R: *Electroconvulsive Therapy*, ed 2. Oxford University Press, New York, 1992.

*Abrams R: Electroconvulsive therapy in the medically complicated patient. Psychiatr Clin North Am *14:*871, 1991.

Alexopoulus GS, Young RC, Abrams RC: ECT in the high-risk geriatric patient. Convuls Ther *5:*75, 1989.

*American Psychiatric Association Task Force: Report on electroconvulsive therapy: The practice of ECT: Recommendations for treatment, training, and privileging. American Psychiatric Association, Washington, DC, 1990.

Cerletti U, Bini L: Un novometodo di shockterapie “L'elettro-shock.” Boll Accad Med Roma *64:*136, 1938.

Clarke TB, Coffey CE, Hoffman GW: Continuation therapy for depression using outpatient electroconvulsive therapy. Convuls Ther *5:*330, 1989.

Coffey CE, Figiel GS, Djang WT, Sullivan DC: Effects of ECT on brain structure: A pilot prospective magnetic resonance imaging study. Am J Psychiatry *145:*701, 1988.

Coffey CE, Lucke J, Weiner RD. Seizure threshold in electroconvulsive therapy (ECT) II. The anticonsultant effect of ECT. Biol Psychiatry *37:*777, 1995.

Costelli I, Steiner A, Kaufman MA: Comparative effects of esmolol and labetalol to attenuate hyperdynamic states after electroconvulsive therapy. Anesth Analg *80:*557, 1995.

Currier MB, Murray GB, Welch CC: Electroconvulsive therapy for post-stroke depressed geriatric patients. J Neuropsychiatry Clin Neurosci *4:*140, 1992.

Figiel GS, Coffey CE, Djang WT, Haffman G Jr, Doraiswamy PM: Brain magnetic resonance imaging findings in ECT-induced delirium. J Neuropsychiatry Clin Neurosci *2:*53, 1990.

*Fink M, Abrams R: Ambulatory electroconvulsive therapy. Report of a task force of the Association for Convulsive Therapy. Convuls Ther *12:*42, 1996.

Flint AJ, Rifat SL: The treatment of psychotic depression in later life: A comparison of pharmacotherapy and ECT. Int J Geriatr Psychiatry *13:*23, 1998.

Folstein MF, Folstein SE, McHugh PR: Mini-mental state: A practical method for grading the cognitive state of patient for the clinician. J Psychiatr Res *12:*189, 1975.

Greenberg L, Fink M: The use of electroconvulsive therapy in geriatric patients. Clin Geriatr Med *8:*349, 1992.

Greenberg RM: ECT in the elderly. New Dir Ment Health Serv *76:*85, 1997.

Greenberg RM, Pettli A: Benzodiazepine and ECT. Convuls Ther *9:*262, 1993.

Hay DP: Electroconvulsive therapy in the medically ill elderly. Convuls Ther *5:*8, 1989.

Hay DP, Hay L, Blackwell B, Spiro HR: ECT and tardive dyskinesia. J Geriatr Psychiatry Neurol *3:*106, 1990.

Hay DP, Hay L, Grossberg G: Electroconvulsive therapy for agitated dementia patients: Treating Alzheimer's and other dementia. In *Clinical Implication of Recent Research Advances*, MD Beyner, editor. Springer, New York, 1995.

*Hay DP, Hay LK: Electroconvulsive therapy. In *Comprehensive Review of Geriatric Psychiatry*, vol 2, ed 2, J Sadovoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

Hay DP, Hay L: The role of ECT in the treatment of depression. In *Depression: New Directions in Theory, Research, and Practice*, CD McCann, NS Endler, editors. Wall and Emerson, Toronto, 1990.

Hay DP, Hay L, Blockwell B, Spiro HR: ECT and tardive dyskinesia. J Geriatr Psychiatry Neurol *3:*106, 1990.

Howie MB, Hiestand DC, Zvara DA, Kim PY, McSweeney TD, Coffman JA: Defining the dose range for esmolol used in electroconvulsive therapy hemodynamic attenuation. Anesth Analg *75:*805, 1992.

Kelsey MC, Grossberg GT: Safety and efficacy of caffeine augmented ECT in elderly depressives: A retrospective study. J Geriatr Psychiatry Neurol *8:*168, 1995.

*Kramer BA: ECT in elderly patients with schizophrenia. Am J Geriatr Psychiatry *7:*2, 1999.

Krystal AD, Weiner RD, McCall V, Shelp FE, Arias R, Smith P: The effect of ECT stimulus dose and electrode placement on ictal electroencephalogram: An intraindividual crossover study. Biol Psychiatry *34:*759, 1993.

Malitz S, Sackeim HA, Deana P, Kauzler M, Kerr B: The efficacy of electroconvulsive therapy. Dose-response interactions with modality. Ann NY Acad Sci *462:*56, 1986.

Manly DT, Swartz CU: Asymmetrical bilateral right frontotemporal left frontal stimulus electrode placement: comparisons with bifrontotemporal and unilateral placement. Convuls Ther *10:*267, 1994.

Monroe R: Maintenance electroconvulsive therapy. Psychiatr Clin North Am *14:*947, 1991.

Norden DK, Siegler E: Electroconvulsive therapy in the elderly. Hosp Pract *28:*59, 1993.

Petrides G, Dhossche D, Fink M, Francis A: Continuation ECT: Relapse prevention in affective disorders. Convuls Ther *10:*189, 1994.

Rasmussen K, Abrams R: Treatment of Parkinson's disease with electroconvulsive therapy. Psychiatr Clin North Am *14:*4, 1981.

Rasmussen RG, Zorumski CF: Electroconvulsive therapy in patients taking theophylline. J Clin Psychiatry *54:*427, 1993.

Rice EH, Sombrotton LB, Markowitz JC, Leon AC: Cardiovascular morbidity in high-risk patients during ECT. Am J Psychiatry *151:*1637, 1994.

Rohland BM, Carroll BT, Jacoby RG: ECT in the treatment of the catatonic syndrome. J Affect Disord *29:*255, 1993.

Sackeim HA, Prudic J, Devanand DP, Decina P, Kerr B, Malitz S: The impact of medication resistance and continuation pharmacotherapy on relapse following response to electroconvulsive therapy in major depression. J Clin Psychopharmacol *10:*96, 1990.

Small JG, Milstein V: Lithium interactions: Lithium and electroconvulsive therapy. J Clin Psychopharmacol *10:*346, 1990.

Stoudemire A, Hill CD, Marquardt M, Dalton S, Lewison BJ: Recovery and relapse in geriatric depression after treatment with antidepressants and ECT in a medical-psychiatric population. Gen Hosp Psychiatry *20:*170, 1998.

Stoudemire A, Hill CD, Morris R, Dalton ST: Improvement in depression-related cognitive dysfunction following ECT. J Neuropsychiatry Clin Neurosci *7:*31, 1995.

Swartz CM: Anesthesia for ECT. Convuls Ther *9:*301, 1993.

Tomac TA, Rummans TA, Pileggi TS, Li H: Safety and efficacy of electroconvulsive therapy in patients over age 85. Am J Geriatr Psychiatry *5:*126, 1997.

Trzepacz PT, Weniger FC, Greenhouse J: Etomidate anesthesia increases seizure duration during ECT: A retrospective study. Gen Hosp Psychiatry *15:*115, 1993.

Weiner RD, Coffey CE: Minimizing therapeutic differences between bilateral and unilateral nondominant ECT. Convuls Ther *2:*261, 1986.

Weiner SJ, Ward TN, Ravaus CG: Headache and electroconvulsive therapy. Headache *34:*155, 1994.

Zielinski RJ, Roose SP, Devanand MD, Woodring S, Sackheim HA: Cardiovascular complications of ECT in depressed patients with cardiac disease. Am J Psychiatry *150:*904, 1993.

Zwil AS, Pelchat RJ: ECT in the treatment of patients with neurological and somatic disease. Int J Psychiatry Med *24:*1, 1994.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4H PSYCHOSOCIAL TREATMENTS: GENERAL PRINCIPLES

JOEL SADAVOY, M.D., F.R.C.P.(C)

Psychotherapy of elderly persons comprises several modalities and is applied to a heterogeneous array of patients. Psychotherapy is delivered within a social context that differs for each of these elderly individuals and has an important influence on treatment that must be considered individually.

Regardless of the type of psychotherapy used, therapists for aged patients must understand, acknowledge, and incorporate into the treatment the array of psychosocial factors that form the context of the patients' problems and the matrix of the patients' life, within which they are supported or fail. In psychodynamic terms, this matrix may be viewed as the holding environment, the adult version of what parents provide an infant. As with children, when the holding environment fails to support and enhance, the individual comes under psychological stress. Persons in their environment are reflected in the environments within the persons. Psychotherapists will be most helpful and effective when they work simultaneously with both the external and internal worlds of elderly patients.

DEFINITION

The term “psychosocial” refers to social and environmental elements in individuals' lives that interact with their intrapsychic processes. Such elements include interpersonal relationships, role functions, societal attitudes, economics, physical environment, culture and ethnicity, and spiritual beliefs.

RELATIONSHIPS

Old age assaults the integrity of the patient's social network. Perhaps most profound are the changes associated with spousal bereavement, producing sudden role change, isolation, abandonment, economic disruption, and altered interpersonal relationships. Beyond bereavement, marital relationships also are changed by physical or cognitive impairment that shifts patterns of dominance and responsibility. For example, the basic nurturing role previously carried out by a loving wife now falls into the unaccustomed hands of her husband in the face of her decline with Alzheimer's disease. He experiences not only loss but the need to offer comfort and care, behavior and feelings with which he has always been awkward. Guilt, resentment, and some anger emerge as a consequence.

Intimacy similarly shifts, often toward distance and isolation. Friendships easily maintained during younger, more active years often become more difficult to sustain. Driving is more challenging, evening and night travel more fearsome, and gathering places less welcoming and accommodating. Illness and other age-related events such as bereavement take their toll on friendship networks. Hence, even for those with naturally outgoing personalities, the avenues for establishing and maintaining new friendships are often blocked. Like Swiss cheese, friendship networks become riddled with holes that gradually coalesce over time. One effect is the erosion of relationships in which old persons' history, accomplishments, desires, and ambitions are known by their friends. Increasingly, aging fragments and isolates, uncontrollably and inexorably. While certainly not universal, this process of change contributes to creating an atmosphere during psychotherapy of elderly persons, a context imposed by old age on the therapeutic process. Old patients, often through no fault of their own and despite their efforts, present with a burden of aloneness. This can create in therapists the anxious sense that they must do more than treat—respond to an unspoken request for friendship and become for the patient the only place and person to offer a safe, intimate, emotional haven.

Beyond friendship, sexual and other physical intimacy may vanish from the life of the old person, leaving no one to touch or be touched by. The literature suggests that sexuality among elderly persons can remain vibrant and pleasurable; however, over a certain age, elderly persons are often widowed and without partners. Hence, for example, intimate touching may then become the domain of the medical encounter during physical examination or the nonsexual hugging of children, grandchildren, or friends. Physical isolation becomes especially evident in retirement or nursing homes or other institutions where predominantly old women live in an unusually gender-homogenous environment. Whatever sexual or intimate interest remains is put in cold storage, often forever. Psychotherapists and other caregivers must be sharply aware of the patient's needs and reactions under these circumstances, elicit the impact on the patient, and incorporate the reality into the treatment, working toward either change or acceptance.

The psychosocial phenomenon of loss of intimate relationships in old age may have special impact on the form of the psychotherapeutic relationship. Therapists are often cast as a wished-for substitute for the lost relationship. Therapists may feel alone with these patients in their depleted world and experience uncomfortable pressures to adopt the role or behaviors of friend, confidant, protector, advisor, or advocate.

While old age is a time of substantial change that requires great adaptability, satisfactions may still abound. However, they tend to be encountered within a narrower frame that is focused more on family relationships and very close friendships. For many, the loss or unavailability of friendships or relationships promotes a kind of centrifugal emotional force that draws some elderly persons into their families and somewhat away from other relationships. The loss of confiding relationships may have clinical relevance, since some research suggests that these relationships protect against depressive disorder. Effective psychotherapy and environmental interventions that support the patient's need for new relationships can counter some of the effects of loss of confidants.

The increasing importance of family relationships with advancing age adds an element to the psychotherapy of older adults not found as prominently among younger adult patients. Therapists must often consider age-induced dependency issues in families and the conflicts these issues engender, such as forced intimacy and caregiving situations with adult children. For example, conflicts arise over the different perceptions of need that the elderly parent has. Children may be overly anxious and protective; aging parents may inappropriately resist needed help or exaggerate their helplessness. Whatever the pattern, the psychotherapist for the elderly person must take the family system into account, sometimes involving the other family members directly in the therapy or helping the elderly person develop appropriate interpersonal patterns of interaction. In this area as in many others, flexibility in psychotherapeutic technique and interventions is essential. The usual adult-therapist stance broadens when treating elderly persons as psychosocial issues associated with frailty, dependency, and relational shifts impinge. Impelled by these realities, the therapist must often combine the overt roles of advocate, intermediary, and educator with the more traditional psychotherapeutic roles and interventions. Sometimes this means reaching out actively to families, social agencies, or medical caregivers, activities that may be beyond the psychotherapist's usual role. While flexibility is essential, therapists should also examine their motivations to intervene on their patients' behalf. Transference and countertransference-induced forces can lead therapists to inappropriately cross therapeutic boundaries, unwittingly adopting the idealized roles of adult child or spouse, seemingly in response to real psychosocial needs, but actually impelled by dimly understood and often unconscious dynamic forces in the therapeutic relationship. Therapists dealing with relationship changes in aged patients must be prepared to deal with their own attendant feelings of loss, grief, fear, and loneliness prompted by the psychosocial stresses of old age.

ROLE CHANGES

Occupational role changes impose relatively unique challenges on elderly persons. This group is almost universally forced to give up important sources of self-esteem, creative activity, social networks, and the organizing effects of job expectations. Reactions differ. Retirement may be eagerly anticipated and handled well, especially by those who used to be called blue-collar workers. Others, such as professionals, often avoid retirement and may find the loss of their occupational role difficult. Either way, the psychotherapist must appreciate the centrality of work-related roles in structuring life and the changes aging imposes on this structure. These factors are especially relevant in the so-called young-old, 65- to 75-year-olds who are more likely to have just encountered the issues of occupational role

change. These issues generally recede later, either adapted to and worked through or supplanted by other, more pressing issues. However, when fresh, the psychosocial pressures of occupational role change, especially when combined with events such as spousal bereavement, often induce vulnerability and fearfulness and impair the capacity to seek new, alternative sources of gratification.

Parenting roles may change dramatically in old age. Well-established patterns of parental dominance and reliance by children on a strong authoritative parent often suddenly reverse, and conflict, guilt, or fear may invade a previously stable system. Particularly vulnerable are family systems whose members adapted to earlier conflicts by establishing distance (e.g., children moving away). The increasing needs of aging parents may impose pressures on children to care for parents for whom they have had long-standing conflicted feelings. The psychotherapist must be careful to examine and understand the parent-child system and beware of taking stereotyped approaches to the aging parents, since the patterns of parental role changes are quite varied. For example, frailty and fear of being alone and incapacitated often leads elderly patients to complain about the inattentiveness of adult children. Taken at face value, the psychosocial needs of the patient often are indisputable. However, the pattern of the family members' reactions to the patients' needs are rooted in the long history of their relationships with each other. To be effective, the therapist must be sensitive to the emotional forces that are preventing the family from adapting to the new, age-induced needs of the parent.

SOCIETAL ATTITUDES AND RESPONSES

Biases about aging and the aged abound, and the older individual is confronted with them. Some of the North American perspectives on aging are quite negative. For example, the elderly are perceived as no longer productive and therefore of marginal importance to mainstream life or, because of frailty, imposing undue economic burdens on overstretched pension and public support systems. The effective psychotherapist for elderly persons incorporates these issues into the therapy, recognizing with the patient the nature and impact of the messages, and working on appropriate responses. In this role, the therapist can help buffer the sometimes harsh messages that old persons hear daily about those in their age group. A particular issue is the difference between the patient's self-perception—more youthful, involved, and competent—and the perception of younger adults who can only see and react to the apparently universal aging appearance—lost skin turgor, decreased muscle mass, gray hair, and postural change. In other words, the somewhat stereotyped view of old persons by the young delays a more complete and deeper understanding of the truer self-perceptions held by the older individual. This phenomenon has been termed the *normative intergenerational empathic lag*. Older patients in psychotherapy often feel both misunderstood by society at large and ashamed of themselves. The therapist will be most effective if these issues are articulated and explored with aging patients to help them counter societally induced loss of self-esteem and feelings of alienation by confronting distortions and developing techniques to defend against inappropriate stereotyping by those around them.

ECONOMIC AND ENVIRONMENTAL FACTORS

For many, old age is a time of economic and environmental change. Retirement and the many difficulties sometimes imposed by living on fixed incomes, social security, or supplements provided by children can induce real or fantasized anxiety and preoccupation about financial security. These factors can influence psychotherapy in a number of ways (e.g., inducing envious transferential feelings in patients or guilty countertransference in therapists).

Financial restrictions (often real but sometimes imagined) lead to restricted social activity; feelings of isolation, humiliation, and shame; and uncertainty and fear of the future. Like advancing frailty or bereavement, economic difficulties must be actively dealt with in psychotherapy. This involves determining with patients their actual financial situation, separating distortions from reality, uncovering latent meanings about financial security, and then introducing both practical approaches to coping and interpretive understanding of the underlying meaning of the situation to these patients.

Living accommodations is another important psychosocial variable that often changes in old age. The key element that induces psychological stress associated with moving to a new setting in old age is that it is generally a forced move, precipitated by physical or cognitive decline, economic factors, or isolation. Even more than death, the specter of abandonment to an impersonal chronic care facility is most feared by the elderly. The prospect of such a move provokes various forms of abandonment anxiety, including the knowledge that this will probably be their last abode before death, or internal conflicts created by forced interaction with other elderly residents, which living in even the best long-term care environment imposes. Aged patients routinely express their anxiety at being confronted with the pain and disability of their age-peers. They see their own future mirrored in them and become afraid. These feelings can prevent patients from accepting realistic moves to settings where they can receive needed care. Pressing a reluctant, frail patient to accept needed help in an institution can pose ethical dilemmas for the therapist. Not the least of these is the reality that many or most of our institutions are barely adequate to the task of satisfying the needs of their residents. Inevitably, no matter what the setting, the new resident entering long-term care will encounter severe psychosocial stressors, including some loss of individual identity, regimentation to the institutional routines, and loss of privacy. Confronting these realities can be difficult not only for patients, but also for the sensitive therapist who is only too aware of the limitations of the resources on which they must rely.

Psychotherapists of aged patients facing environmental dislocation are most effective if they can uncover the patients' fears and anxieties—abandonment, loneliness, shame, infirmity, pain, or death—and work toward both realistic appraisal of the future and development of skills to combat the specific feared outcome of the environmental transition.

ETHNORACIAL ISSUES

Culture, ethnicity, and race have a profound impact on aging that must be understood and dealt with by psychotherapists, especially those practicing in geographical areas made up of diverse communities. Depending on the specific community, the presence of mental illness may carry with it a profound stigma that stops both patients and their families from seeking care. Hence, such patients often carry societally induced feelings of shame at their inability to cope with their emotional problems. Compounding the more universal problematic aspects of aging, aging members of ethnoracial communities also must deal with barriers to mental health care imposed by language and immigration-related features, such as the feelings associated with leaving their homeland, unanticipated family conflicts, loss of status, or inaccessible spiritual support. Often, newly arrived or even well-established but poorly integrated aged immigrants carry a heavy burden of isolation, unfamiliar with how to travel beyond their neighborhood, unduly reliant on often over-stretched family members, feeling like aliens when they may have expected to feel like respected elders, encountering racist attitudes, or struggling with inadequate finances. To establish an effective therapeutic alliance, psychotherapists must make the patients aware that they are sensitive to these issues by inquiring about them in an empathic, caring, and specific manner. Most important is overtly demonstrating to these patients that they are recognized as individuals who will not be stereotyped by virtue of their culture, ethnicity, or race.

Factors associated with immigration may have special relevance to psychotherapy. For example, those in refugee groups often have hidden histories of severe deprivation and trauma associated with war, political oppression, or famine. The late effects of earlier adult trauma frequently endure, but their presence must be actively sought by the therapist, as patients frequently keep silent about such experiences and their persisting effects.

SPIRITUAL ISSUES

Frequently omitted from psychotherapy inquiry, spiritual issues have profound importance in the lives of many elderly patients. Religious or philosophical beliefs often provide great comfort and enable appropriate acceptance of change, aloneness, or death. Aging patients may return to religion in later years both as a source of comfort and as a social resource. Psychotherapy will be more acceptable to the patient and often more effective when it is based on a carefully elicited understanding of the patient's beliefs and uses the strengths inherent in the patient's faith as an ally in treatment.

SUGGESTED CROSS-REFERENCES

The various psychotherapies are covered in Chapter 30. Sexuality is presented in Section 19.1. Death and bereavement are covered in Section 28.5. Minority issues in geriatric psychiatry are presented in Section 51.6e. Section 51.4i covers individual psychotherapy, Section 51.4j covers cognitive-behavioral therapy, Section 51.4k presents family interventions, and Section 51.4l discusses group therapy.

SECTION REFERENCES

*Anderson L: Loneliness research and intervention: A review of the literature. Aging Ment Health *2:*264, 1998.

*Fernandez ME, Mutran EJ, Reitzes DC, Sudha S: Ethnicity, gender and depressive symptoms in older workers. Gerontologist *38:*71, 1998.

Koenig HG: Religious attitudes and practices of hospitalized medically ill older adults. Int J Geriatr Psychiatry *13:*213, 1998.

*McFarland MR: Use of culture care theory with Anglo- and African-American elders in a long-term care setting. Nurs Sci Q *10:*186, 1997.

*Miller MD, Wolfson L, Frank E, Cornes C, Silberman R, Ehrenpries L, Zaltman J, Malloy J, Reynolds CF: Using interpersonal psychotherapy (IPT) in a combined psychotherapy and medication research protocol with depressed elders. A descriptive report with case vignettes. J Psychother Pract Res *7:*47, 1997.

*Sadavoy J: Late life effects of prior psychological trauma. Am J Geriatr Psychiatry *5:*287, 1997.

Turner BS: Aging and generational conflicts: A reply to Sara Levin. Br J Sociol *49:*299, 1998.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4I INDIVIDUAL PSYCHOTHERAPY

JOEL SADAVOY, M.D., F.R.C.P.(C), AND LAWRENCE W. LAZARUS, M.D.

Three elements distinguish geriatric psychotherapy: it is directed to the needs of patients over the age of 65, it requires modification of technique because of the special developmental features of old age, and it requires flexibility to treat the frequent comorbid age-related conditions so prevalent in this age group. Perhaps more often than psychotherapy of other age groups, geriatric psychotherapy must be practiced in a variety of settings. Because of rapid changes during the later years of life, the techniques of geriatric psychotherapy for a given patient must often be attuned to the ebb and flow of relapse, crisis, or chronic decline.

HISTORY

Modern psychotherapy began with Sigmund Freud's attempt to construct a universal model of understanding of human behavior. In the essentially prebiological era of psychiatric treatment, psychoanalytic theory formed the basis for most psychotherapeutic interventions. Unfortunately, this erroneously colored the attitudes of psychiatrists toward psychotherapy of elderly persons. Freud held a particularly gloomy, but highly influential, view of the potential for successful psychoanalytic treatment of patients over the age of 50. He believed that there was too much material to be dealt with in analysis and that the elderly had inelastic mental processes and hence were not educable. These ideas produced inordinate pessimism about the potential for psychological change in elderly adults. As the psychotherapies matured and began to be applied to a broader range of patients, therapists came to recognize that age per se does not preclude therapeutic change. Indeed, in the 1950s and 1960s several analysts wrote enthusiastically of their positive experiences in treating elderly patients. In retrospect, this enthusiasm resulted from treatment of generally intact and relatively vigorous elderly patients.

The quality of evidence supporting psychotherapies for elderly patients is highly variable. While some randomized controlled trials exist, most descriptions of psychotherapeutic outcomes are based on uncontrolled studies, and the bulk of opinions about psychotherapy of elderly patients that appear in the literature reflect the clinical experience of respected authorities. Hence, any recommendations derived from this data should be considered in the early stages of validation and subject to modification as more-careful studies refine the evidence. Nevertheless, the broad consensus of clinicians and investigators in this area of inquiry is that the psychotherapies are important and effective tools for treating the psychiatric problems and disorders of late life.

THEORETICAL ISSUES

A major developmental task of aging is finding restitution for the inevitable biopsychosocial losses associated with this stage of the life cycle. Therefore, loss is one of the most prominent issues discussed by elderly patients. For many elderly persons, it is the cumulative effect of multiple repeated losses, before sufficient time has passed for mourning and resolution, that is so devastating. The older person struggles to maintain ego integrity and a sense of purpose in life at a phase of the life cycle often characterized by diminishing coping resources, increasing self-doubt, and the onslaught of assaults and challenges to self-esteem and a positive self-image. The magnitude of an elderly person's reaction to a loss depends to some degree on the amount of pride and emotional investment in the lost function. Beyond lowered self-esteem, elderly adults experience life changes that can damage their sense of security and safety. For some the world becomes uninviting and dangerous, particularly if they have encountered real events such as falls, personal attack, or inaccessible help at a time of need.

Self-worth, security, and independence depend in part on the individual's ability to maintain physical capacities. Impairment of important psychological functions becomes pronounced when frailty or illness forces restricted function and imposes a new, different, and often unacceptable self-image and personal identity.

Psychological adaptiveness in old age varies widely. Vulnerability to states of depleted self-esteem, insecurity, fear, or anxiety is heightened when an individual cannot adapt flexibly and is too wedded to maintaining or retrieving that which must be relinquished. Problems of psychological adaptation and lost self-esteem occur in late life because of such factors as overdependence on social or work roles, overemphasis on the maintenance of control and mastery in all situations, and unresolved problems with self-esteem. Difficulties with self-esteem may be rooted in maladaptive, unmodified beliefs; unconscious issues such as dependency conflicts; or interpersonally generated psychological reactions like unresolved grief.

Many elderly persons adapt quite well to the challenges imposed by negative societal stereotypes and the losses noted above that threaten to erode self-esteem. Several psychological self-protective mechanisms help to maintain a positive self-image and self-esteem. The elderly individual may focus on past successes, discount messages that do not fit with lifelong self-concepts, refuse to apply general myths and misconceptions about aging to the self, and choose to interact with those who provide an ego-syntonic experience. Some selectively screen out unwanted perceptions such as those that challenge impaired cognitive abilities or produce other anxiety. While often self-protective, such selectivity may also distort reality and promote maladaptive denial.

SUPPORTIVE PSYCHOTHERAPY

Frail, institutionalized, cognitively impaired, or chronically ill elderly patients require such psychotherapeutic techniques as support of healthier defense mechanisms, ventilation, and advice and help with acceptance of diminished capacities and increased dependency needs. In addition, they need caring, nurturing environments and assistance of families with their care giving.

The patient's life history and psychodynamics (obtained from the patient or family members) enables the therapist to institute specific supportive therapeutic strategies.

An elderly woman with progressive cognitive impairment and lifelong personality traits of dependency and passivity is anxious and depressed about entering a nursing home. She may adapt well to the home if assigned to a roommate who is nurturing and protective.

Alvin Goldfarb has described a brief, supportive therapy technique for institutionalized, cognitively impaired patients. The therapist promotes patients' foundering self-esteem, sense of control, and safety by permitting them to develop an apparent special relationship with the therapist, who is perceived as a benevolent and powerful figure. The patients believe they have some control over the benevolent physician. This is accomplished in small, subtle ways. For example, the physician elicits the patient's preferences for the frequency of sessions, daily timetables, diet, or socializing and then acquiesces to the patient's wishes, while maintaining a quiet caution about being unduly manipulative. The technique includes weekly, short (15 minutes) visits and gratifying the patient's realistic requests when possible.

LIFE REVIEW OR REMINISCENCE THERAPY

Robert Butler and others have noted the universal tendency of the aging person to reflect on and reminisce about the past. Reminiscence is characterized by the progressive return of memories of past experiences, especially those that were meaningful and conflictual. To varying degrees, elderly patients in therapy reminisce about the past, search for meaning in their life, and strive for some resolution of past interpersonal and intrapsychic conflicts. Life review therapy systematically

enhances this reminiscing process and makes it more conscious and deliberate. The therapist may guide the process by encouraging the patient to write or tape a biography with review of special events and turning points. Techniques include reunions with family and good friends and looking through memorabilia such as scrapbooks or picture albums. This technique has been reported to resolve old problems, increase tolerance of conflict, relieve guilt and fears, and enhance self-esteem, creativity, generosity, and acceptance of the present.

COGNITIVE-BEHAVIORAL THERAPY

In adapting cognitive-behavioral therapy to elderly patients, the therapist is often more active, redirecting the patient from reminiscence to here-and-now issues. The pace of therapy is sometimes slower than with younger adults, and when appropriate, the therapeutic alliance may be enhanced by permitting the patient to adopt a senior, teaching role with the younger therapist. To compensate for specific deficits such as hearing loss or cognitive slowing, information may be presented in different modalities, such as written material or taped sessions for the patient to take home. With older patients, termination is managed gradually rather than abruptly.

BRIEF (TIME-LIMITED) PSYCHODYNAMIC PSYCHOTHERAPY

Brief, or time-limited, psychodynamic therapy can be considered for elderly patients with clearly defined, age-related problems that can be expected to resolve within a brief period of psychotherapy, such as an adjustment disorder, unresolved grief reaction, or new-onset anxiety disorder. Setting a time limit at the beginning of therapy reinforces the patients' confidence in their ability to resolve problems within a circumscribed period of time, focuses and accelerates the therapeutic process, diminishes their fear of dependency, and accommodates limited financial resources. Brief therapy uses the therapist's psychodynamic understanding of the patient and the patient's transference to the therapist to clarify and help resolve the patient's emotional reactions to a current life stress. During brief dynamic psychotherapy the patient often uses the therapist to validate competency and normalcy, which helps to support and restore feelings of mastery and self-esteem. Symptomatic improvement often exceeds the achievement of insight or self-understanding.

INSIGHT-ORIENTED PSYCHOTHERAPY

The basic approach to insight-oriented therapy for older patients who are physically and cognitively intact is largely the same as that for younger patients. However, most clinicians agree that the themes introduced in therapy by elderly patients are more focused on issues of loss, sexual and physical decline, cumulative trauma, fear of pain and disability, decline in self-esteem, and increasing dependency. During longer-term psychotherapy, the condition and circumstances of the older patient evolve and change. The development of physical illness may require medical referral, collaboration with family members, or advocacy on behalf of the patient for community and other support services. Such active intrusions of the patient's real-life problems into the therapeutic process are not commonly dealt with by psychodynamic therapists, but when treating elderly patients, techniques must be adapted. Therapists are most effective when they learn to integrate different therapeutic modalities while maintaining a psychodynamic focus on the therapeutic relationship, the transference, and the goal of inner conflict resolution. Indeed, the necessity to integrate modalities of treatment is one of the hallmarks of the flexibility necessary to treat and manage the evolving therapeutic issues of elderly patients effectively.

TECHNIQUES

In the initial interview the psychiatrist is active and emphatic in exploring patient and family resistance to therapy. The elderly patient is usually interviewed individually before important information and concerns are obtained from the family. Elderly patients usually respond positively to the therapist's listening, empathic, and encouraging attitude. Some patients like to sit close to the therapist and may welcome such spontaneous warmth as a natural touch on the patient's arm. However, the therapist should evaluate each patient's sensitivities about intimacy and not assume that familiarity is welcome just because the patient is old.

Psychologically less-sophisticated patients who are accustomed to talking about their physical complaints when visiting a physician, initially may feel more at ease if the therapist begins with a medically oriented focus. The therapist may then shift the focus to psychological and social problems. Attention is paid to previously overlooked medical problems and medication adverse effects that may contribute to current psychiatric signs and symptoms.

In general, the therapist of the elderly patient is more active, engaging, supportive, and flexible in the therapeutic approach than therapists of younger patients. The therapist must know the normative, age-related changes that occur in the psychological, cognitive-intellectual, and social life of older people; geriatric medicine; psychopharmacotherapy; family therapy; and community resources. In both office and hospital settings, it is especially useful to work with a mental health colleague, such as a social worker, who is knowledgeable about family therapy and community resources.

Realistic, explicit treatment goals are negotiated with the patient and family as appropriate. Patients must feel that they have obtained something from the initial interview (e.g., empathic understanding, or hopefulness) to increase the likelihood that they will return for subsequent sessions. Telephone reminders to selected patients or family members in advance of the next appointment imparts the therapist's interest and improves patient compliance.

The utility of various forms of psychotherapy for specific clinical indications is slowly being investigated in elderly populations. Overall, for major depressive disorder in elderly patients, psychotherapy may be recommended as an important adjunct to pharmacotherapy or electroconvulsive therapy (ECT) and can be recommended as a primary mode of treatment for some patients with depression without endogenous features. The specific techniques used are tailored to the individual patient. A few clinical trials have examined the combined use of psychotherapy and an antidepressant. While results are mixed, evidence suggests that cognitive-behavioral therapy may augment the effectiveness of antidepressant medication. Some studies strongly suggest that behavioral, cognitive, and brief dynamic psychotherapy are all effective treatments for selected elderly patients with depression. A recent meta-analysis of the efficacy of psychotherapeutic treatments for the elderly supported the conclusion that these treatments are more effective than no treatment under placebo-controlled conditions and established the efficacy of psychosocial interventions in both major depressive disorder and subclinical varieties of depression. However, as in younger patients, no modality of psychotherapy shows clear superiority over another in the treatment of geriatric depression. Thus an integrated approach to psychotherapy for geriatric patients is best, based on their need for cognitive-behavioral, dynamic, and interpersonal psychotherapeutic interventions of various durations.

Behavioral interventions for the elderly have been most commonly used with cognitively impaired residents in institutions, but therapists are increasingly using this approach for outpatients.

CLINICAL ISSUES

Indications for Individual Psychotherapy Elderly adults are a heterogeneous population. The continuum ranges from the normative or healthy older patient whose function is retained and who is relatively unimpaired by physical or cognitive decline to the frail elderly patient. Five categories are normative elderly; normative elderly in crisis; independent frail elderly; dependent, physically frail elderly; and dependent, cognitively impaired, frail elderly. Indications for various modalities of individual therapy are strongly influenced by the position of the patient on the normative-frail continuum. Moreover, because capacities change over time, therapeutic approaches for a given patient may also change.

Even patients at an advanced age can benefit from insight-oriented therapy. This technique is more often effective in patients who retain adequate ego strength and the capacity to mourn losses and tolerate regression and have sufficient time and finances. These qualities are generally found in the normative aging group.

With advancing age, crisis is frequently the stimulus for seeking out psychiatric care, often initiated by a family member or caregiver. Individual therapy during crisis focuses on mourning for lost important relationships or personal capacities and working through conflicts over separation, redirecting energy to more realistic goals, redefining relationships especially with regard to accepting appropriate dependency and reestablishing or developing appropriate defenses to enhance mastery, sense of control, and self-esteem. Psychotherapy during crisis is often integrated with medication and environmental therapies. Brief interpersonal psychotherapy and cognitive-behavioral therapy are useful approaches during the crisis, although longer-term supportive or uncovering (i.e., psychodynamically oriented) therapy may evolve.

Behavioral techniques using operant behavioral conditioning have been applied to a wide range of target symptoms in chronically ill elderly patients, including incontinence, pain, insomnia, wandering, socially inappropriate behaviors, level of activity and participation, communication skills, toileting, bathing, and feeding. Environmental cues are manipulated to increase, decrease, or shape behaviors. Careful analysis of environmental consequences, for example, may uncover inadvertent reinforcers of undesirable behaviors. These techniques are useful for institutionalized and frail elderly patients to reduce the excess disability that often

accompanies dementia. An ABC approach to behavioral management of disturbed behavior may be useful. This technique includes accurate neuropsychological evaluation; defining a specific measurable and observable behavior for modification; determining the antecedents (A) to the onset of the behavior and systematically recording them; recording the type, frequency, and duration of the baseline (B) conditions; tailoring a treatment program (A & B) to alter the behavior, which includes caregiver counseling, introduction of reinforcing, and extinguishing techniques, that is, consequences (C); instituting environmental changes and psychotherapy to assist caregivers; and program evaluation. A major deficiency of behavioral interventions in general is their labor intensiveness and the need to maintain the program. Once the interventions are withdrawn, behaviors tend to return to pretreatment levels in these institutionalized populations.

Various forms of individual psychotherapy have shown therapeutic value for states of bereavement. Psychotherapeutic intervention is particularly effective for bereavement complications such as chronicity, depression, or anxiety disorders. Most interventions include some education, support, attention to the relationship with the deceased as it developed before the death and as it has changed since, and supportive assurances of “normality” for bereaved individuals who need to be reminded that their feelings and experiences are not deviant, abnormal, or fixed.

While psychotherapy of various types has been recommended for treatment of anxiety disorders, the data do not support strong recommendations. Recent preliminary data suggest that cognitive-behavioral therapy can be quite effective in older patients with panic disorder, and may thus be considered as an alternative to pharmacological treatment for selected patients or as an adjunct to pharmacological treatments when patients have behavioral symptoms that can be targeted.

Psychotherapy has shown some value in the management of elderly patients struggling with alcohol abuse. Inclusion of cognitive-behavioral training that addresses themes of self-advocacy, self-esteem, and relapse prevention may aid recovery.

Individual psychotherapy is often helpful in the treatment of sexual dysfunction. For example, premature ejaculation usually responds readily to sex education, reassurance, and sex therapy techniques practiced at home alone or with a cooperative partner.

Limitations Elderly adults are underrepresented in psychiatric outpatient clinics (2 to 4 percent of patients seen) and even more so in private psychiatric practice. The effect of psychotherapy, therefore, appears to be limited by the willingness of practitioners, both psychiatric and primary care, to use it for elderly patients. The completion of a series of regular therapy sessions, whether short or long term, is often limited by such practical considerations as difficulties with finances, transportation, or weather conditions.

Raised in an era when shame was associated with psychiatric problems and treatment, some elderly persons feel embarrassed and indignant when their physician or family suggests consultation with a psychiatrist. Many elderly persons as well as primary care physicians and mental health professionals believe that depression and anxiety are expected concomitants of aging or attribute abnormal psychological states to physical illness. The symptoms themselves—depression, apathy, and hopelessness—prevent older persons from seeking treatment. Patients whose ethnoracial backgrounds differ markedly from that of their therapist may have quite different notions and expectations about a visit to a psychiatrist. Many cultures view mental disorder as shameful. Individuals may fear damage to their families' reputations. For example, elderly patients have expressed fear that their grandchildren may be less marriageable if there is mental illness in the family. Beyond cultural barriers, other access barriers limit care for these groups including language incompatibilities, lack of knowledge of where to obtain psychiatric services, economic issues, a tendency to seek care only after a crisis, and fragmented, difficult-to-access, mental health-care services.

Adult children of an aging parent may minimize the parent's psychiatric symptoms, be ambivalent about assuming a parenting role, and may share their parent's negative stereotypic attitudes toward psychiatry. Practical considerations such as geographical inaccessibility to the psychiatrist's office pose additional barriers to treatment. Physical illness or other crises may arise during the course of longer-term psychotherapies, which sometimes limits the use of psychodynamic psychotherapy because the therapist must shift to a more supportive mode during the crisis period.

Cognitive-behavioral interventions in long-term care settings, while often effective when introduced by a consultant, generally fail if the consultant leaves. Front-line staff in these settings usually lack the training to sustain sophisticated cognitive-behavioral programs on their own.

Complications Whatever the form of psychotherapy, the course may be complicated by issues of transference and countertransference. The patient's feelings about therapy will be influenced by education, socioeconomic status, culture, and ethnicity, which can restrict or enhance the impact of treatment.

Transference is age independent (of both patient and therapist), the determinants stemming from the timeless unconscious. However, developmental theory expands the traditional perspectives by demonstrating that the transference derives not only from significant infantile experience but also from internalized objects encountered in adult life, for example, with a spouse or a child.

While there are no age-specific transferences, the psychodynamic impact of life experiences during old age may increase the emergence of certain forms of transference. For example, frailty leads to dependency conflicts and fears of abandonment, emotions that can easily evoke parental transference deriving from the patient's need for a protector or savior. This form of transference emerges readily in dependent, often-institutionalized patients, even when psychotherapeutic contact is relatively brief and intermittent. Similarly, loss of self-esteem and sources of external admiration, so central to aging, promotes idealizing and mirroring transferences. Bereavement, especially of a spouse, or other loss (e.g., sexual outlets or capacity) may lead to eroticized spousal transference in which the patient unconsciously relates to the therapist as an erotic love object who becomes a fantasy substitute for the lost object. This reaction may be distinguished from classical oedipal transference which also may occur in therapy with elderly patients.

Reverse transference describes a phenomenon of geriatric psychotherapy in which the patient apparently experiences and relates to the therapist as a child and takes a parental stance and role toward the therapist in the treatment dyad. Therapists may experience this transference interaction as demeaning or patronizing and feel uncomfortable with the seeming reversal of dominance. Young therapists or residents in training may find this especially challenging. While therapists of elderly patients often comment on reverse transference, it is generally a transient period of therapy that defensively protects the patient from deeper levels of transference-based affect associated with conflicts over dependency.

Somewhat different is *filial transference*, in which patients deal with the therapist as one of their own children, enacting such unconscious conflicts as rage for not fulfilling wishes for caring to which the patient unconsciously feels entitled or guilt deriving from earlier failures in the parental role.

Countertransference responses of therapists are often intense in treatment of elderly patients, sometimes blocking the therapist's willingness to engage the patient in individual psychotherapy. Less experienced therapists or those in training may initially doubt the applicability of their skills to elderly patients, thus reinforcing preexisting uncertainties about the mastery of their craft. The patient's frailty sometimes induces especially strong withdrawal from using psychotherapy because therapists are unwilling to encounter and confront their own hopelessness and fears of abandonment or mortality. Some therapists withdraw because of fear of being engulfed by the patient's demands and dependency. These factors may be compounded by the added therapeutic burdens of working at a slower pace, overcoming sensory deficits, and dealing with cognitive decline.

Countertransference parallels transference reactions in that therapists may develop strong urges to rescue the patient and do too much, reactions that may derive from unresolved guilt and rage toward their own parents.

Therapists must deal within themselves with the repetitive need for therapy to focus on mourning and acceptance of loss. This is in contrast to the more hopeful perspective of ego development and mastery, which is common in psychotherapy of younger adults and easier for therapists to deal with. Treating elderly patients holds the danger that helpless anxiety or despair may emerge in therapists who have not successfully dealt with their own issues of loss.

Therapists of elderly patients are often surprised by the emergence of erotic countertransference during more intensive therapy and may recoil from or deny its presence or feel uncomfortable about aspects of the patient's sexual feelings, thereby depriving the therapy of crucial areas of work.

The sometimes overwhelming effects of cumulative life trauma are painful for therapists, who may retreat into mechanistic or formulistic approaches, inhibiting their own empathic relatedness.

A young therapist was surprised when a difficult elderly woman complained angrily about feeling manipulated and exposed when sent home to write her life story, a technique the therapist had read was suitable for the elderly. In discussion of the case, it became evident that he felt overwhelmed by the patient's painful life and uncertain of what would be helpful. Rather than engaging her in an attempt to understand her feelings, he prematurely employed a cognitive intervention that made her feel infantilized and rejected.

Idealizing countertransference may inhibit the therapist from recognizing the patient's ongoing conflicts; for example, about sexuality, aggression, self-esteem, and grandiosity.

In approaching termination, many elderly patients need to be assured of the therapist's continued future availability should the stresses associated with advancing age precipitate regression, conflict, and decompensation. While some elderly patients terminate successfully, for many, therapy is an intermittent process to be reactivated as needed. Thus it is often helpful to discuss with patients feelings that may prevent their return to therapy such as shame, fear of disappointing the therapist, and desire for independence.

Contraindications There are few contraindications to psychotherapy, although the form of treatment is carefully tailored to the needs of the patient. Psychoanalytic psychotherapy is approached with caution. The therapist first ensures that the patient retains sufficient ego capacity to deal with regression during therapy, to work through adaptive mourning, and to replace lost sources of self-esteem. Even life review therapy may be contraindicated for patients who have realistic guilt about the past or who become overwhelmed when focusing on unmourned or unresolvable past disappointments and losses.

As in younger populations, cognitive-behavioral therapy is indicated for treatment of acute or chronic depressive disorders, concurrent with antidepressant medications when indicated, and for symptoms associated with personality disorder. Mild cognitive impairment secondary to senile dementia or cerebrovascular accident is not a contraindication to cognitive-behavioral therapy, although severe or suicidal depression or severe cognitive decline requires alternative therapies.

Integrated Therapy Integration of psychotherapies often is the most effective way to proceed.

At the age of 70, Mrs. S. was referred for psychiatric assessment because of her distress over her husband's declining physical state secondary to advanced Parkinson's disease. In the initial interview the therapist tried to determine three separate but interrelated factors: the patient's surface conflicts; the aspects of the conflicts that might be amenable to practical, environmental intervention; and the deeper psychological conflicts that were coloring the patient's ability to deal with the situation.

The manifest problems were loss of her husband's companionship and feeling helpless to cope with his illness. For example, he could not lie comfortably in bed and spent most nights sleeping in a living room chair, a situation that she could not tolerate. Mrs. S. feared her own growing incapacity from many years of arthritis and worried about the fate of her husband should she become immobilized.

In approaching the manifest difficulties, the therapist helped Mrs. S. define the specific areas of emotional conflict (especially her grief at the loss of the husband she once knew) and express and ventilate feelings of fearful anticipation, helplessness, and vulnerability. A cognitive therapy approach was used to help her recognize which parts of her feelings were based on unrealistic thoughts. Focusing on her feelings of failure and aloneness, the therapist helped her understand that others had struggled as she was struggling and that her sense of failure was not her fault but an inevitability of the illness. Concurrently, she was encouraged to distinguish her ideas of what was best for her husband from his actual needs.

At the same time, Mrs. S. examined the practical realities posed by her husband's illness and experimented with alternative methods of coping. For example, she gradually came to accept that there was no harm to her husband if he slept in a chair, as long as he felt comfortable. This simple change in perspective, arising from direct explanation and advice from the therapist, relieved much of her tension and guilt and allowed her to sleep more restfully.

The third concurrent component of therapy was an exploration of deeper sources of feeling. This part of therapy revealed the patient's rage at her husband. She had been unaware of how his illness had reactivated her long-standing belief that she had never been understood or cared for as she had wanted. She spoke of her conflicted feelings about her childhood and her lonely, restricted life with which she appeared to cope on the surface, while actually experiencing abandonment and fear. The therapist actively encouraged her reminiscence in the early phases of therapy, attempting to evoke the emotions that accompanied the memories. Gradually, she moved into a phase of meaningful exploration of her inner state. The resulting improved comfort helped her carry on in her caregiving role but also freed her to begin to seek sources of essential nurturance and gratification for herself.

The successful outcome of this treatment probably delayed Mrs. S.'s own decline and enabled her husband to remain safely at home for a much longer time. This case illustrates both the efficacy of psychotherapy and its potential cost-effective and humanitarian outcome.

Goals of Treatment Psychotherapy is most effective when it is used for specific indications with defined goals. Goals are modified to fit the function, cognition, and psychological capacity of each patient. Generally, the greater the frailty, the more modest are the goals of psychological change. Goals for patients able to undertake dynamic psychotherapy include psychological mastery over the past and its unresolved grief; resolution of adult and childhood sources of shame and guilt; resolution of current interpersonal conflict, especially with spouse or adult children; working through conflicts associated with loss of self-esteem and establishing substitute sources for those that had to be relinquished; coming to terms with failure to achieve ideal aspirations; and, at the most advanced level, processing and integrating unconscious conflicts experienced in the therapeutic transference relationship. George Pollock has termed the creativity and energy that may be released by successful resolution of losses *mourning-liberation*.

In crisis therapy, the goal is to reestablish the patient's best premorbid level of function. Often this requires identifying more realistic levels of function. Examples include acceptance of appropriate levels of dependence or controlling excess, psychologically induced morbidity.

For the more debilitated, frail, elderly patients, psychotherapy goals are often symptom-focused behavioral changes. These include focal interventions; for example, dealing with dementia-related behaviors such as aggression, excess withdrawal, anxious agitation, or delusions. With these patients especially, therapy goals include strengthening the roles of family caregivers and fostering a more appropriate, supportive environment for the patient.

ETHICAL ISSUES

Frail, often-institutionalized elderly patients have limited capacity to use the psychotherapies. They are easily viewed by therapists as beyond help. In allocating scarce resources, therapists struggle with whether to expend effort on this population or reserve it for their more vigorous elderly patients. This is a difficult ethical dilemma since the frail elderly patients are the most dependant, helplessly reliant on caregivers, who in turn are burdened with the role of deciding who shall or shall not receive certain forms of care.

While most forms of psychotherapy require the active participation of patients, and therefore their inherent consent, behavioral interventions may be applied to the patient externally. In using behavioral techniques with frail elderly patients, therapists are advised to exercise caution in distinguishing between externally imposed goals of therapy, that is, a focal symptom that is troublesome to staff, and the overall welfare of an often helpless patient whose ability to consent to, and participate in, setting treatment goals may be limited or impaired.

Clinicians are often confronted with ethical concerns over confidentiality and the patient's decision-making capacity. Both family members and therapists may inadvertently infantilize more frail patients and override their right to participate fully in the therapeutic decision-making process. The psychotherapist must carefully evaluate a patient's ability to make decisions. Similarly, the family's right to know about the patient and participate in therapy is carefully balanced by the preeminent right of the competent elderly patient to confidentiality in the therapeutic relationship. In practice this is often a difficult issue for therapists, especially when families are intrusive, demanding, or litigious.

SUGGESTED CROSS-REFERENCES

Additional perspectives on individual psychotherapy may be read in Section 51.6e on ethnic minority issues, Section 51.4l on group psychotherapy of the elderly,

Section 51.4j on cognitive-behavioral of the elderly, and Section 30.1 on psychoanalytic psychotherapy.

SECTION REFERENCES

Arean P, Perri M, Nezu A, Schein R, Christopher F, Joseph T: Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. J Consult Clin Psychol *61:*1003, 1993.

Atchley RC: The aging self. Psychother Theory Res Prac *9:*338, 1982.

Baker F: Psychiatric treatment of older African-Americans. Hosp Community Psychiatry *45:*32, 1994.

Cohen GD: Psychodynamic perspectives on the clinical approaches to brain disease in the elderly. In *Psychiatric Consequences of Brain Disease in the Elderly*, D Conn, A Grek, J Sadavoy, editors. Plenum, New York, 1989.

Goldfarb AI, Turner H: Psychotherapy of aged persons. Utilization and effectiveness of brief therapy. Am J Psychiatry *109:*916, 1953.

Goldstein M: Taking another look at the older patient and the mental health system. Hosp Community Psychiatry *45:*117, 1994.

Hartman C, Lazarus LW: Psychotherapy with elderly depressed patients. Clin Geriatr Med *8:*355, 1992.

Hussian RA, Davis RL: *Responsive Care: Behavioral Interventions with Elderly Persons*. Research Press, Champaign, IL, 1985.

Kahana R: Geriatric psychotherapy: Beyond crisis management. In *Treating the Elderly Psychotherapy: The Scope For Change in Later Life*, J Sadavoy, M Leszcz, editors. International Universities Press, Madison, CT, 1987.

Knight B: *Psychotherapy with Older Adults*. Sage, London, England, 1986.

Lazarus LW: Self-psychology and psychotherapy with the elderly: Theory and practice. J Geriatr Psychiatry *13:*69, 1980.

Lazarus LW, Cohler BJ, Lesser J: Self-psychology: Its application to understanding patients with Alzheimer's disease. Int Psychogeriatr *8:*243, 1996.

Lazarus LW, Groves L: Brief psychotherapy with the elderly: A study of process and outcome. In *Treating the Elderly with Psychotherapy: The Scope for Change in Later Life*, J Sadavoy, M Leszcz, editors. International Universities Press, New York, 1987.

Lazarus LW, Sadavoy J, Langsley PR: Individual psychotherapy. In *Comprehensive Review of Geriatric Psychiatry*, J Sadavoy, L Lazarus, L Jarvik, editors. American Psychiatric Press, Washington, DC, 1991.

Lewis MF, Butler RN: Life-review therapy: Putting memories to work in individual and group psychotherapy. Geriatrics *29:*165, 1974.

Miller M, Wolfson L, Frank E, Cornes C, Silberman R, Ehrenpreis L, Zaltman J, Malloy J, Reynolds C: Using Interpersonal Psychotherapy (IPT) in a combined psychotherapy/medication research protocol with depressed elders. A descriptive report with case vigaettes. J Psychoth Pract Res *7:*47, 1997.

Mossey JM, Knott KA, Higgins M, Talerico K: Effectiveness of psychosocial intervention, interpersonal counselling, for subdysthymic depression in medically ill elderly. J Gerontol A Biol Sci Med Sci *51:*M172, 1996.

Myers WA: *Dynamic Therapy of the Older Patient*. Aronson, New York, 1984.

Myers WA: Transference and countertransference issues in treatment involving older patients and younger therapists. J Geriatr Psychiatry *19:*221, 1986.

*Myers WA, editor: *New Techniques in the Psychotherapy of Older Patients*. American Psychiatric Press, Washington, DC, 1991.

*Nemiroff RA, Colarusso CA, editors: *The Race Against Time*. Plenum, New York, 1985.

*Newton NA, Lazarus LW: Behavioral and psychotherapeutic interventions. In *Handbook of Mental Health and Aging*, JE Birren, RB Sloane, GD Cohen, editors. Academic Press, San Diego, CA, 1992.

Raphael B, Middleton W, Martinek N: Counselling and therapy of the bereaved. In *Handbook of Bereavement: Theory Research and Intervention*, MS Stroebe, W Stroebe, RO Hansson, editors. Cambridge University Press, Cambridge, 1993.

Rechtschaffen A: Psychotherapy with geriatric patients: A review of the literature. J Gerontol *14:*73, 1959.

Renshaw DC: *Seven Weeks to Better Sex*. Random House, New York, 1995.

Reynolds CF III: Treatment of major depression in later life: A life cycle perspective. Psychiatr Q *68:*221, 1997.

Reynolds CF III, Frank E, Houck PR, Mazmudar S, Dew MA, Cornes C, Buysse DJ, Begley A, Kupfer DJ: Which elderly patients with remitted depression remain well with continued interpersonal psychotherapy after discontinuation of antidepressant medication? Am J Psychiatry *154:*958, 1997.

*Reynolds C, Frank E, Perel S, Imber S, Cornes C, Miller M, Mazumdar S, Houck P, Dew M, Stack J, Pollock B, Kupfer D: Nortriptyline and interpersonal psychotherapy as maintenance therapies for recurrent major depression: A randomized controlled trial in patients older than 59 years. JAMA *281:*39, 1999.

*Reynolds CF, Miller M, Pasternak R, Frank E, Perel J, Cornes C, Houck P, Mazumdar S, Dew M, Kupfer D: Treatment of bereavement-related major depressive episodes in later life: A controlled study of acute and continuation treatment with nortriptyline and interpersonal psychotherapy. Am J Psychiatry *156:*202, 1999.

Rice C, Longabaugh R, Beattie M, Noel M: Age group differences in response to treatment for problematic alcohol use. Addiction *88:*1369, 1993.

Rybarczyk B, Gallagher-Thompson D, Rodman J, Zeiss A, Gantz F, Yesavage J: Applying cognitive-behavioral psychotherapy to the chronically ill elderly: Treatment issues and case illustration. Int Psychogeriatr *1:*127, 1992.

Sadavoy J: Psychodynamic perspectives on Alzheimer's disease and related dementias. Am J Alzheimer Care Related Disord Res *6:*12, 1991.

*Sadavoy J, Leszcz M, editors: *Treating the Elderly with Psychotherapy: The Scope for Change in Later Life*. International Universities Press, Madison, CT, 1987.

Scogin F, McElreath L: Efficacy of psychosocial treatment for geriatric depression—a review. J Consult Clin Psychol *62:*1969, 1994.

Silberschatz G, Curtis JT: Time limited psychodynamic psychotherapy with older adults. In *New Techniques in Psychotherapy of Older Patients*, WA Myers, editor. American Psychiatric Press, Washington, DC, 1991.

Small GW: Recognizing and treating anxiety in the elderly. J Clin Psychiatry *58*(Suppl 3):41, 1997.

Solomon K, Szwabo P: Psychotherapy for patients with dementia. In *Memory Function and Aging-Related Disorders*, J Morley, R Coe, R Strong, editors. Springer, New York, 1992.

Steuer J: Psychotherapy with the elderly. Psychiatr Clin North Am *5:*199, 1982.

*Steury S, Blank ML, editors: *Readings in Psychotherapy with Older People*. National Institute of Mental Health, Washington, DC, 1991.

Swales P, Solfvin J, Sheikh J: Cognitive-behaviour therapy in older panic disorder patients. Am J Geriatr Psychiatry *4:*46, 1996.

Teri L, Logsdon R: The future of psychotherapy with older adults. Special issue: The future of psychotherapy. Psychotherapy *29:*81, 1992.

Thompson LW, Gantz F, Florsheim M, Del Maestro S, Rodman J, Gallagher-Thompson D, Bryan H: Cognitive-behavioral therapy for affective disorders in the elderly. In *New Techniques in Psychotherapy of Older Patients*, WA Myers, editor. American Psychiatric Press, Washington, DC, 1991.

Woods R: Psychosocial management of depression. Int Rev Psychiatry *1:*427, 1993.

Yesavage JA, Karasu TB: Psychotherapy with elderly patients. Am J Psychother *360:*41, 1982.

Zisook S, Downs NS: Diagnosis and treatment of depression in late life. J Clin Psychiatry *59*(Suppl):80, 1998.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4J COGNITIVE-BEHAVIORAL THERAPY

GARRETT C. DAUM, M.D.

Psychopharmacological therapy is often considered the initial treatment of choice for psychiatric illness in the elderly. However, psychotherapy or combined treatment may sometimes be more appropriate and more effective. There has been growing recognition that psychotherapeutic techniques developed for treatment of younger patients can be successfully modified for use with elderly patients. One such technique that has demonstrated success with the older patient, especially those with mood disorders or substance use disorders, is cognitive-behavioral psychotherapy.

HISTORY

Cognitive-behavioral psychotherapy is actually a combination of cognitive and behavioral treatment approaches with considerable overlap in theoretical background and practical implementation. In the 1970s Peter Lewinsohn described a behavioral treatment approach to depression while at about the same time Aaron Beck developed his seminal work in cognitive therapeutic techniques. By the early 1980s these treatment approaches had been modified specifically for use with older depressed patients in large part through the work of Gary Emery, Dolores Gallagher, and Larry Thompson. These techniques have since been applied to patients with diagnoses beyond depression, in areas such as geriatric substance abuse, obsessive-compulsive disorder, and pain management.

THEORETICAL ISSUES

Gerontologists have consistently found greater individual variability, cognitive slowing, and greater correlation of physical and psychological processes in elderly adults than in younger adults. Aging requires adapting to inevitable change and, for many, coping with a succession of losses that come at a time of diminishing resources. Common presenting problems that can lead to depression, anxiety, or substance abuse include changes in lifestyle with loss of societal and financial status, retirement and job loss, relationship problems and loss of familial and social attachments with resultant social isolation, and loss of physical status through somatic changes and illness. Cognitive-behavioral therapy has theoretical advantages for use with elderly adults. Its time-limited nature attends to the limited resources available to many elderly adults. It engages patients by focusing on immediate and relevant concerns and by using concrete tasks. It directly challenges the negative beliefs and stereotypes about aging that may deepen pessimism about treatment success. It can be integrated with other treatment modalities, and, it can counteract inactivity, social withdrawal, and apathy especially common in patients with depression. Further, it has measurable treatment goals and provides individuals with an educational opportunity to develop skills and coping strategies useful for a wide variety of stresses encountered in older life.

TECHNIQUES

Cognitive-behavioral therapy with the older patient is suitable for both individual and group work. The basic principles and techniques used are similar to those developed for younger patients, with the similar goal of identifying and correcting dysfunctional thoughts, beliefs, and self-defeating behavior.

Modifications to traditional techniques have been recommended by researchers in the field to accommodate the unique set of variables presented by persons in this age group. Adaptations suggested for the elderly adult include the need to assess and address potential negative stereotypes about psychotherapy (“I'm not crazy”), presenting treatment as a way to learn new strategies to cope with current life stressors, the need for a thorough psychosocial assessment and medical history to understand the many potential influential factors operating in this age group, the need for the therapist to maintain the structure of the therapy session more actively and prevent patients from reminiscing and digressing from the current focus or becoming overly inclusive in recalling events, and conducting therapy at a slower pace with more repetition of important ideas. Therapists should also be aware of sensory deficits in their patients and make necessary adjustments (e.g., by presenting material using multiple means such as verbal instruction, audio and written handouts, and use of posterboards).

CLINICAL ISSUES

Cognitive-behavioral therapy has been most extensively studied for treatment of depression in cognitively intact, motivated elderly adults. Application of this therapy to older patients suffering from substance abuse, anxiety disorders, personality disorders, and psychosis has been less extensively studied. Additionally, cognitive-behavioral therapy has been used in the treatment of outpatients with Alzheimer's disease (with caregivers included in the treatment process), in geriatric crisis intervention, and as adjunct therapy on a geriatric rehabilitation unit. Behavioral techniques have been found helpful in the treatment of specific behavioral symptoms in cognitively impaired nursing home residents and have been applied to other specific symptoms including insomnia, pain, and chronic obstructive pulmonary disease. When appropriate, cognitive-behavioral therapy adapts well to adjunctive use with other therapeutic modalities such as psychopharmacology, dynamic psychotherapy, and 12-step support groups.

Contraindications to cognitive-behavioral therapy are similar to those of other psychotherapeutic techniques, but cognitive capacity must be assessed especially in terms of executive functioning and ability to understand and retain new material, though cognitively impaired individuals may benefit from therapy geared more toward behavioral strategies. Psychotic symptoms are a relative contraindication, depending on the nature and severity of the symptoms.

Based on limited controlled-research data, some general points can be made concerning successful treatment outcomes. Less successful treatment outcomes have been associated with depressed patients with more severe illness, those with a diagnosis of personality disorders, significantly stressful life events, and patients with a limited social support network. Patients with greater initial motivation to treatment appear to have better outcomes.

> Mr. D., a 62-year-old male veteran with a history of more than 10 years of intravenous heroin and amphetamine dependence, enrolled in a cognitive-behavioral treatment group for elderly substance abusers. Mr. D. had been living on the streets for many months and was plagued by dysfunctional thoughts concerning his self-worth and guilt over continued using of drugs and was very pessimistic about his chance at successful treatment, stating “I'm too old, it's too late for me to change.” Mr. D. was able to engage in treatment and began to develop socialization skills with the older clients in the group. Mr. D. was able to identify and begin to correct his negative cognitions and understand his drug using through the A-B-C chain model of behavior, consisting of antecedents (situations, thoughts, feelings, cues, and urges), behavior (drug using), and consequences. Through a combination of cognitive-behavioral techniques including thought-stopping procedures, anger-tension management coping strategies, and role rehearsals of at-risk situations Mr. D. was able to remain abstinent. As his understanding and coping skills improved, so did his self-confidence and self-esteem. He was eventually able to obtain full-time employment and move into his own apartment.

Goals should be well defined at the start of treatment. They should be mutually agreed upon, realistic, and measurable. Progress of therapy in terms of understanding and learning concepts as well as clinical improvement should be periodically reassessed.

RESEARCH AND EVALUATION

The overall paucity of well-designed research studies, the inherent limitations and difficulties of psychotherapeutic research with older patients, and the

methodological differences between studies make many conclusions preliminary. Studies varied substantially in terms of referral source, treatment duration, individual versus group treatment, attrition rates, and marital status. Additionally, information was usually limited in terms of patients' psychiatric history and comorbid physical conditions. Nonetheless some general points can be made. Overall, cognitive-behavioral therapy is more effective than delayed treatment or wait-list controls. As in younger depressed patients, group and individual cognitive-behavioral therapy appear to have similar efficacies. In addition, efficacy rates are similar in comparisons between older and younger age groups. There are mixed results when psychodynamic techniques are compared with cognitive-behavioral therapy; some studies report no significant differences in improvement while others report more favorable results with cognitive-behavioral therapy. Additionally, one study reported a lower relapse rate after 1 year in the cognitive-behavioral therapy group than in the group receiving psychodynamic therapy. The few studies of cognitive-behavioral therapy with elderly alcohol-abusing patients have demonstrated clear benefit for many patients. Though similar levels of reduction on standardized depression scales are found when studies using cognitive-behavioral treatment are compared with those involving psychopharmacological treatment, differences in the initial severity of illness and treatment duration make comparisons of this type very problematic.

SUGGESTED CROSS-REFERENCES

Mood disorders are covered in Chapter 14 and mood disorders in the elderly in Section 51.3d. Substance-related disorders are covered in Chapter 11 and substance abuse in the elderly in Section 51.3h. Behavior therapy is extensively covered in Section 30.2 and cognitive therapy in Section 30.6.

SECTION REFERENCES

Abraham IL, Onega LL, Reel SJ, Wofford AB: Effects of cognitive group interventions on depressed frail nursing home residents. In *Depression in Long Term and Residential Care: Advances in research and treatment,* RL Rubinstein, MP Lawton, editors. Springer, New York, 1997.

Beutler LE, Scogin F, Kirkish P, Shretlen D, Corbishley A, Hamblin D, Meredith K, Potter R, Bamford CR, Levenson AI: Group cognitive therapy and alprazolam in the treatment of depression in older adults. J Consult Clin Psychol *55:*550, 1987.

Blackburn IM, Bishop S, Glen M, Whalley L, Christie J: The efficacy of cognitive therapy in depression: A treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. Br J Psychiatry *139:*181, 1981.

Brok AJ: A modified cognitive-behavioral approach to group therapy with the elderly. Group *21:*115, 1997.

Carmin CN, Pollard CA, Ownby RL: Obessive-compulsive disorder: Cognitive behavioral treatment of older versus younger adults. Clin Gerontol *19:*77, 1998.

Cook AJ: Cognitive-behavioral pain management for elderly nursing home residents. J Gerontol B Psychol Sci Soc Sci *53:*51, 1998.

Dupree LW, Broskowski J, Schonfeld L: The Gerontology Alcohol Project: A behavioral treatment program for elderly alcohol abusers. Gerontologist *24:*510, 1984.

Gallagher D, Steffen A: Comparative effects of cognitive-behavioral and brief psychodynamic psychotherapies for depressed family caregivers. J Consult Clin Psychol *62:*543, 1994.

Gallagher D, Thompson L: Effectiveness of psychotherapy for both endogenous and nonendogenous depression in older adults. J Gerontol *38:*307, 1983.

Gallagher DE, Thompson L: Treatment of major depressive disorder in older adult outpatients with brief psychotherapies. Psychother Theory Res Pract *19:*482, 1982.

Goldwasser AN, Auerbach S, Harkins S: Cognitive, affective, and behavioral effects of reminiscence group therapy on demented elderly. Int J Aging Hum Dev *25:*209, 1987.

Jarvik LF, Mintz J, Steuer J: Treating geriatric depression: A 26-week interim analysis. J Am Geriatr Soc *30:*713, 1982.

Koder D: Treatment of anxiety in cognitively impaired elderly: Can cognitive-behavioral therapy help? Int Psychogeriatr *10:*173, 1998.

*Koder D, Brodaty H, Anstey K: Cognitive therapy for depression in the elderly. Int J Geriatr Psychol *11:*97, 1996.

*Lazarus LW, Sadavoy J: Individual psychotherapy. In *Comprehensive Review of Geriatric Psychiatry,* ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

Lincoln NB, Flannaghan T, Sutcliffe L, Rother L: Evaluation of cognitive behavioral treatment for depression after stroke: A pilot study. Clin Rehabil *11:*114, 1997.

Lopez MA, Mermelstein R: A cognitive-behavioral program to improve geriatric rehabilitation outcome. Gerontologist *35:*696, 1995.

Meador KC, Davis CD: Psychotherapy. In *Textbook of Geriatric Psychiatry,* ed 2, EW Busse, DG Blazer, editors. American Psychiatric Press, Washington, DC, 1996.

Mintz J, Steuer J, Jarvik L: Psychotherapy with depressed elderly patients: Research considerations. J Consult Clin Psychol *49:*542, 1981.

Morin CM, Kowatch RA, Barry T, Walton E: Cognitive-behavior therapy for late-life insomnia. J Consult Clin Psychol *61:*137, 1993.

Morris RG, Morris L: Cognitive and behavioural approaches with the depressed elderly. Int J Geriatr Psychiatry *6:*407, 1991.

Parham IA, Priddy J, McGovern T, Richman C: Group psychotherapy with the elderly: Problems and prospects. Psychother Theory Res Pract *19:*437, 1982.

Rush AJ, Beck A: Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. Cognit Ther Res *1:*17, 1977.

Sadavoy J: Integrated psychotherapy for the elderly. Can J Psychiatry *39:*S19, 1994.

Scogin F, Jamison C, Gochnauer K: Comparative efficacy of cognitive and behavioral bibliotherapy for mildly and moderately depressed older adults. J Consult Clin Psychol *57:*403, 1989.

Scogin F, McElreath L: Efficacy of psychosocial treatments for geriatric depression: A quantitative review. J Consult Clin Psychol *62:*69, 1994.

Steuer JL, Hammen C: Cognitive-behavioral therapy for the depressed elderly. Cognit Ther Res *7:*285, 1983.

Steuer JL, Mintz J, Hammen C: Cognitive-behavioral and psychodynamic group psychotherapy in the treatment of geriatric depression. J Consult Clin Psychol *52:*180, 1984.

Teri L, Gallagher-Thompson D: Cognitive-behavioral interventions for treatment of depression in Alzheimer's patients. Gerontologist *31:*413, 1991.

*Teri L, Gallagher-Thompson D, Thompson LW: Cognitive-behavioral therapy with depressed older adults. In *Diagnosis and Treatment of Depression in Late Life*, LS Schneider, CF Reynolds III, BD Lebowitz, AJ Friedhoff, editors. American Psychiatric Press, Washington, DC, 1994.

Teri L, Lewinsohn P: Individual and group treatment of unipolar depression: Comparison of treatment outcome and identification of predictors of successful treatment outcome. Behav Ther *17:*215, 1986.

*Thompson LW: Cognitive-behavioral therapy and treatment for late-life depression. J Clin Psychiatry *57:*29, 1996.

Thompson LW, Gallagher D: Efficacy of psychotherapy in the treatment of late-life depression. Adv Behav Res Ther *6:*127, 1984.

*Thompson LW, Gallagher D, Breckenridge J: Comparative effectiveness of psychotherapies for depressed elders. J Consult Clin Psychol *55:*385, 1987.

Thompson LW, Gallagher-Thompson D: Psychotherapeutic interventions with older adults in outpatient and extended care settings. In *Depression in Long Term and Residential Care: Advances in Research and Treatment*, RL Rubinstein, MP Lawton, editors. Springer, New York, 1997.

Wilson KCM, Scott M, Abou-Saleh M, Burns R, Copeland J: Long-term effects of cognitive-behavioural therapy and lithium therapy on depression in the elderly. Br J Psychiatry *167:*653, 1995.

Zeiss AM, Breckenridge JS: Treatment of late life depression: A response to the NIH Concensus Conference. Behav Therapy *28:*3, 1997.

Zerhusen JD, Boyle K, Winslow W: Out of the darkness: Group cognitive therapy for depressed elderly. J Psychosoc Nurs *29:*16, 1991.

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4K INTERVENTIONS AND CONSULTATION WITH FAMILIES OF OLDER ADULTS

DEBORAH A. KING, PH.D., CLEVELAND G. SHIELDS, FH.D., AND LYMAN C. WYNNE, M.D., FH.D.

Clinical gerontologists recognized long ago the importance of addressing the mental health problems of older adults within the context of their social world. More than three decades of research has debunked the myth that older adults (i.e., those over 65 years of age) lead lonely, isolated lives devoid of meaningful contact with adult offspring or other family members. Most elderly individuals live independently but maintain regular contact with family. Four out of five have at least one adult child, and when health problems arise, the adult children are the most common caregivers. A small minority (less than 5 percent) of those over 65 live in nursing homes, but a great percentage of them still have regular contact with family members.

Family therapy with older adults is the process of bringing two or more family members together (often from multiple generations) to address problems encountered in the later stages of the family life cycle. Within this developmental context, the goal of therapy may include any or all of the following: (1) to educate and facilitate communication about the unique medical, psychological, and social challenges of aging as they affect the elderly person and the family, (2) to improve the family's ability to negotiate successfully the normative stresses of aging, and (3) to improve the family's ability to solve major problems associated with late-life medical, neurological, or psychiatric disorders.

HISTORY AND THEORETICAL ISSUES

The emerging field of late-life family therapy has its roots in two distinctive but overlapping theoretical approaches to the understanding of human behavior and behavioral problems. Family systems theory maintains that the family functions as a network of interconnected elements so that change in any one individual's status or behavior will affect the status or behavior of other members of the network. Family life cycle theory views the family system as moving through time, encountering both predictable, normative transitions and unique, family-specific challenges. Combining these perspectives, the late-life family therapist interprets the behavioral or emotional problems of an older adult as indications that the family system is having difficulty negotiating a transition or crisis in the unfolding family life cycle. The family therapist's task is to help family members organize or reorganize their resources in a manner consistent with the late-life challenge they are currently confronting.

The family life cycle has been conceptualized as containing successive stages, each with its own specific set of developmental tasks and emotional processes. Most conceptualizations developed to date are unsatisfactory in that they focus on earlier stages of the life cycle during which the family is forming, raising children, and launching children, with little development or differentiation of the later stages of the cycle, such as adjustment to retirement, attainment of grandparenthood, coping with loss and illness, and passing on the family legacy. With advances in medical technology, elderly individuals may actually live more years after their children have left home than they lived when their children were in the home. As well, most life cycle theories feature a traditional idea of family that includes a husband and wife who have children and remain married for life. This stereotypical depiction does not capture the true heterogeneity of contemporary American family life with its childless families, blended families, families of single mothers, gay couples and parents, or other varieties of family experience. In many instances, family therapists must adapt the traditional life cycle model to the particular population and culture with whom they are working.

Other specific approaches to family therapy with older adults have been developed that build upon systemic and life-cycle themes. The contextual approach emphasizes family ethics and seeks to help balance the intergenerational “ledger” so that patterns of caregiving are congruent with the family's history and resources. Behavioral and cognitive-behavioral approaches work with behavioral contingencies in the family and the correction of cognitive distortions. The symbolic-experiential approach emphasizes the family's perceptual construction of who is in and who is out of the family, to help them adjust to changes in family structure caused by major illness and death.

The strength-vulnerability model of health and illness in older adults also builds upon systemic and life cycle themes by postulating that the mental health or illness of individual family members is determined over time by the interaction of individual and family life cycle processes. The model focuses on strengths and vulnerabilities at the level of the individual older adult (e.g., biological health, personal themes, and individual life history) as well as health-enhancing and risk factors at the larger levels of family and other social systems (e.g., family events, themes, relational processes). Central to the strength-vulnerability model and its clinical applicability is the epigenetic theory of family relational development. According to this theory, *attachment and caregiving* is the bond of affection between family members that must be present as a foundation for the development of other interpersonal processes in the family. *Communication* is the development of shared meaning between family members. This process includes sharing important themes and content as well as constructive communication of affect. *Problem solving* is the ability to develop solutions to problems and carry them out cooperatively. Implicitly or explicitly, this process involves the delineation of family roles as they pertain to decision making and authority. *Mutuality* is the long-term commitment to the relationship and the ability to reinvent the relationship over time in the face of life-cycle transitions. These processes develop sequentially in families, so that some attachment and caregiving is necessary for effective communication. Likewise, some effective communication is needed to successfully and collaboratively solve problems. Finally, effective problem solving is a prerequisite to the ability to reorganize and reinvent with mutual understanding those family relationships that accelerate during periods of crisis and life change. The epigenetic pattern is actually circular: if crises are met with successful mutuality, attachment and caregiving will take a new and strengthened form. With unsatisfactory resolution of change, the bedrock of attachment and caregiving will be undermined.

TECHNIQUES

The late-life family therapist must frame interventions that are developmentally congruent with the family's stage of relational and life cycle development. Thus, careful assessment is a necessary precursor to intervention.

Assessment During the initial sessions the therapist evaluates the family's unique strengths and vulnerabilities, including at a minimum the following key points.

Life cycle transitions faced by the family and by individual family members; for example, retirement, grandparenthood, loss of spouse, onset of chronic or degenerating illness

Individual roles assumed by family members; for example, caregiver, breadwinner, matriarch or patriarch, scapegoat

Family relational processes; that is, the quality of attachment or caregiving bonds, communication, and problem solving

Ethnic or cultural factors impacting the family's functioning or the therapist's own reactions to the family

Intervention When initially working with a family, several levels of intervention are considered, each reflecting work at successively deeper strata of family process. These levels should be used as general guidelines for considering the type and intensity of treatment appropriate for a particular family rather than as discrete stages of treatment to be followed in lockstep fashion with every family.

Level One: Psychoeducation and Community Referral Families who exhibit solid attachment bonds, open communication, and good problem-solving skills congruent with their stage of life cycle development do not require ongoing therapy but instead benefit from brief education and consultation regarding a specific

late-life problem. For example, when an older adult develops a new medical or neurological disorder such as Alzheimer's disease, families benefit from information specific to the disorder and its management as well as referral to community agencies and support groups. Usually these families can then organize themselves for effective problem solving pertaining to the care and support of the elderly person. In the case of chronic, deteriorating illnesses such as Alzheimer's disease, however, the family may benefit from brief, intermittent contact with the therapist at those times when the disease worsens. As pleasant and rewarding as these families are to work with, the family therapist must recognize that typically they do not need regular, ongoing professional support.

Level Two: Restoring Effective Problem Solving Often families have solid, positive bonds of attachment with good communication skills and a history of effective problem solving, yet they are temporarily overwhelmed by the intensity and unfamiliarity of a late-life problem, such as severe illness or the declining functional ability of a previously capable older adult. In these cases, effective problem solving may be temporarily stymied because of paralyzing grief reactions or sudden disequilibrium in the previously stable family structure. The family may need to reorganize the lines of authority and decision making to bring them into accordance with their life-cycle transition. For example, when an elderly matriarch or patriarch experiences decline in health or cognitive capacity, the spouse and adult offspring may be unaccustomed to assuming a decision-making role. After receiving clear and direct information about the status of the elder and having an opportunity to grieve about the changes in the once-powerful individual, adult offspring at this level of development can mobilize their resources to solve family problems. Problems of financial management are a common source of family conflict. Adult offspring may believe or have evidence that their parent or parents are mishandling diminishing resources, but the elder may see the problem quite differently.

Mr. and Mrs. C. had just celebrated their fiftieth wedding anniversary when the family started to notice changes in Mr. C.'s behavior. Typically an outgoing and engaging man who “always liked to be in the driver's seat” in regard to most matters of family and civic life, he began to defer decisions about financial matters to his wife and stopped taking the lead in various church activities and fund raisers. The couple had two twin daughters who were married and raising their own children. Despite their busy lives, the daughters maintained close ties with their parents and visited frequently. In fact, they were both inclined to say that their father was the most important man in their lives in terms of giving them guidance and support. As Mr. C. became more reclusive over the next year, they frequently discussed the situation with their mother but never with their father. All of the family members were afraid to discuss their concerns with Mr. C., as they were concerned about hurting his feelings or damaging his pride. However, they finally sought consultation with the family doctor after Mr. C. had gone off for a walk one day and was discovered by neighbors, tearfully trying to find his way back home. Medical workup failed to reveal any reversible medical causes for Mr. C.'s swiftly declining cognitive functions, and a provisional diagnosis of Alzheimer's disease was made. The family was referred to a local memory disorders clinic for counseling and the opportunity to learn about medications that might help slow Mr. C.'s cognitive decline. During an initial meeting, they were given informational material about Alzheimer's disease, antidementia medications, and support groups in town. However, they declined further meetings because Mr. C. said it was unnecessary and they were reluctant to challenge his opinion.

One month later, they called the clinic requesting an urgent appointment because Mr. C. had been involved in an automobile accident. Only Mrs. C. and the two daughters came to the appointment. All were distraught, saying they did not know what to do or how to handle Mr. C. who always had made decisions for himself and everyone else in the family. Mrs. C. disclosed that she was now being treated for depression and felt unable to summon the energy needed to keep watch over her husband. The therapist listened to their concerns and encouraged each of them to share her feelings about seeing her beloved family member change so much. Eventually, the therapist gently suggested that it was necessary for Mr. C. to retire as active head of the family and hand the baton of authority on to the next generation. One of the daughters stated that she had been thinking of the potential benefits of assuming power of attorney for her parents but was frightened to bring this up for fear it would upset her parents. Mrs. C. reassured her daughters that she appreciated their helping out in any way possible. They spent the rest of the session planning a family meeting during which they would gently confront Mr. C. with their concerns. The therapist encouraged them to frame the problem as needing to have him consider treatment so that he could help them cope with the difficulties posed by his illness.

Over the next 2 months, they met with the family therapist three times to get support and problem solve as a family about how to handle Mr. C.'s illness. Mr. C. himself attended two of these sessions and eventually agreed to give up driving and take some medication that might slow the progression of his dementia. He also agreed to turn the management of the couple's finances over to his daughters. This represented a major shift in the authority structure of the family, with the daughters beginning to play a major role in the caretaking of the older generation. After these initial sessions during which they developed new roles for decision making, the family was able to communicate and problem solve effectively regarding the care and management of Mr. C.'s rapidly progressing dementia. However, they met with clinic staff several times for information, consultation, and support over the course of the illness, especially during periods when Mr. C.'s condition worsened and required new behavioral strategies or, ultimately, a higher level of care (i.e., when he became agitated and aggressive, when his medical and behavioral problems became so severe that the family needed to seek placement).

Level Three: Improving Patterns of Communication Some families who care deeply about each other and have solid bonds of attachment communicate little about serious problems facing the family. Some may have a history of arguments and resentment stemming from past crises. Other families have been able to function with little overt communication in earlier stages of the life cycle because they did not experience major loss, medical illness, or cognitive decline. When faced with the tremendous challenges that come with aging, they cannot develop a common understanding of the problem and thus cannot problem solve effectively. Families at this level benefit from active modeling of open communication as well as active structuring of negative or critical communication. Basic techniques of communication training such as active listening can greatly benefit these families.

Level Four: Improving Attachment and Caregiving Bonds Families with negative, weak, or ambivalent bonds of attachment typically have a history of conflict, marginal functioning, and insufficient nurturance between family members. In some of the most difficult cases, adult offspring who were abused or neglected as children are struggling to care for their former abuser. One of the most damaging mistakes a late-life therapist can make at this level is intervening on the level of education, problem solving, or communication without addressing the negative or ambivalent attachment bonds. These families are at risk for elder abuse, and in some cases professional caregivers must be brought in or referral must be made to a long-term care facility. In many instances, however, the technique of family life review can be used to induce empathy for the elder's position, thereby decreasing negativity and possibly forging more-positive bonds of affection between the generations.

Family life review is the process of bringing multiple generations of family members together to weave a coherent intergenerational story. Typically, this process is facilitated by having the family construct a *family genogram*, which is a graphic, multigenerational picture (family tree) that encapsulates family history, structure, and relationships over time. The therapist encourages the family to begin with the story of the elders' families of origin, gradually working down through the themes and experiences of subsequent generations. Learning about the elder's upbringing and early family circumstances or struggles often gives younger family members a different, more empathic perspective on the elder. Some offspring or grandchildren can then establish an adult relationship with the elder rather than continuing to relate solely from the vantage point of children who have their own needs for nurturance and caretaking.

Mr. D. was an 67-year-old widowed, retired salesman who was admitted to an acute psychiatric ward following a near-lethal suicide attempt by overdose on pain medication. He was found unconscious by a neighbor, having written a note to his children giving them information about his bank accounts. On the hospital unit, he was quite withdrawn and depressed, refusing to accept treatment and giving little information to the therapist except to mutter occasionally that he was “worth more dead than alive.”

Telephone calls were made to his two children (a 45-year-old son and a 42-year-old daughter), who immediately came to meet with the treatment team. (Mr. D. refused to attend the first meeting.) They sadly acknowledged that they had seen their father much less in the past year, as their lives were quite full with their own jobs and families. As well, they noted that they had never been as close to their father as they had been to their mother, who died 5 years earlier of lung cancer. When asked to describe their relationship with their father, they revealed his history of alcohol abuse and described their childhood fears of his unpredictable violent rages, during which he would become loud and verbally abusive. Although he had stopped drinking long ago and no longer treated them aggressively, they viewed him as a solitary man who had lost interest in his extended family since the death of his wife. The daughter volunteered that their mother had died from lung cancer 5 years earlier and they were not told of her diagnosis or the specifics of her illness until after her death, supposedly so they would not worry. When the therapist noticed the tears in the daughter's eyes and asked if they had ever really talked about her death as a family, the daughter began to weep quietly.

Recognizing the ambivalent bonds of attachment and closed communication style of the family, the therapist decided not to move immediately into the potentially overwhelming process of family grieving. Instead, attempts were made to get Mr. D. to attend family sessions to “get some family history that will help us learn how better to help you and your family cope with all that's been happening.” Mr. D. agreed to five sessions to be held over the course of the next 3 weeks, with the understanding that the purpose of the meetings was “to help his children” cope with the changes in the family. (The therapy team suggested that the purpose be framed this way to help Mr. D. save face in the family.) The initial sessions were spent conducting a life review, during which Mr. D. was asked to tell his children about his own family of origin. For the first time, he revealed to his son and daughter that he himself had been physically abused by his father. They were moved by their father's story and, for the first time, were able to see him as a vulnerable person who had been deeply wounded as a child. Rather than the powerful, frightening figure who would rage at them during his intoxicated episodes, they could imagine him as a small boy cowering under the blows of his own father.

During the third session, they were able to confront Mr. D. with their own childhood experiences of his drunken rages. After some initial distress during which he attempted to explain away his previous behavior, the therapist helped him ask for their forgiveness. At this point he began to weep quietly and said that he had also asked their mother for forgiveness before she died. As the family shared remembrances of Mrs. D., they were finally able to grieve her loss as a family. In the fifth and final family meeting, they were ready to turn to the issue of Mr. D.'s depression and suicide attempt. Mr. D. described how purposeless and alone he felt after his wife died. He had very few friends and missed seeing his children and grandchildren. Yet he felt that he was of no use to anyone anymore and did not want to “be a drain” on his children. The children responded emotionally that they were not ready to lose both of their parents. With support from the therapist, they were able to convince their father to comply with the recommendations for a trial of antidepressant treatment and supportive counseling sessions.

CLINICAL ISSUES

Indications Clinical experience and research indicate that most serious mental health problems of older adults are more successfully addressed when family members are involved in the treatment. Although caution and clinical judgment must be exercised as to the type and timing of family intervention, substantial evidence indicates that psychiatric problems worsen or resolve less completely when conflictual or chaotic family environments are not addressed in the treatment. Research shows that family factors may be especially important in the following clinical conditions.

Medical or Neurological Illnesses Requiring Caregiving The provision of regular, ongoing care for an elder with chronic or degenerating illness such as Alzheimer's disease can cause serious emotional distress in the caregiver. Research with families of Alzheimer's patients indicates that a combination of individual and family therapy can decrease the level of caregiver depression and delay nursing home placement of the patient. Effective interventions for caregiving families typically focus on a combination of goals, including increasing knowledge about the illness, developing problem-solving techniques for managing difficult patient behaviors and the caregiver's associated emotions, and increasing the caregiver's sense of support.

Research examining the problem-solving conversations of caregiving families reveals that certain patterns of communication can contribute to caregiver depression. For example, negative affect expressed by family members is directly related to the depression level of the caregiver. Depressed caregivers give more empathic responses to family members than less depressed caregivers do, and these empathic responses often engender angry comments from the family members. These findings suggest that depressed caregivers attempt to provide support to other family members but receive angry responses in return.

Late-Life Depression and Suicide Depression is one of the more common and functionally disabling psychopathological disorders of late life. For men, the risk of suicide increases dramatically with age and is often associated with major depressive disorder. Research with mixed-age samples of adults underscores the relevance of family relational factors in the etiology and treatment of depression and suicide. Negative patterns of communication (typically assessed in terms of high levels of criticism) and poor problem-solving skills have been associated with relapse and poor recovery from depressive disorders as well as with increased suicidal behavior. Although less research has focused specifically on older adults, the research that does exist associates higher levels of critical communication and emotional overinvolvement in the adult offspring with higher rates of depressive relapse in the older adult and lower rates of complete recovery. Additionally, family strain is associated with suicidal behavior in depressed older adults.

When viewed from a family-systems, life-cycle perspective, high levels of critical communication and family strain indicate that the family is having difficulty negotiating later life-cycle transitions such as declining physical or mental health in the older generation. Anecdotal clinical evidence indicates that many depressed older adults become suicidal at a normative juncture or transition point in the family life cycle when the elder needs greater closeness and more support from adult offspring because of significant interpersonal losses, change of role status, or physical decline. Yet the younger generations are frequently quite busy managing the needs of their own families or marriage. This developmental crisis occurs gradually in many families but abruptly in others; for example, when a previously capable elder is diagnosed with terminal or deteriorating illness. In these instances, the family's relational development (i.e., their ability to communicate effectively and problem solve successfully) may lag behind their stage of structural development (i.e., when the older generation can no longer make primary decisions). Older individuals can be at increased risk for suicide if they remain isolated or alienated from adult offspring when they are experiencing a developmental imperative to be connected and somehow useful to younger and future generations. If little contact or open communication exists between the generations, the suicidal behavior may represent a means of communicating despair, bringing the intergenerational family together, or both. In such a closed system, elders may come to view themselves as expendable at best or burdensome at worst. The suicidal act may be perceived as the only way to help their successors; that is, these elders perceive it as a means of relieving the family of the burden of having to support or take care of them. Unfortunately, these clinical impressions have not been validated with empirical research, and there are currently no empirically tested models of family treatment designed specifically to address the problem of suicidal older adults.

ETHICAL ISSUES

Elder Abuse Community-based studies indicate that between 3 and 6 percent of individuals over the age of 65 in the United States and other Western countries have experienced abuse, neglect, or both, most often at the hands of a family member. Many states have mandatory reporting laws for elder abuse as well as state- or county-funded protective service programs that work with older adults who have been abused. Family therapists must know the reporting laws in the state where they are practicing and watch for signs of abuse or neglect. Although estimates indicate that one third of actual cases of child abuse are reported, only one of six cases of elder abuse is reported. Older adults are often reluctant to disclose the problem, for fear of retaliation from the abusing family member.

Elderly individuals who suffer from depression are inclined to view themselves as negative and burdensome, and they may actually believe that they deserve maltreatment. There is a need for increased professional awareness of the problem and its warning signs. Research indicates that elders who are cognitively or functionally disabled are at particular risk for abuse, and some who suffer severe disability are unable to report maltreatment. Advanced age, physical and economic dependency on caregivers, isolation, intergenerational conflict, alcoholism (in the elder or caregivers), and mental illness (in the elder or caregivers) have all been associated with increased risk of abuse. Thus, the family therapist must be alert for signs of abuse or neglect, including repeated, unexplained bruises or injuries, malnutrition, fearfulness around particular family members, or, in the case of elders who are being financially exploited, a discrepancy between their ability to pay for personal commodities and their reported level of income. Although the therapist may fear offending the family members or the elder by inquiring about such issues or referring the family to the local adult protective service, most families who are not abusive welcome careful and comprehensive review of the elder's situation. In some

instances, it is useful to frame the referral to the adult protective services as a means of getting access to useful support services in the community.

Although there is some literature on effective community support programs for abused elders, the field generally lacks outcome studies of family therapy for elder abuse or neglect. In part, this is because many of these families have such severely damaged bonds of attachment that caretaking must be arranged outside the family system. Nevertheless, more research is needed to guide family therapists in deciding which families can be treated and which family treatment techniques are the most effective.

End-of-Life Decision Making As advances in medicine extend the survival of more individuals with chronic illness, more older adults and their families are confronted with decisions about the use of life-sustaining measures. Despite the emphasis on individual autonomy and self-determination that is featured in many formal documents and statutes on the subject, recent research indicates that older adults want family members involved in such decisions. Moreover, many elders with terminal and chronically disabling illnesses cannot make decisions for themselves because of cognitive disorders or loss of consciousness. Investigations asking older adults about who they would choose as a surrogate decision-maker indicate that as many as 90 percent of older adults choose family members over friends, doctors, or lawyers. In descending order, they prefer a spouse, adult offspring, siblings, or other family members such as a cousin or niece to make end-of-life decisions for them. Despite the preference of most physicians to have one family member to deal with in such a circumstance, older adults typically prefer that medical decisions be made collaboratively, with input from multiple family members, especially when adult offspring are involved.

When functioning as surrogate decision-makers, adult offspring want to respect their parents' values and wishes. Yet research indicates little agreement between adult offspring and their parents on the importance of factors such as financial strain or burden of care in making a decision about life-sustaining treatment for the parent. Moreover, it is questionable whether adult children actually know their parents' specific preferences in these matters. There is a great need for family therapists and other health care professionals to help facilitate open communication about these matters while elders can still express their wishes.

SUGGESTED CROSS-REFERENCES

Section 30.5 features discussion of family therapy with the general adult population. Normal aging is covered in Section 51.2c. Late-life neurological and psychiatric disorders, such as Alzheimer's disease and mood disorders, are described in Section 51.3d and Section 51.3e. Ethnic and minority issues are covered in Section 51.6e. Ethical issues of late-life are discussed further in Section 51.6c. Section 51.6g includes more information on elder abuse.

SECTION REFERENCES

Basler SC, King DA: He's sick, but I'm the one who hurts: Our work with a medically ill older couple. In *The Shared Experience of Illness: Stories of Patients, Families, and Their Therapists*, SH McDaniel, J Hepworth, WJ Doherty, editors. Basic Books, New York 1997.

*Blieszner R, Bedford VH, editors: *Handbook of Aging and the Family*. Greenwood Press, Westport, CT, 1995.

Boss P, Caron W, Horbal J, Mortimer J: Predictors of depression in caregivers of dementia patients: Boundary ambiguity and mastery. Fam Process *29:*245, 1990.

Boszormenyi-Nagy I: *Foundations of Contextual Therapy*. Brunner/Mazel, New York, 1987.

Butler R: The life review: An interpretation of reminiscence in the aged. Psychiatry *26:*65, 1963.

Campbell TL, Patterson JM: The effectiveness of family interventions in the treatment of physical illness. J Marital Fam Ther *21:*545, 1995.

*Carter B, McGoldrick M: *The Changing Family Life Cycle: A Framework for Family Therapy*, ed 2. Gardner Press, New York, 1988.

Hargrave TD, Anderson WT: *Finishing Well: Aging and Reparation in the Intergenerational Family*. Brunner/Mazel, New York, 1992.

*Hargrave TD, Hanna SM: *The Aging Family: New Visions in Theory, Practice, and Reality*. Brunner/Mazel, New York, 1997.

Herr JJ, Weakland JH: *Counseling Elders and Their Families: Practical Techniques for Applied Gerontology*. Springer, New York, 1979.

High DM: All in the family: Extended autonomy and expectations in surrogate health care decision-making. Gerontologist *28:*46, 1988.

Hinrichsen GA, Niederehe G: Dementia management strategies and adjustment of family members of older patients. Gerontologist *34:*95, 1994.

Hinrichsen GA, Pollack S: Expressed emotion and the course of late-life depression. J Abnorm Psychol *106:*336, 1997.

Hughston GA, Christopherson VA, Bonjean MJ: *Aging and Family Therapy: Practitioner Perspectives on Golden Pond*. Haworth Press, New York, 1988.

Karel MJ, Gatz M: Factors influencing life-sustaining treatment decisions in a community sample of families. Psychol Aging *11:*226, 1996.

King DA, Bonacci DD, Wynne LC: Families of cognitively impaired elders: Helping adult children confront the filial crisis. Clin Gerontol *10:*3, 1990.

Knight BG, McCallum TJ: Psychotherapy with older adult families: The contextual, cohort-based maturity/specific challenge model. In *Clinical Geropsychology*, IH Nordus, GR VandenBos, editors. American Psychological Association, Washington, DC, 1998.

Lachs MS, Williams C, O'Brien S, Hurst L, Horwitz R: Risk factors for reported elder abuse and neglect: A nine-year observational cohort study. Gerontologist *37:*469, 1997.

Mittelman MS, Ferris SH, Shulman E, Steinberg G, Ambinder A, Mackell JA, Cohen J: A comprehensive support program: Effect on depression in spouse-caregivers of AD patients. Gerontologist *35:*792, 1995.

Neidhardt ER, Allen JA: *Family Therapy with the Elderly*. Sage, Newbury, CA, 1993.

Qualls SH: Aging families: The personal, permanent group. Group *21:*175, 1997.

Reis M, Nahmiash D: When seniors are abused: An intervention model. Gerontologist *35:*666, 1995.

*Richman J: *Preventing Elderly Suicide*. Springer, New York, 1993.

Shields CG: Family interaction and caregivers of Alzheimer's disease patients: Correlates of depression. Fam Process *31:*19, 1992.

*Shields CG, King DA, Wynne LC: Interventions with later-life families. In *Integrating Family Therapy: Handbook of Family Psychology and Systems Theory*, RH Mikesell, DD Lusterman, SH McDaniel, editors. American Psychological Association, Washington, DC, 1995.

Shields CG, Wynne LC: The strength vulnerability model of mental health and illness in the elderly. In *The Aging Family: New Visions in Theory, Practice and Reality*, TD Hargrave, SM Hanna, editors. Brunner/Mazel, New York, 1997.

Sukosky D: Life review in family psychotherapy. J Fam Psychother *5:*21, 1994.

Teri L, Logsdon R, Wagner A, Uomoto J: The caregiving role in behavioral treatment of depression in dementia patients. In *Stress Effects on Family Caregivers of Alzheimer's Patients*, E Light, G Niederehe, BD Lebowitz, editors. Springer, New York, 1994.

Wynne LC: The epigenesis of relational systems: A model for understanding family development. Fam Process *23:*297, 1984.

Zweig RA, Hinrichsen GA: Factors associated with suicide attempts by depressed older adults: A prospective study. Am J Psychiatry *150:*1687, 1993.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.4 TREATMENT OF PSYCHIATRIC DISORDERS

51.4L GROUP THERAPY

MOLYN LESZCZ, M.D., F.R.C.P.(C)

Group treatments play an important role in the comprehensive care of geriatric patients presenting psychological distress. Group therapy entails a broad range of specific models and techniques applied to a heterogeneous population of elderly adults in a variety of settings, and there is a risk of misapplication of models if elderly adults are viewed as a homogeneous population. A unifying principle is that group therapy should provide opportunities for restoration, amelioration, or improved adaptation and coping, related to the participants' age-induced losses of interpersonal, emotional, and cognitive capacities. In all instances, treatment is guided by the therapist's belief in the continued scope for change in later life.

HISTORY

An early tradition of group therapy began with the work of social workers in a variety of settings for seniors including homes for the aged, community centers, and mental institutions in the earlier part of the twentieth century, providing group interventions combining activity and social treatments. Maurice Linden was one of the first psychiatrists to publish in this area. He challenged traditional pessimistic myths about the untreatability of the elderly population. The opportunity he provided psychotic and “senile” institutionalized patients for social engagement and relatedness diminished exaggerated disability, dependence, and withdrawal. The intervention was well tolerated by patients, satisfying to staff, and economical.

Over the last half century, in response to the growing elderly population, the practice of group therapy has evolved to include acute care settings, day hospitals, and ambulatory and community settings. For some elderly patients today, group therapy is a later-life installment of an earlier psychotherapy. Group therapy is regularly integrated with psychopharmacotherapy as part of comprehensive treatment. Within the field of psychotherapy itself, the current approaches synthesize the strengths of cognitive-behavior, dynamic, and developmental models.

Current models include (1) verbal groups for the cognitively intact, including groups for burdened caregivers; (2) verbal groups for the cognitively impaired, including resocialization, remotivation, and reality-orientation groups; (3) activity- and creativity-centered groups, including dance movement, drama, music, art, and poetry groups; and (4) a variety of time-limited homogeneous groups addressing issues such as retirement and bereavement.

THEORETICAL ISSUES

The themes encountered in geriatric group therapy include reactions to loss or decline and the resultant discontinuity with a prior, more valued sense of self. Depression is a frequent disturbance, often comorbid with anxiety, affecting both cognitively intact and cognitively impaired elderly adults.

Loneliness and the lack of social integration are profound risk factors in depression. Higher levels of social support predict a significantly superior clinical course in the treatment of depression. Group therapy provides opportunities to grieve losses and reengage with interests and relationships. It provides real relatedness along with the acquisition of skills needed to generate social integration beyond treatment. Enhanced social support produces a main effect of improved well-being and a secondary, stress-buffering effect, which are particularly important for elderly persons in light of the likelihood of further losses. Combined pharmacotherapy and psychotherapy in the immediate, maintenance, and continuation phases of treatment is an area of important clinical study. It appears that combined treatment produces particularly robust results in treatment of acute problems. Combining drug and psychosocial therapies requires attention to the meaning each treatment has to the individual and cognizance of the risks of devaluing one or the other. Objectives include sustaining durable treatment responses and reducing subsequent affective morbidity, chronic psychosocial invalidism, and poor adherence to nonpsychiatric medical treatment. Treatment compliance can be enhanced through social support, education, and remoralization.

Developmental, psychodynamic, and cognitive-behavior models demonstrate substantial overlap and may be integrated both theoretically and technically. Dynamic or developmental approaches and cognitive-behavior approaches combine into a seamless, integrated whole with the recognition that effective psychotherapy is facilitated by appreciation of the individual's covert, subjective experience of more manifest interpersonal behavior, bridging depth of understanding with skill acquisition. The objectives of treatment include restoration of a sense of the individual's self, which integrates the present with the past; empathic confrontation of cognitive distortions; identification of maladaptive, alienating interpersonal behaviors; acquisition of interpersonal, social, and behavioral skills that the individual can use to enter and maintain a social network; and extending the adaptive cycle outside the in-treatment environment. Within this framework one can work within the same group or even within the same session on a number of interrelated issues.

The integrative model emphasizes the individuals' authorship of their own relational worlds. It combines working in the here-and-now of the group, using interpersonal feedback within the group, with processing of extra-group events and homework, which adds an intensity and relevance that enlivens cognitive and behavioral approaches. Similarly, incorporating the patient's reminiscing about an earlier, more robust sense of self creates opportunities for psychological restoration by recollection and also sets the stage for contemporary use of these valued elements of the self.

Activity- and creativity-centered groups foster social engagement through shared participation in a pleasurable activity and rekindle a sense of self through use of the creative process. Group activities elicit feelings of pleasure and mastery, challenge self-denigrating and constrictive distortions that individuals may have about themselves, and counter the trend toward homogenization within institutional life.

Group therapy may be more acceptable and provoke less anxiety than individual therapy because patients can share concerns and be of mutual help. Regression may also be diminished by participation in the group. Groups generally provide opportunities for behavioral change and may be of particular use for the older person who is less familiar with more abstract psychological issues.

Groups for burdened caregivers have become quite important in response to the increasing number of frail or impaired individuals being cared for by their families in the community. These caregivers are often highly burdened, isolated, overwhelmed, and not infrequently depressed, yet they are the key to implementing any interventions for the identified patient. The most effective groups provide a combination of emotional support and practical information. Groups may give caregivers an opportunity to work through feelings of loss, abandonment, and anger at the burdens imposed on them. Attention within the actual groups to the process of care giving and receiving can help caregivers identify a range of strengths and problems.

Resocialization and remotivation groups for cognitively impaired elderly persons play a significant role, largely in institutional settings. Interpersonal engagement may be achieved through the structure and content of these groups, with an emphasis on maximal sensory stimulation, rather than through active attention to the group process. Further development of nonpharmacological interventions in the treatment of demented individuals is needed to provide maximum rehabilitation opportunities for the retention of cognitive and social abilities.

TECHNIQUES

Technical expertise in the application of group therapies is surpassed in importance only by the therapist's posture of respect, hopefulness, genuine warmth, and empathy. Successful treatment hinges on establishing group cohesion, which may be problematic with demoralized, withdrawn, and self- or other-devaluing patients. Therapist activity must focus on catalyzing group interaction in ways that do not diminish or infantilize the participants. Cotherapy may provide an appropriate level of focus and activation, continuity during therapist absences, and the recognition and working through of countertransferential reactions. Within institutional settings, cotherapy also facilitates transmission of information between the group and other treaters to ensure maximum treatment integrity.

Attention must be paid to maintaining group norms and group integrity, which may also be challenged by patient or institutional resistance. Meetings must be scheduled at an inviolate time in an inviolate location. Active outreach may be required in ambulatory settings to promote regular attendance. Each member should find the therapy self-esteem enhancing and sustaining and a nonfailure experience. The therapist should achieve a balance between support and expression and may need to intervene with active prompting, reinforcement, reframing, and modulation of emotional intensity, emphasizing the invitation for what is desired rather than the rebuke for what is not. The empathic confrontation of projective defenses and resistances to engagement is an important therapeutic task. The therapist may need to be quite active in reinforcing progroup or procohesion exchanges, bolstering patient feelings of worth and belonging. Depressive affect must not overwhelm the group, as depressed elderly persons are particularly prone to the affect of demoralization. Patients must understand that demoralization is a specific reaction and not an obligatory state of aging.

Groups for elderly adults may often be held at midday in ambulatory settings and may be contiguous with lunch, to encourage opportunities for social interaction around mealtime. In institutional settings, group members often are involved with one another outside the group. Opportunities for engagement and interaction outside the group can be used as part of the prescription of behavioral interventions and to facilitate homework. In all instances, a group norm should be established that protects the confidentiality of the group therapy and encompasses extra group contact within the treatment framework.

Therapy groups generally consist of 6 to 10 patients, but activity groups may be much larger and require more than one or two leaders. Session length depends on members' capacities and may vary from 30 minutes daily for cognitively impaired patients to 75 and 90 minutes weekly for more cognitively intact patients. Many group formats are ongoing, but most cognitive-behavioral groups and groups for burdened caregivers are closed and time-limited, lasting from 6 to 20 sessions. The limitations of briefer treatments are being recognized, and treatments are often extended to include maintenance and booster sessions that meet less frequently but over a longer period of time, to extend treatment into maintenance and continuation phases. Special provision may be needed to accommodate hearing impairment through the use of FM amplifying systems.

Self-psychology makes a valuable contribution in psychodynamic group approaches by virtue of the understanding it provides about the inner world of the individual and the consequences of loss of selfobject and self-sustaining relationships, capacities, and functions. Empathic awareness of the impact of loss and decline as experienced subjectively may diminish the risk of collusion with demoralization or engagement in negative interpersonal cycles of criticism and rejection. The group may be used constructively in a self-sustaining mode for mirroring, idealized, adversarial, or twinning selfobject transferences, creating an opportunity for each member of the group to feel valued among a group of peers. Authentic engagement and empathic attunement with peers may diminish patient concerns about not being understood by much younger therapists.

Life review and reminiscing techniques use the natural process of patients' recollections and review of their lives. An empathic appreciation of the subjective meaning of the past to each member of the group is important. Reminiscing can be used as a main treatment focus or as an introductory exercise to stimulate cohesion in beginning groups. Reminiscing can be spontaneous or prescribed in a series of sessions, following a chronological life course sequence. One can use either milestoning (which emphasizes only positive events) or a comprehensive, autobiographical review process that provides an opportunity for mourning, conflict resolution, and acceptance. Encouraging a focus on the past is not without risk. Facilitating the group members' return to dialogue within the here-and-now of the group and eliciting feedback and interaction ensures that morbid self-absorption, preoccupation, and guilt do not exacerbate the individual's disengagement from the group or adherence to the past as a way of avoiding engagement with the present.

Aging often involves a shift from primary control, in which one effects change in the environment to deal with obstructions to one's needs and potentials, to secondary control, in which one's beliefs and attitudes must be realigned to address these obstructions (e.g., decreased mobility and access to social engagement and opportunities). Failing to do so may generate profound difficulties with self-criticism and beliefs about one's incompetence. Cognitive-behavior therapy emphasizes conscious cognition and learning and adapts behavioral strategies and the cognitive therapy model to the group setting, affording opportunities to acquire cognitive and behavioral skills. Dysfunctional cognitions and beliefs that cause depression are identified, confronted, and linked to prescribed changes in behavioral patterns.

This model emphasizes teaching skills, with a persistent focus on patients assuming responsibility and crediting themselves for improvements, stimulating a sense of mastery. Larger problems are deconstructed into smaller, more manageable components. Therapeutic strategies include challenging the patient's processes of generalizing, magnifying, and dichotomous thinking. Patients are helped to generate alternative solutions and responses; reattribute meaning to events; schedule activities with the assertion of pleasurable activities and affects; use time meaningfully; and challenge both the discarding of positive options and the failure to take credit for partial positive gains. Patients may be taught to log the relationship between mood and either pleasurable or aversive experiences and to increase the frequency of pleasure-focused activities. Behavioral prescriptions may be quite explicit and include modeling, role play, didactic sessions, psychoeducation, and review of homework, with feedback provided about each group member's efforts in using prescribed skills to reengage with pleasurable experiences.

A group therapy session for depressed elderly persons in a psychiatric day hospital began with the orientation of a new member to the group, a 79-year-old widower. Mr. J. was responsive to the group members, who discussed the value that they found in open expression to one another and sharing mutual support. They cautioned him that it might take time to feel comfortable within the group and not to be discouraged because like them, he would be likely to feel progressively more comfortable within the group. In fact Mr. J. seized upon this welcome to begin to articulate how he understood his recent onset of depression. Although a widower for some time, he told the group that earlier he had been able to overcome the loss of his wife because she had been ill for some time and they had a long and essentially happy life together. Six months ago, however, his 18-year-old granddaughter was killed by a car when she was jogging. This loss was incomprehensible to him, and the feelings of grief and depression overwhelmed him and made it impossible to find comfort or solace in any activity or relationship. He had disengaged from his life and felt such sadness with the parents of his granddaughter that he withdrew from them as well, refusing invitations to visit. Members of the group immediately expressed support and sadness, recollecting prior losses they had had.

Mrs. G., a 77-year-old woman who had previously participated in the group marginally, was visibly moved in response to Mr. J. This was noted, and with support she articulated her own experience for the first time to the group. Encouraged by Mr. J.'s openness, Mrs. G. started by saying that she was going to talk about some things that she had not talked about previously with anyone in the group. She said that although she too was widowed, a source of recurrent depression was her grief over the loss of her daughter, who had died several years ago at the age of 30. She described how her daughter had been diagnosed with a chronic renal disease at a young age and, in fact, had not been expected to live into her 20s. Her daughter defied the original prognosis, and despite her grim prognosis and over her mother's objection, she married at the age of 22, hoping that she would again defy the grim prognosis she received as a teenager. Mrs. G. told the group that her daughter died a few years later, and they were obviously saddened by this further disclosure. The group leader noted the obvious pain of these losses but seized upon an additional point that Mrs. G. mentioned, namely, that she had felt estranged from her daughter because of her daughter's seemingly “foolish” action of getting married. The greater tragedy she noted was that she had not been able to be as close to her daughter as she had hoped, prior to her daughter's death. She felt pain and guilt about that isolation and estrangement to this day.

With exploration and questioning from group members, she was pushed to try to rethink her daughter's actions as not necessarily foolish and inappropriate but rather as a reflection of somebody who was committed to living life as fully as possible for as long as possible. Living life in the presence of death was underscored as a major challenge that had to be faced. Other members of the group described how they have reacted to the fragility of life by avoidance and withdrawal. The group therapist suggested that this could be an important time to respond to this fragility by moving in the opposite direction—engaging life and the people that mattered to them. It was pointed out that this was happening in this group within the here-and-now, and when asked, Mrs. G. noted her relief in disclosing this guarded secret to the group members. People had chosen to support each other and encourage openness, facilitating necessary psychological work. In so doing, they duly honored the memory of their loved ones. The sadness began to lift in the group as group members were asked to articulate clearly and precisely the areas they could begin to work on, both within the group and outside of the group, and to seize such opportunities and to focus on this in subsequent meetings.

CLINICAL ISSUES

Indications Group therapy may be used as an adjunct to pharmacotherapy or as an independent treatment. Selection criteria should emphasize the homogeneity of

members' ego functions and cognitive capacities, not age alone. Groups homogeneous for these factors develop cohesion more rapidly than groups that must struggle with a member who significantly deviates from the group norm and invites disidentification (e.g., inclusion of a cognitively impaired patient with cognitively intact patients). Although pregroup preparation has not been extensively researched with elderly persons, the evidence supporting its usefulness in other patient populations is quite substantial, and it is routinely incorporated into cognitive and behavioral group treatments. Pretherapy preparation outlining the goals and methods of the group treatment demystifies an otherwise anxiety-provoking situation, helps to establish a therapeutic alliance, and sets group norms regarding regular attendance, confidentiality, and extragroup socialization. It may increase tenure and task adherence and foster hopefulness. Growing successful clinical experience contradicts earlier assumptions regarding elderly adults' capacity to engage effectively in psychotherapy. Many elderly individuals have been able to maintain effective adaptation, until those coping abilities are overwhelmed by loss, transition in roles, or illness. Comprehensive assessment of each individual's emotional, interpersonal, and cognitive status establishes a focus for treatment. Identification of egosyntonic character difficulties, in particular narcissistic pathology or interpersonally alienating behaviors, may be useful in delimiting behaviors that can result in early rejection from the group. Drop-out rates in group therapy in general range from 10 to 50 percent in younger adults and is likely the same in the elderly population. The likelihood of premature termination from treatment increases with increasing severity of depression, comorbid physical illness, and comorbid characteristic difficulties manifested in persistent *devaluation and externalization*.

Contraindications Contraindications are relative and relate more to the specific group than to the general premise of a group intervention. In general, individuals who require acute and active management because they are in crisis, suicidal, or acutely psychotic do not do well in groups. Severe cognitive or sensory impairment may also require exclusion from group therapy.

Goals Notwithstanding the range of group formats described, consensus exists regarding realistic goals of treatment. Although obviously shaped by particular patient needs, in general, the goals of treatment include reduction of symptomatic disturbance; reduction of isolation; restoration of interpersonal skills and engagement within a social network; correction of distortions and cognitive beliefs; restoration of self-esteem; improved cognitive function; acquisition of problem-solving and coping skills; grieving and adapting to loss; and realistic acceptance of one's needs and limitations with effective use of available resources. In institutional settings, group therapy may provide a forum to deal with the imposed changes in life situations and reduced feelings of autonomy, thereby enhancing quality of life by addressing issues of fit or dissonance between the individual and the imposed environment.

Objectives of groups for burdened caregivers include education about the illness process and specific ways of providing care; acquisition of both strategies for dealing with disruptive behavior and problem-solving and stress-reduction skills. Working through loss and grief and facilitating appropriate disengagement that may lead to more realistic decisions about placement are also important objectives. Additional areas of focus include legitimizing the needs of the caregiver and promoting self-care, working through anger and guilt, and addressing relationships with health care professionals. These objectives may be integrated or addressed in a series of modules that emphasize different important aspects sequentially. Quality-of-life concerns for both caregiver and patient are important foci. Groups for the cognitively impaired aim at reducing excess disability and fostering more appropriate social behavior and social functioning, thereby improving quality of life.

ETHICAL ISSUES

Proper patient selection and preparation and competent therapist functioning contribute to a treatment that is generally well tolerated and experienced by participants as humane and supportive. Interventions should be attuned to ethnic and cultural diversity and respect the unique beliefs, customs, and attitudes of identified populations. The therapy group may become very important and meaningful to its members, who may resist ending treatment because of the important social support and self-restorative experience the group provides. This necessitates active focus on transferring skills gained in treatment into the external world, recognizing that treatment is a bridge and not a destination. Therapeutic groups must be well integrated collaboratively into the overall milieu in which the individual resides. Alterations required in the framework of group therapy within institutional settings, along with other factors that contribute to a higher likelihood of extra group contact occurring and even being endorsed, warrants special attention to issues of confidentiality. Extragroup contact and the risk of subsequent subgrouping may be managed by diminishing secretiveness around the extragroup relationships and actively bringing them into the purview of the group. Group therapy is compatible with clinically indicated pharmacotherapy, and its concurrent use is complementary, not antagonistic.

RESEARCH AND EVALUATION

Substantial evidence supports the effectiveness of a range of group therapies with elderly persons. Qualitative reviews note that rationally used and conceptually sound group approaches are as effective as, and significantly superior to, no treatment of geriatric depression. This is clearest for mild or moderate depression, but even for severe depression, group psychotherapy may play an important role allied with pharmacotherapy in the acute, maintenance, and continuation phases of treatment. Structured groups are more effective than unstructured groups. Groups that lack focus or appreciation of the potential to amplify depression by encouraging intense affects beyond the members' capacity to tolerate have the potential for poorer treatment outcomes. Active cognitive and behavioral skill acquisition that addresses interpersonal difficulties and cognitive distortions is superior to reminiscence therapy alone. The focus that cognitive behavior therapy maintains on skill acquisition is particularly effective and leads to significantly improved and durable outcomes at 1-year follow-up.

Meta-analysis of psychosocial interventions in geriatric depression notes that psychosocial interventions (with group therapy particularly prominent) demonstrate a significant effect relative to placebo or no treatment. Positive outcome correlates with the degree of interpersonal engagement and the sense of interpersonal competence and effectiveness experienced by members of the group. Severe comorbid physical illness correlates negatively with successful outcome, and the durability of improvement is reduced by serious physical illness.

Benefits of treatment include reduced anxiety and depression and increased feelings of self-validation, self-acceptance, and self-esteem, with more-realistic awareness of one's strengths and limits. Continuing posttreatment benefits are fostered by improvements in interpersonal and social functioning. Successful outcome in geriatric group therapy is linked to sufficient tenure in treatment, therefore therapeutic strategies that facilitate treatment maintenance are important. Group therapy in the treatment of bereavement generates interesting findings; participants are generally consistent in reporting high satisfaction and positive valuation of the group experience, both for peer- and leader-led groups. However, at 1-year follow-up there appear to be few significant treatment benefits afforded by the intervention. Important questions remain regarding the effective duration and intensity of psychotherapeutic intervention and which patients require what form or intensity of treatment.

Groups for burdened caregivers generally demonstrate satisfaction with the group experience, even though complexities in the measurement of subjective and objective burden, coupled with the progressive deterioration of the care receiver, may confound other outcome measures. The burden of caregiving appears to be substantially greater with cognitively impaired persons than with physically impaired individuals. Treatment benefits reported include significant improvement in feelings of mastery, self-directedness, and problem-solving ability; increased capacity to consider appropriate alternatives to care; delay in institutionalization; improved relationship with the care receiver; improved capacity to grieve; and an enhanced capacity to care for oneself. Both professional- and peer-led groups show significant effects, although the former tends to emphasize the provision of information and problem solving, whereas the latter emphasizes peer support and social networking. Meta-analysis of controlled interventions for burdened caregivers shows that group interventions in general have a small but positive effect on caregiver burden and dysphoria, alongside the participant's subjectively high valuation.

Evaluation of outcomes in the group therapy of cognitively impaired patients are hampered both by a lack of consistency in measures used to determine improvements in cognition, behavior, and orientation and by limited follow-up. Early outcome studies focused on populations of institutionalized, chronic, psychogeriatric patients; documented improved discharge rates, socialization, functioning; and participation in hospital activities. Verbal orientation and socially appropriate behavior may be improved by group interventions, but the gains are generally modest and short-lived without ongoing staff or family reinforcement. Improvements appear to result from behavioral reinforcement rather than being genuine changes in cognition. However, behavioral improvement may endure beyond the active treatment phase and can be further reactivated by the resumption of treatment. Aggressive use of reality orientation techniques in the absence of empathic connection to the patients may result in further depressive reactions and dysphoria. The least equivocal outcome in this area is that staff involved in the more active interventions appear to value highly the opportunity for this form of active and humane engagement with their patients. Indeed, staff's improved morale and involvement may have important therapeutic impact.

SUGGESTED CROSS-REFERENCES

Related topics include cognitive-behavioral therapy (Section 51.4j) and individual psychotherapy (Section 51.4i) with elderly adults. The full spectrum of psychotherapies is discussed in Chapter 30.

SECTION REFERENCES

American Psychiatric Association: Practice guidelines for the treatment of patients with Alzheimer's disease and other dementias of late life. Am J Psychiatry *154*(Suppl):1, 1997.

*Arean PA, Perri MG, Nezu AM, Scheir RL, Christophe F, Joseph TX: Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. J Consult Clin Psychol *61:*1003, 1993.

Baines S, Saxby P, Ehlert K: Reality orientation and reminiscence therapy. Br J Psychiatry *151:*222, 1987.

Baker MN, Baker HS: *Self Psychological Contributions to the Theory and Practise of Group Psychotherapy*, A Alonse, MI Swiller, editors. American Psychiatric Press, Washington, DC, 1993.

Beck J: *Cognitive Therapy: Basics and Beyond*. Guilford, New York, 1995.

Beitman BD: Medications during psychotherapy: Case studies of the reciprocal relationship between psychotherapy processes and medication use. In *Integrating Pharmacotherapy and Psychotherapy*, BD Beitman, GL Klerman, editors. American Psychiatry Press, Washington, DC, 1991.

Brok AJ: A modified cognitive-behavioral approach to group therapy with the elderly. Group *21:*115, 1997.

Burnside I, Haight B: Reminiscence and life review: Therapeutic interventions for older people. Nurse Pract *19:*55, 1994.

Gallagher-Thompson D, Leary MC, Ossinalde C, Romero JJ, Wald MJ, Fernandez-Gamarra E: Hispanic caregivers of older adults with dementia: Cultural issues in outreach and intervention. Group *21:*211, 1997.

*Gallagher-Thompson D, Thompson LW: Psychotherapy with older adults in theory and practice. In *Comprehensive Textbook of Psychotherapy Theory and Practice*, B Bongar, LE Boulter, editors. Oxford University Press, New York, 1995.

Gerber GJ, Prince PN, Snider MC: Group activity and cognitive improvement among patients with Alzheimer's disease. Hosp Community Psychiatry *42:*843, 1991.

Gurland B: Epidemiology of psychiatric disorders. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

Heckhausen J, Schultz R: A life-span theory of control. Psychol Rev *102:*284, 1995.

Knight BG, Lutzky SM, Macofsky-Urban F: A meta-analytic review of interventions for caregiven distress: Recommendations for future research. Gerontologist *33:*240, 1993.

Labrecque MS, Peak T, Toseland RW: Long-term effectiveness of a group program for caregivers of frail elderly veterans. Am J Orthopsychiatry *62:*575, 1992.

*Leszcz M: Group therapy. In *Comprehensive Review of Geriatric Psychiatry II*, ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

Leszcz M: Integrated group psychotherapy for the treatment of depression in the elderly. Group *21:*89, 1997.

Lieberman MA, Yalom ID: Brief group psychotherapy for the spousally bereaved: A controlled study. Int J Group Psychother *42:*117, 1992.

Linden M: Group psychotherapy with institutionalized senile women: Study in gerontologic human relations. Int J Group Psychother *3:*150, 1953.

MacLennan BW, Saul S, Bakur Weiner M: *Group Psychotherapies for the Elderly* (American Group Psychotherapy Association monog no. 5). International Universities Press, Madison, CT, 1988.

Miller DM, Wolfson L, Frank E, Cornes C, Silberman R, Ehrenpreis L, Zaltman J, Malloy J, Reynolds CF III: Using interpersonal psychotherapy (IPT) in a combined psychotherapy/medication research protocol with depressed elders. J Psychother Pract Res *7:*47, 1998.

Müller U, Barash-Kishon R: Psychodynamic-supportive group therapy model for elderly Holacaust survivors. Int J Group Psychother *48:*461, 1998.

NIH Consensus Development Panel on Depression in Late Life: Diagnosis and treatment of depression in late life. JAMA *268:*1018, 1992.

Reynolds CR III, Frank E, Percel JM, Mazumdar S, Kupfer DJ: Maintenance therapies for late-life recurrent major depression: Research and review circa 1995. Int Psychogeriatr *7*(Suppl):27, 1995.

Schmid AH, Rouslin M: Integrative outpatient group therapy for discharged elderly psychiatric inpatients. Gerontologist *32:*558, 1992.

*Scogin F, McElreath L: Efficacy of psychosocial treatments for geriatric depression: A quantitative review. J Consult Clin Psychol *62:*69, 1994.

Sherbourne CD, Hays RD, Wells KB: Personal and psychosocial risk factors for physical and mental health outcomes and course of depression among depressed patients. J Consul Clin Psychol *63:*345, 1995.

*Toseland RW: *Group Work with the Elderly and Family Caregivers*. Springer, New York, 1995.

Wachtel PL: From eclecticism to synthesis: Toward a more seamless psychotherapeutic integration. J Psychother Integrat *1:*43, 1991.

Williams-Barnard CL, Lindell AR: Therapeutic use of “prizing” and its effect on self-concept of elderly clients in nursing homes and group homes. Issues Ment Health Nurs *13:*1, 1992.

Yalom ID: *The Theory and Practice of Group Psychotherapy*, ed 4. Basic Books, New York, 1995.

Yost EB, Beutler KE, Corbichley MA: *Group Cognitive Therapy: A Treatment Approach for Depressed Older Adults*. Pergamon, Oxford, England, 1986.

51.5 HEALTH CARE DELIVERY SYSTEMS

51.5A FINANCIAL ISSUES

GARY L. GOTTLIEB, M.D., M.B.A.

The importance of financial issues in shaping health care delivery for older people continues to intensify. For more than a century, sources of funding and mechanisms of payment for service have influenced mental health policy for older adults. These economic factors are key determinants of access to care, organization of health care delivery services, and the behavior of providers. Over the last decade, both government and industry have sharpened their focus on health care expenditures. This has stimulated the development of new models of care and has fundamentally changed the relationship between patients and providers.

The historic importance of government policy in shaping the status of psychiatric care for older people cannot be overemphasized. During the nineteenth century, financial responsibility for the care of chronically mentally ill persons and those with dementia was transferred from local to state authorities, causing asylums and almshouses to be replaced by the massive state hospital systems of the first part of the twentieth century. Ultimately, deinstitutionalization to limited community resources resulted in the transfer of a substantial number of older persons with severe and persistent mental illnesses from state hospitals to community nursing homes with sparse psychiatric resources. Reimbursement for nursing home care under Medicaid effectively shifted some of the financial responsibility for this population from the states to the federal government. The federal government's sensitivity to these new expenditures and social concerns about inappropriate care led to the Nursing Home Reform Provisions of the latter part of the 1980s and a set of guidelines for the use of psychiatric technologies in that environment.

In the mid-1960s, the introduction of Medicare and Medicaid created a private market for older adult health care as private practitioners and institutions could be reimbursed fee-for-service at market prices by these insurance programs. The lucrative nature of service delivery, reinforced by the growth of commercial indemnity insurance, drove the development of for-profit hospital care, particularly in freestanding psychiatry. Simultaneously, severe limitation in reimbursement for ambulatory psychiatric services under Medicare prevented the growth of private practice interest in geriatric care while creating perverse incentives for the use of hospital resources.

The health care cost reform of the last generation has similarly reshaped psychiatric service delivery. Medicare has become more attractive than commercial payers to hospitals and to individual providers. Nonpsychiatrist mental health specialists have become included in Medicare and other payment programs, and a greater diversity of providers has made psychotherapeutic services more broadly available to elderly adults. However, downward pressure on the growth of Medicare expenditures and incentives for the growth of managed care limit access to state-of-the-art psychiatric care for older adults.

ECONOMIC AND HEALTH POLICY ISSUES

Rational planning for mental health care services for the coming century requires understanding the projected growth and heterogeneity of older America. Forty years ago, only 1 of every 11 Americans was older than age 65. However, in each of the last three decades, this population has grown substantially and now totals about 35 million people, approximately 12 percent of the American population. More striking, the aging of the post–World War II baby boom cohort will yield approximately 66 million people over the age of 65 by the year 2030, accounting for about a quarter of the overall population.

Differences in life expectancy between men and women and the distribution of ethnic minorities among the older adult population will have important economic consequences. Given that there are approximately 1.5 women for every man older than 65 and about 2.5 women for every man older than 85, the economic effects of income discrimination and traditional work roles experienced by women in young and middle age will be exaggerated as retired women depend on pensions and Social Security derived from preretirement income.

Similarly, ethnic minorities make up a growing proportion of the over 65 population. In 1990, only about 11 percent of persons over 65 were from nonwhite ethnic minorities. By the year 2025, 15 percent of this population will be from minority groups. While life expectancy at birth for whites exceeds that of African-Americans by about 8 percent, by the time people reach age 75, mortality rates for blacks are considerably lower than they are for whites. Of critical economic and social importance, much older black adults have considerably higher rates of poverty and illness than white adults in the same age group. The economic and social discrimination of chronic underprivilege associated with American minority status are exaggerated by the socioeconomic consequences of older age. In this population, accrued social and financial resources are limited, access to preventive and acute health care is poor, and the likelihood of exposure to tobacco, alcohol, and other toxins is higher. This causes a cycle of medical, economic, and social poverty.

Retirement has important effects on economic status and access to health care and insurance. Retirement reduces personal income between 30 and 50 percent. Most Americans still retire in their 60s; about 90 percent of men in their early 50s are in the labor force, while only about 40 percent of men between the ages of 62 and 64 continue to work. Only 10 percent of men and 4 percent women over age 70 are in the active work force. Retirement directly affects health insurance premiums, availability of insurance, and access to some components of the health care delivery system. Recent efforts to restructure both Medicare and Social Security have included proposals to delay eligibility for these programs and thereby effectively delay retirement from 65 to 67. The very tight labor market of the end of the twentieth century and the shift from heavy industry to a service economy may support delays in retirement, particularly for lower- and middle-income Americans.

Currently, most Americans over 65 depend on a combination of Social Security benefits, pension, and income from savings and investments after retirement. Reduced income after retirement threatens stability of social class for older persons and increases the likelihood that some will experience economic poverty for the first time during late life. About 1 in 8 persons over age 65 have incomes below the established poverty level. About 20 percent of Americans who live beyond age 85 have incomes at or below that level. In contrast, more than 10 percent of persons over 65 have incomes above $60,000 per year, and 13 percent have assets in excess of a quarter of a million dollars. This heterogeneity of distribution of wealth and income among older adults creates major gaps in health and social policy for some segments of the population, while generating controversy regarding entitlements for those with substantial means. These issues are magnified in geriatric mental health care and in long-term care, in which out-of-pocket copayments, deductibles, and direct purchase of services account for a substantial component of total costs.

MEDICARE SYSTEM

Since 1966, Medicare has been the most important insurance program for older Americans. Medicare was designed to provide medical care benefits for older persons participating in the Social Security program. In 1972, Medicare was expanded to cover younger persons with disabilities and older adults outside of Social Security who were willing to pay a premium. In 1973, the program was further expanded to provide medical coverage for people with end-stage renal disease. As of 1997, about 38 million people, or about 13 percent of the American population, were covered by Medicare. About 90 percent of this population is over age 65.

Growth in the older adult population, advances in medical technology, and other increases in the cost of health care contributed to extraordinary growth in health care expenditures over the last generation. Health care costs now account for approximately 14 percent of the gross domestic product, totaling more than $1 trillion per year. Older adults consume approximately 30 percent of this care, and just under 50 percent of that amount is paid for by Medicare.

Various efforts to contain Medicare costs have successfully slowed health care inflation as the twentieth century ends. Cost reduction has resulted from reductions in reimbursement of hospital and physician providers, increased requirements for out-of-pocket payments by individuals, and a shift from indemnity insurance to managed care. Each of these efforts affects access to appropriate psychiatric care for older adults, and each has affected the size and shape of geriatric mental

health specialty care.

Medicare was designed to provide coverage for treatment of acute conditions. Most preventive services, long-term care, ongoing use of pharmaceuticals, and dental services are excluded from traditional indemnity Medicare. Other services, including mental health care, are subject to payment limitations and substantial copayments. Medicare pays for about three-quarters of older adult hospital care, 2 to 3 percent of nursing home care, and just under 60 percent of physician services.

The traditional, unmanaged Medicare program has two components: Hospital Insurance, or Part A, covers inpatient care, limited home care, skilled nursing facility care, and hospice services. Part B, the Supplemental Medical Insurance program, covers services provided by physicians and other professional providers, durable medical equipment, ambulatory medical laboratory tests, some outpatient hospital care, and group therapy services of partial psychiatric hospitals and comprehensive outpatient rehabilitation facilities.

Hospital Insurance (Part A) Americans over age 65 who participate in the Social Security program are entitled to Part A Medicare coverage. The Part A hospital insurance trust fund is supported by Social Security payroll taxes paid by employers and employees. These payroll taxes totaled $115 billion in 1997 and made up 88 percent of the income of the trust fund. The remainder came from interest earned on the monies in the trust fund and other miscellaneous sources. About 22 percent of the 38 million people eligible for Part A received hospital services in 1997. Their utilization rate is approximately four times that expected in younger, commercially insured populations.

Coverage for inpatient care is limited to relatively brief inpatient stays, presumably for the stabilization of acute symptoms. Part A provides coverage for episodes or spells of illness. A *spell* is an episode of inpatient care that starts with hospital admission and ends after the first period of 60 consecutive days outside the hospital after hospital discharge. Persons who are discharged and readmitted numerous times during a given episode of illness who have not been out for 60 consecutive days are considered to be within the same spell. For treatment of any condition in a general hospital, there is no limit on the number of spells or days covered within an individual's lifetime; however, no more than 150 days are covered during each episode.

The first 60 days of each spell are fully paid for after a deductible equal to the average cost of 1 day of hospitalization ($768 in 1999) has been met. The next 30 inpatient days are paid at the same rate less a 25 percent copayment. The last 60 days of covered hospitalization requires a 50 percent copayment. Days 91 to 150 are designated “lifetime reserve days.” Medicare will pay for those days during any single spell of illness, but only once in a beneficiary's lifetime. For example, patients who require five hospitalizations totaling 110 days without a period of discharge of at least 60 days would have 40 days of remaining coverage in their lifetime for future episodes that lasted more than 90 days. Individuals may save these days for future hospitalizations and use other resources to pay for any part of the costs of days 91 to 150.

When Medicare was signed into law, large populations of older adult patients were in state psychiatric hospitals. Legislators feared that the enactment of Medicare would shift the burden of care for that population from state-supported resources to acute care facilities reimbursed with federal monies. Therefore, Medicare coverage for inpatient treatment in free-standing psychiatric hospitals is limited to a total of 190 days over the course of a person's lifetime. If an individual becomes eligible for Medicare during a psychiatric hospitalization, the Medicare intermediary may deem that less than the full 150 days allocated for the spell of illness will be covered. Given improvements in treatment and increased life expectancy, this limitation can easily create undue hardship and prevent continuity of care. For example, a person with a recurrent major affective disorder who suffers several depressive episodes requiring inpatient care after age 65 could easily exhaust the lifetime free-standing psychiatric hospital benefit.

Part A provides coverage for nonacute services if they are used briefly and for “remediable” conditions. To obtain benefits for skilled nursing facility (SNF) care, a beneficiary must have been hospitalized for at least three consecutive days for an acute condition and admission to the SNF must follow within 30 days of hospital discharge. Medicare may cover up to 100 days of SNF care for each episode of illness. Similarly, coverage for home health services is limited to care for acute and remediable conditions.

From 1966 until 1983, Medicare Part A paid for hospital care using a retrospective, cost-based reimbursement system. This system compensated providers fee-for-service for all of those costs deemed allowable under Medicare without an incentive for utilization management. In 1982, after years of skyrocketing growth in Medicare, the Reagan administration signed into law the Tax Equity and Fiscal Responsibility Act (TEFRA). TEFRA created incentives to reduce hospital payments by setting targets for operating costs, and it provided supplemental payments for compliance with specified objectives. Reimbursement was shifted from a retrospective, cost-based approach to a design that would reflect patient mix, geographical location, and training costs. Most importantly, TEFRA replaced cost-based reimbursement with a prospective payment system.

The prospective payment system established by TEFRA uses diagnosis-related groups (DRGs) to classify patients who require similar care into 23 major diagnostic categories, and assigns each discharge to 1 of 468 diagnostic groups. These groups are derived from principal and secondary diagnoses, procedures rendered, and, to a lesser degree, age, sex, complications, and discharge status. The DRG system was revolutionary in creating a case-based approach to reimbursement. Hospitals are paid a fixed amount for each case according to the DRG, regardless of the resources consumed during the hospitalization. If patients require extraordinary resources or extremely prolonged hospitalization, Medicare classifies the patient as an “outlier” and provides additional payment to the hospital at a rate considerably below actual costs. This set of conditions creates an incentive for hospitals to provide efficient care and create an infrastructure of ambulatory services that could shorten or substitute for hospitalization.

Fourteen of the DRGs apply to the treatment of psychiatric disorders. Several studies indicated that while DRGs are modestly effective in predicting utilization for medical and surgical diagnoses, their ability to predict utilization for psychiatric diagnoses is poor. Therefore, the Health Care Financing Administration (HCFA) exempted free-standing psychiatric hospitals and distinct psychiatric units in general hospitals from the DRG system.

Free-standing psychiatric hospitals and psychiatric units in general hospitals continue to be reimbursed retrospectively by Medicare. However since TEFRA was implemented, payment for treatment in psychiatric units is capped at a rate established for each facility on the basis of utilization during its first fiscal year of operation after October 1, 1983. Now, if the actual cost of care per case exceeds the capped rate, the hospital must absorb the loss. However, if the cost is less than the capped rate, a small incentive payment is retained by the hospital. A minor but extremely important provision of the Balanced Budget Amendment of 1997 (BBA) limited the maximum TEFRA rate for psychiatric facilities to 75 percent of the national median. Additionally, the BBA eroded incentive payments for hospitals that were more efficient than the target rate.

Supplemental Medical Insurance (Part B) Medicare Part B is a voluntary medical insurance program that requires a monthly premium payment ($43.80 in 1998) for participation. About 97 percent of Medicare Part A beneficiaries also participate in the Part B program. Part B is supported by a trust fund that includes general revenues from the Treasury, as well as premiums. Premiums cover about 25 percent of the total cost of the program. Part B benefits are subjected to an annual deductible ($100.00 in 1999) as well as a copayment. Nonpsychiatric physician services and psychiatric inpatient care require a 20 percent copayment. While the number of reimbursable ambulatory psychiatric visits is no longer subject to limitation, Medicare requires a 50 percent copayment for psychotherapeutic services.

From 1984 when TEFRA and prospective payment were initiated, until 1992, the rate of growth of Medicare physician costs was more than double the rate of growth for inpatient hospital costs. Policy analysts attributed this growth to uncontrolled prices and increases in the number and intensity of services consumed. In 1989, inflation in physician-related costs and mounting evidence that physicians' behaviors in treating complex medical illnesses were influenced by the high prices of the procedures they performed, led to a sweeping reform of Medicare Part B through the Omnibus Budget Reconciliation Act of 1989 (OBRA-89).

OBRA-89 created standards for expenditures and procedure volumes in an attempt to control the total costs of physicians' services. The law requires that the secretary of health and human services establish specific volume performance standard growth rates for surgical, primary care, and nonsurgical non–primary care services at the beginning of each fiscal year. These standards must be related to medical inflation and changes in the size of the Medicare population as well as changes in service volume and intensity. At the end of each year, the secretary must compare actual growth and expenditures with this volume performance standard and determine any changes in fee schedules for the following year.

OBRA-89 also created a national uniform Medicare fee schedule to replace the local methods traditionally used by intermediaries to set fees for each procedure and for each provider. In 1992, a Resource Based Relative Value Scale (RBRVS) was developed to determine relative values of about 7000 physician procedures and services. The relative value of each of these services is derived from three components: work, practice and malpractice expenses, and the opportunity costs of postgraduate specialty training required. The proportion of each component used in determining the relative value is based on a weighted average of specialty-specific practice expense and malpractice data. In an effort to change provider behavior and reduce incentives to perform highly reimbursed technical

procedures, work is derived from time intensity and stress associated with delivery of the procedure. About 60 percent of the value of each fee is adjusted for local costs. The product of each relative value and a monetary conversion factor determines the fee for each service.

OBRA-89 was a landmark for behavioral health in its provision for direct reimbursement of services provided by psychologists and social workers. It also set criteria for nonphysician providers regarding consultation with a physician. Nonphysician providers must document that a patient has been informed that consultation with the patient's primary care physician is desirable to evaluate medical conditions that could contribute to a patient's well-being. The provider must also document any written or verbal communications with the primary care physician about treatment unless the patient specifically prohibits contact. The law does not require consultation with a psychiatrist.

In general, direct patient contact is required for any reimbursable service. Consultation or other services rendered by telephone and renewals of prescriptions that do not include direct evaluation of clinical conditions are not reimbursable. Medicare allows billing for services required to obtain clinical information from relatives or close associates of unreliable or uncommunicative patients. Family counseling services are only reimbursable when their purpose is to facilitate patient care.

Providers may participate in Part B of the Medicare program in two ways. First, providers may accept assignment and agree to accept the Medicare fee as payment in full. Medicare's fiscal intermediary will pay the provider the approved amount less copayment and deductible directly. The provider is legally required to bill the patient directly for the copayment and the deductible. Second, a provider may choose not to accept assignment. Providers who do not accept assignment may charge patients up to 115 percent of the approved Medicare fee. Medicare reimbursement for providers who do not participate is limited to 95 percent of the approved fee less deductibles and copayment. The provider may then bill the patient directly for the balance of charges up to the 115 percent limit.

All providers must accept assignment from Medicare beneficiaries who also participate in the Medicaid program. Assignment must be accepted for all Medicare recipients who are at or below the federal poverty level, whether or not they have Medicaid as a secondary payer. In an increasing number of states, medical licensure authorities and insurance commissions require that all physicians who participate in the Medicare program accept assignment for patients seen within that state.

MEDICARE RISK IN MANAGED CARE

Since the early 1970s, Medicare has encouraged older Americans to enroll in capitated health plans. Since that time, health maintenance organizations (HMOs) have been able to contract directly with Medicare. Medicare pays the contractor a premium equal to 95 percent of the local adjusted average per capital cost (AAPCC). The AAPCC is based on the individual's county of residence, age, sex, and whether a patient is covered by Medicaid, lives at home or in an institution, works or is retired or disabled. HMOs, while allowed to earn “normal profit margins,” must use savings above a predetermined rate to provide extra services for elderly members. To penetrate a local market, HMOs frequently waive premiums and offer usually uncovered services to induce enrollment. Unlike traditional Medicare, HMOs allow access to primary care and other preventive services in the absence of a primary diagnosis. Additionally, managed care plans usually provide coverage for some portion of outpatient pharmaceutical costs.

The nature of HMO risk contracting with Medicare provides an incentive for HMOs to market to healthy older adult populations who are likely to consume considerably less than the 95 percent of AAPCC. Healthy populations yield substantially greater profit margins, leaving sicker, more-disabled persons to be cared for through direct payments from HCFA in the traditional Medicare program. Several studies have indicated that capitated plans generally enroll healthier and less costly populations than those who remain in fee-for-service Medicare. In one study, in the 6 months preceding enrollment, new members of capitated plans had health care costs that were 35 percent lower than the average costs of patients in traditional Medicare. Some plans profit through marketing to younger older adults (people ages 65 to 70) in settings with a greater likelihood of health and independence. Studies indicate that Medicare has overpaid capitated plans by 5 to 20 percent because of the better health of their members. To reduce financial incentives created by enrolling healthy persons and to pay more fairly for the care of frailer populations, the BBA required that Medicare link payments to HMOs to the health status of their beneficiaries by the beginning of the year 2000. More-frail populations will be eligible for higher payments than healthier cohorts. This should encourage plans to enroll more chronically ill populations and to bring a broader base of the general older adult population into managed care.

Medicare managed care penetration has increased steadily over the last decade. While only 3.6 percent of older Americans were enrolled in HMOs in 1990, this proportion increased to 10 percent in 1996 and over 15 percent in 1998. The Congressional Budget Office estimates that just under 20 percent of Medicare beneficiaries will be in risk contracts by the year 2000. Approximately 100,000 Medicare beneficiaries are joining managed care plans each month. The Medicare Plus Choices Program, a major component of the BBA, is designed to make it easier for various kinds of organizations, including provider-sponsored organizations and preferred provider organizations (PPOs), to participate in Medicare managed care. Medicare Plus Choices is intended to create greater competition among plans.

Managed care plans caring for Medicare recipients have applied the same paradigms to behavioral health care that they have used for their younger commercial populations; the vast majority carve out the behavioral health benefit and capitate the risk for these services to managed care organization vendors. Most managed care organizations are large, publicly traded for-profit entities that function as psychiatric and substance abuse insurance companies in numerous markets regionally and nationally. Managed care organizations generally establish local networks of ambulatory and inpatient providers; establish fee-for-service pricing with those providers, supply medical management, care management, utilization review, and credentialing functions; and pay claims. These organizations are generally very effective in reducing inpatient utilization, which enables them to negotiate significantly discounted prices for services and limit the size of available networks.

A review of the credentialing and procedure manuals for a group of the largest managed care organization nationally found virtually no mention of requirements for specialty training or expertise in the care of older adults or for special consideration of medical comorbidity or frailty. There were no requirements for services to be sited where older adults reside or where access to care may be facilitated for frail or disabled persons. This review suggests that the current model for delivery of managed behavioral health services to older adult Medicare beneficiaries is unlikely to reflect recent advances in behavioral neuroscience. Similarly, most carve-out programs exclude or limit comprehensive assessment, neuropsychological testing, and other routine components of dementia evaluation.

Recently, a number of not-for-profit entities have developed the capacity to be included in full-risk agreements with their medical and surgical colleagues. This has allowed emergence of a new model of behavioral health care, carving in services while locating providers on-site or adjacent to primary care delivery. These programs are serving considerably smaller populations than most managed care organizations, and differences in actual outcomes have yet to be demonstrated.

Managed care continues to be a priority in Medicare policy. The federal government plans to reduce Medicare expenditures by $115 billion by 2002, by broadening the types of coverage available to enrollees. However, reductions in payments to managed care entities by Medicare coupled with complex paperwork requirements have prompted more than 100 managed care plans to leave the Medicare program over the past year. The tension between the financial effects of the BBA and government efforts to attract beneficiaries to managed care plans is likely to persist over the coming years.

About 42 percent of older adults who remain in the traditional Medicare program purchase Medicare supplementary (Medigap) insurance. In 1992, the federal government created standards for private insurance to supplement the federal program. It provided requirements for 10 basic policies. All of these policies must include payment of the Medicare Part A hospital and Part B provider coinsurances. Additionally, all policies must pay the 50 percent copayment for outpatient psychiatric services. More elaborate policies may cover deductibles, SNF coinsurance, outpatient prescription pharmaceuticals, and preventive in-home services. Medigap regulations provide incentives for the use of PPOs by Medicare recipients. This allows insurers to offer lower premiums to members who agree to use a PPO for all of their care. An increasing number of states have authorized insurers to offer this option.

MEDICAID

Enacted in 1965 as a social insurance program to pay for the medical care of Americans with limited financial means, Medicaid is funded through a combination of federal and state matching funds. To receive these matching funds, states must provide a basic benefit package. The program allows states to impose restrictions on the services that they fund and the level of reimbursement. Medicaid reimbursements for psychiatric services for older adults vary substantially by state. However, Medicaid pays about 25 percent of all psychiatric care in the United States.

Between 10 and 15 percent of older Americans are eligible for both Medicare and Medicaid. While older adults represent about 16 percent of all Medicaid recipients, the care that they consume accounts for about 40 percent of all Medicaid expenditures. Most state Medicaid programs have provisions for purchasing Medicare Part B coverage. This allows states to reduce their risk for payment of professional services. In this way, Medicaid functions as a supplementary (Medigap) policy for many older adults. The program covers services after Medicare benefits have been exhausted, and it also covers a variety of services that may not be covered by Medicare,

including prescription drugs and long-term care.

Medicaid's importance to geriatric health care is greatest in its coverage of long-term care services. Medicaid pays for about 42 percent of SNF and intermediate care facility costs. Most states require that persons have limited income and assets to qualify for Medicaid.

The nursing home is an important site of care for older adults with severe and persistent mental illness, including depression associated with medical illness and disability, schizophrenia, and dementia. Deinstitutionalization from state facilities shifted the financial burden of care for many persons with these diagnoses from the states (in state hospitals) to cost sharing with the federal government in its partial support of Medicaid. The nursing home reform provisions of OBRA-87 require preadmission screening for nursing home applicants to determine if active psychiatric illness is present. If active treatment for an identified disorder cannot be provided appropriately in the nursing home, referral for this care is mandated. Nursing homes must demonstrate efforts to provide appropriate care for behaviorally disturbed patients, and appropriate training in management of difficult behaviors is required of all staff members. OBRA-87 requires monitoring of the use of psychotropic medications and restraints in long-term care. While these regulations have added substantial costs, there has been no incremental reimbursement from Medicaid to support them.

OUT-OF-POCKET EXPENDITURES

Health care costs paid directly by older adults from personal or family income or savings are classified as out-of-pocket expenditures. They include premiums for Medicare Part B, private Medigap policies or enrollment fees for Medicare managed care, as well as copayments and deductibles. Charges for uncovered services, including outpatient pharmaceuticals and long-term care, constitute the largest portion of out-of-pocket expenditures. While per capita out-of-pocket spending has declined because of the growth of managed care, the effects of out-of-pocket payments on older adults, particularly those with limited resources, are considerable. It is estimated that Medicare beneficiaries over age 65 who have incomes below the federal poverty level and who are also eligible for Medicaid spend 35 percent of their incomes on out-of-pocket health care costs. Medicare beneficiaries below the poverty level who do not have Medicaid spend half their incomes on out-of-pocket health care costs. Similarly, as the use of long-term care facilities grows, middle-income and lower-middle-income persons will become more seriously affected by growing payments for uncovered services. Notably, numerous levels of long-term care are not covered by Medicaid or Medicare. For example, the growth of assisted living has created increased demand for a level of service that can only be purchased out of pocket.

Prescription drugs are the fastest growing segment of health care costs. Recently, spending for prescription drugs has increased at two to three times the rate of health care inflation. As the twentieth century ends, policymakers are working furiously to slow this growth while debating expansion of Medicare coverage to include pharmaceuticals.

FUTURE DIRECTIONS

In the twenty-first century the growth of the older adult population is the major challenge faced by health care policymakers. Older adults consume health care services disproportionately, and our ability to extend and improve the quality of life depends on expensive technologies and interventions. The recent emphasis on cost reduction while technology and the size of the frail older population grow in parallel creates a clash among humanity, science, and economic stability. Advances in behavioral neuroscience and particularly in geriatric psychiatry can improve quality of life, extend life, and improve the economic productivity of the fastest growing subsegment of the population. Health policy that deals rationally with these variables must consider all of these economic perspectives without the distortion created by an overemphasis on reducing the direct costs of care.

SUGGESTED CROSS-REFERENCES

Section 51.3 covers psychiatric disorders of late life, and Section 51.4 covers psychiatric treatment of older patients. Managed care is discussed in Section 51.5b and Section 52.2, VA medical centers in Section 51.5c, and community services for elderly psychiatric patients in Section 51.5d. Section 51.6a covers psychiatric aspects of long-term care.

SECTION REFERENCES

Bureau of the Census: *Current population reports. Health insurance coverage, 1997 and 1998*. Government Printing Office, Washington, DC, 1998.

Chulis GS, Eppig FP, Hogan MD, Waldo DR, Arnett RH: Health insurance and the elderly. Health Aff *12:*111, 1993.

Coulam RF, Gaumer GL: Medicare's prospective payment system: A critical review. Health Care Financ Rev *13*(Suppl):45, 1991.

Gallagher TH, Alpers A, Lo B: Health care financing administration's new regulations for financial incentives in Medicaid and Medicare managed care: One step forward? Am J Med *105:*409, 1998.

*Ginzberg E: US health system reform in the early 21st century. JAMA *280:*1539, 1998.

Gottlieb GL: Financial issues. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadavoy, editor. American Psychiatric Press, Washington, DC, 1996.

*Gottlieb GL: *Managed Behavioral Healthcare Standards, Guidelines, and Competencies for Older Adults with Mental Illness*. Center for Mental Health Services, SAMSHA, Rockville, MD, 1998.

Grimaldi P: Medigap insurance policies standardized. Nurs Management *23:*20, 1992.

Hsiao WC, Braun P, Dunn D, Becker ER, Yntema D, Verrill DK, Stamenovic E, Chen S: An overview of the development of the resource-based relative value scale. Med Care *30*(Suppl):1, 1992.

Iezzoni LI, Ayanian JZ, Bates DW, Burstin HR: Paying more fairly for Medicare capitated care. N Engl J Med *339:*1933, 1998.

*Iglehart J: The American health care system—expenditures. N Engl J Med *340:*70, 1999.

*Kuttner R: The American health care system—health insurance coverage. N Engl J Med *340:*163, 1999.

Levit K, Cowan C, Braden B, Stiller J, Sensenig A, Lazenby H: National health expenditures in 1997: More slow growth. Health Aff *17:*99, 1998.

*Mendelson DN, Schwartz WB: The effects of aging and population growth on health care costs. Health Aff *12:*119, 1993.

Rovner J: Medicare privatisations have growing pains. Lancet *352:*1532, 1998.

Scherl DJ, English JT, Sharfstein SS, editors: *Prospective Payment in Psychiatric Care*. American Psychiatric Press, Washington, DC, 1988.

Schneider LS, Reynolds CF, Lebowitz B, Freidhoff A, editors: *Diagnosis and Treatment of Depression in Late Life: Results of the NIH Consensus Development Conference*. American Psychiatric Press, Washington, DC, 1994.

Weiner M, Powell NR, Weller WE, Shaffer TJ, Anderson GF: Alzheimer's disease under managed care: Implications from Medicare utilization and expenditure patterns. J Am Geriatr Soc *46:*762, 1998.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.5 HEALTH CARE DELIVERY SYSTEMS

51.5B MANAGED CARE

JEROME V. VACCARO, M.D., AND S. ALAN SAVITZ, M.D.

Managed care has become a core term in the health care lexicon in a relatively short period of time. Through the mid-1980s, the impact of prospective payment initiatives such as health maintenance organizations (HMOs), diagnosis-related groups (DRGs), or other managed care efforts continued to be eclipsed by the more prevalent fee-for-service system. Physicians and other health care providers became accustomed to receiving reimbursement for their services retroactively and encountering little or no clinical oversight of their work. The exceptions to this were the utilization review efforts in place through various organizations (such as hospitals) and by some payers (notably the Health Care Financing Administration [HCFA]). These reviews, however, were always retrospective and rarely affected an individual patient's care. They were designed instead to influence practice patterns in a collegial, consultative manner. In the late 1980s and early 1990s the health care delivery system began to change, and prospective payment, accountability, and managed care became major forces.

DEFINITIONS

Behavioral Health *Behavioral health* is used in this section to denote the combined areas of mental health and chemical dependency. In both the public and private sectors, this term is used increasingly to spur integration of these two areas, since patients often present with comorbid conditions that require unified approaches to clinical care and unified systems of care.

Managed Care *Managed care* is used to denote an effort to bring business controls or effective business principles to health care delivery. Managed care in its current forms has influenced the delivery of clinical care to be continuous, comprehensive, effective for the condition being treated, and cost efficient. With these goals stated, several issues are important. First, under fee-for-service systems, data were not collected in any meaningful way outside research settings, so it is difficult to ascertain the true effects of managed care efforts in a broad sense. However, data from more limited settings (e.g., several private employer–funded plans and a number of Medicaid plans) support increased overall access (in one study, the probability of using any mental health care increased significantly under a managed care program), increased comprehensiveness (particularly in outpatient settings), cost control, and some accountability for outcomes. As of 1998 managed behavioral health care plans covered about 130 million Americans. This has essentially brought control to the cost of behavioral health care while maintaining or improving access to care. The major change has been the sharp decrease in use of inpatient care and extended outpatient psychotherapy.

Medicare The Medicare program is a federally funded, federally run program that provides basic health insurance coverage to most of the elderly population and a portion of the disabled population (e.g., physically disabled or seriously and persistently mentally ill). It began in 1965, when Congress passed legislation introduced by President Lyndon Johnson in 1964. The Medicare program enjoys widespread support, and groups such as the American Association of Retired Persons (AARP) have become vocal and powerful advocates resisting reform measures such as mandatory enrollment in prepaid plans, increased retiree contributions, or decreased benefits.

The Medicare program is composed of two separate components, parts A and B. Part A pays for hospital care, home health care, hospice care, and some skilled nursing services. Part B pays for physician services, outpatient hospital and clinic care, laboratory and other diagnostic tests, and other services. This A and B distinction has become obsolete because of changes in our health care delivery systems (e.g., the emergence of unified or integrated clinical delivery systems), and it presents barriers to effective management of care in our new environment. In fact, the A and B separation can contribute to fragmented care by providing for separate inpatient and outpatient delivery systems. Medicare Part A provides coverage for retired individuals who receive Social Security benefits, those who have received government disability benefits for more than 2 years, and those with end-stage renal disease. Whether or not they are entitled to part A benefits, all persons age 65, and disabled or end-stage renal disease part A participants, may join part B by paying a monthly premium.

While many see the Medicare program as generous, it offers poor catastrophic coverage; no vision, dental, or outpatient medication benefits; and high copayments for many services. Copayments for services vary. For example, a typical copayment for a psychotherapy session may be as high as $20 to $30 under the traditional Medicare system, while such copayments in typical managed behavioral health care plans rarely exceed $10 to $15. Copayments and deductibles for hospital care can be several hundred dollars per stay, where such payments under a managed plan either do not exist or are quite low. In 1998, Medicare recipients paid an average of about $46 a month in premiums for their part B (nonhospital care) coverage. In all, Medicare covers about 45 percent of the annual health care bill for elderly adults. This has led to the emergence of a second system of coverage, including Medigap insurance, retiree benefits programs, and contributions from the Medicaid program.

Medigap Insurance and Retiree Benefits Many elderly individuals purchase additional insurance coverage that pays for some or all of their Medicare premiums, deductibles, and coinsurance and provides coverage for catastrophic illness. Others receive such coverage as a retirement benefit from previous employment. In all, about 65 percent of elderly adults have one or both of these benefits. In 1996, Medigap premiums averaged about $70 per month. Increasingly, employers are attempting to control their own contributions to retirement plans by using managed care.

Medicaid The Medicaid program provides additional coverage for about 16 percent of the elderly population who meet stringent federal and state eligibility criteria (these criteria mainly involve income limits and health care expenditure considerations). Medicaid pays the premiums, copayments, and deductibles for these “categorically” (e.g., an individual falls into a Medicaid-eligible category) eligible individuals. Because of the increased enrollment of Medicaid recipients in managed care plans, the care of dually eligible elderly individuals becomes ever more complex and fraught with administrative challenges (e.g., as states move Medicaid to managed care, and seniors remain in fee-for-service Medicare, care management and delivery become very difficult).

HISTORY

A series of events occurred that were intended to improve behavioral health care while being cost-effective. First, in the 1960s, new services and technologies (e.g., alternatives to inpatient care, psychiatric rehabilitation) were developed through the community mental health movement. Central features of this included avoiding inpatient care, the commitment that care should be community-based, and focusing on a population's needs (e.g., catchment areas) in service planning and delivery. Then in the 1970s, the availability of mental health benefits in indemnity plans increased. These benefits covered inpatient care and some outpatient treatment. Within HMOs, however, minimal benefits were offered, usually limited to emergency assessment and crisis intervention. At the same time, the number of psychiatric inpatient beds grew significantly, especially in free-standing psychiatric hospitals and rehabilitation facilities. Concurrently, training programs for behavioral health specialists increased, especially for master's level clinicians. State boards registered, certified, or licensed these providers to practice independently.

By the 1980s these factors combined to greatly increase the use of services, which caused a rapid escalation in behavioral health care costs in indemnity insurance (fee-for-service) plans. Large employers, particularly those in the Fortune 500, wanted to maintain behavioral health benefits but saw a lack of access in HMOs and rapid inflation in indemnity plans. For example, one such company saw mental health care costs increase between 12 and 33 percent per year during a 2-year period. By the mid- to late-1980s, this led to the establishment of “carve-out” managed care companies that used utilization management techniques to control costs. These “carve-out companies,” now collectively referred to as managed behavioral health care organizations (MBHOs), made the case to large and midsize employers that they could control costs (and in most cases reduce costs from fee-for-service levels) while enhancing or maintaining quality.

Provider networks were initially composed of practitioners and facilities willing to discount their fees. This situation has evolved so that efforts now place more emphasis on quality management and outcomes management in establishing and maintaining provider networks. For example, the American Managed Behavioral

Healthcare Association (AMBHA) publishes a set of outcome and process indicators, the Performance Measures for MBHOs (PERMS), which is now in its second edition (PERMS 2.0). The PERMS indicators include measures such as access to or availability of care, appropriateness of interventions (e.g., the use of family interventions with children), effectiveness of care, and satisfaction with the experience of care. These indicators supplement those published by the National Committee on Quality Assurance (NCQA), referred to as HEDIS (the Health Plan Employers Data and Information Set), which cover more general health care indicators. The general thrust of NCQA and AMBHA efforts is to have managed care plans demonstrate accountability for improving service and treatment outcomes over time, as NCQA accreditation depends on such improvement.

In 1996, the Medicare program had 38 million enrollees, increasing by about 2.5 percent per year since its inception. The costs associated with the Medicare program were $191 billion in 1996 and are projected to rise to over $314 billion per year by 2002. Almost since its inception, it has been predicted that the Medicare Trust Fund (part A) would be bankrupt within 2 to 17 years of the date of prediction (e.g., by 1972, the prediction was that the fund would be insolvent by 1976). Current estimates are that the trust fund will be bankrupt by the early twenty-first century. This latest estimate is more dire because of the rapid aging of the population, the growing imbalance between elderly and younger populations (e.g., the graying of the baby boomers), fears about returning to a large federal deficit, and antitaxation sentiment.

Cost increases have been commonplace in the Medicare program since its beginnings: 17.5 percent during the 1970s, 12.1 percent during the 1980s, and 10.3 percent during the 1990s. The forces that account for these rapid rises include increased numbers of enrollees (aging of the population, inclusion of special populations such as those with disabilities or end-stage renal disease), increases in covered benefits (e.g., home health, hospice), expenses associated with advanced technology, and contributions to the academic enterprise through higher rates of reimbursement.

During the 1990s, the growth in the number of Medicare enrollees in prepaid managed care plans such as HMOs and social HMOs accelerated, so that by 1996 11 percent of recipients (over 4 million people) were enrolled in such plans. This trend is expected to continue and perhaps accelerate through the first decade of the new century. For the most part, Medicare HMO members are satisfied with their experiences in HMOs. In a survey of members, 96 percent of members rated their overall health care by the plan as good to excellent. A comprehensive review of the Medicare HMO program concluded that Medicare HMOs use resources more efficiently (with minor differences in access, quality, and satisfaction) than fee-for-service members. However, this study and others argue that HMOs do not perform as well for chronically ill patients.

Another driving force behind increased enrollment is the fact that the level of covered services under most, if not all, managed Medicare plans greatly exceeds that of the traditional Medicare plan. For example, most managed plans offer some coverage for outpatient medications, and many offer vision and other benefits not covered under the traditional program. Given the more generous benefits, the cost savings, and the maintenance or improvement in the quality of care, one might expect the growth to be even more rapid. Factors limiting growth include the following:

1. *Perceptions that choice is limited in such plans*. This problem was especially true when staff model HMOs (e.g., choice of physicians is limited to physician-employees of the plan) predominated or when network model plans (e.g., choice is limited to physicians contracted with the plan) had small provider panels because of low physician participation. These concerns are less prominent, as both staff and network models have matured and aggressively seek to provide maximum choice (e.g., many plans now offer free access to a wide variety of specialists, and most plans have greatly expanded the number of contracted physicians on their panels). However, public opinion about this matter has been slow to change.
2. *Concern that not all care options will be made clear or available*, either because of improper physician incentives or contractual restrictions promulgated by HMOs. This criticism was levied at some plans in which physicians were given financial and other incentives to limit care (e.g., monitoring specialist referrals). Widely covered criticisms that plans were introducing gag clauses that prohibited physicians from discussing care options with their patients were not verified during a General Accounting Office (GAO) investigation. These concerns have been addressed by increased public scrutiny and consumer sophistication, to say nothing of increased government regulation.
3. *Inadequate geographical coverage of existing HMOs*. In the early phases of managed Medicare programs, especially with the social HMOs, geographical reach limited HMO expansion. By 1996, however, over 75 percent of Medicare enrollees lived in areas served by managed Medicare programs. In addition, the government has become more sophisticated about access, now specifying numbers of physicians and specialists that a plan must have in its network in a given geographical area. In addition, at the insistence of accrediting bodies such as NCQA, plans aggresively monitor indicators of adequate access and must report these.
4. *Concerns that managed plans would not be able to deal with the frail elderly or disabled populations*. The successes of the social HMOs and mainstream managed care plans in developing effective and efficient models of care for these populations, coupled with the increased skill and awareness of regulatory bodies and consumer groups, have allayed many of these concerns. New risk-adjustment methodologies may encourage plans to enroll these individuals. However, these new risk-adjustment plans led to many HMOs exiting certain markets because of concerns about the fairness of payment.
5. *Public outcry about managed care plans*. Recently, public outcry about quality and access under managed care has increased. The use of anecdotes by media reporters has become commonplace, supporting fears about managed care. In a large effort directed by California elected officials, Alain Enthoven and others investigated the issue of public concern about managed care. They concluded that most individuals are satisfied with their managed care plans. Room for improvement existed in relation to chronic care and some other specific areas. Enthoven and colleagues felt that the increased public backlash reflected the fact that managed care was now the dominant force in health care (fee-for-service or indemnity plans ceased to be sustainable by 1997), with resultant consumer fears and, more importantly, physician outrage at their loss of control. This consumer outcry is now a dominant force in efforts to regulate managed care. Whether these regulations are largely rational and accomplish their stated goals of protecting consumers remains to be seen.

Whatever the final rate of growth, clearly managed Medicare plans will increase their penetration and begin to affect quality and overall cost. When considering the impact of managed Medicare programs on behavioral health care, one must recognize that Medicare beneficiaries, with the exception of the seriously and persistently mentally ill, have used very few behavioral health services. For example, depression and chemical dependency, while recognized as prevalent and costly disorders among elderly adults, are rarely diagnosed and treated in this population. As a result, a relatively small portion (generally less than 3 percent of total HMO costs) of Medicare funds has been spent on behavioral health services. Since most organizations use past use to calculate appropriate capitation rates, the amounts allocated to behavioral health in managed Medicare programs tend to fall short of the real need. In addition, when one considers the impact of the aging of the population, with resultant increases in the prevalence of medical-psychiatric comorbidity (e.g., psychosis among the dementia population), the imbalance between demand or need and available resources becomes even more striking. Balancing these forces are the increasing sophistication of geropsychiatry and gerontology and increased awareness of these problems by general medical clinicians, behavioral health providers, consumers, and public regulatory agencies. It is hoped that this imbalance will begin to correct itself as more data are translated into better information reporting in HCFA and managed care organizations, facilitating more rational allocation of resources. Cost-effectiveness becomes an even more important issue in this context, crying out for efforts to decrease costly and resource-wasteful turf battles (e.g., is dementia really a medical or a psychiatric condition?), decrease inappropriate inpatient use in favor of more flexible community alternatives, and increase recognition of behavioral problems.

In the coming years public policymakers will have to struggle with the regional imbalance in reimbursement rates. Under the current system, managed care plans' rates are set at a percentage of current Medicare expenses in a given market (this is the current adjusted average per capita cost [AAPCC] payment methodology). This leads to regional imbalance, for example, in 1996 the rates paid in Miami, Florida, were 2.4 times greater than those paid in Waco, Texas. This imbalance has historical determinants (the rate relates mainly to variations in the use of more expensive [usually inpatient or home health care] interventions), but as Medicare shifts from a retrospective payment system to a prospective payment system, such differences will no longer be sustainable. In addition, policymakers are discussing lowering the rates paid to plans, at a time when more of the disabled and severely ill population are projected to join managed Medicare plans. Suggesting that lawmakers be advised to increase spending to remedy these problems is unrealistic.

TRENDS AND PRACTICAL ISSUES

Managed behavioral health care has not had its full impact in altering geriatric clinical practice. The changes in the structure and practice patterns in our health care system have affected the geriatric population mainly through a halo effect of changes occurring elsewhere in our health care system. That is, general improvements in health care and accountability for outcomes, rather than a changed approach to care and the delivery system of behavioral health care for the geriatric population, have been the main sources of improved outcomes. Until now, the vast majority of geriatric services have been provided under the traditional system, in which care is monitored retrospectively, and payment is rendered for services without prior involvement of clinical experts or care managers outside the doctor-patient dyad. Under managed care, payments are made prospectively, care is continuously monitored, and payers and clinicians are held accountable for clinical care and service outcomes. Geriatric psychiatry is now entering the maelstrom of health care reform activity, and practices will change dramatically through the first decade of the new century.

Perhaps more than any other group, the elderly population requires extensive integration of general health and behavioral health services. The high degree of

medical-psychiatric comorbidity, the profound social and economic challenges of this population, and the sociobehavioral sequelae of the aging process demand careful, far-reaching integration of care to ensure continuity, comprehensiveness, effectiveness, and efficiency.

Special Behavioral Health Care Needs of the Geriatric Population Behavioral disorders among the geriatric population will receive increased attention within managed systems of care, in part because of increased scrutiny by the NCQA of health plan's approach to behavioral health care. In 1997, NCQA adopted rigorous standards for behavioral health care and access to such care; these standards were applied to general health plans in 1999. Behavioral health disorders are extensive (e.g., some estimates of behavioral disorders among the nursing home population are as high as 80 percent), underdiagnosed, and poorly treated. In addition, the reimbursement restrictions of traditional programs such as Medicare prevent creative problem-solving that could decrease behavioral morbidity and mortality among this group. For example, while Medicare will pay for hospital care and in-home care for a severely depressed, medically ill patient, it will not specifically reimburse a psychiatrist for home visits (outside the construct of a home health program) or pay for assertive clinical case management services that might prevent rehospitalization or enhance functioning. Instead, practitioners must attempt to work within organizational structures that are ill-equipped to meet the specific behavioral health care needs of this population.

Winnie was a 78-year-old woman who had been homebound for over 9 months. She had numerous medical conditions, including diabetes, hypertension, mild congestive heart failure, and uncontrolled hyponatremia. She complained of being severely depressed and socially isolated to her primary care physician, who attempted to have her see a psychiatrist. However, Winnie could not go to the psychiatrist's office because of her difficulties and her concerns about stigmatization. After 6 months of poor progress and frustration, the physician referred her to a home health care agency for “mental status checks.”

The home health agency, the only one in Winnie's small town, had no behavioral health experts on their staff, but sent a nurse to Winnie's home. The nurse performed weekly “mental status checks” that consisted of asking Winnie whether she was depressed or suicidal or both; she also “validated Winnie's experience of feeling socially isolated.” Winnie's physician felt frustrated by her lack of progress and was reluctant to offer her an antidepressant medication because of his lack of experience with these agents in severely medically ill patients. Six months after the home health agency became involved in Winnie's care, Winnie enrolled in a managed Medicare program.

Shortly after Winnie enrolled in the Medicare HMO program, her physician called the managed behavioral healthcare organization (MBHO) responsible for the coordination of her care and asked for assistance in helping Winnie. The MBHO ensured that Winnie was seen by one of its mobile geriatric psychiatrists, who prescribed an antidepressant in coordination with her primary care physician, and by a clinical case manager, who ensured that Winnie was linked to a variety of community services to decrease her social isolation. Three months after these interventions, Winnie's medical illnesses were well controlled; she was euthymic and said she now derived satisfaction from her life. This example demonstrates what can be accomplished when administrative barriers to effective care (e.g., benefits limited to established programs or services such as home health care or office-based psychiatry) are removed. In fact, not only are such interventions clinically effective, they are cost-efficient as well.

Increased Use of Nontraditional Services As in the rest of behavioral health care, creative and flexible programing for this population will increase. Already there is a history of innovation in geriatric health care, through the home health and social HMO programs. In the future, there will be decreased use of services such as partial hospitalization and traditional forms of psychotherapy (both of which have been used less by elderly patients) and increased reliance on community alternatives to them. Such a list of offerings will undoubtedly include use of assertive case management programs and services, individually tailored psychosocial rehabilitation services, community-based chemical dependency detection and treatment services, primary care–based and–coordinated depression detection and treatment, and use of existing providers of geriatric services (e.g., senior centers, peer counseling services).

Chemical Dependency Underdetection Some have suggested that chemical dependency among the elderly population has reached epidemic proportions. Especially when abuse of or dependence upon prescription medications is considered, this set of problems is quite extensive. There are numerous barriers to adequate detection: elderly adults rarely seek behavioral health services, primary care practitioners are notoriously poor at detecting these problems, and chemical dependency problems are generally hidden by patients. Primary care practitioners will increase efforts to involve behavioral health providers in designing and implementing detection programs and then delivering effective chemical dependency treatment programs, probably in community and general health care settings. These interventions will be focused and goal-directed with a cognitive-behavioral base. Residential programs are unlikely to hold much sway for a variety of reasons, not the least of which are patient acceptability and cost. Interventions will be carefully monitored for effectiveness and will be judged on the basis of their inclusion of relapse-prevention techniques and use of geriatric-specific 12-step and other community resources.

Depression Underdetection and Inadequate Treatment Many elderly individuals seek help from their primary care practitioners for behavioral health problems along with their general health concerns. In fact, many primary care visits are used by elderly patients when they need behavioral health care for depression, not for health problems. Finally, many health problems are complicated in this population by unrecognized depression. Primary care practitioners lack the time and expertise to detect depression reliably and accurately. In the Managed Care Outcomes Project, researchers found that within a pool of about 12,000 primary care patients, over 1600 of them either carried a diagnosis of depression or were treated with an antidepressant. Only about half of those diagnosed with depression actually received an antidepressant (over 80 percent of these were prescriptions for non-SSRI antidepressants); about half received a benzodiazepine.

Behavioral health providers will be called upon to design effective and efficient ways to identify depression in primary care settings and then treat these disorders, likely in the medical setting. Of paramount concern to the elderly patient and the primary care physician will be the ability to deliver services in medical settings to avoid stigmatization and enhance treatment coordination.

Through initiatives stimulated by NCQA, emphasis on behavioral health, and primary care integration, health plans have begun to develop programs to address primary care practitioners' inadequate recognition and treatment of depression. Plans can use new information systems to identify specific primary care practitioners who, for example, use inappropriate medications (e.g., benzodiazepines, multiple antidepressants) for the treatment of depression. Once identified, they are targeted for internet-based continuing medical education programs. Following the training, their prescribing practices are tracked for improvement. This last component in the training process, ensuring that the training results in actual practice changes, has been the most difficult to demonstrate; perhaps newly available information technologies will aid in this process.

Medical-Psychiatric Interface Perhaps more than with any other patient group, the coordination of general health and behavioral health care is highly significant for elderly patients. Many have significant, chronic medical problems, and great attention has been paid to this area in recent years. In Veterans Affairs settings, HMOs, and other institutions, models have been used that facilitate efficient coordination of these services. The use of practice guidelines and the increased presence of behavioral health care providers in primary care settings will increase. The critical aspect of guideline development—now that so many of these best practices documents exist—is tracking practitioner behavior to ensure that the guidelines are being followed and best practices are being implemented. Only through managed systems of care, in which accountability is fostered through aggressive reporting, are such improvements possible.

Dementia More vexing than any other will be the challenges of coordinated and comprehensive treatment of individuals with dementia. As treatments for the core cognitive symptoms and associated medical complications of this set of illnesses become more advanced, the need for behavioral health services will surely grow. For example, the ability to target early memory deficits for improvement can be augmented by rehabilitation strategies including skills training, family support and education, and environmental manipulations. Our field must improve performance at the interface between psychiatry and medicine in the dementias to decrease fragmentation of care. Protocols to define the limits of psychiatric and medical expertise are being tested now, enabling practitioners to focus their efforts on improving care. Areas of future concern will include the appropriate role for psychiatry (to provide all care for the cognitive and behavioral aspects of these disorders or only for behavioral abnormalities; or to serve as consultants for gerontology or primary care teams); the most efficient ways to design comprehensive, community-based systems of continued care; and the use of established behavioral interventions to rehabilitate demented patients and prevent further disability.

Special Needs Considerations Because of the population focus inherent in managed care efforts, it is feasible to define the special needs of subpopulations such as frail elderly adults and seriously mentally ill patients. More than a decade of experience in treating frail elderly patients in social HMOs will allow continued refinement of the detection of special needs, accommodations in service access and delivery required by this population, and the effective application of novel techniques.

The aging of the seriously mentally ill population has yet to become the subject of much discussion and debate, aside from the occasional article about long-term outcomes for this population. Managed care initiatives will take account of the special needs of seriously mentally ill elderly patients and effect changes in the way services are delivered. For example, clinical case management (e.g., Programs for Assertive Community Treatment [PACT]) and psychosocial rehabilitation models should be revised so that interventions are tailored to the special needs of these individuals. As a result, knowledge about the effective use of community supports

and resources such as senior centers, meals programs, and peer counseling services for seriously mentally ill elderly patients will enhance service delivery and outcome.

SUGGESTED CROSS-REFERENCES

Section 5.3 discusses mental health services research. Chapter 10 covers dementia and other cognitive disorders, and 51.3e covers dementia in the elderly population. Chapter 11 and Section 51.3h cover substance-related disorders. Section 51.5a discusses Medicare and Medicaid, and Section 51.5d discusses VA psychogeriatric services. A more detailed discussion of managed care is presented in Section 52.2.

SECTION REFERENCES

*Baker LC: Association of managed care market share and health expenditures for fee-for-service Medicare patients. JAMA *281:*432, 1999.

Brown RS: Do health maintenance organizations work for Medicare? Health Care Financ Rev *15:*7, 1993.

*Goldman W, McCulloch J, Sturm R: Costs and use of mental health services before and after managed care. Health Aff *17:*40, 1998.

Goldman: Outpatient utilization patterns of integrated and split psychotherapy and pharmacotherapy for depression. Psychiatr Serv *49:*477, 1998.

Greenberg J, Leutz W, Greenlick M, Malone J, Ervin S, Kodner D: The social HMO demonstration: Early experience. Health Aff *7:*66, 1988.

*Horn SD, Sharkey PD, Gassaway J: Managed care outcomes project: Study design, baseline patient characteristics, and outcome measures. Am J Managed Care *2:*237, 1996.

Iglehart JK: Politics and public health. N Engl J Med *334:*203, 1996.

*Iglehart JK: The struggle to reform Medicare. N Engl J Med *334:*1071, 1996.

*Jensen GA, Morrisey MA, Gaffney S: The new dominance of managed care: Insurance trends in the 1990s. Health Aff *16:*125, 1997.

Meyer JA, Silow-Carroll S, Regenstein M: Managed care and Medicare. Report no. 9614. American Association of Retired Persons, Washington, DC, 1996.

Nelson L: Access to care in Medicare HMOs, 1996. Health Aff *16:*148, 1997.

*Ware J: Differences in four-year health outcomes for elderly and poor chronically ill patients treated in HMO and fee-for-service systems: Results from the medical outcomes study. JAMA *276:*1039, 1996.

Welch HG, Wennberg DE, Welch WP: The use of Medicare home health care services. N Engl J Med *335:*324, 1996.

Wilensky GR, Newhouse JP: Medicare: What's right? What's wrong? What's next? Health Aff *18:*92, 1998.

CHAPTER 51. GERIATRIC PSYCHIATRY

51.5 HEALTH CARE DELIVERY SYSTEMS

51.5C VETERANS AFFAIRS MEDICAL CENTERS AND PSYCHOGERIATRIC SERVICES

WILLIAM W. VAN STONE, M.D., AND THOMAS B. HORVATH, M.D., F.R.A.C.P.

DEFINITION

The Veterans Administration, established by the United States Congress in 1930, was renamed the Department of Veterans Affairs (VA) when it achieved cabinet level status in 1989. The Department, through its Veterans Health Administration (VHA), now administers the largest organized medical care system in the country, including 173 hospitals, 131 nursing homes, 40 domiciliaries, and 470 clinics; 132 VA medical centers are affiliated with 105 accredited medical schools throughout the country. VA health care centers (combinations of hospitals, clinics, nursing homes, and domiciliaries) vary much in size and complexity, but they are all approved by the Joint Commission and their average scores have been 5 to 10 points higher than those of comparable private-sector hospitals. As of January 1999, the VHA employs over 220,000 people and the overall health care budget is 17 billion dollars, of which over $2 billion goes for mental health. The VHA employs 8000 psychiatrists and 1630 psychologists who provide mental health services.

There are 25.1 million veterans, of whom 9.1 percent have service-related disabilities and nearly 4.8 percent are poor. Under the new eligibility rules, over 3.2 million veterans use the VA on an annual basis.

In 1998 the VHA treated 650,000 veterans for mental disorders, including substance abuse, an increase of 20 percent since 1990. Of these, 88,200 needed inpatient care (average length of stay 18 days) and 640,000 received outpatient care with an average of 7 visits per year. On a given working day, there were 8,800 veterans in psychiatric inpatient care and another 117,600 were receiving psychiatric outpatient care.

In recent years numerous government and nongovernment reports have consistently concluded that fundamental changes were needed in the VHA health care system to make it more efficient and responsive to patients, including

- redistribution of VHA health care resources throughout the country to better meet changing regional population needs;
- decentralization of funds, decision making, and clinical care initiatives to regional integrated service networks;
- better access to care through multiple, community-based, ambulatory settings; and
- organization of integrated care delivery around patients' needs with emphasis on primary care teams.

DEMOGRAPHICS

In 1999 veterans age 65 or older numbered 9 million and are estimated to remain in the range of 7.8 to 9 million until the year 2020. Most of these persons will be veterans of World War II and the Korean Conflict until approximately 2010, when aging Vietnam-era veterans will begin to augment that age group. By 2015 the latter cohort will account for 51.9 percent of elderly veterans. Between 1999 and 2010 the number of veterans age 75 years and over will remain stable between 3.7 and 4.4 million.

Throughout the first two decades of the twenty-first century, veterans age 65 years and over will comprise a larger proportion, 37 percent to 46 percent, of the total veteran population than their counterparts in the general population (<20 percent). In 1987 an analysis of discharge data from all VA medical centers showed that approximately one third of all inpatient admissions were geriatric patients who accounted for nearly 41 percent of all hospital days. By 1998 the percentage of discharges of patients over 65 years old had grown to 45 percent. Planning and implementation of programs to meet the needs of the elderly veteran population for physical and mental illness must therefore have a corresponding emphasis.

EPIDEMIOLOGY OF MENTAL DISORDERS IN THE ELDERLY

Although the reported prevalence of mental illness among the elderly varies, conservative estimates for those age 65 years or older include a minimum of 10 percent with dementia of the Alzheimer's type or other dementia, and an additional 15 percent to 30 percent with other psychiatric illness. If an estimate of 30 percent is used, it can be expected that at any time during the period 1999 to 2020 from 2.3 to 2.7 million veterans will need psychogeriatric care. Psychiatric services are generally underused by today's elderly, but that trend is likely to change for VA as the Vietnam-era veterans enter the over-65 age group. Clinical experience suggests that the current, more psychologically oriented generations are less reluctant than their predecessors to acknowledge needing mental health care and thus may be more apt to seek help from the VA.

THEORETICAL ISSUES

Planning for high-quality psychogeriatric care requires the integration of conflicting patient needs in at least two clinical areas: (1) the need for assistance during an ongoing illness versus the need to maintain self-sufficiency and (2) the need for expertise from clinical subspecialists and from multiple clinical disciplines versus the need for one informed primary care provider to integrate the information into a plan. In addition, health care organizations with multiple facilities tend to group patients with similar problems in one location to minimize duplicating scarce specialized services. Travel time to appropriate resources can become a serious barrier to treatment, particularly for elders. These conflicts are playing out in the 22 recently decentralized, cost-conscious VA-integrated service networks (composed of geographically related groups of VA medical centers and clinics).

Levels of Care Older patients presenting with multiple ongoing health challenges must be provided with sufficient support and expertise to manage their symptoms, but not enough to stifle their independence and potential for self-management. In recognition of this wide range of patient needs, in June 1993 the VHA adopted a policy that psychogeriatric programs as well as those in general psychiatry should be organized on a continuum of care that would provide “a full spectrum of interconnected inpatient, outpatient, community, and home-based services which integrate the medical, psychological, social, vocational, and avocational needs of this population in a comprehensive and flexible manner.”

This theoretical continuum of programs has been fleshed out in *Integrated Psychogeriatric Patient Care*, a recent VA program guide created by an ongoing VA psychogeriatric field advisory group. This guide provides recommendations for treatment modalities, organization, space, equipment, staffing, education, research, and quality management. Under today's restricted funding mandates, it is expected that not all potential programs will be vailable or even appropriate at every VA medical center; however, all major components of the continuum must be available within the larger VA integrated service network (VISN). The VHA is experimenting with small, mobile psychogeriatric subspecialty teams to provide consultation, case management, and familial/provider education regarding psychogeriatric issues when needed, along with ongoing care if necessary. Psychogeriatric consultation by teleconference will soon prove to be practical among VA facilities.

TECHNIQUES

In addition to traditional medical and psychiatric assessment, diagnosis, and treatment techniques, the elderly often require additional specialized care to address multiple medical and psychosocial problems. VA psychogeriatric programs are often integrated with those in geriatric medicine as well as with many adjunctive services such as dentistry, dietetics, pharmacy, podiatry, psychology, recreation and rehabilitation therapies, social work, speech and hearing, and the chaplaincy. The Undersecretary for Health has recently declared that VA healthcare is more than just “medical care,” thus approving a whole system of “wraparound services” of a psychosocial nature.

Outreach services and clinical screening tools are often offered in primary care clinics to elderly veterans who may not be inclined to seek mental health services spontaneously. Psychogeriatric consultation provided to the growing numbers of medical primary care teams in VA facilities will enhance their capabilities. Attention to accurate diagnosis and prescription of psychiatric medications to relieve the symptoms of illness, rather than to control behavior, is especially important. In addition to a growing reaction against using medications as a form of restraint for disturbed elderly patients, the use of physical restraints has come under increasing scrutiny, resulting in marked decreases in the use of all restraints in VA settings. Management of disturbing or high-risk behaviors in the elderly requires careful review of diagnoses, the number and dosage levels of medications, and the staff's interpretation of the meaning or intent of the behaviors. Behavioral modification protocols for the elderly tend to be effective only when communication within the caregiving team is open and frequent.

Other psychological interventions for the elderly may include self-awareness training, grief counseling, remotivation therapy, reminiscence review, and intergenerational programs. Staff and family interventions may include behavior management, respite care, environmental modification, and palliative support groups. Organized transportation is considered an essential adjunct to day care programs.

CLINICAL ISSUES

The VA psychogeriatric field advisory group has recommended that clinicians focus more on mild rather than severe mental impairment in the elderly because treatment of the former can prolong the patient's stay at home. The advisory group stressed that training, support, and the availability of respite programs is essential for those who provide care in the patient's home. There is some evidence, however, that respite admissions for those who have dementia may result in a lowered level of functioning and a subsequent increased burden on the caregiver following the patient's return home. The advisory group also recommended a greater emphasis on outreach to patients with untreated depression, because depression can result in intractable physical illness on the one hand or be mistaken for an irreversible dementing illness on the other. Depression may also accompany early clinical dementia.

Clinical practice guidelines for treatment of major depressive disorders including comorbidities of substance abuse and posttraumatic stress disorder were developed by a multidisciplinary group of VA professionals in 1996, distributed to all VA medical centers in 1997, and revised in 1998. These guidelines include a brief screening tool for depression in the elderly to be used in all clinical areas, including primary care, and provide guidance for prescription of psychiatric medications in the elderly as well as special psychosocial interventions. Guidelines were also published for the recognition and treatment of dementia.

Unified Psychogeriatric Biopsychosocial Evaluation and Treatment (UPBEAT), a 3.6 million dollar demonstration project currently being tested at nine VA facilities, is exploring clinical and economic outcomes of screening elderly patients for depression, anxiety, and substance abuse in short-term VA medical and surgical hospital settings. Willing patients with positive symptoms are evaluated by an interdisciplinary psychogeriatric team, after which their progress is followed by case managers. Early findings suggest no reduction in readmissions but a reduction of 5.4 bed days of care for the 12 months following discharge (and consequently lower overall costs) when compared to a randomized “usual care” group.

Recommendations for future psychogeriatric care include using greater care and expertise in the prescription of psychiatric medications, better management of high-risk behaviors, more attention to substance abuse and delayed posttraumatic stress disorder in older veterans, and fuller collaboration with geriatricians regarding management of medical comorbidities, such as chronic obstructive pulmonary disease, diabetes, hypertension, and cardiovascular disease. Psychogeriatric intervention in existing VA and non-VA nursing homes, day care, and home care programs also should be expanded.

ETHICAL ISSUES

Geriatric psychiatrists may be asked for an evaluation of elderly patients' competency and of their ability to understand treatment choices. The VA National Cost Containment Center, in collaboration with the Mental Health Strategic Health Care Group, has published clinical guidelines for the determination of competency. Geropsychiatric advice may also be sought regarding the ailing patient's desires regarding life-sustaining technology where preservation of life versus quality of life issues often present ethical dilemmas. The VA's Clinical Ethics Center at White River Junction published a helpful series of ethical guidelines.

RESEARCH AND EVALUATION

A computer search of medical research articles published since 1990 on geriatric psychiatry and veterans revealed 66 that are relevant. Of these, 27 addressed aspects of posttraumatic stress disorder, including studies of exprisoners of war; 14 explored alcohol abuse, its detection and treatment; 10 studied Alzheimer's disease and related disorders, and the rest addressed other aspects of illness in elderly veterans.

Some major themes expressed in these articles were:

1. Posttraumatic stress disorder is underdiagnosed in elderly veterans, of whom 35 to 52 percent over age 55 were reportedly exposed to combat. Symptoms of posttraumatic stress disorder are less severe in the older veterans as compared to those from the Vietnam era, but are, for veterans of any age, proportional to the actual exposure to combat stress and may be associated with anatomical changes in the brain. Exprisoners of war from any era are especially at risk.
2. Supportive alcohol programs designed especially for elderly patients with alcoholism are more effective than those typically designed for younger persons.
3. Institutionalized veterans with a long-term mental illness fared better when given maximal control over their daily lives.

SUGGESTED CROSS-REFERENCES

An introduction and overview of geriatric society appears in Section 51.1. Dementia of the Alzheimer's type is discussed in Chapter 10 on delirium, dementia, and amnestic and other cognitive disorders and in Section 51.3e on Alzheimer's disease and other dementing disorders. Mood disorders are presented in Chapter 14 and Section 51.3d on mood disorders in late life.

SECTION REFERENCES

Das D, Orengo CA, Kunik ME, Molinari V, Workman RH: Cognitive decline in patients on an acute geropsychiatric unit. J Neuropsychiatry Clin Neurosci *10:*205, 1998.

Department of Veterans Affairs, Health Administration Program Guide 1103.3: *Mental Health Program Guidelines for the New Veterans Health Administration*. Veterans Health Administration, Washington, DC, February 1999.

*Department of Veterans Affairs, Office of Clinical Programs: *Integrated Psychogeriatric Patient Care, Program Guide PG 10–22, M-2, Part X*. Veterans Health Administration, Washington, DC, 1996.

*Druss BG, Rohrbaugh RM, Rosenheck RA: Depressive symptoms and health costs in older medical patients. Am J Psychiatry *156:*477, 1999.

Fortney JC, Booth BM, Blow FC, Bunn JV: The effects of travel barriers and age on the utilization of alcoholism treatment aftercare. Am J Drug Alcohol Abuse *21:*391, 1995.

Green JA, Wagner J, Johnson W: Development of a geropsychiatric unit. South Med J *87:*392, 1994.

Hirsch CH, Davies HD, Boatright F, Ochango G: Effects of a nursing home respite admission on veterans with advanced dementia. Gerontologist *33:*523, 1993.

Kashner TM, Rodell DE, Ogden SR, Guggenheim FG, Karson CN: Outcomes and costs of two VA inpatient treatment programs for older alcoholic patients. Hosp Commun Psychiatry *43:*985, 1992.

*Kunik ME, Snow-Turek AL, Iqbal N, Molinari VA, Orengo CA, Workman RH, Yudofsky SC: Contribution of psychosis and depression to behavioral disturbances in geropsychiatric inpatients with dementia. J Gerontol A Biol Sci Med Sci *54:*M157, 1999.

*Lee KA, Vaillant GE, Torrey WC, Elder GH: A 50-year prospective study of the psychological sequelae of World War II combat. Am J Psychiatry *152:*516, 1995.

*Moos RH, Mertens JR: Patterns of diagnoses, comorbidities, and treatment in late-middle aged and older affective disorder patients: Comparison of mental health and medical sectors. J Am Geriatr Soc *44:*682, 1996.

Morton JL, Jones TV, Manganaro MA: Performance of alcoholism screening questionnaires in elderly veterans. Am J Med *101:*153, 1996.

Palmer AC, Withee BM: Dementia care: Effects of behavioral intervention training on staff perception of their work in veterans' nursing home. Geriatr Nurs *17:*137, 1996.

*Roughan PA: Mental health and psychiatric disorder in older women (Review Article). Clin Geriatr Med *9:*173, 1993.

Sajatovic M, Ramirez LF, Garver D, Thompson P, Ripper G, Lehmann LS: Clozapine therapy for older veterans. Psychiatr Serv *49:*340, 1998.

Spiro A III, Schnurr PP, Aldwin CM: Combat-related post-traumatic stress disorder symptoms in older men. Psychol Aging *9:*17, 1994.

Timko C, Nguyen AT, Williford WO, Moos RH: Quality of care and outcomes of chronic mentally ill patients in hospital and nursing homes. Hosp Community Psychiatry *44:*241, 1993.

Valenstein M, Kales H, Mellow A, Dalack G, Figueroa S, Lawton BK, Blow FC: Psychiatric diagnosis and intervention in older and younger patients in a primary care clinic: Effect of a screening and diagnostic instrument. J Am Geriatr Soc *46:*1499, 1998.

Van Stone WW, Goldstein MZ: Mental health services for older adults in the VA system. Hosp Community Psychiatry *44:*828, 1993.

Veterans Health Administration: *Assessment of Competency and Capacity of the Older Adult: A Practice Guideline for Psychologist*. VHA, Washington, DC, March 1997.

*Veterans Health Administration: Clinical Guidelines for Major Depressive Disorder (MDD), MDD with Post Traumatic Stress Disorder (PTSD), and MDD with Substance Abuse (SA). Veterans Health Administration, Washington, DC, 1998.

Veterans Health Administration: *Dementia Identification and Assessment: Guidelines for Primary Care Practitioners*. Veterans Health Administration, Washington, DC, 1997.

51.5 HEALTH CARE DELIVERY SYSTEMS

51.5D COMMUNITY SERVICES FOR THE ELDERLY PSYCHIATRIC PATIENT

BARRY D. LEBOWITZ, PH.D.

A number of excellent reviews have focused on issues of mental health care for older persons. These reviews reflect widespread agreement on the components of an ideal system: a full multidisciplinary range of clinical, rehabilitative, preventive, and supportive services; comprehensive assessment; case management; aggressive outreach; and seamless coordination of mental health, physical health, and social services. There is also widespread agreement that such a system is a fiction, rarely approached even in the best of circumstances. Five issues are particularly relevant: (1) the growing importance of managed care, (2) the enhanced significance of primary care, (3) the place of comprehensive assessment, (4) the redefinition of outcomes, and (5) the clarification of the role of the family and of the institution.

IMPORTANCE OF MANAGED CARE

The traditional private practice fee-for-service approach to the provision of medical care in the United States is being supplanted by various organized systems referred to as managed care. These systems are characterized by financial arrangements based on capitation, geographical coverage, emphasis on ambulatory care, vertical integration of primary care providers and community hospitals, referral specialists, and restricted use of tertiary care and the academic health center. There are many forms of managed care, including health maintenance organizations and preferred-provider organizations. These organizations use many approaches to control costs by monitoring physician behavior, establishing formal practice guidelines, and routinely applying utilization review and prior approval procedures for all nonemergency care. Another development in managed care has been the proliferation of so-called behavioral health carve out companies. *Behavioral health* defines a broad range of mental health, substance abuse, counseling, employee assistance, and crisis intervention services. These organizations contract directly with employers or with managed care organizations to provide these specialized services under some sort of fixed fee arrangement. This organization then assumes all responsibility for the covered services.

There is great concern that this emerging pattern will result in significant problems because of the financial benefit in failing to recognize, diagnose, and appropriately treat mental disorders. This risk is particularly great with respect to the care of geriatric patients, who characteristically exhibit complex patterns of comorbidity and who may be seriously disadvantaged by the separation of behavioral health services from other health services.

SIGNIFICANCE OF PRIMARY CARE

The scope of responsibility of the primary care sector is being expanded and redefined in the emerging health care system in the United States. At the level of training, emphasis is being moved from specialty and subspecialty care to primary care; the stated goal of many involved in system reform is to allocate as many as one-half the post-medical degree residency positions to primary care disciplines. The primary care physician will assume greater responsibility for diagnosis, treatment, and long-term management in all areas of health care, including care of older patients with mental disorders. Moreover, in most organized systems, access to specialty and subspecialty care will be through the primary care provider who will serve as a referral resource, case manager, and gatekeeper.

A rough rule of thirds characterizes the present situation of care of older persons in need of mental health services: approximately one in three receives care from a mental health professional, another third is treated in primary care, and another third receives no care at all. Compared with midlife adults, older people are more likely to receive mental health care from general physicians than from any other type of professional.

The reliance on nonspecialty care raises concerns about the appropriateness of the treatments provided. General clinicians too often prescribe drugs with significant anticholinergic and mood-altering effects for their elderly patients, even when more appropriate treatments are available. Comorbidity raises other types of concerns. The quality of care and the long-term outcome of treatment for mental disorder are significantly improved in mental health specialty care, medical services are of higher quality, and physical health outcomes are improved in the general health care sector. The widespread use of treatment or practice guidelines, such as those from the Agency for Health Care Policy and Research or those generated within managed care systems or professional organizations, could possibly improve the appropriateness of treatment.

COMPREHENSIVE ASSESSMENT

Comprehensive assessment is one of the foundations of geriatric care. Recent studies clearly demonstrate the benefit of multidisciplinary comprehensive assessment for community-residing older persons. Among the major differences observed between intervention and control groups were significantly improved diagnosis of depression and cognitive impairment and a reduction in long-stay nursing home admissions.

REDEFINITION OF OUTCOMES

As part of the ongoing process to reform the health care system in the United States, treatment approaches in many areas of medicine are subject to intense scrutiny. One such effort is directed toward the efficacy of treatment for severe mental disorders, including disorders of late life. The overwhelming conclusion of this analysis is that a variety of rigorously tested treatments is available, efficacious, and affordable. At the same time, it is recognized that most treatment assessment studies have been too narrowly focused on symptom reduction, and that attention must be paid to the impact of treatment on significant public health factors such as function, disability, quality of life, health care resource use, institutionalization, morbidity, and mortality.

FAMILY AND INSTITUTION

The design and operation of community services for older persons are best seen as complementary to those services provided by the family and by the institution, particularly the nursing home. Using Alzheimer's disease as the model of the effects of chronic stress on the family members responsible for the ongoing care of the disabled older person, research has documented both the psychosocial and psychobiological parameters of the burden of care. Importantly, however, studies have shown that interventions directed at the family caregiver, and not the patient, can have effects on patient outcomes and can delay institutional placement for a significant period.

The nursing home is a critical and unique component of the continuum of community services for older persons with mental disorders. There are no alternatives in the community for patients who require a comprehensive level of care. The proliferation of community based services has resulted in an increased need for nursing home care because the provision of care appropriate to the needs of the chronically disabled results in substantial increases in life expectancy that, in turn, amplifies and expands the role of institutional care. “Alternatives to institutionalization” are not, in fact, alternatives at all and do not displace the need for institutional care.

Community services for the elderly have been developed out of a complex mix of clinical needs, available resources, policy determinations, and political activities. Currently all institutions are under intense scrutiny, and the role and responsibility of government in the provision of services is being reevaluated. Older persons, with their complex patterns of comorbidity and need for ongoing multidisciplinary care, are extremely vulnerable to the plans of those whose notions of health care consist exclusively of neatly balanced graphs, charts, and tables. The challenge is to ensure that the needs of the patients remain at the center of all these discussions.

SUGGESTED CROSS-REFERENCES

An introduction and overview of geriatric society appears in Section 51.1a. Alzheimer's disease and dementia of the Alzheimer's type are discussed in Chapter 10 on delirium, dementia, and amnestic and other cognitive disorders and in Section 51.3e on Alzheimer's disease and other dementing disorders. Mood disorders are presented in Chapter 14 and Section 51.3d on mood disorders in late life. Managed care is discussed in Section 52.2; the impact of managed care on the elderly is discussed in Section 51.5b.

SECTION REFERENCES

Bartels SJ, Colenda CC: Mental health services for Alzheimer's disease: Current trends in reimbursement and public policy. Am J Geriatr Psychiatry *6*(Suppl):S85, 1998.

Burns BJ, Taube C: Mental health services in general medical care and in nursing homes. In *Mental Health Policy for Older Americans: Protecting Minds at Risk*, BS Fogel, A Furino, GL Gottlieb, editors. American Psychiatric Press, Washington, DC, 1990.

*Colenda CC, Streim J, Greene JA, Meyers N, Beckwith E, Rabins P: The impact of OBRA '87 on psychiatric services in nursing homes. Am J Geriatr Psychiatry *7:*12, 1999.

Depression Guidelines Panel: *Depression in Primary Care*, vol 2, *Treatment of Major Depression*. Clinical practice guideline no. 5. Rockville, MD. U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research. AHCPR publ no. 93-0551, April, 1993.

Estes CL: Mental health services for the elderly: Key policy elements. In *Emerging Issues in Mental Health and Aging*, M Gatz, editor. American Psychological Association, Washington, DC, 1995.

Gatz M, Smyer MA: The mental health system and older adults in the 1990s. Am Psychol *47:*741, 1992.

George LK: Community and home care for mentally ill older adults. In *Handbook of Mental Health and Aging*, ed 2, JE Birren, RB Sloane, GD Cohen, editors. Academic Press, San Diego, CA, 1992.

Gold MR, Hurley R, Lake T, Ensor T, Berenson R: A national survey of the arrangements managed-care plans make with physicians. N Engl J Med *333:*1678, 1995.

*Gottlieb GL: Barriers to care for older adults with depression. In *Diagnosis and Treatment of Depression in Late Life*, LS Schneider, CF Reynolds, BD Lebowitz, AJ Friedhoff, editors. American Psychiatric Press, Washington, DC, 1994.

Gruenberg EM: The failure of success. Milbank Mem Fund Q *55:*3, 1977.

*Knight BG, Kaskie B: Models for mental health service delivery to older adults. In *Emerging Issues in Mental Health and Aging*, M Gatz, editor. American Psychological Association, Washington, DC, 1995.

Lebowitz BD: Mental health and aging. Generations *16:*65, 1993.

Lebowitz BD, Gottlieb GL: Clinical research in the managed care environment. Am J Geriatr Psychiatry *3:*21, 1995.

Lebowitz BD, Niederehe G: Concepts and issues in mental health and aging. In *Handbook of Mental Health and Aging*, ed 2, JE Birren, RB Sloane, GD Cohen, editors. Academic Press, San Diego, CA, 1992.

*Lebowitz BD, Pearson JL, Schneider LS, Reynolds CF, Alexopoulos GS, Bruce ML, Conwell Y, Katz IR, Meyers BS, Morrison MF, Mossey J, Niederehe G, Parmelee P: Diagnosis and treatment of depression in late life: Consensus statement update. JAMA *278:*1186, 1997.

Light E, Niederehe G, Lebowitz BD, editors. *Stress Effects on Family Caregivers of Alzheimer's Patients*. Springer, New York, 1994.

Mittelman MS, Ferris SH, Shulman E, Steinberg G, Ambinder A, Mackell JA, Cohen J: A comprehensive support program: Effect on depression in spouse-caregivers of AD patients. Gerontologist *35:*792, 1995.

*Mittelman MS, Ferris SH, Steinberg G, Shulman E, Mackell JA, Ambinder A, Cohen J: An intervention that delays institutionalization of Alzheimer's disease patients: Treatment of spouse caregivers. Gerontologist *33:*730, 1993.

*National Advisory Mental Health Council. *Parity in Financing Mental Health Services: Managed Care Effects on Cost, Access & Quality*. National Institutes of Health NIH 98-4322, Bethesda, MD, 1998.

National Technical Assistance Center for State Mental Health Planning. *Planting the Seeds of Change: Developing Mental Health and Aging Coalitions to Improve Services for Older Persons with Mental Illness*. National Center, Alexandria, VA, 1997.

Niederehe G, Schneider LS: Treatments for depression and anxiety in the aged. In *A Guide to Treatments that Work*, PE Nathan and JM Gorman, editors. Oxford, New York, 1998.

Norquist G, Wells KB, Rogers WH, Davis LM, Kahn K, Brook R: Quality of care for depressed elderly patients hospitalized in the specialty psychiatric units or general medical wards. Arch Gen Psychiatry *52:*695, 1995.

Olfson M, Pincus HA: Outpatient mental health care in nonhospital settings: Distribution of patients across provider groups. Am J Psychiatry *153:*1352, 1996.

Pearson JL, Reynolds CF, Kupfer DJ, Lebowitz BD: Outcome measures in late life depression. Am J Geriatr Psychiatry *3:*191, 1995.

Schulz R, O'Brien AT, Bookwala J, Fleissner K: Psychiatric and physical morbidity effects of dementia caregiving: Prevalence, correlates, and causes. Gerontologist *35:*771, 1995.

Silverman M, Musa D, Martin DC, Lave JR, Adams J, Ricci EM: Evaluation of outpatient geriatric assessment: A randomized multi-site trial. J Am Geriatr Soc *43:*733, 1995.

Stuck AE, Aronow HU, Steiner A, Alessi CA, Bula CJ, Gold MN, Yuhas KE, Nisenbaum R, Rubenstein LZ, Beck JC: A trial of annual in-home comprehensive geriatric assessments for elderly people living in the community. N Engl J Med *333:*1184, 1995.

Willcox SM, Himmelstein DU, Woolhandler S: Inappropriate drug prescribing for the community-dwelling elderly. JAMA *272:*292, 1994.

51.6 SPECIAL AREAS OF INTEREST

51.6A PSYCHIATRIC ASPECTS OF LONG-TERM CARE

IRA R. KATZ, M.D., PH.D., JOEL E. STREIM, M.D., AND BUSTER D. SMITH, M.D.

The concept of long-term care has expanded in recent years to include a spectrum of health services required by persons with illnesses and disabilities that interfere with their ability to care for themselves. It can include convalescent care during periods of recovery as well as life-long care for those with irreversible disability. Long-term care can be provided across a wide variety of settings for delivery of health care, including community settings such as home, day-care centers, partial hospital programs, nursing homes, and other types of assisted living or residential care settings.

With the aging of the population in the United States and the resulting increase in the number of elderly patients with chronic diseases and disability, long-term care has become an increasingly important issue in political, economic, and clinical terms. The magnitude of the problem can be estimated by recognizing the number of people who are dependent with respect to activities of daily living. That figure increases from 3.5 percent of those age 65 to 74, to 35 percent of those 85 years of age and older. Although the proportion of people living in nursing homes increases from 1 percent of those age 65 to 74, to 24 percent of those 85 and older, most older patients with self-care deficits live in the community, and most long-term care for elderly persons is in the form of informal assistance, usually provided in the home by the patient's family.

A number of recent changes have occurred in the delivery of long-term care. Over the 10 years from 1985 to 1995, the number of nursing home beds increased by 9 percent to 1.8 million. However, during this time, the number of homes decreased by 9 percent to 16,600, and their occupancy rate decreased by 4.8 percent to 87.4 percent. In spite of an 18 percent increase in the population over age 65, the number of nursing home residents increased by only 3.8 percent to 1.55 million. As a result of these changes, the ratio of nursing home residents to the general population over age 65 decreased from approximately 46.2 per 1000 in 1985 to 41.3 in 1995.

The decline in the proportion of elderly adults residing in nursing homes probably reflects significant growth in other forms of long-term care. For example, in recent years about 11,000 home health and hospice agencies provided care to nearly 2 million patients, and a number of major corporations have begun to view assisted living for older adults as a major growth industry. Other recent changes in long-term care have occurred within nursing homes in response to both marketplace incentives and clinical needs. Nursing homes now offer many of the rehabilitation services historically provided on inpatient units. Many nursing facilities have subacute units that provide step-down care for medical, surgical, or poststroke patients. Others have established special care units for patients with Alzheimer's disease.

Although long-term care options and settings are becoming more diverse, the major recent advances in scientific knowledge about the mental health components of long-term care are mostly from studies in nursing homes. Nursing homes are important sites for mental health services. Results from research in nursing homes can be applied to comparable patients in other long-term care settings.

NURSING HOMES: PROBLEMS AND REFORMS

Nursing homes account for approximately 4.1 percent of elderly Americans, with 1.55 million older people living in nursing homes at any time. This figure, however, greatly underestimates the number of individuals who utilize nursing home services. In 1995 1.7 million people were admitted to nursing homes; two thirds to three fourths of them were discharged after several months. It has been estimated that the number of nursing home residents will double by the year 2020 and triple by 2040, and that the probability that any person will require nursing home care at some point in his or her life exceeds 40 percent. The total costs of institutional care for elderly patients in the United States were estimated for 1993 at $74.9 million, with $36.9 million coming from Medicaid and $4.8 million from Medicare.

Nursing home care represents approximately 8 percent of the nation's total health care costs, but the burden of payment is markedly different from that of other aspects of health care, with approximately 50 percent of costs borne directly by residents and their families and most of the remainder shared equally by the federal and state governments through Medicaid programs. Medicare payments are limited to subacute care. In spite of recent trends toward a leveling of growth, the number of nursing homes and nursing home beds has increased dramatically since the mid-1960s when Medicaid programs were first developed. In parallel, there has been increasing recognition of the problems with nursing home care and the need for reform. As stated in a 1986 Institute of Medicine report “Improving the Quality of Care in Nursing Homes,” many nursing homes provide good care,

> but in many other government certified nursing homes, individuals who are admitted receive very inadequate—sometimes shockingly deficient—care that is likely to hasten the deterioration of their physical, mental, and emotional health. They also are likely to have their rights violated and may even be subject to physical abuse.

Several major issues raised about the quality of nursing home care were related to its psychiatric aspects; they included concerns that physical and chemical restraints were inappropriately being used to control residents' behavior and that psychiatric disorders (primarily depression) were being undertreated. Estimates of the prevalence of the use of physical restraints, such as wrist or ankle cuffs, belts, vest restraints, or geriatric chairs designed to restrict mobility, ranged from 25 to 85 percent. Use of these devices was endemic in spite of evidence from observational studies that they do not safely control agitated behavior as well as suggestions from cross-national studies that similar patient populations can be managed without their use. Surveys of medication use in long-term care facilities found prevalence rates of up to 74 percent for medications that act on the central nervous system.

Antipsychotic agents have been the most frequently prescribed psychotropic medications, with most reports of prevalence in the range of 20 to 50 percent. Evidence suggesting the misuse of those drugs came from findings that variables unrelated to patient characteristics (e.g., facility size, ratio of patients to staff, and the size of the physician's nursing home practice) were directly associated with use of drugs. Substantial numbers of patients received psychotropic drugs without any diagnosis of mental disorder and without any chart note indicating the presence of relevant target symptoms. The importance of this issue for patients, families, and advocacy groups was poignantly captured by the newspaper headline “America's Other Drug Problem.”

Advocates for improving the quality of long-term care were concerned about issues such as the misuse of restraints and the lack of mental health services for patients who needed long-term care because of dementia or psychiatric disorders associated with medical illnesses. However, as policy evolved, these issues became confounded with questions of whether patients with schizophrenia and other persistent and severe psychiatric disorders belong in nursing homes, whether they could receive appropriate treatment in these settings, and whether they were being admitted to nursing homes by states primarily to shift the costs of their care to the federal government. In 1987 the federal government, the largest single source of payment for nursing home care, used Medicaid and Medicare legislation to mandate nursing home reform that was enacted in the Omnibus Budget Reconciliation Act (OBRA) of 1987. OBRA 1987 (with modification in 1990) required preadmission screening and annual resident review (PASARR) to ensure that patients who belong in psychiatric hospitals were not inappropriately admitted to nursing homes. Other regulations stated that “the resident has a right to be free from any physical and chemical restraints imposed for purposes of discipline or convenience and not required to treat the resident's medical symptoms.” OBRA ‘87 requires that “residents who have not used antipsychotic drugs are not given these drugs unless antipsychotic drug therapy is necessary to treat a specific condition as diagnosed and documented in the clinical record,” and “residents who use antipsychotic drugs receive gradual dose reductions and behavioral interventions, unless clinically contraindicated, in an effort to discontinue these drugs.” Regulations also state that each resident must be free from unnecessary drugs, which are defined as any drug used without appropriate indications, in excessive dose, for excessive duration, without adequate monitoring, or in the presence of adverse consequences that indicate that the medication should be reduced or discontinued. Although the definition of unnecessary drugs is rather general, guidelines used by surveyors to monitor facilities' compliance with OBRA regulations focus on antipsychotic, anxiolytic, and

sedative-hypnotic drugs. The primary goal in the implementation of the antipsychotic drug and unnecessary drug regulations is to decrease the use of medications as chemical restraints.

In addition to the PASARR and psychotropic drug regulations, there are also requirements related to the treatment of mental health problems. Under the general quality-of-care provisions that “each resident must receive and the facility must provide the necessary care and services to attain or maintain the highest practicable physical, mental, and psychosocial well-being,” there is a specific provision related to mental health, stating that a facility must ensure that “a resident who displays mental or psychosocial adjustment difficulty receives appropriate treatment and services to correct the assessed problem.” However, implementation of these OBRA requirements for mental health treatment has lagged behind the other provisions related to psychiatric disorders, despite compelling evidence in the clinical literature that psychiatric treatment is a critical part of nursing home care.

OBRA regulations state that each facility's staff must conduct regularly scheduled assessments of all residents using a standardized instrument, the Minimum Data Set (MDS), which was designed to identify clinical problems, including mental health problems. The MDS includes sections related to mental health and mental illness that evaluate customary routines, cognitive patterns, mood and behavior patterns, psychosocial well-being, and activity pursuit patterns. When certain trigger signs are identified, facilities must complete resident assessment protocols (RAPs) that provide guidelines for determining whether the resident's treatment plan should be modified. RAPs address delirium, cognitive loss dementia, psychosocial well-being, mood state, behavior symptoms, activities, psychotropic drug use, and physical restraints. The MDS has been proposed for use as a tool for outcome assessment and research as well as clinical management.

Research indicates that since OBRA reforms, the use of antipsychotic medications in nursing homes has declined by more than one third. Moreover, available data suggest that the decrease in use of these agents has not been accompanied by substantial increases in use of other medications or marked increases in psychiatric or behavioral symptoms. Although OBRA reforms have had some positive effects on psychotropic drug–prescribing practices, they have not, by themselves, decreased the overuse of psychotropic drugs, nor have they reversed the problem of undertreatment of psychiatric disorders. Controlled studies have demonstrated that even after OBRA, educational interventions can decrease overall use of antipsychotic agents while helping to target appropriate patients for treatment. Research over the past decade has also confirmed both the benefits of psychopharmacological treatment in the nursing home (e.g., for major depressive disorder) and the risks of all classes of psychotropic drugs (e.g., as causes of falls). Thus, both clinical and epidemiological data demonstrate that even after implementation of OBRA regulations intended to limit the use of psychotropic medication, there is still a need for appropriate pharmacological treatment of psychiatric disorders.

Nationwide concern over the growing costs of health care for the elderly has led to ongoing debates about how to pay (and who should pay) for long-term care. Federal initiatives to decrease “entitlement” expenditures are giving the states greater discretion over how they spend limited Medicaid funds. As the states experience growing financial pressures, this could undermine recent advances in access to mental health care for elderly nursing home residents.

Medicare reimbursement for nursing home treatment under the most common mental health procedure codes increased 244 percent, from $20.8 million in 1991 to $71.7 million in 1993. During this time, Medicare costs for these codes across all settings increased by only 57 percent, from $353 million to $555.6 million. Although this might have been interpreted as evidence for improvements in case recognition and treatment in nursing homes, it led instead to concerns by the Department of Health and Human Services that Medicare might have been billed for unnecessary or inappropriate services. These concerns suggest the need to develop guidelines for cost-effective and accountable provision of medically necessary treatment, with standards to ensure that nursing home residents who have psychiatric disorders are not undertreated.

MENTAL HEALTH IN NURSING HOMES: CLINICAL ISSUES

Epidemiology Alvin Goldfarb, who investigated nursing home psychiatric care in 1962, estimated the prevalence of psychiatric disorders among nursing home residents at 87 percent. Subsequent prevalence estimates based upon systematic psychiatric evaluations of residents have been comparable. In 1990 Barry Rovner and coworkers found diagnosable psychiatric disorders in 80 percent of a series of new admissions to nursing homes; and Patricia Parmelee and coworkers, who included milder depressions, found approximately 90 percent. In the study by Rovner's group, newly admitted nursing home residents were interviewed by psychiatrists. The distribution of diagnoses is given in Table 51.6a-1. Although the work was completed before implementation of OBRA regulations, only 8.5 percent of the residents admitted to the nursing facility, or 10.6 percent of those with a diagnosable psychiatric disorder, had a history of previous psychiatric hospitalization. That statistic demonstrates that the high prevalence of psychiatric disorders among nursing home residents cannot be attributed to transinstitutionalization and the admission of chronic psychiatric patients. This is consistent with the findings of Bernard Stotsky, who in 1967 demonstrated a high prevalence of psychiatric disorders among allegedly nonpsychiatric nursing home residents. This is explained by conditions such as dementia and depression related to disabling medical and neurological illnesses that make nursing home care necessary.

Any psychiatric disorder	80.2
Dementia (all)	67.4
Complicated by:	
Depression	6.3
Psychosis	13.5
Delirium	7.0
Depression	10.4
Schizophrenia and other psychoses	2.4

Adapted from Rovner BW, German PS, Broadhead J, Morris RK, Brant LJ, Blaustein J, Folstein MF: The prevalence and management of dementia and other psychiatric disorders in nursing homes. Int Psychogeriatr 2:13, 1990.

Table 51.6a-1 Prevalence of Psychiatric Disorders in the Nursing Home (Percentage)

Dementia Dementia of the Alzheimer's type and other dementias are the most common psychiatric disorders among nursing home residents. There have long been concerns that many patients misdiagnosed as suffering from a dementia are inappropriately admitted to nursing homes. Clearly, all patients should receive a comprehensive evaluation before decisions are made about their need for long-term care. Such evaluations, however, may not always be completed before nursing home admission, especially in patients with acute medical problems; in such cases the evaluations should be conducted as soon as possible after admission. Identification and treatment of reversible cognitive impairment can improve the quality of life of a nursing home resident or facilitate return to the community. Thomas Sabin and coworkers reported that 52 of 111 (47 percent) nursing home residents with dementia had a potentially treatable component of their cognitive impairment. Rovner and colleagues found that approximately 7 percent of a series of new admissions to a nursing home (10 percent of those with dementia) exhibited symptoms of a superimposed delirium. The first author and coworkers found that 6 to 12 percent of the cognitively impaired residents of a long-term care facility actually exhibit significant improvements in cognitive performance over a 1-year period as a result of toxic and metabolic encephalopathies or depression.

Of the irreversible dementias found in nursing home residents, dementia of the Alzheimer's type is the most common, followed by vascular dementia. Estimates of the frequency of vascular causes or components of dementia vary from about one third to as much as one half of the total in various studies. Although both types of dementia are irreversible, diagnostic evaluation that distinguishes vascular dementia from dementia of the Alzheimer's type can trigger appropriate medical interventions to decrease the rate of progression.

Psychoses and depression may occur with dementia. Estimating the prevalence of those syndromes is difficult because standard methods of assessment may fail in the presence of severe cognitive deterioration. Nevertheless, available data suggest that hallucinations and delusions are seen in approximately 20 percent of demented residents, that a major depressive disorder occurs in approximately 10 percent, and that less severe depression occurs in a comparable number. Other, less well-defined behavioral disturbances are common; according to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV), they can be documented in diagnoses of dementia by using the specifier “with behavioral disturbance.” They may include states of both agitation and apathy or withdrawal.

Jiska Cohen-Mansfield and coworkers define agitation as “inappropriate verbal, vocal, or motor activity that is not judged by an outside observer to result directly from the needs or confusion of the agitated individual.” They emphasize that agitation is a symptom, not a diagnostic term. Agitated behaviors can be symptoms of more specific psychiatric syndromes; aggressive behaviors can be symptoms of psychosis, and screaming, a symptom of depression. Agitation can also occur (usually in

patients with dementia) in the absence of other psychiatric symptoms. Factor analysis has identified three types of agitation that are common in nursing home residents: aggressive behavior, nonaggressive physical behavior, and verbally agitated behavior. Although most nursing home residents exhibit one or more agitated behaviors during the course of a week, a smaller number show disturbances of such frequency and severity that they interfere with needed care, represent sources of danger, or cause excess disability.

There is little information on the extent to which agitation can be attributed to biological factors as opposed to problems resulting from unmet needs, environmental irritants, or social conflicts. Evidence for a biological component comes from the finding that tritiated imipramine binding in platelets is lower in Alzheimer's disease patients with agitation or delusions than in controls without behavioral disturbances. Apathy has been less well investigated, but there is evidence that it can be distinguished from depression and that apathetic, withdrawn geriatric patients may respond to stimulant medication.

There has been a movement toward developing special care units for patients with Alzheimer's disease and related disorders that have architecture, administrative structures, staffing, and programing designed to maximize the level of functioning and minimize behavioral disturbances. At present, about 15 percent of nursing homes have some form of special care unit or specialized program for patients with Alzheimer's disease. Although this represents a total of 2500 units and more than 30,000 beds, investigators are still developing specifications for special care units and methods for assessing the outcomes of care. Rovner and his colleagues conducted a randomized clinical trial of a dementia care program that included augmented activities, guidelines for use of psychotropic drugs, and educational rounds and found that active treatment significantly decreased both behavioral symptoms and use of psychotropic drugs. Another type of special care, use of a hospice model for the care of nursing home residents who are in the terminal stages of Alzheimer's disease, may reduce the patients' discomfort, improve their quality of life, and reduce costs.

Depression The National Institutes of Health Consensus Development Conference Statement on the Diagnosis and Treatment of Depression in Late Life emphasized the importance of depression in the nursing home as a problem requiring the attention of clinicians, researchers, and policy makers. Depressive symptoms are associated with increased disability, increased nursing care needs, decreased participation in activities, increased reports of pain, biochemical evidence of subnutrition, increased risk of cognitive deterioration, and treatment refusal.

Recent studies suggest that depressive disorders are the most common functional disorders in the nursing home and related settings. The prevalence of major depression among cognitively intact nursing home residents has been estimated at approximately 25 percent. Although it is an order of magnitude greater than the prevalence observed among older adults in the community, the higher prevalence may have been expected because of the universal presence of chronic illness and disability in long-term care patients. Minor depressive disorder and self-rated dysphoria are also clinically significant in that they appear to represent risk factors for major depressive disorder; hence, treatment of those disorders represents an opportunity for prevention as well as for symptom relief. The importance of depression follows from the fact that it may be a treatable source of distress and increased disability, even in the presence of other irreversible disabling diseases. The coexistence of depression and physical illness, however, raises concern about the validity of the use of standard approaches to psychiatric diagnosis under conditions in which somatic or neurovegetative symptoms of depression may lose their diagnostic specificity. There are also questions about relation between the symptoms of depression and physical illness; and the initial evaluation of nursing home residents with depression must therefore include a medical evaluation seeking specific medical causes of affective and neurovegetative symptoms. When these are present the appropriate DSM-IV diagnosis is mood disorder due to a general medical condition, and the initial intervention should target the medical disorder.

Major depressive disorder in institutional patients is also associated with increased mortality. Although there has been some controversy about whether patients are depressed because they are dying or dying because they are depressed, current evidence suggests that the decreased survival associated with severe major depression remains significant even after controlling for coexisting medical illnesses or disabilities.

Nursing home staff may have difficulties identifying depressive symptoms. Accordingly, there are ongoing needs for staff training in this area. The Geriatric Depression Scale has been validated for use as a screening instrument to identify residents with depressive symptoms. Although there has been controversy about its utility among patients with cognitive impairment, recent findings demonstrate that when administered verbally, it retains acceptable levels of sensitivity and specificity (with a standard cutoff of 10) in patients with Mini-Mental State Examination scores down to 15.

The finding from a double-blind, placebo-controlled study by the first author and coworkers that residential care patients (from a nursing home and congregate apartment setting) with symptoms of major depressive disorder responded specifically to the tricyclic antidepressant nortriptyline (Pamelor, Aventyl) suggests that standard approaches to diagnosis can be applied in the long-term care setting. However, evidence indicates that special problems arise in the treatment of nursing home residents. Low serum albumin concentrations and high levels of self-care disability predict a lack of response to nortriptyline, suggesting that a specific treatment-relevant subtype of major depressive disorder, perhaps related to failure to thrive, may emerge in long-term care patients. The high frequency of adverse effects of nortriptyline (approximately 30 percent) suggests the need for careful monitoring of long-term care patients during treatment and for research designed to optimize the safety and effectiveness of treatment.

Recent reports point to increasing the use of newer antidepressant drugs, such as the selective serotonin reuptake inhibitors (SSRIs) and agents with mixed noradrenergic and serotonergic effects, in nursing home residents. While these agents have adverse effect profiles that might be expected to be better tolerated by frail elderly nursing home patients, there is evidence that they can cause serious adverse events. P.B. Thapa and colleagues have demonstrated that use of SSRIs is associated with a nearly two-fold increase in the risk of falls in nursing home residents. Their study showed little difference in rates of falls between those treated with tricyclic drugs and those treated with SSRIs.

Studies comparing tricyclic drugs, serotonin reuptake inhibitors, and newer antidepressant drugs are in progress. Nevertheless, increasing evidence indicates that major depressive disorder in nursing home residents is a serious, common illness that is treatable. Antidepressants may be the one class of psychopharmacological drugs that is underused in the nursing home, and the Institute of Medicine suggests that one possible indicator of quality care may be the appropriate use of medications to treat residents with depression.

QUALITY OF LIFE

Nursing homes are communities consisting of residents, families, staff, and health care providers in which many patients with chronic disease and disability live out their lives. Accordingly, regulations developed to implement OBRA 1987 include the requirement that “a facility must care for its residents in a manner and in an environment that promotes maintenance or enhancement of each resident's quality of life.” Specific provisions require attention to issues of dignity; individuality; self-determination; participation in resident and family groups; participation in social, religious, and community activities; availability of ongoing activities programs designed in accordance with the resident's interests and well-being, and delivery of all services in a manner that reasonably accommodates the individual needs and preferences of the resident. The high prevalence of psychiatric disorders, primarily dementia and depressive disorders, implies that expertise regarding the psychiatric disorders of late-life is essential to meet quality-of-life and quality-of-care requirements.

A generation of research has established the general principle that both quality of life and health outcomes for nursing home residents are improved when the care environment is designed to foster the resident's sense of autonomy and control. M. Powell Lawton and his colleagues have emphasized that optimizing the resident's quality of life requires attention to the fit between individuals and their care environment. According to his ecological theory of aging and adaptation, the demands from the environment must match the resident's competencies and capabilities. Maximal performance can be elicited when the demands and challenges from the environment slightly exceed the level that would match the resident's current abilities; this may be most appropriate in a rehabilitation program. Alternatively, maximal comfort can be elicited when the environmental pressure is slightly below the resident's current abilities; this may be most appropriate during convalescence from acute illness or injury. According to Lawton, affective or behavioral pathology can occur when environmental demands exceed the zone of maximal performance and become overwhelming or when they fall below the zone of maximal comfort and represent a form of deprivation.

It is an ongoing challenge for nursing home staff to modify demands on the resident appropriately and to determine whether a change in activity level is a matter of changing physical ability or personal preference and when it is a symptom of psychopathology. Nursing homes each day must struggle with the way psychiatric disorders affect medical decision making. Development of institutional policies, procedures, and programs requires that psychiatric consultation be available to nursing home administrators and staff. Nursing homes are in fact neuropsychiatric institutions, and input from mental health professionals is necessary if the homes are to fulfill their missions.

SUGGESTED CROSS-REFERENCES

Alzheimer's disease and other dementing disorders are discussed in Section 51.3e, consultation-liaison psychiatry in Section 25.12, the introduction and overview of geriatric psychiatry in Section 51.1, and mood disorders in geriatric patients in Section 51.3d. Also, clinical assessment and approach to diagnosis in neuropsychiatry is discussed in Section 2.1, psychiatric disorders of late life in Section 51.3, psychiatric problems in the medically ill elderly in Section 51.3a, and psychiatry and medicine in Section 28.1. More-detailed discussions of dementing illnesses can be found in Chapter 10. Further discussion of depression can be found in Chapter 14.

SECTION REFERENCES

Advisory Panel and Alzheimer's Disease: *Alzheimer's Disease and Related Dementias: Acute and Long-Term Care Services*. NIH Publ no. 96-4136, 1996.

*Borson S, Doane K: The impact of OBRA-87 on psychotropic drug prescribing in skilled nursing facilities. Psychiatr Serv *48:*1289, 1997.

Cohen-Mansfield J, Marx MS, Rosenthal AS: A description of agitation in a nursing home. J Gerontol *44:*M77, 1989.

Evans LK, Strump NE: Tying down the elderly. A review of the literature on physical restraint. J Am Geriatr Soc *37:*65, 1989.

Fries BE, Mehr DR, Schneider D, Foley WJ, Burke R: Mental dysfunction and resource use in nursing homes. Med Care *31:*898, 1993.

Gerety MB, Chiodo LK, Kanten DN, Tuley MR, Cornell JE: Medical treatment preferences of nursing home residents: Relationship to function and concordance with surrogate decision-makers. J Am Geriatr Soc *41:*953, 1993.

*Goldfarb AI: Prevalence of psychiatric disorders in metropolitan old age and nursing homes. J Am Geriatr Soc *10:*77, 1962.

Health Care Financing Administration, Department of Health and Human Services: Medicare and Medicaid: Requirements for long term care facilities. Fed Regist *56:*553, 1991.

Institute of Medicine, Committee on Nursing Home Regulation: *Improving the Quality of Care in Nursing Homes*. National Academy Press, Washington, DC, 1986.

Katz IR, Beaston-Wimmer P, Parmelee PA, Friedman E, Lawton MP: Failure to thrive in the elderly: Exploration of the concept and delineation of psychiatric components. J Geriatr Psychiatry Neurol *6:*161, 1993.

*Katz IR, Parmelee PA: Depression in elderly patients in residential care settings. In *Diagnosis and Treatment of Depression in Late Life. Results of the NIH Consensus Development Conference*, LS Schneider, CF Reynolds, BD Lebowitz, AJ Friedhoff, editors. American Psychiatric Press, Washington, DC, 1994.

Katz IR, Simpson GM, Curlik SM, Parmelee PA, Muhly C: Pharmacologic treatment of major depression for elderly patients in residential care settings. J Clin Psychiatry *51:*41, 1990.

Lasser RA, Sunderland T: Newer psychotropic medication use in nursing home residents. J Am Geriatr Soc *46:*202, 1998.

Lawton MP, van Haitsma K, Klapper M: A balanced stimulation and retreat program for a special care dementia unit. Alzheimer's Dis Assoc Dis I>8(Suppl 1):S133, 1994.

Maslow K: Special care units for persons with dementia: Expected and observed effects on behavioral symptoms. Alzheimer's Dis Assoc Dis *8*(Suppl 3):S122, 1994.

McGivney SA, Mulvihill M, Taylor B: Validating the GDS Depression Screen in the nursing home. J Am Geriatr Soc *42:*490, 1994.

Morris JN, Hawes C, Fries BE, Phillips CD, Mor V, Katz S, Murphy K, Drugovich ML, Friedlob AS: Designing the national resident assessment instrument for nursing homes. Gerontologist *30:*293, 1990.

Parmelee PA, Katz IR, Lawton MP: Depression among institutionalized aged: Assessment and prevalence estimation. J Gerontol *44:*M22, 1989.

Ray WA, Federspiel CF, Schaffner W: A study of antipsychotic drug use in nursing homes: Epidemiologic evidence suggesting misuse. Am J Public Health *70:*485, 1980.

Ray WA, Taylor JA, Meador KG, Lichtenstein MJ, Griffith MR, Fought R, Adams ML, Blazer DG: Reducing antipsychotic drug use in nursing homes: A controlled trial of provider education. Arch Intern Med *153:*713, 1993.

Rodin J: Aging and health: Effects of the sense of control. Science *233:*1271, 1986.

Rovner BW, German PS, Broadhead J, Morriss RK, Brant LJ, Blaustein J, Folstein MF: The prevalence and management of dementia and other psychiatric disorders in nursing homes. Int Psychogeriatr *2:*13, 1990.

*Rovner BW, Katz IR: Psychiatric disorders in the nursing home: A selective review of studies related to clinical care. Int J Geriatr Psychiatry *8:*75, 1993.

Rovner BW, Steeke CD, Shmuely Y, Folstein MF: A randomized trial for dementia care in nursing homes. J Am Geriatr Soc *44:*7, 1996.

Sabin TD, Vitug AJ, Mark VH: Are nursing home diagnosis and treatment inadequate? JAMA *248:*321, 1982.

Shorr RI, Fought RL, Ray WA: Changes in antipsychotic drug use in nursing homes during implementation of the OBRA-87 regulations. JAMA *271:*358, 1994.

Siegler EL, Capezuti E, Maislin G, Baumgarten M, Evans L, Strumpf N: Effects of restraint reduction intervention and OBRA '87 regulations on psychoactive drug use in nursing homes. J Am Geriatr Soc *45:*791, 1997.

Smyer MA, Cohn MD, Brannon D: *Mental Health Consultation in Nursing Homes*. New York University Press, New York, 1988.

Stotsky BA: Allegedly nonpsychiatric patients in nursing homes. J Am Geriatr Soc *15:*535, 1967.

Strahan GW: An overview of nursing homes and their current residents. Data from the 1995 National Nursing Home Survey. Center for Disease Control and Prevention. National Center for Health Statistics. Advance Data no. 280, Jan 27, 1997.

Streim JE: Psychiatric problems in nursing homes. In *Treating the Elderly*, Sheikh JI, editor. Jossey-Bass, San Francisco, 1996.

*Streim JE, Oslin D, Katz I, Parmelee PA: Lessons from geriatric psychiatry in the long-term care setting. Psychiatric Quarterly *68:*281, 1997.

*Sullivan-Marx EM, Strumpf NE, Evans LK, Baumgarten M, Maislin G: Predictors of continued physical restraint use in nursing home residents following restraint reduction efforts. J Am Geriatr Soc *47:*342, 1999.

Thapa PB, Gideon P, Cost CW, Milam AD, Ray WA: Antidepressants and the risk of falls among nursing home residents. N Engl J Med *339:*875, 1998.

Volicer L, Collard A, Hurley A, Bishop C, Kem D, Karon S: Impact of special care unit for patients with advanced Alzheimer's disease on patients' discomfort and costs. J Am Geriatr Soc *42:*597, 1994.

51.6 SPECIAL AREAS OF INTEREST

51.6B FORENSIC ISSUES

BENNETT BLUM, M.D., AND SPENCER ETH, M.D.

CRIMINAL MATTERS

The growth in the geriatric population brings new challenges to psychiatry and issues that were once rare are now common. The elderly are often victims of crimes because, as a group, they are easy to overpower or manipulate and they fear retaliation. Seniors are usually willing to press charges when they are victims of violent crime, but generally do not pursue the prosecution of family members and caregivers. Shame regarding their own behavior, fear of losing their independence, physical fragility, and mental incapacity are common deterrents. Criminals find the elderly an attractive victim population for these reasons. Also, people over 65 currently control over $1 trillion in assets in the United States. Unfortunately, many law enforcement agencies, prosecutors, and judges still tend to be swayed by outdated "antiaging" biases. Psychiatrists may become unwitting accomplices in these crimes. They are uniquely positioned to help patients avoid these tragic and often life-threatening events.

Elder Abuse Definitions At least six commonly recognized forms of elder abuse exist: physical abuse, neglect, self-neglect, emotional abuse, financial abuse, and sexual abuse (Table 51.6b-1). Some authorities consider abandonment a separate category of abuse, but many think it is a specific form of neglect. *Physical abuse* is the use of physical force that is likely to produce pain, injury, or impairment. In addition to acts of violence (e.g., hitting, biting, kicking, pushing, burning), physical abuse also includes physical punishment, force-feeding, and the inappropriate use of restraints or drugs. *Neglect* is the failure or refusal to fulfill one's obligations or duties toward an elderly adult. It often refers to a failure to provide adequate care, food, or shelter. *Abandonment* is the desertion or willful forsaking of an elderly person by a person who has physical custody or has assumed responsibility for the elder's care. *Self-neglect* is behavior of elderly persons that threatens their health or safety. *Emotional abuse* is willful infliction of emotional pain or distress, including such behaviors as insults, threats, harassment, intimidation, humiliation, and enforced isolation. *Financial abuse* is the inappropriate or illegal use of an elderly person's money, property, or assets. *Sexual abuse* is nonconsensual sexual contact or activity with an elderly person.

1. Physical abuse
2. Neglect
3. Emotional abuse
4. Psychological abuse
5. Financial abuse
6. Sexual abuse
7. Abandonment (may be considered a subcategory of neglect)

Table 51.6b-1 Types of Elder Abuse

Incidence The true incidence of elder abuse is not known. Early studies suggested that only 1 of every 14 cases of abuse was reported to authorities; however, results of the most recent national study indicate that 1 of every 5 cases is reported. If this is correct, then approximately 500,000 people over age 60 were victims of elder abuse, neglect, or self-neglect in 1996. Because of methodological problems, the actual number of cases is probably much greater.

In substantiated reports of elder abuse provided to adult protective service agencies, approximately half of all cases involve neglect. Emotional abuse is found in more than one third of cases. Financial abuse occurs in slightly less than one third, and physical abuse occurs in one fourth of substantiated cases. Sexual abuse seems to occur infrequently, but inadequate research has been done in this area to allow definitive comments.

Victimology Any elderly person may become a victim; however, the risk increases with advancing age. People over 80 are much more likely to be abused than people in their 60s. The National Elder Abuse Incidence Study (NEAIS) found that more than half the victims of neglect and nearly half of those financially abused were over 80. Similarly more than 40 percent of victims of emotional and physical abuse were over 80. Generally, women are more likely to be abused than men. Approximately half of physical abuse or financial abuse victims had incomes between $5,000 and $9,999. Two thirds of neglect victims had incomes in the same range, as did 38 percent of emotional abuse victims. Most substantiated cases of elder abuse involved white victims. Less than 10 percent of substantiated cases in the NEAIS involved victims from African-American, Latino, Asian, Pacific Islander, American Indian, or Native Alaskan heritage. Approximately three fourths of the victims were described as "physically frail." Most victims experienced some confusion. Over 40 percent had either moderate or severe depression.

It is not unusual to discover that victims are dependent upon their abusers and feel they must protect the abuser. Many victims will not cooperate with medical personnel, investigators from adult protective services, or law enforcement.

Perpetrators Men are the perpetrators in most substantiated cases of elder abuse; however, women are perpetrators in 40 to 50 percent of cases. Perpetrators generally work alone, but cases involving multiple perpetrators are not rare. In the NEAIS, most perpetrators were younger than 59, but the findings suggest that more than one third were older than 60, and more than 25 percent were over the age of 70. Overall, more than three fourths were white, and approximately one fifth were black. Nearly half of the perpetrators were children of the victim, and one fifth were spouses. All the cases analyzed were substantiated incidents that had been reported to adult protective services or other authorities. It is believed that many cases are never reported.

Self-Neglect Current research indicates that most self-neglecting elderly persons are female, nearly half are 80 or older, and more than three fourths are white. Most experience confusion, but the majority are not depressed.

Reporting Elder Abuse Most states mandate elder abuse reporting; however, states vary widely regarding the recognition of "mandated" versus "voluntary" reporters, the procedures and timing for reporting, as well as the definition of *elder abuse*. For these reasons, practitioners are advised to learn the specific laws for their states.

Financial Abuse The term *financial abuse* encompasses both fiduciary abuse and fraud. A *fiduciary* is a person in a position of trust and responsibility, such as a physician, attorney, caregiver, or spouse. *Fiduciary abuse* is the misuse of this position of trust to exploit another. *Fraud* refers to financial exploitation that results from false promises. The perpetrator of a fraud need not have any prior relationship to the victim; this category includes all types of con artists.

One of the most common forms of fraud is fraudulent telemarketing. Although most telemarketing is legitimate, fraudulent telemarketing is pervasive. According to a 1995 study by the American Association of Retired Persons (AARP), fraudulent telemarketers steal approximately $40 billion per year—much of it from the elderly. The most important finding is that many victims are bright, sophisticated, and socially active. They tend to have strong family and social support, are characterized as

“strongly opinionated,” and are unwilling to believe they were defrauded. Other victim groups include those who are shy, isolated, naive, withdrawn, or easily manipulated. Appropriate psychosocial intervention and training may reduce their susceptibility.

Financial Abuse by Family Members Financial abuse often accompanies physical abuse. Researchers estimate that one case of financial abuse occurs for every two or three cases of physical abuse in the United States. Research from Canada and the United Kingdom indicates that family members alone are responsible for the financial exploitation of approximately 3 percent of the population over 65.

The “Civil Mirage” Although financial abuse is a crime, many law enforcement officers, prosecutors, and judges mistakenly believe it is a matter for civil courts. This attitude is changing as more is learned about the methods used by perpetrators. One tactic of offenders is to obtain control by having victims designate them on Durable Power of Attorney forms (Fig. 51.6b-1 and Fig. 51.6b-2). These forms are used to designate an alternate decision-maker for health care and finances. They authorize the perpetrators to make decisions on behalf of their victims if the elderly persons are deemed unable to make such decisions themselves. Unwitting physicians may be asked by perpetrators to provide a letter stating that an elderly person is impaired. The perpetrator uses these and other documents to control or steal the victim's assets. Elder abuse specialists refer to this misuse of legal documents as the “civil mirage.”

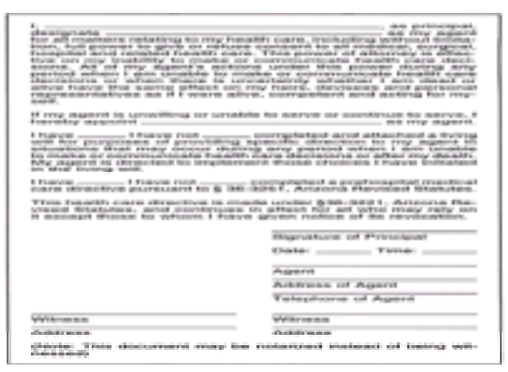

Witness

Witness

Address

Address

(Note: This document may be notarized instead of being witnessed)

FIGURE 51.6b-1 Health care power of attorney.

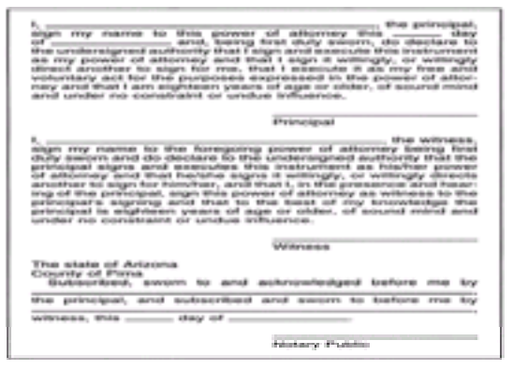

Principal

Witness

The state of Arizona
County of Pima

Subscribed, sworn to and acknowledged before me by _____, the principal, and subscribed and sworn to before me by _____, witness, this _____ day of _____.

Notary Public

FIGURE 51.6b-2 Power of attorney.

Undue Influence During elder abuse cases psychiatrists are commonly asked whether the victim was unduly influenced. Undue influence, a legal term, refers to a situation in which a person is pressured to act in a manner desired by another. Bowing to the undue influence results in behavior that deviates from that normally expected and may be contrary to the person's needs and interests. To determine whether undue influence occurred, the court must decide that a vulnerable person fell prey to coercion, intimidation, or some other form of manipulation, usually because of some mental distress or disability. The psychiatrist's role is to assist the court by providing a clear clinical description of the person's mental state, stressing particular symptoms that may heighten vulnerability to interpersonal pressure. Salient factors to consider include the nature of the relationship (i.e., fiduciary, romantic, or dependent); the degree to which the suspect person created an environment conducive to manipulation; and the presence of recent medical illnesses, traumatic events, and other psychosocial stressors. The presence of an independent social network may mitigate these factors.

Most general or geriatric psychiatrists are not trained to assess susceptibility to undue influence, nor are they skilled at recognizing the conditions necessary for undue influence to exist. Initial analysis should include consideration of the four essential characteristics of undue influence listed in Table 51.6b-2. The authors strongly suggest that detailed evaluations be performed only by experts in geriatric or forensic psychiatry who have extensive knowledge, skills, and experience in this area. A common but inaccurate belief is that assessment of undue influence is equivalent to the assessment of mental capacity. Although people with impaired mental capacity may be easier to manipulate, undue influence may occur in cognitively intact, and even in intellectually superior, people (Fig. 51.6b-3).

1. The victim must be dependent upon the perpetrator
2. The victim must be prevented from learning and using pertinent information before making a major financial decision
3. The victim's trust, fears, or prejudices must be manipulated by the perperator
4. The victim must acquiesce to the perpetrator's desires; in cases of financial abuse, the perpetrator obtains access to the victim's assets

Table 51.6b-2 The Characteristics of Undue Influence

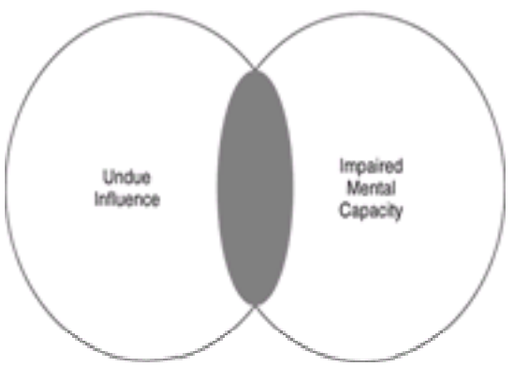

FIGURE 51.6b-3 Relation of undue influence and mental capacity. The overlapping area represents cases most likely to require forensic psychiatric evaluation.

A wealthy widower began dating an attractive woman several decades younger than himself. His children became concerned when he gave his new girlfriend his car and some valuable jewelry. He confided to them that he was ambivalent about the gifts, but she said she would lose interest in him if he did not continue to shower her with expensive presents. He reasoned the money was less important to him than having someone around so as to not feel lonely.

An elderly man became quite reliant upon a young female assistant who lived in his home and made all the arrangements for his medications, meals, and visitors. Eventually, she insisted he allow her to assume responsibility for his business and investment decisions or she would seek other employment. Fearing he would be left alone and unable to provide for his medical needs, he acquiesced to her wishes.

Undue Influence Regarding Wills As part of contesting a will, an attorney may ask a psychiatrist to perform a retrospective analysis of a person who has since died. Although such analyses are inexact, a psychiatrist properly trained and experienced in these evaluations may be able to help the court understand the decedent's mental capacity and susceptibility to undue influence at the time of executing the will. The psychiatrist should clearly distinguish facts known to be true from any assumptions, hypotheses, or opinions. The psychiatrist should also be prepared to confront a rigorous cross-examination.

Mental Capacity The law recognizes different types and degrees of mental capacity. For example, persons may not be able to create contracts or make donations, yet still be able to provide food, shelter, and clothing for themselves. In the context of making financial decisions, psychiatrists must carefully assess all requisite behaviors, besides noting general cognitive or neurological deficits. One model for assessing decision-making capacity, which has been used in state and federal courts, is presented in Table 51.6b-3. The benefit of this approach is the recognition and presentation of the many psychological functions used to make knowledgeable, reasoned, and rational decisions. This presentation helps judges and juries understand how specific impairments affect a patient's decisions, without attempting to negate the patient's ability to perform all tasks. A thorough psychiatric assessment is then used to explain any behavioral or cognitive problems.

To demonstrate mental capacity, persons must
Express desires
Learn new information
Concentrate and remember pertinent information
Identify pertinent parties and their responsibilities
Consider abstract concepts
Consider probable consequences for each major option
Understand the likely impact the decision will have on their life
Choose and maintain a decision
Not make decisions based upon delusions

Table 51.6b-3 Mental Capacity Model

Forensic assessment of undue influence and mental capacity may be aided by use of appropriate screening and evaluation tests described in the section on forensic tests.

Murder of the Elderly Elderly persons are usually considered to be the group least victimized by violent crime; however, most elder deaths are not carefully investigated. Little research has been done in this area, but findings from recent studies suggest that elderly persons have an increased risk of being murdered during a theft-related crime. Beatings are also more likely to occur when the victim is elderly. Because many elderly adults spend most of their time at home, they are more likely to encounter burglars targeting their dwelling. Many perpetrators perceive elderly persons as wealthy, physically vulnerable, isolated, and unable to defend themselves, and therefore a violent confrontation is more likely. Homicides related to interpersonal conflict are less frequent among elderly persons, but they do occur. Elderly spouses are less likely to commit murder than are adult children or acquaintances. The primary motives for elder murder are (1) greed and (2) anger or revenge. An increasing body of anecdotal reports suggests a link between fiduciary abuse and murder. The perpetrator unduly influences an elderly person and then neglects or poisons the victim after obtaining control of the elder's assets. The perpetrator may obtain medication to sedate or poison the victim by posing as a patient, or by posing as the concerned caregiver of a mild-to-moderately impaired elderly adult.

In one particularly heinous case, a boarding home landlady routinely sedated her elderly tenants (causing several deaths) to collect their Social Security checks. She obtained the drugs by complaining of psychiatric symptoms to several physicians. She was later found guilty of multiple counts of murder.

A young woman with a history of theft and drug-related offenses falsely identified herself as a nurse to a lonely, wealthy, elderly man. After becoming his live-in caregiver, she began withdrawing money from his bank accounts. When her actions came under scrutiny by bank officials, she married the man, then proceeded to withdraw several hundred thousand dollars from his accounts. He died 2 months after their wedding from malnutrition and neglect.

A caregiver was inappropriately sedating an elderly patient after the caregiver was granted control of the bank account. The victim did not die but was left penniless.

Elder as Criminal Although it is rarely suspected, elderly adults make up 1 percent of the nation's prison population. Reasons for this include lifelong recidivism, economic hardship, and mental illness. Elderly offenders have committed every known type of crime. Evaluation may be requested to assess an offender's sanity or capacity to waive *Miranda* rights, confess, plea bargain, stand trial, be sentenced, or be executed.

Elder as Witness Elderly witnesses may be questioned on their competence to testify. Assessment of testimonial capacity includes evaluation of the elder's perception, memory, understanding the obligation to testify truthfully, and susceptibility to suggestion.

Suicide People over the age of 65 account for approximately 20 percent of all suicides in the United States. The rate for people aged 80 to 85 is double the overall rate in the United States. The highest risk group are elderly white men who are divorced or widowed. Common life events and risk factors include social isolation, recent bereavement, physical illness, uncontrolled pain, fear of a prolonged death, major change in social status (e.g., retirement), alcohol or other substance abuse, depression, and availability of firearms. Homicide-suicides and assisted suicides are related phenomena.

Homicide-Suicides Homicide-suicides account for approximately 5 percent of all homicides nationally, but regional figures vary widely. Incidence rates range from 0.3 to 0.7 per 100,000 persons under the age of 55 and from 0.4 to 0.9 per 100,000 persons over the age of 55. The phenomenon occurs most often between spouses or people in a consortial relationship. Few studies exist in this area. Data from a recent study suggest that poor health, alcohol abuse, and depression are commonly found in elderly men who kill their wives or girlfriends. In a typical scenario, a depressed elderly husband uses a gun to kill his wife who also has serious health problems.

Mr. X. 96 years old, shot his wife of 72 years then killed himself. His wife was essentially confined to a wheelchair for 5 years and then developed a severe kidney infection. Several months before their deaths, one of Mr. X.'s boyhood friends died. Mr. X. became depressed. Usually known as a friendly and outgoing person, Mr. X. isolated himself in his home. A neighbor noticed that he cried frequently. Mr. X. has told the neighbor there was “no way he would allow himself or his wife to be put in a convalescent home. ‘Not as long as I've got a gun.' ”

Assisted Suicide of the Elderly A recent flurry of attention has been paid to an issue that has traditionally been shrouded in secrecy—physician-assisted suicide as

a health care option for the dying patient. For years some physicians have treated severe pain in terminally ill elderly patients with doses of analgesics that may depress respiration but are necessary for adequate pain control. For example, Sigmund Freud, suffering greatly from advanced oral cancer, requested and received an injection of morphine that probably hastened his death. Although the prescription of medication with the explicit goal of facilitating the patient's death is generally deemed unethical and illegal, public and professional opinions are changing on the issue. At the same time, concern exists that the elderly will become a population at risk for society-condoned murder. As a result, the nation's highest courts have been forced to consider whether informed consent can be extended to include a lethal choice.

When evaluating someone who wants physician-assisted suicide, it is important to remember many treatable medical and psychiatric conditions affect mood and decision-making ability. The public has often heard about the relation between inadequately treated pain and the desire to die—that is, once the pain is treated, suicidal desires abate. Malnutrition also affects mentation and is common among elderly patients. Even slight malnutrition may cause decreased concentration, confusion, and depression, leading to suicidal thoughts. Drugs are well-known precipitants of similar symptoms. Finally, psychiatric diseases, particularly major depressive disorder, impair a person's ability to think clearly or to consider options reasonably. Effective treatment is available for each of these conditions.

The influence of others cannot be overestimated. Vulnerable individuals have been pressured to accede to “suicide” through arguments relating to the expense of staying alive, the depletion of their families' resources, fear of pain or losing control, and the like. An ill person who is frightened, lonely, and perhaps confused can fall prey to authoritative manipulators who sound well meaning.

Crimes against the elderly are sometimes covered up by creating the appearance of a suicide. A 1995 survey performed in the Netherlands indicated that 0.7 percent of all deaths that year were caused by physicians acting without the consent of the decedent. Nearly one fourth of the physicians interviewed said they had ended a patient's life without the patient's explicit request. In the United States, research suggests that many elderly adults who survive so-called suicide attempts—or what could have appeared to have been natural deaths—say that they did not want to die. Many are drugged against their will. Perpetrators have been known to lie to physicians to obtain drugs that are then used to poison an older person.

CIVIL ISSUES

Informed Consent for Medical Treatment Agreeing or declining to provide informed consent regarding health care decisions depends on three elements in the decision-making process: information disclosure by the physician, voluntary participation by the patient, and patient competence.

Information Disclosure Information disclosure involves the communication, in readily understood language, of the nature of the medical problem, the proposed remedy and reasonable alternatives, and the attendant risks and benefits of all the treatment and no-treatment options. Given the amount of possible information any particular patient may wish to know, how should the caring physician explain the necessary and sufficient data? The goal of information disclosure is to enhance the patient's autonomy by encouraging a knowledgeable judgment concerning treatment options. It would seem prudent to offer the patient all pertinent facts.

Determining what constitutes important information depends on what would be relevant and significant for the patient in question or, if that cannot be ascertained, for the hypothetical reasonable patient. A recent California Supreme Court decision suggests that in accord with the standard of practice within the professional community, a physician can limit disclosure to the therapeutic focus and thereby omit information concerning the patient's nonmedical interests (e.g., business affairs).

The pressure of cost containment by the managed care industry has precluded physician members of some managed care organizations from discussing certain innovative or expensive research treatments with patients. This so-called gag rule threatens the informed consent process and consequently jeopardizes the fiduciary relationship between patient and physician.

There are patients whose diseases have progressed to the point at which standard treatments are not expected to confer meaningful benefit. When the chance of improvement is infinitesimal and the cost is exorbitant, is a physician justified in withholding a particular treatment option? Although this futility paradigm seems straightforward, there are inevitable uncertainties concerning the prognosis for any specific clinical intervention. Given this ambiguity, decisions are usually driven by an assessment of the likely burdens and benefits to the patient, the family, other patients, and society at large. For example, given the limited number of organs available for transplantation, should an elderly man who has chronic obstructive pulmonary disease and cirrhosis be offered a new liver? Who should decide? Such prioritizing of acceptable treatments produces a series of protocols or rules that creates a system of health care rationing. To date, American medicine has been reluctant to accept any rigorous form of rationing that eliminates the physician and patient as the responsible decision makers.

Voluntary Participation by the Patient Voluntary participation by the patient refers to the freedom to decide without being subject to undue influence. Voluntary participation is compromised when there is external coercion by family, friends, or even the health care team. The notion of voluntary participation may be expanded to embrace situations of internal influence caused by psychiatric symptoms such as delusion and depression. Undue influence is a legal term indicating that a person has been coerced, intimidated, or manipulated in some fashion to behave according to the desires of a perpetrator.

Competency According to Black's Law Dictionary, *competency* is a legal term that means to be “duly qualified: answering all requirements; having sufficient capacity, ability, or authority; possessing the requisite physical, mental, natural or legal qualifications.” Only a court can decide whether a given person is competent to perform a particular act.

Capacity signifies the “mental ability to understand the nature and effects of one's acts.” Although *capacity* and *competency* are often used interchangeably, *capacity* can also refer to the specific mental functions that are essential to informed decision-making. Psychiatrists may be asked to rely upon their knowledge, skill, and authority to assess a patient's mental capacities to help the court decide an issue of legal competence.

The question of how to address the issue of competency is difficult because there are different ways of organizing the idea. *General competence* implies that a person is functionally capable in all legally relevant domains. For example, in American law an adult is presumed to be generally competent unless a special court proceeding finds otherwise. This application differs from that of “specific competence,” in which it is assumed that particular skills are needed for certain tasks. Common examples of specific competence requiring unique abilities include the competence to marry, to drive a car, and to manage a large corporation. The notion of “specific competence” is fundamental to the assessment of an individual's ability to make informed decisions regarding health care.

Because of their training and practice, medical staff members tend to place a premium on the likely outcome of a patient's treatment decision. This consequentialist perspective can be called *consequence-dependence competency*. Especially in hospital settings patients who adamantly reject their doctors' orders are often deemed by the medical staff as being ipso facto incompetent, until the patient proves otherwise by agreeing to the recommended care. Although the medical team may have good intentions, this response undermines patient autonomy and changes the therapeutic alliance to an adversarial struggle to force the patient to accept what is defined as medically correct. A more legally and ethically appropriate approach to the assessment of competency of the noncompliant patient is to evaluate the patient's ability to comprehend and consider medical information, rather than exclusively focusing on the final decision. Only by emphasizing process over conclusion can consequence-independent competence be determined.

The emotional context of the patient's cognitive repertoire must be included in every assessment. Clearly, depression, anxiety, or mania can affect the patient's ability to listen, understand, and remember new information. In addition, extreme emotional states or lability may render a patient more vulnerable to undue influence.

If the court finds a person incompetent, the judge will appoint a surrogate decision maker. There are two categories of surrogates: guardians and conservators. A *guardian* has the legal authority to make decisions directly affecting the incompetent person's life, such as providing consent for medical treatment or placement into a nursing care facility. In contrast, a *conservator* has the legal responsibility to make decisions about the person's assets and property. Some jurisdictions have combined the two functions so that only one designation is used.

Components of Capacity Three common models of mental capacity exist: the philosophical/legal, medical, and functional models. Each has unique advantages and disadvantages.

PHILOSOPHICAL-LEGAL MODEL According to the philosophical-legal model, mental capacity results from being able to express desires, understand pertinent risks and benefits, appreciate the ramifications of a decision, and think rationally. These four elements are found in all state laws, though they may be combined or rephrased. This model's weakness is the lack of agreement about the definition, and thus the assessment, of each term.

MEDICAL MODEL In the medical model, medical symptoms are linked to incapacity. That is, to the extent that persons have neurological or psychiatric symptoms, they are considered unable to perform certain cognitive tasks and therefore have impaired mental capacity. The strength of this model is its ability to diagnose pathology causing cognitive impairment, recommend treatment, and make prognostic comments. The weaknesses of the model are the questionable association between medical symptoms and real-world functioning and its requirement of medical or neuropsychological expertise, which limits the model's use in routine matters.

One example of a medical model of capacity is found in California's Due Process in Competence Determinations Act. Passed in 1995, this act requires assessment of attention, orientation, memory, thought processes, executive and other cognitive functions, hallucinations, delusions, obsessions, and mood and anxiety disorders when evaluating any type of mental capacity. Although the creators of the legislation emphasize the functional aspects of this assessment, mental health expertise is clearly required. Most police officers and many social workers, public guardian investigators, adult protective service investigators, and attorneys lack this level of expertise, even those who specialize in elder abuse. Specialized training and subsequent long-term supervision is necessary to develop the requisite knowledge, skills, and experience to assess each component appropriately. An inadequate assessment may condemn an elderly person to further abuse; while an overzealous assessment will result in the loss of personal liberties.

FUNCTIONAL MODEL The functional model focuses upon observable behavior. Several mental abilities must be intact to make an informed, reasoned, and rational decision (Table 51.6b-3). Any impairment indicates the need for further investigation.

One of the most important, yet often unrecognized or overlooked, abilities is thinking strategically. To the extent this ability is impaired, a person cannot analyze the probable outcomes of decisions. This is a major component of executive functions, but impairment is often incorrectly attributed to trouble concentrating, mild memory loss, eccentricity, or personal rigidity.

Assessing Capacity The logical and most effective approach to assessing mental capacity follows the models described above. First, use the functional model (Table 51.6b-3) to document the elderly person's behavior, activities of daily living (ADLs), and instrumental activities of daily living (IADLs). ADLs include bathing, grooming, dressing, feeding, bathing, and ambulating; IADLs consist of more complex behaviors, such as shopping, preparing food, taking medications properly, using a telephone, and handling finances responsibly. This assessment is best performed by directly observing elderly persons in their home. If circumstances prevent a home visit, obtain this information from reliable, ancillary sources.

If functional impairments exist, follow the medical model and analyze the elderly person's overall health, medications, and psychiatric condition. A formal mental status examination is essential when evaluating cognitive abilities and should include assessment of attention; concentration; orientation; speech; language; memory; thought processes; emotional, control, and executive functions; insight; judgment; praxis, and constructional abilities. Practitioners may be confused about the technical meaning of the term executive functions. These mental abilities include anticipation, planning, selection, and monitoring of responses. They are critical to attention, memory, and learning and are related to insight and judgment.

Behavioral signs of executive dysfunction include impaired organization of new material, poor recall of recent and remote information, imitation and perseveration, difficulty using tools, and concrete and stimulus-bound thinking. Although not core symptoms of executive dysfunction, several other behaviors commonly coexist: tactlessness, impulsivity, social inappropriateness, mood lability, and irritability. Poverty of speech or profound nonchalance may also exist.

> A wealthy elderly woman had a number of small strokes. Afterward, she wrote multiple checks for the same item, impulsively bought large quantities of items she never used, and gave away her possessions to strangers.

After describing the person's functional impairments, and any psychiatric causes, the psychiatrist may now apply the information to the philosophical/legal model. How do the impairments or pathology affect the person's ability to express desires, understand pertinent risks and benefits, appreciate the ramifications of a decision, or think rationally? Using this approach, psychiatrists can best present clear and cogent reasoning to the court in support of their conclusions.

> An elderly widower wanted to marry a younger woman. His adult children opposed the marriage, afraid that their father might give his money and property to his new wife rather than to them. In the ensuing hearing his children presented an expert witness who testified that the man had dementia of the Alzheimer's type and therefore lacked the capacity to marry. The psychiatrist retained by the man concurred with the diagnosis but proceeded to describe his current level of functioning. The court found that the man had sufficient mental capacity and was therefore competent to marry regardless of his children's opposition.

Neuropsychological Tests Neuropsychological testing may be used to extend and quantify cognitive deficits detected on the mental status examination. Some frequently administered instruments include the Wisconsin Card Sorting Test, the Stroop Color Word Test, the Rey-Osterrieth Complex Figure Copying Test, and the Trails Making Tests of the Halstead-Reitan Neuropsychological Battery. However, these tests are limited by the considerable variability in performance and paucity of normative values across the geriatric age span, from the early 60s to the 90s. Also, only a moderate relation exists between test results and behavior in the real world. The psychiatrist must therefore draw upon as many different sources of data as possible in assessing any patient's functional capacities.

Mini-Mental State Examination The Folstein Mini-Mental State Examination (MMSE), a crude screening tool of cognitive function, has only limited correlation with real-world abilities or behavior. If an elderly person's capacity to make medical decisions; provide self-care; execute wills, trusts, or contracts; or engage in any other financial transaction is at issue, the MMSE should be used only as part of an overall psychiatric evaluation. Relying solely upon the MMSE to assess mental capacity is a misuse of the test. The MMSE has no value for assessing undue influence.

Forensic Tests Several tests have been designed to assess the probability that an elderly person was physically abused. Most have serious practical limitations and are not designed for forensic use, but are instead aids for geriatric practitioners.

Forensic tests are now available to assess the mental capacity to make financial decisions as well as undue influence. The Financial Abuse Screening Tool for Undue Influence (FAST-UI) and the Financial Abuse Screening Tool for Mental Capacity (FAST-MC) are useful for rapid clarification of situations; only the most current editions should be used.

For more in-depth analysis, the Mental Capacity Inventory and Undue Influence Questionnaire are detailed and reliable aids; however, multiple versions exist, and physicians should only use the latest editions of those designated for physicians and psychologists. These tests have been successfully used in state and federal courts.

Although many tests are being developed to assist the assessment of mental capacity to make medical decisions, the MacCAT-T is currently among the most reliable and easiest to use. This test was developed as a result of extensive analysis of laws throughout the country regarding competence to make treatment decisions. The test requires minimal preparation time and can be performed at the bedside.

LIVING WILLS AND DURABLE POWER OF ATTORNEY

The Federal Patient Self-Determination Act of 1990 established the right of patients to execute advance directives if they so choose. One type of advanced directive is the living will (Fig. 51.6b-4), a document prepared by a competent adult that indicates the conditions under which that person wishes to consent to cessation of specific medical treatments. Its purpose is to guide medical staff in delivering care in the event the person becomes terminally ill and incapable of independent decision-making (incompetent). These documents are generally limited in that they cannot provide precise guidance for the myriad of circumstances that might arise while treating a particular patient.

FIGURE 51.6b-4 Sample living will.

The other type of advanced directive is a proxy designation, such as a durable power of attorney that names a preferred surrogate decision-maker in the event the declarant becomes incompetent to make health care decisions (Fig. 51.6b-1). It may also be used to designate a surrogate to oversee financial and other matters (Fig. 51.6b-2). The surrogate can ensure that the preferences of the incompetent patient will be followed in whatever situation may arise, insofar as the surrogate is aware of, or can simulate, the patient's wishes.

A durable power of attorney may be designed to allow people with dementia of the Alzheimer's type to participate in research studies. This disease affects 2 to 5 percent of the population over 65 and 10 to 20 percent of people over 80 years old. Approximately 10 million Americans are expected to be affected by the year 2050. Given this prevalence and the associated personal, social, and economic costs, research into the causes, treatment, and prevention of the disease is necessary. Unfortunately, the disease destroys decision-making capacity; therefore obtaining informed consent for participation in trials is difficult. This problem may be reduced by appropriately appointing a surrogate decision-maker. Obviously, such a durable power of attorney must be assigned before the afflicted person becomes incompetent.

Insufficient attention has been paid to the question of the competence of persons who draw up advance directives. Unscrupulous people have been known to encourage vulnerable elderly persons to appoint them as surrogates in durable power-of-attorney documents, which despite the name may be executed without the involvement of an attorney. Once the document is signed and the person is declared incompetent, surrogates can freely deplete their victims' financial resources.

SUGGESTED CROSS-REFERENCES

Section 7.1 contains basic information on how to conduct a thorough mental status evaluation. Section 51.2c discusses the psychological, sociocultural, and physiological aspects of aging. Section 51.2a has important information on the psychiatric examination of an elderly patient. Section 51.6c and Section 51.6g describe ethical issues and elder abuse. See Section 29.1 for more information on suicide, and Section 28.7 for information on domestic violence.

SECTION REFERENCES

Angell M: The Supreme Court and physician-assisted suicide—the ultimate right. N Engl J Med *336:*50, 1997.

*Blum B: Assessment of mental capacity and undue influence—a guide for California attorneys. *Estate Planning for the Terminally Ill*. CEB Press, Sacramento, CA, 1997.

Christensen K, Haroun A, Schneiderman LJ, Jeste DV: Decision-making capacity for informed consent in the older population. Bull Am Acad Psychiatry Law *23:*353, 1995.

Cohen D, Llorente M, Eisendorfer C: Homicide-suicide in older persons. Am J Psychiatry March 1998.

Dukoff R, Sunderland T: Durable power of attorney and informed consent with Alzheimer's disease patients: A clinical study. Am J Psychiatry *154:*1070, 1997.

Falzon AL, Davis GG: A 15 year retrospective review of homicide in the elderly. J Forensic Sci *43:*371, 1998.

Fenigsen R: Dutch euthanasia revisited. Issues Law Med *13:*301, 1997.

Gostin LO: Deciding life and death in the courtroom: From Quinlan to Cruzan, Glucksberg, and Vacco—a brief history and analysis of constitutional protection of the ‘right to die.' JAMA *278:*1523, 1997.

Hankin M: A brief introduction to the due process in competence determinations act: A statement of legislative intent. Calif Trusts Estates Q *1:*36, 1995.

Hendin H: Physician-assisted suicide and euthanasia in the Netherlands. Lessons from the Dutch. JAMA *277:*1720, 1997.

Hirshfeld EB: Should ethical legal standards for physicians be changed to accommodate new models of rationing for health care? Univ Pa Law Rev *140:*1809, 1992.

High BH, Doole MM: Ethical and legal issues in conducting research involving elderly subjects. Behav Sci Law *13:*319, 1995.

Lachs MS, Williams CS, O'Brien S, Pillemer KA, Charlson ME: The mortality of elder mistreatment. JAMA *280:*428, 1998.

La Rue A: *Aging and Neuropsychological Assessment*. Plenum, New York, 1992.

Lyon GR, Drasnegor N, editors: *Attention, Memory, and Executive Function*. Paul H. Brookes, Baltimore, 1996.

McCormack P: Undernutrition in the elderly population living at home in the community: A review of the literature. J Adv Nurs *26:*856, 1997.

*National Center on Elder Abuse: *The National Elder Abuse Incidence Study; Final Report*. 1998.

Nelsen C, Huff-Corzine L: Strangers in the night: An application of the lifestyle-routine activities approach to elderly homicide victimization. Homicide Stud *2:*130, 1998.

Quill TE, Bernard L, Brock DW: Palliative options of last resort: A comparison of voluntary stopping eating and drinking, terminal sedation, physician-assisted suicide, and voluntary active euthanasia. JAMA *278:*2099, 1997.

Reis M, Nahmaish D: Validation of the indicators of abuse (IOA) screen. Gerontologist *38:*471, 1998.

Schneiderman LJ, Jecker NS: *Wrong Medicine: Doctors, Patients, and Futile Treatment*. Johns Hopkins University Press, Baltimore, 1995.

Scoevers RA, Asmus FP, Van Tilburg W: Physician-assisted suicide in psychiatry: Developments in the Netherlands. Psychiatr Serv *49:*1475, 1998.

Sprehe DJ, Loughridge A: Use of legal terms in will contests: Implications for psychiatrists. Bull Am Acad Psychiatry Law *24:*255, 1996.

*Tatara T: *An Analysis of State Laws Addressing Elder Abuse, Neglect, and Exploitation*. National Center on Elder Abuse, Washington, DC, 1995.

US Department of Justice, Bureau of Justice Statistics, press release, “The Nation's Elderly Experience Less Crime,” October 13, 1992.

US Department of Justice, Bureau of Justice Statistics, press release, “Elderly Crime Victims: National Crime Victimization Survey,” March, 1994.

Van der Maas PJ, van der Wal G, Haverkate I, de Graaff CL, Kester JG, Onwuteaka-Philipsen BD, van der Heide A, Bosma JM, Willems DL: Euthanasia, physician-assisted suicide, and other medical practices involving the end of life in the Netherlands, 1900–1995. N Engl J Med *335:*1699, 1996.

Weisman AM, Sharma KK: Forensic analysis and psycholegal implications of parricide and attempted parricide. J Forensic Sci *42:*1107, 1997.

*White BC: *Competence to Consent*. Georgetown University Press, Washington, DC, 1994.

Wood WP: *The Bone Garden*. Pocket Books, New York, 1994.

*Yarbrough J: *Predators of the Elderly*. Training videotape. Los Angeles Sheriff's Department, Homicide Bureau, January, 1997.

51.6 SPECIAL AREAS OF INTEREST

51.6C ETHICAL ISSUES

DEBORAH B. MARIN, M.D., AND CHRISTINE K. CASSEL, M.D.

The ethical issues that arise in the practice of medicine are particularly salient in geriatric psychiatry. Age alone places older adults at risk of discrimination regarding their medical care. Dementia, depression, and medical illness are common disorders in the older adult. These conditions and sometimes their treatments can impair patients' ability to communicate their wishes and to exercise their best judgment. Cognitive compromise may interfere with several domains of an individual's functioning, including the ability to choose appropriate medical treatment or make sound day-to-day decisions. Impaired function in any of these activities could have an impact on both the patients and their environment. As a consequence, significant ethical issues may arise during the course of treating these patients.

DEPRESSION AND END-OF-LIFE DECISIONS

The ethics literature pays little attention to depression among older patients. Depression can be successfully treated with medication or electroconvulsive therapy. Most geriatricians recommend serotonin reuptake inhibitors more widely than tricyclic drugs because they appear safer and have fewer adverse effects. Yet, depression is significantly underdiagnosed in community-dwelling and institutionalized individuals. Identifying treatable depression becomes exceedingly important for the psychiatrist who is asked to consider whether a patient is making a competent decision when deciding to refuse life-sustaining treatment. A patient's right to refuse care is a fundamental tenet of our legal system's respect for self-determination. However, we are considered justified in overriding a patient's refusal of life-sustaining treatment if the individual's judgment is clouded by suicidal ideation. In older adults, depression may be difficult to distinguish from sadness resulting from the awareness of life's limited possibilities. Some philosophers argue that people in advanced old age may have a perfectly rational assessment that their life journey has been completed and that death is an acceptable next step. Psychiatrists often argue that such thinking is invariably the result of a treatable depression. Many people now live past the age when normal patterns of developmental psychology were established. It may be necessary to reexamine the degree to which death acceptance is normal in advanced old age.

The question of diagnosing depression is nowhere more focused and important than in the euthanasia and assisted-suicide debate. This nation is engaged in a major social discourse about whether it is ever morally acceptable for physicians to help patients die, either by providing them with a prescription for a lethal dose of medication or more actively through an injection. In 1997 the United States Supreme Court rejected two federal circuit court decisions that argued that people had a constitutionally protected right to assisted suicide. The Supreme Court did not argue whether physician-assisted suicide was illegal or immoral but stated that this topic warrants further discussion. The Court also determined that individual states should be allowed their own legislative and statutory conclusions.

In 1994 a statewide referendum in Oregon approved legalization of a strictly defined measure for physician-assisted suicide. It requires that the patient is terminally ill and makes the request multiple times. It must also be established that multiple opinions concur with the diagnosis and that every attempt has been made to reverse treatable causes of suffering. Specifically, these treatments include appropriate relief of pain and psychological distress. Inherent in the statute is a psychiatric review to determine the presence of a treatable depression. Psychiatrists in Oregon and other states have begun to discuss developing standards that would be meaningful in such a situation. A terminally ill individual may have many of the signs and symptoms associated with depression. Such patients may also wish to die to relieve personal suffering or to ease the burden placed on others. These depressive symptoms may warrant a course of pharmacotherapy, yet pharmacological treatments are not always effective. Questions often arise regarding the appropriate role for psychiatrists in the evaluation and treatment of patients seeking to end their lives. No psychiatric guidelines for such decisions exist, but they remain an important challenge in the years ahead.

DEMENTIA

The prevalence of dementia, in particular Alzheimer's disease and vascular dementia, increases with age. These illnesses pose ethical challenges regarding patient autonomy, capacity, and confidentiality.

Autonomy A physician's role is to choose treatments that enhance a patient's autonomy, or self-determination. Physicians must weigh their concerns for patient autonomy with the desire to do what is best for the individual, a concept called paternalism. The medical profession in the United States espouses the preservation of patient self-determination. Yet, paternalism may be justified if a patient is too ill to make an informed treatment choice or if a therapeutic intervention has benefits that significantly outweigh the loss of autonomy. As a dementia progresses, a patient's capacity to make autonomous decisions decreases, and ethical dilemmas arise from the autonomy-paternalism conflict.

Capacity Psychiatrists are often requested to assess whether a patient is capable of making treatment decisions. Capacity determinations consider the magnitude and type of the decision a patient must make. The specificity of capacity decisions is pertinent to the diagnosis of dementia. Individuals with mild-to-moderate dementia often can direct medical decisions. This domain of decision making may be maintained even if another is impaired. For example, Alzheimer's patients may be capable of consenting to antidepressant treatment, yet may not be able to handle their finances. In contrast, individuals in the severe stage of the illness likely cannot make any treatment decisions.

Confidentiality A basic tenet of the physician-patient relationship is the confidentiality of the information obtained by the clinician. In psychiatry, confidentiality usually pertains to the release of information to other professionals or family members. Cognitively intact patients can request that the physician withhold the release of information, as long as they have not expressed the desire to hurt a specific person. In contrast, demented individuals may not be able to make informed decisions regarding information transfer between their family caregivers and the treating physician.

Mild Stage

Mr. C., a 75-year-old married man with one 36-year-old daughter, came for a psychiatric evaluation that was initiated by his wife. Though reluctant, he agreed, yielding to his family's concerns about his difficulty driving and handling the family finances. During the interview, Mr. C. became increasingly upset by his wife's complaints about his cognitive deficits. The physician requested Mr. C.'s permission to speak with his wife privately for the purpose of history gathering. Although Mr. C. had been an accountant, he could no longer balance his checkbook or remember to pay the monthly bills on time. He frequently asked the same questions and consistently forgot recent events. He used to have an excellent sense of direction, but now his wife had to prompt him to make correct turns when driving to familiar places. He had mild word-finding difficulties but could still clearly convey his ideas.

When interviewed, Mr. C. acknowledged his memory problem. He did not believe it to be serious, although he knew it caused his family great concern. He adamantly denied any difficulty driving. Upon mental status testing, Mr. C. displayed impaired recent memory and mild word-finding difficulty. He stated the wrong day and date but knew the month and location of the interview. He could subtract serial threes but could not perform more complex calculations. At the conclusion of the first visit, the physician explained the need to perform some further tests, after which she would review the findings and provide treatment recommendations at the next visit.

The evaluation revealed a diagnosis of Alzheimer's disease. Prior to the second clinic visit, Mrs. C. called to request that Mr. C. not be informed of the diagnosis if it were a dementia, especially dementia of the Alzheimer's type. She worried that he might have a catastrophic reaction since his sister had died 5 years previously in a chronic care facility in the severe stage of the illness. The physician explained that it was Mr. C.'s privilege to know his diagnosis but did not divulge it over the phone. With appropriate information, the physician explained, Mr. C. could then make the most informed decisions regarding his care. At the second visit, the physician reviewed the findings and treatment options with the family. The daughter asked whether her father or she should be genetically tested.

Comment Alzheimer's disease alters the typical physician–adult patient relationship in which the patient initiates the first contact, serves as the primary informant, and determines whether or not to disclose information to others. A patient's autonomy is often disrupted even before the initial physician contact; family caregivers often initiate the treatment because of their concern about the patient's behaviors. Patient confidentiality is also compromised when a family member serves as an informant or receiver of clinical information. It can be argued that the distress patients experience on learning of the severity of their symptoms outweighs the benefit of being present for this portion of the interview. A physician can choose a paternalistic approach to protect the patient from undue distress and request to speak with the family alone to obtain a history. Most, but not all, studies suggest that most relatives of Alzheimer's disease patients prefer that the afflicted family member not be told the diagnosis, yet they themselves would prefer to be told if they were to develop the illness. However, the data suggest that most Alzheimer's patients cope well with the diagnosis.

A dementia patient's autonomy and confidentiality can be conceptualized as somewhat akin to what is encountered in pediatrics, yet one is dealing with an adult. The patient's well-being and desires are the center of the treatment plan, but the family is likely to be included in communication with the physician and in treatment decisions. One of the most difficult decisions pertains to driving. This activity provides mobility and autonomy to many older adults, particularly if they live in an area that does not provide adequate public transportation. Yet, even mildly impaired dementia patients are at increased risk of vehicular accidents. Most states do not require that a physician report the dementia diagnosis to the motor vehicle bureau. Nonetheless, the physician must discuss the potential dangers of driving with the patient and family.

Dementia patients often cannot give a complete history or understand the severity of their impairment. Also, the patient may not be able to comply with a treatment plan without the family's involvement. Yet, mild- to moderate-stage patients can often participate in treatment planning. A dementia patient's strengths defy the common misconception that a cognitively impaired individual has no capacity. Physicians frequently use the clinical history and cognitive testing to assess a patient's capacity. There are scales to assess capacity, yet they are not widely used. When a patient has capacity, physicians should discuss advance directives. Although patients often do not initiate these discussions with their physicians, they are very receptive to this topic. In respect of patient autonomy, the physician is obligated to engage in these discussions to protect patients in the event that they cannot communicate their wishes.

Genetic testing of Alzheimer's patients also raises ethical concerns. The question arises whether genetic screening should be available to patients and family members who are at increased risk for developing the illness. Since genetic testing has no therapeutic value at this time, it is not recommended as part of standard clinical care for Alzheimer's patients. However, as treatments become available for populations at risk, an ethical issue regarding access to these screening programs is likely to arise.

Moderate Stage

During the 3 years following Mr. C.'s diagnosis, he required a home attendant to help with activities of daily living. Mrs. C. also arranged for all his medical care and the family's financial planning. At times he accused her of being an imposter (Capgras's syndrome) who had murdered his wife; he occasionally was so upset about this that he hit her. Mrs. C. was emotionally spent and financially drained by her husband's behavior and care needs.

Mental status testing indicated that Mr. C. was oriented to self only. He denied having any memory problems or behavioral disturbances. When Mrs. C. raised the possibility of chronic-care-facility placement, Mr. C. screamed that he “did not belong in a home.” His wife and physician both knew that he had been against placement from the beginning of his treatment. The physician discussed the possibility of psychotropic medication for Mr. C.'s behavioral symptoms. She also recommended that Mr. C. enter a day program so that Mrs. C. could get more respite.

Comment In the moderate stage of Alzheimer's disease the patient's reliance on others significantly increases. Day-to-day decisions are commonly made by others, thereby further disrupting an individual's autonomy. Physicians are often faced with the ethical dilemma of prescribing psychotropic medications to cognitively impaired patients who have limited capacity to give informed consent. Optimally, the patient would consent to take medication. A physician's recommendation to treat should consider both the patient and their environment. If pharmacotherapy maintains patients in their homes for a longer period of time, the paternalistic decision to treat actually respects the patient's autonomy.

Nursing home placement poses ethical questions beyond the issue of behavioral management. Generally, cognitively impaired individuals are more comfortable in familiar environments. A new setting, such as a day care center or nursing home, can be disruptive and anxiety provoking; however, nursing home placement may be preferable in some circumstances. Although physicians must assume the role of patient advocate, they must also make every effort to ensure that the family is not overburdened with the caregiving and that the patient is safe. If a patient requires more care than can be given at home, the physician may have to take a decisive stance and recommend placement.

Severe Stage

One year later, Mr. C. was placed into a long-term-care facility, where he sustained a hip fracture. He deteriorated significantly and never regained the ability to ambulate. He required spoon feeding to maintain his weight and hydration. His arms were restrained to prevent him from removing an intravenous line. He could only communicate by demonstrating bodily pain when being moved. He then developed a pneumonia that could necessitate a ventilator. Mrs. C. alluded to her husband's health care proxy and stated that he wanted “no heroic measures” if he became terminally ill.

Comment Over the past decade, there has been a significant tendency to avoid physical or chemical restraints in the medical setting. Restraint use is counter to patient autonomy and of limited therapeutic value. Therefore, clinical use of restraints must be clearly justified.

If a patient's health care preferences are documented, those wishes should be followed. A written advance directive (e.g., living will, health care proxy, durable power of attorney) can provide valuable information. Living wills, unfortunately, are often very general. The treating physician may not know the intent, or even the existence, of the living will. This is one of the major shortcomings of written advance directives. A surrogate or proxy may be available when decisions need to be made (as was the case with Mr. C.) and can respond to the specificity of the information and the situation at hand. However, often such proxies have never discussed the patients' values with them. Thus, physicians should encourage advance directive discussions with family members who may subsequently have to make proxy decisions. The widespread discussions about improving care of the dying in the United States and the controversy about assisted suicide may encourage many families to have these discussions in advance of severe disabilities and incapacity.

Once it is established that the patient does not wish to have life-prolonging measures when quality of life is no longer acceptable by standards that are

when decisions need to be made (as was the case with Mr. C.) and can respond to the specificity of the information and the situation at hand. However, often such proxies have never discussed the patients' values with them. Thus, physicians should encourage advance directive discussions with family members who may subsequently have to make proxy decisions. The widespread discussions about improving care of the dying in the United States and the controversy about assisted suicide may encourage many families to have these discussions in advance of severe disabilities and incapacity.

Once it is established that the patient does not wish to have life-prolonging measures when quality of life is no longer acceptable by standards that are personal and consistent with prior expressed values, then it is acceptable to follow that wish. Indeed, one could argue that a broad interpretation of life-sustaining measures be used, including not only traditional high-tech interventions, such as ventilator support, but also feeding tubes, artificial nutrition and hydration, and antibiotics for infections. If a physician authorizes withholding or foregoing treatment of an infectious illness such as pneumonia, every measure must still be taken to ensure patient comfort. These interventions include proper pain control, management of secretions, and provision of a comforting milieu. The principles of appropriate palliative care include human contact, medication if necessary to treat agitation and anxiety, and availability of the physician for the patient and family.

End-of-life decisions often require extensive discussions with the family. Proxies are often asked to make a health care decision of enormous import at a time when they are actively grieving the loss of their family member. Family members may experience extensive guilt when making these decisions. Many ethicists believe it is appropriate, when life-extending treatment really cannot add any more promise of a meaningful quality of life, for physicians to accept some decision on their own shoulders and inform the family. Rather than asking, “Do you want us to keep your mother alive?” they should say, “It is unlikely that your mother will survive this last illness. Our plan is to make her as comfortable as possible. What can you tell us about what she might have wanted?” This puts family members in a much more open position to share that information rather than feeling that their decision will actually cause the parent's death, which is, of course, not true and an unnecessary burden on the family.

SUGGESTED CROSS-REFERENCES

Medical-legal issues in geriatric psychiatry are discussed in Section 51.6b, elder abuse and neglect in Section 51.6g, and financial issues in Section 51.5a. Alzheimer's disease and other dementing disorders are discussed in Sections 51.3e and Chapter 10. Mood disorders are discussed in Chapter 14 and suicide in Section 29.1. Death, dying, and bereavement are discussed in Section 28.5.

SECTION REFERENCES

*Appelbaum PS, Grisso T: Assessing patients' capacities to consent to treatment. N Engl J Med *319:*1635, 1988.

Bahro M, Silber E, Sunderland T: How do patients with Alzheimer's disease cope with their illness. A clinical experience report. J Am Geriatr Soc *43:*41, 1995.

*Barns RC: Telling the diagnosis to patients with Alzheimer's disease. Br Med J *314:*375, 1997.

Block SD, Billings JA: Patient requests for euthanasia and assisted suicide in terminal illness. The role of the psychiatrist. Psychosomatics *36:*445, 1995.

*Brechling BG, Schneider CA: Preserving autonomy in early stage dementia. J Gerontol Soc Work *20:*17, 1993.

Carney MT, Morrison RS: Advance directives: When, why and how to start talking. Geriatrics *52:*65, 1997.

Cassel CK: Ethical dilemmas in dementia. Semin Neurol *4:*92, 1984.

Drickamer MA, Lachs MS: Should patients with Alzheimer's disease be told their diagnosis? N Engl J Med *326:*947, 1992.

Dukoff R, Sunderland T: Durable power of attorney and informed consent in Alzheimer's disease: A clinical study. Am J Psychiatry *154:*1070, 1997.

*Field M, Cassel CK: *Approaching Death: Improving Care at the End of Life*. National Academy Press, Washington, DC, 1997.

*Ganzini L, Lee MA: Psychiatry and assisted suicide in the United States. N Engl J Med *336:*1824, 1997.

Grisso T: *Evaluating Competencies: Forensic Assessments and Instruments*. Plenum, New York, 1986.

Hammes BJ, Rooney BL: Death and end-of-life planning in one midwestern community. Arch Intern Med *158:*383, 1998.

Haycox JA: Social management. In *Alzheimer's Disease: The Standard Reference*, B Reisberg, editor. Free Press, New York, 1983.

Hebert LE, Hennekens CH, Taylor JO: Prevalence of Alzheimer's disease in a community of older persons: Higher than previously reported. JAMA *262:*2551, 1989.

Ingelfinger FJ: Arrogance. N Engl J Med *303:*1507, 1980.

Janofsky JS, McCarthy RJ, Folstein MF: The Hopkins Competency Assessment Test: A brief method for evaluating patients' capacity to give informed consent. Hosp Community Psychiatry *43:*132, 1992.

Landry FJ, Kroenke K, Lucas C, Reeder J: Increasing the use of advance directives in medical outpatients. J Gen Intern Med *12:*412, 1997.

Maguire CP, Kirby M, Coen R, Coakley D, Lawlor BA, O'Neill D: Family members' attitudes toward telling the patient with Alzheimer's disease their diagnosis. Br Med J *313:*530, 1996.

Marson DC, Schmitt FA, Ingram KK, Harrell LE: Determining the competency of Alzheimer's patients to consent to treatment and research. Alzheimer Dis Assoc Disord *8:*5, 1994.

Meier DE, Cassel CK: Nursing home placement and the demented patient. A case presentation and ethical analysis. Ann Intern Med *104:*98, 1986.

Meyers BS: Telling patients they have Alzheimer's disease. Br Med J *314:*321, 1997.

Mezey M, Mitty E, Rappaport M, Ramsey G: Implementation of the Patient Self Determination Act (PSDA) in nursing homes in New York City. J Am Geriatr Soc *45:*43, 1998.

Morrison RS, Zayas LH, Mulvihill M, Baskin SA, Meier DE: Barriers to completion of health care proxies: An examination of ethnic differences. Arch Intern Med *158:*2493, 1998.

Post SG, Whitehouse PJ, Binstock RH, Bird TD, Eckert SK, Farrer LA, Fleck LM, Gaines AD, Juengst ET, Karlinsky H, Miles S, Murray TH, Quaid KA, Relkin NR, Roses AD, St. George-Hyslop PH, Sachs GA, Steinbock, B, Troschke E, Zinn AB: The clinical introduction of genetic testing in Alzheimer's disease. An ethical perspective. JAMA *227:*832, 1997.

Quill TE, Cassel CK: Nonabandonment: A central obligation for physicians. Ann Int Med *122:*368, 1995.

51.6 SPECIAL AREAS OF INTEREST

51.6D SOCIOCULTURAL ISSUES

HUGH C. HENDRIE, M.B., CH.B.

Although interest in the influence of sociocultural factors on the etiology and management of psychiatric disorders has a long history, formal studies in this area are more recent. Cross-cultural research has not been without controversy including whether or not Western concepts of psychopathology can be imposed on non-Western populations. One problem lies within the many different definitions of culture, which range from the broadly conceived "way of life" to the more narrowly conceived "unconscious canons of choice." Certainly, if a broad definition is used to include such lifestyle issues as diet, exercise, and exposure to toxins, cross-cultural research has made large contributions to the understanding of risk factors in nonpsychiatric diseases such as heart disease, stroke, colon cancer, and Parkinson's disease. The distinction between the effects of culture or the effects of social and economic deprivation, particularly in minority populations, is another contentious area. This issue was exemplified recently from the results of the North Manhattan Aging Project where apparent differing rates of dementia between Latino, African-American, and non-Latino white populations were accounted for not by sociocultural membership but by different levels of education across groups. Variations have been reported between minority and nonminority groups on scores on cognitive tests such as the Mini Mental State Examination (MMSE), but again these differences are accounted for to a large extent by lack of educational opportunities in the minority group rather than cultural or genetic differences.

Rapidly increasing ethnic minority populations in the United States have led to a growing interest in and concern about the influence of sociocultural issues on the mental health and well-being of the elderly. Variables such as different languages, values, and beliefs and prejudice and bias between minority groups and health care providers can adversely affect access to care, accurate diagnosis, treatment, and management compliance. Fortunately, these issues are becoming better recognized and educational recommendations for health care providers are included in task force reports from the American Psychiatric Association and the American Society on Aging (Table 51.6d-1). Unfortunately, minority groups still remain underrepresented in most research studies in the United States. Research on sociocultural factors in geriatric psychiatry is still in its infancy. However, a number of interesting and provocative approaches are under way.

Considerations for Developing a Model Program
1. Understand the cultural traditions, historical experiences, and the social and political networks within the community.
2. Identify and involve community leaders, organizations, grantmakers, and advocates in planning and organizing efforts.
3. The planning group and event participants must decide direction and priorities.
4. Define goals based on a realistic timeline.
5. Work to develop community leaders by providing education about aging services and program systems and the skills necessary to effect change.
6. Design interactive sessions that promote discussion and hands-on activities.
7. Be available to provide technical assistance.
8. Conduct monthly follow-up with participants.
9. Form relationships between older participants and community professional leadership.
10. Emphasize issues that directly affect elders' lives.

Adapted from the American Society on Aging: Serving elders of color: Challenges to providers and the aging network. San Francisco, March 1992.

Table 51.6d-1 American Society on Aging: Programs for Minorities

DEVELOPMENT OF CULTURE-FAIR INSTRUMENTS

There are major methodological problems in conducting cross-cultural studies. These include consideration of the age structure of the population, issues of sampling in very different types of communities, some of which lack accurate census data, and the establishment of inter-rater reliability of diagnoses by researchers across sites. A major challenge is the development of culture-free or culture-fair screening and clinical instruments. In addition to translation, a World Health Organization (WHO) study group emphasized the importance of harmonization in this process. *Harmonization* means that the instrument must be harmonious or consistent with the cultural linguistic and educational norms of the subject population. The field of psychiatry has had considerable success in instrument development, for example, the creation of the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) and WHO-sponsored international epidemiological interview, sponsored by the Composite International Diagnostic Interview (CIDI), an expansion of the Diagnostic Interview Schedule (DIS), which has now been tested in worldwide field trials. In geriatric psychiatry, the most attention has been paid to develop instruments to identify dementia, a particularly difficult task because diagnosis is based upon elucidating cognitive decline and social impairment. Reliable screening instruments based upon cognitive performance are now available and have been used successfully in cross-cultural studies, including the Hindi Mental State Exam (HMSE) and the Cognitive Abilities Screening Instrument (CASI). A screening instrument has been developed that includes both cognitive testing and informant data about performance in everyday living, the Community Screening Instrument for Dementia (CSI "D"). This instrument has now been used with good sensitivity and specificity in the Cree, African-American, Chinese, and Yoruba populations. The complicated steps necessary to develop a culture-fair instrument are presented in a simplified outline in Table 51.6d-2.

Criteria		Age 65 Years and Over (%)	95% Confidence Limits
Canadians (English-speaking)—1994	DSM-III-R/ICD-10	8.0	7.5–8.5
African-Americans—1995	ICD-10	8.24	7.09–9.40
Japanese-Americans (Honolulu) males only—1996	DSM-III-R	7.6	5.1–10.8
Japanese-Americans (Seattle)—1996	DSM-III-R	6.3	5.9–6.8

Table 51.6d-2 Age-Standarized Prevalence Rates for Dementia in Ethnic Groups in North America

LIFE SATISFACTION

The now-voluminous literature on life satisfaction, while having some contradictory findings, has one constant feature: elderly subjects generally report higher levels of life satisfaction than younger subjects. As life satisfaction is related to a number of socioeconomic as well as health indicators, it might be anticipated that the well-being of the elderly would be different in differing cultures. A unique, 13-nation international survey on human values and well-being involving both Western and non-Western countries reported that the elderly in all countries indicate more contentment and satisfaction with human relations, material needs, and religious

experience than did younger subjects. The reported level of life satisfaction by the elderly in all countries was remarkably similar regardless of the different economic status of the nations involved in the study. The only major differences between countries was that more non-Western than Western cultures reported homogeneity of satisfaction responses across age groups. They offered two possible explanations for this latter finding. First, Asian societies as opposed to occidental societies have relatively preserved extended-family systems and a greater commonality of religious beliefs and philosophies. Second, presumably because of perceived better opportunities, subjects in developed societies have higher aspirations, which may or may not be fulfilled whereas in less developed countries, such achievement motivation is less likely to occur and therefore satisfaction with the status quo may be more likely. The significance of extended-family relationships and religious involvement in tempering the effects of relatively low income and status on life satisfaction in minority elderly subjects in the United States has also been reported.

LATE-LIFE DEPRESSION

The value of conducting cross-national epidemiological studies in depression has recently been reinforced by the publication of the results of a major population-based international study involving 10 different countries and 38,000 community subjects aged 18 years and older using the DIS. Rates of major disorders between countries varied fifteenfold from 1.4 percent lifetime prevalence of depression in Taiwan to 19 percent lifetime prevalence of depression in Beirut; the latter figure is not surprising in view of the conditions in Beirut at the time of the study. These differences in rates suggest that cultural differences or variations in risk-factor profile could influence disease expression in depression. Unfortunately, the number of elderly subjects included in the study in many countries precluded a direct comparison of rates of late-life depression and few other such studies exist in the literature. However, there is reason to believe that differences in rates of depression in the elderly also exist between countries and that this area would be a fruitful one to explore. In a population-based study of major depressive disorder using the Geriatric Mental Schedule–Automated Geriatric Examination Computer Assisted Taxonomy (GMS-AGECAT) in Zaragoza, Spain, 1-year prevalence rates of 4.8 percent were reported. This contrasted with 3 percent prevalence rates of major depressive disorder reported from Liverpool, England, also using the GMS-AGECAT.

Rates of depression in elderly minority populations in the United States remain controversial. The Epidemiological Catchment Area (ECA) study suggested that 1-year prevalence rates for African-Americans (2.2 percent) were lower than for the white (2.8 percent) population and that Latinos had somewhat higher rates (3.3 percent). However, the ECA study had few elderly minority subjects on whom to base prevalence rates. In a larger population-based study in north Virginia, 1-year prevalence rates of 3.1 percent in elderly African-Americans (65 and older) comparable to the ECA rates in whites were found. Previous reports of varying frequencies of depressive symptoms in ethnic minority elderly populations may be due to differential responses of minority subjects to questionnaires such as the Center for Epidemiological Studies–Depression (CES–D) scale.

DEMENTING DISORDERS

Perhaps the greatest interest in cross-cultural studies in elucidating environmental and cultural risk factors in geriatric psychiatry is now in the area of the dementias and in dementia of the Alzheimer's type in particular. Rather paradoxically, this is at least partly the result of the findings from the recent genetic studies including the discovery of the powerful association between the possession of the e4 allele of apolipoprotein-e (APOE) and late-onset dementia of the Alzheimer's type. Epidemiologists are now able to combine the techniques of population genetics with risk factor research to explore the interaction between genetics and environment in the etiology of the disease. This ability to identify risk factors at the genetic level has other significant implications for epidemiological research. It may no longer be necessary to consider ambiguous and politically charged concepts such as race and ethnicity. A 1998 editorial in *Science* was entitled “DNA Studies Challenge the Meaning of Race.” Sophisticated etiological hypothesis involving specific genes and specific environmental factors can now be formulated to explain geographic or ethnic variations in disease rates.

In the National Advisory Council on Aging 1995 report, a major objective for research into this form of dementia was to “search for selective risk factors through cross-cultural epidemiological studies as a prelude for developing prevention strategies.” If populations with significantly differing rates of this dementia could be identified, risk factor research would be greatly expedited. A particularly fruitful area of research is studying migrant populations. A powerful argument for the presence of social and environmental risk factors could be made if varying rates of dementia are found in populations from the same ethnic group at different levels of development and from different environments. Conversely, different ethnic groups living in the same environment should have similar rates of illness. A few studies are now in progress looking at populations originating from Japan, India, and Africa. The results to date are intriguing but far from definitive. Recently, reported prevalence rates for dementia for Japanese living in Hawaii and in Seattle and African-Americans living in Indianapolis are very similar and remarkably close to the rates reported in a large Canadian study of English-speaking residents (Table 51.6d-3).

1. Identify cognitive and behavioral dimensions to be measured based on diagnostic criteria.
2. Select items for translation with indigenous investigators to make first draft.
3. Pretest first drafts for acceptability with small sample of target subject ($N = 10$).
4. Review and pretest revise for second drafts.
5. Pilot test on larger samples ($N = 50–100$).
6. Statistical analysis of pilot study and revise for final instruments. Each draft undergoes translation, independent back translation, and consensus harmonization before testing.

Table 51.6d-3 Steps to Construct a Culturally Harmonious Test

In a five-county, urban-rural area of North Carolina one study found no difference between African-Americans and whites in either prevalence or incidence rates for Alzheimer's disease. However, in a prospective, longitudinal study of subjects living in north Manhattan, the cumulative risk for Alzheimer's disease was four times higher for African-Americans and two times higher in Hispanics than in the white population for subjects who did not possess the e4 allele. These discrepant findings of Alzheimer's disease rates amongst African-Americans are perhaps not surprising. It is likely that there is a considerable genetic heterogeneity in African-Americans, based upon their country of origin or subsequent intermarriage. It is also likely that ethnic groups living in diverse geographic settings, even in the same country, are subject to different environmental and sociocultural influences. A recent study reported considerable variation in mortality rates due to cardiovascular disease within African-American populations depending on their place of birth. It is possible, therefore, that variations in rates of Alzheimer's disease for African-Americans in New York, Indianapolis, and North Carolina may be explained by a combination of genetic and site-specific environmental and sociocultural differences.

The prevalence rates of dementia in the Japanese-American studies are higher than those reported from Japan although the methodology in these studies is not identical. Whereas in Japan the most predominant form of dementia is vascular dementia, in both Japanese-American studies the predominant form of dementia was dementia of the Alzheimer's type. Age-adjusted prevalence rates of dementia (2.29 percent) and dementia of the Alzheimer's type (1.41 percent) in Ibadan, Nigeria, were significantly lower than the African-Americans in a study using identical methodology at both sites.

It is too early in the evolution of these studies to determine what might be the relevant interaction between genetic, environmental, social, and cultural factors that may increase or lessen the risk of developing Alzheimer's disease. A rather intriguing preliminary finding has been reported in the Seattle-based Japanese-American study where preservation of a Japanese lifestyle, including reading, writing and conversing in Japanese, was associated with low levels of cognitive decline and lower rates of dementia of the Alzheimer's type in that population. This is consistent with studies in the Cree in Northern Canada where the disorder also appeared to be relatively infrequent. The Cree attributed the apparent preservation of intellectual function to the role of the elderly. In this traditional, nonliterate, hunter-gatherer society, elderly Cree acted as the oral transmitters of cultural traditions from grandparents to grandchildren, thus ensuring a secure role for the elderly and ongoing intellectual and social stimulation. This role was still preserved at least in the reserves included in the study.

The importance of continued social activities was also highlighted in the Paquid study in France when regular participation in social and leisure activities such as traveling, odd jobs, knitting, or gardening was associated with a lower risk of subsequent dementia. Continuing intellectual stimulation has also been put forward as a possible explanation for the finding of the apparent protective effect of higher education in many (but not all) risk factor studies on dementia.

Sociocultural factors also have a considerable influence on the behavioral disorders associated with dementia and, in particular, to caregiver response to these disorders. These factors include the perception of the role of the elderly and the caregiver in society, the availability of a supportive social network, and the availability of nursing homes.

It is likely that sociocultural factors have a major influence on the etiology, diagnosis, and treatment of the psychiatric disorders of the elderly although the mechanisms explaining these influences are not always clear. The newer cross-cultural approaches offer great hope in elucidating these mechanisms and shedding light on possible social, environmental, and genetic interactions.

SUGGESTED CROSS-REFERENCES

All sections of Chapter 51, particularly Section 51.1b on the epidemiology of psychiatric disorders, are suggested for additional reading as well as Section 4.4 on cultural psychiatry.

SECTION REFERENCES

American Society on Aging: Serving elders of color: Challenges to providers and the aging network. American Society on Aging, San Francisco, March, 1992.

Brown DR, Ahmed F, Gary LE, Milburn NG: Major depression in a community sample of African Americans. Am J Psychiatr *152:*373, 1995.

*Butt DS, Beiser M: Successful aging: A theme for international psychology. Psychol Aging *2:*87, 1987.

Callahan CM, Wolinsky FD: The effect of gender and race on the measurement properties of the CES-D in older adults. Med Care *32:*341, 1994.

Canadian Study of Health and Aging: Study methods and prevalence of dementia. Can Med Assoc J *150:*899, 1994.

Coke MM: Correlates of life satisfaction among elderly African Americans. J Gerontol B Psychol Sci Soc Sci *47:*P316, 1992.

Copeland JRM, Davidson IA, Dewey ME, Gilmore C, Larkin BA, McWilliam C, Saunders PA, Scott A, Sharma V, Sullivan C: Alzheimer's disease, other dementias, depression and pseudo-dementia: Prevalence, incidence and three-year outcome in Liverpool. Br J Psychiatr *161:*230, 1992.

Fabrigoule C, Letenneur L, Dartigues JF, Zarrouk M, Commenges D, Barberger-Gateau P: Social and leisure activities and risk of dementia: A prospective longitudinal study. J Am Ger Soc *43:*485, 1995.

Fang J, Madhavan S, Alderman MH: The Association between birthplace and mortality from cardiovascular causes among Black and White residents of New York City. N Eng J Med *335:*1545, 1996.

Fillenbaum GG, Heyman A, Huber MS, Woodbury MA, Leiss J, Schmader KE, Bohannon A, Trapp-Moen B: Prevalence and three-year incidence of dementia in older black and white community residents. J Clin Epidemiol *51:*587, 1998.

Friedland RP: Epidemiology, education and the ecology of Alzheimer's disease. Neurology *43:*246, 1993.

Ganguli M, Ratliff G, Chandra V, Sharma S, Gilby J, Pandav R, Belle S, Ryan C, Baker C, Seaberg E, Dekosky S: A Hindi version of the MMSE: The development of a cognitive screening instrument for a largely illiterate rural elderly population in India. Int J Ger Psychiatr *10:*367, 1995.

Graves AB, Larson EB, Edland SD, Bowen JD, McCormick WC, McCurry SM, Rice MM, Wenzlow A, Uomoto JM: Prevalence of dementia and its subtypes in the Japanese American population of King County, Washington State: The *Kame* Project. Am J Epidemiol *144:*760, 1996.

Graves AB, Larson EB, Wenzlow A: Protective factors for cognitive decline in a Japanese American community. Abstract presented at The Fifth International Conference on Alzheimer's Disease and Related Disorders. Neurobiol Aging *17*(Suppl):S40, 1996.

Gurland B, Wilder D, Cross P, Lantigua R, Teresi J, Barrett V, Stern Y, Mayeux R: Relative rates of dementia by multiple case definitions, over two prevalence periods, in three sociocultural groups. Am J Ger Psychiatry *3:*6, 1995.

Hall KS, Hendrie HC, Britain HM, Norton JA: The development of a dementia screening interview in two distinct languages. Int J Meth Psychiatr Res *3:*1, 1993.

Hall KS, Ogunniyi AO, Hendrie HC, Osuntokun BO, Hui SL, Musick BS, Rodenberg CA, Unverzagt FW, Gureje O, Baiyewu O: A cross-cultural community-based study of dementias: Methods and performance of the survey instrument in Indianapolis, U.S.A., and Ibadan, Nigeria. Int J Meth Psychiatr Res *6:*129, 1996.

*Hendrie HC: Alzheimer Disease: A review of cross-cultural studies. In *Epidemiology of Alzheimer's: From Gene to Prevention*, R Mayeux, Y Christen, editors. Springer-Verlag, New York, 1999.

Hendrie HC: Epidemiology of dementia and Alzheimer disease. Am J Ger Psychiatr *6* (Suppl):S3, 1998.

Hendrie HC, Hall KS, Pillay N, Rodgers D, Prince C, Norton JA, Brittain H, Nath A, Blue A, Kaufert J, Shelton P, Postl B, Osuntokun BO: Alzheimer's disease is rare in Cree. Int Psychogeriatr *5:*5, 1993.

*Hendrie HC, Osuntokun BO, Hall KS, Ogunniyi AO, Hui SL, Unverzagt FW, Gureje O, Rodenberg CA, Baiyewu O, Musick BS, Adeyinka A, Farlow MR, Oluwole SO, Class CA, Komolafe O, Brashear A, Burdine VE: Prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. Am J Psychiatry *152:*1485, 1995.

Hendrie HC, Baiyewu O, Eldemire D, Prince C: Cross-cultural perspectives: Caribbean, Native American and Yoruba. Int Psychogeriatr *8:*1996.

Jablensky A, Sartorius N: Is schizophrenia universal? American Psychiatric Association: Ethnic minority elderly. A task force report of the American Psychiatric Association, Washington, DC, 1993.

Lobo A, Saz P, Marcos G, Dia JL, De-la-Camara C: The prevalence of dementia and depression in the elderly community in a southern European population. The Zaragoza study. Arch Gen Psychiatr *52:*497, 1995.

Murden RA, McRae TD, Kaner S, Buchanan ME: Mini-Mental State Exam scores vary with education in blacks and whites. J Am Geriatr Soc *39:*149, 1991.

*Murphy JM: Anthropology and psychiatric epidemiology. Acta Psychiatr Scand *90:*48, 1994.

*Sakauye KM: Ethnocultural aspects. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1997.

Tang MX, Stern Y, Marder K, Bell K, Gurland B, Lantigua R, Andrews H, Feng L, Tycko B, Mayeux R: The *APOE*-e4 allele and the risk of Alzheimer disease among African Americans, whites, and Hispanics. JAMA *278:*751, 1998.

Teng EL, Hasegawa K, Homma A, Imai Y, Larson E, Graves A, Sugimoto K, Yamaguchi T, Sasaki H, Chiu D, White LR: The Cognitive Abilities Screening Instrument (CASI): A practical test for cross-cultural epidemiological studies of dementia. Int Psychogeriatr *6:*45, 1994.

*Weissman MM, Bland RC, Canino GJ, Faravelli C, Greenwald S, Hwu HG, Joyce PR, Karam EG, Lee CK, Lellouch J, Lepine JP, Newman SC, Rubio-Stipec M, Wells E, Wickramaratne PJ, Wittchen HU, Yeh EK: Cross-national epidemiology of major depression and bipolar disease. JAMA *276:*293, 1996.

Welsh KA, Ballard E, Nash F, Raiford K, Harrell L: Issues affecting minority participation in research studies of Alzheimer disease. Alzheimer Dis Assoc Disord *8:*38, 1994.

White L, Petrovitch H, Ross W, Masaki KH, Abbott RD, Teng EL, Rodriguez BL, Blanchette PL, Havlik RJ, Wergowske G, Chiu D, Foley DJ, Murdaugh C, Curb JD: Prevalence of dementia in older Japanese-American men in Hawaii: The Honolulu-Asia Aging Study. JAMA *276:*955, 1996.

Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D: Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. Br J Psychiatr *159:*645, 658, 1991.

51.6 SPECIAL AREAS OF INTEREST

51.6E MINORITY ISSUES

ZF. M. BAKER, M.D., M.P.H.

The population of older United States citizens who are people of color is rapidly changing. Among the four groups that constitute the ethnic elder population, persons age 80 and older are the fastest growing segment of the population. The term “ethnic elder” is used to describe persons age 65 and older who are of African-American, American Indian or Alaskan Native, Asian-American, or Hispanic-American. The United States government has designated these four population groups minorities. By 2010 the combined populations of all ages for these four groups will represent the majority of persons in the Unites States, outnumbering persons of European background. By the year 2030 the combined numbers of older persons in these four groups (ethnic elders) will represent 25 percent of persons age 65 and older in the United States (Fig. 51.6e-1). The term ethnic elder is used to reflect both the age and the unique role that many of these elders hold in their extended families and their communities of residence.

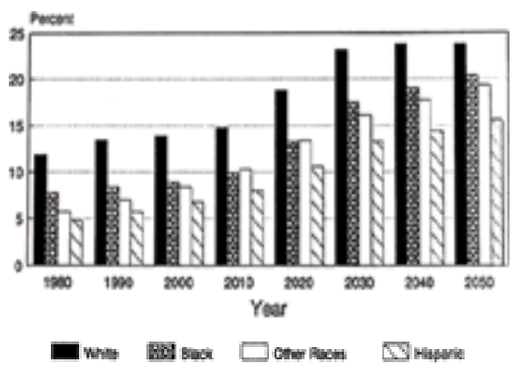

FIGURE 51.6e-1 Proportion of each race and ethnic group, age 65 and older. (Reprinted with permission from Angel JL, Hogan DP: The demography of minority aging populations. In *Minority Elders: Longevity, Economics, and Health*, L Harootyan, editor. Gerontological Society of America, Washington, DC, 1991.)

Preparing mental health professionals to work with an increasingly culturally diverse population of older Americans is an important task for the year 2000. Recent texts and the Glossary of Culture Bound Syndromes in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) are helpful resources. This DSM-IV appendix provides an outline for the assessment of the patient's cultural identity and how it affects the person's perception of illness. The influence of culture on the therapeutic alliance between the patient and the therapist is also addressed. As an initial effort at incorporating the known information about the cultural context of mental illness, the material in Appendix I, coupled with specific cultural content associated with each disorder, provides a useful resource for those unfamiliar with a specific culture and its influence on psychiatric symptoms and the definition of unusual behavior. The references at the end of this chapter provide information on key resources in this field.

The four groups that constitute ethnic minority elders are not homogeneous. The black American population has become increasingly diverse with immigration from the Caribbean islands and Africa, specifically Ghana, Senegal, and Nigeria. The American Indian and Alaskan Native elders include over 400 recognized tribal communities with differing customs and over 250 languages. More than 20 specific Asian-American populations fit the descriptive term “Asian-American elder.” The diversity in language and customs influencing the definition of illness and treatment is seen, also, among Hispanic-American elders, who represent 21 different cultural groups.

Historical perspective experiences of ethnic elders are important because whether a specific patient did or did not directly experience these events, the clinician can determine whether they contributed to the psychiatric symptoms present and whether they influenced the way in which the ethnic elder was brought to treatment.

Although many ethnic elders of the four groups have had limited education, marginal employment, poor access to health care resulting in chronic medical problems, and the psychological stress of living in economically disadvantaged areas, many others have been teachers, successful business owners, lawyers, and doctors. The mental health professional who finds out about the unique history of the ethnic elder being treated will avoid making incorrect generalizations.

African-American and Hispanic-American elders obtain psychological comfort from religion and have an extensive social network deriving from their churches. Among many Native American people and Asian-Americans, men and women are seen as part of an interactive environment with nature. Psychological well-being is defined by these elders as maintenance of an appropriate harmony with the environment. The ying and yang conceptualization of complementary environmental components, a Chinese belief system, is more comparable to the belief systems of the American Indians than to the Judeo-Christian belief system.

One's definition of illness is influenced by one's culture. Understanding how symptoms are associated with specific conceptualizations of illness can facilitate establishment of the therapeutic alliance. The therapist's knowledge can avert drug-drug interactions between Western medicine and non-Western treatments; medicinal teas may include ingredients with psychoactive properties. Ethnic elders may arrive with their families to see a psychiatrist as the “last stop.” They have been to their native healer, their religious leader, and their family doctor, and come only then to a psychiatrist, because of the stigma associated with mental illness among many of these groups. Frequently, psychiatric care is sought because symptoms fail to resolve and the family members and social network of the ethnic elder has become increasingly concerned. The multiple medical problems of many ethnic elders (hypertension, diabetes mellitus, glaucoma) increase the risk for drug-drug interactions or a delirium due to an atropine psychosis.

AFRICAN-AMERICANS

Demographic Characteristics Persons of African origin in the United States are African, African-American, and African-Caribbean. Most are African-Americans. African-Americans are 12 percent of the total United States population, and persons age 65 and older are 11 percent of the total African-American population. Persons age 85 and older are the fastest-growing segment of the African-American population; 17 percent of persons age 65 and older in 2030 and 21 percent of persons age 65 and older in 2050 will be African-American elders (Fig. 51.6e-1). For persons age 65 to 74, there are 67 men for every 100 women. In the age group 85 and older, there are only 46 men for every 100 women. Most elderly women are single (divorced, widowed, or never married), and most elderly men are married.

Nineteen percent of African-American elders completed high school, 4 percent completed college, 11 percent had 5 years or less of formal education, and 24 percent had no formal education. One third of African-American elders are below the poverty level, and 26 percent of households are below the poverty level. The extended family and informal social networks of friends, neighbors, and church members continue to be important social, financial, and spiritual resources for African-American elders.

Immigration History and Patterns of Acculturation Africans arrived on the North American continent as a result of war and the trading of prisoners of war in Africa. Surviving the climate and the heavy toil, Africans became the focus of the slave trade and provided the labor that built America into a major agricultural nation. The rich cultural heritage and value system (Table 51.6e-1) of predominantly West African tribes was being lost until Africans from various tribal groups began to use a new language, English, to communicate. With the formal ending of the brutal system of chattel slavery, the current cohort of African elders learned the history of the Reconstruction Era from their grandparents and great-grandparents (Table 51.6e-2). The extended family and the adoption and absorption of nonrelated persons into the family characterized the structure of African-American families during Reconstruction, and that pattern continues today.

Collectivity
Sharing
Affiliation
Obedience to authority
Belief in spirituality
Respect for the elderly
Respect for the past

Reprinted with permission from Pinderhughes E: Afro-American families and the victim system. In *Ethnicity and Family Therapy*, M McGoldrick, JK Pearce, J Giordano, editors. Guilford, New York, 1982.

Table 51.6e-1 African Cultural Values

TABLE cellSpacing=0 cellPadding=0 align=left border=0 hspace="10" vspace="5">

Table 51.6e-2 African-Americans: Significant Dates and Periods in History

Racial harassment and, following Reconstruction, economic recession, the development of the cotton gin, and the destruction of the cotton crop by the boll weevil stimulated the movement of African-Americans from the Southern states to the Midwest, the Northeast, and the West. Not until the migration to the North and the association of the alcohol abstinence movement with the Ku Klux Klan did many African-Americans begin to consume alcoholic beverages. The development of speakeasies and African-American involvement in entertainment and the arts contributed to the escalation in substance abuse by black Americans. The period of chattel slavery in the United States and the denigration of African-Americans by stereotyping contributed to the negative attitudes that characterized legalized segregation. The civil rights movement in the 1960s provided a rediscovery of the contributions of Africans to the development of the Western world. The specific contributions and strengths of the black community began to be summarized during that period in the literature of several fields.

The specific historical events that have formed the experience of African-American elders are summarized in Table 51.6e-2. The historical events are correlated with specific developmental stages in Table 51.6e-3.

Current Age of Cohorts	Age at Historical Events			
	1900–1920	1920–1940	1940–1960	1960–Present
	Ku Klux Klan Lynchings National Association for the Advancement of Colored People Participation in World War I	Red Summer W. E. B. DuBois Marcus Garvey's back to Africa movement Harlem renaissance Great Depression	World War II: segregated units; factory work in North and West World War II soldiers return to segregated society 1954: Supreme Court ruling desegregating schools	Civil rights movement Rev. Dr. Martin Luther King's nonviolent protests Black power movement Desegregation and affirmative action Vietnam War Crack epidemic
85+	Children and adolescents	Young adults	Middle-aged	Elders
75–85	Children	Adolescents and young adults	Young adults to middle-aged	Middle-aged to old
65–75		Children and adolescents	Adolescents and young adults	Adults to young old
55–65		Children	Children to young adults	Young adults to middle-aged

Reprinted with permission from Baker FM: The black elderly: Biopsychosocial perspectives within an age cohort and adult development context. J Geriatr Psychiatry 15:313, 1982.

Table 51.6e-3 Correlation of Historical Events With Developmental Stages: Cohort Experiences, African-American Elders

Medical Problems The rate of diabetes mellitus in African-Americans is five times higher than that in white Americans. The rates of hypertension and obesity are also significantly higher for African-Americans than for some ethnic elders and white Americans. The 5-year survival figures for black Americans with malignancies of the breast, cervix, uterus, prostate, and esophagus are only one third to one fourth the survival figures for white Americans. The rate of cerebrovascular disease among elderly African-Americans is three times higher than the rate among white Americans. In black elders alcohol is the substance of abuse, with significantly greater morbidity and mortality for black men than for black women. The rates of glaucoma are higher and the progression of the disease faster among African-American elders than among white Americans. Of the chronic diseases, arthritis is a particular concern for elderly African-Americans. The number of chronic medical problems increases the risk for major depressive disorders and episodes of delirium.

Psychiatric Disorders Various authors have questioned whether the rates of dementia differed in African-American elders. A higher than usual prevalence of vascular dementia was expected because of the high rates of hypertension. Recent data from a few investigators provide some support for that hypothesis. Although the prevalence of dementia of the Alzheimer's type, with late onset, was hypothesized to be the same as in white Americans, recent work by investigators at Duke University on a biracial subsample drawn from the Duke University longitudinal survey found a higher prevalence rate of the dementia among black elders. A cross-cultural study comparing African-American elders in Indianapolis, Indiana, with Africans in Ibadan, Nigeria, found that rates of dementia of the Alzheimer's type among African-American elders were similar to those of the total U.S. population, while lower rates of dementia were found among the Ibadan elders. Further studies are needed.

Epidemiological surveys that have screened elderly black Americans for psychiatric symptoms found rates of depressive symptoms that varied from 26 to 60 percent. Since ambulatory medical clinic patients and medical inpatients have rates of depressive illness that range from 29 to 33½ percent, it was hypothesized that because of their multiple chronic health problems, African-American elders would have higher rates of depressive disorder due to medical illness (secondary depression) than the 10 to 15 percent found in community resident elders. Recent studies have confirmed this hypothesis. Older medically ill, physically disabled, and hospitalized African-Americans who were depressed were found to have a remission of their depression at 6-month follow-up compared to elderly whites. This was a consistent finding with no change in the number of medical illnesses and the level of physical disability. Although the National Institute of Mental Health Epidemiologic

Catchment Area (ECA) Survey data reported above-average prevalence rates for phobic disorders among African-Americans in Baltimore, Maryland, and Durham, North Carolina, the prevalence of those disorders among elderly African-Americans awaits further investigation.

The Center for the Establishment of a Registry for Alzheimer's Disease (CERAD) has coordinated a national multicenter evaluation of the most frequently used screening tests to assess cognitive impairment. Norms for African-American elders for the specific neuropsychological tests of the CERAD screening battery were published in 1996.

Diagnosis The period of chattel slavery and Reconstruction resulted in a biased view of African-Americans that was reflected in psychiatric diagnoses. Blacks attempting to run to freedom were said to have the psychiatric disorder of “drapsomania.” In the early 1800s black Americans were reported to lack the intrapsychic structure to become depressed because of their “happy-go-lucky” nature. Although these psychiatric formulations reflected the societal bias of the investigators, misdiagnosis of psychiatric symptoms in African-American patients remains a concern today.

An African-American with a history of alcohol abuse beginning at age 14 may experience an episode of alcohol-induced psychotic disorder with hallucinations at age 24. If the psychotic symptoms were diagnosed as schizophrenia, that diagnostic label may continue for decades. A few authors have described bipolar disorder in poor black populations, a psychiatric disorder previously diagnosed mainly in middle-class white populations. An elderly African-American with a 30-year history of psychiatric illness requires a thorough record review and family interview to clarify the symptoms, the course of the disorder, and the presence or absence of alcohol use during the period of acute illness. Although the recent literature has emphasized the presence of psychotic symptoms in a variety of psychiatric disorders (alcohol hallucinosis, delusions, bipolar disorder), some psychiatrists view those symptoms as pathognomonic for schizophrenia. African-American elders of the 1990s with psychiatric illnesses with an onset in their 20s were diagnosed at a time when American psychiatrists made diagnoses of schizophrenia more often than did psychiatrists in Great Britain. A review of those diagnostic labels is essential.

Suicide Despite psychosocial stressors for poor African-American elders, their rates of completed suicide are significantly lower than those for elderly white Americans. Rates of completed suicide in the United States for 1990 and 1993 for white males ranged from 17.9 per 100 population (ages 65 to 74) to 22.8 per 100 population (ages 85 and older). In contrast, the rates of completed suicide for black men ranged from 14.7 (ages 65 to 74) to a rate too small to be calculated (age 85 and older). Rates of completed suicide for white women for 1990 and 1993 ranged from 7.2 per 100 population (age 65 to 74) to 5.4 per 100 population (age 85 and over). The rates of completed suicide among older black women were one third those of older white women, a rate of 2.6 (1990) and 2.2 (1993) for ages 65 to 74. Too few black women ages 75 to 84 and 85 and older completed suicides to calculate a rate of completed suicide in 1990 and in 1993. Religion and social support networks contribute to the maintenance of self-esteem by African-American elders. Their involvement in their churches and friendship networks give them multiple social roles in their retirement years.

Effects of Drugs Although 48 percent of African-Americans are slow acetylators, the importance of that fact in the metabolism of psychoactive medications is unclear. This fact may explain why older African-Americans elders treated with high dosages of antipsychotic medication may experience increasing confusion and increased severity of parkinsonian adverse effects. A few reports have indicated that African-American elders have been treated with high doses of antipsychotic medication to decrease or to prevent agitation. Further research is needed in older black patients to determine whether lower doses of antipsychotic medication are more effective than higher doses and result in fewer adverse effects.

African-American elders have been less likely to reside in nursing homes because of care provided by their children and extended family members. Those African-American elders in nursing home settings had similar rates of dementia (68 percent) compared to elderly white nursing home residents (70 percent). Rates of psychiatric disorders among elderly African-American nursing-home residents were as high as among white nursing-home residents, but the rates of behavioral disturbances were lower among African-Americans.

AMERICAN INDIANS AND ALASKAN NATIVES

Demographic Characteristics American Indians and Alaskan Natives (Eskimos and Aleuts) comprise more than 400 distinct tribal groups and speak more than 250 languages besides English. The Aleuts reside largely along the Aleutian chain, and various Indian groups reside in the interior and southeastern areas of Alaska. The term “Alaskan Native” also includes the Inuit Eskimos, who reside in north and northwest Alaska, and the Yupik Eskimos, who are found in southwest Alaska. These groups differ in culture, history, and language. Although data on American Indian and Alaskan Native populations are available from the Indian Health Service, less information is available from the U.S. Bureau of the Census.

American Indians and Alaska Native represent less than 1 percent of the total United States population, and persons age 65 and older represent 6 percent of the American Indian and Alaskan Native population. Nationally, 54 percent of American Indians live in predominantly urban areas; less than 24 percent live on reservations. In the 65-to-74 age group, there are 77 men for every 100 women; for age 85 and older, there are only 59 men for every 100 women. In those ages 65 to 74, 25 percent of American Indians and Alaskan Natives completed high school, 4 percent completed college, 11 percent completed the fifth or sixth grade, and 23 percent had no formal education. Twenty-five percent of urban-resident and 39 percent of rural-resident American Indian and Alaskan Native elderly persons live below the poverty level. The extended family or clan, as defined by some American Indian tribes, is an important psychosocial resource for the American Indian and Alaskan Native elder.

Immigration History and Patterns of Acculturation American Indians were the first Americans. Their adjustment and cultural accommodation with the varying populations of immigrants were sometimes accomplished without significant conflict but were more commonly associated with significant tensions that escalated to frank hostility.

Today's American Indian elders experienced the traumatic rupture of their familial and community bonds during their crucial adolescent years. The practice of moving American Indian children to boarding schools to facilitate their education and understanding of American culture ended only recently. The psychosocial stressors of attempting to bridge a dual identity (experienced by all groups of nonwhite Americans) were focused at a crucial developmental age for those American Indian adolescents and youths. The recent economic expansion resulting from oil exploration has produced a similar disruption of cultural continuity for young Alaskan Natives. The historical events that have formed the experience of American Indian elders are detailed in Table 51.6e-4.

Table 51.6e-4 American Indians and Alaska Natives: Significant Dates and Periods in Recent History

Medical Problems Several investigators have shown that an American Indian at age 55 has medical and psychosocial problems similar to those of a white American at age 65. The major health problems of American Indians and Alaskan Natives include non-insulin-dependent diabetes mellitus, hypertension, tuberculosis, heart disease, cancer, liver, and kidney disease.

Psychiatric Disorders Although the Center for Epidemiologic Studies Depression Scale (CES-D) has been used in a study of three tribes, a detailed study of screening instruments for all tribal groups of American Indian and Alaskan Native elders has not been completed. The CES-D was found to have good reliability and

validity, and an estimation of depressive disorder ranging from 30 to 50 percent was based on the evaluation of three tribal groups in the southwestern United States.

The CERAD project has developed a specific culture-fair battery to screen ethnic elder populations for cognitive impairment. That battery has not been applied to a population of community-resident American Indian and Alaskan Native elders to determine the prevalence of symptoms of cognitive impairment, an important area of future research.

The problem of alcohol dependence varies across tribal groups. A few authors have shown that depressive symptoms and mood disorders are associated with patterns of alcohol abuse in four American Indian tribes. Those studies assessed the presence of depressive symptoms after 6 to 8 weeks of alcohol abstinence.

Suicide The mortality caused by alcohol-related motor vehicle accidents among American Indians of all ages and by suicide among Alaskan Natives age 55 and younger underscores the importance of implementing effective primary and secondary preventive strategies. The rates of suicide among American Indian and Alaskan Native elders are significantly lower than those for elderly white Americans.

Effects of Drugs Lower dosages of antipsychotic and antidepressant medications may be indicated. Research to determine whether specific tribal groups of American Indians are slow acetylators needs to be completed.

ASIAN-AMERICANS AND PACIFIC ISLANDERS

Demographic Characteristics Asian-American and Pacific Islander elders comprise more than 20 different ethnicities. Their origins vary widely: East Asia, Southeast Asia, the Indian subcontinent, Polynesia, Melanesia, and Micronesia. Asian-Americans and Pacific Islanders constitute 1 percent of the total United States population. Six percent of the Asian-American and Pacific Islander population are age 65 and older. Sixty-four percent of these ethnic elders live in California, Hawaii, and Washington (in declining order of residence). Sixty-five percent were born outside the United States. The four largest groups are Filipino (26.7 percent), Chinese (26.6 percent), Japanese (24.7 percent), and Asian Indian (14.6 percent) (Fig. 51.6e-2). Ninety percent of Asian-Americans and Pacific Islanders live in urban areas. Thirty-seven percent of elderly Asian-Americans and Pacific Islanders completed high school, 10 percent completed college, 13 percent completed the fifth or sixth grade, and 21 percent had no formal education. Fourteen percent of Asian-American elders live below the poverty level, and 10 percent of rural-resident elders live below the poverty level.

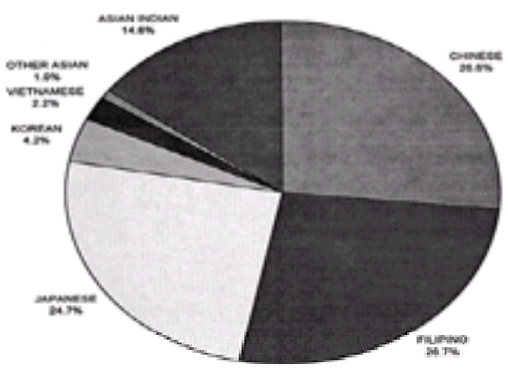

FIGURE 51.6e-2 Asian elderly population by nation of origin, 1980. (Reprinted with permission from Angel JL, Hogan DP: The demography of minority aging populations. In *Minority Elders: Longevity, Economics, and Health*, L Harootyan, editor. Gerontological Society of America, Washington, DC, 1991.)

Immigration History and Patterns of Acculturation A biphasic immigration pattern resulted from discriminatory legislation. In the late 1880s to 1924 often single, poorly educated immigrants from China, Japan, and the Philippines immigrated to the United States to provide cheap manual labor. Intending to make their fortunes before returning home or bringing their families to the United States, they could not fulfill those plans. With the Chinese Exclusion Act of 1882 and the Oriental Exclusion Act of 1924, immigration and naturalization of all Asian groups stopped. Because of legal sanctions and informal sanctions against intermarriage, those early, predominantly male immigrants became a bachelor society, with the exception of the Japanese who selected brides from pictures. Repeal of the Asian Exclusion Acts in 1965 resulted in a second wave of immigration that included many Asian women. Elderly wives who had been excluded in the past received preferential entry by the New Immigration Act of 1965. Key historical events for Chinese-Americans are summarized in Table 51.6e-5.

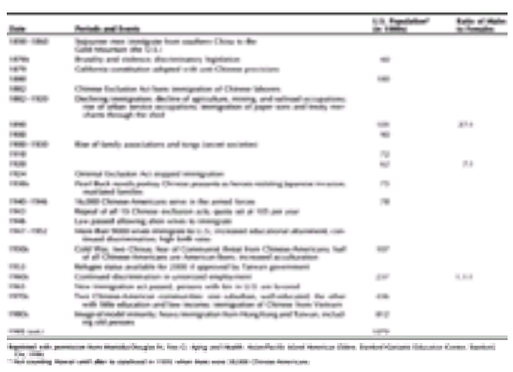

Table 51.6e-5 Chinese Americans: Significant Dates and Periods in Immigration and History

A few investigators have reported that of foreign-born Asian-Americans and Pacific Islanders age 65 and older in 1990, in decreasing order the greatest number was Japanese, followed by Filipino, Asian Indians, and Chinese. One study reported that of 100 low-income Asians age 55 and older in Asian communities in Los Angeles, 75 percent spoke only their native language, with 77 percent reporting that they had resided in the United States for more than 20 years.

Medical Problems Death rates are lower for Chinese, Japanese, and Filipinos (the only statistics on the various Asian and Pacific Islander groups kept by the National Center for Health Statistics) than for whites. Specific population-based studies have found low rates of mortality caused by coronary artery disease and acute myocardial infarction and a higher than average incidence of liver and esophageal cancers in Chinese populations. Second-generation (Nisei) Japanese-Americans were found to have three to six times more non-insulin-dependent diabetes mellitus than white American men, to have lower rates of kidney-related complications among diabetics, and to have a higher incidence of stomach cancer.

Psychiatric Disorders Depression A few investigators have studied the prevalence of depressive disorders among a random sample of Chinese, Filipino, Japanese, and Korean persons who lived in Seattle, Washington. Using the CES-D, they found that the mean score for the Asian-American sample combined was 9.38, in comparison with a mean score for whites that ranged from 7.96 to 9.25. Nineteen percent of Asian-Americans tested positive for depressive symptoms (a CES-D score of 16 or above). Koreans, the most recent immigrants, had the highest mean CES-D scores, followed by Filipinos, Japanese, and Chinese. Those differences were statistically significant. Recent immigrant status and resultant adjustments, low employment levels, and low socioeconomic status contributed to the high CES-D scores. The prevalence of depressive symptoms among these Asian-American residents of Seattle was at least as high as that in white Seattle residents and was consistent with community surveys of the mental health of Asian-Americans in California.

Posttraumatic Stress Disorder In a study of Southeast Asian refugee patients, a systematic reassessment of the population found significant rates of posttraumatic stress disorder. Patients' underreporting of symptoms, the failure of psychiatrists to explore a history of trauma in non-Cambodian groups, and the necessary unhurried, unobtrusive clinical approach to diagnosis and treatment required for working with the population contributed to the failure to identify posttraumatic stress

disorder. Among these refugees somatic complaints (headaches, stomachaches, poor sleep), were more commonly reported than depressive symptoms, which contributed to underdiagnosis of depression. The recent immigration of populations from Southeast Asia requires therapists to evaluate for the presence of posttraumatic stress disorder. Vietnam refugees resettled in cities throughout the United States are at risk.

Suicide The rates of completed suicides by Chinese and Japanese women more than 75 years old exceed the rates for white American women (Table 51.6e-6). Only in the old-old (persons 85 and older) do the rates of completed suicide by Asian-American men exceed the rates for white American men; Japanese men have a suicide rate of 139.76, Chinese men have a suicide rate of 64.10, and white American men have a suicide rate of 53.28. As shown in Table 51.6e-6, the total age-specific and age-adjusted suicide rates per 100,000 population are higher for Asian-American elders than for white Americans over age 65. Factors contributing to the high rates have been linked to the absence of spouses and families for the Asian men who immigrated to the United States to obtain better employment. As a result of the Asian exclusion laws, those Chinese, Japanese, and Filipino men never married, and these bachelor cohorts, resident in ethnic communities, are alone and have minimal social support. These elderly Japanese, Chinese, and Filipino men may be depressed, but they have not been studied for the prevalence of psychiatric disorders, particularly mood disorders. Some Japanese men in these cohorts brought brides to the United States and married before the Oriental Exclusion Act of 1924.

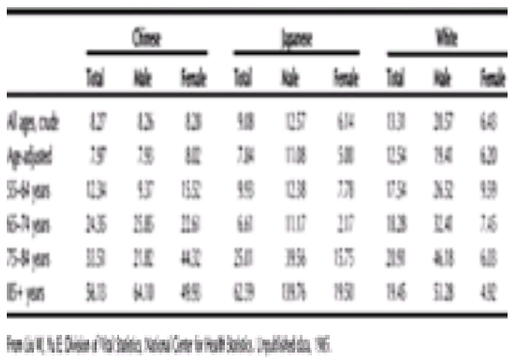

	Chinese			Japanese			White		
	Total	Male	Female	Total	Male	Female	Total	Male	Female
All ages, crude	8.27	8.26	8.28	9.08	12.57	6.14	13.31	20.57	6.40
Age-adjusted	7.59	7.95	8.02	7.84	11.08	5.08	12.54	19.41	6.20
55–64 years	12.34	9.37	15.52	9.93	12.38	7.78	17.54	26.52	9.59
65–74 years	24.35	25.85	22.61	6.61	11.17	2.17	18.28	32.41	7.45
75–84 years	33.51	21.02	44.22	25.01	39.56	15.75	28.91	46.18	6.03
85+ years	56.13	64.10	49.91	42.39	139.76	19.50	19.45	53.28	4.92

From Liu WT, Yu E: Chinese at Risk? Statistics, National Center for Health Statistics. Unpublished data, 1985.

Table 51.6e-6 Average Annual Age-Specific and Age-Adjusted Suicide Rates per 100,000 United States Population for Specified Races, 1980 (Age 55 to Age 85)

Elderly Japanese and Chinese women immigrated to the United States in two waves: before the Asian exclusion acts and after those acts were rescinded in the 1960s. Some investigators suggest that these elderly women completed suicide because of their mutilated family experience, the absence of children, social isolation, and minimal involvement with the surrounding American culture. Although elderly Chinese and Japanese women may be at risk of mood disorders, no studies of those populations for the prevalence of psychiatric disorders have been completed to date. A correlation of specific developmental states with historical events for Chinese-Americans is presented in Table 51.6e-7.

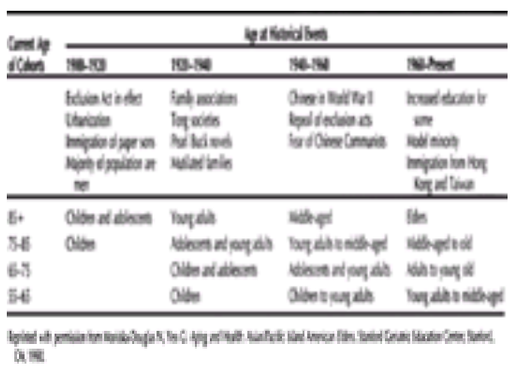

Current Age of Cohorts	Age at Historical Events			
	1888–1920	1920–1940	1940–1960	1960–Present
	Exclusion Act in effect Urbanization Immigration of paper sons Majority of population are men	Family associations Tong societies Pearl Buck novels Mediated families	Chinese in World War II Repeal of exclusion acts Fear of Chinese Communists	Increased education for some Model minority Immigration from Hong Kong and Taiwan
85+	Children and adolescents	Young adults	Middle-aged	Elders
75–85	Children	Adolescents and young adults	Young adults to middle-aged	Middle-aged to old
65–75		Children and adolescents	Adolescents and young adults	Adults to young old

Reprinted with permission from Morishima/Douglass PA, Yee C: Aging and Health: Asian/Pacific Island American Elders. Stanford Geriatric Education Center, Stanford, CA, 1990.

Table 51.6e-7 Correlation of Historical Events With Developmental Stages: Cohort Experiences, Chinese American Elders

Effects of Drugs Several studies found Asian-Americans to require one third to one half of the usual dosage of antidepressant and antipsychotic medications. Using lower-than-usual dosages in elderly patients and slowly titrating increases of psychoactive medications upward is particularly important for these ethnic elders.

HISPANIC-AMERICANS

Demographic Characteristics The earliest Hispanic immigrants to the continental United States were from Mexico and Puerto Rico. They were followed by Cuban immigrants who fled the Castro revolution and by Salvadorans, Nicaraguans, and Colombians. Hispanic-Americans represent 8 percent of the total United States population. They have a younger than average median age, higher fertility rates, and increasing immigration—factors that contribute to the rapid growth of the population in the nation. Among Hispanic-Americans (Fig. 51.6e-3), the percentage of persons age 65 and older varies: for Mexican-Americans, 47.6 percent; for Cuban-Americans, 18.7 percent; for Puerto Ricans, 10.9 percent; and for Central and South Americans, 8.4 percent. For Hispanic Americans age 65 to 74, there are 86 men for every 100 women. At age 85 there are 61 men for every 100 women. Of Hispanic-Americans who were ages 65 to 74 in 1990, 20 percent completed high school, 4 percent completed college, 18 percent completed the fifth or sixth grade, and 35 percent had no formal education. Twenty-six percent of Hispanic-Americans live below the poverty level.

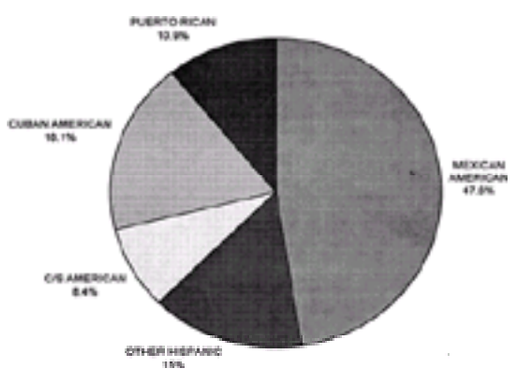

FIGURE 51.6e-3 Hispanic elderly population by nation of origin, 1989. C/S American, Central and South American. (Reprinted with permission from Angel JL, Hogan DP: The demography of minority aging populations. In *Minority Elders: Longevity, Economics, and Health*, L Harootyan, editor. Gerontological Society of America, Washington, DC, 1991.)

Immigration History and Patterns of Acculturation Issues of exclusion based on ethnicity, level of acculturation, and language vary for each component of the Hispanic-American population. Mexican-Americans, who have a unique history in the development of the southwestern United States, include a population ranging from the original landowners and indigenous residents of Texas to the undocumented workers of the 1990s (Table 51.6e-8 and Table 51.6e-9).

Dates	Periods and Events	U.S. Population[1] (in 1000s)
Late 1500s–early 1600s	Gradual colonization of northern provinces of New Spain	
1821	Mexico obtains independence from Spain	
1845–1854	Conquest of northern Mexico by U.S. and annexation of territories of U.S. Southwest	80
	Increased immigration to California for gold rush and to Texas and Arizona ranchos and farms; anti-Mexican laws, rough discrimination, invalidation of land titles for many Mexican land-grant holders	
1850s–1900		
1900		381 to 562
1910	Mexican Revolution followed by political and economic chaos, spurring immigration to the U.S.	
	500,000 enter on permanent visas and thousands more enter informally; cause industry; establishment of the border patrol; increased employment in railroads and manufacturing, some in mid-west; development of urban barrios	
1920–1930	Immigration increases dramatically because of depression and dust bowl; forced repatriation of 500,000 residents to Mexico; family-oriented migrant farming; segregated housing and social organization, including labor unions	
1938–1942	350,000 Mexican-Americans serve in World War II; increased immigration to fill war-related industrial jobs; GI Forums programs anti-Mexican press coverage of conflict between zoot-suiters and Anglo servicemen and police	
1942–1947	Ex-GIs organize for civil rights, declining agricultural jobs, increasing urbanization and education in highly segregated communities; documented immigration of 273,000 in the 1950s; periodic raids and mass deportations of undocumented residents	
1947–1960		
1960s	Chicano movement; increased immigration of women	
	Immigration and Nationality Act sets 20,000 ceiling on annual immigration and issues family reunification of U.S. permanent residents	
1965		
	Bi-directional cultural migration from Texas to the Midwest and the West; greater political organization and passion; greater recognition of Mexican-Americans in U.S. Catholic Church; bilingual education; wide-scale deportation of undocumented workers	
1970s		6,000
	Emphasis on education, continuation of Mexican heritage and Spanish language through arts and the media; amnesty program; continued[2] deportations	
1980s		
1990	Increasing role in American media; increase in bilingual materials and advertising	11,000

[1] Approximate data; permission from Carlos J. Aging and Health: Hispanic American Elders. Stanford Geriatric Education Center, Stanford, CA, 1990.
[2] Estimated numbers are highly controversial; 12 million used.

Table 51.6e-8 Mexican-Americans: Significant Dates and Periods in Immigration and History

Current Age of Cohorts	Age at Historical Events			
	1900–1920	1920–1940	1940–1964	1960–Present
	Heritage of loss of land	Massive immigration	World War II participation	Chicano movement
	Mexican Revolution	Depression	Immigration	Deportations
		Repatriation	Urbanization	
			GI forum	
85+	Children and adolescents	Young adults	Middle-aged	Elders
75–85	Children	Adolescents and young adults	Young adults to middle-aged	Middle-aged to old
65–75		Children and adolescents	Adolescents and young adults	Adults to young old
55–65		Children	Children to young adults	Young adults to middle-aged

Reprinted with permission from Carlos J. Aging and Health: Hispanic American Elders. Stanford Geriatric Education Center, Stanford, CA, 1990.

Table 51.6e-9 Correlation of Historical Events With Developmental Stages: Cohort Experiences, Mexican-American Elders

Medical Problems Although most Hispanic-Americans have higher than average mortality rates—which have been related to the stress and wear and tear of raising large numbers of children—mortality and morbidity data are lacking because no large-scale studies have been done. Most existing studies are case reports or descriptive studies. The National Health and Nutrition Examination Survey that focused on Hispanic-Americans provided the first national data on Hispanic-Americans. United States census data did not separate Hispanics from whites until 1979.

Several investigators reported that arthritis, hypertension, circulation problems, and diabetes mellitus are the primary diseases of concern. Mexican-born persons had a rate of death caused by accidents that was twice that of Puerto Rican–born and Cuban-born populations. Mexican-born persons had the highest cerebrovascular disease mortality rates, and Cuban-born persons had the lowest. The mortality rates for chronic liver disease and cirrhosis in Puerto Ricans were twice those of Mexicans and three times those of Cubans. Mortality rates for homicide were high in all groups of Hispanic-Americans, but highest among Puerto Ricans.

Psychiatric Disorders Few studies have looked at psychiatric disorders in elderly Hispanic-Americans. One study used a regrouped set of Comprehensive Assessment and Referral Evaluation (CARE) items in an probability sample of 700 elderly Hispanics living in Los Angeles County to reflect criteria from the revised third edition of DSM (DSM-III-R). The authors found that more than 26 percent of the population had depression or dysphoria. Mood disorder correlated strongly with medical disability and dementia.

Mexican-American and Puerto Rican populations have been studied to a limited extent for the presence of psychiatric symptoms and disorders. High acculturation was associated with alcohol dependence and suicide among Mexican-American men. A study of a Puerto Rican community found higher-than-average rates of depressive disorders among persons who had little involvement with the American culture. Factors contributing to the rates of alcohol dependence and suicidal behavior were related to a failure of Puerto Rican persons to resolve the transition from rural Puerto Rico to urban centers.

Recent data from the CERAD investigators confirm the findings of ECA investigators that components of the Mini-Mental State Examination (serial sevens, retrograde spelling of *world*, and knowledge of one's county of residence) were more likely to be missed by nonwhite persons, particularly those with less than 8 years of education. In addition, the CERAD investigators identified specific culturally appropriate alternative questions that maintained the intent of the screens for cognitive impairment. Spanish translations of the revised instruments are available from CERAD investigators for the Mini-Mental State Examination and the Blessed Dementia Screen, an assessment of memory and functional level. Whether Hispanic-Americans have biological differences in the metabolism of psychoactive medications is unclear.

Suicide Among Hispanic elderly men the rates of suicide were the second highest among ethnic elders (Table 51.6e-10). Suicide was highest among Cuban and Puerto Rican men and lowest among women, particularly Mexican women.

Race	Number of Deaths	Rate per 100,000
White	6139	22.5
Black	164	6.6
American Indian and Alaskan Native	7	6.7
Asian	53	11.5
Hispanic (any race)	74	10.3

From National Center for Health Statistics. Unpublished data, 1990.

Table 51.6e-10 Completed Suicides by Ethnic Elders Age 65 and Older (1988 Data)

Effects of Drugs The effects of drugs are unknown.

FUTURE DIRECTIONS

Studies to identify factors contributing to the decreased suicide rates among ethnic elders have not been published to date. A few studies have addressed the adjustment of Asian elders in nursing home settings and have assessed the presence of depressive symptoms among ethnic elders. Studies differentiating between the presence of depressive symptoms and depressive illness among various populations of ethnic elders remain to be completed.

Areas of research among African-American elders include further assessment of the prevalence of Alzheimer's disease and vascular dementia in this population. It will be important to determine whether African-American elders have as high a prevalence rate of phobic disorders as African-Americans of younger age cohorts (ECA

survey data). The impact of slow acetylation among African-American elders and its impact on the metabolism of psychoactive medications remains unstudied.

Establishing effective dosage levels of antipsychotic and antidepressant medications is crucial to the effective management of all ethnic elders. Although there are some data for Asian-American populations, few if any published studies address the use of medications in African-American, American Indian and Alaskan Native, and Hispanic-American elders.

SUGGESTED CROSS-REFERENCES

Other pertinent parts of the chapter on geriatic psychiatry include Section 51.2 on assessment; Section 51.3 on psychiatric disorders of late life, (particularly Section 51.3e on Alzheimer's disease and other dementing disorders and Section 51.3d on mood disorders), and Section 51.4 on treatment. A general discussion of psychological sciences appears in Chapter 3, of sociocultural sciences in Chapter 4, of epidemiology in Section 5.1, of alcoholism in Section 11.2, of posttraumatic stress disorder in Section 51.3c, of suicide in Section 29.1, and of community psychiatry in Section 51.5d.

SECTION REFERENCES

Adebimpe VR: Race, racism, and epidemiologic survey. Hosp Community Psychiatry *45:*27, 1994.

Baker FM: Ethnic minority elders: A mental health research agenda. Hosp Community Psychiatry *43:*337, 1992.

Baker FM: Suicide among ethnic minority elderly: A statistical and psychosocial perspective. J Geriatr Psychiatry *27:*241, 1994.

*Baker FM, Lightfoot OB: Geriatric psychiatry: Evaluation and treatment of psychiatric disorders in ethnic elders. In *Culture, Ethnicity, and Mental Illness*, AC Gaw, editor. American Psychiatry Press, Washington, DC, 1993.

Baker FM, Okwumabua J, Philipose V, Wong S: Screening African American elderly for the presence of depressive symptoms: A preliminary investigation. J Geriatr Psychiatry Neurol *9:*127, 1996.

Baker FM, Parker DL, Wiley C, Velli SA, Johnson JT: Depressive symptoms in African American medical patients. Int J Geriatr Psychiatry *10:*9, 1995.

Bernard MA, Lampley-Dallas V, Smith L: Common health problems among minority elders. J Am Diet Assoc *97:*7, 1997.

Bowman KF, Landefeld CS, Quinn LM, Palmer RM, Kowal J, Fortinksy RH: Strain in African American and white American caregivers of hospitalized elderly: implications for discharge planning. Research on Aging *20:*547, 1998.

Class CA, Unverzagt FW, Gao S, Hall KS: Psychiatric disorders in African American nursing home residents. Am J Psychiatry *153:*677, 1996.

Cohen GD: African American issues in geriatric psychiatry: A perspective on research opportunities. J Geriatr Psychiatry Neurol *6:*195, 1993.

*Cuellar J: *Aging and Health: American Indian/Alaska Native*. Stanford Geriatric Education Center, Stanford, CA, 1990.

Cueller J: *Aging and Health: Hispanic American Elders*. Stanford Geriatric Education Center, Stanford, CA, 1990.

del Carmen R: Assessment of Asian-Americans for family therapy. In *Mental Health of Ethnic Minorities*, FC Serafica, Al Schwebel, RK Russell, PD Isaac, LB Meyers, editors. Praeger, New York, 1990.

Fawzi MCS, Pham T, Lin L, Nguyen TV, Ngo D, Murphy E, Mollica RE: The validity of posttraumatic stress disorder among Vietnamese refugees. J Trauma Stress *10:*101, 1997.

Fernandez ME, Mutran EJ, Reitzes DC, Sudha S: Ethnicity, gender, and depressive symptoms in older workers. Gerontologist *38:*71, 1998.

Folstein MF, Bassett SS, Anthony JC, Raymonoski AJ, Nestadt GR: Dementia: Case ascertainment in a community survey. J Gerontol *46:*M132, 1991.

Guarnaccia PJ, Parra P: Ethnicity, social status, and families' experiences of caring for a mentally ill family member. Community Ment Health J *32:*243, 1996.

*Harootyan L, editor: *Minority Elders: Longevity, Economics, and Health*. Gerontological Society of America, Washington, DC, 1991.

Hatton DC: Health perception among older urban American Indians. West J Nurs Res *16:*392, 1994.

Hendrie HC, Osuntokun BO, Hall KS, Ogunniyi AO, Hui SL, Unverzagt FW, Gureje O, Baiyewu O, Rodenberg CS, Musick BS, Farlow MR, Class CA, Brashear A, Burdine VE, Oyewole S, Raji SO, Komolafe O: The prevalence of Alzheimer's disease and dementia in two communities: Nigerian Africans and African Americans. Am J Psychiatry *152:*1485, 1995.

Kinzie JD, Boehnlein JK, Leung PK, Moore LJ, Riley C, Smith D: The prevalence of posttraumatic stress disorder and its clinical significance among Southeast Asian refugees. Am J Psychiatry *147:*913, 1990.

Kleinman A: How is culture important in DSM-IV? In *Culture and Psychiatic Diagnosis. A DSM-IV Perspective*, JE Mezzich, A Kleinman, H Fabrega, D Parron, editors. American Psychiatric Press, Washington, DC, 1996.

Koenig HG, George LK: Depression and physical disability outcomes in depressed medically ill hospitalized older adults. Am J Geriatr Psychiatry *6:*230, 1998.

Kwumabua JO, Baker FM, Wong SP, Pilgram BO: Characteristics of depressive symptoms in elderly urban and rural African Americans. J Gerontol A Biol Sci Med Sci *52:*241, 1997.

Levin JS, Markides KS, Ray LA: Religious attendance and psychological well-being in Mexican Americans: a panel analysis of three-generations data. Gerontologist *36:*454, 1996.

*Lyketsos CG, Chen LS, Anthony JC: Cognitive decline in adulthood: An 11.5 year follow-up of the Baltimore Epidemiologic Catchment Area study. Am J Psychiatry *156:*58, 1999.

Manson SM, Walker RD, Kivlahan DR: Psychiatric assessment and treatment of American Indians and Alaskan Natives. Hosp Community Psychiatry *38:*165, 1987.

Mendoza R, Smith MW, Poland RE, Lin K-M, Strickland TL: Ethnic psychopharmacology: The Hispanic and Native American perspective. Psychopharmacol Bull *27:*449, 1991.

Morioka-Douglas N, Yeo G: *Aging and Health: Asian/Pacific Island American Elders*. Stanford Geriatric Education Center, Stanford, CA, 1990.

Mungas D, Marshall SC, Weldon M, Haan M, Reed BR: Age and education correction of Mini-Mental State Examination for English and Spanish-speaking elderly. Neurology *46:*700, 1996.

Neighbors HW: Mental health. In *Life in Black America*, JS Jackson, editor. Sage, Newbury Park, CA, 1991.

*Quiroga P, Calvo C, Albala C, Urquidi J, Santos JL, Perez H, Klassen G: Apolipoprotein E polymorphism in elderly Chilean people with Alzheimer's disease. Neuroepidemiology *18:*48, 1999.

Richardson J: *Aging and Health: Black American Elders*. Stanford Geriatric Education Center, Stanford, CA, 1990.

Ruiz DS, editor: *Handbook of Mental Health and Mental Disorder Among Black Americans*. Greenwood, New York, 1990.

*Sakauye KM, Baker FM, Chacko RC, Jimenez RG, Nickens HW, Thompson JW, de Figueiredo JM, Liu W, Manson S: Report of the Task Force on Ethnic Minority Elderly. American Psychiatric Association, Washington, DC, 1992.

Sauvaget C, Tsuji I, Minami Y, Fukao A, Hisamichi S, Asano H, Sato M: Dementia-free life expectancy among elderly Japanese. Gerontology *43:*168, 1997.

*Shepardson LB, Gordon HS, Ibrahim SA, Harper DL, Rosenthal GE: Racial variation in the use of do-not-resuscitate orders. J Gen Intern Med *14:*15, 1999.

Stolley JM, Koenig H: Religion/spirituality and health among elderly African Americans and Hispanics. J Psychosoc Nurs Ment Health Serv *35:*32, 1997.

Unverzagt FW, Hall KS, Torke AM, Rediger JD, Mercado N, Gureje O, Osuntokun BO, Hendrie HC: Effects of age, education and gender on CERAD neuropsychological test performance in an African American sample. Clin Neuropsychol *10:*180, 1996.

Whitfield KE, Seeman TE, Miles TP, Albert MS, Berkman LF, Blazer DG, Rowe JW: Health indices as predictors of cognition among older African Americans: MacArthur studies of successful aging.

Ethn Dis *7:*127, 1997.

Yamamoto J, Rhee S, Chang DS: Psychiatric disorders among elderly Koreans in the United States. Community Ment Health J *30:*17, 1994.

*Yeo G, Gwen Y, Gallagher-Thompson D, editors: *Ethnicity & the Dementias*. Taylor & Francis, Washington, DC, 1996.

51.6 SPECIAL AREAS OF INTEREST

51.6F GENDER ISSUES

MARION ZUCKER GOLDSTEIN, M.D.

The later years of life constitute a considerably longer period of time than childhood, adolescence, young adulthood, or adulthood. Late life can span 40 years or more and start at 50, 55, 60, 62, or 65, depending on various organizational criteria. For example, eligibility to join the American Association of Retired Persons (AARP) starts at age 50, most airline and movie discounts start at age 62, Medicare eligibility starts at 65, eligibility to participate in studies on aging can start as early as 55, and admission to a geriatric unit at 60. Late life has a definitive cutoff (namely death) but not a definitive age of onset or end. Since the term *sex* has primarily biological connotations and *gender* refers to a broad spectrum of biopsychosocial attributes that can cross the stereotyped boundaries between male and female, the term “gender” will be used in this section. The gender ratio at onset of late life is about equal; however, with advancing years, women become an ever larger majority.

Both genders in our aging population are extremely heterogeneous in their life experiences. For example, at the turn of the century in the year 2000, the older population will include persons born around the turn of the previous century as well as those born after World War II. The late lifers will have experienced many societal changes including the rapid advances in information technologies. Adults over 50 are one of the fastest-growing segments of computer and Internet users, making up about 16 percent of the total, of which 42 percent are women and 58 percent men. Gender roles have changed from being primarily defined by men in power and authority to being less stereotyped and more self-determined. Communications, behaviors, and responsibilities within and between genders, families, and generations have changed markedly. Modifications in psychodynamic theory and psychotherapeutic approaches for both genders and for young and old are gradually reflecting these societal changes. Gender-specific life-cycle experiences, adaptations, and pathology during late life demand the same, if not more, psychiatric attention and study than earlier periods in life have had during most of the current century. Psychiatric studies of late life and the impact of earlier life on late life are in an early phase of development. However, because they are relatively recent, they have the benefit of the advances of the past three decades of studies on developmental gender differences in earlier life.

Many in our aging population have come to the United States during various immigration waves. Though unique for each gender, cultural heterogeneity is vast for both, and generalizations from one group to another are as flawed as generalizations from one gender to another or one ethnic group to another. By mandate, all federally funded research must now include both genders, be gender specific, and include ethnic variations. The Accreditation Council of Graduate Medical Education Psychiatric Residency Program Requirements mandate clinical experiences and didactics that include gender and ethnic variations in each age group. An increasing number of studies specific to women's health and health care are being funded. Most Veterans Affairs (VA) studies remain male specific, but universal conclusions can no longer be drawn from such studies. The Health and Human Services Department's Office on Women's Health envisions a nationwide network of integrated health care (including mental health care) for women from cradle to grave.

HISTORY OF DEMOGRAPHIC TRENDS AND THEIR SEQUELAE

In 1990 39 percent of births survived till age 65 and 12 percent to age 80. At the end of the century 86 percent of births are expected to survive to age 65 and 58 percent to age 80. Women constitute 63 percent of the population over 65, with 69 men for every 100 women at age 65 and 36 men for every 100 women at age 85. Elderly men who turn 65 can expect to live another 15 years; elderly women who turn 65 can expect to live an additional 19.4 years. At the turn of the century 67 percent of those age 30 will have a grandmother alive and 27 percent a grandfather. Ten percent of persons age 65 and over have living parents who can be great-great grandparents.

In the 1930s 50 percent of women had four or more children; in the 1980s, 25 percent of women had four or more children. Twenty percent of babyboomers have no children, and 25 percent have only one child. When the 76 million babyboomers born between 1946 and 1964 become senior boomers in 2010 to 2028, the network of family caregivers will have declined markedly in terms of numbers and availability. Primary caregivers continue to be predominantly women, be they wives, daughters, daughters-in-law, or paid service. More often, women in each generation will be in the workforce. Most sites of employment, unless owned and run by women with their own family responsibilities, continue to be unfriendly to family caregiving. Though more grandmothers, great-grandmothers, and great-great-grandmothers may be living, their roles will have changed. They may be physicians, professors, lawyers, corporate executives, or be otherwise employed in the next century, thus preventing the economic deprivations many elderly women experienced in the previous century and continue to experience.

Women make up 72 percent of poor elderly people, with twice as many African-American women experiencing poverty as Caucasian women. Women are half as likely to have pensions, and the amount of their pension is half as much as that of men. When men die they are three times as likely as women to leave an impoverished spouse.

Single, widowed, and divorced women often have to purchase the home care that married men get from their spouses. Women have more disabilities and more often live alone; thus they require the services of nursing homes more often than men. Seventy-five percent of nursing home residents are women age 65 and older. Women are four times more likely than men to use Medicaid for nursing home care because of their relatively indigent status. Because of relatively low Medicaid reimbursements to nursing homes, these women have little choice in the selection of a nursing home and must share a room, their beds can only be held for a limited period of time in case of hospitalization, and medications must be generic. Older men with spouses to look after them at home use more acute care hospital Medicare dollars than women. Medicare does not pay for nursing home care for more than 21 days and only following a hospitalization, it does not pay for prescriptions, and it pays only 50 percent of mental health outpatient care. The lure of paying for prescriptions and lower copay for outpatient mental health care by managed care companies often comes at great sacrifice of continuation of care and availability of specialists. Other health care economizing restrictions such as limited formularies are also common.

One of five illnesses with lowest out-of-pocket costs for treatment are most common among men (e.g., lung cancer); three of four illnesses with highest out-of-pocket cost for treatment are more common among women (e.g., arthritis). More women than men Medicare enrollees lack supplemental insurance policies.

Considerably more men than women have opportunities to participate in policy making as it pertains to Medicaid, Medicare, private insurance, or a combination thereof. As more women enter the research, financial, and political arenas, gender-specific health insurance in both the public and private sector is gradually evolving.

HEALTH AND HEALTH CARE OF WOMEN

Eighty percent of elderly persons 65 and over have at least one chronic condition, especially older women. These conditions most frequently include arthritis, hypertension, sensory impairment, heart disease, urinary incontinence, and depression. The prevalence of self-reported arthritis increases with age. Surveys show higher rates of arthritis-related consequences (e.g., limitation of activities) for women than for men. Higher rates are reported by overweight women, those with less than 11 years education, and those with incomes below $20,000.

Women have smaller hearts and heart vessels, more delayed onset of arteriosclerosis, a higher frequency of glucose intolerance and diabetes mellitus, and more-marked osteoporosis than men. Many studies have shown that gender bias prevails in diagnosis and treatment and inclusion of women in clinical studies and trials, which has contributed to long delays in determining these gender differences. Elderly women have more orthostatic hypotension than men and are at increased

risk for falling and subsequent hip fractures and their distressing sequelae. Postmenopausal hormone replacement therapy is now recommended for women to prevent osteoporosis and also because of the favorable effects of estrogen on serum lipids and brain function.

Research has shown that women with recurrent depression in their younger years are at higher risk for reduction of bone density in their postmenopausal years. The mechanism by which depression is associated with bone loss is unclear but may be related to higher urinary cortisol excretion in these women than in nondepressed controls. Decreased bone density is apparently not associated with taking antidepressants. On the other hand, women who have been treated with anticonvulsants such as valproic acid (Depakene) or carbamazepine (Tegretol) were found to be at high risk for lower bone density and were targeted as candidates for postmenopausal estrogen replacement.

New onset of mania has been described in late life in patients who were started on hormone replacement therapy, and awareness of this potential risk is needed in clinical practice. Estrogen replacement therapy has been shown to improve immediate and delayed verbal memory, maintain the ability to learn new material, and enhance mood in a dose-dependent manner. Estrogen replacement therapy has not been found to have any effect on general attention and visuospatial memory. The National Institutes of Health (NIH)-sponsored Multicenter Women's Health Initiative is well under way and will bring more scientific understanding to diagnosis and treatment of elderly women in terms of risk factors and prevention of osteoporosis, cardiovascular disease, and dementia. Because of the increased risk of breast cancer, regular mammograms are needed during the course of hormone replacement therapy. The increase in fibrocystic tissue during hormone replacement therapy makes mammograms by the current state-of-the art precision more difficult to read. Patients need to be warned of the risk they take, and insurance companies need to reimburse for every-half-year mammograms as medically indicated instead of insisting on the annual follow-up. This serious risk factor should limit any research with estrogen replacement therapy that does not include responsible, long-term follow-up of the participants. Because of the increased risk of uterine cancer for most women who have not had a hysterectomy, progesterone maintenance must be added. Synthetic progesterone dampens mood. Though risks and benefits of hormone replacement therapy are under ongoing investigation, the extensive use in clinical practice has surpassed the current knowledge base. The salutary effect of hormone replacement therapy on hot flashes, dry skin and mucous membranes, insomnia, and incontinence cannot be minimized, nor can the current epidemic of breast cancer. The optimal age for starting hormone replacement therapy has not yet been determined and depends on the conditions to be targeted and risks involved. Clinical guidelines for counseling postmenopausal women about preventive hormone replacement have been issued by the American College of Physicians and the American College of Obstetricians and Gynecologists and need to be under ongoing review.

DEPRESSION

The results of the Epidemiologic Catchment Area (ECA) Survey conducted in five different communities in North America have contributed to our knowledge of the gender differences in incidence and prevalence of mental disorders. All forms of depression appeared to occur approximately twice as often in women as in men throughout life. The more frequent occurrence of isolation, poverty, single status, lack of education, and poor health in older women must be considered as contributing to late-life depression. The prevailing underrecognition of depression in primary care practices is detrimental to the quality of life of more elderly women than men. This is compounded by the fact that research has shown that 50 percent of caregivers, most of whom are women, have symptoms of clinical depression and a higher rate of doctor visits than a matched group of noncaregivers. Older women with chronic nonlife-threatening painful conditions seek medical attention more frequently than older men. The range of symptoms with which elderly women exhibit their depression is a challenge to clinicians. Physical symptoms without pathological correlates are a common mask for depression. These masked depressions are more frequently seen in primary care practice than the criteria for depression listed in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). Though primary depressive disorders and secondary depressive disorder are listed in DSM-IV, presentation of a medical symptom with no pathological correlate is not mentioned. Research is investigating whether there is an association between depression, immune function, physical illness, and lack of estrogen in older women.

Older women suffer more from the adverse effects of their medications, which often include benzodiazepines and unnecessary polypharmacy. Ageism and sexism contribute to these practices. Gender-specific issues in patient-physician communications contribute to diagnostic and treatment decisions and need considerably more attention in outcome studies. Hormone replacement therapy and antidepressants together have been found effective in alleviating severe depression. This treatment modality needs further consideration for depressed postmenopausal women, with inclusion of contributions of estrogen to cognitive improvement as well as major attention to risk factors. Treatment of depression can improve cognitive functions; therefore, research on whether estrogen does so as well creates considerable ethical questions. Administering progesterone to the majority of postmenapausal women who have not had a hysterectomy creates an additional research challenge.

The incidence of suicide in the elderly is related to both the nature of the depression and gender. Divorced or widowed men over the age of 65 are four times more likely to commit suicide than other community-dwelling persons, including the younger population. The incidence of death in older men the year following divorce or widowhood is considerably higher than that of women. No evidence indicates that the death rate of women rises after they lose a partner, whereas the mortality rate in women increased after they lost a son in the war. Altogether, men accounted for 81 percent of suicides of persons age 65 and older in 1992, with considerable ethnic variation within this group. The suicide rate for men age 80 to 84 rose from 43.5 per 100,000 to 58.6 per 100,000 and for women from 4.7 per 100,000 to 6.4 per 100,000 between 1980 and 1992. One of the national health objectives for the year 2000 is to reduce the suicide rate of white men age 65 and over, who constitute the highest percentage. Preventive measures include peer-counseling programs, suicide prevention centers, and programs to increase awareness of risk factors. Bereavement has been shown to be a risk factor for increased alcohol consumption, more so in men than in women. Contributing risk factors for this behavior include a past history of major depression, a depressive episode beyond normal bereavement following a loss, high alcohol consumption prior to bereavement, and dissatisfaction with social supports following a loss. Women experience more anxiety than depression following bereavement. A form of separation anxiety manifested by feelings of panic, yearning, pining, sighing, restlessness, insomnia, shortness of breath, palpitations, and trembling has been observed 1 year after a death of a loved one in women with inadequate supports. It seems that many men and many women experience diminished quality of life with longevity, but not all. For example, the rare exceptions are those centenarian survivors who come to the attention of the media because of their sustained well-being. Each decade of elderly women has unique needs. One of the main problems remains inadequate resources, especially lack of transportation.

EVOLUTION OF GENDER-SPECIFIC DEVELOPMENTAL THEORIES

From 1932 to approximately 1972 psychoanalytic theory dictated the universality of male development. Girls and women were perceived as anatomically, morally, and emotionally deficient in male attributes, and their role in the life cycle of men was dedicated to enhancing expression of these male attributes for men. The qualities of adulthood, namely the capacity to think autonomously, make clear decisions, and take responsible actions, were associated with masculinity. The splitting of love and work *(lieben und arbeiten)* was out of balance between the genders, creating many injustices that became justified by this theoretical bias. Cumulative imbalance of economic relations between men and women prevailed throughout the life cycle, with women's work invisible, devalued, and poorly paid or unpaid. These prevailing norms went unquestioned and unchallenged. Men's relationships were (and are) subordinated to rules with male principles and concepts of justice. The emphasis on separation, individuation, hierarchy of authority, competition, and mentoring of other men superseded networking and collaboration between and within each gender with equal authority and reward. Like the situation observed in monkey colonies, individuals on top of the hierarchy have the liberty to perpetrate unretaliated aggression, and the subordinates in the hierarchy are more vulnerable, are at risk when they retaliate, and have fewer rights. These societal value systems have taken their toll on the current cohort of women in late life. These value systems and sexual stereotyping prevailed in the teachings of psychotherapy as long as theory blinded observation. For example, most women who wanted a role in the medical psychoanalytic establishment where these theories evolved and became firmly grounded joined the male ranks at lower levels than men and had to adopt the beliefs of the male establishment. Professional women who did not conform were asked to leave. For example, when Karen Horney evolved a feminist frame of reference of psychoanalytic thinking and later developmental theories for each gender, she had to separate from the male-dominated establishment and found the American Institute for Psychoanalysis. The evolution of gender-specific developmental theories empowered women of all ages. For example, Maggie Kuhn, founder of the Gray Panthers with six other women who had been forced to retire at the age of 65, pressed for innovative work concepts and intergenerational networking; health care systems to protect everyone, old, young, families, and children; and elimination of homelessness.

Among those who made major contributions to gender-specific observation and put them in a positive light for women in particular were Nancy Chodorow, Carol Gilligan, Jean Baker Miller, Alexandra Symonds, Matina Horner, and I. Broverman. The greater capacity for empathy, compassion, intimacy, and attachment in women, which lead to different priorities in decision making than those of most men, have been examined to enhance their appreciation and reward. The fear of adverse consequences of success, such as social rejection and loss of feminine attributes while entering into a male-dominated arena, was studied and found to be greatly ameliorated by support from other women, regardless of age. Mentoring and role modeling by older women in leadership roles has become desirable as has companionship between women throughout the life cycle, which is no longer considered second best. More men are aware that women's productivity and leadership can take place both inside and outside the home if men participate in both on a more egalitarian basis. These changed behaviors can reflect in their aging years. Only a few in the present cohort of elderly men can make such adaptations. The current generation of grandparents at times needs help to adapt to these changes, especially when intergenerational conflicts occur.

AGE AND GENDER-SPECIFIC PSYCHOTHERAPY

Challenges to the still prevailing ageist and sexist stereotyping in psychotherapy continue. Patients and many therapists are likely to be encumbered by sexist and ageist stereotypes. Renewal and generativity can be fostered at any age as long as it is appropriate to gender, mental, sensory, educational, and financial capacities and essential resources are made available.

Therapists must pay special attention to their own fears of aging, disabilities, losses, and death so they do not avoid these vital concerns during the aging process of their patients and can assist in optimal adaptation. Ageist and stereotyped views of elderly adults as shipwrecks at the end of their voyage—too slow, unproductive, and close to death—denies the many who are capable of ongoing growth, development, adaptation, friendships, and enjoyment the opportunity to get professional help in a respectful, reality-oriented, compassionate, and knowledgeable manner.

Elderly patients may need to adapt to greater dependency than they had earlier in life. Older women who were either independent during their adult lives or in a nurturant position that they equated with their worth, usually find dependency on others especially difficult. Excessive resistance to dependency can result in self-injurious or self-neglectful behaviors. Not wanting to be a burden to others (especially not one's own children) is a common concern, especially of older women, and can often be successfully addressed in psychotherapy by pointing out the vital contributions they have made to their family during their lifetimes. Unrealistic and excessive feelings of entitlement or dependency on others beyond actual need and capacity may also need to be addressed. The individual etiologies and meanings of such behaviors need to be explored and clarified to the patient in a gender-specific manner.

Family participation is a vital component of psychotherapy of dependent elderly persons. Again, more women in the advanced years live alone, have more chronic illness, and may need assistance from adult children. The readiness for emotional, physical, and at times financial committment to this parental caretaking task can vary widely among adult children in a family. The skilled and sensitive psychotherapist can assess and effectively distribute tasks and help families function as a quasi-multidisciplinary team while accessing appropriate community resources. By not assuming that the women in the family will take on all the household and caretaking responsibilities, gender barriers to participating in caregiving can at times be overcome. Medicare denials of reimbursement for these professional services is still all too common. Family involvement in medication management and compliance as well as many other activities of daily living for the elderly patient remains a professional standard in the therapeutic interventions for the dependent, frail elderly patient.

The developmental phase of being a grandparent needs attention in theory, research, teaching, clinical practice, and policy making. The stereotyped image of the grandmother as an unemployed or retired, babysitting, meal-preparing, cookie-baking, and ever-available individual becomes less and less of a reality. No such expectations are usually made of the grandfather, whose interest and assistance is often appreciated at whatever level he chooses. Grandfathers, when retired, may enjoy the new role of helper and caregiver. A therapist's role in a grandparent's life can include resolution of intergenerational conflicts and of facilitation of development of this new and challenging identity. Grandmothers who live alone and at a distance are less likely to have the financial resources to visit than grandfathers or both together. Clinical observations have revealed that depriving older women of the opportunity for regular visits with adult children and grandchildren can contribute to depression and considerable distress. Other clinical observations reveal that older men who have survived their alcoholism during younger years tend to get depressed when they are rejected by adult children who cannot forgive their abusive behaviors from earlier years. This rejection inadvertently leads to deprivation of the grandfather role. Gender-specific developmental theory for later life has not yet been established, though various modalities of psychotherapy including cognitive-behavioral approaches have been found quite effective for both men and women in late life.

GENDER- AND AGE-SENSITIVE MEDICATION MANAGEMENT

Medication management must be gender and age sensitive. Gender-related differences in efficacy safety, and dose recommendations have not been adequately researched. Physiological (e.g., 36 to 48 percent increase in adipose tissue in women as they age) and anatomical (e.g., organ size) differences between elderly women and men affect response to antidepressant, antipsychotic, and anxiolytic drugs. Research on gender differences in pharmacokinetics and pharmacodynamics is just beginning, and comparison data between men and women not uniformly available.

The relative lack of gender-sensitive drug studies in the elderly population affects more women than men. The 1993 Food and Drug Administration (FDA) recommendations include the need for data analysis by age, sex, and race; body weight; renal, cardiac, and hepatic status; concomitant illnesses and use of other drugs. The FDA guidelines support informed individualization of treatment while taking into account the needs of children, elderly adults, and each sex. Currently, drug studies with elderly patients usually take place after (rather than before) FDA approval.

MENTAL DISORDERS OTHER THAN DEPRESSION AND DEMENTIA

Except for sexual and gender identity disorders and conditions associated with the reproductive system, DSM-IV does not relate manifestations of disease to gender because of lack of information based on adequate research. Women are at higher risk than men for most disorders at all ages except alcohol and other substance use disorders (other than iatrogenic). Other parts of this section review diagnosis and treatment of psychopathology in the older population.

Anxiety disorders, such as panic disorder, phobias (especially agoraphobia), and obsessive-compulsive disorder, and somatoform disorders occur in more than 4 percent of the elderly community population. There is a higher incidence of agoraphobia in older women than men. Whether there is an association between the higher incidence of victimization of women and the higher incidence of disabilities in the advanced years has not been researched.

Excessive consumption of alcohol by elderly adults is a leading public health problem that originates from and leads to stressful life situations. A man's alcoholism often creates the problem for the woman he lives with. The ECA study revealed a 1-year prevalence of alcohol use disorder in persons age 65 and older of 3.1 percent in men and 0.4 percent in women. Though the prevalence in the older population is considerably lower than that in younger populations, alcohol-related hospitalizations are more frequent in older persons. The tolerance for alcohol diminishes with age, especially in women. Lower consumption than in younger persons can cause severe physical and psychosocial complications including cognitive deficits. Survey data reveal that alcoholism is frequently overlooked clinically in older women because it is not sought. Once detected and diagnosed, few age- and gender-appropriate treatment opportunities are readily available.

Sex and gender differences in psychosis (schizophrenia in particular) and population (bipolar I disorder in the elderly) require research attention in terms of differences in etiology, manifestations, course of illness, biopsychosocial consequences, and risks and benefits of treatment. More women than men have late-life onset of schizophrenia. No data are available to indicate that estrogen may have a protective function earlier in life or that it is effective in the treatment of psychosis—again, an ethical question in view of the advances in psychotropic medications. Women with psychotic disorders are at high risk of receiving less than optimal preventive care for disease. No studies of antipsychotic medication with and without hormone replacement therapy or hormone replacement therapy alone during late-life psychosis are available. Ethical and legal considerations may well delay such studies indefinitely. The fact that women are more apt to get tardive dyskinesia makes a gradual crossover to adequate dosages of the newer serotonin–dopamine antagonists such as risperidone (Risperdal) and olanzapine (Zyprexa) even more urgent than it is for elderly men.

FUTURE DIRECTIONS

Gender issues in mental health and aging require basic and health services research. Women's issues and men's issues have taken separate paths, and women's issues are lagging behind. For example, there is a biologic marker for prostate cancer but no reliable marker for breast cancer. Women can benefit from men's studies and men from women's studies, but women have much catching up to do in research and clinical service. Estrogen is not the panacea for all aging problems of women, though many can be symptomatically relieved and possibly prevented. A high research priority is to have women of all ages who take or have taken hormone replacement with varying frequency and doses and their medical care providers participate in a longitudinal epidemiological outcome study with minimal exclusion criteria. The group would include women on estrogen replacement therapy and hormone replacement therapy, with psychosis, depression, dementia, diminished bone density, breast cancer, uterine cancer, heart attacks and strokes, and hip fractures taking vitamin and mineral supplements and those who do not, and on high-fat, low-fat and other diets. The research population would be large enough and be studied long enough with regular assessment for data analysis to determine what women benefit substantially from replacement therapy and what women are at risk. Is this a gender issue or a women's issue? It is a societal issue that urgently needs attention. Practice patterns of care providers, their decision trees, and gender variations in communication with their patients can be studied simultaneously. Subgroups can be studied for immune and stress status with biologic markers and others for genetic vulnerabilities. In the era of information technologies a 20-year lag between research findings and practice is no longer acceptable. The busy practitioners who must see more and more patients in a shorter time and the patients who take time out of their busy day must be appropriately reimbursed for time spent in this research, regardless of gender so the research does not become yet

another gender issue.

Fragmentation of medical services must be stopped. Health care for both aging men and women, including mental health care, needs to be locally centralized and gender- and culture-sensitive. Centralization of all health care services would greatly facilitate clinical research in the aging population. Can women lead the way? Yes, of course, they may even be more motivated to do so than men if given the appropriate resources.

The effects of earlier life development, losses, and other traumas on late-life adaptations, coping mechanisms, and physiological and pathological responses require gender-based life-cycle studies. Erik Erikson and George Vaillant have been the pioneers. Appropriate application of information technologies will permit study of larger numbers in many subgroups to learn which past experiences enhance and which hinder physiological, interpersonal, and personal adaptation and coping in late life. Life-cycle studies of each gender should be simultaneous, not sequential.

SUGGESTED CROSS-REFERENCES

Section 2.6 discusses neuropsychiatric aspects of movement disorders; Section 3.1, perception and cognition; Section 3.4, biology of memory. Section 51.6g deals with elder abuse. Section 5.4 discusses animal research and its relevance to psychiatry, Chapter 6.3 discusses theories of personality and psychopathology and other psychodynamic schools, Chapter 12 discusses schizophrenia; Chapter 13, other psychotic disorders; Chapter 14, mood disorders; Chapter 15, anxiety disorders; Chapter 25, psychological factors affecting medical conditions; Chapter 28, psychiatry and other specialties. Section 29.1 deals with suicide; Chapter 30 with psychotherapies, and Chapter 31 with biological therapies. Dementia in the elderly and gender issues are covered in Section 51.3e and Section 51.6f; Section 51.3h is on drug and alcohol abuse.

SECTION REFERENCES

Ahern KD: *The Older Woman the Able Self*. Garland Publishing, New York, 1996.

Baker Miller J: *Psychoanalysis and Women: Eminent Psychoanalysists Dispel Myths and Explore Realities*. Brunner/Mazel, New York, 1973.

Broverman I, Vogel S, Broverman D, Clarkson F, Rosenkrantz P: Sex role stereo types: A current appraisal. J Soc Issues *28:*59, 1972.

*Centers of Disease Control and Prevention: Suicide among older persons. JAMA *275:*59, 1996.

Chodorow WN: *Culture and Society*, MZ Rosaldo, Lamphere L, editors. Stanford University Press, Stanford, CA, 1978.

Fletcher CV, Acosta EP, Strykowski JM: Gender differences in human pharmacokinetics and pharmacodynamics. J Adolesc Health *15:*619, 1994.

Friedan B: *The Fountain of Aging*. Simon & Schuster, New York, 1993.

Gallo JJ, Anthony JC, Muthen BO: Age differences in the symptoms of depression: A latent trait analysis. J Gerontol *49:*251, 1994.

Ganzini L, Atkinson RM: Substance abuse. In *Comprehensive Review of Geriatric Psychiatry*, ed 2, J Sadavoy, LW Lazarus, LF Jarvik, GT Grossberg, editors. American Psychiatric Press, Washington, DC, 1996.

*George LK, Gwyther IP: Caregiver well-being, multidimensional examination of family caregivers of demented adults. Gerontologist *26:*253, 1986.

Gilligan C: *In a Different Voice: Psychological Theory and Women's Development*. Howard University Press, Cambridge, MA, 1982.

Goldstein MZ: Mental health concerns in the advanced years. In *Textbook of Women's Health*, L Wallis, editor. Lippincott-Raven, Philadelphia, 1998.

Goldstein MZ, Griswold K: Gender sensitivity in health care for elderly men and women. Psychiatr Serv *49:*323, 1998.

*Grady D, Cummings DR, Petitti D, Rubin S, Audet AM: American College of Physicians guidelines for counseling postmenopausal women and about preventive hormone therapy. Ann Intern Med *11:*1038, 1992.

Harris HW, Lebowitz BD: Clinically oriented basic science: Emerging opportunities for research in late-life mental disorders. Am J Geriatr Psychiatry *5:*284, 1997.

Horney K: *New Ways in Psychoanalysis*. Norton, New York, 1936.

Journal of Women and Aging. The multidisciplinary quarterly of psychosocial practice, theory and research, D Garner, SO Mercer, editors. Hawthorn Press, Binghamton, NY, 1989.

Kelly K: Building aging programs with online information technology. Generations *21:*15, 1997.

Lee GR, Willets MC, Seccombe K: Widowhood and depression-gender differences. Res Aging *20:*611, 1998.

*Lindamer LA, Lohr JB, Harris MJ: Gender related clinical differences in older patients with schizophrenia. J Clin Psychiatry *60:*61, 1999.

McIntyre S, Hunt K, Sweeting H: Gender differences in health: Are things really as simple as they seem. Soc Sci Med *42:*617, 1996.

Michelson D, Statokis C, Hill L, Reynolds J, Gallivent F, Chrousos G, Gold P: Bone mineral density in women with depression. N Engl J Med *335:*1178, 1996.

Miles S, Parker K: Men, women and health insurance. N Engl J Med *336:*218, 1997.

Pravder MM: *Women In Context Towards a Feminist Reconstruction of Psychotherapy*. Guilford, New York, 1994.

Ray RE: A postmodern perspective on feminist gerontology. Gerontologist *36:*674, 1996.

Reference and Research Sources: Contemporary social issues. Women and aging, no. 35, a bibliography. Santa Cruz, CA, 1994.

Roughan PA: *Mental Health and Psychiatric Disorders in Older Women: Clinics in Geriatric Medicine Care of the Older Woman*, Kaiser FE, guest editor. Saunders, Philadelphia, 1993.

*Schneider LS, Small GW, Hamilton SH: Estrogen replacement and response to fluoxetine in a multicenter geriatric depression trial. Am J Geriatr Psychiatry *5:*97, 1997.

*Sherwin B: Menopause, early aging and elderly women. In *Psychopharmacology and Women: Sex, Gender, and Hormones*. MF Jensvold, U Halbreich, JA Hamilton, editors. American Psychiatric Press, Washington, DC, 1996.

Tan M, Brausgrove L: Management of the depressed elderly woman. Clin Geriatr *52:*56, 1997.

Turner B, Troll L: *Women Growing Older: Psychological Perspectives*. Sage, Thousand Oaks, CA, 1994.

Verbrugge LM: Women, men and osteoarthritis. Arthritis Care Res *8:*212, 1995.

*Wolinsky FD, Wyrwich KW, Gurney JG: Gender differences in the sequelae of hospitalization for acute myocardial infarct among older adults. J Am Geriatr Soc *47:*151, 1999.

Zisook S, Mulhill M, Schuchter SR: Widowhood and anxiety. Psychiatr Med *8:*99, 1990.

51.6 SPECIAL AREAS OF INTEREST

51.6G ELDER ABUSE, NEGLECT, AND EXPLOITATION

MARION ZUCKER GOLDSTEIN, M.D.

The American Psychiatric Association (APA) has published a Position Statement on Elder Abuse, Neglect and Exploitation as recommended by the APA Council on Aging and approved by the APA Assembly and Board of Trustees in 1995. The content of this position paper is reflected in the following text and in Table 51.6g-1 and Table 51.6g-2. “Elder abuse, neglect, and exploitation have been identified as major public health problems. All 50 states have mandatory or voluntary reporting laws; however, problems remain.” Varying laws require specific professionals who have reason to suspect elder abuse or neglect to report their observations to designated public health agencies. Failure to report is considered a criminal offense in 42 states and the District of Columbia. Colorado, Illinois, New Jersey, New York, North Dakota, Pennsylvania, South Dakota and Wisconsin make reporting voluntary. Criminal enforcement of these laws which differ in nature from state to state is nonexistent. Clinical incentives to reporting, such as a comprehensive system of services to respond to such reports, providing appropriate follow-up services to victims and overstressed caregivers, and providing reporters with outcome of interventions, are lacking in most jurisdictions. This situation and the relative isolation of most abused elderly from society, have contributed to making mandatory reporting laws ineffective.

Emotional Abuses	Psychiatric Symptoms*
Threats	Resignation
Insults	Ambivalence
Harassment	Fear
Harsh orders	Anger
Infantilization	Cognitive impairment
Restrictions of social activities	Depressed mood
Restrictions of religious activities	Insomnia
Financial exploitation	Substance abuse
Physical abuse	Delirium
Excess stress	Agitation
	Lethargy
	Self-neglect

* Any or several of these symptoms can occur with any or several of these emotional abuses. Adapted from American Psychiatry Association Council on Aging: Position statement on elder abuse, neglect, and exploitation. Am J Psychiatry 152:5, 1995.

Table 51.6g-1 Elder Abuse, Neglect, and Exploitation: Emotional Abuses and Psychiatric Symptoms

- **Participate** in the formulation of protocols for the identification of elder abuse and exploitation.
- **Research** applicable laws concerning reporting of elder abuse.
- **Utilize** available resources in the APA Central Office and state societies to help in assessment and reporting of elder abuse and in interventions.
- **Collaborate** with government and private agencies that support the prevention, identification, and treatment of elder abuse.
- **Participate** actively in multidisciplinary research efforts on the incidence and prevalence of elder abuse and on outcome studies of specific interventions.
- **Develop** curricula for medical schools, residency training programs, and continuing medical education that teach the physician's role in the prevention, identification, and treatment of elder abuse.

Reprinted with permission from American Psychiatric Association Council on Aging: Position paper on elder abuse, neglect, and exploitation. Am J Psychiatry 152:5, 1995.

Table 51.6g-2 Elder Abuse, Neglect, and Exploitation: APA Recommendations for Psychiatrists

Emotional abuses and psychiatric symptoms seen in abused elderly persons, as listed in the APA position paper, are reflected in Table 51.6g-1.

“The APA recommends that a comprehensive biopsychosocial assessment of the victimized elderly person and the perpetrator of the victimization be completed to facilitate effective interventions. These interventions should include utilization of legal, social, and financial resources.” APA encourages psychiatrists to participate in the activities outlined in Table 51.6g-2.

The presenting symptoms of elder abuse are some of the manifestations listed in the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) under a variety of diagnostic categories. Elder abuse as such is not listed in DSM-IV as a possible diagnosis for the symptom complexes noted here. A code for elderly victims of abuse, separate from that for children and adults, is necessary. The feeling of a foreshortened future as a consequence of trauma among elderly patients requires different interventions than those for children, adolescents, and adults. Child, partner, and elder maltreatment need to be clearly distinguished in domains such as prevention, detection, intervention, diagnosis, treatment, and legislative and judicial arenas. For example, the child abuse model for reporting elder abuse reflects the societal stereotype of infantilizing older persons and restricting self-determination and control of their own lives to the limit of capacity. Whereas child abuse was initially brought to public attention by physicians, elder abuse was brought to public and congressional attention by social workers and nurses. In 1996 3000 members of the American College of Emergency Physicians were surveyed about their practice characteristics, availability of elder maltreatment protocols, and their familiarity with local laws and reporting requirements. The results of the 24 percent who responded suggested that practicing emergency physicians are not confident in identifying or reporting the geriatric victims of abuse or neglect. Inadequacies in training, research, and continuing education with regard to mistreatment of older people were implied.

However, the APA Committee on Family Violence under the Council on Children and Adolescents did see to it that elder maltreatment was included in their Guide on Family Violence. The American Medical Association (AMA) has published diagnostic and treatment guidelines specific to elder abuse. The commitment of mainstream medicine to contribute to finding humane solutions to this increasing problem is growing steadily. Elder abuse should be included in medical school curricula, in postgraduate medical training, in Accreditation Council of Continuing Medical Education (ACCME) accredited forums of medical education, and in medical publications. Medical research in collaboration with social scientists and research nurses will enrich and broaden the knowledge base.

In 1991 a small amount of federal funds was specifically designated for the prevention and intervention of elder abuse and neglect in an amendment to the reauthorization of the 1965 Older American Act. These funds are subject to approval of annual renewal. The National Resource Center on Elder Abuse (NRCEA) was operated by a consortium of the American Public Welfare Association (now the American Public Human Services Association [APHSA]), the National Association of State Units on Aging (NASUA), the University of Delaware, and the National Committee for the Prevention of Elder Abuse (NCPEA) with funding from the Administration on Aging (AOA). This consortium has contributed to research in the epidemiology of elder abuse, neglect, and exploitation.

In 1998 the Department of Health and Human Services (DHHS) announced a new National Center on Elder Abuse (NCEA) for the promotion of understanding among

states and local networks of community workers, physicians, elder volunteers and others working to prevent elder abuse. The center is led by the National Association of State Units on Aging and is funded by a $1 million DHHS grant to expand its support of training activities of persons who work with older persons on a regular basis as well as medical and social service professionals to identify and respond appropriately to cases of elder abuse and neglect. The NCEA consists of a partnership of the National Association of State Units on Aging (NASUA), the American Bar Association (ABA) Commission on Legal Problems of the Elderly, Clearing House on Abuse and Neglect of the Elderly (CANE), the Goldman Institute on Aging San Francisco Consortium for Elder Abuse (GIOA), the National Association of Adult Protective Services Administrators (NAAPSA), and the National Committee for the Prevention of Elder Abuse (NCPEA). The NCEA has contributed to the 1996 National Elder Abuse Incidence Study (NEAIS) of domestic elder (age 60 and over) abuse, neglect, and self-neglect with a final report in 1998. This report confirmed the “iceberg theory” of the past 20 years, that for every reported and substantiated incident, at least five still go unreported.

EPIDEMIOLOGY

The true prevalence or incidence of elder abuse is not known because of underreporting and variability of reporting methods, instruments, sites, and reporters. The National Institute on Aging supported a survey of a random sample of 2020 elderly persons in the Boston community in 1985. While this study was being conducted the reported elder abuse prevalence of 32 per 1000 and incidence of 26 per 1000 was found. Nearly three fifths of the perpetrators were spouses, one fifth were adult children, and one fifth were other persons such as siblings, grandchildren, and boarders. Among the abusive adult children, sons outnumbered daughters 2 to 1.

During the same period of time as the study was conducted, the reported incidence of elder abuse in Massachusetts was 1.8 per 1000, indicating that only 1 in 14 cases came to public attention. This discrepency between survey results and reporting occurred in a state with a law that required (with penalty of $25–$1000 for noncompliance) immediate reporting of all physical abuse of persons age 60 and older.

A telephone interview study of 577 staff members in 31 long-term care facilities revealed a substantial rate of abusive behaviors among staff who also reported job dissatisfaction and the perception of elder persons as children. Surveys of elder maltreatment have been undertaken in many countries including Canada, the United Kingdom, Australia, the Scandinavian countries, France, Israel, South Africa, and others. Persons reporting include community dwellers, home health visitors, family caregivers, hospital patients and their families, and retrospective reviewers of maltreatment. International and cross-cultural perspectives have been described. National incidence and prevalence of institutional abuse is particularly difficult to determine. Ombudsmen (volunteers who visit nursing homes at least weekly) have been organized in many areas in each state by federal mandate since 1978. These ombudsmen report complaints by nursing home residents and their families and report their own observations about residents' care to state agencies. However, underdetection and underreporting continue. Ombudsmen are mandated to resolve only short-term situations. Overcrowding, understaffing, unappetizing dining rooms with one aide feeding five people at the same time, screams coming out of closed bathrooms, and insufficiently trained and supervised staff have not been targets for reports by ombudsmen. A 1990 study in six states of a total of 14,018 complaints divided in eight categories revealed a distribution as outlined in Table 51.6g-3.

Category	Percentage
Physical abuse, inadequate hygiene, neglect	42.6
Understaffing, discharge of Medicaid patients	18.1
Theft and loss of personal items, indignities	11.0
Unappetizing food, lack of assistance at meals	8.7
Lack of cleanliness, suitable ventilation, and temperature	7.3
Questionable charges and misuse of personal funds	6.5
Medications not given as ordered, overmedication	3.6
Physician services inaccessible for diagnosis and treatment	2.1

Table 51.6g-3 Categories of Abuses in Long-Term Care as Reported by Ombudsmen

Researchers also addressed elder abuse and neglect in the community. In 1991 NRCEA established reporting methods to and from state adult protective service (APS) agencies and state units on aging across the nation. NRCEA's analysis of existing state and federal definitions of elder abuse, neglect, and exploitation led to concise definitions of seven types of elder abuse: physical, sexual, emotional, and financial exploitation; and neglect, abandonment, and self-neglect. Once these reporting methods were implemented, the data collected revealed reporting of domestic elder abuse to have increased by 106 percent between 1986 and 1994 (117,000 to 241,000). Whereas the total increase in elder population from 1986 to 1996 was 10 percent, the NEAIS revealed an increase of 150 percent (117,000 in 1986 to 293,000 in 1996) of unduplicated Adult Protective Services–reported domestic abuse or neglect incidents during these 10 years. Whether new occurrences or reporting was higher remains to be researched.

Reporters of domestic elder abuse in the NRCEA study were: 21.6 percent physicians or other health care providers, 9.4 percent APS or other agency workers, and 14.9 percent family members and relatives; the remainder were friends and neighbors, law enforcement personnel, clergy, banks and business institutions, elder victims themselves, and others. Agencies that received and investigated reports were state and local APS, social service, and law enforcement agencies, police, and sheriff departments.

In 1994 physical abuse accounted for 15.7 percent, emotional abuse for 7.3 percent, financial and material exploitation for 12.3 percent, neglect for 58.5 percent, and unknown or other types of abuse for the remainder of the total abuse of elders. The median age for self-neglecting elders was 77.2 years and 76.5 years for the remainder of substantiated reports. Sixty-two percent were women, 38 percent were men; 65.4 percent of victims were white, 21.4 percent were black, 9.6 percent were Hispanic, and less than 1 percent were Native-Americans, Asian-American, or Pacific-Islanders. Characteristics of perpetrators were as follows for 1994: 50.6 percent were men and 49.3 percent were women; adult children were the most frequent abusers, spouses ranked next, and other family members were next in frequency.

ABUSED ELDERLY

The profile of an elderly person who is at high risk for abuse is complex and varied. This individual often lacks the opportunity or the physical and mental ability to report the abuse. When the perpetrators of abuses are the adult children, the elderly victims may desire to protect their children, be embarrassed about the situation, and even blame themselves. Common fears of disbelief, reprisal, abandonment, and institutionalization must be recognized and addressed by the professional provider of care who suspects victimization. Psychiatrists working in emergency rooms, acute-care inpatient units, consultation or liason services in general hospitals, and long–term care facilities need to question the possibility that they may be denying, minimizing, or rationalizing the occurrence of the abuse. Being alert to these findings can assist the professional provider of care to prevent recurrences by appropriate interventions in individual situations.

The following case vignettes demonstrate the complex situations reported to a local APS agency, how needs are being met by the APS, and how psychiatric assessment and treatment might have contributed to each situation had this been called upon or been available in a timely and expert fashion with participation of a multidisciplinary team.

Victim: Father (H.G.) 76, diagnosed with Parkinson's disease

Perpetrator: Son (J.G.) 30, diagnosed with schizophrenia

With H.G.'s consent, given with much encouragement from his neurologist, the neurologist was able to report abuse by H.G.'s son J.G., abuse that H.G. had been telling him about. J.G., who had been diagnosed with schizophrenia, had never left the parental home; his mother had died 2 years ago. The neurologist described H.G. as an intelligent and highly educated man whose Parkinson's disease had been exacerbated by the stress of household responsibilities and his son's threatening and uncontrolled behaviors. During a home visit, the APS worker observed that H.G. had been confined to two rooms by his son. APS recommended involuntary hospitalization for J.G. and an order of protection to prevent J.G. from returning home. He felt entitled to “possession” of the house but was unable to look after his own or his father's needs. No further assessment of his condition was described in their report.

J.G.'s phone contacts with his father were threatening both during and after discharge. Upon his return to the community, J.G. did not follow up on psychiatric outpatient treatment and lived in a motel with financial assistance from family members. H.G. sold his home and moved to a retirement home, where he reported feeling safe and contented. In this situation, involvement of mental health professionals would have been helpful for father and son. A more thorough history should have been obtained upon initial APS involvement. The current situation had arisen 2 years prior, after H.G.'s wife's death. She had been her son's advocate in the household and had educated herself on schizophrenia. According to family members, H.G. had disparaged J.G., favored J.G.'s older brother, and denied the diagnosis of schizophrenia. J.G. had been noncompliant with medication treatment because of extrapyramidal adverse effects. J.G. had never received disability income or insurance coverage.

A multidisciplinary mental health team could have provided evaluation and a treatment plan that could have included ongoing case management (with assistance with finances, insurance, and living situation) and medication management. The team could have offered education and support for H.G., as well as referral to community agencies such as the National Alliance for the Mentally Ill (NAMI), which his wife had participated in actively. Improved outcome could have included continuity of psychiatric treatment for J.G., maintenance of H.G. in the community, and belated understanding of his son's condition.

Victim: Wife (E.A.), 75, diagnosed with dementia

Perpetrator: Husband (C.A.), 78, diagnosed with bipolar I disorder

E.A. had been married to C.A. for 50 years. The couple had remained childless, outlived their siblings, and not maintained contact with nieces or nephews. E.A. had progressively become less able to take care of the household or her own physical care. C.A. had bipolar I disorder with grandiosity, delusions, excessive spending or giving money away, bouts of infidelity, and alcohol abuse. He had ignored his wife's needs during most of their married life but claimed “to love her.” He was now quite unable to deal with her ever-increasing dependency or to provide her with a safe home environment.

APS made attempts to provide personal care aids and nursing services for E.A. and even obtained guardianship of her person in order to make medical decisions on her behalf. One of the personal care aids alleged that C.A. had raped his disabled wife, a not unlikely event considering the hypersexuality during the acute phases of his now untreated bipolar I disorder. APS recommended involuntary psychiatric hospitalization of C.A. and hospitalization of E.A. for a delirium and dementia workup and nursing home placement. APS reported that the primary care physician would not “cooperate” with E.A.'s hospitalization, objecting to the potential expense of a prolonged hospitalization on an acute care unit. E.A. had to be admitted on an emergency basis. She died 2 weeks later in the hospital of causes that were not in the APS report. Following E.A.'s death, APS applied for guardianship of C.A.'s income and assests and transported him for psychiatric evaluation. He was hospitalized for several months and discharged to a supervised residence in the community. APS makes monthly visits and found him adapted to his surroundings but depressed. There was no further mention of the psychiatric management of his bipolar I disorder. There was speculation that he suffers from “chronic separation loss” and therefore was thought “to take no interest in life.” Appropriate psychiatric evaluation and treatment would most likely have stabilized his mood swings and improved the quality of his life.

Some areas' protective service agencies do a remarkable job in intervening and dealing with complex practical problems. However, mental health issues are rarely addressed because of lack of expertise, financial limitations, and perceived or actual unavailability of psychiatric consultations.

CHARACTERISTICS OF CAREGIVERS AT RISK FOR ABUSIVE BEHAVIORS

The perpetrators of elder maltreatment require as much attention as the victims. Stresses associated with caretaking are strong predictors of abuse. Separating victims from perpetrators can be very difficult in view of limited formal and informal supports available to the caregiver and alternative resources for the dependent elderly. The following major categories of abusive caregivers have been identified.

Criminals This category includes perpetrators of elder abuse whose activities are criminal in nature and who need to be dealt with by the criminal justice system while the victim is protected. Mental capacity of perpetrator, nature and severity of victimization, motivation, forethought, and malice must be determined. Abandonment, rape, and murder are examples of criminal activity by those who have assumed responsibility for the care of a dependent elderly.

Adult Children The pattern of abuse across generations, from child abuse to elder abuse, has been noted relatively infrequently whereas child abuse from generation to generation is common. In other words, parents who have been abused as children are more likely to abuse their own children than to abuse their frail elderly parents.

Spouses Spousal abuse can be a marital pattern of behavior of many years' duration preceding the disability that made one partner a caregiver of the other or that created increased dependence on each other. Unlike situations in which physical and verbal abuse arise out of stress secondary to resistance or lack of knowledge, access to, or availability of community or institutional resources, cases of long-term spousal abuse require different types of intervention. The relative frailty of the victim or the nature of the disability determines the nature of the protection required when modifications of the abuser's behavior is unlikely.

Persons With Mental Disorders Other perpetrators of abuse are those who because of their own limitations or disorders (e.g., developmental disabilities, schizophrenia, substance abuse) were previously taken care of and provided for by the now-dependent elderly parent. Physicians are in an optimal position to alert families with adult dependent children to plan well in advance for such an eventuality in late life. Situations in which role reversal is not possible need to be attended to in a timely manner.

Other Overstressed Caregivers Who Turn Into Abusers Another category of caregivers at risk for abusive behaviors are those whose behaviors have gradually become abusive or neglectful and whose elderly charge has become more and more dependent over the course of time. Community and family resources for such caregivers may be inadequate to begin with or may have become inadequate over time. Caregiver stresses may have become compounded by added responsibilities for more than one dependent person. Such a caregiver's behavior can most likely be modified with suitable interventions.

ASSESSMENT AND INTERVENTIONS

A multidisciplinary team is necessary to assess the situation when abuse, neglect, or exploitation is involved. Team members should include a geriatric psychiatrist, a geriatrician or other primary care physician who is well schooled in the problems of the elderly, a neuropsychologist, nurse, social worker, and a lawyer familiar with elder law. The team would provide medical and neuropsychiatric workups, information on community resources for caregiving, and for financial and forensic assistance. The primary task of such a team is crisis intervention. Other equally important functions are prevention, training of students and of less skilled workers, and education of the public.

Reporting to an adult protective service agency is effective if the local APS has the resources to provide protection and health care in a timely and ongoing manner. If the local protective service lacks resources, the multidisciplinary team can assume an educational and advocacy role to make sure that the reporting of abuse leads to effective intervention and follow-up. The multidisciplinary team can become a vital resource for a university, medical school, and the community. The team would evaluate the history, risk factors for abuse, medical and neuropsychiatric status, functional status, and psychosocial factors. With the backing of such a team the primary care physician need no longer feel overwhelmed by the complexity of most elder abuse, neglect, and exploitation situations that cannot be dealt with

effectively by any one health care provider.

A major problem workers in this field encounter is client rejection of and resistance to (noncompliance) the recommendations made. In order to establish a working alliance, a skilled and sensitive interviewer is needed who can include the elements outlined in Table 51.6g-4. Protocols and checklists can be of assistance to ensure comprehensive and factual information.

▶ **Assess** sensory and mobility deficits
▶ **Listen** with a nonjudgmental ear, without recoiling when atrocities are revealed
▶ **Assess** nature of physical and mental illnesses
▶ **Assess** suicidal or homicidal intent
▶ **Distinguish** realities from fantasies
▶ **Distinguish** fears from anxieties
▶ **Distinguish** denial from exaggerations
▶ **Assess** developmental disabilities
▶ **Distinguish** early-onset from late-onset mental disorders
▶ **Detect** mental stress secondary to physical abuse
▶ **Assess** perpetrator's psychopathology and other impairments
▶ **Be** culture and gender sensitive

Table 51.6g-4 Essentials for Inclusion in Assessment Interview

The psychiatrist brings to the team expertise in the diagnosis of major mental illness, in the interface between medical and psychiatric conditions, in the effects of nonpsychiatric medications on the elderly individual, and in the prescription of psychiatric medications. A psychodynamic understanding of individual, marital, and family situations can facilitate a working alliance and greatly diminish patient noncompliance.

SUGGESTED CROSS-REFERENCES

Section 51.3e discusses Alzheimer's disease and other dementing disorders, Section 51.6b discusses medical-legal issues that are relevant to the elderly, and Section 27.3 presents adult antisocial behavior and criminality. Section 28.5 deals with death, dying, and bereavement.

SECTION REFERENCES

Adelman RD, Siddiqui H, Foldi N: Approaches to diagnosis and treatment of elder abuse and neglect. In *Handbook of Clinical Geropsychology*, M Hersen, VB Van Hasselt, editors. Plenum, New York, 1998.

American Medical Association: *Diagnostic and Treatment Guidelines on Elder Abuse and Neglect*. American Medical Association, Chicago, 1993.

Blunt PA: Financial exploitation of the incapacitated: Investigation and remedies. J Elder Abuse Neglect *5:*19, 1993.

Coyne AC, Reichman WE, Berbig LJ: The relationship between dementia and elder abuse. Am J Psychiatry *15:*543, 1993.

*Fulmer T, Burkenhauser D: Elder maltreatment assessment as a part of everyday practice. J Gerontol Nurs *18:*42, 1992.

*Goldstein Zucker M: *Elder Maltreatment. Family Violence: A Clinical and Legal Guide*, SJ Kaplan, editor. American Psychiatric Press, Washington, DC, 1996.

Hwalek M, Williamson D, Stahl C: Community-based M-team roles: A job analysis. J Elder Abuse Neglect *3:*45, 1991.

Jones JS, Veenstra TR, Deamon JP, Krohmer J: Elder maltreatment—National Survey of Emergency Physicians. Ann Emerg Med *30:*473, 1997.

*Korbin JE, Anetzberer G, Austin C: The intergenerational cycle of violence in child and elder abuse. J Elder Abuse Neglect *7:*1, 1995.

Kosberg IK, Garcia JL, editors: *Elder Abuse: International and Cross-Cultural Perspectives*. Haworth Press, New York, 1995.

Lachs MS: Preaching to the unconverted: Educating physicians about elder abuse. *7:*1, 1995.

*Lachs MS, Pillemer K: Abuse and neglect of elderly persons. N Engl J Med *332:*437, 1995.

Lachs MS, Williams CS, Obrien S, Hurt L, Kossack A, Siegal A, Tinetti ME: Emergency department use of older victims of family violence. Ann Emerg Med *30:*448, 1997.

McCreadie C, Bennett G, Tinker A: Investigating British general practitioners knowledge and experience of elder abuse: Report of a research study in an inner London borough. J Elder Abuse Neglect *9:*3, 1998.

Moskowitz S: Private enforcement of criminal mandatory reporting laws. J Elder Abuse Neglect *9:*3, 1998.

Moon A, Williams O: Perceptions of elder abuse and help seeking patterns among African American, Caucasian American elderly women. Gerontologist *33:*386, 1993.

Norris FH: Epidemiology of trauma: Frequency and impact of differential potentially traumatic events on different demographic groups. J Consult Clinic Psychol *60:*409, 1992.

Ollinger JP: Elder abuse: The outlook on federal legistation. J Elder Abuse Neglect *3:*43, 1991.

Paton Nelson R, Huber R, Netting FE: The long-term-care ombudsman program and complaints of abuse and neglect: What have we learned. J Elder Abuse Neglect *6:*91, 1994.

Paveza GJ, Cohen D, Eisdorfer C, Freels S: Severe family violence and Alzheimer's disease: Prevalence and risk factors. Gerontologist *32:*439, 1992.

Pillemer K, Finkelhor D: The prevalence of elder abuse: A random sample survey. Gerontologist *28:*51, 1988.

Pillemer K, Hudson B: A model abuse prevention program of nursing assistants. Gerontologist *33:*128, 1993.

Pillemer K, Suitor JJ: Violent feelings: What causes them in family caregivers. J Gerontol *47*(Suppl):165, 1992.

Ramsey-Klawsnik H: Interviewing elders for suspected sexual abuse: Guidelines and techniques. J Elder Abuse Neglect *5:*5, 1993.

Schmidt WC: Accountability of lawyers in serving vulnerable, elderly clients. J Elder Abuse Neglect *5:*39, 1993.

Stein KF: *Elder Abuse and Neglect: A National Research Agenda*. National Aging Resource Center on Elder Abuse, Washington, DC, 1991.

Steinmetz SK: Elder abuse by family caregivers: Process and intervention strategies. Contemp Fam Ther *10:*256, 1988.

Sugg NK, Inui T: Primary care physician response to domestic violence: Opening Pandora's box. JAMA *267:*3151, 1992.

Tartara T, Blumenman MA: *Elder Abuse Information Series No. 1 and 3: Elder Abuse in Domestic Settings*. National Center on Elder Abuse, Washington, DC, 1996.

*Tartara T, editor: *Understanding Elder Abuse in Minority Populations*. Brunner/Mazel, Philadelphia, PA, 1999.

Tinker A, McCreadie C: Elder Abuse. In *Bocklehurst's Textbook of Geriatric Medicine and Gerontology*, ed 5, R Tallis, H Fillit, JC Brockehurst, editors. Churchill Livingstone, New York, 1998.

CHAPTER 52. HOSPITAL AND COMMUNITY PSYCHIATRY

52.1 PUBLIC PSYCHIATRY

JOHN RICHARD ELPERS, M.D.

Public psychiatry encompasses all mental health service systems that are primarily sponsored and funded by governments—federal, state, and local. It is no longer appropriate to conceptualize hospital and community services as separate treatment systems; they are integral components of the spectrum of services essential to any public mental health treatment system. Services may be provided directly by civil servants or contracted by government to either nonprofit or for-profit agencies. The essential feature is that the clientele are the responsibility of government because they do not have the means to provide for their own care.

HISTORY

The history of public psychiatry is the history of psychiatry until the beginning of the twentieth century. Before then there was no private psychiatry except for some small institutions practicing *moral treatment*, a philosophical movement that held that with humane care and close patient-staff interaction, patients could be restored to function. Moral treatment followed the reformations of Philippe Pinel, William Tuke, and Benjamin Rush at the beginning of the nineteenth century. While these small institutions were quite successful, many people with serious mental illness wandered through the towns and countryside; others were held in homes where keepers were paid for their care.

The dismal care of persons with serious mental illnesses led to a national movement in the United States to provide care and treatment for people with mental illness. The intent of the movement, led by Dorothea Dix (1802–1877), was to establish asylums for the insane. Congress passed a bill giving the federal government this responsibility; however, President Franklin Pierce vetoed that legislation in 1854. Thereafter, Dix and her followers convinced all of the existing states to establish such asylums, which marked the beginning of the state hospital systems and the tradition of state responsibility for persons with serious mental illness.

The cause of humane treatment was forgotten as the Civil War absorbed the nation's resources and energy. After the Civil War, apathy and therapeutic nihilism again prevailed. The insane asylums expanded as demand for their services increased. Unfortunately, staffing lagged far behind the patient population, and over the next 80 years, these once-promising facilities evolved into terrible custodial institutions, not unlike those that had been dismantled at the beginning of the nineteenth century.

Modern Public Psychiatry At the beginning of the twentieth century, psychodynamic psychiatry, catalyzed by Sigmund Freud and his many followers, again stirred hope that mental illness might be cured or prevented. The private practice of psychodynamic psychiatry sprang from this belief system and the child guidance movement based on the principle that early intervention in childhood prevents mental illness in the adult. While that movement never accomplished its prevention goals, it nevertheless stimulated interest in more ambitious efforts to apply psychodynamic and public health theories to a broad spectrum of mental illnesses.

Two other movements were taking place at that time, one professional and one lay. The emphasis on treatability and the successes of some progressive institutions again fed a sense of optimism and the idea that people could be treated outside institutions. By 1909, Adolph Meyer (1866–1950) proposed a visionary comprehensive community mental health approach that included geographically congruent service boundaries for human service agencies that serve similar populations; cooperation between psychiatrists, police, teachers, and social workers; and community programs of prevention and rehabilitation. Meyer, a major proponent of drawing the family physician into the care of people with mental illnesses, had many followers and became an acknowledged leader of American psychiatry. His ideas permeate much of the subsequent community mental health movement in America.

In 1905 Clifford W. Beers, a former mental patient, wrote *A Mind that Found Itself*, which exposed the awful conditions in mental hospitals. After its publication, Beers developed a coalition of lay and professional leaders and founded a new movement to improve the care of persons with mental illness. This movement was compatible with other reformist philosophies of the time and led first to the formation of the Connecticut Committee for Mental Hygiene in 1908 and subsequently to the formation of the National Committee for Mental Hygiene which became the National Mental Health Association, an organization that continues to be a leader in the fight for better care for people with mental illnesses.

Social Psychiatry Social psychiatry has been an effort to integrate the knowledge accumulated in clinical psychiatry with that of sociology. The diverse theoretical base of social psychiatry ranges from the interpersonal theories of Harry Stack Sullivan to Talcott Parsons' concept of social roles. Persons considering themselves social psychiatrists use theories and methods of the social scientists as well as clinical knowledge, leading to interests in family and extended family groups, group dynamics, dynamics of the workplace, a wide array of social support systems, and cross-cultural issues. The speculations of social psychiatry have contributed a great deal to the development of the community aspects of public psychiatry, leading psychiatry to address such areas as poverty, racial prejudice, war, and mass migrations. Unfortunately, social psychiatry's efforts were swallowed up by the magnitude of the problems addressed, with little or no impact; therefore, one cannot ascertain the correctness of the theories behind the interventions.

Role of the Federal Government Leaders of American psychiatry involved themselves early in the care of military personnel in World War II. A great deal was learned from mistakes in handling psychiatric casualties during World War I, which left many chronically ill veterans suffering from posttraumatic stress disorders. The treatment of acute stress reactions with sedation and support of the soldiers at the forward aid station and rapid restoration to their fighting units proved highly successful. After the war, this success further fueled the belief that mental illness could be successfully addressed. The National Institute of Mental Health (NIMH) was established in 1948, and the Joint Commission for Mental Illness and Health was convened by President Dwight D. Eisenhower in 1955.

The joint commission proved to be a major turning point in the interest and understanding of mental illness. It involved a tremendous number of persons from a wide array of interest groups and gathered many volumes of data. Its most significant findings and recommendations were condensed into a compact book, *Action for Mental Health*, which quickly became an important, widely read document when published in 1961, early in President John F. Kennedy's administration.

The NIMH leaders became the primary shapers of the federal response to the report of the joint commission. The major early decision was to develop an entirely new system of care rather than attempting reform of the existing state institutions. The belief at that time was that the available federal resources would be only a small fraction of the states' current expenditures in state hospitals; thus, if federal resources were to go to improve state hospitals, they would have little impact on the overall treatment of mental illness. The federal government was primarily interested in prevention, early intervention, and treatment of illnesses before they became chronic, not in assuming the cost of the states' care of those already institutionalized. The rationale was that institutions would disappear as chronic mental illnesses were eliminated. Unfortunately, that theory proved to be incorrect; once again the the lack of commitment of money doomed persons with serious mental illness to continuing suffering.

Community Mental Health Centers In January 1963, President Kennedy's message to Congress on mental health and retardation bore only slight resemblance to the recommendations of the joint commission after the input of the NIMH, but it formed the basis of the federal community mental health centers legislation that ensued. In October 1963, legislation to begin the construction of community mental health centers (Public Law 88-164) was signed into law, and in 1965, the construction act was extended and staffing grants were authorized (Public Law 89-105).

Local community governmental or nonprofit agencies were to be given grants to construct and operate community mental health centers. Technically, state governments were eligible and involved in the planning, but this involvement was more technicality than fact, and grants were directly negotiated between the

recipient agency and the NIMH. Centers had to ensure that a specified essential range of services would be provided—inpatient, outpatient, partial hospitalization, emergency services, and community consultation and education to a defined catchment area of 75,000 to 200,000 people. Constructed facilities were required to provide the essential services for 20 years and continue to be used for some type of public mental health services for 50 years. A multidisciplinary staffing pattern was required and links to welfare workers, clergy, family agencies, schools, and other human service agencies were to be maintained. After 1965, in addition to the essential services listed above, a long list of optional and supplementary services became eligible for staffing grants. Initially these staffing grants started at 80 percent of the projected budget and declined over 5 years to zero. Amendments to the original legislation allowed higher funding ratios for poverty areas and longer periods of support. In 1975 major revisions (PL 94-63) expanded the original five essential services to twelve, adding services for children and aged persons, alcoholism and drug abuse services, screening prior to state hospital admission, follow-up care for patients discharged from institutions, and transitional housing. Other requirements included a community governing board, quality assurance and utilization review, a relationship with any health maintenance organizations serving the catchment area, and services under Medicare and Medicaid. As grants declined and centers incurred financial difficulties, grants were made available for "distress," special programs, and new populations. Thus many programs received federal financial support far longer than 5 years.

By 1981, at the zenith of federal interest in mental health services, there were 758 community mental health centers serving catchment areas comprising over 52 percent of the country's population. At that time, President Carter's even more ambitious Mental Health Systems Act became effective. However, that act was short-lived; it was repealed in the early days of the Ronald Reagan administration in favor of block grants to the states to diminish the federal investment. Those block grants gradually decreased in relative size and importance.

The intent of the declining staffing grant was to help a community mental health center become established, develop local support, acquire multiple revenue sources—especially Medicaid and other third-party payments—and become ingrained into the local fabric of services to ensure its continued functioning after federal funds were withdrawn. This strategy seemed reasonable, since the per capita mental health services expenditures of the NIMH never exceeded 10 percent of the states' expenditures for public mental health services. Unfortunately, the strategy was largely a failure. While many formerly federally funded community mental health centers continue to exist, many operate as private agencies and have little or no influence on the availability of services to people with serious mental illness. Those centers that treat serious mental illness depend heavily on state government support and are either directly operated by or under contract to state mental health systems. The original vision of free-standing agencies that would treat all of the community's mental health problems and eliminate the need for state hospital care was not realized.

Serious mental disorders were neither preventable nor curable by early intervention. Detaching the new system from state mental health systems allowed many centers to avoid handling serious mental illness in favor of treating the "worried well." Much of the third-party money for the treatment of mental illness declined drastically in the late 1970s and 1980s. Medicaid, the federal-state health care financing system for persons in any of the federal financial aid categories, eroded through both fee and service restrictions. This was to have been the major source of revenue to allow the centers to serve people seriously disabled by mental illness. Simultaneously, private insurance coverage for mental illness was restricted as costs escalated. Limits were imposed on both treatment modalities and duration. In addition, the numbers of employees covered for mental health care through their employers' group plans declined. Thus, centers lost much of the revenue that had paid for care of the poor and the middle class.

Deinstitutionalization The community mental health movement was and is far greater than the federal community mental health center effort. Before 1963 a number of states had begun to adopt the philosophy of treating mentally ill persons outside of institutions. Many community mental health programs were established at the local level with state subsidies. Creative ways were found to allow people to use the available Aid to the Totally Disabled (ATD) program to support themselves in the community. Enactment of the Supplemental Security Income (SSI) legislation in 1972 provided more funds for community living, and the movement of seriously ill persons out of institutions accelerated. The reduction in state hospital populations preceded the availability of psychotropic medications; however, the availability of neuroleptic and antidepressant drugs expedited this depopulation of the state hospitals. The neuroleptic medications often were not taken after discharge because of their adverse effects; yet hospital administrators were reassured that their patients would behave well in the community. Eventually the success or failure of state community mental health programs came to be judged by their impact on state hospital populations. Deinstitutionalization superseded providing better services to people with mental illnesses.

Deinstitutionalization was more a civil rights goal than a psychiatric treatment goal. The 1960s were a time when American social systems were in great flux with strong momentum toward freedom, individual rights, and removal of governmental restrictions. These trends were best exemplified in the civil rights movement, the demands for sexual freedom, the rebellion of youth against the older social norms, and major riots caused by racial and economic inequalities. Even before the move to deinstitutionalize mentally ill patients, a major depopulation of the prison systems resulted from shorter sentencing, early release, and an emphasis on rehabilitation rather than punishment. Thus, philosophic trends regarding the care and treatment of the mentally ill continued to follow the social trends in the larger society.

The states' motivation to move patients to the community was not always altruistic. State hospitals required major expenditures. The opportunity to reduce hospital size and budgets by shifting some of the costs of caring for people to the federal subsidy programs was quite attractive. This shift of responsibility was abetted by mental health professionals who underestimated the cost of their new community programs to state legislatures and governors.

The apathy of patients in institutions was attributed to the institutionalization itself. With active rehabilitation, some of those removed from the institutions did become less apathetic. However, when left alone, their apathy returned. Similarly, as the availability of institutional care declined and more persons with severe mental disorders were treated in short-term hospitals and outpatient settings, many of them showed similar apathy. The apathy attributed to institutional care was instead a product of the illness itself—the passive, or negative, symptoms of schizophrenia.

TRANSINSTITUTIONALIZATION Making deinstitutionalization an end in itself was unfortunate. Although many long-term patients in state hospitals had been in those institutions far too long with little or no effort to rehabilitate them, shifting them to the community did not necessarily improve their fate. Their apathy made them easy to place in boarding facilities (or nursing homes if more care was required). However, no evidence shows that these patients were happier after placement than they were in the state institutions.

The shift of chronically and seriously ill patients out of state hospitals, although variable, has drastically reduced the populations of those hospitals since the peak levels of the late 1950s. By 1998 California, an extreme example, reduced its noncriminal, mentally ill, state hospital population to less than 3 percent of its 1957 level of 35,000 patients. Nationwide the state psychiatric hospital population decreased from 560,000 in 1955 to fewer than 88,000 public state and county beds in 1993. Unfortunately, many patients were merely shifted from state hospitals to other settings, most frequently nursing homes—an estimated 750,000 people—and board-and-care facilities. The real advantage of these community facilities was frequently fiscal. They were cheaper for state governments. Initially, the new institutions were often smaller and better than the state hospitals, but this is no longer the case. Care at many state hospitals improved as their staff-to-patient ratios improved, and there is less selectivity in the community facilities recruited with increased demand. Patient populations are more widely dispersed, and their care is more difficult to monitor than in the past.

Initially, as populations in state hospitals decreased, community hospital care increased. General hospital psychiatric beds rose from 7,000 to over 53,000 in 1990. Private, free-standing psychiatric hospitals peaked at 79,000 beds, and the Veterans Administration hospitals provided over 26,000 beds in 1988. The impact of limitations of insurance coverage, utilization review and control, and managed care has decreased the number of free-standing psychiatric hospital beds rapidly since that time; in 1993 there were approximately 43,000 beds. The differences between the severity of illness of patients treated in public and those treated in private institutions are far lower today because all hospitalized patients must demonstrate symptoms too severe to be treated in a less-intensive care setting. The availability of inpatient mental health services will be further complicated as the welfare reforms passed by Congress in 1996 are implemented by the states, modifying who is eligible for Medicaid.

The reduced utilization of hospitals has encouraged a more innovative range of alternative services. These range from good residential care settings to a wide array of outpatient services emphasizing case management, aggressive outreach, treatment of concurrent substance abuse, and a wide array of psychosocial rehabilitation services for persons with serious, persistent mental illnesses. Unfortunately, many jurisdictions cannot afford to provide a comprehensive, competent system of care; thus persons with serious illnesses become homeless or live in inappropriate settings, causing major annoyance to the community and their relatives and friends. These individuals would have been kept in state institutions in the past.

EVALUATION Deinstitutionalization as implemented seems to have been a failure. However, the concept of community care in preference to remote institutional care was not tested. The ability of state governments to reduce or shift the cost of treatment was tested and was quite successful. In spite of innovative treatment programs and new medications, many persons with mental illness now have far fewer treatment options because funds are not available. This lack of treatment coupled with the nationwide shortage of low-income housing leaves persons with severe mental illnesses in desperate straits, with few support systems. That unfortunate outcome illustrates the danger of undertaking a massive change in treatment policy without fully understanding the costs, the benefits, or the moral commitment of society to

the treatment of people with mental illness.

Mental health professionals share the responsibility for the plight of deinstitutionalized mentally ill persons. Many argued, without hard data, that they could treat seriously mentally ill people less expensively in the community than in state institutions. They assumed that through early intervention and judicial selection of the least-expensive alternatives they could perform as promised. Neither their performance nor the available funds measured up to expectations. States did save large sums of money. The people with mental illness remain in the community. Unfortunately many congregate in large numbers in our urban centers and in lesser numbers throughout the country, frequently lacking adequate services. Treatment is available in many parts of the country only when dangerous or disruptive behavior causes patients to be brought to emergency rooms and jails. Patients not on the streets may live in hovels with little hope of rehabilitation, alienated from families, and without support systems. At the end of the twentieth century, the conditions that led reformers to demand acceptable asylums for persons with mental illness at the beginning of the nineteenth century are seen again. The society at large is again turning to restrictive institutions to solve its problems. We have come full circle.

BASIC CONCEPTS

Responsibility Fixed responsibility for the mental health care of a specific individual or group of individuals is basic to public psychiatry. This responsibility must be for the individual, regardless of where the person is at any given time—in a treatment facility or in the community. Preferably that responsibility is assigned before a person's mental illness is detected, so a source of care can be designated at the earliest sign of illness. Thus patients would become the responsibility of a given physician because of their employment status, membership in a particular health insurance plan, or a contractual relationship with a particular provider group in this age of managed care. The payment for services might be wholly or partially guaranteed by membership in an age or disability category (e.g., Medicare or Medicaid), but there must still be a mechanism to ensure the availability of care.

Fixing responsibility for care of persons with severe mental illness is essential but difficult. They frequently leave hospitals, nursing homes, and residential facilities, either because their condition improves or because it deteriorates. Many cannot hold jobs long enough to enroll in health plans, and when they can, the managed care plans usually provide limited medical services without the essential psychosocial supports. Without these supports persons with severe mental illnesses frequently cannot avail themselves of medical care and are often excluded from managed care plans.

When the responsibility for care is not otherwise assigned, governmental jurisdictions tend to assign it geographically. The individual does not select the responsible person or agency; when someone first comes to the public mental health agency, it is already clear who must assume the responsibility for that patient. This fixed responsibility is intended to eliminate shunting away less desirable—meaning more difficult—patients or selecting only patients considered amenable to a particular type of treatment (e.g., psychotherapy).

A mental health agency with responsibility for the population of a defined geographical area must have a way to learn about that population. There should be an assessment of how its limited resources should be deployed to best meet the needs of the population served. This process, called *needs assessment*, can be accomplished directly by sampling treated and untreated populations in the area, gathering the opinions of knowledgeable persons about the area's needs, or inferring from what is known about similar populations. The complexity, cost, and specificity of the assessment will depend upon the approach or approaches chosen.

Community Participation Since the public mental health provider is carrying out some governmental functions, its legitimacy in the community must be established. This problem is generally addressed through the principle of *community participation*, that is, the community served must have a voice in the governance of the program, regardless of who is paying the bill. Defining and obtaining community participation is complex. Usually the elected political leaders of the community are expected to play a major role. They may be a public agency's governing board, appoint some or all of the agency's governing board, or serve in advisory capacities. They extend their authority over private agencies or individuals through contracts.

Indigenous leaders outside the political power structure form an important power structure in every community. These may be business leaders, leaders of the arts, civic crusaders, and, perhaps most importantly, leaders of the community's racial and ethnic groups. These leaders must all participate actively in defining the agency's mandate and approach to providing services.

Service systems need to have a process to develop a plan to serve their respective populations. The plan should evolve from the needs assessment, and if there is adequate community participation in its development, it will have wide public support as the best that can be done with available resources. Planning requires knowledgeable staff, but the plan itself depends on the broad range of inputs from the needs assessors, community participants, knowledgeable professionals, consultants expert in areas in which staff are not, and persons expert in organization, budgets, and management. Patients and their families are essential participants because they have unique knowledge of the needs of persons with mental illness. The process of developing the plan is frequently more important than the product. A successful process draws the diverse elements of the agency and its constituency together into a cooperative, functional relation that will allow efficient future problem solving. A wonderful plan developed without an effective process will not work.

A key element in bringing together the plan and ensuring a satisfying process is educating all participants in the art of the possible. The more the lay leadership knows about the restraints on the treatment system—legislative mandates, regulations, facilities, license requirements, existing contracts, budgets, and staff availability—the more helpful they can be in the planning process and the more realistic will be the final product. Mental health professionals must be willing to learn the values, customs, and expectations of the community. They often find that they know far less than they thought.

Outreach Outreach includes all the mechanisms that might be used to ensure that persons who need services know about their availability and priorities. The mechanisms range from listings in the Yellow Pages and community service advertising to cruising the streets enticing homeless persons with mental illness into treatment by offering them food, clothing, and shelter. In this process the community leadership enlisted in the participation and planning efforts can be valuable help. The outreach processes must reflect the service priorities established in the plan.

Outreach sometimes means that the agency must engage in more assertive social action. In the early days of the community mental health movement, social action was seen as a mechanism to change the negative forces of society and thus prevent mental illness. That objective is seldom raised today; instead, social action today is usually undertaken to improve the chances that people will receive necessary treatment. Reducing stigma is one form of social action. The effort to both eliminate the demeaning slurs and empathically present the plight of the patients and their families in the media can do a great deal to eliminate stigma. This allows patients to seek treatment and makes treatment programs more available. Social action may mean joining with other concerned parties to advocate better low-cost housing so people with mental illness will have a chance at decent shelter or better insurance coverage to allow more services. The social actions of a mental health agency should be judged by the impact the desired change will have on the lives of the people it serves.

System Balance Modern public mental health systems, especially in urban areas, tend to be complex. Every phase of a patient's illness requires a different type or intensity of service. If a service component is not available when the patient needs it, the patient must either be placed in a less appropriate treatment setting or have treatment delayed until the needed program is available. For a system to be balanced, each component of the treatment system must exist in balance with all other components. Bringing about this balance is extremely difficult because mental health professionals and lay persons tend to invest themselves in specific components of the system, fighting to develop their favorite component at the expense of others. When mentally ill persons are on the streets and emergency rooms are full, unsophisticated persons would solve the problem by simply acquiring more beds. When this is attempted, however, the new beds fill rapidly with long-stay patients if the needed community resources are not available to allow them to be discharged from those hospital beds. Thus, this most expensive element of the system again backs up, emergency rooms again become overcrowded, the initial problem has not been rectified, and resources that could have enhanced the community services have been absorbed by the bed costs. Had the system invested in fewer beds, more community placements, and extensive community support systems, the problem would have improved without the costly bottleneck.

To ensure a balanced treatment system, a central management must have authority over the entire integrated system and must have the ability to see that an appropriate plan is implemented to correct imbalances. Trust and respect for the central management are essential if such a complex system is to function as an integrated whole. When resources are limited, system elements tend to declare war on each other, severely taxing the skills of the integrating authority. When funding systems are tied to service components rather than to patient needs (as occurs in most noncapitated systems), fundable services tend to be available regardless of need while those not readily funded are in short supply. Many programs are severely distorted by what the federal Medicaid program does or does not support.

Continuity of Care In the ideal, continuity of care means that patients have the same therapists or caregivers throughout their illness. Thus the staff person who initiates treatment in the emergency room should follow the patient, whether hospitalized, returned home, or placed in an alternative setting. A small system may allow the staff this scheduling flexibility to permit staff to work with their patients wherever they might be. In the real world such continuity of care is rarely available,

regardless of how important it is for persons with serious mental illnesses. In the name of staff efficiency or convenience, service systems tend to become compartmentalized. Staff members are assigned to program units not to patients, an antiquated model difficult to eliminate. A positive aspect of some managed care plans is their insistence on continuity of care; they will not contract for services unless psychiatrists are available to follow their patients as both inpatients and outpatients.

Many paper transfer mechanisms are used to compensate for the lack of continuity of care in compartmentalized service systems. A great many use charts, forms, and documents. Some systems have charts that go with the patient even if the staff do not. Admission and referral forms are used to let newly assigned staff know what the patient has undergone previously. Similarly, discharge summaries and aftercare plans are designed to follow the patient from the hospital back to the community. In many settings computer networks are beginning to help transfer information about patients as they traverse the system. Telephone conversations and face-to-face discussions are very important adjuvants to the paper transfer of information. Unfortunately all of these mechanisms help the staff feel better but do little to reassure the patients.

Legal requirements for confidentiality can be major impediments to continuity of care when the same therapist is not responsible throughout all phases of a patient's illness. A patient's refusal to sign a release of information form when parts of the system are different agencies can be a powerful disincentive to continuity of care. However, confidentiality requirements can be an excuse for not doing all in the professional's power to ensure that essential information is transmitted. Although public stigma and the secrecy of psychotherapy motivate patients to be concerned about their confidentiality, staff members should help them see the greater good in information transfer. With the shift toward more biological treatment of serious mental illness and away from probing the unconscious, such concerns about confidentiality lose relevance.

Case management constitutes a more human effort to maintain continuity of care. Here a new class of providers is used to follow the patient throughout treatment. Their role is a mixture of advocacy, system management, patient support, and sometimes therapy. If the elements of the mental health system are not available or do not cooperate, the role of case manager can be frustrating. To be effective, case managers must have small case loads and wide latitude to meet the patient's needs. When these objectives are met, it is hard to distinguish the case manager from the primary therapist for the person with serious mental illness. Merging the roles of case managers and primary therapists to eliminate unnecessary redundancy may accomplish the original goal of continuity of care—the patient has the same primary therapist throughout the duration treatment. The case manager described here is not to be confused with the case managers working for managed care companies who have minimal patient contact and are better described as cost managers.

Least Restrictive Alternatives The well-accepted concept of *least restrictive alternatives* means that mentally ill persons should be treated in settings that least interfere with their civil rights and freedom to participate in society. Patients should not be hospitalized (especially against their will in locked facilities) if their illness can be treated in a more open setting. Most states now have legislation protecting the rights of patients, and treatment in the least restrictive setting is usually one of those rights.

How far the concept of least restrictive care should be carried remains controversial. Patient rights advocates argue that this is the overriding principle, even if the therapeutic potential of the least restrictive facility is seriously compromised. Clinicians argue that therapeutic efficacy is the greatest concern and that the least restrictive principle can be compromised in the interest of better treatment. One must consider carefully whether any restriction on a patient is for the convenience of the psychiatrist and staff or for the protection of the patient.

Rehabilitation Interest in rehabilitation is enjoying major resurgence in public mental health programs for two reasons. First, deinstitutionalization placed many more persons with serious mental illness in the community under the care of public mental health systems. Without continuous attention to their psychosocial rehabilitation these individuals deteriorate over time and cost the mental health programs great sums of money. Second, the new generation of neuroleptics alleviates the negative aspects of schizophrenia and facilitates rehabilitation. Unfortunately few psychiatrists are available to lead psychosocial rehabilitation efforts. Most of the traditional mental health training programs offer little, if any, training in the techniques of rehabilitation. Disease or illness paradigms used in medical and psychiatric training do not encourage interest in or pursuit of rehabilitation. When one focuses on finding and identifying symptoms and syndromes, one does not concurrently learn to see the patient's strengths or potential. The professions closely allied to psychiatry and medicine have generally adopted views of the patient similar to those of psychiatry. A second problem is that the vast majority of those professionals interested in rehabilitation—in medicine as well other disciplines—are interested in physical rehabilitation and have little knowledge of the needs of persons with mental illness. Not being trained in managing the emotional problems of persons with serious mental illnesses, they have little tolerance for their idiosyncrasies. Similar behaviors in persons not labeled as mentally ill are often better tolerated. Finally, it is very difficult for professionals to agree on the goals of rehabilitation and unfortunately they are reluctant to consult the patients.

Psychosocial rehabilitation goals can range from complete restoration of functioning to limited improvement in the patient's ability to handle self-care. Before rehabilitation can begin, such basics as food, clothing, and shelter must be available to the patient. Thereafter each patient's needs must be individually ascertained. Work is an extremely important component of any individual's sense of self-worth and participation in society. Professionals often underestimate a person's ability to work, seeing symptoms rather than specific functional capacities. That realization has evoked increased interest in work-oriented rehabilitation in recent years, with good results. Many persons with mental illness require continuous rather than short-term efforts to achieve and maintain improved functioning. Some professionals find this frustrating. The concept that once attained, improved functional abilities may erode if continuing professional attention is not available is incompatible with traditional medicine's desire for a cure. In reality, this situation does not differ from that in many other chronic medical conditions, and the outcome is frequently far better.

SYSTEM COMPONENTS

Prevention Mental health programs have traditionally used the public health nomenclature of primary, secondary, and tertiary prevention. However, when applied to mental illnesses, this system has problems.

Primary Prevention *Primary prevention*—avoiding the occurrence of an illness—is elusive in psychiatry. In the past a lot of social engineering was tried in the name of preventing mental illnesses, but there is no evidence that it was successful and so it is now out of favor. The social breakdown syndrome was to be prevented through deinstitutionalization; however, the loss of social skills in persons who have never been in long-term institutions is now considered to be a component of the primary mental illness. Currently, some professionals believe that behaviors such as child abuse are learned in childhood and can be prevented through reeducation; this theory remains unproved. As human genetics becomes better understood, genetic counseling may become much more relevant. Presently it is important in such disorders as Huntington's disease and of some value in advising prospective parents with a strong family history of schizophrenia or other major mental illnesses. Primary prevention of the mental disorders that psychiatry struggles with today awaits better delineation of their causes.

The primary prevention of disorders that used to fill mental hospitals is being practiced every day. Tertiary syphilis is prevented by treatment of the primary infection. Many organic brain disorders are now prevented by making proper vitamins available in foods, especially for high-risk persons, such as those with alcoholism. Serious mental and emotional damage can be prevented in many children by eliminating the use of cocaine and other drugs during pregnancy. Education about acquired immune deficiency syndrome (AIDS) can prevent both the physical and psychiatric aspects of this disease. When primary prevention works, the illnesses are soon forgotten or the responsibility for their prevention is no longer the purview of psychiatry.

Secondary Prevention *Secondary prevention* is the early detection and prompt treatment of a disorder to avoid or reduce permanent damage. In psychiatry this is a major responsibility of children's services. The current best model to accomplish effective secondary prevention is the federal Children's Mental Health Services Initiative and its many permutations in the states. These systems attempt early identification of emotionally disturbed children and carefully target wraparound services to meet a broad spectrum of needs and mitigate their illnesses. This is done through the careful coordination of the work of agencies most likely to come into contact with high-risk children. Once children with various behavioral disorders are identified, these multiagency systems of care bring all of their resources to bear on the needs of the child and family. The effort is to reinforce the family and create the best possible environment to mitigate further deterioration in the child's emotional state. Family structures are preserved, and out-of-home placements are avoided to allow more-normal social development. A large percentage of children with behavioral disorders of childhood demonstrate major mental illnesses as adults. Whether these early interventions will prevent continuing disability or deterioration as the child matures is not yet known.

On a broader scale, secondary prevention is embodied in the work of most psychiatrists and other mental health professionals. All try to initiate treatment and alleviate suffering at the earliest possible moment. Crisis intervention theories and services are important attempts at secondary prevention. Efforts to educate the public and reduce stigma to allow persons to seek treatment earlier is also secondary prevention.

Tertiary Prevention *Tertiary prevention*—reducing the functional impairments caused by disease in persons in the population—in its broadest sense encompasses all forms of treatment of persistent mental illnesses. In its more restrictive and useful sense, tertiary prevention is interpreted to mean reducing the residual defects left after the disease itself has been addressed. In this case the primary thrust of tertiary prevention is rehabilitation. Because most serious mental illnesses follow a long and relapsing course, it is not appropriate to delay rehabilitation until treatment is complete. Consequently secondary and tertiary prevention are often concurrent processes in psychiatry. Indeed, it is often difficult to ascertain how much of the patient's apathy and anhedonia is a symptom of the active disease and how much is residual.

Tertiary prevention or rehabilitation in psychiatry nearly always addresses patients suffering from the most severe and debilitating illnesses—schizophrenia, the most severe affective disorders, and the most disabling personality disorders. All of these illnesses, especially schizophrenia, tend to strike in late adolescence and young adulthood. These individuals are effectively removed from society during those years when most people complete their education, learn trades, establish careers, begin families, and develop social support systems in the community. Consequently, even if the illness were completely cured without any residual functional deficits, these people would still need extensive social rehabilitation. Unfortunately the illness leaves residuals in many cases. These people can have a wide range of psychological deficits that impair their ability to interact with others, handle the usual stresses of daily life, and achieve their potential. Therefore, psychiatric rehabilitation involves a complex process in which the professional attempts to address the psychological, social, and often the medical needs of the patient simultaneously. Obviously, the better the patient's functioning in society can be maintained during the acute exacerbations of illness, the less psychosocial rehabilitation will be necessary. This is why modern public psychiatry attempts to limit the lengths of hospitalizations by rapid interventions and to maintain social support systems even when patients are acutely ill.

Institute of Medicine Prevention Model Because many people find the primary, secondary, tertiary prevention terminology confusing, a new definition of prevention, recommended by the Institute of Medicine's Committee on Prevention of Mental Disorders, is gaining acceptance. Under this system, the term “prevention” would be used only for interventions that occur before the onset of a disorder. Secondary and tertiary prevention would be dropped in favor of the more accepted terms *“treatment”* and *“rehabilitation.’* That system further breaks down prevention into three categories. Universal interventions are targeted to the general public, independent of risk factors. Selective interventions are targeted to individuals or groups thought to have a higher-than-average risk of developing mental illnesses. Indicated interventions are targeted to high-risk individuals with biological markers or early signs of illness who do not yet meet the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) diagnostic criteria.

Crisis and Emergency Services Crisis and emergency services are mechanisms through which patients in need reach the system. In some programs these are the same service, in others they are separate operations. Emergency services are generally the entry point for patients with psychobiological problems who need hospitalization or a similar level of short-term care. Crisis services are for people whose difficulties are psychosocial and require immediate intervention but not hospitalization. The difficulty with this differentiation is the inability of patients to determine which service they need. Those persons and agencies in the community commonly referred to as *gatekeepers* (meaning police, social agencies, families, and others having first contact with individuals in trouble) also have difficulty differentiating an emergency from a crisis. Therefore a public mental health system that separates its emergency and crisis services must educate both the public and the gatekeepers about who should be referred where and provide for rapid transfer of inappropriate referrals to the other entry point. One alternative is to have a single entry point that provides both emergency and crisis services. Another is to have mobile crisis-emergency teams that screen people in the community, determine which services are necessary, and institute the appropriate responses. Mobile teams are most effective when they function integrally with local police agencies.

Crisis services are designed to help someone decompensating under acute psychological stress to reintegrate rapidly, marshal any available environmental resources, and return to function in the community as soon as possible, usually in hours to a few days. To accomplish that task experts in crisis intervention, short-term psychotherapy, psychotropic medications, family intervention, and community resources must be readily available.

Emergency services on the other hand are intended to serve patients decompensating in the course of a major mental illness deemed likely to require hospitalization or a similar intervention. Emergency services should be able to admit patients to the hospital or other appropriate facility without having the patient undergo further screening; otherwise there are intolerable backups of patients in the emergency setting. Both crisis and emergency services must make rapid but accurate diagnoses, rule out medical conditions that might be wholly or partially responsible for the patient's psychiatric symptoms, obtain the necessary laboratory and X-ray studies, and initiate treatment. To ensure the availability of medical care when possible organic causes are suspected, emergency and crisis services are best located in or near a general hospital with a good emergency room. A close working relationship between the psychiatric and medical emergency rooms is essential.

Well-functioning emergency-crisis programs greatly improve the effectiveness and efficiency of the overall system. By limiting access they avoid wasting resources on persons who do not need or cannot benefit from mental health services. When services are in short supply, as they tend to be in many jurisdictions, these programs play an essential rationing function. By assigning patients to the most effective, least restrictive, and least expensive service available, they ensure efficient use of available services, which is helpful to both the patient and the system.

Short-Term Hospital Services After the emergency service, the person in need of intensive treatment is usually admitted to an acute inpatient service in a general or free-standing psychiatric hospital. Licensing definitions and names may vary from state to state, but these services are expected to be well staffed with a full diagnostic and treatment armamentarium. Average stays are usually short, 1 or 2 weeks. After careful diagnosis, the patient's treatment in the acute hospital should be intense and involve psychotropic medications, education about the illness, group and individual psychotherapy, and activity therapies as appropriate. Work with family members and early discharge planning are critical. True continuity of care in an integrated service system will shorten the hospital stay by allowing continued development of the professional relationships, education, and treatment plan after the patient leaves the hospital. Patients clearly benefit from more rapid return to home, community, and jobs if the proper support systems are in place.

Long-Term Services Persons with severe, persistent mental illnesses have continuing service needs. These needs can be broken down into several components: treatment, housing, rehabilitation, social support, and financial support. While these components are discussed separately, they should be provided as parts of an integrated service plan with a single treatment team.

Treatment Patients with severe, persistent illnesses should have a number of alternatives available in a well-functioning public mental health system. Patients who can live in the community but would benefit from intensive treatment should have a full array of services available as intense or more intense than those in the hospital. Included are education about their illness and the effects, adverse effects, risks, and benefits of psychotropic medications; psychotherapy, especially of a supportive nature; group therapy; social activities; and assistance with housing, income maintenance, and transportation. This array of services can be organized into a day hospitalization program, which can be quite effective if it is truly intensive. However, today it is considered more efficient to have these services as front-end components of an integrated psychosocial rehabilitation system.

Patients who cannot live in the community because they remain dangers to themselves or others or who are gravely disabled but no longer need or can tolerate intensive treatment are cared for in state hospitals, skilled nursing facilities with psychiatric treatment staff available, and a variety of residential treatment programs. These programs must be individually assessed in any public mental health system because they vary greatly in intensity and quality of treatment and environment. With the advent of the newer neuroleptics and the success of psychosocial rehabilitation programs it is a mistake to assume that any patients must spend long periods of their lives in these long-term care institutions.

Some long-term programs—both residential and outpatient—are designed to address special problems of persons with serious mental illness and coexisting drug and alcohol dependence. The various forms of substance abuse are major problems for many mentally ill persons; up to 40 percent of persons with severe mental illnesses have coexisting substance abuse problems. Every public mental health system must have resources to treat dual diagnosis patients, preferably integrated with their comprehensive mental health care.

The vast majority of the severely mentally ill persons treated in public mental health systems today can live in the community. However, some will never achieve that goal or will do so only after long periods of illness. These persons require long-term or custodial care, which is still the role of state hospitals in most jurisdictions. With the reduction of state hospital populations, the quality of custodial care in most state institutions has improved greatly. What used to be very oppressive back wards are now usually humane, tolerant settings for people whose illness is severe, persistent, and unresponsive to modern psychiatric care. Skilled nursing homes in the community serve a similar population, especially those with concurrent medical problems and organic mental disorders. Those community facilities are varied and must be monitored individually. Licensing standards and quality-control mechanisms vary among the states.

Housing Securing and maintaining appropriate living situations is a major problem for persons with mental illness. Treatment is impossible if adequate housing, food,

and clothing are not available. Many patients have alienated their families of origin. They are unlikely to have developed supportive relationships with mates or peers; thus they must manage on their own with little help. Their problem is compounded by both the decline in the amount of government-subsidized low-cost housing over the past 25 years and by the disappearance of single-room occupancy hotels caused by urban redevelopment. When they were cared for in institutions, housing was provided. With deinstitutionalization, SSI was supposed to cover basic living costs; however, with the rising cost of even inexpensive apartments and limited or interrupted SSI support, many people with mental illness are forced to live on the streets or in board-and-care facilities that are often dirty and noisy and provide inadequate food and services.

Multiple solutions to the housing problems must be considered. Some mentally ill persons do remain at home. This is a cheap (and therefore desirable) solution for public mental health systems, but remaining at home undermines the self-esteem of young and middle-aged adults and places undue burdens on families who have already suffered a great deal. The presence of a mentally ill son or daughter can raise emotional tensions in the best of homes. Thus other mechanisms must be sought to house people whose illness forces dependency.

A wide array of facilities throughout the country fall under the general rubric of half-way houses. These programs have one thing in common—they are designed to ease the transition from hospital to the community. Most are designed to serve specific types of patients, have unique philosophies, and place specific demands or expectations on their patients. They tend to be selective and hence unavailable to many of the most seriously ill persons. Because no universal funding mechanism covers care in half-way houses, there are not a lot of them, and many cater to people with private means of payment. When available, a well-structured half-way house can be helpful for selected patients.

A major mechanism used in most states is a licensed facility that provides room, board, and some supportive care for chronically mentally ill persons. Some are for-profit, some nonprofit; some are large, some have six or fewer beds; some are very good, some are awful. Most depend upon what SSI pays in their respective states, taking most of the monthly check and giving the patient a small amount of spending money for personal needs. Most board-and-care facilities keep and dispense medications prescribed by the patient's doctor. Some insist that patients be medicated, sometimes heavily, to control behavior. These facilities tend to have little privacy, with shared bathrooms and bedrooms. For persons who have difficulty tolerating others encroaching on their space, as do persons with mental illness, board-and-care facilities may be very unpleasant settings.

One of the peculiarities of most current public mental health systems is that the housing needs of seriously mentally ill persons are tightly linked to their treatment needs. When hospitalized, the patient loses the bed in the board-and-care because the facility is no longer being paid. In the hospital facility they must accommodate to a whole new set of peers and staff members. When the illness recedes, the patient is transferred to another, less intensive facility with another collection of strangers. If patients again improve, they must again move, deal with new people, and try once more to establish some sense of continuity. Those forced moves are designed for the convenience and economic stability of the treating staff and institutions, not for frightened people for whom developing satisfying interpersonal relationships is a problem.

It is better for people with mental illness to live in permanent, stable housing, not linked to their treatment needs, as early as possible in their rehabilitation. Thereafter, should they need services, they go out for them—rehabilitation, socialization programs, work readiness training, supported employment, and psychiatric sessions. Should they need hospitalization, their homes remain, and they can return to a familiar environment on discharge and continue existing relationships. These homes might be an apartment by themselves, apartments with roommates, or apartment buildings for the disabled built with assistance from the federal Department of Housing and Urban Development. The essential features are that persons may remain there as long as they desire, they have privacy and space of their own, they own the furniture and appointments, and they can regulate their own schedules. Many people considered quite ill do much better in such settings than when they are shunted around. They have a secure base from which they can move out and take modest risks in dealing with other people, which reduces their fright. Some patients do require education and support to maintain their own apartments, but that support is usually less than is expected.

Rehabilitation Services With the advent of the serotonin-dopamine antagonists (atypical antipsychotics) many more persons with severe illnesses have had their symptoms controlled enough to make them candidates for rehabilitation. Indeed, today one must assume that all persons with schizophrenia and major mood disorders do or will need rehabilitation services. Therefore rehabilitation is an essential component of the continuum of services necessary for persons served by the public sector.

Rehabilitation focuses on the patient's remaining capacities not the residual symptoms. It emphasizes individual responsibility and self-reliance rather than illness and dependency. Once the acute symptoms are under control, rehabilitation should begin immediately. There is no single approach to all patients. Instead, one uses whatever means are necessary to help recovering patients adapt to their specific environments. One does not help a patient develop skills for no purpose. The rehabilitation needs of patients must be individually assessed within the context of what they wish to do with their life. Patients should expect to improve, but the expectations should be compatible with their capacities.

An essential ingredient in rehabilitation is hope. Staff must believe that improvement is possible and communicate that belief to the patient. Many mental health professionals see symptoms rather than potential and underestimate the patient's capacity for rehabilitation. Usually they think of work when they think of rehabilitation. In our society, work contributes a great deal to one's identity and self-esteem and is thus important. Yet competitive, mainstream, and full employment need not always be sought. If the job stretches the client's capacities, it might be part time, noncompetitive, or even volunteer and still build self-esteem.

Rehabilitation also addresses other skills necessary for survival in the community. These skills may include tolerating others in near proximity, the ability to dress appropriately, polite social interactions, keeping appointments, preparing one's own food, and grooming. The development of any one of these skills may seem mundane, but acquisition of a few may completely change a person's life. Being able to exist in the community without being shunned or appearing bizarre is essential to rehabilitation.

Integrated Services Combining all treatment, rehabilitation, case management, and support services into one agency, so that patients have a single source for all of their needs, is now considered the most efficient way to serve persons with severe mental illnesses. Successful programs are driven by patients' needs. That means that the patients define their needs rather than the staff, although staff do assist in the process. Any treatment goal not shared by the patient is doomed to failure. An integrated service agency must also have the flexibility in the use of its resources to provide what the patient needs rather than some predetermined product. Integrated service programs care for their patients regardless of the level of service they require, ensuring continuity of staff relationships and services. They maintain close relationships with the patients' families, residential facilities, and even employers if appropriate. Employment readiness training, job-finding assistance, housing assistance, ensuring that patients receive their appropriate benefits, and social and recreational opportunities are all key elements of an integrated service program. Because these are costly programs, they are usually reserved for persons with severe illnesses. They function best if provided under a capitated funding system that allows the agency complete freedom in the use of its resources and evaluates the agency on patient outcomes.

Administration Administration is the most essential but least loved component of a public mental health system. It includes establishing the vision, goals, objectives, and implementing plans of the mental health service program, nearly always through complex interactions with a wide array of individuals and agencies with competing interests. Administration requires individuals with strong clinical backgrounds, a vision of what the service should be, strong leadership skills, and a working knowledge of such technical aspects of administration as personnel, budget, fiscal and supply operations, and information systems.

Maintaining public support for mental health services is essential because they depend on public money; thus the public must be continuously educated about mental illness and the importance of public mental health services to the general well-being of the community. Less sophisticated citizens may be concerned only with public safety and decorum, not easy items to ensure. More altruistic citizens will also expect humane care and rehabilitation. Knowledgeable persons who have friends, children, and other relatives in the system have specific demands, some of which are unique. The mental health professional community, both within and outside of the system, also has specific ideas about what the system should be doing for which patients. Reconciling these usually competing demands, especially in the face of limited resources, is a Herculean task. It demands leadership, tact, diplomacy, and the ability to conceptualize and sell an integrated, comprehensive program to a large, skeptical audience. Such skills are in short supply in any one individual, so a team effort is essential.

PSYCHIATRIST REQUIREMENTS

Training An effective psychiatrist in the public mental health system must first be competent and well trained. No area of basic training can be neglected because patients in all diagnostic categories, all age groups, and in a wide array of predicaments will be encountered in any public mental health system. The patients require sophisticated interventions ranging from somatic therapies to insight-oriented psychotherapy. A thorough knowledge of both psychopharmacology and establishing

trusting relationships with patients that allows them to accept one's medication recommendations is critical.

Persons interested in working in public systems should observe the workings of the institutions in which they are trained. How the programs function affects both the patient care provided and the training. Observing the complex interactions of the various professionals in different administrative layers can be highly educational and requires little of the trainee's time. Hospitals, where one spends a great deal of time during residency training, are microcosms of the far more complex social and administrative system one enters in public psychiatry.

Special Interests and Knowledge The most important attribute of a good public system psychiatrist is a deep empathy for and interest in persons with severe mental illness. Intellectual fascination with the complexities of schizophrenia, including research interests, is important for the public psychiatrist's dedication to that difficult population. One should also have similar interests in the major mood disorders and the most disabling personality disorders.

Public mental health systems today must handle the treatment of substance abuse, both primary illnesses and those secondary to other mental disorders, and thus require personnel (including psychiatrists) who are skilled in dealing with these problems. Substance abuse is complex and multidetermined and different patients require different types of interventions. Psychiatrists are now required to have greater sophistication regarding all types of substance abuse, regardless of their practice settings. Because many traditional substance abuse programs abhor the use of any drugs, including psychotropics, they may not be appropriate for patients with both mental illness and substance abuse problems. Thus psychiatrists in public mental health systems may be called upon to help design special programs for these dual-diagnosis patients.

Since psychosocial rehabilitation must now be an integral part of any public mental health system, psychiatrists must be knowledgeable in this area. The public psychiatrist may not be a direct participant in many of the psychosocial rehabilitation activities, but knowledge of their importance to the patient's recovery is essential. When psychiatrists do participate directly in rehabilitation programs, they invariably see patients from a new and valued perspective. Every opportunity to observe patients in real work situations should be seized. Willingness to leave the office, meet patients in their workplace, and help with mundane problems give valuable assistance to the rehabilitation process.

Family and Social Systems An effective public psychiatrist must be willing to work with the families of mentally ill persons in a caring, nonjudgmental manner. Decades of having been blamed inappropriately for the illnesses of their children cause parents to be wary of psychiatrists and other mental health professionals. Yet they are frequently the strongest support systems that patients have in the community. Seeing the deterioration in their children and siblings has caused families excruciating pain. Many have done, and will continue to do, all in their power to help the ill member of the family. At times, those efforts may seem overly intrusive and even counterproductive. Yet the psychiatrist must recognize the basic good intentions and help all members of the family cope with a difficult situation. Parents generally wish to be allies in the treatment of their children and are willing to accept advice about how best to cope with their difficult children. Blaming parents for the illness of their children is totally inappropriate.

Families are hungry for information about the diseases that affect their members. A psychiatrist must be willing to educate family groups in any way possible–speeches, individual meetings, written materials, or telephone conversations. No opportunity should be missed. Families seek information not only about the diseases, but also about community resources and current research in the field. The Alliance for the Mentally Ill, a national association of families, is a sophisticated organization that plays a vital role in the education of both families and professionals and is a strong advocacy body for persons with severe mental disorders.

A working knowledge of social systems and their effects on persons with mental illness is necessary. An understanding of how school systems deal with deviant behavior is particularly important. School systems, like all such social agencies, are variable and unique to their specific region or community. A psychiatrist must be willing to look at each system individually and ascertain how that particular system will affect young patients. As patients age, the number of social agencies with which psychiatrists interact expands exponentially. They include welfare agencies, housing authorities, landlords, police, courts, probation systems, and even transportation and urban renewal. Some public psychiatrists work with many of these agencies to change their methods of doing business to ease the interactions of all mentally ill persons. Others intercede only when their patients have difficulty interacting with the social system. In either case the public psychiatrist must know the functions of interacting community agencies and the attitudes of their personnel. Offering consultation services to social agencies in one's community is an excellent way to gain an understanding of the social services available in one's community. There is helpful literature in the area of community consultation, and experience rapidly brings confidence. Most rewarding is the fact that the consultant nearly always learns as much or more than the consultees.

A psychiatrist who wishes to be a leader of a public mental health program needs a view of social systems beyond that of specific social agencies. This view comes best from being an astute observer of social and political trends and understanding the impact these trends have on attitudes toward mental illness. For example, shifts in attitudes about the treatment of criminals are usually echoed in attitudes toward people with mental illness. Economic conditions and policies certainly affect the availability of mental health and other services essential to patients. The country continuously undergoes shifts in political philosophy; psychiatrists may not be able to affect them, but they must understand them to protect their programs and patients.

Epidemiology Epidemiology has taught us a great deal about the distribution of mental illnesses in communities; that information can be of valuable assistance in planning public mental health service systems. Epidemiological data gathered by the NIMH Epidemiologic Catchment Area studies and the National Comorbity Survey are used to determine the need for public mental health treatment resources in various populations. The interaction of those data with specific communities' particular ethnic and socioeconomic makeup is very complex. A psychiatrist familiar with the psychiatric syndromes, the screening instruments used to identify those syndromes in the population, and the effectiveness of those instruments when used in various ethnic and socioeconomic populations can be helpful in their application to a program or community.

Psychiatrists must know the communities in which they work. That knowledge can be gained through careful inspection of census and other demographic data and by spending time in its neighborhoods. With this knowledge, one can quickly determine whether projections of need made from epidemiological data are correct. The community at large or special populations are frequently surveyed to assist in the planning of local mental health services. Since these surveys are often undertaken with small budgets and limited expertise, one must appreciate both their limitations and their usefulness. A psychiatrist should understand the basics of the application (and misapplication) of statistics and questionnaires.

Leadership The knowledge and skills needed to carry out the functions defined under administration are many and varied. No discipline has a monopoly on high-level management positions; indeed, some are held by lay administrators. However, the vast majority of directors of public mental health programs are members of the major mental health professions. For any discipline, special skills are required to achieve high-level management and leadership positions. Some of these skills are best gained by additional training, some can only be acquired by experience under the tutelage of a good mentor, and others require a combination of experience and education. One can gain the necessary education without pursuing further degrees in public administration or management, but such degrees can be helpful. It is often more efficient to undertake an appropriate reading program and avail oneself of the many short courses on specific topics that are available in most communities.

Public mental health services must be collaborative efforts because of the nature and size of their treatment populations. Modern psychiatric training should take place in institutions where those in other disciplines are also being trained. Shared training experiences help the disciplines work together, but mutual respect is essential. Breadth of technical knowledge does not ensure any professional a particular role in a mental health treatment program beyond those responsibilities defined by training and licensure. Other roles must be earned. Certainly psychiatrists have a distinct advantage, since theirs is the only discipline trained in all aspects of diagnosing and treating mental illness. However, to achieve leadership, they must demonstrate their competence like anyone else. That competence must go beyond the technical areas to encompass interpersonal relations, empathy with patients, the ability to teach, and a willingness to help with whatever needs to be done.

Psychiatrists do not allow themselves to be relegated to only a psychopharmacologist role. Because of the relative cost of psychiatrists and the critical function of psychotropic medications in the treatment of severe mental illness, some systems try to use psychiatrists only for medication management. However, one cannot be a good psychopharmacologist without understanding the patient's functioning in all aspects of life. That knowledge cannot be gained in a 15-minute interview in an office. The psychiatrist must learn to use broader interactions with both patients and colleagues to determine the effects of medications and other interventions on the patients. Whether or not patients take prescribed medications is greatly affected by their relationship with the prescribing psychiatrist.

Management Skills A number of other skills required to manage public mental health systems are rarely discussed or taught. To be a successful public mental health administrator, one must understand how communities are organized, how they choose their leadership, and how the formal and informal power structures work. Clinicians who are administrators must be able to frame their programs to be acceptable to the powers that be. Without that acceptance, the finest program

conceptualizations will go for naught. One must be aware of prevailing public attitudes, where support can be found, the opinions of key community leaders, and when one is wasting time. Public speaking and salesmanship skills are helpful, and neither is learned in most psychiatric training programs.

The ability to work with and respect lay leaders is essential. One must communicate without technical jargon and respect all political philosophies. In public systems politicians change, and administrators must remain nonpartisan to provide continuity of programs in the face of changing political control. Unfortunately, some elected officials feel that all managers, like themselves, are or should be partisan. This attitude can lead to disruptive management changes when political control changes. A leader must know when to resign for the good of the program.

Technical Skills Political skills and agility may be hard to learn in a formal manner, but the skills that concern public administration and management are not. They deal with budget, personnel management, organization, control, and leadership. To manage a public mental health program one must learn both the budget process and how to read and manage the final product. The budget is where all plans for the program come to fruition or fail; planning and budgeting are inseparable. Every governmental jurisdiction has its own budgeting process, but they are in fact quite similar and are easily understood if one simply gets to know the major players and participates actively in the process. A rapidly evolving technical area that requires management attention is the use of computers and information systems. A manager must be able to use the full potential of the information system and know whether that system is state of the art.

Managers must understand and devote time and energy to personnel matters. Careful selection of the management team is essential. Top administrators must recognize their own shortcomings and choose staff that complement them. Insecure administrators tend to choose weak, nonthreatening managers. Strong administrators tend to choose the most competent managers they can find, but they must also fit the organization. The working climate must be open so that all staff can carry out their functions and speak what they believe to be the truth without fear of reprisals. Those who raise the difficult questions about the organization are most likely the strongest members of the management team; however, those who only raise questions and never offer solutions may be less than helpful.

Leadership is a critical quality but it is very difficult to define or teach. It involves both the knowledge of what needs to be done and the ability to communicate that knowledge to a wide array of people. It includes the ability to see how what happens in one part of the organization will eventually affect all other parts, so one can balance competing interests. The ability to see the big picture and act accordingly has been described as the ability to think in general systems theory. Unfortunately this ability may be inborn rather than learned.

EXTERNAL FORCES

Shifting Governmental Roles Mental health services are expensive and are not politically popular in many jurisdictions. In the United States there is a continuous effort by the federal government, the states, and local governments to pass costs to each other. This battle is clearly evident in many of the Medicaid and Medicare rules, exceptions, and distortions; welfare reform; the passing of service responsibilities with or without funds to local governments; and attempts to cut taxes at the federal level by funding fewer state and local programs. These shifts represent both danger and opportunity. Often a shift in funding sources lets one update programs by getting out from under onerous regulations and fiscal strings. On the other hand there is the grave danger of getting new responsibilities without the resources to meet these obligations.

The care of persons with mental illness is not uniform throughout the United States, and persons with chronic mental illness are a highly mobile population. They cannot expect the same or even similar services in various parts of the country. Even financial eligibility for service varies widely. Whether these factors affect patient's choices of where to live has not been determined.

Economic Forces The community's willingness to support mental health and social services is directly related to the health of the general economy. When the national economy is poor, tax revenues fall and are diverted to other projects. In those times public mental health programs are often cut back even though the demand for service increases. In many jurisdictions mental health services are soft priorities for funding, that is, the general public considers them desirable services but, because the taxpayer does not expect to use them, they will not be funded in stiff competition with other priorities. New mental health services can only be funded in good economic times.

The private sector has a major impact on the public sector. Here two issues are involved—the availability of private-sector services and the means to pay for them. The unique services (especially rehabilitative ones) needed by persons with serious mental illness are often unavailable in the private sector, and even persons with the means to pay must come to the public sector. More important is the availability of third-party coverage. The tremendous expenditures necessary to care for someone with schizophrenia are often unmanageable, even for upper-middle-class families, if they are not covered by insurance. Insurance coverage for mental health services has been one of the casualties of the rapid rise in health care costs in the United States. The stringent limits on insurance coverage for mental illness has brought about the current demand for parity of coverage with that for illnesses of other body systems. While some small moves toward parity were made at the federal level, the real battles are being fought at the state level with very mixed results. Should real parity come about, demands on public mental health systems may be lessened.

Managed Care The nation's response to the intolerable escalation of health care costs has been managed care. Under managed care a number of mechanisms are used to limit services to those the payer considers absolutely necessary. In addition, costs are lowered by negotiating lower rates and capitation contracts with selected providers and even by limiting the availability of certain services. Managed care has had a massive impact on both parties and providers in the private sector, forcing many patients to turn to the public sector for longer-term care and rehabilitative services not offered by many managed care entities.

While managed care has had a remarkable impact on the costs of health care in the private sector, its impact on quality and outcomes is quite controversial. Many leaders of mental health managed care companies came from the public sector, and they applied rationing techniques that have been prevalent in the public sector for ages. With the looming saturation of the private managed care market, a number of managed care companies are now looking to contract with states and local governments to manage the health care services to the poor (predominantly the Medicaid population) under recent health care reform legislation.

Leaders of public mental health care systems are concerned about managed care entities taking over the mental health needs of the Medicaid population for a number of reasons. Even though most contracts have differential rates, there is great concern that healthy patients will be skimmed and sicker patients left to a residual, largely unfunded, public sector. Managers of public mental health services have been distributing limited resources to the public patient population for a long time, and they know that there are not enough resources available to service people properly. Thus, they are alarmed about what will be left after management fees, profits, and stockholder dividends are taken out. Perhaps the greatest fear is that these private companies will make attractive offers to local governments to manage their mental health services, replace the existing local mental health management team and at least some providers, and then discover that under the fiscal and administrative restraints of governmental programs, there are no profits to be made. They would then pull out of their contracts, leaving chaos and a service vacuum in their wake. Many experiments in public sector managed care are currently under way. They must be watched very carefully.

Social Attitudes Mental health professionals can do little to influence social attitudes that depend on economic conditions, political leadership, and world affairs. Yet, social attitudes have a greater impact on the delivery of public mental health services than most people realize.

They dictate whether institutional or community programs are supported, determine who is to be cared for in the public mental health systems, and influence the amount of freedom allowed to the clientele. The public's attitude toward criminals and other social deviants is strongly correlated with their attitude about persons with severe mental illnesses. While it is important to try and separate mental illness from other types of social deviancy in the minds of the public, public mental health programs and their leaders probably cannot affect major social attitude changes. Instead, they must attempt to adapt, while preserving their service mission and humanity. Fortunately, humane and effective mental health services can be provided in many different settings if the needs of the patients are kept foremost in the minds of the professional staff.

Mental health professionals must constantly guard against having their patient care responsibilities corrupted by political demands or social fads. It is often difficult to know when the program is appropriately adapting to changing social needs or values and when it is neglecting its responsibilities to its patients in service to an alien ideology.

Human Resources Psychiatrists with the training, aptitude, and interests for the public sector are in short supply in most regions of the country. Many of our psychiatric training programs use public facilities, but little is done to help the resident psychiatrist appreciate a future in the public system. Instead, training is often seen as a period of indentured servitude in public hospitals, a rite of passage to the private sector. This attitude, unfair to both the public systems and the trainees, is often perpetuated by their role models in academic psychiatry. Failing to understand the larger system causes the trainees to get a distorted picture of chronic mental

illness (e.g., during most of their lives, patients with schizophrenia are not as dysfunctional as when they are hospitalized) and of the programs that can profoundly affect the lives of patients outside the hospital. The impressions that trainees gain from dealing with only the most frustrating, difficult patients understandably blunt their interest in public service later in their careers.

One benefit of the current move toward managed care in the private sector is an increased interest in positions in the public sector by young psychiatrists. That increased interest has generated greater demands for training experiences in psychosocial rehabilitation, in correctional facilities, and with seriously ill populations.

CURRENT ISSUES

Homeless People The homeless population is a heterogeneous group with multiple diagnoses; a high percentage suffer from substance-related disorders. They can be categorized into three groups: (1) persistently homeless persons who have a combination of schizophrenia and substance-related disorders, a history of psychiatric hospitalization, and multiple health problems; (2) episodic homeless persons whose psychiatric diagnoses are more likely to be serious personality disorders or mood disorders with a complex treatment history; and (3) situationally homeless persons who suffer from stress rather than a primary psychiatric or substance-related disorder. Persistently homeless persons resist all treatment settings that force them into congregate living or limit their behaviors. Therefore getting or keeping them in traditional treatment settings is usually not possible. Instead, reaching them in their natural settings with medications and support to develop trust and make their lives on the streets more tolerable is a better strategy. Episodic homeless persons present the most complex treatment challenge. They require wraparound services—psychosocial rehabilitation, psychiatric care, community living support—with aggressive case management. Situationally homeless persons require primary social service intervention with secondary psychiatric support.

Underserved People Members of minority populations, especially those who speak English as a second language or who do not speak English at all, require special consideration. Bilingual and bicultural staff are essential to relate to those populations. This is not an easy requirement to meet, since few such professionals have been trained, and this problem may be exacerbated by the recent trend of limiting minority quotas and affirmative action programs in educational institutions. Recent immigrants from underdeveloped countries find mental health services an alien concept. In their homelands they depended on support systems that are frequently not available in the United States. Better understanding by mental health professionals of cultural and familial styles and values is essential to remove these obstacles to their getting care.

Others who are underserved include elderly adults and persons residing in rural areas that offer few mental health services. Both of these groups tend to suffer quietly; their mental illness does not come readily to the attention of others. To reach them, mental health programs will have to work more with other practitioners of medicine—gerontologists, family practitioners, and internists.

Outpatient Commitment The recognition that some persons require medications to remain peacefully in the community has led to a number of proposals for laws to allow involuntary outpatient care. Some might say that the concept is an oxymoron; others feel that it should be tried. It is true that persons on probation or parole who must be in psychiatric care as a condition of their release are more likely to stay in treatment. However, these individuals face the very real threat of being returned to prison or jail to serve out a sentence for a crime they committed.

In the case of persons with mental illness, the situation is different. The strongest threat one would have is returning the patient to a hospital, something all parties are trying to avoid. Furthermore, it might be unconstitutional to detain individuals who are not at that time gravely disabled or a danger to themselves or others, even if they are not complying with treatment. Rehospitalizing an individual for not complying with treatment is equivalent to making noncompliance with treatment a crime. Many feel that such a policy would further criminalize mental illness—something to be avoided.

An alternative to outpatient commitment, conservatorship, is used in many states. Conservators—usually not part of the treatment system and often family members—are given responsibility for the patients' well-being and varying amounts of authority over their life, up to and including the ability to place them in a locked facility if their condition demands such placement. Of course, treatment staff must agree that the patient needs to be admitted for medical care before that individual can be admitted. Thus these patients are being admitted because of their condition, not as punishment for not complying with treatment.

Self-Help Rehabilitation has proceeded hand in hand with other patient empowerment efforts—the self-help and client movements. Patients usually see self-help programs as adjuncts to traditional treatment programs that support the use of psychotropic drugs and regular treatment contacts. Often they are organized by traditional agencies such as mental health associations, although the clients are expected to govern the programs to the extent possible. These programs are valuable in the psychosocial rehabilitation of persons with chronic mental illnesses. They give patients an opportunity to test social skills and experiment with control of their own lives in an accepting environment not available in regular society.

The client movement is difficult to define and characterize. It ranges from self-help groups to organizations of former patients with radically antipsychiatric belief systems. Most client organizations have been constructive and have performed invaluable services in sensitizing professionals to the inhumanities inherent in some treatment programs. Those on the radical fringe, however, have often impeded services to patients in dire need of care. Often the leaders of these more radical groups have themselves been mistreated; they were labeled psychotic and treated with antipsychotics when their problems were primarily personality disorders. They then perceive all patients to be like themselves, a tragic and destructive assumption. In any case, mental health professionals should do their best to help all of the client movements play a constructive role in improving the system and their own members' rehabilitation.

Working With Families Many families remain wary of involvement with the mental health system regardless of mental health professionals' efforts to rectify past errors. In addition, many patients have severed relationships with their families, and because of the trauma they suffered, parents and children are reluctant to reestablish relationships. Reestablishing family relationships is often critical to treatment and rehabilitation.

Families who are involved with their children are tremendous assets to rehabilitation. They provide a wealth of knowledge regarding the patient's treatment history, including what works and what does not. They are an invaluable support system for the patient that is difficult or impossible to replicate. Although it is frequently unwise for the patient to continue to live at home beyond an age when young people usually leave, a concerned family is still an asset.

Psychiatrists must recognize that families have borne a tremendous burden for years with little or no support. That support must now be provided on an ongoing basis. Families who are involved with their children deserve to have readily available professional backup. The love and emotional support that parents can offer their child is far more important than playing the role of caretakers.

Parents have a right to know everything possible about the illness—the prognosis, realistic expectations, and its most optimistic potential resolution. In addition they need to be kept informed about current research and possible implications for their children. The success of the Alliance for the Mentally III shows that well-informed parents constitute one of the strongest forces for improvement in the system of treatment for serious mental illnesses and research to seek cures.

Resources Continuous underfunding is the curse of public mental health systems. The situation ebbs and flows, subject to external economic and political pressures, the ability of the systems to use available resources efficiently, and the pressure brought by organized constituency groups for greater resources. Lobbying is critically important. Successful lobbying requires a coalition of all advocacy groups and professional associations that have a stake in the outcome. Petty differences must be put aside to allow agreement on a single national agenda.

The National Mental Health Association has been a major force for the improvement of care for people with mental illness and the protection of their rights for decades. It is organized with chapters in states and metropolitan areas and is active at all levels where people need support and protection. Its primary function is advocacy for the rights of patients, both civil and treatment; however, its chapters frequently provide social and work programs and housing and other support services for people with mental illnesses.

Much of the effort to improve funding for both services and research is necessarily undertaken by professionals in the field. Most public mental health leaders spend disproportionate amounts of time fighting to obtain an adequate budget, as do the managers of their various program divisions and contract agencies. However, many of these efforts are undermined by the accusation by funding bodies that managers and employees of the provider agencies are simply feathering their own nests. For that reason the agency personnel must work with advocates, professional organizations, the Alliance for the Mentally Ill, patient networks, and any other support groups to present a united front to the decision makers. History shows that programs prosper only when such coalitions are present.

COLLABORATION OF PUBLIC AND ACADEMIC INSTITUTIONS

Academic institutions and public mental health programs are mutually dependent on each other. Public mental health programs need well-trained psychiatrists and other mental health professionals; academic institutions need clinical training facilities and research populations. However, issues of money and control have strained this relation. Public programs fear that the academic institutions are only interested in them for their money and other resources; academic institutions feel that they must exercise control over programs in which either research or training is conducted. In addition, public institutions fear that research projects will jeopardize their reputation or their patients.

Academic institutions have been important in the evolution of public psychiatry. They have undertaken monumental epidemiological studies, operated some outstanding community mental health centers, trained some personnel with unique skills for public psychiatric work, and researched service delivery methods. Yet many argue that academic institutions frequently shirk service delivery that does not directly feed their needs and fail to give their trainees the skills and motivations necessary to work in the public sector.

Doing more training in well-developed public programs would give trainees more of the needed skills and interests. The benefits to the universities beyond training sites and research materials are less tangible but perhaps more important. They can be summarized as image and relevance. Publicly funded institutions must be seen as having special value to the community beyond the benefits to their students and faculty to maintain their levels of support. Public services can enhance that image as readily as research. Similarly, active involvement in public mental health services is one sure way to keep the university relevant to the needs of the community.

The shift of research interests toward the biological and away from psychosocial areas has diminished academic institutions' interest in public mental health programs. This trend should not be allowed to continue. The advance of organized systems of health care has made the interdependence of public mental health systems and training institutions greater than ever.

FUTURE DIRECTIONS

Public mental health systems will exist until schizophrenia, the severe mood disorders, and the more disruptive personality disorders can be prevented or easily cured. Society reluctantly accepts its responsibility to people with mental illness and will continue to address the problem. Current trends indicate that in the next decade much more mental health service will be provided in correctional institutions because of the increasing propensity of society to jail mentally ill persons committing crimes, the lack of adequate mental health treatment facilities in many urban areas, substance abuse by mentally ill persons, and the demand for psychiatric care for persons with severe personality disorders. Prisons are one of the great growth industries of our society, and their mentally ill inmates will increase concomitantly. A fascinating aspect of public psychiatry is its constant evolution.

SUGGESTED CROSS-REFERENCES

The other sections in this chapter on economics, the role of the psychiatric hospital, and psychiatric rehabilitation are of obvious importance to persons interested in public psychiatry. In addition, all of Chapter 12 on schizophrenia is of special interest, especially Section 12.9 on its psychosocial treatment. Chapter 13, Chapter 14, and Chapter 24 dealing with other psychotic disorders, mood disorders, and personality disorders, respectively, are also important since these disorders provide many of the public sector's patients. Section 5.1 discusses epidemiology. Finally, anyone involved in public psychiatry will be interested in the comprehensive coverage of managed care in Section 52.2.

SECTION REFERENCES

*Anthony WA, Liberman RP: The practice of psychiatric rehabilitation: Historical, conceptual, and research base. Schizophr Bull *12:* 542, 1986.

Bray JD, Bevilacqua JJ: A multidisciplinary public-academic liaison to improve public mental health services in South Carolina. Hosp Community Psychiatry *44:* 985, 1993.

Caplan G: *Principles of Preventive Psychiatry*. Basic Books, New York, 1964.

*Chandler D, Meisel J, Hu T, McGowen M, Madison K: Client outcomes in a three-year controlled study of an integrated service agency model. Psychiatr Serv *47:* 1337, 1996.

Cnaan RA, Blankertz L, Messinger K, Gardner J: Psychosocial rehabilitation: Toward a definition. Psychosoc Rehabil J *11:* 61, 1988.

Cohen CI: Down and out in New York and London: A cross-national comparison of homelessness. Hosp Community Psychiatry *45:* 769, 1994.

Cuffel BJ, Snowden, L, Masland M, Piccagli G: Managed care in the public mental health system. Community Ment Health J *32:* 109, 1996.

Essock S, Goldman H: States' embrace of managed mental health care. Health Aff *14:* 3/34. 1995.

*Foley HA, Sharfstein SS: *Madness and Government*. American Psychiatric Press, Washington, DC, 1983.

Glasscote RM, Gudeman JE, Elpers R: *Halfway Houses for the Mentally Ill*. Joint Information Service, Washington, DC, 1971.

Goffman E: *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. Doubleday, Garden City, NY, 1961.

*Griswold KS, Goldstein MZ: Issues affecting the lives of older persons with developmental disabilities. Psychiatr Serv *50:* 315, 1999.

Harding CM, Zubin J, Strauss JS: Chronicity in schizophrenia: Fact, partial fact, or artifact. Hosp Community Psychiatry *38:* 477, 1987.

Joint Commission on Mental Illness and Health: *Action for Mental Health*. Basic Books, New York, 1961.

Kaufmann CL, Ward-Colasante C, Farmer J: Development and evaluation of drop-in centers operated by mental health consumers. Hosp Community Psychiatry *44:* 675, 1993.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshelman S, Wittchen H, Kendler KS: Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Arch Gen Psychiatry *51:* 8, 1994.

Lamb HR, editor: *The Homeless Mentally Ill*. American Psychiatric Association, Washington, DC, 1984.

*Levin BL, Petrilla J, editors: *Mental Health Services, a Public Health Perspective*. Oxford, New York, 1996.

Mirabi M, editor: *The Chronically Mentally Ill, Research and Services*. Spectrum, Jamaica, NY, 1984.

Mrazek PJ, Haggerty RJ, editors: *Reducing Risks for Mental Disorders. Frontiers for Prevention Intervention Research*. National Academy Press, Washington, DC, 1994.

Parsons T: *Social Structure and Personality*. Free Press of Glencoe, New York, 1964.

Petrila J, Ayers K: Enforcing the fair housing amendments act to benefit people with mental disability. Hosp Community Psychiatry *45:* 156, 1994.

Robins LN, Regier DA, editors: *Psychiatric Disorders in America. The Epidemiologic Catchment Area Study*. Free Press, New York, 1991.

Stein LI, Test MA, editors: *Alternatives to Mental Hospital Treatment*. Plenum, New York, 1978.

Sullivan HS: *Fusion of Psychiatry and Social Science*. Norton, New York, 1964.

*Talbott JA, editor: *The Chronic Mental Patient, Five Years Later*. Grune & Stratton, Orlando, FL, 1984.

Talbott JA: Has academic psychiatry abandoned the community? Acad Psychiatry *15:* 106, 1991.

Talbott JA, Robinowitz CB, editors: *Working Together: State-University Collaboration in Mental Health*. American Psychiatric Press, Washington, DC, 1986.

Weisburd DE, editor: The village, a special edition. J Calif Alliance Ment Ill *4:* 2, 1993.

Winters EE, editor: *The Collected Papers of Adolph Meyer*. Johns Hopkins University Press, Baltimore, 1952.

52.2 MANAGED CARE

ROBERT JEAN CAMPBELL, III, M.D.

DEFINITION

Managed care is an organized system of health care providers whose goal is to deliver needed services in an effective, responsible, and cost-efficient manner. It is generally assumed that any such system will also provide increased access to appropriate primary health care. Managed care differs significantly from what most people understand as health insurance, that is, the indemnity fee-for-service system in which a person buys an insurance policy to protect against the unforeseen cost of falling sick and needing care. Managed care organizations consist of networks of providers, such as physicians, hospitals, clinics, and day programs, that often share the financial risks associated with providing care to all of those enrolled in the plan whenever they need it.

Managed care includes an evolving array of cost-containment mechanisms that combined with different types of provider networks, reimbursement systems, provider incentives, plan sponsorship, and gatekeeping design that can create a diversity of plans tailored to employer characteristics and employee demographics. The managed care systems that now predominate are health maintenance organizations (HMOs) and preferred provider organizations (PPOs). Individual HMOs may offer more than 400 different health, mental health, and substance abuse benefit structures, based in part on the specific needs and requests of the individual employer groups.

HISTORICAL DEVELOPMENT

The basic elements of managed care—capitation and integration—had developed in the United States by the 1800s, in the funeral and benevolent societies founded by immigrants to cover their death expenses. Those early models of managed care organizations were developed as a way to fill unmet health needs, not as a way of containing costs, dominating a market, or turning a profit. Since then, the situation has changed radically, in part because the U.S. government, through health programs introduced in the 1960s, became the single most important purchaser of health care.

Only 5 years after Medicare was introduced in 1966, the Nixon administration became alarmed by an unanticipated rise in its cost. Despite mounting health care costs, the rate of uninsured individuals continued to swell, threatening to topple a fragile health care delivery infrastructure. Even for many with insurance, coverage was inadequate to pay the costs of a serious illness. Less money is coming into Medicare, and it is estimated that the fund's reserves will be exhausted by 2001. Without some sort of change, the system may be unable to support the scope of coverage currently allowed.

COSTS OF HEALTH CARE

Development of the managed care system was part of the response to seemingly uncontrollable rises in health care costs. In 1993 business alone was spending over $279 billion yearly ($3781 per employee) on health insurance premiums and payroll tax contributions to the Medicare Hospital Insurance Trust Fund. Arguably, American medicine has become the world's largest business.

In 1970 the government introduced the concept of a health maintenance strategy that allowed beneficiaries to choose between HMO and traditional fee-for-service systems. Not long after, employer purchasers began to encourage their employees to join health plans, whose attractions to the consumer were lower out-of-pocket costs and ease of use.

HMOs were designed to be of two forms: prepaid group practices (such as the Kaiser Permanente Health Plan) and independent practice associations. As of July 1, 1994, there were 593 HMOs in the United States. Nearly three fourths of all enrollees were in independent practice associations, in which physicians, hospitals, and other entities contract with a health plan, or in mixed models that involve more than one structure.

In the current era of managed care, the players on the health care field are not necessarily the willing partners they once were; they are just as likely to be the product of health care power politics, each competing for dominance, control, and ultimately as much profit as can be squeezed out of the system.

PHASES IN THE DEVELOPMENT OF MANAGED CARE ORGANIZATIONS

In the early stages when there is little or no managed care in the area, the medical system is “overbedded” and relies heavily on the use of specialists. Managed care first tries to control costs by limiting the delivery of health care services in inappropriately costly settings. In the next phase, health plan administrators adopt more comprehensive utilization management techniques, and physicians become more fully integrated into managed care plans. As managed care begins to penetrate, discounted fee-for-service is initiated, and plans for consolidation of health services develop.

In these generations of managed care, the emphasis is on managing the utilization of health care services, not on managing the care itself. Any cost savings achieved come primarily from providing care in less costly settings or by rationing services. Taken to the extreme, this approach might indiscriminately reduce both appropriate and inappropriate care.

By the time managed care penetration achieves 30 to 40 percent in an area, reimbursement within a given market declines significantly. Consolidation among hospitals, physician groups, and other large health care entities becomes more prominent. Redundant beds are closed, and there is increasing effort to end duplication in administrative and clinical systems. Then the health care system must find the capital necessary to acquire primary care or other essential groups.

Development of managed care organizations has not been a unified national movement. Different parts of the country are at very different stages of managed care organization development. Because of their variety, their continuing evolution into new forms, their different phases of development in different parts of the country, and the role they play as potentially large contributors to the profit side of big business, managed care organizations are not best understood through the peer-reviewed journals with which most physicians are familiar. They change so rapidly and involve such large amounts of money that their modifications and development are more likely to be reported in the business sections of newspapers and trade journals of the hospital and managed care industries than in professional medical publications.

MAJOR CHARACTERISTICS OF MANAGED CARE ORGANIZATIONS

As noted, no single model can describe the rapid changes in the field. Nonetheless, some models are used more than others, and some common strategies can be identified. Mainstream managed care organization features include (1) a claims payment system; (2) competition; (3) provider networks, particularly medical care systems that strictly control resource consumption; (4) gatekeepers; (5) shifting economic risk from payer to patient and provider; (6) an enrolled population; (7) management and information systems; (8) monitoring provider performance; and (9) influencing patient-care decision making.

Provider Networks The first step in implementing a managed care organization plan is to assemble a network of providers. Networks must be large enough not only in terms of the actual number of hospitals and physicians—primary care physicians and specialists—but also in their willingness to use midlevel practitioners to expand the primary care base. A typical rule of thumb calls for one primary care physician for every 2000 enrollees (and, often, one psychiatrist for every 10,000 covered lives), with primary care physicians making up 50 percent of the total medical staff. The ratio varies according to the characteristics of the population served.

Organization of Providers Once the providers have been selected, they must be organized along some kind of functional model. There are many different possibilities.

In the *staff model*, the professionals are either employed or exclusively retained by the HMO, managed care organization, or network.

In the *assessment and referral model*, assessors are employed or retained by the managed care organization. They provide initial diagnoses and treatment plans, which necessarily reflect their own theoretical orientation.

In the *central access network*, the patient contacts a central point. The managed care organization verifies eligibility, assigns the patient to a primary care physician in its network, and authorizes treatment. All care, including assessment, is rendered by community providers. Many self-insured employers use this model, and most *managed mental health care* or *alcohol, drug, and mental health* companies and insurers operate this type of network.

In *open access* provider panels, the patient may go to any approved provider at any time. The provider must review the case with the managed care organization and obtain continuing authorization after either an assessment or a set number of visits. (The participating provider should be certain to note which company requires what.)

In *opt-out* systems, patients may choose participating providers (in-network) or nonparticipating providers (out-of-network). If they choose the latter, certain increases in deductibles or copayments or reductions in benefits may apply. The financial penalties in an opt-out system may be so great that most enrollees cannot afford to remain outside the network.

Closed networks are those that accept no more providers into their panels. Within a network, the number of selected providers is typically much smaller than the actual number of providers in a given area, especially in metropolitan centers. A managed mental health care or alcohol, drug, and mental disease firm wants a relatively short list of providers. That allows the firm to reward its own providers with more referrals, thereby making it more likely that the providers will remain within the network of discounted fees. (The “any willing provider” laws under consideration in many states would eliminate this advantage.)

PHYSICIAN HOSPITAL ORGANIZATION Formal alliance between physicians and hospitals, usually created to contract for managed care and to increase the physician hospital organization's market share, physician hospital organizations often require capital investment by participating physicians, usually through some kind of arrangement that allows physicians to pay back their share of the initial capitalization over time.

Inherent in the physician hospital organization concept is the assumption that physicians and the hospital will share decision making. Leadership cannot be left in the hands of the hospital's administrative personnel, nor can it be shared by them with handpicked physician representatives who do not command the trust and respect of the medical staff.

PREFERRED PROVIDER ORGANIZATIONS Also called select provider organizations, PPOs are providers who have agreed to discount their fees to enrollees of the plan, meaning lower out-of-pocket expenses for people who choose them. It is the least restrictive form of managed care. The impetus for managed care organizations to contract with mental health providers presupposes the willingness of the providers to share in the company's vision of managed health care.

HEALTH MAINTENANCE ORGANIZATIONS HMOs are the most restrictive type of managed care organization and currently the most popular. HMOs developed early in the twentieth century as prepaid group practice plans. There are many forms, but they have some characteristics in common: assuming contractual responsibility to provide an integrated and comprehensive range of health and mental health services to a voluntarily enrolled population on a prepaid contractual basis. Until recently, it was believed that all HMOs eventually would be acquired by leading indemnity insurance companies. But it now looks as if HMOs will dominate the business, perhaps replacing insurance carriers entirely. *Point of service* plans are similar to HMOs but allow some reimbursement if a patient opts to receive care outside the provider network.

INDEPENDENT PRACTICE ASSOCIATIONS In this model, providers deliver ambulatory services in their own offices and maintain their own staffs; their fees are either discounted or capitated (paid in advance, based on the number of enrollees in the program). Independent practice associations are contracting networks of solo practitioners. HMOs capture their market share by offering a wide choice of physicians, and independent practice associations are a quick and inexpensive way to assemble a large network. Many multispecialty groups and hospitals form independent practice associations to expand their geographical coverage without expending large amounts in capital funds (it can cost tens to hundreds of thousands of dollars for a hospital to acquire a physician's practice).

The HMO contracts for inpatient care from community hospitals in its service area. An HMO may create an independent practice association to serve it and only it (a “married” independent practice association). Or independent practice associations may be formed to capture the business of many HMO and PPO products developed by the insurance industry; these “unmarried” associations can enter into relationships with many organizations simultaneously.

The essential task of this model is to preserve the desirable elements of autonomy of traditional practice and yet give the independent practice association the necessary authority for organizational change and governance. The independent practice association organizes the delivery of care. It contracts with HMOs and PPOs; it assembles, credentials, and inspects the network of provider physicians, institutions, and services; it constructs authorization and referral processes; it establishes gatekeeper and specialist responsibilities, sometimes including treatment protocols; it disburses payment to physicians and ancillary services; it conducts utilization and quality assurance reviews; and it ensures the fiscal integrity of the enterprise. No entity in medicine has ever before had so much power over independently practicing physicians.

Hospitals, because of their financial and organizational resources and their interest in generating admissions, were central to the initial formation of independent practice associations in California, an advanced managed care market with high HMO penetration. There have been exceptions, however, and throughout the country more and more independent practice associations are now owned and managed by physicians.

NETWORK MODEL Also known as an *integrated delivery system*, a network is an organization of primary care, specialty care, and hospital care services under one umbrella. In many respects, the organization of networks mirror that of HMOs except that inpatient facilities are a full partner with ambulatory components in ensuring quality services. Without such integration, inpatient facilities are often in a competitive bidding process to attract managed care dollars by discounting their charges. As an integrated, full partner with ambulatory resources in a network plan, inpatient providers no longer have filling beds as their primary goal but shift instead to appropriate utilization of inpatient and ambulatory capacities.

PROVIDER-SPONSORED NETWORKS Employers seek to extract additional savings by cutting out managed care organizations and turning to provider-sponsored networks, hoping that by dealing directly with providers they will manage healthcare benefits more cost-effectively and will trim the administrative expense of an intermediary. Some provider-sponsored networks are beginning to accept downstream risk from HMOs: the HMO bears full risk for a group of patients and pays the provider-sponsored network on a capitated or percent-of-premium basis to provide certain services.

DISEASE MANAGEMENT *Disease management*, or *disease state management*, is a newer clinical model that integrates, both clinically and fiscally, all health care services for a specific health problem. Its primary purpose is to identify and manage the 20 percent of patients who use 80 percent of total health resources.

The hospital, physicians and other professionals, and intensive case management are integrated into a single program. In behavioral health, disease management programs have been developed for depression and schizophrenia. Disease management gives the flexibility needed for early intervention, prevention programs, integration of clinical services and psychopharmacology, and improvement of outcomes.

MANAGEMENT SERVICE ORGANIZATION Not another integration model, management service organizations provide contract management or practice management services to physician organizations, and they add substantial capital to the group. Some management service organizations purchase the equipment, furniture, and supplies of medical practices and employ the office staff and physicians. The key component missing in most physician organizations has been the administrative component, and management service organizations provide a stong administrative infrastructure.

Primary Care Managers or Gatekeepers Both HMOs and point of service plans rely on gatekeepers, who are usually primary care physicians or, in the case of mental health and substance abuse, a specialty care firm. The gatekeeper is responsible for a longitudinal approach to patient care, integrating the clinical and psychosocial aspects, as well as subspecialty care, health promotion, and disease prevention.

The gatekeeper screens all patients and decides if referral to a specialist is needed; often, this includes a decision about what kind of specialist is appropriate (e.g., psychiatrist, psychologist, psychiatric social worker, or psychiatric nurse practitioner). Many plans carve out mental illness and substance abuse, using a managed mental health care or alcohol, drug, and mental health firm to handle them. The firm sets up a provider network and authorizes needed services. The companies are staffed by mental health professionals, usually nurses and social workers, who use written sets of criteria to determine what services are needed and for how long.

Shifting Economic Risk Benefit design and pricing strategies shift the economic risk from payer to patient and provider. In managed care organizations the contract is between the provider and the payer. If the patient complies with the restrictions imposed by the managed care organization, the cost of medical care will be less than that incurred outside the network of contracted providers that the firm has set up.

Pricing strategies include incentives to enrollees to use the plan's providers, and incentives to providers to be cost-effective in giving care, often using risk and reward options (sometimes called bonus funds).

Reimbursement of the Practitioner Managed care organizations use several methods to reimburse physicians for services. The provider contract should identify the method used and provide detailed information. The most common mechanisms are fixed-fee schedules, discounted fees, and capitation.

In a *fixed-fee schedule*, all physicians receive the same payment for the same procedures, based on either an average of fees in the area or a relative value scale. In a *discounted fee-for-service* arrangement, physicians in the plan agree to discount a certain percentage (usually 10 to 20 percent off their regular charge). In *capitation*, physicians are paid a certain amount per number of enrollees in the plan per month or per year, regardless of the amount of service provided.

Capitation is an incentive-based system of health care in which providers are paid a negotiated sum in advance to cover all required medical and mental health services for each enrolled member over a defined period of time. Under capitation, hospitals are not only financially responsible for their own services; they may also be responsible for services they do not provide themselves, such as tertiary hospitalization, skilled nursing care, home health care, and ambulance services—even though some of those services may be delivered by their competitors.

Many capitation contracts include a *withhold*—a percentage (often in the neighborhood of 10 percent) of the physician's fees are reserved until the nd of the contract or are paid only if the business is profitable. While a capitation system encourages providers to render medically necessary services, it also removes incentives to overuse health care resources. It creates a precarious situation for the risk-bearing provider, who receives a fixed sum for each person covered each month, regardless of whether any services are used. Capitation pits hospitals, specialists, and primary care physicians against each other in reaching a decision about what portion of the original capitation amount each one will get (the “split of the cap” problem).

Meantime, capitation continues to grow, and provider-managed capitated systems may ultimately replace fee-for-service medicine. In the independent practice association model, the managed care organization makes capitated payments to the independent practice association, which decides how and what to pay each physician and assumes some responsibility for managing the costs and quality of care provided by its physician members. Some independent practice associations control use by *contact capitation*—specialists are paid a standard rate for each patient seen, regardless of the number of follow up visits or procedures.

How the total bonus dollars are apportioned among the primary care physicians at the end of the year varies from one independent practice association to another. They may be divided equally among all the primary care physicians; or the referral costs generated by each primary care physician may be tracked, and those with lower referral costs are rewarded with bigger bonuses than the primary care physicians with high referral costs. In December 1996 the government moved to stricter control over bonus payments that could be seen as rewards to physicians for withholding or limiting care.

Specialty capitation is also growing; it reduces the specialist's incentive to provide more care, and by limiting the payout to specialists, it reduces pressure on the primary care physician not to refer. It reduces the risk for primary care physicians, but it also reduces their potential profit. In a system that includes specialty capitation, it is essential to delineate which services will be paid for by the capitated fee and which services are paid for separately by the managed care organization or independent practice association.

If the current trend away from hospital use toward less intensive levels of care continues, the flow chart of premium dollars within a capitated managed care organization might look as shown in Table 52.2-1.

$100 premium, allocated to		
HMO		$17
Administration	$9	
Out of area services	$3	
Profit and reserves (incl. withhold)	$5	
Provider group	$11	$83
Administration: quality assurance, reinsurance, contracting, marketing, information systems		
Hospital, including outpatient surgery, clinics	$29	
Medical group	$43	
Emergency services	$ 4.25	
Primary care physicians	$20.50	
All specialists	$18.25	
Total	$100	

Table 52.2-1 Example of Distribution of Funds in a Capitated Managed Care Organization

Enrolled Population Geographical accessibility to the population served often means using primary care physicians not on the medical staff of the hospitals who refer to specialists on the hospital staff. Some of these referring physicians may hold admitting privileges elsewhere; if so, they must meet preestablished quality standards and demonstrate a commitment to the concepts of managed care.

Management and Information Systems Management and information systems must be sophisticated enough to handle simultaneously the use and other clinical data of the system, financial functions (billing, collection, and distribution of funds), and measurement of performance outcomes. They must have access to current membership information and to services provided by both participating and nonparticipating providers and subcontracted providers. They should be able to coordinate elements of care, conduct population-based risk analysis, monitor changes over time in disease rates, track individual patients rather than procedures, and provide other information for clinical decision making. Current computerized systems primarily collect data on procedures, use, and costs; they are generally ineffective in handling nonhospital data.

Providers also need automated information systems to provide adequate information to subscribers and to managed care organizations on the basis of real-time data

on the treatments they are providing. Practitioners need concise and immediate information about treatments, including guidelines, disease management strategies, and the costs of different diagnostic strategies and medications. A primary care physician in a managed care organization must have information about the complication rate and the experience of the surgeon or other specialist to whom the patient is being referred. The physician also needs data about outcomes at hospitals.

MANAGED CARE ORGANIZATION STRATEGIES

Once the provider network is in place, the next concern of a managed care organization is how to influence the patterns of practice of its providers. The most direct approach is to limit the choices that providers can make in treating their patients. They may restrict the ability of physicians to perform certain procedures or to order certain medications or tests.

Other levers for restricting clinical choices focus on limiting the use of specialists or specialty services. Gatekeepers in one sense serve a rationing function, and that may convert the doctor-patient relationship into an adversarial one.

Monitoring Provider Performance The managed care organization must know how well its providers are functioning. In compiling data about the quality, outcomes, and quantity of treatment provided, the managed care organization constructs a profile of the care delivered by each treatment provider and each facility.

Different purposes are served by different profiling systems. Profiling can be used to analyze the strengths of each provider; it can show the treatments and diagnoses that a provider handles most effectively. This information can be used to match a patient with the provider who has had the greatest success in treating the patient's condition. Ensuring that the skills of the providers match the needs of patients within the system is an important component of network management. A profiling system can also be designed to find the providers who need the most scrutiny or have the greatest difficulty working within a managed care organization.

Assembling providers' profiles begins with a review of their credentials, their malpractice histories, their training, and their own reports on the treatments they provide and the patients they prefer. The convenience of the provider's geographical location, the ease of access for handicapped patients, the waiting time for appointments—all these features of a practice are weighed in establishing the provider profile. The managed care organization may perform a site visit to examine the program and offices where patients are treated.

The willingness of the provider to cooperate with the review process and the quality of treatment reports submitted also contribute significantly to the provider's profile. The profile may also incorporate case managers' evaluations of their interactions with the provider, and patient satisfaction may be tracked.

Utilization Review Utilization review is a specific subset of managed care and the strategy with the most immediate impact on patient care. A third party, interposed between patient and provider, judges the necessity, appropriateness, and adequacy of care and thus becomes the final arbiter of treatment. Utilization review consists of a set of techniques used to manage health care costs by influencing patient care decision making; an important technique is case-by-case assessment of the appropriateness of care before its provision. Utilization review is the most explicit effort to regulate physicians by modifying their treatment decisions; it is a direct intrusion on professional autonomy and judgment.

Under capitation, utilization management involves all the elements of the continuum of care, from the primary care physician's office, specialists' offices, the hospital, skilled nursing facility, and the patient's home. The goal is to control utilization before hospitalization is required; if admission is necessary, appropriate and cost-effective care must be implemented as soon as possible to minimize the probability of readmission. Utilization management must, therefore, be integrated with the medical protocols being followed by physicians who deliver outpatient care and aftercare once patients are discharged. Managed care organizations do not merely deny inappropriate admissions; instead, they direct the patient to the most appropriate level of care.

Prior Review Also known as *prior authorization* or *precertification, prior review* is an assessment of the need for the medical services being proposed for a patient, before they are provided. In the past, prior review was most frequently required for inpatient care; it is now being used more and more for ambulatory care, where it may be required in advance of treatment or after a specified number of visits.

Prior review typically consists of a telephone call between the treating physician and the first-level reviewer (e.g., psychiatric nurse or social worker). A few managed care organizations prefer written prior review requests. Prior review usually authorizes the initial procedures and the initial amount, duration, and scope of service required to evaluate the patient and generate an initial treatment plan.

How It Works Although hundreds of variations are possible, a managed care organization usually works in the following way. The patient calls a toll-free number and requests care. The system addresses emergencies immediately, while most callers follow a more routine path. The computer plays a central role in eligibility verification, data entry, utilization review, provider relations, treatment authorization, and provider assignment.

While the patient is on the telephone, eligibility is verified. The more sophisticated systems review basic demographic data on the patient, the patient's employer, the patient's group number and location, benefit descriptors, benefits used versus benefits available, availability of past treatment history, name of previous or current provider or providers, any “red flag” situations such as impending complaints or previous suicide attempts.

If the record shows that the patient has a therapist, the patient will be allowed to remain with that practitioner if the patient wishes to do so, if the therapist is a participating provider, and if, in the estimation of the reviewer or operator who is at the other end of the line, the provider's services are most appropriate for the patient.

If the patient has no therapist or record, the utilization review nurse, even with no knowledge of the patient's local area, can identify providers rapidly by scrolling down the list of participating providers in the patient's area. The computer thus introduces an element of randomness into provider assignment. The computer system creates a patient file, allowing for direct entry of notes about the patient's condition, requests, and disposition. The operator can either accept an existing provider or request a search for other providers if the therapist of record does not meet the company's requirements.

Initial screening is usually by zip code, because the company is obliged to provide care that is available within 30 minutes of driving time from the patient's home or workplace. Providers within a given zip code are generally displayed alphabetically, but they may be subtyped into particular clinical competencies. Once the reviewer has found a provider acceptable to the patient, the assignment is made and the system moves on to authorization for treatment. Most managed care organizations will not authorize treatment until a provider has been assigned.

In the case of self-referred patients, the psychiatrist is ordinarily responsible for both verifying the patient's eligibility to receive services and obtaining utilization review approval for the course of treatment being recommended. It is often wise to verify eligibility before the patient arrives and then obtain utilization review approval after the initial interview for the course of treatment being recommended.

Patients referred to the psychiatrist directly by the managed care organization usually (but not always) have been checked for their eligibility to receive services. The provider must check back at the end of the authorized period to update the record and to request additional care if necessary. If assessment has not already been done (by the primary care physician, for instance), the psychiatrist must get back to the managed care organization after assessing the patient, to obtain utilization review approval for the treatment planned.

Concurrent Review Concurrent review evaluates the appropriateness of the initial authorization for treatment, the need for the services outlined in the treatment plan, the consistency of the different treatment interventions with the measurable goals of treatment, and the adequacy of discharge planning. Requests for continued inpatient or outpatient treatment beyond the initial amount authorized (rarely more than 1 week in the hospital) trigger additional reviews, which may be required in any event at specified intervals. The psychiatrist is asked to provide updated treatment plans, progress notes, and treatment summaries.

The actual process of obtaining authorization for continued treatment through treatment summaries is much more involved than the foregoing would indicate. In one state, for example, approximately 175 managed care organizations are in operation. Most of them require written treatment summaries, and most have developed their own forms. The protocols associated with each form differ from the others. Providers are required by some managed care organizations to send the information after the assessment; by others, after a set number of visits; and by still others, with each request for continued authorization. Some may be faxed, others can only be

mailed. Several companies require use of their form; others will accept the provider's form.

Discharge Planning Discharge planning begins at the time of prior review. Goals for discharge and follow-up care (or termination, in the case of outpatients) should be contained in the initial treatment plan. Discharge planning may require arranging for additional follow-up medical and nonmedical services; case management may be of great assistance in such instances.

Case Management Also referred to as high-cost case management, case management targets the relatively few beneficiaries who generate (or are likely to generate) large expenditures or need a large number of providers or services. It determines whether extra assistance in planning, arranging, and coordinating a specialized treatment plan will permit appropriate and less costly care. Unlike utilization review, case management is voluntary and may permit coverage not specified in the benefit plan, such as residential treatment or partial hospitalization.

Retrospective Review Retrospective review is the final evaluation of utilization, designed to ensure consistency of the treatment plan, progress report, and medical records. The usual review is a medical record audit performed on-site for a random sample of patients. The results are used to analyze patterns of practitioner or institutional care for use in provider education, profiling, and selective contracting arrangements.

EFFECTS OF MANAGED CARE ORGANIZATIONS ON PHYSICIANS

Physicians in managed care organizations find themselves trying to balance three sides of patient care: patient compliance and lifestyle issues on one side, government regulation and requirements on another, and the cost (to both patient and physician) of the treatment chosen on the third side.

If managed care does in fact work, it will significantly affect the physician workforce requirement. It has been estimated that by the end of the century the surplus of physicians will be one third of the projected requirement. The surplus is particularly marked in the specialties, which account for 85 percent of the total surplus (an extra 139,000 specialists) (Table 52.2-2).

	Number of Enrollees in Plan		
	60,000	300,000	300 million
Primary care physicians	30.0	150.0	150,000
Psychiatrists	6.0	30.0	30,000*
Hospital beds	120	600	600,000

* In the private sector; not included are those needed in the public sector (also unknown are the ultimate effects of shifting costs and patients to the public sector). In 1996, there were more than 40,000 psychiatrists in the United States.

Table 52.2-2 Estimated Number of Full-Time Equivalent Primary Care Physicians, Hospital Beds, and Psychiatrists by Size of Plan

In recent years, the level of clinical autonomy has decreased for all physicians, including self-employed solo practitioners, not only because of the growth of managed care, but also because of concerns about professional liability, the implementation of certain clinical guidelines, and the activities of peer-review organizations.

Continued efforts to slow the growth of health care expenditures have furthered a trend toward large group practices. If current trends persist, most physicians in future years will be employees who are likely to have relatively little clinical autonomy, the hallmark of medical professionalism, with their practices subject to increasing bureaucratic rationalization.

EFFECT OF MANAGED CARE ON HOSPITALS

Like physicians, hospitals face major changes in a managed care organization environment. In the past, 10 patients requiring hospitalization might have received 2 to 3 weeks of inpatient care for a total of 210 patient days. Under a managed care organization only 3 to 5 of those patients will be admitted in the first place, and their average length of stay will range from 10 to 12 days, for a total of 60 inpatient days. In capitation contracts the hospital is committed to reducing the number of services provided, not to increasing the number of admissions.

PUBLIC HOSPITALS

Of all the pressures facing public hospitals, perhaps the greatest is the potentially devastating loss of revenue due to the growth of Medicaid managed care. Public hospitals have relied on fee-for-service Medicaid for as much as 40 percent of their total revenue. In recent years, however, states have begun to turn over their Medicaid programs to private managed care plans. These plans frequently sign up low-cost Medicaid enrollees (most often, women and children), depriving public hospitals of revenue while leaving them with the burden of caring for higher-cost Medicaid patients such as the chronically ill as well as uninsured patients.

MEDICARE AND MEDICAID MANAGED CARE

Because of a legal quirk in the new Medicare Plus Choice law, more and more health plans are pulling out of the Medicare risk-contracting business. In October 1998, 119 health plans notified the Health Care Financing Administration (HCFA) that they were either leaving Medicare completely or reducing their service areas. More than 500,000 beneficiaries were affected. Payment cutbacks and increasingly onerous, and often unreasonable, regulations may prompt a wave of health plans to depart Medicaid, similar to the documented defections in Medicare coverage. Payment cuts have already caused some Medicaid plans to lose millions and have driven others, often large commercial HMOs, from the market.

Some health systems believe the best strategy is to merge with smaller plans, adding fuel to the already existing fires of merger mania. If too many commercial plans leave the market, it could jeopardize competition and raise concern that a two-tiered health system is forming.

The prevailing philosophy in many states is that managed care is an adequate replacement for all public health services, and money is being diverted from categorical programs to managed care without adequate provisions for community health measures. Children constitute more than half the Medicaid population, and they highlight the multiplicity of factors that contribute to poor health, particularly in urban areas: poverty, parental and social neglect, and maldistribution of resources. While managed care can potentially reduce service fragmentation, increase access to individual care, and stimulate the development of more appropriate and less restrictive services in community settings, at the same time capitated payment arrangements inherently provide incentives to limit care. Children with special health care needs—chronic and handicapping conditions—may constitute between 20 and 30 percent of all children, and such medically complex children are the most likely to receive inadequate care or be extruded administratively from the system.

CHANGES IN MEDICAL CARE DELIVERY

The emergence of managed care as the predominant force in the practice of medicine is virtually ensured in the United States. Predictions using reasonable rates of growth indicate that 40 to 65 percent of the U.S. population will be enrolled in managed care plans within the next 5 years. What are the implications for the practice of medicine of the shift from patient-based to population-based medicine? As resources directed to health care are reduced, how will the trade-off between cost and quality be altered? Will quality even remain on the agenda? Who defines the appropriateness of interventions? How will managed care organization physicians make decisions about allocation of scarce resources, and what efforts should be made in tracking the population in terms of mental, physical, and social functioning? Should resources be allocated to assessing and improving the functional status of depressed patients who do not take their antidepressant medications? Should resources be reserved for those who participate in the medical system? Should certain groups (e.g., the chronically ill, the mentally ill, the elderly, children) be given

special resources for prevention and rehabilitation? If managed care organization demands reduce the contact-time between patient and physician, will the quality of the interaction be altered, and will the physician be able to gather enough information to make the correct clinical decision? Are the data now being made available to the public valid enough to guide the consumer in making health care choices?

Today, the typical hospital is 65 percent occupied. In Minneapolis, one of the pioneers of managed care, the number of hospital beds has dropped from 9000 to 4000. Nationally, the number of hospital beds has declined from 1,200,000 in 1989 to 980,000 in 1994, and it will probably dip to 870,000 by 1999.

California is the nation's most competitive market for managed health care services. HMOs there targeted hospitals as the principal locus of preventable health care utilization and expenditure. As a result, rates of admission and lengths of stay have plummeted. Hospitals are merging, reducing bed capacity, or closing. More and more medical and surgical procedures are being shifted to outpatient, subacute, and home health settings.

Between 1983 and 1993, the number of hospitals declined by 19 percent, acute care discharges declined by 11 percent, average length of stay by 17 percent, inpatient days by 30 percent (but to as low as 48 percent in areas of high HMO penetration), and inpatient surgical procedures by 35 percent. Patient visits to hospital-based ambulatory care facilities increased by 69 percent, and outpatient surgical procedures performed in hospital facilities increased by 132 percent.

SHIFT TOWARD CONSOLIDATION

Because their role as facilities for inpatient services is declining, probably irreversibly, hospitals are pursuing a strategy of “vertical integration” to gain control of the full continuum of care in their communities. Some observers predict that competition will drive providers in some major markets to consolidate into two or three systems by the turn of the century and that stand-alone hospitals and sole practitioners will disappear.

In 1994, 87 hospital mergers were publicly announced; in 1995, 126. All of 1995's transactions were eclipsed by MedPartners/Mullikin's $2.5 billion acquisition of Caremark, announced in April 1996. Aetna, a $13.3 billion company, acquired US Healthcare for $8.18 billion in April 1996 and Prudential Health Care for $1 billion in December 1998.

Health care merger and acquisition activity reached a record high in 1997 of 1183 transactions—an increase of 19 percent over the 997 mergers of 1996 and 87 percent over the 633 deals of 1995. The mergers included 56 HMOs, 127 units in the home health sector, 181 hospitals, 108 laboratories and dialysis centers, 108 long-term care facilities, 278 physician medical groups, 59 in the psychiatric sector, 74 rehabilitation facilities, and 192 in all other categories. In 1998, transactions decreased 9 percent to 198 deals, but the number of hospitals involved rose almost 10 percent to 687. The 1998 declines were particularly marked in physician groups, physician practice management companies, and the home health market.

The relations between managed care organizations and integrated delivery systems are still adversarial, and the megamergers may make them more so. The directions taken by the recent mergers have cast some doubt on the future of integrated delivery systems. Vertical and horizontal integration may be a transition phase, rather than the end result. Joint ventures, alliances, and “virtual” integration—consolidations involving less than full-asset mergers—are becoming more common. Rather than consolidate all their facilities at one location, managed care organization executives realize that they must provide multiple access sites.

Although HMOs prospered at first by taking advantage of easy one-time savings, their large profits are shrinking in the face of rising medical and administrative costs. Since 1996, many HMOs have experienced huge losses. Oxford lost $291 million in 1997 and a staggering $553 million in the first half of 1998. Even profitable plans lagged behind past performance. Humana's net income fell 97 percent from 1995 to 1996, Aetna's profits dropped 79 percent, and Cigna's net income fell 11 percent.

In 1997 HMOs accumulated nearly $800 million in losses, on average $1 million each, compared with an average profit of $5.4 million each in 1996. Administrative costs have also risen (8 percent, to $21.95 per enrollee per month in 1998 from $19.53 in 1997). Inpatient costs rose 4.9 percent per day, even though inpatient days per 1000 enrollees dropped 3.2 percent. Most Blue's plans also reported operational losses in 1998.

HMO losses translated into a large increase in premiums in 1998. Consumers who pay for their own care faced especially steep increases. The alternative was to pray for good health and go without insurance. The pool of uninsured people continued to rise by more than one million annually, to 43.1 million in 1997 (the latest information available). Small businesses that cannot self-insure, largely because they have so little leverage with health plans, are also hit hard by insurance rate increases.

MEASURING EFFECTS OF MANAGED CARE ORGANIZATIONS

Despite the dire predictions about its threats to the quality of health care and the occasional horror story reported in the press, managed care has yet to produce unequivocally negative results detectable by the measures so far applied. There is concern that necessary care will be delayed or eliminated, but few data support the position that managed care denies necessary care. What has changed are the ways many physicians and hospitals had practiced medicine under a fee-for-service system; whether that has an adverse effect on the quality of medical care remains to be demonstrated.

Almost all studies since 1980 have found that while HMO members generally had somewhat lower hospital admission rates, shorter lengths of stay, less use of expensive procedures and tests, and more use of preventive procedures and examinations than enrollees in indemnity plans, the HMOs provided a roughly comparable quality of care based on outcomes and process measures. Hospital costs in high managed-care markets are about 11 percent below the national average. Managed care does not give higher mortality rates or complication rates. In fact, the risk-adjusted mortality rates in high managed-care markets are 5.25 percent below the national average. Overall, HMOs are approximately 18 percent more cost-efficient than indemnity plans, and their cost-effectiveness is strongly correlated with the level of HMO penetration in the region. This finding contradicts the common argument that HMOs have only been cost-effective because they cover the best risks.

A study comparing changes in physical and mental health status between fee-for-service and HMO patients over 4 years found no distinguishable changes on average between patients in the two systems. Patients who were elderly, poor, or chronically ill, however, were more than twice as likely to decline in health in an HMO than in a fee-for-service plan.

A survey of New York City Medicaid beneficiaries found that managed care enrollees consistently rated their care better than, or at least as good as, conventional Medicaid beneficiaries. Beneficiaries in managed care and nonmanaged care systems did not differ on measures of use, nor did managed care decrease the use of more expensive settings to obtain care; inpatient stays and emergency department visits were essentially the same in both groups.

More-recent studies have been less positive, however. More and more public opinion surveys report that more than half the American public feel their health system has deteriorated with the advent of managed care. In Montana, HCFA investigated utilization rates for mental health services provided in the Medicaid program after the introduction of for-profit managed care delivery. Authorized inpatient days dropped 96 percent; residential services, 85 percent; authorized partial hospitalization visits, 45 percent; intensive outpatient services, 25 percent; and authorized outpatient visits, 76 percent. In addition, the number of patients being readmitted within 30 days after discharge increased 147 percent in fiscal year 1997 over previous years. The reductions in health care utilization are in no way matched by reductions in cost. Some critics argue that managed care does little more than transfer resources from those who are vulnerable because of illness and who need treatment (to which they are entitled under the Medicaid program) to venture capitalists who run managed care organizations.

Although quality-of-care reporting systems are in an early phase of development, the quality of both processes and outcomes of care can be evaluated. *Process measures* (what is done to a patient) are more sensitive measures of quality of care because poor *outcomes* (what happens to a patient) may not occur each time something is done incorrectly. *Efficacy* (the outcome of a procedure performed under ideal circumstances) may not predict *effectiveness* (the outcome of a procedure performed under usual circumstances).

Outcomes are multidimensional; a complete assessment of the outcome of care requires measuring general, physiological, mental, physical, and social health and patient satisfaction. Disease-specific outcome data must be collected, and patient preferences should be incorporated into these assessments. Patient satisfaction surveys alone are not always valid, however. Treating a patient in a pleasant and respectful manner (i.e., a good bedside manner) can mask poor technical care.

NEXT GENERATION OF MANAGED CARE

HMOs have long talked about introducing information technologies to track and reduce health costs, along with new models to manage and prevent disease. But few HMOs got past being able to market themselves, process claims, and cut checks each month. Health plans had trouble raising rates because big employers and purchasing groups beat them down. In recent years, however, 16 of the nation's 25 biggest HMOs were swallowed by other companies, which are now so big that they call the shots and find it easier to raise prices.

In addition to their need to raise premiums, health insurers and managed care companies are struggling with an aging population, expensive drugs, new medical technologies, the growth of services to patients outside hospitals, disaffected providers, the growing perception among consumers and policymakers that HMOs have put profits over patients, the complexities and unanticipated expenses of outcome research, the need for population-based prevention services, and the possibility of huge losses related to the various patient protection laws that have been passed or are under consideration by Congress and many state legislatures.

Currently, managed care organizations and payers have little incentive to invest dollars saved through short-term interventions in long-term preventive medicine programs. In the next generation of managed care, it is likely that disease management will overtake complications management and that managed care organizations will reinvest the short-term savings realized through their current strategies into a long-range strategy that promotes health, lowers disability, reduces morbidity and mortality, and ultimately achieves long-term savings.

HMO REFORM

The latest version of Rep. Charlie Norwood's bill, the Access to Quality Care Act of 1999, would amend the Employee Retirement Income Security Act (ERISA)—often used as a shield by HMOs against lawsuits brought in state courts by enrollees who claim that their decisions contributed to a patient's injury—to allow malpractice suits to be brought against HMOs and let states impose stricter standards on health plans than ERISA itself imposes. The bill (1) ensures access—plans must have adequate capacity to offer a broad range of services in a reasonable time at a reasonable location; (2) ensures emergency room care (if a “prudent layperson” believes the visit was necessary, the plan must cover it); (3) prohibits gag clauses; (4) regulates cost-control incentives by requiring that the plan disclose coverage, cost structure, provider mix, utilization review requirements, financial arrangement, and grievance procedures; and (5) allows enrollees to sue when they are injured by a plan's decision to deny coverage of health care services.

Managed care companies are beginning to embrace the external review and independent appeals processes they once scorned, in an effort to stave off reforms that will require them to justify their health care decisions to judges and juries. In some states, however, they were too slow to act (Table 52.2-3). In Texas, for example, HMOs can be held liable for malpractice if they make negligent benefits decisions affecting care. The first case under the new law involved a psychiatric patient, whose family claims that a managed care firm had negligently forced his discharge, despite the protests of his treating psychiatrist, by denying benefits. Following his discharge, the patient committed suicide by ingesting a half-gallon of antifreeze.

Table 52.2-3 HMO Consumer Bill of Rights

LEGAL ISSUES

Problems arise because many services have moved from hospital to primary care offices, gatekeepers are allowed limited flexibility yet are pressured to expand their scope of services, the information (and promises) given to the public by managed care organizations is not always reliable, and the choices provided patients in obtaining medical care may be restricted. Managed care limits a patient's freedom of choice, and it is thus inevitable that there will be more lawsuits against providers.

The main dangers for the physician and in particular the gatekeeper in managed care risk contracts are:

- Limits on care: Accepting utilization review denials without objection or delaying care while waiting for authorization, leading to “loss of opportunity for cure.” The physician's contract with the managed care organization should spell out how the physician and the patient can appeal the denial, delay, or location of patient care decisions. Managed care organizations often fail to give the public (and physicians) adequate or accurate information; some also intimate in that the care they will provide is more extensive than it actually is. In both cases, the physician is left holding the bag, because the patient tends to blame the physician rather than the managed care organization. Physicians are expected to provide services that would meet the community's standards of care, whether or not they are authorized by the plan. Lawsuits alleging negligence under managed care have been growing steadily over the last 5 years, and the claims are increasingly targeting physician groups rather than HMOs. Health care plans have been criticized particularly for withholding needed care by excluding or limiting coverage for chronic conditions such as human immunodeficiency virus (HIV) disease or mental illness, by denying reimbursement for experimental procedures, or by setting limits on the duration of inpatient hospital care for certain conditions.
- Financial incentives, as in capitation and bonuses that may lead to conflicts of interest. Many managed care organizations include “gag rules” in their contracts with physicians; they prohibit physicians from doing or saying anything that might undermine the confidence of enrollees in the plan. Yet the doctrine of informed consent requires disclosure by the physician. The enrollee may not lose confidence in the plan; instead, trust in the physician is undermined.
- Expanded scope of practice. The specialist's standard of care will apply even when the special care is rendered by the gatekeeper; this may lead to liability for failure to diagnose as well as failure to perform. Furthermore, even though the panel of specialists may be limited in a plan, gatekeepers are nonetheless expected to ascertain the qualifications of those to whom they refer, and they may be vicariously liable for specialist care.
- Coordination of care, such as failing to follow up on patients who are referred.
- Health screenings, often promised when a plan is being marketed as part of its focus on wellness. The physician who does not follow through could be held responsible.
- Plan exemptions. ERISA limits the types of damage that can be awarded a claimant. A managed care plan cannot be held liable for the consequences of a benefit determination, although it can be held liable for the consequences of a treatment decision. Physicians, however, are always held to their professional responsibilities to the patient. Whether a judgment made by a managed care organization is a benefit determination or a treatment decision has been the subject of many court cases. It is not clear if ERISA preempts state tort actions based on the medical malpractice of physicians or if managed care organizations, as employers, can be held vicariously liable for the physician's negligence. Some managed care organizations have tried to avoid the courts' uncertainty through indemnification (hold-harmless) clauses in their contracts with physicians. Such clauses require the physician to indemnify the managed care organization not only for physician negligence but also for all patient care liability, even denial of care by the managed care organization. Obviously federal law has not yet solved the problems of access and equity in health care, and it may even obstruct states' efforts to do so. ERISA preempts state action to ensure access and nondiscrimination, while the financial disincentives imposed by the managed care organization to provide fewer costly medical interventions, emergency department visits, and specialist referrals are potential threats to the patient-physician relationship and the quality of care provided.
- Termination. The termination policy in the contract between the physician and the managed care organization must clearly spell out the mechanisms for notification of patients and the transfer of patient care responsibilities when the physician resigns or is terminated; otherwise the patient may claim abandonment.
- Antitrust liability. Peer review and credentialing activities may expose the reviewer to antitrust action by a physician who is not admitted to the plan or one who is

dropped from a plan. In many areas, physicians and hospitals lack the immunity from antitrust actions that HMOs enjoy and are in fact proscribed from joining together and working collectively in ways that might benefit their patients. Under the 1994 guidelines to the antitrust enforcement policies of the Federal Trade Commission (FTC) and the Department of Justice (DOJ), physicians in solo practice or in small groups were allowed to negotiate discounted fee-for-service contracts directly with employers only under a cumbersome “messenger model.” In such a model, the physicians had to appoint a third party to be their messenger, who solicited offers from payers and then informed each physician separately (and without disclosing the range of fees that any other physician had said would be acceptable) about the payers' decisions. Any attempts to modify the messenger model were seen as per se violations of antitrust law—meaning they were considered violations merely because they had occurred and not because they were shown to harm competition. Clearly, it would be better for physicians if a rule of reason applied instead, requiring demonstration on a case-by-case basis that antitrust laws were violated and harmed competition. In August 1996 the antitrust guidelines were revised so that physician and multiprovider networks can escape automatic charges of price fixing. They can still be examined and challenged, of course, but it will be on a rule-of-reason basis rather than per se.

Many states are trying to increase control over managed care organizations (Table 52.2-3). Mandated maternity length of stay was the most widespread law in 1996, propelled by media reports that plans were requiring new mothers to be discharged from the hospital in as few as 6 hours after delivery. Next in frequency were laws requiring plans to provide patients with information about terms and conditions, anti–gag clause legislation, “fairness” laws (e.g., requiring plans to credential providers based on quality), standards and safeguards in utilization review, and requiring plans to allow patients to use nonnetwork providers.

Other issues that have come under legislative review in different states include complaint procedures, methods for opting out of a plan, marketing abuses, regulation of emergency services through the “prudent layperson” standard of what is a serious health problem, ensuring the chronically ill direct access to specialists, outside review of coverage decisions, provider due process for termination including the opportunity for a hearing, disclosing information on policies for referral, prior authorization, provider compensation, standards for drug coverage, establishment of a grievance procedure for denials and setting deadlines for resolving grievances, and maintaining confidentiality by establishing security procedures.

ETHICAL ISSUES

Whether health care should be subjected to the values of the marketplace is a fundamental question, because market-driven care tends to alienate physicians and undermine patients' trust of physicians.

Conflict of Interest Beneficence is central to medicine, but it is an impediment to the operation of competition and market forces. Beneficence toward individual patients frustrates the physician's obligatory role as gatekeeper in a managed care plan, which may encourage some gatekeepers to try to manage cases longer than they should, to keep patients at the primary level of care. Within capitated managed care organizations, physicians are expected to provide a wide range of services while keeping expenses to a minimum, typically by limiting use of services, increasing efficiency, shortening the time spent with each patient, and sparing use of specialists.

Managed care involves other conflicting loyalties. Within a managed care organization physicians have a responsibility to the patients they treat, but they also have a responsibility to all the other members of the capitated population. At times, that involves setting limited goals for particular patients in the interest of preserving access to appropriate care for all other patients in the group. It sometimes means drawing a careful distinction between what the patient wants and what the patient needs, between what is the best of all possible treatments in a particular situation and which one will be adequate and acceptable.

Managed care can pit patient needs against the financial interests of their physicians. Managed care organizations may pay bonuses to physicians, with the bonus increasing as expenditures for patient care decrease. Possible results are that some physicians will not provide all the services they should and that some will not always be adequate advocates for their patients in dealing with the managed care system. Under incentives to cut corners, even though there is no actual patient harm, physicians may lose their patients' trust that all necessary care is being provided.

Confidentiality Because health care depends on ready access to data, absolute privacy is unattainable in any kind of health data system. This raises the vexing question of the proper balance between data collection and privacy protection and the best means of achieving it. Justifications for privacy are based primarily on respect for the individual; justifications for using health information are based on the societal or collective good.

At present, data banks exist for specific kinds of data (e.g., genetic databases) and diseases (e.g., acquired immune deficiency syndrome [AIDS], tuberculosis, cancer). Systematic acquisition of a broad range of personal health data automatically results in loss of privacy. Health information can reveal intimate aspects of a person's or a family's life; it may affect one's ability to hold a job, maintain custody of children, secure immigration status, or obtain access to insurance or public benefits. In growing recognition of the effects of behavior on personal health, health agencies now collect and analyze behavioral information regarding, for instance, alcohol and other drug use, seat-belt and bicycle helmet use, smoking, exercise, and sexual practices.

Many forms of surveillance require disclosure of patient information to health departments. Patients, often rendered physically and mentally vulnerable by their illnesses, divulge intimate details of their lives to their physicians. Assurance that such information remains confidential is central to a trusting physician-patient relationship. It promotes patient candor about health and disease risks; uncertainty about confidentiality favors decreased disclosure, less frankness, or worse, reluctance to seek care. Finally, unauthorized disclosure of information can result in embarrassment, stigma, and discrimination.

Most states have nominal safeguards of public health privacy, but they are often incomplete or inadequate. Statutes are often silent about secondary uses of information (i.e., disclosure of data for purposes beyond those used to justify the original collection). Patients are not told whether, or to what extent, data collected for one purpose may be used for an unrelated purpose.

Many states have enacted disease-specific statutes. Some laws rigorously protect certain disease-specific data (e.g., HIV/AIDS) but are silent about guarding information on other conditions (e.g., noncommunicable diseases). As a result, different parts of the same health record may receive different degrees of protection under separate disease-specific statutes.

Electronic Records The change from fee-for-service to managed care has greatly expanded the use of electronic patient records. Manual records per se are by no means a guarantee of privacy. The major differences between the two kinds of records are that electronic records can more easily store much more information, can transmit it rapidly and simultaneously to large numbers of viewers (thousands, or even millions, rather than one or two), can readily link data that would otherwise remain separated indefinitely, and can store data indefinitely. The electronic record offers so many advantages that no credible suggestion of control has ever proposed its elimination. In view of the strengths and desirability of electronic records, what can be done to protect the patient's confidentiality? Among the basic elements of security strategy are the following: (1) data confidentiality, (2) identification and authentication of transmission sources and destinations, (3) data integrity, (4) ability to trace the message to a specific sender and receiver, and (5) access control.

Electronic systems can support selective restrictions on access to data, maintain extensive audit trails of who has seen or modified data, and transmit data in encrypted form. A message can be tagged so the sender cannot deny having sent it and the recipient cannot deny having received it. Digital signatures—unique numerical codes—are used to check the authenticity of sender and receiver. They can also be used for the physician's approval of record summaries and medical orders. The signature identifies location, privileges, and other essentials of secure electronic message exchange.

The *firewall* is a sophisticated software/hardware combination that stops unauthorized messages from the outside, accepting only those approved in advance. It also keeps insiders from using the Internet for unauthorized purposes.

A *message authentication code* (also called a *check sum*) extracts a unique summary of the message, compresses it, and compares it with the actual message. It recognizes any bit that has been changed and exposes the corrupted file.

No perfect system of confidentiality protection (if there be one) has yet been devised. Breaches of security can occur at all levels of communication, and information can be intercepted at different points. The human layer of confidentiality is the hardest to regulate. Hackers who attempt to break security systems invariably do so by duping people into giving out their passwords. They might claim they are system maintainers, impressing the unwary with technical jargon. Another hacker technique is “dumpster diving,” that is, looking through trash cans for paper memos containing the scribbled passwords that computer users write to themselves to help remember how to access their systems.

Informatics specialists strive for a balance, using technical and administrative safeguards to balance efficient access with confidentiality. A further step is to enact laws that impose severe sanctions on those who contribute to the misuse of confidential medical information. Finally, the sanctions must be enforced rigorously (not haphazardly) against all violators, not only a few token scapegoats.

SUGGESTED CROSS-REFERENCES

Primary care psychiatry is covered in Section 28.1. Health care delivery systems are discussed in Section 51.5 which covers Medicare, Medicaid, veterans affairs and community issues for the elderly. Legal issues in psychiatry and ethics in psychiatry are covered in Chapter 54. The future of psychiatry is discussed in Section 55.3. For the editors' view of managed care in psychiatry, see the Preface.

SECTION REFERENCES

Alperin RC, Phillips DG, editors: *The Impact of Managed Care on the Practice of Psychotherapy: Innovation, Implementation, and Controversy*. Brunner/Mazel, New York, 1997.

Anders G: *Health Against Wealth*. Houghton Mifflin, New York, 1996.

*Baker LC: Association of managed care market share and health expenditures for fee-for-service Medicare patients. JAMA *281:* 432, 1999.

Bauer JC: *Not What the Doctor Ordered: Laying a New Foundation for 21st Century Health Care or, How to End the Medical Monopoly in Pursuit of Managed Care*. McGraw-Hill, New York, 1998.

*Bellandi D: Mergers and acquisitions. Mod Healthc *11:* 48, 1999.

Brook RH, Kamberg C, McGlynn E: Health system reform and quality. JAMA *276:* 476, 1996.

Coile RC Jr: *The Five Stages of Managed Care: Strategies for Providers, HMOs, and Suppliers*. Health Administration Press, Chicago, 1997.

Daniels A, Dickman N, Zieman G: *The Comprehensive Managed Care Tool Kit*. CentraLink, Tiburon, CA, 1996.

*Docherty JP: Cost of treating mental illness from a managed care perspective. J Clin Psychiatry *60*(Suppl):49, 1999.

Friedman E: Capitation, integration, and managed care—lessons from early experiments. JAMA *275:* 957, 1996.

Goodman M, Brown J, Deitz P: *Managing Managed Care II*. American Psychiatric Press, Washington, DC, 1996.

Gostin LO: Contempo 1996—law and medicine. JAMA *275:* 1817, 1996.

*Gottlieb S: Managed care update. Am Med News *11:* 13, 1999.

Halverson PK, Kaluzny AD, McLaughlin CP, editors with Mays GP: *Managed Care and Public Health*. Aspen, Gaithersburg, MD, 1998.

Isenberg SF, editor: *Managed Care, Outcomes, and Quality: A Practical Guide*. Thieme, New York, 1997.

*Jacob JA: Blues seeing red. Am Med News *4:* 12, 1999.

Kletke PR, Emmons D, Gillis K: Current trends in physicians' practice arrangements. JAMA *276:* 555, 1996.

Kolb DS: *Assessing Organizational Readiness for Capitation*. American Hospital Publishing, Chicago, 1996.

Lieberman JA, Rush A: Redefining the role of psychiatry in medicine. Am J Psychiatry *153:* 1388, 1996.

*Mechanic D: *Mental Health and Social Policy: The Emergence of Managed Care*. Allyn & Bacon, Boston, 1998.

Morrisey M, editor: *Managed Care and Changing Health Care Markets*. AEI Press, Washington, DC, 1998.

Nash DB: *Total Learning Concepts* [managed care manual]. Boston, 1997.

Poynter WL: *The Textbook of Behavioral Managed Care: From Concept through Management to Treatment*. Brunner/Mazel, New York, 1998.

*Reinhardt UE: The economist's model of physician behavior. JAMA *281:* 462, 1999.

Roberts MC, Hurley LK: *Managing Managed Care*. Plenum, New York, 1997.

Robinson JC: Managed care in California. JAMA *276:* 1060, 1996.

Royce R: *Managed Care: Practice and Progress*. Radcliffe Medical Press, New York, 1997.

Shore JH: Psychiatry at a crossroad: Our role in primary care. Am J Psychiatry *153:* 1398, 1996.

Sisk JE, Gorman S, Reisinger A: Evaluation of Medicaid managed care—satisfaction, access, and use. JAMA *276:* 50, 1996.

Smith DR, Wong HY, Eichert JH: Third generation of managed care. Am J Manage Care *2:* 821, 1996.

Ware WE Jr, Bayliss M, Rogers W, Kosinski M, Tarlov A: Medical outcomes study. JAMA *276:* 1039, 1996.

Welfeld, JA: *Contracting with Managed Care Organizations: A Guide for the Health Care Provider*. American Hospital Publishing, Chicago, 1996.

*Wilkerson JD, Devers KJ, Given RS, editors: *Competitive Managed Care: The Emerging Health Care System*. Jossey-Bass, San Francisco, 1997.

52.3 ROLE OF THE PSYCHIATRIC HOSPITAL IN THE TREATMENT OF MENTAL ILLNESS

W. WALTER MENNINGER, M.D.

DEFINITION

Efforts to control escalating health care costs in the 1990s have focused on the most expensive locus of treatment, the hospital. This has been true for medical and surgical hospitalization as well as psychiatric hospitalization. Nonetheless, the psychiatric hospital represents a significant modality in the treatment of severe mental illness. It is a structured, therapeutic setting that simultaneously provides care, protection, and control while allowing the patient respite from a stressful situation.

From 1970 to 1992 the number of mental health organizations in the United States with inpatient or residential treatment services doubled, but the actual number of psychiatric beds provided by these organizations diminished by nearly half, from 524,878 in 1970 to 270,867 in 1992. Most striking has been the decrease in state hospitals, both in numbers and in beds. In 1990, while the total number of beds was lowered, the number of private psychiatric hospitals had increased threefold over the previous two decades from 150 to 462, with somewhat more than half owned by for-profit multihospital corporations. This same period saw more than a doubling of general hospital psychiatric units from 664 to 1571.

As tracked by the National Institute of Mental Health and National Center for Mental Health Services, patient care episodes in all mental health organizations increased fivefold between the first collection of these data in 1955 (1.7 million episodes) and 1992 (8.8 million episodes). In 1955, the inpatient episodes (1.3 million) represented more than three fourths of the total patient care episodes; the balance was outpatient. Consistent with their relative dominance in 1955, state and county hospitals accounted for 63 percent of those inpatient treatment episodes. By 1992, while inpatient episodes were somewhat increased (to 2.3 million), they accounted for only one fourth of all treatment episodes. Outpatient episodes then accounted for 69 percent of all treatment episodes, and partial care accounted for the remaining 6 percent. By 1992 only 15 percent of the inpatient episodes occurred in state and county hospital facilities. An increased number of psychiatric units in nonfederal general hospitals served 43 percent of the inpatient episodes in 1992 (up from 21 percent in 1955), while the proportion of 1992 inpatient episodes in private hospitals jumped to 24 percent (from 9 percent in 1955). Veterans Administration medical centers provided 9 percent of the inpatient episodes in 1992 (7 percent in 1955), and multiservice centers, not in existence in 1955, accounted for 9 percent of the inpatient episodes in 1992.

Paralleling the overall decrease in hospital beds, the average length of inpatient stay has been halved since the early 1980s. In one university hospital psychiatric service, the length of stay dropped 64 percent between 1985 (39 days) and 1993 (13 days). For Medicare admissions in 1992, the average length of stay was between 16 and 19 days. In short–term-care private hospitals, the average inpatient stay in 1993 ranged from 10 to 28 days, and this downward trend has continued in the years since. In general hospital units, the average stay now is 4 to 6 days. Whatever their governance, all hospital programs are expanding to provide a full continuum of care, ranging from crisis respite to intensive inpatient treatment to partial hospital care and sometimes residential programs for aftercare.

HISTORY

Sanctuaries for the physically and mentally ill date back to sixth century BC Greece. The earliest institutions devoted just to the care of the mentally ill were in the fourth century AD, in Byzantium and in Jerusalem. Over the next 1000 years, various religious orders, both Christian and Mohammedan, established places of refuge for the mentally ill. In Baghdad, the mentally ill were brought from all parts of Persia in 1173 to a House of Grace, where they received medical care until they recovered. Bethlehem Hospital in London, founded in 1247, began to accept “lunatics” in 1377. In Spain, mental hospitals date from 1410; mental patients were designated by the term une innocente (one without malice, easily fooled), and the mental hospital was identified as *hospital de innocentes*. The institutional care of the mentally ill in Europe in the late eighteenth and early nineteenth century was largely one of abuse, neglect, and restraint, with patients poorly clothed and fed, chained in hovels and dungeons, at times exhibited for the amusement of the public. For many years, keepers at London's Bedlam (Bethlehem Hospital) displayed patients like animals in a side show.

Philippe Pinel (1745–1826) is credited with revolutionizing the institutional care of the mentally ill when he liberated over 50 patients from chains and dungeons at the Bicetre hospital in Paris in 1793; his subsequent publication of *Traite Medico-Philosophique Sur L'alienation Mentale* (1801) outlined a humane approach to the care of these patients. In that same era, in England, William Tuke (1732–1819) proposed to the Society of Friends that they establish an institution “where a milder and more appropriate system of treatment than that usually practiced might be adopted.” In 1796 this led to the York Retreat, where abuse and restraint were replaced by kindness, tolerance, gentle exercise, work in the garden, and light recreation.

In early America, patients were cared for in their own or their relatives' homes, allowed to roam the countryside, or locked up in jails. The first institution to provide a specific unit for the mentally ill was the Pennsylvania Hospital, founded in 1752; its charter specifically provided for “the relief of the sick and the reception and cure of lunatics.” The first public facility exclusively for the mentally ill was opened in 1773 in Williamsburg, Virginia, based on the Bedlam Hospital in England. Both these institutions initially treated the mentally ill with procedures similar to the practices in Europe: bodies were chained, scalps were shaved, and bowels were purged. At the time of the American Revolution when Benjamin Rush (1745–1813) was on the staff of the Pennsylvania Hospital patients were being treated with venesection (blood letting), low-salt diets, purges, and emetics.

Early in the nineteenth century, “moral treatment” reached the United States. Modeled after the principles espoused by Tuke at the York Retreat, this treatment emphasized open wards, pleasant surroundings, minimal restraints, regular activity, and a familial relationship between the superintendent and the patients. Several private institutions committed to this humane therapeutic approach were opened: the Friends Asylum at Philadelphia (1817); the McLean Asylum outside Boston (1818), the Bloomingdale Asylum in White Plains (1821), and the Hartford Retreat (1824). Extraordinary treatment outcomes were claimed for this new treatment, with “cure” rates of 40 to 100 percent.

Except for a few early public mental hospitals—Eastern State Hospital in Lexington, Kentucky (1824); Manhattan State Hospital in New York City (1825); Western State Hospital in Staunton, Virginia (1828)—care for the indigent mentally ill was provided in jails. Dorothea Lynde Dix (1802–1887), a school teacher retired because of recurring tuberculosis, encountered the appalling conditions of the mentally ill when she was teaching Sunday School to female prisoners in Boston in 1841. She embarked on a crusade to get the mentally ill out of the jails and into suitable hospitals. By 1850 almost all the northeastern and midwestern states accepted her concept of a state-funded mental hospital; by 1860, 28 of the 33 states had at least one such hospital.

These state hospitals were originally designed to provide moral treatment in rural settings well removed from population centers because the “evils” of urban life were presumed to be a major cause of mental illness. They became large institutions, isolated from patients' families, increasingly impersonal and inhumane. The moral treatment emphasis on silence, orderliness, and regular routines gradually turned into regimentation and control. The initial expectation for treatment success was not achieved, optimism faded, populations began to rise, and hospitals became overcrowded.

In the late nineteenth century in Europe and the United States, special psychiatric hospitals were established for scientific study and treatment of mental illness—psychopathic hospitals. Bellevue Hospital in New York City, completed in 1879, was the first in this country, emphasizing examination, short-term observation, and treatment of acutely ill patients. Subsequent psychopathic hospitals were established in Boston (1912), Baltimore (Phipps Clinic, 1913), and Denver (Colorado, 1923), affiliated with medical schools to facilitate psychiatric education and research.

Modern Hospital Psychiatry The first biological era of modern hospital psychiatry was ushered in by the discovery of an organic etiology for some mental disorders early in the twentieth century, particularly syphilis, and the encephalitis associated with the 1917 influenza epidemic. Physicians prescribed specific medications—bromides, hypnotics, amphetamines, thyroid extract—in the treatment of mental illness in the 1930s. In 1933, coma induced by giving insulin to schizophrenia patients was touted as a new cure for the disorder. In 1934, Hungarian psychiatrist Ladislau von Meduna (1896–1964) reported positive results from chemically induced seizures in schizophrenia patients with severe psychosis. In 1938, Ugo Cerletti (1877–1963) and Lucio Bini (1908–1964) introduced electroconvulsive therapy (ECT). Initially employed for schizophrenia patients, it became recognized as an effective treatment for certain depressions, acute mania, and catatonic excitement. By the 1950s ECT had become the most used organic treatment in hospital psychiatry.

During the 1930s psychoanalytic ideas also were applied to the psychiatric hospital, most notably in the work of William Menninger (1899–1966) at the Menninger Clinic and Harry Stack Sullivan (1892–1949) and Frieda Fromm Reichmann (1889–1957) at Chestnut Lodge. Menninger designed an individualized treatment approach addressing the patient's needs through prescribed staff attitudes and therapeutic activities, activities that helped patients develop more appropriate outlets for the internal conflicts and psychological needs that previously found expression in the illness.

Following World War II, hospital programs were profoundly influenced by concepts of social and community psychiatry. In 1947 Maxwell Jones (1907–1990) established what came to be known as a therapeutic community. Rather than operating in the traditional, hierarchical manner of most hospital programs, he emphasized democratic, egalitarian values, encouraging everyone in the community to maximize their therapeutic potential through community meetings and ward activities. Meanwhile, the explosion of antipsychotic and antidepressant medications prompted the second biological era of hospital psychiatry. The availability of medications to diminish psychotic disorganization and facilitate patients' recovery changed the character of hospital units and reduced hospital stays.

In the 1950s, prompted by writings in the lay press, there was a renewed focus on the scandalously poor care provided in most public mental hospitals. Social psychologists were also criticizing the “institutionalization” associated with care in such facilities. Congressional action in 1955 established a Joint Commission on Mental Illness and Health, which produced a report in 1961, *Action for Mental Health*. This report laid the groundwork for the widespread establishment of community mental health centers. Concomitant with these developments and the introduction of psychiatric medications, there was a steady decrease in the public mental hospital census—a decrease that has continued to the present day.

The day or partial hospital is a modern development. The first reported day hospital was established in Moscow by M. A. Dzhagarov in 1933 because of a bed shortage and inadequate funds; it consisted mostly of work therapy. The first program in the Western Hemisphere was established by Ewen Cameron at the Allen Memorial Institute in Montreal, Canada, in 1947. A partial hospital program was established at Yale in 1948, the Menninger Clinic in 1949, and the Massachusetts Mental Health Center in 1952. The growth of partial hospital programs was stimulated by the deinstitutionalization of the public mental hospitals and the creation of community mental health centers.

The role of the psychiatric hospital in the treatment of mental illness has been maligned in recent years, with criticism stemming from several quarters. As a byproduct of the civil right's movement in the 1960s, civil libertarians challenged the restriction of patient movement and activity that accompanies most psychiatric hospitalization. Legislators viewed hospitals as expensive and outdated, needing replacement by less costly alternatives. Professionals, policy makers, and the public all looked at the community mental health movement as a less expensive, more effective, and more humane treatment alternative.

Meanwhile, when psychiatric hospital services were granted an exemption from the cost controls dictated by prospective reimbursement for hospitalized patients based on diagnosis-related groups, many general and psychiatric hospitals expanded capacity. A resulting dramatic increase in mental health care costs in the 1980s prompted third-party payers for psychiatric services to institute independent utilization review and to challenge the justification for hospital treatment. Following the proliferation of private psychiatric hospital beds in the late 1980s, some for-profit psychiatric hospitals engaged in unethical practices to fill those beds—paying for patient admissions, charging for services not rendered, using fraudulent diagnosis data, waiving the patient's deductible, and overcharging the insurance company.

THEORETICAL ISSUES

Traditionally, the functions of the psychiatric hospital have been: (1) the protection of the patients against their possible destruction of themselves, or harm to their homes, their work, or others; (2) the simplification of existence for patients by affording them protection from factors in the environment that precipitated and aggravated their illness; and (3) the provision of supportive care, personal contacts, and opportunities for expression and psychiatric counsel. To facilitate “healing” and recovery from mental illness, the hospital functions as a holding environment—secure, safe, predictable, warm, caring, and nurturing. Limits are clearly defined, particularly with respect to interpersonal boundaries. As a general rule, if these functions can be provided in a less restrictive treatment alternative, a patient should not be hospitalized.

These traditional functions and utilization of the hospital in psychiatric treatment have been challenged in recent years, primarily because of the greater expense of hospital care. In today's managed-care environment, one approach to reducing cost has been to limit psychiatric hospital treatment to patients who meet the criteria for “medical necessity”—suicidal behavior, aggressive behavior, or symptoms or behaviors that impair the person's ability to function outside an inpatient unit. Intervention is warranted only when the patient's life is in jeopardy or if the patient's behavior threatens to harm others. Critics of hospital treatment have also cited the limited data regarding the efficacy of inpatient treatment over alternative care. Indeed, experimental studies of hospital versus alternative treatment for mental disorders during the 1970s prompted a conclusion that outcomes of hospitalization were no more positive than the outcomes of alternative treatment.

Under pressure to reduce the cost of treatment, largely by reducing the length of hospitalization, the goal of hospital treatment has been significantly modified—to resolve the crisis that prompted the hospitalization. In the past, extended hospitalization provided an “in vitro” setting in which patients play out the symptomatic conflicts that they had manifested in everyday life “in vivo,” prior to hospitalization. In the milieu of the treatment setting, patients recreate their internal object world and problematic relationship patterns. Through the process of projective identification, staff members replicate important figures from the patient's past and current relationships. Careful examination of transference and countertranference developments in the patient's relationships with staff members reveals the dynamic conflictual issues in relationships that were a source of distress and malfunction in the patient's everyday life. Thus, the hospital is an artificial environment where the patient's unconscious conflicts can be expressed and perhaps resolved.

The therapeutic milieu in the hospital seeks to address patients' needs by containment, structure, involvement, support, and validation. As outlined by John Gunderson, *containment* sustains the physical well-being of patients and removes the unaccepted burdens of self-control or feelings of omnipotence, that is, it prevents patients from damaging themselves or others. *Structure* encompasses all aspects of the setting that provide a predictable organization of time, place, and person, allowing patients to feel safely attached without being overcontrolled or alone. *Involvement* in group activities and meetings with staff and other patients force patients to utilize and strengthen their ego and modify aversive interpersonal patterns as they interact with the therapeutic environment. *Support* is provided by the conscious efforts of the treatment team to make patients feel better and to enhance their self-esteem. *Validation* occurs when the therapeutic encounters and individualized treatment plan affirm the patient's individuality.

From the perspective of ego psychology, which regards mental illness as an ego disruption and disorganization, John and Elaine Cumming suggest that the hospital

> milieu must offer to the patient a clear, organized, and unambiguous social structure, problems to solve in protected situations, and a variety of settings in which to solve these problems. It should also offer him a peer group and a helpful staff to encourage and assist him to live more effectively. The program should aim at equipping the patient to act in clearly defined roles powered by a variety of motivating forces and governed by different cultural values.

In 1972 D. L. Rosenhan exposed some of the weaknesses and limitations of psychiatric hospitalization after arranging for pseudopatients to be admitted to twelve psychiatric hospitals in five different states. While there were significant methodological shortcomings in his study, his observations are nonetheless reflective of the criticism of the psychiatric hospital unit. He concluded:

> The hospital itself imposes a special environment in which the meanings of behavior can be easily misunderstood. The consequences to patients hospitalized in such an environment—the powerlessness, depersonalization, segregation, mortification, and self-labeling—seem undoubtedly counter-therapeutic.

TECHNIQUES

Hospital treatment characteristically involves a multidisciplinary group of mental health professionals. Each team member addresses different elements of the patient's difficulty: psychiatrist (medical or psychiatric, medication, psychotherapy), nurse and mental health technician (personal care and interaction), psychologist (diagnostic assessment of cognitive strengths and weaknesses, psychotherapy), social worker (relations with family), activity therapist—occupational, recreational, music (avocational and vocational skills). The diagnostic assessment and treatment of the patient are enhanced by the collaboration and integration of these multidisciplinary perspectives.

The phases of hospital treatment include (1) admission, (2) immediate assessment and intervention, (3) continued diagnostic evaluation and refinement of treatment, (4) clarification of treatment goals and discharge criteria, (5) progress toward and achievement of goals, (6) discharge, and (7) appropriate aftercare or follow-up. The impact of managed care review has shortened the time allotted for these phases. When the hospital has a full continuum of care, ranging from intensive inpatient treatment to partial hospital programs for aftercare, the phases will continue into the partial hospital period.

Each patient should have a well-defined master treatment plan that addresses the reason for admission and the need for hospitalization, an assessment of patient problems, patient's strengths and assets, formal diagnoses, discharge criteria, and anticipated placement and posthospital treatment. Further, the treatment plan should identify problems to be focused on by the members of the treatment team, particularly problems that reflect the primary reason for hospitalization. Generally, the problems will be evident in one or more of the following areas: self-concept, interpersonal relationships, thinking and cognition, emotional functioning, impulse regulation or addiction, adaptive skills, or family. Goals should be articulated in measurable terms so that the patient's progress can be readily observed by all members of the treatment team. Prescribed diagnostic and treatment modalities should be recorded, as well as which staff members are responsible for carrying out these tasks (Table 52.3-1).

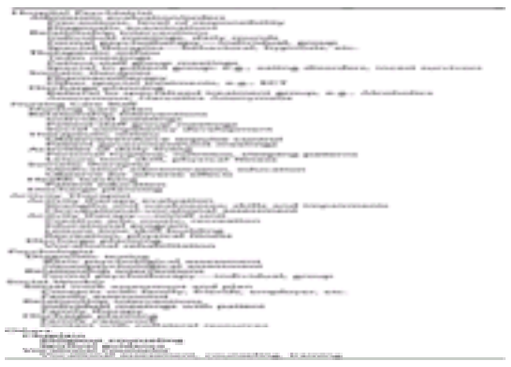

Table 52.3-1 Hospital Treatment—Responsible Staff and Modalities

Hospital treatment programs may be characterized by their length of stay, primary treatment modality, and treatment goals. Inpatient programs are generally categorized as (1) short-term, crisis intervention (1 to 4 weeks); or (2) extended treatment (greater than 4 weeks). Partial hospitalization programs are generally open-ended, and may serve as a transition from inpatient to outpatient care, or they may be entered directly. These programs include various combinations of day hospital supervision and activities and night accommodations—night hospital, halfway house, family care, cooperative housing, and independent living.

Short-Term Inpatient Programs Short-term or brief hospitalization seeks to arrest and reverse an emotional decompensation, alleviate an acute mental illness, reduce symptoms, restore defenses, and facilitate a readjustment of the patient to the prior environment. These goals are achieved through the separation of the patient from the environment by admission to the hospital, administration of appropriate psychiatric medication, exploration, and negotiation with significant others in the patient's life, and supportive and exploratory psychotherapeutic intervention with the patient. Sometimes specialized treatment procedures like ECT may be indicated. Short-term hospitalization may also be utilized for an in-depth diagnostic patient evaluation. Short-term stays are possible when no untoward treatment complications arise and when posthospital resources or support systems are available, accessible, or rapidly developed.

Specialized or focused programs have been developed for a wide range of mental illnesses, such as alcohol or other substance abuse, eating disorders, and dissociative disorders. In the 1970s the prototype of these programs was developed for the treatment of alcohol use disorders, with a time-limited treatment program consisting of patient education about all facets of the condition, group confession and confrontation, self-help groups (Alcoholics Anonymous), and social work with family. Now, these programs may utilize the controlled environment of the hospital to control the patient's impulse to pursue symptomatic behavior while providing psychotherapeutic, educational, social work, and rehabilitative support.

Extended-Treatment Programs Extended-treatment programs remain necessary for difficult-to-treat illnesses that do not respond to brief intervention or short-term treatment. Lloyd Sederer and Paul Summergrad note that longer-term treatment may be indicated when briefer treatments have failed, treatment complications are severe, a treatment alliance is absent or resisted, life stressors are so profound or persistent that they impair recompensation, or an extended period of treatment is needed to achieve any beneficial effect. Extended-treatment programs may be modeled after a psychoanalytic psychotherapeutic model, a therapeutic community model, a token economy behavioral treatment model, or a moral treatment custodial care model.

The psychoanalytic perspective views symptoms as a product of unconscious psychic conflict within the patient, a compromise formation that expresses both the unconscious drive derivative and the ego's defense against the impulse. Further, it is assumed that all behavior is determined and has meaning. The psychoanalytic model of hospital treatment seeks to identify and resolve the conflicts and symptomatic ego defenses of the individual as they are played out in the interactions with other staff and patients in the hospital. As articulated by William Menninger:

> It is our hates and our mismanaged loves that account for most if not all of our emotional sickness. In other words, our reactions to the people about us and their reactions to us result in adjustment or in varying degrees of maladjustment. Similarly, within the hospital, it is the relationships with other human beings—with other patients and with the hospital personnel—and the understanding of these relationships which lead the patient back to emotional health. The hospital must provide a variety of opportunities for resolution of the patient's conflicts through his expressing his hate in appropriate or constructive ways, in his learning to find new or renewed satisfactions in his activities, in his learning to love and be loved.

The therapeutic community was originally introduced in England, conceived by T. F. Main and popularized by Maxwell Jones. This model looks at the hospital unit as a social system. It seeks to optimize the patient's healthy functioning in a setting that levels the hierarchy of the hospital and involves patients in decision-making on the unit. It places responsibility on the patients to serve as agents of change, having a significant role in the rehabilitation of other patients on the unit. This model depends upon older patients transmitting to newer patients the norms, skills, and values essential for meaningful participation in the therapeutic community. Jones' treatment program was primarily designed for the treatment of patients with personality disorders. Therapeutic community approaches are more difficult to implement in units that treat a large number of psychotically disturbed patients because groups may be overstimulating to them.

The token economy model applies principles derived from experimental psychology and operant conditioning, with token rewards used as positive reinforcement for desired behaviors. These programs are used with more regressed, lower-functioning chronically ill patients. As the patients receive rewards for the preferred, adaptive behavior, they can exchange the tokens for desired privileges and activities.

In all models of inpatient hospital treatment, psychiatric nursing plays an essential role. Nursing personnel provide the direct supervision of the patients and administer prescribed medication. They also provide for basic needs of the patient, offer organization and structure, give support and aid intrapsychic growth, assist the patient with interpersonal and interactional skills, and promote coping skills in a teaching and learning environment.

Partial Hospitalization Partial hospitalization offers an alternative to 24-hour hospitalization, with some structured treatment programs that represent a transitional care setting between hospital and community. For some chronically ill patients the partial hospitalization may continue for an extended period. Most programs provide an informal, clublike setting that facilitates social interaction, features therapeutic activities, and diminishes dependency and isolation. Patients have a case manager or treatment coordinator who functions as an advocate and facilitator for the patient. Rehabilitation and vocational programs include assistance with daily living skills,

recreational activity, and job training. Other services provided are medication review, group therapy, family treatment, and crisis treatment.

CLINICAL ISSUES

Indications for Hospitalization Indications for hospital treatment are determined by factors within the individual patient—severity of the illness, level of awareness or insight regarding the illness, and the capacity to contain impulsive, destructive behavior—and factors in the environment such as presence or absence of support and intensity of ongoing stressors. As a general rule hospital treatment is indicated when the patient is so severely disturbed that it is necessary for someone else to step in and protect the patient from harming self or others, and the patient's environment is not in a position to provide this protection (Table 52.3-2). In recent years various checklists of admission criteria have been published; essentially, a patient merits hospitalization when there are serious problems in three parameters: the degree of dangerousness, the adequacy of the support system, and the ability to cooperate with care.

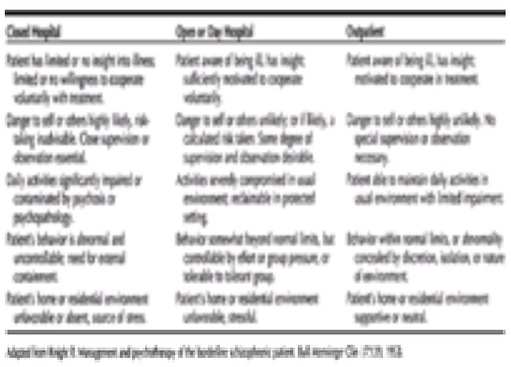

Table 52.3-2 Criteria for Hospitalization

Glen Gabbard has outlined comparative indications for brief and extended hospitalization according to a grid of (1) severity of illness factors, (2) treatment factors, and (3) environmental factors. Severity of illness factors include the danger to self or others, the degree to which symptoms interfere with functioning in work of family and are unresponsive to outpatient treatment, impulsive or noncompliant behavior, and dual diagnosis with substance abuse complicated by another psychiatric disorder. Relevant treatment factors are the history of any previous hospitalization and the response to any prior treatment, the presence of any treatment complications, and the goals of treatment. Environmental factors are the family system, the posthospital support system, and the existence of clear precipitating stressors (Table 52.3-3).

Table 52.3-3 Indications for Brief Hospitalization and Extended Hospitalization

With regard to criteria for hospitalization for children, a review of 375 children seen at the Western Psychiatric Institute and Clinic at the University of Pittsburgh identified six items that discriminated between those children who were hospitalized and those who were not:

1. Aggressive outburst toward other people.
2. Aggressive outbursts toward animals or objects.
3. Patient's condition was deteriorating rapidly or failing to improve despite adequate outpatient treatment.
4. Physical or neurological conditions or a psychotic, disorganized state requires hospitalization to initiate treatment or to establish a diagnosis.
5. A pathological or noxious situation exists among the patient's family or associates that makes treatment without hospitalization impossible; or the patient's disordered state creates such difficulties for the family that the patient has to be hospitalized.
6. Evaluation of the patient's condition requires 24-hour observation and evaluation which only a hospital can provide, including stabilization or reevaluation of medication or treatment of drug or alcohol dependence.

Complications of Hospitalization Complications of hospital treatment—inpatient or partial hospital—are in large part a function of the multidisciplinary treatment team interacting with patients' pathology. Many forces impinge on the work of this team, reflecting group dynamics, patients' transference reactions, and staff countertransference. In their classic study of Chestnut Lodge, Alfred Stanton and Morris Schwartz review “institutional participation in psychiatric illness and treatment,” and they identify breakdowns in staff communication and splitting, which resulted in treatment complications. Gabbard identifies four features of the process of splitting in hospital treatment:

> (1) The process occurs at an unconscious level. (2) The patient perceives different staff members in dramatically different ways, based on projections of the patient's internal object representations, and treats each staff member differently according to those projections. (3) Staff members react to the patient, through projective identification, as though they actually were the projected aspects of the patient. (4) As a result, treaters assume highly polarized positions in staff discussions about the patient and defend those positions with extraordinary vehemence.

T.F. Main has described troublesome countertransference reactions when patients fail to respond to earnest treatment efforts. Treatment staff “tend to become more passionate, to be reinforced by aggression and then to deteriorate in maturity, with sadism invading the situation, together with its concomitants of anxiety, guilt, depression and compulsive reparative wishes, until ultimate despair can ensue.” Other countertransference phenomena include rejection of the unresponsive or difficult patient, benign neglect of undemanding patients, and inappropriate acting out by staff through expression of anger or romantic involvement with provocative patients. These countertransference patterns are best managed by creating a climate and expectation that permits treatment staff working in the milieu to openly express countertransference feelings in staff meetings. Clear communication of disagreements between staff members on the team is essential to cope successfully with the complications that result from splitting and projective identification enactments. Finally, ongoing staff education about splitting and projective identification is important to help staff members recognize these phenomena when they occur.

Patient's Family The family both influences and is influenced by hospitalization of a family member. Rather than focus on a pathological interaction between patient and family, modern hospital treatment attempts to engage the family as a caring and concerned resource in the treatment process. Models have been developed for inpatient family intervention, initiating family therapy with the patient's hospitalization. This approach does not assume that the etiology of the major psychotic disorders lies in family functioning or communication, but rather that the present-day functioning of a family with which the patient is living or is in frequent contact can be a major source of stress or support. The intervention seeks to help the family to understand, live with, and deal with the patient and the patient's illness. Psychoeducational workshops and multiple family group meetings are especially helpful to families of patients who are chronically mentally ill.

CONTEMPORARY ISSUES

Access to Treatment It has become increasingly evident that economic forces determine which patients gain admission to public hospitals, general hospitals, or private psychiatric hospitals. Rather than a specific diagnosis, the driving factor behind admission to a particular hospital is the patient's source of funding. A review of sample surveys from the National Institute of Mental Health for 1975, 1980, and 1986 found state hospitals increasingly admitting patients with no source of payment. Although there was an overall decline in patients with private insurance across all diagnostic and racial groups, patients with mood and mild mental disorders were more likely to have private insurance or access to Medicare or Medicaid dollars and to gain admission to either private general hospitals or private psychiatric hospitals.

What happens with fewer and fewer state hospital beds? Richard Lamb and Roderick Shaner reviewed the experience in California, which had only 8.3 state hospital beds per 100,000 population in 1993, compared to the national average of 40 per 100,000 (in 1955 the national average was 339 beds per 100,000 population). They cite a relatively small but important minority of patients for whom the most effective ongoing psychosocial and psychopharmacological interventions are not enough. They identify four major effects: (1) The new generation of chronically and severely mentally ill persons pose the most difficult problems in community treatment and do not readily benefit from brief hospitalization. (2) In the absence of intermediate and long-stay psychiatric beds, there is a revolving door syndrome, leaving many persons untreated or inadequately treated. (3) The homeless mentally ill population has increased, forcing many to live lives of danger and deprivation on the streets. (4) Society's limited tolerance of mentally disordered behavior results in pressure to institutionalize persons; in the absence of hospital beds, these persons behave such that they are arrested, incarcerated, and transferred to the criminal justice system.

For-Profit Versus Nonprofit Hospitals In 1988 over 64 percent of the inpatient psychiatric facilities were privately owned, with a substantial increase in the number of hospitals in multifacility for-profit corporations. Since the financial goal of these facilities is to maximize the wealth of the stockholders, there are clear incentives for these institutions to have higher prices, lower expenses, and more profitable patients than the nonprofit facilities. With a careful analysis of data from the Health Care Financing Administration's Minimum Cost Data tapes for 1986 to 1990, Michael McCue and Jan Clement compared operational performance and productivity of matched pairs of nonprofit and for-profit psychiatric hospitals. They found significant differences in length of stay (for-profit average stay was 5 days longer, 23.71 days versus 18.44 days nonprofit average stay), number of discharges per bed (for-profit average was 4 fewer, 10.53 versus 15.11 nonprofit average), and full-time employees per adjusted census (for-profit 1 less full-time employee, 2.53 versus 3.65 nonprofit). This resulted in nonprofit hospitals averaging 46 percent lower total profit margin and a 14 percent lower return on total assets that the for-profit, investor-owned group.

The for-profit hospitals had significantly higher net revenue and lower salary expenses per patient day, presumably achieved through reduced staffing levels. The for-profit and nonprofit hospitals did not differ significantly in expenses for routine care and ancillary services; they did differ in their capital structure. The system size and reputation provided the for-profit hospitals greater access to debt capital and newer facilities, with higher depreciation.

Patient Violence With a narrowing of criteria for hospitalization, which emphasizes danger to self or others, the threat of violent behavior by patients in hospitals has received greater attention. It is of note that there were practically no articles on violence and assault in the psychiatric literature and no references in leading textbooks until the mid to late 1960s. That was a period of increasing violence in American society, and the problem has continued to increase, with the Surgeon-General and the U.S. Public Health Service officially identifying violence as a public health concern.

Violent patients present a challenge to staff and patients in hospital units. One prospective study in a large managed mental health program in the metropolitan Chicago area looked at 270 consecutive admissions to seven participating hospitals. It was found that 14.8 percent of those admissions (40 patients) had engaged in a behavior in the 30 days prior to admission that threatened great bodily harm to another and 15.2 percent (41 patients) had engaged in violent or threatening behavior at some point in the previous year. These patients were more disruptive during their hospital stays, engaged in more outbursts and more verbally threatening behavior than the other patients who ranked low on a scale that measured danger toward others. Acute symptom severity and prior dangerousness were the most useful predictors of disruption on the unit. Eileen Morrison studied nine units in four hospitals—private, state, federal, and county—and found that over half of the patient sample was violent toward others on admission. From a series of studies, Morrison has identified a different view of violence from the traditional idea of the violent patient being a schizophrenia patient who is out of control. Rather, the violence in patients in the 1990s parallels the violence in society in the 1990s, reflecting issues of power and control. Predictors of violence in hospital patients are a history of violence; a coercive or exploitative style with others, that is, using others for self-gain; and length of hospital stay. Schizophrenia patients are less likely to be violent than substance abusers and antisocial patients.

Clearly, hospital staff must be prepared to deal with conflict that goes with an aggressive population, and the hospital should have clear policies and guidelines for dealing with outbursts and emergency situations in the least restrictive but most efficient manner. To minimize violence in the hospital, staff needs to be trained in communication skills of negotiation and conflict resolution, and to use creative interventions aimed at preventing aggression and violence before it starts.

Impact of Managed Care Utilization review by outside, managed-care entities has intruded into hospital treatment in a profound manner, but few careful studies have documented the impact of this development. The data on 2265 utilization management reviews conducted during a 3-year period (1989–1992) for patients insured by a single large commercial insurance company were analyzed and it was found that managed care clearly restricted inpatient care, primarily through limiting length of stay. The insurance company reviewers initially approved inpatient psychiatric treatment for nearly all (98.8 percent) of the patients, but authorized, on average, only one third of the days requested (6.9 versus 19). The utilization management relied on standardized treatment protocols that do not include consideration of individual patient or clinical factors and the continuous review and approval placed a considerable burden on the clinician.

Partial Hospital Care and Wraparound Services With the pressure to reduce hospital inpatient utilization, there has been a surge in partial hospital programs and elaboration of a new concept of “wraparound” services developed in the community with the initial assistance and supervision of hospital staff. In 1982 the American Association for Partial Hospitalization published a set of standards to narrow the concept and restrict the term to comprehensive, multidisciplinary programs with medical, psychiatric, psychosocial, and prevocational components, targeted toward patients with serious mental disorders. Despite concerns that the majority of persons traditionally referred to inpatient settings could not get better outside of the hospital, repeated studies have demonstrated that day treatment as an alternative modality is as effective or more so in facilitating adjustment and reintegration of patients into the community. Despite this evidence, partial hospital care has not replaced inpatient treatment for a number of reasons, not the least of which is the entrenched professional commitment to the hospital as a locus of treatment. Also, insurers have been reluctant to cover the costs, for fear the new costs would simply explode in magnitude while hospital costs continue unabated.

This resistance is likely to change as more studies point to the cost effectiveness of partial hospital treatment programs. William Sledge and his colleagues at Yale followed 197 patients in an experimental program designed as an alternative to initial, short-term inpatient hospitalization. The clinical, functional, social adjustment, quality of life, and satisfaction outcome measures were as effective as acute hospital care for these urban, poor, acutely ill voluntary patients with severe mental illness. Further, the alternative care cost roughly 20 percent less in total direct costs, on average, than the inpatient hospitalization. The two programs had roughly equal direct service staff and capital costs, but the day hospital and crisis respite operating costs were 51 percent of the inpatient hospital costs. This study demonstrated again that many people with major mental illnesses, especially voluntary patients, can be treated safely, effectively, and less expensively in alternative settings but this program requires an effective triage and crisis service as well as back-up from a hospital. The authors of the experimental program observed:

> There are times when the structure of a hospital is quite necessary; alternatives to hospitalization do not entirely replace the need for psychiatric hospitals. Essential hospital-based specialty functions are the provision of safety and security in the management of patients who are out of control and quick access to skilled medical nursing care of patients with mental complications.

RESEARCH AND EVALUATION

Concern about the increasing costs associated with hospitalization has prompted a search for reliable data regarding treatment outcomes. Such studies are difficult to design and execute. Traditionally, outcome studies have focused on changes in a patient's clinical status. Equally important is the level of the patient's interpersonal and occupational functioning in daily life.

Researchers in Israel sought to identify variables associated with improved functioning after psychiatric hospitalization, reviewing 25 clinical and demographic variables, improvements and functioning at discharge of 309 consecutive patients in an open psychiatric ward over a 2-year period. Of particular note, they found that diagnosis had little value in predicting which patients would improve during hospitalization. The best predictors of a return to a preadmission level of functioning were psychotherapy, a hospital stay of less than 4 months, and attendance at occupational therapy.

There are an increasing number of significant long-term outcome studies of mental illness like Thomas McGlashan's Chestnut Lodge follow-up study and Michael Stone's follow-up of patients with borderline personality disorder treated at the New York State Psychiatric Institute. Findings reported from controlled studies comparing short-term and long-term hospital treatment have been equivocal. A major finding is that aftercare arrangements are more important in the patient's outcome than the actual length of hospital stay.

A recent study explored the outcomes of nursing home residents admitted to the Houston Veterans Medical Center geropsychiatric inpatient unit. In general, the hospitalization resulted in improvement in overall psychiatric function and problem behaviors of nursing home residents with and without dementia. Of the hospitalized nursing home patients, 10 percent were able to be discharged to less restrictive settings than those from which they were admitted. However, the investigators acknowledged that their 37-day average length of stay was longer than most current health insurance companies allow.

Early studies surveyed attitudes of patients and doctors regarding what was most helpful about hospitalization. One survey at the UCLA neuropsychiatric Institute found that patients ranked the most helpful treatment modality to be (1) being in a new environment, (2) interaction with nurses, (3) interaction with patients, (4) individual psychotherapy, and (5) medication. The doctors ranked medication first, followed by being in a new environment, individual psychotherapy, interaction with nurses, and nursing care programs (Table 52.3-4).

Patients' Ranking		Doctors' Ranking
1	Being in new environment	2
2	Interaction with nurses	4
3	Interaction with patients	6
4	Individual psychotherapy	3
5	Medication	1
6	Recreational therapy	10
7	Work therapy	8
8	Nursing care program	5
9	Group therapy	9
10	Occupational therapy	11
11	Ward social activities	7
12	Psychodrama	12

Adapted from Leonard CV: What helps most about hospitalization? Compr Psychiatr 15:363, 1973.

Table 52.3-4 What Helps in Hospitalization (Rank Order)

Outcome studies of partial hospital treatment programs have found equivalent outcomes in symptom improvement, relapse reduction, and family adjustment for patients who could be treated safely in partial hospital settings. In some studies, better social functioning was achieved in partial hospital treatment. Several programs, like the Massachusetts Mental Health Center day hospital, have demonstrated effective treatment for patients who would otherwise have been admitted to an inpatient unit but these programs typically have high staff-to-patient ratios, highly structured programming, and the capacity to handle clinical emergencies.

One randomized controlled trial of day patient versus inpatient psychiatric treatment in the United Kingdom concluded: “Roughly 40 percent of all acutely ill patients presenting for admission to a psychiatric unit may be treated satisfactorily in a well-staffed day hospital.” At 3 months' follow-up, the only significant difference was a greater improvement in social role performance of inpatients. By the 1-year follow up, there was no significant difference between day patients and inpatients in mental state, social role performance, burden or behavior.

A number of outcome studies prompted by deinstitutionalization in the public sector in the United States, United Kingdom, and Italy conclude that a new long-stay group is replacing the old long-stay group. The residual patients are older, mostly with organic syndromes, who are difficult to accommodate in the community. A follow up study of Rhode Island patients with chronic mental illness and long histories of hospitalization found that 55 percent did require at least one hospital readmission. At the 7-year follow up, 57 percent were functioning effectively in structured, community residential settings; 28 percent had moved on to independent living; and 16 percent had returned to an institutional setting.

In the United States there also appears to be a transinstitutionalization phenomenon, with an increasing number of mentally ill persons being processed by the criminal justice system and incarcerated in correctional facilities rather than mental hospitals.

SUGGESTED CROSS-REFERENCES

Additional perspective on the history and future of psychiatry are found in Section 55.1 and Section 55.3. The role of the structure and protection of patients from destructive consequences of their illness is addressed in the treatment sections for specific diagnoses as well as in Section 29.1 on suicide. Section 48.8 addresses day treatment and residential and inpatient treatment of children and adolescents; Section 52.3 discusses treatment settings for geriatric patients.

SECTION REFERENCES

Belcher JR, DeForge BR, Thompson JW, Myers CP: Psychiatric hospital care and changes in insurance coverage strategies: A national study. J Ment Health Admin *22:* 377, 1995.

Citrome L, Green L, Fost R: Length of stay and recidivism on a psychiatric intensive care unit. Hosp Comm Psychiatry *45:* 74, 1994.

Costello AJ, Dulcan MK, Kalas R: A checklist of hospitalization criteria for use with children. Hosp Comm Psychiatry *42:* 823, 1991.

Cumming J, Cumming E: *Ego and Milieu*. Atherton, New York, 1962.

Druss BG, Bruce ML, Jacobs SC, Hoff RA: Trends over a decade for a general hospital psychiatry unit. Adm Policy Ment Health *25:* 427, 1998.

Fenton WS, Mosher LR, Herrell JM, Blyler CR: Randomized trial of general hospital and residential alternative care for patients with severe and persistent mental illness. Am J Psychiatry *155:* 516, 1998.

Gabbard GO: Splitting in hospital treatment. Am J Psychiatry *146:* 444, 1989.

Goldsmith HF, Manderscheid RW, Henderson MJ, Sacks AJ: Projections of inpatient admissions to specialty mental health organizations: 1990 to 2010. Hosp Community Psychiatry *44:* 478, 1993.

*Group for the Advancement of Psychiatry: *The Family, the Patient, and the Psychiatric Hospital: Toward a New Model*. Brunner/Mazel, New York, 1985.

*Gunderson JG: *Principles and Practice of Milieu Therapy*. Jason Aronson, New York, 1983.

*Hawthorne WB, Green EE, Lohr JB, Hough R, Smith PC: Comparisons of outcomes of acute care in short-term residential treatment and psychiatric hospital settings. Psychiatry Serv *50* :401, 1999.

Hoge MA, Davidson L, Hill WL, Turner VE, Ameli R: The promise of partial hospitalization: A reassessment. Hosp Comm Psychiatry *43:* 345, 1992.

*Husted JR: Population density and re-hospitalization for the seriously and persistently mentally ill. Arch Gen Psychiatry *56:* 285, 1999.

Knight R: Management and psychotherapy of the borderline schizophrenic patient. Bull Menninger Clin *17:* 139, 1953.

Lamb HR, Shaner R: When there are almost no state hospital beds left. Hosp Community Psychiatry *44:* 973, 1993.

Lansing AE, Lyons JS, Martens LC, O'Mahoney MT, Miller SI, Obolsky A: The treatment of dangerous patients in managed care, psychiatric hospital utilization and outcome. Gen Hosp Psychiatry *19:* 112, 1997.

Lieberman PB, Wiitala SA, Elliott B, McCormick S, Goyette SB: Decreasing length of stay: Are there effects on outcomes of psychiatric hospitalization? Am J Psychiatry *155:* 905, 1998.

McGlashan TH: Selective review of recent North American long-term follow-up studies of schizophrenia. In *Psychiatric Treatment: Advances in Outcome Research*, SM Mirin, JT Gossett, MC Grob, editors. American Psychiatric Press, Washington, DC, 1991.

Mechanic D, McAlpine MA, Olfson, M: Changing patterns of psychiatric inpatient care in the United States, 1988–1994. Arch Gen Psychiatry *55:* 785, 1998.

Menninger WC: The functions of the psychiatric hospital. Bull Menninger Clin *6:* 109, 1942.

Morrison EF: The evolution of a concept: Aggression and violence in psychiatric settings. Arch Psychiatr Nurs *8* :245, 1994.

*Munich RL, Gabbard GO: Hospital psychiatry. In *American Psychiatric Press Review of Psychiatry, vol 11*, A Tasman, M Riba, editors. American Psychiatric Press, Washington, DC, 1992.

O'Driscoll C: The TAPS Project. 7: Mental hospital closure—A literature review of outcome studies and evaluative techniques. Br J Psychiatry *162* (Suppl):7, 1993.

Olfson M: The array of psychiatric services in general hospitals. Gen Hosp Psychiatry *15:* 277, 1993.

Rosenhan DL: On being sane in insane places. Science *179:* 250, 1973.

Ross JL: Principles of psychoanalytic hospital treatment. Bull Menninger Clin *49:* 409, 1985.

Rothbard AB, Schinnar AP, Hadley TP, Foley KA, Kuno E: Cost comparison of state hospital and community-based care for seriously mentally ill adults. Am J Psychiatry *155:* 523, 1998.

*Sederer LI, Rothschild AJ: *Acute Care Psychiatry, Diagnosis and Treatment*. Williams & Wilkins, Baltimore, 1997.

Sederer LI, Summergrad P: Criteria for hospital admission. Hosp Community Psychiatry *44:* 116, 1993.

Sledge WH, Tebes J, Rakfeldt J, Davidson L, Lyons L, Druss B: Day hospital/crisis respite care versus inpatient care, Part I: Clinical outcomes. Am J Psychiatry *153:* 1065, 1996.

Sledge WH, Tebes J, Wolff N, Helminiak TW: Day hospital/crisis respite care versus inpatient care, Part II: Service utilization and costs. Am J Psychiatry *153:* 1074, 1996.

Snyder W: Hospital downsizing and increased frequency of assaults on staff. Hosp Community Psychiatry *45:* 378, 1994.

Squire MB, Stout CE, Ruben DH, editors: *Current Advances in Inpatient Psychiatric Care—A Handbook*. Greenwood Press, Westport, CT, 1993.

Stamm I: Countertransference in hospital treatment. Bull Menninger Clin *49:* 432, 1985.

*Stanton AH, Schwartz MS: *The Mental Hospital*. Basic Books, New York, 1954.

Stone MH, Stone DK, Hurt SW: Natural history of borderline patients treated by intensive hospitalization. Psychiatr Clin North Am *10:* 185, 1987.

Wickizer TM, Lessler D, Travis KM: Controlling inpatient psychiatric utilization through managed care. Am J Psychiatry *153:* 339, 1996.

52.4 PSYCHIATRIC REHABILITATION

ROBERT PAUL LIBERMAN, M.D., ALEX KOPELOWICZ, M.D. AND THOMAS E. SMITH, M.D.

The growing recognition that a large proportion of persons with schizophrenic and mood disorders experience a poor quality of life with long-term disability, persisting symptoms, or a relapsing course of illness has given birth to the field of psychiatric rehabilitation. Early intervention and effective treatment of acute episodes of symptom exacerbations are important for minimizing long-term disability. On the other hand, workers in psychiatric rehabilitation emphasize continuous, comprehensive, coordinated, and indefinite treatment of lifelong mental disorders to maintain symptom control, prevent or reduce relapses, and optimize psychosocial performance. The goal of psychiatric rehabilitation is to teach skills and provide community supports so that individuals with mental disabilities can function in social, vocational, educational, and familial roles with the least amount of supervision from the helping professions.

Treatment and rehabilitation are interconnected seamlessly, as are the full range of biopsychosocial services, in a continuous and comprehensive effort to reduce impairment, disability, and handicap among the mentally disabled. In addition, a major goal of psychiatric rehabilitation is enabling the mentally disabled person and family members to be actively involved in treatment decisions and to achieve the highest feasible quality of life in the community.

An enormous population of mentally disabled persons needs psychiatric rehabilitation to bolster their quality of life. For example, the National Plan for the Chronically Mentally Ill used the triad of diagnosis, disability, and duration to identify persons who suffer from persistent or recurrent organic, schizophrenic, mood, anxiety, and other disorders that become chronic and erode or prevent the development of their functional capacities in relation to three or more primary aspects of daily life. These functional areas of daily life include personal hygiene and self-care, self-direction, interpersonal relationships, social transactions, learning, recreation, and economic self-sufficiency.

Since many childhood developmental, psychotic, and mood disorders that persist for many years and impede normal psychosocial development are included in the target population, the total number of persons in the United States who need psychiatric rehabilitation in any 6 month period may exceed 10 million, based on data compiled by the National Institute of Mental Health (NIMH) Epidemiologic Catchment Areas study.

The terms “chronically,” “severely,” or “seriously mentally ill” have been used interchangeably to identify persons with long-term disabilities resulting from schizophrenia, persisting depressions, and bipolar disorder, delusional disorders, borderline personality disorders, and dementias. Estimates of the prevalence of persisting mental disabilities, such as those shown in Table 52.4-1, are now considered gross underestimates, given more recent data from the NIMH Epidemiologic Catchment Area and Co-morbidity studies, the difficulty of counting homeless and jailed mentally ill persons, and the number of persons with dual diagnoses as well as those disabled by serious mood and anxiety disorders. Although this section focuses on rehabilitation of persons with psychotic disorders, persons with disabling mood, anxiety, substance abuse, and personality disorders can also benefit from many of the same structured and skills-training modalities. Behavioral rehabilitation can also slow the functional deterioration of persons with dementias, using modalities found effective with younger, mentally disabled persons.

Prevalence estimates for chronic mentally ill (2.1–2.6%)	4,000,000–5,000,000
Community estimates of chronic mentally ill	2,800,000–3,500,000*
Treated prevalence of chronic mentally ill patients	
In mental hospitals	50,000–60,000
In other specialty facilities†	250,000
In nursing homes	1,000,000–1,100,000
In prisons and jails	100,000

* Excluding 1.2–1.5 million living in institutes.
† Includes halfway houses and other large community residential facilities not covered in community survey.
Data from Barker PR, Manderscheid RW, Hendershot GE, Jack SS, Schoenborn CA, Goldstrom ID: *Serious Mental Illness and Disability in the Adult Household Population of the United States, 1989.* Center for Mental Health Services, Rockville, MD, 1992.

Table 52.4-1 Point Prevalence of Persisting, Long-Term Mental Illness and Disability in the United States

The challenge to psychiatric rehabilitation is amplified by the inadequate resources and poor organization of service delivery for the target population, which results in thousands of homeless, mentally ill persons in urban centers of the United States and a concentration of mentally disabled persons in state and local prisons. With over 5000 mentally ill prisoners, the Los Angeles County Jail is said to be the largest psychiatric institution in the United States. The transinstitutionalization of the seriously mentally ill from state hospitals to jails, custodial board-and-care homes, and inner city streets is a condemnation of our society's humane values. Our failure to provide high-quality, continuous psychiatric treatment is made more tragic by the availability of new rehabilitative technologies that, when systematically organized and delivered, have the potential to reduce morbidity, disability, and handicaps among serious and chronic mentally ill persons.

Organization and delivery of therapeutic and rehabilitative interventions for persons experiencing serious mental disorders requires a longitudinal, informed, mutually respectful and collaborative therapeutic alliance between the patient and caregiver. Whether the care-giving function is fulfilled by a psychiatrist, allied mental health professional, or paraprofessional case manager, the qualities inherent in a supportive, empathic, individually tailored relationship must be present. Given the primary role of the family and other natural caregivers (e.g., operators of residential care homes) in maintaining treatment and compliance, plus the enormity of the family burden experienced by relatives, professionals must also involve families in treatment planning and implementation. Rehabilitation is a labor-intensive and person-to-person venture. Even the pharmacotherapy that must accompany psychiatric rehabilitation is effective only to the extent that the doctor-patient relationship contains trust, problem solving, and shared goals. Medication is never prescribed, ingested, or metabolized in a socioenvironmental vacuum; both drug and psychosocial treatments depend upon a working alliance between treater and treated.

The terminology used to refer to an individual receiving mental health services can vary depending on the site at which the individual is being treated. For example, *patient* may be a suitable term for inpatients in a hospital; *client* may be appropriate in mental health centers; and *consumers* or *members* may be useful in outpatient programs or vocational and residential settings. The term *patient* is used in this chapter for consistency with other chapters in this book. However, readers need to be mindful of the appropriate terminology when developing and implementing rehabilitation programs.

CONCEPTUAL FRAMEWORK FOR REHABILITATION

Studies in Europe, the United States, and Japan that followed up persons who experienced disabling forms of schizophrenia during early adulthood found, 20 to 40 years later, a remarkable 50 to 66 percent functioning actively in their communities with few symptoms, a reasonably good subjective quality of life, and only limited dependence on professional caregivers. These findings have spurred interest in psychiatric rehabilitation as a way to facilitate social and symptomatic recovery of seriously mentally ill persons. With an attachment to data-based empiricism and hypothesis testing, an interdisciplinary cadre of specialists in psychiatric rehabilitation derived new assessment and intervention methods from the vulnerability-stress-protective factors model of psychiatric impairment, disability, and handicap (Fig.

52.4-1). The course and outcome of major mental disorders are defined by:

1. *Impairments*, the characteristic positive and negative symptoms and associated cognitive and affective abnormalities of disorders such as schizophrenia, autistic disorder, and bipolar disorder
2. *Disabilities*, the restrictions impairments impose on such functional life domains as personal hygiene, medication self-management, recreation for leisure, and family and social relationships
3. *Handicap*, the disadvantage experienced by an individual with impairments and disabilities that limits or prevents the fulfillment of normal roles, such as worker, student, friend, citizen, and family member

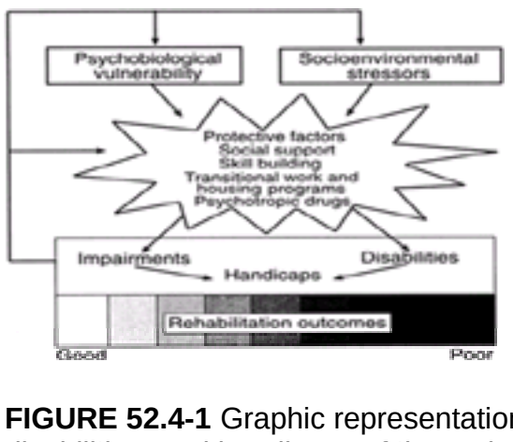

FIGURE 52.4-1 Graphic representation of the vulnerability-stress-protective factors framework for understanding the determinants of symptom impairments, disabilities, and handicaps of the major mental disorders. Protective factors buffer or filter the noxious effects of stressors superimposed on enduring vulnerability in producing psychopathology. There is a feedback loop from impairments and disabilities back to vulnerability and stress; for example, improved social skills and coping capacity can mitigate stressors and may even modify neurobiological vulnerability.

Moving mentally disabled persons along the spectrum of impairment, disability, and handicap from poor to good outcomes requires orchestration of protective factors in treatment and community support services. As long as the psychobiological vulnerability factors responsible for the specific syndrome are unknown, interventions cannot directly modify them. Those vulnerability factors (most likely genetic and neurodevelopmental) endure. They likely exist before the symptoms of the disorder emerge as well as during periods of symptom remission and relapse. Similarly, as long as the principles of community care of the mentally ill are adhered to, vulnerable individuals cannot be isolated from socioenvironmental stressors. Stressors, whether drugs of abuse or social overstimulation, are a fact of life for mentally ill persons, even in the so-called asylums, where privacy is absent and violence is omnipresent. Even in the absence of major life events or the noxious effects of illicit drugs and alcohol, vulnerable persons can succumb to tension or conflict in their environments or to microstressors and daily hassles if they lack the protection conferred by medication, coping abilities, and social support.

Psychiatric rehabilitation must harness the protective factors in both the treatment and natural environments to offset and buffer the adverse effects of stress superimposed upon vulnerability. Those protective factors include optimal psychopharmacological interventions that can raise the threshold at which environmental stressors precipitate symptoms in an individual with a given level of vulnerability to a serious mental disorder. But even prescribing pharmacotherapy with attention to the highest possible benefit to risk ratio (i.e., using atypical antipsychotic drugs to maximize therapeutic effects while minimizing side effects) does not ensure protection against relapse. Medications must be supplemented by psychosocial interventions that (1) equip patients with personal, social, and vocational coping and competence and (2) galvanize necessary social support services to compensate for the intrusion of symptoms, deficits, and disabilities that even the best system of care cannot remediate.

MULTIDIMENSIONAL REHABILITATION KEYED TO DISORDER STAGE

The clinical practice of psychiatric rehabilitation joins three major sets of factors that protect against stress and vulnerability: (1) pharmacotherapy tailored to the type and severity of psychopathology at dosages that do not produce sedation or adverse effects that interfere with positive engagement in rehabilitation; (2) development of skills so the patient can meet the challenge of stressors and life situations that demand adaptation and independence; and (3) a range of supportive social services, including housing, transitional and supported employment, financial support, and case management to sustain a mentally disabled person in the community. Psychiatric rehabilitation assumes that disabled persons need maintenance medication, social and independent living skills, and environmental resources and supports to fulfill the role demands of community life.

The organization of truly comprehensive psychiatric rehabilitation is diagramed in Figure 52.4-2. One dimension of the cube represents the stage of the disorder, ranging from prodromal, to acute, to residual phases. The second dimension contains treatment and rehabilitation modalities, such as drug therapy, family and cognitive therapies, social skills training, and vocational rehabilitation. The third dimension holds requisite support programs aimed at compensating for disability and minimizing handicap (e.g., family education and support, social service entitlements, case management, housing, and psychosocial clubs). The psychiatric rehabilitation practitioner or team following a person for many years through various phases of a disorder should be able to use this multidimensional framework to match treatments and support services to the changing needs of the patients as dictated by their impairments, disabilities, and handicap at any point in time.

FIGURE 52.4-2 Complex cube of psychiatric rehabilitation.

Rehabilitation of individuals with chronic and serious mental disorders had to respond to the transition in locus and focus of care from institution to community and from custodial care to functioning in society. Community support programs, psychosocial self-help clubs, and community mental health centers are now the hubs of psychiatric rehabilitation services where new modes of intervention are coordinated and delivered by case managers in league with multidisciplinary teams of clinicians who assume long-term responsibility for sustaining seriously mentally ill patients in the community. As is shown in Figure 52.4-3, the tripartite roles and competencies of case managers are the key to comprehensive and continuous rehabilitation.

FIGURE 52.4-3 Case management functions for serving the comprehensive needs of seriously mentally ill persons.

Engagement The successful engagement of a patient in a therapeutic relationship is the first step in psychiatric treatment and rehabilitation, no matter what the venue for service delivery. The clinician or clinical team attempting to engage seriously mentally ill persons in treatment must build a collaborative relationship with patients and their families. Professionals and patients must recognize each other's agenda and seek some common ground in developing a comprehensive treatment plan. Engaging homeless, mentally ill persons often requires active outreach, with mental health practitioners meeting prospective consumers of services on their own turf, frequently the streets and back alleys of urban centers.

The process of engagement is contemporaneous with the setting of overall rehabilitation goals. Helping individuals define their desired life roles can be a vehicle for engaging and motivating them to participate in a lengthy rehabilitation process. Patients who actively participate in goal setting are more likely to abandon denial of their illness, acknowledge ownership of their goals, seek reinforcement from significant others when they work toward their goals, and customize or modify their goals as their rehabilitation proceeds. Successful engagement of the patient and family or other natural caregivers in rehabilitation often requires not only active outreach by the professional, but also a nonjudgmental and nonauthoritarian approach and even toleration of nonadherence to medication for periods of time. Engagement, a long-term effort, is often challenged by obstacles in service delivery, reflected in the current terms supplanting “patient” such as “consumer” or “customer.” The following case vignette illustrates engagement.

Mark was a 23-year-old man who experienced the onset of schizophrenia during his last year in college. He was referred to the rehabilitation team after efforts to engage him in treatment had previously failed. He was found to be severely psychotic, evidencing disorganized thinking, paranoid delusions, auditory hallucinations, and bizarre behavior. He had refused earlier efforts to have him take antipsychotic medications because of adverse effects he experienced. The information about his psychotic symptoms gathered by using the Brief Psychiatric Rating Scale (BPRS), particularly that relating to disorganized thinking and the hallucinations, was shared with Mark. He was specifically told about the impact the symptoms had on his daily functioning and was shown the ratings for his symptoms. He was told that his ability to use his social and independent living skills was limited as a result of his psychosis. He agreed to a trial of antipsychotic medication, during which time he would self-monitor his psychotic symptoms using a scale similar to that used in the BPRS. He and his parents were sent home with this scale, and were asked to complete the Independent Living Skills Survey (Fig. 52.4-4), a self-report assessment of daily functioning.

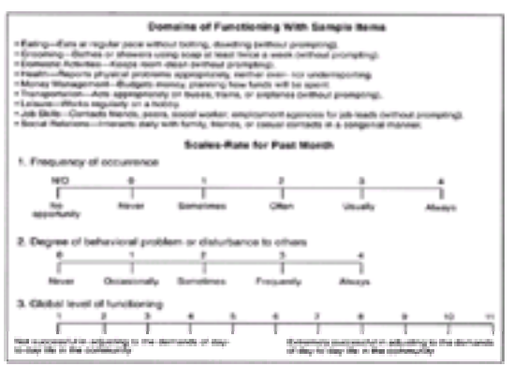

FIGURE 52.4-4 Independent Living Skills Survey.

When Mark returned, his psychotic symptoms had decreased from the extremely severe level to mild and moderate levels. He reported enhanced self-efficacy since taking the medication and especially noted his improved attention and concentration, which allowed him again to read for reasonably long periods of time. That effect was important to Mark because it allowed him to pursue his longer-term rehabilitation goal of completing his college education. Close examination of the Independent Living Skills Survey revealed difficulties in grooming and self-care and basic conversational skills. He and the rehabilitation team agreed to target those deficiencies during their initial skills-training efforts.

Community Reentry Program Ideally, engagement should begin when patients first encounter mental health services, regardless of treatment setting. For instance, inpatient treatment for individuals with serious and persistent mental illness should be seamlessly integrated with community services so that hospitalization is brief and minimally disruptive and the community's rehabilitation services can be implemented as quickly as possible. Unfortunately, this ideal is often not achieved, and many psychiatric inpatients do not follow through with community care after discharge. Often, the result is a cycle of relapse and rehospitalization, known as the “revolving door” phenomenon.

The poor rates of transition to community-based treatments are likely to be both system and patient related. Efforts to alter the structure and function of mental health systems to make them more cohesive and accessible (e.g., the Program for Assertive Community Treatment) have had some success in reducing recidivism, but model programs have not been widely adopted and are unavailable to most patients who might benefit. Fewer efforts have targeted the other half of the aftercare equation, namely, the patient. The Community Re-entry Program—a 16-session modularized, skills-training program—was designed to teach inpatients how to actively collaborate with the mental health system in creating and implementing their own tailor-made aftercare plans. Once engaged in this fashion, educated patients might acknowledge the need for continuing care and actively participate in treatment, discharge, and rehabilitation planning. Studies at New York Hospital–Cornell Medical Center and UCLA/Olive View Medical Center have shown that inpatients who participated in this program improved their knowledge and performance of the skills presented in the program, had significantly higher rates of engagement with aftercare services after discharge, and had improved levels of community functioning 2 months after discharge. These results suggest that engaging patients in psychiatric rehabilitation procedures such as skills training can begin in the hospital and continue without interruption into the community, linking inpatient and outpatient facilities in a common treatment philosophy and a matching set of treatment techniques.

ASSESSMENT

If engagement is the first (albeit continuing) step on the road to successful psychiatric rehabilitation, then assessment is the roadmap for the journey. A thoroughgoing, carefully conducted assessment pinpoints the areas for rehabilitation and provides a baseline for monitoring its effects. Furthermore, assessment within the framework of psychiatric rehabilitation, with its emphasis on positive goals and functioning linked to the current phase of illness and symptoms, fosters a collaborative relationship between patient and practitioner that greatly facilitates treatment.

Assessment is part and parcel of any therapeutic interaction, and some might argue that the clinical sensitivity of an experienced mental health practitioner is all that is needed to conduct an adequate assessment. Indeed, only minimal sensitivity may be needed to detect certain problems such as assaultiveness, which if present must be a target of treatment. Comprehensive assessment not only reveals factors that subtly influence even the most obvious problems, it also establishes both the baseline for monitoring the effects of treatment and the collaborative relation that sustains it. This comprehensive assessment process is illustrated by a recently

developed, integrated and coordinated assessment and treatment-planning instrument, Client Assessment of Strengths, Impairments, and Goals (CASIG).

Client Assessment of Strengths, Impairments, and Goals CASIG is a multidimensional assessment tool that helps practitioners plan, document, and evaluate psychiatric rehabilitation. The acronym of this instrument takes into account the empowerment of the client or consumer in identifying service needs, while following the World Health Organization's International Classification of Impairments, Disabilities [the reciprocal of capabilities or “Strengths”], and Handicaps [the reciprocal of participation in setting interests and goals]. CASIG is based on the model of psychiatric rehabilitation outcomes presented above, with the additional assumption that the plan for services must solicit the active participation and integrate the goals, assets, deficits, needs, resources, and constraints of all the relevant stakeholders. Stakeholders include client or patient and significant others including family members and other caregivers and those paying for the services.

As shown depicted in Figure 52.4-5, CASIG assesses performance in 10 areas of functional living skills, subjective quality-of-life in 11 areas, presence of 5 symptoms, 20 medication adverse effects, compliance with medication, and performance of 10 unacceptable community behaviors. The information is collected from the patient during a 60- to 90-minute interview, with corroborating information collected from the patient's significant others and treatment personnel who know the client well. Additionally, the patient's preferences to change his/her behavior in each area are elicited.

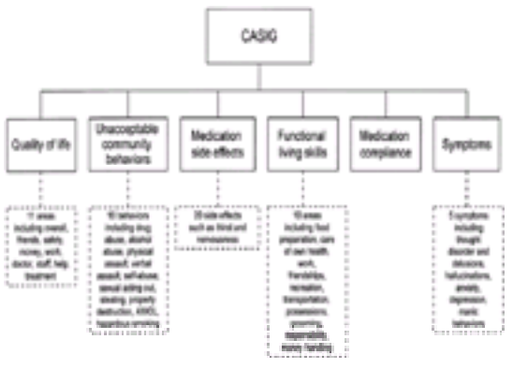

FIGURE 52.4-5 Areas assessed by Client Assessment of Strengths, Impairments, and Goals (CASIG).

Regardless of the area being assessed or the source of the assessment information, CASIG items have all been designed to achieve several objectives:

to be reliably administered by any of the paraprofessional and professional staff who typically provide services for patients (even high-functioning consumers can reliably administer CASIG to other patients)

to be easily incorporated into programs with diverse staffing patterns, resources, locations, and clinical responsibilities

to include multiple sources of information about particular social deficits or symptoms of the patient so that any disagreements can be resolved to plan services better

to bridge the many facilities in which patients are served

to monitor (with repeated administration) the progress of individual patients in their treatment and rehabilitation programs and in the aggregate monitor the effectiveness and changing characteristics of programs

Two clinical examples for the CASIG—one assessing a symptom (depression), and the other assessing a functional living skill (money management)—are presented in Figure 52.4-6 and Figure 52.4-7, respectively.

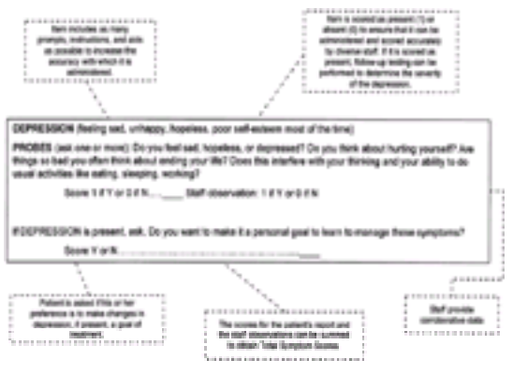

FIGURE 52.4-6 Example of a symptom—depression—assessed by CASIG.

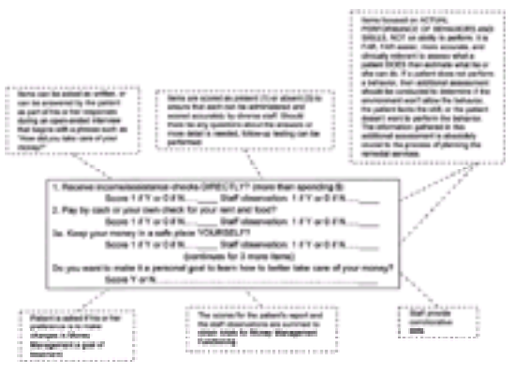

FIGURE 52.4-7 Example of functional living skills: three items from the money management section of the CASIG.

The assessment results are summarized and then interpreted by comparing them with norms developed from other patients and with the requirements for successful functioning in various environments. As indicated in Figure 52.4-8 these comparisons, combined with the expectations of the stakeholders, provide the basis for treatment planning.

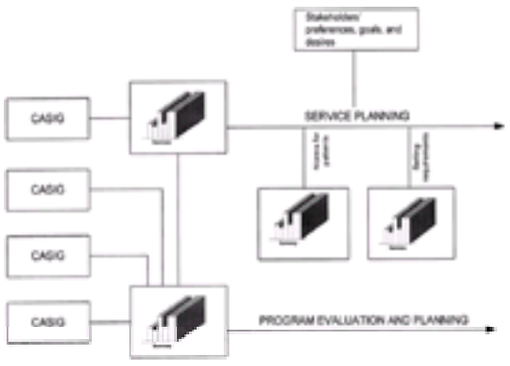

FIGURE 52.4-8 Output of CASIG. Service planning, and program evaluation and planning.

SHORT- AND LONG-TERM GOALS

To guard against deviating from the rehabilitation pathway, both the engagement and assessment processes emphasize the importance of focusing on the rehabilitation goals endorsed and validated by the individual. However, making the patient a full and equal partner in the treatment and rehabilitation process is not always easy. Negative symptoms, disabilities, intrusive hallucinations, institutionalization, and cognitive dysfunction may impede the patient's ability to participate actively and helpfully in this crucial endeavor. The setting and resetting of short- and long-term goals provides the road markers by which clinician and patient can judge whether (or not) they are traveling in the same desired direction.

Within a few months, Mark had noticeably improved his grooming and self-care habits. However, he still had trouble initiating and maintaining conversations. Because of this persistent deficit, the rehabilitation team encouraged Mark to reformulate his goal of pursuing a college education into a long-term plan with several intermediate steps. The team helped Mark identify a series of short-term tasks that would build toward his matriculation at a local community college. Mark began by attending a daily group focused on basic conversation skills at the mental health center. Within a few months, his progress led the rehabilitation team to offer Mark the opportunity to colead the skills-training group, which Mark readily accepted. Within a few weeks Mark was not only cofacilitating the group but also conducting ad hoc tutorial sessions for patients who lagged behind. At a subsequent meeting with the rehabilitation team Mark announced that rather than go to college, he wanted to be hired by the mental health center as a peer advocate. Once again, Mark and the team designed a step-by-step approach that allowed him to work toward this new goal.

The clinician and patient are guided by the continuous interaction of the patient's overall rehabilitation goals with the current stage of the disorder. During the acute, florid stage of the illness, the task for the clinician may be to help the patient move from grandiose, unrealistic fantasies to articulating the more proximal and immediate realistic changes and steps that must be accomplished before the longer-term goals can be reached. As symptoms stabilize and (hopefully) ameliorate, the focus shifts to helping patients define their goals in terms of occupational, student, friendship, familial, and residential roles. Another shift occurs during the late rehabilitation and recovery phase, when individuals begin to look for goals which transcend their status as patients qua patients. In this phase, the goal is to develop and nurture the wellness factors (e.g., hope, courage, self-esteem) that enable patients not only to manage but sometimes to overcome the illness.

Logistically, an individual's preferences, desires, and motivations should be operationalized in behavioral terms and divided into long- and short-term goals. Long-term goals involve months to years and should correlate with overall rehabilitation goals to progress toward functional life roles. Long-term goals should be comprehensive in subserving progress in all relevant domains of life functioning (Social/Interpersonal, Familial, Financial, Recreational, Medical/Psychiatric, Activities of Daily Living/Independent Living Skills, Vocational/Educational, Spiritual, and Housing/Residential). Short-term goals involve days, weeks, or months and should articulate with long-term goals as stepping stones, or subgoals. Short-term goals also must be prioritized and endorsed by all parties participating in the rehabilitation. To facilitate accomplishment, individuals should prioritize their goals and, when possible, have them endorsed by the family, caregivers, and responsible clinicians.

Keith was a 26-year-old man recently diagnosed as having paranoid schizophrenia. When psychotic symptoms were controlled with antipsychotic medications, he engaged in a rehabilitation planning effort with an occupational therapist who was a member of a community mental health center treatment team. When first asked to identify his goals, he said he wanted to enter college and complete his bachelor's degree. He hadn't been in school for 6 years, yet he said he wanted to matriculate as a full-time engineering student in a prestigious school. The clinician thought that goal was unrealistic. On further exploration it was learned that Keith had felt happiest when he was in school, and he now thought this would be the best way to realize his goal of being “happy and independent.”

Once this goal of becoming independent and more pleased with himself was identified, Keith was helped by his therapist to set as a long-term goal resumption of his schooling at a local trade school. He then began to work with his therapist to target skills and deficits in the areas of time management, organizing scheduled activities, and self-assertion. He saw those as essential to prepare for his goal of entering school. He had been an organized person with good study habits prior to developing his illness, so those were seen as potential assets. His basic conversational skills were more than adequate, but he needed training in speaking up to obtain his needs in the classroom, with teachers and with peers. He had the prerequisite resources to engage in the recommended training sessions (i.e., time, a car, and the availability of social skills training services in his local community support program). Six months later, after benefiting from structured and systematic skills training to improve his self-advocacy, he entered a course in automotive mechanics at a vocational school.

MONITORING THE QUALITY OF INTERVENTIONS AND CLINICAL PROGRESS

Biopsychosocial assessment and intervention is an ongoing process requiring the psychiatrist to titrate medications and psychosocial services on the basis of an individual's fluctuating need. In fact, the effectiveness of drug treatment depends on the clinician's ability to elicit, rate, and monitor symptoms, leading to appropriate adjustments (or maintenance) of treatment (e.g., medications, crisis services, skills training). Instruments such as the BPRS have been used by a wide variety of practitioners as sensitive indicators of impending psychotic relapse. Behavioral observations (e.g. time sampling or experience sampling) can be even more sensitive indicators of the need for psychiatric intervention than interview-based ratings, but these are more demanding to implement.

The CASIG evaluation system was designed with the presumption that treatment planning does not end with the collaborative development of the patient's service plan. As shown in Figure 52.4-9, treatment consists of recurring cycles of assessment, planning, and service delivery. These cycles reflect the ever-changing nature of treatment; as the services achieve their effects, all of the stakeholders change. In such a fluid environment, treatment cannot be successful unless assessment and services are intertwined in the on-going process. New information obtained from an updated assessment engenders new treatment plans designed to make the information obsolete and hence require yet another assessment. This pattern proceeds through the patient's journey from acute episode to stabilization, rehabilitation, and (if successful) recovery.

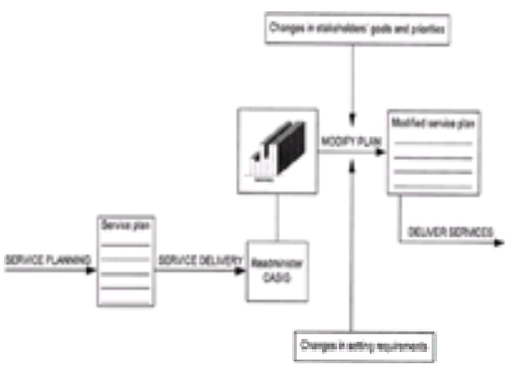

FIGURE 52.4-9 Cycle of service plan, evaluation, and new service plan. This cycle is repeated as often as needed to fulfill service goals.

According to his parents, Matt had been unable to “fit in” since his early teenage years. Now 35 years old, Matt's daily routine consisted of eating at fast food restaurants, working as a stock boy at a local retailer, and spending most of his time alone in his apartment. During the CASIG interview, Matt stated that he had no friends, attended no social gatherings, and was bored most of the time that he was not at work. His parents complained that Matt failed to groom himself regularly and was being threatened with eviction from his apartment for not cleaning properly.

The initial treatment goal derived from the CASIG interview focused on increasing Matt's opportunities for social engagement. He was initially given a homework assignment to count the number of chances he had each day to interact briefly with someone. It quickly became clear to Matt that he had 15 to 25 opportunities each day to say “Hello” to other employees or residents at the apartment building. Consequently, Matt and his therapist updated the CASIG plan, adding social skills training to increase his ability to interact appropriately when the opportunity presented itself.

After learning some brief conversation skills and when to apply them, Matt was next assigned to count the number of times each day that he actually engaged people in brief conversations as well as the percentage of these engagement attempts that were successful. As his percentage of successful engagements increased, he developed more confidence in his ability to “fit in” and thus modified his CASIG plan once again to include attendance and participation in social gatherings. He told his therapist that he wanted to invite a few coworkers to his apartment to watch television after work. This idea required adding goals that he had not considered in his original CASIG plan, such as improved grooming and better housekeeping. Using the recurrent, interconnected technique of the CASIG evaluations allowed Matt to see how his self-identified goal of increased social relationships could be facilitated by improvement in self-care and daily living skills.

FITTING TREATMENT AND REHABILITATION INTERVENTIONS INTO THE GOALS

One must move from the identification of symptoms and functional deficits and formulation of positive goals to the delivery of the appropriate biopsychosocial treatment modalities. Although setting specific, concrete interpersonal goals is perhaps the most challenging step in a patient's rehabilitation, the failure to reinforce these goals with consistent, yet modest accomplishments may lead to discouragement, frustration, and ultimately treatment discontinuation. The clinician choosing and using the needed intervention must encourage patients to reach just beyond their grasp and help facilitate the success of their efforts.

In preparation for his “party,” Matt rehearsed with his therapist each conversation he was to have with his coworkers. Matt had little difficulty in the role play with his therapist, but he reported several failures when he tried to invite his peers to his apartment. This was useful information for Matt's therapist who, rather than simplifying the goal increased his level of intervention by accompanying Matt into the field to provide closer prompts and feedback during the peer contacts. The in vivo support provided by his therapist enabled Matt to make the invitations and successfully host his party. Soon thereafter, Matt began to receive and accept invitations to social gatherings hosted by others, including members of his family.

REHABILITATION SERVICES

Case management services are required to coordinate a wide range of entitlements, community support, and rehabilitation services now available to mentally disabled persons. Specific rehabilitation services include social skills training, family psychoeducation, cognitive remediation, vocational rehabilitation, and self-help programs. The effectiveness of these and other services has been established, and their importance is highlighted by recent notable changes in the nature and organization of care for severely mentally ill persons. For example, introduction of the atypical antipsychotic medications including clozapine (Clozaril), risperidone (Risperdal), and olanzapine (Zyprexa) has led to hypotheses regarding potential enhancement of cognitive and learning capacities that underlie individuals' abilities to benefit from rehabilitation interventions. A wide range of social factors have highlighted the importance of rehabilitation in effecting community reintegration and functioning, including the dramatic reorganization of health care services and funding over the past decade, the growing problem of homelessness and criminalization of the mentally ill, and the Americans With Disabilities Act. The utility and effectiveness of rehabilitation strategies have been recognized, and they are prominently included in a number of treatment guidelines for schizophrenia published by organizations such as the American Psychiatric Association.

As illustrated by the Practice Guidelines for Treatment of Schizophrenia by the American Psychiatric Association:

> [P]sychiatric rehabilitation aims to optimize the recovery of individuals with schizophrenia and other disabling mental disorders. This is accomplished through the use of the full array of biopsychosocial interventions, strengthening the supports and resources available for the person in the community, a collaborative approach with the patients and the natural caregivers, and an emphasis on function rather than symptoms. There is an attempt to improve the patient's performance in social, vocational, educational, and familial roles to achieve the highest quality of life and productivity attainable.

Social Skills Training Social skills training involves behavioral techniques or learning activities that enable patients to acquire instrumental and affiliative skills in domains required to meet the interpersonal, self-care, and coping demands of community living. While skills training can be done with individuals and families, the typical skills-training session is conducted as a group or class with one or two cotherapists as trainers and 5 to 10 mentally ill patients. Sessions last from 45 to 90 minutes, depending on the patients' level of concentration and symptom control, and meet one to five times a week. Since most mentally disabled persons have pervasive deficits in social functioning, skills training should become a central element in their long-term rehabilitation. It is not unusual for skills training to be offered for years, as the person's abilities, goals, and values ascend a hierarchy of community adaptation. Trainers draw on learning activities like those listed in Figure 52.4-10 to help patients master each skill area.

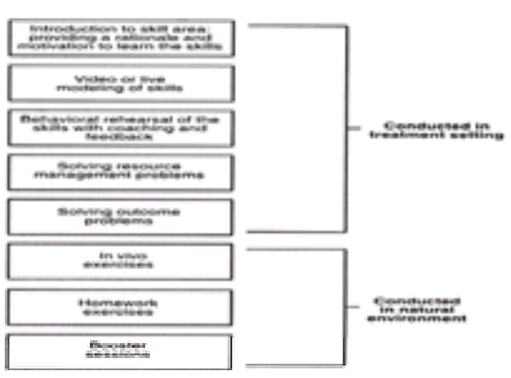

FIGURE 52.4-10 Learning activities used to teach social skills.

In conducting social skills training, therapists engage patients in goal setting and motivational enhancement. Patients are given an understanding of how the skills to be learned can help them achieve their own personal needs and desires. For example, in generating motivation to participate in a conversation skills group, patients are given examples of how conversation can mediate friendships, dating, and job success. The component skills in the skill area are next presented to the patients as learning targets. For example, a trainer leading a basic conversational skills session aimed at teaching patients how to start a friendly chat would say:

> In this skill area, you'll learn three skills that you need to start a friendly conversation. First, you'll learn how to find places where there are people to talk to. Second, you'll learn how to find people who are willing to talk to you. Finally, you'll learn how to find topics of conversation. For example, one way to find a topic is by observing what another person is doing and use that activity as your topic to start a friendly conversation.

Modeling is used next to give patients an opportunity for vicarious learning. Using prearranged situations or situations taken from participants' everyday lives, the

skills are demonstrated by the trainer or by peers in the group. Modeling can be provided by video samplers as well as by role playing by the trainer or another group member. Generalization can be enhanced by having the model or trainer annotate or speak aloud what steps are involved in successfully completing the skills:

> Wow, I'm supposed to go up and start a conversation with that girl. But I'm scared she might turn me down. Wait a minute. I remember in class that it is a good idea to start a conversation about what another person is doing. Let's see, she's sitting over there knitting a sweater so I could start by talking to her about knitting.

Thus modeling and other steps in social skills training combine cognitive techniques with more-overt behavioral techniques into a cognitive-behavior therapy.

Patients are then provided opportunity to try the new skills in guided behavioral rehearsal, or role play. Specifically, patients practice interpersonal situations from their lives that involve the targeted skills and role play a scenario using those skills. Trainers provide active coaching during the role playing and give generous positive feedback afterward by rewarding approximations to the targeted behaviors. Critical feedback can be counterproductive and should generally either be avoided or be couched in constructive terms. One particularly effective training technique involves videotaping the patient's role playing and replaying the video, highlighting the positive signs of progress by the patient.

After patients have demonstrated some mastery of skills in the supportive context of role playing, they are given in vivo homework assignments to facilitate transfer of newly acquired skills into the community. Homework assignments are more likely to be successful when patients are given specific tasks to do in their community (e.g., “Go home tonight and ask your roommate to play cards with you for 10 minutes”) and when they are helped, in advance, to anticipate obstacles to success and to engage in problem-solving exercises to overcome these obstacles (e.g., “Let's anticipate some of the things that might interfere with your inviting your roommate to play cards and what you might do about them.”). Trainers can also accompany patients into their real-life settings to provide coaching in vivo as well as liaison with the patients' natural caregivers to ensure opportunities and encouragement to use the skills. Subsequently, patients need to be reinforced for homework-related efforts. One advantage of doing social skills training in groups is the amplified reinforcement that occurs when patients give their peers positive feedback for implementing their skills in real-life situations.

Variants of Social Skills Training Social skills training can be conducted in the context of individual, family, or group therapy. Goals can be individualized to fit the functional and symptomatic needs of each patient or can follow a prescriptive format in which preset goals are pursued that have general relevance for a large proportion of the patient population. Most patients with chronic schizophrenia can benefit from educational curricula that teach medication and symptom self-management, grooming, social conversation, self-directed recreation, management of personal finances, and family communication. Four variations on the standard skills-training approach have been used effectively to remediate the deficits of persons with severe mental disorders: training of verbal and nonverbal expressive skills in response to specific situational expectations and personal needs; training to improve the accuracy of social perception; training in problem-solving skills; and enhancement of skill generalization via environmental intervention.

Behavioral Response Training In behavioral response training, a task analysis is conducted of the interpersonal situation that is problematic, and socially acceptable responses are targeted as goals for the patient. Each situation is assumed to call for a set of discrete verbal, paralinguistic, and nonverbal skills that combine to form a competent social response. For example, to be a proficient conversational partner, an individual needs to use appropriate forms and amounts of eye contact, facial expression, voice intonation, and interpersonal distance as well as judicious amounts of self-disclosure and social curiosity. Helping patients learn to combine these specific behaviors closes the gaps in their social ineptness.

Abundant research has shown the efficacy of this variant of skills training in reducing deficits in expressive behavior. In addition, the importance of conducting individualized task analyses linked to the phase of a patient's disorder merits emphasis. For example, inpatients struggling with persistent psychotic symptoms who have difficulty finding adequate housing have different behavioral training goals than stable outpatients who wish to improve their vocational functioning. On the other hand, behavioral response training has been criticized for its unidimensional, limited view of social interactions. Impaired processing of social information has been proposed as a more comprehensive model for understanding the social deficiencies of patients with chronic mental disorders. Thus, diminished sensitivity to social cues, inability to comprehend interpersonal problems and generate relevant solutions, poor comprehension of rule-governed social behavior, and deficits at the discourse level of linguistics have been implicated as core limitations in the social behavior of persons with schizophrenia and other chronic mental disorders.

Training in Social Perception Skills Recently efforts have been made to develop strategies for training patients in affect and social cue recognition. Patients with chronic psychotic disorders often have difficulty in accurately perceiving and interpreting the subtle affective and cognitive cues that are critical elements of communication. Social perception abilities are considered the first step in effective interpersonal problem solving; difficulties in this area are likely to lead to a cascade of deficits in social behavior. Training skills in social perception address these deficits and help provide a foundation for developing more-specific social and coping skills.

> Despite attending several social gatherings, Matt felt apart from the rest of the group. He reported that he experienced these events like “a jumble of sights and sounds.” His therapist, recognizing Matt's difficulty with social perception, provided him with a series of questions designed to help organize and give meaning to the social stimuli he encountered. For example, when Matt was confused about a conversation someone was having with him, he would ask himself, “What is this person's short-term goal? At what level of disclosure should I be? Should I be talking now or listening?” Identifying the rules and goals of a particular social interaction provided a template for Matt to recognize and react to a greater variety of social cues, thus enhancing his behavioral repertoire.

Information-Processing Model of Training Methods of training that follow a cognitive perspective teach patients to use a set of generative rules that can be adapted for use in various situations. For example, a six-step problem-solving strategy has developed as an outline for helping patients overcome interpersonal dilemmas: (1) adopt a problem-solving attitude; (2) identify the problem; (3) brainstorm alternative solutions; (4) evaluate solutions and pick one to implement; (5) plan the implementation and carry it out; (6) evaluate the efficacy of the effort and, if ineffective, choose another alternative.

While the step-wise, structured, linear process of problem solving occurs more intuitively and without conscious awareness in normal persons, it can be viewed as a useful interpersonal crutch to help cognitively impaired mental patients cope with the information needed to fill their social and personal needs.

Skill Enhancement via Environmental Organization Skills training is most effective if the patient's environment allows repeated opportunities and encouragement to practice newly learned skills. One example is in vivo amplified skills training. Here the patient's case manager or treatment team, together with the patient, makes direct and repeated efforts to create opportunities, encouragement, and reinforcement for using the acquired skills in the patient's living, working, and social environments. These efforts involve contacts with family, employers, agency officials, peers, and other persons who can facilitate the patient's use of specific skills in the environments where they are most relevant.

A related strategy involves teaching individuals to modify their home environment to incorporate specific cognitive and behavioral cues for skills. Reminder cards, notes, schedules, prompting devices (e.g., remotely activated beepers), and other concrete cues can be systematically used to augment skills learned in training programs.

Modules for Training Social and Independent Living Skills Based on the cognitive model of social skills training, a set of psychoeducational modules was developed at the University of California at Los Angeles (UCLA) Clinical Research Center for Schizophrenia and Psychiatric Rehabilitation for use by interdisciplinary professionals and paraprofessionals in a wide array of mental health facilities and natural environments to teach patients the social and instrumental competencies of key domains of community functioning. The domains include self-management of antipsychotic medication, identifying warning signs of relapse, coping with psychotic symptoms, grooming and personal hygiene, recreation for leisure, interpersonal problem solving, job finding, community reentry, safe and satisfying sex, family coping, street smarts, and engaging in friendly conversations. Those domains were chosen because competency in them has been associated with better social adjustment, longer community tenure, and lowered risk for relapse and rehospitalization.

Each module uses the same highly structured and thoroughly specified instructional techniques, so even paraprofessionals can use them with a minimum of training and consultation. The techniques also compensate for patients' cognitive dysfunctions by using overlearning, visual as well as auditory instruction, and in vivo training. Each module is a self-contained package that can be used alone or in combination with other modules in comprehensive rehabilitation programs.

The training modules were constructed to teach patients specific functional skills, to solve problems they might encounter while attempting to use newly learned skills, and to practice the skills in the natural environment. That model of social skills training offers considerable promise for patients who have the cognitive capacity for

learning social skills in small groups. Each module is divided into separate skill areas, and each area has specific behaviors taught to achieve personal effectiveness and competence. For example, as shown in Figure 52.4-11, the Medication Management Module contains skill areas for (1) understanding the benefits of antipsychotic medication, (2) mastering medication self-administration, (3) coping with the adverse effects of medication, and (4) negotiating medication issues with health care providers. Similarly, the Symptom Management Module comprises four specific skill areas: (1) identifying the early warning signs of relapse; (2) managing a prodrome and developing a relapse prevention plan; (3) coping with persistent symptoms; and (4) avoiding street drugs and alcohol.

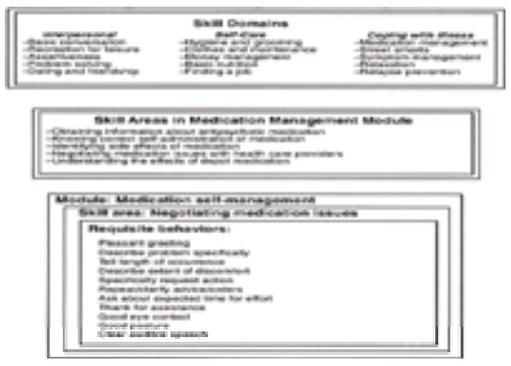

FIGURE 52.4-11 Domains of skills training with skill areas in the Medication Management Module. The fourth skill area comprises specific verbal and nonverbal behaviors that are the educational objectives taught in this skill area.

Each module contains a prescriptive clinician's manual or guide, a videotape demonstrating the skills to be learned, and a patient's workbook containing practice exercises and monitoring forms. Skills are trained by using a combination of focused instructions, videotaped demonstrations, role-playing rehearsals, social and videotape feedback, problem solving, and practice in the natural environment through in vivo exercises and homework assignments. Patients proceed through each skill area in a specific sequence of learning exercises and activities, starting with an introduction that explains the values and advantages of the module. The training procedure emphasizes discovery learning rather than expository or didactic teaching.

INTRODUCTION TO THE MODULE The objective of the first learning activity in the module is to help patients identify the goals of the module, the consequences of achieving the goals, and the steps needed to achieve each goal. Patients are also introduced to the language used in various aspects of training. The introductory exercise consists of a brief description of the module and the topics and goals to be covered in each skill area. Questions such as the following are asked until all patients can answer them correctly:

1. What is the goal of this module?
2. What problems are addressed by this module?
3. If you achieve the goal, what will happen?
4. Do you have time, money, skills, and people to help?
5. What steps are required to achieve the goal?

The purpose of the introduction is to inculcate realistic and favorable expectations and motivation to continue module training. The major goal at this stage is to encourage the patients to think and talk about the material they are about to learn. The therapist concentrates on establishing a highly reinforcing environment rather than on the correctness of patients' responses. However, if patients respond incorrectly to any question, the trainer asks an additional set of ever-more-leading questions designed to prompt patients to give the correct answer.

TRAINING SKILLS Following the introduction, training progresses through the various skills that comprise each domain or area of the module. Training involves two basic sets of procedures. First, patients view a videotaped demonstration of the correct performance of the skills they are about to learn. The tape is stopped periodically, and patients are asked questions to assess their attentiveness and comprehension of the information presented in the videotape demonstration. Incorrect answers result in replay of the videotape and highlighting of the information needed to answer the question correctly when it is repeated.

Next, patients use a role-playing exercise to practice the skills they have just learned. This performance is videotaped for subsequent review by patients and the therapist. The therapist evaluates the performance and provides positive feedback and suggestions for improvement. Role playing with video feedback is depicted in Figure 52.4-12. The role playing is then reenacted, and the process is repeated until the patient exhibits mastery.

FIGURE 52.4-12 A social skills training group at Camarillo State Hospital in which video feedback of the role-playing learning activity provides self-modeling and positive reinforcement for the participating patients. (Photograph courtesy of Thomas Smith, M.D.)

PROBLEM SOLVING The training protocol recognizes that patients may encounter obstacles that make it difficult for them to achieve expected outcomes as they perform their newly acquired skills. Training in problem-solving skills is designed to teach patients methods they can use to overcome the obstacles. The problem-solving model used is a five-step procedure: defining the problem, generating alternative solutions to it, evaluating the alternatives in terms of their potential positive and negative consequences, choosing an alternative based on the evaluation, and implementing the chosen alternative. Patients are taught how to overcome two types of obstacles: resource management problems and outcome problems.

The training in solving resource management problems is designed to teach patients how to gather the resources necessary to implement a particular medication or symptom management skill. For example, even if training in the Symptom Management Module has taught an individual the skills needed to request an appointment with the psychiatrist to evaluate a prodrome, the patient must have access to certain resources such as a telephone to make an appointment and transportation to visit the doctor. A set of resource management problems is presented during the training in each skill area in each of the modules. The therapist describes a skill and asks the following questions:

1. What is your goal in using this skill?
2. What resources must you have to carry out the skill?
3. How would you obtain the resources?
4. If you were to obtain these resources, what positive consequences would happen?
5. If you were to obtain these resources, what negative consequences might happen?
6. Do the positive consequences outweigh the negative consequences? If yes, the method generated by patient is role played; if not, the patient is asked what else

he or she would do.

The training in solving outcome problems teaches patients how to respond when the environment fails to provide the expected outcome following the performance of a particular skill. For example, if an individual arrives for an appointment with the doctor and finds that the doctor has been called away to an emergency, what must the patient do to solve the problem? The training methods are similar to those used during the training in resource management problems.

Training begins as the therapist describes an obstacle that might be encountered as patients attempt to use their skills. Patients are then asked a series of questions that engage them in the problem-solving model.

1. What is the problem?
2. Do you have the time, money, skills, and people to help?
3. What can you do to solve the problem?
4. Is the chosen method feasible?
5. If you use the chosen method, are you likely to achieve your goals?
6. If you were to use that method, what positive consequences would happen?
7. If you were to use that method, what negative consequences would happen?
8. Do the positive consequences outweigh the negative consequences? If yes, role play the alternative generated by the patient; if no, what else would you do?

PROMOTING GENERALIZATION Patients must have the opportunity to practice newly learned skills in the natural environment as an additional step toward programing for generalization. Skills that can only be used in the context of the clinic or hospital are of little use to the patient. In vivo exercises are used to facilitate the transfer of training. Essentially, patients perform the skills in their own world. However, a therapist accompanies the patient to prompt and reinforce the patient's performance and to provide corrective feedback. The exercises are arranged to present increasingly difficult situations that require extending skills beyond the training provided in the clinic. An example in the Medication Management Module is an exercise in which the patient (and therapist) generate questions regarding the patient's antipsychotic medication, after which the patient questions a nurse on his or her treatment team to get the answers. Following that, patients may be encouraged to ask the same or similar questions of their physician or a local pharmacist. In each in vivo exercise the therapist helps patients fine-tune their performance while identifying needed resources and anticipating obstacles that may be encountered during the exercise.

Homework gives patients an opportunity for performance independent of the skills they have learned. Because the module's goal is to teach patients to function independently, homework represents the ultimate step in training. Wherever possible, patients are asked to return from the assignment with tangible evidence that the assignment was completed. For example, if the assignment is to obtain information from a pharmacist about one's medication, the patient can verbally report the information to the therapist and bring back the pharmacist's business card. The role of the therapist in the exercises is to provide feedback to patients regarding their performance.

Validation Meta-analyses and reviews of the more than 50 controlled studies of social skills training have shown that individuals with schizophrenia can acquire and retain skills, and that training results in significant favorable effects on social adjustment, symptoms, relapse, and rehospitalization rates. The most recent skills training studies showed significant methodological advances, including the use of control or comparison groups, increased reliability of assessments, and a better understanding of the cognitive psychology of schizophrenia and the distortions in learning processes commonly encountered. Skills training has consistently led to diminished symptom levels and improved social and interpersonal skills. Strong evidence indicates that skills training leads to skill acquisition and maintenance in schizophrenia, especially if training is intensive (more than two sessions per week) and of sufficient duration (at least 6 months). While schizophrenic patients with even high levels of hallucinations and delusions can acquire skills through systematic training, cognitive disorganization (e.g., severe distractibility and thought disorder) and the deficit syndrome (i.e., primary negative symptoms) appear to interfere markedly with the training process.

Recent advances in rehabilitation include the integration of skills training programs with other standard treatment interventions for schizophrenia. One program combines modular skills training with low-dose antipsychotic medication. This therapeutic strategy is aimed at minimizing neuroleptic adverse effects that can intrude on social functioning while equipping patients with the knowledge and skills to do a better job of managing their illness. Patients participating in this skills training program demonstrated significant skill acquisition and also showed gains in social activities, personal well-being, and overall adjustment. When patients learn through skills training to identify the early warning signs (prodromes) of relapse and to seek early intervention to avert relapse, outcomes are markedly improved, especially with lower doses of medication in patients who had an early onset of their illness.

Other recent developments in skills training involve the use of briefer, highly focused training protocols for commonly encountered clinical situations. In one program, the Community Re-Entry Module, hospitalized patients are taught the skills necessary to collaborate in their discharge planning. These skills are a critical aspect of treatment, given the obstacles in engaging with community-based treatment programs and the increased pressure on clinicians to keep inpatient treatment as brief as possible. A training protocol for chronically mentally ill individuals with comorbid substance abuse problems aims to teach skills necessary to manage the cravings, high-risk situations, and stressors contributing to substance abuse.

The impact and utility of social skills training can also be documented through professional validation, that is, the extent to which it is being implemented by practitioners and mental health facilities. The modules for training social and independent living skills developed at the UCLA Clinical Research Center for Schizophrenia and Psychiatric Rehabilitation have been implemented in 48 of the United States by over 1000 private and public hospitals, community support programs, mental health centers, and practitioners in private practice. A survey of module users found that 61 percent of the respondents used the modules regularly, and another 36 percent planned to institute or reinstitute the modules when administrative and logistical obstacles were overcome. In addition, the modules have been translated into 15 languages and are in use in Asia, Europe, South America, and Africa. Controlled studies have validated the efficacy of the modules in Quebec, France, Switzerland, Germany, Finland, Bulgaria, Poland, and Japan. The growing use of skills training methods around the world and their adaptation to local culture and conditions heightens the value of this modality as a major technique in psychiatric rehabilitation.

Three types of rehabilitative interventions overlap with social skills training in terms of treatment goals and behavioral learning principles: (1) family skills training and psychoeducation, (2) token economy and social learning programs, and (3) cognitive remediation. The choice of a particular rehabilitative modality depends on the needs of an individual as determined through symptomatic and functional assessment; the phase of illness; level of incapacity, structure, and supervision required; and the availability of resources to deliver the services. Thus, individuals with serious cognitive, functional, and symptomatic problems who need frequent prompting and reinforcement to learn basic life skills will benefit from a token economy. Patients who have contact with their family members will benefit from a psychoeducational and skill-building program aimed at improving communication and problem solving. Cognitive therapies and remediation, newer treatment modalities that have yet to prove their clinical validity, represent the cutting edge of treatment development for the attentional, memory, symptomatic, and executive dysfunctions of persons with schizophrenia that endure, limit rehabilitation, and are refractory to pharmacotherapy.

Family Psychoeducation and Behavioral Management Removing the barriers to the generalization of social and independent living skills can be facilitated by involving both the family and the patient in a collective enterprise of (1) education about the patient's particular mental disorder and the ways to obtain professional and community services for it, (2) training in communication skills, and (3) using the communication skills in systematic application of problem solving with the stressors and daily hassles that inevitably accompany mental disability. Because the family comprises an important part of the patient's natural support system and living environment (even chronic mentally ill patients living apart from their relatives may maintain significant personal and phone contact with them), improving the family's attitudes, knowledge, and coping skills regarding serious mental illness can be rapidly translated into better opportunities, encouragement, and reinforcement for using skills learned in the family therapy arena and other rehabilitation settings.

Many international studies have replicated the findings that family stress—reflected in criticism of and emotional overinvolvement with the mentally ill relative—is the most powerful predictor of relapse in schizophrenia and mood disorders. Therefore, several modes of family intervention were designed and empirically validated for the ability to equip relatives with coping skills and thereby change the emotional climate of the family and reduce the incidence of relapses and rehospitalizations. From the vantage point of the vulnerability-stress-protective factors model of serious mental illness, families are viewed as struggling with the stress of positive and negative symptoms and disability of their severely mentally ill relatives. In an interactive process the stress load produces demoralization, tension, and emotional and financial burdens on the family, while it can precipitate relapse in the index patient with fragile coping capacity and significant psychobiological vulnerability to stress.

Disavowing the anachronistic and antitherapeutic theories of the past that implicated family interaction in the etiology of schizophrenia, clinicians now engage relatives as essential allies in treatment and rehabilitation. The psychoeducational model has five basic aims. First, the treatment team must develop collegial relationships with family members and other support persons. Since many mentally disabled persons live apart from their families in group homes or half-way houses, the operators and staff of these facilities in psychoeducation must also be involved. Family members and support persons need coping skills and social support, and

active engagement in the educational process is encouraged by involving them at all stages as colleagues and collaborators in the patient's treatment. Secondly, a specific psychoeducational program is used to teach what is known scientifically about the patient's mental disorder and direct the family and other caregivers to locally available treatment and rehabilitation services. Topics might include information about positive and negative symptoms, effects of psychotropic medications, relapse prevention, crisis management, and treatment options for psychiatric rehabilitation. As much as possible, empirically validated treatment principles drawn from the clinical psychiatric literature are used in the psychoeducation, carefully translated to a layperson's level of comprehension.

Factual material is typically presented during the first half of an education session followed by a guided discussion in which families are helped to personalize their learning with comments or questions regarding their own experiences. A third principle underlying the psychoeducational approach is critical at this point, the assumption of least pathology. Problems understanding or managing a family member's illness are assumed to be a consequence of ignorance or the lack of necessary skills for managing specific situations, not of family dysfunction. Family members are assumed to be doing the best they can and acting in the interests of the patient and family, given their coping capacities and biobehavioral vulnerabilities. With this approach, setbacks are considered to result from lack of knowledge, treatment services, or skills, and the demoralization and extreme self-criticism that severely inhibit the family's efforts to help are minimized. Instead, treatment alliances are strengthened, and families can successfully pursue the next step, which involves developing step-by-step communication and problem-solving skills, the fourth major aim of the psychoeducational approach.

To produce durable clinical effects, family interventions must go beyond education and train family members in such necessary coping skills as basic communication and contingency management. Examples of communication skills that form the basis for effective problem solving include active listening, giving positive feedback, and making positive requests. With these skills, families can better address recurring difficulties associated with patients' symptoms and disabilities and the stressors of everyday life. Skills training strategies similar to those reviewed above are used to help individual families learn to interact in more constructive and goal-oriented ways. Thus, family psychoeducation is a specific form of social skills training, using the family system as the vehicle for targeting goals and skills. For example, the six steps of problem solving are taught using real-life problems faced by the family. Often one family member is chosen as leader of a problem-solving session who guides the family through a process, as shown by the family group in Figure 52.4-13.

FIGURE 52.4-13 A family participating in a problem-solving session of behavioral family management, led by a therapist. (Photograph courtesy of Thomas Smith, M.D.)

Most of the members of the Hanrahan family had been concerned about John's smoking in his room at night. His parents, Jackie and George, were particularly worried that John might fall asleep smoking in bed and start a fire. Using the problem-solving steps they learned in the family skills training sessions, the parents decided to call a meeting to discuss the matter. The Hanrahan family rotated leadership of the meeting each session; today it was John's sister Beth's turn to review the steps.

Although John first saw the problem as his parents' nagging him about his smoking habits, with gentle pressure from Jackie and Beth, John soon agreed with the others that he had a problem with smoking in his room at night. Beth then encouraged everyone to generate solutions to the problem. Jackie said that John should be encouraged to learn new ways to stop smoking and thought he should be rewarded $1.00 each day he went without smoking in his room. George did not like the idea because he thought this would only coddle his son. Beth reminded George that the family should only evaluate the solutions after several had been generated and challenged him instead to come up with a solution. George retorted that they should take away John's smoking privileges altogether.

After listing six solutions, the group reviewed the pros and cons of each alternative. Family members were then asked to vote on the worth of each solution. Four of the five agreed that “finding John another place to smoke at night” would be a good alternative. To implement the solution, Beth volunteered to help John walk through the house after the meeting to find a safe and quiet place to smoke between 8:00 and 10:00 PM. They agreed that the basement family room would be just such a place. The plan was to start the next night, and a week later the group would reconvene and discuss progress.

Next week it was George's turn to lead the family meeting. Under his guidance, the group discussed whether the frequency of John's bedroom smoking had decreased. John reported that he had smoked in his room at night only once the preceding week, which pleased everyone. Jackie, however, said she now found watching TV in the family room a bit unpleasant at night because of John's smoke. Because the original target of bedroom smoking had decreased markedly, the family decided to wait a week before deciding whether and how to tackle the smoke problem in the family room.

The last basic aim of psychoeducation programs involves developing a network of like-thinking people. Chronic mental illness presents substantial burden, stigma, and isolation to families, and a network of involved, understanding support persons and resources is of tremendous help. Psychoeducational programs encourage not only collaboration regarding the treatment process, but also active socialization and the development of lasting, personal relationships among families and treatment staff. Relationships are also encouraged with self-help groups like the affiliates of the National Alliance of Mentally Ill or local mental health associations.

Validation Over two dozen well-controlled studies in the past decade from several countries have examined the effects of family psychoeducation and coping skills training on patients and their families. Results showed that the rate of relapse and subsequent hospitalization for patients who participated in that type of treatment was significantly lower than for patients who completed various comparison treatments. In one study, only 6 percent of patients who participated in behavioral family management's problem-solving treatment with their families relapsed in a 9-month period following treatment, while 44 percent of the contrast group receiving individual supportive therapy relapsed. Over 2 years, only 11 percent relapsed in the family program, versus 78 percent of those who received supportive therapy. Moreover, the patients who participated in behavioral family management also gained significantly more in social adjustment and required less overall neuroleptic medication.

Recent studies show that family psychoeducation programs have significant benefits in areas other than symptom and relapse management for the family member with a chronic psychiatric disorder. When psychoeducation and skills training are conducted in a multiple family group format, family burden is lowered, and relatives' sense of self-efficacy regarding their ill family member improves. Participating in a multiple family group can reduce stress, isolation, and stigma experienced by family members; the burden of care no longer seems unique, and families can exchange helpful suggestions and coping techniques with each other. Evaluation of the clinical effects of multiple family groups suggests that the format may be at least as effective as single-family psychoeducation in reducing relapse while also offering special advantages in case management and vocational rehabilitation. The components of behavioral family management, now replicated in two subsequent studies, are listed in Table 52.4-2.

Behavioral analysis of all members of the family
Defining individual and familywide problems
Setting goals for each individual and family system
Identifying reinforcers
Education of all family members on nature of schizophrenia and currently available treatment and rehabilitation modalities
Training in communication skills
Expressing positive feelings to others and acknowledging when others do or say something positive toward you
Active, reflective listening
Making positive requests and asking for what you want
Expressing negative feelings in constructive ways
Training in problem-solving skills
Be specific and objective in describing the problem
Express how you feel directly and subjectively about the problem
Listen to each other actively and reflectively as the problem is described and feelings are expressed
Help each other generate alternatives and options in dealing with the problem
Weigh the potential consequences or outcomes (risks and benefits, pros and cons) of each alternative
Choose a reasonable alternative
Decide how to implement the alternative
Behavioral interventions for specific problems
Contingency management for negative symptoms
Job-finding skills training
Friendship skills training
Independent living skills training
Realistic and incremental expectations for performance
Reinforcing process through successive approximations (shaping)

Table 52.4-2 Component Interventions of Behavioral Family Management

Research on various forms of family intervention suggests that training in communication and problem-solving skills may be necessary to reduce family burden and stress-induced relapse as well as being prerequisite for achieving gains in social functioning. Family psychoeducation and skills training can be conducted with success in the clinic or in the family home. Recent work suggests that the modality may be as applicable for bipolar disorder and recurrent depressions as for schizophrenia. Studies have not yet isolated the specific, effective ingredients of behavioral family intervention or determined how much of the therapeutic effect results from the nonspecific social support that ensues from discussing common problems.

Cognitive Remediation The cognitive disorganization and neuropsychological deficits commonly seen in schizophrenia play a major role in determining the success of psychiatric rehabilitation strategies. Enduring thought disorder, short-term memory, and verbal learning deficits are more predictive of skill acquisition than psychotic symptoms. In addition, deficits in vigilance, memory, and executive functioning have been repeatedly associated with social skill and overall social adjustment. Further evidence of the importance of neurocognitive deficits in determining social adjustment in schizophrenia comes from studies of antipsychotic medications, especially the novel medication clozapine. Clozapine has greater efficacy than typical antipsychotic medications, which has been attributed in part to improved neurocognitive capacities that mediate the development and use of social skills. Recent studies suggest that clozapine and newer antipsychotics such as risperidone and olanzapine exert a beneficial effect on executive functioning, short-term memory, and attention systems mediated by basal ganglia function.

While many clinicians have assumed that the cognitive deficits of schizophrenia represent an irreversible, enduring form of dementia that cannot be mitigated by rehabilitation, evidence is accumulating to justify strategies for remediation of these basic cognitive deficits. Two types of cognitive remediation strategies have been developed that differ in the behavioral targets for remediation and the nature of the intervention. Both strategies aim to elevate the individual's capacity to learn a wide variety of social and independent living skills, thereby permitting psychiatric rehabilitation to begin at a higher plateau of rehabilitation readiness.

Direct Remediation of Basic Cognitive Deficits Individuals with schizophrenia have pervasive cognitive deficits that are substantially independent of psychotic symptoms. In addition to treating psychotic symptoms with traditional methods, programs have been developed to address these more basic information-processing deficits. Demonstrations have shown that laboratory-based measures of cognitive dysfunctions (e.g., reaction time, memory, card sorting, span of apprehension, and vigilance) can improve significantly with behavioral training. A variety of remediation strategies can be used, including repeated practice, instructional modification such as providing ongoing detailed instructions with cues and immediate feedback throughout a training session, positive reinforcement in the form of money or points that can be redeemed for desired incentives, and errorless learning that teaches discrimination and problem solving in small steps that maximize success and minimizes trial-and-error learning. As shown in Figure 52.4-14, the patient participating in this form of cognitive remediation is typically seated in front of a computer screen and instructed to repeat multiple trials on vigilance, memory, reaction time, discrimination, or concept tasks. As a rehabilitation instrument, the computer provides consistent and repeated trials, while monitoring patients' responses and adjusting test stimuli and consequences to shape gradually improving performance on the cognitive task.

FIGURE 52.4-14 Computer-assisted cognitive remediation. A patient is viewing the video monitor during a abstract reasoning task. (Photograph courtesy of Robert Kern, Ph.D.)

Recent efforts have incorporated more complex models of social cognition into remediation protocols for schizophrenia. In one approach, patients are involved in a hierarchical series of training subprograms, beginning with cognitive differentiation and proceeding through social perception exercises to verbal communication, social skills, and interpersonal problem-solving phases. A modification of this strategy involves methods for assessing discrete cognitive styles (e.g., reality distorters, impoverished, or disorganized), with the subsequent training program designed to address deficits in social cognition hypothesized to underlie the specific cognitive style. Data from these and other similar studies will help elucidate the nature of social cognition and the most appropriate interventions. Another approach adds vocational relevance to the cognitive training by having patients maintain a focus on concentration while performing quasi-job tasks such as folding, cutting, stapling, and sorting papers.

Empirical evaluation of the efficacy of these direct approaches to cognitive training suggests that striking improvements—approaching normalization of some functions—are possible in the cognitive tasks, but the links between improvements in laboratory-based molecular levels and in molar social and clinical status remain to be documented. Current research aims to promote clinical generalization of direct cognitive remediation by identifying and strengthening the cognitive, behavioral, and social processes that mediate learning of adaptive skills. For example, short-term verbal memory and sustained attention are possibly rate-limiting in the acquisition of social and instrumental skills. Vocational outcomes are predicted much more accurately by the patients' attentional capacities than by their symptoms of disorder; hence, current cognitive remediation efforts are directed at normalizing those deficits of attention, concentration, memory, and problem solving that appear to be the core obstacles to a person's responsiveness to rehabilitation. However, the ecological validity of laboratory-based remediation of cognitive dysfunctions awaits demonstration.

Cognitive-Behavioral Therapy of Psychotic Symptoms Another remediation strategy ameliorates psychotic symptoms through cognitive restructuring and behavioral learning principles. To date, strategies have been developed for treating delusions, hallucinations, and negative symptoms. Sometimes referred to as cognitive strategy enhancement, these approaches involve identifying specific symptoms and subsequent training in the use of cognitive coping strategies. Commonly used strategies include distraction (e.g., listening to music or humming when auditory hallucinations occur) or methods such as reframing, self-reinforcement, reality testing, and verbal challenging, all of which have the patient use cognitive protocols to test the logic of specific delusional beliefs. For example, delusional beliefs can be altered by collaborative, reframing methods with an investigative style that leads the patient to discover and test the dysfunctional quality of the irrational beliefs.

Mr. Smith believed he was being poisoned by his father. As a result he was anxious and paranoid during home visits. Rather than confronting the patient's delusion with reality testing, Mr. Smith's therapist engaged him in an examination of how these beliefs had an adverse impact. For example, harboring the persecutory ideas prevented him from feeling close to his father and from enjoying his home visits. The therapist helped Mr. Smith identify times when he felt safe and secure in the presence of his father, such as when they watched a football game together on TV and had some spirited conversation. The contradiction between happy times and fearful times became the opening wedge for questions about the validity of the patient's delusional beliefs.

Self-instruction and self-reinforcement can be used to counter delusions and hallucinations and to improve concentration.

A lawyer with a paranoid delusional disorder was chronically fearful that his colleagues were trying to extrude him from the firm. He was constantly on guard, vigilant, and tense when interacting with them. He was taught to take a more proactive stance at the firm and, instead of waiting for someone to say something, instructed himself to approach his colleagues each day, smile, and exchange some pleasantries. When these social initiatives had succeeded in disarming them, he was given practice in rewarding himself by saying, “That was a good gambit!”

A final therapeutic variant in this category is the use of cognitive-behavioral modification techniques to increase adaptive behavior and reduce the frequency and preoccupation of thought disorder, delusions, and hallucinations that are refractory to antipsychotic medication. Contingent reinforcement, reinforcer sampling, discrimination training, fading procedures, and discrete trials learning have all been effective, at least during short-term follow-up in hospital and clinic settings. Much of the emphasis is on increasing adaptive behavior that is incompatible with the symptoms and that, used frequently, will displace the psychotic symptoms. For example, in the attention-focusing procedure, repeated prompting, modeling, and reinforcement of appropriate verbalizations during many conversational learning trials enhances acquisition of verbal skills and reduces incoherence. The high density of prompts, modeling, and reinforcement makes it almost impossible for a thought-disordered patient to make incorrect responses.

Token Economy While systematic, planned provision of social and tangible reinforcers for patient's participation and progress is an indispensable element of social skills training, family psychoeducation, cognitive remediation, and vocational rehabilitation, contingent use of reinforcement becomes a round-the-clock necessity in token economy and social learning programs for the most seriously mentally disabled patients. If the pervasive positive and negative symptoms, instrumental role deficits, and intolerable, acting-out, and bizarre behaviors of treatment-refractory patients are to be remediated, the entire therapeutic milieu and its staff cadre must be united and consistently deliver systematic positive and negative consequences to patients' behaviors and encourage adaptive and prosocial behaviors from them.

The use of tokens, points, or credits as secondary or generalized reinforcers can be seen as normalizing a mental hospital or day hospital environment with a program mimicking society's use of money to meet instrumental needs. Token economies establish the rules and culture of a hospital inpatient unit or partial hospitalization program, offering coherence and consistency to the interdisciplinary team as it struggles to promote therapeutic progress in difficult patients. These programs are challenging to establish, however, and their widespread dissemination has suffered because of the organizational prerequisites and the additional resources and rewards needed to create a truly positively reinforcing environment.

Elements The first step in designing a token economy is targeting specific behaviors to be increased or decreased in frequency. For example, self-care and appropriate social participation in therapeutic activities would be targets for positive reinforcement, while aggressive and intrusive behaviors might result in token fines. Each behavior needs to be described well enough to be recognized reliably by patients and staff. For example, showering includes gathering soap, shampoo, towel, and clean clothes; preparing the shower area; washing the body; shampooing the hair; rinsing and drying the body; and cleaning up the area after use. Without extreme specificity, patients and staff may not reliably agree upon what behaviors justified token rewards.

After identification of token economy targets, contingencies need to be created that govern the consequences of those behaviors. Contingencies describe if-then rules connecting a target behavior with a reinforcer. Staff specify payoffs much like the contingencies in Table 52.4-3, with the size of the payoff differing across behaviors and proportional to the difficulty of changing that behavior. As the token economy progresses, specific contingencies can be adjusted depending on the frequency with which individual behaviors are performed by the patient group and the fluctuating rate of commodity purchases.

Token earnings	
Morning rising from bed and getting dressed on time	3
Satisfactory completion of morning activities of daily living	3
Satisfactory participation in a social skills training group or recreational therapy activity	10
Satisfactory participation in individual behavioral therapy session	10
Satisfactory participation in leisure time activities (per activity)	5
Meets criteria for dress and grooming checks during day (per check)	3
Showers satisfactorily	3
Completes assigned jobs or tasks on unit (per job or task)	4
Participates in off-unit vocational rehabilitation or adult education activity (per half-day)	10
Token fines	
Smoking rule violation	5
Lying on floor	5
Stealing	10
Forgery of token credit card	10
Assault or property destruction	20
Late return from grounds privileges	20
Reinforcers available for tokens	
Cigarettes	4
Drinks (coffee, tea, sodas, hot chocolate)	10
Snacks (potato chips, pretzels, ice cream, candy)	10
Grounds privileges (per half-hour)	4
Music time (per half-hour)	4
Private room time (per half-hour)	4
Nintendo, Walkman stereo, private TV (per half-hour)	4

* This token economy uses a card that can be punched with holes to document token earnings and purchases. The token economy has three levels, which differ in the flexibility and type of reinforcers and privileges. At the highest level of performance, the patient carries a “credit card” and has full access to all unit privileges and rewards without having to pay with tokens.

Table 52.4-3 Contingencies of Reinforcement in the Token Economy Used at the Camarillo-UCLA Clinical Research Unit*

Frequently, behavioral contingencies are confounded by opportunity. Staff must be able to observe the patient perform the skill so the patient can be reinforced. Hygiene activities are easily scheduled into the daily routine so that staff know when to observe patients perform the target. However, other skills like social behaviors are performed continually throughout the day. Staff may miss a learning opportunity if they do not observe the skill or behavior being exhibited.

Patients must also have the opportunity to perform the targeted skill. Patients cannot be reinforced for leisure activities with peers if they do not have access to games or other recreational activities. Similarly, if the goal in a medication management program is to increase the rate of appropriate drug self-administration, then patients need to be given independent access to their drugs.

After targets are fully defined and reinforcers are in place, rules must be developed to designate how patients receive and return tokens. Administrators of token economies need to define where and when tokens can be exchanged for commodities and other reinforcers. Typically, treatment units dedicate one small room to store cigarettes, decaffeinated coffee, candy, pop, personal radios, videos, and magazines. The store needs to open at least every 2 to 3 hours during waking hours so that patients have ample opportunities throughout the day to exchange tokens for reinforcers.

The prices for commodities are listed at the token store and vary depending on the demand for the item and the cost of purchasing the item from outside retailers. Thus, prices of frequently purchased items may be periodically inflated to maintain the attractiveness of all items in the store.

Empirical Findings Token economies are among the best replicated modality in the treatment of chronically ill psychiatric patients. Studies have shown that symptoms, bizarre behaviors, aggression, and self-care skills of psychiatric patients in token economies have improved significantly even after other treatments were ineffective. Many early studies of token economies focused on chronically institutionalized psychiatric patients and found that many patients in token economies were discharged from the institutions with positive effects that were maintained for up to 2 years. While token economies remain widely used in long-term inpatient programs, recent research shows that these strategies can also be used effectively on acute inpatient psychiatric units and in community-based residential and rehabilitation programs. Token economies are also useful treatment programs for geriatric, chronically aggressive, and dual-diagnosis patient populations.

An exemplary token economy, embedded in a comprehensive social learning program that included skills training, reinforcer sampling, and individualized goal setting, was studied in a rigorous experimental design by Gordon Paul and Robert Lentz. The absolute efficacy of this social learning program (in contrast to milieu therapy or custodial care) was documented in the realms of adaptive functioning, psychotic symptoms, bizarre behaviors, aggression, and cost-effectiveness. For example, the average social-learning patient increased interpersonal and communicative skills by over 1200 percent of entry level. Twenty-five percent of these severely disabled patients improved enough to be indistinguishable from the normal population on a blind, time-sampled behavioral checklist. Over 97 percent of the social-learning patients were successfully discharged to the community with tenure exceeding 120 days, compared with 71 percent of the milieu therapy group and 29 percent of the

custodial patients.

A token economy operated continuously for 26 years at the Camarillo State Hospital, cosponsored by the California Department of Mental Health and the UCLA Clinical Research Center for Schizophrenia and Psychiatric Rehabilitation. Patients were recruited from among the most impaired, treatment-refractory population in the hospital. An integrated pharmacological and behavioral program of interventions was applied. Table 52.4-4 shows the results of a cumulative empirical evaluation of the program, based on direct observation of problem behaviors during patients' baseline assessments and the frequency of the same behaviors in the month prior to discharge. A wide array of positive and negative symptoms, psychosocial deficits, and bizarre and intolerable behaviors were successfully modified in this exemplary model of public-academic liaison. More importantly, follow-up evaluations found that over 80 percent of gains were maintained, and half of the patients discharged from the unit maintained residence in the community.

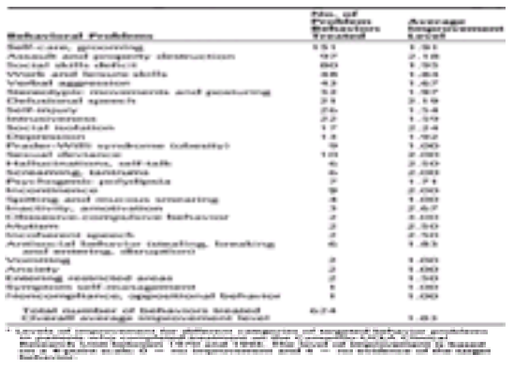

Table 52.4-4 Behavioral Outcomes*

For the increasing reservoir of treatment-refractory psychotic patients living in state hospitals, forensic institutions, or community residential facilities, the token economy is the psychosocial treatment of choice. To ensure a high-quality, ethical token economy (1) staff must be well trained in providing positive reinforcement and shaping successive approximations to adaptive behavior, (2) competent supervision by a behavioral psychologist is essential, (3) administration must provide extra support so that the environment is enriched with reinforcers rather than resorting to behavioral control through deprivation and response costs, (4) the token contingencies should be supplemented with approximately 30 hours per week of planned and scheduled rehabilitation activities consistent with social learning methods, and (5) maintenance and generalization of clinical improvements must be promoted through well-planned aftercare programs and gradual fading of the structure and frequency of the contingencies. As an example of the principle of fading, patients being readied for discharge can be shifted from daily receipt of tokens to a credit card level in which they have free access to all the rewards and privileges of the unit noncontingently as long as they maintain a reasonable level of performance of self-care, social skills, and instrumental skills.

While a faithfully designed and implemented token economy has predictable and salutary effects on chronic and refractory schizophrenia patients, few ideal programs exist. Administrative support is essential for a token economy to succeed. Other variables contributing to patients' benefiting from a token economy include higher intelligence quotient, female sex, and involvement of patients more actively with staff in determining the specific tasks and goals of the program. Patients who negotiate aspects of their contingencies are more likely to comply with the behavior therapy program.

Psychosocial Clubhouses and Self-Help Programs Fellowship clubs of formerly hospitalized patients were developed in the early years of deinstitutionalization. Many psychosocial clubs now exist in most large cities in the United States. The most influential organizations that spearheaded the clubhouse movement include Portals in Los Angeles, Fountain House in New York City, Horizon House in Philadelphia, Center Club in Boston, and Thresholds in Chicago. Clubhouses help satisfy their members' basic needs for acceptance, friendship, advocacy, housing, destigmatization, social and recreational activities, and emotional support. Through the years, the goal of those programs has gone from stabilizing individuals in a normalizing environment by providing peer support to offering opportunities for people to participate actively in planning their own rehabilitation program, including the chance to engage in meaningful and gainful work. Although the psychosocial clubhouse movement developed separately from the more medically oriented community mental health centers, in the past few years an increasing number of community mental health centers have converted their traditional day-treatment programs to the clubhouse model.

Central to the psychosocial self-help philosophy is the belief that individuals with mental disabilities have a fundamental right to work, socialization, and a home and that these basic needs, when satisfied, generate self-esteem and the positive identity necessary for community adjustment. Thus, psychosocial clubhouses focus on developing employment opportunities, peer support, and housing programs tailored to the capabilities of their members. Recently, the increasing influence of consumer empowerment as a philosophy, with its emphasis on patients as consumers of services as well as choice vis-à-vis mental health planning and service delivery, has spurred the growth and popularity of psychosocial self-help programs. The success of those programs, including those operated by patients themselves (so-called consumer-run agencies), is reflected by reimbursement of the services by Medicaid and Medicare and state departments of mental health and rehabilitation, by their certification by accreditation agencies, and by the vitality of the International Association of Psychosocial Rehabilitation Services, an organization spawned by the clubhouse network. Increasingly, the psychosocial self-help programs are evolving toward full-service enterprises with assertive community management, pharmacotherapy, and community support.

Validation For the most part, research on the psychosocial clubhouse model has been impeded by the regrettable barriers between academia and the clubhouses, resulting in studies that lack diagnostic rigor, random assignment of subjects, and reports of subjects' medication status. The outcome research that has been published provides supportive evidence for increased community tenure, increased adaptation, and enhanced satisfaction of members, compared with traditional approaches. However, there is a relative lack of empirical research about the effectiveness of the clubhouse model, in spite of the recognition and acceptance that it has received.

Fountain House, the prototypical clubhouse, was also the first to evaluate outcome systematically. The outcome measures selected were consistent with clubhouse goals, namely, to create a restorative environment within which individuals disabled by mental illness could be helped to lead vocationally productive and socially satisfying lives. Thus, early studies focused on preventing hospitalizations. For example, a 2-year study found that individuals randomly assigned to receive peer outreach services from Fountain House members had a 37 percent rehospitalization rate, compared with the 77 percent recidivism rate of their counterparts who did not receive peer outreach services. Other clubhouses have achieved similar outcomes in randomized, controlled studies aimed at decreasing rehospitalization rates.

The establishment of a Rehabilitation Research and Training Center at the Thresholds Psychosocial Clubhouse has permitted the study of a broader variety of outcome measures. For example, a number of programs that focused on different aspects of psychosocial functioning have been successfully implemented at Thresholds including assertive community treatment, supported education, supported housing, and human immune deficiency virus (HIV) risk reduction. However, studies of transitional employment have found that this prolongs artificial work settings and impedes eventual gainful employment.

Vocational Rehabilitation Vocational rehabilitation has traversed three distinct eras in the pursuit of improved employment outcomes for seriously mentally ill persons. The first was institution bound, with community- or hospital-based, sheltered work programs as the mainstay of intervention. Sheltered approaches were found counterproductive; their dead-end jobs offered no future hope for real jobs and taught few marketable work skills.

The second era was ushered in by the community mental health movement of the 1960s and by the psychosocial rehabilitation clubs. While the community mental health movement set the necessary conditions for building gainful vocational programs, its potential was never realized because most vocational efforts were sponsored by state departments of rehabilitation whose rehabilitation counselors' activities took place outside the main body of the mental health treatment sector. State-sponsored vocational rehabilitation programs tended to focus their efforts on physically disabled and on less seriously mentally ill persons.

Transitional employment programs, such as those in the Fountain House type of psychosocial clubs, evolved independently and often without much input from mental health professionals. Transitional employment involved prevocational work activities or work enclaves in industrial settings with the philosophy of train-then-place in a real job. A job in the competitive work sector was considered stressful, requiring gradual work hardening through jobs intrinsic to the psychosocial clubhouses: clerical services, food preparation, janitorial services, and other maintenance tasks. Research has shown that lengthy periods of work hardening in transitional employment

actually reduce the likelihood of competitive employment and that accelerated placement—as occurs in supported employment—yields better outcomes. That isolation unfortunately still exists today, and most psychiatric treatment systems neglect the area of vocational rehabilitation.

A renaissance in vocational rehabilitation, sparked by emergence of the supported employment movement, may be developing in the 1990s. The federally sponsored supported employment initiative springs from the understanding that persons with psychiatric disabilities require ongoing services, such as training in the skills necessary to maintain employment, once they secure competitive employment. It deemphasizes the importance of what has been labeled prevocational training, advocating instead a place-then-train approach. Based on their interests, skills, experience, and deficits, clients are placed in employment settings and then offered the training and supports necessary to maintain their positions. In its fully applied form, persons are offered services indefinitely, with job coaches visiting them at their workplaces to help them learn and retain the technical, interpersonal, and problem-solving skills they need to sustain employment. Supported employment requires close collaboration and communication among the client in the job, the vocational specialist (or job coach), and the community-based, interdisciplinary treatment team responsible for medication management and case management.

For example, the vocational specialist or job coach assesses a client's stamina, emergent psychopathology, medication adverse effects, and interpersonal relations directly on the job site. In league with the other members of the community mental health team, the job coach or case manager ensures that interventions are delivered to strengthen stamina, nip relapses in the bud, consult with concerned employers, control adverse effects, and improve social skills. In mastering the technical aspects of a job, the job coach carries out a task analysis of the requisite work and subdivides the training process into component, incremental steps. In contrast with the supported employment of mentally retarded persons, in which hands-on, coach-client contact is important, job coaching for mentally ill individuals demands more crisis intervention and liaison with psychiatrists who provide medication to promote clients' tenure on the job.

What are the determinants of successful vocational rehabilitation? The following factors make independent contributions to an individual's prospects for obtaining and retaining a job: psychiatric diagnosis and severity of psychopathology (especially conceptual disorganization and negative symptoms), stress in the family (expressed emotion), neurocognitive functioning, being on a disability pension (disincentive to work), poor premorbid social and work adjustment, no recent work experience, age and gender (older individuals and males are less likely to find jobs), interest and desire to work, and availability of supported employment imbedded in a comprehensive array of community-based psychiatric services.

The Individual Placement and Support (IPS) Program, developed at the New Hampshire–Dartmouth Psychiatric Research Center, is a model of supported employment for individuals with long-term impairments due to severe mental illness. The essence of this model is integrating employment specialists into case management or mental health teams to provide clients with practical assistance in finding and maintaining competitive employment. A critical component of this model is that it is time-unlimited; follow-along support from the mental health team, rather than closure of services by a state's vocational rehabilitation department, is often necessary to sustain employment. Controlled evaluations of the IPS model of supported employment have revealed successful job placements in competitive, community-based work for over 50 percent of the seriously mentally ill persons who participate in the program—more than twice as many as achieved these outcomes through transitional employment.

To minimize stress-induced relapses that can defeat the best-intentioned forms of vocational rehabilitation, practitioners and service systems must ensure that (1) occupational goals are realistically linked to patients' assets and deficits, (2) progress is promoted incrementally with abundant supports and reinforcement, (3) social skills training is made available to help the worker develop social support inside and outside the workplace, and (4) pharmacotherapy and crisis intervention services are kept accessible.

Sex and AIDS Education Sexual issues have been neglected in most work with seriously mentally ill persons. That neglect has historical roots and reflects a tendency of psychiatric practitioners to avoid exploration of sexual problems or to assume that seriously mentally ill persons do not have active sex lives. The sexual behavior of persons with schizophrenia differs from that of the general population, especially male patients, whose predominant sexual activity is autoerotic, with few sexual partners. Their median number of lifetime sexual partners is below normative profiles. However, schizophrenic individuals are at increased risk for HIV infection, perhaps related to their infrequent use of condoms; number of anonymous sexual partners; hypersexuality during psychotic episodes; poor ability to decline sexual overtures; and male homosexual activity, which may be more prevalent than in their counterparts without schizophrenia. Most antipsychotic and antidepressant medications influence sexual drive and activity adversely. Known adverse effects (which undoubtedly influence compliance) include ejaculatory difficulties, erectile dysfunction, decreased libido, orgasmic dysfunction, and priapism.

Education programs on sex and acquired immunodeficiency syndrome (AIDS) have been designed with the following goals: increasing patients' knowledge and comfort about sexuality, helping patients identify and clarify their values and attitudes about sexuality, helping patients acquire decision-making skills regarding sexual activity, overcoming medication-related sexual dysfunction, preventing deterioration of sexual functioning, improving intimacy skills, providing basic AIDS education, and teaching proper condom use. To compensate for patients' attentional and cognitive difficulties, materials were kept graphic and simple. Videotaped presentations were used, as were behavioral techniques such as role playing and structured problem-solving methods. Although careful design of such programs compensates for the cognitive and motivational deficits of chronically mentally ill persons, data are not available to guide future efforts in sex education.

Clinical Case Manager as the Linchpin of the Service Delivery System Successful coordination of the elements and modalities of any service delivery system depends on effective case management. Soon after the onset of the deinstitutionalization movement, many clinicians realized that ensuring continuity of care for seriously mentally ill persons in a system that is complex, fragmented, and frequently inaccessible would require professionals trained to navigate the murky waters of social services. Successful service delivery models with a case-management framework have increased the community tenure of previously institutionalized patients in a variety of settings.

Effective case management can integrate psychiatric treatment and rehabilitation services, incorporating the elements listed in Table 52.4-5. Most individuals with schizophrenia, as part of the nature of their enduring symptomatology, do not readily travel for medical and psychiatric appointments nor do they eagerly embrace the social contact that is offered by therapeutic services. It is often necessary to deliver services through a mobile, outreach form of clinical case management.

The client or patient is the central focus for the case manager, who works collaboratively with the client to formulate long- and short-term goals and create opportunities for treatment, rehabilitation and support
Caseloads for assertive-outreach, intensive, clinical case managers must be limited to less than 20 clients or patients per manager
Case management services should be delivered primarily in the client's community and natural habitat; thus, transportation, mobile communications, and in vivo work are key requisites
Case managers are expected to be accessible, either directly by being on call or indirectly through rotation with other case managers on the team; they often provide crisis intervention, either directly or indirectly
Case managers serve as advocates or brokers developing medical, psychiatric, dental, medication, financial, housing, and social support for clients or patients
Case managers serve as teachers, guides, role models and problem solvers
Case management services are not time limited; thus, long-term nonprofessional supports must be developed by case managers within informal and natural support networks to relieve themselves of cumulative stress and strain from increasing caseload demands

Table 52.4-5 Elements of Clinical Case Managers' Functions and Role

As experience with case management has grown, its prototype has changed from a broker of services to an intensive clinical service provider and advocate. Intensive case managers attempt to wrap needed services for community function and adaptation around each patient, with the type and amount of services matched to an individual's clinical status, values, and phase of illness. The Training in Community Living or the Program of Assertive Community Treatment (PACT) developed in Madison, Wisconsin, is a model of broad-spectrum case management organized in round-the-clock continuous-treatment teams. The program has been replicated throughout the United States and abroad in both rural and urban settings.

The PACT model has been adopted in several areas where funding for public psychiatric services faces greater constraints by state governments that are cutting back on budget support for human services. For example, in Los Angeles County, 7 percent of the mental health budget of $400 million has been funneled into agencies providing PACT-like services to recidivistic mentally disabled persons. These individuals, who are the highest users of hospitalization and costly crisis services, expend 15 percent of the mental health budget and yet make up only 1 percent of the population served in the county (averaging $50,000 per year). By providing the

service agency with a capitated budget (usually between $15,000 and 20,000 per patient a year) the funding agency (in this case the Mental Health Department of Los Angeles County) can maintain fiscal solvency through its ability to assess annual expenses more accurately while allowing the treatment provider maximum flexibility in the use of resources and selection of services such as housing, crisis intervention, vocational rehabilitation, medications, and continuous case management.

Although it is generally agreed that case management is a desired service for individuals with serious, persistent mental illness, there is little consensus about which elements of case management are most clinically useful. A case manager in a psychiatric rehabilitation setting has several specific tasks to perform including assisting patients in building social networks; facilitating procurement of housing and employment; helping patients interact with the various service organizations to meet ongoing needs; teaching patients the skills they require for illness self-management; monitoring the clinical progress of the patients; and, when necessary, undertaking timely clinical interventions. Each of these tasks requires specific competencies, illustrated in Figure 52.4-15. By acting as the fixed point of responsibility within a continuum of care, case managers have influenced patient outcomes toward improved vocational functioning, less social isolation, and more independent living.

FIGURE 52.4-15 Clinical competencies of assertive, outreach case managers.

Validation The current state of research inquiry into case-management services includes several well-designed, randomized, controlled studies with clearly operationalized and faithfully delivered models of case management and careful examination of outcomes such as symptomatic status, relapses, acquisition and generalization of social and independent living skills, vocational outcomes, quality of life, and service utilization. These studies showed that different models of case management may be more effective with system outcomes (e.g., reducing the number and length of hospitalizations and consequently cost-effectiveness) than with clinical outcomes (e.g., symptoms, social functioning, and quality of life). For example, a 2-year study conducted at Department of Veteran Affairs (VA) medical centers found that high users of inpatient services who were randomly assigned to receive intensive psychiatric community care (IPCC) had 33 percent fewer inpatient hospitalization days than their counterparts assigned to standard VA care. The impact of the IPCC intervention was greatest at long-term hospitals, with older patients, and with patients with higher levels of preintervention inpatient service usage.

Despite its proven cost-effectiveness, case management as a clinical intervention cannot be justified solely by decreased service use and increased cost containment. For example, a case-management system that increases service use by homeless, mentally ill persons by moving them from bus stations, shelters, and alleys to hospitals and housing, should be seen as a success, not a failure. Unhappily, current concerns with cost savings have led statewide mental health systems to abandon clinical markers of success in favor of the least-expensive alternative—often leading to transinstitutionalization of mentally disabled persons to prisons and jails.

Ultimately, the success of any case-management approach depends on whether or not patients receive the appropriate services to meet their needs. The problems encountered in attempting to draw conclusions from the current generation of case-management studies include a failure to assess the needs of the individuals receiving the services, inadequate definition of the activities in a given model of case management, and little consideration given to the relation between process and outcome measures (i.e., how a particular intervention leads to a salutary result). Improvements in the yield from empirical evaluations of case management require better specification of case-management methods and objectives, characterization of the symptomatic impairments and psychosocial deficits of the study population, use of multiple outcome measures, and longer studies that follow clients over 3- and 5-year periods.

One such study began in the early 1990s when the California legislature created two pilot integrated service agencies (ISAs), The Village, located in urban Long Beach, and the Stanislaus County ISA, located in a small city within a rural county. The ISAs received a fixed payment for each enrolled patient to cover all mental health services (approximately $15,000 per year). The treatment program incorporated the components of the PACT model, with interdisciplinary teams providing core services and assuming responsibility for each client, offering 24-hour coverage from a single service source for a 3-year period. Individualized treatment was based on a personal services plan developed jointly by patients, family members, and staff.

At each site a team of state, county, and ISA clinicians screened referrals to the programs. The authorizing legislation required that participants in the ISAs have a serious and persistent mental illness, demonstrated by a revised third edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-III-R) diagnosis other than a primary substance use disorder, substantial functional impairment due to the mental disorder, and eligibility for public assistance because of the functional impairment. Eligible patients at each site were randomly assigned to receive ISA services or a comparison group. Patients in the comparison groups continued to receive usual county mental health services, such as outpatient, day treatment, case management, inpatient, and minimal rehabilitative services.

After 12 months of the 36-month program a total 426 patients were assessed on a variety of outcome variables using subjective information from patient interviews and objective data from patients' records. Patients in both ISAs were more likely to remain in treatment, were less likely to use psychiatric hospital care, and were more likely to work for pay. However, much of the employment activity was part time or temporary, and no outcome differences were observed in several other socially and clinically important domains such as rates of long-term hospitalization, arrest, or conviction; symptomatology; substance abuse; homelessness; or quality of life. Data for the full 36 months will be important in either confirming these results or showing the increased impact of mature programs.

In summary, most studies of case management undertaken thus far have been first-generation studies concerned with comparing case management with more traditional models of service delivery. While helpful, this type of research often provides unwieldy comparisons of different models of clinical case management in markedly different care systems, with poorly defined or heterogeneous patient populations. The second-generation of case-management research will compare different models of case management or specific interventions within case-management services to determine which service elements are responsible for the various outcome domains. One hypothesis supported with data from case-management teams that incorporated specific expertise in supported employment and social skills training holds that reaching specified treatment outcomes (e.g., improvements in independent living skills) requires specific therapeutic intervention in that domain (e.g., training in independent living). In the larger perspective, researchers must determine whether adding competency in symptomatic and functional assessment, vocational rehabilitation, social skills training, and family psychoeducation to the repertoire of case managers is translated into superior, long-term improvement in the person's symptomatic and functional status.

Community Support Programs The signing of the Community Mental Health Centers Act by President Kennedy in 1963, and the extension of Social Security disability benefits to mentally ill individuals 10 years later, set the stage for the transition of care of seriously mentally ill patients from large state hospitals to less-restrictive settings. However, by the mid-1970s, the fragmentation and complexity of deinstitutionalization-oriented services and supports were apparent. The prevailing psychotherapeutic paradigms were ill equipped to help individuals with serious and persistent mental illness cope with life in community settings. Most programs were facility or office based; and they expected people to come in to seek services and to negotiate a great deal of red tape. Thus, while the locus of care had shifted, no complementary shift in treatment focus had occurred, nor had new and appropriate modes of care been articulated and implemented.

An overarching conceptual framework was needed to counteract the pervasive fragmentation in the mental health field. The Community Support Program (CSP) was developed as a national model by the NIMH in 1976, with the input of field-based knowledge from practitioners, families, policymakers, and patients. According to this model, interventions were to be delivered in a coordinated fashion by multidisciplinary teams of clinicians who assumed long-term responsibility for the care of the seriously mentally ill patients across the spectrum of mental health services including inpatient units, outpatient programs, partial hospitals, and psychosocial

rehabilitation centers. The community support system had several essential functions or components, listed in Table 52.4-6.

Outreach and engagement
Support of basic needs
Mental health care and treatment
Crisis and emergency services
Comprehensive psychosocial services
Range of housing options
Support and education for provides and families
Development of natural supports
Advocacy and protection
Case management

Table 52.4-6 Elements of a Community Support System

Several principles were to guide the development of local community support programs. People with disabling mental disorders were to be treated with dignity and their privacy and right to confidentiality respected. Services were to be adapted to the changing needs and preferences of each person on the basis of self-determined goals. The service system was to provide comprehensive, accessible services for as long as, and whenever, a person needed them, in settings that were the least restrictive and most normalized.

The impact of the CSP model, though difficult to quantify, has been extensive. Through national working conferences and a focused program of planning and demonstration grants directed to the states, the CSP model has been incorporated into a wide variety of successful community-based programs. One example of the success of the CSP initiative is with young adults diagnosed with both severe mental illness and substance abuse disorder. The results of 13 demonstration projects funded by the CSP branch of the Center for Mental Health Services showed that substance abuse treatments could be integrated with components of outpatient mental health services and that motivational interventions were fundamental. Patients could be engaged in integrated, community-based treatment, with the possibility of concomitant reduced hospital use, improved severity of substance abuse, and enhanced quality of life. These CSP-driven service demonstrations and related program evaluations provided a watershed in the development of dual-disorder services and research.

With the passage of Public Law 99-660, the Americans with Disabilities Act, states were required to incorporate CSP principles into their plans for providing vocational services to persons with serious and persistent mental illness. The act outlaws discrimination against persons with disabilities in nearly every domain of public life: employment, transportation, communication, and recreational activities. With respect to employment, the most important definition of discrimination is an employer's refusal to make a reasonable accommodation. When requested by a qualified applicant or employee with a disability, an employer must provide a reasonable accommodation unless doing so would impose undue hardship. A common example of a reasonable accommodation is building a ramp for an individual confined to a wheelchair. Similarly, individuals with psychiatric disabilities may require accommodations to address their functional limitations. For example, an individual who has difficulty concentrating may ask the employer to provide a private space for work, to limit interruptions and noise. Other possible accommodations for such an individual include maintaining a structured schedule with well-defined daily tasks, eliminating nonessential or secondary tasks that might be distracting, and minimizing supervisor or coworker interruptions of the employee. By invoking the principles of the Americans with Disabilities Act, many communities have been able to secure the mandate and resources needed to design and implement comprehensive, continuous systems of mental health care.

Housing Safe, acceptable housing is a critical element in the complex array of basic and extended services that persons with serious, disabling mental illnesses require to establish and maintain themselves in the community. In fact, rehabilitative efforts are doomed to failure unless affordable, adequate housing is available. Some of the most time-honored needs met in past eras by psychiatric hospitals—offering protection, social support, nutrition, security, supervision, and sanctuary for the mentally disabled—are now provided by the spectrum of community-based housing options. While most psychiatric patients desire such housing, obstacles prevent them from securing and maintaining satisfactory shelter.

Impairments related to the patients' illnesses, such as deficits in social and independent living skills, severe psychotic and affective symptoms, and cognitive dysfunctions; environmental barriers such as stigma, poverty, and discrimination; and service system inadequacies such as inattention to patients' need for support in their vocational and independent living endeavors, have compounded the problem. Consequences of that inattention include homelessness, inappropriate use of intensive psychiatric and general hospital services, family burden, and increased involvement of mentally disordered persons with the legal and penal systems.

Efforts to redress these problems have taken many forms over the past decade. Attention has been given to examining the level of support needed for seriously mentally ill persons to secure and maintain housing. Initially this involved developing comprehensive arrays of transitional living arrangements, so that patients could move from more- to less-supervised, supportive settings. Such elements as quarterway, halfway, and three-quarterway houses; board-and-care and other group living arrangements; crisis and homeless shelters; foster family settings; and supported apartment programs were popularized. The best efforts allowed consumer choice, easy access, and free movement among the service elements depending on patients' needs for supervision and support. However, many efforts suffered from the tendency for clinicians and systems to attend to their own needs, leading to fixed lengths of stay in each of the housing elements, with patients moving to meet systems' needs rather than having services flexibly tailored to their own waxing and waning impairments and personal resources.

Over the past several years, partnership between mental health and public housing agencies have become more common. What has been lacking are approaches that attend to patients' individualized desires for housing. Service providers and policymakers have not implemented systems that actively ask patients what their housing preferences are and then teach them the skills they need to find adequate, acceptable housing. Future efforts will seek to solicit patients' goals for shelter and basic needs, identify their personal resources and deficits in the pursuit of those aims, and help them acquire the coping skills and competencies they require to attain their housing goals.

A shift in conceptual framework has already begun with the increased availability of supported housing. One element of the supported housing movement is a change from reliance on a facility-based residential treatment setting or a series of such specialized settings as the focus for treatment and rehabilitation to the need for a safe, secure home of one's own as a basis for a stable life in the community. In this new paradigm, professionals no longer select the setting or determine what type of placement is best for the patient, nor do they place a person on the basis of open beds or slots in the residential service system. Rather, the person is helped to choose an appropriate living situation on the basis of personal criteria, preferences, resources, and needs. As such, the patient assumes the role of tenant, householder, neighbor, and mainstream community member, working together with staff on mutually agreed on goals and tasks geared toward the individual's success and stability in the home chosen. Additionally, social support, case management, crisis intervention, in-home skills training, and accessible psychiatric consultation are flexibly wrapped around the changing needs of the patient. Financial assistance is available through subsidized rental vouchers (Section 8 grants) provided by the United States Department of Housing and Urban Development.

Two large research efforts have provided some information on the value of the supported housing approach. The Program for Chronic Mental Illness was sponsored by the Robert Wood Johnson Foundation in concert with the United States Department of Housing and Urban Development. The program was targeted toward overcoming fragmentation in the mental health system so that homeless persons with psychiatric disabilities were better housed and served. The nine cities selected developed comprehensive housing plans and designated a single mental health authority with defined, continuous responsibility for those in the target population. This single mental health authority had budgetary control, responsibility for targeting services and supports, and access to rental subsidies.

The demonstration program showed that comprehensive supports and housing subsidies had a major impact on the lives of persons who would otherwise be homeless, including decreased hospitalization and improved quality of life. However, these results were predicated on the patients receiving ongoing, intensive case management and practical help. Normalized housing with intermittent support from visiting case managers did not suffice to buffer the stressors attendant on living independently.

The second demonstration project was a joint venture of the United States Department of Health and Human Services and the Department of Housing and Urban Development. The McKinney Research Demonstration Program for Homeless Mentally Ill Adults consisted of five longitudinal, experimentally designed studies

serving a total of nearly 900 participants. These sites tested a variety of methods of reaching, housing, and serving the target population. In general, the studies demonstrated that homeless persons with serious, persistent mental illness can be engaged, will accept the services that meet their needs, and can successfully become stabilized in community-based housing with appropriate help. Participants displayed significant decline in symptoms of mental disorder and use of psychiatric hospitalization and a slight improvement in social and role functioning. Moreover, individuals' perception of their quality-of-life improved significantly. However, for a large minority of the participants, continuing substance abuse, social isolation, boredom, and overall difficulty adjusting to living alone negatively affected the stability of their lives.

Taken together, these demonstration projects indicate that innovative supportive housing programs do seem to allow many persons with long-term psychiatric disorders to develop a stable home in the community. However, an ambitious policy of independent housing—without consideration of such factors as substance abuse, an individual's social network, poverty, quality of housing, criminal activity in the neighborhood, and the amount and quality of ongoing, personalized support and assistance from community-based mental health teams—is not likely to maintain long-term residential stability.

Combined Psychosocial and Drug Therapies The biopsychosocial approach to comprehensive care including training in social and independent living skills, family psychoeducation, self-management of medication and symptoms, assertive clinical case management, and supported housing and employment can amplify the impact of medication in fostering better outcomes and higher levels of personal functioning. Optimal psychiatric treatment and rehabilitation offered in a coordinated, comprehensive, and continuous fashion can clearly facilitate symptomatic and social recovery from schizophrenia and other disabling mental disorders in many more individuals than are currently helped. Clinicians must teach patients the skills they need to be effective collaborators in the treatment planning process. The following case vignette illustrates this point.

Bill was a 23-year-old man with a 5-year history of schizophrenia. He had been resistant to taking medication due to the adverse effects he had experienced, including severe akathisia, tremors, and muscle stiffness. He acknowledged that these medications had diminished his psychotic symptoms and improved his attention and concentration but resisted efforts to become more compliant with prescribed regimens. He also related that some of his psychotic symptoms, particularly auditory hallucinations, persisted at a low level despite altered dosages and types of medications. Bill had been hospitalized an average of three times a year for psychotic exacerbations since the onset of his illness. He said that all efforts to treat him “missed the boat” and that staff had not attended to his own goals and ambitions.

At this point, staff engaged Bill in a goal-setting process, in which he identified targets of living independently without being hospitalized, taking little or no medication, and eventually getting a job. His psychiatrist and social worker accepted these goals as laudable and set out to establish clear and measurable landmarks to gauge his success in pursuing these aims. They then worked with Bill to identify his personal resources and the obstacles he faced in attaining his goals. Bill said that the most frustrating problem was his lack of understanding about his illness and its treatment, because this led to frequent relapses and concomitant life disruption.

Bill's psychiatrist and social worker next enrolled him in a class designed to increase his understanding of his illness and the medications used to treat it. He gained a good working knowledge of how these medications caused the adverse effects he found so intolerable, and he learned communication skills that would help him effectively negotiate the type and dosage of his medication. Following this, he worked with his psychiatrist and agreed to a trial of an atypical antipsychotic. He noticed that the new medication resulted in much less discomfort from extrapyramidal adverse effects. Finally, he was taught coping methods—such as humming and reducing social stimulation—to manage the persisting auditory hallucinations he experienced. This gave Bill a sense of mastery over his illness, and he adhered to his medication regimen. During the next year, he experienced two minor relapses but sought help from treatment personnel early in the prodromes of these relapses and did not require rehospitalization.

The key features of the foregoing example were that Bill's goals and personal desires were solicited, respected, and incorporated into the treatment plan; his resistance to medication was confronted in a straightforward, nonjudgmental, problem-solving manner; and he was taught skills he needed for effective collaboration in his own treatment and rehabilitation. Treatment occurred in a multidisciplinary team context, and Bill was seen continuously by the same group of treatment providers.

The new generation of antipsychotic medications offers distinct advantages over conventional antipsychotics in facilitating rehabilitation. These agents, which include clozapine, risperidone, olanzapine, quetiapine (Seroquel), and sertindole (Serlect), are superior to traditional antipsychotics in ameliorating negative symptoms such as social withdrawal and anhedonia and are less likely to produce extrapyramidal adverse effects, thus minimizing the use of anticholinergic agents that tend to impair attention and memory capabilities. They may also have direct salutary effects on information processing and attentional capacities in the central nervous system. For all of these reasons, increased use of these newer agents offers hope of greater recovery from the illness if continuous access to treatment and rehabilitation is provided.

Findings to date indicate that comprehensive treatment programs for persons with serious mental disorders should combine drug and behavioral, learning-based interventions. The following principles, distilled from the results of many studies and practice guidelines, summarizes current clinical wisdom regarding drug-psychosocial treatment interactions.

1. Psychosocial treatment is most helpful for patients in reasonably good partial or full remission from florid symptoms who have reached stable levels of maintenance medication. Psychosocial treatment during acute flare-ups of symptoms should be aimed at calming the patient, reducing levels of social and physical stimulation, and helping the patient integrate and understand the symptoms as part of an illness process.
2. The most effective psychosocial treatment—whether provided by individual, group, or family therapy; day hospital; or inpatient milieu therapy—contains elements of practicality, concrete problem solving for everyday challenges, low-key socialization and recreation, engagement of attainable tasks, and specific goal orientation.
3. A continuing positive relationship is central in the overall strategy for treating the patient with schizophrenia, no matter how much drug or psychosocial treatment contributes. That relationship may be with the prescribing psychiatrist or with a paraprofessional case manager.
4. The critical time to offer psychosocial treatment is during the aftercare period, when the patient can absorb rehabilitation and needs assistance in surmounting the problems and stresses of readjusting to family and community.
5. Psychosocial treatment should be long term; benefits rarely become apparent before 12 months and are even greater after 2 years. Indefinite, if not lifelong, psychosocial support, guidance, and training are probably optimal for most patients with chronic schizophrenia. As antipsychotic drugs are most effective in maintaining symptomatic improvements when continued indefinitely, it is not surprising that psychosocial rehabilitation efforts are similarly optimized by continuity.
6. Psychosocial treatment should focus on stressors in the environment and deficits in personal characteristics that seem to play specific roles in relapse and community maladjustment. Schizophrenic relapse is common, even when drug compliance is firmly established. No evidence suggests that a patient's level of manifest psychopathology at hospitalization or discharge predicts subsequent relapse. The best explanation, based on converging lines of evidence from empirical studies, is that the patient's personal assets and deficits, the social environment, and the type of psychosocial therapy are the most powerful influences on relapse, even in the face of reliably administered maintenance medication. Medication is never prescribed, ingested, metabolized, or excreted in a socioenvironmental vacuum.

FUTURE DIRECTIONS

With improved and more reliable methods of diagnosis, symptom assessment, and functional assessment, the population and needs of patients requiring psychiatric rehabilitation are becoming better defined. Because of its growing technology and its conceptual and empirical base, psychiatric rehabilitation has attracted an interdisciplinary group of practitioners and investigators whose work has contributed to its evolving scope and stature as a subspecialty of the mental health and rehabilitation professions. The challenges to community-based care for seriously and chronically mentally ill persons have been met by new initiatives at the clinical and policymaking levels, for example, social skills training, case-management models for coordinating continuing care, supported housing and employment services, greater advocacy and involvement by families, and the Americans with Disabilities Act.

Despite advances in assessment and treatment, implementation of psychiatric rehabilitation in the public sector has been frustrated by inadequate funding and resources, lack of trained personnel, limited dissemination of the new technology, insufficient affordable housing, unattractiveness of voluntary treatment, and a lack of administrative and financial integration of state and local community mental health and substance abuse services. The consequences of those systemwide problems have meant that too few mentally disabled individuals are actually receiving state-of-the-art services, and too many thousands are living on the streets of our

cities and in jails and prisons.

Innovations for the Twenty-First Century How might the future of psychiatric rehabilitation be charted? Will the field continue to flourish or merely become a historical footnote along with deinstitutionalization and community mental health centers? Some ideas for clinicians and researchers emphasizing domains of rehabilitative activity are currently taking root.

Illness Self-Management Techniques Skills training modules, such as those in the UCLA Social and Independent Living Skills series, will enable an increasing array of practitioners help patients acquire the abilities to monitor their psychopathologic and functional progress, to identify stress through changes in skin conductance and prodromal symptoms, and to seek early, flexible levels of intervention from professionals and caregivers. With patients also more knowledgeable about the benefits and adverse effects of their medication, a true partnership in managing pharmacotherapy will yield improved adherence and self-efficacy and fewer relapses, and more individuals will achieve symptomatic and functional recovery.

Cognitive Remediation As researchers learn more about the cognitive substrates and mediators of psychopathology and social dysfunction, techniques will be developed from applied behavior analysis and amplified by a new generation of atypical antipsychotic drugs that can effectively remediate and overcome the attentional, memory, and information-processing deficits. The brain is a plastic organ with many compensatory mechanisms that can be activated by structured, systematic environmental intervention; thus, the biopsychosocial nature of serious mental disorders will illuminate the bidirectional pathways among brain-behavior-environment interactions.

Assessment Technology Functional and symptom assessment will become better integrated, with mathematical models derived from large-scale research studies showing how personal, social, and psychopathological variables interrelate. One study of over 400 mentally ill persons showed that long-term employment outcomes were determined by a confluence of psychiatric symptoms, social competence, work capacity, and pension disincentives. When practitioners adopt quantifiable and repeated measurements of symptoms and functioning in their patients, interventions will be selected more judiciously and their effects monitored more precisely. Improving the quality of care ultimately depends on reliable and clinically relevant assessment instruments that give informational feedback and guide decision making.

Integration of Social Learning With Assertive Community Treatment Proponents of intensive forms of outreach case management and clinical intervention will adopt behavioral therapy techniques to promote greater gains in psychosocial functioning and independence of patients. Life adjustment teams will use structured skills-training techniques—which themselves will have been infused with new principles from research on cognitive science—while engaging patients in naturalistic settings in the community. Demonstration studies of this type of integrated mental health care have already accelerated and broadened the symptomatic and social recovery of persons with serious mental disorders.

Normalization of the lives of long-term mentally ill individuals will improve with a wider array of supported employment, housing options, and access to educational, social, and recreational opportunities that promote community adaptation. The anticipated contraction in the work force because of lowered birth rates, continuing growth of the economy, and the impact of the Americans with Disabilities Act will make jobs more accessible to disabled persons. The type and frequency of follow-along psychiatric services will be tailored to the individual's symptom patterns, work tolerance, stress vulnerability, and social disabilities. Volunteers, students, and retired persons will take on a greater role in providing social support, friendship, and community linkages for mentally ill individuals. Telephones, answering machines, beepers, and informatics made possible by fiber optics, television, and computers will reduce the isolation and stigma of mental illness.

Family and Patient Participation With the growing size and political impact of the National Alliance for the Mentally Ill, the National Alliance for Research on Schizophrenia and Depression, the National Depression and Manic-Depression Association, and the consumer self-help movement, patients and family members will become more effective advocates for needed services and partners in treatment planning and intervention.

Financial Integration of Mental Services New initiatives will flow from the various states and from demonstration projects sponsored by the federal government to improve the cost-effectiveness of mental health services. With pressures to reform the nation's health delivery systems and the advocacy of the National Alliance for the Mentally Ill, efficient and efficacious psychiatric services will be reimbursed at levels commensurate with other medical and surgical services. Incentives will be offered to encourage community-based continuity of care through such mechanisms as capitation, health maintenance organizations, regional consortia of service providers, performance contracts, clinical information systems, and patient registers.

SUGGESTED CROSS-REFERENCES

Public psychiatry is the focus of Section 52.1, and the role of the psychiatric hospital in the treatment of mental illness is the subject of Section 52.3. The individual psychotherapy of schizophrenia is discussed in Section 12.10, and the psychosocial treatment of schizophrenia is discussed in Section 12.9.

SECTION REFERENCES

Anthony WA, Blanch A: Research on community support services: What have we learned? Psychosoc Rehabil J *12:* 55, 1989.

Benton MK, Schroeder HE: Social skills training with schizophrenics: A meta-analytic evaluation. J Consult Clin Psychol *58:* 741, 1990.

*Bond GR, Drake RE, Mueser KT, Becker DR: An update on supported employment for people with severe mental illness. Psychiatr Serv *48:* 335, 1997.

Brenner HD, Roder V, Hodel B, Kienzle N, Reed D, Liberman RP: *Integrated Psychological Therapy for Schizophrenic Patients*. Hogrefe & Huber, Seattle, 1994.

Corrigan PW: Use of a token economy with seriously mentally ill patients: Criticisms and misconceptions. Psychiatr Serv *46:* 1258, 1995.

Corrigan PW, Yudofsky SC, editors: *Cognitive Rehabilitation for Neuropsychiatric Disorders*. American Psychiatric Press, Washington, DC, 1996.

*DeSisto M, Harding CM, McCormick RV, Ashikaga T, Brooks GW: The Maine and Vermont three-decade studies of serious mental illness: Longitudinal course comparisons. In *Historical and Geographical Influences on Psychopathology*, P Cohen, C Slomkowski, editors. Lawrence Erlbaum, Mahwah, NJ, 1999.

Dilk MN, Bond GR: Meta-analytic evaluation of skills training research for individuals with severe mental illness. J Consult Clin Psychol *64:* 1337, 1996.

Drake RE, Mercer-McFadden C, Mueser KT, McHugo GJ, et al: Review of integrated mental health and substance abuse treatment for patients with dual disorders. Schizophr Bull *24:* 589, 1998.

Falloon IR, Coverdale JH, Laidlaw TM, Merry S, Kydd RR, Morosini P: Early intervention for schizophrenic disorders. Implementing optimal treatment strategies in routine clinical services. OTP Collaborative Group. Br J Psychiatry *172* (Suppl): 33, 1998.

Falloon IRH, Boyd JL, McGill G: *Family Care of Schizophrenia*. Guilford, New York, 1984.

Franco H, Galanter M, Castaneda R, Patterson J: Combining behavioral and self-help approaches in the inpatient management of dually diagnosed patients. J Subst Abuse Treat *12:* 227, 1995.

Frese FJ. Advocacy, recovery, and the challenges of consumerism for schizophrenia. Psychiatr Clin North Am *21:* 233, 1998.

Hogarty GE, Kornblith SJ, Greenwald D, DiBarry AL, Cooley S, Flesher S, Reiss D, Carter M, Ulrich R: Personal therapy: A disorder-relevant psychotherapy for schizophrenia. Schizophr Bull *21:* 379, 1995.

Lamb HR, Bachrach LL, Kass FI, editors: *Treating the Homeless Mentally Ill*. American Psychiatric Press, Washington, DC, 1992.

Liberman RP, editor: *Psychiatric Rehabilitation of Chronic Mental Patients*. American Psychiatric Press, Washington, DC, 1988.

*Liberman RP, editor: *Handbook of Psychiatric Rehabilitation*. Macmillan, New York, 1992.

Liberman RP, editor: *Effective Psychiatric Rehabilitation*. New Directions for Mental Health Services, no. 53. Jossey-Bass, San Francisco, 1992.

Liberman RP, DeRisi WJ, Mueser KT: *Social Skills Training for Psychiatric Patients*. Pergamon, Elmsford, NY, 1989.

Liberman RP, Green MF: Whither cognitive-behavior therapy for schizophrenia. Schizophr Bull *18:* 27, 1992.

McFarlane WR, Lukens E, Link B, Dushay R, Deakins SA, Newmark M, Dunne EJ, Horen B, Toran J: Multiple-family groups and psychoeducation in the treatment of schizophrenia. Arch Gen Psychiatry *52:* 79, 1995.

*Mueser KT, Bond GR, Drake RE, Resnick SG: Models of community care for severe mental illness: A review of research on case management. Schizophr Bull *24:* 37, 1998.

*Mueser KT, Tarrier N, editors: *Handbook of Social Functioning in Schizophrenia*. Allyn & Bacon, Boston, 1998.

Paul GL, Lentz R: *Psychosocial Treatment of Chronic Mental Patients*. Harvard University Press, Cambridge, MA, 1977.

Penn DL, Mueser KT: Research update on the psychosocial treatment of schizophrenia. Am J Psychiatry *153:* 607, 1996.

*Smith TE, Liberman RP, Bellack AS: Social skills training for schizophrenia: Review and future directions. Clin Psychol Rev *16:* 599, 1996.

Solomon P, Draine J, Mannion E, Meisel M: Impact of brief family psychoeducation on self-efficacy. Schizophr Bull *22:* 41, 1996.

Stein LI, Test MA, editors: *The Training in Community Living Model: A Decade of Experience*. New Directions for Mental Health Services, no. 26. Jossey-Bass, San Francisco, 1985.

Wyatt RJ, Henter ID: The effects of early and sustained intervention on the long-term morbidity of schizophrenia. J Psychiatr Res *32:* 169, 1998.

Wykes T, Tarrier N, editors: *Outcome and Innovation in Psychological Treatment of Schizophrenia*. Wiley, Chichester, England, 1998.

CHAPTER 53. PSYCHIATRIC EDUCATION

53.1 GRADUATE PSYCHIATRIC EDUCATION

STEPHEN C. SCHEIBER, M.D.

The goal of the American Board of Psychiatry and Neurology (ABPN) is to enhance the quality of psychiatric and neurological care through the voluntary certification process and thereby serve the public interest. To accomplish this goal, the ABPN works closely with the Residency Review Committees for both Psychiatry and Neurology, the two bodies that accredit programs for general residencies and subspecialties. The ABPN also works closely with the chairmen of departments of psychiatry and neurology as well as residency and postresidency training directors of both disciplines. By issuing time-limited certificates rather than lifetime certificates, the ABPN is forging closer relationships with the appropriate specialty and subspecialty societies to encourage these organizations to develop continuing education programs for their constituents. The advances in the biological and social sciences, the advent of new technologies, and the dramatic economic changes in the health care delivery system, particularly as they affect the patient care, contribute to the complexities of maintaining standards of excellence and confidence in the practice of medicine. The history of the specialty boards parallels the expansion of new knowledge and technology.

HISTORY

The need to establish standards of competence in the field of psychiatry was first publicized by Adolf Meyer during his presidential address at the 1928 annual meeting of the American Psychiatric Association (APA). He urged establishing a system of standardization of education similar to the British system. The APA, through its Committee on Medical Services, took a leadership role in 1929 by urging that the field establish formal requirements for the practice of psychiatry. The committee supported the initiation of a plan to sanction competent psychiatrists. Edward Strecker presented the committee report at the 1929 APA annual meeting. The Committee on Graduate Education was formed and given the responsibility of investigating such a plan. Meyer chaired the committee, and Edward Strecker and George Kirby were appointed as members. Despite numerous reports and a name change to the Committee on Psychiatry and Graduate Education, no action was taken.

In 1931, the American Medical Association (AMA), through its Section on Nervous and Mental Diseases, also called attention to the need for a certification process for psychiatrists and suggested involving the National Board of Medical Examiners in establishing such a process.

At the 1933 annual meeting of the APA, the National Committee on Mental Hygiene held a national conference at which the 28 participants urged the formation of an examining board of the APA. The Council of the APA adopted this suggestion and appointed a board of examiners. The appointees were Meyer, William White, C. Macfie Campbell, and Franklin Ebaugh. They were charged with preparing a plan for certification of psychiatrists.

In 1933, the AMA also adopted a plan for the certification of neuropsychiatrists. The AMA appointed Walter Freeman as chairman and Lloyd Ziegler, Edwin Zabriskie, J. Allen Jackson, and George W. Hall as members of its committee.

To avoid duplication of effort, a third group, the American Neurological Association (ANA), was invited to join the AMA and the APA to participate forming an examining body. The ANA Committee consisted of J. Ramsey Hunt, I. S. Wechsler, and H. A. Riley.

At a 1933 meeting the following was established: (1) each of the two specialties would choose four members, and the AMA would choose two from each specialty; (2) each discipline would establish separate qualifications; (3) a separate examination would be administered for each discipline; and (4) to achieve certification in both disciplines, a candidate would need to pass both examinations.

In 1934, the following training requirements for psychiatry were established: (1) the applicant needed to graduate from an approved medical school and possess a license to practice medicine; (2) basic training in psychiatry would include 3 years of study of psychiatry after the internship, with at least 18 months in a clinical setting in an approved hospital, clinic, or laboratory; and (3) 2 years of practice in a hospital or 1 year largely limited to psychiatry would be required.

In October, 1934, the first organizational meeting of the board was held with Meyer as chairman. He, along with Clarence O. Cheney, Campbell, and Ebaugh were APA appointees; Louis Casamajor, Lewis J. Pollock, H. Douglas Singer, and Zabriskie were ANA appointees; and Hall, Jackson, Freeman, and Ziegler were AMA appointees. They became the first directors of the board. The board was incorporated under a Delaware charter and bylaws were adopted. Singer was elected president, Campbell was elected vice president, and Freeman was elected secretary-treasurer. Standing committees were established, and directors were assigned to these committees.

At the December, 1934 meeting, issues of fees, classes of applicants, a “grandfather clause,” and the name of the Board were decided. The name engendered a great deal of controversy over whether neurology or psychiatry should appear first. Psychiatry, with a larger constituency, prevailed over the alphabet.

The first certifying meeting was held in February 1935, and on June 7, 1935, the board issued its first certificates: 22 in psychiatry and neurology and 12 solely in psychiatry. The first 11 went to the founding directors though Pollock refused to apply for or accept a diploma.

In 1974, the number of directors increased from 12 to 16 with the addition of two representatives from each discipline. Sixteen directors currently serve the board (Table 53.1-1). The American Academy of Neurology (AAN) became the fourth sponsoring society. In 1981, the board voted to no longer recognize the AMA as a sponsoring society for neurology directors. Currently the APA provides nominations for five psychiatry directors, the AMA for the remaining three. The ANA and the AAN each nominate four neurology directors. The directors of the board review the nominees of the sponsoring societies and conduct their own election of new directors. The usual period of service is a 4-year term, and a director cannot serve more than two terms. The autonomous functioning of the Board was stated by Henry W. Brosin in 1962 when he wrote:

Darryl C. De Vivo, M.D., President
Sheldon I. Miller, M.D., Vice President
John E. Schowalter, M.D., Secretary
Rosalie A. Burns, M.D., Treasurer
José Biller, M.D., Member-At-Large
Harold P. Adams, Jr., M.D.
Glenn C. Davis, M.D.
Michael H. Ebert, M.D.
Larry R. Faulkner, M.D.
H. Royden Jones, M.D.
Edgar J. Kenton III, M.D.
Alan K. Percy, M.D.
Pedro Ruiz, M.D.
James H. Scully, Jr., M.D.
Nicholas A. Vick, M.D.
Elizabeth B. Weller, M.D.

Table 53.1-1 ABPN Directors—1999

Although the ABPN is composed of representatives from three (now four) constituent societies to whom they report and are responsible, it is a working principle that the Directors vote solely in the best interest of the Board. In the final analysis, they are responsible only to themselves.

American Board of Medical Specialties The ABPN is one of 24 specialty boards. The other boards are Allergy and Immunology, Anesthesiology, Colon and Rectal Surgery, Dermatology, Emergency Medicine, Family Practice, Internal Medicine, Medical Genetics, Neurological Surgery, Nuclear Medicine, Obstetrics and Gynecology, Ophthalmology, Orthopaedic Surgery, Otolaryngology, Pathology, Pediatrics, Physical Medicine and Rehabilitation, Plastic Surgery, Preventive Medicine, Radiology, Surgery, Thoracic Surgery, and Urology. In 1970, the American Board of Medical Specialties (ABMS) was founded as the umbrella organization for all of the boards. It is responsible for reviewing and approving applications for any new boards and for any new certificates of specialization and of special qualifications and added qualifications for subspecialty certification. It also maintains standards of organization and operation for all specialty boards. The executive vice president of the ABMS is J. Lee Dockery, who replaced Donald G. Langsley, in 1991. Dockery has announced his retirement effective December 1997. Langsley, a psychiatrist who had served as a director of the ABPN and as president of the APA, was the executive officer of the ABMS for 10 years. M. J. Martin, a former vice president and director of the ABPN, was elected president-elect of the ABMS in 1992, assumed the presidency in 1994, and completed his term in 1996. Barbara Schneidman, a psychiatrist from Seattle, Washington, was appointed first deputy executive vice president of the ABMS in 1993. She had previously served as president of the Federation of State Licensing Bodies. Figure 53.1-1 shows the 1998 directors.

FIGURE 53.1-1 1998 Directors for the ABPN. Back row from left: Glenn C. Davis, M.D., Pedro Ruiz, M.D., Nicholas A. Vick, M.D., Michael H. Ebert, M.D., Alan K. Percy, M.D., Darryl C. De Vivo, M.D., Sheldon I. Miller, M.D., Robert Harmon, M.D., John E. Schowalter, M.D. Front row from left: Jos Biller, M.D., Edgar J. Kenton III, M.D., Peter M. Silberfarb, M.D., Elizabeth B. Weller, M.D., Stephen C. Scheiber, M.D., Rosalie A. Burns, M.D., William T. McKinney, M.D., Burton A. Sandok, M.D.

The Advisory Board for Medical Specialties, founded in 1933, was the predecessor of the ABMS. Ophthalmology, Otolaryngology, Obstetrics and Gynecology, and Dermatology and Syphilology were the four existing specialty boards in 1933. Along with these four boards, the American Hospital Association (AHA), the Association of American Medical Colleges (AAMC), the Federation of State Medical Boards, and the National Board of Medical Examiners were the founding sponsoring organizations of the advisory board. ABMS is a nonprofit corporation that is supported by dues from each of the 24 boards. Dues and the number of voting representatives allotted to each board are determined by the number of general certificates issued by a board. Through 1992 biannually the ABMS published the *ABMS Compendium of Certified Medical Specialists,* the only official biographical directory authorized by all 24 medical specialty boards. In 1993 the *ABMS Compendium* merged with the *Marquis Who's Who Directory in Medical Specialties* and became the *Official ABMS Directory of Board Certified Medical Specialists*, now published by Marquis Who's Who. The directory is now available on a CD Rom. Information regarding certification in a specialty, subspecialty certification, or recertification is provided by the certifying boards. Other information, such as medical education and specialty training, is provided by the diplomate. Information about each of the 24 boards is included in the directory. The ABMS also helps individual boards by providing assistance in conducting research projects.

The ABMS assisted with a project of getting the Yellow Pages telephone directories to list the diplomates of the 24 ABMS Boards as board-certified. Diplomates wishing to be listed pay a fee for the listing. The ABMS also initiated a telephone answering service to answer inquiries from callers about whether a physician is certified by one of the ABMS Boards. In addition to the ABMS, in 1972 the Council of Medical Specialties (CMSS), the AAMC, the AMA and the AHA established the Liaison Committee on Graduate Medical Education (LCGME) and the Coordinating Council on Medical Education. The LCGME was given ultimate responsibility for approving all graduate medical education programs. The ABMS was also one of the founders of the Liaison Committee on Continuing Medical Education (LCCME), which accredits programs in continuing medical education. This subsequently became the Accreditation Council on Continuing Medical Education. In 1997 Peter Tanguay chaired the committee. Tanguay served as vice president of ABPN in 1997 in his eighth and final year as a psychiatry director.

American Medical Association The AMA historically provided leadership in accrediting American medical education, beginning with voluntary accreditation of medical schools and internships by 1914. To prepare for accreditation of residency programs, in 1920 the AMA's Council on Medical Education formed committees to establish procedures.

Residency review committees for each of the specialties are responsible for accreditation of residency programs. Until 1975, each committee had final authority for accreditation. Beginning in 1975, each of the review committees reported to the LCGME which, in turn, had final authority. In addition to the five founding organizations noted above, the LCGME has representatives from the public and from the federal government. In 1981, the Accreditation Council for Graduate Medical Education (ACGME) replaced the LCGME as the final authority.

Residency Review Committees for Psychiatry and Neurology Until 1983, the Residency Review Committee (RRC) for Psychiatry and Neurology was a single committee for both disciplines. Since then, each of the disciplines has had its own committee. The Psychiatry Residency Review Committee has 15 voting members: five from the ABPN, five representing the APA, and five representing the AMA, recommended by the Council on Medical Education. The executive vice president of the ABPN and the deputy medical director for education of the APA serve in an ex officio capacity. The APA also recommends one nonvoting resident member of the Psychiatry RRC. Each of the three sponsoring organizations selects three general psychiatrists and two child and adolescent psychiatrists to serve. Daniel Winstead chairs the RRC. Sheldon Miller, psychiatry director of the ABPN, chaired the Psychiatry RRC from January 1, 1997 to December 31, 1998. James Shore, president of the ABPN in 1994, was chairman of the Psychiatry RRC in 1992 and was also elected chairman of the Council of Chairmen of all the RRCs in 1992. He was also appointed chairman of the Council on Graduate Medical Education and Career Development of the APA from 1992 to 1997. James Sculley, Jr., a board director, now chairs this council.

The Neurology Residency Review Committee has six members: two from the ABPN, two from the AMA, and two from the American Academy of Neurology (AAN). The AAN was approved as a third parent of the Neurology RRC in 1992. The usual composition of the Neurology RRC is four adult neurologists and two child neurologists.

The Psychiatry RRC is responsible for reviewing training programs in general psychiatry, child and adolescent psychiatry, geriatric psychiatry, addiction psychiatry, and forensic psychiatry. The Neurology RRC reviews programs for neurology, child neurology, and clinical neurophysiology. Procedures were established for separate review of child neurology programs, and actions were taken on each child neurology application since October 1988. Each program must meet the requirements listed in the annual *Directory of Graduate Medical Education Programs*. The Psychiatry RRC established requirements for programs in geriatric psychiatry, which were approved by the ACGME in 1994 after being rejected in the spring of 1993. These postresidency programs must last at least 1 year and must begin at the fifth postgraduate (PGY-5) year of training or higher. The Psychiatry RRC was advised in the fall of 1991 of the need to establish training requirements for addiction psychiatry. These were approved by the ACGME in 1995. The Neurology RRC prepared requirements for clinical neurophysiology that were initially turned down by the ACGME in 1993 and were revised and approved in 1994. In 1996 the requirements submitted by the Psychiatry RRC for forensic psychiatry were approved.

To be listed in the directory, a program must be reviewed and approved by the RRCs. These committees are responsible for establishing the necessary requirements for programs and submitting them for ultimate approval by the ACGME.

PROCEDURES OF THE ABPN

Executive Vice President The ABPN has had five executive directors since its inception in 1934. The name of the office has been changed from secretary of the board, to executive secretary-treasurer, to executive director, to executive secretary, to its current designation, executive vice president. The first four—Freeman (1934–1946), Francis Braceland (1947–1951), David A. Boyd, Jr. (1951–1971), and Lester H. Rudy (1972–1985)—had served as board directors before serving in their part-time administrative capacity. In 1985, the board selected Stephen C. Scheiber, director of psychiatric residency training and professor of psychiatry at the University of Arizona, to serve as its first full-time executive secretary. He assumed his duties on January 1, 1986. His title was changed in 1989 to executive vice president.

The executive vice president is responsible for the supervision and administration of the affairs of the board. He must establish an effective staff organization to carry out the responsibilities of the board. He coordinates all approved projects, programs, and major activities of the office staff. He ensures that appropriate administrative and adequate staff support exists for all activities of the board and its components, including scheduling business, policy, and committee meetings and preparing and distributing agendas and all background materials before such meetings. He supervises and keeps board records and keeps and distributes minutes of the board meetings and other official meetings under the direction of the secretary. He is responsible for responding to inquiries regarding credentials of individual candidates, for implementing examination procedures by setting up and arranging for sites for the written and oral examinations, for informing candidates of their performance in accordance with the policies of the board, for issuing certificates to successful candidates, and for submitting names of certified neurologists and psychiatrists to the appropriate bodies for publication. He maintains an administrative relationship with the American Board of Medical Specialties, the RRCs, and other relevant organizations. He explicates the policies of the board, articulates the board's position on a variety of issues, and oversees financial and legal matters related to the board.

Credentials Over the years the board has reviewed and revised its training requirements for candidates seeking to sit for its certification examinations. Factors considered about credentials include training program requirements, basic science instruction, knowledge of the specialty, clinical education and training, and cultural differences. In 1976 the board no longer required that candidates complete 2 years of experience in their specialty before admission to the examination. Since 1993 candidates can submit applications during the last 6 months of their final year of training if all training requirements were fulfilled no later than June 30th of the year in which the application was submitted.

To sit for the examination the candidate must submit an application and application fee by February 1st for the fall examinations (an additional late fee is required for those who submit applications after February 1st but before March 1st) along with documentation of satisfactory completion of the board's specialized training requirements and a copy of a current, unlimited license to practice medicine in a state, commonwealth, or territory of the United States or a province of Canada. Candidates who cannot submit a copy of the license before sitting for the written examination must submit one before sitting for the oral examination. Otherwise they are not admitted to the oral examination. Before fall 1994 all part I written examinations were held in the spring; they were changed to the fall in response to a request from the field to move the examination closer to the completion of residency training.

An applicant who seeks admission to examination for certification in psychiatry must have satisfactorily completed an approved PGY-1 and 3 full years of postgraduate, specialized training in a program accredited by the ACGME or The Royal College of Canada. This training must be in the specialty in which the applicant seeks certification. No residency training credit can be given for time spent in a residency program in another specialty. Training programs approved by the RRC and accredited by the ACGME are listed in the *Directory of Graduate Medical Education Programs,* published by the AMA.

Training may be completed on a part-time basis, provided that it is no less than half-time, and credit is not given for less than 1-year blocks of training except when approved by the ABPN Credentials Committee for Psychiatry. To ensure continuity of training, the board requires that 2 of the 3 years of residency following the PGY-1 be spent in a single program.

The board has given much thought to the minimum requirements for PGY-1. Currently it must include a minimum of 4 months of primary care experience and 2 months of neurology. The primary care can be in internal medicine, family practice, pediatrics, or combinations of two or three of these specialties. Emergency medicine rotations must emphasize medical rather than surgical experiences and cannot exceed 1 month of the required primary care training.

Four-Year Training Programs A 4-year program in psychiatry is acceptable if no fewer than 4 months during the first year were spent in an approved program providing supervised, direct responsibility for the general medical care of children, adults, or both. The candidate who spends 12 or more months in such an approved program may do so at any time in the 4 years. Since July 1, 1988, all osteopathic physicians entering a PGY-2 in psychiatry must have successfully completed an ACGME-approved first postgraduate year as described above.

The RRC for Psychiatry, in its special requirements for residency training, has tried to both maintain flexibility and originality in programs and yet set minimum standards. Thus the committee requires that training programs give residents responsibility for the diagnosis and treatment of an appropriate number and variety of psychiatric inpatients for not less than 9 months but no more than 18 months (or its full-time equivalent if done on a part-time basis) in 4 years of training. The 4-year programs must also provide for at least 1 year (or its full-time equivalent if done on a part-time basis) in an organized, well-supervised outpatient program that includes experience with (1) a wide variety of disorders and patient and treatment modalities and (2) brief and long-term care of patients, using both psychodynamic and biological approaches to outpatient treatment. Special requirements include a minimum of 2 months in child and adolescent psychiatry, 2 months in consultation psychiatry, and 2 months in the diagnosis and treatment of neurological patients as well as experience in emergency services, geriatric psychiatry, addiction psychiatry, forensic psychiatry, and others.

Combined Training Programs With Psychiatry Applicants seeking certification in both psychiatry and neurology who began postgraduate training after January 1, 1987, must complete a PGY-1 year that meets the requirements for entry into a neurology program. Additionally, applicants must complete (1) 6 full years of postgraduate training in approved programs in the United States, including 3 full years in psychiatry and 3 full years in neurology, or (2) at the discretion and approval of both training directors and in accordance with RRC requirements, the applicant may complete 5 full years of postgraduate training in approved programs in the United States, including 2 full years in psychiatry, 2 full years in neurology, and a fifth full year of training designed to be eligible for simultaneous credit in psychiatry and neurology. The ABPN has established guidelines for programs for combined training in psychiatry and neurology. Proposals are reviewed by the Credentials Committee for Psychiatry and the Credentials Committee for Neurology. Programs that are approved by both committees are submitted for publication in the AMA's *Directory of Graduate Medical Education*. To date 8 have been approved.

The ABPN and the American Board of Internal Medicine have agreed to a combined training program for Internal Medicine and Psychiatry in 5 years instead of the current 6 years as well as a combined training program for Internal Medicine and Neurology in 5 years instead of the current 6 years. The guidelines for these programs were distributed to interested program directors in the respective disciplines who then must propose a combined program to each of the two boards. Once approved by the two boards, the programs are listed in the *Directory of Graduate Medical Education Programs* of the AMA as combined programs. Twenty-eight such programs have been approved.

The ABPN and the American Board of Family Practice have agreed to a combined training program for family practice and psychiatry in 5 years instead of the current 6 years. Guidelines for these programs were distributed to interested program directors in the respective disciplines who must then propose a combined training program to each of the two boards. Once approved by the two boards, the programs are listed in the *Directory of Graduate Medical Education* of the AMA as combined programs. Twelve such programs have been approved.

Similar combined training programs are available for Neurology and Physical Medicine and Rehabilitation in 6 years instead of the current 7 years, and in Neurology, Radiology, and Neuroradiology in 7 years instead of the current 8.

Internship One of the board's most controversial decisions in the last 30 years was the elimination of the internship requirement in 1970. This move was a product of a 1967 recommendation by the Millis Commission (a citizen's commission on graduate medical education of the AMA) to abolish the internship. The policy of eliminating the internship in psychiatry was adopted in October 1969 for implementation in July 1970. Part of the furor that followed resulted from the board's failure to consult adequately with constituent organizations and educational organizations and to allow sufficient time for the field to debate the decision before its implementation.

In 1974 the board reversed its decision and enacted a new policy to become effective in July 1976, namely that the internship year would be reinstated. Again, the

board faced opposition to its decision. This time the APA requested delaying implementation until 1977 to give programs sufficient time to arrange for the necessary medical experiences and to arrange for appropriate financing of the PGY-1 year. The board consulted sponsoring societies and other relevant professional organizations and delayed reinstatement of the internship until 1977, although programs were advised to implement this reinstatement of the internship as soon as possible.

The ABPN works closely with a number of constituent organizations to prevent a recurrence of the furor that surrounded the internship decision. The ABPN works closely with the APA through its Council on Medical Education and Career Development and through regular meetings of the leadership of the two organizations. The executive vice president served as a consultant through 1993 and in 1994 was appointed a corresponding member of the Graduate Medical Education Committee, and the directors of the board serve on many APA councils and committees. In addition, the board worked closely with the APA Task Force on Recertification, the Working Group on Recertification of the APA, and the Commission on Subspecialization of the APA.

The board meets regularly with the American Association of Directors of Psychiatric Residency Training at its annual meetings, presents updated information to the residency training directors, and participates in workshops at these annual meetings. Similarly, the Residency Review Committee in Psychiatry has representation at this meeting and conducts RRC workshops.

The American College of Psychiatrists, which has primary administrative responsibility for the PRITE (Psychiatric Residency In-Training Examination), sought and obtained board agreement on a project correlating the results of the PRITE with subsequent performance on board examinations. The board cooperated on condition that the confidentiality of information on any individual sitting for the board examinations was not violated. The results of the study revealed a good correlation.

The board also participates in the training directors meeting of the American Academy of Child and Adolescent Psychiatry and has participated in meetings of national organizations of its subspecialties such as the American Association for Geriatric Psychiatry, the American Academy of Addiction Psychiatry, and the American Academy of Psychiatry and the Law.

The ABPN also works closely with the residency training directors. Approximately every 5 years, the ABPN circulates the outline of the part I examination to residency training directors for feedback to help update this important outline. The board sponsored a series of meetings of a task force for written examinations in psychiatry with representation from the APA, in particular relating to the Psychiatric Knowledge and Self-Assessment Examination, the PRITE examination of the American College of Psychiatrists, and the National Board of Medical Examiners for its part I, II, and III examinations for medical students. Representatives from the various committees of the different subspecialties in psychiatry and the neurology written committee also participated. One of the major accomplishments of the task force was the establishment of a common outline regarding topics in psychiatry, which was particularly useful for the PRITE and the part I examination. One of the aims of the task force was to determine whether a common outline regarding topics in psychiatry could be used by all of the related organizations without breaching confidentiality.

Examination Policies and Procedures Surpassing its humble beginnings in June 1935, when only 31 candidates sat for the first examination at the Philadelphia General Hospital, the board in 1998 examined 2330 psychiatrists, 860 neurologists, and 135 child neurologists in November for its part I written examination, and 1830 psychiatrists, 556 neurologists, and 83 child neurologists at three part II oral examinations (Table 53.1-2 and Table 53.1-3). As of January 1, 1999, the board had issued 35,764 certificates in psychiatry, 1,159 in neuropsychiatry, 8,928 in neurology, and 1,151 in child neurology. The board added a fourth examination to the part II examination cycle in 1992, 1993, 1994, and 1995 to accommodate the increasing number of candidates.

Candidates (N)	Year
850	1969
779	1972
1784	1975
1696	1978
1228	1981
1277	1984
1670	1987
2106	1990
2422	1991
2974	1992
2982	1993
2181	Spring 1994
742	Fall 1994
2424	1995
2462	1996

Table 53.1-2 Candidates Taking Part I Psychiatry Examinations

Year	Examined (N)	Passed (N)	Passed (%)
1987	1670	921	55
1988	1787	1079	60
1989	1946	1203	62
1990	2108	1244	59
1991	2422	1406	58
1992	2974	1788	60
1993	2982	1681	56
1994 (spring)	2181	1225	56
1994 (fall)	742	604	81
1995	2424	1341	55
1996	2462	1321	54
1997	2432	1255	52
1998	2331	1248	54

Table 53.1-3 Pass Rates—Part I Psychiatry (1987–1998)

Some 67 examination centers in the United States, Puerto Rico, Canada, and Germany were used to administer the 1998 written examinations, including the part I examination and the added qualification examinations. Despite the large numbers, every effort is made to schedule candidates for the part I examination at their chosen location.

In spite of differences between neurology and psychiatry, the board has maintained the strong alliance of the two specialties over the years in the face of periodic and cyclical reviews by the directors regarding splitting the board into two separate boards. The last such serious discussion occurred in 1982. In the first 10 years of the board's history, each candidate was given the same oral examination and performance was evaluated on the basis of the certificate the candidate was seeking. In 1946, the board introduced separate examinations for each of the specialties, and this remains the current policy.

Written Examinations With the increasing numbers of applicants in the 1960s, the board voted to administer a written examination to screen candidates for the oral examination. In 1964 the board contracted with the National Board of Medical Examiners to develop the written examination. Over a 5-year period the written examination was used to test basic knowledge in the two disciplines. The board conducted a study to test whether the written examination predicted how well candidates would perform on the oral examination. Results of the 1965 and 1966 written examinations, convinced the board in 1967 of the validity of the written examination as a screening test. The board also determined that simultaneous administration of the written and oral examinations was too time consuming and costly. Hence, in May 1967, a separate written examination was introduced as part I of a two-step process. The successful completion of part I, which was administered once a year, then led to the scheduling of candidates for the oral part II examination. Unsuccessful candidates for the part I examination can reapply for the written examination annually, with no limit on the number of times a candidate can take the examination. The Part I Written Examination Committee reviews and refines the examination yearly.

In April 1972, a new 4-hour written examination was introduced. All of basic psychiatry and basic neurology was included, which eliminated basic material from the

part II oral examination for the minor specialty. In 1975, the content for psychiatrists and neurologists was separated. For the 1976 examination, the board decided to honor a passing grade on part I for 7 years from the date of passing, provided it was dated after April 1, 1969. Those who passed part I could apply or reapply for the part II examination anytime during the 7-year period. Since March 1987, a passing grade on the part I examination dated March 1987 or thereafter has been valid for 6 years or three attempts to complete the part II examination successfully, whichever comes first. Those who are unsuccessful after three attempts or who fail to pass part II within the 6 years are asked to retake the part I examination.

In 1981, clinical and basic neurology for psychiatry candidates was incorporated into the part I written examination, and neurology was eliminated from the part II oral examination for psychiatry candidates who took the written examination in 1981 or later. A comparable change for neurologists occurred in 1982.

When candidates apply for the part I examination, they receive comprehensive information about the examination, including its length, how the time is divided between the two disciplines, the approximate number of questions devoted to each discipline, and the distribution percentage of specific subject matter on which the candidate will be tested. The Part I Written Examination Psychiatry Committee solicits input from psychiatric educators about the content and uses this input to review the examination content. The board distributes the content in the subspecialty examinations to the subspecialty psychiatric educators to obtain their input as well. Additionally, candidates receive general information about the format of the examination questions and how to maximize their performance during the examination.

Oral Examinations The part II oral examination has sparked more criticism than the written examination. The ABPN is the only one of the 24 boards that continues to use patients as part of the oral examination. All patients used to examine psychiatrists have psychiatric disorders and those for neurologists have neurological disorders. Because of difficulties in recruiting patients and the rising costs of administering the oral examination, the board introduced the audiovisual section of the part II examination in 1976. This format yielded a more standardized examination. Studies evaluating the relation between the patient and audiovisual sections after a 1-year trial period showed a high concordance rate in both specialties: 84 percent for psychiatry and 86 percent for neurology candidates.

Acceptance of the audiovisual section as an alternative to a patient interview led to the current examination procedure for psychiatrists. After successful completion of the part I written examination (which includes basic neurology and psychiatry and clinical neurology for psychiatry candidates), the psychiatry candidates are invited to sit for the two sections of the oral examination: one with a patient and one audiovisual. The audiovisual tape shows a psychiatrist interviewing a patient with a psychiatric disorder; no simulated patients are used. This taped interview is the basis for questions asked by the examiners. The neurologists have since replaced the audiovisual section with a patient vignette section.

Examiners The board devotes considerable time and energy to the selection, orientation, and training of part II examiners. The eight directors select the members of their own examining team from a recommended list of board-certified psychiatrists, including some from the local region where the examination is being administered. Examiners represent a wide spectrum of backgrounds, including academic, institutional, and private practice psychiatrists.

Each examiner receives information regarding the examination and its procedures before the examination. This material includes guidelines on how to evaluate candidates. In 1979, a special 5-hour orientation session was instituted for new examiners. In 1984, the board made this orientation session mandatory for all new examiners. The following year, the board extended this policy and made attendance at the orientation session mandatory for all examiners who had not examined in the previous 4 years. In 1994, this was further modified to include all those who had not examined in the last 2 years. The orientation session focuses on the examination process from the perspective of an examiner. Appropriate and inappropriate examination techniques are demonstrated. New examiners then meet in small groups with senior examiners to role-play examining, and the examiners receive critiques on their examining techniques. The mandatory orientation sessions for new examiners eliminated a plenary session for all examiners with the board of directors. All examiners, old and new, attend a mandated orientation session with a team of examiners under the direction of a psychiatry director who will serve as their team leader during the 2 days of examining.

The board has continued its efforts to standardize the examination. Examination teams now include four to five senior examiners and 20 to 25 primary examiners. Approximately one-half of a director's team is chosen from a national pool, and they examine at all sites during an examination year. The additional examiners are invited from the local region. For each examination, two examiners are scheduled with a single candidate throughout the examination and a senior examiner supervises the experiences for two pairs of primary examiners.

Grading Candidate grading is based on a composite of their performance in both sections of the part II psychiatry examination. If candidates perform marginally in 1 hour, they can redeem themselves in a second hour with an above-average or exceptional performance. Numerical grading was replaced in 1978 by grades of High Pass, Pass, Condition, and Fail. In 1981 the board further refined the grading to Pass, Condition, and Fail. The neurologists have eliminated the Condition grade. All candidate grades for individual sections are then reviewed by the psychiatry directors in a grading session, and they assign a final grade. The passing rates have remained fairly stable over the past decade (Table 53.1-4).

Year	Examined (N)	Passed (N)	Passed (%)
1987	1403	779	56
1988	1434	881	61
1989	1560	930	60
1990	1629	967	59
1991	1881	1132	60
1992	2051	1256	61
1993	2671	1641	61
1994	2396	1467	61
1995	1793	1066	59
1996	1562	940	60
1997	1709	973	57
1998	1830	1058	58

Table 53.1-4 Pass Rates—Part II Psychiatry (1987–1998)

Dissemination of Board Information The board continuously reviews, updates, and changes its policies and procedures, and it communicates changes to its constituency in a number of ways. One method is by an annual report that is circulated to its sponsoring bodies. The report sent to the APA is in turn sent to the *American Journal of Psychiatry*, where it appears under *Official Actions* of the APA with the subtitle “Annual Report of the ABPN, Inc.” The board also participates in the APA Task Force on Communication between the APA and the ABPN, which was established in 1976. The composition of the task force has varied. Originally it consisted of ABPN psychiatry directors, representatives from the APA Council on Medical Education and Career Development, the APA Committee on Graduate Education, and the APA Assembly of District Branches. In 1981, the task force expanded to include representatives from the American Association of Directors of Psychiatric Residency Training (AADPRT) and the American Association of Chairmen of Departments of Psychiatry (AACDP). Subsequently, the size of the Task Force was reduced considerably. The ABPN representative, as recently as 1988, was a former director and president of the board. Beginning in 1987 a full-time psychiatry director was appointed as a member of the task force, and the executive vice president serves in an exofficio capacity. The task force allows the APA to transmit recommendations to the ABPN and the board to inform the APA of current activities and policies. In addition, the leadership of the APA meets regularly with the psychiatry leadership of the ABPN. The executive vice president also meets with the board of trustees and the Assembly of the APA on invitation.

The ABPN has also offered a workshop, “ABPN Update,” at the annual meeting of the APA, which is chaired by the executive vice president with the psychiatry directors as the participants. The most recent workshop was “ABPN Update: Training Requirements to Sit for the ABPN Examination and Certification Process.” The executive vice president and psychiatry board directors have also participated in plenary sessions and other workshops of the APA and the AADPRT. These meetings help minimize misunderstandings and publicize current issues under review and policy revisions by the board.

Child and Adolescent Psychiatry In 1959 the ABPN with the concurrence of the ABMS instituted a certification process for its first subspecialty, Child Psychiatry. The mechanics of the certification process and the establishment of policy is the responsibility of the Committee on Certification in Child Psychiatry, changed in 1987 to the Committee on Certification in Child and Adolescent Psychiatry.

This committee is under the supervision of the board and follows its general policies and guidelines. The committee consists of six board-certified child and adolescent psychiatrists, one board director, and one board-certified pediatrician. Each member serves a 6-year, nonrenewable term. Nominations for the child and

adolescent psychiatrists are solicited from the American Academy of Child and Adolescent Psychiatry (AACAP) and the APA, those for the pediatrician are sought from the American Board of Pediatrics, and the board director is appointed by the board president. The committee elects its members and submits its selections to the board for final approval. The committee is responsible for administering its annual examination. In 1993, 359 candidates sat for the examination. Because of the marked increase in the number of applicants for 1993, a second examination was held in the spring of 1994. Similarly a second examination was conducted in the spring of 1996. As of fall 1998, 4589 candidates have been certified in Child and Adolescent Psychiatry (Table 53.1-5). Since 1987 the examination is conducted on weekends. Approximately 116 ACGME-accredited child and adolescent psychiatry training programs exist in the United States. The passing rate on the child and adolescent psychiatry examination has been fairly constant for the last decade (Table 53.1-6).

Psychiatry Examination Candidates (N)	Year
0	1972*
78	1975
162	1978
225	*1981*
183	1984
194	1987
284	1990
342	1991
435	1992
359	1993
367	Spring 1994
407	Fall 1994
287	1995
154	Spring 1996
239	Fall 1996

* [illegible] examination groups, as 1972.

Table 53.1-5 Candidates Taking Child and Adolescent Psychiatry

Year	Examined (N)	Passed (N)	Passed (%)
1987	194	129	66
1988	179	94	53
1989	177	135	57
1990	284	190	67
1991	342	205	60
1992	435	269	62
1993	359	220	61
1994 (spring)	367	233	64
1994 (fall)	407	246	60
1995	287	184	64
1996 (spring)	154	90	58
1996 (fall)	239	151	63
1997	284	159	56
1998	270	140	52

Table 53.1-6 Pass Rates—Child and Adolescent Psychiatry (1987–1998)

Applications for the examination in child and adolescent psychiatry must be filed by May 1st preceeding the date of examination for which the candidate is applying. The certifying examination in child and adolescent psychiatry has undergone periodic review and revision that have resulted in major changes. The committee contracted with the Center for Educational Development at the University of Illinois College of Medicine to help conduct studies of the examination process. A study in 1976 reviewed the six sections of the oral examination. As a result, the two basic science sections—History and Literature; and Growth and Development—were eliminated from the oral examination and incorporated into a 2-hour written examination.

In 1973 a 3-year study funded by the Ittleson Foundation, the Grant Foundation, and the Maurice Falk Medical Fund was initiated to “assess systematically and broadly the basic competencies required by a child psychiatrist.” The results of this study were published in 1976 in the monograph “Roles and Functions of Child Psychiatrists.”

This study resulted in several major changes in the examination. Candidates who fail only one section of the examination need only repeat the failed section, not the entire examination. Numerical grades were replaced by grades of High Pass, Pass, Marginal Fail, and Fail. Written clinical vignettes were used to assess consultation skills. Candidates were given a one-paragraph description of a situation in which a child and adolescent psychiatrist is called upon to serve as a consultant in a school, hospital, or courtroom setting. They were then asked to describe orally how they would conduct the consultation. The monograph describes criteria for evaluating this section. In 1981, this section (previously the Interprofessional and Community Relations section) was renamed the Child Psychiatry Consultation section. Beginning in 1999 the consultation section was eliminated and questions regarding consultation will be included in the other three parts of the examination.

Beginning with the February 1978 examination, a passing grade on the written examination was valid for a maximum of 7 years. In 1980 the committee voted that a passing grade on all sections of the oral examination would be valid for a maximum of 4 years. As of 1988, passage of the written examination is valid for 5 years and passage of all portions of the oral examination is valid for 5 ears. In 1978, the committee initiated a project to assess the validity of the child psychiatry examination. With the permission of candidates, training directors were asked to respond to a questionnaire to predict how graduates of their program would perform on the certification examination. They were also asked to grade the performance of candidates during their 2 years of training. The data from this study suggest that training directors can accurately predict the performance of graduates of their programs on the certifying examination.

In 1986 the committee approved combining the format of the Preschool and Grade School sections of the examination. Both clinical vignette and audiovisual tapes are used for the oral examination. The committee continues to evaluate its examination, with recent special attention paid to the sections using videotaped interviews.

Combined Training In 1984 the committee established the Articles of Agreement for the Pilot Project on Training in Pediatrics, Psychiatry and Child Psychiatry. The proposal was approved by the American Board of Pediatrics, the APA, the AACAP, the AADPRT, the AACDP, the Society of Professors of Child and Adolescent Psychiatry (SPCAP), and the Psychiatry RRC. All members of the Pediatrics-Psychiatry Joint Training Committee (PPJTC) were selected by the organizations involved in the project, the ABPN, the Committee on Certification in Child and Adolescent Psychiatry, the American Board of Pediatrics, and the National Institute of Mental Health (NIMH). The executive vice president of the ABPN and the deputy medical director for education of the APA were ex officio members. The NIMH and the University of Illinois funded the first meeting of the PPJTC, in November 1984. Members of the PPJTC made site visits to eight facilities for evaluation of their proposed programs. Six training programs were selected: Albert Einstein College of Medicine, Brown University, Mt. Sinai College of Medicine in New York, Tufts/New England Medical Center, the University of Kentucky, and the University of Utah. The first residents accepted for this triple-board project began their PGY-1 in July 1986. The programs last 5 years, and graduates can sit for the certifying examinations in pediatrics, general psychiatry, and child and adolescent psychiatry upon successful completion of this combined program. The first class graduated in June 1991. Several have chosen academic careers, and there is a trend toward working in consultation-liaison child and adolescent psychiatry.

The ABPN, in conjunction with the American Board of Pediatrics and the NIMH, funded an evaluation of the triple-board project in 1985. The project was studied and evaluated over a 10-year period. Both boards agreed to give permanent status to the triple-board program and have distributed guidelines to programs wishing to submit proposals to both boards. Currently 10 approved programs are listed in the AMA *Directory of Graduate Medical Education*.

Recertification The board has reviewed the issue of recertification. In 1975, the following principles were accepted; (1) the board is responsible for administering a recertification examination, (2) the board is not responsible for educational activities related to the preparation for recertification, and (3) the board is responsible for advising specialty organizations about specific educational deficiencies identified by examination results.

In 1976 the board established a Steering Committee on Recertification with representation from the board and its constituent organizations. Separate subcommittees were established for the two disciplines. Constituent psychiatry organizations included the APA, the Committee on Certification in Child and Adolescent Psychiatry, and the AACAP. Constituent neurology organizations included the American Neurological Association, the American Academy of Neurology, and the Child Neurology

Society. Representatives from the private sector also served on the committee. Recertification plans for psychiatry and for child psychiatry and neurology were submitted to the ABMS in 1979. The proposal included the following principles: (1) candidates must hold a valid certification from the ABPN and a current medical license; (2) recertification would be voluntary; (3) plans for psychiatry, child and adolescent psychiatry, and neurology would be similar but not identical; and (4) the process would not be a repetition of the original certification process.

The ABMS Committee on Certification, Subcertification and Recertification (COCERT) reviewed the proposal and suggested some modifications. The board continued to review recertification issues but did not resubmit a proposal until 1988. The board worked closely with the Task Force of the APA. In 1989, the ABPN established a new policy; beginning in October 1994, all certificates would be 10-year, time-limited certificates, and a recertification mechanism would be put in place no later than 2000. As a minimum, recertification will include a test of cognitive knowledge, and the board will also try to make its examination practice sensitive. In February 1994 the ABPN approved a commission on recertification consisting of three psychiatry directors, two APA appointees, and an AACAP appointee from the board's Child and Adolescent Psychiatry Committee. The commission reports to the Psychiatry Council of the ABPN. The commission endorsed using the APA Practice Guidelines as an educational resource for recertification. It also voted for a written, open-book, take-home examination, with psychiatrists given up to 2 months to submit their answers for scoring. The examination is expected to be given every 6 months.

In the meantime, all 24 boards have been approved for recertification for their specialties. In addition, several boards have begun issuing time-limited certificates that are only valid for a specific time period after passing a board examination. At that point, the diplomate must successfully complete a recertification examination to remain board certified. All new certificates issued by the ABPN are 10-year time limited. New certificates in Geriatric Psychiatry, Clinical Neurophysiology, Addiction Psychiatry, and Forensic Psychiatry are all time limited.

Subspecialty Certification The only subspecialty certificate examination offered by the board in psychiatry for three decades was the one for child and adolescent psychiatry. The board also issues a certificate in Neurology with Special Qualification in Child Neurology. Special requirements must be completed by candidates to qualify for the child neurology examination.

Over the last three decades, the board has at various times reviewed requests for subspecialty certification from other psychiatry groups, such as psychoanalysis, administrative psychiatry, and adolescent psychiatry. Similar requests in neurology have been forthcoming for subspecialty certification in electrophysiology and electrodiagnosis. These requests were either withdrawn or turned down by the board in the past.

In July 1986, the board hosted a conference on subspecialty certification. Six representative groups were invited to attend to initiate or renew a dialogue with the board about subspecialty certification: the American Board of Forensic Psychiatry, which administered its own certifying examination; the American Society for Adolescent Psychiatry; the American Association for Geriatric Psychiatry (AAGP); the American Academy of Psychiatrists in Alcoholism and Addiction (AAPAA); the APA Committee on Administrative Psychiatry, which administers its own certifying examination; and the American Psychoanalytic Association, which has its own certifying process.

Constituent organizations and the board continue to debate the wisdom of offering subspecialty certificates in psychiatry in addition to the child and adolescent psychiatry certificate as well as the recently approved geriatric psychiatry, clinical neurophysiology, addiction psychiatry, and forensic psychiatry certificates. These organizations include the APA, the AADPRT, the AACDP, and the Association for Academic Psychiatry, among others. The APA recommended that the board issue a certificate of added qualifications in geriatric psychiatry in 1987. The board appointed a task force on geriatric psychiatry consisting of representatives from the board, the APA, the American Association of Geriatric Psychiatry, the AADPRT, and an internist, to advise the directors about this proposal. The task force held its first meeting in April 1988, when the American Board of Internal Medicine and the American Board of Family Practice issued their first added qualifications examination in geriatric medicine. In 1988, the ABPN approved geriatric psychiatry as a subspecialty. Since the ABMS must approve any certificates for subspecialization by its member boards, the request for geriatric psychiatry was then submitted to the ABMS. In addition to general certificates, the ABMS had provisions for two types of subspecialty certificates: a certificate in special qualifications and recognition of added qualifications. In 1996 the ABMS voted to simplify the names of certificates and discontinued using the qualifying terms “added qualifications” and “special qualifications.” In the future the ABMS will approve certificates for subspecialization.

A certificate for added qualifications indicated that a candidate had completed at least 1 year of full-time formal training and satisfactorily completed an additional examination in that field administered by the ABMS member board. The training program cannot be freestanding but must be associated with an ACGME or equivalent Canadian-accredited residency. The training program must incorporate a specific and identifiable body of knowledge. The purpose, however, cannot be to train principally for additional technical skills. Also, a specialty board must give the ABMS assurance that it has evaluated the impact and effect of a modified certificate before receiving ABMS approval.

Intent to submit a proposal must be announced to all member boards at least 1 year before the meeting at which the formal proposal will be submitted for a vote. A two-thirds majority of voting members of the ABMS must approve any proposal before it can be effected. The proposal must be distributed 180 days before the vote. The certificate of special qualifications differed from added qualifications in requiring a minimum of 2 years of additional training.

The increased emphasis on subspecialty training and practice has led to the 24 boards issuing 76 special certificates beyond their general certificates as of September 1997. Pediatrics and Internal Medicine had the most (15 and 16, respectively). In March 1999 the ABMS approved a new subspecialty of pediatrics and child neurology: neurodevelopmental disabilities and a new pediatric subspecialty: developmental/behavioral pediatrics.

GERIATRIC PSYCHIATRY The ABMS approved the Geriatric Psychiatry Added Qualifications in September 1989. The first Committee on Certification for Added Qualifications in Geriatric Psychiatry was appointed in the fall of 1989. Recommendations for membership on the committee were submitted by the APA and the American Association for Geriatric Psychiatry (AAGP). The committee consisted of five psychiatrists and one neurologist, and has since increased the number of geriatric psychiatrists and added an internist with added qualifications in geriatric medicine. The committee completed its work on the first examination in July 1990, and the first examination, a 157-item written, multiple-choice examination was administered to 661 candidates in the spring of 1991. Performance results for candidates in 1991 and subsequent years are given in Table 53.1-7. The criteria for taking the geriatric psychiatry examination include being board-certified in psychiatry, having an unlimited license to practice medicine, and completion of a residency in geriatric psychiatry. Through 1996, diplomates devoting 25 percent or more of their professional time to geriatric psychiatry could take the examination without a postresidency training program. The same was true for addiction psychiatry through 1998 and is for forensic psychiatry through 1999.

Year	Examined (N)	Passed (N)	Passed (%)
1991	661	490	74
1992	598	359	62
1994	641	422	66
1995	559	376	67
1996	1023	713	69
1998	72	65	90

Table 53.1-7 Performance—Geriatric Psychiatry Candidates

The ABPN requested that the RRC establish guidelines for postresidency training in geriatric psychiatry at the PGY-5 year level or higher. These guidelines were approved by the ACGME. Programs can now submit proposals for geriatric psychiatry programs for RRC approval. Only graduates of geriatric psychiatry programs have been eligible for the geriatric psychiatry examination since 1997. Only graduates of ACGME-approved geriatric psychiatry programs are now eligible for the geriatric psychiatry examination.

CLINICAL NEUROPHYSIOLOGY The neurologists received ABPN approval (1989) and subsequent ABMS approval (1990) to issue a certificate in added

qualifications in clinical neurophysiology. The test-writing committee includes 12 board-certified neurologists and one board-certified psychiatrist. The first 4-hour, multiple-choice examination was administered in March 1992 at six sites. Of the 306 candidates eligible for the examination (including 7 psychiatrists), 278 were examined, and 227 (82 percent) passed; Examinations were also administered in 1994, 1996 and 1997. A total of 1001 candidates participated and 735 (73%) passed. Candidates who have not completed fellowships must have worked the equivalent of 3 years in clinical neurophysiology following completion of a residency training program. The Neurology Residency Review Committee established requirements for postresidency training in clinical neurophysiology; only those who graduate from programs are currently eligible for this examination, and only those who graduate from ACGME-approved programs will be eligible after 2000.

ADDICTION PSYCHIATRY In 1990 the APA recommended addiction psychiatry for added qualifications. The ABPN accepted the recommendation and its proposal was approved in September 1991 by the ABMS. Nominations were sought for membership on the test-writing committee from the APA and the AAPAA, since renamed the Academy of Addiction Psychiatry (AAAP). The Committee on Certification for Added Qualifications in Addiction Psychiatry held its first meeting in December 1991, and the first written, multiple-choice examination was administered in March 1993. Performance results are given in Table 53.1-8. Only graduates of addiction psychiatry programs are currently eligible for this examination, and only those who graduate from ACGME-approved addiction psychiatry programs will be eligible beginning in 2001.

Year	Examined (N)	Passed (N)	Passed (%)
1993	632	465	74
1994	402	302	75
1996	347	290	84
1997	347	278	80
1998	525	431	82

Table 53.1-8 Performance—Addiction Psychiatry Candidates

FORENSIC PSYCHIATRY The APA recommended that the ABPN issue a certificate in added qualifications in forensic psychiatry. The ABPN accepted the recommendations and submitted a proposal which was approved in September 1992. The first Forensic Psychiatry Committee met in December 1993 and the first examination was administered in October 1994. Performance results are shown in Table 53.1-9.

Year	Examined (N)	Passed (N)	Passed (%)
1994	316	256	82
1996	292	236	81
1998	452	322	71

Table 53.1-9 Performance—Forensic Psychiatry Candidates

PUBLICATIONS

The ABPN published the history of its first 50 years, with Marc Hollender, former president and director, as editor. Shore and Scheiber have edited a book on certification, recertification, and continuing medical education for APPI Press.

BOARD RESEARCH

In 1989 the ABPN approved the new full-time position of Research Coordinator for the Board. Dorthea Juul from the University of Illinois accepted this full-time position in 1990. She works closely with the Research and Development Committees of the board in studying the reliability and validity of board examinations. In addition, she contributes regularly to *Academic Psychiatry*, providing abstracts on articles pertaining to medical education and evaluation.

SUGGESTED CROSS-REFERENCES

Related information can be found in Section 53.2 on examining psychiatrists and in Section 55.1 on the history of psychiatry.

SECTION REFERENCES

*Accreditation Council for Graduate Medical Education: *1998–1999 Directory of Graduate Medical Education Programs*. American Medical Association, Chicago, 1998.

*American Board of Medical Specialties: *Annual Report and Reference Handbook—1998*. American Board of Medical Specialties, Evanston, IL, 1998.

*American Board of Medical Specialties Research & Education Foundation: 1997 ABMS Report on Specialty Board Certification, American Board of Medical Specialties, Evanston, IL, 1997. *Information for Applicants*. 1939, 1944, 1946, 1957, 1972, 1978, 1981, 1982, 1983, 1986, 1987, 1988, 1991, 1992, 1993/1994, 1995, 1996, 1997, 1998 and 1999. Deerfield, IL.

*American Board of Psychiatry and Neurology: *Information for Applicants for Certification in Child and Adolescent Psychiatry*. 1999.

*American Board of Psychiatry and Neurology: *Information for Applicants for Certification in Added Qualification in The Subspecialty of Geriatric Psychiatry, Clinical Neurophysiology, Addiction Psychiatry, Forensic Psychiatry*. 1999. Deerfield, IL.

Braceland FJ, Boyd DA: Secretary of the board: Apologia Pro Vita Sua. JAMA *148:*708, 1952.

Brosin HW: Working agreements (1953–1962). In *ABPN Workbook*, HW Brosin, editor. American Board of Psychiatry and Neurology, Rochester, MN 1962.

Carmichael HT, Small SM, Regan PF: *Prospects and Proposals: Lifetime Learning for Psychiatrists*. American Psychiatric Association, Washington, DC, 1972.

Freeman W, Ebaugh FG, Boyd DA: The founding of the American Board of Psychiatry and Neurology. Am J Psychiatry *115:*769, 1959.

*Hollender M, editor: *The American Board of Psychiatry and Neurology: The First Fifty Years*. American Board of Psychiatry and Neurology, Deerfield, IL, 1991.

Langsley DG: Changing patterns of psychiatry specialty certification in the English-speaking countries. Am J Psychiatry *138:*493, 1981.

Langsley DG, Darragh JH, editors: *Trends in Specialization: Tomorrow's Medicine*. American Board of Medical Specialties, Evanston, IL, 1985.

Langsley DG, Yager J: The definition of a psychiatrist: Eight years later. Am J Psychiatry *145:*469, 1988.

*Mancall EL, Bashook PG, editors: *Evaluating Residents For Board Certification*. American Board of Medical Specialties, Evanston, IL, 1998.

Mancall EL, Bashook PG, editors: *Recertification: New Evaluation Methods and Strategies*. American Board of Medical Specialties, Evanston, IL, 1994.

*Marquis Who's Who: *The Official Directory of Board Certified Medical Specialists*. Marquis Who's Who, New Providence, NJ, 1999.

McDermott JF, McGuire C, Berner ES: *Roles and Functions of Child Psychiatrists*. Committee on Certification in Child Psychiatry of the American Board of Psychiatry and Neurology, Evanston, IL, 1976.

McDermott JF, Tanguay PE, Scheiber SC, Juul D, Shore JH, Tucker GH, McCurdy L, Terr LC: *Reliability of the Part II Board Certification Examination in Psychiatry: Interexaminer Consistency*. Am J Psychiatry *148:*1672, 1991.

Nadelson CC, Robinowitz CB, editors: *Training Psychiatrists for the '90s: Issues and Recommendations*. American Psychiatric Press, Washington, DC, 1987.

Rudy LH: Reasons given for success after initial failure on the American Board of Psychiatry and Neurology Part II examination. Am J Psychiatry *138:*1612, 1981.

Scheiber SC: Balancing validity and reliability by using live patients. In *Assessing Clinical Reasoning: The Oral Examination and Alternative Methods*, EL Mancall, PG Bashook, editors. American Board of Medical Specialties, Evanston, IL, 1995.

Scheiber SC: *Certification and Recertification*. Psychiatry Q *62:*2, 1991.

Scheiber SC: Frequently asked questions about the American Board of Psychiatry and Neurology. Acad Psychiatry *17:*43, 1993.

Scheiber SC: Recertification: Implementation strategies. In *Recertification: New Evaluation Methods and Strategies*, EL Mancall, PG Bashook, editors. American Board of Medical Specialties, Evanston, IL, 1994.

Scheiber SC: Specialty board assessment issues and opportunities. In *Recertification: New Evaluation Methods and Strategies*, EL Mancall, PG Bashook, editors. American Board of Medical Specialties, Evanston, IL, 1994.

Shapiro T, Juul D, Scheiber SC: Exploration of failure on the subspecialty examination for child and adolescent psychiatry. Am J Psychiatry *153:*693, 1996.

*Shore JH, Scheiber SC, editors: *Certification, Recertification and Lifetime Learning in Psychiatry*. American Psychiatric Press, Washington, DC, 1994.

Small SM: Continuing certification in psychiatry. Psychiatr Opin *14:*19, 1977.

Talbott JA: Is the “live patient” interview on the boards necessary? Am J Psychiatry *140:*890, 1983.

Talbott JA: Opposition to ‘coercive continuing education and mandatory recertification.' Am J Psychiatry *136:*923, 1979.

Tucker GJ, Martin MJ, Scheiber SC: Subspecialization in psychiatry. Am J Psychiatry *148:*11, 1991.

Webb LC, Juul D, Reynolds CF, Ruiz B, Ruiz P, Scheiber SC, Scully J: How well does the psychiatry residency in-training examination predict performance on the American Board of Psychiatry and Neurology part I examination? Am J Psychiatry *153:*831, 1996

CHAPTER 53. PSYCHIATRIC EDUCATION

53.2 EXAMINING PSYCHIATRISTS AND OTHER TRAINEES

JAMES MORRISON, M.D., AND RODRIGO A. MUÑOZ, M.D.

By the time training is completed, a psychiatrist (or any other mental health professional) has probably been examined many hundreds of times on material that has consumed tens of thousands of study hours. The uses and consequences of all this testing are varied. Some of them are especially important to the individual student as a spur to greater study effort, a signal that the examinee is ready for the next stage in sequential learning, an evaluation of subject matter mastery, a rite of passage or initiation, or status or advertising for those who have passed the examination. In some ways, the examination serves the institution or fellow professionals. It provides feedback regarding teaching effectiveness, measures training program effectiveness, maintains discipline among students, and limits the competition. Finally, testing may serve society as a whole by ensuring quality of the licensee.

During a psychiatrist's march through medical school, postgraduate training, and on to board certification, most of the above-mentioned goals are likely to be a focus of one examination or another. With the maturation of the science of statistics, the story of psychiatric testing has increasingly been that of national examinations. Through the midtwentieth century, individual facilities constructed their own tests; but each year the expense of producing new testing material grew, as did the need to compare the abilities of clinicians who trained at different institutions. Examinations that are standardized throughout the United States are now a central feature of nearly every psychiatry training program.

TEST MAKING, TEST TAKING

Designed to test the extent of clinical knowledge acquired by the end of medical school, the National Board of Medical Examiners (NBME) step II examinations is given twice each year, in August and March of the fourth year of training. Its construction begins with a content outline. Although the NBME will not release the percentage of step II questions devoted to psychiatry, it is one of the five major disciplines represented in this examination. Officials point out that many questions are interdisciplinary.

The psychiatry section development committee comprises eight members who serve 3-year staggered terms. When they come to national board headquarters for 2 days of orientation, each member is assigned to write 50 questions for a total of 350 (the subcommittee chair does not write any). With the help of staff, these questions are then edited to conform to NBME style. Parallel committees and processes apply to each of the other major disciplines: medicine, obstetrics-gynecology, pediatrics, and surgery.

In a group meeting, each question is read aloud by its author. Others in the group may suggest changes in wording, answers, or distractors (the incorrect choices in a multiple-choice examination) before an item is finally accepted. Whenever a question is rejected outright, the author must supply another that tests the same area of knowledge. Questions are also tested by reading them aloud in front of the chairpersons of each of the major subcommittees. No specific degree of difficulty is prescribed. Rather, there must be consensus that each item tests fairly the knowledge that graduating physicians should possess prior to beginning the first postgraduate year (internship). Many questions pose problems that cross the boundaries of two disciplines, for example, the distractors might suggest that the answer could belong to either medicine or psychiatry.

Ultimately, the completed items are pretested in an actual examination. Some 40 different versions of step II will be given during any 1 year. Besides the test questions that contribute to a candidate's score (the NBME calls them *live items*), each version includes a number of items that are being pretested. The live items are the same for all examinations; the pretest items are distributed so that each will be answered by a representative sample of the 50,000 candidates who take the examination each year. The board evaluates the pretest items against the live items and so determines the composition of the next year's examination. The pretest also determines whether any of the distractors are left unchosen; if so, they may be discarded in favor of alternatives that will appeal to more candidates.

For each item a statistic is calculated called *the point biserial discrimination index* (PBDI), which quantifies the correlation between the item and the examinees' overall test performance. If it is positive, a high-scoring candidate tends to do well on the item. A PBDI above 0.20 is desired. New test items may be discarded if they are answered correctly by fewer than 30 percent of examinees or have a negative PBDI.

In recent years two types of question format have emerged as most favored on national tests: the best-single-answer multiple-choice question and the extended matching items (see examples below). This is not to say that multiple choice questions have no drawbacks. Even when composed by professional test writers, they can occasionally be interpreted ambiguously; candidates sometimes complain that subjects are controversial or worse, trivial. But increasingly, several old standbys are now discarded because they are confusing or they assess the examinee's ability to take tests rather than knowledge of the subject material. The true-false question is written in terms of absolutes. Any item that involves a shade of gray requires the examinee to guess what the author had in mind. (Example: “The prevalence of schizophrenia is 1%” might be answered differently by a clinician and an epidemiologist.) The multiple true-false question is a complicated vehicle that requires logic as well as knowledge of subject content. Its instructions are Mark “A” if choices 1, 2, and 3 are correct; “B” if 1 and 3 are correct; “C” if 2 and 4 are correct; “D” if only 4 is correct; “E” if 1, 2, 3, and 4 are correct. (Example: Typical symptoms of somatization disorder include 1. Abdominal pain 2. Seizures 3. Intolerance to foods 4. Impotence.)

Constructing a Test

1. Decide on the pass-fail criterion. A *norm-referenced* test (each candidate's score is reported as it relates to all others' in the group) is designed to produce variance and highlight differences in examinees. A *criterion-referenced* test (each score is given in terms of the absolute percentage correct) yields specific information about what the individual knows. Because both the NBME and American Board of Psychiatry and Neurology (ABPN) Part I examinations are criterion referenced, theoretically all candidates taking either examination could pass.
2. Select the knowledge areas and question content to be tested. Each area will require a certain number of items (about 30) to provide a reliable evaluation. Reliability is increased by greater test length and longer time for the test.
3. Question content should cover important problems that are common or potentially high risk. Questions should assess knowledge likely to be needed by a clinician in the field; avoid minutiae and disorders that are vanishingly rare. Additionally, focus on important areas of clinician knowledge and behavior, such as health and health maintenance, disease mechanisms, diagnosis, and treatment.
4. Determine the question format. Both the national boards and Part I of the ABPN examination rely mainly on multiple choice questions of several types: (1) single best answer, a statement followed by several possible responses from which the candidate must choose; (2) extended matching, as of a series of numbered statements to a series of lettered answers; (3) a vignette followed by several single best-choice questions.
5. Write the questions. Clarity and brevity are desirable qualities (but see no. 6 below). Avoid vague terms such as *usually, frequently, often*, and *may*. Questions should not just measure the ability to memorize; they should require the examinee to retrieve, select, organize, and apply knowledge.
6. Question stems (the information that serves as background to the actual question) that contain detailed information discriminate better between examinees who score well and those who score poorly on the entire test.
7. Devise distractors. These should be neither too hard nor too easy. A distractor that is never selected should be discarded from future editions of the test.
8. Evaluate the new questions for their ability to discriminate top from bottom candidates (i.e., have a positive PBDI). A pretest sample or previous edition of the test will help to determine which candidates will answer an item correctly.

9. After the examination has been given, some items may be found that are often answered incorrectly by top examinees (i.e., they have negative PBDIs). These flawed items must be discarded.

Anatomy of a Single-Best-Answer Test Question

A 31-year-old single woman is admitted to the hospital complaining of voices in her head that tell her to kill herself. She says she has had several prior admissions over the past 6 years, but her clinical chart cannot be found. She has heard voices on and off for many years and believes they may be the result of “interplanetary influence.” An actress by training, she has not worked in her profession since age 24. She has sustained herself by cleaning offices at night and with general relief. Upon close questioning she thinks she might have bumped her head several weeks ago. “But I was drinking,” she says with a giggle, “and I don't remember much about it.” She then denies habitual use of alcohol or other substances.

Fully oriented and coherent, the patient appears distracted during the interview, several times gazing around the room and cocking her head to one side. She interprets the proverb “A stitch in time saves nine” to mean that she should “sew up my ragged sleeve of carelock.” Except for occasional laughter when she says something that does not seem especially funny, her affect is unvarying and she gazes straight at the examiner.

The subject concerns a major area of concern to psychiatrists (psychosis is frequently encountered and fraught with risk).

- The stem is a relatively long, detailed clinical vignette.
  1. What disorder most likely has caused this patient's symptoms?
     A. Alcohol dependence with hallucinosis
     B. Bipolar mood disorder
     C. Schizoaffective disorder
     D. Schizophrenia
     E. Subdural hematoma
  2. Which medication will most likely provide rapid relief of symptoms?
     A. Amitriptyline
     B. Diazepam
     C. Divalproex
     D. Haloperidol
     E. Methylphenidate
  3. The patient's use of the term “carelock” is an example of
     A. Inappropriate affect
     B. Neologism
     C. Paranoia
     D. Perseveration
     E. Tangentiality
- The questions pose several types of problem related to the information provided in the stem, but these are not linked (one answer does not depend on knowing prior answers). Questions require deduction from the facts given, not just recall of isolated facts.
- The distractors are all logical and plausible (they have the same length, format, and type of word or phrase as the correct answer).
- Options are presented alphabetically so as not to favor a given letter or position in the list.
- The third question is actually an older, less satisfactory type of question: it draws on rote memory rather than ability to assess and apply principles.
- The answers (including distractors) are all relatively brief.

If desired, the same vignette could serve as the basis for additional questions about psychopathology, further testing, psychotherapy, family history, and the like.

Sample Extended-Matching Question

For each patient in items 1–5, select from the common set of lettered choices (A–O) the *most likely* diagnosis for the given chief complaint. Each choice may be selected once, more than once, or not at all.

A. Antisocial personality disorder
B. Attention-deficit/hyperactivity disorder
C. Brief psychotic disorder
D. Cocaine withdrawal
E. Dementia of the Alzheimer's type
F. Female sexual arousal disorder
G. Major depressive disorder
H. Mental retardation
I. Posttraumatic stress disorder
J. Pyromania
K. Schizophrenia, disorganized type
L. Schizophreniform disorder
M. Somatization disorder
N. Substance-induced delirium
O. Vascular dementia

1. A 6-year-old whose mother complains of poor progress in the first grade
2. A young married woman with a 6-month history of anorgasmia
3. A college student who cheerfully admits that he has heard voices for 2 months
4. A 24-year-old man arrested for setting fires for building owners who want to collect on insurance
5. An elderly businessman increasingly forgetful of his appointments

Tips for Surviving Written Examinations

1. Answer all questions.
2. Use a watch to help allot sufficient time to each section of the examination.
3. Use the full time allotted.
4. In the absence of definitive knowledge, guess on all questions, unless the test subtracts points for wrong answers (rare today).
5. *Never* and *always* are red flags signaling incorrect distractors.
6. *Possibly, sometimes*, and *maybe* may signal a correct answer.
7. Try to narrow the options by eliminating known incorrect answers.
8. Test writers take pains not to give away correct answers by placing them at any one point on a list.
9. Candidates with time left at the end of the examination should check their work. Research clearly establishes that examinees are far more likely to change an incorrect answer to a correct one than the other way around.

THE PSYCHIATRY BOARDS

The ABPN examination in psychiatry has been administered yearly since 1935. Almost since inception it was given orally; even knowledge in the basic sciences was ascertained by face-to-face query. By the mid-1960s the logistics of an oral examination were complicated by increasing numbers of candidates; an exploding volume of information about psychiatry; and a persisting, increasing demand for an examination procedure that was both impartial and fair. The logical solution seemed to be

a written examination, and during its more than quarter century of use, it has become a relatively inexpensive screening device for the vastly more complicated and expensive oral examinations of Part II.

Part I (Written) Boards The first ABPN examinations depended in part upon essays. Essay questions are easy and inexpensive to write, and they can potentially plumb the depth of a candidate's knowledge. They also have significant disadvantages. Because of the length of time it takes to answer them, relatively few areas of knowledge can be tapped in a single examination. A candidate unfamiliar with just one of, say, five questions may automatically hover perilously close to failure. Essay grading is time consuming and nearly impossible to do without bias. Factors such as poor handwriting or language skills (the latter a problem for some international trainees), which have nothing to do with knowledge of psychiatry, may figure too prominently in the grading process.

The speedy demise of the essay examination and a brief flirtation with a true-false component in the early 1940s cleared the way for various types of multiple-choice question. Several variations on these questions now form the bulk of standardized written psychiatry examinations administered throughout the Western world.

The Part I examination is used partly as a screening tool and partly to assess the candidate's knowledge of basic sciences. It is administered in the autumn of each year at many medical schools and hospitals across the United States. Currently, the Part I examination comprises 7 hours (half each in the morning and afternoon) of computer-scored questions covering both psychiatry and neurology. For psychiatrists, the examination covers the material listed in Table 53.2-1.

For candidates in psychiatry, the psychiatric (major) portion of the examination consists of approximately 260 questions from the following areas:
- Growth and development
- Biological dimensions of psychiatry
- Psychosocial aspects of psychiatry
- Clinical psychiatry, including
  - Phenomenology
  - Diagnosis
  - Treatment

The neurology section (as taken by candidates in psychiatry) consists of approximately 150 questions from the following areas:
- Clinical neurology
  - Neuroanatomy
  - Neuropathology
  - Neuropharmacology
- Clinical diagnostic procedures, including cerebral spinal fluid
- Neuro-ophthalmology
- Electroencephalography
- Neuroimaging
- Therapy and management of patients with neurological disorders

Table 53.2-1 Written Examination for Psychiatry Candidates

Previous editions of this text included somewhat more detailed descriptions of the content of the written examinations. The ABPN no longer permits the publication of copyright material it distributes to candidates who have been accepted to take the Part I examination. This change results from recent litigiousness on the part of a few unsuccessful candidates. The board now finds itself spending inordinate amounts of time and money fighting an increasing (though still small in absolute terms) number of lawsuits brought by unsuccessful candidates who prefer a legal solution to the effort of further preparation as a means for achieving certification.

Board officials have also noted an increase in irregular behavior associated with the certification process. Some candidates have refused to hand in their question booklets at the end of the written examination; some have been found copying answers from the papers of others. The use of forged board certificates also appears to be on the rise. In some cases, psychiatrists have testified in court that they are board certified when they were not.

Standards to Pass To pass Part I, a candidate must pass both the neurology and psychiatry sections. The number of candidates who will pass is not established in advance; rather, the writers of the examination determine what percentage of the questions a successful candidate should be able to answer correctly. However, over the years the pass-fail ratio has remained approximately constant at about three to two. Unsurprisingly, someone taking Part I for the first time is much more likely to pass than is a candidate who has previously failed this examination. A related finding, true for the ABPN examinations as well as for other specialty boards (e.g., Family Practice and Internal Medicine), is that younger candidates consistently perform better than those who are older.

Besides a statement of overall pass or fail for the Part I examination, candidates receive a report giving their subscore test results for the areas mentioned in Table 53.2-1 and for the psychiatry minor (neurology) portion of the examination. Correlations between subscores for the psychiatry major range from about 0.5 to 0.8, indicating that these subtests tap different aspects of learning.

Passing Part I is the only way a candidate can qualify to take Part II. Despite the potential for an oral examination to induce emotional trauma, many candidates see Part I as the real barrier to certification. As far back as 1977 it was noted that 88 percent of psychiatrists who passed Part I on the first try went on to become board certified within the next 4 years; this was true for only 12 percent of those who failed their first try at Part I. Apparently, many candidates who fail on the first attempt either become discouraged and quit or fail again (either part) on a subsequent attempt. Candidates who successfully complete Part I, then fail Part II three times or who allow 6 years to elapse without passing Part II must retake the Part I examination. Table 53.1-2 and Table 53.1-3 list the successful completion rates for Part I and Part II, respectively, over a recent 10-year period.

Part II (Oral Examination) Perhaps in part because it is the most verbal of all the medical sciences, psychiatry has clung tenaciously to the oral examination format. As of this writing it remains the only specialty board to require a witnessed examination of a live patient for certification.

After passing Part I, candidates are scheduled for a Part II examination within the succeeding 12 months. Because scheduling is so complicated, candidates cannot choose the date or place for the Part II examination. However, travel is minimized to some extent by the fact that examinations are offered three times a year in various parts of the country. The location of the Part II examination changes from one year to the next to avoid placing too great a burden on any city's mental health facilities. Most years, one of these examinations is so large (roughly 1000 candidates) that only five cities can accommodate it: Boston, Chicago, New York, Philadelphia, and Washington, D.C. The other two examinations are generally smaller (roughly 500 candidates each) and are distributed geographically around the country.

Part II actually comprises two examinations, each of which involves personal interaction of the candidate with two or more examiners. In one of these encounters, the candidate is questioned about a videotaped interview; in the other, questioning concerns an interview with a live patient that the candidate has just completed under the watchful eyes of the examiners.

Live Patient Examination Nearly everyone regards the interview of an actual patient as the more difficult half of the oral boards. This activity, though in substance no different from what most psychiatrists do every workday, carries enough emotional baggage to make the process far different from any normal patient interview. During this interview, candidates have the opportunity to demonstrate their ability to establish rapport, gather information, elicit and interpret signs and symptoms, and use the information obtained to arrive at a reasoned diagnosis.

This examination is conducted by two psychiatrist examiners unacquainted with the candidate, who has exactly 30 minutes to interview an unfamiliar patient for whom no records or other information have been provided. After the interview has been concluded and the patient has been excused, the examiners offer the candidate a minute or two to review notes and prepare the presentation. During the remaining 28 minutes, the candidate presents a précis of the patient's history; offers a formulation about the diagnosis, dynamics, and recommended therapy; and answers examiner's questions about any of these matters—or any other subject pertaining to clinical psychiatry.

During this entire exercise, the examiners evaluate the candidate's ability to interact with the patient, organize the material, understand and interpret the patient's psychopathology, and think through problems pertaining to diagnosis and treatment. How well the candidate handles these tasks depends in part on having an effective plan for organizing these 60 minutes.

A half-hour witnessed interview without the benefit of collateral information challenges any interviewer. In gathering data that adequately address diagnosis, the candidate must simultaneously demonstrate compassion and tightly control the patient's verbal output. Although the time schedule depends to some extent on the

individual patient, Table 53.2-2 suggests approximate times that should be allotted to the various portions of the interview.

5 minutes	Chief complaint and patient's free discussion of the issues
7 minutes	Specific questions about diagnosis, pertinent negatives, and history of suicidal ideas, violence, alcohol, and drug use
5 minutes	Medical history, review of systems, family history
5 minutes	Personal and social history, evaluate character pathology
5 minutes	Mental status examination
3 minutes	Buffer time: pursue other issues, late-developing leads

Reprinted with permission from Morrison J, Muñoz RA: *Boarding Time*, ed 2. American Psychiatric Press, Washington, DC, 1996.

Table 53.2-2 Time Allotment for Patient Interview

Careful control over the interview process is vital to obtain the necessary information within the brief time available yet establish and maintain rapport with the patient. After an initial few minutes when the patient speaks freely about the various areas of concern, the candidate should begin using closed-ended questions that rapidly review all pertinent portions of the standard psychiatric history and mental status examination. Loquacious patients and those who tend to wander from the point must be gently, but firmly, herded back onto the path. Although patients who are vague, circumstantial, hard to understand, hostile, or complicated uniformly prove more difficult to interview, candidates should try to regard them as an opportunity to demonstrate a capacity for showing empathy while eliciting information in trying circumstances.

Candidates are especially advised to keep their own verbal output to a minimum; if they are talking, the patient cannot give them information. They should not discuss their own feelings, offer therapy, disagree with any aspect of the patient's own treatment, or show irritation or any other negative affect if the patient is slow to provide the information needed. Confrontation should be avoided at almost any cost, except when a patient's restlessness or other signs of discomfort indicate significant resistance to the entire interview process. Then, the candidate may have to make the patient explain the lack of cooperation. Offering the opportunity to explain a sad or hostile affect might break through the resistance.

Of course, the emphasis in any interview is altered by the presumptive diagnosis. A patient who is cognitively impaired may need a much more concrete style of interviewing that focuses on the mental status evaluation and the activities in a typical day; an older patient needs more time to describe losses (e.g., family, friends, mobility) and difficulties performing activities of daily living. Patients who have a psychosis or an anxiety or mood disorder require special attention to the criteria for the diagnosis according to the fourth edition of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV). But for every patient, one must carefully review personal and social information, including current living situation, work, family, marital status, and leisure activities.

PRESENTING THE PATIENT In presenting the just-gathered information to the examiners, organization is again the key. Examiners expect a brief (5 minutes or so) recapitulation of pertinent facts in the case followed by a formulation, not a verbatim retelling of the entire interview. A suggested outline is given in Table 53.2-3.

Identifying data
Present illness
Past history
Psychiatric history
Medical history
Review of systems
Family history
Personal and social history
Mental status evaluation
Most likely diagnosis
Differential diagnosis
Formulation
Predisposing factors (e.g., genetic)
Psychosocial stressors (e.g., abuse as child)
Stresses that precipitate present illness
Characteristic defenses—adaptive style
Proposed workup
Treatment plan
Prognosis

Table 53.2-3 Outline for Case Presentation

During the presentation, candidates sometimes discover to their chagrin that they have forgotten to obtain an important piece of information. Happily, they get a chance to recover, even from failing to request such vital data as suicidal ideas or substance use. If during the presentation they realize that they omitted something important, they can compensate by stating what they should have asked (with appropriate follow-up questions) and how they would use that information. Sometimes an examiner may even give a candidate a third chance to discuss important information such as suicidal ideas or substance use.

> While presenting a patient whose principal symptoms seemed to be characterological, Dr. A. had sailed right through to the formulation when one of the examiners coughed and asked, “Did you note any thoughts of self-destruction?” After a moment of horror-struck silence, the candidate blurted out, “You mean, like I'm having right now?”
>
> Quickly recovering, Dr. A. continued as follows: “I completely forgot to ask my patient whether she had ever tried to kill herself or ever had any thoughts to that effect. If I were seeing her in my practice, I'd immediately go out, bring her back into the office, and ask, ‘Have you ever had any thoughts about harming yourself, made any suicidal attempts.' I'd want to know all the details of any attempts—methods, degree of suicidal intent, any physical harm. This is really important stuff. Suicidal behavior figures in the diagnostic criteria for the mood disorders as well as borderline personality disorder. And it is quite common in a number of other conditions where it isn't a DSM-IV criterion—schizophrenia, substance abuse, and somatization disorder, for example. I can't imagine forgetting to ask, except that I am pretty nervous.”
>
> At the end of the session, the examiners instantly agreed that Dr. A. deserved to pass.

The presentation is the time for candidates to show how well they can manage clinical problems, set priorities, and recognize and cope with problems resulting from therapy. They must be ready to offer a differential diagnosis, even if the patient's most likely diagnosis seems open and shut. They will also be expected to show that they can defend diagnoses with criteria. Because the diagnosis of any one patient is less important than all the patients the candidate will evaluate throughout a professional lifetime, the examiners are hoping to find a candidate who can reason through a clinical problem.

The formulation organizes into a brief paragraph the biological, psychological, and social factors that influence the evolution of the disease process and produce the current symptoms. It should be followed by suggestions for workup, which might include such procedures as a physical examination, laboratory testing, imaging and such simple devices as obtaining records from previous clinicians and interviewing informants. Of course, all discussions of treatment should be based on the biopsychosocial model.

After an excellent presentation of a patient who had described many symptoms of borderline personality disorder, candidate B. would only consider behavior modification as a treatment option. The examiners repeatedly asked whether medication could be a consideration (the patient did report some depressive symptoms), but Dr. B. wouldn't hear of it: “There is little solid evidence that medication has any long-range effect in BPD.” Even when one examiner postulated an extreme case in which such a patient developed severe depression, the response was the same—refusal to consider another course of action. Despite an outstanding interview, the best the examiners could justify for this otherwise excellent candidate was a grade of Condition.

Interactions with the examiners should be as candid as possible. Candidates should avoid using gamesmanship, trumpeting their own expertise or ignoring examiners' interests and opinions. Respectful acceptance of other points of view suggests the ability to think flexibly. A candidate who does not understand a question should request clarification. Not only is this perfectly in order, it is the safest course to follow. Candidates who cannot answer a question are well advised to say so, then add how they would go about finding the answer. No one is expected to know everything, and many examiners feel that good physicians know their own limits.

Videotape Examination Some candidates consider the videotape examination to be a comparatively easy ride, and in some ways it is. Candidates do not have to manage an interview, only absorb a circumscribed amount of information, and they are not subject to the vagaries of temperament displayed by sometimes highly individualistic patients. Still, the videotape examination, a staple of the Part II examinations since 1982, presents its own challenges that recommend mental preparation, as well as outright practice, prior to the actual event.

Each video comprises about 25 minutes of interview with an actual patient. The quality of the tapes is usually less than professional; if they were not followed by half an hour of intense questioning by two examiners, many would not even be very interesting.

Often the vignette seems sliced from a larger interview. It may appear to have no definite beginning or end, and it will probably omit certain vital facts about the patient, such as age, martial status, or occupation. Some interviews seem almost maddeningly unfocused or even trivial and would surely earn a failing grade if conducted by a candidate for the Part II examination.

With the prospect of a relative information drought, candidates should sit as close as they can to the source and be prepared to absorb every drop. Because it is dangerous to look away from the screen for even an instant, it is probably best to keep note taking to the barest minimum. Otherwise, something important in the way of a behavioral mannerism—a tear, a sidelong glance at the camera—could escape unnoticed. During practice sessions, candidates might try using a mental filing system to storing and later recall pertinent observations. This system depends on filling in an outline similar to the one followed with a live patient interview, except that at the end many more bare spots will exist. Bits and pieces of historical material are likely to be scattered throughout the 20 to 25 minutes the camera spends focused on this patient; these should be mentally recorded under present illness for later use in the presentation. Other sorts of information (e.g., family history, personal and social history, past psychiatric history) should be similarly filed.

EXAMINATION PROCESS The video interviewer will probably not conduct a formal mental status evaluation, effectively restricting firm conclusions about recent memory, capacity for calculations, and the ability to tell similarities and differences. Instead, inferences can be drawn from evidence supplied by cognition and behavior, including focus and attention (Does the patient stick to the topic at hand?), language (Is there any apparent aphasia? What does word use and accent tell about upbringing, education, and place of origin?), memory (Can the patient refer back to early portions of the interview?), insight (Is there an appreciation that something is wrong?), and judgment (What does the patient say that indicates compliance with treatment recommendations?).

After the videotape ends, the candidate and examiners will discuss the case in a private room. The presentation of the videotaped patient should proceed along the same general lines as those used for any other patient (Table 53.2-3). The fragmentary nature of this interview means that many data will be missing. It may be tempting to overinterpret those data that are present. Candidates should avoid this temptation and instead volunteer what questions they would like to ask or further observations they would want to make to answer an important diagnostic question.

Noting that the videotaped patient had several times glanced off screen, Dr. C. hypothesized that the behavior could be caused by hallucinations. “Of course, there could have been activity off-camera that we couldn't see” the candidate volunteered. “I'd need to ask the usual sorts of questions about auditory and visual hallucinations, delusions, and so forth before I could say for sure. Would you like me to go through them?”

Dr. C. was playing it safe by mentioning other reasonable interpretations. This is an excellent practice to follow in discussing the videotape interview, which may well not present sufficient information for a definitive diagnosis. As in a live patient examination, flexible thinking is the key to success.

The examiners usually ask the candidate to select for discussion one diagnosis from the list of differentials. Candidates occasionally come to grief because they are unable to select one; some even refuse. The examiners are not trying to set a trap, but merely want a diagnosis—any one will do—upon which to center a discussion of evaluation and treatment.

Finally, any other question about clinical psychiatry can be raised during the course of the discussion of the videotape examination. Examiners might want to hear the candidate's assessment of the rapport between the patient and interviewer, what steps should be pursued for further evaluation, or how psychotherapy might best succeed with such a patient.

SUCCESS AND FAILURE ON THE ORAL BOARDS

Using the official ABPN grading categories, examiners from one Part II team in the mid-1990s developed a list of areas they regarded as crucial to a candidate's success on the examination as a whole. These criteria examined how the candidate handled two basic types of problems. As a group, candidates who passed the overall examination handled each of these 18 items better than those who failed.

Content. Did the candidate:

1. Use appropriate opening and closing strategies
2. Follow cues to new information
3. Appropriately explore affective cues
4. Demonstrate language and cultural sensitivity in communications
5. Use a questioning technique that was open-ended but appropriately structured
6. Control and direct the course of the interview into a cohesive whole Process. Did the candidate:
7. Obtain enough data about the chief complaint and history of the present illness
8. Begin to sketch in relevant data concerning developmental, family, medical, and social histories
9. Sensitively explore any history of alcohol and drug abuse
10. Sensitively assess risk of suicide and homicide
11. Present an organized, concise summary of important data
12. Recount an organized accurate mental status examination
13. Consider emergency issues such as alcohol or drug use, suicide, and violence
14. Present a rational, organized workup that included additional history, collateral information, and testing
15. Choose an appropriate working diagnosis from a sufficiently broad differential for Axes I, II, and III
16. Use all three dimensions of a biopsychosocial formulation
17. Present a comprehensive treatment plan that included drugs, psychotherapy, and criteria for hospitalization and show a knowledge of family systems and milieu principles
18. Discuss negative and positive prognostic factors and transference and countertransference issues

For the live patient examination, four criteria best predicted outcome in a stepwise logistic regression. In descending order of importance they were treatment plan (item 17), information cues (2), mental status examination (12), and interview control (6). Failing candidates scored lowest on information cues (2) and affective cues (3).

For the videotape examination, the four criteria that best predicted outcome were treatment plan (17), summary (11), differential diagnosis (15), and handling of

emergency issues (13). Failing candidates scored lowest on treatment plan (17) and biopsychosocial formulation (16).

Although as of this writing these preliminary findings had not yet been verified by a larger study, decades of observations and anecdotes related by dozens of board examiners have verified their essential truth.

Problems With Content Issues of content can involve virtually any clinical topic covered in this textbook, but candidates typically encounter difficulties with some of the following.

Organization Organizing and verbally presenting a complicated history is a skill that medical students learn early in training. Unaccountably, it also lies at the core of the difficulties many psychiatrists encounter during the board examinations.

> When asked to present the patient “as best you can” from a somewhat murky videotape examination, Dr. D. encountered disaster. Instead of doing a focused presentation, Dr. D. attempted to recount the undigested interview, pretty much as it had just occurred. Jumping around from present illness to historical data to mental status examination, the candidate became more and more anxious, compulsively—and futilely—shuffling through pages of scrawled notes.

Differential Diagnosis For generations the elementary clinical exercise known as the differential diagnosis has been a mainstay of medical decision making, but the stress of the witnessed interview sometimes causes candidates to narrow their focus, rather than broaden it. Some appear to fear they are being tricked or trapped; others cling to a favorite diagnosis as if it were a life preserver. However, examiners only want to find a pattern of sound diagnostic thinking that will serve future patients well.

Knowledge of Criteria Most candidates are reasonably well versed in the latest diagnostic criteria, but sometimes one will mix up items from two or more criteria sets. Some overinterpret the data and make a diagnosis they cannot defend with the material at hand. Examiners usually try to help the candidate work through such a difficulty. However, they will probably show very little tolerance for the occasional candidate who tries to compensate for knowledge defects by inventing criteria. Here, as everywhere else in the oral examination, it is generally much better to recognize and acknowledge a deficiency than run the risk of appearing intellectually dishonest.

> After doing a fine interview, Dr. E. made the flat statement, “The diagnosis is anxiety disorder not otherwise specified,” and then stopped. The combined efforts of two examiners could not persuade this candidate to admit that another diagnosis was possible. The senior examiner noted, “Once many candidates get the patient to report a very well-defined psychiatric syndrome, they tend to forget that there is a differential diagnosis, with several Axes to consider. Even though the issue has been well addressed [in reading material concerning the boards], many people forget.”

Formulation Even if they have done a reasonably competent job of information gathering, some candidates appear puzzled about what to do with it. They seem incapable of condensing what they have learned into the few sentences required.

> After Dr. F. presented the basic observations about a videotape interview, the examiners asked for a formulation of the patient. What they got was an unaltered repetition of what they had just heard from the patient. With prompting, Dr. F. eventually mentioned diagnosis (two Axes only) but seemed to recognize no opportunities for further investigation.
>
> Later that afternoon, after viewing the same videotape, Dr. G. told the same examiners: “In summary, the patient is a middle-aged man, occasionally tearful, who spoke slowly and mentioned other symptoms that would support a diagnosis of depression. I didn't hear anything about previous episodes or manias, so my best Axis I diagnosis would be major depressive disorder, single episode. However, I'd want to quiz him more carefully about mania. I thought he sort of glossed over the question about drinking, so I'd also have to know more before completely ruling out a substance-related mood disorder. I'd defer an Axis II diagnosis—not enough information. Contributing to his depression could be the fact that his mother died several months ago, which I'd mention on Axis IV, as well as the fact that he has been worried about his job because of missing work. That would also be reflected in his Axis V score. I didn't think he looked hypothyroid so I wouldn't mention it on Axis III, but in addition to routine labs I'd order a thyroid screen in my workup, just to be safe. My treatment plan would be biopsychosocial, and I'd consider...”

Further Evaluation Some candidates appear uncertain about how to proceed in a patient workup. Some suggest laboratory tests that are inappropriate; others cannot explain how they would use the information once they had obtained it. Many ignore that most useful (and inexpensive) of procedures, the collateral interview, which is especially vital when patients are psychotic or have cognitive or personality disorders.

Treatment Recommendations The importance of psychopharmacology to the treatment process and its emphasis in training curriculums have ensured that most candidates are reasonably well prepared in this area. But now and again the specter of treatment complications arises to haunt candidates who should have been better prepared to deal with potential complications of a drug they themselves have recommended. For example, unfamiliarity with the symptoms of neuroleptic malignant syndrome is sure-fire trouble for a candidate who has just recommended the use of haloperidol (Haldol).

> Discussing treatment of a depressed, middle-aged man, Dr. H. advocated trazodone (Desyrel) but failed to mention the possibility of priapism. “Try as we might to suggest that a discussion of sexual complications might be in order, we got nothing for our efforts,” said one of Dr. H.'s examiners. “Now, not knowing the side effect of a particular drug isn't usually a calamity, but it is an infamous side effect, and it was the candidate who suggested it in the first place!”

In recent years, both junior and senior examiners have reported feeling enraged when candidates tell them that they “never do psychotherapy.” “Despite all the good advice about addressing all forms of treatment, many candidates' formulations are completely unidimensional,” one examiner reported. “They just kiss off psychotherapy by saying they refer all patients who need it to a social worker or psychologist. They seem to regard this as relieving them of all responsibility for knowing anything about interpersonal aspects of treatment. For some of us, this attitude is a nail in the coffin.”

Fatal Mistakes Most examiners agreed that a single error is unlikely to cause a Part II board failure. Such an error would have to demonstrate an abysmal ignorance of facts or an overwhelming lack of concern for patient well-being; for example, offering a suicidal patient a month's prescription of potentially lethal medication or, for some examiners, not asking a patient who uses drugs about suicide or failing to ask about a history of violence or substance abuse.

The examiners must be confident that the candidate can competently assess a patient for these behaviors crucial to diagnosis, treatment and, indeed, to life itself. Despite the logic of this principle (not to mention the publicity it has received in books, articles, and board preparation courses), some candidacies still founder on a failure to follow up clues the patient may present. These can include hints about suicide or death, tearfulness, abnormally reduced or excited activity level, and references to substance use.

> “I remember one candidate who didn't pick up on what the patient was saying,” an experienced examiner commented. “This patient had made reference to drug abuse and the candidate just didn't get it. My partner and I hated to do it—this was an obviously bright, young psychiatrist embarking on a promising career, who has since become an internationally known leader in psychiatry. But we were forced to issue a failure. The candidate passed the next year, but I still cringe when we run into one another at meetings.”

Relying on a Cookbook Rather than not knowing drugs, some candidates seem acquainted with little else. The emphasis of the boards is biopsychosocial, which means that, candidates must both be prepared to discuss social interventions (placement, self-help groups, family interventions, and the like) and the myriad iterations of psychotherapy and be conversant with the dosages and adverse effects of the latest psychotropic medications.

An unnecessarily uncomfortable half-hour with examiners was the fate of Dr. I., whose patient had a personality disorder with borderline features.

“For 10 minutes all this candidate would talk about was antidepressant medications,” one of the examiners later complained. “Sure, we all saw the depressive symptoms. But we practically had to beat the candidate up to get a treatment approach that was anything but drugs.”

Dr. I. did pass the boards, but according to the examiner it was “a near thing.”

Despite numerous warnings that there are still no pathognomonic signs in psychiatry and that no disorders have only one treatment, some candidates still pursue a single-minded approach to patients and their problems. They may quote the latest “truth” from a journal article and, when asked to explore alternative possibilities, respond with the journal title, page, and year. In the words of one examiner, this is “tantamount to dying with their rights on.”

Problems With the Process Lack of information (as described just above) is serious but to an extent expected; nobody can be expected to know everything. Difficulty with the interpersonal aspects of the examination is another matter, for it suggests something inherent in the approach to people that could negatively affect the candidate's ability to care for patients.

Inadequate Rapport From the opening moments of the patient interview, some candidates show behaviors that suggest an inability to subject their own needs to those of the patient. They ignore the bits of social behavior that should put the patient at ease: making introductions all around, offering a choice of chairs, explaining the purpose of the examination, giving good eye contact, showing empathy for a patient who is in pain. As one examiner put it, “You rarely see a caring candidate who doesn't do well overall in the examination.”

Rapport is especially an issue when a candidate's behavior indicates insufficient concern for the feelings or emotional state of the patient. Some examinees seem incapable of putting themselves in the patient's shoes; a rare candidate may show an almost unbelievably callous disregard for the feelings of a fellow human.

A recent board patient with an advanced case of acquired immune deficiency syndrome (AIDS) had lost his lover and been abandoned by his family. During the interview he admitted that he felt depressed and often thought about death. When asked about the importance of the family's absence, the candidate responded that it made “little difference, because the patient was dying anyway.”

Inadequate Control Although examiners criticize the excessively controlled interview in which a patient is given insufficient scope to discuss problems freely, they far more often find cause to criticize the inadequately controlled, directionless interview.

While the patient talked, Dr. J. spent most of the interview sessions furiously scribbling notes. Actually, the interview started out well enough, with the patient discussing the anxiety symptoms that had brought him to treatment. But soon he wandered off into a discussion of his wife's recent operation, then the problems the family had had with her surgeon's billing practices. The candidate went right along, transcribing his words with passive diligence. Although one of the examiners did finally try to refocus the interview, little change resulted, and Dr. J. failed.

Examiners hope to witness an interview in which all of the important topics are covered and the candidate identifies and follows up on important questions in a way that makes the patient feel party to a conversation rather than the subject of an interrogation.

Dr. K. had to examine a woman who for 18 months had been an inpatient on a state mental hospital ward. Her problems included bulimia nervosa, obsessive-compulsive disorder, major depressive disorder, and multiple substance-related disorders. “And that was just the Axis I stuff,” the candidate reported later.

Without adequate time to evaluate any of these issues fully, Dr. K. spent a few minutes addressing each of them before easing the patient into a discussion of the next problem. “My examiners seemed to recognize that I could get my arms around each of these issues sufficiently to formulate a plan for further evaluation and management. I didn't feel especially good about my interview, but I passed.”

Anxiety Anxiety can affect all aspects of performance during the board examination. Nearly every candidate feels some anxiety, but a significant minority are virtually incapacitated by it. Anxiety reduces some candidates' ability to respond in a warm and sympathetic manner and may drive others to ignore important questions (e.g., hints of suicide or evidence of cognitive disorder) that must be followed up for a passing performance. Still other candidates report that their minds “go blank” when they try to recall the most elementary didactic material concerning diagnosis or management.

Examiners are human and do take into account the anxiety and stress that nearly all candidates experience. The case of Dr. A., cited above, demonstrates how the candid admission of extreme anxiety helped one candidate recover from a potentially grievous error. However, candidates have used a variety of approaches besides candor to deal with their anxiety. These range from behavior modification therapy to, in extreme circumstances, the use of drugs such as propranolol (Inderal).

Differences With Examiners Most examiners seldom note interpersonal clashes with candidates, but it is not unusual for candidates to report feeling that they have in some way run afoul of one or more examiners. In part this may be fed by anxiety; in part it may be a reaction to the neutral stance examiners must take toward examinees. Whatever the cause, some candidates find themselves arguing with examiners. Others pontificate, attempting to demonstrate special expertise or knowledge that they hope will carry the day. Examiners sometimes consider such behavior arrogant.

Long-time examiners report a recent marked change in the knowledge and stated interests of young psychiatrists. Candidates of a generation ago might know too little about psychopharmacology; now, they may know little else. One report from a senior examiner exemplifies many such reports:

We are seeing a change in candidates' attitudes toward patients and toward medicine in general. Maybe this has been caused by managed care, but some of these young doctors seem disengaged—they'd rather not take responsibility for a patient. They seem to feel that clinical concerns are just a part of the portfolio for making the money that they are hoping for.

Other Problems Examiners report a seemingly never-ending supply of anecdotes that attest to candidates' ability to bring trouble down upon themselves. One fussy candidate began the session by insisting upon moving the furniture and changing the lighting, all the while complaining about the examination room's ambiance. These behaviors created almost instant hostility in the patient, who called the candidate “stupid.” Flustered, the candidate inadvertently kicked the patient. The discussion after the patient left the room focused largely on issues of hostility and countertransference. Only after long debate, with active participation of the team leader, was the candidate given a passing grade.

Some candidates have become so attached to their little plastic bottles of purified drinking water that they bring them into the examination room. Such a distraction invites patients to comment or even ask for a drink.

A Sun Belt candidate journeyed north fully togged out for winter, including a pair of large leather mittens. He seemingly had not decided prior to the examination what to do with the mittens, which rested in his lap throughout the interview. The patient's obvious fascination with the mittens so flustered the candidate that he ended up offering them to the patient as a souvenir. The patient accepted, and the candidate spent valuable time justifying this behavior to the examiners, who had not requested an explanation.

Unusually enthusiastic about taking notes, a candidate spent several minutes of her interview time fighting with her pen. The ink simply would not flow and, flushing, she kept mumbling to herself and scratching holes in the paper. Eventually she asked to borrow an examiner's pen and carried on with the examination.

A candidate spent the first part of the interview weeping copiously in front of the patient. After collecting himself, he explained that he usually wore glasses but had chosen to use contacts for this interview.

An examination location in California was not excessively hot, but it was unseasonably humid. A candidate who was wearing a three-piece suit appeared increasingly

uncomfortable and progressively tense. Before making his case presentation, the examiners encouraged him to make himself as comfortable as possible, but he retained his vest and jacket. Perspiring profusely throughout the question-and-answer period, he appeared to consider it a major victory just to get out of the room into fresh air.

EXAMINERS

ABPN examiners possess one qualification in common: each must be board certified in psychiatry. Each team of examiners is headed by one of the eight psychiatrist members of the ABPN. Each of these team leaders chooses 4 senior and 16 primary examiners, so a team comprises a total of 21 board-certified psychiatrists. Other than expenses, examiners are not reimbursed for their time. Participation is generally regarded as an honor that most examiners anticipate from one year to the next

Two primary examiners stay with the candidate for the entire hour (half-hour for the videotape examination). The senior examiner floats, going in and out of the several rooms as time and fancy dictate.

Other than their names, examiners know nothing about the candidates—not where they trained, what kind of work they do, where they are employed nor, perhaps most significantly, whether a candidate has previously taken Part II. If by chance a candidate and examiner are acquainted, the pairings are rearranged so no question of either favoritism or discrimination can arise. As a further safeguard against any appearance of impropriety, no psychiatrist can work as an examiner who during the previous year has participated in any for-profit board review course.

Candidates sometimes complain that an examiner's attitude seems cold or hostile. Examiners are instructed to maintain a professional distance from the examinees, and they are specifically told not to teach or give feedback. But their attitudes are also expected to be respectful, never hostile, sarcastic, or condescending. The need for absolute confidentiality that pervades the entire examination process prompts the instruction to examiners never to discuss the examination or candidates in public—on the street, in hallways, or on elevators.

GRADING PROCESS

For the live patient examination, each primary examiner grades the candidate on five categories (Table 53.2-4):

Table 53.2-4 Performance Guides for Patient Examination

The physician-patient relationship
The conduct of the interview
Organization and presentation of data
Phenomenology, diagnosis, and prognosis
Etiological, pathogenic, and therapeutic issues (which are subdivided into biological, psychological, and social categories)

Only the last three categories are used for the videotape examination.

Immediately after the 60 minutes are up and the candidate has left the room, each primary examiner independently assigns a preliminary grade for each of these five categories. (Because the first two categories can only be scored *Pass* or *Fail;* and the remaining three can be assigned a grade of *Condition*, the first two categories have greater weight than the others.) Using these preliminary grades, each examiner then independently rates the overall performance as Pass, Condition, or Fail. Only at this point do the primary examiners compare notes to reach a consensus grade (also Pass, Condition, or Fail) for this half of the Part II examination. Usually they agree at once; if not, they may engage in vigorous debate, perhaps involving the senior examiner, until they reach consensus.

The final grade for the entire examination is determined that evening, when the eight team leaders meet. A candidate who earns a passing mark on both the live patient interview and the videotape passes the examination and becomes a diplomate of the ABPN. A failing grade on either examination or a conditional grade on both results in failing the entire examination. When a Condition is combined with a Pass, the team leaders compare the grading card notes written by the primary examiners. The candidate will probably pass if the conditional grade was overall not too low and the comments on the grading cards indicate that weak areas in one examination were handled satisfactorily on the other. The live patient examination offers more opportunity for a compensating good performance, so a videotape conditional grade is more likely to result in an overall pass than a conditional grade on the live patient examination.

PREPARING FOR THE BOARDS

Preparation for Part I should include careful reading of a standard American textbook of psychiatry, preferably one that offers a study guide to provide familiarity with the sort of test question that may be asked. Part II preparation should include several practice interviews with patients of the type likely to be found in public institutions, which provide most of the actual board examination patients. In addition, at least one mock board examination should be undertaken.

How much time should candidates spend in preparation? Data for Part I are entirely lacking. A report two decades ago found that the average Part II candidate put in about 120 hours in preparation. Of course, this time is highly variable; for weak candidates, 120 hours is not nearly enough, whereas a strong candidate just out of training may require little preparation at all. Unhappily, all too often the former learn who they are only as a result of taking the examination itself.

Mock Board Examination For a generation, the mock board examination has been standard preparation for Part II. No psychiatric candidate, even one who has just graduated from residency, should consider taking the oral board examination without one.

To be effective, the mock examination should be as close to the real thing as possible. Although the mock candidate would ideally be evaluated by two examiners, time and fiscal restraints often mandate only one. When this is the case, that lone psychiatrist should:

Have served as an official ABPN examiner. That experience will convey the necessary insight into the rules, process, and spirit of the examination process.

Be someone not well known to the candidate. Friends or colleagues may exaggerate the merits and minimize the deficiencies of a candidate's performance.

Be willing to spend the 90 minutes necessary for this exercise, including the hour-long examination itself followed by 30 minutes or so for critique. The critique should include a point-by-point evaluation of the five areas in which candidates are graded as well as an overall score (Pass, Condition, Fail) and comments on every relevant item of content or process raised during the half-hour patient interview.

Provide a patient. For obvious reasons, this patient must be one with whom the candidate has never had prior contact.

Although the outcome of a mock board interview still depends to some extent on the patient's cooperation, any interview carefully witnessed by a thoughtful, impartial observer should prove a pretty good predictor of the course of the actual board examination. The critique, which the candidate might well (with permission of examiners and patient) want to tape record for later study, may reveal difficulties in one or more of three possible areas, any of which should set in motion an immediate plan of action. The sort of problems noted in this critique will closely parallel problems during the actual board examinations discussed in an earlier section.

Inadequacies of Factual Information Lack of facts concerning any modality of treatment or any psychiatric diagnosis should stimulate further review of relevant sections of a standard psychiatric textbook such as this one. Although no one expects candidates to have memorized DSM-IV criteria, a thorough understanding of their structure and basic content is expected. Because Part II candidates have already passed the rigorous Part I evaluation covering multiple content areas (Table 53.2-1), problems with content should be minimal. Indeed, isolated knowledge deficits may not be devastating if candidates can satisfy the examiners about their ability to provide safe psychiatric care.

> After interviewing a young woman with a first-break psychosis, candidate L. admitted to “complete naiveté” regarding the novel antipsychotic agent she was taking. “If I inherited the care of this patient, I'd look that drug up in *AHFS Drug Information* or PDR. And I'd also discuss it with the drug information pharmacist at our medical center.” During the debriefing, the mock examiner specifically praised this cautious, information-seeking approach to uncharted territory.

Conduct of the Interview Almost any mock board interview is likely to yield some suggestions for improvement—more eye contact, fewer interruptions from the candidate, a different focus or depth when following up on earlier questions. Other findings may require much more extensive remediation.

> “I've got to say, I felt really uncomfortable during much of your interview,” the examiner told Dr. M. during the mock board debriefing. “The patient had a story he was trying to tell, but you kept interrupting to ask questions about which medicines he had taken. And some of the important ideas he did manage to get across—his life didn't ‘seem worth living,' for example—you didn't follow up.”

Interpersonal Differences Although board examiners rarely express negative feelings about the psychiatrists they examine, the reverse is not uncommon. These feelings extend even to mock examinations, where the stress of the moment may provoke feelings of distrust or hostility—the examiner might even feel threatened somehow by the candidate! Under stress, candidates are likely to react by becoming defensive and rigid. They may argue with examiners and cling to a position so tenaciously that they cannot acknowledge, much less accept, the fact that an examiner may be trying to nudge them toward an alternative view of the patient. The fact that a given diagnostic or treatment option may be perfectly justifiable and even correct does not excuse any candidate from the need to demonstrate flexible thinking.

A poor performance on a mock board examination should lead to an immediate reexamination with a different examiner. A conflicting opinion could mean that the problem has been solved or that it is confined to certain types of patients (or examiners)—and justifies yet a third mock examination. A second opinion that confirms the first suggests the need for extensive remediation in one or more of the areas just mentioned. Attendance at one of the numerous board review courses (many of them proprietary) offered throughout the country may be indicated. Table 53.2-4 suggests performance standards that apply to the patient examination.

Review Courses How helpful are the seemingly ubiquitous, definitely expensive review courses offered by universities and private organizations? Other than anecdotal evidence, most of which is only of testimonial quality, essentially no data exist for psychiatry. One study in the field of surgery reported that formal review courses may assist the candidate at high risk of board failure, especially those who have already experienced one or more failures.

To be maximally helpful, a review course should have faculty with previous experience as board examiners and give each student the opportunity to participate in a mock board examination. The course should be taken far enough in advance of the actual examination to allow the opportunity to practice new behaviors.

SUBSPECIALTY BOARD EXAMINATIONS

In addition to the board examinations in general psychiatry, the ABPN offers certificates in five subspecialties. The oldest of these, dating to 1959, is the child and adolescent psychiatry examination. It is a full-fledged board in its own right; the others are qualifications added to the basic general psychiatry boards.

Child and Adolescent Psychiatry To sit for the examination in child and adolescent psychiatry an applicant must have had at least 2 years of training in adult and 2 years in child psychiatry, all in the United States, and must already hold the ABPN certificate in general psychiatry. The structure of the examination parallels that for general psychiatry: a qualifying written examination followed by an oral examination that includes a live patient interview.

The written examination covers core knowledge in the field: child and adolescent development, history, psychopathology, and theory. It contains approximately 100 multiple-choice or matching questions and is administered once a year in a 2-hour session as a part of the overall examination.

The oral examination has three segments, each conducted by two examiners. In the preschool and grade school age segment candidates evaluate two patients, one on videotape and the other in a written vignette, and are in turn evaluated on their ability to observe, to describe the child's important behaviors, to determine what additional information is needed, and to formulate a treatment plan.

In the adolescence segment the candidate interviews an adolescent psychiatric patient in the presence of the two examiners. The evaluation is based on the candidate's ability to conduct the interview, obtain important information, formulate a differential diagnosis, and construct a treatment plan.

In the consultation segment the candidate must discuss two written vignettes, each of which describes a situation in which a judge, physician, school, or social agency requests consultation about a child or adolescent. Candidates are rated on their ability to identify problems and needs, formulate goals, and plan interventions. They must also demonstrate familiarity with the important community resources and discuss relevant ethical, interprofessional, and legal issues.

Success on the Child Boards Compared with the adult boards, the success rate for the child boards is high, about 90 percent overall; 95 percent pass the Part I written examination. Because all candidates have first passed the general psychiatry boards, these figures should not be surprising. In fact, it has been argued that certification rates in child psychiatry should be even higher. Those who disagree cite an inability to synthesize and integrate clinical material as the principal cause for failure.

Certificates of Added Qualifications By examination, diplomates in general psychiatry can earn certificates of added qualifications in various subspecialty fields. Added incrementally since the examination for geriatric psychiatry was introduced in 1991, these now include addiction psychiatry, clinical neurophysiology, and forensic psychiatry. Each requires one additional year of specialty training after the fourth postgraduate year (PGY-4). If obtained at a single center, training may be completed on a half-time basis for all subspecialty areas but clinical neurophysiology, which requires full-time training. By the year 2002, all will require that training be obtained in an institution approved by the Accreditation Council for Graduate Medical Examination (ACGME). Each is a written, multiple-choice examination that requires approximately half a day.

Candidates for these certificates are tested in the following areas:

Addiction psychiatry—biological and behavioral basis of practice, evaluation and consultation, laboratory assessment, pharmacology of drugs, pharmacotherapy, and psychosocial treatment

Clinical neurophysiology—autonomic testing, central electroencephalography (EEG) and electromyography (EMG), evoked potentials, instrumentation, physiology, polysomnography, signal analysis, and surgery intensive care unit (ICU) monitoring

Forensic psychiatry—civil and criminal law and landmark cases, legal regulation of psychiatry, and special issues

Geriatric psychiatry—biological aspects of aging, development and psychosocial aspects of aging, diagnostic methods, the interface between general medicine and psychiatry, neuropsychiatry and neurological aspects of aging, psychopathology and psychiatric diagnosis, and treatment

The numbers and success rates of candidates in recent years are given in separate tables in Section 53.1.

OTHER PSYCHIATRY EXAMINATIONS

The ABPN assesses only graduate psychiatrists, not those in training. It forthrightly claims that it has no educational mandate and that it is not intended to be a way of evaluating training strategies. Because of the time and expense of construction, in-house examinations of cognitive material have been virtually eliminated from most training programs.

PRITE Nationally standardized examinations for residents who are still in training were adopted by many specialties during the last third of the twentieth century. The Psychiatry Residency In-Training Examination (PRITE) has been in use constantly since 1982. Although it is voluntary, the Residency Review Committee for psychiatry requires that each program use a formal, standardized test as a part of its evaluation of trainees. With PRITE the only game in town, by the 1990s nearly every training program nationwide had adopted it, and it was being administered to over 5000 residents each year.

This written examination offers residents the opportunity to track their progress throughout training and to gauge their own performance against that of colleagues. To accommodate the schedules of individual training programs, the PRITE can be administered on any 2 half-days within a 1-week window during the fall of each year; the window reduces the likelihood that questions will be circulated by the first examinees to take it.

The examination is administered with the signed understanding that scores will not be used to decide whether a resident will be dropped from a training program. However, many training directors use the results for individual counseling and for assessing the effectiveness of their programs in teaching the cognitive aspects of psychiatry.

Comparative percentile scores are reported for both the psychiatry and the neurology components of this examination. For the overall test and for each subsection, training directors and each trainee receive the individual's exact numerical scores, ranking compared with peers in that program, and nationwide peer group percentile rank. Training directors learn the grouped percentile ranks on each subtest for the residents in each training year as well as how their residents compare nationally on each question.

Resident input into test item development has been encouraged, causing some feeling that those who so participate reap a competitive advantage for their efforts. But in the words of one participant, “I submitted my questions so long before they were scheduled for use that by the time of the exam, I couldn't remember what I had written!”

Roughly 18 percent of the PRITE emphasis is on neurology and the neurosciences; the remainder is on 10 principal psychiatric knowledge areas: adult psychopathology, alcoholism and substance abuse, behavioral science, child psychiatry, consultation and liaison psychiatry, emergency psychiatry, growth and development, patient evaluation, psychosocial therapies and somatic treatment methods. The sponsors of this examination include the American College of Psychiatrists, American Academy of Child and Adolescent Psychiatry, American Association of Directors of Psychiatric Residency Training, and the Association for Academic Psychiatry.

In contrast to the board examinations, the goal of the PRITE is truly to educate. Not only do trainees receive exact numerical scores referenced to all examinees in that individual's year of training, they also receive copies of the questions and answers, complete with references for further study. Of course, this mandates yearly rewriting of the examination itself, which retains only enough questions from the previous year to allow comparison with examinations given in the years before. Development and administration costs of the examination are borne by the training programs, which pay a flat fee of $80 per resident tested per year. Residents invest only their time.

Some residency training directors report that extreme scores on the PRITE strongly predict outcome of the ABPN Part I examination: “I've hardly ever seen a resident who does extremely well—above the 90th percentile—on the PRITE who has any trouble at all with the boards,” one training director reported. “They hardly even need to study. But I counsel graduating residents who score below 50 percent that they'd better take at least one board review course.”

Since 1996 there has also been a child PRITE, a 50-item test whose purpose and structure are essentially the same as those for the general PRITE examination.

Multiple-Station Examination Format A form of the multiple-station examination, or practical examinations, used for decades in basic science courses, the objective structured clinical examination (OSCE), sparingly reported in the past few years, is a novel attempt to use objective criteria to evaluate trainees' practical clinical skills. Typically, several stations are set up, each involving a clinical problem: a videotaped patient's interview to diagnose, laboratory values to interpret, or an emergency scenario to evaluate. Test results are scored according to criteria the faculty determines in advance. Advantages are that a number of trainees can be evaluated within a short time, evaluations are objective, the material tested is relevant to actual clinical situations, and, in terms of both time and money, the examinations are relatively inexpensive to set up and administer.

MEASURING THE SUCCESS OF EXAMINATIONS

Does interview control really equate to safe practice of psychiatry? How well do the boards (or any other cognitive examination) test what clinical psychiatrists need to know? An analogous question might be, Does success in the 100-yard dash predict a marathon victory?

Over the years, a number of criticisms have been leveled at the board's process. For example, after seven decades there still is no attempt to link residents' performance in training with success on the boards. Eligibility doesn't require training directors to certify that graduates possess skill, knowledge, ethics, or a patient-focused attitude—surviving 4 years of training suffices. And, although it is a principal sponsor of the ABPN, the American Psychiatric Association does not require board certification for full membership. Some writers have alleged that this is because the organization might lose members, money, and political influence; however, no evidence supporting such an accusation has ever been adduced.

Over the years the question of what the board examination predicts has generated far more heat than light. Neither of two important qualities, validity and reliability, has been adequately demonstrated for the psychiatry boards.

Reliability In this context reliability means consistency in a test-retest situation. Although some information about examination reliability is available, much of it was published over a quarter-century ago. ABPN Part I reliability has been rated high on the basis of having a small standard error of the mean (2 to 4 percent); a narrow interval increases the confidence that the examination results are accurate.

Two aspects of reliability as it relates to Part II of the board examination have been addressed: agreement of the two primary examiners and agreement between the videotape and live patient examinations.

Examiner Agreement A 1991 study examined interexaminer consistency for 1422 candidates to evaluate the reliability of the Part II examination. In 67 percent of examinations there was perfect agreement about the candidate's performance. That is, without discussion the two examiners initially each gave the candidate the same Pass, Condition, or Fail grade on either the patient interview or the audiovisual examination. For 26 percent of candidates, there were minor disagreements (i.e., a Condition coupled with either a Pass or a Fail); for only 7 percent of candidates were there major disagreements (a Fail coupled with a Pass). The overall weighted kappa (k) statistic was 0.56, which was judged to show “fair to good agreement beyond chance.” When examiners disagreed, discussion (sometimes involving the senior examiner) usually produced consensus with relative ease.

A study of examination stability drawn from the same data found that in the 7 percent that involved major disagreement (Pass-Fail) between examiners, a thorough review of the candidate's entire performance was usually necessary to reach consensus. Most of these major disagreements turned on the concern of one examiner about something specific the candidate had done or not done. Reporting that the examination process was stable across seasons of the year and different

examination sites, these authors found “no evidence that factors such as examiner fatigue to increasing experience influence interexaminer consistency.”

In a 1990 assessment of the reliability of evaluative judgment, 94 medical school psychiatry department faculty members rated a videotape portraying the first 210 seconds of three separate interviews. The psychiatrists were asked to rate the interviews using a data checklist, a global rating scale, and an estimate of time allotted to various interviewing activities. Only a low level of agreement among raters was found for any of these methods, and this performance did not improve subsequently when videotape segments of twice the length were used.

Finally, in the pilot study that used 18 explicit grading criteria, agreement between raters on both the audiovisual and live patient examinations yielded a broad range of intraclass correlation coefficients (0.37 to 0.77). As of this writing, those authors had not further examined what could indicate a significant opportunity to improve reliability.

Videotape-Patient Agreement Most examineees agree that the videotape examination is a useful learning and testing device. This assessment is clearly endorsed by the ABPN. Several factors are relevant to its utility. It increases the efficiency of the examination by using only half the number of patients that would be needed if every candidate interviewed two patients. Some candidates (and examiners) may find it easier to discuss a videotape, where the data are already provided, than an actual patient, where data collection depends on the vagaries of the clinical interview. In any given half-day, all candidates see the same interview tape, which permits a more uniform, hence more reliable, examination. However, the videotape examination offers no opportunity to evaluate interaction with patients, which probably reduces its overall value as a measurement of clinical competence.

In the early 1980s John Talbott reported data concerning the concordance rate between the live patient and videotape sessions. He defined *nonconcordance* as a Pass coupled with either a Fail or a Condition. *Concordance* was either identical scores on both tests or a Fail coupled with a Condition. (The latter results in an overall failing grade. Because a Pass coupled with a Condition usually results in an overall Pass, this definition yields a conservative estimate of concordance.) Of the 2236 candidates studied, 52 percent had identical grades and 61 percent were concordant, an acceptable rate considering that the two formats test somewhat different skills and carry different weights in the judgment of many examiners. Of the 869 nonconcordant candidates, 303 (35 percent) were highly discordant (they received a Pass coupled with a Fail). A candidate with dissimilar scores was twice as likely to score higher on the videotape session. Possible explanations for these findings: examiners may grade the videotape examination more liberally, residency programs may better prepare their graduates to do well in the cognitive aspects of psychiatry than in the performance skills of patient interviewing, or it may be easier to discuss a patient for whom the data are already provided.

Talbott summarized the distinctions between the two examination formats by noting that the patient interview tests whether “the candidate can set the patient at ease, establish good therapeutic contact, maintain interpersonal relatedness, elicit pertinent information, monitor countertransferential reactions, ask direct and open-ended as well as sympathetic and productive questions, and shift modes as warranted in the interview.” The videotape interview “can obviously ascertain none of these, but instead should concentrate directly on the cognitive aspects of the clinical problems presented (e.g., the mental status, diagnosis, and therapy).”

Limits of Reliability The great strength of the ABPN examination continues to be its weakness: the subjectivity of Part II. Considerable effort has been expended to make it fair and, within the limits of the genre, standardized. Yet from time to time, stories surface that suggest the fragility of its impartiality. These stories do not come solely from erstwhile candidates.

The day before each oral examination, all new examiners and those who have not examined recently are required to attend a half-day course, one purpose of which is to ensure uniform scoring. In one exercise, the novice examiners were divided into groups to view and discuss a videotape of a candidate interviewing a difficult, manic-appearing patient. One group unanimously agreed to pass the candidate. An examiner from that group was later astonished when a friend who had viewed the same video with a different group said, “We all agreed that that candidate should fail.”

Of course, no one can expect all examiners to agree about candidates who appear marginal. But the anecdote suggests how much the power of group thinking can influence even board diplomates who are trying their best to render objective, independent judgments about their fellow psychiatrists. Indeed, for all the efforts to achieve validity, the weight of published evidence has so far demonstrated validity for the board process that, while definite, is modest.

Validity *Validity* is defined here as the ability of an examination to yield an accurate measure of a psychiatrist's clinical adequacy. Several types of validity that have been described are relevant to these examinations, including content, construct, and criterion-based validity.

Content Validity Does the examination evaluate material that experts consider vital for assessing psychiatric capability? The countless hours contributed by many academicians to writing questions would seem to defend a positive answer to this question. In a 1982 survey of 482 psychiatric clinicians and academicians, 92 to 99 percent agreed on which 10 skills were most important for a clinical psychiatrist. In descending order, these were conducting a comprehensive interview, evaluating the need for hospitalization, demonstrating reliability, conscientiousness and integrity, accurate diagnosis, formulating and implementing a treatment plan, assessing potential for suicide and violence, providing supportive psychotherapy, medication use, compassion for patient and family, and recognition of countertransference problems. Although this is exactly the sort of information sought in Part II of the boards, the concept of content validity carries with it the taint of circular reasoning.

Construct Validity A test with construct validity actually measures what it claims to measure. For example, if an examination measures competence in a medical specialty, then test performance should correlate positively with amount of training. This was indeed the finding in a study conducted in 1982 by the American Board of Emergency Medicine. Similar findings are reported yearly for the PRITE. The fact that first-time examinees are far more likely than repeaters to pass either part of the ABPN can be taken to indicate that examination's construct validity.

Criterion Validity In criterion validity a new test is validated against an established measure of supposed great accuracy. When the established measure is made at a future time, as in a follow-up study, it is called *predictive validity*.

A 1975 study of 200 candidates evaluated the predictive validity of Part I. Of candidates who passed Part I on the first try, 88 percent ultimately became board certified, compared with ultimate success by only 12 percent of those who failed their first attempt at Part I. Only about half of all psychiatry candidates passed both parts of the boards on the first try, eliciting this comment from the authors: “This may mean that the exams are deficient, that candidates are insufficiently prepared, or that the exams, despite an admitted need for further improvement, do serve to maintain essential professional standards by not permitting overly easy access to specialty status.”

Comparison of the test scores of 701 physicians who had taken both the 1992 PRITE and 1994 ABPN Part I examination revealed a Pearson correlation of 0.67 between the psychiatry portions of the two examinations. Although the authors considered this correlation “moderate,” one wonders why it wasn't higher. After all, the two examinations used similar methods to test supposedly similar aspects of cognitive learning.

Other Correlations In 1985 Donald Langsley and John Lloyd found a low correlation between performance on Part I and the ratings of training directors, and a 1977 study in Canada failed to show any correlation between a psychiatry in-training evaluation report and other measures of candidate performance such as multiple-choice questions and oral examinations. As with most other cognitive examinations in medicine, extremely limited data validate board diplomates' ability to render care that is better than that given by nondiplomates. The only such mental health study yet reported compared psychotherapy outcome in nine patients with the following evaluations of their therapists: didactic examination, global ratings by trainers, supervisors' ratings based on therapists' recollections of sessions, self-ratings, and independent evaluators' ratings of videotaped sessions. Patient outcome only correlated with the supervisors' ratings, which (as the authors acknowledged) could have been influenced by knowledge of that outcome. This halo effect was subsequently diminished when, after viewing videotapes of their trainees' actual therapy sessions, supervisors revised their overall evaluations downward. It appeared that trainees' personality characteristics and ability to discuss theoretical material could significantly enhance supervisors' opinions of them.

Nonmental health studies also suggest the validity of board certification: (1) lower mortality rates for Medicare populations were found in hospitals that had the highest proportion of board-certified physicians: (2) board-certified surgeons had lower morbidity and mortality rates than surgeons who were not board certified: (3) board-certified emergency specialists performed better in an examination in their specialty: (4) internists who were board certified received higher peer ratings than their colleagues without board status, which may be nothing more than halo effect: and (5) ethics and morals are hard enough to define, let alone measure: However, one study evaluated the possible effect of board certification for physicians disciplined by the Medical Board of the State of California. Board certification was significantly less common ($P < .001$) among physicians against whom any disciplinary measure had been taken (57 percent) than among a comparison group of physicians who had not been disciplined in any way (70 percent). Although the trend for psychiatrists was in the same direction, the difference was not significant (59 vs. 63 percent, P = NS). In this study, physicians who identified their main area of specialization as psychiatry constituted 14 percent of the disciplined physicians,

compared with only 6 percent of the comparison group. Another study found that board certification did not predict medical board discipline for sex-related offenses.

The ability of ABPN certification in general psychiatry to signify quality mental health delivery is often regarded almost as a tenet of faith. With much to recommend it in terms of logic, it remains largely unproved. One review of the evaluation of trainees' problem-solving skills concluded that observation of a single patient-provider encounter was unlikely to predict accurately how that trainee would perform in a different clinical situation. Currently, few data would suggest a more optimistic conclusion about the ability of any examination now employed by psychiatry training programs.

RECERTIFICATION AND THE FUTURE

When the American Board of Family Practice sponsored its first examination in 1969, it issued certificates good for 7 years. The ABFP was only the first; increasingly, time-limited certification is now the norm. Since 1994 the ABPN certificate issued in general psychiatry has been good for 10 years.

The recertification examination in psychiatry is planned as a voluntary, user-friendly, open-book examination to be taken at home. Such a structure, far simpler than the original certifying examination, should boast an extremely high rate of candidate success. Indeed, this must be so; consider the potential alternative, a nightmare of many once-certified psychiatrists who must announce themselves decertified. Who would deny them the right to claim only that they had passed the ABPN some years ago? And what would be the status of such an individual's added qualifications in forensics, geriatrics, or substance abuse psychiatry?

Ideally, certification or recertification for any profession as complicated and high-risk as the care of psychiatric patients requires the evaluation of many qualities. These span a range from highly objective to the most subjective, including knowledge, judgment, skills, and attitudes. Only the first of these can be readily assessed by written examination; the second and third can be evaluated (for a given patient) by observing the candidate conduct an interview with an actual patient, then discuss and defend a formulation and treatment plan for that patient. Although examiners may obtain some indication of attitudes from either a written or oral examination, they will probably learn only as much as the candidate is willing to reveal. What is revealed may apply only to a particular patient or on a given day.

From the experiences and data so far available about the testing of psychiatrists, it appears that whatever methods of testing are used, they present only an imperfect representation of the true qualities any candidate possesses. Because training programs, professional organizations, and examining boards deal with the same relatively small group of individuals, efforts at close collaboration should increasingly be undertaken to meld multiple evaluations of trainees' competence, including residency directors' evaluations, in-residence training examination scores, and written and oral performance on the boards.

SUGGESTED CROSS-REFERENCES

Section 53.1 contains information about the history and current operations of the American Board of Psychiatry and Neurology. The standard psychiatric history and mental status examination are covered in Section 7.1 and Section 7.2. A general discussion of reliability and validity can be found in Section 5.2.

SECTION REFERENCES

Borus JF, Yager J: Ongoing evaluation in psychiatry: The first step toward quality. Am J Psychiatry *143:*1415, 1986.

*Case SM, Swanson DB: *Constructing Written Test Questions for the Basic and Clinical Sciences*. National Board of Medical Examiners, Philadelphia, 1996.

Chevron ED, Rounsaville BJ: Evaluating the clinical skills of psychotherapists. Arch Gen Psychiatry *40:*1129, 1983.

Dehlendorf CE, Wolfe SM: Physicians disciplined for sex-related offenses. JAMA *279:*1883, 1998.

Fox G, Herzog DB: Resolved: The pass rate for the child and adolescent psychiatry board examination should be higher than it is present. J Am Acad Child Adolesc Psychiatry *30:*685, 1991.

Greenblatt M, Carew J, Pierce CM: Success rates in psychiatry and neurology certification examinations. Am J Psychiatry *134:*1259, 1977.

Hartz AJ, Krakauer H, Kuhn EM, Young M, Jacobsen SJ, Greer G, Nuenz L, Katzoff M, Bailey RC, Rimm AA: Hospital characteristics and mortality rates. N Engl J Med *321:*1720, 1989.

Hodges B, Regehr G, Hanson M, McNaughton N: Validation of an objective structured clinical examination in psychiatry. Acad Med *73:*910, 1998.

Juul D, Martin MJ, Scheiber SC: The examination for added qualifications in geriatric psychiatry. Am J Geriatr Psychiatry *4:*24, 1996.

Juul D, Tucker GJ: The part I psychiatry examination: Facts about the written examination. In *Certification, Recertification and Lifetime Learning in Psychiatry*, JH Shore, SC Scheiber, editors. American Psychiatric Press, Washington, DC, 1994.

Kelly JV, Hellinger FJ: Physician and hospital factors associated with mortality of surgical patients. Med Care *24:*785, 1986.

Langsley DG: Certification in psychiatry and neurology: Past, present, and future. In *Certification, Recertification and Lifetime Learning in Psychiatry*, JH Shore, SC Scheiber, editors. American Psychiatric Press, Washington, DC, 1994.

*Loschen EL: Using the objective structured clinical examination in a psychiatry residency. Acad Psychiatry *17:*95, 1993.

*McDermott JF, Streltzer J, Lum KY, Nordquist CR, Danko G: Pilot study of explicit grading criteria in the American Board of Psychiatry and Neurology part II examination. Am J Psychiatry *153:*1097, 1996.

McDermott JF, Tanguay PE, Scheiber SC, Juul D, Shore JH, Tucker GJ, McCurdy L, Terr LC: Reliability of the Part II board certification examination in psychiatry: Interexaminer consistency. Am J Psychiatry *148:*1672, 1991.

McDermott JF, Tanguay PE, Scheiber SC, Juul D, Shore JH, Tucker GJ, McCurdy, L, Terr LC: Reliability of the Part II board certification examination in psychiatry: Examination stability. Am J Psychiatry *150:*1077, 1993.

McGuire CH: Medical problem-solving: A critique of the literature. Research in Medical Education: 1984 proceedings of the 23rd Annual Conference, Washington, DC, 1984.

Morrison J, Wickersham P: Physicians disciplined by a state medical board. JAMA *279:*1889, 1998.

*Morrison J, Muñoz RA: *Boarding Time*, ed 2. American Psychiatric Press, Washington, DC, 1996.

Ramsey P, Carline J, Inui T: Predictive validity of certification by the American Board of Internal Medicine. Ann Intern Med *110:*719, 1989.

Shapiro T, Juul D, Scheiber SC: Exploration of failure on the subspecialty examination for child and adolescent psychiatry. Am J Psychiatry *153:*693, 1996.

*Shore JH, Scheiber SC, editors: *Certification, Recertification and Lifetime Learning in Psychiatry*. American Psychiatric Press, Washington, DC, 1994.

Small SM: Limitations and values of evaluation techniques in psychiatric education. Am J Psychiatry *132:*52, 1975.

Smeltzer DJ, Jones BA: Reliability and validity of the psychiatry resident in-training examination. Acad Psychiatry *14:*115, 1990.

Strauss GD, Yager J, Liston EH, Strauss GE: A comparison of national and in-house examinations of psychiatric knowledge. Am J Psychiatry *138:*636, 1981.

Talbott JA: Is the “live patient” interview on the boards necessary? Am J Psychiatry *140:*890, 1983.

Tanguay PE, McDermott JF, Philips I: A study of the board certification examination in child and adolescent psychiatry. J Am Acad Child Adolesc Psychiatry *29:*821, 1990.

Templeton B, Allen MM: Interrater reliability in evaluating trainee interviewing skills. Acad Psychiatry *14:*188, 1990.

Wade TP, Kaminski DL: Comparative evaluation of educational methods in surgical resident education. Arch Surg *130:*83, 1995.

*Warnock JK, Nelson DA: Training psychiatric residents for Part II of the American Board of Psychiatry and Neurology Examination. Bull Menninger Clin *63:*111, 1999.

Webb LC, Juul D, Reynolds CF III, Ruiz B, Ruiz P, Scheiber SC, Scully J: How well does the psychiatry residency in-training examination predict performance n the American Board of Psychiatry and Neurology Part I examination? Am J Psychiatry *153:*831, 1996.

CHAPTER 54. ETHICS AND FORENSIC PSYCHIATRY

54.1 LEGAL ISSUES IN PSYCHIATRY

ROBERT I. SIMON, M.D.

Four overlapping spheres of influence affect psychiatric practice: (1) the psychiatrist's professional, moral, ethical, and legal duties to provide competent care to patients; (2) patients' right of self-determination to receive or reject care; (3) the decisions and directives of courts, legislatures, governmental regulatory agencies, and boards of licensure; and (4) the ethical codes and practices of professional organizations. Psychiatrists must comfortably manage the clinical-legal tensions and conflicts that inevitably arise in everyday practice to practice competently.

In many clinical situations, legal requirements do not necessarily conflict with the provision of good clinical care. Whenever possible, the psychiatrist should attempt to incorporate a legal requirement into the clinical intervention. For example, if a patient is threatening harm to another person, can the psychiatrist enlist the patient in the process of warning and protecting the endangered third party? If not, can the psychiatrist use effective clinical options rather than relying solely upon warning?

Clinical risk management combines the psychiatrist's expertise, a thorough knowledge of the patient, and a clinically useful understanding of the legal requirements governing clinical practice to provide competent clinical care to the patient and (only secondarily) to reduce malpractice risk. For example, keeping adequate patient records is good clinical practice; also, courts generally take the position that what is not recorded is not done. Moreover, without a record, it is much easier for the plaintiff to win a malpractice suit, despite the psychiatrist's provision of good clinical care. For this reason, some liability insurers will not insure psychiatrists who do not keep outpatient records. Ignorance of the law, as it pertains to everyday clinical practice, can unnecessarily limit the psychiatrist's treatment options and lead to destructive defensive practices.

Paradoxically, defensive psychiatry may result in provision of substandard treatment of patients that increases the risk of a malpractice suit. *Defensive psychiatry* refers to any acts or omissions by a psychiatrist that are performed not for the benefit of the patient but rather to avoid malpractice liability or to provide a legal defense against a malpractice claim. Typical examples of defensive psychiatry include unnecessarily hospitalizing a suicidal patient who could be safely treated as an outpatient or refusing to use a neuroleptic such as clozapine (Clozaril) for a chronic, treatment-resistant schizophrenia patient with prominent negative symptoms because of rare reports of agranulocytosis. Such defensive practices invariably harm the therapeutic alliance and may lead to provision of substandard care.

PSYCHIATRIC MALPRACTICE

Psychiatric malpractice is medical malpractice. Medical malpractice is a tort, or civil wrong. It is a noncriminal, noncontract wrong resulting from a physician's negligence that causes injury to a patient under his or her care. Simply put, negligence means doing something that a physician with a duty to care for the patient should not have done, or failing to do something that should have been done as defined by current medical practice. Usually, the standard of care in malpractice cases is established by expert witnesses. The standard of care is also determined by reference to journal articles, professional textbooks and treatises, professional practice guidelines, and ethical practices promulgated by professional organizations.

To prove malpractice, the plaintiff (e.g., patient, family, or estate) must establish by a preponderance of the evidence that (1) a doctor-patient relationship existed that created a *duty* of care, (2) there was a *deviation* from the standard of care, (3) the patient was *damaged*, and (4) the deviation *directly* caused the damage.

These elements of a malpractice claim are sometimes referred to as the 4 Ds (duty, deviation, damage, direct causation). Proof by a preponderance of the evidence as required in a malpractice suit means simply more likely than not. Although the law does not assign a percentage, a preponderance of the evidence is akin to 51 to 49 percent, or just enough evidence to tip the scale one way or the another.

Each of the four elements of a malpractice claim must be present or there can be no finding of liability. For example, a psychiatrist whose negligence is the direct cause of harm to an individual (physical, psychological, or both) is not liable for malpractice if no doctor-patient relationship existed to create a duty of care. Psychiatrists are not likely to be successfully sued if they give negligent advice on a radio program that is harmful to a caller, particularly if a caveat was given to the caller that no doctor-patient relationship was being created. No malpractice claim will be sustained against a psychiatrist if a patient's worsening condition is unrelated to negligent care. Finally, if a psychiatrist treats a patient who is then harmed, no malpractice exists if the psychiatrist did not deviate from the standard of care. In the clozapine example above, proper patient selection, appropriate drug administration, and careful monitoring of the white blood count would likely defeat a malpractice claim if the patient developed agranulocytosis and died.

Not every bad outcome is the result of negligence. Psychiatrists cannot guarantee correct diagnoses and treatments. When the psychiatrist provides due care, mistakes may be made without necessarily incurring liability. Most psychiatric cases are complicated. Psychiatrists make judgment calls when selecting a particular treatment course among the many options that may exist. In hindsight, the decision may prove to be wrong but not a deviation in the standard of care.

This knowledge creates a powerful risk management tool, the risk or benefit note. Recording risk and benefit notes in the patient's chart that carefully describe the rationale for treatments and procedures limits liability exposure. Current clinical judgments that may prove to be wrong later cannot be shown to be negligent when the psychiatrist used reasonable care and documented the reasoning. The law is not interested in punishing errors in judgment per se.

In addition to negligence suits, psychiatrists can be sued for the intentional torts of assault, battery, false imprisonment, defamation, fraud or misrepresentation, invasion of privacy, and intentional infliction of emotional distress. In an intentional tort, wrongdoers are motivated by the intent to harm another person or realize or should have realized that such harm is likely to follow results from their actions. For example, telling a patient that sex with the therapist is therapeutic perpetrates a fraud. Most malpractice policies do not provide coverage for intentional torts. Other legal theories of liability include breach of contract and civil rights violations of the United States Constitution, state constitutions, or federal civil rights statutes.

Malpractice Claims Malpractice claims usually arise when bad outcomes are combined with bad feelings. A good doctor-patient relationship is the most effective natural defense against a lawsuit.

Managed care has transformed the relationship between psychiatrist and patient. Psychiatrists are treating more patients for shorter periods of time. Much less time is available to develop a therapeutic alliance with the patient. Split treatments are common, in which the psychiatrist provides medication and a nonmedical therapist conducts the psychotherapy. The psychiatrist may bear the liability burden in a split-treatment situation.

Psychiatrists are treating more chronic, severely ill patients, frequently as outpatients. Under managed care, the restrictive criteria for psychiatric hospitalization approach the substantive standards for involuntary commitment—severe mental illness and dangerousness. Psychiatrists with large practices and psychiatrists who practice at a number of locations are at higher risk of suit. Supervision of other nonmedical mental health professionals also increases the risk of being sued. Psychiatrists are specializing in geriatric psychopharmacology, dissociative disorders, adolescent addiction medicine, pain management, and adult children of alcoholics, resulting in concentration of risk not usually present in a general psychiatric practice. Some psychiatrists are taking on primary care roles such as managing patients' hypertension and diabetes as well as a variety of acute general medical conditions. The changing health care marketplace also leaves psychiatrists more vulnerable to suits as their relationships with patients are undermined. Many opportunities for bad feelings and bad outcomes exist, including poor

communication, a perceived lack of caring or interest, unavailability during critical events, and perceived unresponsiveness to a patient's particular treatment needs.

The recent malpractice claims experience of the Psychiatrist's Purchasing Group, the liability insurer of members of the American Psychiatric Association, reveals the following approximate frequencies of alleged claims by percentage:

1. Incorrect treatment—33 percent
2. Attempted or completed suicide—20 percent
3. Incorrect diagnosis—11 percent
4. Improper supervision—7 percent
5. Medication error or drug reaction—7 percent
6. Improper commitment—5 percent
7. Breach of confidentiality—4 percent
8. Unnecessary hospitalization—4 percent
9. Undue familiarity—3 percent
10. Libel or slander—2 percent
11. Other (e.g., abandonment, electroconvulsive therapy [ECT], third party injury)—4 percent

The pace of suits alleging the implantation and recovery of false memories of childhood sexual abuse by therapists is accelerating. Borderline patients pose great clinical challenges and associated liability risks. For instance, suicide, violence toward others, the capacity to induce therapist boundary violations, alcohol and drug abuse, impulsive actions, comorbidity, and unstable transferences combined with a tenuous therapeutic alliance all contribute to the increase risk of a bad outcome when treating the borderline patient.

Pursuing a successful claim against a psychiatrist is usually difficult. The plaintiff must prove by a preponderance of the evidence that all four elements necessary to bring a malpractice suit are present. A draw means the plaintiff losses in court. It is comforting that only about 20 to 30 percent of cases brought against psychiatrists are successful.

National Practitioner Data Bank On September 1, 1990, the Health Care Quality Improvement Act of 1986 established the National Practitioner Data Bank. The data bank tracks disciplinary actions, malpractice judgments, and settlements against physicians, dentists, and other health care professionals.

Hospitals, health maintenance organizations, professional societies, state medical boards, and other health care organizations are required to report any disciplinary action taken against providers that lasts more than 30 days. Disciplinary actions include limitation, suspension, or revocation of privileges or professional society membership. Under the Health Care Quality Improvement Act, health care entities and providers are granted immunity from liability when making good-faith peer review reports.

Hospitals must request information from the data bank regarding all physicians applying for staff privileges. Every 2 years, a check of the data bank is required for each physician or other practitioner on the hospital staff. Hospitals that do not comply with these requirements face loss of immunity for professional peer-review activities.

The public does not have access to the data bank. Plaintiffs' attorneys can have access to the data bank only if they can prove that the hospital failed to check the data bank regarding the physician in question. The information obtained can be used only to sue the hospital for negligent credentialing. Physicians can request information from the data bank about their own file at a nominal cost.

Somatic Therapies—Standard of Care The therapeutic use of somatic therapies, including ECT, is evaluated no differently than any other medical or psychiatric procedure in malpractice litigation. The same general standard of ordinary and reasonable care governs whether a psychiatrist's use or failure to use a somatic intervention deviated from the accepted standard of care.

Within the psychiatric profession, no absolute standard protocol exists for the administration of psychotropic medication or ECT. However, clinicians should give careful consideration to the existence of procedures, clinical resources, and practice guidelines, regularly accepted or used by a significant percentage of psychiatrists. For example, the American Psychiatric Association (APA) has published comprehensive findings as task force reports concerning electroconvulsive therapy (1990) and tardive dyskinesia (1992).

Less professional autonomy and flexibility is associated with the use of ECT, which appears to be more rigidly regulated than most other psychiatric treatments. The Joint Commission on Accreditation of Health Care Organizations (JCAHO) considers ECT a *special treatment* procedure, which requires hospitals to have written, informed consent policies concerning its use. In addition, state statutory or regulatory standards, coupled with any specific regulations of a psychiatric facility regarding ECT, could establish a basis for liability, if violated. Nevertheless, no official guidelines should be interpreted as a substitute for sound clinical judgment.

The standard for judging the use and administration of medication, however, is consistent with the more flexible and general “reasonable care” requirement. Use of the *Physicians' Desk Reference* (PDR) to establish or dispute a psychiatrist's pharmacotherapy procedures is a case in point. The PDR is a commercially distributed, privately published reference of medication products used in the United States. The Food and Drug Administration (FDA) requires that drug manufacturers report their official package inserts in the PDR. Although a number of courts have cited the PDR as a credible source of information about medication usage by the medical profession, it does not, by itself, establish the standard of care. Instead, the PDR can be used as one piece of evidence to establish the standard of care in a particular situation.

Most courts now follow the ruling in *Ramon v Farr*, holding that drug inserts alone do not set the standard of care. They are only one factor to be considered, in addition to previous personal experience, the scientific literature, expert testimony, approvals in other countries, and other pertinent factors. The FDA, the PDR, or any other reference source cannot substitute for the psychiatrist's clinical judgment. A substantial body of scientific literature that justifies the clinician's treatment is vastly more persuasive. Similarly, in managed care settings, psychiatrists must vigorously resist efforts to restrict their choice of drugs by predetermined, limited formularies, nor should the choice of generic or proprietary drugs or drug dosages be dictated by others. The treating psychiatrist must determine the specific drug to be prescribed according to the clinical needs of the patient.

Courts recognize the importance of professional judgment and allow psychiatrists and other medical specialists wide latitude in explaining any special diagnostic or treatment considerations that guided their decision making. For example, research in the pharmacological treatment of aggression shows a variety of potentially useful drug therapies—some considered experimental or cutting edge. However, no drug is specifically approved by the FDA for the treatment of aggression.

The courts will consider the fact that rarely does only one treatment choice exist. Moreover, evidence that a treatment procedure is used by at least a respectable minority of professionals in the field could help establish it as a reasonable professional practice. Practice guidelines recommend a variety of treatment approaches but defer to the clinical judgment of the practitioner about the ultimate treatment decision.

The standard of care associated with the use of a somatic therapy to treat a psychiatric patient, at a minimum, should include some variation of the following:

A. **Before treatment**
    1. Complete clinical history (medical, psychiatric)
    2. Current physical examination, when indicated, (performed by the patient's physician or the psychiatrist)
    3. Ordering of necessary laboratory tests and review of all test results
    4. Disclosure of sufficient information to obtain informed consent
    5. Thorough documentation of all decisions, informed consent, procedures, and treatments
B. **During treatment**
    1. Careful monitoring of the patient's response to treatment, including appropriate follow-up and laboratory testing
    2. Prompt adjustments in medications or treatment strategies, as needed
    3. Obtaining a renewed informed consent when treatment is substantially altered or a new treatment is initiated.

Depending on the patient, different treatment interventions require different preparations and precautions. Final treatment decisions rest upon the clinician's medical judgment and a working knowledge of legal issues governing informed consent and patients' rights to refuse treatment.

Somatic Therapies—Legal Liabilities Although no thoroughly reliable data on malpractice claims have been published, lawsuits involving medications appear to constitute a significant share of the litigation filed against psychiatrists. The previously cited malpractice claims experience of the APA reveals that medication error and drug reactions accounted for 7 percent of all claims. Incorrect treatment was the largest malpractice claim category (33 percent). Most malpractice cases fall into one of four categories: suicide, medications, sex, and what some defense attorneys cynically described as “flavor of the month.”

A variety of omissions and poor drug treatment practices commonly result in malpractice actions against psychiatrists or other physicians. The discussion below provides ways to identify problem areas associated with medication treatment.

Failure to Evaluate Properly Sound clinical practice requires that the patient be properly evaluated before any somatic treatment is started. At the very minimum, the psychiatrist should obtain a clinical history and perform a mental status examination. A recent medical examination or continuing medical follow-up by the patient's physician may suffice in lieu of the physical examination or the patient may be referred for a medical evaluation, when indicated.

A number of lawsuits have resulted from the failure to evaluate a patient properly before prescribing psychotropic medication. As a result of this omission, a patient's condition may be misdiagnosed, remain untreated, or worsen. Also, the patient may be exposed to unnecessary drug adverse effects and risks. In managed care settings, psychiatrists must be careful to spend enough time to do a competent evaluation, particularly in split-treatment arrangements.

Failure to Monitor or Supervise Probably the most common act of negligence associated with pharmacotherapy is the failure to supervise the patient's progress while taking the medication, including monitoring the patient for adverse effects. Monitoring may require laboratory testing, physical examination, and medical referral. Serum drug concentrations can be determined for a number of psychotropic medications. The primary indications for obtaining drug concentrations include assessing the therapeutic and toxic levels of medication and patient compliance with treatment. The use of carbamazapine (Tegretol), valproic acid (Depakene), and clozapine requires careful monitoring of the hematopoietic system and the liver. Failing to properly supervise patients taking psychotropic medications may unnecessarily subject them to harmful adverse effects and can delay a change to more effective treatment. A malpractice action may result if a patient is harmed by these omissions.

The question is often asked: How frequently should patients be seen for medication follow-up? The answer is that patients should be seen according to their clinical needs. No stock answer about the frequency of visits can be given. The longer the time interval between visits, however, the greater the likelihood of adverse drug reactions and clinical developments. Patients taking medications should probably not go beyond 6 months for follow-up visits. Managed care policies that do not reimburse for frequent follow-up appointments may result in psychiatrists prescribing large amounts of medications. The psychiatrist is duty bound to provide appropriate treatment to the patient quite apart from managed care or other payment policies.

Negligent Prescription Practices The selection of a medication, initial dosage, form of administration, and other related procedures are all decisions left to the clinical discretion of the treating psychiatrist. The law recognizes that the treating psychiatrist is in the best position to “know the patient” and to determine the best course of treatment under the circumstances. Accordingly, a psychiatrist's prescription practices will be evaluated by a reasonable practice standard. In prescribing psychotropic medication, psychiatrists are required to conform their procedures and decision making with those ordinarily practiced by other psychiatrists under similar circumstances.

Negligent prescription practices usually include exceeding recommended dosages and then failing to adjust the medication level to therapeutic levels, unreasonable mixing of drugs, prescribing medication that is not indicated, prescribing too many drugs at one time, and failing to disclose medication effects. Elderly patients frequently take a variety of drugs prescribed by different physicians. Multiple psychotropic medications must be prescribed with special care because of possible harmful interactions and adverse effects.

Psychiatrists who prescribe medications must explain the diagnosis, risks, and benefits of the drug within reason and as circumstances permit (Table 54.1-1). Obtaining competent informed consent may be problematic if a psychiatric patient has diminished cognitive capacity because of mental illness or chronic brain impairment; a substitute healthcare decision maker may need to provide consent.

Although there exists no consistently accepted standard for information disclosure for any given medical or psychiatric situation, as a rule of thumb, five areas of information are generally provided:

1. Diagnosis—description of the condition or problem
2. Treatment—nature and purpose of proposed treatment
3. Consequences—risks and benefits of the proposed treatment
4. Alternatives—viable alternatives to the proposed treatment including risks and benefits
5. Prognosis—projected outcome with and without treatment

Reprinted with permission from Simon RI: *Clinical Psychiatry and the Law*, ed 2. American Psychiatric Press, Washington, DC, 1992.

Table 54.1-1 Informed Consent: Reasonable Information to be Disclosed

Informed consent should be obtained each time a medication is changed and a new drug is introduced. If patients are injured because they were not properly informed of the risks and consequences of taking a medication, sufficient grounds may exist for a malpractice action.

Other Other areas of negligence involving medication that have resulted in malpractice actions include failure to treat adverse effects that have or should have been recognized; failure to monitor a patient's compliance with prescription limits; failure to prescribe medication or appropriate levels of medication according to the treatment needs of the patient; prescribing addictive drugs to vulnerable patients; failure to refer a patient for consultation or treatment by a specialist, and negligent withdrawal of medication treatment.

Split Treatment The following clinical vignette illustrates a common problem with split treatment in managed care settings.

> A psychiatrist provided medications for a depressed 43-year-old woman. A master's level counselor saw the patient for outpatient psychotherapy. The psychiatrist saw the patient for 20 minutes during the initial evaluation and prescribed a tricyclic drug. Tricyclic drugs have proved effective in alleviating the patient's depressions. As required by the managed care arrangement, the patient was prescribed enough drugs for follow-up in 3 months. The psychiatrist's initial diagnosis was recurrent major depression. The patient denied suicidal ideation. Appetite and sleep were markedly diminished. The patient had a long history of recurrent depression with suicide attempts. The patient's borderline personality disorder was not diagnosed. No further discussions were held after the initial discussion with the counselor. The counselor, who saw the patient once a week for 30 minutes, was unaware of the borderline personality disorder. The patient also had a chronic alcohol problem. Within 3 weeks, after a failed romantic relationship, the patient stopped taking her antidepressant medication during an alcohol binge. Her depression worsened rapidly. The patient committed suicide with an overdose of antidepressant drugs. The counselor and psychiatrist were sued for negligent diagnosis and treatment.

Psychiatrists must do an adequate evaluation, obtain prior medical records, and understand that there is no such thing as a partial patient. Split treatments are potential malpractice traps because patients can “fall between the cracks” of fragmented care. The psychiatrist retains full responsibility for the patient's care in a split treatment situation. This does not preempt the responsibility of the other mental health professionals involved in the patient's treatment. Section V, annotation 3 of the *Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry* states: “When the psychiatrist assumes a collaborative or supervisory role with

another mental health worker, he/she must expend sufficient time to assure that proper care is given."

In managed care or other settings, a marginalized role of merely prescribing medication apart from a working doctor-patient relationship does not meet generally accepted standards of good clinical care. The psychiatrist must be more than just a medication technician. Fragmented care in which the psychiatrist only dispenses medication while remaining uninformed about the patient's overall clinical status, constitutes substandard treatment that may lead to a malpractice action. At a minimum, such a practice diminishes the efficacy of the drug treatment itself or may even lead to the patient's failure to take the prescribed medication.

Split-treatment situations require that the psychiatrist remain fully informed of the patient's clinical status as well as the nature and quality of treatment the patient is receiving from the nonmedical therapist. In a collaborative relationship, the responsibility for the patient's care is shared according to the qualifications and limitations of each discipline. The responsibilities of each discipline do not diminish those of the other disciplines. Patients should be informed of the separate responsibilities of each discipline. The psychiatrist and the nonmedical therapist must periodically evaluate the patient's clinical condition and requirements to determine whether the collaboration should continue. Upon termination of the collaborative relationship, both treaters should inform the patient either separately or jointly. In split treatments, if the nonmedical therapist is sued, the collaborating psychiatrist will likely be sued, and vice versa.

Psychiatrists who prescribe medications in a split-treatment arrangement should be able to hospitalize a patient, if that should become necessary. If the psychiatrist does not have admitting privileges, prearrangements should be made with other psychiatrists who can hospitalize patients if emergencies arise. Split treatment is increasingly used by managed care companies and is a potential malpractice minefield.

Tardive Dyskinesia It is estimated that at least 10 to 20 percent of patients and perhaps as high as 50 percent of patients treated with neuroleptic drugs for more than 1 year exhibit some probable tardive dyskinesia. These figures are even higher for elderly patients. Despite the possibility for a large number of tardive dyskinesia–related suits, relatively few psychiatrists have been sued. In addition, patients who develop tardive dyskinesia may not have the physical energy and psychological motivation to pursue litigation.

Allegations of negligence involving tardive dyskinesia are based on a failure to evaluate a patient properly, a failure to obtain informed consent, a negligent diagnosis of a patient's condition, and a failure to monitor.

Most of the above-noted allegations of negligence were claimed in the landmark case, *Clites v State*. The plaintiff was a mentally retarded man who was institutionalized from age 11 and treated with major tranquilizers from age 18 to 23. After tardive dyskinesia was diagnosed at age 23, the plaintiff's family sued. The family claimed that the defendants negligently prescribed medication, did not inform the patient of the possibility of developing tardive dyskinesia, and failed to monitor and subsequently treat the patient for the adverse effects of the drugs. The jury found for the plaintiff and awarded damages in the amount of $760,165. This award was affirmed on appeal.

The appellate court ruled that the defendants were negligent and deviated from the standards of the "industry." Among the "deviations" the court noted were the failure to conduct regular physical examinations and laboratory tests, the failure to intervene at the first signs of tardive dyskinesia, the inappropriate use of multiple medications at the same time, the use of drugs for convenience (e.g., "behavior management") rather than for therapy, and the failure to obtain the plaintiff's informed consent.

Electroconvulsive Therapy Although many psychiatrists believe that ECT is a viable treatment for carefully selected patients with certain mental disorders, it has been estimated that only 3 to 5 percent of all psychiatric inpatients in the United States receive ECT. Thus, the potential number of legal actions alleging negligence associated with ECT is low. Nevertheless, lawsuits involving ECT are occasionally brought against psychiatrists. ECT-related injuries have occurred in a variety of circumstances in which negligence has been alleged. These cases fall into three groups: pretreatment, treatment, and posttreatment.

The APA Task Force on ECT recommends the following pretreatment procedures: (1) a psychiatric history and examination to evaluate the indications for ECT, (2) a medical examination to determine risk factors, (3) anesthesia evaluation, (4) written informed consent, and (5) an evaluation by a physician privileged to administer ECT. These recommendations do not define, in any absolute sense, the standard of care for ECT. Nonetheless, the task force report will likely be proffered as evidence of the standard of care in malpractice suits involving ECT. Official treatment guidelines must not be considered a substitute for the psychiatrist's sound clinical judgment, but failure to adequately conduct pretreatment procedures could endanger the welfare of the patient and secondarily result in a lawsuit for negligence.

Cases of ECT-related injuries in which negligence has occurred in the actual treatment process include (1) failure to use a muscle relaxant to reduce the chance of a bone fracture, (2) negligent administration of the procedure, and (3) failure to conduct an evaluation of the patient including the use of X-rays before initiating or continuing treatment.

Patients commonly experience certain adverse effects such as temporary confusion, disorientation, and memory loss following ECT. Sound clinical practice requires that psychiatrists provide reasonable posttreatment care and safeguards. Courts have held that the failure to properly attend to a patient for a period of time following ECT can result in malpractice liability. Posttreatment circumstances in which negligence may be claimed include (1) failure to evaluate complaints of pain or discomfort following treatment, (2) failure to evaluate a patient's condition before resuming ECT treatments, (3) failure to monitor a patient properly to prevent falls, and (4) failure to supervise a patient properly who had been injured as a result of ECT.

ECT-related law suits today are infrequent and do not represent a significant litigation problem for psychiatrists. Administering ECT is no longer associated with increased liability insurance premiums. However, recent developments in right-to-refuse-treatment law and increased statutory regulation of intrusive therapies ensure continued close legal scrutiny of ECT.

Suicidal Patients Psychiatrists are more likely to be sued when their patients commit suicide, particularly when psychiatric inpatients kill themselves. Psychiatrists are assumed to have more control over inpatients, making the suicides preventable.

The evaluation of suicide risk is one of the most complex, dauntingly difficult clinical tasks in psychiatry. Suicide is a rare event. In our current state of knowledge, clinicians cannot accurately predict when or if a patient will commit suicide. No professional standards exist for predicting who will or will not commit suicide. Professional standards do exist for assessing suicide risk, but at best, only the degree of suicide risk can be judged clinically following a comprehensive psychiatric assessment.

A variety of clinically useful assessment approaches are available to the practitioner. Suicide risk assessments are here-and-now evaluations, whose usefulness erodes quickly. Psychiatrists should frequently assess the risk of violence and update risk assessments at significant clinical junctures (e.g., room and ward changes, increased privileges, off-ward passes, discharge). Assessing suicide risk has been compared with weather forecasting. Short-term assessments (24 to 48 hours) are more accurate than longer-term assessments. In the short term, factors that influence future patient behaviors can be identified more precisely.

Assessing suicide risk involves 3 steps: (1) identifying patients at suicide risk, (2) assessing the overall suicide risk after evaluating specific suicide risk factors, and (3) providing treatment interventions that have a logical nexus to the overall suicide risk assessment.

Suicide risk evaluation should be standard practice with all patients, even if they do not exhibit overt suicidal symptoms. A significant number of patients with major affective disorders who commit suicide do not initially display suicidal ideation. Simply asking patients if they are suicidal and obtaining a "no harm contract" is insufficient. A layman could just as easily ask these same questions. Especially with new patients, there is no credible basis for relying upon such reassurances. The clinician's expertise is in taking an adequate psychiatric history and performing and recording a competent suicide risk assessment to direct appropriate clinical interventions.

A review of the case law on suicide reveals that certain affirmative precautions should be taken with a suspected or confirmed suicidal patient. For example, failing to perform a reasonable assessment of a suicidal patient's risk for suicide or implement an appropriate precautionary plan will likely render a practitioner liable. The law tends to assume that suicide is preventable if it is foreseeable. Courts closely scrutinize suicide cases to determine if a patient's suicide was foreseeable. *Foreseeability* is a deliberately vague legal term that has no comparable clinical counterpart, a common-sense rather than a scientific construct. It does not (and should not) imply that clinicians can predict suicide. Foreseeability should not be confused with preventability, however. In hindsight, many suicides seem preventable that were clearly not foreseeable. When considering treatment interventions for the suicidal patient, the clinician must assess both risks and benefits. The clinician

who is guided only by risk assessment is practicing defensive psychiatry to the possible detriment of the patient's care.

When making decisions about off-ward privileges, passes, and discharges, psychiatrists should record the risk-benefit assessment that guides their judgment. The risks and benefits of greater freedom must be weighed against those of continued hospitalization. This is both good clinical practice and sound clinical risk management. If a liability claim arises later, carefully documented risk-benefit assessments that inform the court about the clinician's contemporaneous decision-making process can provide the best malpractice defense.

Suicide Prevention Contracts Suicide prevention, or no-harm, contracts (oral or written) between therapists and patients are designed to encourage patients to call for help rather than commit suicide. These contracts have no legal authority. The danger is that the psychiatrist will place too great a reliance upon the contract and be falsely reassured that the patient is safe. Suicide prevention pacts between psychiatrists and patients should never replace an adequate assessment of suicide risk.

Defenses There are a number of legal defenses against a suicide malpractice claim. An open door policy that allows patients more therapeutic freedom of movement has produced a split in court decisions. Courts consider the specific facts of the case and the hospital staff's reasonableness in applying open door policies. However, juries and judges have difficulty giving credence to an abstract treatment idea when confronted with a dead patient.

The doctrine of sovereign or governmental immunity may bar by statute a finding of legal liability against a state or federal facility.

A suicide may be caused by factors unrelated to the psychiatrist's treatment of the patient. For example, a borderline patient may experience an overwhelming rejection between sessions and commit suicide without first attempting to contact the psychiatrist. The unforeseen rejection, a superseding, intervening variable, may have precipitated the suicide, not any negligent treatment.

The best-judgement defense can often be successfully argued when the suicidal patient was properly assessed and treated. In essence, the defense asserts that the patient committed suicide despite receiving reasonable care. The following clinical vignette illustrates the importance of conducting adequate suicide risk assessments.

A 33-year-old, single male patient committed suicide while on a 4-hour therapeutic pass from the hospital prior to anticipated discharge. The patient was hospitalized with a diagnosis of major depression, single episode and suicidal ideation. The managed care company authorized 2 weeks of inpatient treatment. The patient steadfastly denied suicidal thoughts or impulses after admission. He experienced moderate-to-severe depression, anhedonia, global insomnia, hopelessness, agitation, and loss of appetite. The patient signed a suicide prevention contract, promising to inform the psychiatrist or staff immediately of any suicidal ideation or impulses. After antidepressant treatment was started, the patient's energy level improved.

The psychiatrist and the hospital were sued by the patient's parents for wrongful death. In court, the plaintiff's expert testified that a suicide risk assessment would have determined that the patient was a significant suicide risk and that no pass should have been issued. The expert found no evidence in the psychiatric record that a formal suicide risk assessment was conducted before the pass was issued. She opined that not recording a suicide risk assessment was a violation of the standard of care. She further testified that the patient was at greater risk of suicide after administration of an antidepressant drug because the patient now had more energy to carry out his suicidal intent. Finally, the expert opined that the psychiatrist should not have relied on a no harm contract to prevent suicide in lieu of performing a suicide risk assessment.

The defendant psychiatrist testified that the patient's willingness to sign a no harm contract indicated the presence of a therapeutic alliance and the motivation to get well. The psychiatrist further testified that he had performed a formal suicide risk assessment that indicated a low-to-moderate suicide risk. It was an oversight on his part not to record the suicide risk assessment. In the psychiatrist's judgment, the therapeutic benefit of a pass outweighed the low-to-moderate suicidal risk.

The defense psychiatric expert testified that the psychiatrist met the standard of care by assessing the risk of suicide and weighing the suicide risk against the therapeutic benefits of a pass. The expert noted that the psychiatrist had not placed total reliance on the no-harm contract, but had used it appropriately to assess the working alliance with the patient. On cross examination, the defense expert admitted that the suicide risk assessments should have been recorded, but he felt that this was not the cause of the patient's suicide. The expert concluded by stating that a patient who is intent upon committing suicide cannot be stopped from doing so by the psychiatrist or the hospital, even if the patient is “totally locked up.”

The jury found for the plaintiffs, awarding $350,000 in monetary damages. When the jurors were polled by the defense attorney immediately after rendering the verdict, they indicated that a lack of documentation of the psychiatrist's decision-making process prevented their giving credibility to his testimony.

Violent Patients Psychiatrists who treat violent or potentially violent patients may be sued for failure to control aggressive outpatients and for the discharge of violent inpatients. Psychiatrists may be sued for failing to protect society from the violent acts of their patients if it was reasonable for the psychiatrist to have known about the patient's violent tendencies and if the psychiatrist could have done something that could have safeguarded the public. In the landmark case, *Tarasoff v Regents of the University of California*, the California Supreme Court ruled that mental health professionals have a duty to protect identifiable, endangered third parties from imminent threats of serious harm made by their outpatients. Since then, courts and state legislatures have increasingly held psychiatrists to a fictional standard of having to predict the future behavior (dangerousness) of their potentially violent patients. Research has consistently demonstrated that psychiatrists cannot predict future violence with any dependable accuracy.

The duty to protect patients and endangered third parties should be considered primarily a professional and moral obligation and, only secondarily, a legal duty. Most psychiatrists acted to protect both their patients and threatened others from violence long before *Tarasoff*.

If a patient threatens harm to another person, most states require that the psychiatrist perform some intervention that might prevent the harm from occurring. In states with duty-to-warn statutes, the options available to psychiatrists and psychotherapists are defined by law. In states offering no such guidance, health care providers are required to use the clinical judgment that will protect endangered third persons. Typically, a variety of options to warn and protect are clinically and legally available, including voluntary hospitalization, involuntary hospitalization (if civil commitment requirements are met), warning the intended victim of the threat, notifying the police, adjusting medication, and seeing the patient more frequently. The duty to protect allows the psychiatrist to consider a number of clinical options. Warning others of danger, by itself, is usually insufficient. Psychiatrists should consider the *Tarasoff* duty to be a national standard of care, even if they practice in states that do not have a duty to warn and protect.

A number of states have enacted immunity statutes limiting the responsibility of therapists for their patients' violent acts. Most of these statutes provide immunity for disclosures made to fulfill the duty to protect. The number of states passing such statutes increases every year. Definitions vary considerably concerning when the duty arises and how to discharge it. Most statutes require an actual threat made against a clearly identifiable victim before a duty to warn or protect arises. Most often, discharging the duty involves warning the intended victim and law enforcement authorities. Unfortunately, the duty to warn rather than protect is more often relied upon by immunity statutes. Thus, the duty to warn may be defensively invoked as a risk-management tool. Sometimes, warning by itself may induce violence. Thus immunity statutes may encourage reflexive rather than reflective patient management.

If a patient gives the psychiatrist good reason to believe that a warning should be issued to an endangered third party, the confidentiality of the communication that gave rise to the warning may be lost. The warning of endangered third parties has resulted in psychiatrists being compelled to testify in criminal cases.

The *Tarasoff* duty was originally applied to the outpatient setting. However, the same legal duty to protect individuals and society from harm by mental patients arises concerning the release of violent patients. The scope of the duty to warn is usually narrower than the duty not to release a violent patient. In cases involving failure to warn and protect an endangered third party, the violence was serious and imminent, and the victim was usually identifiable and foreseeable. The duty not to release a violent patient has a broader scope because these patients often do not express specific threats toward persons or groups and thus pose a threat to the general public. In release cases, courts have extended the therapist's duty beyond that owed to readily identifiable victims. The number of malpractice suits alleging negligent release is at least five to six times the number of outpatient cases alleging a *Tarasoff* duty.

The physician has a general duty to protect other members of society. This duty is not new to the law. The doctor-patient relationship often gives the physician insider

information about the potential for violence that no one else possesses. The duty to warn and protect arises from the concept of latent dangers to which the physician is privy but not the endangered persons. To practitioners, this reasoning makes more clinical sense than the legal argument of a special relationship between doctor and patient creating the duty to protect endangered persons.

Psychiatrists should not discharge patients without arranging for adequate aftercare. Doomed-to-fail discharges must be avoided by considering what is different in the patient's clinical condition and situational circumstances. The patient's willingness to cooperate with the psychiatrist is critical to follow-up treatment. The psychiatrist should structure the follow-up visits to encourage compliance. Scheduling the patient as soon as possible for appropriate treatment programs that are readily accessible facilitates patient adherence. A study of Veterans Affairs (VA) outpatient referrals showed that when inpatients were referred to the VA mental health clinic, approximately 50 percent failed to keep their first appointments. However, the psychiatrist's ability to ensure follow-up care is limited. Most patients retain the right to refuse treatment after discharge. This must be acknowledged by both the psychiatric and legal communities.

The treatment of psychiatric inpatients has changed dramatically in the managed care era. Most psychiatric units, particularly those in general hospitals, have become short-stay, acute-care, psychiatric facilities. Under managed care, only the sickest, most acute patients are admitted, and then only for short-term hospitalization. Strict scrutiny by utilization reviewers ensures that these patients are hospitalized for brief periods. The purpose of hospitalization is crisis-intervention management, safety, and stabilization of the patient. Treatment is being provided by a variety of mental health professionals, yet the psychiatrist often bears the burden of liability for treatments gone awry. Little opportunity is usually available during the hospital stay to develop a therapeutic alliance with patients. The psychiatrist's stock-in-trade, the ability to communicate with patients, is severely curtailed. All these factors greatly increase the risk for a malpractice suit alleging premature or negligent discharge of potentially violent patients. Managed care restrictions on length of hospital stay must not be allowed to interfere with clinically based decisions indicating a need for continued hospitalization. The American Medical Association has issued guidelines for discharging inpatients, entitled “Evidence-Based Principles of Discharge and Discharge Criteria.”

In either outpatient or inpatient settings, psychiatrists will likely meet the standard of care if they reasonably assess the patient's potential risk for violence, which then directs the choice of clinical interventions. For example, a history of past violence should alert the clinician to future violence. Professional standards are described in the psychiatric literature for the assessment of the risk of violence. However, no standard of care exists for the prediction of individual violent behavior. Psychiatrists should assess the risk of violence frequently, updating risk assessments at significant clinical junctures (e.g., room and ward changes, increased privileges, off-ward passes, discharge). Assessing the risk of violence is a here-and-now determination. Probability risk assessments of violence become progressively less accurate beyond the immediate short term (e.g., 24 to 48 hours). A risk-benefit assessment should be conducted and recorded that lays out the clinician's reasoning both for and against issuing a pass or discharge.

A psychiatrist evaluated a 43-year-old man and made the diagnosis of delusional disorder, jealousy type, after obtaining the following history:

Approximately 6 months before the evaluation, the patient began to believe that his wife was having an affair with her boss. This idea grew into a conviction. The wife vehemently denied his accusations and demanded that the patient see a psychiatrist after he began making threats against her boss. He acquiesced, though he felt there was nothing wrong with him.

After making the diagnosis, the psychiatrist recommended continued outpatient visits and medication and discussed the possibility of hospitalization. Involuntary hospitalization was not considered because of a requirement of a recent overt violent act in the commitment statute. The psychiatrist assessed the risk of violence as high. The patient refused to come back for any more visits, saying he would make no further threats. There is no history of prior violence. When the psychiatrist informed the patient that she had a duty to warn the boss of the danger, the patient stated, “Go ahead, he knows anyway.” As required by state law, the psychiatrist called the threatened individual and informed him of the patient's threats and high risk for violence. The boss dismissed the threats “as so much nonsense.” In addition, the psychiatrist informed the law enforcement authorities.

Six weeks later, the psychiatrist reads in the paper that the patient shot and killed the boss, his wife, and himself. A year later, the psychiatrist was sued by the boss's family for a failure to warn and protect the murdered individual. In the suit, the estate claimed that the psychiatrist neither stressed the seriousness of the threat sufficiently nor properly assessed the patient for involuntary hospitalization. The complaint stated that a medical certification for involuntary hospitalization should have been obtained by the psychiatrist so that the court could decide on the appropriateness of the petition. Her records clearly documented the diagnosis, treatment recommendations, the serious warning to the boss, and the informing of law enforcement authorities.

The state statute governing *Tarasoff*-type cases contained an immunity provision that precluded the imposition of legal liability against mental health professionals when endangered third parties are warned of potential violence threatened by a patient and the police are also informed. The lawsuit was dismissed by the court on a motion for summary judgment made by the psychiatrist's attorney. The lawsuit was precluded by the state immunity statute. The claim that the psychiatrist did not consider involuntary hospitalization was also dismissed because the state's commitment statute required an overt act of violence within 30 days of an evaluation. Thus, the patient did not meet substantive criteria for civil commitment.

A state superior court jury awarded Wendell Williamson, a former University of North Carolina law student, $500,000, holding Myron B. Liptzin liable for negligent treatment that resulted in his former patient's violent acts on January 26, 1996, after treatment ended on May 25, 1994. In the criminal trial, Williamson was found not guilty by reason of insanity in the random shooting murders of 2 men and the serious injury of a police officer near the campus. Williamson was institutionalized with a diagnosis of paranoid schizophrenia.

In the civil suit, the patient alleged that Liptzin, then director of student health psychiatric services, failed to correctly diagnose and explain the seriousness of his illness. Despite improvement with antipsychotic medication while under Liptzin's care, the plaintiff claimed that his treatment was mismanaged by Liptzin not properly arranging for a referral to another psychiatrist following Liptzin's retiement. Liptzin countered that he encouraged Williamson to attend a mental health clinic during the summer; however, he did not make a specific referral. He had given the 26-year-old Williamson a 30-day supply of antipsychotic medication and urged the patient to have his family physician refill the medication if he did not contact a psychiatrist. After discharge, Williamson stopped taking his medication. A judge upheld the $500,000 jury award but the decision is being appealed.

The patient did not appear violent nor make any violent threats at the time of termination. Psychiatrists and other mental health professionals cannot provide society with guarantees of saftey either during or after treatment. Most patients retain the right to refuse treatment and to stop their medications. Violent behaviors are the result of complex, dynamic interactions among a variety of clinical, personality, social, and environmental factors that vary across time and situations. The Liptzin result is unprecedented. It is doubtful that this case will establish a new standard of care for the termination of patients.

RIGHTS OF PATIENTS

Involuntary Hospitalization An individual may be involuntarily hospitalized only if certain statutorily mandated criteria are met. Three fundamental substantive criteria serve as the foundation for all statutory commitment requirements: the individual must be (1) mentally ill, (2) dangerous to self or others, or (3) unable to provide for one's basic needs (gravely disabled). Each state determines which criteria are required and defines what each means. Terms such as “mentally ill” are often loosely described, which shifts the responsibility for proper definition to the clinical judgment of the petitioner.

In addition to persons with mental illness, a number of states have enacted legislation that provides for involuntary hospitalization of three other distinct groups: developmentally disabled (mentally retarded), substance addicts (alcohol, drugs), and mentally disabled minors. Special commitment provisions usually govern the requirements for admission and discharge of mentally disabled minors in addition to numerous due process rights afforded these individuals.

Currently, most mentally ill patients are being treated as outpatients. Most involuntary hospitalizations are for short periods. Involuntary hospitalization of psychiatric patients usually occurs when violent behavior threatens to erupt toward self or others and when patients become unable to care for themselves. These patients often have mental disorders that readily meet the substantive criteria for involuntary hospitalization. Violent individuals who do not suffer from a mental disorder are not candidates for involuntary hospitalization; they are the responsibility of law enforcement authorities.

Clinicians do not actually commit patients. Civil commitment comes under the sole jurisdiction of courts or other, quasi-judicial bodies. The psychiatrist initiates medical certification that brings the patient before the court, usually after a brief period of evaluation. In seeking medical certification the psychiatrist must be guided

by the treatment needs of the patient and the protection of the patient or others who may be endangered by the patient.

The most common type of lawsuit for involuntary hospitalization alleges that a psychiatrist failed to act in good faith and to adhere to statutory requirements, resulting in a wrongful commitment. Generally, these lawsuits are brought under the theory of false imprisonment. Other areas of liability for alleged wrongful commitment include assault and battery, malicious prosecution, abuse of authority, and intentional infliction of emotional distress.

In most states, psychiatrists are granted qualified immunity from liability if they use reasonable professional judgment and act in good faith when petitioning for commitment. The psychiatrist may be sued if evidence exists of willful, blatant, or gross failure to adhere to statutorily defined commitment procedures.

Most states recognize the right of inpatients to refuse treatment. Even though the patient is involuntarily hospitalized, the hospitalization does not negate a presumption of competence. In most states, involuntarily hospitalized patients who refuse medication require a separate court hearing for adjudication of incompetence and the provision of substituted consent by the court. Recently, persons hospitalized under criminal commitment have been accorded a qualified right to refuse treatment. The courts have found that patients' constitutional right to due process is “adequately protected by the exercise of professional judgment” within the medical peer review process of the institution.

Hospitalized patients possess other rights. Patients possess rights of visitation, although these rights can be temporarily suspended for proper cause relating to the patient's care and treatment. Free communication by hospitalized patients through mail, telephone, or visitors is considered a right, unless protection of the patient or others requires supervision of communications. The right to privacy includes allowing patients to have secure locker space, private toilet and shower facilities, and minimum square footage of floor space. Confidentiality is also protected. Economic rights include the right to have and spend money and to handle one's own financial affairs responsibly. In most jurisdictions, involuntarily hospitalized patients do not lose their civil rights, such as the right to manage their own money. Hospitalized patients must be paid for their work in certain jurisdictions unless it is truly therapeutic labor (i.e., work not connected with maintenance of the hospital). “Patient rights” are not absolute and often must be tempered by the clinical judgment of the mental health professional. Inevitably, disputes over perceived or real violations of patients' rights arise. In some jurisdictions, a civil rights officer or ombudsman is mandated by statute to mediate these disputes.

Managed Care Psychiatrists have certain responsibilities toward patients treated in managed care settings, including the responsibilities to disclose all treatment options, to exercise appeal rights, to continue emergency treatment, and to cooperate reasonably with utilization reviewers.

Responsibility to Disclose Psychiatrists have a continuing responsibility to the patient to obtain informed consent for treatments or procedures. All treatment options should be fully disclosed, even those not covered under the terms of a managed care plan. A number of states are enacting legislation making gag rules that limit information about treatment provided to patients under managed care illegal. Managed care organizations cannot require affiliated physicians to limit or withhold advice about treatment options to Medicare beneficiaries, even if those treatments are not covered. Congress has banned gag rules that prohibit health care providers from advising Medicare beneficiaries who choose the health maintenance organization (HMO) option about medical treatments that the HMO does not cover.

Responsibility to Appeal The American Medical Association Council on Ethical and Judicial Affairs states that physicians have an ethical obligation to advocate for any care that they believe will materially benefit their patients, regardless of any allocation guidelines or gatekeeper directives.

Responsibility to Treat Physicians are liable for failure to treat their patients within the defined standard of care. The treating physician has sole responsibility to determine what is medically necessary. Managed care programs generally limit or deny payment for services not the actual services. Psychiatrists must be careful not to discharge suicidal or violent patients prematurely merely because continued coverage will not be approved by a managed care company.

Responsibility to Cooperate With Utilization Review As an ally of the patient and in the provision of good clinical care, the psychiatrist should cooperate with utilization reviewers' requests for information upon proper authorization from the patient. When benefits are denied, one must understand and follow grievance procedures carefully, return telephone calls from review agencies, and provide documented, solid justification for continued treatment.

Boundary Violations Increasingly, psychiatrists and other mental health professionals are being sued for treatment boundary violations that lead to patient exploitation and harm. Monetary and sexual exploitation are the most common.

The therapist establishes treatment boundaries to define and secure the professional relationship that promotes a therapeutic alliance with the patient. The boundary guidelines listed in Table 54.1-2 are accepted by most clinicians, regardless of their treatment orientation. Absolute boundary standards do not exist. Much of the variability in establishing boundaries arises from the nature of the patient, the type of treatment, the nature of the therapist, and the status of the therapeutic alliance. Sound treatment boundaries facilitate a fluctuating, reasonably flexible, psychological interaction between therapist and patient.

Maintain relative therapist neutrality
Foster psychological separateness of patient
Protect confidentiality
Obtain informed consent for treatments and procedures
Interact verbally with patients
Ensure no previous, current or future personal relationship with the patient
Minimize physical contact
Preserve relative anonymity of therapist
Establish a stable fee policy
Provide consistent, private and professional setting
Define time and length of sessions

Reprinted with permission from Simon RI: Treatment boundary violations: Clinical, ethical and legal considerations. Bull Am Acad Psychiatry Law 20: 269, 1992.

Table 54.1-2 Boundary Guidelines for Psychotherapy

Once treatment boundaries are established, problems about boundaries (boundary issues) that inevitably arise from the patient are grist for the therapeutic mill. Boundary violations are strictly created by the therapist. Boundary crossings occur commonly in every therapy and are quickly identified and rectified, causing no harm to the patient.

Sexual exploitation of patients is preceded by progressively serious boundary violations. The earliest boundary violations often begin at the end of the session “between the chair and the door.” Patient and therapist doff their treatment roles to begin a social relationship. Therapy should be considered to be in progress until the patient leaves the office. The time and space “between the chair and the door” can be scrutinized for early warning indicators of inchoate boundary violations.

The principle of abstinence is central to keeping sound treatment boundaries. Therapists must abstain from using a patient for their personal advantage. The therapist's personal gratification should come from the work with the patient. Payment received for the therapist's professional services is the only material gratification permitted.

When making interventions with patients, therapists can spot-check the soundness of their boundaries by asking two basic questions: (1) Is the intervention being made to advance the patient's treatment or for the personal gratification of the therapist? (2) Is the intervention part of progressive boundary violations? Boundary violations need not prove fatal to the patient's treatment if they are caught in time, turned to account for the sake of the patient's treatment, and self-scrutinized by the therapist.

Confidentiality Confidentiality refers to the right of a patient to have communications spoken or written in confidence to a psychiatrist kept undisclosed to outside parties without authorization. Four general sources provide the bases for recognizing and safeguarding patient confidences. First, all 50 states and the District of Columbia have acknowledged this right of protection by creating some form of confidentiality provisions in either professional licensure laws or confidentiality and privilege statutes. In 1996 the United States Supreme Court (in *Redmond v Jaffe*) ruled that communications between psychotherapist and patient are confidential and need not be disclosed in federal trials. The second source, with the longest tradition, comprises the ethical codes of the various mental health professions. Third,

common law has long recognized an attorney-client privilege; developing case law has established similar protection for physicians and psychotherapists. Fourth, the right of confidentiality may be subsumed under the right of privacy.

Breach of Confidentiality Once the doctor-patient relationship is created, the professional has an automatic duty to safeguard a patient's disclosures. This duty is not absolute because circumstances exist in which breaching confidentiality is both ethical and legal.

Patients may waive confidentiality in a variety of situations including managed care settings. Medical records routinely go to potential employers or to insurance companies to obtain benefits. A limited waiver of confidentiality usually exists when a patient participates in group therapy. Whether a group member can be compelled in court to disclose information shared by another group member during group therapy remains unsettled. Many state confidentiality statutes provide statutory exceptions to confidentiality between the psychiatrist and the patient in one or more situations. Apart from statutory disclosure requirements and judicial compulsion, no legal obligation exists to provide information, even to law enforcement officials. In section 4, annotation 8, The *Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry* states “Psychiatrists at times may find it necessary, in order to protect the patient or the community from imminent danger, to reveal confidential information disclosed by the patient.”

Patients' access to their own records is usually controlled by state statutes. These statutory provisions can be found under the heading of “medical records” or the much broader term “privilege.” The rules governing patient's access to their own records are complex and vary considerably in each jurisdiction.

The material contained in the psychiatric record can cause an iatrogenic exacerbation of a patient's mental disorder. Every effort should be made to be present when the patients review their records to clarify statements and answer questions. The physical record is the property of the psychiatrist. The information contained in the record, however, belongs to the patient. The original record should not be relinquished to the patient; only a copy should be provided.

TESTIMONIAL PRIVILEGE

Testimonial privilege, a statutorily created rule of evidence, permits the holder of the privilege (e.g., the patient) to exercise the right to prevent the person to whom confidential information was given (e.g., the psychiatrist) from disclosing it in a judicial proceeding. The patient—not the psychiatrist—holds the privilege that controls the release of confidential information. It is called “testimonial privilege” because it applies only to the judicial setting. Privilege statutes represent the most common recognition by the state of the importance of protecting information provided by a patient to a psychotherapist. This recognition departs from the essential truth-finding purpose of the American system of justice by insulating certain information from disclosure in court. This protection is justified by the special need for privacy in the doctor-patient relationship, that outweighs the quest for an accurate outcome in court.

Privilege statutes cover most of the mental health professions. Cases have been successfully litigated in which the broader physician-patient category was applied to the psychotherapist in the absence of an applicable statute.

Exceptions to Testimonial Privilege Privilege statutes also specify numerous exceptions to testimonial privilege. The most common exceptions include child abuse reporting, civil commitment proceedings, court-ordered examinations, competency proceedings, and cases in which a patient's mental state is in question as part of litigation. This last exception, known as the patient-litigant exception, commonly occurs in will contests, workers' compensation cases, child custody disputes, personal injury litigation, and malpractice actions.

Liability An unauthorized or unwarranted breach of confidentiality may cause a patient serious emotional harm. Consequently, a psychiatrist may be held liable for such a breach based on at least four legal theories: malpractice (breach of confidentiality), breach of statutory duty, invasion of privacy, and breach of (implied) contract.

SEXUAL MISCONDUCT

Therapist-patient sex is almost always preceded by progressive boundary violations in treatment. Patients are often psychologically damaged by both the precursor boundary violations and the end-stage sexual misconduct of the therapist. Sexual misconduct does not occur in isolation; it usually involves a variety of negligent acts of omission and commission.

Civil Liability Psychiatrists who sexually exploit their patients are subject to civil and criminal actions in addition to ethical and professional licensure revocation proceedings. Malpractice is the most common legal action (Table 54.1-3)

Civil lawsuit
Negligence
Loss of consortium
Breach-of-contract action
Criminal sanctions (e.g., statutory, adultery, sexual assault, rape)
Civil action for intentional tort (e.g., battery, fraud)
License revocation
Ethical sanctions
Dismissal from professional organizations

Reprinted with permission from Simon RI, *Clinical Psychiatry and the Law*, ed 2. American Psychiatric Press, Washington, DC, 1992.

Table 54.1-3 Sexual Exploitation: Legal and Ethical Consequences

In a sexual exploitation case, the plaintiff has the difficult burden of proving that the exploitation actually took place. This burden is met if the plaintiff can provide corroborating evidence to support the claim such as testimony from other abused (former) patients, letters, pictures, hotel or motel receipts, and identification of distinctive therapist body markings.

If the defendant practitioner admits to the sexual misconduct, then the plaintiff is left with the responsibility of showing that he or she sustained injuries as a result. Usually, patient injury occurs in the form of emotional damage such as trust problems and a worsened psychiatric condition. Expert psychiatric testimony establishes the type and extent of psychological damage and affirms a breach in the standard of care. The number of states with civil statutes proscribing sexual misconduct is steadily increasing.

Three basic types of remedies are codified in the statutes of a number of states: reporting, civil liability, and criminal penalties. Reporting statutes require the disclosure to state authorities by a therapist who learns of any past or current therapist-patient sex. States that have civil statutes proscribing sexual misconduct incorporate a standard of care that makes malpractice suits easier to pursue. Some of these statutes restrict unfettered discovery of the plaintiff's past sexual history. Criminal sanctions are often the only remedy for exploitative therapists without malpractice insurance, for therapists who are unlicensed, or for therapists who do not belong to professional organizations.

Criminal Sanctions A number of states have passed statutes that make sexual activity by a therapist with a patient a criminal act. Moreover, sexual exploitation of a patient, under certain circumstances, may be considered rape or some analogous sexual offense and may therefore be criminally actionable. Typically, the criminality is determined by one of three factors: the practitioner's means of inducement, the age of the victim, or the availability of a relevant state criminal code.

Sex with a current patient may be prosecuted under sexual assault statutes if the state can prove beyond a reasonable doubt (i.e., with 90 to 95 percent certainty) that the patient was coerced into engaging in the sexual act. Usually, this type of evidence is limited to the use of some form of substance to either induce compliance or reduce resistance. However, a variety of methods have been used by therapists to coerce patients into sexual submission such as anesthesia, ECT, hypnosis, drugs,

force, and threats of harm. To date, claims of “psychological coercion” through manipulation of transference phenomena have not successfully met the level of proof necessary to prosecute a criminal case. In cases involving a minor patient, the issues of consent or coercion are irrelevant, since minors and incompetent persons (including adult incompetent persons) are considered unable to provide valid consent. Therefore, sex with a child or an incompetent person is automatically considered a criminal act.

Licensure and Ethical Codes State licensing organizations, unlike professional associations, may discipline offending professionals more effectively and punitively by suspending or revoking their license. Licensing boards are not as constrained as the courts, which require rigorous rules of evidence in trial procedures. It is generally less difficult for the patient to seek redress through this means. In the vast majority of cases in which the evidence was reasonably sufficient to substantiate a claim of sexual exploitation, the professional's license was revoked or the professional was suspended from practice for a period of time.

Patients can bring ethical charges against psychiatrists before the district branches of the APA. There is no time limit for bringing ethical charges. Ethical violators may be reprimanded, suspended, or expelled from the APA. All national organizations of mental health professionals proscribe sexual relations between therapist and patient. Ethical charges can be filed only against members of a professional group; this option is not available to patients of therapists who do not belong to a professional organization.

RECOVERED MEMORIES

The current clamorous controversy concerning recovered memories of sexual abuse threatens to undermine the credibility of the mental health professions. The debate has generated intense passions that have driven an increasing number of recovered memory cases into the courts. Patients alleging recovered memories of abuse have sued parents and other alleged perpetrators. In a number of instances, the alleged victimizers have sued therapists who, they claim, negligently induced false memories of sexual abuse. In an aboutface, some patients have recanted and joined forces with others (usually their parents) to sue therapists.

The memory debate has polarized a number of therapists into believers and disbelievers. Most therapists hold personal beliefs about the validity of recovered memories of sexual abuse that are somewhere between the extremes. Strongly held personal biases about recovered memories represent a new occupational hazard for clinicians. Such feelings can undermine the therapists' duty of neutrality to their patients, creating deviant treatment boundaries and the provision of substandard care.

Litigation in recovered memory cases is expected to soar in the coming years. Some predict multimillion dollar verdicts against mental health practitioners. A jury awarded over $10 million to the plaintiffs in a case alleging that therapists implanted memories of satanic sexual abuse. A fundamental allegation in these cases is that the therapist abandoned a position of neutrality to suggest, persuade, coerce, and implant false memories of childhood sexual abuse. The guiding principle of clinical risk management in recovered memory cases is maintenance of therapist neutrality and establishment of sound treatment boundaries.

Further complicating the matter is the empirical evidence about memory mechanisms, which (as is typical for any emerging science) reveals contradictory findings about how and what persons retain in memory and forget in various settings. Empirical studies often fail to distinguish whether allegedly repressed memories are not retrieved or simply not reported to researchers.

Valid risk management has a solid clinical footing and is secondarily informed by awareness of the legal issues. The following risk management principles should be considered when evaluating or treating a patient who recovers memories of abuse in psychotherapy:

1. Maintain therapist neutrality—do not suggest abuse
2. Stay clinically focused—provide adequate evaluation and treatment for patients presenting problems and symptoms
3. Carefully document the memory recovery process
4. Manage personal bias and countertransference
5. Avoid mixing treater and expert witness roles
6. Closely monitor supervisory and collaborative therapy relationships
7. Clarify nontreatment roles with family members
8. Avoid special techniques (e.g., hypnosis or sodium amytal) unless clearly indicated; obtain consultation first
9. Stay within professional competence—do not take cases you cannot handle
10. Distinguish between narrative truth and historical truth
11. Obtain consultation in problematic cases
12. Foster patient autonomy and self-determination—do not suggest lawsuits
13. In managed care settings, inform patients with recovered memories that more than brief therapy may be required
14. When making public statements, distinguish personal opinions from scientifically established facts
15. Stop and refer, if uncomfortable with a patient who is recovering memories of childhood abuse
16. Do not be afraid to ask about abuse as part of a competent psychiatric evaluation

SECLUSION AND RESTRAINT

Seclusion and restraint raise complex psychiatric legal issues. Seclusion and restraint have both indications and contraindications (Table 54.1-4 and Table 54.1-5). Seclusion and restraint have become increasingly regulated over the past decade.

1. Prevent clear, imminent harm to the patient or others
2. Prevent significant disruption to treatment program or physical surroundings
3. Assist in treatment as part of ongoing behavior therapy
4. Decrease sensory overstimulation*
5. At patient's voluntary reasonable request

* Seclusion only.
Reprinted with permission from Simon RI: *Concise Guide to Psychiatry and the Law for Clinicians*, ed 2. American Psychiatric Press, Washington, DC, 1998.

Table 54.1-4 Indications for Seclusion and Restraint

1. Extremely unstable medical and psychiatric conditions*
2. Delirious or demented patients unable to tolerate decreased stimulation*
3. Overtly suicidal patients*
4. Patients with severe drug reactions, overdoses or requiring close monitoring of drug dosages*
5. For punishment or convenience of staff

* Unless close supervision and direct observation are provided.
Reprinted with permission from Simon RI: *Concise Guide to Psychiatry and the Law for Clinicians*, ed 2. American Psychiatric Press, Washington, DC, 1998.

Table 54.1-5 Contraindications to Seclusion and Restraint

Legal challenges to the use of restraints and seclusion have been brought on behalf of institutionalized mentally ill and mentally retarded persons. Typically, these lawsuits do not stand alone but are part of a challenge to a wide range of alleged abuses.

Generally, courts hold, or consent decrees provide, that restraints and seclusion be implemented only when a patient creates a risk of harm to self or others and no less-restrictive alternative is available. Additional restrictions include the following:

1. Restraint and seclusion can only be implemented by a written order from an appropriate medical official
2. Orders are to be confined to specific, time-limited periods
3. A patient's condition must be regularly reviewed and documented
4. Any extension of an original order must be reviewed and reauthorized

The acceptability of restraint or seclusion for the purposes of training was recognized by the Supreme Court in *Youngberg v Romeo*, which challenged the "treatment" practices at the Pennhurst State School and Hospital in Pennsylvania. The Court held that patients could not be restrained except to ensure their safety or, in certain undefined circumstances, "to provide needed training." Although recognizing that the defendant had a liberty interest in safety and freedom from bodily restraint, the Court noted that these interests were not absolute nor in conflict with the need to provide training. The Court also held that decisions made by appropriate professionals regarding restraining the patient would be presumed correct. Psychiatrists and other mental health professionals have lauded the decision because the Court recognized that professionals are better able than the courts to determine the needs of patients, including deciding when restraint is appropriate.

Most states have enacted statutes that regulate the use of restraints, often specifying the circumstances in which restraints can be used—usually when a risk of harm to self or danger to others is imminent. Statutory regulation of the use of seclusion is much less common. About one-half of states have laws governing seclusion. Most states with laws regarding seclusion and restraint require some type of documentation of their usage. A number of courts and state statutes outline certain due process procedures that must be followed before restraint or seclusion can be used for nonclinically indicated, disciplinary purposes. These include some form of notice, a hearing, and involvement of an impartial decision-maker.

The APA Task Force on the Psychiatric Uses of Seclusion and Restraint has developed guidelines for the appropriate use of seclusion and restraints, and the JCAHO has promulgated guidelines for hospitals regarding seclusion and restraint requirements. Professional opinion concerning the clinical uses of physical restraints and seclusion varies considerably among psychiatrists. Seclusion can be justified on both clinical and legal grounds for a variety of uses, unless precluded by state freedom from restraint and seclusion statutes.

COMPETENCY: A CORNERSTONE CONCEPT

Competency is a broad concept, encompassing many different legal issues and contexts. Its definition, requirements, and application can vary widely depending upon the context such as health care decision making, executing a will, standing trial, or confessing to a crime.

In general, competency refers to some minimal mental capacity required to perform a specific, legally recognized act or to assume some legal role. When patients' mental competency is in doubt, a court scrutinizes their perception of reality and memory functions, since these abilities bear on the reliability of their responses or testimony.

Incompetent is a legal term applied to individuals who are considered by law not to be mentally capable of performing a particular act or assuming a particular role. The adjudication of incompetence is issue specific. A psychiatric patient who is adjudicated incompetent to execute a will may not be automatically incompetent to do other things, such as consenting to treatment, testifying as a witness, marrying, driving, or making a legally binding contract.

Competency, in both the civil and the criminal context, is commonly raised in two situations—when the person is either a minor (under age 18) or mentally disabled. Minors are not considered legally competent in some situations and require the consent of a parent or designated guardian. Minors who are considered emancipated or mature or who are competent to consent in some cases of medical need or emergency are generally an exception to this rule.

Mentally disabled persons present more complex issues in evaluating competency. A lack of competency cannot be presumed from a person's treatment for mental illness or from institutionalization. Mental disability does not necessarily render a person incompetent or incompetent in all areas of functioning. Foolish decisions by themselves do not denote mental incompetence but are merely an inevitable consequence of the human condition. More often, the psychiatrist must determine if any specific functional incapacities exist that render a person unable to make a particular kind of decision or to perform a particular type of task.

Health Care Decision Making Competency to consent and the right to refuse treatment go hand in hand. Under most circumstances, competent patients have a right to refuse treatment. Forcing treatment against the wishes of a competent patient may result in a suit for assault and battery as well as malpractice. Merely because patients are psychotic does not necessarily mean that they lack sufficient health care decision-making capacity. Nor does involuntary hospitalization deprive a patient of the right to refuse treatment in most states.

Some psychiatric patients who suffer from impaired mental capacity may have difficulty giving valid informed consent to proposed diagnostic procedures and treatments. One study of informed consent found that the schizophrenia and depression groups demonstrated poorer understanding of treatment disclosures, poorer reasoning in decision making regarding treatment, and a greater likelihood of failing to appreciate their illnesses or the potential treatment benefits. The legal requirement that the psychiatrist obtain competent informed consent for proposed treatments and procedures is not negated simply because the patient appears to need medical intervention or would likely benefit from it. Psychiatrists must ensure that the patient or an appropriate substitute decision-maker has given a competent consent before proceeding with treatment.

For a patient's consent to be informed, three essential elements must be satisfied: competency, information, and voluntariness. The patient must be given enough information to make a knowledgeable decision. The decision (consent) must be given voluntarily by a person who is legally competent. Gag rules imposed on psychiatrists in managed care settings to keep patients from being fully informed about treatment options are antithetical to obtaining a valid informed consent.

Several exceptions exist to the requirement of informed consent. The most common is the emergency exception, in which consent is implied when the patient is unable to give consent (e.g., unconsciousness) and is suffering from an acute, life-threatening crisis that requires immediate medical attention. Other exceptions include incompetency (substitute decision-maker necessary), therapeutic privilege (withholding full disclosure if informing would seriously worsen patient's condition or foreclose rational decision making), and waiver (patient competently declines being informed). Clinicians must be especially careful not to use these exceptions to circumvent the requirement of obtaining informed consent. When asserting an exception to obtaining informed consent, the clinician's reasons should be carefully documented.

Levels of Competency Competency is usually narrowly defined legally as cognitive capacity. No firmly established criteria exist for determining a patient's competence. A minimal level of decision making must exist in which the patient can at least (1) understand the particular treatment being offered; (2) make a discernible decision, one way or another, regarding the treatment that has been offered; and (3) communicate that decision verbally or nonverbally.

This minimal standard of decision-making capacity obtains only a simple consent. The patient does not give an informed consent, because alternative treatment choices are not provided. There are generally four standards for determining incompetency in decision making. Based upon levels of mental capacity required, these standards are (1) communication of choice, (2) understanding the information provided, (3) appreciation of one's situation and the risks and benefits of options available, and (4) rational decision-making

Psychiatrists usually prefer the rational decision-making standard in determining whether an individual is incompetent, but most courts accept the first two standards. Rational decision-making occurs when a patient's consent reflects that patient's freedom of choice, personal needs, and values applied to the risks and benefits of appropriate treatment options. The clinician should consider obtaining legal advice if the competency issue cannot be resolved by additional medical and psychiatric

consultation.

Except in an emergency, patients lacking health care decision-making capacity need an authorized representative or guardian appointed to make health care decisions on their behalf. Table 54.1-6 lists a number of consent options that may be available, depending on the jurisdiction.

1. Proxy consent of next of kin
2. Right-to-die statutory surrogate laws (spouse or court appointed guardian when treatment wishes of patient unstated)
3. Advance directives (living will or durable power of attorney)
4. Adjudication of incompetence; appointment of guardian
5. Institutional administrators or committees
6. Treatment review panels
7. Substituted consent of the court

Reprinted by permission from Simon RI: Ethical and legal issues in neuropsychiatry. In *American Psychiatric Press Textbook of Neuropsychiatry*, ed 3, RE Hales, SC Yudofsky, editors. American Psychiatric Press, Washington, DC, 1997.

Table 54.1-6 Common Consent Options for Patients Lacking the Mental Capacity for Health Care Decision-Making

Guardianship Guardianship is a method of substitute decision-making for individuals who have been judicially determined to be unable to act on their own behalf. Historically, the state or sovereign possessed the power and authority to protect the estate of incompetent persons.

This traditional role is carried forward in the purpose of guardianship today. In a number of states, separate provisions exist for the appointment of a “guardian of one's person” (e.g., substitute health care decision-maker) and a “guardian of one's estate” (e.g., someone authorized to make contracts to sell one's property). The latter guardian is frequently called a *conservator*, although this designation is not uniformly used throughout the United States. In some jurisdictions a further distinction is made between *general* (plenary) and *specific* guardianship. The latter guardian is restricted to exercising decisions about a particular area such as major or emergency medical procedures. The disabled person retains the freedom to make decisions about all other medical matters. By contrast, general guardians have total control over the disabled individual's person, estate, or both.

Under our system of law, an individual is presumed competent unless adjudicated incompetent. In many states, the Uniform Guardianship and Protective Proceeding Act or the Uniform Probate Code is used as a model for laws governing competency. The threshold requirement of incompetency is defined by the Uniform Guardianship and Protective Proceeding Act as:

impaired by reason of mental illness, mental deficiency, physical illness or disability, advanced age, chronic use of drugs, chronic intoxication, or other cause (except minority) to the extent of lacking sufficient understanding or capacity to make or communicate reasonable decisions.

Patients suffering from severe psychiatric disorders may meet the above definition. The standard of proof required for a judicial determination of incompetence is clear and convincing evidence. Although the law does not assign percentages to proof, clear and convincing evidence is in the range of 75 percent certainty.

States vary in the extent of their reliance on psychiatric assessments in determining incompetence. Nonmedical personnel including social workers, psychologists, family members, friends, colleagues, and even the individual who is the subject of the proceeding may testify.

Substituted Judgment Psychiatrists may find the time and effort required to obtain an adjudication of incompetence unduly burdensome. It can be inimical to providing competent treatment in a timely manner. Often, families are reluctant to face the formal court proceedings necessary to declare their family member incompetent, particularly when sensitive family matters may be disclosed. A common solution is to seek the legally authorized proxy consent of a spouse or relative acting as a de facto guardian when the refusing patient is thought to be incompetent. Proxy consent by next of kin, however, is not available in every state.

The President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research (1982) recommends that the relatives of incompetent patients be chosen as proxy decision-makers for three reasons: (1) the family is generally most concerned about the good of the patient; (2) the family is usually the most knowledgeable about the patient's goals, preferences, and values; and (3) the family deserves recognition as an important social unit to be treated, within limits, as a single decision-maker in matters that intimately affect its members. Some states allow proxy decision-making by statute. Some statutes merely state that another person may authorize consent on behalf of an incompetent patient; other statutes mention specific relatives.

Unless proxy consent is provided by state statute or by case law, good faith consents by next of kin should not be relied upon by psychiatrists treating patients believed to be incompetent. The appropriate procedure is to seek judicial recognition of a family member or other substitute decision-maker. Increasingly, state statutes are eliminating proxy consent as an option for mentally disordered individuals lacking health care decision-making capacity.

For psychiatric patients who continue to lack sufficient mental capacity to make health care decisions, an increasing number of states provide administrative procedures authorized by statute that permit involuntary treatment of incompetent and treatment-refusing mentally ill patients. All 50 states and the District of Columbia permit individuals to create a durable power of attorney, that is, one that endures even if the competence of the creator does not. A durable power of attorney agreement permits next of kin to consent to treatment on behalf of the patient who can no longer make such a decision.

Physician-Assisted Suicide With increasing legal recognition of physician-assisted suicide for terminally ill persons, psychiatrists are likely to be called upon to become gatekeepers. This role would be a radical departure from the physician's code of ethics that prohibits participation by an ethical doctor in any intervention that hastens death. Every proposal for physician-assisted suicide requires a psychiatric screening or consultation to determine the person's competency to commit suicide. The presence of psychiatric disorders associated with suicide, particularly depression, must be ruled out as the driving factor behind the request for physician-assisted suicide. The ethics of this controversial gatekeeping function are hotly disputed.

Lower court decisions in New York and Washington State have struck down prohibitions against physician-assisted suicide. Currently, most states have such prohibitions. In *Washington v Glucksberg*, the Supreme Court unanimously ruled that the Constitution does not guarantee Americans a right to commit suicide with the help of a physician, leaving the issue for state legislatures to decide. While acknowledging that terminally ill persons can endure great agony, the Court relied upon moral and legal arguments in its ruling, placing greater emphasis on traditional American condemnation of suicide and valuing life.

CRIMINAL PROCEEDINGS

Under the common law, the basic elements of a crime are (1) the mental state or level of intent to commit the act (known as the *mens rea* or guilty mind), (2) the act itself or conduct associated with committing the crime (known as *actus reus* or guilty act), and (3) a concurrence in time between the guilty act and the guilty mental state. The state must prove beyond a reasonable doubt that the defendant committed the criminal act with the requisite intent in order to convict a person of a particular crime.

A person's mental status and reality testing can play a critical role in determining whether a defendant is required to stand trial to face criminal charges, is acquitted of the alleged crime, is sent to prison, is hospitalized or be sentenced, or (in some extreme cases) is executed. Before any defendant can be criminally prosecuted, the court must establish that the accused is competent to stand trial, that is, defendants must understand the charges brought against them and have sufficient rational mental capacity to assist counsel with the defense. The United States Supreme Court in *Dusky v United States* established the legal standard for determining pretrial competency.

Competency to Stand Trial An impairment that puts into question a defendant's competence is usually associated with a mental disorder or defect. However,

persons may be found incompetent to stand trial even if they do not suffer from a mental disease or defect. The presence or absence of a mental illness is irrelevant if the defendant can meet competency requirements. Legal criteria, not medical or psychiatric diagnoses, govern competency. Diagnosis is relevant only to the question of restoring, with treatment, the defendant's competency to stand trial.

Checklists and structured interviews are available that assess specific psychological factors applicable to competency standards. The *Interdisciplinary Fitness Interview* can be used by both lawyers and mental health professionals. It permits a detailed examination of psychopathology and legal knowledge, using explicit scales to rate each response in the competency evaluation. *Evaluating Competencies: Forensic Assessments and Instruments* is also a standard reference in the field.

The presence of mental impairment does not automatically render the defendant incompetent. The impairment must be considered in the context of the particular case or proceeding. For example, mental impairment that renders an individual incompetent to stand trial in a complicated tax fraud case may not render that individual incompetent for a misdemeanor trial. Psychiatrists must be able to apply their clinical findings to the legal standards enunciated in *Dusky* when assessing competency to stand trial.

Insanity Defense The insanity defense is one of the most controversial issues in American jurisprudence. Defendants with mental impairments who are found competent to stand trial may seek acquittal on the claim of insanity, alleging that they were not criminally responsible for their actions at the time the offense was committed. The term “insanity” is a legal construct, not a psychiatric diagnosis.

The M'Naghten test of insanity was used in the vast majority of states until the 1960s. This test originated in 1843 from the first English case where the burgeoning science of psychiatry played a significant role. In a celebrated murder trial, Daniel M'Naghten (Fig. 54.1-1) was found to be insane and thus not guilty. M'Naghten shot and killed Edward Drummond, secretary to British Prime Minister Robert Peel. M'Naghten suffered from the delusion that Peel and the pope were conspiring against him. He intended to kill Peel, but shot Drummond by mistake. The judicial response to a parliamentary inquiry into the case produced the official M'Naghten test:

FIGURE 54.1-1 Daniel M'Naghten. His 1843 murder trial led to the establishment of rules still generally observed in legal insanity pleas. (Courtesy of Culver Pictures.)

The jurors ought to be told in all cases that every man is to be presumed to be sane, and to possess a sufficient degree of reason to be responsible for his crimes, until the contrary be proved to their satisfaction; and that to establish a defense on the ground of insanity, it must be clearly proved that, at the time of the committing of the act, the party accused was labouring under such a defect of reason, from disease of the mind, as not to know the nature and quality of the act he was doing; or, if he did know it, that he did not know he was doing what was wrong.

Criminals commit crimes for a variety of reasons. The law presumes that they do so rationally and of their own free will and thus merit some form of punishment. However, some offenders are so mentally disturbed that they are found to be incapable of acting rationally. Civilized societies have deemed it a violation of fundamental principles of fairness and morality to punish such persons. To do so would thwart two major tenets of punishment—retribution and deterrence.

No precise, generally accepted definition of legal insanity exists. Tests of insanity have always been controversial and have undergone much modification and refinement over the years. The insanity defense standard has four basic elements:

1. Presence of a mental disorder
2. Presence of a defect of reason
3. A lack of knowledge of the nature or wrongfulness of the act
4. An incapacity to refrain from the act

The presence of a mental disorder has remained the consistent core of the insanity defense; the other elements have varied in importance over time. Thus, the insanity defense standard varies throughout the United States, depending upon which state or jurisdiction has control over the defendant raising the defense.

The most stringent insanity standard is applied to defendants tried in a federal court who are governed by the Comprehensive Crime Control Act of 1984. This act was passed following public outrage over the acquittal of John Hinckley, Jr. (Fig. 54.1-2), by reason of insanity, after his attempt to assassinate President Ronald Reagan. The Act provides an affirmative defense to all federal crimes if, at the time of the offense, “the defendant, as a result of a severe mental disease or defect, was unable to appreciate the nature and quality or the wrongfulness of his acts. Mental disease or defect does not otherwise constitute a defense.” This standard eliminates the volitional or irresistible impulse prong of the four-part insanity defense stated above. It does not permit a defense based on a defendant's inability to conform his or her conduct to the requirements of the law. The defense is limited to defendants who cannot appreciate the wrongfulness of their acts (i.e., the cognitive prong of the defense). In a number of jurisdictions as well as in federal law (Comprehensive Crime Control Act), the burden of proof has shifted from the prosecution to the defense.

FIGURE 54.1-2 John Hinckley, Jr. In 1980, he shot President Ronald Reagan and Press Secretary James Brady. Hinckley was found not guilty by reason of insanity in 1982. The Brady Bill, a gun control bill, became law in 1993. (Courtesy of Wide World Photos.)

Despite the restrictions placed upon insanity defense standards following the Hinckley verdict, a number of states follow some version or accept totally the American Law Institute (ALI) test of insanity. The ALI test has two parts: a person is not responsible for criminal conduct if at the time of such conduct as a result of mental disease or defect he lacks substantial capacity either to appreciate the criminality (wrongfulness) of his conduct or to conform his conduct to the requirements of law.

As used in this article, the terms “mental disease or defect” do not include an abnormality manifested only by repeated criminal or otherwise antisocial conduct.

Thus, the ALI test contains both a cognitive and a volitional prong. Both John Hinckley, Jr., and Jeffrey Dahmer (Fig. 54.1-3) were tried under the ALI test. For example, Dahmer had struggled hard against his aberrant sexual impulses in the 7 years that elapsed between his first and second killings. However, the fact that Dahmer could plan his murders and systematically dispose of the bodies convinced the jury that he was able to control his behavior. All the testimony bolstered the notion that like most serial killers, Dahmer knew what he was doing and knew right from wrong. Finally, the jury did not accept the defense that Dahmer suffered from a mental illness to the degree that it had disabled his thinking or behavioral controls.

FIGURE 54.1-3 Jeffrey Dahmer. His murder trial for the deaths of 17 young men and boys gained widespread notoriety after accusations of cannibalistic practices were made. Dahmer was killed in prison by a psychotic inmate in 1994. (Courtesy of Wide World Photos.)

Nevada recently joined Montana, South Dakota, and Utah in abolishing the special plea of insanity altogether. At the same time, it established the plea of guilty but mentally ill. In Montana, evidence of insanity is admissible to negate *mens rea*.

In making an insanity determination, the threshold issue is not the existence of a mental disease or defect per se, but the lack of substantial capacity caused by it. The lack of mental capacity due to causes other than mental illness may suffice. For example, under certain circumstances, mental retardation may provide an adequate basis for an insanity defense.

Diminished Capacity The law recognizes shades of mental impairment that can affect *mens rea*, but not necessarily to the extent of completely nullifying it. The concept of “diminished capacity” was developed for these situations.

Diminished capacity allows the defendant to introduce medical and psychological evidence that relates directly to the *mens rea* for the crime charged, without having to assert a defense of insanity. For example, in the crime of assault with the intent to kill, psychiatric testimony may be permitted to address whether the offender acted with the purpose of committing homicide at the time of the assault. When a defendant's *mens rea* for the crime charged is not supported by clinical evidence, the defendant is acquitted only of that charge. The diminished capacity defense does not lead to total exculpation of criminal responsibility nor to automatic commitment to a mental institution.

Exculpatory and Mitigating Conditions Automatisms, intoxication, seizure disorders, and metabolic conditions have met with very limited success as exculpating or mitigating defenses to criminal charges.

Automatisms The automatism (or unconscious) defense recognizes that some criminal acts may be committed involuntarily. *Automatism*, defined as “having performed in a state of mental unconsciousness or dissociation without full awareness,” is applied to actions or conduct occurring “without will, purpose, or reasoned intention.”

For conviction for a crime, a criminal state of mind (*mens rea*) must be accompanied by the commission of a prohibited act (*actus reus*). The physical act must be conscious and volitional. Statutory and common law in most jurisdictions specifically excludes from the *actus reus* a reflex or convulsion, a bodily movement occurring during unconsciousness or sleep, acts during hypnosis or resulting from hypnotic suggestion, and a bodily movement that is not conscious and voluntary.

The classic, though rare, example of an automatism is the person who commits an offense while sleepwalking. Courts have held that such individuals do not have conscious control of their physical actions and therefore act involuntarily. This defense exists for persons committing a crime during a state of unconsciousness caused by a concussion following a head injury, involuntary ingestion of drugs or alcohol, hypoxia or hypoglycemia, or epileptic seizures.

There are clear limitations to the automatism defense. The automatism defense is not available if the person was aware of the condition prior to the offense and failed to take reasonable steps to prevent the crime. If, for example, a defendant with a known history of poorly controlled epileptic seizures loses control of a car during a seizure and kills another person, that defendant cannot assert the defense of automatism.

Generally, intoxication is not a defense to a criminal charge because it usually results from a person's own actions. However, most states consider voluntary alcoholism relevant in determining whether the defendant possessed the *mens rea* necessary to commit a specific-intent crime or whether there was premeditation in a crime of murder. The mere fact that the defendant was voluntarily intoxicated does not justify a finding of automatism or insanity.

A distinct difference exists when, because of chronic, heavy alcohol use, the defendant is suffering from an alcohol-induced organic mental disorder. If credible clinical evidence is presented that an alcohol-related psychiatric disorder such as alcohol hallucinosis, withdrawal delirium, amnestic disorder, or dementia associated with alcoholism caused significant cognitive or volitional impairment, a defense of insanity or diminished capacity might be upheld.

Another mental state defense occasionally raised by criminal defendants is that the assaultive behavior was involuntarily precipitated by a seizure disorder. This condition is often diagnosed as temporal lobe epilepsy. Aggressive behavior is asserted to primarily result from an uncontrollable, randomly occurring, abnormal brain dysrhythmia. The legal argument that such individuals should not be held accountable for their actions may prevail. Despite its simplicity and rare success in the courts, little significant empirical data support these claims.

Defenses based on metabolic disorders rarely succeed. However, the notorious “Twinkie defense” was part of a successful strategy to defend Dan White (Fig. 54.1-4), charged with two counts of murder in the shooting of San Francisco Mayor George Moscone and Supervisor Harvey Milk. The jury returned a verdict of voluntary manslaughter. White's defense was based on the theory that the ingestion of large amounts of sugar contributed to a state of temporary insanity. Public outrage over the verdict in the Dan White case led to the repeal of the diminished capacity defense by the California legislature in 1981.

FIGURE 54.1-4 Dan White. The former San Francisco supervisor killed San Francisco Mayor George Moscone and Supervisor Harvey Milk at City Hall in 1978. White's “Twinkie defense” helped reduce his crime for murder to manslaughter, for which he served 5 years. White committed suicide after he was released from prison. (Courtesy of Wide World Photos.)

Guilty But Mentally Ill A number of states have established an alternative verdict of guilty but mentally ill. Under guilty but mentally ill statutes, if the defendant pleads not guilty by reason of insanity, this alternative verdict is available to the jury. Under an insanity plea, four outcomes are possible: not guilty, not guilty by reason of insanity, guilty but mentally ill, and guilty.

The problem with guilty but mentally ill is that it is an alternative verdict without a difference. It is basically the same as finding the defendant just plain guilty. The court must still impose a sentence upon the convicted person. Although the convicted person supposedly will receive psychiatric treatment if necessary, this treatment provision is available to all prisoners. Moreover, the frequent unavailability of appropriate psychiatric treatment for prisoners adds an additional element of spuriousness to the guilty but mentally ill verdict. Guilty but mentally ill is essentially no different than finding a defendant “guilty but suffering from lumbago.” Moreover, research shows that persons found guilty but mentally ill have fewer furloughs and longer sentences.

CIVIL LITIGATION

Psychiatrists provide consultation and testimony in a number of areas of civil litigation including malpractice, personal injury, discrimination, child custody, will contests, workers' compensation, disability claims, toxic torts, and a wide variety of other cases where psychiatry and law intersect. Psychiatrists are frequently asked by lawyers to evaluate individuals with alleged posttraumatic stress disorder. Specific guidelines are available for the forensic assessment of posttraumatic stress disorder in litigation.

Forensic psychiatry is now an officially recognized subspecialty. The American Board of Psychiatry and Neurology now offers added qualifications in forensic psychiatry. Forensic psychiatry is defined by the American Academy of Psychiatry and the Law as follows: “Forensic psychiatry is a subspecialty in which scientific and clinical expertise is applied to legal issues in legal contexts embracing civil, criminal, correctional or legislative matters.”

Psychiatrists who enter the legal arena must be aware of the fundamentally different roles of a treating psychiatrist and the forensic psychiatric expert. Treatment and expert roles do not mix. Unlike the orthopaedist, for example, who usually possesses objective, concrete information such as the X-ray of a broken limb to demonstrate physical injuries in court, the treating psychiatrist relies primarily upon the subjective reporting of the patient. In the clinical context, psychiatrists work with the patient's perception of the problems (narrative truth), not necessarily objective reality (historical truth). To maintain confidentiality, most psychiatrists do not speak to third parties to corroborate their statements. The law, however, is concerned with proofs that can be reasonably established by facts. Uncorroborated, subjective patient reporting will likely be attacked in court as speculative, self-serving, and unreliable.

The treating psychiatrist is, and should be, a total ally of the patient. Psychiatrists cannot effectively treat patients that they dislike. This positive bias in favor of the patient is a proper treatment stance that fosters a working therapeutic alliance. In the clinical situation, the psychiatrist's attention is directed toward the diagnosis and treatment of mental disorders. This again is an appropriate focus for the treating psychiatrist.

In court, credibility is a critical asset. Opposing counsel will pursue every opportunity to portray the treating psychiatrist as a subjective mouthpiece for the plaintiff—which may or may not be true. Also, court testimony by the treating psychiatrist may disclose information that may not be legally privileged but is perceived as private and confidential by the patient. Such a disclosure by the therapist is likely to seriously harm the therapeutic relationship. In addition, psychiatrists should inform patients about the consequences of releasing treatment information, particularly in legal matters. Section 4, annotation 2 of the *Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry* (1995) states:

The continuing duty of the psychiatrist to protect the patient includes fully apprising him/her of the connotations of waiving the privilege of privacy. This may become an issue when the patient is being investigated by a government agency, is applying for a position, or is involved in legal action.

When the treating psychiatrist testifies concerning the patient's need for future treatment, an inherent conflict of interest emerges. Therapists stand to benefit economically from treatment recommendations for their patients. Although this may not be the intention of the psychiatrist at all, opposing counsel will be sure to point out that the psychiatrist has a financial interest in the case.

Although opposing counsel may depict forensic experts as hired guns, they are free of conflicting treatment roles. No doctor-patient relationship is created during forensic evaluation with its accompanying treatment biases toward the patient. The expert usually can review a wide variety of records and interview numerous people who know the litigant. Furthermore, the forensic expert considers the possibility of exaggeration or malingering because of the litigation context and the absence of treatment bias. The forensic psychiatrist does not have a conflict of economic interest when recommending treatment for the litigant.

Treating psychiatrists should remain solely in treatment roles with their patients. If testimony on behalf of a patient is required, the psychiatrist should testify as a fact witness rather than as an expert witness. As a fact witness, the psychiatrist will be questioned about the number and length of visits, diagnosis, and treatment. No opinion evidence concerning causation of the injury or extent of damages is usually given. In some jurisdictions, however, the court may attempt to convert a fact witness into an expert witness when the treating psychiatrist takes the stand. Psychiatrists must remain ever mindful of the many double-agent roles that can develop when mixing psychiatry and litigation.

SUGGESTED CROSS-REFERENCES

Competence is linked to delirium and dementia, discussed in Chapter 10; to schizophrenia, discussed in Chapter 12; to delusional disorders, discussed in Section 13.2; to impulse-control disorders, discussed in Chapter 22; and to Alzheimer's disease, discussed in Section 51.3c. Factitious disorders are discussed in Chapter 17 and malingering in Section 27.2. Criminality is discussed in Section 27.3. Emergency psychiatry, including suicide, is discussed in Chapter 29. Biological therapies are discussed in Chapter 31. Mental retardation is discussed in Chapter 34. Child abuse is discussed in Section 49.4. Medical-legal issues and ethical issues in geriatric care are discussed in Section 51.6b and Section 51.6c, respectively. Hospital and community psychiatry are discussed in Chapter 52. Ethics in psychiatry is discussed in Section 54.2.

SECTION REFERENCES

American Psychiatric Association: *The Practice of Electroconvulsive Therapy: Recommendations for Treatment, Training and Privileging*. A task force report. American Psychiatric Association, Washington, DC, 1990.

American Psychiatric Association: *Tardive Dyskinesia: A Task Force Report of the American Psychiatric Association*. American Psychiatric Association, Washington, DC, 1992.

American Psychiatric Association: *The Principles of Medical Ethics with Annotations Especially Applicable to Psychiatry*. American Psychiatric Association, Washington, DC, 1995.

*Appelbaum PS: *Almost a Revolution: Mental Health Law and the Limits of Change*. Oxford University Press, New York, 1994.

Appelbaum PS, Gutheil TG: *Clinical Handbook of Psychiatry and the Law*. Williams & Wilkins, Baltimore, 1991.

Epstein RS: *Keeping Boundaries: Maintaining Safety and Integrity in the Psychotherapeutic Process*. American Psychiatric Press, Washington, DC, 1994.

Grisso T: *Evaluating Competencies: Forensic Assessments and Instruments*. Plenum, New York, 1986.

Grisso T, Appelbaum PS: The MacArthur Treatment Competence Study. III: Abilities of patients to consent to psychiatric and medical treatments. Law Hum Behav *19:*149, 1995.

Grisso T, Appelbaum PS: Comparison of standards for assessing patients' capacities to make treatment decisions. Am J Psychiatry *152:*1033, 1995.

Gutheil TG, Simon RI: Between the chair and the door: Boundary issues in the therapeutic “transition zone.” Harvard Rev Psychiatry *2:*336, 1995.

Gutheil TG, Simon RI: Risk management principles in recovered memories cases: The importance of the clinical foundation. Psychiatr Serv J *48:*1403, 1997.

*Jacobs DG, editor: *The Harvard Medical School Guide to Suicide Assessment and Intervention*. Jossey-Bass, San Francisco, 1999.

Melton GB, Petrilla J, Poythress NG, Slobogin C: *Psychological Evaluations for the Courts*, ed 2. Guilford, New York, 1997.

Miller RD: Criminal responsibility. In *Principles and Practice of Forensic Psychiatry*, R Rosner, editor. Chapman & Hall, New York, 1994.

Mishkin B: Determining the capacity for making health care decisions. Adv Psychosom Med *19:*151, 1989.

Monahan J: Mental disorder and violent behavior: Perceptions and evidence. Am Psychol *47:*511, 1992.

Pope KS: *Sexual Involvement with Therapists: Patient Assessment, Subsequent Therapy, Forensics*. American Psychological Association, Washington, DC, 1994.

Quen JM, editor: *The Psychiatrist in the Courtroom*. Selected papers of Bernard L. Diamond, M.D. Analytic Press, Hillsdale, 1994.

Reisner R, Slobogin C: *Law and the Mental Health System. Civil and Criminal Aspects*, ed 2. West, St. Paul, MN, 1990.

*Rosner R, editor: *Principles and Practice of Forensic Psychiatry*. Chapman & Hall, New York, 1994.

Sederer LI, Ellison J, Keyes C: Guidelines for prescribing psychiatrists in consultative, collaborative, and supervisory relationships. Psychiatr Serv *49:*1197, 1998.

Simon RI: *Bad Men Do What Good Men Dream: A Forensic Psychiatrist Illuminates the Darker Side of Human Behavior*. American Psychiatric Press, Washington, DC, 1996.

*Simon RI: Chronic posttraumatic stress disorder: a review and checklist of factors influencing prognosis. Harv Rev Psychiatry *6:*304, 1999.

Simon RI: Clinical risk management of the rapid cycling bipolar patient. Harv Rev Psychiatry *4:*245, 1997.

Simon RI: *Clinical Psychiatry & the Law*, ed 2. American Psychiatric Press, Washington, DC, 1992.

Simon RI: Clinical risk management of suicidal patients: Assessing the unpredictable. In *American Psychiatric Press Review of Clinical Psychiatry and the Law*, vol 3, RI Simon, editor. American Psychiatric Press, Washington, DC, 1992.

Simon RI: The credible forensic psychiatric evaluation in sexual harassment litigation. Psychiatr Ann *26:*139, 1996.

Simon RI: Discharging sicker, potentially violent psychiatric inpatients in the managed care era: Standard of care and risk management. Psychiatr Ann *26:*726, 1997.

Simon RI: Managing malpractice liability in managed care settings: Risks and remedies in psychiatric care. Primary Psychiatry *5:*91, 1998.

Simon RI: Psychiatrists awake! Suicide risk assessments are all about a good night's sleep. Psychiatr Ann *28:*479, 1998.

Simon RI: Psychiatrists' duties in discharging sicker and potentially violent inpatients in the managed care era. Psychiatr Serv J *49:*62, 1998.

*Simon RI: The suicidal patient: Accepting patient risk and therapist anxiety. In *The Mental Health Clinician and the Law: A Comprehensive Handbook*, LE Lifson, RI Simon, editors. Harvard University Press, Cambridge, MA, 1997.

Simon RI: The Suicide Prevention Pact: Clinical and legal considerations. In *American Psychiatric Press Review of Clinical Psychiatry and the Law*, vol 2, RI Simon, editor. American Psychiatric Press, Washington, DC, 1992.

Simon RI: Treatment boundary violations: Clinical, ethical and legal considerations. Bull Am Acad Psychiatry Law *20:*269, 1992.

Simon RI, Gutheil TG: Ethical and risk management principles in recovered memories cases: Maintaining therapist neutrality. In *Trauma and Memory: Clinical and Legal Controversies*, P Appelbaum, L Uyehara, M Ellin, editors. Oxford University Press, Cambridge, 1997.

*Simon RI, Sadoff RL: *Psychiatric Malpractice: Cases and Comments for Clinicians*. American Psychiatric Press, Washington, DC, 1992.

Simon RI, Wettstein RM: Toward the development of guidelines for the conduct of forensic psychiatric examinations. J Am Acad Psychiatry Law *25:*17, 1997.

*Slovenko R: *Psychiatry and Criminal Culpability*. Wiley, New York, 1995.

*Stone AA: The verdict against Myron Liptzin—Who sets the standard of care? Psychiatr Times *16:*32, 1999.

Strasburger LH, Gutheil TG, Brodsky A: On wearing two hats: Role conflict in serving as both psychotherapist and expert witness. Am J Psychiatry *154:*448, 1997.

54.2 ETHICS IN PSYCHIATRY

PETER B. GRUENBERG, M.D.

As used in this section, ethics in psychiatry is not the same as psychiatric ethics. While there is a body of wisdom referred to as ethics, and a subset of ethics called medical ethics, and a further subset of ethics called psychiatric ethics, ethics in psychiatry deals mainly with the ideal behavior of the ethical psychiatrist and the ethical principles that determine the expected behavior of the ethical psychiatrist.

During the past decade, psychiatric ethics has, unfortunately, had to define itself. Psychiatric ethics has had to resist politicization from many diverse quarters. Recently, some opponents of managed care have suggested that anyone—psychiatrist or businessman—who advocates or participates in a managed care plan is, per se, unethical. At the other extreme, some proponents of managed care presume that a psychiatrist practicing fee-for-service medicine is unethical.

The public became painfully aware of psychiatric ethics because of several celebrated cases that were the object of national attention. Pundits told their viewers and readers what is, and what is not, ethical behavior for a psychiatrist, and sadly, they were often correct. Thus one must review the basic principles of ethics and trace the development of current thinking as it relates to psychiatric ethics.

BASIC ETHICAL PRINCIPLES

A survey of basic ethical principles in psychiatry was summarized by James Rosenberg and Spencer Eth as follows:

Psychiatric ethics is a body of rules and principles in a state of flux, adapting to changes in psychiatry and psychiatry's place in the world at large.

Ethics and Morals The semantic distinction between morals and ethics generally refers to those values a society accepts as universal principles contrasted with the principles that are relative to a particular context or era. General moral principles govern the behavior of all citizens and are deemed both universal and immutable. In Judeo-Christian society the prevailing standard is that each person should treat every other person as that person would wish to be treated by others. Not surprisingly, overlap with the law is considerable.

The term “ethics” is usually reserved for the moral principles restricted to certain groups, such as those in a profession. That role-bound morality can consist of internal or external standards of ethical conduct. For the psychiatric profession, *The Principles of Medical Ethics with Annotations Especially for Psychiatry* (thereafter, *The Principles*), developed by the American Psychiatric Association (APA), is an example of an internal standard used by the profession's major organization to regulate the behavior of its members. Judicial, legislative, or executive bodies may impose external standards as well.

Distinguishing between an ethical dilemma and an ethical violation is important. One is faced with an ethical dilemma when asked to choose between two ethically legitimate alternatives, such as preserving patient confidentiality or protecting endangered third parties. An *ethical conflict* involves the compromise of an ethical principle, usually because of self-interest, such as patient-therapist sexual relations.

APA Sanctions For ethical violations the APA may expel members from the organization or, for less severe violations, suspend membership for a time. During that time a member may be required to undergo supervision or extra training. For still less severe violations a member may be reprimanded or admonished, with no effect on membership status. Expulsion or suspension from the APA is publicly reported. Further, such actions must be reported to the National Practitioners Data Bank.

Core Principles Which of the four core ethical principles—autonomy, nonmaleficence, beneficence, and justice—dominates depends on the particular situation. On one hand, respect for patient autonomy is salient when psychosurgery is proposed, but the principle of social justice may be in force in deciding who receives an expensive new medication that is scarce. The most pervasive ethical conflict in psychiatry and in medicine as a whole is that between *autonomy*, the right of patients to self-determination, and *beneficence*, the duty of physicians to act in the best interest of their patients.

Autonomy The principle of *patient autonomy* is of central importance and, conceptually, is in many ways coextensive with the legal concept of competence. A patient makes an autonomous choice by giving informed consent when that choice is (1) intentional, (2) free of undue outside influence, and (3) made with rational understanding. Usually, when patients respond to a choice by saying “yes,” the desire to comply is assumed. However, that assumption may not be valid with a highly confused patient.

Nonmaleficence *Nonmaleficence* is the duty of the psychiatrist to avoid either inflicting physical and emotional harm on the patient or increasing the risk of such harm. That principle is captured by *primum non nocere*, “first, do no harm.”

Beneficence The principle of *beneficence*—to prevent or remove harm and promote well-being—was the primary driving principle of medical and psychiatric practice throughout history until the rise of consumerism and other factors in the late 1960s. The expression of the principle is paternalism, the use of the psychiatrist's judgment as to what course of action is best for the patient or research subject.

Psychiatrists have historically justified beneficent paternalism on the basis of their superior knowledge, greater objectivity, and powerful desire to help their patients. A common example of the primacy of beneficence is the decision to hospitalize a suicidal patient involuntarily. That unilateral act implies either that the patient's autonomy is intact but must be overridden by the potential for dangerous behavior or that the patient's capacity for autonomous choice is impaired by the mental illness and must be ignored. Here is an example of the distinction between *weak paternalism*, acting beneficently when the patient's impaired faculties prevent an autonomous choice, and *strong paternalism*, acting beneficently despite the patient's intact autonomy.

Guidelines have been proposed for permitting beneficence to overrule patient autonomy: when the patient faces substantial harm or risk of harm, the paternalistic act is chosen that ensures the optimal combination of maximal harm reduction, low added risk, and minimum necessary infringement on patient autonomy. For example, convincing an acutely suicidal patient to enter the hospital voluntarily confers the environment necessary to contain the suicidal behavior, poses little added risk for most patients, and preserves more patient autonomy than involuntary hospitalization or a breach of confidentiality to an intervening third party on an outpatient basis.

Justice Like the principles of autonomy, nonmaleficence, and beneficence, the principle of justice in psychiatry does not operate in a vacuum but is responsive to the ever-changing social, political, religious, and legal mores of the moment.

HISTORY

The changing ethical landscape of psychiatric research and practice can perhaps best be understood as the continuous revision of the basic skeleton of Hippocratic teachings. The elements of the original Hippocratic oath include the following: medicine is a closed profession; use those treatments in which you are skilled and defer to specialists those in which you are not; maximize patient health and minimize harm; avoid sexual relations and other exploitative self-interests; confidentiality; and penalties for violations. Several factors were responsible for the explosion of interest in the psychiatric profession's ethical conduct in the 1970s. In the long

history of mental health care, that era marked the first time that patient autonomy—in the form of consumerism and civil rights activism, including the women's movement—encroached on the tradition of medical paternalism.

Modern psychiatry is dotted with well-publicized cases of patient abuse under the guise of paternalism. Examples include lobotomy, the shock therapies, lysergic acid diethylamide (LSD) experimentation, and the commitment of Soviet dissidents to mental institutions for the purpose of political oppression. The profession has seen major judicial and governmental intrusions into its practice in such areas as confidentiality and involuntary hospitalization and treatment.

During the Second World War, charges were made that involved the behavior of physicians acting under the direction of the Third Reich of Germany. Some medical leaders felt that the Hippocratic Oath did not adequately address some grossly inhumane behavior of the German doctors. In the 1950s and 1960s the Oath of Geneva at medical school graduations sometimes replaced the Oath of Hippocrates. This was, perhaps, the first time in 2000 years that the medical profession began to question and modify its own ethical models.

In the 1970s two social movements came together to sound a wake-up call to the medical profession. The rise of both consumerism and feminism took the entire medical profession to task. Paternalism gave way to autonomy. Patients had rights and they knew it! A hospital patient who asked a nurse “What's that pill?” would no longer tolerate the answer: “It's what Doctor is giving you to make you better.” Women would no longer tolerate sexual behavior from or with their doctors. They protested, they formed self-help groups, and they played a large role in establishing ethics committees at the district branch levels of the APA.

PROFESSIONAL CODES

Most professional organizations and many business groups have codes of ethics. Such codes reflect a consensus about the general standards of appropriate professional conduct. The American Medical Association's (AMA) Principles of Medical Ethics, the APA's *Principles*, and the American College of Physicians Ethics Manual articulate ideal standards of practice and professional virtues of practitioners. These codes include exhortations to use skillful and scientific techniques, to self-regulate misconduct within the profession, and to respect the rights and needs of patients, families, colleagues, and society. Such exhortations are reinforced by ethical principles such as beneficence, utility, autonomy, respect, and justice.

In recent years, there has been increased interest in the use of professional codes of ethics as a standard of criticism and as a means of regulating professional misconduct. Some hospitals withdraw privileges from a psychiatrist who has been expelled for ethical reasons, and malpractice coverage may be denied.

The 1995 edition of *The Principles* provides a useful and comprehensive example of a professional code geared to the psychiatric profession. The manual covers a broad array of psychiatric ethical issues, from fee splitting and sex with former patients to psychiatrists' participation in executions, all of which are unethical (Table 54.2-1). Another useful manual is the 1995 edition of the *Opinions of the Ethics Committee on the Principles of Medical Ethics With Annotations Especially Applicable to Psychiatry*. This manual, developed by APA Ethics Committee members and consultants, provides answers to commonly asked ethical questions. A summary of ethical questions and answers is provided in Table 54.2-2.

Table 54.2-1 The Principles of Medical Ethics With Annotations Especially Applicable to Psychiatry

Table 54.2-2 Ethical Questions and Answers

The Principles, if carefully read, will suggest to the student and practitioner certain behaviors that are expected of an ethical psychiatrist and certain behaviors deemed by the profession to be unethical and unacceptable. The APA and its district branches have a carefully worked-out set of procedures to be followed if a member of that organization is accused of violating the ethical principles. While these procedures are not perfect, they seem to have served the profession well for more than 20 years. Further, they are changed in response to changes in federal and state laws and the oversights that have become apparent.

From a practical point of view, the transgressions that psychiatrists most frequently commit fall into a few categories, and certain activities may fit in more than one category:

1. Sexual boundary violations;
2. Nonsexual boundary violations;
3. Violations of confidentiality;
4. Mistreatment of the patient (incompetence, double agentry); and
5. Illegal activities (insurance, billing, insider stock trading).

SEXUAL BOUNDARY VIOLATIONS

Sexual activity with a current or former patient is unethical. The rest is commentary.

By far the most publicized ethical transgression of psychiatrists involves sexual activity with a patient or a former patient. A variety of defenses have been advanced by psychiatrists accused of such activity; yet it should be obvious what a sexual activity is, what a patient is, and what a former patient is.

The Principles is quite clear on the subject. Sexual activity with a current or former patient is unethical. Glen Gabbard has studied a great many psychotherapists who have gotten into trouble because of sexual involvement with patients. He categorized them into some interesting groups including psychotic disorders, predatory

psychopathy and paraphilias, lovesickness, and masochistic surrender.

Clearly, with such different types of psychotherapists involved in sexual activity with patients, no one pattern of behavior can provide a danger signal. Yet, if one studies complaints brought to professional societies, a rather distinctive pattern does emerge, the slippery slope of smaller violations leading to greater violations.

> A male therapist is treating a female patient who may have some borderline personality traits. During a particularly emotional session, the doctor holds the patient's hand or puts his arm around her shoulder to comfort her. Perhaps there is even a hug; later, another emotional session and another hug, this time a bit longer. Soon each session ends with a hug, maybe a hug and a kiss on the cheek. Later the patient calls the doctor in the evening, sobbing. The doctor makes a house call; more hugs, more kisses. A suicide gesture is often catalytic. This scenario can get quite complicated and involved but it is not uncommon. At this point the situation is, without question, a psychiatric emergency. The doctor must obtain an immediate consultation from an experienced colleague. Unfortunately, this option is rarely taken. Whether due to embarrassment, fear of criticism from the community, denial, or unwarranted omnipotence, the scenario usually progresses down the slippery slope, and sexual activity replaces treatment.

An argument occasionally put forward is that the sexual activity is consensual. That is, it takes place between two autonomous adults, and consequently they should be able to do as they please. Yet even a rudimentary knowledge of dynamic psychiatry reveals that the two adults are rarely autonomous. Powerful, unconscious forces are at work in all doctor-patient relationships. The patient does not have to lie on a couch five times a week to develop transference. Transference occurs in every encounter with a patient. Even the brief medication-management visit produces transference (and countertransference). *The Principles* is quite clear in Section 2.1.

Although not spelled out in *The Principles*, sexual activity with a patient's family member is also unethical. This is most important when the psychiatrist is treating a child or adolescent. Most training programs in child and adolescent psychiatry emphasize that the parents are patients too and that the ethical and legal proscriptions apply to parents (or parent surrogates) as well as to the child. Nevertheless, some psychiatrists misunderstand this concept. Sexual activity between a doctor and a patient's family member is also unethical.

NONSEXUAL BOUNDARY VIOLATIONS

The preceding section points out that any sexual activity with a current or former patient is unethical. That is simple, straightforward, and unambiguous. Nonsexual boundary violations are far less clear and far less obvious. The overriding principle is that exploitation of the doctor-patient relationship or therapist-client relationship for purposes other than treatment is probably unethical. In their definitive book *Boundaries and Boundary Violations in Psychoanalysis* Gabbard and Eva Lester describe many types of boundaries in doctor-patient relationships. The relationship between a doctor and a patient for the purposes of providing and obtaining treatment is what is usually called the *doctor-patient relationship*. That relationship, for that purpose, has a boundary around it and boundaries within it. Either person may cross the boundary.

Not all boundary crossings are boundary violations. For example, a patient may say to a doctor at the end of an hour “I have left my money at home and I need a dollar to get my car out of the garage. Will you lend me a dollar until next time?” The patient has invited the doctor to cross the doctor-patient boundary and set up a lender-borrower relationship as well. Depending upon the doctor's theoretical orientation, the clinical situation with the patient, and other factors, the doctor may elect to cross the boundary. One can then debate whether the boundary crossing is also a boundary violation. A *boundary violation* is a boundary crossing that is also exploitative. It gratifies the doctor's needs at the expense of the patient. The doctor is responsible for preserving the boundary and for seeing to it that boundary crossings are held to a minimum and that exploitation does not occur. Harm to the patient is not a component of a boundary violation. For example, using information supplied by the patient, (e.g., a stock tip) is an unethical boundary violation although no obvious harm may come to the patient. For purposes of discussion, nonsexual boundary violations may be grouped into several arbitrary (overlapping and not mutually exclusive) categories.

Business Almost any business relationship with a former patient is problematic, and almost any business relationship with a current patient is unethical. Naturally, the circumstance and location may play a significant role in this admonition. In a rural area or a small community, one might be treating the only pharmacist (or plumber or couch upholsterer) in town; then one does business with the pharmacist/patient and tries to keep boundaries in check. Ethical psychiatrists try to avoid doing business with a patient or a patient's family member or asking a patient to hire one of their family members. Ethical psychiatrists avoid investing in a patient's business or collaborating with a patient in a business deal. A recent nonfiction book, *Hollywood on the Couch*, describes in embarrassing detail a series of alleged boundary violations committed by famous psychiatrists whose patients were prominent in motion pictures, television, and the stage. Psychiatrists in this well-documented book became involved in the businesses of their patients, much to the detriment of the treatment. Psychiatrists served as “advisors” to their patients' movies, and celebrities were appointed to “Scientific Advisory Boards” of their therapists.

> A successful businessman, Mr. Z. was treated by Dr. A. Sometime after termination, Dr. A. contacted Mr. Z. to initiate a social relationship. Dr. A. called Mr. Z. “Charlie” and Mr. Z. called Dr. A. “Dr. A.” Mr. Z. told Dr. A. that several potential investors were interested in a business that Mr. Z. was starting. Dr. A. said that he would be interested in becoming a principal investor. A deal was struck. Slowly Dr. A. bought more of the business, and despite his promise not to do so, Dr. A. took over control of the business. The business grew because of the acumen of Mr. Z., but Dr. A. became even greedier and attempted a total coup. In the process of taking over the business, Dr. A. used knowledge gained while treating Mr. Z. (some of his strengths and some of his weaknesses). He used the still-present transference in convincing Mr. Z. to turn over still more control to Dr. A., (Mr. Z. still called him “doctor,” and Dr. A. still called him “Charlie”). A lawsuit resolved matters in Mr. Z.'s favor. Mr. Z. now has an abiding mistrust of psychiatrists.

Solicitation of Funds Solicitation of funds from current or former patients presents many ethical problems and financial dilemmas. Psychiatrists' knowledge of interpersonal relationships and especially their knowledge of unconscious factors that affect behavior, make this the only medical specialty troubled by the solicitation of funds from current and former patients. During the past decade, as public funding for research and the support of mental health facilities diminished, directors had to look to other sources for money. One source that has been tapped is the affluent patient. Medical and surgical colleagues have sought financial support from their patients for decades and evidently thought little of the ethical implications of doing so. Now, private psychiatric hospitals and psychiatric departments of medical schools and general medical hospitals have gotten on the fund-raising train. Foundations have been set up by local and national mental health organizations. The problem arises when psychiatrists associated with such organizations seek donations from their patients or former patients. *The Principles* states that psychiatrists should not exploit information furnished by their patients (e.g., financial status and philanthropic proclivities) and should not use their position of power to influence patients (e.g., what to do with their spare cash). Notice that in these examples, the doctor might not even directly benefit from such solicitations. However, there have been instances in which a psychiatrist has attempted to influence a patient to contribute funds to the doctor's own foundation or research projects. In one highly publicized case, a psychiatrist persuaded a patient to give him a generous annual stipend so that he could be free to do his uncompensated research. The APA sanctioned him. The chair of a university psychiatry department wrote to his faculty urging them to contact some of their affluent and grateful patients and solicit money for his department. An ethical psychiatrist-professor wrote to the APA Ethics Committee asking for guidance. The committee suggested that this request was inappropriate because it encouraged faculty members to violate *The Principles* through undue influence and breaking confidentiality. The scheme was dropped.

Ideological Issues Ideological issues can cloud one's judgment and may lead to ethical lapses. Any clinical decision should be based on what is best for the patient; the psychiatrist's ideology should play as little a part as possible in such a decision. The ethical psychiatrist, then, must attend to current research, new approaches, and the results of such approaches. An obvious example of this problem is the psychiatrist who is convinced that an analytic approach is the only way to treat such conditions as major depression, bipolar disorder, and schizophrenia. Current research suggests that other approaches are as useful or more useful in many cases. The patient's welfare and the patient's wishes must be considered in such a situation. A psychiatrist who is consulted by a patient with an illness such as this should tell the patient what forms of treatment are available to treat the illness and allow the patient to decide upon a course of treatment. Naturally, psychiatrists should recommend the treatment that they feel is in the best interest of the patient, but ultimately, the patient should be free to choose.

Social The particular locale and circumstances must be considered in any discussion of the behavior of an ethical psychiatrist in social situations. The overarching principle is that the boundaries of the psychiatrist-patient relationship should be respected. Further, if options exist, they should be exercised in favor of the patient. Problems often arise in treatment situations when friendships develop between the psychiatrist and the patient. Objectivity is compromised, therapeutic neutrality is impaired, and factors outside the consciousness of either party may play a destructive role. Such friendship should be avoided during treatment. Similarly, psychiatrists should not treat their social friends for the same set of reasons. Obviously, in an emergency, one does what one must.

Hollywood on the Couch is an entertaining look at how not to conduct a practice. The authors describe many doctor-patient relationships in which the boundaries between social relationships, business relationships, sexual relationships, and the putative doctor-patient relationship became so blurred that meaningful treatment

was impossible.

Special attention must be paid to the temptation to socialize with the patient's family or to allow a patient to establish a relationship with the doctor's family. This is especially worrisome when the patient is a child and social or sexual relationships are established with the patient's parent.

It is tempting to ask favors of a patient, especially if the favor involves no great effort from the patient. Yet the unintended results may be upsetting. The slippery slope phenomenon operates in this area as well. Asking a patient to mail a letter may seem innocuous. To whom is the letter addressed? To another patient? To the Medical Board? Just to the phone company? If it is mailing a letter this week, is it getting the doctor a coffee next week? If the patient is a car dealer might the doctor ask for a special deal?

A psychoanalyst treated an attorney for several years and both felt that the treatment went well. About a year after termination, the doctor's son graduated from law school. Lawyer positions were hard to obtain. By then the former patient was in charge of a large legal organization of just the type that the doctor's son wanted to join. Would it have been proper for the doctor to contact his former patient to see if a door could be opened? Could the doctor have suggested that his son contact the former patient and mention the fact that he was that doctor's son? Certainly no direct harm would have come to the former patient (assuming the doctor's son was competent). Yet, *The Principles* prohibits such a boundary crossing because it would be exploitative and would be influencing the former patient in a way not directly relative to treatment. (See Principles 1.1 and 2.2 in Table 54.2-1.)

Financial For psychiatrists who practice in the private sector, dealing with the patient about money is a part of treatment. Issues surrounding setting the fee, collecting the fee, and other financial matters are grist for the mill. Even so, ethical concerns must be observed. *The Principles* advises the doctor on such matters as charging for missed appointments and other contractual problems. Ethics complaints against doctors are frequently precipitated by financial issues; thus the doctor must recognize the power that these issues have in the therapeutic relationship. Since the psychotherapeutic relationship is so much like a social relationship—the office looks like a living room, the doctor wears regular clothes, some patients might, without recognizing it, assume that a friendship exists that forgives payment of a fee. When the bill is presented, feelings, even though they are unconscious, are ruffled. The idea that psychiatric services are dispensed in a contractual context cannot be emphasized enough. Early in their careers psychiatrists are often reluctant to discuss fees openly out of a sense of embarrassment over discussing money or a sense of protecting the patient. A prominent psychiatrist once commented that most of his patients could discuss naked aggression, murderous fantasies, and unusual sexual reveries more easily than they could discuss their personal finances.

How an ethical psychiatrist handles the situation when a patient temporarily or permanently runs out of money is important. There are many options—some more problematic than others. The psychiatrist can certainly lower the fee, but caution is needed because a fee lowered to the point where the treatment is not somehow being compensated may evoke countertransference resentment. The number of patients being seen at a reduced fee is a similar consideration. Running up a bill can also be a problem. Is there an expectation of eventually being paid? Is the hypertrophic bill a sham? The frequency of sessions may have to be altered. Any psychiatrist who sees private patients will definitely face these problems.

Some psychiatrists have entered into barter arrangements. Such arrangements are fraught with technical, legal, and ethical potholes. How, for example, can the value of a barter be established? Does an hour of a contractor's time value have the same as an hour of the psychiatrist's time? If not, how will the doctor feel and how will the contractor feel? How will this affect the objectivity of the doctor in decisions about the patient? What if the contractor installs the light switches a bit out of place? What feelings are stimulated in the doctor? Does he ask that they be fixed and risk angering the patient? Does he swallow it and risk having his own anger come out in some unintended way? Does he make an interpretation? The “cleaner” course is obviously to charge the patient a fair fee and expect the patient to pay it. The Internal Revenue Service takes a dim view of barter arrangements and feels that two sets of taxes are being avoided. Technically, tax must be paid by each party on the fair market value of the bartered services received. Overlooking this may be collusion with the patient in an illegal activity.

Insurance *The Principles* states that one must deal honestly with patients and colleagues. That would certainly extend to insurance companies and managed care entities. Since psychiatric services are compensated according to contracts between patients and their insurance companies, psychiatrists must respect both parties to the contract. It is tempting to exaggerate a patient's symptoms to obtain payment for certain services for them. Nevertheless, in doing this, one is substituting compassion for honesty. Similarly, listing a procedure code that does not accurately reflect the services performed violates not only ethical principles but the law as well. Listing a diagnosis that is covered rather than a more accurate one is similarly problematic. The rationalization is that since the insurance and managed care companies are a priori unprincipled in their policies, psychiatrists may hedge a bit in their reports. It is equivalent to saying “taxes are too high” and so not reporting income.

CONFIDENTIALITY

> All that may come to my knowledge in the exercise of my profession or outside of my profession or in the daily commerce with men, which ought not to be spread abroad, I will keep secret and will never reveal.

Hippocrates

> No provider of health care shall disclose medical information regarding a patient of the provider without first obtaining an authorization

California Civil Code 56.10(a)

> The identification of a person as a patient must be protected with extreme care... A psychiatrist may release confidential information only with the authorization of the patient or under proper legal compulsion.

The Principles

Although the medical profession overall is bound by rules of confidentiality—from Hippocrates to current state law—such rules seem to apply especially to the field of psychiatry. Psychiatrists are generally not troubled by hearing an internist ask a patient at a cocktail party how the new antihypertensive medicine is working out. However, psychiatrists are quite concerned if a psychiatrist in a similar situation asks a patient how he or she is doing with his new antipsychotic medicine. Psychiatrists take in stride the orthopaedist who comments on an athlete's rotator cuff injury (his or her patient or not) but are outraged if a psychiatrist comments on a celebrity's substance abuse problem (his or her patient or not). Both psychiatrists and the public have every right to be outraged and one can only wonder why the outrage does not apply to the internist and the orthopaedist. Patients expect confidentiality from psychiatrists, and patients own the confidentiality. That confidentiality survives even the death of the patient and is subsequently owned by the executor, not by the psychiatrist. A confidence cannot be broken just because the patient died. Psychiatrists must diligently guard against breaking confidences even with a release if that release was not signed by someone in an informed state of mind. A blanket release of information, often signed upon applying for insurance, submitting a claim, entering a hospital, and similar circumstances, is not an informed consent. The psychiatrist must ensure that a release of information applies to the situation at hand. It is often suggested that whenever a request for information about a patient appears, the doctor should call the patient to verify consent to the disclosure.

Frequently psychiatrists get official-looking documents with the words “Subpoena Duces Tecum” on top and a signature at the bottom directing them to allow certain records to be copied or commanding them to produce the records somewhere. Most such documents do not come from a court, but from a lawyer's office. If there is any question that the patient has not consented to such disclosures and has not signed an informed consent to release such records, the ethical psychiatrist should contact the patient and come to some understanding. Patients who do not wish such records released may wish to have their lawyer attempt to quash the subpoena. Failing that, the psychiatrist may wish to ask the court to uphold the patient's right to privacy or at least determine what must (and what must not) be released. In certain jurisdictions patients may waive their right to privacy and confidentiality under certain circumstances (e.g., bringing their own mental state into a lawsuit). It is not within the psychiatrist's field of expertise to determine if the patient has waived that right. It is the job of the patient's lawyer or a judge or both.

Overall, psychiatrists should never discuss their patients outside the office. Some psychiatrists feel that it is all right to discuss cases at the dinner table with their spouse and family. After all, they say, “I trust my spouse and kids.” However, trust is beside the point. Patients assume that what they tell the psychiatrist stays inside the consulting room. Merely informing a spouse of the identity of one's patient violates the ethical principles. Attempting to disguise an identity to keep a patient's case history anonymous is tempting fate.

A recent ethics case involved a psychiatrist who mentioned one of his patients to a social group. Details of the case were revealed although not the patient's name. Some members of the social group knew that one of their friends was in treatment with the psychiatrist, and they easily deduced the patient's identity. This led to considerable embarrassment to the patient. The doctor in attempting to defend himself said that if the patient wanted privacy he should not have revealed that he was in treatment with the doctor. This argument was refuted. It is always the doctor's job to protect the privacy of the patient.

Mandatory Reporting The Tarasoff case is an example of the legal system's attempt to solve a social problem—the need to safeguard life—by creating an ethical dilemma for the psychiatrist. This case, which began as a civil lawsuit, ended up with the California Supreme Court ruling that a pyschotherapist has a duty to warn and protect a potential victim of a potentially dangerous patient. The court proclaimed that the principles of justice and nonmaleficence outweighed the principle of beneficence (confidentiality). Most states have agreed with the Tarasoff court and have enacted laws requiring psychotherapists to warn potential victims or to warn the police when an identified person is threatened.

Tarasoff I Does establishment of a therapist-patient relationship oblige the therapist to care for the safety of not only the patient but also others? This issue was raised in 1976 in the case of *Tarasoff v Regents of University of California* (now known as *Tarasoff I*). In this case, Prosenjit Poddar, a student and voluntary outpatient at the mental health clinic of the University of California, told his therapist that he intended to kill a student readily identified as Tatiana Tarasoff. Realizing the seriousness of the intention, the therapist, with concurrence of a colleague, concluded that Poddar should be committed for observation under a 72-hour emergency psychiatric detention provision of the California commitment law. The therapist notified the campus police both orally and in writing that Poddar was dangerous and should be committed.

Concerned about the breach of confidentiality, the therapist's supervisor vetoed the recommendation and ordered all records relating to Poddar's treatment destroyed. At the same time, the campus police temporarily detained Poddar but released him on his assurance that he would “stay away from that girl.” Poddar stopped going to the clinic when he learned from the police about his therapist's recommendation to commit him. Two months later, he carried out his threat to kill Tatiana. The young woman's parents then sued the university for negligence.

The California Supreme Court deliberated the case for the unprecedented time of about 14 months and ruled that a physician or a psychotherapist who has reason to believe that a patient may injure or kill someone must notify the potential victim, the victim's relatives or friends, or the authorities.

The discharge of the duty imposed on the therapist to warn intended victims about danger may take one or more various steps, depending on the case. The court stated that it may call for the therapist to notify the intended victim or others likely to notify the victim of the danger, to notify the police, or to take whatever other steps are reasonably necessary under the circumstances.

The *Tarasoff I* decision has not drastically affected psychiatrists; it has long been their practice to warn the appropriate people or law enforcement authorities when a patient presents a distinct and immediate threat to someone. According to the APA, confidentiality may, with careful judgment, be broken when (1) a patient probably will commit murder, and the act can be stopped only by the psychiatrist's notification of the police; (2) a patient probably will commit suicide, and the act can be stopped only by the psychiatrist's notification of the police; and (3) a patient, such as a bus driver or an airline pilot, who has potentially life-threatening responsibilities shows markedly impaired judgment.

The *Tarasoff I* ruling does not require that therapists report a patient's fantasies; instead, it requires them to report an intended homicide; it is the therapist's duty to exercise good judgment.

Tarasoff II In 1982 the California Supreme Court issued a second ruling in the case of *Tarasoff v Regents of University of California* (now known as *Tarasoff II*), which broadened its earlier ruling, the duty to warn, to include the duty to protect.

The *Tarasoff II* ruling stimulated intense debates in the medicolegal field. Lawyers, judges, and expert witnesses argue the definition of protection, the nature of the relationship between the therapist and the patient, and the balance between public safety and individual privacy. Clinicians argue that the duty to protect hinders treatment because a patient may not trust a doctor if confidentiality is not maintained. Furthermore, because it is not easy to determine whether a patient is dangerous enough to justify long-term incarceration, unnecessary involuntary hospitalization may occur because of a therapist's defensive practices.

As a result of such debates in the medicolegal field since 1976, the state courts have not made a uniform interpretation of the *Tarasoff II* ruling (the duty to protect). Generally, clinicians should note whether a specific identifiable victim seems to be in imminent and probable danger from the threat of an action contemplated by a mentally ill patient; the harm, in addition to being imminent, should be potentially serious or severe. Usually, the patient must be a danger to another person, not to property, and the therapist should take clinically reasonable actions.

In a few cases (none successful so far) claims have already been advanced that a *Tarasoff*-like duty applies to potential infection of partners with human immunodeficiency virus (HIV) by patients under mental health treatment. The breach of confidentiality in *Tarasoff* cases is justified only by the threat of violence. Laws vary confusingly by jurisdiction. The ideal solution is to persuade the patient to make the disclosure and report the matter to the public health authorities.

Dealing With the Press It is quite flattering (at first) for a psychiatrist to be called by the news media and be asked about a person who may be in the news. The person may or may not be the patient of that doctor. Section 7.3 of *The Principles* is quite specific. Psychiatrists may comment on mental illness generally, but not offer opinions about either their patients or a person who is not their patient.

In 1964 Barry Goldwater, an outspoken, conservative man, ran for President of the United States. A magazine sent a questionnaire to nearly all the psychiatrists in the United States asking: “Do you believe Goldwater is psychologically fit to serve as President of the United States?” The magazine sent out 12,356 questionnaires and received 2417 responses. Of these, 1189 psychiatrists said the candidate was not psychologically fit to serve as president. The cover of the magazine ran its main story with the title “1,189 Psychiatrists Say Goldwater is Psychologically Unfit to be President!” How many of these psychiatrists actually examined the candidate? How many of them obtained permission to reveal the results of their examination? On what basis, then, did so many outstanding psychiatrists lose their moral bearings and offer a psychiatric or medical opinion on this man's mental state. By reading about him in the newspaper? By watching him on TV? The authors of *The Principles*, which was written nine years later in 1973, considered this event when they wrote Section 7.3.

In the mid-1990s the wife of a famous athlete and a friend of hers were murdered. Her former therapist immediately contacted the media to describe the psychological events that were going on with the victim and with her husband. The press alternated between being shocked that a therapist would reveal such material and being fascinated at having such a loquacious source. Many psychiatrists and other mental health professionals commented about the victim and the victim's husband, the suspect. The psychiatrist who was attending to the husband-suspect was appropriately silent.

ETHICS IN MANAGED CARE

Recent changes in how psychiatrists are compensated for their services have stimulated much criticism of both the managed care industry and the doctors who work in and for it. Some claim that the inner conflicts caused by working in a managed care environment are insurmountable. Some have declared that participating in a managed care scheme is per se unethical. Others, while lamenting the existence of managed care, have said that one can function ethically in such an environment. A significant number of psychiatrists believe that managed care provides the answer to overuse of limited services, contains costs, and offers the opportunity for outcome studies. There may be some truth to each point of view.

Conflict of Interest There is nothing new about a physician having a conflict of interest regarding how services are provided and how payments are received. In a fee-for-service environment, the more services a physician provides, the greater is the fee. That leads to the conflict of interest in which physicians may provide an abundance of service to increase their fee. Psychiatrists take that conflict as a given and advise trainees to put the interest of their patients before their own and provide only the service that is needed. Yet all psychiatrists are aware of egregious examples of interminable treatments, prolonged hospitalizations, and treatments of unconscionable intensity that are clearly in the interest of the physician or the hospital or both. There have been a few outcries in the press, and some governmental agencies took some miscreants to task, but few complaints of ethics violations were made. Recently, a proposal was made to the APA that any psychiatrist who limited services to a patient covered by a managed care program should be declared guilty of unethical behavior and expelled. This proposal was not accepted, but it received a great deal of support during the debate. (No analogous proposal was made during the era of multiyear hospitalization for adolescent conduct disorder.) These anecdotes underline the difference in the perception of conflict of interest when restrictions are imposed on psychiatrists from the outside

(managed care) and when restrictions must be self-imposed (fee for service). Yet, the ethical principles are the same. *The Principles* section 1, annotation 1, states: "A psychiatrist shall not gratify his or her own needs by exploiting a patient." This clearly applies in both the fee-for-service environment and the managed care environment. Thus, it would be unethical for psychiatrists working as a managed care reviewer to deny needed care to patients on the basis of their employer's cost-containment policy rather than on the honest belief that they did not require the service. On the same principle, it would be unethical for a treating psychiatrist to dishonestly claim that a patient was suicidal (if that were not the case) to obtain payment for services that the treating psychiatrist believes is necessary.

Contracts In the not too distant past, patients came to see doctors and paid out of their pockets for their treatment. This practice continued for psychiatrists somewhat longer than for our other medical colleagues. Under managed care, both doctor and patient operate under a set of rules written by insurance companies and managed care companies. Anyone who is a party to such a contract is bound by its provisions. Doctors have, at least, a slight advantage. They can and should study each contract and understand all the provisions in it. If a managed care contract has provisions they cannot live with, doctors can and should reject the contract. For example, if a contract has a gag provision that prohibits the doctor from telling patients of the existence of procedures or medicines that might help them but are not paid for by the company, the doctor might well reject the contract. Sadly, most patients cannot pick their contract. Employers buy health coverage on the basis of the least costly package, not the one with the fairest benefit plan. Employees get what is offered. Still, doctors must operate under the contract that they have agreed to and which the patient has agreed to albeit reluctantly. Section 2, annotation 5 of *The Principles* states: "Psychiatric services, like all medical services, are dispensed in the context of a contractual arrangement between the patient and the treating physician. The provisions of the contractual arrangement, which are binding on the physician as well as on the patient, should be explicitly established."

APA GUIDELINES FOR ETHICAL PRACTICE IN ORGANIZED SETTINGS

The APA has issued the following addendum to *The Principles* to help physicians clarify existing ethical standards contained in that document and excerpted above.

Psychiatrists have a long and valued tradition of being essential participants in organizations that deliver health care. Such organizations can enhance medical effectiveness and protect the standards and values of the psychiatric profession by fostering competent, compassionate medical care in a setting in which informed consent and confidentiality are rigorously preserved, conditions essential for the treatment of mental illness. However, some organizations may place the psychiatrist in a position where the clinical needs of the patient, the demands of the community and larger society, and even the professional role of the psychiatrist are in conflict with the interests of the organization.

The psychiatrist must consider the consequence of such role conflicts with respect to patients in his/her care and strive to resolve these conflicts in a manner that is likely to be of greatest benefit to the patient. Whether during treatment or a review process, a psychiatrist shall respect the autonomy, privacy and dignity of the patient and his/her family.

These guidelines are intended to clarify existing standards. They are intended to promote the interests of the patient, and should not be construed to interfere with the ability of a psychiatrist in an organized setting. The principles and annotations noted in this communication conform to the statement in the preamble to *The Principles of Medical Ethics*. These are not laws but standards of conduct which define the essentials of honorable behavior for the physician.

1. *Appropriateness of Treatment and Treatment Options*
   a. A psychiatrist shall not withhold information that the patient needs or reasonably could use to make informed treatment decisions, including options for treatment not provided by the psychiatrist. [sect 2, annotation 5; section 5, annotations 2 and 3].
   b. A psychiatrist's treatment plan shall be based on clinical, scientific or generally accepted standards of treatment. This applies to the treating and the reviewing psychiatrist.
   c. A psychiatrist shall strive to provide beneficial treatment which shall not be limited to minimum criteria of medical necessity.
2. *Financial Arrangements*
   When a psychiatrist is aware of financial incentives or penalties which limit the provision of appropriate treatment for that patient, the psychiatrist shall inform that patient and/or the designated guardian.
3. *Review Process*
   A psychiatrist shall not conduct reviews or participate in reviews in a manner likely to demean the dignity of the patient by asking for highly personal material not necessary for the conduct of the review. A reviewing psychiatrist shall strive as hard for a patient he/she reviews as for one he/she treats to prevent the disclosure of sensitive patient material to anyone other than for clear clinical necessity.

SUGGESTED CROSS-REFERENCES

Ethics pertains to all domains of psychiatric research and practice. Of special interest are Section 4.2 on sociology and psychiatry; Chapter 9 on the classification of mental disorders; Chapter 19 on sexuality; Section 28.5 on death, dying, and bereavement; Section 29.1 on suicide; Section 30.11 on evaluation of psychotherapy; Section 28.4 on genetic counseling; Section 31.31 on electroconvulsive therapy, Section 31.22 on psychosurgery; Section 51.6b and Section 51.6c on legal and ethical issues in geriatric psychiatry; Section 52.1 on public psychiatry; Section 51.5a on financial issues; on the economics of psychiatry, Section 54.1 on legal issues in psychiatry; and Section 55.1 on the history of psychiatry.

SECTION REFERENCES

*American Psychiatric Association: *The Principles of Medical Ethics with Annotations Especially for Psychiatry*. American Psychiatric Association, Washington, DC, 1995.

*American Psychiatric Association: *Opinions of the Ethics Committee on the Principles of Medical Ethics with Annotations Especially for Psychiatry*. American Psychiatric Association, Washington, DC, 1993.

Block S, Chodoff P, editors: *Psychiatric Ethics*. Oxford University Press, New York, 1993.

Epstein RS: *Keeping Boundaries: Maintaining Safety and Integrity in the Psychotherapeutic Process*. American Psychiatric Press, Washington, DC, 1994.

Epstein RS, Simon RI, Kay GG: Assessing boundary violations in psychotherapy: Survey results with the Exploitation Index. Bull Menninger Clin *56:*150, 1992.

*Farber S, Green M: *Hollywood on the Couch*. Morrow, New York, 1993.

Gabbard GO: *Sexual Exploitation in Professional Relationships*. American Psychiatric Press, Washington, DC, 1989.

*Gabbard GO, Lester EP: *Boundaries and Boundary Violations in Psychoanalysis*. Basic Books, New York, 1996.

Gruenberg PB: Nonsexual exploitation of patients—An ethical perspective. J Am Acad Psychoanal *23:*425, 1995.

Gutheil TG, Gabbard GO: The concept of boundaries in clinical practice: Theoretical and risk-management dimensions. Am J Psychiatry *150:*188, 1993.

*Lindsay G, Clarkson P: Ethical dilemmas of psychotherapists. Psychologist *12:*182, 1999.

Schoener GR, Milgrom JH, Gonsiorek JC, Luepker ET, Conroe RM: *Psychotherapists' Sexual Involvement with Clients: Intervention and Prevention*. Walk-in Counseling Center, Minneapolis, 1989.

*Simon RI: *Clinical Psychiatry and the Law*. American Psychiatric Press, Washington, DC, 1987.

Waldinger RJ: Boundary crossings and boundary violations: Thoughts on navigating a slippery slope. Harv Rev Psychiatry *2:*225, 1994.

CHAPTER 55. PSYCHIATRY: PAST AND FUTURE

55.1 HISTORY OF PSYCHIATRY

RALPH COLP, JR., M.D.

BEGINNINGS OF PSYCHIATRY

In the first civilizations in Egypt and the Near East, mental illness was attributed to the magical forces of malevolent deities, and the main therapists were priests who used religious and magical rites to counter those forces. Belief in those evil deities—called demons after the advent of Christianity—prevailed in many primitive societies; in the history of civilized societies, credence in demons fluctuated in complex ways.

GREEK AND ROMAN PSYCHIATRY

In the civilizations of ancient Greece and Rome, psychiatry was a province of religious cults for healing illnesses and of schools of philosophy and medicine (often part of philosophy), which viewed mental illness as mainly psychological, mainly somatic, or a mixture of both psyche and soma; thus, the three enduring psychiatric viewpoints were established.

Greek Psychiatry

Hippocrates The writings of the physician Hippocrates (460–370 BC)—now thought to be the work of unknown authors—propounded the view that the human body contains four essential humors—phlegm, yellow bile, black bile, and blood—which are secreted by different organs, possess different qualities, and vary with the seasons. The brain was considered the seat of life, and its normal functioning required a balance between the humors. Large excesses of phlegm caused a form of dementia, yellow bile caused manic rage, and black bile caused melancholia. Small excesses of these three humors and blood also produced phlegmatic, choleric, and sanguine personalities. It was the first attempt to explain differences in temperaments and personalities.

DISEASES AND TREATMENTS The Hippocratic authors were the first to rationally describe and classify such diseases as epilepsy, mania, paranoia, organic toxic delirium, postpartum psychosis, phobias, and hysteria, which they named and believed to be confined to women and to be caused by a wandering uterus. They associated melancholia with an aversion to food, despondency, sleeplessness, irritability, and restlessness, and they believed that personal hygiene, bathing, and dieting were essential to keep the humors in balance; in cases of insanity, they tried to eliminate excess humors by purgatives, cathartics, and (if necessary) bleeding.

Plato and Aristotle The philosopher Plato (427–347 BC) divided the soul into three parts: rational, appetitive (lusts and greed), and spirited-affective. Madness occurred when the appetitive soul lost the influence of the rational soul or when a divine disturbance of the soul produced either inspired or destructive behavior. For treating disease, Plato advocated a verbal dialectic between a patient and a philosopher or a physician, a question-and-answer dialogue that enabled the patient to alleviate an illness by developing a state of philosophical knowledge.

Plato's pupil Aristotle (384–322 BC) believed that mental illness occurred when the soul was subjected to changes in temperature, black bile, and the emotions. He was the first to describe accurately the affections of desire, anger, fear, courage, envy, joy, hatred, and pity.

Phrenitis The term “phrenitis” was used in ancient Greece to refer to an inflammation that produced disturbances in both the mind and the body, and whose location was never anatomically or conceptually well defined. In the nineteenth century it was replaced by “delirium,” “confusion,” and “clouding.” The clinical behavior it describes continues to be important to modern psychiatry.

Psychological Treatments In addition to Hippocratic somatic and medical treatments, the Greeks developed three psychological treatments: methods of inducing sleep, interpreting dreams (often performed by priests in temples, most notably those of Aesculapius, the god of healing), and using words to encourage, console, and gain knowledge of an illness (e.g., when using Plato's dialectics).

Roman Psychiatry

Stoicism and Epicurianism The Romans generally followed the psychiatric teachings of the Greeks and were influenced by Greek Stoic and Epicurean beliefs that passions and unsatisfied desires act on the soul to produce mental illnesses and that those illnesses can be controlled by achieving, through thoughts and conduct, a mental state of *ataraxia*, or lack of perturbation. (Some modern tranquilizers have been classified as ataractics.)

Galen The greatest Roman physician was Galen (AD 130–200), who consolidated and augmented Greek medical and psychological thinking. He conceptualized that the humors exist in normal and abnormal forms and that four qualities—hot, cold, dry, and moist—are essential in forming human temperaments. To the previous concept of *pneuma* (a vital spirit) he added the concepts of natural spirits and animal spirits. He divided the soul into three parts: reason and intellect, courage and anger, and carnal appetites and desires. He believed that diseases are caused by adverse external influences (e.g., bad diet and bad air) acting on an existing predisposition (an abnormality of a humor or pneuma) and that psychological disorders cause physical disorders and vice versa. Treatment should oppose what is unnatural in a disease (cooling the feverish, warming the chilled) and should be in accord with the patient's age, temperament, and way of life.

Galen's writings, especially his treatise *On Melancholia,* influenced medical thinking for nearly 15 centuries. After his death the creativity of ancient medical writing went into a profound and protracted decline.

Treatment of the Insane Neither Greece nor Rome took social responsibility for treating the insane. The exceptions were a few persons who were admitted to Aesculapian sanctuaries and Roman soldiers who were admitted to military hospitals. Most of the seriously insane were kept under restraint at home by their families. The Romans had laws for appointing guardians to protect the holdings of the insane and for declaring them legally incompetent. Insanity was evaluated by judges, not by physicians. The insane were feared and hated (except those thought to be divinely inspired), beaten, imprisoned, and driven away.

MIDDLE AGES

The 1000 years after the fall of the western Roman empire, AD 400–1400, saw an Arab Islamic empire extending from the Near East to Spain, an eastern Christian Byzantine empire, and Christian western European feudal kingdoms.

Christian Western Europe

Causes of Illness The main supernatural causes of illness were thought to be evil astrological influences (astrology was the main scientific guide for physicians) and demons entering the body and possessing the soul. The writings of Hippocrates and Galen that had survived in Arabic translations and were then translated into Latin

influenced some physicians. Some observed that mental illness could be caused by somatic or psychological conditions.

Illness and Treatment Some often-observed psychiatric reactions were (1) symptoms of possession by the devil, who was supposed to choke and throw its victims about; (2) dance mania, characterized by an irresistible urge to dance and make noises; and (3) *acedia,* a form of depression occurring in anchorites when they experienced distressful doubts about being able to live a religiously meaningful life. Treatment was by physicians or lay persons claiming to possess special powers. Treatment varied and included exorcism of demons by magical techniques, medicinal treatments recommended by Hippocrates and Galen, and time-honored folk remedies (including the use of plants, animal parts, and such esoteric medicines as bark from a tree from Paradise).

Hospitals and Asylums Since Islam held that society is responsible for the kindly care of the insane, the Arabs built hospitals with psychiatric divisions in Baghdad (750) and Cairo (873); they also built special insane asylums in Damascus (800), Aleppo (1270), and the Muslim-ruled Spanish city of Grenada (1365).

Christian attitudes toward the insane fluctuated between rejection and toleration. Because the insane harbored demons, they were incarcerated in madmen's towers and were periodically expelled from towns. An insane person might also be condoned as a fool who lacked a mind because of God's favor and who needed to be cared for. Thus, in the fourteenth century, several hospitals in western Europe cared for the mentally ill, in addition to those with physical illness. Historians have usually recorded that the first Christian western European asylum caring exclusively for the insane was in Valencia Spain in 1409, and that this was followed by other Spanish asylums that were built under the influence of Islam. Michael Stone has recently stated that the first asylum in Europe was built in Hamburg in 1375. While some of Hamburg's insane were confined in a tower little is known about their treatments, or how these differed from those in other medieval cities.

“Bedlam” It was probably around 1400 that Bethlem hospital in London began for unknown reasons to admit the insane, who perhaps because they were verbally and physically violent appear to have displaced the other sick; so that by the sixteenth century Bethlem had assumed its role as a madhouse. Over the following centuries, while becoming known in popular culture as “Bedlam” (defined as “a place of uproar”), it became a special and unique institution, pioneering in the asylum management and care of the insane. *The History of Bethlem* (1997), by Jonathan Andrews, Asa Briggs, Roy Porter, and Penny Tucker, is a 750-page epic narrative, that often has vivid passages, and that will become an essential reference for all who are interested in the history of psychiatry and the asylum.

RENAISSANCE

Witches In 1486 fear of witchcraft—injury inflicted by someone using supernatural devices—was stimulated by the publication of *Malleus maleficarum (Witches' Hammer)* by Henry Kramer and James Sprenger, two theologians who had been designated by the Pope to act as papal inquisitors into witchcraft. *Malleus* described witches as mainly women who showed psychotic or hysterical symptoms and sexual delusions, but it also suggested that any affliction in a male or female could be a sign of witchcraft. It prescribed inquisitorial tortures that would force confessions of guilt from those who were accused. The ideas of *Malleus* then fused with the ideas of other theologians (Catholic and Protestant) of a conspiracy against Christendom and with popular beliefs in demonology. The result was a witch-hunting craze that led to the execution of many thousands of people (mostly women) on charges of being witches and only abated after about 150 years.

Physicians and Books

Paracelsus Born in Switzerland of German parents, Paracelsus (ca. 1493–1541) was the most famous physician of his time. His medical ideas were influenced by medieval philosophy, astrology, magic, and myths and were often confused, contradictory, and controversial. Rejecting the prevailing humoral theory of diseases, he postulated that man had a divine spirit and base animal instincts (lust, covetousness, and the passions of the soul) and that psychiatric illness was the triumph of instincts over spirit, caused by the stars stimulating the instincts. Consequently the physician must study both the instinct dominant in a patient and the star that corresponds to it. Treatments included psychotherapy, venesection, trephination (perforation of the skull), and the administration of sulfur—used because its sedative effects were thought to promote the healing actions of “Sympathy” (one of the main forces in the cosmos). Because he believed that emotions caused physical illness, Paracelsus has been called one of the first “psychosomaticists.”

Books Five sixteenth- and early seventeenth-century authors published new views in different areas of psychiatry. Juan Luis Vives (1493–1540) was born in Spain and lived mainly in Holland and Belgium where his work as a scholar was influenced by the humanism of the North European Renaissance. (His Spanish Jewish parents had converted to Christianity and then been condemned by the Inquisition.) In his *Of Soul and Life* (1538) he anticipated modern dynamic psychology and described for the first time the importance of psychological associations in forming emotions and in recollecting and forgetting past events. The Spanish physician Juan Huarte de San Juan (ca. 1530–1592) in *The Examination of Men's Wits* (1575) made an early attempt to evaluate characteristics and differences in human intelligences. The Italian writer Giambattista Porta (1535–1615) made a pioneer attempt in *The Human Physiognomy* (1586) to compare human morphological and psychological traits.

Weyer In 1563 the German-Dutch physician Johann Weyer (1515–1588) published *De Praestigiis Daemonum (The Deception of Demons),* which was based on 12 years of observing witches. In it Weyer emphasized the difference between magicians who willingly took instruction from the devil, and witches who were not responsible for their deeds and were often mentally unstable and ill from medical causes, who needed the help of physicians. Weyer criticized the *Malleus* and stated that some women only confessed to being witches because of torture. George Mora observed in his introduction to the English translation of *De Praestigius* that Weyer's role as a pioneer of modern psychiatry is “difficult to assess”; that for centuries his work was ignored by those interested in psychiatric illness; and that only recently, with the advent of psychodynamic theories and psychotherapies, was Weyer seen as a precursor of these approaches.

Plater The Swiss physician Felix Plater (1536–1614) published two large books, *Practice of Medicine* (1602) and *Observations of Diseases Injurious to Body and Mind* (1614), that classified and described all known diseases and advanced the process of separating psychological knowledge from philosophy and theology. He also was a pioneer in recognizing degrees of mental handicap (i.e., that some mentally handicapped persons may have exceptional ability in one area, that some are mentally disturbed and some are not) and in making autopsies on and studying the neuropathology of handicapped individuals.

Advent of Syphilis During the years 1530–1546 Girolamo Fracastoro (1484–1553), an Italian physician and humanist, published accounts of a new epidemic illness with loathsome and disfiguring symptoms and significant mortality. He named it syphilis and correctly stated that it was caused by sexual contact between two persons. The concept of sexual contagion was new, and syphilis caused a unique and unprecedented mass fear about the danger of sexual activity—a fear that was later sustained when chronic syphilis was recognized as a major cause of dementia. That fear was equaled only by recent fears of acquired immune deficiency syndrome (AIDS).

SEVENTEENTH CENTURY

Seventeenth-century thought on mental illness was influenced by the work of two eminent English physicians—Thomas Willis (1621–1675) and Thomas Sydenham (1624–1689)—an Italian papal physician, Paolo Zacchia (1584–1659), and an Oxford dean of divinity, Robert Burton (1577–1640).

Willis (who performed autopsies on some of his patients) recognized the difference between mental symptoms of gross brain disease and diseases in which the brain seemed normal, which he attributed to disturbed animal spirits. Sydenham gave a comprehensive account of the many symptoms of hysteria, observed that it occurred in both males and females, and ascribed it to abnormal animal spirits. His classification of the causes of disease symptoms into “remote” or “external” (inordinate actions of the body and mind), immediate (a disorder of animal spirits), and antecedent (a “faulty disposition of the animal spirits”) resembled the modern psychiatric endeavor of constructing atheoretical classifications of mental disease. Sydenham's classification may also have been the first attempt to incorporate contemporary physiological and psychological knowledge into a systematic analysis of insanity.

Zacchia in *Questiones medico-legales* (1621–1635) insisted that the physician, rather than the lawyer, should evaluate the mental state of a person. That view was the beginning of forensic psychiatry. Burton in *The Anatomy of Melancholy* (1621) provided authoritative summaries of writings on melancholy from ancient times through the Renaissance. *The Anatomy* was widely read and caused some forms of melancholy to become fashionable in England.

Incarceration of the Insane In 1656 an edict of the French monarchy created a new hospital administrative organization for Paris that consisted of asylums for the insane and other institutions, each with directors who were empowered to detain persons for indefinite periods and who over the next century locked up groups of insane persons along with indigents, orphans, prostitutes, homosexuals, aged persons, and those chronically ill. Similar incarcerations occurred in Germany.

However, few insane persons were confined in other European countries. In England after 1660 most insane persons were kept at home or were left to roam; only a

few were sent to private madhouses or new public asylums in great towns.

EIGHTEENTH CENTURY

With the influence of the eighteenth-century Enlightenment, belief in the power of rational thinking replaced belief in Hippocratic, Galenic, and demonic ideas and led to new theories of mental illness.

Sauvages and Nosology The French physician Boissier de Sauvages (1706–1767) in his *Nosologia methodica* (1763) divided all known diseases into natural classes, orders, genera, species, and varieties (similar to the way naturalists had divided plants and animals). There were 10 classes of diseases, with class 8 containing species of psychiatric illnesses. The nosology was artificial, but it stimulated physicians to rethink their concepts of disease.

Cullen and Neurosis Two years after Sauvages's nosology the English physician William Cullen (1710–1790) published his own nosology. He used for the first time the term “neurosis” and its adjective “neurotic” to describe a variform group of mental diseases (including apoplexy, paralysis, dyspepsia, hypochondriasis, epilepsy, and hysteria), which he thought were caused when parts of the brain were simultaneously in unequal states of excitement and collapse. Later the meaning of the term “neurosis” would undergo several changes.

Gall and Phrenology Franz Joseph Gall (1758–1828) was a German physician who taught that the brain contained separate organs that occupied specific areas and shaped the personality and that those organs were revealed in the configurations of the head and could be mapped on its surface. This doctrine, with modifications by Gall's pupil Johann Gaspar Spurzheim (1776–1832) and the English lawyer George Combe (1788–1858), became the science of phrenology. Because it offered views on human nature, phrenology became popular during the early nineteenth century in England, France, and the United States.

Treatment of the Insane For most of the eighteenth century, the insane inmates in public mental asylums were often regarded as having incurable diseases and were subjected to physical restraints, beatings, and constant fear (in some private asylums treatment was more lenient). Toward the end of the century that harsh treatment was changed by reforms that abolished most restraints and created an asylum regimen in which the mental patient was respected as a person. These reforms were carried out by the Italian physician Vincenzo Chiarugi (1759–1820) in 1789 in the hospitals of Santa Maria Nuova and of Bonifazio in Florence, Italy; by the English layman and Quaker, William Tuke (1732–1822), who in 1796 founded the York Retreat outside the English city of York; and by Jean Baptiste Pussin (1746–1809?), assistant to the French physician Phillippe Pinel (1745–1826), who in 1797 removed the fetters from the inmates in the Paris hospital of Bicêtre.

Pinel's Empiricism and Moral Treatment In his influential 1801 book *A Treatise on Insanity,* based on empirical observations of patients in the Bicêtre, Pinel made a new classification of mental illnesses—mania, melancholia, idiocy, and dementia—and stated that they were caused mainly by heredity and influences from the environment. He also described how (through an asylum regimen of education, reasoning, and persuasion) many symptoms of insanity could be alleviated. He called this regimen the “moral treatment of insanity.”

EARLY NINETEENTH CENTURY

During the first half of the nineteenth century, the view that mental retardation was a defect of intellectual function different from insanity or dementia became well established; psychiatry became a specialty apart from medicine; and leading psychiatrists in France, England, the United States, and Germany were usually guided by principles of moral therapy in their work with patients.

France

Esquirol Jean Etienne Dominique Esquirol (1782–1840) was the most influential psychiatrist of his time. He further developed the theory and practice of moral therapy, seeing the asylum as the most powerful weapon against mental illness, and was one of the authors of an 1838 law providing an asylum in every department of France, with detailed provisions for patient care, a truly comprehensive law that is still in force in the late twentieth century. His influential book *Mental Maladies: A Treatise on Insanity* (1838), guided by Pinel's empirical approach, differentiated between *hallucination* (a term he coined) and illusion; he divided insanities into monomania, a partial insanity that included paranoia and obsessive-compulsive disease (an early clinical recognition of this disease), and a general delirium-like mania. He inaugurated what was probably the first course in psychiatry and trained physicians to be asylum directors. Among his pupils, Jean-Pierre Falret (1794–1870) and Jules Baillarger (1809–1890) independently described and named circular insanity, and Jacques-Joseph Moreau de Tours (1804–1884) described the psychotic effects of taking the drug hashish. He was the first psychiatrist to delineate the symptoms of a drug-induced psychosis.

England In 1813 Samuel Tuke (1784–1857), grandson of William Tuke, made the York Retreat into one of the most renowned moral treatment asylums in the world by publishing a widely read book, *Description of the Retreat*. In 1839 the physician John Conolly (1794–1866) started a movement to abolish mechanical restraints and later chemical restraints such as bromide and chloral hydrate for overactive asylum patients; practitioners of moral treatment had sometimes used restraints. Nonrestraint was accepted at the first (1841) meeting of the Association of Medical Officers of Hospitals for the Insane, today the Royal College of Psychiatrists, which in 1853 founded the *Asylum Journal,* today the *British Journal of Psychiatry*.

United States

Rush Benjamin Rush (1745–1813) was the most famous American physician of his time and became known as the father of American psychiatry, and his book *Medical Inquiries and Observations upon the Diseases of the Mind* (1812) was the first comprehensive book on mental illness by an American. However, Rush's belief in somatic causation (what he believed to be the morbid actions of cerebral blood vessels) and his nonacceptance of moral treatment caused his influence to wane after his death, when prominent superintendents of American asylums were practicing moral treatment.

American Psychiatric Association In 1844 13 asylum superintendents founded the Association of Medical Superintendents of American Institutions for the Insane (today called the American Psychiatric Association) and its official periodical, the *American Journal of Insanity* (today called the *American Journal of Psychiatry*). Imbued with a belief in moral treatment, those 13 founders (in the view of Constance McGovern) “took on the responsibility for the mentally ill, convinced themselves they could manage and cure the insane, and set out to persuade everyone else.” Their work was, at first, enhanced by lay reformers—most notably Dorothea Dix (1802–1887)—who influenced legislators to create or expand more than 30 institutions for the insane.

Germany After Johann Reil (1759–1813) in *Rhapsodies about the Application of Psychotherapy to Mental Disturbances* (1803) and Johann Christian Heinroth (1773–1843) in *Disturbances of the Mind* (1818) had emphasized the role of psychological factors in mental illness, Wilhelm Griesinger (1817–1868) in *Mental Pathology and Therapeutics* (1845) took an opposing somatic view and argued that mental diseases are brain diseases, even though it was not yet possible to correlate specific brain damage with a specific psychiatric disease. Griesinger later became professor of psychiatry and neurology in Berlin, directing university clinics and laboratories in studies of the anatomy and the pathology of the brain. This work and the discovery in 1861 of a speech center in the brain by the French physician Paul Broca (1824–1880), led to the school of brain psychiatry, represented by Theodore Meynert (1833–1892) in Vienna and Carl Wernicke (1848–1905) in Breslau, which attempted to localize psychiatric phenomena in real or imagined brain structures.

LATE NINETEENTH CENTURY AND EARLY TWENTIETH CENTURY

Theory of Degeneration A theory of mental illness, propounded by two French psychiatrists, Benedict-Augustin Morel (1809–1873) and Valentin Magnan (1835–1916), held that a variety of mental illnesses ranging from insanity to obsessions and compulsions could be congenitally inherited from parents who were afflicted by similar diseases and that the predisposition could slowly become activated into a disease by repeated transmission from parent to child or it could suddenly be activated by the effects of such external events as social traumas, alcoholism, syphilis, and other infections.

The concept of degeneration was, for a time, accepted by prominent psychiatrists. The German psychiatrist Richard von Krafft-Ebing (1840–1902) in *Psychopathia sexualis* presented various cases of sexual perversions (he first coined the terms “sadism,” “masochism,” and “sexual bondage”) and homosexuality, stating that they were often caused by degeneration. His book won respect from lawyers and scientists, and it founded the science of sexology. The Italian psychiatrist Cesare Lombroso (1836–1909) in *The Delinquent Man* (1876) and *The Female Offender* (1893) postulated that criminals represent a degenerate biological phenomenon that can be identified on the basis of physiognomy. Although those postulations were incorrect, Lombroso's books inspired a school of criminal anthropology in Italy, out of

which developed the science of criminality.

Classification of Functional Psychoses

Kraepelin When Emil Kraepelin (1856–1926) was a professor of psychiatry at the universities of Dorpat (then a part of Russia) and at Heidelberg and Munich in Germany, he studied the objective symptoms and clinical life histories of thousands of psychotic patients. In 1899, in the sixth edition of his textbook of psychiatry, Kraepelin classified the major psychoses into two groups. Manic-depressive psychosis (formerly called circular insanity) patients usually recovered; dementia precox, which deteriorated to dementia and which consisted of three subgroups—hebephrenia and catatonia, described by Ewald Hecker (1843–1909) and Karl Kahlbaum (1828–1899) in 1871 and 1874, and dementia paranoia, described by Kraepelin and separated from paranoia and paraphrenia. The classification excluded causes and psychological influences and was based on morbid outcome, although Kraepelin later conceded that some patients with dementia precox recovered. It gave clinical meaning to many disconnected facts and created new disease entities.

Bleuler The Swiss psychiatrist Eugen Bleuler (1857–1939) in *Dementia Praecox or the Group of Schizophrenias* (1911) showed that because dementia precox does not always become dementia, it can more accurately be called “schizophrenia.” He also introduced the terms “autism” and “ambivalence” and described the schizoid personality.

Shock Treatments In the 1920s and 1930s convulsions produced by metrazol given by the Hungarian psychiatrist Ladislas Meduna (1896–1964) and by insulin used by the Viennese physician Manfred Sakel (1900–1957) resulted in remissions in the symptoms of schizophrenic patients. In April 1938 two Italian psychiatrists, Ugo Cerletti (1877–1963) and Lucio Bini (1908–1964), produced convulsions by means of electroshock. Electroconvulsive treatment (ECT) largely replaced metrazol and insulin in the treatment of schizophrenia in the 1940s and early 1950s and was found to be most effective in treating mood disorders. Today (while its use has declined because of the use of antipsychotic and antidepressant medications) modified forms of ECT are still used to treat mood disorders.

Descriptions of Psychoses

Korsakoff's Psychosis In 1889–1890 the Russian psychiatrist Sergei Korsakoff (1853–1900) described a syndrome caused by chronic alcoholism, characterized by multiple neuritis, failure of memory, pseudoreminiscences, and disorientation. Korsakoff is usually considered the first notable psychiatrist in Russia.

Alzheimer's Disease In 1906 the German neuropathologist Alois Alzheimer (1864–1915) reported a case of progressive psychosis and dementia in a 51-year-old woman with neurofibrillary degeneration and plaques in the brain. The disorder was named after him as a form of presenile dementia.

General Paresis In 1826 the French physician Antoine Bayle (1799–1858) stated that general paresis, then probably the most common organic psychotic disease, was a separate clinical entity. That assertion was then ignored or disputed. Only in 1912, after advances in bacteriology, immunology, and pathology, was the syphilitic origin and pathogenesis of general paresis demonstrated.

Medical Treatment of Psychoses

Fever In 1917 the Austrian psychiatrist Julius von Wagner-Jauregg (1857–1940) caused remissions in general paresis with malaria-induced fever. For that he became (in 1927) the first psychiatrist to receive the Nobel Prize. In the 1940s malaria therapy for general paresis was superseded by penicillin.

Pellagra and Vitamin Deficiency Pellagra, a major cause of dementia in the southern United States, was cured between 1912 and 1937 by the addition of foods containing pellagra-preventive factor, now known as niacin or nicotinic acid.

Classification of Neuroses In the late nineteenth century *neurosis* had come to mean a psychiatric illness caused by changes in brain function in which no organic lesion could be found and that (unlike a psychosis) was not inherited. The two major neuroses were neurasthenia and hysteria.

Neurasthenia The term *neurasthenia* was first used by the American neurologist George Miller Beard (1840–1883) in 1869 to describe a syndrome of physical and mental exhaustion that had previously been called *hypochondriasis*. The syndrome was prevalent and fashionable among American and European middle-class persons until about 1920 and was treated by rest cures (often consisting of isolation from the patient's parents), exercise, massage, and electrical stimulation of various parts of the body.

Hysteria Concepts of hysteria were influenced by the development of hypnosis. An unrecognized form of hypnosis was practiced in the eighteenth century by the Austrian physician Franz Anton Mesmer (1734–1815), who successfully treated hysteria and many other illnesses by putting patients in a trance and invoking the healing powers of an invisible and impalpable fluid that he called “animal magnetism.” In 1843 the English surgeon James Braid (1795–1860) showed that Mesmer's influence could be self-induced by fixating on an inanimate object and that it was caused not by animal magnetism but by a subjective impression made on the nervous center. Braid called it “neurohypnotism,” which was subsequently shortened to hypnosis. During the 1880s Jean-Martin Charcot (1825–1893), first professor of diseases of the nervous system at the University of Paris (who made Paris an international center for neurological studies), began using hypnosis to study the symptoms of the female patients in the Salpetrière. Through hypnosis he was able to produce many of these symptoms and he erroneously claimed that hypnosis was a pathological state that occurred only in hysterics. That claim was disputed by two leading French hypnotists, Ambroise August Liébeault (1823–1904) and Hippolyte Bernheim (1840–1919), who correctly claimed that most persons could be hypnotized and that hypnosis could be used as a psychotherapeutic procedure in neurotic illnesses.

Freud and the Creation of Psychoanalysis The disputes about neurotic patients and hypnosis stimulated the young Viennese physician Sigmund Freud (1856–1939), who had studied with Charcot and observed the work of Liébeault and Bernheim, to study these patients. In the 1890s he began diagnosing them as suffering from separate clinical illnesses which he called “anxieties,” “phobias,” and “obsessions”; terms that would subsequently influence the diagnosis and classification of neuroses.

At first Freud, along with his physician friend Joseph Breuer (1841–1925), treated these patients with hypnosis. Then by himself he originated a new method that consisted of free thought association and interpretation of dreams, which in 1896 he first called “psychoanalysis.” Using this method, he was able to show that neurotic patients suffer from memories of past traumatic events that they repressed into the unconscious parts of their mind and that when those memories were brought into consciousness by psychoanalysis and reexprienced with the support of the psychoanalyst, the neurotic symptoms were alleviated. The method revealed that the illnesses of these patients (their hysterical symptoms, anxieties, phobias, and obsessions) were caused by unconscious memories of past illicit sexual encounters. The concept of the unconscious had been known to philosophers since the seventeenth century; however, no one had thought that it could be used to investigate and treat psychiatric illnesses.

In 1900 *The Interpretation of Dreams* was published. It was Freud's most insightful work, and it revolutionized psychiatric thought and practice as well as other areas of thought. In it Freud showed how processes of condensation, distortion, disguise, and secondary elaboration transform unconscious latent dream content into conscious manifest content. His analysis of his own dreams, which was part of his self-analysis, was one of the most revealing psychological introspections ever recorded. Freud's *Three Essays on the Theory of Sexuality* (1905), with its explanations of the stages of sexual development and the psychological origins of what were then called the sexual perversions, altered scientific and moral views of sexuality.

Freud was about 50 years old when he formed his main psychoanalytic ideas. For the remaining 30 years of his life, he further developed those ideas and trained a growing number of adherents to become psychoanalysts. Psychoanalytic societies were formed in Europe and the United States; members were required to complete a didactic personal analysis and to adhere to Freud's main tenets. Freud's work became well known in the U.S. after the Clark University conference in 1909 (Fig. 55.1-1).

FIGURE 55.1-1 A: Psychological Conference, Clark University, Worcester, Massachusetts, September 1909. **B:** 1, Franz Boas; 2, E.B. Titchener; 3, William James*; 4, William Stern; 5, Leo Burgerstein; 6, G. Stanley Hall*; 7, Sigmund Freud*; 8, Carl G. Jung*; 9, Adolf Meyer*; 10, H.S. Jennings; 11, C.E. Seashore; 12, Joseph Jastrow; 13, J. McK. Cattell; 14, E.F. Buchner; 15, E. Katzenellenbogen; 16, Ernest Jones*; 17, A.A. Brill*; 18, W.H. Burnham; 19, A.F. Chamberlain; 20, Albert Schinz; 21. J.A. Magni; 22, B.T. Baldwin; 23, Lyman Wells*; 24, G.M. Forbes; 25, E.A. Kirkpatrick; 26, Sandor Ferenczi*; 27, E.C. Sanford; 28, J.P. Porter; 29, Sakyo Kando; 30, Hikoso Kakise; 31, G.E. Dawson; 32, S.P. Hayes; 33, E.B. Holt*; 34, C.S. Berry; 35, G.M. Whipple; 36, Frank Drew; 37, J.W.A. Young; 38, L.N. Wilson; 39, K.J. Karlson; 40, H.H. Goddard; 41, H.L. Klopp; 42, S.C. Fuller (*subsequently continued their interest in psychoanalysis). (Courtesy of American Psychiatric Association.)

Other Concepts of Psychoanalysis and of the Unconscious Several of Freud's followers—Alfred Adler (1870–1937), Carl Gustav Jung (1875–1961), Otto Rank (1884–1939), Karen Horney (1885–1952), and Sandor Rado (1890–1972)—had theoretical differences with him that often centered on the importance that he had assigned to the effects of early sexual experiences on the adult personality. These differences led them to create their own versions of psychoanalysis and psychotherapy. Two psychiatrists made observations of the unconscious that were independent of psychoanalysis: the American, Morton Prince (1854–1929), gave an account of a multiple personality; the Frenchman, Pierre Janet (1859–1947), originated the concept of *psychasthenia,* a weakness in the nervous system that resulted in the dissociation of parts of consciousness, obsessions, phobias, and anxieties. He also postulated that traumatic experiences caused hysteria with dissociation of feelings and memories related to experiences.

Projective Psychological Tests Concepts of the unconscious were used in projective psychological tests, in which persons projected unconscious feelings and conflicts into their perceptions of various visual stimuli. Two tests used frequently in diagnostic evaluations of psychiatric patients were the 1921 inkblot test of the Swiss psychiatrist Hermann Rorschach (1884–1922) and the 1935 Thematic Apperception Test (TAT) of the American physician and psychologist Henry Murray (1893–1988).

American Psychiatry

Treatment of the Insane

ASYLUMS In the late nineteenth and early twentieth centuries mental asylums changed from practicing the therapeutic ideas of moral therapy into institutions for custodial care where patients lived crowded together in unsanitary conditions; suffered from diseases, hopelessness, and abuse; and had little therapeutic contact with those who cared for them. One reason for these changes was an increase in patients who were chronically ill, consisting in part of poor immigrants who were unable to work, elderly persons who were transferred from almshouses to mental hospitals, and an influx of patients with neurosyphilis. Further, psychiatrists and those who worked in asylums had developed negative views about their work and the incurability of those with chronic psychoses.

PSYCHIATRIC HOSPITALS AND MENTAL HYGIENE In response to the need for hospital treatment of persons with acute psychoses, psychiatric hospitals were founded. In New York City, Bellevue Hospital (1879) and the Pathological Institute of the New York state hospital system (1895), later called New York State Psychiatric Institute, were founded. Boston Psychopathic Hospital (later called the Massachusetts Mental Health Center), was founded (1912) as a department of Boston State Asylum, the first psychiatric hospital connected with an asylum and the first to open an outpatient department to study the psychiatric illnesses of children. Similar hospitals were founded in Albany, New York, and in Illinois, Wisconsin, and Michigan. The staffs of those hospitals later included such new psychiatric professionals as psychiatric social workers, clinical psychologists, and occupational therapists.

In 1908 Clifford Beers's book *A Mind That Found Itself* was published. It was a widely read memoir of the bad treatment he had received as a mental patient in three asylums. In 1909 Beers founded the National Committee for Mental Hygiene, which aimed at controlling and preventing mental disorders and which began what was called “the mental hygiene movement” by the rising Swiss-born American psychiatrist Adolf Meyer.

Meyer When Adolf Meyer (1866–1950) was about 30, he stated in an article in the *American Journal of Insanity* his fundamental psychiatric belief that neither a purely somatic nor a psychological approach would suffice by itself but “that all mental activity must have its physiological side and its anatomical substratum... combinations of [nerve] cells....A disease of these cells... means at the same time a *physiological* and a *psychological* disorder; destruction of these cells, a destruction of *physiological* and *psychological* ‘function.' ”

From 1912 to 1940 Meyer was the director of the Henry Phipps Psychiatric Clinic and professor of psychiatry at the Johns Hopkins School of Medicine in Baltimore. As the dominant figure in American psychiatry, he developed his previous belief into the concept of psychobiology, which regarded a psychiatric patient as a biological entity who experienced unique reactions to biological and social influences. When those reactions were comprehensively delineated by taking the patient's life history, they revealed (in the words of Leston Havens) a “broadening perception of the significant reality for understanding pathological human nature” and “the gradual admission of factors hitherto unrelated to mental illness into an understanding of it.”

In treating patients, Meyer stressed common-sense counseling, social service, psychotherapy administered in community clinics by psychiatrists and social workers, child-guidance clinics, and aftercare for discharged asylum and hospital patients. His work has become so much a part of contemporary American psychiatry that his influence is often overlooked.

National Mental Health Act and the National Institute for Mental Health During World War II the need for psychiatry was shown when Selective Service boards rejected more than a million men because of mental disorders. Brief psychotherapy was able to return many servicemen who had experienced psychological problems back to active duty. At the end of the war, many psychiatrists were imbued with therapeutic optimism, and there was a popular belief that environmental stress contributed to mental maladjustment and that purposeful human interventions could alter psychological outcomes.

That attitude led Congress in 1946 to pass the National Mental Health Act, which for the first time in American history provided generous financial support for both psychiatric education and research. It also founded the National Institute of Mental Health (NIMH) with the goals of conducting research into mental illness (especially schizophrenia) and training large numbers of mental health personnel to provide psychiatric care. The above events positioned the United States in the forefront of advances in psychiatry.

MIDDLE AND LATE TWENTIETH CENTURY

During the past five decades the history of psychiatry, especially in America, has been characterized by numerous and unprecedented advances and some failures.

Lobotomies During the late 1940s and early 1950s, under the influence of two neurologists—Egas Moniz (1874–1955) and Walter Freeman (1895–1972)—many intractable psychotic and obsessive-compulsive patients were treated by lobotomies, the destruction of the white matter of the frontal lobes of the brain. Some patients improved, but others suffered severe and irreversible personality deterioration. With the advent of the psychotropic drugs, the use of lobotomies declined markedly. In 1949 Moniz along with the Swiss neurophysiologist Walter Hess was awarded the Nobel Prize in Physiology and Medicine.

Child Psychiatry In 1945, following the growth of child guidance clinics, the American Association of Psychiatric Clinics for Children was founded. Its aims were

maintaining certain standards in the clinics and in the training of the child psychiatrists, who along with psychologists and social workers staffed the clinics. This led to the development of child psychiatry as a specialty that then separated from adult psychiatry with the formation of the American Academy of Child Psychiatry (1953) and the establishment, by the American Board of Psychiatry and Neurology, of child psychiatry as a board-certifiable medical specialty (1959).

Double-Blind Tests In double-blind tests a drug or a procedure and a placebo are compared in such a way that neither the patient nor the therapist involved in the treatment know which preparation is being administered. The tests were used in medicine in 1937 by the American physician Harry Gold and in psychiatry in 1954 by Mogens Schou (1918–) to evaluate lithium, and by John Hampson, David Rosenthal, and Jerome Frank to evaluate the effectiveness of mephensin in relieving psychiatric symptoms. Since then the tests have been used in psychiatry to evaluate the effectiveness of various drugs and procedures.

Psychotropic Drugs

Chlorpromazine In 1952 two French psychiatrists, Jean Delay (1907–1987) and Pierre Deniker (1917–1999), reported that a new drug, chlorpromazine (Thorazine), significantly tranquilized hospitalized patients with chronic schizophrenia without causing undue depression. Chlorpromazine was called a “neuroleptic” because it reduced, rather than paralyzed nervous activity; it was also called an “antipsychotic.”

In the United States double-blind tests conducted by the NIMH showed that the use of chlorpromazine and other newly developed antipsychotics shortened the hospitalizations of some acutely psychotic patients and helped some chronically psychotic patients to either live outside an asylum for the first time or better tolerate living in an asylum.

Benzodiazepines In 1960–1963 the American pharmaceutical company of Hoffman-La Roche marketed two benzodiazepine compounds, chlordiazepoxide (Librium) and then diazepam (Valium), which largely replaced meprobamate (Equanil, Miltown) as methylphenidate anxiolytic drugs in cases of nonpsychotic anxiety. Hoffman-La Roche and other companies then developed other anxiolytic benzodiazepines, including clonazepam (Klonopin) in 1975, lorazepam (Ativan), in 1977 and alprazolam (Xanax) in 1981. They became among the most widely prescribed drugs in America.

Tricyclic Drugs and Monoamine Oxidase Inhibitors After the 1957 reports of the Swiss psychiatrist Roland Kuhn (1912–) and the American psychiatrist Nathanial Kline (1916–1983) and his associates, two classes of drugs were found effective in the treatment of depressions: imipramine (Tofranil)-type drugs (the tricyclic antidepressants) and the monamine oxidase inhibitors (MAOIs). The MAOIs caused episodes of severe hypertensive crisis. British psychiatrist Barry Blackwell (1934–) showed that the episodes were related to eating certain foods, especially cheese containing tyramine, but the MAOIs could be used if those foods were omitted.

In the late 1950s and early 1960s American and English psychiatrists found that tricyclic drugs and MAOIs alleviate certain severe states of anxiety.

Selective Serotonin Reuptake Inhibitors These have been developed for the treatment of depression by different American pharmaceutical companies since 1988, and include fluoxetine (Prozac), paroxetine (Paxil), and sertraline (Zoloft). They have shown fewer adverse effects than the tricyclics and MAOIs, and they largely superseded the use of the older antidepressants and have been therapeutically successful in various recurrent mood disorders, including chronic depressive and dysthymic disorders. They have also been used for obsessive-compulsive disorders. The diagnostic advances of the last two editions of *Diagnostic and Statistical Manual of Mental Disorders* (DSM) have made the clinical indications for their use more precise, so that many cases of depression—once the province of psychiatrists—can now be adequately managed by primary care physicians.

Lithium In 1949 the Australian psychiatrist John Cade (1912–1980) wrote an article for the *Medical Journal of Australia* showing that lithium quieted a group of hospitalized patients with mania and that their mania recurred when lithium was withdrawn. In 1954 the Danish psychiatrist Mogens Schou (1918–) validated Cade's work in a double-blind study. However, many psychiatrists questioned lithium's therapeutic effectiveness and emphasized its toxicity, and not until 1970 did the Food and Drug Administration (FDA) approve the use of lithium in America. Over the past two decades the following drugs carbamazepine (Tegretol), valproate (Depakene), neurontin (Gabapentin), lamictal (Lamotrigne), and topamax (Topiramate) have been used to treat patients with mania who did not respond well to or were intolerant of lithium.

Psychostimulants In 1937 the American pediatrician Charles Bradley (1902–1979) reported in the *American Journal of Psychiatry* that treatment with benzedrine had improved the behavior of a group of children who had been hospitalized because of (what was then diagnosed as) “behavior disorders.” For decades afterward there were occasional reports of stimulants being used to treat disturbed behavior in children and adults. In the 1980s and 1990s with the development of the diagnosis of attention-deficit disorders in children and in adults, stimulants such as amphetamine and methylphenidate (Ritalin) began being regularly used to successfully treat some patients with these disorders.

Serotonin-Dopamine Antagonists Clozapine (Clozaril) was developed in the 1970s, but because of the risk of agranulocytosis, the FDA did not permit its use in the United States for years. In 1988 the American psychiatrist John Kane (1945—) showed that it was an efficacious antipsychotic medication, and in 1989 the FDA permitted its American marketing, along with safeguards of weekly white blood cell counts and weekly drug dispensations. Risperidone (Risperdal) was developed in 1984, and in the 10 years before the FDA permitted its marketing, it was tested in large populations and shown to be effective and safe in the treatment of psychoses. Since 1994 clozapine and risperidone have been successful in treating persons with schizophrenia who were refractory to other medications and in ameliorating positive and negative symptoms of schizophrenia (including delusions, hallucinations, blunting of affect, and avolition).

Olanzapine (Zyprexa) was approved for use by the FDA in 1996. Like clozapine and risperidone, it was effective against both positive and negative symptoms of schizophrenia. It has been found to be more effective than haloperidol in the treatment of schizoaffective disorder and in the treatment of general psychotic symptoms. It has a lower potential than haloperidol to evoke tardive dyskinesia. There are no published data on switching from clozapine or risperidone to olanzapine, and until data to the contrary are available it should not be assumed that olanzapine has better therapeutic effects than clozapine or rispiridone.

Chemical Theories of Mental Illness The clinical advances in psychopharmacology stimulated research in neurochemistry and the development of new concepts about the chemical causes of psychiatric diseases. Since successful psychopharmacological drugs decreased the brain chemical dopamine in patients with schizophrenia and increased the brain chemicals norepinephrine and serotonin in patients with depression, neuroscientists theorized that some cases of schizophrenia were caused by increased dopamine concentration and that two subtypes of depressive illness exist: one caused by norepinephrine deficiency and the other by serotonin deficiency. Although these theories are unproved and perplexing, they have continued to guide neuroscientists in their search for new drugs.

The concept that mental health is influenced by the fluctuations of body chemicals resembled in some ways the 2500-year-old Hippocratic concept that mental and physical health depends on the equilibrium of the four body humors.

Deinstitutionalization In the United States some of the influences for the deinstitutionalization of mental patients were burgeoning public criticisms of the snake-pit environments of state asylums; the 1949–1956 open-door policy of several English hospitals (meaning both unlocked doors and a higher discharge rate of hospital inmates); a reduction of asylum patients, which began in 1956 (for decades previously the asylum population had been increasing) and which was only partly caused by the use of the new antipsychotic drugs; and a 1958 presidential address to the American Psychiatric Association by Harry Solomon, asserting that the state asylums were bankrupt and should be liquidated.

In 1961 the Joint Commission on Mental Illness and Mental Health (an interdisciplinary organization representing 36 medical and paramedical organizations, including the American psychiatric, psychological, and medical associations) issued a report, Action for Mental Health, recommending that the large state asylums be converted to small regional centers for intensive patient evaluation and treatment and that community mental health clinics, general hospitals, and mental hospitals be established. In October 1963 President John F. Kennedy signed the Community Mental Health Centers Act, which directed that these recommendations be carried out.

Community Care Over the past three decades, as the number of patients treated in asylums declined (there was an 82 percent reduction in long-term asylum inmates in the period 1950–1989), other institutions for treatment appeared. Acutely ill patients are now mainly treated in the psychiatric units of general hospitals and sometimes in nonhospital settings of private homes or specially formed households. Most chronically ill patients, especially aged adults, are cared for in nursing homes and also in board-and-care homes, halfway houses (so called because, in the course of a patient's therapy, they were located between the hospital and full community living), supervised or semisupervised apartments, and other facilities (including partial hospital programs, psychosocial clubhouses, supported

employment, vocational rehabilitation, and supported education).

While many aspects of community care are problematic, fragmented, and in need of integration and while (for unknown reasons) the homeless population (which contains many mentally ill individuals) has been growing, these shortcomings (in the words of Gerald Grob) “should not be permitted to conceal the more important fact that a large proportion of severely and persistently mentally ill persons have [because of the programs and institutions of community care] made a more or less successful transition to community life.”

National Alliance for the Mentally Ill Although community care often increased contacts between mentally ill persons and their families, relatively little attention had been given to the needs of such families. In September 1979 the National Alliance for the Mentally Ill (NAMI) was founded, and it has since become (in the opinion of David A. Rochefort) “perhaps the most potent consumer lobbying force ever to manifest itself in the mental health arena.” It is a passionate advocate for research into the biological causes of insanity and for better care and support for both the insane and their families. NAMI also functions as a mutual-help organization.

Nosology In 1974 a task force (later called a “work group”) of the American Psychiatric Association (APA), chaired by Robert Spitzer, prepared the third edition of DSM (DSM-III); a revision, DSM-III-R, with Janet Williams as text editor, was published in 1987. Those two editions differed radically from the two previous editions by approaching psychiatric disorders in two ways. First, the objective clinical manifestations of a disorder were carefully delineated; with emphasis on symptoms and deemphasis on causes. Since this approach resembled that of Kraepelin it was called the “neo-Kraepelin revival.” Second, disorders were given new classifications and new names, which often conveyed both the precise attributes of the disorder and advances in knowledge. For example, “inadequate personality” was replaced by three personality disorders; narcissistic, borderline, and dependent; bulimia was first classified; advances in sexology necessitated new delineations of paraphilias and types of sexual dysfunctions; advances in child psychiatry necessitated describing and naming many new disorders of childhood; alcoholic disorders were classified into a number of new diagnostic categories, each with its list of specific diagnostic symptoms; from a mixture of cognitive and behavioral abnormalities a group of symptoms in adults and children were designated attention-deficit disorder while other behavioral symptoms were grouped as conduct disorder; and, as a result of the psychiatric symptoms suffered by American soldiers in the Vietnam War, the diagnosis of posttraumatic stress disorder was described (in DSM-II traumatic stress response had been subsumed under the rubric “adjustment disorder of adult life.”)

Although these two DSMs had some limitations (many nonpsychotic illnesses presented not as distinct symptoms but as fluctuating mixtures of anxiety and mood disturbances that did not run distinct courses), they were the first diagnostic manuals to be widely accepted both academically and clinically, and they became the diagnostic “Bible” to those who treated psychiatric patients.

From 1988 to 1994 a new APA task force, chaired by Allen Frances, with Harold Pincus as vice-chairperson and Michael First as text editor, worked on preparing the fourth edition of DSM (DSM-IV). Preparation of DSM-IV drew on the research on diagnosis that was generated in part by DSM-III and DSM-III-R and involved a comprehensive, systematic review of the relevant literature, analysis and reanalysis of the collected data, and field trials that compared alternative options and studied the possible impact of suggested changes. The resulting volume, published in May 1994, is the most up-to-date classification of mental disorders for health professionals. It continued the descriptive, atheoretical approach of the two previous DSMs. Some of its changes were discussions of differential diagnosis, course, prevalence, and other aspects of certain illnesses; elimination of the term “organic mental disorders,” because it incorrectly implied that other mental disorders do not have a biological component; deletion of passive-aggressive personality disorder and addition of borderline personality disorder; renaming multiple personality disorder “dissociative identity disorder”; and the addition of an appendix that noted the influence of culture and ethnicity on psychiatric illnesses and diagnosis.

Psychotherapies The aims and scope of psychotherapy have grown from the treatment and alleviation of symptoms in psychiatric patients to the management of a wide variety of personality distresses (such as unhappiness and unfulfillment, social malaise, and social isolation), with the aim of providing greater self-assertion and greater emotional fulfillment.

Types of Psychotherapy Because psychotherapists are often guided by theories that they then modify to meet the particular needs of the patients, the history of their work has been characterized by great diversity and the evolution of perhaps 100 types of psychotherapy.

PSYCHOANALYTIC THERAPIES Melanie Klein (1882–1960) and Anna Freud (1895–1982) formed two different ways of applying psychoanalysis to play therapy with children and different concepts of child development. Harry Stack Sullivan (1892–1949) formed the interpersonal theory, stressing the interpersonal reactions of persons, and coined the terms “participant observer” and “consensual validation” to describe the reactions between patient and therapist. Erik H. Erikson (1902–1994) developed concepts of adult identity, identity diffusion, and crisis. Donald Winnicott (1896–1971) observed transitional objects and transitional phenomena to which persons become deeply attached and was part of the Independent Group of Psychoanalysts in England. Jacques Lacan (1901–1981) emphasized language and the need to make contact with the prelanguage period in the unconscious and was influential in French psychiatry. Heinz Kohut (1913–1981) originated self-psychology and three new views of transference. James Masterson (1926–) and Otto Kernberg (1928–) developed two different concepts for diagnosing and treating borderline and narcissistic disorders.

BEHAVIOR THERAPY Behavior therapy originated in the experiments of the American psychologist John Watson (1878–1958) on behaviorism and the Russian physiologist Ivan Pavlov (1849–1936) on conditioned reflexes. It was then shaped by several trends, including Andrew Salter's (1914–) conditioned reflex therapy, B. F. Skinner's (1904–1990) operant conditioning, Joseph Wolpe's (1915–) psychotherapy by reciprocal inhibition, and trends that emphasized cognitive processes and their adaptive nature. Today behavior therapy consists of a blend of cognitive and behavioral approaches that are aimed at overcoming maladaptive behaviors.

OTHER THERAPIES Brief psychotherapy aims at producing insight and personality changes in a short time. Aaron Beck's (1921–) cognitive therapy aims at rearranging a person's maladaptive thoughts and actions. Carl Roger's (1902–1987) client-centered psychotherapy has patients take responsibility for their feelings and direct the course of their therapy. Fritz Perls's (1893–1970) Gestalt therapy holds that a neurotic patient is repressing one part of the whole (German *Gestalt*) and that, by therapeutically reliving a painful experience, the patient can restore wholeness. Eric Berne (1910–1970), in transactional analysis, postulates that a lifescript created in early childhood by parents shapes and inhibits a person's life and that inhibition can be overcome by deciphering that script. Existential psychotherapy, propounded by the German psychiatrist and philosopher Karl Jaspers (1883–1969), emphasized here-and-now interactions and feelings rather than past events and rational thinking. Milton Erickson (1901–1980) developed hypnotic psychotherapy, in which hypnosis was used to overcome patients' resistance to therapeutic changes.

The above therapies involve one patient; other therapies involve couples (marital and sex), families, and groups. In 1935 two American alcoholics founded a 12-step movement of group psychotherapy for the treatment of alcoholism (Alcoholics Anonymous), which has since evolved to cover a wide range of addictive or assumed-addictive behaviors (including food-related disorders, codependency, gambling, drug, and tobacco addiction) as well as the impact of addictive behavior on family members (Al-Anon), and on the teen-aged children of alcoholics (Alateen).

Evaluations of Psychotherapy Over the past two decades psychological treatments have been shown effective in patients with neurotic symptoms and moderate personality disorders. However, patients benefit to variable degrees; much remains to be learned about the type of psychotherapy and the characteristics of patient and therapist in each particular treatment.

Biofeedback Through the use of instruments, biofeedback provides persons with representations of involuntary or subthreshold physiological processes that they may then be able to control voluntarily by changing their behavior. Biofeedback has been used to treat various psychosomatic diseases. Because it can detect the physiological markers of a patient's denial, resistance, and other defense mechanisms, biofeedback may also be considered as an adjunct for most of the psychotherapies mentioned above.

Social Influences The practice of psychotherapy was influenced by changes in mental health laws and managed care.

MENTAL HEALTH LAWS The late 1960s saw radical reforms in mental health laws; they peaked in the 1970s and subsided by the mid-1980s, but they left changes. They emphasized the rights of the mentally ill to contest involuntary hospitalization and refuse treatments, so that it became, in the words of Paul Applebaum, “axiomatic that [psychiatric] services must be provided in a manner that minimizes intrusions on individual rights and maximizes recourse to independent review when infringements occur.” A shift of the insanity defense from “not guilty by reason of insanity” to “guilty but mentally ill,” allowed the courts to monitor an individual's posttrial psychiatric treatment. Two notable cases in which psychotherapists were sued by third parties who were involved with patients that they were treating—*Tarasoff,* 1976, and *Ramona*, 1994—alerted therapists to the risks posed to them by third parties and the need to protect themselves against these risks, sometimes by warning potential victims about the danger posed by patients they are treating.

MANAGED CARE The increasing costs of psychotherapy led to development of managed care, which curtails costs in various ways, including requiring that a treating therapist submit frequent, written progress reports to a managed care panel and sometimes limiting the number of psychotherapy sessions to 5 to 20 a year. Thus the confidential psychiatrist-patient relationship that excluded all third parties, and the therapeutic options of psychiatrists to treat patients in ways they deemed most appropriate (doing insight-oriented therapy that often required frequent sessions over a prolonged period of time), were constrained in the interests of cost-effectiveness. While one psychiatrist has suggested that some of his colleagues may be able to adapt to managed care by developing business expertise and creating "physician-run managed care programs that can bring medical ethics and patient care values to the market," many other psychiatrists, whose expertise (and professional success) has been in patient-centered therapy, face an uncertain and troubled future.

Histories of Psychiatry General histories of psychiatry were published by Nolan D. C. Lewis (1941), Gregory Zilboorg and George Henry (1941), Jerome Schneck (1960), Franz Alexander and Sheldon Selesnick (1966), Erwin Ackernecht (1959, 1968), and Michael Stone (1996). A history of psychiatry from the eighteenth century to the present was published by Edward Shorter in 1996. John Howells edited *World History of Psychiatry* (1975), consisting of accounts by 42 psychiatrists of the histories of psychiatry in most of the countries and regions of the world; Benjamin Wolman's *International Encyclopedia of Psychiatry, Psychology, Psychoanalysis, and Neurology* (1977) contained short, concise, biographies and histories; and Sharon Romm and Ron Friedman edited *History of Psychiatry* (1994), which offers scholarly histories of psychiatric theories, treatments, and institutions. There have been English translations of the works (in whole or in part) of Hippocrates, Galen, Paracelsus, Weyer, Pinel, Esquirol, Mesmer, Heinroth, Charcot, Griesinger, Kahlbaum, Meynert, and Kraepelin, along with their biographies and new biographies of Wagner-Jauregg by Magda Whitrow (1993), Charcot by Christopher Goetz, Michael Bonduelle, and Toby Gelfand (1995), and Dorothea Dix by Thomas Brown (1998). Ida Macalpine and Richard Hunter edited *Three Hundred Years of Psychiatry: 1535–1860* (1964), a sourcebook of selected English texts and translations of influential European texts that chronicles the growth of English psychiatry up to the time when it became an established branch of medicine. Chris Thompson edited *The Origins of Modern Psychiatry* (1987), containing important psychiatric texts from 1856 to 1938.

Special Areas of Psychiatry The past decades saw a proliferation of histories confined to particular areas, by authors from various disciplines: social, intellectual, and legal historians; sociologists; criminologists; anthropologists; historians of science and medicine; and psychiatrists and other mental health professionals. The histories that focus on countries in a historical period emphasize the intellectual, social, and cultural matrices from which ideas about mental illness and mental deviance have developed, for example, Bennett Simon's *Mind and Madness in Ancient Greece: The Classical Roots of Modern Psychiatry;* Michael W. Dols's *Majnun; The Madman in Medieval Islamic Society;* Michael MacDonald's *Mystical Bedlam: Madness, Anxiety, and Healing in 17th Century England;* Roy Porter's *Mind-Forg'd Manacles: A History of Madness in England from the Restoration to the Regency;* McGovern's *Masters of Madness: Social Origins of the American Psychiatric Profession;* Jan Goldstein's *Console and Classify: The French Psychiatric Profession in the 19th Century;* George Drinka's *The Birth of Neurosis: Myth, Malady, and the Victorians;* Janet Oppenheim's *Shattered Nerves: Doctors, Patients and Depression in Victorian England;* Tom Lutz's *American Nervousness, 1903: An Anecdotal History;* and Francis George Gosling's *Before Freud: Neurasthenia and the American Medical Community, 1870–1910*.

The above intellectual-cultural emphasis was also applied to the following histories of theories, treatments, and illnesses: Wesley Smith's *The Hippocratic Tradition;* Oswei Temkin's *Galenism: Rise and Decline of a Medical Philosophy, The Falling Sickness: A History of Epilepsy from the Greeks to the Beginnings of Modern Neurology*, and *Hippocrates in a World of Pagans and Christians;* Jose Pinero's *Historical Origins of the Concept of Neurosis;* Stanley Jackson's *Melancholia and Depression: From Hippocratic Times to Modern Times;* Adam Crabtree's *From Mesmer to Freud: Magnetic Sleep and the Roots of Psychological Healing;* Alan Gauld, *A History of Hypnotism;* and Léon Chertok and Isabelle Stengers, *A Critique of Psychoanalytic Reason: Hypnosis as a Scientific Problem from Lavoisêr to Lacan*. Three recent books, Sander Gilman, Helen King, Roy Porter, G. S. Rousseau, and Elaine Showalter, *Hysteria Beyond Freud* (1993), Mark Micale, *Approaching Hysteria: Disease and Its Interpretations* (1995), and Helen King, *Hippocrates' Woman: Reading the Female Body in Ancient Greece* (1998), have revolutionized the historiography of hysteria and ancient Greek ideas about the nature of womens' bodies.

German Berrios and Roy Porter edited *A History of Clinical Psychiatry: The Origin and History of Psychiatric Disorders* (1995), which delineates the history of major psychiatric disorders (dementia, stroke, Parkinson's disease, schizophrenia, depression, neuroses, and personality disorders) from both the clinical and the social-historical perspectives. Stanley Finger wrote *Origins of Neuroscience: A History of Explorations into Brain Function* (1993).

ANOREXIA NERVOSA Sonja van't Hof's *Anorexia Nervosa: The Historical and Cultural Specificity: Fallacious Theories and Tenacious Facts* (1994) gives a history of how—from 1860 to the 1940s—the illness came to be recognized as a clinical entity. She also discusses its epidemiology and contends that (contrary to some opinions) it has not increased.

FORENSIC PSYCHIATRY Daniel Robinson's *Wild Beasts & Idle Humours* (1996) gives a history of the insanity defense from antiquity to the present, and Applebaum's *Almost a Revolution* (1994) discusses the changes in American mental health laws in the past 25 years.

CHILD PSYCHIATRY Stone's "Child Psychiatry before the Twentieth Century" (1973) gives some of the opinions on the psychology of normal (and sometimes aberrant) children from the sixteenth through the nineteenth centuries, before child psychiatry became a clinical specialty.

In 1997 two histories of military psychiatry were published: Hans Binneveld, *From Shell Shock to Combat Stress. A Comparative History of Military Psychiatry* (translated from the Dutch by John O'Kane); and Anthony Babington, *Shell-Shock. A History of the Changing Attitudes to War Neurosis*.

HISTORIES OF ASYLUMS Studies of the changes in the characteristics of the asylums and mental institutions for the insane in nineteenth-century England and America have been published by Andrew Scull, *Museums of Madness: The Social Organization of Psychiatry in Nineteenth Century England, The Most Solitary of Afflictions: Madness and Society in Britain 1700–1900, Masters of Bedlam: The Transformation of the Mad-Doctoring Trade;* Gerald Grob, *Mental Institutions in America; Social Policy to 1875* and *Mental Institutions and Mental Illness and American Society, 1875–1940;* and David Rothman, *The Discovery of the Asylum*. These three authors disagreed about the meanings of the changes that they described. Book-length studies have been published on the following nineteenth and early twentieth century mental hospitals: Ticehurst private asylum and the York Retreat, both in England (Charlotte MacKenzie and Anne Digby); the Pennsylvania Hospital for the Insane in Philadelphia (Nancy Tomes); the New York State Lunatic Asylum at Utica and the Willard Asylum for the Chronic Insane (Ellen Dwyer); and the Eastern State Hospital in Williamsburg, Virginia (Norman Dain); the New York State Psychiatric Institute (Lawrence Kolb and Leon Roizen); and the Boston Psychopathic Hospital (Elizabeth Lunbeck).

VIEWS OF DEINSTITUTIONALIZATION Two different views on the causes and meaning of deinstitutionalization have been propounded, first by Scull, *Decarceration: Community Treatment and the Deviant—A Radical View;* and then by Grob, *From Asylum to Community: Mental Health Policy in Modern America* and *The Mad Among Us: A History of the Care of America's Mentally Ill,* and (in England) by Kathleen Jones, *Asylums and After; A Revised History of the Mental Health Services: From the Early 18th Century to the 1990s*.

ACCOUNTS OF FREUD, PSYCHOANALYSIS, AND PSYCHOANALYSTS Freud has become the most written about figure in the history of psychiatry. Following Peter Gay's comprehensive biography. *Freud: A Life for Our Time* (1988), there has been an unceasing outpouring of books on special aspects of Freud's life and work. Some of the more recent of these are Lisa Appignanesi and John Forrester, *Freud's Women* (1992) on Freud's relationships with women including family members, patients, disciples, and friends and how he developed his theories of sexual differences and female sexuality. Gilman's two 1993 books, *Freud, Race, and Gender* and *The Case of Sigmund Freud: Medicine and Identity at the Fin de Siècle,* offer new views on Freud's attitudes toward his Jewish origins and on the influences that shaped his identity as a physician. John Kerr, *A Most Dangerous Method: The Story of Jung, Freud, and Sabina Spielrein* (1993), presents new information on the conflict between Freud and Jung and on early psychoanalytic ideas and treatments; and Peter Newton's *Freud: From Youthful Dream to Mid-Life Crisis* (1995) describes how psychoanalysis originated during a series of developmental periods and transitions in Freud's early to middle years.

Reuben Fine wrote *History of Psychoanalysis* up to 1979, and Nathan Hale's two volumes—*Freud and the Americans* and *The Rise and Crisis of Psychoanalysis in the United States*—give a history of psychoanalysis in America from 1876 to 1985. Edith Kurzwell, in *The Freudians: A Comparative Perspective* (1989), compares the varying impacts of psychoanalysis on areas of culture, medicine, and psychiatry, in five countries—Germany, Austria, France, England, and the United States.

Book length biographies have been written about the following psychoanalysts: Alfred Adler (Paul Stepansky), Marie Bonoparte (Celia Bertin), Anna Freud (Elisabeth Young-Bruehl), Helene Deutsch (Paul Roazen), Erik Erikson (Lawrence Friedman), Ernest Jones (Vincent Brome), Otto Rank (James Lieberman), Carl Jung (Frank McLynn), Karen Horney (Bernard Paris), Harry Stack Sullivan (Helen Perry), Wilhelm Reich (Myron Sharaf), Sandor Ferenczi (Martin Stanton), Melanie Klein (Phyllis Groskurth), Henry A. Murray (Forest Robinson), Sandor Rado (Roazen and Bluma Swerdloff), and Donald Winnicott (Adam Phillips). Paul Stepansky compiled and edited *The Memoirs of Margaret S. Mahler* (1988), and Geoffrey Cocks edited *The Curve of Life: Correspondence of Heinz Kohut, 1923–1981* (1994). Elisabeth Roudinesco, in *Jacques Lacan and Co. A History of Psychoanalysis in France, 1925–1985* (1990), (an English translation of the second volume of Roudinesco's

1982–86, two-volume history of psychoanalysis in France from 1885–1985, entitled *La Bataille de Cent Ans*, “The Hundred Years Battle”), gives an account of Lacan and his work that touches on much of the culture and intellectual life of the century.

Today, while the influence of psychoanalysis on American psychiatry has declined—as shown by the creation of the last three DSMs—and while there continues to be an intense controversy about its value and validity, Micale and Porter have commented that: “Freud's place as one of the supreme makers of the 20th-century mind, alongside Darwin, Marx, and Einstein, remains secure.”

Accounts of Behavior Therapy Alan E. Kazdin wrote *History of Behavior Modification* (1978), and book-length biographies of Ivan Pavlov (B. P. Babkin), John Watson (Kerry Buckley), and B. F. Skinner (Daniel Bjork) have appeared.

Histories of Psychiatric Drugs Frank Ayd and Barry Blackwell edited *Discoveries in Biological Psychiatry* (1984), which contains reminiscences by the discoverers or those close to them of the main psychiatric drugs, and Ayd added the essay “The Early History of Modern Psychopharmacology” (1990). Robert Maxwell and Shohreh Eckhardt in the section on psychiatric drugs in *Drug Discovery: A Casebook and Analysis* (1990) provided detailed accounts of the scientific processes that revealed the therapeutic uses of chlorpromazine, haloperidol, imipramine, the MAOIs, lithium, and the benzodiazepines. Solomon Snyder's *Drugs and the Brain* (an updated 1996 edition of a book first published in 1986) provides a comprehensive survey of the major psychoactive drugs, how they act on the brain and affect behavior, and the history of how they came to be used. David Healy has published two important books: in *The Psychopharmacologists. Interviews by Dr. David Healy* (1996), several pioneer psychopharmacologists discuss, along with their particular discovery, aspects of the professionalization of their work, with the founding of new journals, institutes, and societies; in *The Antidepressant Era* (1997) Healy delineates the history of psychopharmacology from the discovery of chlorpromazine to the current uses of fluoxetine and sertraline, emphasizing what the antidepressants reveal both about the workings of the brain and the sociology of drug marketing.

The quarterly, multidisciplinary *Journal of the History of the Behavioral Sciences* has appeared since 1965. In March 1990 the quarterly journal *History of Psychiatry* appeared, the first English-language journal devoted entirely to the history of psychiatry, published in England, and edited by two English scholars, Berrios (1940–) and Porter (1946–). Since its inception it has been (as stated in its first issue) “completely international, both in terms of contributions and contributors,” and has published articles by varied contributors (including historians, historians of medicine and psychiatry, practitioners of medicine and psychiatry) that have provided important primary information on all aspects of illness.

In 1993 Micale (1957–) and Porter edited and published *Discovering the History of Psychiatry,* the first booklength collection of essays by leading authorities on the lives and work of the major historians of psychiatry and psychoanalysis. It also contained essays on other themes and topics in the history of psychiatry, including the origin of myths and the position of women. Dora Weiner in “‘Le geste de Pinel': The History of a Psychiatric Myth,” shows that the famous episode of Pinel freeing the French insane from their chains in 1793 was a myth created largely by Pinel's son and students; that the insane were first freed in 1797 by Pussin, an assistant to Pinel, and that the latter then approved and followed up on Pussin's action. Tomes, in “Feminist Histories of Psychiatry,” emphasizes the absence of women in the histories of psychiatry before the 1960s and then shows how the recent work of feminist historians has depicted Victorian women as psychiatric patients in and outside of asylums, madwomen in literature, scientific objects in male-authored textbooks of psychology and medicine; and as early practitioners of psychiatry and psychoanalysis. Those feminist historians have formed a new, and still growing, subspecialty in the history of psychiatry.

In the introduction Micale and Porter indicate the many diverse factors that have influenced writings on the history of psychiatry.

Psychiatric history may inform, instruct, and entertain. It may reflect the intellectual personality, the cultural background, or the emotional temperament of the historian. It may be informed by internal disciplinary dimensions of a clinical, theoretical, or professional nature. And it may be shaped externally by contemporary social movements or political exigencies. It may also embody cultural structures and processes of a deeper and longer-term nature. Psychiatric historiography has served purposes that are antiquarian, pedagogical, justificatory, and ideological.

Micale and Porter further state that a study of the methodology and epistomology of the history of psychiatry is much needed.

Future Histories of Psychiatry Bennett Simon (1933–), in a 1996 review of several recent histories of psychiatry, contends that there is no good one-volume history of psychiatry because the field is in a state of “great creative flux,” with so many varied and different histories that no one history can please all. Simon further believes that “we are also living ‘in a time of fallen heroes'—the idealizations that once allowed certain unified histories to be written have been analyzed, deconstructed, demythologized, and at times found to be based on more fiction than fact. These same idealizations also allowed a more unified field of psychiatry... My hope is that the kinds of history of psychiatry now being written and the attendant debates about that history may contribute to a renewal and revitalization of the currently ‘decentered' and at times demoralized field of psychiatry. Figures like Pinel and Charcot, and perhaps even Freud, may once again become centers of historical discourse, but this time for an audience aware of the pitfalls of idealization and the obfuscation to which idealization can be put. I believe we will begin to see a return to an appreciation of heroes and heroines in the past, but an appreciation more solidly grounded in an appreciation of the complexities and ambiguities within which they worked. History might help us thus see the new kinds of heroism required in the confusing and difficult state of psychiatry today, the ability to have vision and purpose and steadfastness but not be blind to the legitimate tensions and disagreements that are part and parcel of our enterprise.”

Some of the major individuals and events in psychiatry, past and present, are described in Table 55.1-1.

Table 55.1-1 Persons and Events in Psychiatry

SUGGESTED CROSS-REFERENCES

Many sections of this textbook contain historical information related to their specific subjects. Of particular interest are Section 4.1 on anthropology and psychiatry, Section 4.2 on sociology and psychiatry, Section 5.1 on epidemiology, Section 6.1 on psychoanalysis, Chapter 7 on other psychodynamic schools, Section 11.1 on the classification of mental disorders, and Section 29.2 on the psychiatric aspects of AIDS. Chapter 31 covers drugs used in psychiatry and other biological therapies. Section 33.1 gives an overview of child psychiatry, Section 52.1 discusses legal issues in psychiatry, Section 26.1 gives a history of psychosomatic medicine, Section 31.10 gives an evaluation of psychotherapy, and Section 53.2 discusses the future of psychiatry.

SECTION REFERENCES

Ackernecht EH: *A Short History of Psychiatry*. Hafner, New York, 1968.

Alexander FG, Selsnick ST: *The History of Psychiatry*. Harper & Row, New York, 1966.

Andrews J, Briggs A, Porter R, Tucker P, Waddington K: *The History of Bethlem*. Routledge, London and New York, 1997

Applebaum PS: *Almost a Revolution: Mental Health Law and the Limits of Change*. Oxford University Press, New York, 1994.

Ayd FJ Jr: The early history of modern psychopharmacology. Neuropsychopharmacology *5:*71, 1991.

Barton WE: *The History and Influence of the American Psychiatric Association*. American Psychiatric Press, Washington, DC, 1987.

Bernstein DM: The discovery of the child: A historical perspective on child and adolescent psychiatry. In Lewis M, editor: *Child and Adolescent Psychiatry. A Comprehensive Textbook*, ed 2, M Lewis, editor. Williams & Wilkins, Baltimore, 1996.

Chessick RD: *Great Ideas in Psychotherapy*. Aronson, New York, 1977.

Ducy C, Simon B: Ancient Greece and Rome. In *World History of Psychiatry,* JG Howell, editor. Bruner/Mazel, New York, 1975.

Ellenberger HF: *The Discovery of the Unconscious: The History and Evolution of Dynamic Psychiatry*. Basic Books, New York, 1970.

Fine R: *A History of Psychoanalysis*. Columbia University Press, New York, 1979.

Finger S: *Origins of Neuroscience: A History of Explorations into Brain Function*. Oxford University Press, New York, 1994.

Freeman MA, Trabin T: *Managed Behavioral Health Care: History, Models, Key Issues, and Future Course*. US Department of Health and Human Services, Public Health Service Substance Abuse and Mental Health Services, Rockville, MD, 1994.

*Friedman L: *Identity's Architect: A Biography of Erik H. Erikson*. Scribner, New York, 1999.

Gauld A: *A History of Hypnotism*. Cambridge University Press, Cambridge, 1992.

*Gay P: *Freud: A Life for Our Time*. Norton, New York, 1988.

Gilman SL, King H, Porter R, Rousseau GS, Showalter E: *Hysteria Beyond Freud*. University of California Press, Berkeley, 1993.

*Grob GN: *The Mad Among Us: A History of the Care of America's Mentally Ill*. Free Press, New York, 1994.

Grob GN: Psychiatry's Holy Grail: The search for the mechanisms of mental diseases. Bull Hist Med *72:*189, 1998.

Harms E: *Origins of Modern Psychiatry*. Thomas, Springfield, IL, 1967.

*Healy D: *The Antidepressant Era*. Harvard University Press, Cambridge, Mass, 1997.

Hunter R, Macalpine I, editors: *Three Hundred Years of Psychiatry, 1535–1860: A History Presented in Selected English Texts*. Oxford University Press, London, 1964.

*Jackson SW: *Melancholia and Depression: From Hippocratic Times to Modern Times*. Yale University Press, New Haven, 1986.

*Kandel ER: A new intellectual framework for psychiatry. Am J Psychiatry. *155:*457, 1999.

Kaptchuk TJ: Intentional ignorance: A history of blind assessment and placebo controls in medicine. Bull Hist Med *72:*389, 1998.

Kazedin AE: *History of Behavior Modification.* University Park Press, Baltimore, 1978.

McGovern CM: *Masters of Madness: Social Origins of the American Psychiatric Profession*. University Press of New England, Hanover, NH, 1985.

Micale MS: *Approaching Hysteria: Disease and Its Interpretations*. Princeton University Press, Princeton, 1995.

Micale MS, Porter R, editors: *Discovering the History of Psychiatry*. Oxford University Press, New York, 1994.

Porter R: *Mind-Forg'd Manacles. A History of Madness in England from the Restoration to the Regency*. Athlone Press, London, 1987.

Robinson DN: *Wild Beasts & Idle Humours: The Insanity Defense from Antiquity to the Present*. Harvard University Press, Cambridge, MA, 1996.

Romm S, Friedman RS, editors: History of psychiatry. Psychiatr Clin North America *17:*453, 1994.

Roth MS, editor: *Freud: Conflict and Culture.* Alfred A Knopf, New York, 1998. (The occasion for this book was the 1998 exhibition of Freud artifacts at the Library of Congress).

Sabashin M: Turning points in twentieth-century American psychiatry. Am J Psychiatry *147:*1267, 1990.

Scull A: *Decarceration: Community Treatment and the Deviant,* ed 2. Rutgers University Press, New Brunswick, NJ, 1984.

Scull A: *The Most Solitary of Afflictions: Madness and Society in Britain, 1700–1900*. Yale University Press, New Haven, CT, 1993.

Shorter E: *From Paralysis to Fatigue: A History of Psychosomatic Illness in the Modern Era*. Free Press, New York, 1992.

Shorter E: *A History of Psychiatry: From the Era of the Asylum to the Age of Prozac*. Wiley, New York, 1996.

Simon B: The history of psychiatry: An opportunity for self-reflection and interdisciplinary dialogue. Psychiatry *59:*336, winter 1996.

Snyder SH: *Drugs and the Brain*. Scientific American Library, New York, 1996.

Stone MH: Child psychiatry before the twentieth century. Int J Child *2:*264, 1973.

*Stone MH: *Healing the Mind: A History of Psychiatry from Antiquity to the Present*. Norton, New York, 1996.

*Szaluta J: *Psychohistory: Theory and Practice*. Lang, New York, 1999.

Zilboorg G: *A History of Medical Psychology*. Norton, New York, 1941.

55.3 THE FUTURE OF PSYCHIATRY

PETER D. KRAMER, M.D.

The course of American clinical psychiatry for the last century would have humbled any prognosticator. Every one of its major developments was surprising. First, the rise to dominance of Sigmund Freud's talking cure, followed by the splintering of psychoanalysis into competing schools. Then, the development at the height of psychoanalytic influence, of a biopsychosocial paradigm, the transformation of military psychiatry into the community mental health movement. Antihistamine-based antipsychotic and antidepressant medications; the lithium revolution; the resurgence, beginning when psychoanalytic influence was at its peak, of a Germanic diagnostic system, the acceptance of the therapeutic milieu as a standard for hospital care was followed by the large-scale deinstitutionalization of the chronically mentally ill, the influence of outcome studies, bringing popularity for simple cognitive and behavioral therapies, and the erosion of many forms of mental health care delivery under pressure from insurers, just as these same outcome studies had bolstered the case for psychotherapy.

The course of change in psychiatry is especially hard to predict because the field is so responsive to external influence. Freud turned to physics to give shape to his metapsychology, and subsequent theorists have borrowed from systems theory and cybernetics. Progress in the science of psychiatry has relied on advances in a variety of disciplines, from engineering to genetics. Meanwhile, the practice of psychiatry is reshaped constantly by the ambient social, political, and economic organization. The very substrate of psychiatry, the object it acts upon and intends to influence, has an erratic history. The *self* or *person* is a concept constantly reshaped by changes in technology, social organization, and spiritual belief. To foretell the course of psychiatry would entail knowing the future of humankind and its achievements.

Even contemporary trends are hard to gauge. For most of the twentieth century, the core discipline in psychiatry was psychoanalysis. Today, cutting-edge discoveries arise most often from biological psychiatry. Perhaps the profession is in the midst of a paradigm shift. In that case, psychotherapy may disappear while psychiatry merges with neurology to form a uniform medical specialty based on brain sciences. Alternatively, if current changes are incremental, psychiatry may continue for some time to embrace both physiological and psychological interventions.

PSYCHOPHARMACOLOGY

Perhaps the safest prediction as the century closes is that the coming decades will see a continued proliferation of biological therapies for disorders of brain and mind. Clinically, psychiatry has been transformed in recent years by the availability of more potent and more specific medications. Conditions long considered refractory are approached routinely with medication: clomipramine (Anafranil) and the selective serotonin reuptake inhibitors (SSRIs) often ameliorate obsessive-compulsive disorder and the negative symptoms of schizophrenia sometimes respond to clozapine (Clozaril) or other serotonin-dopamine antagonists. Medications such as the SSRIs are now mainstays in the treatment of conditions such as dysthymic disorder and the minor mood disorders, that were once the exclusive province of psychotherapy.

The scope of psychopharmacology has been further expanded by three concomitant developments: the incorporation into psychiatric practice of medications traditionally associated with other specialties (e.g., antiepileptic agents including valproate [Depakene] and carbamazepine [Tegretol]), the legitimation of medications previously out of favor (notably the psychostimulants); and a renewed interest in polypharmacy, especially the augmentation of antidepressant drugs.

Barring a series of therapeutic disasters, certain trends will continue. New medications will be introduced, especially medications specific to receptor subtypes or particular brain pathways or regions. Existing medications will find broadened application and the indications for psychopharmacology will be expanded. The pace may accelerate as advances in computing and molecular biology (especially techniques for characterizing receptors) facilitate drug development. These changes will be accompanied by the development of novel biological interventions, perhaps including transplantation of brain elements, introduction of psychoactive genetic fragments into adult cells, and innovations in the ways medications are delivered to parts of the brain. The flowering of psychobiological therapeutics may combine with advances in the basic sciences to affect every aspect of psychiatry, from therapeutics to metapsychology.

Diagnosis A new range of choices for clinicians will create a demand for biologic markers of illness and for ways to predict and monitor treatment success. In psychiatry as in the rest of medicine, genetic profiles may be used to predict liability to illness and responsiveness to specific interventions. Brain imaging may move into the clinical arena as an aid to guiding treatment—perhaps supplemented by chemical or electrophysiological tests or other biological or psychological indicators yet unknown. The field may yet see the secular analogue of the ontological lapsometer imagined by the novelist Walker Percy, a hand-held brain imager that could diagnose the state of the soul.

The increased scope of biological treatment will create pressures on the diagnostic system. In the latter half of this century, acceptance of lithium (Eskalith) for the prophylaxis of manic-depressive illness (bipolar I disorder) gave an important spur to the predominance of what Walter Menninger called the Linnean tradition of diagnosis, that is, division of illness into ever finer subcategories. Recent trends indicate a continued proliferation of diagnostic categories—perhaps to the point where computers become necessary diagnostic aids.

Many of the newer medications already available have broad-spectrum utility. The SSRIs are variously reported to have roles in the treatment of obsessive-compulsive disorder, depressive disorders, bulimia nervosa, impulsive aggression, premenstrual syndrome, and the personality disorders, to mention only a few of many indications. A tension has thus developed between the diagnostic system and the commonest tools of psychiatry. The medications seem to cry out for a different sort of taxonomic system, one based on the state of neuroreceptors or neural pathways.

Perhaps the discovery of biologic markers of the relevant brain states will give rise to a new axis of psychiatric diagnosis or a new grouping of syndromes. In the interim, theorists have speculated that medications typically affect functions or dimensions of mind (e.g., aggression) rather than illness, and one can imagine a taxonomy that would address the relevant functions or dimensions. The future may even include a return to a diagnostic tradition that is (in Menninger's terms) Hippocratic, expressing the clinician's assessment of the complex state of the individual patient—but from a psychopharmacologic perspective rather than, or in addition to, a psychotherapeutic one. Such a system might require that clinicians exercise the subtlety and attention to detail that once characterized the anamnesis preparatory to psychoanalytic treatment.

Prophyaxis New psychiatric tools may revitalize another dream, mental illness prevention. Clinicians are already debating whether to use medication prophylactically in defined situations (e.g., just after delivery by a mother with a history of postpartum psychosis). As the pathology of psychic trauma is further elucidated, much broader forms of medication-based prevention may become possible, such as interrupting the cascade of stress hormones produced in response to important stressors. The death of a parent predisposes a child to future depression, but perhaps the profession will find pharmacological interventions to avert that association. Such capabilities would evoke profound ethical challenges; is prevention of illness always humane or can it be dehumanizing?

Treatment Similar issues are bound to arise in response to the broadened scope of pharmacological treatment of existing conditions. Researchers have already turned their attention to *formes frustes* of psychiatric disorders—levels of mood disorder, autistic disorder, and attention-deficit disorders that would previously have been deemed subsyndromal. Emerging evidence indicates that available medications can mitigate subsyndromal states. At the same time, the field is witnessing a resurgence of Darwinism, in the guise of evolutionary psychology. In the Darwinian perspective, much of what otherwise would be labeled illness is reconceptualized as an adaptive response, perhaps to bygone challenges. More popularly, learning disabilities have been reframed as “learning differences”—what was once called “minimal brain dysfunction” is now deemed a normal variant. These trends will demand reconsideration of the definition of health and illness. For the moment, the tendency in psychiatry is to expand the realm of pathology, for example, by conceptualizing subsyndromal symptomatic depression as an illness. The field may have to choose between, on the one hand, recognizing an ever-expanding domain of illness and, on the other, acknowledging that psychiatric interventions can be applied

appropriately in the treatment of variant normal states.

Indeed, the field will most likely find itself facing the ethical questions raised by the ability to enhance normal functioning. Preliminary research indicates that SSRIs can affect the social behavior of people who suffer no psychiatric disorder whatsoever, perhaps in an adaptive fashion. And the greatly increased prescribing of methylphenidate (Ritalin) for underperforming schoolchildren has underlined the need to distinguish treating defect from heightening performance.

Since drug development largely targets neural receptors and pathways and since the receptors and pathways implicated in illness are presumed to underlie normal temperament, issues of what has been called cosmetic psychopharmacology (the use of medication to change normal but undesirable or unrewarded traits for normal and desired or rewarded ones) will likely become real and even common in the profession. The use of psychotherapy to improve the mood or social functioning of healthy people is widely accepted. As the scope of psychopharmacology becomes indistinguishable from that of psychotherapy, the profession will need to grapple with the use of medication for enhancement and with the distinction between such use and drug abuse. Issues of prevention, definition of illness, and enhancement may elicit a comprehensive conceptual resolution that will be part of the process of self-definition of the psychiatric profession.

TRENDS

Of course, progress in psychopharmacology will come in tandem with developments in basic and clinical research. These are even harder to foresee, but to the extent that psychiatry mimics the rest of medicine, researchers will search for causes of mental illness in such disciplines as genetics, infectious disease, endocrinology, and immunology.

Surprises may include a new overarching theory of mind and mental illness, such as Sigmund Freud elaborated at the turn of the twentieth century, but for the moment, the future seems to hold a varied collection of explanations. This eclecticism may result in further subspecialization within psychiatry or in the transfer to other disciplines of illnesses previously considered psychiatric, as occurred in the past with tertiary syphilis, the vitamin deficiencies, and endocrine disorders. Regarding subspecialization, recent years have seen increased interest in sex-related aspects of pharmacology, a development with vague parallels to the feminist psychotherapy movements of the 1960s. The future likely holds an increased interest in the role of gender, race, and ethnicity in optimal psychiatric care. At the same time, computer-assisted consultation, either by algorithm or with colleagues at distant sites, may sustain the viability of general psychiatric practice.

PSYCHOTHERAPY

Perhaps the greatest mystery is the future of psychotherapy. Particularly in the last half of the twentieth century, psychotherapies proliferated to include treatments based on behavioral and cognitive principles, strategic and systems theory approaches, and many others. Although outcome research is limited in its findings, it has gained in influence, so that although research rarely shows distinctions in efficacy between psychotherapies, treatments that are most easily “manualized” and tested have become increasingly popular. These include cognitive psychotherapy, the interpersonal approach of Gerald Klerman, and (most recently) dialectical-behavioral therapies.

Within the line of therapies that begins with classical psychoanalysis, the recent movement has been away from approaches that attempt to remedy unconscious conflict and toward those (e.g., self psychology and the intersubjective school) that attempt to remedy posited deficits in personality structure, in part by applying empathy as a diagnostic and therapeutic instrument. A series of therapies take into account the research on attachment pioneered by John Bowlby and others. These treatments are more in tune with current theories of mind and of child development than were their predecessors, and they hold forth the possibility of a conceptual and clinical integration of psychotherapeutic and psychopharmacologic approaches. Meanwhile, research of the most preliminary sort points to the possibility of documenting—and in the future perhaps quantifying or monitoring—the central physiological effects of psychotherapy.

Considering scientific and intellectual opportunity only, the future of psychotherapy in psychiatry might look bright. But this view ignores other trends. Psychotherapy has suffered a decline in reputation and status for various reasons including a harsh reassessment of the character and achievements of many of the founders of psychotherapy, a series of scandals involving undue intimacy and other abusive behaviors on the part of practitioners, and controversy regarding the efficacy and scientific integrity of the discipline. At the same time, psychotherapy within psychiatry has been vulnerable to changes in health care delivery systems, notably cost-cutting measures on the part of third-party payers and fourth-party reviewers of health care. The result is a reliance in all disciplines on brief and unambitious treatments, marginalization of psychotherapy within training programs, and widespread discouragement about the ability of psychiatry to maintain psychotherapy as a central element in its identity. One notable dissenter from this pessimism is Leston Havens, who argues that members of a growing and vigorous middle- and upper-middle class throughout the world will ask themselves, “What's for me?” a question best answered by psychotherapy.

This debate reveals the extent to which the future of psychiatry rests on social and economic forces external to the discipline. Even progress in psychopharmacology depends on continued financial opportunities for drug manufacturers, a circumstance that was briefly called into doubt during the health care policy debates of the early 1990s. Other areas of the field, such as public-sector psychiatry, including community mental health care, are particularly at the mercy of shifts in public policy.

Despite multiple sources of uncertainty, a picture of a possible future psychiatry emerges. It is more closely allied with the rest of medicine through the basic sciences and biologically based treatments. Yet, it maintains a distinctive character, both because of its particular requirements for Hippocratic diagnosis and its ongoing use of a distinctive technology, psychotherapy (whether or not primarily administered by psychiatrists). It continues to comprehend tensions between mind and brain, while continuing its theoretical work to minimize that division. It remains intimately tied to a host of human interests, from its sister sciences to philosophy to the social forces it lives amidst. There is even the possibility that psychiatry, with its insistent dialectical attention to mind and brain, will become a model for the rest of medicine, so that psychiatry will not so much respond to its environment as reshape it.

SUGGESTED CROSS-REFERENCES

Biological therapies are covered in Chapter 31. Psychoanalysis and other psychological therapies are covered in Chapter 30. Evolutionary biology is discussed in Section 4.3.

SECTION REFERENCES

Bloom JD: Psychiatry: Three models in search of a future Psychiatr Serv *47:*874, 1996.

*Charlton BG: Psychopharmacology and the human condition. J Royal Soc Med *91:*599, 1998.

*Detre T, McDonald MC: Managed care and the future of psychiatry. Arch Gen Psych *54:*201, 1997.

*Eisenberg L: Past, present and future of psychiatry: Personal reflections. Can J Psychiatry *42:*705, 1997.

*Huang MP, Alessi NE: The internet and the future of psychiatry. Am J Psychiatry *153:*861, 1996.

*Kandel ER: Biology and the future of psychoanalysis: A new intellectual framework for psychiatry revisited. Am J Psychiatry *156:*505, 1999.

Kandel ER: A new intellectual framework for psychiatry Am J Psychiatry *155:*457, 1998.